

Bayesian stochastic volatility models for high-frequency data

Georgi Dinolov, Abel Rodriguez, and Hongyun Wang

Department of Applied Mathematics and Statistics, University of California, Santa Cruz, CA 95064, USA

Abstract

We formulate a discrete-time Bayesian stochastic volatility model for high-frequency stock-market data that directly accounts for microstructure noise, and outline a Markov chain Monte Carlo algorithm for parameter estimation. The methods described in this paper are designed to be coherent across all sampling timescales, with the goal of estimating the latent log-volatility signal from data collected at arbitrarily short sampling periods. In keeping with this goal, we carefully develop a method for eliciting priors. The empirical results derived from both simulated and real data show that directly accounting for microstructure in a state-space formulation allows for well-calibrated estimates of the log-volatility process driving prices.

Keywords: Discrete-time stochastic volatility; Intradaily estimation; Microstructure noise; Integrated variance

1 Introduction

Estimating asset price volatilities is a common problem in finance; for example, accurate estimates of volatility paths play a key role in both option pricing and portfolio design. Traditionally, financial models have used low-frequency returns (e.g., daily, weekly or monthly returns) to investigate price volatility. Early attempts at incorporating higher-frequency information focused on using intra-period maximum and minimum prices (e.g., see Alizadeh et al., 2002, Brandt and Diebold, 2003, and Chou et al., 2010). However, as high-frequency price data has become widely available, interest has turned to using all intra-period prices to generate high-resolution estimates of the volatility path and to improve estimates of the integrated volatility over higher frequencies.

A popular approach to estimating integrated volatilities from high-frequency price data is the realized variance estimator [Comte and Renault, 1998, Andersen et al., 2001, Barndorff-Nielsen and Shephard, 2002]. The realized variance is defined as the sum of squared high-frequency log returns over the period of interest. Under strict stationarity and some other weak regularity conditions for the volatility process, the realized variance converges in probability to the integrated variance of the true diffusion process as the sampling frequency increases. An alternative to the realized variance estimator is to use standard parametric models such as the Generalized Autoregressive Conditionally Heteroscedastic (GARCH) model [Bollerslev, 1986, Andersen and Bollerslev, 1997]. However, the assumptions behind the classical GARCH model are not robust with respect to the specification of the sampling interval, and therefore the model is not invariant to temporal aggregation [Drost and Nijman, 1993, Andersen and Bollerslev, 1997, Zumbach, 2000]. To address this issue some authors have turned to estimating low-frequency GARCH and stochastic volatility models using relevant summaries of the high-frequency prices. For example, Takahashi et al. [2009] and Shirota et al. [2014] use both high-frequency returns as well as the realized variance to estimate stochastic volatility models, while Hansen et al. [2012] does the same for GARCH models. Similarly, Maneeasoonthorn et al. [2014] use the realized volatility and the bipower variation estimators to estimate stochastic volatility models with jumps, while Bollerslev and Zhou [2002] use high order powers of the realized volatility as approximations to higher orders of integrated volatility.

A key challenge in working with high frequency prices is that they are often contaminated with microstructure noise. Indeed, as the sampling period shrinks down to the transaction-by-transaction frequency, irregular spacing between transactions, discreteness in transaction prices, and very short term temporal dependence become dominant features of the data [Stoll, 2000]. One consequence of the presence of microstructure noise is that the realized variance becomes a biased and inconsistent estimator of the true integrated variance [Zhou, 1996]. Possible solutions to this issue have been proposed by Zhang et al. [2005], who suggest sampling data sparsely at an optimally determined frequency and then averaging across the possible grids over the data, Ait-Sahalia et al. [2011], who propose combining estimators based on subsampling data at different frequencies, and Hansen and Lunde [2006] and Barndorff-Nielsen et al. [2008], who employ a class of kernel-based methods similar to those used for estimating the long-run variance of a stationary time-series in the presence of autocorrelation. In the

context of model-based approaches it is common to assume that the summaries of the high-frequency returns used to estimate the model are noisy versions of the true realized volatilities (e.g., see Venter and de Jongh, 2012, Shirota et al., 2014).

This paper describes a Bayesian stochastic volatility model for high-frequency data that explicitly accounts for the presence of microstructure noise. Unlike other approaches in the literature, we estimate our model directly using the high-frequency price data rather than summaries of the high-frequency returns. To account for the effect of microstructure noise we introduce a hierarchical specification in which the observed high-frequency prices are noisy versions of the true unknown prices. One appealing feature of our proposed model is that it is (approximately) coherent across all sampling frequencies, which is in line with previous efforts to validate the application of discrete-time models for volatility in high-frequency settings [Andersen et al., 1999]. Coherency is achieved by starting with a continuous-time model and then carefully discretizing the exact solution to the stochastic differential equations for the price and volatility processes, and by carefully eliciting prior distributions for the parameters of the continuous-time model.

The remainder of the paper is structured as follows: Section 2 describes the continuous- and discrete-time version of the model. Section 3 details the priors used and the method through which they were derived. Section 4 outlines the Bayesian Markov chain Monte Carlo (MCMC) algorithm used to fit the model. Section 5 examines the effect of certain model parameters on the posterior variance of the mean volatility level in our model. Finally, Section 6 includes simulation results demonstrating the robustness of our inferential procedure to microstructure noise.

2 Model Formulation

We begin with the continuous-time stochastic volatility model of Hull and White [1987], where the price \hat{S}_t of an asset follows a Geometric Brownian motion and the time-varying log-volatility process $\log(\hat{\sigma}_t)$ follows a mean-reverting Ornstein-Uhlenbeck (OU) process,

$$d \log(\hat{S}_t) = \hat{\mu} dt + \hat{\sigma}_t \sqrt{dt} \hat{\epsilon}_t, \quad (1)$$

$$d \log(\hat{\sigma}_t) = -\hat{\theta}(\log(\hat{\sigma}_t) - \hat{\alpha}) dt + \hat{\tau} \sqrt{dt} \hat{\epsilon}_{t,1}, \quad (2)$$

where $\hat{\epsilon}_t$ and $\hat{\epsilon}_{t,1}$ are dependent Wiener processes with instantaneous correlation ρ . This model not only allows for the volatility to evolve over time, but also captures leverage effects through the correlation between $\hat{\epsilon}_{t,1}$ and $\hat{\epsilon}_{t,2}$ (e.g., see Black, 1976). However, since at least the papers by Barndorff-Nielsen and Shephard [2001] and Chernov et al. [2003], it is understood that single-factor volatility models cannot reproduce the dependence structure of asset returns. Further, at high sampling frequencies, jumps have been shown to be an important part of asset price variance [Huang and Tauchen, 2005]. As a consequence of these results, we extend the model in (1)-(2) to include two factors for the volatility and jumps in the returns:

$$d \log(\hat{S}_t) = \hat{\mu} dt + \sqrt{\hat{\sigma}_{t,1} \hat{\sigma}_{t,2}} \sqrt{dt} \hat{\epsilon}_{t,1} + dJ_t, \quad (3)$$

$$d \log(\hat{\sigma}_{t,1}) = -\hat{\theta}_1(\log(\hat{\sigma}_{t,1}) - \hat{\alpha}) dt + \hat{\tau}_1 \sqrt{dt} \hat{\epsilon}_{t,1}, \quad (4)$$

$$d \log(\hat{\sigma}_{t,2}) = -\hat{\theta}_2(\log(\hat{\sigma}_{t,2}) - \hat{\alpha}) dt + \hat{\tau}_2 \sqrt{dt} \hat{\epsilon}_{t,2}. \quad (5)$$

The volatility exhibited by the log-prices in (3) is specified as a product of two OU processes. For the purposes of this paper, we will think of the volatility processes as distinct, $\hat{\sigma}_{t,1}$ being slow and $\hat{\sigma}_{t,2}$ being fast. By slow versus fast we mean that the mean-reversion timescale of (4) is greater than that of (5): $1/\hat{\theta}_1 > 1/\hat{\theta}_2$. We allow for leverage between the innovations of returns and the fast volatility, but the slow volatility and the price innovations are independent; the two volatility processes are also independent:

$$\mathbb{E}[\hat{\epsilon}_t \hat{\epsilon}_{t,2}] = \rho, \quad \mathbb{E}[\hat{\epsilon}_t \hat{\epsilon}_{t,1}] = 0, \quad \mathbb{E}[\hat{\epsilon}_{t,1} \hat{\epsilon}_{t,2}] = 0.$$

Finally, we model the jump process dJ_t as a compound Poisson process with constant jump intensity λ and i.i.d. jump size $Z_t \sim N(\mu_J, \sigma_J^2)$. In other words, over a finite interval Δ , $J(t + \Delta) - J(t) = \sum_{j=1}^{N(\Delta)} Z_{t_j}$, where

$$(t_j - t_{j-1}) \sim \text{Exp}(\lambda), \quad N(\Delta) \sim \text{Pois}(\lambda\Delta), \quad Z_{t_j} \sim N(\mu_J, \sigma_J^2), \quad t_j \in (t, t + \Delta). \quad (6)$$

We assume that jump sizes and arrival times are independent of either the price or volatility processes.

To generate a discretization of the model in (3) - (5), consider the (strong) solution of the Ornstein-Uhlenbeck process governing the evolution of the log-volatility in (2),

$$\log(\hat{\sigma}_t) \sim N \left(\hat{\alpha} + \exp \{ -\hat{\theta} t \} \{ \log(\hat{\sigma}_0) - \hat{\alpha} \}, \frac{\hat{\tau}^2}{2\hat{\theta}} \{ 1 - \exp(-2\hat{\theta} t) \} \right), \quad (7)$$

with stationary distribution

$$\log(\hat{\sigma}_t) \sim N\left(\hat{\alpha}, \frac{\hat{\tau}^2}{2\hat{\theta}}\right). \quad (8)$$

For an arbitrary time interval Δ (which, for the purpose of this paper, we measure in milliseconds) we can use (7) to generate the finite-difference equations

$$\log(S_j) = \log(S_{j-1}) + \mu(\Delta) + \sqrt{\sigma_{j,1}\sigma_{j,2}}\varepsilon_j + J_j(\Delta), \quad (9)$$

$$\log(\sigma_{j+1,1}) = \alpha(\Delta) + \theta_1(\Delta) \{\log(\sigma_{j,1}) - \alpha(\Delta)\} + \tau_1(\Delta) \varepsilon_{j,1}, \quad (10)$$

$$\log(\sigma_{j+1,2}) = \alpha(\Delta) + \theta_2(\Delta) \{\log(\sigma_{j,2}) - \alpha(\Delta)\} + \tau_2(\Delta) \varepsilon_{j,2}, \quad (11)$$

where $j = 0, 1, \dots, \lfloor T/\Delta \rfloor$, $i = 1, 2$, and

$$\sigma_{j+1,i} = \hat{\sigma}_{(j+1)\Delta,i} \sqrt{\Delta}, \quad S_j = \hat{S}_{j\Delta}, \quad J_j(\Delta) = J((j+1)\Delta) - J(j\Delta) \quad (12)$$

$$\alpha(\Delta) = \hat{\alpha} + \frac{1}{2} \log(\Delta), \quad \mu(\Delta) = \hat{\mu}\Delta, \quad \theta_i(\Delta) = \exp\{-\hat{\theta}_i\Delta\}, \quad \tau_i(\Delta) = \hat{\tau}_i \sqrt{\frac{1 - \exp\{-2\hat{\theta}_i\Delta\}}{2\hat{\theta}_i}}, \quad (13)$$

and

$$\begin{pmatrix} \varepsilon_j \\ \varepsilon_{j,1} \\ \varepsilon_{j,2} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 & \rho \\ 0 & 1 & 0 \\ \rho & 0 & 1 \end{pmatrix}\right).$$

We write $\alpha(\Delta)$, $\mu(\Delta)$, $\theta_i(\Delta)$, and $\tau_i(\Delta)$ to emphasize that we have a different set of parameters depending on the choice of Δ . The random variable $J_j(\Delta)$ is defined in (6)

Using the strong solution in (7) to derive the finite difference equations (9) - (11) allows us to take any step size Δ irrespective of the relative magnitude of the continuous-time model parameters. Indeed, the more standard forward-Euler discretization provides a poor approximation to the continuous-time model when $\Delta > 1/\hat{\theta}$, the timescale of inertia of the log-volatility process. Even in cases where $\Delta < 1/\hat{\theta}$, discretizing the model SDE using the exact solution of the OU process is a more accurate approximation of the finite-time transition density of the continuous process in (3) - (5) [Elerian et al., 2001]. The discretization (9) - (11) is a discrete-time approximation of the stochastic process in (3) - (5) and the likelihood based on this discretization introduces a bias when estimating continuous time paramters. Elerian et al. [2001] and Eraker [2001] both show that a way to correct for this bias (in the context of MCMC analysis) is to introduce latent, intra-interval subsamples of the process. We expect to achieve a similar, but better, result by decreasing Δ and accounting for microstructure noise. Moreover, the inverse of the transformations in (13) make it possible to meaningfully compare parameters inferred from different sampling frequencies and thereby check the coherency of our inferential procedure across different timescales.

In order to account for the effect of microstructure noise, we extend the previous model by differentiating between the *true* log asset price $\log(S_j)$ and the *discretely observed* log price $Y_j = \log(P_j)$. We treat these observed log prices as a noise-contaminated version of the true log price which is observed discretely only $n(\Delta) = \lfloor T/\Delta \rfloor$ times and whose index j corresponds to $j\Delta$ in continuous-time. More specifically, we let

$$Y_j = \log(S_j) + \zeta_j, \quad (14)$$

where ζ_1, ζ_2, \dots are independent and identically distributed errors with mean zero and standard deviation ξ . To motivate (14), consider one possible source of microstructure noise, the bid-ask spread. We can think of the idealized, “true” equilibrium price as evolving continuously in time by being moved by market supply and demand. Real-time order arrival and market friction makes it so that transaction prices are recorded at the highest bid or lowest ask levels only, thereby bounding the equilibrium market price in the bid-ask range. In this case it is natural to assume that $P_j = S_j + v_j$, where $v_j \sim U[-D_p/2, D_p/2]$ and D_p is the size of the bid-ask spread. Using a first-order Taylor approximation then leads to $Y_j = \log(P_j) \approx \log(S_j) + \zeta_j$, where $\zeta_j = \frac{1}{S_j} v_j$. A similar argument can be used to account for the effect of price discretization.

