Bayesian Nonparametric Estimation of Ex-post Variance

Griffin, Jim and Liu, Jia and Maheu, John M

University of Kent, McMaster University, McMaster University

10 May 2016

Bayesian Nonparametric Estimation of Ex-post Variance[*]

Jim Griffin[†] Jia Liu[‡] John M. Maheu[§]

April, 2016

Abstract

Variance estimation is central to many questions in finance and economics. Until now ex-post variance estimation has been based on infill asymptotic assumptions that exploit high-frequency data. This paper offers a new exact finite sample approach to estimating ex-post variance using Bayesian nonparametric methods. 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.

[*] We are grateful for comments from seminar participants at the 2016 NBER-NSF Seminar on Bayesian Inference in Econometrics and Statistics. Maheu thanks the SSHRC of Canada for financial suport.

[†] School of Mathematics, Statistics and Actuarial Science, University of Kent, UK, J.E.Griffin-28@kent.ac.uk

[‡] DeGroote School of Business, McMaster University, Canada, liuj46@mcmaster.ca

[§] DeGroote School of Business, McMaster University, Canada and RCEA, Italy, maheujm@mcmaster.ca

1 Introduction

This paper introduces a new method of estimating ex-post volatility from high-frequency data using a Bayesian nonparametric model. The proposed method allows the data to cluster under a flexible framework. In contrast to existing classical estimation methods, it 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 & 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 & Lunde 2006, Aït-Sahalia et al. 2011). The impact of market microstructure noise on forecasting is explored in Aït-Sahalia & 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 & 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 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

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

(MCMC) techniques by Escobar & 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 three 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 mixture 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 paper is organized as follows. In section 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. Section 4 extends the Bayesian nonparametric model to deal with heteroskedastic and serially correlated microstructure noise. Section 5 provides an extensive simulation and comparison of the estimators. Applications to IBM and Disney data are found in Section 6. Section 7 concludes followed by an appendix.

2 Existing Ex-post Volatility Estimation

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 & 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 papers results apply to other time intervals.

Under the assumption of frictionless market and semimartingle, considering the following log-price diffusion,

$$\mathrm{d}p(t) = \mu(t)\,\mathrm{d}t + \sigma(t)\,\mathrm{d}W(t), \tag{1}$$

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,

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

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

$$RV_t = \sum_{i=1}^{n_t} r_{t,i}^2, \tag{3}$$

and $RV_t \xrightarrow{p} IV_t$, as $n_t \to \infty$ (Andersen, Bollerslev, Diebold & Labys 2001).

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

$$\sqrt{n_t}\frac{1}{\sqrt{2IQ_t}}(RV_t - IV_t) \xrightarrow{d} \mathrm{N}(0,1), \qquad \text{as} \quad n_t \to \infty, \tag{4}$$

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

$$RQ_t = \frac{n_t}{3}\sum_{i=1}^{n_t} r_{t,i}^4 \xrightarrow{p} IQ_t, \qquad \text{as} \quad n_t \to \infty. \tag{5}$$

2.1.1 Flat-top Realized Kernel

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

$$\tilde{p}_{t,i} = p_{t,i} + \epsilon_{t,i}, \tag{6}$$

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].

$$RK_t^F = \sum_{i=1}^{n_t} \tilde{r}_{t,i}^2 + \sum_{h=1}^{H} k\left(\frac{h-1}{H}\right)(\gamma_{-h} + \gamma_h), \qquad \gamma_h = \sum_{i=1}^{n_t} \tilde{r}_{t,i}\tilde{r}_{t,i-h}, \tag{7}$$

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 & 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) show that the asymptotic distribution of RK_t^F is

$$n_t^{1/4}\left(RK_t^F - IV_t\right) \xrightarrow{d} \mathrm{MN}\left\{0, 4IQ_t^{3/4}\omega\left(ck_\bullet^{0,0} + 2c^{-1}k_\bullet^{1,1}\frac{IV_t}{\sqrt{IQ_t}} + c^{-3}k_\bullet^{2,2}\right)\right\}, \tag{8}$$

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

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

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

$$\check{\omega}^2 = \exp\left[\log(\hat{\omega}^2) - RK_t/RV_t\right]. \tag{9}$$

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

$$TPQ_t = n_t \mu_{4/3}^{-3} \sum_{i=1}^{n_t-2} |\tilde{r}_{t,i}|^{4/3} |\tilde{r}_{t,i+1}|^{4/3} |\tilde{r}_{t,i+2}|^{4/3}, \tag{10}$$

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, $\check{\omega}^2$ and TPQ_t in equation (8), the asymptotic variance of RK_t^F can be calculated.

2.1.2 Non-negative Realized Kernel

The flat-top realized kernel discussed in previous subsection is based on the assumption that error term is white noise. However, the white noise assumption is restrictive and error term can be serial dependent or dependent with returns in reality. Another drawback of the RK_t^F is that it may provide negative volatility estimates, all be it very rarely. Barndorff-Nielsen et al. (2011) further introduced the non-negative realized kernel (RK_t^N) which is more robust to these assumptions of error term and is calculated as

$$RK_t^N = \sum_{h=-H}^{H} k\left(\frac{h}{H+1}\right)\gamma_h, \qquad \gamma_h = \sum_{i=|h|+1}^{n_t} \tilde{r}_{t,i}\tilde{r}_{t,i-|h|}. \tag{11}$$

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

$$n_t^{1/5}\left(RK_t^N - IV_t\right) \xrightarrow{d} \mathrm{MN}(\kappa, 4\kappa^2), \tag{12}$$

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 (9) and (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 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 4.

3.1 Model of High-frequency Returns

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

$$r_{t,i} = \mu_t + \sigma_{t,i} z_{t,i}, \quad z_{t,i} \overset{iid}{\sim} \mathrm{N}(0,1), \; i = 1, \ldots, n_t, \tag{13}$$

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

$$r_t = \sum_{i=1}^{n_t} r_{t,i} \tag{14}$$

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

$$V_t \equiv \mathrm{Var}(r_t | \mathcal{F}_t) = \sum_{i=1}^{n_t} \sigma^2_{t,i}. \tag{15}$$

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 $\sigma^2_{t,i}$.

3.2 A Bayesian Model with Pooling

In this section we discuss a nonparametric prior for the model of (13) that allows for pooling over common values of $\sigma^2_{t,i}$. 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

$$\begin{aligned} r_{t,i} \big| \mu_t, \sigma_{t,i} &\overset{iid}{\sim} \mathrm{N}(\mu_t, \sigma^2_{t,i}), \; i = 1, \ldots, n_t, && (16) \\ \sigma^2_{t,i} \big| G_t &\overset{iid}{\sim} G_t, && (17) \\ G_t \big| G_{0,t}, \alpha_t &\sim \mathrm{DP}(\alpha_t, G_{0,t}), && (18) \\ G_{0,t}(\sigma^2_{t,i}) &\equiv \mathrm{IG}(v_{0,t}, s_{0,t}), && (19) \end{aligned}$$

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 $\sigma^2_{t,i}\big|G_t \sim G_t$ 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 (17) is given as follows.

$$
\begin{aligned}
p(r_{t,i}\big|\mu_t, \Psi_t, w_t) &= \sum_{j=1}^{\infty} w_{t,j}\mathrm{N}(r_{t,i}|\mu_t, \psi^2_{t,j}), && (20)\\
w_{t,j} &= v_{t,j}\prod_{l=1}^{j-1}(1 - w_{t,l}), && (21)\\
v_{t,j} &\overset{iid}{\sim} \mathrm{Beta}(1, \alpha_t), && (22)
\end{aligned}
$$

where $\mathrm{N}(\cdot|\cdot,\cdot)$ denotes the density of the normal distribution, $\Psi_t = \{\psi^2_{t,1}, \psi^2_{t,2}\cdot\ldots,\}$ 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 $\sigma^2_{t,i}$ is $G_{0,t}$. For $\sigma^2_{t,i} \sim \mathrm{IG}(v_{0,t}, s_{0,t})$, the mean and variance of $\sigma^2_{t,i}$ are

$$
\mathrm{E}(\sigma^2_{t,i}) = \frac{s_{0,t}}{v_{0,t} - 1} \quad \text{and} \quad \mathrm{var}(\sigma^2_{t,i}) = \frac{s^2_{0,t}}{(v_{0,t} - 1)^2(v_{0,t} - 2)}. \qquad (23)
$$

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

$$
v_{0,t} = \frac{\left[\mathrm{E}(\sigma^2_{t,i})\right]^2}{\mathrm{var}(\sigma^2_{t,i})} + 2 \quad \text{and} \quad s_{0,t} = \mathrm{E}(\sigma^2_{t,i})(v_{0,t} - 1). \qquad (24)
$$

We use sample statistics $\widehat{\mathrm{var}}(r_{t,i})$ and $\widehat{\mathrm{var}}(r^2_{t,i})$ calculated with three days intraday returns (day $t-1$, day t, and day $t+1$) to set the values of $\mathrm{E}(\sigma^2_{t,i})$ and $\mathrm{var}(\sigma^2_{t,i})$, then use equation (24) to find $v_{0,t}$ and $s_{0,t}$. A shrinkage prior $\mathrm{N}(0, v^2)$ 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, $\alpha_t \sim \mathrm{Gamma}(a, b)$.

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

$$
E[V_t|\{r_{t,i}\}_{i=1}^{n_t}] = E\left[\sum_{i=1}^{n_t} \sigma^2_{t,i}\middle|\{r_{t,i}\}_{i=1}^{n_t}\right]. \qquad (25)
$$

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 $\sigma^2_{t,i}$ as separate and corresponds to no pooling. We discuss estimation of the model next.

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 (20). It introduces an auxiliary variable $u_{t,1:n_t} = \{u_{t,1}, \dots, u_{t,n_t}\}$ 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

$$f\left(r_{t,i}, u_{t,i} | w_t, \mu_t, \Psi_t\right) = \sum_{j=1}^{\infty} \mathbb{1}\left(u_{t,i} < w_{t,j}\right) \mathrm{N}\left(r_{t,i} | \mu_t, \psi^2_{t,j}\right), \tag{26}$$

and integrating out $u_{t,i}$ recovers (20).

