# FORECASTING REALIZED VOLATILITY OF BIST INDICES WITH HAR-TYPE MODELS

EMRE AHMET KÖKSAL

YEDITEPE UNIVERSITY

GRADUATE SCHOOL OF SOCIAL SCIENCES

İSTANBUL, 2023

# FORECASTING REALIZED VOLATILITY OF BIST INDICES WITH HAR-TYPE MODELS

EMRE AHMET KÖKSAL

SUPERVISOR
PROF. DR. ŞÜKRİYE TÜYSÜZ

SUBMITTED TO GRADUATE SCHOOL OF SOCIAL SCIENCES
IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF ARTS
IN
DEPARTMENT OF FINANCIAL ECONOMICS

YEDITEPE UNIVERSITY

GRADUATE SCHOOL OF SOCIAL SCIENCES

İSTANBUL, 2023

# PLAGIARISM

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Date: 25.01.2023

Name/Surname: Emre Ahmet Köksal

# ÖZET

Bu tezde, seçimlik BIST endekslerinde gerçekleşen oynaklık, Heterojen Otoregresif Model (HAR) ve türevleri kullanılarak tahmin edilmiştir. Bu amaç için, 2001 ve 2021 yılları arasında oluşmuş fiyatlar kullanılarak 5 dakikalık getiri değerleri elde edilmiş, bu değerler gerçekleşen oynaklık değerleri ve diğer ölçümlerin hesaplanmasında temel olarak kullanılmıştır. Çalışmada 1 gün sonraki oynaklığın tahmini için kayan pencereler yöntemi kullanılmıştır. Sonrasında bu tahminler asıl gerçekleşen oynaklık değerleri ile kıyaslanmıştır. Tez, HAR cinsi modellerin kapsamlı bir karşılaştırmasını sunmakta olup tahmin için kullanılan verinin özelliklerinin önemini vurgulamaktadır. İlaveten, tezdeki bulgular endekslerde oynaklık tahmini hususunda ağırlıklar bağlamında dengesiz portföy oluşturma kaynaklı sorunların mevcudiyetine de işaret etmektedir. Sonuç olarak, HAR Modelleri Borsa İstanbul zaman serileri için başarılı bir tahminci olduğunu ispat etmiştir.

*Anahtar kelimeler: Borsa İstanbul, Gerçekleşen Oynaklık, HAR Modelleri, Oynaklık Tahmini, Yüksek Frekanslı Veri*

**ABSTRACT**

In this thesis, realized volatility of a selection of BIST Indices are forecasted with Heterogeneous Autoregressive Model (HAR) and its variations. For this purpose, ticks between 2001 and 2021 are used to generate 5-minute returns, which formed the basis for calculations of realized volatility and other realized measures. In the study, rolling windows are utilized for forecasting the volatility of one day ahead. These predictions are then compared to the actual realized volatilities. The thesis provides a thorough comparison of HAR-type models, and emphasizes the importance of underlying time series' characteristics in forecasting. Moreover, the findings of this thesis also hint at matters of diversification particular to index volatility forecasting. In overall, HAR Models proved to be a successful estimator for Turkish Stock Exchange time series.

*Keywords: Borsa İstanbul, HAR Models, High Frequency Data, Realized Volatility, Volatility Forecasting*

To my family

# ACKNOWLEDGMENTS

I wish to express my deepest gratitude to my supervisor Prof. Dr. Şükriye Tüysüz for her guidance, encouragement and feedback throughout the research.

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# 1. INTRODUCTION

The term volatility is an essential feature of finance. It denotes a financial security's degree of deviation from its measured, expected levels of return. Investment classes, therefore, are naturally expected to move up or down in price, while the unexpected rates of change at price levels have a say for the investors from all the complexity spectrum ranging from individual investors to institutions such as hedge funds. To this end, forecasting future volatility is crucial as it may help the decision-making process in portfolio allocation, assessing the risk versus return of holdings, and providing an anchor for various essential methods in financial risk management and derivatives pricing.

The introduction of volatility to the finance literature was no earlier than the 1950s. Although until that time, it was known that an investment opportunity should provide better returns if it were to involve higher risks, the conceptual framework it implies wasn't theorized until Markowitz's famous paper (1952). In his research, Markowitz defined and compared two possible views that an investor can have on the verge of portfolio making and concludes that a regular investor not only seeks the highest return generating portfolio among a specific selection of investment classes, but he also considers the highest returns adjusted with the risk it involves. Markowitz measured that risk as the variance of the portfolio at hand. The findings of this paper will later be marked as the Modern Portfolio Theory, where the risk, measured by volatility, lies in its core. The theoretical framework also paved the way for the Sharpe Ratio, which was put forth by Sharpe (1966) and has become an important parameter for investors while evaluating the comparative performance in regard to unit risks of different investment options such as mutual funds.

The journey of volatility did not stop at the portfolio optimization and benchmarking processes in finance. Investment managers and academic economists have always sought to include in their models the gauge of market anxiousness in the form of volatility, as can be seen in the famous Black-Scholes Model of derivative pricing, risk management tools such as Value at Risk (VAR), and methods to evaluate this anxiety such as CBOE Volatility Index, VIX. In the modern world of quantitative finance, volatility has a fundamental spot, and it is translated into every application with regard to the needs of that specific operation.

In general, volatility can take up two different definitions with a time-based perspective; historical and implied. Historical volatility, as the name suggests, turns to historical records (i.e. prices) of financial security at hand and tries to calculate an average out of differing time windows of, for example, 30, 60, 90 days, one year, or entire data available, to give an idea of what might be the future volatility. Implied volatility, on the other hand, uses strike prices of derivatives of a specific security to measure its future volatility. Since the implied volatility levels to an extent needs historical volatility as a benchmark for the volatility of a particular investment option, a security's historical values stand as vital ingredients for a mathematical formulation. At this point, the question revolved around how to handle this data at hand. Forecasting methodologies delved into many options, such as parametric models (ARCH, GARCH, Stochastic Volatility) and non-parametric ones, as finance literature on the topic grew. A recent trending approach is the calculation of realized variance by using high-frequency data. With the eventual evolution of computing performances to very high speed, the realized variance applications have made it into the spotlight and become the focus of attention for their capabilities in capturing and forecasting this deviation with exceptional precision.

This thesis uses high-frequency data for Turkish Stock Exchange to forecast future volatility. Variations of the Heterogeneous Autoregressive Model (HAR) are used to forecast the forward realized variance of the major indices of Borsa Istanbul; BIST 100 (XU100), BIST 30 (XU030), BIST Financials (XUMAL), BIST Industrials (XUSIN), BIST Banks (XBANK), BIST Chem. Petrol Plastic (XKMYA), BIST Food Beverage (XGIDA), BIST Basic Metal (XMANA), BIST Metal Products Mach. (XMESY), BIST Nonmetal Min. Product (XTAST), and BIST Textile Leather (XTEKS). The research covered 21-year data between 2001 and 2021, from which the realized volatilities of 5-minute prices of those indices regressed per chosen models with a rolling window approach. The thesis provides a thorough comparison of HAR-type models, and emphasizes the importance of underlying data's characteristics in forecasting. Moreover, the findings of this thesis also hint at matters of diversification particular to index volatility forecasting. Study also suggests that using HAR-type models demonstrate successful performance as an estimator of volatility according to a selection of test metrics.

## 2.LITERATURE REVIEW

The methodology of financial modelling of volatility forecasting has taken off by the ARCH Model by Engle (1982), where he conditioned the volatility of a time series with its previous values, thus highlighting the heteroskedastic nature of volatilities, i.e. non-stationarity in variances of a time series. The model was later extended to the GARCH model by Bollerslev (1986), in which the auto-regressive nature of the error term was also introduced to the already defined moving average component. There also have been various versions of GARCH and ARCH through the incorporation of thresholds or quadratic equations to account for the residual shocks.

The probability distributions of these models were assumed as normal distributions. However, literature in the area united on that the return distributions of financial assets display skews and fat-tails. There are probability distributions that take care of the excess kurtosis, such as Student's t distribution and generalized error distribution. There are also probability distributions for skewness problems, usually called skewed distributions. To handle both cases, there are skewed Student's t distribution and skewed generalized error distribution.

In the literature, various GARCH-type models were incorporated with different probability distributions; notable examples are Giot and Laurent (2003), Giot (2003), Bali and Theodossiou (2007) and Bali et al. (2008). The majority of studies on forecasting volatility with different GARCH variations with classical and skewed probability distributions showed the higher performance of asymmetric GARCH models (such as gjrGARCH, APARCH, TGARCH) as well as skewed probability distributions (such as skewed Student-t, skewed generalized error distribution). Moreover, there are studies (Angelidis et al., 2004; Braione & Scholtes, 2016; Giot & Laurent, 2003; Kuester et al., 2006) which also concluded that the choice of volatility model is less relevant than the choice of the probability distribution.

Volatility forecasting was also approached as a stochastic problem rather than a deterministic one, as GARCH or ARCH suggests. Stochastic volatility models, as they were termed, also emphasize time-varying volatility for a particular series, like GARCH models. The most popular stochastic volatility model (SV) was presented by Heston (1993), where volatility was assumed as a long-term mean-reverting process. The model also refers to volatility smile, a phenomenon for volatilities of option derivatives where marginal in-the-money or out-of-the-money values have the greatest implied volatilities.

Volatility, in its essence, is a proxy of market anxiety, a latent quantity. Price, unlike volatility, can be seen throughout a market session in a nearly continuous manner as transaction speed converges to split seconds. Drawing parallels from the price process to the problem of volatility forecasting, high-frequency data of price returns are utilized in a non-parametric framework. The volatility derived from this application was named as the realized volatility. The valuable nature of the high-frequency return data first drew the attention of Merton (1980), who argued for the instrumentality of the sum of squared high-frequency returns in forecasting conditional variance. Throughout that decade, high-frequency data has not seen interest possibly due to computing constraints in real-world practice. However, the last years of the millennium have seen increasing numbers of research papers in the area. Notable examples are Schwert (1998), who relied on 15-minute price returns for stock market volatility; Taylor and Xu (1997), who used 5-minute price returns of Deutsche Mark quotations to forecast hourly and daily volatility of returns; and Andersen, Bollerslev, Diebold, and Labys (2001), who used 5-minute returns of Deutsche Mark and Japanese Yen quotations, in the measurement of daily volatilities. The usage of high-frequency data for realized volatility is not confined to its suitability in forecasting, as underlined in various research papers (Andersen, Bollerslev, Diebold, & Labys, 2000; Andersen, Bollerslev, Diebold, & Ebens, 2001) where its abilities were pointed out in standardizing returns and its co-movement with correlation.

Starting with the new millennium, realized volatility (RV) has begun to be the subject of the focusing glass to be covered in depth in terms of pros and cons with different implementations. Barndorf-Nielsen and Shephard (2002) used RV in an SV model. Andersen et al. (2003) established a forecasting model using RV and compared it with traditional models such as GARCH and SV. Giot and Laurent (2004) analyzed RV to predict one day ahead value-at-risks. Finally, Martens et al. (2009) forecasted S&P 500 volatility with RV, underlining its long-memory characteristics.

The seminal paper of Corsi (2009) has had the most credit in modelling the RV for prediction purposes, as his research proposed a heterogeneous autoregressive model of RV named HAR-RV. The term heterogeneity is derived from the heterogeneous market hypothesis of Müller et al. (1993), wherein it refers to the heterogeneity of the market actors with different trading motives and time preferences. For example, day traders are more accountable for intra-day volatility. In contrast, position traders move their positions with weekly adjustments, and institutional investors act upon longer time horizons for their investments, thus creating volatility in different levels. Grounding on this intuition, Corsi parameterized daily, weekly, and monthly aggregated realized volatilities to forecast the one-day ahead volatility and compared this model against simple autoregressive models and ARFIMA, marking its success in predicting future volatility and capturing long-memory characteristics as good as its counterparts, while being effort-efficient.

Following this first step, the HAR model has seen variations to cover various real sample problems. For instance, Andersen et al. (2007) defined the daily RV in two parts; the continuous path of price volatility and spikes in that path named jumps. Therefore, they proposed a series of HAR models to consider the contribution of jumps and continuous paths together as well as treating jumps separately for the daily RV: HAR-RV-CJ and HAR-RV-J. Corsi and Reno (2012) proposed an extension to the model with a leverage component (Leverage HAR model or LHAR), which included negative returns as a distinct, independent variable in the model for future volatility. Finally, Bollerslev et al. (2016) tried to separate the integrated variance and error term in the realized variance formulation with the proposal of the HARQ model, pointing out the conditionality of the variance of the error term. Quarticity, as the paper named those errors, lies in the course of realized variance naturally and has a changing density in the process, in which sometimes it relieves and gives way to realized variance to forecast this integrated variance efficiently, and when in times it becomes noisier and makes the forecasting inaccurate with RV. The paper also included a selection of variations of this new model, wherein it incorporates continuous part (CHARQ model) or modelling of jumps (HARQ-J).

