COMILLAS
UNIVERSIDAD PONTIFICIA
ICAI

BACHELOR OF INDUSTRIAL TECHNOLOGIES ENGINEERING

FINAL DEGREE PROJECT

FORECASTING INTERVAL TIME SERIES AND CANDLESTICK CHARTS. COMPARING DIFFERENT REGRESSION APPROACHES BASED ON KERNEL SMOOTHING IN FINANCIAL AND ENERGY MARKETS.

Author: Alejandro Agustín de Diego Rodríguez

Director: Carlos Maté Jiménez

Madrid

2021

Declaro, bajo mi responsabilidad, que el Proyecto presentado con el título

Forecasting interval time series and candlestick charts. Comparing different regression approaches based on kernel smoothing in financial and energy markets.

en la ETS de Ingeniería - ICAI de la Universidad Pontificia Comillas en el

curso académico 2020/21 es de mi autoría, original e inédito y

no ha sido presentado con anterioridad a otros efectos.

El Proyecto no es plagio de otro, ni total ni parcialmente y la información que ha sido

tomada de otros documentos está debidamente referenciada.

Fdo.: Alejandro Agustín de Diego Rodríguez Fecha: 09/ 07/ 2021

Autorizada la entrega del proyecto

EL DIRECTOR DEL PROYECTO

Fdo.: Carlos Maté Jiménez Fecha: 09/ 07/ 2021

“Divide et impera”

Gaius Julius Caesar

Agradecimientos

Me gustaría dedicar este espacio para agradecer a las personas que, de una forma u otra, me han permitido llegar hasta aquí.

En primer lugar, agradezco al Dr. Carlos Maté Jiménez por haberme dado la oportunidad de llevar a cabo el proyecto que no hubiera podido finalizar sin su ayuda y paciencia durante las largas horas de aprendizaje. Su guía, aún con siete horas de diferencia horaria, me ha permitido mantener la hoja de ruta en los momentos en los que parecía desvanecer.

A mis padres, María y Alejandro, por vuestro esfuerzo, sacrificio y amor para darme una educación inmejorable e invaluable. Siempre pensando en lo mejor para mí, sois un verdadero ejemplo a seguir. Os estaré eternamente agradecido.

A mis hermanas, por haberme acompañado a lo largo de estos años, apoyándome en todas mis decisiones y esbozando mi forma de ser con su cariño, ejemplo y alegría.

También quería agradecer especialmente a Mimi, por haber sido el pilar que me ha sustentado inquebrantablemente en los momentos difíciles, depositando toda su confianza en mí. Gracias por tenderme un puente fuera de las preocupaciones de la carrera todos estos años.

Por último y no menos importante, a todos mis amigos. Por haberme brindado momentos de diversión, por el gran compañerismo vivido afrontando retos a lo largo de la carrera, y por su inestimable apoyo.

COMILLAS
UNIVERSIDAD PONTIFICIA
ICAI

BACHELOR OF INDUSTRIAL TECHNOLOGIES ENGINEERING

FINAL DEGREE PROJECT

FORECASTING INTERVAL TIME SERIES AND CANDLESTICK CHARTS. COMPARING DIFFERENT REGRESSION APPROACHES BASED ON KERNEL SMOOTHING IN FINANCIAL AND ENERGY MARKETS.

Author: Alejandro Agustín de Diego Rodríguez

Director: Carlos Maté Jiménez

Madrid

2021

PREDICCIÓN DE SERIES TEMPORALES DE INTERVALOS Y GRÁFICOS DE VELA. COMPARACIÓN DE DISTINTOS MODELOS DE REGRESIÓN BASADOS EN ALISADO DE NÚCLEO EN LOS MERCADOS FINANCIEROS Y ENERGÉTICOS.

Autor: de Diego Rodríguez, Alejandro
Director: Maté Jiménez, Carlos
Entidad Colaboradora: ICAI – Universidad Pontificia Comillas

Resumen del proyecto

En este proyecto, varios modelos de regresión y predicción de núcleo para series de tiempo de valores puntuales y de intervalo (STI) son introducidos, probados, y comparados en los mercados financieros y energéticos. Una alta personalización de los modelos es conseguida gracias al amplio rango de alternativas ofrecido para cada uno de los tres componentes: estimador, función de núcleo, y ancho de banda. Se explora por primera vez el campo de la predicción de núcleo para STI con la aplicación de dos modelos de predicción originales basados en los métodos de centro y centro y rango. En la evaluación del rendimiento se tienen en cuenta, entre otras, 7 series temporales de cambio de divisas muy concurridas, así como el petróleo Brent. Las precisiones de las predicciones se evalúan mediante 9 medidas de error diferentes.

Palabras Clave: Regresión kernel, Predicción kernel, datos valorados por intervalo, Mercados energéticos, Mercados financieros, Mercados de divisas

1. Introducción

Desde su primer uso primitivo en el antiguo Egipto, los métodos de previsión y regresión han sido una herramienta indispensable para la organización, la optimización y la evolución de la humanidad. Con el reciente avance del almacenamiento masivo de datos en las últimas décadas, la necesidad de dar valor a esta colosal cantidad de información ha aumentado exponencialmente. Los datos son el nuevo petróleo y el análisis de datos su industria de refinamiento.

Los modelos de regresión y previsión no paramétricos desempeñan un papel importante cuando los métodos paramétricos clásicos no consiguen explicar el comportamiento de una variable. Están más presentes en el mercado financiero debido a su mayor adaptabilidad y rendimiento. Entre ellos, los modelos de regresión kernel han captado la atención de muchos investigadores debido a su equilibrada complejidad y rendimiento. Se pueden encontrar muchos estudios sobre la regresión kernel y la previsión. Sin embargo, no existe literatura académica sobre la previsión de series temporales con valores de intervalo.

Los fenómenos del mundo real, como las temperaturas meteorológicas, la velocidad del viento, o los mercados bursátiles, tienen sus límites diarios (o semanales, o mensuales) y varían en cada periodo (ya sea un día, una semana o un mes). La predicción del precio de cierre de una acción fracasa en la transmisión de información clave de las variables de alta frecuencia. En cambio, las series temporales con valores de intervalo permiten representar características importantes, así como la variabilidad de éstas. En consecuencia, la previsión de las ITS tiene un futuro prometedor en un área de investigación inmadura.

Esta investigación tiene tres objetivos principales. El primero es introducir al lector no familiarizado con la regresión y previsión kernel para series de tiempo de valores puntuales y de intervalo. El segundo objetivo es adaptar la teoría de kernel para datos de valores puntuales a los datos con valores de intervalo. Finalmente, el último objetivo es probar y comparar el rendimiento de los diferentes modelos de regresión y previsión en los mercados financieros y energéticos con datos actuales. El tercer objetivo también pretende investigar si un modelo de predicción prometedor puede tener resultados similares en dos campos diferentes, en este caso los mercados financieros y energéticos.

2. Regresión y predicción kernel para valores puntuales

La regresión kernel, que cuenta con un enfoque no paramétrico, calcula una forma funcional y altamente flexible conformada por los valores vecinos de la variable estudiada en lugar de los parámetros, ajustando un modelo diferente y simple por separado para cada punto observado x_0. Esto permite hacer pocas suposiciones sobre la forma, convirtiéndola en un modelo muy versátil y empleable para una enorme gama de escenarios. Es un campo maduro, especialmente para datos puntuales, con importantes investigaciones como las de Wand y Jones (1995), Hastie et al. (2009), o Chacón y Duong (2018).

El enfoque utiliza una función de estimación, también llamada estimador, que calcula una media ponderada suavizada de la muestra de datos disponibles para cada punto de interés. Este método cuenta con tres parámetros principales de ajuste: estimador, función kernel y ancho de banda. El primero tiene la capacidad de establecer el método de ponderación. Se introducen dos variantes de este parámetro. La primera es la de Nadaraya-Watson y ofrece un ajuste constante local, mientras que la segunda, el estimador lineal local, utiliza un ajuste lineal local y mejora considerablemente la predicción junto a los límites. En segundo lugar, el kernel tiene el papel de función de ponderación. Se presentan once alternativas diferentes, y cada una de ellas pondera según su contorno. Por último, el ancho de banda es un parámetro que regula el rango de influencia de los distintos datos vecinos, dictando la amplitud de la vecindad. Se introducen dos métodos de selección.

Según Vilar-Fernández y Cao (2007), la predicción kernel para datos puntuales utiliza el mismo principio de la regresión kernel. El valor se predice por su estado presente y pasado, que forman los vecinos ponderados circundantes. La principal diferencia es que para la previsión los vecinos sólo existen en un lado. Si se quieren hacer previsiones de más de un paso adelante, los valores ya predichos se añadirían a los datos de la muestra original en un proceso recursivo.

3. Regresión y predicción kernel para ITS

Como se ha observado anteriormente, el análisis de regresión modela el comportamiento de una variable de respuesta Y para una determinada variable explicativa X. Sin embargo, no todos los fenómenos del mundo real pueden representarse mediante valores de respuesta fijos únicos. La forma más sencilla de representación de datos con valores de intervalo se basa en los límites superior e inferior del intervalo. Estas dos variables se utilizan para los modelos de regresión y previsión de ITS kernel. Se presentan dos métodos.

Según Fagundes et al. (2014), el primero, denominado método de información de centro, apdlica la regresión kernel para datos puntuales a cada límite por separado. Como resultado, la regresión por intervalos se descompone en dos regresiones kernel puntuales.

El segundo, también introducido por los mismos autores, se denomina método de información de centro y rango e introduce un paso adicional al representar el intervalo con su centro y rango en lugar de los límites.

De forma similar a la previsión de datos puntuales, los modelos de regresión pueden transformarse en modelos de previsión.

4. Resultados

En consideración del rendimiento de los modelos, la combinación de funciones de núcleo de límites cerrados con el estimador lineal local mostró los niveles de precisión más altos para la regresión kernel para datos puntuales. En el lado opuesto, la previsión kernel para datos puntuales se optimizó cuando se combinaron funciones de núcleo de límite abierto con el estimador Lineal Local.

De los cuatro modelos de previsión kernel ITS, los dos basados en el método de información de centro y rango elaboraron las mejores predicciones. La combinación de este método con el estimador lineal local y las funciones de núcleo gaussiana o sigmoide mostró el mejor rendimiento en la mayoría de los casos (véase la figura 17). En escenarios especiales de tendencias laterales prolongadas, el estimador Nadaraya-Watson mostró los mejores resultados.

Table 6

Performance comparison of superior ITS forecasting models

	iARV		iUTHEIL		CR		ER	
Series	CC-NW	CC-LL	CC-NW	CC-LL	CC-NW	CC-LL	CC-NW	CC-LL
1	1.0039	1.0015	1.0150	1.0041	0.4772	0.4090	0.4310	0.4181
2	0.9868	0.9869	0.9991	0.9996	0.5770	0.5903	0.5395	0.5673
3	0.9659	0.9658	1.0131	1.0046	0.4504	0.5076	0.4268	0.4628

Note. Sigmoid and Gaussian kernel functions and optimized bandwidths were used.

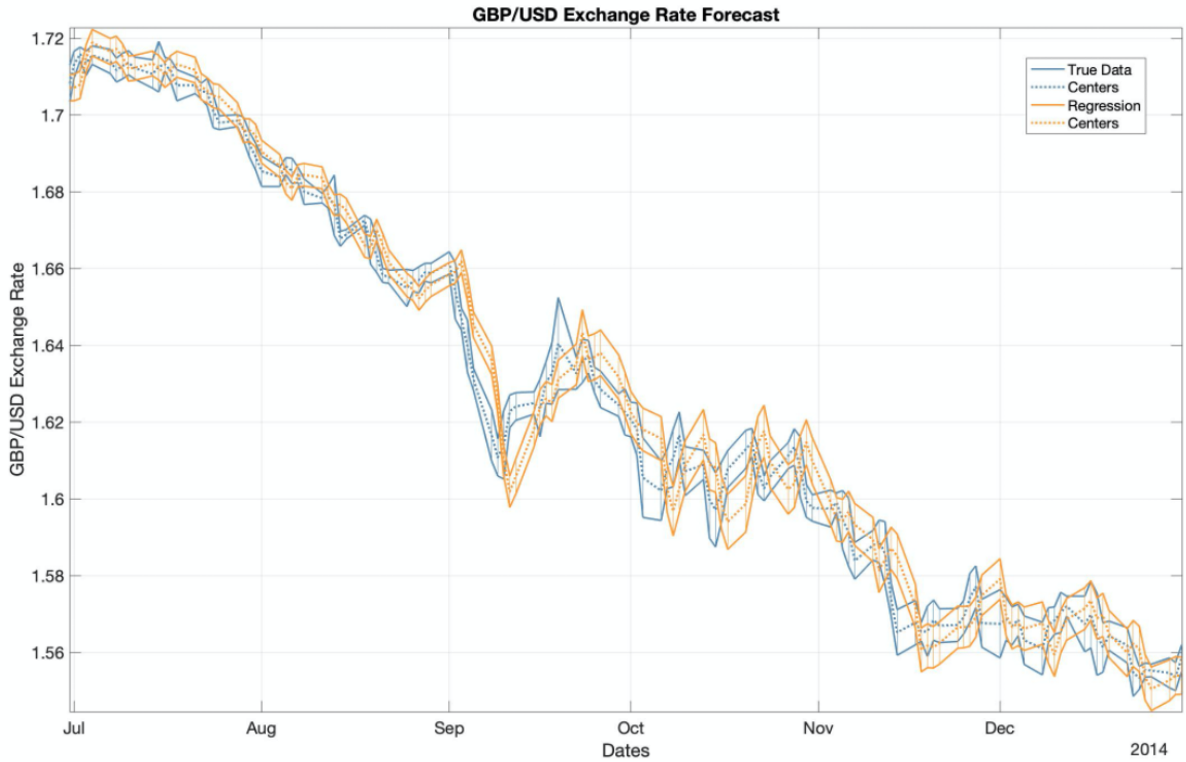


Figure 17. GBP/USD exchange rate CR-LL ITS kernel forecast.

La tabla 6 muestra algunos de los resultados del caso de estudio de los modelos de previsión ITS. Por otra parte, la figura 17 es un ejemplo del modelo de previsión ITS CR-LL que intenta predecir el tipo de cambio GBP/USD.

5. Conclusiones

En este proyecto se han introducido, probado y comparado modelos de regresión kernel y de previsión para series temporales de valores puntuales y de intervalo. Se ha conseguido una alta personalización de los modelos introduciendo y ofreciendo una diversa gama de alternativas de parámetros. Se explora por primera vez el campo de la predicción de núcleo para STI con la aplicación de dos modelos de predicción originales basados en los métodos de centro y centro y rango. Una representación de intervalos basada en los precios bajo, alto y de cierre, ha sido propuesta como suplemento de los gráficos de vela en los mercados financieros. Además, se ha llevado a cabo la implementación de todas estas funciones en MATLAB. Se han escrito un total de 6500 líneas de código distribuidas en 37 funciones, que podrían servir como base potencial para una futura herramienta visual de regresión y previsión kernel.

El modelo random walk fue superado sólo en algunos casos excepcionales, en los que el iUTHEIL era inferior a 1 (véase la Tabla 6). Como se ha visto en otros estudios de previsión no paramétrica, superar al random walk es una tarea difícil que se vuelve aún más compleja cuando se trata de series temporales con valores de intervalo. Maté y Jiménez (2021) proponen el iMLP, cuya variante para series temporales de valores puntuales muestra uno de los mejores rendimientos para la previsión del tipo de cambio euro-dólar. A pesar de ser considerablemente más complejo, este modelo también tiene dificultades al tratar de superar al random walk, lográndolo sólo en ocasiones especiales. Esto apunta a un rendimiento satisfactorio de los modelos kernel cuando se considera un equilibrio entre los resultados y la complejidad del modelo.

Las investigaciones futuras podrían centrarse en la implementación del estimador polinómico local, en un estudio más detallado del criterio de selección del ancho de banda y en la comprobación de los modelos en otras disciplinas, como la predicción meteorológica

6. Referencias

Chacón, J. E. and Duong, T. (2018), '*Multivariate Kernel Smoothing and Its Applications*' 1st edition, *Chapman and Hall/CRC.*

Fagundes, R., De Souza, R., and Cysneiros F. (2014),*' Interval kernel regression', Neurocomputing*, 128, 371-388.

Hastie, T., Tibshirani, R. and Friedman, J. (2009), *'The Elements of Statistical Learning; Data Mining, Inference and Prediction'* 2nd edition, *Springer*.

Maté, C. and Jimeńez, L. (2021), *'Forecasting exchange rates with the iMLP: New empirical insight on one multi-layer perceptron for interval time series (ITS)', Engineering Applications of Artificial Intelligence,* 104358

Vilar-Fernández, J. & Cao, R. (2007*), 'Nonparametric Forecasting in Time Series—A Comparative Study', Communications in Statistics—Simulation and Computation*®, 36:2, 311-334

Wand, M. P. and Jones, M.C. (1995), '*Kernel Smoothing*' 1st edition, *Chapman and Hall/CRC.*

.

FORECASTING INTERVAL TIME SERIES AND CANDLESTICK CHARTS. COMPARING DIFFERENT REGRESSION APPROACHES BASED ON KERNEL SMOOTHING IN FINANCIAL AND ENERGY MARKETS.

Author: de Diego Rodriguez, Alejandro Agustin
Supervisor: Mate Jimenez, Carlos
Collaborating Entity: ICAI – Comillas Pontificial University

Abstract

In this research, kernel regression and forecasting models for crisp and interval-valued time series (ITS) are introduced, tested, and compared in the financial and energy markets. A high customization of the models is achieved offering a diverse range of the three main components: estimator, kernel function, and bandwidth. The field of ITS kernel forecasting is explored for the first time with the implementation of two original forecasting models based on the center and center and range information methods. 7 time series of heavily traded currency exchange rates as well as the Brent oil are considered among others during the performance evaluation. The predictions accuracies are assessed by 9 different error measures.

Keywords: Kernel regression, Kernel forecasting, Interval-valued data, Finance, Energy, Currency Exchange

1. Introduction

Since their first primitive use in ancient Egypt, forecasting and regression methods have been an indispensable tool for organization, optimization and evolution. Within the recent progress of massive data storage in the last decades, the need for giving value to this colossal amount of information has risen. Data is the new crude oil and data analysis its refining industry.

Nonparametric regression and forecasting models play a major role where classical parametric methods fail to explain the behaviour of a variable. They are more present in the financial market due to their superior adaptability and performance. Among them, kernel regression models have caught the attention of many researchers due to their equilibrated complexity and performance balance. Plenty of research can be found for crisp kernel regression and forecasting. However, no academic literature exists for interval-valued time series kernel forecasting.

Real-world phenomena, such as weather temperatures, wind speed, or stock values, have their daily (or weekly, or monthly) bounds and vary in each period (be it a day, week, or month). Only predicting the closing price of a stock fails in transmitting different key information of high-frequency variables. In contrast, interval-valued time series allow representing important characteristics as well as the variability of these. As a result, forecasting ITS has a promising future in an immature research area.

This research has three main objectives. The first one is to introduce the unfamiliar reader to interval time series and crisp kernel regression and forecasting. The second objective is to adapt the crisp kernel theory to interval-valued data. Finally, the last objective is to test and compare the performance of the different crisp and interval-valued kernel regression and forecasting models in the financial and energy markets with current data. The third objective also aims to investigate whether a promising prediction model can have similar results in two different fields, in this case the financial and energy markets.

2. Crisp kernel regression and forecasting

Kernel regression, which is a nonparametric approach, calculates a widely flexible functional form shaped by neighbour values of the studied variable rather than parameters, fitting a different but simple model separately at each observed point x_0. This enables to make little assumption about the form, converting it into a highly versatile and employable model for a huge range of scenarios. It is a mature field, especially for crisp data, with important research such as Wand and Jones (1995), Hastie et al. (2009), or Chacón and Duong (2018).