It is convenient to rewrite the model in terms of a latent state variable $s_{t,i} \in \{1, 2, \dots\}$ that maps each observation to an associated component and parameter $\sigma^2_{t,i} = \psi^2_{t,s_{t,i}}$. Observations with a common state share the same variance parameter. For finite dataset the number of states (clusters) is finite and ordered from $1, \dots, K$. Note that the number of clusters K, is not a fixed value over the MCMC iterations. A new cluster with variance $\psi^2_{t,K+1} \sim G_{0,t}$ 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

$$p(\mu_t) \prod_{j=1}^{K} \left[p(\psi^2_{t,j})\right] p(\alpha_t) \prod_{i=1}^{n_t} \mathbb{1}(u_{t,i} < w_{t,s_{t,i}}) \mathrm{N}(r_{t,i} | \mu_t, \psi^2_{t,s_{t,i}}). \tag{27}$$

Each MCMC iteration contains the following sampling steps.

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

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

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

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

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

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

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

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 & 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.

3.4 Ex-post Variance Estimator

Conditional on the parameter vector θ the estimate of V_t is

$$E[V_t|\theta] = \sum_{i=1}^{n_t} \sigma_{t,s_i}^2. \tag{28}$$

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

$$\hat{V}_t = \frac{1}{M}\sum_{m=1}^{M}\sum_{i=1}^{n_t} \sigma_{t,i}^{2(m)}, \tag{29}$$

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 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

$$\widehat{\log(V_t)} = \frac{1}{M}\sum_{m=1}^{M}\log\left(\sum_{i=1}^{n_t}\sigma_{t,i}^{2(m)}\right). \tag{30}$$

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

4 Bayesian Estimator Under Microstructure Error

An early approach to deal with market microstructure noise was to prefilter with a time-series model (Andersen, Bollerslev, Diebold & Ebens 2001, Bollen & Inder 2002, Maheu & 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 specifications to account for noisy high-frequency returns. A significant difference is that we allow for heteroskedasticity in the noise process.

4.1 DPM-MA(1) Model

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

$$\tilde{p}_{t,i} = p_{t,i} + \epsilon_{t,i},\ \epsilon_{t,i} \sim \mathrm{N}(0, \omega_{t,i}^2), \tag{31}$$

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.

$$\begin{aligned}
\tilde{r}_{t,i}|\mu_t, \theta_t, \delta_{t,i}^2 &= \mu_t + \theta_t \eta_{t,i-1} + \eta_{t,i}, \quad \eta_{t,i} \sim \mathrm{N}(0, \delta_{t,i}^2) && (32)\\
\delta_{t,i}^2|G_t &\sim G_t, && (33)\\
G_t|G_{0,t}, \alpha_t &\sim \mathrm{DP}(\alpha_t, G_{0,t}), && (34)\\
G_{0,t}(\delta_{t,i}^2) &\equiv \mathrm{IG}(v_{0,t}, s_{0,t}). && (35)
\end{aligned}$$

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. 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.

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,

$$\tilde{p}_{t,i} = p_{t,i} + \epsilon_{t,i} - \rho_1 \epsilon_{t,i-1} - \cdots - \rho_{q-1}\epsilon_{t,i-q+1},\ \epsilon_{t,i} \sim \mathrm{N}(0, \omega_{t,i}^2). \tag{36}$$

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

$$\begin{aligned}
\tilde{r}_{t,i}|\mu_t, \{\theta_{t,j}\}_{j=1}^q, \delta_{t,i}^2 &= \mu_t + \sum_{j=1}^{q} \theta_{t,j}\eta_{t,i-j} + \eta_{t,i}, \quad \eta_{t,i} \sim \mathrm{N}(0, \delta_{t,i}^2) && (37)\\
\delta_{t,i}^2|G_t &\sim G_t, && (38)\\
G_t|G_{0,t}, \alpha_t &\sim \mathrm{DP}(\alpha_t, G_{0,t}), && (39)\\
G_{0,t}(\delta_{t,i}^2) &\equiv \mathrm{IG}(v_{0,t}, s_{0,t}). && (40)
\end{aligned}$$

The joint prior of $(\theta_{t,1}, \ldots, \theta_{t,q})$ is $\mathrm{N}(M_\Theta, V_\Theta)\mathbb{1}_{\{\Theta\}}$[5] and $(\eta_{t,0}, \ldots, \eta_{t,-(q-1)}) = (0, \ldots, 0)$.

[5] 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.

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. $\pi\left(\mu_t|\tilde{r}_{t,1:n_t}, \{\psi^2_{t,j}\}^K_{j=1}, \theta_t, s_{t,1:n_t}\right) \propto p\left(\mu_t\right) \prod_{i=1}^{n_t} \mathrm{N}\left(\tilde{r}_{t,i}|\mu_t + \theta_t \eta_{t,i-1}, \psi^2_{t,s_{t,i}}\right)$.
2. $\pi\left(\theta_t|\tilde{r}_{t,1:n_t}, \mu_t, \{\psi^2_{t,j}\}^K_{j=1}, s^t_{1:n_t}\right) \propto p\left(\theta_t\right) \prod_{i=1}^{n_t} p\left(\tilde{r}_{t,i}|\mu_t + \theta_t \eta_{t,i-1}, \psi^2_{t,s_{t,i}}\right)$.
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 $j = 1, \dots, K$.
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 $j = 1, \dots, K$.
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 $i = 1, \dots, n_t$.
6. Find the smallest K such that $\sum_{j=1}^K w_{t,j} > 1 - \min(u_{t,1:n_t})$.
7. $\pi\left(s_{t,i}|\tilde{r}_{1:n_t}, s_{t,1:n_t}, \mu_t, \theta_t, \{\psi^2_{t,j}\}^K_{j=1}, u_{t,1:n_t}, K\right) \propto \sum_{j=1}^K \mathbb{1}(u_{t,i} < w_{t,j}) \mathrm{N}(\tilde{r}_{t,i}|\mu_t + \theta_t \eta_{t,i-1}, \psi^2_{t,j})$ for $i = 1, \dots, n_t$.
8. $\pi(\alpha_t|K) \propto p(\alpha_t)p(K|\alpha_t)$.

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.

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.[6] 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

$$\hat{V}_{t,\mathrm{MA(1)}} = \frac{1}{M} \sum_{m=1}^{M} (1 + \theta_t^{(m)})^2 \sum_{i=1}^{n_t} \delta^{2(m)}_{t,i}, \tag{41}$$

where $\delta^{2(m)}_{t,i} = \psi^{2(m)}_{t,s^{(m)}_{t,i}}$ 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.

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

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

$$\hat{V}_{t,\mathrm{MA(q)}} = \frac{1}{M}\sum_{m=1}^{M}\left(1+\sum_{j=1}^{q}\theta_{t,j}^{(m)}\right)^2\sum_{i=1}^{n_t}\delta_{t,i}^{2(m)}. \tag{42}$$

Next we consider simulation evidence on these estimators.

5 Simulation Results

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 & Bollerslev (1998). The log-price follows

$$\mathrm{d}p(t) = \mu \mathrm{d}t + \sigma(t)\mathrm{d}W_p(t), \tag{43}$$

$$\mathrm{d}\sigma^2(t) = \alpha(\beta - \sigma^2(t))\mathrm{d}t + \gamma\sigma^2(t)\mathrm{d}W_\sigma(t). \tag{44}$$

where $W_p(t)$ and $W_\sigma(t)$ are two independent Wiener processes. The values of parameters follow Andersen & 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 & 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

$$\mathrm{d}p(t) = \mu \mathrm{d}t + \exp\left(\beta_0 + \beta_1 v(t)\right)\mathrm{d}W_p(t), \tag{45}$$

$$\mathrm{d}v(t) = \alpha v(t)\mathrm{d}t + \mathrm{d}W_v(t) \tag{46}$$

and the process for SV1FJ is

$$\mathrm{d}p(t) = \mu \mathrm{d}t + \exp\left(\beta_0 + \beta_1 v(t)\right)\mathrm{d}W_p(t) + dJ(t)\ , \tag{47}$$

$$\mathrm{d}v(t) = \alpha v(t)\mathrm{d}t + \mathrm{d}W_v(t), \tag{48}$$

where $\mathrm{corr}(\mathrm{d}W_p(t), \mathrm{d}W_v(t)) = \rho$, 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 & 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 & Tauchen (2005).[7]

$$\mathrm{d}p(t) = \mu \mathrm{d}t + s\text{-}\exp\left(\beta_0 + \beta_1 v_1(t) + \beta_2 v_2(t)\right)\mathrm{d}W_p(t), \tag{49}$$

$$\mathrm{d}v_1(t) = \alpha_1 v_1(t)\mathrm{d}t + \mathrm{d}W_{v_1}(t), \tag{50}$$

$$\mathrm{d}v_2(t) = \alpha_2 v_2(t)\mathrm{d}t + \left(1 + \psi v_2(t)\right)\mathrm{d}W_{v_2}(t), \tag{51}$$

[7] The function $s\text{-}\exp$ is defined as $s\text{-}\exp(x) = \exp(x)$ if $x \le x_0$ 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 $x_0 = \log(1.5)$.

where $\mathrm{corr}(\mathrm{d}W_p(t), \mathrm{d}W_{v_1}(t)) = \rho_1$ and $\mathrm{corr}(\mathrm{d}W_p(t), \mathrm{d}W_{v_2}(t)) = \rho_2$. 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 & 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 startup conditions 500 initial days are dropped from the simulation.

5.1.1 Independent Noise

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

$$\begin{aligned} \tilde{p}_{t,i} &= p_{t,i-1} + \epsilon_{t,i}, \\ \epsilon_{t,i} &\sim \mathrm{N}(0, \sigma_\omega^2), \\ \sigma_\omega^2 &= \xi^2 \mathrm{var}(r_t). \end{aligned} \tag{52}$$

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 & Russell (2008).

5.1.2 Dependent Noise

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

$$\begin{aligned} \tilde{p}_{t,i} &= p_{t,i-1} + \epsilon_{t,i}, \\ \epsilon_{t,i} &\sim \mathrm{N}\left(\mu_{\epsilon_{t,i}}, \sigma_\omega^2\right), \\ \mu_{\epsilon_{t,i}} &= \sum_{l=1}^{\phi} (1 - l/\phi)\left(p_{t,i-l} - p_{t,i-1-l}\right), \\ \sigma_\omega^2 &= \xi^2 \mathrm{var}(r_t), \end{aligned} \tag{53}$$

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.

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)

$$\sigma_t^2 \equiv \sum_{i=1}^{23400} r_{t,i}^2. \tag{54}$$

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

$$\begin{aligned} \text{RMSE}(\widehat{\sigma_t^2}) &= \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(\widehat{\sigma_t^2} - \sigma_t^2\right)^2}, & (55) \\ \text{Bias}(\widehat{\sigma_t^2}) &= \frac{1}{T}\sum_{t=1}^{T}\left(\widehat{\sigma_t^2} - \sigma_t^2\right). & (56) \end{aligned}$$

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 (4), (8) and (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

$$IQ_t^{\text{true}} = 23400 \sum_{i=1}^{23400} \sigma_{t,i}^4, \tag{57}$$

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 (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.

