QATAR UNIVERSITY

COLLEGE OF ARTS AND SCIENCES

# NON-DETERMINISTIC MODELING USING QUANTILE REGRESSION

BY

AMEENA H. AL ABDULLA

A Thesis Submitted to

the College of Arts and Sciences

in Partial Fulfillment of the Requirements for the Degree of

Masters of Science in Applied Statistics

June 2022

# COMMITTEE PAGE

The members of the Committee approve the Thesis of
Ameena H. Al abdulla defended on 15/05/2022.

Dr. Esam Mahdi
Thesis/Dissertation Supervisor

Dr. Faiz Elfaki
Committee Member

Dr. Hussain Al-Qassem
Committee Member

Dr. Reza Pakyari
Committee Member

Dr. Safwati Ibrahim
Committee Member

Approved:

Ahmed Elzatahry, Dean, College of Arts and Sciences

# ABSTRACT

AL ABDULLA, AMEENA, H., Masters : June : 2022, Applied Statistics

Title: Non-Deterministic Modeling Using Quantile Regression

Supervisor of Thesis: Esam, B, Mahdi.

In this thesis, we utilize quantile regression to model the conditional quantile of the dependent variable given independent variables to capture more details about the conditional distribution. In addition, we apply the quantile-on-quantile regression model to estimate the impact of an independent variable's quantiles on the conditional quantiles of the dependent variable to uncover the dependence between the independent and dependent variables.

We consider the RavenPack news-based index associated with the coronavirus outbreak (Panic, Media Hype, Fake News, Sentiment, Infodemic, and Media Coverage) and the returns of Bitcoin and gold as real-world applications. Our findings demonstrate that the bearish and bullish on Bitcoin and gold are affected by the daily positive and negative shocks in indices caused by coronavirus news asymmetrically. Sentiment induced by coronavirus plays a major role in driving Bitcoin and gold values than other indices. Bitcoin and gold can act as a hedge against coronavirus-related news.

# DEDICATION

*It is my pleasure to dedicate my master's thesis work to my family. To my great parents and siblings, who have always pushed me to achieve my ambitions. To my helpful friends for their unwavering support throughout the master program period.*

ACKNOWLEDGMENTS

Firstly, I would like to express my gratitude to my supervisor, Dr. Esam Mahdi, for taking me under his supervision and for his unlimited support, patience, and encouragement during the period of writing this thesis. Moreover, I would like to thank the committee members, Dr. Faiz Elfaki and Dr. Reza Pakyari, for their valuable comments during the preparation of this thesis.

I would like to express my heartfelt gratitude towards the faculty and staff at the Mathematics, Statistics and Physics Department and the Graduate Studies staff at Qatar University for their assistance and guidance during my graduate studies.

My grateful thanks goes to my parents for their support, encouragement, and prayers for me to complete my master's studies, and to my friends and colleagues at Qatar University during the journey of my graduate studies. A special thanks to my managers at work, Dr. Moumen Hasnah and Dr. Mahmood Ahmed, for their unwavering assistance and support.

# TABLE OF CONTENTS

# LIST OF TABLES

LIST OF FIGURES

# CHAPTER 1: INTRODUCTION

The analysis of financial data has attracted substantial attention in the field of statistics. In recent years, several researchers have used the quantile regression and quantile-on-quantile regression approaches to explore the relationship between the variables in such a data.

A quantile regression model was proposed by Koenker and Bassett (1978) to reveal the conditional quantile of the dependent variable (response), given the values of the independent variables (predictors). It has piqued the interest of several researchers, particularly when the data contain outliers (which cause inconstant variance) and the response variable has an unknown or non-Gaussian distribution. In fact, it is the best alternative candidate model that can be used when the assumptions of the linear regression model are not met. It is used to estimate the parameters through different quantiles of the dependent variable to capture more details about the conditional distribution. The parameters of the quantile regression model are estimated using linear programming methods, which propose different parameters for different quantiles.

A quantile-on-quantile regression model was suggested by Sim and Zhou (2015) to estimate the impact of an independent variable's quantiles on the conditional quantiles of the dependent variable, where this impact can be contingent on the performance of the dependent variable as well as the sign and size of the independent variable shocks. Moreover, the proposed approach shows how different quantiles of an independent variable affect different quantiles of the dependent variable, thereby showing the dependence relationship more deeply between the independent and dependent variables.

The ongoing global pandemic of coronavirus (COVID-19) outbreak has received extensive media coverage over the last two years. The panic-inducing news about the daily number of confirmed COVID-19 cases and deaths has a significant effect on financial markets in almost every country around the world, influencing investor decisions. In response to the novel pandemic, the RavenPack data science team created a live and interactive six indices (Panic, Media Hype, Fake News, Sentiment, Infodemic, and Media Coverage) by monitoring positive and negative news articles and public posts about the COVID-19 outbreak on a global and local scale (Coronavirus Media Monitor, 2021). The coronavirus-related news indices were proposed to assist data-driven professionals in assessing the effect of the coronavirus outbreak on global affairs.

In this thesis, we use quantile regression and quantile-on-quantile regression models to investigate the relationship between RavenPack COVID-19 news and Bitcoin and gold returns by modeling the quantiles of daily returns as a response variable on the news-based index associated with the coronavirus outbreak as an explanatory variable. We consider Bitcoin to be a proxy for the cryptocurrency market because it is well-known as the most prominent digital currency with the highest market capitalization. To the best of our knowledge, this is the first research to investigate the influence of indices induced by coronavirus news on Bitcoin and gold returns using quantile regression and quantile-on-quantile techniques.

The dependence between the RavenPack news-based index connected with coronavirus epidemic (Panic, Media Hype, Fake News, Sentiment, Infodemic, and Media Coverage) and the returns of Bitcoin and gold, is investigated in this study. The results reveal that the positive and negative news related to the pandemic affect the bearish and bullish on Bitcoin and gold asymmetrically. Among all the indices, sentiment plays a larger role in driving Bitcoin and gold prices. Both commodities, Bitcoin and gold, can serve as a hedge against pandemic-related news. More details about time series analysis are given in Appendix A.

For the rest of this thesis, Chapter 2 includes literature review of quantile regression model, quantile-on-quantile regression model and Bitcoin and gold returns and RavenPack coronavirus-related news indices. Chapter 3 provides the quantile at a specific level $\tau$, the procedure for estimating quantile regression and quantile-on-quantile regression coefficients. The procedure of checking the validity of quantile-on-quantile regression model as well. In addition, it studies the local linear quantile regression model and the Jarque-Bera test. Chapter 4 performs the descriptive statistics of the variables and reports the findings based on the quantile regression and quantile-on-quantile regression models that are used to reveal the association between coronavirus pandemic-related news and Bitcoin and gold returns. To sum up, Chapter 5 gives an overview of the entire thesis and proposes some ideas for future work.

# CHAPTER 2: LITERATURE REVIEW

## 2.1 Quantile Regression (QR) Model

Quantile regression has been widely used in recent years for research in several domains, including ecology, financial economics, and healthcare. It was introduced by Koenker and Bassett (1978) to explore the relationship between the predictors and the conditional quantiles of the response variables, by examining the results obtained from both the mean regression and quantile regression models, to obtain beneficial information from the data. The quantile regression model has two major benefits compared to the ordinary least squares' regression model: it does not impose any assumptions about the target variable distribution and it can resist the effect of extreme observations.

Koenker and Bassett (1982) suggested a new method to estimate linear models for the conditional quantile functions. Their proposed approach omits the parametric assumptions on the error distribution and applied the regression quantile model suggested previously by Koenker and Bassett (1978). In addition, they proposed tests for heteroscedasticity using a large sample approach of the proposed estimators. These methods are meant to be used as diagnostic tools in linear model applications.

Koenker and Bassett (1986) showed the robust consistency of regression quantile models in the applications of linear models with independent and identically distributed (iid) errors. Significantly, moderate regularity conditions were required on the distribution of the errors and the regression design sequence. In addition, the robust consistency of the related empirical quantile procedure was demonstrated in similar circumstances. Alternatively, the estimate of the conditional distribution function without any required regularity constraints on the dispersal of the errors for the uniform strong convergence, therefore using Glivenko-Cantelli-type theorem was nominated for his estimator.

Hendricks and Koenker (1992) proposed new approaches based on the quantile regression model to estimate nonparametric models for the conditional quantiles, that were proposed previously by Koenker and Bassett (1978). In addition, the spline parameterizations method of the conditional quantile functions was employed. Their proposed approaches were applied to the data obtained from the Chicago metropolitan area to estimate hierarchical models for the domestic demand of electricity. The primary results showed that the lower quantiles of base load demand alter a bit across domestic housing types, where this variability is hard to clarify using housing characteristics. On the other hand, the upper quantiles of demand alter significantly within residential households, and they are systematically associated with household features and proprietary appliances. Moreover, the analysis of mean demand behavior was implemented instead of different quantiles of demand distribution.

Koenker and Park (1996) suggested a developed algorithm method to estimate the quantile regression estimators when the model includes a nonlinear response function, it is a special case of the median. It can be illustrated as an alternative to the famous iteratively reweighted least squares. Significantly, the effectiveness of the proposed algorithm has been examined on plenty of test problems, in addition to the censored linear quantile regression models.

Koenker and Machado (1999) proposed the goodness of fit procedure for the quantile regression model. It is similar to the conventional coefficient of determination statistic $R^2$ of the ordinary least squares' regression model. Significantly, some inferences were proposed to examine the composite hypotheses on the combined impact of some explanatory variables through the conditional quantile functions. The behavior of the proposed inferences is very related to the previous p-sample goodness of fit in the theory of the Bessel procedure. Some hypothetical examples were illustrated using their proposed approach. An application based on the earliest empirical models of international economic growth was presented.

Koenker and Bilias (2002) showed that quantile regression can significantly help in analyzing duration (survival) data. It can concentrate on a particular region of the conditional duration distribution. In addition, the explanatory variables are presumed to have a clear location-shift influence. The methods were illustrated using the Pennsylvania Reemployment Bonus Experiment. This study showed that Pennsylvania bounces minimize the median period of joblessness by approximately 10 to 15%. However, this influence was reduced-off from the median and was insignificant for the lower and upper tails.

He et al. (2003) took longitudinal data and applied a median regression; they constructed three estimators based on weighting, decorrelating, and the assumption of independence. The performance of the proposed estimators has been evaluated using asymptotic efficiency computations, as well as finite sample Monte Carlo. The results showed that the nature of covariates affects the performance of the estimators. Specifically, the performance of the estimators under the assumption of independence is relatively good when the independent variables are constant in terms of time, or when the correlations within the subject are minor. Moreover, its accomplishment is more favorable using finite samples Monte Carlo than the asymptotic efficiency calculations.

Koenker and Xiao (2006) provided a quantile autoregressive model to illustrate the conditional effects of the factors on the location and scale of the response distribution. Their proposed model is an extension of the conventional linear time series models where the conditioning effect is restricted to the location shift. Their model can be construed as a special form of the parameter autoregressive model with highly dependent parameters. It seeks to handle this constraint and to consider the linear quantile autoregression model, whose coefficients may differ with different levels of quantiles. It studies asymmetric dynamics and local persistency in the analysis of time series.

Yu and Stander (2007) employed the application of Bayesian on Tobit quantile regression model using likelihood function depending on asymmetric Laplace distribution. The advantage of their approach is to avoid solving the problem of non-convex minimization and the density estimation problem. A group of prior distributions as was applied to a quantile regression vector, which leads to an appropriate posterior distribution with a finite moment. In addition, their paper illustrated how Markov Chain Monte Carlo approaches can be applied to sample and summarize the posterior distribution. Moreover, a method was provided to compare substitutional quantile regression models. The ideas were presented using simulation and real data application. As an empirical comparison, the proposed approach out-performed two other common classical estimators.

Xue et al. (2016) discussed the importance of the censored quantile regression method for analyzing time to event data. The application of censored quantile regression on clinical research is limited because of the difficulties in interpreting its results, and its advantages are hard to appreciate compared to the Cox proportional hazards approach, and the lack of an adequate validation procedure. The target of their study is to address the provided limitations by simulation and application on the data of the National Wilms' Tumor clinical trials and propose a validation procedure for the predicted censored quantile regression approach.

Lemonte and Moreno-Arenas (2020) provided a parametric quantile regression method for a finite set of response variables using the distribution of two-parameter heavy-tailed, to model the bounded response variables through diverse quantiles with the availability of atypical observations. The procedure of the frequentist was applied to implement inferences. Moreover, the maximum likelihood methodology was used to obtain the estimators, and the residual analysis was proposed to determine departures gained from model assumptions. In addition, the method of local effect as was discussed, also under a particular perturbation scheme, the normal curvature was employed to study the local effect on the estimates obtained from the maximum likelihood method.

### 2.2 Quantile-on-Quantile Regression (QQR) Model

Sim and Zhou (2015) suggested an original model of quantile-on-quantile regression to study the link between stock returns in the United States and oil prices, it is used to obtain the estimates of the impact of the shock due to oil price quantiles and the United States stock return quantiles. Their methodology demonstrates the dependence in the distributions of oil prices and the United States stock returns; additionally, it identifies two features of the oil stock relationship. The results showed that negative shocks in the oil price (at the low quantiles of the oil price shock) have a positive impact on the United States equities when the United States market performs well (at the high quantiles of the US return). On the other hand, the positive shocks in the oil price had a feeble effect on the United States stock returns, which indicates an asymmetric relationship between United States equities and oil prices.

The approach of quantile-on-quantile regression was employed by Shahzad et al. (2017) to study the experimental validity of the assumption that growth is caused by tourism for the top ten tourist countries in the world over twenty-five years (from 1990 to 2015). These countries are France, Germany, the UK, Italy, the US, Spain, Turkey, Mexico, Russia, and China. The proposed approach describes the organized dependency between tourism improvement and economic growth. Obviously, the primary outcomes showed there was a positive relationship between economic growth and tourism development for all considered countries, but there is a significant variation between countries and between quantiles within each country. In particular, both countries Germany and China had a minimal link between tourism improvement and the growth of the economy, the possible motive might be due to less attention to the tourism sector compared to other economic activities.

### 2.3 Bitcoin and Gold Returns and RavenPack Coronavirus-Related News Indices

The impact of negative news on investor mood is much stronger than that of positive news, as the reaction to negative and positive news is asymmetric (Soroka, 2006). Finding a safe commodity investment in uncertain situations is a matter of course for most investors. There is a growing body of recent financial literature on the use of the quantile-on-quantile regression model in applications related to cryptocurrencies, gold, oil, and other financial markets and assets. For instance, Bouri et al. (2017) used two approaches based on quantile regression and quantile-on-quantile regression, to demonstrate that Bitcoin can perform as a hedge against global uncertainty over shorter investment periods.

Urquhart (2018) researched investor attention using Bitcoin Trend Search on Google Data. He found that Bitcoin's attention the day before did not affect Bitcoin's volatility or return forecasts. In fact, the opposite is true. Meanwhile, he showed that the high realization volatility and returns of Bitcoin over the past few days can have a significant impact on investors' attention to Bitcoin. Klein et al. (2018), and Smales (2019) have shown that Bitcoin is more volatile than other assets and should not be treated as a potentially safe haven compared to gold. Shahzad et al. (2019) considered, in some cases, that gold and Bitcoin could be classified as weak and safe havens at best.

Broadstock and Zhang (2019) showed that economic and political news, particularly on social media, can have a significant impact on stock price movements. They developed a measure of emotions using Twitter's social media messages and showed that emotions have the power to determine the price of equity returns for some U.S. companies. Chen et al. (2020) used an hourly Google search for COVID19-related words to investigate the influence of fear induced by the outbreak of COVID-19 on Bitcoin price fluctuations. They concluded that, during the COVID-19 pandemic Bitcoin failed to perform as a safe haven.

Cepoi (2020) designed a model comprising lower, middle, and upper quantiles of stock return on COVID-19 RavenPack news from six pandemic-affected countries (the US, UK, Germany, France, Spain, and Italy). The researcher used a quantile regression model and pointed out that fake news has a negative non-linear effect on the lower and middle quantiles of earnings, and the gold does not serve as a safe haven when COVID-19 hits. Haroon and Rizvi (2020) showed that panic-laden caused by coronavirus news could lead to a significant increase in stock market volatility, which is believed to be severely affected by the pandemic. Mnif et al. (2020) discussed the fluctuations of the top five cryptocurrencies market during the crisis of the coronavirus. The primary results showed that the coronavirus crisis affected the efficiency of the cryptocurrency market positively. In addition, the index of the magnitude of long memory demonstrated that Bitcoin behaves more efficiently than the other cryptocurrencies before the coronavirus outbreak, but after the epidemic, it becomes less efficient than Ethereum. However, all of the cryptocurrencies under the study behave more efficiently after the crisis of the coronavirus.

Shi and Ho (2021) used RavenPack's Dow Jones News Analysis Database to explore the influence of news sentiment on changes in Dow Jones stock volatility. They proposed a Markov regime-switching fractionally integrated exponential GARCH (MRS-FIEGARCH) model, showing that negative news sentiment significantly increased the volatility of equity returns within the day (intraday). Bouri et al. (2021) found that the newspaper-based Infectious Disease Uncertainty Index could improve the prediction of gold realization variability from short-term, medium-term, and long-term perspectives. Sun et al. (2021) found that economic announcements and news related to the coronavirus did not cause inappropriate investment behavior.

Several research studies have suggested that gold and digital cryptocurrencies, especially Bitcoin, are classified as strategic commodity evidence assets during stress (Dyhrberg, 2016; Urquhart & Zhang, 2019; Mariana et al., 2021). In this regard, Mahdi et al. (2021) provided a new method based on a support vector machine (SVM) algorithm for predicting cryptocurrency prices according to the quantile of gold prices during the coronavirus outbreak.

Banerjee et al. (2021) investigated the effects of several categories of coronavirus news sentiment on cryptocurrency returns using the top 30 cryptocurrencies using nonlinear technique. The results showed a positive nonlinear relationship between the top 30 cryptocurrencies' returns and COVID–19 news sentiment. Most cryptocurrency returns are influenced by media portrayals of COVID–19 circumstances. Their study showed that the benefits of adopting cryptocurrency might be reduced because prices reflect significant volatility during the epidemic period.

Kakinaka and Umeno (2021) showed that the coronavirus pandemic improves short-term, but not long-term, investor crowd behavior. However, Mnif and Jarboui (2021) observed that the COVID-19 pandemic reduced herd prejudice. Bitcoin and gold have always been very volatile commodities, as in most cases of financial data (Kim et al., 2020; Kim et al., 2021).

Iqbal et al. (2021) applied the quantile-on-quantile regression to model the asymmetric nexus among the highly adopted 10 widely cryptocurrencies and the COVID-19. They showed that the relationship between the top ten cryptocurrencies' revenues and the COVID-19 pandemic is not symmetric. The majority of the cryptocurrencies in their research ingest the tiny trauma of COVID-19 by gaining profitable returns; however, they were unable to stand in front of massive shocks at the exception of the cryptocurrency Bitcoin, Cardano, Crypto.com Coin, and up to some degree Ethereum. The results revealed that the quantile-on-quantile regression approach provides similar results as using the quantile regression model.

Table 2.1 and Table 2.2 summarize the reviews of the quantile regression model, quantile-on-quantile regression model, and the independent variables on coronavirus pandemic-related news indices, respectively.

