

Bayesian Comparison of Spot Volatility Models in High Frequency^{*}

Yaohan Chen
Anhui University

Jia Li
Singapore Management University

Jun Yu
University of Macau

November 10, 2025

Abstract

Building on the fixed- k inference framework developed by Bollerslev, Li, and Liao (2021), this paper introduces a class of alternative models for spot volatility in high-frequency, all of which fall within a nonlinear and non-Gaussian state-space framework. We propose Bayesian and particle filter methods for parameter estimation, latent spot volatility extraction, and model comparison. Simulation studies demonstrate the effectiveness of our approach, while empirical applications suggest new directions for spot volatility modeling.

Keywords: High-frequency econometrics, Spot volatility, Nonlinear non-Gaussian state-space model, Markov Chain Monte Carlo, Model selection.

JEL Classification: C14, C22, C32, C58.

^{*}Yaohan Chen, Anhui University, Hefei, China. Email: yaohan.chen@ahu.edu.cn. Jia Li, School of Economics, Singapore Management University, Singapore. Email: jiali@smu.edu.sg. Jun Yu, Faculty of Business Administration, University of Macau, Avenida da Universidade, Macao, China. Email: junyu@um.edu.mo.

1. Introduction

Financial market volatility, as a key measure of risk, plays a vital role in both financial theory and practical applications of asset pricing (Engle, 2004). Early literature recognized the time-varying nature of daily volatility and focused on parametric volatility modeling using daily returns. Seminal contributions include the ARCH model by Engle (1982), the GARCH model by Bollerslev (1986), and the stochastic volatility model by Taylor (1982). These models provide daily volatility estimates as a byproduct of parameter estimation.

More recent literature employs daily realized volatility (RV), a nonparametric estimator of daily integrated volatility (IV), which is constructed from intraday returns, typically at 5-minute intervals. By leveraging high-frequency data, RV delivers more accurate volatility estimates than daily returns. This has spurred extensive research on modeling and forecasting RV, as seen in Andersen et al. (2001a), Andersen et al. (2001b), Andersen et al. (2003), Gatheral, Jaisson, and Rosenbaum (2018), and Wang, Xiao, and Yu (2023), among others. Beyond volatility estimation, RV has broad applications, such as constructing GMM estimators for diffusion models (Bollerslev and Zhou, 2002).

While parametric models for daily volatility are ill-suited for high-frequency spot volatility, which exhibits more complex dynamics. Developing accurate high-frequency volatility models is crucial for understanding intraday and interday volatility behavior, with implications for asset pricing, forecasting, trading, and risk management. For instance, portfolio managers increasingly require intraday rebalancing, necessitating sub-daily volatility and covariance forecasts.

Stroud and Johannes (2014) pioneered high-frequency volatility modeling by proposing a multiplicative specification for returns, incorporating autoregressive stochastic volatility, diurnal patterns, and announcement effects. Their Bayesian estimation using 5-minute S&P 500 futures data confirmed the significance of all three components. This work laid the foundation for further advances, particularly with ultra-high-frequency (UHF) data.

For example, Bekierman and Gribisch (2021) and Watanabe and Nakajima (2024) adopt the same multiplicative specification but introduce alternative dynamic models for high-frequency volatility. However, while RV improves daily volatility estimation, constructing high-frequency models from 5-minute returns may be less effective than using 5-minute spot volatility estimates derived from UHF data. This paper aims to develop high-frequency volatility models using such estimates.

Recent years have seen significant progress in spot volatility estimation. Foster and Nelson (1996) introduced nonparametric estimators for diffusion models, while Kristensen (2010) proposed a kernel-weighted IV estimator with vanishing bandwidth. Zu and Peter Boswijk (2014) advanced the field with an estimator building on Zhang, Mykland, and Ait-Sahalia (2005) and Mykland and Zhang (2008), establishing asymptotic theory under the assumption of an increasing block size k . Departing from this, Bollerslev, Li, and Liao (2021) developed a fixed- k inference framework, where estimation error follows a scaled chi-squared distribution.

Building on Bollerslev, Li, and Liao (2021), we propose a class of high-frequency spot volatility models within a nonlinear, non-Gaussian state-space framework. The observation equation links the fixed- k spot volatility estimator to the true latent volatility, ensuring theoretical rigor.

We develop a Bayesian estimation approach via Markov Chain Monte Carlo (MCMC), enabling parameter estimation, latent volatility extraction, and model comparison. The posterior mean of the latent volatility serves as a smoothed estimate. For model selection, we design a particle filter to approximate the marginal likelihood, facilitating comparisons via information criteria. Additionally, this filter facilitates one-step-ahead volatility forecasting.

The paper proceeds as follows: Section 2 reviews the fixed- k framework and presents our proposed models. Section 3 details the Bayesian methodology, while Section 4 and

Section 5 cover simulations and empirical applications. Section 6 concludes. Technical details are deferred to the Appendix. Throughout the paper, for two random sequences a_n and b_n , we write $a_n \asymp b_n$, if $a_n/C \leq b_n \leq Ca_n$ for some finite constant $C \geq 1$.

2. Model Setup

2.1. Individual asset price process and spot volatility

Before we introduce our high-frequency volatility models, we first clarify some relevant mathematical notations and related concepts. All random variables are defined on a fixed (complete) probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with filtration $(\mathcal{F}_t)_{t \geq 0}$. Following Andersen et al. (2001b), we adopt the assumption that logarithmic asset prices follow a univariate diffusion. In particular, for the a specific individual asset, the logarithmic return is modeled as

$$p_t - p_{t-1} \equiv r_t = \int_{t-1}^t \mu_s ds + \int_{t-1}^t \sigma_s dW_s, \quad (1)$$

where W_s stands for the standard Wiener process and hence, the corresponding volatility measure is based on the quadratic variation process, denoted by Qvar_t , which yields

$$\text{Qvar}_t = [p, p]_t - [p, p]_{t-1} = \int_{t-1}^t \sigma_s^2 ds. \quad (2)$$

This is commonly referred to as the integrated volatility in the literature. It is broadly known in literature (as in Andersen et al., 2001b; Barndorff-Nielsen and Shephard, 2002) that the integrated volatility over non-trivial time interval, such as a day, is an important quantity of interest in finance. Many nonparametric estimators for daily IV have been proposed ever since then, and among them the most widely used estimator is the daily RV based on 5-minute returns.

With the increasing availability of data sampled at ultra high frequencies, how to estimate and forecast the spot volatility, that is σ_t^2 , has drawn a growing interest in the literature. Based on the mathematical foundation in extant literature, we focus on the

following model for (log) price process as in the literature,

$$X_t = X_0 + \int_0^t b_s ds + \int_0^t \sigma_s dW_s + J_t. \quad (3)$$

This is a continuous-time Itô semimartingale process with drift (b_s) , diffusion (σ_s) and jump (J_t) . Making inference of the time-varying σ_t^2 serves as one major target of the present paper.

2.2. Fixed- k inference for volatility

When the logarithmic price of an asset follows model (3), Jacod, Li, and Liao (2021) suggest a way to estimate “spot covariance” in the general multivariate setting, $c_t = \sigma_t \sigma_t^\top$, nonparametrically and uniformly as follows

$$\hat{c}_{n,j} \equiv \frac{1}{k_{n,j} \Delta_n} \sum_{i \in \mathcal{I}_{n,j}} \Delta_i^n X \Delta_i^n X^\top \mathbf{1}_{\{\|\Delta_i^n X\| \leq u_n\}}, \quad (4)$$

where

$$\begin{aligned} \Delta_n &: T/n \\ \Delta_i^n X &: X_{i\Delta_n} - X_{(i-1)\Delta_n} \\ u_n &: \text{truncation threshold satisfying } u_n \asymp \Delta_n^\varpi, \varpi \in (0, 1/2). \\ \mathcal{I}_{n,j} &: \text{set collecting indices of consecutive increments in } j\text{-th block,} \\ &\text{such that } \{1, \dots, n\} = \bigcup_{j=1}^{m_n} \mathcal{I}_{n,j} \text{ and } |\mathcal{I}_{n,j}| = k_{n,j}. \\ \mathcal{T}_{n,j} &: \text{Correspondingly, } [0, T] \text{ can be dissected as } [0, T] = \bigcup_{j=1}^{m_n} \mathcal{T}_{n,j} \\ t(n, j) &\equiv (\min \mathcal{I}_{n,j} - 1) \Delta_n, \end{aligned}$$

and

$$\mathcal{T}_{n,j} \equiv \begin{cases} [t(n, j), t(n, j+1)) & \text{if } 1 \leq j < m_n \\ [t(n, m_n), T] & \text{if } j = m_n. \end{cases}$$

Specifically, T can be interpreted as the total length of time or the number trading days so that $\Delta_n = T/n$ usually refers to the sampling interval and $k_{n,j}$ denotes the block size. While it is commonly assumed that $k_{n,j} \rightarrow \infty$ in the literature, Bollerslev, Li, and Liao

(2021) advocate a way of making inference for spot volatility with $k_{n,j} = k$ fixed. To see the link between the setting of Bollerslev, Li, and Liao (2021, henceforth BLL2021QE) and the model to be established in our paper, note that BLL2021QE set

$$\begin{aligned}\mathcal{I}_{n,j} &\equiv \{(j-1)k+1, \dots, jk\}, \\ \mathcal{T}_{n,j} &\equiv [(j-1)k\Delta_n, jk\Delta_n),\end{aligned}$$

which is a special case with $k_{n,j} = k$. Making the inference of spot volatility in a nonparametrically refers to the fixed- k inference for volatility. In the univariate case, $c_t = \sigma_t^2$ is estimated by $\hat{c}_{n,j}$ with the j -th local block size fixed such that $k_{n,j} = k$, that is,

$$\hat{c}_{n,t} \equiv \hat{c}_{n,j} = \frac{1}{k\Delta_n} \sum_{i \in \mathcal{I}_{n,j}} (\Delta_i^n X)^2 \mathbf{1}_{\{|\Delta_i^n X| \leq u_n\}}, \quad (5)$$

for $t \in \mathcal{T}_{n,j}$ and $j \in \{1, \dots, m_n\}$.