More generally, in order to account for the approximation error as well as for other sources of microstructure noise, we let $\zeta_t \sim N(0, \xi^2)$ where $\xi \approx \frac{D}{2Q}$, $D = \max\{D_p, D_s\}$, D_p represents a rough estimate of the bid-ask spread over the period of interest, D_s represents another possible source of microstructure noise (such as price truncation to the nearest cent), and Q is a rough guess of the average price of the asset over the period of interest. Note that the distribution of ξ is independent of the time scale Δ used for the discretization of the continuous-time process, and therefore independent of the frequency at which prices are observed.

To summarize, our hierarchical discrete-time stochastic volatility model reduces to

$$Y_j = \log(S_j) + \zeta_j, \quad (15)$$

$$\log(S_j) = \log(S_{j-1}) + \mu(\Delta) + \sqrt{\sigma_{j,1}\sigma_{j,2}}\varepsilon_j + J_j(\Delta), \quad (16)$$

$$\log(\sigma_{j+1,1}) = \alpha(\Delta) + \theta_1(\Delta) \{\log(\sigma_{j,1}) - \alpha(\Delta)\} + \tau_1(\Delta) \varepsilon_{j,1}, \quad (17)$$

$$\log(\sigma_{j+1,2}) = \alpha(\Delta) + \theta_2(\Delta) \{\log(\sigma_{j,2}) - \alpha(\Delta)\} + \tau_2(\Delta) \varepsilon_{j,2}, \quad (18)$$

where

$$\zeta_j \sim N(0, \xi^2), \quad \begin{pmatrix} \varepsilon_j \\ \varepsilon_{j,1} \\ \varepsilon_{j,2} \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 & \rho \\ 0 & 1 & 0 \\ \rho & 0 & 1 \end{pmatrix} \right),$$

and initial conditions

$$\log(\sigma_{0,i}) \sim N \left(\alpha, \frac{\tau_i(\Delta)^2}{1 - \theta_i(\Delta)^2} \right), \quad \log(S_0) \sim N(\eta, \kappa^2).$$

Table 1 summarizes our notation:

	Symbol	Interpretation
Data	P_j	Observed asset price at time $j\Delta$.
	Y_j	Logarithm of the observed asset price at time $j\Delta$.
Parameters	S_j	True asset price at time $j\Delta$.
	$\sigma_{j,i}$	Discrete-time approximation of the slow/fast volatility i of the true asset price at time $j\Delta$.
	$\alpha(\Delta)$	Discrete-time approximation of the stationary mean of the log volatility.
	$\mu(\Delta)$	Discrete-time approximation to the mean asset return.
	$\theta_i(\Delta)$	Discrete-time approximation to the autocorrelation associated with the log volatility i .
	$\tau_i(\Delta)$	Discrete-time approximation to the volatility of volatility i .
	ρ	Correlation coefficient between volatility and price innovations for the slow volatility process.
	ξ	Standard deviation associated with the microstructure noise.
	λ	Arrival rate of price jumps.
	μ_J	Expected value of price jump sizes.
	σ_J	Standard deviation of price jump sizes.
Hyper	Δ	Time step between observations (Fixed).
	η	Mean for the true asset price at time 0. (Fixed; no inference performed)
	κ	Standard deviation for the true asset at time 0. (Fixed; no inference performed)

Table 1: Notational summary for our high-frequency stochastic volatility model, including data, parameters and hyperparameters.

3 Prior Elicitation

We approach the problem of estimation and prediction for the model described above using Bayesian methods. This requires that we elicit priors for the unknown parameters ρ , ξ^2 , $\alpha(\Delta)$, $\mu(\Delta)$, $\theta(\Delta)$ and $\tau(\Delta)$. Eliciting a prior for the correlation parameter ρ and the microstructure variance ξ^2 is relatively straightforward since their value and interpretation are independent of the time step Δ . On the other hand, ensuring that the priors for $\alpha(\Delta)$, $\mu(\Delta)$, $\theta(\Delta)$ and $\tau(\Delta)$ are coherent across scales, i.e., that the priors provide the same information no matter what the time step Δ is, is non trivial. To address this problem we proceed to elicit priors on the continuous-time parameters $\hat{\alpha}$, $\hat{\mu}$, $\hat{\theta}$, $\hat{\tau}$ and then use the formulas in (13) to obtain the implied priors on $\alpha(\Delta)$, $\mu(\Delta)$, $\theta(\Delta)$ and $\tau(\Delta)$ for any time step Δ . Ideally, such priors would be invariant to the transformations in (13). However, fully invariant priors

are difficult to elicit and would, in any case, lead to computationally complicated models even using simulation-based methods such as Markov chain Monte Carlo algorithms. Hence, we settle for the more modest goal of assigning priors that belong to families that are conditionally conjugate and therefore lead to computationally tractable models, but whose first two moments are (approximately) coherent across scales.

Prior for ρ : We assign $(\rho + 1)/2$ a symmetric beta distribution with mean $1/2$ and precision c . This prior ensures that $\rho \in [-1, 1]$ as required and implies that $E(\rho) = 0$ a priori. Furthermore, for large values of c , this means that we believe a priori that the leverage effect is relatively small.

Prior for $\mu(\Delta)$: For the mean of the asset returns a prior in the normal family leads to a simple full conditional distribution for MCMC sampling. If we let $E[\hat{\mu}] = \hat{a}_{\hat{\mu}}$ and $\text{Var}[\hat{\mu}] = \hat{b}_{\hat{\mu}}^2$, then $\mu(\Delta) = \hat{\mu}\Delta$ leads to $E[\mu(\Delta)] = \Delta\hat{a}_{\hat{\mu}}$ and $\text{Var}[\mu(\Delta)] = \Delta^2\hat{b}_{\hat{\mu}}^2$. Hence, in our analysis we use the prior

$$\mu(\Delta) \sim N(\Delta\hat{a}_{\hat{\mu}}, \Delta^2\hat{b}_{\hat{\mu}}^2)$$

where values of $\hat{a}_{\hat{\mu}}$ and $\hat{b}_{\hat{\mu}}^2$ are elicited from historical data.

Prior for $\theta(\Delta)$: The discrete-time autocorrelation coefficient $\theta(\Delta)$ of the volatility process is bounded above by 1 and below by 0 such that $\log(\sigma_j)$ is bounded as $j \rightarrow \infty$. Hence, we employ a truncated normal prior for $\theta(\Delta)$,

$$p(\theta(\Delta)) \propto N(a_{\theta}(\Delta), b_{\theta}^2(\Delta)) \mathbb{1}(\theta(\Delta) \in [0, 1]),$$

which leads again to a tractable computational algorithm. Note that because of the truncation,

$$E[\theta(\Delta)] = a_{\theta}(\Delta) + \frac{\phi\left(-\frac{a_{\theta}(\Delta)}{b_{\theta}(\Delta)}\right) - \phi\left(\frac{1-a_{\theta}(\Delta)}{b_{\theta}(\Delta)}\right)}{\Phi\left(\frac{1-a_{\theta}(\Delta)}{b_{\theta}(\Delta)}\right) - \Phi\left(-\frac{a_{\theta}(\Delta)}{b_{\theta}(\Delta)}\right)} b_{\theta}(\Delta) \quad (19)$$

$$\begin{aligned} \text{Var}[\theta(\Delta)] = b_{\theta}^2(\Delta) & \left[1 + \frac{-\frac{a_{\theta}(\Delta)}{b_{\theta}(\Delta)}\phi\left(-\frac{a_{\theta}(\Delta)}{b_{\theta}(\Delta)}\right) - \frac{1-a_{\theta}(\Delta)}{b_{\theta}(\Delta)}\phi\left(\frac{1-a_{\theta}(\Delta)}{b_{\theta}(\Delta)}\right)}{\Phi\left(\frac{1-a_{\theta}(\Delta)}{b_{\theta}(\Delta)}\right) - \Phi\left(-\frac{a_{\theta}(\Delta)}{b_{\theta}(\Delta)}\right)} \right. \\ & \left. + \left\{ \frac{\phi\left(-\frac{a_{\theta}(\Delta)}{b_{\theta}(\Delta)}\right) - \phi\left(\frac{1-a_{\theta}(\Delta)}{b_{\theta}(\Delta)}\right)}{\Phi\left(\frac{1-a_{\theta}(\Delta)}{b_{\theta}(\Delta)}\right) - \Phi\left(-\frac{a_{\theta}(\Delta)}{b_{\theta}(\Delta)}\right)} \right\}^2 \right] \quad (20) \end{aligned}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ denote the density and the cumulative distribution functions of the standard normal distribution. Now, given the prior mean $\hat{a}_{\hat{\theta}}$ and variance $\hat{b}_{\hat{\theta}}^2$ for $\hat{\theta}$, we choose the values of $a_{\theta}(\Delta)$ and $b_{\theta}(\Delta)$ so that the mean and variance of $\theta(\Delta)$ above are approximately equal to the mean and variance of $\exp\{-\hat{\theta}\Delta\}$. To simplify calculation of the moments of $\exp\{-\hat{\theta}\Delta\}$ we use a second-order Taylor expansion of $\exp\{-\hat{\theta}\Delta\}$ to approximate the first two moments of $\theta(\Delta)$ in terms of $\hat{a}_{\hat{\theta}}$ and $\hat{b}_{\hat{\theta}}^2$, an approach known as the Delta-Method (e.g., see Casella and Berger, 2002):

$$E[\exp\{-\hat{\theta}\Delta\}] \approx \exp(-\hat{a}_{\hat{\theta}}\Delta) \left(1 + \frac{1}{2}\hat{b}_{\hat{\theta}}^2\Delta^2\right), \quad (21)$$

$$E[\exp\{-2\hat{\theta}\Delta\}] \approx \exp(-2\hat{a}_{\hat{\theta}}\Delta) \left(1 + 2\hat{b}_{\hat{\theta}}^2\Delta^2\right). \quad (22)$$

Using (19), (20), (21), and (22), and by setting $E[\theta(\Delta)] = E[\exp\{-\hat{\theta}\Delta\}]$ and $\text{Var}[\theta(\Delta)] = \text{Var}[\exp\{-\hat{\theta}\Delta\}]$, we obtain a system of two equations with two unknowns that can be solved numerically to find the values of $a_{\theta}(\Delta)$ and $b_{\theta}^2(\Delta)$ in terms of $\hat{a}_{\hat{\theta}}$, $\hat{b}_{\hat{\theta}}^2$, and Δ .

To elicit $\hat{a}_{\hat{\theta}}$ and $\hat{b}_{\hat{\theta}}^2$, recall that $\hat{\theta}$ is the inverse of the time scale of inertia for $\log(\hat{\sigma}_t)$ in the continuous-time formulation, which can be thought of as the characteristic time length, or unit of time, over which the process for the diffusion of $\log(\hat{\sigma}_t)$ “forgets” about an endogenous shock. The two hyper-parameters can be chosen so that the prior probability mass for $\hat{\theta}$ permits a reasonable range for the timescale of inertia.

Prior for $\alpha(\Delta)$: For the mean log-volatility level $\hat{\alpha}$, we once again use a computationally convenient prior in the normal family. Letting $E[\hat{\alpha}] = \hat{a}_{\hat{\alpha}}$ and $\text{Var}[\hat{\alpha}] = \hat{b}_{\hat{\alpha}}^2$, and recalling that $\alpha(\Delta) = \hat{\alpha} + \frac{1}{2}\log(\Delta)$, we have

$$\alpha(\Delta) \sim N\left(\hat{a}_{\hat{\alpha}} + \frac{1}{2}\log(\Delta), \hat{b}_{\hat{\alpha}}^2\right).$$

To elicit the values of $\hat{a}_{\hat{\alpha}}$ and $\hat{b}_{\hat{\alpha}}^2$, recall that $\hat{\alpha}$ is the stationary (long-term) median of the volatility process. Hence, for most assets these parameters could be elicited by looking at the time series of the asset’s implied volatility (e.g., the VIX index if the asset is the S&P500 index).

