UNIVERSITY OF LJUBLJANA
SCHOOL OF ECONOMICS AND BUSINESS

MASTER'S THESIS

FORECASTING REALIZED VARIANCE: A COMPARISON BETWEEN HAR AND ARFIMA MODELS

Ljubljana, March 2021 AJDA MARJANOVIČ

AUTHORSHIP STATEMENT

The undersigned **Ajda Marjanovič**, a student at the University of Ljubljana, School of Economics and Business, (hereafter: SEB), author of this written final work of studies with the title **Forecasting Realized Variance: A comparison between HAR and ARFIMA models**, prepared under supervision of **prof. dr. Igor Masten**

D E C L A R E

1. this written final work of studies to be based on the results of my own research;
2. the printed form of this written final work of studies to be identical to its electronic form;
3. the text of this written final work of studies to be language-edited and technically in adherence with the SEB's Technical Guidelines for Written Works, which means that I cited and / or quoted works and opinions of other authors in this doctoral dissertation in accordance with the SEB's Technical Guidelines for Written Works;
4. to be aware of the fact that plagiarism (in written or graphical form) is a criminal offence and can be prosecuted in accordance with the Criminal Code of the Republic of Slovenia;
5. to be aware of the consequences a proven plagiarism charge based on this written final work could have for my status at the SEB LU in accordance with the relevant SEB LU Rules;
6. to have obtained all the necessary permits to use the data and works of other authors which are (in written or graphical form) referred to in this written final work of studies and to have clearly marked them;
7. to have acted in accordance with ethical principles during the preparation of this written final work of studies and to have, where necessary, obtained permission of the Ethics Committee;
8. my consent to use the electronic form of this written final work of studies for the detection of content similarity with other written works, using similarity detection software that is connected with the SEB LU Study Information System;
9. to transfer to the University of Ljubljana free of charge, non-exclusively, geographically and time-wise unlimited the right of saving this written final work of studies in the electronic form, the right of its reproduction, as well as the right of making this doctoral dissertation available to the public on the World Wide Web via the Repository of the University of Ljubljana;
10. my consent to publication of my personal data that are included in this written final work of studies and in this declaration, when this written final work of studies is published.

Ljubljana, 8th March 2021 Author's signature: _____

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

ACF	Autocorrelation Function
ADF	Augmented Dickey – Fuller
AR	Autoregressive
AR(FI)MA	Autoregressive (Fractionally Integrated) Moving Average
BIC	Bayesian Information Criterion
BV	Bipower Variation
CEV	Constant Elasticity of Variance
CIR	Cox – Ingersoll – Ross
COMP	NASDAQ Composite
DJIA	Dow Jones Industrial Average
DM	Diebold – Mariano

EWMA	Exponential Weighted Moving Average
(G)ARCH	(Generalised) Autoregressive Conditional Heteroskedasticity
GHAR	Generalised Heterogeneous Autoregressive
GPH	Geweke and Porter – Hudak
HAR	Heterogeneous Autoregressive
HF	High – Frequency
HMH	Heterogeneous Market Hypothesis
ICSS	Iterated Cumulative Sum of Squares
IQ	Integrated Quarticity
IV	Integrated Variance
MA	Moving Average
MAE	Mean Absolute Error
MSE	Mean Squared Error
MEM	Multiplicative Error Model
MIDAS	Mixed Data Sampling
MLE	Maximum Likelihood Estimation
NASDAQ	National Association of Securities Dealers Automated Quotation
NYSE	New York Stock Exchange
OLS	Ordinary Least Squares
OOS	Out-of-sample
QV	Quadratic Variance
RMSE	Root Mean Squared Error
RV	Realized Variance
S&P 500	Standard & Poor's 500
SV	Stochastic Volatility
VARFIMA	Vector Autoregressive Fractionally Integrated Moving Average

INTRODUCTION

Volatility of asset returns plays a central role in financial markets. In one of the first attempts to define it more than a hundred years ago, Louis Bachelier (1900) called it the 'coefficient of instability or nervousness'. Much research has since been devoted to understanding the characteristics of this quantity, which is closely related to the notions of risk and uncertainty. In fact, volatility research has been one of the most active, prominent, and successful areas of research in financial econometrics in the last couple of decades.

The fluctuations in prices and returns are of great importance also for practitioners, since volatility is a key element of risk and portfolio management, option pricing and economic policy. Müller et al. (1995) argued that the state of the market is often best described by looking precisely at its variation. Therefore, having a good understanding of volatility behaviour is important for both researchers and practitioners.

One of the first discoveries about volatility dates back to 1960's, when Mandelbrot (1963) observed that large changes in the markets tend to be followed by large changes and vice versa. Later the term volatility clustering was coined to describe this phenomenon, which lies at the core of many models. Although the concept of volatility is rather intuitive, its statistical analysis is far from trivial due to its unobservable or latent nature.

Recently, the availability of high-frequency data has created new opportunities for research. Nonparametric realized measures of volatility, such as realized variance, constructed from intraday returns have been proposed as a consistent measure of the true volatility, which is unobservable. One of the properties of realized variance is long memory, which is usually defined in terms of a slow, hyperbolically decaying autocorrelation function.

The most commonly used models for long memory processes are fractionally integrated models. The field of long memory began with Hurst, it was further developed by Mandelbrot, but prospered only after Granger and Joyeux (1980) and Hosking (1981) introduced the ARFIMA model. This model is simply the extension of an ARIMA model, which is used for non-stationary or integrated series. While ARIMA allows only for integer values of the differencing parameter, which represents the order of integration, in an ARFIMA model the differencing parameter can take any fractional value.

Another interesting feature, documented by Granger (1980), is that by accumulating simple processes this aggregated series may display long-range dependence. LeBaron (2001) shows that a combination of just three AR(1) processes can simulate long memory, which cannot be empirically distinguished from a true long memory process.

These results have encouraged the development of models based on an additive process with heterogeneous components, for example the HAR model (Corsi, 2009). As an additive model it is able to replicate long memory behaviour, represented by the slow, hyperbolic decay of the autocorrelation function, although it is not formally a long memory model. Because its estimation is rather simple compared to the ARFIMA model, HAR is widely used in practice.

The essence of forecasting is quantification of the mapping from the past and present into the future. Better estimates of the present state thus translate into superior forecasts of the future. Realized volatility excels in this aspect: by exploiting valuable intraday information it provides a relatively precise and quickly-adapting estimate of current volatility (Andersen, Bollerslev, Diebold & Labys, 2003).

The purpose of this master's thesis is to evaluate, which of the two models produces more accurate forecasts of the realized variance. Based on the predictive accuracy and the ease of model estimation and forecasting process we hope to identify the model, which would be recommended for risk management and asset allocation.

We aim to do so by first generating the realized variance from high-frequency prices. To estimate the HAR model a simple ordinary least squares method will be applied, whereas the ARFIMA models will be estimated by maximum likelihood estimation. After verifying that the estimated parameters are statistically significant, we shall proceed with forecasting. We will compare forecasting performance based on the forecasting horizon, different market conditions and different return sampling frequencies. A common measure of the model error will be calculated for each forecasting exercise. Finally, to confirm if one model significantly outperforms the other a statistical test will be used.

Based on recent empirical literature on the topic, our hypotheses are that the HAR model is less sensitive to changes in market conditions and changes in forecasting horizon than the ARFIMA model. This means that HAR's predictive accuracy will drop less for medium- and long-horizon forecasts and it will not vary a lot between different market regimes. Additionally, we hypothesise that the HAR model performs better at lower sampling frequencies, whereas ARFIMA outperforms HAR at the highest sampling frequencies.

The thesis is structured in the following way. First, we provide a short review of literature on volatility modelling and forecasting, with particular emphasis on empirical research of realized variance. In the second chapter we present the theoretical framework, which is essential to understand the principles of the two models, which are presented in the third section. We then describe the data used in the empirical part of the research and the methodology for obtaining forecasts. In section five we present the results of our empirical work, while the sixth section includes some of the limitations of our study and recommendations for further research. Finally, the seventh section concludes this thesis.

1 LITERATURE REVIEW

The concept of realized variance (RV) is often traced back to Merton (1980), who was the first to note that variation over a period of time can be estimated fairly accurately by summing up the squares of intra-day returns if the returns are sampled at sufficiently high frequencies. The first detailed study on the properties of RV came twenty years later, when Andersen, Bollerslev, Diebold and Labys (2001) showed that returns standardised by a sum of squared intra-day returns (realized variance) and the logarithmic transformation of RV

(denoted $lnRV$) are approximately Gaussian distributed, and that realized volatilities appear to be fractionally integrated.

The same authors published a research paper, which set up the foundations for modelling and forecasting realized variance by providing a framework for the use of high-frequency data in measuring and modelling RV and its distribution. This is the first study, which uses fractionally integrated ARMA to model the realized variance. The authors further showed that under the assumption of no microstructure noise, realized variance is a consistent estimator of quadratic variance, which measures the realized sample-path variation of squared returns. It is generated solely by return innovations and represents a unique ex-post realized volatility measure. However, markets are not frictionless, and thus, quadratic variance cannot really be observed (Andersen et al., 2003).

Because this true volatility is latent the choice of volatility measure used for modelling and forecasting is important. Hansen and Lunde (2006a) analysed stock returns and compared eight volatility models with different proxies for unobserved variance. Their results show that realized variance is the preferred measure for comparing various volatility models. Patton (2011) also compared various proxies for volatility and concluded that realized variance leads to the least distorted tests and rankings.

The important issue with realized measures, which exploit high-frequency data is the presence of microstructure noise. Its effects on volatility estimation in high-frequency environments were addressed a couple of years earlier by Zhou (1996). He showed that there is strong negative autocorrelation in tick returns due to the microstructure noise and suggested appropriate adjustments should be made when estimating and forecasting realized variance.

A formal analysis of how market microstructure noise obstructs the use of the limit theory (the convergence of realized variance to quadratic variance) was conducted by Barndorff-Nielsen and Shephard (2002). They refuted the idea that integrated (true) variation can be consistently measured by realized variance and found that the measurement error can be quite large.

Some solutions to construct a measure, which will be optimal in terms of the trade-off between information and noise, were proposed by Zhang, Mykland and Aït-Sahalia (2005) and Hansen and Lunde (2006b). A simple solution is to sample over a slightly longer interval and not use all the tick data (most studies use realized variance constructed from 5-minute returns). More sophisticated methods include subsampling, using kernel-based estimators or pre-filtering intraday returns (McAleer & Medeiros, 2008).

When analysing markets, which are not open 24 hours per day (most of them, except foreign exchange markets), there is another important issue to address – the overnight return[1]. The simplest approach one can take is to simply ignore the overnight return. Hansen and Lunde

[1] As a result of after-hours or pre-market trading.

(2006b) claim that by doing so we do not estimate the true volatility. However, most studies do exactly that. Still, some alternative approaches have been proposed, such as different weighting schemes for overnight and intraday returns. Ahoniemi and Lanne (2013) find that usually a combination of squared overnight return and squared intra-day returns is the most accurate measure of realized volatility, but that in some cases considering only intra-day information is more accurate.

Financial asset prices and returns share some statistical properties, which persist over many markets and asset classes. They are often referred to as stylised facts. These include absence of autocorrelation in returns but dependence in absolute and squared returns, return distribution with heavy tails, leverage effect and gain-loss asymmetry, among others.

Some of the most popular volatility models such as the generalised autoregressive conditional heteroskedasticity (GARCH) or stochastic volatility (SV) models cannot replicate all the stylised facts described above (Breidt, Crato & De Lima, 1998; Nelson, 1990). Especially the 'first generation' of SV models like the Cox-Ingersoll-Ross (CIR) or Constant Elasticity of Variance (CEV) models were rejected in practice (Andersen & Lund, 1997).

But correctly identifying return dynamics is crucial for accurate forecasting, so researchers have turned to the analysis of high-frequency data and realized variance as a measure that can capture true volatility. Realized variance exhibits properties of a fractionally integrated process, and therefore much research has shifted towards long-memory models.

Traditional long-memory models are fractionally integrated GARCH and autoregressive fractionally integrated moving average (ARFIMA), which was introduced by Granger and Joyeaux (1980) and Hosking (1981). Some of the more prominent studies advocating the use of long-memory models for realized variance are Areal and Taylor (2002), Andersen et al. (2003), Koopman, Jungbacker & Hol (2005) and Martens, van Dijk & de Pooter (2009).

Of course, there are alternative ways to accommodate for the long-memory property. One of these is an additive cascade model by Corsi (2009) named heterogeneous autoregressive (HAR) model. It builds on ideas by Granger (1980) that a combination of short-memory processes can generate long-memory features, but it is remarkable that this can be achieved by summation of only three different AR(1) processes as shown by LeBaron (2001). The HAR model is able to reproduce the hyperbolic decay of autocorrelations, while remaining a simple model to estimate in contrast to fractionally integrated models.

There have been numerous extensions of both ARFIMA and HAR models in the recent years. In a multivariate setting, Chiriac and Voev (2011) used VARFIMA model for RV, while Čech and Barunik (2017) introduced the GHAR (generalized HAR) model. Paye (2012) and Christiansen, Schmeling and Schrimpf (2012) extended the HAR model by including exogenous variables. Their results suggest that such variables have limited ability to improve forecasts. Especially macroeconomic variables have not been found to improve

results, while some financial indicator proxies for credit risks and illiquidity are slightly more informative about future volatility.

Other models for realized measures that are worth mentioning are the Multiplicative Errors Model (MEM) by Engle and Gallo (2006), HEAVY model by Shephard and Sheppard (2010), Realized GARCH by Hansen, Huang and Shek (2012) and the MIDAS model by Bai, Ghysels and Wright (2013).

Since so many models were developed, the researchers quickly started comparing their performance. Pong et al. (2004) were the first to compare GARCH forecasts on daily returns, ARMA and ARFIMA models for intraday returns and implied volatilities from option prices. Their results depend importantly on the length of the forecasting horizon. For daily and weekly forecasts, the ARFIMA and ARMA models are superior, whereas for longer horizons (monthly and quarterly) implied volatilities are the most accurate forecasts.

According to Martens et al. (2004) ARFIMA models outperform models from the ARCH class, the Riskmetrics' exponentially weighted moving average (EWMA) model and stochastic volatility (SV) models and can compete with implied volatility forecasts. Koopman et al. (2005) came to the same conclusion by comparing unobserved ARMA components, SV, GARCH and ARFIMA models. In their study, the ARFIMA-lnRV model generated the most accurate forecasts.

When Corsi (2009) proposed the HAR model, he claimed that it outperforms the existing volatility models. He compared its performance to a specification of an ARFIMA model and concluded that HAR is more accurate based on some series and comparable with ARFIMA in other cases.

Although a significant body of literature has found that the HAR and ARFIMA models outperform other volatility models like GARCH and SV, to our knowledge only two studies have explicitly compared the two long memory models. Degiannakis and Floros (2013) compared the performance of ARFIMA and HAR based on daily forecasts and found ARFIMA to be superior. On the other hand, Izzeldin, Hassan, Pappas & Tsionas (2019) argued that HAR forecasts are more stable across different horizons. Their results suggest that HAR is less sensitive to variations in market regimes, but ARFIMA is superior if returns are sampled at ultra-high frequencies. Their conclusion refutes the idea that one of the two models is generally superior.

2 THEORETICAL FRAMEWORK

To better understand the econometric models used in the empirical part we briefly present the underlying theory in this section. We start by describing some characteristics of financial data, we then define the variation processes and conclude with two particular topics, which play an important role in realized variance forecasting: market microstructure effects and the long memory property.

2.1 Stylised facts of financial data

In financial markets we observe asset prices, but most of the empirical research is focused on returns. This allows for normalisation, that is, we can compare the values across assets regardless of the starting price level. Additionally, prices are usually non-stationary, while the returns are stationary (Jondeau, Poon & Rockinger, 2007). It is standard practice to work with continuously compounded or log returns r_t, which have the advantage that the n-period return is just the sum of the corresponding one-period returns:

$$r_t = \ln(P_t) - \ln(P_{t-1}) = p_t - p_{t-1} \tag{1}$$

$$r_{t,n} = \sum_{k=0}^{n-1} r_{t-k} = p_t - p_{t-n} \tag{2}$$

Usually, simple (arithmetic) returns and prices are written in capital letters, while log prices and returns are denoted in small letters. We shall follow this convention throughout this thesis.

The first important common property of financial data or stylised fact is high kurtosis in asset return distributions. The unconditional return distributions exhibit what are generally known as heavy tails. Even after correcting returns for volatility clustering, the conditional distribution still exhibits heavy tails although to a lesser extent than the unconditional one. However, by increasing the time horizon over which the returns are calculated, the distribution becomes more and more Gaussian (Cont, 2001).

This non-Gaussian distribution of financial returns has been known for more than 50 years. Mandelbrot (1963) and Fama (1965) have already shown that the return distributions have a higher kurtosis than a normal distribution and are negatively skewed (more mass is in the left tail of the distribution than in the right one). These two properties have important implications for risk and portfolio management. The first one implies more extreme values or events than would be expected if returns were normally distributed, while the second property implies that there will be more extreme negative shocks than positive ones (Jondeau et al., 2007).

Along with the distributional properties it is in our interest to look also at the dependence properties of returns. The first widely observed property is the absence of linear autocorrelation in liquid markets (Fama, 1970). This is due to the fact that if there was significant autocorrelation in the markets, it would be easy to implement a trading strategy with positive expected returns. Such strategies called statistical arbitrage reduce the autocorrelation (Cont, 2001). In high-frequency return series, however, negative autocorrelation is observed between transaction prices. This is due to the so-called bid-ask bounce, where the transactions are carried out close to the bid or ask price and tend to bounce between these two quotes (Campbel, Lo & MacKinlay, 1997).

Nevertheless, the independence of returns does not imply the independence of the increments. Even simple nonlinear functions of returns like absolute values or squares of returns exhibit noteworthy positive autocorrelation or persistence (Taylor, 1986; Ding, Granger & Engle, 1993). This phenomenon, known as volatility clustering, has been documented already in the middle of the last century: "*large changes tend to be followed by large changes - of either sign - and small changes tend to be followed by small changes…*" (Mandelbrot, 1963, 418).

These observations of statistical properties lead to a decomposition of the return into a product of a white noise, uncorrelated in time, ε_t, and a conditional volatility factor $\sigma_{t,h}$.

$$r_{t,h} = \sigma_{t,h}\varepsilon_t \tag{3}$$

In this decomposition the volatility component is latent, which means it is not directly observable, whereas the returns are observable. Consequently, volatility depends on the model, while correlations of absolute returns can be computed from data (Cont, 2001).

Finally, other stylised facts related to volatility are the correlation of trading volume and volatility and the asymmetry in volatility propagation. Volatility measured over longer time scales is better at predicting volatility measured at short time scales (Cont, 2001). This last property is an important motivation for the heterogeneous autoregressive model, which will be used as one of the forecasting models in this thesis.

2.2 Definition of variation processes

We must introduce and define some concepts of volatility, which will be useful for the discussion of the theory of realized measures. These concepts are quadratic, notional and integrated variance.

To measure return volatility, we must be able to identify a component of the price increment, which does not describe an expected price movement, but rather its innovation. In a discrete-time framework, this can be achieved through an asset pricing model. In a continuous-time setting, because of the no-arbitrage condition, the return innovation is one order of magnitude larger than the average return (Andersen et al., 2010).

Let us first define a univariate log price process $p(t)$ defined on $(\Omega, \mathcal{F}, P)$, that evolves in continuous time over $[0, T]$ as

$$dp(t) = \mu(t)dt + \sigma(t)dW(t), \qquad t \geq 0 \tag{5}$$

where $dp(t)$ is the change in the log price[2], $\mu(t)$ is the drift term with continuous and locally bounded volatility, $\sigma(t)$is a strictly positive volatility process (instantaneous volatility) and $W(t)$ is Brownian motion (Andersen et al., 2010).

[2] We use round brackets to define variables in continuous time and subscripts for variables in discrete time.

By recalling equation (2) the return over the $[0, t]$ interval is then given by

$$r(t) \equiv r(t,t) = p(t) - p(0), \qquad 0 \leq t \leq T \tag{6}$$

The continuously compounded return over the interval $[t-h, t]$ of length h is similarly

$$r(t,h) = r(t) - r(t-h), \qquad 0 \leq h \leq t \leq T \tag{7}$$

We will keep the notation of the time interval $[t-h, t]$, for which $0 \leq h \leq t \leq T$ always holds.

Some assumptions about the price and return processes in the frictionless framework imply that the log-price process is a special semimartingale (Andersen et al., 2010). A special semimartingale can be decomposed uniquely into a sum of a finite variation right-continuous process and a local martingale. A stochastic process is a martingale with respect to the filtration $\mathbb{F} = (\mathcal{F}_t)_{0 \leq t \leq \infty}$ if the expected value of the realisation at time $t+1$ is equal to the realisation at time t and does not depend on any previous realisations (Protter, 2004).

