Imperial College London

Department of Mathematics

Financial Volatility Forecasting in Exchange Rate Market

JIEXIU ZHU

CID: 01714151

Supervised by Professor Almut Veraart

30 August 2023

Submitted in partial fulfilment of the requirements for the MSc in Statistics of Imperial College London

The work contained in this thesis is my own work unless otherwise stated.

Signed: JIEXIU ZHU Date: 30 August 2023

Abstract

I provide a thorough analysis on the volatility forecasting in the exchange rate market. High frequency data is used to define realized variance which is the key indicator for market volatility in this project. Theoretical distribution results of quadratic and bipower variation as well as the statistical test for jumps provide foundations and intuitions for forecasting model design and construction. I begin by considering the statistical properties of realized variance. Then four main areas are discussed - the univariate realized variance point forecast, the univariate probabilistic forecast, the multivariate portfolio variance point forecast and the multivariate distributional forecast. I will focus my discussion on constructing and evaluating models which aim to improve the forecasting performances with the empirical exchange market data.

Key words: realized variance forecasting, high-frequency data, exchange rate, quadratic variation, bipower variation, jumps, heterogeneous autoregressive model, probabilistic forecasting, generalised hyperbolic distribution, copula

Acknowledgements

I am profoundly grateful to all those who have supported and guided me throughout this journey of completing my master's thesis. Their encouragement, wisdom, and unwavering belief in my abilities have been invaluable.

I extend my heartfelt gratitude to my supervisor, Prof Almut Veraart , whose guidance and expertise have been instrumental in shaping the course of my research. Your dedication to my academic growth and your insightful feedback have enriched my work beyond measure.

I would also like to express my sincere appreciation to the MSc Statistics Program Director, Dr Oliver Ratmann, for his consistent help and guidance in resolving the scheduling conflicts I have encountered during the research period.

To my loving family, thank you for your unconditional support, understanding, and sacrifices for the past eight years while I am abroad. To my best friends, Yue Sun, Yue Ge and Shuyang Dai, thank you for your company and support throughout my entire master year. And to my dearest roommate and friend, Zelin Xiong, thank you for always being so warm, kind, and understanding. Your presence has made this journey even more meaningful and memorable.

Lastly, I want to extend my gratitude to all the tutors, mentors, and peers who have contributed to my education and personal growth. To all those mentioned above and the countless others who have contributed to my growth, I extend my heartfelt thanks. Your support has been indispensable, and I am honored to have had you by my side on this rewarding endeavor.

Contents

1. Introduction

1.1. Background

In today's intricate landscape of financial markets, the ability to anticipate and manage risk is essential. Volatility, the key indicator of market dynamics which is a measure of uncertainty and risk in the markets, therefore plays an essential role in determining investment strategies, portfolio allocations, and risk management techniques. Among all asset classes, the exchange rate market stands as a dynamic arena where currencies fluctuate in response to an intricate interplay of economic, geopolitical, and market-specific factors. As a result, accurate forecasting of exchange rate volatility has garnered significant attention from researchers, practitioners, and policymakers alike. Over the past 50 years, researchers and financial institutions have been trying to study the pattern of volatility and design models which can forecast its movement accurately. The parametric GARCH model proposed by Engle (1982) is one of the first few research to model time-varying volatility. However, in recent two decades, one approach that has emerged as a robust tool in volatility measurement and prediction involves the concept of realized variance.

Realized variance, a measure of intra-day price fluctuations, captures the true underlying volatility of an asset constructed from high-frequency data. This is different from the traditional methods which use daily closing prices and therefore often result in incomplete representation of the dynamic volatility patterns within financial markets. The application of realized variance has opened new avenues for exploring volatility dynamics, and now it has already become the most common representation for market dynamic. Daily realized variance then start to become the object that researchers aim to model and forecast. The Heterogeneous Autoregressive model of Realized Variance proposed by Corsi (2009) has become one of the most popular models for realized variance forecasting. Many attempts to adjust and tune the model are made all aiming to improve its forecasting performance. Bollerslev et al. (2020) has proposed the idea of realized semivariance where the realized variance could be split into positive and negative components corresponding to the 'good' and 'bad' risks. This measure was been proven to work effectively among stock and index markets, but further exploration of such method on different asset classes is still needed.

From a theoretical perspective, the analysis of volatility can be perceived within the probability and statistical framework with the support of Itô semimartingale assumption and the theory of quadratic variation. Then the asymptotic distributions of realized variance and other relevant indicators were discovered in Barndorff-Nielsen and Shephard (2004a), and those results can all be extended from the univariate case to multivariate

setting. At the same time, those theoretical results reveal the existence of continuous and discontinuous components in the asymptotic distributions of realized variance, which is also reflected in the empirical market data. The existing literature has mainly focus on the continuous part as it represents the natural dynamics of the asset market without any big noises caused by external factors. However, it is still important to perceive and model realized variance as a whole since in practical application, the entire picture of realized variance is the object of interest for investors. Andersen et al. (2007) has acknowledged such need and proposed to add a 'jump' variable to the Heterogeneous Autoregressive model for Realized Variance. In our project, we will assess a different approach which separately model the different components of realized variance.

In addition to point forecast, probabilistic forecast for realized variance seems more applicable in monitoring Value-at-Risk and expected shortfall levels. These are essential steps in financial risk management (Timmermann, 2000). The existing methods of probabilistic forecasting on volatility includes the classical Bayesian inferential approach (Maneesoonthorn et al., 2012) and more recently, the machine learning approaches with long short-term memory (LSTM) models (Thavaneswaran et al., 2022). Both methods will be considered as computationally expensive in practice. In this project, we will experiment on an alternative probabilistic forecaster which is simpler and more easily implemented.

1.2. Project Aim

The general aim of this project is to explore the possible methods to improve the forecasting performance for realized variance in the exchange rate market. Four main areas are studied and discussed - the univariate realized variance point forecast, the univariate probabilistic forecast, the multivariate portfolio variance point forecast and the multivariate distributional forecast.

We first research into the properties of realized variance of exchange rate pairs, including the popular financial stylized facts, theoretical asymptotic distributions and relevant statistical tests. Motivated by the theoretical results, we will propose two main techniques to improve the forecasting accuracy of an existing volatility forecast model - semivariance measures and continuous-discontinuous components separation. Several models are proposed while implementing the methods, and these models are tested and assessed to evaluate which framework is the most effective one. Forecast aggregation of these proposed models is also demonstrated to see if there are any further improvements. Besides point forecast, we will also investigate the possible model improvements in the probabilistic forecast. Then we will move on to the multivariate setting where the concept of portfolio variance is introduced. To improve the portfolio variance forecast, multivariate models which account for the inter-correlation between variables are carefully compared with the univariate model to find the optimum method. Finally, we briefly discuss the methods and the challenges for multivariate probabilistic forecasts.

We will end the first chapter with the introduction of basic notations which will be

consistently used throughout the report as well as the formal definitions for realized variance in financial mathematics. Chapter 2 studies the statistical properties and significant theoretical results regarding the realized variance which will later serve as the important intuition for model design. Chapter 3 contains the main contribution of this project where various methods are developed to improve the forecasting performance in the univariate point and probabilistic forecast. Chapter 4 further extends our discussion and methods to bivariate setting where portfolio variance forecast and multivariate distributional forecast are performed and evaluated.

This project is programmed with R and all the relevant code is stored in the GitHub repository (`https://github.com/jz4919/MSc_Research_Project`). The following R packages are loaded: `dplyr`, `magrittr`, `tsibble`, `forecast`, `ggplot2`, `opera`, `ghyp`, `copula`, `VineCopula`.

1.3. Basic Notations and Definitions

Let Y_t be the logarithmic price of a single asset for continuous time $t \geq 0$, and it is assumed that this price process can be described by an Itô semimartingale as the form below:

$$Y_t = \int_0^t \mu_s \, \mathrm{d}s + \int_0^t \sigma_s \, \mathrm{d}W_s + J_t \quad t \geq 0 \tag{1.1}$$

(Ghysels et al., 1996) where μ_s represents the drift, σ_s corresponds to the diffusive volatility processes, W_s represents a standard Brownian motion and J_t is a discontinuous jump process. We will be mainly focusing on analysing the last two components of Equation 1.1 which gives the continuous and discontinuous components of the stochastic volatility respectively.

Definition 1.3.1. The **quadratic variation (QV)** process of Y is defined as

$$[Y]_t = \mathrm{p} - \lim_{n \to \infty} \sum_{j=1}^{n} (Y_{t_j} - Y_{t_{j-1}})^2$$

(Jacod, 1987) for any sequence of partitions $t_0 = 0 < t_1 < \cdots < t_n = t$ with $\sup_j \{t_j - t_{j-1}\} \to 0$ for $n \to \infty$ ($\mathrm{p} - \lim$ represents the probability limit). Furthermore, with the assumption of Y_t established in Equation 1.1, QV can alternatively be written as

$$[Y]_t = \int_0^t \sigma_s^2 \, \mathrm{d}s + \sum_{0 \leq s \leq t} (\Delta J_s)^2 \tag{1.2}$$

(Barndorff-Nielsen and Shephard, 2004a) where ΔJ_s for $0 \leq s \leq t$ captures the discontinuous jumps of the price process happened between time 0 and time t. Note that if there are no jumps, we simply have the QV being equal to the integrated diffusive/ continuous volatility.

To estimate this QV process, we use the high-frequency intra-day prices [1] at constant intervals over a unit day. In this project, we observe the currency rate at 5-min interval over each day where all the data are extracted from the open resource TRUEFX (`https://www.truefx.com/`). Let $Y_{j,i}$ denotes the observed asset price at j-th time interval on day i, we have the **logarithmic discrete-time return of δ-interval** ($\delta = 1/288$ given the 5-min interval in our context) at j-th time interval on day i as

$$y_{j,i} = Y_{j\delta,i} - Y_{(j-1)\delta,i}\,,\ j = 1, 2, \cdots, [1/\delta]$$

Then we have the following formal definition for asset realized variance:

Definition 1.3.2. The daily **realized variance (RV)** of the asset price on day i is defined as the summation of high-frequency squared returns within day i:

$$RV_i = \sum_{j=1}^{[1/\delta]} (Y_{\delta j,i} - Y_{\delta(j-1),i})^2 = \sum_{j=1}^{[1/\delta]} y_{j,i}^2$$

and the daily **realized volatility** on day i is simply denoted by $\sqrt{RV_i}$.

With the definition for daily RV_i, we can extend to define weekly and monthly realized variance by accounting for the average variance performance over a period. Since there are on average 5 business days per week and 22 business days per month, on day i, we denote the weekly and monthly realized variance as $RV_{i:i-4}$ and $RV_{i:i-21}$ respectively where

$$RV_{i:i-t} = \frac{1}{t+1} \sum_{m=0}^{t} RV_{i-m} \tag{1.3}$$

[1] the high-frequency price is calculated by taking mean of the bid and ask price at each time point

2. Stylized Facts and Asymptotic Properties

2.1. Volatility Stylized facts

We will start to understand some significant properties of realized variance/ volatility by first introducing the basic stylized empirical facts that summarize the general statistical properties regarding the realized volatility of asset classes. The two most well-known stylized fact of the realized volatility is volatility clustering and persistent auto-correlation (Cont, 2001).

Volatility clustering is a famous phenomenon often observed among the realized volatility trend of most asset classes. It suggests that a large/small price change is more likely to be followed by another large/small price change. Figure 2.1 below helps to illustrate this stylized fact as it plots the high-frequency log return y_t for EURUSD from 2022 Nov 14-16. We can see clearly from the plot that the red circled region is the high volatility cluster where price changes drastically, and the green circled area shows the small volatility region.

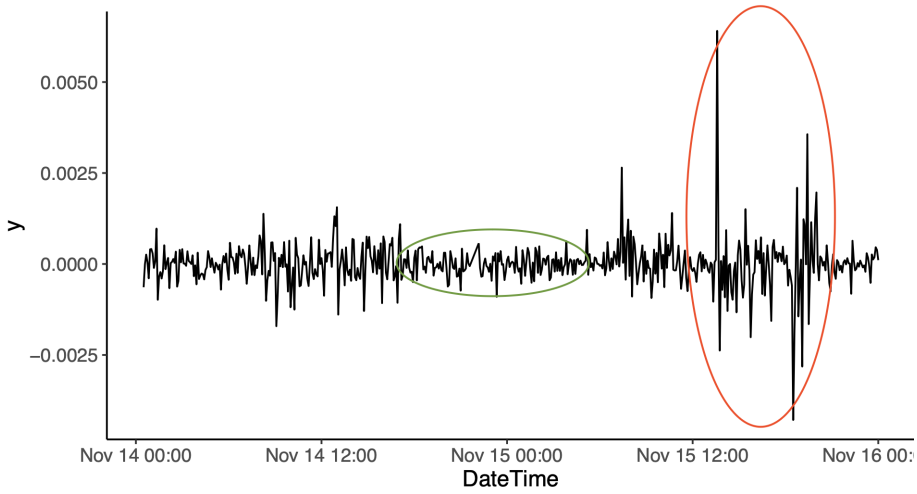


Figure 2.1.: Log return y_t of EURUSD from 2022 Nov 14 - 16

Another very important stylized fact is volatility persistence. Unlike the asset returns whose autocorrelation often decays rapidly (Cont, 2001), the autocorrelation of volatility demonstrates certain level of significance in many scenarios. We can see this phenomenon from Figure 2.2 which plots the autocorrelation function of the daily real-

ized volatility of EURUSD from 2022 May - 2023 Apr. The ACF plot clearly shows that the autocorrelation of the exchange rate is still statistically significant even at lag k = 30, highlighting its persistence.

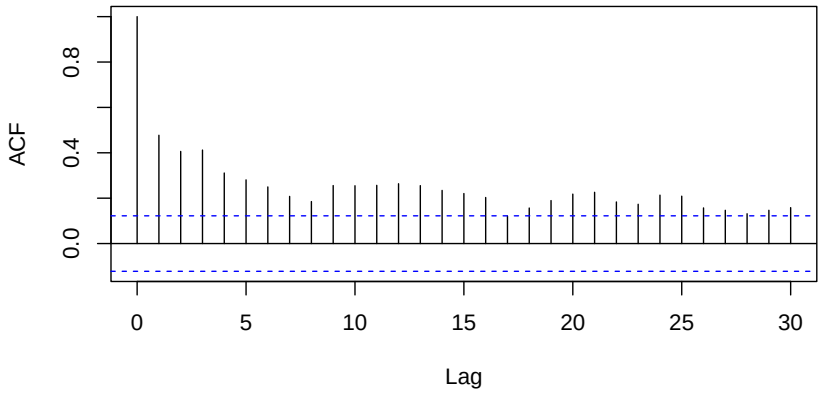


Figure 2.2.: ACF plot of daily realized volatility of EURUSD

2.2. Asymptotic Properties

Here, we introduce a few very important asymptotic distributions of realized variance and realized bipower variation. Recall from Definition 1.3.2 that the daily realized variance is the summation of the high-frequency log returns. Followed from the theory of quadratic variation for semimartingales (Barndorff-Nielsen and Shephard, 2002), for $\delta \downarrow 0$, we have:

$$RV_i \equiv \sum_{j=1}^{[1/\delta]} y_{j,i}^2 \xrightarrow{\mathbb{P}} \int_i^{i+1} \sigma_s^2 ds + \sum_{i \le t \le i+1} (\Delta J_t)^2 \tag{2.1}$$

which implies that the realized daily variance converges in probability to the sum of the 'continuous' diffusive volatility integral and the increments of the squared 'discontinuous' price jumps over the day.

This asymptotic distribution gives one the intuition to separate the diffusive volatility and jump components of $[Y]_t$. Then we introduce the concept of bipower variation proposed by Barndorff-Nielsen and Shephard (2004b).

Definition 2.2.1. The 1-1 order **bipower variation** is defined as

$$\{Y\}_t^{[1,1]} = \mathrm{p} - \lim_{\delta \downarrow 0} \sum_{j=2}^{[1/\delta]} |y_{j-1}||y_j|$$

with Y_t following Equation 1.1, and further assuming that $\mu \equiv 0$ and σ being independent of W, we have

$$\{Y\}_t^{[1,1]} = \mu_1^2 \int_0^t \sigma_s^2 \, \mathrm{d}s$$

(Barndorff-Nielsen and Shephard, 2004b) where $\mu_1 = \sqrt{2}/\sqrt{\pi}$.

Remark 2.2.2. Comparing Definition 2.2.1 and 1.3.1, it is found that one $|y_j|$ in quadratic variation has been replaced by $|y_{j-1}|$ to form the bipower variation. The intuition behind such a definition is the assumption that if there is a jump at time $t-1$, then with $\delta \downarrow 0$, there is unlikely to be a consecutive jump at the next consecutive time-point t (Christensen, 2016). Therefore, each jump term is paired with a consecutive continuous return in the formation of bipower variation, bringing down the effect of jumps and making bipower variation jump-robust.

Similar to how we estimate QV, we have the realized BPV as follows:

Definition 2.2.3. The daily **realized bipower variation (BPV)** on day i is:

$$BPV_i = \sum_{j=2}^{[1/\delta]} |y_{j-1,i}||y_{j,i}|$$

It can be shown that when $\delta \downarrow 0$, the realized BPV_i converges uniformly in probability to the continuous component of QV_i multiplied by a constant (Barndorff-Nielsen and Shephard, 2004b)

$$BPV_i \xrightarrow{\mathbb{P}} \mu_1^2 \int_i^{i+1} \sigma_s^2 \, \mathrm{d}s \quad \text{where} \quad \mu_1 = \sqrt{\frac{2}{\pi}} \tag{2.2}$$

Combining Equation 2.1 and 2.2, Barndorff-Nielsen and Shephard (2004b) pointed out that when $\delta \downarrow 0$

$$RV_i - \mu_1^{-2} BPV_i \xrightarrow{\mathbb{P}} \sum_{i \le t \le i+1} (\Delta J_t)^2 \tag{2.3}$$

suggesting that the increments of the jump components on the trading can be consistently estimated by the difference between realized RV and standardized realized BPV.

Equation 2.1, 2.2, 2.3 are three very important asymptotic distributions which will serve as one of the key intuitions behind our forecast model design and construction. In the following subsection, we will look deeper into how one can identify the occurrence of jumps with statistical testing.

2.3. Adjusted Ratio Jump Test

Recall from Equation 1.1 and 1.2 that the jump component is a significant part of the pricing and quadratic variation process. A price jump is distinctive from the natural diffusive volatility and can usually be observed directly from the high-frequency time series plot. We can see such an example in Figure 2.3 which shows the high-frequency EURUSD rate on 2022 Nov 10. From the plot, one can clearly see a sharp positive jump in the exchange rate on Nov 10 around 12:30 pm GMT.