According to Theorem 9.3.2 of Jacod and Protter (2012), $k_{n,j} \rightarrow \infty$ and $k_{n,j}\Delta_n \rightarrow 0$ are needed to ensure the consistency of $\hat{c}_{n,t}$. The required conditions for the consistency is intuitive as they require the local estimation block contain an increasing number of observations (i.e. $k_{n,j} \rightarrow \infty$), while at the same time the size of local estimation block shrinks to zero asymptotically (i.e. $k_{n,j}\Delta_n \rightarrow 0$). However, the required conditions for ensuring the desired consistency of nonparametric spot volatility estimation (i.e., $\hat{c}_{n,t}$) are stringent and can hardly be met simultaneously in practice. Set against this background, the fixed- k inference with local block size fixed can alleviate this issue by focusing on a single asymptotic scheme that only requires $\Delta_n \rightarrow 0$, which is easy to be achieved in the high-frequency setting.

By setting the estimation block size k fixed, the resulting spot volatility estimator of is not consistent, but easy-to-calculate pointwise confidence intervals are available at any given point in time given the distribution that characterizes the ratio of the nonparametric estimator $\hat{c}_{n,t}$ to the true unobserved volatility process c_t . The main distribution theory associated with the fixed- k inference for spot volatility is established in BLL2021QE, which

is summarized as follows for discussion.

Theorem 1 (Bollerslev, Li, and Liao (2021)). *Suppose that ASSUMPTION 1 imposed in BLL2021QE holds. Then for any finite subset $\mathcal{M} \subseteq \{1, \dots, m_n\}$, there exists a collection of independent random variables $(\bar{S}_j)_{j \in \mathcal{M}}$ such that for any $j \in \mathcal{M}$ and $t \in \mathcal{T}_{n,j}$,*

$$\frac{\hat{c}_{n,t}}{c_t} - \bar{S}_j = O_p(\Delta_n^{(2-r)\varpi \wedge (1/2)}) = o_p(1), \quad (6)$$

where

$$\bar{S}_j = (k\Delta_n)^{-1} \sum_{i \in \mathcal{I}_{n,j}} (W_{i\Delta_n} - W_{(i-1)\Delta_n})^2$$

is a $\bar{\chi}_k^2$ -distributed random variable with $\bar{\chi}_k^2$ referring to the scaled chi-squared distribution such that

$$\bar{\chi}_k^2 \equiv Z_k/k, \text{ with } Z_k \sim \chi_k^2. \quad (7)$$

In companion with this definition, we have the scaled inverse-chi-squared distribution

$$\bar{\chi}_k^{-2} \equiv k/Z_k, \text{ with } Z_k \sim \chi_k^2. \quad (8)$$

This distribution theory suggests that the distribution corresponding to the induced noise can be characterized properly. There is an textbook treatment of the log chi-squared distribution (see Lee, 2012), $\ln \chi_k^2$, associated with $\ln Z_k$ by taking logs on both sides of $Z_k \sim \chi_k^2$ in (8) such that $\ln Z_k$ has following probability density function in analytical form (in logarithmic form),

$$\ln f(z) = -\frac{k}{2} \ln 2 - \ln \Gamma(k/2) + \frac{1}{2}kz - \frac{1}{2} \exp(z).$$

We will show how to use the derived distribution theory to establish nonlinear non-Gaussian state-space models for spot volatility. Before we proceed to the corresponding discussion, we first demonstrate why we establish a non-Gaussian state-space models by showing that the distribution of $\hat{c}_{n,t}/c_t$ is different from Gaussian distribution when the estimation block size k is small. For Z_k , we have $\mathbb{E}(Z_k) = k$ and $\text{Var}(Z_k) = 2k$, where $\mathbb{E}(\cdot)$ and $\text{Var}(\cdot)$ denote

the mean and variance operator respectively. Then for the scaled chi-squared distribution $\bar{\chi}_k^2$, we have $\mathbb{E}(\bar{\chi}_k^2) = \frac{1}{k}\mathbb{E}(Z_k) = 1$, and $\text{Var}(\bar{\chi}_k^2) = \frac{1}{k^2}\text{Var}(Z_k) = \frac{2}{k}$. We plot the density of $\bar{\chi}_k^2$ against the density of a normal distribution with mean $\mathbb{E}(\bar{\chi}_k^2) = 1$ and variance $\text{Var}(\bar{\chi}_k^2) = \frac{2}{k}$, for $k = 5, 10, 30, 60$, respectively, in Figure 1. The choice of $k = 5$ and $k = 10$ may correspond to the empirical scenario when the price data is sampled at 1-minute frequency while the researchers focus on the estimation of 5-minutes spot volatility and 10-minutes spot volatility. By contrast, the choice of $k = 30$ and $k = 60$ may correspond to the empirical scenario when the tick-by-tick returns are used while the researches focus on half-a-minute (30 seconds) spot volatility and 1-minute (60 seconds) spot volatility. The choice of $k = 5$ and $k = 10$ with 1-minute returns is more realistic, but the distribution of fixed- k spot volatility estimator deviates significantly from normal distribution with the same mean and variance. This fact motivates the establishment of nonlinear non-Gaussian state-space models.

Remark 1. *It is important to note that the nonparametric estimator as in (5) requires a selected sequence of truncation threshold u_n , which is usually selected via thumb rule, as suggested by BLL2021QE, based on the bipower variation estimator in Barndorff-Nielsen and Shephard (2004). Instead, one alternative for filtering price jumps in the fixed- k inference setting is considering a leave-one-out spot variance estimator for each local estimation window with k returns. To illustrate the key idea, we write the leave-one-out spot variance estimator as*

$$\hat{c}_{n,j}^{\text{leave-one}} = \frac{1}{(k-1)\Delta_n} \sum_{i^* \in \mathcal{I}_{n,j}^*} (\Delta_{i^*}^n X)^2, \quad (9)$$

where $\mathcal{I}_{n,j}^* = \{i^* \in \mathcal{I}_{n,j} : |\Delta_{i^*}^n X| \leq \max_{i \in \mathcal{I}_{n,j}} |\Delta_i^n X|\}$ and $|\mathcal{I}_{n,j}^*| = (k-1)$, given our assumption that within each local estimation block there exists at most one price-level jump. This assumption is reasonable in that for a fixed k , the length of the local estimation block, $k\Delta_n \rightarrow 0$ when $\Delta_n \rightarrow 0$. By construction, it is not hard to derive the distribution of

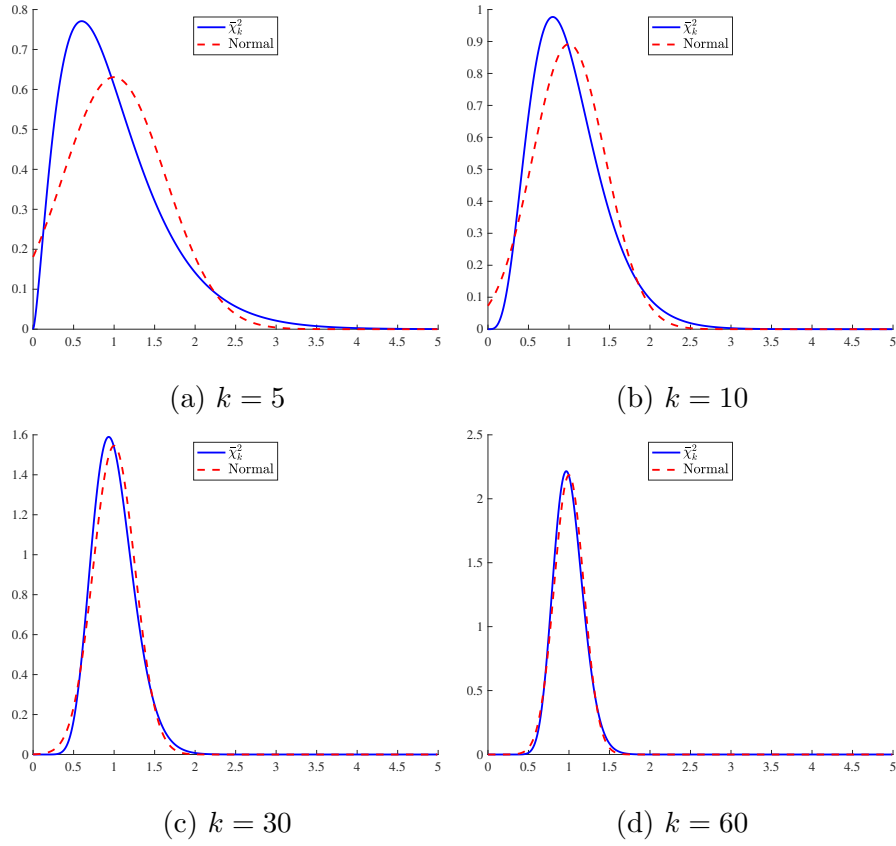


Figure 1: In each panel, we show the density of $\bar{\chi}_k^2$ (blue solid line) and the density of a normal distribution with the same mean and variance (red dashed line) for various values of the local estimation window size k .

$\hat{c}_{n,j}^{\text{leave-one}}$ from “volatility coupling” theory,

$$\frac{\hat{c}_{n,j}^{\text{leave-one}}}{c_t} \approx \frac{1}{k-1} \sum_{i=1}^{k-1} \zeta_{(i)} = \frac{1}{k-1} \left(\left(\sum_{i=1}^k \zeta_i \right) - \zeta_{(k)} \right),$$

where $\zeta_1, \dots, \zeta_k$ are i.i.d. χ_1^2 variables and $\zeta_{(1)} \leq \zeta_{(2)} \leq \dots \leq \zeta_{(k)}$ are the corresponding order statistics. We refer such a kind of construction of spot volatility estimate to the jump-robust spot volatility estimate. The corresponding approximation distribution $\frac{1}{k-1} \left(\left(\sum_{i=1}^k \zeta_i \right) - \zeta_{(k)} \right)$ does not have closed-form probability density function, but it is easy to simulate samples from $\frac{1}{k-1} \left(\left(\sum_{i=1}^k \zeta_i \right) - \zeta_{(k)} \right)$.

