

Universidad de Murcia
Facultad de Matemáticas

TRABAJO FIN DE MÁSTER



Modelos predictivos para
el mercado FOREX

Realizado por:
Alejandro Huertas López

Septiembre 2015

Modelos predictivos para el mercado FOREX

Alejandro Huertas López

dirigido por

Francisco Esquembre Martínez, Javier G. Marín-Blázquez y
Sergio Navarro Sánchez

En memoria de Juan Huertas Ruiz

Resumen

Los modelos predictivos permiten al usuario estar preparado frente a posibles eventos futuros, aportándole estimaciones de lo que sucederá. De esta forma, cuando hablamos de modelos predictivos para el mercado FOREX nos referimos a herramientas que pronostiquen (con mayor o menor precisión) el movimiento futuro de los elementos del mercado.

El mercado FOREX o mercado de divisas es considerado el mercado financiero más importante del mundo. Mueve diariamente sumas de más de mil millones de dólares. El objetivo de este trabajo es elaborar una serie de herramientas y modelos que pronostiquen, a pequeña escala, el movimiento de los elementos de este mercado. Con pequeña escala nos referimos a pronósticos hora a hora, minuto a minuto, de forma que las estimaciones que llegásemos a dar no se viesen muy afectadas por decisiones económicas externas al mercado.

Se hablará principalmente de dos tipos de modelos predictivos: los modelos de series temporales y los modelos basados en técnicas de aprendizaje automático.

Los modelos de series temporales se basan en la teoría de procesos estocásticos. Suponen que toda serie temporal es una realización de un proceso estocástico, y en consecuencia, elaboran modelos que expliquen el proceso estocástico generador de la serie. Este trabajo se centrará en la familia de modelos ARIMA, fundamental en el estudio de series temporales y pilar básico para la mayoría del resto de modelos. A su vez, y por la naturaleza del problema, se indagará en las técnicas de automatización de estos modelos, elaborando un algoritmo propio que realice todo el estudio serie temporal.

Las técnicas basadas en el aprendizaje automático tratan de crear programas y rutinas capaces de generalizar comportamientos a partir de una información suministrada en forma de ejemplos experiencia pasada. Se crea así un programa que parte de una serie de modelos de aprendizaje automático (cada uno definido por sus respectivos parámetros) y mediante el aprendizaje computacional se optimiza cada modelo y se elige aquel que menor error genere en sus predicciones para unos determinados datos. El algoritmo creado permite que para cada elemento del mercado y cada momento se encuentre el modelo adecuado para realizar predicciones.

En conclusión, con este trabajo se obtienen diferentes modelos predictivos que presumiblemente detectarán el movimiento futuro del mercado: dos herramientas útiles para todo analista experto de mercados financieros.

Abstract

Predictive models help users to get ready against possible future events, providing forecasts about what may happen. In this sense, when we speak of predictive models for FOREX market, we refer to tools that try to forecast (with more or less precision) future moves of the elements of the market.

The FOREX market, or foreign exchange market, is considered as the most important financial market in the world. It manages millions of dollars every day. The main objective of this document is to create tools and models that forecast moves of the elements of the market. These predictions will be at a small scale, i.e. our interest is to predict moves in a short time period, therefore allowing that our predictions are not very affected by external economic decisions.

We will work with two types of predictive models: time series models and machine learning techniques based models.

Time series models are based on stochastic process theory. They assume that each time series is a sample of a stochastic process and, therefore, they create models for the underlying stochastic process that generates the time series. In this work we will focus on ARIMA models, one of the main time series models and the basis for the development of the rest of time series models. Due to the type of problem we work with, we will look for techniques that automatize these models too. Thus, we will create our own algorithm that will perform the whole study.

Machine learning techniques explore the construction and study of algorithms that can learn from data and make predictions on new cases. Such algorithms operate by building a model from example inputs in order to make data-driven predictions or decisions. On this way, we will create a program that, using computational learning, starts with a set of potential models (defined by some parameters) and selects the best model with the the most suitable parameters for the particular data. This complete and automatic algorithm allows making predictions for each element of the market, and for each moment also.

In conclusion, this work shows different predictive models that will detect (hopefully) future moves of the market, providing two useful tools for expert analyst for the financial markets

Índice general

Resumen	i
Abstract	ii
1. Introducción	1
1.1. Planteamiento y objetivos del trabajo	2
1.2. Contexto en el que se desarrolla el trabajo	3
1.2.1. La empresa A.I. Talentum	3
1.3. Organización del documento	3
2. El mercado FOREX	5
2.1. Conceptos básicos del mercado FOREX	5
2.1.1. Un poco de historia del FOREX	5
2.1.2. ¿Cómo funciona el mercado FOREX?	6
2.2. El análisis técnico de los mercados financieros	7
3. Modelos de series temporales	12
3.1. Series temporales. Las nociones de estacionariedad y volatilidad	13
3.2. La metodología Box-Jenkins. El modelo ARIMA(p,d,q)	18
3.2.1. Modelos ARMA(p,q). Las nociones de autorregresión y media-móvil	19
3.2.2. El modelo ARIMA(p,d,q).	24
3.2.3. La metodología Box-Jenkins	24
4. Automatización de los modelos de series temporales	33
4.1. Criterios de información como criterios de aceptación para modelos de series temporales	34
4.2. Algoritmo propuesto	36
4.3. Los programas TRAMO y SEATS. Otros métodos de automatizar los modelos de series temporales	40
4.4. Conclusiones y mejoras del algoritmo	44
5. Técnicas de aprendizaje automático	45
5.1. Aprendizaje automático	45
5.2. Técnicas de aprendizaje supervisado para la clasificación	48
5.2.1. El proceso de predicción usando herramientas de <i>machine learning</i> .	52
5.2.2. Algunos métodos de aprendizaje automático	53
5.3. Pruebas, resultados y conclusiones	56
6. Conclusiones	60
6.1. Aportaciones a la empresa	61
6.2. Futuras líneas de trabajo	62

Bibliografía**63**

Capítulo 1

Introducción

Disponer de una herramienta que proporcione información acerca de lo que ocurrirá en el futuro bajo determinadas condiciones ha sido siempre uno de los principales objetivos de la ciencia. Esta herramienta otorgaría información privilegiada a aquél que la tuviese y, por ende, una ventaja extraordinaria frente a sus competidores.

Por ello, tanto en el ámbito público como en el privado, se invierten grandes sumas de dinero en pos de encontrar modelos, métodos y algoritmos que predigan el futuro con cierta fiabilidad. Pero, ¿cómo se pueden construir estos modelos? Todo modelo destinado a la predicción ha de partir de unos supuestos. La única información de la que disponemos para diseñarlos es la que ya conocemos, es decir, el pasado y el presente. Luego todos estos modelos se rigen por una máxima: el futuro es un reflejo del pasado. De esta forma, un modelo para la predicción se nutre de información pasada y presente para plantear sus pronósticos. La tarea del investigador queda en analizar la información disponible para extraer de ella las relaciones causa-efecto entre pasado, presente y futuro.

En este proyecto se expondrán una batería de modelos y técnicas para el cálculo de predicciones dentro del mercado FOREX. El trabajo está planteado en busca de dos objetivos:

- El desarrollo y diseño de los modelos y técnicas empleados para el cálculo de pronósticos.
- La obtención de buenos resultados.

Por ello, el lector puede entender este documento como un texto de carácter didáctico, donde se exponen y explican las herramientas que se van a utilizar; o bien como un texto de corte divulgativo, donde se explica una aplicación de las herramientas presentadas a un problema real.

1.1. Planteamiento y objetivos del trabajo

El mercado FOREX es un mercado vivo y en continuo movimiento. Decir que nuestro objetivo es predecir el mercado es un término demasiado ambiguo. Vamos a especificar. Puesto que el problema de calcular pronósticos dentro de un mercado financiero no es un tema baladí, hemos de fijarnos unos objetivos realistas (aunque no por ello fáciles de alcanzar). En primer lugar, definamos nuestro problema:

Problema: Dada una serie de datos del mercado FOREX (por ejemplo, la cotización del euro frente al dolar) predecir si, en un determinado instante de tiempo, la serie de datos subirá o bajará.

Es decir:

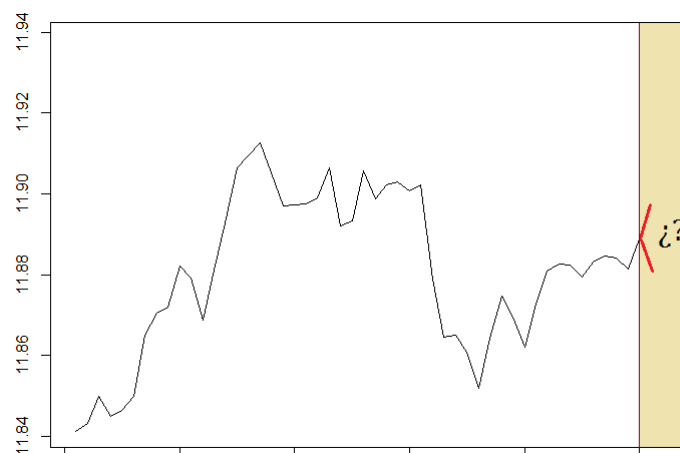


Figura 1.1: El problema. Fuente: elaboración propia.

No nos interesa saber cuánto subirá o bajará, éste es un problema a resolver una vez solventado el problema planteado. Lo que nos interesa es detectar los cambios de tendencia a pequeña escala, donde un pronóstico correcto puede aportar muchas ganancias.

Teniendo siempre presente nuestro problema principal, en este trabajo se presentan otros objetivos, entre los que destaca el estudio de los métodos usados, a saber: los modelos de series temporales (en particular, la familia de modelos ARIMA) y los modelos basados en el aprendizaje automático.

Esta dicotomía entre modelos caracterizará la estructura del trabajo, de la que hablaremos más tarde. Por último, para la elaboración de los modelos y algoritmos se usaron los programas R y Matlab. R presenta un entorno idóneo para la elaboración del código, dada la ingente cantidad de paquetes de que dispone. Por otro lado, Matlab sirve de gran ayuda para la recolección de datos y el tratamiento con las distintas bases de datos de la empresa, gracias al cálculo matricial del que hace gala y sus aplicaciones para bases de datos.

1.2. Contexto en el que se desarrolla el trabajo

Este proyecto se encuentra dentro de un conjunto de trabajos realizados por alumnos de la Universidad de Murcia para la empresa A.I. Talentum, entre los que podemos encontrar los trabajos “Análisis técnico de mercados financieros basado en técnicas de inteligencia artificial” de Ángel Pina (ver [13]) o “Electricidad, matemáticas e inteligencia” de Juan Manuel Espín (ver [4]).

El trabajo se realizó junto con el equipo de A.I. Talentum con un único objetivo: encontrar herramientas que proporcionen distintos pronósticos para el mercado FOREX y que sean de utilidad para expertos en este campo.

1.2.1. La empresa A.I. Talentum

La bolsa se considera un indicador adelantado de la economía. El mercado FOREX es una oportunidad de negocio, tanto para los pequeños inversores como para empresas especializadas en crear herramientas para estos. En el mercado hay que comprar expectativas y vender hechos, por lo que son necesarias herramientas que proporcionen ventajas al inversor.

Los sistemas de trading ¹ automatizados, traders automáticos o robots, permiten establecer una serie de reglas automáticas que dictan cuándo realizar una operación y cuándo cerrarla, de forma que ésta se ejecute de forma autónoma sin intervención humana posterior. Las reglas que en éstos se programan pueden ser relativamente sencillas, como las medias móviles, o mucho más complejas.

Entre las ventajas del trading automático se encuentra la rapidez a la que puede detectar y realizar las operaciones, la facilidad de diversificación o evitar que las emociones del broker influyan en las decisiones del sistema.

A.I. Talentum es una empresa dedicada, entre otras muchas cosas, a este campo. Fundada en Febrero de 2013, Talentum se dedica principalmente al estudio y desarrollo de sistemas automáticos para invertir en bolsa. Cuenta con profesionales de distintas ramas, principalmente economistas, informáticos, químicos y matemáticos, cuyas técnicas específicas combinadas se programan en los robots. En la actualidad, A.I. Talentum busca otras oportunidades de negocio, creando otro tipo de herramientas que sirvan a la tarea del analista experto, entre las que se encuentran las detalladas en este trabajo.

1.3. Organización del documento

Este documento se encuentra estructurado de la siguiente manera: el capítulo 2 trata sobre el mercado FOREX y sirve como introducción para todo aquel que no esté familiarizado con él. Los capítulos 3 y 4 se centran en los modelos de series temporales usados en

¹El trading financiero consiste en comprar o vender un valor subyacente en un mercado financiero con la intención de obtener un beneficio especulativo, intentando aprovechar pequeñas diferencias de precio para obtener una rentabilidad inmediata minimizando el riesgo que conlleve esta operación.

este trabajo para la predicción. El capítulo 3 asienta la base teórica de esta metodología, centrándose en la familia de modelos ARIMA, sin entrar en profundidad en los aspectos teóricos, con el fin de no distanciarse mucho del problema principal. El capítulo 4 se centra en la adecuación de estos modelos a nuestro problema, indagando en los métodos de automatización de los modelos de series temporales.

Por otra parte, el capítulo 5 recoge la otra gran familia de modelos usados: los métodos basados en el aprendizaje automático. Comienza introduciendo al lector en la teoría del aprendizaje automático, dando a conocer las nociones básicas y el proceso de aprendizaje llevado a cabo.

Además, al final de cada capítulo se exponen los resultados obtenidos con los métodos usados.

Por último, el capítulo 6 recoge las conclusiones del trabajo, así como las contribuciones realizadas a la empresa y futuras líneas de investigación.

Capítulo 2

El mercado FOREX

El mercado FOREX (abreviatura del inglés *Foreign Exchange*), también conocido como mercado de divisas, o simplemente FX, es el mercado global de compra-venta de divisas. Se trata del mercado que más movimiento de capitales registra a diario, más de mil millones de dólares, cantidad mayor que la suma de todos los demás mercados juntos. Decimos que es un mercado global, porque está disponible para los inversores de todo el mundo 24 horas al día y 5 días a la semana. Existen aproximadamente unas 4500 instituciones que participan activamente en el FOREX, y millones de inversores que lo hacen a través de estas instituciones.

Este capítulo se presenta como una introducción al contexto que engloba nuestro problema. El objetivo es exponer los conocimientos básicos relativos al mercado y las herramientas frecuentemente usadas para el estudio del mercado FOREX con el fin de que el lector entienda cada una de las futuras gráficas y ejemplos que se expongan. Cualquier lector familiarizado con este tema puede saltarse este capítulo e ir directamente al núcleo de este trabajo. Sin embargo, no deja de ser interesante su lectura para adaptarse al lenguaje usado en los ejemplos que vendrán posteriormente.

2.1. Conceptos básicos del mercado FOREX

2.1.1. Un poco de historia del FOREX

Para hablar de los comienzos del mercado FOREX, es necesario remontarse al año 1944, cuando tuvo lugar el convenio de Bretton-Woods, cuyo objetivo era dotar a las naciones vinculantes al pacto de una estabilidad monetaria que evitara la fuga de capitales entre países y la especulación ¹.

¹La especulación consiste en la compra-venta de bienes con vistas a su posterior reventa o recompra con el único motivo de esperar un cambio en los precios de dichos bienes (un aumento o disminución desmedida) y no la ganancia derivada de su uso.

Antes de la firma de este convenio, el valor de las monedas se fijaba en función de las reservas de oro de cada país, originando un sistema muy inestable. Claro es, cuando un país compraba mercaderías importándolas de fuera, perdía parte de sus reservas de oro, por lo que perdía masa monetaria y el precio de su moneda subía.

Era necesario encontrar un sistema con mayor estabilidad monetaria que permitiera a su vez una sostenibilidad del crecimiento económico y una suavización de los ciclos de crecimiento y decrecimiento. En el convenio de Bretton-Woods se fijó una tasa de cambio de todas las monedas respecto al dólar, y del dólar a su vez respecto al oro, (35\$ por onza de oro) y cada país vinculante se comprometió a mantener sus monedas en un estrecho margen respecto al dolar.

Pero estas condiciones no se cumplieron, puesto que en la década de los 50, una población con carencias notables tras la Segunda Guerra Mundial hizo que se diera un flujo enorme de capitales a nivel internacional que desestabilizó los tipos de cambio pactados en Bretton-Woods. Finalmente, en 1971 se abandonó el Convenio por ser imposible su cumplimiento. Las monedas comenzaron a fluctuar entonces libremente, basándose en la ley de la oferta y demanda. Aumentó notablemente el volumen de capitales en circulación, así como la velocidad y la volatilidad de las operaciones de cambio de divisas.

A partir de los años 80, la introducción de nuevas tecnologías empezó a propiciar la globalización del mercado de cambio de divisas, y su ininterrupción, provocando un mercado a tres bandas (Asia, Europa y América), abierto las 24 horas.

Con el auge de las nuevas tecnologías y la extensión de Internet, el cambio de siglo trajo consigo la transformación del mercado en un mercado global, continuo y accesible para todos los inversores. Este mercado global de compra-venta de divisas es lo que se conoce actualmente como FOREX (Foreing Exchange Currencies Market), o mercado internacional de cambio de divisas.

2.1.2. ¿Cómo funciona el mercado FOREX?

En el FOREX, todos los movimientos y las cotizaciones se expresan en pares de monedas: EUR/USD (el euro frente al dólar americano), GBP/JPY (la libra frente al yen japonés) y así. La explicación de estos pares es sencilla, por ejemplo: EUR/USD es el valor del euro en función del dolar (un euro vale x dolares). Aquí, la divisa base sería el dolar.

Una compra de una divisa implica la venta de otra simultáneamente. Para obtener beneficio, se compra una divisa con la expectativa de que el tipo de cambio con la divisa base suba. Cuando esto ocurre, hemos de realizar la operación de compra-venta inversa para obtener beneficio. Por ejemplo, compro 10 euros con 10 dolares cuando el tipo es 1/1, es decir, EUR/USD = 1,00. Sube el tipo de cambio a 1,2/1 (esto es, por un euro, obtengo 1,20 dolares, o lo que es lo mismo, EUR/USD = 1,20), lo que significa que por los 10 euros que tengo ahora me darían 12 dólares. Vendo mis euros y compro dólares, y como empecé con 10 y ahora dispongo de 12, he ganado 2 dólares en la operación. ¿Fácil, verdad?.

Ahora bien, si hay alguien que quiere vender, debe de haber otra persona que quiera comprar, y esto no siempre ocurre. En cualquier mercado existen empresas, instituciones y grupos de éstas que son quienes crean el mercado, es decir, quienes compran y venden según las leyes de oferta y demanda. En el mercado FOREX, los *market makers* son los bancos y las plataformas de comercio de divisas que actúan junto con los inversores particulares para crear el comercio de divisas.

El *market maker* es la contrapartida al inversor. Los *market makers* interactúan con sus clientes, independientemente de sus posiciones en el mercado, y con otros bancos en calidad de cliente, creando flujos de liquidez en ambos sentidos. Las ganancias provienen de la diferencia de los precios entre compra de una divisa y su venta (a esa diferencia se le llama *spread*). Es decir, las plataformas de inversión en FOREX ponen en contacto, mediante órdenes electrónicas, a inversores y bancos, haciendo que funcione la circulación de capitales que dan liquidez al mercado.

Pero, ¿cómo operan los *market makers*? Cuando un inversor da una orden de compra-venta de divisas, debe existir alguien que acepte esa oferta al mismo tiempo y al mismo precio. Aún tratándose de un mercado gigantesco, esto no siempre ocurre. Sin embargo, los inversores pueden ejercer órdenes en cualquier momento. Aquí es donde desarrollan su papel los *market makers*.

Los *market makers* están ofreciendo a cada segundo del día unos precios firmes de compra y venta, precios que son renovados cada 30 segundos según la oferta y la demanda. Entonces, el inversor puede aceptar esos precios incluso cuando no hay una parte contraria, porque esa oferta es asimilada por los *market makers*. Lo que hacen pues, es asegurar la continuidad del mercado al participar cuando es necesario.