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.

5.3 No Microstructure Noise

Figure 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 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 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 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 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.

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 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 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 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 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 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.

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 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 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 $\hat{V}_{t,\mathrm{MA}(1)}$ and $\hat{V}_{t,\mathrm{MA}(2)}$ 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 6 supports these findings.

Table 9 shows $\hat{V}_{t,\mathrm{MA}(1)}$ and $\hat{V}_{t,\mathrm{MA}(2)}$ have smaller bias if return frequency is one minute or higher.

Table 10 shows the coverage probabilities of all the five estimators. The finite sample results of $\hat{V}_{t,\mathrm{MA}(2)}$ are all very close to the optimal level, no matter the data frequency.

5.6 Evidence of Pooling

Figure 7-9 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.

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.

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 Kibot website[8]. 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.

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

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 suggest 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 = 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.

6.1.1 Ex-post Variance Estimation

Table 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 $\hat{V}_{t,\text{MA}(1)}$ for the 5-minute data and the $\hat{V}_{t,\text{MA}(4)}$ for the 1-minute data in the remainder of the analysis. Comparison with the kernel estimators is found in Figures 10 and 11. Except for the extreme values they are very similar.

Interval estimates for two sub-periods are shown in Figures 12 and 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[9] are also provided with some differences remaining.

The degree of pooling from the Bayesian estimators is found in Figure 14 and 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.

6.1.2 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

$$\hat{V}_t = \beta_0 + \beta_1 \hat{V}_{t-1} + \beta_2 \hat{V}_{t-1|t-5} + \beta_3 \hat{V}_{t-1|t-22} + \epsilon_t, \tag{58}$$

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

$$RV_t = \beta_0 + \left(\beta_1 + \beta_{1Q} RQ_{t-1}^{1/2}\right) RV_{t-1} + \beta_2 RV_{t-1|t-5} + \beta_3 RV_{t-1|t-22} + \epsilon_t \tag{59}$$

[9] 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 & Shephard (2002), Barndorff-Nielsen et al. (2008) and Barndorff-Nielsen et al. (2011).

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 HAR model in forecasting.[10]

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 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

$$\hat{V}_t = \beta_0 + \left(\beta_1 + \beta_{1Q}\widehat{\text{var}}(V_{t-1})^{1/2}\right)\hat{V}_{t-1} + \beta_2\hat{V}_{t-1|t-5} + \beta_3\hat{V}_{t-1|t-22} + \epsilon_t, \tag{60}$$

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

Table 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 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 $t-1$.[11]

The first column of Table 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.

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.

[10] 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.

[11] 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.

Table 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 16 and 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 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.

7 Conclusion

This paper 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 three 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 the realized variance and kernel estimators in terms of average statistic properties as well as time-series characteristics. The Bayesian estimators can be used with confidence and have several benefit 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.

References

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

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

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

Andersen, T. & Bollerslev, T. (1998), 'Answering the skeptics: Yes, standard volatility models do provide accurate forecasts', *International Economic Review* **39**(4), 885–905.

Andersen, T. G. & Benzoni, L. (2008), Realized volatility, Working Paper Series WP-08-14, Federal Reserve Bank of Chicago.
URL: *https://ideas.repec.org/p/fip/fedhwp/wp-08-14.html*

Andersen, T. G., Bollerslev, T., Diebold, F. X. & Ebens, H. (2001), 'The distribution of realized stock return volatility', *Journal of Financial Economics* **61**(1), 43–76.

Andersen, T. G., Bollerslev, T., Diebold, F. X. & Labys, P. (2001), 'The Distribution of Realized Exchange Rate Volatility', *Journal of the American Statistical Association* **96**, 42–55.

Andersen, T. G., Bollerslev, T., Diebold, F. X. & Labys, P. (2003), 'Modeling and Forecasting Realized Volatility', *Econometrica* **71**(2), 579–625.

Andersen, T. G., Bollerslev, T. & Meddahi, N. (2011), 'Realized volatility forecasting and market microstructure noise', *Journal of Econometrics* **160**(1), 220–234.

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

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A. & 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. & Shephard, N. (2009), 'Realized kernels in practice: trades and quotes', *Econometrics Journal* **12**(3), 1–32.

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

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

Barndorff-Nielsen, O. E. & Shephard, N. (2006), 'Econometrics of Testing for Jumps in Financial Economics Using Bipower Variation', *Journal of Financial Econometrics* **4**(1), 1–30.

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

Bollerslev, T., Patton, A. & Quaedvlieg, R. (2016), 'Exploiting the errors: A simple approach for improved volatility forecasting', *Journal of Econometrics* **192**, 1–18.

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

Corradi, V., Distaso, W. & 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. & 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. & Meddahi, N. (2009), 'Bootstrapping realized volatility', *Econometrica* **77**(1), 283–306.

Hansen, P., Large, J. & Lunde, A. (2008), 'Moving Average-Based Estimators of Integrated Variance', *Econometric Reviews* **27**(1-3), 79–111.

Hansen, P. R. & Lunde, A. (2006), 'Realized Variance and Market Microstructure Noise', *Journal of Business & Economic Statistics* **24**, 127–161.

Huang, X. & 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. & Walker, S. G. (2011), 'Slice sampling mixture models', *Statistics and Computing* **21**(1), 93–105.

Maheu, J. M. & 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* pp. 639–650.

Zhang, L., Mykland, P. A. & Aït-Sahalia, Y. (2005), ‘A Tale of Two Time Scales: Determining Integrated Volatility With Noisy High-Frequency Data’, *Journal of the American Statistical Association* **100**, 1394–1411.

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

8 Appendix

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

$$\widetilde{p}_{t,i} = p_{t,i} + \epsilon_{t,i}, \qquad E(\epsilon_{t,i}) = 0 \text{ and } \mathrm{var}(\epsilon_{t,i}) = \omega^2_{t,i}. \tag{61}$$

The log return process is constructed as follows,

$$\widetilde{r}_{t,i} = \widetilde{p}_{t,i} - \widetilde{p}_{t,i-1} = p_{t,i} - p_{t,i-1} + \epsilon_{t,i} - \epsilon_{t,i-1} = r_{t,i} + \epsilon_{t,i} - \epsilon_{t,i-1}, \tag{62}$$

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

$$\mathrm{var}(\widetilde{r}_{t,i}) = \sigma^2_{t,i} + \omega^2_{t,i} + \omega^2_{t,i-1}, \tag{63}$$

$$\mathrm{cov}(\widetilde{r}_{t,i}, \widetilde{r}_{t,i-1}) = -\omega^2_{t,i-1}. \tag{64}$$

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

$$\widetilde{r}_{t,i} = \mu_t + \theta_t \eta_{t,i-1} + \eta_{t,i}, \qquad \eta_{t,i} \sim N(0, \delta^2_{t,i}), \tag{65}$$

which will be used to recover an estimate of ex-post variance for the pure return process, $V_t = \sum_{i=1}^{n_t} \sigma^2_{t,i}$. The corresponding moments of this process are

$$\mathrm{var}(\widetilde{r}_{t,i}) = \theta_t^2 \delta^2_{t,i-1} + \delta^2_{t,i}, \tag{66}$$

$$\mathrm{cov}(\widetilde{r}_{t,i}, \widetilde{r}_{t,i-1}) = \theta_t \delta^2_{t,i-1}. \tag{67}$$

Equating (63) and (66), we have

$$\sigma^2_{t,i} + \omega^2_{t,i} + \omega^2_{t,i-1} = \theta_t^2 \delta^2_{t,i-1} + \delta^2_{t,i} \tag{68}$$

Equating (64) and (67), we have

$$-\omega^2_{t,i-1} = \theta_t \delta^2_{t,i-1} \qquad \text{and} \qquad -\omega^2_{t,i} = \theta_t \delta^2_{t,i}. \tag{69}$$

Based on the result in (69), the summation of $\delta^2_{t,i}$, over $i = 1, \ldots, n_t$, equals

$$\sum_{i=1}^{n_t} \delta^2_{t,i} = -\frac{1}{\theta_t} \sum_{i=1}^{n_t} w^2_{t,i}. \tag{70}$$

Plugging both terms in (69) into (68), yields

$$\sigma^2_{t,i} + \omega^2_{t,i} + \omega^2_{t,i-1} = -\theta_t \omega^2_{t,i-1} - \frac{\omega^2_{t,i}}{\theta_t} \tag{71}$$

$$\sigma^2_{t,i} + \left(1 + \frac{1}{\theta_t}\right) \omega^2_{t,i} + (1 + \theta_t) \omega^2_{t,i-1} = 0. \tag{72}$$

Using the results in (72), the summation of $\sigma^2_{t,i}$, over $i = 1, \ldots, n_t$, equals

$$\sum_{i=1}^{n_t} \sigma^2_{t,i} + \left(1 + \frac{1}{\theta_t}\right) \sum_{i=1}^{n_t} \omega^2_{t,i} + (1 + \theta_t) \sum_{i=1}^{n_t} \omega^2_{t,i-1} = 0 \tag{73}$$

$$V_t = -\left(1 + \frac{1}{\theta_t}\right) \sum_{i=1}^{n_t} \omega^2_{t,i} - (1 + \theta_t) \sum_{i=1}^{n_t} \omega^2_{t,i-1}. \tag{74}$$

The ratio between (70) and (74) is

$$\begin{aligned}
\frac{V_t}{\sum_{i=1}^{n_t} \delta^2_{t,i}} &= \frac{-\left(1 + \frac{1}{\theta_t}\right) \sum_{i=1}^{n_t} \omega^2_{t,i} - (1 + \theta_t) \sum_{i=1}^{n_t} \omega^2_{t,i-1}}{-\frac{1}{\theta_t} \sum_{i=1}^{n_t} \omega^2_{t,i}} && (75)\\
&= \frac{(1 + \theta_t) \sum_{i=1}^{n_t} \omega^2_{t,i} + (\theta_t + \theta_t^2) \sum_{i=1}^{n_t} \omega^2_{t,i-1}}{\sum_{i=1}^{n_t} \omega^2_{t,i}} && (76)\\
&= \frac{(1 + \theta_t)^2 \sum_{i=1}^{n_t-1} \omega^2_{t,i} + (1 + \theta_t) \omega^2_{t,n_t} + (\theta_t + \theta_t^2) \omega^2_{t,0}}{\sum_{i=1}^{n_t-1} \omega^2_{t,i} + \omega^2_{t,n_t}} && (77)\\
&= (1 + \theta_t)^2, \quad \text{if} \quad \omega_{t,n_t} = \omega_{t,0}. && (78)
\end{aligned}$$