Therefore, by assuming that asset prices are strictly positive and finite, the price and return processes are square-integrable, we are operating in the arbitrage-free framework and expected returns are finite (Back, 1991) we can decompose the no-arbitrage log price process in the following way

$$r(t) \equiv p(t) - p(0) = \mu(t) + M(t) \tag{8}$$

where $\mu(t)$ is a predictable, finite-variation process, and $M(t)$ is a local martingale, which can be further decomposed into a continuous sample path $M^C(t)$ and a jump martingale $M^J(t)$ (Andersen et al., 2010)

$$M(t) = M^C(t) + M^J(t) \tag{9}$$

The expected return process $\mu(t)$ may also be decomposed into a finite-variance continuous and predictable jump component. The continuous part, let us denote it $\mu^c(t)$, must be locally an order of magnitude smaller than the corresponding continuous part of the martingale innovation $M^C(t)$. To explain it simply, if this was not the case, a continued long position in a risky asset would be diversified due to the law of large numbers and would distort the risk-return trade-off. The no-arbitrage condition holds only if the innovations are large or if there is a nontrivial jump martingale component (Andersen et al., 2010).

We need continuous price data to observe the $M(t)$ process and analyse the innovation component. Due to the market microstructure effects, which will be explained in the section 2.4, such data cannot be obtained. The result is that research is oriented towards measuring volatility over some time interval rather than the instantaneous volatility at a precise moment in time.

If we have a special semimartingale $X(t)$ then the quadratic variance associated with it is defined by

$$[X,X]_t \equiv X(t)^2 - 2\int_0^t X(-s)dX(s), \qquad 0 < t \leq T \tag{10}$$

where the stochastic integral of a left-continuous process $X(-s)$ is well-defined. The quadratic variance is an increasing and adapted (non-anticipative) process (Protter, 2004).

By dividing the time interval $[0,T]$ into m partitions ($0 = \tau_{m,1} \leq \tau_{m,2} \ldots \leq \tau_{m,m} = T$) then for $t \in [0,T]$ and $t \wedge \tau \equiv \min(t,\tau)$

$$\lim_{m\to\infty}\left(\sum_{j\geq 1}[X(t \wedge \tau_{m,j}) - X(t \wedge \tau_{m,j-1})]^2\right) \to [X,X]_t \tag{11}$$

This is a crucial result, because it states that the quadratic variance of the process represents the cumulative realised sample-path of a semimartingale over a time interval (Andersen et al., 2010).

If we take into account that the process $X^2 - [X,X]$ is a local martingale if X is a locally square integrable martingale, and we recall that the quadratic variation of finite variation processes (e.g. the continuous component of expected return) is zero then under the assumptions of no predictable jumps we obtain the definition of notional variance. Sometimes referred to as actual variance it is equal to the increment in quadratic variance

$$\begin{aligned} v^2(t,h) &\equiv [r,r]_t - [r,r]_{t-h} \\ &= [M^c,M^c]_t - [M^c,M^c]_{t-h} + \sum_{t-h<s\leq t}\Delta r^2(s) \end{aligned} \tag{12}$$

Equation (12) also implies that only the return innovations influence the quadratic variance regardless of the conditional mean of the process. Equations (11) and (12) therefore suggest that the quadratic variation can be approximated by a sum of the instantaneous squared returns without imposing a specific model on the price process (Andersen et al., 2010).

2.3 Realized variance

In a continuous framework the continuously compounded return is thus given by

$$r(t,h) = p(t) - p(t-h) = \int_{t-h}^{t}\mu(s)ds + \int_{t-h}^{t}\sigma(s)dW(s) \tag{13}$$

An equivalent representation is the stochastic differential equation for the log price process given already by equation (5). In this framework the quadratic variance (QV) is equal to the integrated variance (IV)

$$QV_t = [M,M]_t - [M,M]_{t-h} = \int_{t-h}^{t} \sigma(s)^2 ds = IV_t \tag{14}$$

It is important to note that in a more general framework, where a diffusion model includes jumps, the quadratic and integrated variance do not coincide (Andersen et al., 2010).

The above results suggest that we can estimate the unobservable integrated variance relatively precisely over a period of time by accumulation of squared returns over infinitesimal periods of time. This estimator is the realized variance.

To construct it, daily data is split into n subintervals of equal length. The intra-day returns are then defined as $r_{t,i} = p_{t,i} - p_{t,i-1}$ for $i = 1, \dots n$. These returns are first squared and then summed to obtain the realized variance

$$RV_t = \sum_{i=1}^{n} r_{t,i}^2 \tag{15}$$

Barndorff-Nielsen and Shephard (2002) and Andersen et al. (2003) demonstrated that the realized variance is a consistent estimator of the integrated variance in the absence of microstructure noise. They also derived the asymptotic distribution of RV

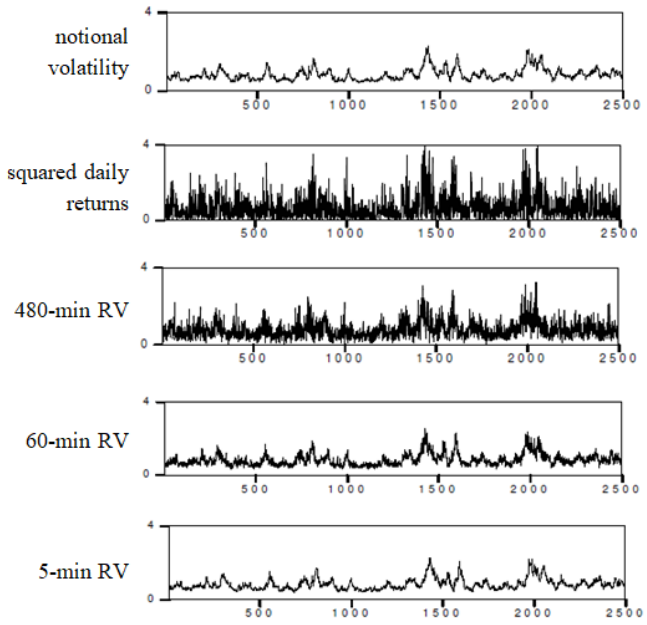
$$\sqrt{n}(RV_t - IV_t) \xrightarrow{d} N(0, 2IQ_t) \tag{16}$$

where IQ_t is the integrated quarticity, defined as

$$IQ_t = \int_{t-h}^{t} \sigma^4(s) ds \tag{17}$$

Figure 1 shows the convergence of realized variance to the integrated variance as $n \to \infty$. Squared daily returns (second panel, $n = 1$) are a noisy measure of notional variance. It retains the general shape, but is far from being a precise measure of true volatility. However, by sampling at higher and higher frequencies, the realized variance is becoming more similar to the notional variance. In the last panel ($n = 288$, 5-minute returns in a 24-hour foreign exchange market) the realized variance is almost identical to the integrated variance in the first panel (Andersen et al., 2010).

Figure 1: Convergence of realized volatility to notional volatility



Source: Andersen et al. (2010).

The theory of realized variance allows us to directly look at the distribution of the true ex-post variance. One of its advantages is that it holds across different asset classes (Andersen, Bollerslev, Diebold and Ebens, 2001; Andersen, Bollerslev, Diebold and Labys, 2001).

The simple RV model can be extended to include the jump components. However, in the presence of jumps, RV_t converges in probability to a quantity that includes also a jump process and is thus not a consistent estimator of the integrated variance. Therefore, the realized variance provides an ex-post measure of the true volatility accounting also for the discontinuous jump part (Baillie, Calonacci, Cho & Rho, 2019).

In the presence of jumps in the price process, the continuous-time jump diffusion becomes

$$dp(t) = \mu(t)dt + \sigma(t)dW(t) + \kappa(t)dq(t) \tag{18}$$

where $q(t)$ is a counting process of number of jumps up until time t and $\kappa(t)$ is the size of a jump at time t. As explained above, the realized variance then converges to the true variance including the jump part (Barndorff-Nielsen & Shephard, 2004)

$$RV_t \rightarrow \int_{t-1}^{t} \sigma^2(s)ds + \sum_{t-1<s\leq t} \kappa^2(s) \tag{19}$$

The two components can be separated by making specific parametric assumptions about the diffusion process or by using the realized bipower variation (BV). In presence of jumps, the

BV can be used to consistently estimate integrated variance (Barndorff-Nielsen & Shephard, 2004). However, little predictability comes from jumps (Andersen et al., 2010).

A serious impediment to the use of the limit theory by Andersen et al. (2003) and Barndorff-Nielsen and Shephard (2002) is that it is based on the assumption of absence of microstructure noise. However, at ultra-high frequencies the true price (the signal) is contaminated by the market microstructure noise.

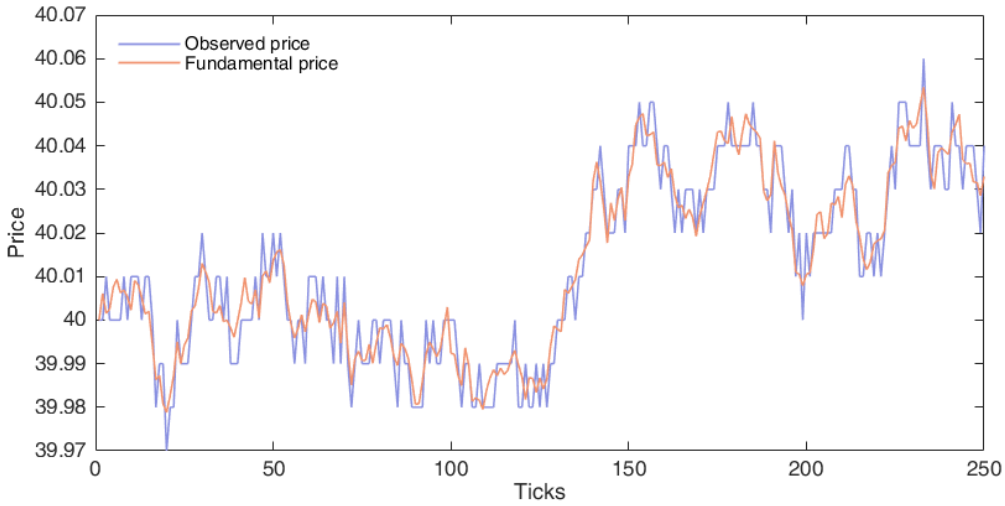
2.4 Market microstructure noise

Market microstructure effects can have a significant impact on statistical properties of asset prices, especially in the high-frequency framework. The assumption of continuous prices is invalid as the prices we observe in the markets are discrete, often quoted to two decimal places and analogously increments cannot be infinitesimally small (Harris, 1990, 1991). Besides, securities sometimes do not trade at evenly spaced intervals or do not trade at all for a period of time, a feature especially pronounced in more illiquid markets (Amihud & Mendelson, 1987; Lo & MacKinlay, 1990). Finally, the bid-ask bounce is responsible for negative serial correlation in returns (Amihud & Mendelson, 1986; Roll, 1984).

The bid-ask bounce is presented in Figure 2, where we simulate a random walk as a fundamental price process and plot simulated observed prices, which randomly bounce between the bid and ask quotes.

To eliminate the bid-ask bounce effect, Zhou (1996) used only bid prices to calculate tick-by-tick returns yet he still found negative autocorrelation. To understand why the high frequency data is so contaminated by noise he spent several days on a trading floor. He found additional sources of noise such as fighting-screen effect, where traders constantly update their quotes to remain on the trading screen even if the fundamental price has not changed. Another very trivial sources of noise are typographical errors and delayed quotes.

Figure 2: Simulated fundamental and observed prices



Source: Own work.

Hansen and Lunde (2006b) found that the presence of microstructure noise significantly complicates volatility estimation and, as a result, estimators like realized variance become unreliable. Market microstructure puts under question the validity of theoretical results, which are based on the assumption of the absence of market microstructure. Below we demonstrate why microstructure noise poses an obstacle to infinitesimal sampling.

By dividing the day t into n sub-intervals of length $\Delta = \frac{1}{n}$, we have on the i-th sub-interval

$$p_{t,i}(\Delta) = p^*_{t,i}(\Delta) + \varepsilon_{t,i}(\Delta) \tag{20}$$

where $p_{t,i}$ is the observed price, $p^*_{t,i}$ is the fundamental (unobserved) price and $\varepsilon_{t,i}$ is the noise term with variance σ^2_ε. The intra-day return is then

$$r_{t,i}(\Delta) = r^*_{t,i}(\Delta) + \upsilon_{t,i}(\Delta) \tag{21}$$

We assume that $\varepsilon_{t,i}$ is independent and identically distributed (i.i.d.) and the variance of $\upsilon_{t,i}$ is therefore $\sigma^2_\upsilon = 2\sigma^2_\varepsilon$.

The realized variance based on noisy returns is then

$$\mathrm{RV_t}(\Delta) = \sum_{\mathrm{i=1}}^{\mathrm{n}} \mathrm{r}^2_{\mathrm{t,i}} = \sum_{\mathrm{i=1}}^{\mathrm{n}} (\mathrm{r}^*_{\mathrm{t,i}} + \upsilon_{\mathrm{t,i}})^2 = \mathrm{IV_t} + \mathrm{u_t}(\Delta, \sigma^2_\upsilon) \tag{22}$$

where the error u_t that is generated every time we observe realized variance is given by

$$\mathrm{u_t}(\Delta, \sigma^2_\upsilon) = \eta_\mathrm{t}(\Delta) + 2\sum_{\mathrm{i=1}}^{\mathrm{n}} \sigma_{\mathrm{t,i,\Delta}}\, \mathrm{z_{t,i}} \upsilon_{\mathrm{t,i,\Delta}} + \sum_{\mathrm{i=1}}^{\mathrm{n}} \upsilon^2_{\mathrm{t,i}}(\Delta) \tag{23}$$

where $\eta_t(\Delta)$ is the discretisation error, which depends only on the sampling frequency Δ. In the limit this component will go to zero. For a given $\Delta > 0$ we have

$$E[RV_t(\Delta)] = E[IV_t] + E[u_t(\Delta)] \tag{24}$$

with $E[IV_t] = \omega$ and $E[u_t(\Delta, \sigma^2_\upsilon)] = \frac{1}{\Delta}\sigma^2_\upsilon$. If we sampled at infinitesimally small intervals (that is $\Delta \to 0$) we would get infinite amount of noise (Zhang et al., 2005)

$$\lim_{\Delta \to 0} E[u_t(\Delta, \sigma^2_\upsilon)] = \lim_{\Delta \to 0} \frac{1}{\Delta}\sigma^2_\upsilon = \infty \tag{25}$$

At ultra-high frequencies, the discreteness of the price grid makes the continuous-time models inappropriate. When tick data (data of all the transactions carried out) is not available at evenly spaced intervals, some kind of interpolation can be used to calculate returns (Andersen & Bollerslev, 1997). However, as shown by Hansen & Lunde (2006b) if realized variance is computed based on this linear interpolation method, it converges to 0 as sampling

frequency increases. In other words, "*the quadratic variation of a straight line is zero*" (Hansen & Lunde, 2006b, 130). Even though this only holds in the limit, the previous tick method is preferred to linear interpolation when constructing returns.

Intraday returns can therefore be computed by using different sampling schemes. Most often, calendar time sampling is used, where prices used for return construction are equidistant in calendar time (n-minute sampling). Returns can be computed between every n^{th} transaction, which is referred to as tick time sampling. Lastly, we have business time sampling, where sampling times are unobservable because they are defined in relation to integrated variance (Hansen & Lunde, 2006b).

Hansen and Lunde (2006b) claimed that the best way to limit the impact of market microstructure noise is to study and understand its properties. In their study they discovered some interconnected facts about microstructure noise, namely, that the noise is time-dependent, correlated with the efficient price, and that its properties have changed considerably over time.

All of the results above suggest that when we generate the realized variance we are met with a trade-off between sampling at the highest possible frequency following the theory of quadratic variation and sampling at a frequency low enough to get rid of strong market microstructure effects. Andersen, Bollerslev, Diebold and Ebens (2001) argued that the decision ultimately depends on the market liquidity. The more liquid the market, the smaller the sampling interval can be, so that the measurement is not too contaminated by the microstructure noise. The most frequent interval used in studies is sampling over 5 minutes, which ensures that RV is quite accurately measured (Andersen & Bollerslev, 1998) and not too affected by market microstructure.

Zhang et al. (2005) proposed an alternative to consistently estimate integrated variance. By using all data through subsampling, averaging and then correcting for the bias, they introduced a consistent estimator, albeit it converged rather slowly. The authors found that there is little difference between estimators based on sparse sampling (e.g. 5-minute) and estimators based on subsampling and averaging. On the other hand, their two-scale estimator, which explicitly accounts for the microstructure bias, produced more accurate results.

Contrarily, Hansen and Lunde (2006b) discovered that the bias due to the microstructure noise in realized variance is relatively small and without a dramatic impact. Therefore, since the bias is negligible, sampling at higher frequencies is preferred. They argued that the market microstructure effects are very complex and cannot be resolved altogether by a simple specification for noise.

In a similar manner Liu, Patton and Sheppard (2015) compared many estimators of the integrated variance and concluded that by taking 5-minute RV as a benchmark none of the competing measures outperformed it significantly. When using the model confidence set by

Hansen et al. (2011) the 1-minute subsampled realized variance and some kernel-based methods were superior. Finally, 5- and 15-minute realized variances were found to be the most accurate when comparing forecasting performance.

In this thesis we will follow the law of parsimony and work with n −minute RV without subsampling, since the subsampling does not recognisably improve forecasts, we are not analysing tick-by-tick data and many studies suggest that simpler measures outweigh more complicated ones with specific noise specifications.

2.5 Long memory

Harold Edwin Hurst studied the hydrological properties of the irregular flows of the river Nile in the middle of the 20th Century. He observed long stretches of either high or low floods, even if there was no regularity in the occurrence of such episodes. His observations motivated many studies, which led to the development of the long memory concept. Hurst's work inspired Mandelbrot (1963), which found that his observations can be described by self-similarity, a typical property of fractals. By using this concept, the foundations were laid for the development of the long-memory models (Graves, Gramacy, Watkins & Franzke, 2017).

Beran (1994) nicely described the qualitative characteristics of a typical long-memory process. Such a time series exhibits relatively long periods, where realizations tend to remain at low or high values. Furthermore, looking at short time periods, cycles or trends could be recognised, which disappear when looking at the series as a whole. Finally, the sample path seems stationary.

There are also quantitative properties, which characterise such long-range persistent processes. Firstly, the variance of the sample mean decays to zero at an approximate rate of $n^{-\alpha}$, where $\alpha \in (0,1)$, which is a slower rate than n^{-1}, where n is the sample size. Secondly, sample autocorrelations $\widehat{\rho_h}$ decay to zero at a rate, which is approximately proportional to $h^{-\alpha}$ for $\alpha \in (0,1)$, where h is the autocorrelation order. Lastly, near the origin, the plot of log periodogram $I(\lambda)$ against the log frequency is scattered around a negatively sloped straight line (Beran, 1994).

By assuming that the series is stationary these latter features can be reformulated. The variance of the sample mean is in the limit equal to a constant multiplied by $n^{-\alpha}$ for $\alpha \in (0,1)$. The autocorrelations ρ_h are asymptotically equal to a constant multiplied by $h^{-\alpha}$ for $\alpha \in (0,1)$. Finally, the spectral density $f(\lambda)$ has a pole at zero, which equals a constant multiplied by $\lambda^{-\beta}$ where $\beta \in (0,1)$. If these properties hold, the series is said to be a stationary process with slowly decaying correlations or long memory[3] (Beran, 1994).

We introduce the Hurst parameter as a measure of long memory. It is related to the property of the rate at which the autocorrelations decrease, and it is given by $H = 1 - \frac{\alpha}{2}$. In terms of

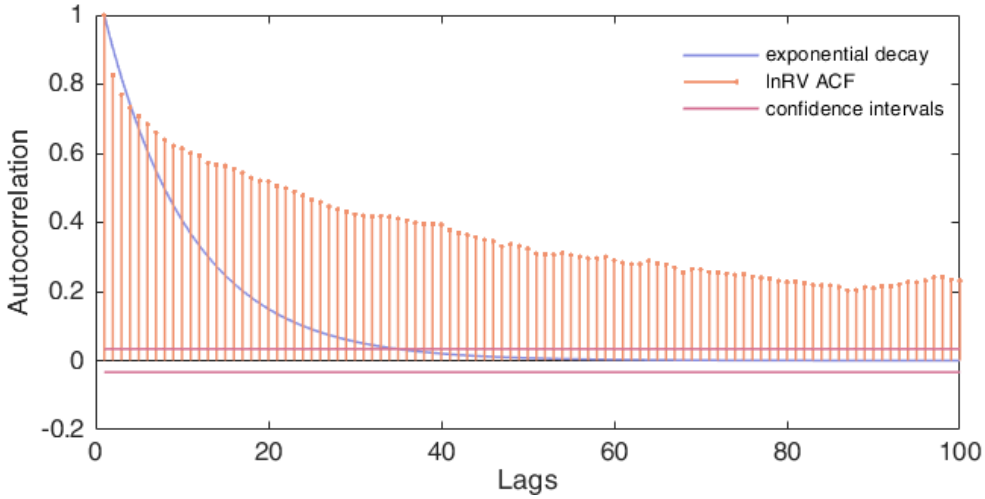
[3] Other terms for long memory used in the literature are long-range or strong dependence or persistence.

the Hurst exponent, long memory property arises for $H \in (0.5,1)$. We will later use the fractional differencing parameter d, which is related to the Hurst parameter by the formula $d = H - \frac{1}{2}$. By substitution $-\alpha = 2d - 1$, so (Beran, 1994)

$$\lim_{h\to\infty} \rho_h = c(d, \varphi, \psi) \cdot |h|^{2d-1} \tag{4}$$

In Figure 3 we present the sample autocorrelation function (ACF) for the logarithm of realized variance. It is clear from the figure that autocorrelations decay at a hyperbolic rate, typical of a long memory process and do not follow an exponential decay, which is characteristic of a short memory process.