The approach uses an estimating function, also called estimator, which calculates a smooth weighted average of the available data sample for each point of interest. This method counts with three main tuning parameters: estimator, kernel function, and bandwidth. The first one has the ability to establish the weighting method. Two variants of this parameter are introduced. The first one is the Nadaraya-Watson and offers a local constant fit, whereas the second one, the Local Linear estimator, uses a local linear fit and considerably improves the prediction next to the boundaries. Secondly, the kernel has the role of weighting function. Eleven different alternatives are presented, and each of them weights according to its weighting shape. Finally, the bandwidth is a parameter that regulates the influence range of the different weights, dictating the width of the neighbourhood. Two selection methods are introduced.

Following Vilar-Fernández and Cao (2007), crisp kernel forecasting uses the same principle of kernel regression. The value is predicted by its present and past state, which are the surrounding weighted neighbours. The main difference is that for forecasting, neighbours only exist on one side. If forecasts of more than one step ahead want to be done, the already predicted values would be added to the original sample data in a recursive process.

3. ITS kernel regression and forecasting

As observed above, regression analysis models the behaviour of a response variable Y for a given explanatory variable X. Nevertheless, not all real-world phenomena can be represented by single fixed response values. The simplest form of interval-valued data representation is based on the upper and lower interval boundaries. These two variables are used for the ITS kernel regression and forecasting models. Two methods are presented.

Following Fagundes et al. (2014), the first one called center information method applies crisp data kernel regression to each boundary separately. As a result, the interval-valued regression is decomposed into two singular kernel regressions.

The second one, also introduced by the same authors, is called the center and range information method and introduces an additional step by representing the interval with its center and range instead of the boundaries.

In a similar way to crisp data forecasting, the regression models can be transformed into forecasting ones.

4. Results

Talking about the performance of the models, the mixture of closed boundary kernel functions with the Local Linear estimator showed the highest accuracy levels for crisp kernel regression. On the opposite side, crisp kernel forecasting was optimized when combining open boundary kernel functions with the Local Linear estimator.

Out of the four ITS kernel forecasting models, the two based on the center and range information method elaborated the best predictions. The combination of this method with the Local Linear estimator and the Gaussian or Sigmoid kernel functions showed the best performance in most of the cases (see Figure 17). In special scenarios of prolonged side trends, the Nadaraya-Watson estimator showed better results.

Table shows some of the results of the ITS forecasting models case study. On the other hand, figure 17 is an example of the ITS CR-LL forecasting model trying to predict the GBP/USD exchange rate.

Table 6

Performance comparison of superior ITS forecasting models

	iARV		iUTHEIL		CR		ER	
Series	CC-NW	CC-LL	CC-NW	CC-LL	CC-NW	CC-LL	CC-NW	CC-LL
1	1.0039	1.0015	1.0150	1.0041	0.4772	0.4090	0.4310	0.4181
2	0.9868	0.9869	0.9991	0.9996	0.5770	0.5903	0.5395	0.5673
3	0.9659	0.9658	1.0131	1.0046	0.4504	0.5076	0.4268	0.4628

Note. Sigmoid and Gaussian kernel functions and optimized bandwidths were used.

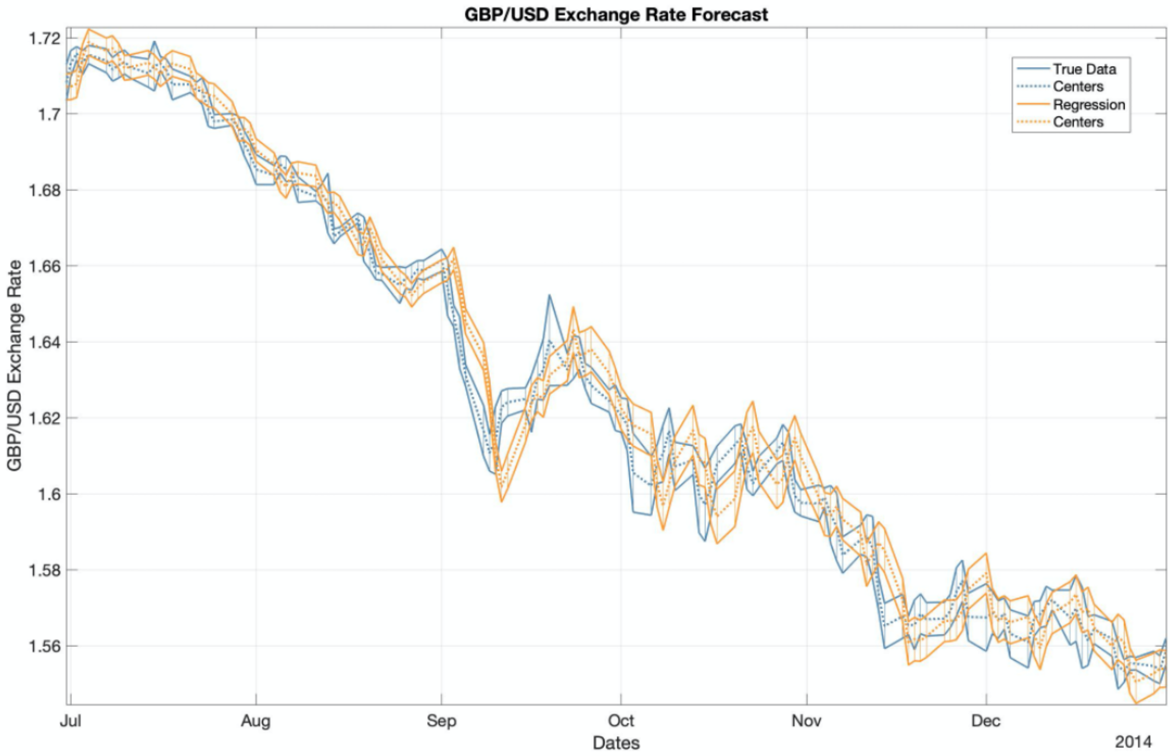


Figure 17. GBP/USD exchange rate CR-LL ITS kernel forecast.

5. Conclusions

In this research, kernel regression and forecasting models for crisp and interval-valued time series have been introduced, tested, and compared. A high customization of the models has been achieved by introducing and offering a diverse range of parameters' alternatives. The field of ITS kernel forecasting is explored for the first time with the implementation of two original forecasting models based on the center and center and range information methods. An interval representation based on the high, low and closing prices has been proposed as a supplement of the candlestick charts in the financial markets. The implementation of all these features to MATLAB has been carried out. A total of 6500 lines of code distributed in 37 scripts have been written, serving as a potential foundation of a future kernel regression and forecasting visual tool.

The random walk was beaten only in some exceptional cases, where the iUTHEIL was lower than 1 (see Table 6). As seen in other nonparametric forecasting studies, to outperform the random walk is a difficult task which becomes even more complex when dealing with interval-valued time series. Maté and Jiménez (2021) proposed the iMLP, which variant for crisp time series shows one of the best performances for the Euro US Dollar exchange rate forecast. Despite of being considerably more complex, this model also suffered when trying to outperform the random walk, only achieving it in special occasions. This points towards a satisfactory performance of the kernel models when considering a balance between results and complexity.

Future research could be focused on the implementation of the local polynomial estimator, a more detailed study of the bandwidth selection criterion, and a testing of the model in other disciplines such as weather forecasting.

6. Bibliography

Chacón, J. E. and Duong, T. (2018), '*Multivariate Kernel Smoothing and Its Applications*' 1st edition, *Chapman and Hall/CRC.*

Fagundes, R., De Souza, R., and Cysneiros F. (2014),*' Interval kernel regression', Neurocomputing*, 128, 371-388.

Hastie, T., Tibshirani, R. and Friedman, J. (2009), *'The Elements of Statistical Learning; Data Mining, Inference and Prediction'* 2nd edition, *Springer*.

Maté, C. and Jimeńez, L. (2021), *'Forecasting exchange rates with the iMLP: New empirical insight on one multi-layer perceptron for interval time series (ITS)', Engineering Applications of Artificial Intelligence,* 104358

Vilar-Fernández, J. & Cao, R. (2007*), 'Nonparametric Forecasting in Time Series—A Comparative Study', Communications in Statistics—Simulation and Computation*®, 36:2, 311-334

Wand, M. P. and Jones, M.C. (1995), '*Kernel Smoothing*' 1st edition, *Chapman and Hall/CRC.*

Table of contents

List of figures

List of tables

CHAPTER 1. INTRODUCTION

1.1 BACKGROUND AND MOTIVATION

Since their first primitive use in ancient Egypt, forecasting and regression methods have been an indispensable tool for organization, optimization and evolution. At that time, the water level of the Nile River during the flooding season was used as a predictor to harvest. Within the recent progress of massive data storage in the last decades, the need for giving value to this colossal amount of information has risen. Forecasting models have experienced a remarkable expansion due to their outstanding capacity to establish relationships and predict future scenarios in a huge variety of fields. Countless types of prediction models for different scenarios have been developed. Among all, forecasting the financial market is a major challenge in both the academic and professional world. Due to its noisy nature, financial time series are among the most difficult signals to forecast, which naturally leads to the debate on market predictability among academics and market traders. Furthermore, their interval-valued nature adds another complexity layer to their study.

Nonparametric regression and forecasting models play a major role where classical parametric methods fail to explain the behaviour of a variable. They are more present in the financial market due to their superior adaptability and performance. Among them, kernel regression models have caught the attention of many researchers due to their equilibrated complexity and performance balance. Plenty of research can be found for crisp kernel regression, and some for crisp kernel forecasting. Some papers about kernel regression for interval-valued time series are also available. However, no academic literature exists for interval-valued time series kernel forecasting. Therefore, work still needs to be done to further comprehend and test this approach.

The main motivation of the project comes from the increasing weight of data analysis and prediction in the academic and professional world. Data is the new crude oil and data analysis its refining industry. The second motive is linked to the nature of daily observations. Real-world phenomena, such as weather temperatures, wind speed, or stock values, have their daily (or weekly, or monthly) bounds and vary in each period (be it a day, week, or month). Only predicting the closing price of a stock fails in transmitting different key information of high-frequency variables. In contrast, interval time series allow representing important characteristics as well as the variability of these. As a result, forecasting ITS has a promising future in an immature research area. The third and last motive focuses on the financial aspect. The sudden rise in stock interactions density throughout the last years due to the increasing involvement of small-budget investors, who now have more practical and intuitive access to the trading world, have risen the volatility and risk of the market. This creates an urgent need for more accurate models. Developing more precise forecasts would enable faster market reactions, hence more control over the situation.

1.2 OBJECTIVES

This research has three main objectives. The first one is to introduce the unfamiliar reader to interval time series and crisp kernel regression and forecasting. The second objective is to adapt the crisp kernel theory to interval-valued data. Finally, the last objective is to test and compare the performance of the different crisp and interval-valued kernel regression and forecasting models in the financial and energy markets with current data. The third objective also aims to investigate whether a promising prediction model can have similar results in two different fields, in this case the financial and energy markets.

1.3 STRUCTURE

The first chapter, which contains this section, forms an introduction where the background, motivation, objectives, structure, methodology, and resources of this study are presented. Afterwards, the kernel regression and forecasting methodology for crisp time series is introduced in the second chapter. The main tuning parameters of this methodology, estimator, function, and bandwidth, are explained in detail. The third chapter is based on an adaptation of the previous theory to interval-valued time series. The second and third chapters will be introduced with a short state of the art description. In the fourth chapter, a case study testing and comparing the performance of the different models is contained. Finally, some conclusions and guidelines for future research are exposed in the fifth chapter.

1.4 METHODOLOGY

As part of the research's nature, which is analytical and experimental, the different models will be firstly formulated and then tested and compared. Current financial and energy market data will be needed. The data will be gathered in Excel and imported for analysis into MATLAB, which will be used to code and test the different models. In the case study, the different tuning parameters will be compared to select the optimal ones and use them in the performance comparison of the models. The models will have to work under different conditions represented by a diverse variety of financial and energy time series.

1.5 RESOURCES

Due to the project's nature, only a few resources and programs will be required. On the one hand, academic articles and current data of financial and energy markets will be necessary. This data is usually open on the internet and no problems should appear when gathering it. Access to academic papers is becoming easier over time. There may be no other option than buying an academic research paper when needed, but most of them will be available. On the other hand, three main computer programs will be needed; MATLAB, Excel and Word, and all of them are provided by the university. Throughout the development of the project, the need for additional programs may arise. A computer will be the only physical resource needed. The project will be mainly assembled and coded on a private laptop. In case of insufficient processing power, another computer may be used.

1.6 ODS CONTRIBUTION GOALS

The sustainable development goals are the fundamental pillars that will sustain our future society. All research should be directed towards the fulfilment of these objectives and so does this one. Some of the goals are more present than others in the project. They are gathered by topic resemblance.

First of all, the outcome of this research will be involved with the goals of quality education, innovation and partnerships (4, 9 and 17 respectively). The research will broaden the academic boundaries and offer more information about the discussed topic. This will allow transmitting more knowledge to future generations. Furthermore, innovation is a keystone in a final degree project. The final outcome will be an original and innovative system for forecasting financial and energy markets. Testing the multidisciplinary skills of the model is another original research area. When it comes to partnership, the project will be sustained by past academic papers and will contribute to future research. The whole academic network is a global partnership program with the aim of pushing human knowledge further.

Moreover, the project outcome will try to contribute to affordable and clean energy production (7), as the forecasting research and its multidisciplinary approach could help in the field of weather prediction. More accurate models allow a wider use of renewable energy sources, which volatility is a considerable impediment still today. In a similar way, the research may help with responsible production (12), as better prediction models reduce the surplus. By contributing to the two goals mentioned, the research would also help indirectly in the climate action (13).

Important to remember is the initial purpose of the project to have a meaningful value, but most important is the final use of the results.

CHAPTER 2. STATE OF THE ART

Kernel smoothing is a mature field with important research such as Wand and Jones (1995) or Chacón and Duong (2018). The use of Kernel regression in financial market forecasting has been proved to outperform various previous models. For example, Wang and Zhu (2008) propose a two-step kernel learning method based on the support vector regression (SVR) which learns a sparse linear combination of different kernels to optimize its performance. Contrary to the final aim of this research, they focus on individual variables forecasting, lacking the broader perspective given by interval time series. The model shows promising results surpassing the cumulative log return of other models, i.e., buy-and-hold or SVR when tested in the S&P500 index among others.

According to the actual knowledge, there are several methodologies for ITS forecasting. These can be divided into two main categories regarding how data is handled. In the first one, individual variables of an ITS, i.e., upper or lower boundary, are treated individually using single-valued methods. For the second category, all variables of an interval are treated as symbolic data using interval arithmetic. Symbolic variables are thought better suited for describing complex real-life situations (Billard and Diday 2003, Billard and Diday 2006 and Noirhomme-Fraiture and Brito 2011). Some of these methods are exponential smoothing, the kNN-algorithm and the multilayer perceptron (Arroyo et al. 2011). More recently, Tao et al. (2015) investigated the possibility of forecasting interval time series by denoting the lower and upper bounds of the interval as real and imaginary parts of a complex number respectively. Maceda (2018) presents a wide perspective of eight commonly used linear parametric regression models for ITS prediction and compares their performance. One of his conclusions is that in terms of accuracy, the more complex and advanced the model, the slightly better results, but sacrificing computational power.

Although some of these methodologies had promising results, a change of perspective has taken place over the last years. Instead of developing more complex models, new research points towards combining simpler forecasts, which has been proved to reduce the final forecasting error as shown in some forecasting competitions such as M4 (see Makridakis et al. 2020). Maté (2021) summarizes the existing literature concerning this new approach and outlines several open lines of research. Compared to forecasts combination in classical statistics, a mature field with key papers such as Bates and Granger (1969) and Clemen (1989), forecasts combination for ITS is a young research area with a promising future. The improved numerical accuracy of combining suggests continuing in this direction.

At the beginning of the following chapters, additional information about the state of the art of the developed fields will be introduced. This was decided to make the reading and understanding of the development more fluent and easier to follow.

CHAPTER 3. CRISP KERNEL REGRESSION AND FORECASTING

3.1 INTRODUCTION TO KERNEL REGRESSION

Estimating a number of parameters and using them to predict is a powerful method when there is enough information available. Unfortunately, a lack of complete information about the functional form and behaviour of many real-world phenomena causes misspecifications leading to inaccurate and biased predictions. One clear example is the financial world. Its uncertainty and volatility originate a need for constant creation of new parametric regression models as the influence of their parameters changes over time.

Kernel regression, which is a nonparametric approach, avoids this situation by calculating a widely flexible functional form shaped by neighbour values of the studied variable rather than parameters, fitting a different but simple model separately at each observed point x_0. This enables to make little assumption about the form, converting it into a highly versatile and employable model for a huge range of scenarios. It is a mature field, especially for crisp data, with important research such as Wand and Jones (1995), Hastie et al. (2009), or Chacón and Duong (2018).

The approach uses an estimating function $\hat{m}$, also called estimator, which calculates a smooth weighted average of the available data sample $(X, Y) \in \mathbb{R}^d \times \mathbb{R}$ for each point of interest. This method counts with three main tuning parameters: estimator, kernel function, and bandwidth. Each of them can considerably change the regression outcome. The first one has the ability to establish the weighting method. Secondly, the kernel has the role of weighting function. Finally, the bandwidth is a parameter that regulates the influence range of the different weights, dictating the width of the neighbourhood. A brief formulation and description of each of them will follow to introduce the unfamiliar reader into the method.

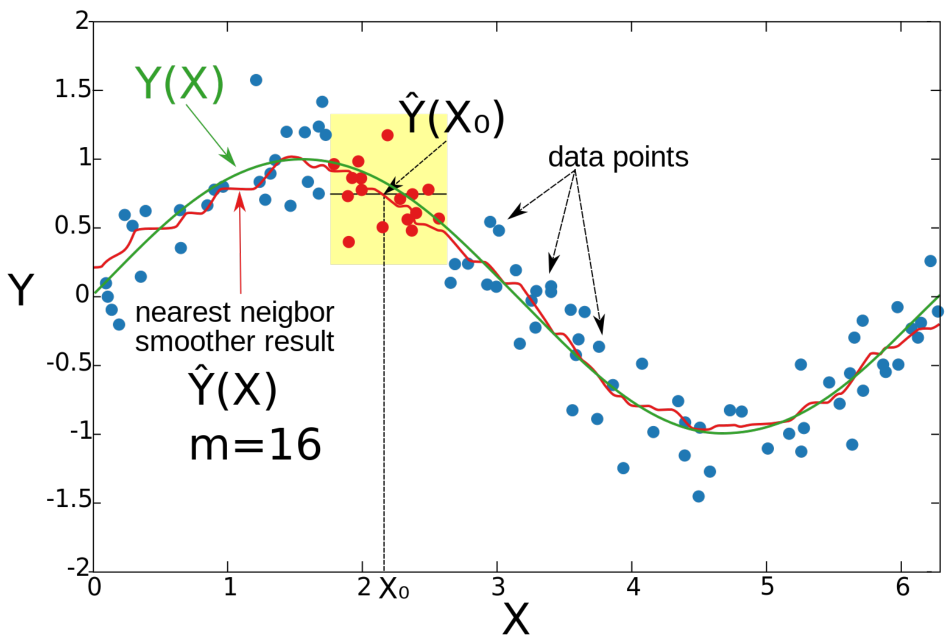


Figure 1. Example of Kernel regression.

In Figure 1, a first example of kernel regression can be observed. The red line represents the regression whereas the green line is the true function. Furthermore, $\hat{Y}(x)$ is the estimating function which calculates an average value of the 16 nearest neighbours as estimation for each x_0. The number of the influencing neighbours is controlled by a mixture of the kernel function and the bandwidth. In this case, an overfitted result can be seen.

