REPUBLIC OF TURKEY
ÇUKUROVA UNIVERSITY
INSTITUTE OF SOCIAL SCIENCES
DEPARTMENT OF ECONOMETRICS

NONPARAMETRIC AND QUANTILE REGRESSION APPROACHES: ENERGY AND COMMODITY MARKET LINKS

Sera ŞANLI

DOCTORAL DISSERTATION

ADANA / 2020

REPUBLIC OF TURKEY
ÇUKUROVA UNIVERSITY
INSTITUTE OF SOCIAL SCIENCES
DEPARTMENT OF ECONOMETRICS

NONPARAMETRIC AND QUANTILE REGRESSION APPROACHES: ENERGY AND COMMODITY MARKET LINKS

Sera ŞANLI

Supervisor: Prof. Dr. Mehmet ÖZMEN

Second Supervisor: Prof. Dr. Mehmet BALCILAR

Member of Examining Committee: Prof. Dr. H. Altan ÇABUK

Member of Examining Committee: Prof. Dr. Erkut DÜZAKIN

Member of Examining Committee: Prof. Dr. Bülent GÜLOĞLU

Member of Examining Committee: Prof. Dr. A. Özlem ÖNDER

DOCTORAL DISSERTATION

ADANA / 2020

To Directorate of the Institute of Social Sciences of Çukurova University;

We certify that this dissertation is satisfactory for the award of the degree of Doctor of Philosophy in the Department of Econometrics.

Chairperson: Prof. Dr. Mehmet ÖZMEN
(Supervisor)

Second Supervisor: Prof. Dr. Mehmet BALCILAR

Member of Examining Committee: Prof. Dr. H. Altan ÇABUK

Member of Examining Committee: Prof. Dr. Erkut DÜZAKIN

Member of Examining Committee: Prof. Dr. Bülent GÜLOĞLU

Member of Examining Committee: Prof. Dr. A. Özlem ÖNDER

I certify that this dissertation conforms to the formal standards of the Institute of Social Sciences and I confirm that these signatures above belong to the academics mentioned. …/…/2020

Prof. Dr. Serap ÇABUK
Director of the Institute

NOTE: The uncited usage of the reports, charts, figures and photographs in this dissertation, whether original or quoted from other sources, is subject to the Law of Works of Arts and Thought No: 5846.

ETHICS DECLARATION

In this dissertation study that I have prepared in accordance with the Çukurova University Institute of Social Sciences Thesis Writing Guidelines; I hereby declare that

- I have obtained all data, information and documents that I have presented in this dissertation within the framework of academic rules and ethical conduct,
- I have presented all information, documentation, evaluations and findings in accordance with scientific ethical and moral principles,
- I have properly cited and referenced all sources that I have used in the dissertation,
- I have not made any changes in the data used and the findings obtained,
- This dissertation represents my original work and has been written entirely by me,

I hereby acknowledge all possible losses of rights that will be able to arise against me in the case of a contrary circumstance (in the case of any circumstance contradicting with my declaration). …. / …. / 2020

Sera ŞANLI

ÖZET

PARAMETRİK OLMAYAN VE KANTİL REGRESYON YAKLAŞIMLARI: ENERJİ VE EMTİA PİYASASI İLİŞKİLERİ

Sera ŞANLI

Doktora Tezi, Ekonometri Ana Bilim Dalı
Danışman: Prof. Dr. Mehmet ÖZMEN
İkinci Danışman: Prof. Dr. Mehmet BALCILAR
Eylül 2020, 143 sayfa

Parametrik olmayan regresyon, tahmin edilecek nesneler için fonksiyonel biçimlerin açık bir tanımlamasını gerektirmeyerek veri modellemesinde büyük esneklik sunması bakımından büyük önem taşımaktadır; böylelikle veri üretim süreci üzerine konan parametrik varsayımları gevşetmektedir. Parametrik olmayan regresyonun önemi ele alınarak, bu tez üç çalışmadan oluşmuştur ve çağdaş ekonomiler için büyüme ve rekabetçi olma kalitesini etkileyen itici güçlerden biri olarak kabul edilebilen enerji ve emtia piyasalarını analiz etmek için esas olarak parametrik olmayan regresyon, parametrik olmayan kantil tahmini ve parametrik olmayan kantil regresyon yaklaşımlarını uygulamayı amaçlamaktadır. Birinci çalışma; ham petrol, teknoloji ve borsa kompozit fiyat endekslerinden temiz enerji fiyat endekslerine doğru giden çok değişkenli nedensellikleri saptamak ve seri yaklaştırmaları yöntemine dayanarak koşullu kantil fonksiyonunun ortalama türev tahminlerini elde etmek için parametrik olmayan kantilde nedensellik yaklaşımını uygulamayı amaçlamıştır. İkinci çalışma, Klemelä (2017) tarafından yapılan çalışmaya dayanarak Bitcoin getirilerinin volatilitesini farklı tahmin ufukları için karesi alınmış ve orijinal getirileri kullanarak tahmin etmekle -ve aynı zamanda kantil tahminini uygulamakla- ilgilenmiştir. Üçüncü çalışma ise altın, ham petrol ve S&P500 hisse senedi fiyatları arasındaki dinamik davranışları 7-gün ilerisi için tahminlerde parametrik olmayan vektör otoregresyon yaklaşımını kullanarak ve bazı rekabetçi modellerle bir karşılaştırma sunarak araştırmıştır.

Anahtar kelimeler: Parametrik olmayan regresyon, kantilde nedensellik, ortalama türev, volatilite tahmini, enerji ve emtia piyasaları.

ABSTRACT

NONPARAMETRIC AND QUANTILE REGRESSION APPROACHES: ENERGY AND COMMODITY MARKET LINKS

Sera ŞANLI

Ph.D. Thesis, Department of Econometrics
Supervisor: Prof. Dr. Mehmet ÖZMEN
Second Supervisor: Prof. Dr. Mehmet BALCILAR
September 2020, 143 pages

Nonparametric regression is of great importance with respect to offering great flexibility in data modelling by not requiring a clear description of functional forms for estimated objects, thus relaxing the parametric assumptions imposed on the data generating process. Considering its significance, this dissertation comprises three essays and mainly aims to carry out nonparametric regression, nonparametric quantile estimation and nonparametric quantile regression approaches in order to analyze energy and commodity markets which can be regarded as one of the driving forces affecting the quality of being competitive and growth for contemporary economies. The first essay has aimed to perform the nonparametric causality-in-quantile framework in order to detect the multivariate causations running from crude oil, technology and stock market composite price indexes towards clean energy price indexes and also obtain the average derivative estimates of the conditional quantile function based on series approximations method. The second essay has dealed with predicting the volatility of Bitcoin returns using squared and original returns as proxies for volatility based on the study by Klemelä (2017) -and also performing the quantile estimation- for different prediction horizons. The third essay has explored the dynamic behaviors between gold and the drivers of gold -namely, crude oil and S&P500 stock prices- using the nonparametric vector autoregression approach in predictions and presenting a comparison with some competitive models for 7-day ahead forecasting.

Keywords: Nonparametric regression, causality-in-quantile, average derivative, volatility prediction, energy and commodity markets.

ACKNOWLEDGEMENTS

First of all, I would like to express my endless thanks to my esteemed supervisor Dear Prof. Dr. Mehmet ÖZMEN who has always showed understanding towards me in my dissertation process, always supported my studies with his valuable opinions and contributions and whom I am honored to know him as my supervisor since my master's degree.

I would like to pour out my thanks to my second supervisor Dear Prof. Dr. Mehmet BALCILAR who always stood by me with his scientific contributions and excellent guidance from the title of the thesis to its content and analysis, made me feel privileged because of working together and whom I am proud of him for being my supervisor. I would never have been able to finish this dissertation without the guidance of him.

In addition, I am sincerely grateful to Dear Prof. Dr. H. Altan ÇABUK, Prof. Dr. Erkut DÜZAKIN, Prof. Dr. Bülent GÜLOĞLU and Prof. Dr. Özlem ÖNDER who accept to take place in my dissertation defense.

I am also deeply grateful to TUBITAK (The Scientific and Technological Research Council of Turkey) – BIDEB (Scientist Support Department) and their personnel for their financial support throughout my doctoral process by awarding me within the scope of 2211-E Direct National Scholarship Programme for PhD Students (Grant No: 2211-E Doğrudan Yurt İçi Doktora Burs Programı). They deserve a great appreciation for being a great motivation source to me to focus on my studies.

Finally, I am infinitely indebted to my family who has always supported me throughout my life with their loves and best wishes and contributed so much to the preparation of me to the future.

This thesis is dedicated to my mother and my brother.

TABLE OF CONTENTS

CHAPTER II

PREDICTING THE VOLATILITY OF BITCOIN RETURNS BASED ON KERNEL REGRESSION

CHAPTER III

NONPARAMETRIC MARKET DYNAMICS AMONG GOLD, CRUDE OIL AND S&P500 STOCKS

LIST OF ABBREVIATIONS

aGARCH: Semiparametric Additive Autoregressive Structure

AIC: Akaike Information Criterion

AMISE: Approximate Mean Integrated Square Error

ARCH: Autoregressive Conditional Heteroscedasticity

ARMA: Autoregressive Moving Average

ASH: Average Shifted Histogram

CCS: Carbon Capture and Storage

CHARN: Conditional Heteroscedastic Autoregressive Nonlinear

CI: Confidence Interval

CSSPE: Cumulative Sums of Squared Prediction Errors

DGP: Data Generating Process

ECO: WilderHill Clean Energy Index

EGARCH: Exponential Generalized Autoregressive Conditional Heteroscedasticity

ETF: Exchange Traded Fund

EWMA: Exponentially Weighted Moving Average

GARCH: Generalized Autoregressive Conditional Heteroscedasticity

GDP: Gross Domestic Product

GHG: Greenhouse Gas

HN: Heston-Nandi

i.i.d.: Independent and Identically Distributed

IMSE: Integrated Mean Square Error

IRLS: Iteratively Reweighted Least Squares

KDE: Kernel Density Estimation

LA-VAR: Lag-Augmented Vector Autoregression

LP: Local Polynomial

LSCV: Least Squares Cross Validation

MA: Moving Average

MR: Many Regressors

MSE: Mean Square Error

NCHARN: Nonparametric Conditional Heteroscedastic Autoregressive Nonlinear

NP: Nonparametric

NR: Normal Reference

NYMEX: New York Mercantile Exchange

NYSE: New York Stock Exchange

PDF: Probability Density Function

QR: Quantile Regression

QTARCH: Qualitative Threshold Autoregressive Conditional Heteroscedasticity

RF: Random Forest

RMSE: Root Mean Square Error

SIC: Schwarz Information Criterion

TGARCH: Threshold Generalized Autoregressive Conditional Heteroscedasticity

VaR: Value-at-Risk

VAR: Vector Autoregressive (Autoregression)

VAR-NN: Vector Autoregressive Neural Network

VIX: Implied Volatility Index

WTI: West Texas Intermediate

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER I

CLEAN ENERGY, OIL, TECHNOLOGY AND STOCK MARKET LINKS: EVIDENCE FROM NONPARAMETRIC QUANTILE CAUSALITY TESTS AND AVERAGE DERIVATIVE ESTIMATES

1.1. Introduction

Density estimation approaches can be expressed as parametric or nonparametric. In parametric approach where the data come from a known distribution family, unknown parameter values are tried to be estimated. Under holding assumptions, estimates generated through parametric approaches will be more accurate than the ones generated from nonparametric approaches. However, when assumptions are not satisfied, the usage of parametric method will lead to a larger error (Tu & Li, 2017, p. 447). Nonparametric approaches have an advantage in that they do not necessitate to clearly describe the functional forms of the models in interest by relaxing the parametric assumptions and therefore are very flexible in usage. Through these methods, one can understood which knowledge the data present about the underlying data generating process (DGP) in a specific way in cases where parametric specifications presumed are detected not to be coherent with the dataset at hand and do not give an insight about outperforming parametric models after implementing model adequacy tests. Kernel smoothing methods -where the concept of kernel implies a weighting function- as one of the mostly used nonparametric procedures perform in a good manner in order to be able to cope with large data sets (Racine, 2008, p. 1) and these methods have gained an intense attention recently -especially as one of the mainstream approaches to machine learning researches apart from being employed also in many far-reaching application fields like artificial intelligence, bioengineering, computational biology, image and audio processing, genomics, pattern recognition, computer vision, finance etc.- by reason of yielding a sound performance of real world applications relative to their rivals. One can regard kernel methods as two-stage state-of-the-art procedures which contain transforming the input patterns from observational space into a high-dimensional kernel feature space where the data set is likely to become linearly separable by capturing nonlinear

associations of the original data first and then carrying out linear algorithms on the feature space (Camps-Valls, Rojo-Álvarez, & Martínez-Ramón, 2007, p. viii).

By and large, communities having high-energy flows hold the opportunity to acquire more equity through taxation of high incomes (or taxation on carbon-based energy consumption by targeting affluent consumers first) (Heinberg & Friedly, 2016, pp. 152-153) or government redistributive programs. However, some least-industrialized nations (especially in parts of sub-Saharan Africa and Central America) are not so lucky due to having a limited access to basic needs such as food, clean water, health opportunities, clothing, electricity or clean cookstoves etc. and therefore suffering from a severe poverty. This situation requires for increasing their total and per capita energy consumption to be able to obtain a barely sufficient standard of living. In order to eradicate the issue of poverty, traditional economic development targets to utilize from the routes that affluent countries follow to get their wealth like resource extraction, production, urbanization, free-market policies or debt - each depending on continuously increasing fossil-fuel consumption. Those some rapidly industrializing countries (e.g., China, India and parts of Southeast Asia) -which adopt the understanding of competitive production using cheap labor and cheap energy (usually from indigenous coal), experience accordingly increasing levels of greenhouse gas (GHG) emissions and are heavily connected to fossil fuel-dependent infrastructure- have significantly increased local economic inequality apart from health and environmental costs (Heinberg & Friedly, 2016, pp. 146-150). Therefore, the methods of fulfilling the existing goals of such societies require an overall revision. Depending also on the rapid exhaustion of high-qualified and cheap fossil fuels (Heinberg & Friedly, 2016, p. 130), a better target for overcoming these negativities will be to catch the sustainable level of energy consumption per capita in the long run through renewable energy (solar panels, wind turbines, energy storage) sources in a way that will help ensure economic equality (Heinberg & Friedly, 2016, pp. 150-153). Getting rid of fossil fuel dependency and adopting renewable energy technologies have certain advantages and disadvantages in terms of society. Renewable energy has many facilities that can be accessed with the use of fossil fuels, e.g. solar and wind for electricity generation, battery powered cars or biofuels for transportation, solar thermal system for heating etc. Moreover, hereby it will be exempted from financial and social costs in processes such as the extracting, refining or transporting of exhausting fossil fuels and the environmental externalities that may arise as a result of combustion of these fuels. Solar concentrators or hydrogen (generated from renewable electricity) can

be used to operate industrial processes with high-heat. In addition, the electricity generation from solar and wind does not necessitate a conversion process like the one required in the case of coal and natural gas; thus, not mentioning great losses of primary energy and keeping energy efficiency at a high level (Heinberg & Friedly, 2016, p. 124). However, most of these substitutions are likely to require significant research and development (R&D) and therefore a cost which creates a disadvantage (Heinberg & Friedly, 2016, pp. 139-141). Besides; although the decrease in power plant emissions as a result of the enlargement of renewable energy creates an improvement in the level of welfare due to the drop of hospital stays and medicine prescriptions associated with asthma crises, it gives rise to a decline in gross domestic product (GDP) which is also the case naturally having a direct impact on economic growth (Heinberg & Friedly, 2016, p. 130). One of the situations making the transition from fossil-fuel energy to renewable energy compelling is undoubtedly their effects on economic growth such that while the previous that is inexpensive and plentiful makes the development of consumption-oriented growth economy possible, the next will likely fail to maintain this type of economy (Heinberg & Friedly, 2016, p. 196). To sum up, a rise in energy consumption is said to be related to economic growth so that generally, energy consumption tends to fall in economic recession periods and unfortunately, what the policy makers truly desire is not to abandon their economic growth expectations for obtaining a trade-off between fossil energy and solar & wind energy by cutting down on fossil energy resources (Heinberg & Friedly, 2016, p. 126).

It is remarkable to say that energy is closely related to social evolution, inequality and justice concepts. Fix (2019) has revealed important findings regarding the relationship between modern societies and energy such that increased energy flows for such societies which are organised hierarchically encourage institutions to become larger and the growth in institution size results in societies becoming more hierarchical which also implies the accumulation of resources in the top layer. Through this accumulation, the increase in the energy usage leads to the increase in income which is proportional to hierarchical power, therefore creating an increase in (income) inequality that is nearly realized via a transition from subsistence to agrarian levels of energy use. Briefly, an inequality boom manifests itself in societies with high governing factor which is defined as the rate that income scales with hierarchical power. Thus, Fix (2019) suggests the "energy-hierarchy-inequality (EHI)" hypothesis that will reveal the origin of inequality. According to the findings of the model based on this hypothesis, there exists a common

relationship between energy use, hierarchy and inequality. That study touches an important point as well that energy usage exceeding agrarian levels generates a negligible impact on inequality removing the concentration of hierarchical power (Fix, 2019, pp. 2, 6, 9). In addition, the philosopher named Ivan Illich in his book “Energy and Equity” states too that more energy abundance implies a decrease in the distribution of control over that energy (Heinberg & Friedly, 2016, p. 146).

Doubtless energy -and specifically access to clean energy- is of great significance for many nations. To state clearly; clean energy consists of multifarious industries, not just an individual industry. Some large sub-sectors under clean energy can be expressed as solar power, wind power, fuel cells, geothermal power, biofuels, cleaner coal or power efficiency. One core point is that these sectors should be evaluated seperately in terms of their causal agents, technologic infrastructures and thus profit models (Asplund, 2008, p. 2). It is known that the dream of clean coal has adorned the advertising and projects of Americans for years. Nowadays, however, almost no nation has acquired coal-fueled "clean" power plants due to high-priced “carbon capture and storage (CCS)” process [CCS does not represent a type of clean energy, but a path for making dirty energy cleaner -a transition to “cleaner coal”- (Heal, 2009, p. 21).] which may require laying pipelines or compressors in accordance with the size of the oil industry and unfortunately, some societal effects of coal (for instance, coal dust exposed in the mining process and high rates of lung disease) cannot be accounted for by CCS in the strict sense due to not removing all pollutants from the combustion process. On the other hand, because of the exposure of non-renewable resources (such as coal and natural gas) to the risk of exhaustion, the mining of resources with low-quality and depending on this complicated extraction technologies will be required resulting in increased production costs (Heinberg & Friedly, 2016, pp. 135-137). To sum up, obtaining energy from cleaner sources is of vital importance in order to cope with various issues arisen from catalysts for the clean energy industry which include *fossil fuel negatives* that can be identified with pollution and global warming threat -created via burning fossil fuels and subsequently realized CO_2 emissions, GHGs or distinct pollutants that are harmful for atmosphere-, drastic price spikes and surges (specifically, U.S. has gone through five recessions directly due to the crude oil price spikes for the periods 1973-1975, 1980, 1990-1991) or cartels [e.g. Organization of Petroleum Exporting Countries (OPEC) cartel making high prices guarantee for revenue-maximization by restricting the supply of crude-oil]; *energy security; growing global energy demand; GHG emissions and climate change; clean*

energy technology improvements or rising electricity prices (Asplund, 2008, pp. 27-41). Therefore, potent policies should be enacted by governments through energy taxation of fossil fuels (a carbon tax or taxation of GHG emissions), subsidies to renewable energy or improvisations in technological innovations that serve for cutting down the GHG emissions in order to be able to offer better investment capabilities which will provide an efficient transition towards clean energy.

Considering the importance of clean energy, this study makes use of the recently developed nonparametric testing procedure of Granger-causality in quantiles in order to cover the multivariate causations of New York Mercantile Exchange (NYMEX) Light Crude Oil Continuous Contract Settlement Prices, New York Stock Exchange (NYSE) Arca Technology Price Index and S&P 500 Composite Price Index on WilderHill Clean Energy Price Index and S&P Global Clean Energy Price Index variables (all expressed in U.S. Dollars) that have been extracted out of Thomson Reuters database and the framework of nonparametric quantile estimation has been based upon series approximations scheme. To this end, this paper has been structured in that way: Section 1.2 mainly presents a summary of literature studies on energy and factors affecting it, Section 1.3 introduces the approaches to nonparametric density estimation, Section 1.4 reports the empirical findings and Section 1.5 gives a summary of concluding remarks.

1.2. Literature Review

As related to the nonparametric estimation, Stone (1977) has examined the consistency conditions of probability weight functions with respect to nearest neighbors and convergence rates for a large family of nonparametric estimators. Chaudhuri (1991) has introduced the asymptotic properties for nonparametric estimation of conditional quantile function and its derivatives through locally polynomial fits. Bahraoui et al. (2018) have aimed to improve the convergence rate of mean square error (MSE) for Nadaraya quantile estimator via a modified estimator by making a comparison of MSEs of the estimated quantile for Weibull, Lognormal and Mixture of Lognormal-Pareto distributions. Simulation studies have confirmed that the improvement becomes larger by increasing the sample size.

There exist a vast number of studies on energy and factors affecting it in the literature. Henriques and Sadorsky (2008) have handled the empirical relationship between alternative energy stock prices (WilderHill Clean Energy Index (ECO)), technology stock prices (Arca Technology Index), oil prices (U.S. West Texas

Intermediate (WTI) crude oil futures prices) and interest rates. Results have revealed that there exists a Granger causality running from technology and oil prices towards the alternative energy stock prices depending on the computations from Toda and Yamamoto (1995)'s lag-augmented vector autoregression (LA-VAR) Wald tests. Besides, alternative energy stock prices are more influenced by a shock in technology stock prices relative to a shock in oil prices based on simulation results implying that oil prices shocks are not as important as technology stock price shocks. On the other hand, only shocks in interest rates have a statistically significant effect on oil prices. Kumar, Managi and Matsuda (2012) have analyzed the linkages between oil prices and alternate energy stock prices with a LA-VAR model consisting of stock price indexes of clean energy firms, technology stock prices, oil prices, carbon prices and short-term interest rates. They have reported that the stock prices of clean energy firms are affected by stock prices of high technology firms and oil prices each individually but not by carbon prices. All indexes of clean energy stocks have also been explained by past movements in interest rates according to Granger causality Wald test statistics. Technology stock prices having a positive effect on clean energy stocks in the study have also been affected by the clean energy stocks represented by the WilderHill New Energy Global Innovation Index & ECO and by the variations in oil prices and interest rates. Any of the variables covered in the study could not account for carbon prices. A Granger causal relationship has also been detected from interest rates towards oil prices. Besides, the movements in interest rates have been explained by technology and clean energy stocks apart from oil prices. Sadorsky (2012a) has aimed to analyze conditional correlations & volatility spillovers between oil prices and the stock prices of clean energy companies & technology companies using four different multivariate generalized autoregressive conditional heteroscedasticity (GARCH) models -BEKK, diagonal, constant conditional correlation & dynamic conditional correlation- where the data covered in the research are the daily closing prices of ECO, NYSE Arca Technology Index & the nearest contract to maturity on the WTI crude oil futures contract. The best performing model has been detected as dynamic conditional correlation model. According to the results, the stock prices of clean energy companies are much more associated with technology stock prices when compared to oil prices and oil futures (but not technology stocks) can be a useful hedge for clean energy stocks. Sadorsky (2012b) has aimed to examine the determinants of renewable energy company risk where the list of renewable energy companies was obtained from WilderHill Clean Energy Exchange Traded Fund (ETF) using a variable

beta model which is an extension of the capital asset pricing model and found that oil price increases have a positive effect on company risk. It is also noted that in the case of positive and moderate oil returns, the impact of oil price returns is balanced through the increases in firm sales growth by producing a lower systematic risk. Managi and Okimoto (2013) have aimed to investigate the dynamic relationships among oil prices, clean energy stock prices and high-technology stock prices taking possible structural breaks in the market into consideration by carrying out Markov-Switching vector autoregressive (VAR) models, also using the short-term interest rates. For measuring the stock market performance of clean energy firms, it has been utilized from ECO. Results have shown that structural changes have been detected to exist during late 2007 corresponding exactly to the period in which a remarkable rise in oil price was observed and the U.S. economy entered a recession. Contrary to the past studies existing in the literature, they have reported a positive relationship between oil prices and clean energy prices subsequent to structural changes implying a movement from conventional energy to clean energy. In brief, the link between oil prices and clean energy markets has been influenced by structural changes. Ercen (2015) has analyzed the relationship between U.S. Dollar, WTI crude oil price, solar index, wind index and ECO through VAR model. Results have shown that a long-run comovement has been detected between oil and renewable energy index prices. According to the results, it is stated that the market share of renewable energy has shown an increase through technological development and renewable index prices have been affected by general demand shocks; however, declining oil prices do not create a damaging effect on renewables - particularly on solar index prices. Milioti (2017) has investigated whether alternative energy stock prices are influenced by fossil fuel price fluctuations (specifically, natural gas and WTI oil prices) or not by performing a LA-VAR analysis and it has been found no significant relationship between the alternative energy index and fossil fuels for the weekly and monthly frequency excluding the daily frequency according to the Granger causality test results. In addition, the performance of the alternative energy index is explained by the past movements in WTI oil prices (but not in natural gas prices). Reboredo, Rivera-Castro and Ugolini (2017) have investigated co-movement and causal relationships between oil and renewable energy stock returns at different time scales utilizing from continuous and discrete wavelets to examine the dynamic correlations over time and both linear and nonlinear Granger causal links using the tests -proposed by Hiemstra and Jones (1994) and later modified by Diks and Panchenko (2006)- in the time-frequency domain. Data covered in the analysis for the

period January 2006 to March 2015 are specifically renewable energy indices (ECO, S&P Global Clean Energy Index and European Renewable Energy Index), renewable energy sectoral indices (the NYSE Bloomberg Global Wind, Solar Energy and Smart Technologies indices) and WTI oil prices. Results have revealed that the dynamic interaction between oil prices and renewable energy returns is more powerful towards the long-term even though there exist some differences between general and sectoral indices and different time periods. Linear causal relations have been detected at lower frequencies -but not for higher frequencies- as bidirectional. However, nonlinear causality from renewable energy indices towards oil prices has been observed at distinct time horizons. Besides, oil can be said to be able to be used as a hedge for renewable energy investments but not to forecast renewable energy stock returns. As a result, policymakers should invest greater efforts in promoting short-run development of renewables; reinforce these efforts in the long run in case oil prices are low and alleviate them when oil prices are high. Dutta (2018) has investigated the association between crude oil and U.S. energy sector stock prices using implied volatility indexes (VIX) for totally 338-week observations. Autoregressive Distributed Lag (ARDL) Bound Test results have revealed the presence of a long-term comovement between variables and according to the Toda-Yamamoto test, there is a bi-directional causal relationship among the U.S. energy sector VIX and crude oil VIX. Chang, McAleer and Wang (2018) have aimed to investigate latent volatility Granger causality derived from the diagonal BEKK multivariate conditional volatility model and partial volatility spillovers for crude oil ETF and four renewable energy ETFs which are solar, wind, water and nuclear. For deriving diagonal BEKK model, they have utilized from the study by McAleer, Chan, Hoti and Lieberman (2008) who derive the multivariate extension from Tsay (1987)'s univariate random coefficient autoregressive process. Results have shown that significant positive latent volatility Granger causality relationships have been observed between solar, wind, nuclear and crude oil ETFs; in particular, the volatilities of return shocks of wind and water have potent effects on solar than do solar shocks on water and wind. Apart from these, all combinations of renewable resources and crude oil ETFs can be said to have negative partial impacts of covolatility spillovers from the shocks. Kanamura (2019) has analyzed the relationship between environmental value embedded in clean energy indices and energy value by focusing on the influence of energy risk on clean energy business through a suggestion of a supply and demand-based correlation (CR) model of clean energy indices & energy prices and a market risk model based on this. Variables handled in the research are the stock indices

and energy prices consisting of S&P Global Clean Energy Index (GCE), ECO, S&P/TSX Renewable Energy and Clean Technology Index (TXCT), S&P 500, WTI crude oil prices and Henry Hub (HH) natural gas prices. According to the findings, correlations between GCE or ECO and WTI crude oil or HH natural gas prices have been detected to be positive and an increasing function of the energy prices. However, although the correlations between S&P 500 and WTI or HH prices are positive; they represent a decreasing function of the corresponding energy prices. In addition, when the results associated with GCE and ECO are considered; TXCT is seen to develop as a clean energy index despite the fact that it functions partially.

1.3. Approaches to Nonparametric Density Estimation

In the literature of nonparametric estimators, there are two of them which are frequently encountered; that is to say, the histogram and frequency estimators. Although the two have a non-smoothness feature in common, there is a clear distinction in that the former is preferred for estimating the probability density function (PDF) of a continuous variable while the latter is preferred for estimating the probability function of discrete events (Racine, 2008, p. 5). Unless otherwise specified, let us assume that we try to estimate the unknown density $f(x)$ of a random variable X at a point x given n real observations $X_1, ..., X_n$ in a data set and specify the notation $\hat{f}$ in a way to represent the density estimator to be considered (Pagan & Ullah, 1999, p. 6).

1.3.1. Histogram

One of the most straightforward and oldest approaches to nonparametric density estimation is the histogram. Considering the case where X is discrete, a consistent estimator of $f(x)$ is expressed as $\hat{f}(x) = n^{-1}\sum_{i=1}^{n} I(X_i = x)$ with $I(X_i = x)$ denoting an indicator function which takes on the value 1 in case X_i is equal to x, 0 otherwise. If X is continuous, the probability that $X_i = x$ is zero and the estimation of $f(x)$ requires averaging X_is which take place in the interval around x, say $(x - h/2,\ x + h/2)$ where any x values are placed on the midpoint of the sampling interval. In this case, $\hat{f}(x)$ is stated as $\hat{f}(x) = (nh)^{-1}\sum_{i=1}^{n} I\left(x - \frac{h}{2} \le X_i \le x + \frac{h}{2}\right)$ or equivalently

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} I(-1/2 \leq \phi_i \leq 1/2) \tag{1.1}$$

where $\phi_i = (X_i - x)/h$, $I(A) = 1$ in case A holds true and 0 otherwise. Using the property $I(\phi) = I(-1/2 < \phi < 1/2)$ inferred from $\int_{-\infty}^{\infty} I(\phi)d\phi = \int_{-1/2}^{1/2} I(\phi)d\phi = \int_{-1/2}^{1/2} d\phi = 1$, one can confirm that the empirical density estimate is nonnegative and integrates to unity (Pagan & Ullah, 1999, pp. 7-8). If necessary to mention more specifically, the construction of an histogram is feasible through an origin x_0 and a bin width h where the latter controls the degree of smoothness of the estimate and the bins of the histogram are expressed as the intervals $[x_0 + mh,\ x_0 + (m+1)h)$ that are closed on the left and open on the right to be certain about definiteness property for positive and negative integers m. Then, the resulting estimate will be susceptible to the choice of both x_0 and h (Silverman, 1986, pp. 7-9). Despite its computationally easiness, discontinuous nature of the histogram results in hardships for cases where taking derivatives of the estimate is necessary (Racine, 2008, p. 7). Amongst the earliest studies, Van Ryzin (1973) has proposed a class of histogram-type density estimators based on spacings of order statistics and presented the proofs regarding pointwise consistency conditions of such estimators. An extended form of this paper to the bivariate estimator case is also discussed in Kim and Van Ryzin (1985). A differently modified version of the histogram introduced by Scott (1985) is the Average Shifted Histogram (ASH) which gives an average of m shifted histograms depending on the calculation of these histograms $\hat{f}_1, \hat{f}_2, ..., \hat{f}_m$ with equal sized bin width h and origins $t_0 = 0, \frac{h}{m}, \frac{2h}{m}, ..., \frac{(m-1)h}{m}$. When m tends to infinity, the ASH approximates a triangular kernel estimator. Apart from having a larger computational efficiency relative to a kernel estimator, an ASH provides the advantages of a better approximation & smoothness features over the histogram and overcomes the problem of sensitivity to the choice of origin (Deng & Wickham, 2011, p. 2; Klemelä, 2009, p. 365; Scott, 1992, p. 113). For a detailed description of ASH, see Chapter 5 in Scott (1992).

1.3.2. Kernel Density Estimation

Kernel density estimation (KDE) technique is a nonparametric way which is structured on estimating unknown PDFs from a given data set utilizing from a kernel function and can be considered as the generalization over the conventional histogram

method towards continuously differentiable density estimators. At the core of KDE, it lies to detect the effect of each point in the data set by weighting for all sample points. Good approximation properties of a kernel density estimator are captured in the case of a small number of regressors so that an increase in the number of regressors leads to a decrease in the convergence rate of the estimator and partial derivatives of the conditional mean estimator as well in a drastic manner (Gencay, Selcuk, & Whitcher, 2002, p. 273). A considerable amount of literature is available on KDE. The main idea of kernel density smoothing beyond the histogram-type estimators has been based upon smoothing periodograms and come into the picture with the preliminary studies by Einstein (1914), Daniell (1946) and Yaglom (1987). Besides, first suggestions regarding density estimates and nonparametric discriminant analysis have been put forward through a technical report by Fix and Hodges (1951) though the first paper on density estimation has been published by Rosenblatt (1956) who has first handled the asymptotical characteristics of the kernel density estimator with respect to MSE. Following the Rosenblatt (1956), one of the earliest attempts in explaining the mathematical theory of KDE has appeared thanks to Parzen (1962) who uses Fourier transform of the kernel estimators. For a more detailed theoretical investigation on KDE, see the books by Devroye and Györfi (1985), Silverman (1986), Devroye (1987), Härdle (1990), Scott (1992), Wand and Jones (1995), Devroye and Lugosi (2000).

1.3.2.1. Properties of Kernel Estimator

The univariate kernel density estimator has been formed in order to tackle with some constraints arising from the histogram -which has an inherent discrete nature stemmed from the demolished data structure as a result of relocating their individual positions with an interval where its cut-off points are located- in such a way that it provides the smoothness of PDF using the original locations of all data points in a sample. It could also be called the Rosenblatt-Parzen estimator (Rosenblatt (1956), Parzen (1962)) and expressed as follows:

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{X_i - x}{h} \right) \tag{1.2}$$

where $K(\cdot)$ denotes a symmetric weight function (i.e. kernel) and should have small values due to smaller weights designated for data points with large distances $|X_i - x|$, h is called the bandwidth (i.e. smoothing parameter or window width) and defined as a

function of sample size n. As n increases and h decreases, $\hat{f}(x)$ will converge to the true underlying density function $f(x)$ with probability one (Pagan & Ullah, 1999, p. 10; Racine, 2008, pp. 8-9). While a large value of h generates very smooth estimates, a small-valued h parameter generates twisted estimates (Deng & Wickham, 2011, p. 2). A comprehensive discussion of KDE could be reached in Silverman (1986) and Scott (1992). A general p th order kernel is required to satisfy the conditions $\int K(z)dz = 1$, $\int z^m K(z)dz = 0\,(m = 1,\ldots, p-1)$ and $\int z^p K(z)dz = \kappa_p \neq 0$ where p is greater than or equal to 2 as an integer. Although higher-order kernels offer strategies for bias reduction of $\hat{f}(x)$, their usage creates a dissatisfactory repercussion due to the possible negative values for density estimates (Racine, 2008, p. 9). In brief, obtaining a density estimate that is smoother than the histogram or the naive estimator is highly possible through KDE and the choice of the kernel plays a key role in this stage. To give an example, one can get more smooth estimators using the Gaussian kernel rather than the uniform kernel (Ghosh, 2018, pp. 19-20).

Statistical features of most kernel estimators are determined via the pointwise MSE criterion (or the pointwise risk). The MSE of $\hat{f}(x)$ is expressed by

$$\text{mse}\,\hat{f}(x) = E\{\hat{f}(x) - f(x)\}^2 = \text{var}\,\hat{f}(x) + \{\text{bias}\,\hat{f}(x)\}^2 \tag{1.3}$$

For sufficiently smooth densities; assuming that the kernel has finite fourth moment and utilizing from the Taylor series expansion, the bias can be given as

$$\text{bias}\{\hat{f}(x)\} = \frac{h^2}{2}\kappa_2 f''(x) + o(h^2) \tag{1.4}$$

while the variance is

$$\text{var}\{\hat{f}(x)\} = \frac{1}{nh} f(x)\int K^2(z)dz + o\left(\frac{1}{nh}\right) \tag{1.5}$$

(e.g. Wand & Jones, 1995, pp. 20-21) (Sheather, 2004, p. 589). We can say that curvature of the density function ($f''(x)$) leads to the bias of density estimate. Kernel density estimate has a leading bias term which is proportional to h^2 while the bias is of order h for a histogram estimator (Sheather, 2004, p. 589).

As more information about the sample becomes available ($n \to \infty$), h tends to zero (bias $\to 0$) and nh (i.e. efficient sample size) goes to infinity (var $\to 0$). Based on these consistency conditions for the given density (also MSE($\hat{f}(x)$) $\to 0$), the bias caused by

the curvature $f''(x)$ will shrink and vanish in the limit. These formulas state pointwise properties. However, the integrated mean square error (IMSE) represents a global error measure aggregating the MSE over the entire domain of the density. By minimizing IMSE with respect to the bandwidth and kernel function, the optimal bandwidths and optimal kernels are derived. With given variance and bias formulas in Equation (1.4) and Equation (1.5), IMSE is identified as follows:

$$\operatorname{imse}\hat{f}(x) = \int \operatorname{mse}\hat{f}(x)dx = \int \operatorname{var}\hat{f}(x)dx + \int \{\operatorname{bias}\hat{f}(x)\}^2 dx$$

$$\approx \int \left[\frac{f(x)}{nh}\int K^2(z)dz + \left\{ \frac{h^2}{2}\kappa_2 f''(x) \right\}^2 \right] dx = \frac{\Phi_0}{nh} + \frac{h^4}{4}\kappa_2{}^2\Phi_1 \qquad (1.6)$$

where $\Phi_0 = \int K^2(z)dz$ and $\Phi_1 = \int \{f''(x)\}^2 dx$ (Racine, 2008, pp. 10-11).

Hall, Racine and Li (2004) proposed an estimator for nonparametric estimation of the conditional density $g(y \mid x) = f(x, y) / f(x)$ which is defined as

$$\hat{g}(y \mid x) = \hat{f}(x, y) / \hat{f}(x) \qquad (1.7)$$

where $\hat{f}(x, y)$ and $\hat{f}(x)$ denote the kernel estimators of the joint and marginal densities $f(x, y)$ and $f(x)$ respectively (Racine, 2008, pp. 24-25). An alternative method for density estimation is the penalized likelihood method where the density is approximated as a mixture of m "basis functions" (or densities) with equal weights in a common manner. Unlike the KDE, in this method the bases/densities are not required to be centered at each data point; the basis functions differ only by a location parameter that is mostly called knots (Deng & Wickham, 2011, p. 3).

1.3.3. The Choice of Smoothing Parameter

1.3.3.1. Reference Rule of Thumb

In determining the bandwidth value using the reference rule of thumb, it has been utilized from a standard family of distributions for appointing a value to the unknown constant $\int f''(z)^2 dz$. To give an example; for the normal family, this unknown constant becomes equivalent to $\int f''(z)^2 dz = \frac{3}{8\sqrt{\pi}\sigma^5}$. In the case of Gaussian kernel, then

$$\int K^2(z)dz = \frac{1}{\sqrt{4\pi}}, \quad \int z^2 K(z)dz = 1 \qquad (1.8)$$

therefore the optimal bandwidth becomes

$$h_{opt} = (4\pi)^{-1/10}\left(\frac{3}{8}\right)^{-1/5}\pi^{1/10}\sigma n^{-1/5} = 1.059\sigma n^{-1/5} \tag{1.9}$$

Briefly, the rule-of-thumb bandwidth is approximately equivalent to $1.06\sigma n^{-1/5}$ and the sample standard deviation $\hat{\sigma}$ is used in practical sense (Racine, 2008, p. 13).

1.3.3.2. Plug-In Methodology

Sheather and Jones (1991) have proposed a data-based bandwidth selection method -also called "solve-the-equation" approach- for KDE which is an extension of Park and Marron (1990)'s 'plug-in' algorithm advancing the remarks discovered by Hall (1980) and Sheather (1983, 1986). Their bandwidth selection procedure exhibits a reliable good performance for smooth density shapes in simulations and depends on selecting the bandwidth in a way to minimize the approximate mean integrated square error (AMISE) in case $n \to \infty$ and $h = h(n) \to 0$:

$$AMISE(\hat{f}_h) = \frac{1}{nh}R(K) + \frac{h^4}{4}{\sigma_K}^4 R(f'') \tag{1.10}$$

(e.g. Silverman (1986), section 3.3) where $R(g) = \int g^2(x)dx$ and ${\sigma_g}^2 = \int x^2 g(x)dx$. $AMISE(\hat{f}_h)$ mentioned here is a general representation which is valid for all plug-in methods and as a common view, plug-in methods are considered to date back to Woodroofe (1970). The difference between these methods shows itself in the exact form of the estimate of unknown term $R(f'')$ in AMISE. Using the kernel-based estimate of $R(f'')$ represented by $\hat{S}(\alpha)$ where α denotes some suitable smoothing parameter, an estimate of the asymptotically optimal bandwidth value that minimizes AMISE is obtained through

$$\hat{h}_{AMISE} = [R(K)/\{{\sigma_K}^4 \hat{S}(\alpha)\}]^{1/5} n^{-1/5} \tag{1.11}$$

In estimating smooth densities, the performance of the estimator proposed by Park and Marron (1990) is superior in an obvious manner when compared to the other contemporary methods in the literature such as 'least squares cross-validation' and for a sufficiently smooth f, their plug-in method has a convergence rate of order $n^{-4/13}$. A useful side of the procedure proposed by Sheather and Jones (1991) comes through using a non-stochastic term for reducing bias in estimation process of $R(f'')$ without increasing variance (Sheather & Jones, 1991, pp. 683-685).

Sheather and Jones (1991)'s method adopts a bandwidth value α_1, namely $\alpha_1(h) = c_1 h^{5/7}$, to be used in the estimate $R(\hat{f}'')$ which is different from h and includes $R(f''')$ term in c_1; then estimates the resulting unknown functionals of f via kernel estimates with smoothing parameters based on a normality assumption on f. In this method, a negligibly small bias is likely through the estimation of $R(f''')$ by some $\hat{V}$ such that $\hat{V} = R(f''') + o_p(n^{-1/14})$ (Jones & Sheather, 1991) (Sheather & Jones, 1991, p. 686). Thus, the Sheather-Jones plug-in bandwidth is obtained as the solution to Equation (1.11) by replacing $\hat{S}(\alpha)$ with $R(\hat{f}''_{\alpha_1(h)})$. Apart from being a method advocated to a large extent (e.g. Bowman & Azzalini, 1997, p. 34; Simonoff, 1996, p. 77; Venables & Ripley, 2002, p. 129); when density estimation is not straightforward -for instance, depending on the presence of a large number of modes-, this method is inclined to result in over-smoothing and therefore, least-squares cross validation could be preferred (Sheather, 2009, p. 374).

1.3.3.3. Least Squares Cross Validation

Least squares (or unbiased) cross-validation method which is widely studied in the literature represents a fully automatic and data-driven approach in choosing the smoothing parameter. At the core of this method, it lies to choose the bandwidth which minimizes the IMSE of the resulting estimate. The integrated squared difference between $\hat{f}(x)$ and $f(x)$ is expressed as

$$\int \left\{ \hat{f}(x) - f(x) \right\}^2 dx = \int \hat{f}(x)^2 dx - 2\int \hat{f}(x) f(x) dx + \int f(x)^2 dx \qquad (1.12)$$

These values are replaced with sample counterparts and adjusted for bias. Then the bandwidth value can be computed based on the obtained objective function through minimization. For more detailed information, see Rudemo (1982) and Bowman (1984) who proposed this method (Racine, 2008, p. 14).

1.3.4. Nonparametric Quantile Regression

1.3.4.1. Local Polynomial Quantile Regression

Local polynomial kernel approaches (Stone, 1977; Cleveland, 1979) depend on the idea of approximating the unknown density function through a low-order polynomial imposing more weight on the values of response variable (generally, cubic) which

correspond to x_i observations closest to x and less weight to the points farther than x (Sheather, 2009, p. 375). Cleveland (1979) has proposed a robust locally weighted regression for scatterplot smoothing (generally speaking LOWESS smoothing method) based on the nearest neighbour bandwidths by using a tricube weight function shown as $w(x) = (1-|x|^3)^3$ for $|x|<1$ (and $w(x) = 0$ otherwise) and including a robustness step in which small (large) weights have been put on large (small) residuals. Cleveland and Devlin (1988) have broadened the scope of the methodology regarding locally weighted regression estimator using the tricube weight function and nearest neighbour bandwidths as in Cleveland (1979), but not handled the robustness in this study (thus, the estimator mentioned here is called LOESS) (Cleveland, 1979; Cleveland and Devlin, 1988; Sheather, 2009, p. 378).

Considering the conventional nonparametric model $Y_i = m(X_i) + \varepsilon_i$ for $i = 1,....,n$ under the assumption of independent and identically distributed (i.i.d.) errors having constant variance and being independent of X_i, we aim to estimate the τ -th conditional quantile of the response variable y given x defined by $m(x) = Q_Y(\tau \mid x)$ and the derivatives of this function. Local polynomial regression can be regarded as an extension of local weighted averaging by assigning a weight to each observation and weighted quantile regression is expressed as

$$\min_{\beta \in R^{p+1}} \sum_{i=1}^{n} w_i(x)\rho_\tau[y_i - \beta_0 - \beta_1(x_i - x_0) - \beta_2(x_i - x_0)^2 - \cdots - \beta_p(x_i - x_0)^p] \quad (1.13)$$

where $w_i(x) = K((X_i - x)/h)/h$, assuming K denotes a positive, symmetric & unimodal function and $\rho_\tau(v) = v\{\tau - I(v<0)\}$. Equation (1.13) allows for estimating higher order derivatives provided that the function m is smooth enough. The selection of smoothing parameter h is of vital importance in local polynomial fitting so that a very large value of h results in a local structure not understood obviously by over-smoothing while a very small value results in excessive variability based on too few observations. In a common manner, the bandwidth selection is founded upon the bias & variance trade-off (Hastie & Tibshirani, 1990) and the best value of it can be determined through an investigation on graphical representations for different h values or an automatic selection procedure depending on cross-validation or Mallow's C_p criterion (Mallows, 1973) (Davino, Furno, & Vistocco, 2014, p. 165). ε_i s have a strictly positive Hölder continuous density around

their τ -th quantile at 0. Let $\mathbf{V}$ denote a nonempty open subset of R^d including the point $0 \in R^d$, $u = (u_1, ..., u_d)$ denote a vector with d-dimension containing nonnegative integers and D^u denote the differential operator $\partial^{[u]} / (\partial x_1^{u_1} \cdots \partial x_d^{u_d})$, where $[u] = u_1 + \cdots + u_d$. For some real numbers $c > 0$, $\varphi \in (0,1]$ and some fixed nonnegative integer ℓ; consider $\Phi(c, \ell, \varphi)$ as the collection of functions m on $\mathbf{V}$ which are continuously differentiable up to order ℓ. The function m is smooth by assumption satisfying $| D^u m(x) - D^u m(0) | \leq c | x |^{\varphi}$ for all $x \in \mathbf{V}$ and $[u] = k$. The degree of smoothness of the functions m, represented by $p = \ell + \varphi$, has a crucial role in establishing the convergence rate of the estimator.

Up to now, a fixed value of smoothing parameter h has been considered. However, when x has a skewed distribution; it is likely to come across outlier values and some areas of the x-axis could contain very few points (maybe none) precluding the local-fitting of polynomials for specific x values. In such a case, it is required for adjusting the bandwidth parameter according to x values measured at not-equally spaced time points and nearest neighbour bandwidth is appropriate to make sure about the usage of the same number of points effectively for all x observations in estimating $m(x)$. How far x_i points are located from x is measured through the distance defined by $d_i(x) = | x - x_i |$ for $i = 1, 2, \ldots, n$ and the nearest neighbour bandwidth is considered as the *k*'th smallest $d_i(x)$ (Sheather, 2009, pp. 376, 378).