Prior for $\tau^2(\Delta)$: We use a prior in the Inverse-Gamma family for $\tau^2(\Delta)$, so that

$$\tau^2(\Delta) \sim \text{Inv-Gamma}(a_{\tau^2}(\Delta), b_{\tau^2}(\Delta)).$$

To find the values of $a_{\tau^2}(\Delta)$ and $b_{\tau^2}(\Delta)$ recall that $\tau^2(\Delta) = \hat{\tau}^2 (1 - \exp\{-2\hat{\theta}\Delta\}) / (2\hat{\theta})$. If we let $E[\hat{\tau}^2] = \hat{a}_{\tau^2}$ and $\text{Var}[\hat{\tau}^2] = \hat{b}_{\tau^2}^2$, and if we use the prior mean and variance of $\hat{\theta}$ as before, we can again apply the Delta-Method to approximate the prior first and second moments of $\tau^2(\Delta)$ by performing a second-order Taylor expansion of $\tau^2(\Delta)$ and $(\tau^2(\Delta))^2$ about the prior means of $\hat{\tau}^2$ and $\hat{\theta}$, leading to

$$\begin{aligned} E\left[\hat{\tau}^2 \left(\frac{1 - \exp\{-2\hat{\theta}\Delta\}}{2\hat{\theta}}\right)\right] &\approx \hat{\tau}^2 \left(\frac{1 - \exp\{-2\hat{\theta}\Delta\}}{2\hat{\theta}}\right) \Big|_{\hat{\tau}^2=\hat{a}_{\tau^2}, \hat{\theta}=\hat{a}_{\hat{\theta}}} \\ &\quad + \frac{1}{2} \hat{b}_{\tau^2}^2 \frac{\partial^2}{\partial^2 \hat{\tau}^2} \left[\hat{\tau}^2 \left(\frac{1 - \exp\{-2\hat{\theta}\Delta\}}{2\hat{\theta}}\right)\right] \Big|_{\hat{\tau}^2=\hat{a}_{\tau^2}, \hat{\theta}=\hat{a}_{\hat{\theta}}} \\ &\quad + \frac{1}{2} \hat{b}_{\hat{\theta}}^2 \frac{\partial^2}{\partial^2 \hat{\theta}} \left[\hat{\tau}^2 \left(\frac{1 - \exp\{-2\hat{\theta}\Delta\}}{2\hat{\theta}}\right)\right] \Big|_{\hat{\tau}^2=\hat{a}_{\tau^2}, \hat{\theta}=\hat{a}_{\hat{\theta}}} \end{aligned} \quad (23)$$

and

$$\begin{aligned} E\left[\left\{\hat{\tau}^2 \left(\frac{1 - \exp\{-2\hat{\theta}\Delta\}}{2\hat{\theta}}\right)\right\}^2\right] &\approx \left\{\hat{\tau}^2 \left(\frac{1 - \exp\{-2\hat{\theta}\Delta\}}{2\hat{\theta}}\right)\right\}^2 \Big|_{\hat{\tau}^2=\hat{a}_{\tau^2}, \hat{\theta}=\hat{a}_{\hat{\theta}}} \\ &\quad + \frac{1}{2} \hat{b}_{\tau^2}^2 \frac{\partial^2}{\partial^2 \hat{\tau}^2} \left\{\hat{\tau}^2 \left(\frac{1 - \exp\{-2\hat{\theta}\Delta\}}{2\hat{\theta}}\right)\right\}^2 \Big|_{\hat{\tau}^2=\hat{a}_{\tau^2}, \hat{\theta}=\hat{a}_{\hat{\theta}}} \\ &\quad + \frac{1}{2} \hat{b}_{\hat{\theta}}^2 \frac{\partial^2}{\partial^2 \hat{\theta}} \left\{\hat{\tau}^2 \left(\frac{1 - \exp\{-2\hat{\theta}\Delta\}}{2\hat{\theta}}\right)\right\}^2 \Big|_{\hat{\tau}^2=\hat{a}_{\tau^2}, \hat{\theta}=\hat{a}_{\hat{\theta}}} \end{aligned} \quad (24)$$

The right sides of (23) and (24) are functions of $a_{\tau^2}(\Delta)$ and $b_{\tau^2}(\Delta)$, so that the above system of equations can be solved numerically to find $a_{\tau^2}(\Delta)$ and $b_{\tau^2}(\Delta)$ in terms of the other known prior hyperparameters. Finally, to elicit \hat{a}_{τ^2} and \hat{b}_{τ^2} , we have to recall that the ratio $\frac{\hat{\tau}^2}{2\hat{\theta}}$ represents the long-run variance of the log-volatility process $\log(\hat{\sigma})$. The prior for $\hat{\tau}^2$ can therefore be elicited from market-traded approximations of the volatility process, such as the VIX.

Prior for ξ^2 : For computational convenience, the variance of the microstructure noise is assigned an inverse Gamma with shape parameter a_{ξ} and rate parameter b_{ξ} . The mean of the prior can be elicited from information about the bid-ask spread, the tick size and the average price of the stock as discussed in Section 2, while its standard deviation is selected so that we stay within an order of magnitude (above and below) of the mean.

Prior for $(\lambda, \mu_J, \sigma_J^2)$: We assign a $\text{Gamma}(a_{\lambda}, b_{\lambda})$ prior for the arrival rate of jumps, and a Normal-Inverse-gamma prior for the expected value and variance of the jump sizes with hyperparameters $(m_J, 1, a_J, b_J)$. The hyperparameters for the arrival rate of jumps are elicited from weekly dynamics, assuming an average of once-per-week jump and a variance an order to magnitude greater:

$$E[\lambda \Delta^*] = \Delta^* \frac{a_{\lambda}}{b_{\lambda}} = 1, \quad \text{Var}[\lambda \Delta^*] = \Delta^{*2} \frac{a_{\lambda}}{b_{\lambda}^2} = 10, \quad \Delta^* = 1 \text{ week (milliseconds)}. \quad (25)$$

The jump size hyperparameters are elicited from historical data of the S&P 500 index. In particular, m_J is the average observed weekend jump size from 1990 to 2015 on the S&P 500, whereas $b_J/(a_J - 1)$ is the observed variance of the said jumps. Finally, we set $a_J = 1.1$ so that the prior admits a valid expectation for the jump variance σ_J^2 but is minimally informative (equivalent to 1.1 data points in the prior).

4 Computation

The posterior distribution of high-frequency stochastic volatility is analytically intractable, so we perform parameter inference and prediction using a Markov chain Monte Carlo (MCMC) algorithm. Our sampler extends

Component	p_l	m_l	v_l^2
1	0.00609	1.92677	0.11265
2	0.04775	1.34744	0.17788
3	0.13057	0.73504	0.26768
4	0.20674	0.02266	0.40611
5	0.22715	-0.85173	0.62699
6	0.18842	-1.97278	0.98583
7	0.12047	-3.46788	1.57469
8	0.05591	-5.55246	2.54498
9	0.01575	-8.68384	4.16591
10	0.00115	-14.65000	7.33342

Table 2: Parameters of the mixture representation of the log Chi-squared distribution, provided in Omori et al. [2007].

the ideas introduced in Omori et al. [2007], which used a mixture of normals approximation to the distribution of a log χ^2 distribution. More specifically, our algorithm alternates between 1) integrating out the true latent prices with the forward component of the Forward-Backward algorithm [Carter and Kohn, 1994, Frühwirth-Schnatter, 1994] and sampling $(\rho, \xi, \mu(\Delta))$ jointly given all other parameters 2) sampling the true asset prices $S_0, S_1, \dots, S_{n(\Delta)}$ from their joint full conditional distribution using the Forward-Backward algorithm, 3) jointly sampling the mixture indicators $\gamma_1, \dots, \gamma_{n(\Delta)}$ (to be introduced below) given all other parameters, 4) given the mixture indicators, integrating out the volatilities using the forward component of the Forward-Backward algorithm and jointly sampling all of the volatility parameters except ρ , 5) jointly sampling the volatilities $(\sigma_{1,1}, \sigma_{1,2}), \dots, (\sigma_{n(\Delta),1}, \sigma_{n(\Delta),2}), (\sigma_{n(\Delta)+1,1}, \sigma_{n(\Delta)+1,2})$ using a second Forward-Backward algorithm.

For Steps 1 and 2 in our inferential procedure, note that, given the mean return $\mu(\Delta)$, the microstructure variance ξ^2 and the volatilities $(\sigma_{1,1}, \sigma_{1,2}), \dots, (\sigma_{n(\Delta),1}, \sigma_{n(\Delta),2}), (\sigma_{n(\Delta)+1,1}, \sigma_{n(\Delta)+1,2})$, equations (15) and (16) - (17) define a linear state-space model with state variable $x_j = \log(S_j)$ and Gaussian innovations. Hence, using a Forward-Backward algorithm to sample the true asset prices is straightforward. For Steps 4 and 5, we note that

$$\log \{ |\log(S_j/S_{j-1}) - \mu(\Delta) - J_j(\Delta)| \} = \frac{1}{2} \log(\sigma_{j,1}) + \frac{1}{2} \log(\sigma_{j,2}) + \log(\epsilon_{j,1}^2)/2.$$

Following Omori et al. [2007] we approximate the error term using a mixture of Gaussian distributions,

$$\log(\epsilon_{j,1}^2)/2 \sim \sum_{l=1}^{10} p_l N\left(\frac{m_l}{2}, \frac{v_l^2}{4}\right)$$

(see Table 2 for the values of $\{p_l\}$, $\{m_l\}$ and $\{v_l\}$). The mixture can be rewritten by introducing auxiliary indicators $\gamma_1, \dots, \gamma_{n(\Delta)}$ such that

$$\log(\epsilon_{j,1}^2)/2 \mid \gamma_j \sim N\left(\frac{m_{\gamma_j}}{2}, \frac{v_{\gamma_j}^2}{4}\right), \quad \Pr(\gamma_k = l) = p_l.$$

The auxiliary indicators are sampled jointly conditional on all other parameters, and $\Pr(\gamma_k = l) = p_l$ is interpreted as the prior probability that observation k belongs to mixture element l .

Conditionally on the true prices, the indicators $\gamma_1, \dots, \gamma_{n(\Delta)}$, we have again a linear state-space model with Gaussian innovations so that we can integrate out the volatilities in the posterior and sample all of the volatility parameters except for ρ to complete Step 3. Conditional on this sample, the volatilities can be sampled using another Forward-Backward algorithm for Step 4. Details of the algorithm are given in Appendix 7.

Once the algorithm has converged and the burn-in samples have been discarded, point and interval estimates can be easily obtained using empirical estimates. For example, given a sample of the volatility path $((\sigma_{1,1}^{(b)}, \sigma_{1,2}^{(b)}), (\sigma_{2,1}^{(b)}, \sigma_{2,2}^{(b)}), \dots, (\sigma_{n(\Delta),1}^{(b)}, \sigma_{n(\Delta),2}^{(b)}))$ for $b = 1, \dots, B$, a sample of the in-sample (approximate) integrated variance $IV = \int_0^T \hat{\sigma}_t^2 dt$ can be obtained as

$$IV^{(b)} \approx \sum_{j=1}^{n(\Delta)} (\sigma_{j,1}^{(b)} \sigma_{j,2}^{(b)}).$$

A similar approach can be used to make out-of-sample predictions of the integrated volatility.

5 The effect of the mean-reverting rate $\hat{\theta}$ and the observational duration on the posterior variance of the mean log-volatility $\hat{\alpha}$

When estimating model parameters, the common intuition is that an increase in sample size leads to a decrease in posterior uncertainty. When dealing with the estimation of stochastic volatility models for high-frequency data, one may apply this thinking when the sample size is increased by obtaining more frequent price path samples for a fixed observational period. However, in the case where the volatility process has a finite non-zero mean-reversion timescale (as is the case for the Ornstein-Uhlenbeck processes used here), an increase in the number of intraperiod observations does not add information about the mean-level of the process. Rather, the posterior uncertainty for this model parameter can only be decreased by increasing how long we observe the process.

To demonstrate this feature of the model, we study analytically the relationship between the mean-reverting rate $\hat{\theta}$, the time duration of observation T , and the posterior variance of mean log-volatility $\hat{\alpha}$ for a single volatility process. To proceed analytically, we consider a simplified inference problem described by the following assumptions: 1) the mean log-volatility $\hat{\alpha}$ is the only parameter to be inferred – all other parameters are known; 2) the prior distribution for the mean log-volatility $\hat{\alpha}$ is normal and is denoted by $N(\hat{a}_{\hat{\alpha}}, \hat{b}_{\hat{\alpha}}^2)$ (previously there was no parametric assumption made on the prior for $\hat{\alpha}$); c) the log-volatility $\log(\hat{\sigma}_t)$ is observed exactly (without error) on a uniform grid $\{0, \Delta, 2\Delta, \dots, N\Delta\}$ in time duration $[0, T]$ where Δ is the sampling period and $N = T/\Delta$.

The exact solution of the Ornstein-Uhlenbeck process (2) is given in (7). Applying the exact solution (7) to the time interval $[j\Delta, (j+1)\Delta]$, we obtain

$$\log(\hat{\sigma}_{(j+1)\Delta}) = \theta(\Delta) \log(\hat{\sigma}_{j\Delta}) + (1 - \theta(\Delta))\hat{\alpha} + \tau(\Delta)\epsilon_j, \quad 0 \leq j \leq N-1$$

where $\epsilon_j \sim N(0, 1)$, and $\theta(\Delta)$ and $\tau(\Delta)$ are given in (13). Recall the stationary distribution of the continuous-time log-volatility process in (8)

$$\log(\hat{\sigma}_t) \sim N(\hat{\alpha}, \tau(\infty)^2), \quad \tau(\infty)^2 = \frac{\hat{\tau}^2}{2\hat{\theta}}.$$

Here we denote the stationary variance as $\tau(\infty)^2$ for mathematical convenience. The likelihood of $\hat{\alpha}$ given the observation $\{\log(\hat{\sigma}_0), \log(\hat{\sigma}_1), \dots, \log(\hat{\sigma}_N)\}$ is

$$L(\hat{\alpha} | \log(\hat{\sigma}_0), \log(\hat{\sigma}_1), \dots, \log(\hat{\sigma}_N)) \propto \exp\left(\frac{-(\log(\hat{\sigma}_0) - \hat{\alpha})^2}{2\tau(\infty)^2}\right) \times \prod_{j=0}^{N-1} \exp\left(\frac{-(\log(\hat{\sigma}_{(j+1)\Delta}) - \theta(\Delta)\log(\hat{\sigma}_{j\Delta}) - (1 - \theta(\Delta))\hat{\alpha})^2}{2\tau(\Delta)^2}\right). \quad (26)$$

Since we assume $N(\hat{a}_{\hat{\alpha}}, \hat{b}_{\hat{\alpha}}^2)$ as the prior for $\hat{\alpha}$, the posterior distribution of $\hat{\alpha}$ is normal and the reciprocal of the posterior variance of $\hat{\alpha}$ has the expression

$$\frac{1}{\text{Var}[\hat{\alpha}]} = \frac{1}{\hat{b}_{\hat{\alpha}}^2} + \frac{1}{\tau(\infty)^2} + \sum_{j=0}^{N-1} \frac{(1 - \theta(\Delta))^2}{\tau(\Delta)^2} = \frac{1}{\hat{b}_{\hat{\alpha}}^2} + \frac{2\hat{\theta}}{\hat{\tau}^2} \left(1 + N \cdot \tanh\left(\frac{\hat{\theta}\Delta}{2}\right)\right). \quad (27)$$

In the above, we have used the expressions of $\tau(\Delta)$ and $\theta(\Delta)$ given in (13).