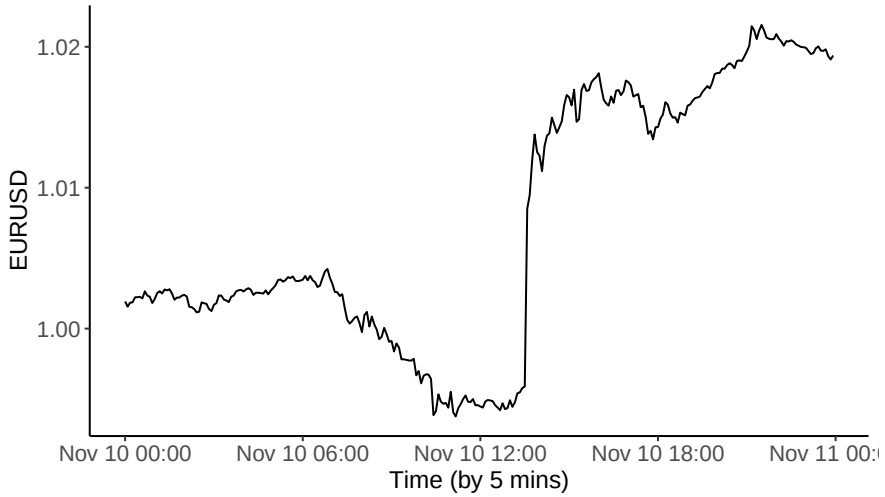


Figure 2.3.: EURUSD rate on 2022 Nov 10

Under this section, we perform the adjusted ratio test proposed by Barndorff-Nielsen and Shephard (2006a) on the EURUSD exchange rate to test the existence of jumps over the 257 trading days of the past year. This test allows one to determine if a jump has occurred on a specific trading day with the null hypothesis assuming that there is no jump. In the original paper, Barndorff-Nielsen and Shephard (2006a) derived the asymptotic distribution for linear jump statistic, ratio jump statistic and adjusted ratio jump statistic. Later it is found that the adjusted ratio jump statistic, Z, gives the best performance in both simulation and case studies. Therefore, in our experiment, we will directly use the adjusted ratio jump statistic to perform the test. We first introduce the daily realized quadpower variation which is an important component used to compute the daily jump test statistic:

Definition 2.3.1. The daily **realized quadpower variation (QPV)** of Y on day i is

$$QPV_i = \delta^{-1} \sum_{j=4}^{[1/\delta]} |y_{j-3,i}| \, |y_{j-2,i}| \, |y_{j-1,i}| \, |y_{j,i}| \quad \text{where} \quad \delta = \frac{1}{288}$$

and under Assumption 1.1, this realized quadpower variation converges in probability to the integrated quarticity multiplied by a constant

$$QPV_i \xrightarrow{\mathbb{P}} \mu_1^4 \int_{i-1}^{i} \sigma_s^4 \, \mathrm{d}s = \{Y\}_t^{[1,1,1,1]}, \quad \mu_1 = \sqrt{\frac{2}{\pi}} \tag{2.4}$$

Remark 2.3.2. The integrated quarticity $\int_{i-1}^{i} \sigma_s^4 \, \mathrm{d}s$ determines the asymptotic distribution for the ratio jump statistic $\mu_1^{-2}\{Y\}_t^{[1,1]}/[Y]_t$ (Barndorff-Nielsen and Shephard, 2006a). To make the statistical test feasible, one uses realized RV, BPV and QPV to approximate $[Y]_t$, $\{Y\}_t^{[1,1]}$ and $\{Y\}_t^{[1,1,1,1]}$ respectively. The probability convergence in

Equation 2.4 highlights that QPV_i is also robust to jumps just like BPV_i. The jump-robustness of quadpower variation is achieved by taking the products of four consecutive log returns. The intuition behind such definition is similar to Remark 2.2.2.

Then the daily adjusted ratio test statistic of day i is computed as below:

$$Z_i = \frac{\delta^{-1/2}}{\sqrt{\vartheta \max\left(1, \frac{QPV_i}{BPV_i^2}\right)}} \left(\frac{\mu_1^{-2} BPV_i}{RV_i} - 1\right) \quad \text{where} \quad \vartheta = \frac{\pi^2}{4} + \pi - 5\,,\ \mu_1 = \sqrt{\frac{2}{\pi}} \tag{2.5}$$

and $Z_i \xrightarrow{L} N(0,1)$ as $\delta \downarrow 0$ (Barndorff-Nielsen and Shephard, 2006a).

The jump statistic Z_i is computed for all 257 trading days, and the rejection rate at 5% level and 1% level are 0.2568 and 0.1478 respectively. These rejection rates are relatively high suggesting that there are quite a number of trading days where jumps were observed over the past year. This is somewhat under our expecation as the exchange rate market has indeed been quite volatile for the past year due to several rate hikes by the Federal Reserve and the European Central Bank. The result is also illustrated by Figure 2.4 where the ratio statistic $\frac{\mu_1^{-2} BPV_i}{RV_i}$ ($\times$ marker) and the respective 1% significance level (blue line) are plotted. We can see from the figure that, on some days there are really significantly small ratio statistics suggesting a very high probability of jump occurrence.

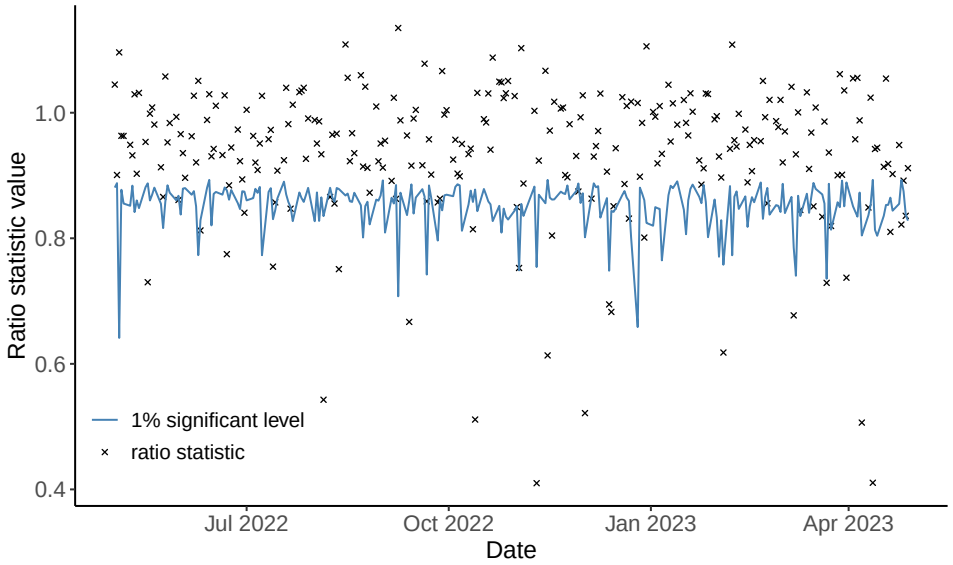


Figure 2.4.: Ratio statistic and corresponding adjusted 1% significance levels for EURUSD

To briefly investigate the possible cause of jumps, we selected 6 extreme days where we observe both a large value of RV_i ($> 1\text{e}-4$) and a small jump statistic (rejected at 0.1% level). Figure 2.5 plots the log return and the ratio statistic graphs side by

side for the 6 chosen days. For example, zooming into the left two plots on the first row which give information from 2022 July 12-14, we can see the log return plot shows a big price jump around 12:30 pm GMT on July 13, and the ratio statistic on that day is also well below the 1% level while those of the other two days are above the critical level. A similar pattern can be found in all other 5 days, indicating that this adjusted ratio jump statistic is relatively reliable in capturing the occurrence of jumps. By checking the business calendar for these 6 days, we found that all the price jumps are closely linked to macroeconomic news announcements such as inflation rate and GDP announcements. Table 2.1 summarises the statistical information of those 6 days and identifies the corresponding market news related to the price jumps.

In general, we see that most price jumps are induced by big macroeconomic news announcements. From the log return plots in Figure 2.5, it can be observed that such price jumps seem to restore quickly and has very little persistent effect on the log return even in high-frequency data. Therefore, we can see that the properties of the jumps can be quite different from the properties of the continuous volatility.

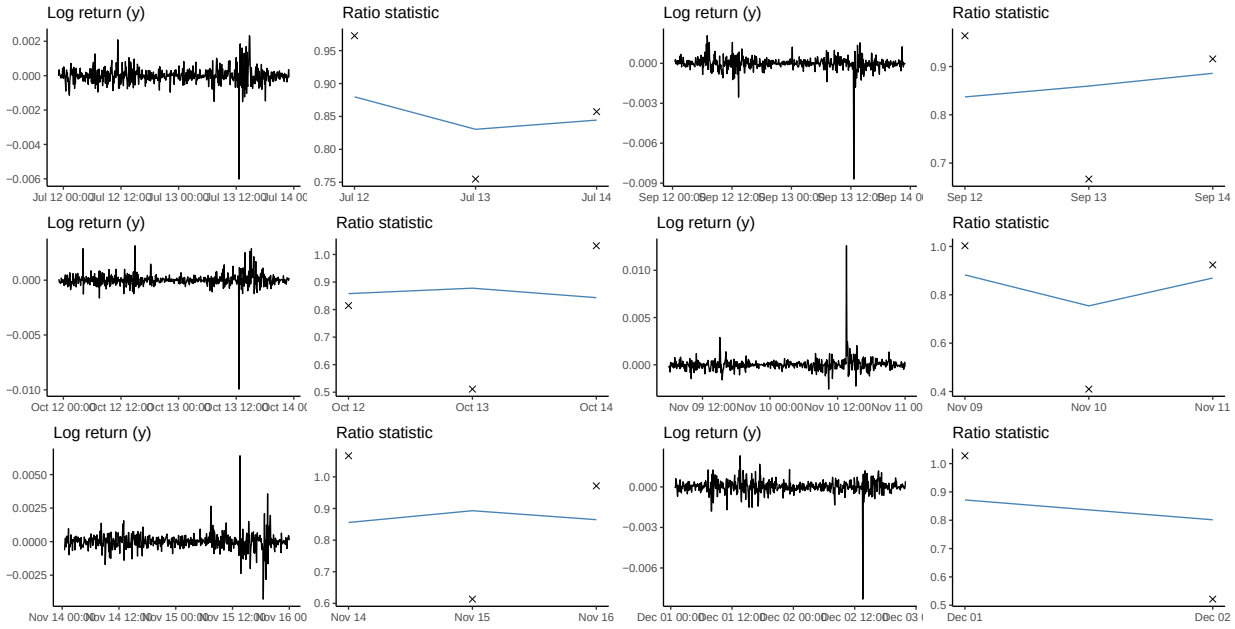


Figure 2.5.: Log-return plots of 6 days with significant jump in EURUSD and the corresponding ratio statistic plots

Date	RV_i	Z_i	Market event
2022-07-13	1.0626e-04	-3.3640	US CPI[1] announcement for Jun 2022
2022-09-13	1.2691e-04	-5.5173	US CPI announcement for Aug 2022
2022-10-13	1.7410e-04	-9.2907	US CPI announcement for Sep 2022
2022-11-10	2.4141e-04	-5.5850	US CPI announcement for Oct 2022
2022-11-15	1.6667e-04	-8.4062	EU GDP[2], US PPI[3] announcement for Oct 2022
2022-12-02	1.0439e-04	-5.6110	US Employment Rate announcement for November 2022

Table 2.1.: Summary of information on the 'extreme' days with big jumps

[1]Consumer Price Index - important indicator for inflation rate
[2]Gross Domestic Product - important indicator for economic growth
[3]Producer Price Index - important indicator for inflation rate

3. Univariate Variance Forecast

This chapter marks the main contribution of this article where we design and construct models to improve the forecasting performance on univariate volatility. The chapter starts with point forecast models and then moves on to discuss probabilistic forecasting. We will start with the standard heterogeneous autoregressive (HAR) model (Corsi, 2009), and then further develop the model with various techniques such as semivariance measures and diffusive-jump separation. The performance of all models/ techniques will be summarized and discussed together in Subsection 3.1.5. Here, we use the EURUSD exchange rate of 257 trading days from 2022 May to 2023 April throughout all model fitting and evaluation.

3.1. Realized Variance Point Forecast

3.1.1. Exploratory analysis

With the high-frequency exchange rate data, one could easily compute the daily realized variance with Definition 1.3.2. However, before any model fitting on RV_t, it is first worth exploring the data and considering if there is any need for transformation on the raw realized variance. The HAR model which will be applied later is a linear model that assumes the data to follow a normal distribution with constant variance. However, the raw realized variance data do not usually exhibit such nice statistical properties. Therefore, some suitable transformations could be applied to the data to help approximate a more normal distribution and stabilize the variance, allowing for better model fitting.

To understand which transformation better approximates the normal distribution, we perform the Box-Cox transformation experiment on the raw realized variance and the result is shown in Figure 3.1. Box and Cox (1964) proposed the following power transformation function which helps to correct asymmetry and non-linearity in variables and is very commonly used in data transformation context:

$$y^{(\lambda)} = \begin{cases} \frac{y^{\lambda}-1}{\lambda} & \lambda \neq 0 \\ \log(y) & \lambda = 0 \end{cases} . \tag{3.1}$$

From Figure 3.1, we can see that the 95% level of the normality test restricts the value of λ to fall between around -0.4 and 0, suggesting that the logarithmic transformation might be a suitable transformation for RV_t. Figure 3.2 gives three Q-Q plots for RV_t, $\sqrt{RV_t}$ and $\log(RV_t)$ respectively, which are the three most commonly used forms for financial market data analysis. From the subplots, we can see that $\log(RV_t)$ demonstrates the

least deviations from the theoretical normal distribution, while the other two Q-Q plots show great asymmetry and heavy tails on the right ends. This further confirms our choice of logarithmic transformation. $\log(RV_t)$ will then consistently be used for model fitting and evaluation in the rest part of this chapter.

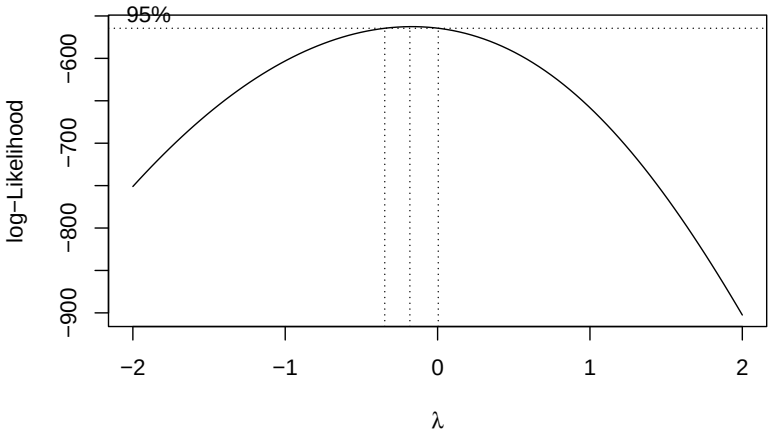


Figure 3.1.: Box-Cox normality plot

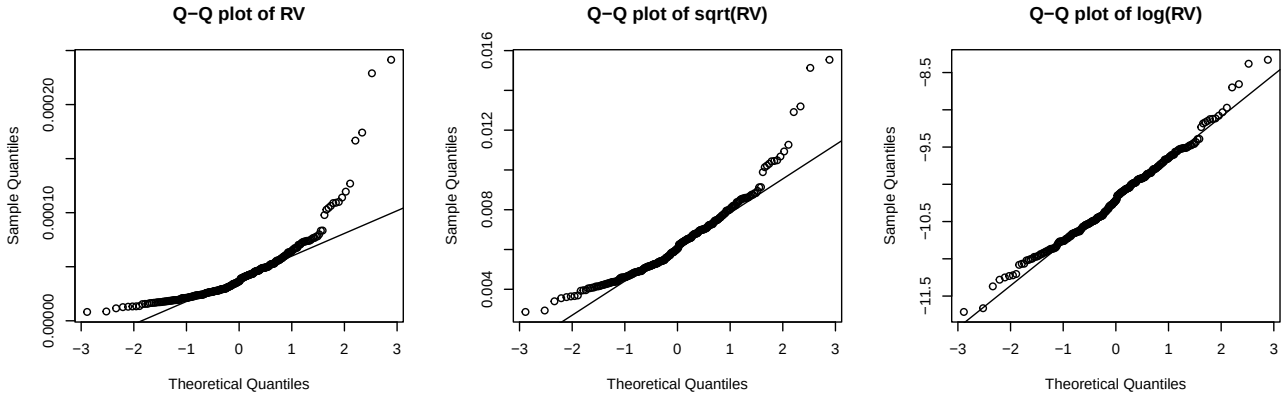


Figure 3.2.: Q-Q plot of three transformations of RV_t

3.1.2. HAR model

We will start our model fitting with the standard heterogeneous autoregressive (HAR) model for realized variance proposed in Corsi (2009). The model can be used to fit the $\log RV$ data to form the **HAR_RV** model as follows:

$$\log RV_{t+1} = \phi_0 + \phi^{(d)} \log RV_t + \phi^{(w)} \log RV_{t:t-4} + \phi^{(m)} \log RV_{t:t-21} \tag{3.2}$$

It is a simple linear model stating that the value of RV on the next day is dependent on the daily, weekly and monthly RV on today (recall the definition of $RV_{i:i-t}$ in Equation 1.3). This model considers the volatility persistence by taking the weekly and monthly RV into account. We first fit this model with all the trading days data, and the estimated coefficients and their respective p-values are shown in Table 3.1. It can be observed from the table that $\phi^{(d)}$ has the highest statistical significance suggesting that the previous day RV plays a great part in determining the next day RV. One can also read the value for adjusted R^2 of this fitted model as 0.3709.

ϕ_0	$\phi^{(d)}$	$\phi^{(w)}$	$\phi^{(m)}$	adj.R^2
-1.2995 (0.1581)	0.2997 (0.0001)	0.2965 (0.0237)	0.2769 (0.0537)	0.3709

Table 3.1.: Estimated coefficient (p-value) for each predictor for HAR_RV

The fitted model is then used to perform the in-sample forecasting by evaluating the model on the original whole dataset and computing the in-sample predictions. Figure 3.3 below shows the predicted values (blue line) together with the respective true log RV_t. It can be seen from the plot that that the model generally captures the overall movements of the variance, however, it also fails to model most of the extreme values (spikes of the grey line) of log RV_t.

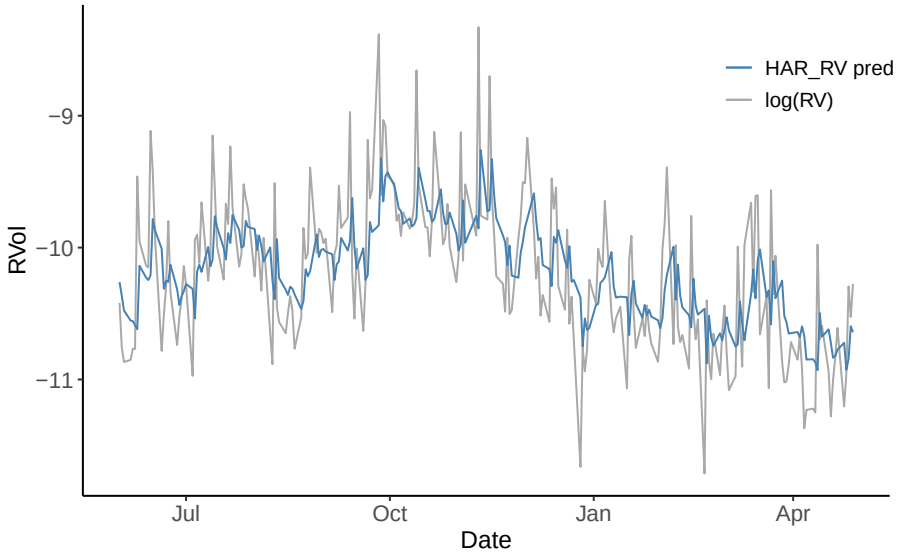


Figure 3.3.: RV_HAR in-sample forecast on EURUSD from 2022 May - 2023 Apr

This RV_HAR model will serve as the base and the benchmark model, and in the following subsections, several techniques will be explored and modifications will be made to the RV_HAR model to see if those proposed methods actually demonstrate an improvement in the forecasting performance compared to RV_HAR.

