Advancing Financial Market Volatility Measurement and Forecasting

ADVANCING FINANCIAL MARKET VOLATILITY MEASUREMENT AND FORECASTING

BY

JIA LIU, B.S., M.A.

A THESIS

SUBMITTED TO THE DEGROOTE SCHOOL OF BUSINESS

AND THE SCHOOL OF GRADUATE STUDIES

OF MCMASTER UNIVERSITY

IN PARTIAL FULFILMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

© Copyright by Jia Liu, June 2017

All Rights Reserved

Doctor of Philosophy (2017) McMaster University

(Business Administration - Finance) Hamilton, Ontario, Canada

TITLE:	Advancing Financial Market Volatility Measurement and Forecasting
AUTHOR:	Jia Liu B.Sc., (Tianjin University of Finance and Economics) M.A., (University of South Florida)
SUPERVISORS:	Dr. John Maheu Dr. Ronald Balvers Dr. Peter Miu
NUMBER OF PAGES:	xv, 191

Abstract

This thesis proposes several new and useful financial econometric tools to facilitate risk analysis, portfolio choice and forecasting.

This thesis starts with an introduction in Chapter 1. The research background, motivation and the structure of the thesis are illustrated in this chapter.

Chapter 2 proposes a class of models that jointly model returns and ex-post variance measures under a Markov switching framework. Both univariate and multivariate return versions of the model are introduced. Estimation can be conducted under a fixed dimension state space or an infinite one. The proposed models can be seen as nonlinear common factor models subject to Markov switching and are able to exploit the information content in both returns and ex-post volatility measures. Applications to equity returns compare the proposed models to existing alternatives. The empirical results show that the joint models improve density forecasts for returns and point predictions of return variance. Using the information in ex-post volatility measures can increase the precision of parameter estimates, sharpen the inference on the latent state variable and improve portfolio decisions.

Chapter 3 offers a new exact finite sample approach to estimating ex-post variance using Bayesian nonparametric methods. Until now ex-post variance estimation has been based on infill asymptotic assumptions that exploit high-frequency data. In

contrast to the classical counterpart, the proposed method exploits pooling over high-frequency observations with similar variances. Bayesian nonparametric variance estimators under no noise, heteroskedastic and serially correlated microstructure noise are introduced and discussed. Monte Carlo simulation results show that the proposed approach can increase the accuracy of variance estimation. Applications to equity data and comparison with realized variance and realized kernel estimators are included.

Chapter 4 extends the third chapter to estimate the ex-post covariance matrix of asset returns from high-frequency data. As before, pooling is used to improve estimation accuracy and the method does not rely on infill asymptotic assumptions. In addition, the proposed covariance estimator is guaranteed to be positive definite. Furthermore, a new synchronization method of observations based on data augmentation is introduced. The Bayesian estimator is made robust to independent microstructure noise and nonsynchronous trading. According to Monte Carlo simulations, the new estimator is very competitive with existing estimators. Empirical applications evaluate the new estimator from an economic perspective.

Finally, Chapter 5 concludes and summarizes the contribution of this thesis to the literature.

Acknowledgements

I would like to express my deepest gratitude to my Ph.D. supervisor, Dr. John Maheu for his guidance, support and encouragement during my Ph.D. journey. Dr. Maheu is a very knowledgeable and experienced researcher and a very patient mentor who really cares about his students. He not only helps me to overcome the challenges I faced in my Ph.D. research, but also guides me step by step to build my research, teaching and presentation skills. Dr. Maheu is very nice and approachable. It is very enjoyable to talking with him about either research topics or personal life. It has been my fortune to have Dr. Maheu as my supervisor. Without his guidance and persistent help, this thesis would not have been completed and I may not be at the same position as I am today.

I am also very grateful to my other supervisory committee members, Dr. Ronald Balvers and Dr. Peter Miu for their never-ending support. Their valuable suggestions significantly improve this thesis. Dr. Balvers and Dr. Miu also offer me lots of helpful advice on job search. In addition, I really appreciate all the teaching and presentation opportunities that Dr. Balvers provided to me.

Furthermore, I extend special thanks to Dr. Jiaping Qiu, Dr. Narat Charupat and Dr. Richard Deaves. I have learned a lot through working with them on research projects.

I also want to thank my Ph.D. colleagues, in particular Arthur Luo, Donghui Chen, Azam Shamsi, Robin Li, Adam Stivers and Yuchuan Jin, for their support. I am also thankful to those at Ph.D. office, especially Deb Randall Baldry.

Finally, I would like to thank my parents, Xiaochang Liu and Yan Wang for their endless love and support in every stage of my life. I owe special thanks to my wife, Beibei Liu. She sacrificed a lot to support my academic path. Her love and support make my life and Ph.D. journey more enjoyable and colourful.

Chapter 2 is based on joint work with John Maheu and I am the main contributor. Chapter 3 is based on joint work with John Maheu and Jim Griffin and I am the main contributor.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

This thesis proposes new and improved approaches of estimating volatility and covariation of asset returns from high frequency data and new joint Markov switching models with ex-post risk measures. This chapter will highlight the research background, motivation, contribution and structure of this thesis.

Chapter 2 proposes a way of jointly modelling asset returns and ex-post volatility measures under a Markov switching framework. In the recent decades, the estimation of volatility has benefited from the availability of high frequency data and financial econometricians have proposed model-free estimators for ex-post volatility. The most famous examples are the realized variance by Andersen *et al.* (2001) and Barndorff-Nielsen and Shephard (2002) and its multivariate extension: the realized covariance by Barndorff-Nielsen and Shephard (2004). Those ex-post volatility measures provide "observable" data for latent volatility, thereby allowing the analysis of volatility data directly.

Another important econometric tool is the Markov switching model. Markov switching models and its variations have been widely applied in the field of finance

and economics to analyze stock returns, foreign exchange rates, interest rates and other macroeconomic variables. Existing Markov switching models only use one series of data, such as asset returns, as the data input. It is worth investigating if the additional information contained in ex-post variance measures provides additional benefits to Markov switching models.

Chapter 2 contributes to the literature by integrating Markov switching models with ex-post volatility measures. A class of new Markov switching models that can jointly model both returns and ex-post volatility measures are proposed. The empirical applications show that the proposed joint models outperform the benchmark model in density forecasts of returns, variance prediction and portfolio choice. The parameter estimation and the identification of latent state variables are improved potentially.

Chapter 2 applies existing methods of computing ex-post risk measures and show those risk measures are beneficial to financial and economics data modelling. The estimation of ex-post volatility is a very active area of research. Existing methods rely on infill asymptotic assumptions and require data frequency to be high enough to get accurate estimates. In other words, the existing estimator works well theoretically but may not be very accurate in reality, which leaves room for improvement. Chapter 3 and Chapter 4 takes a completely new approach and apply Bayesian financial econometric tools to develop new and improved ex-post volatility estimators.

With the help of Bayesian nonparametric tools, Chapter 3 proposes an ex-post variance method that delivers exact finite sample results. Another benefit is the proposed method allows pooling in the estimation of ex-post variance. Return observations with similar variance parameter can be pooled automatically by the model for

the purpose of increasing estimation accuracy. The Bayesian nonparametric variance estimation approach is also extended to fit the data with independent and dependent microstructure noise.

Based on simulation evidence, the Bayesian nonparametric variance estimator has lower error in estimating integrated variance and better finite sample results, especially when the data frequency is low. Empirical applications show that the forecast of future volatility can be potentially improved, if the volatility measures are estimated using the proposed method.

In Chapter 4, the univariate Bayesian nonparametric variance estimator introduced in Chapter 3 is extended to its multivariate version to allow pooling in covariance estimation. The estimation of covariance matrix is an important problem since covariance serves as the key input to many problems in finance and economics. In practice, the estimation of the covariance matrix suffers from two major challenges, which are microstructure noise and nonsynchronous trading. Existing works have proposed different ways of forming covariance estimators and several time schemes to synchronize irregularly-spaced observations.

As in the univariate case, the Bayesian nonparametric approach of estimating the ex-post covariance matrix allows pooling in covariance estimation and delivers exact finite sample results. Another improvement is that the proposed estimator is guaranteed to be positive definite, which is a desirable property. More than that, a new synchronization method is proposed to deal with non-synchronous trading. All of those benefits lead to a more accurate estimator, which is confirmed by Monte Carlo simulation results. In real data applications, the Bayesian nonparametric estimator results in better portfolio choice outcomes.

Chapter 3 and Chapter 4 contribute to the literature by providing the first exact finite sample approach to ex-post volatility estimation. By linking the Bayesian non-parametric analytical tools with the estimation of ex-post volatility, new versions of ex-post variance and covariance estimators are introduced. The proposed methods deliver improved estimators and provide the entire posterior distributions of volatility, which benefit risk measuring, analysis and forecasting.

Bibliography

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

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

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

Chapter 2

Improving Markov Switching Models Using Realized Variance

2.1 Introduction

This chapter proposes a new way of jointly modelling return and ex-post volatility measures under a Markov switching framework. Parametric and nonparametric versions of the models are introduced in both univariate and multivariate settings. The models are able to exploit the information content in both return and ex-post volatility series. Compared to existing models, the joint models improve density forecasts of returns, point predictions of realized variance and portfolio decisions.

Since the pioneering work by Hamilton (1989) the Markov switching model has become one of the standard econometric tools in studying various financial and economic data series. The basic model postulates a discrete latent variable governed by a first-order Markov chain that directs an observable data series. This modelling approach has been fruitfully employed in many applications. For instance, Markov

switching models have been used to identify bull and bear markets in aggregate stock returns (Maheu and McCurdy, 2000; Lunde and Timmermann, 2004; Guidolin and Timmermann, 2006; Maheu *et al.*, 2012), to capture the risk and return relationship (Pastor and Stambaugh, 2001; Kim *et al.*, 2004), for portfolio choice (Guidolin and Timmermann, 2008), for interest rates (Ang and Bekaert, 2002; Guidolin and Timmermann, 2009) and foreign exchange rates (Engel and Hamilton, 1990; Dueker and Neely, 2007). Recent work has extended the Markov switching model to an infinite dimension. The infinite hidden Markov model (IHMM), which is a Bayesian nonparametric model, allows for a very flexible conditional distribution that can change over time. Applications of IHMM include Jochmann (2015), Dufays (2016), Song (2014), Carpantier and Dufays (2014) and Maheu and Yang (2016).

Realized variance (RV), constructed from intraperiod returns, is an accurate measure of ex-post volatility. Andersen *et al.* (2001) and Barndorff-Nielsen and Shephard (2002) formalized the idea of using higher frequency data to measure the volatility of lower frequency data and show RV is a consistent estimate of quadratic variation under ideal conditions. Barndorff-Nielsen and Shephard (2004b) generalized the idea of RV and introduced a set of variance estimators called realized power variations. Furthermore, RV has been extended to realized covariance (RCOV), which is an ex-post nonparametric measure of the covariance of multivariate returns, by Barndorff-Nielsen and Shephard (2004a). A survey of RV and related volatility proxies is Andersen and Benzoni (2009).

This chapter is the first to exploit the information content of ex-post volatility measures in a Markov switching context to improve estimation, forecasting and portfolio decisions.[1] This is done for finite and infinite state models. We assume that

[1]Other papers that have used high frequency data to improve estimation and forecasting include

regime changes in returns and realized variance are directed by one common unobserved state variable. Closely related to our approach is Takahashi *et al.* (2009) who propose a stochastic volatility model in which unobserved log-volatility affects both RV and the variance of returns. They find improved fixed parameter and latent volatility estimates but do not investigate forecast performance or portfolio choice.

There is no reason to confine attention to RV, and therefore we investigate the use of other volatility measures and in the multivariate setting realized covariance. Four versions of the univariate return models are proposed. We consider RV, log(RV), realized absolute variation (RAV), or log(RAV) as ex-post volatility measures coupled with returns to construct joint models. We then extend the MS-RV specification to its multivariate version with RCOV.

It is more flexible to drop the finite state assumption and let the data determine the number of states needed to fit the data. Using Bayesian nonparametric techniques, we extend the finite state joint MS models to nonparametric versions. These models allow the conditional distribution to change more flexibly and accommodate any nonparametric relationship between returns and ex-post volatility.

Markov switching models have been particularly useful at monthly and quarterly frequencies as the Markov chain dynamics are a dominant feature of the data.[2] Therefore, the proposed joint MS and joint IHMM models are compared to existing models in empirical applications to monthly U.S. stock market returns including forecasting and portfolio applications.

Based on the log-predictive Bayes factors, the proposed joint models strongly

Alizadeh *et al.* (2002), Blair *et al.* (2001), Shephard and Sheppard (2010), Noureldin *et al.* (2012), Hansen *et al.* (2012) Hansen *et al.* (2014), Jin and Maheu (2013, 2016) and Maheu and McCurdy (2011).

[2]Our models could be applied to higher frequency data but in this case additional structure would be needed to capture the strong persistence in high frequency volatility Rydén *et al.* (1998).

dominate the models that only use returns. Moreover, we find the gains from joint modelling are particularly large during high volatility episodes. The empirical results also show that the joint models reduce the error in predicting realized variance. With the help of additional information offered by RV, RAV and RCOV, posterior density intervals for parameters are tighter and the inference on the unobservable state variables is improved. In general, adding RV to a model improves the forecast performance of any MS model, finite or infinite, but the best performing models are the joint infinite hidden Markov models.

Exploiting measures of ex-post volatility also lead to better portfolio choice outcomes. Several portfolio exercises that use the models are considered over two sample periods. A robust result is larger Sharpe ratios for models that incorporate realized variance or realized covariance. In addition, investors are always willing to pay a positive performance fee to obtain forecasts from these models for their investment decisions.

This chapter is organized as follows. In Section 2.2, we show how to incorporate ex-post measures of volatility into Markov switching models. The joint MS models are extended to the nonparametric versions in Section 2.3. Benchmark models used for comparison are found in Section 2.4. Section 2.5 illustrates the Bayesian estimation steps and model comparison. A univariate return application to the market index is found in Section 2.6 while a multivariate return application to equity follows in Section 2.7. The next section concludes followed by an appendix that gives detailed steps of posterior simulation.

2.2 Joint Markov Switching Models

In this section, we will focus on simple specifications of the conditional mean but dynamic models with lags of the dependent variables could be used. We will first discuss the four versions of univariate return joint models, then introduce the multivariate version.

Higher frequency data is used to construct ex-post volatility measures. Let $r_{t,i}$ denote the i^{th} intraperiod continuously compounded return in period t, $i = 1, \ldots, n_t$, where n_t is the number of intraperiod returns. Then the return and realized variance from $t-1$ to t is

$$r_t = \sum_{i=1}^{n_t} r_{t,i}, \tag{2.1}$$

$$RV_t = \sum_{i=1}^{n_t} r_{t,i}^2. \tag{2.2}$$

Andersen *et al.* (2001) and Barndorff-Nielsen and Shephard (2002) formalized the idea of using higher frequency data to measure the volatility of r_t. They show that RV_t is a consistent estimate of quadratic variation under ideal conditions.[3] Similarly, for multivariate returns $R_{t,i}$ is the i^{th} intraperiod $d \times 1$ return vector at time t and the time t return is $R_t = \sum_{i=1}^{n_t} R_{t,i}$. $RCOV_t$ denotes the associated realized covariance (RCOV) matrix which is computed as follows,

$$RCOV_t = \sum_{i=1}^{n_t} R_{t,i} R_{t,i}^{'}. \tag{2.3}$$

[3]We have not made adjustments for market microstructure dynamics since our high-frequency data consists of daily returns and are relatively clean. Nevertheless, any of the existing approaches that correct for microstructure dynamics in computing ex-post volatility measures could be used.

2.2.1 MS-RV Model

We first use RV as the proxy for ex-post volatility to build a joint MS-RV model. The proposed K-state MS-RV model is given as follows.

$$r_t\big|s_t \sim \mathrm{N}(\mu_{s_t}, \sigma^2_{s_t}), \tag{2.4}$$

$$RV_t\big|s_t \sim \mathrm{IG}(\nu+1, \nu\sigma^2_{s_t}), \tag{2.5}$$

$$P_{i,j} = p(s_{t+1}=j|s_t=i), \tag{2.6}$$

where $s_t \in \{1,\ldots,K\}$. Conditional on state s_t, RV_t is assumed to follow an inverse-gamma distribution[4] $\mathrm{IG}(\nu+1, \nu\sigma^2_{s_t})$, where $\nu+1$ is the shape parameter and $\nu\sigma^2_{s_t}$ is the scale parameter.

The basic assumption of this model is that RV_t is subject to the same regime changes as r_t and share the same parameter $\sigma^2_{s_t}$.[5] Note, that RV_t and the other volatility measures used in this chapter, are assumed to be a noisy measure of the state dependent variance $\sigma^2_{s_t}$. Conditional on the latent state, the mean and variance of RV_t are

$$E(RV_t|s_t) = \frac{\nu\sigma^2_{s_t}}{(\nu+1)-1} = \sigma^2_{s_t}, \tag{2.7}$$

$$\mathrm{Var}(RV_t|s_t) = \frac{(\sigma^2_{s_t})^2}{(\nu-1)}. \tag{2.8}$$

[4]If $x \sim \mathrm{IG}(\alpha,\beta)$, $\alpha>0$, $\beta>0$ then it has density function:

$$g(x|\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-\alpha-1}\exp\left(-\frac{\beta}{x}\right).$$

The mean of x is $E(x)=\frac{\beta}{\alpha-1}$ for $\alpha>1$ and $\mathrm{Var}(x)=\frac{\beta^2}{(\alpha-1)^2(\alpha-2)}$ for $\alpha>2$.

[5]Formally, the high frequency data generating process is assumed to be $r_{t,i} = \mu_{s_t}/n_t + (\sigma_{s_t}/\sqrt{n_t})z_{t,i}$, with $z_{t,i} \sim \mathrm{NID}(0,1)$. Then $E\left[\sum_{i=1}^{n_t} r^2_{t,i}|s_t\right] = \mu^2_{s_t}/n_t + \sigma^2_{s_t} \approx \sigma^2_{s_t}$ when the term $\mu^2_{s_t}/n_t$ is small due to n_t being large.

Therefore RV_t is centered around $\sigma^2_{s_t}$, but in general, not equal to it. The variance of the distribution of RV_t is positively correlated with the realized variance itself. During high volatility periods, the movements of realized variances are more volatile. Both the return process and realized variance process are governed by the same underlying Markov chain with transition matrix P.

Since $\sigma^2_{s_t}$ influences both the return process and RV_t process, the model can be seen as a nonlinear common factor model.[6] Exploiting the information content of RV_t for $\sigma^2_{s_t}$ can lead to more precise estimates of model parameters, state variables and forecasts.

2.2.2 MS-logRV Model

Another possibility is to model the logarithm of RV as normally distributed. The MS-logRV model is shown as follows,

$$r_t \big| s_t \sim \mathrm{N}\left(\mu_{s_t}, \exp(\zeta_{s_t})\right), \tag{2.9}$$

$$\log(RV_t) \big| s_t \sim \mathrm{N}\left(\zeta_{s_t} - \frac{1}{2}\delta^2_{s_t}, \delta^2_{s_t}\right), \tag{2.10}$$

$$P_{i,j} = p(s_{t+1} = j | s_t = i), \tag{2.11}$$

where $s_t \in \{1, \ldots, K\}$. In this model there are three state-dependent parameters: μ_{s_t}, ζ_{s_t} and $\delta^2_{s_t}$, which enable both the mean and variance of returns and $\log(RV_t)$ to be state-dependent. $\zeta_{s_t} - \frac{1}{2}\delta^2_{s_t}$ is the mean of $\log(RV_t)$ and $\exp(\zeta_{s_t})$ is the variance of returns. Since RV_t is log-normal, $E[RV_t|s_t] = \exp(\zeta_{s_t})$ which is assumed to be the variance of returns.

[6] This is in contrast to linear dynamic factor models such as Forni and Reichlin (1998), Kose *et al.* (2003) and Stock and Watson (2010).

2.2.3 MS-RAV Model

Now we consider using realized absolute variation (RAV), instead of RV in the joint MS model. Calculated using the absolute values of intraperiod returns, RAV is robust to jumps and may be less sensitive to outliers Barndorff-Nielsen and Shephard (2004b). RAV_t is computed using intraperiod returns as

$$RAV_t = \sqrt{\frac{\pi}{2}}\sqrt{\frac{1}{n_t}}\sum_{i=1}^{n_t}|r_{t,i}|, \tag{2.12}$$

where $r_{t,i}$ denotes the i^{th} intraperiod log-return in period t, $i = 1, \ldots, n_t$. It can be shown that RAV_t provides an estimate of the standard deviation of r_t.[7]

Consistent with the inverse-gamma distribution to model the variance or its proxy, we assume RAV follows a square-root inverse-gamma distribution (sqrt-IG). The density function of sqrt-IG(α, β) is given by

$$f(x) = \frac{2\beta^{\alpha}}{\Gamma(\alpha)}x^{-2\alpha-1}\exp\left(-\frac{\beta}{x^2}\right),\ x > 0, \alpha > 0, \beta > 0, \tag{2.13}$$

and the first and second moments of sqrt-IG(α, β) are given as follows

$$E[x] = \sqrt{\beta}\cdot\frac{\Gamma(\alpha - \frac{1}{2})}{\Gamma(\alpha)} \text{ and } E[x^2] = \frac{\beta}{\alpha - 1},\ \text{for } \alpha > 1. \tag{2.14}$$

These results can be found in Zellner (1971).

[7] As before, if $r_{t,i} = \mu_{s_t}/n_t + (\sigma_{s_t}/\sqrt{n_t})z_{t,i}$, with $z_{t,i} \sim \text{NID}(0,1)$ and μ_{s_t}/n_t is small, then we have $E\left[\sqrt{\frac{\pi}{2}}\sqrt{\frac{1}{n_t}}\sum_{i=1}^{n_t}|r_{t,i}||s_t\right] \approx \sqrt{\frac{\pi}{2n_t}}\frac{1}{\sqrt{n_t}}\sum_{i=1}^{n_t}E\left[|\sigma_{s_t}z_{t,i}||s_t\right] = \sqrt{\frac{\pi}{2}}\frac{\sigma_{s_t}}{n_t}\sum_{i=1}^{n_t}E\left[|z_{t,i}||s_t\right] = \sigma_{s_t}$, since for a standard normal variate x, $E\left[|x|\right] = \sqrt{\frac{2}{\pi}}$.

We define the joint MS model of return and RAV as,

$$
\begin{aligned}
r_t\big|s_t &\sim \mathrm{N}(\mu_{s_t}, \sigma^2_{s_t}), && (2.15)\\
RAV_t\big|s_t &\sim \text{sqrt-IG}\left(\nu, \sigma^2_{s_t}\left[\frac{\Gamma(\nu)}{\Gamma(\nu-\frac{1}{2})}\right]^2\right), && (2.16)\\
P_{i,j} &= p(s_{t+1}=j|s_t=i), && (2.17)
\end{aligned}
$$

where $s_t \in \{1,\ldots,K\}$ and $\nu > 1$. As in the MS-RV model, the mean and variance of both r_t and RAV_t are state-dependent. In each state, the return follows a normal distribution with mean μ_{s_t} and variance $\sigma^2_{s_t}$. The mean and variance of RAV_t conditional on state s_t are given as follows.

$$
\begin{aligned}
E(RAV_t\big|s_t) &= \sigma_{s_t}, && (2.18)\\
\mathrm{Var}(RAV_t\big|s_t) &= \frac{\sigma^2_{s_t}}{\nu-1}\left[\frac{\Gamma(\nu)}{\Gamma(\nu-\frac{1}{2})}\right]^2 - \sigma^2_{s_t}. && (2.19)
\end{aligned}
$$

2.2.4 MS-logRAV Model

Similar to the MS-logRV model discussed in Section 2.2.2, the logarithm of RAV can be modelled as opposed to RAV. The MS-logRAV specification is

$$
\begin{aligned}
r_t\big|s_t &\sim \mathrm{N}\left(\mu_{s_t}, \exp(2\zeta_{s_t})\right), && (2.20)\\
\log(RAV_t)\big|s_t &\sim \mathrm{N}\left(\zeta_{s_t}-\frac{1}{2}\delta^2_{s_t}, \delta^2_{s_t}\right), && (2.21)\\
P_{i,j} &= p(s_{t+1}=j|s_t=i), && (2.22)
\end{aligned}
$$

where $s_t \in \{1,\ldots,K\}$. The model is close to the MS-logRV parametrization, but now $\zeta_{s_t}-\frac{1}{2}\delta^2_{s_t}$ is the mean of $\log(RAV_t)$ and $\exp(2\zeta_{s_t})$ is the state-dependent variance

of returns. Since RAV_t is log-normal, $E(RAV_t|s_t) = \exp(\zeta_{s_t})$ which is the standard deviation of returns.

2.2.5 Multivariate MS-RCOV Model

The univariate return models can be extended to the multivariate setting by including realized covariance matrices. The multivariate MS-RCOV model we consider is

$$R_t\big|s_t \sim \mathrm{N}\left(M_{s_t}, \Sigma_{s_t}\right), \tag{2.23}$$

$$RCOV_t\big|s_t \sim \mathrm{IW}\left(\Sigma_{s_t}(\kappa - d - 1), \kappa\right), \kappa > d + 1, \tag{2.24}$$

$$P_{i,j} = p(s_{t+1} = j|s_t = i). \tag{2.25}$$

where $s_t \in \{1, \ldots, K\}$. M_{s_t} is a $d \times 1$ state-dependent mean vector and Σ_{s_t} is the $d \times d$ covariance matrix. $RCOV_t$ is assumed to follow an inverse-Wishart distribution $\mathrm{IW}(\Sigma_{s_t}(\kappa - d - 1),\ \kappa)$, where $\Sigma_{s_t}(\kappa - d - 1)$ is the scale matrix and κ is the degree of freedom.

Σ_{s_t} is the covariance of returns as well as the mean of $RCOV_t$ since

$$E[RCOV_t\big|s_t] = \frac{1}{\kappa - d - 1}\Sigma_{s_t}(\kappa - d - 1) = \Sigma_{s_t}, \tag{2.26}$$

assuming $\kappa > d + 1$. The parameter κ controls the variation of the inverse-Wishart distribution and the smaller κ is, the larger spread the distribution has. Both R_t and $RCOV_t$ are governed by the same Markov chain with transition matrix P.

2.3 Joint Infinite Hidden Markov Model

2.3.1 Dirichlet Process and Hierarchical Dirichlet Process

All of the Markov switching models we have discussed require the econometrician to set the number of states. An alternative is to incorporate the state dimension into estimation. The Bayesian nonparametric version of the Markov switching model is the infinite hidden Markov model, which can be seen as a Markov switching model with infinitely many states. Given a finite dataset, the model selects a finite number of states for the system. Since the number of states is no longer a fixed value, the Dirichlet process, an infinite dimensional version of Dirichlet distribution, is used as a prior for the transition probabilities.

There are several benefits to using an infinite Markov chain. First, this flexible structure can accommodate both recurring regimes as well as structural changes. As new data arrives if a new state of the market occurs the model can introduce a new state and associated parameter to account for this. Since the model is applied to a finite dataset a finite set of states will be used and this is estimated along with the rest of the parameters. Since as many states can be used as needed the model is nonparametric in nature and able to capture an unknown continuous density that changes over time.

The Dirichlet process $\mathrm{DP}(\alpha, H)$, was formally introduced by Ferguson (1973) and is a distribution of distributions. A draw from a $\mathrm{DP}(\alpha, H)$ is a distribution and is almost surely discrete and centered around the base distribution H. $\alpha > 0$ is the concentration parameter that governs how close the draw is to H.

We follow Teh *et al.* (2006) and build an infinite hidden Markov model (IHMM)

using a hierarchical Dirichlet process (HDP). This consists of two linked Dirichlet processes. A single draw of a distribution is taken from the top level Dirichlet processes with base measure H and precision parameter η. Subsequent to this, each row of the transition matrix is distributed according to a Dirichlet processes with base measure taken from the top level draw. This ensures that each row of the transition matrix governs the moves among a common set of model parameters. In addition, each row of the transition matrix is centered around the top level draw but any particular draw will differ. If Γ denotes the top level draw and P_j the j^{th} row of the transition matrix P then the previous discussion can be summarized as

$$\Gamma\big|\eta \sim \text{DP}(\eta, H), \tag{2.27}$$

$$P_j\big|\alpha, \Gamma \overset{iid}{\sim} \text{DP}(\alpha, \Gamma),\ j = 1, 2, ... \tag{2.28}$$

This formulation provides a prior over the natural numbers (states) such that each P_j has $E[P_j] = \Gamma$.[8] For more details and examples of the HDP used in the infinite hidden Markov model see Maheu and Yang (2016).

Combining the HDP, (2.27)-(2.28), with the state indicator s_t and the data density, forms the infinite hidden Markov model,

$$s_t\big|s_{t-1}, P \sim P_{s_{t-1}}, \tag{2.29}$$

$$\theta_j \overset{iid}{\sim} H,\ j = 1, 2, ..., \tag{2.30}$$

$$y_t\big|s_t, \theta \sim F(y_t\big|\theta_{s_t}), \tag{2.31}$$

[8]To make this concrete, consider the following simple example. Abstracting from an infinite chain to a three dimensional one, suppose $\Gamma = (0.51, 0.32, 0.17)$. Each row of the transition matrix is obtained as $P_j \sim \text{Dir}(\Gamma)$ where $\text{Dir}(\Gamma)$ is the Dirichlet distribution with parameter Γ and $E[P_j] = \Gamma$. Then, for instance, the rows sampled could be $P_1 = (0.55, 0.2, 0.25)$, $P_2 = (0.42, 0.41, 0.17)$ and $P_3 = (0.49, 0.30, 0.21)$.

where $\theta = \{\theta_1, \theta_2, ...\}$ and $F(\cdot|\cdot)$ is the data distribution. The two concentration parameters η and α control the number of active states in the model. Larger values favour more states while small values promote a parsimonious state space. Rather than set these hyperparameters they can be treated as parameters and estimated from the data. In this case, the hierarchical prior for η and α are

$$\eta \sim \mathrm{G}(a_\eta, b_\eta), \tag{2.32}$$

$$\alpha \sim \mathrm{G}(a_\alpha, b_\alpha), \tag{2.33}$$

where $\mathrm{G}(a, b)$ stands for the gamma distribution with shape parameter a and rate parameter b.

The models can be estimated with MCMC methods. We discuss the specific details below for each model.

2.3.2 IHMM with RV and RAV

RV or RAV can be jointly modelled in the IHMM model as we did in the finite Markov switching models. The joint IHMM is constructed by replacing the Dirichlet distributed prior of the MS model by a hierarchical Dirichlet process. Hierarchical priors are used for concentration parameter α and η and allow the data to influence the state dimension. For example, the IHMM-RV model is given as (2.27)-(2.28),

(2.32)-(2.33) with

$$
\begin{aligned}
s_t \big| s_{t-1}, P &\sim P_{s_{t-1}}, && (2.34)\\
\theta_j &= \{\mu_j, \sigma_j^2\} \overset{iid}{\sim} H,\ j = 1, 2, ..., && (2.35)\\
r_t \big| s_t, \theta &\sim \mathrm{N}(\mu_{s_t}, \sigma_{s_t}^2), && (2.36)\\
RV_t \big| s_t &\sim \mathrm{IG}(\nu + 1, \nu\sigma_{s_t}^2). && (2.37)
\end{aligned}
$$

The base distribution is $H(\mu) \equiv \mathrm{N}(m, v^2), H(\sigma^2) \equiv \mathrm{IG}(v_0, s_0)$. The parameter $\sigma_{s_t}^2$ is common to the distribution of r_t and RV_t.

The IHMM-logRV and IHMM-logRAV models are formed similarly by replacing the fixed dimension transition matrix with infinite dimensional versions with a HDP prior. For instance, the IHMM-logRV specification replaces (2.35)-(2.37) with

$$
\begin{aligned}
\theta_j &= \{\mu_j, \zeta_j, \delta_j^2\} \overset{iid}{\sim} H,\ j = 1, 2, ..., && (2.38)\\
r_t \big| s_t, \theta &\sim \mathrm{N}\left(\mu_{s_t}, \exp(\zeta_{s_t})\right), && (2.39)\\
\log(RV_t) \big| s_t &\sim \mathrm{N}\left(\zeta_{s_t} - \frac{1}{2}\delta_{s_t}^2, \delta_{s_t}^2\right). && (2.40)
\end{aligned}
$$

The base distribution is $H(\mu) \equiv \mathrm{N}(m_\mu, v_\mu^2), H(\zeta) \equiv \mathrm{N}(m_\zeta, v_\zeta^2)$ and $H(\delta) \sim \mathrm{IG}(v_0, s_0)$. The parameter ζ_{s_t} is common to the distribution of r_t and $\log(RV_t)$. Similarly, the IHMM-logRAV model, replaces (2.35)-(2.37) with

$$
\begin{aligned}
\theta_j &= \{\mu_j, \zeta_j, \delta_j^2\} \overset{iid}{\sim} H, j = 1, 2, ..., && (2.41)\\
r_t \big| s_t, \theta &\sim \mathrm{N}\left(\mu_{s_t}, \exp(2\zeta_{s_t})\right), && (2.42)\\
\log(RAV_t) \big| s_t &\sim \mathrm{N}\left(\zeta_{s_t} - \frac{1}{2}\delta_{s_t}^2, \delta_{s_t}^2\right). && (2.43)
\end{aligned}
$$

The base distribution is $H(\mu) \equiv \mathrm{N}(m_\mu, v_\mu^2), H(\zeta) \equiv \mathrm{N}(m_\zeta, v_\zeta^2)$ and $H(\delta) \sim \mathrm{IG}(v_0, s_0)$. Now, ζ_{s_t} affects both r_t and $\log(RAV_t)$.

2.3.3 Multivariate IHMM with RCOV

The multivariate MS-RCOV model can be extended to its nonparametric version, labelled IHMM-RCOV as follows. In this model (2.27)-(2.28), (2.32)-(2.33) is combined with

$$
\begin{aligned}
s_t \big| s_{t-1}, P &\sim P_{s_t-1}, && (2.44)\\
\theta_j = \{M_j, \Sigma_j\} &\overset{iid}{\sim} H,\ j = 1, 2, ..., && (2.45)\\
R_t \big| s_t, \theta &\sim \mathrm{N}(M_{s_t}, \Sigma_{s_t}), && (2.46)\\
RCOV_t \big| s_t &\sim \mathrm{IW}(\Sigma_{s_t}(\kappa - d - 1), \kappa). && (2.47)
\end{aligned}
$$

The base distribution is $H(M) \equiv \mathrm{N}(m, V), H(\Sigma) \equiv \mathrm{W}(\Psi, \tau)$, where $\mathrm{W}(\Psi, \tau)$ denotes Wishart distribution, Ψ are $d \times d$ positive definite matrices and $\kappa > d + 1$ being the degree of freedom. Σ_{s_t} is a common parameter affecting the distributions of R_t and $RCOV_t$.

2.4 Benchmark Models

Each of the new models are compared to benchmark models that do not use ex-post variance measures. The benchmark specifications are essentially the same model with RV_t or RAV_t omitted. For example, in the univariate application we compare to the

following MS specification.

$$r_t|s_t \sim \mathrm{N}(\mu_{s_t}, \sigma^2_{s_t}), \tag{2.48}$$

$$P_{i,j} = p(s_{t+1} = j|s_t = i), \tag{2.49}$$

where $s_t \in \{1, \ldots, K\}$. The IHMM comparison model is given with (2.27)-(2.33) with (2.30) and (2.31) specified as

$$\theta_j = \{\mu_j, \sigma^2_j\} \overset{iid}{\sim} H, \ j = 1, 2, ..., \tag{2.50}$$

$$r_t|s_t, \theta \sim \mathrm{N}(\mu_{s_t}, \sigma^2_{s_t}). \tag{2.51}$$

The benchmark models for the multivariate application are similarly derived by omitting $RCOV_t$.

2.5 Estimation and Model Comparison

For notation, let $r_{1:t} = \{r_1, \ldots, r_t\}$, $RV_{1:t} = \{RV_1, \ldots, RV_t\}$, $y_{1:t} = \{y_1, \ldots, y_t\}$ where $y_t = \{r_t, RV_t\}$. We further define $R_{1:T} = \{R_1, \ldots, R_T\}$, $RCOV_{1:T} = \{RCOV_1, \ldots, RCOV_T\}$ and $Y_{1:t} = \{Y_1, \ldots, Y_t\}$ where $Y_t = \{R_t, RCOV_t\}$.

2.5.1 Estimation of Joint Finite MS Models

The joint finite MS models are estimated using Bayesian inference. Taking the MS-RV model as an example, model parameters include $\theta = \{\mu_j, \sigma^2_j\}_{j=1}^K$, $\phi = \{\nu\}$ and transition matrix P. By augmenting the latent state variable $s_{1:T} = \{s_1, s_2, \cdots, s_T\}$, MCMC methods are used to simulate from the conditional posterior distributions.

The prior distributions are listed in Table 3.1. One MCMC iteration contains the following steps.

1. $s_{1:T}\big|y_{1:T}, \theta$
2. $\theta_j\big|y_{1:T}, s_{1:T}, \phi$, for $j = 1, 2, \ldots, K$
3. $\phi\big|y_{1:T}, s_{1:T}, \theta$
4. $P\big|s_{1:T}$

The first MCMC step is to sample the latent state variable $s_{1:T}$ from the conditional posterior distribution $s_{1:T}\big|y_{1:T}, \theta, P$. We follow Chib (1996) and use the forward filter backward sampler. In the second step, μ_j is sampled using the Gibbs sampler for the linear regression model. The conditional posterior of σ_j^2 is of unknown form and a Metropolis-Hasting step is used. The proposal density follows a gamma distribution formed by combining the likelihood for $RV_{1:T}$ and the prior. $\nu\big|y_{1:T}, \{\sigma_j^2\}_{j=1}^K$ is sampled using the Metropolis-Hasting algorithm with a random walk proposal. Finally, each row of P is drawn from its conditional posterior, which is a Dirichlet distribution. Additional details of posterior sampling are collected in the appendix.

After an initial burn-in of iterations are discarded we collect N additional MCMC iterations for posterior inference. Simulation consistent estimates of posterior quantities can be formed.[9] For example, the posterior mean of θ_j is estimated as,

$$E[\theta_j|y_{1:T}] \approx \frac{1}{N}\sum_{i=1}^{N}\theta_j^{(i)}, \tag{2.52}$$

[9]For a good textbook treatment of MCMC estimation see Greenberg (2014).

where $\theta_j^{(i)}$ is the i^{th} iteration from posterior sampling of parameter θ_j. The smoothed probability of s_t can be estimated as follows.

$$p(s_t = k|y_{1:T}) \approx \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}(s_t^{(i)} = k), \tag{2.53}$$

where $\mathbb{1}(A) = 1$ if A is true and otherwise 0.

The estimation of MS-RAV, MS-logRV, MS-logRAV and MS-RCOV models are done in a similar fashion. Detailed estimation steps of the models are in the appendix.

2.5.2 Estimation of Joint IHMM Models

In the IHMM-RV model, estimation of the unknown parameters $\theta = \{\mu_j, \sigma_j^2\}_{j=1}^{\infty}$, $\phi = \{\nu, \alpha, \eta\}$, P, Γ and $s_{1:T}$ are different given the unbounded nature of the state space. The beam sampler, introduced by Gael *et al.* (2008), is an extension of the slice sampler by Walker (2007), and is an elegant solution to estimation challenges that an infinite parameter model presents. An auxiliary variable $u_{1:T} = \{u_1, u_2, \cdots u_T\}$ is introduced that randomly truncates the state space to a finite one at each MCMC iteration. Conditional on $u_{1:T}$ the number of states is finite and the forward filter backward sampler previously discussed can be used to sample $s_{1:T}$.

The key idea behind the beam sampling is to introduce the auxiliary variable u_t that preserves the target distributions, and has the following conditional density

$$p(u_t|s_{t-1}, s_t, P) = \frac{\mathbb{1}(0 < u_t < P_{s_{t-1},s_t})}{P_{s_{t-1},s_t}}, \tag{2.54}$$

where $P_{i,j}$ denotes element (i,j) of P. The forward filtering step becomes

$$p(s_t|y_{1:t},u_{1:t},P) \propto p(y_t|y_{1:t-1},s_t)\sum_{s_{t-1}=1}^{\infty}\mathbb{1}(0<u_t<P_{s_{t-1},s_t})p(s_{t-1}|y_{1:t-1},u_{1:t-1},P) \quad (2.55)$$

$$\propto p(y_t|y_{1:t-1},s_t)\sum_{s_{t-1}:u_t<P_{s_{t-1},s_t}} p(s_{t-1}|y_{1:t-1},u_{1:t-1},P), \quad (2.56)$$

which renders an infinite summation into a finite one. Conditional on u_t, only states satisfying $u_t < P_{s_{t-1},s_t}$ are considered and the number of states become a finite number, say K. The same considerations hold for the backward sampling step.

Each MCMC iteration loop contains the following steps.

1. $u_{1:T}\big|s_{1:T},P,\Gamma$
2. $s_{1:T}\big|y_{1:T},u_{1:T},\theta,\phi,P,\Gamma$
3. $\Gamma\big|s_{1:T},\eta,\alpha$
4. $P\big|s_{1:T},\Gamma,\alpha$
5. $\theta_j\big|y_{1:T},s_{1:T},\phi$ for $j=1,2,\ldots,$
6. $\phi\big|y_{1:T},s_{1:T},\theta$

$u_{1:T}$ is sampled from its conditional densities $u_t\big|s_{1:T},P \sim \text{Uniform}(0,P_{s_{t-1},s_t})$ for $t=1,\cdots,T$. Following the discussion above, conditional on $u_{1:T}$ the effective state space is finite of dimension K and $s_{1:T}$ is sampled using the forward filter backward sampler. Γ and each row of transition matrix follow a Dirichlet distribution after additional latent variables are introduced. The sampling of μ_j, σ_j^2 and ν are the same as in the joint finite MS models. Posterior sampling of the IHMM-logRV, IHMM-logRAV and IHMM-RCOV models can be done following similar steps. The appendix

provides the detailed steps.

Given N MCMC iterations collected after a burn-in period are discarded, posterior statistics can be estimated as usual. The estimation of state-dependent parameters suffer from a label-switching problem. This means that the states and the associated parameters can be permuted over different MCMC iterations while maintaining the same likelihood value. Thus, it is not possible to track state j over the MCMC iterations as the definition of state j can change. Therefore we focus on label invariant quantities. For example, the posterior mean of θ_{s_t} is computed as

$$E[\theta_{s_t}|y_{1:T}] \approx \frac{1}{N}\sum_{i=1}^{N} \theta^{(i)}_{s_t^{(i)}}. \tag{2.57}$$

2.5.3 Density Forecasts

The predictive density is the distribution governing a future observation given a model $\mathcal{M}$, prior and data. It is computed by integrating out parameter uncertainty. The predictive likelihood is the key quantity used in model comparison and is the predictive density evaluated at next period's return

$$p(r_{t+1}\big|y_{1:t}, \mathcal{M}) = \int p(r_{t+1}\big|y_{1:t}, \Lambda, \mathcal{M}) p(\Lambda\big|y_{1:t}, \mathcal{M})\ \mathrm{d}\Lambda, \tag{2.58}$$

where $p(r_{t+1}\big|y_{1:t}, \Lambda, \mathcal{M})$ is the data density given $y_{1:t}$ and parameter Λ and $p(\Lambda\big|y_{1:t}, \mathcal{M})$ is the posterior distribution of Λ.

To focus on model performance and comparison it is convenient to consider the log-predictive likelihood and use the sum of log-predictive likelihoods from time $t+1$

to $t+s$ given as

$$\sum_{l=t+1}^{t+s} \log p(r_l \big| y_{1:l-1}, \mathcal{M}). \tag{2.59}$$

The log-predictive Bayes factor between $\mathcal{M}_1$ and model $\mathcal{M}_2$ is defined as

$$\sum_{l=t+1}^{t+s} \log p(r_l \big| y_{1:l-1}, \mathcal{M}_1) - \sum_{l=t+1}^{t+s} \log p(r_l \big| y_{1:l-1}, \mathcal{M}_2). \tag{2.60}$$

A log-predictive Bayes factor greater than 5 provides strong support for $\mathcal{M}_1$.

Predictive Likelihood of MS Models

Both parameter uncertainty and state uncertainty need to be integrated out in order to calculate the predictive likelihood. The predictive likelihood of a K-state joint MS model can be estimated as follows

$$p(r_{t+1} \big| y_{1:t}) \approx \frac{1}{N} \sum_{i=1}^{N} \sum_{s_{t+1}=1}^{K} \mathrm{N}(r_{t+1} \big| \mu_{s_{t+1}}^{(i)}, \sigma_{s_{t+1}}^{2(i)}) P^{(i)}_{s_{t+1}, s_t^{(i)}}, \tag{2.61}$$

where $\mu_{s_{t+1}}^{(i)}$ and $\sigma_{s_{t+1}}^{2(i)}$ are the i^{th} draw of $\mu_{s_{t+1}}$ and $\sigma_{s_{t+1}}^2$ respectively, and $P_{j1,j2}^{(i)}$ denotes element $(j1, j2)$ of $P^{(i)}$ all based on the posterior distribution given data $y_{1:t}$.

The calculation of the predictive likelihood for the multivariate MS models follows the same method,

$$p(R_{t+1} \big| Y_{1:t}) \approx \frac{1}{N} \sum_{i=1}^{N} \sum_{s_{t+1}=1}^{K} \mathrm{N}(R_{t+1} \big| M_{s_{t+1}}^{(i)}, \Sigma_{s_{t+1}}^{(i)}) P^{(i)}_{s_{t+1}, s_t^{(i)}}. \tag{2.62}$$

Predictive Likelihood of IHMM

For the IHMM models the state next period may be a recurring one or it may be new, the calculation of predictive likelihood is sightly different and is estimated as follows,

$$p(r_{t+1}|y_{1:t}) \approx \frac{1}{N}\sum_{i=1}^{N} \mathrm{N}(r_{t+1}|\mu_i, \sigma_i^2) \tag{2.63}$$

where the parameter values μ_i and σ_i^2 are determined using the following steps. Given $s_t^{(i)}$, draw $s_{t+1} \sim \text{Multinomial}(P^{(i)}_{s_t^{(i)}}, K^{(i)}+1)$.

1. If $s_{t+1} <= K^{(i)}$, set $\mu_i = \mu^{(i)}_{s_{t+1}}$, $\sigma_i^2 = \sigma^{2(i)}_{s_{t+1}}$.

2. If $s_{t+1} = K^{(i)}+1$, draw a new set of parameter values from the prior: $\mu_i \sim \mathrm{N}(m, v^2)$ and $\sigma_i^2 \sim \mathrm{IG}(v_0, s_0)$.

In multivariate IHMM models, the predictive likelihood is calculated exactly the same way except that the base measure draw is from a multivariate normal and an inverse-Wishart distribution.

2.5.4 Point Predictions for Returns and Volatility

In addition to density forecasts we evaluate the predictive mean of returns and the predictive variance (covariance) of returns. For the finite state MS models, conditional on the MCMC output, the predictive mean for r_{t+1} is estimated as

$$E[r_{t+1}|y_{1:t}] \approx \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{K} \mu_j^{(i)} P^{(i)}_{s_t^{(i)},j}. \tag{2.64}$$

The second moment of the predictive distribution can be estimated as follows.

$$E[r_{t+1}^2|y_{1:t}] \approx \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{K}(\mu_j^{(i)2} + \sigma_j^{2(i)})P^{(i)}_{s_t^{(i)},j}, \tag{2.65}$$

so that the variance can be estimated from

$$\mathrm{Var}(r_{t+1}|y_{1:t}) = E[r_{t+1}^2|y_{1:t}] - (E[r_{t+1}|y_{1:t}])^2. \tag{2.66}$$

For the multivariate model we have

$$E[R_{t+1}|Y_{1:t}] \approx \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{K} M_j^{(i)} P^{(i)}_{s_t^{(i)},j}, \tag{2.67}$$

$$E[R_{t+1}R_{t+1}^{'}|Y_{1:t}] \approx \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{K}(\Sigma_j^{(i)} + M_j^{(i)}M_j^{(i)'})P^{(i)}_{s_t^{(i)},j} \tag{2.68}$$

which can be used to estimate

$$\mathrm{Cov}(R_{t+1}|Y_{1:t}) = E[R_{t+1}R_{t+1}^{'}|Y_{1:t}] - E[R_{t+1}|Y_{1:t}]E[R_{t+1}|Y_{1:t}]^{'}. \tag{2.69}$$

For the IHMM models, the predictive mean and variance are derived from the following.

$$E[r_{t+1}|y_{1:t}] \approx \frac{1}{N}\sum_{i=1}^{N}\mu_i, \tag{2.70}$$

$$E[r_{t+1}^2|y_{1:t}] \approx \frac{1}{N}\sum_{i=1}^{N}(\sigma_i^2 + \mu_i^2). \tag{2.71}$$

The parameters μ_i, σ_i^2 are selected following the steps in Section 2.5.3. Similar results

hold for the multivariate versions.