3.2 KERNEL ESTIMATORS

3.2.1 NADARAYA-WATSON ESTIMATOR

Following the idea of a moving average, this estimator predicts x_0 based on the value y_i of its neighbours x_i and the distance between them. Let us consider a random pair $(X, Y) \in \mathbb{R}^d \times \mathbb{R}$ and an estimator

$$m_0(x) = \mathbb{E}(Y|X = x)$$

which is the regression function of Y on X. The basic objective is to construct an estimate $\hat{m}$ of m_0, from i.i.d. samples $(x_1, y_1), \dots (x_n, y_n) \in \mathbb{R}^d \times \mathbb{R}$ that have the same joint distribution as (X, Y). In this research, the univariate condition is considered, therefore $d = 1$. The variable X is often called predictor, input, or feature, whereas Y is the output, response, or outcome. In the specific case of time series, X is always a time interval, i.e., seconds, hours, or weeks. Important to remember is that in the nonparametric approach a certain parametric form of m_0 is not assumed. Due to its definition, $m_0(x)$ can be rewritten as

$$m_0(x) = \int y f_{Y|X=x}(y)\, dy$$

$$= \frac{\int y f(x, y)\, dy}{f_X(x)}$$

The joint and marginal densities can be replaced by kernel density estimators

$$f(x, y) = \hat{f}(x, y, h) = \frac{1}{n} \sum_{i=1}^{n} K_{h_1}\left(\frac{x - x_i}{h}\right) K_{h_2}\left(\frac{y - y_i}{h}\right)$$

$$f_X(x) = \hat{f}_X(x, h_1) = \frac{1}{n}\sum_{i=1}^{n} K_{h_1}(\frac{x - x_i}{h})$$

After replacing them in m_0 and simplifying, the Nadaraya-Watson estimator is formed

$$\widehat{m_0}(x, h) = \sum\nolimits_{i=1}^{n} \frac{K_h\left(\frac{x - x_i}{h}\right)}{\sum_{i=1}^{n} K_h\left(\frac{x - x_i}{h}\right)} y_i \, , \qquad (1)$$

where K_h is the kernel function, and h the bandwidth. The weights are defined as

$$W_i(x) = \frac{K_h\left(\frac{x - x_i}{h}\right)}{\sum_{i=1}^{n} K_h\left(\frac{x - x_i}{h}\right)} .$$

This estimator can be considered a weighted average of $y_1, y_2, \dots y_n$ by means of the set of weights $w_1, w_2, \dots\ w_n$, which add to one. The varying weights depend on the evaluation point x_0, converting the Nadaraya-Watson estimator into a local mean of Y about $X = x_0$.

Figure 2 illustrates a second kernel regression example using the Nadaraya-Watson estimator introduced by the equation (1) combined with the Epanechnikov kernel function. The function $\hat{f}(x_0)$ stands for the estimator $\widehat{m}(x_0)$. The yellow shape reflects the weight assigned to observations following the shape of the Epanechnikov kernel. Data points with a minimum influence predicting x_0 are red coloured. The green curve is the kernel-weighted average.

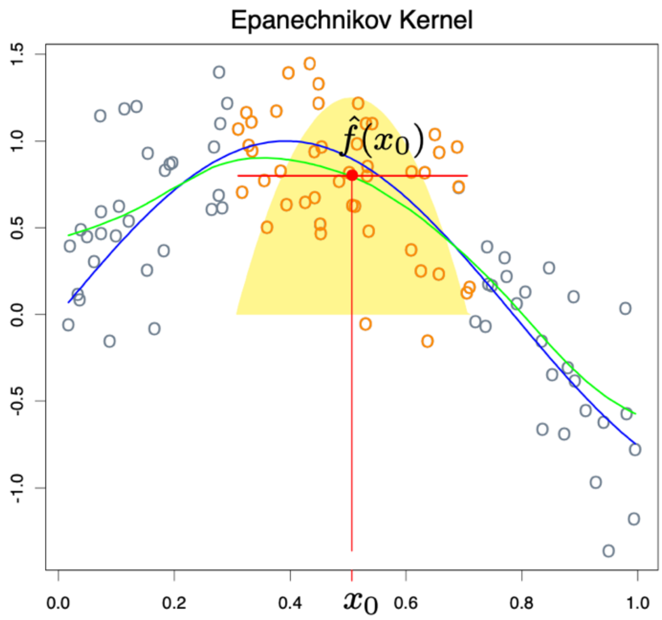


Figure 2. Example of the Nadaraya-Watson estimator [Hastie et al. (2009)].

3.2.2 LOCAL LINEAR ESTIMATOR

The Nadaraya-Watson estimator, a locally-weighted average, is a powerful and flexible regression tool, but suffers from poor bias at the boundaries. Figure 2, where the blue curve represents the true function and the green curve illustrates the kernel-weighted average, shows this behaviour in the extremes. The reason behind is the asymmetry of the kernel in that region. In the case of the left extreme, there are many neighbours on the right side but none on the opposite. Therefore, the predicted value lacks of the weight of its left-sided neighbours. Moreover, this bias can be present in the interior of the domain if the data is not equally spaced, creating areas of higher information density which overweight in a similar manner the prediction.

Local linear regression offers a solution by moving from a local constant fit to a local linear fit. This is done by solving an additional weighted least squares problem at each target point x_0

$$\min_{\alpha(x),\beta(x)} \sum_{i=1}^{N} K_h(\frac{x-x_i}{h})[y_i - \alpha(x) - \beta(x)x_i]^2 \,.$$

Therefore, the estimate has the form

$$\hat{m}(x) = \hat{\alpha}(x) + \hat{\beta}(x)x_0 \,,$$

and can be rewritten as

$$\hat{m}(x) = b(x)^T(\boldsymbol{B}^T\boldsymbol{W}(x)\boldsymbol{B})^{-1}\boldsymbol{B}^T\boldsymbol{W}(x)y \qquad (2)$$

$$= \sum_{i=1}^{N} l_i(x)y_i \,,$$

where $b(x) = (1, x_i)$, B is a $n \times 2$ matrix with ith row $b(x_i) = (1, x)$, and Ω is a $n \times n$ diagonal matrix with ith diagonal element $K(\frac{x-x_i}{h})$. The weights l_i combine the weighting kernel K_h and the least squares operations.

Looking closer to Figure 3, the differences between the two estimators can be observed. The green estimated line adjusts better to the blue true function. The behaviour in the boundaries is considerably improved by implementing the local linear estimator represented by equation (2).

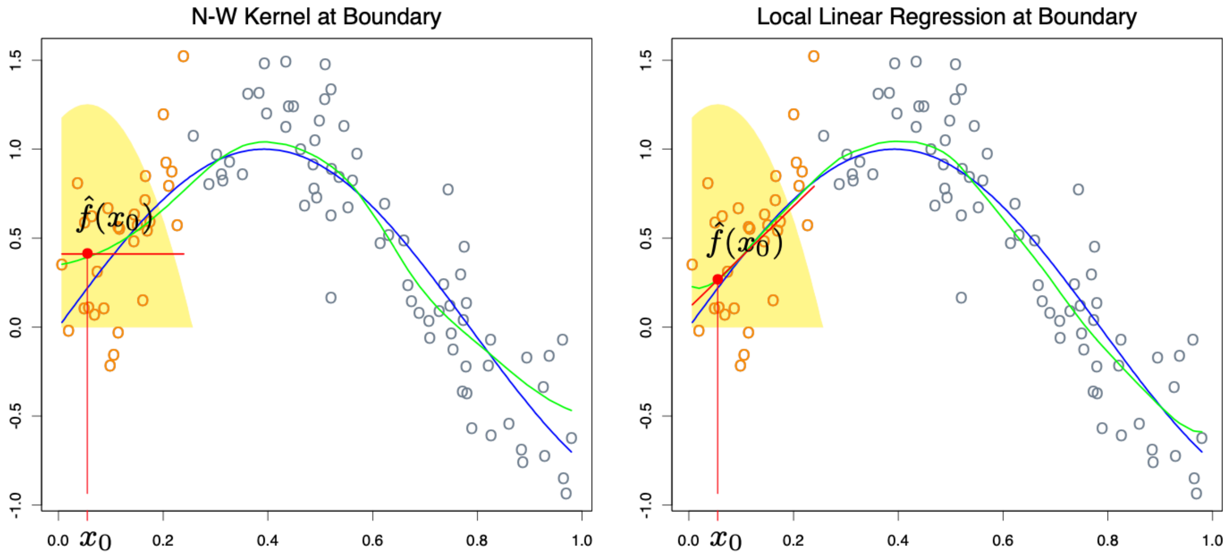


Figure 3. Comparison of the Nadaraya-Watson and Local Linear estimators [Hastie et al. (2009)].

3.3 KERNEL FUNCTIONS

The Kernel function is the second main tuning parameter of Kernel regression. This function K is a k-multivariate probability function which is continuous, symmetric, and integrable which integrates to 1. Therefore, it satisfies

$$K(-u) = K(u), \quad \int K(u)\,du = 1, \quad \int uK(u)\,du = 0, \quad 0 < \int u^2 K(u)\,du < \infty \,.$$

There is a considerable variety of them available and each function weights differently in their area of influence. In this section, eleven commonly used kernel functions will be briefly introduced and a general criterion for choosing the most adequate one will follow.

3.3.1 GAUSSIAN KERNEL

Being one of the most commonly used kernel functions, the Gaussian kernel is defined by

$$K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}$$

This function applies the idea of giving large weights to x_i close to x following the shape of a gaussian distribution.

3.3.2 EPANECHNIKOV KERNEL

The second function is the Epanechnikov Kernel which is defined by

$$K(u) = \frac{3}{4}(1-u^2), \quad |u| \le 1$$

This one is also a commonly used function due to its parabolic shape and weights akin to the Gaussian kernel, but establishes influence boundaries.

3.3.3 QUARTIC KERNEL

The third one is the Quartic Kernel and follows the form

$$K(u) = \tfrac{15}{16}(1-u^2)^2, \quad |u| \le 1$$

Its shape is also parabolic but narrower, concentrating more weight in the closest neighbours.

3.3.4 TRIWEIGHT KERNEL

The Triweight Kernel follows the equation

$$K(u) = \tfrac{35}{32}(1-u^2)^3, \quad |u| \le 1$$

Compared to the Quartic, this one has a narrower and sharper shape gathering more weight in the centre.

3.3.5 TRICUBE KERNEL

The fifth is the Tricube Kernel which is defined by

$$K(u) = \frac{70}{81}(1 - |u|^3)^3, \quad |u| \leq 1$$

It follows a similar shape compared to the Quartic but with a flatter tip, which distributes the weights slightly more to the sides.

3.3.6 COSINE KERNEL

The sixth is the Cosine Kernel which is defined by

$$K(u) = \frac{\pi}{4}\cos(\frac{\pi}{2}u), \quad |u| \leq 1$$

The weights of this kernel are similarly distributed and follow the shape of a cosine function.

3.3.7 LOGISTIC KERNEL

This function follows the equation

$$K(u) = \frac{1}{e^u + 2 + e^{-u}}$$

The Logistic kernel spreads horizontally and distributes weight to further neighbours extending its area of influence. It is the flattest out of all functions.

3.3.8 SIGMOID KERNEL

The eighth kernel is defined by

$$K(u) = \frac{2}{\pi}\frac{1}{e^u + e^{-u}}$$

Its shape is akin to the Logistic Kernel extending smoothly along the horizontal axis, but gathering, to a small degree, more weight close to the y axis.

3.3.9 SILVERMAN KERNEL

The Silverman Kernel is defined by

$$K(u) = \frac{1}{2} e^{-\frac{|u|}{\sqrt{2}}} \sin\left(\frac{|u|}{\sqrt{2}} + \frac{\pi}{4}\right)$$

Although it spreads as far as the last mentioned, this function has the peculiarity of assigning negative weights to the furthest influential neighbours, and higher weights to the closest ones.

3.3.10 TRIANGULAR KERNEL

The Triangular kernel follows the equation

$$K(u) = (1 - |u|), \quad |u| \leq 1$$

This is the first non-parabolic mentioned and has a triangular shape, extending rougher along the horizontal axis and concentrating more influence in the centre. x_1

3.3.11 UNIFORM KERNEL

The last one is the Uniform Kernel which is defined by

$$K(u) = \frac{1}{2}, \quad |u| \leq 1 \,.$$

It spreads horizontally giving the same weight to all the neighbours in the area of influence. Its nature makes it similar to the k-nearest-neighbours method.

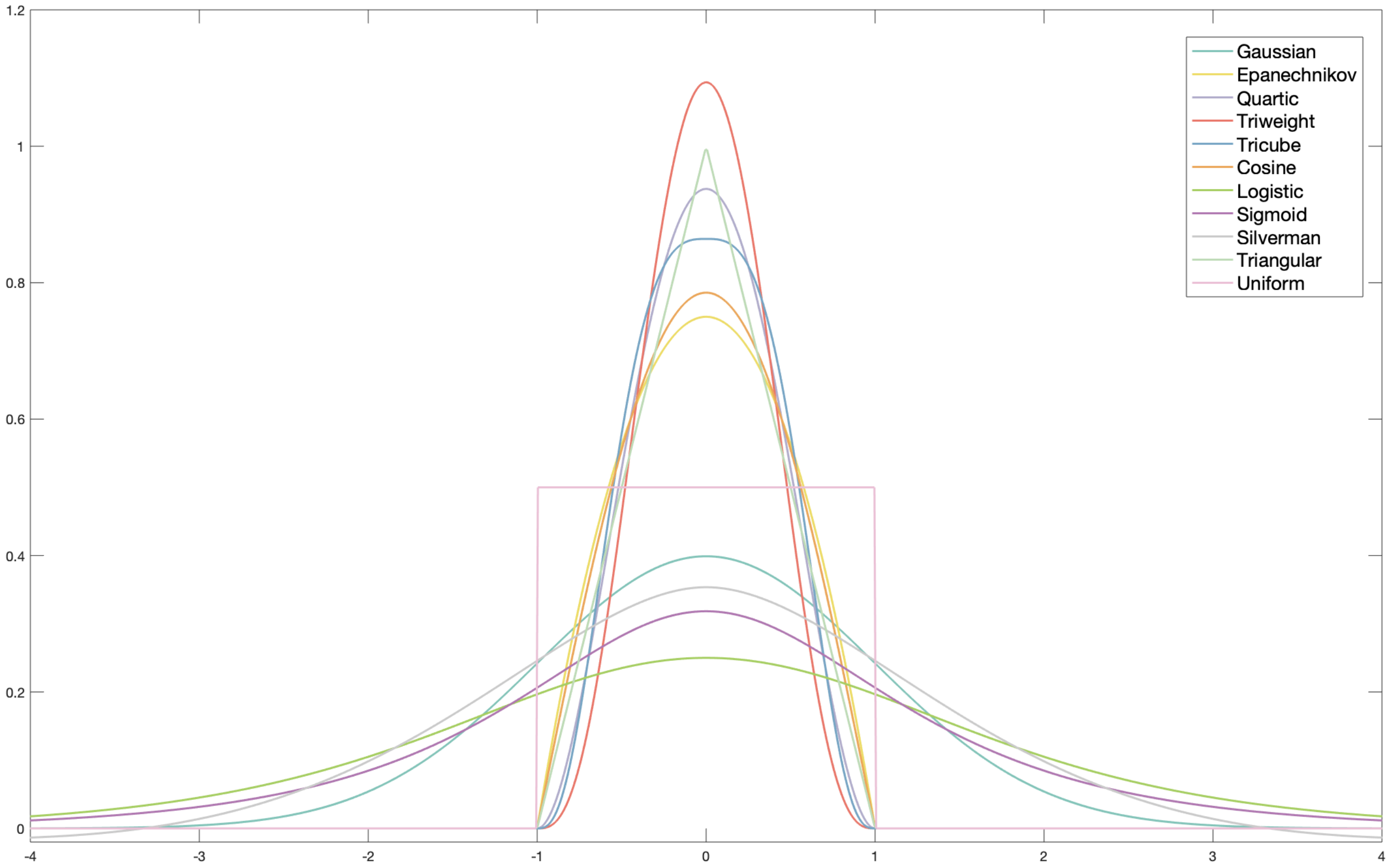


Figure 4. Comparison of the above enumerated kernel functions in a common coordinate system.

3.3.12 KERNEL FUNCTION'S SELECTION CRITERION

An accurate selection of the kernel function needs to take into account the other tuning parameters of the regression. Figure 4 shows a shape comparison of the enumerated kernel functions in a common coordinate system. For example, if constant estimator and bandwidth are applied, a sharper kernel gathering more weight close to X_n will affect the model akin to selecting a smaller bandwidth, creating a rougher regression model. On the other hand, if a flatter one is selected, the model will follow a smoother regression. Figure 5 illustrates this behaviour by comparing two different regression models, both with the same bandwidth and estimator. The first one applies the Triweight kernel function whereas the second one uses the Logistic function. A Time Series of the Abbott Laboratories (ABT) daily closing price in the NYSE stock market during January 2019 was used.

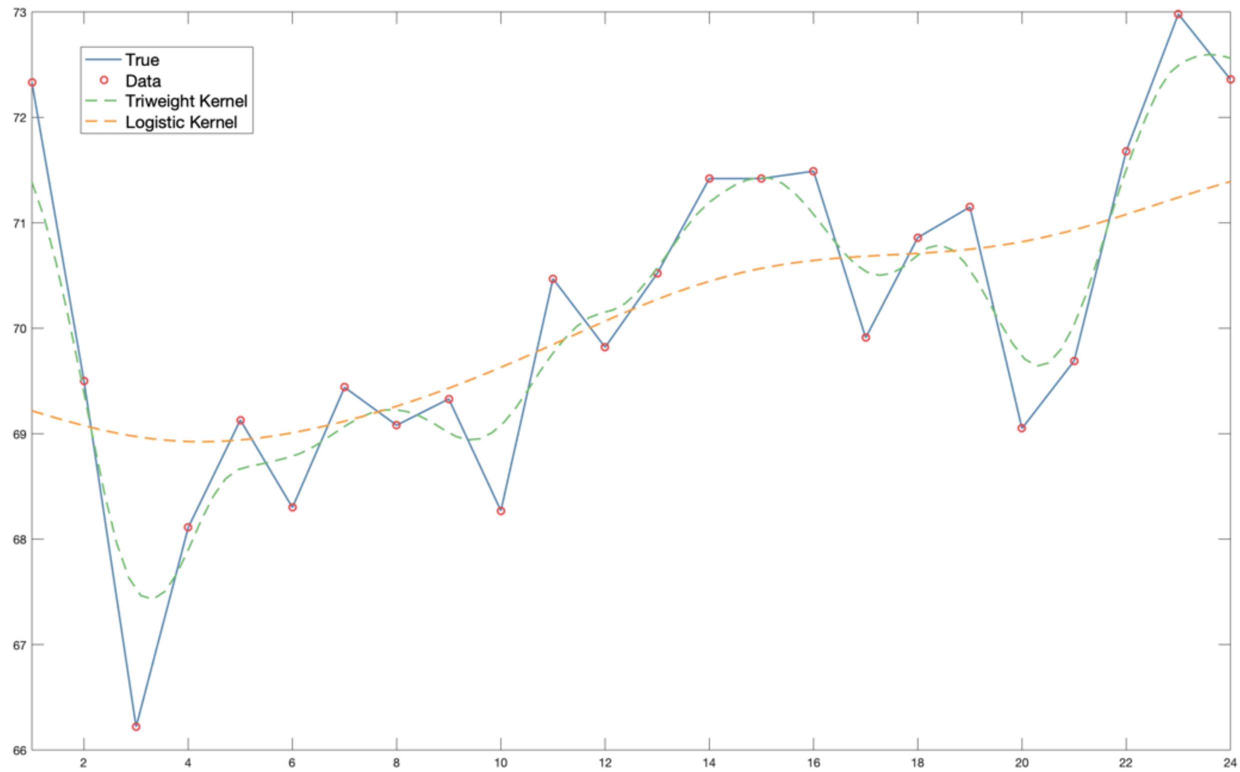


Figure 5. Regression comparison of two kernel functions.

A balance must be found to not overfit neither underfit the model, and the kernel function selection is an important step of the process. Compared to linear regression models, the overfitting caused by kernel regression does not rise the computational time or cost, as the calculations complexity remains equal. On the other side, overfitting may decrease the regression error in the training set, but increase it in the test sets. Being aware of the kernel function influence in the final outcome, its selection relies on the model designer. In this thesis, the minimization of error measures will be applied in order to decide which is the most adequate kernel function for different situations.