.

Table 2.1. Summary of the Reviews of the Quantile and Quantile-on-Quantile Regression Models

| Authors | Data/Factors | Model | Distribution | Findings |
|---|---|---|---|---|
| Koenker and Bassett (1978) | | Quantile Regression | Gaussian and Non-Gaussian | They provided a novel approach to the linear model, which they called regression quantiles. |
| Koenker and Bassett (1982) | 1- Engel's Food Expenditure Data:<br>y: Annual Food Expenditure.<br>x: Annual Household Income.<br>2- Demand for Admen Data:<br>y: Number of Employees.<br>x: Annual U.S. Billings. | Quantile Regression | | They proposed a new approach to estimate linear models for the conditional quantile functions. They proposed tests for heteroscedasticity using a large sample approach of the proposed estimators. |
| Koenker and Bassett (1986) | | Quantile Regression | Moderate regularity conditions were required on the distribution of the errors and the regression design sequence | They illustrated the strong consistency of the regression quantile models in linear approaches with independent and identically distributed errors. |
| Hendricks and Koenker (1992) | Household Electricity Demand Data Obtained from Chicago Metropolitan Area. | Quantile Regression | Nonparametric | They suggested approaches to estimate nonparametric models for the conditional quantiles. The application showed that lower quantiles of base load demand vary a bit across domestic housing types. While, the upper quantiles of demand alter significantly within residential households. |
| Koenker and Park (1996) | | Quantile Regression | | They provided a developed algorithm approach to estimate the quantile regression estimators when the model contains a nonlinear response function. |

| Authors | Data/Factors | Model | Distribution | Findings |
| --- | --- | --- | --- | --- |
| Koenker and Machado (1999) | International Economic Growth Data:<br>y: GDP Growth.<br>x: Initial Per-Capita GDP, Male Secondary Education, Life Expectancy, Public Consumption/GDP, Female Secondary Education, Human Capital, Black Market Premium, Female Higher Education, Education/GDP, Political Instability, Male Higher Education, Investment/GDP, and Growth Rate Terms Trade. | Quantile Regression | Gaussian Mixture | They proposed a goodness of fit procedure for the quantile regression model. It is analogous to the coefficient of determination of the ordinary least squares' regression model. |
| Koenker and Bilias (2002) | Pennsylvania Reemployment Bonus Experiment. | Quantile Regression | | They illustrated how quantile regression can be applied to analyze duration (survival) data. The application showed that Pennsylvania bounces minimizes the median period of joblessness by approximately 10 to 15%. |
| He et al. (2003) | 1- Labor Pain Data:<br>$y_{it}$: The amount of pain for the $i^{th}$ individual at period $t$.<br>$x_i$: Treatment indicator that takes 1 for the placebo and 0 for the treatment.<br>2- Weight-Lifting Data: | Median Regression | | They applied median regression to the longitudinal data. They constructed three estimators based on weighting, decorrelating, and the assumption of independence. |

| Authors | Data/Factors | Model | Distribution | Findings |
|---|---|---|---|---|
| | $y_{it}$: The strength for the $i^{th}$ person at period $t$.<br>$x_{i1}$: Treatment indicator that takes 1 for a program (RI) and 0 otherwise.<br>$x_{i2}$: Treatment indicator that takes 1 for a program (WI) and 0 otherwise. | | | |
| Koenker and Xiao (2006) | 1- Unemployment Rates in the US Data.<br>2- US Retail Gasoline Price Dynamics Data. | Quantile Autoregression | | They proposed a quantile autoregressive model to show the conditional effects of the covariates on the location and scale of the response distribution. |
| Yu and Stander (2007) | Women's Labor Force Participation of Mroz (1987) Data. | Tobit Quantile Regression | Asymmetric Laplace | The proposed approach avoids solving the problem of non-convex minimization and density estimation. Some prior distributions were applied to a quantile regression vector, which led to an appropriate posterior distribution with a finite moment. |
| Sim and Zhou (2015) | y: Stock Returns in the United States.<br>x: Oil Prices. | Quantile-on-Quantile Regression | Nonparametric | They suggested an original model of a quantile-on-quantile regression model. The negative shocks in the oil price have a positive impact on the US equities when the US market performs well. However, the positive shocks in the oil price had a feeble effect on the US stock returns. |

| Authors | Data/Factors | Model | Distribution | Findings |
|---|---|---|---|---|
| Xue et al. (2016) | The National Wilms' Tumor Clinical Trials Data. | Censored Quantile Regression | For the simulation studies, generated time-to-event data using a piecewise exponential distribution and generated a censoring time using an independent uniform distribution | They provided the analysis of time-to-event data based on the censored quantile regression method. They showed how to address the limitations of censored quantile regression model using simulation and application. |
| Shahzad et al. (2017) | y: Gross Domestic Product.<br>x: Tourism Activity. | Quantile-on-Quantile Regression | Nonparametric | They study the validity of the assumption that the growth is caused by tourism in the top ten tourist countries in the world. They showed there was a positive relationship between economic growth and tourism development for all considered countries. |
| Bouri et al. (2017) | y: Bitcoin Returns.<br>x: VIX. | Quantile Regression and Quantile-on-Quantile Regression | | They concluded that the Bitcoin can perform as a hedge against global uncertainty over shorter investment periods. |
| Lemonte and Moreno-Arenas (2020) | Peruvian General Election In 2006:<br>y: Proportion of Blank Votes.<br>x: Human Development Index (HDI). | Quantile Regression | t-distribution | They proposed a parametric quantile regression model for a finite set of response variables based on the distribution of two-parameter heavy-tailed. |

Table 2.2. Summary of the Reviews of the Independent Variables on Coronavirus Pandemic-Related News Indices

| Authors | Data/Factors | Model | Findings |
|---|---|---|---|
| Broadstock and Zhang (2019) | y: Stock Price.<br>x: Social Media (Twitter) Messages. | Capital Asset Pricing | They showed that emotions have the power to determine the price of equity returns for some U.S. companies. |
| Chen et al. (2020) | y: Bitcoin Price.<br>x: Hourly Google Search for COVID19-Related Words. | Vector Autoregressive | They showed that during the COVID-19 pandemic Bitcoin failed to perform as a safe haven. |
| Cepoi (2020) | y: Stock Return.<br>x: RavenPack News from Six Pandemic-Affected Countries (US, UK, Germany, France, Spain, and Italy). | Quantile Regression | They illustrated that fake news has a negative non-linear effect on the lower and middle quantiles of earnings, and gold does not serve as a safe haven when COVID-19 hits. |
| Haroon and Rizvi (2020) | y: Stock Market.<br>x: Panic-Laden Caused by Coronavirus News. | GARCH | They concluded that the stock market was severely affected by the pandemic. |
| Shi and Ho (2021) | RavenPack's Dow Jones News Analysis Database:<br>y: Dow Jones Stock.<br>x: News Sentiment. | Markov Regime-Switching Fractionally Integrated Exponential GARCH | They showed that the negative news sentiment significantly increased the volatility of equity returns within the day. |
| Bouri et al. (2021) | y: Gold Returns.<br>x: Newspaper-Based Infectious Disease Uncertainty Index. | Heterogeneous Autoregressive Realized Variance | They found that the newspaper-based Infectious Disease Uncertainty Index could improve the prediction of gold realization variability. |
| Sun et al. (2021) | y: Medical Stock Portfolios.<br>x: Coronavirus Related News, and Economic Related Announcements. | Regression | They showed that the economic announcements and news related to the coronavirus did not cause inappropriate investment behavior. |
| Banerjee et al. (2021) | y: Top 30 Cryptocurrencies.<br>x: Panic, Media Hype, Fake News, Infodemic, Sentiment, and Media Coverage Indices. | Nonlinear | The benefits of adopting cryptocurrency might be reduced because prices reflect significant volatility during the epidemic period. |

| Authors | Data/Factors | Model | Findings |
|---|---|---|---|
| Iqbal et al. (2021) | y: Top 10 Cryptocurrencies.<br>x: COVID-19 Cases and COVID-19 Deaths. | Quantile Regression and Quantile-on-Quantile Regression | They showed that the majority of cryptocurrencies ingest the tiny trauma of COVID-19 by gaining profitable returns. However, they were unable to stand in front of massive shocks, with the exception of Bitcoin, Cardano, Crypto.com Coin, and up to some degree Ethereum. |

# CHAPTER 3: METHODOLOGY

## 3.1 Quantiles

Quantiles are points in a distribution, each corresponding to the rank order of its values in that distribution (Frey, 2018). The most common quantiles are 25%, 50%, and 75%. For instance, the 75% quantile means that three-quarters of the sorted points are less than or equal to it, and one-quarter of the sorted data is above that quantile.

Consider $X$ is a random variable with a cumulative distribution function (CDF) of $F(x) = P(X \leq x)$, then the $\tau^{th}$quantile of $X$ is referred to as:

$$Q(\tau) = F^{-1}(\tau) = inf\{x: F(x) \geq \tau\}, \quad (1)$$

where $\tau$ is the quantile level, and its possible values belong to the interval (0,1). The definition of the asymmetric quantile loss function is as follows:

$$\rho_\tau(u) = u(\tau - I(u < 0)), \quad (2)$$

where $\rho$ (.) is the quantile loss function, and $I(.)$ is the indicator function. Figure 3.1 demonstrates the loss function depicted by the linear function (Koenker, 2005).

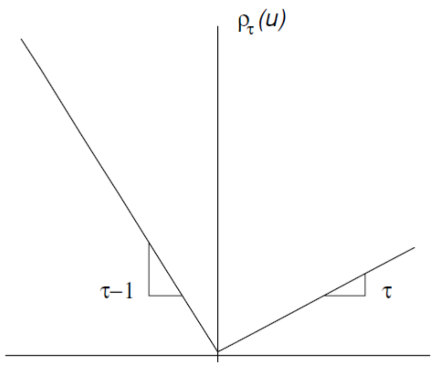


Figure 3.1. The quantile regression loss function

The procedure for estimating quantiles is given in Appendix B.

### 3.2 Quantile Regression Model

The quantile regression model is defined as:

$$y_i = \boldsymbol{x}_i^T \boldsymbol{\beta}(\tau) + \varepsilon_i(\tau)$$
$$= \beta_0(\tau) + \beta_1(\tau)x_1 + \ldots + \beta_k(\tau)x_k + \varepsilon_i(\tau), \qquad (3)$$

where $\boldsymbol{x}_i^T = [1, x_1, \ldots, x_k]$ are independent variables, $\boldsymbol{\beta}(\tau) = [\beta_0(\tau), \beta_1(\tau), \ldots, \beta_k(\tau)]^T$ are $(k+1)$ quantile regression coefficients, which describes the change in the $\tau^{th}$ quantile of the dependent variable with respect to the independent variable, and $\tau \in (0,1)$.

The conditional quantile function in accordance with the $\tau^{th}$ quantile is given through the following equation:

$$Q_y(\tau|\boldsymbol{x}) = \boldsymbol{x}_i^T \boldsymbol{\beta}(\tau), \qquad (4)$$

the parameter vector $\boldsymbol{\beta}(\tau)$ estimated by solving the following optimization problem:

$$\hat{\beta}(\tau) = \underset{\{\beta \in R^{k+1}\}}{\operatorname{argmin}} \sum_{i=1}^{n} \rho_\tau(y_i - \boldsymbol{x}_i^T \boldsymbol{\beta}), \qquad (5)$$

where $\rho_\tau(.)$ is the quantile loss function illustrated in Equation (2).

The quantile regression model in Equation (3) can be seen as:

$$y_i = \boldsymbol{x}_i^T \boldsymbol{\beta}(\tau) + (u_i - v_i), \qquad (6)$$

by proposing the $2n$ artificial variables $u_i$, $v_i$, where $u_i = \varepsilon_i I(\varepsilon_i > 0)$ and $v_i = |\varepsilon_i| I(\varepsilon_i < 0)$ for $i = 1, \ldots, n$ (Koenker, 2005). The residual vector is divided into positive and negative parts. As a result, Equation (5) becomes:

$$\min_{\{\beta \in R^{k+1}\}} \tau \boldsymbol{1}_n^T \boldsymbol{u} + (1-\tau) \boldsymbol{1}_n^T \boldsymbol{v}, \qquad (7)$$

subject to $$y - \boldsymbol{X}^T\boldsymbol{\beta} = \boldsymbol{u} - \boldsymbol{v}, \qquad for\, u\, and\, v\, \geq 0$$

where $\boldsymbol{X}$ is a design matrix of $n \times (k+1)$, $\boldsymbol{u}$ and $\boldsymbol{v}$ are vectors of $n \times 1$ with elements of $u_i, v_i,\ i = 1, \dots, n$. The terms $\tau\mathbf{1}_n^T\boldsymbol{u}$ and $(1-\tau)\mathbf{1}_n^T\boldsymbol{v}$ are to penalize for under-prediction and over-prediction, respectively (Marasinghe, 2014). Equation (7) can be solved using linear programming methods, such as the Simplex algorithm, which was suggested by Barrodale and Roberts (1974) and adapted by Koenker and d'Orey (1987). It is the default option in many software, including RStudio, and it is usually preferred among linear programming methods.

### 3.3 Local Linear Quantile Regression

Local linear quantile regression is a non-parametric approach for smoothing quantile regression curves and for estimating the derivatives of a specific estimate. The local regression and the quantile regression together summarize information about the smooth quantile curves. Chaudhuri (1991) proposed the asymptotic behavior of regression quantiles, and Chaudhuri et al. (1997) then used these findings to estimate the average derivatives on the local quantile regression. Based on that, consider that the sample $(x_i, y_i), i = 1, \dots, n$, adopts the following model:

$$y_i = m_\tau(x_i) + \varepsilon_i(\tau), \tag{8}$$

where $m_\tau(.)$ is an unknown function and $x$ is a uni-dimensional covariate. The Taylor expansion in the neighborhood of $x$ can be used to approximate the quantile function $m_\tau(x)$ locally with a polynomial:

$$m_\tau(x_i) \approx \sum_{j=1}^{k} \frac{m_\tau^j(x)}{j!} (x_i - x)^j \equiv \widetilde{\boldsymbol{X}}_i^T \boldsymbol{\beta}_\tau, \tag{9}$$

where $m_\tau^j$ is the $j^{th}$derivative of $m_\tau$, $\widetilde{\boldsymbol{X}}_i^T = (1, (x_i - x), (x_i - x)^2, \dots, (x_i - x)^k)$, and $\boldsymbol{\beta}_\tau = (\beta_{0\tau}, \beta_{1\tau}, \dots, \beta_{k\tau})$. Consequently, the function can be estimated by

$$\widehat{m}_\tau(x) = \hat{\beta}_{0\tau}, \tag{10}$$

then, the first derivative is calculated as follows:

$$\widehat{m}'_\tau(x) = \hat{\beta}_{1\tau}. \tag{11}$$

The estimates of the local polynomial quantile regression $\boldsymbol{\beta}_\tau$ are computed using the weighted objective function:

$$\underset{\{\beta \in R^{k+1}\}}{\operatorname{argmin}} \sum_{i=1}^{n} w_i(x)\rho_\tau(y_i - \widetilde{\boldsymbol{X}}_i^T \boldsymbol{\beta}), \tag{12}$$

where $w_i(x) = K((x_i - x)/h)$, $K$ is the bounded kernel function, and $h$ is the bandwidth parameter.

### *3.3.1 Bandwidth Selection*

Bandwidth is the degree of smoothness of the curve. In non-parametric regression, selecting the best bandwidth is crucial. For the non-parametric mean regression, there are numerous approaches for bandwidth selection, including plug-in, rule-of-thumbs, and cross-validation. These approaches obtain the asymptotic optimal bandwidth by minimizing the mean integrated squared error, or the mean square error. The conventional techniques for selecting the bandwidth extend to the quantile regression field.

Then, adjusted the cross-validation approach to the kernel quantile regression by replacing the squared loss criterion with the quantile loss function obtained in Equation (2), using the following formula:

$$CV(h) = \sum_{i=1}^{n} \rho_\tau(Y_i - Q_n^{(-i)}(\tau|x_i)), \tag{13}$$

where $Q_n^{(-i)}(\tau|x_i)$ is an estimator for the conditional quantile estimate $Q_n(\tau|x_i)$ obtained in Equation (4). The cross-validation method has a poor relative convergence rate: $O(n^{-1/10})$ (Loader, 1999).

Yu and Jones (1998) provided the rule of thumb based on the concept of the plug-in approach to select the regression quantile smoothing parameters by minimizing the local linear quantile function in accordance with Equation (12), and with $k = 1$. Consider $f$ as a marginal density of $X$, $Q(\tau|\mathrm{x})$ is the conditional quantile estimate, and $g(H(Y)|X = x)$ is the conditional density of the function $H(Y)$ based on $\tau$, then the optimal bandwidth would be given through the following equation:

$$h_\tau^5 = \frac{R(K)\tau(1-\tau)}{n\mu_2(K)^2 Q''(\tau|\mathrm{x})^2 f(x) g(Q(\tau|\mathrm{x})|x)^2}, \tag{14}$$

where $\mu_2(K) = \int u^2 K(u)du$ and $R(K) = \int K^2(u)du$. In addition, $Q''(\tau|\mathrm{x})$ and $g(Q(\tau|\mathrm{x})|x)$ are unknown functions. Yu and Jones (1998) provided the following steps for determining the optimal bandwidth:

- Calculate the ratio $\left(\frac{h_{\tau_1}}{h_{\tau_2}}\right)^5$ by employing the optimal bandwidths at various quantiles $\tau_1$ and $\tau_2$

$$\left(\frac{h_{\tau_1}}{h_{\tau_2}}\right)^5 = \frac{\tau_1(1-\tau_1)Q''(\tau_2|\mathrm{x})^2 g(Q(\tau_2|\mathrm{x})|x)^2}{\tau_2(1-\tau_2)Q''(\tau_1|\mathrm{x})^2 g(Q(\tau_1|\mathrm{x})|x)^2}. \tag{15}$$

- As per their rule-of-thumb procedure let $Q''(\tau_1|\mathrm{x}) = Q''(\tau_2|\mathrm{x})$.
- For the term $g(Q(\tau|\mathrm{x})|x)$, the standard normal distribution was employed.
- Thus, the bandwidth formula is given through the following equation:

$$h_\tau^5 = \pi^{-1} 2\tau(1-\tau)\phi(\Phi^{-1}(\tau))^{-2} h_{1/2}^5, \tag{16}$$

where $h_{1/2}$ is the median's optimal bandwidth.

- The following expression is a combination of the plug-in approach and the rule of thumb, which can be used for computing $h_{1/2}$

$$\left(\frac{h_{mean}}{h_{1/2}}\right)^5 = \frac{2}{\pi}, \tag{17}$$

where $h_{mean}$is the optimal bandwidth for the mean regression. The optimal choice for $h_{mean}$ can be found using the plug-in rule (Fan & Gijbels, 1992; Ruppert et al., 1995):

$$h_{mean}^5 = \frac{R(K)\sigma^2(x)}{n\mu_2(K)^2\{m''(x)\}^2 f(x)}, \tag{18}$$

where $m(x)$ is the conditional mean function.

Under the normality assumption, the relative rate of convergence of the provided rule-of-thumb procedure is $O(n^{-1/7})$ (Yu & Jones, 1998).