The literature on forecasting volatility dependent on realized measures and the HAR model has tested its validity in numerous experiments. Wang (2009) applied regular variations of the HAR model (HAR-RV, HAR-RV-CJ, HAR-RV-J) on the FTSE 100 index between September 2000 and August 2005 with 5-minute intervals and saw HAR-RV-CJ as the best performer in forecasting daily, weekly and monthly horizons. The study evaluated forecasting performances through Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Theil Inequality Coefficient (TIC). Todorova and Soucek (2014) studied the incorporation of overnight external market information (MSCI World Index) into a simple HAR model (HAR-RV) for ASX 200 index between March 2007 and January 2014 with 5-minute intervals. The study saw its benefits in forecasting daily horizon according to the Mincer-Zarnowitz $R^2$ Statistic (MZR), Mean Error (ME), MAE, Mean Relative Error (MRE) and RMSE. Agermark and Hoti (2016) have combined the HAR model with implied volatility to forecast foreign exchange rates of SEK against the US Dollar and Euro for a 1-week and 1-month horizon. The study involved 5-minute high-frequency data between January 2008 and February 2016. They tested HAR-RV, HAR-RV-CJ, and the addition of implied volatility (IV) as an estimator to the two models mentioned earlier, and concluded that there are benefits in adding IV as a parameter to the HAR models, depending on RMSE and MAE tests. Chen et al. (2018) have studied a HAR model variation with time-varying coefficients (TVC-HAR) on S&P 500 returns between May 1999 and October 2010 with 5-minute returns, comparing the subject model with simple HAR-RV, HAR-GARCH (1,1) (see Corsi et al., 2008), HAR-RV-J, HAR-RV-CJ, and HAR-RV-J with GARCH (1,1) for 1-day, 5-day and 22-day forecasting horizons on Conditional Predictive Ability (CPA) test. The research found out that although for a one-day horizon all the models show similar performances, the proposed model exhibits its benefits in weekly and monthly forecasting horizons. Özbekler et al. (2021) researched the predictive performance of HAR models (HAR-RV, HAR-RV-J, HAR-RV-CJ) in the European sovereign bond markets (UK, Germany, France, and Switzerland) using intraday data between January 2005 and October 2019 with 10-minute intervals and for 1-day, 5-day and 22-day forecasting horizons, finding a satisfying display for models which try to capture the effects of jumps, under Quasi–Likelihood loss function (QLIKE) and MZR. Kambouroudis et al. (2021) studied the extensions of the HAR model with implied volatility, overnight returns, leverage effects and conditional heteroskedasticity (HAR-IV, HAR-O, LHAR, HAR-G, and combinations of two or more features such as HAR-IVLO for both implied volatility, leverage and overnight returns) on the ten most significant stock indices (S&P 500, DJIA, Nasdaq 100, Euro STOXX 50, CAC 40, DAX, AEX, SMI, FTSE 100, Nikkei 225) from February 2001 to June 2019 with 5-minute returns, finding tangible benefits of implied volatility variations against the simple HAR model and other extensions, per QLIKE and Model Confidence Set (MCS).

There are also studies that review commodities and HAR Models together. Prokopczuk et al. (2016) have tested a family of HAR models' (HAR-RV, HAR-RV-J, HAR-RV-RJ (HAR Model with Realized Jumps, see Tauchen & Zhou, 2011), HAR-RV-ARJ (decomposition of realized jumps into negative and positive parts), HAR-RV-CJ) forecasting performance for 1-, 5- and 22-day ahead. The research was on four energy commodities (crude oil, gasoline, heating oil and natural gas) between January 2007 and June 2012 for 15-minute returns, concluding that the inclusion of jumps into models has limited effect on the forecasting performance of the HAR variations regarding six different loss functions (Mean Squared Error (MSE), Mean Squared Percentage Error (MSPE), MAE, MAPE, Logarithmic Loss (LL), and QLIKE). Tang et al. (2021) analyzed S&P 500 and WTI futures' 5-minute intraday returns between January 2007 and April 2017 with HAR-RV, an extension of the HAR-RV with S&P 500 RV and oil futures RV (HAR-RV-OIL), and VHAR-RV-OIL-DCC model; a vector HAR model with the dynamic conditional correlation of Engle (2002). They also repeated this modification for HAR-RV-TCJ (see Corsi et al., 2010), thus pooling a 6-model contest. The models are evaluated by their out-of-sample $R^2$ and MSPE scores. The research concluded that adding more variables to the simple HAR model has tangible benefits. Degiannakis et al. (2022) have compared the simple HAR model and its variations (HAR-RV, HAR-RV-J, HAR-RV-CJ, HAR-PS (see Patton & Shephard, 2015) for 1 to 66 days forward predictions, including leverage effects, continuous parts, and jumps on agricultural commodities (Corn, Rough Rice, Soybeans, Sugar, and Wheat) between January 2010 and June 2017 for 5-minute returns. They concluded that there is no use in adding complexity to the primary model as per the Mean Squared Prediction Error, MAPE, and QLIKE loss function tests. In a similar research on agricultural commodities (Corn, Cotton, Indica Rice, Palm oil and Soybean) between January 2011 and December 2015 with a 5-minute frequency to forecast one-day, one-week and one-month ahead, Luo et al. (2022) outperformed a selection of HAR models (HAR-RV, HAR-RV-CJ, HAR-RV-TCJ, HAR-RV-ΔJ (see Patton & Shephard, 2015, and Centralized HAR (see Bollerslev et al., 2018) by adding hidden Markov regime switching structures, based on the MZR, Mean Absolute Forecast Error (MAFE), and MCS.

The HAR model has also been tested on cryptocurrencies. Bergsli et al. (2022) compared traditional GARCH models (GARCH, EGARCH, IGARCH, GJR-GARCH, and APARCH) and the HAR-RV model on bitcoin volatility between January 2014 and September 2018 for 10-, 15-, and 30-minute frequencies and for forecasts on horizons of 1, 2, 5, 10, and 15 days, confirming the simple HAR model's superiority in terms of MSE, MAE, MAPE, QLIKE and MCS. Ftiti et al. (2021) advanced further by testing HAR-RV, HAR-RV-CJ, HAR-SRV, and HAR-RV-ΔJ between April 2018 and June 2020 for 5-minute returns, to capture the optimal forecasting performance on Bitcoin, Ethereum, Ethereum Classic, and Ripple. They found that including jumps into simple HAR model gives the best forecasting outcomes. The models are compared with MCS for daily, weekly, and monthly volatility forecasts.

There are few studies on Turkish high-frequency data for volatility modelling and forecasting. Çelik and Ergin (2014) were the first to utilize intraday values of BIST 30 and its futures contracts, at a frequency of 1, 5, 10 and 15 minutes, from February 2005 to April 2010. The dataset was used to compare volatility modelling performances of HAR-RV, HAR-RV-J, HAR-RV-CJ, MIDAS and traditional forecasting models of GARCH and RGARCH for the daily forecasting horizon, and found that with high-frequency data, HAR-RV-CJ and MIDAS models outperform their counterparts depending on RMSE, MAE, MAPE, and TIC tests. They also used logarithmic and square root transformations along with the original data to see the effects of a normal probability distribution of the data. Eroğlu et al. (2021) analyzed the BIST 100 index between April 2016 and January 2019, with the help of the HAR-RV, LHAR-RV, HAR-RV-CJ, and LHAR-RV-CJ model, to find out that the jumps in the return series are stationary and there is a leverage effect for the negative returns in small amounts. They used both the original 1-minute returns and log returns and discovered that the data has non-normal distribution with fat-tails. The comparison criteria for these models were RMSE, MAE, Mean Percentage Error (MPE) and MAPE for the one-day forward value of volatility for an out-of-sample period of the last year of the data. Finally, Türensal (2021) have utilized HAR-RV, HAR-RV-J, HARQ, HARQ-J, CHAR (Continuous HAR model with bipower variation, see Bollerslev et al., 2016), and CHARQ model of 5-minute log percentage return data of BIST 100 between the start of 2018 and August 2020, and found HAR-RV and CHAR as the best performers in out-of-sample data as per the QLIKE and MSE tests for one day forward forecasts. The research has found that the data at hand is not normally distributed owing to fat tails and skewness. The out-of-sample forecast period was the last month of the data and covered through a 20-day rolling window.

## 3.THEORETICAL FRAMEWORK AND METHODOLOGY

### 3.1. Theoretical Framework

#### 3.1.1. Realized Volatility (RV)

The price process of a sample financial asset $P_t$ can be determined by the stochastic differential equation:

$$dP_t = \mu_t d_t + \sigma_t \, dW_t, \tag{1}$$

Where $\mu_t$ and $\sigma_t$ denote the drift and the instantaneous volatility processes, respectively, and $W_t$ is a standard Brownian motion. From this formulation, the return for the $i$-th interval of a trading day can be defined as:

$$r_{t,i} = \; P_{t-1+i\Delta}/P_{t-1+(i-1)\Delta} - 1, \tag{2}$$

$$i = 1,2, \ldots, M,$$

Where $M$ equals $1 \,/\, \Delta$ to denote the sampling frequency. As this frequency goes to infinity, the formulation will start to cover the infinitesimal price movements, enabling to formulate the integrated volatility for a given day as:

$$IV_t = \int_{t-1}^{t} \sigma_s^2 \, \mathrm{d}s. \tag{3}$$

However, this application is impossible in normal conditions due to the discrete nature of the price process and market anomalies that appear as this process converges to continuous state. Instead, realized volatility RV formulation replaces to become a depiction of integrated volatility as:

$$RV_t = \; \sum_{i=1}^{M} r_{t,i}^2 \tag{4}$$

The $RV_t$ here is the summation of the square of intraday returns for a day $t$. As $M \rightarrow \infty$, the number of return calculations within a day creates enough samples to replicate and give an idea of the integrated volatility for a given day. As mentioned earlier, it is important to limit the number of return instances to a reasonable degree in order to provide a sound analysis that is not diluted with market inefficiencies. For that reason, this study involves a 5-minute intra-day returns in measurement of the realized volatility.

### 3.2. Forecasting Methodology

#### 3.2.1. HAR Model (HAR-RV)

In his landmark paper, Corsi (2009) had done an analysis of the RV with his proposed heterogeneous autoregressive model, named HAR-RV. Intuition in this model was derived from asymmetric relation between long-term and short-term volatility. For instance, day traders can be held accountable for daily volatility, but they base their actions on the course of long-term volatility. Therefore, anything affects the long-term structure will have its reflections on shorter terms, i.e. there is a presence of long-memory in financial time series. Corsi designed this phenomenon as a cascade of daily, weekly, and monthly RVs to forecast the one day ahead volatility:

$$RV_{t+1}^{d} = c + \beta^{d} RV_{t}^{d} + \beta^{w} RV_{t}^{w} + \beta^{m} RV_{t}^{m} + \varepsilon_{t+1}^{d}, \tag{5}$$

where $RV_t^d$ , $RV_t^w$ and $RV_t^m$ represent the daily, weekly and monthly realized volatility. The daily RV is the sum of squared returns in one day, whereas weekly and monthly realized volatility are the average of the past values of daily RVs according to the size of the time period. The model is estimated with an ordinary least squares method.

#### 3.2.2. HAR Model with Jumps (HAR-RV-J)

Andersen et al. (2007) separated the daily RV into two different parts; the continuous path of price volatility and spikes in that path named jumps. To consider and define the effects of jumps to future volatility, they proposed two HAR model variations, specified as HAR-RV-J and HAR-RV-CJ. The first model they proposed was HAR-RV-J, where:

$$RV_{t+1}^{d} = c + \beta^{d} RV_{t}^{d} + \beta^{w} RV_{t}^{w} + \beta^{m} RV_{t}^{m} + \beta^{jd} {J_t}^{d} + \varepsilon_{t+1}^{d}, \tag{6}$$

In addition, they defined Jumps ($J_t$) as:

$$J_t = max[RV_t - BPV_t; 0], \tag{7}$$

where $RV_t$ stands for the realized volatility, and $BPV_t$ stands for the bipower variation. Bipower variation (Barndorff-Nielsen & Shephard, 2004) is defined as:

$$BPV_t = \mu_1^{-2} \sum_{i=k+1}^{M} |r_{t,i}||r_{t,i-k}|, \tag{8}$$

$$k \geq 0,$$

where $\mu_1 = \sqrt{\frac{2}{\pi}} = E(Z)$ is a representation of the mean of the absolute value of the Standard Gaussian random variable.