El objetivo de todo inversor es, por tanto, adivinar el “movimiento” (crecimiento o decrecimiento) de un par de divisas para dar una orden de compra o venta maximizando el beneficio de la operación. El cómo predecir este movimiento es una historia aparte, y existen muchos procedimientos para intentarlo, algunos de los cuales se darán a continuación.

2.2. El análisis técnico de los mercados financieros

El objetivo de todo buen analista de mercados financieros es saber interpretar la situación del mercado, ya sea a través de noticias económicas que engloben al mercado (como puede ser una subida en el precio del petróleo) o estudiando el movimiento del mercado y siendo capaz de extraer patrones de conducta de él. En general, podemos englobar el análisis de los mercados financieros en dos grandes ramas: el análisis fundamental y el análisis técnico.

El *análisis fundamental* se realiza teniendo en cuenta la economía en un sentido amplio de la palabra. Es el análisis más utilizado por la gran mayoría de profesionales del mercado. Parte de la creencia de que el precio de los productos no es necesariamente representativo de su valor. El mercado Forex es un mercado de flujos que

se mueven en distintas direcciones dependiendo de quién sea el actor que lo lleva a cabo y que toma las decisiones de intervenir influenciado por la información de la que dispone, por criterios de eficiencia económica, por decisiones políticas o por criterios de oportunidad. Por lo tanto, se trata de averiguar cuáles son los datos de la economía que van a influir en la toma de decisiones de estos grandes actores.

El *análisis técnico* busca patrones y tendencias de comportamiento en la cotización de los productos financieros, basándose en la historia de comportamientos de éstos, y trata con ellas de predecir el movimiento futuro del mercado en busca de maximizar el beneficio obtenible.

El estudio del mercado que se hará en este trabajo será puramente técnico, por lo que indagaremos un poco más en qué consiste este tipo de análisis.

El análisis técnico de los mercados financieros en general, y del mercado Forex en particular, como ya se ha dicho, consiste en el estudio de los movimientos del mercado (principalmente desde un punto de vista gráfico) con el propósito de pronosticar las futuras tendencias de los precios. En general, este análisis parte de los siguientes supuestos, cuyo origen se encuentra en la *teoría de Dow* (ver [10]):

1. Los movimientos del mercado lo descuentan todo.
2. Los precios se mueven por tendencias.
3. El mercado tiene memoria.

(1.) La primera afirmación viene a decir que cualquier cosa que pueda afectar al precio de un objeto financiero se refleja realmente en su precio en el mercado, es decir, el precio del producto en el mercado refleja su valor real, por lo que uno es capaz de predecir el devenir del objeto únicamente observando el mercado. (2.) La segunda afirmación justifica el uso de gráficas en el análisis técnico del mercado. El único propósito de la representación gráfica de los movimientos de precios de un mercado es identificar tendencias en las primeras etapas de su desarrollo, con el fin de que las transacciones vayan en la dirección de dichas tendencias. Por último, (3.) “el mercado tiene memoria.”^{es} equivalente a la frase “la clave para comprender el futuro está en el estudio del pasado”. Es decir, refleja la esperanza de que el futuro será solamente una repetición del pasado, y por ello, un estudio de los datos actuales y del pasado serán la clave para el futuro.

Por otro lado, todo sea dicho, existe una amplia teoría que afirma que todas estas asunciones son falsas. ¿Es posible obtener beneficios dentro de un mercado financiero basándose únicamente en su comportamiento pasado? Según la teoría del camino aleatorio (*Random Walk Theory*) el movimiento de los distintos activos financieros del mercado, si los agentes que participan en éste son inteligentes y tienen acceso a la misma información, no sigue patrones ni tendencias. Por tanto, no se pueden usar los datos pasados para obtener beneficios o para predecir movimientos futuros.

De no ser así, si los precios no se movieran de forma caótica o difícilmente predecible, alguien podría ser capaz de modelar su evolución disponiendo de más información que los demás participantes. La ausencia de patrones o de tendencias es lo que haría que el mercado fuese “justo”.

Como es de esperar, asumiremos un enfoque del mercado más parecida al del análisis técnico, dejando la *Random Walk Theory* como una alternativa en caso de que el análisis técnico no proporcione resultados concluyentes.

Volviendo al análisis técnico, podemos agrupar todas las técnicas que usa en dos vertientes: el **análisis chartista** y el **análisis de osciladores**.

Sin entrar mucho en detalle, el *chartismo* es un sistema basado únicamente en el estudio de gráficos. Supone que los movimientos de los precios son debidos a una combinación de expectativas y sentimientos de los inversores, y que estos actuarán de la misma manera en un futuro, por lo que trata de descubrir patrones en las figuras que forman las evoluciones de los precios (ver figura 2.1).

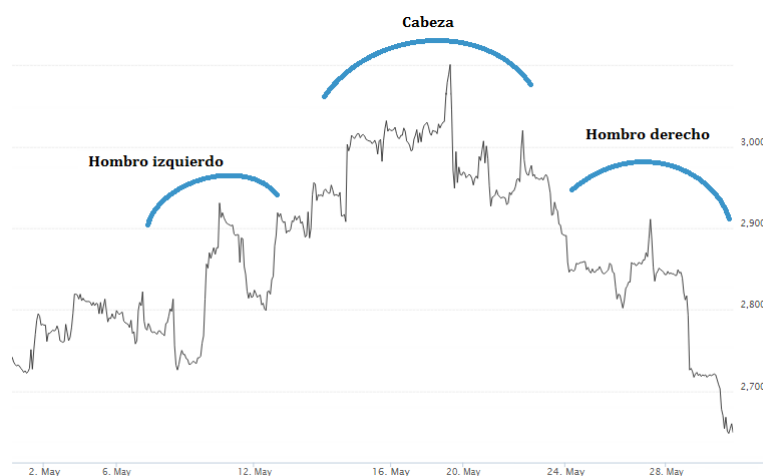


Figura 2.1: Formación hombro-cabeza-hombro: caracterizada por la presencia de un pico alto, el cual es seguido por un pico de mayor tamaño, el cual a su vez es seguido por otro pico cuya altura es menor a la de este pero aproximadamente igual a la del primero. Se suele utilizar como método para predecir un cambio en la tendencia de alcista a bajista. Fuente: elaboración propia.

Por otro lado, el *análisis de osciladores* utiliza herramientas matemáticas para determinar señales de compra o venta en los gráficos, utilizando como variable principal las cotizaciones de los precios. En el siguiente apartado podemos ver un ejemplo de análisis de osciladores.

Algunas técnicas del Análisis de Osciladores: la media-móvil

El ejemplo más sencillo de oscilador (y uno de los más recurridos) es la media móvil. La media móvil de un periodo es el promedio de los precios de un determinado título durante ese periodo. Matemáticamente, si queremos calcular la media móvil de periodo p de los datos X_t en el momento d , lo obtendremos con la siguiente fórmula:

$$M_d = \frac{\sum_{i=0}^{p-1} X_{d-i}}{p}$$

Al tomar las medias móviles de unos determinados datos, fijado el periodo de antemano, se consigue suavizar la curva de precios, permitiendo observar de una forma más sencilla las tendencias del mercado.

Además, resulta ser una forma muy intuitiva de obtener señales de compra y de venta. Estas señales se darán cuando la gráfica de la cotización del valor corta la gráfica formada por la media móvil, lo que será indicador de que el mercado está sufriendo una tendencia bajista o alcista a corto plazo. Por ejemplo, si una cotización ascendente corta la media móvil significa que el precio ha ascendido de forma más rápida de lo normal, por lo que podemos suponer que estamos ante una tendencia alcista, siendo conveniente comprar.

Un ejemplo más visual podemos encontrarlo en la figura 2.2, donde se puede observar el movimiento suave de la gráfica de la media móvil (en rojo) frente a la gráfica de los valores reales (en azul), además de los principales puntos donde se cortan ambas gráficas y cuyo significado es susceptible a interpretar.



Figura 2.2: Medias Móviles: Cotización del IBEX 35 tomado cada minuto entre las 09:01 y las 13:48 del 28 de Mayo de 2015. Con un 1 se marcan las principales señales de compra y con un 2 las señales de venta. Fuente: elaboración propia.

Podemos encontrar varias modificaciones que alteran la simplicidad de este método. Por una parte, tomar medias móviles de periodos más largos o más pequeños varía el riesgo y la rentabilidad esperada: Tomando medias móviles más largas las tendencias que detectemos serían más seguras pero de menor rentabilidad que con medias móviles más cortas. Por otra parte, hay múltiples mejoras a este método, que van desde el uso de varias medias móviles simultáneamente (para detectar por ejemplo los cortes entre ellas en lugar de con la cotización) hasta otras formas de calcular estas medias, dando por ejemplo más importancia a las cotizaciones más recientes.

En conclusión, existen una gran cantidad de herramientas que todo buen analista técnico de mercados financieros debería conocer, como son:

- Bandas de Bollinger.
- Velas japonesas.
- Gráficos de barras y gráficos de puntos.
- Índices Arms, índices Tick.
- Osciladores de McClellan.
- Índice de demanda (ID).
- ...

Para todo aquel interesado en invertir su tiempo y su dinero en el mercado FOREX de forma eficiente, una lectura muy valiosa de todas estas técnicas puede encontrarse en [10], donde se expone de forma rigurosa y detallada distintas herramientas del análisis técnico de mercados financieros.

Capítulo 3

Modelos de series temporales

En este capítulo se estudiará como construir un modelo que pueda explicar la estructura y prever la evolución de una variable que observamos a lo largo del tiempo, lo que se conoce como *serie temporal* o *serie de tiempo*. Dicha variable puede ser macroeconómica (como el índice de precios al consumo, demanda de electricidad, exportaciones, etc.), microeconómica (ventas de una empresa, consumo de electricidad en un hospital, existencias en un almacén), física (temperatura en un proceso, caudal de un río, concentración en la atmósfera de un agente contaminante) o social (número de nacimientos, matrimonios o votos a un partido).

Para realizar este estudio hemos de suponer que los datos se presentan en intervalos regulares de tiempo (horas, días, meses ...) y se desea utilizar la “inercia” en el comportamiento de la serie para prever su evolución futura. A este estilo de análisis se le llama *univariante*, debido a que únicamente usa la información de la propia historia de la serie de datos recogidas, basándose en la hipótesis central de que las condiciones futuras serán análogas a las pasadas. Este tipo de análisis es de gran utilidad para el cálculo de previsiones a corto plazo. Para obtener pronósticos fiables a medio y largo plazo hemos de tener en cuenta otras variables, relacionándolas con las que nos interesa de alguna forma. Por lo general, mediante un modelo de regresión dinámica (ver [11]).

Entre los objetivos que se persiguen con el estudio de modelos de series temporales se encuentra:

- Obtener una descripción concisa del fenómeno generador de la serie de datos.
- Construir un modelo que aproxime de la forma más fielmente posible el comportamiento de la serie de datos.
- Predecir valores desconocidos de la serie a partir de la información disponible, así como predecir valores futuros o pasados.
- Controlar el proceso que genera la serie, examinando qué puede ocurrir cuando se alteran algunos parámetros del modelo o estableciendo políticas de intervención

cuando el proceso se desvíe de un objetivo preestablecido más de una cantidad determinada.

En este capítulo se dará a conocer una de las familias más famosas de modelos univariantes de series de tiempo, los modelos ARIMA (acrónimo del inglés *autoregressive integrated moving average*, lo que en español sería modelo autorregresivo integrado de media móvil), desarrollado por George Box y Gwilym Jenkins a finales de los sesenta, quienes también sistematizaron su estudio. Con el fin de no alejarnos del objeto principal del trabajo, ilustraremos cada una de las explicaciones dadas en esta parte con aplicaciones prácticas en el mercado FOREX.

Comenzaremos el capítulo introduciéndonos en el marco teórico para el estudio de variables dinámicas, la teoría de procesos estocásticos. La primera parte dará a conocer los conceptos necesarios para el análisis de series temporales. En la segunda parte nos introduciremos en el estudio del modelo ARIMA, siguiendo la metodología de Box y Jenkins.

3.1. Series temporales. Las nociones de estacionariedad y volatilidad

Supongamos que disponemos de la cotización del par de divisas EURUSD tomada hora a hora desde el 8 hasta el 28 de mayo de 2015 (figura 3.1). Los datos aquí recogidos se encuentran equidistantes unos de otros y todos ellos representan una misma variable: la cotización del EURUSD en un instante de tiempo (una hora en este caso). Este tipo de colección de datos se conoce como *serie temporal*. Las series temporales no son más que una realización o muestra de un proceso estocástico. La definición que adoptaremos de proceso estocástico es la siguiente:

Definición 3.1.1. *Un proceso estocástico es un conjunto de variables aleatorias $\{z_t\}$, $t = 1, \dots, n$, ordenadas en el tiempo y que siguen una ley de distribución conjunta.*

En este contexto, una serie temporal es una muestra del vector n -dimensional $\{z_1, \dots, z_n\}$ que define el proceso estocástico (una sucesión de observaciones donde cada observación corresponde a una variable aleatoria distinta). Además, la ordenación de la sucesión de observaciones es esencial para el análisis de la misma. No se puede alterar, porque se cambiarían las características de la serie que se quiere estudiar.

La teoría de procesos estocásticos se basa en la teoría de la probabilidad, de forma que intenta ajustar una estructura probabilística a la muestra de datos que define al proceso. De esta forma se definen la *función de medias* y la *función de varianzas* para el proceso que define la serie:

$$\mu_t = E[z_t], \quad \sigma_t^2 = \text{Var}[z_t]. \quad (3.1)$$



Figura 3.1: Cotización del euro frente al dolar tomado hora a hora desde el 8 al 28 de mayo de 2015. Fuente: <http://www.elrincondeltrader.com/>.

Además, existen otras funciones que nos permiten medir la estructura de dependencia lineal entre las variables aleatorias del proceso, como la *función de autocovarianzas*:

$$\text{Cov}(t, t + j) = \text{Cov}[z_t, z_{t+j}] = E[(z_t - \mu_t)(z_{t+j} - \mu_{t+j})], \quad (3.2)$$

o la *función de autocorrelación*, que se presenta como la estandarización de la función de autocovarianzas:

$$\rho_{(t,t+j)} = \frac{\text{Cov}(t, t + j)}{\sigma_t \sigma_{t+j}}. \quad (3.3)$$

Cada una de estas funciones nos permite estudiar y clasificar cada serie temporal que vayamos a estudiar¹. Por ejemplo, un proceso puede ser estable en media y no en varianza, o al revés (es decir, alguna de las funciones de (3.1) son constantes). Por lo general, las series de datos a las que haremos frente no tendrán estas particularidades, y su estudio será más complicado.

Por tanto, cuando se nos presenten unos datos como los de la figura 3.1, éstos los interpretaremos como muestras tomadas de un proceso estocástico $\{z_t\}$, donde un valor en el instante de tiempo t_0 no será más que una muestra de la variable aleatoria z_{t_0} . Por eso, nuestro fin no es buscar un modelo que se ajuste bien a una serie temporal (lo que se podría hacer con un modelo de regresión), sino buscar un modelo “estocástico” que explique correctamente el proceso estocástico de donde procede nuestra serie.

¹En lo que sigue, se hará un abuso de notación y se tomarán las nociones de serie temporal y de proceso estocástico como sinónimos.

Procesos estacionarios

Una noción básica en el estudio de métodos basados en modelos de procesos estocásticos (metodología de Box-Jenkins) es la noción de estacionariedad, que se introduce a continuación:

Definición 3.1.2. Diremos que un proceso estocástico (o serie temporal) es estacionario si existen y son estables la media, la varianza y las covarianzas. Es decir, si para todo t :

- (1) $\mu_t = \mu = \text{cte.}$
- (2) $\sigma_t^2 = \sigma^2 = \text{cte.}$
- (3) $\text{Cov}(t, t+k) = \text{Cov}(t, t-k) = \gamma_k \quad k = 0, \pm 1, \pm 2, \dots$

Resulta que todo el desarrollo estadístico de la metodología de Box-Jenkins se realiza a partir de un proceso estocástico estacionario, en función del cual se puede determinar qué proceso estocástico ha sido capaz de generar la serie temporal bajo estudio con el fin de caracterizar el comportamiento de la serie y poder predecir su futuro. La filosofía que subyace en la teoría de la predicción es siempre la misma: se aprende de las regularidades del comportamiento pasado de la serie y se proyectan hacia el futuro. Por lo tanto, es preciso que los procesos estocásticos generadores de las series temporales tengan algún tipo de estabilidad.

La pregunta relevante ahora es si las series temporales económicas presentan, en general, un comportamiento estacionario. Una simple inspección visual de muchas series temporales económicas permite comprobar que presentan cambios de nivel, tendencias crecientes o estacionalidades, como ocurría en la figura 3.1. Se puede concluir, por lo tanto, que la mayoría de las series económicas no van a presentar comportamientos estacionarios, tanto en media, como en varianza.

¿Qué podemos hacer entonces para tratar series no estacionarias? Ante este problema, se tienen estandarizadas algunas soluciones, que consisten básicamente en transformar la serie en una estacionaria.

No estacionariedad en media

Ante la presencia de una componente tendencial (es decir, un comportamiento creciente o decreciente, exponencial o aproximadamente lineal) alcista o bajista (o alternando una y otra) en nuestra serie, un recurso muy utilizado es la **diferenciación**, es decir, aplicar diferencias consecutivas a la serie:

$$w_t^{(1)} = z_t - z_{t-1}$$

Si el resultado de esta operación no genera una serie estacionaria, se volvería a diferenciar, esta vez sobre la serie diferenciada $\{w_t^{(1)}\}$:

$$w_t^{(2)} = w_t^{(1)} - w_{t-1}^{(1)} = z_t - 2z_{t-1} + z_{t-2}$$

y así, sucesivamente. Se dirá que un proceso es *integrado de orden h* si al diferenciarlo h veces, se obtiene un proceso estacionario. No olvidemos esta noción, se volverá a usar más adelante.

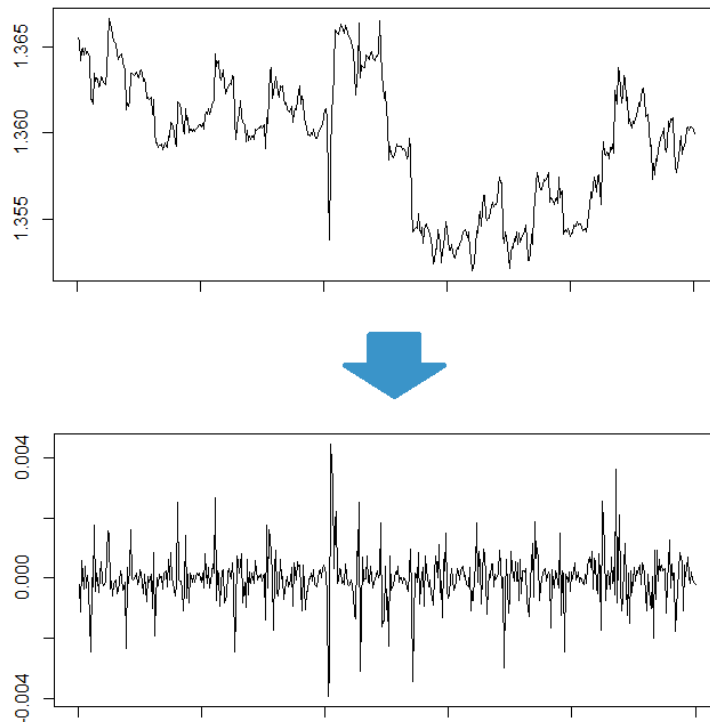


Figura 3.2: Efecto de una diferenciación sobre una serie no estacionaria en media. Se puede observar como la serie resultante tiene media constante igual a 0. Fuente: elaboración propia.

No estacionariedad en varianza

Cuando una serie no es estacionaria en varianza, es decir, no se puede sostener el supuesto de que ha sido generada por un proceso con varianza constante en el tiempo, la solución es transformar la serie mediante algún método que estabilice la varianza.