3.1.3. Semivariance measures

The first technique introduced here is the semivariance measures. The intra-day high-frequency price changes rapidly and the sign of the return $y_{j,i}$ also alters actively. Definition 1.3.2 of realized variance tends to summarize the overall variation performance of the price over the day while failing to account for the signs of $y_{j,i}$. However, it has been long proposed by many scholars that different signs of returns should be treated differently - both Roy (1952) and Markowitz (1970 - 1959) pointed out that the investors tend to be more concerned with the negative returns as such downside risk is closely related to loss incurred. The idea of separating the positive and negative returns was first proposed by Barndorff-Nielsen et al. (2010) which splits the realised variance into two components (also known as semivariance measures):

$$RV_t^+ = \sum_{j=1}^{[1/\delta]} y_{j,t}^2 \mathbb{I}_{\{y_{j,t} \geq 0\}} \ , \quad RV_t^- = \sum_{j=1}^{[1/\delta]} y_{j,t}^2 \mathbb{I}_{\{y_{j,t} < 0\}}$$

From the above definition, one can clearly see that summing up RV_t^+ and RV_t^- gives the value of the original RV_t. Figure 3.4 below plots the daily RV_t^+ and RV_t^- for EURUSD for the past year. We can see from the plot that on most of the days, RV_t^+ and RV_t^- have similar levels of values.

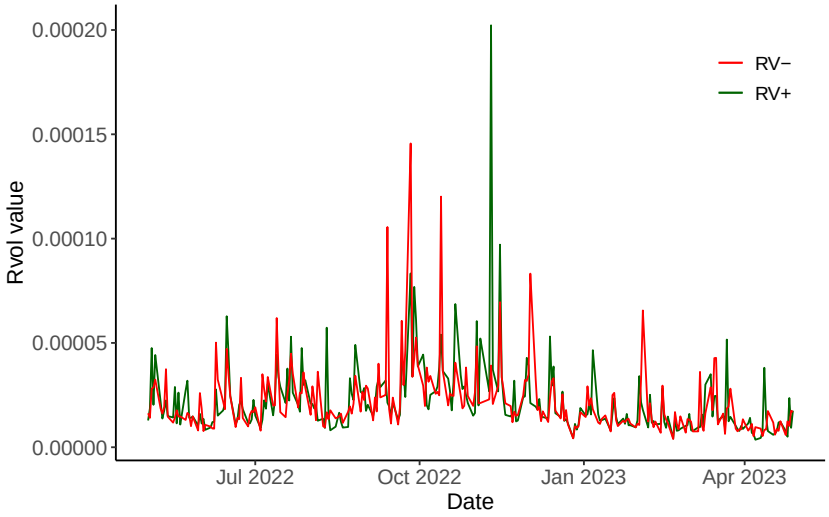


Figure 3.4.: Daily realized semivariance measures of EURUSD, 2022 May - 2023 Apr

The semivariance measures can be implemented to modify the RV_HAR model and to form RV_SemiHAR (**RV_SHAR**) model shown in Equation 3.3.

$$\log RV_{t+1} = \phi_0 + \phi^{(d+)} \log RV_t^+ + \phi^{(d-)} \log RV_t^- + \phi^{(w)} \log RV_{t:t-4} + \phi^{(m)} \log RV_{t:t-21} \tag{3.3}$$

Bollerslev et al. (2020) has tested and evaluated this modified SHAR model with stock index data and confirmed that SHAR had a better fitting and forecasting performance

than the original HAR model. However, so far very few experiments have been seen to test this practice on the exchange rate data. Since different asset classes often demonstrate very different price and volatility patterns due to the difference in the market nature, here we will fit and evaluate the SHAR model on our exchange rate EURUSD to see if this practice also has a better performance under our setting.

Model 3.3 was fitted with all the trading days data, and the estimated coefficients and their respective p-values are shown in Table 3.2 below. Through comparing the value and p-value for each coefficient, so far there is no clear sign that $\phi^{(d-)}$ plays a more important part in determining $\log RV_t$. However, the adjusted R^2 value for SHAR_RV model is slightly higher than that of HAR_RV. Further evaluation of the in-sample and out-of-sample prediction performance will be discussed in detail in Section 3.1.5.

ϕ_0	$\phi^{(d+)}$	$\phi^{(d-)}$	$\phi^{(w)}$	$\phi^{(m)}$	adj.R^2
-1.1341	0.2601	0.0731	0.2646	0.2680	0.3785
(0.2180)	(0.0009)	(0.3460)	(0.0450)	(0.0604)	

Table 3.2.: Estimated coefficient (p-value) for each predictor for SHAR_RV

3.1.4. Diffusive-jump separation

This technique is motivated by the asymptotic properties introduced in Section 2.2 in the previous chapter. The asymptotic distribution of RV_t is the summation of a continuous (diffusive) volatility integral C_t and discontinuous jump components D_t. It is natural to consider modelling the two components separately and later combining them to compute the final RV_t. There are several ways to separate the continuous and discontinuous parts of RV_t, and in this paper, we have proposed three separation methods.

Constant threshold

The most naive way of separation would be using a constant threshold α - where any high-frequency return whose absolute value is greater than α is considered as a 'jump' while the rest are considered as diffusive volatility. C_t and D_t are formally written as the equations below:

$$C_t = RV_t^{(c)} = \sum_j y_{j,t}^2 \mathbb{I}_{\{|y_{j,t}| \leq \alpha\}}$$

$$D_t = RV_t^{(d)} = \sum_j y_{j,t}^2 \mathbb{I}_{\{|y_{j,t}| > \alpha\}}$$

Then we have the following models together named as **HAR_RVc+RVd**:

$$\log RV_{t+1}^{(c)} = \phi_0 + \phi^{(d)} \log RV_t^{(c)} + \phi^{(w)} \log RV_{t:t-4}^{(c)} + \phi^{(m)} \log RV_{t:t-21}^{(c)} \tag{3.4}$$

$$\log RV_{t+1}^{(d)} = \beta_0 + \beta^{(d)} \log RV_t^{(d)} + \beta^{(w)} \log RV_{t:t-4}^{(d)} + \beta^{(m)} \log RV_{t:t-21}^{(d)} \tag{3.5}$$

where we fitted the HAR model (Equation 3.2) to $\log RV^{(c)}$ and $\log RV^{(d)}$ separately. The final predicted value for $\log RV$ can be back calculated as below:

$$\log RV = \log(\exp(\log RV^{(c)}) + \exp(\log RV^{(d)})) \tag{3.6}$$

The semivariance measures can also be applied in this case to get **SHAR_RVc+RVd**:

$$\log RV_{t+1}^{(c)} = \phi_0 + \phi^{(d+)} \log RV_t^{(c+)} + \phi^{(d-)} \log RV_t^{(c-)} + \phi^{(w)} \log RV_{t:t-4}^{(c)} + \phi^{(m)} \log RV_{t:t-21}^{(c)} \tag{3.7}$$

$$\log RV_{t+1}^{(d)} = \beta_0 + \beta^{(d+)} \log RV_t^{(d+)} + \beta^{(d-)} \log RV_t^{(d-)} + \beta^{(w)} \log RV_{t:t-4}^{(d)} + \beta^{(m)} \log RV_{t:t-21}^{(d)} \tag{3.8}$$

With a few experiments trying different levels of thresholds, it is found that by setting $\alpha = 0.002$, the model gives the best result in terms of mean squared error. Hence, $\alpha = 0.002$ will be used throughout. Table 3.3 and 3.4 below present the estimated coefficients for Equation 3.4 - 3.8. It can be seen from the tables that the adjusted R^2 has a great improvement for $\log RV_{t+1}^{(c)}$ (as compared to the adjusted R^2 when the HAR model is fitted to RV_{t+1} as a whole), and the $\phi^{(d)}$ has become even more significant especially shown in HAR_RVc+RVd. However, it is also noticed that the adjusted R^2 for $\log RV_{t+1}^{(d)}$ is very low, suggesting a relatively poor fit for the jump component.

Model	Fitted term	ϕ_0/ β_0	$\phi^{(d)}$/ $\beta^{(d)}$	$\phi^{(w)}$/ $\beta^{(w)}$	$\phi^{(m)}$/ $\beta^{(m)}$	adj.R^2
HAR_RVc+RVd	$\log RV_{t+1}^{(c)}$	-0.9772	0.4332	0.2669	0.2061	0.5215
		(0.1893)	(2.73e-08)	(0.0260)	(0.0968)	
	$\log RV_{t+1}^{(d)}$	-5.5909	0.1788	0.0617	0.3197	0.0554
		(0.0197)	(0.0175)	(0.6904)	(0.1622)	
HAR_RVc+dRVd	$\log RV_{t+1}^{(d)}$	-9.8024	0.2295	-	-	0.0486
		(2e-16)	(0.0004)			

Table 3.3.: Estimated coefficient (p-value) for each predictor for HAR_RVc+RVd and HAR_RVc+dRVd

In addition, it is worth noting that the discontinuous component usually demonstrates a very different property from the continuous volatility. This was already noted in the case study in Section 2.3 where we see the big 'jumps' observed in Figure 2.5 usually get absorbed very quickly and show no sign of any persistent impact. This seems to be different from the volatility persistence we have discussed. To further see this difference, we have plotted the ACF for $RV_t^{(c)}$ and $RV_t^{(d)}$ separately shown in Figure 3.5 below. It

Fitted term	ϕ_0/ β_0	$\phi^{(d+)}$/ $\beta^{(d+)}$	$\phi^{(d-)}$/ $\beta^{(d-)}$	$\phi^{(w)}$/ $\beta^{(w)}$	$\phi^{(m)}$/ $\beta^{(m)}$	adj.R^2
$\log RV_{t+1}^{(c)}$	-0.6986	0.2406	0.1968	0.2575	0.2084	0.5191
	(0.3519)	(0.0088)	(0.0169)	(0.0357)	(0.0945)	
$\log RV_{t+1}^{(d)}$	-4.3353	0.1991	0.0853	0.0471	0.3178	0.0579
	(0.0785)	(0.0239)	(0.3247)	(0.7606)	(0.1646)	

Table 3.4.: Estimated coefficient (p-value) for each predictor for SHAR_RVc+RVd

can be easily observed that the two ACF plots demonstrate very different traits. The ACF for $RV_t^{(c)}$ presents a similar pattern as the ACF for RV_t (Figure 2.2) which both shows a slow decay in serial correlations. However, the ACF for $RV_t^{(d)}$ on the right shows that there is hardly any autocorrelation for $k > 1$ - the 'jump' effect did not even last for more than one day. This is consistent with what we have observed from the high-frequency return plots with 'jumps'.

Bearing this in mind, we may need to make certain adjustments to the HAR model fitted to $RV_t^{(d)}$ as the weekly and monthly components hardly play any significant part in determining $RV^{(d)}$ on the next day. Therefore, we have proposed the dHAR model which only involves the daily $RV_t^{(d)}$ in predicting $RV_{t+1}^{(d)}$. The modified model for $RV_t^{(d)}$ is shown below:

$$\log RV_{t+1}^{(d)} = \beta_0 + \beta^{(d)} \log RV_t^{(d)} \tag{3.9}$$

Equation 3.9 is then combined with Equation 3.4 to form what we named as the **HAR_RVc+dRVd** model and then compute the final $\log RV$ together. The estimated coefficient for Equation 3.9 is also reported in Table 3.3.

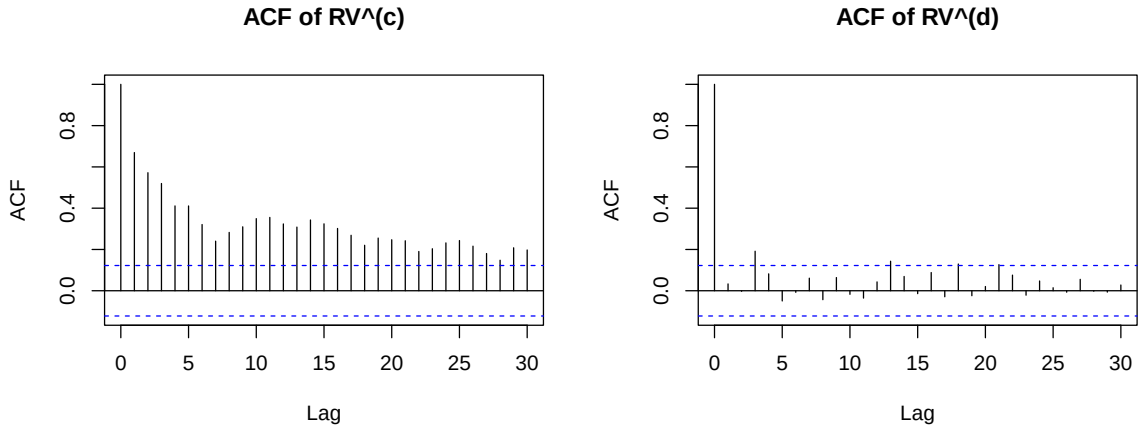


Figure 3.5.: ACF plot for $RV^{(c)}$ (left) and $RV^{(d)}$ (right)

Bipower variance separation

The second diffusive-jump separation method we explored here makes use of the asymptotic distributions of RV_t, $\mu_1^{-2}BPV_t$ and $(RV_t - \mu_1^{-2}BPV_t)$ discussed in Section 2.2. Recall from Equation 2.2 and 2.3 that RV_t converges in probability to QV_t; $\mu_1^{-2}BPV_t$ converges in probability to the continuous volatility of QV_t; and $(RV_t - \mu_1^{-2}BPV_t)$ converges in probability to the discontinuous jumps of QV_t. Therefore, it is natural to think of splitting RV_t into $\mu_1^{-2}BPV_t$ and $(RV_t - \mu_1^{-2}BPV_t)$. However, note that in real practice, there exist situations where $RV_t < \mu_1^{-2}BPV_t$, and hence $RV_t - \mu_1^{-2}BPV_t < 0$. In order to avoid 'negative' estimation for $\sum(\Delta J_t)^2$, we will use the truncated result which takes the maximum between 0 and $(RV_t - \mu_1^{-2}BPV_t)$. The formal representation for C_t and D_t with the bipower variance separation is then written as:

$$C_t = BV_t = \mu_1^{-2} \sum_{j=1}^{[1/\delta]} |y_{j-1,t}||y_{j,t}| \quad \text{where} \quad \mu_1 = \sqrt{\frac{2}{\pi}}$$
$$D_t = J_t = \max(RV_t - BV_t,\, 0)$$

One can easily transform BV_t to $\log BV_t$ and fit it with the HAR model. However, we cannot do the direct logarithmic transformation on J_t as it could be zero sometimes (28.8% of the J_t is 0). The adjustment made here was to add a small constant ϵ to J_t before the logarithmic transformation to make the final output as $\log(J + \epsilon)$. Then, by fitting $\log BV_t$ and $\log(J + \epsilon)$ (will be written as $\log^* J_t$ for simplicity) with the HAR model, we have the following **HAR_BV+J** model:

$$\log BV_{t+1} = \phi_0 + \phi^{(d)} \log BV_t + \phi^{(w)} \log BV_{t:t-4} + \phi^{(m)} \log BV_{t:t-21} \tag{3.10}$$
$$\log^* J_{t+1} = \beta_0 + \beta^{(d)} \log^* J_t + \beta^{(w)} \log^* J_{t:t-4} + \beta^{(m)} \log^* J_{t:t-21} \tag{3.11}$$

Note that in this case, when we attempt to obtain the predicted value for $\log RV_t$, one needs to subtract the ϵ in the back calculation as follows:

$$\log RV_t = \log(\exp(\log BV_t) + \exp(\log(J_t + \epsilon)) - \epsilon) \tag{3.12}$$

To determine the most suitable value for ϵ, we have conducted several experiments of model fitting and evaluation with ϵ ranging from $1e-5$ to $1e-10$. The real values for J_t mostly fall between $1e-5$ and $1e-6$. It is found that the most suitable value for $\epsilon = 1e-6$, which is slightly but not much below the overall mean of the non-zero J_ts. The value for ϵ should not be too small as if it becomes too close to zero, the logarithmic transformation will produce very negative values and introduce many noises and instabilities to the model. Figure 3.6 plots the time series for J_t (top), $\log(J+1e-6)$ (middle) and $\log(J+1e-10)$ (bottom) together. The bottom plot confirmed our concern regarding using a too-small value for ϵ as there are too many downward spikes and the original pattern of J_t is no longer reserved. On the contrary, for $\log(J + 1e-6)$, the general shape of the original graph is largely preserved while the log transformation shrinks the extreme values closer to the mean.

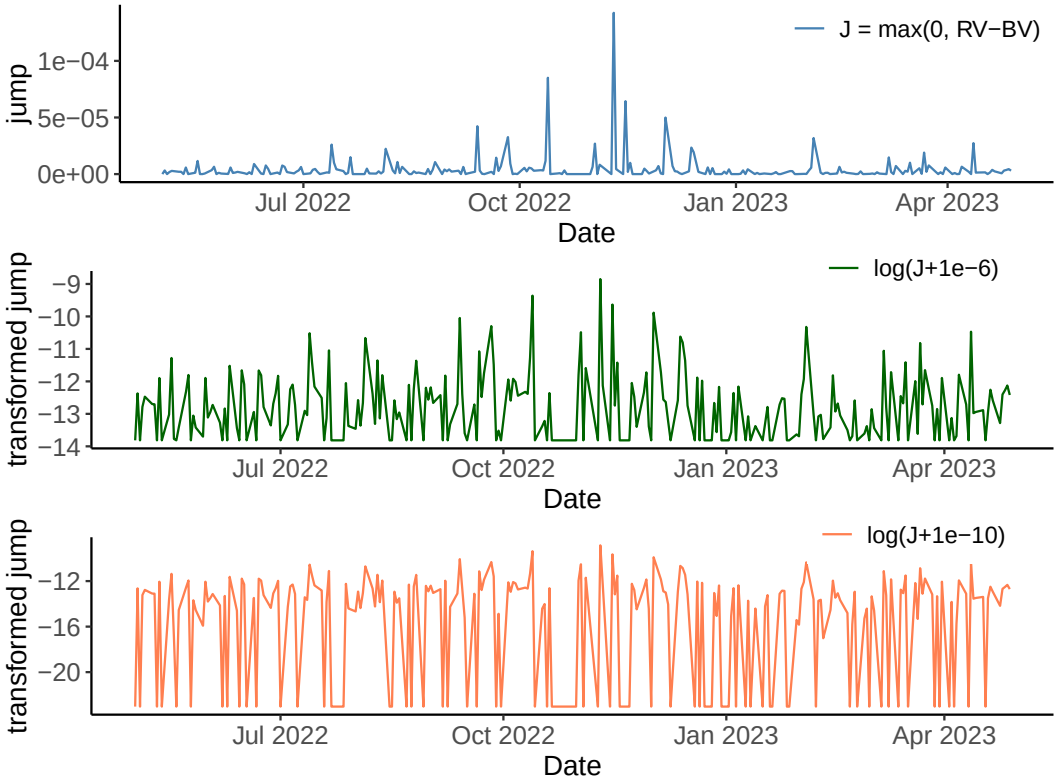


Figure 3.6.: Comparison between J_t and the transformed J_t

In addition, through investigating the statistical properties for BV_t and J_t, we found that the ACF plots for these two are very similar to the ACF plots for $RV_t^{(c)}$ and $RV_t^{(d)}$ respectively shown in Figure 3.5 - BV_t has autocorrelation which is statistically significant for very large k while the autocorrelation for J_t decays very fast within one day. Therefore, the dHAR model is also implemented on $\log^* J_t$ (Equation 3.13) and together with Equation 3.10 to form the **HAR_BV+dJ** model.

$$\log^* J_{t+1} = \beta_0 + \beta^{(d)} \log^* J_t \tag{3.13}$$

Fitting with the data of 257 trading days, the estimated coefficients for the two models using this separation method are shown in Table 3.5. This result shows similar signs as Table 3.3 with improvement on the adjusted R^2 for the diffusive volatility model but rather poor performance for the 'jump' model.

