THREE ESSAYS ON FINANCIAL VOLATILITY MODELING

THREE ESSAYS ON FINANCIAL VOLATILITY MODELING

By Efthymios NIKOLAKOPOULOS,

A Thesis Submitted to the School of Graduate Studies in the Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy

McMaster University
Doctor of Philosophy (2022)
Hamilton, Ontario (DeGroote School of Business)

TITLE: Three essays on financial volatility modeling
AUTHOR: Efthymios NIKOLAKOPOULOS (McMaster University)
SUPERVISOR: Dr. John M. MAHEU
COMMITTEE: Dr. Ronald J. BALVERS
COMMITTEE: Dr. Narat CHARUPAT
NUMBER OF PAGES: xiii, 132

Abstract

This thesis studies three important topics in modeling financial volatility. First, the jump clustering in ex post variance and its implications on forecasting, second, the underlying distribution of stochastic volatility and third, the role of non-Gaussian multivariate return distribution combined with a realized GARCH framework.

The first chapter is on variance jumps. Financial markets present unexpected and large jumps, due to unobserved news flow. I focus on modeling the ex post variance jumps, their time-dependent arrivals and their sizes. I use a discrete-time bivariate model, with two autoregressive components which capture the long and short-run memory of the ex post variance measures. I estimate contemporaneous and time-dependent jumps in the log-measures of realized variance and bipower variation. The results from S&P500 show that the variance jumps are frequent and persistent. I examine the ability of jumps to forecast returns and ex post variance densities over horizons of up to 50 days out-of-sample. Modeling jumps significantly improves ex post variance density forecasts for all horizons and improves forecasts of the returns density.

In the second chapter I explore the empirical non-Gaussian features of stochastic volatility. The standard assumption in a stochastic volatility specification is typically a restrictive Gaussian AR(1) structure. I drop this assumption and instead I assume that latent log-volatility follows an infinite mixture of normals with a Dirichlet process prior. The ex post measure of realized variance is used as a source of information to help identify the unknown distribution of log-volatility. Results from major stock indices show strong evidence of non-Gaussian distributional behaviour of volatility. The proposed framework captures asymmetry and thick tails in returns as well as realized variance. In out-of-sample forecasting, the new model provides improved density forecasts for returns, negative returns and log-realized variance.

In the third chapter a new approach for multivariate realized GARCH models is proposed. Two new extensions that have non-Gaussian innovations are developed. The first one is a parametric version, with multivariate-t innovations. The second one is a nonparametric approximation of the return distribution using an infinite mixture of multivariate normals given a Dirichlet process prior. The proposed models are based on the assumption that the realized covariance follows an Inverse Wishart distribution with conditional mean set to the conditional covariance of returns. The benefits of the proposed models are demonstrated from density forecasting and portfolio applications. Results from two equity datasets indicate that modeling the tail behaviour improves return density forecasting compared to the Gaussian assumption. The proposed models produce the least volatile global minimum variance portfolios out-of-sample and provide improved forecasts of Value-at-Risk and Expected Shortfall.

Acknowledgements

I would have never been able to complete this thesis without the guidance of my advisor and the committee members, the continuous support from my beloved family and last but not least, the help from all the wonderful friends I had the privilege to make throughout this journey.

Foremost, I would like to express my sincere gratitude and thank my advisor Professor John M. Maheu. This thesis would not have been possible without his patiently guidance and support. He has been a great mentor to me. He introduced me to the world of Bayesian econometrics and taught me how to be a scholar and researcher. Every discussion with him was a new lesson for me. I feel blessed that I had the chance to work with him.

I would also like to express my sincere gratitude and thank the committee members Professors Ronald J. Balvers and Narat Charupat for all their help and support during my studies and their useful comments on this thesis. A special mention to Dean C. Mountain who was a part of the committee but left this world so early and suddenly. It was a great pleasure to meet him. I also want to thank all the participants of the DeGroote School of Business Ph.D. seminars for their helpful comments on this thesis.

I would like to thank my parents Dimitri and Irene, and my brother George for all their support and help to pursue my goal. I also want to thank my great friends who have been close to me even though we live far. A special thanks to all my family in Canada who have been incredibly helpful ever since I moved there to start my studies.

Last but not least, I would like to thank all the friends I made down this road. Lulu, Parastoo, Firat, Armagan, Onur, Ramy, Javier, Robin, Zoe, Shu, Hamid and Reda, all of you guys made this journey special.

Contents

Abstract iii

Acknowledgements v

Declaration of Authorship xiii

Introduction xiv

1 Modeling jumps in ex post variance. Does it improve density forecasts? 1

1.1 Introduction 1

1.2 Ex post variance 4

1.3 Model specification 6

1.3.1 Heterogeneous variance jumps 8

1.3.2 Independent variance jumps 11

1.3.3 Density forecasting 13

1.4 Empirical application 16

1.4.1 Data 16

1.4.2 Selection of priors 17

1.4.3 Posterior estimation results 17

1.4.4 Forecasting results 19

1.5 Concluding remarks 20

2 Is stochastic volatility Gaussian? A Bayesian semiparametric analysis 34

2.1 Introduction . . . 34
2.2 Realized variance . . . 37
2.3 Model specification . . . 38
2.3.1 Estimation . . . 41
2.3.2 Benchmark models . . . 42
2.3.3 Predictive density . . . 43
2.4 Empirical application . . . 45
2.4.1 Data . . . 45
2.4.2 Selection of priors . . . 46
2.4.3 Posterior estimation results . . . 46
2.4.4 Forecasting results . . . 47
2.5 Concluding remarks . . . 49

3 The role of non-Gaussian innovations in multivariate realized GARCH models 68
3.1 Introduction . . . 68
3.2 Realized Covariance . . . 72
3.2.1 Data . . . 73
3.3 Multivariate realized GARCH . . . 73
3.3.1 Gaussian innovations . . . 73
3.3.2 Thick tailed innovations . . . 77
3.3.3 Semiparametric approach . . . 79
3.3.4 Restricted MRGARCH-DPM . . . 83
3.3.5 Predictive density . . . 85
3.3.6 Benchmark models . . . 88
3.4 Empirical applications . . . 90
3.4.1 Selection of priors . . . 90
3.4.2 Posterior estimation results . . . 90
3.4.3 Forecasting results . . . 92

3.4.4 Portfolio applications 93
3.5 Concluding remarks 98

4 Conclusion **110**

A **112**

A1 Sampling steps of 2Comp-RM-HJ 112
A2 Sampling steps of RSV-DPM-V 116
A3 Sampling steps of MRGARCH-DPM 121
A4 Covariance targeting 123

Bibliography **125**

List of Figures

1.1 Estimated jumps in **SPX-RV** from the **2Comp-RV-HJ** model. 25

1.2 Estimated jumps in **SPX-BP** from the **2Comp-BP-HJ** model. 26

1.3 (a) long-run and (b) short-run variance components of SPX-RV. 27

1.4 (a) long-run and (b) short-run variance components of SPX-BP. 28

1.5 One-step ahead log predictive densities for SPX (a) returns and (b) logRV. The densities are calculated by evaluating grids with the full sample posterior draws. 29

1.6 One-step ahead log predictive densities for SPX (a) returns and (b) logBP. The densities are calculated by evaluating grids with the full sample posterior draws. 30

1.7 h-ahead cumulative log predictive likelihoods for SPX (a) returns and (b) logRV. The likelihood evaluations are on the out-of-sample period: 30/12/2019 - 31/12/2021 (500 daily observations). 32

1.8 h-ahead cumulative log predictive likelihoods for SPX (a) returns and (b) logBP. The likelihood evaluations are on the out-of-sample period: 30/12/2019 - 31/12/2021 (500 daily observations). 33

2.1 **DAX** logRV and estimated log-volatility: $\mathbb{E}(h_t|\mathcal{I}_T)$. 54

2.2 **FTSE** logRV and estimated log-volatility: $\mathbb{E}(h_t|\mathcal{I}_T)$. 55

2.3 **SPX** logRV and estimated log-volatility: $\mathbb{E}(h_t|\mathcal{I}_T)$. 56

2.4 Mixture parameters for **DAX**. 57

2.5 Mixture parameters for **FTSE**. 58

2.6 Mixture parameters for **SPX**. 59

2.7 Log predictive densities of (a) **DAX returns**, (b) **DAX logRV** and (c) **DAX stochastic log-volatility** from RSV-N and RSV-DPM-V. 60
2.8 Log predictive densities of (a) **FTSE returns**, (b) **FTSE logRV** and (c) **FTSE stochastic log-volatility** from RSV-N and RSV-DPM-V. 61
2.9 Log predictive densities of (a) **SPX returns**, (b) **SPX logRV** and (c) **SPX stochastic log-volatility** from RSV-N and RSV-DPM-V.. 62
2.10 (a) **DAX** returns (top) and cumulative logBF of RSV-DPM-V vs SV-DPM and RSV-DPM-V vs RSV-N (bottom). (b) **DAX** logRV (top) and cumulative logBF of RSV-DPM-V vs RSV-N (bottom). 64
2.11 (a) **FTSE** returns (top) and cumulative logBF of RSV-DPM-V vs SV-DPM and RSV-DPM-V vs RSV-N (bottom). (b) **FTSE** logRV (top) and cumulative logBF of RSV-DPM-V vs RSV-N (bottom). 65
2.12 (a) **SPX** returns (top) and cumulative logBF of RSV-DPM-V vs SV-DPM and RSV-DPM-V vs RSV-N (bottom). (b) **SPX** logRV (top) and cumulative logBF of RSV-DPM-V vs RSV-N (bottom). 66

3.1 Log predictive density of the EWP returns. 101
3.2 Log determinant of estimated Covariance matrices. 102
3.3 Realized returns, forecasted 1% Value-at-Risk (top) and 1% Expected Shortfall (bottom) for the EWP of **IBM-MSFT-XOM**. 105
3.4 Realized returns, forecasted 1% Value-at-Risk (top) and 1% Expected Shortfall (bottom) for the EWP of **AXP-DD-GE-KO**. 106

List of Tables

1.1 Summary statistics of SPX 22

1.2 Posterior estimation results for **SPX-RV** data. 23

1.3 Posterior estimation results for **SPX-BP** data. 24

1.4 Cumulative log predictive likelihoods of out-of-sample forecasts for **SPX** returns and **logRV** on different forecasting horizons h. 31

1.5 Cumulative log predictive likelihoods of out-of-sample forecasts for **SPX** returns and **logBP** on different forecasting horizons h. 31

2.1 Summary statistics 50

2.2 Posterior estimation results for **DAX**. 51

2.3 Posterior estimation results for **FTSE**. 52

2.4 Posterior estimation results for **SPX**. 53

2.5 Density forecasting results. 63

2.6 Return density region forecasting results. 67

3.1 Summary statistics of IBM, MSFT and XOM. 99

3.2 Summary statistics of AXP, DD, GE and KO. 99

3.3 Posterior Means (and 95% Density Interval) for **IBM(1)-MSFT(2)-XOM(3)**. . 100

3.4 Forecasting results. 103

3.5 Portfolio forecasting results. 104

3.6 VaR violation rates. 107

3.7 Average out-of-sample losses. 108

3.8 Average out-of-sample variance of GMVP returns. 109

Declaration of Authorship

I, Efthymios NIKOLAKOPOULOS, declare that this thesis titled, "Three essays on financial volatility modeling" and the work presented in it are my own.

Introduction

This thesis is formed from three studies on financial volatility models. The studies have two common characteristics. First, they focus on joint models of financial returns and ex post measures of (co)variance. Second, they use parametric or semiparametric methods to explore deviations of the financial data from the Gaussian distribution. The emphasis of the newly developed time series models is to provide improvements in forecasting the distribution of returns. Also it is shown that modeling the non-Gaussian features of financial data is important for risk management and portfolio formation.

Modeling financial volatility is of high importance in the areas of asset pricing, portfolio optimization, risk measurement and management. However, volatility is unobserved in financial markets. It has to be estimated based on return observations. The most popular way to capture the unobserved volatility is with parametric models. The Autoregressive Conditional Heteroskedasticity (ARCH) framework introduced by Engle (1982) is the cornerstone of volatility models. Numerous extensions of this model have been introduced with the most popular being the GARCH of Bollerslev (1986). The key characteristic of these models and their extensions is that future variance is conditional and known based on past return information. Another parametric model based approach is the class of stochastic volatility models, introduced by Taylor (1982). In these, future volatility is unknown conditional on past information.

During the last two decades, the availability of high-frequency (intraday) returns gave access to nonparametric measures of ex post variance. The simplest measure is the realized variance which is defined as the summation of squared high-frequency returns for a fixed time interval. For instance, daily realized variance is constructed from high-frequency (e.g. 5-10 minutes) intraday returns. Realized variance was firstly used empirically by Andersen and Bollerslev (1998)

and was formalized econometrically by Barndorff-Nielsen and Shephard (2002a). They show that in a frictionless market, realized variance consistently estimates the underlying quadratic return variation.

In the first chapter of this thesis I focus on modeling the jumps of the ex post variance measures. Due to the nature of the financial markets, the unexpected news flow information causes prices to jump. Chan and Gray (2018) have shown that ex post variance presents jumps which are the impact of macroeconomic news announcements. Caporin et al. (2015) model ex post variance jumps and show that they can be explained by the credit default swap rates and the variance risk premium.

The contribution of the first chapter is on the forecasting benefits of modeling jump clustering in ex post variance. I use a bivariate discrete-time model from Maheu and McCurdy (2011), for returns and ex post variance. The model has two autoregressive components to capture the long and short-run persistence observed empirically in ex post variance. I extend this model by estimating contemporaneous and time-dependent ex post variance jumps with the methodology of Maheu and McCurdy (2008). The newly developed model captures: variance shocks, asymmetry and thick tails in the ex post variance distribution and thick tails in the return distribution.

I use the new model to estimate time-dependent variance jumps in the S&P500 equity index. I test empirically two ex post variance measures: realized variance and bipower variation (Barndorff-Nielsen and Shephard 2004). Between the two measures, realized variance gives better forecasts of returns. The results show large and persistent variance jumps which occur at least once in a week.

Modeling the time-dependent variance jumps provides improvements in forecasting returns and variance. I test the predictive power of the new model for out-of-sample horizons of up to 50 days. I focus on forecasting the whole density of returns and ex post variance. The results show that the time-dependent variance jumps provide significant improvements in forecasting the ex post variance density, for all horizons, and improve forecasts of the return density.

The ex post measures of variance are a useful source of information to improve parametric conditional variance estimation. Engle (2002b) used the lagged realized variance in a GARCH equation and showed that it contains significant information, more than lagged squared returns, for the conditional variance construction. Hansen et al. (2012) jointly model returns and realized variance to improve the GARCH constructed variance. Takahashi et al. (2009) use realized variance along with returns to help the estimation of stochastic volatility.

The second chapter of the thesis, is a semiparametric extension of the framework by Takahashi et al. (2009), to capture non-Gaussian features of returns and stochastic volatility. Jensen and Maheu (2010) have extended the discrete-time stochastic volatility model with a Dirichlet Process (Ferguson 1973) mixture of infinite normal kernels which captures the underlying unknown return distribution. Jensen and Maheu (2014) use the Dirichlet Process mixture to capture the asymmetric feedback of returns and stochastic volatility. Their model shows the existence of non-Gaussian features in stochastic volatility but does not include ex post variance data which are highly informative about returns variance.

In the second chapter, I use a Dirichlet Process mixture model of infinite normal kernels along with realized variance data to provide a flexible framework for estimating stochastic volatility. The newly develop model can capture asymmetry and thick tails in the distribution of returns and stochastic volatility. The model is empirically tested on major equity indices and the results show strong non-Gaussian behaviour in the estimated volatility. The plots of estimated time varying volatility of volatility show that this gives the model the necessary flexibility to capture shocks in realized variance, as well as extreme returns.

Predictive density plots of volatility show strong evidence of asymmetry and thick tails. Interestingly, the flexible volatility estimation captures asymmetry and tails of ex post variance. This is validated by out-of-sample forecasts, in which the proposed model outperforms the benchmark of Takahashi et al. (2009) in forecasting realized variance. The new model shows improvement in risk measurement as it performs better in forecasting the most volatile equity returns. Due to

the flexible mixture, estimation assigns more probability mass to the far left side of the return density than the benchmarks. For most of the series tested it produces better density forecasts when returns are negative or exceed losses of 1%.

The third chapter of the thesis is a study on modeling the covariance of financial returns with non-Gaussian extensions of multivariate realized GARCH models. For several decades multivariate extensions of the GARCH framework have been the workhorse models in estimating the conditional covariance of asset returns for portfolio and risk management applications. Recently, the ex post measures of covariance and correlation have been used by Hansen et al. (2014), Gorgi et al. (2019) and Archakova et al. (2019) to improve the construction of conditional covariance with GARCH models. However, they do not account for asymmetry and thick tails that are empirically observed in returns.

In the third chapter, I use a new approach to jointly model returns and realized covariance in a multivariate GARCH framework. The realized covariance is assumed to follow an Inverse Wishart distribution, an assumption based on the findings of Jin and Maheu (2016). For the returns distributional assumption, I use the multivariate Gaussian kernel and I also develop two new extensions that capture the empirically observed non-Gaussian features in returns. The first one is a multivariate Student-t distribution to capture thick tails in returns distribution. The second one is an approximation of the underlying returns distribution with a Dirichlet Process mixture of infinite multivariate normal kernels.

I apply the new models to two different equity datasets. The results show the existence of thick tails but little evidence of asymmetry in both datasets. The proposed models provide improved out-of-sample return forecasts. Compared to standard multivariate GARCH models, which do not include realized covariance information, the developed realized GARCH models give better return density forecasts. The two non-Gaussian specifications outperform the Gaussian assumption in forecasting the density of returns and realized covariance.

I test the models on financial applications which focus on risk measurement and portfolio

optimization. For an equal weight portfolio investment, the novel models outperform the benchmarks in forecasting Value-at-Risk and Expected Shortfall. The non-Gaussian models, improve the construction of the global minimum variance portfolio since they produce portfolio weights which give the least volatile realized returns. These results highlight the importance of non-Gaussian return distributional assumptions and realized covariance information in portfolio risk management.

The thesis is organized as follows: Chapter 1 is the the first essay, on modeling jumps in ex post variance, Chapter 2 is the second essay, on the semiparametric stochastic volatility, Chapter 3 is the third essay, on the non-Gaussian multivariate realized GARCH models, while Chapter 4 concludes.

Chapter 1

Modeling jumps in ex post variance. Does it improve density forecasts?

1.1 Introduction

Financial asset prices present unexpected and large discontinuities defined as jumps, due to unexpected news flow in the market. In this chapter, I focus on modeling jumps in ex post variance. I extend a bivariate model for returns and ex post variance in which I model contemporaneous and time varying jumps in the ex post measures of realized variance (RV) and bipower variation (BP). Results show frequent and persistent variance jumps. I test the ability of jumps to forecast returns and variance densities over horizons of up to 50 days out-of-sample. The heterogeneous jumps provide improvements in return density forecasting and strongly improve ex post variance density forecasts.

The jumps in financial returns have been extensively studied in the literature. The framework proposed by Press (1967) is the basic jump model used in finance. It accounts for multiple price jumps in a single period. The jumps arrive under a Poisson counting process and their size is independently normally distributed. This model has the capability to capture asymmetry and thick tails in returns. Oldfield Jr et al. (1977) extend the framework with autoregressive jump sizes. Chan and Maheu (2002) develop a discrete-time Poisson jump model with autoregressive

jump intensity (ARJI) and conditional jump dynamics combined with a GARCH specification (Engle 1982; Bollerslev 1986) for return volatility. The ARJI model has also been used by Maheu and McCurdy (2004), Chan and Feng (2012) and Maheu et al. (2013). These studies identify a form of dependence in the jump arrival times.

A simplification of the Poisson jump model was introduced by Ball and Torous (1983). They model stock price jumps, from news arrivals, with a Bernoulli jump process. The key characteristic of this specification is that only one significant jump is allowed per period. Johannes et al. (1999) found that the probability of a jump in equity indices depends on past jumps and the absolute value of returns. They also provide a methodology for extracting the jump times and sizes. Maheu and McCurdy (2008) show that a heterogeneous jump specification can explain the dynamics of conditional return distribution better than the standard variance models: GARCH and stochastic volatility (SV) (Taylor 1994).

Apart from returns, financial volatility is also characterized by instantaneous and persistent movements. Eraker et al. (2003) extend the continuous-time SV models to include Bernoulli jump processes in returns and latent volatility. They find strong jump evidence in both. Eraker (2004) extends these models with state dependent jump arrivals.

The development of the realized variance (RV) estimators provides consistent and model-free ex post data of the latent variance. The above models that have been used to estimate return jumps can also be used to estimate ex post variance jumps. Caporin et al. (2015) use the ARJI framework to estimate jumps in the bipower realized range. They also find that these ex post variance jumps can be explained by the credit default swap rates and the variance risk premium. Chan and Gray (2018) estimate volatility jumps and show that these coincide with news announcements.

How important are variance jumps to forecasts? What the literature is missing so far is a study for the forecasting information of the variance jumps. With this work I intend to estimate time-varying ex post variance jumps and examine if they provide improved returns and variance

forecasts. I focus on the financial index of S&P500 and predicting the whole distribution of returns and variance.

I use the 2 Component - Observable Stochastic Variance (**2Comp-OSV**) model of Maheu and McCurdy (2011) to capture the time series dynamics of ex post variance. Typically, the Heterogeneous Autoregressive (HAR) model of Corsi (2009) is used to capture the time dependence of the ex post variance measures. This simple time series model explains ex post variance movements based on its daily, weekly and monthly past values. Studies which use extensions of the HAR model are from Andersen et al. (2007), Bollerslev et al. (2009), Andersen et al. (2011), Corsi et al. (2012), Corsi and Renò (2012) and Caporin et al. (2015). Maheu and McCurdy (2011) find that their 2Comp-OSV, which uses long and short-run conditional variance components, explains RV better than a HAR equation and provides improved return density forecasts. The 2Comp-OSV is a suitable framework for multiperiod forecasts of variance and returns.

I extend the 2Comp-OSV to include time-varying ex post variance jumps. The focus of this chapter is only on the variance jumps and an extension with return jumps is left for future work. I use the heterogeneous Poisson jump model of Maheu and McCurdy (2008) to estimate variance jumps, their sizes and time-dependent arrivals. The distinct characteristic of this model is that jump arrivals are governed by a stochastic autoregressive process. This can be interpreted as the unobserved news arrival in the market. It has the capability to model jump persistence and jump clusters. The proposed model can explain the conditional distribution of ex post variance, by capturing tails and asymmetry. It also allows a thick tailed distribution for the returns.

I test empirically two ex post variance measures, realized variance and bipower variation. Results show that between the two measures, realized variance produces better return forecasts. I find large and persistent jumps in both the realized measures of S&P500 volatility. The proposed heterogeneous jump model shows that variance jumps appear at least once per week. A restricted version with constant jump intensity shows variance jumps on average once per week. Predictive

density estimation of log RV and BP shows that the two ex post measures have non-Gaussian distribution. I find strong thick tails and minor evidence of asymmetry. The heterogeneous jump model shows thick tails in returns distribution.

I examine the predictive performance of the jumps by multiperiod density forecasts over horizons of up to 50 days out-of-sample. The results show that the modelled heterogeneous jumps significantly improve ex post variance density forecasts for all the horizons. They also provide improvements in forecasting the returns density.

The chapter is organized as follows: Section 1.2 is a brief introduction to the ex post variance measures used, Section 1.3 presents the specification of the proposed models, their estimation steps and the forecasting process, Section 1.4 presents the empirical application results and Section 1.5 concludes.

1.2 Ex post variance

The ex post variance of an asset's returns is estimated with the nonparametric realized measures. Realized variance (RV) is the simplest measure. The important papers that provide a foundation of the theory and application of RV are from Andersen et al. (2001a), Andersen et al. (2001b), Barndorff-Nielsen and Shephard (2002a), Barndorff-Nielsen and Shephard (2002b) and Andersen et al. (2003). The theoretical foundation starts with a continuous-time stochastic volatility model. Including a jump component, as Press (1967), then the logarithmic asset price at time t, p_t, evolves under the following process

$$dp_t = \mu_t dt + \sigma_t dW_t + \xi_t dq_t, \tag{1.1}$$

where μ_t and $\sigma_t (\sigma_t > 0)$ are the stochastic drift and diffusion processes, W_t is a standard Brownian motion, q_t is a Poisson process, uncorrelated with W_t, with Prob $[dq_t = 1] = \lambda$, Prob $[dq_t = 0] = 1 - \lambda$ and ξ_t is the jump size. Under (1.1) jumps are finite, rare and their frequency depends on λ. The daily continuously compounded logarithmic return at time t, r_t, is

defined as

$$r_t = p_t - p_{t-1} = \int_{t-1}^{t} \mu_\tau d\tau + \int_{t-1}^{t} \sigma_\tau dW_\tau + \sum_{t-1<\tau<t|dq=1} \xi_\tau \tag{1.2}$$

and the quadratic return variation is defined as the summation of integrated variance (IV) and the cumulative squared return jump (RJ) term

$$\text{QV}_t = \text{IV}_t + \text{RJ}_t = \int_{t-1}^{t} \sigma_\tau^2 d\tau + \sum_{t-1\leq\tau\leq t|dq=1} \xi_\tau^2 \tag{1.3}$$

For day t, given intraday returns $r_{t,i}$, $i = 1, ..., n$, Barndorff-Nielsen and Shephard (2002a) and Andersen et al. (2003) show that QV_t can be approximated by the realized variance (RV) estimator defined as

$$\text{RV}_t = \sum_{i=1}^{n} r_{t,i}^2 \to \text{QV}_t, \quad \text{as } n \to \infty. \tag{1.4}$$

RV_t is a consistent estimator of QV_t under no market microstructure noise. In practice, high-frequency returns contain market microstructure noise which makes RV_t biased and inconsistent, as Zhang et al. (2005), Hansen and Lunde (2006) and Bandi and Russell (2008) document.

A popular way to overcome noise in RV_t is the use of *subsampling* as proposed by Zhang et al. (2005). With this statistical technique, the grid of available intraday observations is separated into multiple subgrids. For each subgrid, an estimator of realized variance is calculated. Then, RV_t is set as an average of the subgrid estimators. This gives an approximately unbiased RV_t. Aït-Sahalia and Mancini (2008) show that RV_t from *subsampling* outperforms the simple one, from the summation of squared log returns, in contained bias, variance, RMSE and forecasting.

A measure of the integrated return variance is proposed by Barndorff-Nielsen and Shephard (2004) who define the realized bipower variation (BP) which is an approximation to IV_t and is

calculated as

$$\text{BP}_t \equiv \frac{\pi}{2}\sum_{i=2}^{n}|r_{t,i}||r_{t,i-1}| \to \text{IV}_t, \quad \text{as } n \to \infty. \tag{1.5}$$

BP_t is robust to return jumps unlike RV_t which contains the return jump information. By subtracting BP_t from RV_t, return jumps are indirectly estimated. I use both measures to estimate ex post variance jumps and test empirically which measure gives better return density forecasts.

1.3 Model specification

As discussed in the introduction of the chapter, I use a bivariate discrete-time model specification, of returns and ex post variance. This integrated framework would enable modeling jumps (shocks) in the variance dynamics and examine how these affect (multiperiod) forecasts of ex post variance and returns.

The study of Maheu and McCurdy (2011) showed that a bivariate framework with conditional variance components (Maheu and McCurdy 2007) outperforms models with a HAR specification for logRV. The variance components decay at different rates which helps to capture long and short-run variance dependence. I use their proposed 2Comp-OSV model as the basis of this work. For the rest of this chapter I use the notation of **2Comp-RM** with RM being the realized measure of variance examined in the applications, RV and BP. The **2Comp-RM** model is defined as

$$r_t = \mu + \sqrt{\text{RM}_t}u_t, \quad u_t \sim \text{NID}(0,1), \tag{1.6}$$

$$\text{logRM}_t = \omega + \sum_{i=1}^{2}\phi_i c_{i,t} + \rho u_{t-1} + \sigma v_t, \quad v_t \sim \text{NID}(0,1), \tag{1.7}$$

$$c_{i,t} = (1-\alpha_i)\text{logRM}_{t-1} + \alpha_i c_{i,t-1}, \quad 0 < \alpha_i < 1, \quad i = 1,2, \quad \alpha_1 > \alpha_2, \tag{1.8}$$

with r_t being the log returns at time t, $t = 1, ..., T$, that follow a normal distribution with mean

μ and variance RM_t. RM_t is the realized measure of ex post variance at time t. Eq.(1.7) sets a Gaussian assumption for the natural logarithm (log) of RM with a constant variance σ^2. The mean of logRM is a structure based on the 2 conditional components $c_{i,t}$ with different decay rates α_i with $\alpha_1 > \alpha_2$.

Parameter ρ captures the feedback due to return shocks, which is often called leverage effect. The parameter is expected to have a negative sign. This implies that a return drop will cause increase in future variance and vice versa.

The 2Comp-RM model has a parsimonious and convenient specification. It can produce mean-reverting forecasts for returns and variance. Also, is suitable for multiperiod forecasts as done by Maheu and McCurdy (2011). They show that this specification outperforms other functions in the realized measure equation, in return density forecasting. One of these other functions is the HAR specification of Corsi (2009), which is typically used to model RM.

2Comp-RM can be estimated by standard Likelihood Maximization (MLE). In the following sections, I extend the 2Comp-RM to include jumps in logRM. I estimate these jumps under a Bayesian framework and hence I use the same for the 2Comp-RM. The estimation of model parameters in $\theta = \{\mu, \omega, \phi_1, \phi_2, \alpha_1, \alpha_2, \rho, \sigma^2\}$ is done by collecting a large sample (R) of draws $\{\theta\}_{l=1}^{R}$ from the posterior density. These draws are used for inference. For instance, from $\{\mu\}_{l=1}^{R}$, posterior moments can be calculated such as the posterior mean $\mathbb{E}(\mu|\mathcal{I}_T) = R^{-1}\sum_{l=1}^{R}\mu^{(l)}$ with $\mathcal{I}_T = \{r_{1:T}, \text{logRM}_{1:T}\}$.

Under the Bayes rule, model parameters in θ have the following posterior density

$$p(\theta|\mathcal{I}_T) \propto p(\theta)\prod_{t=1}^{T}\text{N}(r_t|\mu, \text{RM}_t)\,\text{N}\left(\text{logRM}_t\Big|\omega + \sum_{i=1}^{2}\phi_i c_{i,t} + \rho u_{t-1}, \sigma^2\right), \quad i = 1, 2, \quad (1.9)$$

with $p(.)$ being the prior density and $\text{N}(.|)$ the normal density. In this chapter, I use as prior the multivariate normal distribution, $\text{N}(0_k, 100 \times I_k)$, with k being the number of parameters in θ and I_k the identity matrix. The prior is uninformative in order to let the data drive the model estimation. The posterior in (1.9) does not have a known form. To estimate the parameters, I

collect $\{\theta^{(l)}\}_{l=1}^{R}$ draws from the above posterior with a random walk Metropolis-Hastings (MH) algorithm with a thick tailed proposal $\theta^{'}$ which is from

$$h(\theta^{'}) \sim \begin{cases} \mathrm{N}(\theta^{(l-1)}, \kappa \hat{V}_h), & \text{with probability } p = 0.9, \\ \mathrm{N}(\theta^{(l-1)}, 10\kappa \hat{V}_h), & \text{with probability } 1-p. \end{cases} \tag{1.10}$$

where $\hat{V}_h$ is the inverse Hessian matrix evaluated at the posterior mode $\hat{\theta}$, which is computed once at the beginning of estimation. Scaling $\hat{V}_h$ by κ helps to achieve a desired acceptance rate around 0.2-0.4 and explore the posterior. The draw θ' is accepted with probability

$$\min \left\{ p(\theta^{'}|\mathcal{I}_T)/p(\theta^{(l-1)}|\mathcal{I}_T), 1 \right\}. \tag{1.11}$$

Draws which do not satisfy: $0 < \phi_i, < 1, 0 < \alpha_i < 1, \alpha_1 > \alpha_2$ and $\sigma^2 > 0$ are rejected.

1.3.1 Heterogeneous variance jumps

As described in Section 1.2, return jumps are assumed to follow a Poisson process. This means that multiple jumps can occur from $t-1$ to t. A parsimonious discrete model for returns with an autoregressive Poisson jump process has been developed by Chan and Maheu (2002). Their model has been used by Caporin et al. (2015) to estimate jumps in bipower variation.

In this chapter I focus only on the ex post variance jumps. I extend the 2Comp-RM to include ex post variance jumps which arrive according to a heterogeneous Bernoulli process. The jump intensity is governed by a latent and stochastic autoregressive process. This can capture jump clustering and nests non-persistent (iid) jumps.

I use the model by Maheu and McCurdy (2008) to estimate ex post variance heterogeneous

(time-dependent) jumps (HJ) in the 2Comp-RM. Conditional on $\mathcal{I}_T$, the proposed **2Comp-RM-HJ** model is specified as follows

$$r_t = \mu + \sqrt{\text{RM}_t} u_t, \quad u_t \sim \text{NID}(0,1), \tag{1.12}$$

$$\text{logRM}_t = \omega + \sum_{i=1}^{2} \phi_i c_{i,t} + \rho u_{t-1} + J_t \xi_t + \sigma v_t, \quad v_t \sim \text{NID}(0,1), \tag{1.13}$$

$$c_{i,t} = (1-\alpha_i)\text{logRM}_{t-1} + \alpha_i c_{i,t-1}, \quad 0 < \alpha_i < 1, \quad i = 1,2, \quad \alpha_1 > \alpha_2, \tag{1.14}$$

$$\xi_t \sim \text{NID}(\mu_\xi, \sigma_\xi^2), \qquad J_t \in \{0,1\}, \tag{1.15}$$

$$\text{P}(J_t = 1|z_t) = \lambda_t \quad \text{and} \quad \text{P}(J_t = 0|z_t) = 1 - \lambda_t, \tag{1.16}$$

$$\lambda_t = \frac{\exp(z_t)}{1+\exp(z_t)}, \tag{1.17}$$

$$z_t = \gamma_0 + \gamma_1 z_{t-1} + e_t, \quad e_t \sim \text{NID}(0,1). \tag{1.18}$$

J_t is the jump indicator and ξ_t is the jump size which follows an independent normal distribution but can also be extended to have conditional mean and variance. The true J_t and ξ_t are unobserved. They are estimated using data information. If the data indicate a jump, then the posterior mean of J_t would be estimated close to 1. If there is no jump, the posterior mean of J_t would be approximately 0. In some occasions there is not a strong jump/no-jump signal from the data which will give $\mathbb{E}(J_t|\mathcal{I}_T) \in (0,1)$.

λ_t is the time-varying jump intensity which is governed by the latent autoregressive process z_t. Conditional on z_t a jump occurrence follows a Bernoulli distribution with probability λ_t. The logistic function in (1.17) ensures that z_t is mapped in the interval (0,1) for λ_t. z_t can be interpreted as the unobserved news flow into the market that impact the return variance. AR(1) parameter γ_1 captures the jump persistence and lies inside the unit circle, $|\gamma_1| < 1$.

Since RM_t is treated as an observable measure of variance, the 2Comp-RM-HJ model can also stand without the equation (1.12). Its purpose is to examine the effect of persistent variance jumps in forecasting the returns density. The jump component in (1.13) would produce different predictive density for logRM, than the Gaussian 2Comp-RM, since it can capture asymmetry

and thick tails. This can be seen by the first two conditional moments of logRM_t which, given the information set $\mathcal{I}_{t-1} = \{(r_1, \text{RM}_1), ..., (r_{t-1}, \text{RM}_{t-1})\}$ and z_t, are

$$\mathbb{E}(\text{logRM}_t|\mathcal{I}_{t-1}, z_t) = \omega + \sum_{i=1}^{2} \phi_i c_{i,t} + \rho u_{t-1} + \lambda_t \mu_\xi, \tag{1.19}$$

$$\text{Var}(\text{logRM}_t|\mathcal{I}_{t-1}, z_t) = \sigma^2 + \lambda_t \sigma_\xi^2. \tag{1.20}$$

Through the asymmetric and thick tailed forecasts of RM, the model is capable of producing thicker tailed return forecasts, compared to the 2Comp-RM.

Estimation

Conditional on the information set $\mathcal{I}_T$ and z_t, the $t-$th likelihood component of (1.13) is as a 2-state normal mixture model

$$p(\text{logRM}_t|\mathcal{I}_T, z_t) = \sum_{j=0}^{1} \text{N}\left(\text{logRM}_t \middle| \omega + \sum_{i=1}^{2} \phi_i c_{i,t} + \rho u_{t-1} + j\mu_\xi,\ \sigma^2 + j\sigma_\xi^2\right) \text{P}(J_t = j|z_t). \tag{1.21}$$

To estimate the model, we need to augment the parameters $\Theta = \{\theta, \mu_\xi, \sigma_\xi^2, \gamma\}$, where $\gamma = \{\gamma_0, \gamma_1\}$ and $\theta = \{\mu, \omega, \phi_1, \phi_2, \alpha_1, \alpha_2, \rho, \sigma^2\}$, with the latent vectors of the jump sizes $\xi = \{\xi_1, ..., \xi_T\}$, the jump indicators $J = \{J_1, ..., J_T\}$ and the jump intensities $z = \{z_1, ..., z_T\}$. These latent vectors need to be integrated out and this cannot be done with MLE. Alternatively, by applying the Bayes rule, we can get the following posterior density of the model parameters

$$p(\Theta, \xi, J, z|\mathcal{I}_T) \propto p(r|\text{logRM}, \Theta, \xi, J, z)\ p(\text{logRM}|r, \Theta, \xi, J, z)\ p(\xi, J, z|\Theta)\ p(\Theta). \tag{1.22}$$

By augmenting the parameter space with the auxiliary vectors I can estimate the jumps and obtain smoothed estimates of them from the posterior. Since there are no analytical results for the posterior in (1.22), I use Markov chain Monte Carlo (MCMC) techniques to get draws from

it. The MCMC methods deliver a set of posterior draws from a series of conditional distributions. This allows a straightforward estimation with a set of conditional steps. By collecting a large sample (R) of posterior draws for $\{\Theta, \xi, J, z\}_{l=1}^{R}$ we can draw inference for the model parameters.

The draws of $\{\Theta, \xi, J, z\}$ from the posterior are obtained through the following steps:

1. sample $p(\theta|\Theta_{-\theta}, r, \text{logRM}, z)$
2. sample $p(\mu_\xi|\sigma_\xi^2, \xi)$
3. sample $p(\sigma_\xi^2|\mu_\xi, \xi)$
4. sample $p(\xi_t|\sigma^2, \mu_\xi, \sigma_\xi^2, J_t, y_t), \quad t = 1, ..., T.$
5. sample $p(J_t|\sigma^2, \xi_t, \lambda_t, y_t), \quad t = 1, ..., T.$
6. sample $p(z_t|z_{-t}, \gamma, J_t), \quad t = 1, ..., T.$
7. sample $p(\gamma|\Theta_{-\gamma}, z)$

Details of the sampling steps are in appendix A1.

1.3.2 Independent variance jumps

I also develop a restricted version of the 2Comp-RM-HJ, in which jump arrivals in RM are independent. In this version, there is no jump persistence and jumps arrive randomly, with a

fixed probability λ. The **2Comp-RM-IJ** (Independent Jumps) is defined as

$$r_t = \mu + \sqrt{\text{RM}_t} u_t, \quad u_t \sim \text{NID}(0,1), \tag{1.23}$$

$$\text{logRM}_t = \omega + \sum_{i=1}^{2} \phi_i c_{i,t} + \rho u_{t-1} + J_t \xi_t + v_t, \quad v_t \sim \text{NID}(0, \sigma^2), \tag{1.24}$$

$$c_{i,t} = (1-\alpha_i)\text{logRM}_{t-1} + \alpha_i c_{i,t-1}, \quad 0 < \alpha_i < 1, \quad i = 1,2, \quad \alpha_1 > \alpha_2, \tag{1.25}$$

$$\xi_t \sim \text{NID}(\mu_\xi, \sigma_\xi^2), \qquad J_t \in \{0,1\}, \tag{1.26}$$

$$\text{P}(J_t = 1) = \lambda \quad \text{and} \quad \text{P}(J_t = 0) = 1-\lambda, \quad \lambda \in (0,1). \tag{1.27}$$

This jump model sets a Bernoulli (2-state) mixture of normal densities in logRM_t. It can capture thick tails and asymmetry in ex post variance. A simplified version of this jump structure has been introduced by Ball and Torous (1983) and applied to stock returns.

Estimation

Similar to the estimation of the 2Comp-RM-HJ model, conditional on $\mathcal{I}_{t-1}$, the $t-$th likelihood component of (1.24) is as a 2-state normal mixture model

$$p(\text{logRM}_t|\Theta,\lambda) = \sum_{j=0}^{1} \text{N}\left(\text{logRM}_t \middle| \omega + \sum_{i=1}^{2}\phi_i c_{i,t} + \rho u_{t-1} + j\mu_\xi,\ \sigma^2 + j\sigma_\xi^2\right) \text{P}(J_t = j|z_t). \tag{1.28}$$

To estimate the model, we need to augment the parameters $\Theta = \{\theta, \mu_\xi, \sigma_\xi^2, \lambda\}$, where $\theta = \{\mu, \omega, \phi_1, \phi_2, \alpha_1, \alpha_2, \rho, \sigma^2\}$. Following the Bayes rule, the model parameters have the following posterior

$$p(\Theta, \xi, J, \lambda|\mathcal{I}_T) \propto p(r|\text{logRM}, \Theta, \xi, J, \lambda)\ p(\text{logRM}|r, \Theta, \xi, J, \lambda)\ p(\xi, J, \lambda|\Theta)\ p(\Theta). \tag{1.29}$$

To collect MCMC draws for $\{\Theta, \xi, J\}_{l=1}^{R}$ from the above posterior I iterate through the following steps:

1. sample $p(\theta|\Theta_{-\theta}, r, \text{logRM})$
2. sample $p(\mu_\xi|\sigma_\xi^2, \xi)$
3. sample $p(\sigma_\xi^2|\mu_\xi, \xi)$
4. sample $p(\xi_t|\sigma^2, \mu_\xi, \sigma_\xi^2, J_t, y_t), \quad t = 1, ..., T.$
5. sample $p(J_t|\sigma^2, \xi_t, \lambda, y_t), \quad t = 1, ..., T.$
6. sample $p(\lambda|J)$.