1.3.4.2. Quantile Smoothing Splines

A spline, as a commonly used smoothing method, represents a piecewise polynomial function (or curve) interpolating a sequence of control (join) points called knots (the locations where the lines are combined together) and being differentiable up to a specific order. Increasing flexibility in curve modelling is provided through an increment in the number of knots. If the aim is to obtain a smooth fit, cubic splines have the most widespread usage among splines. Intrinsic piecewise structure of a spline is an indicator of locality structure. Regression splines are constructed by combining a lot of separate smoothed polynomial functions in each interval between any two knots and frequently present more superior results relative to the polynomial regression which requires a high-degree polynomial to obtain flexibility in the fits with respect to revealing the flexible fits with an augmentation in the number of knots -but leaving the degree fixed- and usually

providing stability in estimates (James, Witten, Hastie, & Tibshirani, 2013, p. 276). Most literature studies have shown that determining the optimal number of control points is likely through taking areas where the density function seems to vary more quickly into consideration, utilizing from user-defined quantiles of the predictor variable or detecting if the polynomial functions have the continuity feature in their first and second derivatives at the knots - which is a desirable feature in order to be able to obtain a smooth curve.

Apart from Hendricks and Koenker (1992) who have suggested quantile smoothing splines, different representations of quantile smoothing splines have also been proposed by Koenker, Portnoy and Ng (1992) as solutions to the minimization problem given as:

$$\min \sum_{i=1}^{n} \rho_{\tau}[y_i - m(x_i)] + \lambda \left(\int_0^1 | m''(x) |^p \, dx \right)^{1/p} \tag{1.14}$$

where $p \geq 1$, $\rho_{\tau}(v) = v\{\tau - I(v < 0)\}$ gives the check function of Koenker and Bassett (1978) [$I(\bullet)$ shows the indicator function of $(-\infty, 0)$] and τ represents the quantile of interest satisfying $\tau \in [0,1]$. The parameter λ is regarded as the bandwidth parameter included in the kernel function or locally polynomial regression. A very small value of λ (imposing no constraint in the case of $\lambda = 0$) results in an estimated spline interpolating (passing through) the τ-th quantiles, while a sufficiently large value of λ ($\lambda = \infty$) generates a linear regression quantile fit to the observations (Koenker & Bassett, 1978) based on outweighing penalty relative to fidelity. First and second terms in Equation (1.14) are called *fidelity* and *roughness* (or penalty) respectively; while the former gives a measure regarding the goodness of fit of the estimated function, the latter exhibits the smoothness degree of the estimated curve (penalizing curvature). An effective computation for τ and λ can be implemented via parametric linear programming approaches. Cox and Jones in the discussion of the paper by Cole (1988) have suggested quantile smoothing splines as the minimization of the objective function:

$$\sum_{i=1}^{n} \rho_{\tau}[y_i - m(x_i)] + \lambda \int \{m''(x)\}^2 dx \tag{1.15}$$

where λ represents the smoothing parameter in cubic spline smoothing, therefore draws general inferences regarding the intense literature review on smoothing splines leaded by Wahba (1990). In terms of providing a computational easiness, it has been considered to replace $\{m''(x)\}^2$ in the penalty term by $|m''(x)|$ and the median case of this minimization problem has been handled by Schuette (1978) (Koenker, Ng, & Portnoy, 1994, pp. 673-674).

Koenker et al. (1994) have introduced an objective function including a total variation roughness penalty term on $m'(x)$ instead of the L_1 penalty on $m''(x)$:

$$\sum_{i=1}^{n} \rho_\tau [y_i - m(x_i)] + \lambda V(m') \tag{1.16}$$

where the function m which minimizes Equation (1.16) represents a linear spline having control points at x_i s for $i = 1,....,n$. The total variation of a function $g : [a,b] \to \mathbb{R}$ is formulated as

$$V(g) = \sup \sum_{i=1}^{n} \left| g(x_{i+1}) - g(x_i) \right| \tag{1.17}$$

For absolutely continuous g, $V(g)$ could be expressed as $V(g) = \int_a^b \left| g'(x) \right| dx$ and for functions g having an absolutely continuous first derivative g', the total variation roughness penalty term $V(m')$ becomes equivalent to

$$V(m') = \int_a^b \left| g''(x) \right| dx \tag{1.18}$$

(Davino et al, 2014, pp. 169-170; Koenker, 2005, pp. 230-231). For a more detailed description about the topic, see Koenker et al. (1994) and for penalty methods concerning bivariate smoothing, see Koenker (2005, pp. 235-248).

1.3.4.3. Penalized Spline Estimation in Quantile Regression

Penalized linear regression splines take place among the currently most famous approaches for scatterplot smoothing. A linear spline regression model presents a flexible way to accommodate many nonlinear trends which could not be approximated well by simple polynomial regression model (Fitzmaurice, Laird, & Ware, 2004, p. 147) and could be expressed as

$$y = \beta_0 + \beta_1 x + \sum_{i=1}^{K} b_i (x - c_i)_+ + e \tag{1.19}$$

where $(x - c_i)_+$ appears as a predictor in Equation (1.19) indicating the ith linear function with a break point (knot) at c_i ($i = 1,\ldots,K$):

$$(x - c_i)_+ = \begin{cases} x - c_i, & if \ x > c_i \\ 0, & if \ x \le c_i \end{cases} \tag{1.20}$$

and b_i represents the weight pertaining to the ith linear function. In the logic behind the linear spline, it underlies to split the time axis into a sequence of segments having distinct

slopes of linear functions in each segment (between two consecutive knots), but combined together at fixed times (Fitzmaurice et al., 2004, p. 147).

In the presence of large and complex data sets and when the weights b_i are considered as random effects, the usage of penalized splines has a remarkable superiority in the determination of the optimal number of knots such that it does not lead to underfitting or overfitting the data by reducing the required time for choosing the optimal number and location of knots and obtaining a considerably better fit to the data than the case without penalized estimation. For this reason, the weights b_i $(i = 1,\ldots,K)$ in Equation (1.19) are imposed a penalization restriction of $\sum_{i=1}^{K} b_i^2 < D$ for some constant D in order to reduce the aggregate effect of piecewise functions and penalize the splines for overfitting the data, thus to optimize the fit. The resulting estimator is stated as *penalized linear regression spline* and $\beta_0, \beta_1, b_1, b_2, \ldots, b_K$ are determined as the solution to minimizing the objective function

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2 + \lambda \sum_{i=1}^{K} b_i^2 \tag{1.21}$$

for some $\lambda \geq 0$ controlling the smoothness of the obtained fit. The second term in Equation (1.21) is called *roughness penalty* which penalizes too rough fits and therefore, the minimization of Equation (1.21) will reduce all the coefficients b_i towards zero (Griggs, 2013, pp. 26-31; Sheather, 2009, pp. 379-380).

Let $\omega_\tau(x_i)$ be an unknown true conditional quantile function of the response Y_i given $X_i = x_i$ satisfying

$$\omega_\tau(x) = \arg\min_{m} E[\rho_\tau(Y - m(x)) \mid X = x] \tag{1.22}$$

and a B-spline model can be handled for approximating $\omega_\tau(x)$ as represented by:

$$s_\tau(x) = \sum_{k=-p+1}^{K} B_k^{[p]}(x) b_k(\tau) \tag{1.23}$$

where $B_k^{[p]}(x)(k = -p+1,\ldots,K)$ denotes the p-th degree B-spline basis functions which can be specified as follows:

$$B_k^{[0]}(x) = \begin{cases} 1, & c_{k-1} < x \leq c_k \\ 0, & \text{otherwise} \end{cases}$$

$$B_k^{[p]}(x) = \frac{x - c_{k-1}}{c_{k+p-1} - c_{k-1}} B_k^{[p-1]}(x) + \frac{c_{k+p} - x}{c_{k+p} - c_k} B_{k+1}^{[p-1]}(x) \tag{1.24}$$

where c_k represents knots for $k = (-p+1, \ldots, K+p)$ and $b_k(\tau)$ represents unknown parameters (B-spline coefficients vector) for $k = (-p+1, \ldots, K)$. For a detailed description and features of B-splines, see de Boor (2001). The estimator of $b(\tau) = (b_{-p+1}(\tau) \ldots b_{K_n}(\tau))'$ can be specified as

$$\begin{aligned} \hat{b}(\tau) &= (\hat{b}_{-p+1}(\tau) \ldots \hat{b}_{K_n}(\tau))' \\ &= \arg\min_{b(\tau)} \left\{ \sum_{i=1}^{n} \rho_\tau (y_i - B(x_i)' b(\tau)) + \frac{\lambda_\tau}{2} b(\tau)' D_m' D_m b(\tau) \right\} \end{aligned} \tag{1.25}$$

where $B(x_j) = (B_{-p+1}(x_j) \ldots B_{K_n}(x_j))'$, λ_τ indicates (non-negative) smoothing parameter and $(K_n + p - m) \times (K_n + p)$ th matrix D_m indicates the m-th difference matrix expressed by $D_m = (d_{ij}^{(m)})_{ij}$ where

$$d_{ij}^{(m)} = \begin{cases} (-1)^{|i-j|}{}_m C_{|i-j|}, & i \le j \le m+1 \\ 0, & \text{otherwise} \end{cases} \tag{1.26}$$

The difference penalty term $b(\tau)' D_m' D_m b(\tau)$ incorporated into Equation (1.25) is very feasible to be carried out in the mean regression and controls the roughness of the curvature. The solution path for $\hat{b}(\tau)$ can be found by implementing linear programming methods - for instance, simplex or interior points methods. Besides this, the iteratively reweighted least squares (IRLS) can be regarded as a feasible method for nonparametric quantile regression. See also Reiss and Huang (2012) for a detailed investigation on the penalized spline estimator obtained through IRLS method. Taking IRLS into account, the penalized spline estimator of $\omega_\tau(x)$ using $\hat{b}(\tau)$ can be expressed as:

$$\hat{\omega}_\tau(x) = \sum_{k=-p+1}^{K} B_k^{[p]}(x) \hat{b}_k(\tau) = B(x)^T \hat{b}(\tau) \tag{1.27}$$

(Yoshida, 2013, pp. 2-4).

1.3.4.4. Nonparametric Quantile Regression Series Approximation

In quantile regression (QR) framework, the effect of covariates on outcome variables is qualified by the conditional quantile function and its functionals, especially in the case of heterogenous effects. Belloni, Chernozhukov, Chetverikov and Fernández-

Val (2011) have approximated the entire conditional quantile function of Y given $X = x$ for τ-th quantile denoted as $Q(\tau, x)$ or $Q_{Y|X}(\tau \mid x)$ by a linear combination of series terms $Z(x)'\beta(\tau)$ where the vector $Z(x)$ contains transformations of x exhibiting solid approximation features like powers, trigonometric terms, splines or local polynomials and τ is the element of a quantile set ϑ ($\vartheta \subset (0,1)$) satisfying $\tau \in [0,1]$. Belloni et al. (2011) have also developed the nonparametric QR-series framework presenting inference on the entire conditional quantile function $(\tau, x) \mapsto Q(\tau, x)$ and its function-valued linear functionals which are the partial derivative function $(\tau, x) \mapsto \partial_{xr} Q(\tau, x)$, the average partial derivative function $\tau \mapsto \int \partial_{xr} Q(\tau, x) d\eta(x)$ and the conditional average partial derivative function $(\tau, x_r) \mapsto \int \partial_{xr} Q(\tau, x) d\eta(x \mid x_r)$ where η denotes a specific measure and x_r denotes the rth element of x. Based on the two uniform strong approximations (which are conditionally pivotal and Gaussian processes) to the QR-series coefficient process of an increasing dimension, Belloni et al. (2011) have developed four resampling methods (pivotal, gradient bootstrap, Gaussian and weighted bootstrap) where the first two techniques enable to approximate the distribution of the pivotal process while the last two techniques are performed for Gaussian process. These proposed methods make a contribution to deriving an integrated large sample inference theory for the linear functionals of $x \mapsto Q(\tau, x)$, $\tau \in \vartheta$. Quantile specific coefficients are incorporated in the function $\tau \mapsto \beta(\tau)$ and estimated through the QR estimator proposed by Koenker & Bassett (1978). An increase in the number of series terms will result in diminishing QR series approximation error given as

$$E(\tau, x) = Q(\tau, x) - Z(x)'\beta(\tau) \tag{1.28}$$

which will tend to approach zero in the limit or in other words, $\sup_{x \in \chi, \tau \in \vartheta} | E(\tau, x) | \to 0$[1] as $n \to \infty$. This reveals that the true conditional quantile function $Q(\tau, x)$ will be well-approximated by the approximating function $Z(x)'\beta(\tau)$ (Belloni et al., 2011, p. 10). Note that $\hat{\beta}(\tau)$ represents the QR estimator of $\beta(\tau)$ where inference on the latter is hard to

[1] The supremum (i.e. sup) of a nonempty set of real numbers $B \subset \mathbb{R}$ is the least upper bound such that a real number $k \in \mathbb{R}$ exists - which is called the supremum of B denoted by $k = \sup B$ - in case every $x \in B$ satisfies $x \leq k$; but for each $b < k$, an x is available in the set B providing $x > b$ (Bagby, 2001, p. 12).

manage depending on the information that generally speaking, a limit distribution does not exist for the process $\tau \mapsto \sqrt{n}(\hat{\beta}(\tau) - \beta(\tau))$. Inferences having superior quality have been built up under the coupling framework where a coupling is described as a solid technique in probability theory which is the joint construction of two random components on the same probability space being uniformly close to one another with high probability. In their study, Belloni et al. (2011) have developed pivotal and Gaussian couplings implying that these two processes display a strong approximation to $\sqrt{n}(\hat{\beta}(\tau) - \beta(\tau))$. Pivotal coupling has been established upon the gradient bootstrap method which is chiefly presented by Parzen, Wei & Ying (1994) for the parametric case and a valid inference on the basis of the pivotal & gradient bootstrap methods is likely through only a moderate constraint over the increase of the number of series terms as associated with the sample size for a feasible estimation. On the other hand, a strong Gaussian approximation has been constructed utilizing from Yurinskii's coupling (see Yurinskii (1977)) which was also applied by Chernozhukov, Lee and Rosen (2013). In their study, Belloni et al. (2011) have also pointed out that either spline or polynomial-based QR series estimators ($x \mapsto Z(x)'\hat{\beta}(\tau)$) achieve the fastest possible convergence rate in the L_2 norm while spline-based estimators achieve this rate in the sup norm in Hölder smoothness classes (Belloni et al., 2011, pp. 1-4). The core of the proof techniques mentioned in the study by Belloni et al. (2011) consists of maximal inequalities not previously known for function classes and uniform entropy. Series estimators mentioned in the study display some attractive characteristics relative to kernel estimators. Primarily, such estimators enable us to estimate the entire conditional quantile function and its linear functionals using a rather small parameter set so that the values of the estimates could be computed through quantile-specific coefficient estimates $\hat{\beta}(\tau)$. Besides, one can impose shape restrictions with ease utilizing from series estimation and in the case of high dimensional vector including many covariates apart from the main one, series methods are practical in order to restrain the curse of dimensionality by assigning a partially linear form on the functions $x \mapsto Q(\tau, x)$ (Belloni et al., 2011, p. 6).

For a more detailed series approximation, the series of models indexed by the sample size n can be expressed as

$$Y_{i,n} = Q_n(T_{i,n}, X_{i,n}), \tag{1.29}$$

where $Y_{i,n}$ represents an outcome variable, $X_{i,n}$ represents a d_n dimensional vector of covariates, $T_{i,n}$ represents unobserved individual features with $T_{i,n} \mid X_{i,n} \sim$ Uniform $(0,1)$ that is independent of $X_{i,n}$. $Q_n : [0,1] \times \chi_n \to \mathbb{R}$ represents an unknown function so that for all $x \in \chi_n$, the function $Q_n(\tau, x)$ is strictly increasing in τ. $Q_n(\tau, x)$ is also called the "conditional τ-quantile function" and $(X_{i,n}, T_{i,n}, Y_{i,n})_{i=1}^{n}$ is supposed to be random sample from the distribution (X^n, T^n, Y^n). When the covariate vector X is stated as $X = (U, W)$ where U denotes the main covariate while W denotes the vector of covariates excluding U, Model (1.29) in the partially linear quantile model form can also be written differently as

$$Q(\tau, x) = f(\tau, u) + w'\gamma(\tau), \quad \tau \in [0,1] \tag{1.30}$$

Two descriptions are contained in Model (1.29) as "*nonparametric* (NP) *model*" and "*many regressors* (MR) *model*". According to the NP model which presents flexibility by not requiring any parametric assumptions on $Q(\tau, x)$, the functions $x \mapsto Q(\tau, x)$ are assumed to be smooth; X defines a d-dimensional vector of observable covariates with $\chi \subset \mathbb{R}^d$ and T meets the condition $T \mid X \sim$ Uniform $(0,1)$. In addition, DGP is not influenced by n such that for any sample size; Y^n, X^n, T^n, Q_n are treated equally with Y, X, T and Q respectively and QR series approximation error given in Equation (1.28) will disappear in an asymptotical manner under some suitable conditions provided that $m \to \infty$ as $n \to \infty$ (However, for series functions like B-splines and compactly supported wavelets; the entire collection of these terms is based on the sample size n). On the other hand, in MR model, the dimension d_n increases with sample size and the second argument of $Q_n(\tau, x)$ shows linearity: $Q_n(\tau, x) = w'\gamma_n(\tau)$. Combining both models together, nonparametric QR series approximation developed by Belloni et al. (2011) is expressed as follows:

$$Q(\tau, x) \approx Z(x)'\beta(\tau), \quad \beta(\tau) = \left(\alpha(\tau)', \gamma(\tau)'\right)', \quad Z(x) = \left(Z(u)', w'\right)' \tag{1.31}$$

Relying on the information in Equation (1.31), approximating the unidentified argument $f(\tau, u)$ in Equation (1.30) is realized through $Z(u)'\alpha(\tau)$ where the vector $Z(u)$ consists of transformations of u with good approximation features for the NP model. However, $Z(x) = x$ will be assumed for all $x \in \chi$ for the MR part (represented by w) of Model

(1.30) and approximation error $E(\tau, x)$ will be zero in the MR model for all $x \in \chi$. The process of QR-series coefficients $\beta(\tau) = (\beta_1(\tau), ..., \beta_m(\tau))'$ is estimated through the QR estimator proposed by Koenker and Bassett (1978) as follows:

$$\hat{\beta}(\tau) \in \arg\min_{\beta \in \mathbb{R}^m} \sum_{i=1}^{n} \rho_\tau \left(Y_i - Z(X_i)'\beta \right), \quad \tau \in \vartheta \subseteq (0,1) \tag{1.32}$$

where ρ_τ and m represent the check function and the dimension of $\beta(\tau)$ respectively (Belloni et al., 2011, pp. 8-10; Lipsitz, Belloni, Chernozhukov, & Fernández-Val, 2016, p. 370). Considering the QR-series approximation, the *QR-series estimator* can be identified as

$$\hat{Q}(\tau, x) = Z(x)'\hat{\beta}(\tau), \tag{1.33}$$

As the sample size n expands, both the approximation error $E(\tau, x)$ and the estimation error $\hat{Q}(\tau, x) - Z(x)'\beta(\tau)$ will asymptotically vanish. There exist some principal regularity requirements to be considered for the QR-series framework:

Condition S.

S.1. "*The random variables are described on a triangular array so that for any* n, $D_n = \{(X_i, Y_i) : 1 \le i \le n\}$ *represents an i.i.d. random sample from the distribution of the pair* (X, Y).

S.2. *(i) The conditional density* $f_{Y|X}(y \mid x)$ *has a uniform boundedness from above over* $y \in \Upsilon_x$ *where* Υ_x *is the support of the conditional distribution of Y given* $X = x$, $x \in \chi$ and n; *(ii)* $f_{Y|X}(Q(\tau, x) \mid x)$ *has uniform boundedness away from zero over* $\tau \in \bar{\vartheta}$ *where* $\bar{\vartheta}$ *shows the convex hull of* ϑ, $x \in \chi$ and n; *and (iii) the derivative of* $y \mapsto f_{Y|X}(y \mid x)$ *is continuous and uniformly bounded from above in absolute value over* $y \in \Upsilon_x$, $x \in \chi$ and n.

S.3. *The eigenvalues of the Gram Matrix* $\sum = E[Z(X)Z(X)']$ *are bounded uniformly from above and away from zero over* n.

S.4. *The approximation error* $E(\tau, x)$ *satisfies* $\sup_{x\in\chi,\tau\in\vartheta} |E(\tau,x)| \lesssim m^{-\kappa}$ *where* $\kappa \in (0,\infty]$ *is a constant which is independent of* n[2]." (Belloni et al., 2011, p. 10).

Condition S.2. which enables us to keep the track of whether or not the first [for (i) and (ii)] and second [for (iii)] partial derivatives of $Q(\tau, x)$ with respect to τ also display uniformly bounded features as mentioned over $\tau \in [0,1]$, $x \in \chi$ and n for (i) and (iii); $\tau \in \bar{\vartheta}$, $x \in \chi$ and n for (ii). On the other hand, Condition S.3. suggests a logical inference for the MR model regarding that the problem of perfect multicollinearity does not exist among covariates and because the approximation error $E(\tau, x)$ is equal to zero for any $\tau \in \vartheta$ and $x \in \chi$, Condition S.4. holds for $\kappa = \infty$ (Belloni et al., 2011, p. 11). Which approximating functions will be used in $Z(x)$ apart from the dimension of it is of great importance while determining the characteristics of the QR-series estimator and coefficient process. At the core of couplings and resampling methods, Jacobian matrix described as $J(\tau) = E\left[f_{Y|X}(Q(\tau, X) \mid X) Z(X) Z(X)' \right]$ takes on a significant role. For carrying out inference methods and providing algorithms for resampling methods, Belloni et al. (2011) have utilized from the quantity of $\xi_m = \sup_{x\in\chi} \|Z(x)\|$ and the estimator of Jacobian matrix proposed by Powell (1984) which is described as

$$\hat{J}(\tau) = \frac{1}{2h} \mathbb{E}_n \left[1\{ |Y_i - Z_i' \hat{\beta}(\tau)| \le h \} \cdot Z_i Z_i' \right] \tag{1.34}$$

Here $\mathbb{E}_n$ represents the expectation with respect to the empirical measure and h denotes the bandwidth which satisfies $h = h_n \to 0$. With the assumption of $\chi = [0,1]^d$; in case tensor products of polynomials and tensor products of B-splines are incorporated into the vector Z, $\xi_m \lesssim m$ and $\xi_m \lesssim m^{1/2}$ hold respectively (Belloni et al., 2011, p. 14).

1.3.4.4.1. Asymptotic Theory for QR-Series Coefficient Process

In the study by Belloni et al. (2011), mainly two couplings have been constructed in a way to generate two processes whose distributions could be simulated through four resampling methods mentioned before and which are uniformly close to normalized coefficient process shown as in Equation (1.35):

---

[2] What the notation $a^n \lesssim b^n$ implies is that the inequality $a^n \le K b^n$ is satisfied for any given n with a constant K that does not depend upon n (Belloni et al., 2011, p. 7).

$$\sqrt{n}(\hat{\beta}(\cdot)-\beta(\cdot))=\left\{\sqrt{n}(\hat{\beta}(\tau)-\beta(\tau)):\tau\in\vartheta\right\} \tag{1.35}$$

Based on the i.i.d. sample $(X_i,Y_i)_{i=1}^n$, the function $\beta(\tau)$ is estimated through the QR-series coefficient process $\hat{\beta}(\tau)$ as mentioned above. Relying on the assumptions $m\xi_m{}^2\log^2 n=o(n)$ and $m^{-\kappa}\log n=o(1)$ apart from the Condition S., one of the principal conclusions is that $\hat{\beta}(\tau)$ has a convergence rate which holds uniformly over $\tau\in\vartheta$ as follows:

$$\sup_{\tau\in\vartheta}\left\|\hat{\beta}(\tau)-\beta(\tau)\right\|\lesssim_P\sqrt{m/n} \tag{1.36}$$

This theorem also implies that the QR-series estimator achieves a convergence rate uniformly over τ in the L^2 norm for the NP model.

1.3.4.4.1.1. Uniform Strong Approximations (Couplings) and Resampling Methods

Pivotal Coupling: The process presented as

$$\mathbb{T}(\tau)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}Z_i(\tau-\mathbf{1}\{T_i\le\tau\}),\ \ \tau\in\vartheta \tag{1.37}$$

is called (conditionally) pivotal; because as being conditional on $(Z_i)_{i=1}^n$, $(T_i)_{i=1}^n$ includes i.i.d. Uniform$(0,1)$ random variables.

Approximating the process $\sqrt{n}(\hat{\beta}(\cdot)-\beta(\cdot))$ in a strong manner is likely through the pivotal process expressed as $J^{-1}(\cdot)\mathbb{T}(\cdot)=\{J^{-1}(\tau)\mathbb{T}(\tau):\tau\in\vartheta\}$. Relying on the assumptions $m^3\xi_m{}^2=o(n^{1-\varepsilon})$ and $m^{-\kappa+1}=o(n^{-\varepsilon})$ for a constant value $\varepsilon>0$ apart from the Condition S., the approximation to the original process could be expressed as:

$$\sqrt{n}(\hat{\beta}(\tau)-\beta(\tau))=J^{-1}(\tau)\mathbb{T}(\tau)+p(\tau),\ \ \tau\in\vartheta \tag{1.38}$$

where

$$\sup_{\tau\in\vartheta}\left\|p(\tau)\right\|\lesssim_P\frac{m^{3/4}\xi_m{}^{1/2}\log^{1/2}n}{n^{1/4}}+\sqrt{m^{1-\kappa}\log n}=o(n^{-\varepsilon'}) \tag{1.39}$$

for some $\varepsilon'>0$. This theorem also introduces some practical outcomes which one of these is uniformity in τ convergence rate of QR-series estimator in the sup norm for the NP model.

Resampling Methods Based on Pivotal Coupling: There are two resampling methods for a strong approximation based on the pivotal process, namely the pivotal method and the gradient bootstrap method. Approximating the distribution of the process

$\sqrt{n}(\hat{\beta}(\cdot)-\beta(\cdot))$ through the pivotal method based on a pre-determined number of bootstrap repetitions -denoted by B - is possible by accounting for such an algorithm: Before all else, an i.i.d. sequence $\{T_i^b\}_{i=1}^n$ for $b=1,....,B$ is simulated from $T \sim$ Uniform $(0,1)$ and the process

$$\mathbb{T}^*(\tau)=n^{-1/2}\sum_{i=1}^{n} Z_i(\tau-\mathbf{1}\{T_i^b \le \tau\}),\ \ \tau\in\mathcal{G} \tag{1.40}$$

is determined by calculation. Subsequent to this, the estimators $\hat{J}(\tau)$ are computed as in Equation (1.34) and the distributions of both $J^{-1}(\tau)\,\mathbb{T}^*(\tau)$ and $\sqrt{n}(\hat{\beta}(\tau)-\beta(\tau))$ for $\tau\in\mathcal{G}$ are approximated by the conditional distribution $\{\hat{J}^{-1}(\tau)\,\mathbb{T}^*(\tau):\tau\in\mathcal{G},\ 1\le b\le B\}$. The theorem regarding *the pivotal method* can be defined in that way: Relying on the assumptions $h\sqrt{m}=o(n^{-\varepsilon})$, $m^2{\xi_m}^2=o(n^{1-\varepsilon}h)$ and $m^{-\kappa+1/2}=o(n^{-\varepsilon})$ for some constant $\varepsilon>0$ apart from the Condition S.; define

$$\hat{J}^{-1}(\tau)\,\mathbb{T}^*(\tau)=J^{-1}(\tau)\,\mathbb{T}^*(\tau)+p(\tau),\ \ \tau\in\mathcal{G} \tag{1.41}$$

where

$$\sup_{\tau\in\mathcal{G}}\|p(\tau)\| \lesssim_P \sqrt{\frac{{\xi_m}^2 m^2\log n}{nh}}+m^{-\kappa+1/2}+h\sqrt{m}=o(n^{-\varepsilon'}) \tag{1.42}$$

for some $\varepsilon'>0$. This pivotal method is tightly pertinent to the gradient bootstrap approach which was presented by Parzen et al. (1994) for parametric models having a constant dimension and Belloni et al. (2011) have broadened the scope of this approach in a way to be feasible for a general series framework. In order to be able to carry out *gradient bootstrap method*, firstly an i.i.d. sequence $\{T_i^b\}_{i=1}^n$ for $b=1,....,B$ is simulated from $T \sim$ Uniform $(0,1)$ and the process $\mathbb{T}^*(\tau)$ is computed as in Equation (1.40). Then, for each bootstrap repetition $b=1,....,B$ and all τ 's; the method necessitates to obtain the gradient bootstrap estimator $\hat{\beta}^*(\tau)$ as a solution to the quantile optimization problem given as:

$$\min_{\beta\in\mathbb{R}^m}(\mathbb{E}_n[\rho_\tau(Y_i-Z(X_i)'\beta)]-\mathbb{T}^*(\tau)'\beta/\sqrt{n}) \tag{1.43}$$

and the distribution of $\sqrt{n}(\hat{\beta}(\cdot)-\beta(\cdot))$ is approximated utilizing from the process $\sqrt{n}(\hat{\beta}^*(\cdot)-\hat{\beta}(\cdot))=\left\{\sqrt{n}(\hat{\beta}^*(\tau)-\hat{\beta}(\tau)):\tau\in\mathcal{G}\right\}$ that could be simulated. Solving the quantile optimization problem for each simulation is not a valid case for the pivotal method which provides an advantage over the gradient bootstrap method with regard to

computational flexibility and ease. The theorem regarding the gradient bootstrap method can be defined like that: Based on the assumptions, $m^3 \xi_m^{\ 2} = o(n^{1-\varepsilon})$ and $m^{-\kappa+1/2} = o(n^{-\varepsilon})$ for some constant $\varepsilon > 0$ apart from the Condition S.; define

$$\sqrt{n}\left(\hat{\beta}^*(\tau) - \hat{\beta}(\tau)\right) = J^{-1}(\tau)\,\mathbb{T}^*(\tau) + p(\tau), \tag{1.44}$$

where

$$\sup_{\tau\in\vartheta}\left\|p(\tau)\right\| \lesssim_P \frac{m^{3/4}\xi_m^{\ 1/2}\log^{1/2} n}{n^{1/4}} + m^{-\kappa+1/2} = o(n^{-\varepsilon'}) \tag{1.45}$$

for some $\varepsilon' > 0$. As understood from the given theorem, the principal superiority of the gradient bootstrap method when compared to the pivotal method is that the Jacobian matrices $J(\tau)$ do not have to be estimated since such an estimation considers the choice of the smoothing parameter h obligatory in a subjective manner as perceived from Equation (1.34) (Belloni et al., 2011, pp. 14-20; 42).

Gaussian Coupling: Assuming that Condition S. holds and under the assumptions $m^7 \xi_m^{\ 6} = o(n^{1-\varepsilon})$ and $m^{-\kappa+1} = o(n^{-\varepsilon})$ for some constant $\varepsilon > 0$, the strong approximation to the process $\sqrt{n}(\hat{\beta}(\cdot) - \beta(\cdot))$ using the Gaussian process can be expressed as

$$\sqrt{n}(\hat{\beta}(\tau) - \beta(\tau)) = J^{-1}(\tau)G(\tau) + p(\tau), \;\; \tau \in \vartheta \tag{1.46}$$

where $G(\cdot) = G_n(\cdot)$ represents a zero-mean Gaussian process on ϑ as being conditional on $(Z_i)_{i=1}^n$ and

$$\sup_{\tau\in\vartheta}\left\|p(\tau)\right\| = o_p(n^{-\varepsilon'}) \tag{1.47}$$

for some $\varepsilon' > 0$.

Resampling Methods Based on Gaussian Coupling: There are two resampling methods for a strong approximation based on the Gaussian process, namely the Gaussian method and the weighted bootstrap method. Approximating the distribution of the process $\sqrt{n}(\hat{\beta}(\cdot) - \beta(\cdot))$ using the Gaussian method is feasible through such an algorithm: Primarily, a sequence including m independent scalar Brownian bridges that are independent of the data are simulated and denoted by $B_m^b(\cdot) = \{B_m^b(\tau) : \tau \in \vartheta\}$ for $b = 1, ..., B$. As being conditional on $(Z_i)_{i=1}^n$; identify the process

$$G^*(\tau) = G_n^{\ *}(\tau) = \hat{\Sigma}^{1/2} B_m^b(\tau), \;\; \tau \in \vartheta \tag{1.48}$$

as an imitation of $G(\cdot)$ where the estimator of Σ is $\hat{\Sigma} = \mathbb{E}_n [Z_i Z_i']$ (Belloni et al., 2011, p. 14). The theorem regarding *the Gaussian method* can be expressed like that: Considering that Condition S. holds and based on the assumptions $h\sqrt{m} = o(n^{-\varepsilon})$, $m^2 \xi_m^{\ 2} = o(n^{1-\varepsilon} h)$ and $m^{-\kappa+1/2} = o(n^{-\varepsilon})$ for some constant $\varepsilon > 0$; approximating the distributions of both the processes $J^{-1}(\cdot)G^*(\cdot) = \{J^{-1}(\tau)G^*(\tau) : \tau \in \vartheta\}$ as a copy of $J^{-1}(\cdot)G(\cdot)$ and $\sqrt{n}(\hat{\beta}(\cdot) - \beta(\cdot))$ using the Gaussian process is likely through

$$\hat{J}^{-1}(\tau)G^*(\tau) = J^{-1}(\tau)G^*(\tau) + p(\tau), \quad \tau \in \vartheta \tag{1.49}$$

where

$$\sup_{\tau \in \vartheta} \| p(\tau) \| \lesssim_P \sqrt{\frac{m^2 \xi_m^{\ 2} \log n}{nh}} + m^{-\kappa+1/2} + h\sqrt{m} = o(n^{-\varepsilon'}) \tag{1.50}$$

for some $\varepsilon' > 0$ (Belloni et al., 2011, pp. 20-21; 42). For more detailed explanations about the inference processes, see Belloni et al. (2011).

1.3.5. Nonparametric Quantile Causality Approach

In this study, a novel methodology describing nonlinear causality in an efficient manner has been put forward by Balcilar, Gupta & Pierdzioch (2016) and Balcilar, Bekiros & Gupta (2017) through a hybrid approach which combines the frameworks of Nishiyama, Hitomi, Kawasaki & Jeong (2011) and Jeong, Härdle & Song (2012). For the application part of this first essay in the dissertation; clean energy returns have been denoted as y_t while technology, oil and S&P 500 returns have been denoted as x_t. Following Jeong et al. (2012), the quantile-type causality is specified as follows: x_t does not cause y_t in the θ-quantile given the lags of x and y $\left\{ y_{t-1}, ..., y_{t-p}, x_{t-1}, ..., x_{t-p} \right\}$ in the case of

$$Q_\theta(y_t \mid y_{t-1}, ..., y_{t-p}, x_{t-1}, ..., x_{t-p}) = Q_\theta(y_t \mid y_{t-1}, ..., y_{t-p}) \tag{1.51}$$

and x_t causes y_t in the θ-quantile given the lags of x and y $\left\{ y_{t-1}, ..., y_{t-p}, x_{t-1}, ..., x_{t-p} \right\}$ if

$$Q_\theta(y_t \mid y_{t-1}, ..., y_{t-p}, x_{t-1}, ..., x_{t-p}) \neq Q_\theta(y_t \mid y_{t-1}, ..., y_{t-p}) \tag{1.52}$$

where $Q_\theta \{ y_t \mid \bullet \}$ represents the θ-th (conditional) quantile of y_t which depends on t with the restriction identified as $0 < \theta < 1$.

In order to provide a concise focusing on the causality-in-quantiles tests; Y_{t-1}, X_{t-1} and Z_t vectors are defined as $Y_{t-1} \equiv (y_{t-1},...,y_{t-p})$, $X_{t-1} \equiv (x_{t-1},...,x_{t-p})$, $Z_t \equiv (X_t, Y_t)$ while $F_{y_t|Z_{t-1}}(y_t \mid Z_{t-1})$ and $F_{y_t|Y_{t-1}}(y_t \mid Y_{t-1})$ denote the distribution functions of y_t conditioned on Z_{t-1} and Y_{t-1} respectively. Furthermore, it is assumed that the conditional distribution $F_{y_t|Z_{t-1}}(y_t, Z_{t-1})$ is entirely continuous in y_t for nearly all Z_{t-1}. The condition $F_{y_t|Z_{t-1}}\{Q_\theta(Z_{t-1}) \mid Z_{t-1}\} = \theta$ is satisfied with the probability value 1 where $Q_\theta(Z_{t-1})$ and $Q_\theta(Y_{t-1})$ are equivalent to $Q_\theta(y_t \mid Z_{t-1})$ and $Q_\theta(y_t \mid Y_{t-1})$ respectively. Thus, on the basis of Equations (1.51) and (1.52), the hypotheses to be tested for the quantile-based causality-in-mean (i.e. causality in the conditional 1st moment) can be specified as

$$H_0 = P\left\{F_{y_t|Z_{t-1}}\{Q_\theta(Y_{t-1}) \mid Z_{t-1}\} = \theta\right\} = 1 \tag{1.53}$$

$$H_1 = P\left\{F_{y_t|Z_{t-1}}\{Q_\theta(Y_{t-1}) \mid Z_{t-1}\} = \theta\right\} < 1 \tag{1.54}$$

For the causality-in-quantiles tests to be carried out in a feasible manner through a measurable metric, $J = \{\varepsilon_t E(\varepsilon_t \mid Z_{t-1}) f_Z(Z_{t-1})\}$ distance measure is performed in the study by Jeong et al. (2012) where ε_t indicates the regression error term that appears depending on the null hypothesis expressed in Equation (1.53) and $f_Z(Z_{t-1})$ indicates the marginal density function of Z_{t-1}. Thus, the regression error ε_t establishes the core conceptual framework of the causality-in-quantile tests. The null hypothesis in Equation (1.53) can be regarded as true only if $E\left[\mathbf{1}\{y_t \le Q_\theta(Y_{t-1}) \mid Z_{t-1}\}\right] = \theta$ or correspondingly $\mathbf{1}\{y_t \le Q_\theta(Y_{t-1})\} = \theta + \varepsilon_t$ which exhibits the regression error term as clearly apparent. Here $\mathbf{1}\{\bullet\}$ denotes a signal function. Utilizing from ε_t, Jeong et al. (2012) restates the distance metric as follows:

$$J = E\left[\left\{F_{y_t|Z_{t-1}}\{Q_\theta(Y_{t-1}) \mid Z_{t-1}\} - \theta\right\}^2 f_Z(Z_{t-1})\right] \tag{1.55}$$

In Equation (1.55), it is of high importance to know that J is undoubtedly equal to or greater than zero when Equations (1.53) and (1.54) are considered. The equality $J = 0$ is ensured under the null H_0 in Equation (1.53) while $J > 0$ is satisfied only in case the alternative H_1 holds true in Equation (1.54). The practicable causality-in-quantile test

statistics for J in Equation (1.55) takes a somewhat different form in Jeong et al. (2012) for a constant quantile level θ by considering weighting functions given by:

$$\hat{J}_T = \frac{1}{T(T-1)h^{2p}} \sum_{t=p+1}^{T} \sum_{s=p+1, s\neq t}^{T} K\left(\frac{Z_{t-1} - Z_{s-1}}{h}\right) \hat{\varepsilon}_t \hat{\varepsilon}_s \tag{1.56}$$

where $K(\cdot)$ is a known kernel function, h is the bandwidth parameter for the KDE, T is the sample size and p is the lag-order chosen for specifying the vector Z_t while $\hat{\varepsilon}_t$ gives the estimate of the unknown regression error which represents the most vital component of the test statistics $\hat{J}_T$ and can be obtained in conformity with quantile-based causality as follows:

$$\hat{\varepsilon}_t = \mathbf{1}\left\{ y_t \leq \hat{Q}_\theta(Y_{t-1}) \right\} - \theta \tag{1.57}$$

Here the quantile estimator $\hat{Q}_\theta(Y_{t-1})$ is an estimate of the θ-th conditional quantile of y_t given Y_{t-1} and by performing the nonparametric kernel method, it is evaluated as

$$\hat{Q}_\theta(Y_{t-1}) = \hat{F}^{-1}_{y_t|Y_{t-1}}(\theta \,|\, Y_{t-1}) \tag{1.58}$$

where $\hat{F}_{y_t|Y_{t-1}}(y_t \,|\, Y_{t-1})$ represents the *Nadaraya Watson* kernel estimator expressed by

$$\hat{F}_{y_t|Y_{t-1}}(y_t \,|\, Y_{t-1}) = \frac{\sum_{s=p+1, s\neq t}^{T} L\left(\frac{Y_{t-1} - Y_{s-1}}{h}\right) \mathbf{1}(y_s \leq y_t)}{\sum_{s=p+1, s\neq t}^{T} L\left(\frac{Y_{t-1} - Y_{s-1}}{h}\right)} \tag{1.59}$$

where $L(\cdot)$ denotes a known kernel function and h is the smoothing parameter applied in the kernel estimation. Besides, in their study, Jeong et al. (2012, p. 10) have put forth the re-scaled statistics $Th^p \hat{J}_T / \hat{\sigma}_0$ being asymptotically distributed as standard normal where $\hat{\sigma}_0 = \sqrt{2\theta(1-\theta)} \sqrt{\frac{1}{T(T-1)h^{2p}}} \sqrt{\sum_{t\neq s} K^2\left(\frac{(Z_{t-1} - Z_{s-1})}{h}\right)}$.

In this study, the aim is specifically to examine the causality regarding the volatility spillover between WilderHill Clean Energy Price Index or S&P Global Clean Energy Price Index returns and NYMEX Light Crude Oil Continuous Contract Settlement Prices, NYSE Arca Technology Price Index or S&P 500 Composite Price Index returns; therefore, a testing approach for the 2nd moment has been adopted by expanding the framework proposed by Jeong et al. (2012). Due to the fact that the rejection of causality in the moment m does not suggest a logical inference regarding non-causality in the moment k for $m < k$; some hardships can be come across while testing for Granger-

causality in the 2nd or higher order moments. Hence, testing procedures for causality have to be described cautiously (Balcilar, Bouri, Gupta, & Roubaud, 2016, p. 9).

Following Nishiyama et al. (2011), primarily the nonparametric Granger causality-in-quantiles test will be carried out by making the definition of a y_t process in order to be able to test causality in higher order moments:

$$y_t = g(Y_{t-1}) + \sigma(X_{t-1})\varepsilon_t \tag{1.60}$$

Here the vector X_{t-1} is defined as $X_{t-1} = (x_{t-1}, x_{t-2}, ..., x_{t-p})$, ε_t denotes a white noise process, $g(\cdot)$ and $\sigma(\cdot)$ denote unknown functions providing adequate stationarity conditions for y_t. Even though this specification does not take Granger-type causality testing from x_t to y_t into consideration, it is likely to reveal a predictive power of x_t for y_t^2 in case $\sigma(\cdot)$ is structured as a nonlinear function without having to add the squared terms for X_{t-1} as clearly discernible into the function $\sigma(\cdot)$. Reexpressing the formulations in Equations (1.53) and (1.54) in a way to be able to be convertible into the null and alternative hypotheses for testing causality-in-variance (second moment) take the form of

$$H_0 = P\left\{F_{y_t^2|Z_{t-1}}\left\{Q_\theta(Y_{t-1}) \mid Z_{t-1}\right\} = \theta\right\} = 1 \tag{1.61}$$

$$H_1 = P\left\{F_{y_t^2|Z_{t-1}}\left\{Q_\theta(Y_{t-1}) \mid Z_{t-1}\right\} = \theta\right\} < 1 \tag{1.62}$$

For testing the null in Equation (1.61), the process y_t has been substituted in Equations (1.56)-(1.59) with y_t^2. On the other hand, when causality exists in both mean and variance, the description of causality is more appropriate to be rearranged as in the following model for higher-order moments:

$$y_t = g(X_{t-1}, Y_{t-1}) + \varepsilon_t \tag{1.63}$$

and also depending on the framework proposed by Nishiyama et al. (2011), the hypotheses regarding causality-in-quantile testing approach for higher-order moments can be stated as

$$H_0 = P\left\{F_{y_t^k|Z_{t-1}}\left\{Q_\theta(Y_{t-1}) \mid Z_{t-1}\right\} = \theta\right\} = 1 \text{ for } k = 1, 2, ..., K \tag{1.64}$$

$$H_1 = P\left\{F_{y_t^k|Z_{t-1}}\left\{Q_\theta(Y_{t-1}) \mid Z_{t-1}\right\} = \theta\right\} < 1 \text{ for } k = 1, 2, ..., K \tag{1.65}$$

$\hat{\varepsilon}_t$ term in Equation (1.57) can be respecified in order to be used in Equation (1.56) as modified for the k -th order causality in the following way:

$$\hat{\varepsilon}_t^{(k)} = \mathbf{1}\left\{ y_t^{\,k} \le \hat{Q}_\theta(Y_{t-1}) \right\} - \theta \tag{1.66}$$

Here $\hat{Q}_\theta(Y_{t-1})$ represents an estimate of the θ -th conditional quantile of y_t considering Y_{t-1} and through the nonparametric kernel method, it is estimated as

$$\hat{Q}_\theta(Y_{t-1}) = \inf\left\{ y_t^{\,k} : \hat{F}_{y_t^{\,k}|Y_{t-1}}(y_t^{\,k} \mid Y_{t-1}) \ge \theta \right\} \tag{1.67}$$

and in this situation, the *Nadaraya Watson* kernel estimator $\hat{F}_{y_t^{\,k}|Y_{t-1}}(y_t^{\,k} \mid Y_{t-1})$ will be

$$\hat{F}_{y_t^{\,k}|Y_{t-1}}(y_t^{\,k} \mid Y_{t-1}) = \frac{\sum_{s=p+1, s\neq t}^{T} L\left(\frac{Y_{t-1} - Y_{s-1}}{h}\right) \mathbf{1}(y_s^{\,k} \le y_t^{\,k})}{\sum_{s=p+1, s\neq t}^{T} L\left(\frac{Y_{t-1} - Y_{s-1}}{h}\right)} \tag{1.68}$$

To find out whether x_t is the Granger cause of y_t in θ -th quantile up to the K -th moment or not, the test statistic in Equation (1.56) is constructed for each k allowing for the information covered in Equation (1.64). However, bringing the different test statistics for each k together in a way to form one general statistic for the joint null H_0 in Equation (1.64) is an unattainable task because of the statistics being correlated in a reciprocal manner (Nishiyama et al., 2011). In order to cope with this concern, a sequential-testing approach described by Nishiyama et al. (2011) has been pursued. Initially, the nonparametric Granger causality in the mean (i.e. $k = 1$) is tested. In case the null of non-causality is rejected, the process ends at this point and it can be regarded as a robust indicator for the existence of a probable quantile-based causality-in-variance (i.e. $k = 2$). On the other hand, the absence of causality-in-mean does not spontaneously signify that there is also no causality for $k = 2$. Thus, as a result of not being able to reject the null for 1st moment; we can continue performing the tests for the 2nd moment (Balcilar, Akadiri, Gupta, & Miller, 2017, pp. 6-8; Balcilar, Gupta, & Kyei, 2018, p. 79).

1.4. Empirical Findings

In this application part, it has been utilized from "*quantreg.nonpar*" package which is used to detect the point estimates of the conditional quantile function and its derivatives based on series approximations method to the nonparametric argument of the partially linear conditional quantile model (Lipsitz et al., 2016, p. 370). This part tries to reveal the

multivariate causal linkages in the nonparametric quantile framework running from three different variables to clean energy price indexes. In this paper, monthly time series data covering 2003m11-2018m1 period -which are WilderHill Clean Energy Price Index (WCE), S&P Global Clean Energy Price Index (SPCE), NYMEX Light Crude Oil Continuous Contract Settlement Prices (OIL), NYSE Arca Technology Price Index (TECH) and S&P 500 Composite Price Index (SP500)- have been utilized with totally 171 observations and extracted out of Thomson Reuters database. All of these variables are integrated of order 1; that's why their differences have been taken in a way to represent logarithmic returns. In the entire analysis of the application part, the lag order p of the VAR has been determined as 1 by Schwarz (i.e. Bayesian) Information Criterion (SIC) which is asymptotically consistent contrary to the Akaike Information Criterion. Table 1 summarizes the descriptive statistics which are the mean, median, standard deviation, minimum and maximum, skewness and kurtosis values, JB test, Ljung Box test and ARCH-LM (Autoregressive Conditional Heteroscedasticity – Lagrange Multiplier) test results for WCE, SPCE, TECH, SP500 and OIL variables.