Finally, estimation and forecasting for the benchmark model follow along the same lines as discussed for the joint models with minor simplifications.

2.6 Univariate Return Application

Four versions of joint MS models (MS-RV, MS-logRV, MS-RAV and MS-logRAV), joint IHMM-RV and the benchmark alternatives are applied to model monthly U.S. stock market returns from March 1885 to December 2013. The data from March 1885 to December 1925 are the daily capital gain returns provided by Bill Schwert, see Schwert (1990). The rest of returns are from the value-weighted S&P 500 index excluding dividends from CRSP, for a total of 1542 observations. The daily simple returns are converted to continuous compounded returns and are scaled by 12. The monthly return r_t is the sum of the daily returns. RV_t and other ex-post volatility measures are computed according to the definitions previously stated. Table 2.2 reports the summary statistics of monthly returns along with the summary statistics of monthly RV_t, $\log(RV_t)$, RAV_t and $\log(RAV_t)$.

Table 2.1 lists the priors for the various models. The priors provide a wide range of empirically realistic parameter values. The benchmark models have the same prior. Results are based on 5000 MCMC iterations after dropping the first 5000 draws.

2.6.1 Out-of-Sample Forecasts

Table 2.3 reports the sum of log-predictive likelihoods of 1 month ahead returns for the out-of-sample period from January 1951 to December 2013 (756 observations).

Recursive out-of-sample forecasts are computed beginning at the start of the out-of-sample period. As new data arrives the models are re-estimated and forecasts computed one period ahead. This is repeated until the end of the sample is reached.

In each case of a finite state assumption, the joint MS specifications outperform the benchmark model that do not use ex-post volatility measures. The log-predictive Bayes factors between the best finite joint MS model and the benchmark model are greater than 15, which provide strong evidence that exploiting higher frequency data leads to more accurate density forecasts. Using ex-post volatility data offer little to no gains in forecasting the mean of returns, however, it does lead to better variance forecasts as measured against realized variance.

The lower panel of Table 2.3 report the same results for the infinite hidden Markov models with and without higher frequency data. The overall best model according to the predictive likelihood is the IHMM-RV specification. This model has a log-predictive Bayes factor of 20.8 against the best finite state model that does not use high-frequency data. It has about the same log-predictive Bayes factor against the IHHM. The IHMM-RV has the lowest RMSE for RV_t forecasts. It is 10.6% lower than the best MS model that uses returns only.

All of the joint models that include some form of ex-post volatility lead to improved forecasts but generally the best performance comes from using RV_t.

Table 2.4 provides a check on these results over the shorter sample period from January 1984 to December 2013 (360 observations). The general results are the same as the longer sample, high-frequency data offer substantial improvements to density forecasts of returns and gains on forecasts of RV_t.

Figures 2.1 gives a breakdown of the period-by-period difference in the predictive

likelihood values. Positive values are in favour of the model with RV_t. The overall sum of the log-predictive likelihoods is not due to a few outliers or any one period but represent ongoing improvements in accuracy. The joint models do a better job in forecasting return densities when the market is in a high volatility period, such as the period of 1973-1974 crash, the period before and after the internet bubble and the 2008 financial crisis. Table 2.5 reports the log-predictive Bayes factors using data from four major bear regimes. The data windows are small, nevertheless, the log-predictive Bayes factors are all positive and support the joint models which exploit RV.

2.6.2 Parameter Estimates and State Inference

Table 2.6 reports the posterior summary of parameters of the 2 state MS, MS-RV and MS-RAV models based on the full sample. To avoid label switching issues, we use informative priors $\mu_1 \sim \text{N}(\text{-}1, 1)$, $\mu_2 \sim \text{N}(1, 1)$, $P_1 \sim \text{Dir}(4, 1)$, $P_2 \sim \text{Dir}(1, 4)$ and $\sigma_j^2 \sim \text{IG}(5, 5)$ for $j = 1, 2$ and restrict $\mu_1 < 0$ and $\mu_2 > 0$. The results show that all three models are able to sort stock returns into two regimes. One regime has a negative mean and high volatility, the other regime has positive mean return mean along with lower variance. This is consistent with the results of Maheu *et al.* (2012) and several other studies.

Compared with the benchmark MS model, the joint models specify the return distribution more precisely in each state, as can be seen the smaller estimated values $\sigma_{s_t}^2$. For instance, in the first state the innovation variance is 1.6127 for the MS model, while the estimates of variance are 0.5633 and 0.6581 in the MS-RV and MS-RAV models, respectively. The variance estimates in the positive mean regimes drop from

0.2187 to 0.1556 and 0.1460 after joint modelling RV and RAV. We would expect this reduction in the innovation variance to result in better forecasts which is what we found in the previous section.

Another interesting result is that MS-RV and MS-RAV models provide more precise estimates of all the model parameters. As shown in Table 2.6, all the parameter estimators have smaller posterior standard deviations and shorter 0.95 density intervals. For example, the length of the density interval of μ_1 from the benchmark model is 0.437, while the values are 0.187 and 0.167 from the MS-RV and MS-RAV models, respectively.

Figure 2.2 plots $E\left[\sigma_{s_t}^2 \middle| y_{1:T}\right]$ for the IHMM-RV and IHMM models. Volatility estimates vary over a larger range from the IHMM-RV model. For example, it appears that the IHMM overestimates the return variance during calm market periods and underestimates the return variance in several high volatile periods, such as the October 1987 crash and the financial crisis in 2008. In contrast, the return variance from the IHMM-RV model is closer to RV during these times. The differences between the models is due to the additional information from ex-post volatility.

Figure 2.3 plots the smoothed probability of the high return state from the 2 state MS, MS-RV and MS-RAV models. The benchmark MS model does a fairly good job in identifying the primary downward market trends, such as the big crash of 1929, 1973-1974 bear market and the 2008 market crash, but it ignores a series of panic periods before and after 1900, the internet bubble crash and several other relatively smaller downward periods. The joint MS-RV and MS-RAV models not only identify the primary market trends but also are able to capture a number of short lived market drops. The main difference is that the joint model appears to have more frequent

state switches and state identification is more precise. One obvious example is the joint models identify the dot-com collapse from 2000 to 2002 and the market crash of 2008-2009.

In summary, the joint models lead to better density forecasts, better forecasts of realized variance, improved parameter precision and minor differences in latent state estimates.

2.6.3 Market Timing Portfolio

As shown in Section 2.6.1, the joint MS model improves density forecast of returns. We further investigate if the better forecasts lead to actual economic gain. The proposed joint models (MS-RV and IHMM-RV) are compared with benchmark models from a portfolio allocation perspective.

Suppose an investor uses a market timing strategy to manage her portfolio. Let r_t^m, r_t^p and r_t^f denotes the monthly simple return of the market (index), a market-timing portfolio and the risk-free asset, respectively. The trading strategy is designed as follows. Each model $\mathcal{M}$ is used to forecast the direction of the market next month. That is, if $P(r_{t+1}^m > r_t^f | \mathcal{M}, r_{1:t}, RV_{1:t}) > 0.5$, one dollar is invested in the market and held for one month to receive that month's return. In this case, the portfolio return $r_{t+1}^{\mathcal{M}} = r_{t+1}^m$. Otherwise, the risk-free asset is held and the monthly portfolio return is $r_{t+1}^{\mathcal{M}} = r_{t+1}^f$. This is repeated for several different models $\mathcal{M}_1, \ldots, \mathcal{M}_p$ and each produces its own portfolio of returns, $r_t^{\mathcal{M}_1}, \ldots, r_t^{\mathcal{M}_p}$, over time.

The evaluation of portfolios (models) is based on risk-return tradeoff and utility-based approach. Two types of utility functions are used. The quadratic utility function used in Fleming *et al.* (2001) has the following form

$$U_q(r_t^{\mathcal{M}}) = (1 + r_t^{\mathcal{M}}) - \frac{\gamma}{2(1+\gamma)}(1 + r_t^{\mathcal{M}})^2, \tag{2.72}$$

where γ denotes the risk aversion coefficient and is set to be $\gamma = 5$.

Following Skouras (2007) and Clements and Silvennoinen (2013), the second utility function is exponential, given as follows

$$U_e(r_t^{\mathcal{M}}) = -\exp(-\gamma(1 + r_t^{\mathcal{M}})). \tag{2.73}$$

The performance fee Δ that an investor would pay to switch from one portfolio to another is used to evaluate the competing portfolios. Δ is a constant equalizing the ex-post utility from models $\mathcal{M}_1$ and $\mathcal{M}_2$ in

$$\sum_{t=1}^{T} U(r_t^{\mathcal{M}_1}) = \sum_{t=1}^{T} U(r_t^{\mathcal{M}_2} - \Delta). \tag{2.74}$$

Table 2.7 reports the summary statistics of the portfolio returns. The out-of-sample market timing is conducted from 1951-2013 and for a shorter sample 1984-2013. The portfolios based on the joint MS or IHMM models yield higher average returns, lower risks and result in higher Sharpe ratios in all the cases, compared with the ones based on models without RV. It is only possible to beat the Sharpe ratio from the buy and hold strategy if a model uses RV information. For instance, in all cases except one, the Sharpe ratio from the joint models is higher than from a buy

and hold strategy. The last two columns of Table 2.7 record the annualized basis point performance fees that an investor would be willing to pay to switch from the portfolio based on the 2-state MS benchmark model to the one based on another model. It shows incorporating RV in MS models always improves the utility level of an investor with either quadratic or exponential utility. Independent of the number of states or the sample period an investor is always willing to move from the simple MS two state model to a specification that exploits RV.

2.7 Multivariate Return Application

We also evaluate the multivariate joint models through multivariate applications to a vector of equity returns. The daily prices of three equities (stock symbol: IBM, XOM and GE) listed in NYSE are obtained from CRSP. These firms were chosen since they have been actively traded over the full sample period, January 1926 to December 2013 (1056 observations). The continuous compounded returns are constructed and the monthly RCOV is computed using daily values following equation (2.3). The summary statistics of monthly returns R_t and $RCOV_t$ are found in Table 2.9. The prior specification is found in Table 2.8.

2.7.1 Out-of-Sample Forecasts

Table 2.10 reports the results of density forecasts and the root mean squared error of predictions based on 756 out-of-sample observations. We found the larger finite state models are the most competitive and therefore do not include results for small dimension models. The 8 and 12 state models that exploit RCOV are all superior to

the models that do not according to log-predictive values. The improvement in the log-predictive likelihood is 30 or more. Further improvements are found on moving to the Bayesian nonparametric models. The IHMM-RCOV model is the best over the alternative models.

As for point predictions of return and realized covariance, the results is similar to the univariate return applications. The proposed joint models improve predictions of $RCOV_t$ but offer no gains for return predictions.

Figure 2.4 displays the posterior average of active states in both IHMM and IHMM-RCOV models at each point in the out-of-sample period. It shows that more states are used in the joint return-RCOV model in order to better capture the dynamics of returns and volatility.

2.7.2 Portfolio Performance

Beyond the forecasts of return density and covariance, we also evaluate the out-of-sample performance of models in portfolio allocation. Suppose an investor forms her portfolio using three equities (IBM, XOM and GE) and applies the modern portfolio theory by Markowitz (1952) to select a portfolio on the efficient frontier. The weight of a minimum variance portfolio given required portfolio return μ_p can be calculated by solving the problem below.

$$\min_{w_{t+1}} \quad w_{t+1}' \mathrm{Cov}_{t+1|t} w_{t+1} \quad \text{s.t.} \quad w_{t+1}' \boldsymbol{\mu} + (1 - w_{t+1}' \mathbf{1}) r_t^f = \mu_p, \tag{2.75}$$

where w_{t+1} is the portfolio weight, $\mathrm{Cov}_{t+1|t}$ denotes the predictive covariance from a model using data up to time t, $\mathbf{1}$ is a vector of ones and $\boldsymbol{\mu}$ is the return mean vector

set to the sample average and is the same across models. The solution is

$$w_{t+1} = \frac{(\mu_p - r_t^f)\text{Cov}_{t+1|t}^{-1}(\boldsymbol{\mu} - r_t^f \mathbf{1})}{(\boldsymbol{\mu} - r_t^f \mathbf{1})'\text{Cov}_{t+1|t}^{-1}(\boldsymbol{\mu} - r_t^f \mathbf{1})}. \tag{2.76}$$

The portfolio return for model $\mathcal{M}$ is $r_{t+1}^{\mathcal{M}} = w_{t+1}' R_{t+1} + (1 - w_{t+1}'\mathbf{1}) r_{t+1}^f$. R_{t+1} and r_{t+1}^f are all expressed as simple returns. $\boldsymbol{\mu}$ is set to be the sample average of returns and is common to all models.[10] $\text{Cov}_{t+1|t}$ is computed following equation (2.69). The utility functions and performance fee calculations are the same as in the univariate portfolio application.

A second portfolio application compares the models on their ability to produce a global minimum variance portfolio Engle and Colacito (2006). The global minimum variance portfolio can be determined by solving the following minimization problem.

$$\min_{w_{t+1}} \quad w_{t+1}' \text{Cov}_{t+1|t} w_{t+1} \quad \text{s. t.} \quad \text{and} \quad w_{t+1}'\mathbf{1} = 1. \tag{2.77}$$

The weight of global minimum-variance portfolio is given by

$$w_{t+1} = \frac{\text{Cov}_{t+1|t}^{-1}\mathbf{1}}{\mathbf{1}'\text{Cov}_{t+1|t}^{-1}\mathbf{1}}. \tag{2.78}$$

The model parameters are re-estimated each month after new data arrives and the out-of-sample period is from 1951–2013 and 1984–2013. Tables 2.11 and 2.12 summarizes the performance of minimum-variance portfolios for two sample periods based on the benchmark model and proposed joint models. Panel A and B show the

[10] This is done to focus on the differences in the the predictive covariance. When both the predictive mean and predictive covariance are used from each model the IHMM-RCOV specification is strongly favored compared to other models.

results for required annual portfolio returns of $\mu_p = 10\%$ and $\mu_p = 20\%$, respectively.[11]

The benefit of jointly modelling both return and RCOV is clear. In all the cases but one using the predictive covariance from a model incorporating RCOV increases the Sharpe ratio relative to the model without RCOV. The investor is always willing to pay to move to a model that exploits RCOV information. For instance the investor using an 8-state MS model would pay 7.08 basis points to use the forecasts from an 8-state MS-RCOV model. In Table 2.12 performance fees are considerably larger for the more recent sample period.

Finally, Table 2.13 reports the variances of global minimum-variance portfolios based on competing models over the two sample periods. The specifications that jointly model returns and RCOV always lead to portfolios with smaller variances, compared to the model without RCOV, no matter the number of states.

In summary, the joint modelling of returns and RCOV leads to better out-of-sample forecasts and improved portfolio decisions.

2.8 Conclusion

This chapter shows how to incorporate ex-post measures of volatility with returns to improve forecasts, parameter and state estimation under a Markov switching assumption. We show how to build and estimate joint nonlinear factor models. Markov switching can be specified as fixed and finite or countably infinite. In empirical applications the new models give dramatic improvements in density forecasts for returns, forecasts of realized variance and lead to improved portfolio decisions.

[11]Note that the annualized mean return for IBM, XOM and GE are 13.4%, 11.6% and 10.3%, respectively.

2.9 Appendix

2.9.1 MS-RV Model

(1) $s_{1:T}\big|y_{1:T}, \theta, \phi, P$

The latent state variable $s_{1:T}$ is sampled using the forward filter backward sampler (FFBS) in Chib (1996). The forward filter part contains the following steps.

i. Set the initial value of filter $p(s_1 = j\big|y_1, \theta, \phi, P) = \pi_j$, for $j = 1, \ldots, K$, where π is the stationary distribution, which can be computed by solving $\pi = P^\intercal \pi$.

ii. Prediction step: $p(s_t\big|y_{1:t-1}, \theta, \phi, P) \propto \sum_{j=1}^{K} P_{j,s_t} \cdot p(s_{t-1} = j\big|y_{1:t-1}, \theta, \phi, P)$.

iii. Update step:

$$p(s_t\big|y_{1:t}, \theta, \phi, P) \propto f(r_t\big|\mu_{s_t}, \sigma_{s_t}^2) \cdot g(RV_t\big|\nu + 1, \nu\sigma_{s_t}^2) \cdot p(s_t\big|y_{1:t-1}, \theta, \phi, P),$$

where function $f(\cdot)$ and $g(\cdot)$ denote normal density and inverse-gamma density, respectively.

The underlying states are drawn using backward sampler as follows.

i. For $t = T$, draw s_T from $p(s_T\big|y_{1:T}, \theta, \phi, P)$.

ii. For $t = T-1, \ldots, 1$, draw s_t from $P_{s_t, s_{t+1}} \cdot p(s_t\big|y_{1:t}, \theta, \phi, P)$.

Let $n_j = \sum_{t=1}^{T} \mathbb{1}(s_t = j)$ denotes the number of observations belong to state j.

(2) $\mu_j\big|r_{1:T}, s_{1:T}, \sigma_j^2$ for $j = 1, \ldots, K$

μ_j is sampled using the Gibbs sampling for linear regression model. Given prior $\mu_j \sim \mathrm{N}(m_j, v_j^2)$, μ_j is sampled from conditional posterior $\mathrm{N}(\overline{m_j},\ \overline{v_j}^2)$, where

$$\overline{m_j} = \frac{v_j^2 \sum_{s_t=j} r_t + m_j \sigma_j^2}{\sigma_j^2 + n_j v_j^2}, \quad \text{and} \quad \overline{v_j}^2 = \frac{\sigma_j^2 v_j^2}{\sigma_j^2 + n_j v_j^2}.$$

(3) $\sigma_j^2 \big| y_{1:T}, \mu_j, \nu, s_{1:T}$ for $j = 1, \ldots, K$

The prior of σ_j^2 is assumed to be $\sigma_j^2 \sim \mathrm{G}(v_0,\ s_0)$. The conditional posterior of σ_j^2 is given as follows,

$$\begin{aligned} p(\sigma_j^2 \big| y_{1:T}, \mu_j, \nu, s_{1:T}) \quad \propto \quad & \prod_{s_t=j} \left\{ \frac{1}{\sigma_j} \exp\left[-\frac{(r_t - \mu_j)^2}{2\sigma_j^2} \right] \cdot (\nu\sigma_j^2)^{(\nu+1)} \exp\left(-\frac{\nu\sigma_j^2}{RV_t} \right) \right\} \\ & \cdot (\sigma_j^2)^{v_0 - 1} \exp\left(-s_0 \sigma_j^2 \right). \end{aligned}$$

The conditional posterior of σ_j^2 is not of any known form, therefore Metropolis-Hasting algorithm is applied to sample σ_j^2. Combining RV likelihood function and prior provides a good proposal density $q(\cdot)$ for sampling σ_j^2, which follows a gamma distribution and is derived as follows,

$$\begin{aligned} p(\sigma_j^2 \big| y_{1:T}, \mu_j, \nu, s_{1:T}) \quad \propto \quad & \prod_{s_t=j} \left\{ (\nu\sigma_j^2)^{(\nu+1)} \exp\left(-\frac{\nu\sigma_j^2}{RV_t} \right) \right\} \cdot (\sigma_j^2)^{v_0 - 1} \exp\left(-s_0 \sigma_j^2 \right) \\ \sim \quad & \mathrm{G}\left(n_j(\nu+1) + v_0, \nu \sum_{s_t=j} \frac{1}{RV_t} + s_0 \right) \equiv q(\sigma_j^2). \end{aligned}$$

(4) $\nu \big| y_{1:T}, \{\sigma_j^2\}_{j=1}^K, s_{1:T}$

The prior of ν is assumed to be $\nu \sim \mathrm{IG}(a, b)$. The posterior of ν is given as follows.

$$p(\nu \big| y_{1:T}, \sigma_{s_t}^2, s_{1:T}) \propto \prod_{t=1}^{T} \left\{ \frac{(\nu\sigma_{s_t}^2)^{(\nu+1)}}{\Gamma(\nu+1)} RV_t^{-\nu-2} \exp\left(-\frac{\nu\sigma_{s_t}^2}{RV_t} \right) \right\} \cdot (\nu)^{-a-1} \exp\left(-\frac{b}{\nu} \right)$$

ν is drawn from a random walk proposal and negative draws will be dropped.

(5) $P\big|s_{1:T}$

Using conjugate prior for rows of the transition matrix P: $P_j \sim \text{Dir}(\alpha_{j1}, \cdots, \alpha_{jK})$, the posterior is given by $\text{Dir}(\alpha_{j1}+n_{j1}, \cdots, \alpha_{jK}+n_{jK})$, where vector $(n_{j1}, n_{j2}, \cdots, n_{jK})$ records the numbers of switches from state j to the other states.

2.9.2 MS-logRV Model

Forward filter backward sampler is used to sample $s_{1:T}$. The sampling of P is same as step (5) in MS-RV model estimation. The sampling of μ_j is same as step (2) in MS-RV model except replacing σ_j^2 with $\exp(\zeta_j)$.

Let $n_j = \sum_{t=1}^{T} \mathbb{1}(s_t = j)$ denotes the number of observations belong to state j. Assuming the prior of ζ_j is $\zeta_j \sim N(m_{\zeta,j}, v_{\zeta,j}^2)$, the conditional posterior of ζ_j is given as follows,

$$p(\zeta_j\big|y_{1:T}, \mu_j, \delta_j^2, s_{1:T}) \propto \prod_{s_t=j} \left\{ \exp\left[-\frac{\zeta_j}{2} - \frac{(r_t-\mu_j)^2}{2\exp(\zeta_j)}\right] \cdot \exp\left[-\frac{(\log RV_t - \zeta_j + \frac{1}{2}\delta_j^2)^2}{2\delta_j^2}\right] \right\} \cdot \exp\left[-\frac{(\zeta_j - m_{\zeta j})^2}{2v_{\zeta,j}^2}\right]$$

Metropolis-Hasting algorithm is applied to sample ζ_j. The proposal density is

formed as follows,

$$\begin{aligned} p(\zeta_j|y_{1:T},\mu_j,\delta_j^2,s_{1:T}) &\propto \prod_{s_t=j}\left\{\exp\left[-\frac{\zeta_j}{2}-\frac{(r_t-\mu_j)^2}{2\exp(\zeta_j)}\right]\right\}\cdot\exp\left[-\frac{(\zeta_j-\mu^*)^2}{2\sigma^{2*}}\right] \\ &< \exp\left[-\frac{n_j\zeta_j}{2}-\frac{\sum_{s_t=j}(r_t-\mu_j)^2\exp(-\mu^*)(-\zeta_j)}{2}-\frac{(\zeta_j-\mu^*)^2}{2\sigma^{2*}}\right] \\ &\sim \mathrm{N}(\mu_j^{**},\sigma_j^{*2})\equiv q(\zeta_j), \quad \text{where} \end{aligned}$$

$$\mu_j^*=\frac{v_{\zeta,j}^2\sum_{s_t=j}\log(RV_t)+\frac{1}{2}n_jv_\zeta^2\delta_j^2+\delta_j^2m_{\zeta,j}}{n_jv_{\zeta,j}^2+\delta_j^2}, \qquad \sigma_j^{*2}=\frac{\delta_j^2v_{\zeta,j}^2}{n_jv_{\zeta,j}^2+\delta_j^2},$$
$$\mu_j^{**}=\mu^*+\frac{1}{2}\sigma^{*2}\left[\sum_{s_t=j}(r_t-\mu_j)^2\exp(-\mu^*)-n_j\right].$$

Using conjugate prior $\delta_j^2\sim\mathrm{IG}(v_0,s_0)$, the posterior density of δ_j^2 is given by

$$p(\delta_j^2|y_{1:T},\mu_j,\sigma_j^2)\propto\prod_{s_t=j}\left\{\frac{1}{\delta_j}\exp\left[-\frac{(\log RV_t-\zeta_j+\frac{1}{2}\delta_j^2)^2}{2\delta_j^2}\right]\right\}\cdot\ \delta_j^{-v_0-1}\exp\left(-\frac{s_0}{\delta_j^2}\right)$$

Metropolis-Hasting is used to sample δ_j with the following proposal

$$q(\delta_j^2)\equiv\mathrm{IG}\left(\frac{n_j}{2}+v_0,\frac{\sum_{s_t=j}\left(\log RV_t-\zeta_j\right)^2}{2}+s_0\right).$$

2.9.3 MS-RAV Model

The sampling step of $s_{1:T}$, $\{\mu_j\}_{j=1}^K$ and P are same as in MS-RV model estimation. Let $y'_{1:t}=\{y'_1,\ldots,y'_t\}$, where $y'_t=\{r_t,RAV_t\}$. $\{\sigma_j\}_{j=1}^K$ and ν are sampled as follows.

The prior of σ_j^2 is assumed to be $\sigma_j^2\sim\mathrm{G}(v_0,s_0)$. The conditional posterior of σ_j^2

is given as follows,

$$p(\sigma_j^2|y'_{1:T},\mu_j,\nu,s_{1:T}) \propto \prod_{s_t=j}\left\{\frac{1}{\sigma_j}\exp\left[-\frac{(r_t-\mu_j)^2}{2\sigma_j^2}\right]\cdot\sigma_j^{2\nu}\exp\left[-\left(\frac{\sigma_j\Gamma(\nu)}{\Gamma(\nu-\frac{1}{2})}\right)^2\frac{1}{RAV_t^2}\right]\right\} \cdot(\sigma_j^2)^{v_0-1}\exp\left(-s_0\sigma_j^2\right).$$

$\sigma_j^{2\prime}$ is drawn from random walk proposal and negative draws discarded.

The prior for ν is assumed to be $\nu \sim \text{IG}(a,b)$. The posterior of ν is given as follows,

$$p(\nu|RAV_{1:T},\{\sigma_j^2\}_{j=1}^K,s_{1:T}) \propto \prod_{t=1}^{T}\left\{\left[\frac{\sigma_j\Gamma(\nu)}{\Gamma(\nu-\frac{1}{2})}\right]^{2\nu}\frac{RAV_t^{-2\nu-1}}{\Gamma(\nu)}\exp\left[-\left(\frac{\sigma_j\Gamma(\nu)}{\Gamma(\nu-\frac{1}{2})}\right)^2\frac{1}{RAV_t^2}\right]\right\} \cdot\nu^{-a-1}\exp\left(\frac{b}{\nu}\right).$$

Random walk proposal is used to sample ν and negative values discarded.

2.9.4 MS-logRAV Model

The estimation of MS-logRAV model is very similar to that of MS-logRV model except changing the return variance $\exp(\zeta_{s_t})$ to $\exp(2\zeta_{s_t})$.

2.9.5 MS-RCOV

See step (1) and (5) in Section 2.9.1 for the estimation of $s_{1:T}$ and P. Let $n_j = \sum_{t=1}^{T}\mathbb{1}(s_t=j)$ denotes the number of observations belong to state j.

Given conjugate prior $M_j \sim \mathrm{N}(G_j, V_j)$, the posterior density of M_j is given by

$$M_j \big| R_{1:T}, s_{1:T}, \Sigma_j \sim \mathrm{N}(\overline{M}, \overline{V}), \quad \text{where}$$

$$\overline{V} = \left(\Sigma_j^{-1} n_j + V_j^{-1}\right)^{-1}, \qquad \overline{M} = \overline{V}\left(\Sigma_j^{-1} \sum_{s_t=j} R_t + G_j V_j^{-1}\right).$$

The prior of Σ_j is assumed to be $\Sigma_j \sim \mathrm{W}(\Psi, \tau)$. The conditional posterior of Σ_j is given as follows,

$$\begin{aligned} p(\Sigma_j \big| Y_{1:T}, M_j, \kappa, s_{1:T}) \;\propto\; & \prod_{s_t=j} \left\{ |\Sigma_j|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(R_t - M_j)^\intercal \Sigma_j^{-1}(R_t - M_j)\right]\right\} \\ & \cdot \prod_{s_t=j} \left\{ |\Sigma_j|^{\frac{\kappa}{2}} |RCOV_t|^{-\frac{\kappa+d+1}{2}} \exp\left[-\frac{1}{2}\mathrm{tr}(\Sigma_j RCOV_t^{-1})\right]\right\} \\ & \cdot |\Sigma_j|^{\frac{\tau-d-1}{2}} \exp\left[-\frac{1}{2}\mathrm{tr}(\Psi^{-1}\Sigma_j)\right]. \end{aligned}$$

Metropolis-Hasting algorithm is applied to sample Σ_j. The proposal density $q_j(\cdot)$ is formed as follows,

$$\begin{aligned} p(\Sigma_j \big| Y_{1:T}, M_j, \kappa, s_{1:T}) \;\propto\; & \prod_{s_t=j} \left\{ |\Sigma_j|^{\frac{\kappa}{2}} \exp\left[-\frac{1}{2}\mathrm{tr}(\Sigma_j RCOV_t^{-1})\right]\right\} \cdot |\Sigma_j|^{\frac{\tau-d-1}{2}} \exp\left[-\frac{1}{2}\mathrm{tr}(\Psi^{-1}\Sigma_j)\right] \\ \sim\; & \mathrm{W}\left(\left[(\kappa - 1 - d)\sum_{s_t=j} RCOV_t^{-1} + \Psi^{-1}\right]^{-1},\; n_j\kappa + \tau\right) \equiv q_j(\cdot). \end{aligned}$$

Assuming the prior of κ is $\kappa \sim \mathrm{G}(a, b)$, the posterior density of κ is given as follows,

$$\begin{aligned} p(\kappa \big| Y_{1:T}, M_{s_t}, \Sigma_{s_t}, S) \propto \prod_{t=1}^{T} \Bigg\{ & \frac{|\Sigma_{s_t}(\kappa - d - 1)|^{\frac{\kappa}{2}}}{2^{\frac{\kappa d}{2}}\Gamma(\frac{\kappa}{2})} |RCOV_t|^{-\frac{\kappa+d+1}{2}} \\ & \cdot \exp\left[-\frac{1}{2}\mathrm{tr}\left((\kappa - d - 1)\Sigma_{s_t} RCOV_t^{-1}\right)\right]\Bigg\} \cdot \kappa^{a-1}\exp(-b\kappa). \end{aligned}$$

Metropolis-Hasting algorithm with random walk proposal is used to sample κ.

2.9.6 Univariate Joint IHMM-RV

Define vector $C = \{c_1, \cdots, c_K\}$ and another $K \times K$ matrix A, which will be used in sampling Γ and η. The MCMC steps are illustrated as follows. Several estimation steps are based on Maheu and Yang (2016) and Song (2014).

(1) $u_{1:T}\big|s_{1:T}, \Gamma, P$

Draw $u_1 \sim \text{Uniform}(0, \gamma_{s_1})$ and draw $u_t \sim \text{Uniform}(0, P_{s_{t-1},s_t})$ for $t = 2, \ldots, T$.

(2) Adjust the Number of States K

i. Check if $\max\{P^r_{1,K+1}, \cdots, P^r_{K,K+1}\} > \min\{u_{1:T}\}$. If yes, expand the number of clusters by making the following adjustments (ii) - (vi), otherwise, move to step (3).

ii. Set $K = K + 1$.

iii. Draw $u_\beta \sim \text{Beta}(1, \eta)$, set $\gamma_K = u_\beta \gamma^r_K$ and the new residual probability equals to $\gamma^r_{K+1} = (1 - u_\beta)\gamma^r_K$.

iv. For $j = 1, \cdots, K$, draw $u_\beta \sim \text{Beta}(\gamma_K, \gamma_{K+1})$, set $P_{j,K} = u_\beta P^r_{j,K}$ and $P^r_{j,K+1} = (1 - u_\beta)P^r_{j,K}$. Also, add an additional row to transition matrix P. $P_{K+1} \sim \text{Dir}(\alpha\gamma_1, \cdots, \alpha\gamma_K)$.

v. Expand the parameter size by 1 by drawing $\mu_{K+1} \sim \text{N}(m, v^2)$ and $\sigma^2_{K+1} \sim \text{IG}(v_0, s_0)$.

vi. Go back to step(i.).

(3) $s_{1:T}|y_{1:T}, u_{1:T}, \theta, \phi, P, \Gamma$

In this step, the latent state variable is sampled using the forward filter backward sampler Chib (1996). The forward filter part contains the following steps:

i. Set the initial value of filter $p(s_1 = j\big|y_1, u_1, \theta, \phi, P) = \mathbb{1}(u_0 < \gamma_j)$ and normalize it.

ii. Prediction step:

$$p(s_t\big|y_{1:t-1}, u_{1:t-1}, \theta, \phi, P) \propto \sum_{j=1}^{K} \mathbb{1}(u_t < P_{j,s_t}) \cdot p(s_{t-1} = j\big|y_{1:t-1}, u_{1:t-1}, \theta, \phi, P).$$

iii. Update step:

$$p(s_t\big|y_{1:t}, u_{1:t}, \theta, \phi, P) \propto f(r_t\big|\mu_j, \sigma_j^2) \cdot g(RV_t\big|\nu+1, \nu\sigma_j^2) \cdot p(s_t\big|y_{1:t-1}, u_{1:t-1}, \theta, \phi, P).$$

The underlying states are drawn using backward sampler as follows.

i. For $t = T$, draw s_T from $p(s_T\big|y_{1:T}, u_{1:T}, \theta, \phi, P)$.

ii. For $t = T-1, \cdots, 1$, draw s_t from $\mathbb{1}(u_t < P_{j,s_{t+1}}) \cdot p(s_t\big|y_{1:t}, u_{1:t}, P, \theta, \phi)$.

Then we count the number of active clusters and removing inactive states by making following adjustments.

i. Calculate the number of active states (states with at least one observation assigned to it) denoted by L. If $L < K$, remove the inactive states by adjusting the value of states.

ii. Adjust the order of state-dependent parameters μ, σ^2 and Γ according to the adjusted state $s_{1:T}$.

iii. Set $K = L$. Recalculate the residual probabilities of Γ^r_{K+1} for $j = 1, \cdots, K$. Then set the values of parameter μ_j, σ_j^2 and γ_j, to be zero for $j > K$.

(4) $\Gamma \big| s_{1:T}, \eta, \alpha$

i. Let $n_{j,i}$ denotes the number of state moves from state j to i. Calculate $n_{j,i}$ for $i = 1, \cdots, K$ and $j = 1, \cdots, K$.

ii. For $i = 1, \cdots, K$ and $j = 1, \cdots, K$, if $n_{j,i} > 0$, then for $l = 1, \cdots, n_{j,i}$, draw $x_l \sim \text{Bernoulli}(\frac{\alpha\gamma_i}{l-1+\alpha\gamma_i})$. If $x_l = 1$, set $A_{j,i} = A_{j,i} + 1$.

iii. Draw $\Gamma \sim \text{Dir}(c_1, \ldots, c_K, \eta)$, where $c_i = \sum_{j=1}^{K} A_{ji}$.

(5) $P \big| s_{1:T}, \Gamma, \alpha$

For $j = 1, \cdots K$, draw $P_j \sim \text{Dir}(\alpha\gamma_1 + n_{j,1}, \cdots, \alpha\gamma_k + n_{j,K}, \alpha\gamma^r_{K+1})$.

(6) $\theta \big| y_{1:T}, s_{1:T}, \nu$

See the step (2) and step (3) in Appendix 9.1. for the estimation of state-dependent parameters μ_j, σ_j^2, for $j = 1, \ldots, K$.

(7) $\nu \big| y_{1:T}, s_{1:T}, \{\sigma_j^2\}_{j=1}^K, \nu$

Same as the step (4) in Appendix 9.1.

(8) $\eta \big| s_{1:T}, \Gamma, \alpha$

Recompute C vector again as in step (4) and define ν and λ, where $\nu \sim \text{Bernoulli}(\frac{\sum_{i=1}^{K} c_i}{\sum_{i=1}^{K} c_i + \eta})$ and $\lambda \sim \text{Beta}(\eta+1, \sum_{i=1}^{K} c_i)$. Then draw a new value of $\eta \sim \text{G}(a_1 + K - \nu, b_1 - \log(\lambda))$.

(9) $\alpha \big| s_{1:T}, C$

Define ν'_j, λ'_j, for $j = 1, \cdots, K$, where $\nu'_j \sim \text{Bernoulli}(\frac{\sum_{i=1}^{K} n_{j,i}}{\sum_{i=1}^{K} n_{j,i} + \alpha})$ and $\lambda'_j \sim \text{Beta}(\alpha + 1, \sum_{i=1}^{K} n_{j,i})$. Then draw $\alpha \sim \text{G}(a_2 + \sum_{j=1}^{K} c_j - \sum_{j=1}^{K} \nu'_j, b_2 - \sum_{j=1}^{K} \log(\lambda'_j))$.

2.9.7 Univariate Joint IHMM-logRV and Joint IHMM-logRAV

See step (1) - (5), (8) and (9) in Appendix 9.6 for the estimation of auxiliary variable $u_{1:T}$, latent state variable $s_{1:T}$, Γ, transition matrix P, DP concentration parameter η and α. The estimation of $\theta = \{\mu_j, \zeta_j, \delta_j^2\}_{j=1}^{\infty}$ in IHMM-logRV are same as the MS-logRV model, see Appendix 9.3. The parameter estimation of IHMM-logRAV can be done similarly.

2.9.8 Joint IHMM-RCOV

See step (1) - (5), (8) and (9) in Appendix 9.6 for the estimation of $u_{1:T}$, $s_{1:T}$, Γ, P, η and α. The estimation of $\theta = \{M_j, \Sigma_j\}_{j=1}^{\infty}$ and κ are same as the estimation of MS-RCOV model, see Section 2.9.5.

Bibliography

Alizadeh, S., Brandt, M. W., and Diebold, F. (2002). Range-based estimation of stochastic volatility models. *The Journal of Finance*, **57**(3), 1047–1091.

Andersen, T. and Benzoni, L. (2009). Realized volatility. In *Handbook of Financial Time Series*. Springer.

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

Ang, A. and Bekaert, G. (2002). Regime switches in interest rates. *Journal of Business & Economic Statistics*, **20**(2), 163–182.

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

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

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

Blair, B. J., Poon, S.-H., and Taylor, S. J. (2001). Forecasting S&P 100 volatility: the incremental information content of implied volatilities and high-frequency index returns. *Journal of Econometrics*, **105**(1), 5 – 26.

Carpantier, J.-F. and Dufays, A. (2014). Specific Markov-switching behaviour for arma parameters. CORE Discussion Papers 2014014, Universit catholique de Louvain, Center for Operations Research and Econometrics (CORE).

Chib, S. (1996). Calculating posterior distributions and modal estimates in markov mixture models. *Journal of Econometrics*, **75**(1), 79–97.

Clements, A. and Silvennoinen, A. (2013). Volatility timing: How best to forecast portfolio exposures. *Journal of Empirical Finance*, **24**, 108 – 115.

Dueker, M. and Neely, C. J. (2007). Can Markov switching models predict excess foreign exchange returns? *Journal of Banking & Finance*, **31**(2), 279–296.

Dufays, A. (2016). Infinite-state Markov-switching for dynamic volatility. *Journal of Financial Econometrics*, **14**(2), 418–460.

Engel, C. and Hamilton, J. D. (1990). Long swings in the dollar: Are they in the data and do markets know it? *American Economic Review*, **80**, 689–713.

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

Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, **1**(2), 209–230.

Fleming, J., Kirby, C., and Ostdiek, B. (2001). The economic value of volatility timing. *The Journal of Finance*, **56**(1), 329–352.

Forni, M. and Reichlin, L. (1998). Let's get real: A factor analytical approach to

disaggregated business cycle dynamics. *The Review of Economic Studies*, **65**(3), 453–473.

Gael, J. V., Saatci, Y., Teh, Y. W., and Ghahramani, Z. (2008). Beam sampling for the infinite hidden Markov model. In *In Proceedings of the 25th International Conference on Machine Learning*.

Greenberg, E. (2014). *Introduction to Bayesian Econometrics*. Cambridge University Press.

Guidolin, M. and Timmermann, A. (2006). An econometric model of nonlinear dynamics in the joint distribution of stock and bond returns. *Journal of Applied Econometrics*, **21**(1), 1–22.

Guidolin, M. and Timmermann, A. (2008). International asset allocation under regime switching, skew, and kurtosis preferences. *Review of Financial Studies*, **21**(2), 889–935.

Guidolin, M. and Timmermann, A. (2009). Forecasts of US short-term interest rates: A flexible forecast combination approach. *Journal of Econometrics*, **150**(2), 297–311.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, **57**(2), 357–384.

Hansen, P., 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.

Jin, X. and Maheu, J. M. (2013). Modeling realized covariances and returns. *Journal of Financial Econometrics*, **11**(2), 335.

Jin, X. and Maheu, J. M. (2016). Bayesian semiparametric modeling of realized covariance matrices. *Journal of Econometrics*, **192**(1), 19 – 39.

Jochmann, M. (2015). Modeling U.S. inflation dynamics: A bayesian nonparametric approach. *Econometric Reviews*, **34**(5), 537–558.

Kim, C.-J., Morley, J. C., and Nelson, C. R. (2004). Is there a positive relationship between stock market volatility and the equity premium? *Journal of Money, Credit and Banking*, **36**(3), pp. 339–360.

Kose, M. A., Otrok, C., and Whiteman, C. H. (2003). International business cycles: World, region, and country-specific factors. *American Economic Review*, **93**(4), 1216–1239.

Lunde, A. and Timmermann, A. G. (2004). Duration dependence in stock prices: An analysis of bull and bear markets. *Journal of Business & Economic Statistics*, **22**(3), 253–273.

Maheu, J. M. and McCurdy, T. H. (2000). Identifying bull and bear markets in stock returns. *Journal of Business & Economic Statistics*, **18**(1), 100–112.

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. Realized Volatility.

Maheu, J. M. and Yang, Q. (2016). An infinite hidden Markov model for short-term interest rates. *Journal of Empirical Finance*, **38**, 202–220.

Maheu, J. M., McCurdy, T. H., and Song, Y. (2012). Components of bull and bear markets: Bull corrections and bear rallies. *Journal of Business & Economic Statistics*, **30**(3), 391–403.

Markowitz, H. (1952). Mean-variance analysis in portfolio choice and financial markets. *The Journal of Finance*, **7**(1), 77–91.

Noureldin, D., Shephard, N., and Sheppard, K. (2012). Multivariate high-frequency-based volatility (HEAVY) models. *Journal of Applied Econometrics*, **27**(6), 907–933.

Pastor, L. and Stambaugh, R. F. (2001). The equity premium and structural breaks. *The Journal of Finance*, **56**(4), 1207–1239.

Rydén, T., Teräsvirta, T., and Åsbrink, S. (1998). Stylized facts of daily return series and the hidden Markov model. *Journal of Applied Econometrics*, **13**(3), 217–244.

Schwert, G. W. (1990). Indexes of u.s. stock prices from 1802 to 1987. *Journal of Business*, **63**(3), 399–426.

Shephard, N. and Sheppard, K. (2010). Realising the future: forecasting with high-frequency-based volatility (HEAVY) models. *Journal of Applied Econometrics*, **25**(2), 197–231.

Skouras, S. (2007). Decisionmetrics: A decision-based approach to econometric modelling. *Journal of Econometrics*, **137**(2), 414 – 440.

Song, Y. (2014). Modelling regime switching and structural breaks with an infinite hidden markov model. *Journal of Applied Econometrics*, **29**(1), 825–842.

Stock, J. and Watson, M. (2010). *Dynamic Factor Models*. Oxford University Press, Oxford.

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.

Teh, Y. W., Jordan, M. I., Beal, M. J., and Blei, D. M. (2006). Hierarchical dirichlet processes. *Journal of the American Statistical Association*, **101**(476), pp. 1566–1581.

Walker, S. G. (2007). Sampling the dirichlet mixture model with slices. *Communications in Statistics - Simulation and Computation*, **36**(1), 45–54.

Zellner, A. (1971). *An introduction to Bayesian inference in econometrics*. John Wikey and Sons.

Table 2.1: Prior Specifications of Univariate Return Models

Panel A: Priors for MS and Joint MS Models

Model	μ_{s_t}	$\sigma^2_{s_t}$	ν	$\delta^2_{s_t}$	P_j
MS	$\mathrm{N}(0,1)$	$\mathrm{IG}(2, \widehat{\mathrm{var}}(r_t))$	-		$\mathrm{Dir}(1,\ldots,1)$
MS-RV	$\mathrm{N}(0,1)$	$\mathrm{G}(\overline{RV_t}, 1)$	$\mathrm{IG}(2,1)$		$\mathrm{Dir}(1,\ldots,1)$
MS-RAV	$\mathrm{N}(0,1)$	$\mathrm{G}(\overline{RV_t}, 1)$	$\mathrm{IG}(2,1)$		$\mathrm{Dir}(1,\ldots,1)$
MS-logRV	$\mathrm{N}(0,1)$	$\mathrm{N}(\overline{\log(RV_t)}, 5)$	-	$\mathrm{IG}(2, 0.5)$	$\mathrm{Dir}(1,\ldots,1)$
MS-logRAV	$\mathrm{N}(0,1)$	$\mathrm{N}(\overline{\log(RAV_t)}, 5)$	-	$\mathrm{IG}(2, 0.5)$	$\mathrm{Dir}(1,\ldots,1)$

Panel B: Priors for IHMM and Joint IHMM Models

Model	μ_{s_t}	$\sigma^2_{s_t}$	ν	$\delta^2_{s_t}$	η	α
IHMM	$\mathrm{N}(0,1)$	$\mathrm{IG}(2, \widehat{\mathrm{var}}(r_t))$	-	-	$\mathrm{G}(1,4)$	$\mathrm{G}(1,4)$
IHMM-RV	$\mathrm{N}(0,1)$	$\mathrm{G}(\overline{RV_t}, 1)$	$\mathrm{IG}(2,1)$	-	$\mathrm{G}(1,4)$	$\mathrm{G}(1,4)$
IHMM-logRV	$\mathrm{N}(0,1)$	$\mathrm{N}(\overline{\log(RV_t)}, 5)$	-	$\mathrm{IG}(2, 0.5)$	$\mathrm{G}(1,4)$	$\mathrm{G}(1,4)$
IHMM-logRAV	$\mathrm{N}(0,1)$	$\mathrm{N}(\overline{\log(RAV_t)}, 5)$	-	$\mathrm{IG}(2, 0.5)$	$\mathrm{G}(1,4)$	$\mathrm{G}(1,4)$

$\widehat{\mathrm{var}}(r_t)$ is the sample variance, $\overline{RV_t}$, $\overline{\log(RV_t)}$ and $\overline{\log(RAV_t)}$ are the sample means. All are computed using in-sample data.

Table 2.2: Summary Statistics for Monthly Equity Returns and Volatility Measures

Data	Mean	Median	Stdev	Skewness	Kurtosis	Min	Max
r_t	0.047	0.097	0.612	-0.539	9.123	-4.154	3.884
RV_t	0.328	0.156	0.621	6.853	68.499	0.010	8.580
RAV_t	0.470	0.394	0.287	2.807	14.358	0.103	2.747
$\log(RV_t)$	-1.720	-1.856	0.964	0.714	3.992	-4.608	2.149
$\log(RAV_t)$	-0.882	-0.931	0.476	0.682	3.869	-2.274	1.010

This table reports the summary statistics for monthly returns and various ex-post proxies of volatility. See the text for definitions. The sample period is from March 1885 to December 2013 and the number of observations is 1542. (Note: Market closed between July 1914 and December 1914 due to World War I).