The steps 1-5 are the same with the 2Comp-RM-HJ estimation steps with a constant intensity λ. Details of the sampling steps are in appendix A1.

1.3.3 Density forecasting

To examine the forecasting benefits of the modelled variance jumps, I focus on density forecasts. The target measure to estimate for the model comparison is the predictive likelihood, as suggested by Geweke (1994). For the following, I discuss the 2Comp-RM-HJ model but the computations can easily be modified for 2Comp-RM with no jumps and for 2Comp-RM-IJ with constant jump intensity λ.

Given the posterior draws $\Psi = \{\Theta^{(l)}, z^{(l)}\}_{l=1}^R$, conditional on the information set $\mathcal{I}_t = \{(r_1, \text{RM}_1), ..., (r_t, \text{RM}_t)\}$, the predictive density of logRM can be approximated as

$$\hat{p}(\text{logRM}_{t+1}|\mathcal{I}_t, \Psi) = \int p(\text{logRM}_{t+1}|\mathcal{I}_t, \Theta, z_t)\, p(\Theta, z_t|\mathcal{I}_t)\, d\Theta\, dz_t \tag{1.30}$$

$$= \int p(\text{logRM}_{t+1}|\mathcal{I}_t, \Theta, z_{t+1})\, p(z_{t+1}|z_t, \Theta)\, p(\Theta, z_t|\mathcal{I}_t)\, d\Theta\, dz_t\, dz_{t+1} \tag{1.31}$$

$$\approx \frac{1}{R}\sum_{l=1}^{R}\frac{1}{M}\sum_{m=1}^{M} p\left(\text{logRM}_{t+1}|\Theta^{(l)}, z_{t+1}^{(m)}, z_t^{(l)}, \mathcal{I}_t\right) \tag{1.32}$$

where l is the lth posterior draw from (1.22), $l = 1, ..., R$, and M simulated values of z_{t+1} with $m = 1, ..., M$. The density of logRM_{t+1} is

$$p\left(\text{logRM}_{t+1}|\Theta^{(l)}, z_{t+1}^{(m)}, z_t^{(l)}, \mathcal{I}_t\right) = \lambda_{t+1}^{(m)} \, \text{N}\left(\text{logRM}_{t+1}\Big|\omega^{(l)} + \sum_{i=1}^{2}\phi_i^{(l)} c_{i,t+1}^{(l)} + \rho^{(l)} u_t + \mu_\xi^{(l)}, \sigma^{2(l)} + \sigma_\xi^{2(l)}\right) + \left(1 - \lambda_{t+1}^{(m)}\right) \, \text{N}\left(\text{logRM}_{t+1}\Big|\omega^{(l)} + \sum_{i=1}^{2}\phi_i^{(l)} c_{i,t+1}^{(l)} + \rho^{(l)} u_t, \sigma^{2(l)}\right), \tag{1.33}$$

$$\text{with} \quad \lambda_{t+1}^{(m)} = \frac{\exp\left(z_{t+1}^{(m)}\right)}{1 + \exp\left(z_{t+1}^{(m)}\right)}, \tag{1.34}$$

$$z_{t+1}^{(m)} = \gamma_0^{(l)} + \gamma_1^{(l)} z_t^{(l)} + e_{t+1}, \quad e_{t+1} \sim \text{N}(0, 1), \tag{1.35}$$

$$c_{i,t+1}^{(l)} = (1 - \alpha_i^{(l)})\text{logRM}_t + \alpha_i^{(l)} c_{i,t}^{(l)}, \quad i = 1, 2, \tag{1.36}$$

$$\text{and} \quad u_t = \left(r_t - \mu^{(l)}\right) / \sqrt{\text{RM}_t}. \tag{1.37}$$

Similarly, the predictive density of returns can be approximated as

$$\hat{p}\left(r_{t+1}|\widehat{\text{RM}}_{t+1}, \mathcal{I}_t, \Psi\right) = \int p(r_{t+1}|\text{RM}_{t+1}, \mathcal{I}_t, \Theta) \, p\left(\text{logRM}_{t+1}, \Theta, z_t|\mathcal{I}_t\right) \, d\text{logRM}_{t+1} \, d\Theta \, dz_t \tag{1.38}$$

$$\approx \frac{1}{R}\sum_{l=1}^{R}\frac{1}{M}\sum_{m=1}^{M} \text{N}\left(r_{t+1}\Big|\mu^{(l)}, \widehat{\text{RM}}_{t+1}^{(m)}\right), \tag{1.39}$$

where $\widehat{\text{RM}}_{t+1}^{(m)} = \exp(x)$ and x is drawn as

$$x \sim \lambda_{t+1}^{(m)} \, \text{N}\left(\omega^{(l)} + \sum_{i=1}^{2}\phi_i^{(l)} c_{i,t+1}^{(l)} + \rho^{(l)} u_t + \mu_\xi^{(l)}, \sigma^{2(l)} + \sigma_\xi^{2(l)}\right) + \left(1 - \lambda_{t+1}^{(m)}\right) \, \text{N}\left(\omega^{(l)} + \sum_{i=1}^{2}\phi_i^{(l)} c_{i,t+1}^{(l)} + \rho^{(l)} u_t, \sigma^{2(l)}\right) \tag{1.40}$$

The predictive densities are the building blocks for the calculation of the predictive likelihood (PL) with which I evaluate each model's forecasting ability out-of-sample. Specifically, for each

model, by performing a set of τ (with $1 < \tau < T$), recursive posterior estimations, the log predictive likelihood of $\text{logRM}_{T-\tau+1}, ..., \text{logRM}_T$, is the summation of individual log-predictive densities of each estimation

$$\text{logPL}_v(\text{logRM}_{T-\tau+1:T}|\mathcal{I}_T, \Psi) = \sum_{t=T-\tau}^{T-1} \log\left(\hat{p}(\text{logRM}_{t+1}|\mathcal{I}_t, \Psi)\right), \tag{1.41}$$

where the predictive density is given from (1.32). Similarly, the log predictive likelihood for returns, $r_{T-\tau+1}, ..., r_T$, is calculated observation by observation as

$$\text{logPL}_r(r_{T-\tau+1:T}|\text{RM}_{t+1}, \mathcal{I}_T, \Psi) = \sum_{t=T-\tau}^{T-1} \log\left(\hat{p}\left(r_{t+1}|\widehat{\text{RM}}_{t+1}, \mathcal{I}_t, \Psi\right)\right), \tag{1.42}$$

with the density given from (1.39).

The model structure of 2Comp-RM along with the heterogeneous jump process specification, are suitable for multiperiod forecasts. For different forecasting horizons h, $h = 1, ..., H$, the density forecasts are evaluated over an identical set τ of out-of-sample realized measures $\text{logRM}_{T-\tau+1}, ..., \text{logRM}_T$ and returns $r_{T-\tau+1}, ..., r_T$. The cumulative log predictive likelihoods for logRM and returns for a forecasting horizon h, are calculated as

$$\text{logPL}_{h,v}(\text{logRM}_{T-\tau+1:T}|\mathcal{I}_T, \Psi) = \sum_{t=T-\tau-h}^{T-h} \log\left(\hat{p}_h(\text{logRM}_{t+h}|\mathcal{I}_t, \Psi)\right), \tag{1.43}$$

$$\text{logPL}_{h,r}(r_{T-\tau+1:T}|\mathcal{I}_T, \Psi) = \sum_{t=T-\tau-h}^{T-h} \log\left(\hat{p}_h(r_{t+h}|\widehat{\text{RM}}_{t+h}, \mathcal{I}_t, \Psi)\right). \tag{1.44}$$

The following steps are used to calculate the above cumulative likelihoods:

1. Conditional on $\mathcal{I}_t = \{(r_1, \text{RM}_1), ..., (r_t, \text{RM}_t)\}$, obtain a posterior draw (l) of $\Psi = \{\Theta, z\}$

2. For $h = 1, ..., H$ repeat:

(a) Simulate M values of $z_{t+h}^{(m)} = \gamma_0^{(l)} + \gamma_1^{(l)} z_{t+h-1}^{(m)} + v_{t+h}, \quad v_{t+h} \sim N(0,1), \quad m = 1, ..., M$.

(b) Propagate $c_{i,t+h}^{(m)} = (1 - \alpha_i^{(l)}) \log\widehat{\text{RM}}_{t+h-1}^{(m)} + \alpha_i^{(l)} c_{i,t+h-1}^{(m)}$.

(c) Evaluate $p_h^{(m)} \left(\log\text{RM}_{t+h} | \mathcal{I}_t, \Theta, z_{t+h}^{(m)}, c_{i,t+h}^{(m)} \right)$ and save the result.

(d) Simulate $\widehat{\text{RM}}_{t+h}^{(m)} = \exp(x)$ with x drawn from the jump/no-jump mixture in (1.40).

(e) Evaluate $p_h^{(m)} \left(r_{t+h} | \widehat{\text{RM}}_{t+h}^{(m)}, \mathcal{I}_t, \Theta \right)$ and save the result.

(f) Simulate $\hat{r}_{t+h}^{(m)} \sim \text{N} \left(\mu^{(l)}, \widehat{\text{RM}}_{t+h}^{(m)} \right)$.

(g) Save $\hat{u}_{t+h}^{(m)} = \left(\hat{r}_{t+h}^{(m)} - \mu^{(l)} \right) / \sqrt{\widehat{\text{RM}}_{t+h}^{(m)}}$.

3. Save $p_h^{(l)}(\log\text{RM}_{t+h} | \mathcal{I}_t, \Psi) = \frac{1}{M} \sum_{m=1}^{M} p_h^{(m)} \left(\log\text{RM}_{t+h} | \mathcal{I}_t, \Theta, z_{t+h}^{(m)}, c_{i,t+h}^{(m)} \right)$ and $p_h^{(l)} \left(r_{t+h} | \widehat{\text{RM}}_{t+h}, \mathcal{I}_t, \Psi \right) = \frac{1}{M} \sum_{m=1}^{M} p_h^{(m)} \left(r_{t+h} | \widehat{\text{RM}}_{t+h}^{(m)}, \mathcal{I}_t, \Theta \right)$.

After repeating the above steps R times we can calculate the predictive likelihoods

$$\hat{p}_h(\log\text{RM}_{t+h} | \mathcal{I}_t, \Psi) = \frac{1}{R} \sum_{l=1}^{R} p_h^{(l)}(\log\text{RM}_{t+h} | \mathcal{I}_t, \Psi), \tag{1.45}$$

$$\hat{p}_h(r_{t+h} | \widehat{\text{RM}}_{t+h}, \mathcal{I}_t, \Psi) = \frac{1}{R} \sum_{l=1}^{R} p_h^{(l)} \left(r_{t+h} | \widehat{\text{RM}}_{t+h}, \mathcal{I}_t, \Psi \right), \tag{1.46}$$

which are used to calculate the cumulative log predictive likelihoods in (1.43) and (1.44).

1.4 Empirical application

1.4.1 Data

Data are daily open-to-close log returns of S&P500 Index (**SPX**) and the ex post variance measures of realized variance (**RV**) and bipower variation (**BP**) for the period: January 3rd, 2000 - December 31st, 2021 (5,515 daily observations). Summary statistics are in Table 1.1. The data are obtained from Oxford-Man Institute's Realized Library of Heber et al. (2009). The realized

measures have been calculated from 5-minute returns using *subsampling* in order to reduce the market microstructure noise. The returns have been converted to percentages while RV and BP data have been scaled by 100^2 .

1.4.2 Selection of priors

For the application of the models, I select uninformative priors, in order to let the data drive the parameter estimation. Returns mean has a normal prior, $\mu \sim \text{N}(0, 100)$. The same is for the constant of logRM, $\omega \sim \text{N}(0, 100)$ and the leverage effect, $\rho \sim \text{N}(0, 100)$. Variance component parameters have restricted normal priors, $\phi_i \sim \text{N}(0, 100)\mathbf{1}_{\{0<\phi_i<1\}}$ and $\alpha_i \sim \text{N}(0, 100)\mathbf{1}_{\{0<\alpha_i<1\}}$, $i = 1, 2$, with $\alpha_1 > \alpha_2$, for identification. The variance parameter σ^2 has an inverse gamma prior, $\sigma^2 \sim \text{IG}(3/2, 1/2)$.

For the jump size, the mean has a normal prior, $\mu_\xi \sim \text{N}(0, 100)$ and the variance is from, $\sigma_\xi^2 \sim \text{IG}(3/2, 1/2)$. The constant jump intensity parameter has a beta prior, $\lambda \sim \text{B}(1, 1)$. Lastly, the parameters in γ have a bivariate normal prior, $\gamma \sim \text{N}(0, 100 \times I_2)\mathbf{1}_{\{|\gamma_1|<1\}}$.

1.4.3 Posterior estimation results

Results from $R = 10,000$ MCMC posterior draws, after $5,000$ burnin, for SPX-RV and SPX-BP are in Tables 1.2 and 1.3.

Regarding the variance jumps, the jump intensity λ, from 2Comp-RM-IJ, is 0.2359 in RV and 0.2141 in BP. This means one variance jump per week, on average. The unconditional z_t of 2Comp-RM-HJ, calculated from posterior means, is -0.947 for RV and -0.617 for BP. The unconditional jump intensity, calculated from the posterior average of λ_ts, is approximately 0.4082 for RV and 0.4509 for BP. It is more likely to observe a jump in BP variation of S&P500 since its jumps are more persistent. The AR(1) parameter γ_1, which captures the jump persistence, is estimated as 0.9799 in BP and 0.957 in RV.

The frequent and persistent variance jumps are not only due to a volatile market. There are frequent variance drops as well. Figures 1.1 and 1.2 display the modelled jumps (size, occurrence, intensity) in RV and BP from the 2Comp-RM-HJ model. The jumps have a low but significantly positive mean, μ_ξ, 0.0988 in RV and 0.0457 in BP. The model is capable of estimating both positive and negative variance discontinuities. The 2Comp-RM-IJ model puts more focus on the positive variance shocks. It estimates jumps with average size 0.2833 in RV and 0.2020 in BP. Based on the theoretical foundation of the two measures, the modelled jumps in BP are shocks in the variance. In RV they are shocks in variance and/or returns.

The time-dependent specification of the 2Comp-RM-HJ, provides less volatile estimation of the jumps, compared to the constant jump probability. Specifically, the estimated jumps in logRV, from the 2Comp-RM-HJ, have a variance, σ^2_ξ, of 0.3748 which is significantly less than the estimated variance of 0.4257, from the 2Comp-RM-IJ. Similarly, the independent jump variance of logBP is 0.4009 and drops to 0.3244 in the heterogeneous jump specification.

The modelled contemporaneous jumps improve the ex post variance data fit. The variance parameter, σ^2, of logRV is 0.3453 in the Gaussian specification, without jumps. With modelled independent jumps, σ^2 drops to 0.2328 and with heterogeneous jumps declines further to 0.1903. Similar results are observed for the variance of logBP.

Figures 1.3 and 1.4 display the conditional components of RV and BP respectively. Component c_1 is the smooth one and captures the long-run movements of variance. It is persistent with its coefficient α_1 estimated from 0.7965 to 0.826 in RV and from 0.7849 to 0.815 in BP. The second variance component captures the short-run variance dynamics and is less persistent. The coefficient α_2 is estimated from 0.3605 to 0.4029 in RV and from 0.2766 to 0.3039 in BP. This component is strongly affected by the lag of the realized measure. In the models has been imposed the restriction of $\phi_1 = \phi_2$ as this has been found to give the best forecasts, similar to the findings of Maheu and McCurdy (2011).

1.4.4 Forecasting results

Regarding the forecasted distribution of the realized measures, Figures 1.5 and 1.6 display the models one-step ahead out-of-sample logarithmic predictive densities from the full sample posterior estimation. For both the realized measures, the jump models show deviations from the Gaussian shape produced by the 2Comp-RM model. Both logRV and logBP have thick tails and asymmetry. These findings support the results of Barndorff-Nielsen and Shephard (2002a) that empirically RV is not log-normally distributed.

By forecasting the realized measures one-step ahead we can see how the probability of variance jumps impact the market return predictive density. The 2Comp-RM-HJ, with time-dependent jump probability, accounts for thick tails in returns log predictive density, compared to the Gaussian benchmark 2Comp-RM. A constant jump probability, in the 2Comp-RM-IJ, does not affect the tails of the returns predictive distribution.

The impact of the modelled ex post variance jumps in forecasting is demonstrated by density forecasts of 500 out-of-sample daily observations from December 30th, 2019 to December 31st, 2021. I test the models on multiperiod forecasts with horizons of up to 50 days ahead. For each run, I collect $R = 10,000$ posterior draws after $5,000$ burnin and set $M = 100$. Tables 1.4 and 1.5 report the cumulative log predictive likelihoods of market SPX returns, logRV and logBP for the forecasting horizons $h = 1, 5, 10, 50$. Figures 1.7 and 1.8 display the cumulative log predictive likelihoods for all the forecasting horizons h.

The modelled ex post variance jumps provide improved density forecasts of the realized measures. The flexible 2Comp-RM-HJ gives the best forecasts of both logRV and logBP for all the forecasting horizons. The 2Comp-RM-IJ gives the second best forecasts. These results highlight the importance of accounting for volatility shocks in variance forecasts. Since both jump models are mixtures of normals the results validate that the logarithmic ex post variance is not normally distributed.

The modelled heterogeneous jumps in ex post variance give improvements in return density

forecasts. When RV is used as the observable variance, all models give marginally the same returns predictive likelihoods in one-step ahead forecasts (h = 1). For longer horizons, 2Comp-RM-HJ performs the best among the models. When BP is used as the observable variance, 2Comp-RM-HJ outperforms the rest in return forecasts for the forecasting horizons of up to 30 days. The empirical comparison between the two realized measures shows that RV gives marginally better return predictive likelihoods compared to BP.

The 2Comp-RM-IJ model is similar to the 2Comp with a two-mixture of normals used by Maheu and McCurdy (2011). Their model does not provide improved forecasts of S&P500 returns compared to the Gaussian specification. Their results are validated here since the independent variance jumps do not provide improved return forecasts compared to the benchmark 2Comp-RM model. However, the results from the 2Comp-RM-HJ show that a time-varying mixture in the ex post variance can improve density forecasts. Variance jumps are found to be persistent and modeling this improves the forecasting of returns and ex post variance distribution.

1.5 Concluding remarks

In this Chapter I model the contemporaneous jumps of the S&P500 ex post variance. These jumps are the impact of the unobserved news flow in the market. The jumps are heterogeneous. They are modelled with a time-varying intensity which is driven by a latent autoregressive process. I use a bivariate model framework for the market returns and observable variance. The model uses two conditional variance components to capture the long and short-run memory of the ex post variance measures. I examine two ex post measures, realized variance and bipower variation.

The results show significant and frequent variance jumps, which occur at least once per week. The jumps are highly persistent and they improve the ex post variance data fit. The proposed

model shows that ex post variance is a mixture of normals with thick tails. This mixture in observable variance results in an empirically thick tailed density for returns.

The models are tested on density forecasting. I examine their out-of-sample performance in multiperiod forecasts over horizons of up to 50 days. The heterogeneous jumps provide improved density forecasts of ex post variance measures for all the horizons. When BP is used as the variance measure, the heterogeneous jumps provide improved return forecasts. When RV is used, the heterogeneous jumps improve return forecasts for the horizons longer than a single day. Between the two realized measures, RV gives marginally better return predictive likelihoods compared to BP.

TABLE 1.1: Summary statistics of SPX

	Mean	Variance	Skewness	Kurtosis	Min	Max
r_t	0.0076	1.2396	−0.2289	11.3874	−9.3511	10.2202
RV_t	1.0855	6.7016	11.0013	205.2279	0.0122	77.4774
logRV_t	−0.6943	1.3112	0.3762	3.3814	−4.4079	4.3500
BP_t	0.8931	4.8610	10.6917	176.2002	0.0106	60.1815
logBP_t	−0.8929	1.2868	0.4333	3.4402	−4.5506	4.0974

The data are for the period: January 3rd, 2000 - December 31st, 2021, 5,515 daily observations.

TABLE 1.2: Posterior estimation results for **SPX-RV** data.

RM = RV	2Comp-RV		2Comp-RV-IJ		2Comp-RV-HJ	
	Mean	95% D.I.	Mean	95% D.I.	Mean	95% D.I.
μ	0.0943	[0.080, 0.108]	0.0938	[0.079, 0.107]	0.0949	[0.083, 0.105]
ω	-0.0458	$[-0.064,-0.028]$	-0.1033	$[-0.132,-0.076]$	-0.0774	$[-0.093,-0.050]$
ϕ_1	0.4731	[0.464, 0.481]	0.4786	[0.470, 0.487]	0.4790	[0.472, 0.487]
ϕ_2	0.4731		0.4786		0.4790	
α_1	0.7965	[0.758, 0.830]	0.8247	[0.792, 0.853]	0.8260	[0.806, 0.850]
α_2	0.3605	[0.308, 0.414]	0.3641	[0.316, 0.413]	0.4029	[0.362, 0.438]
ρ	-0.1418	$[-0.157,-0.127]$	-0.1425	$[-0.158,-0.129]$	-0.1412	$[-0.152,-0.129]$
σ^2	0.3453	[0.332, 0.358]	0.2328	[0.204, 0.257]	0.1903	[0.179, 0.207]
μ_ξ			0.2833	[0.178, 0.399]	0.0988	[0.043, 0.153]
σ_ξ^2			0.4257	[0.335, 0.549]	0.3748	[0.318, 0.438]
λ			0.2359	[0.151, 0.341]		
γ_0					-0.0407	$[-0.090, 0.001]$
γ_1					0.9570	[0.929, 0.976]

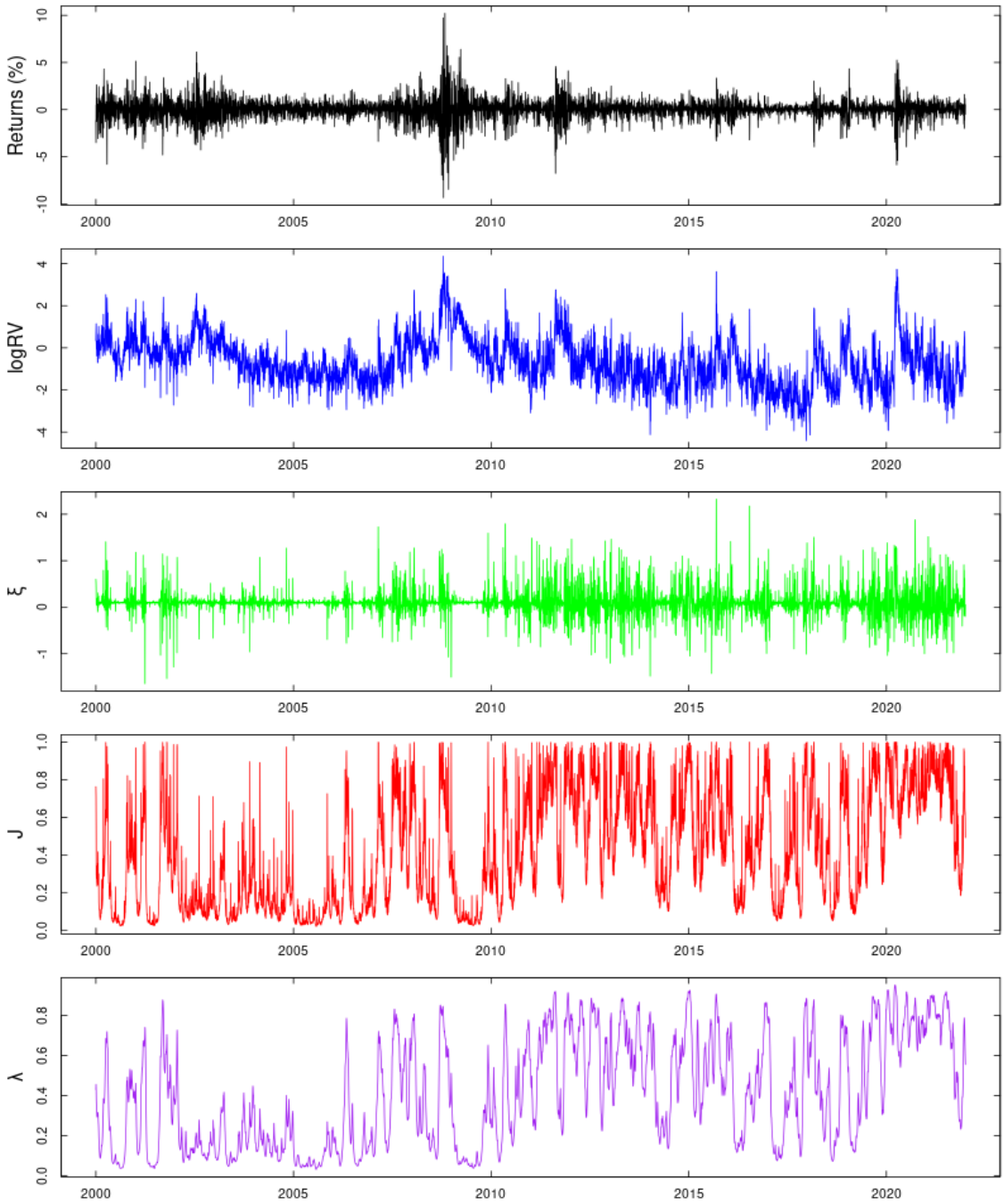
Notes: Results are from 10,000 MCMC posterior draws, after 5,000 burnin sweeps. $\phi_1 = \phi_2$ during model estimation as it has been found to give improved density forecasts of the ex post variance.

TABLE 1.3: Posterior estimation results for **SPX-BP** data.

RM = BP	2Comp-BP		2Comp-BP-IJ		2Comp-BP-HJ	
	Mean	95% D.I.	Mean	95% D.I.	Mean	95% D.I.
μ	0.1139	[0.100, 0.127]	0.1141	[0.101, 0.128]	0.1118	[0.100, 0.126]
ω	-0.0542	$[-0.076,-0.033]$	-0.0843	$[-0.112,-0.059]$	-0.0645	$[-0.087,-0.040]$
ϕ_1	0.4737	[0.466, 0.482]	0.4803	[0.472, 0.489]	0.4795	[0.473, 0.487]
ϕ_2	0.4737		0.4803		0.4795	
α_1	0.7849	[0.751, 0.815]	0.8136	[0.783, 0.839]	0.8150	[0.788, 0.853]
α_2	0.2766	[0.222, 0.335]	0.2754	[0.219, 0.332]	0.3039	[0.252, 0.356]
ρ	-0.1022	$[-0.115,-0.089]$	-0.1026	$[-0.116,-0.090]$	-0.0991	$[-0.113,-0.088]$
σ^2	0.3150	[0.304, 0.328]	0.2248	[0.199, 0.254]	0.1682	[0.155, 0.182]
μ_ξ			0.2020	[0.104, 0.330]	0.0457	[0.004, 0.088]
σ_ξ^2			0.4009	[0.302, 0.529]	0.3244	[0.290, 0.362]
λ			0.2141	[0.118, 0.324]		
γ_0					-0.0124	$[-0.042, 0.018]$
γ_1					0.9799	[0.973, 0.987]

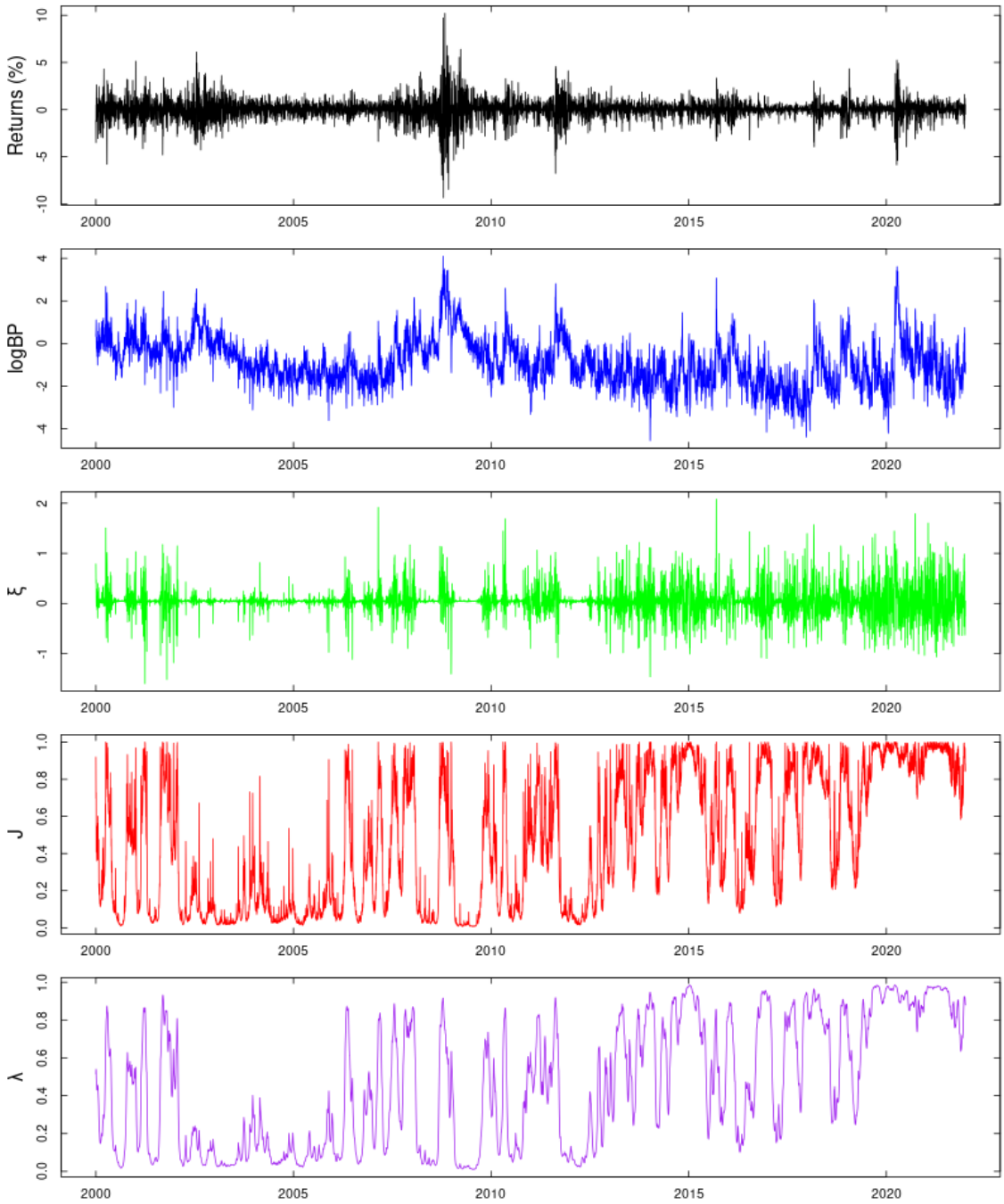
Notes: Results are from 10,000 MCMC posterior draws, after 5,000 burnin sweeps. $\phi_1 = \phi_2$ during model estimation as it has been found to give improved density forecasts of the ex post variance.

FIGURE 1.1: Estimated jumps in **SPX-RV** from the **2Comp-RV-HJ** model.



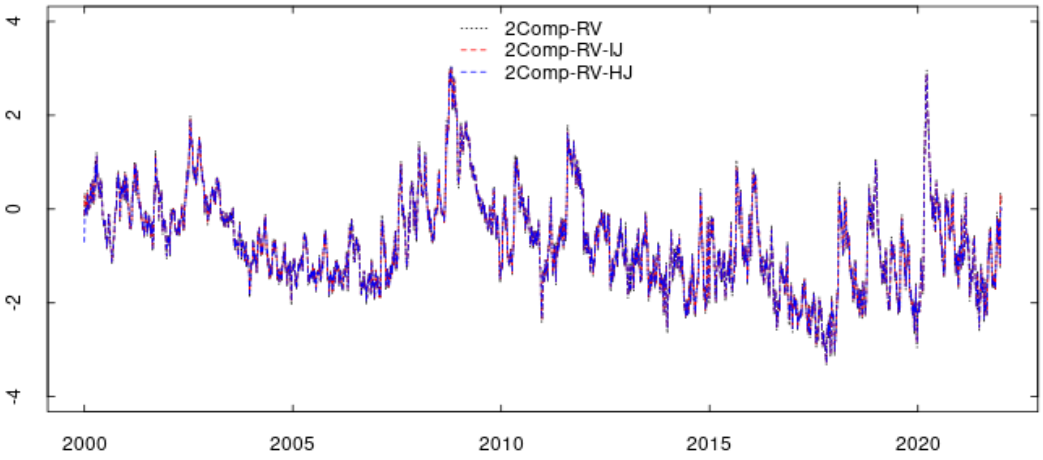
From top to bottom: SPX returns (black), logRV (blue), estimated jump size ξ_t (green), jump indication J_t (red) and estimated jump intensity λ_t (purple).

FIGURE 1.2: Estimated jumps in **SPX-BP** from the **2Comp-BP-HJ** model.



From top to bottom: SPX returns (black), logBP (blue), estimated jump size ξ_t (green), jump indication J_t (red) and estimated jump intensity λ_t (purple).

(a) Realized Variance component c_1



(b) Realized Variance component c_2

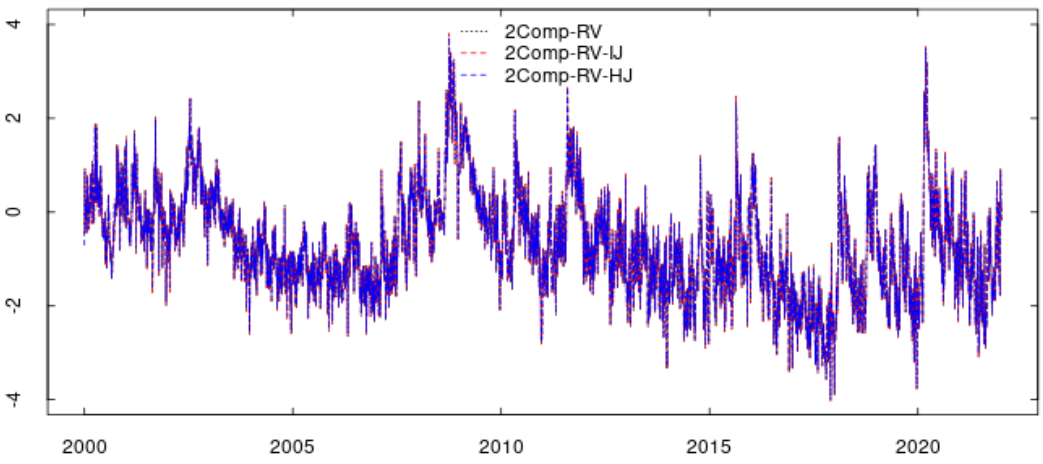
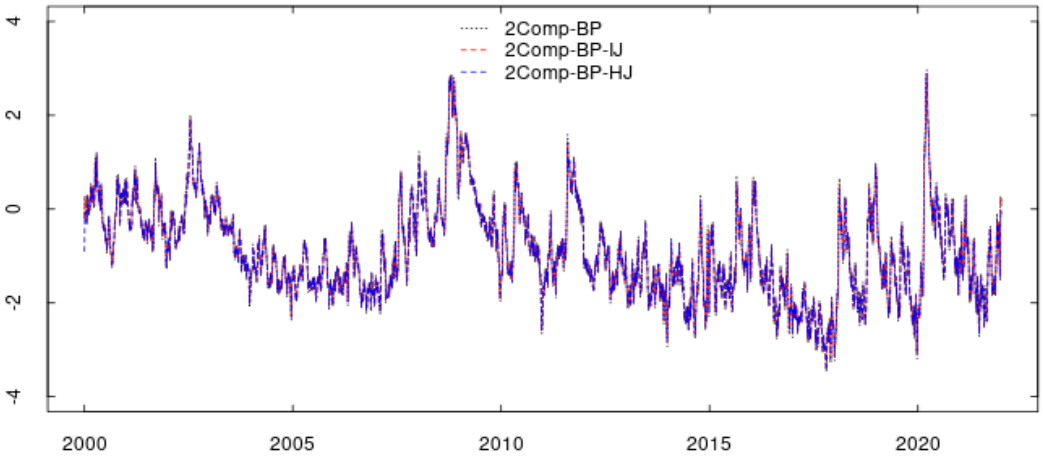


FIGURE 1.3: (a) long-run and (b) short-run variance components of SPX-RV.

(a) Bipower Variation component c_1



(b) Bipower Variation component c_2

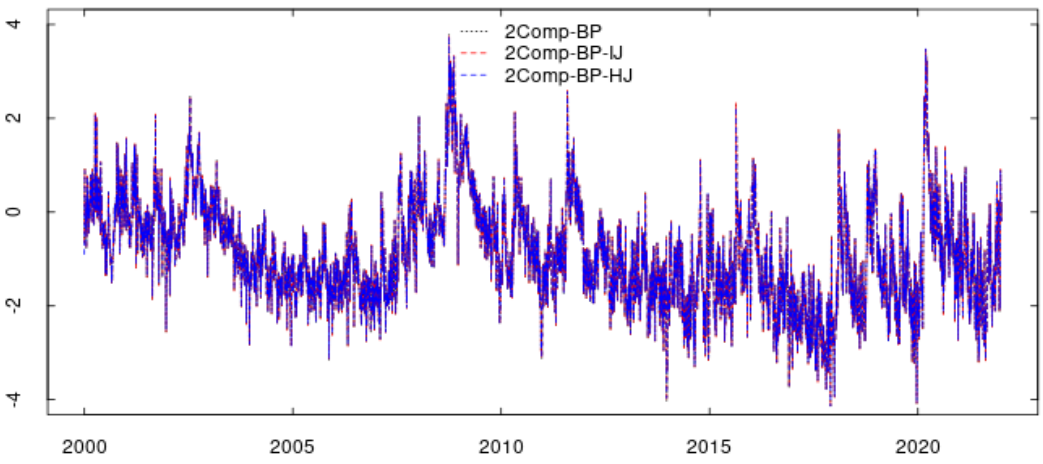
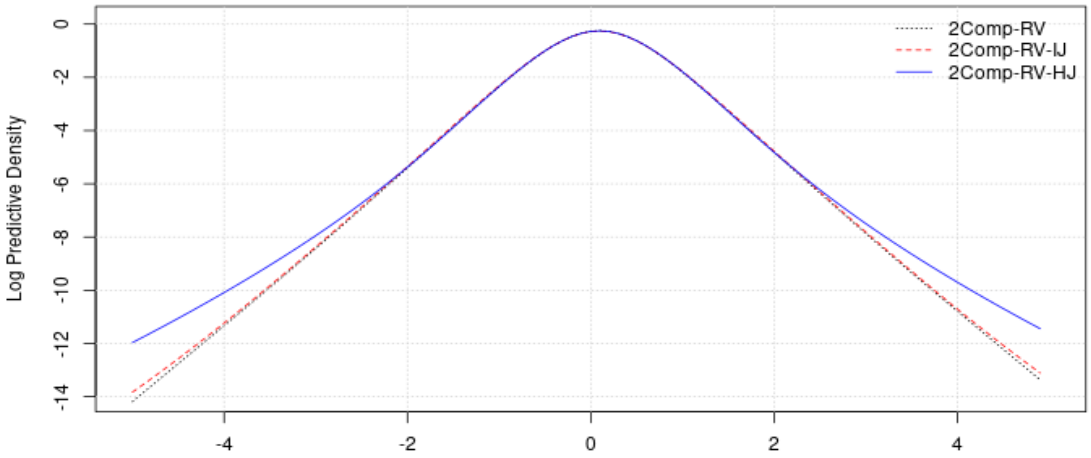


FIGURE 1.4: (a) long-run and (b) short-run variance components of SPX-BP.