Table 1
Descriptive Statistics

	WCE	SPCE	TECH	SP500	OIL
Mean	106.9868	1196.178	300.8308	1522.382	71.16567
Median	86.44000	1007.214	257.2600	1365.740	68.05000
Std.Dev.	63.71144	803.3323	118.1607	456.3081	23.71913
Min	38.01000	398.5320	143.1900	773.1400	29.74000
Max	288.3600	3845.249	685.1900	2837.540	137.0000
Skewness	0.737021	1.447813	1.142909	0.752833	0.313481
Kurtosis	2.281411	4.223255	3.592520	2.626564	2.202959
JB	19.16033*	70.40211*	39.72930*	17.14620*	7.327031*
Q(1)	165.66*	165.15*	160.25*	161.66*	154.50*
Q(12)	1526.0*	1436.7*	1379.2*	1440.9*	803.48*
ARCH(1)	154.92933*	154.98373*	168.24951*	167.61847*	151.39112*
ARCH(12)	147.3771*	146.39686*	157.53020*	156.90692*	142.45096*

Note: * denotes significant values at 5% level of significance. JB represents Jarque-Bera normality test statistics; Q(1) and Q(12) represent the first and the twelfth Ljung-Box autocorrelation test statistics, ARCH(1) and ARCH(12) represent the first and the twelfth order ARCH-LM (Autoregressive Conditional Heteroscedasticity–Lagrange Multiplier) statistics.

Kurtosis values reveal that SPCE and TECH follow a higher peak phenomenon with heavier tails in the distribution compared to WCE, SP500 and OIL which represent platykurtic distribution pattern with values less than three. It is also apparent that all variables display a non-normal distribution with positive skewness (skewed right) and not-mesokurtic structures. Amongst clean energy indexes, probability distribution of SPCE is more highly skewed with the value 1.447813 than WCE implying that SPCE at each time point can display more deviations from the optimal case compared to WCE and the median of SPCE is lower than the expected value of it as in all other positively skewed variables. As a result of the test proposed by Jarque and Bera (1980), JB statistics also verify the non-normality in all series by rejecting the null of normal distribution at 5% level of significance. Besides, the first Q(1) and the twelfth Q(12) order Ljung-Box autocorrelation test statistics reveal a serious autocorrelation problem and significant ARCH effects are observed at the first and the twelfth orders for all variables according to ARCH(1) and ARCH(12) statistics.

Linear Granger causality test results as based upon the linear vector autoregressive (VAR) model have been presented in Table 2 under the null hypothesis saying that there exists no Granger causality from given series towards two clean energy price indexes WCE and SPCE.

Table 2
Linear Granger Causality Tests

	F-statistic[1]	*F*-statistic[2]	*F*-statistic[3]	Order of the VAR(p)
Δ lnWCE	2.30086	1.19761	5.29553**	1
Δ lnSPCE	1.11587	0.40128	2.88726*	1

Notes: [1] H_0 : There is no Granger causality from Δ lnTECH towards given indexes.

[2] H_0 : There is no Granger causality from Δ lnSP500 towards given indexes.

[3] H_0 : There is no Granger causality from Δ lnOIL towards given indexes.

The table records the *F*-statistic values under the null hypotheses specified above. ** and * indicate the rejection of the null of no Granger causality at 5% and 10% significance levels respectively.

In Table 2, findings have shown that there is no predictive linear relationship running from TECH and SP500 to two clean energy price indexes WCE and SPCE. The null of no Granger causality has been rejected only for OIL series concluding that Δ lnOIL Granger causes Δ lnWCE and Δ lnSPCE at 5% and 10% significance levels respectively.

Table 3

BDS Test Results

Epsilon Method: Fraction of Pairs			
	TECH	SP500	OIL
WCE	(m = 2) 2.971191***	(m = 2) 2.739529***	(m = 2) 2.044976**
	(m = 3) 2.720213***	(m = 3) 2.465361**	(m = 3) 1.834369*
	(m = 4) 2.943724***	(m = 4) 2.625743***	(m = 4) 1.807079*
	(m = 5) 3.073375***	(m = 5) 2.816539***	(m = 5) 1.856532*
	(m = 6) 3.246857***	(m = 6) 2.979666***	(m = 6) 1.955709*
SPCE	(m = 2) -0.082389	(m = 2) -0.080439	(m = 2) -0.079475
	(m = 3) -0.111912	(m = 3) -0.109234	(m = 3) -0.107925
	(m = 4) -0.135406	(m = 4) -0.132131	(m = 4) -0.130547
	(m = 5) -0.156050	(m = 5) -0.152234	(m = 5) -0.150408
	(m = 6) -0.174974	(m = 6) -0.170649	(m = 6) -0.168600
Epsilon Method: Standard Deviations			
	TECH	SP500	OIL
WCE	(m = 2) 2.022800**	(m = 2) 1.893409*	(m = 2) 1.674223*
	(m = 3) 2.183939**	(m = 3) 1.991869**	(m = 3) 2.131912**
	(m = 4) 2.051284**	(m = 4) 1.795338*	(m = 4) 2.297218**
	(m = 5) 1.448287	(m = 5) 1.386519	(m = 5) 2.024156**
	(m = 6) 1.265267	(m = 6) 1.797117*	(m = 6) 2.304888**
SPCE	(m = 2) 1.308811	(m = 2) 0.902877	(m = 2) 0.139695
	(m = 3) 2.268503**	(m = 3) 1.743586*	(m = 3) 1.301957
	(m = 4) 3.229567***	(m = 4) 2.800103***	(m = 4) 2.276836**
	(m = 5) 3.870278***	(m = 5) 3.785576***	(m = 5) 3.266217***
	(m = 6) 4.869490***	(m = 6) 4.712145***	(m = 6) 4.066282***

Note: Z-statistics based on the BDS test [Broock et al. (1996)] -where the null hypothesis states the i.i.d. residuals- have been presented according to the two epsilon methods. The default value of epsilon is 0.7. ***, ** and * indicate significant values at 1%, 5% ve 10% significance levels respectively. m represents the embedding dimension of the BDS test. All series in VAR(1) specification have been used in their logarithmic and first-differenced forms.

Now we investigate the nonlinear dependencies between TECH, SP500, OIL series and WCE, SPCE series performing the BDS test suggested by Broock, Scheinkman, Dechert and LeBaron (1996) based on the residuals derived from the first-order

unrestricted bivariate VAR model specification in this research. The null hypothesis in the BDS test says that residuals from VAR model are i.i.d..

Table 3 displays BDS test results. According to the results, the null of i.i.d. assumption is rejected under 'fraction of pairs' and 'standard deviations' epsilon methods for distinct embedding dimensions and significance levels as a solid indicator of both nonlinearity and therefore the requirement of examining for the presence of possible nonlinear causality between TECH, SP500, OIL and clean energy indexes except the case that residuals for SPCE index have been found to be i.i.d. in all equations under 'fraction of pairs' method.

Relying on linear Granger causality findings can give rise to unstable and contradictory results under strong nonlinear structures of the variables. For this reason, before implementing the causality-in-quantile framework which pays regard to outlier values, nonlinear associations or structural instabilities; non-parametric Hiemstra & Jones (1994) and Diks & Panchenko (2006) tests -which capture nonlinear causalities and will be named H-J and D-P tests respectively here- have been applied.

Table 4

Diks and Panchenko (2006) Nonlinear Causality Test

	$\Delta \ln TECH \rightarrow \Delta \ln WCE$		$\Delta \ln SP500 \rightarrow \Delta \ln WCE$		$\Delta \ln OIL \rightarrow \Delta \ln WCE$	
m	t-statistic	p-value	t-statistic	p-value	t-statistic	p-value
2	0.943	0.1728	0.686	0.2463	-0.124	0.5498
3	0.206	0.4185	0.129	0.4488	-0.625	0.7339
4	-0.128	0.5507	0.711	0.2385	-0.865	0.8065
	$\Delta \ln TECH \rightarrow \Delta \ln SPCE$		$\Delta \ln SP500 \rightarrow \Delta \ln SPCE$		$\Delta \ln OIL \rightarrow \Delta \ln SPCE$	
m	t-statistic	p-value	t-statistic	p-value	t-statistic	p-value
2	1.336	0.0908*	0.972	0.1656	-0.505	0.6930
3	0.532	0.2973	0.261	0.3972	-0.809	0.7906
4	-0.148	0.5588	-0.994	0.8398	0.655	0.2561

Note: The default bandwidth value is 0.5. * denotes the significant value and m represents the embedding dimension as "1+maximum lag order".

Table 4 reports the D-P test results so that the null of nonlinear Granger non-causality could not be rejected for all embedding dimensions except for the case that there exists nonlinear Granger causality running from technology to S&P global clean energy

at 10% level of significance only for $m = 2$. As known, the D-P test copes with the issue of over-rejection of the null hypothesis encountered in the H-J test which appears as a modified test of Baek and Brock (1992). For comparative purposes, the H-J test results have also been included in the analysis. Table 5 presents the H-J test outcomes.

Table 5

Hiemstra and Jones (1994) Causality Test

Ly=Lx	$\Delta \ln TECH \rightarrow \Delta \ln WCE$		$\Delta \ln SP500 \rightarrow \Delta \ln WCE$		$\Delta \ln OIL \rightarrow \Delta \ln WCE$	
Lags	t-statistic	p-value	t-statistic	p-value	t-statistic	p-value
1	0.89437	0.81444	1.11878	0.86838	0.52643	0.70071
2	0.82323	0.79481	1.72639	0.95786	0.69786	0.75736
3	1.07489	0.85879	1.77919	0.96239	0.63677	0.73786
4	1.57640	0.94253	1.78046	0.96250	0.71683	0.76326
5	1.36107	0.91325	1.67449	0.95298	0.31507	0.62364
6	1.34019	0.90991	1.61165	0.94648	-0.50387	0.69283
7	1.28705	0.90096	1.41904	0.92205	0.03473	0.51385
8	1.11921	0.86847	1.14576	0.87405	-0.22230	0.58796
Ly=Lx	$\Delta \ln TECH \rightarrow \Delta \ln SPCE$		$\Delta \ln SP500 \rightarrow \Delta \ln SPCE$		$\Delta \ln OIL \rightarrow \Delta \ln SPCE$	
Lags	t-statistic	p-value	t-statistic	p-value	t-statistic	p-value
1	1.06723	0.85706	1.33449	0.90898	0.27156	0.60702
2	1.68833	0.95432	1.99192	0.97681	0.62219	0.73309
3	1.52921	0.93689	1.72867	0.95806	0.74069	0.77056
4	1.68854	0.95434	1.52670	0.93658	0.66113	0.74573
5	1.58742	0.94379	1.63018	0.94846	0.50567	0.69345
6	1.42030	0.92224	1.40216	0.91956	0.18597	0.57376
7	1.34197	0.91019	1.09565	0.86338	0.74189	0.77092
8	1.60520	0.94577	0.79096	0.78551	0.61769	0.73161

Note: The value of epsilon has been chosen as 0.7.

H-J test findings have been seen to support the D-P test results with respect to failing to reject the null at both 5% and 10% levels of significance for all various lags. Since we could not detect a nonlinear Granger causal association in general, as a next step we conduct the nonparametric causality test in the quantile framework which enables to investigate the entire distribution rather than just mean or median of the response variable.

Nonparametric causality-in-quantile framework relies on stationary data so that first differences of the natural logarithmic forms of all price index series have been taken and multiplied by 100 in order to calculate as percentage value. For each quantile index in nonparametric causality-in-quantile application, the bandwidth method has been chosen as 'Leave-One-Out Least Squares Cross Validation (LSCV)' proposed by Racine & Li (2004) and Li & Racine (2004). In addition, 'Local Polynomial Kernel Regression (LPRQ)' has been used as nonparametric quantile regression method and Gaussian type kernels have been utilized in the research.

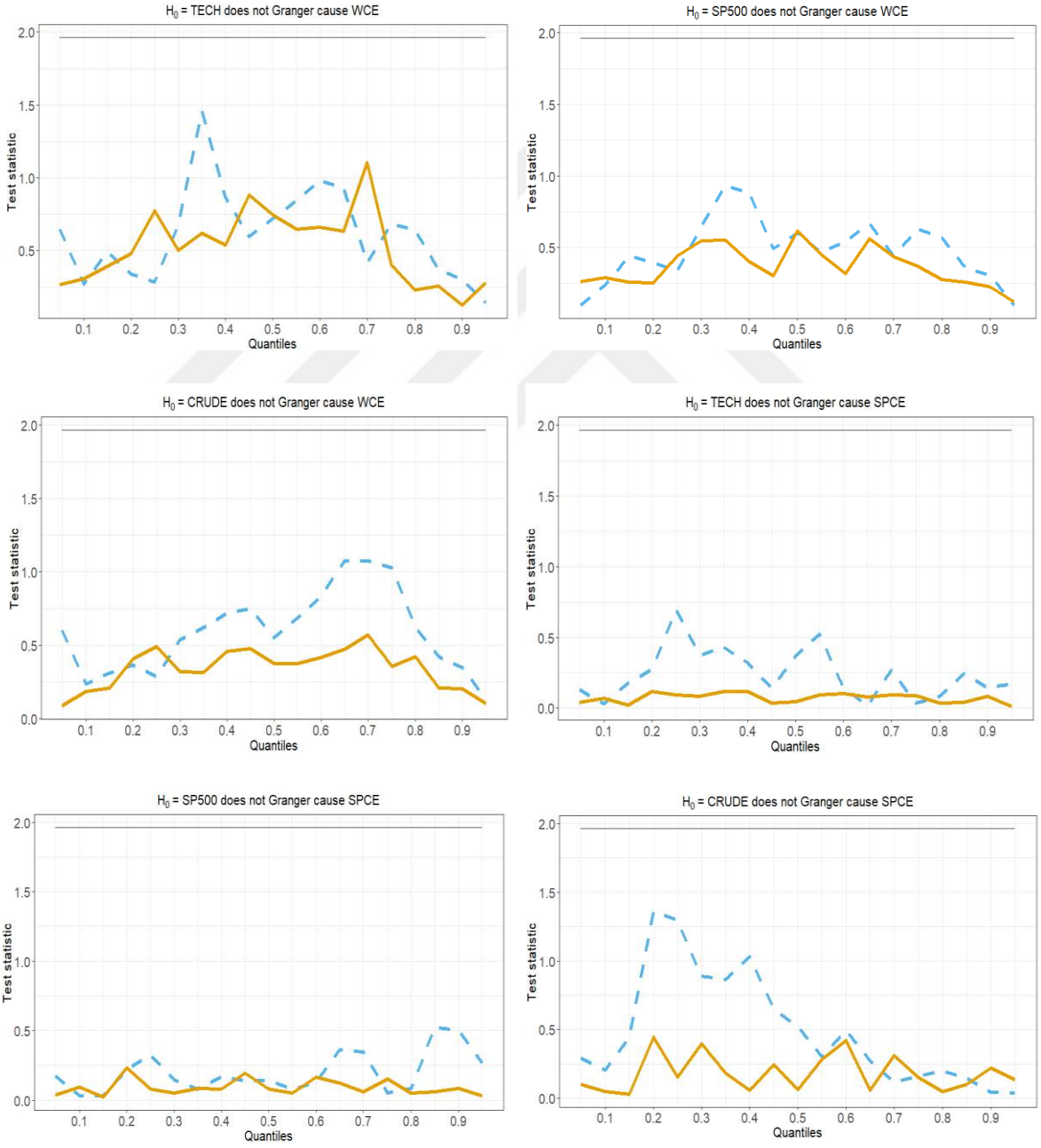

Figure 1. Test outcomes of Granger causality-in-quantile based on LPRQ with LSCV

Figure 1 reveals the causality-in-quantile test outcomes from TECH, SP500 and OIL to clean energy indexes which are WCE and SPCE based on LPRQ using the bandwidth method LSCV. In Figure 1 and other figures subsequent to this, NYMEX Light Crude Oil (OIL) series has been expressed by the abbreviation "CRUDE". The horizontal and vertical axes define quantile levels and nonparametric Granger causality test statistics respectively along with the horizontal thin line specifying critical value at 5% level of significance. Besides, while the solid line in orange color shows the causality-in-mean results; the blue dashed line shows the results for causality-in-variance. According to Figure 1 results, there has been found no evidence of causality from TECH, SP500 and OIL to WCE and SPCE across all quantile indexes as a result of not rejecting the null. As a general look, we realize higher test statistics for nonparametric Granger causality in variance rather than in mean across a broad range of quantiles.

In order to investigate the causality linkages amongst variables more deeply, we try different combinations of nonparametric quantile regression methods (local polynomial kernel regression [LPRQ], spline smoothing [BSRQ], spline smoothing with penalty [RQSSQ], non-crossing quantile regression using B-splines with L1-norm difference penalties [PSRQ]) and bandwidth selection methods (least squares cross validation [LSCV], normal-reference rule-of-thumb [NR], Scott's normal reference rule [SCOTT] – (For Scott's normal reference rule, see Scott (1979, p. 608))). However, to save space, the causality-in-mean and variance test results based on the SCOTT bandwidth and test results based on PSRQ regression method will not be included graphically here.

Figure 2 depicts causality-in-mean test results for a broad range of quantiles based on smoothing with B-spline using the bandwidth LSCV. Generally, strong causality evidence have not been found. To state specifically, S&P 500 composite price index and NYMEX light crude oil price affect the mean returns of WilderHill clean energy price index only at the lower and upper extreme quantiles respectively. NYSE Arca technology price index has also a limited effect on the mean returns of S&P global clean energy price index only at the upper extreme quantiles. The least effect on the mean returns of S&P global clean energy price index has come from NYMEX light crude oil price at quantile levels between about 0.50-0.60.

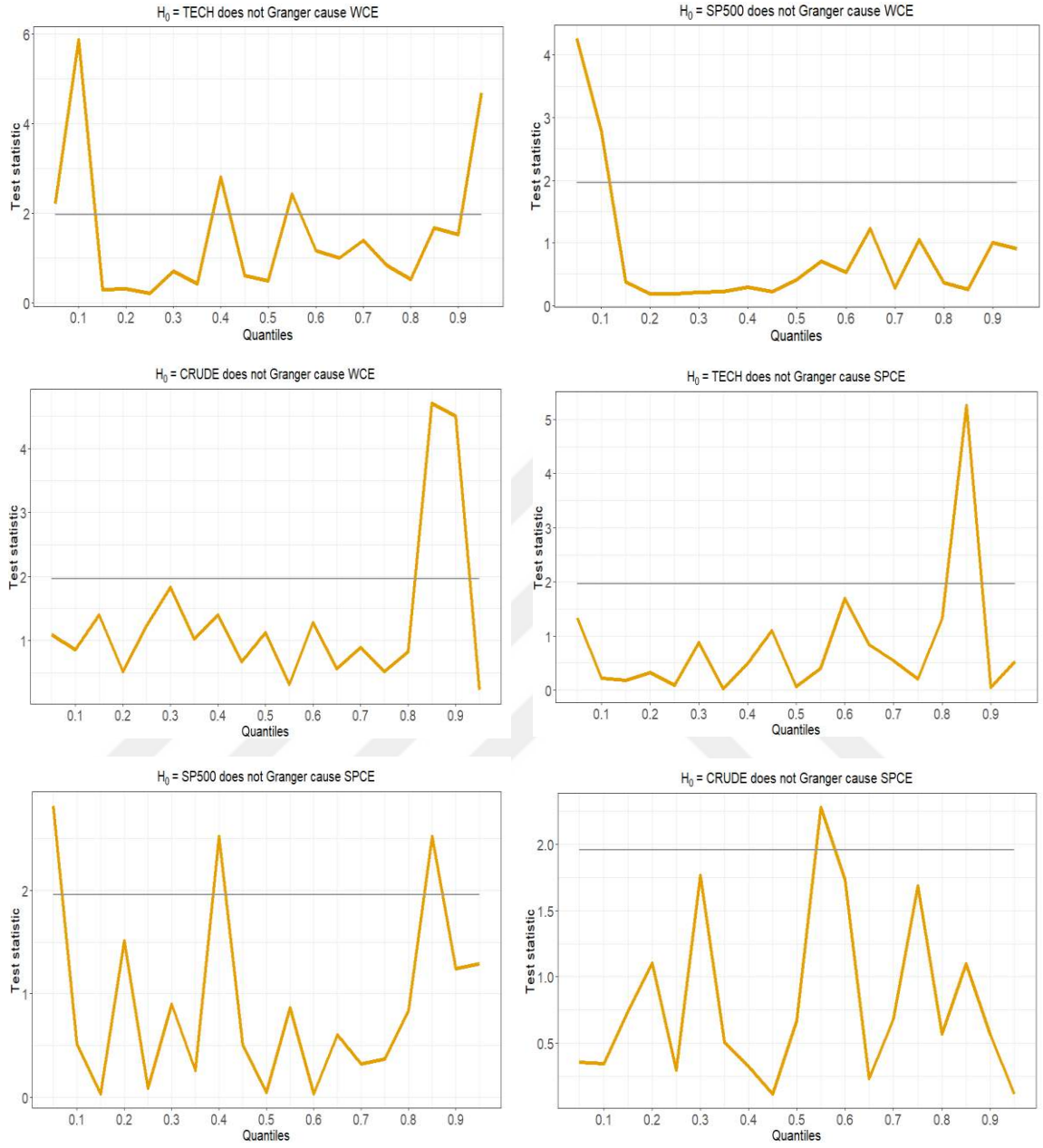

Figure 2. Quantile causality-in-mean test results based on BSRQ with LSCV

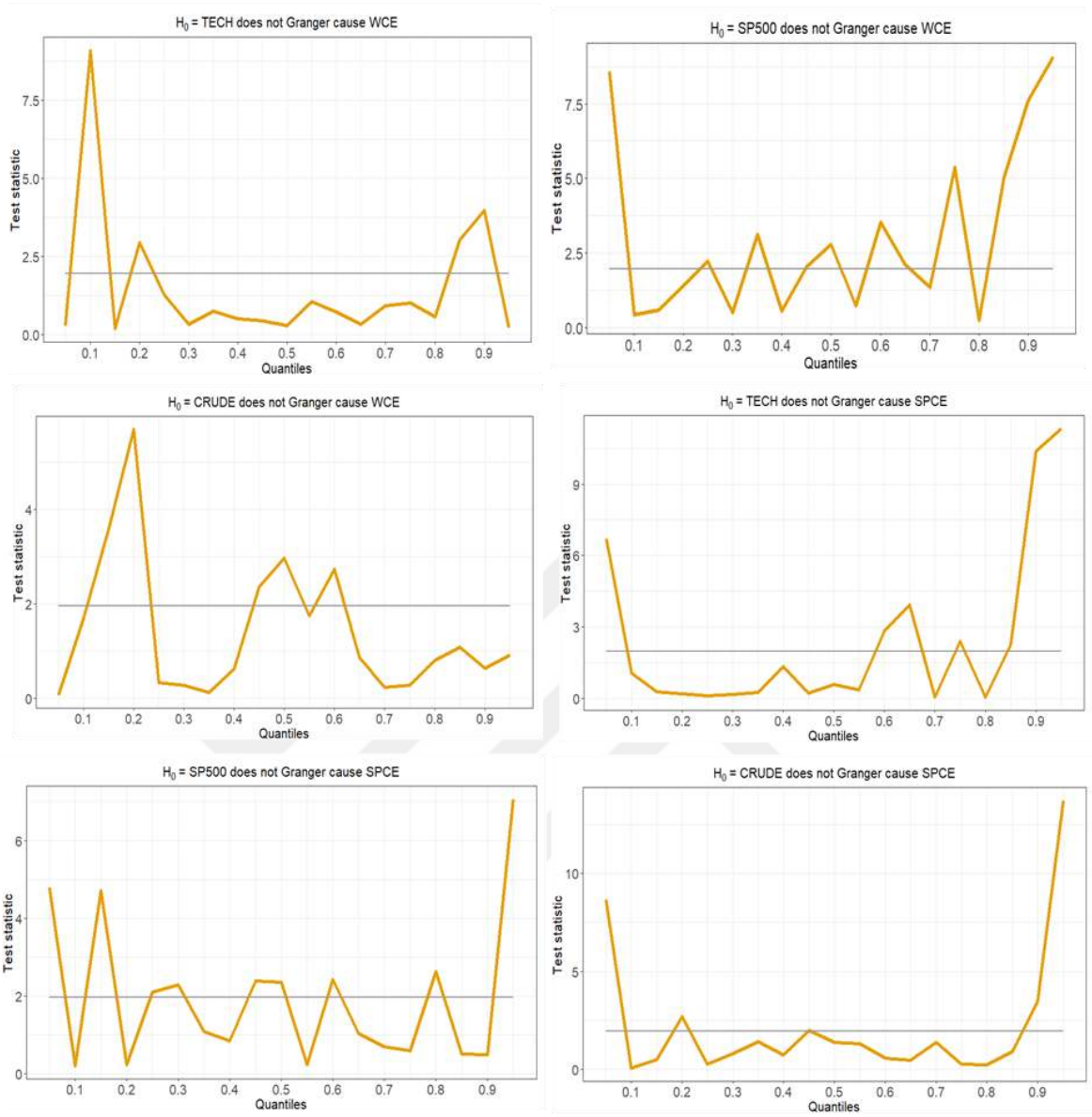

Figure 3. Quantile causality-in-variance test results based on BSRQ with LSCV

Figure 3 shows the nonparametric quantile test outcomes regarding volatility based on smoothing with B-spline using the bandwidth LSCV and reveals that technology, S&P 500 and crude oil Granger cause the volatility of both clean energy indexes at various quantile levels.

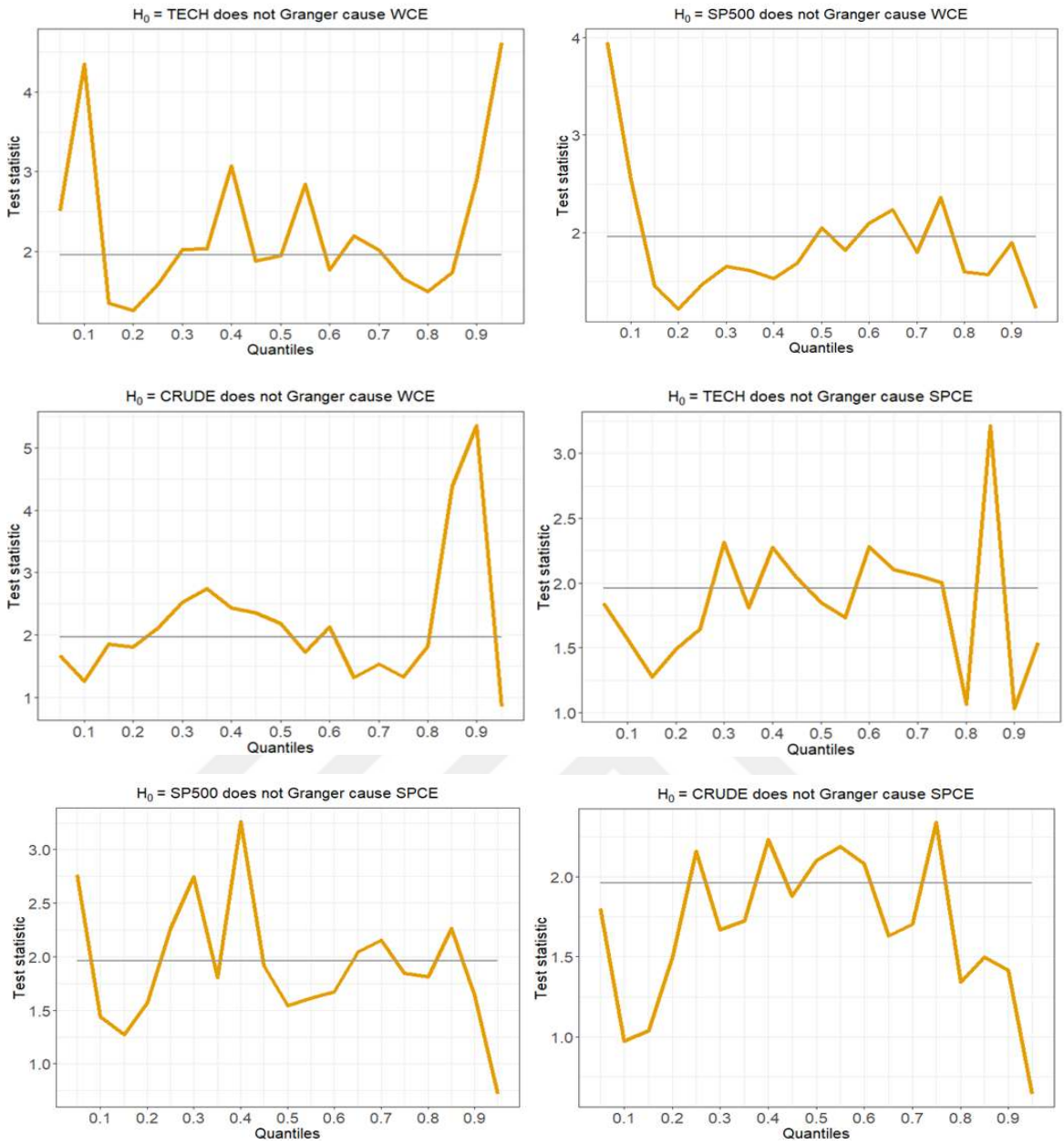

Figure 4. Quantile causality-in-mean test results based on BSRQ with NR

Figure 4 shows quantile causality test results for the 1st moment based on BSRQ method with the bandwidth NR and the findings based on smoothing with B-spline have shown strong causality linkages towards the mean returns of both clean energy indexes using the normal-reference rule-of-thumb bandwidth.

Figure 5 presents the causality-in-variance test results based on smoothing with B-spline using the normal-reference bandwidth.

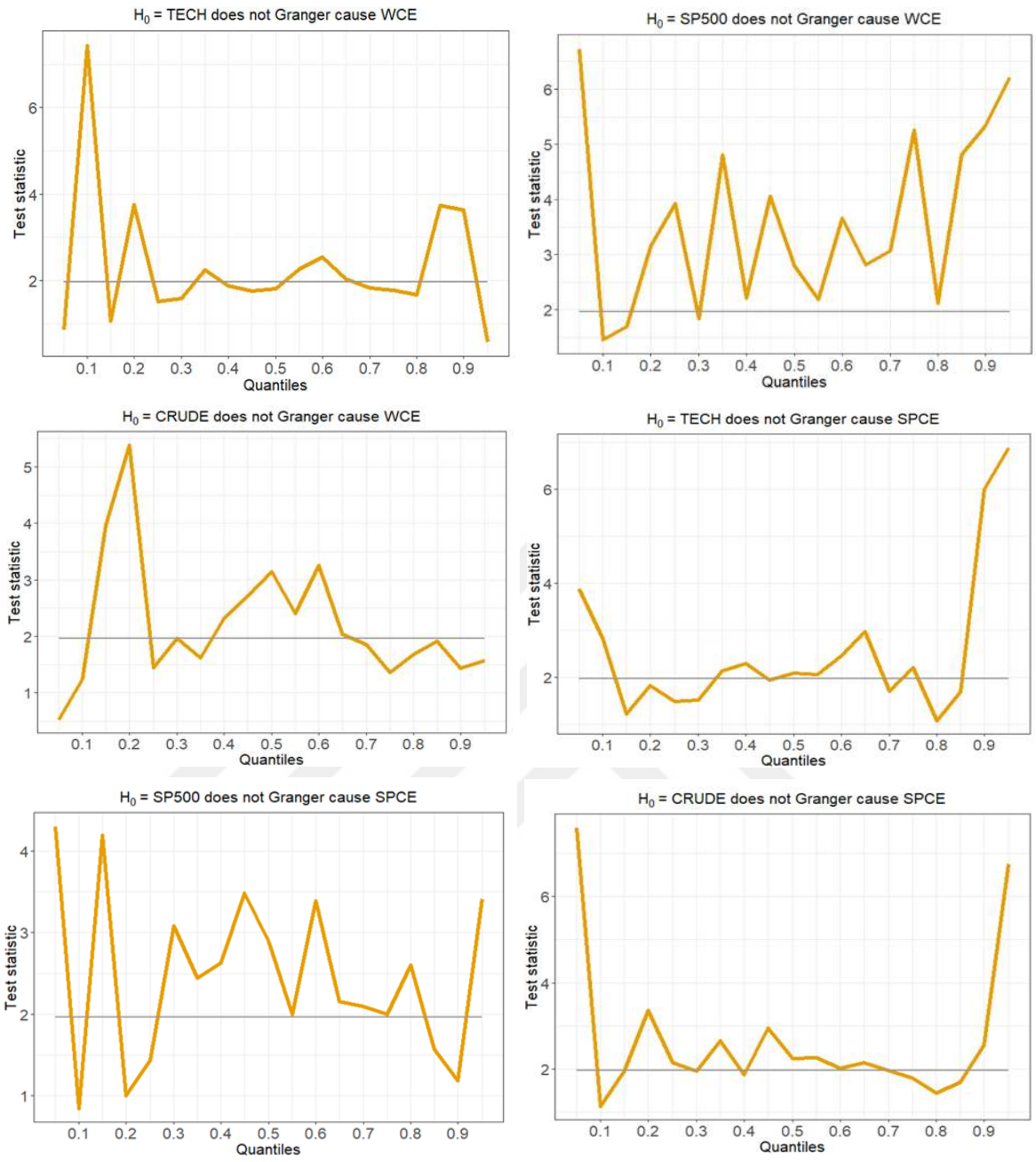

Figure 5. Quantile causality-in-variance test results based on BSRQ with NR

To speak generally, the volatilities of both clean energy indexes have been influenced by technology, S&P 500 and crude oil much more when compared to the mean returns of both clean energy indexes. The evidence support strong causality linkages at the 2nd moment amongst all given variables. It is remarkable to say that S&P 500 composite price causes WilderHill clean energy for all quantile levels at upper than about 0.15.

When the nonparametric regression method is chosen as spline smoothing with penalty, no causality finding has been come across for both 1st and 2nd moments using

the bandwidth LSCV which implies that technology, S&P 500 and crude oil prices do not Granger cause both clean energy indexes at 5% level of significance.

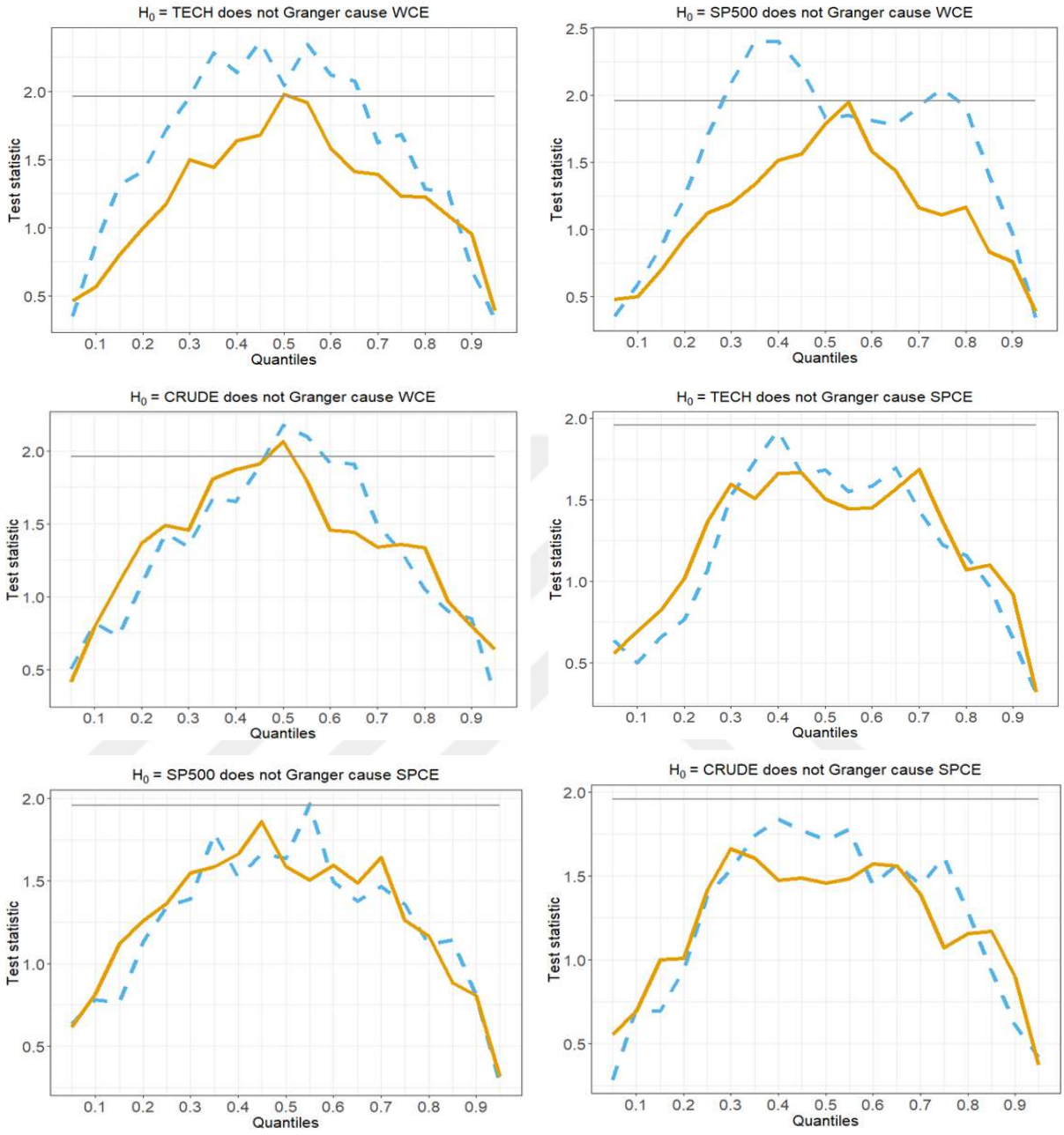

Figure 6. Test outcomes of Granger causality-in-quantile based on spline smoothing with penalty based on the normal reference bandwidth

Another application regarding causality-in-quantile tests has been founded upon spline smoothing with penalty based on the normal-reference rule-of-thumb bandwidth as given in Figure 6 where the solid line and blue dashed line represent test statistics for causality-in-mean and variance respectively. According to the findings, technology and crude oil price do not Granger cause S&P global clean energy under no circumstances for 1st and 2nd moments; however, only the volatility of S&P global clean energy price is

weakly influenced by S&P 500 composite price index for 0.55th quantile level. Typically, there is no Granger causality running from technology to the mean returns of WilderHill clean energy with a minor exception at 0.5th quantile level which can be virtually regarded as a break-even point with the horizontal thin line showing the critical value at 5% level of significance. On the other hand, technology Granger causes the volatility of WilderHill clean energy for the quantile levels specifically between 0.30-0.65. Apart from this, the mean returns of WilderHill clean energy have not been affected by S&P 500 composite price contrary to its volatility where there is a causation at some quantile levels. Also, crude oil price Granger causes both the mean returns and volatility of WilderHill clean energy price with a higher effect on the volatility for the quantiles between 0.45-0.60.

Table 6

A Summary of Nonparametric Granger Causality Test Results for Various Regression Methods with Least Squares Cross Validation

Bandwidth Method: LSCV						
	Δ lnTECH $\rightarrow$ Δ lnWCE		Δ lnSP500 $\rightarrow$ Δ lnWCE		Δ lnOIL $\rightarrow$ Δ lnWCE	
Method:	Mean	Variance	Mean	Variance	Mean	Variance
LPRQ	**No**	**No**	**No**	**No**	**No**	**No**
BSRQ	various	various	lower q.	various	upper q.	Various
RQSSQ	**No**	**No**	**No**	**No**	**No**	**No**
PSRQ	**No**	**No**	**No**	**No**	**No**	**No**
	Δ lnTECH $\rightarrow$ Δ lnSPCE		Δ lnSP500 $\rightarrow$ Δ lnSPCE		Δ lnOIL $\rightarrow$ Δ lnSPCE	
Method:	Mean	Variance	Mean	Variance	Mean	Variance
LPRQ	**No**	**No**	**No**	**No**	**No**	**No**
BSRQ	upper q.	various	Various	various	0.50-0.60	Various
RQSSQ	**No**	**No**	**No**	**No**	**No**	**No**
PSRQ	**No**	**No**	**No**	**No**	**No**	**No**

Note: "**No**" states "There is no Granger causality". Entries in the table represent the approximate quantile levels where test statistics are significant so that "various" means the existence of causality for different quantile levels. In addition, "lower q." and "upper q." mean lower and upper extreme quantiles respectively.

A summary of causality-in-quantile test results based on LSCV bandwidth has been presented in Table 6. In general terms, it is obvious that there is no Granger causality

association towards clean energy indexes for both 1st and 2nd moments in case a regression method apart from BSRQ is used.

Table 7

A Summary of Nonparametric Granger Causality Test Results for Various Regression Methods with Normal-Reference Rule-of-Thumb

Bandwidth Method: NR						
	Δ lnTECH $\rightarrow$ Δ lnWCE		Δ lnSP500 $\rightarrow$ Δ lnWCE		Δ lnOIL $\rightarrow$ Δ lnWCE	
Method:	Mean	Variance	Mean	Variance	Mean	Variance
LPRQ	**-**	**-**	**-**	**-**	**-**	**-**
BSRQ	various	various	various	Various	various	Various
RQSSQ	**No***	0.30-0.65	**No**	various	0.45-0.50	0.45-0.60
PSRQ	**No**	0.30-0.70	**No****	0.30-0.65	0.50-0.60	0.30-0.75
	Δ lnTECH $\rightarrow$ Δ lnSPCE		Δ lnSP500 $\rightarrow$ Δ lnSPCE		Δ lnOIL $\rightarrow$ Δ lnSPCE	
Method:	Mean	Variance	Mean	Variance	Mean	Variance
LPRQ	**-**	**-**	**-**	**-**	**-**	**-**
BSRQ	various	various	various	Various	various	Various
RQSSQ	**No**	**No**	**No**	**No****	**No**	**No**
PSRQ	**No*****	**No**	**No**	**No******	**No**	0.10-0.35

Note: "**No**" states that: There is no Granger causality. Other entries in the table represent the approximate quantile ranges where test statistics are significant so that "various" means the existence of causality for different quantile levels. "**-**" means that no calculations have been made depending on singular design matrix error. *, **, *** and **** denote the exceptions for 0.50th, 0.55th, 0.35th and 0.20th quantile levels where Granger causality exists.

Table 7 reveals a summary of nonparametric causality test results when normal-reference rule-of-thumb bandwidth is taken into consideration. Keeping LPRQ out in the interpretations depending on the singular design matrix error; findings have shown that technology, S&P 500 and crude oil prices have an influence on the volatility of WilderHill clean energy index at specified quantiles regardless of whatever the regression method is. However, there has been found no Granger causality-in-mean only from technology and S&P 500 prices towards WilderHill clean energy index when RQSSQ and PSRQ regression methods are implemented. BSRQ regression method has always recorded causality linkages for both 1st and 2nd moments from all given variables towards both clean energy indexes. On the other hand, in general sense no causality-in-quantile

relationship for mean and variance has been detected towards S&P global clean energy prices for RQSSQ and PSRQ regression methods excluding causalities at some quantile levels shown with * and causality-in-variance from crude oil prices to S&P global clean energy prices using PSRQ method expressed in Table 7.

Table 8

A Summary of Nonparametric Granger Causality Test Results for Various Regression Methods with Scott's Normal-Reference Rule

Bandwidth Method: SCOTT						
	Δ lnTECH $\rightarrow$ Δ lnWCE		Δ lnSP500 $\rightarrow$ Δ lnWCE		Δ lnOIL $\rightarrow$ Δ lnWCE	
Method:	Mean	Variance	Mean	Variance	Mean	Variance
LPRQ	**-**	**-**	**-**	**-**	**-**	**-**
BSRQ	various	various	various	Various	various	Various
RQSSQ	**No**	very limited	**No**	0.30-0.45	**No**	**No**
PSRQ	**No**	0.50-0.60	**No**	0.40-0.60	**No**	0.40-0.60
	Δ lnTECH $\rightarrow$ Δ lnSPCE		Δ lnSP500 $\rightarrow$ Δ lnSPCE		Δ lnOIL $\rightarrow$ Δ lnSPCE	
Method:	Mean	Variance	Mean	Variance	Mean	Variance
LPRQ	**-**	**-**	**-**	**-**	**-**	**-**
BSRQ	upper q.	various	various	Various	**No**	Various
RQSSQ	**No**	**No**	**No**	**No**	**No**	**No**
PSRQ	**No**	**No**	**No**	**No**	**No**	**No**

Note: "**No**" states that: There is no Granger causality. Other entries in the table represent the approximate quantile ranges where test statistics are significant so that "various" means the existence of causality for different quantile levels. "**-**" means that no calculations have been made depending on singular design matrix error. "upper q." means upper extreme quantiles. "very limited" means that there are causal linkages at two or three quantile levels.

A summary of causality-in-quantile test based on Scott's normal-reference bandwidth has been presented in Table 8. Keeping LPRQ out in the interpretations again, there exist Granger causality associations from technology, S&P 500 and crude oil prices towards the mean returns and volatility of two clean energy indexes using BSRQ regression method except that there is no causality-in-mean from crude oil price to SPCE. For RQSSQ and PSRQ methods; SPCE has not been affected by technology, S&P 500 and crude oil prices for both 1st and 2nd moments. On the other hand, no causality-in-mean from given variables towards WCE has been determined; but, the volatility of WCE

has been affected by given variables and this is not a valid case for the linkage from crude oil price to WCE when RQSSQ method is considered.

In this paper, it has been utilized from Fisher's g test statistic proposed by Fisher (1929) to examine the existence of periodicities and thus to design which basis functions will be used.

Table 9

The Fisher's g Periodicity Test Results

	Test-statistic	p-value
WCE	0.047438	0.831568
SPCE	0.053333	0.634267
TECH	0.057013	0.507919
SP500	0.050923	0.718602
OIL	0.079867	0.082073

Note: The null hypothesis implies no periodicities.

According to the Table 9 results, the fourier basis whose usage is better in the case of periodic structures has not been preferred in this research due to the variables with non-periodic framework; rather B-spline and polynomial bases have been performed.

The optimal number of B-spline bases has been considered as 13 which is a result of $\text{nbasis} = \text{nbreaks} + \text{norder} - 2$ equation where the arguments nbasis, nbreaks and norder define the number of basis functions; the length of knot points between piecewise polynomial segments describing the B-spline and the order of specified basis (Here the default value is 4 which is equivalent to one plus the degree of basis whose value is 3 for cubic splines) respectively (Ramsay et al. 2018, p. 48). The break points (knots) have been located at quantiles ranging between 0 and 1 with the increments of 0.10, therefore the number of breaks is equal to 11. Another basis that we consider for measuring the average first derivative of the conditional quantile function related to clean energy indexes is the polynomial basis of degree 3 which serves as one of series approximation regressors like B-spline basis that will be incorporated into the nonparametric regression. In all applications of this paper, the sequence of analyzed quantiles has been defined as [0.05, 0.95] by the increments with 0.05 and 5000 bootstrap repetitions have been performed for the Pivotal and Gaussian methods. Comparisons of four inference processes based on the B-spline and polynomial bases have been displayed in Figure 7 and Figure 8 respectively where the average derivative estimates of the conditional quantile function are depicted. In all applications, uniform confidence bands and

unconditional standard errors have been formed for inference methods and confidence intervals (CI) have been constructed under 5% level of significance.

As we realize in both Figure 7 and Figure 8, the confidence bands generated are roughly similar for all methods except the gradient bootstrap. On the other hand, point estimates have not varied for all inference processes based on given series approximation technique.