Ratio jump test separation

Recall that in the last chapter, following the asymptotic distribution, the statistical test for jumps was also discussed in Section 2.3. Our third diffusive-jump separation technique was motivated by using the result from the statistical test statistic Z_t. We define the adjusted 'jump' component using $(RV_t - \mu_1^{-2} BPV_t)$ multiplied with an indicator

Model	Fitted term	ϕ_0/ β_0	$\phi^{(d)}$/ $\beta^{(d)}$	$\phi^{(w)}$/ $\beta^{(w)}$	$\phi^{(m)}$/ $\beta^{(m)}$	adj.R^2
HAR_BV+J	$\log BV_{t+1}$	-1.1807	0.3499	0.3174	0.2184	0.4428
		(0.1584)	(8.26e-06)	(0.0112)	(0.0990)	
	$\log^* J_{t+1}$	-10.2876	0.0145	0.0849	0.0958	-0.0092
		(0.0056)	(0.8433)	(0.6264)	(0.7713)	
HAR_BV+dJ	$\log^* J_{t+1}$	-12.3119	0.0370	-	-	-0.0029
		(2e-16)	(0.5720)			

Table 3.5.: Estimated coefficient (p-value) for each predictor for HAR_BV+J and HAR_BV+dJ

function which determines whether one rejects the jump test at 1% significance level (where $Z_t < \Phi_{0.01}$). The rest will then be considered as the adjusted continuous volatility. Note that we also truncate this adjusted jump at zero to avoid situations where $(RV_t - \mu_1^{-2} BPV_t) < 0$. The formal separation of this method is defined as follows:

$$C_t = \text{adj.}C_t = RV_t \mathbb{I}_{\{Z_t \geq \Phi_{0.01}\}} + \mu_1^{-2} BPV_t \mathbb{I}_{\{Z_t < \Phi_{0.01}\}}$$
$$D_t = \text{adj.}J_t = \max\big((RV_t - \mu_1^{-2} BPV_t)\mathbb{I}_{\{Z_t < \Phi_{0.01}\}}, 0\big)$$

Similar to what we have encountered in the bipower variance separation, the log function cannot be directly applied to adj.J_t, therefore the alternative transformation $\log(\text{adj.}J_t + \epsilon)$ is used where $\epsilon = 1\text{e} - 6$. Then fitting $\log \text{adj.}C_t$ and $\log^* \text{adj.}J_t$ together gives what we named as **HAR_adj.C+J**: ($\log RV_t$ will be back computed in similar way as Equation 3.12)

$$\log \text{adj.}C_{t+1} = \phi_0 + \phi^{(d)} \log \text{adj.}C_t + \phi^{(w)} \log \text{adj.}C_{t:t-4} + \phi^{(m)} \log \text{adj.}C_{t:t-21} \quad (3.14)$$
$$\log^* \text{adj.}J_{t+1} = \beta_0 + \beta^{(d)} \log^* \text{adj.}J_t + \beta^{(w)} \log^* \text{adj.}J_{t:t-4} + \beta^{(m)} \log^* \text{adj.}J_{t:t-21} \quad (3.15)$$

Again, with no surprise, the ACF plots for adj.C_t and adj.J_t resemble what we have seen in Figure 3.5. Hence, the hHAR model will be also fitted to adj.J_t to see if it is a more suitable model than HAR. The estimated parameters for the models are reported in Table 3.6 where one can see there is not much difference to the results in Table 3.5 in terms of both coefficient values and adjusted R^2 results. However, this is not sufficient to determine which separation technique is the best. The in-sample and out-of-sample model evaluation metrics will be compared to further investigate their effectiveness.

Separation within HAR model

What we have been doing so far is to construct separate models for the continuous and discontinuous parts and combine them together to compute the final estimation for RV_t. Meanwhile, other methods have been proposed (Andersen et al., 2007) that one can split RV_t into the continuous and discontinuous parts within the HAR model (similar to what was done in the semivariance measures). The original model proposed by Andersen et al. (2007) is named as **HAR_RVJ** written in Equation 3.16 where recall

Model	Fitted term	ϕ_0/ β_0	$\phi^{(d)}$/ $\beta^{(d)}$	$\phi^{(w)}$/ $\beta^{(w)}$	$\phi^{(m)}$/ $\beta^{(m)}$	adj.R^2
HAR_adj.C+J	$\log \text{adj.}C_{t+1}$	-1.2377	0.3468	0.3277	0.2054	0.4371
		(0.1423)	(9.61e-6)	(0.0087)	(0.1205)	
	$\log^* \text{adj.}J_{t+1}$	-9.5159	-0.0088	-0.1455	0.4449	-0.0064
		(0.0318)	(0.9036)	(0.4346)	(0.2399)	
HAR_adj.C+dJ	$\log^* \text{adj.}J_{t+1}$	-13.6137	-0.0142	-	-	-0.0041
		(2e-16)	(0.8270)			

Table 3.6.: Estimated coefficient (p-value) for each predictor for HAR_adj.C+J and HAR_adj.C+dJ

that $J_t = \max(0, RV_t - \mu_1^{-2} BPV_t)$. In this model, the jump component has been added to the model as a predictor with a newly assigned coefficient.

$$\log RV_{t+1} = \phi_0 + \phi^{(d)} \log RV_t + \phi^{(w)} \log RV_{t:t-4} + \phi^{(m)} \log RV_{t:t-21} + \beta^{(d)} \log^* J_t \quad (3.16)$$

Motivated by Equation 3.16, we also proposed the three new models **HAR_RVcd** (3.17), **HAR_BVJ** (3.18), **HAR_adj.CJ** (3.19) which use the three separation methods discussed above:

$$\log RV_{t+1} = \phi_0 + \phi^{(d)} \log RV_t^{(c)} + \phi^{(w)} \log RV_{t:t-4}^{(c)} + \phi^{(m)} \log RV_{t:t-21}^{(c)} + \beta^{(d)} \log^* RV_t^{(d)} \quad (3.17)$$

$$\log RV_{t+1} = \phi_0 + \phi^{(d)} \log BV_t + \phi^{(w)} \log BV_{t:t-4} + \phi^{(m)} \log BV_{t:t-21} + \beta^{(d)} \log^* J_t \quad (3.18)$$

$$\log RV_{t+1} = \phi_0 + \phi^{(d)} \log \text{adj.}C_t + \phi^{(w)} \log \text{adj.}C_{t:t-4} + \phi^{(m)} \log \text{adj.}C_{t:t-21} + \beta^{(d)} \log^* \text{adj.}J_t \quad (3.19)$$

Table 3.7 reports the estimated coefficients for Model 3.16 to 3.19. All the models again confirm the statistical significance of $\phi^{(d)}$. As for the jump term, except for HAR_BVJ, the coefficient $\beta^{(d)}$ for the rest three models, all demonstrate certain levels of significance from their p-values. Note the estimated $\beta^{(d)} < 0$ for HAR_RVJ, this is probably because the impact of volatility persistence in RV_t is partly reduced by the jump components - as we have previously discussed, the jump components do not show lagged impacts. On the other hand, the estimated $\beta^{(d)} > 0$ for the rest three models. Note that in these three cases, we have split RV_t into two separate components, and therefore the $\beta^{(d)}$ term here does not serve as a component for offsetting. In addition, these four models all have a much higher adjusted R^2 value than the original HAR_RV in Table 3.1, suggesting possible better fits. However, to confirm this conclusion, one should also check the evaluation metrics of the models which will be discussed in the following subsection.

Model	ϕ_0	$\phi^{(d)}$	$\phi^{(w)}$	$\phi^{(m)}$	$\beta^{(d)}$	adj.R^2
HAR_RVJ	-1.7914	0.3858	0.2483	0.2677	-0.0613	0.4062
	(0.0627)	(3.45e-5)	(0.0625)	(0.0612)	(0.0867)	
HAR_RVcd	-0.9232	0.3859	0.2729	0.2170	0.0290	0.5259
	(0.2131)	(2.22e-6)	(0.0223)	(0.0795)	(0.0771)	
HAR_BVJ	-1.1315	0.3449	0.3204	0.2181	0.0056	0.4405
	(0.1952)	(2.88e-5)	(0.0111)	(0.1001)	(0.8404)	
HAR_adj.CJ	-0.9232	0.3859	0.2729	0.2170	0.0290	0.5259
	(0.2131)	(2.22e-6)	(0.0223)	(0.0795)	(0.0771)	

Table 3.7.: Estimated coefficient (p-value) for each predictor for HAR_RVJ, HAR_RVcd, HAR_BVJ and HAR_adj.CJ

3.1.5. Forecasting models summary and comparison

Under this section, we summarize and compare all the methods we have proposed in the previous part. For better reference, a complete list of all the models we have proposed for univariate point forecast can be found in Appendix A.

To compare the predictive performance of the models, we first performed the in-sample forecast and out-of-sample forecast with our dataset.

- In-sample forecast: the data for all 257 trading days are used to train the model parameters. Then the training data is again evaluated by the model to give the forecasting results.
- Out-of-sample: start from day 23 (the first day where monthly $RV_{t:t-21}$ is available for prediction), a rolling window of 191 days (10 months) will be used as the training set which trains the model and predicts for the next day $\log RV_t$. In total, we have the out-of-sample forecasts for the last 43 days (last two months), which can alternatively be considered as the testing set.

Then, by comparing the forecasting results with the true dataset, we are able to compute the mean squared error (MSE) for the model predictions which is the most commonly used model evaluation metric.

In addition to MSE, we also implemented the Diebold-Mariano (DM) test (Diebold and Mariano, 1995) to test for predictive accuracy. The DM test is used to compare the predictive performance of two forecasting methods, say A and B. Here, we perform the one-side DM test whose null hypothesis is that model A's forecast is not 'better' (have lower forecast errors) than model B. The test will be conducted with following steps:

1. Let $e_t^{(a)}$ and $e_t^{(b)}$ be the forecast errors at time point t of model A and B respectively, and let $d_t = (e_t^{(a)})^2 - (e_t^{(b)})^2$, $\bar{d} = \frac{1}{N}\sum_{t=1}^{N} d_t$.
2. Calculate the standard error for $\bar{d}$: $SE(\bar{d}) = \sqrt{\frac{1}{N(N-1)}\sum_{t=1}^{N}(d_t - \bar{t})^2}$

3. Compute the Diebold-Mariano test statistic $DM = \frac{\bar{d}}{SE(\bar{d})}$, which is then compared to the upper quantile of $N(0,1)$, and the respective p-value can be calculated. A small p-value suggests that there is sufficient evidence to reject the null hypothesis and conclude that model A has better predictive accuracy than model B.

Table 3.8 reports the evaluation metrics for all the proposed models including the in-sample and out-of-sample MSE, and the p-values of the 1-side DM test for in-sample and out-of-sample forecasts. Note that we use HAR_RV as the base case, each newly proposed model is then tested against the HAR_RV model with the alternative hypothesis that the new model has better predictive performance than HAR_RV. The p-values for the DM test are then recorded. Here are a few remarks from the table that worth highlighting:

- By observing the first two columns, one can see that most of the newly proposed models have a lower MSE than HAR_RV for both in-sample and out-of-sample cases. HAR_adj.C+J and HAR_adj.C+dJ both has the lowest in-sample MSE, and HAR_RVc+RVd has the lowest out-of-sample MSE.

- Moving on to the last two columns, p-values which are less than 0.05 are highlighted. One can see that **HAR_BV+J** and **HAR_BV+dJ** are the two models whose tests are rejected for both in-sample and out-of-sample forecasts, suggesting that these two models have better prediction performance than HAR_RV under both the training and testing context. In addition, **HAR_adj.C+J** and **HAR_adj.C+dJ** also show relatively good results with very low DM-test p-values for the in-sample forecast and those for the out-of-sample forecast are only slightly above 0.05.

- The semivariance measure (SHAR_RV) does not show a very significant improvement on the original HAR model from the DM test p-values.

- The models that separate the components within themselves (last four rows) do not demonstrate a significant improvement as compared to methods where the components are separately modelled.

It is therefore able to conclude that for this EURUSD exchange rate dataset, the bipower variance separation and ratio jump test separation are the two most effective methods in terms of forecast accuracy improvement.

3.1.6. Forecast Aggregation

In this subsection, we will briefly explore the expert aggregation for realized variance forecasting in our context. Forecast aggregation is, as its name suggests, to combine the prediction of individual forecasters, also known as 'experts', together to produce an aggregated forecast. The aggregated forecast is found to have a significant improvement on the individual forecasts in electricity forecasting studies (Gaillard and Goude, 2015). Under the financial forecasting context, Carl et al. (2021) has applied expert aggregation on different machine learning models to construct stock return predictions. So far, very

Model	in-samp MSE	out-samp MSE	in-samp $DM_{1\text{-side}}$	out-samp $DM_{1\text{-side}}$
HAR_RV	0.2063	0.2188	NA	NA
SHAR_RV	0.2030	0.2136	0.0929	0.2209
HAR_RVc+RVd	0.2033	0.2102	0.2695	0.2314
HAR_RVc+dRVd	0.2070	**0.2064**	0.5422	0.0953
SHAR_RVc+RVd	0.2066	0.2102	0.5142	0.2314
HAR_BV+J	0.2008	0.2094	**0.0298**	**0.0416**
HAR_BV+dJ	0.2006	0.2087	**0.0281**	**0.0315**
HAR_adj.C+J	**0.2002**	0.2114	**0.0183**	0.07366
HAR_adj.C+dJ	**0.2002**	0.2109	**0.0196**	0.0685
HAR_RVJ	0.2037	0.2125	0.1780	0.1145
HAR_RVcd	0.2205	0.2182	0.9391	0.4858
HAR_BVJ	0.2061	0.2159	0.4824	0.4171
HAR_adj.CJ	0.2027	0.2155	0.1823	0.3854

Table 3.8.: Table to compare performance metrics of all fitted models for $\log RV$

little literature has analysed how such forecast aggregation would perform on volatility forecasting. By selecting a few well-performed models from the previous section, we will aggregate the individual expert forecasts and assess if there are any improvements in the predictive performance.

Suppose at time t, we have a finite number K experts to form a forecast vector $\mathbf{x}_t = (x_{1,t}, \cdots, x_{K,t})$. In forecast aggregation, one assigns weights to each expert prediction, $\mathbf{w}_t = (w_{1,t}, \cdots, w_{K,t}) \in \mathbb{R}^k$. Then the aggregated forecast $\hat{x}_t$ is the inner products of the experts' predictions and their corresponding weights $\hat{x}_t = \mathbf{w}_t \cdot \mathbf{x}_t = \sum_{k=1}^{K} w_{k,t} x_{k,t}$. Note that under most circumstances, weight vector $\mathbf{w}_t$ would be consistently updated over time with knowledge of the past observations and of the new expert advice. The algorithms to decide how w_t is computed and how $\hat{x}_t$ is formed are known as aggregation rules. There are various types of aggregation rules, and here we will introduce and apply one commonly used rule - Bernstein Online Aggregation (BOA) - proposed in Wintenberger (2017). Suppose $\ell_{k,t}$ is the square loss of expert k at time t, BOA will update the weights for each expert according to its prediction error compared to other forecasters. Algorithm 1 (Carl et al., 2021) below presents how weights are updated for each expert at each time-point t.

Then we would perform BOA aggregation with the models we have developed in the previous part of this chapter. We select the base HAR_RV model, together with six other models whose DM test p-values are less than 0.1 for either in-sample or out-of-sample tests - SHAR_RV, HAR_RVc+dRVd, HAR_BV+J, HAR_BV+dJ, HAR_adj.C+J, and HAR_adj.C+dJ. The BOA rule can be implemented with the `opera` package in R.

Figure 3.7 below presents the weights assigned to each expert throughout all training days. From the plot, we can see that four models stand out and take the majority of the

Algorithm 1 Bernstein Online Aggregation (BOA) with learning rate

Input: Equal initial weights $w_{k,0} = \frac{1}{K}$ so that $\sum_{k=0}^{K} w_{k,0} = 1$ and parameter learning rate $\eta > 0$

for each t, compute the weight vector $\mathbf{w_t}$ **do**

$$w_{k,t} = \frac{\exp(-\eta \ell_{k,t}(1 + \eta \ell_{k,t}))w_{k,t-1}}{\sum_{s=1}^{t-1} \exp(-\eta \ell_s (1 + \eta \ell_s))}$$

end for

weight across the timeline. Around the first 90 days, **HAR_BV+J** and **HAR_BV+dJ** take most of the weights while for the days after **HAR_adj.C+J** and **HAR_adj.C+dJ** occupy the majority of the weights. On the last day, these four experts together take around 90% of the weights in the aggregated forecast. This result again helps to confirm our previous comment on the forecasting evaluation results - these four models are indeed the most well-performed models compared to HAR_RV. Recall that from Table 3.8, model HAR_BV+J and HAR_BV+dJ are found to be the ones with the best performance in terms of DM test results. However, Algorithm 1 suggests that weights vector $\mathbf{w}_t$ depends largely on the loss (MSE) of the experts, and the model HAR_adj.C+J and HAR_adj.C+dJ give the least in-sample MSE which probably explains why these two models take a majority of the weights throughout most of the time. The last row of table 3.9 presents the evaluation metrics of the aggregated forecast with the BOA rule. For comparison convenience, we have copied the first three rows from Table 3.8. By comparing the metrics of BOA aggregated forecast with that of HAR_RV, it can be clearly seen that there is an improvement in predictive accuracy in both in-sample and out-of-sample forecasts. However, while comparing these metrics with the individual expert HAR_BV+dJ and HAR_adj.C+dJ, it is found that model HAR_BV+dJ outperforms the aggregated forecast in terms of both MSE and DM test p-values.

Unlike electricity forecasting, so far there is no sign suggesting that the aggregated forecast of realized variance will have a significant improvement on the predictive performance of the individual expert. Instead, our result suggests the opposite - the 'best' experts will take the dominant weights and give better evaluation results than the aggregated forecast. This is likely because we have very similar experts in this case - all are variations of the HAR_RV model, and there is very limited room for the aggregated rules to explore between different performances of the experts. It is advised that in future research, one could incorporate different types of volatility forecasting methods such as machine learning models or other stochastic models in the forecast aggregation. More variation of experts could possibly bring more dynamics to the aggregation process, and it is worth exploring if this could lead to any predictive improvements.