Table 2.3: Equity Forecasts: Jan. 1951 - Dec. 2013

No. of States	Models	Log-predictive Likelihoods	RMSE$[r_{t+1}]$	RMSE$[RV_{t+1}]$
2 States	MS	-548.409	0.5268	0.5285
	MS-RV	-535.003	0.5242*	0.5338
	MS-logRV	-534.914	0.5276	0.5229
	MS-RAV	-533.370*	0.5263	0.5263
	MS-logRAV	-534.256	0.5269	0.5199*
3 States	MS	-538.437	0.5244*	0.5240
	MS-RV	-523.000*	0.5290	0.5070
	MS-logRV	-524.754	0.5286	0.5087
	MS-RAV	-523.171	0.5276	0.5032*
	MS-logRAV	-525.353	0.5283	0.5048
4 States	MS	-535.454	0.5232*	0.5193
	MS-RV	-520.363*	0.5273	0.5029
	MS-logRV	-528.631	0.5284	0.4902*
	MS-RAV	-527.708	0.5277	0.4976
	MS-logRAV	-530.697	0.5290	0.4920
-	IHMM	-535.165	0.5229	0.5348
	IHMM-RV	**-514.662**	**0.5216**	0.4724
	IHMM-logRV	-516.643	0.5228	**0.4647**
	IHMM-logRAV	-517.148	0.5244	0.4775

This table reports the sum of 1-period ahead log-predictive likelihoods of return $\sum_{j=t+1}^{T} \log(p(r_j|y_{1:j-1}, \text{Model}))$, root mean squared error for return and realized variance predictions over period from Jan 1951 to Dec 2013 (756 observations). The symbol $*$ identifies the largest log-predictive likelihood and the smallest RMSE given a fixed number of states. Bold text indicates the largest log-predictive likelihood and the smallest RMSE in the entire column.

Table 2.4: Equity Forecasts: Jan. 1984 - Dec. 2013

No. of States	Models	Log-Predictive Likelihoods	RMSE$[r_{t+1}]$	RMSE$[RV_{t+1}]$
2 States	MS	-300.019	0.5542	0.6863
	MS-RV	-293.311	0.5512*	0.7130
	MS-logRV	-291.794	0.5563	0.6938*
	MS-RAV	-290.353*	0.5543	0.7050
	MS-logRAV	-290.709	0.5553	0.6968
3 States	MS	-294.914	0.5522*	0.6817
	MS-RV	-283.877	0.5570	0.6794
	MS-logRV	-284.135	0.5568	0.6781
	MS-RAV	-281.126*	0.5556	0.6764*
	MS-logRAV	-282.150	0.5561	0.6772
4 States	MS	-292.397	0.5506*	0.6755
	MS-RV	-281.211*	0.5553	0.6749
	MS-logRV	-285.383	0.5570	0.6564*
	MS-RAV	-282.523	0.5559	0.6696
	MS-logRAV	-284.563	0.5580	0.6614
-	IHMM	-291.091	0.5529	0.7002
	IHMM-RV	**-279.504**	**0.5475**	0.6344
	IHMM-logRV	-281.344	0.5503	**0.6209**
	IHMM-logRAV	-280.019	0.5529	0.6434

This table reports the sum of 1-period ahead log-predictive likelihoods of return $\sum_{j=t+1}^{T} \log(p(r_j|y_{1:j-1}, \text{Model}))$, root mean squared error for return and realized variance predictions over period from Jan 1984 to Dec 2013 (360 observations). The symbol $*$ identifies the largest log-predictive likelihood and the smallest RMSE given a fixed number of states. Bold text indicates the largest log-predictive likelihood and the smallest RMSE in the entire column.

Table 2.5: Log-Predictive Bayes Factors for Market Declines

Market Declines	Period	$\sum \log \frac{p(r_{t+1}\|r_{1:t},\text{IHMM-RV})}{p(r_{t+1}\|r_{1:t},\text{IHMM})}$
1973-74 stock market crash	Feb. 1973 - Dec. 1974	0.2878
Black Monday	Oct. 1987 - Dec. 1987	1.4915
Dot-com bubble	Jan. 2000 - Dec. 2002	2.3788
Financial crisis of 2007-08	Jul. 2007 - Dec. 2008	1.5615

This table reports log-predictive Bayes factors for the IHMM-RV model versus the IHMM over several sample periods.

Table 2.6: Estimates for Stock Market Returns

	MS		MS-RV		MS-RAV	
Parameter	Mean	Stdev	Mean	Stdev	Mean	Stdev
μ_1	-0.2995 (-0.528, -0.089)	0.1104	-0.0840 (-0.159, -0.020)	0.0354	-0.1372 (-0.229, -0.050)	0.0445
μ_2	0.0875 (0.060, 0.116)	0.0140	0.1224 (0.096, 0.149)	0.0139	0.1130 (0.089, 0.138)	0.0125
σ_1^2	1.6127 (1.191, 2.217)	0.2630	0.5633 (0.481, 0.678)	0.0486	0.6581 (0.597, 0.748)	0.0394
σ_2^2	0.2187 (0.196, 0.241)	0.0115	0.1556 (0.143, 0.171)	0.0049	0.1460 (0.138, 0.154)	0.0044
ν	-	-	1.3431 (1.183, 1.501)	0.0824	2.1529 (2.009, 2.316)	0.0758
$P_{1,1}$	0.8775 (0.793, 0.943)	0.0396	0.9023 (0.862, 0.937)	0.0193	0.8716 (0.828, 0.910)	0.0210
$P_{2,2}$	0.9849 (0.972, 0.994)	0.0058	0.9442 (0.924, 0.962)	0.0099	0.9538 (0.937, 0.969)	0.0083

This table reports the posterior mean, standard deviation and 0.95 density intervals (values in brackets) of parameters of selected 2 state models. The prior restriction $\mu_1 < 0$ and $\mu_2 > 0$ is imposed. The sample period is from March 1885 to December 2013 (1542 observations).

Table 2.7: Performance of Market Timing Portfolios

Panel A: Jan. 1951 - Dec. 2013						
Number of states	Model	Mean	St. Dev.	Sharpe Ratio	Δ_q	Δ_e
Buy-Hold		0.0826	0.5148	0.1604	-	-
2 States	MS	0.0663	0.4198	0.1578	-	-
	MS-RV	0.0720	0.3929	**0.1831**	**92.256**	**96.192**
4 States	MS	0.0690	0.4427	0.1560	-7.188	-4.512
	MS-RV	0.0698	0.4040	**0.1729**	**55.860**	**60.372**
-	IHMM	0.0657	0.4414	0.1487	-37.608	-37.884
	IHMM-RV	0.0770	0.4348	**0.1770**	**82.392**	**87.888**
Panel B: Jan. 1984 - Dec. 2013						
Number of states	Model	Mean	St. Dev.	Sharpe Ratio	Δ_q	Δ_e
Buy-Hold		0.0911	0.5408	0.1684	-	-
2 States	MS	0.0564	0.4755	0.1187	-	-
	MS-RV	0.0772	0.3909	**0.1975**	**325.536**	**337.728**
4 States	MS	0.0684	0.4845	0.1412	100.476	108.540
	MS-RV	0.0645	0.4291	**0.1502**	**145.476**	**156.852**
-	IHMM	0.0563	0.4975	0.1132	-37.212	-37.692
	IHMM-RV	0.0794	0.4590	**0.1729**	**247.956**	**261.276**

The summary statistics are based on annualized returns. The values in the last two columns are annualized basis point performance fees that an investor is willing to pay to switch from the 2-state Markov switching model. The risk aversion coefficient is $\gamma = 5$. Bold numbers indicate the largest Sharpe ratio and the largest performance fee for a class of models.

Table 2.8: Prior Specification of Multivariate Models

Panel A: Priors for Multivariate MS Models					
Model	M_{s_t}	Σ_{s_t}	κ	P_j	
MS	$\mathrm{N}(\mathbf{0}, 5\mathbf{I})$	$\mathrm{IW}(\widehat{\mathrm{Cov}}(R_t), 5)$	-	$\mathrm{Dir}(1, \ldots, 1)$	
MS-RCOV	$\mathrm{N}(\mathbf{0}, 5\mathbf{I})$	$\mathrm{W}(\frac{1}{3}\overline{RCOV_t}, 3)$	$\mathrm{G}(20, 1)\mathbb{1}_{\kappa>4}$	$\mathrm{Dir}(1, \ldots, 1)$	
Panel B: Priors for Multivariate IHMM Models					
Model	M_{s_t}	Σ_{s_t}	κ	η	α
IHMM	$\mathrm{N}(\mathbf{0}, 5\mathbf{I})$	$\mathrm{IW}(\widehat{\mathrm{Cov}}(R_t), 5)$	-	$\mathrm{G}(1, 4)$	$\mathrm{G}(1, 4)$
IHMM-RCOV	$\mathrm{N}(\mathbf{0}, 5\mathbf{I})$	$\mathrm{W}(\frac{1}{3}\overline{RCOV_t}, 3)$	$\mathrm{G}(20, 1)\mathbb{1}_{\kappa>4}$	$\mathrm{G}(1, 4)$	$\mathrm{G}(1, 4)$

$\mathbf{0}$ denotes zero vector, $\mathbf{I}$ is the identity matrix. $\widehat{\mathrm{Cov}}(R_t)$, and $\overline{RCOV_t}$ are computed using in sample data.

Table 2.9: Summary Statistics of Returns (IBM, XOM, GE)

Panel A: Summary of Returns							
Data	Mean	Median	St. Dev	Skewness	Kurtosis	Min	Max
IBM	0.134	0.135	0.824	-0.192	5.169	-3.644	3.635
XOM	0.116	0.096	0.707	-0.152	6.942	-3.930	3.773
GE	0.103	0.088	0.940	-0.324	7.755	-5.265	5.336

Panel B: Return Covariance and RCOV Mean						
	Covariance of Return			Average of RCOV		
Data	IBM	XOM	GE	IBM	XOM	GE
IBM	0.678	0.236	0.411	0.742	0.250	0.370
XOM	0.236	0.500	0.338	0.250	0.641	0.394
GE	0.411	0.338	0.882	0.370	0.394	0.970

The panel A of above table reports the summary statistics of the monthly return of IBM, XOM and GE. The reported data are annualized values after scaling the raw returns by 12. The panel B reports the covariance matrix calculated from monthly return vectors and the averaged RCOV matrix, which are calculated using daily returns. The sample period is from Jan 1926 to Dec 2013 (1056 observations).

Table 2.10: Multivariate Equity Forecasts

No. of States	Models	Log Predictive Likelihoods	$\|\|\text{RMSE}(R_{t+1})\|\|$	$\|\|\text{RMSE}(RCOV_{t+1})\|\|$
8 States	MS	-2294.571	**1.2843**	2.4230
	MS-RCOV	-2264.490*	1.2857	2.2049*
12 States	MS	-2315.345	1.2849*	2.4979
	MS-RCOV	-2270.969*	1.2856	2.2320*
-	IHMM	-2274.063	1.2877	2.3651
	IHMM-RCOV	**-2262.383**	1.2873*	**2.1956**

This table summarizes the sum of 1 month log predictive likelihoods of return, $\sum_{j=t+1}^{T} \log(p(R_j|y_{1:j-1}, \text{Model}))$, root mean squared errors of mean and covariance prediction over Jan 1951 to Dec 2013 (totally 756 predictions), when the models are applied to analyze IBM, XOM, GE jointly. The root mean squared errors provided in this table are matrix norms. $||A|| = \sqrt{\sum_i \sum_j a_{ij}^2}$. The symbol $*$ identifies the largest log-predictive likelihood and the smallest RMSE given a fixed number of states. Bold text indicates the largest log-predictive likelihood and the smallest RMSE in the entire column.

Table 2.11: Performance of Minimum Variance Portfolios (Jan. 1951 - Dec. 2013)

Panel A: Required Annualized Portfolio Return: 10%						
		Mean	St. Dev.	Sharpe Ratio	Δ_q	Δ_e
8 States	MS	0.1103	0.3296	0.3345	-	-
	MS-RCOV	0.1090	0.3155	**0.3455**	**7.080**	**3.888**
12 States	MS	0.1103	0.3369	0.3273	-10.704	-10.188
	MS-RCOV	0.1084	0.3137	**0.3456**	**3.744**	**0.900**
-	IHMM	0.1081	0.3257	0.3320	-15.924	-16.224
	IHMM-RCOV	0.1089	0.3170	**0.3436**	**4.188**	**0.600**
Panel B: Required Annualized Portfolio Return: 20%						
		Mean	St. Dev.	Sharpe Ratio	Δ_q	Δ_e
8 States	MS	0.1956	0.8916	0.2194	-	-
	MS-RCOV	0.1962	0.8660	**0.2265**	**106.752**	**83.784**
12 States	MS	0.1934	0.9161	0.2111	-123.468	-116.364
	MS-RCOV	0.1970	0.8626	**0.2284**	**128.772**	**103.104**
-	IHMM	0.1933	0.8715	0.2218	56.940	45.876
	IHMM-RCOV	0.1985	0.8695	**0.2283**	**116.520**	**90.996**

The summary statistics are based on annualized return (scaled by 12). Δ_q is annualized basis point performance fees that an investor with quadratic utility is willing to pay to switch from the 8-state MS model. Δ_e is the fee for investor with exponential utility. The risk aversion coefficients are both utility function are $\gamma = 5$. Panel A and B shows the results of portfolio with 10% and 20% required return, respectively. Out of sample period: Jan 1951 - Dec 2013, totally 756 months. Bold numbers indicate the largest Sharpe ratio and the largest performance fee for a class of models.

Table 2.12: Performance of Minimum Variance Portfolios (Jan. 1984 - Dec. 2013)

Panel A: Required Annualized Portfolio Return: 10%						
		Mean	St. Dev.	Sharpe Ratio	Δ_q	Δ_e
8 States	MS	0.0907	0.3370	0.2692	-	-
	MS-RCOV	0.0925	0.3239	**0.2856**	**36.612**	**33.060**
12 States	MS	0.0889	0.3491	0.2545	-36.708	-36.336
	MS-RCOV	0.0938	0.3222	**0.2912**	**51.936**	**48.348**
-	IHMM	0.0936	0.3354	0.2791	30.972	**33.420**
	IHMM-RCOV	0.0925	0.3232	**0.2860**	**36.948**	32.880
Panel B: Required Annualized Portfolio Return: 20%						
		Mean	St. Dev.	Sharpe Ratio	Δ_q	Δ_e
8 States	MS	0.1788	0.9140	0.1956	-	-
	MS-RCOV	0.1833	0.8934	**0.2052**	**128.520**	**99.360**
12 States	MS	0.1728	0.9542	0.1811	-229.368	-217.260
	MS-RCOV	0.1869	0.8857	**0.2110**	**195.588**	**161.340**
-	IHMM	0.1871	0.8966	**0.2087**	**153.588**	**154.812**
	IHMM-RCOV	0.1831	0.8967	0.2042	113.532	75.228

The summary statistics are based on annualized return (scaled by 12). Δ_q is annualized basis point performance fees that an investor with quadratic utility is willing to pay to switch from the 8-state MS model. Δ_e is the fee for investor with exponential utility. The risk aversion coefficients for both utility function are $\gamma = 5$. Panel A and B shows the results of portfolio with 10% and 20% required return, respectively. Out of sample period: Jan 1984 - Dec 2013, totally 360 months. Bold numbers indicate the largest Sharpe ratio and the largest performance fee for a class of models.

Table 2.13: Return Variances of Global Minimum Variance Portfolios

Panel A: Jan. 1951 - Dec. 2013		
No. of States	Model	Variance
8 States	MS	0.31123
	MS-RCOV	**0.29601**
12 States	MS	0.32271
	MS-RCOV	**0.29191**
-	IHMM	0.30325
	IHMM-RCOV	**0.29675**
Panel B: Jan. 1984 - Dec. 2013		
No. of States	Model	Variance
8 States	MS	0.31498
	MS-RCOV	**0.31112**
12 States	MS	0.33399
	MS-RCOV	**0.30378**
-	IHMM	0.31098
	IHMM-RCOV	**0.31047**

This table reports variances of annualized return of global minimum variance portfolios based on competing models. Bold numbers indicate the smallest portfolio variance for a class of models.

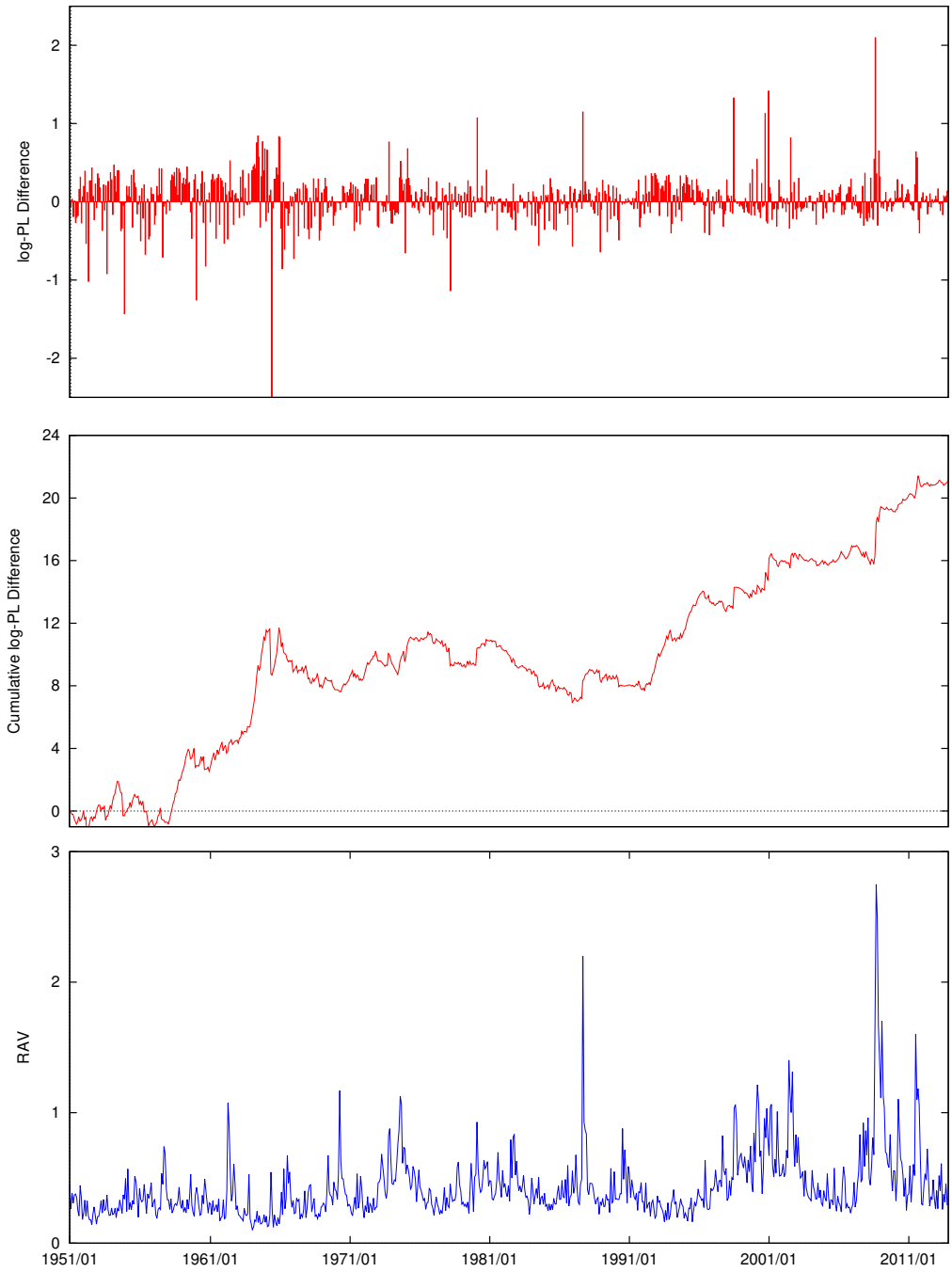

Figure 2.1: The top panel shows $\log\left(p(r_{t+1}|y_{1:t}, \text{IHMM-RV})\right) - \log\left(p(r_{t+1}|y_{1:t}, \text{IHMM})\right)$ over January 1984 to December 2013. The second panel plots the cumulative log-predictive likelihood difference. The final panel is the time series plot of realized absolute variation.

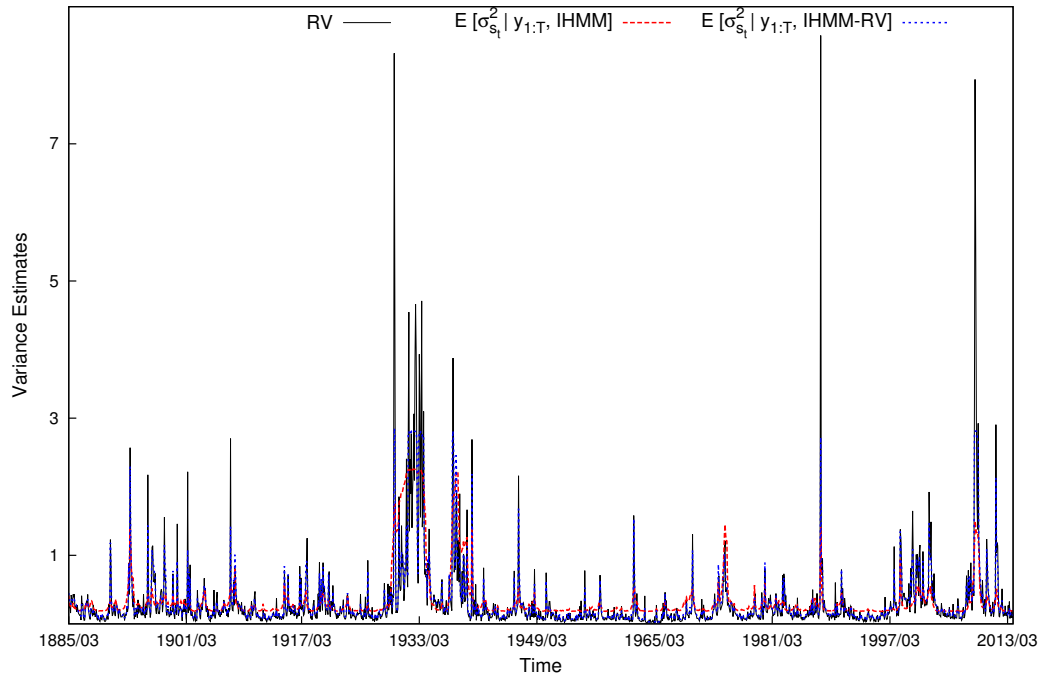

Figure 2.2: $E\left[\sigma_{s_t}^2 \middle| y_{1:T}, \text{IHMM}\right]$, $E\left[\sigma_{s_t}^2 \middle| y_{1:T}, \text{IHMM-RV}\right]$ and Realized Variance

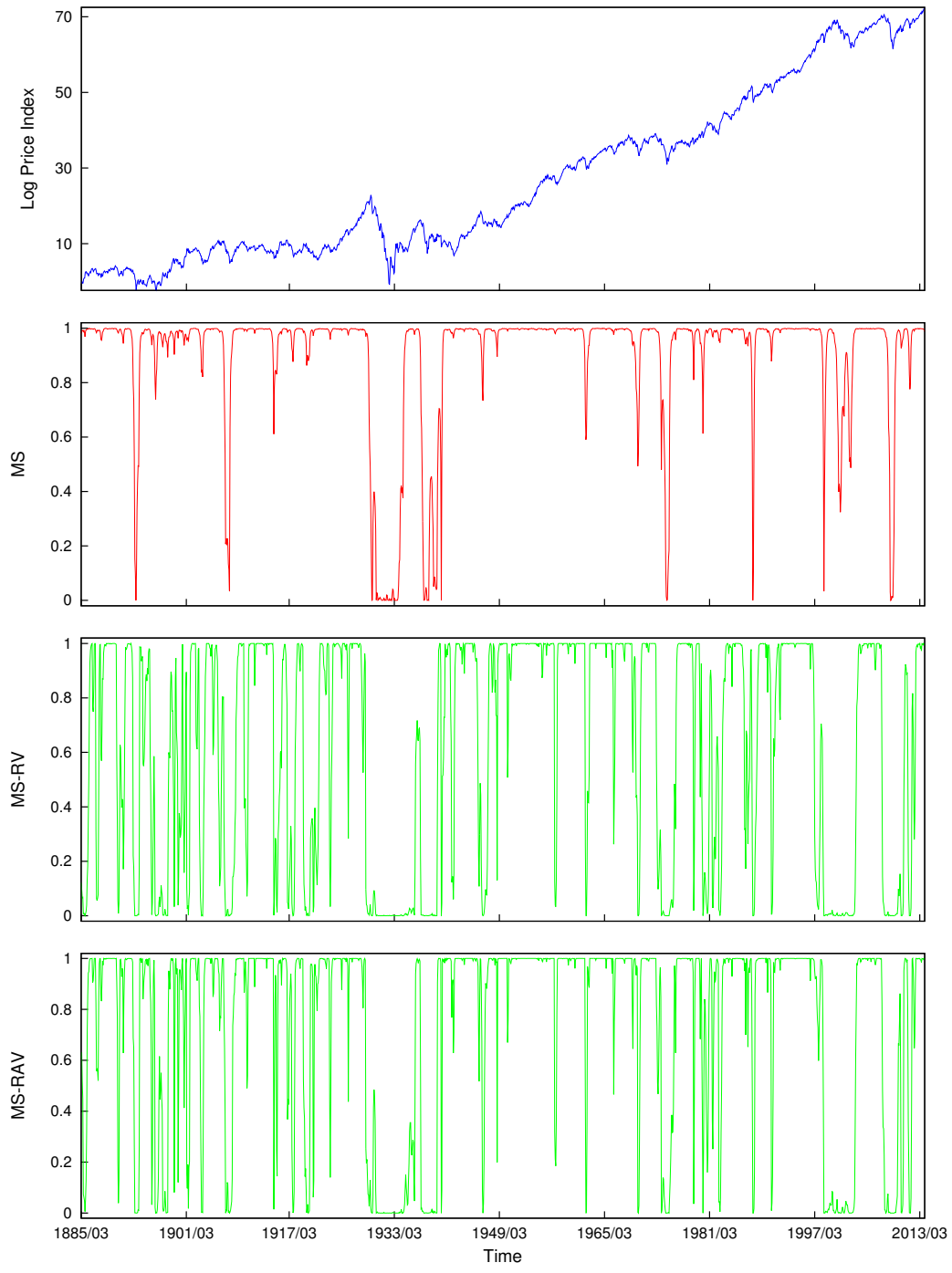

Figure 2.3: Smoothed Probability of High Return State

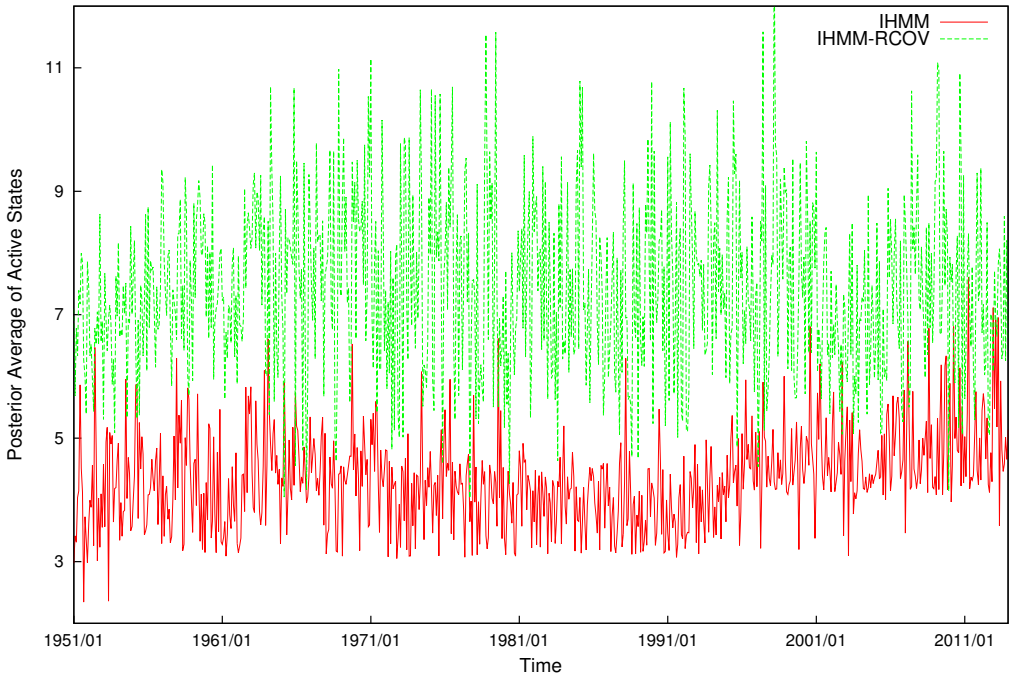

Figure 2.4: Number of Active Clusters: IHMM and IHMM-RCOV Models

Chapter 3

Bayesian Nonparametric Estimation of Ex-post Variance

3.1 Introduction

This chapter introduces a new method of estimating ex-post volatility from high-frequency data using a Bayesian nonparametric model. In contrast to existing classical estimation methods, such as those applied in Chapter 2, the proposed method allows the data to cluster under a flexible framework and delivers an exact finite sample distribution for the ex-post variance or transformations of the variance. Bayesian nonparametric variance estimators under no noise, heteroskedastic and serially correlated microstructure noise are proposed.

Volatility is an indispensable quantity in finance and is a key input into asset pricing, risk management and portfolio management. In the last two decades, researchers have taken advantage of high-frequency data to estimate ex-post variance using intraperiod returns. Barndorff-Nielsen and Shephard (2002) and Andersen *et al.* (2003)

formalized the idea of using high frequency data to measure the volatility of lower frequency returns. They show that realized variance (RV) is a consistent estimator of quadratic variation under ideal conditions. Unlike parametric models of volatility in which the model specification is important, RV is a *model free* estimate of quadratic variation in that it is valid under a wide range of spot volatility dynamics.[1]

RV provides an accurate measure of ex-post variance if there is no market microstructure noise. However, observed prices at high-frequency are inevitably contaminated by noise in reality and returns are no longer uncorrelated. In this case, RV is a biased and inconsistent estimator Hansen and Lunde (2006); Aït-Sahalia *et al.* (2011). The impact of market microstructure noise on forecasting is explored in Aït-Sahalia and Mancini (2008) and Andersen *et al.* (2011).

Several different approaches have been proposed to estimating ex-post variance under microstructure noise. Zhou (1996) first introduced the idea of using a kernel-based method to estimate ex-post variance. Barndorff-Nielsen *et al.* (2008) formally discussed the realized kernel and showed how to use it in practice in a later paper (Barndorff-Nielsen *et al.* (2009)). Another approach is the subsampling method of Zhang *et al.* (2005). Hansen *et al.* (2008) showed how a time-series model can be used to filter out market microstructure to obtain corrected estimates of ex-post variance. A robust version of the predictive density of integrated volatility is derived in Corradi *et al.* (2009). Although bootstrap refinements are explored in Goncalves and Meddahi (2009) all distributional results from this literature rely on in-fill asymptotics.

Our Bayesian approach introduces a new concept to this problem, pooling. The

[1]For a good survey of the key concepts see Andersen and Benzoni (2008), for an in-depth treatment see Aït-Sahalia and Jacod (2014).

existing ex-post variance estimators treat the information on variance from all intraperiod returns independently. However, the variance of intraperiod returns may be the same at different time periods. Pooling observations with common variance level may be beneficial to daily variance estimation.

We model intraperiod returns according to a Dirichlet process mixture (DPM) model. This is a countably infinite mixture of distributions which facilitates the clustering of return observations into distinct groups sharing the same variance parameter. The DPM model became popular for density estimation following the introduction of Markov chain Monte Carlo (MCMC) techniques by Escobar and West (1994). Estimation of these models is now standard with several alternatives available, see Neal (2000) and Kalli *et al.* (2011). Our proposed method benefits variance estimation in at least two aspects. First, the common values of intraperiod variance can be pooled into the same group leading to a more precise estimate. The pooling is done endogenously along with estimation of other model parameters. Second, the Bayesian nonparametric model delivers exact finite inference regarding ex-post variance or transformations such as the logarithm. As such, uncertainty around the estimate of ex-post volatility is readily available from the predictive density. Unlike the existing asymptotic theory which may give confidence intervals that contain negative values for variance, density intervals are always on the positive real line and can accommodate asymmetry.

By extending key results in Hansen *et al.* (2008) we adapt the DPM models to deal with returns contaminated with heteroskedastic noise and serially correlated noise.

Monte Carlo simulation results show the Bayesian approach to be a very competitive alternative. Overall, pooling can lead to more precise estimates of ex-post variance and better coverage frequencies. We show that the new variance estimators

can be used with confidence and effectively recover both the average statistical features of daily ex-post variance as well as the time-series properties. Two applications to real world data with comparison to realized variance and kernel-based estimators are included.

This chapter is organized as follows. In Section 3.2, we provide a brief review of some existing variance estimators which serve as the benchmarks for later comparison. The Bayesian nonparametric model, daily variance estimator and model estimation methods are discussed in Section 3.3. Section 3.4 extends the Bayesian nonparametric model to deal with heteroskedastic and serially correlated microstructure noise. Section 3.5 provides an extensive simulation and comparison of the estimators. Applications to IBM and Disney data are found in Section 3.6. Section 3.7 concludes followed by an appendix.

3.2 Existing Ex-post Volatility Estimation

3.2.1 Realized Variance

Realized variance (RV), which equals the summation of squared intraperiod returns, is the most commonly used ex-post volatility measurement. Andersen *et al.* (2003) and Barndorff-Nielsen and Shephard (2002) formally studied the properties of RV and show it is a consistent estimator of quadratic variation under no microstructure noise. We will focus on variance estimation over a day t but all of the results apply to other time intervals.

Under the assumption of frictionless market and semimartingale, considering the

following log-price diffusion,

$$\mathrm{d}p(t) = \mu(t)\,\mathrm{d}t + \sigma(t)\,\mathrm{d}W(t), \tag{3.1}$$

where $p(t)$ denotes the log-price at time t, $\mu(t)$ is the drift term, $\sigma^2(t)$ is the spot variance and $W(t)$ a standard Brownian motion. If the price process contains no jump, the variation of the return over $t-1$ to t is measured by IV_t,

$$IV_t = \int_{t-1}^{t} \sigma^2(\tau)\mathrm{d}\tau. \tag{3.2}$$

Let $r_{t,i}$ denotes the i^{th} intraday return on day t, $i = 1, \ldots, n_t$, where n_t is the number of intraday returns on day t. Realized variance is defined as

$$RV_t = \sum_{i=1}^{n_t} r_{t,i}^2, \tag{3.3}$$

and $RV_t \xrightarrow{p} IV_t$, as $n_t \to \infty$ (Andersen *et al.*, 2001a).

Barndorff-Nielsen and Shephard (2002) derive the asymptotic distribution of RV_t as

$$\sqrt{n_t}\frac{1}{\sqrt{2IQ_t}}(RV_t - IV_t) \xrightarrow{d} \mathrm{N}(0,1), \qquad \text{as} \quad n_t \to \infty, \tag{3.4}$$

where IQ_t stands for the integrated quarticity, which can be estimated by realized quarticity (RQ_t) defined as

$$RQ_t = \frac{n_t}{3}\sum_{i=1}^{n_t} r_{t,i}^4 \xrightarrow{p} IQ_t, \qquad \text{as} \quad n_t \to \infty. \tag{3.5}$$

3.2.2 Flat-top Realized Kernel

If returns are contaminated with microstructure noise, RV_t will be biased and inconsistent (Zhang *et al.*, 2005; Hansen and Lunde, 2006; Bandi and Russell, 2008). The observed log-price $\tilde{p}_{t,i}$, is assumed to follow

$$\tilde{p}_{t,i} = p_{t,i} + \epsilon_{t,i}, \tag{3.6}$$

where $p_{t,i}$ is the true but latent log-price and $\epsilon_{t,i}$ is a noise term which is independent of the price.

Barndorff-Nielsen *et al.* (2008) introduced the flat-top realized kernel (RK_t^F), which is the optimal estimator if the microstructure error is a white noise process[2].

$$RK_t^F = \sum_{i=1}^{n_t} \tilde{r}_{t,i}^2 + \sum_{h=1}^{H} k\left(\frac{h-1}{H}\right)(\gamma_{-h} + \gamma_h), \qquad \gamma_h = \sum_{i=1}^{n_t} \tilde{r}_{t,i}\tilde{r}_{t,i-h}, \tag{3.7}$$

where H is the bandwidth, $k(x)$ is a kernel weight function.

The preferred kernel function is the second order Tukey-Hanning kernel[3] and the preferred bandwidth is $H^* = c\xi\sqrt{n_t}$, where $\xi^2 = \omega^2/\sqrt{IQ_t}$ denotes the noise-to-signal ratio. ω^2 stands for the variance of microstructure noise and can be estimated by $RV_t/(2n_t)$ by Bandi and Russell (2008). RV_t based on 10-minute returns is less sensitive to microstructure noise and can be used as a proxy of $\sqrt{IQ_t}$. $c = 5.74$ given Tukey-Hanning kernel of order 2.

Given the Tukey-Hanning kernel and $H^* = c\xi\sqrt{n_t}$, Barndorff-Nielsen *et al.* (2008)

[2]Another popular approach to dealing with noise is subsampling. See Zhang *et al.* (2005), Aït-Sahalia and Mancini (2008) for the Two Scales Realized Volatility (TSRV) estimator.

[3]Tukey-Hanning kernel with order 2: $k(x) = \sin^2\left[\frac{\pi}{2}(1-x)^2\right]$.

show that the asymptotic distribution of RK_t^F is

$$n_t^{1/4}\left(RK_t^F - IV_t\right) \xrightarrow{d} \text{MN}\left\{0, 4IQ_t^{3/4}\omega\left(ck_\bullet^{0,0} + 2c^{-1}k_\bullet^{1,1}\frac{IV_t}{\sqrt{IQ_t}} + c^{-3}k_\bullet^{2,2}\right)\right\}, \quad (3.8)$$

where MN is mixture of normal distribution, $k_\bullet^{0,0} = 0.219$, $k_\bullet^{1,1} = 1.71$ and $k_\bullet^{2,2} = 41.7$ for second order Tukey-Hanning kernel.

Even though ω^2 can be estimated using $RV_t/(2n_t)$, a better and less biased estimator suggested by Barndorff-Nielsen *et al.* (2008) is

$$\breve{\omega}^2 = \exp\left[\log(\hat{\omega}^2) - RK_t/RV_t\right]. \quad (3.9)$$

The estimation of IQ_t is more sensitive to the microstructure noise. The tri-power quarticity (TPQ_t) developed by Barndorff-Nielsen and Shephard (2006) can be used to estimate IQ_t,

$$TPQ_t = n_t\mu_{4/3}^{-3}\sum_{i=1}^{n_t-2}|\tilde{r}_{t,i}|^{4/3}|\tilde{r}_{t,i+1}|^{4/3}|\tilde{r}_{t,i+2}|^{4/3}, \quad (3.10)$$

where $\mu_{4/3} = 2^{2/3}\Gamma(7/6)/\Gamma(1/2)$. Replacing IV_t, ω^2 and IQ_t with RK_t^F, $\breve{\omega}^2$ and TPQ_t in equation (3.8), the asymptotic variance of RK_t^F can be calculated.

3.2.3 Non-negative Realized Kernel

The flat-top realized kernel discussed in previous subsection is based on the assumption that the error term is white noise. However, the white noise assumption is restrictive and the error term can be serial dependent or dependent on returns in reality. Another drawback of the RK_t^F is that it may provide negative volatility estimates,

albeit very rarely. Barndorff-Nielsen *et al.* (2011) further introduced the non-negative realized kernel (RK_t^N) which is more robust to these error term assumptions and is calculated as

$$RK_t^N = \sum_{h=-H}^{H} k\left(\frac{h}{H+1}\right)\gamma_h, \qquad \gamma_h = \sum_{i=|h|+1}^{n_t} \tilde{r}_{t,i}\tilde{r}_{t,i-|h|}. \tag{3.11}$$

The optimal choice of H is $H^* = c\xi^{4/5}n_t^{3/5}$ and the preferred kernel weight function is the Parzen kernel[4], which implies $c = 3.5134$. ξ^2 can be estimated using the same method as in the calculation of RK_t^F.

Barndorff-Nielsen *et al.* (2011) show the asymptotic distribution of RK_t^N based on $H^* = c\xi^{4/5}n_t^{3/5}$ is given by

$$n_t^{1/5}\left(RK_t^N - IV_t\right) \xrightarrow{d} \mathrm{MN}(\kappa, 4\kappa^2), \tag{3.12}$$

where $\kappa = \kappa_0(IQ_t\omega)^{2/5}$, $\kappa_0 = 0.97$ for Parzen kernel function, ω and IQ_t can be estimated using equation (3.9) and (3.10).

Note that RK_t^N is no longer a consistent estimator of IV_t and the rate of convergence is slower than that of RK_t^F. If the error term is white noise, RK_t^F is superior to RK_t^N, but RK_t^N is more robust to deviations from independent noise and is always positive.

[4]Parzen kernel function:

$$k(x) = \begin{cases} 1 - 6x^2 + 6x^3, & 0 \le x \le 1/2 \\ 2(1-x)^3, & 1/2 < x \le 1 \\ 0, & x > 1 \end{cases}$$

3.3 Bayesian Nonparametric Ex-post Variance Estimation

In this section, we introduce a Bayesian nonparametric ex-post volatility estimator. After defining the daily variance, conditional on the data, the discussion moves to the DPM model which provides the model framework of the proposed estimator. The approach discussed in this section deals with returns without microstructure noise and an estimator suitable for returns with microstructure noise is found in Section 3.4.

3.3.1 Model of High-frequency Returns

First we consider the case with no market microstructure noise. The model for log-returns is

$$r_{t,i} = \mu_t + \sigma_{t,i} z_{t,i}, \quad z_{t,i} \overset{iid}{\sim} \mathrm{N}(0,1),\ i = 1, \ldots, n_t, \tag{3.13}$$

where μ_t is constant in day t. The daily return is

$$r_t = \sum_{i=1}^{n_t} r_{t,i} \tag{3.14}$$

and it follows, conditional on the unknown realized volatility path $\mathcal{F}_t \equiv \{\sigma^2_{t,i}\}_{i=1}^{n_t}$, the ex-post variance is

$$V_t \equiv \mathrm{Var}(r_t|\mathcal{F}_t) = \sum_{i=1}^{n_t} \sigma^2_{t,i}. \tag{3.15}$$

In our Bayesian setting V_t is the target to estimate conditional on the data $\{r_{t,i}\}_{i=1}^{n_t}$. Note, that we make no assumptions on the stochastic process generating $\sigma_{t,i}^2$.

3.3.2 A Bayesian Model with Pooling

In this section we discuss a nonparametric prior for the model of (3.13) that allows for pooling over common values of $\sigma_{t,i}^2$. The Dirichlet process mixture model (DPM) is a Bayesian nonparametric mixture model that has been used in density estimation and for modeling unknown hierarchical effects among many other applications. A key advantage of the model is that it naturally incorporates parameter pooling.

Our nonparametric model has the following hierarchical form

$$
\begin{aligned}
r_{t,i}\big|\mu_t, \sigma_{t,i} &\overset{iid}{\sim} \mathrm{N}(\mu_t, \sigma_{t,i}^2),\ i = 1, \ldots, n_t, && (3.16)\\
\sigma_{t,i}^2\big|G_t &\overset{iid}{\sim} G_t, && (3.17)\\
G_t\big|G_{0,t}, \alpha_t &\sim \mathrm{DP}(\alpha_t, G_{0,t}), && (3.18)\\
G_{0,t} &\equiv \mathrm{IG}(v_{0,t}, s_{0,t}), && (3.19)
\end{aligned}
$$

where the base measure is the inverse-gamma distribution denoted as $\mathrm{IG}(v, s)$, which has a mean of $(s/v - 1)$ for $v > 1$. The return mean μ_t is assumed to be a constant over i.

The Dirichlet process was formally introduced by Ferguson (1973) and is a distribution over distributions. A draw from a $\mathrm{DP}(\alpha_t, G_{0,t})$ is an almost surely discrete distribution which is centered around the base distribution $G_{0,t}$. Therefore, a sample from $\sigma_{t,i}^2\big|G_t \sim G_t$ has a positive probability of repeated values. The concentration parameter $\alpha_t > 0$ governs how closely a draw G_t resembles $G_{0,t}$. Larger values of

α_t lead to G_t having the more unique atoms with significant weights. As $\alpha_t \to \infty$, $G_t \to G_{0,t}$ which implies that every $r_{t,i}$ has a unique $\sigma^2_{t,i}$ drawn from the inverse-gamma distribution. In this case there is no pooling and we have a setting very close to the classical counterpart discussed above. However, for finite α_t, pooling can take place. The other extreme is complete pooling for $\alpha_t \to 0$ in which there is one common variance shared by all observations such that $\sigma^2_{t,i} = \sigma^2_{t,1}$, $\forall i$. Since α_t plays an important role in pooling we place a prior on it and estimate it along with the other model parameters for each day.

A stick breaking representation (Sethuraman (1994)) of the DPM in (3.17) is given as follows.

$$p(r_{t,i}\big|\mu_t, \Psi_t, w_t) = \sum_{j=1}^{\infty} w_{t,j} \mathrm{N}(r_{t,i}|\mu_t, \psi^2_{t,j}), \tag{3.20}$$

$$w_{t,j} = v_{t,j} \prod_{l=1}^{j-1}(1 - w_{t,l}), \tag{3.21}$$

$$v_{t,j} \overset{iid}{\sim} \mathrm{Beta}(1, \alpha_t), \tag{3.22}$$

where $\mathrm{N}(\cdot|\cdot, \cdot)$ denotes the density of the normal distribution, $\Psi_t = \{\psi^2_{t,1}, \psi^2_{t,2} \ldots, \}$ is the set of unique values of $\sigma^2_{t,i}$, $w_t = \{w_{t,1}, w_{t,2}, \ldots, \}$ and $w_{t,j}$ is the weight associated with the j^{th} component. This formulation of the model facilitates posterior sampling which is discussed in the next section.

Since our focus is on intraday returns and the number of observations in a day can be small, especially for lower frequencies such as 5-minute. Therefore, the prior should be chosen carefully. It is straightforward to show that the prior predictive

distribution of $\sigma_{t,i}^2$ is $G_{0,t}$. For $\sigma_{t,i}^2 \sim \text{IG}(v_{0,t}, s_{0,t})$, the mean and variance of $\sigma_{t,i}^2$ are

$$\text{E}(\sigma_{t,i}^2) = \frac{s_{0,t}}{v_{0,t}-1} \quad \text{and} \quad \text{var}(\sigma_{t,i}^2) = \frac{s_{0,t}^2}{(v_{0,t}-1)^2(v_{0,t}-2)}. \tag{3.23}$$

Solving the two equations, the values of $v_{0,t}$ and $s_{0,t}$ are given by

$$v_{0,t} = \frac{\left[\text{E}(\sigma_{t,i}^2)\right]^2}{\text{var}(\sigma_{t,i}^2)} + 2 \quad \text{and} \quad s_{0,t} = \text{E}(\sigma_{t,i}^2)(v_{0,t}-1). \tag{3.24}$$

We use sample statistics $\widehat{\text{var}}(r_{t,i})$ and $\widehat{\text{var}}(r_{t,i}^2)$ calculated with three days intraday returns (day $t-1$, day t, and day $t+1$) to set the values of $\text{E}(\sigma_{t,i}^2)$ and $\text{var}(\sigma_{t,i}^2)$, then use equation (3.24) to find $v_{0,t}$ and $s_{0,t}$. A shrinkage prior $\text{N}(0, v^2)$ is used for μ_t since μ_t is expected to be close to zero. v^2 is small and adjusted according to the data frequency. Finally, $\alpha_t \sim \text{Gamma}(a, b)$.