En una serie económica normal, lo general es que la varianza cambie conforme el nivel de la serie cambia. Ante esta situación, para estabilizar la varianza se utilizan las *transformaciones Box-Cox*²

$$z_t^{(\lambda)} = \begin{cases} \frac{z_t^\lambda - 1}{\lambda} & \lambda \neq 0, \\ \log(z_t) & \lambda = 0, \end{cases}$$

²¿Por qué estas transformaciones estabilizan la varianza? En [7], páginas 43 y 44, se expone una demostración de la validez de estas transformaciones para series cuya varianza depende de su media. Ver también http://dm.udc.es/assignaturas/estadistica2/sec4_5.html para más información.

donde λ es el parámetro de la transformación, cuya elección se realiza de entre un conjunto predeterminado de valores de los cuales se selecciona el de máxima verosimilitud. Es interesante señalar que, usualmente, las transformaciones Box-Cox no solo estabilizan la varianza sino que también mejoran la aproximación a la distribución normal del proceso.

Sin embargo, las series financieras no muestran el comportamiento particular del resto de series económicas, donde la varianza varía en función del nivel de la serie. De nuevo, si revisamos la figura 3.1, podemos ver como la varianza de los datos va alternando valores grandes y pequeños, ajena a los movimientos de la serie.

La noción de volatilidad

La suposición de varianza constante puede suponer diversos problemas estadísticos a la hora de estimar modelos. En general, se trata de problemas ligados a la eficiencia de los parámetros estimados y su fuerte inestabilidad ante el amplio intervalo de confianza en el que se mueven.

El fenómeno de varianza no constante en los datos es comunmente conocido como *volatilidad* de los datos. Es decir, la alta frecuencia de la serie ocasiona periodos de gran varianza frente a otros de poca varianza (esto se aprecia, por ejemplo, en la figura 3.1). Este término, aplicado al ajuste de modelos a series financieras, suele venir acompañado por las nociones de *homocedasticidad* (la varianza del error del modelo ajustado es constante) y *heterocedasticidad* (la varianza de las perturbaciones no es constante a lo largo de las observaciones).

Pero vayamos poco a poco. Los modelos basados en la metodología de Box-Jenkins no tienen presente el efecto de la volatilidad, lo que suele ocasionar fallos en el cálculo de predicciones. Existen muchos artículos que ponen de manifiesto la ineficacia del modelo ARIMA ante series que presentan un movimiento caótico (en el sentido liviano de la palabra).

Robert F. Engle, en 1982, propuso una familia de modelos que estudiaban el comportamiento estadístico de la varianza: los modelos ARCH (acrónimo del inglés *autoregressive conditional heteroskedasticity*). Justificó el uso de estos modelos citando varias situaciones donde el estudio de la varianza condicional (a los datos) puede explicar propiedades que no pueden ser explicados por los modelos ARIMA. En [14], se muestra el comportamiento dinámico de los activos financieros, realizando un estudio de éstos mediante modelos que estudian la varianza condicional de los mercados.

El lector se preguntará ahora por qué se realiza un estudio de los mercados financieros desde el punto de vista de la metodología de Box-Jenkins sabiendo de antemano que esta metodología suele dar malos resultados en este campo. Daré tres respuestas a esta pregunta, la primera teórica y las otras dos más prácticas:

1. La familia de modelos ARIMA son la base de estudio del resto de modelos, tanto ARCH, como GARCH (generalización de los modelos ARCH) o como cualquier

otro tipo de modelos basados en la teoría de procesos estocásticos. Por lo tanto, su estudio es fundamental y necesario, siendo el resto de modelos complementos de éste. En otras palabras, no empezaremos a construir la casa por el tejado.

2. En un principio, no se tiene pensado la aplicación de estos modelos directamente a los datos del mercado, sino que se aplicarán sobre la idea del *vector de fuerzas de la moneda*, una idea propia de la empresa A.I. Talentum, cuya correcta interpretación permite deducir el movimiento del mercado.
3. Por último, la aplicación del modelo ARIMA a los datos no se hará de forma directa, sino que los datos se preprocesarán de forma adecuada. Esto es, adecuamos los datos para el modelo ARIMA para después adecuar un modelo ARIMA a los datos.

3.2. La metodología Box-Jenkins. El modelo ARIMA(p,d,q)

La descomposición clásica de las series temporales es el pilar maestro de los modelos ARIMA. Éstos suponen que toda serie temporal se puede descomponer en una componente de tendencia, una componente estacional y un proceso estacionario.

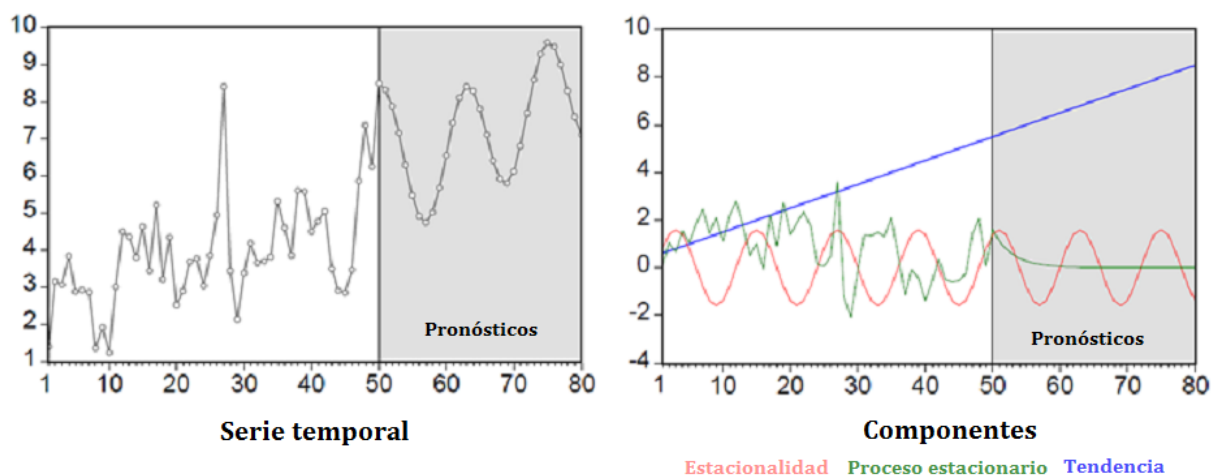


Figura 3.3: Descomposición clásica de las series temporales. Fuente: [9].

Por tanto, la idea de trasfondo en la identificación de un modelo ARIMA(p,d,q) para ajustar a una serie se basa en los siguientes puntos:

1. Si la serie no es estacionaria, transformarla hasta “convertirla” en estacionaria, haciendo uso de transformaciones Box-Cox para la no estacionariedad en varianza y diferencias para la no estacionariedad en media (eliminación de las componentes tendencial y estacional).

2. Una vez obtenido el proceso estacionario que define la serie, identificar el modelo que se ajuste mejor al proceso estacionario (identificación de los órdenes p y q).
3. Obtenido el modelo que se ajusta a la componente estacionaria de la serie, volvemos a incluir las componentes tendenciales (identificación del orden d) y estacionales.

Por último, el lector ha de saber que la componente estacional no se estudiará en este trabajo, tanto por motivos técnicos como empíricos. Podemos suponer que las series financieras no muestran comportamientos estacionales. Este hecho implicaría la existencia de patrones fácilmente reconocibles en el mercado, lo cual no ocurre en la realidad. Por otra parte, un estudio inicial de nuestras series (tanto de su comportamiento gráfico como un estudio de correlaciones) desvela la no existencia de estacionalidad en las mismas. Si el lector está interesado en modelos estacionales, le dirigimos a la lectura de [11], lectura casi obligatoria para todo aquel interesado en la teoría de series temporales.

3.2.1. Modelos ARMA(p,q). Las nociones de autorregresión y media-móvil

Comenzaremos estudiando modelos estacionarios (modelos para ajustar procesos estacionarios). Dentro de este conjunto, podemos destacar los *Modelos Autorregresivos* y los *Modelos de Media-Móvil*. Antes de comenzar con el estudio de estos modelos, hemos de introducir un tipo especial de procesos estacionarios.

Definición 3.2.1. *Un proceso estacionario $\{a_t\}$ que cumple las siguientes propiedades:*

- (1) $\mu_t = \mu = 0$
- (2) $\sigma_t^2 = \sigma^2$
- (3) $Cov(t, t+k) = 0 \quad k = 0, \pm 1, \pm 2, \dots$

*se denomina **proceso de ruido**. Si suponemos que todas las variables del proceso tienen distribución normal, se denomina **proceso de ruido blanco**.*

El proceso de ruido blanco simula el término de error en los modelos que vayamos a definir, dándole el toque estocástico a los modelos.

Modelos autorregresivos. AR(p)

La forma de dependencia más simple es relacionar la variable $\{z_t\}$ con $\{z_{t-k}\}$ ($k = 1, \dots, p$) linealmente mediante la ecuación de «autorregresión»:

$$z_t = \phi_0 + \phi_1 z_{t-1} + \dots + \phi_p z_{t-p} + a_t. \quad (3.4)$$

La idea de los modelos autorregresivos es explicar la variable de estudio en el momento t mediante combinaciones lineales de las observaciones de ella misma correspondientes a periodos anteriores (teniendo en cuenta un término de error a_t).

La ecuación (3.4) define un modelo autorregresivo de orden p (representado por las siglas AR(p)). Usualmente, se suele usar la notación abreviada:

$$\phi_p(L)z_t = \phi_0 + a_t, \quad (3.5)$$

donde $\phi_p(L)$ es el operador polinomial de retardos:

$$\phi_p(L) = 1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p,$$

y, a su vez, el termino L^k es el operador de k retardos, que aplicado a una variable en el momento t da como resultado el valor de esa variable en el momento $t - k$:

$$L z_t = z_{t-1} \quad \text{y} \quad L^k z_t = z_{t-k}.$$

Modelos de medias-móviles. MA(q)

Un modelo de medias móviles es aquel que explica el valor de una determinada variable en un período t en función de un término independiente y una sucesión de errores correspondientes a períodos precedentes (ponderados convenientemente). La ecuación que define el modelo de medias-móviles de orden q (representado por las siglas MA(q)) es la siguiente:

$$z_t = \theta_0 + a_t + \theta_1 a_{t-1} + \dots + \theta_q a_{t-q}, \quad (3.6)$$

que de nuevo se puede abreviar como en el caso de los modelos AR como sigue, siendo en este caso el operador polinomial de retardos $\theta_q(L)$:

$$z_t = \theta_q(L)a_t + \theta_0,$$

Así como un modelo autorregresivo es intuitivamente sencillo de comprender, la formulación de un modelo de medias móviles resulta no ser tan intuitiva. Uno puede preguntarse de dónde proceden esos errores o cuál es la justificación de un modelo de este tipo.

En realidad, un modelo de medias móviles puede obtenerse mediante un modelo autorregresivo, y viceversa. Este resultado es una consecuencia del *teorema general de descomposición de Wold* (ver [11]). Para ilustrarlo, únicamente hemos de pensar que, si tomamos la ecuación (3.4) de un modelo AR(1) (sin término constante) considerada en la etapa $t - 1$ en lugar de t , sustituyendo en la ecuación (3.4) obtenemos:

$$z_t = \phi z_{t-1} + a_t \quad \implies \quad z_t = \phi(\phi z_{t-2} + a_{t-1}) + a_t = a_t + \phi a_{t-1} + \phi^2 z_{t-2}.$$

Si repetimos la operación indefinidamente, lo que obtenemos es

$$z_t = a_t + \theta a_{t-1} + \theta^2 a_{t-2} + \dots + \theta^j a_{t-j} + \dots$$

obteniendo la expresión, sin término constante, de un modelo de medias móviles de orden infinito. Los procesos de medias móviles se suelen denominar procesos de memoria corta. Esto se debe a que, a diferencia de los modelos autorregresivos (que se denominan procesos de memoria larga), la acción de la perturbación a_t en el modelo MA(q) solo influye en z_t ,

z_{t+1} y así hasta z_{t+q} , es decir, solo permanece q periodos en el sistema. Este fenómeno de impulso finito es lo que motiva el uso de estos modelos.

En la teoría, no todos los procesos autorregresivos o de medias móviles son aptos para modelar un proceso estacionario, esto es, son estacionarios o invertibles³. Han de cumplir ciertas condiciones (condiciones impuestas a las raíces del operador polinomial que define el proceso). A aquellos interesados en los aspectos teóricos de estos modelos, les referenciamos a [11]⁴.

Modelos autorregresivos de media móvil. ARMA(p,q)

Como el lector habrá supuesto, los modelos ARMA(p,q) son la superposición de un modelo AR(p) y un modelo MA(q). Estos modelos determinan el valor de z_t en función de su pasado hasta el retardo p , de la innovación contemporánea y del pasado de la innovación hasta el retardo q . La fórmula explícita es:

$$z_t = c + \phi_1 z_{t-1} + \dots + \phi_p z_{t-p} + a_t + \theta_1 a_{t-1} + \dots + \theta_q a_{t-q}.$$

De nuevo, podemos simplificar esta fórmula haciendo uso del operador polinomial de retardos:

$$\phi_p(L)z_t = \theta_q(L)a_t. \quad (3.7)$$

Al igual que ocurría con los modelos MA y AR, el modelo ARMA ha de satisfacer ciertas condiciones para ser estacionario o invertible. De nuevo, no nos centraremos en la teoría (para ello, nuevamente se referencia al lector a [11]). Nos interesaremos más en los aspectos prácticos de estos modelos.

¿Como calcular el orden de los modelos?

Pese a conocer la idea tras los modelos antes mencionados, no es costumbre utilizar esta idea para modelar las series que se nos presenten. Es posible que sabiendo (por motivos externos a este estudio) que los datos tiene una fuerte dependencia lineal, y que cada dato se encuentra muy relacionado con los dos datos anteriores a él, uno pueda verse tentado a modelar la serie mediante un proceso autorregresivo AR(2), sin mayor estudio que ese. Ese no es el modo de actuar que seguiremos.

En la práctica, el cálculo del orden y el modelo a ajustar se lleva a cabo mediante la *función de autocorrelación simple* (FAS) y la *función de autocorrelación parcial* (FAP).

³Se dice que un proceso es *invertible* si el presente depende de forma convergente de su propio pasado, lo que implica que la influencia de z_{t-k} en z_t ha de ir disminuyendo conforme nos alejemos en el pasado. Desde un punto de vista matemático, un proceso es invertible si todas las raíces de la ecuación polinomial

$$1 - \theta_1 x - \theta_2 x^2 \dots - \theta_q x^q = 0,$$

están fuera del círculo unidad.

⁴En general, se tiene que todos los modelos MA son estacionarios y todos los modelos AR son invertibles. La generalidad no se da al revés.

En primer lugar, recuperamos la noción de función de autocorrelación ya introducida al principio de esta sección (ecuación (3.3)), adecuándola a procesos estacionarios.

Definición 3.2.2. Dado $\{z_t\}$ un proceso estacionario, se define la **función de autocorrelación** como

$$\varrho_j = \frac{\text{Cov}(t, t+j)}{\sigma_t \sigma_{t+j}} = \frac{\gamma_k}{\sigma^2},$$

donde σ^2 es la varianza del proceso.

La función de autocorrelación para un proceso estacionario presenta las siguientes propiedades fácilmente comprobables:

- $\varrho_0 = 1$.
- $\varrho_j \leq 1$.
- $\varrho_j = \varrho_{-j}$.
- $\lim_{k \rightarrow \infty} \varrho_k = 0$, dándose la convergencia de forma rápida.

Se define la función de autocorrelación simple (también conocida como *correlograma*) como la representación gráfica de la función de autocorrelación en función del retardo. Usualmente, esta representación gráfica se lleva a cabo mediante un gráfico de barras donde en el eje x se representan los retardos (ver figura 3.4).

Esta representación no es suficiente para determinar el orden de un proceso. Ilustremos este hecho: Si comparamos un AR(1) con un AR(2) vemos que aunque en ambos modelos cada observación está relacionada con las anteriores, el tipo de relación entre observaciones separadas dos periodos es distinto, como se puede ver a continuación (las flechas representan la relación directa):

$$\begin{array}{l} \text{AR(1): } z_{t-3} \rightarrow z_{t-2} \rightarrow z_{t-1} \rightarrow z_t \\ \text{AR(2): } z_{t-3} \rightarrow z_{t-2} \rightarrow z_{t-1} \rightarrow z_t \end{array}$$

Es decir, en el AR(1), el efecto de z_{t-2} sobre z_t es siempre a través de z_{t-1} y no existe efecto directo entre ambas. Sin embargo, en AR(2), además de esto, existe un efecto diferencial directo entre ambas. La función de autocorrelación simple tiene sólo en cuenta que z_t y z_{t-2} están relacionadas en ambos casos, pero si medimos la relación directa entre z_t y z_{t-2} (eliminando el efecto de z_{t-1}) encontraremos que para un AR(1) este efecto es nulo y para un AR(2) no.

Este fenómeno se puede observar mediante la función de autocorrelación parcial, que es una representación gráfica del *coeficiente de autocorrelación parcial* (cuyo cálculo se puede ver, por ejemplo, en [11], página 567) en función del retardo (ver figura 3.4).

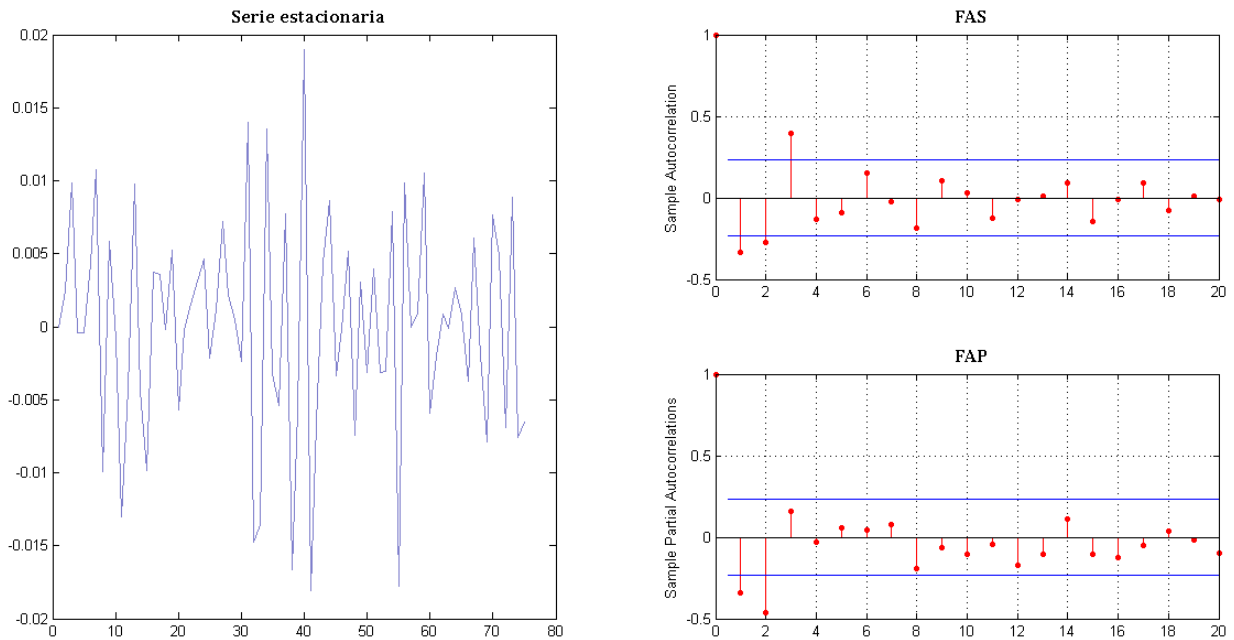


Figura 3.4: Función de autocorrelación simple (FAS) y función de autocorrelación parcial (FAP) de una serie estacionaria. Las bandas azules son conocidas como las *bandas de Bartlett*. Si la correlación está por debajo de dichas bandas, se supone estadísticamente 0. Imagen sacada con el programa Matlab. Fuente: elaboración propia.

La forma de actuar es la siguiente: en primer lugar, se calculan la FAS y la FAP de la serie estacionaria a estudiar. Una vez obtenidas, se comprueba con qué modelo se ajustan más estas gráficas.

Los modelos $AR(p)$, $MA(q)$ y $ARMA(p,q)$ se encuentran caracterizados según su FAS y su FAP como se expone en la tabla 3.1.