(a) SPX returns



(b) SPX logRV

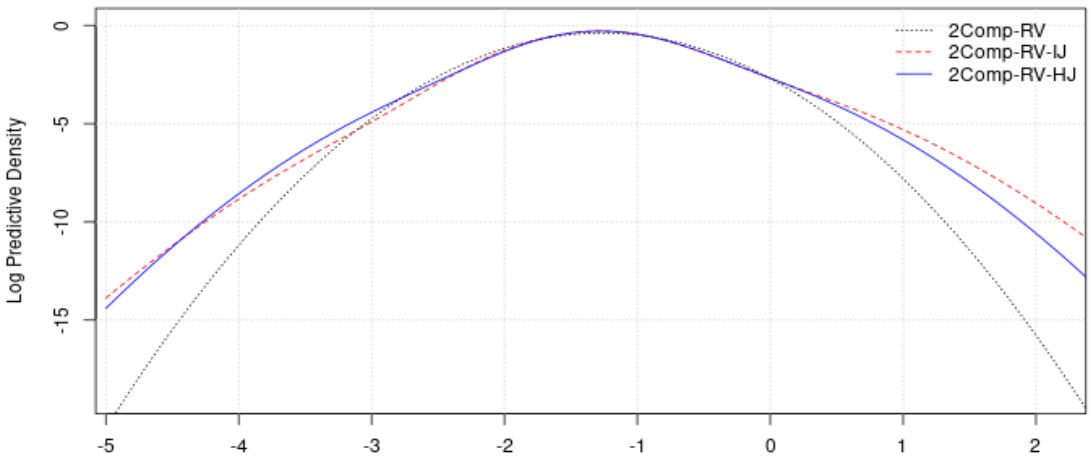
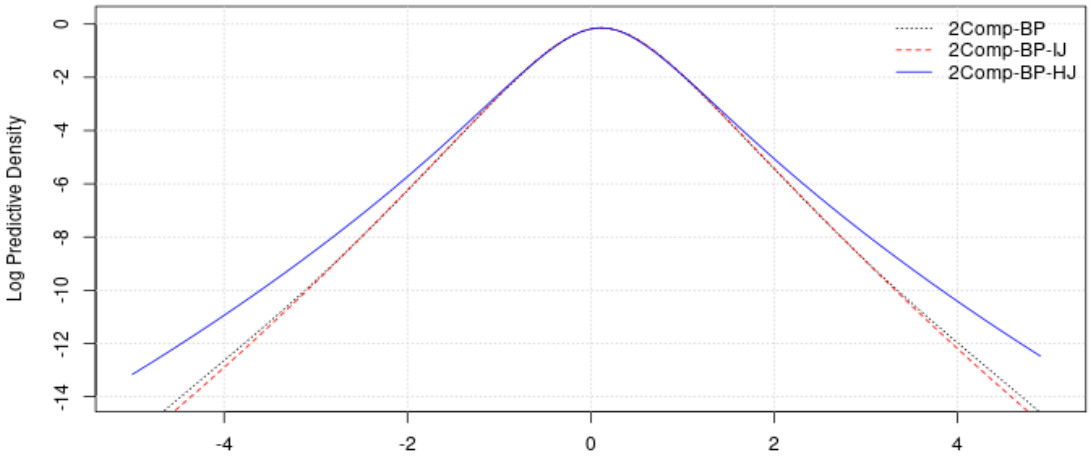


FIGURE 1.5: One-step ahead log predictive densities for SPX (a) returns and (b) logRV. The densities are calculated by evaluating grids with the full sample posterior draws.

(a) SPX returns



(b) SPX logBP

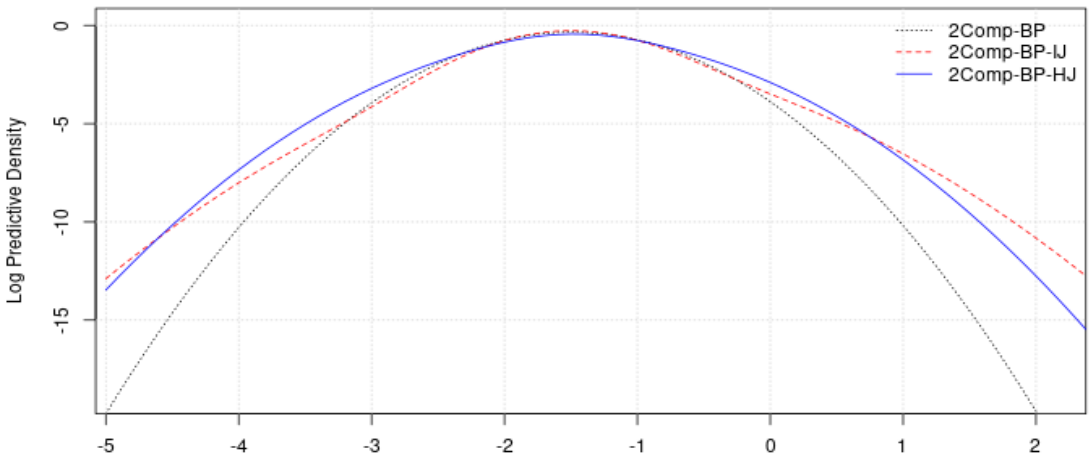


FIGURE 1.6: One-step ahead log predictive densities for SPX (a) returns and (b) logBP. The densities are calculated by evaluating grids with the full sample posterior draws.

TABLE 1.4: Cumulative log predictive likelihoods of out-of-sample forecasts for **SPX** returns and **logRV** on different forecasting horizons h.

	$h=1$		$h=5$		$h=10$		$h=50$	
RM = RV	returns	logRV	returns	logRV	returns	logRV	returns	logRV
2Comp-RV	**−590.81**	−547.69	−657.81	−745.76	−684.89	−819.69	−744.15	−965.06
2Comp-RV-IJ	−591.11	−544.50	−658.65	−724.32	−686.52	−790.16	−752.41	−938.49
2Comp-RV-HJ	−591.09	**−536.43**	**−654.79**	**−709.89**	**−678.86**	**−770.98**	**−739.52**	**−907.17**

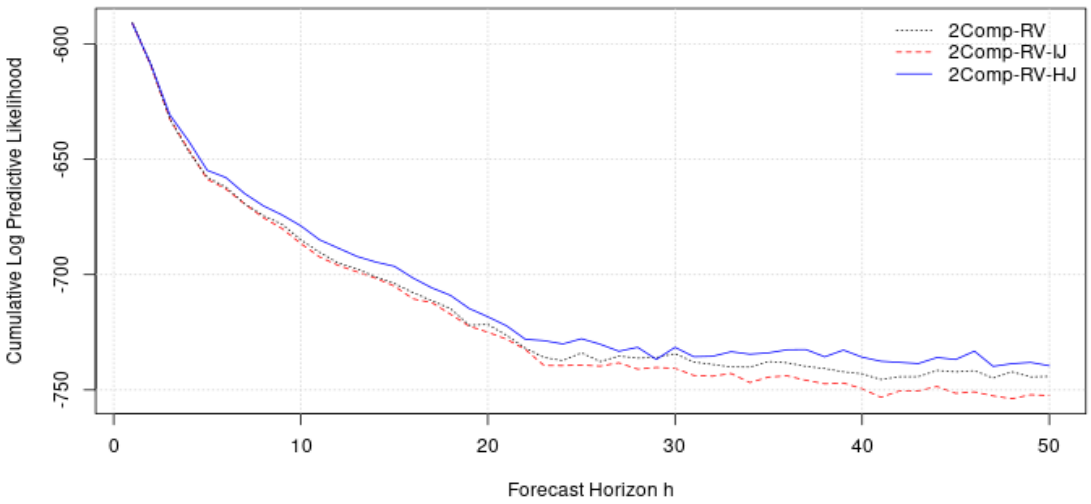
Notes: The likelihood evaluations are on the period: 30/12/2019 - 31/12/2021 (500 out-of-sample forecasts).

TABLE 1.5: Cumulative log predictive likelihoods of out-of-sample forecasts for **SPX** returns and **logBP** on different forecasting horizons h.

	$h=1$		$h=5$		$h=10$		$h=50$	
RM = BP	returns	logBP	returns	logBP	returns	logBP	returns	logBP
2Comp-BP	−592.53	−584.93	−665.66	−774.29	−695.20	−849.30	−753.25	−990.91
2Comp-BP-IJ	−592.66	−582.44	−667.86	−756.21	−698.84	−827.02	−773.79	−981.56
2Comp-BP-HJ	**−590.97**	**−554.15**	**−659.22**	**−727.53**	**−683.34**	**−792.85**	**−751.12**	**−949.99**

Notes: The likelihood evaluations are on the period: 30/12/2019 - 31/12/2021 (500 out-of-sample forecasts).

(a) SPX returns



(b) SPX logRV

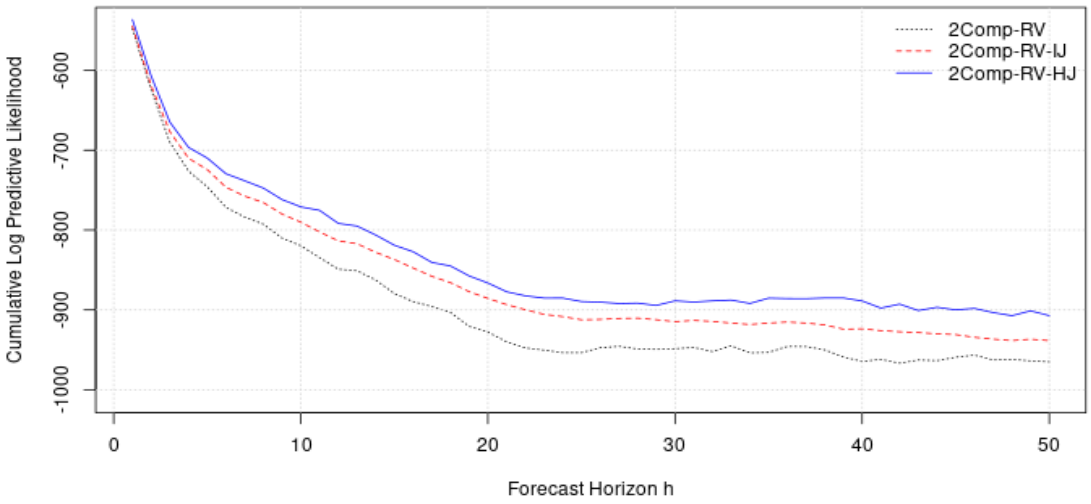
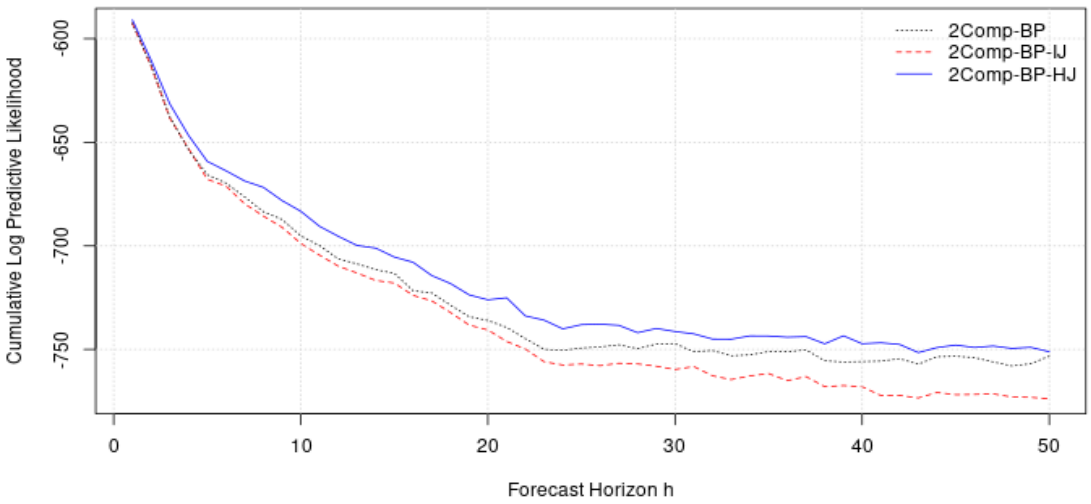


FIGURE 1.7: h-ahead cumulative log predictive likelihoods for SPX (a) returns and (b) logRV. The likelihood evaluations are on the out-of-sample period: 30/12/2019 - 31/12/2021 (500 daily observations).

(a) SPX returns



(b) SPX logBP

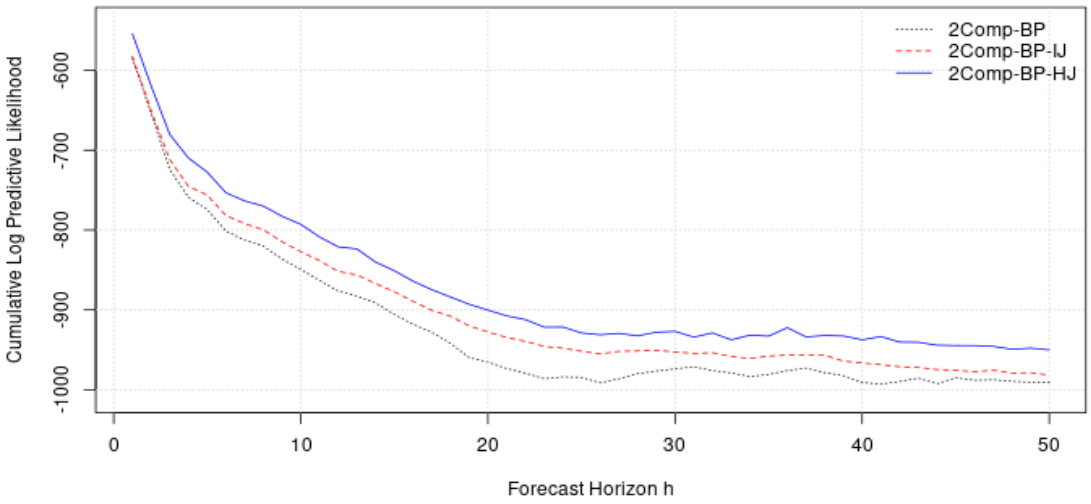


FIGURE 1.8: h-ahead cumulative log predictive likelihoods for SPX (a) returns and (b) logBP. The likelihood evaluations are on the out-of-sample period: 30/12/2019 - 31/12/2021 (500 daily observations).

Chapter 2

Is stochastic volatility Gaussian? A Bayesian semiparametric analysis

2.1 Introduction

In this chapter a new semiparametric stochastic volatility model is proposed in which the latent log-volatility has a flexible distributional assumption. I use a Dirichlet process mixture model to approximate volatility's underlying distribution. To improve the identification process I use the ex post measure of realized variance. I find strong evidence on non-Gaussian distribution in the stochastic volatility of equity indices. The model also captures asymmetry and tails of log-realized variance and returns. In out-of-sample forecasting the new model provides improved density forecasts for returns, negative returns and log-realized variance.

There are two main approaches in modeling the volatility of financial returns. The first is the (G)ARCH-type models (Engle 1982; Bollerslev 1986). In these the volatilty is assumed to be deterministic and conditional on past return information. The second approach is the stochastic volatility (SV) models introduced by Taylor (1982). These differ than GARCH models since they assume stochastic and latent variance given past information. SV models treat volatility as the impact of an unobserved news flow process. Kim et al. (1998) develop a Bayesian approach to estimate the discrete time SV model.

In the simplest form SV models assume that returns are a continuous mixture of normals governed by their latent log-volatility which follows an AR(1) process. Early parametric extensions of the SV models focus on the non-Gaussian features of the return distribution. Mahieu and Schotman (1998) use a finite mixture of normals for return density modeling. Liesenfeld and Jung (2000) use normal mixtures and Student-t assumption on return innovations. Chib et al. (2002) use Student-t errors and a jump component. Durham (2006) uses two SV processes to model the return variance. Abanto-Valle et al. (2010) and Nakajima and Omori (2012) also focus on modeling the tail behaviour of returns. In the above studies the volatility equation remains a parametric AR(1) specification.

Jensen and Maheu (2010) develop a semiparametric SV model in which the return variance is decomposed into two parts. One is a parametric SV and the second is an infinite mixture of normals with a Dirichlet process (DP) prior (Ferguson 1973). This model can approximate the underlying continuous return distribution while capturing the volatilty clustering with the SV part. Other semiparametric SV models have been developed from Delatola and Griffin (2011), Yu (2012), Delatola and Griffin (2013), Jensen and Maheu (2014) and Zaharieva et al. (2020) in a multivariate framework.

A different extension of the SV models was introduced by So et al. (1998) who estimated the latent volatility under a finite number of regimes. Regime-switching SVs have also been applied by Kalimipalli and Susmel (2004) and Vo (2009). Vo (2009) extend the SV-DPM model of Jensen and Maheu (2010) by making the constant of volatility equation a two-state Markov-switching variable. Li et al. (2022) use an infinite hidden Markov mixture instead of the independent Dirichlet process mixture (DPM) in return innovations. These studies indicate that a Gaussian AR(1) structure of latent volatility is a strong assumption.

In this work I explore the non-Gaussian features of stochastic volatility. I drop the restrictive assumption of normal innovations in the volatility equation. Instead, I estimate latent log-volatility semiparametrically with a DPM model. I use an infinite mixture of normals in the

constant and the variance of stochastic volatility. This enables to capture shocks, jumps and clustering that occur in return variance. A similar infinite mixture has been used by Jensen and Maheu (2014). The model I develop has two differences. First, I include the mean of returns in the mixture. This allows the model to capture asymmetry and thick tails of the return distribution. Second, to improve the identification of latent volatility I use realized variance (RV) data as an extra source of information along with returns.

RV data, calculated from high-frequency returns, is more informative than squared returns about their variance. Takahashi et al. (2009) first used RV data to assist the volatility estimation in SV. Other studies that incorporate RV data are from Yamauchi and Omori (2020) in the multivariate SV, along with realized correlation data, Liu (2021) who models returns and logRV with DPM and Hansen et al. (2012) in the GARCH model. I extend the framework of Takahashi et al. (2009) to a model with a DPM mixture in the latent volatility equation and the returns mean.

The model application on major equity indices shows strong non-Gaussian behaviour in the estimated volatility. Results indicate that latent volatility is a mixture of at least 4 normal clusters. The plots of estimated time varying volatility of volatility shows that this gives the model the necessary flexibility to capture variance shocks, in RV, as well as extreme returns.

Predictive density plots of volatility show strong evidence of asymmetry and thick tails. Interestingly, the flexible volatility estimation captures asymmetry and tails of logRV data. This is validated by out-of-sample forecasting, in which the proposed model outperforms the Gaussian benchmark in logRV density forecasts. The new model performs better in forecasting the most volatile equity returns. Due to the flexible mixture it assigns more probability mass to the far left side of the return density than the Gaussian benchmark. For most of the series tested it produces better density forecasts when returns are negative and exceed losses of 1%. This is important in risk management applications and these will be examined in future work.

The chapter is organized as follows: Section 2.2 is a brief discussion on RV, Section 2.3 presents the proposed model, its estimation steps and the forecasting process, Section 2.4 is the

empirical application and in Section 2.5 are the concluding remarks.

2.2 Realized variance

The ex post variance of an asset's return is estimated with the nonparametric realized measures. Realized variance is the simplest measure. The important papers that provide a foundation of the theory and application of RV are from Andersen et al. (2001a), Andersen et al. (2001b), Barndorff-Nielsen and Shephard (2002a), Barndorff-Nielsen and Shephard (2002b) and Andersen et al. (2003). The theoretical foundation starts with a continuous-time stochastic volatility model. The simplest form, without jumps, is discussed here, however the proposed model can capture return jumps. The logarithmic asset price at time t, p_t, evolves under the following process

$$dp_t = \mu_t dt + \sigma_t dW_t, \tag{2.1}$$

where μ_t and $\sigma_t (\sigma_t > 0)$ are the stochastic drift and the diffusion process W_t is a standard Brownian motion. The daily continuously compounded logarithmic return at time t, r_t, is defined as

$$r_t = p_t - p_{t-1} = \int_{t-1}^{t} \mu_\tau d\tau + \int_{t-1}^{t} \sigma_\tau dW_\tau \tag{2.2}$$

and the quadratic return variation is the integrated variance (IV)

$$\text{QV}_t \equiv \text{IV}_t = \int_{t-1}^{t} \sigma_\tau^2 d\tau \tag{2.3}$$

For day t, given intraday returns $r_{t,i}$, $i = 1, ..., n$, Barndorff-Nielsen and Shephard (2002a) and Andersen et al. (2003) show that QV_t can be approximated by the RV estimator defined as

$$\text{RV}_t = \sum_{i=1}^{n} r_{t,i}^2 \to \text{QV}_t, \quad \text{as } n \to \infty. \tag{2.4}$$

RV_t is a consistent estimator of QV_t under no market microstructure noise. In practice, high-frequency returns contain microstructure noise which makes RV_t biased and inconsistent, as Zhang et al. (2005), Hansen and Lunde (2006) and Bandi and Russell (2008) document.

A popular way to overcome noise in RV_t is the use of *subsampling* as proposed by Zhang et al. (2005). With this statistical technique, the grid of available intraday observations is separated into multiple subgrids. For each subgrid, an estimator of realized variance is calculated. Then, RV_t is set as an average of the subgrid estimators. This gives an approximately unbiased RV_t. Aït-Sahalia and Mancini (2008) show that RV_t from *subsampling* outperforms the simple one, from the summation of squared log returns, in contained bias, variance, RMSE and forecasting.

2.3 Model specification

To examine the distributional behaviour of stochastic volatility, I develop an SV model where log-volatility does not follow the standard Gaussian AR(1) process which is usually adopted in literature. Instead, I estimate it under the assumption of an infinite mixture of normals, given a Dirichlet process (DP) prior, of Ferguson (1973). This flexible mixture framework has the ability to approximate any underlying continuous distribution.

I extend the framework proposed by Takahashi et al. (2009) which uses RV data to help the estimation of stochastic volatility. I include a Dirichlet process mixture (DPM) in the mean, the variance of stochastic volatility and the returns mean. That way the model has the flexibility to estimate volatility under a free distributional assumption and also capture the asymmetry and thick tails of returns distribution. The proposed model offers a parsimonious way to include ex post variance information and extend the SV-DPM of Jensen and Maheu (2010).

Conditional on the information set $\mathcal{I}_{t-1} = \{(r_1, \text{logRV}_1), ..., (r_{t-1}, \text{logRV}_{t-1})\}$ the **RSV-DPM-V** (Realized SV with DPM in Volatility) model has the following hierarchical specification

$$r_t = m_t + \exp(h_t/2)u_t, \quad u_t \sim \text{NID}(0,1), \tag{2.5}$$

$$\text{logRV}_t = c + dh_t + bz_t, \quad z_t \sim \text{NID}(0,1), \tag{2.6}$$

$$h_t = g_t + \delta h_{t-1} + v_t e_t, \quad e_t \sim \text{NID}(0,1), \tag{2.7}$$

$$m_t, g_t, v_t^2 | G \overset{iid}{\sim} G, \tag{2.8}$$

$$G|G_0, \alpha \sim \text{DP}(\alpha, G_0), \tag{2.9}$$

$$G_0(m_t, g_t, v_t^2) \equiv \text{N}(m_0, u_0^2) - \text{N}(g_0, q_0^2) - \text{IG}(v_0, s_0), \tag{2.10}$$

with model parameters: $\theta = \{\beta, b^2, \delta, \alpha\}$, $\beta = \{c, d\}$.

r_t are log returns at time $t, t = 1, ..., T$ and logRV_t denotes the natural logarithm of the returns realized variance at time t. h_t is the stochastic log-volatility of returns at time t. (2.5) and (2.6) are usually referred as measurement equations. (2.6) uses logRV_t as informational signal to assist the estimation of stochastic volatility. Parameter d captures deviations between h_t and logRV_t due to market frictions and is expected to be close to 1. Takahashi et al. (2009) restrict $d = 1$. My results show this to be a strict assumption since I find d to be close but significantly different than 1.

Eq.(2.8)-(2.10) place in infinite mixture of normals in returns and log-volatility. The mixing parameters are returns mean m_t, stochastic volatility constant g_t and volatility of volatility v_t. Stochastic volatility specification in (2.7) is an extension of the one used by So et al. (1998) who use a 3-state Markovian mixture in the volatility constant. In this specification the path dependence of volatility is not explicitly stated.

The mixture parameters, m_t, g_t and v_t, are distributed according to the latent G which is nonparametrically modelled with a DP prior. A draw from a DP, $G \sim \text{DP}(\alpha, G_0)$, is almost surely a discrete distribution and has two parameters, the base measure G_0 and the precision

parameter $\alpha > 0$. The DP is centred around G_0 since $\mathbb{E}[G] = G_0$ and the precision parameter α determines how close is G to G_0 since $\text{Var}[G] = G_0[1 - G_0]/(\alpha + 1)$. In this case the base measure of DP in (2.10) is a normal-normal-inverse gamma (N-N-IG) prior.

G has a support of infinite distributions. Let $\mu = \{\mu_1, \mu_2, ...\}$, $\gamma = \{\gamma_1, \gamma_2, ...\}$ and $\sigma_t^2 = \{\sigma_1^2, \sigma_2^2, ...\}$ denote the unique points of support in G with

$$\mu_j \overset{iid}{\sim} \text{N}(m_0, u_0^2), \quad j = 1, 2, ... \tag{2.11}$$

$$\gamma_j \overset{iid}{\sim} \text{N}(g_0, q_0^2), \quad j = 1, 2, ... \tag{2.12}$$

$$\sigma_j^2 \overset{iid}{\sim} \text{IG}(v_0, s_0), \quad j = 1, 2, ... \tag{2.13}$$

The conditional distributions of returns and volatility have the following stick-breaking representation of Sethuraman (1994),

$$p(r_t|\mathcal{I}_{t-1}, h_t, W, \mu) = \sum_{j=1}^{\infty} w_j \text{N}\left(r_t|\mu_j, \exp(h_t)\right), \tag{2.14}$$

$$p(h_t|\mathcal{I}_{t-1}, W, \gamma, \sigma^2) = \sum_{j=1}^{\infty} w_j \text{N}\left(h_t|\gamma_j + \delta h_{t-1}, \sigma_j^2\right), \tag{2.15}$$

where $W = \{w_1, w_2, ...\}$ is the infinite set of weights associated with the mixing normal densities, with $\sum_{j=1}^{\infty} w_j = 1$ and a stick-breaking prior which is generated as

$$w_1 = v_1, \quad w_j = v_j \prod_{l=1}^{j-1}(1 - v_l), \quad j > 1, \tag{2.16}$$

$$v_j \overset{iid}{\sim} \text{B}(1, \alpha). \tag{2.17}$$

where $\text{B}(.)$ denotes the Beta distribution. A finite set $\{(r_1, h_1), ..., (r_T, h_T)\}$ will be associated with a finite set $\{(m_1, g_1, v_t^2), ..., (m_T, g_T, v_T^2)\}$ of draws from G in (2.8). The DPM allows data clustering in identical sets of (m_t, g_t, v_t^2). This truncates the infinite mixture into a finite one with k unique clusters, $\{\mu_j, \gamma_j, \sigma_j^2\}_{j=1}^k$, $k < T$. This means that the model learns from the data and uses a mixture of k normal kernels to approximate the underlying data distribution.

The flexible nonparametric DPM framework nests different parametric specifications. The precision parameter α controls the number of mixture clusters. When $\alpha \to 0$ then $w_1 = 1$, $w_j = 0, \forall j > 1$ and stochastic volatility will follow a Gaussian AR(1) with variance σ_1^2. In this case RSV-DPM-V will be equivalent with the proposed RSV-N of Takahashi et al. (2009).

2.3.1 Estimation

A straightforward method to estimate DPM models is by using the stick-breaking formulation and the slice sampler by Walker (2007) and Kalli et al. (2011), which truncates the infinite mixture into a finite number κ, $\kappa \leq k < T$, of unique clusters, $\{\mu_j, \gamma_j, \sigma_j^2\}_{j=1}^{\kappa}$, with at least one data observation assigned in each cluster. To do so, the parameter space is expanded by introducing two latent vectors. The first one is a cluster or state indicator $s_{1:T} = \{s_1, ...s_T\}$ which maps each set (r_t, h_t) to a cluster j. The second auxiliary vector $u_{1:T} = \{u_1, ...u_T\}$, with $u_t \in (0, 1)$, helps to convert the infinite sum in (2.14) and (2.15) into a finite mixture.

The joint posterior of the RSV-DPM-V model $p(\{\mu_j, \gamma_j, \sigma_j^2\}_{j=1}^{\infty}, s_{1:T}, u_{1:T}, h_{1:T}, \theta | \mathcal{I}_T)$ is proportional to

$$
\begin{aligned}
& p(\theta)p(w_{1:k}) \prod_{j=1}^{k} p(\mu_j, \gamma_j, \sigma_j^2) \\
& \times \prod_{t=1}^{T} \mathbf{1}\{u_t < w_{s_t}\} \, \mathrm{N}(r_t | \mu_{s_t}, \exp(h_t)) \; \mathrm{N}\left(\mathrm{logRV}_t | c + dh_t, b^2\right) \; \mathrm{N}\left(h_t | \gamma_{s_t} + \delta h_{t-1}, \sigma_{s_t}^2\right).
\end{aligned}
\tag{2.18}
$$

with the first line being the prior and the second one the likelihood. $\mathbf{1}\{.\}$ is the indicator function and k is the smallest positive integer that satisfies the condition $\sum_{j=1}^{k} w_j > 1 - \min(u_{1:T})$. The above posterior does not have a known form. To sample the parameters and estimate the stochastic volatility I follow standard Markov chain Monte Carlo (MCMC) techniques to sample from a series of conditional distributions. This allows a straightforward estimation with a set of conditional steps.

At first, I initialize $\theta, k, w_{1:k}, s_{1:T}, \mu_{1:k}, \gamma_{1:k}, \sigma^2_{1:k}, \alpha$ and $h_{1:T}$. Then I collect a large number of posterior draws $\{\theta^{(i)}, k^{(i)}, s^{(i)}_{1:T}, \mu^{(i)}_{1:k}, \gamma^{(i)}_{1:k}, \sigma^{2(i)}_{1:k}, \alpha^{(i)}, h^{(i)}_{1:T}\}_{i=1}^{R}$ by iterating through the following MCMC steps:

1. Sample $h_t, t = 1, ..., T$ from $p(h_t|h_{-t}, r_{1:T}, \text{logRV}_{1:T}, \theta, \mu_{1:k}, \gamma_{1:k}, \sigma^2_{1:k}, s_{1:T})$.
2. Sample δ from $p(\delta|h_{1:T}, \gamma_{1:k}, \sigma^2_{1:k}, s_{1:T})$.
3. Sample $\mu_{1:k}$ from $p(\mu_{1:k}|r_{1:T}, h_{1:T}, s_{1:T})$, $\gamma_{1:k}$ from $p(\gamma_{1:k}|r_{1:T}, h_{1:T}, \sigma^2_{1:k}, s_{1:T}, \delta)$ and $\sigma^2_{1:k}$ from $p(\sigma^2_{1:k}|r_{1:T}, h_{1:T}, \gamma_{1:k}, s_{1:T}, \delta)$.
4. Update $w_{1:k}, u_{1:T}, k|s_{1:T}$.
5. Sample $s_t, t = 1, ..., T$ from $p(s_t|r_{1:T}, \text{logRV}_{1:T}, h_{1:T}, \theta, \mu_{1:k}, \gamma_{1:k}, \sigma^2_{1:k}, w_{1:k}, u_{1:T}, k)$.
6. Sample α from $p(\alpha|\kappa, T)$ with κ being the number of active clusters.
7. Sample β from $p(\beta|\text{logRV}_{1:T}, h_{1:T}, b^2)$ and b^2 from $p(b^2|\text{logRV}_{1:T}, h_{1:T}, \beta)$.

Details of the sampling steps are in the appendix A2. Repeating the above steps R times, after discarding R_0 burnin sweeps, I get the posterior draws for inference.

2.3.2 Benchmark models

As benchmark models I use the **RSV-N** of Takahashi et al. (2009) which has a normal innovations in returns and stochastic volatility. This is defined as

$$r_t = \mu + \exp(h_t/2)u_t, \quad u_t \sim \text{NID}(0, 1), \tag{2.19}$$

$$\text{logRV}_t = c + dh_t + bz_t, \quad z_t \sim \text{NID}(0, 1), \tag{2.20}$$

$$h_t = \gamma + \delta h_{t-1} + \sigma_v e_t, \quad e_t \sim \text{NID}(0, 1). \tag{2.21}$$

The second benchmark used is the **SV-DPM** of Jensen and Maheu (2010) which approximates the return distribution with an infinite mixture of normal kernels and a Gaussian AR(1)

stochastic volatility. This has the following hierarchical form

$$r_t = m_t + \lambda_t \exp(h_t/2)u_t, \quad u_t \sim \text{NID}(0,1), \tag{2.22}$$

$$h_t = \delta h_{t-1} + \sigma_v e_t, \quad e_t \sim \text{NID}(0,1), \tag{2.23}$$

$$m_t, \lambda_t^2 | G \overset{iid}{\sim} G, \tag{2.24}$$

$$G | G_0, \alpha \sim \text{DP}(\alpha, G_0), \tag{2.25}$$

$$G_0(m_t, \lambda_t^2) \equiv \text{N}(m_0, u_0^2) - \text{IG}(v_l, s_l). \tag{2.26}$$

The SV-DPM is a natural benchmark to the RSV-DPM-V without logRV_t signal to show the importance of RV_t.

2.3.3 Predictive density

In this section I present the prediction process for the RSV-DPM-V model since it nests the benchmark specifications. In forecasting the whole distribution of a variable, such as returns, the measure of interest is the predictive density $p(r_{t+1}|\mathcal{I}_t)$. The out-of-sample predictive density plots for $p(h_{t+1}|\mathcal{I}_t, \text{RSV-DPM-V})$ can give a visualization of the stochastic volatility's underlying distribution.

The key task when forecasting with DPM models is to integrate out the uncertainty about the future state of the mixture parameters. Conditional on $\mathcal{I}_t = \{r_{1:t}, \text{logRV}_{1:t}\}$, the predictive density of returns, logRV and log-stochastic volatility can be approximated with the use of R

posterior draws as

$$p(r_{t+1}|\mathcal{I}_t, h_{t+1}) \approx \frac{1}{R}\sum_{i=1}^{R} \mathrm{N}\left(r_{t+1}\middle|\mu^{(i)}_{s^{(i)}_{t+1}}, \exp\left(h^{(i)}_{t+1}\right)\right), \tag{2.27}$$

$$p(\mathrm{logRV}_{t+1}|\mathcal{I}_t, h_{t+1}) \approx \frac{1}{R}\sum_{i=1}^{R} \mathrm{N}\left(\mathrm{logRV}_{t+1}\middle|c^{(i)} + b^{(i)}h^{(i)}_{t+1}, b^{2(i)}\right), \tag{2.28}$$

$$p(h_{t+1}|\mathcal{I}_t) \approx \frac{1}{R}\sum_{i=1}^{R} \mathrm{N}\left(h_{t+1}\middle|\gamma^{(i)}_{s^{(i)}_{t+1}} + \delta^{(i)}h^{(i)}_{t}, \sigma^{2(i)}_{s^{(i)}_{t+1}}\right), \tag{2.29}$$

$$\text{where} \quad s^{(i)}_{t+1} = \begin{cases} j, \text{if} \quad \sum_{l=0}^{j-1} w^{(i)}_l < \phi < \sum_{l=0}^{j} w^{(i)}_l, \\ k^{(i)}+1, \text{if} \quad \phi \geq \sum_{l=0}^{k^{(i)}} w^{(i)}_l, \end{cases} \tag{2.30}$$

with $w^{(i)}_o = 0$, $j \leq k^{(i)}$, $\phi \sim \mathrm{U}(0,1)$ and h_{t+1} is simulated as

$$h^{(i)}_{t+1} \sim \mathrm{N}\left(\gamma^{(i)}_{s^{(i)}_{t+1}} + \delta^{(i)}h^{(i)}_{t}, \sigma^{2(i)}_{s^{(i)}_{t+1}}\right). \tag{2.31}$$

The above means that the future value of $s^{(i)}_{t+1}$ is one of the existing clusters with probability equal to the associated weights and there is a nonzero probability of introducing a new cluster $\left(\mu^{(i)}_{k^{(i)}+1}, \gamma^{(i)}_{k^{(i)}+1}, \sigma^{2(i)}_{k^{(i)}+1}\right)$ from the base measure G_0.

To test a model's forecasting ability, the predictive density calculations are used to evaluate out-of-sample realizations of the modelled variables and construct the predictive likelihood. Specifically, for each model, by performing a set of τ (with $1 < \tau < T$), recursive posterior estimations, the log predictive likelihood of $r_{T-\tau+1:T}$, from each model, is the summation of individual log predictive densities of each estimation

$$\mathrm{logPL}(r_{T-\tau+1:T}|\mathcal{I}_T) = \sum_{t=T-\tau}^{T-1} \log\left(p(r_{t+1}|\mathcal{I}_t, h_{t+1})\right). \tag{2.32}$$

Similarly, the log predictive likelihood of $\mathrm{logRV}_{T-\tau+1:T}$ can be calculated as

$$\mathrm{logPL}(\mathrm{logRV}_{T-\tau+1:T}|\mathcal{I}_T) = \sum_{t=T-\tau}^{T-1} \log\left(p(\mathrm{logRV}_{t+1}|\mathcal{I}_t, h_{t+1})\right). \tag{2.33}$$

The difference between two log predictive likelihoods, from two different models ($\text{logPL}_1 - \text{logPL}_2$), gives the log Bayes Factor (logBF) (Kass and Raftery 1995) and shows the significance in their density forecasting performance. A rough guide to the values of logBF = $\text{logPL}_1 - \text{logPL}_2$, is as follows: $0-1$ not worth more than a bare mention, $1-3$ positive evidence for model 1, $3-5$ strong evidence for model 1 and >5 very strong evidence for model 1.

The proposed RSV-DPM-V model has the ability to accommodate extreme return events due to the flexibility in estimating the latent volatility. This is important for risk management applications where forecasting the left tail of the distribution is often the task of interest. I test the model's ability to forecast extreme return events with the predictive density for specific distributional regions. For a value η, with $\eta \in \mathbb{R}$, the predictive density of $r_{t+1} < \eta$ is defined as

$$p(r_{t+1}|r_{t+1}<\eta,\mathcal{I}_t) = \frac{p(r_{t+1}|\mathcal{I}_t)\mathbf{1}\left\{r_{t+1}<\eta\right\}}{\int_{-\infty}^{\eta} p(r_{t+1}|\mathcal{I}_t)dr_{t+1}}$$

$$\approx \frac{1}{R}\sum_{i=1}^{R} \frac{\mathrm{N}\left(r_{t+1}\middle|\mu_{s_{t+1}^{(i)}}^{(i)}, \exp\left(h_{t+1}^{(i)}\right)\right)\mathbf{1}\left\{r_{t+1}<\eta\right\}}{\Phi\left(\left(\eta-\mu_{s_{t+1}^{(i)}}^{(i)}\right)/\exp\left(h_{t+1}^{(i)}/2\right)\right)}, \tag{2.34}$$

where $\Phi(.)$ denotes the standard Gaussian c.d.f. The denominator in (2.34) is an integrating constant ensuring that the predictive density integrates to one. The cumulative log-predictive likelihood can be calculated as of (2.32), conditional on $r_{t+1} < \eta$.

2.4 Empirical application

2.4.1 Data

Data are the daily open-to-close log returns of: **DAX**, **FTSE** (FTSE 100) and **SPX** (S&P500) indices and their ex post variance measure of realized variance (**RV**). Summary statistics are in Table 2.1. The data are obtained from Oxford-Man Institute's Realized Library of Heber

et al. (2009). RV is calculated from 5-minute returns using *subsampling* in order to reduce the microstructure noise. The returns have been converted to percentages and RV data have been scaled by 100^2.

2.4.2 Selection of priors

The base measure of DPM, G_0, for the return mean is set around zero, $G_0(\mu_j) \equiv \text{N}(0, 0.1)$. The same is for the constant of log-volatility with $G_0(\gamma_j) \equiv \text{N}(0, 0.1)$. For the volatility of volatility I use $G_0(\sigma_j^2) \equiv \text{IG}(5/2, 1/2)$. In the benchmarks I use $\mu \sim \text{N}(0, 0.1), \gamma \sim \text{N}(0, 0.1)$ and $\sigma_v^2 \sim \text{IG}(5/2, 1/2)$. In the SV-DPM I use $G_0(\lambda_j^2) \equiv \text{IG}(10/2, 10/2)$. The precision parameter of DPM has a Gamma prior, $\alpha \sim \Gamma(2, 8)$, following Jensen and Maheu (2010). For the measurement equation parameters I use: $\beta \sim \text{N}(\beta_0, B_0)$ with $\beta_0 = [0 \quad 1]'$, $B_0 = 100 \times I_2$ and $b^2 \sim \text{IG}(5/2, 1/2)$. AR(1) parameter is from $\delta \sim \text{N}(0.9, 0.01)\mathbf{1}\{|\delta| < 1\}$.

2.4.3 Posterior estimation results

Results from 10,000 MCMC posterior draws, after 5,000 burnin, for DAX, FTSE and SPX are in Tables 2.2, 2.3 and 2.4. The key result for all the data series is that RSV-DPM-V shows a mixture of normals in volatility. The model uses on average 4.53 clusters for DAX, 5.64 clusters for FTSE and 4.34 clusters for SPX data. SV-DPM model has a broader range in the number of clusters. It uses more clusters for DAX, 9.6, and SPX, 7.45, and less clusters for FTSE, 2.44.

In the measurement equations between RSV-N and RSV-DPM-V, the constant c is consistently estimated at -0.174 for DAX, -0.457 for FTSE and -0.699 for SPX. These values are close to the sample means of logRV data in Table 2.1. Noticeable is the decline in variance b^2 in RSV-DPM-V compared to the benchmark RSV-N. In DAX is 0.145 from 0.1616, in FTSE is 0.2 from 0.22 and in SPX is 0.18 from 0.183. This indicates the effect of the flexible mixture in latent volatility which helps to estimate h_t closer to logRV_t compared to the Gaussian RSV, especially for the more volatile series of DAX and FTSE. Coefficient d is slightly declined in RSV-DPM-V for DAX and FTSE, and has a minor increase for SPX.

The effect of the volatility mixture is clear in the posterior estimation of h_t. In Figures 2.1 and 2.2 for the series DAX and FTSE, RSV-DPM-V is able to estimate the latent volatility closer to the outlier values of logRV, compared to the restricted Gaussian RSV-N. The volatility mixture effect is less obvious in SPX.