### 3.4 Quantile Autoregressive (QAR) Model

Consider a series of independent and identically distributed standard uniform random variables, with an autoregressive process of order $p$:

$$y_t = \beta_0(U_t) + \beta_1(U_t)y_{t-1} + \cdots + \beta_p(U_t)y_{t-p}, \tag{19}$$

where $y_t$ is the response variable and $\boldsymbol{\beta}_j$ are unknown functions that will be estimated. It is a monotone increasing in $U_t$, then the $\tau^{th}$ conditional quantile of $y_t$ is defined as:

$$Q_{y_t}(\tau|y_{t-1}, \dots, y_{t-p}) = \beta_0(\tau) + \beta_1(\tau)y_{t-1} + \cdots + \beta_p(\tau)y_{t-p},$$

compactly $$Q_{y_t}(\tau|\mathcal{F}_{t-1}) = \boldsymbol{x}_t^T\boldsymbol{\beta}(\tau), \tag{20}$$

where $\boldsymbol{x}_t^T = [1, y_{t-1}, \dots, y_{t-p}]$ and $\mathcal{F}_t$ is $\sigma$-field produced by $\{y_r, r \le t\}$. The $\tau^{th}$ conditional quantile of $y_t$ is the linear function of the response's lagged values. The transition from Equation (19) to Equation (20) is an instantaneous outcome of the postulate that for every monotone increasing function $g$ and the standardized uniform random factor $U$, gives:

$$Q_{g(U)}(\tau) = g\big(Q_U(\tau)\big) = g(\tau), \tag{21}$$

where $Q_U(\tau) = \tau$ is the quantile of $U$. The parameters of the autoregressive model might be dependent on the quantile level $(\tau)$. Therefore, the parameters can differ across quantiles. The conditioning factors can shift the location and alter the scale or shape of the conditional distribution of the response variable $y_t$.

The provided model plays a significant role in extending the modeling region between the conventional stationary time series approaches and the alternatives of their unit root. The first order of the quantile autoregressive model is used to illustrate the procedure:

$$Q_{y_t}(\tau|\mathcal{F}_{t-1}) = \beta_0(\tau) + \beta_1(\tau)y_{t-1}, \tag{22}$$

where $\beta_0(\tau) = \sigma\Phi^{-1}(\tau)$ and $\beta_1(\tau) = \min\{\gamma_0 + \gamma_1\tau, 1\}$ for $\gamma_0 \in (0,1)$ and $\gamma_1 > 0$. This model derives the $y_t$ based on the standard Gaussian unit root model when $U_t > (1 - \gamma_0) / \gamma_1$. However, for smaller values of $U_t$, less than $(1 - \gamma_0) / \gamma_1$ a mean reversion tendency will be presented and stationarity will be salvaged (Koenker, 2017). A series of strong positive innovations bolster its unit root, such as behavior, but negative realizations support mean reversion, hence undermining the process's persistence. Setting $\beta_1(\tau)$ to a constant yields the first order of the conventional Gaussian autoregressive model.

### 3.5 Quantile-on-Quantile Regression Model

A nonparametric approach is used to explore the asymmetric impacts of positive and negative shocks in the independent variable on the distribution of the dependent variable. To achieve this, a local linear model is constructed to link the quantile of the dependent variable to the quantile of the independent variable. As a result, the association between the independent variable and the dependent variable may be altered at various points in their respective distributions. The $\tau$-quantile of the dependent variable as a function of the independent variable is given through the following equation:

$$y_t = \beta^{\tau}(x_t) + \alpha^{\tau} y_{t-1} + \varepsilon_t^{\tau}, \tag{23}$$

where $y_t$ and $y_{t-1}$ are the dependent variable at period $t$ and $t-1$ respectively, while $x_t$ is the independent variable at period $t$, and $\beta^{\tau}(.)$ is an unknown function because there is no prior hypothesis or assumption suggested on the association between $x_t$ and $y_t$. Then, $\varepsilon_t^{\tau}$ is the random error term of a zero $\tau$-quantile, and $\tau$ is the $\tau^{th}$quantile that corresponds to the conditional distribution of $y_t$.

To investigate the relationship between the $\tau$-quantile of the dependent variable and the $\theta$-quantile of the independent variable, represented by $x^{\theta}$, Equation (23) has been inspected in the neighborhood of $x^{\theta}$. Given the unknown quantity $\beta^{\tau}(.)$, this function could be linearized using the first-order Taylor series expansion of $\beta^{\tau}(.)$ around $x^{\theta}$. This results in:

$$\beta^{\tau}(x_t) \approx \beta^{\tau}\left(x^{\theta}\right) + \beta^{\tau'}(x^{\theta})(x_t - x^{\theta}), \tag{24}$$

where $\beta^{\tau\prime}$ is the partial derivative of $\beta^{\tau}(x_t)$, $x_t$ relating the marginal effect, $\beta^{\tau}\left(x^{\theta}\right)$ and $\beta^{\tau'}(x^{\theta})$ are parameters that doubly indexed in $\tau$ and $\theta$. This is because the $\theta$-quantile of the independent variable is a function of $\theta$ only. Since $\beta^{\tau}\left(x^{\theta}\right)$ and $\beta^{\tau'}(x^{\theta})$ are functions of both $\tau$ and $x^{\theta}$, and since $x^{\theta}$ is a function of $\theta$, this indicates that $\beta^{\tau}\left(x^{\theta}\right)$ and $\beta^{\tau'}(x^{\theta})$ are both functions of $\tau$ and $\theta$. Therefore, $\beta^{\tau}\left(x^{\theta}\right)$ and $\beta^{\tau'}(x^{\theta})$ can be redefined as $\beta_0(\tau, \theta)$ and $\beta_1(\tau, \theta)$ respectively. As a result, Equation (24) can be represented as follows:

$$\beta^{\tau}(x_t) \approx \beta_0(\tau,\theta) + \beta_1(\tau,\theta)\left(x_t - x^{\theta}\right), \tag{25}$$

next, substituting Equation (25) into Equation (23) to get

$$y_t = \beta_0(\tau,\theta) + \beta_1(\tau,\theta)\left(x_t - x^{\theta}\right) + \alpha(\tau)y_{t-1} + \varepsilon_t^{\tau},$$

For simplicity, let $\beta_0(\tau,\theta) + \beta_1(\tau,\theta)\left(x_t - x^{\theta}\right) + \beta_2(\tau)y_{t-1} = (*)$ (26)

where $\beta_2(\tau) \equiv \alpha(\tau) \equiv \alpha^{\tau}$, and part $(*)$ of Equation (26) is the $\tau$-conditional quantile of the dependent variable that shows the relationship between the $\tau$-quantile of the dependent variable and the $\theta$-quantile of the independent variable, conditionally that the binary index of $\beta_0$ and $\beta_1$ are associated with $\tau$ and $\theta$ respectively. Moreover, the general dependence between the dependent variable and the independent variable can be measured by the dependence between their distributions. The obtained parameters may produce distinct outputs according to the $\tau$-quantile of $y_t$ and $\theta$-quantile of $x_t$. To estimate the parameters in Equation (26), the $x_t$ is replaced with its estimated counterpart $\hat{x}_t$, and the $x^{\theta}$ is replaced with its experimental quantile of $\hat{x}_t^{\theta}$. Then, the local linear regression method on the minimization problem is used:

$$\min_{b0,b1,\mathrm{b2}} \sum_{i=1}^{n} \rho_{\tau}\left[y_t - b_0 - b_1\left(\hat{x}_t - \hat{x}^{\theta}\right) - b_2(\tau)y_{t-1}\right] K\left(\frac{F_n(\hat{x}_t) - \theta}{h}\right), \tag{27}$$

where $b_0, b_1$ and $b_2$ are the estimated coefficients of $\beta_0, \beta_1$ and $\beta_2$ respectively, and $\rho_{\tau}(.)$ is the slanted absolute value function that creates the $\tau$-conditional quantile of $y_t$. The Gaussian kernel function $K(.)$ is employed to describe the local effect presented by the $\theta$-quantile of the independent variable. This kernel is used as a minimal weighting criterion in the minimization problems to enhance the estimation efficiency based on the bandwidth $h$. These weights are associated with the distance of $\hat{x}_t$ from $\hat{x}^{\theta}$. Therefore, $F_n(\hat{x}_t) = \frac{1}{n}\sum_{k=1}^{n} I(\hat{x}_k < \hat{x}_t)$ is the distance of empirical distribution function, where $I$ is the indicator function, and $\theta$ is the distribution function's value that is associated with $x^{\theta}$.

### 3.6 Robustness Check for the QQR Approach

The validity of the QQR technique is checked based on the procedures proposed by (Sim & Zhou, 2015). The QQR model can be considered as a technique that decomposes the estimates of the conventional QR approach, where it can be used to obtain a particular estimated parameter at different quantiles of an independent variable. In this thesis, the QR model regresses the $\tau^{th}$ quantile of the dependent variable on the independent variable; therefore, the QR parameters are indexed by $\tau$ only. However, the QQR model regresses the $\tau^{th}$ quantile of the dependent variable on the $\theta^{th}$ quantile of the independent variable; hence, its parameters are indexed by $\tau$ and $\theta$.

As a result, the QQR approach has more detailed information about the association between independent and dependent variables than the QR approach would, as this relationship is perceived by the QQR approach to be inhomogeneous across $\theta$. Because of this inherent property of decomposition in the QQR model, the conventional QR estimates can be recovered using QQR estimates. In particular, the parameters of the QR model, which are indexed by $\tau$ only, can be constructed by averaging the QQR coefficients along $\theta$ using the following formula:

$$\hat{\beta}_k(\tau) = \frac{1}{Q}\sum_{\theta} \hat{\beta}_k(\tau, \theta), \quad k = 0,1,2, \ldots \tag{28}$$

where $Q$ is the length of the quantiles vector and $\theta$ is a vector of quantile levels.

In this regard, the validity of the QQR model can be checked by comparing the $\hat{\beta}_k(\tau)$ estimated by the quantile regression with the averaged of $\hat{\beta}_k(\tau, \theta)$ estimated by the quantile-on-quantile regression.

## 3.7 Jarque-Bera (JB) Test

Jarque and Bera (1980) provided the JB test as a goodness-of-fit test to measure whether the data has skewness and kurtosis. In other words, it tests the normality distribution. Under the normality assumption, the theoretical values of skewness and kurtosis are zero and three, respectively. The value of the test statistic is always positive. Several authors, including Jarque and Bera (1987), and Urzua´ (1996), have noted that the JB test performs well in comparison to various other normality tests when the alternatives to the normal distribution are relevant to the Pearson family. The test statistic for JB is given through the following equation:

$$\mathrm{JB} = \frac{n}{6}\left(S^2 + \frac{(K-3)^2}{4}\right), \tag{29}$$

where $n$ is the sample size, $S$ is the sample skewness, $S = \hat{\mu}_3/\hat{\mu}_2^{3/2}$, and it is an estimator of the parameter $\beta_1 = \mu_3/\mu_2^{3/2}$. In addition, $K$ is the sample kurtosis, $K = \hat{\mu}_4/\hat{\mu}_2^2$, and it is an estimator of the parameter $\beta_2 = \mu_4/\mu_2^2$. The terms $\mu_2, \mu_3$ and $\mu_4$ are the second, third, and fourth moments, respectively. Their estimates are obtained through the equation $\hat{\mu}_j = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^j\,, j = 2{,}3{,}4$ (Thadewald & Büning, 2007). According to Bowman and Shenton (1975), the asymptotic distribution of the JB test is a chi-squared distribution with two degrees of freedom. Thus, the null hypothesis that the distribution of the data is normal has to be rejected at a significant level $\alpha$ when $JB \geq \chi^2_{1-\alpha,2}$.

# CHAPTER 4: APPLICATION

## 4.1 Coronavirus Pandemic News, Bitcoin, and Gold Returns

Cryptocurrency is a system that uses cryptography to provide secure transactions or exchanges of digital currencies in a decentralized manner using a peer-to-peer mechanism. Investors can trade cryptocurrencies on markets for fiat currencies. Significantly, the first successfully created cryptocurrency was Bitcoin. In 2008, an individual or a group of hackers with the pseudonym Satoshi Nakamoto published a white paper to illustrate the applications of a digital currency called Bitcoin using blockchain technology (Härdle et al., 2019). The concept of the blockchain can be described as a distributed database to record all digital events such as trades that have been implemented and share them among participating parties. These transactions in the public registry must be confirmed by most of the participants in that system when the information entered cannot be erased. It encloses a provable record of each possible transaction. After ten years, hundreds of cryptocurrencies were created using similar innovations that Bitcoin proposed, with some changes to their governing algorithms. Moreover, an untold number of applications have been implemented using blockchain technology (Crosby et al., 2016).

Investing in cryptocurrency offers several potential benefits, including high profits, diversification, security against currency depreciation and inflation, and future adoption and usage in many nations throughout the world. On the other hand, there are some potential drawbacks, such as high volatility and large losses (for some cryptocurrencies), which are positively correlated with gold and equities (also for some cryptocurrencies). This is because they have a low store of value due to the fluctuation and bounded usage being uncontrolled and open to unscrupulous behavior. Meanwhile, the value of cryptocurrencies is affected by news and rumors (Bunjaku, 2017). For instance, in May 2021, Tesla CEO Elon Musk announced in a tweet that "Tesla has suspended vehicle purchases using Bitcoin due to climate change concerns," which led to a reduction in Bitcoin's value by more than 10%, while Tesla's shares dipped (BBC, 2021).

Gold occupies a unique position among metals for its high value and long history of being intertwined with cultures for decades. It has a significant impact on the global economy. Gold does not corrode, rust or decay. Unlike paper money, coins, and other items, gold has retained its worth over time. It is seen as a way for people to carry on and protect their wealth by keeping it in bank vaults. Therefore, it is important to provide research about gold prices and the factors that may affect their exchange rates. Gold price history should be reviewed to consider investment opportunities with this valuable metal. Furthermore, the price history may offer information that assists in purchasing or selling options (Adams, 2016).

The coronavirus pandemic-related news indices are provided by the RavenPack team and published on the Coronavirus Media Monitor website. The RavenPack database monitors recent news and trending events surrounding the COVID-19 disease outbreak to generate a more analytical way of tracking information about the novel COVID-19 pandemic and to measure people's reactions around the world, which include enormous social distancing measures and scare buying. The database summarizes emotions and sentiments from millions of news articles and posts them into simple indicators that resemble stock tickers. In addition, the database provides users with more information and forewarning about changes in the virus's news, changes that frequently reflect or foretell real-world events. These indices are Panic, Media Hype, Fake News, Sentiment, Infodemic, and Media Coverage.

According to the RavenPack coronavirus database, the coronavirus Panic Index (PI) tracks the amount of news coverage that makes reference to panic or hysteria, and COVID-19. The larger values indicate more references to panic published in the media. The coronavirus Media Hype Index (MHI) shows the percentage of news stories that mention the COVID-19. The coronavirus Fake News Index (FNI) summarizes the amount of media news about the novel virus that mentions misstatement or fake news about coronavirus. Larger values indicate more fake news published in the media. The coronavirus worldwide Sentiment Index (SI) shows the amount of sentiment in all entities that were reported in the news, along with COVID-19. The coronavirus Infodemic Index (II) computes the percentage of all existence, including places, companies, etc., that are associated with COVID-19 in some way. The coronavirus Media Coverage Index (MCI) computes the proportion of all news sources that mention the novel coronavirus (Coronavirus Media Monitor, 2021). More details about coronavirus-related news indices are posted in Appendix C.

As a result, Bitcoin and gold returns and coronavirus pandemic-related news indices are the best illustrations of non-determinism. The daily Bitcoin and gold prices were collected from the online source (https://www.investing.com/). The daily returns of Bitcoin and gold were calculated and converted into logarithmic values (log differences). Let $P_t$ be the price at day $t$, $P_{t-1}$ be the price at day $t-1$, and $r_t$ be the log return defined as $r_t = log(\frac{P_t}{P_{t-1}})$. The indices induced by the daily coronavirus pandemic news were collected from the online source (https://coronavirus.ravenpack.com/). In this study, the coronavirus pandemic-related news indices were divided by 100.

This research aims to evaluate how Bitcoin and gold returns respond to the coronavirus pandemic-related news indices throughout a sample period from January

23, 2020, to September 01, 2021. The data include nine variables: daily dates, Bitcoin returns, gold returns, PI, MHI, FNI, SI, II, and MCI.

This study focuses on exploring the unprecedented impact of coronavirus pandemic-related news indices on Bitcoin and gold returns separately. It shows how the different levels of each coronavirus pandemic- related news index influence Bitcoin and gold returns. It examines whether there are any differences in the pattern of small, medium, and large changes in the level of each coronavirus pandemic-related news index when Bitcoin and gold returns are low, middle, and high. In this research, RStudio software is used to analyze the returns and the indices associated with coronavirus outbreak data.

## 4.2 Analysis

### *4.2.1 Descriptive Statistics*

Table 4.1 shows that Bitcoin returns ranged between -0.48 and 0.179. Gold returns ranged between -0.05 and 0.051, where 25% of the gold returns amounted to less than or equal to -0.004 and 75% were less than or equal to 0.005. In contrast, 25% of the Bitcoin returns were less than or equal to -0.015, and 75% of the Bitcoin returns were less than or equal to 0.022. Bitcoin's daily mean and median returns are around 0.003. This amounts to a monthly return of approximately 0.09 percent per month, and an annualized return of approximately 1.06 percent per year. Daily mean and median gold returns are around 0.001, resulting in a monthly return of roughly 0.3 percent per month and an annualized return of about 0.37 percent per year. Bitcoin and gold annualized returns have relatively small positive values, indicating that both commodities can act as safe havens during the current pandemic period.

The highest percentage of the panic index was 9.21, and the lowest percentage was 0.63. The percentages of the Media Hype Index ranged from 4.27 to 69.27. Among all the indices that had their values ranging from 0 and 100, the Fake News Index had the lowest average of 0.638 percent, while the Media Coverage Index had the highest average of 69.65 percent. The Sentiment Index has a negative average of 16.45 percent, while 25% of the Infodemic Index is less than or equal to 43.285 percent, and 75% is less than or equal to 54.47 percent.

Table 4.1. Descriptive Statistics

| Var | Mean | Min | p25 | Med | p75 | Max | Skew | Kurt | JB (p-value) |
|---|---|---|---|---|---|---|---|---|---|
| BTC | 0.003 | -0.48 | -0.015 | 0.003 | 0.022 | 0.179 | -2.26 | 26.7 | 18102 (2.2e-16) |
| Gold | 0.001 | -0.05 | -0.004 | 0.001 | 0.005 | 0.051 | -0.52 | 5.2 | 805.09 (2.2e-16) |
| PI | 2.624 | 0.63 | 1.82 | 2.42 | 3.045 | 9.21 | 1.8 | 4.86 | 631.04 (2.2e-16) |
| MHI | 31.93 | 4.27 | 23.7 | 31.89 | 37.59 | 69.27 | 0.56 | 0.57 | 27.597 (1.017e-06) |
| FNI | 0.638 | 0.05 | 0.41 | 0.56 | 0.8 | 2.24 | 1.14 | 1.56 | 131.67 (2.2e-16) |
| SI | -16.45 | -69.9 | -29.66 | -11.67 | 0.93 | 12.96 | -0.82 | -0.28 | 46.909 (6.514e-11) |
| II | 48.42 | 9.79 | 43.285 | 49.38 | 54.47 | 67.67 | -1.03 | 1.61 | 118.47 (2.2e-16) |
| MCI | 69.65 | 21.9 | 66.735 | 72.69 | 74.58 | 82.61 | -2.33 | 6.06 | 1007.2 (2.2e-16) |

Var, Min, p25, Med, p75, Max, Skew, Kurt and JB stand for variable, minimum value, first quantile, median, third quantile, maximum value skewness, kurtosis, and Jarque-Bera test statistics, respectively. The descriptive statistics of the coronavirus related six indices are before dividing them by 100.