For a finite dataset $i = 1, \ldots, n_t$ our target is the following posterior moment

$$E[V_t | \{r_{t,i}\}_{i=1}^{n_t}] = E\left[\sum_{i=1}^{n_t} \sigma_{t,i}^2 \middle| \{r_{t,i}\}_{i=1}^{n_t}\right]. \tag{3.25}$$

Note that the posterior mean of V_t can also be considered as the posterior mean of realized variance, $RV_t = \sum_{i=1}^{n_t} r_{t,i}^2$ assuming μ_t is small. As such, RV_t treats each $\sigma_{t,i}^2$ as separate and corresponds to no pooling. We discuss estimation of the model next.

3.3.3 Model Estimation

Estimation relies on Markov chain Monte Carlo (MCMC) techniques. We apply the slice sampler of Kalli *et al.* (2011), along with Gibbs sampling to estimate the DPM model. The slice sampler provides an elegant way to deal with the infinite states

in (3.20). It introduces an auxiliary variable $u_{t,1:n_t} = \{u_{t,1}, \ldots, u_{t,n_t}\}$ that randomly truncates the state space to a finite set at each MCMC iteration but marginally delivers draws from the desired posterior.

The joint distribution of $r_{t,i}$ and the auxiliary variable $u_{t,i}$ is given by

$$f\left(r_{t,i}, u_{t,i} | w_t, \mu_t, \Psi_t\right) = \sum_{j=1}^{\infty} \mathbb{1}\left(u_{t,i} < w_{t,j}\right) \mathrm{N}\left(r_{t,i} | \mu_t, \psi_{t,j}^2\right), \tag{3.26}$$

and integrating out $u_{t,i}$ recovers (3.20).

It is convenient to rewrite the model in terms of a latent state variable $s_{t,i} \in \{1, 2, \ldots\}$ that maps each observation to an associated component and parameter $\sigma_{t,i}^2 = \psi_{t,s_{t,i}}^2$. Observations with a common state share the same variance parameter. For a finite dataset the number of states (clusters) is finite and ordered from $1, \ldots, K$. Note that the number of clusters K, is not a fixed value over the MCMC iterations. A new cluster with variance $\psi_{t,K+1}^2 \sim G_{0,t}$ can be created if existing clusters do not fit that observation well and clusters sharing a similar variance can be merged into one.

The joint posterior is

$$p(\mu_t) \prod_{j=1}^{K} \left[p(\psi_{t,j}^2)\right] p(\alpha_t) \prod_{i=1}^{n_t} \mathbb{1}(u_{t,i} < w_{t,s_{t,i}}) \mathrm{N}(r_{t,i} | \mu_t, \psi_{t,s_{t,i}}^2). \tag{3.27}$$

Each MCMC iteration contains the following sampling steps.

1. $\pi\left(\mu_t | r_{t,1:n_t}, \{\psi_{t,j}^2\}_{j=1}^K, s_{t,1:n_t}\right) \propto p\left(\mu_t\right) \prod_{i=1}^{n_t} p\left(r_{t,i} | \mu_t, \psi_{t,s_{t,i}}^2\right)$.

2. $\pi\left(\psi_{t,j}^2 | r_{t,1:n_t}, s_{t,1:n_t}, \mu_t\right) \propto p\left(\psi_{t,j}^2\right) \prod_{t:s_{t,i}=j} p\left(r_{t,i} | \mu_t, \psi_{t,j}^2\right)$ for $j = 1, \ldots, K$.

3. $\pi\left(v_{t,j} | s_{t,1:n_t}\right) \propto \mathrm{Beta}\left(v_{t,j} | a_{t,j}, b_{t,j}\right)$ with $a_{t,j} = 1 + \sum_{i=1}^{n_t} \mathbb{1}\left(s_{t,i} = j\right)$ and $b_{t,j} =$

$\alpha_t + \sum_{i=1}^{n_t} \mathbb{1}(s_{t,i} > j)$ and update $w_{t,j} = v_{t,j} \prod_{l<j} (1 - v_{t,l})$ for $j = 1, \ldots, K$.

4. $\pi(u_{t,i}|w_{t,i}, s_{t,1:n_t}) \propto \mathbb{1}\left(0 < u_{t,i} < w_{t,s_{t,i}}\right)$.

5. Find the smallest K such that $\sum_{j=1}^{K} w_{t,j} > 1 - \min(u_{t,1:n_t})$.

6. $\pi\left(s_{t,i}|r_{1:n_t}, s_{t,1:n_t}, \mu_t, \{\psi_{t,j}^2\}_{j=1}^{K}, u_{t,1:n_t}, K\right) \propto \sum_{j=1}^{K} \mathbb{1}(u_{t,i} < w_{t,j})\, p\left(r_{t,i}, |\mu_t, \psi_{t,j}^2\right)$ for $i = 1, \ldots, n_t$.

7. $\pi(\alpha_t|K) \propto p(\alpha_t)\, p(K|\alpha_t)$.

In the first step μ_t is common to all returns and this is a standard Gibbs step given the conjugate prior. Step 2 is a standard Gibbs step for each variance parameter $\psi_{t,j}^2$ based on the data assigned to cluster j. The remaining steps are standard for slice sampling of DPM models. In 7, α_t is sampled based on Escobar and West (1994).

Steps 1-7 give one iteration of the posterior sampler. After dropping a suitable burn-in amount, M additional samples are collected, $\{\theta^{(m)}\}_{m=1}^{M}$, where $\theta = \{\mu_t, \psi_{t,1}^2, \ldots, \psi_{t,K}^2, s_{t,1:n_t}, \alpha_t\}$. Posterior moments of interest can be estimated from sample averages of the MCMC output.

3.3.4 Ex-post Variance Estimator

Conditional on the parameter vector θ the estimate of V_t is

$$E[V_t|\theta] = \sum_{i=1}^{n_t} \sigma_{t,s_i}^2. \tag{3.28}$$

The posterior mean of V_t is obtained by integrating out all parameter and distributional uncertainty. $E\left[V_t|\{r_{t,i}\}_{i=1}^{n_t}\right]$ is estimated as

$$\hat{V}_t = \frac{1}{M}\sum_{m=1}^{M}\sum_{i=1}^{n_t}\sigma_{t,i}^{2(m)}, \tag{3.29}$$

where $\sigma_{t,i}^{2(m)} = \psi_{t,s_{t,i}^{(m)}}^{2(m)}$. Similarly other features of the posterior distribution of V_t can be obtained. For instance, a (1-α) probability density interval for V_t is the quantiles of $\sum_{i=1}^{n_t}\sigma_{t,s_{t,i}}^2$ associated with probabilities $\alpha/2$ and $(1-\alpha/2)$. Conditional on the model and prior these are exact finite sample estimates, in contrast to the classical estimator which relies on infill asymptotics[5] to derived confidence intervals.

If $\log(V_t)$ is the quantity of interest, the estimator of $E\left[\log(V_t)|\{r_{t,i}\}_{i=1}^{n_t}\right]$ is given as

$$\widehat{\log(V_t)} = \frac{1}{M}\sum_{m=1}^{M}\log\left(\sum_{i=1}^{n_t}\sigma_{t,i}^{2(m)}\right). \tag{3.30}$$

As before, quantile estimates of the posterior of $\log(V_t)$ can be estimated from the MCMC output.

3.4 Bayesian Estimator Under Microstructure Error

An early approach to deal with market microstructure noise was to prefilter with a time-series model Andersen *et al.* (2001b); Bollen and Inder (2002); Maheu and McCurdy (2002). Hansen *et al.* (2008) shows that prefiltering results in a bias to realized variance that can be easily corrected. We employ these insights into moving average

[5]Infill asymptotics refers to the increasing rate of sampling within a fixed time interval.

specifications to account for noisy high-frequency returns. A significant difference is that we allow for heteroskedasticity in the noise process.

3.4.1 DPM-MA(1) Model

The existence of microstructure noise turns the intraday return process into an auto-correlated process. First consider the case in which the error is white noise:

$$\tilde{p}_{t,i} = p_{t,i} + \epsilon_{t,i},\ \epsilon_{t,i} \sim \mathrm{N}(0, \omega_{t,i}^2), \tag{3.31}$$

where $\tilde{p}_{t,i}$ denotes the observed log-price with error, $p_{t,i}$ is the unobserved fundamental log-price and $\omega_{t,i}^2$ is the heteroskedastic noise variance.

Given this structure it can be shown that the returns series $\tilde{r}_{t,i} = \tilde{p}_{t+1,i} - \tilde{p}_{t,i}$ has non-zero first order autocorrelation but zero higher order autocorrelation. That is $\mathrm{cov}(\tilde{r}_{t,i+1}, \tilde{r}_{t,i}) = -\omega_{t,i}^2$ and $\mathrm{cov}(\tilde{r}_{t,i+j}, \tilde{r}_{t,i}) = 0$ for $j \geq 2$. This suggest a moving average model of order one.

Combining MA(1) parameterization with our Bayesian nonparametric framework yields the DPM-MA(1) models.

$$\begin{aligned}
\tilde{r}_{t,i}|\mu_t, \theta_t, \delta_{t,i}^2 &= \mu_t + \theta_t \eta_{t,i-1} + \eta_{t,i}, \quad \eta_{t,i} \sim \mathrm{N}(0, \delta_{t,i}^2) && (3.32)\\
\delta_{t,i}^2|G_t &\sim G_t, && (3.33)\\
G_t|G_{0,t}, \alpha_t &\sim \mathrm{DP}(\alpha_t, G_{0,t}), && (3.34)\\
G_{0,t} &\equiv \mathrm{IG}(v_{0,t}, s_{0,t}). && (3.35)
\end{aligned}$$

The noise terms are heteroskedastic. Note that the mean of $r_{t,i}$ is not a constant term

but a moving average term. The MA parameter θ_t is constant for i but will change with the day t. The prior is $\theta_t \sim \mathrm{N}(m_\theta, v_\theta^2)\mathbb{1}_{\{|\theta_t|<1\}}$ in order to make the MA model invertible. The error term $\eta_{t,0}$ is assumed to be zero. Other model settings remain the same as the DPM illustrated in Section 3.3. Later we show how estimates from this specification can be be used to recover an estimate of the ex-post variance V_t of the true return process.

3.4.2 DPM-MA(q) Model

For lower sampling frequencies, such as 1 minute or more, first order autocorrelation is the main effect from market microstructure. As such, the MA(1) model will be sufficient for many applications. However, at higher sampling frequencies, the dependence may be stronger. To allow for a more complex effect on returns from the noise process consider the MA(q-1) noise affecting returns,

$$\tilde{p}_{t,i} = p_{t,i} + \epsilon_{t,i} - \rho_1\epsilon_{t,i-1} - \cdots - \rho_{q-1}\epsilon_{t,i-q+1},\ \epsilon_{t,i} \sim \mathrm{N}(0, \omega_{t,i}^2). \tag{3.36}$$

For returns, this leads to the following DPM-MA(q) model,

$$\begin{aligned}
\tilde{r}_{t,i}|\mu_t, \{\theta_{t,j}\}_{j=1}^q, \delta_{t,i}^2 &= \mu_t + \sum_{j=1}^{q} \theta_{t,j}\eta_{t,i-j} + \eta_{t,i}, \quad \eta_{t,i} \sim \mathrm{N}(0, \delta_{t,i}^2) && (3.37)\\
\delta_{t,i}^2|G_t &\sim G_t, && (3.38)\\
G_t|G_{0,t}, \alpha_t &\sim \mathrm{DP}(\alpha_t, G_{0,t}), && (3.39)\\
G_{0,t} &\equiv \mathrm{IG}(v_{0,t}, s_{0,t}). && (3.40)
\end{aligned}$$

The joint prior of $(\theta_{t,1}, \ldots, \theta_{t,q})$ is $\mathrm{N}(M_\Theta, V_\Theta)\mathbb{1}_{\{\Theta\}}$[6] and $(\eta_{t,0}, \ldots, \eta_{t,-(q-1)}) = (0, \ldots, 0)$.

3.4.3 Model Estimation

We discuss the estimation of DPM-MA(1) model and the approach can be easily extended to the DPM-MA(q). The main difference in this model is that the conditional mean parameters μ_t and θ_t require a Metropolis-Hasting (MH) step to sample their conditional posteriors. The remaining MCMC steps are essentially the same. As before, let $\psi^2_{t,i}$ denote the unique values of $\delta^2_{t,j}$ then each MCMC iteration samples from the following conditional distributions.

1. $\pi\left(\mu_t|\tilde{r}_{t,1:n_t}, \{\psi^2_{t,j}\}_{j=1}^K, \theta_t, s_{t,1:n_t}\right) \propto p\left(\mu_t\right) \prod_{i=1}^{n_t} \mathrm{N}\left(\tilde{r}_{t,i}|\mu_t + \theta_t\eta_{t,i-1}, \psi^2_{t,s_{t,i}}\right)$.

2. $\pi\left(\theta_t|\tilde{r}_{t,1:n_t}, \mu_t, \{\psi^2_{t,j}\}_{j=1}^K, s^t_{1:n_t}\right) \propto p\left(\theta_t\right) \prod_{i=1}^{n_t} p\left(\tilde{r}_{t,i}|\mu_t + \theta_t\eta_{t,i-1}, \psi^2_{t,s_{t,i}}\right)$.

3. $\pi\left(\psi^2_{t,j}|\tilde{r}_{t,1:n_t}, \mu_t, \theta_t, s_{t,1:n_t}\right) \propto p\left(\psi^2_{t,j}\right) \prod_{t:s_t=j} p\left(\tilde{r}_{t,i}|\mu_t + \theta_t\varepsilon_{t,i-1}, \psi^2_{t,j}\right)$ for $j = 1, \ldots, K$.

4. $\pi\left(v_{t,j}|s_{t,1:n_t}\right) \propto \mathrm{Beta}\left(v_{t,j}|a_{t,j}, b_{t,j}\right)$ with $a_{t,j} = 1 + \sum_{i=1}^{n_t} \mathbb{1}(s_{t,i} = j)$ and $b_{t,j} = \alpha_t + \sum_{i=1}^{n_t} \mathbb{1}(s_{t,i} > j)$ and update $w_{t,j} = v_{t,j} \prod_{l<j}(1 - v_{t,l})$ for $j = 1, \ldots, K$.

5. $\pi\left(u_{t,i}|w_{t,i}, s_{t,1:n_t}\right) \propto \mathbb{1}(0 < u_{t,i} < w_{t,s_{t,i}})$ for $i = 1, \ldots, n_t$.

6. Find the smallest K such that $\sum_{j=1}^K w_{t,j} > 1 - \min(u_{t,1:n_t})$.

7. $\pi\left(s_{t,i}|\tilde{r}_{1:n_t}, s_{t,1:n_t}, \mu_t, \theta_t, \{\psi^2_{t,j}\}_{j=1}^K, u_{t,1:n_t}, K\right) \propto \sum_{j=1}^K \mathbb{1}(u_{t,i} < w_{t,j})\mathrm{N}(\tilde{r}_{t,i}|\mu_t + \theta_t\eta_{t,i-1}, \psi^2_{t,j})$ for $i = 1, \ldots, n_t$.

8. $\pi(\alpha_t|K) \propto p(\alpha_t)p(K|\alpha_t)$.

[6]Restrictions on MA coefficients: all the roots of $1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q = 0$ are outside of the unit circle.

In steps 1 and 2 the likelihood requires the sequential calculation of the lagged error as $\eta_{t,i-1} = \tilde{r}_{t,i-1} - \mu_t - \theta_t \eta_{t,i-2}$ which precludes a Gibbs sampling step. Therefore, μ_t and θ_t are sampled using a MH with a random walk proposal. The proposal is calibrated to achieve an acceptance rate between 0.3 and 0.5.

3.4.4 Ex-post Variance Estimator under Microstructure Error

Hansen *et al.* (2008) showed that prefiltering with an MA model results in a bias in the RV estimator.[7] In the Appendix it is shown that the Hansen *et al.* (2008) bias correction provides an accurate adjustment to our Bayesian estimator in the context of heteroskedastic noise. From the DPM-MA(1) model the posterior mean of V_t under independent microstructure error is

$$\hat{V}_{t,\mathrm{MA}(1)} = \frac{1}{M} \sum_{m=1}^{M} (1 + \theta_t^{(m)})^2 \sum_{i=1}^{n_t} \delta_{t,i}^{2(m)}, \tag{3.41}$$

where $\delta_{t,i}^{2(m)} = \psi_{t,s_{t,i}^{(m)}}^{2(m)}$ The log of V_t, square-root of V_t and density intervals can be estimated as the Bayesian nonparametric ex-post variance estimator without microstructure error.

In the case of higher autocorrelation the DPM-MA(q) model adjusted posterior estimate of V_t is

$$\hat{V}_{t,\mathrm{MA(q)}} = \frac{1}{M} \sum_{m=1}^{M} \left(1 + \sum_{j=1}^{q} \theta_{t,j}^{(m)} \right)^2 \sum_{i=1}^{n_t} \delta_{t,i}^{2(m)}. \tag{3.42}$$

[7] If $\tilde{r}_t = \theta_1 \eta_{t-1} + \cdots + \theta_q \eta_{t-q+1} + \eta_t$, then under their assumptions the bias corrected estimate of ex-post variance is $RV_{\mathrm{MAq}} = (1 + \theta_1 + \cdots + \theta_q)^2 \sum_{i=1}^{n_t} \hat{\eta}_i^2$, where $\hat{\eta}_i$ denotes a fitted residual.

Next we consider simulation evidence on these estimators.

3.5 Simulation Results

3.5.1 Data Generating Process

We consider four commonly used data generating processes (DGPs) in the literature. The first one is the GARCH(1,1) diffusion, introduced by Andersen and Bollerslev (1998). The log-price follows

$$\mathrm{d}p(t) = \mu \mathrm{d}t + \sigma(t)\mathrm{d}W_p(t), \tag{3.43}$$

$$\mathrm{d}\sigma^2(t) = \alpha(\beta - \sigma^2(t))\mathrm{d}t + \gamma\sigma^2(t)\mathrm{d}W_\sigma(t). \tag{3.44}$$

where $W_p(t)$ and $W_\sigma(t)$ are two independent Wiener processes. The values of parameters follow Andersen and Bollerslev (1998) and are $\mu = 0.03$, $\alpha = 0.035$, $\beta = 0.636$ and $\gamma = 0.144$, which were estimated using foreign exchange data.

Following Huang and Tauchen (2005), the second and third DGP are a one factor stochastic volatility diffusion (SV1F) and one factor stochastic volatility diffusion with jumps (SV1FJ). SV1F is given by

$$\mathrm{d}p(t) = \mu \mathrm{d}t + \exp\left(\beta_0 + \beta_1 v(t)\right)\mathrm{d}W_p(t), \tag{3.45}$$

$$\mathrm{d}v(t) = \alpha v(t)\mathrm{d}t + \mathrm{d}W_v(t) \tag{3.46}$$

and the price process for SV1FJ is

$$\mathrm{d}p(t) = \mu \mathrm{d}t + \exp\left(\beta_0 + \beta_1 v(t)\right)\mathrm{d}W_p(t) + dJ(t)\ , \tag{3.47}$$

where $\text{corr}(\text{d}W_p(t), \text{d}W_v(t)) = \rho$, and $J(t)$ is a Poisson process with jump intensity λ and jump size $\delta \sim N(0, \sigma_J^2)$. We adopt the parameter settings from Huang and Tauchen (2005) and set $\mu = 0.03$, $\beta_0 = 0.0$, $\beta_1 = 0.125$, $\alpha = -0.1$, $\rho = -0.62$, $\lambda = 0.014$ and $\sigma_J^2 = 0.5$.

The final DGP is the two factor stochastic volatility diffusion (SV2F) from Chernov *et al.* (2003) and Huang and Tauchen (2005).[8]

$$\text{d}p(t) = \mu \text{d}t + s\text{-}\exp\left(\beta_0 + \beta_1 v_1(t) + \beta_2 v_2(t)\right) \text{d}W_p(t), \tag{3.48}$$

$$\text{d}v_1(t) = \alpha_1 v_1(t) \text{d}t + \text{d}W_{v_1}(t), \tag{3.49}$$

$$\text{d}v_2(t) = \alpha_2 v_2(t) \text{d}t + (1 + \psi v_2(t)) \, \text{d}W_{v_2}(t), \tag{3.50}$$

where $\text{corr}(\text{d}W_p(t), \text{d}W_{v_1}(t)) = \rho_1$ and $\text{corr}(\text{d}W_p(t), \text{d}W_{v_2}(t)) = \rho_2$. The parameter values in SV2F are $\mu = 0.03$, $\beta_0 = -1.2$, $\beta_1 = 0.04$, $\beta_2 = 1.5$, $\alpha_1 = -0.00137$, $\alpha_2 = -1.386$, $\psi = 0.25$ and $\rho_1 = \rho_2 = -0.3$, which are from Huang and Tauchen (2005).

Data is simulated using a basic Euler discretization at 1-second frequency for the four DGPs. Assuming the length of daily trading time is 6.5 hours (23400 seconds), we first simulate the log price level every second. After this we compute the 5-minute, 1-minute, 30-second and 10-second intraday returns by taking the difference every 300, 60, 30, 10 steps, respectively. The initial volatility level, such as v_{1t} and v_{2t} in SV2F, at day t is set equal to the last volatility value at previous day, $t-1$. $T = 5000$ days of intraday returns are simulated using the four DGPs and used to report sampling properties of the volatility estimators. In each case, to remove dependence on the

[8]The function $s\text{-}\exp$ is defined as $s\text{-}\exp(x) = \exp(x)$ if $x \leq x_0$ and $s\text{-}\exp(x) = \frac{\exp(x_0)}{\sqrt{x_0}} \sqrt{x_0 - x_0^2 + x^2}$ if $x > x_0$, with $x_0 = \log(1.5)$.

startup conditions 500 initial days are dropped from the simulation.

Independent Noise

Following Barndorff-Nielsen *et al.* (2008), log-prices with independent noise are simulated as follows

$$\begin{aligned} \tilde{p}_{t,i} &= p_{t,i} + \epsilon_{t,i}, \\ \epsilon_{t,i} &\sim \mathrm{N}(0, \sigma_\omega^2), \\ \sigma_\omega^2 &= \xi^2 \mathrm{var}(r_t). \end{aligned} \tag{3.51}$$

The error term is added to the log-prices simulated from the 4 DGPs every second. The variance of microstructure error is proportional to the daily variance calculated using the pure daily returns. We set the noise-to-signal ratio $\xi^2 = 0.001$, which is the same value used in Barndorff-Nielsen *et al.* (2008) and close to the value in Bandi and Russell (2008).

Dependent Noise

Following Hansen *et al.* (2008), we consider the simulation of log-prices with dependent noise as follows,

$$\begin{aligned} \tilde{p}_{t,i} &= p_{t,i} + \epsilon_{t,i}, \\ \epsilon_{t,i} &\sim \mathrm{N}\left(\mu_{\epsilon_{t,i}}, \sigma_\omega^2\right), \\ \mu_{\epsilon_{t,i}} &= \sum_{l=1}^{\phi} \left(1 - l/\phi\right) \left(p_{t,i-l} - p_{t,i-1-l}\right), \\ \sigma_\omega^2 &= \xi^2 \mathrm{var}(r_t), \end{aligned} \tag{3.52}$$

where $\phi = 20$, which makes the error term correlated with returns in the past 20 seconds (steps). If past returns are positive (negative) the noise term tends to be positive (negative). All other settings, such as σ_ω^2 and ξ^2, are the same as in the independent error case.

3.5.2 True Volatility and Comparison Criteria

We assess the ability of several ex-post variance estimators to estimate the daily quadratic variation (QV_t) from the four data generating processes. QV_t is estimated as the summation of the squared intraday pure returns at the highest frequency (1 second)

$$\sigma_t^2 \equiv \sum_{i=1}^{23400} r_{t,i}^2. \tag{3.53}$$

The competing ex-post daily variance estimators, generically labeled $\hat{\sigma}_t^2$, are compared based on the root mean squared errors (RMSE), and bias defined as

$$\text{RMSE}(\widehat{\sigma_t^2}) = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(\widehat{\sigma_t^2} - \sigma_t^2\right)^2}, \tag{3.54}$$

$$\text{Bias}(\widehat{\sigma_t^2}) = \frac{1}{T}\sum_{t=1}^{T}\left(\widehat{\sigma_t^2} - \sigma_t^2\right). \tag{3.55}$$

The coverage probability estimates report the frequency that the confidence intervals or density intervals from the Bayesian nonparametric estimators contain the true ex-post variance, σ_t^2. The 95% confidence intervals of RV_t, RK_t^F and RK_t^N reply on the asymptotic distribution, which are provided in equation (3.4), (3.8) and (3.12). We take the bias into account to compute the 95% confidence interval using RK_t^N.

The estimation of integrated quarticity is crucial in determining the confidence

interval for the realized kernels. We consider two versions of quarticity, one is to use the true (infeasible) IQ_t which is calculated as

$$IQ_t^{\text{true}} = 23400 \sum_{i=1}^{23400} \sigma_{t,i}^4, \tag{3.56}$$

where $\sigma_{t,i}^2$ refers to spot variance simulated at the highest frequency. The other method is to estimate IQ_t using the tri-power quarticity estimator, see formula (3.10). The confidence interval based on IQ_t^{true} is the infeasible case and the confidence interval calculated using TPQ_t is the feasible case.

Table 3.1 list the priors specification of models. For each day 5000 MCMC draws are collected after 1000 burn-in to compute the Bayesian posterior quantities. A 0.95 density interval is the 0.025 and 0.975 sample quantiles of MCMC draws of $\sum_{i=1}^{n_t} \sigma_{t,i}^2$, respectively.

3.5.3 No Microstructure Noise

Figure 3.1 plots 500 days of σ_t^2 and estimates RV_t and $\hat{V}_t$ based on returns simulated from the GARCH(1,1) DGP at 5-minute, 1-minute, 30-second and 10-second. Both estimators become more accurate as the data frequency increases.

In Table 3.2, $\hat{V}_t$ has slightly smaller RMSE in 12 out of the 16 categories. For example, for the 5-minute data $\hat{V}_t$ reduces the RMSE by over 5% for the SV2F data. This is remarkable given that RV_t is the gold standard in the no noise setting. Figure 3.2 plots the difference between RMSE of RV_t and $\hat{V}_t$ in 100 subsamples for GARCH(1,1) and SV1F returns at different frequencies. $\hat{V}_t$ is superior to RV_t in most of the subsamples, especially for low frequency returns.

Table 3.3 shows the bias to be small for both estimators. The Bayesian estimator reduces the bias for data simulated from GARCH and SV1F diffusion, while RV_t has smaller bias in the other cases.

Table 3.4 shows that coverage probabilities for 95% confidence intervals of RV_t and 0.95 density intervals of $\hat{V}_t$. The Bayesian nonparametric estimator produces fairly good coverage probabilities for both low and high frequency data, except for the SV2F data. For RV_t, data frequencies higher than 5-minutes are needed to obtain good finite sample coverage when the asymptotic distribution is used.

In summary, under no microstructure noise, the Bayesian nonparametric estimator is very competitive with the classical counterpart RV_t. $\hat{V}_t$ offers smaller estimation error and better finite sample results than RV_t when the data frequency is low. Performance of RV_t and $\hat{V}_t$ both improve as the sampling frequency increases.

3.5.4 Independent Microstructure Noise

In this section we compare RV_t, RK_t^F, $\hat{V}_t$ and $\hat{V}_{t,\mathrm{MA}(1)}$. Figure 3.3 displays the time-series of RK_t^F, $\hat{V}_{t,\mathrm{MA}(1)}$ along with the true variance for several sampling frequencies for data from the SV1F DGP. Both estimators become more accurate as the sampling frequency increases.

Table 3.5 shows the RMSE of the various estimators for different sampling frequencies and DGPs. RV_t and $\hat{V}_t$ produce smaller errors in estimating σ_t^2 than RK_t^F and $\hat{V}_{t,\mathrm{MA}(1)}$ for 5-minute data. However, increasing the sampling frequency results in a larger bias from the microstructure noise. As such, RK_t^F and $\hat{V}_{t,\mathrm{MA}(1)}$ are more accurate as the data frequency increases. Compared to RK_t^F, $\hat{V}_{t,\mathrm{MA}(1)}$ has a smaller RMSE in all cases, except for 30-second and 10-second SV2F return. Figure 3.4 shows

that $\hat{V}_{t,\mathrm{MA}(1)}$ outperforms RK_t^F in most of the subsamples.

The bias of the estimators is found in Table 3.6. Again, RV_t and $\hat{V}_t$ overestimate the ex-post variance by a significant amount unless the data frequency is low. Both RK_t^F and $\hat{V}_{t,\mathrm{MA}(1)}$ produce better results as more data is used. The bias of RK_t^F is smaller than that of $\hat{V}_{t,\mathrm{MA}(1)}$, but the differences are minor.

As can be seen in Table 3.7, $\hat{V}_{t,\mathrm{MA}(1)}$ has the best finite sample coverage among all the alternatives except for the SV2F data. For example, the coverage probabilities of 0.95 density intervals are always within 0.5% from the truth. Note that the density intervals are trivial to obtain from the MCMC output and do not require the calculation IQ_t. The coverage probabilities of either infeasible and feasible confidence intervals of realized kernels are not as good as those of $\hat{V}_{t,\mathrm{MA}(1)}$. Moreover, RK_t^F requires larger samples for good coverage, while density intervals of $\hat{V}_{t,\mathrm{MA}(1)}$ perform well for either low or high frequency returns.

3.5.5 Dependent Microstructure Noise

The last experiment considers the performances of the estimators under dependent noise. RK_t^N, RV_t, $\hat{V}_t$, $\hat{V}_{t,\mathrm{MA}(1)}$ and $\hat{V}_{t,\mathrm{MA}(2)}$ are compared. Figure 3.5 plots the estimators for different sampling frequencies. It is clear that estimation is less precise in this setting.

The RMSE of estimators can be found in Table 3.8. Again, RV_t and $\hat{V}_t$ provide poor results if high frequency data is used. Except for one entry in the table, a version of the Bayesian estimator has the smallest RMSE in each case. The $\hat{V}_{t,\mathrm{MA}(1)}$ estimator is ranked the best if return frequency is 30 seconds, followed by $\hat{V}_{t,\mathrm{MA}(2)}$ and RK_t^N. For 10 seconds returns, $\hat{V}_{\mathrm{MA}(2)}$ provides the smallest error. Compared to

RK_t^N the $\hat{V}_{t,\text{MA}(1)}$ and $\hat{V}_{t,\text{MA}(2)}$ can provide significant improvements for 30 and 10 second returns. For instance, at 30 seconds, reductions in the RMSE of 10% or more are common while at the 10 second frequency reductions in the RMSE are 25% or more. The subsample analysis shown in Figure 3.6 supports these findings.

Table 3.9 shows $\hat{V}_{t,\text{MA}(1)}$ and $\hat{V}_{t,\text{MA}(2)}$ have smaller bias if return frequency is one minute or higher.

Table 3.10 shows the coverage probabilities of all the five estimators. The finite sample results of $\hat{V}_{t,\text{MA}(2)}$ are all very close to the optimal level, no matter the data frequency.

3.5.6 Evidence of Pooling

Figure 3.7-3.9 display the histograms of the posterior mean of the number of clusters in three different settings. There are: the DPM for 5-minute SV1F returns (no noise), the DPM-MA(1) for 1-minute SV1FJ returns (independent noise) and the DPM-MA(2) for 30-second SV2F returns (dependent noise). The figures show significant pooling. For example, in the 1-minute SV1FJ return case, most of the daily variance estimates of V_t are formed by using 1 to 5 pooled groups of data, instead of 390 observations (separate groups) which is what the realized kernel uses. This level of pooling can lead to significant improvements for the Bayesian estimator.

In summary, these simulations show the Bayesian estimate of ex-post variance to be very competitive with existing classical alternatives.

3.6 Empirical Applications

For each day, 5000 MCMC draws are taken after 10000 burn-in draws are discarded, to estimate posterior moments. All prior setting are the same as in the simulations.

3.6.1 Application to IBM Return

We first consider estimating and forecasting volatility using a long calendar span of IBM equity returns. The 1-minute IBM price records from 1998/01/03 to 2016/02/16 were downloaded from the Kibot website[9]. We choose the sample starting from 2001/01/03 as the relatively small number of transactions before year 2000 yields many zero intraday returns. The days with less than 5 hours of trading are removed, which leaves 3764 days in the sample.

Log-prices are placed on a 1-minute grid using the price associated with closest time stamp that is less than or equal to the grid time. The 5-minute and 1-minute percentage log returns from 9:30 to 16:00(EST) are constructed by taking the log price difference between two close prices in time grid and scaling by 100. The overnight returns are ignored so the first intraday return is formed using the daily opening price instead of the close price in previous day. The procedure generates 293,520 5-minute returns and 1,467,848 1-minute returns.

We use a filter to remove errors and outliers caused by abnormal price records. We would like to filter out the situation in which the price jumps up or down but quickly moves back to original price range. This suggests an error in the record. If $|r_{t,i}| + |r_{t,i+1}| > 8\sqrt{\text{var}_t(r_{t,i})}$ and $|r_{t,i} + r_{t,i+1}| < 0.05\%$, we replace $r_{t,i}$ and $r_{t,i+1}$ by $r'_{t,i} = r'_{t,i+1} = 0.5 \times (r_{t,i} + r_{t,i+1})$. The filter adjusts 0 and 70 (70/1,467,848 =

[9]http://www.kibot.com

0.00477%) returns for 5-minute and 1-minute case, respectively.

From these data several version of daily $\hat{V}_t$, RV_t and RK_t are computed. Daily returns are the open-to-close return and match the time interval for the variance estimates. For each of the estimators we follow exactly the methods used in the simulation section.

Ex-post Variance Estimation

Table 3.11 reports summary statistics for several estimators. Overall the Bayesian and classical estimators are very close. Both the realized kernel and the moving average DPM estimators reduce the average level of daily variance and indicate the presence of significant market microstructure noise. Based on this and an analysis of the ACF of the high-frequency returns we suggest the $\hat{V}_{t,\mathrm{MA}(1)}$ for the 5-minute data and the $\hat{V}_{t,\mathrm{MA}(4)}$ for the 1-minute data in the remainder of the analysis. Comparison with the kernel estimators is found in Figures 3.10 and 3.11. Except for the extreme values they are very similar.

Interval estimates for two sub-periods are shown in Figures 3.12 and 3.13. A clear disadvantage of the kernel based confidence interval in that it includes negative values for ex-post variance. The Bayesian version by construction does not and tends to be significantly shorter in volatile days. The results of log variance[10] are also provided with some differences remaining.

The degree of pooling from the Bayesian estimators is found in Figure 3.14 and 3.15. As expected, we see more groups in the higher 1-minute frequency. In this case, on average, there are about 3 to 7 distinct groups of intraday variance parameters.

[10] 95% confidence intervals using $\log(RV_t)$, $\log(RK_t^F)$ and $\log(RK_t^N)$ are based on the asymptotic distributions in Barndorff-Nielsen and Shephard (2002), Barndorff-Nielsen *et al.* (2008) and Barndorff-Nielsen *et al.* (2011).

Ex-post Variance Modeling and Forecasting

Does the Bayesian estimator correctly recover the time-series dynamics of volatility? To investigate this we estimate several versions of the Heterogeneous Auto-Regressive (HAR) model introduced by Corsi (2009). This is a popular model that captures the strong dependence in ex-post daily variance. For $\hat{V}_t$ the HAR model is

$$\hat{V}_t = \beta_0 + \beta_1 \hat{V}_{t-1} + \beta_2 \hat{V}_{t-1|t-5} + \beta_3 \hat{V}_{t-1|t-22} + \epsilon_t, \tag{3.57}$$

where $\hat{V}_{t-1|t-h} = \frac{1}{h}\sum_{l=1}^{h} \hat{V}_{t-l}$ and ϵ_t is the error term. $\hat{V}_{t-1}$, $\hat{V}_{t-1|t-5}$ and $\hat{V}_{t-1|t-22}$ correspond to the daily, weekly and monthly variance measures up to time $t-1$. Similar specifications are obtained by replacing $\hat{V}_t$ with RV_t or RK_t.

Bollerslev *et al.* (2016) extend the HAR model to the HARQ model by taking the asymptotic theory of RV_t into account. The HARQ model for RV_t is given by

$$RV_t = \beta_0 + \left(\beta_1 + \beta_{1Q} RQ_{t-1}^{1/2}\right) RV_{t-1} + \beta_2 RV_{t-1|t-5} + \beta_3 RV_{t-1|t-22} + \epsilon_t \tag{3.58}$$

The loading on RV_{t-1} is no longer a constant, but varying with measurement error, which is captured by RQ_{t-1}. The model responds more to RV_{t-1} if measurement error is low and has a lower response if error is high. Bollerslev *et al.* (2016) provide evidence that the HARQ model outperforms the HAR model in forecasting.[11]

An advantage of our Bayesian approach is that we have the full finite sample posterior distribution for V_t. In the Bayesian nonparametric framework, there is no need to estimate IQ_t with RQ_t, instead the variance, standard deviation or other

[11] A drawback of this specification is that it is possible for the coefficient on RV_{t-1} to be negative and produce a negative forecast for next period's variance. To avoid this when $\beta_1 + \beta_{1Q} RQ_{t-1}^{1/2} < 0$ it is set to 0.

features of V_t can be easily estimated using the MCMC output. Replacing RQ_{t-1} with $\widehat{\text{var}}(V_{t-1})$, the modified HARQ model for $\hat{V}_t$ is defined as

$$\hat{V}_t = \beta_0 + \left(\beta_1 + \beta_{1Q}\widehat{\text{var}}(V_{t-1})^{1/2}\right)\hat{V}_{t-1} + \beta_2\hat{V}_{t-1|t-5} + \beta_3\hat{V}_{t-1|t-22} + \epsilon_t, \tag{3.59}$$

where $\widehat{\text{var}}(V_{t-1})^{1/2}$ is an MCMC estimate of the posterior standard deviation of V_t.

Table 3.12 displays the OLS estimates and the R^2 for several model specifications. Coefficient estimates are comparable across each class of model. Clearly the Bayesian variance estimates display the same type of time-series dynamics found in the realized kernel estimates.

Finally, out-of-sample root-mean squared forecast errors (RMSFE) of HAR and HARQ models using both classical estimators and Bayesian estimators are found in Table 3.13. The out-of-sample period is from 2005/01/03 to 2016/02/16 (2773 observations) and model parameters are re-estimated as new data arrives. Note, that to mimic a real-time forecast setting the prior hyperparameters $\nu_{0,t}$ and $s_{0,t}$ are set based on intraday data from day t and $t-1$.[12]

The first column of Table 3.13 reports the data frequency and the dependent variable used in the HAR/HARQ model. The second column records the data used to construct the right-hand side regressors. In this manner we consider all the possible combinations of how RK_t^N is forecast by lags of RK_t^N or $\hat{V}_{t,\text{MA}}$ and similarly for forecasting $\hat{V}_{t,\text{MA}}$. All of the specifications produce similar RMSFE. In 7 out of 8 cases the Bayesian variance measure forecasts itself and the realized kernel better.

[12]Data from day $t+1$ would not be available in a real-time scenario. Using only data from day t to set $\nu_{0,t}$ and $s_{0,t}$ gives very similar results.

3.6.2 Applications to Disney Returns

The second application considers ex-post variance estimation of Disney returns. Transaction and quote data for Disney was supplied by Tickdata. The quote data is NBBO (National Best bid/ask Offer). We follow the same method of Barndorff-Nielsen *et al.* (2011) to clean both transaction and quote datasets and form grid returns at 5-minute, 1-minute, 30-second and 10-second frequencies using transaction prices. The sample period is from January 2, 2015 to December 30, 2015 and does not include days with less than 6 trading hours. The final dataset has 247 daily observations.

We found weaker evidence of serial correlations in Disney returns and therefore focus on lower order moving average specifications. Our recommendation would be to use $\hat{V}_t$ for 5-minute and 1 minute data and $\hat{V}_{t,\text{MA(1)}}$ for 30-second data.

Table 3.14 displays the summary statistics of daily variance estimators of Disney returns. The sample average of the different variance estimators is quite different than the sample variance of daily returns. This is more of a small sample issue than anything else. We do see that both the kernel and the Bayesian models with MA terms generally reduce the average variance level compared to the unadjusted versions (RV_t and $\hat{V}_t$).

Figure 3.16 and 3.17 display box plots of the daily variance estimates for the classical and Bayesian estimators for the 5-minute and 30-second data. There are several important points to make. First, both estimators recover the same general pattern of volatility in this period. Second, the Bayesian density interval is often shorter and asymmetric compare to the classical counterpart. Although there is general agreement, the high variance days of December 15,21 and 22 indicate some differences particularly in Figure 3.17. Finally, both estimates become more accurate with the

higher frequency 30-second data and also make a significant downward revision to the variance estimates on December 21 and 22.

3.7 Conclusion

This chapter offers a new exact finite sample approach to estimate ex-post variance using Bayesian nonparametric methods. The proposed approach benefits ex-post variance estimation in two aspects. First, the observations with similar variance levels can be pooled together to increase accuracy. Second, exact finite sample inference is available directly without replying on additional assumptions about a higher frequency DGP. Bayesian nonparametric variance estimators under no noise, heteroskedastic and serially correlated microstructure noise cases are introduced. Monte Carlo simulation results show that the proposed approach can increase the accuracy of ex-post variance estimation and provide reliable finite sample inference. Applications to real equity returns show the new estimators conform closely to the realized variance and kernel estimators in terms of average statistical properties as well as time-series characteristics. The Bayesian estimators can be used with confidence and have several benefits relative to existing methods. The Bayesian estimator can capture asymmetric density intervals, always remains positive and does not rely on the estimation of integrated quarticity.

3.8 Appendix

3.8.1 Adjustment to DPM-MA(1) Estimator

Let $p_{t,i}$ denotes the latent intraday price and $\epsilon_{t,i}$ is the microstructure noise which is independently distributed and heteroskedastic. The observed intraday price $\widetilde{p}_{t,i}$ is

$$\widetilde{p}_{t,i} = p_{t,i} + \epsilon_{t,i}, \qquad E(\epsilon_{t,i}) = 0 \text{ and } \text{var}(\epsilon_{t,i}) = \omega^2_{t,i}. \tag{3.60}$$

The log return process is constructed as follows,

$$\widetilde{r}_{t,i} = \widetilde{p}_{t,i} - \widetilde{p}_{t,i-1} = p_{t,i} - p_{t,i-1} + \epsilon_{t,i} - \epsilon_{t,i-1} = r_{t,i} + \epsilon_{t,i} - \epsilon_{t,i-1}, \tag{3.61}$$

where $\widetilde{r}_{t,i}$ and $r_{t,i}$ are the observed return and pure return. The variance and first autocovariance of $\{r_{t,i}\}_{i=1}^{n_t}$ are

$$\text{var}(\widetilde{r}_{t,i}) = \sigma^2_{t,i} + \omega^2_{t,i} + \omega^2_{t,i-1}, \tag{3.62}$$

$$\text{cov}(\widetilde{r}_{t,i}, \widetilde{r}_{t,i-1}) = -\omega^2_{t,i-1}. \tag{3.63}$$

Consider the following heteroskedastic MA(1) model for the observed $\tilde{r}_{t,i}$,

$$\widetilde{r}_{t,i} = \mu_t + \theta_t \eta_{t,i-1} + \eta_{t,i}, \qquad \eta_{t,i} \sim N(0, \delta^2_{t,i}), \tag{3.64}$$

which will be used to recover an estimate of ex-post variance for the pure return

process, $V_t = \sum_{i=1}^{n_t} \sigma_{t,i}^2$. The corresponding moments of this process are

$$\text{var}(\widetilde{r}_{t,i}) = \theta_t^2 \delta_{t,i-1}^2 + \delta_{t,i}^2, \tag{3.65}$$

$$\text{cov}(\widetilde{r}_{t,i}, \widetilde{r}_{t,i-1}) = \theta_t \delta_{t,i-1}^2. \tag{3.66}$$

Equating (3.62) and (3.65), we have

$$\sigma_{t,i}^2 + \omega_{t,i}^2 + \omega_{t,i-1}^2 = \theta_t^2 \delta_{t,i-1}^2 + \delta_{t,i}^2 \tag{3.67}$$

Equating (3.63) and (3.66), we have

$$-\omega_{t,i-1}^2 = \theta_t \delta_{t,i-1}^2 \qquad \text{and} \qquad -\omega_{t,i}^2 = \theta_t \delta_{t,i}^2. \tag{3.68}$$

Based on the result in (3.68), the summation of $\delta_{t,i}^2$, over $i = 1, \ldots, n_t$, equals

$$\sum_{i=1}^{n_t} \delta_{t,i}^2 = -\frac{1}{\theta_t} \sum_{i=1}^{n_t} w_{t,i}^2. \tag{3.69}$$

Plugging both terms in (3.68) into (3.67), yields

$$\sigma_{t,i}^2 + \omega_{t,i}^2 + \omega_{t,i-1}^2 = -\theta_t \omega_{t,i-1}^2 - \frac{\omega_{t,i}^2}{\theta_t} \tag{3.70}$$

$$\sigma_{t,i}^2 + \left(1 + \frac{1}{\theta_t}\right) \omega_{t,i}^2 + (1 + \theta_t) \omega_{t,i-1}^2 = 0. \tag{3.71}$$

Using the results in (3.71), the summation of $\sigma_{t,i}^2$, over $i = 1, \ldots, n_t$, equals

$$\sum_{i=1}^{n_t} \sigma_{t,i}^2 + \left(1 + \frac{1}{\theta_t}\right) \sum_{i=1}^{n_t} \omega_{t,i}^2 + (1+\theta_t) \sum_{i=1}^{n_t} \omega_{t,i-1}^2 = 0 \tag{3.72}$$

$$V_t = -\left(1 + \frac{1}{\theta_t}\right) \sum_{i=1}^{n_t} \omega_{t,i}^2 - (1+\theta_t) \sum_{i=1}^{n_t} \omega_{t,i-1}^2. \tag{3.73}$$

The ratio between (3.69) and (3.73) is

$$\frac{V_t}{\sum_{i=1}^{n_t} \delta_{t,i}^2} = \frac{-\left(1 + \frac{1}{\theta_t}\right) \sum_{i=1}^{n_t} \omega_{t,i}^2 - (1+\theta_t) \sum_{i=1}^{n_t} \omega_{t,i-1}^2}{-\frac{1}{\theta_t} \sum_{i=1}^{n_t} \omega_{t,i}^2} \tag{3.74}$$

$$= \frac{(1+\theta_t) \sum_{i=1}^{n_t} \omega_{t,i}^2 + (\theta_t + \theta_t^2) \sum_{i=1}^{n_t} \omega_{t,i-1}^2}{\sum_{i=1}^{n_t} \omega_{t,i}^2} \tag{3.75}$$

$$= \frac{(1+\theta_t)^2 \sum_{i=1}^{n_t-1} \omega_{t,i}^2 + (1+\theta_t)\, \omega_{t,n_t}^2 + (\theta_t + \theta_t^2)\, \omega_{t,0}^2}{\sum_{i=1}^{n_t-1} \omega_{t,i}^2 + \omega_{t,n_t}^2} \tag{3.76}$$

$$= (1+\theta_t)^2, \quad \text{if} \quad \omega_{t,n_t} = \omega_{t,0}. \tag{3.77}$$

Finally, we have

$$(1+\theta_t)^2 \sum_{i=1}^{n_t} \delta_{t,i}^2 = V_t, \qquad \text{if} \quad \omega_{t,n_t} = \omega_{t,0}. \tag{3.78}$$

3.8.2 Adjustment to DPM-MA(2) Estimator

If the observed intraday price $\widetilde{p}_{t,i}$ is

$$\widetilde{p}_{t,i} = p_{t,i} + \epsilon_{t,i} - \rho \epsilon_{t,i-1}, \qquad E(\epsilon_{t,i}) = 0 \text{ and } \operatorname{var}(\epsilon_{t,i}) = \omega_{t,i}^2. \tag{3.79}$$

Then log return process is constructed as follows.

$$\begin{aligned}\widetilde{r}_{t,i} &= \widetilde{p}_{t,i} - \widetilde{p}_{t,i-1} \\ &= p_{t,i} - p_{t,i-1} + \epsilon_{t,i} - \rho\epsilon_{t,i-1} - \epsilon_{t,i-1} + \rho\epsilon_{t,i-2} \\ &= r_{t,i} + \epsilon_{t,i} - (1+\rho)\epsilon_{t,i-1} + \rho\epsilon_{t,i-2}.\end{aligned} \tag{3.80}$$

Using the following heteroskedastic MA(2) model for $\tilde{r}_{t,i}$,

$$\widetilde{r}_{t,i} = \mu_t + \theta_{1t}\eta_{t,i-1} + \theta_{2t}\eta_{t,i-2} + \eta_{t,i}, \qquad \eta_{t,i} \sim N(0, \delta_{t,i}^2) \tag{3.81}$$

it can be shown the adjustment term is

$$(1+\theta_{1t}+\theta_{2t})^2 \sum_{i=1}^{n_t} \delta_{t,i}^2 = V_t, \qquad \text{if} \quad \omega_{t,n_t-1} = \omega_{t,0} \quad \text{and} \quad \omega_{t,n_t} = \omega_{t,-1}. \tag{3.82}$$

Similar results hold for higher order MA models.