The flexibility of the volatility mixture is clear in the Figures 2.4, 2.5 and 2.6. In these it can be seen the time-varying returns mean, volatility constant and volatility of volatility. The RSV-DPM-V seems to capture efficiently, return and variance jumps as well as volatility clustering. The key difference of RSV-DPM-V from the models in literature such as the SV-DPM of Jensen and Maheu (2010) is that the variance of volatility σ_t^2 is in fact time-varying. This accommodates shocks and jumps in latent variance. Especially for DAX and FTSE, $\mathbb{E}(\sigma_{s_t}^2|\text{RSV-DPM-V})$ has rare extreme spikes which are associated with bursts in logRV. For SPX, which is less volatile than the other two series, volatility of volatility is time-varying as well, but at a smaller range.

2.4.4 Forecasting results

The posterior estimation results and plots show that stochastic volatility has a time changing variance, indicating a non-Gaussian structure. Predictive density plots validate that stochastic volatility from RSV-DPM-V is not Gaussian. It is clear for all the data series in Figures 2.7, 2.8 and 2.9 that $p(h_{T+1}|\mathcal{I}_T)$ is different than a Gaussian shape. It has strong asymmetry and thick tails. Moreover, from the $p(\text{logRV}_{T+1}|\mathcal{I}_T)$ plots can be seen that the flexibility of RSV-DPM-V in the estimation of h_t captures the non-Gaussian features of logRV. The empirical skewness and excess kurtosis of returns are also captured by the RSV-DPM-V as seen in all the $p(r_{T+1}|\mathcal{I}_T)$ plots.

How does the capture of volatility's non-Gaussian features impact the return modeling? Does it improve density forecasts? This is answered by the cumulative log predictive likelihood results of 500 daily out-of-sample predictions, from recursive posterior estimations, in Table 2.5. For DAX, the RSV-DPM-V model outperforms both benchmarks in forecasting returns and logRV. For FTSE, RSV-DPM-V is better that RSV-N in forecasting logRV and second best in predicting

return density. SV-DPM is marginally the best, with a logBF of 1.95 compared to RSV-DPM-V. For SPX, RSV-DPM-V outperforms again RSV-N in forecasting logRV. RSV-N is marginally better than RSV-DPM-V in return density forecasting with a logBF of 1.23. The fact that for all series RSV-DPM-V is better in predicting logRV density than RSV-N, validates that the flexible estimation of h_t helps to better model the non-Gaussian features of logRV.

Some insight to the forecasting results of Table 2.5 is given in Figures 2.10, 2.11 and 2.12. These show how the logBFs of the RSV-DPM-V model against the benchmarks change throughout the out-of-sample forecasting period. In returns density forecasting, RSV-DPM-V model tends to perform better when returns have large negative drops. This indicates the model's flexibility to account for variance and returns shocks. This is also shown in logRV density forecasting where RSV-DPM-V performs better than RSV-N in periods of high variance such as during the start of the pandemic in March 2020.

I also test the proposed model on the ability to forecast the left region of the return distribution. This is an important task in financial risk management. I calculate the cumulative log predictive likelihood based on the region predictive density in (2.34), for out-of-sample returns less than 0 and -1. The results are in Table 2.6. For DAX, RSV-DPM-V is the best model in forecasting negative returns and second best when returns are less than -1%, following closely SV-DPM. RSV-DPM-V is better than both benchmarks in predicting the left side returns of FTSE. For the least volatile series of SPX, RSV-N is the best at forecasting negative returns and performs close to RSV-DPM-V when returns are less than -1%.

The proposed RSV-DPM-V identifies non-Gaussian volatility features for all the data series. In forecasting, RSV-DPM-V seems to perform well for the more volatile series of DAX and FTSE, compared to SPX in which is close to the Gaussian framework of RSV-N. A model application to individual stocks, which tend to be more volatile than indices, is left for future work. Also, due to its flexibility in variance estimation and in the left tail prediction, RSV-DPM-V can be used for Value-at-Risk and Expected Shortfall forecasts.

2.5 Concluding remarks

This chapter explores the empirical distributional features of stochastic volatility. A new semi-parametric SV model is proposed where latent log-volatility has no underlying assumption about its distribution. Instead is estimated as an infinite mixture of normal with a Dirichlet process prior. RV information is used to assist the estimation of the unobserved volatility. The model includes in the mixture the returns mean which gives the ability to capture the underlying return density.

Results from three equity indices indicate that stochastic volatility is not Gaussian. It is a mixture of normals and reveals strong evidence of asymmetry and thick tails. The flexible mixture estimation of volatility helps to capture non-Gaussian features in logRV and returns. In out-of-sample forecasting, the proposed model outperform the benchmarks in logRV density predictions. In return forecasting the proposed model performs the best for the most volatile series. It also assigns higher probability mass to negative returns for the two most volatile series. When returns are less than -1% the proposed model gives better forecasts for two of the three examined series.

TABLE 2.1: Summary statistics

DAX	Mean	Variance	Skewness	Kurtosis	Min	Max
r_t	-0.0277	1.5343	-0.1313	8.3761	-9.4122	9.9934
RV_t	1.5914	8.2967	7.7525	102.1241	0.0414	58.8347
logRV_t	-0.1642	1.0871	0.4308	3.2182	-3.1842	4.0747

Data period: January 3rd, 2000 - December 31st, 2021, 5,574 days.

FTSE	Mean	Variance	Skewness	Kurtosis	Min	Max
r_t	-0.0028	1.3268	-0.3702	10.3209	-10.2995	9.3872
RV_t	1.2045	8.4831	15.8272	415.6667	0.0133	106.0012
logRV_t	-0.4569	1.0166	0.5833	3.6644	-4.3182	4.6635

Data period: January 4th, 2000 - December 31st, 2021, 5,547 days.

SPX	Mean	Variance	Skewness	Kurtosis	Min	Max
r_t	0.0076	1.2396	-0.2289	11.3874	-9.3511	10.2202
RV_t	1.0855	6.7016	11.0013	205.2279	0.0122	77.4774
logRV_t	-0.6943	1.3112	0.3762	3.3814	-4.4079	4.3500

Data period: January 3rd, 2000 - December 31st, 2021, 5,515 days.

TABLE 2.2: Posterior estimation results for **DAX**.

	SV-DPM		RSV-N		RSV-DPM-V	
	Mean	95% D.I.	Mean	95% D.I.	Mean	95% D.I.
μ			0.0574	[0.053, 0.062]		
c			-0.1739	[-0.185,-0.163]	-0.1738	[-0.184, -0.164]
d			0.9652	[0.951, 0.980]	0.9629	[0.949, 0.977]
b^2			0.1616	[0.152, 0.172]	0.1450	[0.136, 0.154]
γ			-0.0008	[-0.008, 0.006]		
δ	0.9849	[0.978, 0.991]	0.9669	[0.959, 0.975]	0.9649	[0.957, 0.973]
σ_v^2	0.0373	[0.027, 0.049]	0.0649	[0.055, 0.075]		
α	0.6477	[0.225, 1.256]			0.3928	[0.099, 0.918]
κ	9.6135	[5.000, 16.000]			4.5828	[3.000, 8.000]

Notes: Results are from 10,000 MCMC posterior draws, after 5,000 burnin. α is the presicion parameter of the DPM and κ is the number of active normal clusters.

TABLE 2.3: Posterior estimation results for **FTSE**.

	SV-DPM		RSV-N		RSV-DPM-V	
	Mean	95% D.I.	Mean	95% D.I.	Mean	95% D.I.
μ			0.0499	[0.046, 0.054]		
c			-0.4572	[-0.470, -0.445]	-0.4572	[-0.469,-0.445]
d			0.9618	[0.944, 0.980]	0.9533	[0.936, 0.970]
b^2			0.2203	[0.209, 0.233]	0.2001	[0.188, 0.212]
γ			-0.0008	[-0.007, 0.005]		
δ	0.9773	[0.969, 0.985]	0.9654	[0.957, 0.974]	0.9561	[0.946, 0.965]
σ_v^2	0.0439	[0.032, 0.057]	0.0589	[0.050, 0.068]		
α	0.2047	[0.021, 0.629]			0.4836	[0.112, 1.123]
κ	2.4433	[1.000, 8.000]			5.6436	[3.000, 10.000]

Notes: Results are from 10,000 MCMC posterior draws, after 5,000 burnin. α is the presicion parameter of the DPM and κ is the number of active normal clusters.

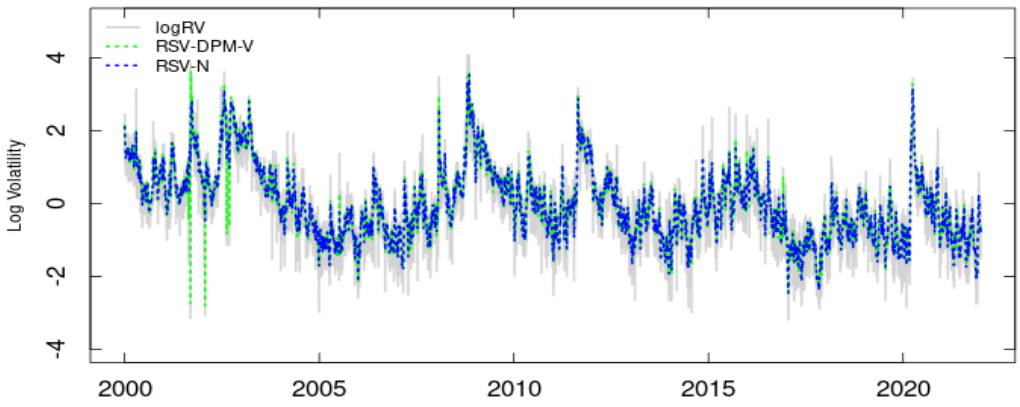
TABLE 2.4: Posterior estimation results for **SPX**.

	SV-DPM		RSV-N		RSV-DPM-V	
	Mean	95% D.I.	Mean	95% D.I.	Mean	95% D.I.
μ			0.0736	[0.070, 0.077]		
c			-0.6989	$[-0.710, -0.688]$	-0.6990	$[-0.710, -0.688]$
d			0.9527	[0.937, 0.968]	0.9561	[0.941, 0.971]
b^2			0.1832	[0.171, 0.196]	0.1801	[0.168, 0.192]
γ			-0.0007	$[-0.010, 0.008]$		
δ	0.9828	[0.976, 0.989]	0.9545	[0.945, 0.964]	0.9581	[0.949, 0.967]
σ_v^2	0.0514	[0.039, 0.068]	0.1105	[0.096, 0.126]		
α	0.5151	[0.141, 1.096]			0.3723	[0.094, 0.880]
κ	7.4500	[3.000, 14.000]			4.3404	[3.000, 8.000]

Notes: Results are from 10,000 MCMC posterior draws, after 5,000 burnin. α is the presicion parameter of the DPM and κ is the number of active normal clusters.

FIGURE 2.1: **DAX** logRV and estimated log-volatility: $\mathbb{E}(h_t|\mathcal{I}_T)$.

(a) $\mathbb{E}(h_t|\mathcal{I}_T)$ from the models: RSV-DPM-V and RSV-N.



(b) $\mathbb{E}(h_t|\mathcal{I}_T)$ from the models: RSV-DPM-V and SV-DPM.

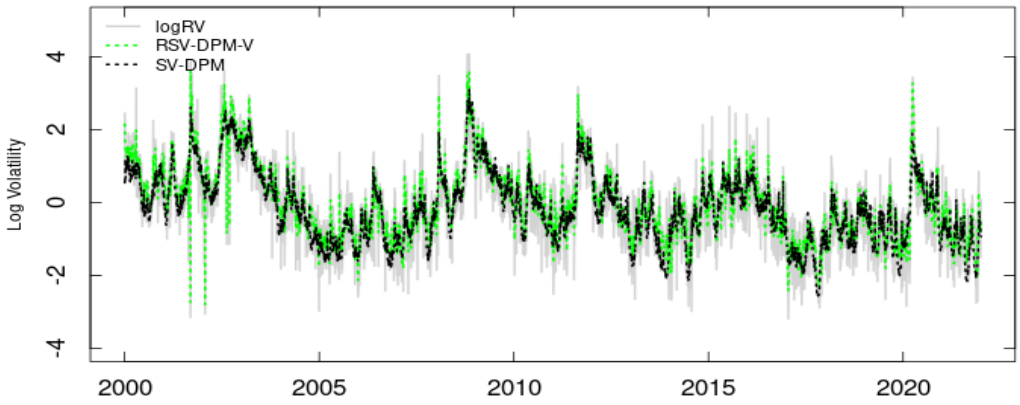
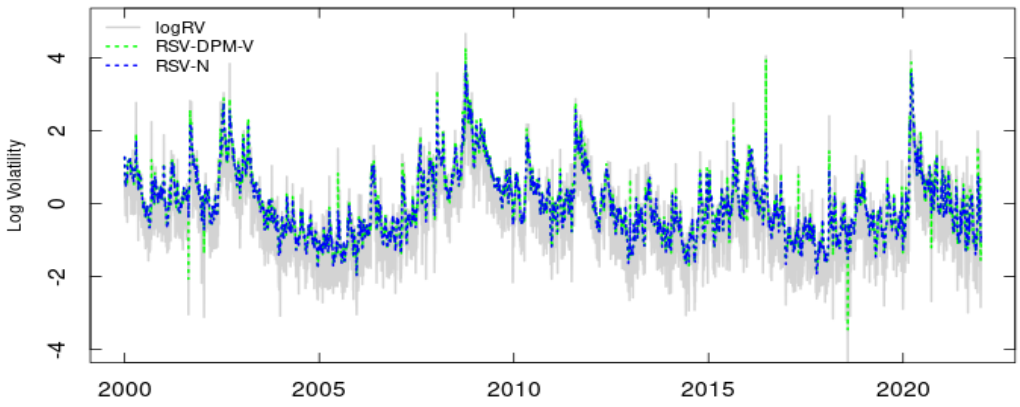


FIGURE 2.2: **FTSE** logRV and estimated log-volatility: $\mathbb{E}(h_t|\mathcal{I}_T)$.

(a) $\mathbb{E}(h_t|\mathcal{I}_T)$ from the models: RSV-DPM-V and RSV-N.



(b) $\mathbb{E}(h_t|\mathcal{I}_T)$ from the models: RSV-DPM-V and SV-DPM.

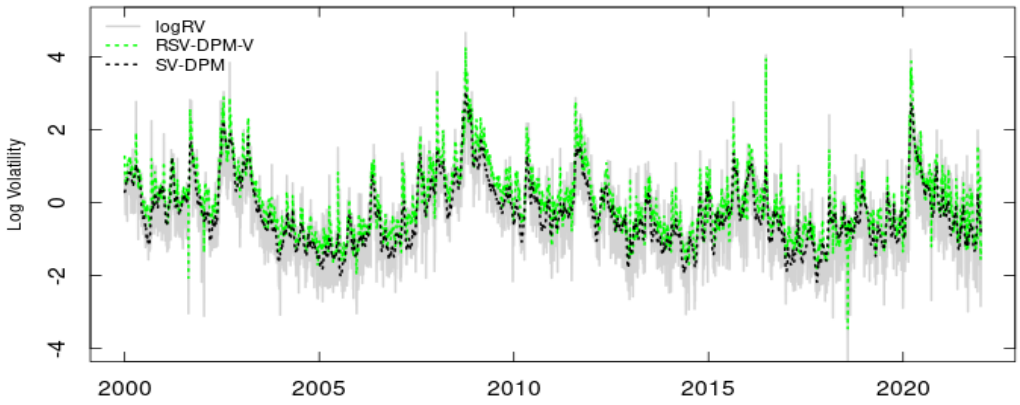
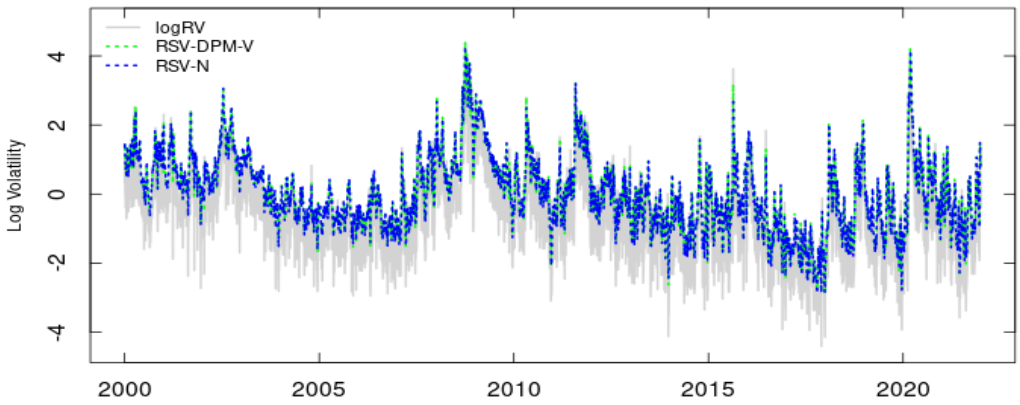


FIGURE 2.3: **SPX** logRV and estimated log-volatility: $\mathbb{E}(h_t|\mathcal{I}_T)$.

(a) $\mathbb{E}(h_t|\mathcal{I}_T)$ from the models: RSV-DPM-V and RSV-N.



(b) $\mathbb{E}(h_t|\mathcal{I}_T)$ from the models: RSV-DPM-V and SV-DPM.

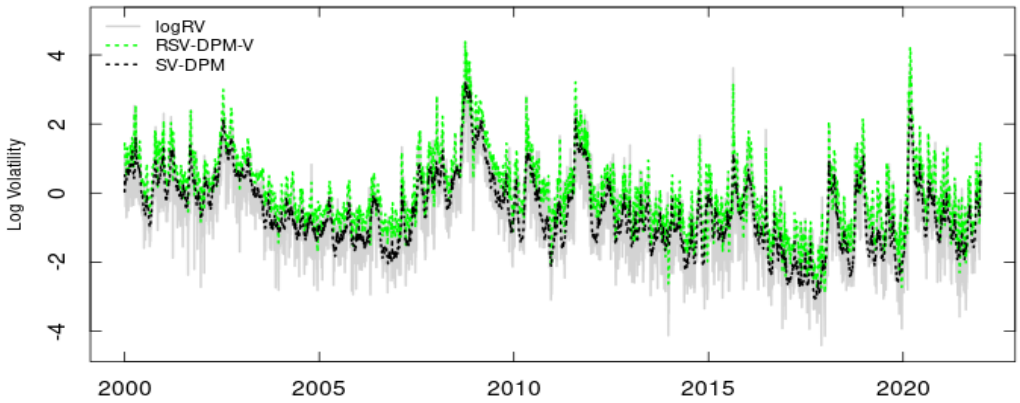
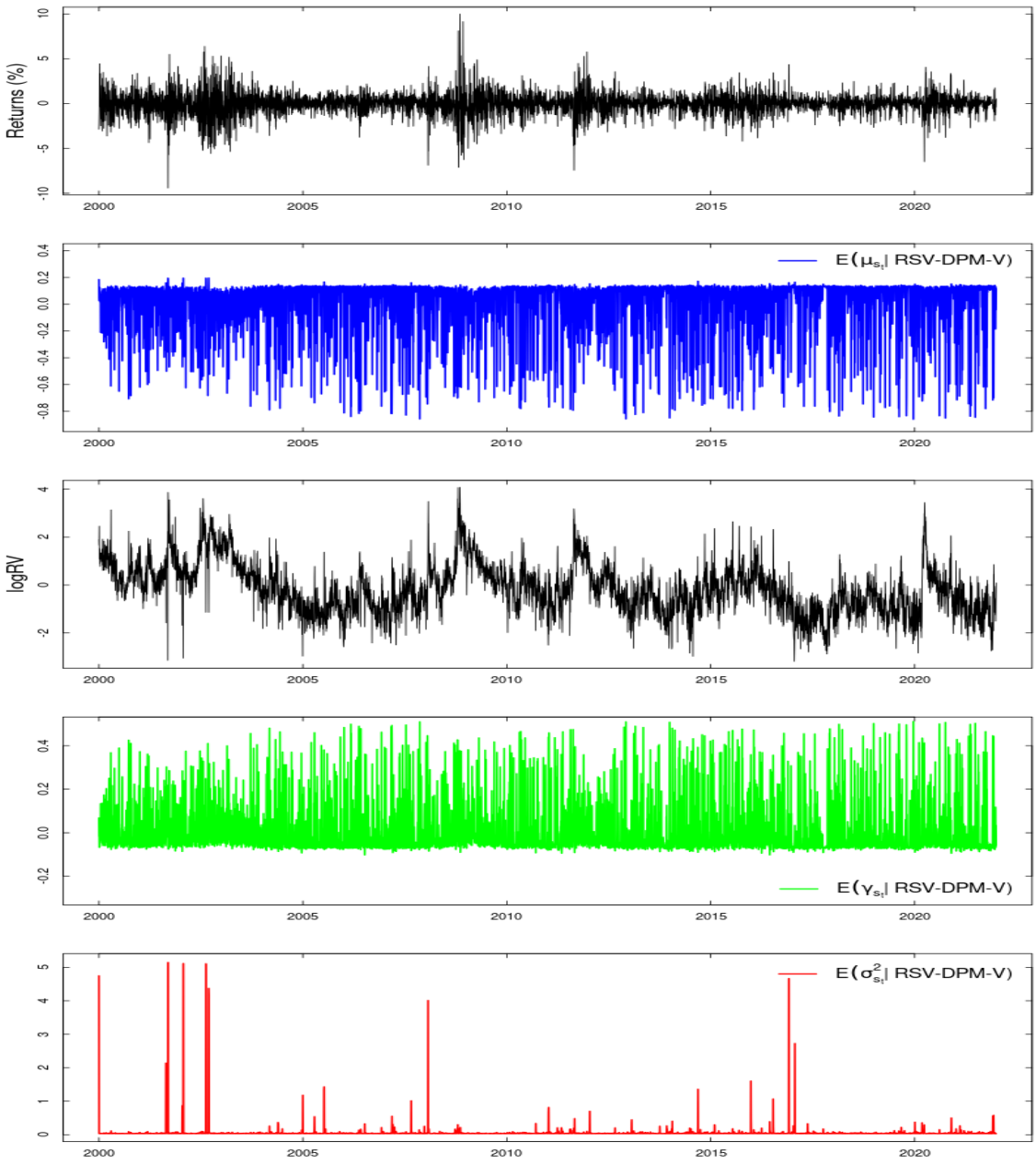
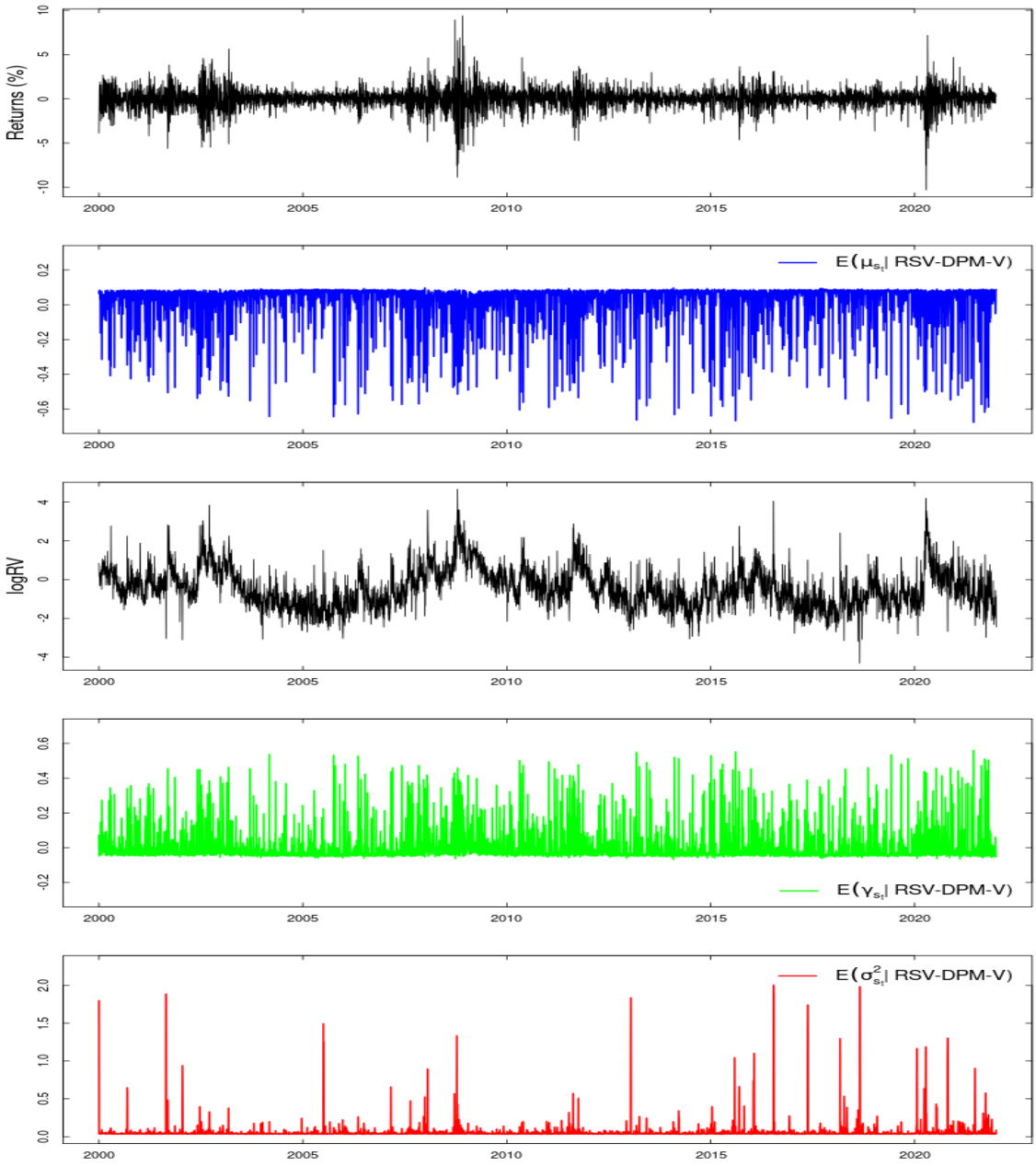


FIGURE 2.4: Mixture parameters for **DAX**.



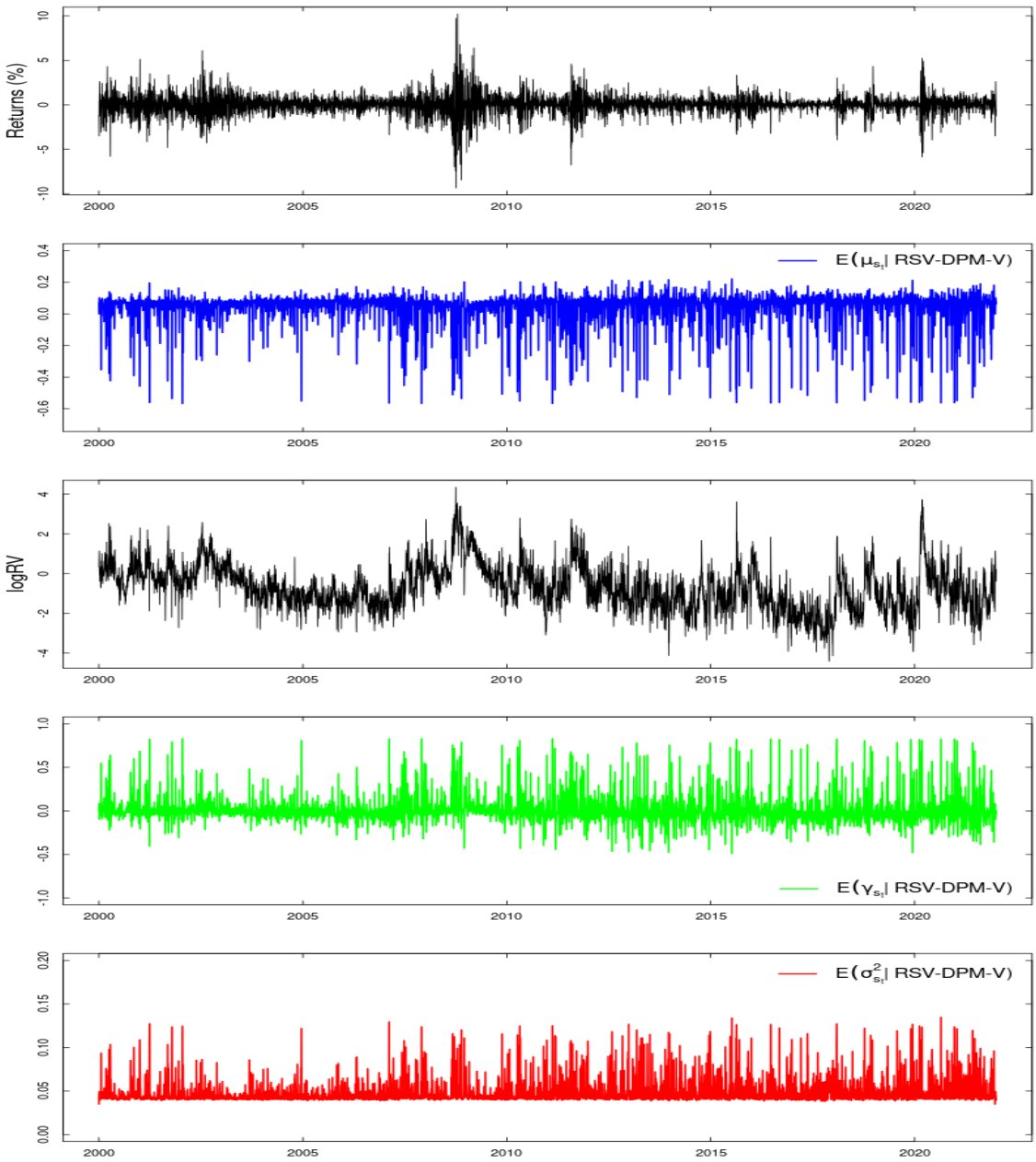
From top to bottom: **DAX** returns (black), estimated return mean $\mathbb{E}(\mu_{s_t}|\mathcal{I}_T, \text{RSV-DPM-V})$ (blue), **DAX** logRV (black), estimated st. vol. constant $\mathbb{E}(\gamma_{s_t}|\mathcal{I}_T, \text{RSV-DPM-V})$ (green) and estimated volatility of volatility $\mathbb{E}(\sigma^2_{s_t}|\mathcal{I}_T, \text{RSV-DPM-V})$ (red).

FIGURE 2.5: Mixture parameters for **FTSE**.



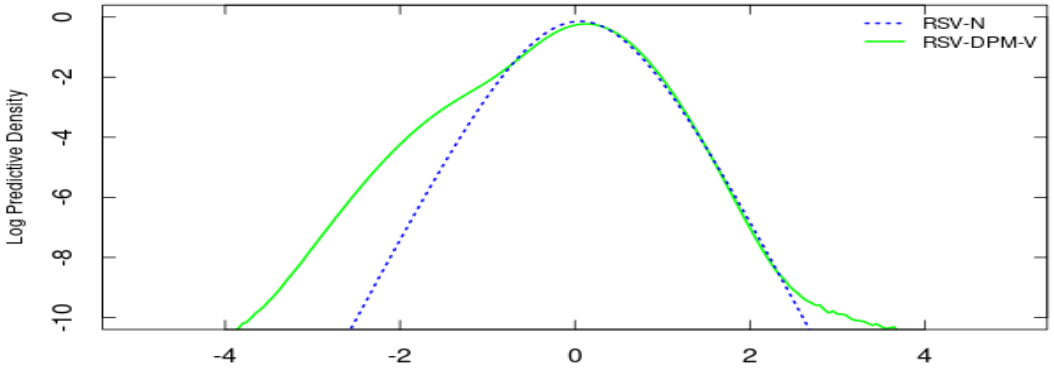
From top to bottom: **FTSE** returns (black), estimated return mean $\mathbb{E}(\mu_{s_t}|\mathcal{I}_T, \text{RSV-DPM-V})$ (blue), **FTSE** logRV (black), estimated st. vol. constant $\mathbb{E}(\gamma_{s_t}|\mathcal{I}_T, \text{RSV-DPM-V})$ (green) and estimated volatility of volatility $\mathbb{E}(\sigma^2_{s_t}|\mathcal{I}_T, \text{RSV-DPM-V})$ (red).

FIGURE 2.6: Mixture parameters for **SPX**.

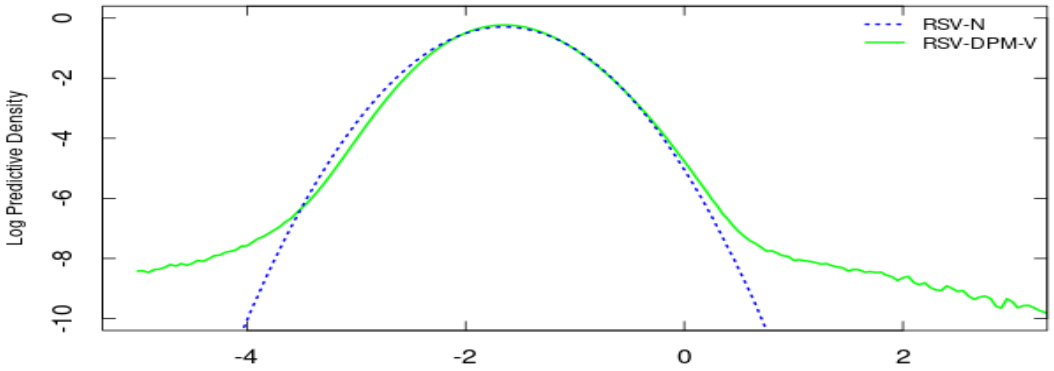


From top to bottom: **SPX** returns (black), estimated return mean $\mathbb{E}(\mu_{s_t}|\mathcal{I}_T, \text{RSV-DPM-V})$ (blue), **SPX** logRV (black), estimated st. vol. constant $\mathbb{E}(\gamma_{s_t}|\mathcal{I}_T, \text{RSV-DPM-V})$ (green) and estimated volatility of volatility $\mathbb{E}(\sigma^2_{s_t}|\mathcal{I}_T, \text{RSV-DPM-V})$ (red).

(a) Log predictive density of **DAX returns**: $\log p(r_{T+1}|\mathcal{I}_T)$



(b) Log predictive density of **DAX logRV**: $\log p(\text{logRV}_{T+1}|\mathcal{I}_T)$



(c) Log predictive density of **DAX stochastic log-volatility**: $\log p(h_{T+1}|\mathcal{I}_T)$

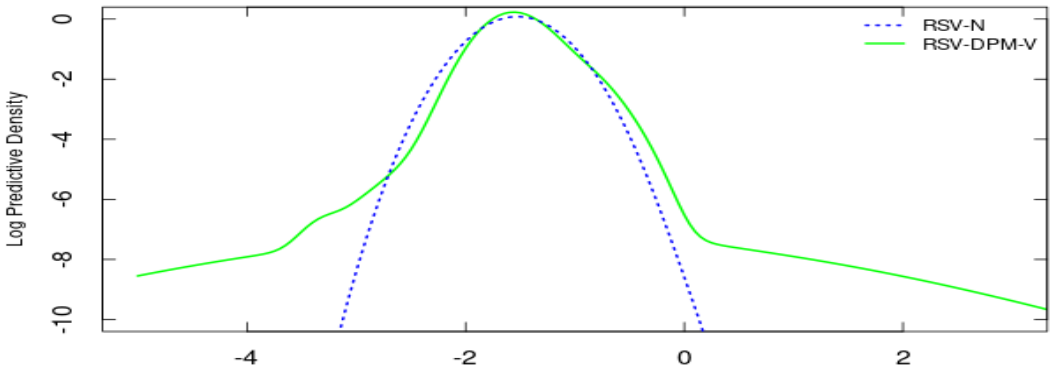
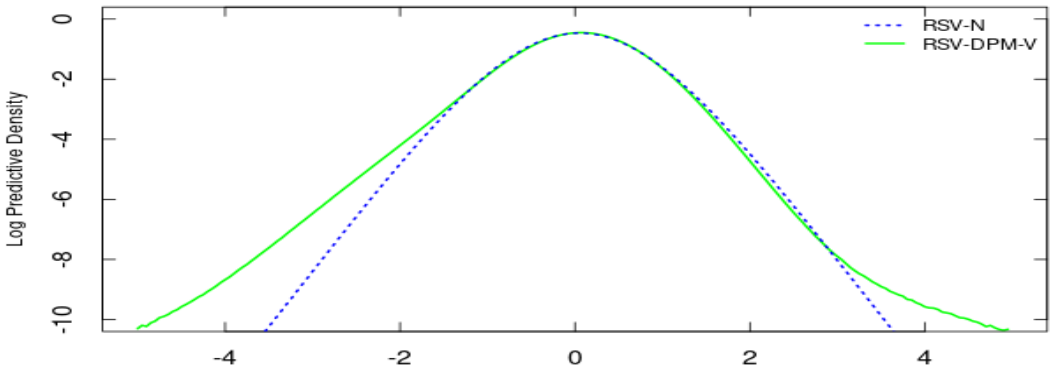
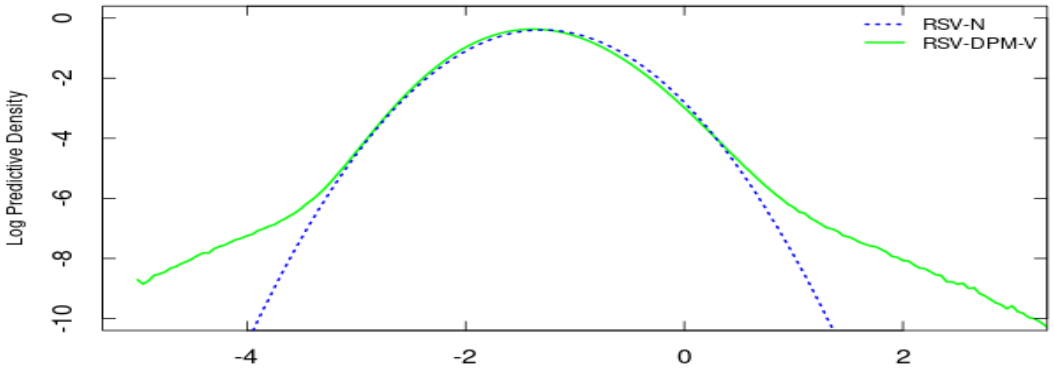


FIGURE 2.7: Log predictive densities of (a) **DAX returns**, (b) **DAX logRV** and (c) **DAX stochastic log-volatility** from RSV-N and RSV-DPM-V.

(a) Log predictive density of **FTSE returns**: $\log p(r_{T+1}|\mathcal{I}_T)$



(b) Log predictive density of **FTSE logRV**: $\log p(\text{logRV}_{T+1}|\mathcal{I}_T)$



(c) Log predictive density of **FTSE stochastic log-volatility**: $\log p(h_{T+1}|\mathcal{I}_T)$

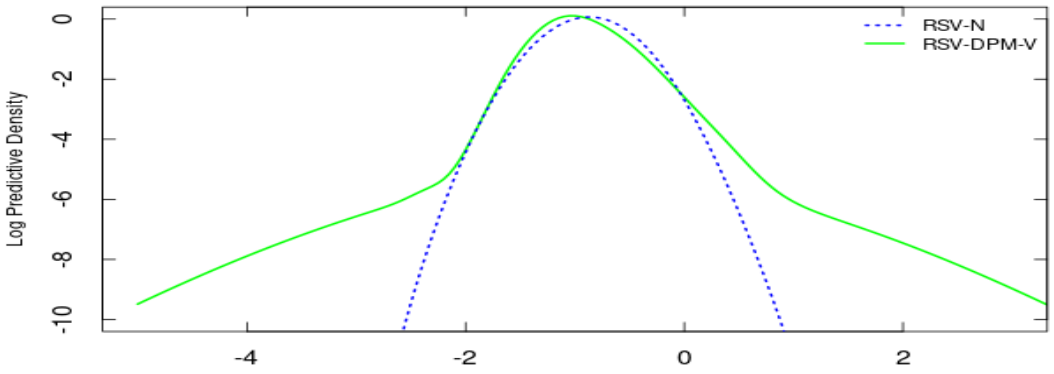
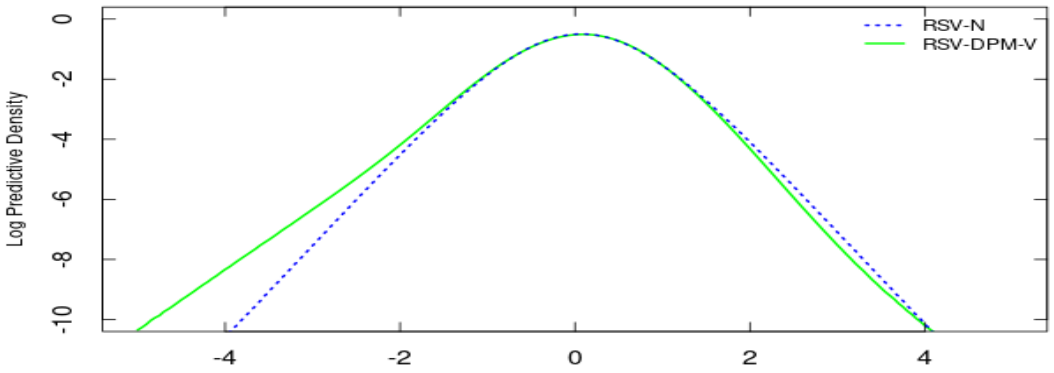
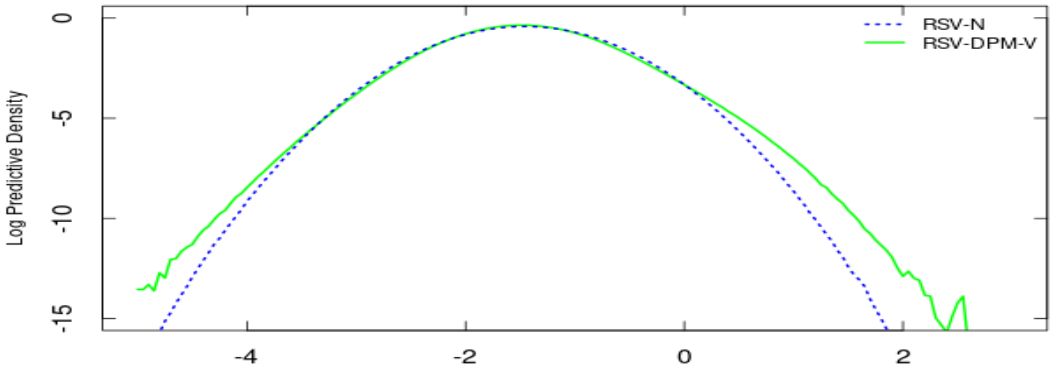


FIGURE 2.8: Log predictive densities of (a) **FTSE returns**, (b) **FTSE logRV** and (c) **FTSE stochastic log-volatility** from RSV-N and RSV-DPM-V.