3.4 KERNEL BANDWIDTH

The last tunning parameter is the kernel bandwidth h. It has the ability to extend the neighbourhood, and, consequently, the influence area of each kernel. The selection of the smoothing parameter is a fundamental task throughout the design process, and in some cases, especially among practitioners, it is overseen, and set to a default value.

The choice of the bandwidth is a process involving bias and variance. When choosing a narrow neighbourhood, small h value, the estimation becomes a weighted average of a small number of values y_i close to x_0, increasing the variance, but reducing the bias. In contrast, by selecting a wider neighbourhood, the bias rises and the variance decreases. This is because more distant observations, which may be completely different to the estimation, have a bigger influence.

Vast research has been published concerning this issue. Bowman and Azzalini (1997) suggest an optimal bandwidth which takes the standard deviation and the number of observations into account. Köhler et al. (2014) revise and compare 20 different selection methods for univariate distributions in an extensive simulation study. Being kernel bandwidth selection a wide and complex field, its analysis goes beyond the scope of this research. Therefore, only two different approaches to select the optimal bandwidth will be mentioned: Bowman and Azzalini and error measure optimization.

3.4.1 BOWMAN AND AZZALINI'S OPTIMAL BANDWIDTH

As mentioned before, this first methodology calculates the optimal bandwidth regression based on the standard deviation and the number of observations. Following Bowman and Azzalini's, the optimal bandwidth selection is found by

$$h_x = \frac{|x - \bar{x}|}{0.6745 \times (\frac{4}{3 \times n})^{\frac{1}{5}}}$$

$$h_y = \frac{|y - \bar{y}|}{0.6745 \times (\frac{4}{3 \times n})^{\frac{1}{5}}}$$

$$h = \sqrt{h_y \times h_x}\,.$$

This method offers good bandwidth values when the sample size is big enough and has a minimum variance. However, if the data values are too small, and consequently their variances, the method may return inoperative values. For the kernel functions with boundaries, the lack of enough data and its resultant small bandwidth value may even cause the unproper working of the model.

3.4.2 ERROR MEASURE OPTIMAL BANDWIDTH

This simple method is based on an iteration to find the minimum error measures by trying different bandwidth values. A code was written to automatize the process. Compared to the other method, this one does not rely on sample size neither does it on variance or value size. Therefore, it becomes a more polyvalent selection method. There is a need for a training set, as well as in the other one. The error to be minimized can be freely chosen. In this study, MSE and RMSSE have been selected as main errors to optimize the bandwidth value. The only disadvantage offered by this method may be the computational cost and time when the sample size is large.

3.5 CRISP TIME SERIES KERNEL FORECASTING

After introducing the unfamiliar reader to kernel regression, it is time to approach the use of kernel regression as a forecasting tool. Creating Kernel regression models is a mature field, but forecasting future values using them is a less explored path. One may think that the process of calculating a function to analyse the behaviour of a variable is similar to the parametric model's approach, where a data sample is used to find out whether there is a correlation between at least two variables or not. If a correlation is verified, the data values are used to calculate the parameters of the model's function. Finally, the function can be used to predict future behaviours of the output regarding the predictive variable. Billard and Diday (2000), pioneers in linear regressions models for interval-valued data, introduced this methodology for crisp data for the first time and illustrated an example from cardiology. This approach is useful for parametric models, but nonparametric forecasting is unable to apply the same simple method due to its own nature.

3.5.1 FORMULATION OF THE PROBLEM

As mentioned before, the main idea behind Kernel regression is to calculate a weighted average of the data sample for the prediction. Following Vilar-Fernández and Cao (2007), let us consider a stationary and univariate time series Z_t observed on the time interval $1 \leq t \leq n$. Z_t follows a Markov process, which means that the predictions can be made based on its present state

$$Z_{t+1} = m\left(Z_t, Z_{t-1}, Z_{t-2}, \dots, Z_{t-p}\right) + \varepsilon_{t+1} \tag{3}$$

where m denotes the weighting function, the estimator, p denotes the number of past time steps influencing the forecast, and ε represents the forecasting error. Similar to the nonparametric regression approach, the value is predicted by the surrounding weighted neighbours. The main difference is based on the fact that for forecasting, neighbours only exist on one side.

If forecasts of more than one step ahead want to be done, the already predicted values would be added to the original sample data

$$Z_{t+2} = m\left(Z_{t+1}, Z_t, Z_{t-1}, Z_{t-2}, \dots, Z_{t-p}\right) + \varepsilon_{t+2}. \tag{4}$$

This process can be recursively repeated until the horizon prediction is reached. Important to remember is that the more steps ahead are considered in the forecast, the more dependent on the predicted values it will be. The ideal is updating the data whenever new information is available, substituting the latest predictions for the incoming real data. As a result, the future predictions have a more precise and realistic understanding of the past.

3.5.2 PARAMETERS STUDY

A total of 4 time series of the Euro US Dollar exchange rate throughout 2010 and 2019 have been analysed following the crisp forecasting method presented above. The first one follows a bullish trend, whereas the others follow bullish and sideways behaviours. The influence of each parameter in the prediction outcome was studied by comparing performance indicators based on four different error measures: MAPE, MAE, RMSE, and MSE. This accuracy measure will be introduced in chapter 5.

The main 4 parameters studied were kernel function, kernel estimator, bandwidth, and number of forecasting steps. The first two were tested methodically including all different options mentioned above. In comparison, three bandwidth values were analysed for each period. One of the values was selected through the second bandwidth selection method presented above, the error measures optimal bandwidth. This method may appear primitive and slow, but the sample size allowed to use it without considerably extending the computational time or cost with remarkable results. The other two values were intentionally chosen to observe the behaviour of the model in extreme situations. Other bandwidth optimization methods had to be declined as issues appeared due to the small sample size as well as the value and little variance of the exchange rate data, especially for kernel functions with influence boundaries. For the last parameter, two and three days ahead were tested to compare the performance in each situation.

The first parameter to be analysed is the number of steps ahead, as it offers the simplest performance observation. The more steps ahead are predicted, the higher error measures take place. When trying to forecast further future values, more weight lies on the recently not-updated predicted values, which have the strongest influence for being the closest neighbours. Additionally, the long-term forecasts tend to stabilize around a constant value for the Nadaraya-Watson estimator (Figure 6), whereas they tend to follow a line with a constant slope for the Local Linear estimator (Figure 7). This phenomenon comes from the own weighting nature of the estimator function. A superior performance can be observed for a two-steps-ahead forecasting (Figure 8). Therefore, the model shows the best performance for one step ahead predictions, followed by increasing steps. The number of sample points out of the 2% prediction margin in Figure 6, Figure 7, and Figure 8 are a good indicator of the models' performance.

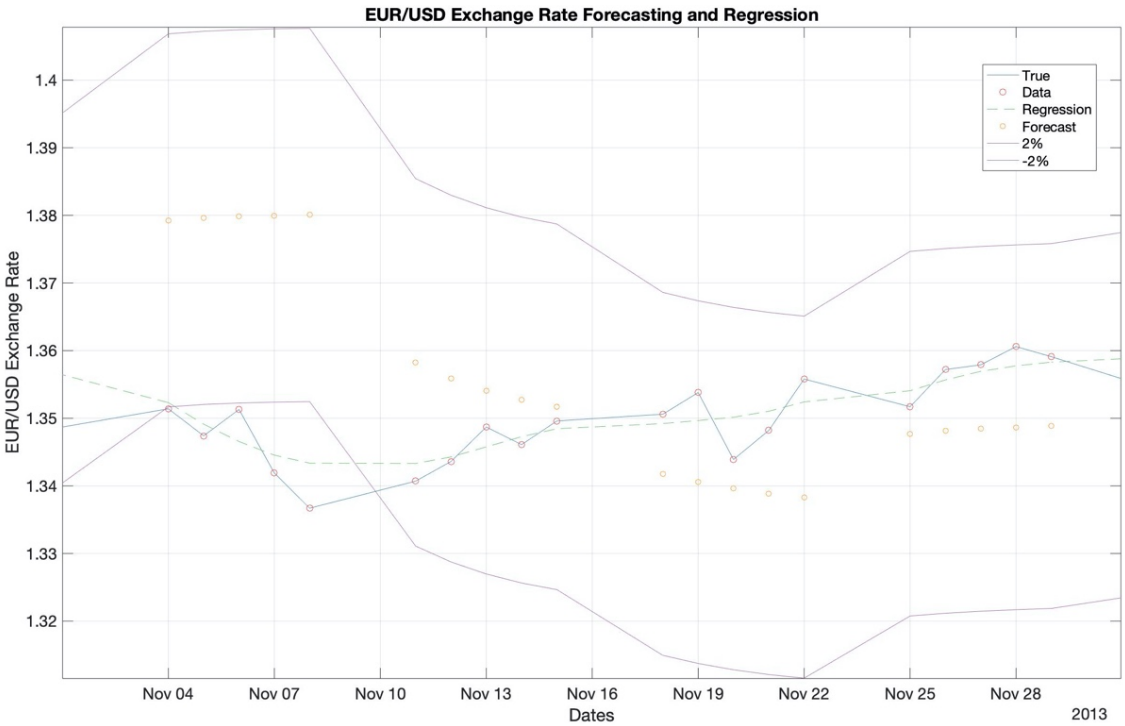


Figure 6. Five-steps-ahead crisp kernel forecast using the Nadaraya-Watson estimator.

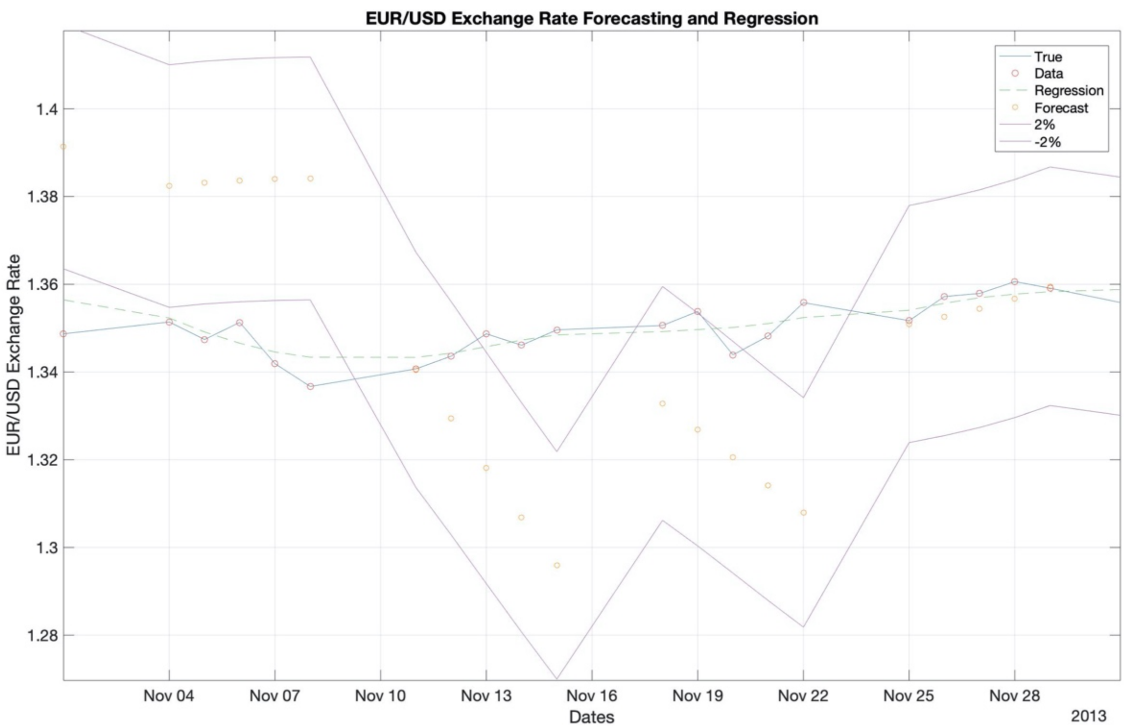


Figure 7. Five-steps-ahead crisp kernel forecast using the Local Linear estimator.

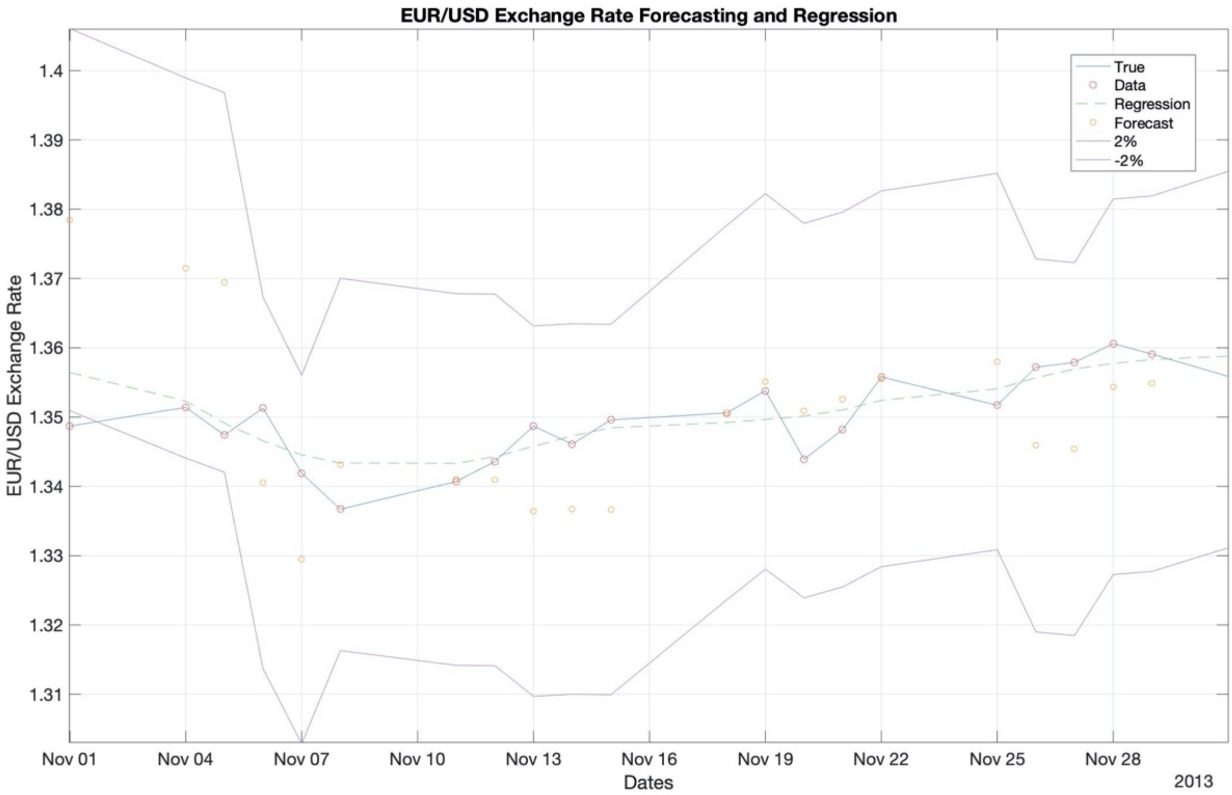


Figure 8. Two-steps-ahead crisp kernel forecast using the Local Linear estimator.

When it comes to the estimator, best forecasting results are achieved using the Local Linear estimator. This is natural, as its initial purpose was to correct the border bias of the Nadaraya-Watson estimator. When observing pure regression, it is difficult to state which of both has better results, which vary depending on the sample data and other model parameters.

Talking about the behaviour of the different kernel functions, they can be divided into two main groups. The first ones are those without boundary limits: Gaussian, Logistic, Sigmoid, and Silverman. All the others form the second group. On the one hand, the first group achieves better forecasting performances with little exceptions when combined with the Local Linear estimator. Surprisingly, the second group has smaller error values when the Nadaraya-Watson estimator is in use. On the other hand, regression results are usually superior for the second group.

The bandwidth optimal value is the most complex to analyse as it varies considerably depending on the sample data and other model parameters. The main conclusion that can be made is that one narrow range of optimal values can be found for each function group mentioned above. Each group tends to have a preferent bandwidth value, sometimes being similar while other times being considerably distant. Usually, the second group has higher optimal values due to the boundary limits, which create the need of widening the influence area through the bandwidth. Figure 9 shows the prediction of the third time series with an optimized bandwidth value of 3, which gives the smallest error measures for the Gaussian function and Local Linear estimator in these conditions. In contrast, Figure 10 shows the same conditions but with an oversmoothed bandwidth value of 9.

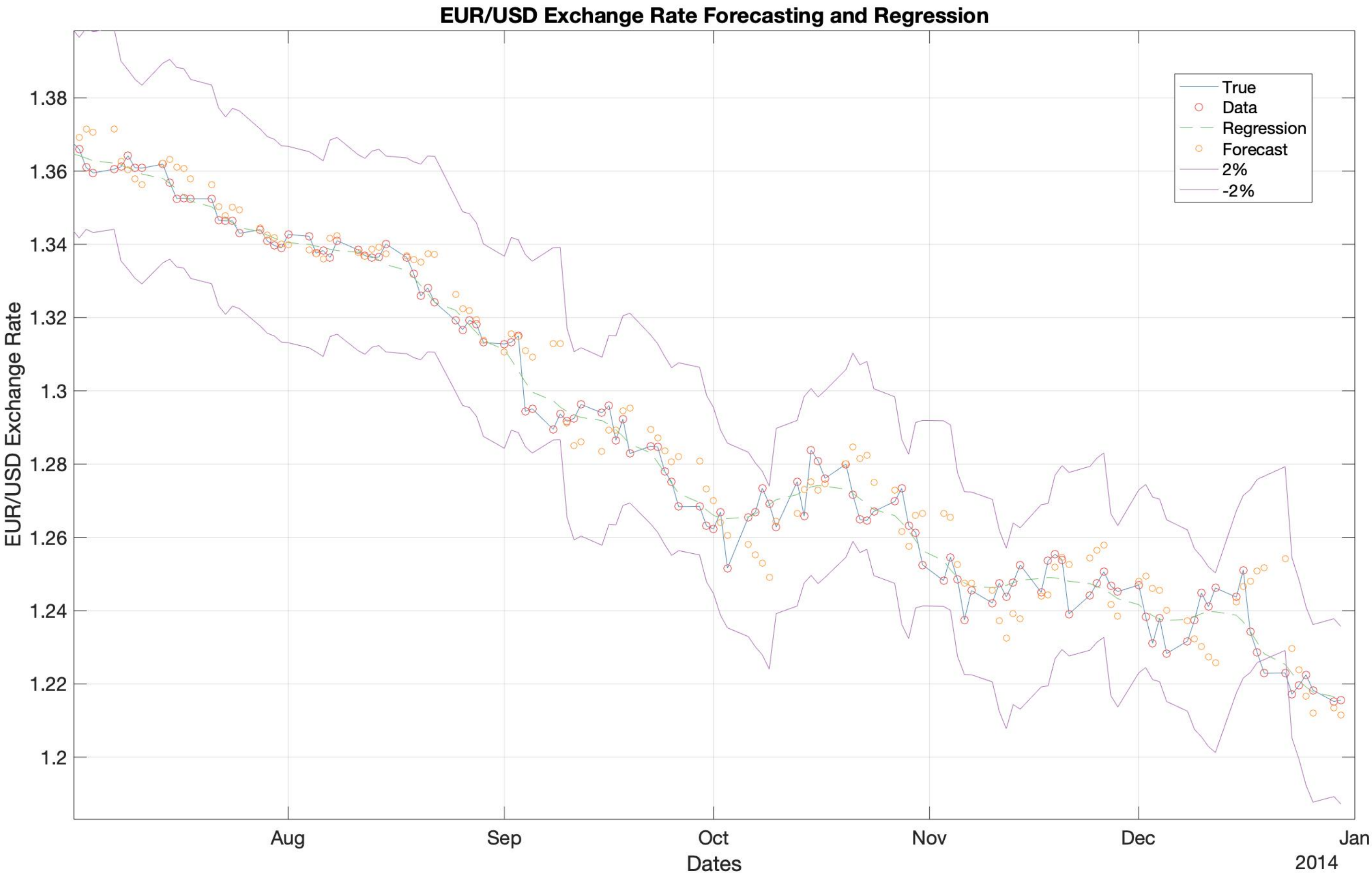


Figure 9. Two-steps-ahead crisp kernel forecast with an optimized bandwidth value.