Within the fixed- k volatility inference modeling framework, the distribution usually

does not have a closed form or may have analytical forms with complicated expressions; however, one can easily obtain simulated samples of reasonably large size. Based on the simulated samples, we can approximate the probability density function using Gaussian Mixture Models (GMM). This GMM approximation method can also be applied in other fixed- k inference setting (e.g., the “Optimal CandlestickK” estimate in Li, Wang, and Zhang (2022), and more recently in Bollerslev et al. (2024a) and Bollerslev et al. (2024b), among others) to approximate the “gap” distribution that characterizes the ratio of nonparametric estimate of spot volatility and the true unobserved volatility process.

To sum up, the theoretical foundation for the fixed- k inference theory lies in the “volatility coupling” theory in Jacod, Li, and Liao (2021), from which many distributions of different variants can be derived, including the just mentioned “jump-robust” version, the “Optimal CandlestickK” version, and the most recent one in Bollerslev, Li, and Ren (2024), among others. These distributions may or may not have closed-form characterizations and we generally advocate GMM approximation of the simulated samples as the proxy distribution in our spot volatility model.

2.3. State-space models for spot volatility

Although unobserved spot volatility is indexed continuously in our model, to facilitate nonparametric estimation of spot volatility, following (5), we assume there exists a surjective function that maps $t \in [0, T]$ to $j \in \{1, \dots, m_n\}$. As implied by the fixed- k inference theory, instead of estimating $\ln(c_t)$ for $t \in \mathcal{T}_{n,j}$, $\ln(c_t)$ is approximated by a local constant $\ln(c_{n,j})$ for the j -th block with $j \in \{1, \dots, m_n\}$. In the state-space framework, therefore, we model blocks dynamics. To ensure our notations to be consistent with those in the literature (such as Chernov et al., 2003), we split each day into M disjoint blocks and the size of each block is fixed as k . If T represents the total number of trading days, then $n = k(MT)$ and $\Delta_n = T/n = 1/(kM)$. In this case, the total number of blocks for the

T trading days is $m_n = MT$. For instance, if the price data is sampled minute by minute within one day, and in each day trading hours start from 09:30 to 16:00. Besides, every 5 minutes are treated as a local estimation block for this case. Then by specification we have $T = 1$, $\Delta_n = 1/(6.5 \times 60) = 1/390$, $k = 5$, and $M = 78$.

Based on the fixed- k inference theory, we set up the following class of state-space models

$$\begin{cases} \ln(\hat{c}_{n,j}) = \ln(c_{n,j}) + \epsilon_j, & \epsilon_j \sim \ln \bar{\chi}_k^2 \text{ or GMM approximation,} \\ \ln(c_{n,j}) = \text{alternative models.} \end{cases} \quad (10)$$

$$(11)$$

Clearly, the observation equation comes from the fixed- k theory.

Since ϵ_j is not a Gaussian variable, all models in this class belong to the class of a nonlinear non-Gaussian state-space models.¹

In all alternative models, we assume that the log of the latent volatility process can be decomposed as

$$\ln(c_t) = \mu + h_t + s_t, \quad (12)$$

where h_t is a stochastic volatility process, s_t the intraday seasonal component. Correspondingly, for each block, we have the following decomposition

$$\ln(c_{n,j}) = \mu + h_j + s_j, \quad s_j = \tilde{s}(r), \quad j = 1, \dots, m_n, \quad (13)$$

where $r = (\text{mod}(j-1, M) + 1)/M$ and $\text{mod}(x, y)$ refers to the modulo operation that takes the remainder of any $x \in \mathbb{Z}$ divided by $y \in \mathbb{Z}$.

To capture the diurnal pattern, we assume $\tilde{s}(r)$ is a quadratic function. In particular, it is assumed that $\tilde{s}(r) = 12(1-b)\left(r - \frac{1}{2}\right)^2 + b$ as this is the only function within the class $f(r) = c(r-a)^2 + b$ that satisfies (i) $\int_0^1 (c(r-a)^2 + b) dr = 1$; (ii) $\text{argmin}_r c(r-a)^2 + b = \frac{1}{2}$.

¹As we have mentioned earlier, the distribution of ϵ_j does not necessarily have a closed-form characterization. GMM approximation ensures that the conditional likelihood can be calculated using the corresponding particle filter, allowing us to cast the established spot volatility model into a Gaussian state-space model so that the Kalman filter can be applied, which we will discuss later.

The first condition is imposed for identification. The second condition assumes that the diurnal pattern reaches the minimum in the middle of a trading day, an empirical regularity that has been found in the literature.² There is a restriction in using the quadratic function. That is, it implies a symmetric diurnal pattern. In a recent study, Christensen, Hounyo, and Podolskij (2018) propose a nonparametric method to estimate the diurnal pattern and find an asymmetric diurnal pattern. However, our approach can be easily extended to cover more complicated deterministic functions for diurnal pattern. Besides, we want to point out that given the function form of $\tilde{s}(r)$, b serves as the key parameter determining the intraday diurnal pattern.³

We impose AR(1) dynamic with a jumps structure on the volatility process. Specifically, we assume that the transition dynamics from the j -th block to the $(j+1)$ -th block follows

$$h_{j+1} = \phi h_j + e_j + J_j \eta_j, \quad e_j \sim \mathcal{N}(0, \sigma_e^2), \quad \eta_j \sim \mathcal{N}(\mu_\eta, \sigma_\eta^2), \quad (14)$$

where J_j is a jump indicator following Bernoulli distribution $\mathcal{Bern}(\kappa)$, defined by

$$J_j = \begin{cases} 1 & \text{with probability } \kappa \\ 0 & \text{with probability } 1 - \kappa, \end{cases}$$

with κ being the jump probability and η_j determining the jump size. Given the specified volatility process dynamics, we briefly discuss here why we advocate GMM to approximate the distribution of ϵ_j . In fact, using GMM to approximate a non-Gaussian random variable is a technique broadly adopted in literature; see, for example, Kim, Shephard, and Chib

²To satisfy the condition that $\arg\min_r c(r-a)^2 + b = \frac{1}{2}$, we have $c > 0$ and $a = \frac{1}{2}$. Substituting $a = \frac{1}{2}$ into $\int_0^1 (c(r-a)^2 + b) dr = 1$ yields $\frac{1}{12}c + b = 1$ and $c = 12(1-b)$. Thus, the quadratic function is uniquely determined by b . The larger the value of b is approaching 1, the less pronounced becomes the quadratic volatility pattern.

³The model can be extended to incorporate announcement effects, for which an MCMC algorithm can be designed to conduct posterior analysis.

(1998); Omori et al. (2007). The main advantage of using GMM is that one can cast the non-Gaussian state-space model into a standard linear Gaussian state-space model so that the Kalman filter along with the standard simulation smoother can be applied for extracting the latent stochastic spot volatility. Given the specification in (10) and (14), it is easy to note that both ϵ_j and $e_j + J_j \eta_j$ follow a mixture normal distribution. We introduce an auxiliary latent state variable $S_j \in \{1, \dots, K\}$ to indicate the mixture component of ϵ_j that is approximated through GMM.

$$f(\epsilon_j) = \sum_{\omega=1}^K q_{\omega} f_N(\epsilon_j \mid \tilde{m}_{\omega}, v_{\omega}^2), \quad (15)$$

where $f_N(\epsilon_j \mid \tilde{m}_{\omega}, v_{\omega}^2)$ denotes the ω -th Gaussian component with mean equal to $\tilde{m}_{\omega}$ and variance equal to v_{ω}^2 , assigned with probability q_{ω} . $\{q_{\omega}, \tilde{m}_{\omega}, v_{\omega}^2\}$ are determined by the EM algorithm. We demonstrate the GMM structure for $\ln \chi_5^2$ and $\ln \chi_{10}^2$ using Table 1 and Figure 1. We approximate $\ln \chi_5^2$ using a GMM with 7 components and $\ln \chi_{10}^2$ using 10 components. The detailed structure of the components is summarized in Table 1. Figure 1 shows that both approximations are accurate.

2.4. Alternative model specifications

In this section, we introduce the following alternative model specifications corresponding to (13) and (14).

- **Model 1**

$$\ln(c_{n,j}) = \mu + h_j,$$

$$h_{j+1} = \phi h_j + e_j, \quad e_j \sim \mathcal{N}(0, \sigma_e^2).$$

- **Model 2**

$$\ln(c_{n,j}) = \mu + h_j,$$

$$h_{j+1} = \phi h_j + e_j + J_j \eta_j, \quad e_j \sim \mathcal{N}(0, \sigma_e^2), \quad \eta_j \sim \mathcal{N}(\mu_{\eta}, \sigma_{\eta}^2).$$

Table 1: GMM structure

ω	$k = 5$			$k = 10$		
	q_ω	$\tilde{m}_\omega$	v_ω^2	q_ω	$\tilde{m}_\omega$	v_ω^2
1	0.1231	0.9874	0.4817	0.0696	1.9946	0.2476
2	0.1943	1.5760	0.2741	0.0973	2.2823	0.1570
3	0.1417	1.1123	0.4403	0.0842	2.1102	0.2131
4	0.1839	1.7175	0.2223	0.0941	2.2082	0.1816
5	0.0469	0.2557	0.7775	0.0927	2.3577	0.1318
6	0.1696	1.3002	0.3742	0.0277	1.5782	0.3603
7	0.1407	1.8683	0.1677	0.0903	2.1651	0.1957
8				0.0799	2.4201	0.1117
9				0.0950	2.3374	0.1385
10				0.0506	1.8391	0.2891
11				0.0969	2.2597	0.1645
12				0.0614	2.4732	0.0954
13				0.0604	1.9219	0.2676

• **Model 3**

$$\ln(c_{n,j}) = \mu + h_j + s_j,$$

$$h_{j+1} = \phi h_j + e_j + J_j \eta_j, \quad e_j \sim \mathcal{N}(0, \sigma_e^2), \quad \eta_j \sim \mathcal{N}(\mu_\eta, \sigma_\eta^2),$$

$$s_j \equiv \tilde{s}(r) = 12(1-b) \left(r - \frac{1}{2}\right)^2 + b, \quad r = (\text{mod}(j-1, M) + 1)/M.$$