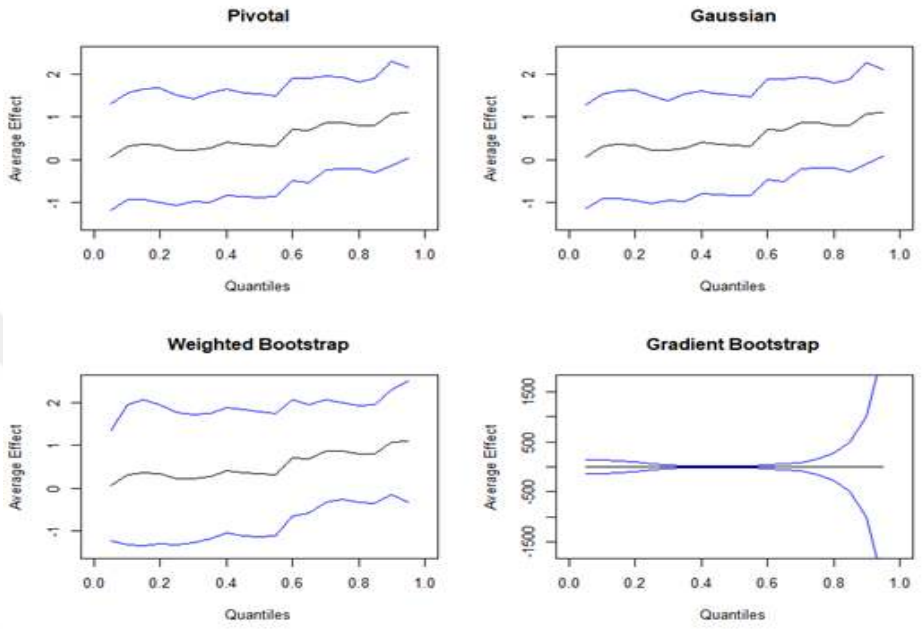

Figure 7: Average derivatives of the conditional quantile function of WCE in terms of TECH with 95% CI based on B-Spline basis

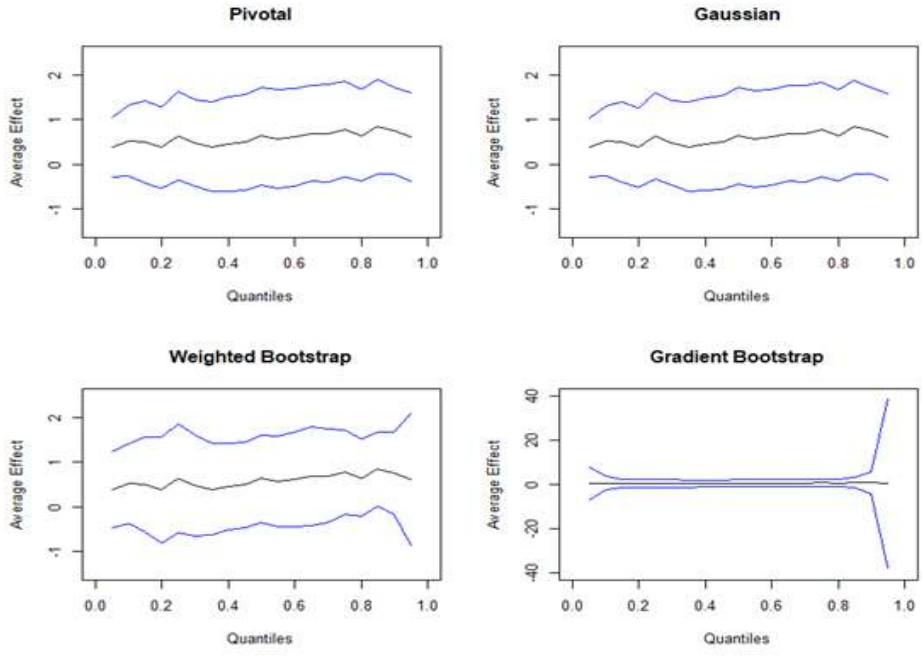

Figure 8: Average derivatives of the conditional quantile function of WCE in terms of TECH with 95% CI based on polynomial basis

The findings also reveal that uniform confidence intervals are narrower and therefore closer to point estimates in the case of polynomial basis as shown in Figure 8 when compared to the B-spline basis. To save space and set an example, only the pivotal inference output will be given place here.

Table 10

Quantile Estimates for WCE in terms of TECH under the Pivotal Inference with Polynomial Basis

Quantile	Point Estimate	Standard Error	Uniform 95% Confidence Interval		One-Sided 95% CIs Lower	Upper
0.05	0.3834	0.2355	-0.2776	1.044	-0.2175	0.9843
0.1	0.5288	0.2821	-0.2632	1.321	-0.1912	1.249
0.15	0.4989	0.3257	-0.4154	1.413	-0.3323	1.33
0.2	0.3745	0.3197	-0.5228	1.272	-0.4412	1.19
0.25	0.6357	0.3501	-0.3471	1.618	-0.2578	1.529
0.3	0.4754	0.3426	-0.4862	1.437	-0.3988	1.35
0.35	0.3927	0.3579	-0.6119	1.397	-0.5206	1.306
0.4	0.4535	0.3736	-0.5951	1.502	-0.4998	1.407
0.45	0.4916	0.3791	-0.5725	1.556	-0.4758	1.459
0.5	0.6307	0.3893	-0.4621	1.723	-0.3628	1.624
0.55	0.5658	0.3903	-0.5296	1.661	-0.4301	1.562
0.6	0.6075	0.386	-0.4761	1.691	-0.3776	1.593
0.65	0.6953	0.3817	-0.3762	1.767	-0.2788	1.669
0.7	0.6924	0.3896	-0.4013	1.786	-0.3019	1.687
0.75	0.7812	0.3804	-0.2866	1.849	-0.1896	1.752
0.8	0.6493	0.3668	-0.3802	1.679	-0.2866	1.585
0.85	0.8476	0.3763	-0.2086	1.904	-0.1126	1.808
0.9	0.7566	0.3445	-0.2104	1.724	-0.1225	1.636
0.95	0.6177	0.3491	-0.362	1.598	-0.273	1.508

Hypothesis Testing

Null Hypothesis:	*p-value*
theta (tau, w) <= 0 for all tau, w	3.158e-05
theta (tau, w) >= 0 for all tau, w	1
theta (tau, w) = 0 for all tau, w	0.2844

Table 10 presents point estimates which are equivalent to the average derivative estimates for WilderHill clean energy index with respect to technology under pivotal inference using polynomial basis approximation. According to the Table 10 results, the

null hypothesis which states that the average effect of TECH on WCE is negative has been rejected; however, we fail to reject the null hypotheses saying that the average effect of TECH on WCE is positive or equal to zero with p-values 1 and 0.2844 respectively at the 5% significance level for all quantile indexes of the distribution.

Table 11

A Comparison of P-Values Constructed by Different Inference Methods and Bases Approximations for the Average Derivatives of WCE with respect to TECH

Combinations	H_0 : Average Effect <= 0	H_0 : Average Effect >= 0	H_0 : Average Effect = 0
Pivotal, B-Spline	0.0001	0.9999	0.1056
Pivotal, Polynomial	0.0001	0.9999	0.2792
Gaussian, B-Spline	0.0001	0.9999	0.1059
Gaussian, Polynomial	0.0001	0.9999	0.2789
Weighted Bootstrap, B-Spline	0.0004	0.9996	0.1521
Weighted Bootstrap, Polynomial	0.0001	0.9999	0.1302
Gradient Bootstrap, B-Spline	0.1360	0.8640	0.4449
Gradient Bootstrap, Polynomial	0.0039	0.9961	0.2131

Table 11 shows p-values related to the average effects for each combination of inference methods & bases approximations and reveals that the null hypotheses stating that the average effect is positive or the average effect is zero cannot be rejected for all combinations at the 5% level of significance. On the other hand, the null hypothesis which implies that TECH affects WCE negatively has been rejected regardless of the type of bases and inference methods chosen when uniform 95% confidence bands are taken into account except for the gradient bootstrap procedure with B-Spline basis in which an insignificant p-value (0.1360) has been observed.

Figure 9 and Figure 10 depict a comparison of four inference processes based on the B-spline and polynomial bases respectively to examine the effect of SP500 on WCE. It is apparent to realize that the confidence bands are roughly similar for inference methods excluding the gradient bootstrap procedure and the point estimates are the same for four inference methods. Both figures with average derivative estimates being bounded away from zero reveal that SP500 affects WCE positively in the general sense.

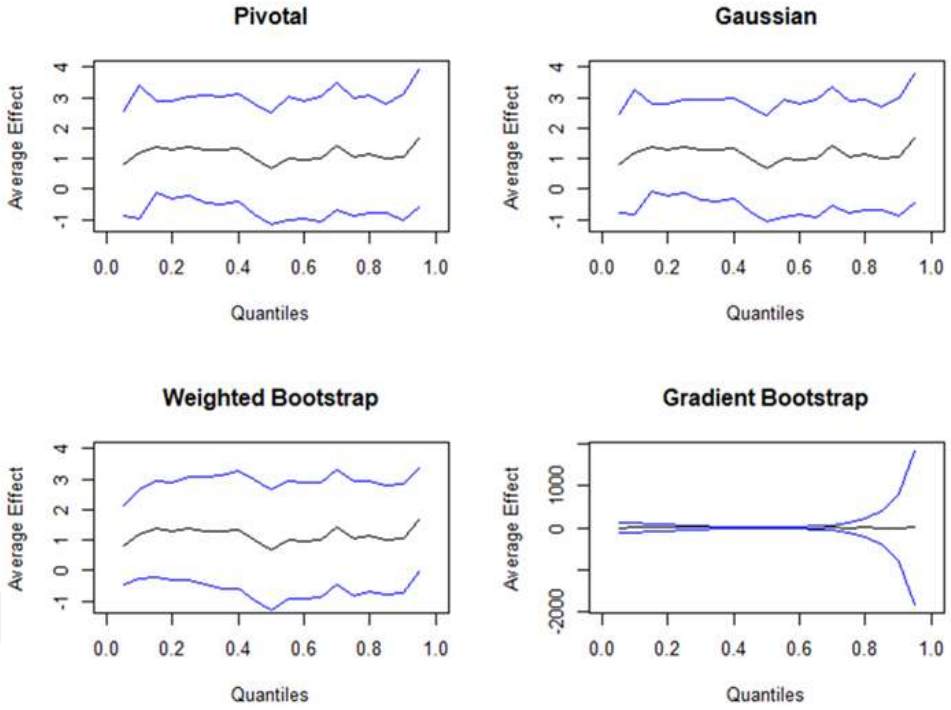

Figure 9. Average derivatives of the conditional quantile function of WCE in terms of SP500 with 95% CI based on B-Spline basis

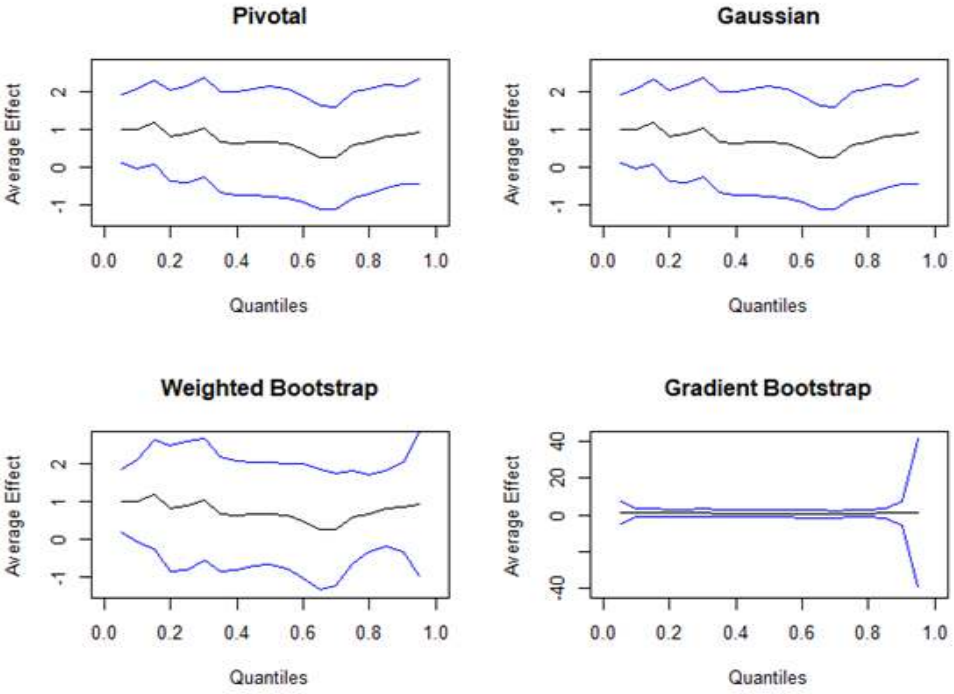

Figure 10. Average derivatives of the conditional quantile function of WCE in terms of SP500 with 95% CI based on polynomial basis

Table 12 presents a comparison of p-values regarding distinct nonparametric quantile regressions for evaluating the impact of SP500 on WCE. The findings show that

we fail to reject the null hypothesis saying that the average effect is positive and the null hypothesis saying that the average effect is equal to zero according to insignificant p-values at 5% level of significance except for the combination of weighted bootstrap method and polynomial basis in performing the zero average effect with a significant p-value (0.0743) at 10% significance level . On the other hand, the null of negative average effect has been rejected for all combinations except for the combination of Gradient bootstrap and B-spline basis.

Table 12

A Comparison of P-Values Constructed by Different Inference Methods and Bases Approximations for the Average Derivatives of WCE with respect to SP500

Combinations	H_0 : Average Effect <= 0	H_0 : Average Effect >= 0	H_0 : Average Effect = 0
Pivotal, B-Spline	0.0002	0.9998	0.2084
Pivotal, Polynomial	0.0000	1.0000	0.1126
Gaussian, B-Spline	0.0000	1.0000	0.2192
Gaussian, Polynomial	0.0000	1.0000	0.1091
Weighted Bootstrap, B-Spline	0.0000	1.0000	0.2256
Weighted Bootstrap, Polynomial	0.0000	1.0000	0.0743
Gradient Bootstrap, B-Spline	0.1024	0.8976	0.2956
Gradient Bootstrap, Polynomial	0.0026	0.9974	0.1417

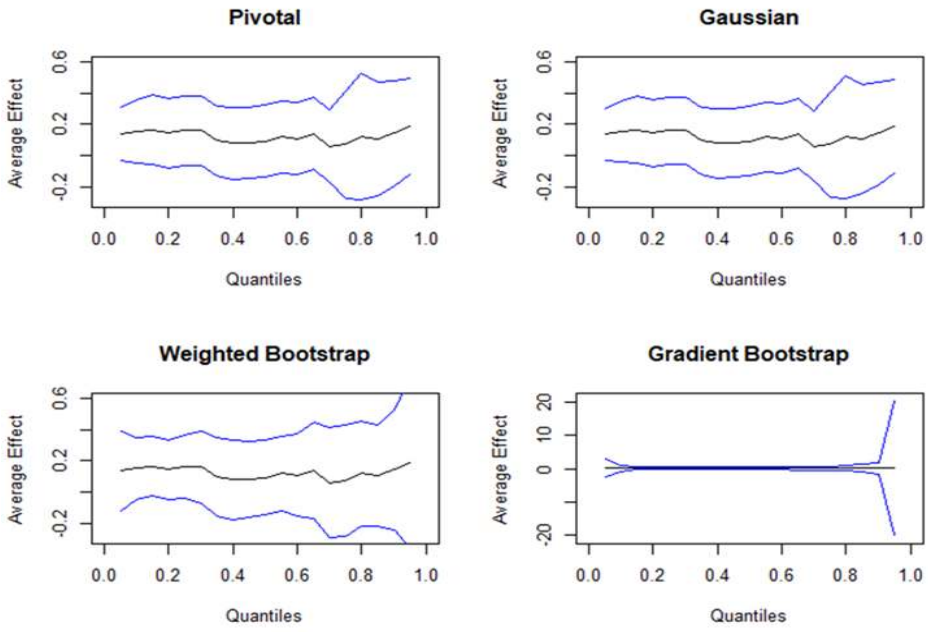

Figure 11. Average derivatives of the conditional quantile function of WCE in terms of OIL with 95% CI based on B-Spline basis

Figure 11 and Figure 12 depict a comparison of average derivative estimation results for WCE in terms of OIL under four inference processes with B-spline and polynomial bases respectively.

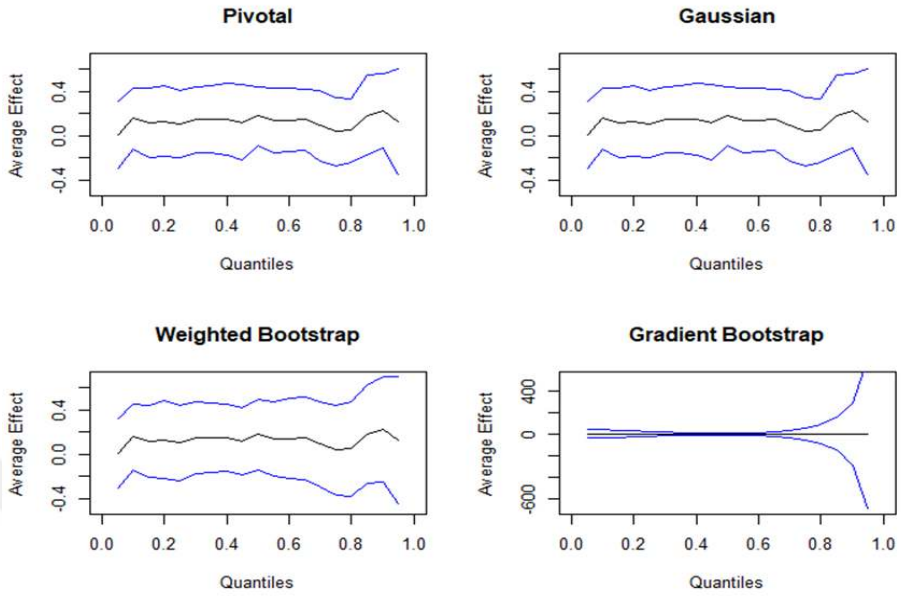

Figure 12. Average derivatives of the conditional quantile function of WCE in terms of OIL with 95% CI based on polynomial basis

As in the previous cases, confidence bands can again be expressed to be roughly similar to one another excluding the gradient bootstrap procedure. To speak generally, point estimates given in these figures can be inferred to take smaller values relative to Figure 7 - Figure 10. Therefore, OIL is concluded to have less impact on WCE when compared to TECH and SP500.

Table 13

A Comparison of P-Values Constructed by Different Inference Methods and Bases Approximations for the Average Derivatives of WCE with respect to OIL

Combinations	H_0 : Average Effect <= 0	H_0 : Average Effect >= 0	H_0 : Average Effect = 0
Pivotal, B-Spline	0.0045	0.9955	0.2506
Pivotal, Polynomial	0.0001	0.9999	0.2265
Gaussian, B-Spline	0.0032	0.9968	0.2416
Gaussian, Polynomial	0.0002	0.9998	0.2254
Weighted Bootstrap, B-Spline	0.0098	0.9902	0.3223
Weighted Bootstrap, Polynomial	0.0003	0.9997	0.1570
Gradient Bootstrap, B-Spline	0.1225	0.8775	0.4622
Gradient Bootstrap, Polynomial	0.0055	0.9945	0.1389

A comparison of p-values related to various nonparametric quantile regressions to examine the impact of OIL on WCE has been given in Table 13. According to the evidence, we fail to reject the null hypothesis saying that the average effect is positive and the null hypothesis saying that the average effect is equal to zero at 5% significance level; however, we reject the null of negative average effect in all cases except for the Gradient bootstrap & B-spline combination with an insignificant p-value (0.1225).

After examining the average derivative estimation results for WCE, the results for SPCE will be presented here. However, plots depicting the average derivative estimates will not be given place here this time for the sake of saving space.

Table 14

A Comparison of P-Values Constructed by Different Inference Methods and Bases Approximations for the Average Derivatives of SPCE with respect to TECH

Combinations	H_0 : Average Effect <= 0	H_0 : Average Effect >= 0	H_0 : Average Effect = 0
Pivotal, B-Spline	0.0016	0.9984	0.2018
Pivotal, Polynomial	0.0067	0.9933	0.0727
Gaussian, B-Spline	0.0013	0.9987	0.2034
Gaussian, Polynomial	0.0064	0.9936	0.0737
Weighted Bootstrap, B-Spline	0.0011	0.9989	0.1282
Weighted Bootstrap, Polynomial	0.0447	0.9553	0.3034
Gradient Bootstrap, B-Spline	0.0893	0.9107	0.3243
Gradient Bootstrap, Polynomial	0.0187	0.9813	0.2640

Table 14 shows the p-value results for the average impact of TECH on SPCE and it can be inferred that the only significant average effect has been obtained in Gaussian inference & polynomial basis approximation combination at 10% level of significance with a p-value of 0.0737. Besides, the average effect of TECH on SPCE is positive for all combinations. The negative effect has been come across only in Gradient bootstrap & B-spline combination for 5% significance level.

According to the Table 15 findings in which the average impact of SP500 on SPCE is investigated, the null hypotheses saying that the average effect is positive and the average effect is equal to zero cannot be rejected for all combinations at 5% significance level; however, the null of negative average effect has been rejected in all cases except for the Gradient bootstrap & B-spline combination.

Table 15

A Comparison of P-Values Constructed by Different Inference Methods and Bases Approximations for the Average Derivatives of SPCE with respect to SP500

Combinations	H_0 : Average Effect <= 0	H_0 : Average Effect >= 0	H_0 : Average Effect = 0
Pivotal, B-Spline	0.0000	1.0000	0.1493
Pivotal, Polynomial	0.0000	1.0000	0.1376
Gaussian, B-Spline	0.0000	1.0000	0.1501
Gaussian, Polynomial	0.0000	1.0000	0.1393
Weighted Bootstrap, B-Spline	0.0008	0.9992	0.2975
Weighted Bootstrap, Polynomial	0.0002	0.9998	0.1123
Gradient Bootstrap, B-Spline	0.1590	0.8410	0.3966
Gradient Bootstrap, Polynomial	0.0026	0.9974	0.1281

Table 16

A Comparison of P-Values Constructed by Different Inference Methods and Bases Approximations for the Average Derivatives of SPCE with respect to OIL

Combinations	H_0 : Average Effect <= 0	H_0 : Average Effect >= 0	H_0 : Average Effect = 0
Pivotal, B-Spline	0.0010	0.9990	0.1438
Pivotal, Polynomial	0.0113	0.9887	0.2539
Gaussian, B-Spline	0.0006	0.9994	0.1493
Gaussian, Polynomial	0.0128	0.9872	0.2666
Weighted Bootstrap, B-Spline	0.0091	0.9909	0.2320
Weighted Bootstrap, Polynomial	0.0168	0.9832	0.2206
Gradient Bootstrap, B-Spline	0.1064	0.8936	0.4567
Gradient Bootstrap, Polynomial	0.0180	0.9820	0.1617

Table 16 presents p-values concerning the average impacts of OIL on SPCE which has a conformity with Table 15 results so that the null hypotheses saying that the average effect is positive or equal to zero cannot be rejected at 5% significance level; but the null hypothesis implying negative average effect has been rejected for all cases except for the Gradient bootstrap & B-spline combination.

1.5. Concluding Remarks

Renewable energy has become an increasingly important trend over time. To this end, in this paper, it has been aimed to reveal the multivariate causality relationships based on the nonparametric quantile framework running from three different variables - namely NYMEX Light Crude Oil Continuous Contract Settlement Prices (OIL), NYSE Arca Technology Price Index (TECH) and S&P 500 Composite Price Index (SP500)- towards two clean energy price indexes which are WilderHill Clean Energy Price Index (WCE) and S&P Global Clean Energy Price Index (SPCE) for the period 2003m11-2018m1. Thus, six regression models have been handled. All variables have been joined into the analysis in the form of logarithmic returns. Firstly, linear Granger causality test results have pointed out that a predictive linear relationship has been detected running from OIL (but not from TECH & SP500) towards WCE and SPCE. Under the causality-in-quantile framework, Gaussian type kernels have been used in the whole analysis. The causality relationships amongst variables have been handled using different combinations of nonparametric quantile regression methods (local polynomial kernel regression [LPRQ], spline smoothing [BSRQ], spline smoothing with penalty [RQSSQ], non-crossing quantile regression using B-splines with L1-norm difference penalties [PSRQ]) and bandwidth selection methods (least squares cross validation [LSCV], normal-reference rule-of-thumb [NR], Scott's normal reference rule [SCOTT]).

In the light of causality-in-quantile test findings based on LSCV, causality linkages running from TECH, SP500 and OIL each individually towards both WCE and SPCE have been come across for both 1st and 2nd moments only when BSRQ method is performed. If necessary to express clearly, S&P 500 composite price index and NYMEX light crude oil price have an infuence on the mean returns of WCE only at the lower and upper extreme quantiles respectively. NYSE Arca technology price index causes the mean returns of SPCE only at the upper extreme quantiles and NYMEX light crude oil price has the least impact on the mean returns of SPCE at quantile levels only between approximately 0.50-0.60. The other causal linkages for all variables including causality-in-mean and causality-in-variance have been obtained for various quantile levels. Under the test results based on NR bandwidth and all regression methods excluding LPRQ, the evidence of Granger causality-in-mean has not been detected only from technology and S&P 500 prices towards WCE as a result of carrying out RQSSQ and PSRQ regression methods implying that the mean returns of WCE have been affected by only crude oil

prices. On the other hand, among all regression methods, in the general sense the causality relationships towards SPCE have not been recorded for RQSSQ and PSRQ regression methods excluding some quantile levels when both 1st and 2nd moments are considered. One important point here is that even when RQSSQ and PSRQ regression methods in which causality findings are almost none are considered together, it is seen that crude oil price has been the only variable which has an impact on the volatility of SPCE using PSRQ method.

Nonparametric Granger causality test results based on the Scott's Normal-Reference rule have revealed that taking into account all bandwidth methods; according to the BSRQ method in which causality relationships are encountered in all conditions, the only situation in which the causality relationship cannot be determined is the situation where a relationship from OIL to SPCE is examined. In brief, oil is the only variable that has no effect on the mean returns of SPCE. Besides; TECH, SP500 and OIL (each individually) have a greater impact on the volatility of WCE rather than on the mean returns of it.

To sum up the important details, when all bandwidth methods are taken into consideration in the nonparametric quantile causality framework, the method that reveals the causality findings in the best way has been determined as the smoothing with B-spline (BSRQ) method. Even in the case of regression methods where causality findings are almost absent, it is seen that to say the least crude oil price has been the only variable which influences the volatility of SPCE when normal reference rule-of-thumb is taken into account. Therefore, the importance of crude oil prices on SPCE is remarkable and much more than TECH & SP500. On the other hand, in case SCOTT bandwidth method is chosen, crude oil has not been able to create an influence on the mean returns of SPCE. So, crude oil is of great significance in explaining the volatility of SPCE rather than the mean returns of it.

In addition, the comparison of four inference processes -namely Pivotal, Gaussian, Weighted Bootstrap and Gradient Bootstrap- based on the B-spline and polynomial bases where the average derivative estimates of the conditional quantile function are computed has revealed that the null hypotheses implying that the average effect of TECH on WCE is positive or zero cannot be rejected and the only exception for negative effect has been come across in Gradient bootstrap method with B-spline where the null could not be rejected. These same results are valid for the average derivatives of WCE with respect to both SP500 and OIL when 5% significance level is considered. TECH, SP500 and OIL each seperately have affected WCE positively for all combinations; only in case Gradient

bootstrap with B-spline is performed, WCE has been influenced negatively by these variables. It is notable to say that OIL is seen to have less impact on WCE compared to TECH and SP500 when the values of point estimates are evaluated.

As to the results for SPCE, the only significant average effect (in terms of testing the equality to zero) of TECH on SPCE has been obtained in Gaussian inference & polynomial basis approximation combination at 10% level of significance. Generally, the average effect of TECH on SPCE is positive for all combinations. The negative effect has been detected only for Gradient bootstrap & B-spline combination and 5% level of significance and this result is valid also for the average effect of SP500 on SPCE. So, the average effects of SP500 on SPCE can also be said to be either positive or zero. The same results for SP500 can also be attributed to OIL regarding the average derivatives of SPCE.

As a conclusion, this research paper has been thought to contribute to the literature in terms of handling the multivariate causations of TECH, SP500 and OIL on two clean energy indexes -namely WCE and SPCE variables- on the nonparametric quantile framework and average derivative estimations all together presenting the results for many methods. The average derivative estimations have revealed important findings so that for all combinations of the two series approximating functions and four inference processes, the average effects of TECH, SP500 and OIL on WCE and SPCE are not negative -except the Gradient bootstrap and B-spline combination- thus showing consistency with general findings in the literature. No doubt, energy sector is highly dependent on oil and it can be inferred from our research findings that especially, this dependence on crude oil has affected the volatility of clean energy stock indexes significantly and positively depending on the expectation that price increases realized in crude oil which are usually an indicator of inflationary pressures -therefore affected also by interest rate shocks and may be stemmed from business cycles- form an inclination towards adopting (non-petroleum based) renewable energy sources. Depending on business cycle effects arising from oil prices, technology stock prices are also expected to create a positive impact on clean energy stocks. Briefly, energy security issues or environmental & climatic issues like global warming caused by carbon dioxide emissions or other contaminants trigger oil price movements and afterwards observed price peaks in oil create a surge for substituting oil by clean energy. In this process, renewable energy emerges as one of the prerequisites for energy independence and long-run investments on clean energy companies surge due to the need for new technologies to obtain clean energy, so depending also on economic development. Since the growing demand for renewable energy heavily depends on oil

prices and technology stock prices, characterizing the linkages among them is of great importance for policymakers.

REFERENCES

Asplund, R. W. (2008). *Profiting from clean energy: A complete guide to trading green in solar, wind, ethanol, fuel cell, carbon credit industries, and more*. Hoboken, New Jersey: John Wiley & Sons.

Baek, E., & Brock, W. (1992). *A general test for Granger causality: Bivariate model.* Technical Report, Iowa State University and University of Wisconsin, Madison.

Bagby, R. J. (2001). *Introductory analysis: A deeper view of calculus.* San Diego, CA: Harcourt/Academic Press.

Bahraoui, Z., Bahraoui, F., Bahraoui, M. A., & Fassi, K. E. (2018). Improving the rate convergence of nonparametric quantile. *International Journal of Contemporary Mathematical Sciences, 13*(6), 255-262.

Balcilar, M., Akadiri, S., Gupta, R., & Miller, S. (2017). *Partisan conflict and income distribution in the United States: A nonparametric causality-in-quantiles approach* (No. 2017-11). University of Connecticut, Department of Economics Working Paper Series.

Balcilar, M., Bekiros, S., & Gupta, R. (2017). The role of news-based uncertainty indices in predicting oil markets: A hybrid nonparametric quantile causality method. *Empirical Economics*, *53*(3), 879-889.

Balcilar, M., Bouri, E., Gupta, R., & Roubaud, D. (2016). *Can volume predict Bitcoin returns and volatility? A nonparametric causality-in-quantiles approach* (No. 2016-62). University of Pretoria, Department of Economics.

Balcilar, M., Gupta, R., & Kyei, C. (2018). Predicting stock returns and volatility with investor sentiment indices: A reconsideration using a nonparametric causality-in-quantiles test. *Bulletin of Economic Research*, *70*(1), 74-87.

Balcilar, M., Gupta, R., & Pierdzioch, C. (2016). Does uncertainty move the gold price? New evidence from a nonparametric causality-in-quantiles test. *Resources Policy*, *49*, 74-80.

Belloni, A., Chernozhukov, V., Chetverikov, D., & Fernández-Val, I. (2011). *Conditional quantile processes based on series or many regressors* (No.arXiv: 1105.6154). Retrieved December, 10, 2019, from https://arxiv.org/pdf/1105.6154.pdf.

Bowman, A. W. (1984). An alternative method of cross-validation for the smoothing of density estimates. *Biometrika*, *71*(2), 353-360.

Bowman, A. W., & Azzalini, A. (1997). *Applied smoothing techniques for data analysis: The kernel approach with S-Plus illustrations*. Oxford: Oxford University Press.

Broock, W. A., Scheinkman, J. A., Dechert, W. D., & LeBaron, B. (1996). A test for independence based on the correlation dimension. *Econometric Reviews*, *15*(3), 197-235.

Camps-Valls, G., Rojo-Álvarez, J. L., & Martínez-Ramón, M. (2007). *Kernel methods in bioengineering, signal and image processing*. Hershey, PA: Idea Group Publishing.

Chang, C. L., McAleer, M., Wang, Y. A. (2018). *Latent volatility Granger causality and spillovers in renewable energy and crude oil ETFs* (Tinbergen Institute Discussion Paper No. 2018-052/III). Available at SSRN: https://ssrn.com/abstract=3184772.

Chaudhuri, P. (1991). Global nonparametric estimation of conditional quantile functions and their derivatives. *Journal of Multivariate Analysis*, *39*(2), 246-269.

Chernozhukov, V., Lee, S., & Rosen, A. M. (2013). Intersection bounds: Estimation and inference. *Econometrica*, *81*(2), 667-737.

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association, 74*(368), 829-836.

Cleveland, W. S., & Devlin, S. J. (1988). Locally weighted regression: An approach to regression analysis by local fitting. *Journal of the American Statistical Association, 83*(403), 596-610.

Cole, T. J. (1988). Fitting smoothing centile curves to reference data. *Journal of the Royal Statistical Society Series A, 151*(3), 385-418.

Daniell, P. J. (1946). Discussion on symposium on autocorrelation in time series. *Journal of the Royal Statistical Society*, *8*(1946), 88-90.

Davino, C., Furno, M., & Vistocco, D. (2014). *Quantile regression: Theory and applications.* John Wiley & Sons.

de Boor, C. (2001). *A practical guide to splines.* New York: Springer-Verlag.

Deng, H., & Wickham, H. (2011). *Density estimation in R.* http://vita.had.co.nz/papers/density-estimation.pdf. Retrieval Date: 03.11.2019.

Devroye, L. (1987). *A course in density estimation.* Boston: Birkhäuser.

Devroye, L., & Györfi, L. (1985). *Density estimation: The L1 view*. New York: John Wiley & Sons.

Devroye, L., & Lugosi, G. (2000). *Combinatorial methods in density estimation*. New York: Springer-Verlag.

Diks, C., & Panchenko, V. (2006). A new statistic and practical guidelines for nonparametric Granger causality testing. *Journal of Economic Dynamics and Control*, *30*(9-10), 1647-1669.

Dutta, A. (2018). Oil and energy sector stock markets: An analysis of implied volatility indexes. *Journal of Multinational Financial Management*, *44*, 61-68.

Einstein, A. (1914). Méthode pour la détermination de valeurs statistiques d'observations concernant des grandeurs soumises à des fluctuations irrégulières. *Archives des Sciences Physiques et Naturelles, 37,* 254–256.

Ercen, H. İ. (2015). *The interaction between renewable energy market and oil prices*. Master's Thesis, Near East University. Retrieved from http://docs.neu.edu.tr/library/6349714328.pdf.

Fisher, R. A. (1929). Tests of significance in harmonic analysis. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, *125*(796), 54-59.

Fitzmaurice, G. M., Laird, N. M., & Ware J. H. (2004). *Applied longitudinal analysis*. Hoboken: John Wiley & Sons.

Fix, B. (2019). Energy, hierarchy and the origin of inequality. *PLoS ONE, 14*(4): e0215692. https://doi.org/10.1371/journal.pone.0215692.

Fix, E., & Hodges, J. L. (1951). *Discriminatory analysis, nonparametric estimation: consistency properties* (Technical Report No. 4, Project No. 21-49-004). Randolph Field, Texas: USAF School of Aviation Medicine.

Gencay, R., Selcuk, F., & Whitcher, B. J. (2002). *An introduction to wavelets and other filtering methods in finance and economics*. San Diego, California: Academic Press.

Ghosh, S. (2018). *Kernel smoothing: Principles, methods and applications*. John Wiley & Sons.

Griggs, W. (2013). *Penalized spline regression and its applications.* Whitman College Report. https://www.whitman.edu/Documents/Academics/Mathematics/Griggs.pdf. Retrieval Date: 01.12.2019.

Hall, P. (1980). *Objective methods for the estimation of window size in the nonparametric estimation of a density*. Unpublished manuscript.

Hall, P., Racine, J. S., & Li, Q. (2004). Cross-validation and the estimation of conditional probability densities. *Journal of the American Statistical Association, 99*(468), 1015–1026.

Härdle, W. (1990). *Applied nonparametric regression*. Cambridge: Cambridge University Press.

Hastie, T. J., & Tibshirani, R. J. (1990). *Generalized additive models.* Monographs on Statistics and Applied Probability (Vol.43), New York: Chapman & Hall/CRC Press.

Heal, G. (2009). *The economics of renewable energy* (No. w15081). National Bureau of Economic Research (NBER) Working Paper Series, Cambridge.

Heinberg, R. & Friedly, D. (2016). *Our renewable future: Laying the path for one hundred percent clean energy.* Washington, DC: Island Press.

Hendricks, W., & Koenker, R. (1992). Hierarchical spline models for conditional quantiles and the demand for electricity. *Journal of the American Statistical Association*, *87*(417), 58-68.

Henriques, I., & Sadorsky, P. (2008). Oil prices and the stock prices of alternative energy companies. *Energy Economics*, *30*(3), 998-1010.

Hiemstra, C., & Jones, J. D. (1994). Testing for linear and nonlinear Granger causality in the stock price-volume relation. *The Journal of Finance*, *49*(5), 1639-1664.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning (with applications in R).* New York: Springer.

Jarque, C. M., & Bera, A. K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters*, *6*(3), 255-259.

Jeong, K., Härdle, W. K., & Song, S. (2012). A consistent nonparametric test for causality in quantile. *Econometric Theory*, *28*(4), 861-887.

Jones, M. C., & Sheather, S. J. (1991). Using non-stochastic terms to advantage in kernel-based estimation of integrated squared density derivatives. *Statistics & Probability Letters*, *11*(6), 511-514.

Kanamura, T. (2019). *A model of price correlations between clean energy indices and energy commodities*. Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3391003.

Kim, B. K., & Van Ryzin, J. (1985). A bivariate histogram density estimator: Consistency and asymptotic normality. *Statistics & Probability Letters*, *3*(3), 167-173.

Klemelä, J. S. (2009). *Smoothing of multivariate data: Density estimation and visualization* (Vol. 737). John Wiley & Sons, Inc.

Koenker, R. (2005). *Quantile Regression.* Cambridge: Cambridge University Press.

Koenker, R., & Bassett, G. (1978). Regression quantiles. *Econometrica: Journal of the Econometric Society*, *46*(1), 33-50.

Koenker, R., Ng, P., & Portnoy, S. (1994). Quantile smoothing splines. *Biometrika*, *81*(4), 673-680.

Koenker, R., Portnoy, S. & Ng, P. (1992). Nonparametric estimation of conditional quantile functions. In Y. Dodge (Ed.), *Statistical data analysis based on the L1 norm and related methods* (pp. 217-229). Elsevier Science.

Kumar, S., Managi, S., & Matsuda, A. (2012). Stock prices of clean energy firms, oil and carbon markets: A vector autoregressive analysis. *Energy Economics*, *34*(1), 215-226.

Li, Q., & Racine, J. (2004). Cross-validated local linear nonparametric regression. *Statistica Sinica*, *14*(2), 485-512.

Lipsitz, M., Belloni, A., Chernozhukov, V., & Fernández-Val, I. (2016). Quantreg.nonpar: An R package for performing nonparametric series quantile regression. *The R Journal, 8*(2), 370-381.

Mallows, C. L. (1973). Some comments on C_p. *Technometrics, 15*(4), 661-675.

Managi, S., & Okimoto, T. (2013). Does the price of oil interact with clean energy prices in the stock market?. *Japan and the World Economy*, *27*, 1-9.

McAleer, M., Chan, F., Hoti, S., & Lieberman, O. (2008). Generalized autoregressive conditional correlation. *Econometric Theory*, *24*(6), 1554-1583.

Milioti, R. (2017). *The impact of fossil fuel prices on alternative energy stocks.* Doctoral dissertation, Duke University Durham. Retrieved from: https://pdfs.semanticscholar.org/8803/40915661621a65f84c195d5dee4b92f6c413.pdf.

Nishiyama, Y., Hitomi, K., Kawasaki, Y., & Jeong, K. (2011). A consistent nonparametric test for nonlinear causality - specification in time series regression. *Journal of Econometrics*, *165*(1), 112-127.

Pagan, A., & Ullah, A. (1999). *Nonparametric econometrics*. New York, USA: Cambridge University Press.

Park, B. U., & Marron, J. S. (1990). Comparison of data-driven bandwidth selectors. *Journal of the American Statistical Association*, *85*(409), 66-72.

Parzen, E. (1962). On estimation of a probability density function and mode. *The Annals of Mathematical Statistics*, *33*(3), 1065-1076.

Parzen, M. I., Wei, L. J., & Ying, Z. (1994). A resampling method based on pivotal estimating functions. *Biometrika*, *81*(2), 341-350.

Powell, J. L. (1984). Least absolute deviations estimation for the censored regression model. *Journal of Econometrics, 25*(3), 303-325.

Racine, J. S. (2008). Nonparametric econometrics: A primer. *Foundations and Trends® in Econometrics*, *3*(1), 1-88.

Racine, J., & Li, Q. (2004). Nonparametric estimation of regression functions with both categorical and continuous data. *Journal of Econometrics*, *119*(1), 99-130.

Ramsay, J. O., Wickham, H., Graves, S., & Hooker, G. (2018). *Package'fda': Functional data analysis.* Retrieved January 12, 2020, from https://cran.r-project.org/web/packages/fda/fda.pdf.

Reboredo, J. C., Rivera-Castro, M. A., & Ugolini, A. (2017). Wavelet-based test of co-movement and causality between oil and renewable energy stock prices. *Energy Economics*, *61*, 241-252.

Reiss, P. T., & Huang, L. (2012). Smoothness selection for penalized quantile regression splines. *The International Journal of Biostatistics*, *8*(1), Article 10.

Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics, 27*(3), 832-837.

Rudemo, M. (1982). Empirical choice of histograms and kernel density estimators. *Scandinavian Journal of Statistics*, *9*(2), 65-78.

Sadorsky, P. (2012a). Correlations and volatility spillovers between oil prices and the stock prices of clean energy and technology companies. *Energy Economics*, *34*(1), 248-255.

Sadorsky, P. (2012b). Modeling renewable energy company risk. *Energy Policy*, *40*, 39-48.

Schuette, D. R. (1978). A linear programming approach to graduation. *Transactions of Society of Actuaries*, *30*, 407-445.

Scott, D. W. (1985). Averaged shifted histograms: Effective nonparametric density estimators in several dimensions. *The Annals of Statistics*, *13*(3), 1024-1040.

Scott, D. W. (1979). On optimal and data-based histograms. *Biometrika*, *66*(3), 605-610.

Scott, D. W. (1992). *Multivariate density estimation: Theory, practice, and visualization.* New York: John Wiley & Sons.

Sheather, S. (1983). A data-based algorithm for choosing the window width when estimating the density at a point. *Computational Statistics & Data Analysis*, *1*, 229-238.

Sheather, S. J. (1986). An improved data-based algorithm for choosing the window width when estimating the density at a point. *Computational Statistics & Data Analysis*, *4*(1), 61-65.

Sheather, S. J. (2004). Density estimation. *Statistical Science*, *19(*4), 588-597.

Sheather, S. (2009). *A modern approach to regression with R*. New York: Springer Science & Business Media.

Sheather, S. J., & Jones, M. C. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society: Series B (Methodological)*, *53*(3), 683-690.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. New York: Chapman & Hall / CRC Press.

Simonoff, J. S. (1996). *Smoothing methods in statistics*. New York: Springer.

Stone, C. J. (1977). Consistent nonparametric regression. *The Annals of Statistics*, *5*(4), 595-645.

Toda, H. Y., & Yamamoto, T. (1995). Statistical inference in vector autoregressions with possibly integrated processes. *Journal of Econometrics*, *66*(1-2), 225-250.

Tsay, R. S. (1987). Conditional heteroscedastic time series models. *Journal of the American Statistical Association*, *82*(398), 590-604.

Tu, C. H., & Li, C. (2017). A novel entropy-based approach to feature selection. In N. T. Nguyen, S. Tojo, L. M. Nguyen & B. Trawinski (Eds.), 9th *Asian Conference on Intelligent Information and Database Systems* (pp. 445-454). Cham, Switzerland: Springer.

Van Ryzin, J. (1973). A histogram method of density estimation. *Communications in Statistics-Theory and Methods*, *2*(6), 493-506.

Venables, W. N. & Ripley, B. D. (2002). *Modern applied statistics with S* (4th ed.). New York: Springer.

Wahba, G. (1990). *Spline models for observational data* (Vol. 59). Philadelphia: SIAM (Society for Industrial and Applied Mathematics).

Wand, M. P., & Jones, M. C. (1995). *Kernel smoothing*. London: Chapman & Hall.

Woodroofe, M. (1970). On choosing a delta-sequence. *The Annals of Mathematical Statistics*, *41*(5), 1665-1671.

Yaglom, A. M. (1987). Einstein's 1914 paper on the theory of irregularly fluctuating series of observations. *IEEE Acoustoustics Speech Signal Process Magazine, 4*(4), 7-11.

Yoshida, T. (2013). Asymptotics for penalized spline estimators in quantile regression. *Communications in Statistics - Theory and Methods*, DOI:10.1080/03610926.2013.765477. Available at: https://arxiv.org/pdf/1209.1156.pdf.

Yurinskii, V. (1977). On the error of the Gaussian approximation for convolutions. *Theory of Probability and Its Applications*, *22*(2), 236–247.

CHAPTER II

PREDICTING THE VOLATILITY OF BITCOIN RETURNS BASED ON KERNEL REGRESSION

2.1. Introduction

Volatility represents an inherently latent quantity presenting information about market uncertainty, therefore it is not directly observable and its interpretation should be grounded on the movements in market prices so that large fluctuations in prices imply a high volatility. However, it is hard to detect how high this volatility is depending on huge price shocks that cannot be made a distinction about if they are temporary or permanent. It can be expressed as one of the theoretical descriptions of the most prevalent risk measures (Danielsson, 2011, pp. 31, 73). In the more specific sense, according to financial economics literature, the frequently used description for volatility is "the (instantaneous) standard deviation (or "sigma") of the random Wiener-driven component in a continuous-time diffusion model" (Andersen, Bollerslev, Christoffersen, & Diebold, 2006, p. 780). In standard ARCH model and its extensions, the conditional variance is characterized by a function of lagged squared returns and thus, these lagged squared returns may be considered as approximate potential components leading to the volatility (Engle, 2004, p. 409). Application areas of volatility prediction take a comprehensive place in literature ranging from risk management towards portfolio allocation, options trading or quantile estimation. In financial economics, return volatility and therefore risk-return effects have always played a central role. However, volatility and risk are two concepts that have different meanings (Poon & Granger, 2003, p. 478) and the relationship between them can be explained indirectly through the leverage effect which has been mainly suggested by Black (1976). The leverage effect reveals the negative relationship between current asset returns and future volatility so that it refers to an unexpected increase in future volatility in response to an unexpected negative return leading to an increase in debt-to-equity ratio and therefore an increase in the risk situation (Bollerslev, Chou, & Kroner, 1992, p. 24; Schwert, 1990, p. 86). In addition to the contribution of Cox and Ross (1976) as well, this concept has been also widely discussed in Christie (1982); French, Schwert and Stambaugh (1987); Kupiec (1989); Schwert (1990) and many other papers. As expressed by Poon and Granger (2003), "the link between volatility and risk is tenuous;

in particular, risk is more often associated with small or negative returns, whereas most measures of dispersion make no such distinction". The vast majority of volatiliy models is comprised of the models with squared returns. However, discusssions about whether squared returns or absolute returns should be used as a proxy for actual volatility have a large place in the literature. Poon and Granger (2003) have noted that using squared returns as a proxy for volatility will generate a low value of determination coefficient and weaken the inference process with respect to forecast accuracy. Thus, some studies advocating the usage of absolute returns in terms of forecast accuracy can be exemplified as Taylor (1986); Ederington & Guan (2000) and McKenzie (1999) (Poon & Granger, 2003, pp. 480, 492).

Bitcoin -as the most familiar digital cryptocurrency with its intrinsic decentralized feature compared to its alternatives like BitShares, Ethereum, Litecoin, Ripple etc.- has maintained to be the largest cryptocurrency by total market capitalization value & the number of transactions per day since its creation in 2009 and got an important place in global financial markets depending on low transaction costs mostly & its more versatile structure as a virtual currency relative to the prices of standard currencies (Ciaian & Rajcaniova, 2018, pp. 1-2). In 2019, cryptocurrencies have outperformed other conventional asset classes such as commodities, bonds or equities and particularly, Bitcoin has been extremely volatile throughout its historical path. Measuring and predicting price fluctuations of Bitcoin have always been a center of interest by many researchers and the volatility of Bitcoin is driven by a great number of factors. As one of the virtual currencies, total supply of Bitcoins is limited with 21 million coins and debates related to Bitcoin have focused on whether it should be counted as a currency, security or a commodity. The issue of which asset class the cryptocurrencies will be considered as is very important. Because in order for the regulatory agencies of the states to make various regulations and taxations on the assets, they must first classify the assets that these assets belong to. By U.S. Commodity Futures Trading Commission (CFTC), Bitcoin and other digital currencies have been declared as commodities in September 2015. In general sense, commodity is defined as "the name given to all tradable goods such as gold, silver, copper, oil, natural gas, wheat, corn, cotton" and by definition, the acceptance of Bitcoin as a commodity may seem to be reasonable. The lethal coronavirus (COVID-19) pandemic starting in China is increasing day by day with more than 30,000,000 confirmed cases around the world. Since cash money is considered to pose a risk of carrying and therefore facilitating the spread of the virus, in such a conjuncture some people have a

tendency for using virtual currencies in a way to minimize the usage of cash money. Due to the coronavirus outbreak, a slowdown in production has brought along an economic slowdown and not only traditional stock markets but also cryptomarkets have experienced a remarkable collapse. Specifically, Bitcoin has been influenced drastically depending on the pandemic as well and the determinants of its highly volatile structure have been discussed to a large extent by many researchers.