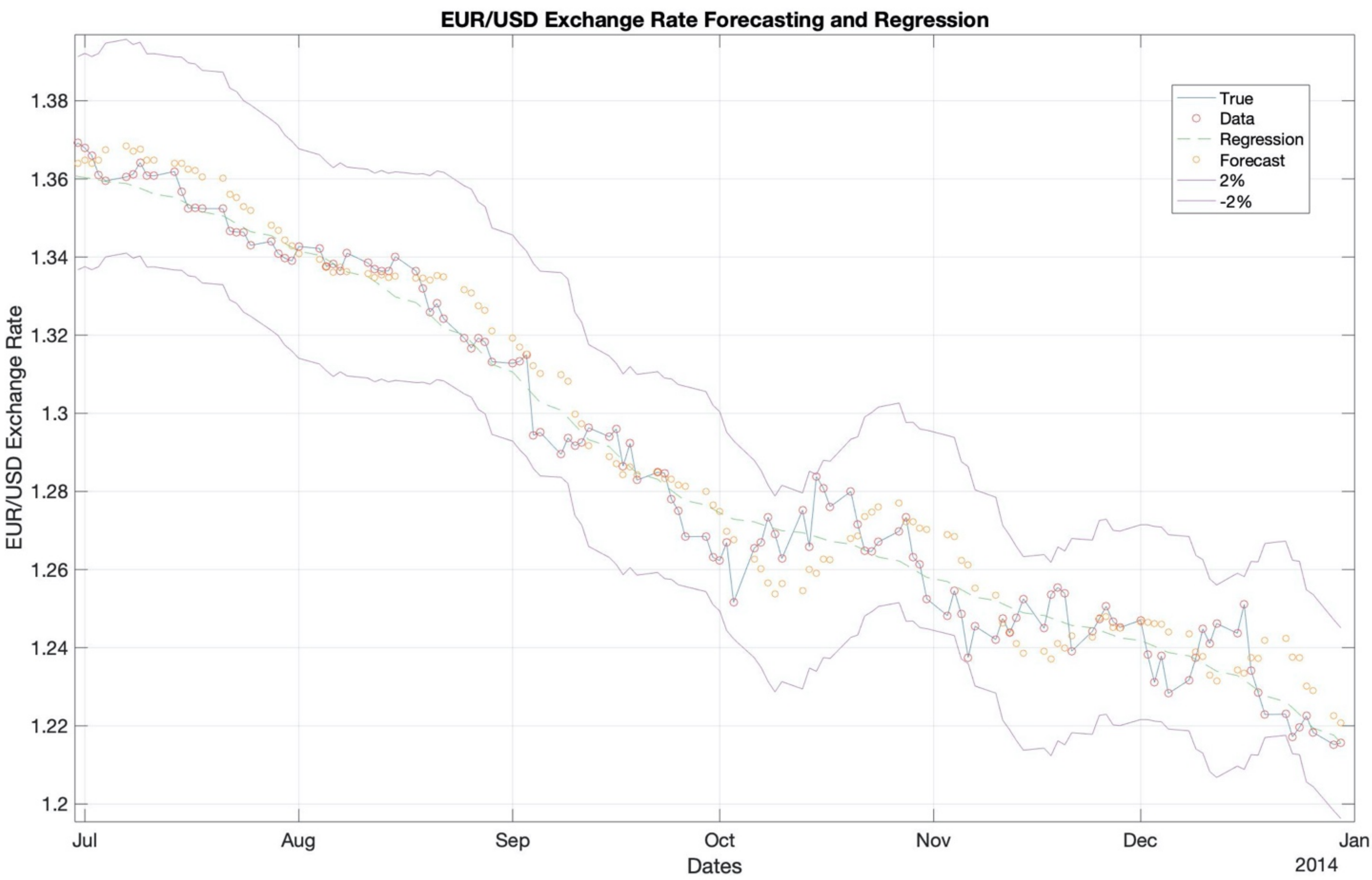


Figure 10. Two-steps-ahead crisp kernel forecast with an oversmoothed bandwidth value.

The difference between both predictions is immediately visible in the shape of the margin lines, which is softer for Figure 10. Using a high bandwidth does not allow the model to properly predict sharper and more aggressive changes, such as those observable around October and November of 2014.

The last noteworthy observation is related to the data trend mentioned at the beginning of this section. The performance of the model is slightly enhanced when the sample follows a down- or uptrend instead of remaining constant around a value.

To summarize all the mentioned conclusions, the best forecasting results are achieved when using the Local Linear estimator in addition to open Kernel functions. Furthermore, the less steps ahead are predicted, the more precise the forecast will be. Additionally, a universal optimal bandwidth value does not exist and a selection process has to be carried out for each situation. Finally, the results point towards better performance levels for up- or downtrend time periods.

CHAPTER 4. ITS KERNEL REGRESSION AND FORECASTING

4.1 INTRODUCTION TO ITS KERNEL REGRESSION

As observed above, regression analysis models the behaviour of a response variable Y for a given explanatory variable X. Nevertheless, not all real-world phenomena can be represented by single fixed response values. When trying to model or predict from the crisp data perspective, large amounts of information can be lost. One clear example could be the temperature forecast, where just predicting the maximum daily temperature fails to explain the variation throughout the day. A common solution for this problem is shortening the time interval, which in this case translates into predicting by hours instead of days. This process can be repeated as much as desired for time series, as time is a continuous variable. However, it creates the second main inconvenience of crisp data models, which is the exponential increase of sample size and computational cost. The more divisions are made, the more data has to be handled. The simple change of predicting by hours instead of daily raises the computational cost and the sample size 24 times. In the case of weather forecasting, this level of information may be detailed enough, but it may not be for other real-life phenomena such as financial markets, where a more detailed picture of the future behaviour of the market gives a privileged position. Therefore, a balance between complexity and cost needs to be carried out.

Interval-valued time series offer a good solution for both problems by converting extremely large crisp datasets into smaller more manageably sized ones without lacking the internal variation and structure of the variable. Their use is highly present in today's society, e.g. weather forecasting and stock market values. Moreover, all continuous variables can benefit from them. Giving a closer look to the engineering field, all designs and measures count with tolerances. These can be regarded as interval-valued data as well. It is true that the explanatory variable X may not be time, but some regression models mentioned below could be applied to this field as well. The advantages that interval-valued series offer and their extensive use create a need for further study.

Whereas linear regression analysis of interval-valued data began two decades ago with some key papers such as Billard and Diday (2000, 2002), Lima Neto et al. (2004), and Im and Kang (2018), limited study of nonparametric regression models for interval-valued time series has been carried out.

Fagundes et al. (2014) were among the first studying this field, proposing two interval-valued kernel regression model families. The first one is based on estimating the interval boundaries through multiple kernel regressions, while the second one uses regression mixtures by combining kernel and linear regression. Moreover, Lima Neto and De Carvalho (2017) presented a nonlinear regression method for interval-valued data. More recently, Jang and Kang (2020) introduced a different alternative for nonparametric regression by implementing the Local Linear estimator, extending Fagundes et al.' research, which was limited to the Nadaraya-Watson estimator.

In this chapter, some of the methods mentioned above will be briefly introduced. Many of them are based on combining crisp data kernel regression models for key representing values of the interval. Therefore, having a good understanding of the preceding chapter is a solid foundation for ITS kernel regression. Furthermore, the application of these methods for forecasting will be introduced and analysed in the last section.

4.1 ITS KERNEL REGRESSION

The simplest form of interval-valued data representation is based on the upper and lower interval values in the form of

$$y_i = [y_{Li}, y_{Ui}]. \tag{5}$$

In this expression, y_{Li} and y_{Ui} denote the lower and upper boundaries of the i^{th} observation of the variable y respectively. This form of representation captures the variance as well as the range of values of the variable, giving a more complete picture when compared to crisp data for the same interval period.

4.1.1 CENTER INFORMATION ITS KERNEL REGRESSION

Following Fagundes et al. (2014), this first method applies crisp data kernel regression to each boundary separately. As a result, the interval-valued regression is decomposed into two singular kernel regressions.

Let (x_i, y_i) be a data set of n objects where y_i is an interval-valued variable following the structure mentioned above. The upper and lower boundaries are separated and treated as crisp variables, applying an estimator as seen in chapter 2 for each of them

$$m_{0L}(x) = \mathbb{E}(Y_L|X = x)$$

$$m_{0U}(x) = \mathbb{E}(Y_U|X = x).$$

The final estimation of the interval follows the form

$$\widehat{\mathbb{E}}(Y|X) = [\widehat{m}^{y_L}(x), \widehat{m}^{y_U}(x)\,]. \tag{6}$$

As seen in chapter 3, bandwidth, kernel function, and estimator have to be determined for each variable. The estimator is used as a weighting function and weights have to be calculated for each value of X. No additional steps are involved. The authors select the Nadaraya-Watson (1) estimator in this case. Furthermore, the Gaussian kernel function is chosen. The method offers a simple adaptation of crisp data models for interval-valued variables.

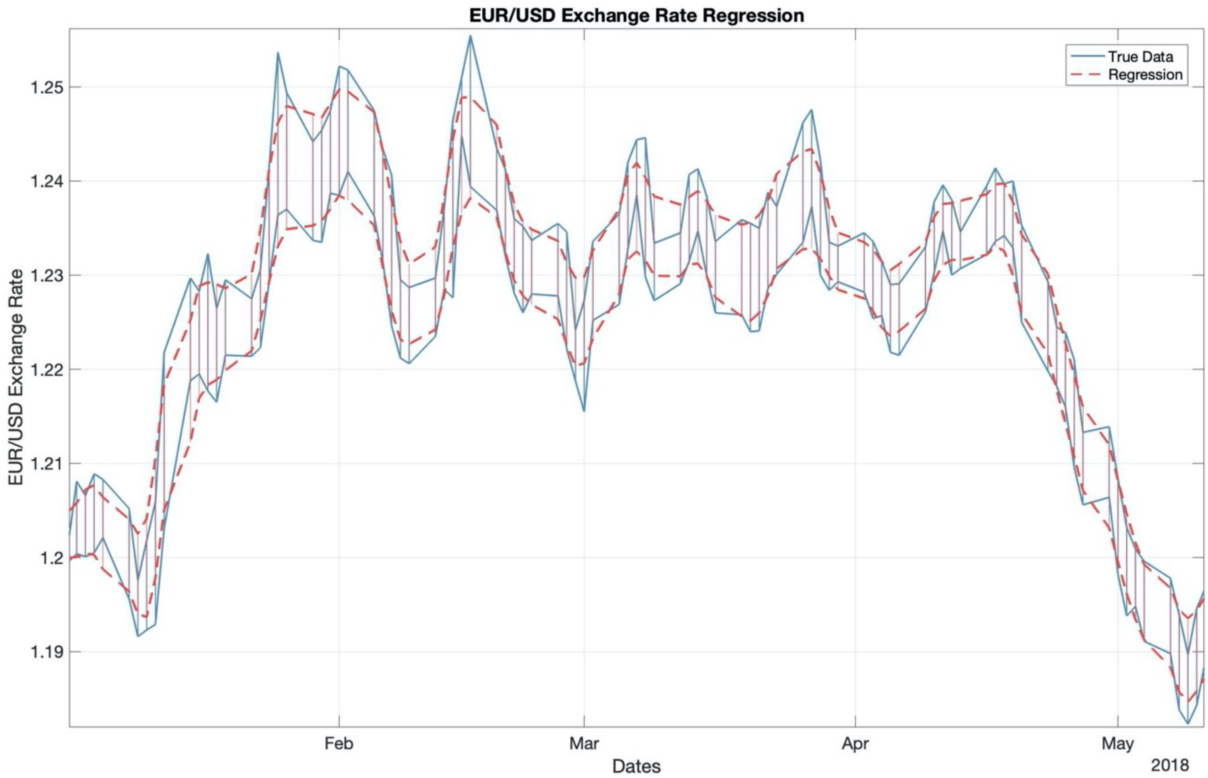


Figure 11. Center information ITS kernel regression.

Figure 11 illustrates an example of the center information ITS kernel regression (6) applied to the Euro US Dollar exchange rate during the beginning of 2018 using the Nadaraya-Watson estimator, the Tricube kernel function, and a bandwidth value of 3 for both crisp variables.

4.1.2 CENTER INFORMATION ITS KERNEL REGRESSION

The same paper of Fagundes et al. (2014) proposes a second model based on estimating the center and range of the interval instead of the upper and lower boundaries. In this case, the response interval can be rewritten as

$$y_i = [y_{Li}, y_{Ui}] = \left[y_i^c - \frac{y_i^r}{2}, y_i^c + \frac{y_i^r}{2}\right], \tag{7}$$

where y_i^c and y_i^r are the center and range of the i^{th} observation respectively and are computed as

$$y_i^c = \frac{y_{Li} + y_{Ui}}{2}$$

$$y_i^r = y_{Ui} - y_{Li} \ .$$

Similar to the first method, the next step is to decompose the interval-valued estimations into two singular crisp data models, but this time for the center and range values

$$m_{0c}(x) = \mathbb{E}(Y_c | X = x)$$

$$m_{0r}(x) = \mathbb{E}(Y_r | X = x)$$

with the final estimation following the form

$$\widehat{\mathbb{E}}(Y|X) = \left[\widehat{m}^c(x) - \frac{\widehat{m}^r(x)}{2}, \widehat{m}^c(x) + \frac{\widehat{m}^r(x)}{2}\right]. \tag{8}$$

The range could be considered as the diameter. This approach adds or subtracts the radius of the range to the center by dividing the diameter by two. The complexity of this method is slightly higher as more middle steps are involved. However, it is still a distinctively practical method to adapt crisp data regression models for interval-valued variables. The selection of the estimator and the kernel function by the authors remains the same as above.

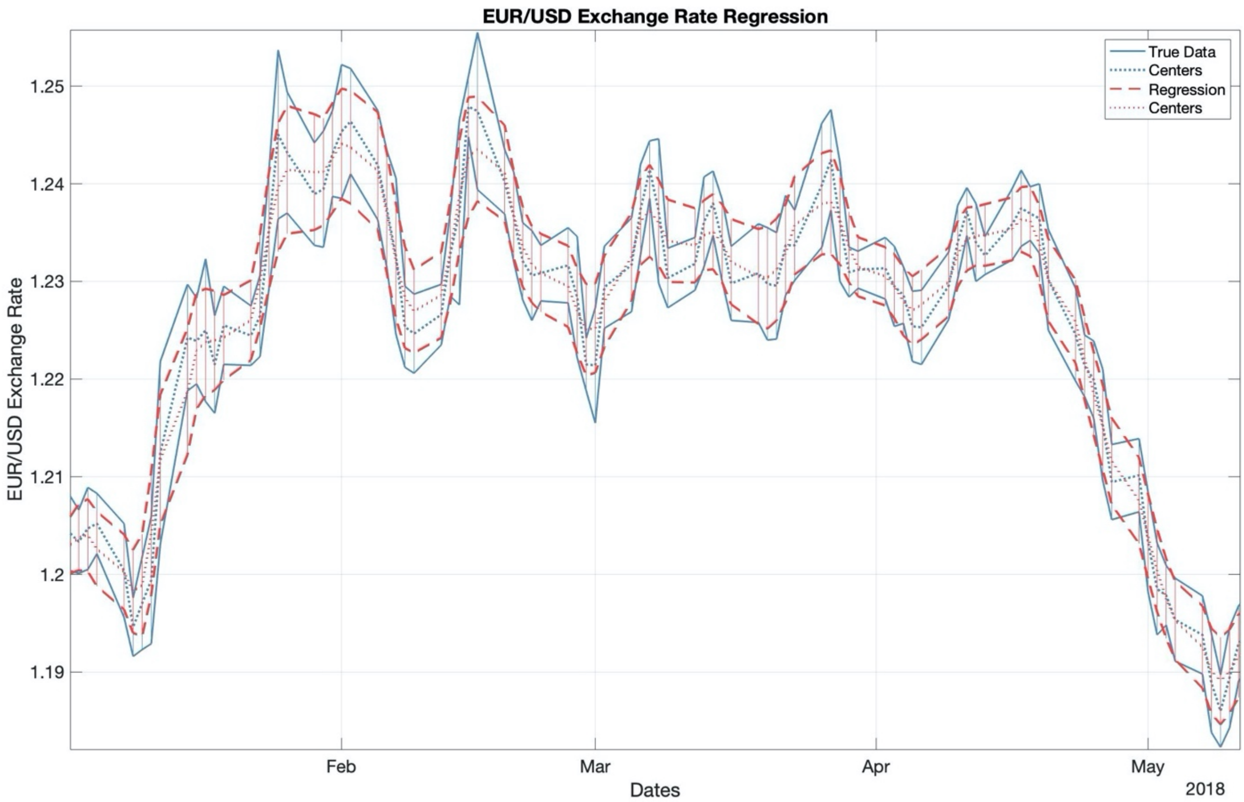


Figure 12. Center and range information ITS kernel regression.

Figure 11 depicts an example of the center and range information ITS kernel regression (8) applied to the same time series and with the same parameters' selection as in Figure 11.

4.1.3 ITS KERNEL REGRESSION MIXTURE AND OTHERS

The two last methods presented in Fagundes et al. (2014) are based on the center and range information method used with a combination of kernel regression and linear regression. These will be shortly introduced but not tested in the case study, as parts of them go beyond the scope of this study. The first one applies a linear form to model the centers and the nonparametric Nadaraya-Watson estimator to the ranges. The final estimation follows the form

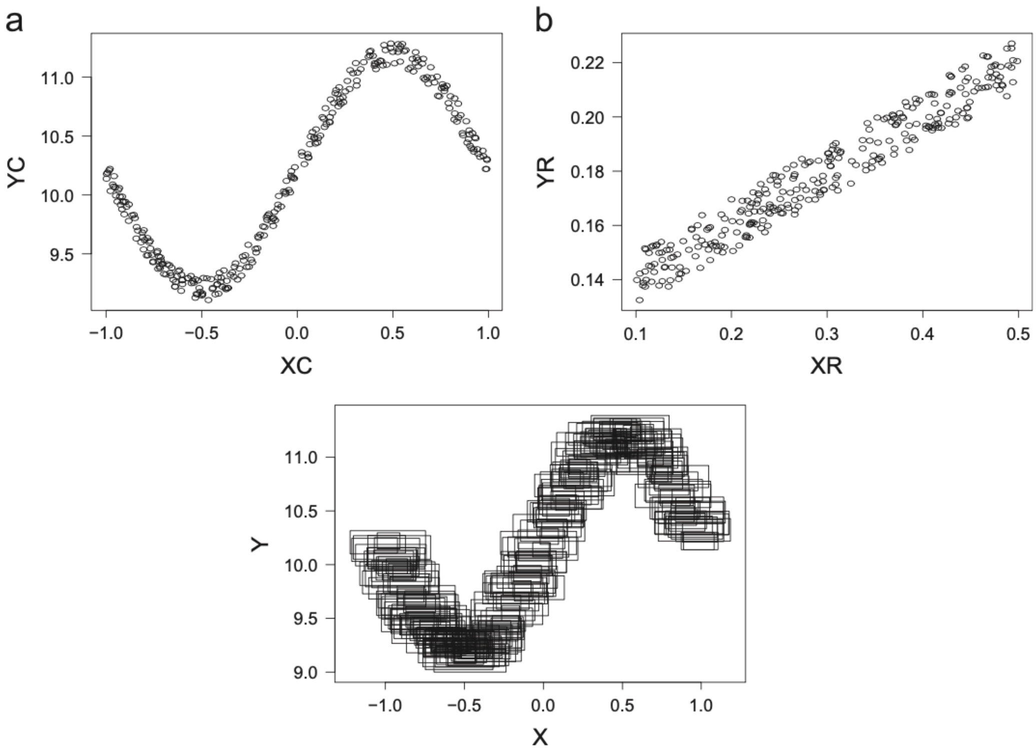
$$\widehat{\mathbb{E}}(Y|X) = \left[\hat{m}_L^c(x) - \frac{\hat{m}_{NW}^r(x)}{2}, \hat{m}_L^c(x) + \frac{\hat{m}_{NW}^r(x)}{2}\right],$$

where $\hat{m}_L^c$ and $\hat{m}_{NW}^r$ denote the linear and nonparametric estimator respectively. The parameters of the linear regression are calculated using a least square estimator. Meanwhile, the second mixture method does the opposite, employing the Nadaraya-Watson estimator for the centers and the linear regression for the ranges

$$\widehat{\mathbb{E}}(Y|X) = \left[\hat{m}_{NW}^c(x) - \frac{\hat{m}_L^r(x)}{2}, \hat{m}_{NW}^c(x) + \frac{\hat{m}_L^r(x)}{2}\right].$$

Figure 13 shows an example where the upper left subfigure illustrates the estimation of the centers calculated by kernel regression using the Nadaraya-Watson estimator. Meanwhile, the upper right subfigure depicts the linear regression of the center of the intervals. The lower plot is a combination of the both. Rectangles are formed for each x_0 as a range of possible results is formed by combining both estimations.