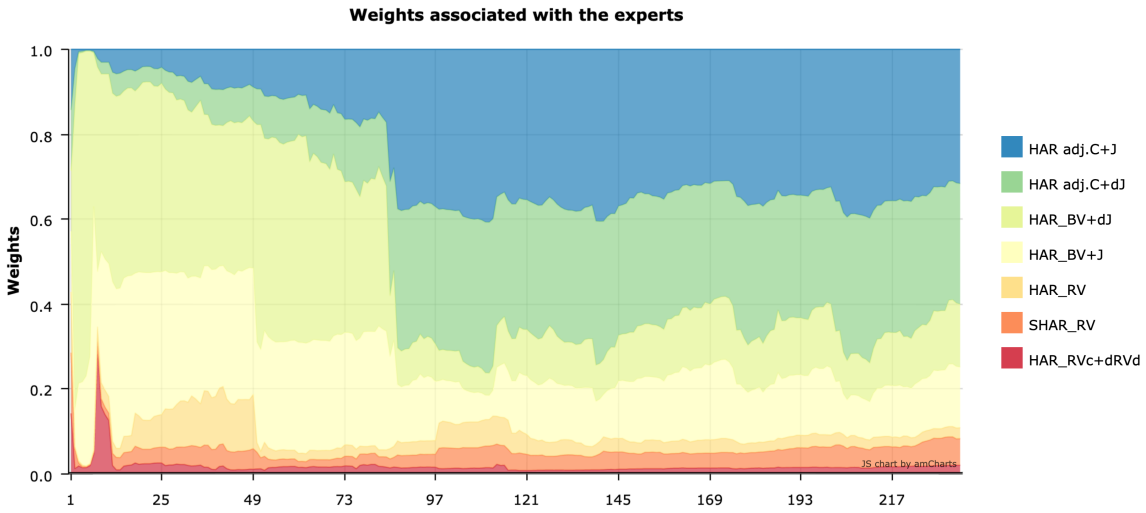


Figure 3.7.: Weights associated with the experts over time in in-sample aggregated forecast

Model	in-samp MSE	out-samp MSE	in-samp $DM_{1\text{-side}}$	out-samp $DM_{1\text{-side}}$
HAR_RV	0.2063	0.2188	NA	NA
HAR_BV+dJ	0.2006	0.2087	0.0281	0.0315
HAR_adj.C+dJ	0.2002	0.2109	0.0196	0.0685
BOA agg	0.2014	0.2102	0.0315	0.0344

Table 3.9.: Performance metrics comparison of aggregated forecast with individual experts

3.2. Probabilistic Forecast

So far, we have been focusing on the point forecast which aims to predict the exact value for the next day's realized variance. However, it is very common for the point forecast to deviate from the true value to a certain extent. Therefore under such circumstances, a probabilistic forecast seems to be more applicable and informative. Unlike point forecast, the probabilistic forecast actually quantifies the uncertainties in prediction (Gneiting and Katzfuss, 2014), and instead of outputting a single prediction value, it will predict the distribution of the realized variance. In fact, Timmermann (2000) has already pointed out that the density forecast of the financial time series plays a large role in risk management. The predictive quantiles deliver useful information for financial intuitions that constantly monitor their Value-at-Risk (VaR) levels. So far, several attempts have been made to conduct probabilistic forecasts for asset class volatility - methods include the classical Bayesian inferential approach (Maneesoonthorn et al., 2012) and more recent machine learning models (Thavaneswaran et al., 2022). In this project, we will make use of the distribution of the residuals from the point forecast models to predict the distribution of the realized variance in future days.

We start with exploring the residual distribution of the in-sample fitted HAR_RV

model. Figure 3.8 presents the density histogram of the residuals and the blue line represents the fitted Gaussian distribution to the residuals data. One can clearly see from the histogram that the distribution of the residuals is quite asymmetric and it has rather long tails on the right. Therefore, the Gaussian distribution is probably not the most suitable distribution to describe the residuals.

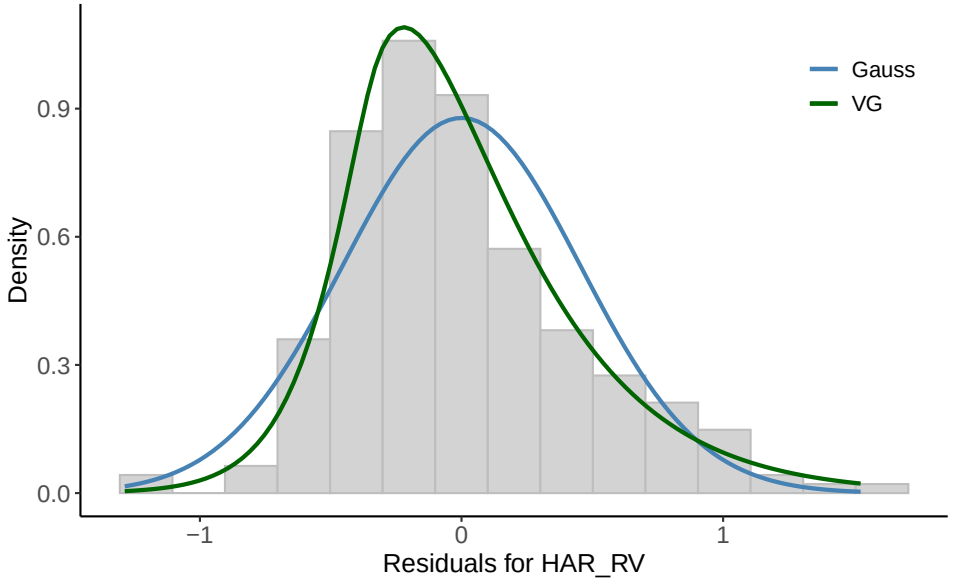


Figure 3.8.: Density histogram of the residuals for HAR_RV with the fitted distributions

Here, we introduce the class of Generalized Hyperbolic (GH) distribution which was first proposed in Barndorff-Nielsen (1977). The density function for GH distribution in univariate case is shown below (there are few parametrizations and here we present the most widely used one):

$$f_{\mathbf{X}}(\boldsymbol{x}) = \frac{\left(\alpha^2-\beta^2\right)^{\lambda/2}}{\sqrt{2\pi}\alpha^{\lambda-\frac{1}{2}}\delta^{\lambda}\mathrm{K}_{\lambda}\left(\delta\sqrt{\alpha^2-\beta^2}\right)} \times \frac{\mathrm{K}_{\lambda-\frac{1}{2}}\left(\alpha\sqrt{\delta^2+(x-\mu)^2}\right)}{\left(\sqrt{\delta^2+(x-\mu)^2}\right)^{\frac{1}{2}-\lambda}} e^{\beta(x-\mu)} \tag{3.20}$$

where $\mathrm{K}_{\lambda}(\cdot)$ is the modified Bessel function of the third kind, β is the asymmetry parameter, λ is the scale parameter and μ is the location parameter. The GH distribution is suitable for characterising semi-heavy tail and asymmetry behaviours. Special cases of the GH distribution include the Gaussian distribution, Student's t-distribution, the hyperbolic distribution, the normal-inverse Gaussian (NIG) distribution and the variance-gamma (VR) distribution (Marc et al., 2022). Therefore, we fit the residuals of HAR_RV again with the GH distributions this time, and it is found that the VG distribution gives the lowest Akaike Information Criterion (AIC) score among all cases of GH distribution family. The green line in Figure 3.8 plots the fitted VG distribution on the density histogram. One can see that the green line generally captures the overall empirical distribution of the data. It is also observed during the experiment that asymmetric cases

of GH distribution always have much better performance (lower AIC) compared to their respective symmetric version (eg. asymmetric fitted NIG has a much lower AIC than the fitted symmetric NIG in this case).

With the pool of distributions for residual fitting, here is how the probabilistic forecast is implemented:

1. Pick a model, say HAR_RV, and train the model on data from time 1 to T, and obtain the 1-day point forecast for time $T+1$, say $\hat{X}_{T+1}$.

2. Obtain the residuals of the fitted model from the training set (e_1 to e_T), and pick a GH distribution to fit the residuals and estimate the relevant parameters (can be a fixed distribution such as Gaussian or NIG, or alternatively, one can always fit the data with all distributions and pick the one with the lowest AIC score).

3. The point forecast $\hat{X}_{T+1}$ is then added to the location parameters of the fitted GH distribution. The shifted GH distribution is then the probabilistic forecast for time $T+1$.

Similar to the point forecast, we perform both the in-sample and out-of-sample probabilistic forecast where

- In-sample forecast: A point forecast model is trained with the entire dataset and the in-sample point forecast is obtained. Compute the residuals of the in-sample model and fit the appropriate GH distribution. The location parameters will be adjusted according the point forecast at each time point.

- Out-of-sample forecast: Rolling window of 191 days is used as the training set which estimates the parameters for the point-forecast model and predicts for the next day $\log RV_{T+1}$. Compute the residuals of the model and fit the appropriate GH distribution whose location parameters will be added with the point forecast for day $T+1$. The adjusted GH distribution is then the out-of-sample probabilistic forecast for day $T+1$. In total, we have the out-of-sample forecast for the last 43 days, which can alternatively be considered as the testing set.

Here we will try three different cases for the fitted GH distribution. We first fix the distribution to be the Gaussian distribution and the asymmetric NIG distribution. In the third case, we will choose the 'best' fitted GH distributions with the lowest AIC score at each time point. It is expected that the third method will give the best output as it constantly caters to any possible change in the distribution of the residuals. However, the method itself is very computationally expansive and time-consuming as it requires fitting all GH distributions at each time point in order to pick the 'best' one to proceed.

In order to assess the performance of the probabilistic forecasters, we first introduce the evaluation metrics that will be used. Probabilistic calibration and sharpness are the two main criteria to assess when evaluating a probabilistic forecast. Calibration checks

the statistical compatibility between the forecast and the realizations, and sharpness assesses the concentration of the predictive intervals (Gneiting and Katzfuss, 2014). The probability integral transform (PIT) is widely used to check if the probabilistic forecast is well calibrated, and it is formally defined as below (Gneiting and Ranjan, 2013):

Definition 3.2.1. Let F denotes a fixed, non-random predictive CDF for an observation Y. Let V be a standard uniformly distributed variable that is independent of the CDF-valued random quantity F and the observation Y. For $y \in \mathbb{R}$, let $F(y-) = \lim F_{x \uparrow y}(x)$. Then

$$Z_F = F(Y-) + V(F(Y) - F(Y-))$$

is the PIT of the probabilistic forecast F.

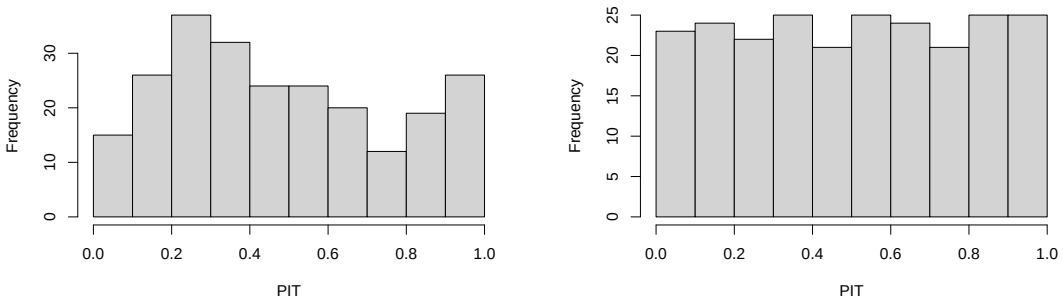
It is then pointed out that the F is statistically calibrated if Z_F has a standard uniform distribution. Therefore, the closer the PIT histogram is to a uniform distribution, the more well-calibrated our probabilistic forecast is. In addition, we also employed the local proper scoring rule (Gneiting and Raftery, 2007) to assess calibration and sharpness at the same time. The forecasters are aiming to minimize the score.

Definition 3.2.2. The logarithmic score (LS) is defined as:

$$LS(f, y) = -\log f(y)$$

With the method and evaluation metrics properly defined, we shall continue with the discussion on the probabilistic forecast with the model HAR_RV. Recall that Figure 3.8 reveals that the VG distribution fits the residuals better than the Gaussian distribution. We will then first perform the in-sample probabilistic point forecast with the model HAR_RV. As mentioned earlier, we have fit the residuals with Gaussian distribution, NIG distribution and the 'best' GH distribution chosen by AIC score (VR distribution in this case). Figure 3.9 presents the PIT histograms for the probabilistic forecast models (note that the PIT histogram for NIG distribution is the same as VR distribution). It can be seen from the plots that the PIT histogram of the probabilistic forecast with Gaussian distribution is far from statistically uniform, on the contrary, the probabilistic forecasts with asymmetric NIG and VR distribution have PIT histograms which are very close to a uniform distribution. It suggests that these two probabilistic forecasts are well-calibrated and are much better than the performance of forecast using the Gaussian distribution. The LS scores of the three probabilistic forecasts with HAR_RV can be found in the first row of Table 3.10. It is clearly seen that the LS score for using asymmetric NIG and the best-fitted GH distribution is lower than that with the Gaussian distribution.

We then extend our analysis to all the rest models. To avoid any redundant and repetitive work, we will work on selective models in this section. Based on the result from point forecasts, the base model HAR_RV together with the four outperforming models are selected to further investigate the distributional forecast - HAR_BV+J, HAR_BV+dJ, HAR_adj.C+J, HAR_adj.C+dJ. We then conduct the probabilistic forecast with each of the five point-forecast models. As mentioned earlier, the probabilistic forecast will



(a) Gauss distribution

(b) NIG/VR distribution

Figure 3.9.: HAR_RV in-sample PIT histogram for probabilistic distribution models

be done with Gaussian, asymmetric NIG and the 'best' GH distribution selected at each time point. Both the in-sample and out-of-sample forecasts are performed and the corresponding LS scores are reported in Table 3.10. It can be observed from the table that:

- For both in-sample and out-of-sample forecasts of all models, the probabilistic forecast with Gaussian distribution has the largest LS score, suggesting the poorest performance.
- For the in-sample forecast, the LS score for asymmetric NIG is the same as using the 'best' selected GH distribution. However, when it comes to out-of-sample forecasts, the probabilistic forecast with the more flexible 'best' selected GH distribution still outperforms the one which fixes the asymmetric NIG distribution. As more uncertainties exist, the forecast with a more flexible and adaptable choice of distribution tends to have better performances.
- When comparing different models, the in-sample LS score table (top) seems to suggest that the improvement in the predictive ability of the selected models is extended to the probabilistic forecasting context. For all the columns (different distributions used in the probabilistic forecast), the four selected models all have a lower score than the base HAR_RV model. However, such improvement does not move on to out-of-sample forecasts. For the out-of-sample LS scores (bottom), except for the Gaussian distribution column, there is no sign of great improvement in the LS score while using the selected four models compared to HAR_RV. This might suggest that the selected models are better in goodness-of-fit, but have limited improvement in predictive ability.
- The probabilistic forecast with HAR_BV+dJ using the best-fitted GH distribution has the overall lowest LS score for both in-sample and out-of-sample forecasts.

Model	Gauss	NIG	'best'
HAR_RV	0.6298	0.5787	0.5787
HAR_BV+J	0.6161	0.5583	0.5583
HAR_BV+dJ	0.6157	0.5578	**0.5578**
HAR_adj.C+J	0.6147	0.5589	0.5589
HAR_adj.C+dJ	0.6147	0.5592	0.5592

(a) In-sample LS score

Model	Gauss	NIG	'best'
HAR_RV	0.6636	0.6320	0.6189
HAR_BV+J	0.6439	0.6290	0.6184
HAR_BV+dJ	0.6423	0.6278	**0.6170**
HAR_adj.C+J	0.6488	0.6378	0.6291
HAR_adj.C+dJ	0.6475	0.6373	0.6290

(b) Out-of-sample LS score

Table 3.10.: LS score for different probabilistic distribution forecast

To take a closer look at the best probabilistic forecast, Figure 3.10 and 3.11 plots the in-sample and out-of-sample distributional forecast with HAR_BV+dJ using the best fitted GH distribution respectively. The outer layer represents the 90% quantile, and each layer of colour represents 10% quantile change with the darkest inner layer representing 10% quantile. Through simple computation, 90.7% and 88.4% of the true values lie within the predicted 90% quantiles of the in-sample and out-of-sample forecast respectively, suggesting that the proposed probabilistic distribution is relatively reasonable. The in-sample probabilistic forecast generally captures the trend of the real-time series for $\log RV_t$, and except for a few very extreme spikes, most of the true values fall within the bound of 90% prediction levels. As for the out-of-sample forecast, though we can see the inner 10% quantiles do not seem to follow the true values very closely, in general, the greater quantile layers are able to enclose most of the true values. This could probably provide an informative message for the investors and risk managers when they are monitoring the Value-at-Risk or expected shortfall levels.

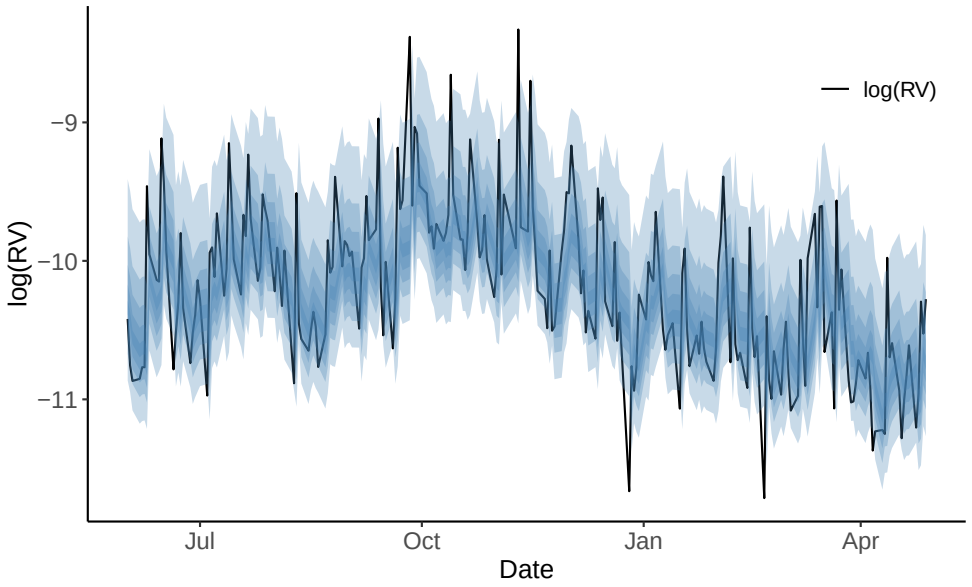


Figure 3.10.: In-sample probabilistic forecast plot for HAR_BV+dJ with best fitted GH distributions

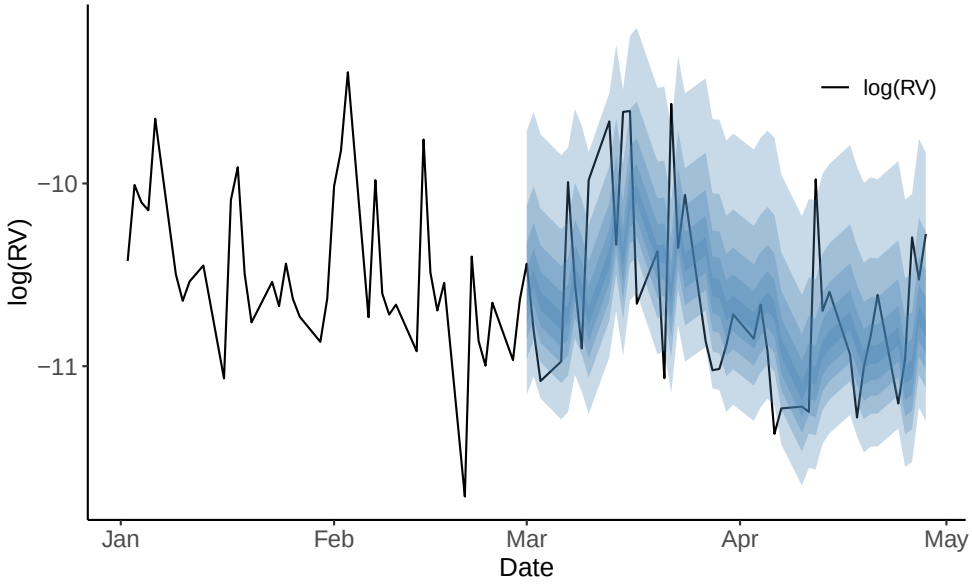


Figure 3.11.: Out-of-sample probabilistic forecast for HAR_BV+dJ with the best fitted GH distributions

4. Multivariate Variance Forecast

After a thorough analysis of the univariate forecast, we now move on to consider the volatility forecast in the multivariate context. In the real world, instead of investing in a single exchange rate, investors tend to make investments in multiple assets simultaneously or build an investment portfolio that includes different assets. Therefore, it might be more helpful to solve multivariate problems than univariate cases as the price of different assets usually demonstrates certain levels of inter-correlations. In this chapter, a new currency pair AUDUSD is introduced, and together with EURUSD, we will perform volatility forecasting from a bivariate perspective.

