

**PROPUESTA PARA MEJORAR LA PRECISIÓN EN EL PRONÓSTICO DE LOS
RETORNOS DE TIPO DE CAMBIO MEDIANTE UN MODELO HÍBRIDO QUE
INCLUYE VECTORES AUTORREGRESIVOS Y REDES NEURONALES**

Tesis de Grado presentado por

Alejandro Javier Parot Alvarado

como requisito parcial para optar al título de

Ingeniero Civil Industrial

y al grado de

Magíster en Ciencias de la Ingeniería Industrial

Profesor Referente:

Dr. Werner Kristjanpoller

Profesor Coreferente Interno:

Dr. Javier Scavia

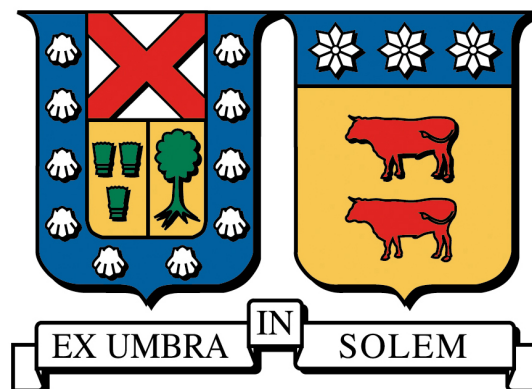
Profesor Coreferente Externo:

Dr. Marcelo Villena

—

AGOSTO 2017

UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA
DEPARTAMENTO DE INDUSTRIAS



**PROPUESTA PARA MEJORAR LA PRECISIÓN EN EL PRONÓSTICO DE LOS
RETORNOS DE TIPO DE CAMBIO MEDIANTE UN MODELO HÍBRIDO QUE
INCLUYE VECTORES AUTORREGRESIVOS Y REDES NEURONALES**

Tesis de Grado presentado por

Alejandro Javier Parot Alvarado

como requisito parcial para optar al título de

Ingeniero Civil Industrial

y al grado de

Magíster en Ciencias de la Ingeniería Industrial

Profesor Referente:

Dr. Werner Kristjanpoller

Profesor Coreferente Interno:

Dr. Javier Scavia

Profesor Coreferente Externo:

Dr. Marcelo Villena

—

VALPARAÍSO, AGOSTO 2017

TITULO DE LA TESIS:

**PROPUESTA PARA MEJORAR LA PRECISIÓN EN EL PRONÓSTICO DE LOS
RETORNOS DE TIPO DE CAMBIO MEDIANTE UN MODELO HÍBRIDO QUE
INCLUYE VECTORES AUTORREGRESIVOS Y REDES NEURONALES**

AUTOR:

Alejandro Javier Parot Alvarado

TRABAJO DE TESIS, presentado en cumplimiento parcial de los requisitos para el Grado de Magíster en Ciencias de la Ingeniería Industrial y de Ingeniero Civil Industrial de la Universidad Técnica Federico Santa María.

Dr. Werner Kristjanpoller

Dr. Javier Scavia

Dr. Marcelo Villena

VALPARAÍSO, CHILE. AGOSTO 2017

AGRADECIMIENTOS

Primeramente, agradezco a mi familia y polola por el cariño y apoyo incondicional que me han brindado durante mi paso por la universidad y a lo largo de mi vida. En especial a mi mamá, por formarme como persona, entregarme todo su amor y estar conmigo siempre. También, a mis amigos y compañeros del magíster, sin los cuales podría haber terminado esta tesis mucho antes pero no hubiese sido un proceso tan satisfactorio. Finalmente, agradezco a mi profesor guía y a la Universidad Técnica Federico Santa María, por brindarme la oportunidad de formarme profesionalmente y crecer como persona.



RESUMEN EJECUTIVO

La metodología propuesta muestra que efectivamente es posible mejorar la precisión en el pronóstico de los retornos del tipo de cambio EUR/USD al componer modelos autorregresivos y redes neuronales en un modelo híbrido que analiza los residuales del modelo lineal para modelar la componente no lineal de corto y largo plazo entre las series.

En general, el mejor pronóstico de ANN está dado por el tamaño de ventana de 252 días, mientras que los peores pronósticos se encuentran para la ventana de 42 días. Más aún, los mejores pronósticos se concentran en el caso de 2 capas ocultas para la ventana de 252 días donde se encuentra la mayor cantidad de variaciones negativas del MSE. En particular, el modelo VEC-ANN con una configuración de 2 capas ocultas y 20 neuronas por capa para una ventana corrediza de 252 días mejora el MSE del mejor modelo VAR-VECM un 26,883 % y las pruebas DM y MCS confirman estadísticamente que dicho modelo tiene mayor capacidad predictiva que el modelo de referencia y que todos los modelos en estudio, respectivamente. Además, puede verse durante el análisis que para todos los tamaños de ventana corrediza, a medida que se aumenta el número de capas ocultas los pronósticos se vuelven más planos, es decir, se concentran dentro de cotas mucho más pequeñas. Por otro lado, al aumentar el número de neuronas en cada capa oculta los pronósticos tienden a extender levemente las cotas que restringe el aumento del número de capas ocultas.

Esta metodología tiene la ventaja de ser simple y bastante certera para mejorar los pronósticos. Los resultados son importantes en el contexto de los mercados financieros donde manejar mayor información es fundamental para obtener beneficios. Además, es una nueva metodología propuesta para aportar al desarrollo de procesos de investigación científica.

Palabras Clave: Tipo de cambio, VAR, VECM, Cointegración, Red Neuronal Artificial.

ABSTRACT

Este estudio propone el uso de un modelo híbrido para mejorar la precisión del pronóstico del retorno de tipo de cambio EUR/USD para un día hacia adelante. Específicamente, el uso de diferentes configuraciones de Red Neuronal Artificial (ANN) junto a dos modelos econométricos, VAR y VECM. Este modelo híbrido es capaz de capturar la relación no lineal de corto y largo plazo entre las series. La motivación para usar modelos híbridos recae en el hecho de que la integración de diferentes modelos logra recoger más información y mejorar la habilidad predictiva en comparación a cada modelo por separado. En particular, la ANN modela la relación no lineal de corto y largo plazo que los modelos econométricos no pueden capturar (análisis residual). La capacidad predictiva de los modelos se compara de acuerdo al (MSE), como función de pérdida y la significancia se analiza usando las pruebas estadísticas de Diebold-Mariano (DM) y MCS. Los resultados muestran que el modelo VEC-ANN supera al econométrico.

Keywords: Tipo de cambio, VAR, VECM, Cointegración, Red Neuronal Artificial.

ÍNDICE DE CONTENIDOS

1. INTRODUCCIÓN	1
1.1. MOTIVACIÓN	1
1.2. OBJETIVOS	3
1.2.1. OBJETIVO PRINCIPAL	3
1.2.2. OBJETIVOS SECUNDARIOS	3
1.3. ALCANCE	3
2. MARCO TEÓRICO	7
2.1. Estudios previos	7
2.2. Mercados Financieros	10
2.2.1. Mercado de divisas	10
2.3. Series de tiempo financieras	11
2.4. Estacionariedad en las series de tiempo	12
2.4.1. Prueba de raíz unitaria	15
2.5. Modelos Econométricos	16
2.5.1. Modelos Autorregresivos	17
2.5.1.1. Autorregresivo de orden 1	17
2.5.1.2. Autorregresivo de orden 2	20
2.5.1.3. Autorregresivo de orden p	20
2.5.2. Vector Autorregresivo	21
2.5.2.1. Función de verosimilitud condicional	26
2.5.2.2. Causalidad de Granger	29
2.5.2.3. Vector Autorregresivo con raíces unitarias	31
2.5.3. Cointegración	32
2.5.3.1. Caracterización general del vector de cointegración	33
2.5.3.2. Implicancias de la Cointegración en el VAR	34
2.5.3.3. Representación del VECM	35
2.6. Redes Neuronales	37
2.6.1. Estructura de una Neurona	38
2.6.2. Estructura de una ANN	40
2.6.3. Métodos de aprendizaje	42
2.6.3.1. Aprendizaje No Supervisado	42
2.6.3.2. Aprendizaje Supervisado	42
2.6.4. Representación de la ANN	42
2.6.5. Backpropagation	46
2.7. Medidas de desempeño	48
2.7.1. Criterio de Información de Akaike	48
2.7.2. Función de Pérdida	49
2.7.3. Pruebas estadísticas	50

2.7.3.1. Prueba Diebold-Mariano	50
2.7.3.2. Model Confidence Set	51
3. METODOLOGÍA	53
3.1. Datos	53
3.2. Modelo Propuesto	57
4. RESULTADOS Y DISCUSIÓN	65
5. CONCLUSIONES	81
Bibliografía	85
A. ANEXOS	89



ÍNDICE DE TABLAS

4.1. Estadístico t y Valor-p para cada serie en la prueba ADF. Fuente: Elaboración propia.	66
4.2. Cantidad de ventanas que presentan Cointegración para cada tamaño de ventana corrediza. Fuente: Elaboración propia	66
4.3. MSE del modelo VAR-VECM en cada tamaño de ventana (días). Fuente: Elaboración propia.	70
4.4. Variación del MSE entregado por el modelo VEC-ANN de 1 capa oculta respecto del modelo VAR-VECM. Fuente: Elaboración propia.	73
4.5. Variación del MSE entregado por el modelo VEC-ANN de 2 capas ocultas respecto del modelo VAR-VECM. Fuente: Elaboración propia.	74
4.6. Variación del MSE entregado por el modelo VEC-ANN de 3 capas ocultas respecto del modelo VAR-VECM. Fuente: Elaboración propia.	75
4.7. Variación del MSE entregado por el modelo VEC-ANN de 4 capas ocultas respecto del modelo VAR-VECM. Fuente: Elaboración propia.	76
4.8. Valor-p para cada modelo de 1 capa oculta en la prueba MCS.	79
4.9. Valor-p para cada modelo de 2 capas ocultas en la prueba MCS. Fuente: Elaboración propia	80
4.10. Valor-p para cada modelo de 3 capas ocultas en la prueba MCS.	80
4.11. Valor-p para cada modelo de 4 capas ocultas en la prueba MCS.	80



ÍNDICE DE FIGURAS

2.1.	Estructura de una Neurona Artificial. Fuente: Elaboración propia.	39
2.2.	Estructura de una Red Neuronal Artificial. Fuente: Elaboración propia. . .	43
2.3.	Estructura general de una Red Neuronal Artificial Multicapas. Fuente: Elaboración propia.	45
3.1.	Evolución del tipo de cambio EUR/USD en el tiempo expresado en días. Fuente: Elaboración propia.	54
3.2.	Evolución del tipo de cambio GBP/USD en el tiempo expresado en días. Fuente: Elaboración propia.	54
3.3.	Evolución del tipo de cambio JPY/USD en el tiempo expresado en días. Fuente: Elaboración propia.	55
3.4.	Evolución del retorno del tipo de cambio EUR/USD en el tiempo expresado en días. Fuente: Elaboración propia.	56
3.5.	Evolución del retorno del tipo de cambio GBP/USD en el tiempo expresado en días. Fuente: Elaboración propia.	56
3.6.	Evolución del retorno del tipo de cambio JPY/USD en el tiempo expresado en días. Fuente: Elaboración propia.	57
3.7.	Estructura de Red Neuronal Artificial Multicapas implementada. Fuente: Elaboración propia.	63
4.1.	Comparación entre la serie del retorno de tipo de cambio EUR/USD y su pronóstico generado por el modelo VAR-VECM en la ventana de 42 días. Fuente: Elaboración propia.	67
4.2.	Comparación entre la serie del retorno de tipo de cambio EUR/USD y su pronóstico generado por el modelo VAR-VECM en la ventana de 63 días. Fuente: Elaboración propia.	67
4.3.	Comparación entre la serie del retorno de tipo de cambio EUR/USD y su pronóstico generado por el modelo VAR-VECM en la ventana de 84 días. Fuente: Elaboración propia.	68
4.4.	Comparación entre la serie del retorno de tipo de cambio EUR/USD y su pronóstico generado por el modelo VAR-VECM en la ventana de 126 días. Fuente: Elaboración propia.	68
4.5.	Comparación entre la serie del retorno de tipo de cambio EUR/USD y su pronóstico generado por el modelo VAR-VECM en la ventana de 252 días. Fuente: Elaboración propia.	69
4.6.	Comparación entre la serie del retorno de tipo de cambio EUR/USD y su pronóstico generado por el modelo VAR-VECM en la ventana de 378 días. Fuente: Elaboración propia.	69

4.7. Residuales del pronóstico del modelo VAR-VECM con ventana corregida de 126 días para el retorno del EUR/USD. Fuente: Elaboración propia. 70

4.8. Pronóstico para la ventana de 252 días con 2 capas ocultas y 20 neuronas. Fuente: Elaboración propia. 77

4.9. Zoom entre los días 3100 y 3200 del pronóstico para la ventana de 252 días con 2 capas ocultas y 20 neuronas. Fuente: Elaboración propia. 78

A.1. Pronósticos del modelo VEC-ANN para el retorno del tipo de cambio EUR/USD con una ventana de 42 días, para 1 capa oculta con las configuraciones (a), (b), (c) y (d). Fuente: Elaboración propia. 90

A.2. Pronósticos del modelo VEC-ANN para el retorno del tipo de cambio EUR/USD con una ventana de 42 días, para 2 capas ocultas con las configuraciones (a), (b), (c) y (d). Fuente: Elaboración propia. 91

A.3. Pronósticos del modelo VEC-ANN para el retorno del tipo de cambio EUR/USD con una ventana de 42 días, para 3 capas ocultas con las configuraciones (a), (b), (c) y (d). Fuente: Elaboración propia. 92

A.4. Pronósticos del modelo VEC-ANN para el retorno del tipo de cambio EUR/USD con una ventana de 42 días, para 4 capas ocultas con las configuraciones (a), (b), (c) y (d). Fuente: Elaboración propia. 93

A.5. Pronósticos del modelo VEC-ANN para el retorno del tipo de cambio EUR/USD con una ventana de 63 días, para 1 capa oculta con las configuraciones (a), (b), (c) y (d). Fuente: Elaboración propia. 94

A.6. Pronósticos del modelo VEC-ANN para el retorno del tipo de cambio EUR/USD con una ventana de 63 días, para 2 capas ocultas con las configuraciones (a), (b), (c) y (d). Fuente: Elaboración propia. 95

A.7. Pronósticos del modelo VEC-ANN para el retorno del tipo de cambio EUR/USD con una ventana de 63 días, para 3 capas ocultas con las configuraciones (a), (b), (c) y (d). Fuente: Elaboración propia. 96

A.8. Pronósticos del modelo VEC-ANN para el retorno del tipo de cambio EUR/USD con una ventana de 63 días, para 4 capas ocultas con las configuraciones (a), (b), (c) y (d). Fuente: Elaboración propia. 97

A.9. Pronósticos del modelo VEC-ANN para el retorno del tipo de cambio EUR/USD con una ventana de 84 días, para 1 capa oculta con las configuraciones (a), (b), (c) y (d). Fuente: Elaboración propia. 98

A.10. Pronósticos del modelo VEC-ANN para el retorno del tipo de cambio EUR/USD con una ventana de 84 días, para 2 capas ocultas con las configuraciones (a), (b), (c) y (d). Fuente: Elaboración propia. 99

A.11. Pronósticos del modelo VEC-ANN para el retorno del tipo de cambio EUR/USD con una ventana de 84 días, para 3 capas ocultas con las configuraciones (a), (b), (c) y (d). Fuente: Elaboración propia. 100

A.12. Pronósticos del modelo VEC-ANN para el retorno del tipo de cambio EUR/USD con una ventana de 84 días, para 4 capas ocultas con las configuraciones (a), (b), (c) y (d). Fuente: Elaboración propia. 101

A.13. Pronósticos del modelo VEC-ANN para el retorno del tipo de cambio EUR/USD con una ventana de 126 días, para 1 capa oculta con las configuraciones (a), (b), (c) y (d). Fuente: Elaboración propia. 102

A.14. Pronósticos del modelo VEC-ANN para el retorno del tipo de cambio EUR/USD con una ventana de 126 días, para 2 capas ocultas con las configuraciones (a), (b), (c) y (d). Fuente: Elaboración propia. 103

A.15. Pronósticos del modelo VEC-ANN para el retorno del tipo de cambio EUR/USD con una ventana de 126 días, para 3 capas ocultas con las configuraciones (a), (b), (c) y (d). Fuente: Elaboración propia.	104
A.16. Pronósticos del modelo VEC-ANN para el retorno del tipo de cambio EUR/USD con una ventana de 126 días, para 4 capas ocultas con las configuraciones (a), (b), (c) y (d). Fuente: Elaboración propia.	105
A.17. Pronósticos del modelo VEC-ANN para el retorno del tipo de cambio EUR/USD con una ventana de 252 días, para 1 capa oculta con las configuraciones (a), (b), (c) y (d). Fuente: Elaboración propia.	106
A.18. Pronósticos del modelo VEC-ANN para el retorno del tipo de cambio EUR/USD con una ventana de 252 días, para 2 capas ocultas con las configuraciones (a), (b), (c) y (d). Fuente: Elaboración propia.	107
A.19. Pronósticos del modelo VEC-ANN para el retorno del tipo de cambio EUR/USD con una ventana de 252 días, para 3 capas ocultas con las configuraciones (a), (b), (c) y (d). Fuente: Elaboración propia.	108
A.20. Pronósticos del modelo VEC-ANN para el retorno del tipo de cambio EUR/USD con una ventana de 252 días, para 4 capas ocultas con las configuraciones (a), (b), (c) y (d). Fuente: Elaboración propia.	109
A.21. Pronósticos del modelo VEC-ANN para el retorno del tipo de cambio EUR/USD con una ventana de 378 días, para 1 capa oculta con las configuraciones (a), (b), (c) y (d). Fuente: Elaboración propia.	110
A.22. Pronósticos del modelo VEC-ANN para el retorno del tipo de cambio EUR/USD con una ventana de 378 días, para 2 capas ocultas con las configuraciones (a), (b), (c) y (d). Fuente: Elaboración propia.	111
A.23. Pronósticos del modelo VEC-ANN para el retorno del tipo de cambio EUR/USD con una ventana de 378 días, para 3 capas ocultas con las configuraciones (a), (b), (c) y (d). Fuente: Elaboración propia.	112
A.24. Pronósticos del modelo VEC-ANN para el retorno del tipo de cambio EUR/USD con una ventana de 378 días, para 4 capas ocultas con las configuraciones (a), (b), (c) y (d). Fuente: Elaboración propia.	113



CAPÍTULO 1

INTRODUCCIÓN

1.1. MOTIVACIÓN

El poder pronosticar de forma precisa el comportamiento del tipo de cambio, tanto en términos de retorno como de volatilidad, es un problema de importancia global que ha sido estudiado por diversos investigadores y expertos, especialmente en el sector financiero, donde conocer la evolución de una serie es fundamental para rescatar información, tomar decisiones y disminuir los riesgos de inversión.

Debido a la eficiencia que presentan los mercados financieros, es difícil conseguir buenos pronósticos a corto y largo plazo [1]. Además, en general, las series de tipo de cambio son series de tiempo no lineales, caóticas, no paramétricas y dinámicas [2], dificultando especialmente su predicción. Por lo tanto, con el paso del tiempo, han sido desarrolladas distintas técnicas y metodologías para pronosticar el tipo de cambio de una mejor manera que el modelo de caminata aleatoria. Los modelos económicos se han aplicado en gran medida para esta tarea, pero también han sido altamente criticados porque la mayoría de ellos son de tipo lineal y trabajan bajo supuestos que los restringen demasiado, lo cual puede implicar que se vean superados por el modelo de la caminata aleatoria al momento de realizar pronósticos de la serie y, por lo tanto, no representan la mejor opción para trabajar. Debido a esto, se ha impulsado el uso de técnicas de inteligencia artificial de tipo paramétrico, como las Redes Neuronales Artificiales (abreviadas como ANN, por las

siglas en inglés para Artificial Neural Networks), que pueden captar de mejor manera el comportamiento no lineal de las series de tiempo a través de sus propiedades de aprendizaje, adaptabilidad y entrenamiento, sin la necesidad de conocer previamente la estructura de los datos con los que se la alimenta, logrando posicionarse como una técnica que favorece en gran medida a los modelos enfocados en pronosticar ([3]; [4]; [5]; [6]).

A pesar de las discrepancias que existen entre diversos estudios acerca de la linealidad del tipo de cambio, la literatura relacionada muestra un gran interés en la utilización de modelos no lineales para analizar el comportamiento de esta serie de tiempo por sobre los modelos clásicos ([7]; [8]; [9]).

Debido a lo anterior, y con el objetivo de lograr pronósticos más precisos, el uso de híbridos se presenta como una gran alternativa. La motivación para usar este tipo de modelos recae en el hecho de que, teórica y empíricamente, la integración de diferentes modelos permite ocupar las características propias de cada uno de ellos en cadena, logrando reconocer distintos patrones, recoger más información y mejorar la habilidad predictiva en comparación a usarlos por separado.

Este estudio pretende verificar que existe una mejora en la precisión de los pronósticos del retorno de tipo de cambio al usar modelos híbridos en comparación al uso de un modelo econométrico, todo en el marco de un análisis de ventanas corredizas sobre los datos. Para lograr esto, la metodología que se propone primero implementa un modelo de Vectores Autorregresivos (VAR) usando una ventana corrediza diaria sobre los datos para determinar la relación a largo plazo entre las variables. Junto a lo anterior, se analiza la presencia de vectores de cointegración en cada una de las ventanas para implementar el modelo de Vectores de Corrección del Error (abreviado como VECM, por las siglas en inglés de Vector Error Correction Model) cuando corresponda para obtener el pronóstico de este modelo. De este modo, el pronóstico del modelo econométrico VAR-VECM en cada ventana se determina por el modelo VECM o VAR dependiendo de si es que existen vectores de cointegración, o no, respectivamente, resultando en un modelo más flexible

para ser utilizado como referencia. Luego, se calculan los errores del pronóstico del modelo econométrico en comparación al valor real al que se debió llegar, para unirlos a rezagos de la serie y alimentar en conjunto diferentes configuraciones de la ANN en otra ventana corrediza diaria para obtener los pronósticos finales del modelo híbrido de inteligencia artificial.

1.2. OBJETIVOS

1.2.1. OBJETIVO PRINCIPAL

El objetivo principal de este estudio es proponer una metodología para mejorar la precisión en los pronósticos del retorno de tipo de cambio mediante el uso de un modelo híbrido en comparación al uso de un modelo econométrico de referencia.

1.2.2. OBJETIVOS SECUNDARIOS

- Aportar a la literatura existente en relación a modelos enfocados en pronósticos mediante los resultados de la nueva metodología desarrollada.
- Sensibilizar el análisis para diferentes tamaños de ventana corrediza (42, 63, 84, 126, 252 y 378 días).
- Sensibilizar el análisis para diferentes configuraciones de la Red Neuronal Artificial, variando la cantidad de capas ocultas y cantidad de neuronas en cada una de ellas.
- Validar los resultados obtenidos mediante pruebas estadísticas disponibles en la literatura.

1.3. ALCANCE

La innovación de esta propuesta recae en el análisis por ventana corrediza diaria para aplicar vectores de cointegración y diferentes configuraciones de redes neuronales para

pronosticar el retorno del tipo de cambio EUR/USD, a partir del uso de las paridades EUR/USD, GBP/USD y JPY/USD en términos diarios desde el 4 de Enero de 1999 hasta el 30 de Diciembre de 2015. Específicamente, se analiza la presencia de vectores de cointegración en cada ventana para asignar un pronóstico a través del modelo VAR o VECM a cada una, según corresponda, y obtener un modelo econométrico VAR-VECM más flexible para usarlo como punto de comparación al tener los resultados del modelo propuesto. Además, se usa la sensibilización de ventana corrediza para alimentar la ANN con diferentes cantidades de datos y para examinar diversas configuraciones de esta en cada escenario.

Las principales diferencias con respecto a estudios previos que muestra la literatura son, por un lado, la alimentación de la ANN porque ellos, a pesar de trabajar con modelos híbridos que combinan modelos econométricos con redes neuronales artificiales, no desarrollan un análisis de ventanas corredizas, lo cual entrega una ventaja importante al modelo propuesto en cuanto a dinamismo, mejor capacidad de ajuste a los cambios y, además, le permite rescatar la evolución de la serie en el tiempo en lugar de considerar solo un marco referencial, esto es, el modelo propuesto captura la relación de corto y largo plazo entre las series. Como segundo elemento diferenciador, se presenta los datos con los que se alimenta la ANN, ya que ellos incluyen los residuales del modelo lineal y sus rezagos, entre otros elementos, permitiendo reconocer la componente no lineal que el modelo clásico no es capaz de procesar. Por otro lado, un valor agregado en este estudio es la sensibilización de diferentes configuraciones para la ANN con el objetivo de mostrar una variación empírica de los resultados al cambiar la cantidad de capas ocultas y la cantidad de neuronas en cada una. Esta metodología tiene la ventaja de ser simple y bastante certera para mejorar los pronósticos.

Esta nueva metodología propuesta para mejorar los pronósticos del retorno de tipo de cambio es de gran importancia para accionistas, compañías internacionales y el sector financiero en general, porque el éxito de un modelo de este tipo ayuda a ofrecer información adicional acerca del comportamiento del mercado con el fin de evitar pérdidas y disminuir la exposición a cambios en el tipo de cambio. Por otro lado, también es importante para

investigadores, porque se entrega la posibilidad de tener más herramientas analíticas para mejorar procesos investigativos y enriquecer la literatura enfocada en pronósticos.

Es importante notar que el enfoque de esta investigación es de corte ingenieril y no económico. La importancia radica en la capacidad que tiene la metodología de mejorar la precisión del pronóstico al usar modelos híbridos, y no hay un interés primordial por obtener el mejor pronóstico posible para el tipo de cambio, es decir, se intenta lograr la mejora en la precisión de los pronósticos del tipo de cambio mediante experimentos dentro de los parámetros, herramientas y componentes del modelo pero considerando que el tipo de cambio se puede pronosticar a partir de sus rezagos y sin considerar elementos económicos externos que pueden mejorar aún más el pronóstico.



CAPÍTULO 2

MARCO TEÓRICO

En este capítulo se analizan los modelos en los que se basa el presente estudio, partiendo en el punto 2.1 con una revisión extensa de la literatura para conocer estudios previos relacionados a los tópicos que se trabajan. Luego, en los puntos 2.2 y 2.3 se contextualiza lo que son los Mercados Financieros y las Series de Tiempo para introducir al punto 2.5 donde se muestran y definen teóricamente los Modelos Econométricos utilizados. Luego, en el punto 2.6 se explican las Redes Neuronales para finalizar con las medidas de desempeño en el punto 2.7.

2.1. Estudios previos

La literatura existente muestra la complejidad de pronosticar el tipo de cambio, especialmente después de que Meese & Rogoff [10] mostraron que la mayoría de los modelos lineales no tienen la capacidad de superar los pronósticos del modelo de caminata aleatoria para el tipo de cambio. Tiempo después, Álvarez-Díaz [11] hizo notar que el trabajo de Meese & Rogoff parte del supuesto de linealidad, lo cual no se muestra empíricamente, por lo que existe una justificación para probar nuevas herramientas de pronóstico que consideren una estructura no lineal del tipo de cambio. Así, entran en juego las Redes Neuronales Artificiales (ANN, por las siglas en inglés para Artificial Neural Network), cuya principal característica es la habilidad para resolver problemas no lineales a través del aprendizaje sobre los datos con los que se la alimenta, algo que otras herramientas no

pueden lograr. Empíricamente, las ANN son capaces de modelar con precisión funciones continuas y no lineales, sin la necesidad de conocer previamente la estructura de los datos que se le entregan [12]. Más aún, las series de tiempo financieras son de corte numérico y no requieren una manipulación previa para ser utilizadas de forma nominal como datos de entrada en la ANN. Esto implica que no existan problemas como la pérdida de información o el uso de métodos de conversión que alteren los resultados, entre otras cosas; por lo tanto, las ANN se sitúan especialmente bien en la aplicación de este tipo de series, tal como reporta Lam [13].

Las ANN han sido utilizadas como alternativas a los modelos clásicos de predicción para series de tiempo financieras en diversos estudios, como por ejemplo: Kuan & White [14] y Swanson & White [15], quienes mostraron las implicancias y utilidad de aplicar Redes Neuronales para pronosticar mercados financieros; Leigh et al. [16] y Zhang & Wu [2], quienes mostraron la aplicación exitosa de redes Neuronales para pronosticar el mercado accionario; Dhamija & Bhalla [17] que estudiaron la comparación de varias arquitecturas de ANN para pronosticar el tipo de cambio, encontrando que la aplicación es muy efectiva para ese propósito; Khashei et al. [18] que mostraron una mejora en los resultados predictivos al utilizar ANN luego de modelos autorregresivos, concluyendo que su modelo híbrido obtiene mejores resultados que los modelos ARIMA; Dunis & Huang [19] quienes aplicaron el modelo Multilayer Perceptron (MLP) de inteligencia artificial para pronosticar un periodo hacia adelante la paridad EUR/USD, usando términos autorregresivos como entrada a la Red, concluyendo que el MLP se desempeña de mejor forma que el modelo tradicional fuera de la muestra cuando se analizan los retornos anualizados; Galeshchuk [12] comparó el poder predictivo de la ANN en 3 paridades (EUR/USD, GBP/USD, y JPY/USD) en términos diarios, mensuales y trimestrales; además, Yao & Tan [6] mostraron las ventajas de usar Redes Neuronales para capturar los componentes no lineales del tipo de cambio.

La composición de modelos es una idea que ha ido en crecimiento desde la publicación de Bates & Granger [20] y, luego, con Clemen [21], quienes propusieron un análisis y com-

pilación bibliográfica al respecto. La motivación para unir modelos recae en la posibilidad de recoger, de cada uno, la habilidad de reconocer diferentes elementos y patrones de una serie, para luego combinar estas capacidades y mejorar el desempeño en la predicción de cada modelo por separado ([22]; [23]). Los modelos híbridos pueden ser homogéneos, es decir, compuestos de una sucesión de modelos con características similares; o heterogéneos, compuestos de modelos con características diferentes, como por ejemplo, modelos lineales y no lineales [24]. Claros ejemplos de la efectividad de usar modelos híbridos se muestran en los estudios de: Wedding & Cios [25], quienes describieron una metodología que combina un modelo ARIMA con Radial Basis Function Network; Tsaih et al. [26], que presentan un modelo híbrido que integra Redes Neuronales con técnicas basadas en sistemas de reglas para pronosticar con éxito el índice S&P 500; Bildirici et al. [27], quienes propusieron una familia de Redes Neuronales para componerlas con un modelo TAR-VEC para mejorar el modelamiento de cointegración no lineal; Fatima & Hussain [28], que hicieron una comparación entre modelos ARIMA, ARCH/GARCH y ANN, tanto en conjunto como de forma separada, concluyendo que el modelo híbrido tiene el mejor desempeño; Li et al. [29], que propusieron la aplicación de un modelo AR-GRNN para pronosticar series de tiempo financieras, mostrando que el modelo combinado es una manera más efectiva que cada modelo de forma individual; Khashei & Bijari [30], donde se propone un modelo híbrido, compuesto de un modelo ARIMA y Redes Neuronales, lo cual muestra mejores resultados en términos predictivos que la Red Neuronal por sí sola para 3 conjuntos de series de tiempo; Hajizadeh et al. [31] que propone un modelo híbrido ANN-PSO para mejorar la habilidad que tiene la familia de modelos GARCH en pronosticar la volatilidad del tipo de cambio EUR/USD; y en el estudio de Sarica et al. [32], que proponen un modelo híbrido para pronosticar series de tiempo usando AR-ANFIS y mostrando que el híbrido produce pronósticos certeros.

Con todos estos antecedentes, es posible notar el gran aporte que los modelos compuestos significan para las metodologías enfocadas en pronósticos, ya que la evidencia empírica muestra que, si son utilizados de manera eficiente, estas herramientas mejoran los pronósticos que cada modelo por separado puede ofrecer. Por lo tanto, uno de los

objetivos de este estudio es continuar contribuyendo a la literatura y al desarrollo de nuevas metodologías, basándose en lo que ya se ha hecho y lo que se ha expuesto previamente.

2.2. Mercados Financieros

Un mercado se define económicamente como el conjunto de mecanismos mediante los cuales los compradores y vendedores de un bien o servicio están en contacto para realizar el comercio. De este modo, el mercado financiero se define como un espacio que logra reunir y contactar a oferentes y demandantes dispuestos a realizar transacciones de instrumentos financieros. Este espacio no es siempre físico pero se caracteriza por reunir a todos los oferentes y así facilitar la búsqueda de instrumentos a los demandantes. Los instrumentos financieros que se transan, como acciones, opciones de compra, bonos, divisas, etc., deben ser definidos por un precio de acuerdo a la ley de oferta y demanda.

Las principales características de estos mercados son su: Amplitud, en cuanto a la gran cantidad de títulos financieros que se negocian dentro de ellos; Profundidad, en cuanto a las diversas eventualidades que cubren estos títulos financieros, como alzas y caídas en el precio de algún activo; Libertad, en cuanto a que no existen barreras de entrada o salida del mercado; Flexibilidad, en cuanto a los cambios en los precios de activos según los movimientos de la economía; y Transparencia, en cuanto a la información existente respecto a los precios de cada activo.

2.2.1. Mercado de divisas

El mercado de divisas es un mercado mundial y descentralizado en el que se negocian divisas con el objetivo de facilitar el flujo monetario proveniente del comercio internacional. Corresponde al mercado financiero más grande del mundo, llegando a mover un volumen diario de transacciones de alrededor de US\$ 5 mil millones.

En el mercado de divisas, las monedas se negocian en pares. Cada paridad constituye un producto individual y es tradicionalmente anotado como XXX/YYYY, donde YYY es

el código internacional de tres letras (Según ISO 4217) en el cual el precio de una unidad de XXX se expresa. Por ejemplo: EUR/USD es el precio del Euro (EUR) expresado en dólares americanos (USD).

Este mercado, aunque cuenta con acceso las 24 horas del día, en la práctica se ve limitado por el cierre de los fines de semana. La hora del día en la que se accede a este mercado tiene un impacto directo en la liquidez para operar en una o varias divisas, ya que los momentos en los que abren las principales bolsas del mundo (Londres, Nueva York y Tokio) son los que presentan una mayor liquidez y movimiento.

El hecho de no ser un mercado centralizado significa que no existe una sola cotización para las divisas que se negocian, ya que estas van a depender de los agentes que participan en él. Las fluctuaciones que se presentan en los tipos de cambio son causadas, generalmente, por flujos monetarios reales así como por las expectativas de cambios en ellos debido a los cambios en las variables económicas como el crecimiento del PIB, inflación, tipo de interés, los déficit o superávit comerciales, desempleo, entre otros. Fuera de un contexto relacionado a las variables económicas, la estabilidad política y la psicología de mercado son también factores importantes al momento de analizar fluctuaciones del tipo de cambio.

2.3. Series de tiempo financieras

En la mayoría de los fenómenos que se estudian a través del tiempo, se debe tomar en cuenta su componente dinámica para comprenderlos de manera general. Una herramienta útil para esto es el análisis de series de tiempo. Una serie de tiempo es una secuencia de datos indexados por su marca temporal. Estas series encuentran aplicación, y son estudiadas, en un gran número de disciplinas, como por ejemplo: ingeniería, sociología, economía, finanzas, medicina, entre otras. El propósito fundamental es encontrar modelos que permitan representar los datos y hacer inferencias del proceso en estudio, incluyendo su predicción.

En particular, para una serie de tiempo de tipo financiera, el análisis puede enfocarse

en examinar de forma teórica y práctica la valoración de cierto instrumento en el tiempo, como también los volúmenes transados, entre otros factores. Existen varios tipos de precios que se pueden analizar para resumir la información de la variación del instrumento en un día, como por ejemplo: los precios de apertura, cierre, el mínimo, el máximo y el precio promedio. Con esta información es posible realizar estudios técnicos, los cuales proporcionan pronósticos de tendencias (como alzas o bajas de precio), basándose en la inspección visual y búsqueda de patrones en los precios anteriores del instrumento.

Una característica importante de las series de tiempo a tener en consideración al momento de aplicar herramientas para analizarlas es la estacionariedad. Se dice que una serie de tiempo es estacionaria si su media, su varianza y su autocovarianza (en los diferentes rezagos) permanecen constantes sin importar el momento en el cual se midan, es decir, son invariantes respecto al tiempo. Por consiguiente, una serie de tiempo estacionaria no presenta tendencia de crecimiento o declinación prolongada. Por otro lado, una serie de tiempo no estacionaria tendrá una media que varía con el tiempo, o una varianza que varía con el tiempo, o ambas. En consecuencia, una serie de tiempo que tiene una tendencia de crecimiento o declinación en el tiempo se dice que no es estacionaria.

2.4. Estacionariedad en las series de tiempo

Una serie de tiempo se puede modelar de la forma [33]

$$y_t = \mu + \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \dots = \mu + \psi(L)\varepsilon_t \quad (2.1)$$

donde $\sum_{j=0}^{\infty} |\psi_j| < \infty$, las raíces de $\psi(z) = 0$ están fuera del círculo de radio 1, y ε_t es ruido blanco con media 0 y varianza σ^2 . Se pueden notar dos puntos importantes de estas series: Primero, la esperanza de la variable es constante e independiente del tiempo:

$$E(y_t) = \mu$$

Segundo, los pronósticos a futuro convergen a la media:

$$\lim_{s \rightarrow \infty} \hat{y}_{t+s|t} = \mu$$

Pero muchas de las series económicas y financieras presentan tendencias a través del tiempo que deben ser reflejadas en su modelamiento. Dos de las especificaciones para describir estas tendencias son: Primero, incluir una tendencia determinística temporal:

$$y_t = \alpha + \delta t + \psi(L)\varepsilon_t \quad (2.2)$$

Entonces, la media μ del proceso estacionario 2.1 se reemplaza por una función lineal del tiempo t . Estos procesos se describen como de tendencia estacionaria, porque si se elimina la tendencia $\alpha + \delta t$ de la ecuación 2.2, el resultado es un proceso estacionario.