Figure 3: Autocorrelation function for lnRV



Source: Own work.

3 MODELS

Volatility of asset returns has often been modelled by GARCH or stochastic volatility models, where volatility is treated as a latent variable. The theory of realized variance, which is based on a different approach of modelling volatility by constructing a time series allows the researchers to treat volatility as an observed variable and thus, standard time-series models like ARFIMA can be used for RV modelling (Corsi et al., 2008).

In this section we present the two models, which we use in the empirical part. The ARFIMA model has long been the most accurate model for realized variance but has seen its dominance fade with the development of the rather simple HAR model. We derive the models, describe the model selection and estimation procedure, and provide the framework for forecasting.

3.1 Autoregressive Fractionally Integrated Moving Average (ARFIMA) model

By recalling Hurst's observations about the behaviour of volatile river flows, Granger vaguely put forward the idea of 'fractional' differencing in one of his papers in 1980. He and Hosking, then independently of one another created the foundations for a new class of ARFIMA models, which are widely used for long memory processes in economics, finance, earth sciences and in other areas. We will use it to model realized variance, as has been done by Andersen et al. (2001), Martens et al. (2004), Koopman et al. (2005), Baillie et al. (2019) and many others.

3.1.1 Derivation

We start with a standard ARMA(p, q) model, which is one of the most common models used in time series analysis

$$y_t = \varphi_1 y_{t-1} + \cdots + \varphi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} \tag{26}$$

The value of the process at time t, depends on the past values of the process (up to a certain lag) and on the past values of the error terms (with the additional term, which represents the error at time t). By introducing lag polynomials we can rewrite the model as

$$\Phi(L)(y_t - \mu) = \theta(L)\varepsilon_t \tag{27}$$

where μ is the process mean, ε_t is a white noise process, $\Phi(L)$ is the autoregressive lag polynomial of the form

$$\Phi(L) = 1 - \varphi_1 L - \varphi_2 L^2 - \ldots - \varphi_p L^p \tag{28}$$

and $\theta(L)$ is the moving average lag polynomial of the form

$$\theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \ldots + \theta_q L^q \tag{29}$$

L is the lag operator, which transforms the series so that: $Ly_t = y_{t-1}$. Lag operator can be applied multiple times to the series. In general, $L^k y_t = y_{t-k}$ (Hamilton, 1994).

ARMA models are used to model stationary series. For non-stationary data, where the series must be differenced d-times[4] to obtain a stationary series, the ARIMA model was introduced by Box and Jenkins (1970), which is given by the following expression

$$\Phi(L)(1 - L)^d(y_t - \mu) = \theta(L)\varepsilon_t \tag{30}$$

[4] Here, d is an integer.

where d is an integer and usually equal to 1. The ARIMA can be seen as a generalisation of ARMA, since we obtain the latter by setting $d = 0$, which corresponds to the fact that stationary series are said to be integrated of order 0 (an $I(0)$ series).

By the same logic, fractional ARIMA models are a generalisation of the classic ARIMA models by allowing the d parameter to be fractional. The fractional difference operator $(1 - L)^d$ is defined by the following binomial expansion

$$(1 - \mathrm{L})^{\mathrm{d}} = \sum_{\mathrm{k}=0}^{\infty} \delta_{\mathrm{k}} \mathrm{L}^{\mathrm{k}} = \sum_{\mathrm{k}=0}^{\infty} \binom{\mathrm{d}}{\mathrm{k}} (-\mathrm{L})^{\mathrm{k}} \tag{31}$$

The ARFIMA(p, d, q) model for y_t is then given as

$$\Phi(\mathrm{L})(1 - \mathrm{L})^{\mathrm{d}}(\mathrm{y_t} - \mu) = \theta(\mathrm{L})\varepsilon_{\mathrm{t}} \tag{32}$$

where d takes on fractional values.

Since the series we model is the realized variance, the ARFIMA-RV model is given by

$$\Phi(\mathrm{L})(1 - \mathrm{L})^{\mathrm{d}}(RV_{\mathrm{t}} - \mu) = \theta(\mathrm{L})\varepsilon_{\mathrm{t}} \tag{33}$$

The ARMA part of the model is assumed to be invertible (moving average polynomial parameters $|\theta| < 1$) and stationary (autoregressive polynomial parameters $|\varphi| < 1$) (Hamilton, 1994). If the lag polynomials do not share common roots, then the demeaned process $y_t - \mu_t$ is said to be integrated of order d, abbreviated as $I(d)$. The properties of this process depend importantly on the value of the fractional differencing parameter d (Doornik & Ooms, 2004).

If $d > 0$ the process exhibits long memory and for $d < 1$ the process is mean reverting. If $d = 0$ the process is a white noise, with zero correlations. For $d = 0.5$ the process is stationary but not invertible. We are especially interested in the following two cases: when $0 < d < 0.5$ the autocovariance function decays hyperbolically, whereas if $-0.5 < d < 0$ the process is over-differenced and the inverse autocorrelations decay hyperbolically. The latter is often referred to as intermediate memory or an anti-persistent process (Hosking, 1981). Both cases where the fractional difference parameter is in absolute terms smaller than 0.5 are covariance stationary (Doornik & Ooms, 2004).

The process is covariance-stationary or weakly stationary if its mean and autocovariances do not depend on date t, that is

$$\mathrm{E}[\mathrm{Y_t}] = \mu\,, \quad \text{for all t} \tag{34}$$

$$\mathrm{E}\big[(\mathrm{Y_t} - \mu)(\mathrm{Y_{t-j}} - \mu)\big] = \gamma_{\mathrm{j}}, \quad \text{for all t and any j} \tag{35}$$

Equation (34) indicates that if a process is weakly stationary, the covariance between two observations depends only on the distance between them (j) and not on their position in the series (t) (Hamilton, 1994).

3.1.2 Model selection and estimation

There are two main approaches for estimating an ARFIMA(p, d, q) model. The first one is by exact maximum likelihood estimation (MLE), as described by Sowell (1992). The disadvantage of this method is that the number of autoregressive and moving average parameters must be known beforehand. This creates the problem of how to select the appropriate number of parameters.

The alternative is to use a semi-parametric approach. The most popular of these is the Geweke and Porter-Hudak (GPH) approach (Geweke & Porter-Hudak, 1983). While all the parameters can be estimated simultaneously by MLE, the semi-parametric approach consists of two steps: firstly, the fractional difference parameter d is estimated and, secondly, all the other parameters are estimated later (Reisen, Abraham & Lopes, 2001).

The GPH estimation of the fractional differencing parameter d is based on a log-periodogram regression. The second step of the estimation involves fitting an ARMA model to the data filtered according to the estimate of d from the first step. However, this estimator has been shown to suffer from a considerable finite sample bias and inefficiency if ε_t is persistent (Agiakloglou et al., 1993). Nevertheless, the GPH approach is quite widely used in the literature.

Another popular semi-parametric estimator is the Whittle estimator, which is also based on a periodogram analysis. Other estimators are the rescaled range estimator, which builds on the idea that a ratio of range rescaled by the standard deviation converges in probability to a certain constant. Lo (1991) proposed a modified rescaled range estimator, where the standard deviation is replaced by a heteroskedasticity and autocorrelation consistent estimator of volatility.

To summarise, these estimation methods estimate the fractional differencing parameter d in the first step. Then, an ARMA model is fitted to the filtered data. In this second step the Bayesian Information Criterion (BIC) is used for lag selection and maximum likelihood estimation is used to estimate all the parameters (Bhardwaj & Swanson, 2006).

In the literature, each study analysing RV selects the number of lags in an ARFIMA model arbitrarily: Koopman et al. (2005) claim there is no significant difference between different specifications and thus estimate ARFIMA$(1, d, 0)$, Degiannakis & Floros (2013) estimate 6 different models with $p = \{0, 1, 2\}$ and $q = \{0, 1\}$, Corsi, (2009) compares his HAR model to an ARFIMA$(5, d, 0)$, while Izzeldin et al. (2019) set p and q parameters to 0, to estimate only the long-memory effect.

In this thesis we decide to follow the methodology of Baillie et al. (2019), which is to ex-ante select the number of autoregressive (p) and moving average (q) parameters. Then, the fractional integration parameters d, p and q can be estimated following the maximum likelihood method in a more unbiased manner.

We estimate $p \cdot q$ models with $p = \{1, \dots, 9\}$ and $q = \{1, \dots, 9\}$ and 3 additional models with AR or MA, or both set to 0: ARFIMA$(0, d, 0)$, ARFIMA$(0, d, 1)$ and ARFIMA$(1, d, 0)$. We therefore estimate $81 + 3 = 84$ models. For each model the Bayesian information criterion, which is more appropriate for time series model selection, is calculated by the following formula

$$\mathrm{BIC} = \mathrm{k} * \ln(\mathrm{n}) - 2\mathrm{LL} \tag{36}$$

where LL, is the log-likelihood of the function, k is the number of parameters estimated and n is the number of observations. The table with BIC values for all the estimated models can be found in Appendix 2.

The model with the lowest BIC is chosen as the best one and used for forecasting. The selected model was ARFIMA$(1, d, 1)$. It is extremely unlikely that the true model would have autoregressive and moving average parameters of an order higher than 9 (this can also be deducted from increasing BIC values as the number of parameters increases).

In our case the ARFIMA$(1, d, 1)$ is therefore

$$(1 - \varphi\mathrm{L})(1 - \mathrm{L})^{\mathrm{d}}(\ln \mathrm{RV_t} - \mu) = (1 - \theta\mathrm{L})\varepsilon_{\mathrm{t}} \tag{37}$$

As we have seen there are many estimation procedures discussed and compared in Bhardwaj and Swanson (2006), Doornik and Ooms (2004) and Reisen et al. (2001) among others. However, since we have decided to follow Baillie et al. (2019) by estimating a variety of possible ARFIMA specifications by using the maximum likelihood estimation and then select the best model using BIC, we will describe only the MLE method in more detail.

3.1.3 Maximum likelihood estimation

To generally define the maximum likelihood estimation, we start with random variables $(Y_1, Y_2, \dots Y_n)$ with a density function $f(\cdot\,|\theta)$, where θ is a vector of parameters. Let y_i be the observed value of Y_i. The likelihood function of θ given the vector of observed values is then $L = \prod_{i=1}^{n} f(y_i|\theta)$. We often work with the log-likelihood function, which is simply the logarithm of L, because it is easier to work with sums than products. Finally, the value of θ that maximises the (log) likelihood function is the maximum likelihood estimator (Amemiya, 1994).

To derive the maximum likelihood estimator for ARFIMA we start with an ARMA process. The autocovariances for an ARMA process are given as $\gamma(\mathrm{i}) = \mathrm{E}[(\mathrm{y_t} - \mu)(\mathrm{y_{t-i}} - \mu)]$. The

autocorrelations are then $\rho(i) = \gamma(i)/\gamma(0)$, where $\gamma(0)$ is simply the variance of the process. The covariance matrix V for the joint distribution of $y = (y_1, \ldots\ y_T)'$ is given by (Doornik & Ooms, 2003)

$$\mathbf{V} = \begin{bmatrix} \gamma_0 & \gamma_1 & \gamma_2 & \cdots & \gamma_{T-1} \\ \gamma_1 & \gamma_0 & \gamma_1 & \ddots & \vdots \\ \gamma_2 & \gamma_1 & \gamma_0 & \ddots & \gamma_2 \\ \vdots & \vdots & \vdots & \ddots & \gamma_1 \\ \gamma_{T-1} & \cdots & \gamma_2 & \gamma_1 & \gamma_0 \end{bmatrix} \tag{38}$$

The V matrix has a Toeplitz structure, where each descending left-to-right diagonal is constant. Many covariance-stationary processes involve the application of Toeplitz matrices (Gray, 2006).

Under normality $y \sim N(\mu, V)$. The log-likelihood for the ARFIMA(p, d, q) model is then

$$\begin{aligned} &\log L(d, \varphi, \theta, \beta, \sigma_\varepsilon^2) \\ &\quad = -\frac{1}{2}[T \cdot \log(2\pi) + \log(\det(V)) + (y - X\beta)'V^{-1}(y - X\beta)] \end{aligned} \tag{39}$$

This log-likelihood function can be maximised with respect to parameters $d, \varphi, \theta, \beta$, and σ_ε^2 by numerical optimization. The maximum likelihood estimates are those parameter values, which maximise the function. The key issue with function evaluation is the computation of the autocovariance matrix (Doornik & Ooms, 2003).

Assuming that roots of the AR polynomial are distinct, the autocovariance function for an ARFIMA(p, d, q) model is given by

$$\gamma_i = \sigma^2 \sum_{j=-q}^{q} \sum_{k=1}^{p} \psi(j)\xi_k C(d, p + j - i, \rho_k) \tag{40}$$

where

$$\psi(j) = \sum_{l=\max(0,j)}^{\min(q,q+j)} \theta_l\, \theta_{l-j}, \tag{41}$$

$$\xi_k = \left[\rho_k \prod_{i=1}^{p} (1 - \rho_i \rho_k) \prod_{m \neq k} \rho_k - \rho_m \right]^{-1} \tag{42}$$

and

$$C(d, i, \rho) = \frac{\gamma_i^{\,*}}{\sigma^2}[\rho^{2p}\beta(i) + \beta(-i) - 1] \tag{43}$$

where $\beta(i) = F(d + i, 1, 1 - d + i, \rho)$ and $F(a, b, c, x)$ is a Gaussian hypergeometric function

$$F(a,b,c,x) = \sum_{n=0}^{\infty} \frac{(a)_n (b)_n}{(c)_n} \frac{x^n}{n!}, \quad \text{where} \quad (z)_n = \begin{cases} 1 & , n = 0 \\ z(z+1)(z+n-1), & n > 0 \end{cases} \tag{44}$$

(Palma, 2007), and $\gamma_i{}^*$ is the autocovariance of an ARFIMA$(0, d, 0)$ process given by

$$\gamma_i{}^* = \sigma^2 \frac{\Gamma(1-2d)}{\Gamma(1-d)\Gamma(d)} \frac{\Gamma(i+d)}{\Gamma(1+i-d)} \tag{45}$$

where $\Gamma(\cdot)$ is a gamma function, which is defined for all complex numbers, except negative integers, by $\Gamma(x) = \lim_{k\to\infty} \frac{k!k^{x-1}}{(x)_k}$, where $(x)_k$ is a shifted factorial as defined in equation (44) (Andrews, Askey & Roy, 1999; Doornik & Ooms, 2003).

To compute the determinant and the inverse of the covariance matrix the Durbin-Levinson algorithm is often used, although Cholesky decomposition can also be applied (Doornik & Ooms, 2003; Palma, 2007).

3.2 Heterogeneous Autoregressive (HAR) model

The heterogeneous autoregressive (HAR) model was developed as an alternative to fractionally integrated models for modelling realized variance. The advantage of the HAR model is that it can be estimated by a simple ordinary least squares method. Moreover, HAR is able to generate long memory behaviour displayed by realized measures using very few parameters (Christoffersen, 2012).

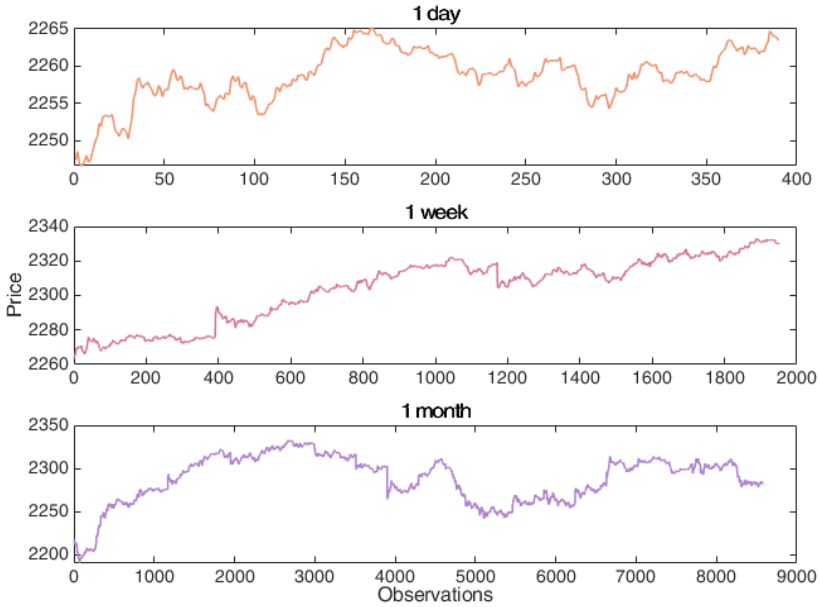
3.2.1 Motivation

The motivation behind the HAR model is the heterogeneous market hypothesis (HMH) by Müller et al., (1995), first presented already in 1993, which takes into account the heterogeneity across traders.

The heterogeneous market hypothesis is based on fractal theory, so a time series is analysed on different scales and with different levels of resolution. It suggests that different agents in the market have different time horizons and dealing frequencies and therefore respond differently to past events and news arriving to the market. The HMH is validated by the success of trading strategies, which account for various market participants with different dealing frequencies and risk propensity. In contrast to standard time series analysis dealing with regularly spaced observations, in this context intrinsic time is used (Müller et al., 1995).

Fractals display similar patterns at many different scales. This property is called self-similarity and has been observed in many phenomena. The heterogeneous markets hypothesis suggests that the market itself is fractal, since observing prices over one day, one week or over one year shows a similar pattern (Figure 4).

Figure 4: Self-similarity of the price process



Source: Own work.

Another crucial empirical finding by Müller et al. (1995), which characterises the HMH is that precisely these heterogeneous market agents generate volatility. If the market was homogeneous, the speed of convergence of the price to its true value would be directly proportional to the number of agents in the market. Volatility and market presence would be negatively correlated. However, empirically the opposite thing is observed. The two measures are positively correlated, which suggests that the markets are indeed heterogeneous. Due to many different participants, each with its own target price and carrying out transactions in different market situations, the volatility is higher.

Another theory based on the idea that the volatility process can be decomposed into multiple components has been developed by Andersen and Bollerslev (1997). The Mixture of Distribution Hypothesis focuses on the heterogeneous nature of information arrivals.

Hence, the heterogeneity in the financial markets can arise from differences in agents' income or risk profiles, institutional constraints, geographical locations or different time horizons. Corsi (2009) based the HAR model precisely on the heterogeneity of temporal horizons.

There are many types of financial market participants with different aims in the market and therefore, with different trading frequencies and information processing. At one end of the spectrum we have the daily traders: market makers, dealers and active speculators. Their trading horizon is usually only a couple of minutes or at most some hours. On the opposite side we have institutional investors like funds or insurance companies, which rebalance their portfolios for possibly larger amounts but much less frequently as their trading horizon can

span one or even more months. These participants can be roughly divided into three groups: short-horizon (1 day), medium-horizon (1 week) and long-horizon (1 month) (Corsi, 2009).

The motivation for the model is that these groups of market participants "*perceive, react to, and cause different types of volatility components"* (Corsi, 2009, p.178). This type of categorization has been quite neglected in econometric modelling although it has a compelling economic interpretation and can be deduced by simply observing the market.

Another important foundation for the model is the asymmetric propagation of volatility. Some studies (Müller, 1997; Arneodo et al., 1998; Lynch & Zumbach, 2003, Gençay & Selçuk, 2004), have explored the interplay between these different volatility components and the researchers came to the conclusion that the influence of long-horizon volatility has a much stronger influence on short-horizon volatility than vice-versa. The economic interpretation of this volatility behaviour is that long-run volatility is important for daily traders as a guidance of the future trends and risks. Contrarily, daily volatility does not influence the long-term trading strategies of institutional investors to a large extent (Corsi, 2009).

3.2.2 Derivation and estimation

Corsi (2009) proposed a simple model with three volatility components, which is based on the observation that heterogeneity in the market structure leads to a volatility cascade. The model tries to replicate the stylised characteristics of financial data. The last step in model derivation is the use of realized volatility measures, which make volatility observable.

To develop the model, Corsi (2009) first defined latent partial volatilities ${\tilde{\sigma}_t}^{(d)}$, ${\tilde{\sigma}_t}^{(w)}$ and ${\tilde{\sigma}_t}^{(m)}$, which represent the volatility generated by daily, weekly and monthly traders respectively.