The daily Bitcoin and gold prices (left panels) in Figure 4.1 show that at the beginning of the COVID-19 pandemic, Bitcoin prices had their lowest value of 4927, which then increased gradually to reach a peak of 63518 until the mid-April 2021, before they had a steep decline. Gold prices soared from the start of the pandemic until mid-August 2020, which then began to fluctuate and gradually drop.

As observed in the histogram plots of Bitcoin and gold returns (right panels) in Figure 4.1, the dynamics of their returns exhibit significant evidence of volatility clustering with many outliers and excess positive kurtosis, implying leptokurtic (heavy-tailed) distributions. Bitcoin returns have a negatively skewed distribution. It seems that gold returns are normally distributed, but the JB test rejects that assumption with respect to the p-value (based on α=0.01).

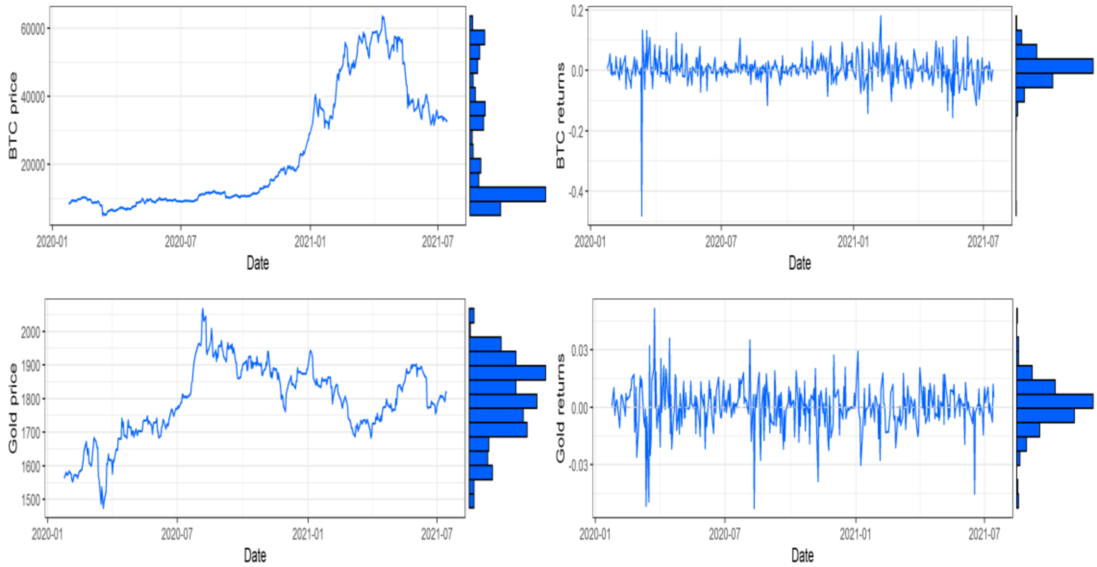


Figure 4.1. Evolution of the Bitcoin and gold daily prices and returns throughout the sample period

At the beginning of the COVID-19 pandemic, the Panic, Media Hype, Fake News, Infodemic, and Media Coverage Indices increased gradually to reach their highest points of 0.092, 0.693, 0.018, 0.677, and 0.826, respectively. Meanwhile, the Sentiment Index reached a peak of 0.13 by mid-2020. After March 2020, the Panic Index showed a sharp decline, followed by random fluctuations around 0.03, with a slight increase at the end of December 2020. The behavior of the Fake News Index is approximately similar to that of the Panic Index, but with more spikes. Both the

Infodemic and Media Coverage Indices decreased gradually and slightly after the period of initial COVID-19 spread around April 2020. Similarly, the Media Hype Index declined gradually, with a slight increase from October 2020 to December 2020. The Sentiment Index decreased to its lowest value during March 2020, and then started to rise again in the latter half of 2020 to reach its highest value. It kept hovering around that level until seeing a sharp decline in December 2020 before finally rising to slightly below the peak levels. All the indices (except the Sentiment Index) reached their lowest points by the beginning of the sample period, as shown in Figure 4.2.

The distribution of the Panic, Media Hype, and Fake News Indices is skewed to the right. While the Sentiment, Infodemic and Media Coverage Indices are all skewed negatively, all of the indices have leptokurtic distributions except for the Sentiment Index, whose distribution is platykurtic.

The JB test is used to test the null hypothesis that data have a normal distribution with zero skewness and excess kurtosis. The results show that the variables are significant with respect to their p-value (based on $\alpha = 0.01$). The distributions of the variables are not normally distributed, as shown in Figures 4.1, 4.2, and Table 4.1, which is a solid reason to use the quantile regression technique to handle the heavy tails data. Positive relationships can be seen when comparing the daily evolution of Bitcoin and gold returns in Figure 4.1 with the trending patterns of all the coronavirus-related indices in Figure 4.2 (except for the Sentiment Index).

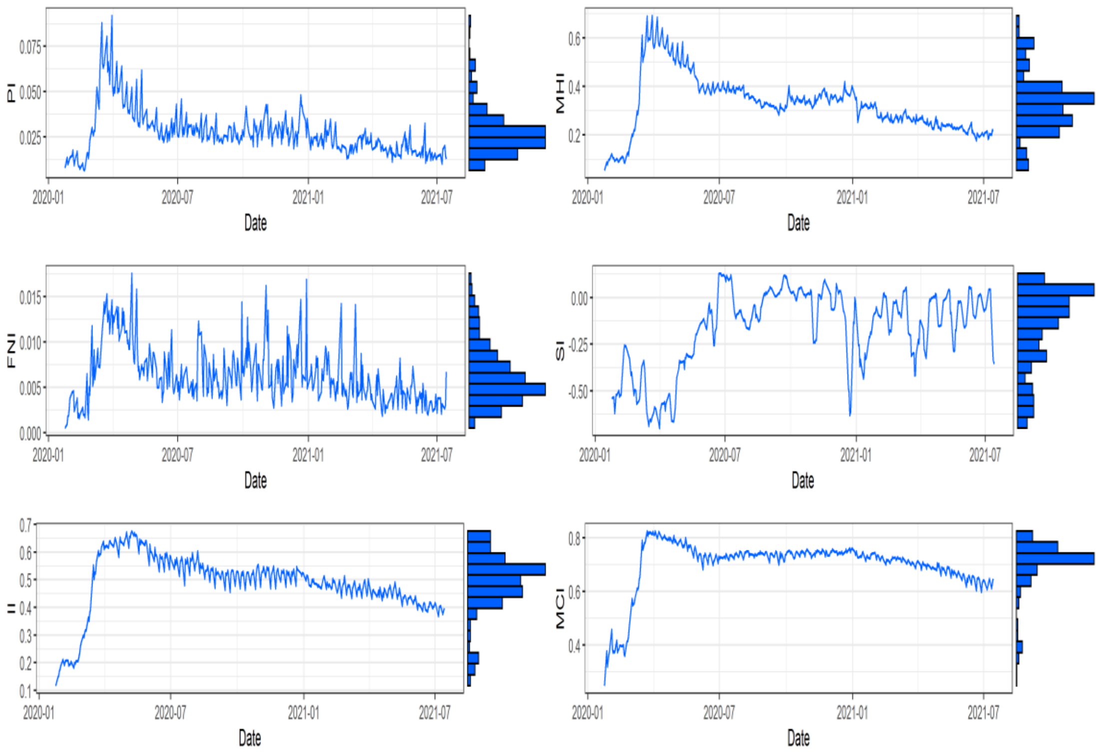


Figure 4.2. Trends and histograms of indices induced by the daily coronavirus pandemic news throughout the sample period

As shown in Figure 4.3, the correlation coefficients between Bitcoin returns and the coronavirus pandemic news indices ranged from -0.03 (between Bitcoin returns and the Sentiment index) to 0.15 (between Bitcoin returns and the Fake News index). Similarly, the lowest correlation is between the gold returns and the Sentiment Index of -0.02, while the highest correlation of 0.07 occurred between gold returns and the Fake News Index. The relatively poor correlations between all the pandemic news indices and Bitcoin and gold returns suggest that the conventional ordinary least squares method will not be able to capture the relationship between coronavirus-related news and Bitcoin (and gold) returns. The correlation between Bitcoin (and gold) returns and the Sentiment Index is very weak negative, indicating that positive sentiment leads to lower returns, while negative sentiment is linked to high returns.

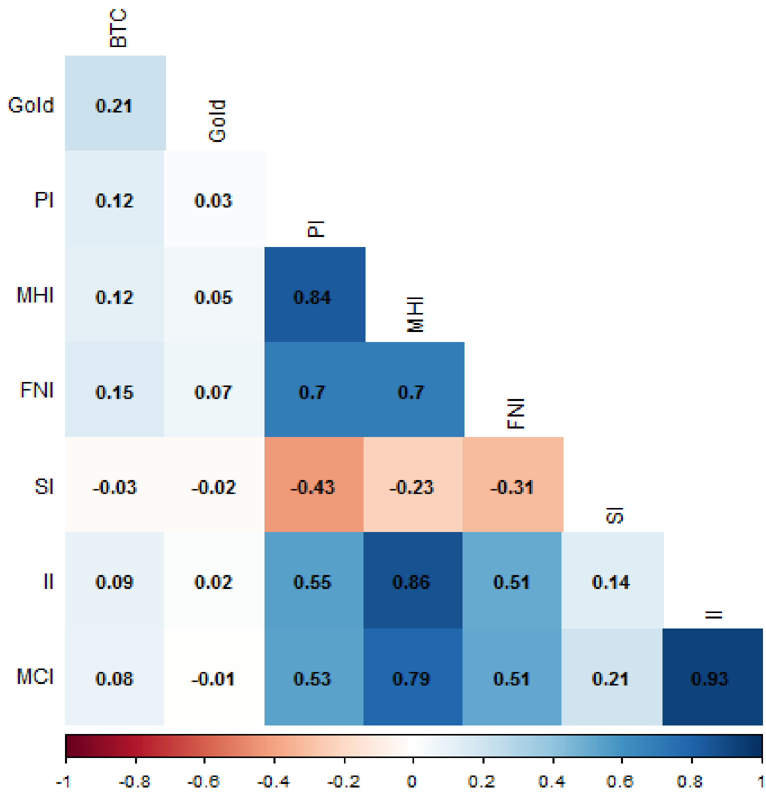


Figure 4.3. Correlation coefficients matrix

Graphs in Figures 4.4 and 4.5 give an idea of the quantiles of key variables. Quantiles for Bitcoin and gold returns reveal similar patterns, with returns below approximately 35$^{th}$ percentile being negative. In addition, similar graph pattern for all the indices (except for the Sentiment Index) induced by the daily coronavirus pandemic news. Due to its nature, the Sentiment Index's quantile plot can be seen in the same manner as other indices. For instance, a large negative sentiment value is interpreted similarly to a large positive Fake News value.

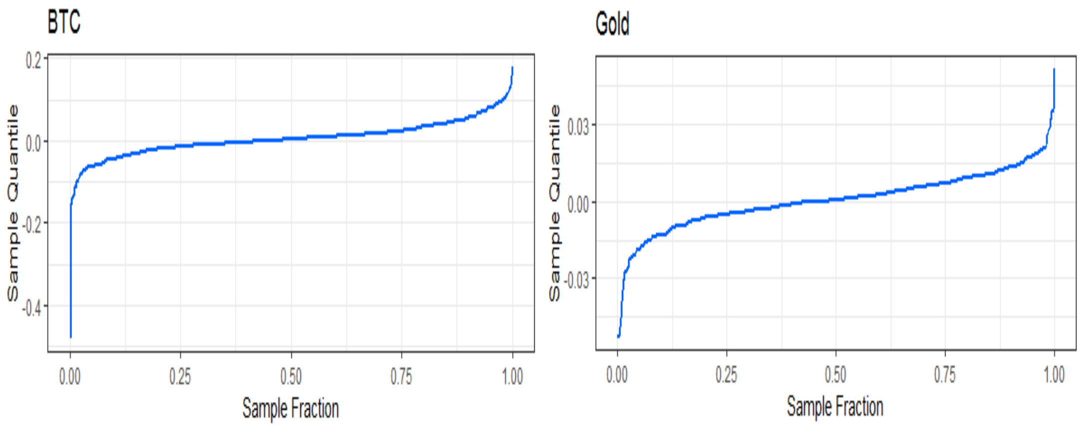


Figure 4.4. Quantile plots of Bitcoin and gold returns

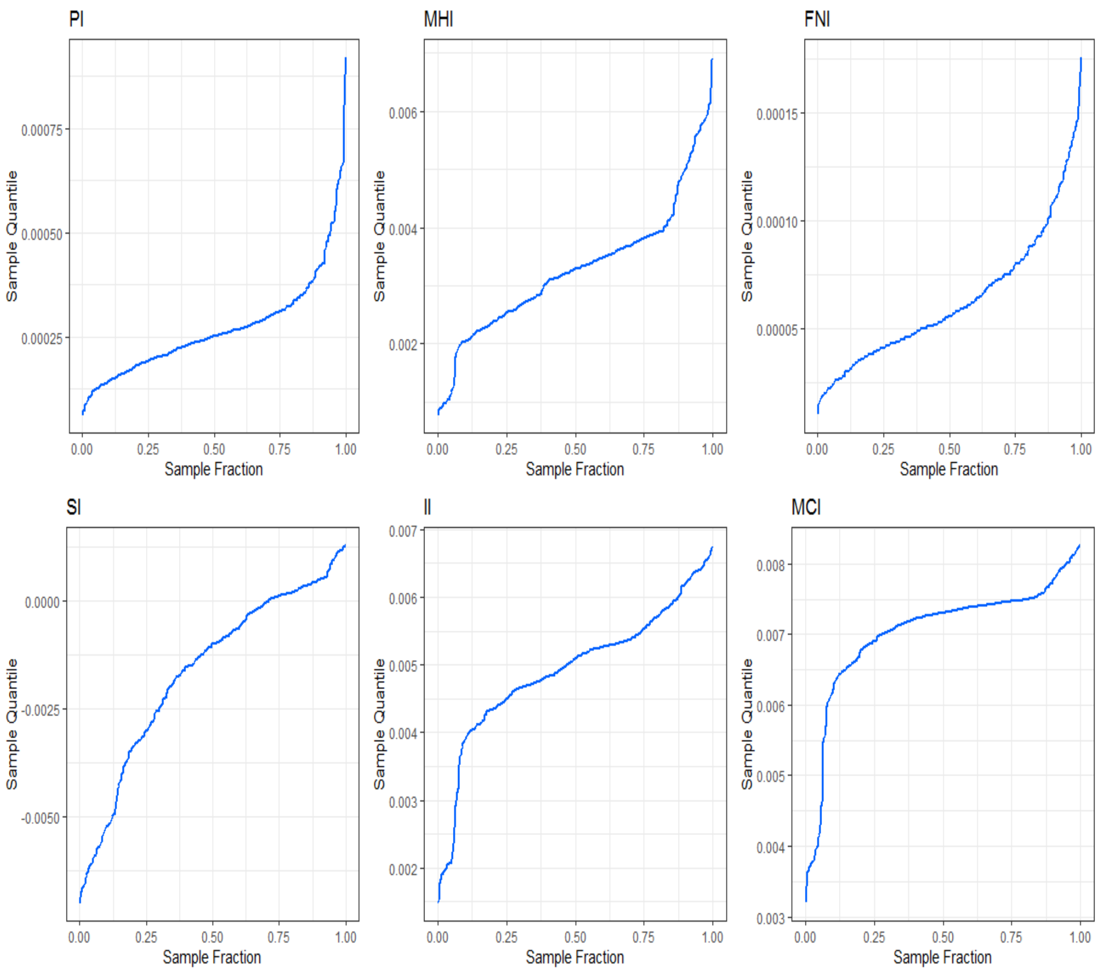


Figure 4.5. Quantile plots of coronavirus pandemic-related news indices

*4.2.2 Quantile Regression Model*

Several researchers have shown that negative and positive news reactions are asymmetric, with bad news (as independent variables) having a far bigger effect on stock returns (as a dependent variable) than good news. The QR model is designed to uncover the distribution of the dependent variable based on the symmetrical (not asymmetric) effects of independent variables. Therefore, such asymmetric responses are difficult to capture with the quantile regression technique. In this situation, the QQR approach is a more robust method for providing a more complete picture of dependence.

The QR model is employed to explore the effect of good and bad news associated with the COVID-19 epidemic on Bitcoin and gold returns. The estimated parameters and their corresponding p-values are given in Tables 4.2 and 4.3. For example, the QR equation of the Panic Index and Bitcoin return at the 5$^{th}$ quantile based on Equation (22) can be written as:

$$btc_t = -0.07 + 0.22PI - 0.1btc_{t-1},$$

when the Panic Index and Bitcoin returns at time $t-1$ are both zero, Bitcoin returns fall by 0.07 unit. For one additional panic value, the Bitcoin return at time $t$ will increase by 0.22 as long as the Bitcoin return at time $t-1$ remains constant. Similarly, one additional unit in Bitcoin returns at time $t-1$ will decrease the Bitcoin returns at time $t$ by 0.1, as long as the panic value remains constant for those with low Bitcoin returns (at the 5% quantile).

The estimated intercept coefficient makes a significant contribution to the model. This is justifiable as the intercept coefficient reflects the level of returns at time $t$ for given levels of coronavirus-related news indices at time $t$ and returns at time $t-1$; hence, a higher return quantile have a larger intercept parameter. In almost all cases, the effects of COVID-19-related news and $\hat{\beta}_2(\tau)$ on gold returns are very weak, as the associated p-values indicate insufficient evidence to conclude that $\hat{\beta}_1(\tau)$ and $\hat{\beta}_2(\tau)$ are significantly different from zero, while this is not the case for Bitcoin returns.

The Sentiment Index has a negative impact on Bitcoin returns, the magnitude of that effect increases from the lowest to the upper quantiles. The negative and significant effect on Bitcoin returns appears between the $75^{th}$ and $95^{th}$ quantiles. However, the Sentiment Index influences the gold return at the $90^{th}$ quantile only negatively and significantly. The Fake News Index has a positive and significant effect on Bitcoin returns across all quantiles except for the $90^{th}$ quantile, while the effect of that index is positive and significant on gold returns at the $90^{th}$ quantile only. The Infodemic and Media Coverage Indices have a positive and significant effect on gold returns at the highest quantile only ($95^{th}$ quantile).