Figure 13. Example of ITS kernel regression mixture [Fagundes et al. (2014)].



The combination of two different regression methods rises the complexity of the process, sometimes without considerable improvements. The conclusions of the research indicated higher performance levels for the second and fourth methods followed by the first one. A comparative study with other nonlinear regression models also showed superior performances of the nonparametric and mixture models.

More recently, Jang and Kang (2020) proposed a slight variation of the center and range information method (8) by applying the Local Linear Estimator (2) to both variables

$$\widehat{\mathbb{E}}(Y|X) = \left[\widehat{m}^{c}_{LL}(x) - \frac{\widehat{m}^{r}_{LL}(x)}{2}, \widehat{m}^{c}_{LL}(x) + \frac{\widehat{m}^{r}_{LL}(x)}{2}\right].$$

In a similar way to the Nadaraya-Watson estimator, the steps for modelling the variables through the Local Linear estimator can be found in chapter 2. No additional computations apart from calculating the centers and ranges and combining them at the end to form the intervals are required. The main conclusion of the study pointed towards a superior performance of this method compared to the ones mentioned in Fagundes et al. (2014). This makes sense looking back to the conclusions of chapter 2, where the Local Linear estimator also had better performance levels for crisp data kernel regression.

The implementation of the Local Linear estimator to the center and range information method can be observed in Figure 14 on the Euro US Dollar exchange rate during a bullish trend of May and June 2018. The Quartic kernel function and a bandwidth value of 3 were used.

The detailed development of the crisp data kernel regression and its parameters in chapter 2 allows to easily rise the degree of customization of the interval-valued regression models mentioned above. By substituting the Gaussian function, the performance of these ITS regression models could improve, because, as already seen before, different kernel functions may be better suited for different scenarios. Furthermore, decomposing the interval-valued data into several variables enables to personalize and optimize each variable's model separately by selecting optimal bandwidth values, estimators and kernel functions for each of them.

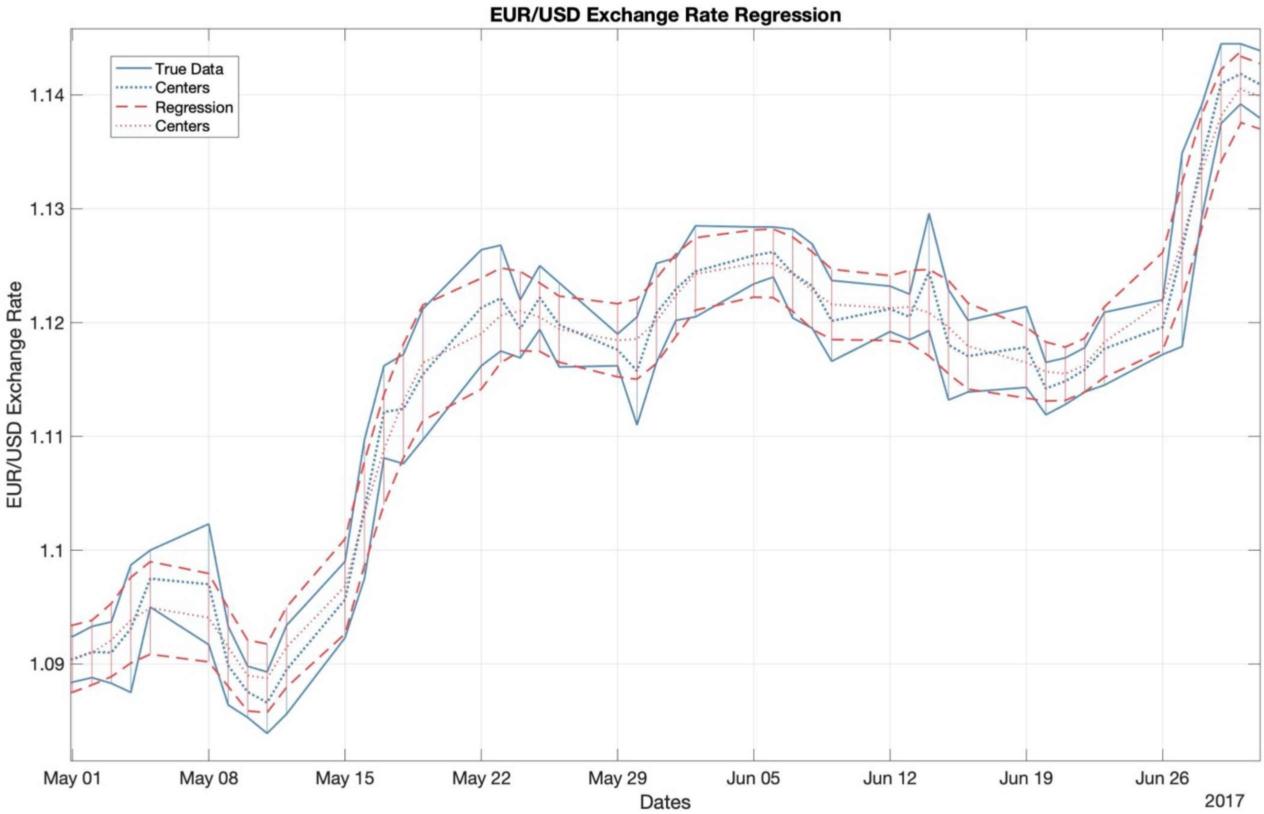


Figure 14. Center and range information ITS kernel regression with the Local Linear estimator.

4.2 ITS KERNEL FORECASTING

The interval-valued kernel regression discussed above serves as a foundation for ITS kernel forecasting, which will be introduced in this section. In a similar way to crisp data forecasting, the regression models can be easily transformed into forecasting ones. There is little to almost no existing academic literature related to kernel forecasting methods applied to interval-valued time series. This thesis introduces three different kernel forecasting methods. The first one is based on the center information regression method (6), whereas the second one is inspired by the center and range information method (8). The last one is an own proposal thought for stock market forecasting.

4.2.1 CENTER INFORMATION ITS KERNEL FORECASTING

As seen in equation (3), the main idea behind every kernel forecasting method is to calculate a weighted average of the data sample for the prediction. Therefore, similar to ITS kernel regression, the interval-valued response variable Y has to be decomposed into two crisp variables (6).

In a similar way to equation (3), let us consider two stationary and univariate time series Z_t^U and Z_t^L observed on the time interval $1 \leq t \leq n$. Both follow a Markov process, which means that the predictions can be made based on their present state

$$Z_{t+1}^U = m^U\left(Z_t^U, Z_{t-1}^U, Z_{t-2}^U, \dots, Z_{t-p}^U\right) + \varepsilon_{t+1}^U$$

$$Z_{t+1}^L = m^L\left(Z_t^L, Z_{t-1}^L, Z_{t-2}^L, \dots, Z_{t-p}^L\right) + \varepsilon_{t+1}^L.$$

Z_t^U and Z_t^L denote the upper and lower time series respectively, whereas m^U and m^L represent the estimator functions for each of them. The variable p denotes the number of past time intervals influencing the forecast and ε^U and ε^L are the respective forecasting errors.

Following the recursive crisp data forecasting equation (4), if forecasts of more than one step ahead want to be done, the already predicted values would be added to the original sample data

$$Z_{t+2}^U = m^U\left(Z_{t+1}^U, Z_t^U, Z_{t-1}^U, \dots, Z_{t-p}^U\right) + \varepsilon_{t+1}^U$$

$$Z_{t+2}^L = m^L\left(Z_{t+1}^L, Z_t^L, Z_{t-1}^L, \dots, Z_{t-p}^L\right) + \varepsilon_{t+1}^L .$$

Modelling the upper and lower boundaries separately gives the freedom to select different estimators, kernel functions and bandwidths for each of them. The different parameters selection can be guided by crisp data error measures minimization or interval-valued error measures minimization. These last error measures will be introduced in chapter 5 before the case study.

A two-steps-ahead ITS center information kernel forecast of the Euro US Dollar exchange rate during a bearish trend of the second semester 2014 can be seen in Figure 15. The Local Linear estimator was used in combination with the Gaussian kernel function and optimized bandwidth values.

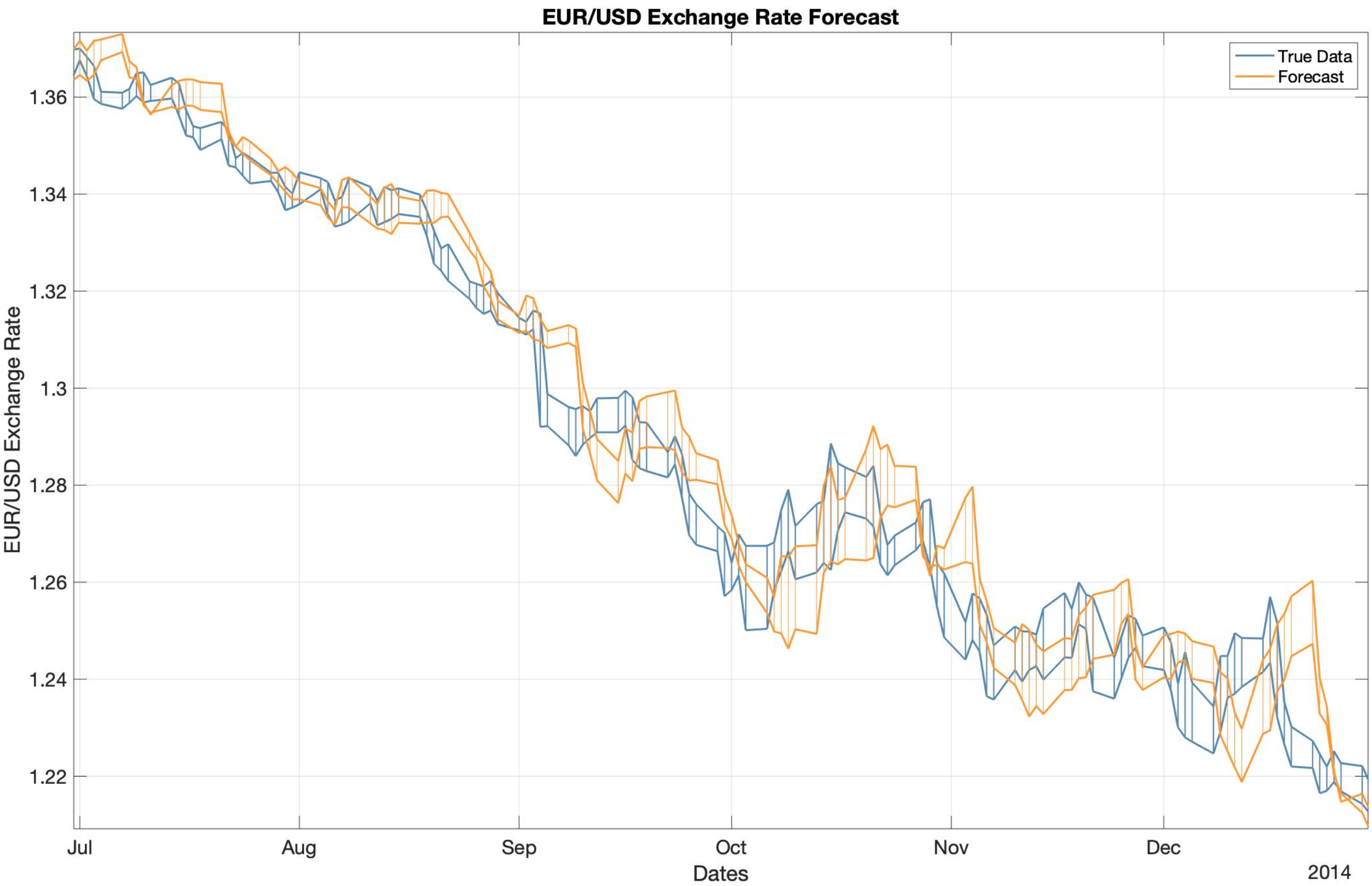


Figure 15. ITS center information kernel forecast.

4.2.2 CENTER AND RANGE INFORMATION ITS KERNEL FORECASTING

As seen in the equation (7), in this method the interval-valued response variable Y can be represented by

$$Y = [Y_L, Y_U] = \left[Y^c - \frac{Y^r}{2}, Y^c + \frac{Y^r}{2}\right].$$

Similar to the first ITS forecasting method, two stationary and univariate time series Z_t^c and Z_t^r representing the centers and ranges can be considered. Following the same development, the individual predictions have the form

$$Z_{t+1}^c = m^c\left(Z_t^c, Z_{t-1}^c, Z_{t-2}^c, \dots, Z_{t-p}^c\right) + \varepsilon_{t+1}^c$$
$$Z_{t+1}^r = m^r\left(Z_t^r, Z_{t-1}^r, Z_{t-2}^r, \dots, Z_{t-p}^r\right) + \varepsilon_{t+1}^r.$$

Once the prediction is completed, only one more step is necessary, which consists in calculating back the lower and upper boundaries based on the center and range predictions. This second method also allows a high customization range, which can be guided by error measures' minimization.

Figure 16 illustrates an example of the ITS center and range information kernel forecasting in the same conditions as the preceding figure.

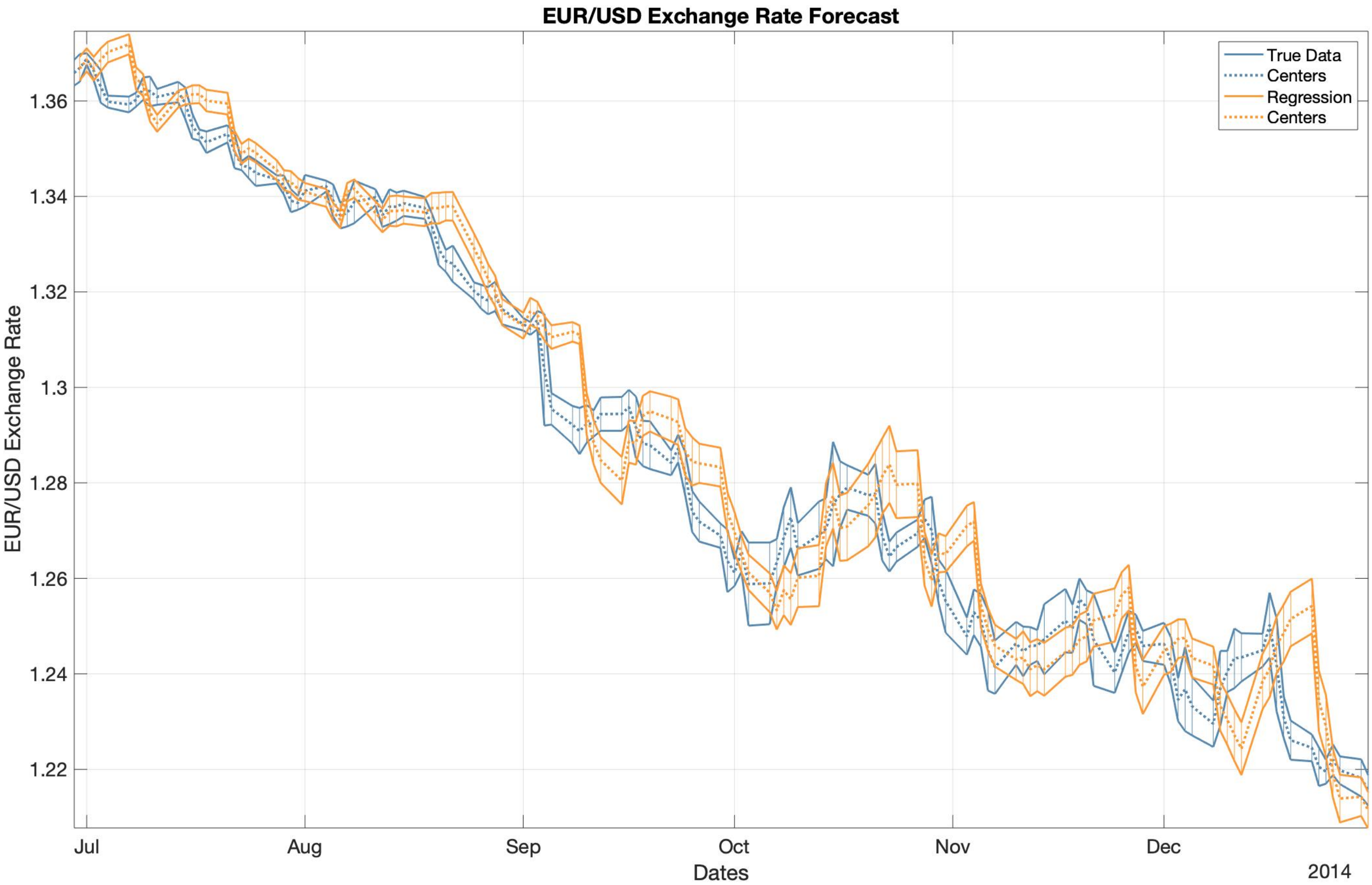


Figure 16. ITS center and range information kernel forecast.

4.2.3 FINANCIAL MARKETS ITS KERNEL FORECASTING

Two forecasting methods based on already existing regression models have been proposed. In this section, a third one is introduced. This model, or better said, combination of previous models, is thought to better suit the financial markets, as it gives a more detailed picture of the behaviour of a financial asset. When trying to predict the future value of, for example, a stock, both models fail to transmit key information. This key information is the closing stock price, which many investors, traders, financial institution, regulators, and stakeholders use as a reference point for determining performance over a specific time. Information about the closing price added to the high and low values can give a better understanding of when to sell or buy a stock.