The return process is determined by the daily volatility component, which is the highest-frequency component in the model. By defining the daily integrated volatility ${\sigma_t}^{(d)} = {\tilde{\sigma}_t}^{(d)}$, the return process is

$$r_t = {\sigma_t}^{(d)} z_t \tag{46}$$

where z_t is a white noise (Corsi, 2009).

At each level of the cascade the latent partial volatility process is modelled as a function of the past realized volatility and the expectation of the longer horizon partial volatilities because of the asymmetric propagation. The model for daily partial volatility is therefore

$${\tilde{\sigma}_{t+1d}}^{(d)} = c^{(d)} + \varphi^{(d)} {RV_t}^{(d)} + \gamma^{(d)} E_t\left[{\tilde{\sigma}_{t+1w}}^{(w)}\right] + {\tilde{e}_{t+1d}}^{(d)} \tag{47}$$

where ${RV_t}^{(d)}$ is the daily ex post realized volatility and ${\tilde{e}_{t+1d}}^{(d)}$ is the volatility innovation, which is serially and cross-sectionally independent with zero mean (Corsi, 2009).

Weekly partial volatility process has the following structure

$$\tilde{\sigma}_{t+1w}{}^{(w)} = c^{(w)} + \varphi^{(w)} RV_t{}^{(w)} + \gamma^{(w)} E_t\left[\tilde{\sigma}_{t+1m}{}^{(m)}\right] + \tilde{e}_{t+1w}{}^{(w)} \tag{48}$$

Whilst for the longest monthly horizon only the 'AR(1)' structure remains

$$\tilde{\sigma}_{t+1m}{}^{(m)} = c^{(m)} + \varphi^{(m)} RV_t{}^{(m)} + \tilde{e}_{t+1m}{}^{(m)} \tag{49}$$

Since the quantities on the right hand side of the equations are not the lags of latent partial volatilities, these model cannot be said to have a proper AR(1) structure. However, as the realized volatilities are close proxies for the corresponding partial latent volatilities, the structure is very similar to an AR(1) process (Corsi, 2009).

Each volatility component therefore has a factor, which corresponds to the market's expectation of realized volatility in the next period based on the current level of volatility and a second factor, which represents the expectation of the long-term volatility (Corsi, 2009). We recall that due to the asymmetric propagation of volatility, long-term volatility affects the short-term volatility level.

By substituting recursively the above formulae and using $\sigma_t{}^{(d)} = \tilde{\sigma}_t{}^{(d)}$ the model becomes a three-factor volatility model with past realized volatilities over different horizons as factors

$$\sigma_{t+1d}{}^{(d)} = c + \beta^{(d)} RV_t{}^{(d)} + \beta^{w} RV_t{}^{(w)} + \ \beta^{(m)} RV_t{}^{(m)} + \tilde{e}_{t+1d}{}^{(d)} \tag{50}$$

By recalling that

$$\sigma_{t+1d}{}^{(d)} = RV_{t+1d}{}^{(d)} + e_{t+1d}{}^{(d)} \tag{51}$$

and combining equtions (50) and (51) the HAR model is

$$RV_{t+1d}{}^{(d)} = c + \beta^{(d)} RV_t{}^{(d)} + \beta^{w} RV_t{}^{(w)} + \ \beta^{(m)} RV_t{}^{(m)} + e_{t+1d} \tag{52}$$

where (Corsi, 2009)

$$e_{t+1d} = \tilde{e}_{t+1d}{}^{(d)} - e_{t+1d}{}^{(d)} \tag{53}$$

To summarise, the HAR model is based on an additive cascade of partial volatilities from high frequencies to low frequencies where each level is similar to an AR(1) process.

Corsi (2009) developed the model for the square-root of realized variance (realized volatility) but suggested that the model can be analogously written for the variance or logarithmic variance. Since the natural logarithm of realized variance has the distribution that is closely similar to the normal distribution (based on our data we calculate skewness equal to 0.7 and kurtosis equal to 3.9) we will forecast using the natural logarithm of realized variance ($lnRV$), which is used in many studies.

By replacing realized variance with the natural logarithm of realized variance ($lnRV$) we obtain the HAR-$lnRV$ model

$$lnRV_t^{(d)} = c + \beta^{(d)} lnRV_{t-1}^{(d)} + \beta^{(w)} lnRV_{t-1}{}^{(w)} + \beta^{(m)} lnRV_{t-1}{}^{(m)} + \varepsilon_t \tag{54}$$

where weekly and monthly log realized variances are defined as

$$lnRV_{t-1}{}^{(w)} = \frac{1}{5}\sum_{i=1}^{5} lnRV_{t-\mathrm{i}}^{(d)} \tag{55}$$

$$lnRV_{t-1}{}^{(m)} = \frac{1}{22}\sum_{i=1}^{22} lnRV_{t-\mathrm{i}}^{(d)} \tag{56}$$

The HAR model uses realized variance aggregated over a week and a month. To allow for a direct comparison between these quantities, which are aggregated over differently long horizons, the multi-period variances must be normalised. Weekly RV is therefore just an arithmetic average of 5 daily realized variances, while monthly RV is the average of the last 22 daily realized variances (Corsi, 2009).

Based on the theory of realized variance, all the terms in equations (52) and (54) are considered observable, therefore we can easily estimate all the β parameters by the ordinary least squares (OLS) method. To account for potential autocorrelation and heteroskedasticity in error terms, the Newey-West covariance correction can be applied (Corsi, 2009).

Following the martingale theory, the optimal forecast for the $t+1$ period is the expected value of the notion at time t. The one-day ahead forecast is then simply

$$\widehat{lnRV_{t+1}^{(d)}} = \widehat{\beta_0} + \widehat{\beta^{(d)}} lnRV_t^{(d)} + \widehat{\beta^{(w)}} lnRV_t{}^{(w)} + \widehat{\beta^{(m)}} lnRV_t{}^{(m)} \tag{57}$$

Many studies have extended the HAR model in several ways: by assuming various distribution for the innovations, by assuming GARCH effects in innovations, Corsi and Renò (2012) have added estimated price jumps by using (threshold) bipower variation (BV) and added leverage. The leverage effect is a phenomenon studied by Black (1976), Pagan and Schwert (1990) and Engle and Ng (1993) among others, who found a negative correlation between lagged excess returns and volatility. Excess returns are also defined as news, where bad news represent lower excess returns than expected and good news imply higher excess returns than expected. The leverage effect therefore implies that a large negative return will increase future volatility by more than a positive return of the same magnitude would (Martens et al., 2009).

More recently, Audrino et al. (2018) proposed a flexible HAR with flexible lag structure and not the [1,5,22] lag selection, although Craioveanu and Hillebrand (2014) show through

heavy computational techniques that other flexible versions do not significantly outperform the HAR(1,5,22) model, which is therefore acceptable to use.

4 DATA AND FORECASTING PROCEDURE

4.1 Data

When analysing financial markets, high-frequency (HF) data should be the primary source for research. The reason is that many agents in financial markets use high-frequency or even tick data when making decisions. Nonetheless, many studies in financial literature are still based on daily data. This is because low-frequency data is easier to analyse. The main downside when dealing with high-frequency observations is precisely the difficulty of gathering, storing and processing HF data (Gençay, Dacorogna, Müller, Pictet & Olsen, 2001). Datasets with tick or high-frequency prices often include erroneously recorded observations, missing data and irregularly spaced data points (Zivot, 2005).

Our empirical analysis was carried out using the NASDAQ Composite index high-frequency prices from January 3, 2006 to May 8, 2020. Data was acquired from FirstRateData. Days with late market opening or early market closure (mainly around the Fourth of July, Christmas and New Year) were removed following Oomen (2006) as is standard practice when cleaning high-frequency data (Zivot, 2005).

There were altogether 49 days removed following this procedure, which represented less than 1.5 percent of the whole dataset. We were left with 3,562 days for further analysis. Intra-day price data was not recorded at tick level, but at 1-minute intervals, which guaranteed us 390 observations for each full trading day, starting at 9:30 and finishing at 16:00. Our analysis was therefore based on 1,389,180 data points from 3,562 trading days.

The most widely used stock index research data is the Standard and Poor's 500 index (S&P 500), which includes 500 leading corporations from all the 11 market sectors[5]. In this thesis NASDAQ Composite (COMP) index data is used. Together with the S&P 500 and the Dow Jones Industrial Average (DJIA), COMP is among the three most followed stock indices in the United States.

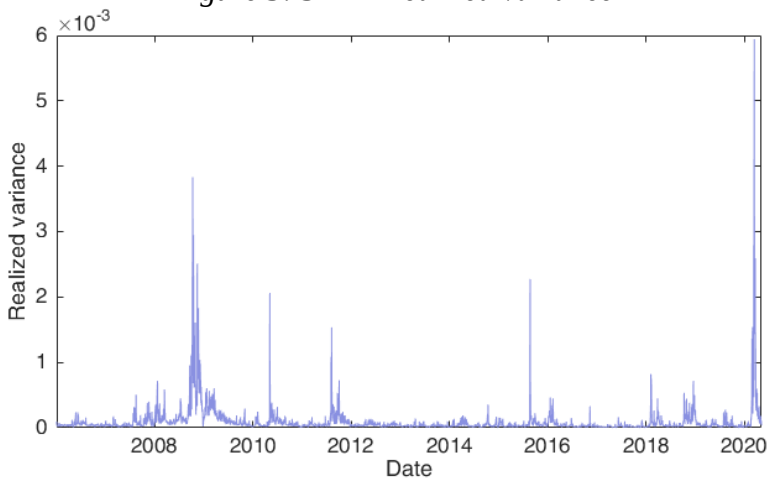
The NASDAQ Composite is an index traded on the New York Stock Exchange (NYSE), which included 2718 securities as of June 30, 2020. Unlike the S&P 500, the COMP index is much more specialised in certain market sectors. Companies from the information technology sector account for almost 50% of the market capitalisation of the index (NASDAQ, 2020). COMP is generally more volatile than S&P 500 since it is not as diversified. It is also specialised in the technology sector, which is among the most volatile

[5] The Global Industry Classification Standard broad sectors are Consumer Discretionary, Consumer Staples, Energy, Financials, Health Care, Industrials, Information Technology, Materials, Real Estate, Telecommunication Services and Utilities.

sectors. To our knowledge no study comparing long-memory models has been based on data from this index.

We constructed realized variance by first computing 1-minute log returns as a difference between log prices in time $t+1$ and t. The 1-minute log returns were then summed to obtain the 2-, 3-, 5- and 10-minute returns following the additive property defined in equation (2). To construct the daily n-minute RV we summed the squared n-minute log returns for each day.

Figure 5: 5-min Realized variance



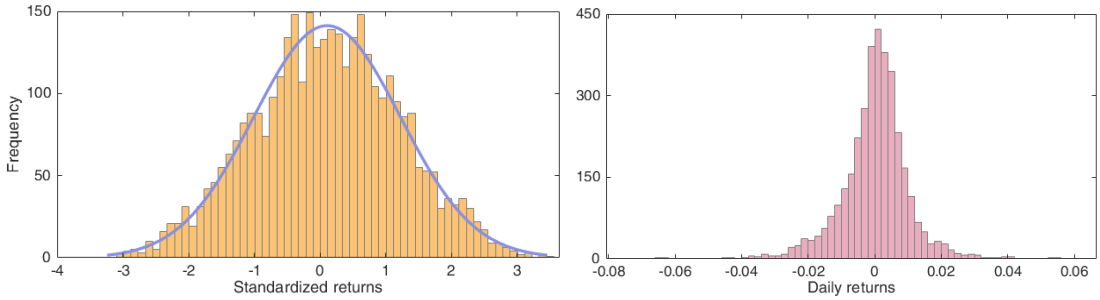
Source: Own work.

A nice property of realized variance is that if we use it to standardise the returns, they will be approximately normally distributed. This is because returns standardised by integrated volatility are normally distributed and realized variance converges to integrated variance (Andersen et al., 2000)

$$\frac{r_t}{\sqrt{RV_t}} \sim N(0,1) \tag{58}$$

This can be observed in Figure 6, where the distribution of daily returns on the right is visibly leptokurtic (exhibits fat tails), while the standard normal density function fits the standardized returns on the left well.

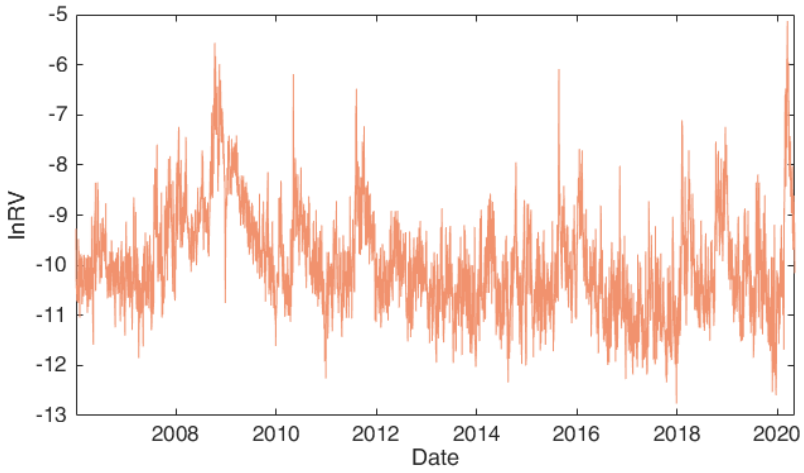
Figure 6: Distributions of standardized and non-standardized daily returns



Source: Own work.

We decide to model and forecast the natural logarithm of RV following many other studies, first and foremost Andersen, Bollerslev, Diebold & Ebens (2001). These authors analysed the distribution of realized volatility for stock returns and suggested the use of $lnRV$ instead of RV precisely because realized variance is log-normally distributed.

Figure 7: 5-min lnRV



Source: Own work.

Based on our data, the distribution of $lnRV_t$ has kurtosis equal to 3.87 and skewness equal to 0.71 (Table 1). The normal distribution has these parameters equal to 3 and 0, respectively. In this sense, the distribution of $lnRV$ is much more similar to the normal distribution than that of RV_t, which has a kurtosis of 187 and skewness over 10.

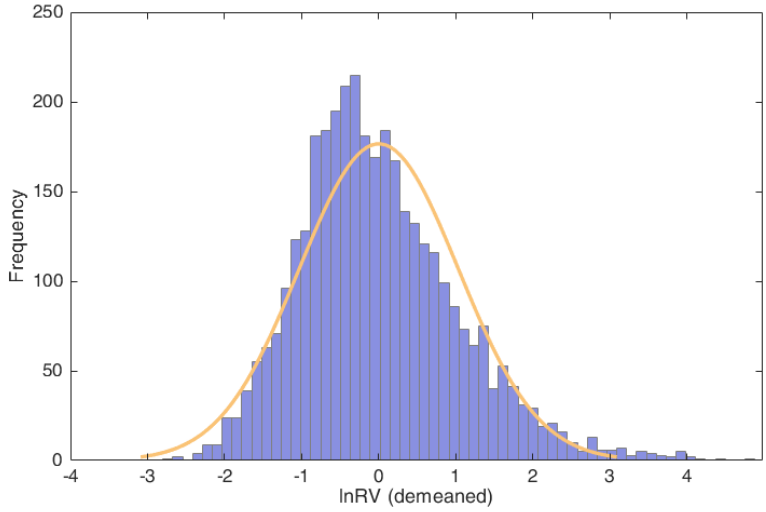
Table 1: Summary statistics

	1-min returns	daily returns r_t	5-min RV_t	5-min $lnRV_t$
Mean	2.3385e-07	9.120e-05	9.494e-05	-9.9596
Std. deviation	4.4857e-04	0.0107	2.2716e-04	1.0299
Skewness	-1.6098	-0.5281	10.8575	0.7054
Kurtosis	330.46	9.124	187.26	3.8684
Min	-0.060756	-0.0733	2.8757e-06	-12.7592
Max	0.0344	0.0589	0.0059	-5.1256
Observations	1,389,180	3,562	3,562	3,562

Source: Own work.

In Figure 8 we compare the distribution of demeaned $lnRV$ to a standard normal distribution. We observe a relatively good fit, although leptokurtosis and positive skewness are visible. Indeed, the Jarque – Bera test for normality rejects the hypothesis that $lnRV_t$ is normally distributed ($JB = 407.3; p - value = 0.001$).

Figure 8: Distribution of demeaned lnRV



Source: Own work.

To test the stationary of the series (the mean and the variance of the process are constant over time) the Augmented Dickey Fuller (ADF) test is used. The null hypothesis of the ADF test is that the series is non-stationary, i.e. there is a unit root. Therefore, by rejecting the null hypothesis we could conclude that the series is stationary.

We cannot reject the null for the $lnRV$ series, although the p-value is 0.07, which is close to the standard significance level of 0.05. However, Diebold and Rudebusch (1991) show that if the true data generating process is fractionally integrated, the tests proposed by Dickey and Fuller (1979, 1981) can often lead to the incorrect conclusion about the presence of a unit root.

Similarly, Baillie (1996) argues that the ADF and Phillips-Perron unit root tests distinguish rather poorly between the null hypothesis of a unit root ($I(1)$, non-stationary series) and the alternative of a fractionally integrated $I(d)$ series.

4.2 Forecasting procedure and forecast evaluation

To evaluate the hypotheses about sensitivity of the two models to changes in forecasting horizon, market conditions and sampling frequencies, 10 forecasting exercises are performed for each of the models. We first generate daily, weekly, and monthly forecasts for the 5-minute $lnRV$ series, and then we split the series into three subsamples representing different market conditions. Finally, we compare daily forecasts based on 5-min $lnRV$ to the forecasts based on 1-, 2-, 3- and 10-min $lnRV$ series.

We adopt the pseudo out-of-sample (OOS) forecasting method, which involves estimating a model on a subsample of data and then forecasting on a reserved sample. In contrast to in-sample forecasting, the OOS method strictly separates the observations into two parts. A smaller portion is reserved for generating forecasts, while the majority of the sample is used

for model estimation. This method is called *pseudo* OOS because reserved observations are not genuinely out-of-sample (Stock & Watson, 2020).

The sample can be split into the estimation and forecasting parts in a variety of proportions, although reserving 20 percent of data for forecast evaluation and estimating the model on the initial 80 percent of observations is the most common approach (Hyndman & Athanasopoulos, 2018).

We adopt a rolling approach for forecasts, which keeps the estimation window fixed, in contrast to the recursive approach, which uses all available data up to time t to make a forecast for the following period. We therefore use the first 80 percent of observations ($R = 2850$) from the full sample ($T = 3562$) to estimate model parameters in the first step and make a forecast for the $R + 1$ period. By keeping the rolling window size R fixed, we re-estimate the parameters in the second step based on observations from 2 to $R + 1$, while the forecast is generated for the $R + 2$ period. This process is repeated k -times ($k = 712$) so that $R + k = T$.

We adopt a combination of iterated and direct methods for multi-period ahead forecasting, so that at every step the forecasts from previous steps are included and the model is re-estimated. According to Sorjamaa & Lendasse (2006), who proposed this combined method, it outperforms iterated or direct approaches.

Therefore, when forecasting one-week (5 days) or one-month (22 days) ahead the parameters were initially estimated using the first R observations and a forecast was generated for the $R + 1$ period. In the second step the one-period ahead forecast was treated as a data point, so the estimation window ranged from 2 to R from the initial sample plus the forecast obtained in the first step. This is repeated five times for weekly forecast and 22 times for the monthly forecast. So, for each $t + n$ forecast we used realised values only up to time t, while the other $(n - 1)$ values needed for the forecast at time $t + n$ were one-period ahead forecasts from each iteration.

At every iteration the forecast $\widehat{lnRV_t}$ is compared to the actual realised value. To evaluate forecast precision, the root mean squared error (RMSE) measure was calculated

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(\widehat{lnRV_t} - lnRV_t)^2} \tag{59}$$

RMSE is one of the most widely used error measures beside mean absolute error (MAE). The size of the metric depends on the order of magnitude of observations in the series. Since we are comparing two models based on the same series, we are only interested in determining, which model has a smaller RMSE value.