La segunda especificación es un proceso de raíz unitaria,

$$(1 - L)y_t = \delta + \psi(L)\varepsilon_t \quad (2.3)$$

donde se define $\psi(L) \neq 0$, para eliminar la posibilidad de que la serie original, y_t , sea estacionaria. Además, $(1 - L)$ corresponde al operador de primera diferencia, denotado como Δ y que cumple:

$$\Delta y_t = y_t - y_{t-1}$$

El ejemplo más claro de un proceso de raíz unitaria es la caminata aleatoria que se obtiene al fijar $\psi(L) = 1$ en 2.3:

$$y_t = y_{t-1} + \delta + \varepsilon_t \quad (2.4)$$

Una representación alternativa de 2.3 es la siguiente especificación:

$$y_t = \alpha + \delta t + u_t \quad (2.5)$$

donde u_t sigue un proceso ARMA de media cero:

$$(1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p)u_t = (1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q)\varepsilon_t \quad (2.6)$$

y donde el operador de media móvil $(1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q)$ es invertible. Al factorizar el operador autorregresivo de 2.6 para obtener valores propios de la forma:

$$(1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p) = (1 - \lambda_1 L)(1 - \lambda_2 L)\dots(1 - \lambda_p L)$$

Si todos los valores propios $\lambda_1, \lambda_2, \dots, \lambda_p$ están dentro del círculo de radio unitario, entonces 2.5 puede expresarse como

$$u_t = \frac{1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q}{(1 - \lambda_1 L)(1 - \lambda_2 L)\dots(1 - \lambda_p L)}\varepsilon_t = \psi(L)\varepsilon_t$$

con $\sum_{j=0}^{\infty} |\psi_j| < \infty$ y las raíces de $\psi(z) = 0$ fuera del círculo de radio unitario. Entonces, cuando $|\lambda_i| < 1$ para todo i , el proceso 2.5 sería un caso especial del proceso de tendencia estacionaria 2.2. Si $\lambda_1 = 1$ y $|\lambda_i| < 1$ para $i = 2, 3, \dots, p$ entonces 2.6 especificaría que

$$u_t(1 - L) = \frac{1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q}{(1 - \lambda_2 L)(1 - \lambda_3 L)\dots(1 - \lambda_p L)}\varepsilon_t = \psi^*(L)\varepsilon_t \quad (2.7)$$

con $\sum_{j=0}^{\infty} |\psi_j^*| < \infty$ y las raíces de $\psi^*(z) = 0$ fuera del círculo de radio unitario. Entonces si se le aplica primera diferencia a 2.5, el resultado es

$$(1 - L)y_t = (1 - L)\alpha + [\delta t - \delta(t - 1)] + (1 - L)u_t = 0 + \delta + \psi^*(L)\varepsilon_t$$

que corresponde a la forma del proceso de raíz unitaria 2.3. Entonces, el término proceso de raíz unitaria, representado en 2.5, se refiere a que una de las raíces o valores propios (λ_i) del polinomio autorregresivo 2.6 es la unidad, mientras que el resto se mantienen dentro del círculo de radio unitario. Otra expresión para el proceso 2.3 es que es integrado de orden 1 ($y_t \sim I(1)$).

Si el proceso tiene 2 valores propios (λ_1 y λ_2) iguales a la unidad con los otros dentro

del círculo de radio unitario, entonces se debe tomar la segunda diferencia para lograr la estacionariedad en la serie de tiempo, lo que se denota como un proceso integrado de orden 2 ($y_t \sim I(2)$). En general, cuando se debe tomar d diferencias para llegar a la estacionariedad, el proceso se denomina como integrado de orden d ($y_t \sim I(d)$).

Por conveniencia, es común usar los datos en términos de su logaritmo natural antes de modelarlos según 2.5 o 2.6, ya que para pequeños cambios, la primera diferencia del logaritmo natural de una variable es aproximadamente igual al porcentaje de cambio en la variable:

$$\begin{aligned} (1 - L) \log(y_t) &= \log(y_t/y_{t-1}) \\ &= \log(1 + [(y_t - y_{t-1})/y_{t-1}]) \\ &\approx (y_t - y_{t-1})/y_{t-1} \end{aligned}$$

donde se usa el hecho de que para un x cercano a cero se cumple que $\log(1 + x) \approx x$. Entonces, si la primera diferencia del logaritmo natural de una variable sigue un proceso estacionario, se asume que la tasa de crecimiento de la serie es un proceso estacionario.

2.4.1. Prueba de raíz unitaria

Es importante revisar si una serie es estacionaria antes de usarla en una regresión, ya que los procedimientos estándar de inferencia no se aplican si existen regresores integrados. El método formal para evaluar estacionariedad es la prueba de raíz unitaria.

Para la prueba de raíz unitaria Dickey-Fuller (DF) [34], se considera la siguiente expresión:

$$\Delta y_t = \alpha y_{t-1} + \delta x_t + \varepsilon_t \quad (2.8)$$

Donde x_t es una variable exógena, δ es su coeficiente, y ε_t es error ruido blanco con media $E(\varepsilon_t) = 0$ y varianza σ^2 . Las hipótesis nula corresponde a la existencia de raíz unitaria, mientras que la hipótesis alternativa corresponde a que no existe evidencia de raíz unitaria.

Estas pueden ser expresadas de la forma $H_0 : \alpha = 0$ y $H_1 : \alpha < 0$, respectivamente. Para evaluar, se aplica el estadístico t_α :

$$t_\alpha = \frac{\hat{\alpha}}{SE(\hat{\alpha})} \quad (2.9)$$

Donde $\hat{\alpha}$ es el estimador por mínimos cuadrados ordinarios de α , y $SE(\hat{\alpha})$ es el error estándar asociado a α . Bajo la hipótesis nula de raíz unitaria, el estadístico no sigue una distribución t-student, sino una distribución asintótica. Esta prueba estadística es válida solo para series autorregresivas de 1 rezago, ya que si la serie tiene una mayor cantidad de rezagos se viola el supuesto de ruido blanco para ε_t .

La prueba de Dickey-Fuller Aumentada (ADF) [34] considera una corrección paramétrica para una serie autorregresiva con p rezagos de la forma:

$$\Delta y_t = \alpha y_{t-1} + \delta x_t + \sum_{i=1}^p \beta_i \Delta y_{t-i} + \varepsilon_t \quad (2.10)$$

Entonces esta especificación se aplica para las hipótesis mencionadas y el estadístico t_α .

2.5. Modelos Econométricos

La econometría es una rama de la economía que utiliza modelos matemáticos y estadísticos para analizar, interpretar y hacer predicciones sobre sistemas económicos, prediciendo variables como el precio de bienes y servicios, tasas de interés, tipos de cambio, las reacciones del mercado, el coste de producción, entre otros. A pesar de ser una rama de esta ciencia, la econometría no necesariamente requiere o presupone una teoría económica para desarrollar el análisis econométrico. De hecho, en algunos casos, considerarla puede ser una restricción que imposibilita obtener un análisis riguroso.

Un modelo econométrico se define a partir de las variables (endógenas, exógenas) que explican y determinan el modelo, los parámetros estructurales que acompañan a las variables, las ecuaciones y su formulación en forma matemática, la perturbación aleatoria

que explica la parte no sistemática del modelo, y los datos estadísticos. A partir del modelo econométrico especificado, en una segunda etapa se procede a la estimación, fase estadística que asigna valores numéricos a los parámetros de las ecuaciones del modelo. Para ello se utilizan métodos estadísticos como pueden ser: mínimos cuadrados ordinarios, máxima verosimilitud, etc. La tercera etapa en la elaboración del modelo es la verificación y contrastación, donde se someten los parámetros y la variable aleatoria a contrastes estadísticos para cuantificar en términos probabilísticos la validez del modelo estimado. La cuarta etapa consiste en la aplicación del modelo conforme al objetivo del mismo.

Los mayores problemas que se presentan son la escasez de datos, los sesgos que pueden presentar los datos existentes, y los sesgos del propio investigador. Aun así, los modelos econométricos son la única aproximación científica al entendimiento de muchos fenómenos.

2.5.1. Modelos Autorregresivos

Un modelo autorregresivo (AR) es una representación de un tipo de proceso aleatorio que describe ciertos procesos variables en el tiempo. El modelo autorregresivo especifica que la variable de salida depende linealmente de sus propios valores anteriores, llamados rezagos.

2.5.1.1. Autorregresivo de orden 1

Un modelo autorregresivo de orden 1, denotado como AR(1), satisface la siguiente ecuación de primera diferencia:

$$y_t = c + \phi y_{t-1} + \varepsilon_t \quad (2.11)$$

Donde c es una constante y ε_t corresponde al error ruido blanco. Del análisis para ecuaciones de primeras diferencias de [33], se sabe que si $|\phi| \geq 1$ entonces no existe un proceso estacionario para y_t que satisface 2.11. Por el contrario, si $|\phi| < 1$, sí existe un proceso estacionario para y_t que satisface la ecuación 2.11. La estacionariedad asegura que el proceso sea estable y mantenga su media y su varianza independiente del tiempo t en la

cual se mide. Asumiendo que el proceso es estacionario para y_t , se pueden obtener los momentos directamente de 2.11. Tomando el valor esperado de ambos lados de la ecuación,

$$E(y_t) = c + \phi E(y_{t-1}) + E(\varepsilon_t) \quad (2.12)$$

Como se está asumiendo que $|\phi| < 1$, entonces,

$$E(y_t) = E(y_{t-1}) = \mu \quad (2.13)$$

Sustituyendo 2.13 en 2.12 y considerando que ε_t es ruido blanco, se logra,

$$\mu = c + \phi\mu + 0$$

Que se puede reescribir para obtener la ecuación de la media de un AR(1),

$$\mu = c/(1 - \phi) \quad (2.14)$$

Para obtener el segundo momento, se puede reescribir 2.11 a partir de 2.14,

$$y_t = \mu(1 - \phi) + \phi y_{t-1} + \varepsilon_t$$

o también,

$$(y_t - \mu) = \phi(y_{t-1} - \mu) + \varepsilon_t \quad (2.15)$$

Al elevar al cuadrado ambos lados de 2.15 y aplicar el valor esperado,

$$E((y_t - \mu)^2) = \phi^2 E((y_{t-1} - \mu)^2) + 2\phi E((y_{t-1} - \mu)\varepsilon_t) + E(\varepsilon_t^2) \quad (2.16)$$

Si se tiene en cuenta que, al aplicar el operador de rezagos sobre 2.11, se tiene la siguiente solución estable:

$$y_t = c/(1 - \phi) + \varepsilon_t + \phi\varepsilon_{t-1} + \phi^2\varepsilon_{t-2} + \phi^3\varepsilon_{t-3} + \dots \quad (2.17)$$

Entonces, por extensión, el término $(y_{t-1} - \mu)$ corresponde a una función lineal de los errores $\varepsilon_{t-1}, \varepsilon_{t-2}, \varepsilon_{t-3}, \dots$:

$$(y_{t-1} - \mu) = \varepsilon_{t-1} + \phi\varepsilon_{t-2} + \phi^2\varepsilon_{t-3} + \dots$$

Pero como ε_t no está correlacionado con $\varepsilon_{t-1}, \varepsilon_{t-2}, \varepsilon_{t-3}, \dots$, entonces tampoco lo está con $(y_{t-1} - \mu)$ y el segundo término del lado derecho de 2.16 se hace 0,

$$2\phi E((y_{t-1} - \mu)\varepsilon_t) = 0 \quad (2.18)$$

Nuevamente, como se asume que $|\phi| < 1$,

$$E((y_t - \mu)^2) = E((y_{t-1} - \mu)^2) = \gamma_0 \quad (2.19)$$

Reemplazando 2.18 y 2.19 en 2.16,

$$\gamma_0 = \phi^2\gamma_0 + 0 + \sigma^2$$

Que se puede reescribir para tener la ecuación de la varianza de un AR(1),

$$\gamma_0 = \sigma^2 / (1 - \phi^2) \quad (2.20)$$

De manera similar, al multiplicar $(y_{t-j} - \mu)$ a ambos lados de 2.15 y aplicar valor esperado,

$$E((y_t - \mu)(y_{t-j} - \mu)) = \phi E((y_{t-1} - \mu)(y_{t-j} - \mu)) + E((y_{t-j} - \mu)\varepsilon_t) \quad (2.21)$$

Pero, por extensión de 2.17, el término $(y_{t-j} - \mu)$ es función lineal de $\varepsilon_{t-j}, \varepsilon_{t-j-1}, \varepsilon_{t-j-2}, \dots$, cuando $j > 0$, por lo que no está correlacionado a ε_t y el último término del lado derecho de 2.21 se hace 0. Notando que el primer término del lado derecho de 2.21 es la autocovarianza de observaciones de y separadas en $j - 1$ periodos,

$$E((y_{t-1} - \mu)(y_{t-j} - \mu)) = \gamma_{j-1}$$

Entonces, cuando $j > 0$, 2.21 se transforma en la autocovarianza de orden j :

$$\gamma_j = \phi\gamma_{j-1} \quad (2.22)$$

2.5.1.2. Autorregresivo de orden 2

Un modelo autorregresivo de orden 2, denotado como AR(2), satisface

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t \quad (2.23)$$

Siguiendo un procedimiento similar al usado en el AR(1) [33], la ecuación de la media para un AR(2) se escribe como:

$$\mu = c/(1 - \phi_1 - \phi_2) \quad (2.24)$$

La ecuación de la varianza para AR(2) se escribe como:

$$\gamma_0 = \frac{(1 - \phi_2)\sigma^2}{(1 + \phi_2)((1 - \phi_2)^2 - \phi_1^2)} \quad (2.25)$$

Y la autocovarianza de orden j , para $j > 0$ se expresa como

$$\gamma_j = \phi_1 \gamma_{j-1} + \phi_2 \gamma_{j-2} \quad (2.26)$$

2.5.1.3. Autorregresivo de orden p

Por extensión, un modelo autorregresivo de orden p , denotado por AR(p), se puede representar mediante la ecuación [33]

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t \quad (2.27)$$

Siguiendo un procedimiento similar al usado en el AR(1), la ecuación de la media para un AR(p) se escribe como:

$$\mu = c/(1 - \phi_1 - \phi_2 - \dots - \phi_p) \quad (2.28)$$

Y las autocovarianzas para $AR(p)$ se escriben como:

$$\gamma_j = \begin{cases} \phi_1\gamma_{j-1} + \phi_2\gamma_{j-2} + \dots + \phi_p\gamma_{j-p} & \text{para } j = 1, 2, \dots \\ \phi_1\gamma_1 + \phi_2\gamma_2 + \dots + \phi_p\gamma_p + \sigma^2 & \text{para } j = 0 \end{cases} \quad (2.29)$$

2.5.2. Vector Autorregresivo

El modelo Vector Autorregresivo (VAR, por las siglas en inglés para Vector Autoregressive) es una extensión natural de los modelos autorregresivos univariados para series de tiempo multivariadas. Este modelo muestra las interacciones simultáneas que se dan entre un grupo de variables a través de un sistema de ecuaciones que considera que el conjunto de variables explicativas de cada ecuación se forma a partir de los rezagos de cada una de las variables endógenas del modelo, además de sus propias variables exógenas o determinísticas, y sus respectivos rezagos. Una de las principales motivaciones detrás del uso de los modelos VAR es la dificultad en identificar las variables exógenas a incluir en los modelos, como es preciso hacer en un modelo de ecuaciones simultáneas. El VAR fue introducido a la aplicación en series de tiempo económicas por Sims [35], quien mostró que los modelos VAR entregan especial flexibilidad al analizar y pronosticar series de tiempo económicas y financieras.

Un Vector Autorregresivo de orden p , denotado como $VAR(p)$, es una generalización de 2.27 expresada en términos matriciales:

$$\mathbf{y}_t = \mathbf{c} + \Phi_1\mathbf{y}_{t-1} + \Phi_2\mathbf{y}_{t-2} + \dots + \Phi_p\mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t \quad (2.30)$$

Donde las interacciones dinámicas entre un conjunto de variables se recolectan en el vector \mathbf{y}_t de dimensiones $(n \times 1)$. Por ejemplo, el primer elemento de \mathbf{y}_t (denotado y_{1t}) representa una variable en el tiempo t , mientras que el segundo elemento (y_{2t}) representa otra variable en el tiempo t , y así sucesivamente para el resto de las n variables. El vector \mathbf{c} , de dimensiones $(n \times 1)$, almacena las constantes de cada ecuación; la matriz Φ_k , de dimensiones $(n \times n)$, contiene los coeficientes autorregresivos para $k = 1, 2, \dots, p$; el vector $(n \times 1)$ $\boldsymbol{\varepsilon}_t$ es una generalización del ruido blanco que cuenta con un primer momento $E(\boldsymbol{\varepsilon}_t) = \mathbf{0}$ y un

segundo momento $E(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t') = \boldsymbol{\Omega}$ para $t = \tau$, donde $\boldsymbol{\Omega}$ se define como una matriz $(n \times n)$ simétrica positiva.

Sean c_i el i -ésimo elemento del vector \mathbf{c} y $\phi_{ij}^{(k)}$ el elemento de la fila i -ésima en la columna j -ésima para la matriz $\boldsymbol{\Phi}_k$. Entonces la primera fila del sistema 2.30 se especifica

$$\begin{aligned} y_{1t} = & c_1 + \phi_{11}^{(1)}y_{1,t-1} + \phi_{12}^{(1)}y_{2,t-1} + \dots + \phi_{1n}^{(1)}y_{n,t-1} \\ & + \phi_{11}^{(2)}y_{1,t-2} + \phi_{12}^{(2)}y_{2,t-2} + \dots + \phi_{1n}^{(2)}y_{n,t-2} \\ & + \dots + \phi_{11}^{(p)}y_{1,t-p} + \phi_{12}^{(p)}y_{2,t-p} + \dots + \phi_{1n}^{(p)}y_{n,t-p} + \varepsilon_{1t} \end{aligned} \quad (2.31)$$

El modelo de Vector Autorregresivo es un sistema en el cual cada variable responde a una regresión con respecto a sus p rezagos y a una constante, así como también respecto a p rezagos de cada una de las otras variables consideradas en el modelo, notando que cada regresión tiene las mismas variables explicativas.

Usando la notación del operador de rezagos, la ecuación 2.30 puede reescribirse en la forma

$$[\mathbf{I}_n - \boldsymbol{\Phi}_1 L - \boldsymbol{\Phi}_2 L^2 - \dots - \boldsymbol{\Phi}_p L^p] \mathbf{y}_t = \mathbf{c} + \boldsymbol{\varepsilon}_t \quad (2.32)$$

O, equivalentemente

$$\boldsymbol{\Phi}(L) \mathbf{y}_t = \mathbf{c} + \boldsymbol{\varepsilon}_t$$

Donde $\boldsymbol{\Phi}(L)$ representa una matriz polinomial $(n \times n)$ en el operador de rezagos L .

Similar a lo que ocurre en un modelo AR, asumiendo que el VAR(p) es estacionario, se puede obtener la media $\boldsymbol{\mu}$ del proceso:

$$\boldsymbol{\mu} = (\mathbf{I}_n - \boldsymbol{\Phi}_1 - \boldsymbol{\Phi}_2 - \dots - \boldsymbol{\Phi}_p)^{-1} \mathbf{c} \quad (2.33)$$

Entonces 2.30 puede reescribirse en términos de la desviación de la media

$$(\mathbf{y}_t - \boldsymbol{\mu}) = \boldsymbol{\Phi}_1(\mathbf{y}_{t-1} - \boldsymbol{\mu}) + \boldsymbol{\Phi}_2(\mathbf{y}_{t-2} - \boldsymbol{\mu}) + \dots + \boldsymbol{\Phi}_p(\mathbf{y}_{t-p} - \boldsymbol{\mu}) + \boldsymbol{\varepsilon}_t \quad (2.34)$$

Esta representación conviene reescribirla como un VAR(1), por lo que se define

$$\xi_t = \begin{bmatrix} \mathbf{y}_t - \boldsymbol{\mu} \\ \mathbf{y}_{t-1} - \boldsymbol{\mu} \\ \vdots \\ \mathbf{y}_{t-p+1} - \boldsymbol{\mu} \end{bmatrix}$$

$$\mathbf{F} = \begin{bmatrix} \boldsymbol{\Phi}_1 & \boldsymbol{\Phi}_2 & \dots & \boldsymbol{\Phi}_{p-1} & \boldsymbol{\Phi}_p \\ \mathbf{I}_n & 0 & \dots & 0 & 0 \\ 0 & \mathbf{I}_n & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & \mathbf{I}_n & 0 \end{bmatrix}$$

$$\mathbf{v}_t = \begin{bmatrix} \boldsymbol{\varepsilon}_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Entonces el VAR(p) en 2.34 se puede reescribir como el siguiente VAR(1):

$$\xi_t = \mathbf{F}\xi_{t-1} + \mathbf{v}_t \quad (2.35)$$

Donde

$$E(\mathbf{v}_t, \mathbf{v}'_\tau) = \begin{cases} \mathbf{Q} & \text{para } t = \tau \\ \mathbf{0} & \text{para Otro caso} \end{cases}$$

con

$$\mathbf{Q} = \begin{bmatrix} \boldsymbol{\Omega} & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix}$$

La ecuación 2.35 implica que

$$\xi_{t+s} = \mathbf{v}_{t+s} + \mathbf{F}\mathbf{v}_{t+s-1} + \mathbf{F}^2\mathbf{v}_{t+s-2} + \dots + \mathbf{F}^{s-1}\mathbf{v}_{t+1} + \mathbf{F}^s\xi_t \quad (2.36)$$

Para que el proceso VAR sea estacionario, todos los valores propios de \mathbf{F} deben estar incluidos dentro del círculo de radio 1. Los valores propios de \mathbf{F} satisfacen

$$|\mathbf{I}_n\lambda^p - \Phi_1\lambda^{p-1} - \Phi_2\lambda^{p-2} - \dots - \Phi_p| = 0 \quad (2.37)$$

Entonces, un VAR(p) es estacionario mientras $|\lambda| < 1$ para los valores propios de 2.37. De forma equivalente, el VAR(1) es estacionario si todos los valores de z que satisfacen

$$|\mathbf{I}_n - \Phi_1z - \Phi_2z^2 - \dots - \Phi_pz^p| = 0 \quad (2.38)$$

están fuera del círculo de radio 1.

Las primeras n filas del sistema de vectores representado en 2.35 corresponden a una generalización

$$\begin{aligned} \mathbf{y}_{t+s} = & \boldsymbol{\mu} + \boldsymbol{\varepsilon}_{t+s} + \Psi_1\boldsymbol{\varepsilon}_{t+s-1} + \Psi_2\boldsymbol{\varepsilon}_{t+s-2} + \dots + \Psi_{s-1}\boldsymbol{\varepsilon}_{t+1} \\ & + \mathbf{F}_{11}^{(s)}(\mathbf{y}_t - \boldsymbol{\mu}) + \mathbf{F}_{12}^{(s)}(\mathbf{y}_{t-1} - \boldsymbol{\mu}) + \dots + \mathbf{F}_{1p}^{(s)}(\mathbf{y}_{t-p+1} - \boldsymbol{\mu}) \end{aligned} \quad (2.39)$$

Aquí $\Psi_j = \mathbf{F}_{11}^j$ y \mathbf{F}_{11}^j denota el bloque superior izquierdo de \mathbf{F}^j , donde \mathbf{F}^j es la matriz \mathbf{F} elevada a la j -ésima potencia, esto es, la matriz \mathbf{F}_{11}^j de dimensiones $(n \times n)$ considera las filas desde la 1 a la n y columnas desde la 1 a la n de la matriz \mathbf{F}^j de dimensiones $(np \times np)$. Similarmente, \mathbf{F}_{12}^j denota el bloque de \mathbf{F}^j que consiste en las filas desde la 1 a la n y columnas desde la $(n+1)$ a la $2n$, mientras \mathbf{F}_{1p}^j considera las filas desde la 1 a la n y columnas desde la $[n(p-1)+1]$ a la np de \mathbf{F}^j . Si todos los valores propios de \mathbf{F} caen dentro del círculo de radio 1, entonces $\mathbf{F}^s \rightarrow \mathbf{0}$ a medida que $s \rightarrow \infty$ y el vector \mathbf{y}_t se puede expresar como una suma convergente de la historia de los errores $\boldsymbol{\varepsilon}$:

$$\mathbf{y}_t = \boldsymbol{\mu} + \boldsymbol{\varepsilon}_t + \Psi_1\boldsymbol{\varepsilon}_{t-1} + \Psi_2\boldsymbol{\varepsilon}_{t-2} + \dots = \boldsymbol{\mu} + \Psi(L)\boldsymbol{\varepsilon}_t \quad (2.40)$$

que es una representación vectorial de $MA(\infty)$.

Es importante notar que \mathbf{y}_{t-j} es una función lineal de $\boldsymbol{\varepsilon}_{t-j}, \boldsymbol{\varepsilon}_{t-j-1}, \dots$, donde ningún término está correlacionado con $\boldsymbol{\varepsilon}_{t+1}$ para $j = 0, 1, \dots$. Luego, $\boldsymbol{\varepsilon}_{t+1}$ no está correlacionado con \mathbf{y}_{t-j} para $j \geq 0$. Entonces, el pronóstico lineal de \mathbf{y}_{t+1} basado en $\mathbf{y}_t, \mathbf{y}_{t-1}, \dots$, está dado por

$$\hat{\mathbf{y}}_{t+1|t} = \boldsymbol{\mu} + \boldsymbol{\Phi}_1(\mathbf{y}_t - \boldsymbol{\mu}) + \boldsymbol{\Phi}_2(\mathbf{y}_{t-1} - \boldsymbol{\mu}) + \dots + \boldsymbol{\Phi}_p(\mathbf{y}_{t-p+1} - \boldsymbol{\mu}) \quad (2.41)$$

y $\boldsymbol{\varepsilon}_{t+1}$ puede ser interpretado como el error al pronosticar \mathbf{y}_{t+1} basándose en la función lineal de una constante y de $\mathbf{y}_t, \mathbf{y}_{t-1}, \dots$. De forma más general, se desprende de 2.39 que un pronóstico de \mathbf{y}_{t+s} basándose en $\mathbf{y}_t, \mathbf{y}_{t-1}, \dots$ tomará la forma

$$\hat{\mathbf{y}}_{t+s|t} = \boldsymbol{\mu} + \mathbf{F}_{11}^{(s)}(\mathbf{y}_t - \boldsymbol{\mu}) + \mathbf{F}_{12}^{(s)}(\mathbf{y}_{t-1} - \boldsymbol{\mu}) + \dots + \mathbf{F}_{1p}^{(s)}(\mathbf{y}_{t-p+1} - \boldsymbol{\mu}) \quad (2.42)$$

Las matrices de media móvil $\boldsymbol{\Psi}_j$ equivalentemente se pueden calcular como sigue. Los operadores $\boldsymbol{\Phi}(L)$ y $\boldsymbol{\Psi}(L)$ se relacionan a través de

$$\boldsymbol{\Phi}(L) = [\boldsymbol{\Psi}(L)]^{-1} \quad (2.43)$$

cumpliendo que

$$[\mathbf{I}_n - \boldsymbol{\Phi}_1 L - \boldsymbol{\Phi}_2 L^2 - \dots - \boldsymbol{\Phi}_p L^p][\mathbf{I}_n - \boldsymbol{\Psi}_1 L - \boldsymbol{\Psi}_2 L^2 + \dots] = \mathbf{I}_n \quad (2.44)$$

Si se fija el coeficiente que acompaña a L^1 igual a la matriz cero, se tiene

$$\boldsymbol{\Psi}_1 - \boldsymbol{\Phi}_1 = 0 \quad (2.45)$$

De forma general, al fijar el coeficiente que acompaña a L^s igual a la matriz cero, se tiene

$$\boldsymbol{\Psi}_s = \boldsymbol{\Phi}_1 \boldsymbol{\Psi}_{s-1} + \boldsymbol{\Phi}_2 \boldsymbol{\Psi}_{s-2} + \dots + \boldsymbol{\Phi}_p \boldsymbol{\Psi}_{s-p} \quad \text{para } s = 1, 2, \dots \quad (2.46)$$

Con $\boldsymbol{\Psi}_0 = \mathbf{I}_n$ y $\boldsymbol{\Psi}_s = \mathbf{0}$ para $s < 0$.

Dentro de los supuestos implícitos que se aplican en los Vectores Autorregresivos, se puede notar que: Cuando el VAR es estacionario, los parámetros \mathbf{c} y Φ_1, \dots, Φ_p en la ecuación 2.39 pueden definirse como los coeficientes de la proyección de \mathbf{y}_t bajo una constante y los vectores $\mathbf{y}_{t-1}, \dots, \mathbf{y}_{t-p}$. En consecuencia, ε_t no está correlacionado con los términos $\mathbf{y}_{t-1}, \dots, \mathbf{y}_{t-p}$ por la definición de Φ_1, \dots, Φ_p . Los parámetros de un VAR se pueden estimar de forma consistente a través de n regresiones de Mínimos Cuadrados Ordinarios (MCO) de la forma 2.45. Además, se asume que el error ε_t definido por esta proyección no está correlacionado con $\mathbf{y}_{t-p-1}, \mathbf{y}_{t-p-2}, \dots$. El supuesto de que \mathbf{y}_t se puede representar mediante un Vector Autorregresivo es básicamente el supuesto de que p rezagos son suficientes para resumir todas las correlaciones dinámicas entre los elementos de \mathbf{y} .

2.5.2.1. Función de verosimilitud condicional

Considerando la ecuación 2.35, se asume que las n variables se observan durante $(T+p)$ periodos de tiempo. La aproximación más simple es condicionar a las primeras p observaciones (denotadas como $\mathbf{y}_{t-p+1}, \mathbf{y}_{t-p+2}, \dots, \mathbf{y}_0$) y estimar las últimas T observaciones (denotadas como $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_T$). El objetivo es formar la función de verosimilitud condicional

$$f_{\mathbf{Y}_T, \mathbf{Y}_{T-1}, \dots, \mathbf{Y}_1 | \mathbf{Y}_0, \mathbf{Y}_{-1}, \dots, \mathbf{Y}_{-p+1}}(\mathbf{y}_T, \mathbf{y}_{T-1}, \dots, \mathbf{y}_1 | \mathbf{y}_0, \mathbf{y}_{-1}, \dots, \mathbf{y}_{-p+1}; \boldsymbol{\theta}) \quad (2.47)$$

y maximizar con respecto a $\boldsymbol{\theta}$ llamado el estimador de máxima verosimilitud, donde $\boldsymbol{\theta}$ es un vector que contiene los elementos de \mathbf{c} , Φ_1, \dots, Φ_p , y $\boldsymbol{\Omega}$. Los Vectores Autorregresivos se estiman basándose en la función 2.47 en lugar de una función de verosimilitud no condicional de toda la muestra.

Condicionado a los valores de \mathbf{y} observados hasta el tiempo $t - 1$, el valor de \mathbf{y} para el dato t es igual a la constante

$$\mathbf{c} + \Phi_1 \mathbf{y}_{t-1} + \Phi_2 \mathbf{y}_{t-2} + \dots + \Phi_p \mathbf{y}_{t-p} \quad (2.48)$$

Además de una variable $N(\mathbf{0}, \mathbf{\Omega})$. Entonces,

$$\mathbf{y}_t | \mathbf{y}_{t-1}, \mathbf{y}_{t-2}, \dots, \mathbf{y}_{-p+1} \sim N(\mathbf{c} + \mathbf{\Phi}_1 \mathbf{y}_{t-1} + \mathbf{\Phi}_2 \mathbf{y}_{t-2} + \dots + \mathbf{\Phi}_p \mathbf{y}_{t-p}, \mathbf{\Omega}) \quad (2.49)$$

La media condicional 2.48 se puede escribir de una forma más compacta. Sea \mathbf{x}_t un vector que contiene una constante y p rezagos de cada uno de los elementos de \mathbf{y} :

$$\mathbf{x}_t = \begin{bmatrix} 1 \\ \mathbf{y}_{t-1} \\ \mathbf{y}_{t-2} \\ \vdots \\ \mathbf{y}_{t-p} \end{bmatrix}$$

$((np+1) \times 1)$

Y sea $\mathbf{\Pi}'$ la matriz:

$$\mathbf{\Pi}' = \begin{bmatrix} \mathbf{c} & \mathbf{\Phi}_1 & \mathbf{\Phi}_2 & \dots & \mathbf{\Phi}_p \end{bmatrix}$$

$((np+1) \times 1)$

Entonces, la media condicional 2.48 se puede reescribir como $\mathbf{\Pi}' \mathbf{x}_t$. La j -ésima fila de $\mathbf{\Pi}'$ contiene los parámetros de la j -ésima ecuación del VAR. Usando esta notación, 2.49 se puede escribir de forma compacta como

$$\mathbf{y}_t | \mathbf{y}_{t-1}, \mathbf{y}_{t-2}, \dots, \mathbf{y}_{-p+1} \sim N(\mathbf{\Pi}' \mathbf{x}_t, \mathbf{\Omega}) \quad (2.50)$$

Así, la densidad condicional de la observación t es

$$f_{\mathbf{Y}_t | \mathbf{Y}_{t-1}, \mathbf{Y}_{t-2}, \dots, \mathbf{Y}_{-p+1}}(\mathbf{y}_t | \mathbf{y}_{t-1}, \mathbf{y}_{t-2}, \dots, \mathbf{y}_{-p+1}; \boldsymbol{\theta}) = (2\pi)^{-n/2} |\mathbf{\Omega}^{-1}|^{1/2} \exp\left[(-1/2)(\mathbf{y}_t - \mathbf{\Pi}' \mathbf{x}_t)' \mathbf{\Omega}^{-1} (\mathbf{y}_t - \mathbf{\Pi}' \mathbf{x}_t)\right] \quad (2.51)$$

La densidad conjunta de las observaciones 1 hasta t condicionada a $\mathbf{y}_0, \mathbf{y}_{-1}, \dots, \mathbf{y}_{-p+1}$ satisface

$$\begin{aligned} & f_{\mathbf{Y}_t, \mathbf{Y}_{t-1}, \dots, \mathbf{Y}_1 | \mathbf{Y}_0, \mathbf{Y}_{-1}, \dots, \mathbf{Y}_{-p+1}}(\mathbf{y}_t, \mathbf{y}_{t-1}, \dots, \mathbf{y}_1 | \mathbf{y}_0, \mathbf{y}_{-1}, \dots, \mathbf{y}_{-p+1}; \boldsymbol{\theta}) \\ &= f_{\mathbf{Y}_{t-1}, \dots, \mathbf{Y}_1 | \mathbf{Y}_0, \mathbf{Y}_{-1}, \dots, \mathbf{Y}_{-p+1}}(\mathbf{y}_{t-1}, \dots, \mathbf{y}_1 | \mathbf{y}_0, \mathbf{y}_{-1}, \dots, \mathbf{y}_{-p+1}; \boldsymbol{\theta}) \\ &\quad \times f_{\mathbf{Y}_t | \mathbf{Y}_{t-1}, \mathbf{Y}_{t-2}, \dots, \mathbf{Y}_{-p+1}}(\mathbf{y}_t | \mathbf{y}_{t-1}, \mathbf{y}_{t-2}, \dots, \mathbf{y}_{-p+1}; \boldsymbol{\theta}) \end{aligned}$$

Aplicando esta fórmula de manera recursiva, la verosimilitud para la muestra completa $\mathbf{y}_T, \mathbf{y}_{T-1}, \dots, \mathbf{y}_1$ condicionada a $\mathbf{y}_0, \mathbf{y}_{-1}, \dots, \mathbf{y}_{-p+1}$ es el producto de las densidades condicionales individuales:

$$\begin{aligned} & f_{\mathbf{y}_T, \mathbf{y}_{T-1}, \dots, \mathbf{y}_1 | \mathbf{y}_0, \mathbf{y}_{-1}, \dots, \mathbf{y}_{-p+1}}(\mathbf{y}_T, \mathbf{y}_{T-1}, \dots, \mathbf{y}_1 | \mathbf{y}_0, \mathbf{y}_{-1}, \dots, \mathbf{y}_{-p+1}; \boldsymbol{\theta}) \\ &= \prod_{t=1}^T f_{\mathbf{y}_t | \mathbf{y}_{t-1}, \mathbf{y}_{t-2}, \dots, \mathbf{y}_{-p+1}}(\mathbf{y}_t | \mathbf{y}_{t-1}, \mathbf{y}_{t-2}, \dots, \mathbf{y}_{-p+1}; \boldsymbol{\theta}) \end{aligned} \quad (2.52)$$

La verosimilitud logarítmica de la muestra se encuentra al reemplazar 2.51 en 2.52 y aplicando logaritmo natural:

$$\begin{aligned} \ell(\boldsymbol{\theta}) &= \sum_{t=1}^T \log f_{\mathbf{y}_t | \mathbf{y}_{t-1}, \dots, \mathbf{y}_1 | \mathbf{y}_0, \mathbf{y}_{-1}, \dots, \mathbf{y}_{-p+1}}(\mathbf{y}_t, \mathbf{y}_{t-1}, \dots, \mathbf{y}_1 | \mathbf{y}_0, \mathbf{y}_{-1}, \dots, \mathbf{y}_{-p+1}; \boldsymbol{\theta}) \\ &= -(Tn/2) \log(2\pi) + (T/2) \log |\boldsymbol{\Omega}^{-1}| \\ &\quad - (1/2) \sum_{t=1}^T [(\mathbf{y}_t - \boldsymbol{\Pi}' \mathbf{x}_t)' \boldsymbol{\Omega}^{-1} (\mathbf{y}_t - \boldsymbol{\Pi}' \mathbf{x}_t)] \end{aligned} \quad (2.53)$$

El estimador de máxima verosimilitud para $\boldsymbol{\Pi}$, que considera la constante \mathbf{c} y los coeficientes autorregresivos $\boldsymbol{\Phi}_j$ está dado por

$$\hat{\boldsymbol{\Pi}}'_{(n \times (np+1))} = \left[\sum_{t=1}^T \mathbf{y}_t \mathbf{x}_t' \right] \left[\sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t' \right]^{-1}$$

Donde la fila j -ésima de $\hat{\boldsymbol{\Pi}}'$ es

$$\hat{\boldsymbol{\pi}}'_{(1 \times (np+1))} = \left[\sum_{t=1}^T y_{jt} \mathbf{x}_t' \right] \left[\sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t' \right]^{-1}$$

Entonces, los estimadores de máxima verosimilitud de los coeficientes para la ecuación j -ésima de un VAR se encuentran a través de una regresión de MCO de y_{jt} sobre una constante y p rezagos de todas las variables en el sistema.