Let us consider an interval-valued variable Y which can be represented as

$$Y = [Y_L, Y_{CL}, Y_U] = \left[Y^c - \frac{Y^r}{2}, Y_{CL},\ Y^c + \frac{Y^r}{2}\right]$$

where Y_U *and* Y_L denote the upper and lower boundaries, Y_{CL} depicts the closing price, and Y^c and Y^r represent the center and ranges. The closing price has to satisfy

$$Y_L \leq Y_{CL} \leq Y_U.$$

In a similar way to both forecasting methods discussed above, the interval-valued response variable can be decomposed into three instead of two representative crisp variables. A superposition of any of the two preceding methods and a third crisp kernel forecast is possible. Each of them can be predicted individually following

$$Z_{t+1}^U = m^U\left(Z_t^U, Z_{t-1}^U, Z_{t-2}^U, \dots, Z_{t-p}^U\right) + \varepsilon_{t+1}^U$$

$$Z_{t+1}^L = m^L\left(Z_t^L, Z_{t-1}^L, Z_{t-2}^L, \dots, Z_{t-p}^L\right) + \varepsilon_{t+1}^L$$

$$Z_{t+1}^{CL} = m^{CL}\left(Z_t^{CL}, Z_{t-1}^{CL}, Z_{t-2}^{CL}, \dots, Z_{t-p}^{CL}\right) + \varepsilon_{t+1}^{CL}$$

for the center information method or

$$Z_{t+1}^{c} = m^{c}\left(Z_{t}^{c}, Z_{t-1}^{c}, Z_{t-2}^{c}, \dots, Z_{t-p}^{c}\right) + \varepsilon_{t+1}^{c}$$

$$Z_{t+1}^{r} = m^{r}\left(Z_{t}^{r}, Z_{t-1}^{r}, Z_{t-2}^{r}, \dots, Z_{t-p}^{r}\right) + \varepsilon_{t+1}^{r}$$

$$Z_{t+1}^{CL} = m^{CL}\left(Z_{t}^{CL}, Z_{t-1}^{CL}, Z_{t-2}^{CL}, \dots, Z_{t-p}^{CL}\right) + \varepsilon_{t+1}^{CL}$$

for the center and range information method. No difficult steps are added to the process, but a more detailed glance of the future variable is achieved. Whenever the closing price prediction is lower or higher than the lower or upper forecasted boundaries, the closing price will be equal to those values. Figure 17 provides an example of the financial markets ITS kernel forecasting applied to a 2014 summer period of the Euro US Dollar exchange rate with the Local Linear estimator, the Gaussian function and optimized bandwidth values. The forecast experiences a high accuracy from the end of July until the 20th of August, but fails to follow the fast and aggressive changes taking place around the 20th of August and beginning of September.

The main novelty of this model does not depend upon the calculation of the forecast, which is based on the center and center and range information methods, but upon the additional visual information provided by the results.

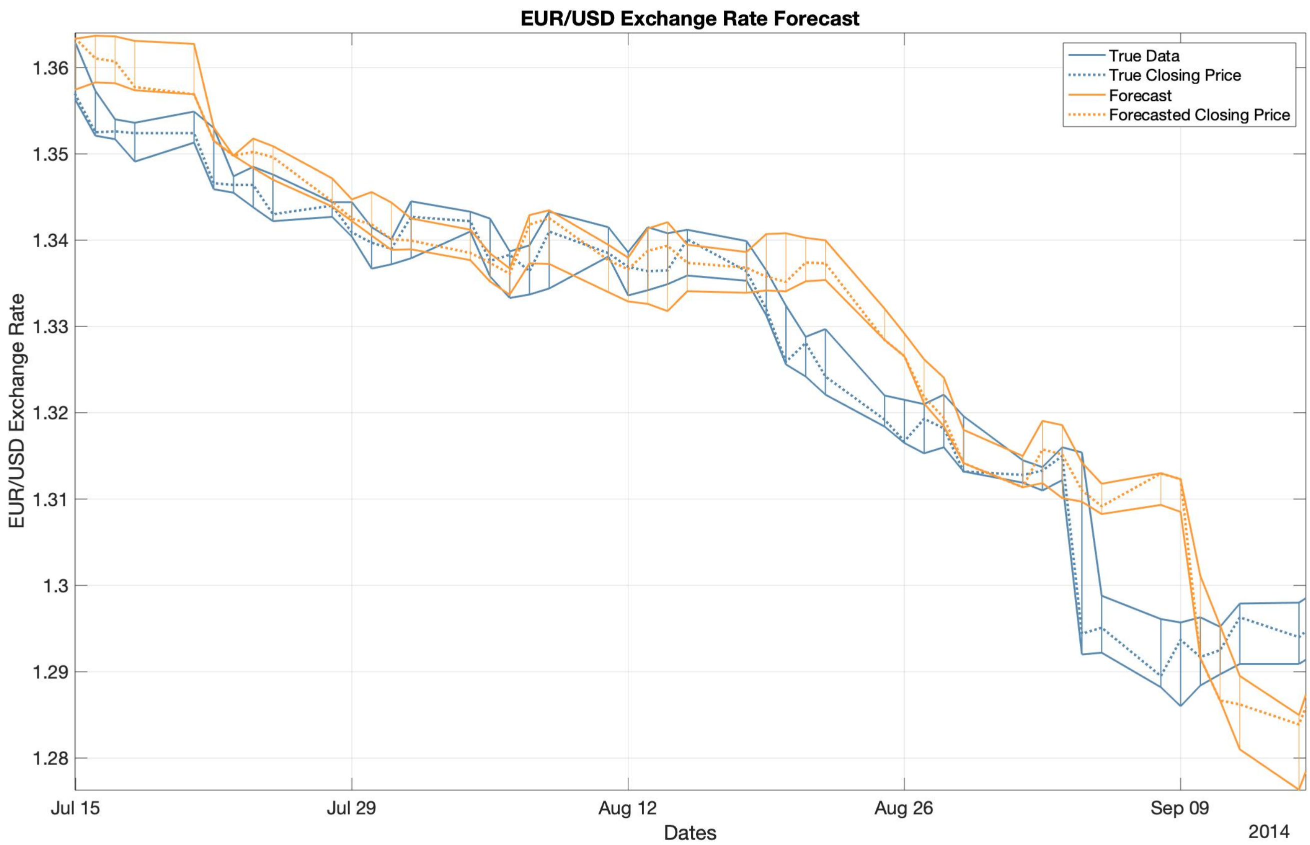


Figure 17. ITS financial markets kernel forecast.

CHAPTER 5. CASE STUDY

5.1 ACCURACY MEASURES

In order to compare the performance of the different models, accuracy measures are needed. There are measures specially thought for crisp data whereas others are directed towards interval-valued data. Each of them has its own advantages and disadvantages. In this section, the error measures that are used throughout the case study are introduced.

5.1.1 CRISP KERNEL REGRESSION

A good indicator to compare several regression methods is the mean squared error (MSE) defined as

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

where $\hat{y}_t$ stands for predicted value and y_t for observed value. In addition, the RMSE, the square root of the MSE, is another commonly used measure which will be used as well. The third error measure implemented in this thesis is the mean absolute error (MAE), which is represented by

$$MAE = \frac{\sum_{t=1}^{T}|\hat{y}_t - y_t|}{T}.$$

The last error measure for crisp data regression is the mean absolute percentage error (MAPE) and is calculated by

$$MAPE = \frac{\sum_{1=t}^{T} \left| \frac{\hat{y}_t - y_t}{y_t} \right|}{T} \times 100 \, .$$

This metric is another commonly used accuracy ratio due to its intuitive interpretation in terms of relative error. Out of the four error measures, the RMSE is usually the most informative for comparison, as it reflects a better performance metric when dealing with large error values by penalizing them proportionally.

5.1.2 CRISP KERNEL FORECASTING

Apart from those mentioned above, one more error measure will be added for crisp time series forecasting. The root mean squared scaled error (RMSSE) introduced by Makridakis et al. (2020), which is a variant of the famous mean absolute scaled error (MASE), is calculated as follows

$$RMSSE = \sqrt{\frac{\frac{1}{h}\sum_{t=n+1}^{n+h}(y_t - \hat{y}_t)^2}{\frac{1}{n-1}\sum_{t=2}^{n+h}(y_t - y_{t-1})^2}} \, ,$$

where n is the length of the training sample and h the forecasting horizon. This measure is scale independent and penalizes positive and negative forecast errors, thus being symmetric.

5.1.3 ITS KERNEL REGRESSION AND FORECASTING

Several error measures have been introduced for crisp data regression. As our ITS regression models are based on composed crisp data methods, the same error measures can be respectively applied to the different decomposed variables. However, 4 more measures oriented towards interval-valued data are introduced. Following Maté (2021), the interval average relative variance (iARV) compares the predictions of the model with the predictions given by average values of the interval variable and is calculated as follows

$$iARV = \frac{\sum_{t=1}^{T}(\hat{y}_t^L - y_t^L)^2 + \sum_{t=1}^{T}(\hat{y}_t^U - y_t^U)^2}{\sum_{t=1}^{T}(\bar{y}_t^L - y_t^L)^2 + \sum_{t=1}^{T}(\bar{y}_t^U - y_t^U)^2}.$$

The lower the value of iARV, the better the estimation. The second performance metric, the interval Theil's U statistic (iUTheil), compares the predictions of the model with a naïve model in a similar way to the RMSSE. As explained by Maté (2021) and Maia and de Carvalho (2011), it balances the model with the random walk and is given by the equation

$$iUTheil = \sqrt{\frac{\sum_{t=2}^{T}(y_t^L - \hat{y}_t^L)^2 + \sum_{t=2}^{T}(y_t^U - \hat{y}_t^U)^2}{\sum_{t=2}^{T}(y_t^L - y_{t-1}^L)^2 + \sum_{t=2}^{T}(y_t^U - y_{t-1}^U)^2}}.$$

For values equal to 1, the model shows the same performance as the random walk. If the values are lower than 1, it performs better than the naïve model. The opposite occurs for values higher than 1.

On the other hand, two more recent accuracy measures have been introduced by Rodrigues and Salish (2015). Following Maté (2021), who implements them as forecasting accuracy measures, the coverage rate (CR), the first of them, calculates the existing intersection between the predicted and actual intervals and is represented by

$$CR = \frac{1}{T}\sum_{t=1}^{T} \frac{\omega([y]_t \cap [\hat{y}]_t)}{\omega([y]_t)}.$$

The closer the value gets to 1, the more real interval is covered by the predicted value. Nevertheless, this measure fails seizing how much wider the predicted interval is compared to the actual one. Therefore, the second measure called efficiency rate (ER) is needed. Its equation has the form

$$ER = \frac{1}{T}\sum_{t=1}^{T} \frac{\omega([y]_t \cap [\hat{y}]_t)}{\omega([\hat{y}]_t)}$$

and has the only difference of using the predicted interval as divisor. A combination of both is a good indicator of a prediction's accuracy. For a forecast or regression to be reasonably accurate, it is necessary that both measures are as high as possible.

5.2 PERFORMANCE COMPARISON

All the different models as well as the accuracy measures presented above were coded in MATLAB. Furthermore, additional functions generating regression and forecasting illustrations were programmed. MATLAB was the main coding language used throughout the project. A total of around 6500 lines of code distributed throughout 37 MATLAB scripts were written. One of the main reasons for this large size comes from the high level of parameters selection available for all models. The already existing code could serve as foundation for a possible future app development introducing kernel regression and forecasting methods for crisp and interval-valued time series.

The crisp and interval-valued models introduced in this study were applied to different financial time series. Additionally, the accuracy measures presented above were used as performance indicators among the different models. Regression and forecasting models for crisp and interval-valued data were tested separately.

5.2.1 CRISP KERNEL REGRESSION

An initial parameters study was introduced in section 3.5.2. In this section, 5 additional closing price time series were used to expand the performance comparison of the different kernel regression models. Four real random stock assets (*Series 1-4*) dating back from August 2015 to January 2021, and the EURUSD exchange rate (*Serie 5*) were used.

In Table 1, the Local Linear and Nadaraya-Watson estimators are compared using the same kernel function and bandwidth values for crisp kernel regression. All error measures are considerably low, which reflects the high performance of the kernel regression models. Out of the two estimators, a slight improvement is always observed for the Local Linear estimator. This is natural and follows the analysis of chapter 3, as the Local Linear estimator offers a better regression next to the boundaries.

Table 2 shows the comparison of the different kernel functions in 3 out of 5 time series. Only these three series and the RMSE were included, as no other important observations were introduced by the additional information. For regression, it is clear that the kernel functions with boundaries have a better performance than those without, which are the Gaussian, Logistic, Sigmoid, and Silverman. Among the ones of the first type, the functions giving the most weight to the closest points have the smallest RMSE values. These are the Triweight, Tricube, Quartic, and Triangular. These observations support those mentioned in *section 3.5.2.*

Table 1

Performance comparison of the kernel estimators for crisp kernel regression

	MSE		RMSE		MAE		MAPE	
Series	NW	LL	NW	LL	NW	LL	NW	LL
1	0.0360	0.0359	0.1898	0.1895	0.1114	0.1111	2.8793	2.8749
2	0.0206	0.0205	0.1434	0.1433	0.1034	0.1032	2.4381	2.4341
3	0.0219	0.0218	0.1478	0.1478	0.0895	0.0895	3.3679	3.3675
4	0.0610	0.0607	0.2469	0.2465	0.1538	0.1535	2.4712	2.4689
5	1.61e-05	1.61e-05	0.004	0.004	0.0031	0.0031	0.2761	0.2757

Note. Constant bandwidth (2.5) and constant kernel function (Gaussian) were used.

Table 1. Performance comparison of the kernel estimators for crisp kernel regression.

Table 2

Performance comparison of the kernel functions for crisp kernel regression

	RMSE		
Functions	Series 1	Series 2	Series 5
Gaussian	0.1895	0.1433	0.004
Logistic	0.24	0.1958	0.0052
Sigmoid	0.2171	0.1731	0.0046
Silverman	0.2042	0.1577	0.0043
Epanechnikov	0.1328	0.0969	0.0028
Uniform	0.1626	0.1164	0.0034
Triangular	0.1140	0.0837	0.0024
Quartic	0.1131	0.0840	0.0024
Triweight	**0.1001**	**0.0751**	**0.0021**
Tricube	0.1172	0.0874	0.0025
Cosine	0.1290	0.0944	0.0027

Note. Constant bandwidth (2.5) and estimator (LL) were used.

Table 2. Performance comparison of the kernel functions for crisp kernel regression.

5.2.2 CRISP KERNEL FORECASTING

For the forecasting performance comparison, the historical data of the 5 time series is divided into two groups. The first one represents past data which is already known and spans from January until November 2019. The second time period, December 2019, is considered as future and unknown and a two-steps-ahead recursive forecasting is applied to predict the values. The time division was guided considering the variety of data trends observed throughout these time periods.

Table 3 shows a performance comparison of the different kernel estimators. Surprisingly, only in 3 out of 5 time series the Local Linear estimator has superior results. The Nadaraya-Watson may be implemented for special scenarios, but on average the Local Linear estimator has better performance. The fifth time series has considerably lower error measures although having similar RMSSE values. The reason behind is the small value and variation of the Euro US Dollar exchange rate throughout time compared to the equities.

Talking about kernel functions, Table 4 gives a look at their differences when compared with a constant bandwidth value and the same estimator. In a similar way to

Table 2, only three series and two error measures were included as no more insights were provided by the additional information. Reaffirming the statements of *section 3.5.2*, the functions without boundaries have the best performances in the time series 1 and 5. The time series 4 exhibits once more an unusual behaviour. This may be due to the high volatility of the financial equity, which experiments sharp and radical changes during this period. However, the error measures obtained for this time series in combination with the Triangular kernel function are the best of the table. Among the open functions, the Gaussian and the Silverman have the best results.

Table 3

Performance comparison of the kernel estimators for crisp kernel forecasting

	MSE		RMSE		MAE		MAPE		RMSSE	
Series	NW	LL	NW	LL	NW	LL	NW	LL	NW	LL
1	0.1167	0.1140	0.3416	0.3377	0.2662	0.2947	10.945	12.19	1.8462	1.8188
2	0.0514	0.03558	**0.2268**	**0.1891**	0.1662	0.1591	4.0113	3.8280	**1.9651**	**1.6388**
3	0.0405	0.0363	0.2012	0.1906	0.1495	0.1499	6.2321	6.5261	2.4709	2.3411
4	0.7331	1.2714	0.8562	1.1276	0.7042	1.2714	13.257	17.098	**1.4183**	1.8678
5	3.38e-5	4.52e-5	0.0058	0.0067	0.0049	0.0057	0.4380	0.6092	2.0624	2.1842

Note. Optimized bandwidth and constant kernel function (Gaussian) were used.

Table 3. Performance comparison of the kernel estimators for crisp kernel forecasting

Table 4

Performance comparison of the kernel functions for crisp kernel forecasting

	RMSE			RMSSE		
Functions	Series 1	Series 4	Series 5	Series 1	Series 4	Series 5
Gaussian	**0.3377**	1.4741	0.0065	**1.8188**	2.4418	2.3810
Logistic	0.4033	0.8513	0.0059	2.1798	1.4101	2.1809
Sigmoid	0.4070	0.8409	0.0063	2.1996	1.3930	2.3164
Silverman	0.4034	0.9112	**0.0046**	2.1804	1.5094	**1.6778**
Epanechnikov	0.5864	0.7646	0.0094	3.1694	1.2665	3.4356
Uniform	0.5637	0.9015	0.0101	3.0466	1.4934	3.7175
Triangular	0.5714	**0.7576**	0.0092	3.0881	**1.2550**	3.3829
Quartic	0.5444	0.7897	0.0090	2.9424	1.3081	3.2867
Triweight	0.5098	0.8686	0.0086	2.7549	1.4388	3.1660
Tricube	0.5353	0.8213	0.0089	2.8927	1.3605	3.2493
Cosine	0.5803	0.7634	0.0093	3.1363	1.2647	3.4138

Note. Constant bandwidth (5) and estimator (LL) were used.

Table 4. Performance comparison of kernel functions for crisp kernel forecasting.

5.2.3 ITS KERNEL FORECASTING

In this section, two main experiments were carried out. For the first one, a total of 6 time series were selected to test and compare all the different ITS kernel forecasting models. The datasets of the first two, Apple's and Amazon's high and low prices throughout the year 2019, were divided into two and the second halves were used as two-steps-ahead forecasting periods. Furthermore, the other 4 time series belonged to the Euro US Dollar exchange rate throughout 4 different time periods. In a similar way, part of the datasets were used as two-steps-ahead forecasting periods. The first half of the datasets were used as training sets to tune the bandwidth values. The accuracy measures for interval-valued variables described above were employed as performance indicators.

Table 5 shows the performance comparison of the different ITS forecasting models in the first experiment. C and CR stand for center information and center and range information method respectively. Moreover, LL and NW denote the Local Linear and Nadaraya-Watson estimators. The bandwidths were selected following the error measure optimal bandwidth criteria. Additionally, the Gaussian function was set as generic kernel function for two main reasons. The first one is related to its superior performance compared to the others when forecasting. The other one is because of being the most commonly used kernel function in the academic field. Therefore, the results will be easier to compare with other studies in the future.

The results do not show a clear superior forecasting model. Ranked by total number of best error measures, the CR-LL would be the first one closely followed by the CR-NW and C-LL models. The worst performance can be observed for the C-NW model. Furthermore, the change from the Nadaraya-Watson to the Local Linear estimator brings slight improvements in around half of the cases.

Talking about the coverage and efficiency rates, both indicators stay low around 0.2 to 0.4 apart from a few exceptions. On the one hand, the first two time series, which represent two out of the five big tech companies, show poor results for these accuracy measures. Both experience an exponential rise throughout the forecasting period which is even difficult to follow for the Local Linear estimator. On the other hand, the third time series, which is shaped by a mixture of up- and downtrends of the exchange rate throughout the years 2012 and 2013, shows the best results of the table for the CR-NW model, reaching coverage and efficiency rates of 0.79 and 0.81 respectively, thus becoming an exceptional and unique case.

When it comes to comparing the forecasting models with the random walk, the naïve model is only beaten by the C-LL model in the fifth time series. All models closely follow the performance of the random walk in the third, fourth, and fifth time series. Furthermore, the last error measure, the interval average relative variance, stays usually close to 1 and points towards the CR-LL as the superior method.

The second experiment focused on the two superior forecasting models, the CR-NW and CR-LL, to give a deeper insight into the performance of these models in varied currency exchange markets scenarios. Therefore, two out of the four major pairs, the world's most heavily traded currency pairs in the forex (FX) market, were selected. The first one is the US Dollar - Japanese Yen exchange rate whereas the second represents the British Pound Sterling - US Dollar exchange rate. Additionally, the Brent Oil Contract was used to test the models in the energy market and analyse their interdisciplinary behaviour. A comparison of diverse market trends was also achieved by selecting diverse time periods of the three time series. The first one follows a volatile side trend during the year 2019 whereas the second contains a mild down trend throughout the year 2014. The last time series suffers a sharp downtrend during the same year.

In order to test the influence of the forecasting horizon, a one-step-ahead recursive forecast was carried out. The kernel functions were selected by comparing their error measures for each of the scenarios. Furthermore, the error measure optimal bandwidth criterion was employed to select the bandwidth.