Now, using the linear approximation $\tanh\left(\frac{x}{2}\right) \approx \frac{x}{2}$ and setting $T = N\Delta$, we can write (27) as

$$\frac{1}{\text{Var}[\hat{\alpha}]} \approx \frac{1}{\hat{b}_{\hat{\alpha}}^2} + \frac{2\hat{\theta}}{\hat{\tau}^2} \left(1 + \frac{\hat{\theta}T}{2}\right). \quad (28)$$

This expression is valid for $\hat{\theta}\Delta \leq 1$, i.e. when the timescale of inertia of the log-volatility process is *greater* than the spacing between observations. The important consequence of (28) is that decreasing Δ does not decrease the posterior variance of $\hat{\alpha}$. In other words, an increase in the number of intraperiod observations does not add information about $\hat{\alpha}$. Rather, the posterior uncertainty for $\hat{\alpha}$ can only be decreased by increasing T (increasing how long we observe the process) or increasing $\hat{\theta}$ (on average, increasing the number of reversions to the mean). The rate of information increase for $\hat{\alpha}$ with respect to T and $\hat{\theta}$ is examined under two conditions.

When $\hat{\tau}$ is fixed, $1/\text{Var}[\hat{\alpha}]$ increases linearly with the time duration T and increases quadratically with the mean-reverting rate $\hat{\theta}$. The quadratic increase of $1/\text{Var}[\hat{\alpha}]$ with respect to $\hat{\theta}$ is the combined result from two contributions: i) for larger $\hat{\theta}$, the variance of $\log(\hat{\sigma}_t)$ is smaller and consequently each data point is a more accurate approximation to $\hat{\alpha}$; and ii) for larger $\hat{\theta}$, the time duration $[0, T]$ covers more rounds of $\log(\hat{\sigma}_t)$ fluctuating away from $\hat{\alpha}$ and relaxing back toward $\hat{\alpha}$.

When $\frac{\hat{\tau}^2}{2\hat{\theta}}$ (the stationary variance of log-volatility) is fixed, $1/\text{Var}[\hat{\alpha}]$ increases linearly with $\hat{\theta}T$. In this case, if the prior is wider than the stationary distribution ($\hat{b}_\alpha^2 \geq \frac{\hat{\tau}^2}{2\hat{\theta}}$) and the time duration is much larger than the time scale of inertia ($T \gg 1/\hat{\theta}$), then the posterior variance of $\hat{\alpha}$ is inversely proportional to the time duration:

$$\text{Var}[\hat{\alpha}] \approx \frac{\hat{\tau}^2}{\hat{\theta}} \cdot \frac{1}{\hat{\theta}T}. \quad (29)$$

When $\hat{\theta}\Delta \gg 1$ (i.e. when the timescale of inertia of the log-volatility process is much *smaller* than the spacing between observations), the linear approximation $\tanh(\hat{\theta}\Delta/2) \approx \hat{\theta}\Delta/2$ is invalid. Instead, we have the approximation $\tanh(\hat{\theta}\Delta/2) \approx 1$, which leads to

$$\frac{1}{\text{Var}[\hat{\alpha}]} \approx \frac{1}{\hat{b}_\alpha^2} + \frac{2\hat{\theta}}{\hat{\tau}^2}(1+N), \quad \text{for } \hat{\theta}\Delta \gg 1$$

Under this regime, $1/\text{Var}[\hat{\alpha}]$ is approximately proportional to the number of observations, N , provided that $\hat{\theta}\Delta \gg 1$ is preserved as N is increased. This occurs when the spacing between observations, Δ , is fixed and the increase in N comes from extending the observational duration T . When T is fixed, as N increases Δ decreases, which eventually will carry the system from the regime of $\hat{\theta}\Delta \gg 1$ to that of $\hat{\theta}\Delta \leq 1$. The behavior of the posterior variance of $\hat{\alpha}$ for $\hat{\theta}\Delta \leq 1$ with N increasing through either lowering Δ or increasing T is illustrated in Section 6.4 below.

The results above, which indicate that $\hat{\alpha}$ cannot be estimated consistently under in-fill asymptotics, apply to any stochastic volatility model based on the Ornstein-Uhlenbeck process. Note, however, that they do not contradict standard asymptotic results from the realized volatility literature, which focus on the integrated variance during a finite period of time and not on the long-term median volatility of the process.

6 Illustrations

6.1 Effect of microstructure noise and sampling frequency on estimates: simulation studies

In this section we examine the effects of microstructure noise and sampling frequency on our inference of the model parameters. We first consider a one-day (6.5 hours) simulated dataset in which the true log-prices were generated according to (16) - (17) with $\Delta = 10^{-3}$ seconds, but where the microstructure noise was incorporated by adding to each point (exponentiated to transform from log-price to price level) a uniformly distributed random variable between -0.05 and 0.05 , simulating a \$0.1 bid-ask spread. The price was rounded to the nearest 100th, then transformed back to the log scale.

The true parameters used in the simulations were set to be reasonably close to typical values on the S&P500 market. The instantaneous return per millisecond $\hat{\mu}$ was set to $\hat{\mu} = 1.7 \cdot 10^{-12}$ /millisecond, corresponding to an annual return of 1%, based on 251 trading days per year, 6.5 trading hours per trading day, excluding jumps between trading sessions. Assuming a characteristic timescale of inertia of the slow continuous log-volatility process to be 3.5 hours (measured in milliseconds) and that of the fast log-volatility process to be 10 minutes, $\hat{\theta}_1$ and $\hat{\theta}_2$ were set to $\hat{\theta}_1 = 7.94 \cdot 10^{-8}$ /milliseconds and $\hat{\theta}_2 = 1.57 \cdot 10^{-6}$ /milliseconds. The remaining parameters $\hat{\tau}_i^2$ and $\hat{\alpha}$ governing the behavior of the log-volatility processes were set using the publicly traded VIX index. The VIX is the square root of the risk-neutral market expectation of the S&P 500 variance over the next 30 days on an annualized scale, such that

$$\log \hat{\sigma}_t \approx \log(\log(1 + \text{VIX}_t/100)) - \frac{1}{2} \log(T_{\text{year}})$$

where $T_{\text{year}} = 1 \text{ year} = 251 \times 6.5 \times 3.6 \cdot 10^6 \text{ ms}$. We can transform between VIX_t and $\log(\hat{\sigma}_t)$, obtaining approximations of the historical log-volatility path. With $\hat{\theta}_i$ set, the parameters $\hat{\alpha}$ and $\hat{\tau}_i^2$ determine the stationary distribution of the log-volatility process, $\log(\hat{\sigma}_{t,i}) \sim N\left(\hat{\alpha}, \frac{\hat{\tau}_i^2}{2\hat{\theta}_i}\right)$. Thus we set $\hat{\alpha}$ to be the mean of this VIX-derived log-volatility path, and $\hat{\tau}_i^2$ the variance thereof, multiplied by $2\hat{\theta}_i$. A summary of the true model parameters used in this simulation is presented in Table 3.

We fit our model to the simulated data using priors whose means equal the true simulation values and whose standard deviations are roughly one order of magnitude higher than the prior mean (recall that the prior specification procedure described in Section 3 only requires the first two moments of each continuous-time parameter). For example, since $\hat{\mu}_T = 1.7 \cdot 10^{-12}$, we let $E[\hat{\mu}] = \hat{\mu} = 1.7 \cdot 10^{-12}$ and $\text{Var}[\hat{\mu}] = \hat{b}_\mu^2 = (1 \cdot 10^{-11})^2$. In this case, the prior region covered by three standard deviations to the left and right of the prior mean approximately corresponds to a range from -15% to a 20% annual return. In general, specifying the prior standard deviation

Parameter	Value	Interpretation
$\hat{\mu}$	$1.7 \cdot 10^{-12}/\text{ms}$	Annual asset return of 1%, based on 251 trading days per year, 6.5 trading hours per trading day, excluding jumps between trading sessions.
$\hat{\theta}_1$	$7.94 \cdot 10^{-8}/\text{ms}$	Timescale of inertia for the slow log-volatility process equal to 3.5 hours.
$\hat{\theta}_2$	$1.67 \cdot 10^{-6}/\text{ms}$	Timescale of inertia for the slow log-volatility process equal to 10 minutes.
$\hat{\alpha}$	$-13 - \log(\sqrt{\text{ms}})$	Average of daily closing VIX values from 1/2/1990 to 4/10/2015 (18.9% a year), transformed to the $\log(\hat{\sigma}_t)$ scale with time measured in milliseconds.
$\hat{\tau}_1^2$	$1.86 \cdot 10^{-8}/\text{ms}$	$2\hat{\theta}_1$ times the variance of daily closing VIX values from 1/2/1990 to 4/10/2015, transformed to the $\log(\hat{\sigma}_t)$ scale with time measured in milliseconds. In terms of the annualized volatility for the price, the corresponding distribution of annualized volatility has (1st, 10th, 50th, 90th, 99th) percentiles given by (8.16%, 11.8%, 18.9%, 30.7%, 46.6%)
$\hat{\tau}_2^2$	$3.9 \cdot 10^{-7}/\text{ms}$	$2\hat{\theta}_2$ times the variance of daily closing VIX values from 1/2/1990 to 4/10/2015, transformed to the $\log(\hat{\sigma}_t)$ scale with time measured in milliseconds. In terms of the annualized volatility for the price, the corresponding distribution of annualized volatility has (1st, 10th, 50th, 90th, 99th) percentiles given by (8.16%, 11.8%, 18.9%, 30.7%, 46.6%)
ξ_T^2	$2.5 \cdot 10^{-7}$	Bid-ask spread of \$0.1 for an average price of \$100.
ρ_T	0	Innovations in the price and log-volatility process are independent, no leverage effect.

Table 3: Summary for model parameters in simulation data, along with associated market interpretation of these parameters.

in this manner usually leads to a relatively wide but reasonable prior coverage of parameter model values. A summary of the prior means and standard deviations used to fit the model is provided in Table 4.

Parameter	Prior mean	Prior standard deviation
$\hat{\mu}$	$1.7 \cdot 10^{-12}$	$1 \cdot 10^{-11}$
$\hat{\theta}$	$5.6 \cdot 10^{-7}$	$1 \cdot 10^{-6}$
$\hat{\alpha}$	-13	10
$\hat{\tau}^2$	$1.3 \cdot 10^{-7}$	$1 \cdot 10^{-6}$
ξ^2	$2.5 \cdot 10^{-7}$	$1 \cdot 10^{-6}$
ρ	0	1

Table 4: Parameters of the prior distributions used for inference with simulated data.

We fit three slightly different versions of our model to the simulated dataset. In the first version, the microstructure noise parameter was set to $\xi^2 = 0$, so our model does not take into account microstructure noise and reduces to the standard SV models used in the literature. In the second version, ξ^2 was fixed to $2.5 \cdot 10^{-7}$, a level of microstructure noise roughly consistent with the true level of microstructure noise added in the data. Finally, the third version corresponds to our full model where ξ^2 is estimated from the data by assigning it a Gamma prior with mean $2.5 \cdot 10^{-7}$ and standard deviation $1 \cdot 10^{-6}$ (see Table 4). The dataset was analyzed assuming sampling periods of 300, 30, 15, and 5 seconds (note that, because the size of the microstructure noise is assumed to be the same at every sampling scale, we use the same prior for ξ^2 for all sampling periods).

The plots for the posterior mean and 95% probability bands for the estimated log-volatility paths are given in Figure 1, along with the true signal from the simulated data. As the sampling period decreases, we see that the first version of the model (which ignores microstructure noise by fixing $\xi^2 = 0$) fails to capture the latent signal. On the other hand, the other two versions of our models produce much better posterior estimates for the latent signal. In particular, we see that, as expected, the naive choice $\xi^2 = 0$ model overestimates the volatility signal for higher sampling frequencies where the time interval Δ diminishes enough so that the microstructure noise

dominates for the volatility signal. We can also see that, although the second and third versions of the model tend to smooth out the true volatility path, the reconstruction generated by the model that estimates ξ^2 from the data is somewhat more accurate.

In addition to estimating the volatility path, we also investigate the ability of the model to infer model parameters. In particular, Figure 2 shows the posterior density estimates for the continuous-time parameters $\hat{\alpha}$, $\hat{\tau}^2$, $\hat{\theta}$ and $\hat{\mu}$ (which are comparable across scales), as well as the posterior distribution for ξ^2 (in the case of the third version of the model, which is the only one in which it is estimated from the data). Note that, when the model is estimated with $\xi^2 = 0$ fixed, the posterior densities for $\hat{\alpha}$, the mean level of log volatility, show a reduction in variance with increasing sampling frequency: posterior draws become more centered around a wrong, over-estimated value for mean log-volatility level. These results are consistent with those obtained for the volatility path and show that the model fails to capture the constant information content in the data regarding $\hat{\alpha}$. We also note that learning $\hat{\tau}^2$ and $\hat{\theta}$ is difficult whether we do or do not include microstructure noise. However, due to the constant information in the data with respect to $\hat{\alpha}$, the posterior uncertainty for $\hat{\theta}$ and $\hat{\tau}^2$ seems to remain constant even with increasing sampling frequency when ξ is not fixed.

6.2 Estimating Integrated Variance

As described in Section 4, the posterior draws for σ_j^2 allow us to approximate the posterior distributions for the integrated variance of the latent volatility process. In this Section we extend the previous simulation study to compare the 95% intervals generated by the three versions of our model with those generated from a realized variance estimate. The literature on realized volatility estimators for high-frequency data is vast (for a review, refer to Pigorsch et al., 2012), but the construction of confidence intervals for the realized volatility estimators can be challenging. Here we compare the coverage properties of our model-based credible intervals against bootstrap-based confidence intervals of the the kernel-based realized variance estimator introduced by Zhou [1996] and Hansen and Lunde [2006]. The idea behind the bootstrapping method is to periodically extend the available data set and randomly reselect a new data set to construct a bootstrap sample (see Hwang et al., 2013 for a full description of the procedure).