El estimador de máxima verosimilitud para Ω está dado por

$$\hat{\Omega} = (1/T) \sum_{t=1}^T \hat{\varepsilon}_t \hat{\varepsilon}_t' \quad (2.54)$$

Donde el elemento de la fila i y la columna j de $\hat{\Omega}$ es

$$\hat{\sigma}_{ij} = (1/T) \sum_{t=1}^T \hat{\varepsilon}_{it} \hat{\varepsilon}_{jt} \quad (2.55)$$

lo cual corresponde al producto promedio del MCO del residual para la variable i y el MCO del residual de la variable j . Si $i = j$, este término se reescribe simplemente como el promedio del residual, proveniente de la regresión de la variable i en el VAR sobre una constante y p rezagos de todas las variables, al cuadrado

$$\hat{\sigma}_i^2 = (1/T) \sum_{t=1}^T \hat{\varepsilon}_{it}^2 \quad (2.56)$$

2.5.2.2. Causalidad de Granger

Un elemento clave para obtener un Vector Autorregresivo es saber si es que una variable y puede ayudar a pronosticar otra variable x . Si no puede, entonces se dice que y no causa por Granger a x , como plantea Granger [36]. Más formalmente, y no causa por Granger a x si para cada $s > 0$ el error cuadrado medio (abreviado MSE, por las siglas en inglés para Mean Squared Error) de un pronóstico de x_{t+s} basado en (x_t, x_{t-1}, \dots) es el mismo que el MSE de un pronóstico de x_{t+s} que se basa en (x_t, x_{t-1}, \dots) y (y_t, y_{t-1}, \dots) . Restringiéndose a funciones lineales, lo anterior se puede escribir como

$$MSE[\hat{E}(x_{t+s}|x_t, x_{t-1}, \dots)] = MSE[\hat{E}(x_{t+s}|x_t, x_{t-1}, \dots, y_t, y_{t-1}, \dots)] \quad (2.57)$$

Para implementar pruebas econométricas que verifican si una serie y causa por Granger a otra x , se estima por MCO la expresión autorregresiva con p rezagos

$$x_t = c_1 + \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \dots + \alpha_p x_{t-p} + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \dots + \beta_p y_{t-p} + u_t \quad (2.58)$$

Entonces se aplica una prueba F de hipótesis nula

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0 \quad (2.59)$$

Una forma de implementar esta prueba es calcular la suma del cuadrados de los residuales de 2.57,

$$RSS_1 = \sum_{t=1}^T \hat{u}_t^2 \quad (2.60)$$

y compararlo con la suma del cuadrado de los residuales de un autorregresivo univariado para x_t ,

$$RSS_0 = \sum_{t=1}^T \hat{e}_t^2 \quad (2.61)$$

donde

$$x_t = c_0 + \gamma_1 x_{t-1} + \gamma_2 x_{t-2} + \dots + \gamma_p x_{t-p} + e_t \quad (2.62)$$

también se estima por MCO. Si

$$S_1 = \frac{(RSS_0 - RSS_1)/p}{RSS_1/(T - 2p - 1)} \quad (2.63)$$

es mayor que el valor crítico de 5 % para una distribución $F(p, T - 2p - 1)$, entonces se rechaza la hipótesis nula de que y no causa por Granger a x . La prueba estadística 2.63 tendría una distribución F exacta para una regresión con regresores fijos y residuales Gaussianos, pero con variables dependientes rezagadas como en este caso, la prueba es válida solamente de forma asintótica. Una prueba asintóticamente equivalente está dada por

$$S_2 = \frac{T(RSS_0 - RSS_1)}{RSS_1} \quad (2.64)$$

y se rechaza la hipótesis nula que y no causa por Granger a x si S_2 es mayor que un valor crítico de 5 % para una variable $\chi^2(p)$.

2.5.2.3. Vector Autorregresivo con raíces unitarias

Suponiendo que un vector y_t se puede describir por un VAR en las diferencias Δy_t , es posible estimar dicho VAR en niveles. Considerando la expresión 2.32, para cualquier valor de $\Phi_1, \Phi_2, \dots, \Phi_p$ los siguientes polinomios son equivalentes:

$$(I_n - \Phi_1 L - \Phi_2 L^2 - \dots - \Phi_p L^p) = (I_n - \rho L) - (\zeta_1 L + \zeta_2 L^2 + \dots + \zeta_{p-1} L^{p-1})(1 - L) \quad (2.65)$$

donde

$$\rho = \Phi_1 + \Phi_2 + \dots + \Phi_p \quad (2.66)$$

$$\zeta_s = -[\Phi_{s+1} + \Phi_{s+2} + \dots + \Phi_p] \quad \text{para } s = 1, 2, \dots, p-1 \quad (2.67)$$

Entonces, cualquier VAR(p) 2.32 puede reescribirse de la forma

$$(I_n - \rho L)y_t - (\zeta_1 L + \zeta_2 L^2 + \dots + \zeta_{p-1} L^{p-1})(1 - L)y_t = c + \varepsilon_t \quad (2.68)$$

o, de forma equivalente:

$$y_t = \zeta_1 \Delta y_{t-1} + \zeta_2 \Delta y_{t-2} + \dots + \zeta_{p-1} \Delta y_{t-p+1} + c + \rho y_{t-1} + \varepsilon_t \quad (2.69)$$

Luego, la primera diferencia de y sigue un VAR($p-1$) de la forma

$$\Delta y_t = \zeta_1 \Delta y_{t-1} + \zeta_2 \Delta y_{t-2} + \dots + \zeta_{p-1} \Delta y_{t-p+1} + c + \varepsilon_t \quad (2.70)$$

requiriendo de 2.66 que

$$\rho = \Phi_1 + \Phi_2 + \dots + \Phi_p = I_n \quad (2.71)$$

Entonces, el VAR expresado en 2.32 contiene al menos una raíz unitaria si se cumple para el siguiente determinante:

$$|I_n - \Phi_1 - \Phi_2 - \dots - \Phi_p| = 0 \quad (2.72)$$

2.5.3. Cointegración

Los procesos cointegrados son una clase particular de vectores de raíz unitaria implícitos en los modelos de Corrección del Error, propuestos por Davidson et al. [37] y trabajados posteriormente por Engle & Granger [38].

Un vector de series de tiempo \mathbf{y}_t de dimensión $(n \times 1)$ se denomina cointegrado si cada una de las series individuales son no estacionarias, $I(1)$, mientras que combinaciones de las series $\mathbf{a}'\mathbf{y}_t$ son estacionarias, $I(0)$, para un vector de cointegración \mathbf{a} de dimensión $(n \times 1)$ distinto de cero. Como ejemplo de un vector cointegrado, se puede considerar el siguiente sistema:

$$y_{1t} = \gamma y_{2t} + u_{1t} \quad (2.73)$$

$$y_{2t} = y_{2,t-1} + u_{2t} \quad (2.74)$$

con u_{1t} y u_{2t} procesos ruido blanco no correlacionados. La ecuación 2.74 se puede representar como una caminata aleatoria,

$$\Delta y_{2t} = u_{2t} \quad (2.75)$$

mientras que al diferenciar 2.73 se tiene

$$\Delta y_{1t} = \gamma \Delta y_{2t} + \Delta u_{1t} = \gamma u_{2t} + u_{1t} - u_{1,t-1} \quad (2.76)$$

lo cual tiene una representación como media móvil de orden 1 ($MA(1)$):

$$\Delta y_{1t} = v_t + \theta v_{t-1} \quad (2.77)$$

donde v_t es un proceso ruido blanco y $\theta \neq 1$ mientras $\gamma \neq 0$ y $E(u_{2t}^2) > 0$. Así, tanto y_{1t} como y_{2t} son $I(1)$ aunque la combinación lineal $(y_{1t} - \gamma y_{2t})$ es estacionaria. Entonces, se dice que $\mathbf{y}_t = (y_{1t}, y_{2t})'$ está cointegrado con $\mathbf{a}' = (1, -\gamma)$. Cointegración significa que aunque existan cambios permanentes en el tiempo para cada elemento individual de \mathbf{y}_t ,

existe una relación de equilibrio en el largo plazo que afecta en conjunto estos elementos, representada por la combinación lineal $\mathbf{a}'\mathbf{y}_t$.

Cuando \mathbf{y}_t está cointegrado, entonces no es correcto ajustar un modelo VAR a los datos en su diferencia. Un VAR de orden finito en diferencias logra una mala aproximación al sistema cointegrado de 2.73 y 2.74 ya que y_2 en su nivel contiene información útil para pronosticar y_1 más allá de la contenida en un número finito de rezagos al variar solamente y_2 . Un sistema cointegrado debe incluir rezagos de la variable en su nivel junto a rezagos de las diferencias en un vector autorregresivo que explica a $\Delta\mathbf{y}_t$.

2.5.3.1. Caracterización general del vector de cointegración

Recordando que un vector \mathbf{y}_t de dimensiones $(n \times 1)$ está cointegrado si cada uno de sus elementos individuales es $I(1)$ y si existe un vector \mathbf{a} de dimensiones $(n \times 1)$ distinto de cero tal que $\mathbf{a}'\mathbf{y}_t$ sea estacionario. En este caso, \mathbf{a} se denomina vector de cointegración.

Este vector de cointegración no es único. Para cualquier escalar b distinto de cero, si \mathbf{a} es un vector de cointegración, entonces $b\mathbf{a}$ también lo es. De hecho, podrían existir $h < n$ vectores $(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_h)$ de dimensiones $(n \times 1)$ linealmente independientes tales que $\mathbf{A}'\mathbf{y}_t$ sea un vector estacionario de dimensiones $(h \times 1)$, donde \mathbf{A}' es la siguiente matriz:

$$\mathbf{A}' = \begin{bmatrix} \mathbf{a}'_1 \\ \mathbf{a}'_2 \\ \vdots \\ \mathbf{a}'_h \end{bmatrix}$$

$(h \times n)$

Para un vector \mathbf{c}' de dimensiones $(1 \times n)$ que es linealmente independiente de las filas de \mathbf{A}' , se tiene que $\mathbf{c}'\mathbf{y}_t$ es un escalar no estacionario. Entonces, existen exactamente h relaciones de cointegración entre los elementos de \mathbf{y}_t basándose en el espacio de vectores de cointegración $(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_h)$.

2.5.3.2. Implicancias de la Cointegración en el VAR

Aunque un VAR en diferencias no es consistente con un sistema cointegrado, un VAR en su nivel sí puede serlo. Si \mathbf{y}_t puede ser representada en su nivel como un vector autorregresivo de orden p de la forma 2.30:

$$\mathbf{y}_t = \mathbf{c} + \Phi_1 \mathbf{y}_{t-1} + \Phi_2 \mathbf{y}_{t-2} + \dots + \Phi_p \mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t$$

o equivalentemente

$$\Phi(L)\mathbf{y}_t = \mathbf{c} + \boldsymbol{\varepsilon}_t \quad (2.78)$$

donde

$$\Phi(L) = \mathbf{I}_n - \Phi_1 L - \Phi_2 L^2 - \dots - \Phi_p L^p \quad (2.79)$$

Entonces $\Delta \mathbf{y}_t$ tiene una representación

$$(1 - L)\mathbf{y}_t = \boldsymbol{\delta} + \Psi(L)\boldsymbol{\varepsilon}_t \quad (2.80)$$

donde $\boldsymbol{\delta} = E(\Delta \mathbf{y}_t)$ [33]. Agregando $\Phi(L)$ a 2.80 se tiene

$$(1 - L)\Phi(L)\mathbf{y}_t = \Phi(L)\boldsymbol{\delta} + \Phi(L)\Psi(L)\boldsymbol{\varepsilon}_t \quad (2.81)$$

Reemplazando 2.78 en 2.81 y sabiendo que $(1 - L)\mathbf{c} = \mathbf{0}$, se tiene

$$(1 - L)\boldsymbol{\varepsilon}_t = \Phi(1)\boldsymbol{\delta} + \Phi(L)\Psi(L)\boldsymbol{\varepsilon}_t \quad (2.82)$$

Como 2.82 debe mantener la igualdad para todo $\boldsymbol{\varepsilon}_t$, se requiere que

$$\Phi(1)\boldsymbol{\delta} = \mathbf{0} \quad (2.83)$$

además de que $(1 - L)\mathbf{I}_n$ y $\Phi(L)\Psi(L)$ sean polinomios idénticos en L . Lo que implica, en particular para $L = 1$, que

$$\Phi(1)\Psi(1) = \mathbf{0} \quad (2.84)$$

Denotando como π' a cualquier fila de $\Phi(1)$. Entonces 2.83 y 2.84 implicarían que $\pi' \delta = \mathbf{0}$ y que $\pi' \Psi(1) = \mathbf{0}$, por lo que π' es un vector de cointegración [33] que cumple con:

$$\pi = [\mathbf{a}_1 \ \mathbf{a}_2 \ \dots \ \mathbf{a}_h] \mathbf{b}$$

o, equivalentemente

$$\pi' = \mathbf{b}' \mathbf{A}'$$

para la matriz \mathbf{A}' de dimensiones $(h \times n)$ cuya fila i -ésima es \mathbf{a}'_i , y para el vector \mathbf{b} de dimensiones $(h \times 1)$. Aplicando este razonamiento para cada fila de $\Phi(1)$, entonces existe una matriz \mathbf{B} de dimensiones $(n \times h)$ tal que

$$\Phi(1) = \mathbf{B} \mathbf{A}' \quad (2.85)$$

2.5.3.3. Representación del VECM

Una representación final para un sistema cointegrado se obtiene recordando de la ecuación 2.69 que cualquier VAR en la forma 2.30 se puede escribir como

$$\mathbf{y}_t = \mathbf{c} + \zeta_1 \Delta \mathbf{y}_{t-1} + \zeta_2 \Delta \mathbf{y}_{t-2} + \dots + \zeta_{p-1} \Delta \mathbf{y}_{t-p+1} + \rho \mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t$$

Restando \mathbf{y}_{t-1} a ambos lados se tiene

$$\Delta \mathbf{y}_t = \mathbf{c} + \zeta_1 \Delta \mathbf{y}_{t-1} + \zeta_2 \Delta \mathbf{y}_{t-2} + \dots + \zeta_{p-1} \Delta \mathbf{y}_{t-p+1} + \zeta_0 \mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t \quad (2.86)$$

donde

$$\zeta_0 = \rho - \mathbf{I}_n = -(\mathbf{I}_n - \Phi_1 - \Phi_2 - \dots - \Phi_p) = -\Phi(1) \quad (2.87)$$

Si \mathbf{y}_t tiene h relaciones de cointegración, entonces al reemplazar 2.85 y 2.87 en 2.86 se tiene

$$\Delta \mathbf{y}_t = \mathbf{c} + \zeta_1 \Delta \mathbf{y}_{t-1} + \zeta_2 \Delta \mathbf{y}_{t-2} + \dots + \zeta_{p-1} \Delta \mathbf{y}_{t-p+1} - \mathbf{B} \mathbf{A}' \mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t \quad (2.88)$$

Al definir $\mathbf{z}_t = \mathbf{A}' \mathbf{y}_t$ y notando que \mathbf{z}_t es un vector estacionario de dimensiones $(h \times 1)$, entonces 2.88 se puede reescribir como la representación de la Corrección del error del

vector cointegrado [33] de la forma

$$\Delta \mathbf{y}_t = \mathbf{c} + \zeta_1 \Delta \mathbf{y}_{t-1} + \zeta_2 \Delta \mathbf{y}_{t-2} + \dots + \zeta_{p-1} \Delta \mathbf{y}_{t-p+1} - \mathbf{B} \mathbf{z}_{t-1} + \boldsymbol{\varepsilon}_t \quad (2.89)$$

Este modelo también se conoce como VECM por las siglas en inglés para Vector Error Correction Model. Aquí, a diferencia de lo que haría un VAR incluyendo solo rezagos de las diferencias para representar $\Delta \mathbf{y}_t$, las diferencias en cada variable tienen una regresión a partir de una constante, de $(p - 1)$ rezagos del cambio de la misma variable, de $(p - 1)$ rezagos del cambio en cada una de las otras variables del sistema, y de los niveles originales de cada elemento h de \mathbf{z}_{t-1} .

Para el ejemplo de 2.73 y 2.74, este sistema se puede escribir matricialmente de la forma

$$\begin{bmatrix} \Delta y_{1t} \\ \Delta y_{2t} \end{bmatrix} = \begin{bmatrix} -1 & \gamma \\ 0 & 0 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} \gamma u_{2t} + u_{1t} \\ u_{2t} \end{bmatrix}$$

lo que es un caso particular de 2.86 con $p = 1$,

$$\zeta_0 = \begin{bmatrix} -1 & \gamma \\ 0 & 0 \end{bmatrix},$$

$\varepsilon_{1t} = \gamma u_{2t} + u_{1t}$, $\varepsilon_{2t} = u_{2t}$ y todos los otros parámetros son iguales a cero. Entonces la forma de Corrección del Error es

$$\begin{bmatrix} \Delta y_{1t} \\ \Delta y_{2t} \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \end{bmatrix} \mathbf{z}_{t-1} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}$$

donde $\mathbf{z}_t = y_{1t} - \gamma y_{2t}$ representa el error de la relación de equilibrio a largo plazo y el vector que lo antecede es la corrección a Δy_t causada por ese error.

Las pruebas de Johansen, propuestas por Johansen [39] y Johansen & Juselius [40], evalúan la existencia de cointegración a través de 2 pruebas diferentes: La prueba de Traza y la prueba del Máximo Valor Propio. En la primera, la hipótesis nula asume que existen a lo más h vectores de cointegración, lo cual se contrasta con la hipótesis alternativa de

que existen exactamente h . Por otro lado, en la segunda prueba, la hipótesis nula de que existen exactamente h vectores de cointegración se contrasta con la hipótesis alternativa que plantea la existencia de $h + 1$ vectores de cointegración.

2.6. Redes Neuronales

El interés por la investigación sobre Redes Neuronales ha aumentado en los últimos años. Desde la publicación de McCulloch & Pitts en 1943 [41], donde se presentó el primer modelo de neuronas artificiales, se han desarrollado propuestas nuevas y más sofisticadas con el paso del tiempo. La Red Neuronal Artificial (abreviada ANN, por las siglas en inglés para Artificial Neural Network) es un intento de modelar la capacidad de procesamiento de información que tiene el sistema nervioso basándose en la neurona. El cerebro se compone de miles de millones de neuronas interconectadas y, aunque individualmente cada neurona necesita varios segundos para reaccionar a diferentes estímulos, su capacidad de dar respuesta es mucho más eficiente a la de sistemas computacionales digitales, debido a que su estructura de procesamiento se da a través de interconexiones en múltiples capas que trabajan en paralelo. Las ANN son sistemas de procesamiento de información de tipo no algorítmico que tienen la ventaja de poder reconocer patrones no lineales en una serie de tiempo, además de poder realizar inferencia incluso en datos que contienen ruido, y poder aproximar un gran número de funciones con un alto grado de precisión [42]. Dentro de las principales características de las ANN se encuentran:

- Son sistemas con una gran adaptabilidad a la información con la que se trabaja, lo que las transforma en una herramienta flexible.
- Tienen un alto grado de aprendizaje sobre las estructuras y patrones que guían a la información que procesan, lo que facilita su uso a la hora de pronosticar respuestas.
- Tienen la capacidad de realizar procesamientos en paralelo.

2.6.1. Estructura de una Neurona

De forma general, la estructura de una neurona biológica se compone principalmente de *dendritas*, que se encargan de recibir la señal al contactarse con otras células a través de la *sinapsis*; el *cuerpo*, que almacena y procesa lo que reciben las *dendritas*; y el *axón*, que se encarga de transmitir una respuesta. Considerando esto, una neurona artificial se modela mediante:

- *Inputs* o Nodos de entrada: que recogen los datos proporcionados a la neurona.
- Pesos: que expresan el grado de importancia con el que cada *input* se conecta a una cierta neurona. Se definen como valores numéricos entre 0 y 1.
- Punto de suma ponderada: que realiza una combinación lineal entre todas las entradas que llegan a ella y sus respectivos pesos.
- Función de Activación o Transferencia: que se encarga de computar, detectar patrones y limitar el rango de salida de la neurona. Esta función puede ser lineal o no lineal.
- Sesgo: que representa un cierto desplazamiento de la función de activación con respecto al origen del eje que representa el valor de la entrada neta. Originalmente se define con un valor 1.
- *Output* o Nodo de salida: que arroja la respuesta de la neurona, como resultado del procesamiento y aprendizaje de la información que se recibe en los *inputs*.

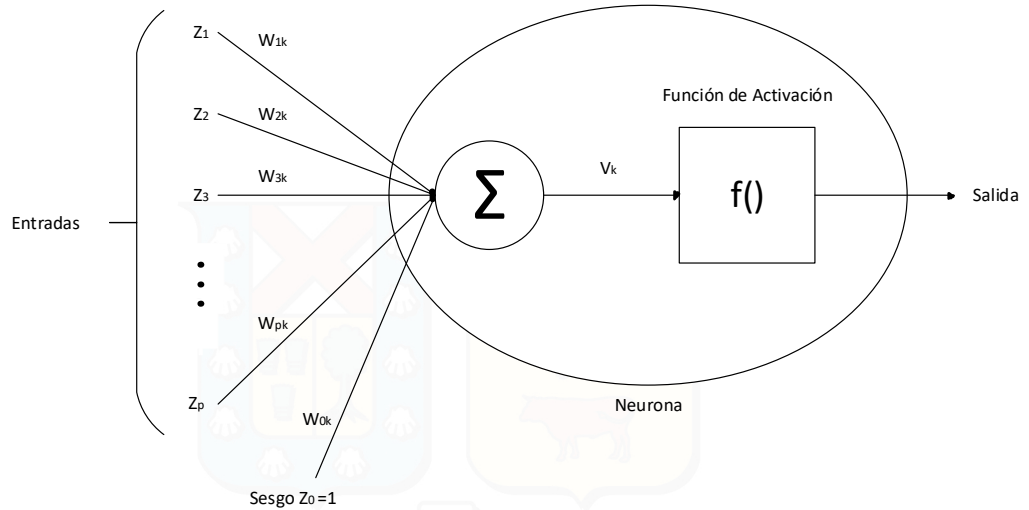


Figura 2.1: Estructura de una Neurona Artificial. Fuente: Elaboración propia.

En la fig 2.1 se puede apreciar la estructura para una Neurona Artificial k , donde Z_1, Z_2, \dots, Z_p son las p entradas de la neurona k ; $W_{1k}, W_{2k}, \dots, W_{pk}$ son los pesos de las entradas a la neurona k ; W_{0k} es el peso del sesgo Z_0 de la neurona k ; V_k es la suma de las entradas multiplicadas por los pesos respectivos; $f()$ es la función de activación para la neurona k que arroja una salida.

La Función de Activación a utilizar dependerá de cada investigación y datos en estudio. La mayoría de las veces, se seleccionan por prueba y error rescatando la que entrega mejores resultados. Entre las más usadas se encuentran:

- **Función Lineal:** Esta función entrega una salida igual a la entrada que recibe. Generalmente se usa en la capa de salida de una red neuronal. Matemáticamente puede expresarse como:

$$f(x) = x \quad (2.90)$$

- **Función Logística o Sigmoide (*Sig*):** Esta función entrega una salida con valores entre 0 y 1. Es una de las más usadas en redes neuronales debido a su gran capacidad

para resolver problemas de predicción. Matemáticamente puede expresarse como:

$$f(x) = \frac{1}{1 + e^{-x}} \quad (2.91)$$

- Función Tangente hiperbólica (*Tanh*): Aunque es similar a la función logística en cuanto a forma, entrega una salida con valores entre -1 y 1 . Se utiliza frecuentemente en redes neuronales de múltiples capas. Matemáticamente puede expresarse como:

$$f(x) = \frac{1 - e^{-2x}}{1 + e^{-2x}} \quad (2.92)$$

- Función Tangente hiperbólica Sigmoide (*Tansig*): Esta función entrega una salida con valores entre -1 y 1 . Recoge las ventajas de las funciones *Tanh* y *Sig* para realizar pronósticos en redes de múltiples capas. Matemáticamente puede expresarse como:

$$f(x) = \frac{2}{1 + e^{-2x}} - 1 \quad (2.93)$$

2.6.2. Estructura de una ANN

Una Red Neuronal es un sistema compuesto por varias de estas neuronas ordenadas y dispuestas de forma tal que se pueda procesar de la mejor manera la información con la que se la alimenta para obtener una respuesta deseada. Por lo general, la ANN se separa en 3 tipos de capas:

- Capa de entrada: Este es el primer nivel de la red. La capa de entrada recibe la información desde fuentes externas y las lleva al próximo nivel. El número de entradas dependerá del caso en estudio ya que no existe un método o regla que garantice encontrar un valor óptimo general. Los datos de entrada son numéricos y en muchos casos puede ser conveniente escalarlos y/o pre procesarlos para que sean trabajados de mejor manera en los siguientes pasos de la red sin perder su estructura e identidad.
- Capas ocultas: También se les denomina capas intermedias, por encontrarse entre

la capa de entrada y de salida de la red. Aquí se realiza el procesamiento de la información gracias a una función de activación que trabaja sobre los datos ingresados. El número de capas de la red, así como la cantidad de neuronas incluidas en cada una, es propio a cada problema y frecuentemente se obtienen basándose en el método de prueba y error hasta obtener una configuración que entregue los menores errores generalizados [43]. El número de capas ocultas se relaciona con el grado de linealidad y complejidad que presentan los datos, por lo tanto un mayor número de capas puede captar de mejor manera la componente no lineal de ellos.

- **Capa de Salida:** Esta capa recibe las neuronas o nodos de salida y las deriva la respuesta de la red al receptor externo. El número de salidas de la red dependerá del problema en estudio. La salida de la red deberá estar expresada en la misma escala de los datos originales, es decir, si los datos fueron escalados, deberán ser regresados a su escala inicial. Lo mismo ocurre para el caso del pre procesamiento.

La dirección en la que fluye la información dentro de la red neuronal determinará la clasificación que se le asigna. Existen varios tipos de Redes Neuronales, pero dentro de las más destacables se encuentran:

- **Redes de alimentación hacia adelante (*Feedforward*):** En este tipo de redes, las salidas de las neuronas de una capa solo se propagan a las neuronas de la capa siguiente, es decir, la información fluye solamente de la entrada a la salida.
- **Redes retroalimentadas:** En este tipo de redes, las salidas de las neuronas de una capa pueden ser entradas de las neuronas de las capas anteriores.
- **Redes de alimentación lateral:** En este tipo de redes, las salidas de las neuronas de una capa pueden ser entradas para las neuronas de la misma capa.
- **Redes recurrentes:** En este tipo de redes, existen lazos cerrados. Los datos se propagan hacia adelante, pero también hacia atrás.

2.6.3. Métodos de aprendizaje

Existen dos métodos básicos de aprendizaje: el No Supervisado y el Supervisado. Las reglas de aprendizaje supervisadas suelen ser más complejas pero también más exactas en sus resultados.

2.6.3.1. Aprendizaje No Supervisado

Este tipo de aprendizaje considera un entrenamiento sin guía y sólo se usa información local durante todo el proceso. Es un modelo más cercano al sistema biológico, no se utiliza vector de salida esperada, y sólo hay vectores de entrada en el conjunto de entrenamiento. El algoritmo modifica los pesos de forma que las salidas sean consistentes, es decir, que a entradas muy parecidas, la red compute la misma salida.

2.6.3.2. Aprendizaje Supervisado

Este tipo de aprendizaje considera un entrenamiento guiado donde se utiliza información global. En su implementación se presentan dos vectores (uno de entrada y otro de salida deseada). La salida computada por la red se compara con la salida deseada, y los pesos de la red se modifican en el sentido de reducir el error cometido. Se repite iterativamente, hasta que la diferencia entre la salida computada y la deseada sea aceptablemente pequeña, comparada con algún parámetro de error.

2.6.4. Representación de la ANN

El modelo de ANN más utilizado para pronosticar series de tiempo es la Red Neuronal Supervisada alimentada hacia adelante (Llamada también Multilayer Perceptron o Supervised Feed Forward Neural Network en inglés), con una sola capa oculta y un nodo de salida [44]. Esta red se caracteriza por utilizar 3 capas que contienen unidades de procesamiento conectadas hacia adelante, es decir, cada nodo de la capa de entrada (z_1, z_2, \dots, z_P) está conectado a cada nodo de la capa oculta (h_1, h_2, \dots, h_Q) y estos últimos,

a su vez, se conectan con el nodo de la capa de salida que entrega la respuesta (\hat{y}_t) como muestra la figura 2.2

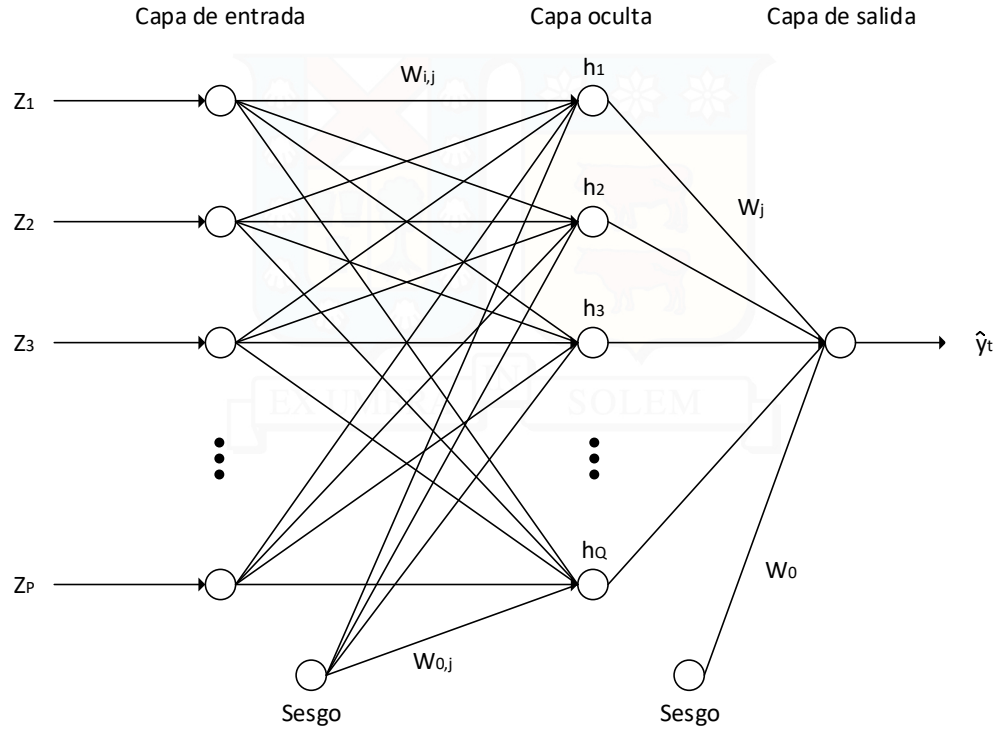


Figura 2.2: Estructura de una Red Neuronal Artificial. Fuente: Elaboración propia.

El modelo representado en la figura 2.2 se expresa matemáticamente en la ecuación 2.94

$$\hat{y}_t = w_0 + \sum_{j=1}^Q w_j f \left(\sum_{i=1}^P w_{i,j} z_i + w_{0,j} \right) + u_t \quad (2.94)$$

Donde \hat{y}_t indica el pronóstico de la variable y en el tiempo t ; la entrada z_i responde a variables exógenas en el tiempo t o rezagadas, variables endógenas, residuales en el tiempo t o rezagados, entre otras (para $i = 1, \dots, P$); $w_{i,j}$ representa el peso desde el nodo i en la capa de entrada hasta el nodo j en la capa oculta (para $i = 1, \dots, P$ y $j = 1, \dots, Q$); w_j representa los pesos desde el nodo j en la capa oculta hasta el nodo \hat{y}_t en la capa de salida; $w_{0,j}$ responde al peso desde la unidad de Sesgo hasta el nodo j en la capa oculta; w_0 representa el peso desde la unidad de Sesgo al nodo de la capa de salida; u_t es el error en el que incurre el modelo en el tiempo t ; f representa la función de transferencia que se aplica

en la capa oculta. Se puede notar que el término

$$f\left(\sum_{i=1}^P w_{i,j}z_i + w_{0j}\right)$$

corresponde a lo que ocurre en una neurona, como se representa en la sección 2.6.1, donde la función de activación se aplica sobre la suma ponderada de todas las entradas a la neurona por sus respectivos pesos, además del peso del sesgo.

Si el modelo que se debe implementar es más general, ampliando la cantidad de capas ocultas o la cantidad de neuronas en la capa de salida, esta representación ya no es válida y, por la complejidad que adquiere la notación, el análisis debe hacerse desde otro enfoque más centrado en las neuronas por cada capa. Considerando la figura 2.3, se generaliza en caso anterior para las capas $l = 1, \dots, L$ donde $l = 1$ es la primera capa oculta y $l = L$ es la capa de salida, la interpretación matemática para la primera capa oculta es:

$$h_j^{(1)} = f^{(1)}\left(\sum_{i=1}^P w_{i,j}^{(1)}z_i + w_{0j}^{(1)}\right) \quad (2.95)$$

donde el superíndice (1) indica que se trata de una variable o función asignada a la capa $l = 1$, es decir: $f^{(1)}$ representa la función de activación en la capa 1; $w_{i,j}^{(1)}$ es el peso del input i hacia el nodo j en la capa 1; $w_{0j}^{(1)}$ es el sesgo para el nodo j en la capa 1; y el término $h_j^{(1)}$ corresponde a la salida de la neurona j en la capa 1. Esto es válido para $j = 1, \dots, Q_1$ donde Q_1 representa la cantidad total de nodos en la capa 1. Entonces, para las siguientes capas ocultas, $l = 2, \dots, L - 1$ la interpretación matemática será:

$$h_j^{(l)} = f^{(l)}\left(\sum_{k=1}^{Q_{l-1}} w_{k,j}^{(l)}h_k^{(l-1)} + w_{0j}^{(l)}\right) \quad (2.96)$$

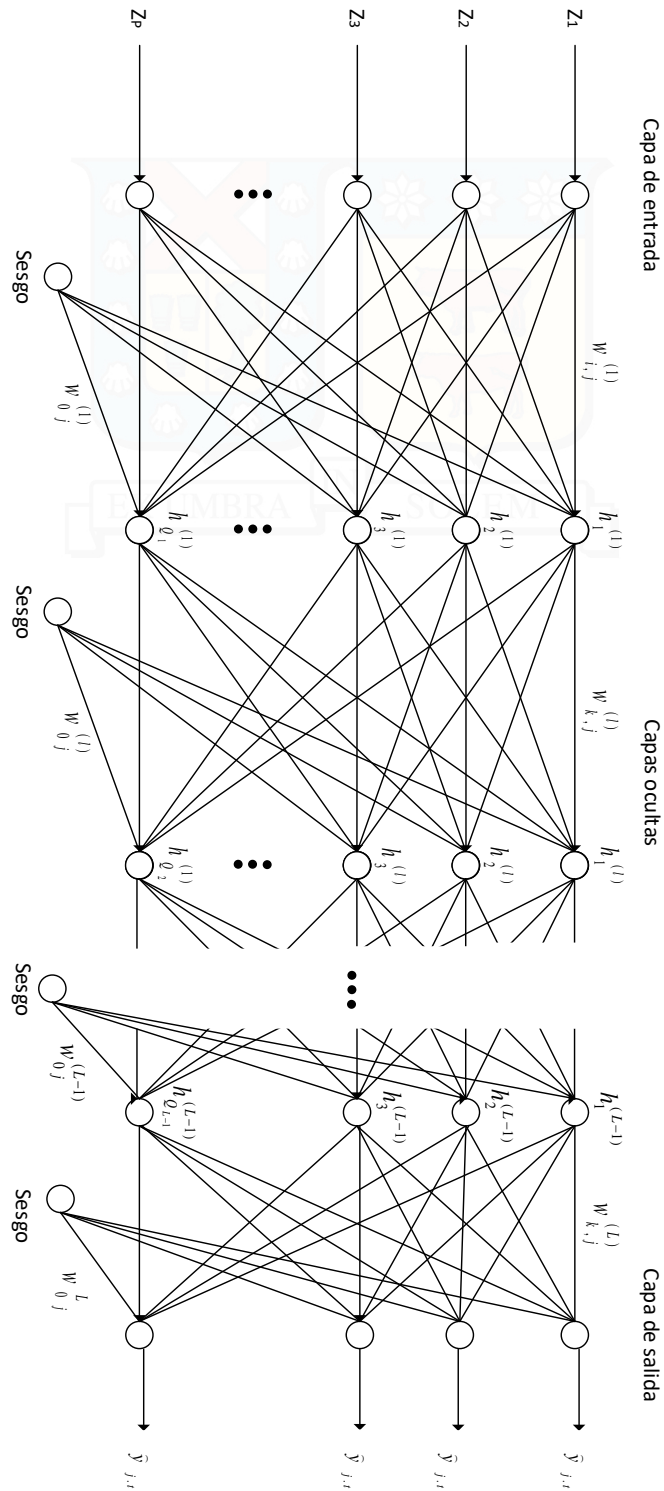


Figura 2.3: Estructura general de una Red Neuronal Artificial Multicapas. Fuente: Elaboración propia.

donde el superíndice (l) indica que se trata de una variable o función asignada a la capa l , es decir: $f^{(l)}$ representa la función de activación en la capa l ; $w_{k,j}^{(l)}$ es el peso que va desde el nodo $k = 1, \dots, Q_{l-1}$ en la capa inmediatamente anterior a l hacia el nodo j en la capa l ; el término $h_k^{(l-1)}$ es la salida de la neurona k en la capa inmediatamente anterior a l ; $w_{0j}^{(l)}$ es el sesgo para el nodo j en la capa l ; y el término $h_j^{(l)}$ corresponde a la salida de la neurona j en la capa l . Esto es válido para $j = 1, \dots, Q_l$ donde Q_l representa la cantidad total de nodos en la capa l . Finalmente, la capa de salida se interpreta como:

$$\hat{y}_{j,t} = h_j^{(L)} = f^{(L)} \left(\sum_{k=1}^{Q_{L-1}} w_{k,j}^{(L)} h_k^{(L-1)} + w_{0j}^{(L)} \right) + u_t \quad (2.97)$$

donde el superíndice (L) indica que se trata de una variable o función asignada a la capa de salida L , es decir: $f^{(L)}$ representa la función de activación en la capa L , que por lo común es lineal; $w_{k,j}^{(L)}$ es el peso que va desde el nodo $k = 1, \dots, Q_{L-1}$ en la capa inmediatamente anterior a L hacia el nodo j en la capa L ; el término $h_k^{(L-1)}$ es la salida de la neurona k en la capa inmediatamente anterior a L ; $w_{0j}^{(L)}$ es el sesgo para el nodo j en la capa L ; y el término $\hat{y}_{j,t} = h_j^{(L)}$ corresponde a la salida de la neurona j en la capa final L . Esto es válido para $j = 1, \dots, Q_L$ donde Q_L representa la cantidad total de nodos en la capa L .

2.6.5. Backpropagation

A partir de un entrenamiento y reconocimiento de patrones, la red es capaz de aprender de la información con la cual se la alimenta. Esto es posible gracias a un algoritmo de aprendizaje llamado *Backpropagation*, propuesto para el uso en Redes Neuronales por Rumelhart et al. [45]. Este algoritmo es el más utilizado en Redes Neuronales multicapas al momento de enfrentar problemas de clasificación y pronósticos de datos debido a la posibilidad de reconocer irregularidades, excepciones y patrones en una muestra de entrenamiento.

Este algoritmo de optimización ajusta los pesos mediante un proceso iterativo que se

basa en el método del gradiente descendente para minimizar la siguiente función de costos

$$C = \frac{1}{T} \sum_{t=1}^T (y_t - \hat{y}_t)^2 \quad (2.98)$$

Donde T es el tiempo total a pronosticar; y_t es el valor real de la serie (o *target* en inglés) en el tiempo t ; \hat{y}_t es el valor pronosticado por la ANN en el tiempo t .