Table 4.2. Estimation Results Based on QR Model, Where Bitcoin Quantiles is the Response Variable

| Index | $\hat{\beta}_k(\tau)$ | 5th | 10th | 25th | 50th | 75th | 90th | 95th |
|---|---|---|---|---|---|---|---|---|
| PI | $\hat{\beta}_0(\tau)$ | -0.07*** | -0.06*** | -0.02*** | -0.01 | 0.01 | 0.04*** | 0.05*** |
| | $\hat{\beta}_1(\tau)$ | 0.22 | 0.53* | 0.34 | 0.41** | 0.55* | 0.57 | 0.95* |
| | $\hat{\beta}_2(\tau)$ | -0.1 | -0.18 | -0.14** | -0.14** | -0.18*** | -0.18** | -0.17** |
| MHI | $\hat{\beta}_0(\tau)$ | -0.08*** | -0.06*** | -0.03*** | -0.00 | 0.02* | 0.04*** | 0.05*** |
| | $\hat{\beta}_1(\tau)$ | 0.07* | 0.07* | 0.05** | 0.03 | 0.02 | 0.06* | 0.08* |
| | $\hat{\beta}_2(\tau)$ | -0.12 | -0.19 | -0.15** | -0.15*** | -0.18** | -0.25*** | -0.19** |
| FNI | $\hat{\beta}_0(\tau)$ | -0.08*** | -0.06*** | -0.03*** | -0.01* | 0.01* | 0.03*** | 0.06*** |
| | $\hat{\beta}_1(\tau)$ | 3.5*** | 2.96*** | 2.23*** | 2*** | 1.85** | 3.26 | 3.35** |
| | $\hat{\beta}_2(\tau)$ | -0.07 | -0.26* | -0.17*** | -0.15*** | -0.22*** | -0.16** | -0.16 |
| SI | $\hat{\beta}_0(\tau)$ | -0.07*** | -0.04*** | -0.01*** | 0 | 0.02*** | 0.05*** | 0.06*** |
| | $\hat{\beta}_1(\tau)$ | -0.01 | 0 | -0.00 | -0.01 | -0.04*** | -0.05* | -0.07* |
| | $\hat{\beta}_2(\tau)$ | -0.04 | -0.14 | -0.08 | -0.14*** | -0.19*** | -0.26*** | -0.22** |
| II | $\hat{\beta}_0(\tau)$ | -0.1*** | -0.08*** | -0.04*** | -0.00 | 0.02 | 0.03** | 0.05* |
| | $\hat{\beta}_1(\tau)$ | 0.07 | 0.08 | 0.05* | 0.02 | 0.01 | 0.05* | 0.06 |
| | $\hat{\beta}_2(\tau)$ | -0.06 | -0.18 | -0.13* | -0.14*** | -0.17** | -0.29*** | -0.15 |
| MCI | $\hat{\beta}_0(\tau)$ | -0.1 | -0.06 | -0.05* | -0.02 | 0.02 | 0.01 | 0.01 |
| | $\hat{\beta}_1(\tau)$ | 0.05 | 0.03 | 0.05 | 0.03* | 0.01 | 0.07** | 0.09** |
| | $\hat{\beta}_2(\tau)$ | -0.07 | -0.14 | -0.12** | -0.14*** | -0.17** | -0.26*** | -0.16* |

* Reject the null hypothesis: $H_0: \beta_k(\tau) = 0$ $(k = 0,1,2)$ at 10% significance level.
** Reject the null hypothesis: $H_0: \beta_k(\tau) = 0$ $(k = 0,1,2)$ at 5% significance level.
*** Reject the null hypothesis: $H_0: \beta_k(\tau) = 0$ $(k = 0,1,2)$ at 1% significance level.

Table 4.3. Estimation Results Based on QR Model, Where Gold Quantiles is the Response Variable

| Index | $\hat{\beta}_k(\tau)$ | 5th | 10th | 25th | 50th | 75th | 90th | 95th |
|---|---|---|---|---|---|---|---|---|
| PI | $\hat{\beta}_0(\tau)$ | -0.01* | -0.01 | -0.0** | 0 | 0* | 0.01*** | 0.01*** |
| | $\hat{\beta}_1(\tau)$ | -0.23 | -0.28* | -0.04* | -0.01 | 0.08 | 0.22*** | 0.24 |
| | $\hat{\beta}_2(\tau)$ | -0.04 | -0.03 | 0 | 0.01 | 0.02 | 0.05*** | 0.06* |
| MHI | $\hat{\beta}_0(\tau)$ | 0.01* | -0.01** | -0.00** | 0 | 0* | 0.01*** | 0.01*** |
| | $\hat{\beta}_1(\tau)$ | -0.01 | -0.01 | -0.00 | -0.00 | 0.01 | 0.02*** | 0.02 |
| | $\hat{\beta}_2(\tau)$ | -0.03 | -0.02 | 0 | 0.01* | 0.03 | 0.04 | 0.05 |
| FNI | $\hat{\beta}_0(\tau)$ | -0.02*** | -0.01*** | -0.00*** | 0 | 0* | 0.01*** | 0.01*** |
| | $\hat{\beta}_1(\tau)$ | -0.28 | -0.46 | -0.15 | -0.01 | 0.34 | 0.71*** | 0.78 |
| | $\hat{\beta}_2(\tau)$ | -0.03 | -0.02 | 0 | 0.01 | 0.02 | 0.04*** | 0.05 |
| SI | $\hat{\beta}_0(\tau)$ | -0.02*** | -0.01*** | -0.00*** | 0 | 0.01*** | 0.01*** | 0.01*** |
| | $\hat{\beta}_1(\tau)$ | 0.01 | 0.01 | 0 | 0 | -0.01 | -0.01*** | -0.01 |
| | $\hat{\beta}_2(\tau)$ | -0.04 | -0.02 | -0.00 | 0.01 | 0.02 | 0.05* | 0.06 |
| II | $\hat{\beta}_0(\tau)$ | -0.01*** | -0.01* | -0.00 | 0 | 0.01*** | 0.01*** | 0.01* |
| | $\hat{\beta}_1(\tau)$ | -0.01 | -0.01 | -0.00 | -0.00 | 0 | 0 | 0.02* |
| | $\hat{\beta}_2(\tau)$ | -0.02 | -0.01 | 0 | 0.01 | 0.02 | 0.04 | 0.04 |
| MCI | $\hat{\beta}_0(\tau)$ | -0.01 | -0.01 | -0.00 | 0.01 | 0.01 | 0.01 | 0.01 |
| | $\hat{\beta}_1(\tau)$ | -0.01 | -0.01 | -0.01 | -0.01 | 0 | 0 | 0.02* |
| | $\hat{\beta}_2(\tau)$ | -0.02 | -0.02 | 0 | 0.01 | 0.03 | 0.04 | 0.04 |

* Reject the null hypothesis: $H_0\colon \beta_k(\tau) = 0$ $(k = 0,1,2)$ at 10% significance level.
** Reject the null hypothesis: $H_0\colon \beta_k(\tau) = 0$ $(k = 0,1,2)$ at 5% significance level.
*** Reject the null hypothesis: $H_0\colon \beta_k(\tau) = 0$ $(k = 0,1,2)$ at 1% significance level.

### *4.2.3 Quantile-on-Quantile Regression (QQR) Model*

Figures 4.6, 4.7, 4.8, and 4.9 illustrate the empirical outcomes from the quantile-on-quantile regression approach used to model the daily news-based indices linked to the COVID-19 outbreaks and the daily returns of Bitcoin and gold. The colored vertical bars on the right side of the 3D graphs represent the direction and magnitude of the associations between the coefficients. The $x$, $y$, and $z - axis$ display the $\theta^{th}$ quantiles of indices induced by the COVID-19 news, the $\tau^{th}$ quantiles of Bitcoin or gold returns, and the estimated parameters $\hat{\beta}_0(\tau,\theta)$ or $\hat{\beta}_1(\tau,\theta)$ or $\hat{\beta}_2(\tau,\theta)$ , respectively. The values of the estimated parameters and the relationship shifted from the bottom to up as the color shifted from dark blue (downward), green (middle) to dark yellow (upward). All the values in the vertical bars in Figure 4.6 (except for the Sentiment Index) have positive values.

Figure 4.6 shows a predominantly weak relationship between the Panic Index and Bitcoin returns, as seen by the light blue color that spreads across the graph with the exception of the presented yellow color in the Panic Index's uppermost quantiles ($90^{th}$ to $95^{th}$) and all quantiles of Bitcoin returns. The association between the Panic Index's uppermost quantiles and Bitcoin returns reveals a strong positive relationship. This suggests that extreme panic news has a beneficial influence on Bitcoin returns across most of Bitcoin return quantiles. As a result, Bitcoin can act as a hedge against panic caused by the media.

The Media Hype and Fake News Indices reveal essentially identical patterns of association with Bitcoin returns at several quantiles with some exceptions at the upper-middle quantile of the Media Hype Index and the lowermost quantiles of Bitcoin returns. The grid combining the lowermost ($5^{th}$ to $15^{th}$) and the uppermost ($90^{th}$ to $95^{th}$) quantiles of the news that induced hype and the Fake News Index with all the quantiles of Bitcoin returns illustrates a strong positive relationship. On the other hand, Bitcoin return decreases in the regions that combine the ($60^{th}$ to $80^{th}$) quantiles of the Media Hype Index and the ($5^{th}$ to $55^{th}$) quantiles of Bitcoin returns, where Bitcoin recovered its lost value outside these regions. This advantage reveals that varying levels of media-induced hype and fake news have asymmetrical effects on Bitcoin's returns, and that Bitcoin acts as a hedge against excessive levels of media-induced hype and fake news.

The graphical depiction of the Infodemic and the Media Coverage Indices illustrates almost identical patterns of predominantly weak and positive links with Bitcoin returns for the vast majority of different quantile combinations. The area indicated by the lowermost ($5^{th}$ to $15^{th}$) quantiles of the Infodemic Index and the medium-to-higher ($40^{th}$ to $95^{th}$) quantiles of Bitcoin returns show a strong relationship. The grid that combines the lowermost quantiles ($5^{th}$ to $10^{th}$) of the Media Coverage Index and the uppermost quantiles ($70^{th}$ to $95^{th}$) of Bitcoin returns shows a strong relationship. Such an asymmetric effect confirms prior findings: Bitcoin acts as a hedge with a weak safe haven against small and large shocks of Infodemic Index and Media Coverage induced by COVID-19-related news.

In contrast to other indices, the Sentiment Index has a strong positive impact on Bitcoin returns at the lower to middle ($5^{th}$ to $60^{th}$) quantiles of the Sentiment Index and all quantiles of Bitcoin returns. Only after then, does it become negative from the middle to the uppermost ($60^{th}$ to $95^{th}$) quantiles of the Sentiment Index and all quantiles of Bitcoin returns. These outcomes imply that the sentiment induced by COVID-19-related news plays a more significant role in leading Bitcoin than other indices induced by coronavirus-related news.

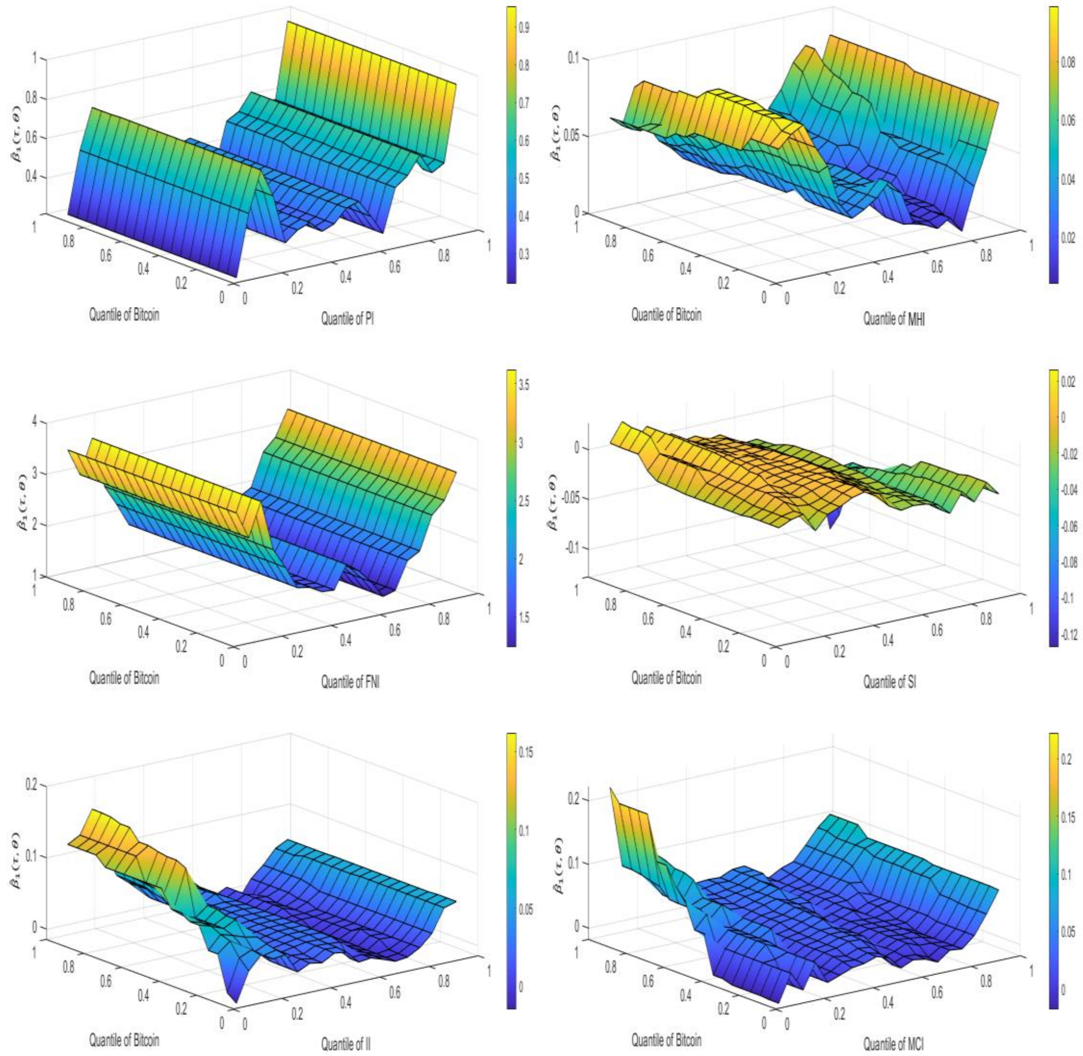


Figure 4.6. QQ estimates of the parameter $\beta_1(\tau, \theta)$. The response variable is the quantiles of Bitcoin returns

The graphs in Figure 4.7 show the cross-sectional dependence between the quantiles of daily news-based indices linked to COVID-19 outbreaks and the quantiles of daily gold returns. The visual pattern of the Panic Index shows a clear association between gold returns and the panic induced by coronavirus-related news. The value of gold plummeted sharply at the lowermost quantile ($5^{th}$ to $10^{th}$) of the Panic Index and the uppermost quantile ($80^{th}$ to $95^{th}$) of gold returns. The association between the Panic Index at lower quantiles ($10^{th}$ to $30^{th}$) and all quantiles of gold returns is displayed in blue (dark and light) and light green colors, indicating a gradual decline of the negative impact of panic on the value of gold. This negative impact becomes positive in the area combining the (30$^{th}$ to 80$^{th}$) quantiles of the Panic Index, followed by strong positive corresponding uppermost (80$^{th}$ to 95$^{th}$) quantiles of the Panic Index and all quantiles of gold. As a result, as the level of panic news rises, gold returns rise as well, implying that gold served as a safe haven and a hedge against media-induced panic.

The graphs of the Media Hype, Fake News, Infodemic, and Media Coverage Indices show that these indices have almost identical effects on gold returns at different quantiles. The strong negative effect of the MHI, FNI, II, and MCI indices on gold returns can be seen through the lowermost (5$^{th}$ to 25$^{th}$) quantiles of most of these indices. This negative effect gradually declines in the region that combines the (25$^{th}$ to 85$^{th}$) quantiles, after which it becomes strongly positive corresponding to the uppermost (85$^{th}$ to 95$^{th}$) quantiles. These findings suggest that gold acted as a hedge against Media Hype, Fake News, Infodemic, and Media Coverage Indices caused by coronavirus-related news.

The links between the Sentiment Index and gold returns are strong positive in the region that combines the lowermost (5$^{th}$ to 25$^{th}$) quantiles of both variables. They are weak positive in the area that combines the (25$^{th}$ to 60$^{th}$) quantiles of the Sentiment Index and nearly every quantile of gold returns. The links are negative in the grid that combines the 5$^{th}$ quantile of the Sentiment Index and the uppermost quartiles (60$^{th}$ to 95$^{th}$) of gold returns and the region that combines the (60$^{th}$ to 95$^{th}$) quantiles of the Sentiment Index as well as all quantiles of gold returns. This supports the prior outcome that media-induced sentiment plays a significant role in leading the value of gold more than other indices induced by coronavirus-related news. The findings obtained from Figures 4.6 and 4.7 confirm that both Bitcoin and gold can act as a hedge against the daily index of level of pandemic news.

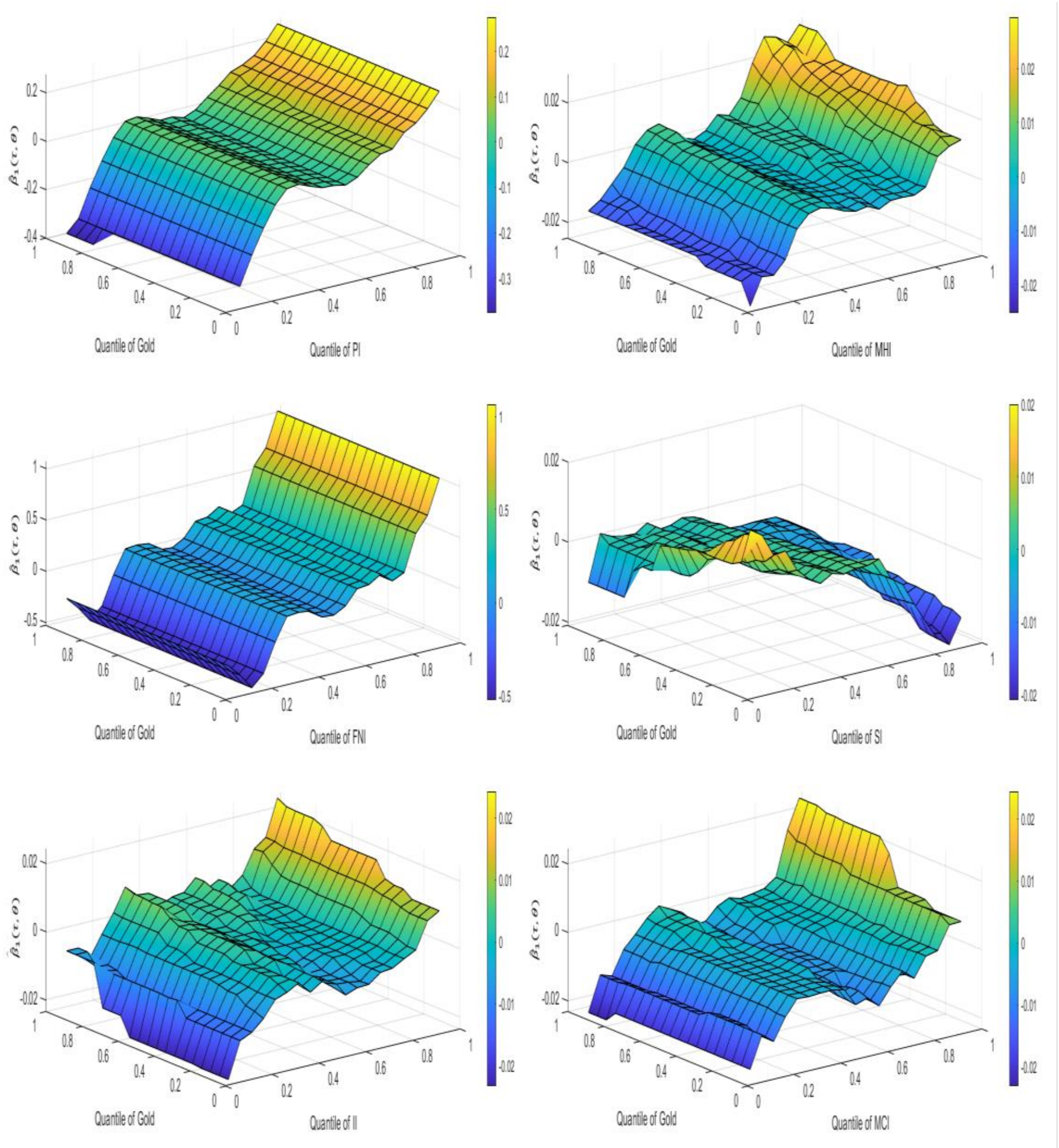


Figure 4.7. QQ estimates of the parameter $\beta_1(\tau, \theta)$. The response variable is the quantiles of gold returns

Examining the graphical depiction of the indices generated by the COVID-19 pandemic news in Figures 4.8 and 4.9 shows almost identical patterns of predominantly weak and negative associations with Bitcoin (and gold) returns for the great majority of combinations of various quantiles. The association between the indices at lower quantiles and all quantiles of Bitcoin (and gold) returns is represented by blue (dark and light) and light green colors, showing a gradual decrease in the negative effect of the indices on the value of Bitcoin (and gold). However, this negative effect turns positive in the area that combines the upper-middle quantiles of the indices, followed by strong positive corresponding uppermost quantiles of the indices and almost all quantiles of Bitcoin (and gold). Thus, the increase in the level of indices induced by coronavirus-related news led to an increase in the returns of Bitcoin (and gold), which suggests that Bitcoin (and gold) performed as a safe havens and a hedge against the daily index of the level of pandemic news. The figures of the estimated parameter $\hat{\beta}_2(\tau)y_{t-1}$ of the quantile-on-quantile are posted in Appendix D.