Bibliography

Aït-Sahalia, Y. and Jacod, J. (2014). *High-Frequency Financial Econometrics*. Princeton University Press.

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

Aït-Sahalia, Y., Mykland, P. A., and Zhang, L. (2011). Ultra high frequency volatility estimation with dependent microstructure noise. *Journal of Econometrics*, **160**(1), 160–175.

Andersen, T. and Bollerslev, T. (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review*, **39**(4), 885–905.

Andersen, T. G. and Benzoni, L. (2008). Realized volatility. Working Paper Series WP-08-14, Federal Reserve Bank of Chicago.

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

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Ebens, H. (2001b). 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. (2003). Modeling and forecasting realized volatility. *Econometrica*, **71**(2), 579–625.

Andersen, T. G., Bollerslev, T., and Meddahi, N. (2011). Realized volatility forecasting and market microstructure noise. *Journal of Econometrics*, **160**(1), 220 – 234.

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

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

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

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (2008). Designing realized kernels to measure the ex post variation of equity prices in the presence of noise. *Econometrica*, **76**(6), 1481–1536.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (2009). Realized kernels in practice: trades and quotes. *Econometrics Journal*, **12**(3), C1–C32.

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.

Bollen, B. and Inder, B. (2002). Estimating daily volatility in financial markets utilizing intraday data. *Journal of Empirical Finance*, **9**(5), 551–562.

Bollerslev, T., Patton, A., and Quaedvlieg, R. (2016). Exploiting the errors: A simple approach for improved volatility forecasting. *Journal of Econometrics*, pages 1–18.

Chernov, M., Ronald Gallant, A., Ghysels, E., and Tauchen, G. (2003). Alternative models for stock price dynamics. *Journal of Econometrics*, **116**(1-2), 225–257.

Corradi, V., Distaso, W., and Swanson, N. R. (2009). Predictive density estimators for daily volatility based on the use of realized measures. *Journal of Econometrics*, **150**(2), 119 – 138.

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

Escobar, M. D. and West, M. (1994). Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association*, **90**, 577–588.

Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, **1**(2), 209–230.

Goncalves, S. and Meddahi, N. (2009). Bootstrapping realized volatility. *Econometrica*, **77**(1), 283–306.

Hansen, P., Large, J., and Lunde, A. (2008). Moving average-based estimators of integrated variance. *Econometric Reviews*, **27**(1-3), 79–111.

Hansen, P. R. and Lunde, A. (2006). Realized variance and market microstructure noise. *Journal of Business & Economic Statistics*, **24**, 127–161.

Huang, X. and Tauchen, G. (2005). The relative contribution of jumps to total price variance. *Journal of Financial Econometrics*, **3**(4), 456–499.

Kalli, M., Griffin, J. E., and Walker, S. G. (2011). Slice sampling mixture models. *Statistics and Computing*, **21**(1), 93–105.

Maheu, J. M. and McCurdy, T. H. (2002). Nonlinear features of realized FX volatility. *The Review of Economics and Statistics*, **84**(4), 668–681.

Neal, R. M. (2000). Markov chain sampling methods for dirichlet process mixture models. *Journal of Computational and Graphical Statistics*, **9**(2), 249–265.

Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica*, pages 639–650.

Zhang, L., Mykland, P. A., and Aït-Sahalia, Y. (2005). A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association*, **100**, 1394–1411.

Zhou, B. (1996). High frequency data and volatility in foreign exchange rates. *Journal of Business and Economic Statistics*, **14**(1), 45–52.

Table 3.1: Prior Specifications of Models

Model	μ_t	$\sigma^2_{t,i}$	Θ_t	α_t
DPM	$\mathrm{N}(0, v^2)$	$\mathrm{IG}(v_{0,t}, s_{0,t})$	-	$\mathrm{Gamma}(2, 8)$
DPM-MA(q)	$\mathrm{N}(0, v^2)$	$\mathrm{IG}(v_{0,t}, s_{0,t})$	$\mathrm{N}(\mathbf{0}, I)\mathbb{1}_{\{\lvert\Theta_t\rvert\}}$	$\mathrm{Gamma}(2, 8)$

[1.] $v_{0,t}$ and $s_{0,t}$ are calculated using equation (3.24).
[2.] $\mathbb{1}_{\{\lvert\Theta_t\rvert\}}$ denotes the invertibility condition for the MA(q) model.
[3.] v^2 is adjusted according to data frequency: $v = 0.001, 0.0002, 0.0001, 0.00002$ for 5-minute, 1-minute, 30-second and 10-second returns.

Table 3.2: RMSEs of RV_t and $\hat{V}_t$ (No Microstructure Noise Case)

Data Freq.	Estimator	GARCH	SV1F	SV1FJ	SV2F
5-minute	RMSE(RV_t)	0.12352	0.21226	0.21471	0.45601
	RMSE($\hat{V}_t$)	**0.12010**	**0.20608**	**0.20981**	**0.43154**
1-minute	RMSE(RV_t)	0.05368	0.09283	**0.09771**	0.23296
	RMSE($\hat{V}_t$)	**0.05330**	**0.09228**	0.10104	**0.22675**
30-second	RMSE(RV_t)	0.03886	0.06530	**0.06741**	0.14178
	RMSE($\hat{V}_t$)	**0.03867**	**0.06503**	0.07321	**0.14021**
10-second	RMSE(RV_t)	0.02177	0.03601	**0.03662**	**0.09535**
	RMSE($\hat{V}_t$)	**0.02175**	**0.03594**	0.04747	0.09645

This table reports the root mean squared error (RMSE) of estimating 5000 daily ex-post variances using RV_t and Bayesian nonparametric estimator $\hat{V}_t$ under different frequencies and DGPs. Microstructure noise is not considered.

Table 3.3: Biases of RV_t and $\hat{V}_t$ (No Microstructure Noise Case)

Data Freq.	Estimator	GARCH	SV1F	SV1FJ	SV2F
5-minute	Bias(RV_t)	-0.00258	-0.00315	**-0.00348**	**-0.00187**
	Bias($\hat{V}_t$)	**-0.00223**	**-0.00256**	-0.00414	-0.01841
1-minute	Bias(RV_t)	-0.00170	-0.00120	**-0.00152**	**0.00097**
	Bias($\hat{V}_t$)	**-0.00125**	**-0.00043**	-0.00229	-0.00294
30-second	Bias(RV_t)	-0.00105	-0.00086	**-0.00105**	0.00159
	Bias($\hat{V}_t$)	**-0.00051**	**0.00010**	-0.00166	**-0.00031**
10-second	Bias(RV_t)	-0.00028	-0.00049	**-0.00001**	**-0.00103**
	Bias($\hat{V}_t$)	**0.00017**	**0.00031**	-0.00105	-0.00161

This table reports bias estimates using 5000 daily ex-post variances using RV_t and Bayesian nonparametric estimator $\hat{V}_t$ under different frequencies and DGPs. Microstructure noise is not considered.

Table 3.4: Coverage Probability (No Microstructure Noise Case)

Data Freq.	Interval Estimator	GARCH	SV1F	SV1FJ	SV2F
5-minute	RV_t	93.00%	92.90%	92.84%	89.66%
	$\hat{V}_t$	94.88%	95.04%	95.10%	87.32%
1-minute	RV_t	94.64%	94.42%	94.28%	93.70%
	$\hat{V}_t$	95.44%	95.14%	95.22%	91.72%
30-second	RV_t	95.20%	95.24%	94.86%	94.72%
	$\hat{V}_t$	95.86%	95.76%	95.46%	92.30%
10-second	RV_t	95.96%	96.14%	95.84%	95.56%
	$\hat{V}_t$	96.42%	96.44%	96.28%	92.80%

This table reports the coverage probabilities of 95% confidence intervals using RV_t and 0.95 density intervals using $\hat{V}_t$ based on 5000 days results for different data generating processes. Microstructure noise is not considered.

Table 3.5: RMSE of RV_t, RK_t^F, $\hat{V}_t$ and $\hat{V}_{t,\mathrm{MA}(1)}$ (Independent Microstructure Error Case)

Data Freq.	Estimator	GARCH	SV1F	SV1FJ	SV2F
5-minute	RMSE(RV_t)	0.16003	0.29182	0.30651	0.47509
	RMSE(RK_t^F)	0.22988	0.42318	0.43993	0.84092
	RMSE($\hat{V}_t$)	**0.15724**	**0.28671**	**0.30132**	**0.46281**
	RMSE($\hat{V}_{t,\mathrm{MA}(1)}$)	0.21529	0.38812	0.40768	0.74354
1-minute	RMSE(RV_t)	0.48607	0.85374	0.94598	0.59132
	RMSE(RK_t^F)	0.11157	0.20184	0.20822	0.46638
	RMSE($\hat{V}_t$)	0.48678	0.85495	0.94626	0.58876
	RMSE($\hat{V}_{t,\mathrm{MA}(1)}$)	**0.10530**	**0.18777**	**0.19456**	**0.41457**
30-second	RMSE(RV_t)	0.95855	1.69544	1.87445	1.10124
	RMSE(RK_t^F)	0.08483	0.15200	0.15743	**0.26357**
	RMSE($\hat{V}_t$)	0.95990	1.69788	1.87556	1.10106
	RMSE($\hat{V}_{t,\mathrm{MA}(1)}$)	**0.07882**	**0.14016**	**0.15151**	0.27695
10-second	RMSE(RV_t)	2.86639	5.06382	5.60527	3.26388
	RMSE(RK_t^F)	0.05575	0.10097	0.10683	**0.16911**
	RMSE($\hat{V}_t$)	2.86891	5.06833	5.60855	3.26612
	RMSE($\hat{V}_{t,\mathrm{MA}(1)}$)	**0.05374**	**0.09600**	**0.10539**	0.20980

This table reports the root mean squared error (RMSE) of estimating 5000 daily ex-post variances using RV_t, RK_t^F and Bayesian nonparametric estimators $\hat{V}_t$ and $\hat{V}_{t,\mathrm{MA}(1)}$ based on returns at different frequencies and simulated from 4 DGPs. The price is contaminated with white noise.

Table 3.6: Biases of RV_t, RK_t^F, $\hat{V}_t$ and $\hat{V}_{t,\text{MA}(1)}$ (Independent Microstructure Error Case)

Data Freq.	Estimator	GARCH	SV1F	SV1FJ	SV2F
5-minute	Bias(RV_t)	0.09390	0.16795	0.18381	0.11193
	Bias(RK_t^F)	**0.00075**	**-0.00355**	**-0.00621**	**-0.00011**
	Bias($\hat{V}_t$)	0.09427	0.16859	0.18348	0.10298
	Bias($\hat{V}_{t,\text{MA}(1)}$)	0.01784	0.03396	0.03391	-0.00022
1-minute	Bias(RV_t)	0.47833	0.83981	0.93104	0.54949
	Bias(RK_t^F)	**0.00277**	**0.00509**	**0.00671**	**0.00162**
	Bias($\hat{V}_t$)	0.47915	0.84124	0.93114	0.54769
	Bias($\hat{V}_{t,\text{MA}(1)}$)	0.00743	0.01270	0.01027	-0.01415
30-second	Bias(RV_t)	0.95446	1.68793	1.86666	1.08855
	Bias(RK_t^F)	**0.00145**	**0.00240**	**0.00490**	**-0.00352**
	Bias($\hat{V}_t$)	0.95990	1.69045	1.86771	1.08861
	Bias($\hat{V}_{t,\text{MA}(1)}$)	0.00542	0.00970	0.00662	-0.01960
10-second	Bias(RV_t)	2.86404	5.05938	5.60035	3.26016
	Bias(RK_t^F)	**0.00040**	**-0.00079**	**0.00146**	**-0.00229**
	Bias($\hat{V}_t$)	2.86891	5.06392	5.60360	3.26243
	Bias($\hat{V}_{t,\text{MA}(1)}$)	0.00367	0.00763	0.00407	-0.02415

This table reports bias estimates from 5000 daily ex-post variances using RV_t, RK_t^F and Bayesian nonparametric estimators $\hat{V}_t$ and $\hat{V}_{t,\text{MA}(1)}$ based on returns at different frequencies and simulated from 4 DGPs. The price is contaminated with white noise.

Table 3.7: Coverage Probabilities (Independent Microstructure Error Case)

Data Freq.	Interval Estimator	GARCH	SV1F	SV1FJ	SV2F
5-minute	RV_t	87.60%	85.00%	84.42%	22.52%
	RK_t^F - Infeasible	87.84%	87.66%	87.94%	93.48%
	RK_t^F - Feasible	84.28%	96.20%	83.68%	97.72%
	$\hat{V}_t$	81.18%	78.06%	76.68%	18.80%
	$\hat{V}_{t,\mathrm{MA}(1)}$	94.24%	94.48%	94.24%	89.84%
1-minute	RV_t	0.46%	0.82%	0.78%	5.64%
	RK_t^F - Infeasible	88.50%	89.78%	89.02%	93.32%
	RK_t^F - Feasible	99.30%	97.76%	95.26%	97.86%
	$\hat{V}_t$	0.42%	0.07%	0.52%	4.92%
	$\hat{V}_{t,\mathrm{MA}(1)}$	94.90%	95.06%	95.00%	86.54%
30-second	RV_t	0.00%	0.00%	0.02%	1.86%
	RK_t^F - Infeasible	89.80%	90.46%	90.74%	92.80%
	RK_t^F - Feasible	77.44%	99.48%	99.52%	97.94%
	$\hat{V}_t$	0.00%	0.00%	0.00%	1.66%
	$\hat{V}_{t,\mathrm{MA}(1)}$	94.92%	95.34%	94.88%	85.76%
10-second	RV_t	0.00%	0.00%	0.00%	0.04%
	RK_t^F - Infeasible	92.08%	92.68%	92.90%	92.10%
	RK_t^F - Feasible	99.98%	99.98%	99.98%	98.62%
	$\hat{V}_t$	0.00%	0.00%	0.00%	0.04%
	$\hat{V}_{t,\mathrm{MA}(1)}$	94.90%	95.42%	95.12%	82.22%

This table reports the coverage probabilities of 95% confidence intervals using RV_t, RK_t^F and 0.95 density intervals using $\hat{V}_t$ and $\hat{V}_{\mathrm{MA}(1)}$ based on 5000 days results for different data generating processes. The price is contaminated with white noise.

Table 3.8: RMSE of RV_t, RK_t^N, $\hat{V}_t$, $\hat{V}_{t,\mathrm{MA}(1)}$ and $\hat{V}_{t,\mathrm{MA}(2)}$ (Dependent Microstructure Error Case)

Data Freq.	Estimator	GARCH	SV1F	SV1FJ	SV2F
5-minute	RMSE(RV_t)	0.21825	0.39266	0.41585	0.58520
	RMSE(RK_t^N)	0.23575	0.44343	0.45080	0.89975
	RMSE($\hat{V}_t$)	**0.21593**	**0.38823**	**0.41013**	**0.54514**
	RMSE($\hat{V}_{t,\mathrm{MA}(1)}$)	0.22309	0.40177	0.42160	0.84072
	RMSE($\hat{V}_{t,\mathrm{MA}(2)}$)	0.29148	0.53858	0.57819	1.17410
1-minute	RMSE(RV_t)	0.84121	1.48399	1.60189	1.6954
	RMSE(RK_t^N)	0.14158	0.25780	0.26987	**0.52030**
	RMSE($\hat{V}_t$)	0.84222	1.48565	1.60134	1.67811
	RMSE($\hat{V}_{t,\mathrm{MA}(1)}$)	**0.11496**	**0.20471**	**0.21227**	0.52199
	RMSE($\hat{V}_{t,\mathrm{MA}(2)}$)	0.13738	0.24860	0.26145	0.62091
30-second	RMSE(RV_t)	1.66229	2.95397	3.19560	3.37090
	RMSE(RK_t^N)	0.11918	0.21559	0.22306	0.42729
	RMSE($\hat{V}_t$)	1.66431	2.95754	3.19640	3.35928
	RMSE($\hat{V}_{t,\mathrm{MA}(1)}$)	**0.08864**	**0.15827**	**0.16826**	**0.34777**
	RMSE($\hat{V}_{t,\mathrm{MA}(2)}$)	0.10526	0.18883	0.19355	0.39105
10-second	RMSE(RV_t)	4.40694	7.81961	8.49852	7.85934
	RMSE(RK_t^N)	0.09850	0.18004	0.18376	0.34594
	RMSE($\hat{V}_t$)	4.41064	7.82610	8.50079	7.85264
	RMSE($\hat{V}_{t,\mathrm{MA}(1)}$)	0.16416	0.30819	0.30435	0.89896
	RMSE($\hat{V}_{t,\mathrm{MA}(2)}$)	**0.06928**	**0.12831**	**0.13609**	**0.25218**

This table reports the root mean squared error (RMSE) of estimating 5000 daily ex-post variances using RV_t, RK_t^N and Bayesian nonparametric estimators $\hat{V}_t$, $\hat{V}_{t,\mathrm{MA}(1)}$ and $\hat{V}_{t,\mathrm{MA}(2)}$ based on returns at different frequencies and simulated from 4 DGPs. The observed prices contains microstructure noise that is dependent with returns.

Table 3.9: Biases of RV_t, RK_t^N, $\hat{V}_t$, $\hat{V}_{t,\text{MA}(1)}$ and $\hat{V}_{t,\text{MA}(2)}$ (Dependent Microstructure Error Case)

Data Freq.	Estimator	GARCH	SV1F	SV1FJ	SV2F
5-minute	Bias(RV_t)	0.16032	0.28455	0.30733	0.17262
	Bias(RK_t^N)	**0.01349**	**0.02985**	**0.03232**	0.00733
	Bias($\hat{V}_t$)	0.16078	0.28532	0.30711	0.16238
	Bias($\hat{V}_{t,\text{MA}(1)}$)	0.02134	0.03879	0.04075	**0.00062**
	Bias($\hat{V}_{t,\text{MA}(2)}$)	0.05647	0.10334	0.11660	0.04083
1-minute	Bias(RV_t)	0.81057	1.42504	1.54563	0.87166
	Bias(RK_t^N)	0.02421	0.04351	0.04360	0.01839
	Bias($\hat{V}_t$)	0.81167	1.42695	1.54552	0.86862
	Bias($\hat{V}_{t,\text{MA}(1)}$)	**0.00909**	**0.01674**	**0.01661**	-0.01113
	Bias($\hat{V}_{t,\text{MA}(2)}$)	0.01724	0.03180	0.02990	**-0.00565**
30-second	Bias(RV_t)	1.61481	2.85837	3.10192	1.72912
	Bias(RK_t^N)	0.02791	0.04940	0.05114	0.02369
	Bias($\hat{V}_t$)	1.61861	2.86192	3.10304	1.72808
	Bias($\hat{V}_{t,\text{MA}(1)}$)	**0.00740**	**0.01384**	**0.00991**	-0.01316
	Bias($\hat{V}_{t,\text{MA}(2)}$)	0.01097	0.02012	0.01847	**-0.01156**
10-second	Bias(RV_t)	4.32800	7.65381	8.34221	4.67328
	Bias(RK_t^N)	0.04034	0.07209	0.07321	0.04327
	Bias($\hat{V}_t$)	4.33163	7.66022	8.34491	4.65505
	Bias($\hat{V}_{t,\text{MA}(1)}$)	0.10993	0.20159	0.20140	0.13645
	Bias($\hat{V}_{t,\text{MA}(2)}$)	**0.00656**	**0.01333**	**0.00872**	**-0.01857**

This table reports the bias estimates from 5000 daily ex-post variances using RV, RK^N and Bayesian nonparametric estimators $\hat{V}$, $\hat{V}_{\text{MA}(1)}$ and $\hat{V}_{\text{MA}(2)}$ based on returns at different frequencies and simulated from 4 DGPs. The observed prices contains microstructure noise that is dependent with returns.

Table 3.10: Coverage Probabilities (Dependent Microstructure Error Case)

Data Freq.	Interval Estimator	GARCH	SV1F	SV1FJ	SV2F
5-minute	RV_t	76.22%	74.00%	73.12%	21.14%
	RK_t^N - Infeasible	87.26%	87.62%	87.64%	76.72%
	RK_t^N - Feasible	91.16%	91.34%	92.02%	96.42%
	$\hat{V}_t$	65.43%	63.34%	73.12%	21.14%
	$\hat{V}_{t,\mathrm{MA}(1)}$	94.28%	94.42%	93.94%	89.40%
	$\hat{V}_{t,\mathrm{MA}(2)}$	94.72%	94.72%	94.14%	89.74%
1-minute	RV_t	0.00%	0.00%	0.10%	0.06%
	RK_t^N - Infeasible	90.02%	90.40%	89.98%	71.70%
	RK_t^N - Feasible	99.80%	99.80%	99.70%	99.46%
	$\hat{V}_t$	0.00%	0.00%	0.04%	0.04%
	$\hat{V}_{t,\mathrm{MA}(1)}$	94.68%	95.08%	94.76%	87.20%
	$\hat{V}_{t,\mathrm{MA}(2)}$	94.70%	94.66%	94.34%	86.80%
30-second	RV_t	0.00%	0.00%	0.00%	0.00%
	RK_t^N - Infeasible	91.50%	91.72%	91.26%	70.94%
	RK^N - Feasible	100.00%	100.00%	100.00%	99.96%
	$\hat{V}$	0.00%	0.00%	0.00%	0.00%
	$\hat{V}_{\mathrm{MA}(1)}$	94.76%	95.30%	94.90%	85.40%
	$\hat{V}_{\mathrm{MA}(2)}$	94.90%	94.50%	94.76%	85.84%
10-second	RV_t	0.00%	0.00%	0.00%	0.00%
	RK_t^N - Infeasible	91.90%	92.44%	92.30%	69.72%
	RK_t^N - Feasible	100.00%	100.00%	100.00%	100.00%
	$\hat{V}_t$	0.00%	0.00%	0.00%	0.00%
	$\hat{V}_{t,\mathrm{MA}(1)}$	64.94%	65.70%	67.90%	78.84%
	$\hat{V}_{t,\mathrm{MA}(2)}$	94.42%	95.34%	95.12%	82.36%

This table reports the coverage probabilities of 95% confidence intervals of RV, RK^N and 0.95 density intervals of Bayesian nonparametric estimators $\hat{V}$, $\hat{V}_{\mathrm{MA}(1)}$ and $\hat{V}_{\mathrm{MA}(2)}$ based on 5000 days results. The observed prices contains microstructure noise that is dependent with returns.

Table 3.11: Summary Statistics: IBM Variance Measures

Frequency	Data	Mean	Median	Var	Skew.	Kurt.	Min	Max
Daily	r_t	0.0673	0.0656	1.6046	0.2069	8.3059	-6.4095	12.2777
	r_t^2	1.6091	0.4352	18.9654	13.9087	387.4812	0.0000	150.7429
5-minute	RV_t	1.8353	0.9458	11.9867	9.5887	148.2622	0.1032	76.2901
	RK_t^F	1.6613	0.8447	9.3647	8.5539	124.8480	0.0375	71.9626
	RK_t^N	1.6670	0.8476	8.8872	8.0467	109.1098	0.0556	66.3995
	$\hat{V}_t$	1.7839	0.9211	10.5246	8.5070	116.9235	0.1080	70.2483
	$\hat{V}_{t,\mathrm{MA}(1)}$	1.6686	0.8422	9.2512	7.3019	79.49682	0.0284	52.9256
	$\hat{V}_{t,\mathrm{MA}(2)}$	1.6997	0.8486	10.1324	8.4690	118.2698	0.0118	72.2393
1-minute	RV_t	2.0004	1.0468	13.5019	10.5704	202.6835	0.1535	103.8773
	RK_t^F	1.7952	0.9163	10.8043	8.3092	113.5727	0.1006	73.8576
	RK_t^N	1.7425	0.8973	9.6499	7.7187	94.7830	0.0897	60.2024
	$\hat{V}_t$	1.9653	1.0306	13.0413	10.7078	209.5183	0.1523	103.2695
	$\hat{V}_{t,\mathrm{MA}(1)}$	1.8326	0.9028	11.2766	7.4122	82.9604	0.1077	61.9220
	$\hat{V}_{t,\mathrm{MA}(2)}$	1.7895	0.8946	10.8954	8.4448	120.2637	0.1058	75.0062
	$\hat{V}_{t,\mathrm{MA}(3)}$	1.7392	0.8814	9.6590	7.8399	102.1530	0.0968	63.0524
	$\hat{V}_{t,\mathrm{MA}(4)}$	1.7103	0.8688	9.0700	7.2766	84.7264	0.0971	55.1365

This table reports the summary statistics of ex-post variance estimators based on 5-minute and 1-minute returns, along with the summary statistics of daily return and daily squared return. The number of daily observation is 3764.

Table 3.12: HAR and HARQ Model Regression Results Based on IBM Ex-post Variance Estimators

Data Freq.	Parameter	HAR		HARQ	
		RK_t^F	$\hat{V}_{t,\text{MA(1)}}$	RK_t^F	$\hat{V}_{t,\text{MA(1)}}$
5-minute	β_0	0.1322	0.1233	0.1015	-0.0124
		(0.0374)	(0.0376)	(0.0382)	(0.0394)
	β_1	0.1926	0.2432	0.2341	0.4585
		(0.0196)	(0.0197)	(0.0224)	(0.0284)
	β_2	0.5649	0.4871	0.5664	0.4350
		(0.0332)	(0.0330)	(0.0331)	(0.0329)
	β_3	0.1598	0.1929	0.1422	0.1475
		(0.0286)	(0.0282)	(0.0289)	(0.0282)
	β_{1Q}	-	-	-0.0012	-0.0197
				(0.0003)	(0.0019)
	R-squared	57.74%	59.33%	57.90%	60.45%
Data Freq.	Parameter	HAR		HARQ	
		RK_t^N	$\hat{V}_{t,\text{MA(4)}}$	RK_t^N	$\hat{V}_{t,\text{MA(4)}}$
1-minute	β_0	0.1246	0.1297	0.0065	-0.0337
		(0.0365)	(0.0374)	(0.0367)	(0.0390)
	β_1	0.2493	0.2464	0.4464	0.5138
		(0.0195)	(0.0196)	(0.0242)	(0.0288)
	β_2	0.5435	0.5154	0.5033	0.4531
		(0.0318)	(0.0321)	(0.0312)	(0.0319)
	β_3	0.1331	0.1598	0.0708	0.0914
		(0.0265)	(0.0271)	(0.0263)	(0.0272)
	β_{1Q}	-	-	-0.0031	-0.0328
				(0.0002)	(0.0026)
	R-squared	62.71%	60.38%	64.39%	61.96%

[1] This table reports OLS regression results for the HAR and HARQ model. The results in top panel are based on RK_t^F and $\hat{V}_{t,\text{MA(1)}}$ calculated using 5-minute returns and the bottom panel shows the results of 1-minute RK_t^N and $\hat{V}_{t,\text{MA(4)}}$. The values in brackets are standard error of coefficients.

[2] Sample period: 2001/01/03 - 2016/02/16, 3764 observations.

Table 3.13: Out-of-Sample Forecasts of IBM Variance

Panel A: 5-minute Return			
Dependent Variable	Regressors	HAR	HARQ
5-minute RK_t^F	RK_t^F	1.84113	1.84444
	$\hat{V}_{t,\text{MA(1)}}$	**1.84083**	**1.81263**
5-minute $\hat{V}_{t,\text{MA(1)}}$	RK_t^F	1.86262	1.86699
	$\hat{V}_{t,\text{MA(1)}}$	**1.85581**	**1.83220**
Panel B: 1-minute Return			
Dependent Variable	Regressors	HAR	HARQ
1-minute RK_t^N	RK_t^N	1.87539	**1.82881**
	$\hat{V}_{t,\text{MA(4)}}$	**1.87006**	1.83173
1-minute $\hat{V}_{t,\text{MA(4)}}$	RK_t^N	1.93618	1.88542
	$\hat{V}_{t,\text{MA(4)}}$	**1.92428**	**1.87654**

[1] This table reports the root mean squared forecast error (RMSFE) of forecasting next period ex-post variance using both classical and Bayesian nonparametric variance estimator. Both HAR and HARQ model are considered. The forecasting target is the dependent variable one period out-of-sample.

[2] On each day, the model parameters are re-estimated using all the data up to that day.

[3] Out of sample period: 2005/01/03 - 2016/02/16, 2773 days.

Table 3.14: Summary Statistics: Disney Variance Measures

Frequency	Data	Mean	Median	Var	Skew.	Kurt.	Min	Max
Daily	r_t	-0.0389	0.0181	0.9636	-0.5750	5.8989	-4.1621	3.4451
	r_t^2	0.9651	0.2398	4.6847	4.6843	28.5354	0.0000	17.3236
5-minute	RV_t	1.2881	0.8274	4.0558	7.8379	85.1183	0.1740	25.3443
	RK_t^F	1.2949	0.7435	5.3934	7.9331	83.5863	0.0692	28.54962
	$\hat{V}_t$	1.2485	0.7907	3.7617	7.9542	87.3439	0.1812	24.5851
	$\hat{V}_{t,\mathrm{MA}(1)}$	1.3119	0.8102	5.7790	8.2469	88.7450	0.0833	30.0731
1-minute	RV_t	1.3018	0.9177	2.7262	7.6683	83.5494	0.2024	20.9397
	RK_t^F	1.2727	0.8019	4.4924	8.7883	102.4300	0.1373	27.8559
	$\hat{V}_t$	1.2783	0.8984	2.6164	7.6427	83.0711	0.2033	20.4880
	$\hat{V}_{t,\mathrm{MA}(1)}$	1.2587	0.8365	3.6509	8.7573	102.9476	0.1751	25.2803
30-second	RV_t	1.3077	0.9536	2.4310	7.2835	76.3258	0.2232	19.3896
	RK_t^F	1.2558	0.8224	3.3762	8.2074	92.6638	0.1744	23.7206
	RK_t^N	1.2559	0.8255	3.8733	8.4716	96.7899	0.1352	25.5732
	$\hat{V}_t$	1.2876	0.9366	2.3030	7.1143	73.1945	0.2238	18.6778
	$\hat{V}_{t,\mathrm{MA}(1)}$	1.2154	0.8546	2.6036	7.5516	80.0820	0.1854	20.1312
	$\hat{V}_{t,\mathrm{MA}(2)}$	1.2478	0.8557	3.3850	8.4036	95.4195	0.1618	23.8698

This table reports the summary statistics of ex-post variance estimators based on 5-minute, 1-minute and 30-second Disney returns, along with the summary statistics of daily return and daily squared return. Sample period: 01/03/2015 - 12/29/2015.

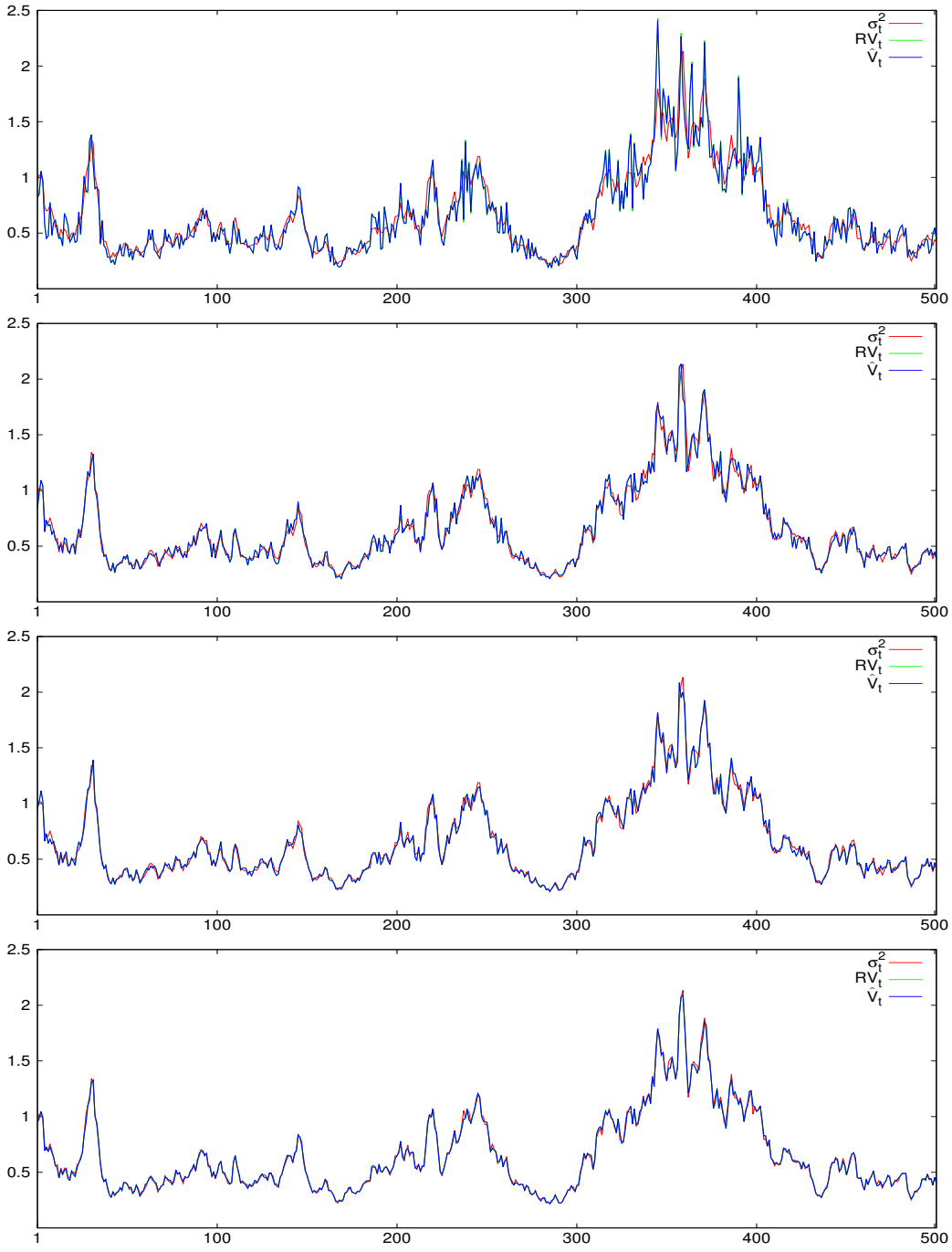

Figure 3.1: True Variance σ_t^2, RV_t and $\hat{V}_t$ (No Microstructure Noise Case). From top to bottom: 5-minute, 1-minute, 30-second, 10-second returns simulated from GARCH(1,1) DGP without noise.

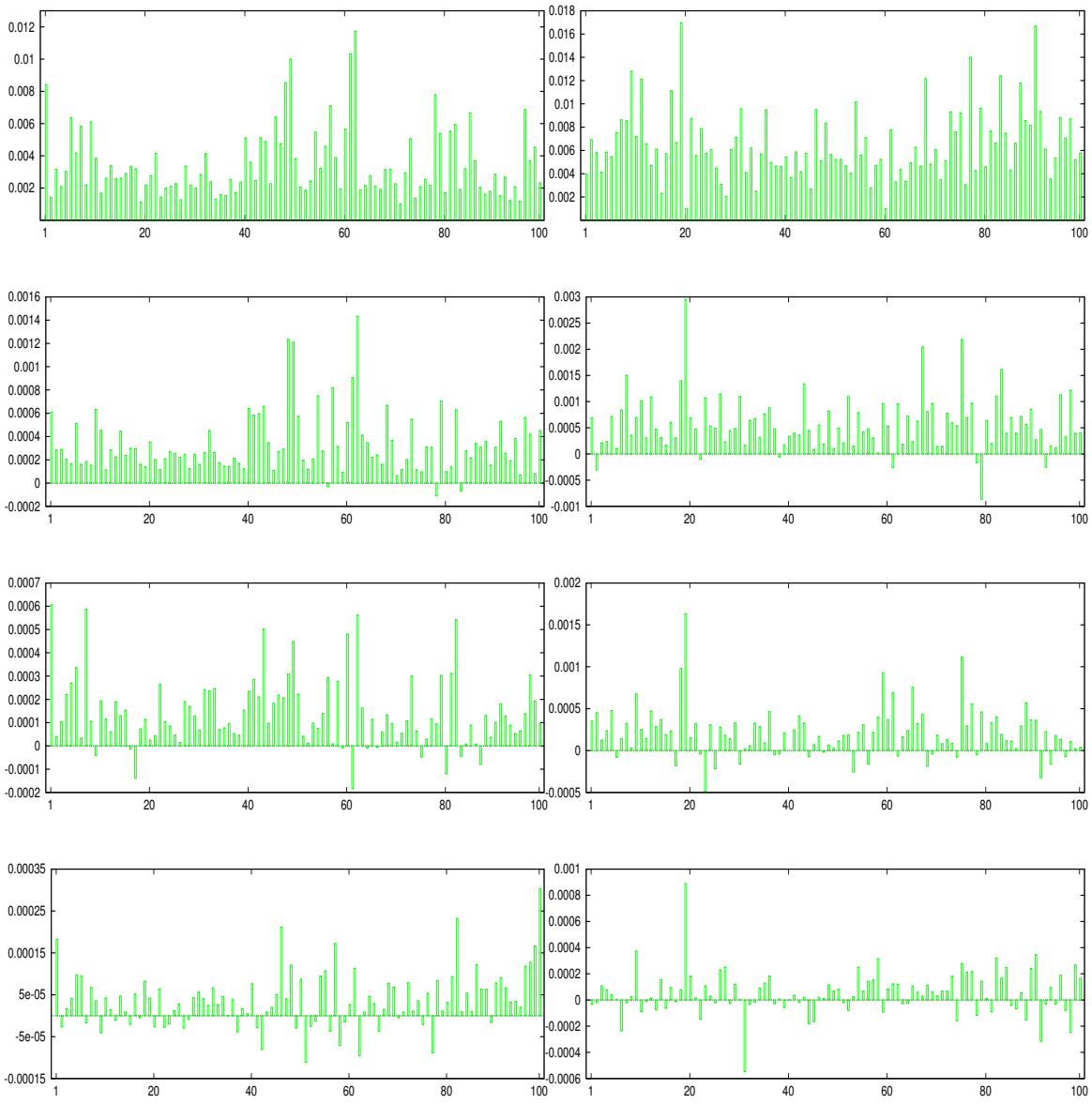

Figure 3.2: RMSE(RV_t)-RMSE($\hat{V}_t$) in 100 subsamples (No Microstructure Noise Case). Left: GARCH(1,1) DGP, Right: SV1F DGP, From top to bottom: 5-minute, 1-minute, 30-second, 10-second returns.

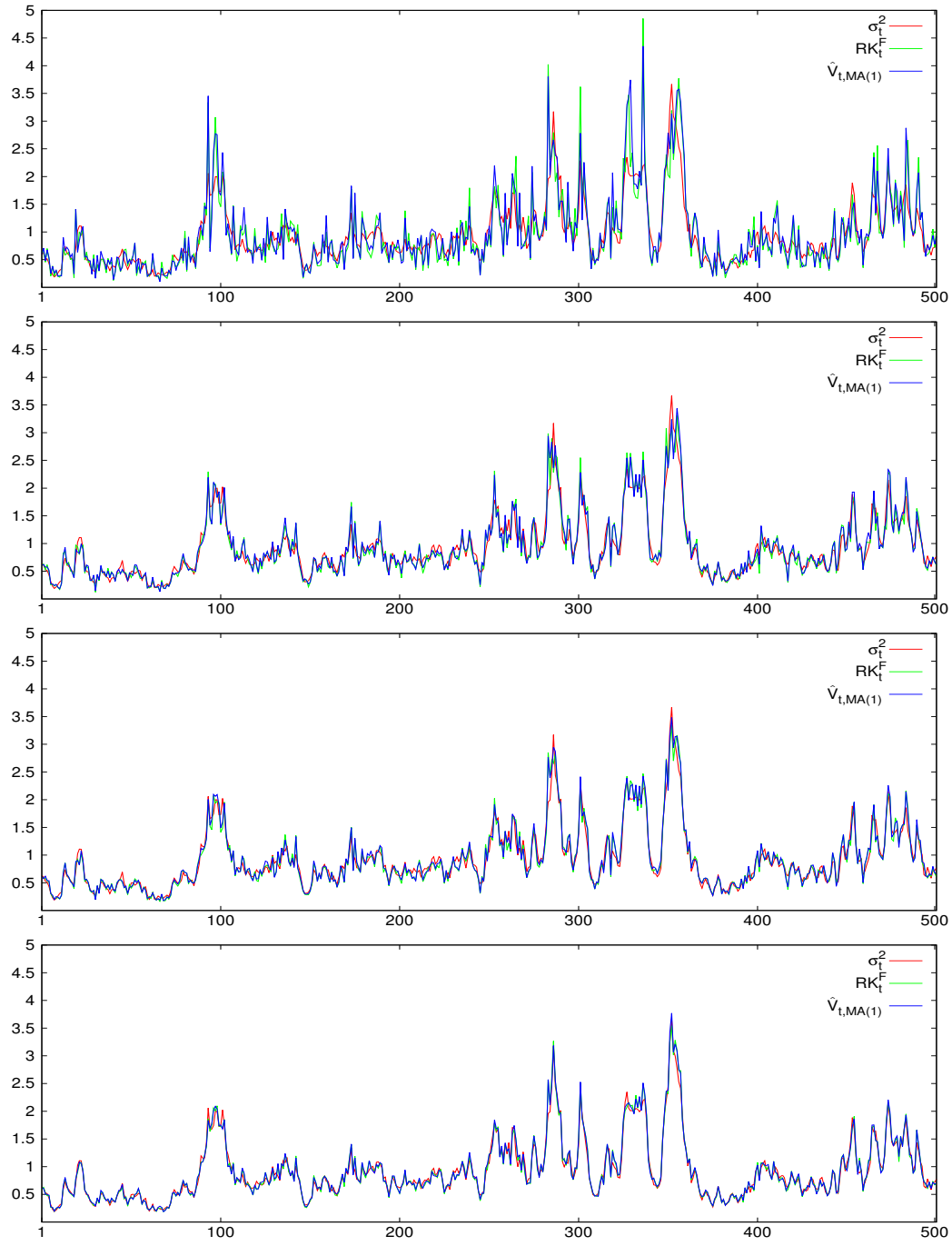

Figure 3.3: True Variance σ_t^2, RK_t^F and $\hat{V}_{t,\mathrm{MA}(1)}$ (Independent Microstructure Noise Case). From top to bottom: 5-minute, 1-minute, 30-second, 10-second returns simulated from SV1F DGP with independent noise.

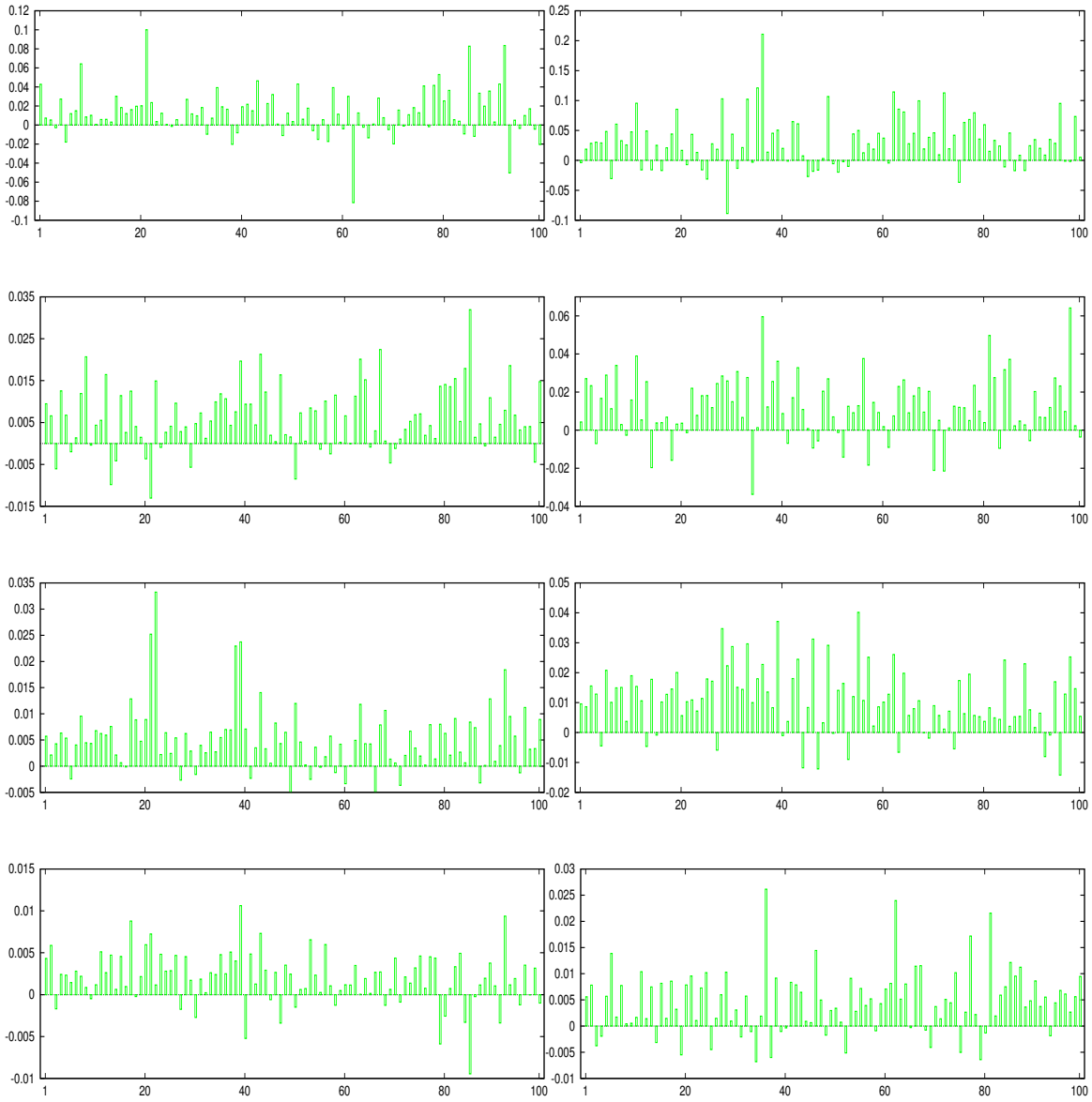

Figure 3.4: RMSE(RK_t^F)-RMSE($\hat{V}_{t,\mathrm{MA}(1)}$) in 100 subsamples (Independent Microstructure Noise Case). Left: GARCH(1,1) DGP, Right: SV1F DGP, From top to bottom: 5-minute, 1-minute, 30-second, 10-second returns.