El algoritmo comienza con la inicialización de todos los pesos a un valor aleatorio pero pequeño, cercano a cero. Luego, se toma un vector de la entrada y se lo carga a la Red Neuronal de modo que pase a través de todas las capas, considerando los pesos iniciales, hasta obtener un valor en la salida. Este valor pronosticado se compara al valor real de la serie y el error resultante se devuelve a la Red Neuronal para minimizar la función de costos C al variar el peso $w_{i,j}$ en una cantidad $\Delta w_{i,j}$ de acuerdo a la siguiente ecuación:

$$w_{i,j}^* = w_{i,j} + \Delta w_{i,j} \quad (2.99)$$

Lo que es equivalente a escribir:

$$w_{i,j}^* = w_{i,j} - R \frac{\partial C}{\partial w_{i,j}} \quad (2.100)$$

Donde $w_{i,j}^*$ es el peso ajustado; R responde a la tasa de aprendizaje y la fracción que lo acompaña corresponde a la derivada parcial de la función de costo respecto al peso, que muestra el papel que desempeña el gradiente descendente en este algoritmo. Finalmente, este proceso se repite para cada peso de forma iterativa hasta obtener el mínimo en la función C , esto significa que, al estar bien entrenada, la Red Neuronal provee salidas próximas a los valores deseados cuando se le proporcionan nuevos datos fuera del conjunto de entrenamiento. Estos valores de salida se conocen como pronósticos fuera de la muestra (*out-of-sample* en inglés).

2.7. Medidas de desempeño

Para poder entender si el pronóstico que entrega un modelo es mejor al que entrega otro, es necesario comparar la capacidad predictiva que tiene cada uno. No basta solamente con ver qué pronóstico se acerca más al valor real, sino que se requiere de una validación estadística para ello. Frente a esto, se han desarrollado diversas medidas de desempeño, métricas de error y pruebas estadísticas que permiten establecer evidencia estadística para elegir los resultados de un modelo por sobre otro.

2.7.1. Criterio de Información de Akaike

El Criterio de Información de Akaike (abreviado AIC, por las siglas en inglés para *Akaike Information Criterion*), propuesto por Akaike [46], es una medida de la calidad relativa de los modelos estadísticos para un conjunto de datos determinado. Teniendo un conjunto de modelos para aplicar, el AIC estima la calidad de cada modelo, relativo a cada uno de los modelos restantes del conjunto, lo que implica que puede ser usado como herramienta para la selección de modelos. En específico, el criterio funciona estimando la calidad de un modelo, al representar un cierto conjunto de datos, basándose en la pérdida de información que sufre en relación a la compensación entre su bondad de ajuste y la complejidad que presenta en términos de la cantidad de elementos que utiliza. Si bien este criterio ayuda en la selección de modelos, no funciona en base a una hipótesis nula, por lo que no es capaz de referirse a la calidad de un modelo en términos absolutos. Si todos los modelos candidatos en la prueba se ajustan mal a los datos, el AIC no da advertencia de eso, sino de cuál es mejor entre ellos.

Dado un conjunto de modelos candidatos, el de mejor calidad, y por lo tanto el elegido, será aquel que presente el menor valor de AIC. Este valor de AIC para cada modelo se puede obtener de la siguiente expresión:

$$AIC = 2k - 2 \log(\hat{f}(y_i|\hat{\theta})) \quad (2.101)$$

donde k es el número de parámetros estimados en el modelo; $\hat{f}(y_t|\hat{\theta})$ es el valor máximo de la función de verosimilitud del modelo, con el estimador de máxima verosimilitud $\hat{\theta}$; y \log representa la función logaritmo natural. Según la expresión 2.101, el AIC recompensa la bondad de ajuste (representada por la función de verosimilitud), pero también penaliza por el número de parámetros estimados.

2.7.2. Función de Pérdida

En estadística se usa principalmente para estimar parámetros, pero también para comparar resultados ya que cuantifica la cantidad por la cual una predicción se desvía del valor real. Existen muchas funciones de pérdida para considerar, pero su elección dependerá del problema específico que se esté trabajando y de los datos que se manejen. Entre las más utilizadas se encuentra la función de errores cuadrados (abreviada SE por la sigla en inglés para *Squared Errors*) que se representa del siguiente modo

$$SE = (y_t - \hat{y}_t)^2 \quad (2.102)$$

donde la diferencia entre el valor real y el pronóstico es el error de predicción. Por otro lado, también es muy utilizada la función del promedio de los errores cuadrados (abreviada MSE por la sigla en inglés para *Mean Squared Errors*) que se representa del siguiente modo:

$$MSE = \frac{1}{T} \sum_{t=1}^T (y_t - \hat{y}_t)^2 \quad (2.103)$$

Donde T es el total de datos pronosticados; y_t es el valor real de la serie (o *target* en inglés) en el tiempo t ; \hat{y}_t es el valor pronosticado en el tiempo t . Una vez que se tiene el valor de la función de pérdida, es posible comparar la capacidad predictiva de cada modelo y determinar cuál es el mejor basándose en el que entrega el menor valor de la función de pérdida, en este caso SE o MSE .

2.7.3. Pruebas estadísticas

Tener la variación porcentual de la función de pérdida de un modelo sobre otro no es suficiente y se hace necesario establecer la significancia estadística de dicha diferencia. Frente a esto se emplea el uso de 2 pruebas estadísticas mencionadas en la literatura.

2.7.3.1. Prueba Diebold-Mariano

La prueba estadística de Diebold-Mariano (DM), propuesta por Diebold & Mariano [47], se usa para verificar la igualdad de precisión entre 2 modelos al pronosticar bajo el supuesto generales y basándose en la función de pérdida definida como la diferencia entre los residuales al cuadrado de cada modelo. En particular, esta prueba intenta verificar la hipótesis nula

$$H_0 : E(d) = 0 \quad (2.104)$$

mediante la implementación del estadístico

$$DM = \frac{\bar{d}}{\sqrt{\frac{1}{T} \text{var}(d)}} \sim N(0, 1) \quad (2.105)$$

Donde $d = e_1^2 - e_2^2$ es la diferencia entre el residual del modelo 1 al cuadrado y el residual del modelo 2 al cuadrado; \bar{d} es el promedio de d , $\text{var}(d)$ es la varianza de d , T es la cantidad de residuales en cada modelo y se asume que este diferencial de pérdida d es estacionario.

Los valores que entregue este test indicarán estadísticamente si existe o no igualdad de predicción entre los modelos. Por ejemplo, si este valor es comparado con el valor crítico de $\pm 1,96$ para una distribución Normal y se encuentra fuera de estos límites se puede asegurar, para un cierto nivel de significancia, que la hipótesis nula de igualdad se rechaza y, por lo tanto, un modelo es mejor que el otro. De ser este el caso, el signo del valor entregado por el estadístico indicará cuál de los dos modelos es mejor, ya que un valor negativo implica que el residual al cuadrado del segundo modelo es mayor al del primero (considerando la disposición en d) y, por lo tanto, se concluye que el modelo 1 es más preciso al entregar un error menor. Por extensión, si el valor de la prueba DM es positivo y fuera de los límites del

valor crítico, se concluye que el modelo 2 es estadísticamente más preciso que el modelo 1.

2.7.3.2. Model Confidence Set

El Model Confidence Set (MCS), propuesto por Hansen et al. [48] es una mejora del modelo Superior Predictive Ability (SPA) propuesto por Hansen [49] para mostrar robustez estadística en los resultados. El SPA está diseñado para verificar si algún modelo perteneciente a un conjunto M es significativamente mejor para realizar pronósticos que un determinado modelo de referencia contra el cual se compara. En particular, la hipótesis nula es que el modelo de referencia no es inferior a alguna de las $k = 1, \dots, m$ alternativas dentro del conjunto M . El rechazo de la hipótesis nula solo identifica uno o más modelos como significativamente superiores al modelo de referencia. Entonces, la prueba SPA ofrece una guía para identificar los modelos que pertenecen a M^* , es decir, el conjunto de elementos superiores que contiene un número finito de elementos que cumplen la hipótesis nula, pero no especifica cuál de ellos es el mejor en dicho conjunto.

A diferencia del SPA, la prueba MCS no requiere identificar un modelo de referencia, ya que todos los modelos sirven de referencia al realizar una serie de comparaciones. Además, esta prueba tiene la ventaja de que puede ser utilizada para seleccionar el mejor modelo. En particular, una secuencia de pruebas SPA definen el MCS como el conjunto de modelos de referencia que no son significativamente inferiores a las alternativas, excluyendo a los modelos inferiores. Aquí, la hipótesis nula es

$$H_0 : E(d_{ij}) = 0 \quad (2.106)$$

donde $d_{ij} = L_i - L_j$, con L representando la función de pérdida para los modelos i y j en el conjunto M , respectivamente. El objetivo de la prueba MCS es determinar el conjunto M^* donde un modelo i es preferido, en términos de la esperanza de la diferencia entre funciones de pérdida, a un modelo alternativo j si $E(d_{ij}) < 0$ para todo j en el conjunto M que contiene a todos los modelos a analizar. Esto se realiza a través de una secuencia de pruebas de significancia, donde elementos que son significativamente inferiores a otros

elementos de M en términos de predicción son eliminados para determinar finalmente el conjunto M^* .



CAPÍTULO 3

METODOLOGÍA

En el presente capítulo se presenta y explica la metodología desarrollada durante el estudio. En el primer punto, 3.1, se muestran los datos utilizados durante el análisis. Posteriormente, en el punto 3.2 se describe el modelo híbrido propuesto, las ecuaciones que lo gobiernan y los parámetros utilizados.

3.1. Datos

Los datos utilizados en el presente estudio corresponden a los tipos de cambio Euro respecto al dólar americano (EUR/USD), Libra Esterlina respecto al dólar americano (GBP/USD) y Yen respecto al dólar americano (JPY/USD) en términos diarios para el periodo comprendido entre el 4 de Enero de 1999 y el 30 de Diciembre del 2015 ¹, abarcando un total de 4242 datos para cada paridad cambiaria luego de filtrar días no hábiles, feriados y festivos donde la bolsa no transa. Estos datos se representan gráficamente mediante su evolución temporal en las figuras 3.1, 3.2 y 3.3.

¹Los datos fueron extraídos de la página del Banco Central de Chile: <https://goo.gl/jGZCbo>

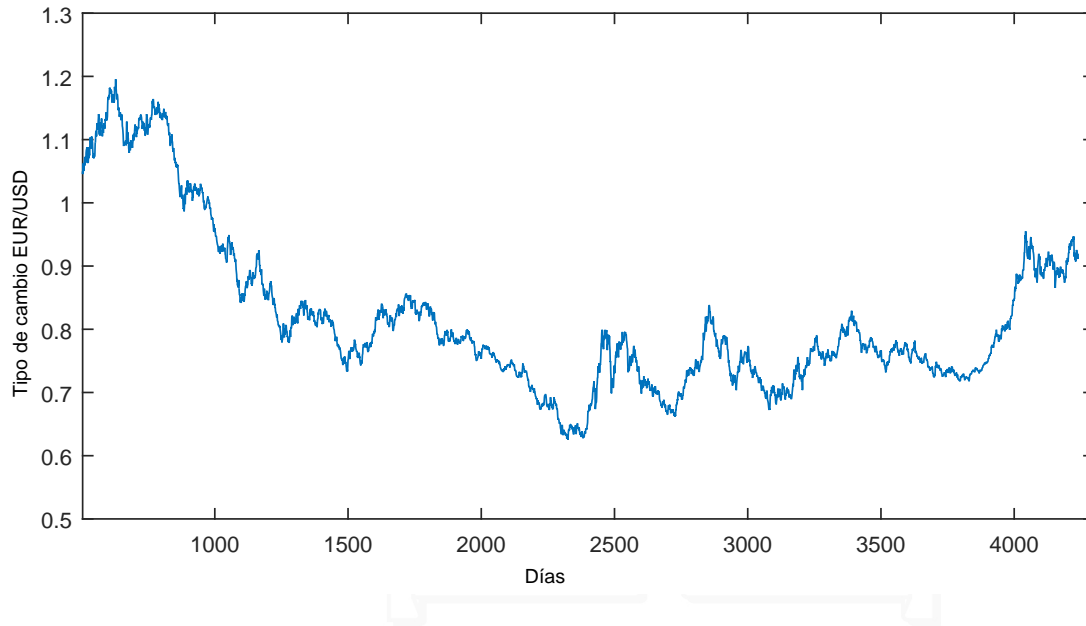


Figura 3.1: Evolución del tipo de cambio EUR/USD en el tiempo expresado en días. Fuente: Elaboración propia.

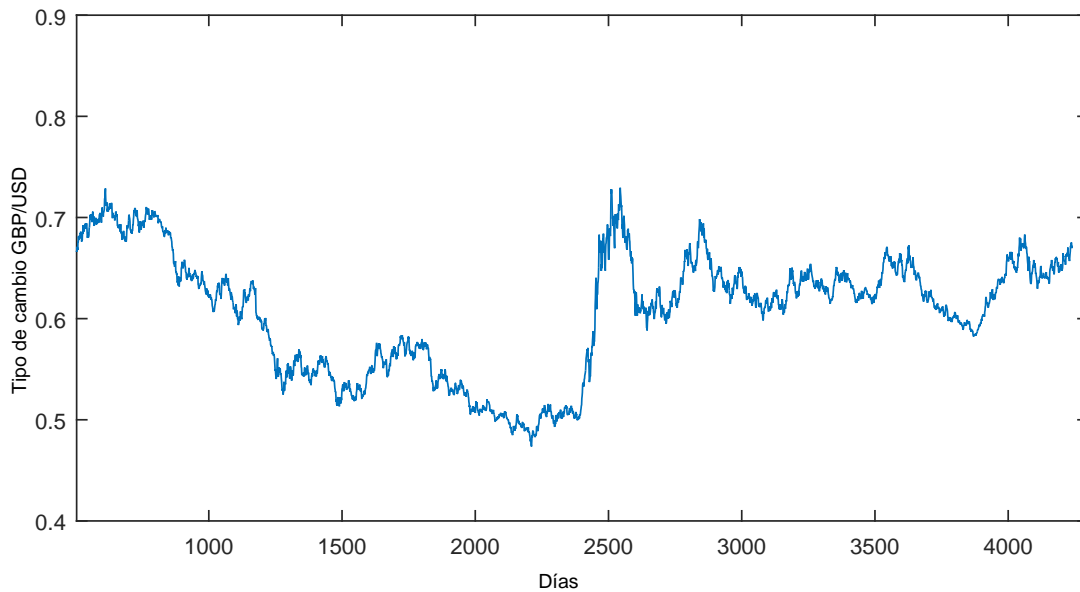


Figura 3.2: Evolución del tipo de cambio GBP/USD en el tiempo expresado en días. Fuente: Elaboración propia.

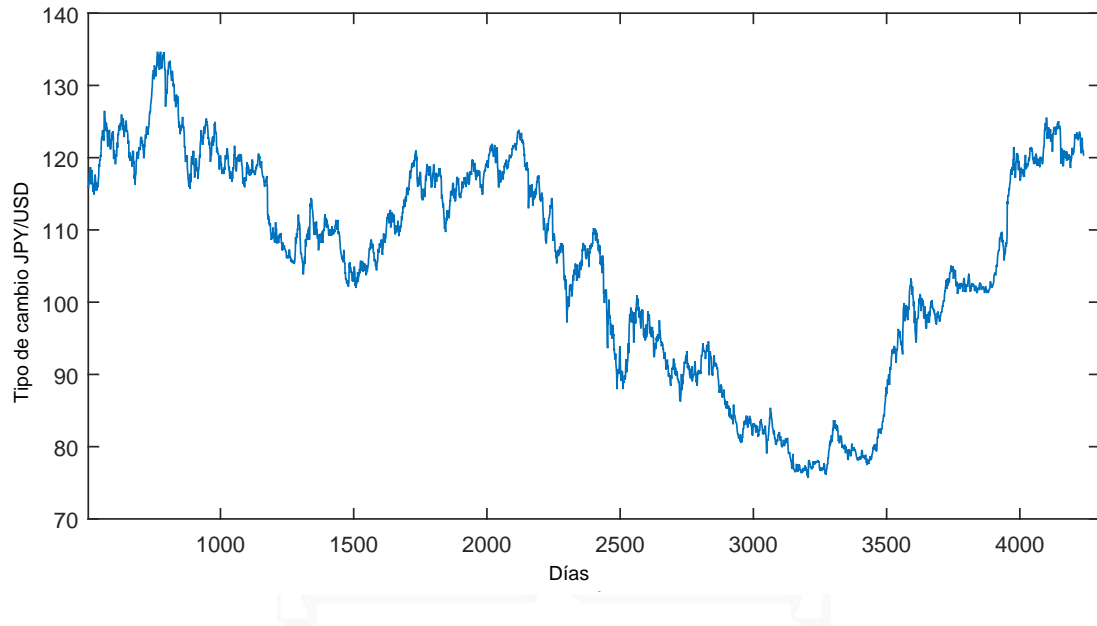


Figura 3.3: Evolución del tipo de cambio JPY/USD en el tiempo expresado en días. Fuente: Elaboración propia.

En las figuras 3.1, 3.2 y 3.3 se puede notar que la tendencia de las tres paridades era hacia una disminución hasta cerca del día 2500, que corresponde a mediados del año 2008 como consecuencia de la crisis Subprime. Luego de este punto, el tipo de cambio tiende a aumentar, excepto para el caso JPY/USD que continúa a la baja hasta cerca del día 3500, donde la tendencia cambia al alza.

Para modelar, las series se usan en términos del logaritmo natural en el tiempo t , como se describió al final de la sección 2.4. Se define entonces

$$Y_{kt} = \log(X_{kt}) \quad (3.1)$$

donde $\log()$ representa la función logaritmo natural y X_{kt} corresponde al valor del precio del tipo de cambio k en el tiempo t , con $k = \{1, 2, 3\}$ representando las series EUR/USD, GBP/USD y JPY/USD, respectivamente. El retorno, o variación, del tipo de cambio entre dos días consecutivos se define como

$$\Delta Y_{kt} = Y_{kt} - Y_{kt-1} = y_{kt} \quad (3.2)$$

donde Y_{kt-1} responde al valor del logaritmo natural de la serie k en el día anterior a t . Estas transformaciones se pueden ver en las figuras 3.4, 3.5 y 3.6.

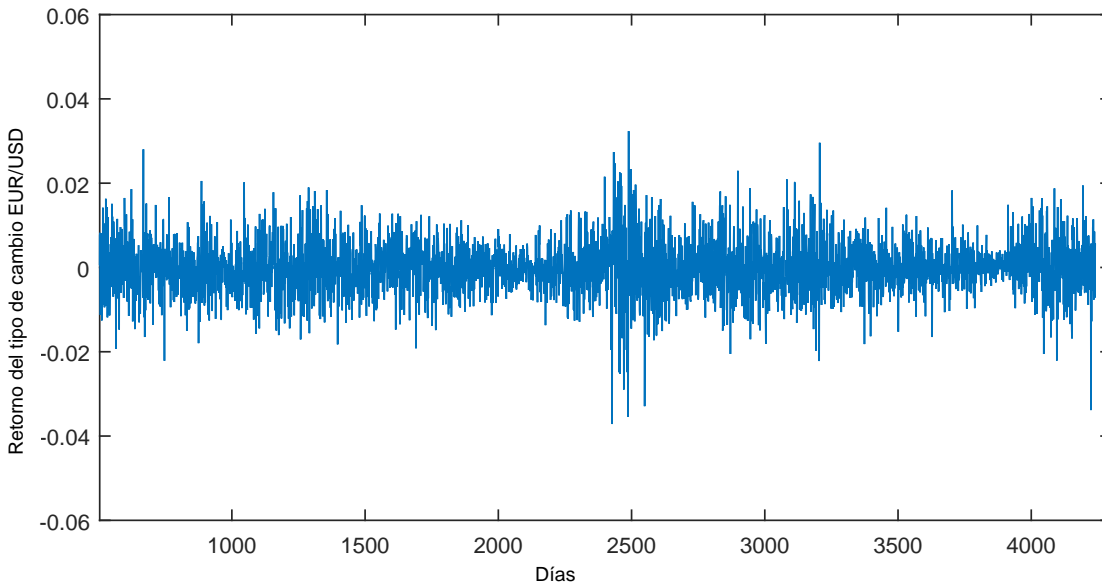


Figura 3.4: Evolución del retorno del tipo de cambio EUR/USD en el tiempo expresado en días.
Fuente: Elaboración propia.

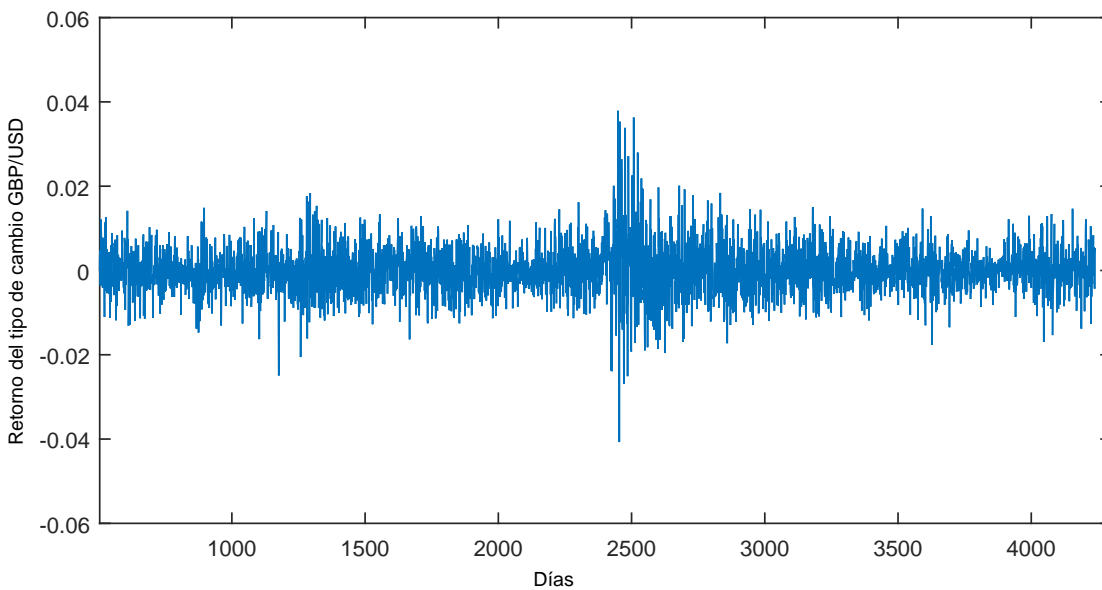


Figura 3.5: Evolución del retorno del tipo de cambio GBP/USD en el tiempo expresado en días.
Fuente: Elaboración propia.

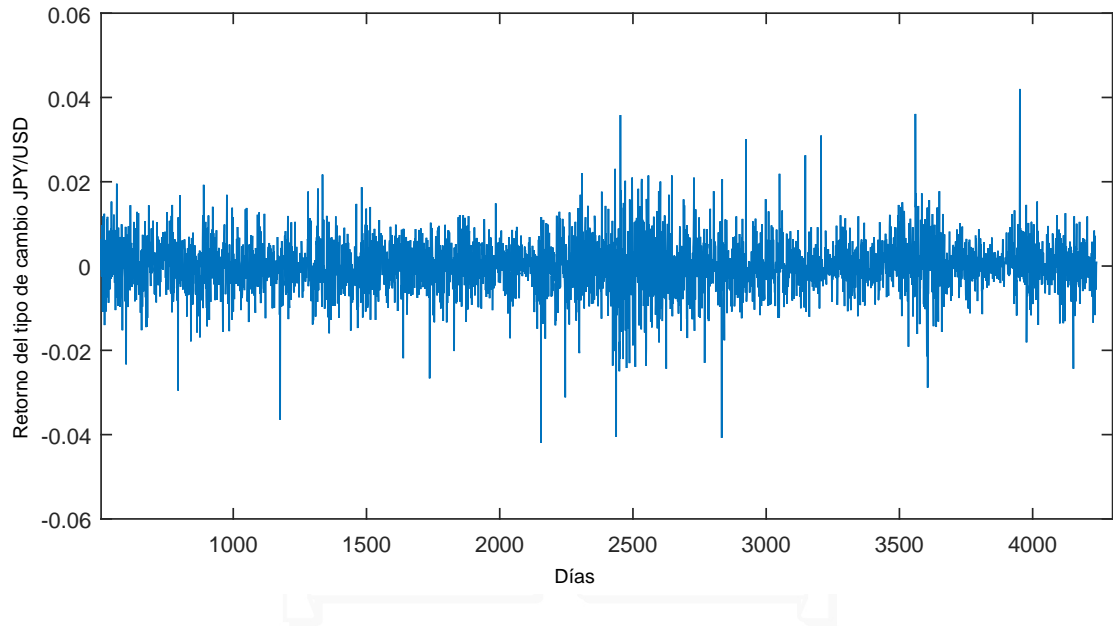


Figura 3.6: Evolución del retorno del tipo de cambio JPY/USD en el tiempo expresado en días. Fuente: Elaboración propia.

En las figuras 3.4, 3.5 y 3.6 se aprecia la evolución del retorno de cada tipo de cambio en el tiempo. Es posible notar que los valores se mantienen dentro de un rango estable pero, como en las figuras 3.1, 3.2 y 3.3, los valores cercanos al día 2500 son interesantes para analizar porque existen variaciones que salen de esos márgenes de acuerdo al movimiento del tipo de cambio en esa fecha correspondiente a la crisis Subprime.

Durante la metodología, se trabajará con estas 3 series de tiempo en términos de la diferencia de logaritmos naturales y_{kt} pero la aplicación de la predicción se realiza solamente sobre la serie y_{1t} , del retorno del EUR/USD.

3.2. Modelo Propuesto

El modelo propuesto corresponde a un híbrido para pronosticar un paso hacia adelante el retorno del tipo de cambio, basándose en la idea de que la combinación de diferentes modelos unifican las características propias de cada uno, permitiendo que se rescate más información de la serie de tiempo. Esto lleva a incrementar la precisión de la predicción con respecto a la utilización de un único modelo aislado.

Específicamente, se propone combinar un modelo lineal, VAR con Corrección del error (VECM), con un modelo no lineal, la Red Neuronal Supervisada Multicapas Alimentada hacia adelante (o *Multilayer Supervised Feed Forward Neural Network* en inglés) aplicando como algoritmo de aprendizaje un *Backpropagation*, y formando un modelo VEC-ANN. Al trabajar con series de tiempo que presentan una componente no lineal, como el tipo de cambio, los modelos econométricos lineales VAR y VECM no son suficientes para recoger toda la información que presenta la serie y, por lo tanto, se vuelve necesario combinar estos modelos con otros que sí puedan reconocer esta componente no lineal, como las redes neuronales, para rescatar más información y generar pronósticos más precisos que los que podría generar el modelo econométrico de referencia al ser utilizado en solitario.

Es importante notar que este modelo se realiza a través de ventanas corredizas de corte diario con el fin de agregar dinamismo en el modelo al considerar el movimiento de la serie en el tiempo, para evitar centrarse solo en un gran marco temporal de referencia. Se considerarán distintos tamaños de ventana corrediza para encontrar aquella que entregue los mejores resultados. Estos tamaños son 42, 63, 84, 126, 252 y 378 días. En términos prácticos, para que no parezca tan abstracta esta elección, estas ventanas representan 2 meses, 3 meses, 1 trimestre, 1 semestre, 1 año, y 1 año y medio, respectivamente. El procedimiento desarrollado para pronosticar el tipo de cambio aplicando el modelo VEC-ANN por ventanas corredizas se describe paso a paso a continuación:

Paso 1: El primer paso consiste en obtener los datos del precio del tipo de cambio para las paridades EUR/USD, GBP/USD y JPY/USD, filtrarlos para los días hábiles, transformarlos aplicando logaritmo natural y pre-procesarlos como el retorno del tipo de cambio según se describe en el punto 3.1. Luego de esto, el tamaño de datos se reduce de 4242 a 4241, ya que el primer elemento se pierde al hacer la primera diferencia.

Paso 2: El siguiente paso consiste en aplicar la prueba de raíz unitaria, según proponen Dickey & Fuller [34], a las series para verificar su estacionariedad, como se

describe en el punto 2.4.1.

Paso 3: Una vez que se comprueba la estacionariedad de las series en su primera diferencia, es decir, en los retornos del tipo de cambio EUR/USD, GBP/USD y JPY/USD, respectivamente, se verifica la existencia de cointegración considerando una relación entre las 3 series en cada una de las 6 ventanas corridizas. Para esto, se aplican las pruebas que plantean Johansen [39] y Johansen & Juselius [40], descritas en el punto 2.5.3.3. Basta con que una de las 2 pruebas implementadas, Traza o Máximo Valor Propio, muestre la existencia de vectores de cointegración para que el modelo de Corrección del Error sea considerado en el análisis de esa ventana.

Paso 4: Con el análisis de cointegración hecho para cada ventana corridiza, se aplica el modelo econométrico que corresponda a los resultados, es decir, si las pruebas de Johansen indican que existen vectores de cointegración para una determinada ventana, entonces el modelo aplicado a dicha ventana será el VECM y, de lo contrario, cuando no existan vectores de cointegración el modelo aplicado será el VAR. Estos modelos econométricos se describen en las secciones 2.5.3 (ecuación 2.30) y 2.5.2 (ecuación 2.88), respectivamente. Por ejemplo, para la ventana de tamaño 252, se analiza cointegración desde el día 1 hasta el día 252 y dependiendo del resultado de las pruebas esa ventana se modela con VAR o VECM para pronosticar el día 253. Luego, la ventana se corre un día y toma los datos desde el día 2 hasta el 253 (este dato 253 corresponde al valor de la serie y no al pronóstico de ese día obtenido antes), se vuelve a analizar cointegración y, dependiendo del resultado, se vuelve a asignar un modelo VAR o VECM a este conjunto de datos para pronosticar el día 254, y así sucesivamente hasta finalizar la muestra y tener todos los pronósticos del modelo VAR-VECM. La cantidad de rezagos utilizados en cada modelo para cada ventana se determinará a través del criterio AIC, descrito en el punto 2.7.1, considerando como máximo de 8 rezagos.

Paso 5: El Paso 4 se repite para cada uno de los 6 tamaños de ventana obteniendo $4241 - w$ pronósticos en cada una, donde w es el tamaño de la ventana considerada, ya que

los primeros elementos de cada serie se pierden al modelar. Así, para el caso de la ventana de tamaño 252, se tendrían 3989 pronósticos.

Paso 6: Teniendo el conjunto de pronósticos del modelo VAR-VECM para el total de datos en cada tamaño de ventana, se debe aplicar la función de pérdida MSE, descrita en el punto 2.7.2 (ecuación 2.103), entre los pronósticos de cada tamaño de ventana y los valores reales de la serie de retorno EUR/USD para determinar el tamaño de ventana que presenta el menor valor de la función de pérdida. El tamaño de ventana que cumpla este objetivo será elegido como el mejor de los modelos VAR-VECM y se utilizará más adelante como punto de comparación. La razón para aplicar el MSE a los pronósticos y al valor real de la serie de retorno EUR/USD es que la aplicación del modelo propuesto va enfocada a esta serie en particular. Así, la expresión 2.103 en este caso particular se describe como

$$MSE = \frac{1}{T} \sum_{t=1}^T (y_{1t} - \hat{y}_{1t})^2$$

Paso 7: Habiendo seleccionado el mejor modelo VAR-VECM según MSE, se rescatan los residuales del modelo en cada ventana, ε_{kt} donde $k = \{1, 2, 3\}$ representa el retorno del tipo de cambio EUR/USD, GBP/USD y JPY/USD, respectivamente. En particular, se necesitan los residuales para el retorno del tipo de cambio EUR/USD, ε_{1t} , que es el caso de interés para la aplicación.

Paso 8: El siguiente paso es la aplicación de Inteligencia Artificial capaz de captar la relación no lineal entre las variables, en particular se implementa la Red Neuronal Artificial Supervisada Multicapas con Alimentación hacia adelante que se describe en la sección 2.6.4. Para la obtención del pronóstico de la serie en un tiempo $t + 1$, la alimentación de esta ANN considera los retornos para el tipo de cambio EUR/USD en el tiempo t y sus rezagos, además de los residuales del mejor modelo VAR-VECM, obtenidos en el **Paso 7**, para el tiempo t y sus rezagos. Al igual que en el caso del modelo econométrico, aquí también se realiza una sensibilización por ventanas corredizas para determinar la mejor opción en cuanto a la cantidad

de datos pasados de cada variable que alimentarán a la ANN, considerando los tamaños de ventana 42, 63, 84, 126, 252 y 378 días. Para ser consistentes con el máximo impuesto al modelo VAR-VECM, se determina que la Red Neuronal se alimentará con un máximo de 8 rezagos de cada variable.

Paso 9: Además de realizar una sensibilización en la cantidad de datos que alimentan a la Red Neuronal por ventana corrediza, también se hace una sencibilización del número de capas ocultas y del número de neuronas en cada una de estas capas. En particular, se considerarán diferentes condiguraciones de la ANN donde el número de capas ocultas, l , puede ser: 1, 2, 3 o 4; mientras que el número de neuronas en cada capa oculta, Q_l puede ser: 5, 10, 15, o 20. Además, se estudiará el caso donde solo existe una neurona en la capa de salida representado en la figura 3.7. Las ecuaciones 2.95, 2.96 y 2.97 se reescriben para este caso como 3.3, 3.4 y 3.5, respectivamente:

$$h_j^{(1)} = f^{(1)} \left(\sum_{i=1}^a w_{i,j}^{(1)} y_{1,t-i} + \sum_{i=a+1}^{a+b} w_{i,j}^{(1)} e_{1,t+a-i} + w_{0j}^{(1)} \right) \quad (3.3)$$

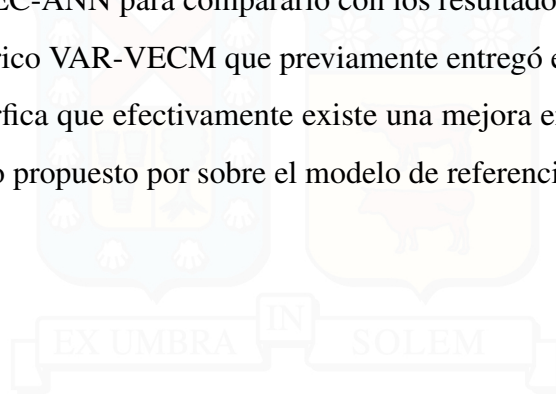
$$h_j^{(l)} = f^{(l)} \left(\sum_{k=1}^{Q_{l-1}} w_{k,j}^{(l)} h_k^{(l-1)} + w_{0j}^{(l)} \right) \quad (3.4)$$

$$\hat{y}_t = h_t^{(L)} = w_0^{(L)} + f^{(L)} \left(\sum_{k=1}^{Q_{L-1}} w_k^{(L)} h_k^{(L-1)} \right) + u_t \quad (3.5)$$

donde se mantienen las variables descritas previamente en la sección 2.6.4 y además se incluyen: la variable de entrada $y_{1,t-i}$ correspondiente al retorno del tipo de cambio EUR/USD en el tiempo $t-i$ (para $i = 1, \dots, a$); la variable de entrada e_{t+a-i} correspondiente a los residuales del modelo VAR-VECM para el tiempo $t+a-i$ (para $i = a+1, \dots, a+b$); y los parámetros a y b son enteros que corresponden al número de rezagos en la serie de retorno y en los residuales, respectivamente, donde $a + b = P$. La razón para considerar los residuales (e) en este análisis es que ellos contienen la relación no lineal que los modelos econométricos no son capaces de reconocer [30], ayudando a obtener un pronóstico más acertado que el

que produce el modelo de referencia.

Paso 10: Finalmente, la métrica de error MSE, descrita en el punto 2.7.2, y las pruebas estadísticas explicadas en el punto 2.7.3, se aplican a los resultados obtenidos en el modelo híbrido VEC-ANN para compararlo con los resultados obtenidos en el modelo econométrico VAR-VECM que previamente entregó el menor MSE. De este modo, se verifica que efectivamente existe una mejora en la capacidad predictiva del modelo propuesto por sobre el modelo de referencia.



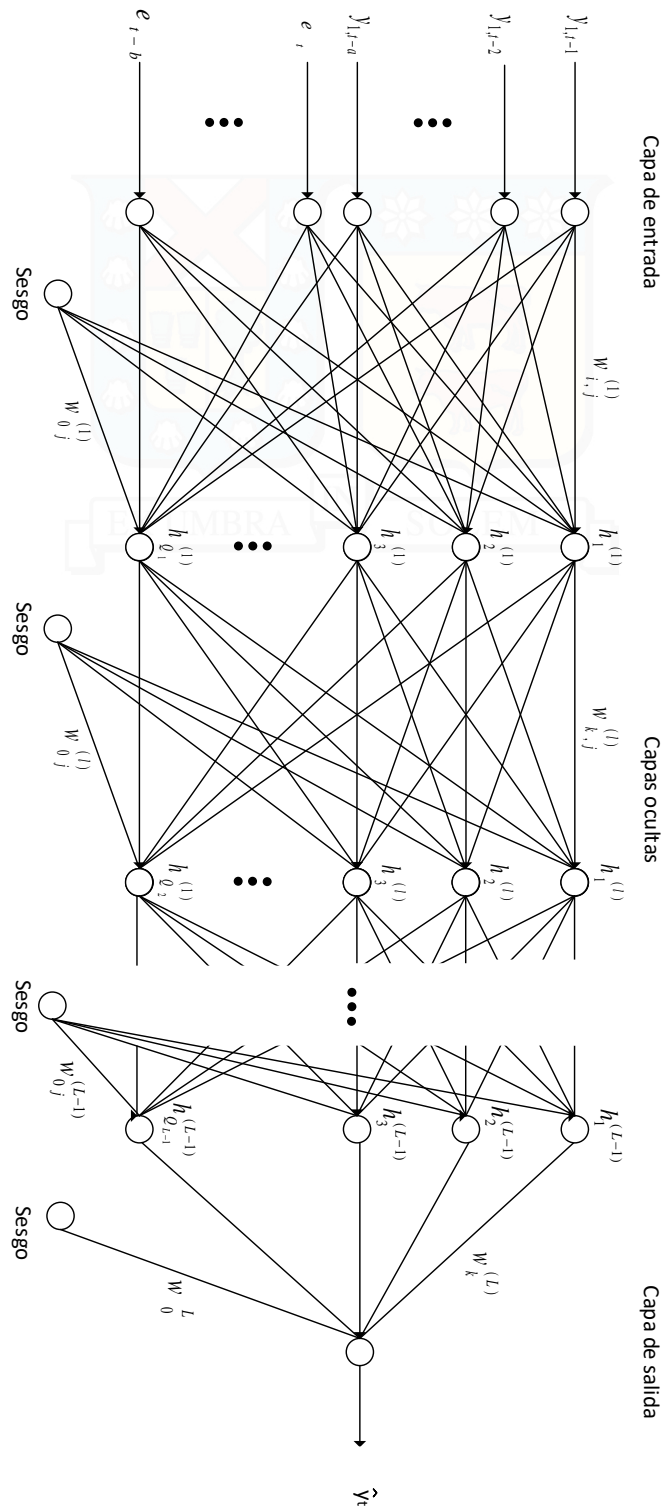


Figura 3.7: Estructura de Red Neuronal Artificial Multicapas implementada. Fuente: Elaboración propia.