The results of the comparison are shown in Table 5. The table is constructed using 300 simulated data sets, each corresponding to a single trading day. We compare the percentage of times the 95% confidence/credible intervals for the integrated variance (IV) estimator covers the true integrated variance value. A well-calibrated interval will produce a 95% coverage on average, and we see that the estimator based on our approach where ξ^2 is fully estimated preforms very well, both when compared to $\xi^2 = 0$, $\xi^2 = 2.5 \cdot 10^{-7}$, and when compared to the kernel-based estimator.

Sampling period	5 min	60 sec	30 sec	15 sec	5 sec
Inference with $\xi^2 = 0$	93	72	28	3	0
Inference with $\xi^2 = 2.5 \cdot 10^{-7}$	95	79	57	23	0
Inference with ξ^2 estimated	95	91	92	96	97
Inference with kernel-based estimator	53	51	48	59	76

Table 5: Coverage table comparing the percentage of covered integrated variance levels by 95% confidence/credible intervals of our estimator and a kernel-based estimator.

6.3 Effect of microstructure noise and sampling frequency on estimates: market data

We perform an analysis for real market data, which consists of a single day of midpoint stock prices of Apple Inc. (NASDAQ:AAPL) on March 6th, 2014, printed on the millisecond from the NYSE TAQ data set. Our a priori estimate of the bid-ask spread driving microstructure noise is centered on \$0.1 as with the simulation data. The volatility paths are shown in Figure 3. In the case where $\xi^2 = 0$, the model estimates the volatility signal to be, on average, higher than in the cases where $\xi^2 > 0$, which is also seen in the posterior means estimates of $\hat{\alpha}$ in Figure 4. This is consistent with the simulation-study results, where the $\xi^2 = 0$ model attributes microstructure noise to the log-volatility signal.

Figure 4 shows the posterior distributions for the model parameters. Note that the posterior for ξ^2 is centered around $8 \cdot 10^{-9}$ – two orders of magnitude smaller than the prior center of $2.7 \cdot 10^{-7}$. This value of the posterior mean is roughly equivalent to a bid-ask spread of \$0.01, which is reasonable for a highly-traded stock like AAPL. Furthermore, note that fixing $\xi^2 = 2.5 \cdot 10^{-7}$ compared to treating ξ^2 as an unknown parameter, leads to volatility paths that are smoother and have greater coherence in posterior estimates of the model parameters across sampling periods. This is especially true for $\hat{\theta}$, since the closer the log-volatility process is to being discontinuous,

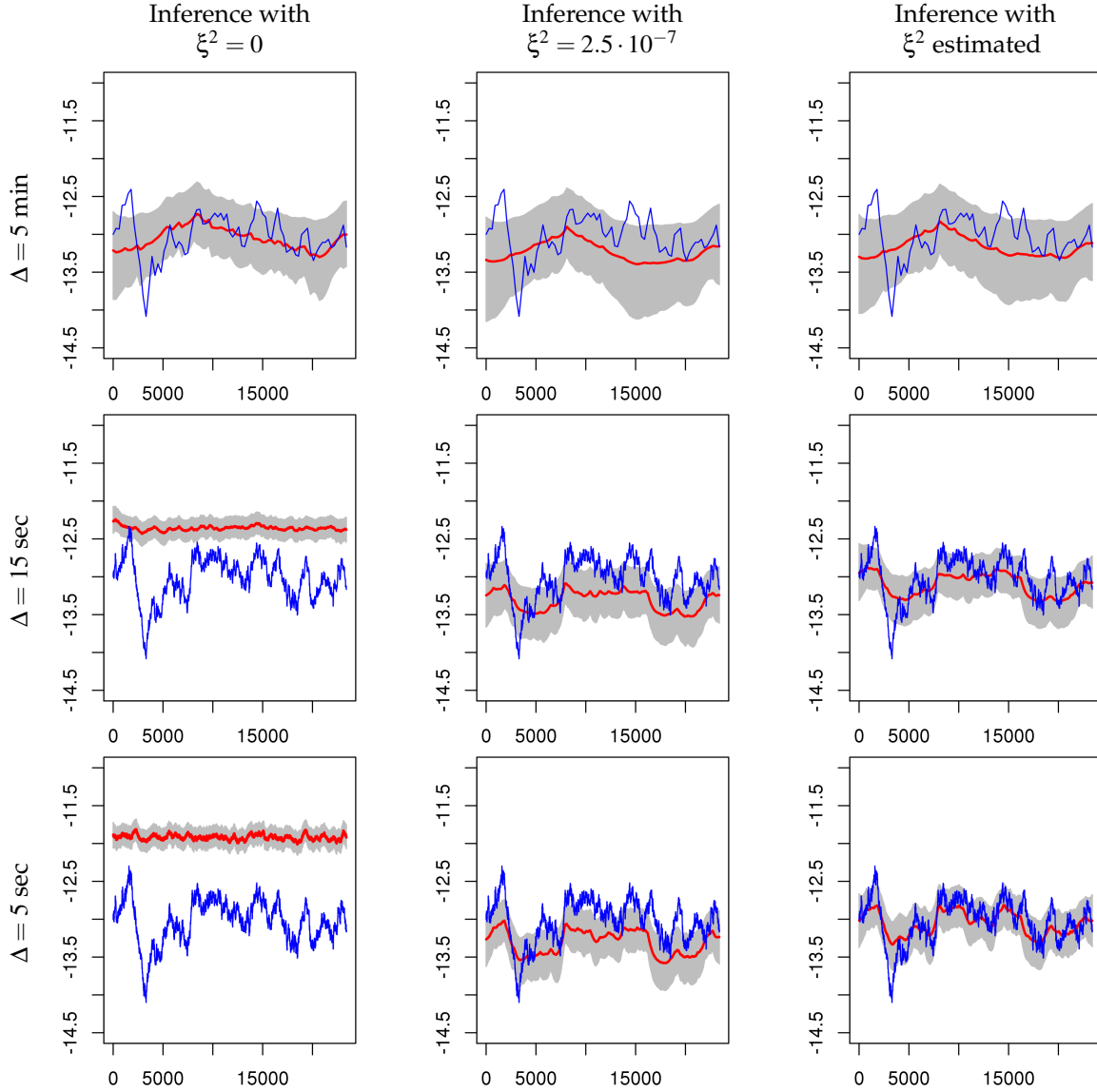


Figure 1: Log-volatility paths for simulated data. All three inference approaches are applied to the same data set that contains microstructure noise. The microstructure noise added in the simulated data is approximately at the level of $\xi^2 = 2.5 \cdot 10^{-7}$. Red denotes the posterior mean of the paths, while the gray region denotes the posterior 95% probability for the log-volatility value. Blue is the true log-volatility signal. We see that when microstructure is ignored ($\xi^2 = 0$), we fail to recover the true signal.

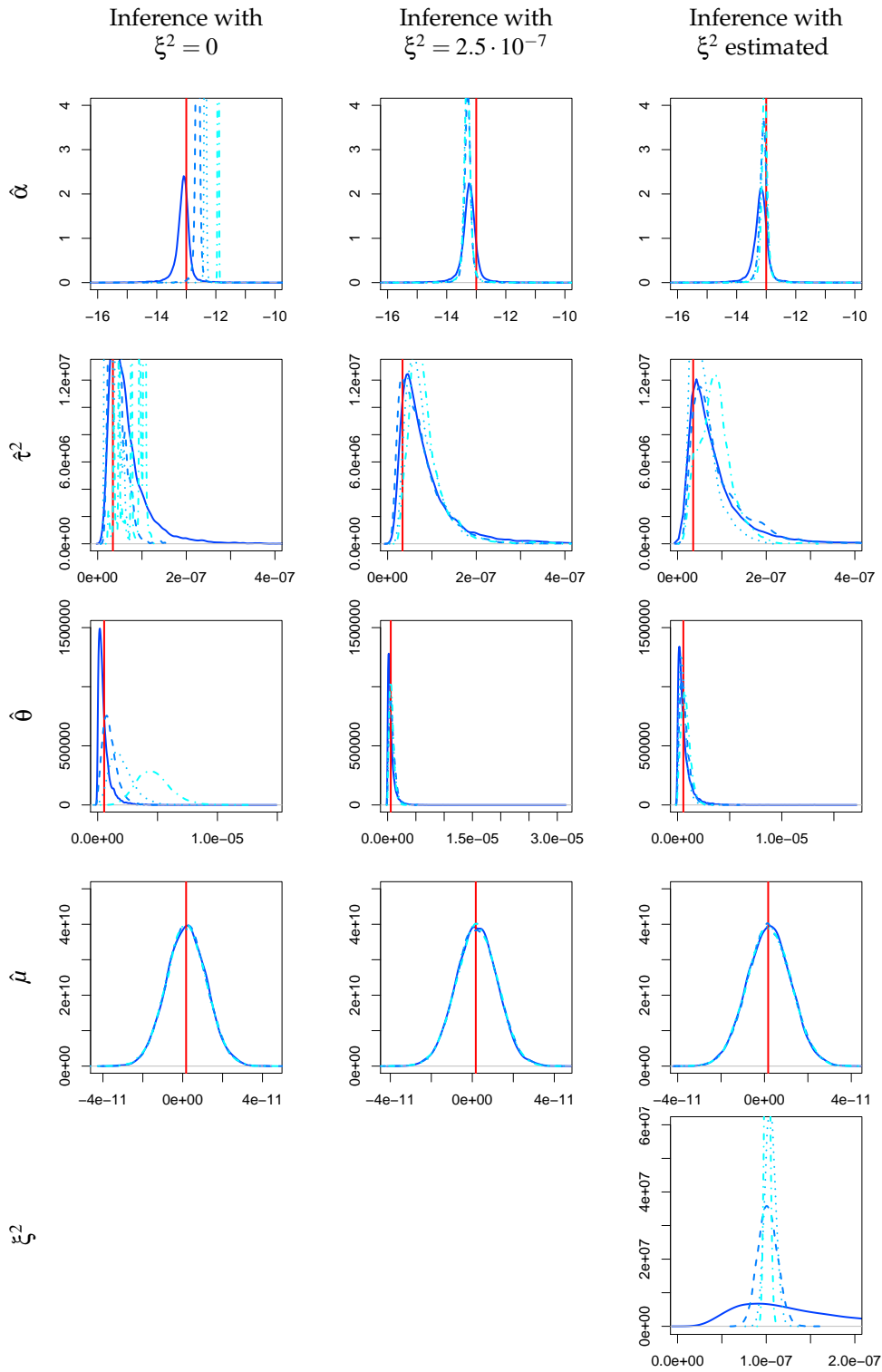


Figure 2: Posterior density approximations of model parameters for simulated data. The red vertical line represents the true parameter value. The sampling periods used are: 5 minutes (—), 30 seconds (— · —), 15 seconds (· · · · ·), and 5 seconds (· · · · ·).

the shorter its timescale must be. We thus see the obvious trade-off when specifying ξ^2 : if ξ^2 is too large, we run the risk of over-smoothing; if ξ^2 is too small, we confound the effects of noise with those of the volatility process.

6.4 Effect of timescale of inertia on estimates of $\hat{\alpha}$

To illustrate the discussion in Section 5 on the effect of the log-volatility timescale on our method's ability to learn $\hat{\alpha}$, we examine the posterior uncertainty for $\hat{\alpha}$ under two scenarios: 1) increasing sample size N by decreasing

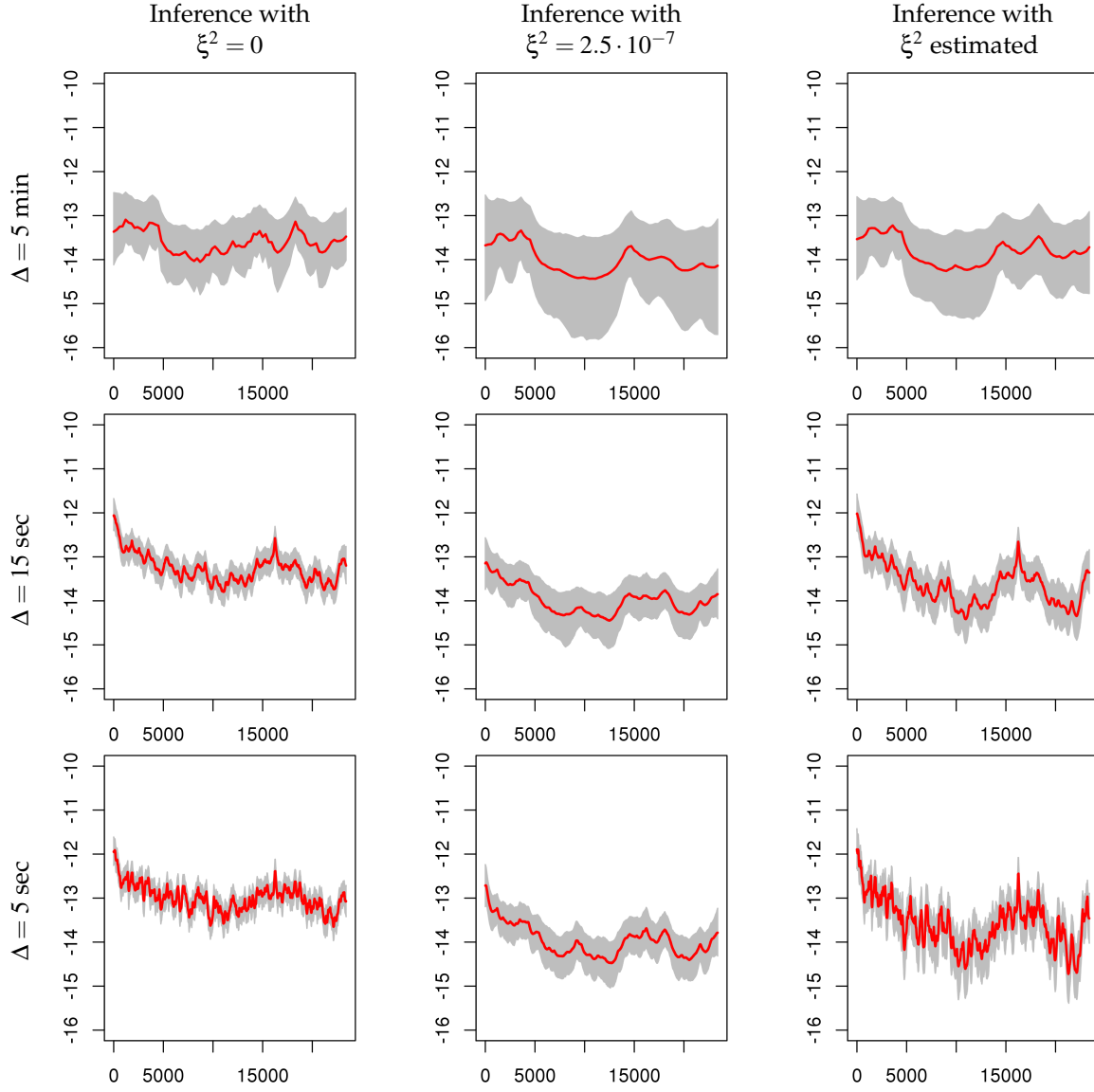


Figure 3: Log-volatility paths for the AAPL 03/06/2014 data. Red denotes the posterior mean of the paths, while the gray region denotes the posterior 95% probability for the log-volatility value.