In this paper, it has been concerned with predicting the volatility of Bitcoin returns using squared returns as proxy for volatility and the examination of volatility prediction has been based upon the study by Klemelä (2017). This paper has been organized as follows: Section 2.2 introduces some preliminary concepts regarding volatility, Section 2.3 discusses the theoretical framework, Section 2.4 presents the application results and the last section discusses an outline of core findings.

2.2. Literature Review

2.2.1. A Review on Researches Related to Nonparametric Volatility Prediction

There are many studies regarding nonparametric volatility prediction. One of the early studies can be exemplified as the study by Pagan and Schwert (1990) who have studied and compared the models with nonparametric kernel estimator & nonparametric flexible Fourier form estimator apart from the parametric ones in conditional stock return volatility estimation covering U.S. data for the period 1834-1925 and concluded that although extending the GARCH and exponential generalized autoregressive conditional heteroscedasticity (EGARCH) models with nonparametric estimators increases the explanatory power dramatically, results reveal that any extensions are best implemented in a parametric structure which is the case confirmed also by the out-of-sample prediction results reporting that the nonparametric models get worse compared to the parametric models. Abberger (1995) has investigated volatility by utilizing from nonparametric methods, taking marginal quantiles into account and comparing nonparametric quantile & GARCH estimations using Gaussian distributions on daily DAX returns. As a result, the kernel estimation of the conditional distribution has displayed weak consistency property and asymmetric effects have been revealed at different quantile levels depending on the kernel estimation. Heid (1996) has tried to estimate the conditional volatility of exchange rates and stock prices on daily basis utilizing from nonparametric regression approaches and examining raw, squared & absolute returns. In conclusion, estimated

volatility functions have been detected to be U-shaped with a minimum close to zero. Yang (2000) has suggested a direct local polynomial estimation technique for a finite nonparametric GARCH model in order to analyze foreign exchange volatility and revealed that the suggested model presents a satisfactory result for out-of-sample prediction. Audrino and Bühlmann (2001) have proposed a GARCH model of $\sigma_t^2 = f(R_{t-1}, \sigma_{t-1}^2)$ with tree-structured iterative estimation algorithms for estimating volatility in daily log-returns where $R_t = \sigma_t \varepsilon_t$ and $f : \mathbf{R}^2 \to \mathbf{R}$ denotes the unknown function parameterized by a threshold function. Hu (2011) has presented a comparison of two nonparametric GARCH volatility models -*Functional Gradient Descent (FGD) method* proposed by Audrino & Bühlmann (2009) and *Semiparametric Additive Autoregressive structure (aGARCH)* proposed by Wang, Feng, Song & Yang (2012) as both are based on B-spline smoothing method- with the classical GARCH (1,1) model by performing simulation studies. Based on the data set generated from standard GARCH(1,1), both nonparametric models have not been found to be superior to the simple GARCH(1,1) process. On the other hand, for the data set generated from a GJR model; aGARCH has shown the best performance amongst three models and for the data set generated from a more complicated nonparametric GARCH model, FGD and aGARCH have outperformed the standard GARCH(1,1) process. Ou and Wang (2012) have proposed a nonparametric financial volatility model based on the relevance vector machine algorithm and compared parametric GARCH-type models with this proposed model. In conclusion, the study has pointed out that the proposed model improves the performance of volatility forecasts relative to the parametric GARCH-type models for both simulated data based on ARMA-GARCH, ARMA-EGARCH & ARMA-GJR and real Dow Jones industrial average stock market data. Hou and Suardi (2012) have applied a nonparametric GARCH model for modelling and forecasting oil price return volatility for Brent and WTI crude oil markets covering 4845 daily observations and compared the nonparametric GARCH model with nine GARCH-class models studied by Wei, Wang & Huang (2010). They have documented that the nonparametric GARCH model outperforms the parametric GARCH models in terms of out-of-sample volatility forecasting accuracy. Klemelä (2017) has employed kernel regression predictor for volatility prediction and reported that kernel predictor outperforms GARCH(1,1) predictor and other related predictors. A recent study by Chikhi and Bendob (2018) for Orange stock returns has revealed that informational shocks create transitory effects on

volatility and nonparametric nonlinear ARCH model using local linear kernel methodology displays an improved performance in predictions.

Other contributions to nonparametric volatility estimation include Pagan and Hong (1991); Artis and Taylor (1994); Bossaerts, Hafner and Härdle (1996); Fan and Yao (1998); Niizeki (1998); Franke, Härdle and Kreiss (1998); Audrino and Barone-Adesi (2003); Randal, Thomson and Lally (2004); Zhang, Wang and Li (2009); Mattiussi (2010); Chung (2012); Cassim (2018); Dobrovidov and Tevosian (2018); Bandi and Renò (2018); Giordano and Parrella (2019).

2.2.2. A Review on Studies Related to Bitcoin

A great number of empirical studies have focused on the volatility of Bitcoin price movements and driving factors of this volatility. Balcilar, Bouri, Gupta & Roubaud (2017) have employed nonparametric causality-in-quantiles test for estimating the causal relation between trading volume and Bitcoin volatility and concluded that Bitcoin return volatility cannot be predicted through volume at any point of the conditional distribution. Jang and Lee (2017) have revealed that Bayesian neural network experiences a good performance in explaining the high volatile structure of the Bitcoin price. Guo and Antulov-Fantulin (2018) have dealt with predicting short-term price fluctuations of Bitcoin for one-step ahead by making a comparison of various approaches including time series models, ensemble methods and the temporal mixture model. They have utilized from the historical hourly volatility data of Bitcoin market and order-book data from the OKCoin from the period September 2015 to April 2017 including 13730 hourly volatility observations and 701892 order book snapshots of different time scales and data types. Results have shown that for most cases, the temporal mixture model of volatility -which is proposed for capturing the dynamic effect of order-book features- is superior to the others with respect to prediction performances. Conrad, Custovic & Ghysels (2018) have implemented the GARCH-MIDAS model in order to capture potential drivers of the long- and short-term volatility components of Bitcoin and found that long-term Bitcoin volatility is negatively affected by the S&P 500 realized volatility. Naimy & Hayek (2018) have compared GARCH(1,1), EWMA and EGARCH(1,1) models for forecasting Bitcoin volatility and concluded that EGARCH(1,1) model is superior to other two models with respect to forecast accuracy in both in-sample and out-of-sample contexts. Aalborg, Molnár & de Vries (2019) have reported that HAR-RV model proposed by Corsi (2009) is suitable in predicting Bitcoin volatility and lagged volatility variables can

account for the variation in realized volatility of Bitcoin on a large scale. Chaim & Laurini (2019) have estimated the volatility function of Bitcoin daily and high frequency prices for three subsamples based on the nonparametric estimator of Florens-Zmirou (1993) and Jiang & Knight (1997) and revealed the presence of a Bitcoin price bubble during the period January 2013-April 2014. Charles & Darné (2019) have suggested a replication research with the aim of verification for the study by Katsiampa (2017) who compares six GARCH-type models (GARCH, EGARCH, TGARCH, APARCH, CGARCH and ACGARCH) for estimating volatility of Bitcoin returns and reports the AR-CGARCH as the best model- utilizing from the same sample. In conclusion, partially different results have been reported such that Bitcoin returns have been characterized by the jumps and all GARCH-type models have been rejected implying all are inappropriate for modelling Bitcoin returns. Kim, Lee & Kang (2019) have examined the effects of Bitcoin futures on the intraday Bitcoin volatility using the discrete Fourier transform. Katsiampa (2019) has used a bivariate diagonal BEKK model for examining the volatility dynamics of Bitcoin. Caporale & Zekokh (2019) have tried to determine the best model combination for modelling volatility of Bitcoin, Ethereum, Ripple and Litecoin using Markov-Switching GARCH models concluding that in general sense mixture distribution models have outperformed Markov-Switching models and standard GARCH models may lead to inaccurate predictions for VaR and ES. Gyamerah (2019) has compared distinct nonparametric GARCH type models (standard (SGARCH), integrated (IGARCH) & threshold (TGARCH)) based on the Students-t distribution, Generalized Error Distribution (GED) & Normal Inverse Gaussian (NIG) distribution covering January 2014-August 2019 and concluded that the best performing model has been found to be tGARCH-NIG for estimating time-varying volatility of Bitcoin returns. In the paper by Khaldi, El Afia & Chiheb (2019); a comparative study on determining the best model to capture bitcoin volatility based on the leverage effect has revealed that nonparametric models -Multilayer Perceptron (MLP) and Standard Recurrent Neural Network (RNN)- are superior to parametric ones for Bitcoin volatility forecasting and amongst all models -be parametric or not- the best performing model is MLP that fails in forecast accuracy as the level of forecasting horizons increases. Then, it compares the best GARCH-type model against the best ANN model with respect to short-term and long-term horizons. Samırkas (2020) has investigated the volatility characteristics of Bitcoin for totally 2398 observations using GARCH(1,2), TGARCH(1,1,2) and EGARCH(1,1,1) models. In conclusion, the effect of shocks on volatility is permanent according to

parameter estimates and EGARCH has been found to be superior to other models in terms of prediction performance. Bouri, Gkillas, Gupta & Pierdzioch (2020) have investigated the effect of the US-China trade war in predicting daily realized volatility of Bitcoin returns by implementing heterogenous autoregressive realized volatility (HAR-RV) model for handling stylized facts and concluded that trade uncertainty enhances forecast accuracy for diverse forecast horizons. Other studies related to Bitcoin volatility include Pichl and Kaizoji (2017); Cermak (2017); Guo, Bifet and Antulov-Fantulin (2018); Acharya, Thomas and Pani (2018); Fang, Bouri, Gupta and Roubaud (2019); Malladi, Dheeriya and Martinez (2019); Senarathne (2019); Yu et al. (2019).

2.3. Theoretical Framework

2.3.1. Some Preliminary Concepts

According to the study by Klemelä (2017), what is meant by the concept "volatility prediction" is the prediction of a squared return denoted by $R_{t+\eta}^2$ where t represents the current time, $\eta \geq 1$ represents the prediction horizon and $R_i = P_i / P_{i-1} - 1$ represents the net return of an asset with prices P_i. In predicting squared returns, it is utilized from the observed past returns $R_1, ..., R_t$ (Klemelä, 2017, p. 1). On the other hand, in financial manner the volatility is exactly defined as the standard deviation of returns (Danielsson, 2011, p. 9).

2.3.1.1. Realized Volatility

Many practitioners also deal with predicting realized variance as a useful tool in volatility measurement which means the sum of η squared returns and can be calculated as follows:

$$RVar = \sum_{i=1}^{\eta} R_{t+i}^2 \tag{2.1}$$

(Klemelä, 2017, p. 3). In Klemelä (2017, p. 15), as an example, the realized volatility is expressed as follows:

$$RV(t, t+1) = \sum_{i=1}^{s} R_{t+(i-1)\Delta,\, t+i\Delta}^2 , \quad R_{t,m} = \log(P_m / P_t) \tag{2.2}$$

where Δ represents five minutes and s represents the number of five-minute periods during the day $[t,\ t+1]$. This realized volatility is practical in approximating the integral

$\int_t^{t+1} \sigma_u^2 du$ in the continuous time model given by $dS_t = \mu_t dt + \sigma_t dW_t$ where $S_t = \log P_t$, μ_t represents the drift term, σ_t represents the volatility process and W_t represents the Brownian motion. For longer horizons, the realized volatility is calculated as

$$RV(t, t+\eta) = \sum_{i=t}^{t+\eta-1} RV(i, i+1) \tag{2.3}$$

where $\eta \geq 1$ (Klemelä, 2017, pp. 15-16). In general, realized volatility is described as the square root of realized variance; but sometimes the term of variance is used in the same sense as volatility (Alexander, 2008, p. 334; Reno, 2007, p. 26). Statistical features associated with realized volatility have been discussed in Barndorff-Nielsen & Shephard (2002) and for an application on forecasting realized volatility in details; see Andersen, Bollerslev, Diebold & Labys (2003).

2.3.1.2. Simple EWMA (Exponentially Weighted Moving Average) Method

A procedure related to the Exponentially Weighted Moving Average (EWMA) control chart -which is also called a geometric moving average (MA) chart- was proposed by Roberts (1959) and can be represented by

$$E_t = (1-r)E_{t-1} + r\bar{X}_t \text{ , for } t > 0 \tag{2.4}$$

where E_t represents EWMA at time t for $t = 0, 1, 2, ..$ and $\bar{X}$ is a control chart which serves the purpose of gathering the information related to the process and thus monitoring the performance of the process with respect to enabling to reveal the presence or in some cases, the reasons of conditions that may necessitate a corrective action. In this procedure, the weight r is designated to the most recent observations ($r \in (0,1]$) and all other observations are appointed the fraction $(1-r)$ of the weight of their consecutives. The initial value E_0 can be considered as equal to μ_0 where μ_0 denotes the ordinate of the central line that belongs to control chart. The central line describes the level at which the process operates as free of problems. A fault detection in the process is realized when the point $\bar{X}_t$ referring to the average of measurements at time t exceeds lower and upper control limits. Equation (2.4) can also be expressed in a way that will display the geometric progression of weights designated to $\bar{X}$ s by taking μ_0 into account for $t \geq 0$:

$$(E_t - \mu_0) = r(1-r)^{t-1}(\bar{X}_1 - \mu_0) + r(1-r)^{t-2}(\bar{X}_2 - \mu_0) + \cdots + r(1-r)(\bar{X}_{t-1} - \mu_0) + r(\bar{X}_t - \mu_0) \tag{2.5}$$

It is apparent to see that in case all $\bar{X}$ s have an expected value which is equal to μ_0 , the expected value of E_t will also be μ_0 (Roberts, 1959, pp. 239-240). Through iteration method, Equation (2.4) can also be written as:

$$E_t = r\sum_{j=0}^{t-1}(1-r)^j \bar{X}_{t-j} + (1-r)^t E_0 \tag{2.6}$$

showing that it has been utilized from the weighted average of all past and present observations in the calculation of EWMA (Gupta & Guttman, 2013, p. 499).

EWMA has an advantageous feature such that it can show robustness property against the non-normality of the observations. Borror, Montgomery and Runger (1999) have revealed that robustness has been detected for $0.05 \le r \le 0.1$ (Fricker, 2013, p. 204). The value of the smoothing parameter of the moving average -which is an indicator of to which degree the forecast is influenced by the present observations- is usually chosen depending on how quickly the mean of the process differs so that when the mean changes faster (slower), the value of the parameter r should be chosen as large (small). Appropriate values for r are chosen through simulation methods in a common manner. Particularly, the sum of squared one-step-ahead forecast errors is calculated and the value of r is selected in a way to minimize this sum. For forecasting studies, r is selected between 0.05 and 0.30 in general (Perry, 2010, pp. 1-3). This interval specified for r has been supported in Abraham and Ledolter (1983, pp. 85-86) as well such that the smoothing parameter r has been described as $1-w$ where w denotes the discount coefficient that should be chosen between 0.7 and 0.95. Forecasting literature states in general that r defined by $1-w$ should be chosen as close to zero. As consistent with this idea, Brown (1962) expresses that w should be chosen between $\sqrt{0.70} = 0.84$ and $\sqrt{0.95} = 0.97$ (Abraham & Ledolter, 1983, p. 109). On the other hand, in Montgomery (2013, p. 436), it has been put forward that the smoothing parameters taking place in the range $0.05 \le r \le 0.25$ perform better in the practical sense and in general the value of $r = 0.20$ has a widespread usage.

2.3.2. Definition of the Kernel Predictor

In our application, we are dealing with the prediction of the response variable $R^2_{t+\eta}$ based on two predictor variables which are the square root of EWMA of past squared returns denoted by X_{1t} and EWMA of past returns denoted by X_{2t} (Klemelä, 2017, p. 1).

The Nadaraya-Watson kernel regression predictor which gives the weighted average of the past squared returns is expressed as:

$$\hat{f}(t) = \hat{f}_h(t,\eta) = \sum_{i=k}^{t-\eta} w_i(t) R_{i+\eta}^2 \tag{2.7}$$

where $1 < k < t-\eta$, $\eta \geq 1$ which denotes the prediction horizon,

$$w_i(t) = \frac{K_h(X_t - X_i)}{\sum_{j=k}^{t-\eta} K_h(X_t - X_j)} \tag{2.8}$$

with $w_i(t) \geq 1$ which represents the weights meeting the condition $\sum_{i=k}^{t-\eta} w_i(t) = 1$, $K : R^d \to R$ represents the kernel function defined by $K_h(x) = K(x/h)/h^d$ where h denotes the smoothing parameter and K has been selected as the density of standard normal distribution. Determining the smoothing parameter is likely through the minimization of

$$CV_t(h) = \sum_{i=t_0}^{t-\eta} \left(R_{i+\eta}^2 - \hat{f}_h(i,\eta) \right)^2 \tag{2.9}$$

which gives the sum of squared prediction errors (SSPE) and makes possible to obtain a different smoothing parameter h_t at each time t where $t_0 + \eta \leq t \leq T$ (Klemelä, 2017, p. 7). It is remarkable to say that a higher weight is appointed to the predictor values which are closer to the current time point. Although the EWMA of squared returns is regarded as a good proxy for volatility, it is more likely to consider the leverage effect by incorporating the EWMA of returns into the regression problem. One of the core matter prior to implementing the kernel regression estimation procedure lies at the suitable transformation of the predictive variables in a way that will display standard normal distributions, otherwise a spatially adaptive smoothing parameter is required to perform the kernel regression (Klemelä, 2017, p. 2). Thus, the explanatory variables $X_t = (X_{1t}, X_{2t})$ where marginal distributions can be considered as transformed form of the original observed data of the predictive variables $Z_t = (Z_{1t}, Z_{2t})$ for $t = 1,....,T$ observations are defined as follows:

$$Z_{t1} = \left(\sum_{i=1}^{t} \ell_i^1(t) R_i^2 \right)^{1/2}, \; Z_{t2} = \sum_{i=1}^{t} \ell_i^2(t) R_i \tag{2.10}$$

where the weight functions are given as

$$\ell_i^{\nu}(t) = \frac{G((t-i)/g_{\nu})}{\sum_{j=1}^{t} G((t-j)/g_{\nu})} \tag{2.11}$$

for $\nu = 1,2$; g_1 and $g_2 > 0$ represent the smoothing parameters and G represents the kernel function. In the application part, the kernel function G has been chosen based on EWMA. In the study by Klemelä (2017), the case in which EWMA of past squared returns is replaced by the GARCH(1,1) predictor of the volatility denoted by $\hat{\sigma}_{t+1}^2$ has also been handled and depending on this, Z_{t1} will be equivalent to $\hat{\sigma}_{t+1}$ (Klemelä, 2017, pp. 5-6).

Utilizing from the kernel regression predictor presents a clear picture about the effects of past squared and past returns on future volatility (Klemelä, 2017, p. 3). Apart from the kernel predictor mentioned here, for a detailed explanation about nonparametric volatility models & nonparametric volatility measurement -including ARCH filters and smoothers used for measuring the volatility over infinitesimally short horizons and realized volatility measures- see Bauwens, Hafner & Laurent (2012, pp. 269-291) and Andersen, Bollerslev & Diebold (2010, pp. 103-124).

2.3.3. GARCH Models

A general standard GARCH (p,q) process suggested by Bollerslev (1986) is expressed in the following way:

$$\sigma_t^2 = \beta_0 + \sum_{i=1}^{p} \beta_i R_{t-i}^2 + \sum_{i=1}^{q} \alpha_j \sigma_{t-j}^2 \tag{2.12}$$

where $R_t = \sigma_t \varepsilon_t$, $\varepsilon_t \sim$ i.i.d. (0,1). The conditional volatility represented by σ_t^2 is explained by the lagged squared terms R_{t-i}^2 and its own lagged values σ_{t-j}^2. Fat tails and volatility clustering can be explained by standard GARCH models; however, these models do not account for asymmetries based on the leverage effect (Berlinger et al., 2015, pp. 28, 30).

In a common manner, GARCH models are performed depending on logarithmic returns. This incentive associated with logarithmic returns is stemmed from the fact that in that case the price P_t takes positive value with probability one. However, the usage of net returns rather than logarithmic ones can be grounded on risk-management applications focusing on the estimation of upper quantiles of the loss distribution (namely, value-at-risk [VaR]) which is represented by

$$L_{t,m} = -(P_{t+m} - P_t) = -P_t(P_{t+m} / P_t - 1) = -P_t R_{t,m} \tag{2.13}$$

and therefore related to the volatility prediction where m-period net return is expressed by $R_{t,m} = P_{t+m} / P_t - 1$ for $m \geq 1$ (Klemelä, 2017, p .4).

2.3.3.1. GARCH(1,1) Model

The GARCH (1,1) model can be expressed in a way to consider the returns as:

$$\sigma_t^{\,2} = \beta_0 + \beta_1 R_{t-1}^2 + \alpha \sigma_{t-1}^2 \tag{2.14}$$

where $R_t = \sigma_t \varepsilon_t$, $\beta_0 > 0$, $\beta_1 \geq 0$, $\alpha \geq 0$ and $\{\varepsilon_t\}$ represents an i.i.d. (0,1) process. The conditional expectation of $R_{t+\eta}^2$ given past returns is defined as

$$E(R_{t+\eta}^2 \mid R_t, R_{t-1}, ...) = \bar{\sigma}^2 + (\beta_1 + \alpha)^{\eta-1}(\sigma_{t+1}^2 - \bar{\sigma}^2) \tag{2.15}$$

where

$$\bar{\sigma}^2 = \frac{\beta_0}{1 - \beta_1 - \alpha} \tag{2.16}$$

and

$$\sigma_{t+1}^2 = \frac{\beta_0}{1-\alpha} + \beta_1 \sum_{k=0}^{\infty} \alpha^k R_{t-k}^2 \tag{2.17}$$

giving the best prediction in the MSE sense. For the one-step prediction ($\eta = 1$), this GARCH(1,1) predictor is close to the EWMA. In the case of predicting squared returns for prediction horizons $\eta \geq 1$ using EWMA; as the prediction horizon η becomes larger, we have to select a larger value of the smoothing parameter g_ν in Equation (2.11) implying that the moving average will converge to the sample mean and therefore it is utilized from the latter in the case of a large prediction horizon (Klemelä, 2017, pp. 9-10).

One advantage of GARCH(1,1) process is that it enables to handle autoregression in leptokurtic distributions of asset returns and volatility clustering phenomenon which implies persistent shocks such that as stated by Mandelbrot (1963, p. 418), “large changes tend to be followed by large changes, of either sign, and small changes tend to be followed by small changes” (Cont, 2007, p. 290). This inclination for large or small price movements to cluster together could be perceived by intuition from the fact that a large positive (negative) shock in ε_t leads to an increase (decrease) in the value of R_t and thus an increase (decrease) in σ_{t+1} which in turn creates a larger (smaller) value for R_{t+1}. Apart

from GARCH models, volatility clustering can also be modelled by stochastic volatility models. However, since GARCH(1,1) has a symmetric structure, asymmetries in distributions and leverage effects cannot be handled by this process (Berlinger et al., 2015, p. 28; Cont, 2007, p. 290.).

2.3.3.2. Asymmetric GARCH Models & News Impact Curve

Pagan & Schwert (1990) and Engle & Ng (1993) have proposed the *news impact curve* investigating how conditional volatility changes in response to shocks (i.e. news affecting the market fluctuations) in different sizes and thus making it possible to define the asymmetric effects in volatility -as a crucial stylized fact of financial time series returns- clearly (Berlinger et al., 2015, p. 30). Some of recent studies regarding news impact curve can be exemplified as Caporin & Costola (2019); Othman, Alhabshi & Haron (2019) and Garcin & Goulet (2019). Some stylized facts related to volatility have been also discussed in Bollerslev, Engle & Nelson (1994). What is meant by the asymmetric effect is that predictable (conditional) volatility becomes larger in the case of an unexpected decrease in price (bad news) when compared to an unexpected increase in price (good news) of similar magnitude (Engle & Ng, 1993, p. 1752). In order to handle asymmetric effects of random shocks, Nelson (1991) has suggested an exponential GARCH (EGARCH) process. A simple demonstration of this process can be stated as

$$\log\sigma_t^2 = \beta_0 + \sum_{i=1}^{p}\left(\beta_i\varepsilon_{t-i} + \gamma(|\varepsilon_{t-i}| - E|\varepsilon_{t-i}|)\right) + \sum_{j=1}^{q}\alpha_j\log(\sigma_{t-j}^2) \qquad (2.18)$$

where E denotes the expectation operator. The parameter β_i allows for the asymmetry feature and a negative value of it implies that volatility gives more reaction against bad news (negative shocks) than good news (positive shocks). Therefore, contrary to the standard GARCH models, good and bad news create different effects on conditional volatility (Berlinger et al., 2015, p. 31; Engle & Ng, 1993, p. 1753) and in addition, EGARCH offers an advantage via modelling the logarithm of the conditional variance and thus removing the non-negativity constraints (Fan & Yao, 2017, p. 145).

There exist also a large number of asymmetric GARCH formulations -some are nested models- capturing the leverage effect in financial returns, for instance TGARCH model suggested by Zakoian (1994); GJR-model by Glosten, Jagannatthan and Runkle (1993); Absolute GARCH (AGARCH) model of Hentschel (1995) (Jondeau, Poon, & Rockinger, 2007, p. 94) or the Generalized Asymmetric Power ARCH (APGARCH)

model introduced by Ding, Granger and Engle (1993). Specification tests for asymmetric GARCH have been introduced in Hagerud (1997). Despite the advantages of EGARCH model, that the model includes a great number of nonlinear algorithms complicates the empirical prediction of the model. On the contrary, the GJR - GARCH model is much simpler (Wang, 2003, p. 38). Other alternative models with detailed content can be come across in the research of GARCH models by Bollerslev, Chou and Kroner (1992). Although some stylized features regarding financial time series can be identified better through the above alternative processes than the standard GARCH model, no findings have been encountered indicating that any alternative model is superior to the standard GARCH model in a regular manner (Wu, 2011, p. 4). Apart from the parametric processes given above, nonparametric GARCH models have been mentioned in Bühlmann & McNeil (1999) and Bühlmann & McNeil (2002).

As an asymmetric GARCH(1,1) model, Heston & Nandi (2000) have proposed a model for price options which considers the leverage effect and will be performed in the application part of this paper:

$$\begin{aligned}\sigma_t^2 &= \beta_0 + \beta_1(\varepsilon_{t-1} - \lambda\sigma_{t-1})^2 + \alpha\sigma_{t-1}^2 \\ &= \beta_0 + \frac{\beta_1(R_{t-1} - \lambda\sigma_{t-1}^2)^2}{\sigma_{t-1}^2} + \alpha\sigma_{t-1}^2\end{aligned} \tag{2.19}$$

where $\lambda \in \mathbf{R}$ denotes the skewness parameter. Apart from this model, the most prevalent GARCH models for taking the leverage effect into consideration have been proposed by Andersen et al. (2006) as asymmetric, threshold and exponential GARCH models. In the case of asymmetric GARCH (1,1), the conditional expectation of $R_{t+\eta}^2$ given past returns is

$$E(R_{t+\eta}^2 \mid R_t, R_{t-1}, ...) = \bar{\sigma}^2 + (\beta_1\lambda^2 + \alpha)^{\eta-1}(\sigma_{t+1}^2 - \bar{\sigma}^2) \tag{2.20}$$

where

$$\bar{\sigma}^2 = \frac{\beta_0 + \beta_1}{1 - \beta_1\lambda^2 - \alpha} \tag{2.21}$$

presenting the best prediction in the MSE sense. The formula for EWMA can be modified in a way to include squared returns as

$$\hat{f}(t) = (1-\delta)R_t^2 + \delta\hat{f}(t-1) \tag{2.22}$$

where δ is the smoothing parameter of the moving average (satisfying $0 \le \delta \le 1$) which can be determined through the cross-validation criterion $CV_t(\delta)$ in Equation (2.9) and the leverage effect is incorporated into the Equation (2.22) like that:

$$\hat{f}_{\delta}(t) = (1-\delta)\hat{f}_{\delta}(t-1)\left(\frac{R_t}{\sqrt{\hat{f}(t-1)}} - \lambda\right)^2 + \delta\hat{f}_{\delta}(t-1) \tag{2.23}$$

This formula is similar in structure to the GARCH model proposed by Engle & Ng (1993) and the smoothing parameter g in Equation (2.11) is equivalent to $g = -1/\log\delta$. When compared to the cross-validation, more stable predictors are generated using the aggregation of predictors depending on the upsurge of $CV_t(\delta)$ criterion during financial attacks. The aggregated predictor is represented by

$$\hat{f}(t) = \sum_{s=1}^{S} q_{t,s}\hat{f}_{\delta s}(t) \tag{2.24}$$

where $\delta_1, ..., \delta_S$ gives the finite collection of smoothing parameters, and

$$q_{t,s} = \frac{\exp\{-CV_t(\delta_s)\}}{\sum_{\nu=1}^{S}\exp\{-CV_t(\delta_\nu)\}} \tag{2.25}$$

(Klemelä, 2017, pp. 10-12).

Klemelä (2017, p. 3) states that asymmetric GARCH models have been formed in order to take the leverage effect into consideration; however, GARCH (1,1) model is superior to these models in handling this stylized fact. More detailed GARCH volatility models are available in Palm (1996).

2.3.4. Quantile Estimation Using Volatility Prediction

Quantiles hold an important place in risk management depending on the concept of value-at-risk (VaR) that is a market risk measurement technique as related to the loss distribution. Measuring performances of two quantile estimators (namely $\hat{q}_1$ and $\hat{q}_2$) is possible through calculating the difference of cumulative sums of losses expressed by

$$D_t = SL_t(\hat{q}_1) - SL_t(\hat{q}_2) \tag{2.26}$$

where

$$SL_t(\hat{q}) = \sum_{i=t_0}^{t-1}\rho_\tau(R_{i+1} - \hat{q}_{i+1}) \tag{2.27}$$

for τ-th conditional quantile $q_{t+1} = F_{t+1}^{-1}(\tau)$ under the restriction $0 < \tau < 1$ where $F_{t+1}(x) = P(R_{t+1} \le x \,|\, R_t, R_{t-1}, ...)$ is the conditional distribution function of the return series. In Equation (2.27), $\rho_\tau(x)$ states the loss function regarding quantile estimation which is equal to $x(\tau - 1)$ for $x < 0$ and $x\tau$ otherwise. Plotting D_t against time has a

practical meaning with respect to giving an insight about detecting all time periods where the first estimator outperforms the second estimator.

In quantile estimation framework using volatility prediction, two different situations can be handled. Firstly, define the return series as $R_t = \mu_t + \sigma_t \varepsilon_t$ where μ_t represents the conditional mean and σ_t represents the conditional standard deviation. In that case, the sample mean is used for estimating μ_t. The conditional quantile estimator as related to the predicted volatility $\hat{\sigma}_{t+1}$ can be expressed as follows:

$$\hat{q}_{t+1} = \hat{\mu}_{t+1} + \hat{\sigma}_{t+1}\hat{q}(\varepsilon) \tag{2.28}$$

where $\hat{\mu}_{t+1}$ represents the prediction of R_{t+1}, $\hat{\sigma}_{t+1}$ represents the estimator of the conditional standard deviation of the return series at time $t+1$ and $\hat{q}(\varepsilon)$ represents the estimator of the τ-th quantile of the distribution regarding $\varepsilon_t = (R_t - \mu_t)/\sigma_t$. First situation for quantile estimation is

$$\hat{q}(\varepsilon) = \sqrt{\frac{g-2}{g}} t_g^{-1}(\tau) \tag{2.29}$$

where t_g denotes the distribution function regarding t distribution with g degrees of freedom being greater than 2. In the second situation, quantile estimation is done through the τ-th empirical quantile of the residuals given by

$$\hat{\varepsilon}_1 = (R_1 - \hat{\mu}_1)/\hat{\sigma}_1, \ldots\ldots, \hat{\varepsilon}_t = (R_t - \hat{\mu}_t)/\hat{\sigma}_t \tag{2.30}$$

as ordered from the smallest to the largest one and the empirical quantile is represented by

$$\hat{q}(\varepsilon) = \hat{\varepsilon}_{(\lceil t\tau \rceil)} \tag{2.31}$$

where $\lceil x \rceil$ gives the smallest integer value being greater than or equal to x (Klemelä, 2017, pp. 26-28).

2.4. Data Set and Application

In this paper, it has been tried to measure the performance of Bitcoin returns through daily historical Bitcoin-dollar closing prices data covering the period September 17th 2014 – March 13rd 2020, totally 2005 observations. The data in question have been extracted out of the web page of Yahoo (https://finance.yahoo.com/) with ticker symbol "BTC-USD" and the Bitcoin (USD) price chart for the given period has been displayed in Figure 13.

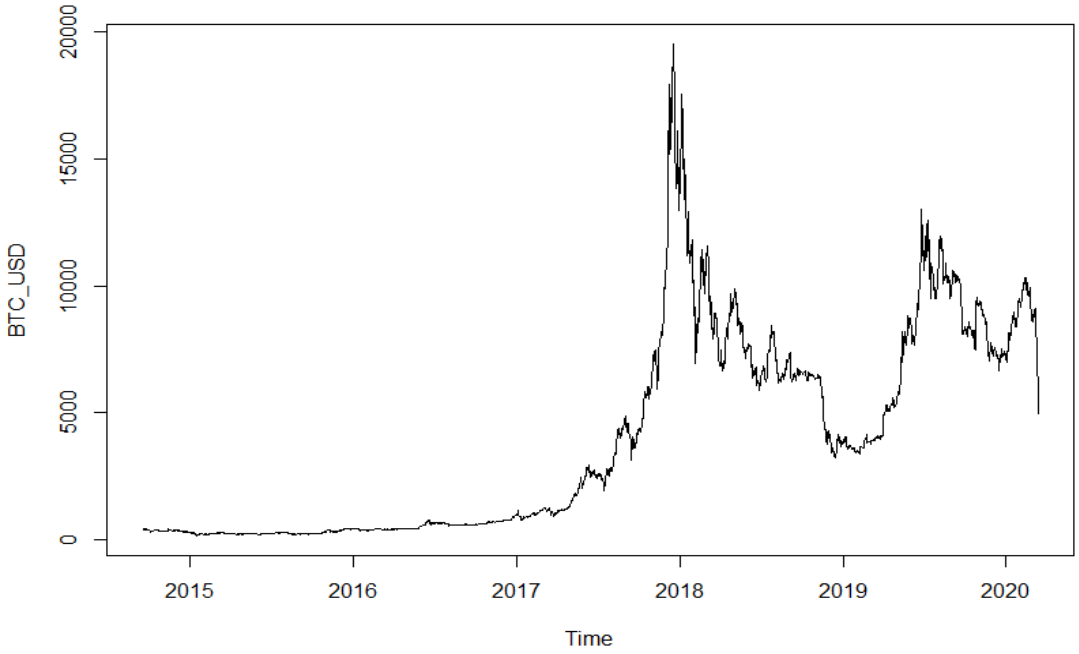

Figure 13. Bitcoin daily price index

Predictor variables of this paper have been derived from Bitcoin daily net returns which are represented in Figure 14.

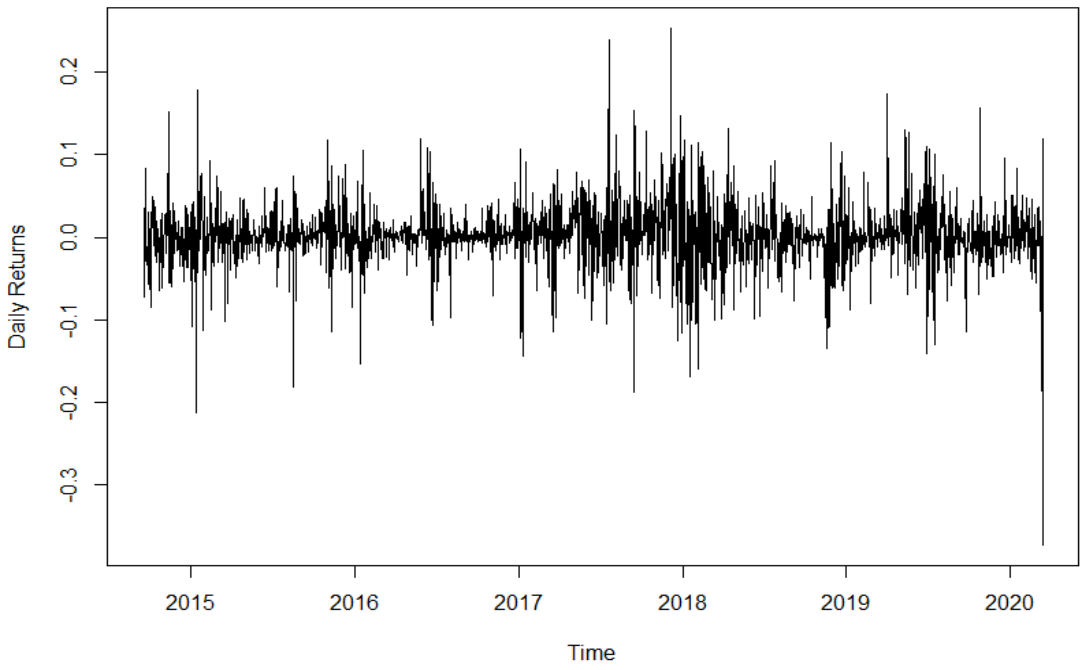

Figure 14. Bitcoin daily net returns

Firstly, the relative performances of kernel regression predictor and GARCH(1,1) predictor have been benchmarked in estimating the conditional quantiles using the

differences of the cumulative sums of squared prediction errors (CSSPE) which are advantageous in preventing the selection of the testing period, therefore preventing dominating events like financial shocks to have an influence on analysis results and facilitating the comparison of predictors over all time periods without requiring to exclude outliers from the analysis. The formula for the differences of cumulative sums of squared prediction errors to make a comparison between two predictors, namely $\hat{f}_1$ and $\hat{f}_2$ is given as follows:

$$D_t = SSPE_t(\hat{f}_1) - SSPE_t(\hat{f}_2) \tag{2.32}$$

or in other expression,

$$D_t - D_u = \sum_{i=u-\eta+1}^{t-\eta} \left(R_{i+\eta}^2 - \hat{f}_1(i) \right)^2 - \sum_{i=u-\eta+1}^{t-\eta} \left(R_{i+\eta}^2 - \hat{f}_2(i) \right)^2 \tag{2.33}$$

where $t > u$. Equation (2.33) gives an insight about the time periods in which the one predictor exceeds another predictor in performance. The case of $D_t - D_u < 0$ indicates that $\hat{f}_1$ outperforms $\hat{f}_2$ for the time period $[u,t]$ where $t > u$ while the case of $D_t - D_u > 0$ indicates the very opposite (Klemelä, 2017, pp. 2, 8-9). In this application, two predictor variables which are the square root of EWMA of past squared returns denoted by X_{1t} and EWMA of past returns denoted by X_{2t} have been used. The smoothing parameters have been obtained using cross-validation and normal reference rule. The smoothing parameter to be used in Equation (2.8) based on the normal reference rule can be expressed as

$$h_t = \left(\frac{4}{d+2} \right)^{1/(d+4)} t^{-1/(d+4)} \tag{2.34}$$

where $d = 2$ giving the smaller values of the smoothing parameter when compared to the cross-validation (Klemelä, 2017, p. 8).

Smoothing parameters for the kernel function G based on EWMA, denoted by g_ν in Equation (2.11), have been designated as $g_1 = g_2 = 40$.

Figure 15 displays the scatter plot of both predictor variables used in kernel regression. Panel (a) shows predictor variables without transformation defined by

$Z_t = (Z_{1t}, Z_{2t})$ in Equation (2.10) and Panel (b) shows transformed predictor variables $X_t = (X_{1t}, X_{2t})$ whose marginals represent approximately standard normal distributions.

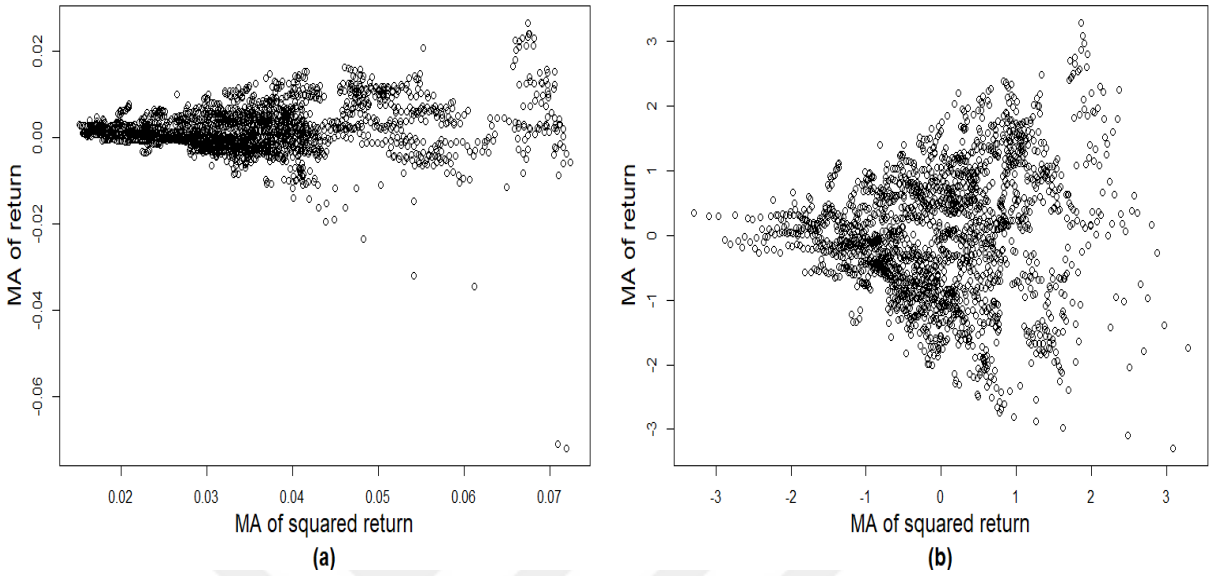

Figure 15. Scatter plots of explanatory variables in (a) original form (b) transformed form as standard normal

In an obvious manner, it can be said that we obtained spatially more homogeneous dispersion of data in the transition from Panel (a) to Panel (b).

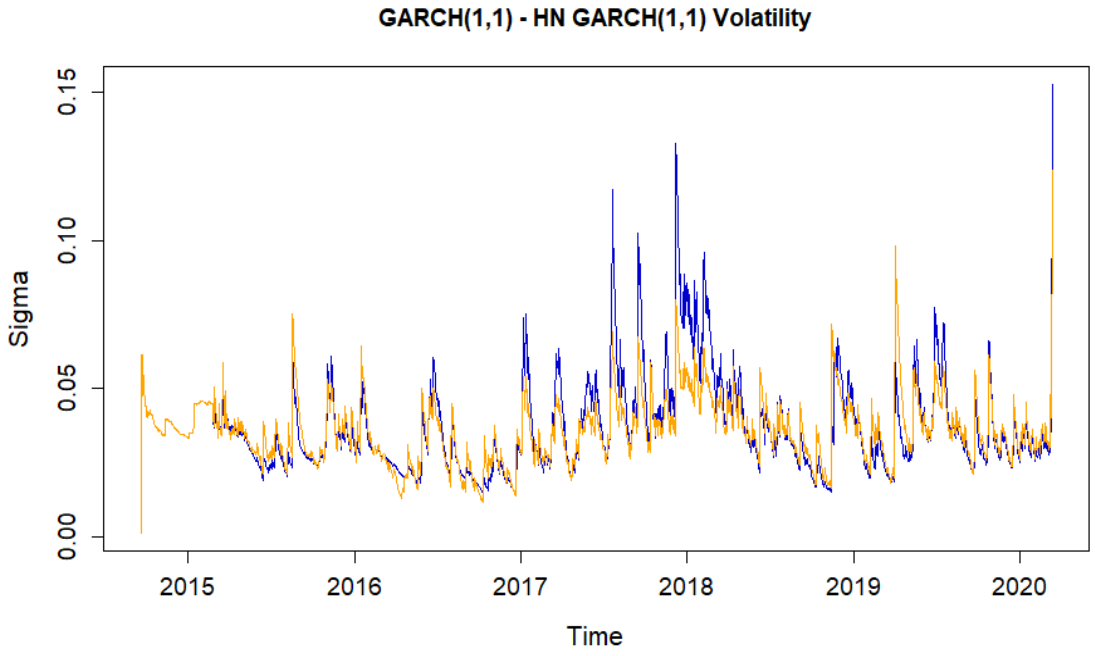

Figure 16. A comparison of GARCH(1,1) & Heston-Nandi (HN) GARCH(1,1) models

Figure 16 presents a comparison of volatilities from the GARCH(1,1) and HN-GARCH(1,1) models displayed by blue and orange lines respectively. Roughly speaking, HN-GARCH(1,1) model reveals a somewhat lower volatility relative to GARCH(1,1) model even though two volatilities seem to follow similar patterns.

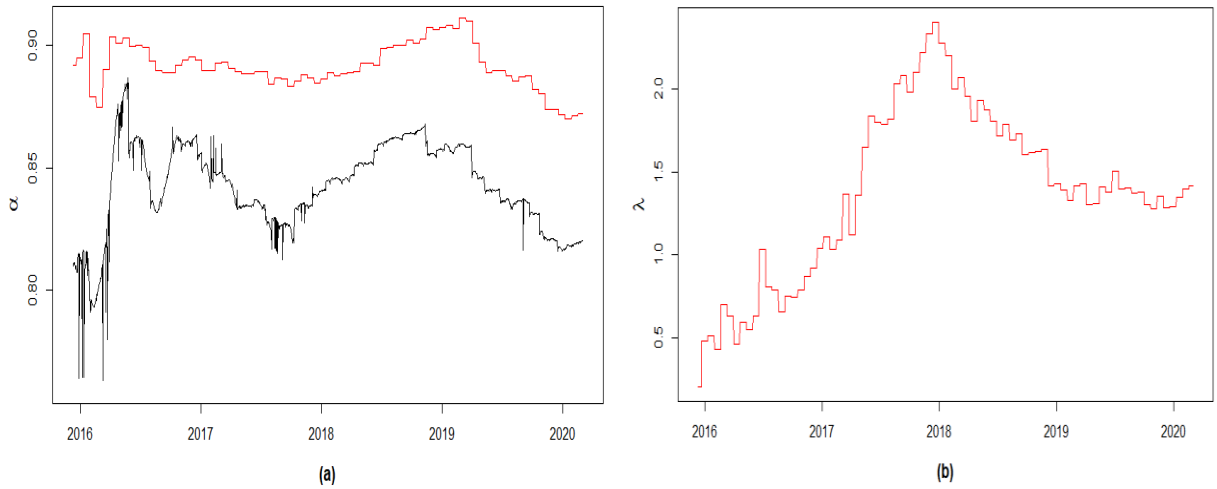

Figure 17. GARCH and asymmetric GARCH parameter estimates

Figure 17 depicts the time series regarding GARCH parameter estimates: The sequential estimates of α included in Equations (2.14) and (2.19) for standard GARCH(1,1) model (black) and Heston-Nandi model (red) have been given in Panel (a); λ estimates of HN-GARCH(1,1) in Equation (2.19) have been shown in Panel (b).