CAPÍTULO 4

RESULTADOS Y DISCUSIÓN

La metodología propuesta para mejorar la precisión en los pronósticos del retorno de tipo de cambio entrega los siguientes resultados, expuestos de acuerdo al orden secuencial en el que se deben obtener al seguir el planteamiento del punto 3.2:

Paso 1: Al filtrar, transformar y pre-procesar los datos del precio del tipo de cambio, los retornos del tipo de cambio EUR/USD, GBP/USD y JPY/USD, a los cuales se aplica la metodología propuesta, se representan en las figuras 3.4, 3.5, 3.6, respectivamente.

Paso 2: La prueba de raíz unitaria ADF, propuesta por Dickey & Fuller [34] y representada en la tabla 4.1, muestra que el valor-p de cada una de las series de precio es muy alto y no permite rechazar la hipótesis nula de que existe raíz unitaria para un 5 % de significancia. Por el contrario, el valor-p que la prueba asigna a cada serie de retorno de tipo de cambio es 0,000, rechazado la hipótesis nula para un 5 % de significancia. Por lo tanto, efectivamente las series de tipo de cambio EUR/USD, GBP/USD y JPY/USD son estacionarias al aplicar la transformación de la diferencia de los logaritmos naturales, es decir, son de orden $I(1)$ y se deben utilizar como retornos.

Paso 3: Debido a que las pruebas de cointegración, propuestas por Johansen [39] y Johansen & Juselius [40], se aplican para cada una de las 6 ventanas corridizas en

Prueba de Raíz Unitaria Dickey-Fuller Aumentada
Hipótesis Nula: La serie tiene raíz unitaria

Serie de Precio	Estadístico t*	Valor-p	Serie de Retorno	Estadístico t*	Valor-p
EUR/USD	-1,211	0,6714	EUR/USD	-27,719	0,0000
GBP/USD	-1,848	0,3572	GBP/USD	-30,007	0,0000
JPY/USD	-1,209	0,6723	JPY/USD	-30,298	0,0000

*Valor crítico para la prueba al 5 %: -2,8647

TABLA 4.1: Estadístico t y Valor-p para cada serie en la prueba ADF. Fuente: Elaboración propia.

estudio a lo largo de todo el tamaño de la muestra, no es posible exponer los resultados de cada prueba y solamente se puede mostrar una forma condensada en la tabla 4.2. Es posible notar que a medida que el tamaño de ventana disminuye, la cantidad de ventanas que presentan cointegración es mayor, lo cual es lógico ya que hay más ventanas para realizar la prueba, entonces en esas ventanas es donde se aplicará el modelo VECM en lugar del VAR. El porcentaje de ventanas cointegradas va entre 13,39 % para la ventana de 378 días y 35,91 % para la ventana de 42 días.

Cantidad de ventanas que presentan Cointegración según Johansen

Tamaño de Ventana (días)	378	252	126	84	63	42
Ventanas cointegradas	568	778	884	992	1245	1523
Porcentaje del total*	13,39 %	18,34 %	20,84 %	23,39 %	29,36 %	35,91 %

*Total de la muestra: 4241 datos

TABLA 4.2: Cantidad de ventanas que presentan Cointegración para cada tamaño de ventana corregida. Fuente: Elaboración propia

Paso 4 y 5: La aplicación de los modelos econométricos, VAR y VECM, entregan los pronósticos para el retorno del tipo de cambio EUR/USD en las ventanas corregidas de tamaño 42, 63, 84, 126, 252 y 378, representados en las figuras 4.1, 4.2, 4.3, 4.4, 4.5 y 4.6, respectivamente. Si bien se puede apreciar que a medida que aumenta el tamaño de la ventana el pronóstico del modelo se vuelve más plano gráficamente, esta observación no es concluyente para

señalar que los resultados de ventanas más pequeñas son los mejores, porque estadísticamente pueden no serlo.

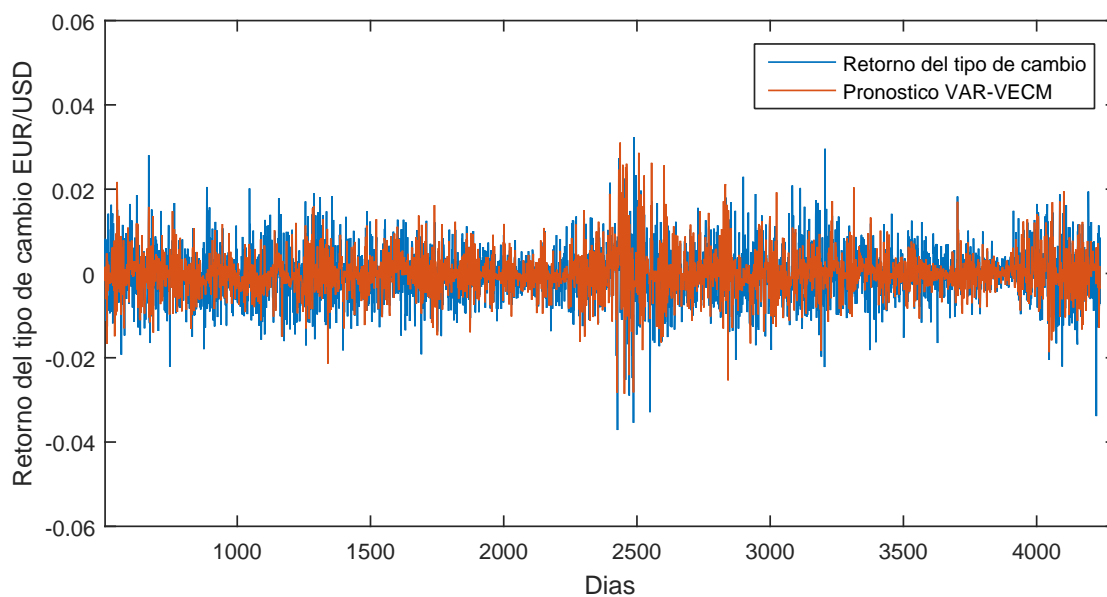


Figura 4.1: Comparación entre la serie del retorno de tipo de cambio EUR/USD y su pronóstico generado por el modelo VAR-VECM en la ventana de 42 días. Fuente: Elaboración propia.

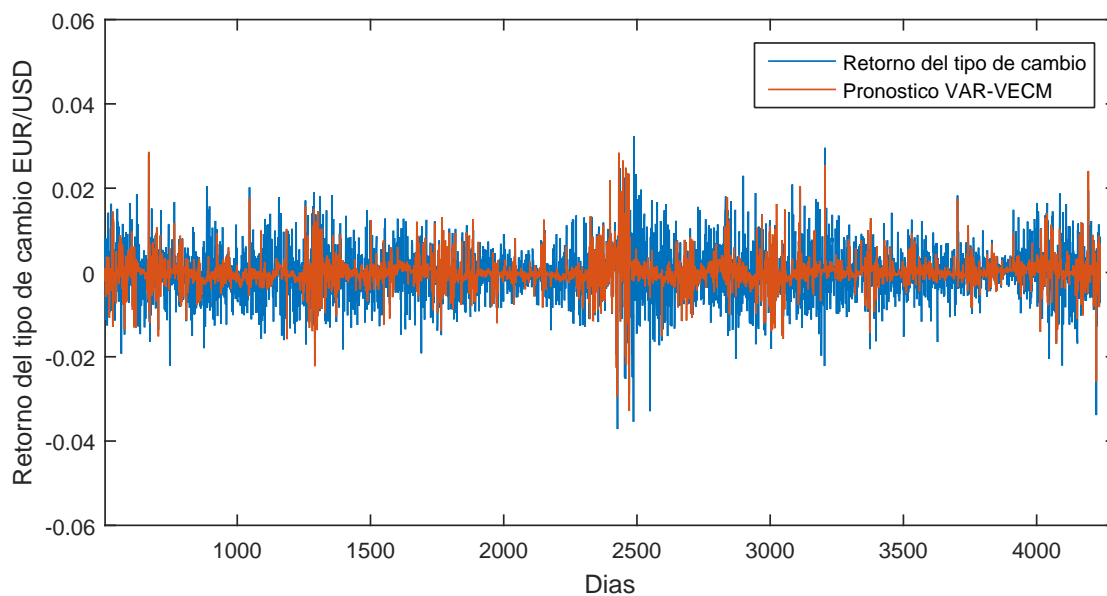


Figura 4.2: Comparación entre la serie del retorno de tipo de cambio EUR/USD y su pronóstico generado por el modelo VAR-VECM en la ventana de 63 días. Fuente: Elaboración propia.

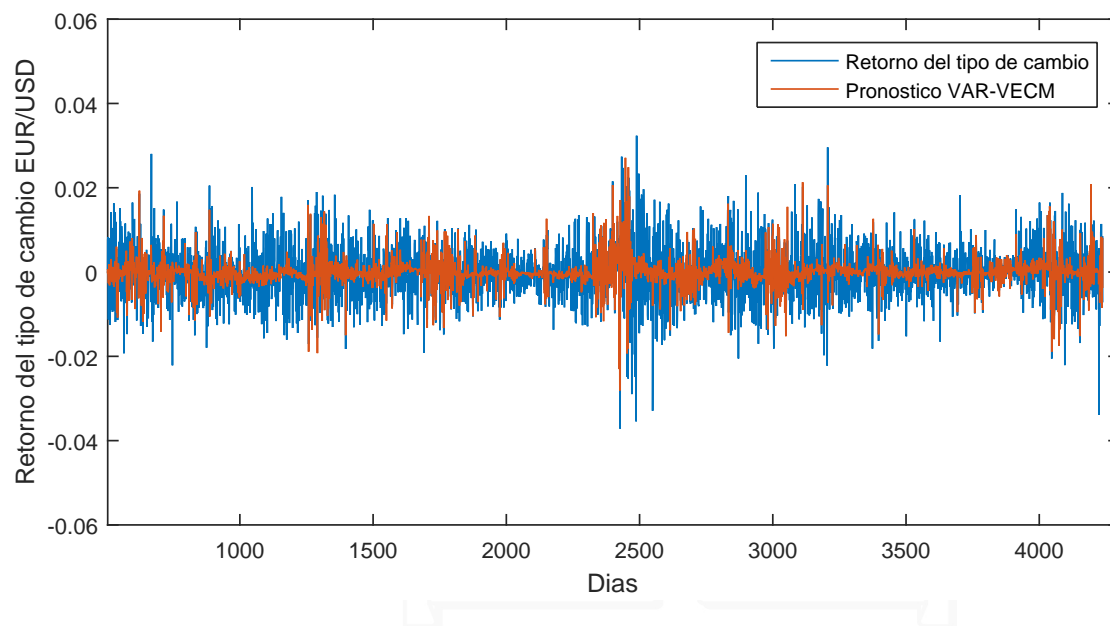


Figura 4.3: Comparación entre la serie del retorno de tipo de cambio EUR/USD y su pronóstico generado por el modelo VAR-VECM en la ventana de 84 días. Fuente: Elaboración propia.

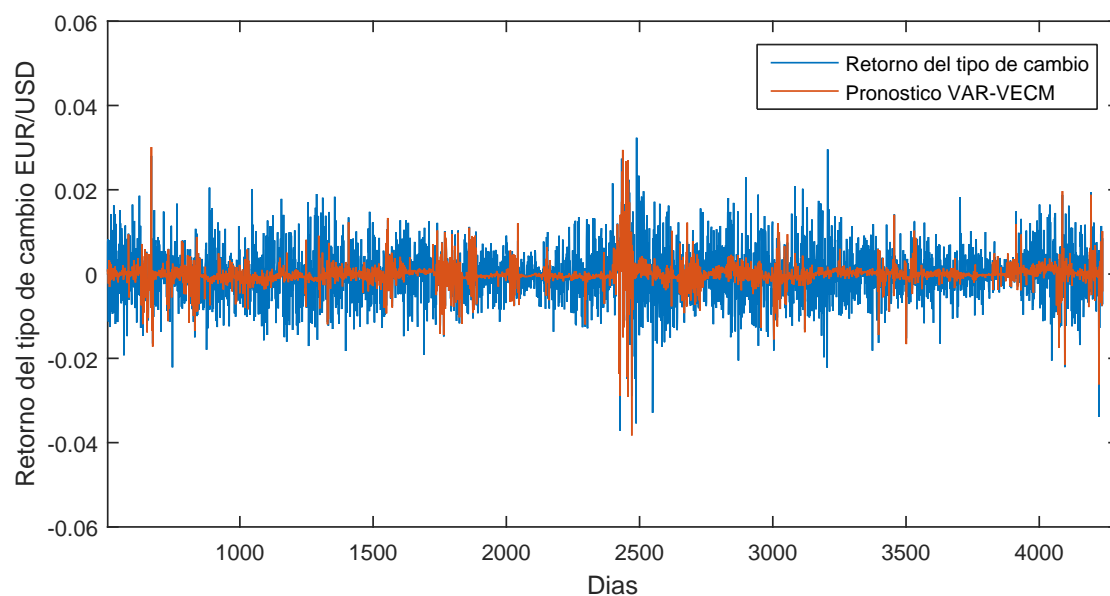


Figura 4.4: Comparación entre la serie del retorno de tipo de cambio EUR/USD y su pronóstico generado por el modelo VAR-VECM en la ventana de 126 días. Fuente: Elaboración propia.

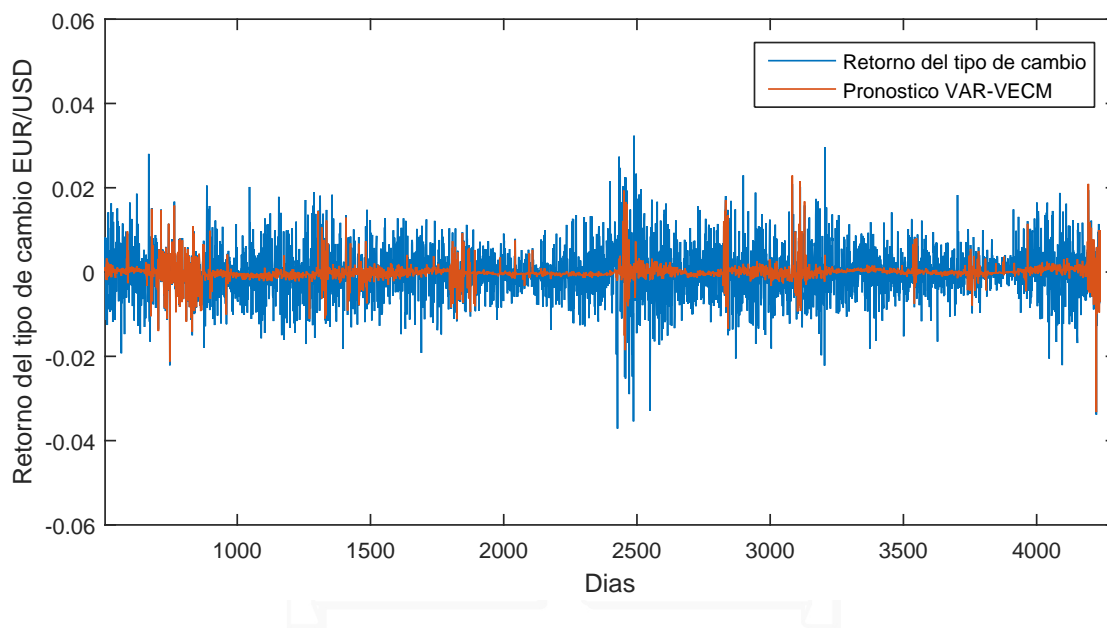


Figura 4.5: Comparación entre la serie del retorno de tipo de cambio EUR/USD y su pronóstico generado por el modelo VAR-VECM en la ventana de 252 días. Fuente: Elaboración propia.

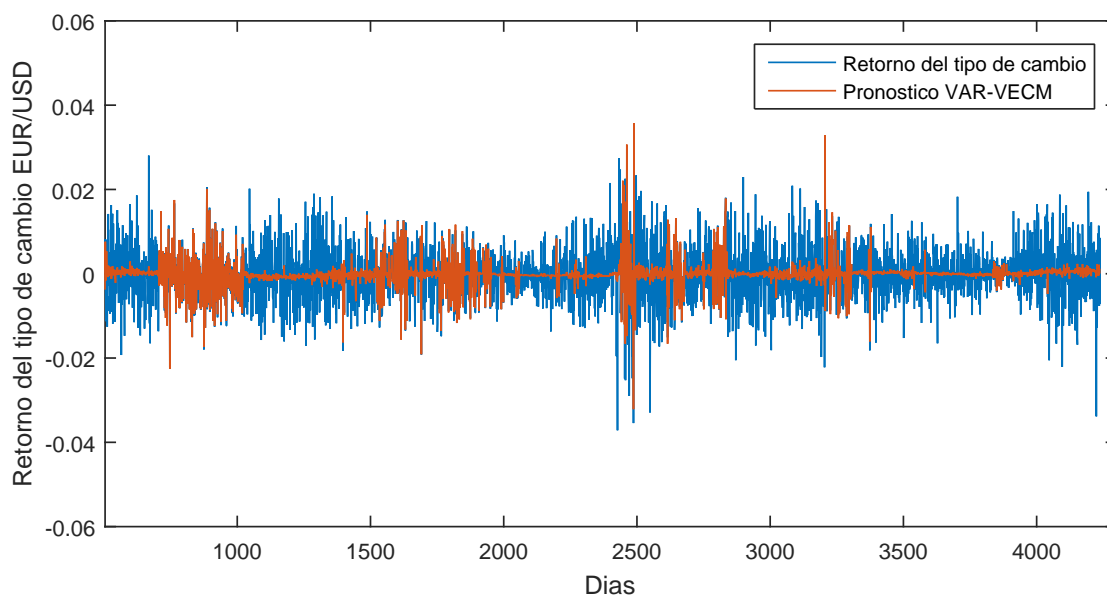


Figura 4.6: Comparación entre la serie del retorno de tipo de cambio EUR/USD y su pronóstico generado por el modelo VAR-VECM en la ventana de 378 días. Fuente: Elaboración propia.

Paso 6: Al aplicar la función de pérdida MSE, para encontrar el modelo que presenta el menor error generalizado, se obtienen los resultados expuestos en la tabla 4.3. De esta tabla se desprende que la ventana corrediza que ofrece el menor error

corresponde a la de 126 días con un MSE de $3,39 \times 10^{-5}$, seguida de la de 84 días con un MSE de $3,47 \times 10^{-5}$ y, contrario a lo que se observa gráficamente, a través de la estadística se muestra que la peor opción corresponde a la ventana corrediza de 42 días con un MSE de $4,37 \times 10^{-5}$ bastante superior al resto de los modelos trabajados. Así, el representante de los modelos econométricos que será usado como modelo de referencia más adelante, es el modelo VAR-VECM con ventana corrediza de 126 días, cuyos pronósticos se ven en la figura 4.4.

MSE del modelo VAR-VECM en cada tamaño de ventana

Tamaño de Ventana	378	252	126	84	63	42
MSE	3,71E-05	3,65E-05	3,39E-05	3,47E-05	3,55E-05	4,37E-05

TABLA 4.3: MSE del modelo VAR-VECM en cada tamaño de ventana (días). Fuente: Elaboración propia.

Paso 7: Los residuales del pronóstico del retorno de tipo de cambio EUR/USD se extraen del modelo VAR-VECM con ventana corrediza de 126 días y alimentarán la ANN. Estos residuales se pueden ver en la figura 4.7.

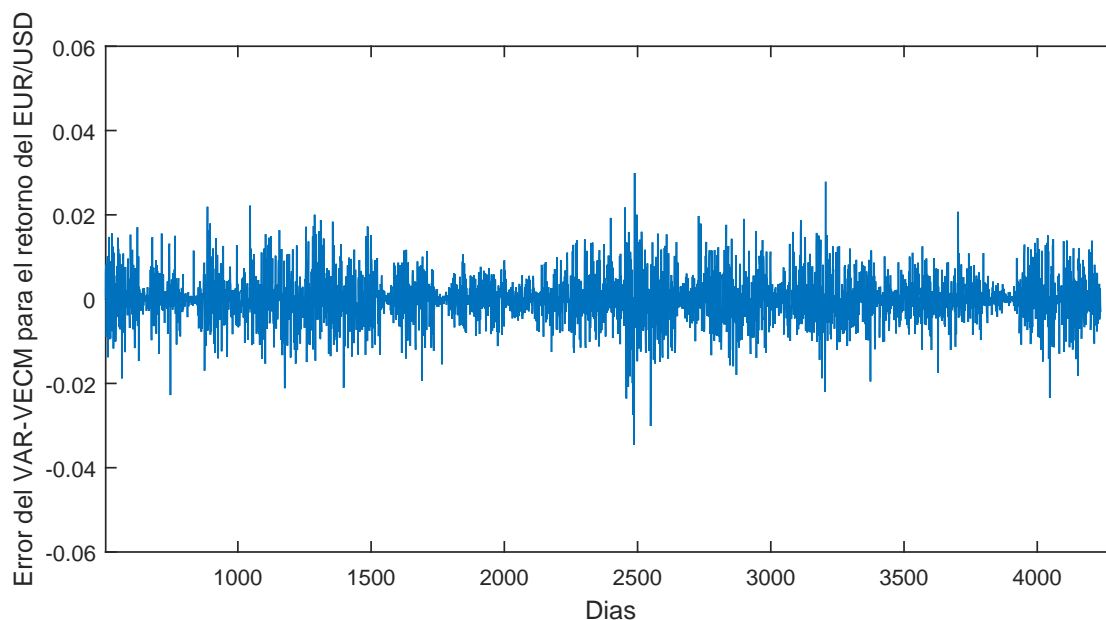


Figura 4.7: Residuales del pronóstico del modelo VAR-VECM con ventana corrediza de 126 días para el retorno del EUR/USD. Fuente: Elaboración propia.

Paso 8 y 9: Los pronósticos que entrega la Red Neuronal para cada ventana corrediza y para cada sensibilización de capas ocultas y neuronas por capa se muestran de color naranja en las figuras del anexo [A](#), [A.1](#), [A.2](#), [A.3](#), [A.4](#), [A.5](#), [A.6](#), [A.7](#), [A.8](#), [A.9](#), [A.10](#), [A.11](#), [A.12](#), [A.13](#), [A.14](#), [A.15](#), [A.16](#), [A.17](#), [A.18](#), [A.19](#), [A.20](#), [A.21](#), [A.22](#), [A.23](#) y [A.24](#). De estas figuras se puede notar, como un análisis gráfico preliminar, que los pronósticos que genera la inteligencia artificial se ajustan de mucho mejor forma a los cambios y patrones en los datos reales que los pronósticos del modelo econométrico, logrando identificar los momentos donde el retorno del tipo de cambio presenta una tendencia más y menos volátil. Además, para todos los tamaños de ventana corrediza, a medida que se aumenta el número de capas ocultas los pronósticos se vuelven más planos, es decir, se concentran dentro de cotas mucho más pequeñas. Por otro lado, también se puede ver de manera gráfica que al aumentar el número de neuronas en cada capa oculta los pronósticos tienden a extender levemente las cotas que restringe el aumento del número de capas ocultas. Entonces, al menos de manera gráfica y a modo de ejemplo, la figura [A.18d](#) que representa los pronósticos para la ventana de 252 días con la configuración de 2 capas ocultas y 20 neuronas parece mostrar peores pronósticos que la figura [A.17d](#) que responde a 1 capa oculta con 20 neuronas, pero mejores que la figura [A.18a](#) para la configuración de 2 capas ocultas y 5 neuronas.

Paso 10: Con los pronósticos para cada ventana corrediza, cada cantidad de capa oculta y cada cantidad de neuronas por capa, se aplica la métrica de error MSE en todos los casos para encontrar la combinación que entrega los mejores pronósticos en términos estadísticos. Los resultados de aplicar MSE para comparar los errores generalizados de cada modelo se muestran en las tablas [4.4](#), [4.5](#), [4.6](#), [4.7](#). Se debe notar que en primera instancia se consideró hasta 20 neuronas por capa oculta, pero los mejores resultados se obtienen justamente en este límite; por consiguiente, el análisis debe extenderse a 30 neuronas por capa oculta, para verificar si se alcanza el mínimo MSE. Esta extensión solo se realiza en los tamaños de ventana que entregan los mejores pronósticos, que corresponden a

378 días (en el caso de 2 capas ocultas) y 252 días (para todos los casos). Esto se debe al hecho de que, si el resto de los tamaños de ventana no son capaces de superar los pronósticos con pocas neuronas, siguiendo la tendencia, tampoco lo harán con más neuronas por capa. Esta extensión en la cantidad de neuronas no muestra una mejora, es decir, los mejores resultados se encontraban desde un comienzo en las 20 neuronas. Además, se debe notar que el MSE del modelo econométrico usado como comparación en cada caso es el MSE de la ventana de tamaño de 126 días extraído de la tabla 4.3. Las tablas 4.4, 4.5, 4.6, 4.7 muestran que para todas las capas ocultas, en general, el mejor pronóstico de ANN está dado por el tamaño de ventana de 252 días, mientras que los peores pronósticos se encuentran para la ventana de 42 días. Más aún, los mejores pronósticos se concentran en el caso de 2 capas ocultas para la ventana de 252 días (ver tabla 4.5), donde se encuentra la mayor cantidad de variaciones negativas del MSE, e incluso se encuentra la máxima variación negativa para el caso de 20 neuronas en estas 2 capas ocultas, con una mejora sobre el MSE del pronóstico econométrico de un 26,883 %, seguido por el modelo de 2 capas ocultas con 20 neuronas en la ventana de 378 días con una mejora de 25,086 %. Por otro lado, los peores pronósticos se concentran en el caso de 4 capas ocultas (ver tabla 4.7), donde no existen variaciones negativas del MSE, es decir, el error general del modelo VEC-ANN es mayor que el del VAR-VECM, lo que indicaría que el modelo híbrido no mejora la precisión del pronóstico.

Entonces, se puede desprender que un mayor número de capas y neuronas no necesariamente entrega mejores resultados predictivos. Esto dependerá de los datos con los que se trabaje. Para este caso en particular, un número alto de neuronas trabaja de mejor manera, pero para un bajo número de capas ocultas. Específicamente, el modelo que entrega los mejores resultados se da con una configuración de 2 capas ocultas y 20 neuronas por capa oculta cuando se trabaja con una ventana corregida de tamaño 252 días. Este modelo se representa en la figura 4.8 y es el candidato de los modelos híbridos que será comparado y

		MSE Ventana (días)							
Capas ocultas	Neuronas	Modelo	378	252	126	84	63	42	
1	5	VEC-ANN	5,07E-05	6,65E-05	3,30E-05	3,18E-05	3,00E-05	3,25E-05	
		VAR-VECM	3,39E-05	3,39E-05	3,39E-05	3,39E-05	3,39E-05	3,39E-05	
		Variación	49,346 %	95,946 %	-2,684 %	-6,118 %	-11,556 %	-4,325 %	
	10	VEC-ANN	3,43E-05	3,09E-05	3,26E-05	3,40E-05	3,88E-05	3,79E-05	
		VAR-VECM	3,39E-05	3,39E-05	3,39E-05	3,39E-05	3,39E-05	3,39E-05	
		Variación	1,081 %	-9,003 %	-3,973 %	0,256 %	14,398 %	11,624 %	
	15	VEC-ANN	3,24E-05	3,14E-05	4,10E-05	4,72E-05	4,71E-05	5,79E-05	
		VAR-VECM	3,39E-05	3,39E-05	3,39E-05	3,39E-05	3,39E-05	3,39E-05	
		Variación	-4,589 %	-7,533 %	20,988 %	39,172 %	38,742 %	70,577 %	
	20	VEC-ANN	3,82E-05	3,63E-05	5,25E-05	6,02E-05	5,98E-05	6,35E-05	
VAR-VECM		3,39E-05	3,39E-05	3,39E-05	3,39E-05	3,39E-05	3,39E-05		
	Variación	12,669 %	6,982 %	54,674 %	77,341 %	76,417 %	87,189 %		
25	VEC-ANN		1,17E-04						
	VAR-VECM		3,39E-05						
	Variación		245,511 %						
30	VEC-ANN		1,33E-04						
	VAR-VECM		3,39E-05						
	Variación		290,77 %						

TABLA 4.4: Variación del MSE entregado por el modelo VEC-ANN de 1 capa oculta respecto del modelo VAR-VECM. Fuente: Elaboración propia.

		MSE									
		Ventana (días)									
Capas ocultas	Neuronas	Modelo	378	252	126	84	63	42			
2	5	VEC-ANN	3,03E-05	2,96E-05	3,06E-05	3,32E-05	3,44E-05	3,72E-05			
		VAR-VECM	3,39E-05	3,39E-05	3,39E-05	3,39E-05	3,39E-05	3,39E-05			
		Variación	-10,797 %	-12,699 %	-9,807 %	-2,207 %	1,495 %	9,636 %			
10	5	VEC-ANN	2,74E-05	2,77E-05	2,94E-05	3,17E-05	3,46E-05	3,80E-05			
		VAR-VECM	3,39E-05	3,39E-05	3,39E-05	3,39E-05	3,39E-05	3,39E-05			
		Variación	-19,143 %	-18,462 %	-13,229 %	-6,553 %	1,958 %	12,154 %			
15	5	VEC-ANN	2,66E-05	2,65E-05	2,90E-05	3,10E-05	3,51E-05	3,82E-05			
		VAR-VECM	3,39E-05	3,39E-05	3,39E-05	3,39E-05	3,39E-05	3,39E-05			
		Variación	-21,513 %	-21,977 %	-14,649 %	-8,725 %	3,512 %	12,711 %			
20	5	VEC-ANN	2,54E-05	2,48E-05	2,93E-05	2,97E-05	3,33E-05	3,68E-05			
		VAR-VECM	3,39E-05	3,39E-05	3,39E-05	3,39E-05	3,39E-05	3,39E-05			
		Variación	-25,086 %	-26,883 %	-13,488 %	-12,529 %	-1,823 %	8,466 %			
25	5	VEC-ANN	6,17E-05	6,52E-05							
		VAR-VECM	3,39E-05	3,39E-05							
		Variación	81,851 %	92,076 %							
30	5	VEC-ANN	6,99E-05	6,56E-05							
		VAR-VECM	3,39E-05	3,39E-05							
		Variación	105,94 %	93,41 %							

TABLA 4.5: Variación del MSE entregado por el modelo VEC-ANN de 2 capas ocultas respecto del modelo VAR-VECM. Fuente: Elaboración propia.

		MSE Ventana (días)						
Capas ocultas	Neuronas	Modelo	378	252	126	84	63	42
3	5	VEC-ANN	3,39E-05	3,30E-05	3,38E-05	3,57E-05	3,77E-05	3,97E-05
		VAR-VECM	3,39E-05	3,39E-05	3,39E-05	3,39E-05	3,39E-05	3,39E-05
	Variación		-0,124 %	-2,651 %	-0,495 %	5,234 %	11,110 %	17,079 %
	10	VEC-ANN	3,25E-05	3,23E-05	3,32E-05	3,48E-05	3,76E-05	4,13E-05
		VAR-VECM	3,39E-05	3,39E-05	3,39E-05	3,39E-05	3,39E-05	3,39E-05
	Variación		-4,324 %	-4,665 %	-2,104 %	2,559 %	10,825 %	21,614 %
	15	VEC-ANN	3,20E-05	3,23E-05	3,34E-05	3,34E-05	3,73E-05	4,13E-05
		VAR-VECM	3,39E-05	3,39E-05	3,39E-05	3,39E-05	3,39E-05	3,39E-05
	Variación		-5,754 %	-4,832 %	-1,454 %	-1,592 %	10,055 %	21,690 %
	20	VEC-ANN	3,20E-05	3,15E-05	3,43E-05	3,28E-05	3,73E-05	4,22E-05
		VAR-VECM	3,39E-05	3,39E-05	3,39E-05	3,39E-05	3,39E-05	3,39E-05
	Variación		-5,752 %	-7,050 %	1,102 %	-3,426 %	9,925 %	24,513 %
	25	VEC-ANN		3,20E-05				
		VAR-VECM		3,39E-05				
Variación			-5,651 %					
30	VEC-ANN		3,19E-05					
	VAR-VECM		3,39E-05					
Variación			-5,98 %					

TABLA 4.6: Variación del MSE entregado por el modelo VEC-ANN de 3 capas ocultas respecto del modelo VAR-VECM. Fuente: Elaboración propia.

		MSE							
		Ventana (días)							
Capas ocultas	Neuronas	Modelo	378	252	126	84	63	42	
4	5	VEC-ANN	3,54E-05	3,56E-05	3,68E-05	3,78E-05	3,95E-05	4,19E-05	
		VAR-VECM	3,39E-05	3,39E-05	3,39E-05	3,39E-05	3,39E-05	3,39E-05	
		Variación	4,374 %	4,864 %	8,429 %	11,387 %	16,347 %	23,406 %	
10	10	VEC-ANN	3,55E-05	3,54E-05	3,70E-05	3,68E-05	3,99E-05	4,21E-05	
		VAR-VECM	3,39E-05	3,39E-05	3,39E-05	3,39E-05	3,39E-05	3,39E-05	
		Variación	4,660 %	4,368 %	9,155 %	8,375 %	17,583 %	24,011 %	
15	15	VEC-ANN	3,53E-05	3,56E-05	3,81E-05	3,59E-05	4,10E-05	4,26E-05	
		VAR-VECM	3,39E-05	3,39E-05	3,39E-05	3,39E-05	3,39E-05	3,39E-05	
		Variación	4,122 %	5,080 %	12,242 %	5,967 %	20,844 %	25,623 %	
20	20	VEC-ANN	3,61E-05	3,56E-05	3,73E-05	3,54E-05	4,12E-05	4,39E-05	
		VAR-VECM	3,39E-05	3,39E-05	3,39E-05	3,39E-05	3,39E-05	3,39E-05	
		Variación	6,331 %	5,039 %	10,003 %	4,260 %	21,594 %	29,346 %	
25	25	VEC-ANN		4,97E-05					
		VAR-VECM		3,39E-05					
		Variación		46,388 %					
30	30	VEC-ANN		3,63E-05					
		VAR-VECM		3,39E-05					
		Variación		7,10 %					

TABLA 4.7: Variación del MSE entregado por el modelo VEC-ANN de 4 capas ocultas respecto del modelo VAR-VECM. Fuente: Elaboración propia.

probado estadísticamente contra el modelo de referencia. De la figura 4.8 se puede notar que los pronósticos de este modelo se ajustan mucho mejor a los datos reales que los pronósticos del modelo econométrico (ver figura 4.4). Sigue patrones, tendencias y cambios que sufren los datos de una mejor manera debido a que recoge más información a partir de la componente no lineal con la cual se alimenta la ANN y la ventana corrediza rescata el dinamismo de la serie. El modelo híbrido es capaz de captar momentos de alta y baja volatilidad a lo largo del tiempo, a excepción de los eventos cercanos al día 2500 (Crisis Subprime) y los cercanos al día 4000, debido a factores ajenos a los datos considerados en este análisis, y que la ANN no es capaz de reconocer. La figura 4.9 corresponde a un acercamiento de la figura 4.8, para 100 días entre los datos 3100 y 3200, donde se puede apreciar claramente lo bien que se ajusta el modelo VEC-ANN para pronosticar en términos generales, no solo en cuanto a la cercanía con el valor real del retorno de tipo de cambio, sino que también reconociendo los cambios de dirección que este experimenta en el tiempo, con muy pocas excepciones como los días 3172 y 3173 donde falló en dirección.

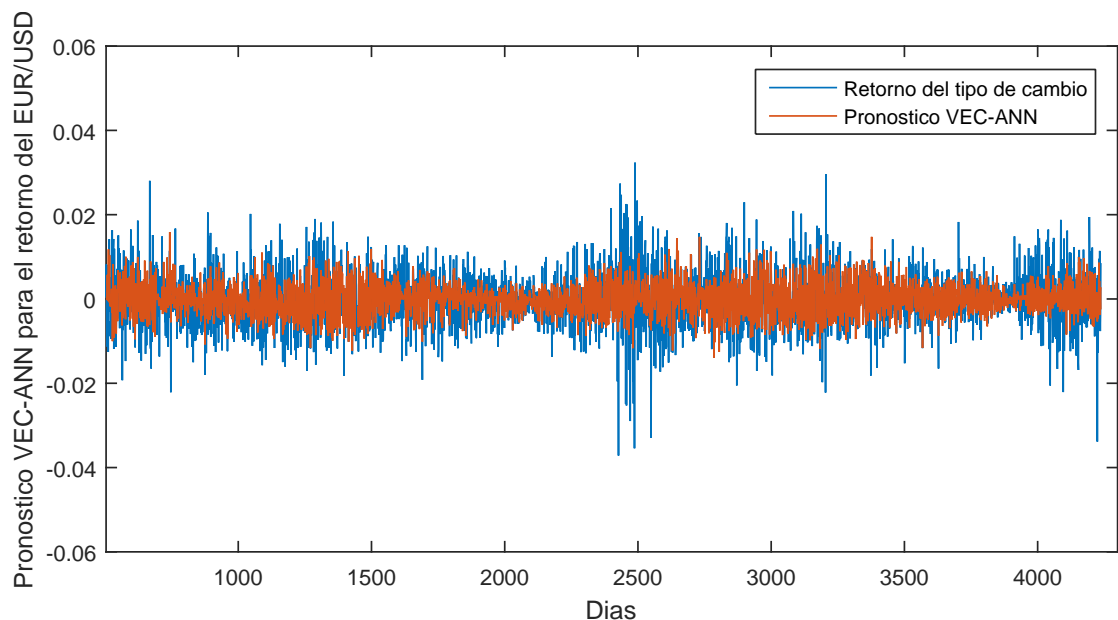


Figura 4.8: Pronóstico para la ventana de 252 días con 2 capas ocultas y 20 neuronas. Fuente: Elaboración propia.

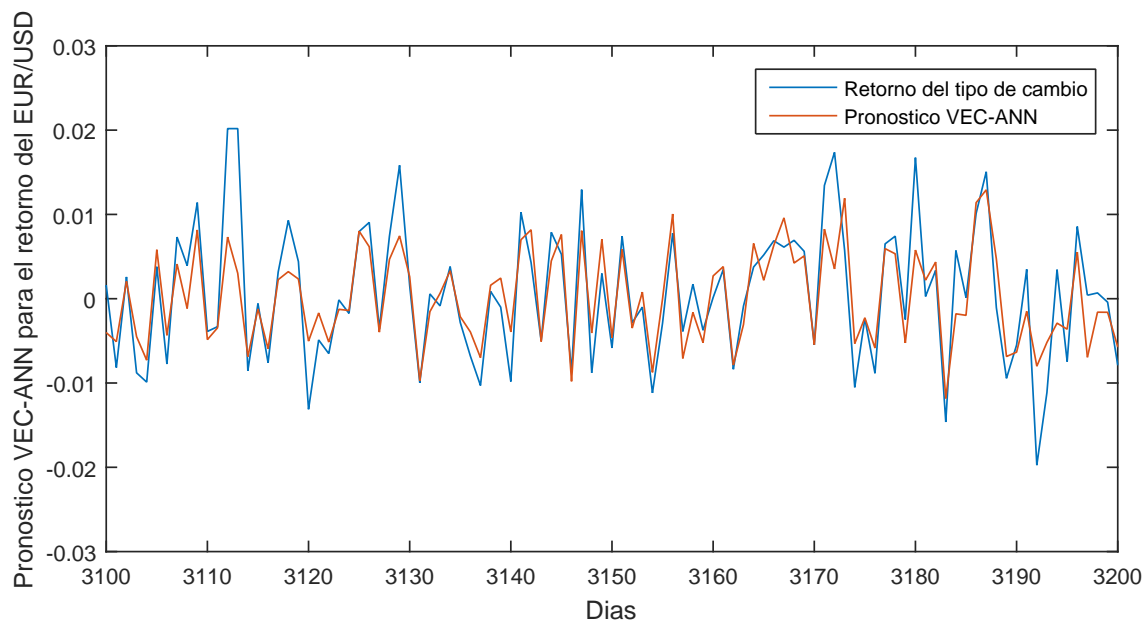


Figura 4.9: Zoom entre los días 3100 y 3200 del pronóstico para la ventana de 252 días con 2 capas ocultas y 20 neuronas. Fuente: Elaboración propia.