Finally, we have

$$(1 + \theta_t)^2 \sum_{i=1}^{n_t} \delta^2_{t,i} = V_t, \qquad \text{if} \quad \omega_{t,n_t} = \omega_{t,0}. \tag{79}$$

8.2 Adjustment to DPM-MA(2) Estimator

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

$$\widetilde{p}_{t,i} = p_{t,i} + \epsilon_{t,i} - \rho\epsilon_{t,i-1}, \qquad E(\epsilon_{t,i}) = 0 \text{ and } \mathrm{var}(\epsilon_{t,i}) = \omega^2_{t,i}. \tag{80}$$

Then log return process is constructed as follows.

$$\begin{aligned}
\widetilde{r}_{t,i} &= \widetilde{p}_{t,i} - \widetilde{p}_{t,i-1}\\
&= p_{t,i} - p_{t,i-1} + \epsilon_{t,i} - \rho\epsilon_{t,i-1} - \epsilon_{t,i-1} + \rho\epsilon_{t,i-2} \\
&= r_{t,i} + \epsilon_{t,i} - (1 + \rho)\epsilon_{t,i-1} + \rho\epsilon_{t,i-2}.
\end{aligned} \tag{81}$$

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

$$\widetilde{r}_{t,i} = \mu_t + \theta_{1t}\eta_{t,i-1} + \theta_{2t}\eta_{t,i-2} + \eta_{t,i}, \qquad \eta_{t,i} \sim N(0, \delta^2_{t,i}) \tag{82}$$

it can be shown the adjustment term is

$$(1 + \theta_{1t} + \theta_{2t})^2 \sum_{i=1}^{n_t} \delta^2_{t,i} = V_t, \qquad \text{if} \quad \omega_{t,n_t-1} = \omega_{t,0} \quad \text{and} \quad \omega_{t,n_t} = \omega_{t,-1}. \tag{83}$$

Similar results hold for higher order MA models.

Table 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 (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 2: RMSE 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: Bias 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 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 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 6: Bias 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	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($\hat{V}_{t,\mathrm{MA}(1)}$)	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($\hat{V}_{t,\mathrm{MA}(1)}$)	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($\hat{V}_{t,\mathrm{MA}(1)}$)	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($\hat{V}_{t,\mathrm{MA}(1)}$)	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 $\hat{V}_{t,\mathrm{MA}(1)}$ based on returns at different frequencies and simulated from 4 DGPs. The price is contaminated with white noise.

Table 7: Coverage Probability (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%
	$\hat{V}_{t,\text{MA}(1)}$	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%
	$\hat{V}_{t,\text{MA}(1)}$	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%
	$\hat{V}_{t,\text{MA}(1)}$	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%
	$\hat{V}_{t,\text{MA}(1)}$	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 $\hat{V}_{\text{MA}(1)}$ based on 5000 days results for different data generating processes. The price is contaminated with white noise.

Table 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 9: Bias 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	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($\hat{V}_{t,\mathrm{MA}(1)}$)	0.02134	0.03879	0.04075	**0.00062**
	Bias($\hat{V}_{t,\mathrm{MA}(2)}$)	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($\hat{V}_{t,\mathrm{MA}(1)}$)	**0.00909**	**0.01674**	**0.01661**	-0.01113
	Bias($\hat{V}_{t,\mathrm{MA}(2)}$)	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($\hat{V}_{t,\mathrm{MA}(1)}$)	**0.00740**	**0.01384**	**0.00991**	-0.01316
	Bias($\hat{V}_{t,\mathrm{MA}(2)}$)	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($\hat{V}_{t,\mathrm{MA}(1)}$)	0.10993	0.20159	0.20140	0.13645
	Bias($\hat{V}_{t,\mathrm{MA}(2)}$)	**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 $\hat{V}$, $\hat{V}_{\mathrm{MA}(1)}$ and $\hat{V}_{\mathrm{MA}(2)}$ based on returns at different frequencies and simulated from 4 DGPs. The observed prices contains microstructure noise that is dependent with returns.

Table 10: Coverage Probability (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 11: Summary Statistics: IBM

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 12: HAR and HARQ Model Regression Result Based on IBM Ex-post Variance Estimators

Data Freq.	Parameter	HAR		HARQ	
		RK_t^F	$\hat{V}_{t,\mathrm{MA}(1)}$	RK_t^F	$\hat{V}_{t,\mathrm{MA}(1)}$
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	$\hat{V}_{t,\mathrm{MA}(4)}$	RK_t^N	$\hat{V}_{t,\mathrm{MA}(4)}$
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 $\hat{V}_{t,\mathrm{MA}(1)}$ calculated using 5-minute returns and the bottom panel shows the results of 1-minute RK_t^N and $\hat{V}_{t,\mathrm{MA}(4)}$. The values in brackets are standard error of coefficients.

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

Table 13: Out-of-Sample Forecasts of IBM Volatility

Panel A: 5-minute Return			
Dependent Variable	Regressors	HAR	HARQ
5-minute RK_t^F	RK_t^F	1.84113	1.84444
	$\hat{V}_{t,\mathrm{MA}(1)}$	**1.84083**	**1.81263**
5-minute $\hat{V}_{t,\mathrm{MA}(1)}$	RK_t^F	1.86262	1.86699
	$\hat{V}_{t,\mathrm{MA}(1)}$	**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**
	$\hat{V}_{t,\mathrm{MA}(4)}$	**1.87006**	1.83173
1-minute $\hat{V}_{t,\mathrm{MA}(4)}$	RK_t^N	1.93618	1.88542
	$\hat{V}_{t,\mathrm{MA}(4)}$	**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 14: Summary Statistics: Disney

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.

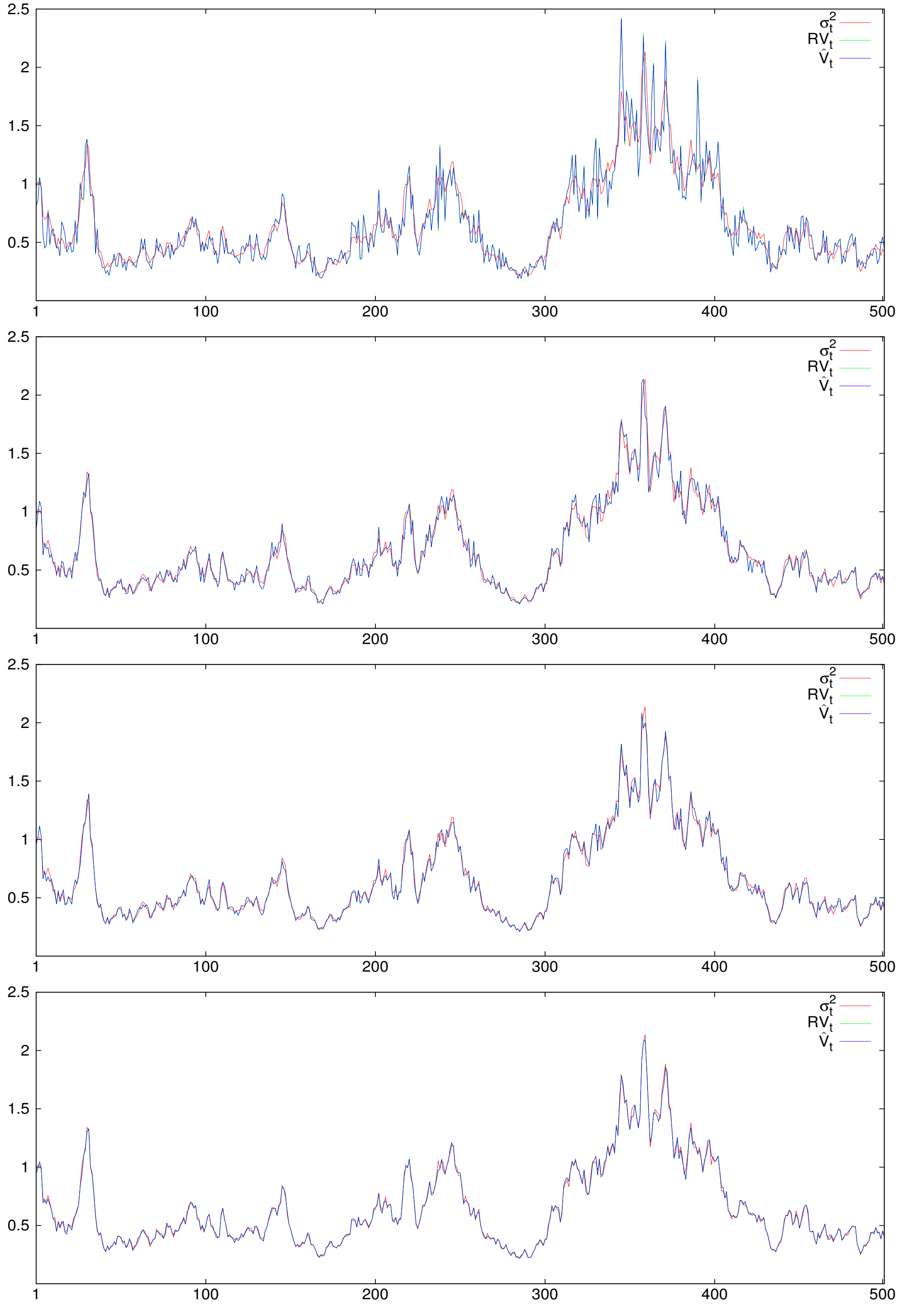

Figure 1: True Variance σ_t^2, RV_t and $\hat{V}_t$ (No Microstructure Noise Case). From top to bottom: 5-minute, 1-minute, 30-second, 10-second returns simulated from GARCH(1,1) DGP without noise.