Table 5

Performance comparison of the different ITS forecasting models

	iARV				iUTHEIL				CR				ER			
Series	C-NW	C-LL	CR-NW	CR-LL	C-NW	C-LL	CR-NW	CR-LL	C-NW	C-LL	CR-NW	CR-LL	C-NW	C-LL	CR-NW	CR-LL
1	1.1740	0.9770	1.1561	0.8733	1.5318	1.1830	1.6049	1.0693	0.0734	0.2592	0.0294	0.2252	0.1643	0.3182	0.0764	0.3285
2	1.2423	1.0142	0.9240	0.9450	1.7613	1.2797	1.0135	1.0833	0.0708	0.0682	0.0528	0.1197	0.0850	0.1238	0.0722	0.1570
3	0.9842	0.9852	0.9845	0.9844	1.009	1.0201	1.0121	1.0115	0.4048	0.3283	**0.7948**	0.7626	0.3188	0.3512	**0.8102**	0.6568
4	0.9974	1.0007	1.0015	1.0002	1.0044	1.0033	1.0063	1.0012	0.2203	0.3381	0.3667	0.3256	0.2488	0.3598	0.3831	0.3250
5	0.9806	0.9800	0.9796	0.9796	1.0304	**0.9998**	1.0079	1.0037	0.1646	0.3727	0.3137	0.2389	0.1725	0.3784	0.3324	0.2376
6	0.9983	1.0059	1.0013	0.9991	1.0528	1.0842	1.0650	1.0559	0.2958	0.1875	0.2642	0.1837	0.3605	0.1888	0.2948	0.2123

Note. Optimized bandwidths and constant function (Gaussian) were used.

Table 5. Performance comparison of ITS kernel forecasting models

Table 6

Performance comparison of superior ITS forecasting models

	iARV		iUTHEIL		CR		ER	
Series	CR-NW	CR-LL	CR-NW	CR-LL	CR-NW	CR-LL	CR-NW	CR-LL
1	1.0038	1.0015	1.0146	1.0041	0.4933	0.4090	0.4304	0.4181
2	0.9868	0.9869	**0.9991**	0.9996	0.5770	**0.5903**	0.5395	**0.5673**
3	0.9659	**0.9658**	1.0131	1.0046	0.4504	0.5076	0.4268	0.4628

Note. Sigmoid and Gaussian kernel functions and optimized bandwidths were used.

Table 6. Performance comparison of superior ITS kernel forecasting models

The results can be found in Table 6. Compared to the previous study, the performance of both forecasting models experienced a considerable improvement, reaching maximum coverage and efficiency rates of 0.5903 and 0.5673 for the CC-LL model in the second time series (Figure 19). Part of the merit can be attributed to the change from two- to one-step-ahead recursive forecasts. This allows the models to react faster to the market changes by updating the data more frequently as seen in equation (4). The customization of the kernel functions as well as the bandwidths is the second main reason for such an improvement. Whereas in the first experiment the Gaussian function was used, in this occasion the Sigmoid and Gaussian were employed due to their more precise forecasts.

Moreover, the random walk was beaten by both models in this time series, which can be observed in the iUTHEIL values of the second time series. The CC-LL model had better accuracy levels for the two last time series, making the model based on the Local Linear estimator the preferred choice.

Talking about the behaviour of the models in different trends, the best performance can be seen in mild down trends, which take place during the second time series (Figure 19). The strong bearish trend of the last time series follows (Figure 20) and the last place is occupied by the side trend represented in the first time series (Figure 18).

It can be deduced that in the periods where the data followed a notable slope, the Local Linear estimator had superior results, whereas the Nadaraya-Watson had more accurate predictions for stable side trend time series. These results support the own nature of both estimator's weighting scheme.

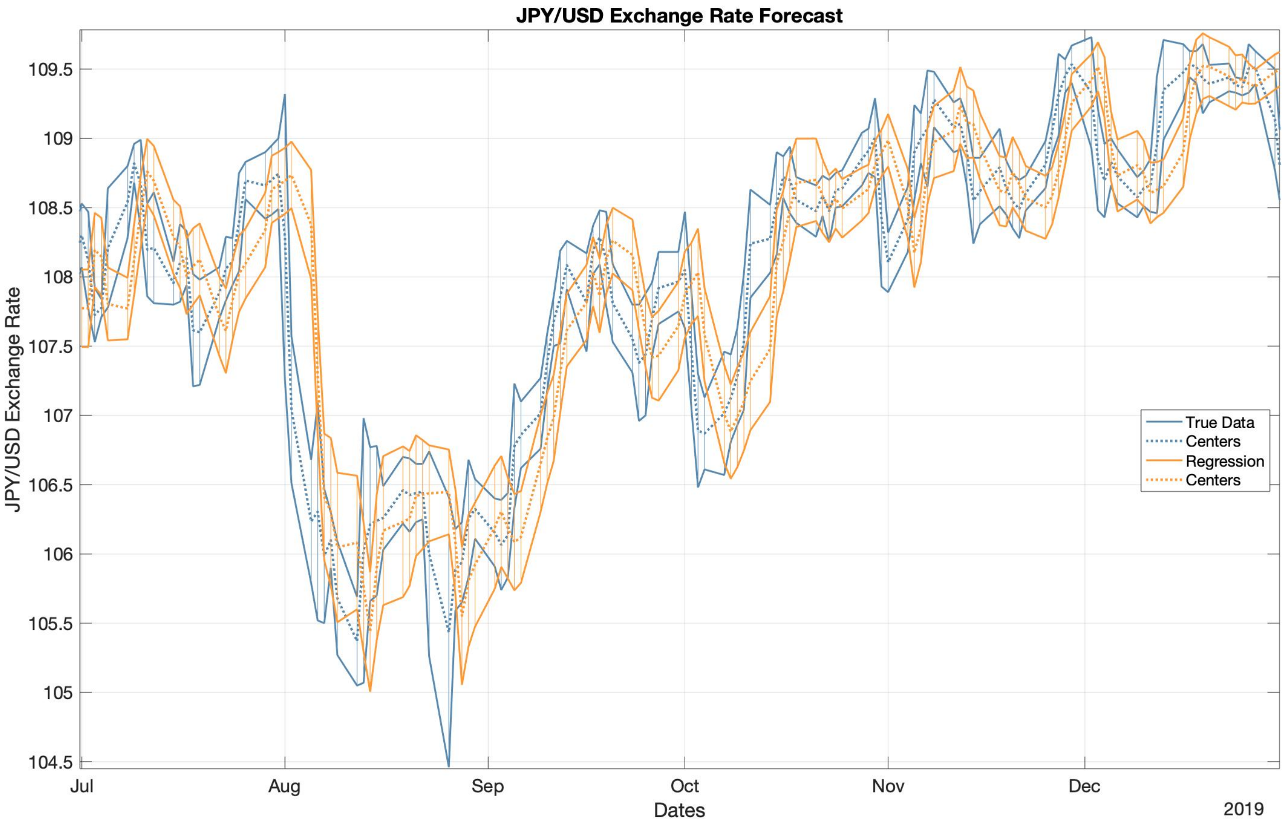


Figure 18. JPY/USD exchange rate CR-NW ITS kernel forecast.

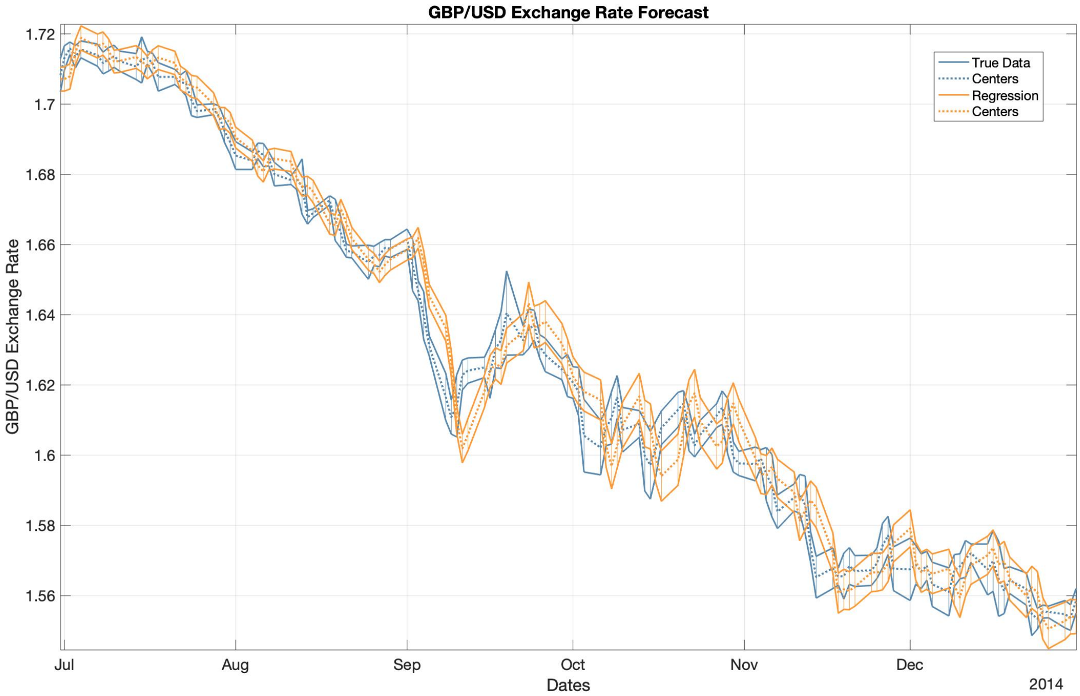


Figure 19. GBP/USD exchange rate CR-LL ITS kernel forecast.

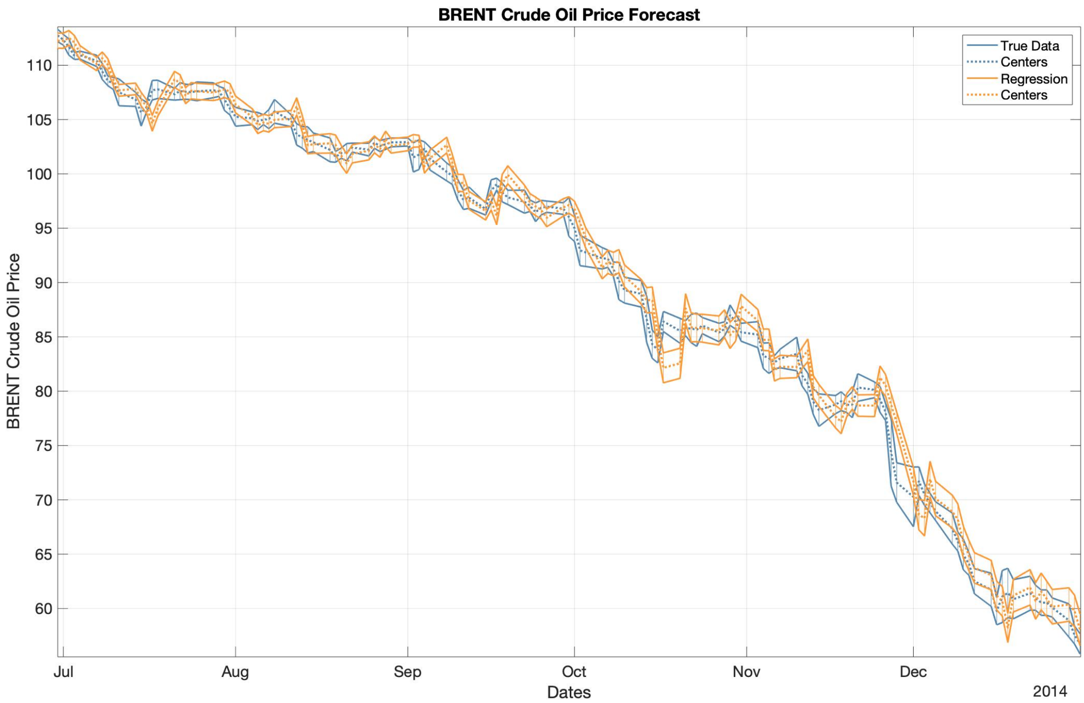


Figure 20. Brent crude oil CR-LL ITS kernel forecast.

The three last figures, Figure 18, Figure 19, and Figure 20, show the final result of the project, which is a mixture of two forecasting models of interval-valued time series in the financial and energy markets. Extending what has been said about them above, a satisfactory covering rate can be seen in all of them, specially for the last two. Excluding some little exceptions, the forecasted interval always accomplishes to match part of the real interval, which gives safety and robustness when using the model in real life. Throughout the three time series, most inaccuracies are found when drastic price changes take place, for example during October of Figure 20, or August of Figure 18. Therefore, when the prices experience mild to medium variances, the models show to have a good reaction.

CHAPTER 6. CONCLUSIONS AND FUTURE RESEARCH

6.1 CONCLUSIONS

In this research, kernel regression and forecasting models for crisp and interval-valued time series have been introduced, tested, and compared. A high customization of the models has been achieved by introducing and offering a diverse range of parameters' alternatives. The implementation of all these features to MATLAB has been carried out. A total of 6500 lines of code distributed in 37 scripts have been written, serving as a potential foundation of a future kernel regression and forecasting visual tool.

Furthermore, under the best knowledge of the author, the field of ITS kernel forecasting has been explored for the first time with the implementation of two original forecasting models. An own elaborated visualization for financial interval-valued time series forecasting showing the closing as well as high and low prices of a financial asset has been proposed and supplements the use of candlestick charts by providing additional information.

Talking about the performance of the models, the mixture of closed boundary kernel functions with the Local Linear estimator showed the highest accuracy levels for crisp kernel regression. On the opposite side, crisp kernel forecasting was optimized when combining open boundary kernel functions with the Local Linear estimator.

Out of the four ITS kernel forecasting models, the two based on the center and range information method elaborated the best predictions. The combination of this method with the Local Linear estimator and the Gaussian or Sigmoid kernel functions showed the best performance in most of the cases. In special scenarios of prolonged side trends, the Nadaraya-Watson estimator showed better results.

There are two common conclusions concerning all models. The first one is related to the number of steps ahead of the predictions. The smaller the number, the better results were observed in the forecasts. Secondly, the error measure optimal bandwidth selection criterion showed superior results compared to the Bowman and Azzalini's criterion in all cases.

The random walk was beaten only in some exceptional cases. As seen in other nonparametric forecasting studies, to outperform the random walk is a difficult task which becomes even more complex when dealing with interval-valued time series. Maté and Jiménez (2021) proposed the Interval Multi-Layer Perceptron forecasting model, which variant for crisp time series shows one of the best performances for the Euro US Dollar exchange rate forecast. Despite of being considerably more complex, this model also suffered when trying to outperform the random walk, only achieving it in special occasions. This points towards a satisfactory performance of the kernel models when considering a balance between results and complexity.

6.2 FUTURE RESEARCH

In this section, some guidelines for possible future research are introduced. The first one is related to the kernel estimator. Studying the implementation of the Local Polynomial estimator could potentially bring considerable improvements. However, this new estimator would rise the complexity of the model.

One of the most important tuning parameters, the bandwidth, could be further studied by comparing a wider range of selection processes. The paper by Köhler et al. (2014) could serve as a guideline.

Furthermore, the ITS kernel regression and forecasting models could be compared with more classical regression and forecasting models. The research of Maceda and Maté (2018) provides a good source of some classical linear parametric models.

Finally, the models presented in this study could be tested in other disciplines such as weather forecasting in order to better understand their performance in less volatile datasets.

CHAPTER 7. BIBLIOGRAPHY

Arroyo, J., Espínola, R. and Maté, C. (2011), 'Different approaches to forecast interval time series: a comparison in finance', *Computational Economics* 37(2), 169–191.

Bates, J. M. and Granger, C. W. J. (1969), '*The Combination of Forecasts*', *Journal of the Operational Research Society* 20, 451–468.

Billard, L. and Diday, E. (2000), *'Regression analysis for interval-valued data'*, *Data Analysis, Classification, and Related Methods*, (pp. 369-374), Berlin, Springer-Verlag.

Billard, L. and Diday, E. (2002), *'Symbolic regression analysis'*, *Classification, Clustering and Data Analysis,* (pp. 281-288), Berlin, Springer-Verlag.

Bowman, A. W. and Azzalini, A. (1997), 'Applied Smoothing Techniques for Data Analysis: The Kernel Approach with S-Plus Illustrations', *Volume 18 of Oxford Statistical Science Series.*

Chacón, J. E. and Duong, T. (2018), '*Multivariate Kernel Smoothing and Its Applications*' 1st edition, *Chapman and Hall/CRC.*

Fagundes, R., De Souza, R., and Cysneiros F. (2014),*' Interval kernel regression', Neurocomputing*, 128, 371-388.

Hastie, T., Tibshirani, R. and Friedman, J. (2009), *'The Elements of Statistical Learning; Data Mining, Inference and Prediction'* 2nd edition, *Springer*.

Im, S. and Kang, K. (2018), '*On regression analysis of interval-valued data'*, *Journal of the Korean Data & Information Science Society*, 29, 351-365.

Jang, J. and Kang, K. (2020), *'Local linear regression analysis for interval-valued data', The Korean Statistical Society, Communications for Statistical Applications and Methods,* 27, 365-376.

Köhler, Max & Schindler, Anja & Sperlich, Stefan, (2014), *'A Review and Comparison of Bandwidth Selection Methods for Kernel Regression'*, International Statistical Review.

Lima Neto, E. A., De Carvalho, F. and Tenorio, C. (2004), *'Univariate and multivariate linear regression methods to predict interval-valued features'*, *Advances in Artificial Intelligence 2004* (Vol. 3339, pp. 526-537), Berlin, Heidelberg, Springer.

Lima Neto, E. A. and De Carvalho, F. (2017), *'Nonlinear regression applied to interval-valued data', Pattern Analysis and Applications*, 20, 809-824.

Maceda, P. Maté, C. (2018), '*Basic Analytical System for Interval Valued Data*', *Comillas Pontificial University – ICAI.*

Maia, A. L. S. and de Carvalho, F. d. A. (2011), *'Holts exponential smoothing and neural network models for forecasting interval-valued time series'*, *International Journal of Forecasting* **27**(3), 740–759.

Makridakis, S., Spiliotis, E. and Assimakopoulos, V. (2020), *'The M4 competition: 100,000 time series and 61 forecasting methods'*, *International Journal of Forecasting* 36(1), 54–74.

Makridakis, S. and Evangelos & Assimakopoulos, V. (2020), *'The M5 Accuracy competition: Results, findings and conclusions', International Journal of Forecasting*

Maté, C. (2021), '*Combining Interval Time Series Forecasts. A First Step in a Long Way (Research Agenda)', Revista Colombiana de Estadística - Applied Statistics* 44, 123-157.

Maté, C. and Jimeńez, L. (2021), *'Forecasting exchange rates with the iMLP: New empirical insight on one multi-layer perceptron for interval time series (ITS)', Engineering Applications of Artificial Intelligence,* 104358

Noirhomme-Fraiture, M. and Brito, P. (2011), 'Far beyond the classical data models: symbolic data analysis', *Statistical Analysis and Data Mining: the ASA Data Science Journal* 4(2), 157–170.

Rodrigues, P. M. and Salish, N. (2015), *'Modeling and forecasting interval time series with threshold models'*, *Advances in Data Analysis and Classification* **9**(1), 41–57.

Tao, X., Yukun, B., Zhongyi, H. and Raymond, C. (2015), '*Forecasting interval time series using a fully complex-valued RBF neural network with DPSO and PSO algorithms*', *Information Sciences* 305, 77-92.

Vilar-Fernández, J. & Cao, R. (2007*), 'Nonparametric Forecasting in Time Series—A Comparative Study', Communications in Statistics—Simulation and Computation®*, 36:2, 311-334

Wand, M. P. and Jones, M.C. (1995), '*Kernel Smoothing*' 1st edition, *Chapman and Hall/CRC.*

Wang, L. and Zhu, J. (2008), '*Financial market forecasting using a two-step kernel learning method for the support vector regression*', Springer 174, 103-120.