Las pruebas estadísticas explicadas en el punto 2.7.3 son necesarias debido a que, a pesar de que se tienen las comparaciones del MSE, existen muchos modelos con indicadores parecidos y no es posible asegurar el mejor de ellos solo analizando el error generalizado. Los resultados de la prueba DM aplicada sobre el modelo econométrico con ventana corrediza de 126 días junto al modelo híbrido de 2 capas, 20 neuronas y ventana corrediza de 252 días indican que, con un 95 % de confianza, la hipótesis nula de igualdad en la precisión de pronósticos para los modelos VAR-VECM y VEC-ANN se rechaza al entregar un valor del estadístico de $-6,2607$. Este valor, al ser comparado con un valor crítico de $\pm 1,96$ para una tabla Normal, estadísticamente confirma el rechazo de la hipótesis nula y, al ser un valor negativo, indica que el modelo econométrico tiene un mayor valor del cuadrado del error que el modelo VEC-ANN; por consiguiente, este último presenta una mayor precisión al pronosticar. Adicionalmente a la prueba DM, se aplica la prueba MCS a una matriz donde cada columna corresponde a los residuales de cada modelo analizado en el estudio, es decir, todos los modelos econométricos para cada tamaño de ventana y todos los modelos híbridos para cada tamaño de ventana, cada cantidad de capas y cada cantidad de neuronas por

capa, y luego se evalúan las funciones de pérdida de cada uno. Los resultados de esta prueba son consistentes con el MSE obtenido en las tablas 4.4, 4.5, 4.6 y 4.7, mostrando que los mejores pronósticos se obtienen para el caso del modelo híbrido de 2 capas ocultas. De hecho, los únicos modelos que considera esta prueba como "mejores", porque no existen otros modelos significativamente inferiores, son para esta cantidad de capas ocultas. Para esta prueba, un modelo que entrega un valor-p mayor a 0,95, para un nivel de significancia del 5 %, rechaza la hipótesis que existe otro modelo en el conjunto que lo iguala o supera en términos predictivos. Entonces, el modelo que entrega un valor-p igual a 1 significa que es el último modelo del conjunto y, por lo tanto, es tan bueno como sí mismo. Así, el modelo para la ventana corrediza de 252 días, con 2 capas ocultas y 20 neuronas, que previamente entregó el menor MSE con una variación de -26,883 % respecto del econométrico, a través de esta prueba de robustez también entrega el mayor valor-p como se puede notar en las tablas 4.8, 4.9, 4.10 y 4.11, demostrando que este modelo pertenece al conjunto de los mejores modelos con un valor-p de 1. Los modelos de 15 capas para las ventanas de 378 días y 252 días también pertenecen al conjunto de mejores modelos pero al usar un nivel de significancia del 5 %, mientras que el modelo de 20 neuronas para la ventana de 378 días pertenecería al conjunto de mejores modelos si el nivel de significancia a usar fuese 1 %, como se puede concluir de la tabla 4.9.

Neuronas	Valor-p de la prueba MCS					
	Tamaño de Ventana (días)					
	378	252	126	84	63	42
5	0	0	0	0	0	0
10	0	0	0	0	0	0
15	0	0	0	0	0	0
20	0	0	0	0	0	0

TABLA 4.8: Valor-p para cada modelo de 1 capa oculta en la prueba MCS.

Valor-p de la prueba MCS						
Tamaño de Ventana (días)						
Neuronas	378	252	126	84	63	42
5	0	0,058	0	0	0	0
10	0,797	0	0	0	0	0
15	0,987*	0,987*	0	0	0	0
20	0,992**	1***	0	0	0	0

TABLA 4.9: Valor-p para cada modelo de 2 capas ocultas en la prueba MCS. * indica que este modelo pertenece al conjunto de mejores modelos con un nivel de significancia del 5 %. ** indica que este modelo pertenece al conjunto de mejores modelos con un nivel de significancia del 1 %. *** indica que este modelo pertenece al conjunto de mejores modelos con un nivel de significancia del 0 %.

Valor-p de la prueba MCS						
Tamaño de Ventana (días)						
Neuronas	378	252	126	84	63	42
5	0	0	0	0	0	0
10	0	0	0	0	0	0
15	0	0	0	0	0	0
20	0	0	0	0	0	0

TABLA 4.10: Valor-p para cada modelo de 3 capas ocultas en la prueba MCS.

Valor-p de la prueba MCS						
Tamaño de Ventana (días)						
Neuronas	378	252	126	84	63	42
5	0	0	0	0	0	0
10	0	0	0	0	0	0
15	0	0	0	0	0	0
20	0	0	0	0	0	0

TABLA 4.11: Valor-p para cada modelo de 4 capas ocultas en la prueba MCS.

CAPÍTULO 5

CONCLUSIONES

Los mercados financieros modernos presentan un alto grado de competencia entre sus participantes; por lo tanto, tener la mayor cantidad de información posible es fundamental. En este escenario, los pronósticos del tipo de cambio juegan un papel especial en la anticipación de los movimientos del mercado. Este estudio aplica un modelo híbrido usando ventana corrediza y análisis de residuales para pronosticar los retornos del tipo de cambio.

La motivación para usar modelos híbridos recae en el hecho de que, teórica y empíricamente, la integración de diversos modelos mejora la habilidad para pronosticar con respecto a cada modelo individual. Entonces, el modelo propuesto que considera una ANN alimentada, entre otros factores, por los residuales provenientes de un modelo econométrico VAR-VECM debería entregar mejores resultados que los modelos clásicos al obtener mayor información de la componente no lineal de corto y largo plazo que ellos no pueden procesar. Se utiliza un análisis de ventana corrediza para obtener un modelo dinámico que se ajuste mejor a cambios en los datos con respecto a un modelo que solo considera un espacio temporal como marco general.

Los resultados muestran que el modelo propuesto entrega mejores pronósticos que el econométrico, especialmente usando una ventana corrediza de 252 días con una configuración de 2 capas ocultas y 20 neuronas por capa. Estos resultados son evaluados en términos de la métrica de error MSE y se prueban estadísticamente a través de las pruebas

Diebold-Mariano (DM) y Model Confidence Set (MCS) ya que un simple análisis gráfico preliminar no es suficiente. El MSE muestra que el modelo propuesto mejora los pronósticos al reducir el error generalizado en un 26,883 % respecto al modelo clásico. La prueba DM muestra que el modelo híbrido (VEC-ANN de 2 capas con 20 neuronas y ventana corrediza de 252 días) sobrepasa estadísticamente al modelo de referencia (VAR-VECM con ventana corrediza de 126 días) en términos de la precisión en los pronósticos, al rechazar la hipótesis nula de que ellos tienen igual capacidad predictiva con un estadístico de $-6,2607$ bastante alejado del $\pm 1,96$ que especifica el valor crítico en una tabla Normal. Finalmente, la prueba MCS muestra que el modelo híbrido elegido (VEC-ANN de 2 capas con 20 neuronas y ventana corrediza de 252 días) supera a todo el resto de los modelos en estudio para un nivel de significancia del 0 %, sin una referencia específica, concluyendo que no existe mejor modelo entre ellos para aumentar la precisión del retorno de tipo de cambio EUR/USD. Todas estas pruebas llevan a verificar que la metodología y el modelo propuesto se condicen con lo establecido en la literatura sobre la mejora que existe en la utilización de modelos híbridos como, por ejemplo, los casos de Bildirici et al. [27], Fatima & Husain [28] y Khashei & Bijari [30], que fueron trabajos en los que se basó el presente estudio.

A través de estos resultados y sus conclusiones es posible determinar el cumplimiento del objetivo principal de este estudio, que corresponde a proponer una metodología para mejorar la precisión en los pronósticos del retorno de tipo de cambio mediante el uso de un modelo híbrido en comparación al uso de un modelo econométrico de referencia. Esto se logró al obtener el modelo híbrido VEC-ANN que es capaz de capturar las relaciones no lineales de corto y largo plazo en las series para mejorar la precisión en los pronósticos que entrega el modelo VAR-VECM. De igual modo, se verifica que los objetivos secundarios también se cumplen: aportando una nueva metodología a la literatura existente en relación a modelos enfocados en pronósticos; sensibilizando el análisis para diferentes tamaños de ventana corrediza (42, 63, 84, 126, 252 y 378 días), para mostrar que de esta manera es posible capturar de mejor forma el dinamismo en los datos; sensibilizando el análisis para diferentes configuraciones de la Red Neuronal Artificial, variando la cantidad de capas ocultas y cantidad de neuronas en cada una de ellas, mostrando que un mayor número de

capas ocultas tiende a acotar y aplanar los pronósticos, mientras que un mayor número de neuronas en cada capa oculta tiende a expandir los pronósticos levemente, acercándolos más a los valores reales por la capacidad de reconocer elementos más complejos con un mayor número de neuronas; validando los resultados obtenidos mediante pruebas estadísticas que verifican que el modelo propuesto es superior al querer aumentar la precisión de los pronósticos.

Futuras investigaciones pueden basarse en estos resultados y enfocarse en agregar más modelos a la cadena híbrida, con diferentes características, para rescatar más información a partir de cada uno y así obtener pronósticos más precisos. Además, sería interesante comprobar la aplicación de este modelo en otros indicadores económicos, como pueden ser: inflación, precio de commodities, etc. para verificar cómo se comportan las distintas configuraciones necesarias de la red y los tamaños de ventana corrediza en esos análisis.



Bibliografía

- [1] Lutz Kilian and Mark P Taylor. Why is it so difficult to beat the random walk forecast of exchange rates? *Journal of International Economics*, 60(1):85–107, 2003.
- [2] Yudong Zhang and Lenan Wu. Stock market prediction of s&p 500 via combination of improved bco approach and bp neural network. *Expert systems with applications*, 36(5):8849–8854, 2009.
- [3] An-Sing Chen and Mark T Leung. Regression neural network for error correction in foreign exchange forecasting and trading. *Computers & Operations Research*, 31(7):1049–1068, 2004.
- [4] Jefferson T Davis, Athanasios Episcopos, and Sannaka Wettimuny. Predicting direction shifts on canadian–us exchange rates with artificial neural networks. *Intelligent Systems in Accounting, Finance and Management*, 10(2):83–96, 2001.
- [5] Vassilis Kodogiannis and A Lolis. Forecasting financial time series using neural network and fuzzy system-based techniques. *Neural computing & applications*, 11(2):90–102, 2002.
- [6] Jingtao Yao and Chew Lim Tan. A case study on using neural networks to perform technical forecasting of forex. *Neurocomputing*, 34(1):79–98, 2000.
- [7] Kenneth W Clements and Yihui Lan. A new approach to forecasting exchange rates. *Journal of International Money and Finance*, 29(7):1424–1437, 2010.
- [8] Lean Yu, Shouyang Wang, and Kin Keung Lai. A novel nonlinear ensemble forecasting model incorporating glar and ann for foreign exchange rates. *Computers & Operations Research*, 32(10):2523–2541, 2005.
- [9] Mark T Leung, An-Sing Chen, and Hazem Daouk. Forecasting exchange rates using general regression neural networks. *Computers & Operations Research*, 27(11):1093–1110, 2000.
- [10] Richard Meese and Kenneth Rogoff. Empirical exchange rate models of the seventies: Do they fit out of sample? *Journal of International Economics*, 14:3–24, 1983.
- [11] M. Alvarez-Diaz, R. Bissoondeal, J. Binner, M. Bhuruth, A. Gazel, and V. Mootanah. Forecasting exchange rates with linear and nonlinear models. *Applied Economics*, 40:1969–1984, 2008.

- [12] Svitlana Galeshchuk. Neural networks performance in exchange rate prediction. *Neurocomputing*, 172:446–452, 2016.
- [13] Monica Lam. Neural network techniques for financial performance prediction: integrating fundamental and technical analysis. *Decision support systems*, 37(4):567–581, 2004.
- [14] Chung-Ming Kuan and Halbert White. Artificial neural networks: an econometric perspective*. *Econometric reviews*, 13(1):1–91, 1994.
- [15] Norman R Swanson and Halbert White. Forecasting economic time series using flexible versus fixed specification and linear versus nonlinear econometric models. *International journal of Forecasting*, 13(4):439–461, 1997.
- [16] William Leigh, Russell Purvis, and James M Ragusa. Forecasting the nyse composite index with technical analysis, pattern recognizer, neural network, and genetic algorithm: a case study in romantic decision support. *Decision support systems*, 32(4):361–377, 2002.
- [17] AK Dhamija and VK Bhalla. Exchange rate forecasting: comparison of various architectures of neural networks. *Neural Computing and Applications*, 20(3):355–363, 2011.
- [18] Mehdi Khashei, Mehdi Bijari, and Gholam Ali Raissi Ardali. Improvement of autoregressive integrated moving average models using fuzzy logic and artificial neural networks (anns). *Neurocomputing*, 72(4):956–967, 2009.
- [19] Christian L Dunis and Xuehuan Huang. Forecasting and trading currency volatility: An application of recurrent neural regression and model combination. *Journal of Forecasting*, 21(5):317–354, 2002.
- [20] John M Bates and Clive WJ Granger. The combination of forecasts. *Or*, pages 451–468, 1969.
- [21] Robert T Clemen. Combining forecasts: A review and annotated bibliography. *International journal of forecasting*, 5(4):559–583, 1989.
- [22] William G Baxt. Improving the accuracy of an artificial neural network using multiple differently trained networks. *Neural computation*, 4(5):772–780, 1992.
- [23] G Peter Zhang. A neural network ensemble method with jittered training data for time series forecasting. *Information Sciences*, 177(23):5329–5346, 2007.
- [24] Tugba Taskaya-Temizel and Matthew C Casey. A comparative study of autoregressive neural network hybrids. *Neural Networks*, 18(5):781–789, 2005.
- [25] Donald K Wedding and Krzysztof J Cios. Time series forecasting by combining rbf networks, certainty factors, and the box-jenkins model. *Neurocomputing*, 10(2):149–168, 1996.

- [26] Ray Tsaih, Yenshan Hsu, and Charles C Lai. Forecasting s&p 500 stock index futures with a hybrid ai system. *Decision Support Systems*, 23(2):161–174, 1998.
- [27] Melike Bildirici, Elçin A Alp, and Özgür Ö Ersin. Tar-cointegration neural network model: An empirical analysis of exchange rates and stock returns. *Expert Systems with Applications*, 37(1):2–11, 2010.
- [28] Samreen Fatima and Ghulam Hussain. Statistical models of kse100 index using hybrid financial systems. *Neurocomputing*, 71(13):2742–2746, 2008.
- [29] Weimin Li, Yishu Luo, Qin Zhu, Jianwei Liu, and Jiajin Le. Applications of ar*-grnn model for financial time series forecasting. *Neural Computing and Applications*, 17(5-6):441–448, 2008.
- [30] Mehdi Khashei and Mehdi Bijari. An artificial neural network (p, d, q) model for timeseries forecasting. *Expert Systems with applications*, 37(1):479–489, 2010.
- [31] Ehsan Hajizadeh, Masoud Mahootchi, Akbar Esfahanipour, and Mahdi Massahi Kh. A new nn-pso hybrid model for forecasting euro/dollar exchange rate volatility. *Neural Computing and Applications*, pages 1–9, 2015.
- [32] Busenur Sarica, Erol Eğrioğlu, and Barış Aşıkçil. A new hybrid method for time series forecasting: Ar–anfis. *Neural Computing and Applications*, pages 1–12, 2016.
- [33] James Douglas Hamilton. *Time series analysis*, volume 2. Princeton university press Princeton, 1994.
- [34] David A Dickey and Wayne A Fuller. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association*, 74(366a):427–431, 1979.
- [35] Christopher A Sims. Macroeconomics and reality. *Econometrica: Journal of the Econometric Society*, pages 1–48, 1980.
- [36] Clive WJ Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: Journal of the Econometric Society*, pages 424–438, 1969.
- [37] James EH Davidson, David F Hendry, Frank Srba, and Stephen Yeo. Econometric modelling of the aggregate time-series relationship between consumers' expenditure and income in the united kingdom. *The Economic Journal*, pages 661–692, 1978.
- [38] Robert F Engle and Clive WJ Granger. Co-integration and error correction: representation, estimation, and testing. *Econometrica: journal of the Econometric Society*, pages 251–276, 1987.
- [39] Søren Johansen. Statistical analysis of cointegration vectors. *Journal of economic dynamics and control*, 12(2-3):231–254, 1988.

- [40] Søren Johansen and Katarina Juselius. Maximum likelihood estimation and inference on cointegration—with applications to the demand for money. *Oxford Bulletin of Economics and statistics*, 52(2):169–210, 1990.
- [41] Warren S McCulloch and Walter Pitts. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4):115–133, 1943.
- [42] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Multilayer feedforward networks are universal approximators. *Neural networks*, 2(5):359–366, 1989.
- [43] H Gholam Hosseini, Dehan Luo, and Karen Jane Reynolds. The comparison of different feed forward neural network architectures for ecg signal diagnosis. *Medical engineering & physics*, 28(4):372–378, 2006.
- [44] Guoqiang Zhang, B Eddy Patuwo, and Michael Y Hu. Forecasting with artificial neural networks:: The state of the art. *International journal of forecasting*, 14(1):35–62, 1998.
- [45] David E Rumelhart, Geoffrey E Hinton, Ronald J Williams, et al. Learning representations by back-propagating errors. *Cognitive modeling*, 5(3):1, 1988.
- [46] Hirotugu Akaike. A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6):716–723, 1974.
- [47] Francis X Diebold and Robert S Mariano. Comparing predictive accuracy. *Journal of Business & economic statistics*, 20(1):134–144, 2002.
- [48] Peter R Hansen, Asger Lunde, and James M Nason. The model confidence set. *Econometrica*, 79(2):453–497, 2011.
- [49] Peter Reinhard Hansen. A test for superior predictive ability. *Journal of Business & Economic Statistics*, 23(4):365–380, 2005.

ANEXO A

ANEXOS



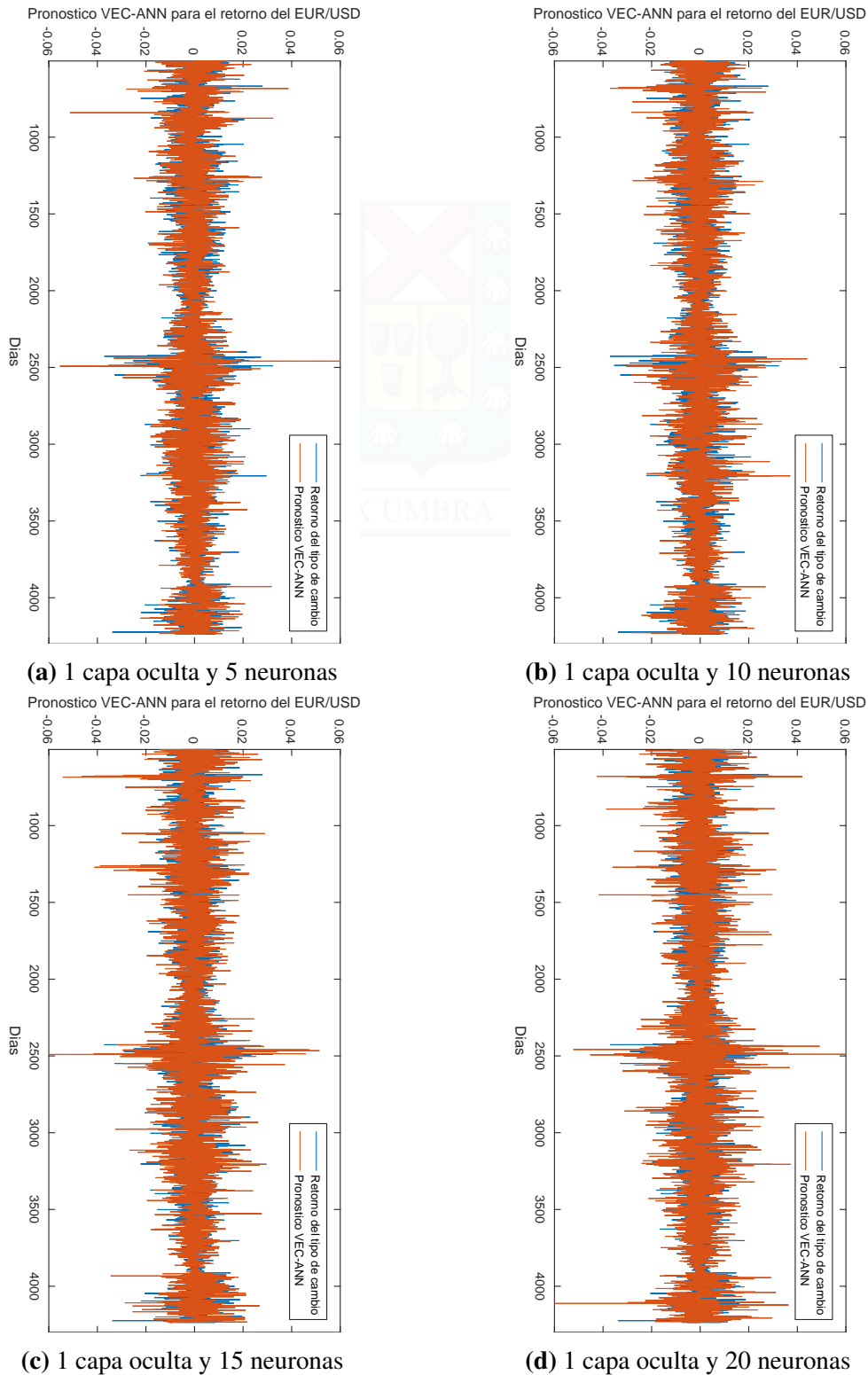


Figura A.1: Pronósticos del modelo VEC-ANN para el retorno del tipo de cambio EUR/USD con una ventana de 42 días, para 1 capa oculta con las configuraciones (a), (b), (c) y (d). Fuente: Elaboración propia.

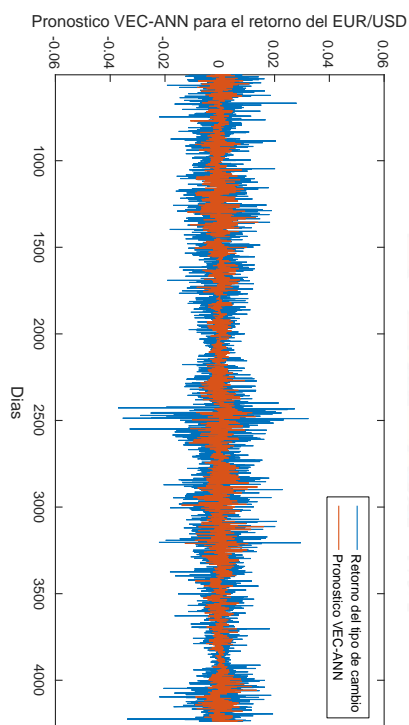
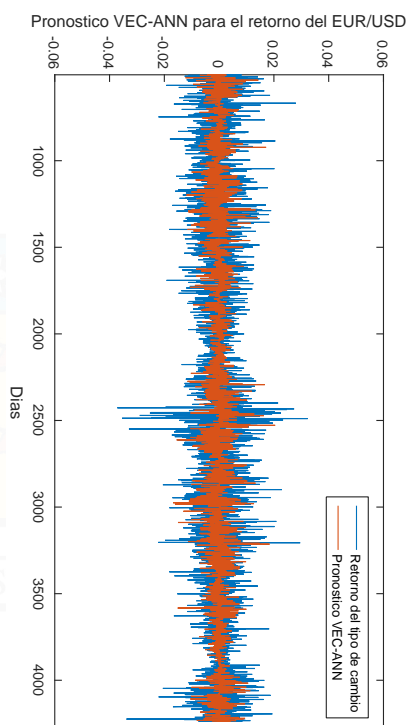
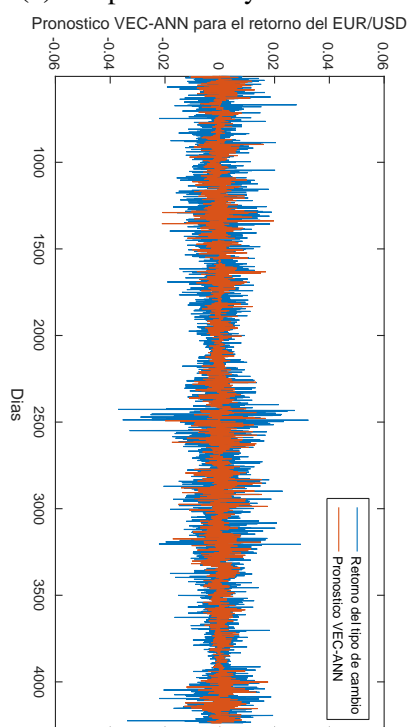
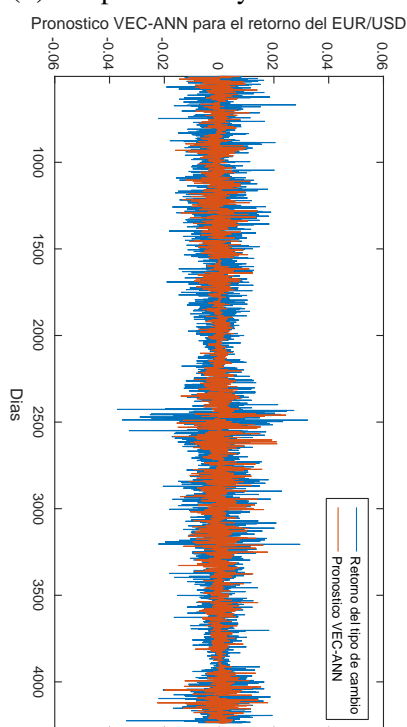
**(a)** 2 capas ocultas y 5 neuronas**(b)** 2 capas ocultas y 10 neuronas**(c)** 2 capas ocultas y 15 neuronas**(d)** 2 capas ocultas y 20 neuronas

Figura A.2: Pronósticos del modelo VEC-ANN para el retorno del tipo de cambio EUR/USD con una ventana de 42 días, para 2 capas ocultas con las configuraciones **(a)**, **(b)**, **(c)** y **(d)**. Fuente: Elaboración propia.

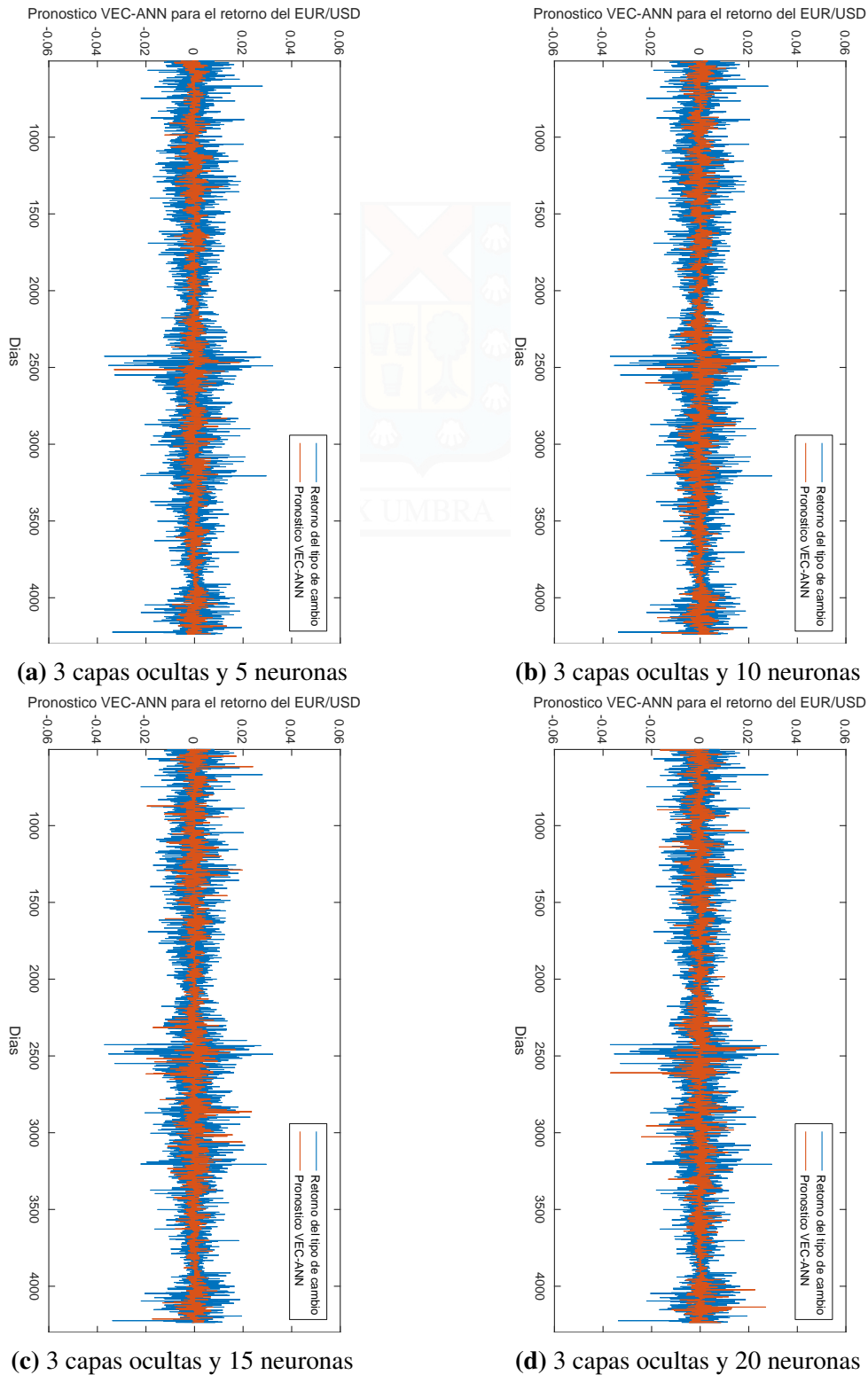


Figura A.3: Pronósticos del modelo VEC-ANN para el retorno del tipo de cambio EUR/USD con una ventana de 42 días, para 3 capas ocultas con las configuraciones (a), (b), (c) y (d). Fuente: Elaboración propia.

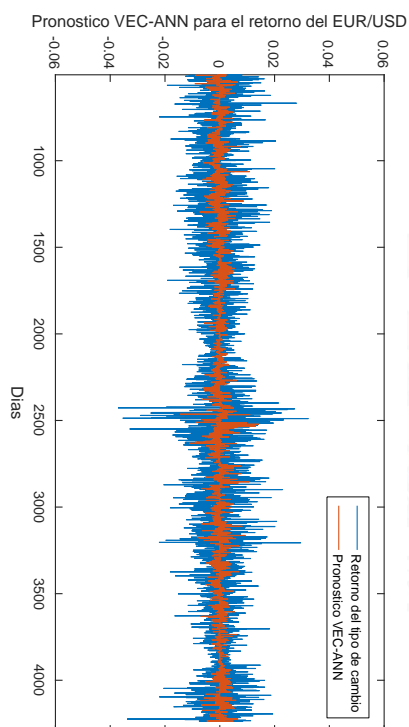
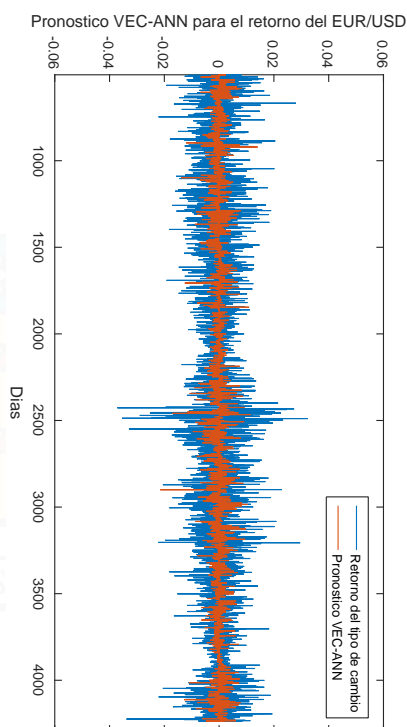
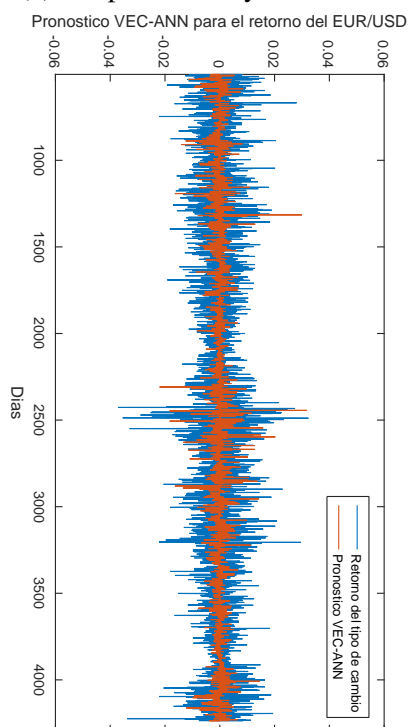
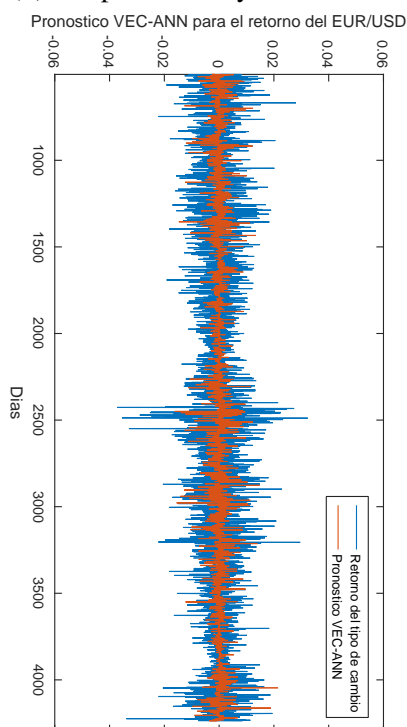
**(a)** 4 capas ocultas y 5 neuronas**(b)** 4 capas ocultas y 10 neuronas**(c)** 4 capas ocultas y 15 neuronas**(d)** 4 capas ocultas y 20 neuronas

Figura A.4: Pronósticos del modelo VEC-ANN para el retorno del tipo de cambio EUR/USD con una ventana de 42 días, para 4 capas ocultas con las configuraciones **(a)**, **(b)**, **(c)** y **(d)**. Fuente: Elaboración propia.

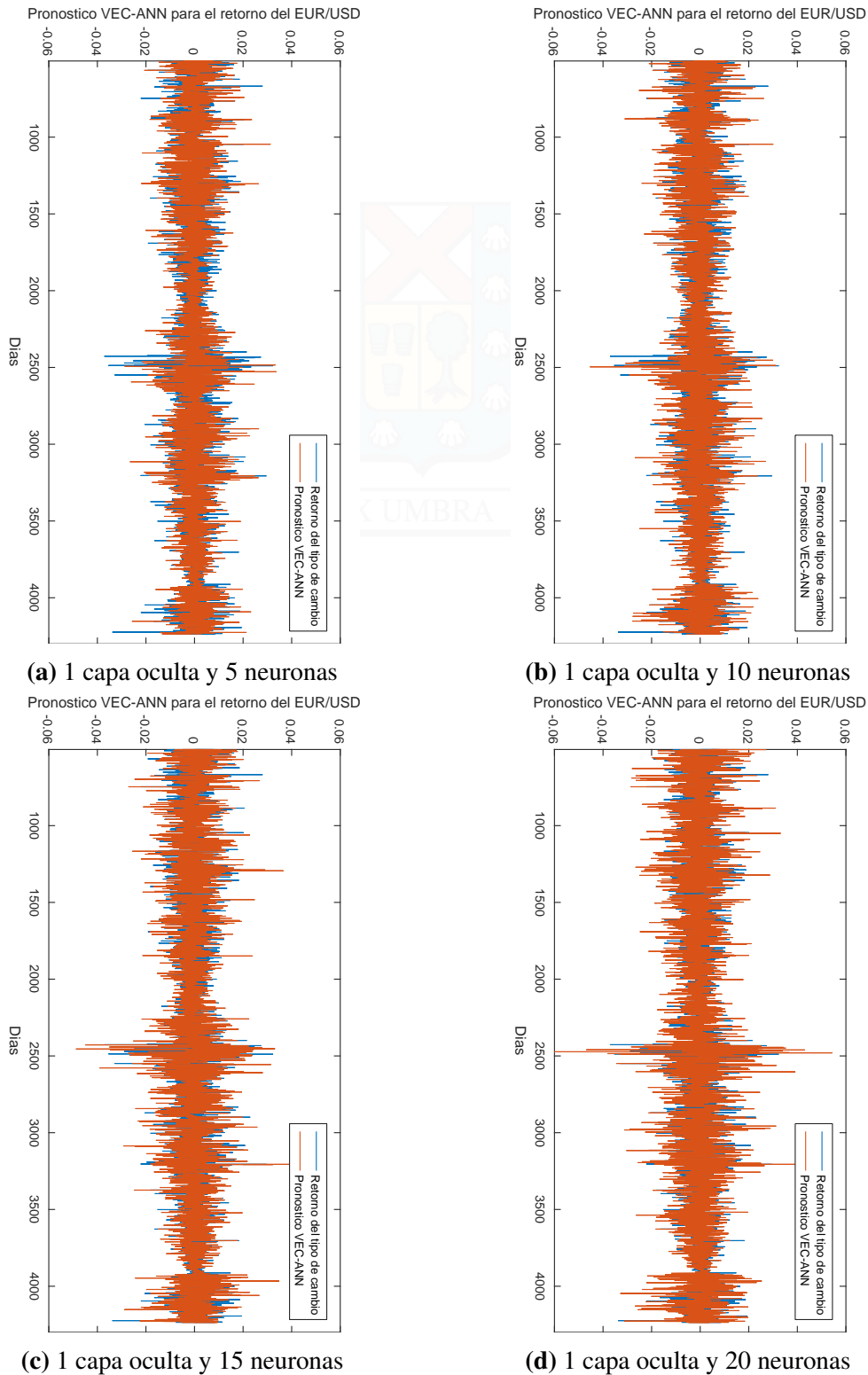
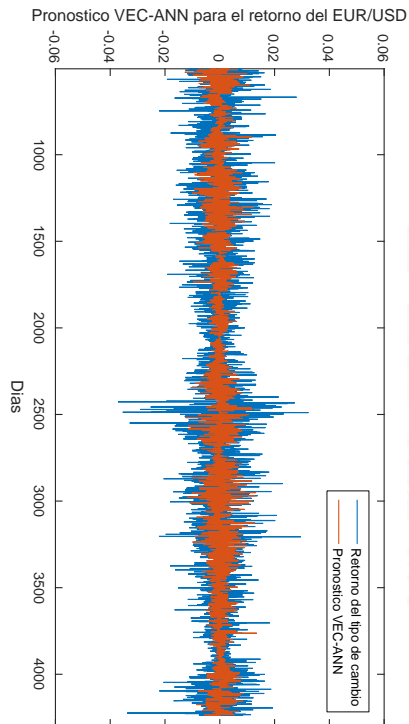
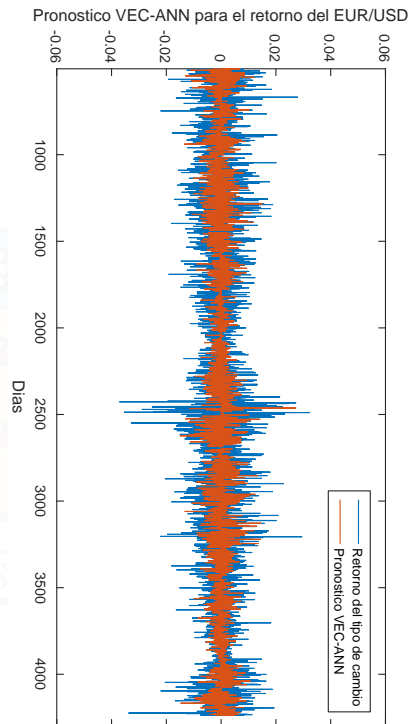


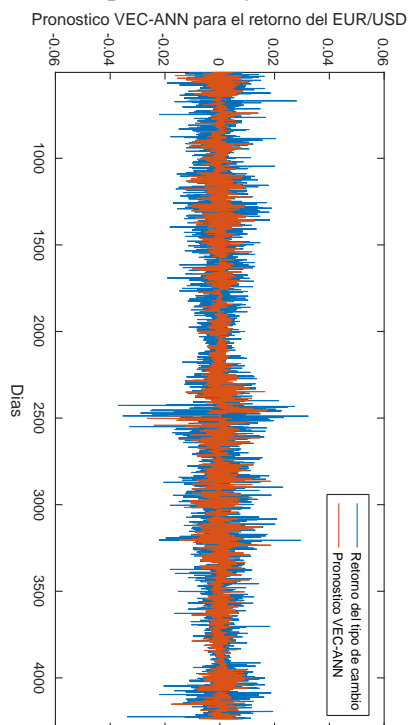
Figura A.5: Pronósticos del modelo VEC-ANN para el retorno del tipo de cambio EUR/USD con una ventana de 63 días, para 1 capa oculta con las configuraciones (a), (b), (c) y (d). Fuente: Elaboración propia.