(a) Log predictive density of **SPX returns**: $\log p(r_{T+1}|\mathcal{I}_T)$



(b) Log predictive density of **SPX logRV**: $\log p(\text{logRV}_{T+1}|\mathcal{I}_T)$



(c) Log predictive density of **SPX stochastic log-volatility**: $\log p(h_{T+1}|\mathcal{I}_T)$

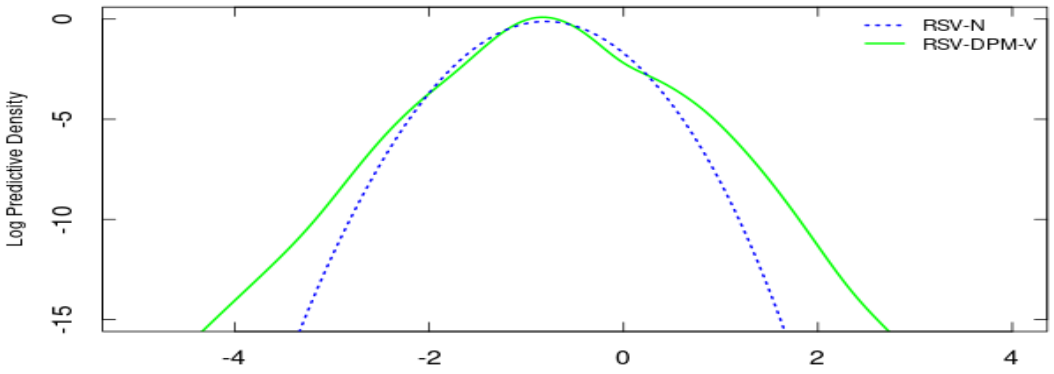


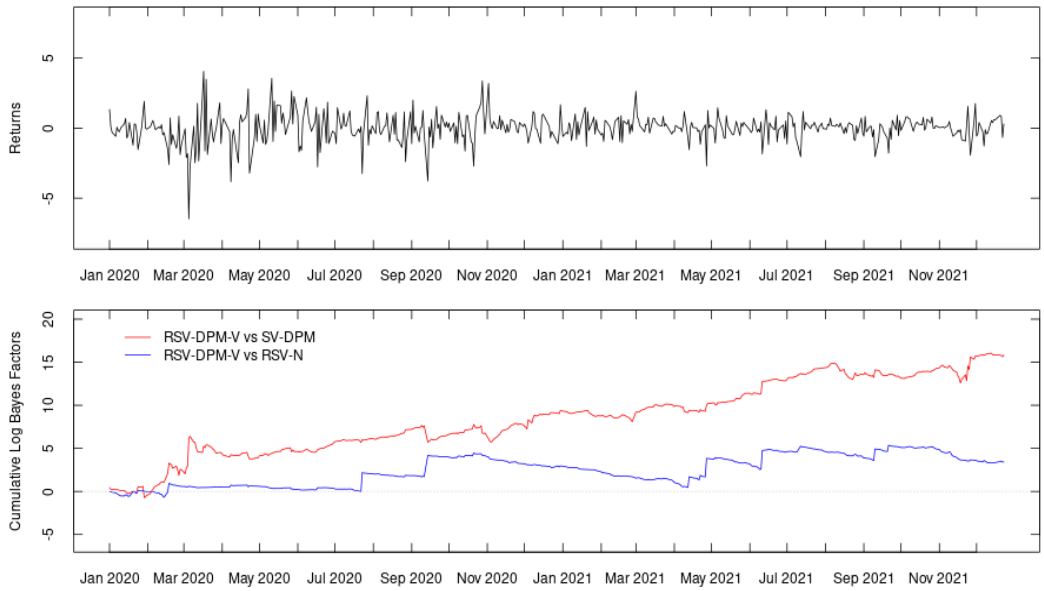
FIGURE 2.9: Log predictive densities of (a) **SPX returns**, (b) **SPX logRV** and (c) **SPX stochastic log-volatility** from RSV-N and RSV-DPM-V..

TABLE 2.5: Density forecasting results.

	DAX		FTSE		SPX	
	returns	logRV	returns	logRV	returns	logRV
SV-DPM	−642.55		**−749.89**		−620.05	
RSV-N	−630.12	−433.62	−752.59	−602.36	**−609.49**	−538.98
RSV-DPM-V	**−626.71**	**−429.09**	−751.84	**−588.77**	−610.76	**−533.36**
Forecasts (τ)	500		500		500	
Time period	08/01/2020 - 31/12/2021		08/01/2020 - 31/12/2021		30/12/2019 - 31/12/2021	

Notes: This table displays the cumulative log predictive likelihood (**logPL**) of returns and logRV. Bold values are the best among the models.

(a) **DAX** returns and cumulative logBF of RSV-DPM-V against the benchmarks.



(b) **DAX** logRV and cumulative logBF of RSV-DPM-V against RSV-N.

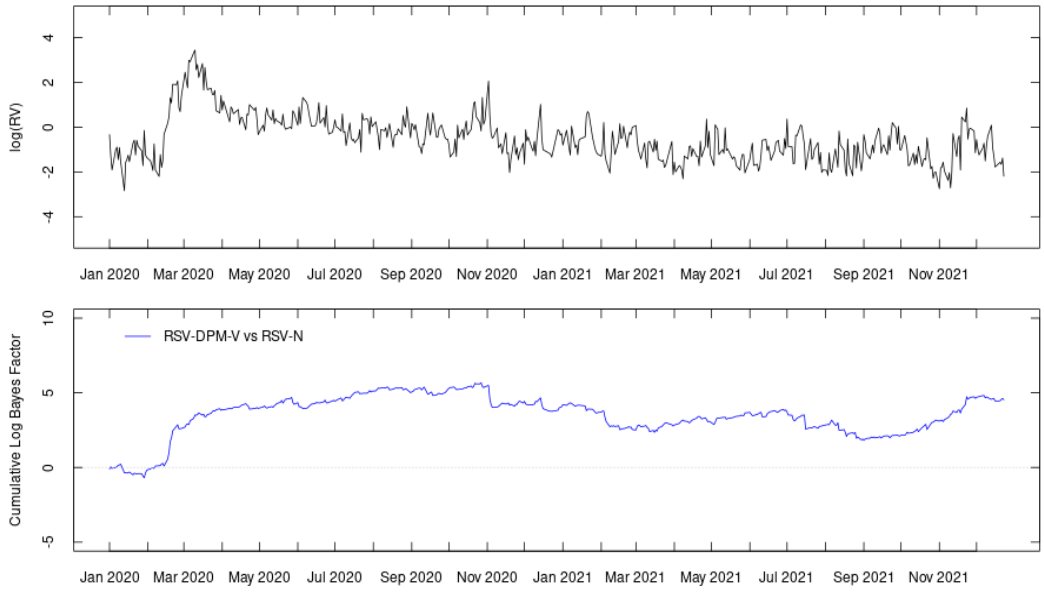
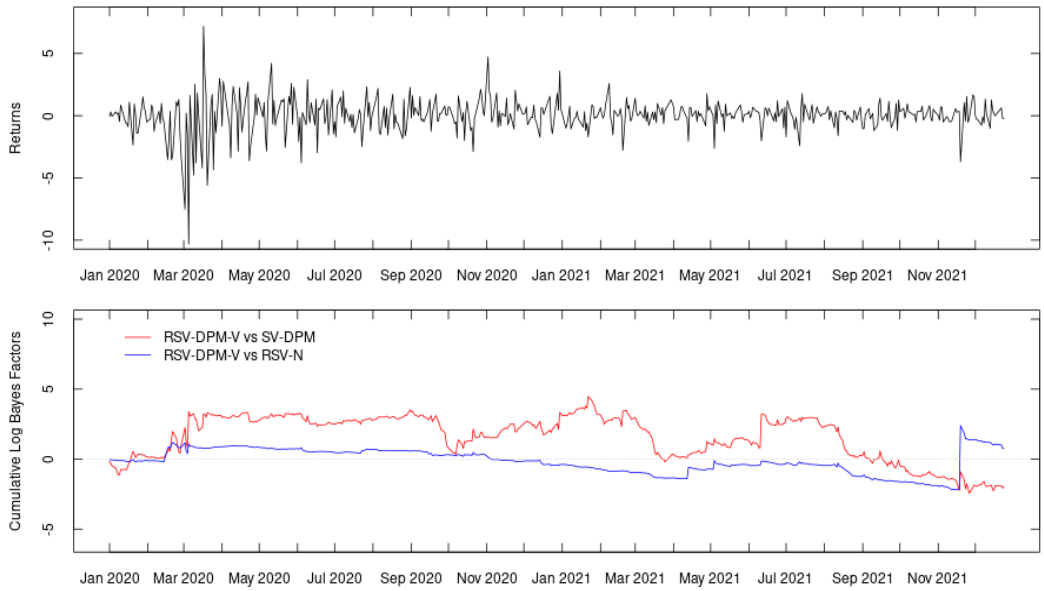


FIGURE 2.10: (a) **DAX** returns (top) and cumulative logBF of RSV-DPM-V vs SV-DPM and RSV-DPM-V vs RSV-N (bottom).
(b) **DAX** logRV (top) and cumulative logBF of RSV-DPM-V vs RSV-N (bottom).

(a) **FTSE** returns and cumulative logBF of RSV-DPM-V against the benchmarks.



(b) **FTSE** logRV and cumulative logBF of RSV-DPM-V against RSV-N.

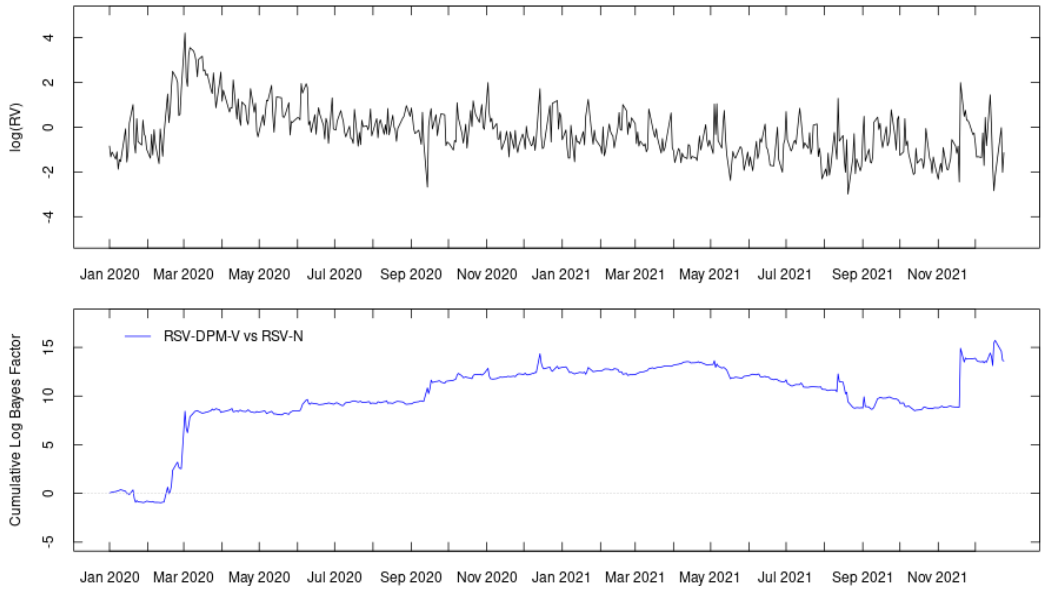
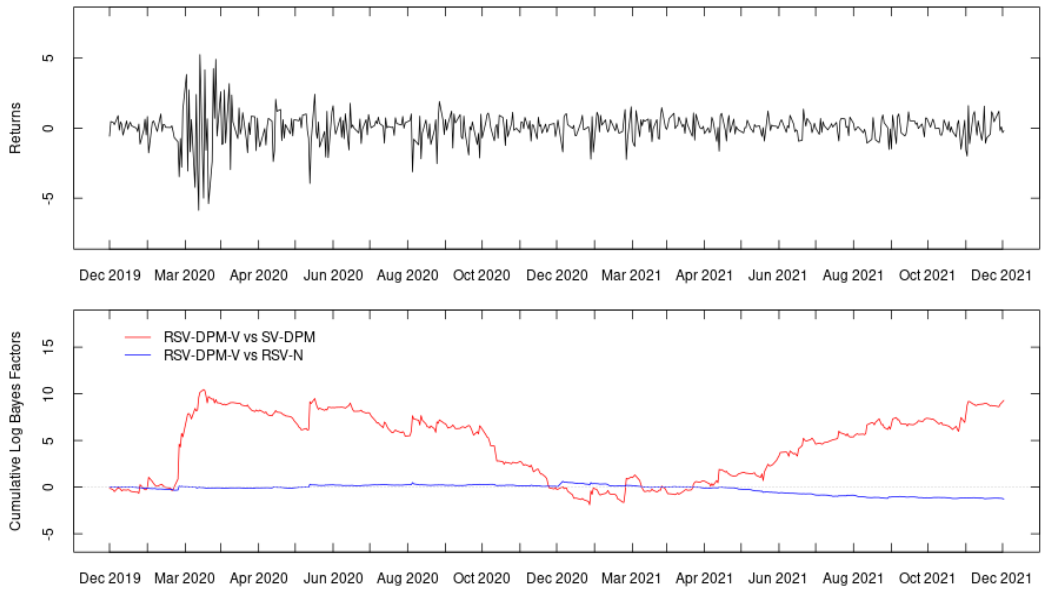


FIGURE 2.11: (a) **FTSE** returns (top) and cumulative logBF of RSV-DPM-V vs SV-DPM and RSV-DPM-V vs RSV-N (bottom).
(b) **FTSE** logRV (top) and cumulative logBF of RSV-DPM-V vs RSV-N (bottom).

(a) **SPX** returns and cumulative logBF of RSV-DPM-V against the benchmarks.



(b) **SPX** logRV and cumulative logBF of RSV-DPM-V against RSV-N.

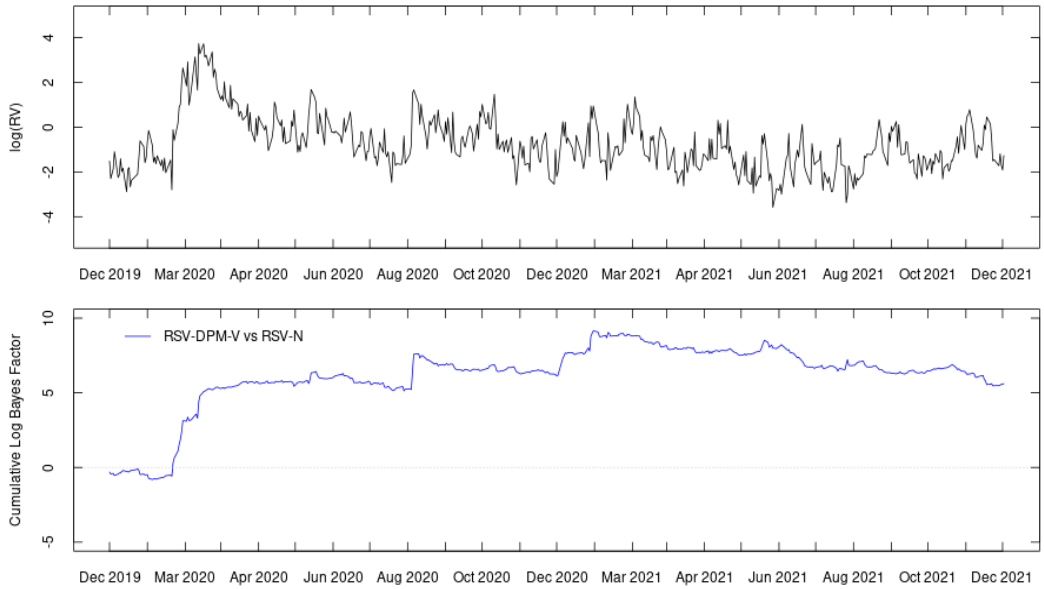


FIGURE 2.12: (a) **SPX** returns (top) and cumulative logBF of RSV-DPM-V vs SV-DPM and RSV-DPM-V vs RSV-N (bottom).
(b) **SPX** logRV (top) and cumulative logBF of RSV-DPM-V vs RSV-N (bottom).

TABLE 2.6: Return density region forecasting results.

	DAX		FTSE		SPX	
η (%):	-1	0	-1	0	-1	0
SV-DPM	$-$**46.88**	-162.06	-86.32	-208.47	-57.90	-156.84
RSV-N	-52.74	-159.82	-84.73	-205.43	-52.62	$-$**135.51**
RSV-DPM-V	-47.43	$-$**153.64**	$-$**79.07**	$-$**200.73**	$-$**52.21**	-137.04
Forecasts (τ)	60	233	72	232	57	219
Time period	08/01/2020 - 31/12/2021		08/01/2020 - 31/12/2021		30/12/2019 - 31/12/2021	

Notes: This table displays the cumulative region log predictive likelihood (**logPL**). It is constructed by the region predictive densities: $p(r_{t+1}|r_{t+1} < \eta, \mathcal{I}_t)$, for $\eta = \{-1, 0\}$. The chosen regions are negative returns ($r_{t+1} < 0$) and more than a 1% loss ($r_{t+1} < -1$). Bold values are the best among the models.

Chapter 3

The role of non-Gaussian innovations in multivariate realized GARCH models

3.1 Introduction

This chapter explores the impact of non-Gaussian innovations in multivariate realized GARCH (MRGARCH) models that use realized covariance (RC) information in the construction of the conditional return covariance. A new approach is proposed to jointly model returns and RC in a multivariate GARCH framework. The empirical non-Gaussian features of returns are captured with two novel models. The first one is a parametric version with multivariate-t innovations. The second one is a nonparametric approximation of the return distribution using an infinite mixture of multivariate normals, given a Dirichlet process (DP) prior. The proposed models are based on the assumption that RC follows an Inverse Wishart (IW) distribution with conditional mean set to the conditional covariance of returns. The benefits of the proposed models are addressed with forecasting and portfolio risk management applications.

Modeling the covariance of asset returns is important in portfolio optimization, asset pricing and risk management. Typically, multivariate extensions of GARCH models (Engle 1982;

Bollerslev 1986) are used to estimate the conditional covariance. Due to the complexity of multivariate modeling different specifications have been proposed in literature. The most popular have been the model of Bollerslev et al. (1988), the Dynamic Conditional Correlation (DCC) model proposed by Engle (2002a) and restricted versions of the BEKK model of Baba et al. (1990) and Engle and Kroner (1995). A comprehensive study on the multivariate GARCH (MGARCH) models is from Bauwens et al. (2006). All these models are based on low frequency (daily, weekly, monthly) squared returns to construct the conditional covariance.

Realized measures of covariance, based on high frequency returns[1], have been used to improve the construction of conditional covariance matrices. The realized beta GARCH of Hansen et al. (2014) is a bivariate model that captures the dynamics of conditional variances and correlations in a DCC framework which links logarithmic realized and conditional measures with linear measurement equations. Under the same framework, the multivariate realized GARCH model of Archakova et al. (2019) is the generalization of realized GARCH models, estimated with a two-step method. Gorgi et al. (2019) developed the realized Wishart GARCH, a multivariate model where RC is assumed to have a Wishart distribution with mean that incorporates the conditional covariance matrix which is constructed from a score-GARCH specification. Other models based on RC have been found to perform better than standard MGARCH in forecasting multivariate return distributions as shown by Noureldin et al. (2012) and Jin and Maheu (2013). RC has also been used as extra data information to improve multivariate stochastic volatility models by Shirota et al. (2017) and Yamauchi and Omori (2020).

Concerning the return distribution, the Gaussian assumption is typically used in the above models. This is an unrealistic assumption given the fact that empirically returns have negative-skewed distribution with thick tails. Diamantopoulos and Vrontos (2010) proposed an MGARCH model with Student-t innovations, a specification that would capture thick tails but it still remains symmetric. Jensen and Maheu (2013) proposed a vector-diagonal semiparametric MGARCH

[1]For instance, daily realized covariance matrices are constructed from high-frequency (e.g. 5-minute) intraday returns.

with no distributional assumption where instead returns follow an infinite mixture of normals using a Dirichlet process prior. This flexible specification can approximate any continuous return distribution and yield better density forecasts than parametric models.

Combining realized measures along with thick tails for returns improves covariance modeling. Opschoor et al. (2018) extended the HEAVY model of Noureldin et al. (2012) by accounting for thick tails in modeling RC, using the matrix-F distribution, along with Student-t distributed returns. Jin and Maheu (2016) model RC using infinite mixtures of IW kernels. This results in a mixture of Student-t kernels in the conditional return distribution. These models provide better returns density forecasts. In this chapter I extend the MRGARCH models with several non-Gaussian return distributional assumptions.

This chapter makes three contributions. First, a Bayesian general framework is developed for MRGARCH models. I model daily returns along with daily RC which is used as a signal for the conditional covariance. In contrast to Hansen et al. (2014) and Archakova et al. (2019) who breakdown the covariance modeling in univariate linear specifications, I jointly model the RC elements with an IW density. My approach on modeling RC is closer to the realized Wishart-GARCH of Gorgi et al. (2019) but with two important differences. The first one is that, compared to Wishart, the IW distribution provides a better fit for realized covariance matrices, with major forecasting benefits, as shown by Jin and Maheu (2016). The second one is that the conditional covariance in my framework is built with a vector-diagonal asymmetric MGARCH extension which, compared to the restricted score-GARCH of Gorgi et al. (2019), allows for richer dynamics in the conditional covariance.

The proposed MRGARCH specification captures the asymmetric volatility feedback in a parsimonious way. The framework is a multivariate extension of the asymmetric GARCH of Engle and Ng (1993). This helps to capture the leverage effect in the MGARCH specification without the use of parametric leverage functions as Hansen et al. (2014) and Archakova et al. (2019). I find that a leverage effect captured through lag returns provides incremental information for the

conditional covariance of returns.

The second contribution is that the proposed general MRGARCH framework is extended to accommodate thick tails and asymmetry in returns. Two new models are developed that drop the restrictive Gaussian assumption. The first one has a parametric multivariate-t distribution which captures the thick tails observed in returns. The second one has no underlying assumption but instead models the latent return distribution using a nonparametric infinite mixture of normals. This specification is an extension of the models developed by Jensen and Maheu (2013), and is flexible enough to approximate any continuous distribution and to capture thick tails and asymmetry in returns.

Results from two equity datasets show deviations from the Gaussian assumption for the returns. I find significant thick tails but little evidence of asymmetry in the return density of my data. The new proposed semiparametric MRGARCH specification can produce covariance matrices closer to RC, than the parametric models, and this is reflected with improved forecasts of RC density. In addition, modeling the tails gives major improvements in forecasting the return density.

The third contribution is the portfolio risk management applications of the MRGARCH models. For an equal weight investment, I examine the ability of the proposed models to forecast Value-at-Risk and Expected Shortfall, out-of-sample. I use the loss function from Fissler and Ziegel (2016) and Patton et al. (2019), to evaluate the models performance on jointly forecasting the aforementioned tail risk measures. The proposed semiparametric models consistently produce the best or the second best forecasts of the two measures. The benefits of non-Gaussian innovations in MRGARCH models are also shown from forming minimum variance portfolios. The new MRGARCH models which accommodate thick tails have the lowest out-of-sample portfolio variance.

The chapter is organized as follows: section 3.2 briefly presents the setup of RC, in section 3.3 is the model description along with the estimation process and the density forecasting setup,

section 3.4 contains the empirical applications and section 3.5 concludes.

3.2 Realized Covariance

Realized measures of covariation are based on high-frequency (intraday) data to estimate the latent low-frequency (daily) covariance matrix. They are multivariate extensions of realized variance measures, Andersen and Bollerslev (1998) and Barndorff-Nielsen and Shephard (2002a).

For a trading day t, we can have Q high frequency observations for the n-length vector of logarithmic asset prices $p_{t,i},\ i = 1, ..., Q$. Based on these, high-frequency log returns are calculated as

$$r_{t,i} = p_{t,i} - p_{t,i-1}, \tag{3.1}$$

with $r_{t,i} \in \mathbb{R}^n$ being the return vector for n series at time i of day t. Given these intraday returns, the nonparametric estimator of realized covariance (RC), for day t, is

$$RC_t = \sum_{i=1}^{Q} r_{t,i} r'_{t,i}, \tag{3.2}$$

with $RC_t \in \mathbb{R}^{n \times n}$ being a positive definite symmetric matrix where the diagonal elements consist the realized variance estimators and the off-diagonal elements are the realized covariances. Banrdorff-Nielsen and Shephard (2004) show that, under no market microstructure noise, when $Q \to \infty$, RC_t converges to the quadratic covariation, which is equal to the conditional return covariance according to Andersen et al. (2003).

The multivariate realized kernel of Barndorff-Nielsen et al. (2011) gives estimations robust to microstructure noise. Calculating RC_t by averaging on different subsampled estimations (subsampling) can reduce the noise from the high-frequency returns.

3.2.1 Data

I use the data of Noureldin et al. (2012), consisting of two different sets of daily open-to-close equity log returns, obtained from Oxford-Man Institute Realized Library. The first one consists of International Business Machines (**IBM**), Microsoft (**MSFT**) and Exxon Mobil (**XOM**) ($n = 3$ series) while the second one includes American Express (**AXP**), DuPont (**DD**), General Electric (**GE**) and Coca-Cola (**KO**) ($n = 4$ series). Datasets are from 2242 trading days for the sample period from February 1st, 2001 to December 31st, 2009. Realized measures of variance and covariance are calculated from open-to-close 5-minute returns, with subsampling using 1-minute returns for each trading day. Returns have been converted to percentages and RC matrices have been scaled by 100^2. Summary statistics are in Tables 3.1 and 3.2.

3.3 Multivariate realized GARCH

The models developed in this work have the vector-diagonal form of Ding and Engle (2001) but other MGARCH specifications could be used. An extension of their specification is used to allow for volatility feedback similar to the asymmetric GARCH of Engle and Ng (1993). Maheu and Shamsi Zamenjani (2021) use a similar MGARCH framework.

3.3.1 Gaussian innovations

MRGARCH-N: This proposed fully parametric multivariate realized GARCH model is an extension of the parametric MGARCH models to include RC information. RC_t is calculated from intraday high-frequency prices, and serves as a signal for the conditional return covariance. Instead of calculating conditional covariance solely based on (daily) squared returns, RC is used as extra data information. Given the information set $\mathcal{I}_{t-1} = \{r_{1:t-1}, RC_{1:t-1}\}$ available at day

$t-1$, the MRGARCH-N model is specified as

$$r_t|\mathcal{I}_{t-1} \sim \mathrm{N}_n(\mu, H_t), \tag{3.3}$$

$$RC_t|\mathcal{I}_{t-1} \sim \mathrm{IW}_n\left(\nu, (\nu-n-1)H_t^{1/2}V\left(H_t^{1/2}\right)'\right), \quad \nu > n+1, \tag{3.4}$$

$$H_t = \Omega + A \odot (r_{t-1}-\lambda)(r_{t-1}-\lambda)' + B \odot H_{t-1} + C \odot RC_{t-1}, \tag{3.5}$$

where r_t is a $n \times 1$ vector of log returns at time t (trading day) for n data series. $\mathrm{N}_n(\mu, H_t)$ denotes the multivariate normal distribution with μ being the length n mean vector and H_t the $n \times n$ covariance matrix, at time t, with $n(n+1)/2$ unique elements. Conditional return covariance from (3.3) is $\mathrm{Cov}(r_t|\mathcal{I}_{t-1}) = H_t$.

Symbol $\odot$ is the Hadamard product. Parameters in (3.5) are parameterized as $A = a\,(a)'$, $B = b\,(b)'$ and $C = c\,(c)'$ where a, b, c and λ are length n vectors with λ being a nonlinear asymmetric response effect to covariance which can be considered as a multivariate version of the asymmetric GARCH model of Engle and Ng (1993). Ω is an $n \times n$ positive definite parameter matrix.

Eq.(3.4) is called measurement equation and models the joint dependence between r_t and RC_t. The measurement assumption is that RC_t follows the Inverse Wishart distribution, with ν degrees of freedom and scale matrix $(\nu-n-1)H_t^{1/2}V\left(H_t^{1/2}\right)'$. $H_t^{1/2}$ is the Cholesky decomposition of H_t. The conditional mean of RC_t is

$$\mathbb{E}[RC_t|\mathcal{I}_{t-1}] = \frac{(\nu-n-1)H_t^{1/2}V\left(H_t^{1/2}\right)'}{(\nu-n-1)} = H_t^{1/2}V\left(H_t^{1/2}\right)', \tag{3.6}$$

with V being an $n \times n$ positive definite parameter matrix which completes the model as it helps to capture deviations between RC_t and H_t. This makes the measurement equation flexible to adjust to possible noise contained in RC_t. If there is no market microstructure noise in RC_t, V would be approximately equal to an identity matrix and RC_t is an unbiased estimator of return covariance, $\mathbb{E}[RC_t|\mathcal{I}_{t-1}] = H_t$.

The degree of freedom parameter ν controls the dispersion of the IW distribution. From the above it can be seen that the scale matrix of IW distribution is positive definite as long as V and H_t are positive definite and $\nu > n + 1$. These conditions are ensured during model estimation.

Regarding the estimation of the positive definite matrix Ω, I use covariance targeting. Let $\bar{\Sigma}$ being the sample returns covariance matrix and $\bar{\mu}$ the sample returns mean vector. For the derivation of targeting I assume that conditional covariance is stationary, $\mathbb{E}(H_t) = \bar{\Sigma}$, and set $V \equiv I_n$. During the model estimation, Ω is targeted as[2]

$$\Omega = \bar{\Sigma} \odot (\iota\iota' - A - B - C) - A \odot (\bar{\mu} - \lambda)(\bar{\mu} - \lambda)'. \tag{3.7}$$

Compared to the existing multivariate realized GARCH models in literature, the proposed specification of MRGARCH-N has three key differences. First, the model developed by Archakova et al. (2019) link log realized and conditional moments through a set of $n(n+1)/2$ linear measurement equations. MRGARCH-N model jointly links RC_t in (3.4) to the full conditional covariance matrix. This parsimoniously allows full dependence between the RC_t and H_t elements. The approach is similar to Gorgi et al. (2019). However, they assume that RC_t follows a Wishart distribution. Empirical evidence, from Jin and Maheu (2016), show that an IW assumption, as in (3.4), is a better fit for realized covariance matrices, compared to a Wishart.

Second, Gorgi et al. (2019) use a score-GARCH framework for the construction of conditional covariance. MRGARCH-N has a standard MGARCH construction of conditional covariance, which is a multivariate extension of the GARCH-X model of Engle (2002b). This allows to estimate different effects, through GARCH parameters, among variances and covariances.

The third key difference is on modeling the asymmetry (leverage) effect between returns and covariance. Archakova et al. (2019) use two parametric leverage functions to capture the effect. In MRGARCH-N, the effect is parsimoniously captured by the vector λ. The use of lag RC in (3.5), is expected to reduce the magnitude of parameters a, based on the empirical results

[2]Details in appendix A4

of Engle (2002b) and Hansen et al. (2012). They found ARCH parameters to be insignificant. I find that λ has an estimation trade-off with a parameters, which are empirically significant.

Estimation

To estimate MRGARCH-N, I follow Bayesian analysis for the model parameters in $\theta = \{\mu, a, b, c, \lambda, \nu\}$ and V, with the following posterior density

$$p(\theta, V|\mathcal{I}_T) \propto p(\theta)p(V) \prod_{t=1}^{T} \mathrm{N}_n\left(r_t|\mu, H_t\right) \mathrm{IW}_n\left(RC_t\Big|\nu, (\nu-n-1)H_t^{1/2}V\left(H_t^{1/2}\right)'\right), \tag{3.8}$$

where $\mathrm{N}_n(.|)$ denotes the multivariate normal density and $\mathrm{IW}_n(.|)$ the IW density, both of dimension n. $p(.)$ denotes the prior density of parameters. I simulate the parameters by collecting $\{\theta^{(i)}\}_{i=1}^{R}$ and $\{V^{(i)}\}_{i=1}^{R}$, draws from the above posterior. Posterior means are calculated as $\mathbb{E}(\theta|\mathcal{I}_\mathcal{T}) = R^{-1}\sum_{i=1}^{R}\theta^{(i)}$. The above posterior does not have a known form. To draw from it, I initialize $\theta^{(0)}$ and $V^{(0)}$, and iterate, R times, the following Markov chain Monte Carlo (MCMC) steps:

1. Draw $\theta^{(i)}$ with a random walk Metropolis-Hastings (MH) algorithm with proposal θ' which is from

$$h(\theta') \sim \mathrm{N}(\theta^{(i-1)}, \varsigma\hat{V}_h), \tag{3.9}$$

where $\hat{V}_h$ is the inverse Hessian matrix evaluated at the posterior mode $\hat{\theta}$, which is computed once at the beginning of estimation. Scaling $\hat{V}_h$ by ς helps to achieve a desired acceptance rate. The draw θ' is accepted with probability

$$\min\left\{p(\theta'|\mathcal{I}_T)/p(\theta^{(i-1)}|\mathcal{I}_T), 1\right\}, \tag{3.10}$$

with $p(.|\mathcal{I}_T)$ the posterior given in (3.8). If θ' is rejected then $\theta^{(i)} = \theta^{(i-1)}$. Setting ς as

0.3 - 0.5 gives an acceptance frequency around 0.2 - 0.4 in the MH step. This allows the sampler to fully explore the posterior. For every $\theta^{(i)}$, $\Omega^{(i)}$ is calculated as in (3.7). Only posterior draws that result in positive definite $\Omega^{(i)}$ and $H_{1:T}^{(i)}$ are accepted.

2. Draw $V^{(i)}|H_{1:T}^{(i)}, RC_{1:T}, \nu^{(i)}$ with a Gibbs sampler. Given a Wishart prior, $V \sim \mathrm{W}_n\left(\psi_0, \psi_0^{-1} I_n\right)$, the posterior is

$$\begin{aligned} p(V|RC_{1:T}, H_{1:T}, \nu) &\propto p(V)p(RC_{1:T}|V, H_{1:T}, \nu) \\ &\propto \mathrm{W}_n\left(V|\psi_0, \psi_0^{-1} I_n\right) \prod_{t=1}^{T} \mathrm{IW}_n\left(RC_t\Big|\nu, (\nu-n-1)H_t^{1/2} V \left(H_t^{1/2}\right)'\right) \\ &\sim \mathrm{W}_n(\nu_p, \Psi_p), \end{aligned} \tag{3.11}$$

with $\nu_p = \psi_0 + T\nu$ and $\Psi_p = \left[(\psi_0^{-1} I_n)^{-1} + (\nu-n-1)\sum_{t=1}^{T}\left(H_t^{1/2}\right)' RC_t^{-1} H_t^{1/2}\right]^{-1}$.

3.3.2 Thick tailed innovations

Empirically, Gaussian innovations is a restrictive assumption for financial returns Richardson and Smith (1993) and Ding and Engle (2001). A more flexible distributional assumption that can capture extreme tails is a better fit for the financial returns. Here the MRGARCH-N is extended to accommodate thick tails in distribution.

MRGARCH-t: This novel fully parametric model allows returns to follow a thick tailed multivariate Student-t distribution. Given the information set $\mathcal{I}_{t-1}$ available at time $t-1$, the MRGARCH-t model is specified as

$$r_t|\mathcal{I}_{t-1} \sim \mathrm{t}_n(\mu, H_t, \zeta), \quad \zeta > 2, \tag{3.12}$$

$$RC_t|\mathcal{I}_{t-1} \sim \mathrm{IW}_n\left(\nu, (\nu-n-1)\frac{\zeta}{\zeta-2} H_t^{1/2} V \left(H_t^{1/2}\right)'\right), \quad \nu > n+1, \tag{3.13}$$

$$H_t = \Omega + A \odot (r_{t-1}-\lambda)(r_{t-1}-\lambda)' + B \odot H_{t-1} + C \odot RC_{t-1}. \tag{3.14}$$

Model notation and parameterization is the same as in MRGARCH-N. $\mathrm{t}_n(\mu, H_t, \zeta)$ is the multivariate Student-t (St-t) distribution of dimension n with mean μ, scale matrix H_t and degree of freedom ζ. Conditional return covariance from (3.12) is $\mathrm{Cov}(r_t|\mathcal{I}_{t-1}) = \frac{\zeta}{\zeta-2}H_t$. From (3.13), the conditional mean of RC_t is related with the returns conditional covariance through

$$\mathbb{E}(RC_t|\mathcal{I}_{t-1}) = \frac{(\nu-n-1)\frac{\zeta}{\zeta-2}H_t^{1/2}V\left(H_t^{1/2}\right)'}{\nu-n-1} = \frac{\zeta}{\zeta-2}H_t^{1/2}V\left(H_t^{1/2}\right)'. \tag{3.15}$$

If there is no market microstructure noise in RC_t, V would be approximately equal to an identity matrix and RC_t is an unbiased estimator of return covariance, $\mathbb{E}[RC_t|\mathcal{I}_{t-1}] = \mathrm{Cov}(r_t|\mathcal{I}_{t-1}) = \frac{\zeta}{\zeta-2}H_t$. I assume covariance stationarity, $\mathbb{E}(H_t) = \bar{\Sigma}$, and set $V \equiv I_n$, to use covariance targeting [3] for Ω as

$$\Omega = \bar{\Sigma} \odot \left(\iota\iota' - \frac{\zeta}{\zeta-2}A - B - \frac{\zeta}{\zeta-2}C\right) - A \odot (\bar{\mu}-\lambda)(\bar{\mu}-\lambda)'. \tag{3.16}$$

MRGARCH-t nests MRGARCH-N, as if the t degree of freedom $\zeta \to \infty$ then the conditional return distribution becomes Gaussian. The posterior inference of ζ in MRGARCH-t helps to explore the tail behaviour of return distribution when modelled along with RC.

Estimation

Estimation for MRGARCH-t is similar to MRGARCH-N. Parameters in vector $\theta = \{\mu, a, b, c, \lambda, \zeta, \nu\}$ and V have the following posterior density

$$p(\theta, V|\mathcal{I}_T) \propto p(\theta)p(V)\prod_{t=1}^{T}\mathrm{t}_n(r_t|\mu, H_t, \zeta)\,\mathrm{IW}_n\left(RC_t\middle|\nu, (\nu-n-1)\frac{\zeta}{\zeta-2}H_t^{1/2}V\left(H_t^{1/2}\right)'\right), \tag{3.17}$$

where $\mathrm{t}_n(.|)$ denotes the multivatriate-t density of dimension n. The above posterior does not have a known form. To draw from it, I initialize $\theta^{(0)}$ and $V^{(0)}$, and iterate, R times, the following

[3] Details in appendix A4

MCMC steps:

1. Get a new draw $\theta^{(i)}$ from the above posterior with a random walk MH algorithm with proposal $\theta^{'}$ which is from $h(\theta^{'}) \sim \mathrm{N}(\theta^{(i-1)}, \varsigma\hat{V}_h)$, with $\hat{V}_h$ the inverse Hessian matrix evaluated at the posterior mode $\hat{\theta}$. The draw is accepted with probability

$$\min\left\{p(\theta^{'}|\mathcal{I}_T)/p(\theta^{(i-1)}|\mathcal{I}_T), 1\right\}, \tag{3.18}$$

   where $p(.|\mathcal{I}_T)$ is the posterior in (3.17). For every $\theta^{(i)}$, $\Omega^{(i)}$ is calculated as in (3.16). Only posterior draws that result in positive definite $\Omega^{(i)}$ and $H_{1:T}^{(i)}$ are accepted.

2. Given prior $V \sim \mathrm{W}\left(\psi_0, \psi_0^{-1} I_n\right)$, draw $V^{(i)}|H_{1:T}^{(i)}, RC_{1:T}, \nu, \zeta$ with a Gibbs sampler as

$$\begin{aligned} p(V|RC_{1:T}, H_{1:T}, \nu) &\propto p(V)p(RC_{1:T}|V, H_{1:T}, \nu, \zeta) \\ &\sim \mathrm{W}(\nu_p, \Psi_p), \end{aligned} \tag{3.19}$$

   with $\nu_p = \psi_0 + T\nu$ and $\Psi_p = \left[(\psi_0^{-1} I_n)^{-1} + (\nu - n - 1)\frac{\zeta}{\zeta - 2}\sum_{t=1}^{T}\left(H_t^{1/2}\right)' RC_t^{-1} H_t^{1/2}\right]^{-1}$.