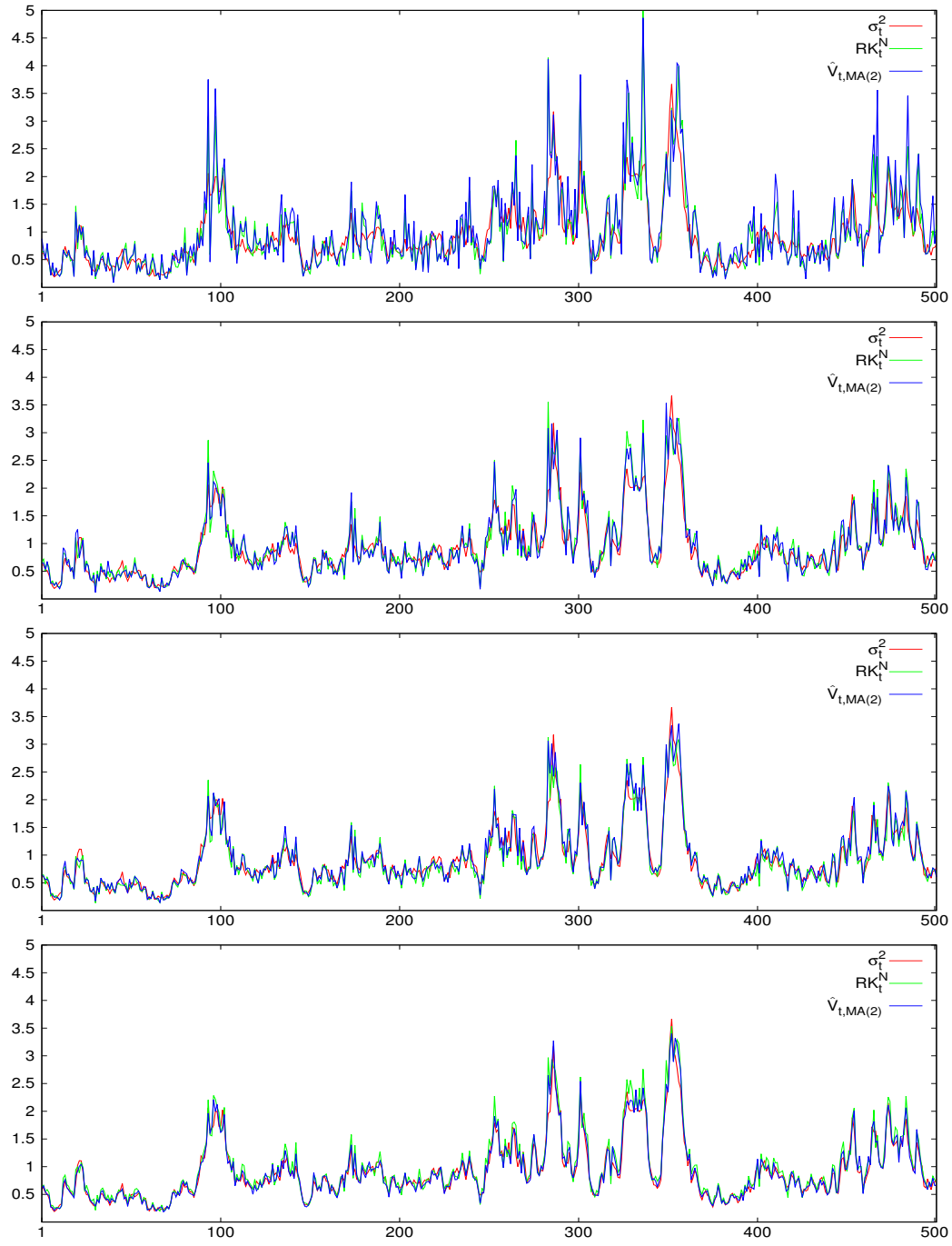

Figure 3.5: True Variance σ_t^2, RK_t^N and $\hat{V}_{t,\mathrm{MA}(2)}$ (Dependent Microstructure Noise Case). From top to bottom: 5-minute, 1-minute, 30-second, 10-second returns simulated from SV1F DGP with noise correlated with returns.

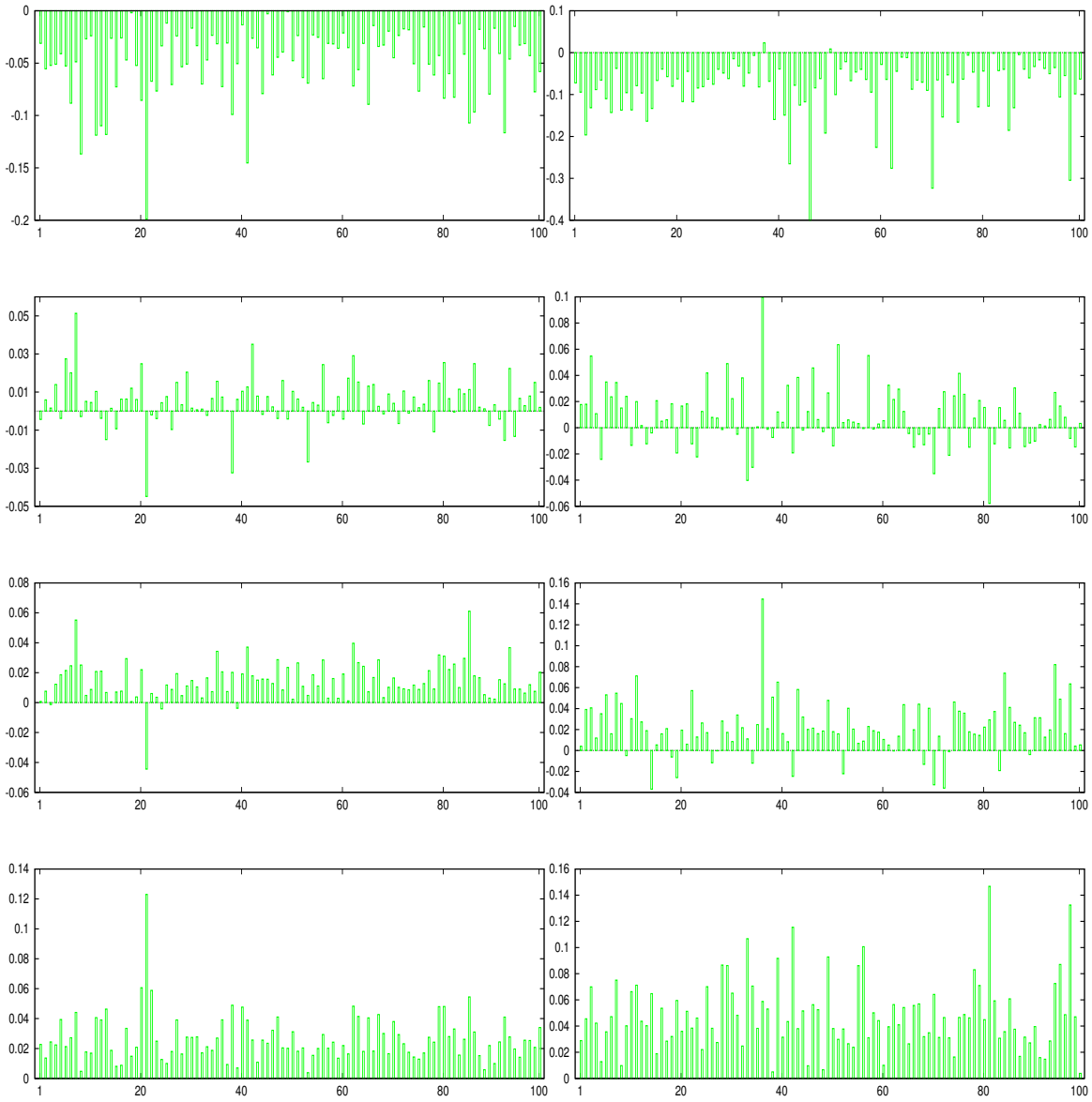

Figure 3.6: RMSE(RK_t^N)-RMSE($\hat{V}_{t,\mathrm{MA}(2)}$) in 100 subsamples (Dependent Microstructure Noise Case). Left: GARCH(1,1) DGP, Right: SV1F DGP, From top to bottom: 5-minute, 1-minute, 30-second, 10-second returns.

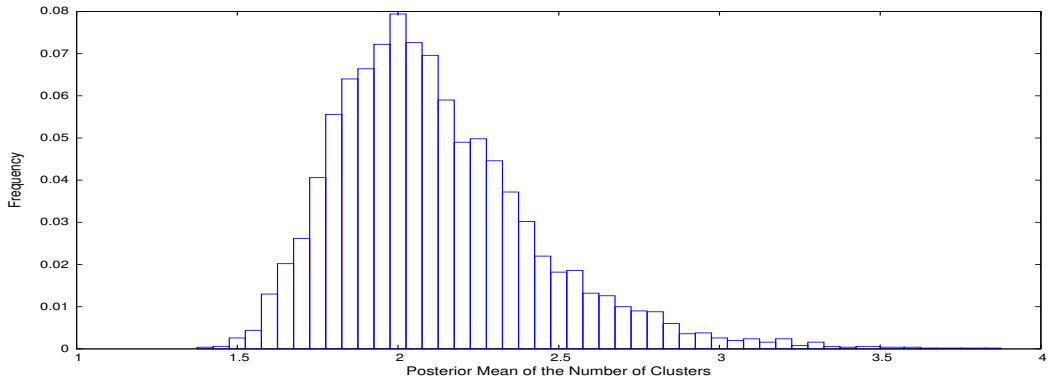

Figure 3.7: Posterior Mean of the Number of Clusters, K. Model: DPM. Data: 5-minute return without microstructure noise from SV1F.

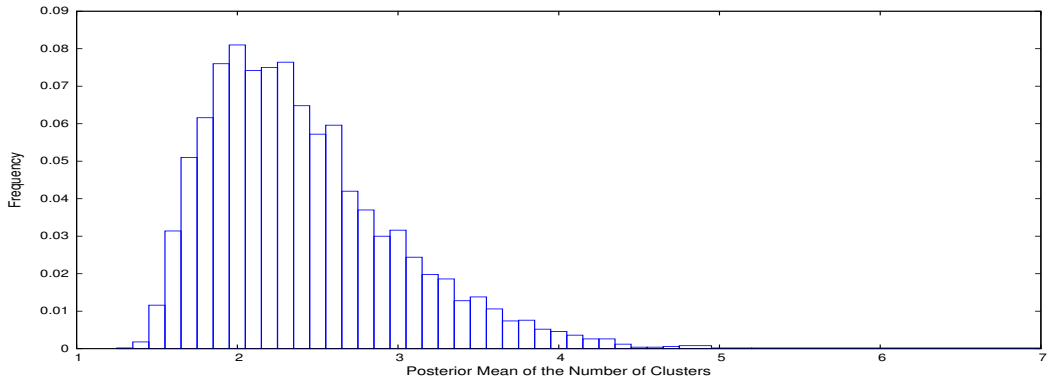

Figure 3.8: Posterior Mean of the Number of Clusters, K. Model: DPM-MA(1). Data: 1-minute return with independent noise from SV1FJ

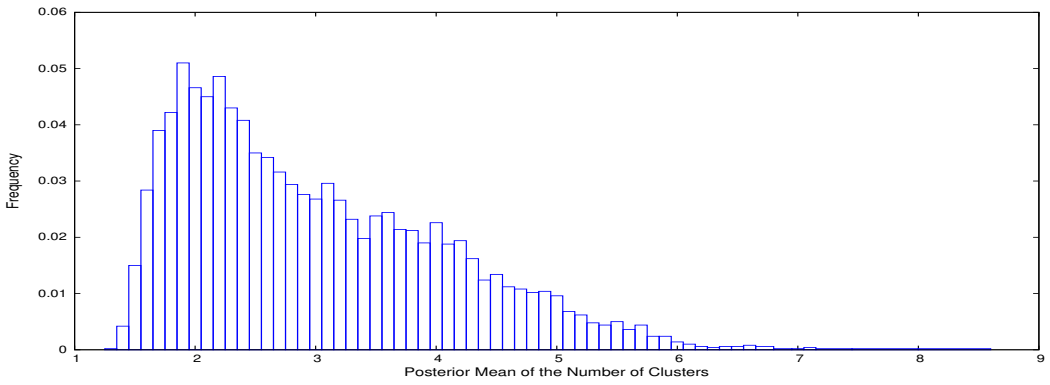

Figure 3.9: Posterior Mean of the Number of Clusters, K. Model: DPM-MA(2). Data: 30-second return with dependent noise from SV2F.

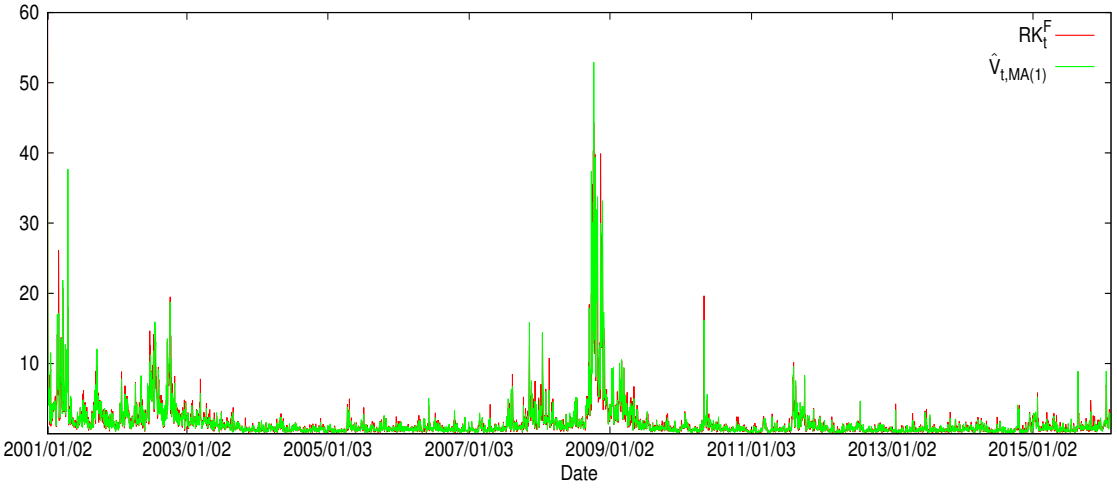

Figure 3.10: RK_t^F and $\hat{V}_{t,\mathrm{MA}(1)}$ based on 5-minute IBM returns

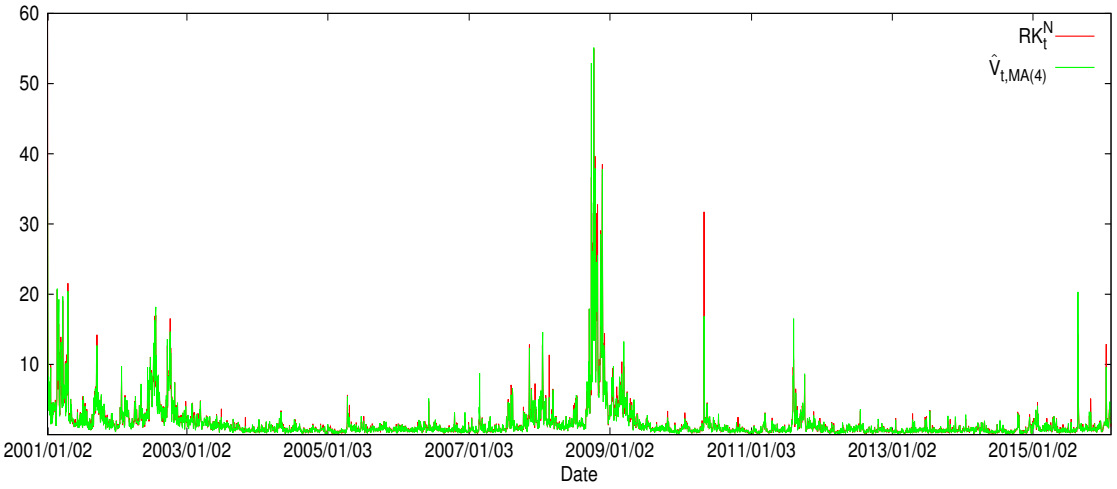

Figure 3.11: RK_t^N and $\hat{V}_{t,\mathrm{MA}(4)}$ based on 1-minute IBM returns

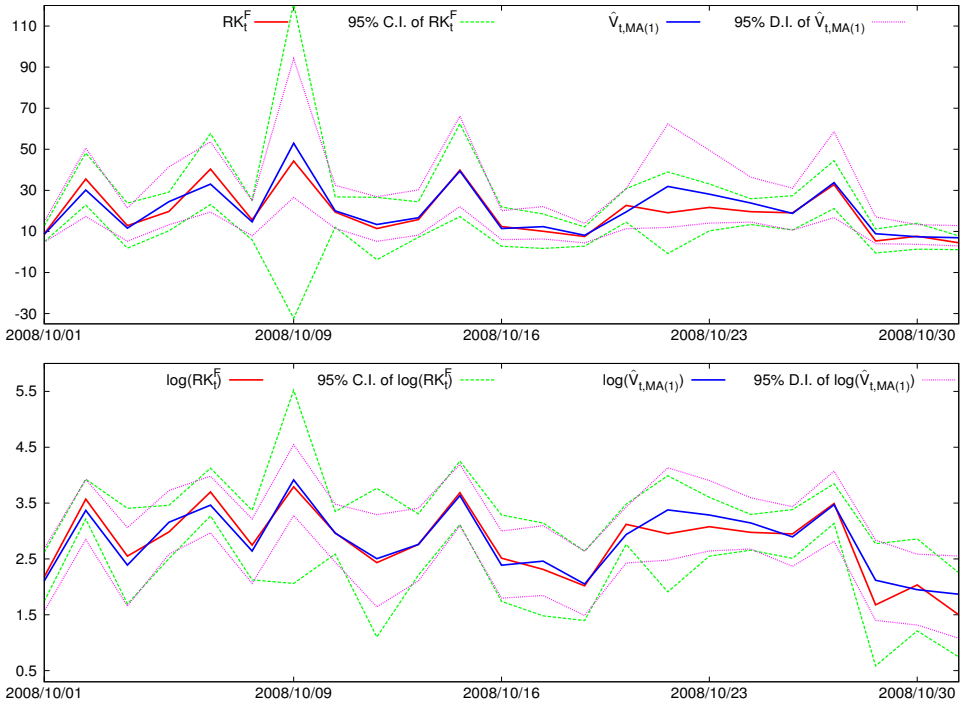

Figure 3.12: High Volatility Period: RK_t^F and $\hat{V}_{t,\mathrm{MA}(1)}$ calculated using 5 minute IBM returns. Top: variance, below: log-variance

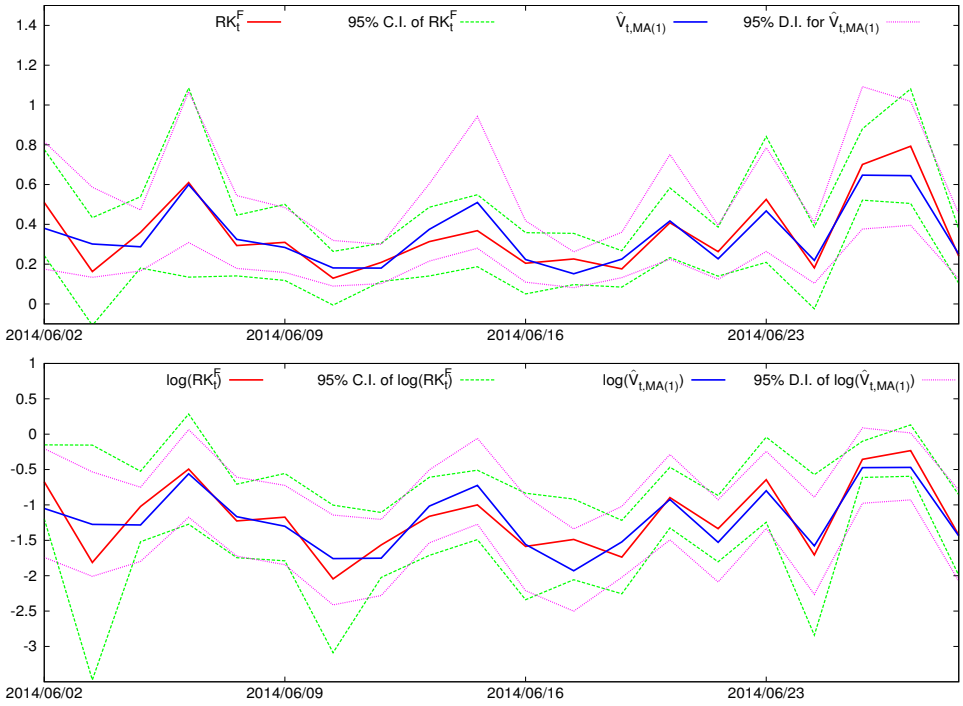

Figure 3.13: Low Volatility Period: RK_t^F and $\hat{V}_{t,\mathrm{MA}(1)}$ calculated using 5 minute IBM returns. Top: variance, below: log-variance

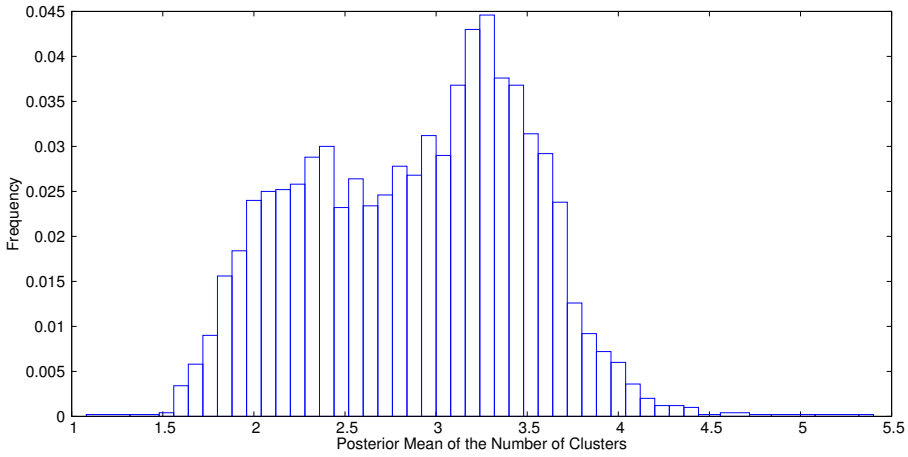

Figure 3.14: Posterior Mean of the Number of Clusters, K (Based on 3764 days results from DPM-MA(1) using 5-minute IBM returns).

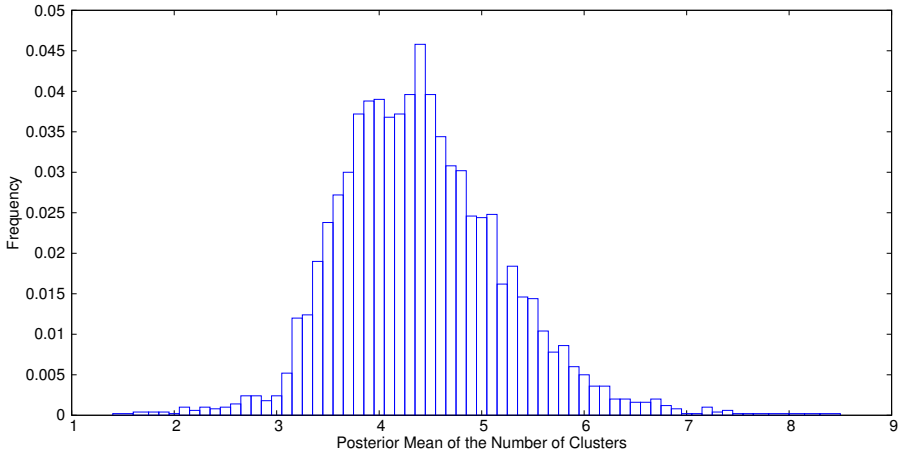

Figure 3.15: Posterior Mean of the Number of Clusters, K (Based on 3764 days results from DPM-MA(4) using 1-minute IBM returns).

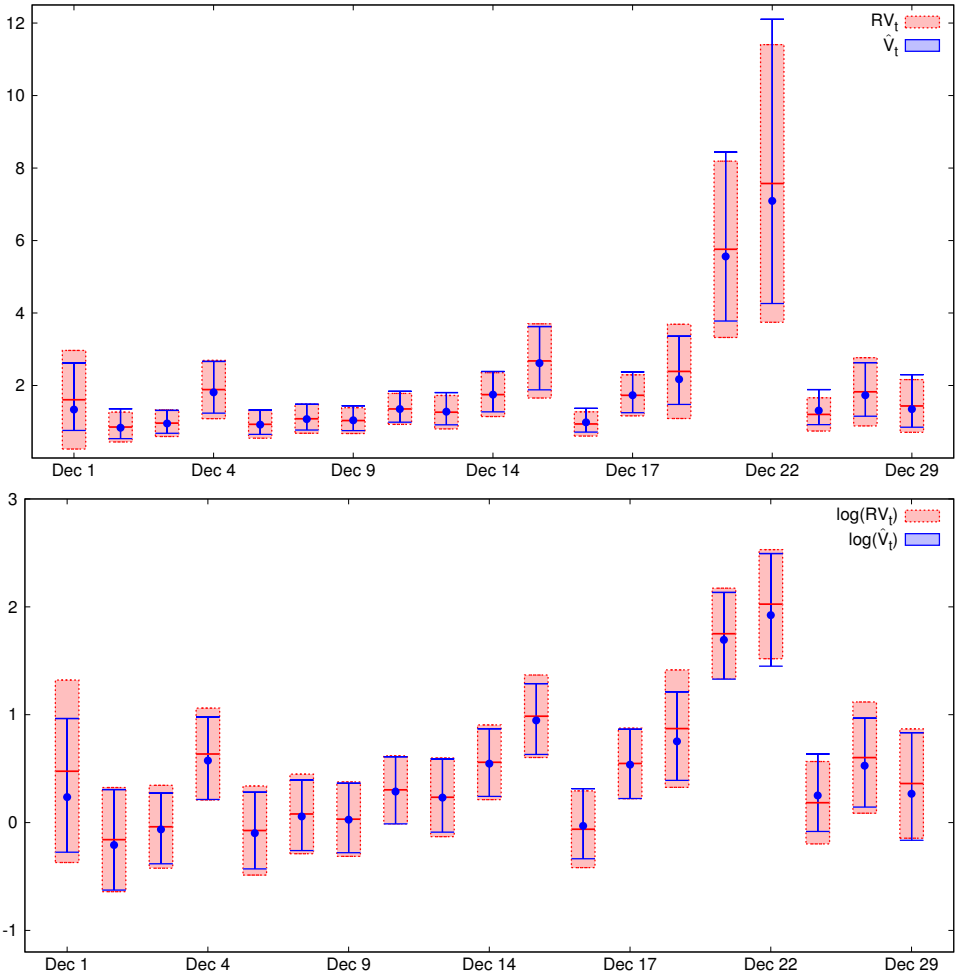

Figure 3.16: RV_t and $\hat{V}_t$ Based on 5-minute Disney Returns in December 2015. Top: Variance, Below: Log-variance

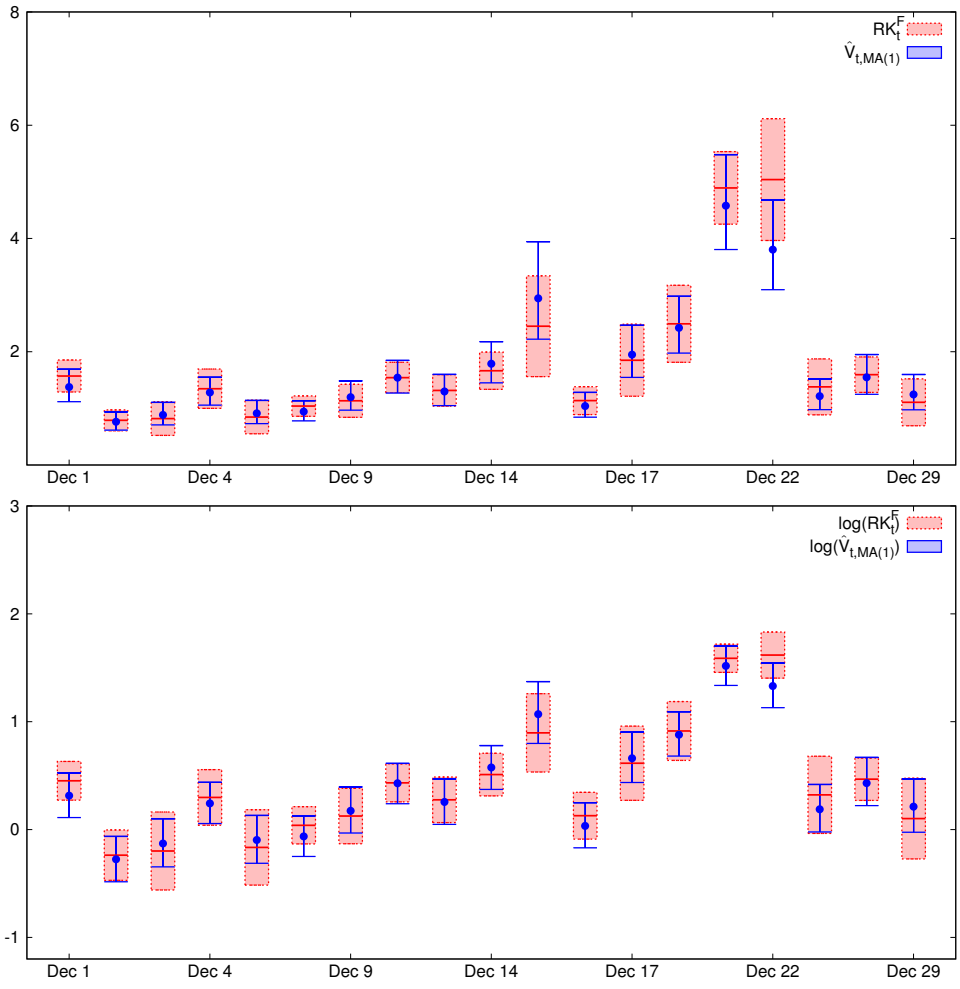

Figure 3.17: RK_t^F and $\hat{V}_{t,\mathrm{MA}(1)}$ Based on 30-second Disney Returns in December 2015. Top: Variance, Below: Log-variance

Chapter 4

Bayesian Nonparametric Covariance Estimation with Noisy and Nonsynchronous Asset Prices

4.1 Introduction

The univariate Bayesian nonparametric variance estimator introduced in Chapter 3 is extended to its multivariate version to allow pooling in covariance estimation. The method delivers exact finite sample inference and the estimated covariance matrix is guaranteed to be positive definite. In addition, a new way of synchronizing observations based on data augmentation is introduced. The estimator is designed for regularly or nonsynchronously spaced price data, with or without independent microstructure noise.

The covariance matrix of asset returns is the key input for many finance problems, such as portfolio allocation and asset pricing. Since the availability of high frequency

data, estimation of covariance using intraday returns has become a very active area of research. An important first step is Andersen *et al.* (2003) and Barndorff-Nielsen and Shephard (2004). They formalized the realized covariance estimator and showed it is an asymptotic consistent estimator of the integrated covariance, under the assumption that observations are free of measurement error. However, in reality, prices are contaminated with market microstructure noise and transactions arrive nonsynchronously, which lead to poor statistical performance of the realized covariance estimator.

Several different approaches have been used to pave the way for covariation estimation of noisy and nonsynchronously spaced prices. One branch of the literature is based on methods to synchronize returns and adjust the bias of estimators. Zhang (2011) suggested the optimal sampling frequency in constructing realized covariance and proposed the two scales estimator. Griffin and Oomen (2011) formally studied the realized covariance with lead-lag adjustments. Barndorff-Nielsen *et al.* (2011) introduced the multivariate realized kernel based on refresh time synchronization. Aït-Sahalia *et al.* (2010) proposed the Quasi-maximum likelihood estimator of covariance as well as the generalized synchronization method. The cumulative covariance estimator proposed by Hayashi and Yoshida (2005) exemplifies another branch of the literature. Their estimator can be applied directly to raw observations is unbiased under a no noise assumption. Voev and Lunde (2007) proposed a bias correction to make the cumulative covariance estimator suitable for noisy prices. Peluso *et al.* (2014) introduced a Bayesian estimator of the covariance of noisy and asynchronous returns based on a parametric model.

This chapter proposes a Bayesian nonparametric approach to estimate the ex-post

covariance matrix of a vector of asset returns. Instead of using the data independently, I exploit pooling among observations with a common covariance matrix. This is achieved with a use of Dirichlet process mixture model. The Bayesian nonparametric framework allows the number of groups or clusters of covariance matrices to vary flexibly and to be determined endogenously. To adjust for bias, I use a vector moving average model for high-frequency data and introduce pooling in this setting. From this model, a covariance estimator that corrects for market microstructure noise and nonsynchronous trading is derived based on Hansen *et al.* (2008).

In related work Peluso *et al.* (2014) introduced data augmentation based on dynamic linear model to synchronize the high-frequency prices. This chapter also uses data augmentation to synchronize the data but exploits pooling to increase estimation accuracy. The proposed synchronization method is built on top of the previous-tick method defined in Hansen and Lunde (2006) but eliminates the zero-return problem caused by stale prices. Missing observations are augmented as unknown variables and are estimated conditional on observed data and model structure. With the proposed synchronization method, the Bayesian nonparametric covariance estimator with moving average adjustment fully accounts for the nonsynchronous bias.

Another advantage of the Bayesian nonparametric covariance estimator is that it is guaranteed to be positive definite. Using an inverse Wishart distribution as the prior guarantees the sampled intraday covariance is always positive definite. In addition, with synchronization based on data augmentation, the zero returns caused by stale prices are removed, which ensures non-singular matrices.

Monte Carlo simulation is conducted to compare the Bayesian nonparametric covariance estimator with realized covariance and multivariate realized kernel given

regularly or nonsynchronously spaced data, with or without the influence of independent microstructure noise. The results shows the proposed estimator yields lower root-mean-square-errors and better finite sample results in estimating both diagonal and off-diagonal elements of the covariance matrix in most cases. Empirical applications to equity data show the Bayesian covariance estimator captures similar time series dynamics of correlation and realized beta as the multivariate realized kernel. Using a volatility-timing strategy, the minimum variance portfolio based on the Bayesian nonparametric estimator outperforms ones based on realized covariance or multivariate realized kernel in terms of Sharpe ratio and utility level of an investor. Moreover, the Bayesian approach provides the exact distribution of the covariance, which allows users to analyze how the optimal weights and return of a portfolio are influenced by the covariance uncertainty.

This chapter is organized as follows. In Section 4.2, the Bayesian nonparametric model, daily covariance estimator and model estimation steps are discussed. Section 4.3 discusses the model and estimator that account for bias caused by noise and nonsynchronous trading. Section 4.4 illustrates the synchronization method with data augmentation. Section 4.5 conducts data simulation and compares the proposed estimator with competing alternatives. Empirical applications are found in Section 4.6. Section 4.7 concludes followed by an appendix.

4.2 Bayesian Nonparametric Covariance Estimation

This section starts by defining the target quantity, then illustrates the Bayesian nonparametric model and the estimator suitable for regularly spaced returns without microstructure noise.

4.2.1 Ex-post Daily Covariance

Suppose the log-prices of d assets are generated from

$$\mathrm{d}P(t) = \mu(t)\mathrm{d}t + \Pi(t)\mathrm{d}W(t), \tag{4.83}$$

where $\mu(t)$ is the drift term, $\Sigma(t) = \Pi(t)\Pi(t)'$ is the instantaneous covariance matrix and $W(t)$ stands for a standard Brownian motion vector.

As the true measure of the ex-post daily covariance, the integrated covariance is the quantity of interest and is defined as

$$\mathbf{V}_t = \int_{t-1}^{t} \Pi(\tau)\Pi(\tau)'\mathrm{d}\tau. \tag{4.84}$$

Let $P_{t,i} = (p_{t,i}^{(1)}, \dots p_{t,i}^{(d)})$ denotes the regularly spaced log-price vector, where $i = 1, \dots, n_t$ and $p_{t,i}^{(j)}$ denotes the i^{th} log-price of asset j on day t. Given intraday return $R_{t,i} = P_{t,i} - P_{t,i-1}$, the realized covariance ($\mathbf{RC}_t$) discussed in Andersen *et al.* (2003) and Barndorff-Nielsen and Shephard (2004) is defined as

$$\mathbf{RC}_t = \sum_{i=1}^{n_t} R_{t,i}R_{t,i}'. \tag{4.85}$$

Under a no microstructure noise setting, $\mathbf{RC}_t$ converges to $\mathbf{V}_t$ as $n_t \to \infty$. In finite samples, the summation of cross products of intraday returns provides a noisy estimator of $\mathbf{V}_t$ and the finite sample distribution of $\mathbf{RC}_t$ is unknown and must be approximated from the asymptotic distribution derived by Barndorff-Nielsen and Shephard (2004).

Instead of treating all intraperiod returns independently, the proposed approach allows data to cluster. Returns with similar intraday covariance can be grouped flexibly under the Bayesian nonparametric framework. Taking advantage of pooling, the proposed method yields a less noisy covariance estimator, compared with existing alternatives.

4.2.2 DPM Model

The proposed Bayesian nonparametric model extends the univariate Dirichlet process mixture (DPM) model used in Chapter 3 to its multivariate version and is given as

$$R_{t,i}\big|\mu_t, \Sigma_{t,i} \overset{iid}{\sim} \mathrm{N}(\mu_t, \Sigma_{t,i}),\ i = 1, \ldots, n_t, \tag{4.86}$$

$$\Sigma_{t,i}\big|G_t \overset{iid}{\sim} G_t, \tag{4.87}$$

$$G_t\big|G_{0,t}, \alpha_t \sim \mathrm{DP}(\alpha_t, G_{0,t}), \tag{4.88}$$

$$G_{0,t} \equiv \mathrm{IW}(\Psi_t, \nu_t), \tag{4.89}$$

where $R_{t,i}$ is assumed to follow a multivariate normal distribution with constant return mean μ_t over i and state-dependent intraday covariance matrix $\Sigma_{t,i}$. The intraday returns within a day are modelled by a multivariate normal mixture and the data within a cluster shares the same intraday covariance.

As the distribution of $\Sigma_{t,i}$, G_t is a discrete distribution with a varying number of clusters. This is achieved with the help of Dirichlet process $\mathrm{DP}(\alpha_t, G_{0,t})$, which is the prior distribution of G_t. A draw from $\mathrm{DP}(\alpha_t, G_{0,t})$ is centred around the base distribution $G_{0,t}$ which is an inverse Wishart distribution denoted as $\mathrm{IW}(\Psi_t, \nu_t)$. The base measure guarantees the positive definiteness of $\Sigma_{t,i}$ as any draw from an inverse Wishart distribution is positive definite.

The hyperparameters Ψ_t and ν_t need to be calibrated day by day as the dynamics of asset volatility changes across time. The following method is suggested to determine the values of Ψ_t and ν_t. It is known that for matrix $X_t \sim \mathrm{IW}(\Psi_t, \nu_t)$, the mean of X_t and variance of diagonal elements of X_t are

$$E(X_t) = \frac{\Psi_t}{\nu_t - d - 1} \quad \text{and} \quad \mathrm{var}(X_t^{(j)}) = \frac{2(\Psi_t^{(jj)})^2}{(\nu_t - d - 1)^2(\nu_t - d - 3)}. \tag{4.90}$$

Expressing Ψ_t and ν_t in terms of $E(X_t)$ and $\mathrm{var}(X_t^{(j)})$ yields

$$\Psi_t = E(X_t)(\nu_t - d - 1) \quad \text{and} \quad \nu_t = \frac{2(E(X_t^{(j)}))^2}{\mathrm{var}(X_t^{(j)})} + d + 3. \tag{4.91}$$

Substituting $E(X_t)$ and $\mathrm{var}(X_t^{(j)})$ with $\frac{1}{n_t}\mathbf{RC}_t$ and $\widehat{\mathrm{var}}(r_{t,i}^{(j)2})$ yields one estimate of ν_t. ν_t is set to be the average of $\nu_t^{(j)}$ based on all d assets. Hyper parameters Ψ_t and ν_t are set to be

$$\Psi_t = \frac{\nu_t - d - 1}{n_t}\mathbf{RC}_t, \tag{4.92}$$

$$\nu_t = \frac{1}{d}\sum_{j=1}^{d} \frac{2(\mathbf{RC}_t^{(jj)})^2}{\widehat{\mathrm{var}}(r_{t,i}^{(j)2})n_t} + d + 3. \tag{4.93}$$

The number of clusters is influenced by the precision parameter α_t in the Dirichlet

process. As α_t increases, the number of cluster increases and the effect of pooling gets diminished. The extreme case is that each observation has its own cluster. In this case, the proposed estimator is analogous to $\mathbf{RC}_t$. To add flexibility, α_t is treated as a parameter and a hierarchical prior $\text{Ga}(a, b)$ is placed on it.

Finally, a shrinkage prior $\text{N}(0, \Lambda)$, where Λ is a diagonal matrix with small variance values, is used for μ_t.

4.2.3 Model Estimation

The model is estimated using Markov chain Monte Carlo (MCMC) techniques. I apply the slice sampler of Kalli *et al.* (2011) to the stick-breaking representation of the DPM. Expressing the DP prior as the stick-breaking representation by Sethuraman (1994), the DPM model can be written as

$$p(R_{t,i}|\mu_t, \{\Phi_{t,j}\}_{j=1}^{\infty}, \{w_{t,j}\}_{j=1}^{\infty}) = \sum_{j=1}^{\infty} w_{t,j}\text{N}(R_{t,i}|\mu_t, \Phi_{t,j}), \tag{4.94}$$

$$w_{t,1} = v_{t,1}, \quad w_{t,j} = v_{t,j}\prod_{l=1}^{j-1}(1 - w_{t,l}), \quad v_{t,j} \overset{iid}{\sim} \text{Beta}(1, \alpha_t), \tag{4.95}$$

where $w_{t,j}$ is the weight associated with the j^{th} component and $\Phi_{t,j}$ denotes the unique covariance matrix in cluster j.

In the slice sampling, a set of auxiliary variables $u_{t,1:n_t} = \{u_{t,1}, \ldots, u_{t,n_t}\}$ is introduced to slice the infinite state space to a finite one so that the sampling of model parameters is feasible. $u_{t,1:n_t}$ is sampled along with other parameters and randomly truncates the state space to $K_t = \sum_{j=1}^{\infty} \mathbb{1}(u_{t,i} < w_{t,j})$ at each MCMC iteration. The

joint distribution of $R_{t,i}$ and $u_{t,i}$ is given by

$$f\left(R_{t,i}, u_{t,i} \middle| \mu_t, \{\Phi_{t,j}\}_{j=1}^{\infty}, \{w_{t,j}\}_{j=1}^{\infty}\right) = \sum_{j=1}^{\infty} \mathbb{1}\left(u_{t,i} < w_{t,j}\right) \mathrm{N}\left(R_{t,i} | \mu_t, \Phi_{t,j}\right). \tag{4.96}$$

The original model (4.94) is recovered by integrating out $u_{t,i}$.

Next, introduce a set of latent state variables $s_{t,1:n_t} = \{s_{t,1}, \ldots, s_{t,n_t}\}$ that label each observation's cluster. Given $s_{t,i} \in \{1, 2, \ldots K_t\}$, then $\Sigma_{t,i} = \Phi_{t,s_{t,i}}$. Note that the number of clusters K_t is adjusted over MCMC iterations. A new cluster with covariance $\Phi_{t,K_t+1} \sim \mathrm{IW}(\Psi_t, \nu_t)$ can be opened and clusters with similar covariances can be merged.

The Gibbs sampler is used to sample μ_t and $\Phi_{t,j}$ for $j = 1, \ldots, K_t$. The estimation of μ_t is based on all $R_{t,i}, i = 1, \ldots, n_t$ and the sampling of $\Phi_{t,j}$ is conditional on data allocated to group j. The concentration parameter α_t is sampled using the method in Escobar and West (1994). The estimation contains the following steps and details can be found in Appendix 4.8.1.

1. Sample $\mu_t \big| R_{t,1:n_t}, \Phi_{t,1:K_t}, s_{t,1:n_t}$.
2. Sample $\Phi_{t,j} \big| R_{t,1:n_t}, s_{t,1:n_t}, \mu_t$ for $j = 1, \ldots, K_t$.
3. Sample $v_{t,j} \big| s_{t,1:n_t}$ for $j = 1, \ldots, K_t$.
4. Sample $u_{t,i} \big| w_{t,i}, s_{t,1:n_t}$ for $i = 1, \ldots, n_t$.
5. Sample $s_{t,i} \big| R_{1:n_t}, s_{t,1:n_t}, \mu_t, \Phi_{t,1:K_t}, u_{t,1:n_t}, K_t$ for $i = 1, \ldots, n_t$.
6. Sample $\alpha_t \big| K_t$.

Each MCMC iteration yields a set of $\{\mu_t, \Phi_{t,1:K_t}, s_{t,1:n_t}, \alpha_t\}$ draws. The model parameters can be estimated using sample averages of the MCMC outputs, after discarding the results in a burn-in period.

4.2.4 Bayesian Nonparametric Covariance Estimator

In the Bayesian nonparametric framework, the estimator of $\mathbf{V}_t$ is the posterior moment

$$E[\mathbf{V}_t \big| R_{t,1:n_t}] = E\left[\sum_{i=1}^{n_t} \Sigma_{t,i} \Big| R_{t,1:n_t}\right]. \tag{4.97}$$

Integrating out all parameter and distributional uncertainty and using M MCMC outputs, $E\left[\mathbf{V}_t | R_{t,1:n_t}\right]$ is estimated as

$$\widehat{\mathbf{V}}_t = \frac{1}{M}\sum_{m=1}^{M}\sum_{i=1}^{n_t} \Sigma_{t,i}^{(m)} = \frac{1}{M}\sum_{m=1}^{M}\sum_{i=1}^{n_t} \Phi^{(m)}_{t,s_{t,i}^{(m)}} \tag{4.98}$$

The posterior distributions of any functions of $\mathbf{V}_t$, such as realized beta or correlation, are readily available from MCMC outputs. For instance, from the sampled $\{(\sum_{i=1}^{n_t} \Sigma_{t,i})^{(m)}\}_{m=1}^{M}$, a (1-$\alpha$) probability density interval for the covariance between asset j and k is interval between the $\alpha/2\%$ and $(1-\alpha/2)\%$ quantiles of $\sum_{i=1}^{n_t} \Sigma_{t,i}^{(jk)}$. Note that the exact finite sample estimates can be obtained directly, while the classical estimator relies on asymptotic distribution to derive the confidence intervals.

4.3 Extensions

This section considers covariance estimation of noisy prices with nonsynchronous trading. After a brief discussion of the bias caused by microstructure noise and nonsynchronous trading, the Bayesian nonparametric covariance estimator designed for prices with independent microstructure noise and nonsynchronous trading is illustrated.

4.3.1 Microstructure Noise and Nonsynchronous Trading

Let $p^{(j)}_{t,\tau_l^j}$ denotes the latent intraday log-price of asset j, τ_l^j denotes the arrival time of the l^{th} observation of asset j. Contaminated with microstructure noise, the observed intraday price is

$$\dot{p}^{(j)}_{t,\tau_l^j} = p^{(j)}_{t,\tau_l^j} + \epsilon^{(j)}_{t,\tau_l^j}, \qquad \epsilon^{(j)}_{t,\tau_l^j} \sim \mathrm{N}(0, \omega_t^{(j)2}), \tag{4.99}$$

where $\epsilon^{(j)}_{t,\tau_l^j}$ is independent with $p^{(j)}_{t,\tau_l^j}$ and $\epsilon^{(k)}_{t,\tau_l^j}$ for $k \neq j$.

Synchronized using a previous-tick scheme with grid length h, the regularly spaced price is defined as

$$\widetilde{p}^{(j)}_{t,i} = \dot{p}^{(j)}_{t,\max(\tau^j|\tau^j \leq ih)}, \quad j = 1, \ldots, d. \tag{4.100}$$

The regularly spaced return vector of the d assets is denoted as $\widetilde{R}_{t,i} = \widetilde{P}_{t,i} - \widetilde{P}_{t,i-1}$, where $\widetilde{P}_{t,i} = \left(\widetilde{p}^{(1)}_{t,i}, \widetilde{p}^{(2)}_{t,i}, \cdots, \widetilde{p}^{(d)}_{t,i}\right)'$.