Como se puede ver, la caracterización de los modelos $AR(p)$ y $MA(q)$ no deja lugar a dudas, aunque en la práctica puede ser muy diferente. Sin embargo, la caracterización de los modelos $ARMA(p,q)$ es más difícil de visualizar. Por lo general, la FAS (FAP) de un modelo $ARMA$ presenta ciertos coeficientes iniciales que dependen del orden de la parte MA (AR) y después un decrecimiento dictado por la parte AR (MA). Esta estructura compleja hace que el orden de un modelo $ARMA$ sea difícil de identificar en la práctica.

Cuadro 3.1: Autocorrelaciones de modelos AR, MA y ARMA.

	<i>f.a. simple (FAS)</i>	<i>f.a. parcial (FAP)</i>
AR(p)	Muchos coeficientes no nulos que decrecen con el retardo como mezcla de exponenciales y senoides.	p primeros coeficientes no nulos y el resto cero.
MA(q)	q primeros coeficientes no nulos y el resto cero.	Muchos coeficientes no nulos que decrecen con el retardo como mezcla de exponenciales y senoides.
ARMA(p,q)	Decrecimiento hacia cero.	Decrecimiento hacia cero.

3.2.2. El modelo ARIMA(p,d,q).

La formulación general que presenta un modelo ARIMA de órdenes p , d y q (esto es, un modelo ARIMA(p,d,q)) es la siguiente:

$$\phi_p(L)(1-L)^d z_t = c + \theta_q(L)a_t. \quad (3.8)$$

donde el elemento $(1-L)^d$ recoge las d raíces unitarias del modelo (3.8). Sin su presencia, el modelo sería estacionario y presentaría una estructura similar a (3.7). En la literatura es usual usar la notación $\Delta^d = (1-L)^d$.

El orden d se denomina *orden de integración*, es decir, es el número de diferencias que hay que tomar de la serie para que sea estacionaria. Como ya se sabe, p es el orden del polinomio autorregresivo estacionario y q el orden del polinomio de medias móviles invertible.

3.2.3. La metodología Box-Jenkins

Box y Jenkins establecieron una metodología sistematizada para identificar el modelo de procesos estocásticos adecuado a ajustar a una serie de datos. Esta metodología se sigue empleando hoy en día, con ligeras variantes. Uno de los primeros objetivos que se tiene en este trabajo es comprobar si mediante un modelo ARIMA se puede modelizar y describir una serie financiera de las que se nos presentan.

En esta parte, se explicará paso a paso la metodología a seguir⁵, ilustrándola con su aplicación directa a la serie de tiempo generada por el *vector de fuerzas del euro*.

Los pasos a seguir se pueden ver en la figura 3.5.

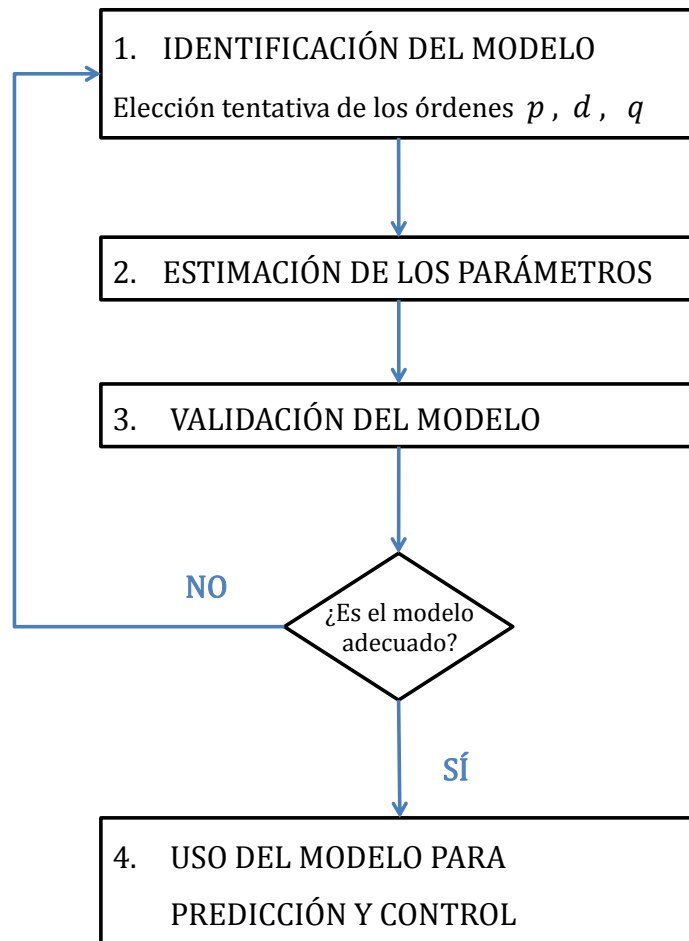


Figura 3.5: Metodología Box-Jenkins paso a paso. Fuente: elaboración propia.

⁵Todos los paquetes estadísticos llevan ya implementados los cálculos a realizar en cada una de las partes. Si el lector está interesado, estos cálculos se pueden ver en [11], capítulo 15, apartados 7, 8, 9 y 10.

1. Identificación del modelo

Supuesta la no existencia de volatilidad en los datos⁶, nuestro primer objetivo es seleccionar el modelo ARIMA(p,d,q) apropiado para la serie, es decir, que reproduce fielmente las características de la serie. La identificación del modelo se lleva a cabo en dos fases:

- a. Análisis de la estacionariedad, donde se determinan las transformaciones que son necesarias aplicar para obtener una serie estacionaria. Incluye dos apartados:
 - Estacionariedad en varianza: transformaciones estabilizadoras de varianza (principalmente, transformaciones Box-Cox).
 - Estacionariedad en media: número de diferencias d que hay que tomar para lograr que la serie sea estacionaria en media (no presente componente tendencial).
- b. Elección de los órdenes p y q . Una vez obtenida la serie estacionaria, el objetivo es determinar el proceso estacionario ARMA(p,q)⁷ que lo haya generado, tal y como se explicó en la sección 3.2.1.

Las transformaciones más frecuentes para estabilizar la varianza suelen ser transformaciones logarítmicas ($\lambda = 0$). La elección de un tipo de transformación u otra se hace, como ya se ha dicho, eligiendo el valor de λ que maximice la verosimilitud en la transformación Box-Cox. Muchos paquetes estadísticos llevan ya implementados el cálculo del valor de λ adecuado para los datos.

Seguiremos una forma particular para contrastar la necesidad de un tratamiento de la varianza. En primer lugar, extraemos la tendencia de la serie (diferencias). A esta serie transformada, le aplicamos el test ADF (test de Dickey-Fuller aumentado⁸), un test de raíz unitaria que contrasta si la serie es estacionaria o no. Si el test rechaza la hipótesis de estacionariedad, tenemos dos opciones:

1. volver a diferenciar (por lo general, dos diferencias es suficiente para eliminar la componente tendencial) y volver a aplicar el test ADF; o
2. asumir que es necesario un tratamiento de la varianza en la serie original, y por tanto, aplicar una transformación Box-Cox a la serie original.

Si el test acepta la hipótesis de estacionariedad, no es necesaria ningún tratamiento sobre la varianza. Para saber el número de diferencias necesarias para lograr que la serie sea estacionaria en media, se aplica sucesivamente el test de Dickey-Fuller sobre la serie diferenciada hasta aceptar estacionariedad.

Sobre la serie estacionaria, calculamos un modelo ARMA(p,q) que se ajuste bien a los datos. El cálculo del orden autorregresivo p y el orden de media móvil q se hace mediante un estudio de las funciones de autocorrelación simple y parcial, tal y como se vio anteriormente.

⁶La existencia de volatilidad, en sentido de varianza condicional, desvirtuaría cualquier estudio ARIMA que queramos ajustar a los datos, como ya se ha mencionado antes.

⁷Obviamente, ARMA($p,0$) = AR(p) y ARMA($0,q$) = MA(q).

⁸Para más información sobre este test estadístico, vea [7], páginas 54 - 60.

Lo usual en la parte de la identificación es obtener una gama de modelos candidatos, e ir descartando modelos una vez llegado a la parte de validación. Ante un empate, se suele recurrir a criterios de información como el AIC (sección 4.1).

Caso práctico: Deseamos ajustar un modelo ARIMA a la serie formada por la *fuerza del euro* durante el 6 al 17 de abril de 2015, tomada hora a hora (ver figura 3.6), y calcular los pronósticos de las 6 siguientes horas. Para ello, haremos uso de los paquetes *tseries* y *forecast* de R.

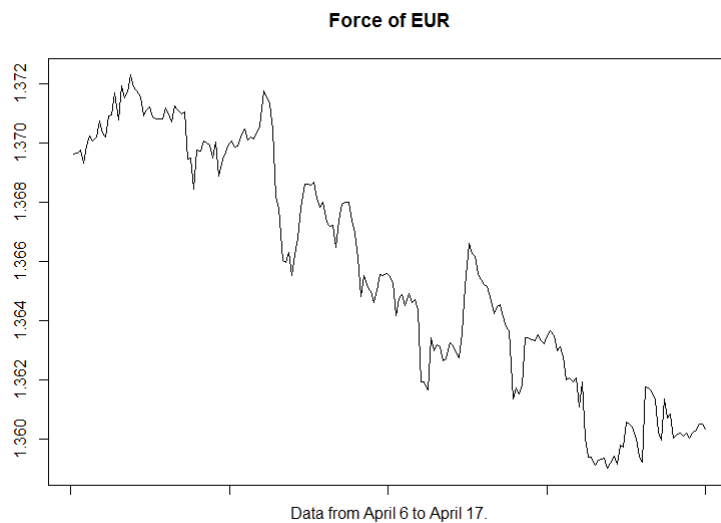


Figura 3.6: Serie de la fuerza del euro tomada hora a hora del 6 al 17 de abril de 2015. FUente: elaboración propia.

Si actuamos tal y como se dijo antes, al aplicar el test ADF a la primera diferencia de la serie (figura 3.7), obtenemos un $p\text{-valor} < 0,05$, lo que implica que la serie transformada es estacionaria (la hipótesis nula del test ADF supone la existencia de raíces unitarias, esto es, la no estacionariedad de la serie).

Transformada la serie en estacionaria, procedemos a visualizar su FAS y su FAP, para identificar los órdenes p y q (ver figura 3.8).

La estructura de la FAS y la FAP de esta serie no es fácilmente identificable, siendo la 2^a y la 4^a correlación las únicas distintas de 0. Esto sugiere que tanto $p > 0$ y $q > 0$. Probamos con un modelo ARMA(3,2), puesto que hay correlaciones mayores que 2 estadísticamente distintas de 0.

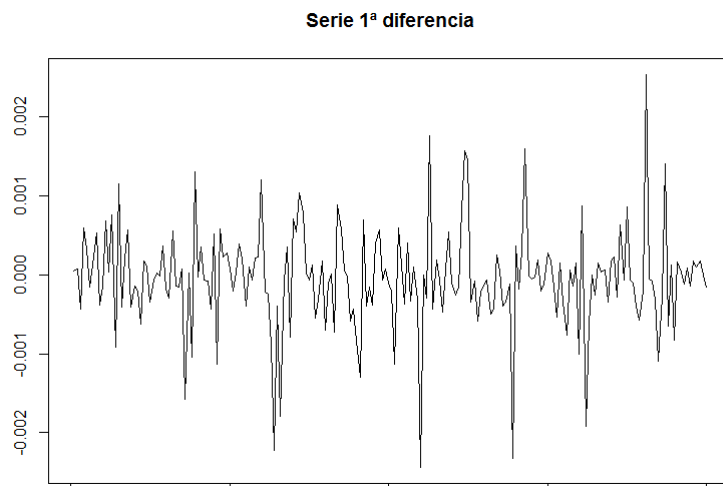


Figura 3.7: Serie 1ª diferencia de los datos. La estacionariedad de esta serie implica que el orden $d = 1$. Fuente: elaboración propia.

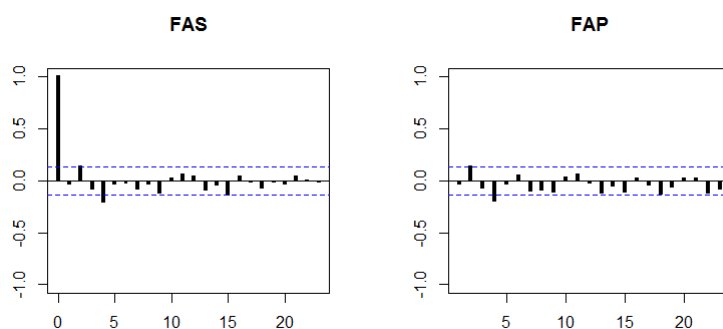


Figura 3.8: FAS y FAP de la serie 1 diferencia. Fuente: elaboración propia.

2. Estimación de los parámetros

La estimación de los parámetros del modelo (ϕ_i y θ_j en (3.8)) se puede hacer a través de diferentes métodos. Las formas más utilizadas son el método de Máxima Verosimilitud y Mínimos Cuadrados. Ambos métodos de estimación se basan en el cálculo de las innovaciones a_t a partir de los valores de la serie estacionaria. La innovación a_t se puede escribir como:

$$a_t = z_t - c - \phi_1 z_{t-1} - \dots - \phi_p z_{t-p} - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q},$$

despejando de la ecuación (3.7). Por lo tanto, para calcular las innovaciones a partir de un conjunto de información y de un vector de parámetros desconocidos, se necesitan un conjunto de valores iniciales z_0, z_1, \dots, z_{p-1} y a_0, a_1, \dots, a_{p-1} .

En la práctica, el procedimiento seguido es aproximar las innovaciones imponiendo una serie de condiciones sobre los valores iniciales, con lo que se obtienen los denominados

estimadores de *Mínimos Cuadrados Condicionados* y *Máximo Verosímiles Condicionados* (ver [7], páginas 69-72 para más información).

Caso práctico: Siguiendo con el caso práctico, el modelo candidato es el ARIMA(3,1,2) (se necesitó solo 1 diferencia). Mediante el método de Máxima Verosimilitud, usando el algoritmo de Hannan-Rissanen para el cálculo de los valores iniciales, se obtienen los coeficientes del modelo:

Parte AR	Parte MA
$\phi_1 = 0,4077141$	$\theta_1 = -0,4329399$
$\phi_2 = -0,6486289$	$\theta_2 = 0,8560216$
$\phi_3 = -0,1418090$	

3. Validación del modelo

La validación del modelo ARIMA se hace en función de los parámetros estimados y de los residuos del modelo:

- Los parámetros del modelo ARMA estimado han de cumplir las condiciones de estacionariedad e invertibilidad propias de modelos estacionarios.
- Los residuos del modelo han de ser ruido blanco.

Cualquier paquete estadístico suele llevar implementado la primera parte en el cálculo de los coeficientes del modelo, por lo que nos restringiremos al estudio de los residuos.

Caso práctico: En la figura 3.9 se puede ver la serie de los residuos del modelo junto con su FAS y su FAP.

Puesto que los residuos han de ser ruido blanco, se han de contrastar las condiciones de ruido blanco⁹:

- **Variables incorreladas.** Para ello, aplicamos el test de Ljung-Box a la serie de residuos, obteniéndose un *p-valor* de 0,97. En la figura 3.9 ya se podía apreciar que los valores estaban incorrelados.
- **Media estadísticamente igual a 0.** Haciendo el test de la T de Student para contrastar que la media de los datos sea igual a 0, obtenemos un *p-valor* de 0,389155, aceptándose la hipótesis de media nula.

⁹Cada uno de los test realizados durante el proceso de validación se pueden encontrar en cualquier texto de estadística, por ejemplo, en [11].

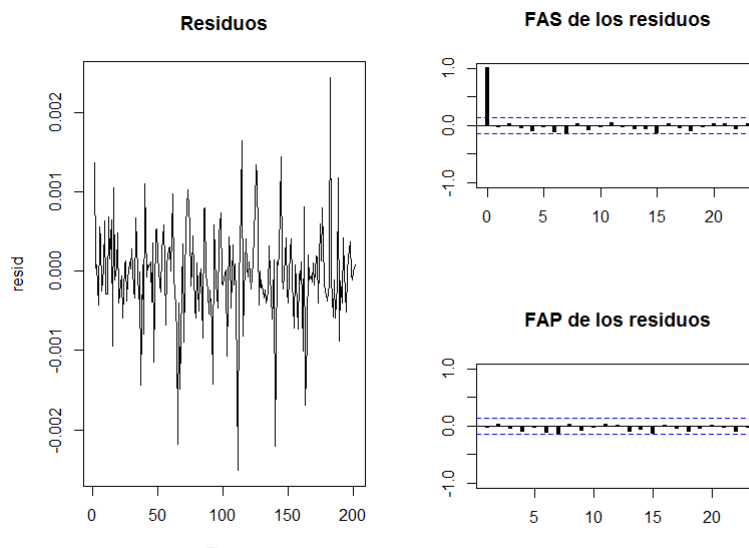


Figura 3.9: Residuos del modelo ARIMA(3,1,2) ajustado a los datos. Fuente: elaboración propia.

- **Varianza constante.** Nuevamente, podemos aplicar el test de Dickey-Fuller aumentado para contrastar si los datos son estacionarios, lo que directamente nos indicaría que la varianza es constante. El *p-valor* de este test es menor que 0,01, por lo que la varianza se supone constante.
- **Los datos siguen una distribución normal.** Hacemos el test de Shapiro-Wilk para contrastar la normalidad de los residuos. Obtenemos un *p-valor* menor que 0,05, por lo que concluimos que los residuos no siguen una distribución normal.

En conclusión, podemos decir que los residuos del modelo constituyen un proceso de ruido (no de ruido blanco). La no normalidad de los residuos puede ocasionar problemas a la hora del cálculo de los intervalos de confianza para los pronósticos. En consecuencia, se debería probar con el siguiente modelo candidato, pero en nuestro caso en particular, ninguno de los otros candidatos satisfacía el requisito de normalidad en los residuos.

4. Cálculo de pronósticos

En los pasos anteriores hemos obtenido el modelo y además hemos comprobado su idoneidad para poder predecir.

La predicción óptima del valor desconocido z_{n+k} , que denotaremos por \tilde{z}_{n+k} , es el valor esperado de z_{n+k} sabiendo que se conocen los valores z_1, z_2, \dots, z_n , es decir, la esperanza condicionada de z_{n+k} conocido z_1, z_2, \dots, z_n . De forma análoga se procede con los residuos (\tilde{a}_{n+k}):

$$\tilde{z}_{n+k} = E[z_{n+k} | z_n, \dots, z_1], \quad \tilde{a}_{n+k} = E[a_{n+k} | z_n, \dots, z_1].$$

Los pronósticos puntuales se calculan tomando estas esperanzas condicionadas en la ecuación en diferencias generada al despejar el valor de z_{n+k} en la ecuación que define el modelo (ecuación (3.7) o (3.8), según el modelo). Por otro lado, los intervalos de predicción se calculan asumiendo que los residuos del modelo siguen una distribución normal, y calculando el intervalo de confianza con media \tilde{z}_{n+k} , fijado el nivel de significación que se desea. Para una información más detallada de este proceso vea [11], capítulo 15.10.

Caso práctico: Los pronósticos obtenidos con nuestro modelo ARIMA(3,1,2) sobre la fuerza del euro durante el 6 al 17 de Abril de 2015 se pueden ver en la figura 3.10.

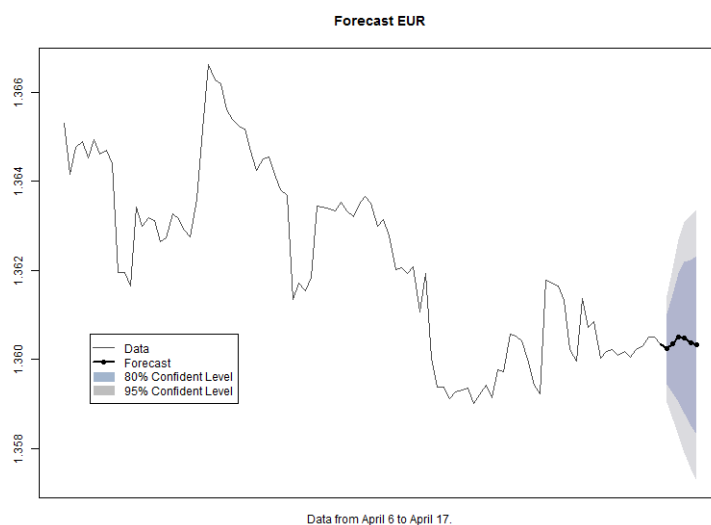


Figura 3.10: Pronósticos de las 6 horas siguientes del modelo ARIMA(3,1,2) e intervalos de confianza al 80 % y al 95 %. Fuente: elaboración propia.