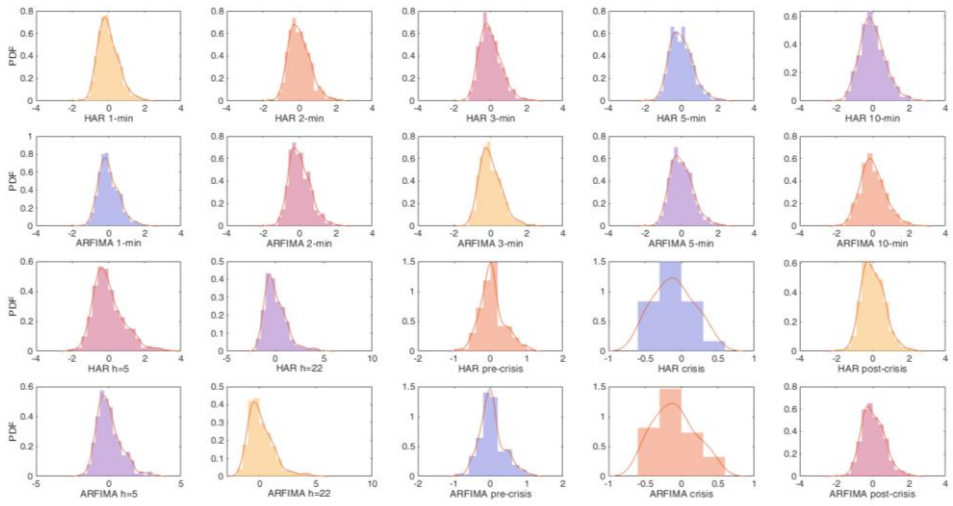
Willmott and Matsuura (2005) suggest that MAE is a more natural measure of the average error than RMSE. Since the latter is a measure based on sum of squares, it depends on the magnitude of errors, the distribution of squared errors and the sample size. As a result, RMSE may be a misleading measure of average error and therefore should not be used to assess average model performance according to the authors.

On the other hand, Chai and Draxler (2014) claim that RMSE is more appropriate when errors are normally distributed. Since squared error measures are more dominated by outliers, RMSE may actually be preferred to MAE. In our case large deviations in forecast errors are particularly undesired. Volatility forecasting is especially important in risk management and asset valuation, therefore large deviations of forecasts from realized values can cause great distress. By underestimating volatility, the exposure to risk may be larger than thought. On the other hand, overestimating volatility may lead to incorrect or suboptimal allocation of resources.

Chai and Draxler (2014) show that if sample size is larger than 100 the RMSE measure can very accurately reconstruct the error distribution. They also prove that Willmott and Matsuura (2005) wrongly claimed that the RMSE metric does not satisfy the triangle inequality, which is required for distance function metrics. Bucci (2017) also suggests that measures based on square errors like mean square error (MSE) and RMSE are both robust measures if the models are used on the same time series.

RMSE is especially robust if the errors are normally distributed. If that is not the case, they need to be distributed similarly to obtain a credible comparison between different models. In Figure 9 we can observe the distribution of errors from all the twenty forecasting exercises. The 5- and 10-minute $lnRV$ forecast errors are close to being normally distributed (the results of the Jarque – Bera tests for the normality of forecast errors are found in Appendix 3).

Figure 9: Distribution of forecasting errors



Source: Own work.

All other error distributions are quite leptokurtic (fat tails), while error distributions for long horizon forecasts display positive skewness. However, the distributions do not change significantly for each forecasting exercise; therefore, the RMSE measure will give us a relatively good indication of the superior model.

We should be careful when comparing the RMSE values for forecasts in different market conditions. In that case the series was split into three parts, which are not of the same length. RMSE tends to be increasingly larger than the average error as sample size increases so RMSE can be compared only for the two models in a certain market condition and not across market conditions (for example, pre-crisis to crisis period).

To assess whether the two models produce forecasts, which differ significantly, the Diebold – Mariano (DM) test was implemented. The test is widely used in the relevant literature. The DM test is very versatile as it allows for potentially non-Gaussian, nonzero mean, serial or cross-sectional correlated forecast errors (Diebold & Mariano, 1995).

We denote forecast errors for each of our models the same as for the RMSE calculation

$$e_t = \left(\widehat{lnRV_t} - lnRV_t\right) \tag{60}$$

The null hypothesis of the Diebold-Mariano test states that there is no difference in accuracy of competing forecasts. By rejecting the null hypothesis, we can therefore assume that the predictive accuracy of the two models differs significantly. By denoting a loss function $g(\cdot)$, the null hypothesis states

$$H_0: E\big(g(e_{it})\big) = E\left(g\big(e_{jt}\big)\right) \Leftrightarrow E(d_t) = 0, \quad \forall t. \tag{61}$$

where the loss differential d_t is defined as

$$d_t \equiv g(e_{it}) - g\big(e_{jt}\big) \tag{62}$$

where i and j represent each of the two models (Diebold & Mariano, 1995).

Under not very strict conditions the Diebold-Mariano test statistic is asymptotically normally distributed

$$DM = \sqrt{T}\frac{\bar{d}}{\sqrt{\omega}} \sim N(0,1) \tag{63}$$

where the average error difference is calculated as

$$\bar{d} = \frac{1}{T}\sum_{t=1}^{T} e_{it} - e_{jt} \tag{64}$$

and

$$\omega = \lim_{t\to\infty} var(\sqrt{T}\,\bar{d}) \tag{65}$$

is the asymptotic variance, which is usually approximated by the sample variance. Since we use the sample variance, the test statistic is not exactly standard normally distributed, but it is nonetheless standard practice to reject the null hypothesis if the test statistic falls outside the $(-1.96, 1.96)$ interval.

To summarise, for each of the ten forecasting exercises we compute RMSE value for both the HAR and ARFIMA model. We then use the Diebold-Mariano test. If we reject the null hypothesis, we assume that the models differ in accuracy and we accept the model with the lower RMSE value as superior.

5 EMPIRICAL RESULTS

In this section we present the estimation and forecast accuracy results. In the first subsection we compare the forecasting performance of the two models over daily, weekly, and monthly horizons. We then present the impact of different regimes or market conditions on predictive accuracy. Finally, we compare the forecasting accuracy based on different sampling schemes for realized variance.

5.1 Long horizons

We begin by estimating the HAR parameters for the full 5-min $lnRV$ sample to check if the parameters are in line with expectations and most importantly if they are statistically significant. We estimate 4 parameters, the intercept and the three betas, which are related to the daily, weekly, and monthly realized variation.

Using the non-logarithmic HAR-RV model Corsi (2009) obtains the following estimates for the S&P 500 index with observations from January 1990 to July 2007

Table 2: HAR-RV estimates by Corsi (2009)

	constant	$\beta^{(d)}$	$\beta^{(w)}$	$\beta^{(m)}$
estimate	-0.781	0.372	0.343	0.224
(t – statistic)	(4.065)	(9.858)	(7.263)	(6.467)

Source: Corsi (2009).

By assuming that the realized variance aggregated over daily, weekly and monthly horizon is a very good approximation of the true volatility generated by these distinct market components, we can obtain an estimate of how much each market component contributes to the market activity. For a series of stock index prices, Corsi (2009) suggests that the contribution of market components decreases for longer horizons. In other words, we expect the daily beta to have the highest and the monthly beta to have the lowest value of the three. Our estimation results are presented in Table 3.

Table 3: HAR-lnRV estimates

coefficient	estimate	standard error	t – statistic	p –value
constant	-0.5916	0.1154	-5.127	$3.101 \cdot 10^{-7}$
$\beta^{(d)}$	0.4751	0.0237	20.046	$8.906 \cdot 10^{-85}$
$\beta^{(w)}$	0.3321	0.0332	10.003	$2.991 \cdot 10^{-23}$
$\beta^{(m)}$	0.1333	0.0253	5.2688	$1.455 \cdot 10^{-7}$

Source: Own work.

The difference between the daily and monthly beta for our sample is larger than the one by Corsi (2009), which is not that important since the sample was different. However, the pattern is similar, since the daily beta has the highest value, which decreases as time horizon lengthens.

Standard errors are computed using the Newey-West estimator. From the t – statistic and the corresponding p – value it is evident that the beta coefficients are highly statistically significant. The model fits the data well since the F – statistic of the full model versus the model with just the intercept is equal to $3.06 \cdot 10^3$ and the adjusted R-square is equal to 0.721.

We then estimate the ARFIMA parameters for the full 5-min $lnRV$ sample in Stata with the following command: **`arfima lnrv, ar(1) ma(1)`**. Based on BIC values for the 84 models (combinations of AR and MA parameters ranging from 1 to 9 and three additional models with AR or MA, or both parameters set to 0), the ARFIMA$(1, d, 1)$ was chosen as the best-fitting model (see Appendix 2). The maximum likelihood estimation results are presented in Table 4.

Table 4: ARFIMA$(1, d, 1)$ MLE results

coefficient	estimate	standard error	t – statistic	p –value
constant	-9.909	1.263	-7.846	$5.637 \cdot 10^{-15}$
φ_1	0.892	0.0454	19.664	$8.365 \cdot 10^{-82}$
θ_1	-0.838	0.0663	-12.637	$7.690 \cdot 10^{-36}$
d	0.462	0.0323	14.298	$3.922 \cdot 10^{-45}$
σ_ε^2	0.294	0.007	42.199	$7.24 \cdot 10^{-316}$

Source: Own work.

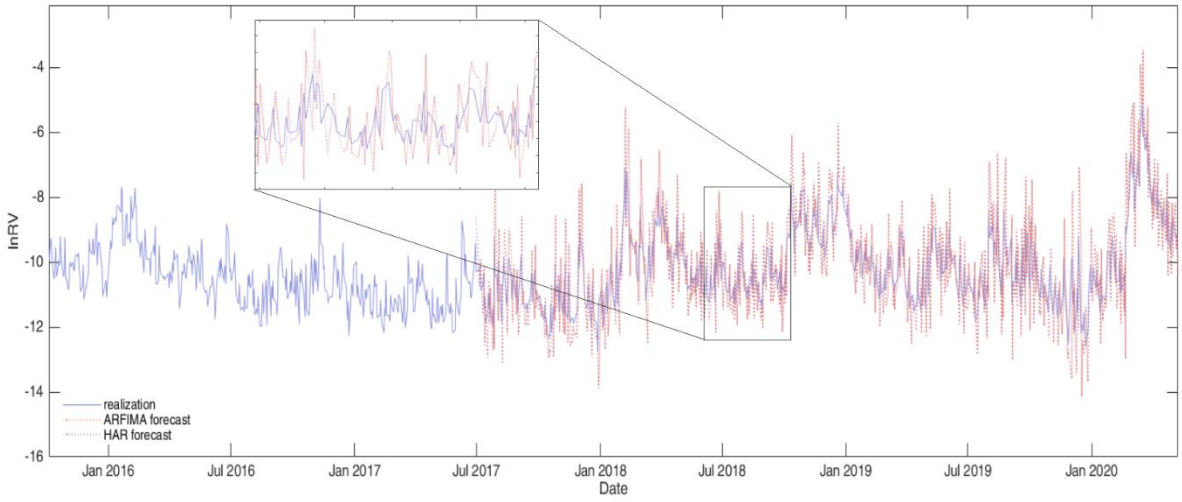
The F – statistic for the ARFIMA model is $3.75 \cdot 10^3$, with p – value zero, so we conclude that also the ARFIMA$(1, d, 1)$ model fits the data well. The log-likelihood value of the model, which was used for computing the information criterion, is -2877.605.

Once again, all the parameters are highly statistically significant, with p-values essentially equal to 0. We therefore generate forecasts using the ARFIMA$(1, d, 1)$ model and assume that the series is fractionally integrated of order 0.462. This degree of fractional integration d is in line with estimates by Andersen et al. (2003). For three distinct realized volatility series they estimate the parameter to be equal to 0.39, 0.41 and 0.43. Most existing literature estimates the d parameter to be between 0.3 and 0.48 (Andersen & Benzoni, 2009).

After estimating the model parameters for the full sample and confirming that the models have a good fit, we implement one-day ahead rolling forecasting as described in subsection 4.2. The command used for obtaining the vectors of parameters of the rolling estimation in Stata is: **`rolling _b, window(R): arfima lnrv, ar(1) ma(1).`** The code used to generate forecasts in MATLAB for both HAR and ARFIMA models is presented in Appendix 4.

In the figure below we display a segment of realised observations and the corresponding daily forecasts generated by the two models. We can observe that the models seem to perform quite well and very similarly.

Figure 10: Observed realizations of lnRV and daily forecasts



Source: Own work.

Finally, weekly and monthly forecasts are generated. In Table 5 we compare the forecasting performance of the HAR and ARFIMA models for the different forecasting horizons. For one-day ahead forecasts, the ARFIMA model outperforms the HAR model, as the RMSE values are 0.6315 and 0.6342 respectively. On the other hand, for weekly and monthly horizons the HAR model is more accurate. By applying the Diebold-Mariano test to see if the models differ significantly, we observe that HAR significantly outperforms ARFIMA only based on the monthly forecasts.

Table 5: RMSE values for different horizons and DM test results

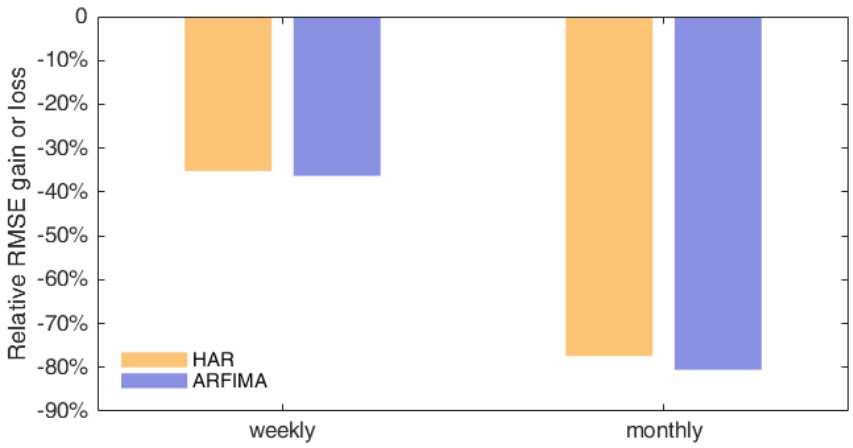
Horizon	RMSE		DM test statistic	p-value	Best performer
	HAR	ARFIMA			
Daily (h=1)	0.6342	0.6315	-1.33	0.18	ARFIMA
Weekly (h=5)	0.8575	0.8608	0.58	0.56	HAR
Monthly (h=22)	1.1253	1.1402	2.15	0.032	HAR**

** 5% significance level

Source: Own work.

The performance of the two models worsens as the forecasting horizon is extended (Figure 11). HAR proves to be less sensitive to longer forecasting horizons, as its accuracy drops by 35.2% for weekly horizons and 77.4% for monthly horizons, while ARFIMA weekly and monthly accuracy decreases by 36.3% and 80.6% respectively.

Figure 11: Relative RMSE loss at longer horizons



Source: Own work.

These results are in line with Izzeldin et al. (2019), who show that HAR forecasts are relatively less sensitive to the lengthening of the forecasting horizon, while the accuracy of ARFIMA forecasts worsens at long horizons. This is in contrast to Bhardwaj and Swanson (2006), who suggest that ARFIMA models are preferred at multi-step ahead forecasts. However, their pool of models does not include HAR and includes mainly various GARCH and SV model specifications. Our results for the daily horizon are in line with those by Degiannakis and Floros (2013). They find that ARFIMA provides more accurate forecasts of $lnRV$. However, they do not test for significance although the loss function values are very close for both models and may imply insignificant superiority (ARFIMA = 0.2727, HAR = 0.2732). Similarly, Hansen et al. (2005) do not find significant evidence that ARFIMA would outperform simpler models.

The existing literature is therefore quite divided on the question of predictive accuracy at different horizons. Our results suggest ARFIMA outperforms HAR at daily forecasts, but

not significantly, whereas HAR is more accurate on a weekly basis and significantly more accurate on a monthly basis.

The results confirm our hypothesis that the HAR model is less sensitive to the changes in forecasting horizon since the drop in predictive accuracy for weekly and monthly horizons compared to the daily one is smaller for HAR than ARFIMA.

5.2 Market conditions

The behaviour in the markets is cyclical with periods of growth and prosperity, when market volatility tends to be lower and periods of downturn, when prices move more frantically. By splitting our sample into periods, where market conditions were favourable or not, we can assess the performance of the two models in specific states of the market.

The iterative cumulative sum of squares (ICSS) algorithm could be used to determine the data points where the regime changes. It was developed by Inclan and Tiao (1994) to detect changes in the variance of the series. However, it was designed for normal and independent processes. Since financial data usually displays heavy tails and, in some cases, serial dependence, this algorithm is too sensitive and finds more than 30 structural breaks in our series of daily returns. There are other alternatives such as the nonparametric change-point algorithm by Ross (2013), which is more robust and does not detect some of the spurious breaks that ICSS does.

We decide to implement a MATLAB function 'findchangepts' to detect structural breaks in the variance of daily returns, which will represent a regime change. The function was set to find at most 3 change points in the standard deviation of the return series to obtain reasonably long periods for estimation and forecasting with the hope, that these periods would have an economic interpretation as well.

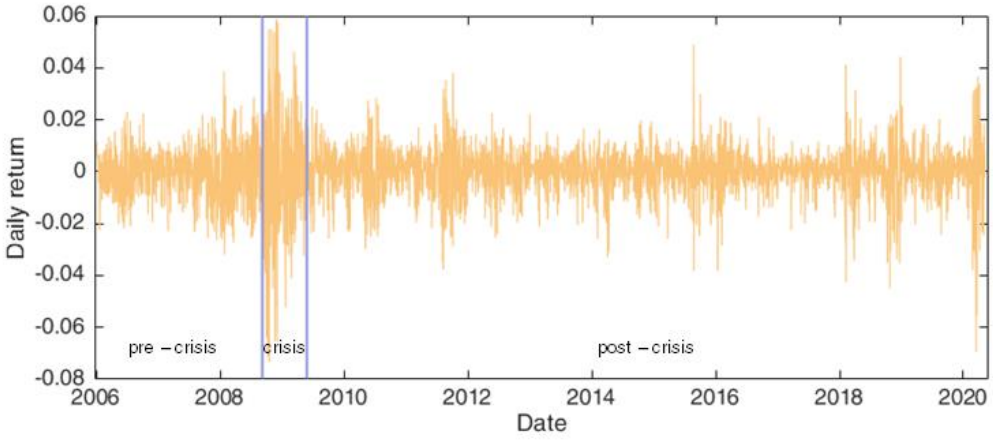
A changepoint is a point in time, when a statistical property, in our case the variation of a signal, changes suddenly. The function 'findchangepts' first divides the signal into two segments and computes the estimate of variance for each part. Then it measures the deviation from the estimate for each point within the segment. Finally, it sums up all the deviations to find the total residual error. This process is repeated for all the possible locations of a division point between segments. The division point becomes a change point where the total residual error attains a minimum (MathWorks, 2020).

We obtain three meaningful changepoints: September 2, 2008, May 27, 2009 and February 25, 2020. Since the last structural break divides only about the last 50 observations from the full sample, we decide not to split the series at that observation and keep only the first two structural breaks. We can think of the two cut-off points as the start and end of the most volatile part of the 2008 financial crisis. Therefore, we divide our sample into three regimes: pre-crisis (January 3, 2006 – September 2, 2008), crisis (September 3, 2008 – May 27, 2009) and post-crisis (May 28, 2008 – May 8, 2020), see Figure 12. The last changepoint on February 25, 2020, which we discard due to the inadequate number of observations for

estimation and forecast evaluation, would represent the start of the new downturn due to the COVID-19 outbreak.

Although, the change points were obtained with a MATLAB function somewhat arbitrarily by allowing for only three structural breaks, these changepoints separate periods with noticeably different volatility and have a compelling economic interpretation, so we continue with the estimation and forecasting for each of these subsamples.

Figure 12: Pre-crisis, crisis and post-crisis periods



Source: Own work.

RMSE values in Table 6 suggest that HAR performed better in the pre-crisis and crisis period, while ARFIMA is superior in the post-crisis period. By looking at the average log realized variance for the three regimes we notice that, while $lnRV$ is expectedly the highest during the crisis period ($\overline{lnRV}_{crisis} = -7.98$), it is higher in the pre-crisis than post-crisis period ($\overline{lnRV}_{pre-crisis} = -9.75$; $\overline{lnRV}_{post-crisis} = -10.14$). We could therefore cautiously suggest that HAR performs better in periods of relatively higher volatility, although another more probable explanation, which has to do with sample size, will be presented later.

Table 6: RMSE values for different regimes and DM test results

Regime	Weights	RMSE		DM test statistic	p-value	Best performer
		HAR	ARFIMA			
Pre-crisis	0.183	0.3458	0.3466	-0.506	0.61	HAR
Crisis	0.051	0.2912	0.2981	1.7609	0.078	HAR*
Post-crisis	0.766	0.6317	0.6290	-0.7974	0.43	ARFIMA
Weighted	-	0.5620	0.5605			ARFIMA

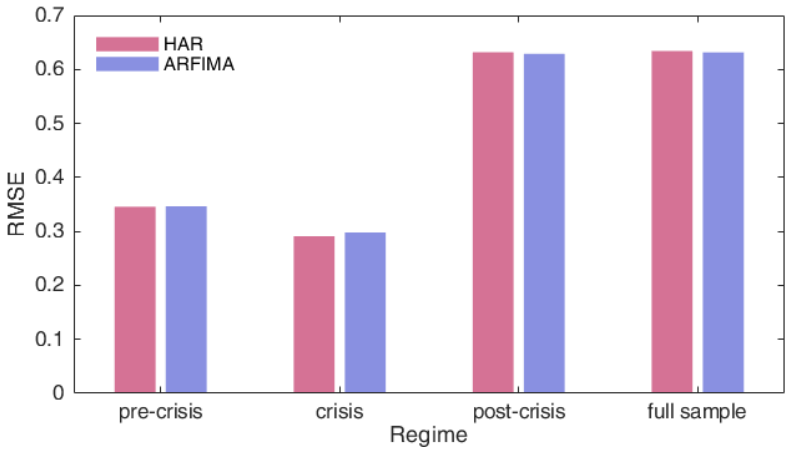
* 10% significance level

Source: Own work.