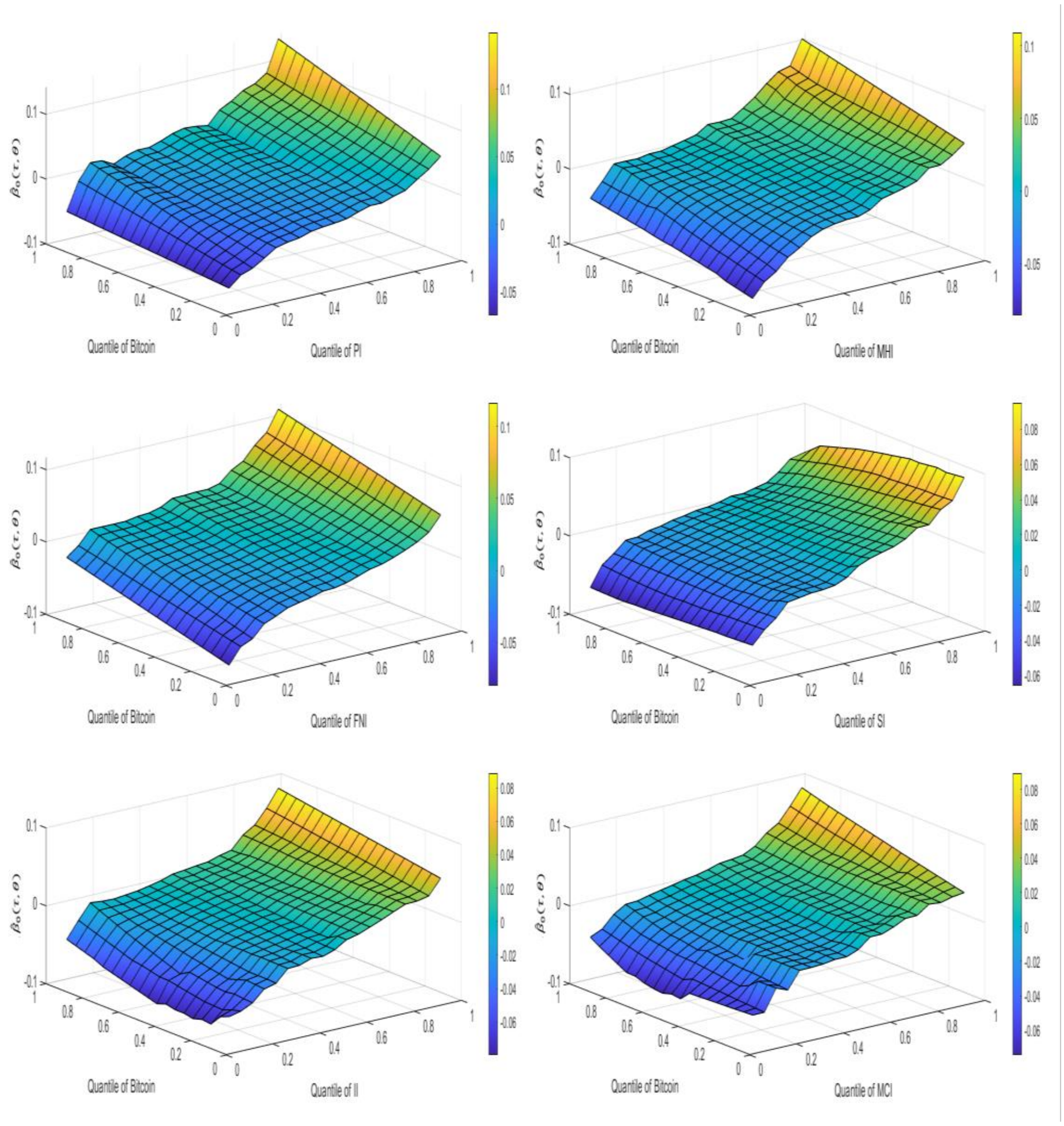


Figure 4.8. QQ estimates of the parameter $\beta_0(\tau,\theta)$. The response variable is the quantiles of Bitcoin returns

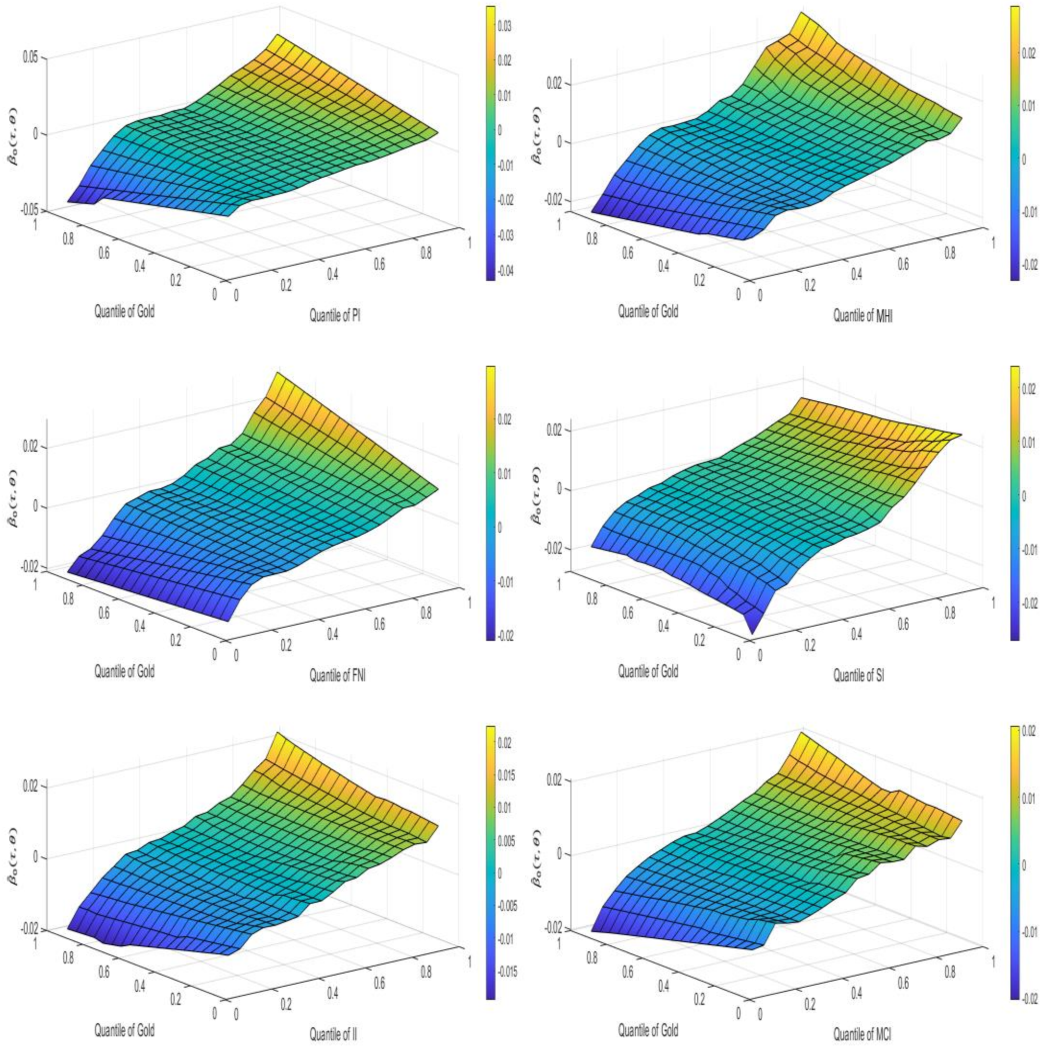


Figure 4.9. QQ estimates of the parameter $\beta_0(\tau, \theta)$. The response variable is the quantiles of gold returns

*4.2.4 Robustness Check for the QQR Approach*

Based on Equation (28), the length of the quantiles vector (Q) is 19, and the vector of the quantile levels ($\theta$) is $[0.05, 0.10, \dots, 0.95]$. In other words, (Q) is the number of quantiles within the vector of the quantile levels ($\theta$). The estimated parameter $\hat{\beta}_1(\tau)$ of the QR model are obtained by regressing the $\tau^{th}$ quantile of Bitcoin or gold returns on the news-based indices associated with COVID-19 outbreaks. Meanwhile, the estimated parameters $\hat{\beta}_1(\tau, \theta)$ of the QQR are obtained by regressing the $\tau^{th}$ quantile of Bitcoin or gold returns on the $\theta^{th}$ quantile of the news-based indices associated with COVID-19 outbreaks. Then the validity of the QQR approach is checked by comparing $\hat{\beta}_1(\tau)$ with the averaged of $\hat{\beta}_1(\tau, \theta)$.

Figures 4.10 and 4.11 compare the estimates of the coefficient $\beta_1(\tau)$ based on the QR and QQR approaches at various quantiles of Bitcoin and gold returns, respectively. The averaged QQR estimates of $\hat{\beta}_1(\tau, \theta)$ are nearly identical to the quantile regression estimates for all indices generated by the COVID-19 pandemic news regardless of the quantile considered, which validates the results that the QR estimates can be recovered from the QQR estimates. As a result, Figures 4.10 and 4.11 confirm the findings obtained from the QQR analysis reported earlier.

The resultant graphs for QQR and robustness of $\hat{\beta}_0(\tau, \theta)$ and $\hat{\beta}_2(\tau, \theta)$ are reported in Appendix E as Figures 5.3, 5.4, 5.5, and 5.6, correspondingly. No essential differences were observed between the graphs in Figures 4.10 and 4.11 and those posted in Appendix E.

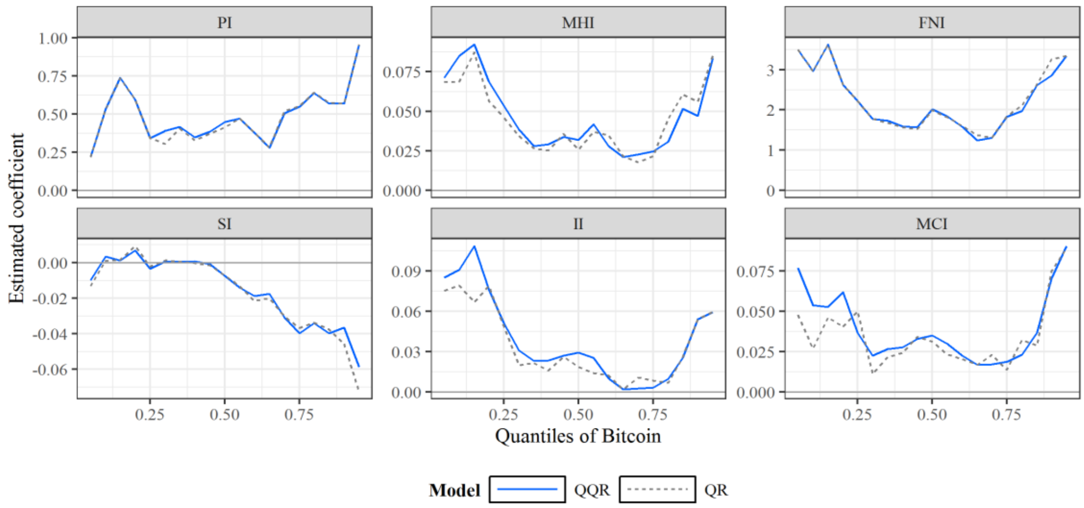


Figure 4.10. Comparison of $\hat{\beta}_1(\tau)$ estimated by QR (dashed black line) and the averaged of $\hat{\beta}_1(\tau, \theta)$ estimated by QQR (continuous blue line) at different quantiles of Bitcoin returns

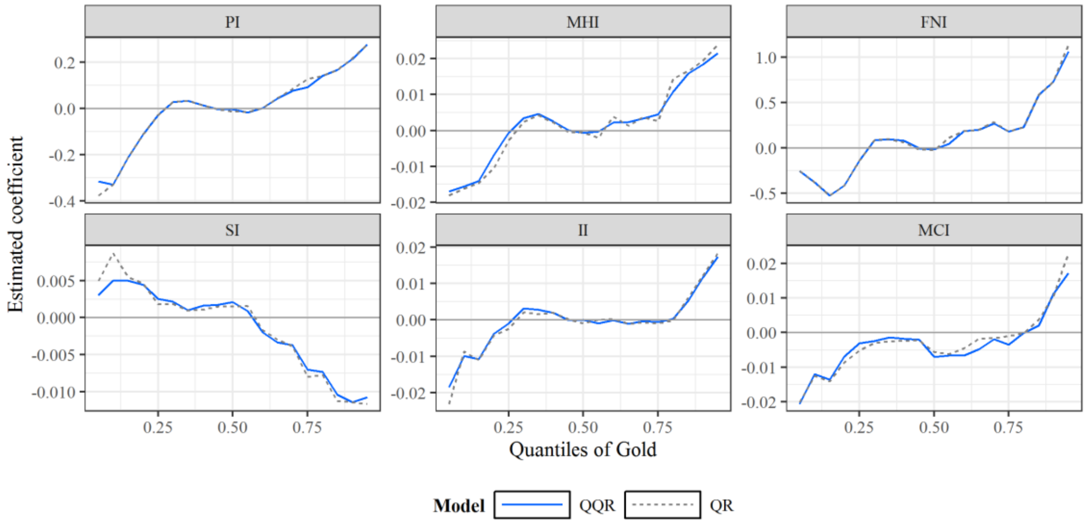


Figure 4.11. Comparison of $\hat{\beta}_1(\tau)$ estimated by QR (dashed black line) and the averaged of $\hat{\beta}_1(\tau,\theta)$ estimated by QQR (continuous blue line) at different quantiles of gold returns

### *4.2.5 Granger Causality Test*

Before applying the Granger causality test, the stationarity of the series must be checked. For this purpose, the augmented Dickey-fuller (ADF) test is used. The test shows that all of the variables are stationary except for the Sentiment Index, with respect to the p-value (based on $\alpha$=0.05). Therefore, the first differences for the Sentiment Index is applied.

The Granger causality test demonstrates unidirectional causality running from the Fake News Index to Bitcoin and gold returns. It also explores, a unidirectional causality from Bitcoin returns to the Panic Index and from the Infodemic Index to Bitcoin returns. Obviously, there is a bidirectional causality between the Media Hype Index and Bitcoin returns as well as between the Media Coverage Index and Bitcoin returns. On the other hand, there is no causality relationship between the Sentiment Index and Bitcoin (and gold) returns.

As a summary, the average daily Bitcoin and gold returns were around 0.003 and 0.001 respectively. The Media Coverage Index had the highest average of 69.65 percent, while the Sentiment Index had the lowest average of -16.45 percent. The JB test showed that the distributions of the variables are not normally distributed which is a solid reason to use the quantile regression (QR) technique. The QR model explored the effect of coronavirus-related news indices on Bitcoin and gold returns separately from the $5^{th}$ to $95^{th}$ quantiles. The results showed that the estimated intercept coefficient makes a significant contribution to the model. In almost all cases, the effects of COVID-19-related news and $\hat{\beta}_2(\tau)$ on gold returns are very weak while this is not the case for Bitcoin returns. The results obtained from the quantile-on-quantile regression (QQR) model showed that Bitcoin and gold acted as a hedge against indices caused by coronavirus-related news. The sentiment induced by COVID-19-related news plays a more significant role in driving Bitcoin and gold than other indices induced by coronavirus-related news. The robustness check for the QQR approach showed that the QR estimates can be recovered from the QQR estimates. The Granger causality test demonstrates a bidirectional causality between the Media Hype Index and Bitcoin returns as well as between the Media Coverage Index and Bitcoin returns.

# CHAPTER 5: CONCLUSION AND SUGGESTION FOR FUTURE RESEARCH

## 5.1 Conclusion

In the above study, quantile regression and quantile-on-quantile regression were used for modeling non-deterministic data, which are difficult to predict due to a lack of familiarity with the cause and effect relationship and are characterized by heavy tails. Quantile regression and quantile-on-quantile regression models have been applied in many different sectors and subjects, including financial data. The estimated coefficients through various quantiles of the dependent variable were derived using RStudio software.

The non-parametric quantile regression and quantile-on-quantile regression models were used to estimate the coefficients. A local linear quantile regression approach was used for smoothing the quantile regression curves based on the best bandwidth. The validity of the quantile-on-quantile regression approach was checked. The Jarque and Bera test was employed to test whether the data had skewness and kurtosis matching a normal distribution.

In the current information era, people might be inundated with a tremendous volume of real or fake news via various media means, including news related to the COVID-19 pandemic, which began at the end of December 2019, 100 years after the Spanish flu, and has spread rapidly over the world, causing havoc. In this thesis, the quantile regression and quantile-on-quantile regression approaches were employed to model the relationship between the RavenPack news-based index associated with the coronavirus pandemic and the returns of two commodities, Bitcoin and gold, over the period from January 23, 2020, to September 01, 2021. The descriptive statistics revealed that Bitcoin prices increased slowly during the beginning of the pandemic and then peaked remarkably until the middle of April 2021 before they steeply declined.