La idea de la empresa tras el estudio ARIMA de los datos era comprobar si se podían extraer características relevantes que facilitarían un estudio posterior más sofisticado.

El hecho de obtener un modelo que refleja de forma más o menos precisa el comportamiento de la serie y que calcula unas predicciones consistentes, suponía un avance en el estudio que la empresa tenía en este campo. Las predicciones se pueden usar como un punto de apoyo a partir del cual tomar decisiones. Aun así, el considerar estas predicciones como definitivas podría conllevar a problemas no deseados. Hemos de tener en cuenta que para que estas predicciones sean fiables desde un punto de vista estadístico, los modelos han de satisfacer todas las hipótesis de validación, cosa que no en todos los casos se daba (como se vio en el caso práctico anterior).

Es necesario pues contrastar la eficacia del modelo para el tipo de estudio que se desea realizar. Para contrastar la efectividad de este estudio se necesita realizarlo sobre muchas

otras series, y en muchas otras situaciones. Este análisis supone un consumo de tiempo y esfuerzo considerables. Por ello, se decidió continuar de otra forma, siguiendo con la misma línea de estudio.

Capítulo 4

Automatización de los modelos de series temporales

El procedimiento Box-Jenkins, a diferencia de otros métodos clásicos de previsión, presenta la particularidad de no estar totalmente automatizado. Este hecho representa en muchos casos una ventaja, pues el trabajo de un identificador experimentado operando interactivamente con el proceso produce generalmente buenos resultados. Sin embargo, hay ocasiones en las que se desearía disponer de un modelador automático de series. Entre las posibles ventajas de esto, destacamos:

- Posibilidad de modelar gran cantidad de series en poco tiempo.
- Posibilidad de tratar series sin disponer de personal suficientemente instruido.
- Conseguir un mayor rigor en la elección de los modelos a estimar. Aun poseyendo una gran experiencia en este campo, es posible identificar modelos con una estructura muy parecida a la de la serie de estudio, pero con valores incompatibles en las autocorrelaciones. Al automatizar el proceso, el ordenador descartaría aquellos modelos incompatibles de forma directa, evitando pérdidas de tiempo innecesarias y elecciones inadecuadas de modelos.
- Mayor rapidez en el tratamiento de las series. Toda la modelización del capítulo anterior se reduciría a un solo paso.

En el capítulo anterior se vio la metodología usada en el estudio de series temporales desde el punto de vista de procesos estocásticos. Los resultados obtenidos proporcionaban un buen punto de partida en el intento de pronósticar el movimiento futuro del mercado FOREX.

Viendo que este tipo de modelización podía ser de ayuda en futuros estudios, se decidió proseguir con esta línea de investigación. Para ello, era necesario automatizar el

estudio realizado en el capítulo anterior. Conocedores de que existen ya herramientas que automatizan este estudio de forma más o menos eficiente, se propuso realizar una herramienta personal que automatizará la metodología Box-Jenkins, pero adecuándola al tipo de datos que se disponen.

En este capítulo se introduce un algoritmo para automatizar los modelos ARIMA siguiendo la metodología presentada por Box y Jenkins. Se comenzará introduciendo los criterios de información de Akaike (*AIC*) y bayesiano (*BIC*), usadas para la identificación automática del modelo ARIMA “óptimo” para la serie de datos introducida. Posteriormente, se expondrá las ideas básicas del algoritmo, así como el esqueleto del mismo, explicando de forma somera cada uno de los pasos implementados.

Por último, hablaremos de otros programas que automatizan el estudio de series temporales, haciendo énfasis en el programa TRAMO/SEATS.

4.1. Criterios de información como criterios de aceptación para modelos de series temporales

Algunos problemas tratados mediante secuencias de test de hipótesis podrían ser tratados más eficazmente mediante la aplicación de criterios de selección de modelos.

Los criterios de selección de modelos se basan, casi en su mayoría, en asignar una “puntuación” a los modelos en función del ajuste que tienen éstos a los datos. Esta “puntuación” se obtiene mediante cálculos que involucran la función de verosimilitud de los datos y el número de parámetros a estimar.

Estudiaremos dos criterios de información muy usados en el análisis de series temporales: el *criterio de información de Akaike (AIC)* y el *criterio de información bayesiano (BIC)*.

En el contexto de la elaboración de un modelo ARIMA a partir de una serie temporal $\{z_t\}$, los criterios de información *AIC* y *BIC* se emplean para decidir los órdenes p y q asociados a la transformación estacionaria de los datos $\{z_t\}$.

Hemos de tener en cuenta que estos criterios no deben usarse nunca para discriminar ni entre diferentes transformaciones alternativas de los datos $\{z_t\}$, ni entre modelos estimados con diferentes tamaños muestrales efectivos (que pueden variar en función del método de estimación empleado). Esto es, siempre se debe comparar el *AIC* y el *BIC* de entre modelos de la misma familia. Además, la selección de un modelo no debería basarse únicamente en el empleo de estos criterios, aunque pueden resultar de gran utilidad durante el proceso de identificación de modelos.

Un modelo es tanto más adecuado cuanto más pequeños son los valores de su *AIC* y *BIC*. Una consulta muy fructífera para aquellos interesados en este tema es [3], donde se explica la teoría de la información (en particular, los criterios basados en la información

estadística) aplicada en el análisis factorial. En este texto, nos centraremos en su estudio en el análisis de series de tiempo.

Criterio de información de Akaike, AIC

El criterio más sencillo basado en la información estadística es el criterio de información de Akaike, desarrollado por Hirotugu Akaike en 1974 e ideado en un principio para aplicarse al estudio de series temporales. Su idea clave es la de penalizar un exceso de parámetros ajustados, cosa que muchos test estadísticos, como el test asintótico de la chi-cuadrado, no tienen en cuenta.

El *AIC* (acrónimo del inglés *Akaike Information Criterion*) es un estimador muestral de $E[\log f(X|\theta)]$, es decir, la esperanza de la log-verosimilitud, que viene dado por la expresión general:

$$AIC(k) = -2 \log \mathcal{L}[\hat{\theta}(k)] + 2k, \quad (4.1)$$

donde $\mathcal{L}[\theta(k)]$ es a función de verosimilitud de las observaciones, $\hat{\theta}(k)$ es la estimación máximo verosimil del vector de parámetros θ ¹, y k es el número de parámetros independientes estimados dentro del modelo.

La idea básica subyacente al uso del *AIC* para la selección de modelos es la maximización de la log-verosimilitud esperada de un modelo determinado, mediante el método de máxima verosimilitud, que puede ser visto como la aplicación del principio de maximización de la entropía².

El objetivo de la selección de modelos mediante el *AIC* es estimar la pérdida de información cuando la función de probabilidad f , asociada al modelo verdadero (el que describe fielmente a los datos), es aproximada mediante la distribución de probabilidad g , asociada con el modelo que va a ser evaluado. Algunas de las ventajas del *AIC* que lo hacen tan utilizado en la práctica son su simplicidad y facilidad para ser implementado, junto con el notable hecho de que no existe el problema de especificar subjetivamente un nivel de significación arbitrario para contrastar los modelos.

Criterio de información bayesiano, BIC

Obsérvese que en la ecuación (4.1), el término de penalización del *AIC* es $2k$, y no depende del tamaño muestral n de la población considerada (el tamaño de la serie). Esto conduce al hecho de que un mismo número de parámetros comunes es seleccionado mediante el *AIC*, tanto para muestras pequeñas como para muestras grandes.

¹En el caso de los modelos ARMA, el vector θ está formado por los parámetros autorregresivos ϕ_i , con $i = 1, \dots, p$; y de media móvil θ_j , con $j = 1, \dots, q$.

²En el ámbito de la teoría de la información, la *entropía* (o entropía de la información) es una medida de la incertidumbre de una fuente de información.

En el contexto de procedimientos basados en la verosimilitud, Gideon E. Schwarz en 1978 sugirió que el *AIC* podría no ser asintóticamente justificable, y presentó un criterio de información alternativo a partir de un enfoque bayesiano, el *BIC* (*Bayesian Information Criterion*). Con este criterio, se busca incorporar la acción del tamaño muestral dentro de la penalización del criterio:

$$BIC(k) = -2 \log \mathcal{L}[\hat{\theta}(k)] + (\log n)k, \quad (4.2)$$

donde cada valor es como antes, a excepción de n , que representa el tamaño muestral de los datos.

Ambos criterios expresan un compromiso entre la obtención de un buen ajuste (a través del primer sumando de cada uno de ellos) y la obtención de un modelo con un número reducido de parámetros (a través del segundo sumando). En este sentido, puede comprobarse que el *BIC* siempre selecciona un modelo más escueto que el *AIC*, puesto que la penalización que establece al sobreajuste es mayor.

Otra notable diferencia entre estos dos criterios se explica por su distinto objetivo. Según diversos autores, el criterio *BIC* trata de seleccionar el modelo correcto y es un criterio consistente, de manera que la probabilidad de seleccionar el modelo correcto tiende a 1 cuando aumenta el tamaño muestral (ver [12]). El criterio *AIC* no pretende seleccionar el modelo correcto, ya que admite que el modelo verdadero puede no estar entre los estimados, y trata de obtener el modelo que proporcione mejores predicciones entre los existentes.

En definitiva, la elección entre un criterio u otro para el algoritmo se hará de forma puramente empírica, eligiendo aquel que dé mejores resultados una vez realizado un número considerable de pruebas. De partida, uno se presta más favorable al *AIC* que al *BIC*, ya que la suposición de que el modelo verdadero esté incluido dentro del conjunto de modelos candidatos quizás es demasiado disparatada, teniendo en cuenta que los datos a los que haremos frente son combinaciones de series financieras y que los modelos ARIMA no son, en principio, muy eficientes ante este tipo de datos.

4.2. Algoritmo propuesto

El algoritmo diseñado consta de 6 partes, las cuales se pueden agrupar en las 4 partes que constituyen la metodología de Box-Jenkins (ver figura 3.5). Cada parte del algoritmo se basa en la modelización de la metodología antes mencionada, y que se puede ver en el capítulo anterior. El esquema del algoritmo usado se puede ver en la figura 4.1.

Las partes de este algoritmo no son más que la automatización lógica de la metodología presentada en la sección 3.2.3. Las elecciones automáticas se hacen a partir de distintos test estadísticos, los cuales permiten ajustar el estudio a lo que haría un identificador experimentado.

La parte más novedosa con respecto a la metodología de Box-Jenkins se encuentra en el primer y tercer paso de este algoritmo. En el primer paso, se procesan los datos introducidos con el fin de encontrar el último conjunto de datos que no presente efecto ARCH ³, esto es, a grosso modo, que no haya una fuerte presencia de volatilidad en los datos.

El tercer paso, el cual consiste en la identificación de los órdenes p y q y la elección de los modelos candidatos, se realiza en función de la cantidad de información de cada modelo presentado (AIC o BIC), en vez de la visualización de las correlaciones de la serie, lo que es inviable automatizar.

En pocas palabras, se toman los datos a modelar y se procesan adecuadamente. Después, tras encontrar la transformación necesaria de los datos para hacerlos estacionarios, de entre un conjunto predeterminado de modelos, se escogen aquellos con menor AIC (o BIC). En la parte de validación, se ve si los residuos de estos modelos satisfacen los criterios de ruido blanco. Una vez hechas todas las comprobaciones, se escoge aquel modelo que más criterios satisface y menor AIC (o BIC) presenta y se usa para pronosticar.

Aplicaciones y resultados

La naturaleza del algoritmo se basa en el supuesto de que la mayor información que podamos obtener de los datos para pronosticar periodos cercanos se encuentra en los últimos datos recolectados. Es decir, suponemos que hay más relación entre datos cercanos que lejanos.

La validez del algoritmo a la hora de elegir el modelo ARIMA correcto se comprobó con series temporales clásicas, como la serie del IPC australiano medido desde 1972 hasta 1991, en la que el algoritmo acertó en la predicción del modelo ARIMA(2,1,0); o la clásica serie usada por Box y Jenkins del total de pasajeros de aerolíneas internacionales desde 1949 hasta 1960, donde el algoritmo predijo un modelo ARIMA(1,1,6) (frente al modelo real ARIMA(0,1,1)(0,1,1)). La elección de órdenes p y q altos en este tipo de series es un intento del algoritmo de suplir la falta de un estudio estacional.

Satisfechos con los resultados obtenidos, se procede a testear la validez del algoritmo frente a las series financieras a las que nos enfrentamos. Tal y como se encuentra definido nuestro problema, nos enfrentamos a un problema de clasificación, donde los valores predichos a contrastar son que el pronóstico sube o baja en función de los periodos anteriores.

Para testear los resultados, lo que haremos será modelar la serie $\{z_t\}$ hasta el instante T , y guardarnos los valores de la serie en instantes mayores que T . Una vez obtenidos los pronósticos, los contrastaremos con los valores guardados, con el fin de comprobar si el modelo supo predecir cuando la serie subía y cuando bajaba. Para medir el error de los

³El efecto ARCH es la suposición de que la varianza de los términos de error de la serie se pueden expresar en función de los términos de error de tiempos pasados, es decir, que la varianza del error se puede modelar mediante procesos AR.

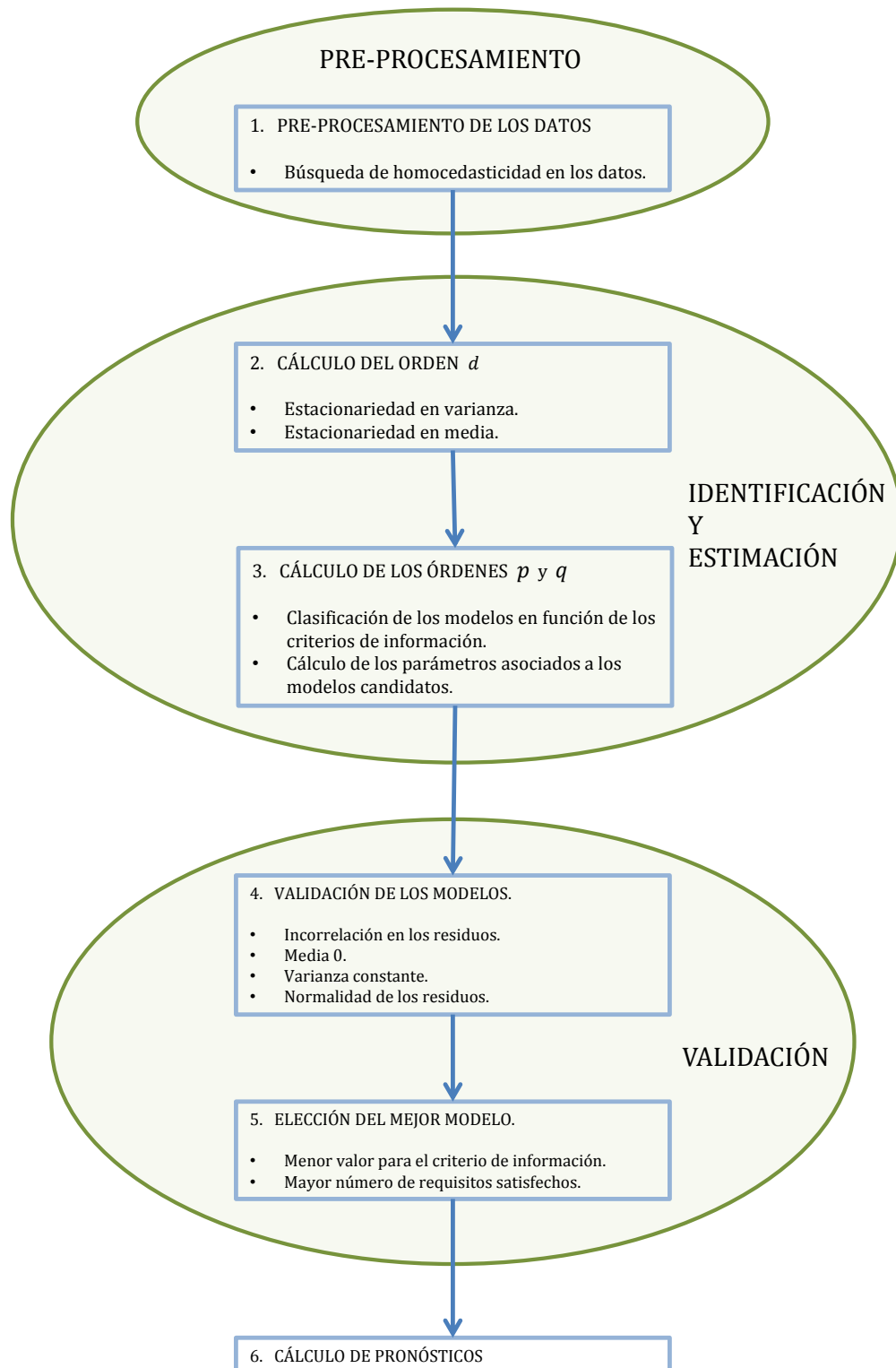


Figura 4.1: Esquema del algoritmo de automatización del análisis ARIMA de series temporales no estacionales. La implementación del algoritmo se llevó a cabo en R (con la ayuda del paquete *forecast*) y en Matlab.

pronósticos, haremos uso de matrices de confusión y nos basaremos en la precisión de los pronósticos (más adelante, en 5.2, se introducirán estas nociones).

¿AIC o BIC?

La primera prueba que se realiza es contrastar qué criterio es más eficiente para la elección de los modelos, el criterio *AIC* o el criterio *BIC*.

Las pruebas que se realizaron no son concluyentes con respecto a qué criterio elegir, ya que ambos proporcionan resultados prácticamente similares. El porcentaje de aciertos converge a valores entre el 42 % y el 59 % para el criterio *AIC*, y el 34 % y el 59 % para el *BIC*.

Además, estas pruebas se realizaron con series de todo tipo, tanto las formadas por vectores de fuerzas de las monedas, como con otras series financieras de las que maneja la empresa. Estas simples pruebas, junto a la teoría tras cada uno de los criterios, fueron suficientes para decantarse a favor del criterio *AIC* frente al *BIC*. Finalmente, otro detalle alentador es que el criterio *AIC* viene implementado en la gran mayoría de programas usados sin necesidad de paquetes o código auxiliar.

Resultados para los vectores de fuerzas

Pasamos ahora a analizar los resultados obtenidos con el algoritmo realizado sobre los vectores de fuerzas. En un principio supondremos que cada vector de fuerzas tiene una naturaleza diferente, por lo que los resultados obtenidos sobre esa serie no tiene por qué trasladarse al resto de series.

El análisis se realizará en tres vertientes. En primer lugar, estudiamos si el algoritmo es eficaz para predecir el movimiento en los 4 primeros pronósticos. Después, estudiamos la eficacia del algoritmo para predecir altas tendencias (tanto bajistas como alcistas) en el sentido de gran pendiente. Por último analizamos los intervalos de confianza para los pronósticos, comprobando si éstos engloban los resultados verdaderos. Los resultados a estos estudios se pueden encontrar en las siguientes tablas:

Precisión 4 primeros pronósticos

	AUD	CAD	CHF	EUR	GBP	JPY	NZD	USD
1° hora	51,8 %	51,5 %	53,2 %	48,3 %	52,5 %	53,7 %	51,7 %	49,9 %
2° hora	49,1 %	52,3 %	41,1 %	51,6 %	49,3 %	52,2 %	49,9 %	51,5 %
3° hora	47,1 %	49,2 %	48,4 %	48,6 %	48,7 %	49,8 %	51,2 %	47,1 %
4° hora	47,5 %	48,1 %	49,8 %	49,2 %	52,3 %	49,4 %	51,9 %	47,7 %

Podemos ver en la tabla anterior como los resultados, tras 1000 pruebas, convergen a un 50,0 % de acierto. Una conclusión a la que podemos llegar tras ver estos resultados es que los modelos ARIMA no son eficaces para captar la frecuencia de estos datos y por ello los resultados obtenidos no son fiables. También hay que preguntarse si ésta es la mejor forma de evaluar el algoritmo.