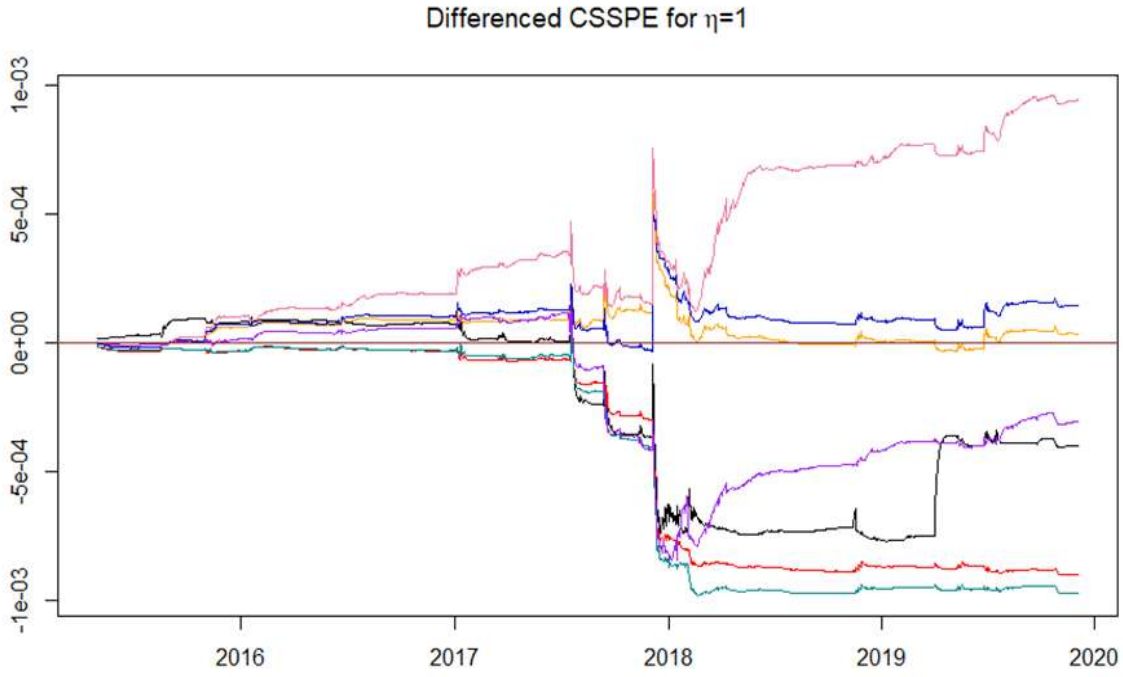

Figure 18. Comparison of (asymmetric) GARCH and (asymmetric) EWMA predictors based on differenced CSSPE for prediction horizon $\eta = 1$

Performances of different predictors based on the differences of the cumulative sums of prediction errors $SSPE_t(\hat{f}) - SSPE_t(\hat{f}_{garch})$ for prediction horizon $\eta = 1$ have been displayed in Figure 18. Analyses have been implemented for $t_0 = 160$ trading days. For differenced CSSPE series described by $SSPE_t(\hat{f}) - SSPE_t(\hat{f}_{garch})$, $\hat{f}_{garch}$ denotes the standard GARCH(1,1) predictor and in Figure 18, the predictor $\hat{f}$ represents the Heston-Nandi (HN) GARCH(1,1) (black); the cross-validated EWMA with $\lambda = 0$ (orange), $\lambda = 0.2$ (blue), $\lambda = 0.5$ (pink); the aggregated EWMA with $\lambda = 0$ (red), $\lambda = 0.2$ (green), $\lambda = 0.5$ (purple). For EWMA with cross-validation and aggregation, it has been utilized from the grid $\delta = \exp\{-1/g\}$ (δ and g are smoothing parameters of the MA and kernel function expressed in Equation (2.22) and Equation (2.11) where $g = 1,5,10,20,30,50$ in the case of prediction horizon $\eta = 1$. For cross-validated and aggregated EWMA, δ has been calculated as 0,980. Figure 18 reveals that the sums of squared prediction errors are dominated by the end of 2017. This finding is undoubtedly likely to be realized depending on Bitcoin's record price climb by the end of the year 2017 which experiences the levels being close to $20000 in December -implying a price bubble as some researchers say- and in turn, experiences eventual drops called the 2018 cryptocurrency crash or the announcement on the closure of cryptocurrency local trading platforms made by China on 15th October, 2017. Figure 18 shows that GARCH(1,1) predictor has recorded a good performance during all time periods. However, from mid-2017; Heston-Nandi and aggregated EWMAs with $\lambda = 0$ (red), $\lambda = 0.2$ (green), $\lambda = 0.5$ (purple) have revealed a better performance than traditional GARCH(1,1) predictor. This situation shows itself prominently since 2018. Starting from the end of 2015, the worst performance among all estimators in all time periods belongs to asymmetric EWMA with cross-validation with $\lambda = 0.5$ (pink). If we evaluate the MAs in itself, aggregation of MAs seems to outperform the cross-validation. Aggregated EWMA with $\lambda = 0.5$ performs worse than aggregated EWMAs with $\lambda = 0$ & $\lambda = 0.2$ and HN-GARCH(1,1) (excluding some time points); however, is better than the cross-validated EWMA in any circumstances. If we evaluate aggregated EWMAs within themselves, predictors can be said to perform better when the λ value is below 0.5. Furthermore, in predicting squared returns, the best performance among all predictors in 2018 and afterwards obviously belongs to the aggregated EWMA with $\lambda = 0.2$ (green).

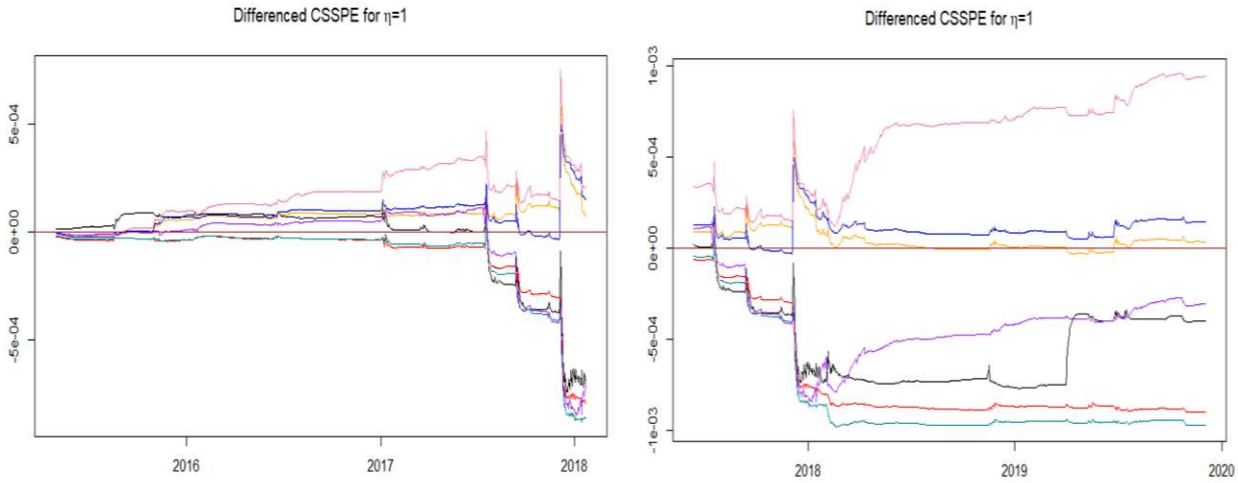

Figure 19. Comparison of (asymmetric) GARCH and (asymmetric) EWMA predictors for two sub-samples with $\eta = 1$

With respect to providing a more clear vision, time-series displayed in Figure 18 have been divided into two parts in Figure 19 where the left panel shows the time series until January 2018 and the right panel shows the time series from June 2017.

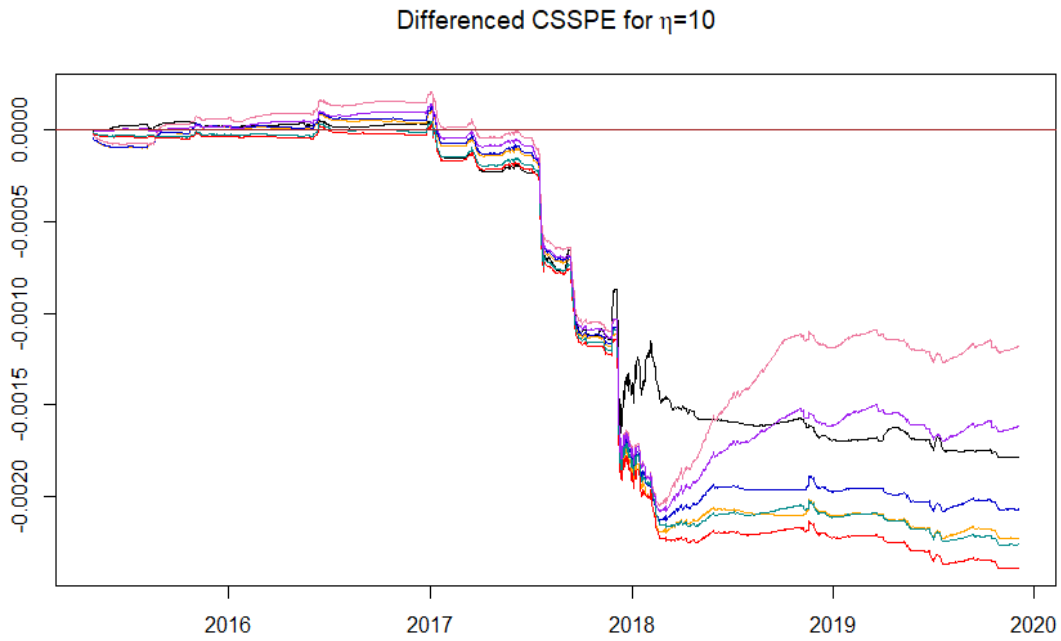

Figure 20. Comparison of (asymmetric) GARCH and (asymmetric) EWMA predictors based on differenced CSSPE for prediction horizon $\eta = 10$

In this paper, EWMA with cross-validation and EWMA with aggregation have been performed using the grid $\delta = \exp\{-1/g\}$ where

$g = 10, 20, 40, 60, 80, 100, 120, 150, 200, 300$ in the case of prediction horizon $\eta = 10$. As a result, δ has been computed as 0.997 and λ values have been chosen as 0, 0.2 and 0.5.

Figure 20 shows that based on the differenced CSSPE series represented by $SSPE_t(\hat{f}) - SSPE_t(\hat{f}_{garch})$, the predictor $\hat{f}$ indicates the Heston-Nandi GARCH(1,1) (black); the cross-validated EWMA with $\lambda = 0$ (orange), $\lambda = 0.2$ (blue), $\lambda = 0.5$ (pink); the aggregated EWMA with $\lambda = 0$ (red), $\lambda = 0.2$ (green), $\lambda = 0.5$ (purple). For some times up to 2017, differenced CSSPE is very slightly greater than 0 implying that GARCH(1,1) predictor performs a little bit better than others. It can also be said that the performance of GARCH(1,1) has been deteriorated in the case of $\eta = 10$ when compared to $\eta = 1$. In the general sense, all predictors except GARCH(1,1) have improved the prediction and if necessary to state, especially since mid-2017, the best predictor performance distinctly belongs to the asymmetric MA with aggregation (red). Also, EWMAs with cross-validation and aggregation for $\lambda = 0$ and $\lambda = 0.2$ have performed better than Heston-Nandi model and EWMA with aggregation & cross-validation for $\lambda = 0.5$. Towards the first half of 2018, while MA predictors are overall better than Heston-Nandi GARCH(1,1); towards 2020, it can be said that Heston-Nandi GARCH(1,1) performs better than EWMAs with aggregation and cross-validation for $\lambda = 0.5$. In brief, all MA and HN predictors have improved the predictability when compared to standard GARCH(1,1) as prediction horizon becomes larger.

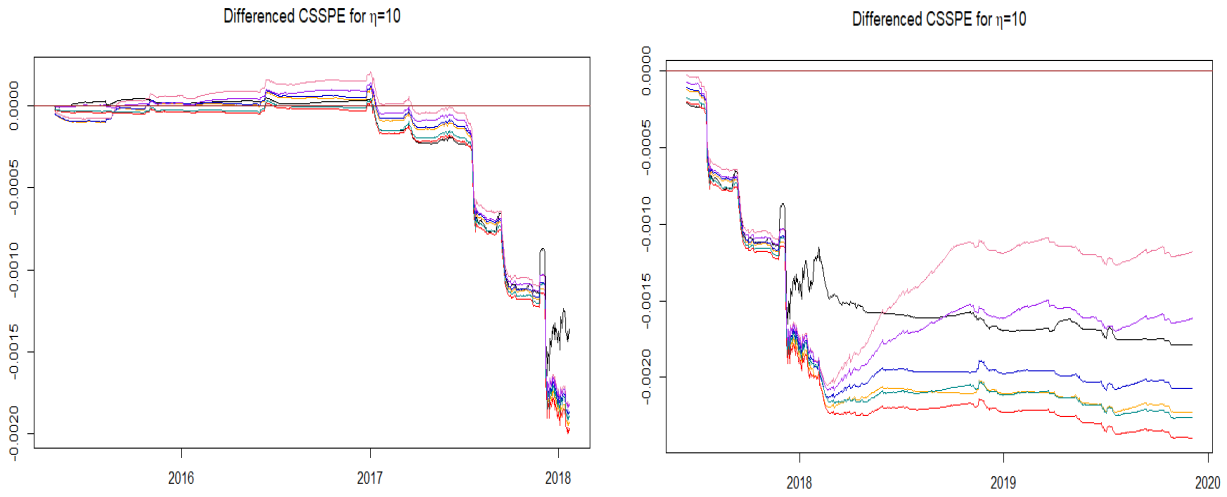

Figure 21. Comparison of (asymmetric) GARCH and (asymmetric) EWMA predictors for two sub-samples with $\eta = 10$

Figure 21 presents the time series given in Figure 20 as seperated into two parts in terms of providing a more clear picture. The left panel displays the time series until January 2018 while the right panel displays the time series starting from June 2017.

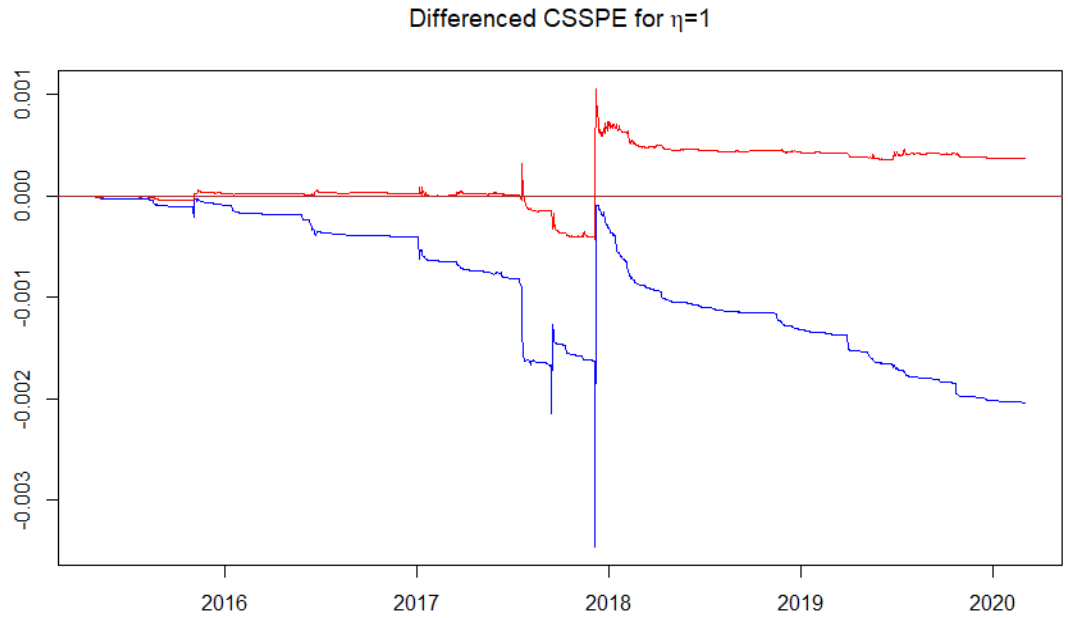

Figure 22. Comparison of GARCH and kernel regression based on differenced CSSPE for prediction horizon $\eta = 1$

In this paper, kernel regression has been handled in two different forms: in the first, the MA of squared returns has been used as a predictor variable as in Equation (2.10) and in the second, GARCH (1,1) volatility has been used as a predictor variable which can be stated as $Z_{t1} = \hat{\sigma}_{t+1}$ (Klemelä, 2017, pp. 6, 17).

Figure 22 presents a comparison of GARCH and kernel regression for the prediction horizon of one day. Based on the differenced CSSPE time series expressed by $SSPE_t(\hat{f}) - SSPE_t(\hat{f}_{garch})$, $\hat{f}$ indicates kernel regression predictor with EWMA volatility as the first predictor variable with smoothing parameter $g_1 = 20$ (represented by red time series) and in the second case, it represents kernel regression predictor with GARCH(1,1) volatility as the first predictor variable (represented by blue time series). The smoothing parameter for the EWMA of the returns has been chosen as $g_2 = 10$. Prediction has been performed using the grid where $g = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7$. Figure 22 reveals that kernel regression predictor with GARCH(1,1) volatility (blue) performs better than the standard GARCH(1,1) predictor during all time periods and kernel predictor with EWMA volatility (red) prevails the standard GARCH(1,1) towards

the end of 2017. It can be inferred from Figure 22 that GARCH(1,1) volatility is superior to EWMA as the kernel regression predictor.

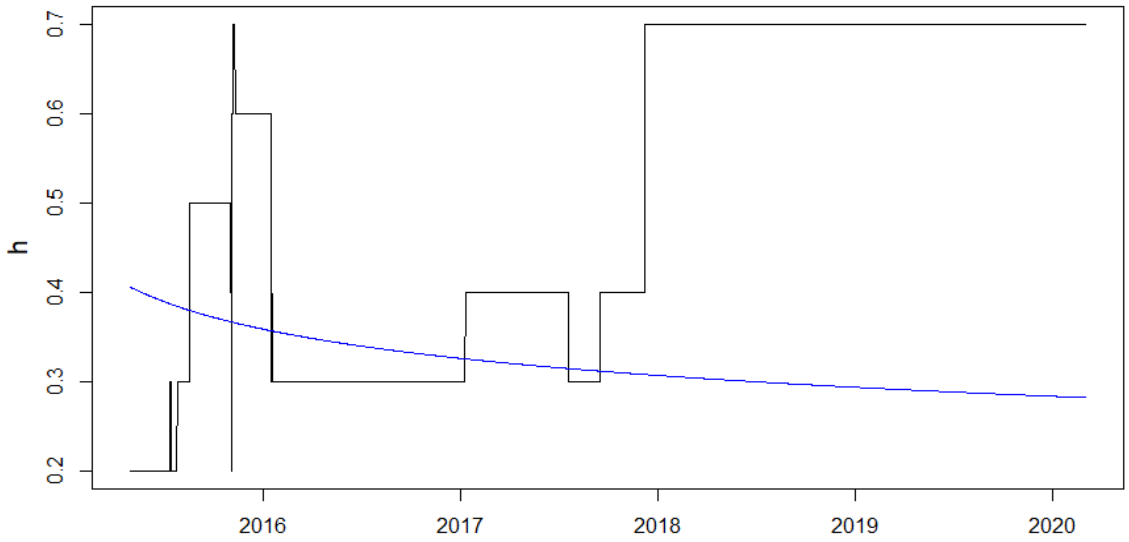

Figure 23. Cross-validation (black) and normal-reference rule (blue) smoothing parameters for $\eta = 1$

Figure 23 depicts the time series regarding the cross-validation smoothing parameters for the kernel regression predictor with standard GARCH(1,1) volatility as the first predictor variable (black) and the time series regarding the normal-reference rule smoothing parameter (blue) expressed in Equation (2.34) (choosing $d = 2$) for the prediction horizon of one day. As seen, smoothing parameter chosen based on the normal-reference rule takes smaller values relative to the smoothing parameter chosen based on the cross-validation in the general sense.

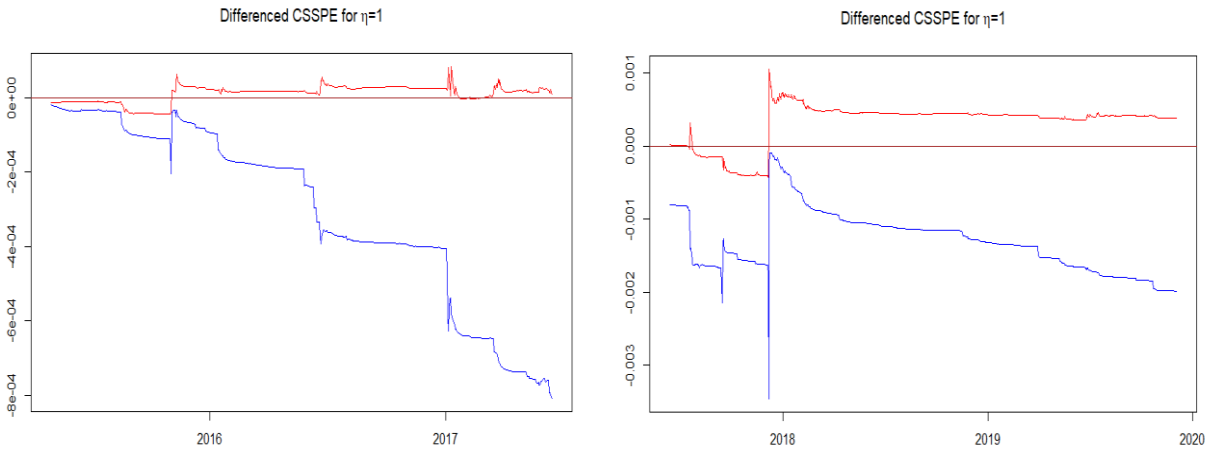

Figure 24. Comparison of kernel regression predictors for two sub-samples with $\eta = 1$

Figure 24 shows the time series of Figure 22 in a detailed view where the left panel depicts the time series until June 2017 and the right panel depicts the time series after June 2017 having a jump in the winter of 2017. Roughly speaking, it can be said that standard GARCH(1,1) predictor performs slightly better than kernel predictor with EWMA volatility (red) for the period from the end of 2015 until June 2017 on the left panel and for the period from the end of 2017 towards 2020 on the right panel.

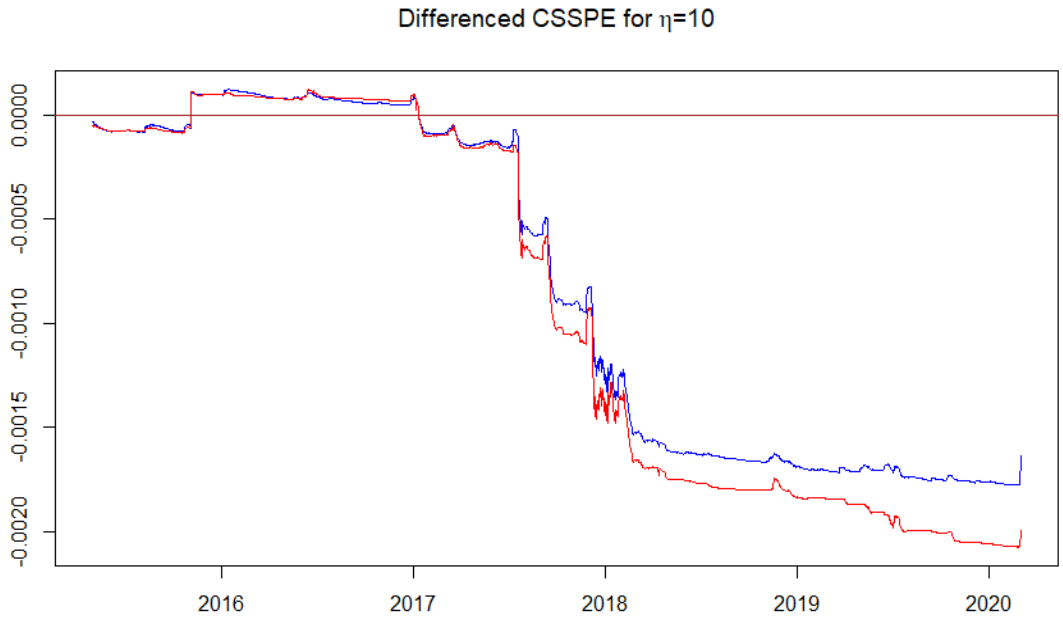

Figure 25. Comparison of GARCH and kernel regression based on differenced CSSPE for prediction horizon $\eta = 10$

Figure 25 presents a comparison of GARCH and kernel regression for the prediction horizon of ten-days. Based on $SSPE_t(\hat{f}) - SSPE_t(\hat{f}_{garch})$, $\hat{f}$ indicates kernel regression predictor with EWMA volatility as the first predictor variable with smoothing parameter $g_1 = 20$ (red line) and in the second situation, it indicates kernel regression predictor with GARCH(1,1) volatility as the first predictor variable (blue line). The smoothing parameter for the EWMA of the returns has been chosen as $g_2 = 20$ as well. Kernel prediction has been done using the grid where $g = 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9$. It can be inferred from Figure 25 that as seen from the end of 2015 until 2017, standard GARCH predictor performs better relative to others; however, after 2017 kernel predictors perform better than GARCH(1,1) predictor and during this period -contrary to Figure 22 which depicts the findings for $\eta = 1$- among kernel predictors; EWMA volatility (red) shows a better performance than GARCH volatility (blue) for the prediction horizon of ten days.

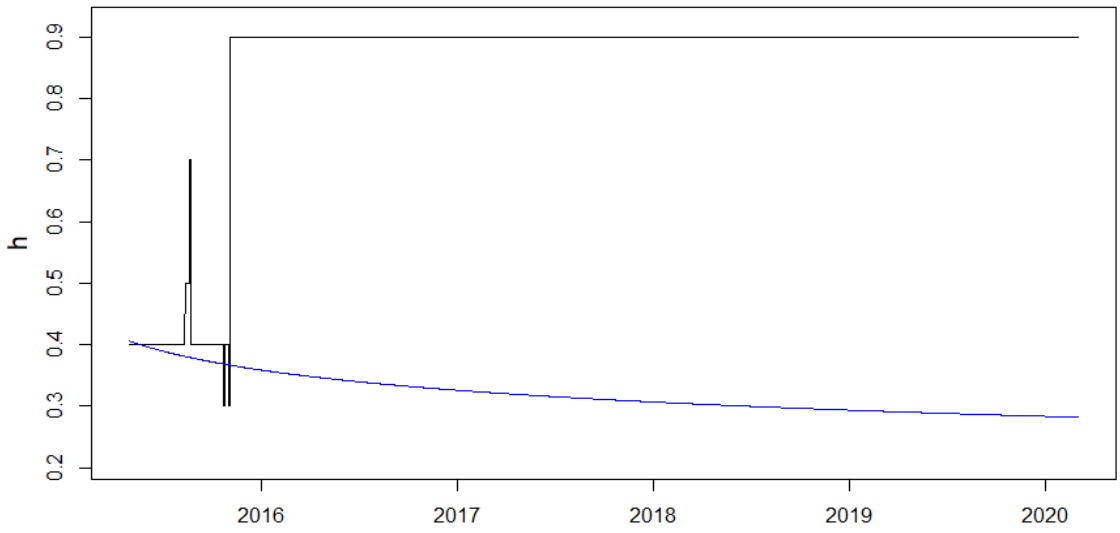

Figure 26. Cross-validation (black) and normal-reference rule (blue) smoothing parameters for $\eta = 10$

Figure 26 displays the time series regarding the cross-validation (black) and normal reference rule (blue) smoothing parameters with the same logic as in Figure 23 for the prediction horizon of ten days and normal reference rule is seen to be superior to cross-validation method in terms of generating smaller smoothing parameters.

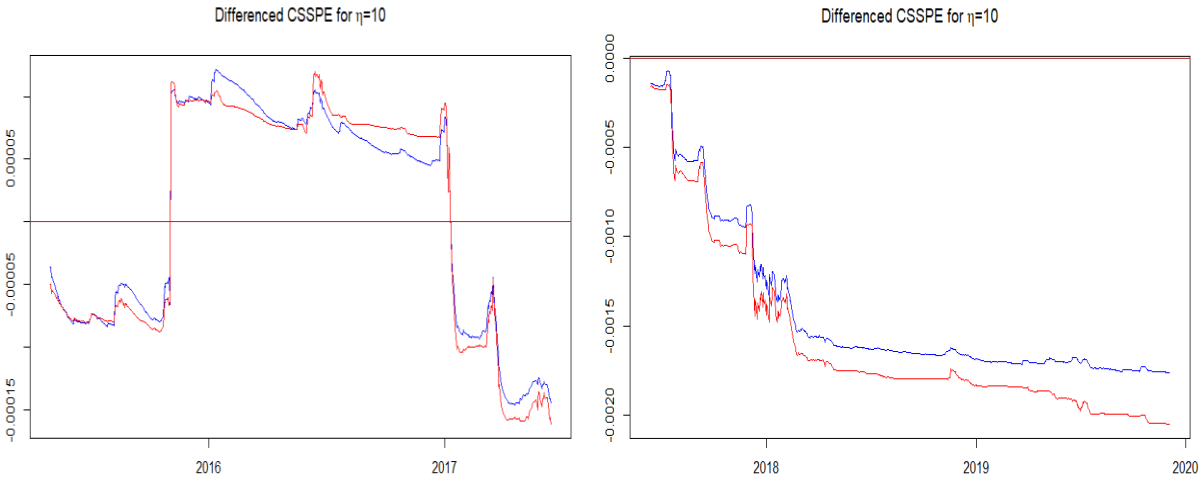

Figure 27. Comparison of kernel regression predictors for two sub-samples with $\eta = 10$

Figure 27 displays the time series of Figure 25 in a zoomed frame where the left panel shows the time series until June 2017 and the right panel shows the time series after June 2017.

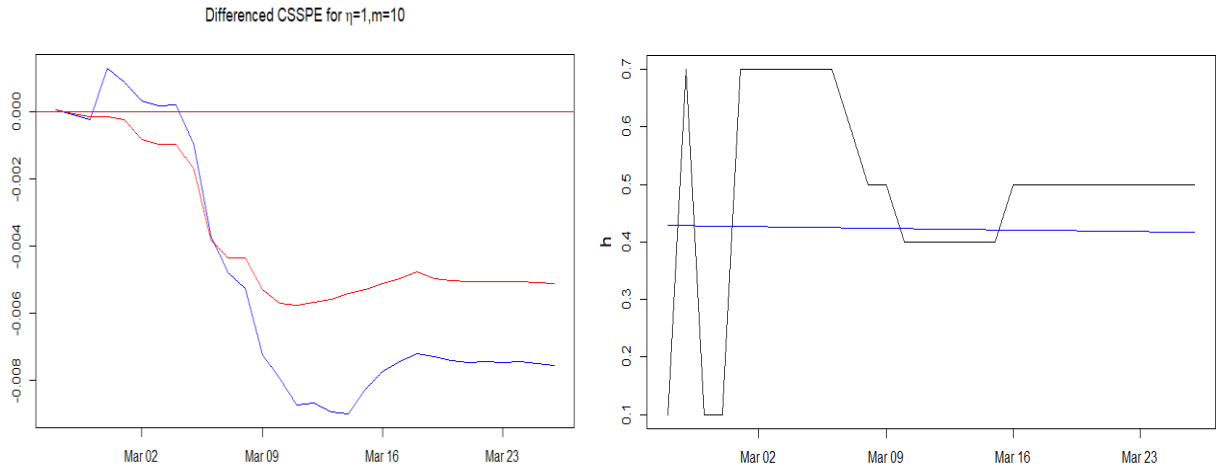

Figure 28. Comparison of kernel regression predictors for return horizon $m = 10$ and prediction horizon $\eta = 1$

Figure 28 shows a comparison between GARCH (1,1) and kernel regression predictors for return horizon of ten days and prediction horizon of one day. The left panel in Figure 28 shows the time series $SSPE_t(\hat{f}) - SSPE_t(\hat{f}_{garch})$ where $\hat{f}$ indicates the kernel regression predictor with EWMA volatility based on past squared returns with smoothing parameter $g_1 = 20$ (red line) and GARCH(1,1) volatility (blue line). The smoothing parameter for the EWMA of the returns has been chosen as $g_2 = 10$. Kernel prediction has been carried out using the grid where $g = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7$ and it can be inferred from the left panel that having negative values of differenced CSSPEs, kernel regression predictor outperforms mostly the standard GARCH(1,1) predictor. But, if necessary to compare two kernel predictors, GARCH(1,1) volatility can be said to be superior to EWMA in general. The right panel in Figure 28 depicts the time series of chosen smoothing parameters h based on the cross-validation (black) and the normal-reference rule (blue) for $m = 10$ and $\eta = 1$.

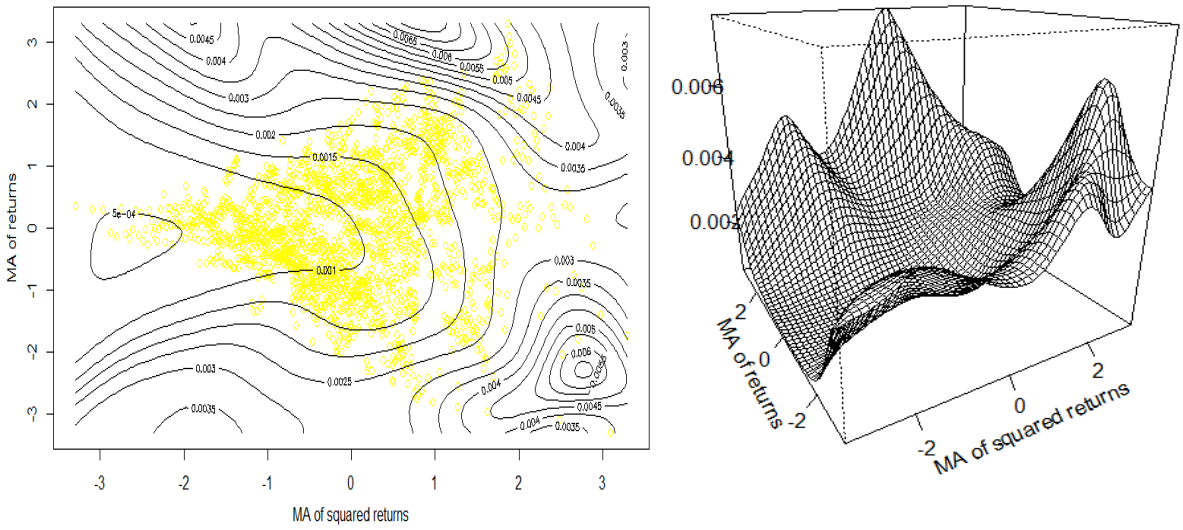

Figure 29. Contour and perspective plots related to estimates of kernel regression

Figure 29 displays plots of kernel regression function estimates covering the whole sample for prediction horizon of one day ($\eta = 1$). Smoothing parameter h of the kernel regression predictor has been taken as 0.5 and smoothing parameters of both MA of squared returns & MA of original returns -denoted by g_ν - have been appointed as $g_1 = g_2 = 40$. The left panel in Figure 29 shows a contour plot while the right panel shows a perspective plot. It can be inferred from the perspective plot in Figure 29 that there has been found no evidence regarding leverage effect mechanism in the general sense, but it works when predicted volatility becomes larger as MA of squared returns increases and MA of original returns decreases only from zero towards approximately the value between -2 and -3.

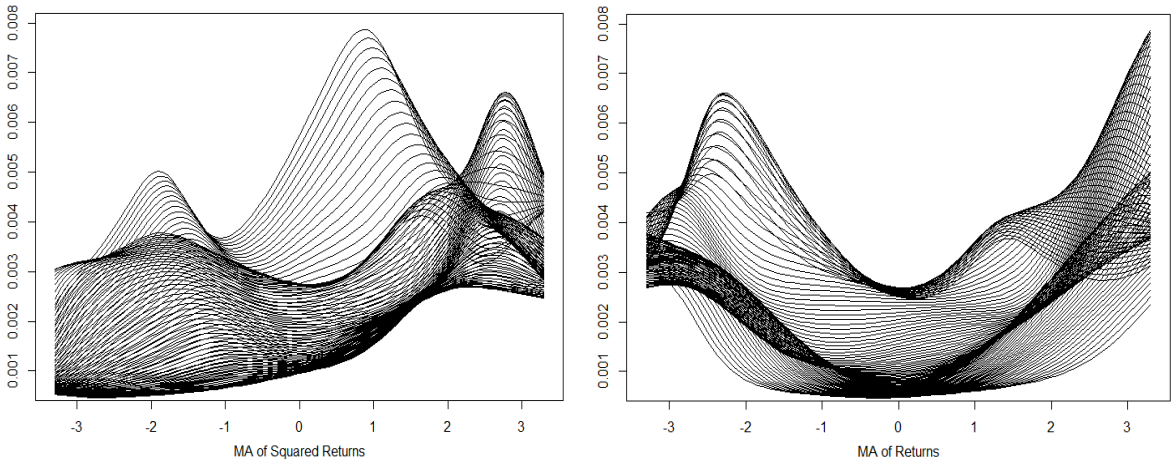

Figure 30. Slices of kernel prediction

A slice can be used as a counterpart of the news impact curve represented by $x_2 \to f(x_1, x_2)$ where x_2 denotes the moving average of past returns keeping x_1 constant (Klemelä, 2017, p. 23). Figure 30 displays slices belonging to the estimated kernel regression function $\hat{f}(x_1, x_2)$ of Figure 29. The left panel shows slices $x_1 \mapsto \hat{f}(x_1, x_2)$ for several values of x_2 while the right panel shows slices $x_2 \mapsto \hat{f}(x_1, x_2)$ for several values of x_1 where x_1 indicates the MA of squared returns and x_2 indicates the MA of original returns.

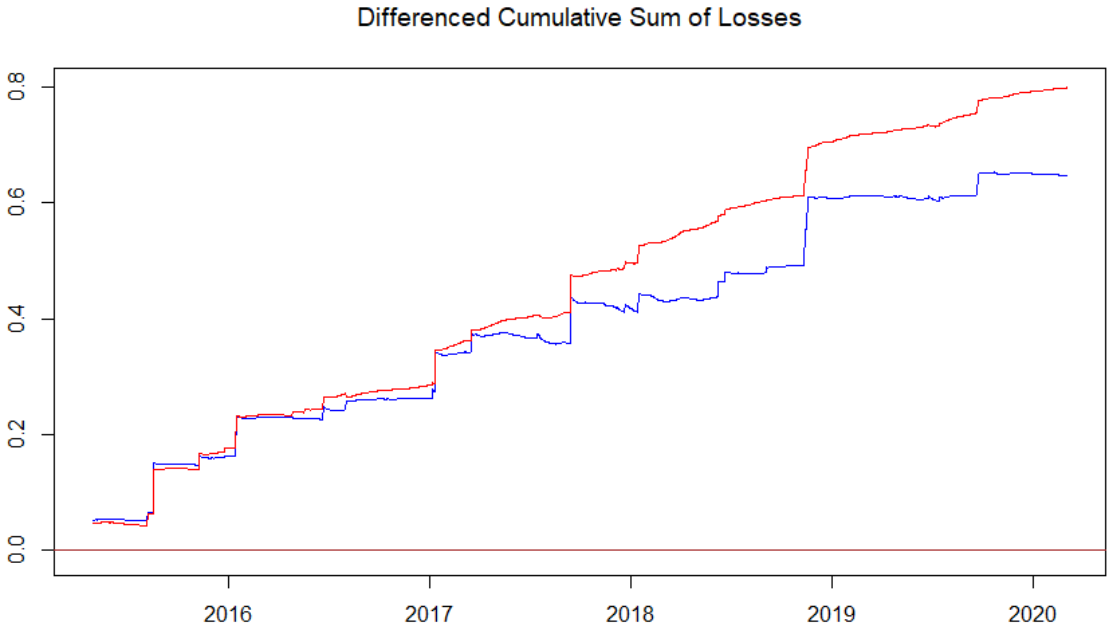

Figure 31. Quantile estimation of one-day returns for $\tau = 0.01$ and $\eta = 1$

Figure 31 considers the quantile estimation of one-day returns for the quantile level $\tau = 0.01$ and the given time series represents the differenced cumulative sum of losses expressed by $D_t = SL_t(\hat{q}) - SL_t(\hat{q}_{garch})$, where $\hat{q}$ represents the conditional quantile estimator with the kernel regression predictor and $\hat{q}_{garch}$ represents the conditional quantile estimator with the GARCH(1,1) predictor. D_t series obtained for the empirical quantiles and student residuals are represented by the red and blue curves respectively. Smoothing parameters for the MA of squared returns and the MA of original returns have been chosen as $g_1 = 20$ and $g_2 = 10$ respectively. Kernel prediction has been done by applying the grid where $g = 0.1,\ 0.2,\ 0.3,\ 0.4,\ 0.5,\ 0.6,\ 0.7$. Time series $D_t = SL_t(\hat{q}) - SL_t(\hat{q}_{garch})$ calculated for both student residuals and empirical quantiles take

place above the brown horizontal reference line at zero point implying that for the prediction horizon of one day, standard GARCH(1,1) predictor outperforms the kernel predictor for all time periods considered and generally speaking, student residuals perform slightly better than the empirical quantiles for one-day returns.

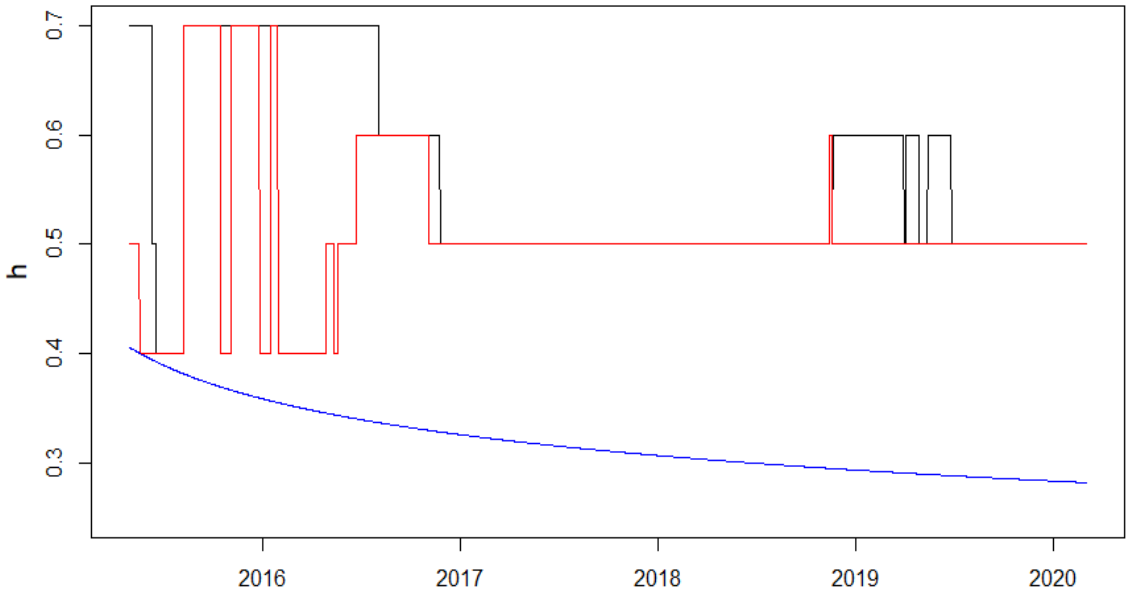

Figure 32. Cross-validation bandwidths for student (black) & empirical (red) quantiles and normal-reference rule (blue) bandwidths for $\eta = 1$

Figure 32 shows the plot of cross-validation smoothing parameters for student residuals (black), empirical quantiles (red) and the smoothing parameters corresponding to the normal reference rule (blue) depending on the quantile estimation results given in Figure 31. It is apparent to say that normal reference rule provides the smallest smoothing parameters when compared to student and empirical quantiles during all time periods.