The presence of independent microstructure noise turns each asset's return series into an autocorrelated process. Moreover, the return series based on the previous-tick scheme have a lead-lag dependence. The summation of cross products of mismatched intraday returns underestimates the daily covariance between two assets as it covers only a proportion of the daily interval. In addition, the downward bias gets larger as

the length of the grid shrinks. This phenomenon is documented as the "Epps effect" (Epps, 1979). Following Kanatani and Renó (2007), the bias can be attributed to nonsynchronous bias, which is caused by the mismatched trading time of assets, and zero-return bias, which is caused by absence of a transaction in one or multiple grid(s). Due to microstructure noise and nonsynchronous trading, the realized covariance and the estimator proposed in Section 4.2 are no longer suitable estimators of the ex-post covariance.

Popular ways of mitigating the influences of microstructure noise and nonsynchronous trading is to add lead-lag adjustment or autocovariance adjustment to the realized covariance estimator. For example, the multivariate realized kernel ($\mathbf{RK}_t$) proposed by Barndorff-Nielsen *et al.* (2011) is defined as

$$\mathbf{RK}_t = \sum_{h=-H}^{H} \left(k\left(\frac{h}{H}\right) \sum_{i=h+1}^{n_t} \widetilde{R}_{t,i} \widetilde{R}'_{t,i-h} \right), \tag{4.101}$$

where their recommended kernel function $k(\cdot)$ is the Parzen kernel[1], the optimal bandwidth is $H = c_0 n_t^{3/5}$ and $\widetilde{R}_{t,i}$ is synchronized using a refresh time scheme[2].

4.3.2 DPM-VMA(1) Model

This chapter adapts the vector moving average with one lag to capture the autocorrelation and propose a Bayesian nonparametric covariance estimator with pooling for

[1]Parzen kernel function:

$$k(x) = \begin{cases} 1 - 6x^2 + 6x^3, & 0 \le x \le 1/2 \\ 2(1-x)^3, & 1/2 < x \le 1 \\ 0, & x > 1 \end{cases}$$

[2]In a refresh time scheme, the prices are sampled at the time that all assets haven been traded. The return series is irregularly spaced.

noisy and nonsynchronous prices.

Combining the vector moving average parametrization with the Bayesian nonparametric framework introduced in Section 4.2 yields the DPM-VMA(1) model.

$$\widetilde{R}_{t,i} = \mu_t + \Theta_t \eta_{t,i-1} + \eta_{t,i}, \eta_{t,i} \sim \mathrm{N}(0, \Delta_{t,i}),\ i = 1, \ldots, n_t, \tag{4.102}$$

$$\Delta_{t,i} \big| G_t \overset{iid}{\sim} G_t, \tag{4.103}$$

$$G_t \big| G_{0,t}, \alpha_t \sim \mathrm{DP}(\alpha_t, G_{0,t}), \tag{4.104}$$

$$G_{0,t} \equiv \mathrm{IW}(\Psi_t, \nu_t), \tag{4.105}$$

where $\eta_{t,i-1} = R_{t,i-1} - \mu_t - \Theta_t \eta_{t,i-2}$, Θ_t is a $d \times d$ matrix and $\eta_{t,i} = (\eta_{t,i}^{(1)}, \eta_{t,i}^{(2)}, \cdots, \eta_{t,i}^{(d)})'$ represents the error term.

The mean of $\widetilde{R}_{t,i}$ is not a constant vector but has a moving average structure. μ_t and Θ_t are constant for i but will change with the day t. The DPM-VMA(1) model implies $\mathrm{cov}(R_{t,i}, R_{t,i-1}) = \Theta_t \Delta_{t,i-1}$. Thus, in order to capture both the autocorrelation and the cross sectional dependence in returns, Θ_t has to be a $d \times d$ full matrix. The initial error vector $\eta_{t,0}$ is assumed to be a zero vector. Other model settings remain the same as in the DPM model except $\Delta_{t,i}$ is used to denote the intraday covariance. Note that the DPM-VMA(1) model allows error terms to be heteroskedastic, so it does not require the covariance of microstructure noise to be identical.

The priors for parameters μ_t, α_t and $\Delta_{t,i}$ are defined the same as in the DPM model. The prior of the elements of Θ_t is assumed to be $\Theta_t^{(jk)} \sim \mathrm{N}(m_\theta, v_\theta^2)$ for $j = 1, \ldots, d$ and $k = 1, \ldots, d$.

4.3.3 Model Estimation

As in the estimation of the DPM model, I apply the slice sampling technique to the stick-breaking representation of the DPM-VMA(1) model. Because of the moving average structure, the Gibbs sampler is not feasible to sample μ_t and Θ_t. Sampling high-dimensional parameters such as μ_t and Θ_t using Metropolis-Hasting results in very low mixing and finding good proposals is also challenging. I apply the Hamiltonian Monte Carlo method of Neal (2011) to sample μ_t and Θ as blocks. The Hamiltonian Monte Carlo adopts the Hamilton dynamics, rather than a probability distribution, to propose draws in Markov chain. Unlike the random walk proposal, the Hamilton dynamics produces distant proposals which explore the target distribution more efficiently, and has a high acceptance rate.

The Gibbs sampler handles the estimation of $\Phi_{t,j}$, which represents the unique values of $\Delta_{t,i}$. The remaining MCMC steps are essentially the same as in the DPM estimation. Each MCMC run contains the following steps. Appendix 4.8.2 provides the estimation procedure in detail.

1. Sample $\mu_t \big| \widetilde{R}_{t,1:n_t}, \Phi_{t,1:K_t}, \Theta_t, s_{t,1:n_t}$.
2. Sample $\Theta_t \big| \widetilde{R}_{t,1:n_t}, \mu_t, \Phi_{t,1:K_t}, s_{t,1:n_t}$.
3. Sample $\Phi_{t,j} \big| \widetilde{R}_{t,1:n_t}, \mu_t, \Theta_t, s_{t,1:n_t}$ for $j = 1, \ldots, K_t$.
4. Sample $v_{t,j} \big| s_{t,1:n_t}$ for $j = 1, \ldots, K_t$.
5. Sample $u_{t,i} \big| w_{t,i}, s_{t,1:n_t}$ for $i = 1, \ldots, n_t$.
6. Sample $s_{t,i} \big| \widetilde{R}_{1:n_t}, s_{t,1:n_t}, \mu_t, \Theta_t, \Phi_{t,1:K_t}, u_{t,1:n_t}, K_t$ for $i = 1, \ldots, n_t$.
7. Sample $\alpha_t \big| K_t$.

4.3.4 Bayesian Nonparametric Covariance Estimator with Bias Adjustment

Hansen *et al.* (2008) point out that the covariance estimator based on the moving average model requires an adjustment in order to obtain an unbiased estimator. Incorporating their adjustment, the ex-post covariance $\mathbf{V}_t$ is estimated as

$$E[\mathbf{V}_t \big| \widetilde{R}_{t,1:n_t}] = E\left[(\mathrm{I} + \Theta_t) \sum_{i=1}^{n_t} \Delta_{t,i} (\mathrm{I} + \Theta_t)' \Big| \widetilde{R}_{t,1:n_t} \right]. \tag{4.106}$$

which can be estimated as

$$\begin{aligned} \widehat{\mathbf{V}}_{\mathrm{MA},t} &= \frac{1}{M} \sum_{m=1}^{M} (\mathrm{I} + \Theta_t^{(m)}) \left(\sum_{i=1}^{n_t} \Delta_{t,i}^{(m)} \right) (\mathrm{I} + \Theta_t^{(m)})' \\ &= \frac{1}{M} \sum_{m=1}^{M} (\mathrm{I} + \Theta_t^{(m)}) \left(\sum_{i=1}^{n_t} \Phi_{t, s_{t,i}^{(m)}}^{(m)} \right) (\mathrm{I} + \Theta_t^{(m)})'. \end{aligned} \tag{4.107}$$

Appendix 4.8.3 proves that the Bayesian estimator provided in equation (4.107) correctly recovers the ex-post covariance in the presence of independent microstructure noise and nonsynchronous trading, assuming no zero-return bias.

The proposed estimator is analogous to the realized covariance with one lead and one lag adjustment[3] in that both are adjusted using autocovariance estimates. The

[3]The realized covariance estimator with one lags and one leads is defined as

$$\mathrm{RCLL}(1,1) = \sum_{i=1}^{n_t} \sum_{l=-1}^{1} R_{t,i+l}^{(1)} R_{t,i}^{(2)}.$$

estimator in equation (4.107) can be decomposed as follows

$$\begin{aligned}
&(\mathrm{I}+\Theta_t)\sum_{i=1}^{n_t}\Delta_{t,i}(\mathrm{I}+\Theta_t') \\
&=\sum_{i=1}^{n_t}(\Theta_t\Delta_{t,i-1}\Theta_t'+\Delta_{t,i}+\Theta_t\Delta_{t,i-1}+\Delta_{t,i}\Theta_t') \\
&=\sum_{i=1}^{n_t}\left[\mathrm{cov}(\widetilde{R}_{t,i})+\mathrm{cov}(\widetilde{R}_{t,i},\widetilde{R}_{t,i-1})+\mathrm{cov}(\widetilde{R}_{t,i},\widetilde{R}_{t,i+1})'\right],
\end{aligned} \tag{4.108}$$

where the last equation is derived using $\mathrm{cov}(\widetilde{R}_{t,i})=\Theta_t\Delta_{t,i-1}\Theta_t'+\Delta_{t,i}$ and $\mathrm{cov}(\widetilde{R}_{t,i},\widetilde{R}_{t,i-1})=\Theta_t\Delta_{t,i-1}$, implied from the vector moving average process in (4.102).

4.4 Synchronization with Data Augmentation

The process of placing observed prices on a grid is referred to as synchronization. Previous-tick method yields equally spaced data but it results in nonsynchronous bias and zero-return bias in covariance estimation using high-frequency data. The refresh time method proposed by Barndorff-Nielsen *et al.* (2011) produces irregularly spaced data and the number of data depends on the most illiquid asset.

This chapter proposes a synchronization method based on data augmentation to improve the previous-tick scheme. The missing observations on common grid points are treated as unknown variables and are estimated along with other model parameters under the Bayesian framework. Figure 4.18 provides one example of three assets to illustrate the mechanism. The solid dots denote the time of transactions and the dashed line represents the regularly spaced sampling time. In this example, there is no transaction in interval $(i+2,i+3]$ for both asset 1 and 2 and in interval $(i,i+1]$ for asset 3. In other words, $p_{t,i+3}^{(1)}$, $p_{t,i+3}^{(2)}$ and $p_{t,i+1}^{(3)}$ are missing and need to

be augmented.

Formally, if b out of d prices are missing for $P_{t,i}$, where $(1 \leq b \leq d)$, the sampling of missing price records are conditional on the $q = d - b$ observed prices, the adjacent prices $P_{t,i-1}$ and $P_{t,i+1}$, the mean vector μ_t and the covariance matrices $\Sigma_{t,s_{t,i}}$ and $\Sigma_{t,s_{t,i+1}}$. The return vectors $R_{t,i}$ and $R_{t,i+1}$ provide the linkage between the model and the missing observations. First splitting $P_{t,i}$, $R_{t,i}$, μ_t and $\Sigma_{t,i}$ into two groups, one corresponds to the b missing observations in $P_{t,i}$, the other matches to the q observed prices.

$$P_{t,i} = \begin{bmatrix} P^b_{t,i} \\ P^q_{t,i} \end{bmatrix}, \quad R_{t,i} = \begin{bmatrix} R^b_{t,i} \\ R^q_{t,i} \end{bmatrix}, \quad \mu_t = \begin{bmatrix} \mu^b_t \\ \mu^q_t \end{bmatrix}, \quad \Sigma_{t,i} = \begin{bmatrix} \Sigma^{bb}_{t,i} & \Sigma^{bq}_{t,i} \\ \Sigma^{bq'}_{t,i} & \Sigma^{qq}_{t,i} \end{bmatrix}.$$

The conditional distribution of $R^b_{t,i}$ given observed $R^q_{t,i}$ is

$$R^b_{t,i} \big| R^q_{t,i} = P^b_{t,i} - P^b_{t,i-1} \big| R^q_{t,i} \sim \mathrm{N}\left(\overline{\mu}_{t,i}, \overline{\Sigma}_{t,i}\right), \tag{4.109}$$

where $\overline{\mu}_{t,i}$ and $\overline{\Sigma}_{t,i}$ are the mean and covariance of distribution of $R^b_{t,i}$ conditional on $R^q_{t,i}$. For the DPM model, $\overline{\mu}_{t,i}$ and $\overline{\Sigma}_{t,i}$ are

$$\overline{\mu}_{t,i} = \mu^b_t + \Sigma^{bq}_{t,i}(\Sigma^{qq}_{t,i})^{-1}(R^q_{t,i} - \mu^q_t), \tag{4.110}$$

$$\overline{\Sigma}_{t,i} = \Sigma^{bb}_{t,i} - \Sigma^{bq}_{t,i}(\Sigma^{qq}_{t,i})^{-1}(\Sigma^{bq}_{t,i})'. \tag{4.111}$$

For the DPM-VMA(1) model, the conditional mean has a moving average dynamics and is derived as

$$\overline{\mu}_{t,i} = \mu^b_t + (\Theta_t \eta_{t,i-1})^b + \Sigma^{bq}_{t,i}(\Sigma^{qq}_{t,i})^{-1}(R^q_{t,i} - \mu^q_t - (\Theta_t \eta_{t,i-1})^q). \tag{4.112}$$

$\overline{\mu}_{t,i+1}$ and $\overline{\Sigma}_{t,i+1}$ can be derived similarly.

If the prices of d assets are all missing, then $\overline{\mu}_{t,i} = \mu_t$ for DPM model, $\overline{\mu}_{t,i} = \mu_t + \Theta_t \eta_{t,i-1}$ for DPM-VMA(1) model and $\overline{\Sigma}_{t,i} = \Sigma_{t,i}$.

The density of $P^b_{t,i}$ conditional on observed prices and model parameters is given as

$$\begin{aligned}\pi\left(P^b_{t,i}\middle|\cdots\right) \propto \exp\Big\{&-\tfrac{1}{2}\left[P^{b\prime}_{t,i}\overline{\Sigma}^{-1}_{t,i}P^b_{t,i} - 2P^{b\prime}_{t,i}\overline{\Sigma}^{-1}_{t,i}P^b_{t,i-1} + \overline{\mu}_{t,i}\right]\\ &-\tfrac{1}{2}\left[P^{b\prime}_{t,i}\overline{\Sigma}^{-1}_{t,i+1}P^b_{t,i} - 2P^{b\prime}_{t,i}\overline{\Sigma}^{-1}_{t,i+1}(P^b_{t,i+1} - \overline{\mu}_{t,i+1})\right]\Big\} \qquad (4.113)\\ \sim \mathrm{N}(M^b, V^b),&\end{aligned}$$

where

$$M^b = V^b\left[\overline{\Sigma}^{-1}_{t,i}(P^b_{t,i-1} + \overline{\mu}_{t,i}) + \overline{\Sigma}^{-1}_{t,i+1}(P^b_{t,i+1} - \overline{\mu}_{t,i+1})\right], \qquad (4.114)$$

$$V^b = \left(\overline{\Sigma}^{-1}_{t,i} + \overline{\Sigma}^{-1}_{t,i+1}\right)^{-1}. \qquad (4.115)$$

Intuitively, by using observed prices and knowing the model structure, we can infer the missing prices. The missing prices are estimated and regularly spaced returns are updated in each MCMC run, resulting in data at all grid points.

A numerical example is provided to illustrate the benefit of augmenting missing observations. The nonsynchronously spaced prices of 3 assets are simulated and the observations arrive every 60 seconds, 40 seconds and 30 seconds on average, respectively. The details of the data generating process will be illustrated in Section 4.5. I set the length of the grid to be 60 seconds and estimate the covariance matrix using the Bayesian nonparametric method with and without data augmentation. Figure 4.19 plots the 100 days' covariance estimates between asset 1 and 2, along with the true

covariance and $\mathbf{RC}_t$ based on 1-minute previous-tick returns. It shows that the $\widehat{\mathbf{V}}_{\mathrm{MA},t}$ without data augmentation and $\mathbf{RC}_t$ suffer from the "Epps effect" and the estimates are far from the truth. On the contrary, with data augmentation, the Bayesian nonparametric covariance estimator recovers the true values very well. The improvement brought by data augmentation comes from two aspects. For one thing, it fills the information gap and removes the bias caused by zero-returns. For another, the nonsynchronous bias is reduced as the missing price gap is filled exactly at each grid point, not prior to it. As a result, only adjacent returns are correlated and one lead and one lag or MA(1) adjustment are adequate to correct the bias. It implies that the proposed Bayesian nonparametric estimator $\widehat{\mathbf{V}}_{\mathrm{MA},t}$, together with the synchronization with data augmentation, fully accounts for bias caused by nonsynchronous trading.

4.4.1 Determination of Grid Length

As the previous-tick method, the proposed synchronization method requires setting a common-time grid. There is a trade-off between the length of the grid and the quality of the covariance estimator. The larger the grid length, the fewer missing prices need to be augmented. But that leads to more information loss as the sampling frequency is lower. Conditional on data, increasing grid frequency results in diminished value in data augmentation since more grid points contain missing observations and data augmentation is necessary at more points.

For the purpose of finding the optimal grid length for the Bayesian approach, I use simulated nonsynchronous data with independent microstructure noise to compare the accuracy of covariance estimation at different grid lengths. A case with ten assets is considered and the average intervals between two transactions vary from 5

seconds to 10 seconds. Figure 4.20 plots the histogram of root mean squared errors (RMSE)s of diagonal element estimates under different grid lengths. Figure 4.21 provides the RMSEs for ten covariance estimates. Let $\overline{D}_t$ denotes the average price change duration of the ten assets on day t. Increasing the length from $\overline{D}_t$ to $3\overline{D}_t$, the RMSEs of both diagonal and off-diagonal estimates of the covariance matrix display U-shape patterns and are smaller than those of $\mathbf{RK}_t$ in almost all the cases. Setting $1.5\overline{D}_t$ as the grid length[4] minimizes the error in estimating both diagonal and off-diagonal elements of the ex-post covariance matrix.

4.4.2 Positive Definiteness

The proposed Bayesian nonparametric covariance estimator is guaranteed to be positive definite, whereas classical approaches deliver positive semi-definite results. In the estimation of large covariance matrices, it is possible that the number of available observations is lower than the number of assets. For example, the number of data points synchronized using refresh time depends on the most inactive asset and can be driven to a number below the data dimension. The previous-tick approach can increase the number of observations by shrinking grid length, but it may result in many zero returns. In those cases, the covariance matrix calculated using traditional methods can be singular and not positive definite. The proposed Bayesian nonparametric approach can solve the difficulty. As the missing price gaps can all be filled using data augmentation, the grid length can be adjusted to ensure that the number of non-zero return vectors is above the data dimension. In addition, using an inverse Wishart distribution as the base distribution of intraday covariance ensures

[4]If the calculated length is not the divisor of 23400, then round it to the smallest number that makes 23400 divisible.

the positive definiteness of the covariance estimator.

4.5 Simulation Results

4.5.1 Data Generating Process

Following Barndorff-Nielsen *et al.* (2011), the fundamental log prices are generated from the following multivariate factor stochastic volatility model.

$$\mathrm{d}p^{(j)} = \mu^{(j)}\mathrm{d}t + \rho^{(j)}\sigma^{(j)}\mathrm{d}B^{(j)} + \sqrt{1-\rho^{(j)2}}\sigma^{(j)}\mathrm{d}W, \tag{4.116}$$

$$\sigma^{(j)} = \exp(\beta_0^{(j)} + \beta_1^{(j)}v^{(j)}), \tag{4.117}$$

$$\mathrm{d}v^{(j)} = \alpha^{(j)}v^{(j)}\mathrm{d}t + \mathrm{d}B^{(j)}, \tag{4.118}$$

where W and $B^{(j)}$ are standard Brownian motions, $\mathrm{cor}(\mathrm{d}W, \mathrm{d}B^{(j)}) = 0$ and the values of the parameters are $(\mu^{(j)}, \beta_0^{(j)}, \beta_1^{(j)}, \alpha^{(j)}, \rho^{(j)}) = (0.04, -0.3125, 0.125, -0.025, -0.3)$ for $j = 1, \ldots, d$.

Following Barndorff-Nielsen *et al.* (2011), the error terms are independent of each other and are added to the fundamental log price as follows

$$\tilde{p}_{t,l}^{(j)} = p_{t,l}^{(j)} + \epsilon_{t,l}^{(j)}, \quad \epsilon_{t,l}^{(j)} \sim \mathrm{N}(0, \sigma_\omega^{(j)2}), \quad l = 1, \cdots, N_t, \tag{4.119}$$

$$\sigma_\omega^{(j)2} = \xi^2\sqrt{\tfrac{1}{N_t}\textstyle\sum_{l=1}^{N_t}(\sigma_{t,l}^{(j)})^4}, \tag{4.120}$$

where ξ^2 stands for the noise-signal ratio and is set to be $\xi^2 = 0.001$, which is a value commonly used in the literature. Assuming the length of daily trading time is 6.5 hours ($N_t = 23400$), the log prices at every second are simulated. The regularly

spaced 5-minute, 1-minute and 10-second prices are generated by taking the records every 300, 60 and 10 steps, respectively.

In the simulation of nonsynchronous prices, the arrival times of observed prices are simulated from independent Poisson processes. Parameter λ in Poisson process governs the trading frequency of simulated data. For example, $(\lambda^{(1)}, \lambda^{(2)}) = (30, 10)$ means the transactions of asset 1 and 2 arrives every 30 seconds and 10 seconds on average, respectively.

The estimation target is the true daily ex-post covariance matrix $\Sigma_t = \sum_{l=1}^{N_t} \Sigma_{t,l}$, where $\Sigma_{t,l}^{(jk)} = \sqrt{1-\rho^{(j)2}}\sigma_{t,l}^{(j)}\sqrt{1-\rho^{(k)2}}\sigma_{t,l}^{(k)}$.

4.5.2 Precision of Estimation

The comparison starts by assessing the precision in estimating the ex-post covariance matrix. The benchmark estimators are realized covariance and multivariate realized kernel. The prior settings for the Bayesian nonparametric estimators are shown in Table 4.15.

Table 4.16 reports the averaged RMSEs of $\mathbf{RC}_t$, $\mathbf{RK}_t$ and $\widehat{\mathbf{V}}_t$ in estimating both diagonal and off-diagonal elements of covariance given regularly spaced 5-minute, 1-minute and 10-second returns. Both a three assets case and a ten assets case are considered. Panel A shows the results in the ideal case in which prices are free of measurement error. In that case, $\mathbf{RC}_t$ provides the "golden" standard as it is the consistent estimator of the integrated covariance. The Bayesian nonparametric estimator $\widehat{\mathbf{V}}_t$, however, offers additional improvement. Using $\widehat{\mathbf{V}}_t$ as the estimator, the RMSEs of both diagonal and off-diagonal elements are reduced no matter the data frequency or the number of dimension. The proposed covariance estimator works

especially well given low-frequency returns. For example, in the 10 assets case given 5-minute returns, switching from $\mathbf{RC}_t$ to $\widehat{\mathbf{V}}_t$ reduces the average RMSE of variance estimates from 0.1871 to 0.1567, and the average RMSE of off-diagonal elements from 0.1217 to 0.0992. The comparison suggests that the finite sample approach, together with taking advantage of pooling, improve the precision of estimation. The results in Panel B are based on prices with microstructure noise. As the frequency increases, $\mathbf{RC}_t$ is not an unbiased estimator for variance estimation as it does not account for the microstructure noise. Both $\mathbf{RK}_t$ and $\widehat{\mathbf{V}}_{\mathrm{MA},t}$ improve as the data frequency increases. In both the three assets and ten assets cases, $\widehat{\mathbf{V}}_{\mathrm{MA},t}$ provides lower error in estimating ex-post covariance than $\mathbf{RK}_t$ given 1-minute and 10-second returns.

Table 4.17 summarizes the RMSEs of the covariance estimation in a more realistic scenario with random arrival times. Both a low frequency case with $\lambda^{(j)} \in \{30, 40, 60\}$ and a higher frequency case with $\lambda^{(j)} \in \{5, 8, 10\}$ are considered. $\mathbf{RC}_t$ is formed using 5-minute return synchronized by previous tick. $\mathbf{RK}_t$ uses returns based on refresh time synchronization. The synchronization with data augmentation is used for $\widehat{\mathbf{V}}_{\mathrm{MA},t}$ and the grid length is set to be 1.5 times the average duration of price changes. The RMSEs of the three estimators in the no noise and independent microstructure noise cases are presented in Panel A and B, respectively. The 5-minute $\mathbf{RC}_t$ uses low frequency data to reduce the bias caused by noise and nonsynchronous trading. As a result, the estimation accuracy is not high since only 78 observations are used. $\mathbf{RK}_t$ increases the estimation accuracy but is not as accurate as the proposed Bayesian nonparametric estimator $\widehat{\mathbf{V}}_{\mathrm{MA},t}$. Among the 16 cases, $\widehat{\mathbf{V}}_{\mathrm{MA},t}$ always yields the smallest RMSE. For instance, in the high frequency case with 10 assets, the average RMSEs of diagonal and off-diagonal elements of $\widehat{\mathbf{V}}_{\mathrm{MA},t}$ are 0.0820 and 0.0541, while the values

are 0.1109 and 0.0739 if $\mathbf{RK}_t$ serves as the estimator. The improvement is greater than 20%.

4.5.3 Finite Sample Results

Existing ex-post covariance estimation approaches do not have finite sample distributional results and the distributions of estimators are approximated from asymptotic distributions. Using the asymptotic distributions derived in Barndorff-Nielsen and Shephard (2004) and Lunde *et al.* (2015), respectively, the 95% confidence interval of variance and covariance based on $\mathbf{RC}_t$[5] and $\mathbf{RK}_t$[6] can be derived. The Bayesian nonparametric approach offers the exact finite sample results, which does not require any approximation. The density intervals of $\widehat{\mathbf{V}}_t$ or $\widehat{\mathbf{V}}_{\mathrm{MA},t}$ can be obtained as the by-product of MCMC outputs.

Given simulated data, $\mathbf{RC}_t$, $\mathbf{RK}_t$ and $\widehat{\mathbf{V}}_{\mathrm{MA},t}$ are compared based on coverage frequencies for diagonal and off-diagonal elements of the ex-post covariance. Table 4.18 and Table 4.19 report the 95% coverage frequencies for regularly spaced return and

[5] For $\mathbf{RC}_t$, the 95% confidence interval of diagonal and off-diagonal elements are

$$\left[\mathbf{RC}_t^{jj} \pm z_{0.975}\sqrt{\tfrac{2}{3}\textstyle\sum_{i=1}^{n_t}(r_{t,i}^j)^4}\right], \tag{4.121}$$

$$\left[\mathbf{RC}_t^{jk} \pm z_{0.975}\sqrt{\textstyle\sum_{i=1}^{n_t}(r_{t,i}^j)^2(r_{t,i}^k)^2 - \sum_{i=1}^{n_t-1} r_{t,i}^j r_{t,i}^k r_{t,i+1}^j r_{t,i+1}^k}\right]. \tag{4.122}$$

[6] For $\mathbf{RK}_t$, the 95% confidence intervals of diagonal and off-diagonal elements are

$$\left[\mathbf{RK}_t^{jj} - \kappa_0(\mathbf{RK}_t^{jj})^{\frac{4}{5}}(\hat{\Omega}_t^{jj})^{\frac{1}{5}}(n_t)^{-\frac{1}{5}} \pm z_{0.975}\cdot 2\kappa_0(\mathbf{RK}_t^{jj})^{\frac{4}{5}}(\hat{\Omega}_t^{jj})^{\frac{1}{5}}(n_t)^{-\frac{1}{5}}\right], \tag{4.123}$$

$$\left[\mathbf{RK}_t^{jk} - \kappa_0\left(\frac{(\mathbf{RK}_t^{jj}\mathbf{RK}_t^{kk})^2+(\mathbf{RK}_t^{jk})^2}{2}\sqrt{\frac{\hat{\Omega}_t^{jk}}{n_t}}\right)^{\frac{2}{5}} \pm z_{0.975}\cdot 2\kappa_0\left(\frac{(\mathbf{RK}_t^{jj}\mathbf{RK}_t^{kk})^2+(\mathbf{RK}_t^{jk})^2}{2}\sqrt{\frac{\hat{\Omega}_t^{jk}}{n_t}}\right)^{\frac{2}{5}}\right], \tag{4.124}$$

where $\kappa_0 = 0.97$. Ω_t can be estimated using a HAC estimator with one lag: $\hat{\Omega}_t = \frac{1}{2n_t}\left(\sum_{i=1}^{n_t} R_{t,i}R_{t,i}' + \frac{1}{2}\sum_{i=1}^{n_t-1} R_{t,i}R_{t,i+1}'\right)$

nonsynchronous return cases, respectively. The no noise and microstructrue noise cases are both included. The Bayesian nonparametric estimator produces fairly good coverage probabilities no matter the data frequency and dimension, especially in the case with noisy and nonsynchronous data. As shown in Panel B of Table 4.19, the coverage probability of the 95% density interval of $\widehat{\mathbf{V}}_{\mathrm{MA},t}$ are closer to 95% than the 95% confidence interval of $\mathbf{RK}_t$ or the 5-minute $\mathbf{RC}_t$ in most of the cases.

4.6 Empirical Applications

This section provides applications of estimating ex-post covariances of equity prices using the Bayesian nonparametric approach. The tick prices and national best bid and offer (NBBO) prices of 10 equities (stock symbols: AA, BAC, CAT, F, GE, GIS, JNJ, T, WMT, XOM) listed on the NYSE and the Standard & Poor's Depository Receipt (SPY) from July 1, 2014 to June 30, 2016 are obtained from Tickdata. The cleaning procedure provided by Barndorff-Nielsen *et al.* (2009) is applied to the data. After eliminating half-tradings days (before Thanks-giving and Christmas), 499 trading days are left in the sample.

The 5-minute realized covariance, multivariate realized kernel and the Bayesian nonparametric covariance estimator based on the DPM-VMA model are applied to estimate the daily covariance matrix of returns of the ten assets. The synchronization methods for $\mathbf{RC}_t$, $\mathbf{RK}_t$ and $\widehat{\mathbf{V}}_{\mathrm{MA},t}$ are previous tick with 5-minute interval, refresh-time and the proposed synchronization with data augmentation. Because of the discreteness of transaction prices, real data contain more zero returns than simulated data. I set the grid length of proposed synchronization to be $3\overline{D}_t$, where $\overline{D}_t$ is the

average duration of price changes of d assets on day t. The prior setting are same as in Section 4.5 and are shown in Table 4.15.

4.6.1 Correlation and Realized Beta

Figure 4.22 to Figure 4.24 plot the correlation series between pairs AA-BAC, CAT-F and GE-GIS, implied by 5-minute $\mathbf{RC}_t$, $\mathbf{RK}_t$ and $\widehat{\mathbf{V}}_{\mathrm{MA},t}$ in the 10×10 case. Overall, the three versions of the correlation estimates share similar dynamics. The correlations based on $\mathbf{RK}_t$ and $\widehat{\mathbf{V}}_{\mathrm{MA},t}$ are close to each other and the correlation implied by 5-minute $\mathbf{RC}_t$ seems more volatile.

The proposed covariance estimator enables the calculation of a Bayesian nonparametric version of realized beta, which is defined as

$$\beta_t^{\widehat{\mathbf{V}}_{\mathrm{MA}}} = \frac{\widehat{\mathbf{V}}_{\mathrm{MA},t}^{jk}}{\widehat{\mathbf{V}}_{\mathrm{MA},t}^{jj}}. \tag{4.125}$$

Three versions of realized beta for the AA-SPY pair based on $\mathbf{RC}_t$, $\mathbf{RK}_t$ and $\widehat{\mathbf{V}}_{\mathrm{MA},t}$ in 2×2 case are calculated. Table 4.20 shows the fitted values of the ARMA(1,1) model for the three versions of realized beta. The estimation result confirms the strong persistence of realized beta. Comparison of parameter estimates implies $\beta_t^{\mathbf{RK}}$ and $\beta_t^{\widehat{\mathbf{V}}_{\mathrm{MA}}}$ share similar time series dynamics.

The traditional way of deriving the distribution of realized beta relies on the delta method applied to the asymptotic distribution of covariance inference. The Bayesian nonparametric approach does not require any approximation. The distribution of realized beta can be obtained using the MCMC outputs. Figure 4.25 plots the realized

beta of the AA-SPY combination provided by $\mathbf{RK}_t$ and $\widehat{\mathbf{V}}_{\mathrm{MA},t}$, along with the corresponding 95% confidence (density) intervals in March 2016. In most of the dates, the Bayesian density interval is shorter, compared with the classical counterpart.

4.6.2 Portfolio Allocation Evaluation

The Monte Carlo experiments illustrated in Section 4.5 show that the Bayesian nonparametric covariance estimator often outperforms the realized kernel from the statistical perspective. It is worth exploring whether the better statistical properties lead to economic gains.

Following Fleming *et al.* (2003), the evaluation is based on the performance of a minimum-variance portfolio formed using a rolling covariance estimator of 5-minute $\mathbf{RC}_t$, $\mathbf{RK}_t$ and $\widehat{\mathbf{V}}_{\mathrm{MA},t}$, respectively. The stock pool includes ten equities (AA, BAC, CAT, F, GE, GIS JNJ, T, WMT and XOM) and the evaluation period is from July 2, 2014 to June 30, 2016. Suppose an investor applies the volatility-timing strategy to adjust the portfolio weights each day by solving the following risk minimization problem given a desired portfolio return μ_0.

$$\text{Min } w_t'\Sigma_t w_t \quad \text{s. t.} \quad w_t'\mu = \mu_0 \quad \text{and} \quad w_t'\mathbf{1} = 1, \tag{4.126}$$

where w_t stands for portfolio weights on day t, Σ_t is the covariance matrix, μ is the daily return mean of the assets and μ_0 is the required return of the portfolio. The solution of the minimization problem is

$$w_t = \frac{\Sigma_t^{-1}\mu}{\mu'\Sigma_t^{-1}\mu}\mu_0. \tag{4.127}$$

Based on the ex-post covariance estimates, the next period covariance is predicted using an exponential smoother.

$$\widehat{\Sigma}_t = \exp(-\kappa)\widehat{\Sigma}_{t-1} + \kappa \exp(-\kappa)\widehat{S}_{t-1}, \tag{4.128}$$

where κ is the decay rate and $\widehat{S}_{t-1}$ is the ex-post covariance estimator, which can be $\mathbf{RC}_t$, $\mathbf{RK}_t$ or $\widehat{\mathbf{V}}_{\mathrm{MA},t}$. μ is predetermined and is set to be the average of the most recent 5 years' daily returns. The only difference between each portfolio is the estimates of $\widehat{\Sigma}_t$ used.

A utility-based approach is used to assess economic gains. The same utility function is adapted as Fleming *et al.* (2003).

$$U(r_t^p) = W_0 \left[(1 + r_t^f + r_t^p) - \frac{\gamma}{2(1+\gamma)}(1 + r_t^f + r_t^p)^2 \right], \tag{4.129}$$

where $r_t^p = w_t' R_t$ is the portfolio return on day t, r_t^f is the daily risk-free rate obtained from Federal Economic Data and γ stands for the risk aversion coefficient.

The performance of two competing strategies can be evaluated through calculating the performance fee Δ that an investor would pay to switch from one to another. Δ is a constant that satisfies the following equation.

$$\sum_{t=1}^{T} U(r_t^{p1}) = \sum_{t=1}^{T} U(r_t^{p2} - \Delta). \tag{4.130}$$

The daily return of the portfolio based on 5-minute $\mathbf{RC}_t$ is set to be r_t^{p1}, which serves as the benchmark. r_t^{p2} can be based on $\mathbf{RK}_t$ or $\widehat{\mathbf{V}}_{\mathrm{MA},t}$.

Table 4.21 shows the Sharpe ratio of the portfolios based on $\mathbf{RC}_t$, $\mathbf{RK}_t$ and $\widehat{\mathbf{V}}_{\mathrm{MA},t}$

under a different decay rate κ and μ_0. Portfolios using a rolling estimator based on $\widehat{\mathbf{V}}_{\mathrm{MA},t}$ always have the highest Sharpe ratios, followed by portfolios based on $\mathbf{RK}_t$ and 5-minute $\mathbf{RC}_t$. Table 4.21 also lists the annualized basis point fees that an investor with quadratic utility would like to pay to switch from the 5-minute $\mathbf{RC}_t$ based strategy to using $\mathbf{RK}_t$ or $\widehat{\mathbf{V}}_{\mathrm{MA},t}$. Given different decay rates and required portfolio returns, both a less risk-averse investor ($\gamma = 1$) and a more conservative investor ($\gamma = 10$) would be willing to pay a higher performance fee in order to choose portfolio based on $\widehat{\mathbf{V}}_{\mathrm{MA},t}$, instead of the $\mathbf{RK}_t$ based one. For example, in the case with $\kappa = 0.1$ and $\mu_0 = 0.1$, an investor with $\gamma = 10$ would like to pay almost 20 extra basis points in order to choose $\widehat{\mathbf{V}}_{\mathrm{MA},t}$, rather than the $\mathbf{RK}_t$ as the covariance estimator to form the portfolio.

Furthermore, the Bayesian approach provides a method to assess how the estimation uncertainty influences the optimal weights, and thereby the portfolio return. The classical method provides only a point estimate of the covariance, which estimates the center of the covariance distribution. However, the parameter realization can deviate from the distribution mean. As a result, the estimated weight may not equal the true optimal weight and the risk and return of the portfolio are influenced. Since the Bayesian model delivers an exact finite sample distribution of covariance, not only the mean estimates of covariance, portfolio weights and return, but also all the possible outcomes of those quantities can be obtained. Using the results from 5000 MCMC iterations, the histograms of portfolio returns and the weights of the ten assets on June 29, 2016 are shown in Figure 4.26 and Figure 4.27. For example, due to the uncertainty in covariance, the weight on BAC varies from -3% to 7% and the possible portfolio returns lie in the range of 0.65% and 1.1%.

4.7 Conclusion

This chapter proposes a Bayesian nonparametric method of estimating a covariance matrix for nonsynchronous prices contaminated with independent microstructrue noise. The proposed estimator provides at least four benefits in covariance estimation. First, pooling observations with similar covariance increases the precision of ex-post covariance estimation. Second, the Bayesian approach delivers exact finite sample results without relying on any infill asymptotic assumption. Third, the estimated covariance estimator is guaranteed to be positive definite. Last but not least, a new synchronization method with data augmentation is introduced to convert nonsynchronous observations to regularly spaced data series without zero-return problem.

Monte Carlo simulation confirms that the Bayesian nonparametric covariance estimator is very competitive with existing estimators given both regularly and nonsynchronously spaced data, with and without microstructure noise. Empirical application to equity returns shows that the correlation and realized beta implied by the Bayesian nonparametric covariance estimator have similar time series dynamics as the multivariate realized kernel. The minimum variance portfolio based on the proposed estimator outperforms the portfolio formed using realized covariance and multivariate realized kernel in terms of Sharpe ratios and the utility level.

4.8 Appendix

4.8.1 Estimation Steps of DPM Model

1. Sampling μ_t:

   Given prior: $\mu_t \sim \mathrm{N}(M_\mu, V_\mu)$, the conditional posterior of μ_t is

$$p(\mu_t \big| R_{t,1:n_t}, \Phi_{t,1:K_t}, s_{t,1:n_t}) \propto p(\mu_t) \prod_{i=1}^{n_t} p(R_{t,i} \big| \mu_t, \Phi_{t,s_{t,i}}) \sim \mathrm{N}(\overline{M}_\mu, \overline{V}_\mu), \quad (4.131)$$

   where $\overline{V}_\mu = (\sum_{i=1}^{n_t} \Phi_{t,s_{t,i}}^{-1} + V_\mu^{-1})^{-1}$ and $\overline{M}_\mu = \overline{V}_\mu \left(\sum_{i=1}^{n_t} \Phi_{t,s_{t,i}}^{-1} R_{t,i} + V_\mu^{-1} M_\mu \right)$.

2. Sampling $\Phi_{t,j}$ for $j = 1, \ldots, K_t$:

   Given prior $\Phi_{t,j} \sim \mathrm{IW}(\Psi_t, \nu_t)$, the conditional posterior of $\Phi_{t,j}$ is

$$\begin{aligned} & p(\Phi_{t,j} \big| R_{t,1:n_t}, s_{t,1:n_t}, \mu_t) \propto p(\Phi_{t,j}) \prod_{s_{t,i}=j} p(R_{t,i} \big| \mu_t, \Phi_{t,j}) \\ & \propto \left| \Phi_{t,j} \right|^{-\frac{1}{2}(n_j+\nu+d+1)} \exp\left[-\frac{1}{2} \mathrm{tr}(\Psi_t + Q_{t,j}) \Phi_{t,j}^{-1} \right] \sim \mathrm{IW}(\Psi_t + Q_{t,j}, n_j + \nu_t), \end{aligned} \quad (4.132)$$

   where $Q_{t,j} = \sum_{s_{t,i}=j} (R_{t,i} - \mu_t)(R_{t,i} - \mu_t)'$ and $n_j = \sum_{i=1}^{n_t} \mathbb{1}(s_{t,i} = j)$.

3. Sampling $s_{t,i}$ for $i = 1, \cdots, n_t$:

$$P(s_{t,i} = j \big| R_{t,i}, \mu_t, \Phi_{t,1:K_t}, w_{t,1:K_t}, u_{t,i}) \propto \sum_{j=1}^{K_t} \mathbb{1}(w_{t,j} > u_{t,i}) N\left(R_{t,i} \big| \mu_t, \Phi_{t,j} \right). \quad (4.133)$$

4. Sampling $v_{t,j}$ and calculate $w_{t,j}$ for $j = 1, \cdots, K_t$:

$$p(v_{t,j}\big|s_{t,1:n_t}, \alpha_t) \sim \text{Beta}\left(1 + \sum_{i=1}^{n_t} \mathbb{1}(s_{t,i} = j), \alpha_t + \sum_{i=1}^{n_t} \mathbb{1}(s_{t,i} > j)\right). \quad (4.134)$$

$w_{t,j}$ are computed as $w_{t,1} = v_{t,1}$, and $w_{t,j} = v_{t,j} \prod_{l=1}^{j-1}(1 - v_{t,l})$.

5. Sampling $u_{t,i}$ for $i = 1, \ldots, n_t$:

$$p(u_{t,i}\big|s_{t,1:n_t}, w_{t,1:K_t}) \sim \text{Unif}(0, w_{t,s_{t,i}}). \quad (4.135)$$

6. Find the smallest K_t such that $\sum_{j=1}^{K_t} w_{t,j} > 1 - \min(u_{t,1:n_t})$.

7. Sampling α_t:

   Given prior $\alpha_t \sim \text{Ga}(a, b)$,

$$p(\alpha_t\big|K_t) \sim q \cdot \text{Ga}(a + K_t, b - \log \xi_t) + (1 - q) \cdot \text{Ga}(a + K_t - 1, b - \log \xi_t), \quad (4.136)$$

   where $q = \frac{a+K_t-1}{a+K_t-1+n_t(b-\log \xi_t)}$ and $\xi_t \sim \text{Beta}(\alpha_t + 1, n_t)$.

4.8.2 Estimation Steps of DPM-VMA(1) Model

1. Sampling μ_t:

Given prior: $\mu_t \sim \mathrm{N}(M_\mu, V_\mu)$. The posterior of μ_t is

$$p\left(\mu_t \middle| \widetilde{R}_{t,1:n_t}, \Theta_t, \Phi_{t,1:K_t}\right) \propto p(\mu_t) \prod_{i=1}^{n_t} p(\widetilde{R}_{t,i} | \mu_t + \Theta_t \eta_{t,i-1}, \Phi_{t,s_{t,i}})$$
$$\propto \exp\left\{ -\frac{1}{2}\mu_t' \left(\sum_{i=1}^{n_t} \Phi_{t,s_{t,i}}^{-1} + V_\mu^{-1} \right) \mu_t - \mu_t' \left(\sum_{i=1}^{n_t} \Phi_{t,s_{t,i}}^{-1} (\widetilde{R}_{t,i} - \Theta_t \eta_{t,i-1}) + V_\mu^{-1} M_\mu \right) \right\}. \tag{4.137}$$

where $\eta_{t,i-1} = \widetilde{R}_{t,i-1} - \mu_t - \Theta_t \eta_{t,i-2}$.

Define $U_1(\mu_t) = -\log\left[p(\mu_t | \widetilde{R}_{t,1:n_t}, \Theta_t, \Phi_{t,1:K_t}) \right]$, we have

$$U_1(\mu_t) = \frac{1}{2}\mu_t' \left(\sum_{i=1}^{n_t} \Phi_{t,s_{t,i}}^{-1} + V_\mu^{-1} \right) \mu_t - \mu_t' \left(\sum_{i=1}^{n_t} \Phi_{t,s_{t,i}}^{-1} (\widetilde{R}_{t,i} - \Theta_t \eta_{t,i-1}) + V_\mu^{-1} M_\mu \right) \tag{4.138}$$

$$\frac{\partial U_1(\mu_t)}{\partial \mu_t} = \left(\sum_{i=1}^{n_t} \Phi_{t,s_{t,i}}^{-1} + V_\mu^{-1} \right) \mu_t - \left(\sum_{i=1}^{n_t} \Phi_{t,s_{t,i}}^{-1} (\widetilde{R}_{t,i} - \Theta_t \eta_{t,i-1}) + V_\mu^{-1} M_\mu \right). \tag{4.139}$$

Introduce an auxiliary d-dimensional vector P_1 and define function $K_1(P_1) = \sum_{j=1}^{d} \frac{(P_1^{(j)})^2}{2}$. The leapfrog method is used to approximate the Hamiltonian dynamics.

(1) Set $\mu_{t,0} = \mu_t^{(m-1)}$, where $\mu_t^{(m-1)}$ is the value of μ_t in previous iteration. Initialize P_1 as $P_1^{(j)} \sim \mathrm{N}(0,1)$.

(2) $P_1' = P_1 - \frac{\epsilon}{2} \frac{\partial U_1(\mu_{t,0})}{\partial \mu_t}$ and $\mu_t' = \mu_{t,0}$.

(3) For l from 1 to L,

   a. $\mu_t' = \mu_t' + \epsilon P_1'$.

   b. If $l < L$, $P_1' = P_1' - \epsilon \frac{\partial U_1(\mu_t')}{\partial \mu_t}$.

c. If $l = L$, $P_1' = P_1' - \frac{\epsilon}{2}\frac{\partial U_1(\mu_t')}{\partial \mu_t}$.

Update $\mu = \mu_t'$ with acceptance rate $\min\left(1, \exp(U_1(\mu_{t,0}) + K_1(P_1) - U_1(\mu_t') - K_1(P_1'))\right)$.

L is the leapfrog step and ϵ is the the stepsize. I set $L = 20$ and adjust ϵ very 10 MCMC iterations. If the average acceptance rate in previous 10 runs is zero, set $\epsilon = 0.9\epsilon$. If the average acceptance rate is above 0.8, adjust $\epsilon = 1.1\epsilon$.