As can be seen, main structure of the primary model was sustained with a slight addition of jump parameter with its coefficient. Jumps here formulated in line with the Barndorff-Nielsen and Shephard (2004); the maximum of the absolute values of difference between $RV_t$ and $BPV_t$ and zero. $BPV_t$ is deemed as robust to jumps, while $RV_t$ process includes both jumps (i.e. discontinuous) and continuous part.

### 3.2.3. HAR Model with Quarticity and Jumps (HARQJ)

In volatility literature, RV is viewed as a proxy of the integrated volatility (IV) process. Therefore, existence of measurement errors in RV causes attenuation bias in forecasting with HAR modelling of the volatility, particularly when the variances of measurement errors are high. To address this variability in error variance, Bollerslev et al. (2016) proposed to modify the daily realized volatility partition of the fundamental HAR model with quarticity and also included jumps, same as in the HAR-RV-J model of Andersen et al. (2007) that is mentioned previously:

$$RV_{t+1}^{d} = c + \left(\beta^{d} + \beta_{Q}^{d} RQ_t^{1/2}\right) RV_t^{d} + \beta^{w} RV_t^{w} + \beta^{m} RV_t^{m} + \beta^{jd} {J_t}^{d} + \varepsilon_{t+1}^{d} \tag{9}$$

Note that the equation in the first parenthesis takes place of the coefficient of the daily RV, and $RQ_t$ denotes the Realized Quarticity, which is formulated as:

$$RQ_t = \frac{M}{3} \sum_{i=1}^{M} r_{t,i}^{4}. \tag{10}$$

The $\beta^d$ (notation in the first parenthesis), $\beta^w$, $\beta^m$coefficients are same as basic HAR model, and $\beta_Q^d$ is coefficient for realized quarticity. As can be inferred, this coefficient enables a dynamic modelling for the last day's realized volatility. When the variance of the error variance is high for daily measures, the $\beta_Q^d$ coefficient gets below zero to reduce this day's effect on forecasted volatility, while the measurement error gets lower the coefficient tends to zero to permit realized volatility of that day to be accounted. The reason behind the modification's confinement to daily realized volatility is measurement errors are being averaged out in weekly and monthly levels, so that coefficients do not vary in significant amounts than the basic HAR model. In terms of RVs and ${J_t}^d$, design and notation of the model was defined identical to that of the original HAR-RV-J model.

### 3.2.4. Continuous HAR Model (CHAR)

Grounding on the findings of the Andersen et al. (2007), Bollerslev et al. (2016) also defined the Continuous HAR Model (CHAR) as:

$$RV_{t+1}^{d} = c + \beta^{d} BPV_{t}^{d} + \beta^{w} BPV_{t}^{w} + \beta^{m} BPV_{t}^{m} + \varepsilon_{t+1}^{d}, \qquad (11)$$

The model above uses bipower variation as the estimator for its robustness against jumps in the volatility process, thus providing as a valid regressor for continuous path of the volatility. The validation to this approach can be perceived from section 3.2.2.

### 3.2.5. Continuous HAR Model with Quarticity (CHARQ)

Bollerslev et al. (2016) also combined their CHAR model with quarticity by formulation:

$$RV_{t+1}^{d} = c + \left(\beta^{d} + \beta_{Q}^{d} TPQ_{t}^{1/2}\right) BPV_{t}^{d} + \beta^{w} BPV_{t}^{w} + \beta^{m} BPV_{t}^{m} + \varepsilon_{t+1}^{d}, \qquad (12)$$

in which BPV notation for daily, weekly, and monthly estimators are identical to the CHAR model. However, the model differentiates from HARQJ model in terms of its treatment towards quarticity. To constitute a jump robust estimator of quarticity, model uses Tripower Quarticity (TPQ) of Barndorff-Nielsen and Shephard (2006):

$$TPQ_{t} \equiv M\mu_{4/3}^{-3} \sum_{i=1}^{M-2} \left|r_{t,i}\right|^{4/3} \left|r_{t+1,i}\right|^{4/3} \left|r_{t+2,i}\right|^{4/3}, \qquad (13)$$

where $\mu_{4/3} \equiv 2^{2/3}\Gamma(7/6) / \Gamma(1/2) = E\left(|Z|^{4/3}\right)$.

### 3.3. Performance Evaluation Methodology

For performance comparison for forecasting methodologies, the Model Confidence Set (MCS) of Hansen et al. (2011) is employed. In its basis, the model aims to find the best fitting model or a selection of best fitting models by executing an iteration of tests until to the point where models that underperform are omitted and models that provide eligible results in predetermined confidence interval survives. It should also be noted that the initial set of models can stay as the best performing models with no worse performing model.

To set out the framework, let $M_0$ be the initial set of models to be tested. For an individual model $M^i$ in the initial set, the ultimate goal is to be among the best performing models, denoted as $M^*$. The first step of the evaluation of $M^*$ is calculations of loss functions for every instance in the forecasting period, revealing the contrasts between the actual value of the RV and the forecasted value for each model $M^i$. The differences of losses between two models $i, j$ in a particular time $t$ can be defined as:

$$d_{i,j,t} = l_{i,t} - l_{j,t} \tag{14}$$

where $i, j = 1,2, \dots, m$ and $t = 1,2, \dots, n$.

The above formulation can be rewritten to capture a sample wide computation for differentials as:

$$\bar{d}_{i.} = \bar{l}_i - \bar{l}_. \tag{15}$$

where $\bar{d}_{i.}$ stands for the difference between average sample loss for model $i$ ($\bar{l}_i = n^{-1} \sum_{t=1}^{n} l_{i,t}$) and average of all sample losses for all models ($\bar{l}_. = m^{-1} \sum_{i \in M_0} \bar{l}_i$).

To transform this calculation into a comparable statistic for all of the models, below formulation:

$$t_{i.} = \frac{\bar{d}_{i.}}{\sqrt{\widehat{var}(\bar{d}_{i.})}} \tag{16}$$

is used, where we defined the numerator part in previous equation, and $\widehat{var}(\bar{d}_{i.})$ denotes the estimate of the variance of $\bar{d}_{i.}$ This estimation was done by block-bootstrapping method. From this point, the null and the alternative hypotheses can be put forth as:

$$H_{0,M} : c_{i.} = 0, \text{ for all } i = 1,2, \dots, m$$
$$H_{A,M} : c_{i.} \neq 0, \text{ for some } i = 1,2, \dots, m \tag{17}$$

The $c_{i.}$ here denotes the $E(d_{i.})$ and is assumed to be finite and not time dependent.

To sum up, the forecasting models will first be compiled, and then compared as per their $t_{i.}$ values, until the null hypothesis holds for every model that is not eliminated, forming the $M^{*}$. The elimination procedure removes the model or models with maximum $t_{i.}$ value in pre-set confidence level, for every loss function occupied.

In this thesis, the 6 loss functions occupied in the MCS process are:

**1.** Squared Error 1: $SE_{1,t} = (\tilde{\sigma}_t - \hat{\sigma}_t)^2$ (18)

**2.** Squared Error 2: $SE_{2,t} = \left({\tilde{\sigma}_t}^2 - {\hat{\sigma}_t}^2\right)^2$ (19)

**3.** QLIKE Loss Function: $QLIKE_t = log\left({\hat{\sigma}_t}^2\right) + {\tilde{\sigma}_t}^2{\hat{\sigma}_t}^{-2}$ (20)

**4.** $R^2$ Log Loss Function: $R^2LOG_t = [log\left({\tilde{\sigma}_t}^2{\hat{\sigma}_t}^{-2}\right)]^2$ (21)

**5.** Absolute Error 1: $AE_{1,t} = |\tilde{\sigma}_t - \hat{\sigma}_t|$ (22)

**6.** Absolute Error 2: $AE_{2,t} = |{\tilde{\sigma}_t}^2 - {\hat{\sigma}_t}^2|$ (23)

Figures with ∼ denote the actual observations, whereas ^ denote the forecasted values.

## 4. DATA

The data used in this thesis spans 5,103 working days of the Borsa Istanbul Stock Exchange between 02.01.2001 and 30.04.2021, with inclusive boundaries. For this period, high-frequency data are obtained from BIST DataStore (datastore.borsaistanbul.com) for the XU100, XU030, XUMAL, XUSIN, XBANK, XKMYA, XGIDA, XMANA, XMESY, XTAST, and XTEKS indices. The reasoning behind the selection among the whole list of BIST indices depends on the constituent number and the market capitalization of the indices. The selection is also pertinent to coverage of firm-level idiosyncrasies and providing both diversified and concentrated analysis of the volatility process, as Campbell et al. (2001) suggest. The intraday values for those indices are then sampled into 5-minute returns for continuous trading sessions in a day, thus resulting in 378,560 observations. The formula for the derivation of the return series is the same as in equation (2), and for the calculation of RV, equation (4) is applied. The log and square root transformations of the RVs are not different from the generic mathematical technique.

The time space between price observations for return calculations is in line with the majority of the RV literature, as well as the detailed research over 11 years and 31 different financial assets by Liu et al. (2015), wherein they concluded that there is no significant advantage of using other realized measures than 5-minute RV. For forecasting methodology, the thesis utilizes 274 days of rolling windows for predicting the one-day ahead volatility, equating to 4829 daily forecasts. The parsing of the data and forecasting are done through R and its package "highfrequency". The performance evaluation sequence is conducted with R and package "MCS".

During this period, Borsa Istanbul have altered its session length by changing start and end times and removing the noon break. The research data of this study starts with 4-hour continuous trading sessions between 10:00 – 12:00 and 14:00 – 16:00. The first notable change was delivered on 13.08.2001, in which continuous trading sessions for the Equity Market stretched to 5 hours by rescheduling the start and end times as 09:30 – 12:00 and 14:00 – 16:30. In the coming years, starting time for the continuous morning session was deviated back and forth between 09:30 and 10:00, until it was set to 10:00 on November 14, 2016. The end of the morning session witnessed a constant expansion until it was seized on October 4, 2019, resulting in one integrated continuous trading session until 18:00. In this study, days with two trading sessions were combined into one whole trading session in resemblance to the current state. The unchanging values between two trading sessions are also omitted from that day's array. In addition, due to the variation in the regimes of the session starts and ends, the derived return data have different numbers of instances per day in different streaks of periods. The values of the closing auctions are also included in the analysis.

Another aspect of the data worth mentioning is the availability of sufficient calculation frequency to account for 5-minute returns. In all of the analysed days, Borsa Istanbul provided at least 10 seconds of calculated values of all of the selection. This element prevents virtual data generation and contributes to the robustness of the research.

The generation of descriptive statistics of the data is operated on the 3 RV series (RV, log RV, square root RV) per each of the indices and with a rolling window basis. To put it differently, no whole series for an index is adopted for the computation of these statistics; every 274-day window is treated separately to provide a series of values for statistics (Mean of the means, mean of the maximums etc.). This approach is necessary to treat each window as a separate sample space, in total alignment with the adopted forecasting procedure. The tables for descriptive statistics of all series are displayed below.