3.3.3 Semiparametric approach

Empirically returns have tails thicker than the normal distribution and are not symmetric. This means that the symmetric distributional assumption of parametric models, such as normal or Student-t, is too restrictive. Jensen and Maheu (2013) proposed an extension of parametric MGARCH models where the return distribution is nonparametrically modelled under an infinite mixture of normals with a Dirichlet Process (DP) prior of Ferguson (1973). Any underlying continuous multivariate distribution can be approximated without having to specify its unknown form.

Unlike parametric MGARCH models which assume that the covariance at time t, H_t, is completely determined by information of $t-1$, the model proposed by Jensen and Maheu (2013) has a two component covariance, a GARCH constructed one and a scale matrix from the DP prior.

In a similar way, I extend the MRGARCH-N model to a semiparametric version that combines the RC enhanced GARCH constructed covariance and the nonparametric approximation of the underlying return distribution that a DP mixture (**DPM**) offers.

MRGARCH-DPM: This novel semiparametric multivariate realized GARCH model, has the following hirerchical form

$$r_t|\mathcal{I}_{t-1}, H_t, m_t, L_t \sim \mathrm{N}_n\left(m_t, H_t^{1/2} L_t \left(H_t^{1/2}\right)'\right), \tag{3.20}$$

$$m_t, L_t|G \overset{iid}{\sim} G, \tag{3.21}$$

$$G|G_0, \alpha \sim \mathrm{DP}(\alpha, G_0), \tag{3.22}$$

$$G_0(m_t, L_t) \equiv \mathrm{N}_n(m_0, M_0) - \mathrm{IW}_n(\nu_0, (\nu_0 - n - 1)V_0), \quad \nu_0 > n + 1, \tag{3.23}$$

$$RC_t|\mathcal{I}_{t-1} \sim \mathrm{IW}_n\left(\nu, (\nu - n - 1)H_t^{1/2} V \left(H_t^{1/2}\right)'\right), \quad \nu > n + 1, \tag{3.24}$$

$$H_t = \Omega + A \odot (r_{t-1} - \lambda)(r_{t-1} - \lambda)' + B \odot H_{t-1} + C \odot RC_{t-1}. \tag{3.25}$$

Eq.(3.20)-(3.23) place an infinite mixture of multivariate normals on the return distribution. The mixing is on the mean vector of returns, m_t, and their covariance component, L_t. L_t is a positive definite matrix which scales H_t to yield a better approximation of the data density. The mixing moments m_t, L_t are distributed according to the latent G which is nonparametrically modeled from a DP prior. A draw from a DP, $G \sim \mathrm{DP}(\alpha, G_0)$, is almost surely a discrete distribution and has two parameters, the base measure G_0 and the precision parameter $\alpha > 0$. The DP is centred around G_0 since $\mathbb{E}[G] = G_0$ and the precision parameter α determines how close is G to G_0 since $\mathrm{Var}[G] = G_0[1 - G_0]/(\alpha + 1)$. In this case the base measure of DP in (3.23) is a normal-Inverse Wishart ($\mathrm{N} - \mathrm{IW}$) prior.

Sethuraman (1994) uses the stick-breaking representation for the infinite mixture of distributions in the DP. The MRGARCH-DPM model in (3.20)-(3.23) can be written as

$$p(r_t|\mathcal{I}_{t-1}, H_t, M, \Lambda, W) = \sum_{j=1}^{\infty} w_j \, \mathrm{N}_n\left(r_t \middle| \mu_j, H_t^{1/2} \Lambda_j \left(H_t^{1/2}\right)'\right), \tag{3.26}$$

where $W = \{w_1, w_2, ...\}$ is the infinite set of weights associated with the normal kernels, with $\sum_{j=1}^{\infty} w_j = 1$ and stick-breaking prior generated as

$$w_1 = v_1, \quad w_j = v_j \prod_{l=1}^{j-1}(1 - v_l), \quad j > 1, \tag{3.27}$$

$$v_j \overset{iid}{\sim} \mathrm{B}(1, \alpha). \tag{3.28}$$

where $\mathrm{B}(.)$ denotes the Beta distribution. Let $M = \{\mu_1, \mu_2, ...\}$ and $\Lambda = \{\Lambda_1, \Lambda_2, ...\}$ denote the unique points of support in G. A given dataset $r_{1:T} = \{r_1, ..., r_T\}$ will be associated with a finite set $\{(m_1, L_1), ..., (m_T, L_T)\}$ of draws from G in (3.21). By construction the DPM permits data to cluster under identical sets of (m_t, L_t). This allows pooling data into a finite number of k unique clusters, $\{(\mu_j, \Lambda_j)\}_{j=1}^{k}$, with $k < T$.

MRGARCH-DPM nests the parametric models MRGARCH-N and MRGARCH-t. The precision parameter α controls the number of clusters in the mixture. When $\alpha \to 0$ then $w_1 = 1$, $w_j = 0, \forall j > 1$, and returns follow a multivariate normal distribution with mean μ_1 and covariance scale Λ_1. When $\alpha \to \infty$, then $G \to G_0$ and if $\mu_j = \mu$, is constant $\forall j$, returns follow a multivariate Student-t distribution.

Covariance of returns is semiparametric and has two components, the parametric conditional H_t and the nonparametric Λ_j. The latter helps in capturing shocks that occur in day t. Typically, the base measure would be set so as $\mathbb{E}[\Lambda] = I_n$. If there are no shocks, then r_t would have $\Lambda_j = I_n$ and $\mathrm{Cov}(r_t|\mathcal{I}_{t-1}) = H_t$. H_t is the connection between r_t and RC_t. The conditional mean of RC_t is

$$\mathbb{E}[RC_t|\mathcal{I}_{t-1}] = H_t^{1/2} V \left(H_t^{1/2}\right)'. \tag{3.29}$$

If there is no market microstructure noise in RC_t, V would be approximately equal to an identity matrix and $\mathbb{E}[RC_t|\mathcal{I}_{t-1}] = H_t$. If there is no shock in day t and no noise in RC_t then $\mathrm{Cov}(r_t|\mathcal{I}_{t-1}) = \mathbb{E}[RC_t|\mathcal{I}_{t-1}] = H_t$. MRGARCH-DPM though, has the flexibility to capture

shocks observed in returns and noise in RC.

In H_t construction I assume covariance stationarity, $\mathbb{E}(H_t) = \bar{\Sigma}$, and set $V \equiv I_n$, to derive the covariance targeting [4] for Ω which is

$$\Omega = \bar{\Sigma} \odot (\iota\iota' - A - B - C) - A \odot (\bar{\mu} - \lambda)(\bar{\mu} - \lambda)'. \tag{3.30}$$

Estimation of MRGARCH-DPM

In order to make estimation feasible I use the stick-breaking formulation and the slice sampler by Walker (2007) and Kalli et al. (2011), to truncate the infinite normals into a finite number k, $k < T$, with the associated unique normal clusters $\{(\mu_j, \Lambda_j)\}_{j=1}^{k}$. To do so, the parameters are expanded by the introduction of two latent vectors. The first one is a cluster or state indicator $s_{1:T} = \{s_1, ...s_T\}$ which maps each observation r_t to a normal cluster j. The second auxiliary vector $u_{1:T} = \{u_1, ...u_T\}$, with $u_t \in (0, 1)$, helps to convert the infinite sum into a finite mixture. u_t is defined such that the joint density of (r_t, u_t) is

$$f(r_t, u_t|\mathcal{I}_{t-1}, M, \Lambda, W) = \sum_{j=1}^{\infty} \mathbf{1}\{u_t < w_j\} \ \mathrm{N}_n\left(r_t \middle| \mu_j, H_t^{1/2}\Lambda_j\left(H_t^{1/2}\right)'\right), \tag{3.31}$$

where $\mathbf{1}\{.\}$ is an indicator function. Slice variable u_t makes the set $\{u_t < w_j\}_{j=1}^{\infty}$ finite. Integrating (3.31) over u_t would give the desired density of r_t. Conditional on $H_{1:T}$, the full likelihood of returns is

$$p(r_{1:T}, s_{1:T}, u_{1:T}|H_{1:T}, \mu_{1:k}, \Lambda_{1:k}, w_{1:k}) = \prod_{t=1}^{T} \mathbf{1}\{u_t < w_{s_t}\} \ \mathrm{N}_n\left(r_t \middle| \mu_{s_t}, H_t^{1/2}\Lambda_{s_t}\left(H_t^{1/2}\right)'\right). \tag{3.32}$$

[4] Details in appendix A4

The posterior density of MRGARCH-DPM parameters is proportional to

$$p(\theta)p(w_{1:k})p(u_{1:T})p(s_{1:T})p(\alpha)p(V)\prod_{j=1}^{k}p(\mu_j)p(\Lambda_j)$$
$$\times\prod_{t=1}^{T}\mathbf{1}\left\{u_t<w_{s_t}\right\}\ \mathrm{N}_n\left(r_t\middle|\mu_{s_t},H_t^{1/2}\Lambda_{s_t}\left(H_t^{1/2}\right)'\right)\ \mathrm{IW}_n\left(RC_t\middle|\nu,(\nu-n-1)H_t^{1/2}V\left(H_t^{1/2}\right)'\right), \tag{3.33}$$

with $\theta=\{a,b,c,\lambda,\nu\}$ and k being the smallest positive integer that satisfies the condition $\sum_{j=1}^{k}w_j>1-\min(u_{1:T})$. The posterior does not have a known form. To sample from it, I initialize $\theta,k,w_{1:k},s_{1:T},\mu_{1:k},\Lambda_{1:k},V,\alpha$ and take posterior draws through the following conditional distributions:

1. Sample $\mu_{1:k},\Lambda_{1:k}|r_{1:T},H_{1:T},s_{1:T}$.
2. Update $w_{1:k}|s_{1:T},\alpha$ with a stick-breaking process.
3. Sample the slice vector $u_{1:T}|w_{1:k},s_{1:T}$.
4. Update k as the smallest integer that satisfies: $\sum_{j=1}^{k}w_j>1-\min(u_{1:T})$.
5. Sample $s_{1:T}|r_{1:T},H_{1:T},\mu_{1:k},\Lambda_{1:k},w_{1:k},\alpha$.
6. Sample $\alpha|T,\kappa.\ \kappa$
7. Sample $\theta|r_{1:T},RC_{1:T},V,\mu_{1:k},\Lambda_{1:k},w_{1:k},s_{1:T}$.
8. Sample $V|RC_{1:T},H_{1:T},\nu$.

Details are provided in A3. Repeating the above steps R times, with R_0 burnin sweeps, I get the posterior draws for inference.

3.3.4 Restricted MRGARCH-DPM

I also develop a restricted version of MRGARCH-DPM, in which the mean vector of returns is constant and not included in the mixture.

MRGARCH-DPM-Λ: This semiparametric MRGARCH model has the following hierarchical form

$$r_t|\mathcal{I}_{t-1}, H_t, L_t \sim \mathrm{N}_n\left(\mu, H_t^{1/2} L_t \left(H_t^{1/2}\right)'\right), \tag{3.34}$$

$$L_t|G \overset{iid}{\sim} G, \tag{3.35}$$

$$G|G_0, \alpha \sim \mathrm{DP}(\alpha, G_0), \tag{3.36}$$

$$G_0(L_t) \equiv \mathrm{IW}_n(\nu_0, (\nu_0 - n - 1)V_0), \quad \nu_0 > n+1 \tag{3.37}$$

$$RC_t|\mathcal{I}_{t-1} \sim \mathrm{IW}_n\left(\nu, (\nu - n - 1)H_t^{1/2} V \left(H_t^{1/2}\right)'\right), \quad \nu > n+1, \tag{3.38}$$

$$H_t = \Omega + A \odot (r_{t-1} - \lambda)(r_{t-1} - \lambda)' + B \odot H_{t-1} + C \odot RC_{t-1}. \tag{3.39}$$

The model has the following Sethuraman (1994) representation

$$p(r_t|\mathcal{I}_{t-1}, H_t, \Lambda, W) = \sum_{j=1}^{\infty} w_j \, \mathrm{N}_n\left(r_t \middle| \mu, H_t^{1/2}\Lambda_j \left(H_t^{1/2}\right)'\right). \tag{3.40}$$

This model captures the thick tails of the return distribution and hence nests both the proposed parametric specifications. When $\alpha \to 0$ then $w_1 = 1$, $w_j = 0, \forall j > 1$, MRGARCH-DPM-Λ is equivalent to MRGARCH-N, returns follow a multivariate normal distribution with mean μ and covariance scale Λ_1. When $\alpha \to \infty$, then $G \to G_0$, returns follow a multivariate Student-t distribution and MRGARCH-DPM-Λ is equivalent to MRGARCH-t.

MRGARCH-DPM-Λ estimation is the same as for MRGARCH-DPM described in Section 3.3.3 with a constant vector μ which is estimated in the MH step 7, in which the parameter block is $\theta = \{\mu, a, b, c, \lambda, \nu\}$ and they have the following posterior

$$\begin{aligned} p(\theta|\mathcal{I}_T, \Lambda_{1:k}, s_{1:k}, V) \propto p(\theta) \prod_{t=1}^{T} & \mathrm{N}_n\left(r_t \middle| \mu, H_t^{1/2}\Lambda_{s_t}\left(H_t^{1/2}\right)'\right) \\ & \times \mathrm{IW}_n\left(RC_t \middle| \nu, (\nu - n - 1)H_t^{1/2} V \left(H_t^{1/2}\right)'\right). \end{aligned} \tag{3.41}$$

Approximation of predictive density with MRGARCH-DPM-Λ is done as in MRGARCH-DPM, described in the following Section 3.3.5, with the use of draws $\{\mu^{(i)}\}_{i=1}^{R}$ from the above posterior in (3.41).

3.3.5 Predictive density

For the parametric models, MRGARCH-N and MRGARCH-t, their posterior draws $\{\theta^{(i)}\}_{i=1}^{R}$ can be used to approximate the predictive return density as

$$p(r_{t+1}|\mathcal{I}_t) \approx \frac{1}{R}\sum_{i=1}^{R} f_n\left(r_{t+1}|\mathcal{I}_t, \theta^{(i)}\right), \tag{3.42}$$

with $f_n(.|)$ being the n dimension multivariate p.d.f. (normal or St-t) associated with each model's return density assumption.

The key task when forecasting with DPM models is to integrate out the uncertainty about the future state of the modelled series. Conditional on $\mathcal{I}_t = \{r_{1:t}, RC_{1:t}\}$, H_{t+1} can be calculated from (3.25) and the predictive density of MRGARCH-DPM can be approximated as

$$p(r_{t+1}|\mathcal{I}_t, \mu_{1:k}, \Lambda_{1:k}, s_{1:k}, w_{1:k}) \approx \frac{1}{R}\sum_{i=1}^{R} \mathrm{N}_n\left(r_{t+1}\middle|\mu^{(i)}_{s^{(i)}_{t+1}}, H^{1/2(i)}_{t+1}\Lambda^{(i)}_{s^{(i)}_{t+1}}\left(H^{1/2(i)}_{t+1}\right)'\right), \tag{3.43}$$

$$\text{where} \quad s^{(i)}_{t+1} = \begin{cases} j, & \text{if} \quad \sum_{l=0}^{j-1} w^{(i)}_l < \phi < \sum_{l=0}^{j} w^{(i)}_l, \\ k^{(i)}+1, & \text{if} \quad \phi \geq \sum_{l=0}^{k^{(i)}} w^{(i)}_l, \end{cases} \tag{3.44}$$

with $w^{(i)}_o = 0$, $j \leq k^{(i)}$ and $\phi \sim \mathrm{U}(0,1)$. The above means that the future value of $s^{(i)}_{t+1}$ is one of the existing clusters with probability equal to the associated weights and there is a nonzero probability of introducing a new normal cluster $\left(\mu^{(i)}_{k^{(i)}+1}, \Lambda^{(i)}_{k^{(i)}+1}\right)$ from the base measure G_0. The predictive return covariance integrates out the uncertainty about the covariance of returns

for the next period by using posterior draws as

$$\begin{aligned}\text{Cov}(r_{t+1}|\mathcal{I}_t) &= \mathbb{E}[r_{t+1}(r_{t+1})'|\mathcal{I}_t] - \mathbb{E}(r_{t+1}|\mathcal{I}_t)[\mathbb{E}(r_{t+1}|\mathcal{I}_t)]' \\ &\approx \frac{1}{R}\sum_{i=1}^{R}\left[H_{t+1}^{1/2(i)}\Lambda_{s_{t+1}}^{(i)}\left(H_{t+1}^{1/2(i)}\right)' + \mu_{s_{t+1}}^{(i)}\left(\mu_{s_{t+1}}^{(i)}\right)'\right] - \left(\frac{1}{R}\sum_{i=1}^{R}\mu_{s_{t+1}}^{(i)}\right)\left(\frac{1}{R}\sum_{i=1}^{R}\mu_{s_{t+1}}^{(i)}\right)'.\end{aligned} \tag{3.45}$$

For the proposed models, conditional on H_{t+1}, the predictive density of RC_{t+1} can be approximated as

$$p(RC_{t+1}|\mathcal{I}_t) \approx \frac{1}{R}\sum_{i=1}^{R}\text{IW}_n\left(RC_{t+1}|\mathcal{I}_t, H_{t+1}^{(i)}\right), \tag{3.46}$$

where $\text{IW}_n(.|)$ is the Inverse Wishart p.d.f with degree of freedom $\nu^{(i)}$ and scale matrix based on each model's specification.

The predictive density is the key measure to calculate the value of predictive likelihood (PL) and evaluate each model's forecasting ability out-of-sample. Specifically, for each model, by performing a set of τ (with $1 < \tau < T$), recursive posterior estimations, the log predictive likelihood of $r_{T-\tau+1:T}$, from each model, is the summation of individual log predictive densities of each estimation

$$\text{logPL}_r(r_{T-\tau+1:T}|\mathcal{I}_T) = \sum_{t=T-\tau}^{T-1}\log\left(p(r_{t+1}|\mathcal{I}_t)\right). \tag{3.47}$$

Similarly, the log predictive likelihood of $RC_{T-\tau+1:T}$ is calculated as

$$\text{logPL}_{RC}(RC_{T-\tau+1:T}|\mathcal{I}_T) = \sum_{t=T-\tau}^{T-1}\log\left(p(RC_{t+1}|\mathcal{I}_t)\right). \tag{3.48}$$

The models' performance on forecasting the covariance of returns can be measured by the root

mean squared forecast error (RMSFE) of the matrix (Frobenius) norm difference between out-of-sample RC_{t+1} and the forecasted covariance of returns

$$\text{RMSFE}_{RC} = \frac{1}{T-\tau+1}\sum_{t=T-\tau}^{T-1}||RC_{t+1} - \text{Cov}(r_{t+1}|\mathcal{I}_t)||, \tag{3.49}$$

where $\text{Cov}(r_{t+1}|\mathcal{I}_t)$ is the estimated predictive covariance matrix based on each model's return distributional assumption, $\text{Cov}(r_{t+1}|\mathcal{I}_t) \equiv \mathbb{E}(H_{t+1})$ for multivariate normal, $\text{Cov}(r_{t+1}|\mathcal{I}_t) \equiv \mathbb{E}(\frac{\zeta}{\zeta-2}H_{t+1})$ for multivariate-t or DPM approximation from (3.45).

It is often of interest to forecast portfolio returns and multivariate GARCH models are typically used for these applications. For a given set of portfolio weights ω, with $\iota'\omega = 1$, portfolio returns are $r^p_{t+1} = \omega' r_{t+1}$, with $r^p_{t+1} \in \mathbb{R}$ and $r_{t+1} \in \mathbb{R}^n$. The conditional portfolio return density can be approximated as

$$p(r^p_{t+1}|\mathcal{I}_t, \omega) \approx \frac{1}{R}\sum_{i=1}^{R} f_1\left(r^p_{t+1}|\mathcal{I}_t, H^{(i)}_{t+1}, \omega\right), \tag{3.50}$$

with $f_1(.|)$ being a univariate p.d.f. based on each model's return distributional assumption: $\text{N}\left(.|\omega'\mu^{(i)}, \omega' H^{(i)}_{t+1}\omega\right)$, $\text{t}\left(.|\omega'\mu^{(i)}, \omega' H^{(i)}_{t+1}\omega, \zeta^{(i)}\right)$ or $\text{N}\left(.\middle|\omega'\mu^{(i)}_{s^{(i)}_{t+1}}, \omega' H^{1/2(i)}_{t+1}\Lambda^{(i)}_{s^{(i)}_{t+1}}\left(H^{1/2(i)}_{t+1}\right)'\omega\right)$ for DPM models. Using the predictive density, the forecasting ability of each model can be evaluated, for a set of τ recursive posterior estimations, with the portfolio log predictive likelihood for $r^p_{T-\tau+1:T}$ is calculated as

$$\text{logPL}_p(r^p_{T-\tau+1:T}|\mathcal{I}_T) = \sum_{t=T-\tau}^{T-1}\log\left(p(r^p_{t+1}|\mathcal{I}_t, \omega)\right). \tag{3.51}$$

The proposed MRGARCH models due to their flexibility can be used to forecast certain areas of the portfolio predictive density such as the tails (Diks et al. (2011)). For a value η, with $\eta \in \mathbb{R}$,

the portfolio predictive density of $r^p_{t+1} < \eta$ is defined as

$$p(r^p_{t+1}|r^p_{t+1} < \eta, \mathcal{I}_t) = \frac{p(r^p_{t+1}|\mathcal{I}_t)\mathbf{1}\left\{r^p_{t+1} < \eta\right\}}{\int_{-\infty}^{\eta} p(r^p_{t+1}|\mathcal{I}_t)dr^p_{t+1}}$$

$$\approx \frac{1}{R}\sum_{i=1}^{R} \frac{f_1\left(r^p_{t+1}|\mathcal{I}_t\right)\mathbf{1}\left\{r^p_{t+1} < \eta\right\}}{\Phi\left(\left(\eta - \mu^{p(i)}_{t+1|t}\right)/\sigma^{p(i)}_{t+1|t}\right)}, \tag{3.52}$$

with $\mu^p_{t+1|t} \equiv \mathbb{E}\left(r^p_{t+1}|\mathcal{I}_t\right)$ and $\mu^{p(i)}_{t+1|t} = \omega'\mu^{(i)}$(normal, St-t and DPM-$\Lambda$), or $\omega'\mu^{(i)}_{s^{(i)}_{t+1}}$ (DPM), $\sigma^{2p}_{t+1|t} \equiv \text{Var}\left(r^p_{t+1}|\mathcal{I}_t\right)$ and $\sigma^{2p(i)}_{t+1|t} = \omega' H^{(i)}_{t+1}\omega$ (normal), $\omega'\frac{\zeta^{(i)}}{\zeta^{(i)}-2}H^{(i)}_{t+1}\omega$ (Student-t) or $\omega' H^{1/2(i)}_{t+1}\Lambda^{(i)}_{s^{(i)}_{t+1}}\left(H^{1/2(i)}_{t+1}\right)'\omega$ (DPM). $\Phi(.)$ denotes the associated c.d.f. of univariate standard distribution of each model's assumption: normal or St-t, with $\zeta^{(i)}$ degrees of freedom. The denominator in (3.52) is an integration constant ensuring that the predictive density integrates to one.

3.3.6 Benchmark models

The proposed MRGARCH models are compared with the following MGARCH models that do not include RC in their specification.

MGARCH-N: This parametric model has a Gaussian assumption for the multivariate return distribution and includes an asymmetric leverage effect in covariance. It is defined as

$$r_t|r_{1:t-1} \sim \text{N}_n(\mu, H_t), \tag{3.53}$$

$$H_t = \Omega + a(a') \odot (r_{t-1} - \lambda)(r_{t-1} - \lambda)' + b(b') \odot H_{t-1}. \tag{3.54}$$

MGARCH-t: This parametric model has the multivariate-t assumption for the return distribution and can capture the extreme tails that empirically are observed in financial returns. It is an extension of the model proposed by Diamantopoulos and Vrontos (2010) and the benchmark of Jensen and Maheu (2013), to include asymmetric leverage effect. This specification is also

used as a benchmark from Maheu and Shamsi Zamenjani (2021) and is defined as

$$r_t|r_{1:t-1} \sim \mathrm{t}_n(\mu, H_t, \zeta), \quad \zeta > 2, \tag{3.55}$$

$$H_t = \Omega + a(a') \odot (r_{t-1} - \lambda)(r_{t-1} - \lambda)' + b(b') \odot H_{t-1}. \tag{3.56}$$

MGARCH-DPM: This semiparametric MGARCH model, developed by Jensen and Maheu (2013), has no underlying distributional assumption for returns but instead, it uses an infinite mixture of normals to approximate their unknown density. This gives the flexibility to capture the asymmetry and the extreme tail behaviour observed in returns. The specification used here is proposed by Maheu and Shamsi Zamenjani (2021) and is defined as

$$r_t|r_{1:t-1}, H_t, m_t, L_t \sim \mathrm{N}_n\left(m_t, H_t^{1/2} L_t \left(H_t^{1/2}\right)'\right), \tag{3.57}$$

$$m_t, L_t|G \overset{iid}{\sim} G, \tag{3.58}$$

$$G|G_0, \alpha \sim \mathrm{DP}(\alpha, G_0), \tag{3.59}$$

$$G_0(m_t, L_t) \equiv \mathrm{N}_n(m_0, M_0) - \mathrm{IW}_n(\nu_0, (\nu_0 - n - 1)V_0), \quad \nu_0 > n + 1 \tag{3.60}$$

$$H_t = \Omega + a(a') \odot (r_{t-1} - \lambda)(r_{t-1} - \lambda)' + b(b') \odot H_{t-1}. \tag{3.61}$$

MGARCH-DPM-Λ: This semiparametric model, developed by Jensen and Maheu (2013), is a restricted version of MGARCH-DPM, with a constant return mean vector. The specification used here is the following

$$r_t|r_{1:t-1}, H_t, m_t, L_t \sim \mathrm{N}_n\left(\mu, H_t^{1/2} L_t \left(H_t^{1/2}\right)'\right), \tag{3.62}$$

$$L_t|G \overset{iid}{\sim} G, \tag{3.63}$$

$$G|G_0, \alpha \sim \mathrm{DP}(\alpha, G_0), \tag{3.64}$$

$$G_0(L_t) \equiv \mathrm{IW}_n(\nu_0, (\nu_0 - n - 1)V_0), \quad \nu_0 > n + 1, \tag{3.65}$$

$$H_t = \Omega + a(a') \odot (r_{t-1} - \lambda)(r_{t-1} - \lambda)' + b(b') \odot H_{t-1}. \tag{3.66}$$

3.4 Empirical applications

3.4.1 Selection of priors

For MRGARCH-N, the mean vector of returns μ is from an independent normal prior $\mu \sim \mathrm{N}_n(0, 0.1 \times I_n)$, with I_n being the identity matrix of dimension n. For the measurement equation, the IW degree of freedom parameter has an exponential prior $\nu \sim \mathrm{Exp}(n+10)$, with $\mathbb{E}(\nu) = n + 10$. Matrix V has a Wishart prior, $V \sim \mathrm{W}_n(n+10, (n+10)^{-1} \times I_n)$ with $\mathbb{E}(V) = I_n$. MRGARCH parameters $\{a, b, c, \lambda\}$ have an independent normal prior, $\mathrm{N}_n(0, 100 \times I_n)$. To ensure model identification, draws which do not satisfy $a_1 \geq 0$, $b_1 \geq 0$, $c_1 \geq 0$ and positive definite Ω and $H_{1:T}$ are rejected. For MRGARCH-t are used the above priors for the parameters. The Student-t degree of freedom has a uniform prior, $\zeta \sim \mathrm{U}(2, 100)$.

For the DPM models, the parameters in measurement equation and MRGARCH construction are the same as in MRGARCH-N. The base measure of DPM, G_0, for the return mean is set around zero, $\mathbb{E}(\mu) = 0$, with $m_0 = 0$ and $M_0 = 0.1 \times I_n$. For DPM-Λ, $\mu \sim \mathrm{N}_n(0, 0.1 \times I_n)$. The scaling covariance matrices Λ are centred around the identity matrix, $\mathbb{E}(\Lambda) = I_n$, by setting $V_0 = (\nu_0 - n - 1)I_n$ making $\Lambda \sim \mathrm{IW}_n(\nu_0, (\nu_0 - n - 1)I_n)$. I use $\nu_0 = n + 10$, but other values of it do not affect the results. The precision parameter of DPM has a Gamma prior, $\alpha \sim \Gamma(2, 8)$, following Jensen and Maheu (2013).

3.4.2 Posterior estimation results

Results from 10,000 MCMC posterior draws, after 5,000 burnin, for IBM-MSFT-XOM are in Tables 3.3. Similar results are observed for the AXP-DD-GE-KO dataset.

Incorporating RC information, results in a change of all GARCH parameter values. RC_{t-1} is highly significant in the conditional covariance construction with c parameters ranging in 0.443-0.478 for MRGARCH-t and 0.517-0.529 for the rest MRGARCH models. The other structural parameters, a and b significantly differ between the benchmarks MGARCH and MRGARCH

that include RC. The values of b range from 0.964 to 0.976 for MGARCH and drop to 0.825-0.833 in MRGARCH models. Parameters in a for MGARCH have values from 0.184 to 0.235 and for MRGARCH decline to 0.1-0.145. Parameters in λ for MGARCH have values from 0.26 to 0.42. For MRGARCH models they rise to 0.6-0.92. This trade-off between a and λ captures the information from return shocks to construct future covariance. Parameters in Ω increase in MRGARCH models compared to their associated benchmarks. This is the effect of RC in the long-run unconditional covariance.

Regarding the shape of the multivariate return distribution, MGARCH-t indicates thick tails, with degree of freedom parameter ζ being on average 9.113 for IBM-MSFT-XOM. Including RC information, in the MRGARCH-t model, makes the tails thicker with the average ζ equal to 7.18 and helps to estimate it more consistently, with a smaller 95% density interval range.

The DPM models show that a finite mixture of normals is required to approximate the underlying return distribution. Without RC, MGARCH-DPM uses on average 4.96 normal clusters to approximate return density while with RC, MRGARCH-DPM uses 5.58 clusters. DPM-Λ models, which have a mixture only in covariance, use more clusters, on average 9.7-10.7. Figure 3.1 displays the daily out-of-sample predictive density of equal weight portfolio (EWP) returns constructed by the data. For both datasets the proposed models indicate thicker tails than the Gaussian model. MRGARCH-DPM shows minor evidence of asymmetry in returns.

In the measurement equation, the IW degree of freedom has a consistent value around 15.7, across the models. The matrix V, which scales the conditional mean of $RC_{1:T}$, is close to the prior mean, the identity matrix. This suggests that RC is near to an unbiased estimation of the covariance of returns, as constructed by the MRGARCH models.

Figure 3.2, displays the log determinant of the conditional covariance matrices produced from the full sample posterior estimations. Models with constant return mean μ are displayed. The proposed models produce covariance matrices with determinant closer to the RC. The MRGARCH-DPM-Λ is the most flexible model in producing covariance matrices that follow

the dynamics of RC. This is validated from the forecasting results when measuring the RMSFE of RC.

3.4.3 Forecasting results

The results from recursively forecasting out-of-sample returns and RC are presented in Table 3.4. Model performance is measured by RMSFE for the mean and logPL for the density forecasting. When comparing two models, logPL difference of 5 and above indicates strong evidence of better density forecasting performance. Also, this comparison favours parsimonious specifications unless complex models provide an improved data description.

Comparing the two groups, MGARCH and MRGARCH forecast results displayed in Table 3.4, the MRGARCH models have lower RMSFE of RC than the MGARCH ones. This means that MRGARCH models produce conditional covariance matrices closer to RC. Because of that they have better performance in forecasting the return density, as indicated by their larger log predictive likelihood values.

Interestingly, the models that accommodate thick tails outperform the Gaussian ones for both datasets. Parametric MRGARCH-t gives the best return density forecasting along with MRGARCH-DPM-Λ. Noteworthy, the MRGARCH models with DPM help to construct the closest conditional covariance matrices to RC. This supports the indications from Figure 3.2. The DPM models also perform the best at forecasting the RC density. All the models perform similarly in forecasting the mean of returns, as indicated by their close RMSFE values.

These results contribute to the existing literature of MRGARCH models and are of high importance as they indicate that the Gaussian assumption is too restrictive. modeling the distribution of returns with the nonparametric DPM not only significantly improves density forecasting, compared to a Gaussian assumption, but also improves predictive covariance forecasts.

3.4.4 Portfolio applications

The models here are tested on their out-of-sample density forecasting as explained in Section 3.3.5. The key result from Table 3.5 is that MRGARCH models outperform the MGARCH ones in forecasting the portfolio density. However, among the MRGARCH models none strongly stands out. For IBM-MSFT-XOM, MRGARCH-DPM-Λ and MRGARCH-t produce marginally better logPL than MRGARCH-N. For AXP-DD-GE-KO, MRGARCH-N, MRGARCH-DPM-Λ and MRGARCH-DPM have marginally larger logPL than MRGARCH-t. Forecasting the tail density when portfolio returns are less than -1.0%, MRGARCH-N has the largest logPL value, followed closely by the rest of the MRGARCH models.

The effect of improved conditional covariance construction and the impact of non-Gaussian innovations is further explored with portfolio applications that put emphasis on risk management, tail events and construction of minimum variance portfolios.

Portfolio tail risk measures

The proposed MRGARCH models, due to their flexibility and their emphasis on the tails of the return distribution, can be more accurate in calculating risk measures of high importance for the areas of investments and risk management. The models are tested on their ability to forecast the tail risk measures of Value-at-Risk (VaR) and Expected Shortfall (ES).

VaR is a quantile measure based on a confidence level ϵ. Given an investment time interval (day, week, etc), VaR is the least potential loss that could happen to an investment $\epsilon\%$ of the time intervals. For the simple case of one period portfolio forecasting, $\text{VaR}^{\epsilon}_{t+1}$ is defined as

$$\text{Prob}\left[r^{p}_{t+1} \leq \text{VaR}^{\epsilon}_{t+1} | \mathcal{I}_t\right] = \epsilon \tag{3.67}$$

where r^{p}_{t+1} is the portfolio return realization $r^{p}_{t+1} = \omega' r_{t+1}$, with $r^{p}_{t+1} \in \mathbb{R}$, $r_{t+1} \in \mathbb{R}^n$ and $\iota'\omega = 1$. VaR does not consider the magnitude of the potential loss. This is better accessed by

ES which is the measure of expected loss conditional on exceeding the VaR value. The $\text{ES}^{\epsilon}_{t+1}$ is

$$\text{ES}^{\epsilon}_{t+1} = \mathbb{E}\left[r^{p}_{t+1} | r^{p}_{t+1} \leq \text{VaR}^{\epsilon}_{t+1}, \mathcal{I}_t\right] \tag{3.68}$$

To explore the impact of RC information on forecasting VaR and ES, I consider the EWP (equal weight portfolio) of each dataset and perform the following steps for the models in test:

1. For each of the τ forecasting periods, simulate R values of $r_{t+1}|\mathcal{I}_t$:
   (a) Given the posterior draws $\{\theta^{(i)}\}_{i=1}^{R}$, construct $H^{(i)}_{t+1}$ according to each model's specification.
   (b) Simulate $r^{(i)}_{t+1}|H^{(i)}_{t+1}$ from each model's return distributional assumption, multivariate normal or Student-t. For the DPM, $r^{(i)}_{t+1}|H^{(i)}_{t+1}, s^{(i)}_{t+1}$ comes from a multivariate normal draw with $s^{(i)}_{t+1}$ drawn as in (3.44).
2. For each vector in $\{r^{(i)}_{t+1}\}_{i=1}^{R}$, compute the portfolio returns $r^{p(i)}_{t+1} = \omega' r^{(i)}_{t+1}$ with $\omega = (1/n, ..., 1/n)'$.
3. The $\lfloor \epsilon R \rfloor$-th least value of $\left\{r^{p(i)}_{t+1}\right\}_{i=1}^{R}$ is the $\text{VaR}^{\epsilon}_{t+1}$.
4. $\text{ES}^{\epsilon}_{t+1} = \dfrac{\sum_{i=1}^{R} r^{p}_{t+1} \mathbf{1}\left\{r^{p(i)}_{t+1} \leq \text{VaR}^{\epsilon}_{t+1}\right\}}{\sum_{i=1}^{R} \mathbf{1}\left\{r^{p(i)}_{t+1} \leq \text{VaR}^{\epsilon}_{t+1}\right\}}$.

I calculate the VaR and ES for the confidence levels: $\epsilon = 0.01, 0.05$ and 0.1. Figures 3.3 and 3.4 display the forecasted 1% ($\epsilon = 0.01$) VaR and ES from the models with a parametric return mean vector. In the figures is noticeable the effect of RC. There are two obvious trends in the forecasted measures which are with and without RC. Models with RC produce more flexible forecasts especially in the volatile period from Oct 2008 to Jan 2009. As expected, the models with t-distributed innovations, produce forecasts larger in magnitude than the Gaussian ones. DPM models are more flexible, their forecasts are mostly between the two parametric ones

for the more volatile portfolio of AXP-DD-GE-KO while for IBM-MSFT-XOM, which is less volatile, DPM gives forecasts close to the Gaussian one.

To evaluate the forecasts of VaR is typically used the Violation Rate (VR), as in Chen et al. (2012) for instance. VR measures the frequency that realized returns are less than the forecasted VaR, for the whole forecasting period and is calculated as

$$VR^{\epsilon} = \frac{1}{\epsilon\tau} \sum_{t=T-\tau}^{T-1} \mathbf{1}\left\{r_{t+1}^{p} \leq \text{VaR}_{t+1}^{\epsilon}\right\}. \tag{3.69}$$

VR^{ϵ} is standardized by the rate ϵ so as the models that give VR closer to one are those which perform the best in forecasting VaR.

VR results are displayed in Table 3.6. The benchmark MGARCH-DPM-Λ has the best VR for $\text{VaR}^{0.01}$ of IBM-MSFT-XOM and for $\text{VaR}^{0.05}$ of AXP-DD-GE-KO. For the rest, MRGARCH models have the best rates. MRGARCH model with DPM have consistently rates close to one for every dataset and VaR level.

Evaluating forecasts of ES is not as straightforward as for VaR. Fissler and Ziegel (2016) propose a class of loss functions that are minimized with the joint true VaR and ES. The benefit of these functions is that there is no need for a distributional assumption, unlike other ES evaluations such as the ones used from Chen et al. (2012) and Gerlach and Chen (2015). I use the function which is derived by Patton et al. (2019), from the aforementioned class, in order to simultaneously evaluate the forecasts of VaR and ES from the models in test. For the realized returns and the forecasted VaR and ES, the loss function is defined as

$$\mathcal{L}(r_{t+1}^{p}, \text{VaR}_{t+1}^{\epsilon}, \text{ES}_{t+1}^{\epsilon}, \epsilon) = -\frac{1}{\epsilon \text{ES}_{t+1}^{\epsilon}} \mathbf{1}\left\{r_{t+1}^{p} \leq \text{VaR}_{t+1}^{\epsilon}\right\}\left(\text{VaR}_{t+1}^{\epsilon} - r_{t+1}^{p}\right) + \frac{\text{VaR}_{t+1}^{\epsilon}}{\text{ES}_{t+1}^{\epsilon}} + \log\left(-\text{ES}_{t+1}^{\epsilon}\right) - 1, \tag{3.70}$$

with the true VaR and ES being the solution of minimizing it,

$$\left(\text{VaR}_{t+1}^{\epsilon}, \text{ES}_{t+1}^{\epsilon}\right) = \arg\min \mathbb{E}\left[\mathcal{L}(r_{t+1}^{p}, \text{VaR}_{t+1}^{\epsilon}, \text{ES}_{t+1}^{\epsilon}, \epsilon)|\mathcal{I}_{t}\right]. \tag{3.71}$$

Thus, from the considered models, the ones that would give the least average loss function values would produce the best forecasts of VaR and ES.

In Table 3.7 are displayed the average out-of-sample losses from (3.70) for the forecasted measures of VaR and ES. MRGARCH models have consistently smaller values than the benchmarks. This is a clear indication that RC improves the forecasting of the risk measures. MRGARCH-N gives the least losses when $\epsilon = 0.01$ for IBM-MSFT-XOM and when $\epsilon = 0.1$ for AXP-DD-GE-KO. MRGARCH-DPM-Λ is the best when $\epsilon = 0.1$ for IBM-MSFT-XOM and for the rest, measures from MRGARCH-DPM give the least losses. Combined with the results from Table 3.6, the last two models, MRGARCH-DPM-Λ and MRGARCH-DPM, that offer the most flexible conditional return covariance, are the ones that consistently produce the best or the second best forecasts about the risk measures of VaR and ES.

Global minimum variance portfolio

In this Section the models are tested on portfolio optimization. I focus on the Global Minimum Variance portfolio (GMVP) problem in which the conditional covariance is the target of the optimization. As addressed by Engle and Colacito (2006) the GMVP weights that are produced from the true conditional covariance give the lowest portfolio variance compared to any other model GMVP weights.