We focus on these model specifications for the following reasons. First, **Model 1** is a straightforward extension of the conventional SV autoregressive process to model the “block” dynamics in a high-frequency setting. The terminology “block” refers to each fixed- k local estimation block for spot volatility. Second, Watanabe and Nakajima (2024) adopted a similar multiplicative specification for the volatility process but did not model the jump effects on the volatility process, as they found that jumps in the volatility process have marginal impact on both in-sample fit and out-of-sample forecast gain. However, the observation equation in the state-space model of this paper is different from that in Watanabe and Nakajima (2024). We aim to model the jump effects potentially caused by

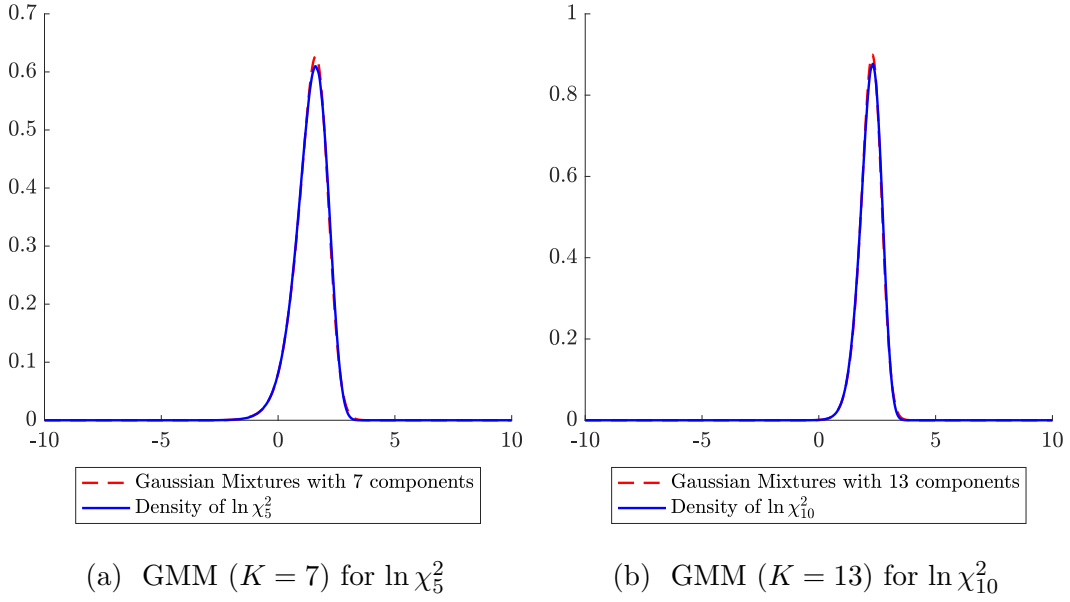


Figure 2: In each panel we plot the density, we plot the density of $\ln \chi_k^2$ (blue solid line) and the density of the corresponding Gaussian mixtures (red dashed line).

system shocks and, therefore, specify the jump components of the volatility process on top of the basic specification of **Model 1** to establish **Model 2** for modeling the spot volatility dynamics. Third, **Model 3** incorporates the basic structures of both **Model 1** and **Model 2** by adding an additional layer of parameter specifications to model the intraday diurnal pattern of spot volatility. In fact, the multiplicative specification of the latent volatility process can be regarded as a superposition of different components. Therefore, the approach for modeling the intraday diurnal pattern can also be extended to other parametric specifications with more complex parameter structures (for instance, modeling announcement effects using a parameterized exponential distribution with proper decay properties), with the corresponding parameters sampled using MCMC. We can complete this extension when necessary. Although the basic multiplicative specifications for the volatility process are adapted from Stroud and Johannes (2014), they did not provide simulation evidence for model comparison performance based on different information criteria. Our simulation studies fill this gap by providing a simulation evidence.

It is obvious that **Model 3** shares the same structure as in (13) and (14) by incorporating jumps on volatility and volatility diurnal pattern simultaneously. In comparison to **Model 3**, **Model 1** has the most parsimonious model structure whereas **Model 2** keeps the single-factor structure with Bernoulli type jumps specifications on volatility process.

To compare models across different specifications, we mainly use two methods: Deviance Information Criterion (DIC, Spiegelhalter et al., 2002, 2014) and marginal-likelihoods. DIC is a popular method for model selection when MCMC output is ready. DIC has a few nice features. First, DIC applies to a wide range of statistical models. Second, it does not suffer from Jeffreys-Lindley-Barlett's paradox. Third, it can be obtained even under improper priors. Finally, Li et al. (2025) justify DIC by showing that DIC is an asymptotically unbiased estimator of the Kullback-Leibler divergence between the data generating process and the plug-in predictive distribution. Besides, there have been studies of volatility using relatively low frequency data demonstrating the performance of DIC (Berg, Meyer, and Yu, 2004).

DIC is given as follows

$$\text{DIC} = D(\bar{\boldsymbol{\theta}}) + 2P_D, \quad (16)$$

where

$$\begin{aligned} D(\boldsymbol{\theta}) &= -2 \ln p(\mathbf{y} \mid \boldsymbol{\theta}), \\ P_D &= -2 \int [\ln p(\mathbf{y} \mid \boldsymbol{\theta}) - \ln p(\mathbf{y} \mid \bar{\boldsymbol{\theta}})] p(\boldsymbol{\theta} \mid \mathbf{y}) d\boldsymbol{\theta}, \end{aligned}$$

$\bar{\boldsymbol{\theta}}$ is the posterior mean of parameter $\boldsymbol{\theta}$, and $\mathbf{y}$ denotes observable data (i.e., fixed- k non-parametric estimate of spot volatility in this setting). One should note that DIC captures trade-off between model fit and complexity, similar to other widely used information criteria such as Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). When making model comparison using DIC, models with smaller DIC values are preferred.

Alternatively, we can calculate the harmonic mean of the log-likelihoods associated with $\boldsymbol{\theta}$ from MCMC runs for each model and make model comparison based on that. This

procedure is similar to the original idea in Newton and Raftery (1994) that uses harmonic mean approximation (HMA) to calculate Bayes factor. The reason for using the Bayes factor for model selection is that it is designed to favor the model that is associated with the true data generating process (DGP), as opposed to DIC, which mainly targets model fitting performance. To formally calculate the Bayes factor, one needs to calculate the marginal likelihood, i.e. $m(\mathbf{y})$, associated with each model. Given the fundamental Bayes theorem, we have

$$m(\mathbf{y}) = \frac{p(\mathbf{y} | \boldsymbol{\theta})\pi(\boldsymbol{\theta})}{p(\boldsymbol{\theta} | \mathbf{y})},$$

where $\pi(\boldsymbol{\theta})$ denotes the prior structure imposed on each model. Taking logarithms on both sides obtains the expression for the logarithm of marginal likelihood,

$$\ln m(\mathbf{y}) = \ln p(\mathbf{y} | \boldsymbol{\theta}) + \ln \pi(\boldsymbol{\theta}) - \ln p(\boldsymbol{\theta} | \mathbf{y}). \quad (17)$$

Chib (1995) and Chib and Jeliazkov (2001) suggest calculating $\ln m(\mathbf{y})$ by evaluating the right-hand side of (17) at an appropriate single point $\boldsymbol{\theta}^*$ that has high probability density in the support of posterior. For the state-space model, one can calculate $\ln p(\mathbf{y} | \boldsymbol{\theta}^*)$ via the auxiliary particle filter discussed in the appendix. By the law of total probability $p(\boldsymbol{\theta}^* | \mathbf{y})$ can be decomposed as

$$p(\boldsymbol{\theta}^* | \mathbf{y}) = p(\theta_1^* | \mathbf{y})p(\theta_2^* | \mathbf{y}, \theta_1^*) \cdots p(\theta_{q_\theta}^* | \mathbf{y}, \theta_1^*, \dots, \theta_{q_\theta-1}^*),$$

where $\boldsymbol{\theta}^* = (\theta_1, \dots, \theta_{q_\theta})$ with q_θ being the dimension of $\boldsymbol{\theta}^*$. Given the property of MCMC run, the decomposition of $p(\boldsymbol{\theta}^* | \mathbf{y})$ can be approximated via a sequence of reduced MCMC run (Chib and Jeliazkov, 2001). Accordingly, one can use $\bar{\boldsymbol{\theta}}$ to approximate $\boldsymbol{\theta}^*$ and calculate $\ln p(\mathbf{y} | \boldsymbol{\theta}^*)$, $\ln \pi(\boldsymbol{\theta}^*)$, and $\ln p(\boldsymbol{\theta}^* | \mathbf{y})$ respectively to obtain $\ln m(\mathbf{y})$. Once $\ln m(\mathbf{y})$ is established, we can compare models by comparing marginal likelihoods.

To compute DIC and marginal likelihoods, one needs to calculate the log-likelihood $\ln(\mathbf{y} | \boldsymbol{\theta})$. For our proposed spot volatility models, we design an auxiliary particle filter to calculate the log-likelihood. Note that the computational burden arises from the need

to calculate the log-likelihood for each MCMC posterior sample of $\boldsymbol{\theta}$. To overcome this computational burden, we implement this part of the algorithm in a parallel scheme using `OpenMP`. The details of the corresponding algorithm are summarized in Appendix B. In the simulation study, we will examine and compare the performance of these two methods.