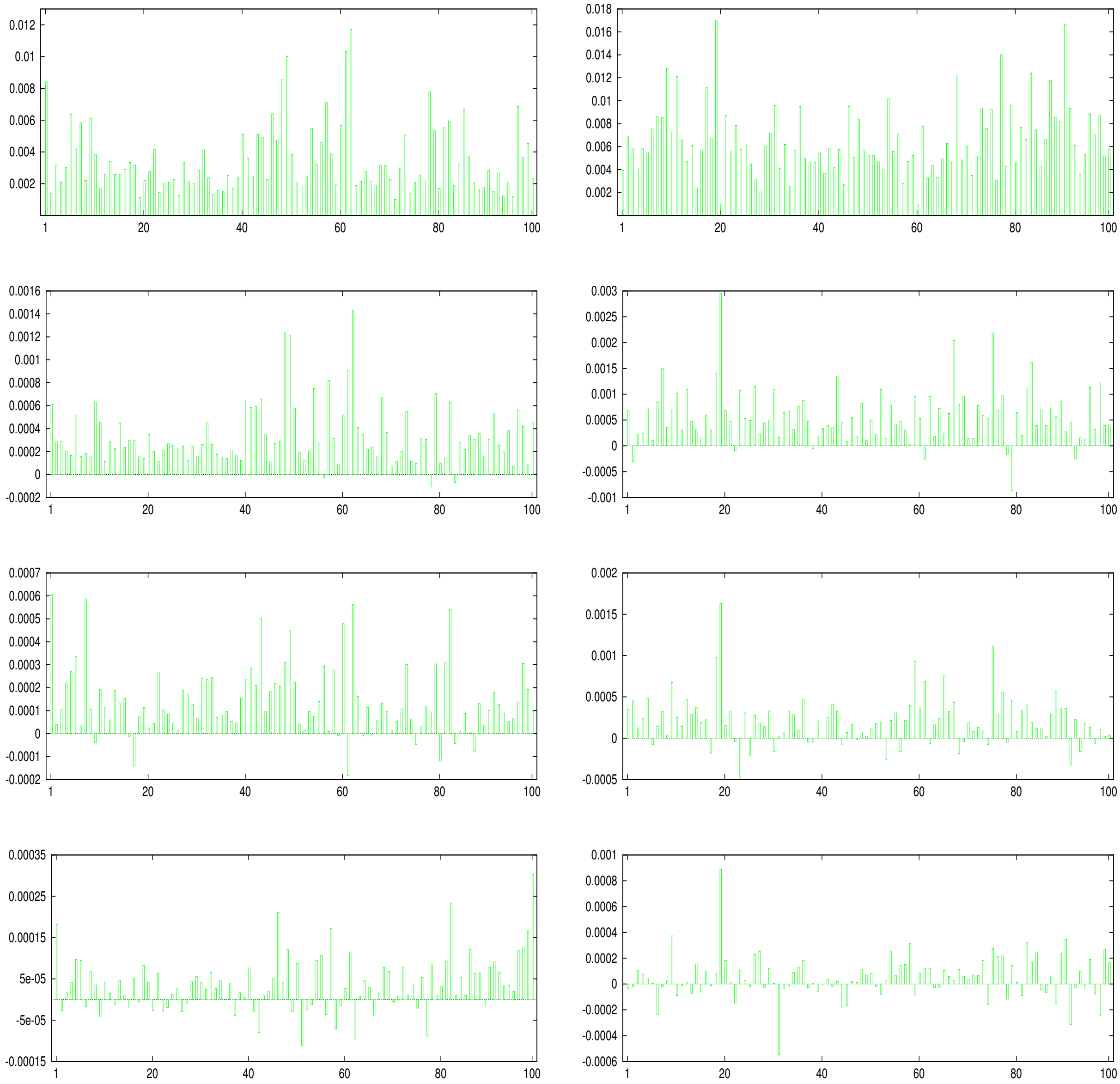

Figure 2: RMSE(RV_t)-RMSE($\hat{V}_t$) in 100 subsamples (No Microstructure Noise Case). Left: GARCH(1,1) DGP, Right: SV1F DGP, From top to bottom: 5-minute, 1-minute, 30-second, 10-second returns.

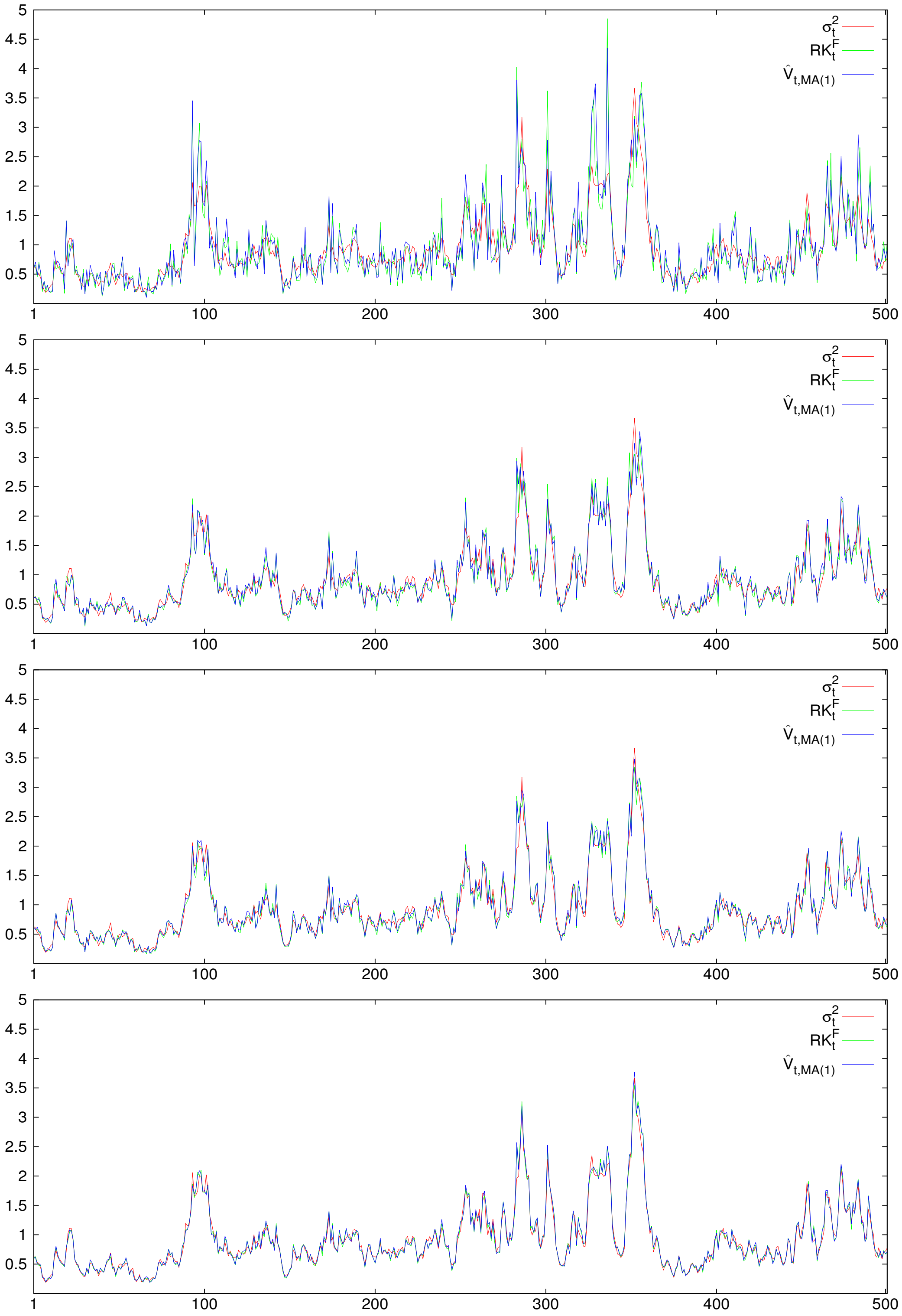

Figure 3: True Variance σ_t^2, RK_t^F and $\hat{V}_{t,\text{MA}(1)}$ (Independent Microstructure Noise Case). From top to bottom: 5-minute, 1-minute, 30-second, 10-second returns simulated from SV1F DGP with independent noise.

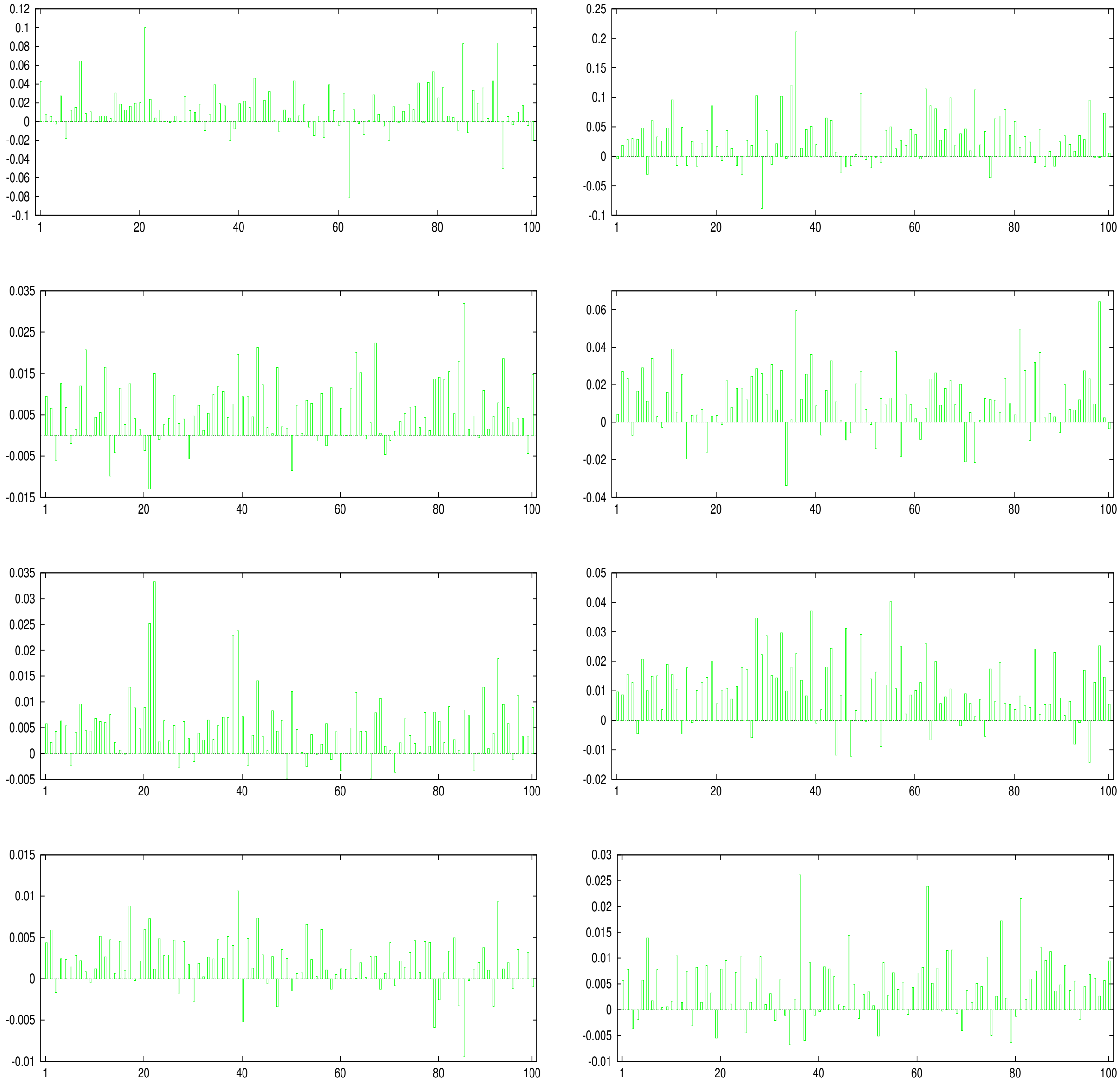

Figure 4: RMSE(RK_t^F)-RMSE($\hat{V}_{t,\mathrm{MA}(1)}$) in 100 subsamples (Independent Microstructure Noise Case). Left: GARCH(1,1) DGP, Right: SV1F DGP, From top to bottom: 5-minute, 1-minute, 30-second, 10-second returns.