Given the information set $\mathcal{I}_t$, available at time t, for n risky assets, investors, who follow a dynamic GMVP investment strategy, are interested in accurate forecasts of the covariance matrix, $\text{Cov}(r_{t+1}|\mathcal{I}_t)$. Based on that, they can adjust their GMVP weights for time $t+1$ by solving the following problem

$$\min_{\omega_{t+1|t}\in\mathbb{R}^n} \left\{ \sigma^{2p}_{t+1|t} = (\omega_{t+1|t})' \text{Cov}(r_{t+1}|\mathcal{I}_t)\omega_{t+1|t} \right\} \tag{3.72}$$

$$\text{s.t.} \quad \iota'\omega_{t+1|t} = 1. \tag{3.73}$$

where $\omega_{t+1|t}$ is the portfolio weights vector constructed for period $t+1$ given information $\mathcal{I}_t$. The solution to the problem is

$$\omega_{t+1|t}^{\text{GMV}} = \frac{\text{Cov}(r_{t+1}|\mathcal{I}_t)^{-1}\iota}{(\iota)'\text{Cov}(r_{t+1}|\mathcal{I}_t)\iota}. \tag{3.74}$$

At the end of the day $t+1$, they realize their return from their position and the information set is expanded to $\mathcal{I}_{t+1}$ so they can readjust portfolio weights for the next day and so on. The estimation of $\text{Cov}(r_{t+1}|\mathcal{I}_t)$ from each MRGARCH model is described in Section 3.3.5.

Models are tested on their GMVP performance out-of-sample. For a given number of days τ (with $1 < \tau < T$), the sample portfolio returns variance, $\overline{\sigma}^{2p}$, of the realized portfolio returns $\{r_t^p\}_{t=T-\tau+1}^{T}$, from each model is calculated as

$$\overline{\sigma}^{2p} = \frac{1}{\tau}\sum_{t=T-\tau+1}^{T}\left(r_t^p - \bar{r}^p\right)^2.$$

where $\bar{r}^p$ is the mean of $\{r_t^p\}_{t=T-\tau+1}^{T}$.

Table 3.8 displays the out-of-sample results for the GMVPs constructed from the models in test. There are two key takeaways from these results. First, MRGARCH models consistently produce lower variance out-of-sample GMVPs than the MGARCH benchmarks. This validates the importance of RC signal for the conditional covariance construction. Second, for the MRGARCH models, Gaussian assumption is restrictive. The least volatile GMVPs are from MRGARCH-DPM(-Λ) for IBM-MSFT-XOM and from MRGARCH-t for AXP-DD-GE-KO. This shows that both RC information and modeling tail behaviour are required in order to achieve the least volatile portfolios out-of-sample and hence to get closer to the true conditional covariance, as Engle and Colacito (2006) have shown.

3.5 Concluding remarks

This study addresses the role of non-Gaussian innovations in the multivariate realized GARCH models. A new approach is used with the assumption that the Realized Covariance follows an Inverse Wishart distribution with conditional mean the constructed conditional return covariance. The multivariate return distribution is explored with a parametric Student-t specification and a nonparametric extension with an independent infinite mixture of multivariate normals given a Dirichlet process prior.

Results from two different equity datasets highlight the strong significance of RC in the construction of conditional covariance. The asymmetry and thick tails of the return distribution are captured by the flexible DPM specification which can produce covariance matrices closer to RC, than the parametric models, and this is reflected in improved forecasts of its density. modeling the tail behaviour gives major return forecasting benefits compared to the Gaussian assumption.

Testing the MRGARCH models in portfolio risk management applications show the importance of non-Gaussian innovations. Specifically, in forecasting the risk measures of VaR and ES, the flexible semiparametric specification of MRGARCH-DPM models, consistently produce the best or the second best out-of-sample forecasts, for a set of confidence levels. In portfolio optimization, the models are tested on producing GMVP weights. The out-of-sample variance of the realized GMVP returns show that MRGARCH models and especially the ones with non-Gaussian innovations give the least volatile global minimum variance portfolios.

TABLE 3.1: Summary statistics of IBM, MSFT and XOM.

	Sample returns			Sample returns covariance			Mean of RC matrices		
	Mean	Skewness	Kurtosis	IBM	MSFT	XOM	IBM	MSFT	XOM
IBM	−0.0002	0.010	6.301	1.689			1.935		
MSFT	0.0004	0.243	6.129	1.156	2.060		1.116	2.460	
XOM	0.0002	−0.191	11.637	0.781	0.871	1.667	0.874	0.974	2.072

Notes: This table displays the summary statistics of IBM, MSFT and XOM sample for the period: 1/2/2001 - 31/12/2009 (2242 trading days).

TABLE 3.2: Summary statistics of AXP, DD, GE and KO.

	Sample returns			Sample returns covariance				Mean of RC matrices			
	Mean	Skewness	Kurtosis	AXP	DD	GE	KO	AXP	DD	GE	KO
AXP	−0.0002	0.317	11.245	4.343				4.419			
DD	−0.0014	0.028	7.243	1.725	2.108			1.457	2.530		
GE	0.0008	0.217	10.954	2.070	1.355	2.676		1.760	1.289	3.194	
KO	0.0009	0.104	6.913	0.892	0.628	0.759	1.131	0.853	0.729	0.824	1.415

Notes: This table displays the summary statistics of AXP, DD, GE and KO sample for the period: 1/2/2001 - 31/12/2009 (2242 trading days).

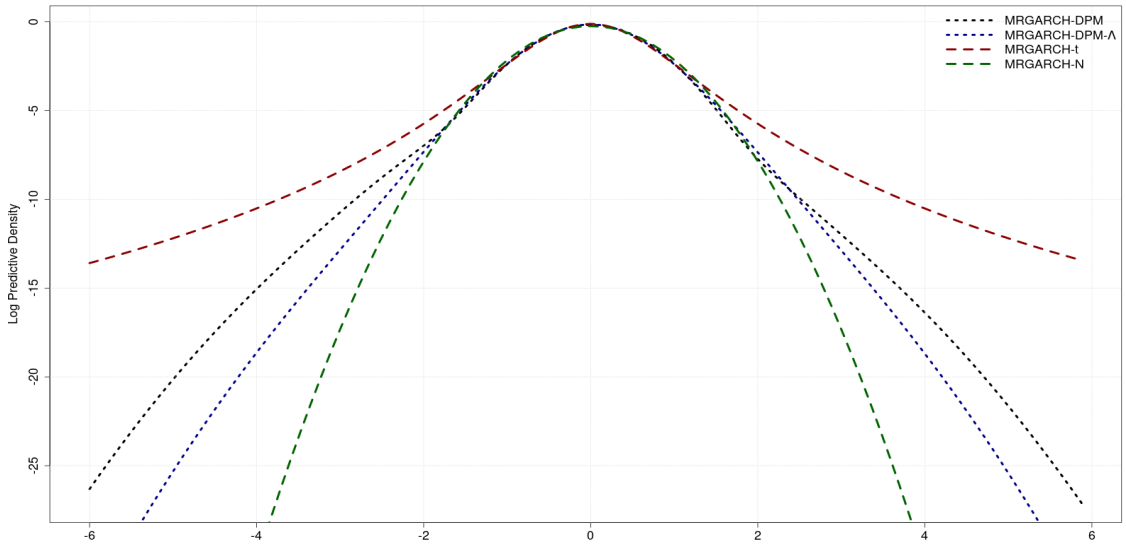
TABLE 3.3: Posterior Means (and 95% Density Interval) for **IBM(1)-MSFT(2)-XOM(3)**.

	MGARCH-N		**MGARCH-t**		**MGARCH-DPM-Λ**		**MGARCH-DPM**		**MRGARCH-N**		**MRGARCH-t**		**MRGARCH-DPM-Λ**		**MRGARCH-DPM**	
μ_1	-0.0077	(-0.049, 0.033)	-0.0081	(-0.046, 0.028)	-0.0071	(-0.046, 0.032)			-0.0105	(-0.052, 0.033)	-0.0039	(-0.045, 0.026)	-0.0114	(-0.047, 0.025)		
μ_2	0.003	(-0.042, 0.044)	-0.0028	(-0.041, 0.041)	-0.0045	(-0.044, 0.034)			0.0091	(-0.039, 0.05)	-0.0002	(-0.034, 0.03)	-0.0014	(-0.043, 0.037)		
μ_3	0.0045	(-0.036, 0.045)	0.0172	(-0.023, 0.06)	0.0119	(-0.03, 0.056)			0.0143	(-0.029, 0.056)	0.0194	(-0.005, 0.044)	0.0135	(-0.025, 0.055)		
Ω_{11}	0.005	(0.001, 0.010)	0.0042	(0.001, 0.009)	0.0069	(0.002, 0.014)	0.0078	(0.002, 0.015)	0.0244	(0.014, 0.033)	0.0179	(0.011, 0.027)	0.0239	(0.013, 0.033)	0.0236	(0.012, 0.034)
Ω_{21}	0.0009	(-0.002, 0.004)	0.0009	(-0.001, 0.004)	0.0014	(-0.002, 0.006)	0.0015	(-0.002, 0.006)	0.0109	(0.004, 0.017)	0.0076	(0.003, 0.013)	0.0112	(0.003, 0.018)	0.0109	(0.002, 0.018)
Ω_{31}	0.0032	(-0.001, 0.007)	0.0028	(0.000, 0.006)	0.004	(-0.001, 0.009)	0.0046	(0.0001, 0.01)	0.0094	(0.003, 0.015)	0.0073	(0.001, 0.012)	0.0094	(0.002, 0.015)	0.0094	(0.002, 0.015)
Ω_{22}	0.0065	(0.002, 0.012)	0.0036	(0.001, 0.008)	0.0051	(0.001, 0.011)	0.0045	(0.001, 0.01)	0.0217	(0.013, 0.031)	0.014	(0.008, 0.02)	0.0235	(0.013, 0.033)	0.0229	(0.012, 0.033)
Ω_{23}	0.0039	(0.000, 0.008)	0.003	(0.000, 0.007)	0.0041	(-0.001, 0.009)	0.0043	(0.0001, 0.009)	0.0097	(0.003, 0.015)	0.0078	(0.003, 0.012)	0.0106	(0.003, 0.017)	0.0106	(0.003, 0.017)
Ω_{33}	0.0198	(0.011, 0.030)	0.0143	(0.008, 0.023)	0.0219	(0.011, 0.034)	0.0231	(0.012, 0.038)	0.0357	(0.024, 0.046)	0.0278	(0.021, 0.034)	0.0381	(0.027, 0.049)	0.0395	(0.029, 0.050)
a_1	0.2147	(0.194, 0.238)	0.1855	(0.166, 0.210)	0.2039	(0.182, 0.229)	0.2092	(0.187, 0.235)	0.1356	(0.108, 0.164)	0.1048	(0.084, 0.122)	0.1230	(0.091, 0.147)	0.1216	(0.095, 0.151)
a_2	0.2161	(0.192, 0.240)	0.1844	(0.164, 0.207)	0.2069	(0.182, 0.239)	0.2059	(0.181, 0.234)	0.1393	(0.104, 0.177)	0.1001	(0.078, 0.118)	0.1211	(0.091, 0.148)	0.1199	(0.089, 0.148)
a_3	0.2348	(0.210, 0.261)	0.1991	(0.176, 0.226)	0.2272	(0.19, 0.261)	0.2289	(0.196, 0.262)	0.1387	(0.095, 0.173)	0.1366	(0.114, 0.154)	0.1450	(0.096, 0.177)	0.1474	(0.111, 0.176)
b_1	0.9728	(0.966, 0.978)	0.9748	(0.968, 0.980)	0.9747	(0.968, 0.98)	0.9733	(0.966, 0.979)	0.8247	(0.807, 0.841)	0.8313	(0.819, 0.845)	0.8293	(0.813, 0.846)	0.8294	(0.814, 0.845)
b_2	0.9727	(0.966, 0.979)	0.9758	(0.969, 0.981)	0.9753	(0.967, 0.981)	0.9754	(0.968, 0.981)	0.8326	(0.817, 0.844)	0.8276	(0.816, 0.840)	0.8303	(0.816, 0.843)	0.8303	(0.817, 0.843)
b_3	0.9644	(0.956, 0.972)	0.9688	(0.960, 0.976)	0.9651	(0.953, 0.975)	0.9644	(0.954, 0.974)	0.8272	(0.813, 0.843)	0.8247	(0.812, 0.835)	0.8251	(0.808, 0.840)	0.8252	(0.809, 0.840)
c_1									0.5283	(0.504, 0.554)	0.4473	(0.428, 0.465)	0.5244	(0.498, 0.549)	0.5247	(0.499, 0.547)
c_2									0.5206	(0.501, 0.542)	0.4579	(0.443, 0.472)	0.5283	(0.508, 0.550)	0.5287	(0.508, 0.549)
c_3									0.5190	(0.493, 0.543)	0.4435	(0.425, 0.459)	0.5183	(0.495, 0.544)	0.5166	(0.493, 0.540)
λ_1	0.4037	(0.272, 0.519)	0.3674	(0.233, 0.490)	0.3999	(0.225, 0.564)	0.3861	(0.221, 0.546)	0.8251	(0.583, 1.152)	0.8922	(0.540, 1.239)	0.9224	(0.590, 1.468)	0.9324	(0.624, 1.318)
λ_2	0.4052	(0.260, 0.533)	0.3619	(0.212, 0.499)	0.3973	(0.231, 0.552)	0.4182	(0.249, 0.554)	0.7492	(0.484, 1.063)	0.8223	(0.450, 1.266)	0.8281	(0.514, 1.232)	0.8505	(0.488, 1.244)
λ_3	0.2813	(0.138, 0.420)	0.2593	(0.089, 0.412)	0.3223	(0.143, 0.492)	0.3091	(0.129, 0.493)	0.6581	(0.393, 1.053)	0.6038	(0.389, 0.916)	0.6929	(0.396, 1.247)	0.6926	(0.436, 1.022)
ζ			9.1132	(7.745, 11.19)							7.1795	(6.639, 7.729)				
ν									15.643	(15.32, 15.99)	15.687	(15.33, 16.11)	15.689	(15.35, 16.04)	15.736	(15.42, 16.12)
α					0.6807	(0.27, 1.274)	0.3799	(0.088, 0.855)					0.7632	(0.302, 1.393)	0.4187	(0.112, 0.918)
κ					9.6096	(6, 14)	4.9638	(2, 9)					10.745	(7, 16)	5.5778	(3, 10)
V_{11}									0.9765	(0.958, 0.996)	0.953	(0.93, 0.977)	0.9726	(0.954, 0.991)	0.9713	(0.952, 0.991)
V_{21}									-0.0499	(-0.061, -0.038)	-0.0435	(-0.055, -0.032)	-0.0488	(-0.06, -0.038)	-0.0488	(-0.06, -0.037)
V_{31}									-0.0247	(-0.036, -0.013)	-0.0275	(-0.039, -0.016)	-0.0256	(-0.037, -0.014)	-0.0261	(-0.037, -0.015)
V_{22}									1.0788	(1.058, 1.101)	1.0629	(1.037, 1.092)	1.0715	(1.053, 1.092)	1.0703	(1.051, 1.09)
V_{23}									-0.0141	(-0.026, -0.002)	-0.0168	(-0.029, -0.005)	-0.0144	(-0.026, -0.003)	-0.0145	(-0.026, -0.003)
V_{33}									1.0788	(1.059, 1.099)	1.0568	(1.031, 1.083)	1.0776	(1.055, 1.101)	1.0762	(1.056, 1.097)

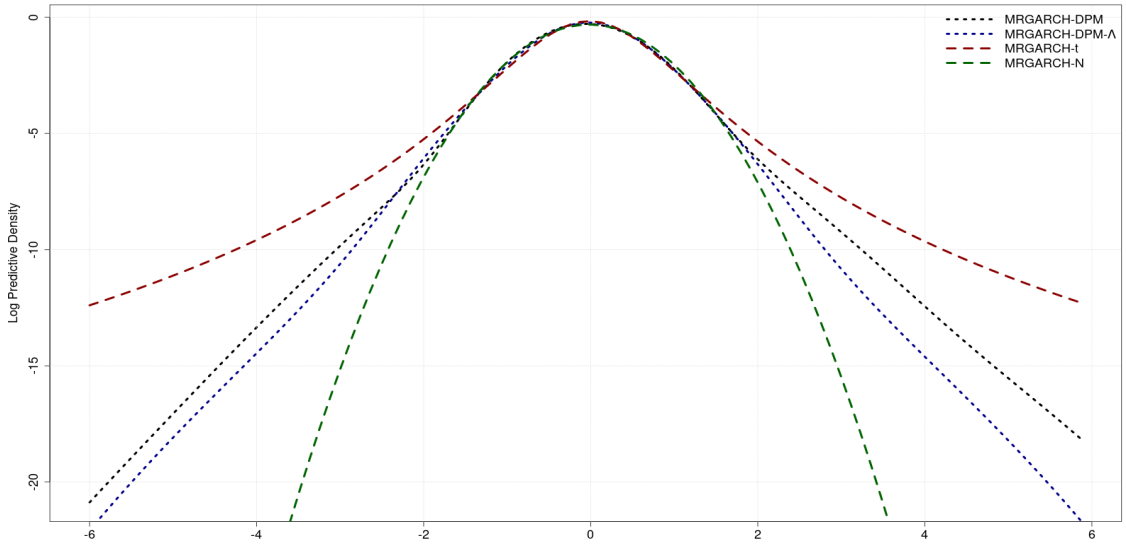
Notes: Results are from 10,000 MCMC posterior draws after 5,000 burnin. a, b, c and λ are the GARCH vectors of parameters. ζ is the St-t degree of freedom, ν is the IW degree of freedom, α is DPM's precision parameter and κ the number of active normal clusters. Parameters in $\{\Omega_{11}, ..., \Omega_{33}\} = \text{vech}(\Omega)$ are calculated from covariance targeting. Parameters $\{V_{11}, ..., V_{33}\} = \text{vech}(V)$ are the elements of the scale matrix in the measurement equation of the models. MGARCH models do not contain RC data in their specification.

FIGURE 3.1: Log predictive density of the EWP returns.

(a) IBM-MSFT-XOM



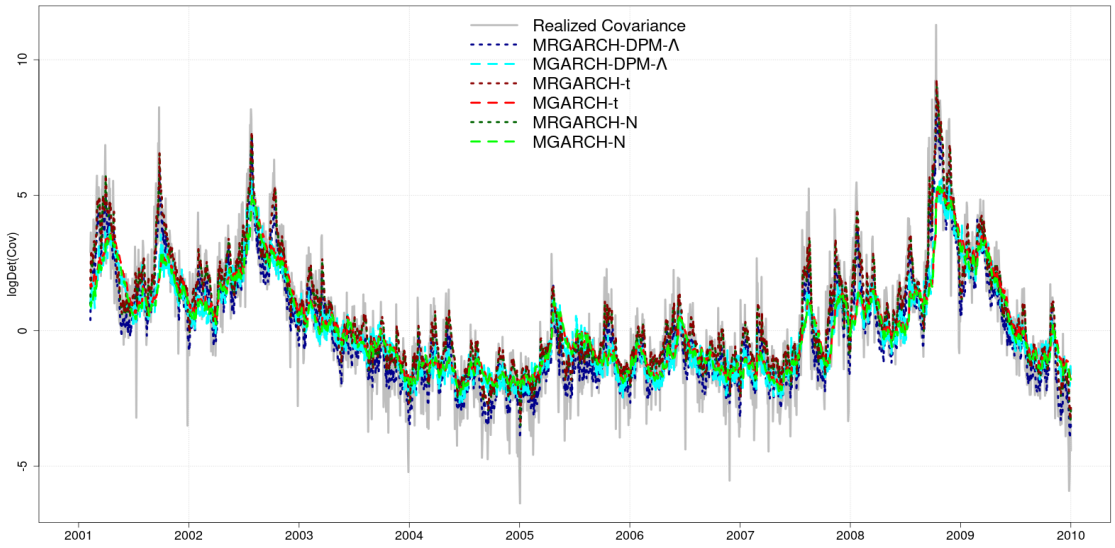
(b) AXP-DD-GE-KO



Notes: The density is calculated by evaluating a grid of one-day-ahead returns with the full sample posterior draws.

FIGURE 3.2: Log determinant of estimated Covariance matrices.

(a) IBM-MSFT-XOM



(b) AXP-DD-GE-KO

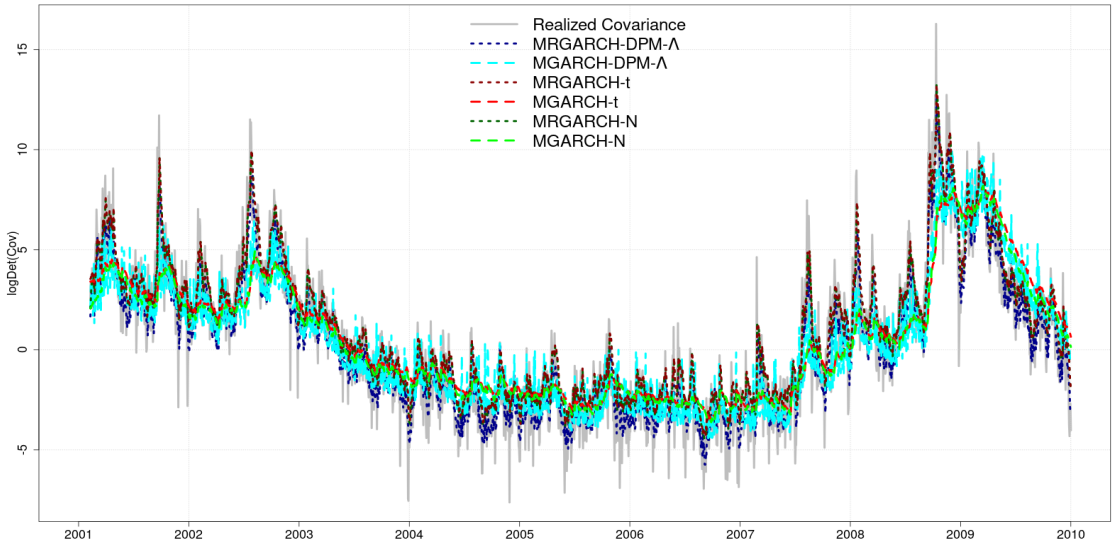


TABLE 3.4: Forecasting results.

	IBM-MSFT-XOM				AXP-DD-GE-KO			
	returns		Realized Covariance		returns		Realized Covariance	
	RMSFE	logPL	RMSFE	logPL	RMSFE	logPL	RMSFE	logPL
MGARCH-N	1.6575	−1417.60	4.0079		**2.6483**	−2407.09	14.515	
MGARCH-t	**1.6571**	−1408.31	4.1726		2.6484	−2378.44	15.577	
MGARCH-DPM-Λ	1.6576	−1410.76	3.8932		2.6486	−2377.17	14.128	
MGARCH-DPM	**1.6571**	−1412.95	3.8698		2.6493	−2381.29	14.053	
MRGARCH-N	1.6575	−1396.13	3.1765	−1038.56	2.6498	−2364.35	9.8772	−3749.95
MRGARCH-t	1.6572	**−1386.73**	3.2113	−1038.79	2.6491	**−2343.03**	9.9644	−3752.71
MRGARCH-DPM-Λ	1.6577	−1387.43	**3.1139**	**−1037.98**	2.6493	−2344.84	9.8258	−3747.94
MRGARCH-DPM	1.6581	−1388.40	3.1211	−1038.18	2.6501	−2355.12	**9.7138**	**−3747.71**
Time period	23/10/2008 - 31/12/2009				23/10/2008 - 31/12/2009			
Out-of-sample forecasts (τ)	300				300			

Notes: This table displays the root mean squared forecast error (RMSFE) and the cumulative log predictive likelihood (logPL) of returns (r) and realized covariance (RC). Bold values are the best among the models.

TABLE 3.5: Portfolio forecasting results.

	EWP of IBM-MSFT-XOM			EWP of AXP-DD-GE-KO		
	RMSFE	logPL	Tail logPL	RMSFE	logPL	Tail logPL
MGARCH-N	1.4534	−458.65	−44.57	**2.0811**	−592.78	−98.93
MGARCH-t	**1.4529**	−456.88	−44.88	2.0812	−594.60	−99.11
MGARCH-DPM-Λ	1.4535	−455.37	−44.62	2.0814	−590.41	−98.83
MGARCH-DPM	1.4531	−456.24	−44.93	2.0822	−592.08	−99.22
MRGARCH-N	1.4534	−452.49	**−42.69**	2.0826	**−578.22**	**−96.76**
MRGARCH-t	1.4529	−449.83	−43.87	2.0819	−580.59	−97.55
MRGARCH-DPM-Λ	1.4536	**−449.64**	−43.17	2.0822	−578.56	−97.01
MRGARCH-DPM	1.4540	−450.59	−43.18	2.0829	−578.89	−97.13
Out-of-sample forecasts (τ)	300		45	300		74
Time period	23/10/2008 - 31/12/2009			23/10/2008 - 31/12/2009		

Notes: This table displays the equal weights portfolio (EWP) root mean squared forecast error (RMSFE) of its returns and the cumulative log-predictive likelihoods (logPL) of its density. Tail predictive likelihood is: $p(r^p_{t+1}|r^p_{t+1} < \eta, \mathcal{I}_t)$, for $\eta = -1.0$. Bold values are the best among the models.

FIGURE 3.3: Realized returns, forecasted 1% Value-at-Risk (top) and 1% Expected Shortfall (bottom) for the EWP of **IBM-MSFT-XOM**.

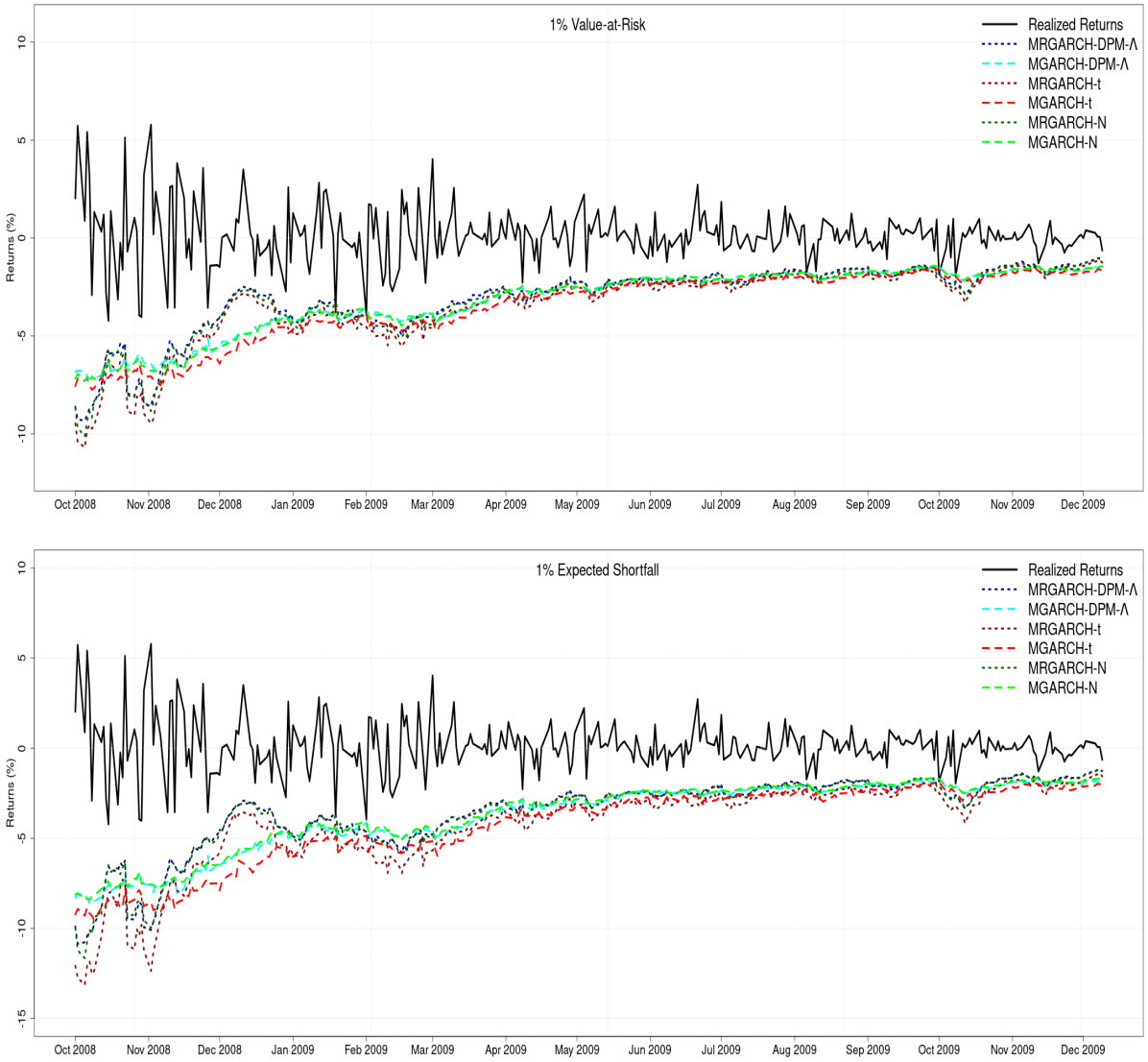


FIGURE 3.4: Realized returns, forecasted 1% Value-at-Risk (top) and 1% Expected Shortfall (bottom) for the EWP of **AXP-DD-GE-KO**.

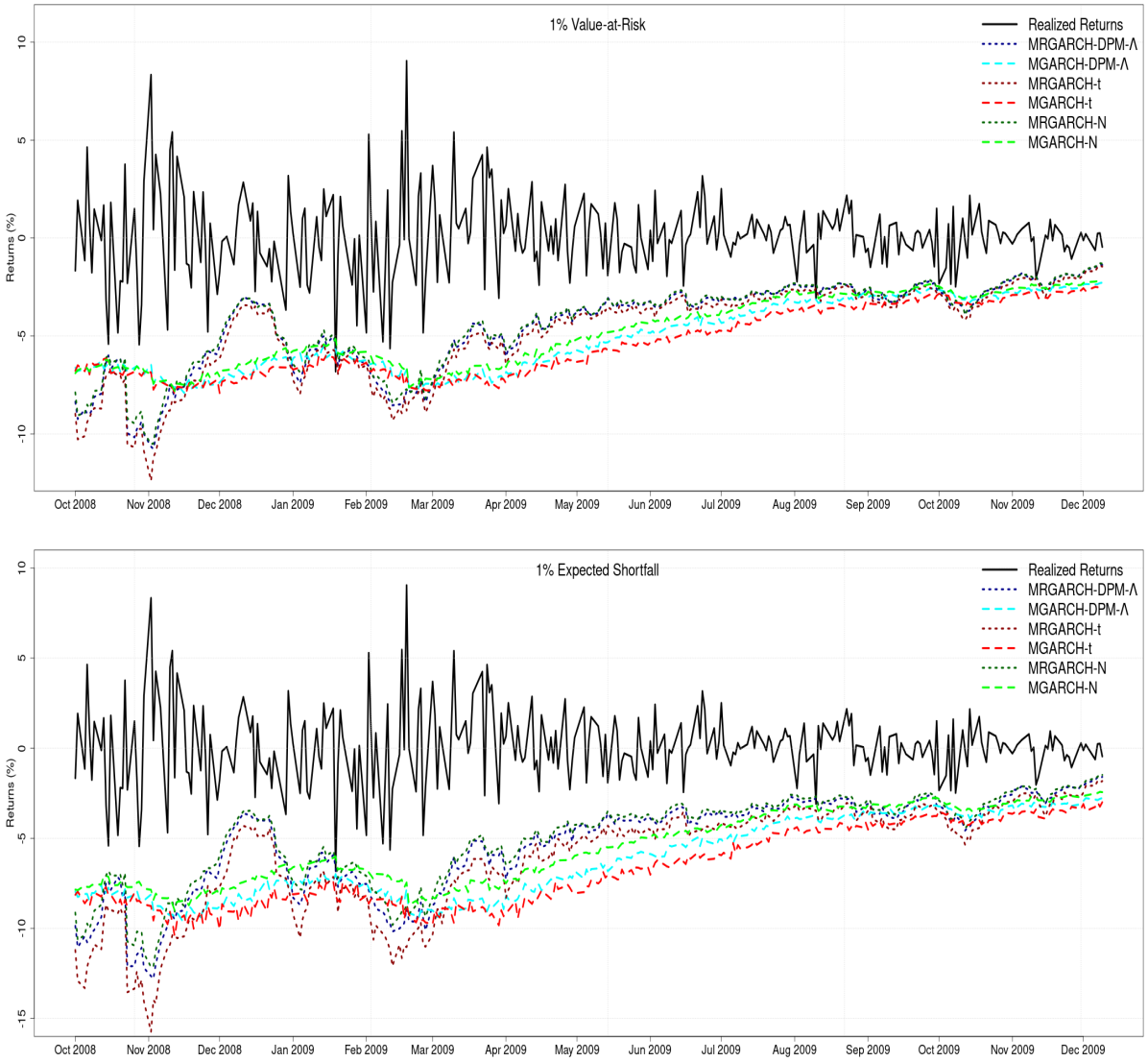


TABLE 3.6: VaR violation rates.

	IBM-MSFT-XOM			AXP-DD-GE-KO		
Value-at-Risk level (ϵ):	0.01	0.05	0.1	0.01	0.05	0.1
MGARCH-N	1.667	0.533	0.800	0.667	0.933	0.733
MGARCH-t	0.667	0.533	0.833	0.333	0.933	0.767
MGARCH-DPM-Λ	**1.000**	0.733	0.833	0.333	**1.000**	0.867
MGARCH-DPM	1.667	0.667	0.833	0.667	1.067	0.833
MRGARCH-N	0.667	0.800	0.767	**1.000**	1.067	**1.067**
MRGARCH-t	0.333	0.867	**0.967**	0.667	1.200	1.167
MRGARCH-DPM-Λ	1.333	**1.000**	**0.967**	**1.000**	1.067	1.133
MRGARCH-DPM	1.333	0.933	**0.967**	**1.000**	1.200	1.100
Time period	23/10/2008 - 31/12/2009					
Out-of-sample forecasts (τ)	300					

Notes: This table displays the violation rates (VR) of VaR calculated from the models for EWPs. Bold indicates the VR values closest to one, among the models. The VR is calculated as: $\frac{1}{\epsilon\tau}\sum_{t=T-\tau}^{T-1}\mathbf{1}\left\{r_{t+1}^{p}\leq \text{VaR}_{t+1}^{\epsilon}\right\}$.

TABLE 3.7: Average out-of-sample losses.

	IBM-MSFT-XOM			AXP-DD-GE-KO		
Level (ϵ):	0.01	0.05	0.1	0.01	0.05	0.1
MGARCH-N	1.174	0.897	0.700	1.630	1.374	1.185
MGARCH-t	1.191	0.903	0.697	1.666	1.384	1.191
MGARCH-DPM-Λ	1.167	0.876	0.688	1.624	1.368	1.173
MGARCH-DPM	1.181	0.871	0.686	1.626	1.373	1.178
MRGARCH-N	**1.066**	0.839	0.663	1.564	1.261	**1.092**
MRGARCH-t	1.142	0.839	0.655	1.581	1.266	1.100
MRGARCH-DPM-Λ	1.069	**0.829**	**0.648**	1.570	1.263	1.096
MRGARCH-DPM	1.082	**0.829**	0.652	**1.540**	**1.257**	1.095
Time period	23/10/2008 - 31/12/2009					
Out-of-sample forecasts (τ)	300					

Notes: This table displays the average out-of-sample losses calulated as: $\frac{1}{\tau}\sum_{t=T-\tau}^{T-1}\mathcal{L}(r_{t+1}^{p}, \text{VaR}_{t+1}^{\epsilon}, \text{ES}_{t+1}^{\epsilon}, \epsilon)$. The lowest values among the models are in bold.

TABLE 3.8: Average out-of-sample variance of GMVP returns.

	IBM-MSFT-XOM	AXP-DD-GE-KO
MGARCH-N	2.1531	2.0441
MGARCH-t	2.0942	2.0071
MGARCH-DPM-Λ	2.1355	2.0256
MGARCH-DPM	2.1329	2.0405
MRGARCH-N	1.9661	1.9020
MRGARCH-t	1.9651	**1.9000**
MRGARCH-DPM-Λ	**1.9643**	1.9215
MRGARCH-DPM	**1.9643**	1.9245
Equal weights portfolio	2.0942	4.3231
Time period	23/10/2008- 31/12/2009	
Out-of-sample forecasts	300	

Notes: This table displays the sample variance ($\overline{\sigma}^{2p}$) of the realized GMV portfolio returns. Bold indicates the lowest values among the models.

Chapter 4

Conclusion

This thesis contributes to three areas of financial volatility modeling in which the ex post measures of variance or covariance are used. The presented studies show that using parametric or semiparametric non-Gaussian assumptions for financial returns and variance, provides statistical benefits in density forecasting and improves risk measurement.

The first study is on modeling jumps in ex post variance, their size and time-dependent arrival. The focus is on the forecasting benefits that the modelled jumps provide for the predictive distribution of returns and ex post variance. The study extends a joint model of returns and ex post variance with two autoregressive components to capture the short and long-run variance persistence.

The heterogeneous variance jumps model is applied on the financial index of S&P500 and the results show frequent and persistent variance jumps which occur at least once per week. Modeling variance jumps significantly improves ex post variance and returns density forecasts for forecasting horizons of up to 50 days out-of-sample.

The second study explores the empirical non-Gaussian features of stochastic volatility. A new semiparametric approach is proposed to estimate the discrete-time stochastic volatility with a Dirichlet Process mixture of infinite normal kernels. The new model has the ability to capture jumps, asymmetry and thick-tails in returns, realized and stochastic volatility.

The new semiparametric stochastic volatility model is applied to a group of equity indices. The results show strong evidence of non-Gaussian behaviour in volatility and returns for all the tested series. In out-of-sample forecasting, the new semiparametric approach gives improved density forecasts of log-realized variance compared to the Gaussian benchmark. In risk measurement forecasting, the new flexible estimation of stochastic volatility gives improve return density forecasts and can allocate more probability mass to negative returns and to losses larger than 1%.

In the third study, a new approach is proposed for multivariate realized GARCH models. Two new extensions are developed which capture the empirically observed non-Gaussian behaviour of financial returns. The first one is a multivariate Student-t assumption which captures the thick-tails in returns distribution. The second one is a nonparametric approximation of the underlying return density with a Dirichlet Process mixture of infinite multivariate normal kernels.

The models application into two different equity datasets, of 3 and 4 stock returns, shows the major forecasting benefits of the non-Gaussian innovations. The proposed models, outperform the benchmarks in returns and realized covariance density forecasts. In portfolio applications, the non-Gaussian multivariate realized GARCH models, improve the forecasting of the equal weight portfolio Value-at-Risk and Expected Shortfall. Finally, it is shown that they can produce the least volatile global minimum variance portfolios. This indicates that the proposed frameworks estimate the returns covariance closer to the unknown true covariance.

Appendix A

A1 Sampling steps of 2Comp-RM-HJ

The algorithm for estimating the 2Comp-RM-HJ model in 1.3.1 is:

1. The parameters $\theta = \{\mu, \omega, \phi_1, \phi_2, \alpha_1, \alpha_2, \rho, \sigma^2\}$ of the 2Comp-RM model have the conditional posterior density which is proportional to

$$p(\theta) \prod_{t=1}^{T} \mathrm{N}\left(r_t|\mu, \mathrm{RM}_t\right) \sum_{j=0}^{1} \mathrm{N}\left(\mathrm{logRM}_t \middle| \omega + \sum_{i=1}^{2} \phi_i c_{i,t} + \rho u_{t-1} + j\mu_\xi,\ \sigma^2 + j\sigma_\xi^2\right) \mathrm{P}(J_t = j|z_t), \tag{A.1}$$

with $p(\theta)$ being the prior density and the rest is the data likelihood. The above posterior does not have a standard form. To get posterior draws, I use a random walk MH algorithm with proposal $\theta^{'}$ which is from the following thick-tailed distribution

$$h(\theta^{'}) \sim \begin{cases} \mathrm{N}(\theta^{(l-1)}, \kappa \times I_k), & \text{with probability } p = 0.9, \\ \mathrm{N}(\theta^{(l-1)}, 10\kappa \times I_k), & \text{with probability } 1 - p, \end{cases} \tag{A.2}$$

with k being the number of parameters in vector θ. The draw θ' is accepted with probability: $\min\left\{p(\theta^{'}|\mathcal{I}_T)/p(\theta^{(l-1)}|\mathcal{I}_T), 1\right\}$. Draws which do not satisfy: $0 < \phi_i, < 1$, $0 < \alpha_i < 1,\ i = 1, 2, \alpha_1 > \alpha_2$ and $\sigma^2 > 0$ are rejected.

Conditional on θ, define $y = y_1, ..., y_T$, where

$$y_t \equiv \text{logRM}_t - \left(\omega + \sum_{i=1}^{2} \phi_i c_{i,t} + \rho u_{t-1} \right) \tag{A.3}$$

which based on eq.(1.13) has the following form

$$y_t = J_t \xi_t + \sigma v_t, \quad v_t \sim \text{NID}(0,1). \tag{A.4}$$

2. Jump size follows a normal distribution, $\xi_t \sim \text{N}(\mu_\xi, \sigma_\xi^2)$. This is essentially a linear model with only a constant. Following the Bayes theorem, the mean has a conjugate normal prior $p(\mu_\xi) \sim \text{N}(m_0, v_0^2)$, and its conditional posterior is

$$\begin{aligned} p(\mu_\xi | \sigma_\xi^2, \xi) &\propto p(\xi | \mu_\xi, \sigma_\xi^2) p(\mu_\xi) \\ &\sim \text{N}(M, V^{-1}) \\ V &= \sigma_\xi^{-2} T + v_0^{-2} \\ M &= V^{-1} (\sigma_\xi^{-2} T \bar{\xi} + m_0 v_0^{-2}) \end{aligned} \tag{A.5}$$

where $\bar{\xi}$ is the mean of jump size vector ξ.