3. Bayesian and MCMC Analysis

With the state-space model summarized in (10) and (11), we design an MCMC algorithm to estimate all the parameters involved (especially the log spot volatility that is latent). MCMC, as the leading modern Bayesian technique, is suitable for analyzing state-space models. Before we formally discuss the MCMC steps, we first fix notations and then briefly discuss the priors we will use. We collect all parameters that specify the state-space model into $\boldsymbol{\theta}$, that is, $\boldsymbol{\theta} = (\phi, \mu, \sigma_e^2, \kappa, \mu_\eta, \sigma_\eta^2, b)$. Let $\mathbf{h}$ denote the sequence of h_j , $\mathbf{S}$ denote the sequence of the auxiliary latent state variables S_j , $\mathbf{J}$ denote all the jump indicators, and $\boldsymbol{\eta}$ denote all the jump sizes. For simplicity, we let $\tilde{\mathbf{h}} = \mathbf{h} + \mu$.⁴

To conduct the Bayesian analysis, we choose the following prior distributions. For the persistence parameter ϕ and the parameter κ that determines the jump probability, we assume Beta prior distributions. For μ and the parameter μ_η that denotes the jump magnitudes, we assume normal prior distributions. The prior distributions of σ_e^2 and σ_η^2 are chosen to be inverse-gamma. Finally, for the parameter b that determines the intraday diurnal pattern of volatility, we impose the truncated normal distribution as the prior distribution. To sum up, we have following prior distributions: $\frac{\phi+1}{2} \sim \mathcal{B}(\alpha_\phi, \beta_\phi)$, $\mu \sim \mathcal{N}(\tilde{\mu}, \tilde{\sigma}^2)$, $\sigma_e^2 \sim \mathcal{IG}(\alpha_{\sigma_e}, \beta_{\sigma_e})$, $\kappa \sim \mathcal{B}(\alpha_\kappa, \beta_\kappa)$, $(\mu_\eta, \sigma_\eta^2) \sim \mathcal{NIG}(\tilde{\mu}_\eta, \lambda_{\sigma_\eta}, \alpha_{\sigma_\eta}, \beta_{\sigma_\eta})$,

⁴Given this specification, the transition dynamics from the j -th block to the $(j+1)$ -th block can be alternatively expressed as

$$\tilde{h}_{j+1} = \phi \tilde{h}_j + (1 - \phi)\mu + e_j + J_j \eta_j,$$

which would be useful for the design of MCMC sampler; see Appendix A.1.

$b \sim \mathcal{TN}(\tilde{\mu}_b, \tilde{\sigma}_b^2, \alpha_b, \beta_b)$, where $\mathcal{B}$ denotes the beta distribution, $\mathcal{N}$ denotes the normal distribution, $\mathcal{IG}$ denotes the inverse-gamma distribution, $\mathcal{NIG}$ denotes the normal-inverse-gamma distribution, and $\mathcal{TN}$ denotes the truncated normal distribution, and

$$\alpha_\phi, \beta_\phi, \tilde{\mu}, \tilde{\sigma}^2, \alpha_{\sigma_e}, \beta_{\sigma_e}, \alpha_\kappa, \beta_\kappa, \tilde{\mu}_\eta, \lambda_{\sigma_\eta}, \alpha_{\sigma_\eta}, \beta_{\sigma_\eta}, \tilde{\mu}_b, \tilde{\sigma}_b^2, \alpha_b, \beta_b$$

are all the corresponding hyperparameters. Given the prior specifications, the MCMC loop is summarized as follows,

1. Initialize $\boldsymbol{\theta}, \mathbf{S}, \mathbf{J}$.
2. Sample $\mathbf{h} = \{h_j\} \mid \{\ln(\hat{c}_{n,j})\}, \boldsymbol{\theta}, \mathbf{S}, \mathbf{J}$.
3. Sample $\mathbf{S} \mid \{\ln(\hat{c}_{n,j})\}, \mathbf{h}, \boldsymbol{\theta}, \mathbf{J}$.
4. Sample $\mathbf{J} \mid \mathbf{h}, \boldsymbol{\theta}$.
5. Sample $\boldsymbol{\eta} \mid \mathbf{h}, \boldsymbol{\theta}, \mathbf{J}$.
6. Sample $\boldsymbol{\theta} \mid \{\ln(\hat{c}_{n,j})\}, \mathbf{h}, \mathbf{S}, \mathbf{J}$.
7. Go to 2.

Iterations over step 2-7 consists of a complete sweep of MCMC sampler. The detailed description of the algorithm is presented in Appendix A. We implement the algorithm using MATLAB and C++ with Eigen (<http://eigen.tuxfamily.org>).

With the extracted latent spot volatility and the corresponding parameters $\boldsymbol{\theta}$ fixed at the posterior mean of MCMC samples, we can further design an auxiliary particle filter to forecast spot volatility. Since our analysis mainly builds in the high-frequency setting, we focus on one-step-ahead forecasting. That is, we forecast $c_{n,j+1}$ based on the established spot volatility model and data $\hat{c}_{n,1:j} := \{\hat{c}_{n,1}, \dots, \hat{c}_{n,j}\}$, which is the fixed- k nonparametric spot volatility estimation updated till the j -th local estimation window. More specifically, one can use the predictive distribution, $p(c_{n,j+1} \mid \hat{c}_{n,1:j}, \boldsymbol{\theta})$ from the auxiliary particle filter

to construct the one-step ahead spot volatility forecasting (posterior mean or posterior mode of the predictive distribution) and the posterior forecasting confidence intervals.

4. Simulation Studies

4.1. Parameter estimation and extracting volatility

We simulate data (i.e., the (log) price process) from three different data generating processes (DGPs) with the (log) volatility process being modeled as either **Model 1**, or **Model 2**, or **Model 3**. The fixed- k inference theory and “volatility coupling” theory guarantees that as long as the specification of the (log) price process follows Itô semimartingale process with drift, diffusion and jump, the corresponding distribution theory for the nonparametric estimates of spot volatility applies when the local estimation block is fixed. Let us denote the three different DGPs by **DGP 1**, **DGP 2**, and **DGP 3**, respectively, and summarize them as follows.

- **DGP 1**

$$\begin{aligned} dX_t &= \exp\left(\tilde{h}_t/2\right) dW_t, \\ \tilde{h}_t &= \mu_h + h_t, \\ dh_t &= -\kappa_h h_t dt + \sigma_h \left(\rho dW_t + \sqrt{1 - \rho^2} dB_t\right). \end{aligned}$$

- **DGP 2**

$$\begin{aligned} dX_t &= \exp\left(\tilde{h}_t/2\right) dW_t, \\ \tilde{h}_t &= \mu_h + h_t, \\ dh_t &= -\kappa_h h_t dt + \sigma_h \left(\rho dW_t + \sqrt{1 - \rho^2} dB_t\right) + \mathbf{1}_{\{t=t^\circ\}} J_t \eta_t, \\ J_t &\sim \text{Bern}(\kappa), \\ \eta_t &\sim \mathcal{N}(\mu_\eta, \sigma_\eta^2). \end{aligned}$$

• DGP 3

$$\begin{aligned}
dX_t &= \exp\left(\tilde{h}_t/2\right) dW_t, \\
\tilde{h}_t &= \mu_h + h_t + s_t, \\
dh_t &= -\kappa_h h_t dt + \sigma_h \left(\rho dW_t + \sqrt{1-\rho^2} dB_t\right) + \mathbf{1}_{\{t=t^\circ\}} J_t \eta_t, \\
J_t &\sim \text{Bern}(\kappa), \\
\eta_t &\sim \mathcal{N}(\mu_\eta, \sigma_\eta^2), \\
s_t &= 12(1-b) \left(t - \lfloor t- \rfloor - \frac{1}{2}\right)^2 + b.
\end{aligned}$$

In the description of DGPs above we introduce the notation t° such that $t^\circ \in [0, T]$, $r^\circ \in \{1/M, 2/M, \dots, M/M = 1\}$, and $t^\circ = \lfloor t- \rfloor + r^\circ$. Meanwhile, $\lfloor t- \rfloor$ denotes the greatest integer less than t and $\mathbf{1}_{\{\cdot\}}$ refers to an indicator function. In the high-frequency setting, we focus on the spot volatility within one day, that is, the intraday volatility in a trading day of each month. Specifically, we assume that in each trading day there are 6.5 trading hours from 09:30 to 16:00 and we consider one-minute returns. Therefore, there are $6.5 \times 60 = 390$ minutes and $\Delta_n = 1/390$ by construction. Besides, we specify $\mu_h = -6.2$.⁵

Remark 2. *Specifying $\mathbb{E}[W_t B_t] = \rho = 0$ implies that in the state-space model that we have established, our main target is to extract the underlying time-varying volatility process and temporarily ignore the correlation between the Brownian motion that drives the (log) price process and the Brownian motion that drives the spot volatility process. Although the ignorance of correlation between error terms may lead to the misspecification of models, if the goal is shifted towards forecasting while admitting model being misspecified, the leverage effect issue may be of secondary concern.*

Remark 3 (From DGP Dynamic to Block Dynamic). *We discuss the dynamics of SV autoregressive process using **Model 1**. In **Model 1**, if $\kappa_h > 0$, h_t is a stationary*

⁵ $\mu_h = -6.2$ is prior mean specified as in Stroud and Johannes (2014) for the mean of the logarithm of volatility process, specifically, the equation (2) and appendix A of Stroud and Johannes (2014).

Ornstein–Uhlenbeck process. At interval $(t, t + dt)$, its Euler scheme approximation is

$$h_{t+dt} \approx (1 - \kappa_h dt)h_t + \sigma_h(B_{t+dt} - B_t) := \phi_h h_t + \varepsilon_{t+dt}.$$

Given the property of Brownian motion, $\varepsilon_{t+dt} \sim \mathcal{N}(0, \sigma_h^2 dt)$. Data is generated from this continuous setting. To apply nonparametric estimation of spot volatility, we need to select k consecutive time intervals ($dt \equiv \Delta_i^n$) to construct “local estimation window”. This implies that we need to move from “observation” dynamics to “block” dynamics. If we simulate data from the system above, we have

$$h_{t+kdt} \approx \phi_h h_{t+(k-1)dt} + \varepsilon_{t+kdt} \approx \dots \approx \phi_h^k h_t + (\phi_h^{k-1} \varepsilon_{t+dt} + \dots + \varepsilon_{t+kdt}).$$

Thus, the “block” dynamics is given by

$$h_{j+1} \approx \phi_h^k h_j + e_{j+1}, \quad (18)$$

where $e_{j+1} = \phi_h^{k-1} \varepsilon_{t+dt} + \dots + \varepsilon_{t+kdt}$. The variance of $e_{j+1} = \phi_h^{k-1} \varepsilon_{t+dt} + \dots + \varepsilon_{t+kdt}$ is

$$\sigma_e^2 = \frac{\sigma_h^2 dt (1 - \phi_h^{2k})}{1 - \phi_h^2}.$$

Since $\varepsilon_t, \dots, \varepsilon_{t+kdt}$ are i.i.d. and follow Gaussian distribution. Given the property of Gaussian distribution, for finite fixed k , e_{j+1} follows Gaussian distribution. This justifies the use of Gaussian distribution to model transition dynamic of volatility process in