## Table 1: Descriptive Statistics for Plain RV Series

| | Mean | Min | Max | Median | Std | Skew | Kurtosis | JB | JB_p_value |
|---|---|---|---|---|---|---|---|---|---|
| **XBANK** | 0.00056 | 0.00013 | 0.00499 | 0.00046 | 0.00045 | 4.72117 | 44.40573 | 49213.78369 | 0.00026 |
| **XGIDA** | 0.00053 | 0.00008 | 0.00491 | 0.00043 | 0.00046 | 4.92166 | 48.15518 | 59297.86088 | 0.00000 |
| **XKMYA** | 0.00048 | 0.00009 | 0.00394 | 0.00040 | 0.00039 | 4.84293 | 45.61071 | 46292.29030 | 0.00000 |
| **XMANA** | 0.00082 | 0.00016 | 0.00700 | 0.00070 | 0.00064 | 4.91936 | 52.86459 | 80858.53881 | 0.00993 |
| **XMESY** | 0.00037 | 0.00009 | 0.00387 | 0.00030 | 0.00034 | 5.44877 | 54.25260 | 70643.76410 | 0.00000 |
| **XTAST** | 0.00019 | 0.00003 | 0.00285 | 0.00013 | 0.00025 | 6.20532 | 61.13431 | 75381.79969 | 0.00000 |
| **XTEKS** | 0.00039 | 0.00008 | 0.00457 | 0.00029 | 0.00042 | 4.80619 | 42.26001 | 48998.24920 | 0.00000 |
| **XUMAL** | 0.00038 | 0.00008 | 0.00397 | 0.00029 | 0.00036 | 5.03754 | 46.74744 | 51568.29105 | 0.00001 |
| **XUSIN** | 0.00020 | 0.00004 | 0.00310 | 0.00014 | 0.00026 | 6.15020 | 60.66172 | 70591.47652 | 0.00000 |
| **XU030** | 0.00031 | 0.00006 | 0.00333 | 0.00023 | 0.00031 | 5.14404 | 47.41936 | 49443.14728 | 0.00000 |
| **XU100** | 0.00025 | 0.00005 | 0.00317 | 0.00019 | 0.00029 | 5.42863 | 50.68495 | 55303.49069 | 0.00000 |

## Table 2: Descriptive Statistics for Log RV Series

| | Mean | Min | Max | Median | Std | Skew | Kurtosis | JB | JB_p_value |
|---|---|---|---|---|---|---|---|---|---|
| **XBANK** | -7.92312 | -9.43929 | -5.69400 | -7.95890 | 0.55133 | 0.56001 | 4.61433 | 78.08776 | 0.01957 |
| **XGIDA** | -8.06785 | -9.71458 | -5.66029 | -8.10775 | 0.62393 | 0.46219 | 4.36226 | 82.09340 | 0.10360 |
| **XKMYA** | -8.18722 | -9.71998 | -5.85564 | -8.22019 | 0.58892 | 0.54930 | 4.65080 | 101.72463 | 0.06390 |
| **XMANA** | -7.53399 | -9.00221 | -5.37320 | -7.56067 | 0.51413 | 0.55610 | 5.67388 | 207.22012 | 0.05601 |
| **XMESY** | -8.38158 | -9.75913 | -5.93949 | -8.43862 | 0.57475 | 0.80422 | 4.95900 | 117.81813 | 0.02186 |
| **XTAST** | -9.01365 | -10.43426 | -6.23396 | -9.10801 | 0.63005 | 1.10494 | 5.59179 | 196.50136 | 0.00120 |
| **XTEKS** | -8.15818 | -9.52453 | -5.85852 | -8.23405 | 0.56475 | 0.87855 | 4.93064 | 120.22107 | 0.00608 |
| **XUMAL** | -8.32164 | -9.80545 | -5.95227 | -8.37389 | 0.57922 | 0.70128 | 4.66592 | 82.92655 | 0.01413 |
| **XUSIN** | -9.03578 | -10.53858 | -6.28293 | -9.11478 | 0.62675 | 0.94926 | 5.18690 | 122.35768 | 0.00812 |
| **XU030** | -8.51552 | -10.06281 | -6.08992 | -8.57491 | 0.59293 | 0.73255 | 4.70994 | 84.52421 | 0.00576 |
| **XU100** | -8.73252 | -10.31706 | -6.17810 | -8.80123 | 0.62057 | 0.77928 | 4.72382 | 85.88539 | 0.00246 |

## Table 3: Descriptive Statistics for Square Root RV Series

| | Mean | Min | Max | Median | Std | Skew | Kurtosis | JB | JB_p_value |
|---|---|---|---|---|---|---|---|---|---|
| **XBANK** | 0.02106 | 0.00999 | 0.06359 | 0.01991 | 0.00653 | 2.22747 | 14.66825 | 4020.93679 | 0.00601 |
| **XGIDA** | 0.02017 | 0.00835 | 0.06452 | 0.01892 | 0.00704 | 2.26512 | 16.17822 | 8229.25303 | 0.01382 |
| **XKMYA** | 0.01910 | 0.00854 | 0.05829 | 0.01805 | 0.00624 | 2.29585 | 15.27003 | 4357.37325 | 0.00630 |
| **XMANA** | 0.02559 | 0.01185 | 0.07572 | 0.02450 | 0.00755 | 2.51447 | 20.91595 | 12432.26633 | 0.01135 |
| **XMESY** | 0.01695 | 0.00842 | 0.05648 | 0.01582 | 0.00572 | 2.68362 | 18.73863 | 9150.53166 | 0.00000 |
| **XTAST** | 0.01224 | 0.00564 | 0.04834 | 0.01100 | 0.00505 | 3.20496 | 21.49277 | 8428.88014 | 0.00000 |
| **XTEKS** | 0.01807 | 0.00880 | 0.05921 | 0.01661 | 0.00632 | 2.52019 | 15.54608 | 4947.27535 | 0.00000 |
| **XUMAL** | 0.01728 | 0.00827 | 0.05631 | 0.01610 | 0.00585 | 2.44171 | 15.52336 | 4344.11242 | 0.00986 |
| **XUSIN** | 0.01222 | 0.00556 | 0.04896 | 0.01111 | 0.00490 | 3.02038 | 19.66731 | 5979.94616 | 0.00000 |
| **XU030** | 0.01563 | 0.00722 | 0.05230 | 0.01446 | 0.00551 | 2.49834 | 15.48702 | 3928.37840 | 0.01213 |
| **XU100** | 0.01409 | 0.00632 | 0.05042 | 0.01290 | 0.00534 | 2.64027 | 16.47770 | 4517.20227 | 0.00118 |

To review and make a general assessment of all these statistics, the most critical comment that needs to be remarked on is the log series' closeness to normal distribution standards. Although not every p value for Jarque-Bera statistics of log series is above usual confidence levels of %1 and %5, the significant proximity of Kurtosis levels to 3 and skew levels to zero in comparison to other counterparts mark its usefulness in these types of analyses. In addition, the plain RV series' underperformance in contrast to its transformed peers should also be another aspect that needs to be paid attention to. Briefly, the log transformation of RV series proves to be an appropriate input for volatility forecasting purposes due to its convergence to normal distribution.

## 5. EMPIRICAL RESULTS

The results of the forecasting procedure will be presented in this section, along with guiding commentary and decisive remarks for findings. The evaluative process for models is indicated in section 3.3. as Model Confidence Set (MCS). The outputs of the MCS are tabulated for every index, thus adding up to 11 separate tables. In addition, a table for combined results is also constituted to exhibit a general outlook. Finally, a series of figures for forecasts are given. In overall, the models have displayed a satisfying performance in MCS tests.

In index tables, the score of every loss function is calculated as in equation (16) and the model's ranking depending on those values are depicted. The average ranking for a model is calculated as the averages of these rankings, and in tests where a model is eliminated, the ranking is interpreted as 15. Cells that include "E" in tables denote the model's elimination in the loss function test. For general results, a general average ranking that shows a model's average ranking across index series, as well as minimums, maximums, and standard deviations of these rankings, are displayed. Counts of elimination from tests are also given.

At first glance, the dominance of models with transformations appears. With log models taking the lead for average ranking among all series, it is evident that using transformed models are better for forecasting volatility. As can be inferred, plain RV series do not emerge as suitable for forecasting RV depending on these findings since they largely form the last rows of general rankings. This feature is also consistent with the previous section's evidence, where the log series demonstrated characteristics of normal distribution.

Secondly, the plain HAR-RV model with log series' occupation of the top row points out that for a robust RV examination, there may be no need for extensions, differing treatments for jumps and attenuation biases. Although using multiple models secures confident forecast analyses and will prevail as a principle approach, that type of judgement was necessitated due to observable performance distinctness. HAR-RV-J model was a close second, also proving itself as a robust estimator.

Another point that can be made out of this study is related to the apparent underachievement of models with quarticity. In general, models with quarticity underperformed their non-quarticity competitors. Although this outcome implies that daily realized volatility values should not be excluded from the RV forecasts even in the possible presence of market microstructure obscurenesses, there are also index series where models with quarticity performed better (XKMYA and XGIDA). However, this performance display can be attributed to the undiversified nature of the indices where quarticity prevails. In these indices, separate modelling of jumps also produced similar outcomes, though the findings deter making a general comment about the advantages of modelling with jumps in terms of diversification, due to the mixed nature of results as models with jumps also acquired good results in diversified indices.

As a final remark forecast results are given with a collection of figures for plain HAR-RV model on log series. Although being short in magnitude, the forecasts follow the actual values even in tail events, as can be seen from the graphs.

The tables and figures aforementioned are presented below.

## Table 4: Combined Results of MCS

| | | General Average Ranking | Minimum Average Ranking | Maximum Average Ranking | Standard Deviation of Average Rankings | Count of Elimination |
|---|---|---|---|---|---|---|
| **1** | **HAR_log** | 6.20 | 4.17 | 8.83 | 1.20 | 9 |
| **2** | **HARJ_log** | 6.27 | 4.83 | 7.67 | 0.89 | 7 |
| **3** | **CHAR_log** | 6.47 | 4.67 | 12.33 | 2.55 | 14 |
| **4** | **HARQJ_log** | 7.68 | 3.00 | 9.50 | 2.00 | 7 |
| **5** | **HARJ_sqrt** | 8.09 | 7.00 | 9.67 | 0.76 | 24 |
| **6** | **CHAR_sqrt** | 8.33 | 7.17 | 12.33 | 1.38 | 23 |
| **7** | **CHAR** | 8.89 | 7.67 | 11.33 | 1.27 | 25 |
| **8** | **HAR_sqrt** | 9.30 | 7.50 | 12.33 | 1.62 | 25 |
| **9** | **CHARQ_sqrt** | 9.79 | 8.17 | 12.50 | 1.40 | 26 |
| **10** | **CHARQ_log** | 10.26 | 7.67 | 13.33 | 2.01 | 16 |
| **11** | **CHARQ** | 11.02 | 9.33 | 12.50 | 0.95 | 33 |
| **12** | **HARQJ_sqrt** | 11.44 | 8.00 | 14.00 | 1.40 | 24 |
| **13** | **HAR** | 11.97 | 9.50 | 14.00 | 1.19 | 42 |
| **14** | **HARQJ** | 12.61 | 10.67 | 14.17 | 0.96 | 34 |
| **15** | **HARJ** | 12.98 | 11.00 | 15.00 | 1.22 | 35 |

## Table 5: Results of MCS for Forecasting RV of XBANK

| | SE1 | | SE2 | | QLIKE | | R2LOG | | AE1 | | AE2 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Rank** | **Value** | **Rank** | **Value** | **Rank** | **Value** | **Rank** | **Value** | **Rank** | **Value** | **Rank** | **Value** | **Average Ranking** |
| **CHAR_log** | 4 | -1.0987 | 7 | -0.9861 | 12 | -0.8786 | 3 | -0.1135 | 1 | -1.0784 | 1 | -2.3379 | 4.67 |
| **HAR_log** | 6 | -1.0156 | 5 | -1.0330 | 11 | -0.9316 | 5 | 1.3698 | 2 | -0.0539 | 2 | -2.1337 | 5.17 |
| **HARJ_log** | 7 | -0.9848 | 9 | -0.6721 | 10 | -0.9982 | 1 | -1.4122 | 3 | 0.0904 | 4 | -1.0130 | 5.67 |
| **CHAR_sqrt** | 2 | -1.3473 | 3 | -1.0983 | 5 | -1.1489 | E | E | E | E | 3 | -1.0713 | 7.17 |
| **HARJ_sqrt** | 1 | -1.3684 | 2 | -1.1455 | 8 | -1.0841 | E | E | E | E | 6 | 0.2485 | 7.83 |
| **HARQJ_log** | 12 | 0.8885 | 13 | 1.0378 | 9 | -1.0022 | 2 | -0.3710 | 4 | 0.5768 | 9 | 0.5400 | 8.17 |
| **HAR_sqrt** | 5 | -1.0299 | 4 | -1.0969 | 3 | -1.2240 | E | E | E | E | 8 | 0.4547 | 8.33 |
| **CHAR** | 8 | -0.9579 | 1 | -1.1568 | 1 | -1.4324 | E | E | E | E | 10 | 1.6070 | 8.33 |
| **CHARQ_log** | 10 | 0.7654 | 15 | 1.2521 | 13 | -0.8724 | 4 | 0.3779 | 5 | 0.7121 | 5 | 0.0908 | 8.67 |
| **CHARQ_sqrt** | 3 | -1.3428 | 8 | -0.9663 | 6 | -1.1301 | E | E | E | E | 7 | 0.4177 | 9 |
| **CHARQ** | 9 | -0.5396 | 6 | -1.0038 | 2 | -1.3081 | E | E | E | E | E | E | 10.33 |
| **HARQJ_sqrt** | 11 | 0.8580 | 12 | 1.0339 | 4 | -1.1676 | E | E | E | E | 11 | 1.7233 | 11.33 |
| **HARQJ** | 14 | 1.5105 | 14 | 1.0871 | 7 | -1.0962 | E | E | E | E | E | E | 13.33 |
| **HAR** | 15 | 1.5821 | 10 | 0.4267 | 14 | 0.6580 | E | E | E | E | E | E | 14 |
| **HARJ** | 13 | 1.3440 | 11 | 1.0172 | 15 | 1.0709 | E | E | E | E | E | E | 14 |