On the other hand, none of the Diebold – Mariano test statistics show significant difference between the models. The p – value for the crisis period is 0.078, which would suggest modest preference for the HAR model in periods of market turbulence.

The RMSE is a function of average error magnitude, the distribution of errors and the sample size. Since our three regime periods differ substantially in length, we can only compare the performance of the two models in the specific regime, while prediction accuracy cannot be compared across regimes.

Figure 10: RMSE values for different regimes



Source: Own work.

The post-crisis forecasting period ends on 8th May 2020 so it includes also the market downturn caused by the COVID-19 pandemic. Therefore, we repeat RMSE calculation for the subsample, where the last 52 forecasting errors corresponding to the period from 25th February 2020 onwards are discarded. The results do not change significantly. ARFIMA still produces a smaller RMSE measure than HAR (0.6135 and 0.6147 respectively), but after performing the Diebold-Mariano test ($p = 0.65$) we conclude that the models' predictive accuracy is not significantly different.

The most interesting result in this section is therefore the superior performance of the HAR model in the crisis period. A more probable explanation than HAR performing superiorly in high-volatility periods is the sensitivity of the ARFIMA model to small samples. A Monte Carlo simulation by Stadnytska and Werner (2006) indicates that at least 1000 observations are needed to estimate ARFIMA$(1, d, 1)$ parameters accurately. If the sample size is smaller the estimation of the fractional differencing parameter d becomes somewhat unreliable. The parameter estimation errors seem to contaminate the estimate of the long memory parameter and the forecasting accuracy drops significantly compared to longer estimation periods (Bhardwaj & Swanson, 2006).

Our crisis subsample has only 180 observations. Furthermore, the pre-crisis subsample has 653 observations, which is still below the desired 1000. The superiority of the HAR model in the pre-crisis and crisis regime is then most probably due to the fact that it can be estimated

more precisely with a relatively small number of observations. Contrarily, the ARFIMA model needs a much longer estimation period for accurate parameter estimation and consequently accurate forecasts.

Izzeldin et al. (2019) found that both ARFIMA and HAR perform worse during the crisis based on MAE measures, but suggested that the HAR model is less sensitive to changes in market conditions, since ARFIMA had a bigger increase in error during the crisis. However, they were able to draw the comparison between different periods since their sample was split in half into a pre-crisis and crisis period with the same number of observations.

We hypothesised that the HAR model is less sensitive to structural breaks than the ARFIMA model, but we cannot confirm this hypothesis since the models do not differ significantly in their performance.

We may use the 10% significance level with the Diebold – Mariano test as has been done by Bhardwaj and Swanson (2006). In this case we could conclude that the HAR model appears to be more sensitive to changes in market conditions, since it outperforms the ARFIMA model during the crisis but is less accurate for the full sample. In any case, the second hypothesis does not hold.

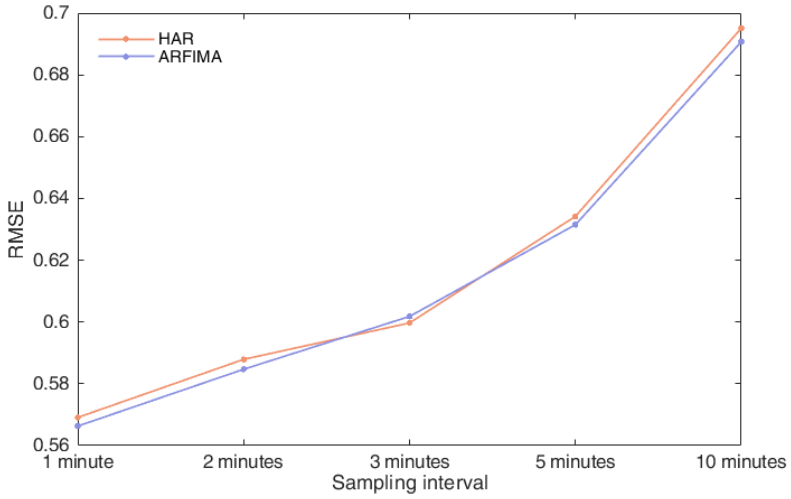
5.3 Sampling frequencies

Realized variance theory tells us that forecasts should become more precise as the sampling frequency increases. However, this holds only up to moderately high frequencies, since the signal is contaminated by the market microstructure noise at ultra-high frequencies. Therefore, realized variance based on 5-minute returns is typically used in the literature. By comparing the forecasts based on $lnRV$ sampled at higher (every 1, 2 and 3 minutes) and lower (every 10 minutes) frequencies we will determine, which sampling frequency produces the most accurate forecasts of $lnRV$.

When comparing daily forecasts of log realized variance computed from returns sampled at different frequencies, we find that both models produce the most accurate forecasts when $lnRV$ is based on returns sampled every minute. The RMSE value then increases for both models as the sampling interval becomes longer.

These results are consistent with Patton and Sheppard (2009), who indicate that forecasts based on 1-minute realized variance are the most accurate, followed by 5-minute, 5-second and finally 1-second RV. On the other hand, Izzeldin et al. (2019) find that the forecasting accuracy peaks at the highest frequency they considered, that is 5 seconds. The forecasts based on 15-second $lnRV$ are also superior to the log realized variance sampled at 1-minute frequency. Unfortunately, our data consists of 1-minute returns so we cannot compare the performance at higher sampling frequencies.

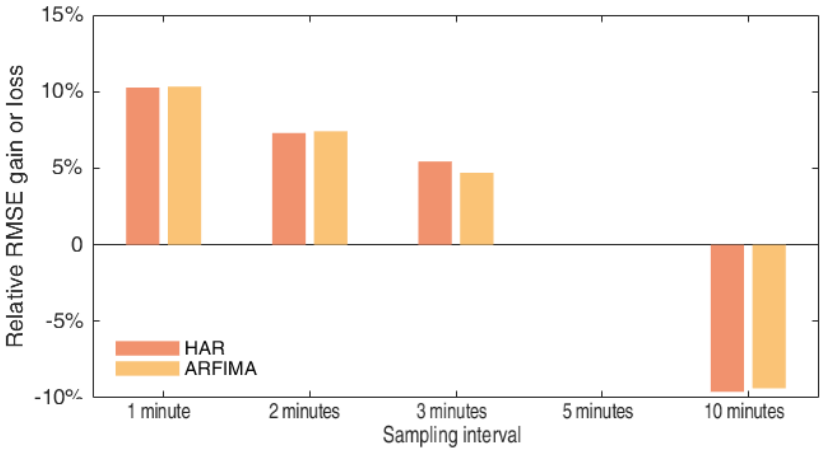
Figure 11: RMSE values for different sampling intervals



Source: Own work.

The greater accuracy of the 1-min $lnRV$ forecasts suggests minimal presence of the microstructure noise at this frequency. The NASDAQ Composite index is one of the most heavily traded indices and thus very liquid, which indicates a relatively low impact of the microstructure noise in the first place. Furthermore, Hansen and Lunde (2006b) discovered that the presence of microstructure noise has decreased over time. Hence, we conclude that the bias in 1-minute returns is negligible and so the 1-min $lnRV$ is preferred to the 5-min realized measures.

Figure 12: Relative RMSE gain or loss for different sampling intervals



Source: Own work.

While there are quite significant differences between forecasts based on different sampling intervals ($\pm 10\%$ with respect to the 5-min $lnRV$), the two models perform rather similarly for the specific sampling interval. In four out of five cases ARFIMA is more accurate than HAR, which is superior only for 3-min $lnRV$ forecasts. The only highly statistically significant difference is at the longest sampling interval, that is, when returns are sampled

every 10 minutes. In this case the ARFIMA model significantly outperforms the HAR model at the 5% significance level.

Table 7: RMSE values for different sampling intervals and DM test results

Sampling interval	RMSE		DM test statistic	p-value	Best performer
	HAR	ARFIMA			
1 min	0.5691	0.5663	- 1.49	0.1348	ARFIMA
2 min	0.5879	0.5847	-1.72	0.0862	ARFIMA*
3 min	0.5997	0.6018	0.46	0.6439	HAR
5 min	0.6342	0.6315	-1.33	0.1846	ARFIMA
10 min	0.6951	0.6907	-2.06	0.0395	ARFIMA**
average	0.6172	0.615	-	-	ARFIMA

* 10% significance level, ** 5% significance level

Source: Own work.

Additionally, ARFIMA beats HAR based on 2-min $lnRV$ forecasts, a result significant at the 10% level. By taking the average RMSE for each model, ARFIMA is the better performing model. This is out of line with the results by Izzeldin et al. (2019), who find that HAR is superior, although the two studies differ in sampling frequencies taken in consideration.

Our third hypothesis, that the ARFIMA model performs better at higher sampling frequencies, while the HAR model performs better at lower sampling frequencies, therefore holds only partially. ARFIMA is superior at higher frequencies (although only at 10% significance level), whereas HAR is certainly not superior at lower frequencies.

6 LIMITATIONS AND FURTHER RESEARCH

There are several ways in which this study could be extended or done differently. Sometimes unexpected obstacles appear during research, which would be addressed differently the second time around. On the other hand, some questions cannot be answered because data needed is not available. In this section we shortly present some of the shortcomings of this study.

There are alternative realized measures which could be taken into account such as the square root of realized variance (realized volatility), which is also nearly Gaussian distributed. Furthermore, the study could be extended by using $lnRV$ with the incorporated overnight return. Since many studies analyse currency volatility and the foreign exchange markets are open around-the-clock, the overnight return does not impact their results, while we cannot claim this for our study. Additionally, we could compare our forecasts with the ones based on $lnRV$ constructed by subsampling or other techniques proposed by Zhang et al. (2005).

Therefore, we could possibly further improve forecast accuracy by considering various realized measures. However, we managed to compare the performance of the HAR and

ARFIMA models by using the logarithm of realized variance, which was actually the purpose of this thesis.

Forecasts based on even higher frequencies should be made for more robust results about the model accuracy. Especially, since the best results were obtained at the highest frequency we considered. Unfortunately, we did not have tick data at our disposal.

The above considerations suggest how the study could be extended and made more robust. However, there is one significant limitation of our results, which prevented us from reaching an unambiguous conclusion. Namely, when analysing the performance of the models in different market conditions, the pre-crisis and crisis samples are too small. By forecasting in a high-volatility period with at least 1000 observations, we could conclude if the HAR model makes more accurate forecasts in such regimes or if the ARFIMA model is superior when it does not suffer from small sample bias.

To carry out such an analysis a larger dataset should be used to find a longer period of higher volatility (e.g. 1997 – 2002). Alternatively, we could define the structural breaks differently, but then the crisis period would not necessarily have a higher volatility than the other periods.

Finally, the study could be extended by constructing and evaluating density forecasts as opposed to point forecasts or by modelling and forecasting in a multivariate framework.

CONCLUSION

Volatility is an essential quantity in finance since it is related to the notions of risk and uncertainty. In the recent years, the wider availability of high-frequency data has motivated research and non-parametric realized measures of volatility have been proposed. They represent a consistent measure of the true volatility, which is unobservable. This allows for accurate forecasts, which are crucial in risk or portfolio management and option pricing.

The realized variance is a simple measure constructed by summing up intraday logarithmic returns generated over short (e.g. 5-minute) intervals. Because RV is approximately lognormal distributed, we decided to compare the models based on forecasts of log realized variance ($lnRV$). In any case, these realized measures, may it be variance, volatility, or log variance, take advantage of intraday data and measure the return volatility more accurately than squared daily returns. And a better estimate of the target quantity translates into improved forecasting performance.

Based on the theory of quadratic variation, realized variance constructed from returns sampled at infinitesimally small intervals would converge to the integrated variance, which is regarded as the ‘true’ volatility generated over some time interval. However, at very high frequencies market microstructure features, such as the discreteness of prices, the bid-ask bounce and infrequent trading, contaminate the signal. Therefore, the aim is to sample at the highest frequency at which the impact of market microstructure noise will be negligible.

The logarithm of realized variance displays long memory, which is characterised by a slowly decaying autocorrelation function. Such processes have usually been modelled with an autoregressive fractionally integrated moving average (ARFIMA) model. Recently, the heterogeneous autoregressive (HAR) model has been proposed as an alternative, which captures the long memory through a sum of simple autoregressive processes.

In this thesis the predictive accuracy of these two autoregressive models, HAR and ARFIMA, is compared. Their performance is compared at various forecasting horizons, in different market conditions and based on realized variance computed from returns sampled at different frequencies. We use high frequency intraday data from January 2006 to May 2020 to construct realized variance of the NASDAQ Composite index.

Although the literature on realized volatility and autoregressive models is extensive, only two studies have used both HAR and ARFIMA models. Degiannakis and Floros (2013) compare only their ability to forecast one-day ahead, while Izzeldin et al. (2019) compare also the accuracy for weekly and monthly forecasting. The latter study compares the models in the pre-crisis and crisis period, but the sample is split very arbitrarily. To compare the forecasting accuracy in different market conditions we instead split the sample by applying an algorithm for detection of structural breaks in variance.

Furthermore, in existing studies only simple ARFIMA models with up to two autoregressive and moving average lags are considered. We estimated more than eighty ARFIMA models to account for the possibility that a model with more lags may fit the data better.

Based on the recent literature we hypothesise that the HAR model is less sensitive to changes in market conditions and changes in forecasting horizon than the ARFIMA model. HAR should perform better when forecasts are based on RV constructed from lower-frequency returns, while ARFIMA should be superior when RV is constructed from high-frequency returns.

In line with our first hypothesis, we find that the HAR model is much less sensitive to weekly and monthly forecasting, since it significantly outperformed the ARFIMA model at the longest horizon considered. On a daily basis, however, the performance of the models was comparable.

We reject the hypothesis that the HAR model is less sensitive to structural breaks since both models performed similarly in all market conditions. HAR was slightly superior during the crisis, but we were not able to distinguish if this is because of better accuracy during high-volatility periods or more probably, HAR being less sensitive to sample size than ARFIMA.

Contrary to expectations, both models performed similarly at the highest sampling frequency, while ARFIMA outperformed HAR at the lowest frequency. There was no significant difference between the accuracy of the models based on 5-min $lnRV$, which had been used for long horizon and market condition forecasting. Since the most accurate

forecasts were achieved at the highest frequency, no model was clearly preferred based on this criterion.

The main conclusion is that the forecast accuracy of the HAR and ARFIMA models is remarkably similar. HAR proved to be significantly superior in some important circumstances, such as long horizon forecasting and forecasting during the crisis period. On the other hand, ARFIMA significantly outperformed HAR only when realized variance was based on 10-minute returns.

Since the performance of the two models is similar, the HAR model is less sensitive to sample size, much more trivial to estimate and the forecast procedure associated with it is much easier to implement, we recommend the HAR model for forecasting realized variance.

REFERENCE LIST

1. Agiakloglou, C., Newbold, P., & Wohar, M. (1993). Bias in an Estimator of the Fractional Difference Parameter. *Journal of Time Series Analysis, 14(3),* 235–246.
2. Ahoniemi, K., & Lanne, M. (2013). Overnight stock returns and realized volatility. *International Journal of Forecasting, 29(4),* 592–604.
3. Amemiya, T. (1994). *Introduction to statistics and econometrics*. Harvard University Press.
4. Amihud, Y., & Mendelson, H. (1986). Asset pricing and the bid-ask spread. *Journal of Financial Economics, 17(2),* 223–249.
5. Amihud, Y., & Mendelson, H. (1987). Trading Mechanisms and Stock Returns: An Empirical Investigation. *Journal of Finance, 42(3),* 533–553.
6. Andersen, T. G., & Benzoni, L. (2009). Realized Volatility. In T. Mikosch, J.P. Kreiss, R. Davis & T. Andersen (Eds.), *Handbook of Financial Time Series*. Berlin: Springer.
7. Andersen, T. G., & Bollerslev, T. (1997). Intraday periodicity and volatility persistence in financial markets. *Journal of Empirical Finance, 4(2),* 115–158.
8. Andersen, T. G., & Bollerslev, T. (1998). Answering the Skeptics: Yes, Standard Volatility Models do Provide Accurate Forecasts. *International Economic Review, 39(4),* 885.
9. Andersen, T. G., & Lund, J. (1997). Estimating continuous-time stochastic volatility models of the short-term interest rate. *Journal of Econometrics, 77(2),* 343–377.
10. Andersen, T. G., Bollerslev, T., & Diebold, F. X. (2010). Parametric and Nonparametric Volatility Measurement. In Y. Aït-sahalia & L. P. Hansen (Eds.), *Handbook of Financial Econometrics: Tools and Techniques* (67–137). North-Holland.
11. Andersen, T., Bollerslev, T., Diebold, F. X., & Ebens, H. (2001). The distribution of realized stock return volatility. *Journal of Financial Economics, 61(1),* 43–76.

12. Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (2000). Exchange Rate Returns Standardized by Realized Volatility are (Nearly) Gaussian. *Multinational Finance Journal, 4*, 159–179.
13. 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,* 42–55.
14. Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (2003). Modelling and Forecasting Realized Volatility. *Econometrica, 71(2),* 579–625.
15. Andrews, G. E., Askey, R., Roy, R. (1999). *Special Functions*. Cambridge University Press.
16. Areal, N. M., & Taylor, S. J. (2002). The realized volatility of FTSE-100 futures prices. *Journal of Futures Markets: Futures, Options, and Other Derivative Products, 22(7),* 627-648.
17. Arneodo, A., Muzy, J. F., & Sornette, D. (1998). Direct causal cascade in the stock market. *The European Physical Journal B-Condensed Matter and Complex Systems, 2(2),* 277-282.
18. Audrino, F., Huang, C., & Okhrin, O. (2018). Flexible HAR model for realized volatility. *Studies in Nonlinear Dynamics & Econometrics, 23(3).*
19. Bachelier, L. (1900). Théorie de la spéculation. *Annales scientifiques de l'École normale supérieure*, 17, 21-86.
20. Back, K. (1991). Asset pricing for general processes. *Journal of Mathematical Economics, 20(4),* 371-395.
21. Bai, J., Ghysels, E., & Wright, J. H. (2013). State space models and MIDAS regressions. *Econometric Reviews, 32(7),* 779-813.
22. Baillie, R. T. (1996). Long memory processes and fractional integration in econometrics. *Journal of Econometrics, 73(1),* 5–59.
23. Baillie, R. T., Calonaci, F., Cho, D., & Rho, S. (2019). Long Memory, Realized Volatility and Heterogeneous Autoregressive Models. *Journal of Time Series Analysis, 40(4),* 609–628.
24. Barndorff-Nielsen, O. E., & Shephard, N. (2004). Power and Bipower Variation with Stochastic Volatility and Jumps. *Journal of Financial Econometrics, 2(1),* 1–37.
25. Barndorff-Nielsen, O. E., & Shephard, N. (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models: Realized Volatility. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 64(2),* 253–280.
26. Beran, J. (1994). *Statistics for long-memory processes*. Chapman & Hall.
27. Bhardwaj, G., & Swanson, N. R. (2006). An empirical investigation of the usefulness of ARFIMA models for predicting macroeconomic and financial time series. *Journal of Econometrics, 131(1–2),* 539–578.
28. Black, F. (1976). The pricing of commodity contracts. *Journal of financial economics, 3(1-2),* 167-179.

29. Box, G. E. P., & Jenkins, G. M. (1970). *Time-series Analysis, Forecasting and control*. Holden-Day.
30. Breidt, F. J., Crato, N., & De Lima, P. (1998). The detection and estimation of long memory in stochastic volatility. *Journal of econometrics, 83(1-2),* 325-348.
31. Bucci, A. (2017). Forecasting Realized Volatility: A Review. *Journal of Advanced Studies in Finance, 8(16),* 94–138.
32. Campbell, J. Y., Lo, A. W., & MacKinlay, A. C. (1997). *The econometrics of financial markets*. Princeton University Press.
33. Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. *Geoscientific Model Development, 7(3),* 1247–1250.
34. Chiriac, R., & Voev, V. (2011). Modelling and forecasting multivariate realized volatility. *Journal of Applied Econometrics, 26(6),* 922-947.
35. Christiansen, C., Schmeling, M., & Schrimpf, A. (2012). A comprehensive look at financial volatility prediction by economic variables. *Journal of Applied Econometrics, 27(6),* 956-977.
36. Christoffersen, P. F. (2012). *Elements of Financial Risk Management* (2nd ed.). Elsevier Academic Press.
37. Cont, R. (2001). Empirical properties of asset returns: Stylized facts and statistical issues. *Quantitative Finance, 1(2),* 223–236.
38. Corsi, F. (2009). A Simple Approximate Long-Memory Model of Realized Volatility. *Journal of Financial Econometrics, 7(2),* 174–196.
39. Corsi, F., Mittnik, S., Pigorsch, C. & Pigorsch, U. (2008). The Volatility of Realized Volatility. *Econometric Reviews*, 27(1-3), 46-78.
40. 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.
41. Craioveanu, M., & Hillebrand, E. (2014). Why It Is OK to Use the HAR-RV(1,5,21) Model. Working Paper.
42. Čech, F., & Baruník, J. (2017). On the modelling and forecasting of multivariate realized volatility: Generalized heterogeneous autoregressive (GHAR) model. *Journal of forecasting, 36(2),* 181-206.
43. Degiannakis, S., & Floros, C. (2013). Modeling CAC40 volatility using ultra-high frequency data. *Research in International Business and Finance, 28,* 68–81.
44. 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,* 427–431.
45. Dickey, D. A., & Fuller, W. A. (1981). Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root. *Econometrica, 49(4),* 1057–1072.
46. Diebold, F. X., & Mariano, R. S. (1995). Comparing Predictive Accuracy. *Journal of Business & Economic Statistics, 13(3),* 134–144.