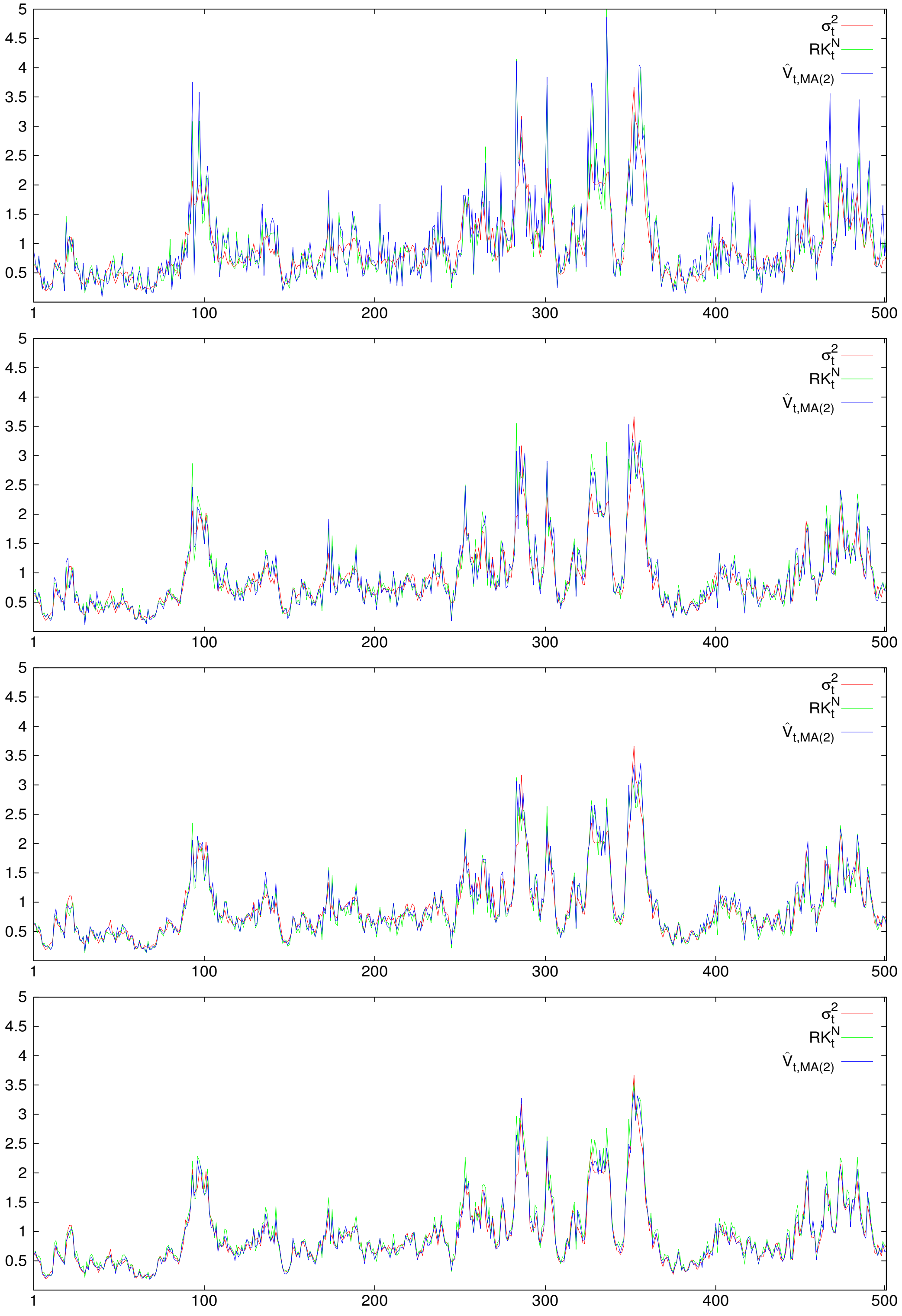

Figure 5: True Variance σ_t^2, RK_t^N and $\hat{V}_{t,\mathrm{MA}(2)}$ (Dependent Microstructure Noise Case). From top to bottom: 5-minute, 1-minute, 30-second, 10-second returns simulated from SV1F DGP with noise correlated with returns.

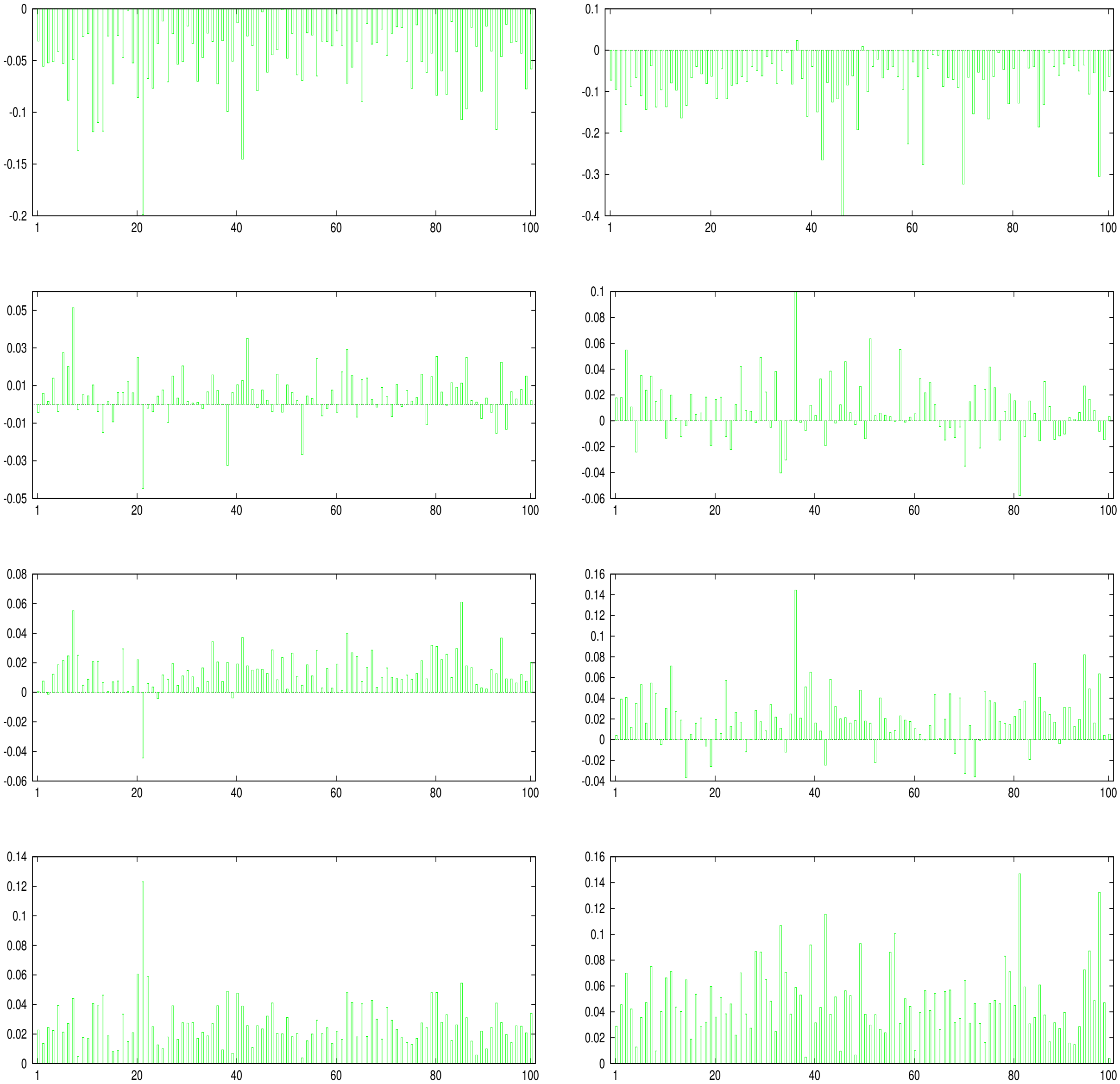

Figure 6: RMSE(RK_t^N)-RMSE($\hat{V}_{t,\mathrm{MA}(2)}$) in 100 subsamples (Dependent Microstructure Noise Case). Left: GARCH(1,1) DGP, Right: SV1F DGP, From top to bottom: 5-minute, 1-minute, 30-second, 10-second returns.

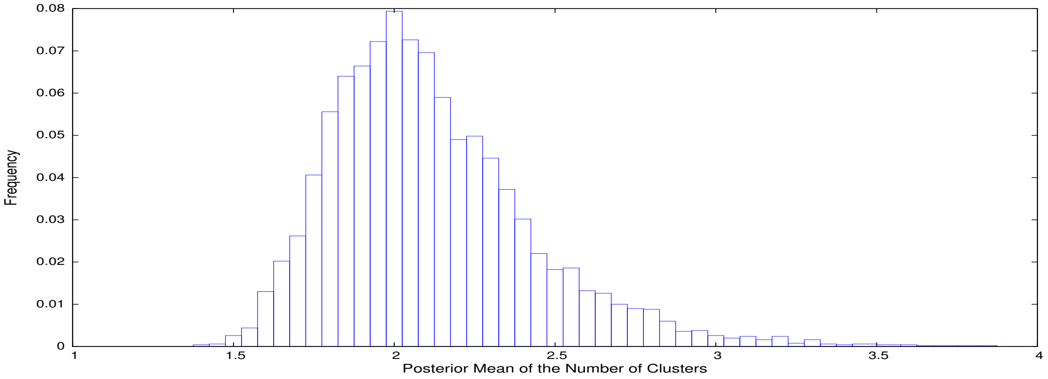

Figure 7: Posterior Mean of the Number of Clusters, K. Model: DPM. Data: 5-minute return without microstructure noise from SV1F.