## Table 6: Results of MCS for Forecasting RV of XGIDA

| | SE1 | | SE2 | | QLIKE | | R2LOG | | AE1 | | AE2 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Rank | Value | Rank | Value | Rank | Value | Rank | Value | Rank | Value | Rank | Value | Average Ranking |
| **HARQJ_log** | 4 | -1.5616 | 4 | -0.8900 | E | E | 1 | -1.5788 | 1 | -1.5448 | 1 | -2.8689 | 4.33 |
| **HARJ_log** | 3 | -1.6825 | 5 | -0.8785 | E | E | E | E | 2 | 0.1695 | 2 | -2.3562 | 7 |
| **HARJ_sqrt** | 1 | -3.4180 | 1 | -1.6438 | 7 | 1.1105 | E | E | E | E | 8 | 1.6620 | 7.83 |
| **HARQJ_sqrt** | 2 | -1.7659 | 6 | -0.8054 | 5 | -0.3268 | E | E | E | E | 5 | 1.1609 | 8 |
| **HAR_sqrt** | 5 | -1.2634 | 2 | -1.2196 | 6 | 0.0841 | E | E | E | E | 7 | 1.4306 | 8.33 |
| **HAR_log** | 6 | -0.4483 | 11 | 0.6540 | E | E | E | E | 3 | 1.0639 | 3 | -1.9494 | 8.83 |
| **HARQJ** | 9 | 0.4410 | 7 | -0.1307 | 3 | -2.0426 | E | E | E | E | E | E | 11 |
| **HARJ** | 10 | 0.6370 | 3 | -1.0049 | 8 | 1.1276 | E | E | E | E | E | E | 11 |
| **HAR** | 12 | 1.1875 | 8 | 0.0203 | 2 | -2.1908 | E | E | E | E | E | E | 11.17 |
| **CHAR** | 13 | 1.2179 | 9 | 0.1605 | 1 | -2.3978 | E | E | E | E | E | E | 11.33 |
| **CHAR_log** | 11 | 0.8995 | 14 | 1.3871 | E | E | E | E | E | E | 4 | 0.4443 | 12.33 |
| **CHAR_sqrt** | 7 | -0.3808 | 12 | 0.7468 | 10 | 1.3082 | E | E | E | E | E | E | 12.33 |
| **CHARQ** | 15 | 1.6975 | 10 | 0.4121 | 4 | -1.2945 | E | E | E | E | E | E | 12 |
| **CHARQ_sqrt** | 8 | 0.4118 | 13 | 0.8663 | 9 | 1.2242 | E | E | E | E | E | E | 12.5 |
| **CHARQ_log** | 14 | 1.5535 | 15 | 1.7114 | E | E | E | E | E | E | 6 | 1.38753 | 13.33 |

## Table 7: Results of MCS for Forecasting RV of XKMYA

| | SE1 | | SE2 | | QLIKE | | R2LOG | | AE1 | | AE2 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Rank | Value | Rank | Value | Rank | Value | Rank | Value | Rank | Value | Rank | Value | Average Ranking |
| **HARQJ_log** | 1 | -2.3910 | 4 | -1.4043 | 10 | -1.1325 | 1 | -0.8349 | 1 | -1.5004 | 1 | -2.9924 | 3 |
| **HARJ_log** | 5 | -1.2409 | 10 | -0.6012 | 8 | -1.1364 | 2 | 0.8349 | 3 | 0.6689 | 4 | -0.9332 | 5.33 |
| **HAR_log** | 4 | -1.7126 | 1 | -1.5678 | 11 | -1.1294 | E | E | 2 | 0.5583 | 2 | -2.3646 | 5.83 |
| **HAR_sqrt** | 3 | -1.8043 | 2 | -1.4656 | 6 | -1.1504 | E | E | E | E | 5 | 0.5962 | 7.67 |
| **CHAR_log** | 6 | -1.1791 | 6 | -1.2886 | 9 | -1.1352 | E | E | E | E | 3 | -1.1320 | 9 |
| **CHAR_sqrt** | 7 | -0.7501 | 7 | -1.0798 | 4 | -1.1552 | E | E | E | E | 6 | 0.8294 | 9 |
| **HARJ_sqrt** | 2 | -1.8272 | 3 | -1.4305 | 5 | -1.1513 | E | E | E | E | E | E | 9.17 |
| **HARQJ_sqrt** | 10 | 0.6173 | 12 | 0.9918 | 3 | -1.1569 | E | E | E | E | 7 | 1.4364 | 10.33 |
| **CHAR** | 8 | -0.1409 | 9 | -0.8877 | 1 | -1.1780 | E | E | E | E | E | E | 10.5 |
| **CHARQ_sqrt** | 11 | 0.9774 | 11 | -0.2265 | 2 | -1.1574 | E | E | E | E | E | E | 11.5 |
| **CHARQ_log** | 14 | 1.4163 | 14 | 1.1188 | 7 | -1.1370 | E | E | E | E | 8 | 1.4605 | 12.17 |
| **HAR** | 12 | 1.0790 | 5 | -1.3091 | 12 | -1.0897 | E | E | E | E | E | E | 12.33 |
| **CHARQ** | 9 | 0.5750 | 8 | -1.0297 | 13 | -1.0847 | E | E | E | E | E | E | 12.5 |
| **HARQJ** | 13 | 1.1313 | 13 | 1.0076 | 14 | 0.8487 | E | E | E | E | E | E | 14.17 |
| **HARJ** | 15 | 1.4239 | 15 | 1.1647 | 15 | 1.0217 | E | E | E | E | E | E | 15 |

# Table 8: Results of MCS for Forecasting RV of XMANA

| | SE1 | | SE2 | | QLIKE | | R2LOG | | AE1 | | AE2 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Rank | Value | Rank | Value | Rank | Value | Rank | Value | Rank | Value | Rank | Value | Average Ranking |
| **HAR_log** | 3 | -1.6491 | 5 | -1.2971 | 11 | 1.3621 | 3 | 0.7465 | 2 | -0.0235 | 1 | -2.0024 | 4.17 |
| **HARJ_log** | 4 | -1.3977 | 6 | -1.2757 | 12 | 1.7081 | 2 | -0.1349 | 3 | 0.1993 | 2 | -1.5694 | 4.83 |
| **HARJ_sqrt** | 1 | -2.6279 | 3 | -1.4163 | 4 | -1.4736 | E | E | E | E | 5 | 0.0598 | 7.17 |
| **HARQJ_log** | 10 | 0.6530 | 12 | 1.0017 | 13 | 1.7419 | 1 | -0.9681 | 1 | -0.1307 | 7 | 0.5030 | 7.33 |
| **HAR_sqrt** | 2 | -2.0075 | 1 | -1.4806 | 9 | 0.7983 | E | E | E | E | 4 | -0.7041 | 7.67 |
| **CHAR_sqrt** | 6 | -0.7842 | 7 | -1.2243 | 5 | -1.0276 | E | E | E | E | 6 | 0.3457 | 9 |
| **HAR** | 9 | 0.2101 | 2 | -1.4341 | 1 | -1.8647 | E | E | E | E | E | E | 9.5 |
| **CHAR_log** | 5 | -0.9694 | 8 | -1.1903 | 14 | 1.7878 | E | E | E | E | 3 | -1.1574 | 10 |
| **CHAR** | 13 | 1.2111 | 4 | -1.3210 | 2 | -1.8094 | E | E | E | E | E | E | 10.67 |
| **CHARQ_sqrt** | 8 | 0.0605 | 10 | -0.2102 | 6 | -1.0134 | E | E | E | E | E | E | 11.5 |
| **CHARQ** | 12 | 1.0615 | 9 | -0.7818 | 3 | -1.6985 | E | E | E | E | E | E | 11.5 |
| **HARQJ_sqrt** | 7 | -0.1269 | 14 | 1.1862 | 10 | 1.0749 | E | E | E | E | 9 | 1.5099 | 11.67 |
| **CHARQ_log** | 11 | 0.9276 | 13 | 1.0369 | 15 | 1.7943 | E | E | E | E | 8 | 0.9904 | 12.83 |
| **HARJ** | 14 | 1.2475 | 11 | 0.6090 | 8 | -0.1241 | E | E | E | E | E | E | 13 |
| **HARQJ** | E | E | 15 | 1.4698 | 7 | -0.4320 | E | E | E | E | E | E | 13.67 |

# Table 9: Results of MCS for Forecasting RV of XMESY

| | SE1 | | SE2 | | QLIKE | | R2LOG | | AE1 | | AE2 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Rank | Value | Rank | Value | Rank | Value | Rank | Value | Rank | Value | Rank | Value | Average Ranking |
| **CHAR_log** | 3 | -1.2666 | 5 | -1.1739 | E | E | 3 | 0.4149 | 1 | -1.4886 | 1 | -2.1257 | 4.67 |
| **HAR_log** | 4 | -1.0229 | 4 | -1.1845 | E | E | E | E | 4 | 0.6440 | 2 | -1.6991 | 7.33 |
| **CHAR_sqrt** | 2 | -1.4028 | 2 | -1.2245 | 8 | 1.3731 | E | E | E | E | 3 | -1.1093 | 7.5 |
| **HARJ_log** | 9 | 0.6918 | 15 | 1.3140 | E | E | 1 | -0.9461 | 2 | -0.1366 | 4 | 0.3282 | 7.67 |
| **CHARQ_log** | 11 | 0.8384 | 10 | 0.1669 | E | E | 2 | 0.1202 | 3 | 0.0314 | 5 | 0.4263 | 7.67 |
| **CHAR** | 5 | -1.0139 | 1 | -1.2632 | 4 | -0.7364 | E | E | E | E | 7 | 0.9056 | 7.83 |
| **HARJ_sqrt** | 1 | -1.5465 | 3 | -1.2121 | 7 | 1.2683 | E | E | E | E | 9 | 1.7462 | 8.33 |
| **HARQJ_log** | 12 | 0.9903 | 13 | 1.1465 | E | E | 4 | 0.4410 | 5 | 1.3289 | 6 | 0.8800 | 9.17 |
| **CHARQ_sqrt** | 6 | -0.8542 | 8 | -0.8377 | 6 | 0.5295 | E | E | E | E | 8 | 1.5653 | 9.67 |
| **CHARQ** | 7 | -0.4376 | 7 | -1.1476 | 3 | -1.2129 | E | E | E | E | E | E | 10.33 |
| **HAR** | E | E | 9 | -0.1030 | 1 | -1.8812 | E | E | E | E | E | E | 11.67 |
| **HARQJ** | 14 | 1.3294 | 11 | 1.0656 | 2 | -1.3076 | E | E | E | E | E | E | 12 |
| **HAR_sqrt** | 8 | -0.4193 | 6 | -1.1651 | E | E | E | E | E | E | E | E | 12.33 |
| **HARJ** | 13 | 1.2068 | 12 | 1.0884 | 5 | 0.33093 | E | E | E | E | E | E | 12.5 |
| **HARQJ_sqrt** | 10 | 0.7942 | 14 | 1.2247 | E | E | E | E | E | E | E | E | 14 |