Viendo que el algoritmo no es bueno para captar la frecuencia de los datos, procedemos a evaluarlo de otra forma. En las siguientes tablas podemos ver otros estudios que se realizaron:

Exactitud de los pronósticos		Aciertos de intervalos	
	¿Capta tendencias altas?	Intervalo de confianza al 80 %	Intervalo de confianza al 95 %
AUD :	53,7 %	AUD :	74,8 %
CAD :	59,2 %	CAD :	78,6 %
CHF :	No concluyente	CHF :	77,6 %
EUR :	54,1 %	EUR :	78,3 %
GBP :	60,5 %	GBP :	77,0 %
JPY :	55,7 %	JPY :	77,6 %
NZD :	56,4 %	NZD :	77,6 %
USD :	65,9 %	USD :	76,6 %

La primera tabla busca averiguar si ante una fuerte tendencia (tanto alcista como bajista) el algoritmo es capaz de adivinarla. Se obtienen por lo general mejores resultados que en la tabla anterior, por lo que, con cierto nivel de confianza, podemos fiarnos del algoritmo cuando pronostique un fuerte cambio de tendencia. Los valores no concluyentes se deben a que la serie no reflejaba suficientes cambios bruscos de tendencia, por lo que no se obtuvieron suficientes resultados como para poder afirmar una solución.

La otra tabla obtenida refleja el porcentaje de aciertos a la hora de pronosticar los intervalos de confianza, esto es, las veces en las que los valores reales de los pronósticos de la hora siguiente se quedaban dentro de los intervalos pronosticados. Este es el mejor resultado obtenido, y por tanto, el principal resultado que uno debe tener en cuenta cuando visualice los pronósticos.

4.3. Los programas TRAMO y SEATS. Otros métodos de automatizar los modelos de series temporales

La automatización de modelos de series temporales es un tema en estudio desde el mismo nacimiento de estos modelos. La creación de programas que automaticen la labor del investigador es totalmente necesaria por razones obvias y ya mencionadas al principio de este capítulo. El principal uso que se buscaba con estos programas era la identificación correcta del modelo adecuado para una muestra de datos solo teniendo en cuenta dicha muestra. Entre el conjunto de programas destinados a tal fin cabe destacar los programas BV4, X11-ARIMA88, X11-UK, X12-ARIMA y X13-ARIMA-SEATS y TRAMO/SEATS. Nosotros nos centraremos en este último.

El programa TRAMO/SEATS fue desarrollado por Agustín Maravall y Victor Gomez en 1996⁴ a partir del programa de J.P. Burman para el Banco de Inglaterra. Las siglas de TRAMO proceden del inglés *Time series Regression with ARIMA, Missing observation and Outliers* y surge de la necesidad de reconocer y corregir los efectos de observaciones atípicas en las series de tiempo, que pueden estar siendo generados por errores en los datos, políticas empresariales o fenómenos de cualquier otro tipo, y que afectan al análisis econométrico de las series. Según los creadores del programa, TRAMO “puede ser visto como un programa que pule series ARIMA contaminadas” produciendo series que pueden ser vistas como la realización de un proceso estocástico, esto es, series originadas por un modelo ARIMA. A modo de ejemplo, podemos ver la acción de la “linealización” realizada por el programa TRAMO sobre la serie fuerza del euro en la figura 4.2.

El enfoque principal de TRAMO es combinar por un lado las facilidades de la detección y corrección automática de valores atípicos (*outliers*), y por otro, la identificación de un modelo SARIMA de forma eficiente.

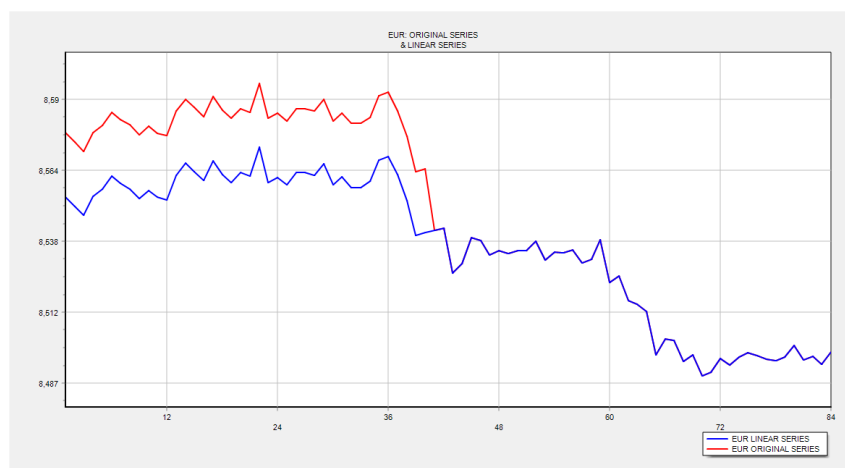


Figura 4.2: Linealización de la serie *fuerza del euro* realizada por TRAMO. Serie original en rojo. Serie “linealizada” en azul.

A su vez, el programa SEATS (del inglés *Signal Extraction in ARIMA Time Series*) sirve como complemento al programa TRAMO (aunque el uso conjunto de ambos no es totalmente necesario), sustituyendo la identificación del modelo realizado por TRAMO por una identificación más eficiente y rápida. SEATS recibe de TRAMO la serie pre-ajustada y la descompone en cuatro componentes: tendencial, cíclica, estacional e irregular. Esta descomposición permite obtener mayor información de la serie a la hora de tratarla y usarla para predecir datos futuros.

Los programas se pueden usar de manera separada, pero lo normal es usarlos conjuntamente de forma secuencial para realizar un análisis en profundidad de las series. El

⁴El procedimiento TRAMO (estudio y predicción de series temporales) y el procedimiento SEATS (obtención de la señal ciclo-tendencia de la serie temporal) se utilizan de forma oficial para el tratamiento de series temporales en multitud de instituciones como EUROSTAT, European Central Bank, Banco de España, INE, Ministerio de Economía y Hacienda, Agencia Tributaria, Banco de Santander Central Hispano, etc.

esquema general del tratamiento de una serie temporal llevado a cabo por el programa es el siguiente:

- TRAMO
 1. Tratamiento de datos no observados (por lo general, mediante interpolación).
 2. Tratamiento del efecto calendario y del efecto Pascua⁵.
 3. Tratamiento de outliers.
 4. Identificación de un modelo ARIMA inicial para la serie “linealizada”.
- SEATS
 5. Diagnóstico de los residuos ocasionados por el modelo inicial
 6. Descomposición de la serie en:
 - Componente de tendencia.
 - Componente estacional.
 - Componente cíclica.
 - Componente irregular.
 7. Obtención del modelo ARIMA final.
 8. Reincorporación de los efectos deterministas removidos por TRAMO.
 9. Obtención de pronósticos.

Todo este proceso se encuentra totalmente optimizado para su rápida ejecución (la ejecución de todo el programa dura poco más de unos segundos). En el manual de TRAMO/SEATS (ver [6]) se puede ver una explicación más detallada de todos estos pasos.

El objetivo ahora es usar esta herramienta para el problema que nos concierne. El lector se estará preguntando ahora por qué no se decidió aplicar desde un principio esta herramienta en vez de desarrollar un algoritmo propio para automatizar el estudio. En primer lugar, hay que apelar al encanto de disponer de un algoritmo propio, del cual se conoce cada uno de los tejemanajes que contiene por el simple hecho de ser su autor. Además, el crear nuestro propio algoritmo supuso adquirir los conocimientos necesarios para ello. En segundo lugar, los programas TRAMO y SEATS son programas independientes escritos en *Fortran* y, pese a que existen versiones de estos para **R** o **Matlab**, su uso no es nada sencillo y reflejan ciertos “errores” de compilación. Por último, hemos de tener en cuenta que los programas TRAMO y SEATS se idearon para series económicas y sociales y su adecuación (automática) para series financieras no es un tema baladí. Además, pese a que estos programas se distribuyen de forma gratuita por el banco de España, no se catalogan como software libre debido a las restricciones de acceso a su código.

⁵El efecto calendario y el efecto Pascua hacen referencia a la aparición continua de observaciones anómalas en los datos, debidas principalmente al periodo temporal en el que se toman. Simulan periodos vacacionales o situaciones que se espera que ocasionen valores anómalos.

Pese a que la importancia que se le da a los programas TRAMO y SEATS en este trabajo no es suficiente, se decidió hacer algunas pruebas con él para ver su utilidad ante series financieras. La idea es (tal y como se hizo anteriormente con el modelo ARIMA básico) corroborar su eficacia ante este tipo de series y en función de los resultados, decidir si investigar en esta línea o no.

A continuación se exponen algunos ejemplos de los resultados obtenidos con este programa⁶. En algunos casos, el programa TRAMO/SEATS captaba, en mayor o menor medida, el movimiento de la serie y conseguía plasmarlo en los pronósticos, llegando incluso a acertar con cierta precisión (ver figura 4.3).

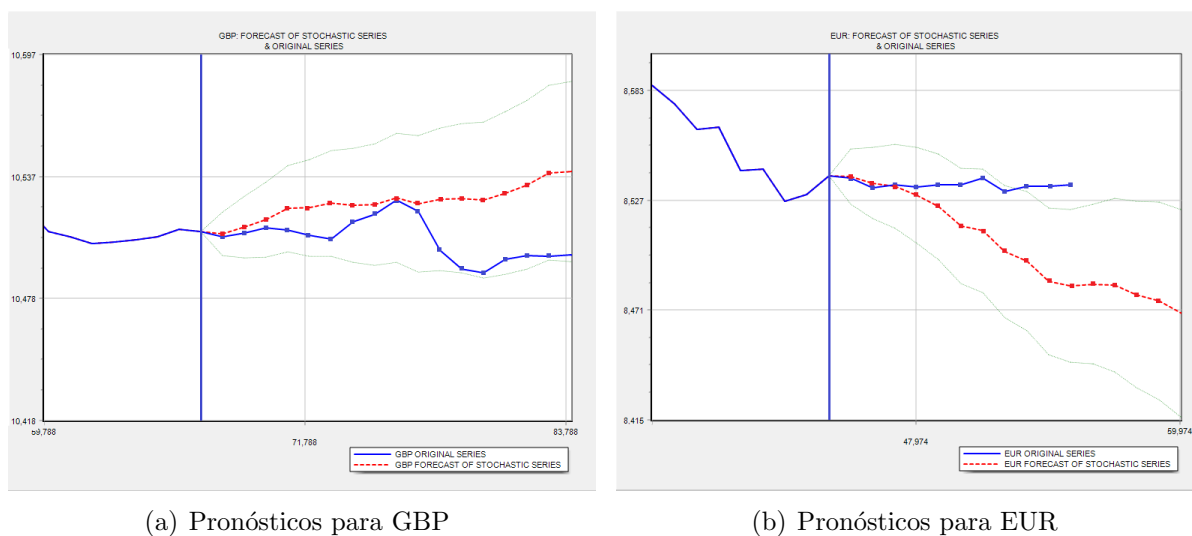


Figura 4.3: Pronósticos con el programa TRAMO/SEATS. Serie original en azul. Pronósticos en rojo.

Sin embargo, en la gran mayoría de casos, el programa TRAMO/SEATS daba resultados insatisfactorios, lo que sugería que era necesario un estudio previo de la serie para adoptar los parámetros y requisitos necesarios para aplicar al programa.

Hemos de considerar que el estudio de este programa se ha hecho desde un punto de vista meramente gráfico, evitando entrar en detalle. Por otro lado, se decidió no alterar el funcionamiento de TRAMO ni de SEATS, dejando los parámetros que venían por defecto en ambos programas y haciendo que su uso fuera totalmente automático. Además, hemos de tener en cuenta que es la primera vez que hablamos de “acertar” en los pronósticos en este texto, por lo que TRAMO/SEATS podría considerarse una futura línea de estudio más que interesante para la empresa.

⁶El programa utilizado es el programa TSW, la versión para Windows de TRAMO/SEATS, la cual se puede descargar de forma gratuita desde el siguiente enlace: http://www.bde.es/bde/es/secciones/servicios/Profesionales/Programas_estadi/Programas.html.

4.4. Conclusiones y mejoras del algoritmo

En este capítulo se ha visto una primera aproximación a los métodos predictivos para el mercado FOREX. El modelo desarrollado, basado en modelos de procesos estocásticos desde el punto de vista de Box y Jenkins (sin tratamiento del factor estacional) no refleja resultados satisfactorios, por lo que se ha de tener cierta precaución en su uso. Sin embargo, se puede utilizar como una primera aproximación, y su uso conjunto con otros modelos puede dar buenos resultados.

Es por ello que este modelo se encuentra implantado en la página web de la empresa A.I. Talentum (<http://www.elrincondeltrader.com/>) junto con muchos otros, donde se utiliza para dar una primera aproximación (visual) a los pronósticos de cualquier serie usada por la empresa, a partir de la cual cualquier analista del mercado puede tomar decisiones e invertir en función de las conclusiones a las que llegue.

No obstante, el principal resultado es la apertura de una nueva línea de investigación para la empresa A.I. Talentum, la cual, recordemos, se dedicaba al estudio de este problema únicamente mediante modelos computacionales (los ya mencionados Traders Automáticos). El enfoque dado por el estudio de series temporales sobre el problema de predecir el futuro del mercado FOREX puede originar resultados sorprendentes, y no se debe dejar de lado.

Queda pendiente, por tanto, el desarrollo de las siguientes líneas de investigación:

- Mejoras del algoritmo incluyendo un pre-procesamiento más sofisticado de los datos (al estilo de TRAMO), así como implementando el efecto estacional.
- El uso de modelos heterocedásticos tales como los modelos ARCH o GARCH.
- La automatización de programas tales como el TRAMO/SEATS, así como la investigación de otros programas (como por ejemplo, el X13-ARIMA-SEATS) para contrastar su eficacia en los mercados financieros.

Por ahora, y en lo que concierne a este trabajo, dejaremos este estudio abierto en este punto, a la espera de poder retomarlo en un futuro, y nos centraremos en otro tipo de modelos predictivos que se verán en el siguiente capítulo.

Capítulo 5

Técnicas de aprendizaje automático

Los métodos usados para la predicción y clasificación son muchos y muy dispares. Engloban una rama muy polivalente de la ciencia. El análisis de series temporales es solo un método más para tal propósito, y como cabe de esperar, no siempre es aplicable a un determinado problema, y no siempre es el que da mejores resultados.

Convencionalmente, se suelen usar herramientas y técnicas computacionalmente más “elaboradas” (aunque quizás sin un trasfondo teórico tan consistente que las avale) para el problema del cálculo de pronósticos. En este capítulo veremos otro tipo de herramientas usadas para la predicción y clasificación. No hablaremos de algoritmos con búsqueda exhaustiva o no informada que resuelven un problema analizando todas las posibilidades, ya que esta metodología es inviable en nuestro problema. No obstante, para tener una búsqueda informada se necesita conocimiento experto sobre el problema, algo de lo que se carece.

Tal y como está definido el problema al que nos enfrentamos, nos centraremos en modelos que “compensan la carencia de conocimiento que se tiene del problema con una gran cantidad de datos” -Ethem Alpaydin: los modelos basados en el *aprendizaje automático*.

5.1. Aprendizaje automático

Supongamos que pudiésemos enseñar al ordenador cómo detectar subidas o bajadas en los precios del mercado. O lo que es más, imaginemos que conseguimos hacer que el ordenador sea autosuficiente y consiga por sí solo identificar patrones en los datos que conducen a un movimiento determinado en los precios, ya sea de subida o de bajada. No sabemos como se comportará el mercado ante determinadas situaciones, pero sí sabemos que ante ciertos “estímulos”, el mercado reacciona. Hay una relación causa-efecto implícita en el mercado. Es decir, creemos que existe un proceso que explica los datos que observamos y confiamos en que este proceso no es completamente aleatorio. Cuando

hay una subida en el precio del petróleo, todos los elementos financieros del mercado se resienten (para bien o para mal); o cuando una divisa presenta un crecimiento continuado, es previsible un decrecimiento brusco en un futuro inmediato.

Existen datos que pueden explicar el comportamiento del mercado, y datos del mismo mercado que reflejan patrones que se repiten históricamente, por lo que a lo único que podemos aspirar es a recolectar estos datos y esperar extraer información suficiente que responda a nuestras preguntas.

Cierto es que no podemos identificar el proceso completo que define los datos, pero creemos poder construir una buena aproximación a éste: confiamos en poder identificar esas regularidades y relaciones. Ésta es la base del *aprendizaje automático*, o más conocido por su nombre en inglés: *machine learning*. La identificación de este comportamiento reiterativo nos ayudará a entender el proceso y a hacer predicciones.

Cuando se aplican los métodos de aprendizaje automático a grandes bases de datos, el proceso se conoce como minería de datos. En este campo, se procesan grandes volúmenes de datos para construir modelos simples que puedan servir, por ejemplo, para obtener predicciones precisas.

Sin embargo, el aprendizaje automático no es solo un problema de base de datos, es parte de la inteligencia artificial. Para ser inteligente, un sistema que está continuamente cambiando de entorno ha de poseer la habilidad de aprender.

Pero, ¿qué es exactamente el aprendizaje automático? Como bien puede intuir el lector, consiste en hacer que una máquina aprenda. Formalmente hablando, trata de crear programas y rutinas capaces de generalizar comportamientos a partir de una información suministrada en forma de ejemplos o experiencia pasada. Partimos de un modelo definido por unos parámetros y mediante el “aprendizaje” (la observación y análisis de ejemplos y experiencia pasada) optimizamos los parámetros del modelo, para luego usarlo en predicción o descripción o ambas.

Por tanto, partimos de unos datos (ejemplos o experiencia pasada), tomamos el modelo que queremos usar, ajustamos los parámetros de este modelo a los datos (de forma que el modelo resultante sea explicativo para los datos) y una vez encontrados los parámetros óptimos del modelo (para estos datos), lo usamos para nuestro propósito. El proceso de aprendizaje que usamos se explicará más adelante (sección 5.2).

Para poder aprender, los datos tienen que estar estructurados. En general, se consideran dos tipos de datos:

- Inputs, también conocidos como datos de entrada o variables predictoras, conforman el conjunto de elementos que el ordenador recibe y analiza.
- Outputs, también conocidos como datos de salida o variables respuesta, son el resultado obtenido por los datos de entrada.

No todos los procesos de aprendizaje son iguales. El aprendizaje depende de los datos que queramos aprender. En función de esto, podemos distinguir tres tipos de aprendizaje automático: aprendizaje supervisado, no supervisado y por refuerzo.

Aprendizaje supervisado: parte de la división clásica de los datos en inputs (X) y outputs (Y). En este tipo de aprendizaje se desarrollan modelos que definan correctamente la relación entre input y output, es decir, que asocien correctamente cada dato de entrada con su respectiva salida. El objetivo es predecir el valor de la variable respuesta basándose en numerosas variables de entrada, de forma que la variable a predecir “supervisa” el proceso de aprendizaje. Formalmente, se buscan los parámetros, θ , del modelo, $g(\cdot)$, que describa la relación entre input y output:

$$Y = g(X | \theta)$$

De forma esquemática, se suelen representar como:



Figura 5.1: Esquema del proceso de aprendizaje supervisado. El cómo aprende el modelo (cómo establece relaciones entre input y output) es algo específico de cada modelo. Fuente: elaboración propia.

Los problemas de aprendizaje supervisado se pueden dividir a su vez en:

- Problemas de clasificación, donde la salida del problema se divide en clases y se busca asociar cada input con su respectiva clase. Problemas de clasificación pueden ser, por ejemplo, clasificar si un correo nuevo es spam o no (output) en función de su contenido (input); decidir si un paciente tiene una enfermedad (output) según sus síntomas e historial clínico (input); o clasificar el riesgo de hacer un crédito a un cliente (output) en función de su capacidad financiera (input).
- Problemas de regresión, donde la salida se representa mediante un valor numérico calculado a partir del input introducido. Problemas de regresión pueden ser la predicción del precio de una casa (output) según sus atributos (input); o la creación de coches autónomos, donde la salida sería el ángulo de giro del volante en cada instante y la entrada los datos tomados por distintos sensores dispuestos por el coche.