3. For the variance, I use a conjugate inverse gamma prior, $p(\sigma_\xi^2) \sim \text{IG}(v_0/2, s_0/2)$, and its conditional posterior is

$$p(\sigma_\xi^2 | \mu_\xi, \xi) \sim \text{IG} \left(\frac{T + v_0}{2}, \frac{\sum_{t=1}^{T} (\xi_t - \mu_\xi)^2 + s_0}{2} \right). \tag{A.6}$$

4. Jump size ξ_t, has a conjugate prior, $p(\xi_t) \sim \mathrm{N}(\mu_\xi, \sigma_\xi^2)$, and its conditional posterior is

$$\begin{aligned} p(\xi_t|\sigma^2, \mu_\xi, \sigma_\xi^2, J_t, y_t) &\propto p(\xi_t|\mu_\xi, \sigma_\xi^2)\; p(y_t|\sigma^2, \xi_t, J_t), \quad t = 1, ..., T \\ &\sim \mathrm{N}(a, A^{-1}) \\ A &= \sigma^{-2} J_t + \sigma_\xi^{-2} \\ a &= A^{-1}\left[\sigma^{-2} J_t\,(y_t) + \mu_\xi \sigma_\xi^{-2}\right] \end{aligned} \tag{A.7}$$

   If there is no jump, $J_t = 0$, then the posterior is identical to the prior. If there is a jump, $J_t = 1$, then posterior contains information about the jump size.

5. Jump indicator, $J_t,\ t = 1, ..., T$, is a Bernoulli variable with associated probabilities λ_t. Its conditional posterior has two outcomes, $J_t \in \{0, 1\}$, with probabilities

$$\begin{aligned} p(J_t = 0|\sigma^2, \xi_t, \lambda_t, y_t) &\propto p(y_t|\sigma^2, \xi_t)\; p(J_t = 0|\lambda_t) \\ &\propto \exp\left(-0.5\sigma^{-2}(y_t)^2\right)(1 - \lambda_t) \end{aligned} \tag{A.8}$$

$$\begin{aligned} p(J_t = 1|\sigma^2, \xi_t, \lambda_t, y_t) &\propto p(y_t|\sigma^2, \xi_t)\; p(J_t = 1|\lambda_t) \\ &\propto \exp\left(-0.5\sigma^{-2}(y_t - \xi_t)^2\right)\lambda_t \end{aligned} \tag{A.9}$$

6. To get the latent process z, I use a single move sampler based on the one from Kim et al. (1998) for the stochastic volatility models. By applying the Bayes rule twice, the conditional posterior of $z_t,\ t = 1, ..., T$, is

$$\begin{aligned} p(z_t|z_{-t}, J_t, \gamma) &\propto p(J_t|z_t)\; p(z_t|z_{-t}, \gamma) \\ &\propto p(J_t|z_t)\; p(z_{t+1}|z_t, \gamma)\; p(z_t|z_{t-1}, \gamma). \end{aligned} \tag{A.10}$$

The above posterior does not have a known form. I use an MH step to sample from it. In eq.(A.10), the last two densities can be combined as

$$p(z_t|z_{-t},\gamma) \propto p(z_{t+1}|z_t,\gamma)\, p(z_t|z_{t-1},\gamma) \propto \mathrm{N}(\mu_h,\delta^2) \tag{A.11}$$
$$\text{with } \mu_h \frac{\gamma_0(1-\gamma_1)+\gamma_1(z_{t+1}+z_{t-1})}{1+\gamma_1^2} \text{ and } \delta^2 = \frac{1}{1+\gamma_1^2}.$$

To draw from the posterior in (A.10), I draw the candidate z' from $q(z'|z_{-t},\gamma) \sim \mathrm{N}(\mu_h,\delta^2)$ and accept the draw with probability

$$\min\left\{\frac{p(z'|z_{-t},J_t,\gamma)/q(z'|z_{-t},\gamma)}{p(z_t|z_{-t},J_t,\gamma)/q(z_t|z_{-t},\gamma)},1\right\}. \tag{A.12}$$

If the draw is accepted then $z_t^{(l)} = z'$, otherwise $z_t^{(l)} = z_t^{(l-1)}$. In the sampler, $z_0^{(l)} \sim \mathrm{N}(\bar{z},1)$, with $\bar{z} = \frac{\gamma_0^{(l)}}{1-\gamma_1^{(l)}}$.

7. Conditional on z we have the linear model

$$z_t = X_t\gamma + e_t, \quad e_t \sim \mathrm{N}(0,1) \tag{A.13}$$

where $\gamma = \gamma_0,\gamma_1$ and $X_t = [1\ z_{t-1}]$. The parameters in γ, have a conjugate bivariate normal prior $p(\gamma) \sim \mathrm{N}(a_0,B_0)$ and are sampled from

$$p(\gamma|\Theta_{-\gamma},z) \sim p(\gamma)\prod_{t=1}^{T} p(z_t|\Theta_{-\gamma},\gamma) \tag{A.14}$$
$$\sim \mathrm{N}(m_g,V_g) \tag{A.15}$$
$$\text{with } V_g = \left(X'X + B_0^{-1}\right)^{-1}$$
$$m_g = V_g^{-1}\left(X'z + B_0^{-1}a_0\right).$$

Draws of γ are accepted if $|\gamma_1| < 1$.

The steps 1-5 for estimating the 2Comp-RM-IJ model in 1.3.2 are the same as the steps of

2Comp-RM-HJ with a constant jump intensity λ. To sample λ in 6, I assume a conjugate Beta prior, $p(\lambda) \sim \mathrm{B}(a_\lambda, b_\lambda)$. Its conditional posterior is

$$p(\lambda|J) \propto p(J|\lambda)p(\lambda) \tag{A.16}$$

$$\sim \mathrm{B}\left(a_\lambda + \sum_{t=1}^{T} J_t, T + b_\lambda - \sum_{t=1}^{T} J_t\right). \tag{A.17}$$

A2 Sampling steps of RSV-DPM-V

The algorithm for estimating the RSV-DPM-V model in 2.3.1 is:

1. To estimate the stochastic volatility vector $h_{1:T}$ I use a single move sampler based on the work of Kim et al. (1998). Following the Bayes rule twice, the conditional posterior $p(h_t|h_{-t}, r_t, \text{logRV}_t, \theta, \mu_{1:k}, \gamma_{1:k}, \sigma^2_{1:k}, s_{1:T})$ is proportional to

$$p(r_t|\mu_{s_t}, h_t)\ p(\text{logRV}_t|h_t, \beta, b^2)\ p(h_t|h_{-t}, \delta, \gamma_{s_t}, \sigma^2_{s_t}) \tag{A.18}$$

$$p(r_t|\mu_{s_t}, h_t)\ p(\text{logRV}_t|h_t, \beta, b^2)\ p(h_t|h_{t-1}, \delta, \gamma_{s_t}, \sigma^2_{s_t})\ p(h_{t+1}|h_t, \delta, \gamma_{s_{t+1}}, \sigma^2_{s_{t+1}}) \tag{A.19}$$

$$p(r_t|\mu_{s_t}, h_t)\ p(\text{logRV}_t|h_t, \beta, b^2)\ f_{\mathrm{N}}(h_t|\bar{h}, v^2_h) \tag{A.20}$$

$$p(r_t|\mu_{s_t}, h_t)\ f_{\mathrm{N}}(h_t|M_h, V) \tag{A.21}$$

with $f_{\mathrm{N}}(.)$ being the normal density. The posterior in (A.19) has an unknown form. By combining the last two densities we get that $p(h_t|h_{-t}, \delta, \gamma_{s_t}, \sigma^2_{s_t}) \propto f_{\mathrm{N}}(h_t|\bar{h}, v^2_h)$ with

$$\bar{h} = \frac{\delta\sigma^2_{s_t}(h_{t+1} - \gamma_{s_{t+1}}) + \sigma^2_{s_{t+1}}(\delta h_{t-1} + \gamma_{s_t})}{\sigma^2_{s_{t+1}} + \delta^2\sigma^2_{s_t}} \tag{A.22}$$

$$v^2_h = \frac{\sigma^2_{s_t}\sigma^2_{s_{t+1}}}{\sigma^2_{s_{t+1}} + \delta^2\sigma^2_{s_t}} \tag{A.23}$$

Using the logRV data in (A.20) we get that $h_t|h_{-t}, \text{logRV}_t, \theta, \gamma_{1:k}, \sigma^2_{1:k}, s_{1:T} \sim \text{N}(M_h, V)$ with

$$M_h = \frac{d(\text{logRV}_t - c)v_h^2 + \bar{h}v_h^2}{d^2 v_h^2 + b^2} \tag{A.24}$$

$$V = \frac{b^2 v_h^2}{d^2 v_h^2 + b^2} \tag{A.25}$$

The posterior in (A.21) does not have a known form so a Metropolis-Hastings (MH) algorithm is used to sample from it. The proposal distribution is found following the results of Kim et al. (1998). They show that $\exp(-h_t)$ is bounded and

$$p(r_t|\mu_{s_t}, h_t) \propto f(r_t, h_t, \mu_{s_t}) = \exp\{-.5h_t - .5\exp(-h_t)(r_t - \mu_{s_t})^2\} \tag{A.26}$$

$$\leq \exp\{-.5h_t - .5\exp(-M_h)(r_t - \mu_{s_t})^2(1 + M_h - h_t)\} \tag{A.27}$$

$$= g(r_t, h_t, \mu_{s_t}, M_h) \tag{A.28}$$

Combining this with (A.21) we get the proposal

$$p(r_t|\mu_{s_t}, h_t)\ f_{\text{N}}(h_t|M_h, V) \leq g(r_t, h_t, \mu_{s_t}, M_h) f_{\text{N}}(h_t|M_h, V) \propto f_{\text{N}}(h_t|M, V) \tag{A.29}$$

with $M = M_h + .5V((r_t - \mu_{s_t})^2 \exp(M_h) - 1)$. The candidate $h'_t \sim \text{N}(M, V)$ is accepted as a draw of h_t with probability

$$\min\left\{\frac{p(h'_t|h_{-t}, r_t, \text{logRV}_t, \theta, \mu_{1:k}, \gamma_{1:k}, \sigma^2_{1:k}, s_{1:T})/\ \text{N}(h'_t|M, V)}{p(h_t|h_{-t}, r_t, \text{logRV}_t, \theta, \mu_{1:k}, \gamma_{1:k}, \sigma^2_{1:k}, s_{1:T})/\ \text{N}(h_t|M, V)}, 1\right\}. \tag{A.30}$$

2. AR(1) parameter δ has a normal prior, $p(\delta) \sim \text{N}(m_\delta, v_\delta)\mathbf{1}\{|\delta| < 1\}$. Its conditional posterior is

$$p(\delta|h_{1:T}, \gamma_{1:k}, \sigma^2_{1:k}, s_{1:T}) \propto \text{N}(M_\delta, V_\delta) \tag{A.31}$$

$$\text{with } V_\delta = \frac{v_\delta}{v_\delta \sum_{t=2}^{T} h^2_{t-1}\sigma_t^{-2} + 1} \tag{A.32}$$

$$\text{and } M_\delta = V_\delta \left(\sum_{t=2}^{T} h_{t-1}(h_t - \gamma_{s_t})\sigma_{s_t}^{-2} + m_\delta / v_\delta \right). \tag{A.33}$$

3. The mixing parameters are drawn from linear model conjugate priors

   (a) $p(\mu_j) \sim \text{N}(m_0, u_0^2),\ j = 1, ..., k$. Its conditional posterior is proportional to

$$p(\mu_j|r_{1:T}, h_{1:T}, s_t = j) \propto p(\mu_j) \prod_{t:s_t=j} p(r_t|\mu_j, h_t) \tag{A.34}$$

$$\propto \text{N}\left(m_\mu, v_\mu^{-1}\right) \tag{A.35}$$

$$\text{with } v_\mu = \sum_{t:s_t=j} \exp(-h_t) + u_0^{-2} \tag{A.36}$$

$$\text{and } m_\mu = v_\mu^{-1} \left(\sum_{t:s_t=j} r_t \exp(-h_t) + m_0 u_0^{-2} \right). \tag{A.37}$$

   (b) $p(\gamma_j) \sim \text{N}(g_0, q_0^2),\ j = 1, ..., k$. Its conditional posterior is proportional to

$$p(\gamma_j|h_{1:T}, \delta, \sigma_j^2, s_t = j) \propto p(\gamma_j) \prod_{t:s_t=j} p(h_t|\gamma_j, \delta, \sigma_j^2) \tag{A.38}$$

$$\propto \text{N}\left(m_\gamma, v_\gamma^2\right) \tag{A.39}$$

$$\text{with } v_\gamma^2 = \frac{\sigma_j^2 q_0^2}{n_j q_0^2 + \sigma_j^2} \tag{A.40}$$

$$\text{and } m_\gamma = v_\gamma^2 \left(\sigma_j^{-2} \sum_{t:s_t=j} (h_t - \delta h_{t-1}) + g_0 q_0^{-2} \right), \tag{A.41}$$

where $n_j = \sum_{t:s_t=j} \mathbf{1}\{s_t = j\}$, is the number of observations in the cluster j.

(c) $p(\sigma_j^2) \sim \text{IG}(v_0/2, s_0/2),\ j = 1, ..., k$. Its conditional posterior is proportional to

$$p(\sigma_j^2|h_{1:T}, \delta, \gamma_j, s_t = j) \propto p(\sigma_j^2) \prod_{t:s_t=j} p(h_t|\gamma_j, \delta, \sigma_j^2) \tag{A.42}$$

$$\propto \text{IG}\left(.5\left(n_j + v_0\right), .5\left(\sum_{t:s_t=j}\left(h_t - \gamma_j - \delta h_{t-1}\right)^2 + s_0\right)\right). \tag{A.43}$$

4. (a) Update the mixture weights in $w_{1:k}|s_{1:T}, \alpha$ with a stick-breaking process as

$$v_j|s_{1:T}, \alpha \sim \text{B}\left(1 + \sum_{t=1}^{T} \mathbf{1}\left\{s_t = j\right\}, \alpha + \sum_{t=1}^{T} \mathbf{1}\{s_t > j\}\right), \tag{A.44}$$

$$w_1 = v_1, \quad w_j = v_j \prod_{l=1}^{j-1}(1 - v_l), \quad j = 2, ..., k. \tag{A.45}$$

(b) Update the slice vector $u_{1:T}|w_{1:k}, s_{1:T}$ from a uniform draw as: $u_t|w_{1:k}, s_{1:T} \sim \text{U}(0, w_{s_t})$.

(c) Update the number of mixture clusters k to the smallest positive integer that satisfies: $\sum_{j=1}^{k} w_j > 1 - \min(u_{1:T})$. If new clusters are needed to satisfy the inequality, their mean and covariance are drawn from the base measure (N-N-IG) in (2.10).

5. Each element s_t of the vector $s_{1:T}$ takes an integer value j which is drawn from a multinomial distribution with probabilities

$$p(s_t = j|r_{1:T}, h_{1:T}, \mu_{1:k}, \gamma_{1:k}, \sigma_{1:k}^2, w_{1:k}, \alpha) \propto \mathbf{1}\left\{u_t < w_j\right\} \text{N}\left(r_t|\mu_j, \exp(h_t)\right) \times \text{N}\left(h_t|\gamma_j + \delta h_{t-1}, \sigma_j^2\right), \tag{A.46}$$

for $j = 1, ..., k$. The number of active clusters κ, can be calculated as the ones with at least one assigned data observation.

6. Draw the DPM precision parameter α with a gamma prior $\alpha \sim \Gamma(a_0, b_0)$ following the two steps algorithm of Escobar and West (1995):

   i. draw the random variable $\xi|\alpha, \kappa \sim \text{B}(\alpha + 1, T)$.

   ii. sample α from

$$\alpha|\xi \sim \pi_\xi \Gamma(a_0 + \kappa, b_0 - \log(\xi)) + (1 - \pi_\xi)\Gamma(a_0 + \kappa - 1, b_0 - \log(\xi)),$$

   with $\frac{\pi_\xi}{1-\pi_\xi} = \frac{a_0+\kappa-1}{T(b_0-\log(\xi))}$.

7. Conditional on h, the measurement equation is the following linear model

$$y_t = X_t\beta + bz_t, \quad z_t \sim \text{N}(0, 1), \tag{A.47}$$

with $y_t = \text{logRV}_t$, $X_t = [1 \quad h_t]$ and $\beta = [c \quad d]'$.

   (a) $p(\beta) \sim \text{N}(\beta_0, B_0)$, β_0 is a 2×1 vector and B_0 a 2×2 matrix. Its conditional posterior is proportional to

$$p(\beta|y_{1:T}, X_{1:T}, b^2) \propto p(\beta) \prod_{t=1}^{T} p(y_t|X_t\beta, b^2) \tag{A.48}$$

$$\propto \text{N}\left(M, V^{-1}\right) \tag{A.49}$$

$$\text{with } V = b^{-2}X'X + B_0^{-1} \tag{A.50}$$

$$\text{and } M = V^{-1}\left(b^{-2}X'y + B_0^{-1}\beta_0\right). \tag{A.51}$$

   (b) $p(b^2) \sim \text{IG}(\nu_0/2, \varsigma_0/2)$. Its conditional posterior is proportional to

$$p(b^2|y_{1:T}, X_{1:T}, \beta) \propto p(b^2) \prod_{t=1}^{T} p(y_t|X_t\beta, b^2) \tag{A.52}$$

$$\propto \text{IG}\left(.5\left(T + \nu_0\right), .5\left((y - X\beta)'(y - X\beta) + \varsigma_0\right)\right). \tag{A.53}$$

A3 Sampling steps of MRGARCH-DPM

1. Draw the state-dependent mean vectors and covariance matrices $\mu_{1:k}, \Lambda_{1:k}|r_{1:T}, H_{1:T}, s_{1:T}$ from linear model conjugate priors with Gibbs sampling as

$$\Lambda_j|r_{1:T}, H_{1:T}, s_{1:T}, \mu_j \sim \text{IW}_n \left(\nu_0 + n_j, V_0 + \sum_{t:s_t=j} \left(H_t^{-1/2}(r_t - \mu_j) \right) \left(H_t^{-1/2}(r_t - \mu_j) \right)' \right), \tag{A.54}$$

$$\mu_j|r_{1:T}, H_{1:T}, s_{1:T}, \Lambda_j \sim \text{N}_n(\tilde{\mu}, \tilde{V}), \tag{A.55}$$

$$\text{with} \quad \tilde{V} = \left(M_0^{-1} + \sum_{t:s_t=j} (H_t^{-1/2})' \Lambda_j^{-1} H_t^{-1/2} \right)^{-1},$$

$$\tilde{\mu} = \tilde{V} \left(\sum_{t:s_t=j} (H_t^{-1/2})' \Lambda_j^{-1} H_t^{-1/2} r_t + M_0^{-1} \mu_0 \right), \; n_j = \sum_{i=1}^{T} \mathbf{1}\{s_t = j\} \text{ and } j = 1, ..., k.$$

2. Update the mixture weights in $w_{1:k}|s_{1:T}, \alpha$ with a stick-breaking process as

$$v_j|s_{1:T}, \alpha \sim \text{B} \left(1 + \sum_{t=1}^{T} \mathbf{1}\{s_t = j\}, \alpha + \sum_{t=1}^{T} \mathbf{1}\{s_t > j\} \right), \tag{A.56}$$

$$w_1 = v_1, \quad w_j = v_j \prod_{l=1}^{j-1} (1 - v_l), \quad j = 2, ..., k. \tag{A.57}$$

3. Update the slice vector $u_{1:T}|w_{1:k}, s_{1:T}$ from a uniform draw as: $u_t|w_{1:k}, s_{1:T} \sim \text{U}(0, w_{s_t})$.

4. Update the number of mixture clusters k to the smallest positive integer that satisfies: $\sum_{j=1}^{k} w_j > 1 - \min(u_{1:T})$. If new clusters are needed to satisfy the inequality, their mean and covariance are drawn from the base measure.

5. Sample the vector $s_{1:T}|r_{1:T}, H_{1:T}, \mu_{1:k}, \Lambda_{1:k}, w_{1:k}, \alpha$, from a multinominal distribution with probabilities

$$p(s_t = j|r_{1:T}, H_{1:T}, \mu_{1:k}, \Lambda_{1:k}, w_{1:k}) \propto \mathbf{1}\{u_t < w_j\} \ \mathrm{N}_n\left(r_t\middle|\mu_j, H_t^{1/2}\Lambda_j\left(H_t^{1/2}\right)'\right), \tag{A.58}$$

for $j = 1, ..., k$. The number of active clusters κ, can be calculated as the ones with at least one assigned data observation.

6. Draw the DPM precision parameter α with a gamma prior $\alpha \sim \Gamma(a_0, b_0)$ following the two steps algorithm of Escobar and West (1995):

   i. draw the random variable $\xi|\alpha, \kappa \sim \mathrm{B}(\alpha + 1, T)$.

   ii. sample α from

$$\alpha|\xi \sim \pi_\xi \Gamma(a_0 + \kappa, b_0 - \log(\xi)) + (1 - \pi_\xi)\Gamma(a_0 + \kappa - 1, b_0 - \log(\xi)),$$

   with $\frac{\pi_\xi}{1-\pi_\xi} = \frac{a_0+\kappa-1}{T(b_0-\log(\xi))}$.

7. MRGARCH parameters in $\theta = \{a, b, c, \lambda, \nu\}$ have the following conditional posterior

$$p(\theta|\mathcal{I}_T, \mu_{1:k}, \Lambda_{1:k}, s_{1:k}, V) \propto p(\theta)\prod_{t=1}^{T}\mathrm{N}_n\left(r_t\middle|\mu_{s_t}, H_t^{1/2}\Lambda_{s_t}\left(H_t^{1/2}\right)'\right) \times \mathrm{IW}_n\left(RC_t\middle|\nu, (\nu - n - 1)H_t^{1/2}V\left(H_t^{1/2}\right)'\right), \tag{A.59}$$

which does not have a standard form. A random walk MH algorithm is used to take a new draw $\theta^{(i)}$ from the above posterior, with proposal θ' which is from $h(\theta') \sim \mathrm{N}(\theta^{(i-1)}, \varsigma\hat{V}_h)$, with $\hat{V}_h$ the inverse Hessian matrix evaluated at the posterior mode $\hat{\theta}$. The draw is accepted

with probability

$$\min\left\{p(\theta^{'}|\mathcal{I}_T)/p(\theta^{(i-1)}|\mathcal{I}_T),1\right\}, \tag{A.60}$$

where $p(.|\mathcal{I}_T)$ is the posterior in (A.59). For every $\theta^{(i)}$, $\Omega^{(i)}$ is calculated as in (3.30). Only posterior draws that result in positive definite $\Omega^{(i)}$ and $H_{1:T}^{(i)}$ are accepted.

8. Update scale matrix $V|RC_{1:T}, H_{1:T}, \nu$ from a conjugate Wishart prior, with a Gibbs draw as

$$\begin{aligned} p(V|RC_{1:T}, H_{1:T}, \nu) &\propto p(V)p(RC_{1:T}|V, H_{1:T}, \nu) \\ &\sim \mathrm{W}_n(\nu_p, \Psi_p), \end{aligned}$$

with $\nu_p = \psi_0 + T\nu$ and $\Psi_p = \left[\Psi_0^{-1} + (\nu - n - 1)\sum_{t=1}^{T}\left(H_t^{1/2}\right)' RC_t^{-1}H_t^{1/2}\right]^{-1}$.

A4 Covariance targeting

Let $\bar{\mu}$ be the population mean which is estimated by the sample mean and $\bar{\Sigma}$ be the population covariance matrix estimated by the sample covariance matrix. I set $V \equiv I_n$, and assume that the conditional covariance matrix is stationary and equal to the sample covariance matrix, $\mathbb{E}(H_1) = ... = \mathbb{E}(H_T) = \bar{\Sigma}$. Instead of drawing Ω, from the posterior, I use the posterior draws (i) of the other parameters in θ to derive covariance targeting as

MRGARCH-N:

$$\begin{aligned} \mathbb{E}(H_t) &= \Omega + A \odot \mathbb{E}\left\{[(r_{t-1} - \bar{\mu}) + (\bar{\mu} - \lambda)]\,[(r_{t-1} - \bar{\mu}) + (\bar{\mu} - \lambda)]'\right\} + B \odot \mathbb{E}(H_{t-1}) + C \odot \mathbb{E}(RC_{t-1}) \\ &= \Omega + A \odot \mathbb{E}\left[(r_{t-1} - \bar{\mu})(r_{t-1} - \bar{\mu})'\right] + A \odot \mathbb{E}\left[(\bar{\mu} - \lambda)(\bar{\mu} - \lambda)'\right] + B \odot \mathbb{E}(H_{t-1}) + C \odot \mathbb{E}\left[H_t^{1/2} V \left(H_t^{1/2}\right)'\right] \\ &= \Omega + A \odot \mathbb{E}(H_{t-1}) + A \odot \mathbb{E}\left[(\bar{\mu} - \lambda)(\bar{\mu} - \lambda)'\right] + B \odot \mathbb{E}(H_{t-1}) + C \odot \mathbb{E}(H_{t-1}) \\ &\Rightarrow \Omega^{(i)} = \bar{\Sigma} \odot \left(\iota\iota' - A^{(i)} - B^{(i)} - C^{(i)}\right) - A^{(i)} \odot \left(\bar{\mu} - \lambda^{(i)}\right)\left(\bar{\mu} - \lambda^{(i)}\right)' \end{aligned} \tag{A.61}$$

MRGARCH-t:

$$\begin{aligned}
\mathbb{E}(H_t) &= \Omega + A \odot \mathbb{E}\left\{[(r_{t-1}-\bar{\mu})+(\bar{\mu}-\lambda)]\,[(r_{t-1}-\bar{\mu})+(\bar{\mu}-\lambda)]'\right\} + B \odot \mathbb{E}(H_{t-1}) + C \odot \mathbb{E}(RC_{t-1}) \\
&= \Omega + A \odot \mathbb{E}\left[(r_{t-1}-\bar{\mu})(r_{t-1}-\bar{\mu})'\right] + A \odot \mathbb{E}\left[(\bar{\mu}-\lambda)(\bar{\mu}-\lambda)'\right] + B \odot \mathbb{E}(H_{t-1}) + C \odot \mathbb{E}\left[\frac{\zeta}{\zeta-2} H_t^{1/2} V \left(H_t^{1/2}\right)'\right] \\
&= \Omega + A \odot \mathbb{E}\left(\frac{\zeta}{\zeta-2} H_{t-1}\right) + A \odot \mathbb{E}\left[(\bar{\mu}-\lambda)(\bar{\mu}-\lambda)'\right] + B \odot \mathbb{E}(H_{t-1}) + C \odot \mathbb{E}\left(\frac{\zeta}{\zeta-2} H_{t-1}\right) \\
&\Rightarrow \Omega^{(i)} = \bar{\Sigma} \odot \left(\iota\iota' - \frac{\zeta}{\zeta-2} A^{(i)} - B^{(i)} - \frac{\zeta}{\zeta-2} C^{(i)}\right) - A^{(i)} \odot \left(\bar{\mu}-\lambda^{(i)}\right)\left(\bar{\mu}-\lambda^{(i)}\right)' \qquad \text{(A.62)}
\end{aligned}$$

The DPM base measure of L is set such as $\mathbb{E}(L) = I_n$.

MRGARCH-DPM / MRGARCH-DPM-Λ :

$$\begin{aligned}
\mathbb{E}(H_t) &= \Omega + A \odot \mathbb{E}\left\{[(r_{t-1}-\bar{\mu})+(\bar{\mu}-\lambda)]\,[(r_{t-1}-\bar{\mu})+(\bar{\mu}-\lambda)]'\right\} + B \odot \mathbb{E}(H_{t-1}) + C \odot \mathbb{E}(RC_{t-1}) \\
&= \Omega + A \odot \mathbb{E}\left[H_t^{1/2} L_t \left(H_t^{1/2}\right)'\right] + A \odot \mathbb{E}\left[(\bar{\mu}-\lambda)(\bar{\mu}-\lambda)'\right] + B \odot \mathbb{E}(H_{t-1}) + C \odot \mathbb{E}\left[H_t^{1/2} V \left(H_t^{1/2}\right)'\right] \\
&= \Omega + A \odot \mathbb{E}(H_{t-1}) + A \odot \mathbb{E}\left[(\bar{\mu}-\lambda)(\bar{\mu}-\lambda)'\right] + B \odot \mathbb{E}(H_{t-1}) + C \odot \mathbb{E}(H_{t-1}) \\
&\Rightarrow \Omega^{(i)} = \bar{\Sigma} \odot \left(\iota\iota' - A^{(i)} - B^{(i)} - C^{(i)}\right) - A^{(i)} \odot \left(\bar{\mu}-\lambda^{(i)}\right)\left(\bar{\mu}-\lambda^{(i)}\right)' \qquad \text{(A.63)}
\end{aligned}$$

For the MGARCH benchmarks apply the same results, for each associated distributional assumption, with the use of $C = 0_n$.

Bibliography

Abanto-Valle, C. A., Bandyopadhyay, D., Lachos, V. H., and Enriquez, I. (2010). Robust Bayesian analysis of heavy-tailed stochastic volatility models using scale mixtures of normal distributions. *Computational Statistics & Data Analysis* 54(12), 2883–2898.

Aït-Sahalia, Y. and Mancini, L. (2008). Out of sample forecasts of quadratic variation. *Journal of Econometrics* 147(1), 17–33.

Andersen, T. G. and Bollerslev, T. (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review*, 885–905.

Andersen, T. G., Bollerslev, T., and Diebold, F. X. (2007). 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.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Ebens, H. (2001a). The distribution of realized stock return volatility. *Journal of Financial Economics* 61(1), 43–76.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2001b). The distribution of realized exchange rate volatility. *Journal of the American Statistical Association* 96(453), 42–55.

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

Andersen, T. G., Bollerslev, T., and Huang, X. (2011). A reduced form framework for modeling volatility of speculative prices based on realized variation measures. *Journal of Econometrics* 160(1), 176–189.

Archakova, I., Hansen, P. R., and Lunde, A. (2019). A Multivariate Realized GARCH Model.

Baba, Y., Engle, R. F., Kraft, D. F., and Kroner, K. F. (1990). Multivariate simultaneous generalized ARCH. *Manuscript, University of California, San Diego, Department of Economics*.

Ball, C. A. and Torous, W. N. (1983). A simplified jump process for common stock returns. *Journal of Financial and Quantitative analysis* 18(1), 53–65.

Bandi, F. M. and Russell, J. R. (2008). Microstructure noise, realized variance, and optimal sampling. *The Review of Economic Studies* 75(2), 339–369.

Banrdorff-Nielsen, O. and Shephard, N. (2004). Econometric analysis of realized covariation: high frequency covariance, regression and correlation in financial economics. *Econometrica* 72, 885–925.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (2011). Multivariate realised kernels: consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. *Journal of Econometrics* 162(2), 149–169.

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

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

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

Bauwens, L., Laurent, S., and Rombouts, J. V. (2006). Multivariate GARCH models: a survey. *Journal of Applied Econometrics* 21(1), 79–109.

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

Bollerslev, T., Engle, R. F., and Wooldridge, J. M. (1988). A capital asset pricing model with time-varying covariances. *Journal of Political Economy* 96(1), 116–131.

Bollerslev, T., Kretschmer, U., Pigorsch, C., and Tauchen, G. (2009). A discrete-time model for daily S & P500 returns and realized variations: Jumps and leverage effects. *Journal of Econometrics* 150(2), 151–166.

Caporin, M., Rossi, E., and Magistris, P. S. d. (2015). Volatility jumps and their economic determinants. *Journal of Financial Econometrics* 14(1), 29–80.

Chan, K. F. and Gray, P. (2018). Volatility jumps and macroeconomic news announcements. *Journal of Futures Markets* 38(8), 881–897.

Chan, W. H. and Maheu, J. M. (2002). Conditional jump dynamics in stock market returns. *Journal of Business & Economic Statistics* 20(3), 377–389.

Chan, W. H. and Feng, L. (2012). Time-varying jump risk premia in stock index futures returns. *Journal of Futures Markets* 32(7), 639–659.

Chen, Q., Gerlach, R., and Lu, Z. (2012). Bayesian Value-at-Risk and expected shortfall forecasting via the asymmetric Laplace distribution. *Computational Statistics & Data Analysis* 56(11), 3498–3516.

Chib, S., Nardari, F., and Shephard, N. (2002). Markov chain Monte Carlo methods for stochastic volatility models. *Journal of Econometrics* 108(2), 281–316.

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

Corsi, F., Audrino, F., and Renó, R. (2012). HAR modeling for realized volatility forecasting.

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

Delatola, E.-I. and Griffin, J. E. (2011). Bayesian nonparametric modelling of the return distribution with stochastic volatility. *Bayesian Analysis* 6(4), 901–926.

Delatola, E.-I. and Griffin, J. E. (2013). A Bayesian semiparametric model for volatility with a leverage effect. *Computational Statistics & Data Analysis* 60, 97–110.

Diamantopoulos, K. and Vrontos, I. D. (2010). A student-t full factor multivariate GARCH model. *Computational Economics* 35(1), 63.

Diks, C., Panchenko, V., and Van Dijk, D. (2011). Likelihood-based scoring rules for comparing density forecasts in tails. *Journal of Econometrics* 163(2), 215–230.

Ding, Z. and Engle, R. F. (2001). Large scale conditional covariance matrix modeling, estimation and testing.

Durham, G. B. (2006). Monte Carlo methods for estimating, smoothing, and filtering one-and two-factor stochastic volatility models. *Journal of Econometrics* 133(1), 273–305.

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

Engle, R. (2002b). New frontiers for ARCH models. *Journal of Applied Econometrics* 17(5), 425–446.

Engle, R. and Colacito, R. (2006). Testing and valuing dynamic correlations for asset allocation. *Journal of Business & Economic Statistics* 24(2), 238–253.

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

Engle, R. F. and Kroner, K. F. (1995). Multivariate simultaneous generalized ARCH. *Econometric Theory*, 122–150.

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

Eraker, B. (2004). Do stock prices and volatility jump? Reconciling evidence from spot and option prices. *The Journal of Finance* 59(3), 1367–1403.

Eraker, B., Johannes, M., and Polson, N. (2003). The impact of jumps in volatility and returns. *The Journal of Finance* 58(3), 1269–1300.

Escobar, M. D. and West, M. (1995). Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association* 90(430), 577–588.

Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, 209–230.

Fissler, T. and Ziegel, J. F. (2016). Higher order elicitability and Osband's principle. *The Annals of Statistics* 44(4), 1680–1707.

Gerlach, R. and Chen, C. W. (2015). Bayesian expected shortfall forecasting incorporating the intraday range. *Journal of Financial Econometrics* 14(1), 128–158.

Geweke, J. (1994). *Bayesian comparison of econometric models*. Tech. rep. Citeseer.

Gorgi, P., Hansen, P. R., Janus, P., and Koopman, S. J. (2019). Realized Wishart-GARCH: A score-driven multi-asset volatility model. *Journal of Financial Econometrics* 17(1), 1–32.

Hansen, P. R. and Lunde, A. (2006). Realized variance and market microstructure noise. *Journal of Business & Economic Statistics* 24(2), 127–161.

Hansen, P. R., Huang, Z., and Shek, H. H. (2012). Realized GARCH: a joint model for returns and realized measures of volatility. *Journal of Applied Econometrics* 27(6), 877–906.

Hansen, P. R., Lunde, A., and Voev, V. (2014). Realized beta GARCH: A multivariate GARCH model with realized measures of volatility. *Journal of Applied Econometrics* 29(5), 774–799.

Heber, G., Lunde, A., Shephard, N., and Sheppard, K. (2009). *Oxford-Man Institute's Realized Library, version 0.1*.

Jensen, M. J. and Maheu, J. M. (2010). Bayesian semiparametric stochastic volatility modeling. *Journal of Econometrics* 157(2), 306–316.

Jensen, M. J. and Maheu, J. M. (2013). Bayesian semiparametric multivariate GARCH modeling. *Journal of Econometrics* 176(1), 3–17.

Jensen, M. J. and Maheu, J. M. (2014). Estimating a semiparametric asymmetric stochastic volatility model with a Dirichlet process mixture. *Journal of Econometrics* 178, 523–538.

Jin, X. and Maheu, J. M. (2013). Modeling realized covariances and returns. *Journal of Financial Econometrics* 11(2), 335–369.

Jin, X. and Maheu, J. M. (2016). Bayesian semiparametric modeling of realized covariance matrices. *Journal of Econometrics* 192(1), 19–39.

Johannes, M., Kumar, R., and Polson, N. G. (1999). State dependent jump models: How do US equity indices jump. *Woking Paper, University of Chicago*.

Kalimipalli, M. and Susmel, R. (2004). Regime-switching stochastic volatility and short-term interest rates. *Journal of Empirical Finance* 11(3), 309–329.

Kalli, M., Griffin, J. E., and Walker, S. G. (2011). Slice sampling mixture models. *Statistics and Computing* 21(1), 93–105.

Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association* 90(430), 773–795.

Kim, S., Shephard, N., and Chib, S. (1998). Stochastic volatility: likelihood inference and comparison with ARCH models. *The Review of Economic Studies* 65(3), 361–393.

Li, C., Maheu, J. M., and Yang, Q. (2022). An Infinite Hidden Markov Model with Stochastic Volatility.

Liesenfeld, R. and Jung, R. C. (2000). Stochastic volatility models: conditional normality versus heavy-tailed distributions. *Journal of Applied Econometrics* 15(2), 137–160.

Liu, J. (2021). A Bayesian Semiparametric Realized Stochastic Volatility Model. *Journal of Risk and Financial Management* 14(12), 617.

Maheu, J. M. and McCurdy, T. H. (2004). News arrival, jump dynamics, and volatility components for individual stock returns. *The Journal of Finance* 59(2), 755–793.

Maheu, J. M. and McCurdy, T. H. (2007). Components of market risk and return. *Journal of Financial Econometrics* 5(4), 560–590.

Maheu, J. M. and McCurdy, T. H. (2008). Modeling foreign exchange rates with jumps. In: *Forecasting in the Presence of Structural Breaks and Model Uncertainty*. Emerald Group Publishing Limited.

Maheu, J. M. and McCurdy, T. H. (2011). Do high-frequency measures of volatility improve forecasts of return distributions? *Journal of Econometrics* 160(1), 69–76.

Maheu, J. M., McCurdy, T. H., and Zhao, X. (2013). Do jumps contribute to the dynamics of the equity premium? *Journal of Financial Economics* 110(2), 457–477.

Maheu, J. M. and Shamsi Zamenjani, A. (2021). Nonparametric dynamic conditional beta. *Journal of Financial Econometrics* 19(4), 583–613.

Mahieu, R. J. and Schotman, P. C. (1998). An empirical application of stochastic volatility models. *Journal of Applied Econometrics* 13(4), 333–360.

Nakajima, J. and Omori, Y. (2012). Stochastic volatility model with leverage and asymmetrically heavy-tailed error using GH skew Student's t-distribution. *Computational Statistics & Data Analysis* 56(11), 3690–3704.

Noureldin, D., Shephard, N., and Sheppard, K. (2012). Multivariate high-frequency-based volatility (HEAVY) models. *Journal of Applied Econometrics* 27(6), 907–933.

Oldfield Jr, G. S., Rogalski, R. J., and Jarrow, R. A. (1977). An autoregressive jump process for common stock returns. *Journal of Financial Economics* 5(3), 389–418.

Opschoor, A., Janus, P., Lucas, A., and Van Dijk, D. (2018). New HEAVY models for fat-tailed realized covariances and returns. *Journal of Business & Economic Statistics* 36(4), 643–657.

Patton, A. J., Ziegel, J. F., and Chen, R. (2019). Dynamic semiparametric models for expected shortfall (and value-at-risk). *Journal of Econometrics* 211(2), 388–413.

Press, S. J. (1967). A compound events model for security prices. *Journal of Business*, 317–335.

Richardson, M. and Smith, T. (1993). A test for multivariate normality in stock returns. *Journal of Business*, 295–321.

Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica*, 639–650.

Shirota, S., Omori, Y., Lopes, H. F., and Piao, H. (2017). Cholesky realized stochastic volatility model. *Econometrics and Statistics* 3, 34–59.

So, M. E. P., Lam, K., and Li, W. K. (1998). A stochastic volatility model with Markov switching. *Journal of Business & Economic Statistics* 16(2), 244–253.

Takahashi, M., Omori, Y., and Watanabe, T. (2009). Estimating stochastic volatility models using daily returns and realized volatility simultaneously. *Computational Statistics & Data Analysis* 53(6), 2404–2426.

Taylor, S. J. (1994). Modeling stochastic volatility: A review and comparative study. *Mathematical Finance* 4(2), 183–204.

Taylor, S. J. (1982). Financial returns modelled by the product of two stochastic processes-a study of the daily sugar prices 1961-75. *Time Series Analysis: Theory and Practice* 1, 203–226.

Vo, M. T. (2009). Regime-switching stochastic volatility: Evidence from the crude oil market. *Energy Economics* 31(5), 779–788.

Walker, S. G. (2007). Sampling the Dirichlet mixture model with slices. *Communications in Statistics — Simulation and Computation®* 36(1), 45–54.

Yamauchi, Y. and Omori, Y. (2020). Multivariate stochastic volatility model with realized volatilities and pairwise realized correlations. *Journal of Business & Economic Statistics* 38(4), 839–855.

Yu, J. (2012). A semiparametric stochastic volatility model. *Journal of Econometrics* 167(2), 473–482.

Zaharieva, M. D., Trede, M., and Wilfling, B. (2020). Bayesian semiparametric multivariate stochastic volatility with application. *Econometric Reviews* 39(9), 947–970.

Zhang, L., Mykland, P. A., and Aıt-Sahalia, Y. (2005). A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association* 100(472), 1394–1411.