4.1. Exploratory analysis

We still start with some basic exploratory analysis between the two exchange rates. Figure 4.1 plots the daily average rate for the two currency pairs on the same graph with the dual axis. It can be observed from the plot that the trends of the two exchange rates are quite similar. They both experience a general decreasing rate from May 2022 to Oct 2022, then there is a turnover for both currency pairs which start to increase from Oct 2022 to Feb 2022 when the trends reverse again around roughly the same time. The correlation between EURUSD and AUDUSD is 0.4693, also indicating a certain level of positive correlation.

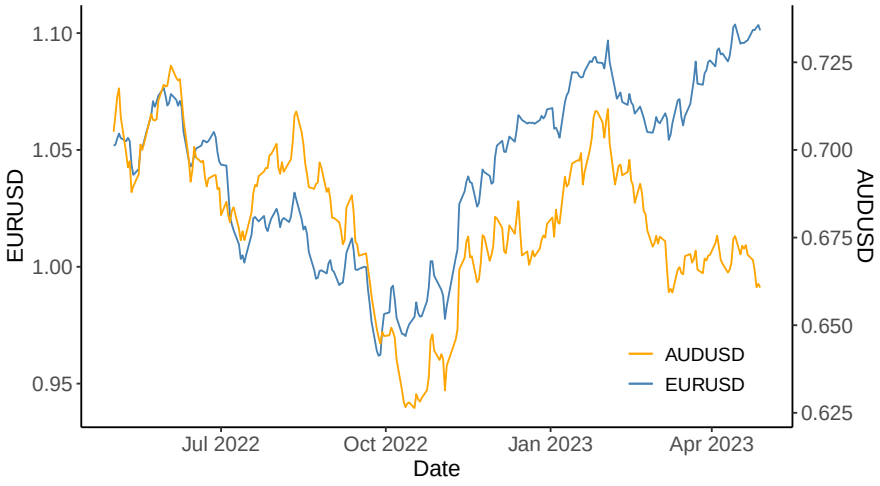


Figure 4.1.: Daily average rate of EURUSD and AURUSD, 2022 May - 2023 Apr

As for the daily realized volatility for the two exchange rates, one can see from Figure 4.2 that AUDUSD generally has a higher daily realized volatility than EURUSD. The

shape of the trends also coincide at most periods. The correlation between $\sqrt{RV_t}$ of the two currency pairs is 0.8162, suggesting that they are highly positively correlated with each other.

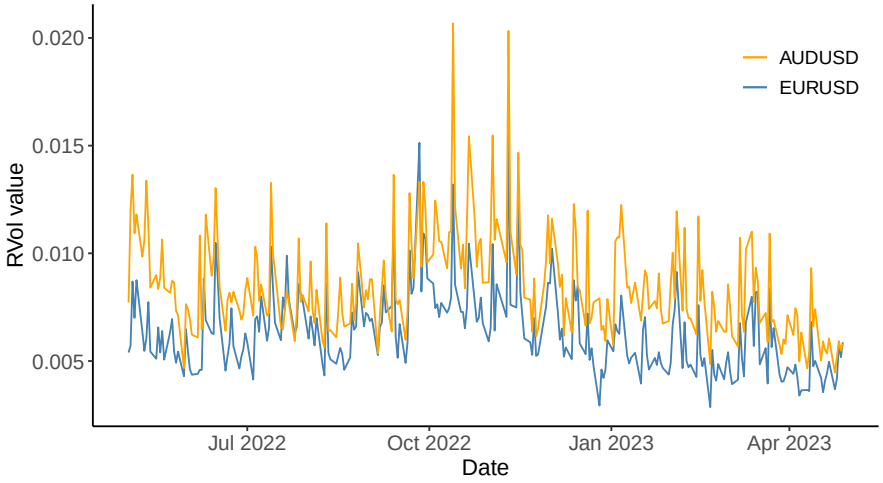


Figure 4.2.: Daily realized volatility of EURUSD and AUDUSD, 2022 May - 2023 Apr

4.2. Portfolio Variance Point Forecast

In this section, we aim to investigate whether employing a multivariate model for portfolio variance prediction would perform better than a single univariate model.

4.2.1. Multivariate HAR model

Assuming we have a portfolio with p different assets where the high-frequency price of individual asset i at time j on day t is represented by $Y_t^{(i)}$, then the high-frequency price of the portfolio at time j on day t can be written as:

$$Y_{j,t}^{(P)} = \mathbf{w}^T \mathbf{Y}_{j,t} = \sum_{i=1}^{p} w_i Y_{j,t}^{(i)} \quad \text{where} \quad \sum_{i=1}^{p} w_i = 1$$

where w_i represents the weight assigned to the asset i in the portfolio. Therefore, one is able to compute the raw realized variance of the portfolio on day t with the high-frequency log return $y_{j,t}^{(P)}$ of $Y_{j,t}^{(P)}$ (similar to Definition 1.3.2):

$$\widetilde{RV}_t^{(P)} = \sum_{j=1}^{[1/\delta]} (y_{j,t}^{(P)})^2$$

then by fitting $\widetilde{RV}_t^{(P)}$ with the univariate HAR model, named as model **HAR_$\widetilde{\text{RV}}$** (Equation 4.1), one will have a univariate forecasting model to fit and forecast the

variance of the portfolio, and this method can be applied to portfolio with as many assets as possible as long as their respective weights are known.

$$RV_{t+1}^{(P)} = \phi_0 + \phi^{(d)}\widetilde{RV}_t^{(P)} + \phi^{(w)}\widetilde{RV}_{t:t-4}^{(P)} + \phi^{(m)}\widetilde{RV}_{t:t-21}^{(P)} \tag{4.1}$$

However, it is noted that this method seems to overlook the inter-correlation between the realized variances of the assets. Instead, it seems to treat each asset in a parallel manner and combine them with simple additions. Given what we have observed in the exploratory analysis (high correlation between realized volatility of currency pairs), such inter-correlation may play an important part in determining the portfolio variance. Therefore, we aim to find a forecast model that accounts for the covariance between asset volatility and assess if it indeed has an improvement on the predictive performance on Model 4.1.

By extending the definition of realized variance to higher dimensions, Barndorff-Nielsen and Shephard (2004a) defined the realized covariance matrix as follows:

Definition 4.2.1. Say a p-dimensional logarithmic price process is sampled on a regular time grid - $\{j\delta : 0 \leq j \leq [1/\delta]\}$ - within a unit day (where $\delta = 1/288$), then let $\mathbf{Y}_{j,t} = (Y_{j,t}^{(1)}, \cdots, Y_{j,t}^{(p)})^T$ denote the n-dimensional price at j-th time interval on day t. Let the jth return of $\mathbf{Y}$ be denoted by $\mathbf{y}_{j,t} = \mathbf{Y}_{j\delta,t} - \mathbf{Y}_{(j-1)\delta,t}$, the **daily realized covariance matrix** on day t is then defined as:

$$\widehat{C}_t \equiv \sum_{j=1}^{[1/\delta]} \left(\mathbf{y}_{j,t}\right)\left(\mathbf{y}_{j,t}\right)^T$$

which implies $(\widehat{C}_t)_{ab} = \sum_{j=1}^{[1/\delta]}(y_{j,t}^{(a)})(y_{j,t}^{(b)})$

Using the covariance matrix $\widehat{C}_t$, an alternative way to compute the realized variance of a portfolio that depends on the realized covariance is shown as (Bollerslev, 2022):

$$\widehat{RV}_t^{(P)} = \mathbf{w}^T\widehat{C}_t\mathbf{w} \tag{4.2}$$

where $\mathbf{w}$ is the weight vector whose entries represents the weight of each asset classes in the portfolio. In bivariate setting of equal weights, Equation 4.2 can be written as:

$$\begin{aligned}\widehat{RV}_t^{(P)} =& 0.5^2(\widehat{C}_t)_{11} + 0.5^2(\widehat{C}_t)_{22} + 0.5(\widehat{C}_t)_{12} + 0.5(\widehat{C}_t)_{21} \\ =& 0.25RV_t^{(1)} + 0.25RV_t^{(2)} + 0.5\text{Cov}(y^{(1)}, y^{(2)})_t \\ \text{where} \quad \text{Cov}(y^{(1)}, y^{(2)})_t =& \sum_j (y_{j,t}^{(1)})(y_{j,t}^{(2)}) = (\widehat{C}_t)_{12} = (\widehat{C}_t)_{21}\end{aligned}$$

Then we propose the following multivariate HAR model fitted on covariance matrix:

$$(\widehat{C}_{t+1})_{ij} = \phi_{ij}^{(0)} + \phi_{ij}^{(d)}(\widehat{C}_t)_{ij} + \phi_{ij}^{(w)}(\widehat{C}_{t:t-4})_{ij} + \phi_{ij}^{(m)}(\widehat{C}_{t:t-21})_{ij} \tag{4.3}$$

Note that this is equivalent to fitting each entry of the covariance matrix $\widehat{C_t}$ with a univariate HAR model. In addition, no transformation is made on the covariance matrix as the matrix could have negative entries on the off-diagonal positions, and therefore logarithmic transformation is infeasible. Several other forms of transformations such as cube root and hyperbolic transformation were tested, but none of these transformations seem to be very effective in normality correction. Therefore, the raw covariance matrix $\widehat{C}_t$ is directly used in model fitting. Model 4.3 is then first applied to forecast the covariance matrix, and the predicted matrix is then used to estimate the portfolio realized variance with Equation 4.2. This alternative method to estimate the portfolio variance is named as **HAR_$\widehat{\text{RV}}$**.

4.2.2. Diffusive-jump separation for multivariate process

Recall that in the univariate case, we have introduced quadratic and bipower variation and the relevant asymptotic distributions of the realized processes, and we further developed the method of diffusive-jump separation for realized variance forecasting. Barndorff-Nielsen and Shephard (2006b) further extended the definitions and the asymptotic properties to higher dimensions. Denote the prices of a p-dimensional vector of assets as:

$$Y_t = (Y_t^{(1)}, Y_t^{(2)}, \cdots, Y_t^{(p)})^T$$

By assuming that Y_t is a Brownian semimartingale plus finite activity jump process (Assumption 1.1), we have the following definition which is the multivariate version of Equation 1.2 (Barndorff-Nielsen and Shephard, 2004a):

Definition 4.2.2. The **quadratic variation (QV)** $p \times p$ **matrix process** can be written as

$$[Y]_t = \int_0^t \Sigma_s \, \mathrm{d}s + \sum_{0 \le s \le t} (\Delta J_s)(\Delta J_s)^T$$

where $\Sigma_s = \sigma_s \sigma_s^T$

When $\delta \downarrow 0$, one has the following asymptotic distribution for the daily realized covariance of day i (Barndorff-Nielsen and Shephard, 2004a):

$$\widehat{C}_i = \sum_{j=1}^{[1/\delta]} \left(\mathbf{y}_{j,i}\right)\left(\mathbf{y}_{j,i}\right)^T \xrightarrow{\mathbb{P}} \int_i^{i+1} \Sigma_s \, \mathrm{d}s + \sum_{i \le s \le i+1} (\Delta J_s)(\Delta J_s)^T \tag{4.4}$$

The multivariate realized bipower process can also be defined as follows:

Definition 4.2.3. The 1-1 order daily **realised bipower variation (BPV)** $p \times p$

matrix process of day i is

$$\{Y\}_i = \begin{pmatrix} \{Y^{(1)}\}_i & \{Y^{(1)}, Y^{(2)}\}_i & \cdots & \{Y^{(1)}, Y^{(p)}\}_i \\ \{Y^{(2)}, Y^{(1)}\}_i & \{Y^{(2)}\}_i & \cdots & \{Y^{(2)}, Y^{(p)}\}_i \\ \vdots & \vdots & \ddots & \vdots \\ \{Y^{(p)}, Y^{(1)}\}_i & \{Y^{(p)}, Y^{(2)}\}_i & \cdots & \{Y^{(p)}\}_i \end{pmatrix}$$

where the l, l-th element of $\{Y\}_i$ is

$$\left\{Y^{(l)}\right\}_i = \gamma_\delta \sum_{j=1}^{n=1/\delta} \left|y_{j-1,i}^{(l)}\right| \left|y_{j,i}^{(l)}\right|, \quad \gamma_\delta = \frac{1}{1-\delta}$$

and the l, k-th element is

$$\left\{Y^{(l)}, Y^{(k)}\right\}_i = \frac{\gamma_\delta}{4}\left(\left\{Y^{(l)} + Y^{(k)}\right\}_i - \left\{Y^{(l)} - Y^{(k)}\right\}_i\right)$$

(Barndorff-Nielsen and Shephard, 2006b), and it is further pointed out that under Assumption 1.1, as $\delta \downarrow 0$, one have the following two asymptotic properties:

$$\{Y\}_i \xrightarrow{\mathbb{P}} \mu_1^2 \int_i^{i+1} \Sigma_s \, \mathrm{d}s \quad \text{where} \quad \mu_1 = \sqrt{\frac{2}{\pi}} \tag{4.5}$$

$$\widehat{C}_i - \mu_1^{-2}\{Y\}_i \xrightarrow{\mathbb{P}} \sum_{i \le s \le i+1} (\Delta J_s)(\Delta J_s)^T \tag{4.6}$$

where Equation 4.5 and 4.6 suggest that the bipower variation separation method employed in the univariate case can again be applied to the multivariate forecasting model. We separate the diffusive and the jump components as below:

$$\begin{aligned} (\widehat{BV}_t)_{ij} &= (\mu_1^{-2}\{Y\}_t)_{ij} \quad \text{for} \quad 1 \le i, j \le p \\ (\widehat{J}_t)_{ii} &= \max\big(0, (\widehat{C}_t - \mu_1^{-2}\{Y\}_t)_{ii}\big) \quad \text{for} \quad 1 \le i \le p \\ (\widehat{J}_t)_{ij} &= (\widehat{C}_t - \mu_1^{-2}\{Y\}_t)_{ij} \quad \text{for} \quad i \ne j\,, \quad 1 \le i, j \le p \end{aligned}$$

Note that we only take the truncated $(\widehat{J}_t)_{ij}$ on diagonal entries as one needs to at least ensure non-negative jumps, and as for the off-diagonal entries, since we allow covariance to be negative, they will be remained as what they are. With the separation being set up, we could fit the continuous and discontinuous components separately with the multivariate HAR model. The combined model below is named as $\mathbf{HAR\_\widehat{BV}+\widehat{J}}$:

$$(\widehat{BV}_{t+1})_{ij} = \phi_{ij}^{(0)} + \phi_{ij}^{(d)}(\widehat{BV}_t)_{ij} + \phi_{ij}^{(w)}(\widehat{BV}_{t:t-4})_{ij} + \phi_{ij}^{(m)}(\widehat{BV}_{t:t-21})_{ij} \tag{4.7}$$

$$(\widehat{J}_{t+1})_{ij} = \beta_{ij}^{(0)} + \beta_{ij}^{(d)}(\widehat{J}_t)_{ij} + \beta_{ij}^{(w)}(\widehat{J}_{t:t-4})_{ij} + \beta_{ij}^{(m)}(\widehat{J}_{t:t-21})_{ij} \tag{4.8}$$

Since there is no transformation being made here, the estimated $\widehat{C}_{t+1}$ is simply computed by adding $\widehat{BV}_{t+1}$ and $\widehat{J}_{t+1}$. Similarly, we can also fit the multivariate dHAR model on $\widehat{\mathrm{J}}_t$ given the lack of persistence of the jump process:

$$(\widehat{J}_{t+1})_{ij} = \beta_{ij}^{(0)} + \beta_{ij}^{(d)}(\widehat{J}_t)_{ij} \tag{4.9}$$

Then Equation 4.9 is combined with 4.7 to form **HAR_$\widehat{\mathbf{BV}}$+$\widehat{\mathbf{dJ}}$** model. The four proposed models are then fitted and evaluated with both in-sample and out-of-sample experiments, and the results are discussed in the coming subsection.

4.2.3. Models summary and comparison

Table 4.1 summarises the evaluation metrics of the portfolio forecasting models. Recall that HAR_$\widetilde{\mathrm{RV}}$ uses the portfolio price to compute the raw portfolio variance and fit the univariate HAR model; HAR_$\widehat{\mathrm{RV}}$ fits the bivariate variance with the multivariate HAR model and compute the portfolio variance with covariance matrix; and HAR_$\widehat{\mathrm{BV}}$+$\widehat{\mathrm{J}}$ and HAR_$\widehat{\mathrm{BV}}$+$\widehat{\mathrm{dJ}}$ employs the diffusive-jump separation method in addition to the multivariate HAR model. Similar to the univariate case, we compare the MSE for both in-sample and out-of-sample forecasts. Besides, we also perform the one-side DM test between HAR_$\widetilde{\mathrm{RV}}$ and the rest three models where the p-values of the tests are reported in the last two columns of the table. We have the following observations from the table:

- For the in-sample prediction, the three models that involve covariance matrix all have MSE smaller than HAR_$\widetilde{\mathrm{RV}}$. This is further verified by the p-values of the in-sample DM tests where the tests are rejected at 5% level for HAR_$\widehat{\mathrm{BV}}$+$\widehat{\mathrm{J}}$ and HAR_$\widehat{\mathrm{BV}}$+$\widehat{\mathrm{dJ}}$, suggesting sufficient evidence to accept the alternative hypothesis that HAR_$\widehat{\mathrm{BV}}$+$\widehat{\mathrm{J}}$ and HAR_$\widehat{\mathrm{BV}}$+$\widehat{\mathrm{dJ}}$ have better predictive accuracy than HAR_$\widetilde{\mathrm{RV}}$. The DM test for HAR_$\widehat{\mathrm{RV}}$ also has a relatively small p-value, and the test will be rejected if the significance level is set to 10%. It could be concluded that for the in-sample portfolio variance forecast of this bivariate exchange rate dataset, the forecasts using the multivariate HAR model and covariance matrix which accounts for the inter-correlation between assets will outperform the univariate HAR model fitted on the raw portfolio variance. Besides, the diffusive-jump variation seems also to be effective under the multivariate case.

- However, for the out-of-sample forecast, the raw HAR_$\widetilde{\mathrm{RV}}$ is found to have the lowest MSE and none of the DM tests are rejected since the p-values are all relatively large. This might suggest that the multivariate HAR model and the use of covaraince matrix could be quite effective in improving goodness-of-fit, but when it comes to predictive ability, so far there is no sign showing that these methods will make a significant improvement on the original HAR_$\widetilde{\mathrm{RV}}$ model.

Then it could be summarized that for this bivariate exchange rate dataset, accounting for the covariance between assets with a multivariate HAR model helps to improve the prediction accuracy of portfolio variance. With the support of the asymptotic properties

of QV and BPV in the multivariate setting, the diffusive-jump separation method is again proven to be effective in portfolio variance forecasting.