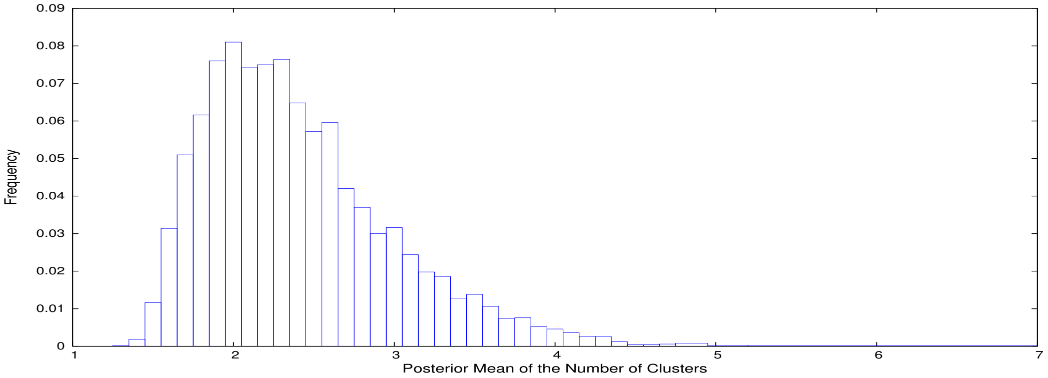

Figure 8: Posterior Mean of the Number of Clusters, K. Model: DPM-MA(1). Data: 1-minute return with independent noise from SV1FJ

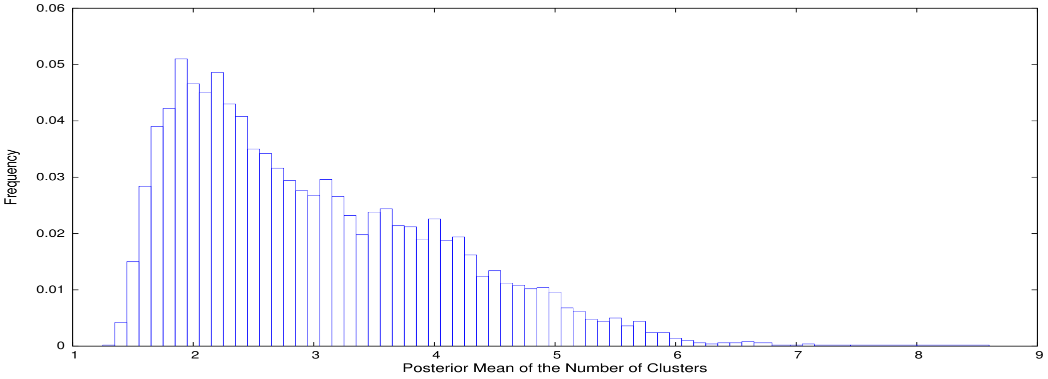

Figure 9: Posterior Mean of the Number of Clusters, K. Model: DPM-MA(2). Data: 30-second return with dependent noise from SV2F.

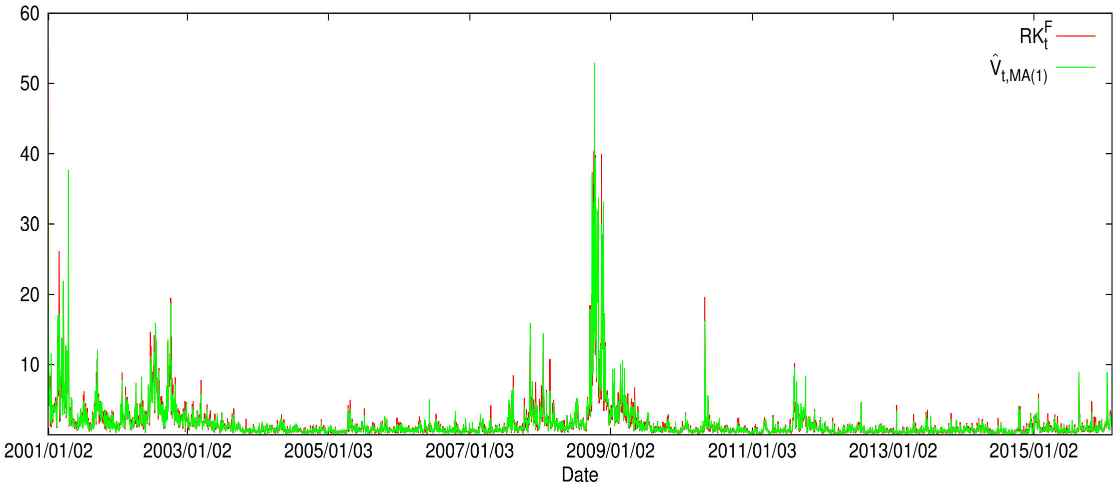

Figure 10: RK_t^F and $\hat{V}_{t,\text{MA}(1)}$ based on 5-minute IBM returns

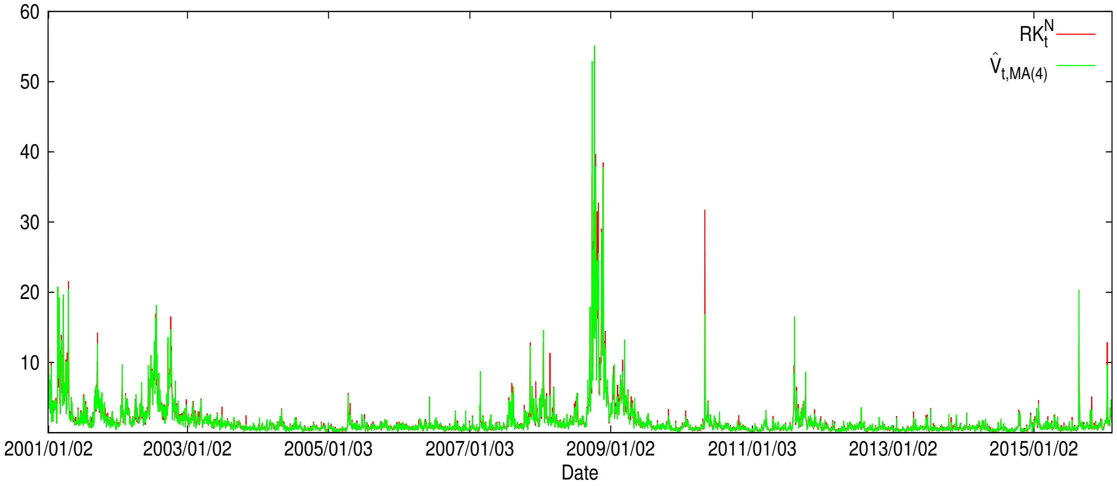

Figure 11: RK_t^N and $\hat{V}_{t,\text{MA}(4)}$ based on 1-minute IBM returns

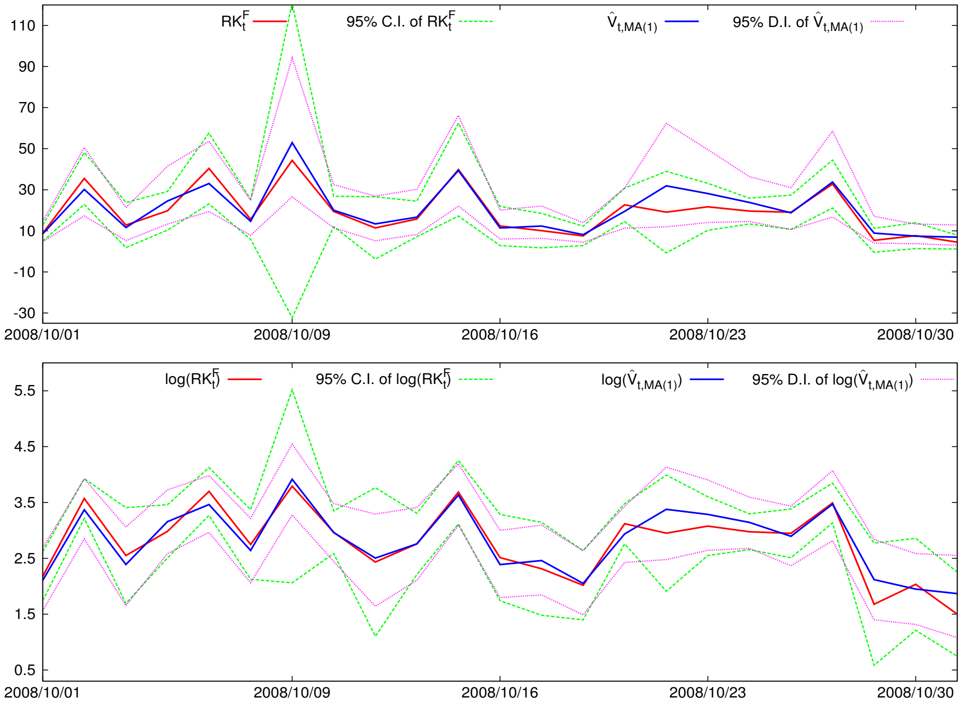

Figure 12: High volatility period: RK_t^F and $\hat{V}_{t,\text{MA}(1)}$ calculated using 5 minute IBM returns. Top: variance, below: log-variance

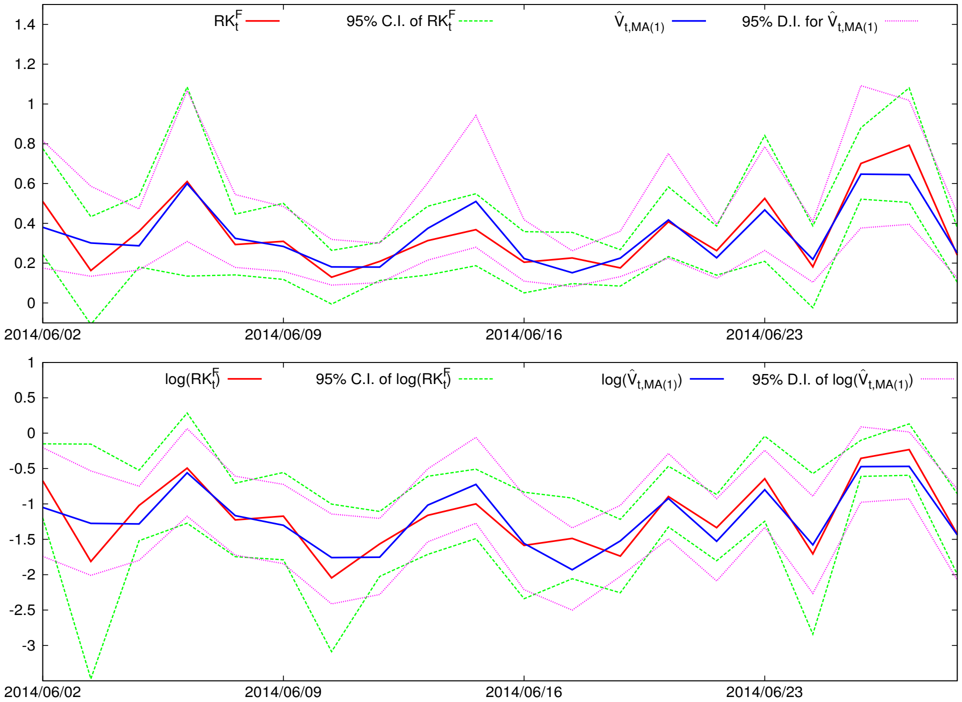

Figure 13: Low volatility period: RK_t^F and $\hat{V}_{t,\text{MA}(1)}$ calculated using 5 minute IBM returns. Top: variance, below: log-variance