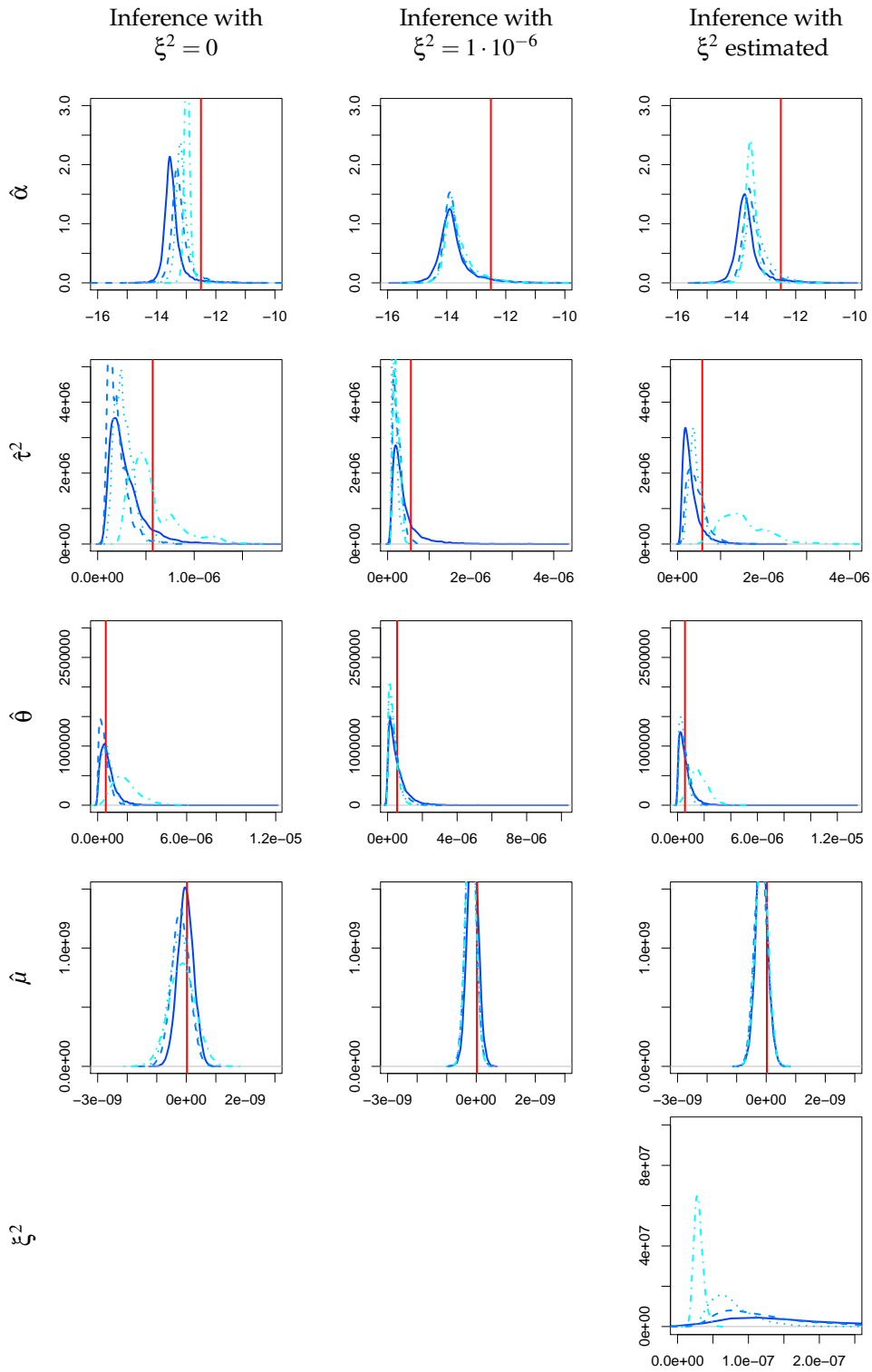


Figure 4: Posterior density approximations of model parameters for the AAPL 03/06/2014 data. The sampling periods used are: 5 minutes (—), 30 seconds (---), 15 seconds (····), 5 seconds (· · · ·), and 1 second (— · —)

the sampling period Δ , and 2) increasing N by increasing the observational period T while keeping Δ constant. The same simulated dataset is used in both 1) and 2), with $1/\hat{\theta} = 15$ min, such that $\hat{\theta}\Delta \leq 1$ and the approximation for the posterior variance of $\hat{\alpha}$ in (28) is applicable.

Under scenario 1), we consider the entire data set and estimate $\hat{\alpha}$ with $\Delta = 1$ min, 30 sec, 15 sec, and 5 sec. The posterior distributions for $\hat{\alpha}$ are shown in the left panel of Figure 5. We see that the posterior uncertainty for $\hat{\alpha}$ remains the same with increasing number of intraperiod samples, as suggested by the analysis in Section 5. Under scenario 2), we fix $\Delta = 1$ min and increase sample sizes by increasing the observational period T , using

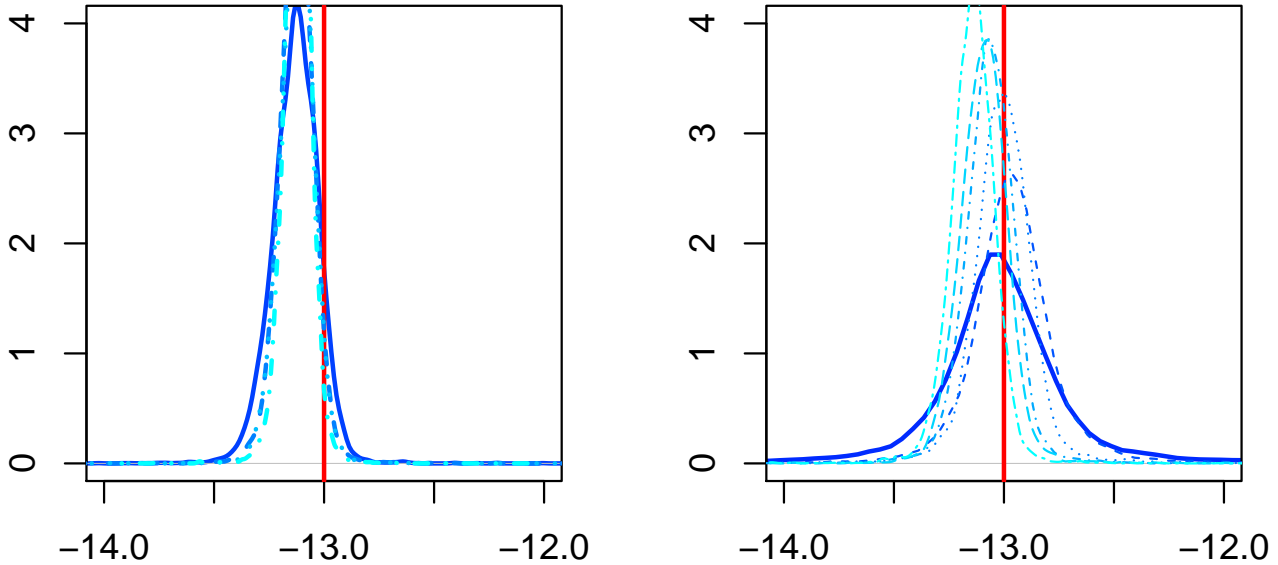


Figure 5: Posterior densities of $\hat{\alpha}$ for two data sets examined under scenarios of increasing sample size N by decreasing Δ (left) and increasing N by increasing the observational duration T (right). The same data set was used for both cases, where $1/\hat{\theta} = 15$ min so that $\hat{\theta}\Delta \leq 1$ for all posterior samples. The red vertical line signifies the true $\hat{\alpha}$ value used in the data-generation process. For a fixed T and decreasing Δ (left) [1 minute (—), 30 seconds (---), 15 seconds (·····), 5 seconds (·-·-·)] the posterior variance of $\hat{\alpha}$ stays approximately the same. For a fixed Δ and increasing T [65 minutes (—), 130 minutes (---), 195 minutes (·····), 260 minutes (·-·-·), 325 minutes (— · —), and 390 (---) minutes], the posterior variance of $\hat{\alpha}$ tends to decrease.

the first 1/6 (65 min) of the data, the first 2/6 (130 min) of the data, and so on through the entirety of the data (390 min). The right panel of Figure 5 shows the posterior densities for $\hat{\alpha}$ under this regime. Confirming the discussion in Section 5, we tend to see a decreasing posterior variance for $\hat{\alpha}$ with increasing observational duration, but not when the sampling frequency increases. The important takeaway point is that a dataset covering a finite observational period contains a finite amount of information, no matter how finely the observational period is sampled.

7 Conclusion

In this paper we have outlined a discrete-time stochastic volatility model for high-frequency data. The model and the algorithm used to estimate it are designed to be coherent across all sampling frequencies. To this end, we elicit priors on the parameters of the continuous-time version of our model and transform them to the discrete-time scale. Both simulation and real data results show that adding the microstructure term in the standard stochastic volatility formulation allows one to fit such models to high-frequency data and extract the log-volatility signal from noisy observations. However, having a good prior estimate of the noise level is an important specification, since attributing some fluctuations in the observed log volatility to microstructure noise has a smoothing effect on the reconstructed log volatility paths. Finally, simulation studies show that the integrated variance estimator derived from our model is well-calibrated and outperforms current kernel-based realized volatility estimators.

Appendix A: Details of the Markov chain Monte Carlo algorithm

We write the full discrete-time model

$$\begin{aligned} Y_j &= \log(S_j) + \zeta_j, \\ \log(S_j) &= \mu(\Delta) + \log(S_{j-1}) + \sigma_j \varepsilon_{j,1}, \\ \log(\sigma_{j+1}) &= \alpha(\Delta) + \theta(\Delta) \{ \log(\sigma_j) - \alpha(\Delta) \} + \tau(\Delta) \varepsilon_{j,2}, \end{aligned} \tag{30}$$

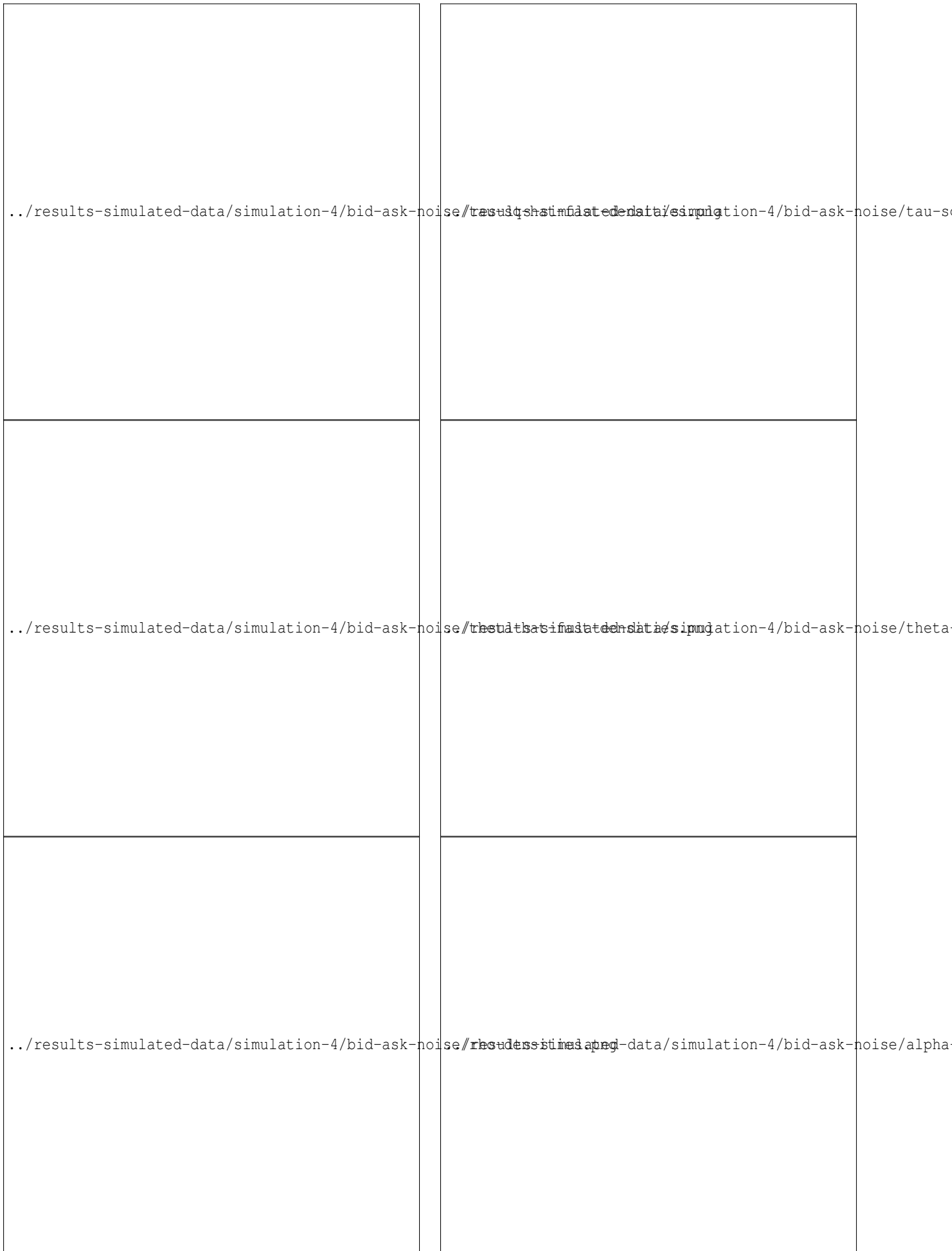



Figure 6: Estimating parameters freely with Data Set 4.