Model	in-samp MSE	out-samp MSE	in-samp $DM_{1\text{-side}}$	out-samp $DM_{1\text{-side}}$
HAR_$\widetilde{RV}$	1.0603e-9	**2.9017e-10**	NA	NA
HAR_$\widehat{RV}$	9.5507e-10	3.0814e-10	0.0737	0.6917
HAR_$\widehat{BV}$+$\widehat{J}$	9.3202e-10	3.0462e-10	**0.0434**	0.6481
HAR_$\widehat{BV}$+$\widehat{dJ}$	**9.3052e-10**	3.0462e-10	**0.0313**	0.6481

Table 4.1.: Table to compare performance metrics of the univariate portfolio with three different multivariate portfolio models for $\widehat{RV}_t^{(P)}$

4.3. Bivariate Probabilistic Modelling

In this last section, we will briefly discuss the possible methods for probabilistic forecast in the bivariate context. Here, we move back from portfolio variance forecast to individual asset variance forecast. However, unlike the univariate case where univariate cumulative distribution functions (CDF) are fitted to the residuals of the models, under the multivariate context, multivariate CDF will be explored where the correlation between the residuals of the model for each individual asset is taken into account. Since now we are mainly concerned with the residuals of the variance point-forecast models, it means that we could again use the logarithmic transformation to approximate normal distribution. This helps to correct the skewness and kurtosis issues of the original dataset where such problems would be magnified under multivariate settings. In this section, two probabilistic forecasts will be explored - fitting residuals with multivariate generalized hyperbolic (GH) distributions and the employment of copula models. We will use the residuals from the basic HAR_RV models to apply and assess the performance of the two methods. Note that the bivariate dataset of EURUSD and AUDUSD exchange rates will still be applied in the case study.

4.3.1. Multivariate generalized hyperbolic distribution fitting

This method is similar to the univariate probabilistic forecast we have discussed in the last chapter where we find the best fitted GH distribution to the univariate variance forecast model. Below we have the bivariate HAR_RV for EURUSD and AUDUSD:

$$\log RV_{t+1}^{(1)} = \phi_1 + \phi_1^{(d)} \log RV_t^{(1)} + \phi_1^{(w)} \log RV_{t:t-4}^{(1)} + \phi_1^{(m)} \log RV_{t:t-21}^{(1)} + \epsilon_1 \tag{4.10}$$

$$\log RV_{t+1}^{(2)} = \phi_2 + \phi_2^{(d)} \log RV_t^{(2)} + \phi_2^{(w)} \log RV_{t:t-4}^{(2)} + \phi_2^{(m)} \log RV_{t:t-21}^{(2)} + \epsilon_2 \tag{4.11}$$

where ϵ_1 and ϵ_2 are the residuals of the two models respectively. Here, we aim to construct a probabilistic forecast for EURUSD and AUDUSD simultaneously by fitting

a multivariate GH distribution on (ϵ_1, ϵ_2). Similar to the univariate case, the point-forecast predicted by Model 4.10 and 4.11 $(\log RV_{t+1}^{(1)}, \log RV_{t+1}^{(2)})$ is then added to the location parameters of the fitted bivariate GH distribution to produce the multivariate distributional forecast for the two exchange rates.

We start with a quick look at the marginal distribution of ϵ_1 and ϵ_2. Figure 4.3 presents the density histograms of ϵ_1 and ϵ_2, and the solid lines represent the best fitted GH univariate distribution (both are variance gamma distribution) for the two residuals respectively. From both the histogram and the fitted density function, one can see that the residuals have similar distributions - asymmetric and long right tails. It is expected that such features will again be reflected in the fitted bivariate distribution.

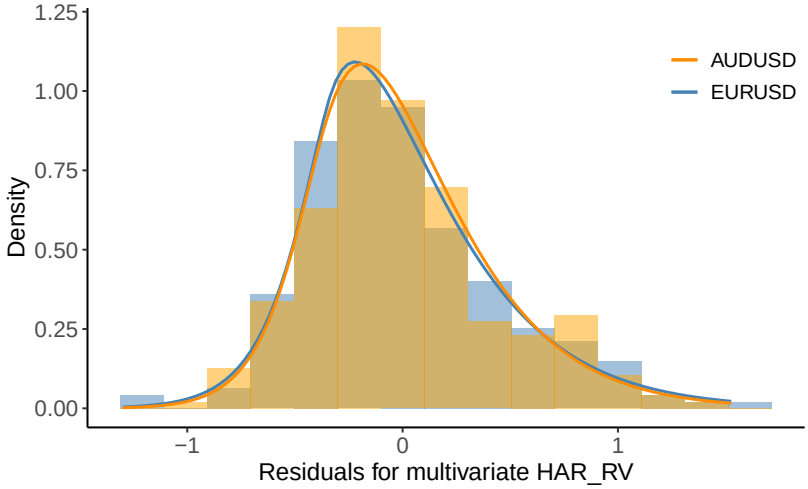


Figure 4.3.: Density histogram and the fitted marginal GH distribution for ϵ_1 and ϵ

The fitted multivariate GH distribution with the lowest AIC is the asymmetric Normal-Inverse Gaussian (NIG) distribution. To take a closer look at how well the GH distributions are fitted to the residuals, we have plotted the scatter plot together with the contour plots of several selected GH distributions. We first fit the residuals pair with the bivariate Gaussian distribution (AIC: 335.43) and the respective contour plot is presented in Figure 4.4. At first glance, it seems that most of the scatter points fall within the concentrated areas of the contour plot, but one could soon spot that the scatter points are more concentrated in the bottom-left area than the top-right area. However, the bivariate Gaussian contour plot is a symmetric distribution, and therefore it can be seen that the yellow area of the contour plot does not match the most concentrated areas for the scatter points. A similar issue can also be found in the left subplot of Figure 4.5 which gives the contour plot of the fitted symmetric NIG distribution (AIC: 318.67). Compared with the contours of Gaussian normal, the contours of symmetric NIG seem to be more compact. However, the symmetry problem still exists, and within the same contour lines, the concentration of scatter points at the bottom left is much greater than those in the top right corner. Last but not least, we shall turn our attention to the 'best' fitted multivariate GH distribution - asymmetric NIG distribution (AIC: 298.20), whose

contours are presented in the left subplot of Figure 4.5. From this graph, we can see that the fitted bivariate distribution is asymmetric with the highest density area (yellow area) being located slightly towards the bottom-left (matching the most concentrated areas of the scatter points) and long tails being found at the top-right end (trying to account for long-right tails for both residuals). It is noted that only when fitted distributions turn from symmetric to asymmetric, the contours can capture the two most important characteristics we observed from the marginal distributions.

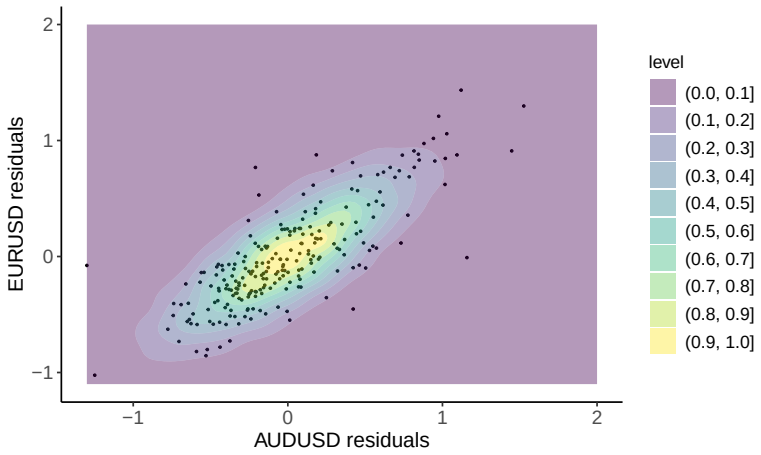


Figure 4.4.: Residuals scatter plot and the fitted bivariate Gaussian distribution

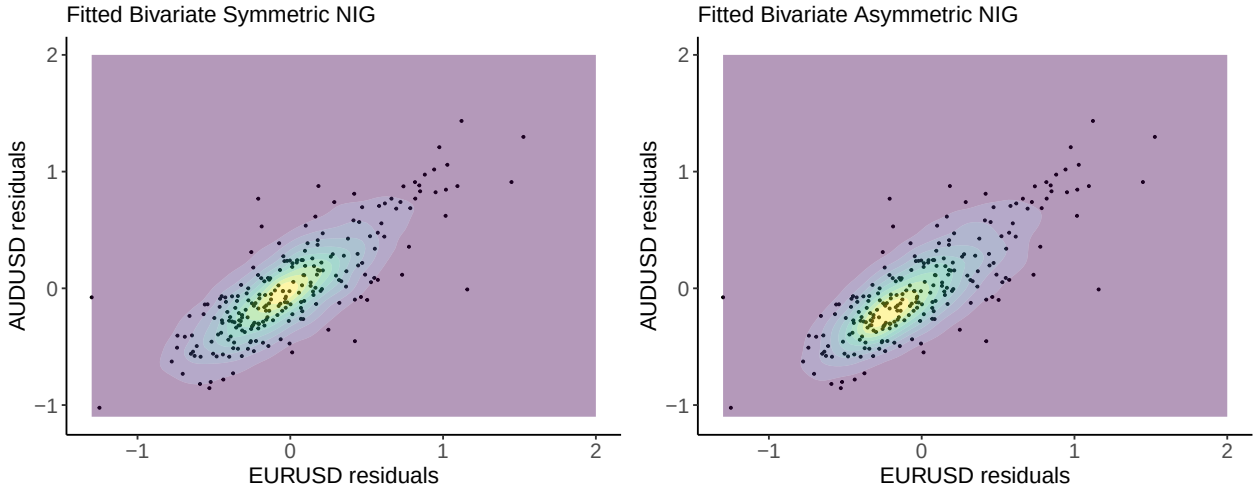


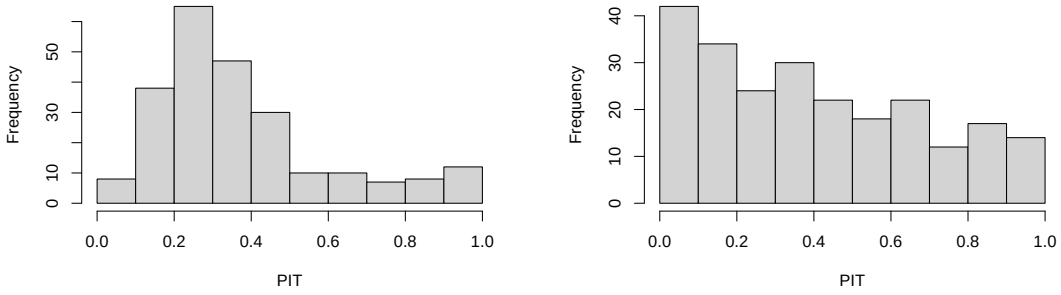
Figure 4.5.: Residuals scatter plot and the fitted bivariate NIG distributions

Then, analogue to the univariate case, we have performed in-sample and out-of-sample probability forecast with the residuals fitted bivariate GH distributions. The probabilistic forecast will again be constructed with bivariate Gaussian, asymmetric NIG and the 'best' fitted GH distribution at each time point. The steps for in-sample and out-of-sample forecasts are the same as in the univariate setting. To more formally access the

performance of the forecast with different fitted GH distributions, we applied the PIT histogram and the in-sample and out-of-sample LS score again (recall the definition of the relevant evaluation metrics in section 3.2).

Figure 4.6 shows the in-sample PIT graph for probabilistic forecasters with Gaussian (left) and NIG (right) distributions respectively. One can clearly see that the PIT diagram for forecasters using Gaussian distribution shows no sign of uniformity. The PIT of the one which uses the asymmetric NIG (best fitted in-sample GH) is more uniform than the left subplot, but the frequency of values between $0.0 - 0.2$ are too high to make the diagram resemble the uniform distribution. Table 4.2 reports the in-sample (top) and out-of-sample (button) average LS score of the probabilistic forecasters for $(\log RV_t^{(1)}, \log RV_t^{(2)})$, and here are the following observations:

- The probabilistic forecasters employing the Gaussian distribution for residual fitting gives the highest LS scores (poorest performance) for both in-sample and out-of-sample prediction.
- The probabilistic forecaster with asymmetric NIG distribution has the lowest LS score for both in-sample and out-of-sample cases. This is different from what we have observed in the univariate case where the 'best' column always gives the 'best' result with the lowerst LS score. This suggests that one could fix asymmetric NIG as the fitted bivariate distribution for the residuals, and only the parameters are updated each time. This would be much more computationally efficient than selecting the distribution with the lowest AIC by fitting every single GH distribution at every time point.



(a) bi-Gauss distribution

(b) bi-NIG distribution

Figure 4.6.: HAR_RV In-sample PIT histogram for multivariate distribution models

Model	gauss	NIG	'best'
bi_HAR	0.6924	**0.6004**	**0.6004**

(a) in-sample LS score

Model	gauss	NIG	'best'
bi_HAR	0.5718	**0.5361**	0.5384

(b) out-of-sample LS score

Table 4.2.: LS score for different probabilistic distribution forecast

4.3.2. Copula model

The second method we propose for multivariate distributional forecast is the use of copula functions. Copula is a commonly used tool to model multivariate distributions. It is a function used to describe the inter-correlation between random variables of their marginal distribution (Nelsen, 1998). Unlike our previous method a multivariate distribution is directly applied to the multivariate dataset, the copula method generally involves two steps - estimate the marginal distribution of each random variable and then estimate the copula which defines the dependence structure between the random variables. Below we have the formal definition for copula (Nelsen, 1998):

Definition 4.3.1. Consider a random variable $\mathbf{X} \in \mathbb{R}^n$, its CDF is the map $\mathbf{F} : \mathbb{R}^n \to [0,1]$ where $\mathbf{F}(\mathbf{x}) = \mathbb{P}(X_1 \leq x_1, \cdots, X_n \leq x_n)$ for any $\mathbf{x} = (x_1, \cdots, x_n) \in \mathbb{R}^n$. Assuming that $F_1, \cdots, F_n$ are all continuous functions, then the random vector $(F_1(x_1), \cdots, F_n(x_n))$ has uniform marginal distribution on $[0,1]$. The **copula** of $\mathbf{X}$ is defined as the joint-CDF of $\mathbf{F}(\mathbf{X})$:

$$C : [0,1]^n \to [0,1] \quad C(u_1, \cdots, u_n) = Pr(F_1(x_1) \leq u_1, \cdots, F_1(x_n) \leq u_n)$$

Sklar (1959) proposed the following key theorem in application of copulas:

Theorem 4.3.2. For any multivariate CDF $\mathbf{F}$, there exists a copula C such that

$$\mathbf{F}(\mathbf{x}) = C(F_1(x_1), F_2(x_2), \cdots, F_n(x_n)), \quad \text{for all} \quad \mathrm{x} \in \mathbb{R}^n$$

Here, we will employ one of the most common copula families - Archimedean copulas to fit our dataset and make the distributional forecast (Nelsen, 2006):

Definition 4.3.3. An **Archimedean copula C** is represented as:

$$C(u_1, \cdots, u_n; \theta) = \psi^{-1}(\psi(u_1; \theta) + \cdots + \psi(u_n; \theta); \theta)$$

where $\psi : [0,1] \to [0, \infty)$ is a continuous monotone decreasing function (as known as the 'generator') with $\psi(0) = 1$, and θ is the parameter of ψ.

Archimedean copulas are models with single parameter and therefore are popular in bivariate settings. In this project, we will apply three commonly-used bivariate Archimedean copulas - Clayton, Gumbel and Frank. Table 4.3 shows the generating function $\psi(\cdot)$, the inverse $\psi^{-1}(\cdot)$ and the bivariate copula models $C_\psi(u, v; \theta)$ for the three Archimedean copulas respectively (Nelsen, 2006).

Name	Generator ψ	Inverse ψ^{-1}	$C_\psi(u,v;\theta)$
Clayton	$\frac{1}{\theta}(t^{-\theta-1})$	$(1+\theta t)^{-\frac{1}{\theta}}$	$\left[\max(u^{-\theta}+v^{-\theta}-1,0)\right]^{-\frac{1}{\theta}}$
Gumbel	$(-\log(t))^\theta$	$\exp(-t^{\frac{1}{\theta}})$	$\exp\left\{-\left[(-\log u)^\theta+(-\log v)^\theta\right]^{\frac{1}{\theta}}\right\}$
Frank	$-\log\left(\frac{\exp(-\theta t)-1}{\exp(-\theta)-1}\right)$	$-\frac{1}{\theta}\log\left(1+\exp(-t)(\exp(-\theta)-1)\right)$	$-\frac{1}{\theta}\log\left[1+\frac{(\exp(-\theta u)-1)(\exp(-\theta v)-1)}{\exp(-\theta)-1}\right]$

Table 4.3.: Table showing the generators and equations for the Archimedean copulas

Then these three copulas will be applied to our model construction. Recall that before fitting the copula function, one first needs to fit the marginal distributions for each random variable. We will again experiment on two cases - fixing asymmetric VR distribution for both currency pairs (the best fitted distribution for in-sample dataset) and updating the 'best' fitted GH distributions at every time point. For each case, we will fit with the Clayton, Gumbel and Frank copula to produce the final multivariate distributions. Similar to all the previous analyses, we perform both the in-sample and out-of-sample forecasts and evaluate the results with PIT diagrams as well as LS scores. Figure 4.7 below presents the PIT histograms for the in-sample distributional forecast using the three copula functions. The three histograms are very similar and they do not show much sign of uniformity, suggesting that the forecasters are not very well calibrated. Therefore, a very similar problem is faced by both the GH residuals fitting methods and the copula models.

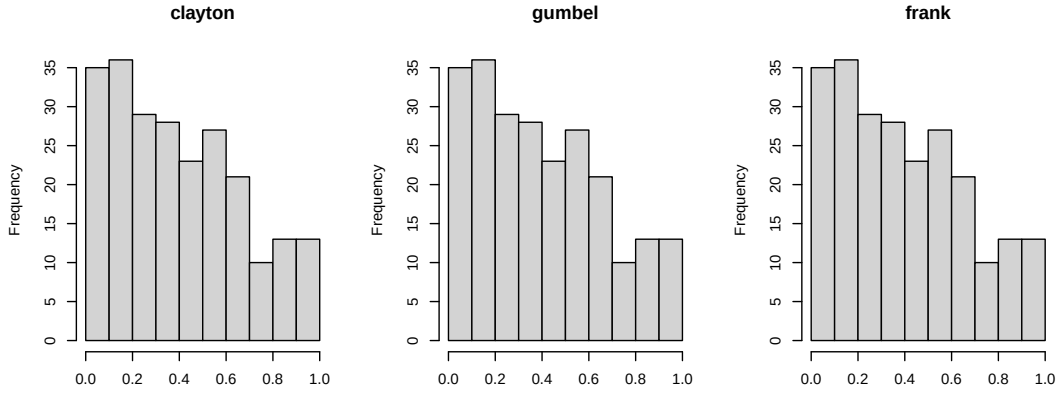


Figure 4.7.: PIT diagrams for the in-sample distributional forecast using different copula functions

Table 4.4 reports the LS scores for both in-sample and out-of-sample tests. The rows provide information on which marginal distributions are fitted to the individual residuals while the column tells which copula function is used. It can soon be observed from the

table that the Gumbel copula outperforms the other two by a great extent for both in-sample and out-of-sample forecasts. For out-of-sample prediction, the usage of the 'best' fitted GH distribution at every time point further improves the LS score. It is worth noting that the best LS scores of the copula model (in-sample: 0.5905; out-of-sample: 0.5303) are very similar to the best LS scores of the previous method found in Table 4.2 (in-sample: 0.6004; out-of-sample: 0.5361). This suggests that the two bivariate probabilistic forecasts have similar performance. At the same time, they also face a very similar problem - the calibration issue indicated by the PIT diagrams. We can notice from the PIT diagrams for both methods (Figure 4.6, 4.7) that the bins on the left (of lower density) tend to have a much higher frequency of occurrence. This could be because our fitted multivariate distributions still tend to overlook the long right tails for two residuals.