Model 1-3. For instance, if $dt := \Delta_i^n = 1/390$, $\sigma_h = 1.2$, $k = 5$, and $\kappa_h = 2$, then $\phi_h = 1 - \kappa_h dt \approx 0.9949$, $\phi_h^k \approx 0.9746$ and

$$\sqrt{\sigma_e^2} = \sqrt{\frac{\sigma_h^2 dt (1 - \phi_h^{2k})}{1 - \phi_h^2}} \approx 0.1345.$$

By comparison, If $dt := \Delta_i^n = 1/23400$, $\sigma_h = 1.2$, $k = 30$, and $\kappa_h = 20$, then $\phi_h = 1 - \kappa_h dt \approx 0.9991$, $\phi_h^k = 0.9747$ and

$$\sqrt{\sigma_e^2} = \sqrt{\frac{\sigma_h^2 dt (1 - \phi_h^{2k})}{1 - \phi_h^2}} \approx 0.0424.$$

Table 2: Posterior results from MCMC under the three DGPs

	True Value	Posterior Mean	Posterior Std dev.	2.5%	97.5%	Inefficiency
DGP 1						
ϕ	0.9746	0.9467	0.0134	0.9175	0.9701	2.5385
σ_e	0.1345	0.1369	0.0159	0.1084	0.1715	3.5158
μ	-6.2000	-6.2556	0.0666	-6.3881	-6.1217	1.0953
DGP 2						
ϕ	0.9746	0.9598	0.0093	0.9404	0.9768	1.9338
σ_e	0.1345	0.1668	0.0157	0.1378	0.1998	3.3181
μ	-6.2000	-6.1938	0.1150	-6.4283	-5.9740	1.0272
κ	0.0047	0.0060	0.0019	0.0029	0.0101	1.1570
σ_η	1.2000	1.1112	0.3587	0.6621	2.0415	1.3829
μ_η	0.8000	0.7518	0.2852	0.1993	1.3142	1.1296
DGP 3						
ϕ	0.9746	0.9567	0.0104	0.9350	0.9759	1.8683
σ_e	0.1345	0.1678	0.0162	0.1374	0.2006	3.1977
μ	-6.2000	-6.1454	0.1056	-6.3594	-5.9430	1.0180
κ	0.0047	0.0056	0.0018	0.0028	0.0098	0.7970
σ_η	1.2000	1.0186	0.3707	0.5858	1.8983	1.1014
μ_η	0.8000	0.7520	0.1508	0.4578	1.0527	0.7298
b	0.7000	0.7561	0.0497	0.6582	0.8532	1.3300

In Table 2, we summarize the posterior estimation results for various DGPs. For **DGP 3**, we set $b = 0.7$. In each MCMC procedure, we run 110,000 iterations, discarding the first 10,000 draws. The 2.5% and 97.5% columns in Table 2 refer to the 2.5% and 97.5% percentiles corresponding to the MCMC outputs. It is obvious that the credible intervals generally cover the true values of the corresponding parameters. The last column of Table 2 reports the inefficiency factors, based on 100 lags of the autocorrelation functions, indicating our MCMC algorithm is very efficient. In addition, we plot the MCMC samples, the corresponding posterior densities, and their sample autocorrelation functions in **DGP 1**

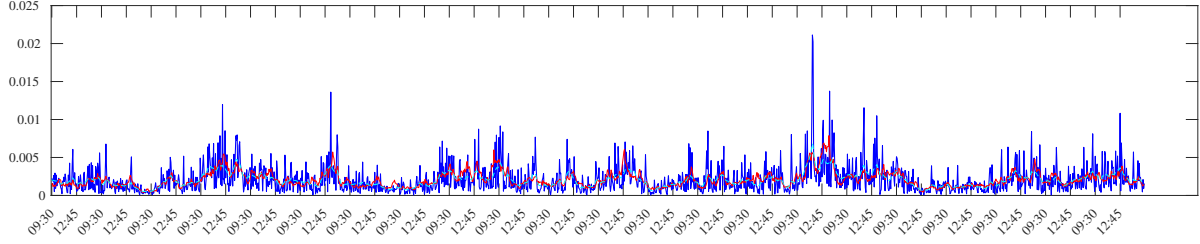
to **DGP 3**. Figure C.1 to Figure C.3, respectively. All the plots suggest that our MCMC samples converge well and our proposed method is effective.

Our method allows us to extract spot volatility by estimating $\mathbb{E} \left(\exp(\tilde{h}_t/2) | \mathcal{F}_T \right)$. This is the smoothed estimate of spot volatility. To see the effectiveness of our spot volatility estimate, in Figure 3 we plot the fixed- k nonparametric estimate of the spot volatility (the blue solid line), the true spot volatility (the red solid line), and the smoothed estimate of spot volatility (the cyan dashed line). It is clear the smoothed estimate of spot volatility is very close to the true spot volatility and much smoother than the fixed- k nonparametric estimate of the spot volatility.

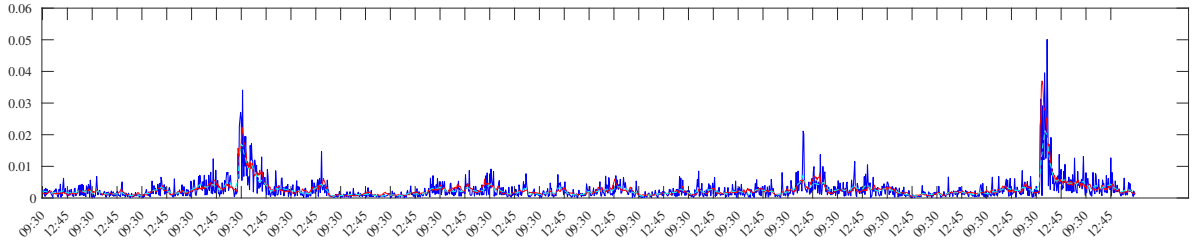
4.2. Model selection

In this part, we discuss the effectiveness of using DIC and marginal likelihoods for model comparison and selection. As we have discussed previously, one should expect that the marginal likelihood method chooses the correct model. However, model selection based on DIC does not guarantee the selection of the true model, since DIC aims to select a model that minimizes the KL “distance” between the DGP and the plug-in predictive distribution asymptotically.

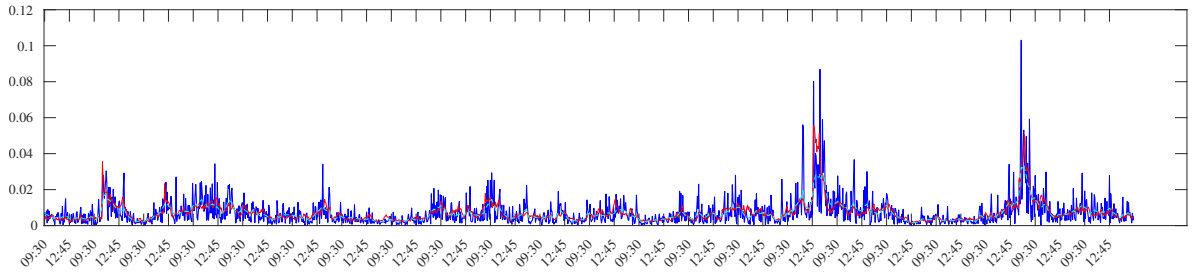
We consider **DGP 1**, **DGP 2**, and **DGP 3** that match **Model 1**, **Model 2**, and **Model 3**. For **DGP 3**, two values for b are considered, $b = 0.7$ and $b = 0.3$. Specifically, we focus on 1-minute returns over daily trading hours from 09:30 to 16:00, assuming 22 trading days within one month, and the local estimation window size is fixed at $k = 5$. Given this specification, we have $T = 22$, $\Delta_n = 1/390$, and accordingly, a total of $m_n = T/(k\Delta_n) = 1716$ local nonparametric estimation blocks, which serve as the observations in the observation equation of the state-space model (More detailed descriptions for this scheme in section 2.3). For each DGP, we simulate a sample path, we then estimate **Model 1**, **Model 2**, and **Model 3** by the MCMC algorithm and calculate DIC and



(a) DGP 1



(b) DGP 2



(c) DGP 3

Figure 3: In each figure we plot the fixed- k nonparametric estimate of the spot volatility (the blue solid line), the true spot volatility (the red solid line), and the smoothed estimate of spot volatility (the cyan dashed line).

marginal likelihood for each model. Finally, we repeat the exercise for 100 times and then calculate the averaged DIC value and the average log marginal likelihood value across 100 replications. Note that DIC prefers a model with a smaller value while marginal likelihood prefer a model with a larger value.

Table 3 reports the averaged DIC value (the left panel) and the average log marginal likelihood value (the right panel).⁶ For each DGP, the smallest DIC value and the largest marginal likelihood value are highlighted in boldface.

We first examine the performance of DIC for model comparison. It is known in the literature, a model with the lowest DIC does not necessarily match the DGP that generates the data. One can see from Table 3 that this is indeed the case. Although under **DGP 1**, **DGP 2**, **DGP 3'** the averaged DIC value is the smallest for the true model, it is not the smallest for the true model under **DGP 3**. In this case, **Model 1** has the smallest DIC value on average, followed by **Model 2**, while the true model, **Model 3**, has the largest DIC value on average. With a large value of b , in **DGP3**, the diurnal pattern is weak. The diurnal pattern attenuates as b draws closer to 1. This study suggests that DIC has a better chance to select correct model when the diurnal pattern is strong than when it is weak.

We then examine the performance of marginal likelihood for model comparison. However, for each DGP, it is expected that marginal likelihood would effectively select the correct model asymptotically. This is indeed reflected by the average log marginal likelihood values shown in Table 3. In each column of the right panel of Table 3, the model with the highest average log marginal likelihood matches the DGP.

Finally, we check the performance of the particle filter for generating forecasts of spot volatility. Figure 4 presents the nonparametric estimate of the spot volatility (the blue solid

⁶We conduct this simulation study on a Linux machine equipped with AMD Ryzen 9950X processor with 16 cores and clock speed of 5.20GHz. A similar Monte Carlo simulation design can also be seen in Li, Yu, and Zeng (2020).