Meanwhile, gold prices had increased from the beginning of the pandemic till the middle of August 2020, and then they started to fluctuate and decline slowly. The Sentiment Index had the lowest average of -16.45 percent, while the Media Coverage Index had the highest average of 69.65 percent.

The quantile regression approach showed that the intercept coefficients had a significant contribution to the model. The quantile regression model failed to reveal the distribution of the Bitcoin or gold return based on the asymmetrical effects of the indices induced by the coronavirus news. Therefore, quantile-on-quantile regression was used as a robust approach to provide a more complete picture of dependence. The quantile-on-quantile regression model demonstrated the association between the quantiles of the coronavirus-related news and the quantiles of the Bitcoin or gold returns by a graphical depiction through the quantiles $(0.05, 0.10, \ldots, 0.95)$.

The quantile-on-quantile regression approach showed that there is a positive relationship between panic news associated with the COVID-19 and Bitcoin and gold returns. As the level of panic news rises, Bitcoin and gold returns rise as well, implying that both commodities serve as a safe haven and hedge against media-induced panic. The effects of Media Hype and Fake News Indices on Bitcoin and gold showed that the different levels of media-induced hype and fake news affected Bitcoin and gold returns asymmetrically. Both Bitcoin and gold performed as a hedge against excessive levels of media-induced hype and fake news.

The findings demonstrated that the sentiment induced by COVID-19 related news played a major role in driving the values of Bitcoin and gold more than other indices induced by coronavirus news. More sentiment induced by the media leads to higher returns. Therefore, the coronavirus-related news indices promoted investing in Bitcoin and gold as the previous findings showed that both commodities, Bitcoin and gold, can perform as a hedge against coronavirus-related news. The robustness of the quantile-on-quantile regression approach was checked. It validates the results that the quantile regression estimates can be recovered from the quantile-on-quantile estimates.

## 5.2 Suggestions for Future Research

This research focused only on Bitcoin and gold returns as dependent variables; future research can be extended to study the effect of coronavirus related news indices on other dependent variables like the stock market or oil price, and then to compare the worldwide results with other state results e.g., the state of Qatar. On the other hand, the properties of other parameters in the model can be studied as well as the impact of other factors on the response variable, such as coronavirus vaccine-related news indices, country-specific variables, etc. In this study, the Granger causality test was used to explore the behavior and directional causality of the variables (see Appendix F). For future research, this work may be extended to discuss the Granger causality test assumptions, lag selection, its validity, etc.

## REFERENCES


Adams, M. (Ed.). (2016). *Gold ore processing: project development and operations*. Elsevier.

Adhikari, R., & Agrawal, R. K. (2013). An introductory study on time series modeling and forecasting. *arXiv preprint arXiv:1302.6613*.

Banerjee, A. K., Akhtaruzzaman, M., Dionisio, A., Almeida, D. M., & Sensoy, A. (2021). Nonlinear nexus between cryptocurrency returns and COVID–19 news sentiment. *Available at SSRN 3923559*.

Barrodale, I., & Roberts, F. D. K. (1974). Solution of an overdetermined system of equations in the l 1 norm [F4]. *Communications of the ACM*, *17*(6), 319-320.

Bassett, G. W., & Koenker, R. W. (1986). Strong consistency of regression quantiles and related empirical processes. *Econometric Theory*, *2*(2), 191-201.

BBC. (2021, May 13). *Tesla will no longer accept bitcoin over climate concerns, says Musk*. https://www.bbc.com/news/business-57096305.

Bouri, E., Gkillas, K., Gupta, R., & Pierdzioch, C. (2021). Forecasting power of infectious diseases-related uncertainty for gold realized variance. *Finance Research Letters*, *42*, 101936.

Bouri, E., Gupta, R., Tiwari, A. K., & Roubaud, D. (2017). Does Bitcoin hedge global uncertainty? Evidence from wavelet-based quantile-in-quantile regressions. *Finance Research Letters*, *23*, 87-95.

Bowman, K. O., & Shenton, L. R. (1975). Omnibus test contours for departures from normality based on√ b 1 and b 2. *Biometrika*, *62*(2), 243-250.

Box, G. E. (1970). GM Jenkins Time Series Analysis: Forecasting and Control. *San Francisco, Holdan-Day*.

Broadstock, D. C., & Zhang, D. (2019). Social-media and intraday stock returns: The pricing power of sentiment. *Finance Research Letters*, *30*, 116-123.

Bunjaku, F., Gjorgieva-Trajkovska, O., & Miteva-Kacarski, E. (2017). Cryptocurrencies–advantages and disadvantages. *Journal of Economics*, *2*(1), 31-39.

Cepoi, C. O. (2020). Asymmetric dependence between stock market returns and news during COVID-19 financial turmoil. *Finance Research Letters*, *36*, 101658.

Chaudhuri, P. (1991). Nonparametric estimates of regression quantiles and their local Bahadur representation. *The Annals of statistics*, *19*(2), 760-777.

Chaudhuri, P., Doksum, K., & Samarov, A. (1997). On average derivative quantile regression. *The Annals of Statistics*, *25*(2), 715-744.

Chen, C., Liu, L., & Zhao, N. (2020). Fear sentiment, uncertainty, and bitcoin price dynamics: The case of COVID-19. *Emerging Markets Finance and Trade*, *56*(10), 2298-2309.

*Coronavirus Media Monitor*. (2021). RavenPack. https://coronavirus.ravenpack.com/

Crosby, M., Pattanayak, P., Verma, S., & Kalyanaraman, V. (2016). Blockchain technology: Beyond bitcoin. *Applied Innovation*, *2*(6-10), 71.

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association*, *74*(366a), 427-431.

Dyhrberg, A. H. (2016). Hedging capabilities of bitcoin. Is it the virtual gold?. *Finance Research Letters*, *16*, 139-144.

Fan, J., & Gijbels, I. (1992). Variable bandwidth and local linear regression smoothers. *The Annals of Statistics*, 2008-2036.

Fedorová, D., & Arltová, M. (2016). Selection of unit root test on the basis of length of the time series and value of ar (1) parameter. *Statistika*, *96*(3), 3.

Frey, B. B. (Ed.). (2018). *The SAGE encyclopedia of educational research, measurement, and evaluation*. Sage Publications.

Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: journal of the Econometric Society*, 424-438.

Härdle, W. K., Harvey, C. R., & Reule, R. C. (2019). Understanding cryptocurrencies. *Journal of Financial Econometrics*, *18*(2), 181-208.

Haroon, O., & Rizvi, S. A. R. (2020). COVID-19: Media coverage and financial markets behavior—A sectoral inquiry. *Journal of Behavioral and Experimental Finance*, *27*, 100343.

He, X., Fu, B., & Fung, W. K. (2003). Median regression for longitudinal data. *Statistics in medicine*, *22*(23), 3655-3669.

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.

Iqbal, N., Fareed, Z., Wan, G., & Shahzad, F. (2021). Asymmetric nexus between COVID-19 outbreak in the world and cryptocurrency market. *International Review of Financial Analysis*, *73*, 101613.

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

Jarque, C. M., & Bera, A. K. (1987). A test for normality of observations and regression residuals. *International Statistical Review/Revue Internationale de Statistique*, 163-172.

Kakinaka, S., & Umeno, K. (2021). Cryptocurrency market efficiency in short-and long-term horizons during COVID-19: An asymmetric multifractal analysis approach. *Finance Research Letters*, 102319.

Kam, K. M. (2014). *Stationary and non-stationary time series prediction using state space model and pattern-based approach*. The University of Texas at Arlington.

Kim, J. M., Jun, C., & Lee, J. (2021). Forecasting the volatility of the cryptocurrency market by GARCH and Stochastic Volatility. *Mathematics*, *9*(14), 1614.

Kim, J. M., Kim, S. T., & Kim, S. (2020). On the relationship of cryptocurrency price with us stock and gold price using copula models. *Mathematics*, *8*(11), 1859.

Klein, T., Thu, H. P., & Walther, T. (2018). Bitcoin is not the New Gold–A comparison of volatility, correlation, and portfolio performance. *International Review of Financial Analysis*, *59*, 105-116.

Koenker, R. (2005). Quantile regression Cambridge University Press New York.

Koenker, R. (2017). Quantile regression: 40 years on. *Annual Review of Economics*, *9*, 155-176.

Koenker, R. W., & d'Orey, V. (1987). Algorithm AS 229: Computing regression quantiles. *Applied statistics*, 383-393.

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

Koenker, R., & Bassett Jr, G. (1982). Robust tests for heteroscedasticity based on regression quantiles. *Econometrica: Journal of the Econometric Society*, 43-61.

Koenker, R., & Bilias, Y. (2002). Quantile regression for duration data: A reappraisal of the Pennsylvania reemployment bonus experiments. *Economic applications of quantile regression* (pp. 199-220). Physica, Heidelberg.

Koenker, R., & Machado, J. A. (1999). Goodness of fit and related inference processes for quantile regression. *Journal of the american statistical association*, *94*(448), 1296-1310.

Koenker, R., & Park, B. J. (1996). An interior point algorithm for nonlinear quantile regression. *Journal of Econometrics*, *71*(1-2), 265-283.

Koenker, R., & Xiao, Z. (2006). Quantile autoregression. *Journal of the American statistical association*, *101*(475), 980-990.

Lemonte, A. J., & Moreno-Arenas, G. (2020). On a heavy-tailed parametric quantile regression model for limited range response variables. *Computational Statistics*, *35*(1), 379-398.

Loader, C. R. (1999). Bandwidth selection: classical or plug-in?. *The Annals of Statistics*, *27*(2), 415-438.

Mahdi, E., Leiva, V., Mara'Beh, S., & Martin-Barreiro, C. (2021). A new approach to predicting cryptocurrency returns based on the gold prices with support vector machines during the COVID-19 pandemic using sensor-related data. *Sensors*, *21*(18), 6319.

Marasinghe, D. (2014). Quantile regression for climate data.

Mariana, C. D., Ekaputra, I. A., & Husodo, Z. A. (2021). Are Bitcoin and Ethereum safe-havens for stocks during the COVID-19 pandemic?. *Finance research letters*, *38*, 101798.

Mnif, E., & Jarboui, A. (2021). COVID-19, bitcoin market efficiency, herd behaviour. *Review of Behavioral Finance*.

Mnif, E., Jarboui, A., & Mouakhar, K. (2020). How the cryptocurrency market has performed during COVID 19? A multifractal analysis. *Finance research letters*, *36*, 101647.

Ruppert, D., Sheather, S. J., & Wand, M. P. (1995). An effective bandwidth selector for local least squares regression. *Journal of the American Statistical Association*, *90*(432), 1257-1270.

Shahzad, S. J. H., Bouri, E., Roubaud, D., Kristoufek, L., & Lucey, B. (2019). Is Bitcoin a better safe-haven investment than gold and commodities?. *International Review of Financial Analysis*, *63*, 322-330.

Shahzad, S. J. H., Shahbaz, M., Ferrer, R., & Kumar, R. R. (2017). Tourism-led growth hypothesis in the top ten tourist destinations: New evidence using the quantile-on-quantile approach. *Tourism Management*, *60*, 223-232.

Shi, Y., & Ho, K. Y. (2021). News sentiment and states of stock return volatility: Evidence from long memory and discrete choice models. *Finance Research Letters, 38*, 101446.

Sim, N., & Zhou, H. (2015). Oil prices, US stock return, and the dependence between their quantiles. *Journal of Banking & Finance, 55*, 1-8.

Smales, L. A. (2019). Bitcoin as a safe haven: Is it even worth considering?. *Finance Research Letters*, *30*, 385-393.

Soroka, S. N. (2006). Good news and bad news: Asymmetric responses to economic information. *The journal of Politics*, *68*(2), 372-385.

*Stock market quotes & financial news*. (2021). http://www.investing.com/

Sun, Y., Bao, Q., & Lu, Z. (2021). Coronavirus (Covid-19) outbreak, investor sentiment, and medical portfolio: Evidence from China, Hong Kong, Korea, Japan, and US. *Pacific-Basin Finance Journal*, *65*, 101463.

Thadewald, T., & Büning, H. (2007). Jarque–Bera test and its competitors for testing normality–a power comparison. *Journal of applied statistics, 34*(1), 87-105.

Urquhart, A. (2018). What causes the attention of Bitcoin?. *Economics Letters*, *166*, 40-44.

Urquhart, A., & Zhang, H. (2019). Is Bitcoin a hedge or safe haven for currencies? An intraday analysis. *International Review of Financial Analysis*, *63*, 49-57.

Urzúa, C. M. (1996). On the correct use of omnibus tests for normality. *Economics Letters*, *53*(3), 247-251.

Xue, X., Xie, X., & Strickler, H. D. (2016). A censored quantile regression approach for the analysis of time to event data. *Statistical Methods in Medical Research*, *27*(3), 955-965.

Yu, K., & Jones, M. (1998). Local linear quantile regression. *Journal of the American statistical Association*, *93*(441), 228-237.

Yu, K., & Stander, J. (2007). Bayesian analysis of a Tobit quantile regression model. *Journal of Econometrics*, *137*(1), 260-276.

Zhang, G. P. (2007). A neural network ensemble method with jittered training data for time series forecasting. *Information Sciences*, *177*(23), 5329-5346.

# APPENDIX

## Appendix A: Time Series

A time series is a set of data that is observed over a period of time. It is represented by a collection of vectors, $\boldsymbol{x}(t), t = 0,1,2, \dots d$, where $t$ reflects the amount of time that has passed, $d$ is the last period of time, and the variables $\boldsymbol{x}(t)$ are defined as random variables. In a time series, the measurements obtained during an event are grouped in chronological order and can be continuous or discrete. The nature of the time series is considered as a non-deterministic.

A time series is influenced by four primary components, which can be isolated from the data of interest. These are the elements: general trend, seasonal variations, cyclical variation, and irregular fluctuations. The general trend is the long-term movement in the series, such as an increase, or decrease. The seasonal variations are the fluctuations in the time series within a year that might be repeated each year. The cyclical variation is the change in the series over a period of time longer than one year. It is caused by circumstances that lead to a repetition in cycles. Irregular variations are produced by uncontrollable factors and do not repeat in a certain pattern. These variations have occurred due to unforeseen incidences such as the COVID-19 pandemic. The effects of these components can be considered using multiplicative and additive decomposition models. The additive model is preferred when the four elements of the time series seem to be independent of each other:

$$Y_t = T_t + S_t + C_t + I_t. \tag{30}$$

On the other hand, the multiplicative model is applied when the elements of the time series are seem to be dependent of each other:

$$Y(t) = T_t \times S_t \times C_t \times I_t, \tag{31}$$

where $Y(t)$ is the original data at time $t$, and $T_t$, $S_t$, $C_t$, and $I_t$ are the trend value, seasonal variations, cyclical variation, and irregular fluctuations at time $t$ respectively (Adhikari & Agrawal, 2013).

Time series analysis has become a commonly used method in various fields, including finance, engineering, medicine, and science. Depending on the nature of the study and the practical need, there are many different types of time series. A time series is commonly represented by a line graph, where the observations are displayed against the corresponding time to visualize the underlying structure of the data. Time series forecasting involves analyzing the previous observations to construct a theoretical model that captures the underlying data generation procedure for the series (Zhang, 2007). The model is then used to forecast future events. Based on the forecast results, valuable strategic decisions and preventative measures are implemented (Kam, 2014).

Stationary models are an important concept in the analysis of time series. In stationary models, the process is statistically stable, with probabilistic features that do not change over a period of time, particularly varying around a fixed mean and a constant variation. The most important models in the linear processes that are used for forecasting the future events of a time series are the autoregressive, moving average, and autoregressive moving average models (Adhikari & Agrawal, 2013).

Appendix B: Quantiles Estimation

According to Koenker (2005) the quantiles are found by minimizing the expected value of $\rho_\tau(X - \delta)$ with respect to $\delta$:

$$E\big(\rho_\tau(X-\delta)\big) = \int_{-\infty}^{+\infty} \rho_\tau(X-\delta)\, dF(x)$$

$$= (\tau - 1) \int_{-\infty}^{\delta} (x-\delta)\, dF(x) + \tau \int_{\delta}^{+\infty} (x-\delta)\, dF(x). \qquad (32)$$

This is then differentiated with respect to $\delta$ as follows:

$$\frac{d}{d\delta}\big[E\big(\rho_\tau(X-\delta)\big)\big] = \frac{d}{d\delta}\int_{-\infty}^{+\infty} \rho_\tau(X-\delta)\, dF(x)$$

$$= (\tau-1)\frac{d}{d\delta}\int_{-\infty}^{\delta}(x-\delta)\, dF(x) + \tau\frac{d}{d\delta}\int_{\delta}^{+\infty}(x-\delta)\, dF(x)$$

$$= (\tau-1)\,\frac{d}{d\delta}\left[\int_{-\infty}^{\delta} x dF(x) - \delta\int_{-\infty}^{\delta} dF(x)\right] - \tau\frac{d}{d\delta}\left[\int_{+\infty}^{\delta} x dF(x) - \delta\int_{+\infty}^{\delta} dF(x)\right]$$

$$= (\tau-1)\left[\delta f(\delta) - \delta f(\delta) - 1\int_{-\infty}^{\delta} dF(x)\right] - \tau\left[\delta f(\delta) - \delta f(\delta) - 1\int_{+\infty}^{\delta} dF(x)\right]$$

$$= (\tau-1)\,[-F(\delta)] - \tau[1 - F(\delta)]$$

$$= -\tau F(\delta) + F(\delta) - \tau + \tau F(\delta)$$

$$= F(\delta) - \tau, \qquad (33)$$

the minimum value of the loss function can be computed by using a unique value of $\delta$

that makes Equation (33) equal to zero. This is because the second derivative of the expected value of $\rho_\tau(X-\delta)$ is the probability density function of $\delta$. Therefore, minimizing the $\rho_\tau$ applied to the residuals leads to estimating the quantiles of the response variable. Generally, the $F(x)$ is unknown. Therefore, it is estimated using the empirical (CDF), and it is obtained using the sample observations:

$$F_n(x) = \sum_{i=1}^{n} I(x_i \le x). \tag{34}$$

Next, minimize the expectation of $\rho_\tau(X-\delta)$ with the empirical distribution

$$E\big(\rho_\tau(X-\delta)\big) = \int_{-\infty}^{+\infty} \rho_\tau(X-\delta)\, dF_n(x) = \frac{1}{n}\sum_{i=1}^{n} \rho_\tau(X-\delta), \tag{35}$$

where $1/n$ is a constant term, and the expectation in the above equation is minimized, which is the same as minimizing $\sum_{i=1}^{n} \rho_\tau(X-\delta)$.

Consider $R(\delta) = \sum_{i=1}^{n} \rho_\tau(X-\delta)$. Assume that the optimal exists at a certain point $\hat{\delta}$. This occurs when both the right and left derivatives of $R$ are positive at a point $\hat{\delta}$. The quantiles are expressed as a solution to the optimization problem.