Aprendizaje no supervisado. En el aprendizaje supervisado, el objetivo es aprender cómo relacionar inputs con outputs, donde las relaciones correctas vienen dadas por un supervisor. En el aprendizaje no supervisado, no existe la figura del supervisor y solo se dispone de los datos de entrada. El objetivo es encontrar regularidades y patrones en los datos que nos permitan asociar unos con otros. Suponemos que existe una estructura en el espacio de datos de entrada donde ciertas conductas son más frecuentes que otras, y nos interesamos en observar lo que generalmente ocurre y lo que no (lo que en estadística se conoce como estimación de la densidad).

Un método muy utilizado en el aprendizaje no supervisado es el *clustering* (ver [5]). La agrupación de datos es útil para, por ejemplo, compañías interesadas en establecer conjuntos de consumidores con el fin de ofrecer un producto que interese a la mayoría.

Aprendizaje por refuerzo. ¿Y si la variable respuesta no fuese un valor, sino una secuencia de acciones? En este caso, una simple acción no es importante. Lo que importa es la secuencia de acciones que generan una correcta política de acción. Una acción es buena si forma parte de una buena política. Los métodos de aprendizaje por refuerzo evalúan la eficiencia de una determinada política y aprenden secuencias de acciones pasadas para generar políticas exitosas.

Ejemplos de aprendizaje por refuerzo los podemos encontrar en la automatización de juegos, métodos heurísticos para la optimización (como la colonia de hormigas) o búsqueda de caminos en robots autónomos.

Dado que nuestro problema consiste en predecir subidas o bajadas en el precio de una divisa dentro del mercado FOREX, lo afrontaremos como un problema de clasificación. Por tanto, profundizaremos un poco más en este tema.

5.2. Técnicas de aprendizaje supervisado para la clasificación

Queremos que el ordenador aprenda a clasificar cuándo unos determinados datos definen una subida en el precio de una divisa y cuándo una bajada.

En este ámbito, el proceso de aprendizaje se basa en deducir una función que asocie unos datos con su respectiva clase (subida o bajada del precio) a partir de la observación de unos ejemplos, llamados *datos de entrenamiento*. Estos ejemplos son, como ya sabemos, pares de objetos, inputs y outputs, correctamente asociados que el algoritmo analiza y de donde saca conclusiones (detecta patrones y relaciones no visibles a simple vista).

Por tanto, queremos crear una función capaz de predecir la clase correspondiente a cualquier objeto de entrada después de haber visualizado una serie de ejemplos. El proceso

de adaptación de las normas y pautas aprendidas por el algoritmo a nuevos ejemplos (desconocidos hasta el momento) se conoce como *generalización*.

Intuitivamente, conforme más ejemplos se analicen, más se conocerá el proceso que define a los datos. Sin embargo, es imposible observar todos los ejemplos, por lo que solo conocemos cual es la salida correcta para un pequeño porcentaje de los datos. Esto conlleva a la no existencia de una única solución: ante datos de entrenamiento distintos, se obtienen clasificadores distintos.

Muchos problemas surgen en el proceso de aprendizaje. Hay que realizar suposiciones que hacen posible el aprendizaje ajenas a la elección de los datos de entrenamiento, como por ejemplo el tipo de clasificador a usar. Sabemos que, para aprender, el algoritmo analiza los datos de entrenamiento. El cómo los interpreta y manipula depende del modelo usado.

Por tanto, hay que manejar dos problemas en el proceso de aprendizaje: la selección del “mejor” conjunto de variables predictoras, esto es, aquel que sea más representativo para los datos y permita diferenciar significativamente las clases (ver figura 5.2) y la selección del “mejor” modelo. La elección de los mejores candidatos en ambos casos se realiza de forma empírica, puesto que depende exclusivamente de los datos que manipulamos.

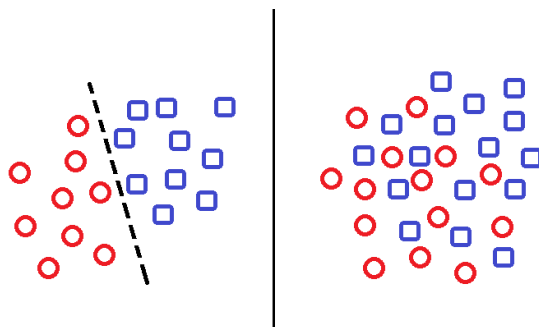


Figura 5.2: Representación gráfica del conjunto de variables predictoras (coordenadas de cada figura). Diferencias entre un buen conjunto de variables predictoras, que permite diferenciar fácilmente entre clases (cuadrado o círculo), frente a uno menos favorable. Fuente: elaboración propia.

De forma paralela, tenemos el problema de cómo evaluar los modelos. La forma general de medir la calidad de un modelo es midiendo su capacidad predictiva, realizando medidas objetivas basadas en la comparación de los valores observados para la variable respuesta y los valores predichos por el modelo. La evaluación del modelo se debe hacer sobre un conjunto de datos distinto al que se ha usado para construir el modelo, evitando así caer en el sobreajuste (*overfitting*¹).

¹El sobreajuste es el fenómeno que ocurre cuando un modelo se ha adecuado demasiado a los datos de entrenamiento, de forma que no aprende de ellos, sino que los memoriza e imita (ver [8]).

Con todo esto se busca medir de forma “honesta” la habilidad de generalización del modelo, también conocido como la calidad de su *sesgo inductivo*², así como la capacidad predictiva del mismo.

La forma de actuar que llevaremos a cabo (y que es la común en este tipo de problemas) para obtener una estimación honesta de la capacidad predictiva de un modelo consistirá en dividir la muestra original de ejemplos en dos submuestras:

- Una muestra de entrenamiento (*training set*), usada para la construcción del modelo.
- Una muestra de test (*testing set*), donde se realizan las predicciones y se evalúa la capacidad predictiva del modelo.

A su vez, para elegir entre el conjunto de modelos candidatos proporcionados por la muestra de entrenamiento, y medir la calidad de su sesgo inductivo, se divide el conjunto de entrenamiento en dos partes: una para el entrenamiento y la otra, llamada muestra de validación (*validation set*), para testear la habilidad de generalización del modelo entrenado.

La forma usual de hacer esto es dividir aleatoriamente el conjunto de entrenamiento en K subconjuntos de similar tamaño y ajustar K modelos, dejando cada vez un elemento de la partición para la validación y construyendo el modelo con los $K - 1$ restantes. Este proceso se conoce como *K-cross-validation* (figura 5.3).

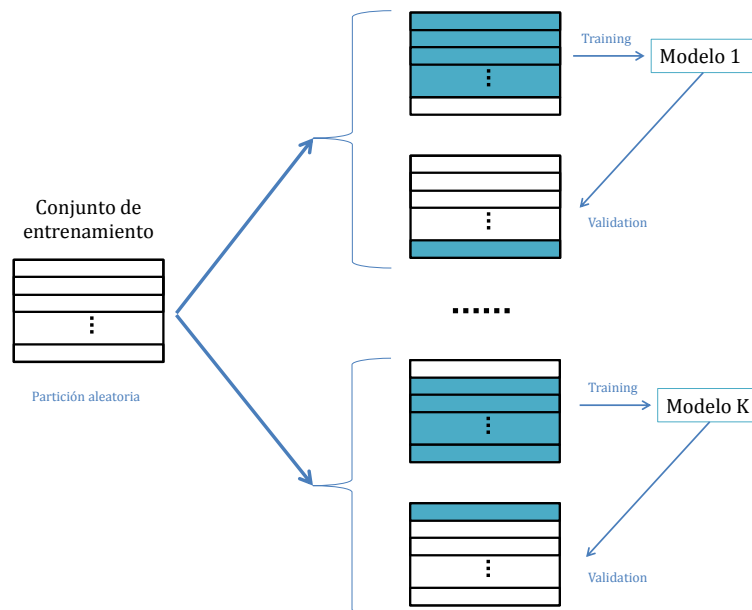


Figura 5.3: Procedimiento para la *cross-validation* o validación cruzada. Fuente: elaboración propia.

²En machine learning, se define el sesgo inductivo (*inductive bias*) como el conjunto de hipótesis y suposiciones que hacen posible el aprendizaje. Por ejemplo, un sesgo inductivo es asumir que ejemplos con datos de entrada muy similares tienen una salida muy similar.

Cada modelo (o conjunto de parámetros de un modelo) evaluado en este proceso tendrá asociado un error, al cual llamaremos error de validación, que se calcula como promedio de los K errores evaluados en las muestras de validación de las K particiones.

Para los problemas de clasificación existen varias formas de medir el error de predicción de un modelo. Para nuestro problema usaremos la matriz de confusión, que, en el caso de dos posibles clases, consiste en lo siguiente:

Cuadro: Matriz de confusión del problema para dos variables respuesta.

Predicho	Observado		Total predicho
	Y = baja	Y = sube	
Y = baja	VP	FP	VP + FP
Y = sube	FN	VN	FN + VN
Total observado	VP + FN	FP + VN	N

donde consideraremos *baja* como nuestra clase positiva y *sube* la negativa. Nos centraremos en el valor de la exactitud (*Accuracy*) para medir el error:

$$\text{Accuracy} = \frac{VP + VN}{N},$$

teniendo en cuenta también los valores de sensibilidad y especificidad de las predicciones:

$$\text{Sensibilidad} = \frac{VP}{VP + FN}, \quad \text{Especificidad} = \frac{VN}{VN + FP}.$$

Otras medidas de error interesantes pueden ser, por ejemplo, las curvas ROC (ver [5]).

Tras elegir el mejor modelo para los datos de entre un conjunto de modelos candidatos, toca estimar un error de generalización para el mismo. Para obtener una idea del error que se espera que cometa el mejor modelo no se usa el error de validación. En este caso se utiliza el conjunto de test, que refleja de forma más precisa el error de generalización del modelo.

Recapitulando, el proceso de aprendizaje consta de los siguientes pasos:

1. Elección de los datos. Variables predictoras (inputs) y respuesta (outputs).
2. Preprocesamiento de los datos y reducción de la dimensionalidad ³.
3. Construcción de las muestras de training y testing.

³Pese a que no se haya hecho hincapié en este paso, no debe ser tomado en balde. El preprocesamiento de los datos puede consistir en un análisis de componentes principales (PCA), un análisis de correlaciones en las variables predictoras u otras formas (ver [1])

4. Entrenamiento del modelo con la muestra de training y elección de los mejores parámetros con la muestra de validación.
5. Preentrenamiento del modelo seleccionado usando todo el training set.
6. Testeo de los resultados con la muestra de testing para estimar el error esperado del modelo elegido.
7. Uso del modelo para predicción o descripción.

Una versión más estricta sobre el proceso de aprendizaje puede verse en [1]. Para aquellos más interesados en los aspectos prácticos del proceso, se les aconseja [2], donde podrán obtener un amplio tutorial sobre el proceso de aprendizaje automático.

5.2.1. El proceso de predicción usando herramientas de machine learning

Todo depende del problema al que hagamos frente, y me gustaría hacer hincapié en esta frase. En los problemas clásicos donde se hace uso de herramientas de aprendizaje automático (algunos citados anteriormente como ejemplos) es la elección del modelo predictor y sus parámetros lo que determina el éxito. Distintos modelos (árboles de decisión, redes neuronales, perceptrones multicapa, regresores lineales, etc.) dan distintos resultados, todo depende de los datos.

Éste es el principal problema al que nos enfrentamos: ¿qué datos son más explicativos para el mercado FOREX? ¿con qué datos uno es capaz de prever un movimiento determinado en el mercado, y cómo interpretarlos? No lo sabemos, y es algo que, por la naturaleza del problema, no se puede saber a priori. Solo podemos hacer suposiciones, experimentos y comparar resultados.

La elección del algoritmo más adecuado para los datos es otro mundo aparte. Por cada par de algoritmos de aprendizaje, hay tantos problemas en el que el primer algoritmo es mejor que el segundo como problemas en el que el segundo es mejor que el primero (*Teorema del No Free Lunch*, ver [8]).

Puesto que gran parte del conocimiento a priori del problema se basa en mera suposición, no se han de descartar opciones sin evidencias empíricas.

El proceso de predicción diseñado usando herramientas de machine learning se describe en la figura 5.4. El valor predicho nos dirá si el precio del elemento financiero del mercado FOREX introducido subirá o bajará en la hora a predecir con respecto a su hora anterior. La parte de aprendizaje automático del código está realizado, en su gran mayoría, con el paquete *caret* del programa R.

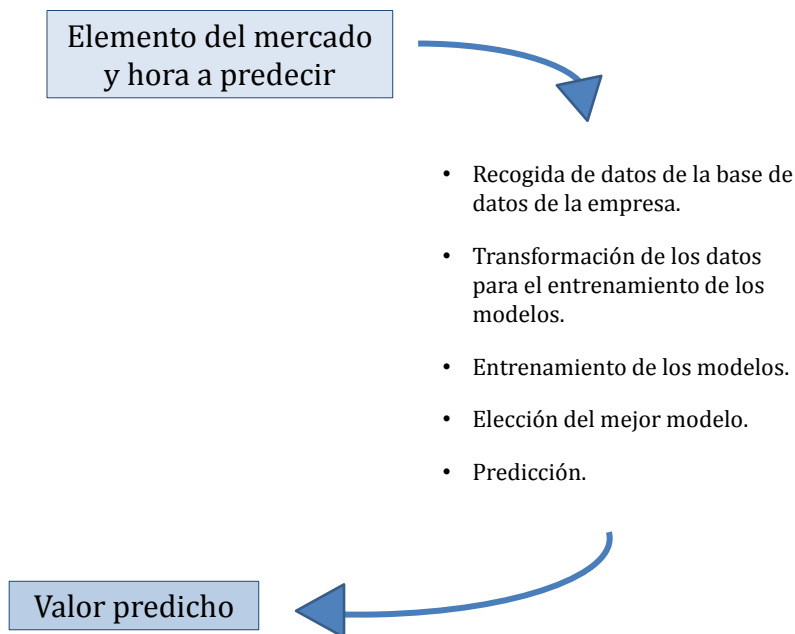


Figura 5.4: Proceso de predicción de un elemento del mercado FOREX en una determinada hora haciendo uso de los métodos de aprendizaje automático. Fuente: elaboración propia.

5.2.2. Algunos métodos de aprendizaje automático

La tarea de un clasificador es separar el espacio de variables en regiones de decisión (clases). Los límites entre diferentes regiones se conocen como fronteras de decisión. El cálculo de estas fronteras de decisión es el objeto de los modelos de clasificación. Ya hemos estado hablando de cómo funciona el proceso de aprendizaje, ahora vamos a ver como se efectúa dicho aprendizaje dentro de los modelos. Para ilustrarlo, expondremos algunos de los modelos implementados.

Arboles de decisión.

Los *árboles de decisión* consisten en una estructura de datos jerárquica basada en la filosofía de “divide y vencerás”. Son modelos jerárquicos basados en la teoría de árboles, formados por nodos internos (*nodos de decisión*) y hojas finales. En cada nodo de decisión se implementa una función $f_m(x)$ (discriminante) que genera un conjunto discreto de salidas. En función del resultado de esta función, se escoge una determinada rama del nodo de decisión.

Uno de los principales atractivos de los árboles de decisión es que son modelos de “caja abierta”, es decir, modelos donde se puede observar directamente las decisiones que va ejecutando (son fáciles de entender e interpretar).

El proceso de construcción del árbol es su proceso de aprendizaje. El modelo analiza los datos y extrae de ellos las reglas lógicas que le permiten clasificar los datos. Se trata de un proceso recursivo, que consta de los siguientes pasos:

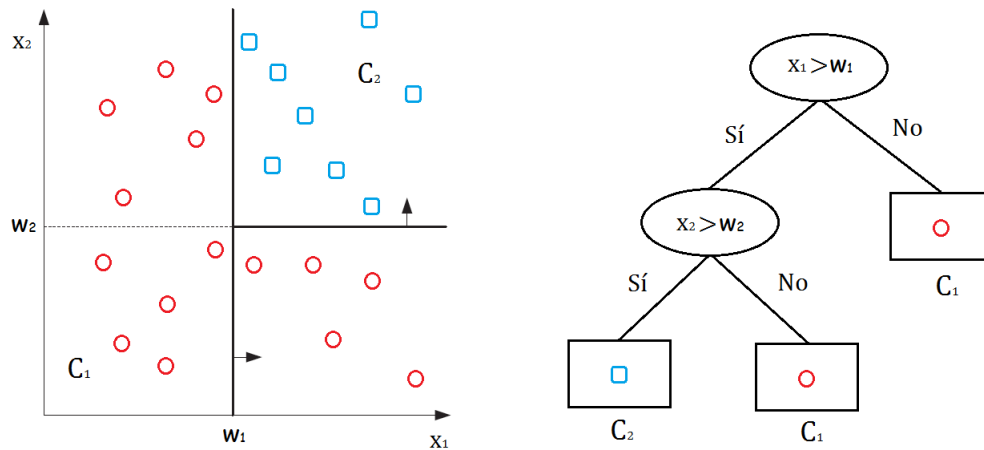


Figura 5.5: Ejemplo de árbol de decisión para dos variables (x_1 y x_2) y dos clases (C_1 y C_2). A la izquierda, el espacio de variables con cada ejemplo. A la derecha, el árbol generado, donde los nodos obalados son nodos de decisión y los rectangulares son nodos hoja. Fuente: elaboración propia.

1. Se analizan las distintas particiones del espacio y se escogen aquellas que generen mejores regiones de decisión.
2. Se optimiza la separación.
3. Se repite el paso 1 con los nodos hijos que sean de decisión.

El algoritmo de creación de los árboles depende del tipo de árbol diseñado. Así, existen distintos algoritmos para tal fin, tales como el *ID3*, el *CART* o el *C5.0*. Por otro lado, existen muchas técnicas que intentan mejorar el rendimiento de los árboles de decisión, como las técnicas de *bagging* o el *Random Forest*, que parten de un conjunto amplio de árboles de decisión válidos para el problema con el fin de mejorar el grado de clasificación, cada uno con sus respectivas particularidades.

K-nearest neighbours. Knn

Con frecuencia, la simplicidad genera mejores resultados que la complejidad. El algoritmo de los *k-vecinos más cercanos*, o *Knn*, es un algoritmo simple que almacena todos los casos disponibles y clasifica en función de una medida de similitud. Un nuevo caso se clasifica en la clase predominante de su vecindario, en función de la “mayoría de votos” de sus vecinos.

Para agrupar los vecinos, se establece una métrica en el conjunto de puntos de entrenamiento, la cual puede ser la distancia Euclídea, la distancia de Manhattan (o rectangular) o la distancia de Minkowski. En el caso de variables categóricas, se suele usar la distancia de Hamming.

Por tanto, el proceso de aprendizaje consiste en: una vez seleccionada una métrica adecuada para el problema, se agrupan los datos de entrenamiento en k “vecindarios”, de tal forma que esta agrupación sea óptima. Cada dato nuevo se clasifica en función de la clase a la que pertenece la mayoría de variables de su vecindario.

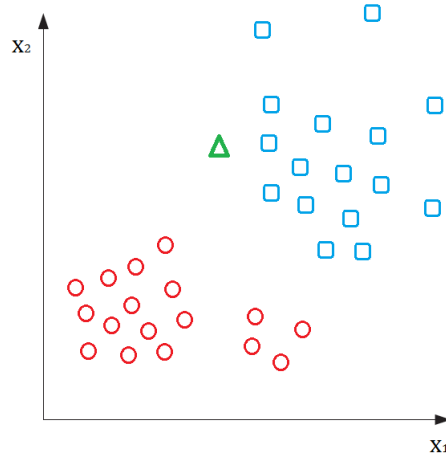


Figura 5.6: Ejemplo del modelo de Knn para el caso de dos variables predictoras (x_1 y x_2). El nuevo dato (triángulo verde) se clasificaría como cuadrado azul debido a que es la clase predominante de su vecindario. Fuente: elaboración propia.