	Clayton	Gumbel	Frank
VR/'best'	0.8611	**0.5242**	0.6504

(a) in-sample LS score

	Clayton	Gumbel	Frank
VR	0.6106	**0.5455**	0.8696
'best'	0.6010	**0.5303**	0.8737

(b) out-of-sample LS score

Table 4.4.: LS score for different probabilistic distribution forecast

Figure 4.8 presents four plots with the contours of different fitted multivariate distributions. The plot on the 2nd row and 2nd column is the residuals scatter plot with the contours of the multivariate distribution with Gumbel copula, while the rest three plots are copied from Figure 4.4 and 4.5 from the last subsection for comparison purposes. Compared to the three contours of the fitted GH distributions, one can see that the outer contours of the distribution constructed with copula models tend to include more points in the top-right corner (the right long tails of the marginal distributions). However, there are still several scatter points that are quite far away from the contours and are left out. In addition, one can also observe from the bottom two subplots that the yellow contour areas are not seen to be the most concentrated area of the scatter points. There is very little difference between the concentrations of scatter plots between the inner contour layers. The lack of distinct high-density regions and the failure to account for extreme right tails together possibly led to the issues we noticed in the PIT diagrams. Therefore, further research is needed to improve the bivariate probabilistic modelling performance.

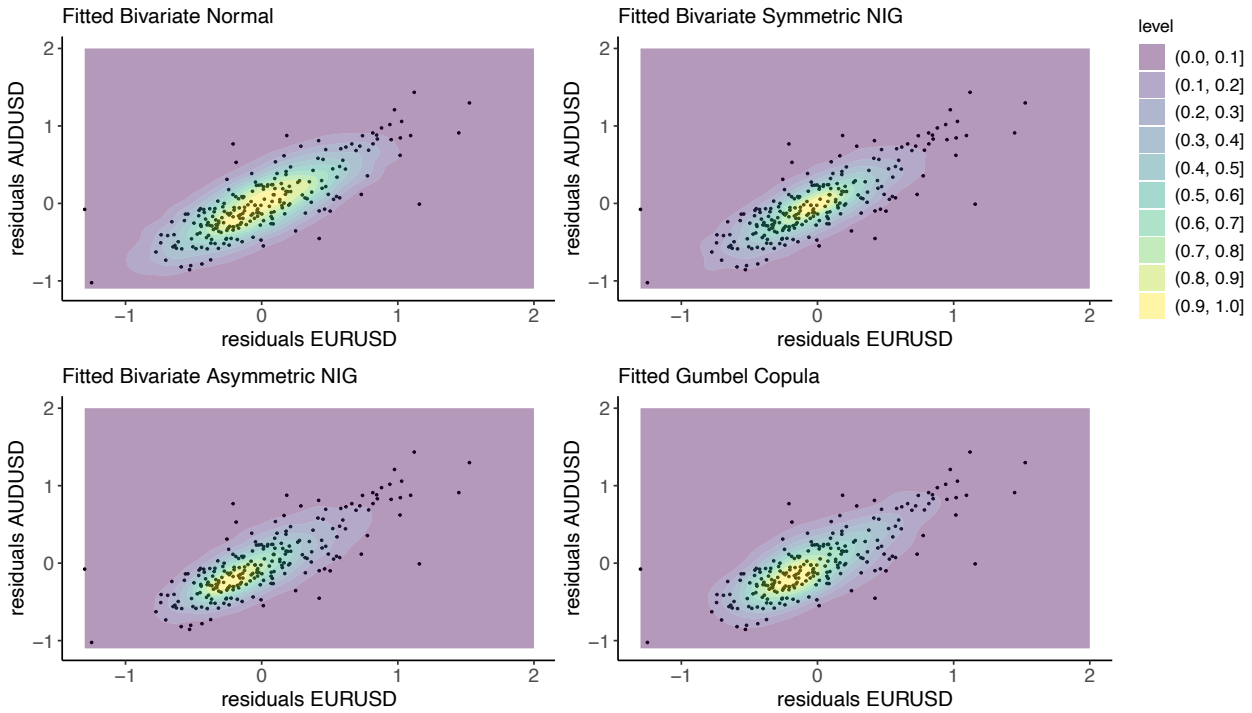


Figure 4.8.: Residuals scatter plot and the fitted bivariate distributions

5. Conclusion

5.1. Summary

In this project, we developed and assessed different models to forecast the realized variance of exchange rate data (EURUSD). We started with the univariate point forecast with the famous HAR model. Working on the theoretical asymptotic distribution results as well as the ratio jump analysis in Barndorff-Nielsen and Shephard (2004a, 2006a), we have proposed diffusive-jump separation methods where the diffusive volatility and jump components of the quadratic variation are separately modelled. Several separation methods were explored and the model fitting results suggest that separating based on the bipower variation or ratio jump test statistic are both effective in helping to improve the goodness-of-fit as well as the predictive performance of the forecast models. Then we move on to perform probabilistic forecasts with the point-forecast models. We have proposed a simple and easily implemented forecaster where generalized hyperbolic distributions are first fitted to the residuals of the point forecast models and then shifted by an amount equal to the point prediction to construct the distributional forecast on the transformed realized variance. This forecaster was verified to be well calibrated as long as the most fitted (from AIC selection) GH distributions are employed and updated at every time point.

We also expanded our discussion to the bivariate setting where another exchange rate (AUDUSD) was introduced. The multivariate HAR model which involves the covariance between assets is found to be more effective in predicting portfolio variance. In addition, the employment of the diffusive-jump separation method again helps to improve the in-sample prediction accuracy of portfolio variance. Last but not least, we have briefly discussed the probabilistic modelling in the bivariate context - residuals GH distribution fitting and copula models are both explored. We found that the best forecasters selected by the two methods produce very close scoring rule results.

5.2. Outlook

The ideas and results presented in this project could still induce further exploration and discussion. First, it was noted for a while that the jump components are not as persistent as the diffusive volatility, and therefore recall that the dHAR model 3.13 is used to model the jump components. However, neither HAR nor dHAR demonstrates good fitting for the jump components (adj.R^2 can be found in Table 3.3, 3.5, 3.6). This is actually under our expectation as the effect of jump components decays very quickly within one hour, hence for any time series model with daily basis data, it is

very hard to capture and effectively predict the movement of jump components. As we have noted in the results of the adjusted ratio jump test, jumps in the pricing process are usually associated with big news announcements, and it might be possible to model the jump components with features other than past volatility. For example, machine learning methods might be used with features including the occurrence of market news announcements and the importance of the news itself. One could be more flexible on the model and feature selection for the jump components modelling since it demonstrates very different properties from the diffusive volatility.

Second, for the multivariate forecasting part, so far only a bivariate dataset has been applied. However, an investment portfolio in reality usually includes more than two assets of varied weights. In future research, more complex models with higher dimensions could be explored to investigate the role of the covariance matrix. In addition, for the probability forecasting section, we have constructed forecasters with fitted GH distribution and copula models. However, both forecasters are not well calibrated given the PIT histograms, and therefore further adjustment is still required. At higher dimensions, vine copula could be used to describe more complex structures between the individual random variables to further improve the prediction accuracy of the multivariate probabilistic modelling.

Last but not least, different exchange rates can be used to test for generalization of the findings and results. In this project, to ensure consistency in analysis, we used EURUSD throughout the experiments. However, to verify the general effectiveness of the proposed techniques, different exchange rate data are suggested to be tested on the models to further verify the results.

A. Univariate point forecast model list

$$
\begin{aligned}
\text{HAR\_RV:}\quad & \log RV_{t+1} = \phi_0 + \phi^{(d)} \log RV_t + \phi^{(w)} \log RV_{t:t-4} + \phi^{(m)} \log RV_{t:t-21} \\
\text{SHAR\_RV:}\quad & \log RV_{t+1} = \phi_0 + \phi^{(d+)} \log RV_t^{+} + \phi^{(d-)} \log RV_t^{-} + \phi^{(w)} \log RV_{t:t-4} + \phi^{(m)} \log RV_{t:t-21} \\
\text{HAR\_RVc+RVd:}\quad & \log RV_{t+1}^{(c)} = \phi_0 + \phi^{(d)} \log RV_t^{(c)} + \phi^{(w)} \log RV_{t:t-4}^{(c)} + \phi^{(m)} \log RV_{t:t-21}^{(c)} \\
& \log^* RV_{t+1}^{(d)} = \beta_0 + \beta^{(d)} \log^* RV_t^{(d)} + \beta^{(w)} \log^* RV_{t:t-4}^{(d)} + \beta^{(m)} \log^* RV_{t:t-21}^{(d)} \\
\text{HAR\_RVc+dRVd:}\quad & \log RV_{t+1}^{(c)} = \phi_0 + \phi^{(d)} \log RV_t^{(c)} + \phi^{(w)} \log RV_{t:t-4}^{(c)} + \phi^{(m)} \log RV_{t:t-21}^{(c)} \\
& \log^* RV_{t+1}^{(d)} = \beta_0 + \beta^{(d)} \log^* RV_t^{(d)} \\
\text{SHAR\_RVc+RVd:}\quad & \log RV_{t+1}^{(c)} = \phi_0 + \phi^{(d+)} \log RV_t^{(c+)} + \phi^{(d-)} \log RV_t^{(c-)} + \phi^{(w)} \log RV_{t:t-4}^{(c)} + \phi^{(m)} \log RV_{t:t-21}^{(c)} \\
& \log^* RV_{t+1}^{(d)} = \beta_0 + \beta^{(d+)} \log^* RV_t^{(d+)} + \beta^{(d-)} \log^* RV_t^{(d-)} + \beta^{(w)} \log^* RV_{t:t-4}^{(d)} + \beta^{(m)} \log^* RV_{t:t-21}^{(d)} \\
\text{HAR\_BV+J:}\quad & \log BV_{t+1} = \phi_0 + \phi^{(d)} \log BV_t + \phi^{(w)} \log BV_{t:t-4} + \phi^{(m)} \log BV_{t:t-21} \\
& \log^* J_{t+1} = \beta_0 + \beta^{(d)} \log^* J_t + \beta^{(w)} \log^* J_{t:t-4} + \beta^{(m)} \log^* J_{t:t-21} \\
\text{HAR\_BV+dJ:}\quad & \log BV_{t+1} = \phi_0 + \phi^{(d)} \log BV_t + \phi^{(w)} \log BV_{t:t-4} + \phi^{(m)} \log BV_{t:t-21} \\
& \log^* J_{t+1} = \beta_0 + \beta^{(d)} \log^* J_t \\
\text{HAR\_adj.C+J:}\quad & \log \text{adj.}C_{t+1} = \phi_0 + \phi^{(d)} \log \text{adj.}C_t + \phi^{(w)} \log \text{adj.}C_{t:t-4} + \phi^{(m)} \log \text{adj.}C_{t:t-21} \\
& \log^* \text{adj.}J_{t+1} = \beta_0 + \beta^{(d)} \log^* \text{adj.}J_t + \beta^{(w)} \log^* \text{adj.}J_{t:t-4} + \beta^{(m)} \log^* \text{adj.}J_{t:t-21} \\
\text{HAR\_adj.C+dJ:}\quad & \log \text{adj.}C_{t+1} = \phi_0 + \phi^{(d)} \log \text{adj.}C_t + \phi^{(w)} \log \text{adj.}C_{t:t-4} + \phi^{(m)} \log \text{adj.}C_{t:t-21} \\
& \log^* \text{adj.}J_{t+1} = \beta_0 + \beta^{(d)} \log^* \text{adj.}J_t \\
\text{HAR\_RVJ:}\quad & \log RV_{t+1} = \phi_0 + \phi^{(d)} \log RV_t + \phi^{(w)} \log RV_{t:t-4} + \phi^{(m)} \log RV_{t:t-21} + \beta^{(d)} \log^* J_t \\
\text{HAR\_RVcd:}\quad & \log RV_{t+1} = \phi_0 + \phi^{(d)} \log \text{adj.}C_t + \phi^{(w)} \log \text{adj.}C_{t:t-4} + \phi^{(m)} \log \text{adj.}C_{t:t-21} + \beta^{(d)} \log^* \text{adj.}J_t \\
\text{HAR\_BVJ:}\quad & \log RV_{t+1} = \phi_0 + \phi^{(d)} \log BV_t + \phi^{(w)} \log BV_{t:t-4} + \phi^{(m)} \log BV_{t:t-21} + \beta^{(d)} \log^* J_t \\
\text{HAR\_adj.CJ:}\quad & \log RV_{t+1} = \phi_0 + \phi^{(d)} \log \text{adj.}C_t + \phi^{(w)} \log \text{adj.}C_{t:t-4} + \phi^{(m)} \log \text{adj.}C_{t:t-21} + \beta^{(d)} \log^* \text{adj.}J_t
\end{aligned}
$$

References

T. G. Andersen, T. Bollerslev, and F. X. Diebold. Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. *The Review of Economics and Statistics*, 89(4):701–720, 2007. ISSN 0034-6535.

O. E. Barndorff-Nielsen. Exponentially decreasing distributions for the logarithm of particle size. *Proceedings of the Royal Society. A, Mathematical, Physical, and Engineering Sciences*, 353(1674):401–, 1977. ISSN 1364-5021.

O. E. Barndorff-Nielsen and N. Shephard. Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society. Series B, Statistical methodology*, 64(2):253–280, 2002. ISSN 1369-7412.

O. E. Barndorff-Nielsen and N. Shephard. Econometric analysis of realized covariation: High frequency based covariance, regression, and correlation in financial economics. *Econometrica*, 72(3):885–925, 2004a. ISSN 0012-9682.

O. E. Barndorff-Nielsen and N. Shephard. Power and bipower variation with stochastic volatility and jumps (with discussion). *Journal of Financial Econometrics*, 2(1):1–48, 2004b. ISSN 1479-8409.

O. E. Barndorff-Nielsen and N. Shephard. Econometrics of testing for jumps in financial economics using bipower variation. *Journal of Financial Econometrics*, 4(1):1–30, 2006a. ISSN 1479-8409.

O. E. Barndorff-Nielsen and N. Shephard. Measuring the impact of jumps in multivariate price processes using bipower covariation. Working Paper, 2006b. URL `http://public.econ.duke.edu/~get/browse/courses/883/Spr16/COURSE-MATERIALS/Z_Papers/BNS_Bipower_multivariate.pdf`.

O. E. Barndorff-Nielsen, K. Silvia, and N. Shephard. Measuring downside risk: realised semivariance. *Volatility and Time Series Econometrics: Essays in Honor of Robert F. Engle, (Edited by T. Bollerslev, J. Russell and M. Watson)*, 4(1):117–136, 2010.

T. Bollerslev. Realized semi(co)variation: Signs that all volatilities are not created equal. *Journal of Financial Econometrics*, 20(2):219–252, 2022. ISSN 1479-8409.

T. Bollerslev, J. Li, A. J. Patton, and R. Quaedvlieg. Realized semicovariances. *Econometrica*, 88(4):1515–1551, 2020. ISSN 0012-9682.

G. E. P. Box and D. R. Cox. An analysis of transformations. *Journal of the Royal Statistical Society. Series B, Methodological*, 26(2):211–252, 1964. ISSN 0035-9246.

R. Carl, M. Bri‘ere, C. Alasseur, and M. Joseph. Expert Aggregation for Financial Forecasting. Papers 2111.15365, arXiv.org, Nov. 2021. URL https://ideas.repec.org/p/arx/papers/2111.15365.html.

K. Christensen. High frequency data econometrics, 2016. URL https://econ.au.dk/fileadmin/site_files/filer_oekonomi/subsites/creates/Diverse_2016/PhD_High-Frequency/Slides_day_2.pdf. Accessed on August 21, 2023.

R. Cont. Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1(2):223–236, 2001. ISSN 1469-7688.

F. Corsi. A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 7(2):174–196, 2009. ISSN 1479-8409.

F. X. Diebold and R. S. Mariano. Comparing predictive accuracy. *Journal of Business and Economic Statistics*, 13(3):253–, 1995. ISSN 0735-0015.

R. F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50(4):987–1007, 1982. ISSN 0012-9682.

P. Gaillard and Y. Goude. Forecasting electricity consumption by aggregating experts: How to design a good set of experts. In A. Antoniadis, J.-M. Poggi, and X. Brossat, editors, *Modeling and Stochastic Learning for Forecasting in High Dimensions*, pages 95–115, Cham, 2015. Springer International Publishing. ISBN 978-3-319-18732-7.

E. Ghysels, A. Harvey, and E. Renault. Stochastic volatility. *Handbook of Statistics*, 74: 128–198, 1996. URL https://EconPapers.repec.org/RePEc:mtl:montec:9613.

T. Gneiting and M. Katzfuss. Probabilistic forecasting. *Annual Review of Statistics and Its Application*, 1(1):125–151, 2014. doi: 10.1146/annurev-statistics-062713-085831. URL https://doi.org/10.1146/annurev-statistics-062713-085831.

T. Gneiting and A. E. Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378, 2007. ISSN 0162-1459.

T. Gneiting and R. Ranjan. Combining predictive distributions. *Electronic Journal of Statistics*, 7(none), 2013. ISSN 1935-7524.

J. Jacod. *Limit Theorems for Stochastic Processes*. Grundlehren der mathematischen Wissenschaften ; 288. Springer, Berlin, 1987. ISBN 3540178821.

W. Maneesoonthorn, G. M. Martin, C. S. Forbes, and S. D. Grose. Probabilistic forecasts of volatility and its risk premia. *Journal of Econometrics*, 171(2):217–236, 2012. ISSN 0304-4076. doi: https://doi.org/10.1016/j.jeconom.2012.06.006. URL https://www.sciencedirect.com/science/article/pii/S0304407612001534. Bayesian Models, Methods and Applications.

W. Marc, L. David, and B. Wolfgang. *ghyp: A package on generalized hyperbolic distributions*, 2022. URL `https://cran.r-project.org/web/packages/ghyp/vignettes/Generalized_Hyperbolic_Distribution.pdf`. Accessed: Aug 09, 2023.

H. M. Markowitz. *Portfolio Selection : Efficient Diversification of Investments*. Monograph /Cowles Foundation for Research in Economics at Yale University ; 16. Yale University Press, New Haven, 1970 - 1959. ISBN 0300013698.

R. B. Nelsen. *An Introduction to Copulas*. Lecture notes in statistics ; 139. Springer, New York, 1998. ISBN 0387986235.

R. B. Nelsen. *An Introduction to Copulas*. Springer Series in Statistics. Springer New York, New York, NY, 2nd ed. 2006. edition, 2006. ISBN 1-280-93833-1.

A. D. Roy. Safety first and the holding of assets. *Econometrica*, 20(3):431–449, 1952. ISSN 0012-9682.

A. Sklar. Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut Statistique de l'Université de Paris*, 8(8):229–231, 1959.

A. Thavaneswaran, Y. Liang, S. Das, R. K. Thulasiram, and J. Bhanushali. Intelligent probabilistic forecasts of vix and its volatility using machine learning methods. In *2022 IEEE Symposium on Computational Intelligence for Financial Engineering and Economics (CIFEr)*, pages 1–8, 2022. doi: 10.1109/CIFEr52523.2022.9776069.

A. Timmermann. Density forecasting in economics and finance. *Journal of Forecasting*, 19(4):231–234, 2000. ISSN 0277-6693.

O. Wintenberger. Optimal learning with bernstein online aggregation. *Machine Learning*, 106(1):119–141, 2017. ISSN 0885-6125.