../results-simulated-data/simulation-4/bid-ask-noise/timeseries-5000.png

../results-simulated-data/simulation-4/bid-ask-noise/timeseries-1000.png

../results-simulated-data/simulation-4/bid-ask-noise/timeseries-200.png

Figure 7: Estimating vol paths freely with Data Set 4.

```
../results-simulated-data/simulation-4/bid-ask-noise/fixed-true-simulated-data/simulations-4/bid-ask-noise/fixed-
```

```
../results-simulated-data/simulation-4/bid-ask-noise/fixed-tsh/simulated-data-fast/simulation-4/bid-ask-noise/fixed-
```

```
../results-simulated-data/simulation-4/bid-ask-noise/fixed-true-simulated-data/sing
```

Figure 8: Estimating parameters freely with Data Set 4 with fixe rho



Figure 9: Estimating vol paths freely with Data Set 4 with fixed rho

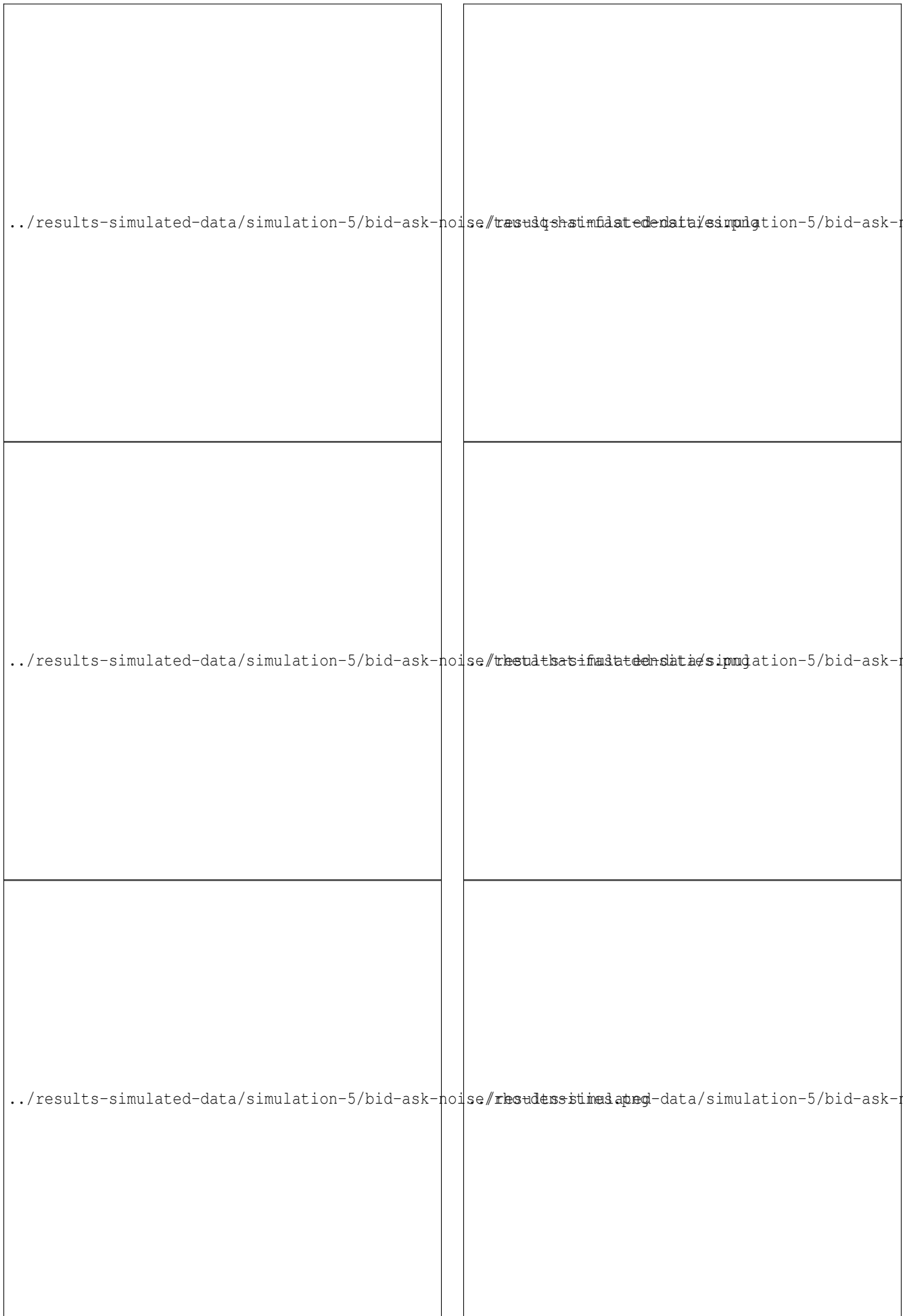


Figure 10: Estimating parameters freely with Data Set 5, switching p .

../results-simulated-data/simulation-5/bid-ask-noise/timeseries-60000.png

../results-simulated-data/simulation-5/bid-ask-noise/timeseries-30000.png

../results-simulated-data/simulation-5/bid-ask-noise/timeseries-5000.png

Figure 11: Estimating vol paths freely with Data Set 5, switching p..

where

$$\zeta_j \sim N(0, \xi^2), \quad \begin{pmatrix} \varepsilon_{j,1} \\ \varepsilon_{j,2} \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right), \quad \log(\sigma_1) \sim N \left(\alpha(\Delta), \frac{\tau(\Delta)^2}{1 - \theta(\Delta)^2} \right), \quad \log(S_0) \sim N(\eta, \kappa^2).$$

This model is nonlinear in terms of the volatility due to the formulation of its evolution on the log-scale. To reparameterize the model to be linear in terms of volatility and thereby use the Kalman Filter and Sampler, we take equation (30) and transform it so that it is linear in terms of $\log(\sigma_j)$,

$$\log(S_j) = \mu(\Delta) + \log(S_{j-1}) + \sigma_j \varepsilon_{j,1} \quad \leftrightarrow \quad \underbrace{\log \left[\left| \log(S_j/S_{j-1}) - \mu(\Delta) \right| \right]}_{y_j^*} = \underbrace{\log(\sigma_j)}_{h_j} + \underbrace{\log(\varepsilon_{j,1}^2)/2}_{\varepsilon_{j,1}^*}.$$

Having defined y_j^* , h_j , and $\varepsilon_{j,1}^*$, the model becomes linear in the terms involving the volatility:

$$Y_j = \log(S_j) + \zeta_j, \quad (31)$$

$$y_j^* = h_j + \varepsilon_{j,1}^*, \quad (32)$$

$$h_{j+1} = \alpha(\Delta) + \theta(\Delta) \{h_j - \alpha(\Delta)\} + \tau(\Delta) \varepsilon_{j,2}, \quad (33)$$

We approximate $\varepsilon_{j,1}^*$ as a mixture of normals

$$\varepsilon_{j,1}^* = \log(\varepsilon_{j,1}^2)/2 \sim \sum_{l=1}^{10} p_l N \left(\frac{m_l}{2}, \frac{v_l^2}{4} \right).$$

We can introduce the mixture indicators $\gamma_1, \dots, \gamma_{n(\Delta)}$ such that

$$\log(\varepsilon_{j,1}^2)/2 \mid \gamma_j \sim N \left(\frac{m_{\gamma_j}}{2}, \frac{v_{\gamma_j}^2}{4} \right), \quad \Pr(\gamma_k = l) = p_l.$$

Hence, conditionally on the true prices, the indicators $\gamma_1, \dots, \gamma_{n(\Delta)}$, and the hyperparameters $\mu(\Delta)$, $\alpha(\Delta)$, $\theta(\Delta)$, $\tau(\Delta)$ and ρ , we have again a linear state-space model with Gaussian innovations. However, due to correlation of the innovations $\varepsilon_{j,1}$ and $\varepsilon_{j,2}$, we need a joint distribution for the transformed and approximated $\varepsilon_{j,1}^*$ and $\varepsilon_{j,2}$. To this end, we directly follow the approach in Omori et al. [2007], beginning with known expression

$$\begin{aligned} p(\varepsilon_{j,2}, \varepsilon_{j,1}^* \mid \gamma_j) &= p(\varepsilon_{j,2} \mid \varepsilon_{j,1}^*, \gamma_j) p(\varepsilon_{j,1}^* \mid \gamma_j) \\ &= p(\varepsilon_{j,2} \mid \underbrace{d_j \exp(\varepsilon_{j,1}^*)}_{\varepsilon_{j,1}}, \gamma_j) p(\varepsilon_{j,1}^* \mid \gamma_j) \\ &= N(\varepsilon_{j,2} \mid \rho d_j \exp(\varepsilon_{j,1}^*), (1 - \rho^2)) N \left(\varepsilon_{j,1}^* \mid \frac{m_{\gamma_j}}{2}, \frac{v_{\gamma_j}^2}{4} \right), \end{aligned}$$

where d_j is the sign of $\varepsilon_{j,1}$. The nonlinear term $\exp(\varepsilon_{j,1}^*)$ is approximated by a linear function, where the constants $(a_{\gamma_j}, b_{\gamma_j})$ are chosen to minimize the expected squared difference between $\exp(\varepsilon_{j,1}^*)$ and its approximation, as done in Omori et al. [2007]

$$\exp(\varepsilon_{j,1}^*) \mid \gamma_j \approx \exp(m_{\gamma_j}/2) (a_{\gamma_j} + b_{\gamma_j} (2\varepsilon_{j,1}^* - m_{\gamma_j})).$$

If $z_j^*, z_j \stackrel{iid}{\sim} N(0, 1)$, the joint distribution for the conditional distribution of the pair $(\varepsilon_{j,1}^*, \varepsilon_{j,2} \mid \gamma_j)$ can be written as

$$\left\{ \begin{pmatrix} \varepsilon_{j,1}^* \\ \varepsilon_{j,2} \end{pmatrix} \mid \gamma_j \right\} = \begin{pmatrix} m_{\gamma_j}/2 \\ d_j \rho \exp(m_{\gamma_j}/2) a_{\gamma_j} \end{pmatrix} + \begin{pmatrix} v_{\gamma_j}/2 & 0 \\ d_j \rho b_{\gamma_j} v_{\gamma_j} \exp(m_{\gamma_j}/2) & \sqrt{1 - \rho^2} \end{pmatrix} \begin{pmatrix} z_j^* \\ z_j \end{pmatrix}. \quad (34)$$

Rearranging equation (32) to express z_j^* in terms of y_j^* , h_j , m_{γ_j} , and v_{γ_j} , then substituting into equation (33) allows us to finally write down the model in the convenient linear state-space form, where the innovations in the state-evolution equations are independent:

$$\begin{aligned} Y_j &= \log(S_j) + \zeta_j, \\ y_j^* &= h_j + \frac{m_{\gamma_j}}{2} + \frac{v_{\gamma_j}}{2} z_j^*, \\ h_{j+1} &= \theta_j(\Delta) h_j + \alpha_j(\Delta) + \tau(\Delta) \sqrt{1 - \rho^2} z_j, \end{aligned}$$

with

$$\begin{aligned}\theta_j(\Delta) &= \theta(\Delta) - d_j \rho b_{\gamma_j} \cdot 2 \exp(m_{\gamma_j}/2), \\ \alpha_j(\Delta) &= \alpha(\Delta)(1 - \theta(\Delta)) + \rho d_j \exp(m_{\gamma_j}/2) a_{\gamma_j} + d_j \rho b_{\gamma_j} v_{\gamma_j} \exp(m_{\gamma_j}/2) \frac{y_j^* - m_{\gamma_j}/2}{v_{\gamma_j}/2}.\end{aligned}$$

The full likelihood for the model can be written as

$$\begin{aligned}p(Y_1, \dots, Y_{n(\Delta)} | h_{1, \dots, n(\Delta)}, n(\Delta) + 1, \log(S_0, \dots, n(\Delta)), \gamma_1, \dots, n(\Delta), \xi, \rho, \tau(\Delta), \theta(\Delta), \alpha(\Delta)) \propto \\ \prod_{j=1}^{n(\Delta)} \xi^{-1/2} \exp \left\{ -\frac{1}{2\xi} (Y_j - \log(S_j))^2 \right\} \\ \times \prod_{j=1}^{n(\Delta)} (v_{\gamma_j}/2)^{-1} \exp \left\{ -\frac{1}{2v_{\gamma_j}^2/4} (y_j^* - h_j - m_{\gamma_j}/2)^2 \right\} \\ \times \prod_{j=1}^{n(\Delta)} \left(\tau(\Delta) \sqrt{1 - \rho^2} \right)^{-1} \exp \left\{ -\frac{1}{2\tau(\Delta)^2(1 - \rho^2)} (h_{j+1} - \theta_j(\Delta)h_j - \alpha_j(\Delta))^2 \right\} \\ \times p(\log(S_0))p(\log(\sigma_1)).\end{aligned}$$