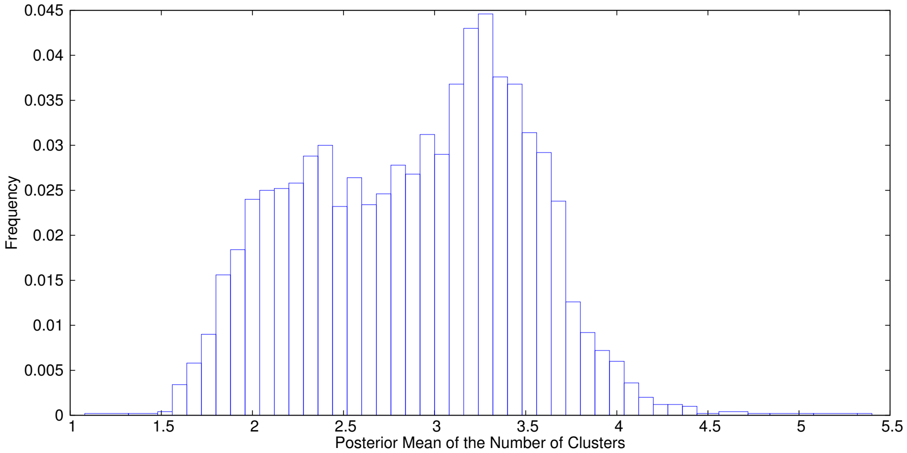

Figure 14: Posterior Mean of the Number of Clusters, K (Based on 3764 days results from DPM-MA(1) using 5-minute IBM returns).

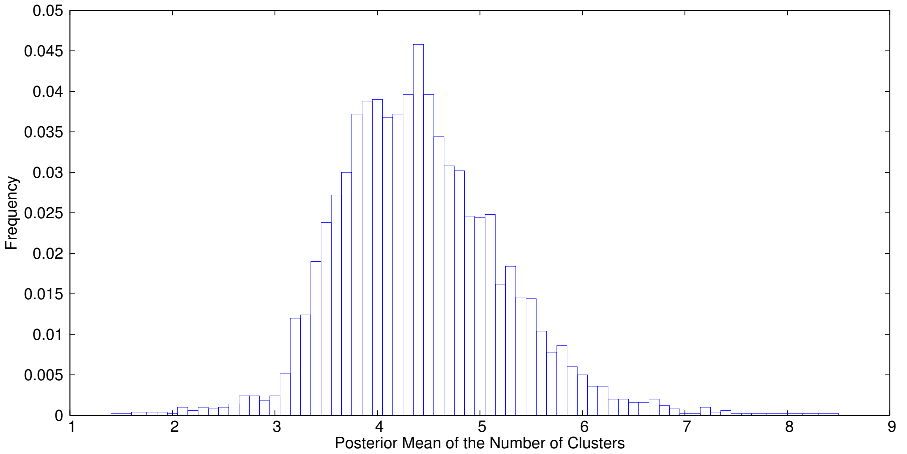

Figure 15: Posterior Mean of the Number of Clusters, K (Based on 3764 days results from DPM-MA(4) using 1-minute IBM returns).

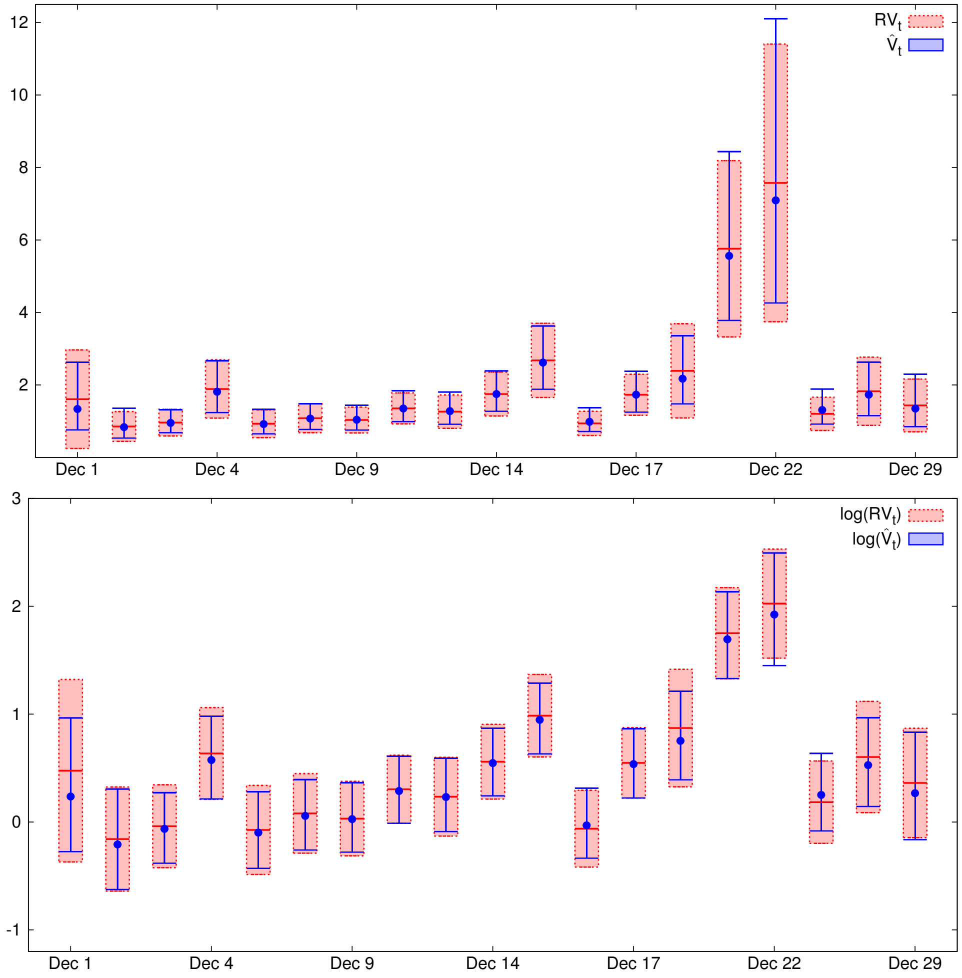

Figure 16: RV_t and $\hat{V}_t$ based on 5-minute Disney returns in December 2015. Top: variance, below: log-variance

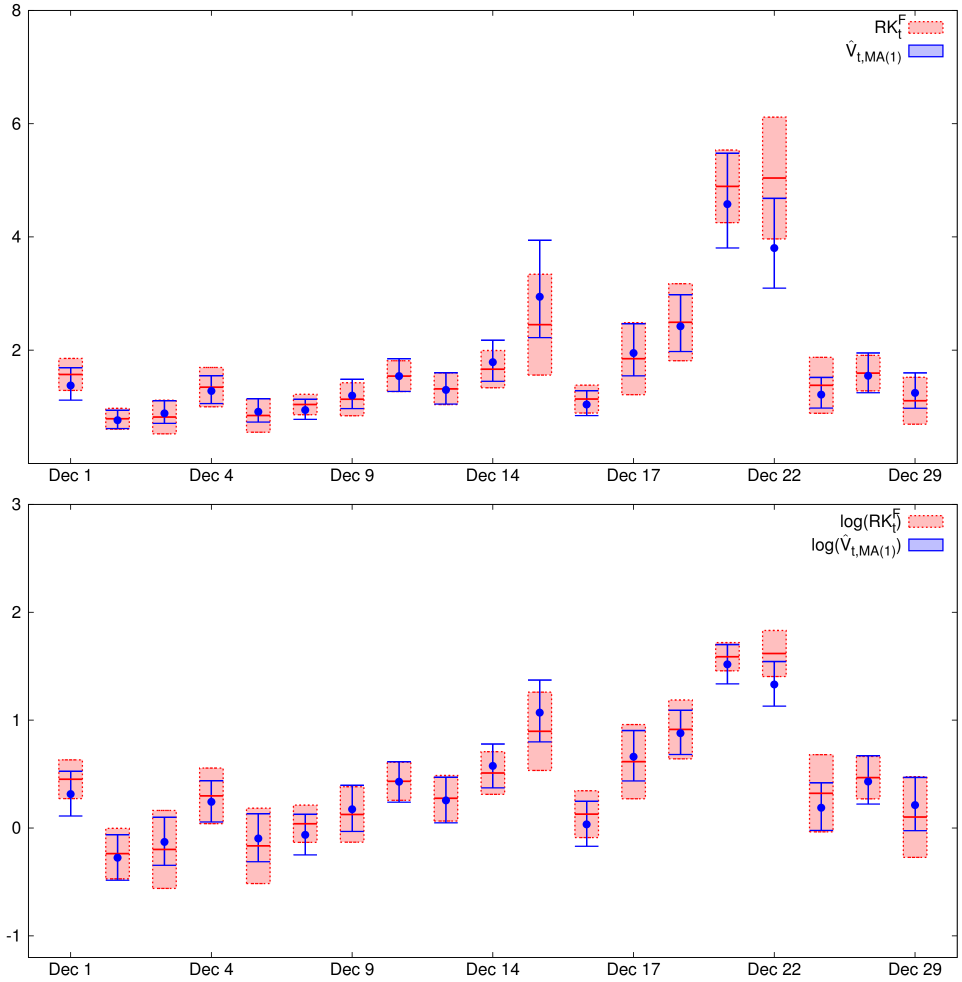

Figure 17: RK_t^F and $\hat{V}_{t,\mathrm{MA}(1)}$ based on 30-second Disney returns in December 2015. Top: variance, below: log-variance