Figure 33 considers the quantile estimation of one-day returns for the quantile level $\tau = 0.05$. In conformity with Figure 31 which considers the quantile estimation for $\tau = 0.01$, standard GARCH(1,1) predictor has been found to be superior to the kernel predictor again for all time periods for the prediction horizon $\eta = 1$. However, the only difference is that the empirical quantiles (red) beat the student residuals -with having smaller differenced cumulative sum of losses relatively- in the case of $\tau = 0.05$ which is contrary to the case in $\tau = 0.01$

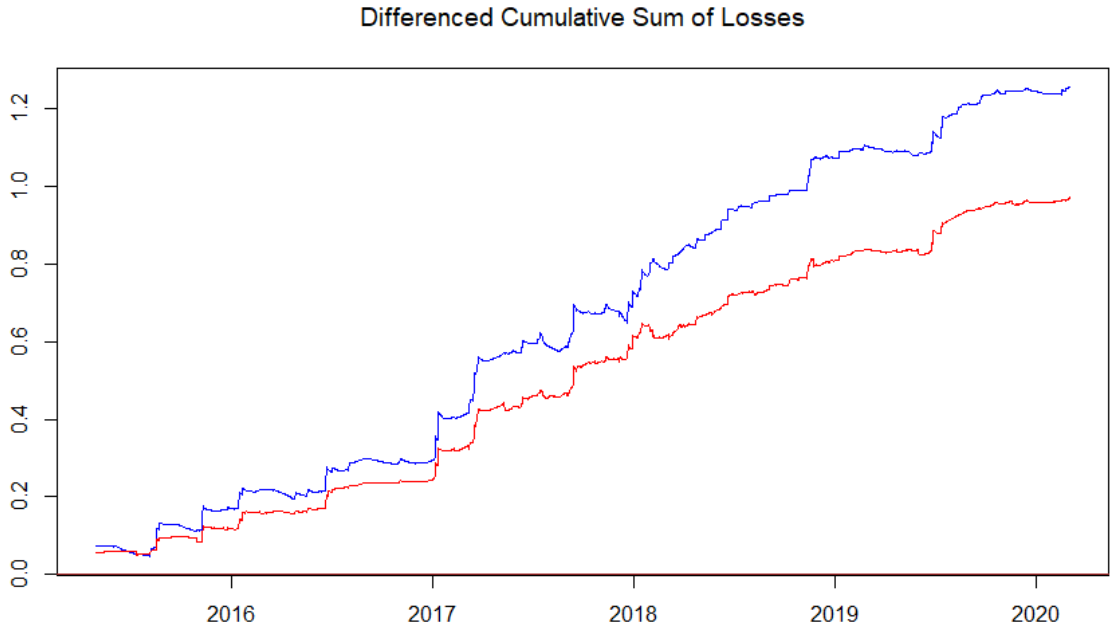

Figure 33. Quantile estimation of one-day returns for $\tau = 0.05$ and $\eta = 1$

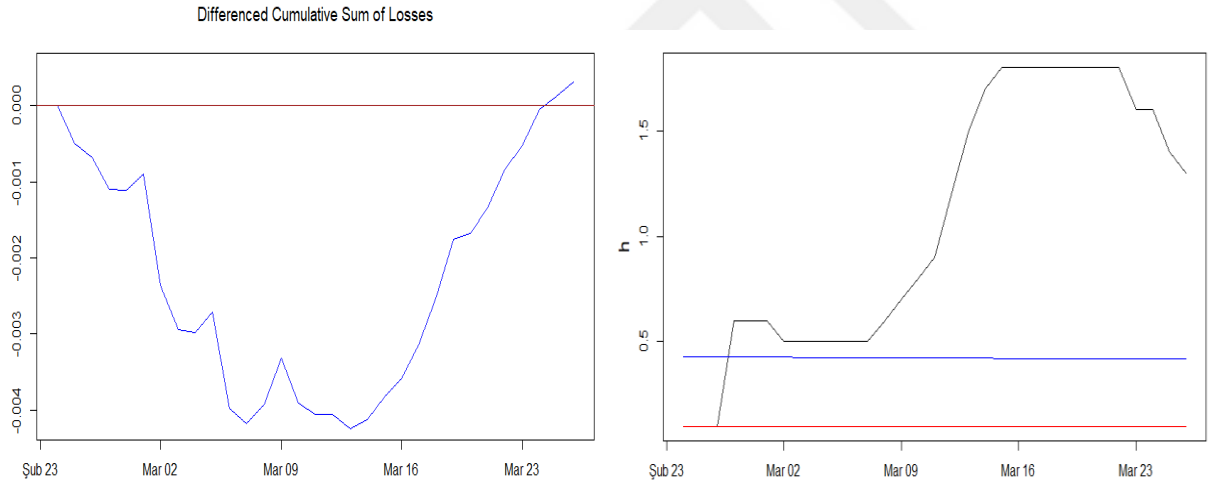

Figure 34. Quantile estimation of ten-day returns for $\tau = 0.01$ and $\eta = 1$

Figure 34 considers the plot regarding the quantile estimation of ten-day returns for $\tau = 0.01$ and the prediction horizon of one day ($\eta = 1$). Smoothing parameters for the MA of squared returns and the MA of original returns have been taken as $g_1 = 20$ and $g_2 = 10$ respectively. Kernel prediction has been done using the grid where $g = \text{seq}(0.1,\ 1.8)$ shows a sequence going from 0.1 towards 1.8 with the increments by 0.1. The left panel shows that for the quantile level $\tau = 0.01$, the time series $D_t = SL_t(\hat{q}) - SL_t(\hat{q}_{garch})$

calculated for student residuals (blue) reveals the superiority of the kernel predictor against GARCH(1,1) predictor during the time periods considered contrary to the case with one-day returns given in Figure 31. However, differenced cumulative sums of losses pertaining to the empirical quantiles (red) have not been plotted due to having infinite as its maximum starting from zero point, therefore implying that the conditional quantile estimator with GARCH(1,1) predictor beats the kernel predictor using empirical quantiles. The right panel in Figure 34 shows the plot of cross-validation smoothing parameters for student residuals (black), empirical quantiles (red) and the smoothing parameters based on the normal reference rule (blue) connected with the quantile estimation of ten-day returns with the prediction horizon of one-day. Estimations based on empirical quantiles can be said to generate smoothing parameters with the smallest values among others. One important note here can be attributed to the blue line showing the smoothing parameters based on normal reference such that it has been drawn as if it is a horizontal (constant) line. Such a situation has been stemmed from y-axis values that the predictors will take lying in a narrow range when all predictors are considered all together. If we plot only the time series regarding the normal reference rule bandwidth, it will seem like in Figure 35:

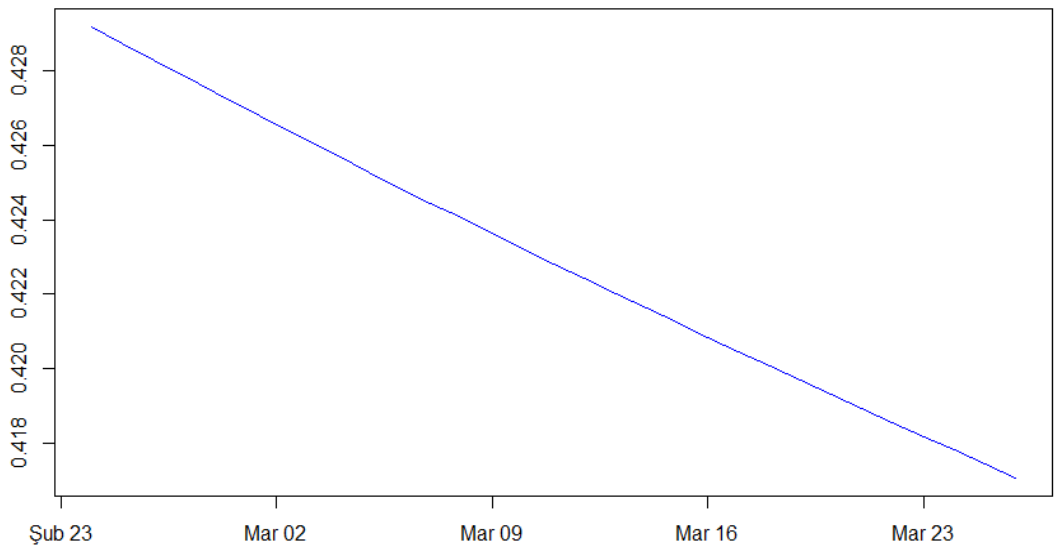

Figure 35. Normal reference rule of thumb smoothing parameter for quantile estimation of ten-day returns when $\tau = 0.01$ and $\eta = 1$

2.5. Concluding Remarks

Bitcoin -depending on its highly volatile structure realized historically- has always been a matter of interest by many researchers. In particular, with the coronovirus epidemic coming on the agenda, it has recently become one of the most controversial issues as a cryptocurrency due to the potential of cash to be able to carry viruses on it. For this reason, this paper has aimed to predict the volatility of Bitcoin returns based upon the study by Klemelä (2017). The core conclusions drawn from this paper can be outlined briefly with a few matters:

- Predictor variables $X_t = (X_{1t}, X_{2t})$ transformed in a way that will represent approximately standard normal distributions have shown a spatially more homogenous dispersion relative to the case in which there is no transformation.
- In comparison of (asymmetric) GARCH and (asymmetric) EWMA predictors for prediction horizon $\eta = 1$, the sums of squared prediction errors have been seen to be dominated by the end of 2017. At the same time, aggregated EWMA and Heston-Nandi GARCH(1,1) predictors have shown better performance than standard GARCH(1,1) predictor. Among EWMA predictors, it turned out that aggregated predictors perform better in case λ is smaller than 0.5. Also, these predictors are superior to cross-validated ones. For the year 2018 and beyond, the best performing one among all predictors considered in predicting squared returns has been determined as aggregated EWMA (with $\lambda = 0.2$).
- When the prediction horion is extended to 10-days ($\eta = 10$), the performance of GARCH(1,1) predictor has been deterioriated relative to the case of $\eta = 1$ in case (asymmetric) GARCH and (asymmetric) MA predictors are handled. In general sense, the best predictor performance from late 2017 belongs to aggregated (asymmetric) EWMA.
- For the prediction horion of one-day, GARCH(1,1) volatility as the kernel predictor prevails standard GARCH(1,1) predictor during all time periods and is also better than EWMA volatility such that this finding is also valid for the case with the return horizon of ten-days & prediction horizon of one day. However; when the prediction horizon becomes larger (the case of $\eta = 10$), EWMA volatility has shown a better performance than GARCH volatility.

- The perspective plot associated with kernel prediction has not revealed an obvious leverage effect. Failure to carry out the intrinsic notion of the leverage effect to Bitcoin can be attributed to the fact that Bitcoin does not have any capital structure (Catania & Sandholdt, 2019, p. 10).

Generally speaking, kernel predictors have improved the performance of volatility prediction for Bitcoin returns. In addition, this improvement has also been recorded against traditional GARCH(1,1) predictor in the quantile estimation of ten-day returns for $\tau = 0.01$ and $\eta = 1$ based on the student residuals contrary to the case with one-day returns.

REFERENCES

Aalborg, H. A., Molnár, P., & de Vries, J. E. (2019). What can explain the price, volatility and trading volume of Bitcoin?. *Finance Research Letters*, *29*, 255-265.

Abberger, K. (1995). *Volatility and conditional distribution in financial markets* (Diskussionsbeiträge - Serie II, No. 252). Universität Konstanz, Sonderforschungsbereich 178 - Internationalisierung der Wirtschaft, Konstanz

Abraham, B., & Ledolter, J. (1983). *Statistical methods for forecasting*. New York: John Wiley & Sons.

Acharya, S., Thomas, A., & Pani, B. (2018). Volatility of Bitcoin and its implication to be a currency. *International Journal of Engineering Technology Science and Research*, *5*(1), 1017-1024.

Alexander, C. (2008). *Market risk analysis, practical financial econometrics* (Vol. 2). John Wiley & Sons.

Andersen, T. G., Bollerslev, T., Christoffersen, P. F., & Diebold, F. X. (2006). Volatility and correlation forecasting. In G. Elliott, C. W. J. Granger & A. Timmerman (Eds,), *Handbook of Economic Forecasting* (pp. 777-878). North-Holland, Amsterdam.

Andersen, T. G., Bollerslev, T., & Diebold, F. X. (2010). Parametric and nonparametric volatility measurement. In Y. Aït-Sahalia & L. P. Hansen (Eds.), *Handbook of Financial Econometrics: Tools and Techniques* (pp. 67-137). North-Holland.

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica*, *71*(2), 579-625.

Artis, M. J., & Taylor, M. P. (1994). The stabilizing effect of the ERM on exchange rates and interest rates: Some nonparametric tests. *Staff Papers*, *41*(1), 123-148.

Audrino, F., & Barone-Adesi, G. (2003). *Semiparametric multivariate GARCH models for volatility asymmetries and dynamic correlations* (Working Paper No. 137). University of Southern Switzerland.

Audrino, F., & Bühlmann, P. (2001). Tree-structured generalized autoregressive conditional heteroscedastic models. *Journal of the Royal Statistical Society Series B (Statistical Methodology)*, *63*(4), 727-744.

Audrino, F., & Bühlmann, P. (2009). Splines for financial volatility. *Journal of the Royal Statistical Society Series B (Statistical Methodology)*, *71*(3), 655-670.

Balcilar, M., Bouri, E., Gupta, R., & Roubaud, D. (2017). Can volume predict Bitcoin returns and volatility? A quantiles-based approach. *Economic Modelling*, *64*, 74-81.

Bandi, F. M., & Renò, R. (2018). Nonparametric stochastic volatility. *Econometric Theory*, *34*(6), 1207-1255.

Barndorff-Nielsen, O. E., & Shephard, N. (2002). Estimating quadratic variation using realized variance. *Journal of Applied Econometrics*, *17*(5), 457-477.

Bauwens, L., Hafner, C. M., & Laurent, S. (2012). *Handbook of volatility models and their applications* (Vol. 3). John Wiley & Sons.

Berlinger, E., Illés, F., Badics, M., Banai, Á., Daróczi, G., Dömötör, B., ... & Márkus, B. (2015). *Mastering R for Quantitative Finance*. Birmingham, UK: Packt Publishing Ltd.

Black, F. (1976). Studies of stock price volatility changes. *Proceedings of the 1976 Meeting of the Business and Economic Statistics Section*, *American Statistical Association*, 177-181.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *31*(3), 307-327.

Bollerslev, T., Chou, R. Y., & Kroner, K. F. (1992). ARCH modeling in finance: A review of the theory and empirical evidence. *Journal of Econometrics*, *52*(1-2), 5-59.

Bollerslev, T., Engle, R. F., & Nelson, D. B. (1994). ARCH models. In R. F. Engle & D. L. McFadden (Eds.), *Handbook of Econometrics: Vol. 4* (pp. 2959-3038). North Holland, Elsevier.

Borror, C. M., Montgomery, D. C., & Runger, G. C. (1999). Robustness of the EWMA control chart to non-normality. *Journal of Quality Technology*, *31*(3), 309-316.

Bossaerts, P., Hafner, C., & Härdle, W. (1996). *Foreign exchange rates have surprising volatility* (No. 1996, 68). Humboldt University of Berlin, Interdisciplinary Research Project 373: Quantification and Simulation of Economic Processes.

Bouri, E., Gkillas, K., Gupta, R., & Pierdzioch, C. (2020). *Forecasting realized volatility of Bitcoin: The role of the trade war.* University of Pretoria, Department of Economics Working Paper Series (No. 2020-03).

Brown, R. G. (1962). *Smoothing, forecasting and prediction of discrete time series.* Englewood Cliffs, NJ: Prentice-Hall.

Bühlmann, P. L., & McNeil, A. J. (1999). Nonparametric GARCH models. In *Research Report/Seminar für Statistik and Department of Mathematics, Eidgenössische Technische Hochschule (ETH)* (Vol. 90). Seminar für Statistik, Eidgenössische Technische Hochschule.

Bühlmann, P., & McNeil, A. J. (2002). An algorithm for nonparametric GARCH modelling. *Computational Statistics & Data Analysis*, *40*(4), 665-683.

Caporale, G. M., & Zekokh, T. (2019). Modelling volatility of cryptocurrencies using Markov-Switching GARCH models. *Research in International Business and Finance*, *48*, 143-155.

Caporin, M., & Costola, M. (2019). Asymmetry and leverage in GARCH models: a News impact curve perspective. *Applied Economics*, *51*(31), 3345-3364.

Cassim, L. (2018). *Non-parametric estimation of GARCH (2, 2) volatility model: A new Algorithm* (MPRA Paper No. 86861). University Library of Munich, Germany.

Catania, L., & Sandholdt, M. (2019). Bitcoin at high frequency. *Journal of Risk and Financial Management, 12*(1), 1-20.

Cermak, V. (2017). *Can Bitcoin become a viable alternative to fiat currencies? An empirical analysis of Bitcoin's volatility based on a GARCH model.* Available at SSRN: https://ssrn.com/abstract=2961405.

Chaim, P., & Laurini, M. P. (2019). Is Bitcoin a bubble?. *Physica A: Statistical Mechanics and Its Applications*, *517*, 222-232.

Charles, A., & Darné, O. (2019). Volatility estimation for Bitcoin: Replication and robustness. *International Economics*, *157*, 23-32.

Chikhi, M., & Bendob, A. (2018). Nonparametric NAR-ARCH modelling of stock prices by the kernel methodology. *Journal of Economics and Financial Analysis*, *2*(2), 105-120.

Christie, A. A. (1982). The stochastic behavior of common stock variances: Value, leverage and interest rate effects. *Journal of Financial Economics*, *10*(4), 407-432.

Chung, S. S. (2012). A class of non-parametric volatility models: Application to financial time series. *Journal of Econometrics*.

Ciaian, P., & Rajcaniova, M. (2018). Virtual relationships: Short-and long-run evidence from Bitcoin and Altcoin markets. *Journal of International Financial Markets, Institutions and Money*, *52*, 173-195.

Conrad, C., Custovic, A., & Ghysels, E. (2018). Long-and short-term cryptocurrency volatility components: A GARCH-MIDAS analysis. *Journal of Risk and Financial Management*, *11*(2), 23.

Cont, R. (2007). Volatility clustering in financial markets: Empirical facts and agent-based models. In G. Teyssière & A. P. Kirman (Eds.), *Long Memory in Economics* (pp. 289-309). Springer, Berlin, Heidelberg.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, *7*(2), 174-196.

Cox, J. C., & Ross, S. A. (1976). The valuation of options for alternative stochastic processes. *Journal of Financial Economics*, *3*(1-2), 145-166.

Danielsson, J. (2011). *Financial risk forecasting: The theory and practice of forecasting market risk with implementation in R and Matlab* (Vol. 588). John Wiley & Sons.

Ding, Z., Granger, C. W., & Engle, R. F. (1993). A long memory property of stock market returns and a new model. *Journal of Empirical Finance*, *1*(1), 83-106.

Dobrovidov, A. V., & Tevosian, V. E. (2018). Nonparametric estimation of volatility and its parametric analogs. *Automation and Remote Control*, *79*(9), 1687-1702.

Ederington, L. H., & Guan, W. (2000). *Forecasting volatility*. Working Paper, University of Oklahoma.

Engle, R. (2004). Risk and volatility: Econometric models and financial practice. *American Economic Review*, *94*(3), 405-420.

Engle, R. F., & Ng, V. K. (1993). Measuring and testing the impact of news on volatility. *The Journal of Finance*, *48*(5), 1749-1778.

Fan, J., & Yao, Q. (1998). Efficient estimation of conditional variance functions in stochastic regression. *Biometrika*, *85*(3), 645-660.

Fan, J., & Yao, Q. (2017). *The elements of financial econometrics*. Cambridge: Cambridge University Press.

Fang, L., Bouri, E., Gupta, R., & Roubaud, D. (2019). Does global economic uncertainty matter for the volatility and hedging effectiveness of Bitcoin?. *International Review of Financial Analysis*, *61*, 29-36.

Florens-Zmirou, D. (1993). On estimating the diffusion coefficient from discrete observations. *Journal of Applied Probability*, *30*(4), 790-804.

Franke, J., Härdle, W., & Kreiss, J. P. (1998). *Nonparametric estimation in a stochastic volatility model*. Report in Wirtschaftsmathematik 37, University of Kaiserslautern.

French, K. R., Schwert, G. W., & Stambaugh, R. F. (1987). Expected stock returns and volatility. *Journal of Financial Economics*, *19*(1), 3-29.

Fricker, R. D. (2013). *Introduction to statistical methods for biosurveillance: with an emphasis on syndromic surveillance*. Cambridge: Cambridge University Press.

Garcin, M., & Goulet, C. (2019). Non-parametric news impact curve: A variational approach. *Soft Computing*, 1-16.

Giordano, F., & Parrella, M. L. (2019). Efficient nonparametric estimation and inference for the volatility function. *Statistics*, *53*(4), 770-791.

Glosten, L. R., Jagannathan, R., & Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance*, *48*(5), 1779-1801.

Guo, T., & Antulov-Fantulin, N. (2018). *Predicting short-term Bitcoin price fluctuations from buy and sell orders* (No.arXiv: 1802.04065). arXiv preprint.

Guo, T., Bifet, A., & Antulov-Fantulin, N. (2018). Bitcoin volatility forecasting with a glimpse into buy and sell orders. In *2018 IEEE International Conference on Data Mining (ICDM)* (pp. 989-994). IEEE.

Gupta, B. C., & Guttman, I. (2013). *Statistics and probability with applications for engineers and scientists.* Hoboken, NJ: John Wiley & Sons.

Gyamerah, S. A. (2019). Modelling the volatility of Bitcoin returns using GARCH models. *Quantitative Finance and Economics, 3*(4), 739-753.

Hagerud, G. E. (1997). *Specification tests for asymmetric GARCH* (Working Paper No. 163). Working Paper Series in Economics and Finance, Stockholm School of Economics.

Heid, F. (1996). *Non-parametric volatility estimation of exchange rates and stock prices*. (Discussion Paper No. A-533). Rheinische Friedrich-Wilhelms-Universität Bonn.

Hentschel, L. (1995). All in the family: Nesting symmetric and asymmetric GARCH models. *Journal of Financial Economics*, *39*(1), 71-104.

Heston, S. L., & Nandi, S. (2000). A closed-form GARCH option valuation model. *The Review of Financial Studies*, *13*(3), 585-625.

Hou, A., & Suardi, S. (2012). A nonparametric GARCH model of crude oil price return volatility. *Energy Economics*, *34*(2), 618-626.

Hu, S. (2011). *Nonparametric GARCH models for financial volatility*. Doctoral dissertation, University of Georgia.

Jang, H., & Lee, J. (2017). An empirical study on modeling and prediction of Bitcoin prices with bayesian neural networks based on blockchain information. *Ieee Access*, *6*, 5427-5437.

Jiang, G. J., & Knight, J. L. (1997). A nonparametric approach to the estimation of diffusion processes, with an application to a short-term interest rate model. *Econometric Theory*, *13*(5), 615-645.

Jondeau, E., Poon, S. H., & Rockinger, M. (2007). *Financial modeling under non-Gaussian distributions*. Springer Science & Business Media.

Katsiampa, P. (2017). Volatility estimation for Bitcoin: A comparison of GARCH models. *Economics Letters*, *158*, 3-6.

Katsiampa, P. (2019). Volatility co-movement between Bitcoin and Ether. *Finance Research Letters*, *30*, 221-227.

Khaldi, R., El Afia, A., & Chiheb, R. (2019). Forecasting of BTC volatility: Comparative study between parametric and nonparametric models. *Progress in Artificial Intelligence*, *8*(4), 511-523.

Kim, W., Lee, J., & Kang, K. (2019). The effects of the introduction of Bitcoin futures on the volatility of Bitcoin returns. *Finance Research Letters*, *33,* 101204.

Klemelä, J. (2017). *Volatility prediction using kernel regression*. Retrieved February 27, 2020, from http://jklm.fi/art/volapred/volapred.pdf.

Kupiec, P. H. (1989). Initial margin requirements and stock returns volatility: Another look. In F. R. Edwards (Ed.), *Regulatory Reform of Stock and Futures Markets* (pp. 189-203). Springer, Dordrecht.

Malladi, R., Dheeriya, P., & Martinez, J. (2019). Predicting Bitcoin return and volatility using gold and the stock market. *Quarterly Review of Business Disciplines*, *5*(4), 357-73.

Mandelbrot, B. (1963). The variation of certain speculative prices. *The Journal of Business*, *36*(4), 394-419.

Mattiussi, V. (2010). *Nonparametric estimation of high-frequency volatility and correlation dynamics*. Doctoral dissertation, City University London.

McKenzie, M. D. (1999). Power transformation and forecasting the magnitude of exchange rate changes. *International Journal of Forecasting*, *15*(1), 49-55.

Montgomery, D. C. (2013). *Introduction to Statistical Quality Control* (7th ed.). New York: John Wiley & Sons.

Naimy, V. Y., & Hayek, M. R. (2018). Modelling and predicting the Bitcoin volatility using GARCH models. *International Journal of Mathematical Modelling and Numerical Optimisation*, *8*(3), 197-215.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, *59*(2), 347-370.

Niizeki, M. K. (1998). A comparison of short-term interest rate models: Empirical tests of interest rate volatility. *Applied Financial Economics*, *8*(5), 505-512.

Othman, A. H. A., Alhabshi, S. M., & Haron, R. (2019). Cryptocurrencies, Fiat money or gold standard: An empirical evidence from volatility structure analysis using news impact curve. *International Journal of Monetary Economics and Finance*, *12*(2), 75-97.

Ou, P., & Wang, H. (2012). Nonparametric financial volatility modelling based on the relevance vector machines. *International Journal of Modelling and Simulation*, *32*(3), 192-197.

Pagan, A. R., & Hong, Y. S. (1991). Nonparametric estimation and the risk premium. In W. A. Barnett, J. Powell & G. E. Tauchen (Eds.), *Nonparametric and Semiparametric Methods in Econometrics and Statistics* (pp. 51-75). Cambridge University Press.

Pagan, A. R., & Schwert, G. W. (1990). Alternative models for conditional stock volatility. *Journal of Econometrics*, *45*(1-2), 267-290.

Palm, F. C. (1996). GARCH models of volatility. In G. S. Maddala & C. R. Rao (Eds.), *Handbook of Statistics* (pp. 209–240). Amsterdam: Elsevier Sciences.

Perry, M. B. (2010). The weighted moving average technique. In J. J. Cochran, L. A. Cox, P. Keskinocak, J. P. Kharoufeh & J. C. Smith (Eds.), *Wiley Encyclopedia of Operations Research and Management Science*. Hoboken, New Jersey: John Wiley & Sons, Inc.

Pichl, L., & Kaizoji, T. (2017). Volatility analysis of Bitcoin. *Quantitative Finance and Economics*, *1*, 474-485.

Poon, S. H., & Granger, C. W. (2003). Forecasting volatility in financial markets: A review. *Journal of Economic Literature*, *41*(2), 478-539.

Randal, J., Thomson, P., & Lally, M. (2004). Non-parametric estimation of historical volatility. *Quantitative Finance*, *4*(4), 427-440.

Reno, R. (2007). *Nonparametric volatility estimation in jump-diffusion models.* Course notes, University of Siena.

Roberts, S. W. (1959). Control chart tests based on geometric moving averages. *Technometrics, 1*(3), 239-250.

Samırkas, M. C. (2020). Modeling and forecasting volatility of Bitcoin. In S. Evci & A. Sharma (Eds.), *Studies at the Crossroads of Management & Economics* (pp. 263-271). London: IJOPEC.

Schwert, G. W. (1990). Stock volatility and the crash of'87. *The Review of Financial Studies*, *3*(1), 77-102.

Senarathne, C. W. (2019). Possible impact of facebook's libra on volatility of Bitcoin: Evidence from initial coin offer funding data. *Management of Organizations: Systematic Research*, *81*(1), 87-100.

Taylor, S. J. (1986). *Modelling financial time series*. Wiley.

Wang, P. (2003). *Financial econometrics methods and models*. Routledge.

Wang, L., Feng, C., Song, Q., & Yang, L. (2012). Efficient semiparametric GARCH modeling of financial volatility. *Statistica Sinica*, *22*(1), 249-270.

Wei, Y., Wang, Y., & Huang, D. (2010). Forecasting crude oil market volatility: Further evidence using GARCH-class models. *Energy Economics*, *32*(6), 1477-1484.

Wu, J., (2011). *Threshold GARCH model: Theory and application*. Working paper, The University of Western Ontario. Retrieved March 23, 2020, from http://publish.uwo.ca/~jwu87/files/JobMarketPaper_JingWu.pdf.

Yang, L. (2000). Finite nonparametric GARCH model for foreign exchange volatility. *Communications in Statistics - Theory and Methods*, *29*(5-6), 1347-1365.

Yu, M., Gao, R., Su, X., Jin, X., Zhang, H., & Song, J. (2019). Forecasting Bitcoin volatility: The role of leverage effect and uncertainty. *Physica A: Statistical Mechanics and Its Applications, 533*, 1–9.

Zakoian, J. M. (1994). Threshold heteroskedastic models. *Journal of Economic Dynamics and Control*, *18*(5), 931-955.

Zhang, X., Wang, Y., & Li, H. (2009). The contrast of parametric and nonparametric volatility measurement based on Chinese Stock Market. In J. Zhou (Ed.), *International Conference on Complex Sciences* (pp. 618-627). Springer, Berlin, Heidelberg.

CHAPTER III

NONPARAMETRIC MARKET DYNAMICS AMONG GOLD, CRUDE OIL & S&P500 STOCKS

3.1. Introduction

Understanding the dynamic relations between economic variables is of great importance with respect to policymakers. Since Sims (1980), VAR models have a comprehensive usage in the sphere of macroeconomics in identifying the dynamic behaviors of economic linkages in multivariate time series context (Bjørnland, 2000, pp. 3-4) depending on the simple linear representation regarding the joint dynamic behaviour of the variables and one reason that VAR models are so prominent is because of the need for capturing main characteristics of the business cycle (Kalli & Griffin, 2015, p. 2). On the other hand, among autoregressive models, conditional heteroscedastic models hold an important place in financial time series analysis with respect to giving an insight about the association between risk and return depending on the observed characteristic of volatility clustering. Discussions on the functional form of the conditional variance have been introduced by Pagan & Schwert (1990) and attracted many researchers to employ nonparametric estimation procedures which present flexibility by not requiring any specific assumptions on the functional form. Besides, nonparametric smoothing of nonlinear autoregressive time series provides practicability in terms of robustness in model selection and prediction (Yang, 2007, p. 85). In this context, as an alternative to the traditional GARCH modelling, a 'Conditional Heteroscedastic Autoregressive Nonlinear' (CHARN) model where the conditional mean and conditional variance (volatility) matrices appear as unknown functions of the past has been put forward (Härdle & Yang, 1996). It is remarkable to note that while GARCH models cannot catch asymmetries that occur due to leverage effect, CHARN models are quite successful in this regard; however, as expressed by Martins-Filho and Yao (2006), since they make it less possible to model the long memory feature of asset return processes depending on their intrinsic Markov characteristic, they offer a more restrictive framework than GARCH models (Brauchler, 2011, p. 19). Moving from the nonparametric VAR approach, this paper aims to analyze the dynamic linkages among gold, oil and stock prices.

Gold which is an enduring commodity that can be recycled as in jewelry form functions as a hedge against inflation and currency devaluation, therefore provides a defensive strategy for investors in the face of increased market uncertainties and can be regarded as a portfolio diversifier for minimizing the systematic risk. It has been a global precious commodity and an investment asset representing a fundamental role throughout the history. After World War II, the US Dollar was pegged to gold at a rate of $35 per troy ounce as a reserve tool through Bretton Woods system -which existed until the year 1971-, thus gold became the backbone of the global monetary system and in 1971, the policy pegged to gold was abandoned and a fiat currency system was adopted (Sujit & Kumar, 2011, p. 145). Undoubtedly, gold prices have experienced dramatic rises from time to time. As an investment asset, gold has been discussed historically as related to if it is connected with fears on rising inflation risk or not such that a weakened dollar is the harbinger of rising dollar prices of commodities (particularly, increasing the price of gold which is a dollar-denominated commodity) against the other currency -as the natural result of rising demand for US commodities- and typically results in high oil prices which are mostly stemmed from geopolitical instabilities, drives up the global inflation and in turn forces the gold price to soar. On the other hand, although it seems that there is a negative relationship between the value of US Dollar and the gold price; the direction of the linkage can turn to positive due to the economic situations of the other countries. For instance, when the currencies of opposite countries are confronted with economic downturns; people holding those currencies in hand will not be able to afford gold even US Dollar depreciates (Al-Ameer, Hammad, Ismail, & Hamdan, 2018, p. 361; Sujit & Kumar, 2011, p. 147).

Main factors that drive gold prices can be expressed as value of the US Dollar, investor behaviors, oil prices, central bank reserves, market conditions. Among the commodities, gold and crude oil are the two strategic ones that are intertwined with each other. Crude oil prices have a structure with a high volatility as the most commonly traded commodity in the global context. On the other hand, the position of gold among the precious metals is of vital importance with its rising prices that move in tandom with the prices of others (Le & Chang, 2011, p. 2). Oil price volatilities create market uncertainty and instability leading to higher inflation and unemployment levels, bringing with it also changes in stock prices. Increasing oil prices would create a decline in the stock price of companies performing based on the usage of oil consumption (Gilbert & Mork, 1984) (Gokmenoglu & Fazlollahi, 2015, p. 479). A general belief regarding the linkage between

gold and crude oil price says that both of them are close substitutes as safe havens against the fluctuations in the US Dollar, thus are in a positive relationship. In their study, Li and Chang (2011) mention two hypotheses for explaining this opinion. According to this, first hypothesis says that gold prices are affected by oil prices. The long recession stemmed from the 1973 petrol crisis can be exemplified for this hypothesis. To state thoroughly, high oil prices put a downward pressure on share prices creating unfavorable effects on economic growth and investors tend to prefer gold as an alternative asset. Because, in case the price of a share is equal to its discounted future cash flows; high oil prices may force the interest rates to rise in order to curb the inflationary pressures, hence result in decreasing profits (Jones, Leiby, & Paik, 2004), making bond investments more appealing against stock investments (Chittedi, 2012) and thus implying an inverse association between oil and stock markets (Akgul, Bildirici, & Ozdemir, 2015, p. 409). Besides, large revenues of prevailing oil exporting countries made from the sale of oil promote the gold investment -by creating a rise in the gold demand and therefore in the gold price implying a movement in the same direction between oil & gold- which is the case being valid if oil exporters allocate a remarkable portion of their asset portfolios to gold and buy gold in proportion to increased oil revenues. In addition, soaring oil prices are conducive to the inflation and in such a situation, an increase in the general price level will increase the gold demand and thus the gold price. As a result, oil prices and inflation usually move in the same direction and gold emerges as an effective inflation hedge tool (Le & Chang, 2011, pp. 4-5). The second hypothesis states that gold and oil prices are only correlated to the movements of common driving factors (e.g. the volatility of US dollar). Briefly, a prevalent saying of "correlation does not imply causation" shows itself here so that among oil and gold prices that are intercorrelated, one does not have to affect the other (Le & Chang, 2011, pp. 6-7). On the other hand, mining high-quality gold is more challenging by composing adverse environmental impacts and therefore the costs for the production of mining such type of golds will be high leading to higher gold prices. Another demand source for gold as related to clean energy concept is also stemmed from solar energy technology which undoubtedly uses gold for minimizing energy costs through golden nanoparticles placed into batteries being capable to transforming the solar energy it absorbs to electrical energy.

The essence of the matter is that in a durable economy where stock markets rise in general, investment demands have a tendency to be directed to assets excluding gold. When the stock market collapses implying an underperforming economy; investors

regard gold as a guard against bad market conditions, this would push the gold demand and also gold prices up. It can be inferred from here that gold prices are typically fed from a high-risk environment. That is, investors flock to gold during economic turmoil times by maintaining their belief regarding that gold is a safe haven and as a result, gold demand goes up. In ancient times especially, the Great Depression gave rise to the soaring gold prices and recently, a good example for this is no doubt the coronovirus agenda. As of the year 2020, the COVID-19 pandemic that could drag the global economy into a possible recession has triggered increases in global gold demand and it has been reported by Reuters that gold price climbed to over seven-year peak because of coronavirus outbreak-led fears of economic downturn by ascending over 1.5%.

Considering the importance of volatility in gold prices for the economy; in this paper, it has been aimed to put forward the dynamic behaviors between gold & the drivers of gold -which are crude oil and S&P500 stock prices here- by utilizing from nonparametric VAR approach in predictions and giving a comparison with some competitive models. To this end, the rest of the paper has been organized as follows: Section 3.2 reviews literature studies, Section 3.3 discusses the theoretical framework regarding the nonparametric vector autoregression, Section 3.4 presents empirical findings and Section 3.5 concludes the paper.

3.2. Literature Review

There exists a growing literature related to nonparametric VAR applications. In their study, Härdle & Tsybakov (1997) have covered the CHARN model as a generalization of the qualitative threshold ARCH model proposed by Gourieroux & Monfort (1992) and applied to DEM/USD foreign exchange rate with a simulation study for analyzing the finite sample properties based on local polynomial (LP) estimators. Fundamentally; it has been concluded that LP estimators of the conditional mean and volatility are pointwise joint asymptotic normal, DEM/USD exchange rate displays a U-shaped volatility function -namely "smiling face"- and risks of returns are much higher for extreme values taken on the past day. Härdle, Tsybakov & Yang (1998) have investigated the convergence rates of the nonparametric estimators of the unknown functions embedded in the conditional mean and volatility matrices in CHARN model in order to carry out these findings in the estimation of the conditional variance matrices for foreign exchange markets, extended the analysis of Härdle & Tsybakov (1997) to the multivariate case and reported that the conditional covariance surface estimations for the Deutsche Mark/US

Dollar & Deutsche Mark/British Pound returns display negative correlation in case they have opposite lagged values and positive correlation elsewhere. Yang, Hardle and Nielsen (1999) have studied the joint estimation of both the additive mean and the multiplicative volatility based on the integrated local polynomial estimation with an application to DEM/USD daily exchange rate returns. Hafner and Herwartz (2002) have introduced a wild bootstrap procedure which is robust under conditionally heteroscedasticity of unknown form for testing parameter restrictions in VAR models. Yang and Tschering (2002) have extended the CHARN model by deterministic seasonal components and suggested non- or semi-parametric estimation approaches for three seasonal nonlinear autoregressive models of varying seasonal flexibility. In their study, Julio-Román, Rodríguez-Niño & Zárate-Solano (2005) have aimed to estimate the conditional mean and volatility smile return functions for COP/USD exchange rate (in Colombian inter-bank market) under two distinct samples -one before the discretional interventions started and the other using the whole sample- at daily frequency utilizing from a nonparametric approach regarding CHARN type of models and local polynomial estimation methods; therefore investigating the effect of interventions on the conditional mean and volatility smile functions. As a result of findings, Julio-Román et al. (2005) have detected essentially a linear expected return function (conditional on the last observed return) with a small slope reflecting some degree of market inefficiency and a volatility smile having lack of symmetry for the pre-intervention sample. On the other hand, discretional interventions do not create a shift in the expected returns function; but have a remarkable impact on volatility smile implying an inclination to increasing the market risk perception nonsymmetrically. Wutsqa (2008) has proposed Vector Autoregressive Neural Network (VAR-NN) approach that shows nonlinear & nonparametric characteristics and belongs to the class of Feed Forward Neural Network. The empirical findings of the research have shown that VAR-NN performs well for approximating nonlinear functions and yields accurate forecasts in Monte Carlo simulation studies compared with VARMA model. Safari and Seese (2009) have dealed with the nonparametric volatility estimation of multiscale CHARN model by applying support vector regression (SVR) method along with wavelet analysis and analyzed daily stock exchange indices consisting of the S&P500, Nikkei225, Hang Seng, FTSE100 & DOW30. As a result, they have reported the superiority of their model by SVR in a single resolution against the technical benchmarks. Araveeporn (2011) has studied a nonparametric CHARN (NCHARN) model with autocorrelated errors utilizing from a class of penalized spline for modelling

the smooth unknown trend and heteroscedasticity with an application to the monthly Stock Exchange Rate of Thailand (SERT) series for totally 414 observations. According to the results, it has been revealed that the NCHARN model with penalized spline under autoregressive process of residuals outperforms the NCHARN model with classical penalized spline with respect to minimizing the mean absolute deviation. Brauchler (2011) has proposed a multivariate method presenting an extension of the local linear estimator of conditional mean & volatility used in CHARN model with an inclusion of an exogenous variable for estimating VaR & expected shortfall (ES) based on the extreme value theory and revealed that the proposed method is superior to the Gaussian GARCH methodology in general. Grobys (2012) has presented an extension of the nonparametric pairs bootstrap methodology to different VAR models considering the FTSE 100, DAX 30 and S&P 500 stock market data. Jeliazkov (2013) has proposed nonparametric dynamic VAR modelling for multivariate time series with an application to U.S. post war data on GDP growth, unemployment, interest rates & inflation by handling the specification, estimation & inference issues from a hierarchical Bayesian perspective and presenting debates on efficient MCMC sampling & model comparison techniques under a new framework in order to describe the unspecified covariate functions apart from taking also heteroscedasticity & other crucial modelling characteristics into account. Wu, Lin, Guan, Harada and Rojas (2017) have studied with Bayesian nonparametric VAR hidden markov models to perform robot introspection. Kapetanios, Marcellino and Venditti (2017) have suggested a nonparametric estimation approach for large dimensional VAR models with time-varying parameters and concluded that incorporating time variation into the VAR model parameters improves the prediction accuracy over models with a constant parameter structure. Kalli and Griffin (2018) have proposed a Bayesian nonparametric VAR (BayesNP-VAR) model which presents a flexibility in terms of explaining both nonlinear structures in conditional mean and heteroscedasticity in conditional variance with comparisons to other models. The short-run out-of-sample predictions have shown that the BayesNP-VAR approach outperforms competing models for all horizons. Considering neural networks -which present flexibility in estimating nonlinear models- as nonparametric in terms of not requiring any assumptions regarding a specific family of distributions for the underlying data; Yasin, Warsito and Santoso (2018) have suggested soft computation tools of Vector Autoregressive Neural network (VAR-NN) model based on graphical user interface for predicting the multivariate time series data with an application to the two series of stock price data from Indonesia Stock

Exchange and reported that VAR-NN model performs well in terms of both in-sample & out-sample for nonlinear function approximation. Another study regarding VAR models using NN algorithm (VAR-NN) has been proposed by Sugito et al. (2019) for predicting rainfall and wave height in Semarang city. Billio, Casarin and Rossini (2019) have introduced a novel Bayesian nonparametric Lasso prior (BNP-Lasso) for high-dimensional VAR models putting Dirichlet process & Bayesian Lasso priors together and presenting a comparison with alternative shrinkage approaches. Findings have shown that their proposed nonparametric Bayesian prior -which copes with overparametrization and overfitting problems by clustering the VAR coefficients into groups and shrinking those coefficients into a common location- performs the best forecasting performance.

3.3. Nonparametric Vector Autoregression (VAR)

Vector autoregression (VAR) approach introduced by Sims (1980) has emerged as a stochastic process which is an extension of the univariate autoregressive model to the class of multivariate linear time series models presenting a practical usage for modelling the dynamics of economic and financial time series to be used for forecasting, basic structural inferences or policy analysis aims (Zivot & Wang, 2006, p. 385). However, assuming fixed or specific linear forms for the conditional covariances in VAR models may not be realistic in an environment where economic theories can also be identified through nonlinear models -e.g. enabling to examine monetary policy shocks which may have asymmetric impacts or fiscal multiplier effects- and have been criticized negatively by many studies e.g. Engle (1982), Robinson (1983, 1984) & Teräsvirta (1994) in the econometric literature. There exist many approaches for modelling nonlinear relationships such as threshold autoregressive (TAR), exponential autoregressive (EXPAR) or smooth-transition autoregressive (STAR) models (Härdle et al., 1998, p. 221; Hubrich & Teräsvirta, 2013, p. 1).

Härdle et al. (1998, p. 221) have also proposed a vector CHARN model where the estimators regarding the conditional mean and the conditional variance matrices representing the unknown functions of past were structured by means of local polynomial (LP) estimator whose statistical features were covered by Tsybakov (1986). CHARN model in particular has a great significance for financial data. A useful approach for asymmetric modelling is a qualitative threshold ARCH (QTARCH) model introduced by Gourieroux and Monfort (1992). In their study, Gourieroux and Monfort (1992) have dealed with flexible parameterizations of the conditional mean & the conditional variance

specified as endogenous piecewise constant functions and put forward a general framework making the testing of the distinct restrictions which may concern both the conditional mean & the conditional variance possible contrary to the case in traditional VAR (Gourieroux & Monfort, 1992, pp. 159-160). CHARN model has appeared as a generalization of QTARCH which includes a constant number of threshold points and is given as

$$Y_t = \sum_{j=1}^{J} k_j I(X_t \in A_j) + \sum_{j=1}^{J} m_j I(X_t \in A_j)\varepsilon_t \ , \quad t = 1,2,.....$$

$$X_t = (Y_{t-1}^T, Y_{t-2}^T, ..., Y_{t-s}^T)^T \in \mathbb{R}^{sd} \ , \ Y_t \in \mathbb{R}^d \tag{3.1}$$

where $\{A_j\}_{j=1}^{J}$ with fixed J represents a partition of the support of the lagged values for Y; d and s denote the dimension and the number of lags respectively; (k_j) and (m_j) denote unknown parameter vectors and matrices respectively; $I(\cdot)$ represents the indicator function and ε_t denotes the white noise process. Based on this, with an enlargement towards a larger class of conditional mean and variance functions, CHARN model is represented as follows:

$$Y_t = f(X_t) + \Sigma^{1/2}(X_t)\varepsilon_t, \ \ t = 1,2,.....$$

$$X_t = (Y_{t-1}^T, Y_{t-2}^T, ..., Y_{t-s}^T)^T \in \mathbb{R}^{sd}, \ \ Y_t \in \mathbb{R}^d \tag{3.2}$$

Equation (3.2) can be said to concern with estimating the conditional mean vector function $f(\cdot)$ & the conditional volatility matrix function $\Sigma(\cdot)$ and to be a limit of Equation (3.1) in a way that the number of threshold points J converges to infinity. In their study, Härdle et al. (1998) handle the model given in Equation (3.2) where $t = s, s+1, ... , \ \ Y_t = (Y_{t1}, Y_{t2}, ..., Y_{td})^T \in \mathbb{R}^d , \ \ \varepsilon_t = (\varepsilon_{t1}, \varepsilon_{t2}, ..., \varepsilon_{td})^T \in \mathbb{R}^d$ and X_t are random variables; ε_t states i.i.d. process with $E(\varepsilon_{1j}) = 0$ for any $1 \le j \le d$ and $E(\varepsilon_{1j}^2) = 1$; the mean vector (f) and volatility matrix (Σ) functions are unknown; $\Sigma(x)$ is a matrix function with symmetric and positive definiteness for any value of x $(x \in \mathbb{R}^{sd})$ and $X_s = (Y_{s-1}^T, Y_{s-2}^T, ..., Y_0^T)^T$ is the initial value that is independent of $\{\varepsilon_t\}$ (Härdle et al., 1998, pp. 222, 226). Masry & Tjøstheim (1995) and Bossaerts, Härdle & Hafner (1996) are other studies related to CHARN models under nonparametric framework. One method for obtaining the estimators can be expressed in that way: First, $\hat{f}(X_t)$ is estimated via

some nonparametric method; then heteroscedastic residuals $\hat{\varepsilon}_t = Y_t - \hat{f}(X_t)$ are estimated and $e_t = \hat{\varepsilon}_t - \bar{\varepsilon}$ is obtained. Finally, $\hat{\sigma}^2(X_t)$ is obtained through $e_t^2 = \sigma^2(X_t) + \eta_t$ (describing $\hat{\sigma}^2(X_t)$ as greater than zero for all t) (Julio-Román et al., 2005, pp. 10-11). The theoretical frameworks of LP regression approach have been suggested by Stone (1977) and Cleveland (1979) (see also Fan & Gijbels (1996) for a detailed explanation). In a nonparametic model like $Y = f(X) + \xi$, the only assumption about regression surface $f(X)$ is the smoothness of it in terms of X. Under this assumption, the unknown $f(\cdot)$ function is locally approximated in a neighbourhood of X_0 through a polynomial of order p (and thus differentiable up to order $p+1$) (Julio-Román et al., 2005, p. 11) by carrying out the Taylor series expansion arguments as follows:

$$f(u) \approx \sum_{j=0}^{p} \frac{f^{(j)}(X_0)}{j!}(u - X_0)^j \overset{def}{=} \sum_{j=0}^{p} \beta_j (u - X_0)^j \tag{3.3}$$

where $f^{(j)}$ represents the j-th derivative of f. So, for X_i that is sufficiently close to X_0; the fit at X_i based on Equation (3.3) can be expressed as

$$f(X_i) \approx \sum_{j=0}^{p} \beta_j (X_i - X_0)^j \overset{def}{=} X_i^T \beta \tag{3.4}$$

where $X_i = (1, (X_i - X_0), ..., (X_i - X_0)^p)'$ and $\beta = (\beta_0, \beta_1, ..., \beta_p)'$. In the case of $p = 0$, LP fitting becomes equivalent to kernel regression (i.e. Nadaraya-Watson estimator emerges as the solution of a local constant fitting). In the case of $p = 1$ (one-dimensional case), we obtain a local linear fit (or regression smoother) which includes an advantage against the kernel regression in terms of estimating not only the regression curve, but also its first derivative. Additionally, the local linear estimator has a pleasant characteristic in terms of being successful at automatic corrections to boundary biases without requiring the usage of specific boundary kernels and according to the study by Fan (1993), local linear estimators have been found to provide the property of full asymptotic minimax efficiency among all linear estimators (Fan, Gasser, Gijbels, Brockmann, & Engel, 1997, p. 80). By intuition, observations that are close to X_0 carry more information about $f(X)$ and this suggests performing a locally weighted polynomial regression given as $\sum_{i=1}^{n} (Y_i - X_i'\beta)^2 K_h(X_i - X_0)$ (for (X_i, Y_i) sample with $i = 1, ..., n$) where K denotes the kernel function and h denotes the bandwidth or smoothing parameter which controls the

smoothness (also controls the overall sizes of the local neighborhoods [Julio-Román et al., 2005, p. 24] and the modelling bias [Fan & Gijbels, 1996, p. 59]) of the regression function. In this context, two points should be considered carefully in LP regression: the choice of h and the choice of p. In case a too large h is selected, the resulting estimate does not capture well characteristics regarding the data and in the case of a too small value of h, it can be faced with spurious sharp forms (Fang, Li, & Sudjianto, 2006, pp. 180-182). For choosing a suitable bandwidth value, plug-in and cross-validation methods can be used. It is also remarkable to note that in LP fitting, not identifying the distributional assumptions on errors ξ presents robust estimation methods and for the fit at point X_0, the regression surface $f(X)$ is well approximated via using suitable polynomial functions in a specifically defined neighbourhood of X_0 -weighted by their distance from X_0- (Julio-Román et al., 2005, pp. 24-25). A treatment regarding nonparametric estimation of the models in Equation (3.2) has been also mentioned by Masry and Tjøstheim (1995) who present strong convergence & asymptotic normality results under mild conditions. For this very general case of the model, estimation is possible only for small dimensions because of "curse of dimensionality" (Hafner, 1998, p. 67).

On the other hand, in order to investigate the behaviors of two time series -namely x_t and y_t [where $x_t = (x_t, ..., x_{t-p})$ and $y_t = (y_t, ..., y_{t-q})$]; Singh and Ullah (1985) have developed nonparametric estimations for joint DGP, conditional DGP & VAR models (i.e. the conditional mean [dynamic regression function] of (x_t, y_t) given their past values $(x_{t-1}, ..., x_{t-p}, y_{t-1}, ..., y_{t-q})$ in the case of $p = q$ for p and $q > 0$) which can be modified to vector processes. In case x_t and y_t exhibit Gaussian processes (GP), two assumptions that should be set in analyzing them -as concerned with the functional form of the model and distribution of error terms- do not pose a problem depending on the fact that the mean and autocovariances of the process can be efficiently estimated in an asymptotical manner under some weak dependence conditions on the process. Coping with issues regarding these two assumptions is possible through nonparametric estimation of the conditional mean (Singh & Ullah, 1985, pp. 27-28). In accordance with the purpose of this paper, only nonparametric VAR model estimation approach among another ones proposed in Singh and Ullah (1985) will be discussed in short.

Singh and Ullah (1985) report some basic assumptions regarding the analysis. Firstly, $\{x_t,\ t = 0, \pm 1, ...\}$ and $\{y_t,\ t = 0, \pm 1, ...\}$ represent (jointly) strictly stationary

real-valued stochastic processes that are defined on a probability space. Let $\{w_t\}$ be a stationary stochastic process $\{w_t,\ t=0,\pm1,.....\}$ and $\mathbf{B}_l^{l'}$ be the σ-field of events generated by stochastic vectors $w_l, w_{l+1},...., w_{l'}$. Then, the process $\{w_t\}$ satisfies the strongly mixing condition -introduced by Rosenblatt (1956) to prove the central limit theorem for weakly dependent random variables (Andrews, 1983, p. 1)- in case α_j that is described as

$$\alpha_j = \sup_{\mathbf{A}\in\mathbf{B}_{-\infty}^{t}\ \mathbf{B}\in\mathbf{B}_{t+j}^{\infty}} \left|\mathbf{P(A\cap B)-P(A)P(B)}\right| \tag{3.5}$$

converges to zero as j goes to infinity. Here the function α_j which is called the mixing coefficient of $\{w_t\}$ measures the dependence of the processes $\{w_t\}_{-\infty<t<l}$ and $\{w_t\}_{l+j\le t<\infty}$ and the joint process $\{(x_t, y_t)\}$ is assumed to be sharply mixing with α_j (Singh & Ullah, 1985, pp. 29-30).