2. Sampling Θ_t:

Given prior $\Theta_{t,jk} \sim \mathrm{N}(m_{jk}, v_{jk}^2)$, the conditional posterior of Θ_t is

$$p\left(\Theta_t \middle| \widetilde{R}_{t,1:n_t}, \Theta_t, \Phi_{t,1:K_t}\right) \propto \exp\left\{-\frac{1}{2}\sum_{i=1}^{n_t}\left[\eta_{t,i-1}'\Theta_t'\Phi_{t,s_{t,i}}^{-1}\left(\Theta_t\eta_{t,i-1} - 2(\widetilde{R}_{t,i} - \mu_t)\right)\right]\right\} \cdot \prod_{j=1}^{d}\prod_{k=1}^{d}\exp\left[-\frac{(\Theta_t^{(jk)} - m_{jk})^2}{2v_{jk}^2}\right] \tag{4.140}$$

Define $U_2(\Theta_t) = -\log\left[p(\Theta_t \middle| \widetilde{R}_{t,1:n_t}, \Theta_t, \Phi_{t,1:K_t})\right]$, we have

$$U_2(\Theta_t) = -\frac{1}{2}\sum_{i=1}^{n_t}\left[\eta_{t,i-1}'\Theta_t'\Phi_{t,s_{t,i}}^{-1}\left(2(\widetilde{R}_{t,i} - \mu_t) - \Theta_t\eta_{t,i-1}\right)\right] + \sum_{j=1}^{d}\sum_{k=1}^{d}\left[\frac{(\Theta_t^{(jk)} - m_{jk})^2}{2v_{jk}^2}\right] \tag{4.141}$$

$$\frac{\partial U_2(\Theta_t)}{\partial \Theta_t} = -\sum_{i=1}^{n_t}\Phi_{t,s_{t,i}}^{-1}\left(\widetilde{R}_{t,i} - \mu_t - \Theta_t\eta_{t,i-1}\right)\eta_{t,i-1}' + \frac{(\Theta_t^{(jk)} - m_{jk})}{v_{jk}^2}\mathbb{1}_{jk} \tag{4.142}$$

Introduce an auxiliary $d \times d$ matrix P_2 and define function $K_2(P_2) = \sum_{j=1}^{d}\sum_{k=1}^{d}\frac{(P_2^{(jk)})^2}{2}$.

(1) Set $\Theta_{t,0} = \Theta_t^{(m-1)}$, where $\Theta_t^{(m-1)}$ is the result from previous iteration. Initialize P_2 that each element is drawn from $\mathrm{N}(0,1)$.

(2) $P_2' = P_2 - \frac{\epsilon}{2}\frac{\partial U_2(\Theta_{t,0})}{\partial \Theta_t}$ and $\Theta_t' = \Theta_{t,0}$.

(3) For l from 1 to L,

a. $\Theta_t' = \Theta_t' + \epsilon P_2'$.

b. If $l < L$, $P_2' = P_2' - \epsilon \frac{\partial U_2(\Theta_t')}{\partial \Theta_t}$.

c. If $l = L$, $P_2' = P_2' - \frac{\epsilon}{2} \frac{\partial U_2(\Theta_t')}{\partial \Theta_t}$.

Accept Θ_t' with acceptance rate $\min\left(1, \exp(U_2(\Theta_{t,0}) + K_2(P_2) - U_2(\Theta_t') - K_2(P_2'))\right)$.

3. Sampling $\Phi_{t,j}$ for $j = 1, \ldots, K_t$:

Given prior $\Phi_{t,j} \sim \text{IW}(\Psi_t, \nu_t)$, the conditional posterior of $\Phi_{t,j}$ is

$$\begin{aligned} p(\Phi_{t,j} \big| \widetilde{R}_{t,1:n_t}, s_{t,1:n_t}, \mu_t, \Theta_t) &\propto p(\Phi_{t,j}) \prod_{s_{t,i}=j} p(\widetilde{R}_{t,i} \big| \mu_t + \Theta_t \eta_{t,i-1}, \Phi_{t,j}) \\ &\propto \left|\Phi_{t,j}\right|^{-\frac{1}{2}(n_j+\nu+d+1)} \exp\left[-\frac{1}{2}\text{tr}(\Psi_t + Q_{t,j})\Phi_{t,j}^{-1}\right] \sim \text{IW}(\Psi_t + Q_{t,j}, n_j + \nu_t), \end{aligned} \tag{4.143}$$

where $Q_{t,j} = \sum_{s_{t,i}=j} (\widetilde{R}_{t,i} - \mu_t - \Theta_t \eta_{t,i-1})(\widetilde{R}_{t,i} - \mu_t - \Theta_t \eta_{t,i-1})'$ and $n_j = \sum_{i=1}^{n_t} \mathbb{1}(s_{t,i} = j)$.

4. Sampling $s_{t,i}$ for $i = 1, \ldots, n_t$:

$$P(s_{t,i} = j \big| R_{t,i}, \mu_t, \Phi_{t,1:K_t}, w_{t,1:K_t}, u_{t,i}) \propto \sum_{j=1}^{K_t} \mathbb{1}(w_{t,j} > u_{t,i}) \text{N}\left(\widetilde{R}_{t,i} \big| \mu_t + \Theta_t \eta_{t,i-1}, \Phi_{t,j}\right). \tag{4.144}$$

5. to 8. Same as step 4 to 7 in the estimation of DPM.

4.8.3 Proof of the Unbiasedness of DPM-VMA(1) Estimator

Let $p_{t,\tau_i}^{(j)}$ denotes the latent intraday log-price of asset j, τ_i denotes the arrival time. Contaminated with microstructure noise, the observed intraday price $\dot{p}_{t,\tau_i}^{(j)}$ is

$$\dot{p}_{t,\tau^j}^{(j)} = p_{t,\tau^j}^{(j)} + \epsilon_{t,\tau^j}^{(j)}, \qquad \epsilon_{t,\tau^j}^{(j)} \sim \mathrm{N}(0, \omega_{t,\tau^j}^{(j)2}).$$

where $\epsilon_{t,\tau^j}^{(j)}$ is independent with $p_{t,\tau^j}^{(j)}$ and $\epsilon_{t,\tau^j}^{(k)}$ for $k \neq j$.

Synchronized using previous-tick scheme with grid length h, the price series is

$$\widetilde{p}_{t,i}^{(j)} = \dot{p}_{t,\max(\tau^j|\tau^j \leq ih)}^{(j)}, \quad j = 1, \ldots, d.$$

The regularly spaced return vector of the j assets is denoted as $\widetilde{R}_{t,i} = \widetilde{P}_{t,i} - \widetilde{P}_{t,i-1}$, where $\widetilde{P}_{t,i} = \left(p_{t,i}^{(1)}, p_{t,i}^{(2)}, \cdots, p_{t,i}^{(d)}\right)'$.

Assuming there is no zero-return bias, the variance and first autocovariance of $\widetilde{R}_{t,i}$ are

$$\mathrm{cov}(\widetilde{R}_{t,i}) = \Sigma_{t,i} + \Omega_{t,i-1} + \Omega_{t,i} - \Gamma_{t,i-1} - \Gamma_{t,i} = \Xi_{t,i-1} + \Xi_{t,i}, \tag{4.145}$$

$$\mathrm{cov}(\widetilde{R}_{t,i}, \widetilde{R}_{t,i-1}) = -\Omega_{t,i-1} + \Gamma_{t,i-1} = -\Xi_{t,i-1}. \tag{4.146}$$

where $\Omega_{t,i} = \mathrm{diag}\left((\omega_{t,i}^{(1)})^2, (\omega_{t,i}^{(2)})^2, \cdots, (\omega_{t,i}^{(d)})^2\right)$, $\Gamma_{t,i}$ is matrix with zero diagonals and measures the lead-lag dependence between every two assets and $\Xi_{t,i} = \Omega_{t,i} - \Gamma_{t,i}$.

Consider the following heteroskedastic VMA(1) model for $\tilde{R}_{t,i}$,

$$\widetilde{R}_{t,i} = \mu_t + \Theta_t \eta_{t,i-1} + \eta_{t,i}, \qquad \eta_{t,i} \sim N(0, \Delta_{t,i}), \tag{4.147}$$

which will be used to recover an estimate of ex-post variance for the fundamental return process, $\mathbf{V}_t = \sum_{i=1}^{n_t} \Sigma_{t,i}$.

The corresponding moments of this process are

$$\operatorname{cov}(\widetilde{R}_{t,i}) = \Theta_t \Delta_{t,i-1} \Theta_t' + \Delta_{t,i}, \tag{4.148}$$

$$\operatorname{cov}(\widetilde{R}_{t,i}, \widetilde{R}_{t,i-1}) = \Theta_t \Delta_{t,i-1}. \tag{4.149}$$

Equating (4.146) and (4.149), we have

$$-\Xi_{t,i-1} = \Theta_t \Delta_{t,i-1} \qquad \text{and} \qquad -\Xi_{t,i} = \Theta_t \Delta_{t,i}. \tag{4.150}$$

Equating (4.145) and (4.148) and using the result in (4.150), we have

$$\begin{aligned} \Sigma_{t,i} + \Xi_{t,i} + \Xi_{t,i-1} &= \Theta_t \Delta_{t,i-1} \Theta_t' + \Delta_{t,i} \\ \Sigma_{t,i} + \Xi_{t,i} + \Xi_{t,i-1} &= -\Xi_{t,i-1} \Theta_t' - \Theta_t^{-1} \Xi_{t,i} \\ \Sigma_{t,i} &= -(\mathrm{I} + \Theta_t') \Xi_{t,i-1} - (\mathrm{I} + \Theta_t^{-1}) \Xi_{t,i} \end{aligned} \tag{4.151}$$

Using the results in (4.151) and (4.150), the summation of $\Sigma_{t,i}$ and $\Delta_{t,i}$, over $i = 1, \ldots, n_t$, are

$$\sum_{i=1}^{n_t} \Sigma_{t,i} = -(\mathrm{I} + \Theta_t') \sum_{i=1}^{n_t} \Xi_{t,i-1} - (\mathrm{I} + \Theta_t^{-1}) \sum_{i=1}^{n_t} \Xi_{t,i}, \tag{4.152}$$

$$\sum_{i=1}^{n_t} \Delta_{t,i} = -\Theta_t^{-1} \sum_{i=1}^{n_t} \Xi_{t,i}. \tag{4.153}$$

The ratio between (4.152) and (4.153) is

$$\begin{aligned}\sum_{i=1}^{n_t}\Sigma_{t,i}(\sum_{i=1}^{n_t}\Delta_{t,i})^{-1} &= (\mathrm{I}+\Theta_t)\sum_{i=1}^{n_t}\Xi_{t,i-1}(\sum_{i=1}^{n_t}\Xi_{t,i})^{-1}\Theta_t + (\mathrm{I}+\Theta_t^{-1})\sum_{i=1}^{n_t}\Xi_{t,i}(\sum_{i=1}^{n_t}\Xi_{t,i})^{-1}\Theta_t \\ &= (\mathrm{I}+\Theta_t)\Theta_t + (\mathrm{I}+\Theta_t^{-1})\Theta_t \\ &= (\mathrm{I}+\Theta_t)(\mathrm{I}+\Theta_t)'\end{aligned} \tag{4.154}$$

Finally, we have

$$\mathbf{V}_t = \sum_{i=1}^{n_t}\Sigma_{t,i} = (\mathrm{I}+\Theta_t)\sum_{i=1}^{n_t}\Delta_{t,i}(\mathrm{I}+\Theta_t), \qquad \text{if} \quad \Xi_{t,n_t} = \Xi_{t,0}. \tag{4.155}$$

Bibliography

Aït-Sahalia, Y., Fan, J., and Xiu, D. (2010). High-frequency covariance estimates with noisy and asynchronous financial data. *Journal of the American Statistical Association*, **105**(492), 1504–1517.

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

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

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (2009). Realized kernels in practice: trades and quotes. *Econometrics Journal*, **12**(3), C1–C32.

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.

Epps, T. W. (1979). Comovements in stock prices in the very short run. *Journal of the American Statistical Association*, **74**(366), 291–298.

Escobar, M. D. and West, M. (1994). Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association*, **90**, 577–588.

Fleming, J., Kirby, C., and Ostdiek, B. (2003). The economic value of volatility timing using "realized" volatility. *Journal of Financial Economics*, **67**(3), 473–509.

Griffin, J. E. and Oomen, R. C. (2011). Covariance measurement in the presence of non-synchronous trading and market microstructure noise. *Journal of Econometrics*, **160**(1), 58–68.

Hansen, P., Large, J., and Lunde, A. (2008). Moving average-based estimators of integrated variance. *Econometric Reviews*, **27**(1-3), 79–111.

Hansen, P. R. and Lunde, A. (2006). Realized variance and market microstructure noise. *Journal of Business & Economic Statistics*, **24**, 127–161.

Hayashi, T. and Yoshida, N. (2005). On covariance estimation of non-synchronously observed diffusion processes. *Bernoulli*, **11**(2), 359–379.

Kalli, M., Griffin, J. E., and Walker, S. G. (2011). Slice sampling mixture models. *Statistics and Computing*, **21**(1), 93–105.

Kanatani, T. and Renó, R. (2007). Unbiased covariance estimation with interpolated data. Department of economics university of siena, Department of Economics, University of Siena.

Lunde, A., Sheppard, K., and Shephard, N. (2015). Econometric analysis of vast covariance matrices using composite realized kernels and their application to portfolio choice. *Journal of Business and Economic Statistics*, **0**(ja), 1–46.

Neal, R. (2011). Mcmc using hamiltonian dynamics. *Handbook of Markov Chain Monte Carlo*, pages 113–162.

Peluso, S., Corsi, F., and Mira, A. (2014). A Bayesian high-frequency estimator of the multivariate covariance of noisy and asynchronous returns. *Journal of Financial Econometrics*.

Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica*, pages 639–650.

Voev, V. and Lunde, A. (2007). Integrated covariance estimation using high-frequency data in the presence of noise. *Journal of Financial Econometrics*, **5**(1), 68–104.

Zhang, L. (2011). Estimating covariation: Epps effect, microstructure noise. *Journal of Econometrics*, **160**(1), 33–47.

Table 4.15: Prior Specifications of Models

Model	μ_t	$\Sigma_{t,i}$	Θ_t^{jk}	α_t
DPM	$\mathrm{N}(0, V)$	$\mathrm{IW}(\Psi_t, \nu_t)$	-	$\mathrm{Gamma}(2, 4)$
DPM-VMA(1)	$\mathrm{N}(0, V)$	$\mathrm{IW}(\Psi_t, \nu_t)$	$\mathrm{N}(0, 0.5)$	$\mathrm{Gamma}(2, 4)$

$\nu_t = \frac{1}{d}\sum_{j=1}^{d} \frac{2(RC_t^{(jj)})^2}{\mathrm{var}(r_{t,i}^{(j)2})} + d + 3$ and $\Psi_t = \frac{\nu_t - d - 1}{n_t}\mathbf{RC}_t$. V is diagonal matrix with $V^{(jj)} = 0.001^2$.

Table 4.16: RMSEs of $\mathbf{RC}_t$, $\mathbf{RK}_t$, $\widehat{\mathbf{V}}_t$ and $\widehat{\mathbf{V}}_{\mathrm{MA},t}$ Given Regularly Spaced Returns

Panel A: No Noise Case ($\xi^2 = 0.0$)

		$\overline{\mathrm{RMSE}}(\Sigma_t^{jj})$			$\overline{\mathrm{RMSE}}(\Sigma_t^{jk})$		
Dimension	Frequency	$\mathbf{RC}_t$	$\mathbf{RK}_t$	$\widehat{\mathbf{V}}_t$	$\mathbf{RC}_t^{5m}$	$\mathbf{RK}_t$	$\widehat{\mathbf{V}}_t$
3 Assets	5-minute	0.1434	0.2648	**0.1366**	0.0859	0.1561	**0.0818**
	1-minute	0.0594	0.1343	**0.0591**	0.0379	0.0831	**0.0376**
	10-second	0.0233	0.0623	**0.0233**	0.0140	0.0394	**0.0140**
10 Assets	5-minute	0.1871	0.3686	**0.1567**	0.1217	0.2324	**0.0992**
	1-minute	0.0789	0.1928	**0.0783**	0.0506	0.1227	**0.0502**
	10-second	0.0329	0.0837	**0.0328**	0.0205	0.0528	**0.0204**

Panel B: Microstructure Noise Case ($\xi^2 = 0.001$)

		$\overline{\mathrm{RMSE}}(\Sigma_t^{jj})$			$\overline{\mathrm{RMSE}}(\Sigma_t^{jk})$		
Dimension	Frequency	$\mathbf{RC}_t$	$\mathbf{RK}_t$	$\widehat{\mathbf{V}}_{\mathrm{MA},t}$	$\mathbf{RC}_t$	$\mathbf{RK}_t$	$\widehat{\mathbf{V}}_{\mathrm{MA},t}$
3 Assets	5-minute	**0.2231**	0.3228	0.3641	**0.1062**	0.1890	0.1936
	1-minute	0.7441	0.1921	**0.1541**	**0.0689**	0.1056	0.0887
	10-second	4.4798	0.1248	**0.0774**	0.0845	0.0674	**0.0436**
10 Assets	5-minute	**0.2609**	0.3245	0.4094	**0.1041**	0.1742	0.2131
	1-minute	0.9197	0.1965	**0.1951**	**0.0666**	0.1021	0.0895
	10-second	5.6987	0.1384	**0.1172**	0.1008	0.0723	**0.0686**

Results are based on 1000 days' simulation. $\overline{\mathrm{RMSE}}(\Sigma_t^{jj})$ stands for the average of RMSEs for diagonal elements of Σ_t. $\overline{\mathrm{RMSE}}(\Sigma_t^{jk})$ is the average of RMSEs for off-diagonal elements. Bayesian nonparametric estimator $\widehat{\mathbf{V}}_t$ and $\widehat{\mathbf{V}}_{\mathrm{MA},t}$ are estimated based on on 5000 MCMC runs, after 1000 burn-in.

Table 4.17: RMSEs of $\mathbf{RC}_t$, $\mathbf{RK}_t$ and $\widehat{\mathbf{V}}_{\mathrm{MA},t}$ Given Nonsynchronous Returns

Panel A: No Noise Case ($\xi^2 = 0.0$)

		$\overline{\mathrm{RMSE}}(\Sigma_t^{jj})$			$\overline{\mathrm{RMSE}}(\Sigma_t^{jk})$		
Dimension	Frequency	$\mathbf{RC}_t^{5m}$	$\mathbf{RK}_t$	$\widehat{\mathbf{V}}_{\mathrm{MA},t}$	$\mathbf{RC}_t^{5m}$	$\mathbf{RK}_t$	$\widehat{\mathbf{V}}_{\mathrm{MA},t}$
3 Assets	$\lambda^j \in \{30, 40, 60\}$	0.1262	0.1355	**0.1057**	0.1112	0.0914	**0.0571**
	$\lambda^j \in \{5, 8, 10\}$	0.4233	0.1471	**0.0995**	0.1511	0.0598	**0.0337**
10 Assets	$\lambda^j \in \{30, 40, 60\}$	0.2613	0.3493	**0.2178**	0.1713	0.1712	**0.0913**
	$\lambda^j \in \{5, 8, 10\}$	0.5685	0.2335	**0.1458**	0.1553	0.0973	**0.0473**

Panel B: Microstructure Noise Case ($\xi^2 = 0.001$)

		$\overline{\mathrm{RMSE}}(\Sigma_t^{jj})$			$\overline{\mathrm{RMSE}}(\Sigma_t^{jk})$		
Dimension	Frequency	$\mathbf{RC}_t^{5m}$	$\mathbf{RK}_t$	$\widehat{\mathbf{V}}_{\mathrm{MA},t}$	$\mathbf{RC}_t^{5m}$	$\mathbf{RK}_t$	$\widehat{\mathbf{V}}_{\mathrm{MA},t}$
3 Assets	$\lambda^j \in \{30, 40, 60\}$	0.3408	0.3474	**0.2581**	0.1955	0.1964	**0.1223**
	$\lambda^j \in \{5, 8, 10\}$	0.2384	0.0761	**0.0563**	0.1057	0.0447	**0.0363**
10 Assets	$\lambda^j \in \{30, 40, 60\}$	0.2715	0.2668	**0.2353**	0.1544	0.1553	**0.0969**
	$\lambda^j \in \{5, 8, 10\}$	0.2418	0.1109	**0.0820**	0.1049	0.0739	**0.0541**

Results are based on 1000 days' simulation. $\overline{\mathrm{RMSE}}(\Sigma_t^{jj})$ stands for the average of RMSEs for diagonal elements of Σ_t. $\overline{\mathrm{RMSE}}(\Sigma_t^{jk})$ is the average of RMSEs for off-diagonal elements. Bayesian nonparametric estimator $\widehat{\mathbf{V}}_{\mathrm{MA},t}$ is estimated based on on 5000 MCMC runs, after 1000 burn-in.

Table 4.18: Coverage Probabilities of $\mathbf{RC}_t$, $\mathbf{RK}_t$, $\widehat{\mathbf{V}}_t$ and $\widehat{\mathbf{V}}_{\mathrm{MA},t}$ Given Regularly Spaced Returns

Panel A: No Noise Case ($\xi^2 = 0.0$)

		$\overline{\mathrm{Prob}}(\Sigma_t^{jj} \in \mathrm{Interval})$			$\overline{\mathrm{Prob}}(\Sigma_t^{jk} \in \mathrm{Interval})$		
Dimension	Frequency	$\mathbf{RC}_t$	$\mathbf{RK}_t$	$\widehat{\mathbf{V}}_t$	$\mathbf{RC}_t$	$\mathbf{RK}_t$	$\widehat{\mathbf{V}}_t$
3 Assets	5-minute	92.97%	87.48%	95.65%	92.44%	79.79%	95.45%
	1-minute	95.25%	87.75%	95.78%	94.91%	72.76%	96.05%
	10-second	96.92%	91.03%	97.12%	97.46%	67.07%	98.13%
10 Assets	5-minute	92.41%	87.11%	97.69%	91.95%	78.09%	98.09%
	1-minute	95.70%	89.64%	95.98%	96.08%	72.65%	96.34%
	10-second	96.45%	92.61%	96.61%	96.48%	68.59%	96.59%

Panel B: Microstructure Noise Case ($\xi^2 = 0.001$)

		$\overline{\mathrm{Prob}}(\Sigma_t^{jj} \in \mathrm{Interval})$			$\overline{\mathrm{Prob}}(\Sigma_t^{jk} \in \mathrm{Interval})$		
Dimension	Frequency	$\mathbf{RC}_t$	$\mathbf{RK}_t$	$\widehat{\mathbf{V}}_{\mathrm{MA},t}$	$\mathbf{RC}_t$	$\mathbf{RK}_t$	$\widehat{\mathbf{V}}_{\mathrm{MA},t}$
3 Assets	5-minute	92.03%	86.88%	93.84%	90.29%	53.95%	94.24%
	1-minute	0.00%	90.43%	93.17%	92.37%	89.16%	93.31%
	10-second	0.00%	93.37%	93.51%	87.55%	95.92%	93.31%
10 Assets	5-minute	92.03%	87.57%	80.44%	92.54%	55.67%	81.07%
	1-minute	0.00%	91.57%	89.26%	93.68%	90.13%	89.39%
	10-second	0.00%	94.42%	89.25%	93.96%	96.39%	73.01%

Results are based on 1000 days' simulation. $\overline{\mathrm{Prob}}(\Sigma_t^{jj} \in \mathrm{Interval})$ stands for the average of coverage frequencies for diagonal elements of Σ_t. $\overline{\mathrm{Prob}}(\Sigma_t^{jj} \in \mathrm{Interval})$ is the averaged coverage frequencies for off-diagonal covariance elements. Bayesian nonparametric estimator $\widehat{\mathbf{V}}_t$ and $\widehat{\mathbf{V}}_{\mathrm{MA},t}$ are estimated based on on 5000 MCMC runs, after 1000 burn-in.

Table 4.19: Coverage Probabilities of $\mathbf{RC}_t$, $\mathbf{RK}_t$ and $\widehat{\mathbf{V}}_{\text{MA},t}$ Given Nonsynchronous Returns

Panel A: No Noise Case ($\xi^2 = 0.0$)

		$\overline{\text{Prob}}(\Sigma_t^{jj} \in \text{Interval})$			$\overline{\text{Prob}}(\Sigma_t^{jk} \in \text{Interval})$		
Dimension	Frequency	$\mathbf{RC}_t^{5m}$	$\mathbf{RK}_t$	$\widehat{\mathbf{V}}_{\text{MA},t}$	$\mathbf{RC}_t^{5m}$	$\mathbf{RK}_t$	$\widehat{\mathbf{V}}_{\text{MA},t}$
3 Assets	$\lambda^j \in \{30, 40, 60\}$	92.44%	91.10%	96.12%	72.96%	23.63%	94.91%
	$\lambda^j \in \{5, 8, 10\}$	93.91%	93.78%	93.04%	91.23%	19.54%	92.17%
10 Assets	$\lambda^j \in \{30, 40, 60\}$	93.11%	91.31%	91.08%	72.76%	21.21%	93.32%
	$\lambda^j \in \{5, 8, 10\}$	93.78%	91.33%	86.18%	91.83%	18.80%	87.94%

Panel B: Microstructure Noise Case ($\xi^2 = 0.001$)

		$\overline{\text{Prob}}(\Sigma_t^{jj} \in \text{Interval})$			$\overline{\text{Prob}}(\Sigma_t^{jk} \in \text{Interval})$		
Dimension	Frequency	$\mathbf{RC}_t^{5m}$	$\mathbf{RK}_t$	$\widehat{\mathbf{V}}_{\text{MA},t}$	$\mathbf{RC}_t^{5m}$	$\mathbf{RK}_t$	$\widehat{\mathbf{V}}_{\text{MA},t}$
3 Assets	$\lambda^j \in \{30, 40, 60\}$	91.43%	87.15%	94.11%	74.90%	65.80%	95.98%
	$\lambda^j \in \{5, 8, 10\}$	90.90%	92.97%	94.44%	91.77%	90.63%	96.53%
10 Assets	$\lambda^j \in \{30, 40, 60\}$	92.97%	88.41%	84.24%	77.75%	64.12%	92.45%
	$\lambda^j \in \{5, 8, 10\}$	90.58%	90.38%	91.67%	90.19%	83.41%	89.86%

Results are based on 1000 days' simulation. $\overline{\text{Prob}}(\Sigma_t^{jj} \in \text{Interval})$ stands for the average of coverage frequencies for diagonal elements of Σ_t. $\overline{\text{Prob}}(\Sigma_t^{jj} \in \text{Interval})$ is the averaged coverage frequencies for off-diagonal covariance elements. Bayesian nonparametric estimator $\widehat{\mathbf{V}}_{\text{MA},t}$ is estimated based on on 5000 MCMC runs, after 1000 burn-in.

Table 4.20: ARMA(1,1) Model Estimation Results

Parameter	$\mathbf{RC}_t^{5m}$	$\mathbf{RK}_t$	$\widehat{\mathbf{V}}_{\text{MA},t}$
ϕ_1	0.9729	0.9569	0.9596
	(0.0138)	(0.0162)	(0.0168)
ρ_1	-0.8358	-0.5951	-0.6524
	(0.0358)	(0.0529)	(0.0552)
μ	1.3821	1.2570	1.2805
	(0.1511)	(0.1121)	(0.1192)

[1] This table reports OLS regression results of ARMA model: $\beta_t = \mu + \phi_1 \beta_{t-1} + \rho_1 \epsilon_{t-1} + \epsilon_t$. The value in the bracket are the standard error. Realized beta for AA-SPY combination are calculated using 2×2 $\mathbf{RC}_t$, $\mathbf{RK}_t$ and $\widehat{\mathbf{V}}_{\text{MA},t}$.

[2] Sample period: 2014/07/03 - 2016/06/29, 499 observations.

Table 4.21: Summary of Minimum Variance Portfolio Performance

	$\mathbf{RC}_t^{5m}$	$\mathbf{RK}_t$			$\widehat{\mathbf{V}}_{\mathrm{MA},t}$		
μ_0	SR	SR	$\Delta(\gamma=1)$	$\Delta(\gamma=10)$	SR	$\Delta(\gamma=1)$	$\Delta(\gamma=10)$
Panel A: Decay Rate $\kappa = 0.05$							
0.02	0.1486	0.1503	-3.6789	-3.7213	0.1518	6.0936	6.0616
0.05	0.1462	0.1479	-8.6373	-8.7157	0.1494	15.2909	15.2397
0.10	0.1454	0.1471	-15.3342	-15.4778	0.1486	30.8584	30.7705
Panel B: Decay Rate $\kappa = 0.1$							
0.02	0.1395	0.1456	8.2759	8.2344	0.1465	12.2842	12.2464
0.05	0.1371	0.1432	21.1735	21.0977	0.1441	31.0270	30.9607
0.10	0.1362	0.1424	44.0385	43.8999	0.1433	63.1898	63.0706
Panel C: Decay Rate $\kappa = 0.2$							
0.02	0.1293	0.1397	17.4046	17.3600	0.1401	19.1291	19.0880
0.05	0.1269	0.1373	43.9367	43.8528	0.1376	48.2247	48.1471
0.10	0.1261	0.1365	89.3768	89.2209	0.1368	97.8582	97.7211

The value listed in this table is the annualized base point fees that an investor with quadratic utility and risk aversion coefficient γ is willing to pay for switching portfolio based on 5-minute $\mathbf{RC}_t$ to portfolio based on $\mathbf{RK}_t$ or $\widehat{\mathbf{V}}_{\mathrm{MA},t}$. The period is from 07/02/2014 to 06/29/2016, totally 499 trading days.

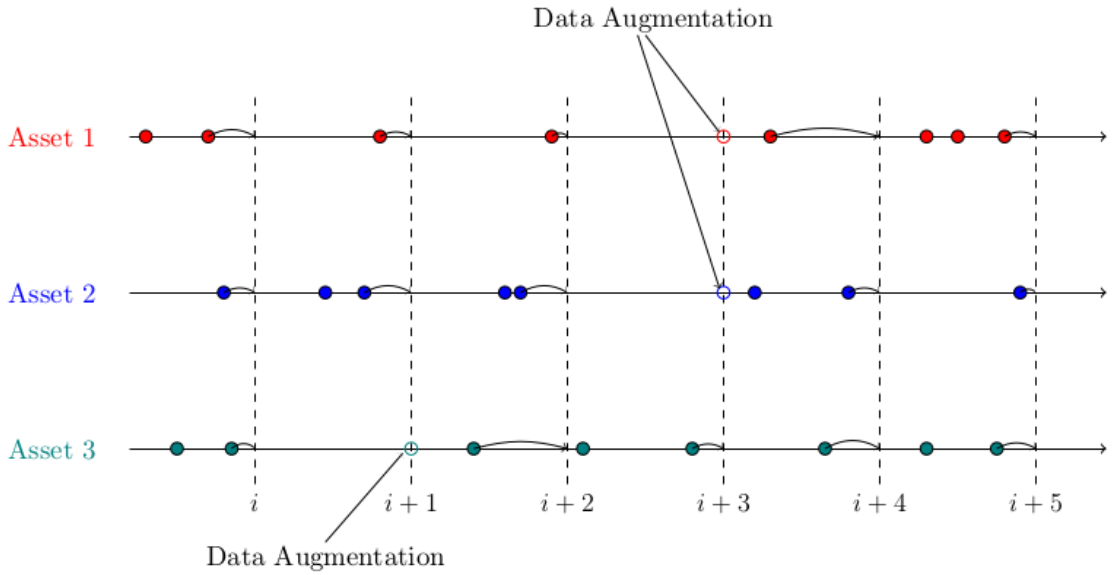

Figure 4.18: Synchronization with Data Augmentation

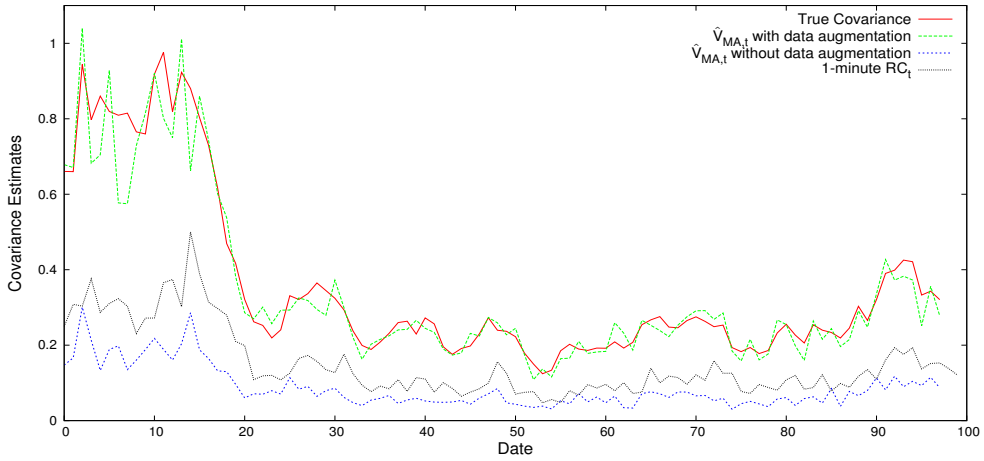

Figure 4.19: Covariance Estimates without Data Augmentation

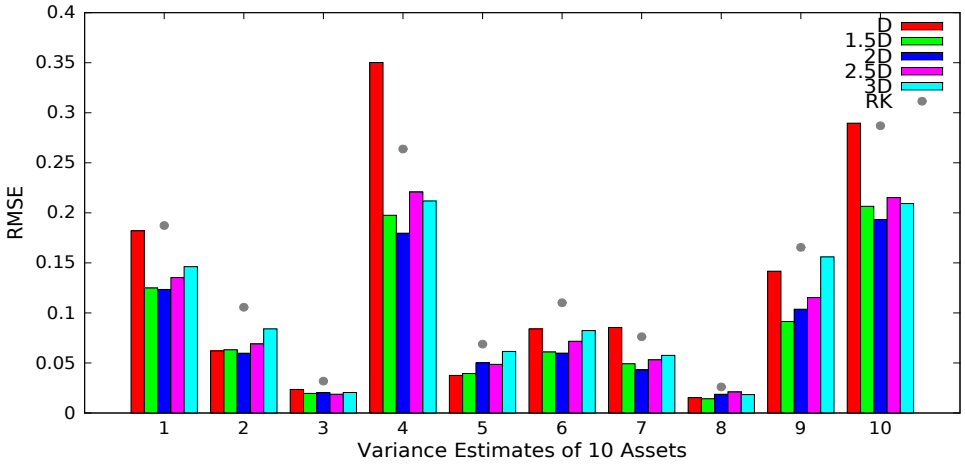

Figure 4.20: Root Mean Squared Error of Variance Estimates with Different Grid Length

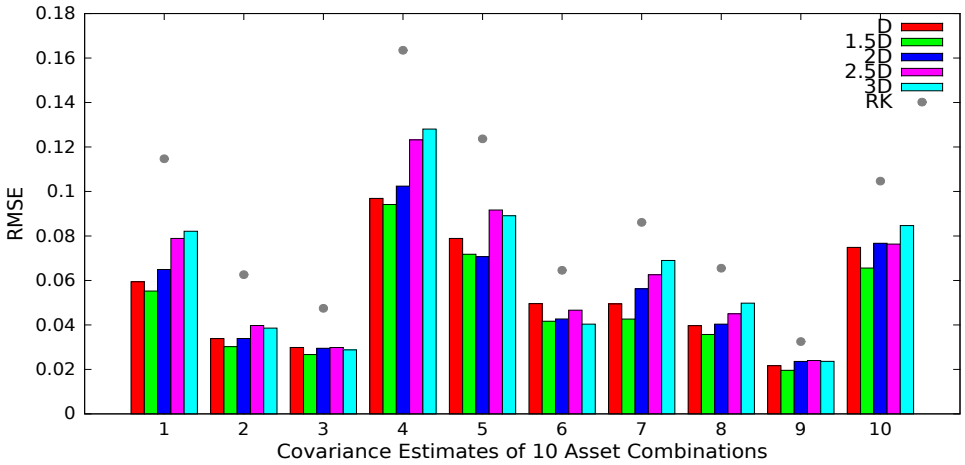

Figure 4.21: Root Mean Squared Error of Covariance Estimates with Different Grid Length

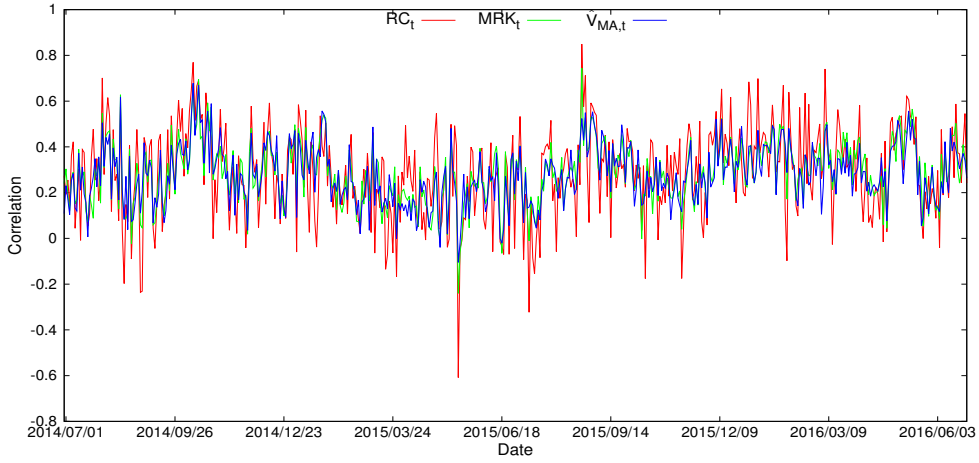

Figure 4.22: Correlation between AA and BAC based on $\mathbf{RC}_t^{5m}$, $\mathbf{RK}_t$ and $\widehat{\mathbf{V}}_{\mathrm{MA},t}$

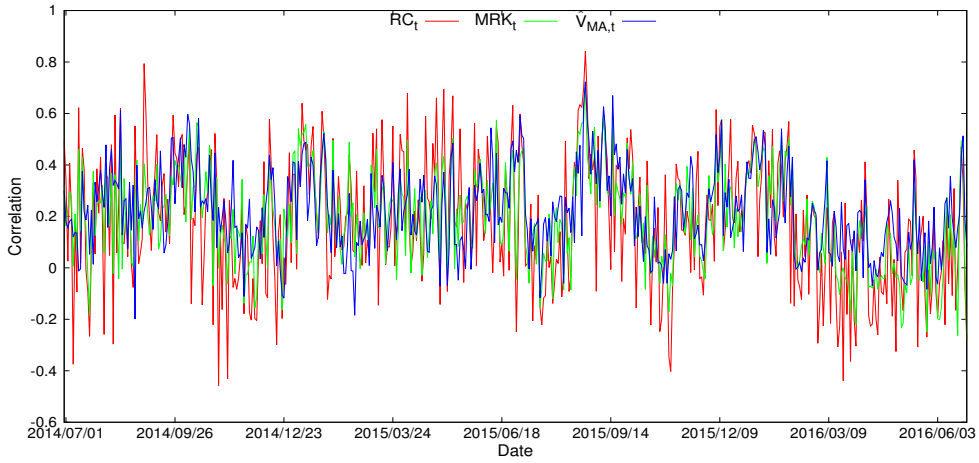

Figure 4.23: Correlation between CAT and F based on $\mathbf{RC}_t^{5m}$, $\mathbf{RK}_t$ and $\widehat{\mathbf{V}}_{\mathrm{MA},t}$

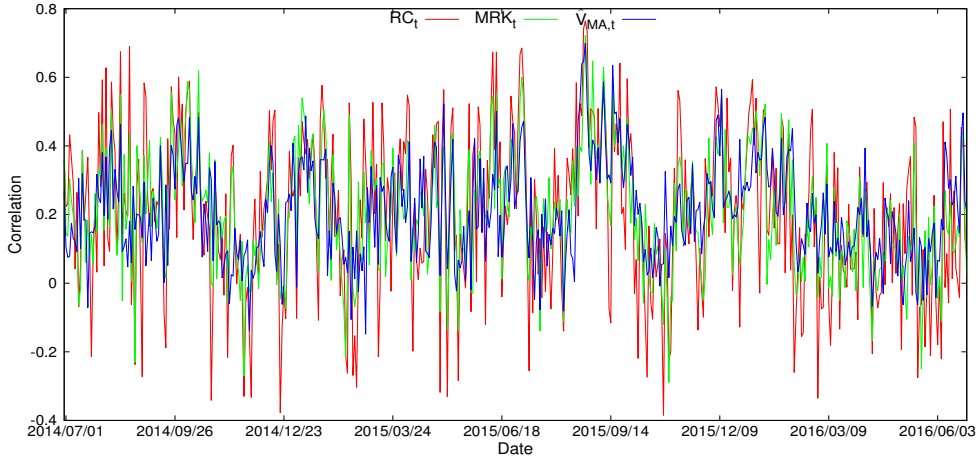

Figure 4.24: Correlation between GE and GIS based on $\mathbf{RC}_t^{5m}$, $\mathbf{RK}_t$ and $\widehat{\mathbf{V}}_{\mathrm{MA},t}$

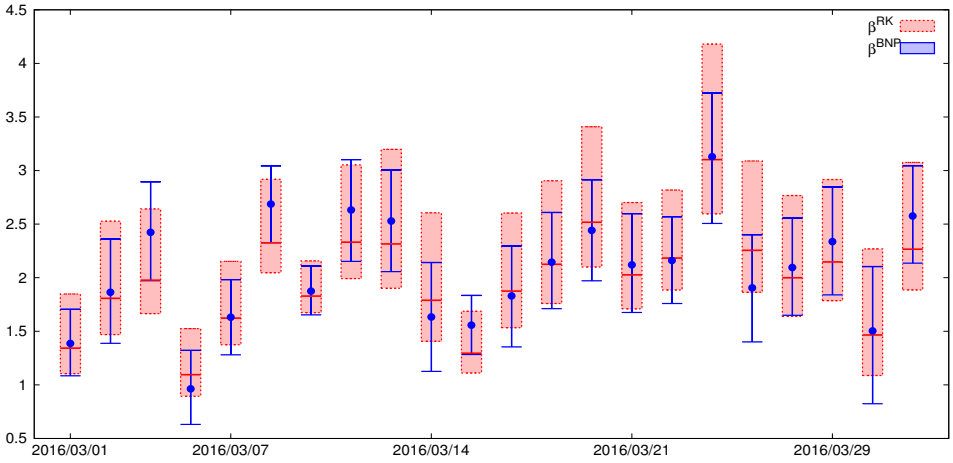

Figure 4.25: Realized Beta and 95% Intervals based on $\mathbf{RK}_t$ and $\widehat{\mathbf{V}}_{\mathrm{MA},t}$

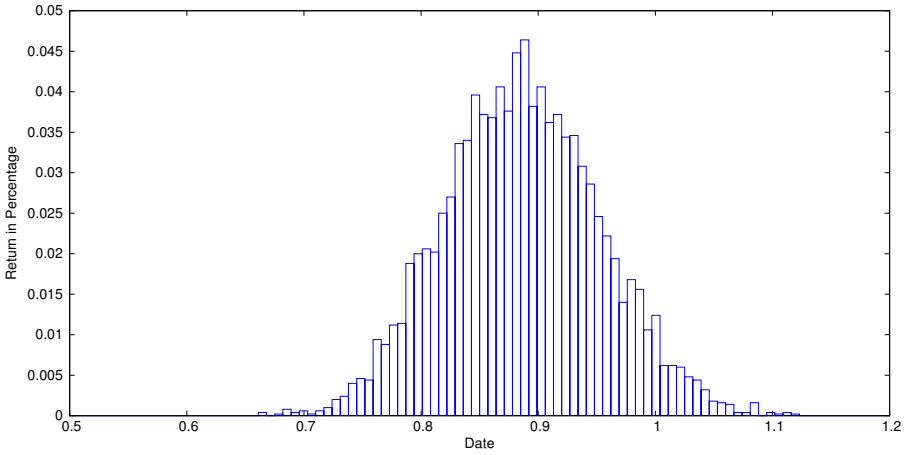

Figure 4.26: Histogram of Portfolio Return Caused by Covariance Uncertainty on June 29, 2016

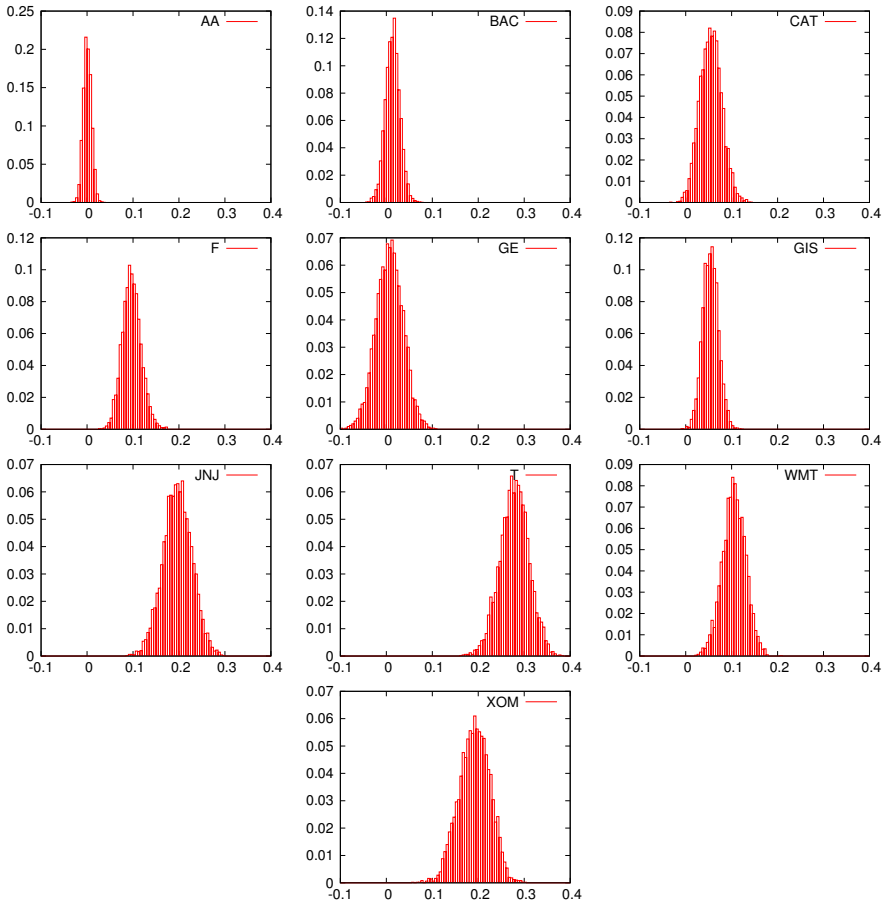

Figure 4.27: Histogram of Weights Caused by Covariance Uncertainty on June 29, 2016

Chapter 5

Conclusion

This thesis proposes new and improved approaches of estimating ex-post variance and covariance of asset returns from high frequency data based on Bayesian nonparametric tools. It also develops a class of new Markov switching models that exploit both return and ex-post volatility measures.

Chapter 2 introduces a way to link ex-post volatility measures with Markov switching models. The proposed models are able to jointly analyze asset returns and ex-post volatility measures and extract information from both data series. Parametric and nonparametric versions of the models are introduced in both univariate and multivariate settings. With the help of extra information provided by ex-post volatility measures, the joint Markov switching models outperform existing ones in density forecast of returns and prediction of volatility. In addition, the estimation of model parameters and the identification of the underlying state variable are improved. The portfolio allocation exercises show the models that incorporate ex-post volatility measures lead to better portfolio choice outcomes. The proposed joint models can be viewed as the strengthened version of Markov switching model and can be widely

applied in the analysis of financial and economic data.

Chapter 3 and Chapter 4 consider the problems of ex-post volatility estimation from a Bayesian prospective and offer alternative estimators which can be better than the existing ones in several aspects. Chapter 3 proposes a new ex-post variance estimation method which takes advantage of pooling and does not rely on asymptotic assumptions. Monte Carlo simulation results support that the proposed estimators are more accurate and have better finite sample results, especially when data frequency is low. Applications to equity returns show that the proposed method leads to a more realistic distribution of the ex-post variance estimator and better forecasting of the variance.

Chapter 4 extends the Bayesian nonparametric ex-post variance estimator introduced in Chapter 3 to its multivariate version and makes the estimator suitable for nonsynchronously spaced data with microstructure noise. As in the univariate case, the Bayesian nonparametric approach of estimating covariance matrix takes advantage of pooling and delivers exact finite sample results. Another important improvement is that the proposed covariance estimators are guaranteed to be positive definite. Furthermore, a new synchronization method based on data augmentation is provided. All of those features make the proposed estimator a better candidate to measure asset covariation. Empirical applications show that the Bayesian nonparametric measure of covariance matrix leads to better portfolio choices.

Chapter 3 and Chapter 4 contribute to the literature by introducing a finite sample approach of estimating ex-post volatility measures. The proposed methods provide both improved volatility measures and the entire distribution of estimators, which benefit not only the measurement of risk, but also the study of the uncertainty of risk

measures.