## Table 10: Results of MCS for Forecasting RV of XTAST

| | SE1 | | SE2 | | QLIKE | | R2LOG | | AE1 | | AE2 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Rank | Value | Rank | Value | Rank | Value | Rank | Value | Rank | Value | Rank | Value | Average Ranking |
| **CHAR_log** | 5 | -1.1575 | 6 | -1.1390 | E | E | 2 | -0.3299 | 1 | -0.8855 | 2 | -1.7714 | 5.17 |
| **HAR_log** | 6 | -1.0541 | 7 | -1.1272 | E | E | 4 | 1.0318 | 3 | -0.6088 | 1 | -1.9212 | 6 |
| **HARJ_log** | 9 | 0.5690 | 10 | 1.1088 | E | E | 1 | -1.7282 | 2 | -0.6092 | 5 | -0.1743 | 7 |
| **HAR_sqrt** | 2 | -1.3012 | 5 | -1.1428 | 4 | -0.4906 | E | E | E | E | 4 | -0.6007 | 7.5 |
| **CHAR_sqrt** | 3 | -1.2941 | 2 | -1.1497 | 8 | 1.3193 | E | E | E | E | 3 | -0.7975 | 7.67 |
| **CHARQ_sqrt** | 4 | -1.2792 | 3 | -1.1497 | 7 | 1.3187 | E | E | E | E | 6 | -0.0734 | 8.33 |
| **CHAR** | 7 | -0.8573 | 1 | -1.1500 | 3 | -0.6446 | E | E | E | E | 10 | 1.3059 | 8.5 |
| **HARQJ_log** | 11 | 1.0115 | 13 | 1.1347 | E | E | 3 | -0.2535 | 4 | 1.0973 | 9 | 0.6717 | 9.17 |
| **CHARQ_log** | 12 | 1.0130 | 12 | 1.1209 | E | E | 5 | 1.2446 | 5 | 1.2722 | 7 | 0.4421 | 9.33 |
| **HARJ_sqrt** | 1 | -1.5518 | 4 | -1.1481 | E | E | E | E | E | E | 8 | 0.5182 | 9.67 |
| **CHARQ** | 8 | -0.7407 | 8 | -1.1235 | 5 | 0.4105 | E | E | E | E | E | E | 11 |
| **HARQJ_sqrt** | 10 | 0.8191 | 11 | 1.1178 | 6 | 0.6200 | E | E | E | E | 11 | 1.3895 | 11.33 |
| **HAR** | E | E | 9 | -1.1171 | 1 | -1.2117 | E | E | E | E | E | E | 11.67 |
| **HARQJ** | 13 | 1.2511 | 14 | 1.1440 | 2 | -0.9955 | E | E | E | E | E | E | 12.33 |
| **HARJ** | 14 | 1.3168 | 15 | 1.1488 | E | E | E | E | E | E | E | E | 14.83 |

## Table 11: Results of MCS for Forecasting RV of XTEKS

| | SE1 | | SE2 | | QLIKE | | R2LOG | | AE1 | | AE2 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Rank | Value | Rank | Value | Rank | Value | Rank | Value | Rank | Value | Rank | Value | Average Ranking |
| **CHAR_log** | 4 | -1.0351 | 6 | -1.1298 | E | E | 3 | 0.3914 | 2 | -0.1787 | 1 | -1.4443 | 5.17 |
| **HAR_log** | 5 | -0.9166 | 8 | -1.1038 | 12 | 1.6895 | 2 | 0.0303 | 3 | -0.1672 | 2 | -1.4199 | 5.33 |
| **HARJ_log** | 8 | 0.4728 | 10 | 1.0266 | E | E | 1 | -1.0743 | 1 | -0.5984 | 6 | 0.0154 | 6.83 |
| **HARJ_sqrt** | 1 | -1.4884 | 3 | -1.1734 | 5 | -1.4190 | E | E | E | E | 3 | -1.0935 | 7 |
| **CHAR_sqrt** | 2 | -1.1805 | 4 | -1.1498 | 7 | -0.2932 | E | E | E | E | 4 | -0.9698 | 7.83 |
| **HARQJ_log** | 10 | 0.9733 | 12 | 1.1099 | 11 | 1.6765 | 4 | 0.6124 | 4 | 0.6493 | 7 | 0.5613 | 8 |
| **HAR_sqrt** | 3 | -1.1732 | 5 | -1.1397 | 10 | 1.2547 | E | E | E | E | 5 | -0.7224 | 8.83 |
| **CHAR** | 6 | -0.7313 | 7 | -1.1216 | 4 | -1.5045 | E | E | E | E | 8 | 0.6220 | 9.17 |
| **CHARQ_sqrt** | 7 | 0.4239 | 1 | -1.1994 | 8 | 0.6996 | E | E | E | E | E | E | 10.17 |
| **HAR** | E | E | 2 | -1.1814 | 1 | -3.5012 | E | E | E | E | E | E | 10.5 |
| **HARQJ_sqrt** | 9 | 0.9115 | 11 | 1.1069 | 9 | 1.1526 | E | E | E | E | 10 | 1.1545 | 11.5 |
| **CHARQ** | E | E | 9 | -1.0278 | 3 | -1.6109 | E | E | E | E | E | E | 12 |
| **HARQJ** | 12 | 1.3366 | 14 | 1.1417 | 2 | -2.6964 | E | E | E | E | E | E | 12.17 |
| **CHARQ_log** | 11 | 1.1337 | 13 | 1.1253 | E | E | E | E | E | E | 9 | 0.9667 | 13 |
| **HARJ** | E | E | 15 | 1.2064 | 6 | -1.0591 | E | E | E | E | E | E | 13.5 |

## Table 12: Results of MCS for Forecasting RV of XUMAL

| | SE1 | | SE2 | | QLIKE | | R2LOG | | AE1 | | AE2 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Rank | Value | Rank | Value | Rank | Value | Rank | Value | Rank | Value | Rank | Value | Average Ranking |
| **CHAR_log** | 4 | -1.0853 | 7 | -1.0031 | E | E | 3 | -0.0702 | 1 | -1.1574 | 1 | -2.0966 | 5.17 |
| **HAR_log** | 6 | -0.9250 | 6 | -1.0060 | 11 | 1.4696 | 5 | 1.0868 | 2 | -0.0639 | 2 | -1.8771 | 5.33 |
| **HARJ_log** | 9 | 0.6893 | 12 | 1.0383 | 13 | 1.6359 | 1 | -1.6746 | 3 | -0.0104 | 5 | -0.1580 | 7.17 |
| **CHAR_sqrt** | 2 | -1.2172 | 2 | -1.0381 | 8 | 0.5323 | E | E | E | E | 3 | -1.0811 | 7.5 |
| **CHAR** | 5 | -0.9883 | 1 | -1.0726 | 2 | -1.7806 | E | E | E | E | 8 | 0.6847 | 7.67 |
| **CHARQ_sqrt** | 1 | -1.2493 | 4 | -1.0179 | 10 | 0.9988 | E | E | E | E | 4 | -0.7640 | 8.17 |
| **HARQJ_log** | 12 | 0.9420 | 11 | 1.0241 | 12 | 1.5634 | 2 | -0.5191 | 5 | 0.7875 | 7 | 0.4992 | 8.17 |
| **HARJ_sqrt** | 3 | -1.1663 | 3 | -1.0298 | 6 | -0.1396 | E | E | E | E | 9 | 0.8086 | 8.5 |
| **CHARQ_log** | 11 | 0.9323 | 13 | 1.0419 | E | E | 4 | 1.0740 | 4 | 0.7507 | 6 | 0.4074 | 8.83 |
| **CHARQ** | 7 | -0.7233 | 5 | -1.0077 | 3 | -1.2874 | E | E | E | E | 11 | 1.1063 | 9.33 |
| **HAR_sqrt** | 8 | -0.5974 | 8 | -0.9787 | 7 | 0.4267 | E | E | E | E | 12 | 1.1545 | 10.83 |
| **HARQJ_sqrt** | 10 | 0.8506 | 10 | 0.9882 | 9 | 0.5335 | E | E | E | E | 10 | 1.0251 | 11.5 |
| **HARQJ** | 14 | 1.2865 | 9 | 0.9840 | 4 | -1.1695 | E | E | E | E | E | E | 12 |
| **HAR** | E | E | E | E | 1 | -2.5432 | E | E | E | E | E | E | 12.67 |
| **HARJ** | 13 | 1.2760 | 14 | 1.0504 | 5 | -0.5699 | E | E | E | E | E | E | 12.83 |

## Table 13: Results of MCS for Forecasting RV of XUSIN

| | SE1 | | SE2 | | QLIKE | | R2LOG | | AE1 | | AE2 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Rank | Value | Rank | Value | Rank | Value | Rank | Value | Rank | Value | Rank | Value | Average Ranking |
| **HARJ_log** | 8 | -0.4054 | 1 | -1.1219 | E | E | 1 | -0.5489 | 2 | -0.3499 | 4 | -0.3287 | 5.17 |
| **CHAR_log** | 5 | -1.0657 | 8 | -1.0178 | E | E | 2 | -0.4971 | 1 | -1.0771 | 1 | -1.5316 | 5.33 |
| **HAR_log** | 6 | -1.0522 | 4 | -1.0292 | 11 | 1.4659 | E | E | 3 | 0.1644 | 2 | -1.4252 | 6.83 |
| **CHAR** | 7 | -0.9076 | 2 | -1.0346 | 1 | -1.5646 | E | E | E | E | 6 | 0.2112 | 7.67 |
| **HARJ_sqrt** | 2 | -1.2181 | 6 | -1.0276 | 5 | -0.4750 | E | E | E | E | 5 | -0.2696 | 8 |
| **CHAR_sqrt** | 3 | -1.0989 | 7 | -1.0223 | 6 | -0.1480 | E | E | E | E | 3 | -1.0107 | 8.17 |
| **CHARQ_sqrt** | 1 | -2.2076 | 5 | -1.0286 | 4 | -0.4791 | E | E | E | E | 10 | 1.0482 | 8.33 |
| **HAR_sqrt** | 4 | -1.0911 | 3 | -1.0306 | 9 | 1.0846 | E | E | E | E | 7 | 0.2665 | 8.83 |
| **HARQJ_log** | 10 | 0.5516 | 11 | 0.9762 | E | E | 4 | 1.1081 | 5 | 1.1301 | 8 | 0.4908 | 8.83 |
| **CHARQ_log** | 11 | 0.8315 | 14 | 1.0185 | E | E | 3 | 0.0946 | 4 | 0.4000 | 9 | 0.6196 | 9 |
| **CHARQ** | 9 | -0.0044 | 9 | -0.9716 | 3 | -1.2215 | E | E | E | E | E | E | 11 |
| **HARJ** | 13 | 0.9411 | 12 | 0.9976 | 2 | -1.5554 | E | E | E | E | E | E | 12 |
| **HARQJ_sqrt** | 12 | 0.9362 | 13 | 1.0171 | 10 | 1.3301 | E | E | E | E | 11 | 1.4451 | 12.67 |
| **HAR** | 14 | 1.2743 | 10 | 0.9564 | 8 | 0.3832 | E | E | E | E | E | E | 12.83 |
| **HARQJ** | 15 | 1.4373 | 15 | 1.0481 | 7 | 0.1278 | E | E | E | E | E | E | 13.67 |

# Table 14: Results of MCS for Forecasting RV of XU030

| | SE1 | | SE2 | | QLIKE | | R2LOG | | AE1 | | AE2 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Rank | Value | Rank | Value | Rank | Value | Rank | Value | Rank | Value | Rank | Value | Average Ranking |
| **CHAR_log** | 4 | -1.1909 | 6 | -0.9911 | E | E | 2 | -0.1460 | 1 | -1.1251 | 1 | -1.9727 | 4.83 |
| **HARJ_log** | 9 | 0.0293 | 9 | 1.0006 | 12 | 1.6237 | 1 | -1.1038 | 2 | -0.7176 | 3 | -0.8135 | 6 |
| **HAR_log** | 5 | -1.0187 | 4 | -1.0271 | 11 | 1.4830 | E | E | 3 | 0.2315 | 2 | -1.8179 | 6.67 |
| **HARJ_sqrt** | 1 | -1.6597 | 2 | -1.1033 | 4 | -1.5190 | E | E | E | E | 9 | 1.2863 | 7.67 |
| **CHAR_sqrt** | 3 | -1.3750 | 3 | -1.0539 | 7 | 0.3498 | E | E | E | E | 4 | -0.6929 | 7.83 |
| **CHAR** | 6 | -0.9436 | 1 | -1.1034 | 2 | -2.1568 | E | E | E | E | 10 | 1.4207 | 8.17 |
| **CHARQ_log** | 11 | 0.9678 | 13 | 1.0945 | E | E | 4 | 1.3111 | 5 | 0.8788 | 5 | 0.4172 | 8.83 |
| **HARQJ_log** | 12 | 1.0898 | E | E | 13 | 1.7073 | 3 | 0.1154 | 4 | 0.8222 | 6 | 0.4513 | 8.83 |
| **CHARQ_sqrt** | 2 | -1.5703 | 8 | -0.9413 | 8 | 0.3663 | E | E | E | E | 7 | 0.9006 | 9.17 |
| **CHARQ** | 7 | -0.6831 | 5 | -1.0082 | 3 | -1.5271 | E | E | E | E | E | E | 10 |
| **HAR_sqrt** | 8 | -0.4575 | 7 | -0.9604 | 9 | 0.5625 | E | E | E | E | E | E | 11.5 |
| **HARQJ_sqrt** | 10 | 0.9639 | 11 | 1.0177 | 10 | 0.9227 | E | E | E | E | 8 | 1.1723 | 11.5 |
| **HARQJ** | 14 | 1.5371 | 10 | 1.0088 | 5 | -0.8923 | E | E | E | E | E | E | 12.33 |
| **HARJ** | 13 | 1.1579 | 12 | 1.0201 | 6 | -0.4235 | E | E | E | E | E | E | 12.67 |
| **HAR** | E | E | E | E | 1 | -3.0114 | E | E | E | E | E | E | 12.67 |