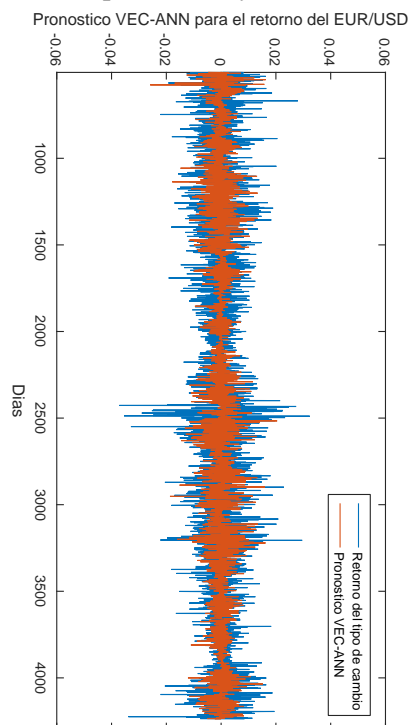
(a) 2 capas ocultas y 5 neuronas



(b) 2 capas ocultas y 10 neuronas



(c) 2 capas ocultas y 15 neuronas



(d) 2 capas ocultas y 20 neuronas

Figura A.6: Pronósticos del modelo VEC-ANN para el retorno del tipo de cambio EUR/USD con una ventana de 63 días, para 2 capas ocultas con las configuraciones (a), (b), (c) y (d). Fuente: Elaboración propia.

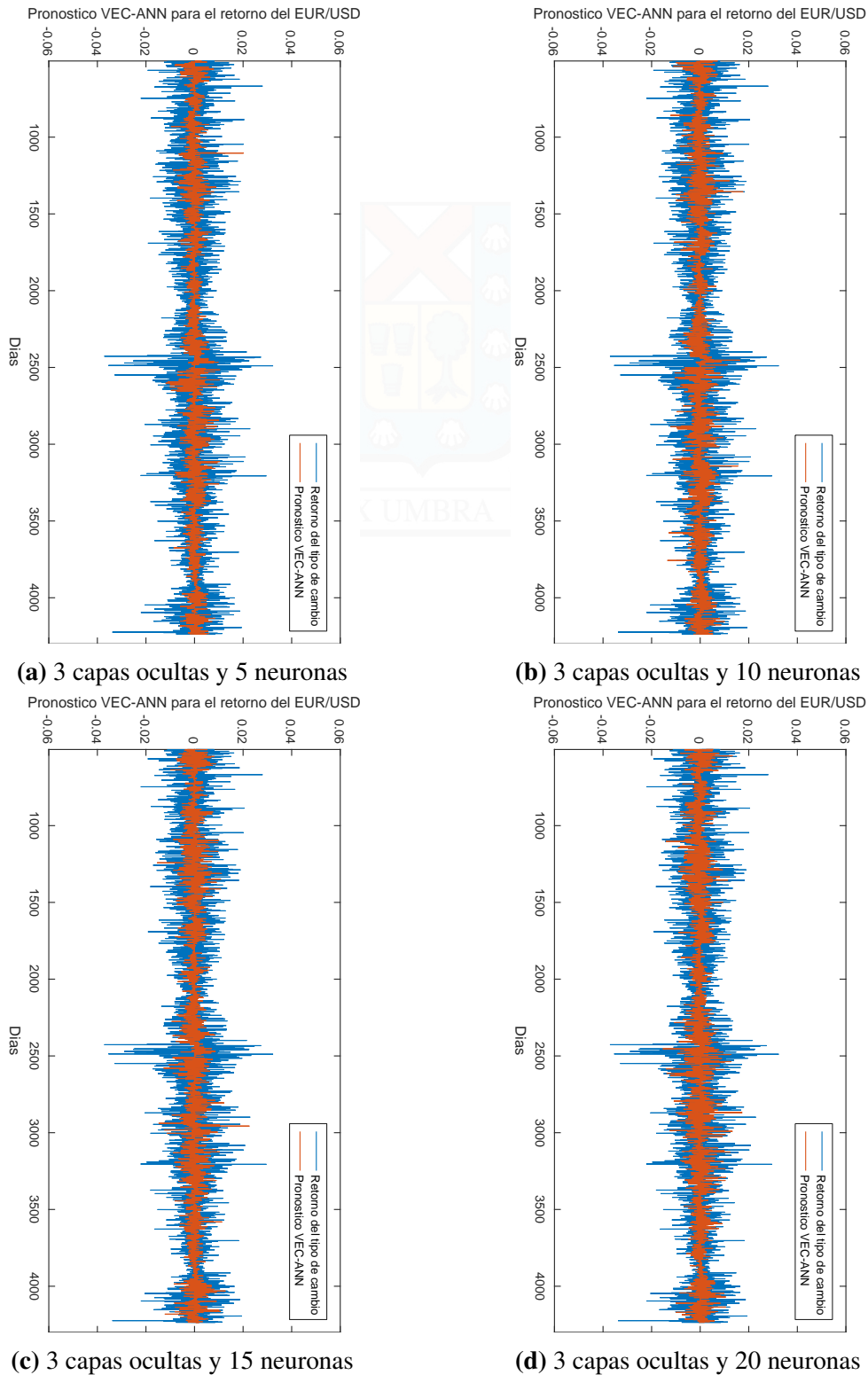


Figura A.7: Pronósticos del modelo VEC-ANN para el retorno del tipo de cambio EUR/USD con una ventana de 63 días, para 3 capas ocultas con las configuraciones (a), (b), (c) y (d). Fuente: Elaboración propia.

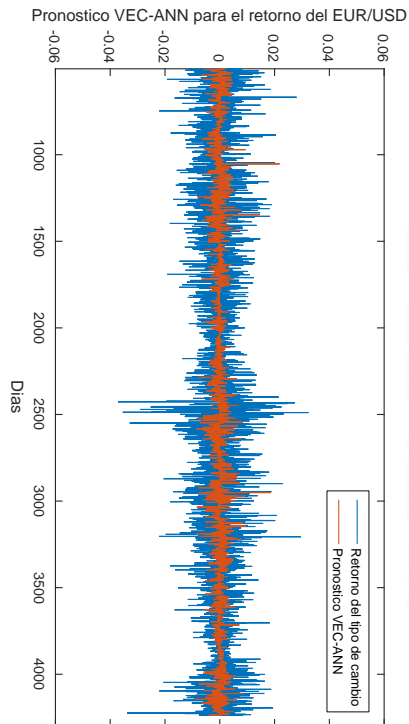
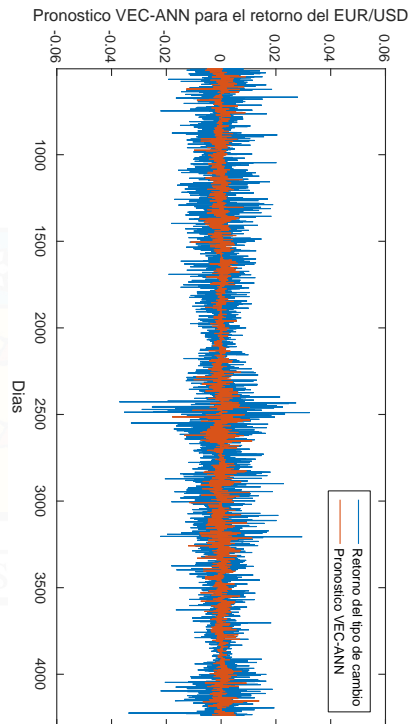
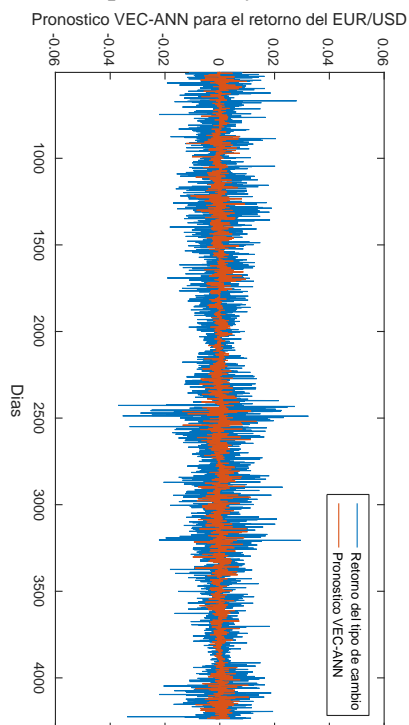
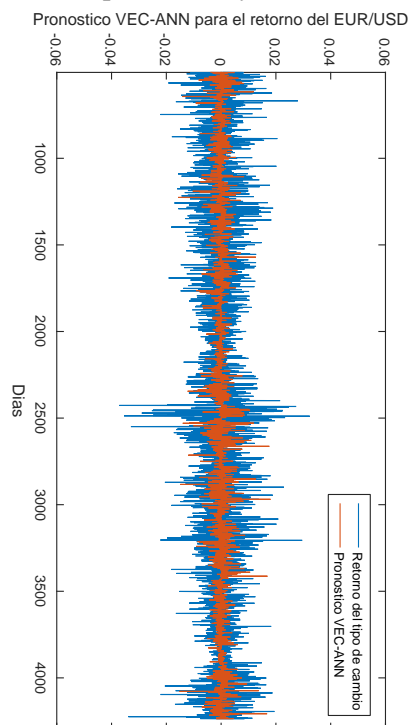
**(a)** 4 capas ocultas y 5 neuronas**(b)** 4 capas ocultas y 10 neuronas**(c)** 4 capas ocultas y 15 neuronas**(d)** 4 capas ocultas y 20 neuronas

Figura A.8: Pronósticos del modelo VEC-ANN para el retorno del tipo de cambio EUR/USD con una ventana de 63 días, para 4 capas ocultas con las configuraciones **(a)**, **(b)**, **(c)** y **(d)**. Fuente: Elaboración propia.

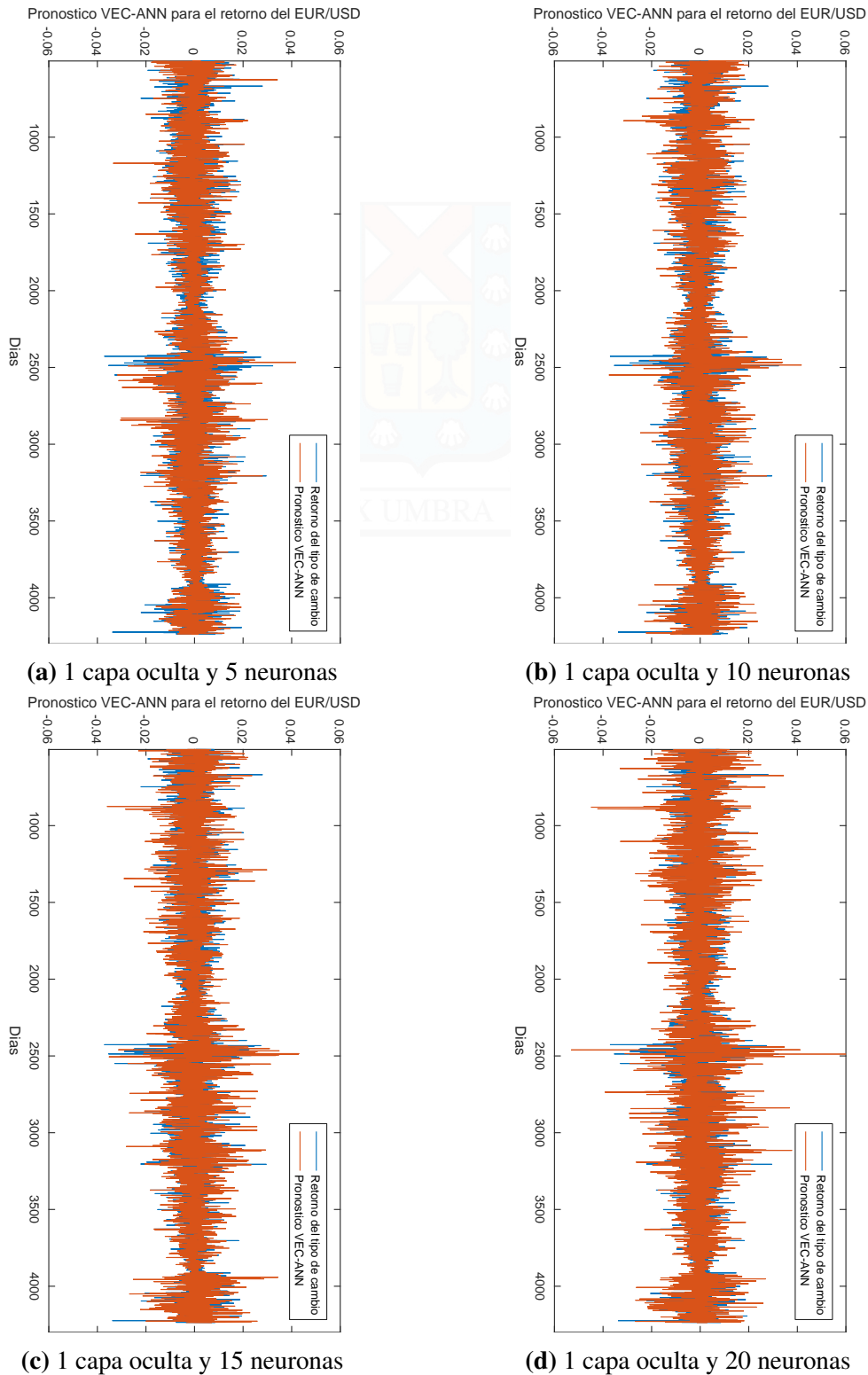
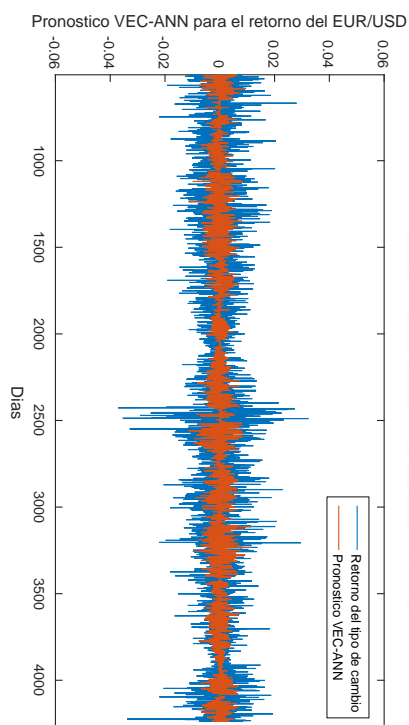
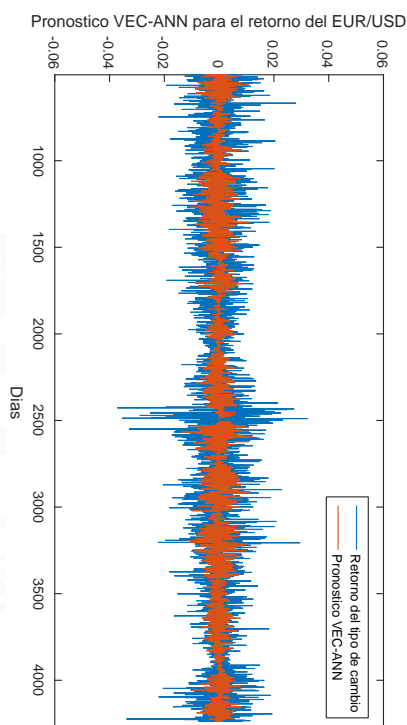


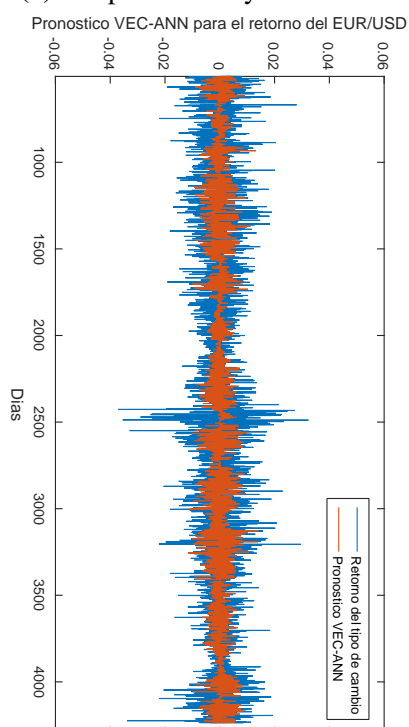
Figura A.9: Pronósticos del modelo VEC-ANN para el retorno del tipo de cambio EUR/USD con una ventana de 84 días, para 1 capa oculta con las configuraciones (a), (b), (c) y (d). Fuente: Elaboración propia.



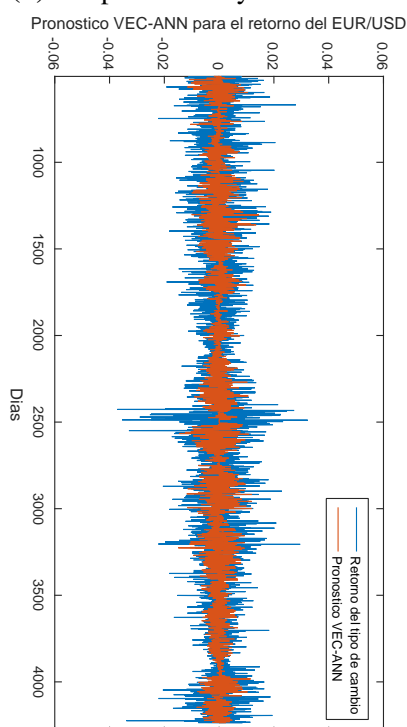
(a) 2 capas ocultas y 5 neuronas



(b) 2 capas ocultas y 10 neuronas



(c) 2 capas ocultas y 15 neuronas



(d) 2 capas ocultas y 20 neuronas

Figura A.10: Pronósticos del modelo VEC-ANN para el retorno del tipo de cambio EUR/USD con una ventana de 84 días, para 2 capas ocultas con las configuraciones (a), (b), (c) y (d). Fuente: Elaboración propia.

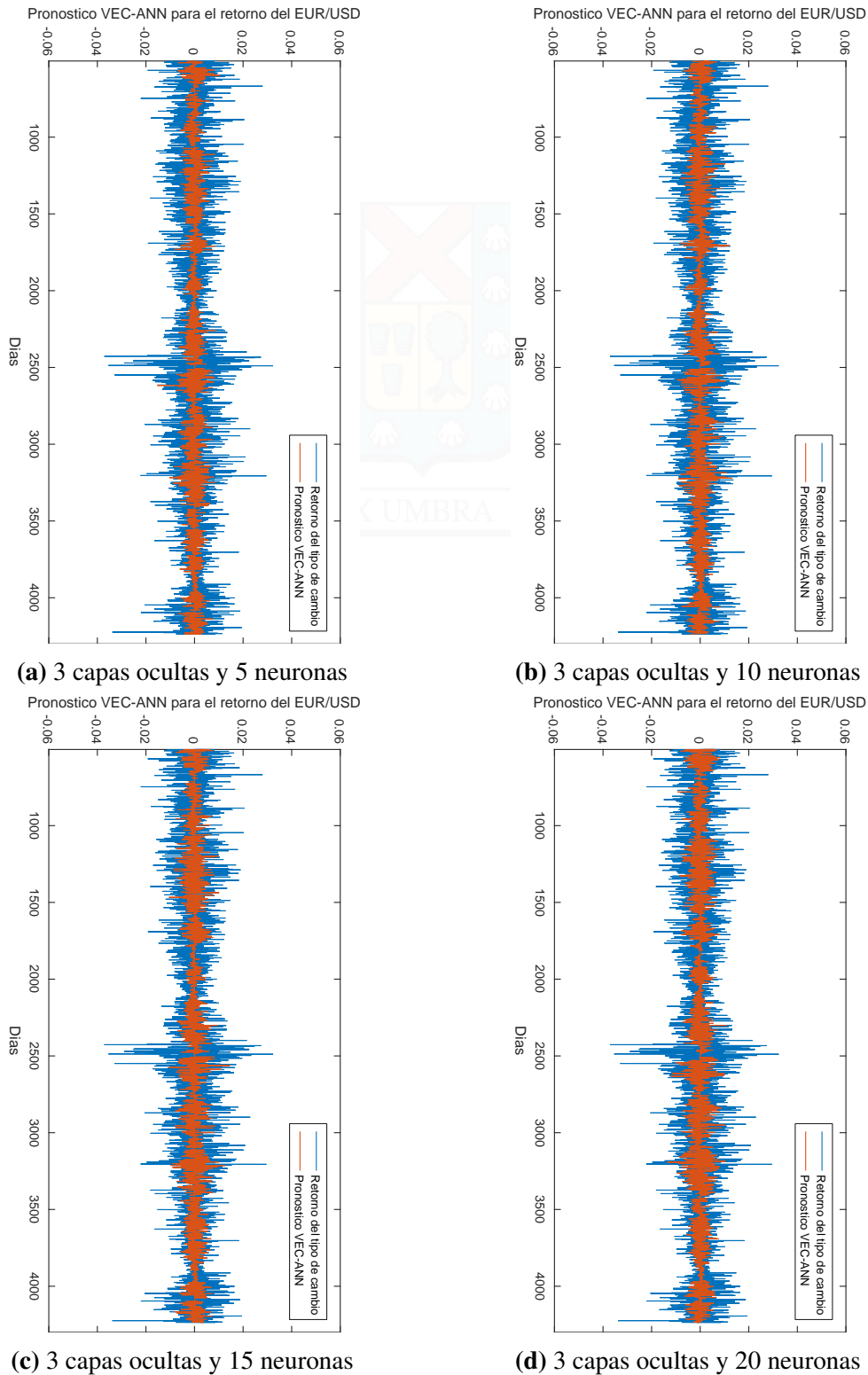


Figura A.11: Pronósticos del modelo VEC-ANN para el retorno del tipo de cambio EUR/USD con una ventana de 84 días, para 3 capas ocultas con las configuraciones (a), (b), (c) y (d). Fuente: Elaboración propia.

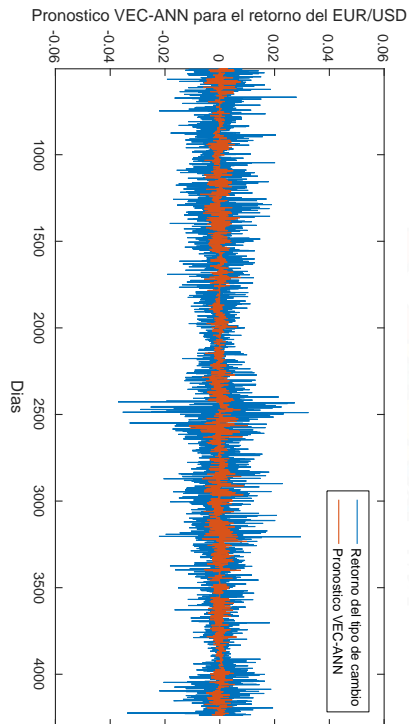
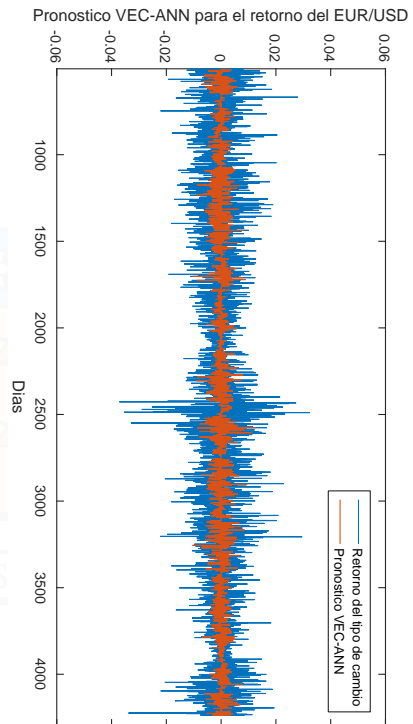
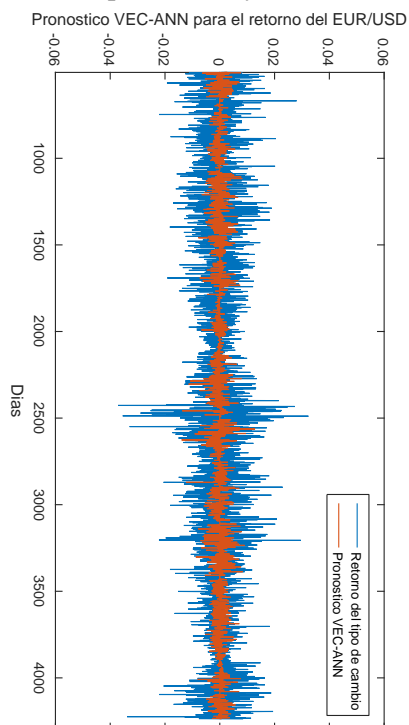
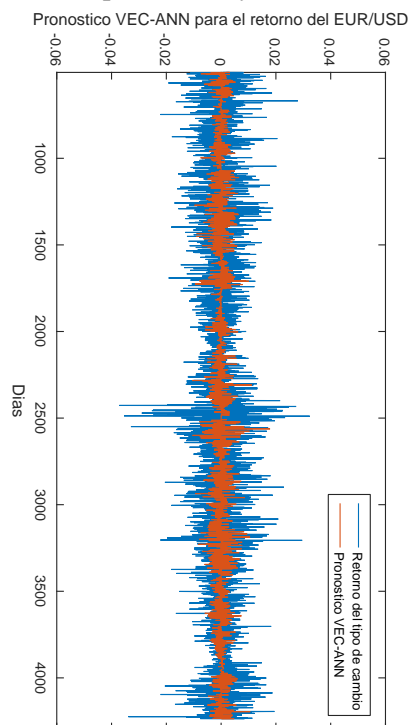
**(a)** 4 capas ocultas y 5 neuronas**(b)** 4 capas ocultas y 10 neuronas**(c)** 4 capas ocultas y 15 neuronas**(d)** 4 capas ocultas y 20 neuronas

Figura A.12: Pronósticos del modelo VEC-ANN para el retorno del tipo de cambio EUR/USD con una ventana de 84 días, para 4 capas ocultas con las configuraciones **(a)**, **(b)**, **(c)** y **(d)**. Fuente: Elaboración propia.

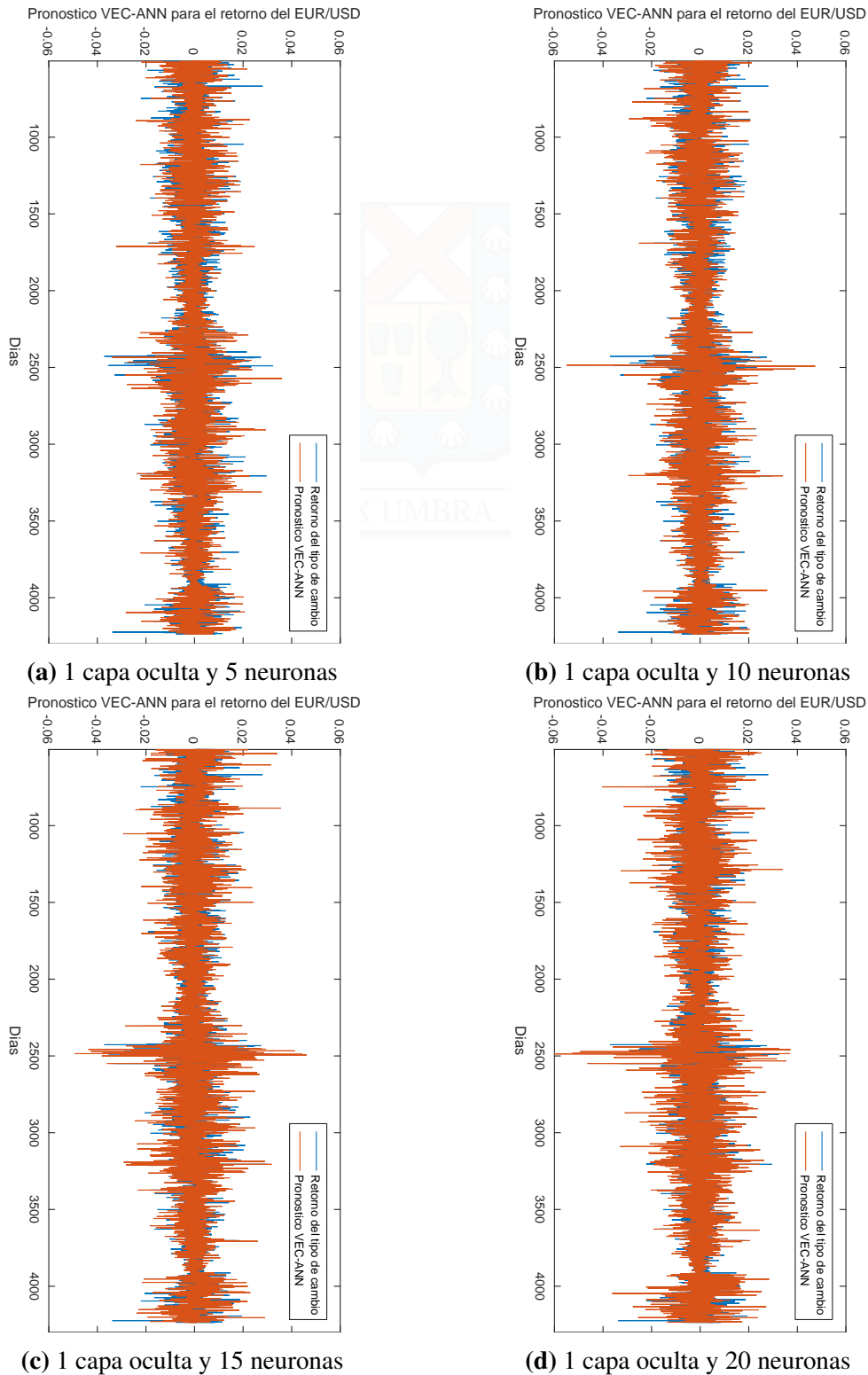


Figura A.13: Pronósticos del modelo VEC-ANN para el retorno del tipo de cambio EUR/USD con una ventana de 126 días, para 1 capa oculta con las configuraciones (a), (b), (c) y (d). Fuente: Elaboración propia.

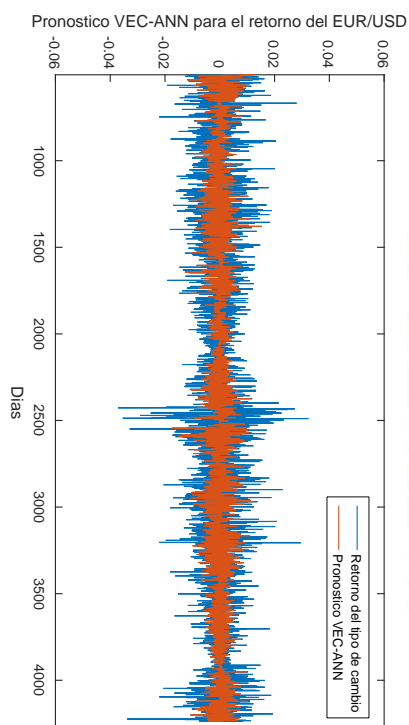
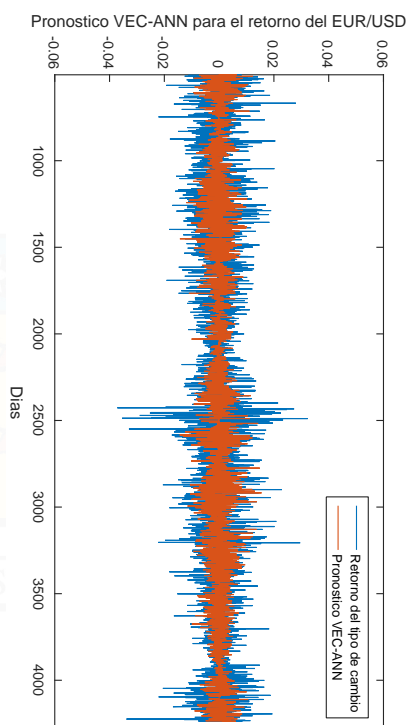
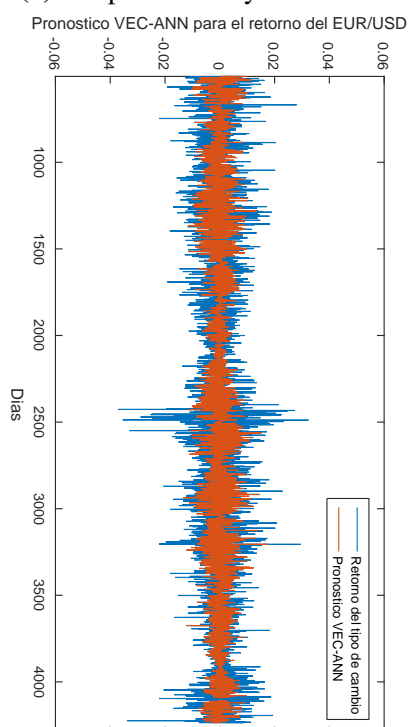
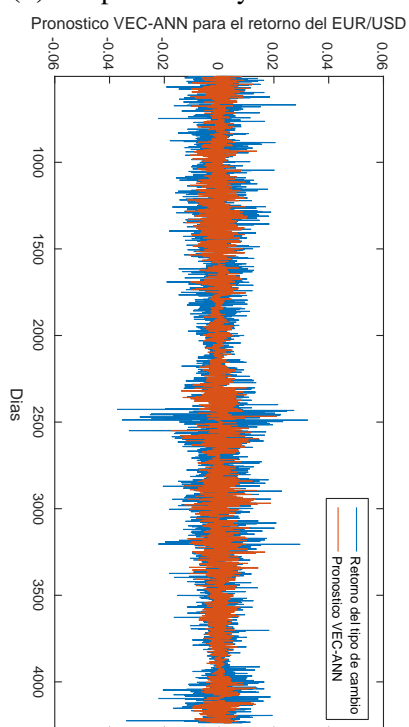
**(a)** 2 capas ocultas y 5 neuronas**(b)** 2 capas ocultas y 10 neuronas**(c)** 2 capas ocultas y 15 neuronas**(d)** 2 capas ocultas y 20 neuronas

Figura A.14: Pronósticos del modelo VEC-ANN para el retorno del tipo de cambio EUR/USD con una ventana de 126 días, para 2 capas ocultas con las configuraciones **(a)**, **(b)**, **(c)** y **(d)**. Fuente: Elaboración propia.