47. Diebold, F. X., & Rudebusch, G. D. (1991). On the power of Dickey-Fuller tests against fractional alternatives. *Economics Letters, 35(2),* 155–160.
48. Ding, Z., Granger, C. W. J., & Engle, R. F. (1993). A long memory property of stock market returns and a new model. *Journal of Empirical Finance, 1(1),* 83–106.
49. Doornik, J. A., & Ooms, M. (1999). A Package for Estimating, Forecasting and Simulating Arfima Models: Arfima package 1.0 for Ox. 31.
50. Doornik, J. A., & Ooms, M. (2003). Computational aspects of maximum likelihood estimation of autoregressive fractionally integrated moving average models. *Computational Statistics & Data Analysis, 42,* 333–348.
51. Doornik, J. A., & Ooms, M. (2004). Inference and Forecasting for ARFIMA Models With an Application to US and UK Inflation. *Studies in Nonlinear Dynamics & Econometrics, 8(2).*
52. Engle, R. F., & Gallo, G. M. (2006). A multiple indicators model for volatility using intra-daily data. *Journal of Econometrics, 131(1-2),* 3-27.
53. Engle, R. F., & Ng, V. K. (1993). Measuring and testing the impact of news on volatility. *Journal of finance, 48(5),* 1749-1778.
54. Fama, E. F. (1965). The behavior of stock-market prices. *The journal of Business, 38(1),* 34-105.
55. Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance, 25(2),* 383–417.
56. Gençay, R., & Selçuk, F. (2004). Extreme value theory and Value-at-Risk: Relative performance in emerging markets. *International Journal of forecasting, 20(2),* 287-303.
57. Gençay, R., Dacorogna, M., Müller, U. A., Pictet, O., & Olsen, R. (2001). *An Introduction to High-Frequency Finance*. Elsevier.
58. Geweke, J., & Porter-Hudak, S. (1983). The estimation and application of long memory time series models. *Journal of time series analysis, 4(4),* 221-238.
59. Granger, C. W. J. (1980). Long memory relationships and the aggregation of dynamic models. *Journal of Econometrics, 14(2),* 227–238.
60. Granger, C. W. J., & Joyeux, R. (1980). An Introduction to Long-Memory Time Series Models and Fractional Differencing. *Journal of Time Series Analysis, 1(1),* 15–29.
61. Graves, T., Gramacy, R., Watkins, N., & Franzke, C. (2017). A Brief History of Long Memory: Hurst, Mandelbrot and the Road to ARFIMA, 1951–1980. *Entropy, 19(9),* 437.
62. Gray, R. M. (2006). Toeplitz and Circulant Matrices: A Review. *Foundations and Trends in Communications and Information Theory, 2(3),* 155–239.
63. Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press.
64. Hansen, P. R., & Lunde, A. (2005). A forecast comparison of volatility models: does anything beat a GARCH (1, 1)?. *Journal of applied econometrics, 20(7),* 873-889.
65. Hansen, P. R., & Lunde, A. (2006a). Consistent ranking of volatility models. *Journal of Econometrics, 131(1),* 97–121.

66. Hansen, P. R, & Lunde, A. (2006b). Realized Variance and Market Microstructure Noise. *Journal of Business & Economic Statistics, 24(2),* 127–161.
67. Hansen, P. R., Huang, Z., & Shek, H. H. (2012). Realized GARCH: a joint model for returns and realized measures of volatility. *Journal of Applied Econometrics, 27(6),* 877-906.
68. Hansen, P. R., Lunde, A., & Nason, J. M. (2011). The Model Confidence Set. *Econometrica, 79(2),* 453–497.
69. Harris, L. (1990). Estimation of Stock Price Variances and Serial Covariances from Discrete Observations. *The Journal of Financial and Quantitative Analysis, 25(3),* 291–306.
70. Harris, L. (1991). Stock Price Clustering and Discreteness. *The Review of Financial Studies, 4(3),* 389–415.
71. Hosking, J. R. M. (1981). Fractional Differencing. *Biometrika, 68(1),* 165–176.
72. Hyndman, R. J. & Athanasopoulos, G. (2018). *Forecasting: Principles and Practice*. OTexts.
73. Inclan, C., & Tiao, G. C. (1994). Use of Cumulative Sums of Squares for Retrospective Detection of Changes of Variance. *Journal of the American Statistical Association, 89(427),* 913–923.
74. Inzelt, G. (2011). *ARFIMA(p,d,q) estimator*. MATLAB Central File Exchange. Retrieved 7th July 2020 from: https://www.mathworks.com/matlabcentral/fileexchange/30238-arfima-p-d-q-estimator
75. Izzeldin, M., Hassan, M. K., Pappas, V., & Tsionas, M. (2019). Forecasting realised volatility using ARFIMA and HAR models. *Quantitative Finance, 19(10),* 1627–1638.
76. Jondeau, E., Poon, S.-H., & Rockinger, M. (2007). *Financial Modelling Under Non-Gaussian Distirbutions*. Springer.
77. Koopman, S. J., Jungbacker, B., & Hol, E. (2005). Forecasting daily variability of the S&P 100 stock index using historical, realised and implied volatility measurements. *Journal of Empirical Finance, 12(3),* 445–475.
78. LeBaron, B. (2001). A builder's guide to agent-based financial markets. *Quantitative finance, 1(2),* 254-261.
79. 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,* 293–311.
80. Lo, A. W. (1991). Long-Term Memory in Stock Market Prices. *Econometrica, 59(5),* 1279–1313.
81. Lo, A. W., & Craig MacKinlay, A. (1990). An econometric analysis of nonsynchronous trading. *Journal of Econometrics, 45(1),* 181–211.
82. Lynch, P. E., & Zumbach, G. O. (2003). Market heterogeneities and the causal structure of volatility. *Quantitative Finance, 3(4),* 320-331.

83. Mandelbrot, B. B. (1963). The variation of certain speculative prices. *Fractals and scaling in finance*, 371-418.
84. Martens, M. P. E., De Pooter, M., & van Dijk, D. J. C. (2004). *Modeling and Forecasting S&P 500 Volatility: Long Memory, Structural Breaks and Nonlinearity.* Tinbergen Institute Discussion.
85. 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.
86. MathWorks. (2020). *findchangepts – Find abrupt changes in signal*. Retrieved 14th September 2020 from: https://www.mathworks.com/help/signal/ref/findchangepts.html
87. McAleer, M., & Medeiros, M. C. (2008). Realized Volatility: A Review. *Econometric Reviews, 27(1–3),* 10–45.
88. Merton, R. C. (1980). *On estimating the expected return on the market: An exploratory investigation (No. w0444).* National Bureau of Economic Research.
89. Müller, U. A., Dacorogna, M. M., Davé, R. D., Pictet, O. V., Olsen, R. B., & Ward, J. R. (1995). *Fractals and Intrinsic Time: A Challenge to Econometricians*. Discussion Paper Presented at the 1993 International Conference of the Applied Econometrics Association.
90. Müller, U. A., Dacorogna, M. M., Davé, R. D., Olsen, R. B., Pictet, O. V., & Von Weizsäcker, J. E. (1997). Volatilities of different time resolutions—analyzing the dynamics of market components. *Journal of Empirical Finance, 4(2-3),* 213-239.
91. NASDAQ. (2020). NASDAQ Composite. Retrieved 4th September 2020 from: https://www.mathworks.com/help/signal/ref/findchangepts.html
92. Nelson, D. B. (1990). ARCH models as diffusion approximations. *Journal of econometrics, 45(1-2),* 7-38.
93. Oomen, R. C. A. (2006). Properties of Realized Variance Under Alternative Sampling Schemes. *Journal of Business & Economic Statistics, 24(2),* 219–237.
94. Pagan, A. R., & Schwert, G. W. (1990). Alternative models for conditional stock volatility. *Journal of econometrics, 45(1-2),* 267-290.
95. Palma, W. (2007). *Long-Memory Time Series: Theory and Methods*. John Wiley & Sons, Inc.
96. Patton, A. J. (2011). Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics, 160(1),* 246-256.
97. Patton, A. J., & Sheppard, K. (2009). Optimal combinations of realised volatility estimators. *International Journal of Forecasting, 25(2),* 218–238.
98. Paye, B. S. (2012). 'Déjà vol': Predictive regressions for aggregate stock market volatility using macroeconomic variables. *Journal of Financial Economics, 106(3),* 527-546.

99. Pong, S., Shackleton, M. B., Taylor, S. J., & Xu, X. (2004). Forecasting currency volatility: A comparison of implied volatilities and AR (FI) MA models. *Journal of Banking & Finance, 28(10),* 2541-2563.
100. Poon, S.-H., & Granger, C. W. J. (2003). Forecasting Volatility in Financial Markets: A Review. *Journal of Economic Literature, 41*, 478–539.
101. Protter, P. E. (2004). *Stochastic integration and differential equations* (2nd ed). Springer.
102. Reisen, V., Abraham, B., & Lopes, S. (2001). Estimation of Parameters in ARFIMA Processes: A Simulation Study. *Communications in Statistics: Simulation and Computation, 30(4),* 787–803.
103. Roll, R. (1984). A Simple Implicit Measure of the Effective Bid-Ask Spread in an Efficient Market. *The Journal of Finance, 39(4),* 1127–1139.
104. Ross, G. J. (2013). Modelling financial volatility in the presence of abrupt changes. *Physica A: Statistical Mechanics and Its Applications, 392(2),* 350–360.
105. Shephard, N., & Sheppard, K. (2010). Realising the future: forecasting with high-frequency-based volatility (HEAVY) models. *Journal of Applied Econometrics, 25(2),* 197-231.
106. Sorjamaa, A. & Lendasse, A. (2006). Time series prediction using DirRec strategy. In Verleysen, M. (Ed.) *European Symposium on Artificial Neural Networks* (143–148).
107. Sowell, F. (1992). Maximum likelihood estimation of stationary univariate fractionally integrated time series models. *Journal of Econometrics, 53(1),* 165–188.
108. Stadnytska, T., & Werner, J. (2006). Sample Size and Accuracy of Estimation of the Fractional Differencing Parameter. *Methodology, 2(4),* 135–141.
109. Stock, J. H., & Watson, M. W. (2020). *Introduction to econometrics*. Boston: Pearson.
110. Taylor, S. J. (1986). *Modelling Financial Time Series*. John Wiley & Sons.
111. Willmott, C., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research, 30,* 79–82.
112. Zhang, L., Mykland, P. A., & Aït-Sahalia, Y. (2005). A Tale of Two Time Scales: Determining Integrated Volatility With Noisy High-Frequency Data. *Journal of the American Statistical Association, 100(472),* 1394–1411.
113. Zhou, B. (1996). High-Frequency Data and Volatility in Foreign-Exchange Rates. *Journal of Business & Economic Statistics, 14(1),* 45–52.
114. Zivot, E. (2005). *Analysis of High Frequency Financial Data: Models, Methods and Software. Part II: Modeling and Forecasting Realized Variance Measures*. Technical Report UWEC-2005, 38

APPENDICES

Appendix 1: Povzetek v slovenskem jeziku

Volatilnost je bistvenega pomena v financah, saj je povezana s tveganjem in negotovostjo. V zadnjih letih je veliko raziskav izkoristilo večjo razpoložljivost visokofrekvenčnih podatkov in razvila se je teorija realiziranih mer volatilnosti, ki dosledno ocenjujejo pravo volatilnost, te pa ne moremo neposredno opazovati. Z realiziranimi neparametričnimi merami lahko tako izdelamo točne napovedi, ki so ključnega pomena pri upravljanju tveganj ali določanju vrednosti opcij.

Realizirana varianca (RV) je preprosta mera, ki jo dobimo s seštevanjem kvadratnih logaritemskih donosov, ustvarjenih v zelo kratkih časovnih obdobjih v dnevu (na primer vsakih 5 minut). Ker je RV porazdeljena približno lognormalno, smo se odločili za primerjavo modelov, ki temeljijo na napovedih logaritma realizirane variance (lnRV). Realizirane mere, naj gre za varianco, volatilnost ali logaritem variance, izkoriščajo visokofrekvenčne podatke znotraj dneva in finančno volatilnost izmerijo natančneje kot kvadrat celodnevnega donosa. Ker bolje ocenijo ciljno količino, je tudi napovedovanje zaradi tega bolj učinkovito.

Realizirana varianca konvergira k integrirani varianci, ki velja za 'pravo' volatilnost, generirano v določenem obdobju, če so donosi, na katerih temelji, izračunani za infinitezimalno kratka obdobja. Vendar to konvergiranje v praksi otežuje prisotnost mikrostrukturnih učinkov, kot so diskretnost cen, odboj med nakupno in prodajno kotacijo in časovno neenakomerno trgovanje. Cilj je donose izračunati v najmanjših možnih intervalih, pri katerih bo vpliv mikrostrukturnih učinkov zanemarljiv.

Logaritem realizirane variance kaže dolgoročno odvisnost oziroma t.i. 'dolg spomin', za katerega je značilna počasi padajoča funkcija avtokorelacije. Za take procese se ponavadi uporabljajo avtoregresijski delno integrirani modeli drsečih sredin (ARFIMA). Kot alternativa je bil predlagan heterogeni avtoregresijski (HAR) model, ki uspe zajeti dolg spomin s seštevanjem komponent preprostih avtoregresijskih procesov.

V magistrskem delu primerjamo natančnost napovedi dveh avtoregresijskih HAR in ARFIMA modelov. Njuno natančnost primerjamo glede na obdobje napovedi, različne tržne razmere in pri realizirani varianci, ki izhaja iz donosov, izračunanih v različnih intervalih. Pri generiranju realizirane variance indeksa NASDAQ Composite uporabljamo visokofrekvenčne podatke od januarja 2006 do maja 2020.

Čeprav je literatura o realiziranih merah in avtoregresijskih modelih obsežna, sta le dve študiji uporabili oba modela, ki ju obravnavamo. Degiannakis in Floros (2013) primerjata samo točnost dnevnih napovedi, Izzeldin et al. (2019) pa tudi natančnost za tedenska in mesečna obdobja. Slednja študija tudi primerja modele v predkriznih in kriznih razmerah, vendar je prelom med obdobjema določen zelo arbitrarno. V tem magistrskem delu pa je pri analizi napovedi v različnih tržnih razmerah časovna serija razdeljena z uporabo algoritma za odkrivanje strukturnih prelomov.

Poleg tega so v obstoječih študijah upoštevani le preprosti ARFIMA modeli do največ drugega reda. V tem delu pa ocenimo več kot osemdeset ARFIMA modelov z namenom upoštevanja možnosti, da model z več odlogi bolje opisuje podatke.

Na podlagi obstoječe literature postavimo hipoteze, da je HAR model manj občutljiv na spremembe tržnih razmer in spremembe napovedovalnega obdobja kot model ARFIMA. Poleg tega predvidevamo, da bo HAR natančnejši, če napovedi temeljijo na RV, temelječi na nižjefrekvenčnih donosih (5- ali 10-minutnih). ARFIMA pa bi morala biti boljša pri realizirani varianci, temelječi na visokofrekvenčnih donosih (1-minutnih).

V skladu s predvidevanji ugotavljamo, da je HAR model veliko manj občutljiv pri dolgoročnem napovedovanju, saj je pri mesečnih napovedih statistično značilno natančnejši od modela ARFIMA. Pri dnevnih napovedih je točnost napovedi za oba modela primerljiva.

Zavračamo hipotezo, da je model HAR manj občutljiv na strukturne prelome, saj v nobeni od tržnih razmer ni statistično značilne razlike med modeloma. HAR model je med krizo nekoliko natančnejši, vendar tega ne moremo pripisati boljši zmogljivosti v obdobjih z visoko volatilnostjo, ampak je bolj verjetno, da je HAR manj občutljiv na velikost vzorca.

V nasprotju s pričakovanji sta oba modela primerljiva pri najvišji frekvenci vzorčenja donosov, medtem ko je ARFIMA uspešnejša od modela HAR, če realizirana varianca temelji na 10-minutnih donosih. Natančnost modelov je primerljiva tudi pri realizirani varianci, ki temelji na 5-minutnih donosih. To je pomembno, ker je bila 5-minutna realizirana varianca uporabljena za dolgoročno napovedovanje in napovedovanje v različnih tržnih razmerah. Najbolj natančne napovedi obeh modelov so dosežene pri najvišji frekvenci donosov. Ker pa sta tudi v tem primeru modela podobno natančna, ne identificiramo modela, ki bi imel prednost pred drugim na podlagi različnih frekvenc vzorčenja donosov.

Glavni sklep je, da je natančnost napovedi HAR in ARFIMA modelov izredno podobna. HAR se je izkazal za značilno boljšega v nekaterih pomembnih okoliščinah, kot sta napovedovanje na dolgi rok in v kriznem obdobju. Po drugi strani pa je ARFIMA znatno boljša od modela HAR le, ko je realizirana varianca temeljila na 10-minutnih donosih. Na podlagi rezultatov sklepamo, da je model HAR najprimernejši za napovedovanje realizirane variance, saj je manj občutljiv na velikost vzorca in veliko enostavnejši pri ocenjevanju parametrov in napovedovanju.

Appendix 2: BIC values for ARFIMA(p, d, q) models

p	q	log-likelihood	BIC
0	0	-2888.3992	5801.33
0	1	-2886.9446	5806.60
1	0	-2886.8397	5806.39
1	1	-2877.6046	5796.1††
1	2	-2877.2857	5803.64
1	3	-2876.8504	5810.95
1	4	-2876.8417	5819.11
1	5	-2876.0073	5825.62
1	6	-2874.8094	5831.40
1	7	-2874.3552	5838.67
1	8	-2874.2501	5846.64
1	9	-2873.6879	5853.38
2	1	-2877.2570	5803.58
2	2	-2875.9677	5809.18
2	3	-2876.8505	5819.13
2	4	-2876.8312	5827.27
2	5	-2874.5099	5830.80
2	6	-2874.4418	5838.84
2	7	-2874.3203	5846.78
2	8	-2871.1082	5848.53
2	9	-2871.0814	5856.66
3	1	-2876.7665	5810.78
3	2	-2875.6557	5816.74
3	3	-2873.8038	5821.21
3	4	-2873.9388	5829.66
3	5	-2870.5693	5831.10
3	6	-2871.9362	5842.01
3	7	-2869.3353	5844.99
3	8	-2871.0925	5856.68
3	9	-2871.0252	5864.72
4	1	-2876.6874	5818.80
4	2	-2876.7639	5827.13
4	3	-2873.6623	5829.11
4	4	-2874.4065	5838.77
4	5	-2869.4603	5837.06
4	6	-2868.0945	5842.50
4	7	-2868.6364	5851.77
4	8	-2866.3648	5855.40
4	9	-2866.2963	5863.44
5	1	-2875.2751	5824.15
5	2	-2874.5112	5830.71
5	3	-2873.4646	5836.89
5	4	-2869.5541	5837.25
5	5	-2866.7662	5839.85
5	6	-2867.9849	5850.46
5	7	-2866.1163	5854.90
5	8	-2865.9483	5862.75
5	9	-2867.2032	5873.43
6	1	-2874.6793	5831.14
6	2	-2874.6760	5839.31
6	3	-2871.9970	5842.13
6	4	-2867.8122	5841.94
6	5	-2867.8330	5850.16
6	6	-2865.0924	5852.86
6	7	-2864.8355	5860.52
6	8	-2866.4916	5872.01
6	9	-2864.5181	5876.24
7	1	-2874.6783	5839.32
7	2	-2874.6837	5847.50
7	3	-2868.8149	5843.95
7	4	-2867.7852	5850.06
7	5	-2866.3740	5855.42
7	6	-2876.1908	5865.23
7	7	-2864.7960	5868.62
7	8	-2864.1661	5875.54
7	9	-2864.1652	5883.71
8	1	-2874.4458	5847.03
8	2	-2871.9213	5850.16

†† smallest BIC value

p	q	log-likelihood	BIC
8	3	-2871.7404	5857.97
8	4	-2865.9258	5854.52
8	5	-2866.8311	5864.51
8	6	-2866.5122	5872.05
8	7	-2864.1664	5875.54
8	8	-2862.4496	5880.28
8	9	-2862.2972	5888.16
9	1	-2873.6270	5853.57
9	2	-2871.1110	5856.72
9	3	-2870.9730	5864.62
9	4	-2865.9255	5862.70
9	5	-2866.7004	5872.43
9	6	-2864.7057	5876.62
9	7	-2864.1645	5883.71
9	8	-2863.5750	5890.71
9	9	-	-

Source: Own work.