For our MCMC algorithm, we implement a Gibbs sampler where we simulate posterior draws from the full conditional posteriors for each set of parameters in the following steps:

1. **Sample $\tau^2(\Delta)$.** With an Inverse-Gamma prior $\tau^2(\Delta) \sim \text{Inv-Gamma}(a_{\tau^2}(\Delta), b_{\tau^2}(\Delta))$, the full conditional posterior for $\tau^2(\Delta)$ is also an Inverse-Gamma with

$$\begin{aligned}p(\tau^2(\Delta) | -) &= \text{Inv-Gamma}(\tau^2(\Delta) | A, B) \\ A &= a_{\tau^2} + \frac{n(\Delta) + 1}{2} \\ B &= b_{\tau^2} + \frac{1}{2(1 - \rho^2)} \sum_{j=1}^{n(\Delta)} (h_{j+1} - \theta_j(\Delta)h_j - \alpha_j(\Delta))^2 + \frac{1 - \theta(\Delta)^2}{2} (h_1 - \alpha(\Delta))^2.\end{aligned}$$

2. **Sample $\theta(\Delta)$.** The full conditional posterior for $\theta(\Delta)$ is given by

$$\begin{aligned}p(\theta(\Delta) | -) \propto \\ p(\theta(\Delta)) \left[\prod_{j=1}^{n(\Delta)} \frac{1}{\tau(\Delta) \sqrt{1 - \rho^2}} \exp \{ (h_{j+1} - \theta_j(\Delta)h_j - \alpha_j(\Delta))^2 \} \right] \\ \times \sqrt{\frac{1 - \theta(\Delta)^2}{\tau^2(\Delta)}} \exp \left\{ -\frac{1 - \theta(\Delta)^2}{2\tau^2(\Delta)} (h_1 - \alpha(\Delta))^2 \right\}.\end{aligned}$$

With a normal prior for $\theta(\Delta)$, the product term

$$p(\theta(\Delta)) \left[\prod_{j=1}^{n(\Delta)} \frac{1}{\tau(\Delta) \sqrt{1 - \rho^2}} \exp \{ (h_{j+1} - \theta_j(\Delta)h_j - \alpha_j(\Delta))^2 \} \right]$$

can be reduced to a Normal-kernel form. This can be used as an efficient proposal distribution in a Metropolis-Hasting step, with the rest of the likelihood used to reject or accept the proposal.

3. **Sample $\alpha(\Delta)$.** The full conditional posterior for $\alpha(\Delta)$ is given by

$$\begin{aligned}p(\alpha(\Delta) | -) \propto \\ p(\alpha(\Delta)) \left[\prod_{j=1}^{n(\Delta)} \frac{1}{\tau(\Delta) \sqrt{1 - \rho^2}} \exp \{ (h_{j+1} - \theta_j(\Delta)h_j - \alpha_j(\Delta))^2 \} \right] \\ \times \sqrt{\frac{1 - \theta(\Delta)^2}{\tau^2(\Delta)}} \exp \left\{ -\frac{1 - \theta(\Delta)^2}{2\tau^2(\Delta)} (h_1 - \alpha(\Delta))^2 \right\}.\end{aligned}$$

Just as above, with a normal prior for $\alpha(\Delta)$, the product term

$$p(\alpha(\Delta)) \left[\prod_{j=1}^{n(\Delta)} \frac{1}{\tau(\Delta) \sqrt{1 - \rho^2}} \exp \{ (h_{j+1} - \theta_j(\Delta)h_j - \alpha_j(\Delta))^2 \} \right]$$

can be reduced to a Normal-kernel form and used an efficient proposal in a Metropolis-Hastings step.

4. **Sample ξ .** With an Inverse-Gamma prior on ξ such that $\xi \sim \text{Inv-Gamma}(a_\xi, b_\xi)$, the full conditional posterior for ξ is also Inverse-Gamma

$$p(\xi|-) = \text{Inverse-Gamma} \left(\xi \left| a_\xi + n(\Delta)/2, b_\xi + \frac{1}{2} \sum_{j=1}^{n(\Delta)} (Y_j - \log(S_j))^2 \right. \right).$$

5. **Sample $\mu(\Delta)$.** For sampling $\mu(\Delta)$, it is convenient to consider the state-space model in terms of the latent log-prices $\log(S_j)$. The parameter $\mu(\Delta)$ appears only in the evolution of the log-price process. Conditional on all other parameters, we have

$$\log(S_j) = \log(S_{j-1}) + \mu(\Delta) + \sigma_j \varepsilon_{j,1},$$

so that the full conditional for $\mu(\Delta)$ is

$$p(\mu(\Delta)|-) \propto p(\mu(\Delta)) N \left(\left(\sum_{j=1}^{n(\Delta)} \frac{\log(S_j/S_{j-1})}{\sigma_j^2} \right) \left(\sum_{j=1}^{n(\Delta)} \frac{1}{\sigma_j^2} \right)^{-1}, \left(\sum_{j=1}^{n(\Delta)} \frac{1}{\sigma_j^2} \right)^{-1} \right).$$

Because each y_j^* is dependent on $\mu(\Delta)$ we must be careful to re-define y_j^* after this sample.

6. **Sample $\gamma_{1,\dots,n(\Delta)}$.** Since each γ_j can take on a finite number of values, for each j we sample the discrete posterior where

$$p(\gamma_j = l|-) \propto p(\gamma = l)(v_l/2)^{-1} \exp \left\{ -\frac{1}{2v_l^2/4} (y_j^* - h_j - m_l/2)^2 \right\} \\ \times \exp \left\{ -\frac{1}{2\tau(\Delta)^2(1-\rho^2)} (h_{j+1} - \theta_j(\Delta)h_j - \alpha_j(\Delta))^2 \right\}$$

7. **Sample the latent log-volatilities.** Conditional on all other parameters, the portion of the state-space model where h_j appears is comprised of the linear system

$$y_j^* = h_j + \frac{m_{\gamma_j}}{2} + \frac{v_{\gamma_j}}{2} z_j^*, \quad z_j^* \sim N(0, 1) \\ h_{j+1} = \theta_j(\Delta)h_j + \alpha_j(\Delta) + \tau(\Delta)\sqrt{1-\rho^2}z_j, \quad z_j \sim N(0, 1) \\ p(h_1) = N \left(h_1 \left| \alpha(\Delta), \frac{\tau(\Delta)^2}{1-\theta(\Delta)^2} \right. \right)$$

We can efficiently obtain posterior samples for $h_{1,\dots,n(\Delta),n(\Delta)+1}$ using the Kalman Forward Filter and Backward Sampler.

8. **Sampling the latent log-prices.** Conditional on all other parameters, the portion of the state-space model where $\log(S_j)$ appears is comprised of the linear system

$$Y_j = \log(S_j) + \zeta_j, \quad \zeta_j \sim N(0, \xi^2) \\ \log(S_j) = \mu(\Delta) + \log(S_{j-1}) + \sigma_j \varepsilon_{j,1}, \quad \varepsilon_{j,1} \sim N(0, 1) \\ p(\log(S_0)) = N(\log(S_0) | \eta, \kappa^2).$$

As before, we sample $\log(S_j)$ using the Kalman Forward Filter and Backward Sampler. Because each y_j^* is dependent on $\log(S_j)$ we must not forget to re-define y_j^* after this sample.

References

- Yacine Ait-Sahalia, Per A Mykland, and Lan Zhang. Ultra high frequency volatility estimation with dependent microstructure noise. *Journal of Econometrics*, 160(1):160–175, 2011.
- Sassan Alizadeh, Michael W Brandt, and Francis X Diebold. Range-based estimation of stochastic volatility models. *The Journal of Finance*, 57(3):1047–1091, 2002.
- Torben G Andersen and Tim Bollerslev. Intraday periodicity and volatility persistence in financial markets. *Journal of empirical finance*, 4(2):115–158, 1997.

- Torben G Andersen, Tim Bollerslev, and Steve Lange. Forecasting financial market volatility: Sample frequency vis-a-vis forecast horizon. *Journal of Empirical Finance*, 6(5):457–477, 1999.
- Torben G Andersen, Tim Bollerslev, Francis X Diebold, and Paul Labys. The distribution of realized exchange rate volatility. *Journal of the American statistical association*, 96(453):42–55, 2001.
- Ole E Barndorff-Nielsen and Neil Shephard. Non-gaussian ornstein–uhlenbeck-based models and some of their uses in financial economics. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(2):167–241, 2001.
- Ole E Barndorff-Nielsen and Neil Shephard. Estimating quadratic variation using realized variance. *Journal of Applied Econometrics*, 17(5):457–477, 2002.
- Ole E Barndorff-Nielsen, Peter Reinhard Hansen, Asger Lunde, and Neil Shephard. Designing realized kernels to measure the ex post variation of equity prices in the presence of noise. *Econometrica*, 76(6):1481–1536, 2008.
- Fischer Black. The pricing of commodity contracts. *Journal of financial economics*, 3(1):167–179, 1976.
- Tim Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3):307–327, 1986.
- Tim Bollerslev and Hao Zhou. Estimating stochastic volatility diffusion using conditional moments of integrated volatility. *Journal of Econometrics*, 109(1):33–65, 2002.
- Michael W Brandt and Francis X Diebold. A no-arbitrage approach to range-based estimation of return covariances and correlations. Technical report, National Bureau of Economic Research, 2003.
- Chris K Carter and Robert Kohn. On gibbs sampling for state space models. *Biometrika*, 81(3):541–553, 1994.
- George Casella and Roger L Berger. *Statistical inference*, volume 2. Duxbury Pacific Grove, CA, 2002.
- Mikhail Chernov, A Ronald Gallant, Eric Ghysels, and George Tauchen. Alternative models for stock price dynamics. *Journal of Econometrics*, 116(1):225–257, 2003.
- Ray Yeutien Chou, Hengchih Chou, and Nathan Liu. Range volatility models and their applications in finance. In *Handbook of Quantitative Finance and Risk Management*, pages 1273–1281. Springer, 2010.
- Fabienne Comte and Eric Renault. Long memory in continuous-time stochastic volatility models. *Mathematical Finance*, 8(4):291–323, 1998.
- Feike C Drost and Theo E Nijman. Temporal aggregation of garch processes. *Econometrica: Journal of the Econometric Society*, pages 909–927, 1993.
- Ola Elerian, Siddhartha Chib, and Neil Shephard. Likelihood inference for discretely observed nonlinear diffusions. *Econometrica*, 69(4):959–993, 2001.
- Bjørn Eraker. Mcmc analysis of diffusion models with application to finance. *Journal of Business & Economic Statistics*, 19(2):177–191, 2001.
- Sylvia Frühwirth-Schnatter. Data augmentation and dynamic linear models. *Journal of time series analysis*, 15(2):183–202, 1994.
- Peter R Hansen and Asger Lunde. Realized variance and market microstructure noise. *Journal of Business & Economic Statistics*, 24(2):127–161, 2006.
- Peter Reinhard Hansen, Zhuo Huang, and Howard Hwan Shek. Realized GARCH: a joint model for returns and realized measures of volatility. *Journal of Applied Econometrics*, 27(6):877–906, 2012.
- Xin Huang and George Tauchen. The relative contribution of jumps to total price variance. *Journal of financial econometrics*, 3(4):456–499, 2005.
- John Hull and Alan White. The pricing of options on assets with stochastic volatilities. *The journal of finance*, 42(2):281–300, 1987.
- Eunju Hwang, Dong Wan Shin, et al. Stationary bootstrapping realized volatility under market microstructure noise. *Electronic Journal of Statistics*, 7:2032–2053, 2013.

- Worapree Maneesoonthorn, Catherine S Forbes, and Gael M Martin. Inference on self-exciting jumps in prices and volatility using high frequency measures. *arXiv preprint arXiv:1401.3911*, 2014.
- Yasuhiro Omori, Siddhartha Chib, Neil Shephard, and Jouchi Nakajima. Stochastic volatility with leverage: Fast and efficient likelihood inference. *Journal of Econometrics*, 140(2):425–449, 2007.
- Christian Pigorsch, Uta Pigorsch, and Ivaylo Popov. Volatility estimation based on high-frequency data. In *Handbook of Computational Finance*, pages 335–369. Springer, 2012.
- Shinichiro Shirota, Takayuki Hizu, and Yasuhiro Omori. Realized stochastic volatility with leverage and long memory. *Computational Statistics & Data Analysis*, 76:618–641, 2014.
- Hans R Stoll. Presidential address: friction. *The Journal of Finance*, 55(4):1479–1514, 2000.
- Makoto Takahashi, Yasuhiro Omori, and Toshiaki Watanabe. Estimating stochastic volatility models using daily returns and realized volatility simultaneously. *Computational Statistics & Data Analysis*, 53(6):2404–2426, 2009.
- JH Venter and PJ de Jongh. Extended stochastic volatility models incorporating realised measures. *Computational Statistics & Data Analysis*, 2012.
- Lan Zhang, Per A Mykland, and Yacine Aït-Sahalia. A tale of two time scales. *Journal of the American Statistical Association*, 100(472), 2005.
- Bin Zhou. High-frequency data and volatility in foreign-exchange rates. *Journal of Business & Economic Statistics*, 14(1):45–52, 1996.
- Gilles Zumbach. The pitfalls in fitting garch (1, 1) processes. In *Advances in Quantitative Asset Management*, pages 179–200. Springer, 2000.