Assume that f represents the joint DGP of (x_t, y_t) which is an $m=(p+q+2)$-variate density function and v is the point in $\Re^m$ (the m-dimensional Euclidean space) at the f function is desired to be estimated. The conditional density $g(v_1 \mid v_2,.....,v_m)$ of x_t at v_1 given $x_{t-1}=v_2,....,x_{t-p}=v_{p+1}, y_t=v_{p+2},...., y_{t-q}=v_{p+q+2}(m=p+q+2)$ can be estimated through the following proposed estimator of $g(v_1 \mid \mathbf{v}^*)$ -where $\mathbf{v}^*$ indicates the vector $(v_2,.....,v_m)$- using f_n which indicates the estimator of f as follows:

$$g_n(v_1 \mid \mathbf{v}^*) = \frac{f_n(v)}{\int f_n(v)dv_1} \tag{3.6}$$

(Singh & Ullah, 1985, pp. 35-36). For computing the dynamic regression function of x_t on $\mathbf{v}_t^* = (x_{t-1},....., x_{t-p}, y_t,....., y_{t-q})$, firstly it is required to know the conditional density of x_t given $\mathbf{v}_t^*$ which is estimated nonparametrically using g_n given in Equation (3.6). The proposed estimator of $f(v)$ based on $v_t=(x_t, y_t)$ for $t=1,.....,n$ is calculated as

$$f_n(v) = \frac{1}{n}\sum_{t=1}^{n}\delta_t(v) \tag{3.7}$$

where

$$\delta_t(v) = h^{-m}K\left(\frac{v_t - v}{h}\right) \tag{3.8}$$

for a fixed K in $\mathcal{K}^r$. Here $h = h_n$ is a function of n and greater than zero (In case $n \to \infty$, h_n converges to zero), K represents all real-valued Borel-measurable bounded functions and is assumed to be symmetric around zero, $\mathcal{K}^r$ represents the class of all K functions on $\mathcal{R}^m$ (Singh & Ullah, 1985, pp. 29-31, 38). Substituting the exact value of f_n into Equation (3.6), we obtain

$$g_n(v_1 \mid \mathbf{v}^*) = \frac{h^{-1}\sum_{t=1}^{n} K((v_t - v)/h)}{\sum_{t=1}^{n} K^*((\mathbf{v}_t^* - \mathbf{v}^*)/h)} = \frac{f_n(v)}{f_n^*(\mathbf{v}^*)} \tag{3.9}$$

where $K^*(\mathbf{v}^*) = \int K(v)dv_1$ and

$$f_n^*(\mathbf{v}^*) = (nh^{m-1})^{-1}\sum_{t=1}^{n} K^*\left(\frac{\mathbf{v}_t^* - \mathbf{v}^*}{h}\right) \tag{3.10}$$

Additionally, when $f^*(\mathbf{v}^*)$ is denoted by $\int f(v)dv_1$, we get $g(v_1 \mid \mathbf{v}^*) = f(v)/f^*(\mathbf{v}^*)$.

Now that the conditional density of x_t given $\mathbf{v}_t^*$ has been specified; the nonparametric estimator of the regression function of x_t on $\mathbf{v}_t^*$, evaluated at a point $\mathbf{v}^*$ of $\mathbf{v}_t^*$ in $\mathcal{R}^{m-1}$ is presented as follows:

$$r_n(\mathbf{v}^*) = \hat{E}(x_t \mid \mathbf{v}_t^* = \mathbf{v}^*) = \int v_1 g_n(v_1 \mid \mathbf{v}^*)dv_1 = \frac{\sum_{t=1}^{n}\{x_t K^*((\mathbf{v}_t^* - \mathbf{v}^*)/h)\}}{\sum_{t=1}^{n} K^*((\mathbf{v}_t^* - \mathbf{v}^*)/h)} \tag{3.11}$$

where $K^*(\mathbf{v}^*) = \int K(v_1, ..., v_m)dv_1$. Let $\int v_1 f(v_1, ..., v_m)dv_1$ be denoted by $l(\mathbf{v}^*)$, $(nh^{m-1})^{-1}\sum_{t=1}^{n} x_t K\left(\frac{\mathbf{v}_t^* - \mathbf{v}^*}{h}\right)$ by $l_n(\mathbf{v}^*)$ and the true value of the regression at $\mathbf{v}^*$ by $r(\mathbf{v}^*)$. In this case,

$$r(\mathbf{v}^*) = \frac{l(\mathbf{v}^*)}{f^*(\mathbf{v}^*)} \text{ and } r_n(\mathbf{v}^*) = \frac{l_n(\mathbf{v}^*)}{f_n^*(\mathbf{v}^*)} \tag{3.12}$$

(Singh & Ullah, 1985, pp. 36, 38-39). For more information, see Singh and Ullah (1985).

A random forest (RF) as one of the effective machine learning algorithms states an ensemble regression and classification method introduced by Breiman (2001) which can also be performed in modelling and forecasting predictions for a vast array of research domains such as stock price forecasting, data mining, bioinformatics, computational biology or ecology. It is a nonparametric regression method where regression trees from

randomly selected feature subsets and bagged samples of the training which form the building blocks of the algorithm are ensembled (Nguyen, Huang, & Nguyen, 2015, p. 326) and random forest does not overfit depending on the 'Law of Large Numbers' (Breiman, 2001, p. 29). Here, bagging refers to the abbreviated form of bootstrap aggregation where bootstrap expresses random selections from the original data with replacement and aggregating expresses averaging all predictions from all trees (Ayyadevara, 2018, p. 107). In this paper, RF algorithm has been also employed for predicting nonparametric VAR. For further readings related to RF, see Breiman (2001) or Cutler, Cutler & Stevens (2012).

3.4. Data and Empirical Findings

This study investigates the dynamic relationships among the prices of the gold futures contract that is traded on NYMEX, NYMEX crude oil futures and S&P 500 stock market index futures contracts for the month June. Data have been extracted from the Yahoo Finance website covering the daily adjusted close prices (in dollars) over March 22, 2000 through May 8, 2020 time period getting to totally 5033 observations. While determining the time period, dates including common observations have been taken into consideration.

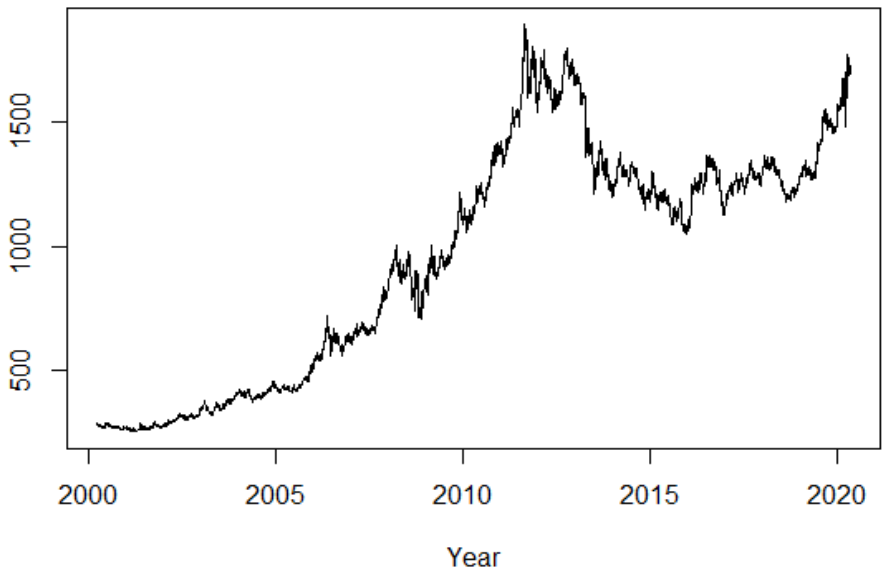

Figure 36. NYMEX gold price per troy ounce in USD

Figure 37. NYMEX crude oil price per barrel

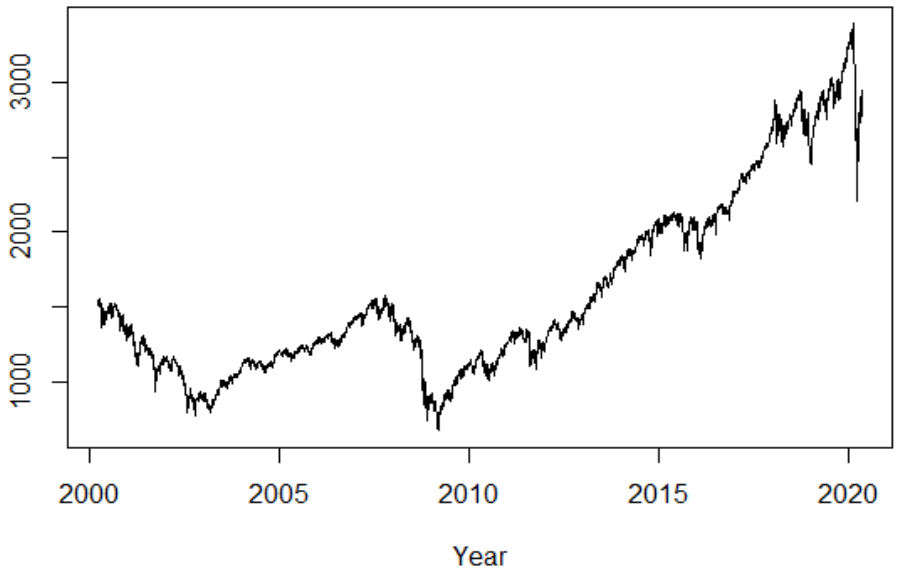

Figure 38. E-Mini S&P500 index futures

The graphs of gold, crude oil and stock market index futures prices have been plotted in Figure 36, Figure 37 and Figure 38 respectively. A conspicuous point emerges among these figures such that it is apparent from Figure 37 that crude oil price takes a negative value for the first time throughout the history and because of that returns of the given series have been computed by taking the percentage change of the futures prices instead of using logarithmic returns. Computed returns for NYMEX gold futures, NYMEX crude oil futures and S&P 500 stock market index futures contracts have been denoted in short as “GOLD”, “OIL” and “S&P500” respectively. The sudden plummet of the June 2020

NYMEX crude oil futures contract closing price below zero -from 17.04 U.S. Dollars on 19th April, 2020 per barrel towards minus 2.72 U.S. Dollars per barrel on 20th April, 2020 (even falling as low as minus $39.44 in the same day)[3]- has been experienced due to the coronavirus pandemic (COVID-19).

The nonparametric VAR application of this paper has been implemented based on the quick note discussed in Viole (2019) using "NNS.VAR" function in the 'NNS' package of R-project software which is the abbreviation of Nonlinear Nonparametric Statistics. In the analysis, it has been aimed to forecast next 7 days (i.e. the prediction horizon $h = 7$) for futures returns using multivariate nonparametric VAR model and then evaluate the accuracies of traditional VAR, random forest & nonparametric VAR model -incorporating an autoregressive model with nonlinear regressions of component series- for the multivariate time series [for more information about the procedure, see Viole (2020)]. To this end, the testing set includes the last seven observations as out-of-sample while the training set for VAR application includes 5025 observations for all return series. Utilizing from the training set, the appropriate number of lags has been determined by considering the different information criteria given as follows:

Table 17

Information Criteria for Determining Optimal Lag Number in VAR Model

Lags	AIC	HQ	SIC	FPE
1	4.484752	4.490219	4.500353	88.654941
2	4.467695	**4.477262**	**4.494998**	87.155571
3	4.468770	4.482438	4.507774	87.249360
4	4.471284	4.489053	4.521990	87.468978
5	4.468250	4.490120	4.530657	87.204026
6	4.468907	4.494877	4.543015	87.261345
7	4.467123	4.497193	4.552932	87.105790
8	4.466272	4.500443	4.563782	87.031684
9	4.465277	4.503548	4.574489	86.945146
10	**4.465256**	4.507627	4.586168	**86.943284**

Note. "AIC", "HQ", "SIC" and "FPE" represent Akaike Information Criterion, Hannan-Quinn Criterion, Schwarz Information Criterion and Final Prediction Error Criterion respectively.

According to Hurvich and Tsai (1989), Akaike Information Criterion (AIC) is biased towards choosing a model with over-parameterization while SIC shows consistency in an asymptotical manner (Balcilar, Saint Akadiri, Gupta, & Miller, 2019, p. 70). Depending

[3] Numerical data have been obtained from https://finance.yahoo.com/quote/CL%3DF/history?p=CL%3DF.

on this information and the given results in Table 17 which summarizes the AIC, HQ, SIC and FPE criteria for detecting the optimal lag order in VAR model, the lag order minimizing SIC has been detected as 2 amongst maximum 10 lags (the values in bold-type in Table 17 show the appropriate lag lengths chosen for each information criteria seperately).

Table 18

A Summary regarding Correlations and RMSE Metrics of Traditional VAR Predictions versus the Actual Values for $h = 7$

	Correlations	RMSEs
GOLD	-0.4285714	1.140067
OIL	0.2500000	14.505741
S&P500	0.8928571	2.107882

Considering multivariate VAR(2) model, Table 18 presents the correlations and root mean square error (RMSE) accuracy measures belonging to the traditional VAR predictions for 7-day ahead against the actual out-of-sample values. Correlations have been computed using the Spearman method. According to the results of traditional multivariate VAR(2) estimates, VAR forecasts of only GOLD equation are negatively correlated with the actual out-of-sample observations in moderate degree. On the other hand, VAR forecasts for S&P500 equation are in a very strong relationship with actual values having a correlation coefficient of 0.8928571 relative to the case with OIL equation. When RMSE metrics are evaluated for three VAR equations, the best & the worst forecasts have been obtained through GOLD (presenting the best fit) and OIL equations with the lowest (1.140067) and the greatest (14.505741) RMSE values respectively. Thus, the forecasts of OIL equation underperform when compared to other two equations and the forecasted values are very close to the actual out-of-sample values for GOLD and S&P500 equations. As a general conclusion, traditional VAR outcomes considering both correlation coefficients and RMSE metrics have shown that more accurate forecasts are obtained in case gold or S&P500 returns are taken as dependent variables in VAR specifications.

Table 19 shows a comparison regarding the forecasts of traditional VAR, NNS.ARMA -which represents the autoregressive model incorporating nonlinear regressions of component series (Viole, 2020, p. 14)- and automatic ARMA (Autoregressive Moving Average) procedures against the actual out-of-sample

observations for 7-day forecast. If necessary to mention correlations first, it can be said that according to NNS.ARMA the forecasts are negatively correlated with the actual values for all equations. The largest correlation coefficient belongs to S&P500 equation while the lowest one belongs to OIL equation. Automatic ARMA results have not recorded any correlation value for GOLD equation (denoted by NA). However, when we fit an ARMA model to the multivariate time series; the forecasts for S&P500 equation are seen to be strongly related with the actual observations also showing consistency with NNS.ARMA results in terms of correlations. Correlations for ARMA forecasts of OIL and S&P500 equations seperately versus the actual values are positive as in traditional VAR correlations. Concisely, when all three procedures are taken into account together, the greatest correlation coefficient of 7-day forecasted values with the actuals belongs to S&P500 returns equation.

Table 19

A Comparison of Correlation Coefficients and RMSEs for Forecasts of Traditional VAR, NNS.ARMA and Traditional ARMA versus the Actual Values for $h = 7$

Equations	Traditional VAR RMSEs	Traditional VAR Correlations	NNS.ARMA RMSEs	NNS.ARMA Correlations	ARMA RMSEs	ARMA Correlations
GOLD	1.1401	-0.4286	1.1587	-0.1071	1.0900	NA
OIL	14.5057	0.2500	19.2860	-0.0714	18.1665	0.4643
S&P500	2.1079	0.8929	2.5621	-0.4286	2.1056	0.7500

On the other hand, RMSE metrics for all methods -namely traditional VAR, NNS.ARMA and automatic ARMA- have the same interpretation such that the forecasted and the actual values are too far from each other with the largest RMSE in the case of OIL equation. Though traditional VAR outperforms the other two methods in terms of RMSE for crude oil returns. Therefore, the GOLD equation with the smallest RMSE and S&P500 equation with a RMSE value close to the former present more accurate forecasts for 7-day horizon when compared to the actual out-of-sample observations. The forecasted values based on the traditional VAR, NNS.ARMA & ARMA procedures and the actual out-of-sample values have been plotted against the forecast period in Figure 39, Figure 40 and Figure 41 for GOLD, OIL and S&P500 equations respectively (where the black, red, blue and brown lines represent the actual, NNS.ARMA, VAR and ARMA forecasts respectively).

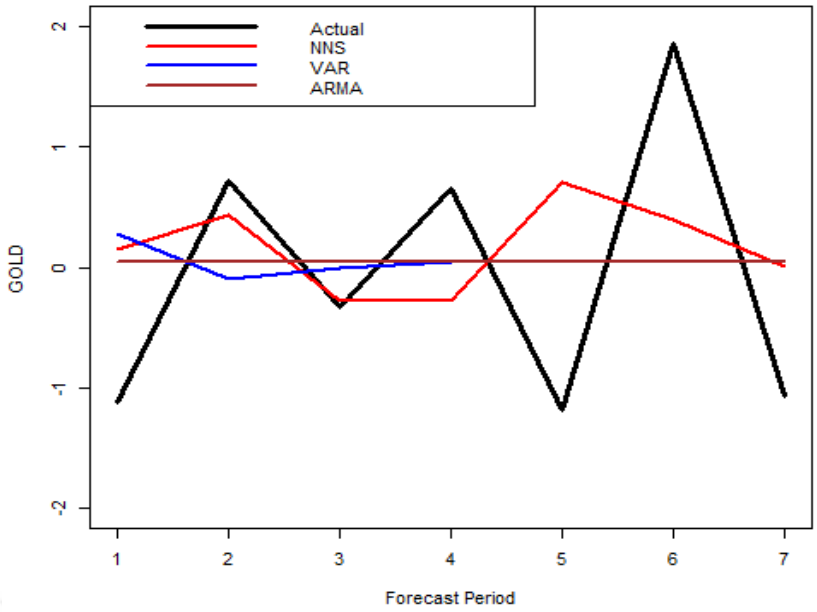

Figure 39. Plots of VAR, NNS.ARMA & ARMA forecasts and the actual out-of-sample observations against the forecast horizon for gold returns

If an overall assessment is made for Figure 39, it is seen that the forecasts of ARMA and VAR models are quite close to each other; even they have matched precisely after the 4th horizon. When compared to the actual values; ARMA for 1-day ahead, NNS.ARMA for 2-day or 3-day ahead, ARMA & VAR for 4-day or 5-day ahead and NNS.ARMA (for 6-day ahead) models generate more accurate forecasts for gold returns. In case the forecast horizon is 7, three methods -namely NNS, VAR and ARMA- can be said to give results that are almost very close to each other.

When compared to the actual values, Figure 40 reveals that the traditional VAR model generates more accurate forecasts for 1-day, 2-day, 4-day and 6-day ahead for the crude oil returns. In case the forecast horizon is 2 or 6, VAR forecasts seem to match up with the actual out-of-sample observations. For 3-day ahead, the forecasts of NNS.ARMA and VAR outperform the ones of ARMA model. On the other hand, the worst forecast results have been revealed through the traditional VAR model for the forecast period 5. When the forecast horizon is 7 (which is the interested case in this paper), VAR model has given the best forecasts for crude oil returns versus the other models. This finding is coherent with the results presented in Table 19. Indeed, the value of RMSE regarding the VAR (14.5057) for crude oil returns is the smallest when compared to the other two methods.

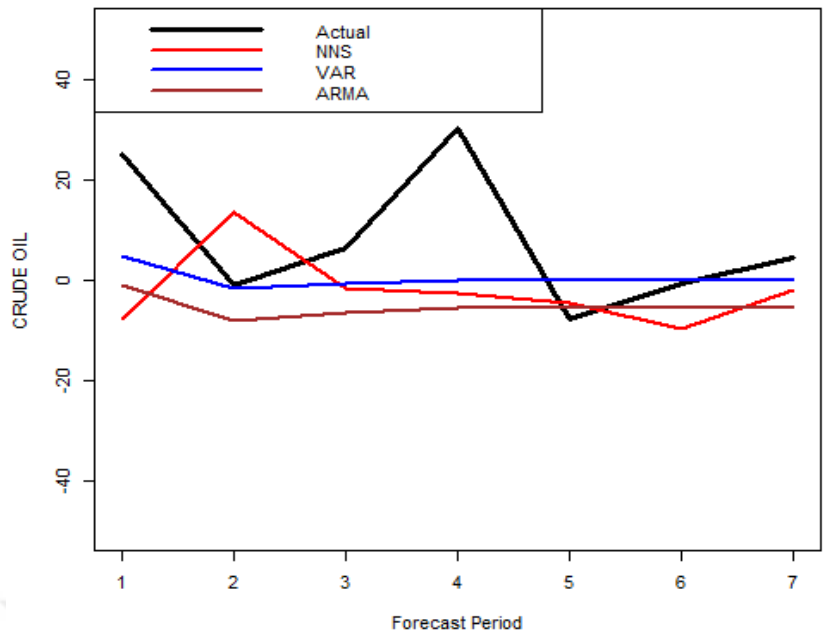

Figure 40. Plots of VAR, NNS.ARMA & ARMA forecasts and the actual out-of-sample observations against the forecast horizon for crude oil returns

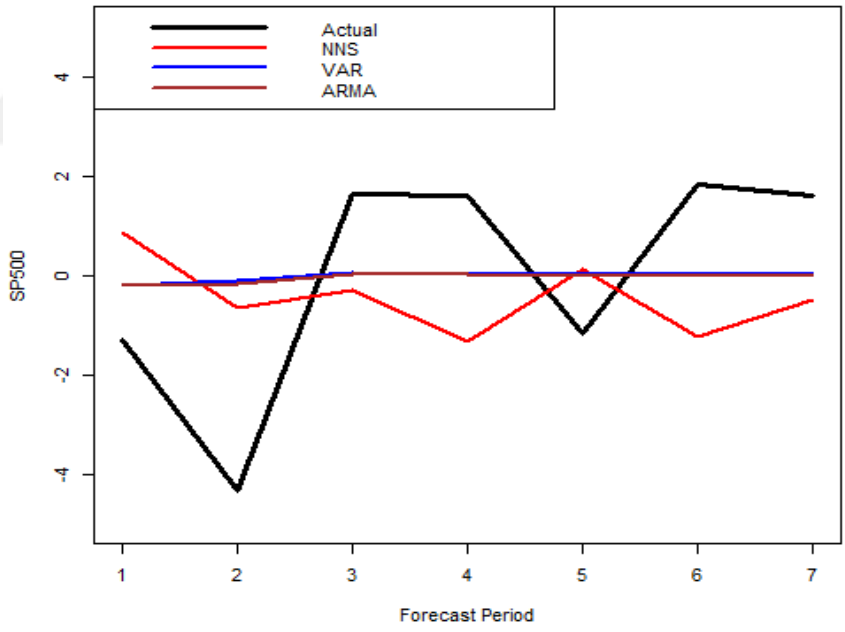

Figure 41. Plots of VAR, NNS.ARMA & ARMA forecasts and the actual out-of-sample observations against the forecast horizon for S&P500 returns

In general sense, Figure 41 shows that the traditional VAR and ARMA forecasts have almost exactly matched with each other for S&P500 returns as it is in gold returns for the forecast period 4 and after. When compared to the actual values, the worst forecasts have been obtained using NNS.ARMA model for 1-day ahead; for 2-day ahead

this situation has reversed and NNS.ARMA has been superior to other models. For 3-day and 4-day ahead, VAR and ARMA models have produced better forecasts relative to NNS.ARMA model. All three models have produced very close estimates for 5-day ahead. As for the 6th and 7th horizons, it can be said that the traditional VAR and ARMA models have performed better than NNS.ARMA in the S&P500 returns prediction.

Table 20

A Comparison of Correlation Coefficients and RMSEs for Forecasts of NNS and RF in addition to VAR and NNS.ARMA versus the Actual Values for 7-day ahead Forecast

Equations	Traditional VAR RMSEs	NNS.ARMA RMSEs	NNS RMSEs	NNS Correlations	RF RMSEs	RF Correlations
GOLD	1.1401	1.1587	0.9760	0.5714	1.3128	0.2143
OIL	14.5057	19.2860	17.3290	-0.0357	15.4847	-0.3571
S&P500	2.1079	2.5621	2.2633	-0.4286	2.7718	-0.2143

As similar to the previous results, RMSE values for NNS and RF in Table 20 besides the traditional VAR and NNS.ARMA methods reveal that these methods generate the forecasts that are far from the actual values based on the largest RMSE metrics when the dependent variable is chosen as crude oil returns rather than gold or S&P500 returns. Nevertheless, apart from the traditional VAR method which performs as the best with the lowest RMSE value of 14.5057, RF is superior to NNS and NNS.ARMA for OIL equation. NNS and RF methods have produced the same signs in terms of correlations which is the case contrary to the traditional VAR correlations given in Table 19. Taking all series into account, it is also remarkable to say that NNS forecasts are more close to the actual values when compared to NNS.ARMA method. Besides, it can be concluded that amongst gold, crude oil and S&P500 returns; it is more logical to choose gold returns as the dependent variable versus other returns depending on the recorded lowest RMSEs for all methods. In case gold and S&P500 returns are dependent variables seperately; RF underperforms VAR, NNS.ARMA and NNS models.

Forecasted values based on the traditional VAR, NNS & Random Forest (RF) models and the actual out-of-sample values have been plotted against the forecast periods up to 7 in Figure 42, Figure 43 and Figure 44 for gold, crude oil and S&P500 returns respectively (where the black, red, blue and green lines represent the actual, NNS, VAR and RF forecasts respectively). While applying RF, the number of trees to grow has been

chosen as 100. In addition, NNS denotes here a multivariate nonlinear nonparametric regression ensembling the forecasted variables and also including NNS.ARMA estimates of the variables, but is not the same as NNS.ARMA.

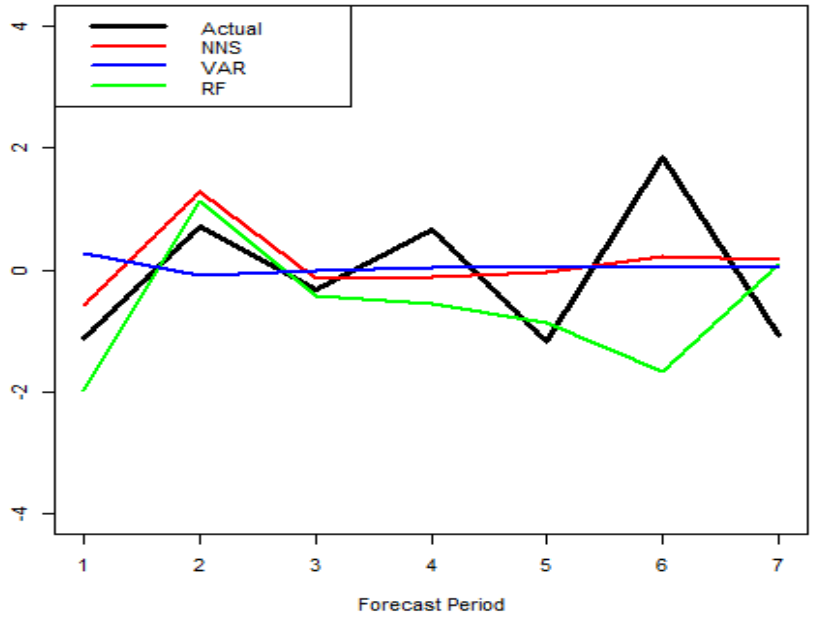

Figure 42. Plots of VAR, NNS & RF forecasts and the actual out-of-sample observations against the forecast horizon for gold returns

For gold returns, Figure 42 reveals that if we make a general evaluation; until 3rd horizon and beyond, NNS and VAR forecasts are very close to each other. When compared to the actual values, the traditional VAR model has produced the worst forecasts for 1-day ahead. For 2-day ahead, RF has given the results that are slightly better than NNS although they seem to be close to each other. Also for 3-day and 5-day ahead, RF has been superior to other models. VAR model for 4-day ahead and NNS model for 6-day ahead have reported the forecasts that are more close to the out-of-sample observations for gold returns against other models. When the forecast horizon is 7; NNS, VAR and RF can be said to reveal the results that are very close to each other.

Figure 43 shows that for 1-day, 2-day and 6-day ahead; VAR model has produced the best forecasts for crude oil returns. When the forecast period is 3; the forecasts of VAR, RF and NS models are quite close to each other. However, VAR and RF methods are a little bit better than NNS. The forecasts of all three models given here almost coincide with each other at the 4th forecast period. Besides, for 5-day ahead, the forecasts of NNS have outperformed against VAR and RF. On the other hand, in case the forecast

horizon is 7; NNS, VAR and RF methods have recorded forecast results that are almost very close to each other.

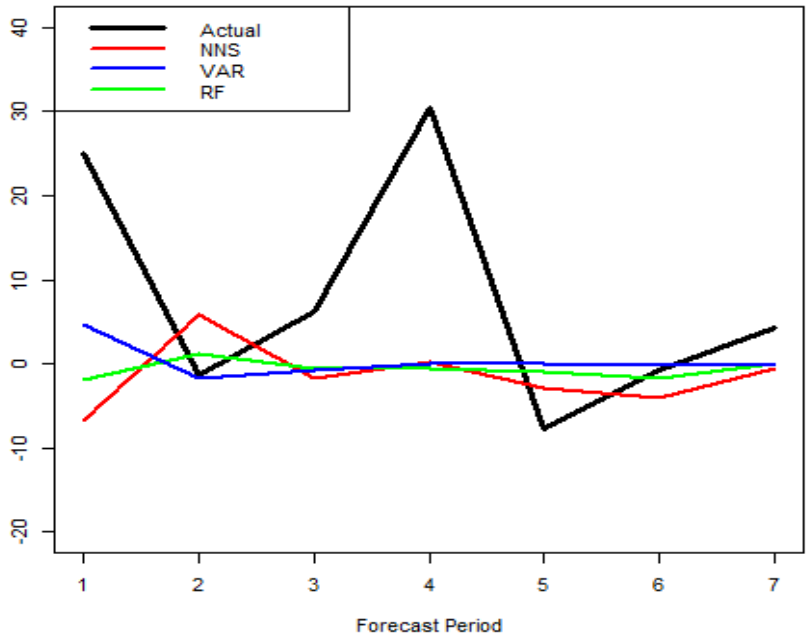

Figure 43. Plots of VAR, NNS & RF forecasts and the actual out-of-sample observations against the forecast horizon for crude oil returns

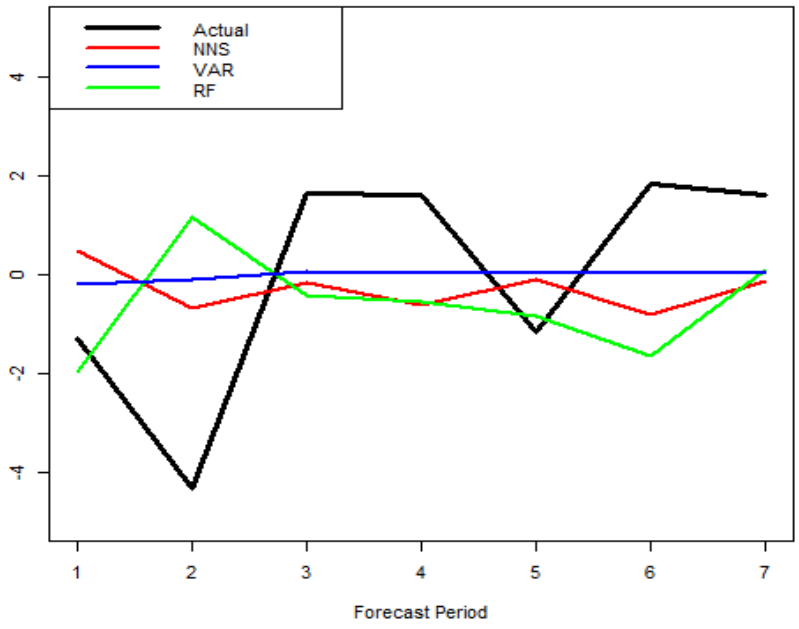

Figure 44. Plots of VAR, NNS & RF forecasts and the actual out-of-sample observations against the forecast horizon for S&P500 returns

According to the Figure 44 findings; RF for 1-day & 5-day ahead, NNS for 2-day ahead and VAR for 3-day, 4-day & 6-day ahead have generated more accurate forecasts for S&P500 returns versus other models. When the forecast horizon is 7, all three methods have presented the results that are almost very close to each other; but still the forecasts of RF and VAR are slightly closer to the actual out-of-sample observations than NNS.

Table 21

A Summary regarding Correlations and RMSE Metrics of NNS.VAR Predictions versus the Actual Values for $h = 7$

	NNS.VAR Correlations	NNS.VAR RMSEs
GOLD	-0.1071	1.1587
OIL	-0.0357	17.3844
S&P500	-0.4286	2.5195

Table 21 shows a summary of correlation coefficients and RMSE values regarding NNS.VAR forecasts versus the actual out-of-sample values for 7-day ahead. NNS.VAR represents here the nonparametric VAR model including NNS.ARMA estimates of series into a nonlinear regression based on partial moment quadrant means for a multivariate time series forecast (Viole, 2020, pp. 38, 51). According to the evidence given in Table 21, the smallest RMSE value (1.1587) implies that choosing gold returns as the dependent variable among the three variables produces forecasts that are closer to the actual values. Results presented here (NNS.VAR) for GOLD and S&P500 are better than RF results given in Table 20 (since RMSEs for NNS.VAR have taken smaller values than the ones for RF). On the other hand, when all variables are taken into consideration, NNS.VAR model has revealed better results than NNS.ARMA model given in Table 20. Besides, it is remarkable to say that the correlation coefficient of the traditional VAR forecasts versus the actual values given in Table 18 had been found to be negative for gold returns and now this negative correlation has been detected only in NNS.VAR & NNS.ARMA models.

3.5. Concluding Remarks

Volatility increments in gold prices create an unsafe investment environment, while the drops in volatility are a sign of safety in terms of investments that will be made in financial markets (Baur, 2012). Thus, increases in the volatility of gold will subject

investors to risk and therefore perceiving volatilities regarding gold prices is of great importance with respect to giving us an insight about the complete economy (Gokmenoglu & Fazlollahi, 2015, p. 479). Two of the important variables that affect gold are crude oil and stock market index. Therefore, in this study, the question of which of these three variables gives a better forecast in the case of being selected as the dependent variable in comparison with the others has been tried to be replied by comparing the accuracies of various methods in which basically the nonparametric VAR model has been considered. Because of having the smallest RMSE values for all models; however, its more logical to use gold returns as the dependent variable against others. Amongst all models, gold returns forecasts have the highest correlation magnitude against the actual values in NNS model with a value of 0.5714 (given in Table 20). Thus, it can be inferred that NNS model should be preferred in explaining gold returns (with the minimum RMSE value of 0.9760). In this research, it can be said how important it is to treat variables as nonlinear-nonparametric. The negative sign of correlation coefficient observed in the traditional VAR for gold returns has also been come across in NNS.VAR model. If necessary to interpret the plots, the main conclusion is that the effective method for having more accurate forecasts differs as the number of horizons changes. Besides, NNS has presented better results relative to NNS.ARMA for all three series in Table 20. To sum up; it can be inferred from this paper that the best methods for 7-day ahead forecasting of gold, crude oil and S&P500 (in case all these variables are each regarded as the dependent variable seperately) in a way that will be close to the actual out-of-sample observations have been determined as NNS, VAR and ARMA respectively (thus, NNS has been superior to the traditional VAR model in forecasting of gold returns).

REFERENCES

Akgul, I., Bildirici, M., & Ozdemir, S. (2015). Evaluating the nonlinear linkage between gold prices and stock market index using Markov-switching bayesian VAR models. *Procedia-Social and Behavioral Sciences*, *210*, 408-415.

Al-Ameer, M., Hammad, W., Ismail, A., & Hamdan, A. (2018). The relationship of gold price with the stock market: The case of Frankfurt Stock Exchange. *International Journal of Energy Economics and Policy*, *8*(5), 357-371.

Andrews, D. (1983). *First order autoregressive processes and strong mixing* (No. 664). Cowles Foundation for Research in Economics, Yale University.

Araveeporn, A. (2011). Developing nonparametric conditional heteroscedastic autoregressive nonlinear model by using maximum likelihood method. *Chiang Mai J. Sci*, *38*(3), 331-345.

Ayyadevara, V.K. (2018). *Pro machine learning algorithms: A hands-on approach to implementing algorithms in Python and R*. New York: Apress.

Balcilar, M., Saint Akadiri, S., Gupta, R., & Miller, S. M. (2019). Partisan conflict and income inequality in the united states: A nonparametric causality-in-quantiles approach. *Social Indicators Research*, *142*(1), 65-82.

Baur, D. G. (2012). Asymmetric volatility in the gold market. *The Journal of Alternative Investments*, *14*(4), 26-38.

Billio, M., Casarin, R., & Rossini, L. (2019). Bayesian nonparametric sparse VAR models. *Journal of Econometrics*, *212*(1), 97-115.

Bjørnland, H. C. (2000). VAR models in macroeconomic research. *Statistics Norway*, *14*, 1-29.

Bossaerts, P., Härdle, W., & Hafner, C. (1996). Foreign exchange rates have surprising volatility. In P. M. Robinson & M. Rosenblatt (Eds.), *Athens Conference on Applied Probability and Time Series Analysis* (pp. 55-72). New York: Springer.

Brauchler, R. (2011). *Multivariate nonparametric estimation of value at risk and expected shortfall for nonlinear returns using extreme value theory.* Available at SSRN: https://ssrn.com/abstract=2147760 or http://dx.doi.org/10.2139/ssrn.2147760.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5-32.

Chittedi, K. R. (2012). Do oil prices matters for Indian stock markets? An empirical analysis. *Journal of Applied Economics and Business Research*, *2*(1), 2-10.

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, *74*(368), 829-836.

Cutler, A., Cutler, D. R., Stevens, J. R. (2012). Random forests. In C. Zhang & Y. Ma (Eds.), *Ensemble Machine Learning: Methods and Applications* (pp. 157-175). Boston, MA: Springer Science Business Media.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the Econometric Society*, *50*(4), 987-1007.

Fan, J. (1993). Local linear regression smoothers and their minimax efficiencies. *The Annals of Statistics*, *21*(1), 196-216.

Fan, J., Gasser, T., Gijbels, I., Brockmann, M., & Engel, J. (1997). Local polynomial regression: optimal kernels and asymptotic minimax efficiency. *Annals of the Institute of Statistical Mathematics*, *49*(1), 79-99.

Fan, J., & Gijbels, I. (1996). *Local polynomial modelling and its applications: monographs on statistics and applied probability 66* (Vol. 66). London: Chapman & Hall, CRC Press.

Fang, K., Li, R. Z., & Sudjianto, A. (2006). *Design and modeling for computer experiments* (*Computer science and data analysis series)*. Chapman & Hall/CRC press.

Gilbert, R. J., & Mork, K. A. (1984). Will oil markets tighten again? A survey of policies to manage possible oil supply disruptions. *Journal of Policy Modeling*, *6*(1), 111-142.

Gokmenoglu, K. K., & Fazlollahi, N. (2015). The interactions among gold, oil, and stock market: evidence from S&P500. *Procedia Economics and Finance*, *25*(Supp. C), 478-488.

Gourieroux, C., & Monfort, A. (1992). Qualitative threshold ARCH models. *Journal of Econometrics*, *52*(1-2), 159-199.

Grobys, K. (2012). A non-parametric approach of heteroskedasticity robust estimation of Vector-Autoregressive (VAR) models. *Journal of Finance and Investment Analysis*, *1*(1), 55-67.

Hafner, C. (1998). *Nonlinear time series analysis with applications to foreign exchange rate volatility*. Springer-Verlag Berlin Heidelberg.

Hafner, C., & Herwartz, H. (2002). *Testing for vector autoregressive dynamics under heteroskedasticity* (EI No. 2002-36). Erasmus University Rotterdam, Erasmus School of Economics (ESE), Econometric Institute.

Härdle, W., & Tsybakov, A. (1997). Local polynomial estimators of the volatility function in nonparametric autoregression. *Journal of Econometrics*, *81*(1), 223-242.

Härdle, W., Tsybakov, A., & Yang, L. (1998). Nonparametric vector autoregression. *Journal of Statistical Planning and Inference*, *68*(2), 221-245.

Härdle, W., & Yang, L. (1996). *Nonparametric time series model selection.* Discussion paper, Humboldt-Universität zu Berlin.

Hubrich, K., & Teräsvirta, T. (2013). *Thresholds and smooth transitions in vector autoregressive models* (CREATES Research Paper No. 2013-18). Department of Economics and Business Economics, Aarhus University.

Hurvich, C. M., & Tsai, C. L. (1989). Regression and time series model selection in small samples. *Biometrika*, *76*(2), 297-307.

Jeliazkov, I. (2013). Nonparametric vector autoregressions: Specification, estimation and inference. In T. B. Fomby, L. Kilian & A. Murphy (Eds.), *VAR Models in Macroeconomics - New Developments and Applications: Essays in Honor of Christopher A. Sims*, *vol. 32* (pp. 327-359). Emerald Group Publishing Limited.

Jones, D. W., Leiby, P. N., & Paik, I. K. (2004). Oil price shocks and the macroeconomy: What has been learned since 1996. *The Energy Journal*, *25*(2), 1-32.

Julio-Román, J. M., Rodríguez-Niño, N., & Zárate-Solano, H. M. (2005). *Estimating the COP exchange rate volatility smile and the market effect of central bank interventions: a CHARN approach.* (Working Paper No. 347). Borradores Banco de la República.

Kalli, M., & Griffin, J. E. (2015). *Bayesian nonparametric vector autoregressive models.* Retrieved May 11, 2020, from https://www.nbp.pl/badania/konferencje/2015/forecasting/pdf/Kalli_paper.pdf.

Kalli, M., & Griffin, J. E. (2018). Bayesian nonparametric vector autoregressive models. *Journal of Econometrics*, *203*(2), 267-282.

Kapetanios, G., Marcellino, M., & Venditti, F. (2017). *Large time-varying parameter VARs: A non-parametric approach* (Working Paper No. 1122). Bank of Italy Temi di Discussione.

Le, T. H., & Chang, Y. (2011). *Oil and gold: Correlation or causation?* (MPRA Paper No. 31795). Munich Personal RePEc Archive.

Martins-Filho & Yao, F. (2006). Estimation of value-at-risk and expected shortfall based on nonlinear models of return dynamics and extreme value theory. *Studies in Nonlinear Dynamics and Econometrics, 10*(2), 107-49.

Masry, E., & Tjøstheim, D. (1995). Nonparametric estimation and identification of nonlinear ARCH time series strong convergence and asymptotic normality: Strong convergence and asymptotic normality. *Econometric Theory*, *11*(2), 258-289.

Nguyen, T. T., Huang, J. Z., & Nguyen, T. T. (2015). Two-level quantile regression forests for bias correction in range prediction. *Machine Learning*, *101*(1-3), 325-343.

Pagan, A. R., & Schwert, G. W. (1990). Alternative models for conditional stock volatility. *Journal of Econometrics*, *45*(1-2), 267-290.

Robinson, P. M. (1983). Nonparametric estimators for time series. *Journal of Time Series Analysis*, *4*(3), 185-207.

Robinson, P. M. (1984). Robust nonparametric autoregression. In J. Franke, W. Härdle & D. Martin (Eds.), *Robust and Nonlinear Time Series Analysis-Lecture Notes in Statistics, vol.26* (pp. 247-255). New York: Springer.

Rosenblatt, M. (1956). A central limit theorem and a strong mixing condition. *Proceedings of the National Academy of Sciences of the United States of America*, *42*(1), 43-47.

Safari, A., & Seese, D. (2009). Non-parametric estimation of a multiscale CHARN model using SVR. *Quantitative Finance*, *9*(1), 105-121.

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica: Journal of the Econometric Society*, *48*(1), 1-48.

Singh, R. S., & Ullah, A. (1985). Nonparametric time-series estimation of joint DGP, conditional DGP, and vector autoregression. *Econometric Theory*, *1*(1), 27-52.

Stone, C. J. (1977). Consistent nonparametric regression. *The Annals of Statistics*, *5*(4), 595-645.

Sugito, Mustafid, Safitri, D., Ispriyanti, D., Hakim, A. R., & Yasin, H. (2019). Rainfall and wave height prediction in Semarang city using vector autoregressive neural network (VAR-NN) methods. In *Journal of Physics: IOP Conference Series*, *vol. 1320, 012017.* IOP Publishing.

Sujit, K. S., & Kumar, B. R. (2011). Study on dynamic relationship among gold price, oil price, exchange rate and stock market returns. *International Journal of Applied Business and Economic Research*, *9*(2), 145-165.

Teräsvirta, T. (1994). Specification, estimation, and evaluation of smooth transition autoregressive models. *Journal of the American Statistical Association*, *89*(425), 208-218.

Tsybakov, A. B. (1986). Robust reconstruction of functions by the local-approximation method. *Problemy Peredachi Informatsii*, *22*(2), 69-84.

Viole, F. (2019). *Multivariate time series forecasting: Nonparametric vector autoregression using NNS.* Available at SSRN: https://ssrn.com/abstract=3489550.

Viole, F. (2020). *Package 'NNS'*. Retrieved May 31, 2020, from https://cran.r-project.org/web/packages/NNS/NNS.pdf.

Wu, H., Lin, H., Guan, Y., Harada, K., & Rojas, J. (2017). Robot introspection with bayesian nonparametric vector autoregressive hidden markov models. In *2017 IEEE-RAS 17th International Conference on Humanoid Robotics* (pp. 882-888). IEEE.

Wutsqa, D. U. (2008). The Var-NN model for multivariate time series forecasting. *MatStat*, *8*(1), 35-43.

Yang, L. (2007). Nonparametric modelling of quarterly unemployment rates. *Journal of Data Science*, *5*(1), 85-101.

Yang, L., Hardle, W., & Nielsen, J. (1999). Nonparametric autoregression with multiplicative volatility and additive mean. *Journal of Time Series Analysis*, *20*(5), 579-604.

Yang, L., & Tschernig, R. (2002). Non-and semiparametric identification of seasonal nonlinear autoregression models. *Econometric Theory*, *18*(6), 1408-1448.

Yasin, H., Warsito, B., & Santoso, R. (2018). Soft computation vector autoregressive neural network (VAR-NN) GUI-based. In *E3S Web of Conferences*, *vol. 73,. 13008*. EDP Sciences.

Zivot, E., & Wang, J. (2006). Vector autoregressive models for multivariate time series. In *Modeling Financial Time Series with S-Plus®* (pp. 385-429). NewYork: Springer.

CURRICULUM VITAE

Personal Information:

Name-Surname: Sera Şanlı

Gender: Female

Nationality: Turkish

Date of birth: October, 30, 1990

Place of birth: Krefeld/West Germany

Work address: Çukurova University, Faculty of Economics and Administrative Sciences, Department of Econometrics, Office No: 301, Sarıçam, ADANA

Telephone Number: +90 (322) 3387254-60/155, 3611/155

E-mail: sanlis@cu.edu.tr

Educational Background:

PhD: 2015-2020, Çukurova University, Institute of Social Sciences, Department of Econometrics, ADANA

Master: 2013-2015, Çukurova University, Institute of Social Sciences, Department of Econometrics, ADANA

Undergraduate: 2009-2013, Çukurova University, Faculty of Economics and Administrative Sciences, Department of Econometrics (English), ADANA

High School: 2005-2008, Istiklal Makzume Anatolian High School, Iskenderun/HATAY

2004-2005, Iskenderun Demircelik Anatolian High School (Preparatory Year), HATAY

Work Experience:

12/2013 – to date : Research Assistant, Çukurova University, F.E.A.S., Department of Econometrics, ADANA

Language Skills:

English (Advanced Level)

German (Intermediate Level)