Appendix 3: Jarque – Bera (JB) test results for forecast errors

forecast	JB_{ARFIMA}	JB_{HAR}
daily	42.8422 (0.0001)	44.5585 (0.0001)
weekly	134.5842 (0.0001)	123.7272 (0.0001)
monthly	279.6811 (0.0001)	279.4659 (0.0001)
pre-crisis	9.2960 (0.0185)	9.1184 (0.0192)
crisis	1.4616 (0.3288)	1.3891 (0.3469)
post-crisis	18.6014 (0.0016)	20.5482 (0.0012)
1 minute	60.3526 (0.0001)	61.3295 (0.0001)
2 minutes	55.3181 (0.0001)	57.3209 (0.0001)
3 minutes	71.5053 (0.0001)	67.6166 (0.0001)
10 minutes	31.0989 (0.0001)	32.9348 (0.0001)

Note: The null hypothesis of the Jarque – Bera test is that data comes from a normal distribution with skewness equal to 0 and kurtosis equal to 3. The test was performed in MATLAB, which uses a modified algorithm to calculate critical values for small samples; p – values are in parentheses.

Source: Own work.

Appendix 4: MATLAB Functions

1. Generating RV series sampled at different frequencies

```
function [lnRV, daily_r] = aggregate_frequencies(r, delta)

% INPUTS:
    %r: an array of 1-minute returns
    %delta: level of aggregation (delta = 5 -> 5-min RV)

% OUTPUTS:
    %lnRV: an array of log aggregated RV series
    %daily_r: an array of daily returns

ndays = size(r,2);

r_delta=reshape(r,delta,390/delta,ndays);
r_delta=squeeze(sum(r_delta,1));

RV=sum(r_delta.^2)';
lnRV = log(RV);
daily_r = sum(r_delta)';
end
```

Source: Own work.

2. Generating HAR regressors (daily, weekly, monthly RV)

```
function [X_matrix] = X_mat(series)
%creates a matrix with daily, weekly and monthly RV from a series of daily RV
% weekly RV
bw = (1/5)*ones(1,5);
RV_w = filter(bw,1,series);

for i=1:5-1
    RV_w(i)=RV_w(i)*5/i;
end

% monthly RV
bm = (1/22)*ones(1,22);
RV_m = filter(bm,1,series);

for i=1:22-1
    RV_m(i)=RV_m(i)*22/i;
end
X_matrix = [series RV_w RV_m];
end
```

Source: Own work.

3. HAR forecasting

```
function [RMSE, error] = har_forecasting(Z, s, R);

%INPUTS:
    %Z: RV series
    %s: forecast horizon
    %R: rolling window length

%OUTPUTS:
    %RMSE: root mean square error
    %errors: an array of forecast errors

r = length(Z) - R;
if s == 1
    for i=1:r-1
        series = Z(i:R+i-1);
        matrix = X_mat(series);
        X = [ones(length(matrix),1) matrix];
        Y = series(2:end);
        beta_hat(1,1:4) = X(1:end-1,:)\ Y;
        prediction (i,1) = X(end,:)*beta_hat(1,:)';
        observed(i,1) = Z(R+i);
        error(i,1) = observed(i,1)-prediction(i,1);
    end
elseif s > 1
      for i=1:r-s
        series = Z(i:R+i-1);
        matrix = X_mat(series);
        X = [ones(length(matrix),1) matrix];
        Y = series(2:end);
        beta_hat(1,1:4) = X(1:end-1,:)\ Y;
        prediction (1,i) = X(end,:)*beta_hat(1,:)';
        observed(1,i) = Z(R+i);
        err(1,i) = observed(1,i)-prediction(1,i);
           for h = 1:s
            series2 = [Z(h+i:R+i-1); prediction(1:h,i)];
            matrix2 = X_mat(series2);
            X2 = [ones(length(matrix2),1) matrix2];
            Y2 = series2(2:end);
            beta_hat(h+1,1:4) = X2(1:end-1,:)\ Y2;
            prediction (h+1,i) = X2(end,:)*beta_hat(h+1,:)';
            observed(h+1,i) = Z(R+i+h);
            err(h+1,i) = observed(h+1,i)-prediction(h+1,i);
        end
        error(i,1) = err(end-1, i);
        yhat(i,1) = prediction(end-1,i);
      end
end
```

```
RMSE = sqrt(mean(error.^2));
end
```

Source: Own work.

4. ARFIMA forecasting

```
function [RMSE, errors] = arfima_forecasting(Z, s, ar, ma, d, sigma, R)

%INPUTS:
    %Z: time series
    %s: forecast horizon
    %ar: an array of AR parameters
    %ma: an array of MA parameters
    %d: an array of d parameters
    %sigma: an array of sigma parameters
    %R: size of rolling window used in estimation

%OUTPUTS:
    % RMSE: root mean square error
    %errors: an array of errors
    %y: an array of observed values
    %yhat: an array of predicted values

r = length(Z) - R;
if s == 1
    for i=1:r
        mu(i,1)= mean(Z(i:R+i-1));
    end
    for i =1:r
        series = Z(i:R-1+i);
        gamma_s = arfima_covs(length(series)+1,[d(i) ar(i) ma(i) sigma(i)],[1
1]);
        [~,L] = durlevML(gamma_s);
        L = reshape(L,length(series)+1,length(series)+1)';
        yhat(i,1) = mu(i) - L(length(series)+1, 1: length(series))*(series -
mu(i));
        obs(i,1) = Z(R+i);
        errors(i,1) = obs(i,1)-yhat(i,1);
    end
elseif s > 1
    for i=1:r+1-s
        series = Z(i:R+i-1);
        mu(1,i) = mean(series);
        gamma_1 = arfima_covs(length(series)+1,[d(i) ar(i) ma(i) sigma(i)],[1
1]);
        [~,L] = durlevML(gamma_1);
        L = reshape(L,length(series)+1,length(series)+1)';
```

```
        y(1,i) = mu(1,i) - L(length(series)+1, 1: length(series))*(series -
mu(1,i));
        for h = 1:s
            Z2 = [Z((h+1):R+i-1); y(1:h,i)];
            mu(h+1,i)= mean(Z2);
            gamma_2 = arfima_covs(length(Z2)+1,[d(i) ar(i) ma(i) sigma(i)],[1
1]);
            [~,L] = durlevML(gamma_2);
            L = reshape(L,length(Z2)+1,length(Z2)+1)';
            y(h+1,i) = mu(h+1,i) - L(length(Z2)+1, 1: length(Z2))*(Z2 -
mu(h+1,i));
        end
        obs(i,1) = Z(R-1+i+s);
        yhat(i,1) = y(end-1,i);
        errors(i,1) = obs(i,1)-yhat(i,1);
    end
end
RMSE = sqrt(mean(errors.^2));
end
```

Source: Own work.

5. Diebold-Mariano test

```
function [DM,p] = dbmar(e_arf,e_har)

%INPUTS:
    %e_arf = an array of ARFIMA forecast errors
    %e_har = an array of HAR forecast errors

%OUTPUTS:
    %DM = Diebold-Mariano test statistic
    %p = p-value of Diebold-Mariano test statistic
d = e_arf.^2 - e_har.^2;
avg_d = mean(d);
var_d = var(d);
T = length(e_arf);
DM = avg_d / sqrt((1/T)*var_d);
[~,p] = ztest(DM, 0, 1);
end
```

Source: Own work.

6. Computing ARMA covariance matrix (Inzelt, 2011)

```
function [gamma_arma] = arma_covs(lag,params,arma_part)
    % Ian McLeod's method for calculating the analytical ACF for linear
stationary ARMA(p,q)processes.
```

```
      % Reference: McLeod, Ian: Derivation of the Theoretical
Autocovariance Function of Autoregressive-Moving Average Time
Series.Applied Statistics(1975) 24 No.2 p255-257
      % (C) György Inzelt 2011
      wn_var = params(length(params));
      if  arma_part(1) == 0 && arma_part(2) == 0, phi = 0   ; theta =0;
        elseif arma_part(1) ==0 && arma_part(2) ~= 0 ,theta =
params(1,1:1+arma_part(2)-1)   ; phi =0;
        elseif arma_part(1) ~=0 && arma_part(2) ~=0, phi =
params(1,1:1+arma_part(1)-1);
               theta = params(1,1+arma_part(1):1+sum(arma_part)-1);
        elseif arma_part(1) ~=0 && arma_part(2)==0, phi =
params(1,1:1+arma_part(1)-1); theta = 0;
      end
     gamma_arma = zeros(lag,1);
    %ARMA(0,q) case
     if arma_part(1) == 0 && arma_part(2)~=0
        for ii = 0:1:lag
           if ii == 0
              gamma_arma(ii+1) = (sum([1 theta].^2))*wn_var ;
           elseif ii > 0 && ii < arma_part(2)
              gamma_arma(ii+1) = (theta(ii) + sum(
theta(ii+1:length(theta)).*theta(1:length(theta)-ii) ))*wn_var;
           elseif ii == arma_part(2)
              gamma_arma(ii+1) = theta(length(theta))*wn_var;
           elseif ii > arma_part(2)
              gamma_arma(ii+1) = 0;
           end
        end
     %ARMA(0,0)-just for completeness
     elseif arma_part(1)==0 && arma_part(2)==0
      gamma_arma(1) = wn_var;
      gamma_arma(2:lags) = 0;
     %ARMA(p,0)
     elseif arma_part(1) ~= 0 && arma_part(2)==0
        if arma_part(1)==1
           gamma_arma = (phi(1).^((0:1:lag)'))*(1/(1-phi(1)^2))*wn_var;%%
        elseif arma_part(1) > 1
           F = zeros(arma_part(1),arma_part(1));
           F(1,:) = phi;
           F(2:length(F),1:length(F)-1 )=eye(length(F)-1)  ;
           G = wn_var*inv(( eye((arma_part(1))^2) - kron(F,F)   )) ;
           gamma_arma(1:arma_part(1)) = G(1:arma_part(1),1);
           for jj = arma_part(1)+1:1:length(gamma_arma)
               gamma_arma(jj) =  phi*gamma_arma(jj-1:-1:jj-arma_part(1));
           end
        end
     %ARMA(p,q)
     elseif arma_part(2) ~=0 && arma_part(2) ~=0
       %calculating the cross-covariances
```

```
      [gamma_pacf, rhs, lhs]= arma_crosscov(phi,theta,arma_part,wn_var);
       gamma_arma(1:arma_part(1)+1,1)= lhs\rhs(1:arma_part(1)+1);
         if arma_part(1) >= arma_part(2)
           for jj = arma_part(1)+2:1:length(gamma_arma)
                   gamma_arma(jj) =  phi*gamma_arma(jj-1:-1:jj-
arma_part(1));
           end
         elseif arma_part(1) < arma_part(2)
             for jj = arma_part(1)+2:1:length(gamma_arma)
                if jj  <=length(rhs)
                   gamma_arma(jj) =  phi*gamma_arma(jj-1:-1:jj-
arma_part(1)) + rhs(jj);
                elseif jj  > length(rhs)
                   gamma_arma(jj) =  phi*gamma_arma(jj-1:-1:jj-
arma_part(1));
                end
             end
         end
     end
 end

 function[gamma_pacf, rhs, lhs]= arma_crosscov(phi,theta,arma_part,wn_var)
      gamma_pacf = zeros(arma_part(2)+1,1);
      rhs = zeros(max(arma_part(1)+1,arma_part(2)+1),1);
      lhs = zeros(arma_part(1)+1,arma_part(1)+1);

      gamma_pacf(1) = wn_var;
      for kk = 2:1:arma_part(2)+1
          for jj = 1:1:min(kk-1,arma_part(1))
          gamma_pacf(kk) = gamma_pacf(kk) + gamma_pacf(kk-jj)*phi(jj);
          end
          gamma_pacf(kk) = gamma_pacf(kk) +  theta(kk-1)*wn_var;
      end
      %rhs
      for ll = 1:1:max(arma_part(1)+1,arma_part(2)+1)
         if ll==1
            rhs(ll) = gamma_pacf(1) + theta*gamma_pacf(2:arma_part(2) + 1)
;
         elseif ll > 1 && ll <= arma_part(2)+1
            rhs(ll) = theta(ll-1:length(theta))*gamma_pacf(1:arma_part(2) +
2-ll) ;
         elseif ll > 1 && ll > arma_part(2)+1
            rhs(ll) =0;
         end
      end
      %lhs
      lhs(1,:) =  [1 -phi];
      lhs(:,1) =  [1 -phi]';
      for ii = 2:1:arma_part(1)+1
          for jj = 2:1:arma_part(1)+1
```

```
            if ii-jj==0 && ii+jj-2 <= arma_part(1)
                phi_ij = 1;
                phi_ij2 = -phi(ii + jj -2);
            elseif ii-jj==0 && ii+jj-2 > arma_part(1)
                phi_ij = 1;
                phi_ij2 = 0;
            elseif ii-jj < 0 && ii+jj-2 <=arma_part(1)
                phi_ij = 0;
                phi_ij2 = -phi(ii + jj -2);
            elseif ii-jj < 0 && ii+jj-2 > arma_part(1)
                phi_ij = 0;
                phi_ij2 =0;
            elseif ii-jj > 0 && ii+jj-2 <= arma_part(1)
                phi_ij = -phi(ii-jj);
                phi_ij2 = -phi(ii+jj-2);
            elseif ii-jj > 0 && ii+jj-2 > arma_part(1)
                phi_ij = -phi(ii-jj);
                phi_ij2 = 0;
            end
                lhs(ii,jj) = phi_ij + phi_ij2 ;
        end
    end
end
```

Source: Inzelt (2011).

7. Computing ARFIMA covariance matrix (Inzelt, 2011)

```
function[gamma_s] = arfima_covs(uptolag,params1,arma_part)
%approximated ACF for ARFIMA(p,d,q) processes
% (c) György Inzelt 2011
%extracting the parameters
wn_var = params1(length(params1));
%muZ = params1(length(params1)-1);
d = params1(1,1);
     if  arma_part(1) == 0 && arma_part(2) == 0, phi =
params1(1,2:2+arma_part(1)-1)    ; theta =0;
       elseif arma_part(1) ==0 && arma_part(2) ~= 0 ,theta =
params1(1,2:2+arma_part(2)-1)   ; phi =0;
       elseif arma_part(1) ~=0 && arma_part(2) ~=0, phi =
params1(1,2:2+arma_part(1)-1);
               theta = params1(1,2+arma_part(1):2+sum(arma_part)-1);
       elseif arma_part(1) ~=0 && arma_part(2)==0, phi =
params1(1,2:2+arma_part(1)-1);theta =0;
     end
gamma_s = zeros(uptolag,1);
switch(sum(arma_part))
    case(0)
%ARFIMA(0,d,0)
```

```
        for h = 0:uptolag-1
            if h < 100
            gamma_s(h+1) =  wn_var*(gamma(1-2*d)*gamma(h+d)/(gamma(1-
d)*gamma(d)*gamma(1+h-d)));
            elseif h>=100
            gamma_s(h+1) = ((h+d)/(1+h-d))*gamma_s(h);
            end
        end
    otherwise
        if arma_part(1) ==0 && arma_part(2)~=0
        %ARFIMA(0,d,q)
        %the MA part
            theta = [1 theta];
            psilag = zeros(2*length(theta)-1,1);
            gamma_term = zeros(uptolag,2*length(theta)-1);
        for ll = -length(theta):1:length(theta)
            for ss = max([0 ll]):min([(length(theta)-1)  (length(theta)-1+
ll)])
                ss_minus_ll = ss - ll;
                if ss_minus_ll < 0, ss_minus_ll = abs(ss_minus_ll) + 1;
                    elseif ss_minus_ll >= 0, ss_minus_ll = ss_minus_ll +1;
                end
                psilag(ll + length(theta))  = psilag(ll + length(theta)) +
theta( ss_minus_ll )*theta(ss+1);
            end
        end
            theta = theta(2:length(theta));
    %the fractional part
        for h = 0:uptolag-1
            if h < 100
                for ll = -length(theta):1:length(theta)
                    gamma_term(h+1,ll + length(theta)+1) = (gamma(1-
2*d)*gamma(d+h-ll))/(gamma(d)*gamma(1-d)*gamma(1+h-d-ll))    ;
                end
                    gamma_s(h+1) =
wn_var*sum(psilag'.*gamma_term(h+1,:),2) ;
            elseif h>=100
                for ll = -length(theta):1:length(theta)
                    gamma_term(h+1,ll + length(theta)+1) =
gamma_term(h,ll+length(theta)+1)*((d+h-ll)/(1+h-d-ll));
                end
                    gamma_s(h+1) =
wn_var*sum(psilag'.*gamma_term(h+1,:),2) ;
            end
        end
    elseif arma_part(1)~=0
        %ARFIMA(p,d,q)
            gamma_s_temp = zeros(uptolag,1);
        for h = 0:uptolag+201
            if h < 100
```

```
            gamma_s_temp(h+1) =  wn_var*(gamma(1-2*d)*gamma(h+d)/(gamma(1-
d)*gamma(d)*gamma(1+h-d)));
            elseif h>=100
            gamma_s_temp(h+1) = ((h+d)/(1+h-d))*gamma_s_temp(h);
            end
        end
        gamma_arma =
arma_covs(201,params1(1,2:length(params1)),arma_part);
        gamma_s_temp =   [gamma_s_temp(201:-1:2) ;
gamma_s_temp(1:length(gamma_s_temp))]   ;

        for k = 0:uptolag-1
          %%approximate algorithm for calculating the ACF of ARFIMA(p,d,q)
          gamma_s(k+1) = (1/wn_var)*sum( [gamma_s_temp(1+k:1+k+400 )
].*[gamma_arma(201:-1:2);gamma_arma(1:201)]);
        end
        end
 end
 end
```

Source: Inzelt (2011).

8. Durbin-Levinson algorithm (Inzelt, 2011)

```
 #include "matrix.h"
 #define square(p) ((p)*(p))
 #define inv(q) (1/(q))
 /* Durbin-Levinson algorithm for linear stationary AR(FI)MA(p,d,q)
processes
    Slightly altered for the maximum likelihood estimation
    (C) György Inzelt 2011
*/
 void levinson_recur1(double* v,double* L, int N, double* gammas,int step)
 {
  int i,k;

  if(step==0)
  {
    *(v + step) = *(gammas + step);
    *(L + step) = 1;
    for(k = step+1;k < N;++k)
    {
      *(L + k) = 0;
    }
  }
  else if(step > 0 && step < N)
  {
     //phi_tt
     *(L + step*N ) =  (-1.00)* *(gammas + step);
```

```
     if(step > 1)
     {
         for(i = 1;i < step ;++i)
         {
           *(L + step*N) -=  *(L + (step-1)*N + (step -1)  - i ) * *(gammas
+ step - i)   ;
         }
     }
      *(L +step*N)  *= inv( *(v + step-1) );
     //v_t
     *(v + step) = *(v + step-1)*(1- square( *(L + step*N) ));
     //phi_tj
     for(i =1; i < step; ++i)
     {
      *(L + step*N + step - i) =  *(L + (step-1)*N + (step -1) - i) + *(L
+ step*N  ) * *(L + (step-1)*N + i -1 ) ;
     }
     //filling L with zeros and ones
     *(L + step*N + step ) = 1;
     if(step != N-1)
     {
         for(k = step*N +step+1 ;k < step*N + N ;++k)
         {
         *(L + k) =0;
         }
     }
  }
    if(step < N-1)
      levinson_recur1(v,L,N,gammas,++step);
 }
 /* The gateway function */
 void mexFunction(int nlhs, mxArray *plhs[], int nrhs, const mxArray
*prhs[])
 {
 int step=0;
 int N;
 double *gammas,*v,*L;
 // getting the autocovariances
 gammas = mxGetPr(prhs[0]);
 N = mxGetM(prhs[0]);
 // v
 plhs[0] = mxCreateDoubleMatrix(0,0,mxREAL);
 mxSetM(plhs[0],N);
 mxSetN(plhs[0],1);
 mxSetData(plhs[0], mxMalloc(sizeof(double)*N*1));
 // L
 plhs[1] = mxCreateDoubleMatrix(0,0,mxREAL);
 mxSetM(plhs[1],square(N));
 mxSetN(plhs[1],1);
 mxSetData(plhs[1], mxMalloc(sizeof(double)*square(N)*1));
```

```
//
v = mxGetPr(plhs[0]);
L = mxGetPr(plhs[1]);
//
levinson_recur1(v, L, N,gammas,step);
//
return;
}
```

Source: Inzelt (2011).