Otros modelos, como el *support vector machine* (SVM), hacen uso de hiperplanos para la separación del espacio de variables cuando la dimensionalidad es muy alta. La idea tras los modelos de clasificación está clara: separar el espacio de variables en distintas regiones pertenecientes a distintas clases, siendo el cómo se hace cuestión de cada modelo en particular.

En total, se han utilizado un total de 15 modelos distintos dentro de nuestro algoritmo (ver tabla 5.1). Como ya se sabe, la elección de un modelo u otro para la predicción se deja en manos de los datos.

Cuadro 5.1: Tabla de modelos

Modelo	Modelo	Modelo	Modelo	Modelo
Support Vector Machines with Linear Kernel	<i>svmLinear</i>	K - Neighbours	<i>knn</i>	C 5.0 <i>C5.0</i>
Neural Network	<i>nnet</i>	Random Forest	<i>rf</i>	CART Trees <i>rpart</i>
Penalized Discriminant Analysis	<i>pda</i>	Naive Bayes	<i>nb</i>	Generalized Additive Model using Splines <i>gam</i>
Linear Discriminant Analysis	<i>lda</i>	Single C5.0 Tree	<i>C5.0Tree</i>	Generalized Linear Model with Stepwise Feature Selection <i>glmStepAIC</i>
Generalized Linear Model	<i>glm</i>	Single C5.0 Ruleset	<i>C5.0Rules</i>	Maximum Uncertainty Linear Discriminant Analysis <i>Mlda</i>

5.3. Pruebas, resultados y conclusiones

En este apartado expondremos los resultados obtenidos con el modelo de predicción basados en técnicas de aprendizaje automático. Las pruebas se realizaron sobre la cotización del EURUSD. En total, se realizaron 24 problemas, uno para cada hora del día.

En la tabla 5.2 se exponen los mejores modelos obtenidos para cada problema, así como su precisión, especificación y sensibilidad.

Cuadro 5.2: Mejores modelos para EURUSD

	Modelo	Precision	Sensitivity	95 % CI	Specificity
00:00	C5.0	0,59	0,75	(0.55, 0.64)	0,42
01:00	nnet	0,54	0,50	(0.49, 0.58)	0,57
02:00	nb	0,58	0,55	(0.54, 0.63)	0,61
03:00	rpart	0,56	0,41	(0.51, 0.61)	0,69
04:00	knn	0,59	0,22	(0.54, 0.63)	0,88
05:00	pda	0,60	0,39	(0.55, 0.64)	0,79
06:00	glm	0,55	0,65	(0.51, 0.59)	0,44
07:00	glmStepAIC	0,56	0,10	(0.51, 0.60)	0,95
08:00	nnet	0,53	0,24	(0.48, 0.57)	0,79
09:00	knn	0,53	0,62	(0.48, 0.58)	0,43
10:00	nb	0,53	0,34	(0.48, 0.58)	0,71
11:00	knn	0,55	0,48	(0.50, 0.59)	0,60
12:00	knn	0,55	0,62	(0.50, 0.60)	0,46
13:00	pda	0,55	0,72	(0.51, 0.60)	0,36
14:00	pda	0,52	0,78	(0.48, 0.57)	0,24
15:00	nnet	0,53	0,66	(0.47, 0.57)	0,38
16:00	knn	0,51	0,51	(0.47, 0.56)	0,51
17:00	rf	0,56	0,55	(0.52, 0.61)	0,57
18:00	glm	0,60	0,55	(0.55, 0.64)	0,63
19:00	knn	0,59	0,56	(0.54, 0.63)	0,61
20:00	rf	0,56	0,48	(0.51, 0.60)	0,61
21:00	knn	0,57	0,57	(0.53, 0.62)	0,58
22:00	nb	0,58	0,33	(0.53, 0.62)	0,76
23:00	glm	0,59	0,92	(0.54, 0.64)	0,16

En estas pruebas se puede apreciar la necesidad de dividir nuestro problema principal en 24 subproblemas, puesto que cada hora es diferente dentro del mercado FOREX. Esto es debido a la influencia de tres mercados distintos: el mercado asiático, el europeo y el americano. Por lo general, el comienzo de cada mercado se marca, según la hora española, a las 00:00 (mercado asiático), 07:00 (mercado europeo) y 14:00 (mercado americano). Esto hace que para medio día, los tres mercados se junten, generando una alta volatilidad en los datos y que a los modelos les sea difícil aprender. Por ello, los mejores resultados se aprecian en las horas de menos conflicto entre mercados.

Los resultados obtenidos con este modelo predictivo reflejan una mejora frente a los obtenidos con el modelo ARIMA, llegando incluso a acercarse al 60% de precisión. En

muchos casos se puede apreciar como los modelos no aprenden de los datos, limitándose a clasificar todos los casos en la clase mayoritaria. Esto sugiere que las variables usadas para el aprendizaje no son las idóneas, por lo que se debería seguir haciendo pruebas en busca de unos datos que aporten mejores resultados.

Por otro lado, y con el fin de no confundir al lector, el hecho de que unos modelos sean elegidos más veces que otros no implica que estos sean mejores. Han dado mejores resultados para los datos de entrenamiento que se le han introducido. Esto no quiere decir que en un futuro, con la incorporación de nuevos ejemplos, sigan siendo los mejores modelos. Por tanto, la elección del mejor modelo se debe hacer, como mínimo, semanalmente.

Por último, se eliminaron de las pruebas los modelos *Mlda* y *lda* debido a fallos computacionales y malos resultados. Al final de este apartado se incluye la tabla completa de resultados para el EURUSD. Como el lector podrá observar, los casos de empate se resuelven arbitrariamente.

	00:00	01:00	02:00	03:00	04:00	05:00	06:00	07:00	08:00	09:00	10:00	11:00
C5.0	0,59275	0,50746	0,53205	0,52239	0,55342	0,56503	0,51064	0,52979	0,51702	0,52017	0,50743	0,51805
C5.0Rules	0,59275	0,50746	0,53205	0,52239	0,55342	0,56503	0,51064	0,52979	0,51702	0,52017	0,50743	0,51805
C5.0Tree	0,59275	0,50746	0,53205	0,52239	0,55342	0,56503	0,51064	0,52979	0,51702	0,52017	0,50743	0,51805
gam	0,44136	0,49893	0,45727	0,46908	0,44658	0,41791	0,44681	0,49787	0,47660	0,46921	0,49045	0,46709
glm	0,54371	0,51173	0,54274	0,53731	0,55342	0,59915	0,55319	0,52128	0,52340	0,51380	0,51805	0,45648
glmStepAIC	0,56503	0,49041	0,54915	0,53731	0,55342	0,58849	0,53830	0,55745	0,52553	0,53079	0,52654	0,51805
knn	0,55864	0,48401	0,53419	0,55864	0,58761	0,57996	0,54681	0,51915	0,51702	0,53079	0,51168	0,54565
nb	0,55011	0,51599	0,58333	0,55437	0,56197	0,57356	0,54681	0,54468	0,50638	0,48620	0,53079	0,53503
nnet	0,56930	0,53518	0,53632	0,53945	0,55342	0,59275	0,54468	0,52979	0,52553	0,52654	0,52654	0,54089
pda	0,53945	0,48401	0,54487	0,53518	0,55342	0,59915	0,54894	0,52128	0,52340	0,51168	0,51805	0,45648
rf	0,54371	0,52878	0,55769	0,52026	0,57265	0,54371	0,47660	0,53404	0,43617	0,52229	0,52229	0,53503
rpart	0,59062	0,52665	0,53205	0,56077	0,55342	0,56503	0,53617	0,53617	0,51489	0,51805	0,51805	0,52017
svmLinear	0,54797	0,47548	0,55769	0,52239	0,55342	0,57996	0,52128	0,52979	0,51702	0,51168	0,49257	0,51805

	12:00	13:00	14:00	15:00	16:00	17:00	18:00	19:00	20:00	21:00	22:00	23:00
C5.0	0,52442	0,51592	0,52017	0,50743	0,50955	0,51168	0,58599	0,53404	0,51923	0,53419	0,55794	0,55733
C5.0Rules	0,52442	0,51592	0,52017	0,50743	0,50955	0,51168	0,58599	0,53404	0,51923	0,53419	0,55794	0,55733
C5.0Tree	0,52442	0,51592	0,52017	0,50743	0,50955	0,51168	0,58599	0,53404	0,51923	0,53419	0,55794	0,55733
gam	0,46709	0,45648	0,47134	0,49682	0,49045	0,51592	0,40127	0,44255	0,45299	0,45299	0,43777	0,42667
glm	0,51805	0,55201	0,52441	0,50955	0,48832	0,48408	0,58872	0,55745	0,53419	0,55556	0,56867	0,58933
glmStepAIC	0,52442	0,54565	0,51592	0,50743	0,48832	0,51168	0,59448	0,54681	0,53846	0,55556	0,55365	0,58133
knn	0,54989	0,53716	0,49045	0,47346	0,51380	0,47983	0,59660	0,58510	0,53632	0,57264	0,56438	0,58400
nb	0,52442	0,52866	0,52442	0,47346	0,50743	0,52442	0,57325	0,55745	0,53632	0,56410	0,57510	0,57333
nnet	0,52442	0,54777	0,52017	0,53078	0,48832	0,48620	0,59236	0,55957	0,54701	0,52564	0,56438	0,55733
pda	0,51805	0,5520	0,52441	0,50106	0,48620	0,48408	0,59660	0,55745	0,53419	0,55128	0,56867	0,58667
rf	0,53716	0,47558	0,46921	0,45648	0,47771	0,56475	0,57113	0,55532	0,55555	0,53846	0,53219	0,55733
rpart	0,53716	0,49682	0,51805	0,48195	0,46709	0,52017	0,58599	0,53617	0,52991	0,54915	0,54936	0,55733
svmLinear	0,52442	0,51592	0,52017	0,50743	0,50955	0,51168	0,58811	0,55106	0,52137	0,54060	0,55794	0,55733

Capítulo 6

Conclusiones

Tal y como nos aventurábamos a decir al principio, el problema al que nos enfrentamos no es en absoluto sencillo, y en este documento queda reflejado. Nuestro objetivo principal era predecir el movimiento a pequeña escala en el mercado de divisas, de tal forma que fijado un instante de tiempo tuviéramos a mano una herramienta que nos dijera si el precio del elemento financiero que estuviéramos observando iba a subir o bajar con respecto a su precio previo.

Como hemos podido observar, las distintas herramientas que se han ido implementando han reflejado resultados insatisfactorios, haciendo que nos planteemos la confianza que podemos depositar en las predicciones que nos dan.

No obstante, hemos de tratar con cuidado estos “malos resultados”. En primer lugar, es posible que los resultados obtenidos con el algoritmo de automatización del modelo ARIMA no se midieran de la forma más sensata posible. Tal y como está diseñado el algoritmo, y debido a la exigencia de arrojar resultados siempre que se le hace una consulta, éste muestra los mejores pronósticos que calcula, y no siempre en estos pronósticos captan la frecuencia de los datos (ver figura 6.1). Por tanto, se aconseja al usuario que tenga en cuenta estos pronósticos siempre y cuando en estos se muestre algo más que la tendencia de la serie. El error del algoritmo recae, por tanto, en que no siempre consigue dar estas predicciones “fiables”.

Por otro lado, los resultados obtenidos por las técnicas de aprendizaje automático son preliminares. Actualmente se está trabajando en la búsqueda del mejor conjunto de entrenamiento para los modelos y por tanto el código diseñado está aun en fase beta. Además, los resultados obtenidos hasta ahora son esperanzadores ya que se aproximan al 60% de acierto, y se espera que con mejores datos se supere esta frontera.

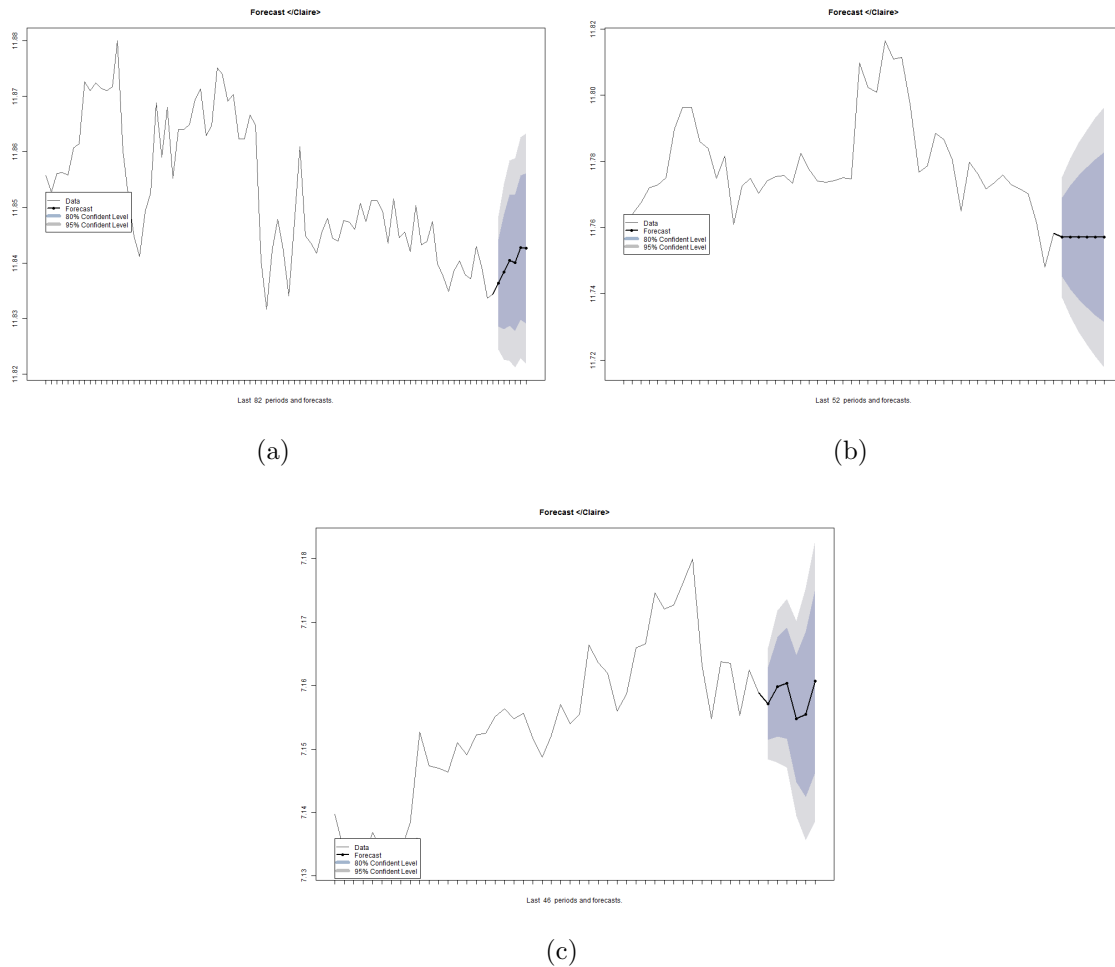


Figura 6.1: Diferentes resultados del algoritmo de automatización de series temporales. Fuente: <http://www.elrincondeltrader.com/>.

6.1. Aportaciones a la empresa

Como ya se ha hecho mención antes, la principal aportación realizada a la empresa es la implantación del algoritmo de automatización del modelo ARIMA en su herramienta web: <http://elrincondeltrader.com>. La idea original de esta herramienta era aportar al analista experto una primera predicción a partir de la cual el usuario pudiese tomar sus propias decisiones.

Junto con esto, se le abrió un campo nuevo de estudio a la empresa introduciéndola en el estudio de series temporales. Partiendo del algoritmo ya en funcionamiento, la empresa puede mejorarlo introduciendo distintas novedades, como pueden ser:

- Una mejora en el preprocesamiento de los datos, al estilo del realizado en el programa TRAMO/SEATS.
- La inclusión del estudio de la componente estacional.
- La adición al modelo ARIMA calculado por el algoritmo de modelos heterocedásticos, como pueden ser los modelos ARCH o GARCH, complementando así el estudio de series temporales.

Por otro lado se encuentra en proceso de implantación el algoritmo de predicción usando técnicas de aprendizaje automático. Se espera que una vez perfeccionado este algoritmo se establezca la opción de obtener predicciones con él en la herramienta web de la empresa.

6.2. Futuras líneas de trabajo

Tal y como se lleva adelantando durante todo este texto, los elementos usados en este trabajo han abierto futuras líneas de investigación para la empresa:

- Mejoras en el algoritmo de automatización del modelo ARIMA tal y como se ha mencionado antes.
- Investigación de otros programas que automatizan el estudio ARIMA, como pueden ser TRAMO/SEATS, X13-ARIMA-SEATS, etc.
- Implementación de más modelos de aprendizaje automático dentro de su respectivo código (en la actualidad hay implementados 15).
- Estudio de los datos de entrenamiento a usar en dicho código.

Pese a todo el trabajo que aún queda por hacer, el lector puede pensar que todo el esfuerzo realizado en este proyecto ha sido en balde dados los resultados obtenidos.

Desde un punto de vista científico, es tan útil encontrar un método que funcione como uno que no funcione, ya que este último guiará el camino de otros investigadores. Quizás los modelos implementados no sean los más idóneos para este problema, o quizás sí, el trabajo futuro lo dirá.

En una entrevista a Thomas Edison, se le preguntó cómo no se había desanimado en su intento por inventar la bombilla habiendo fracasado tantas veces antes de conseguirlo, a lo que el inventor respondió: “no fracasé, sólo descubrí 999 maneras de cómo no hacer una bombilla”. No existen los malos resultados.

Bibliografía

- [1] Alpaydin, E. (2010) *Introduction to Machine Learning*. Second Edition. Massachusetts Institute of Technology. The MIT Press.
- [2] Andrew Ng. (2011). Coursera - Machine Learning. Stanford University. <http://www.coursera.org/course/ml>.
- [3] Caballero Díaz, F.F. (2011) *Selección de modelos mediante criterios de información en análisis factorial. Aspectos teóricos y computacionales*. Tesis Doctoral. Departamento de Estadística e I.O. Universidad de Granada. Granada.
- [4] Espín López, J.M. (2015). *Electricidad, matemáticas e inteligencia*. Trabajo Fin de Máster. Máster en Matemática Avanzada y Profesional. Facultad de Matemáticas. Universidad de Murcia.
- [5] Franco, M. y Vivo, J.M. (2015). Apuntes de asignatura. *Métodos Estadísticos de Predicción y Clasificación*. Máster en Matemática Avanzada y Profesional: Perfil de Empresa. Facultad de Matemáticas. Universidad de Murcia.
- [6] Gómez, V. Maravall A. (1997). *Programs TRAMO and SEATS. Instructions for the User*. Banco de España. Madrid.
- [7] González Casimiro, M.P. (2009). *Análisis de series temporales: modelos ARIMA*. Departamento de Economía Aplicada III. Facultad de Ciencias Económicas y Empresariales. Universidad del País Vasco.
- [8] Marín-Blázquez Gómez, F.J. (2015). Apuntes de asignatura. *Aprendizaje Computacional*. Grado en Ingeniería Informática. Facultad de Informática. Universidad de Murcia.
- [9] Mauricio, J.A.(2007). *Introducción al Análisis de Series Temporales*. Apuntes de la asignatura "Series Temporales". Licenciatura de Economía. Universidad Complutense de Madrid.
- [10] Murphy, J.J. (1999). *Análisis Técnico de los Mercados Financieros*. New York Institute of Finance. New York: Gestión 2000.
- [11] Peña Sánchez de Rivera, D. (1994). *Estadística: Modelos y Métodos 2. Modelos lineales y series temporales*. 2ª edición. Madrid. Alianza Universidad Textos.
- [12] Peña Sánchez de Rivera, D. (2002). *Análisis de datos multivariantes*. Madrid. McGraw-Hill.

- [13] Pina Canelles, A. (2014). *Análisis técnico de mercados financieros basado en técnicas de inteligencia artificial*. Trabajo Fin de Grado. Grado en Matemáticas. Facultad de Matemáticas. Universidad de Murcia.
- [14] Taylor, S. (2007). *Asset Price Dynamics, Volatility, and Prediction*. Princeton and Oxford. Princeton University Press.