# Table 15: Results of MCS for Forecasting RV of XU100

| | SE1 | | SE2 | | QLIKE | | R2LOG | | AE1 | | AE2 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Rank | Value | Rank | Value | Rank | Value | Rank | Value | Rank | Value | Rank | Value | Average Ranking |
| **CHAR_log** | 4 | -1.0675 | 5 | -0.9811 | E | E | 2 | -0.2277 | 2 | -1.0373 | 1 | -1.8112 | 4.83 |
| **HARJ_log** | 9 | 0.0062 | 9 | 0.9263 | E | E | 1 | -1.2069 | 1 | -1.1391 | 3 | -0.8652 | 6.33 |
| **HAR_log** | 5 | -0.9979 | 4 | -1.0123 | 11 | 1.5327 | E | E | 3 | -0.0199 | 2 | -1.8044 | 6.67 |
| **CHAR_sqrt** | 2 | -1.2145 | 3 | -1.0176 | 7 | 0.8277 | E | E | E | E | 4 | -0.7981 | 7.67 |
| **HARJ_sqrt** | 1 | -1.5515 | 2 | -1.0552 | 6 | -0.9388 | E | E | E | E | 8 | 0.8429 | 7.83 |
| **CHAR** | 6 | -0.8985 | 1 | -1.0601 | 2 | -1.8200 | E | E | E | E | 9 | 0.9069 | 8 |
| **CHARQ_log** | 12 | 0.9216 | 13 | 1.0328 | E | E | 4 | 1.3267 | 4 | 0.8532 | 5 | 0.5103 | 8.83 |
| **CHARQ_sqrt** | 3 | -1.1456 | 8 | -0.9301 | 8 | 0.9546 | E | E | E | E | 7 | 0.8317 | 9.33 |
| **HARQJ_log** | 13 | 1.1394 | E | E | E | E | 3 | 0.3666 | 5 | 1.2377 | 6 | 0.6974 | 9.5 |
| **HAR_sqrt** | 7 | -0.7076 | 6 | -0.9772 | 9 | 0.9935 | E | E | E | E | 11 | 1.3593 | 10.5 |
| **CHARQ** | 8 | -0.4055 | 7 | -0.9370 | 5 | -1.2293 | E | E | E | E | E | E | 10.83 |
| **HARJ** | 11 | 0.9081 | 10 | 0.9694 | 3 | -1.5810 | E | E | E | E | E | E | 11.5 |
| **HARQJ_sqrt** | 10 | 0.8886 | 12 | 1.0071 | 10 | 1.4361 | E | E | E | E | 10 | 1.0975 | 12 |
| **HARQJ** | 14 | 1.4030 | 11 | 0.9999 | 4 | -1.4220 | E | E | E | E | E | E | 12.33 |
| **HAR** | E | E | E | E | 1 | -2.9442 | E | E | E | E | E | E | 12.67 |

**Figure 1:Forecasting Results of Plain HAR-RV on Log Series (2002-2006)**

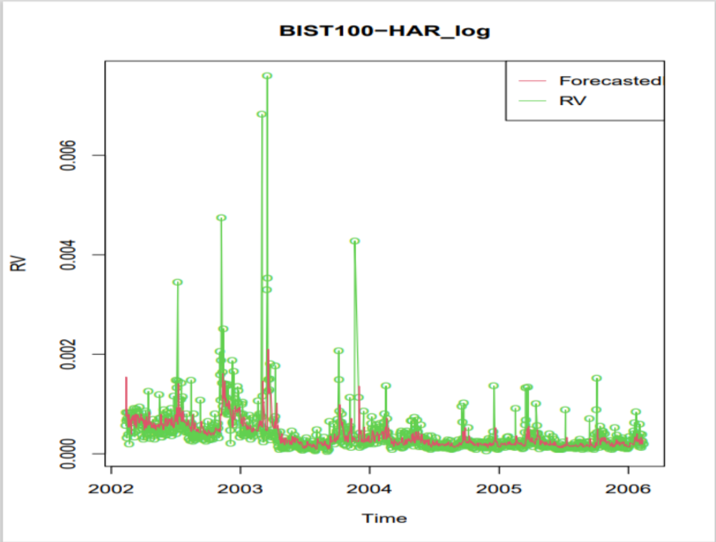



**Figure 2:Forecasting Results of Plain HAR-RV on Log Series (2006-2010)**

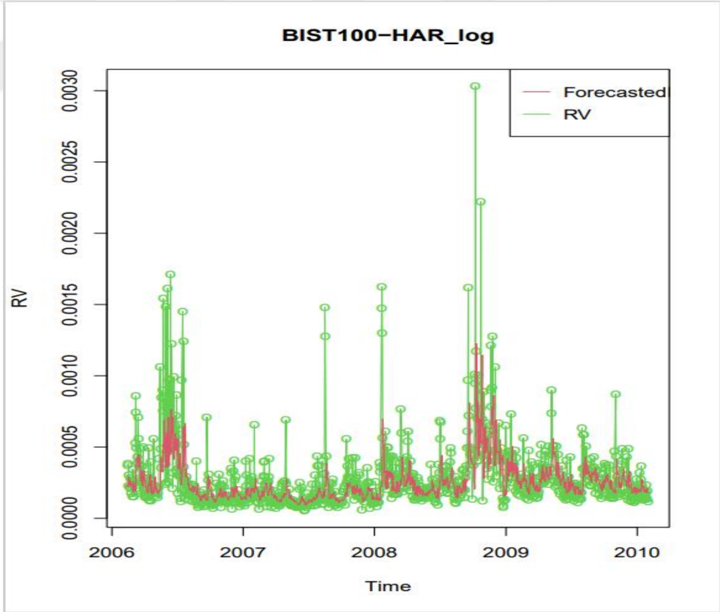


**Figure 3:Forecasting Results of Plain HAR-RV on Log Series (2010-2014)**

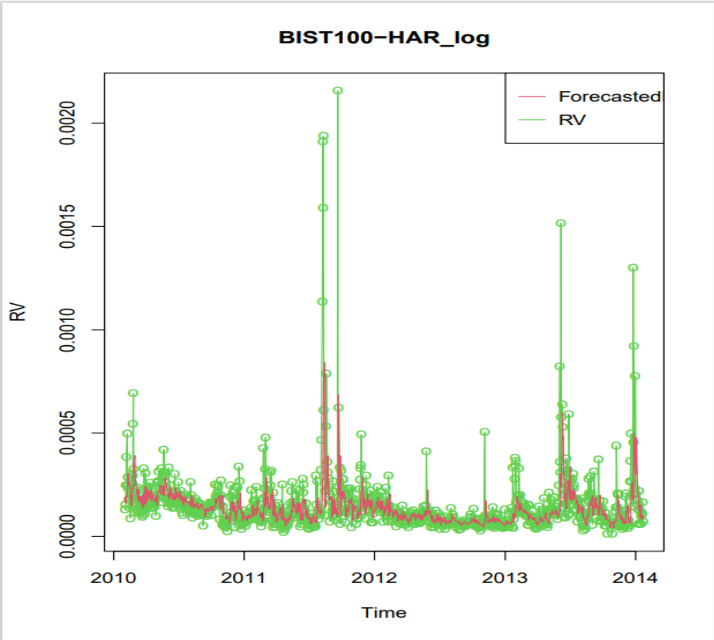



**Figure 4:Forecasting Results of Plain HAR-RV on Log Series (2014-2018)**

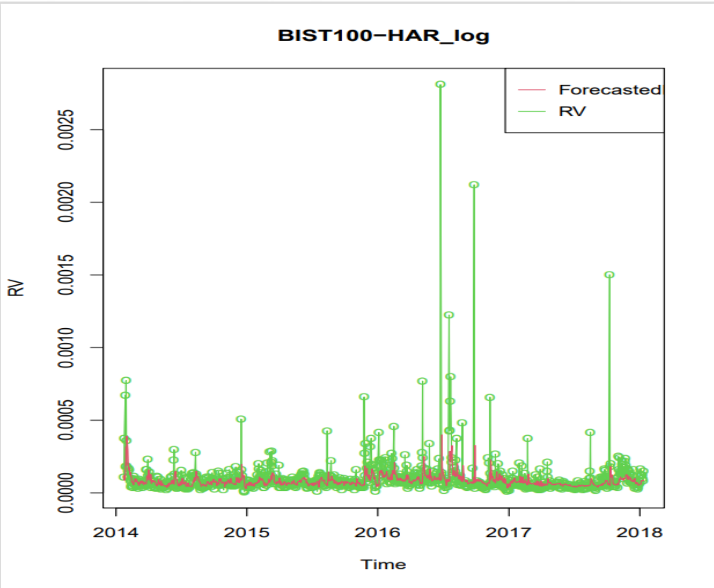
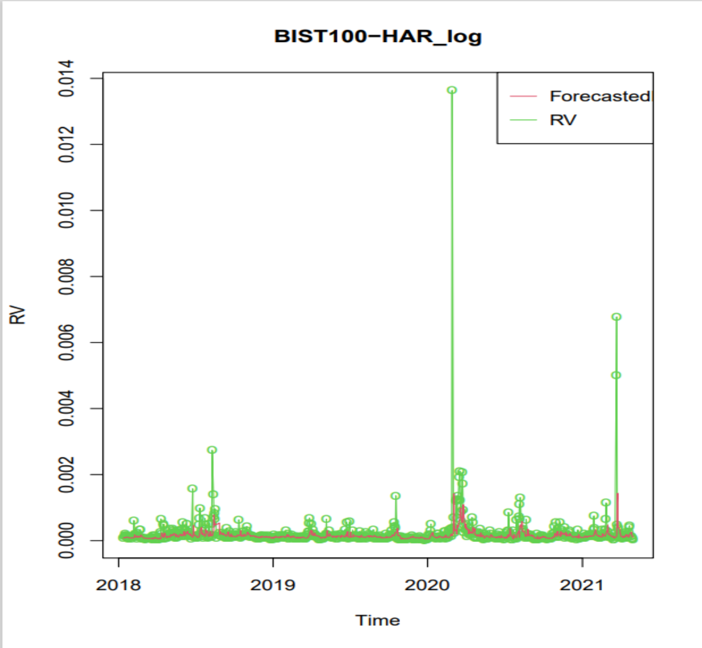


**Figure 5:Forecasting Results of Plain HAR-RV on Log Series (2018-2021)**




## 6. CONCLUSION

The problem of forecasting volatility is an important topic in modern finance. The venture starts with finding the proper definition: there is even dispute over the term volatility's description, and eventually the calculation methodology. The study at hand relies on the high-frequency expansion of the definition of volatility, where the quadratic variation equation finds parity between realized volatility and the ultimate goal: integrated volatility. As the whole quantitative finance literature on volatility suggests, volatility is, in its essence, an indecisiveness of the market while trying to find the right direction. Therefore, it is intuitive to look for the very incremental steps of this random walk.

Grounding on the mindset outlined in the first paragraph, this thesis aimed to provide a comparison of the HAR Model extensions for the purpose of forecasting volatility for BIST indices. As stated in the founding paper of Corsi (2009), the model aligns with the nature of high-frequency data through modelling and averaging the previous RV values, making the long memory count. The study spans 5-minute returns between 2001 and 2021. For forecasting, approximately one-year rolling windows are used. The predictions of the models are compared to the actual realized volatilities.