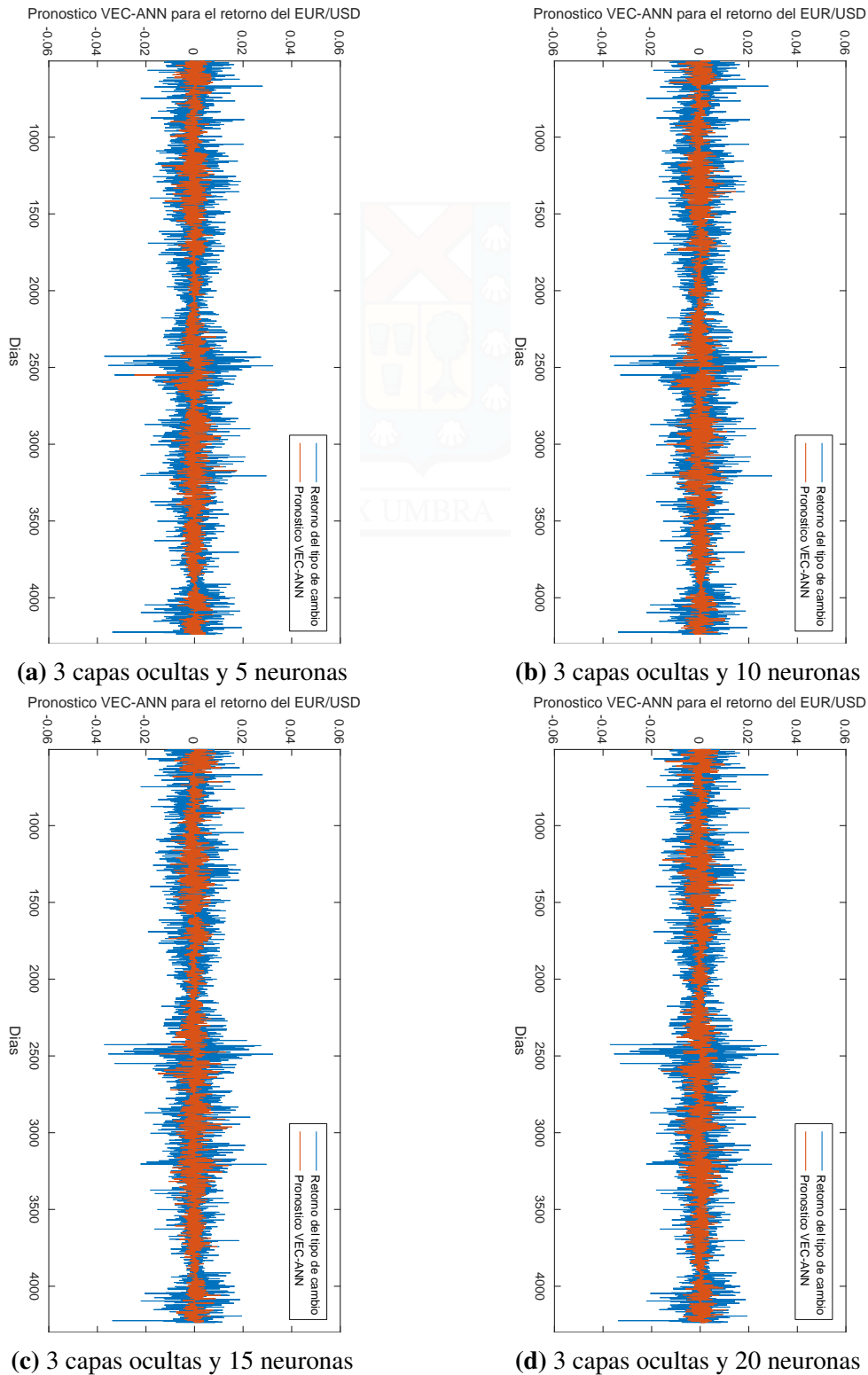


Figura A.15: Pronósticos del modelo VEC-ANN para el retorno del tipo de cambio EUR/USD con una ventana de 126 días, para 3 capas ocultas con las configuraciones (a), (b), (c) y (d). Fuente: Elaboración propia.

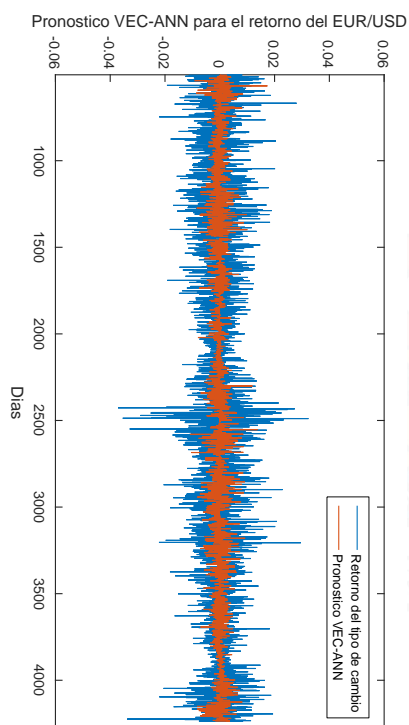
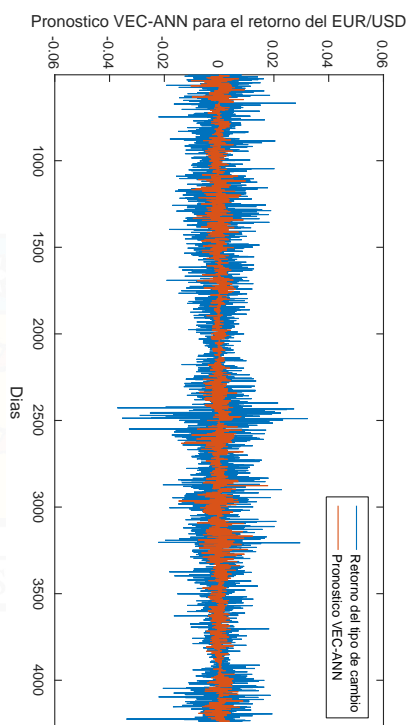
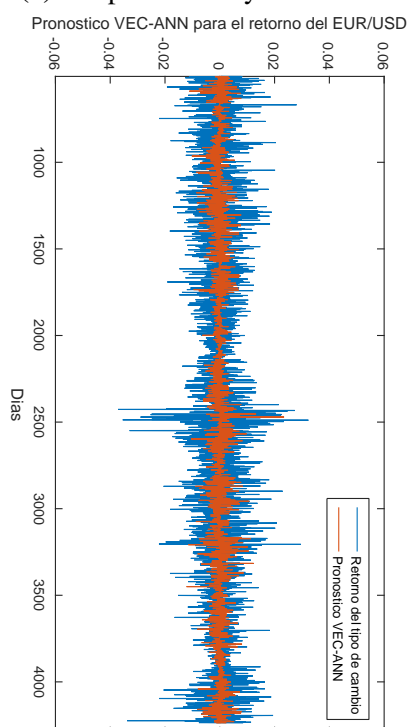
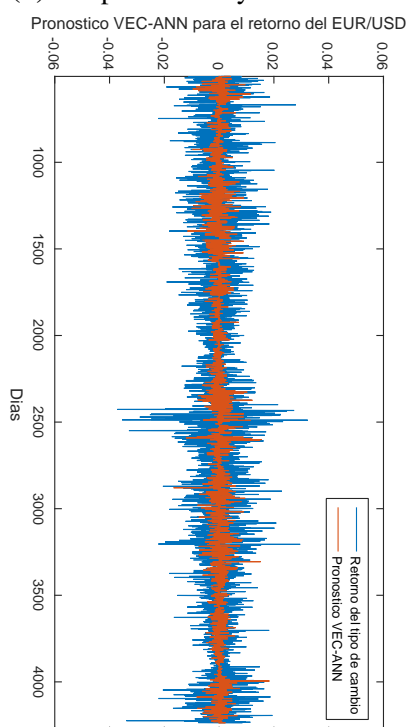
**(a)** 4 capas ocultas y 5 neuronas**(b)** 4 capas ocultas y 10 neuronas**(c)** 4 capas ocultas y 15 neuronas**(d)** 4 capas ocultas y 20 neuronas

Figura A.16: Pronósticos del modelo VEC-ANN para el retorno del tipo de cambio EUR/USD con una ventana de 126 días, para 4 capas ocultas con las configuraciones **(a)**, **(b)**, **(c)** y **(d)**. Fuente: Elaboración propia.

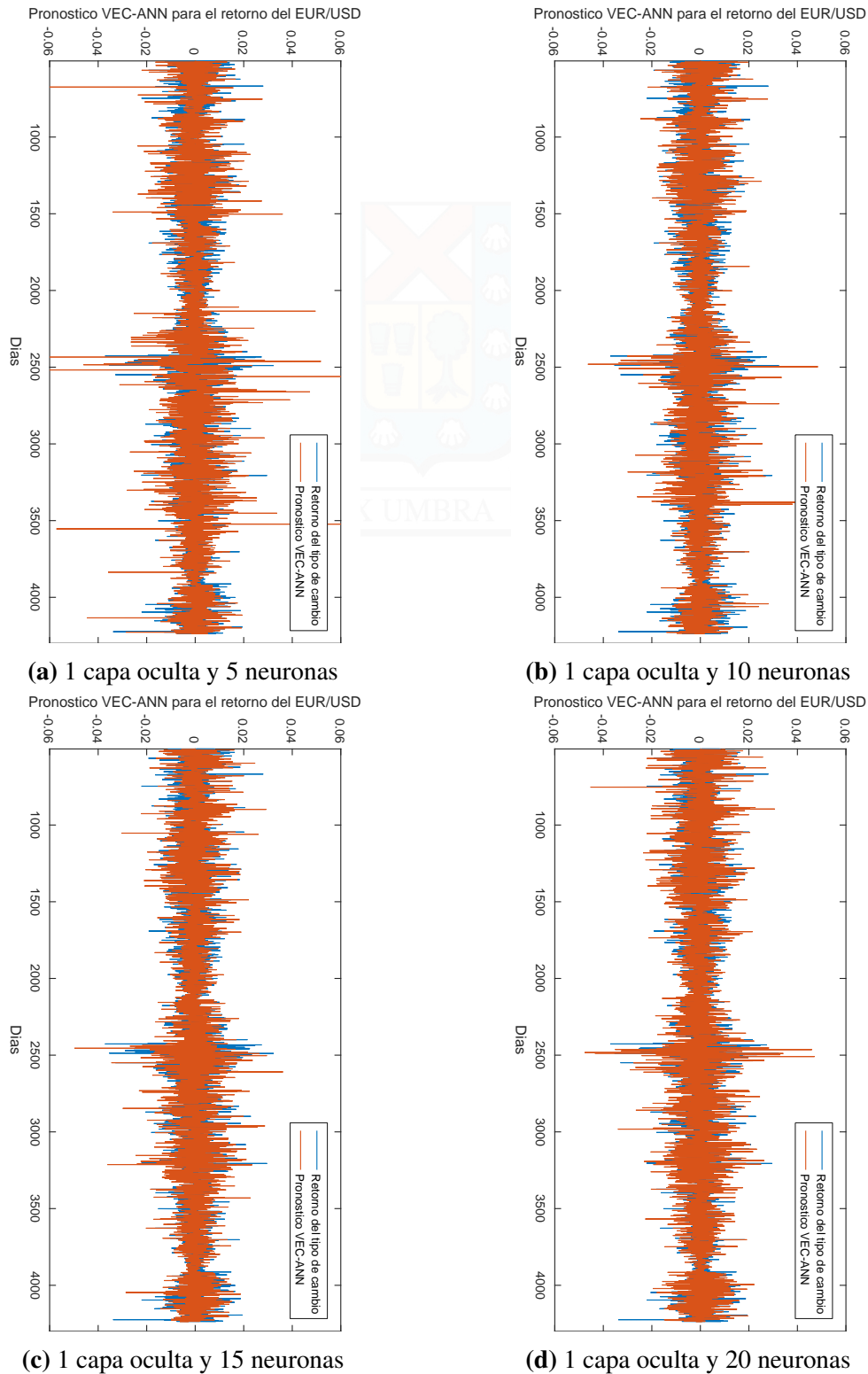


Figura A.17: Pronósticos del modelo VEC-ANN para el retorno del tipo de cambio EUR/USD con una ventana de 252 días, para 1 capa oculta con las configuraciones (a), (b), (c) y (d). Fuente: Elaboración propia.

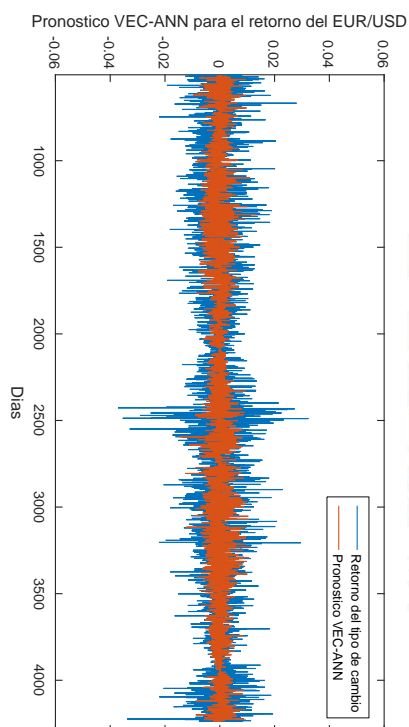
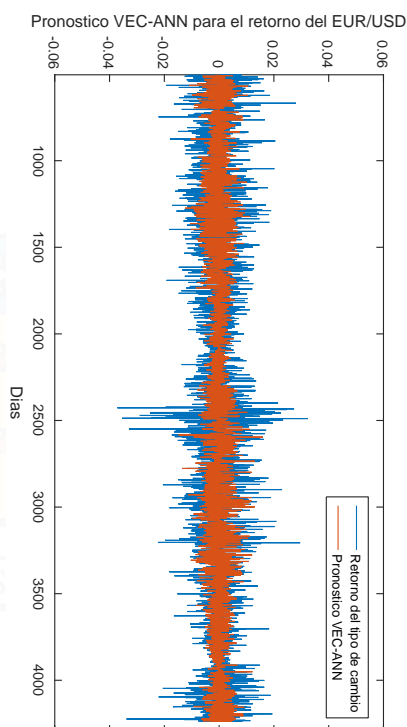
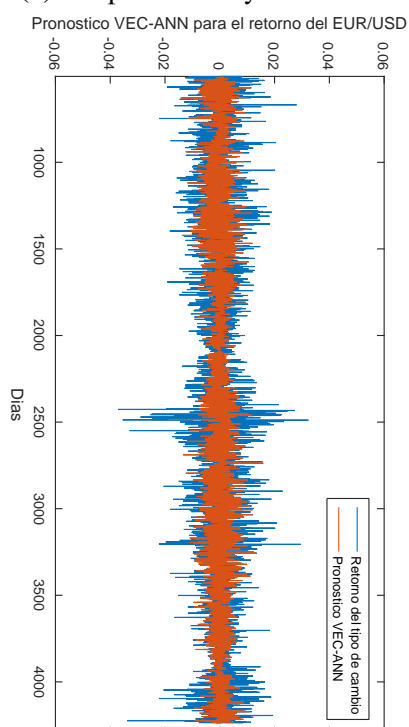
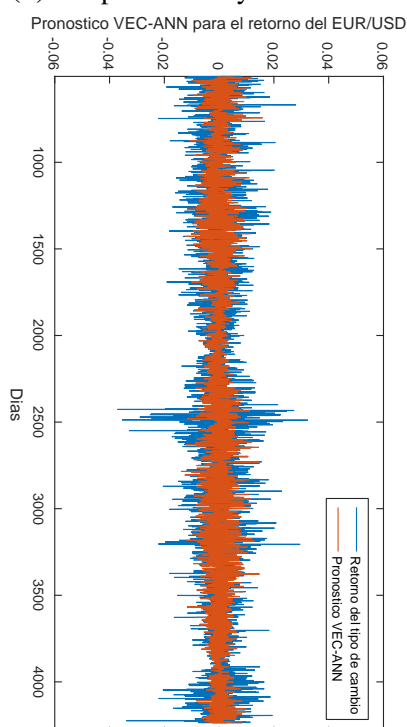
**(a)** 2 capas ocultas y 5 neuronas**(b)** 2 capas ocultas y 10 neuronas**(c)** 2 capas ocultas y 15 neuronas**(d)** 2 capas ocultas y 20 neuronas

Figura A.18: Pronósticos del modelo VEC-ANN para el retorno del tipo de cambio EUR/USD con una ventana de 252 días, para 2 capas ocultas con las configuraciones **(a)**, **(b)**, **(c)** y **(d)**. Fuente: Elaboración propia.

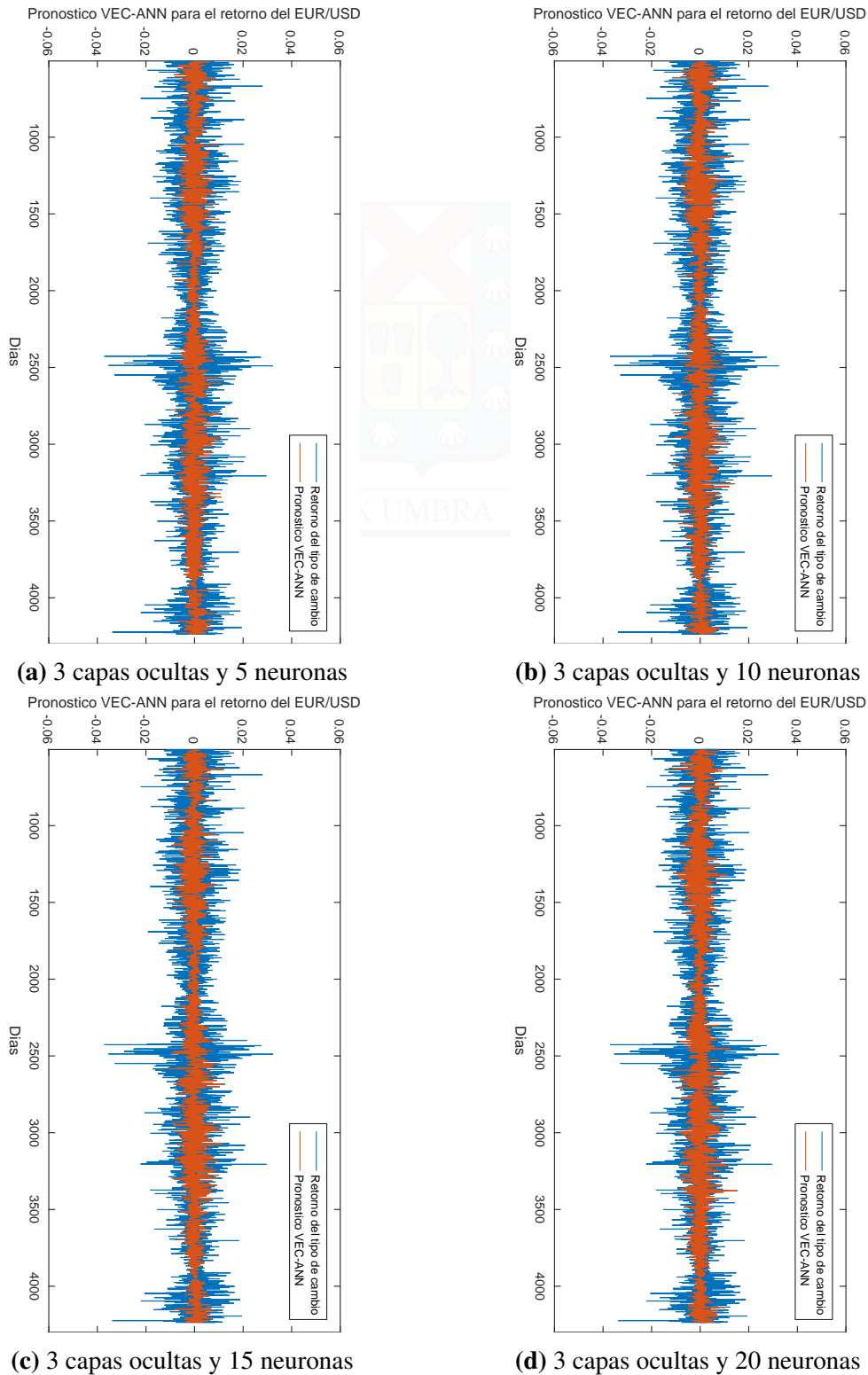


Figura A.19: Pronósticos del modelo VEC-ANN para el retorno del tipo de cambio EUR/USD con una ventana de 252 días, para 3 capas ocultas con las configuraciones (a), (b), (c) y (d). Fuente: Elaboración propia.

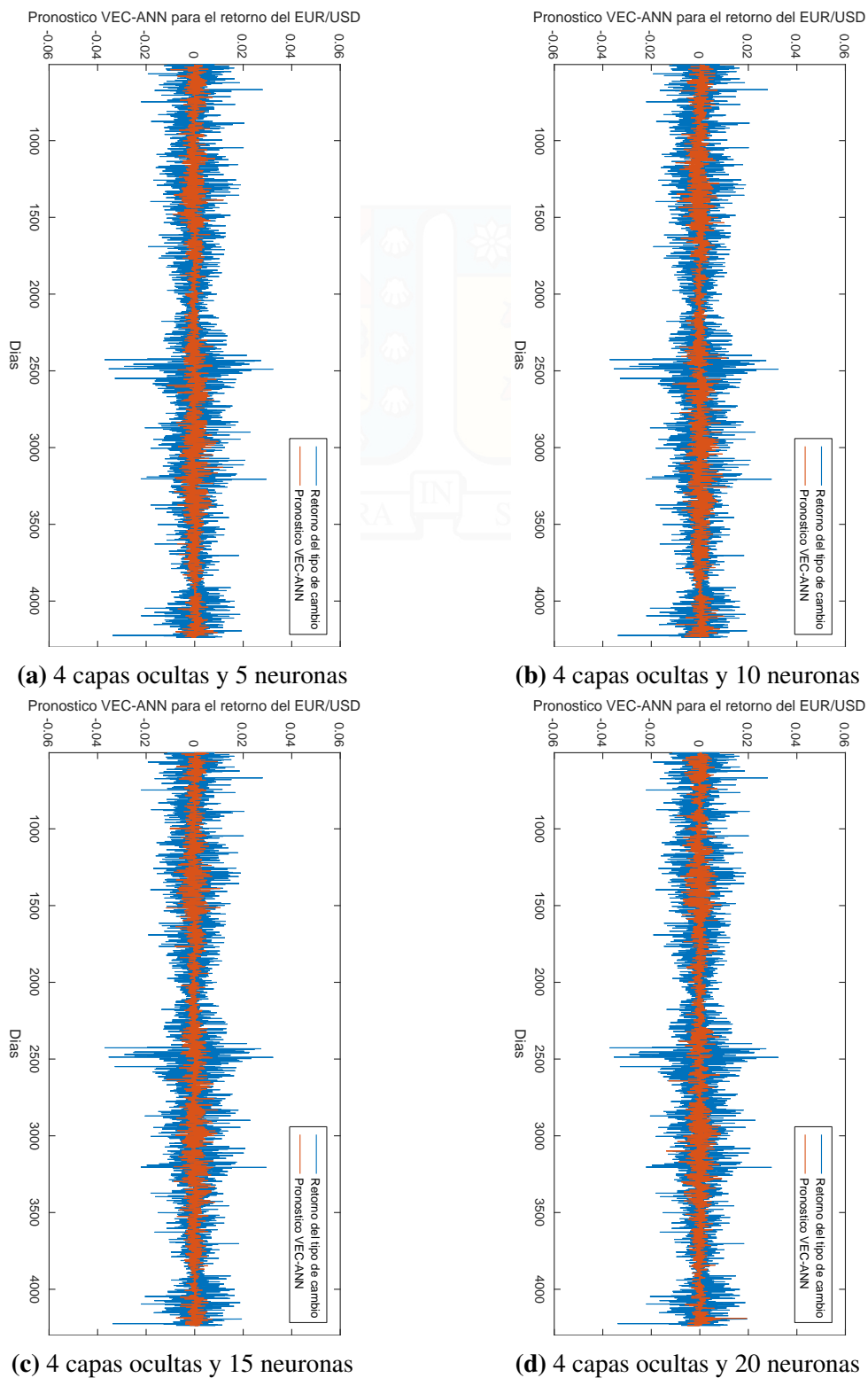


Figura A.20: Pronósticos del modelo VEC-ANN para el retorno del tipo de cambio EUR/USD con una ventana de 252 días, para 4 capas ocultas con las configuraciones (a), (b), (c) y (d). Fuente: Elaboración propia.

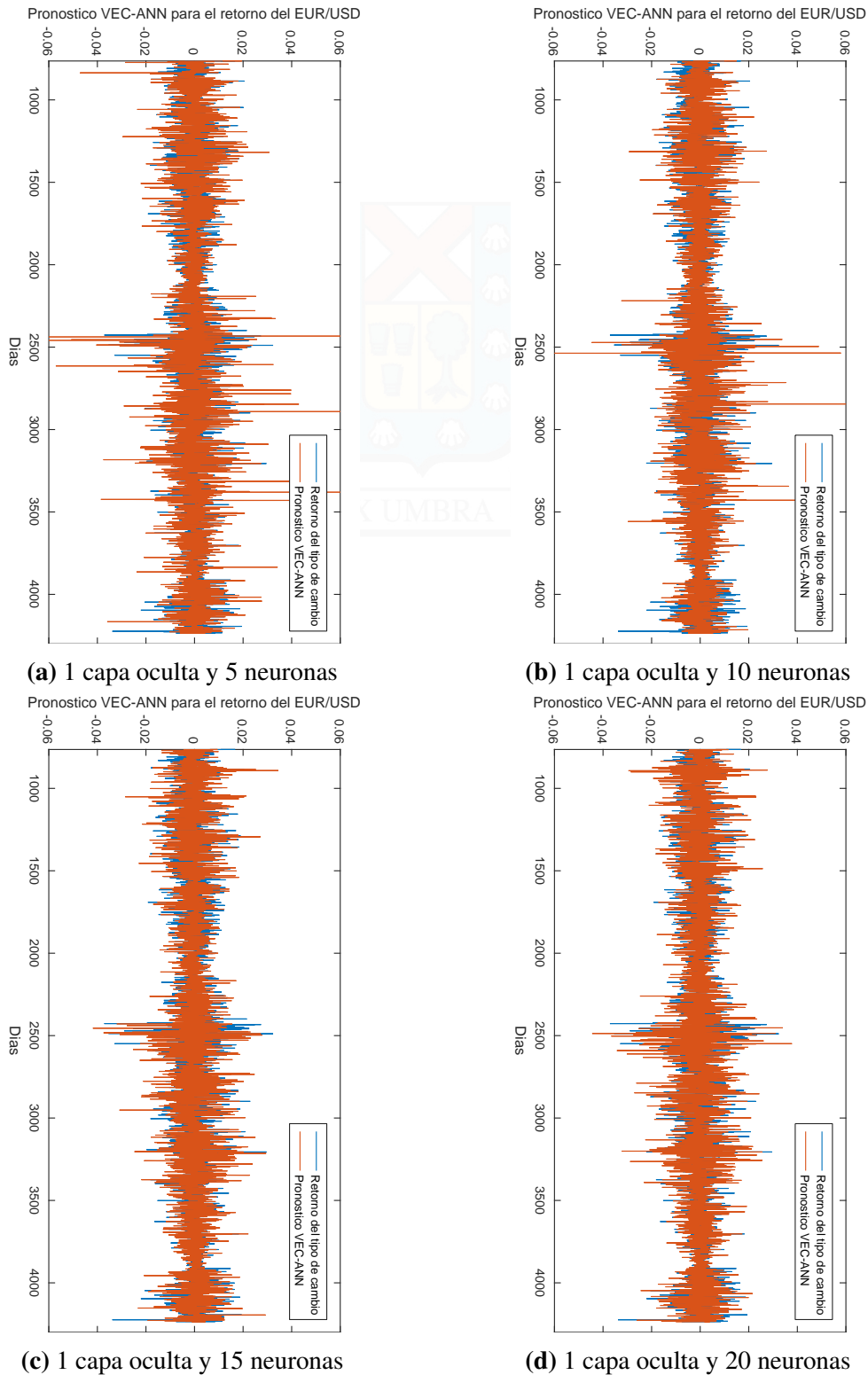
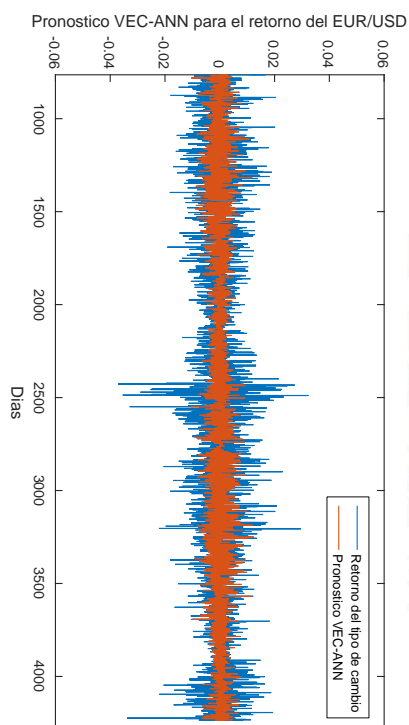
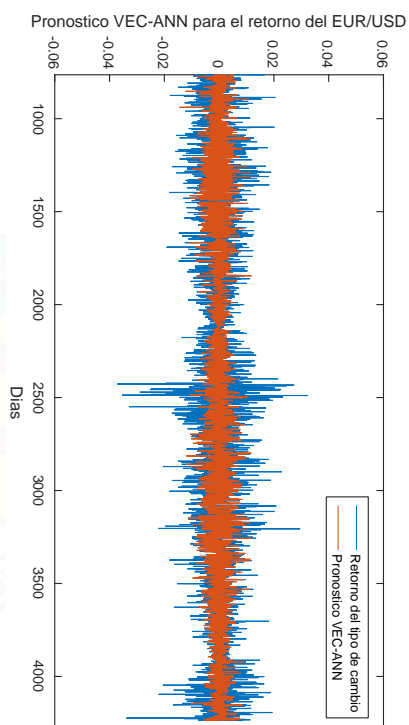


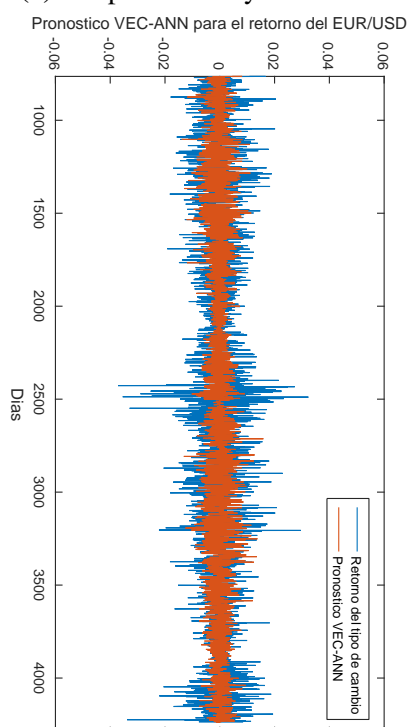
Figura A.21: Pronósticos del modelo VEC-ANN para el retorno del tipo de cambio EUR/USD con una ventana de 378 días, para 1 capa oculta con las configuraciones (a), (b), (c) y (d). Fuente: Elaboración propia.



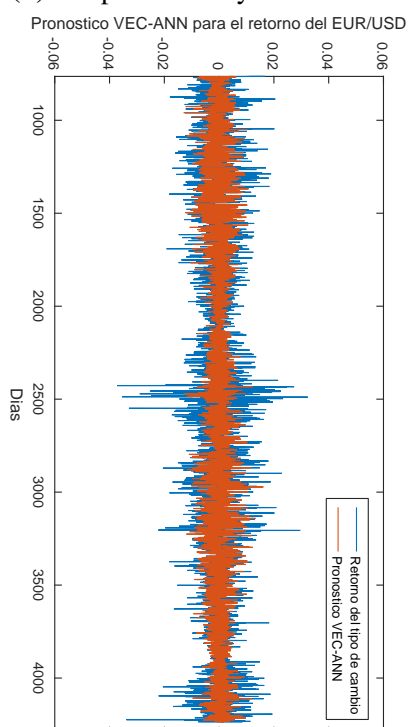
(a) 2 capas ocultas y 5 neuronas



(b) 2 capas ocultas y 10 neuronas



(c) 2 capas ocultas y 15 neuronas



(d) 2 capas ocultas y 20 neuronas

Figura A.22: Pronósticos del modelo VEC-ANN para el retorno del tipo de cambio EUR/USD con una ventana de 378 días, para 2 capas ocultas con las configuraciones (a), (b), (c) y (d). Fuente: Elaboración propia.

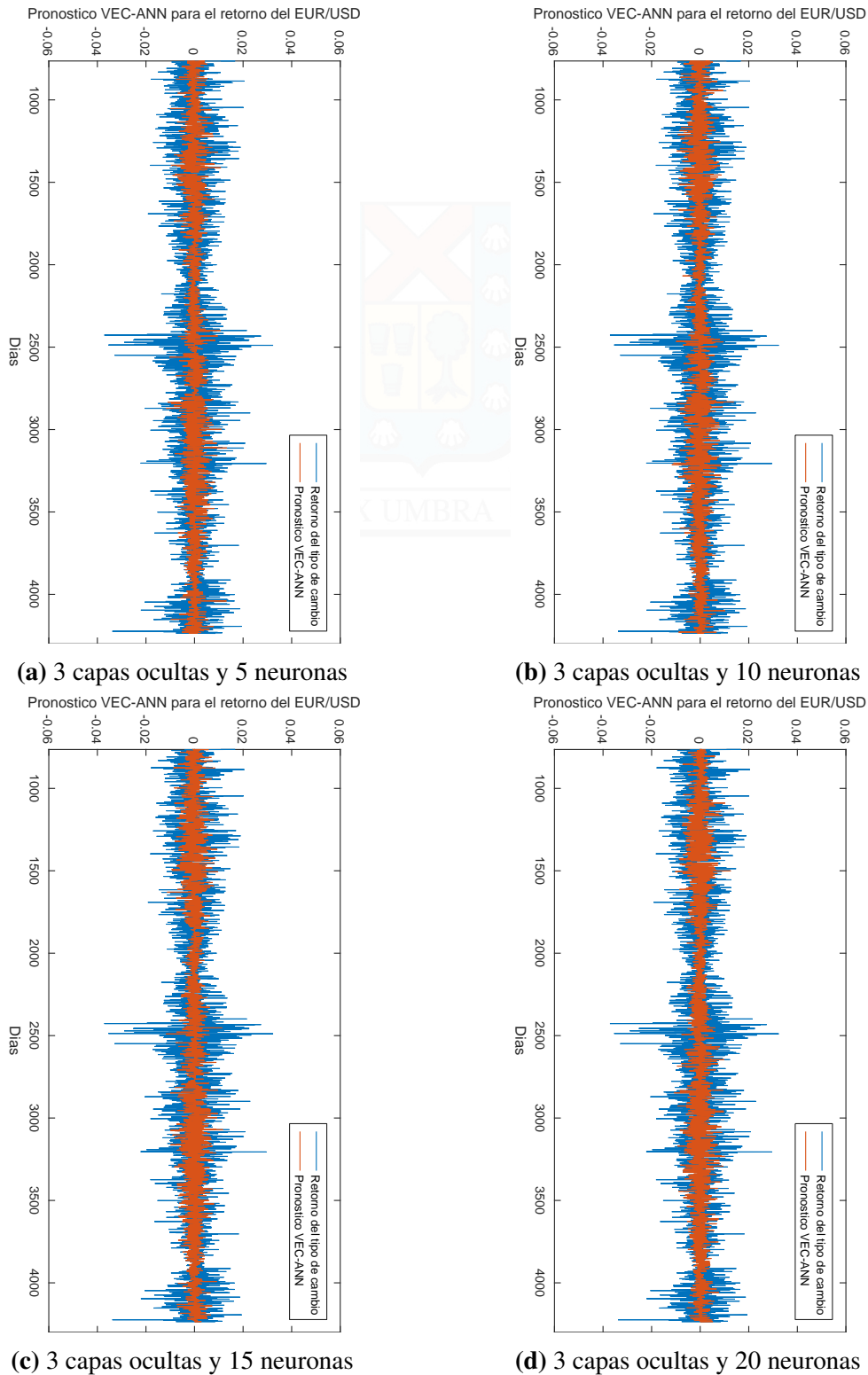


Figura A.23: Pronósticos del modelo VEC-ANN para el retorno del tipo de cambio EUR/USD con una ventana de 378 días, para 3 capas ocultas con las configuraciones (a), (b), (c) y (d). Fuente: Elaboración propia.

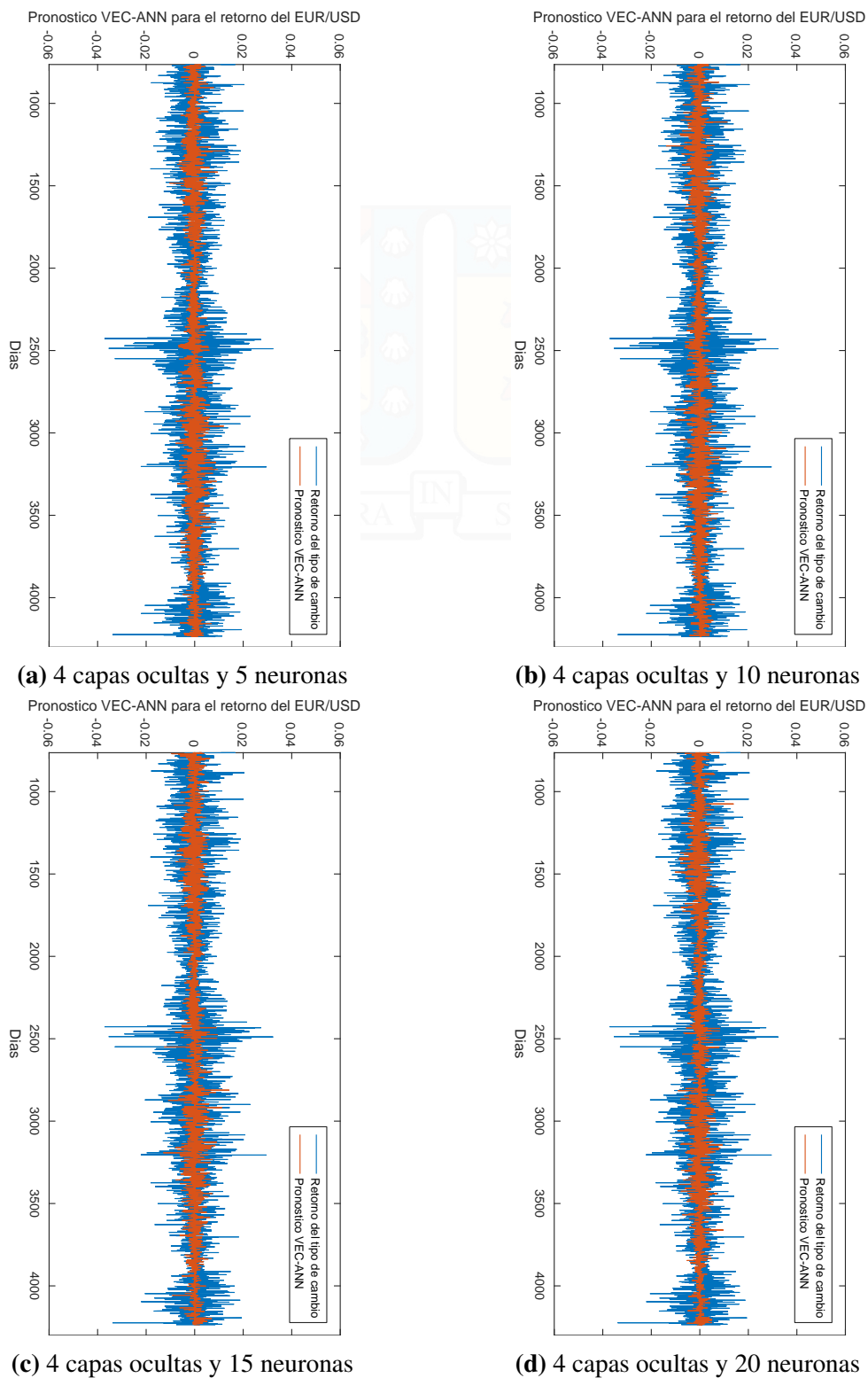


Figura A.24: Pronósticos del modelo VEC-ANN para el retorno del tipo de cambio EUR/USD con una ventana de 378 días, para 4 capas ocultas con las configuraciones (a), (b), (c) y (d). Fuente: Elaboración propia.