Appendix C: Coronavirus Pandemic News Indices

Table 5.1. The Description of News-Based Indices Associated with Coronavirus Outbreaks

| Variable | Description |
|---|---|
| Coronavirus Panic Index (PI) | The values range from 0 to 100, with a value of 2 indicating that panic and COVID-19 are mentioned in 2% of all global news. |
| Coronavirus Media Hype Index (MHI) | The values range from 0 to 100, with a value of 75 indicating that COVID-19 is mentioned in 75% of all global news. |
| Coronavirus Fake News Index (FNI) | The values range from 0 to 100, with a value of 2.00 indicating that fake news and COVID-19 are mentioned in 2% of all news worldwide. |
| Coronavirus Worldwide Sentiment Index (SI) | The values range from -100 to 100. The most positive sentiment is 100, the most negative is -100, and 0 is neutral. |
| Coronavirus Infodemic Index (II) | The values range from 0 to 100, with a value of 60 indicating that 60% of all entities reported by the media are associated with or co-mentioned with the coronavirus. |
| Coronavirus Media Coverage Index (MCI) | The values range from 0 to 100, with a value of 60 indicating that 60% of the sample news presenters are currently reporting stories about the coronavirus. |

The description of the variables is taken from the RavenPack website: https://coronavirus.ravenpack.com/

Appendix D: Quantile-on-Quantile Estimates of the Parameter $\beta_2(\tau)y_{t-1}$

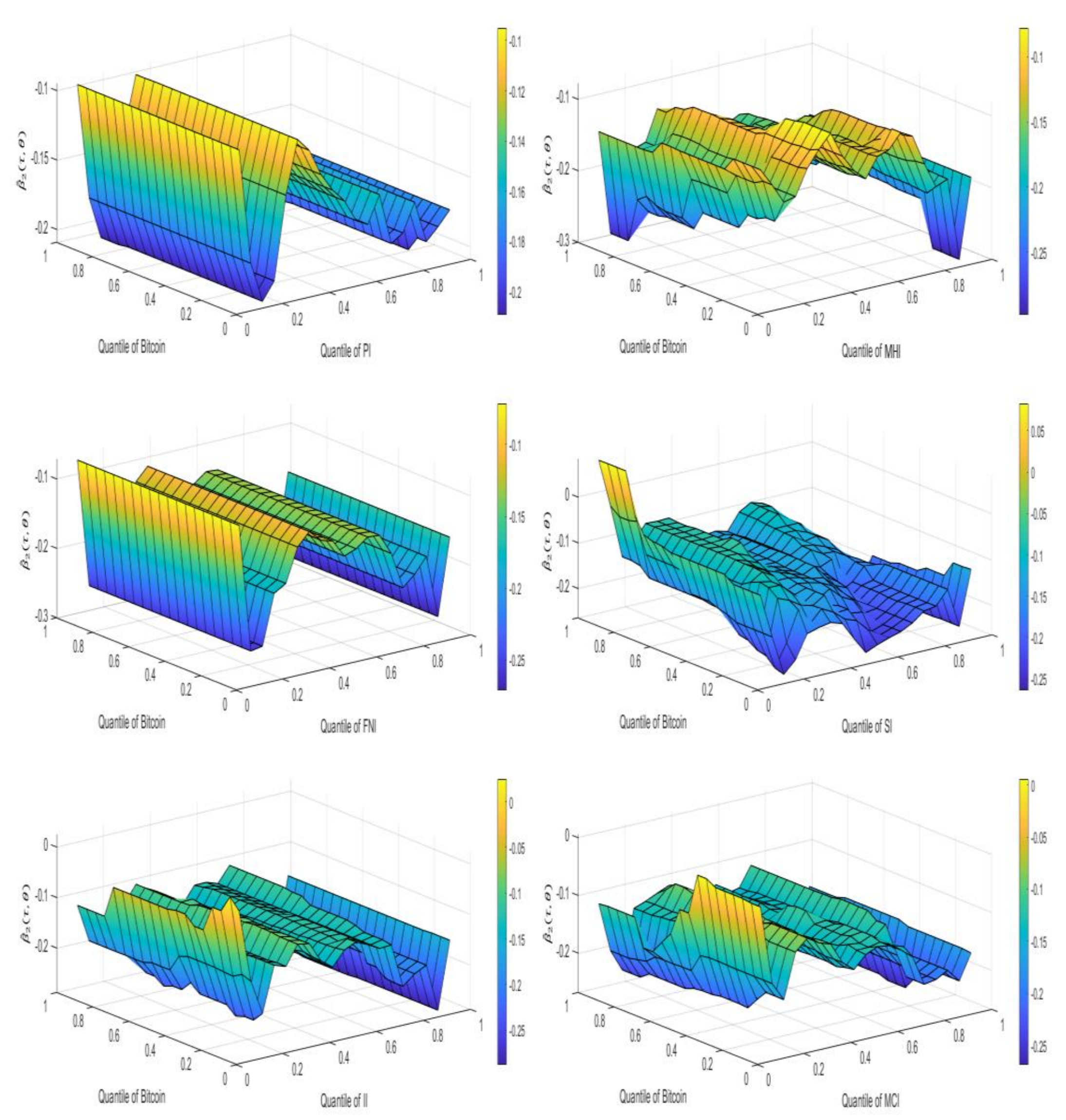


Figure 5.1. QQ estimates of the parameter $\beta_2(\tau)y_{t-1}$. The response variable is the quantiles of Bitcoin returns

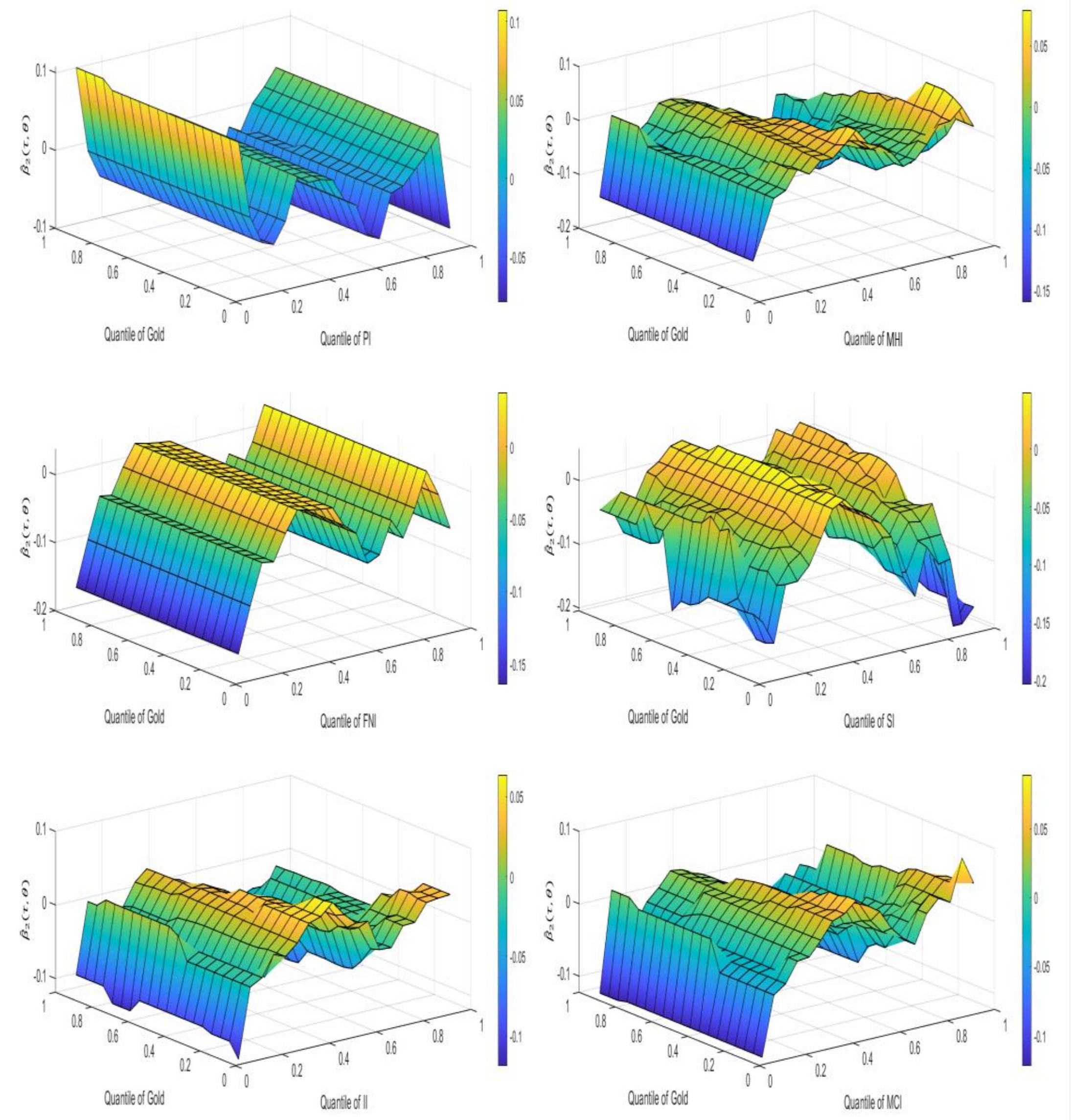


Figure 5.2. QQ estimates of the parameter $\beta_2(\tau)y_{t-1}$. The response variable is the quantiles of gold returns

## Appendix E: Comparison of Estimated Parameters by QR and the Averaged Estimated Parameters by QQR at Different Quantiles of Bitcoin and Gold Returns

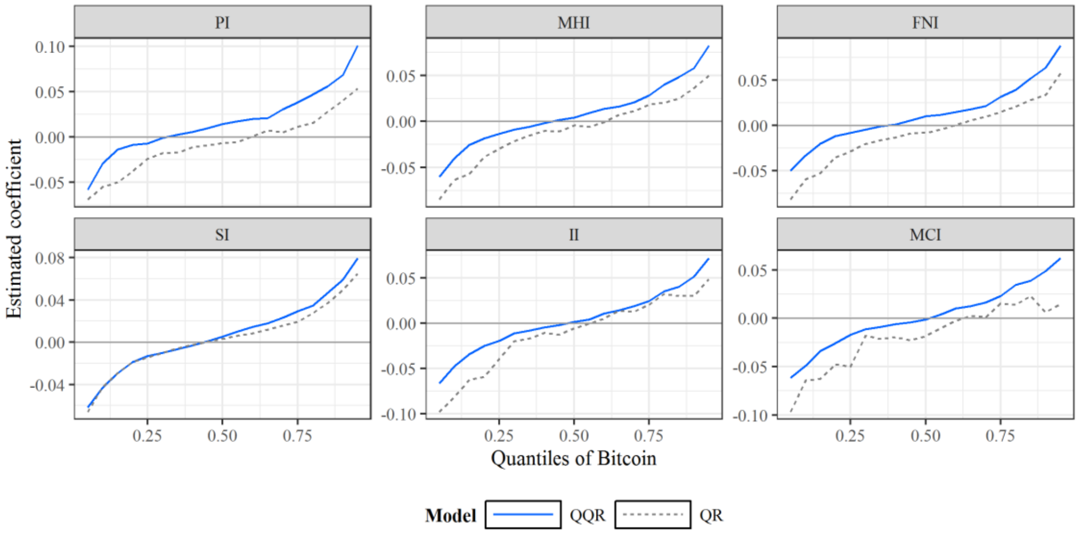


Figure 5.3. Comparison of $\hat{\beta}_0(\tau)$ estimated by QR (dashed black line) and the averaged of $\hat{\beta}_0(\tau,\theta)$ estimated by QQR (continuous blue line) at different quantiles of Bitcoin returns

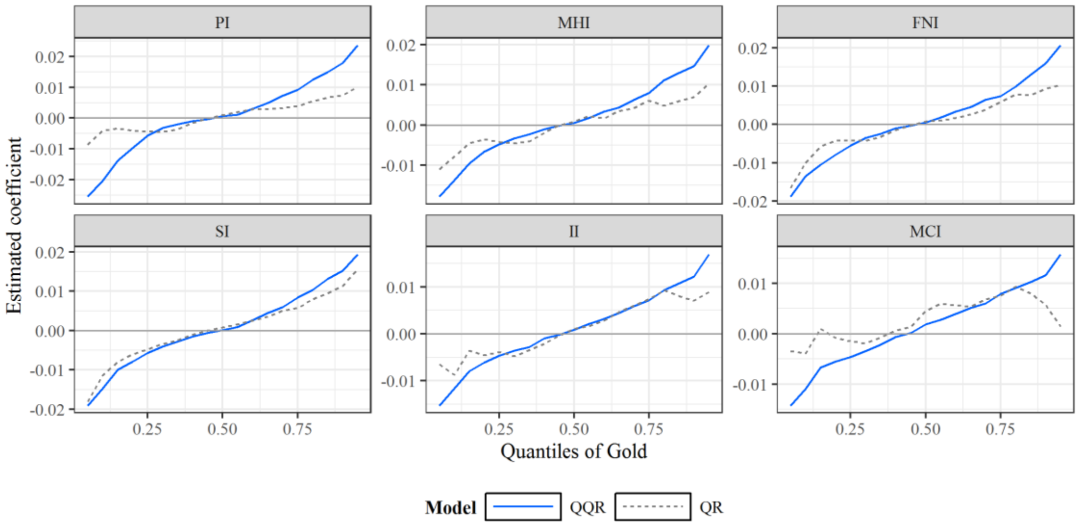


Figure 5.4. Comparison of $\hat{\beta}_0(\tau)$ estimated by QR (dashed black line) and the averaged of $\hat{\beta}_0(\tau,\theta)$ estimated by QQR (continuous blue line) at different quantiles of gold returns

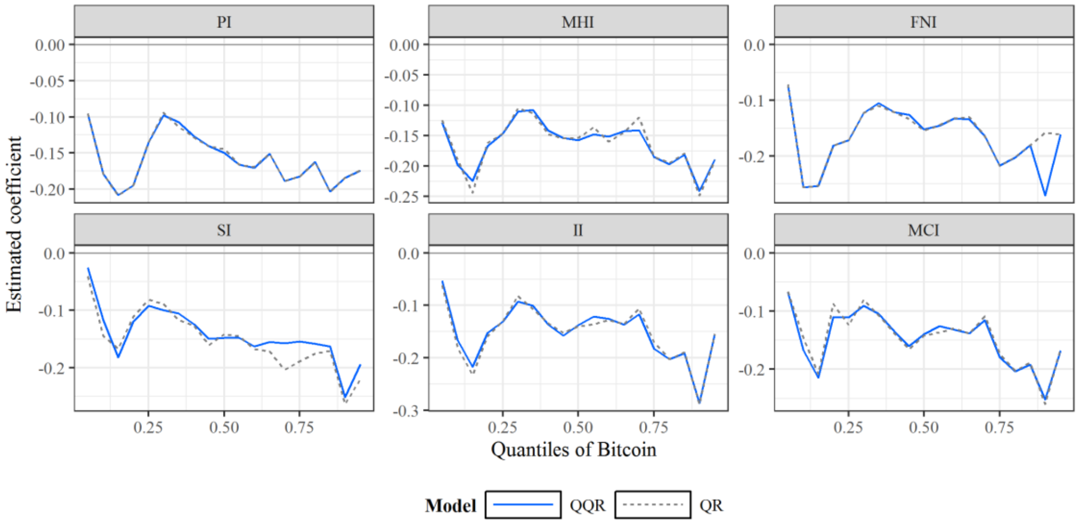


Figure 5.5. Comparison of $\hat{\beta}_2(\tau)$ estimated by QR (dashed black line) and the averaged of $\hat{\beta}_2(\tau,\theta)$ estimated by QQR (continuous blue line) at different quantiles of Bitcoin returns

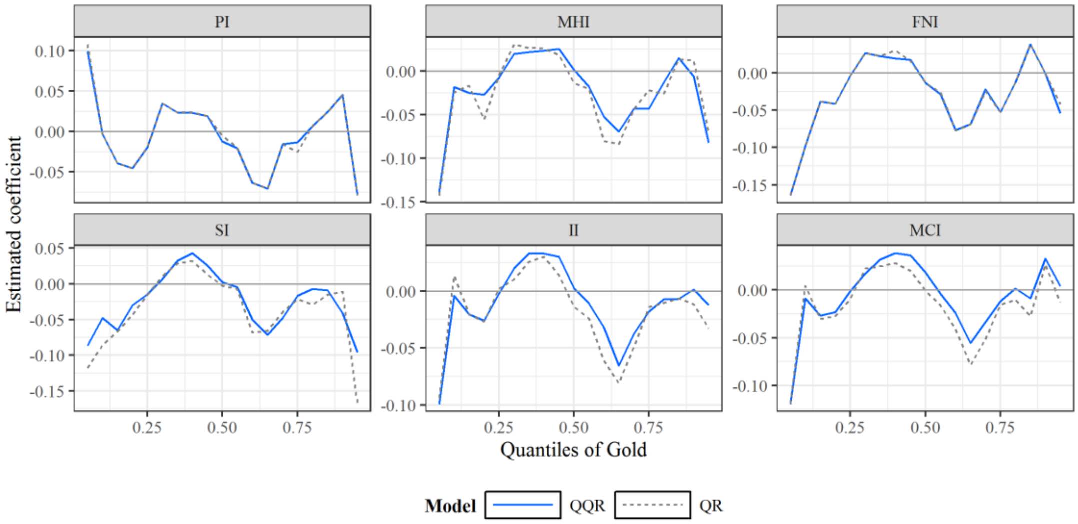


Figure 5.6. Comparison of $\hat{\beta}_2(\tau)$ estimated by QR (dashed black line) and the averaged of $\hat{\beta}_2(\tau,\theta)$ estimated by QQR (continuous blue line) at different quantiles of gold returns

Appendix F: Granger Causality Test

The analysis of causality is used to describe the behavioral and the directional causality of the financial markets. The concept of causality is provided primarily by (Granger, 1969). Granger causality is widely used in the financial field to assess bidirectional relations between variables. The series must be stationary to perform the Granger causality test between the variables $x$ and $y$. The augmented Dickey-Fuller (ADF) test is a common statistical test used to examine the stationarity of a time series (Dickey & Fuller, 1979). It is constructed based on the first order of the autoregressive model:

$$z_t = \emptyset z_{t-1} + \varepsilon_t, \tag{36}$$

where $\emptyset$ is the autoregression parameter and $\varepsilon_t$ is the non-systematic element that meets the attributes of the white noise process in the model (Box, 1970). Then the ADF test of the first order of the autoregressive model without a constant and without trend is given through the following equation:

$$\Delta z_t = \partial z_{t-1} + \varepsilon_t, \tag{37}$$

where $\Delta$ is the first difference operator and $\partial = \emptyset - 1$. If the null hypothesis is true, $H_0\colon \emptyset = 1$, then the process has a unit root and it is considered a non-stationary series (Fedorová & Arltová, 2016).

Generally, the variable $Y$ granger causes the variable $X$, with lags $m$ and $g$, when

$$F\left(x_t \middle| x_{t-1}^{(m)}, y_{t-1}^{(g)}\right) \neq F\left(x_t \middle| x_{t-1}^{(m)}\right), \tag{38}$$

the past values of the variable $Y$ can be used to explain the present value of the variable $X$, considering the past values of the variable $X$. Similarly, the variable $X$ granger causes the variable $Y$, with lags $m$ and $g$, when

$$F\left(y_t \middle| x_{t-1}^{(m)}, y_{t-1}^{(g)}\right) \neq F\left(y_t \middle| y_{t-1}^{(g)}\right). \quad (39)$$

the past values of the variable $X$ can be used to explain the present value of the variable $Y$, considering the past values of the variable $Y$. Granger causality is limited to the analysis of the linear effect between variables, which may lead to some loss of information when the relationship between the variables is based on nonlinear structures. (Banerjee et al., 2021).