The first conclusion of this research is about the underlying time series properties. From descriptive statistics, the transformed series' convergence to normal distribution standards acquired as a fact, with logarithmic transformation stands one step further. This finding is also supported in comparison section, while models with transformations dominated the general statistics, particularly log models. In brief, the benefits of using logarithmic transformations are clear.

Another point that can be made out of this study is related with the plain HAR-RV model's success among all models. While run on log series, the basic model stands out as a robust RV estimator. HAR-RV-J model was a close second, also proving itself as a successful model. Although the thesis' findings demand a remark on the plain model's performance, it would be wiser not to leave the models with extensions because of the reasons laid out in the upcoming paragraph.

The surfacing results on the underperformance of models with quarticity remains as a significant issue. While the tempting nature of commenting on this apparent overall underachievement is understandable, one should also take into account the typicalities of the underlying series. As noted previously in the results section, there are also index series where models with quarticity performed better. The unbalanced nature in terms of weights of these indices where quarticity models retained good rankings inclines the final judgement towards the usefulness of these models in this particular type of index time series. For a well-diversified portfolio, however, it is evident that daily realized volatility values should not be excluded from the RV forecasts even in the possible presence of noise contamination. A similar situation of successfulness in concentrated portfolios also holds for models with jumps, though it would be invalid to state a general comment in this sense about the models with jumps as they were also successful in diversified portfolios.

All in all, with paying attention to the preceding remarks, HAR-type models proved as a reliable estimator in index series. In overall, the models have displayed a satisfying performance in MCS tests. It is also apparent from the forecasting figures in which only outlying values are not fulfilling, where even in these, the direction and magnitude of the forecast hint at marginal values.

## 7. REFERENCES

Agermark, A., & Hoti, V. (2016). Forecasting Exchange Rate Volatility. Applying HAR models and Implied Volatility in SEK denominated markets.

Andersen, T. G., Bollerslev, T., & Diebold, F. X. (2007). Roughing it up: Including jump components in the measurement, modelling, and forecasting of return volatility. *The review of economics and statistics, 89*(4), 701-720.

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (2000). Exchange rate returns standardized by realized volatility are (nearly) Gaussian.

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Ebens, H. (2001). The distribution of realized stock return volatility. *Journal of financial economics, 61*(1), 43-76.

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (2001). The distribution of realized exchange rate volatility. *Journal of the American statistical association, 96*(453), 42-55.

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

Angelidis, T., Benos, A., & Degiannakis, S. (2004). The use of GARCH models in VaR estimation. *Statistical methodology, 1*(1-2), 105-128.

Bali, T. G., & Theodossiou, P. (2007). A conditional-SGT-VaR approach with alternative GARCH models. *Annals of Operations Research, 151*(1), 241-267.

Bali, T. G., Mo, H., & Tang, Y. (2008). The role of autoregressive conditional skewness and kurtosis in the estimation of conditional VaR. *Journal of Banking & Finance, 32*(2), 269-282.

Barndorff-Nielsen, O. E., & Shephard, N. (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 64*(2), 253-280.

Barndorff-Nielsen, O. E., & Shephard, N. (2004). Power and bipower variation with stochastic volatility and jumps. *Journal of financial econometrics*, *2*(1), 1-37.

Barndorff-Nielsen, O. E., & Shephard, N. (2006). Econometrics of testing for jumps in financial economics using bipower variation. *Journal of financial Econometrics*, *4*(1), 1-30.

Bergsli, L. Ø., Lind, A. F., Molnár, P., & Polasik, M. (2022). Forecasting volatility of Bitcoin. *Research in International Business and Finance*, *59*, 101540.

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

Bollerslev, T., Patton, A. J., & Quaedvlieg, R. (2016). Exploiting the errors: A simple approach for improved volatility forecasting. *Journal of Econometrics, 192*(1), 1-18.

Bollerslev, T., Hood, B., Huss, J., & Pedersen, L. H. (2018). Risk everywhere: Modeling and managing volatility. *The Review of Financial Studies, 31*(7), 2729-2773.

Boudt K, Kleen O, Sjørup E (2022). "Analyzing Intraday Financial Data in R: The highfrequency Package." *Journal of Statistical Software*, **104**(8), 1–36. doi:10.18637/jss.v104.i08.

Braione, M., & Scholtes, N. K. (2016). Forecasting value-at-risk under different distributional assumptions. *Econometrics, 4*(1), 3.

Campbell, J. Y., Lettau, M., Malkiel, B. G., & Xu, Y. (2001). Have individual stocks become more volatile? An empirical exploration of idiosyncratic risk. *The journal of finance*, *56*(1), 1-43.

Catania L, & Bernardi M (2017). MCS: Model Confidence Set Procedure. R package version 0.1.3. https://CRAN.R-project.org/package=MCS

Chen, X. B., Gao, J., Li, D., & Silvapulle, P. (2018). Nonparametric estimation and forecasting for time-varying coefficient realized volatility models. *Journal of Business & Economic Statistics*, *36*(1), 88-100.

Corsi, F., Mittnik, S., Pigorsch, C., & Pigorsch, U. (2008). The volatility of realized volatility. *Econometric Reviews, 27*(1-3), 46-78.

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

Corsi, F., Pirino, D., & Reno, R. (2010). Threshold bipower variation and the impact of jumps on volatility forecasting. *Journal of Econometrics, 159*(2), 276-288.

Corsi, F., & Renò, R. (2012). Discrete-time volatility forecasting with persistent leverage effect and the link with continuous-time volatility modeling. *Journal of Business & Economic Statistics, 30*(3), 368-380.

Çelik, S., & Ergin, H. (2014). Volatility forecasting using high frequency data: Evidence from stock markets. *Economic modelling, 36*, 176-190.

Degiannakis, S., Filis, G., Klein, T., & Walther, T. (2022). Forecasting realized volatility of agricultural commodities. *International Journal of Forecasting, 38*(1), 74-96.

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

Engle, R. F. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics, 20*(3), 339-350.

Eroğlu, B. A., İkizlerli, D., & Yener, H. (2021). Reexamination of the BIST 100 Stock Price Volatility with Heterogeneous Autoregressive Realized Volatility Models. *Atatürk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, 25*(2), 457-476.

Ftiti, Z., Louhichi, W., & Ben Ameur, H. (2021). Cryptocurrency volatility forecasting: What can we learn from the first wave of the COVID-19 outbreak? *Annals of Operations Research*, 1-26.

Giot, P. (2003). The information content of implied volatility in agricultural commodity markets. *Journal of Futures Markets: Futures, Options, and Other Derivative Products*, *23*(5), 441-454.

Giot, P., & Laurent, S. (2003). Value-at-risk for long and short trading positions. *Journal of Applied Econometrics*, *18*(6), 641-663.

Giot, P., & Laurent, S. (2004). Modelling daily value-at-risk using realized volatility and ARCH type models. *Journal of empirical finance, 11*(3), 379-398.

Hansen, P. R., Lunde, A., & Nason, J. M. (2011). The model confidence set. *Econometrica*, *79*(2), 453-497.

Heston, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The review of financial studies, 6*(2), 327-343.

Kambouroudis, D. S., McMillan, D. G., & Tsakou, K. (2021). Forecasting realized volatility: The role of implied volatility, leverage effect, overnight returns, and volatility of realized volatility. *Journal of Futures Markets, 41*(10), 1618-1639.

Kuester, K., Mittnik, S., & Paolella, M. S. (2006). Value-at-risk prediction: A comparison of alternative strategies. *Journal of Financial Econometrics, 4*(1), 53-89.

Liu, L. Y., Patton, A. J., & Sheppard, K. (2015). Does anything beat 5-minute RV? A comparison of realized measures across multiple asset classes. *Journal of Econometrics*, *187*(1), 293-311.

Luo, J., Klein, T., Ji, Q., & Hou, C. (2022). Forecasting realized volatility of agricultural commodity futures with infinite Hidden Markov HAR models. *International Journal of Forecasting, 38*(1), 51-73.

Markowitz, H. 1952. "Portfolio Selection," *Journal of Finance, American Finance Association, vol. 7*(1), pages 77-91, March.

Martens, M., Van Dijk, D., & De Pooter, M. (2009). Forecasting S&P 500 volatility: long memory, level shifts, leverage effects, day-of-the-week seasonality, and macroeconomic announcements. *International Journal of forecasting, 25*(2), 282-303.

Merton, R. C. (1980). On estimating the expected return on the market: An exploratory investigation. *Journal of financial economics, 8*(4), 323-361.

Müller, U. A., Dacorogna, M. M., Davé, R. D., Pictet, O. V., Olsen, R. B., & Ward, J. R. (1993). Fractals and intrinsic time: A challenge to econometricians. *Unpublished manuscript, Olsen & Associates, Zürich, 130.*

Özbekler, A. G., Kontonikas, A., & Triantafyllou, A. (2021). Volatility forecasting in European government bond markets. *International Journal of Forecasting, 37*(4), 1691-1709.

Patton, A. J., & Sheppard, K. (2015). Good volatility, bad volatility: Signed jumps and the persistence of volatility. *Review of Economics and Statistics*, *97*(3), 683-697.

Prokopczuk, M., Symeonidis, L., & Wese Simen, C. (2016). Do jumps matter for volatility forecasting? Evidence from energy markets. *Journal of Futures Markets, 36*(8), 758-792.

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

URL http://www.R-project.org/.

Schwert, G. W. (1998). Stock market volatility: Ten years after the crash.

Sharpe, W. F. (1966). Mutual fund performance. *The Journal of business*, *39*(1), 119-138.

Tang, Y., Xiao, X., Wahab, M. I. M., & Ma, F. (2021). The role of oil futures intraday information on predicting US stock market volatility. *Journal of Management Science and Engineering, 6*(1), 64-74.

Tauchen, G., & Zhou, H. (2011). Realized jumps on financial markets and predicting credit spreads. *Journal of Econometrics, 160*(1), 102-118.

Taylor, S. J., & Xu, X. (1997). The incremental volatility information in one million foreign exchange quotations. *Journal of Empirical Finance, 4*(4), 317-340.

Todorova, N., & Souček, M. (2014). Overnight information flow and realized volatility forecasting. *Finance Research Letters, 11*(4), 420-428

Türensal, M. M. (2021). Volatility Estimation Using Heterogeneous Autoregressive Models: The Case of BIST 100. (Master's thesis, Trakya Üniversitesi, Sosyal Bilimler Enstitüsü).

Wang, P. (2009). Modelling and forecasting of realized volatility based on high-frequency data: evidence from FTSE-100 index (Available on Internet) (Master's thesis, Hanken School of Economics).

## APPENDIX A: LIST OF ABBREVIATIONS

**BIST:** Borsa Istanbul

**XU100:** BIST 100 Index

**XU030:** BIST 30 Index

**XUMAL:** BIST Financials Index

**XUSIN:** BIST Industrials Index

**XBANK:** BIST Banks Index

**XKMYA:** BIST Chem. Petrol Plastic Index

**XGIDA:** BIST Food Beverage Index

**XMANA:** BIST Basic Metal Index

**XMESY:** BIST Metal Products Mach. Index

**XTAST:** BIST Nonmetal Min. Product Index

**XTEKS:** BIST Textile Leather Index

**RV:** Realized Volatility

**BPV:** Bipower Variation

**RQ:** Realized Quarticity

**TPQ:** Tripower Quarticity

**HAR:** Heterogeneous Autoregressive Model

**HAR-RV:** Heterogeneous Autoregressive Model with Realized Volatility

**HAR-RV-J:** HAR Model with Jumps

**CHAR:** Continuous HAR Model

**CHARQ:** Continuous HAR Model with Quarticity

**HARQJ:** HAR Model with Quarticity and Jumps

**ARCH:** Autoregressive Conditional Heteroskedasticity Model

**GARCH:** Generalized Autoregressive Conditional Heteroskedasticity Model

**RMSE:** Root Mean Squared Error

**MAE:** Mean Absolute Error

**MAPE:** Mean Absolute Percentage Error

**TIC:** Theil Inequality Coefficient

**MZR:** Mincer-Zarnowitz $R^2$ Statistic

**ME:** Mean Error

**MRE:** Mean Relative Error

**CPA:** Conditional Predictive Ability

**QLIKE:** Quasi–Likelihood Loss Function

**MCS:** Model Confidence Set

**MSE:** Mean Squared Error

**MSPE:** Mean Squared Percentage Error

**LL:** Logarithmic Loss Function