Table 3: Model selection of alternative models by DIC and marginal likelihood under the three DGPs

	DGP 1	DGP 2	DGP 3 ($b = 0.7$)	DGP3' ($b = 0.3$)	DGP 1	DGP 2	DGP 3 ($b = 0.7$)	DGP3' ($b = 0.3$)
	DIC				Marginal Likelihood			
Model 1	3852.5	3937.3	3975.3	4072.8	-1923.6	-1960.5	-1985.9	-2035.6
Model 2	4222.3	3945.6	3977.4	4064.9	-1928.0	-1950.6	-1977.8	-2032.4
Model 3	4011.1	3917.8	3995.0	3972.3	-1930.7	-1953.4	-1959.6	-1960.4

line), the smoothed volatility estimate (the cyan dashed line), the true spot volatility (the red solid line), and the one-step-ahead forecast of volatility from the particle filter (the green solid line) based on simulated data. The gray shaded area refers to the posterior credible intervals for forecasted spot volatility. To demonstrate the performance of our methods for generating forecasts of intraday spot volatility, we simulate ultra-high-frequency tick-by-tick returns, i.e. $\Delta_n = 1/(6.5 \times 60 \times 60) = 1/23400$. We specify the local estimation window size as $k = 5$ and $k = 10$, which correspond to 5-seconds intervals and 10-seconds intervals, respectively. This means that, for generating forecasts of intraday spot volatility, the one-step ahead forecast refers to the next 5-second interval for $k = 5$ and the next 10-second interval for $k = 10$. Meanwhile, we split the trading time in each trading day (09:30 to 16:00, 23400 discrete intervals in total) into two parts and use data collected in the first 70% to estimate model parameters using MCMC. Then we fix the model parameters at the posterior mean and apply particle filter over the remained 30% time to the end of the trading day.

Remark 4. *It is possible to consider $\Delta_n = 1/390$ as in the foregoing discussions. However, local nonparametric estimation based on data would be limited with $\Delta_n = 1/390$ in each trading day. For instance, with $\Delta_n = 1/390$ and $k = 5$, one can obtain $390/5 = 78$ local nonparametric estimations of spot volatility. Therefore, it is advisable to use data from several trading days to first estimate the model parameters, and then perform out-*

of-sample forecasting. Theoretically, one can follow an expanding window scheme and re-estimate model parameters when new return data become available, but this would increase the computational burden.

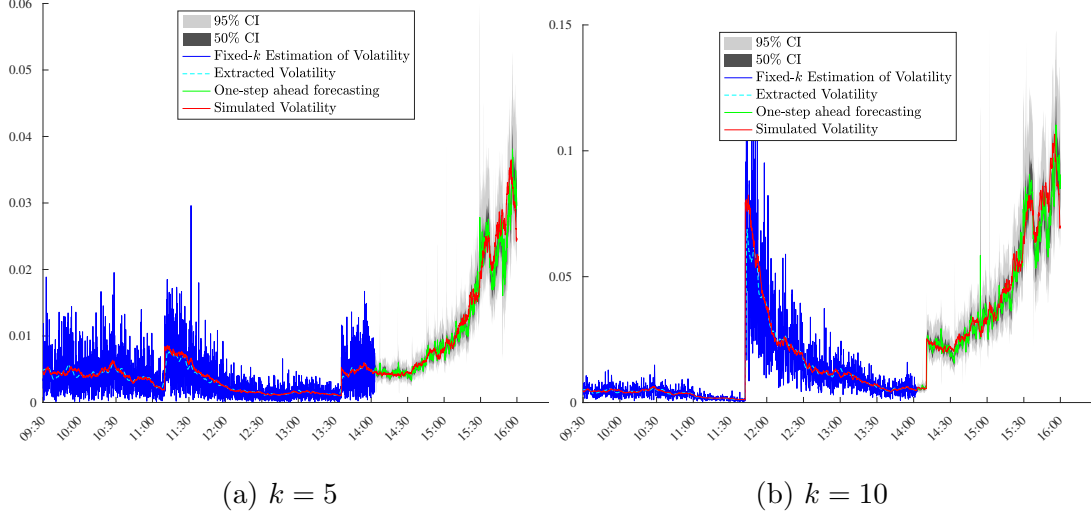


Figure 4: This figure presents the nonparametric estimate of the spot volatility (the blue solid line), the smoothed volatility estimate (the cyan dashed line), the true spot volatility (the red solid line), and the one-step-ahead forecast of volatility from the particle filter (the green solid line) based on simulated data. The gray shaded area refers to the posterior credible intervals for forecasted spot volatility.

5. Empirical Applications

The Bayesian approach we have proposed in this paper is readily applicable to investigate the intraday volatility pattern of the high-frequency factors in Aleti and Bollerslev (2024). Since Fama and French (1992) and Fama and French (1993), there has been a substantial amount of literature discussing factors for modeling equity returns, as they offer a parsimonious statistical description of the returns' cross-sectional dependence structure. Generally speaking, traded factors, broadly discussed in empirical finance literature, are mainly about

characteristic-sorted portfolios, where portfolios of assets are constructed based on cross-sectional characteristics of assets. For instance, the *size* factor in Fama and French (1992) is constructed by monthly sorting observed returns based on market capitalization, dividing assets into portfolios according to this characteristic, and then taking the difference in average returns across these portfolios. In the high-frequency setting, the high-frequency factors or high-frequency factor portfolios are still constructed using conventional methods (such as tercile sorting with monthly rebalancing and value weighting), but using price data, sampled at a 1-minute frequency, to construct high-frequency returns. This is in line with Aït-Sahalia, Jacod, and Xiu (2024), which advocates the use high-frequency data to re-investigate the conventional asset pricing topics. In Aleti (2023) and Aleti and Bollerslev (2024), the authors construct a novel dataset of high-frequency six factors in Fama and French (2015) based on common stocks on NYSE/NASDAQ/NYSEMKT. These factors are Mkt (market risk premium), SMB (size factor), HML (value factor), RMW (profitability factor), CMA (investment factor), and UMD (momentum factor). All these factors are constructed using methodologies that are faithful to the conventional papers, while utilizing high-frequency price data from the NYSE Trade and Quote Database (TAQ) through WRDS. This dataset contains 1-minute returns of these six factors for all common stocks listed on the three primary exchanges (NYSE, NASDAQ, NYSEMKT) from 09:30 to 16:00 each trading day, spanning from January 1996 to December 2020. In this section, we apply the Bayesian spot volatility modeling method to examine how the volatility patterns vary across these six high-frequency factors and which model specifications best fit or explain the data.

Table 4

	DIC					
	Mkt	SMB	HML	RMW	CMA	UMD
September, 1999						
Model 1	5043.0	4888.3	3971.6	3781.4	3835.1	3659.3
Model 2	5062.0	4901.0	3974.9	3782.6	3842.6	3663.3
Model 3	5010.2	4852.6	3909.1	3708.6	3740.5	3545.3
November, 2008						
Model 1	3702.9	3843.8	3656.6	3814.4	3882.3	3802.5
Model 2	21771.0	4018.6	3642.5	3800.3	3878.2	3781.1
Model 3	83919.0	56629.0	9766.5	12164.0	3822.3	5713.9
March, 2016						
Model 1	4397.7	4330.6	4546.6	4541.7	4458.0	4910.8
Model 2	4358.2	4309.0	4491.2	4485.7	4411.4	4877.2
Model 3	4340.5	4259.2	4505.2	4478.8	4386.9	4885.4
May, 2020						
Model 1	4257.0	4136.2	4288.7	4072.6	3902.2	4327.9
Model 2	4247.0	4133.1	4224.8	4025.3	3867.6	4242.2
Model 3	4197.6	4068.0	4231.4	3985.9	3819.3	4252.4

We calculate DIC and model marginal likelihood for each of these high-frequency factors in Table 4 and Table 5, respectively. Each of these two tables has four panels collecting results for four selected periods. The results in these two tables are rich for interpretation. First, since DIC targets model fitting performance while balancing model complexity, Table 4 shows that, for the Mkt and SMB factors, the spot volatility model specification with better fitting performance is not necessarily the most comprehensive one (**Model 3**). In comparison to September 1999 and the other two periods (March 2016 and May 2020), DIC favors a simpler specification (**Model 1** with a smaller DIC) for the spot volatility dynamics. For the HML factor, DIC favors **Model 2**, which does not directly model the polynomial intraday diurnal pattern in November 2008, March 2016, and May 2020, rather

Table 5

	Model Log-marginal Likelihood					
	Mkt	SMB	HML	RMW	CMA	UMD
September, 1999						
Model 1	-2647.8	-2451.6	-1985.3	-1887.3	-1915.8	-1827.7
Model 2	-2649.6	-2450.9	-1986.9	-1889.7	-1919.0	-1832.0
Model 3	-2590.5	-2428.6	-1955.2	-1855.8	-1872.2	-1775.0
November, 2008						
Model 1	-1831.8	-1934.1	-1820.3	-1899.1	-1943.9	-1900.3
Model 2	-1773.9	-1892.6	-1819.3	-1904.1	-1940.9	-1897.1
Model 3	-1720.3	-1818.4	-1794.6	-1859.3	-1909.8	-1867.2
March, 2016						
Model 1	-2197.8	-2162.1	-2281.7	-2274.6	-2236.9	-2462.3
Model 2	-2178.8	-2146.8	-2250.9	-2247.7	-2207.5	-2432.2
Model 3	-2170.2	-2135.0	-2255.1	-2251.6	-2202.9	-2443.9
May, 2020						
Model 1	-2128.5	-2070.9	-2139.2	-2037.2	-1957.0	-2166.3
Model 2	-2131.6	-2061.1	-2099.1	-2019.1	-1928.1	-2116.5
Model 3	-2101.8	-2036.4	-2109.2	-1998.5	-1911.0	-2130.2

than the nested **Model 3** as seen in September 1999. Similarly, for the RMW factor, DIC suggests the nested model specification (**Model 3**) during the three periods (September 1999, March 2016, and May 2020), while in November 2008, **Model 2**, as a simpler model specification, is preferred.

Second, as we previously demonstrated through simulation studies, model log-marginal likelihood targets the model with the best interpretation. Accordingly, Table 5 implies that the intraday volatility patterns of these six high-frequency factors are ubiquitous, but this is not necessarily always the case. There are a few exceptions in which, by comparing model log-marginal likelihoods, a more parsimonious model that does not directly model the intraday volatility pattern is selected. For instance, as shown in Table 5, **Model 2**,

rather than **Model 3**, matches the intraday volatility patterns of the HML, RMW, and UMD factors better in terms of interpretability. Finally, the results in Table 5 also directly imply that the diurnal intraday volatility patterns of high-frequency factors vary over time and differ across factors.

6. Conclusion

A key contribution of Bollerslev, Li, and Liao (2021) is the development of fixed- k inference theory for time-varying spot volatility. While fixed- k inference avoids the need for an expanding local estimation window as the sample size grows and provide a “finite sample” inferential procedure, it introduces inconsistency and noise in nonparametric spot volatility estimation. To mitigate these limitations, we propose a Bayesian framework with multiple model specifications for spot volatility in high-frequency data, grounded in fixed- k theory. All models are formulated as nonlinear non-Gaussian state-space systems, enabling Bayesian estimation, volatility extraction, and model comparison.

Our approach offers two major advantages. First, it significantly reduces the noise inherent in fixed- k spot volatility estimation in high-frequency settings. Second, the Bayesian framework naturally supports model estimation, volatility extraction, model selection, and comparison across competing specifications. Simulations demonstrate the effectiveness of our methods: the designed MCMC and particle filter algorithms accurately recover latent spot volatility, while model performance can be assessed via DIC or marginal likelihoods — depending on whether the focus is volatility fitting or pattern explanation.

Building on this theoretical and simulation-based foundation, we apply our Bayesian spot volatility framework to a novel high-frequency factors dataset, uncovering distinct volatility patterns across factors and over time. Empirically, the most interpretable specification typically includes volatility jumps and an intraday polynomial pattern. However, for pure fitting accuracy, simpler models occasionally outperform. Our framework is readily

extensible to other settings requiring high-frequency volatility analysis, though we defer such empirical applications to future work.

References

- AÏT-SAHALIA, YACINE, JEAN JACOD, AND DACHENG XIU (2024): “Continuous-time fama-macbeth regressions,” Manuscript, forthcoming for *The Review of Financial Studies*.
- ALETI, SAKETH (2023): “The high-frequency factor zoo,” Manuscript.
- ALETI, SAKETH, AND TIM BOLLERSLEV (2024): “News and asset pricing: A high-frequency anatomy of the sdf,” *The Review of Financial Studies*, 38, 712–759.
- ANDERSEN, TORBEN G., TIM BOLLERSLEV, FRANCIS X. DIEBOLD, AND HEIKO EBENS (2001a): “The distribution of realized stock return volatility,” *Journal of Financial Economics*, 61, 43–76.
- ANDERSEN, TORBEN G., TIM BOLLERSLEV, FRANCIS X. DIEBOLD, AND PAUL LABYS (2001b): “The distribution of realized exchange rate volatility,” *Journal of the American Statistical Association*, 96, 42–55.
- ANDERSEN, TORBEN G., TIM BOLLERSLEV, FRANCIS X. DIEBOLD, AND PAUL LABYS (2003): “Modeling and forecasting realized volatility,” *Econometrica*, 71, 579–625.
- BARNDORFF-NIELSEN, OLE E., AND NEIL SHEPHARD (2002): “Econometric analysis of realized volatility and its use in estimating stochastic volatility models,” *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 64, 253–280.
- (2004): “Power and Bipower Variation with Stochastic Volatility and Jumps,” *Journal of Financial Econometrics*, 2, 1–37.
- BEKIERMAN, JEREMIAS, AND BASTIAN GRIBISCH (2021): “A mixed frequency stochastic volatility model for intraday stock market returns*,” *Journal of Financial Econometrics*, 19, 496–530.

- BERG, ANDREAS, RENATE MEYER, AND JUN YU (2004): “Deviance information criterion for comparing stochastic volatility models,” *Journal of Business & Economic Statistics*, 22, 107–120.
- BOLLERSLEV, TIM (1986): “Generalized autoregressive conditional heteroskedasticity,” *Journal of Econometrics*, 31, 307–327.
- BOLLERSLEV, TIM, JIA LI, QIYUAN LI, AND YIFAN LI (2024a): “Optimal candlestick-based spot volatility estimation: New tricks and feasible inference procedures,” Manuscript.
- BOLLERSLEV, TIM, JIA LI, YIFAN LI, AND QIUSHI ZHANG (2024b): “Illuminating important economic news by candlesticks: Optimal testing meets technical analysis,” Manuscript.
- BOLLERSLEV, TIM, JIA LI, AND ZHIPENG LIAO (2021): “Fixed-k inference for volatility,” *Quantitative Economics*, 12, 1053–1084.
- BOLLERSLEV, TIM, JIA LI, AND YUEXUAN REN (2024): “Optimal inference for spot regressions,” *American Economic Review*, 114, 678–708.
- BOLLERSLEV, TIM, AND HAO ZHOU (2002): “Estimating stochastic volatility diffusion using conditional moments of integrated volatility,” *Journal of Econometrics*, 109, 33–65.
- CHERNOV, MIKHAIL, A. RONALD GALLANT, ERIC GHYSELS, AND GEORGE TAUCHEN (2003): “Alternative models for stock price dynamics,” *Journal of Econometrics*, 116, 225–257, *Frontiers of financial econometrics and financial engineering*.
- CHIB, SIDDHARTHA (1995): “Marginal likelihood from the gibbs output,” *Journal of the American Statistical Association*, 90, 1313–1321.

- CHIB, SIDDHARTHA, AND IVAN JELIAZKOV (2001): “Marginal likelihood from the metropolis-hastings output,” *Journal of the American Statistical Association*, 96, 270–281.
- CHRISTENSEN, KIM, ULRICH HOUNYO, AND MARK PODOLSKIJ (2018): “Is the diurnal pattern sufficient to explain intraday variation in volatility? a nonparametric assessment,” *Journal of Econometrics*, 205, 336–362.
- DOUCET, ARNAUD, AND ADAM JOHANSEN (2009): “A tutorial on particle filtering and smoothing: Fifteen years later,” *Handbook of Nonlinear Filtering*, 12.
- ENGLE, ROBERT (2004): “Risk and volatility: Econometric models and financial practice,” *The American Economic Review*, 94, 405–420.
- ENGLE, ROBERT F. (1982): “Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation,” *Econometrica*, 50, 987–1007.
- FAMA, EUGENE F., AND KENNETH R. FRENCH (1992): “The cross-section of expected stock returns,” *The Journal of Finance*, 47, 427–465.
- (1993): “Common risk factors in the returns on stocks and bonds,” *Journal of Financial Economics*, 33, 3–56.
- (2015): “A five-factor asset pricing model,” *Journal of Financial Economics*, 116, 1–22.
- FOSTER, DEAN P., AND DAN B. NELSON (1996): “Continuous record asymptotics for rolling sample variance estimators,” *Econometrica*, 64, 139–174.
- GATHERAL, JIM, THIBAUT JAISSON, AND MATHIEU ROSENBAUM (2018): “Volatility is rough,” *Quantitative Finance*, 18, 933–949.

- GIORDANI, PAOLO, MICHAEL PITT, AND ROBERT KOHN (2011): “Chapter 3 - bayesian inference for time series state space model,” in *The Oxford Handbook of Bayesian Econometrics* ed. by GEWEKE, JOHN, KOOP, GARY, and DIJK, HERMAN VAN, Oxford: Oxford University Press, 61–124.
- JACOD, J., AND P. PROTTER (2012): *Discretization of Process*. Springer.
- JACOD, JEAN, JIA LI, AND ZHIPENG LIAO (2021): “Volatility coupling,” *The Annals of Statistics*, 49, 1982–1998.
- KIM, SANGJOON, NEIL SHEPHARD, AND SIDDHARTHA CHIB (1998): “Stochastic volatility: Likelihood inference and comparison with arch models,” *The Review of Economic Studies*, 65, 361–393.
- KRISTENSEN, DENNIS (2010): “Nonparametric filtering of the realized spot volatility: A kernel-based approach,” *Econometric Theory*, 26, 60–2093.
- LEE, PETER M. (2012): *Bayesian Statistics: An Introduction, 4th edition*. Wiley.
- LI, JIA, DISHEN WANG, AND QIUSHI ZHANG (2022): “Reading the Candlesticks: An OK Estimator for Volatility,” *The Review of Economics and Statistics*, 1–45.
- LI, YONG, SUSHANTA K. MALLICK, NIANLING WANG, JUN YU, AND TAO ZENG (2025): “Deviance information criterion for bayesian model selection: Theoretical justification and applications,” *Journal of Econometrics*, 105978.
- LI, YONG, JUN YU, AND TAO ZENG (2020): “Deviance information criterion for latent variable models and misspecified models,” *Journal of Econometrics*, 216, 450–493.
- MALIK, SHEHERYAR, AND MICHAEL K. PITT (2011): “Particle filters for continuous likelihood evaluation and maximisation,” *Journal of Econometrics*, 165, 190–209.

- MYKLAND, PER A., AND LAN ZHANG (2008): “Inference for volatility-type objects and implications for hedging,” *Statistics and Its Interface*, 1, 255–278.
- NEWTON, MICHAEL A., AND ADRIAN E. RAFTERY (1994): “Approximate bayesian inference with the weighted likelihood bootstrap,” *Journal of the Royal Statistical Society: Series B (Methodological)*, 56, 3–26.
- OMORI, YASUHIRO, SIDDHARTHA CHIB, NEIL SHEPHARD, AND JOUCHI NAKAJIMA (2007): “Stochastic volatility with leverage: Fast and efficient likelihood inference,” *Journal of Econometrics*, 140, 425–449.
- SPIEGELHALTER, DAVID J., NICOLA G. BEST, BRADLEY P. CARLIN, AND ANGELIKA VAN DER LINDE (2002): “Bayesian measures of model complexity and fit,” *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 64, 583–639.
- (2014): “The deviance information criterion: 12 years on,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76, 485–493.
- STROUD, JONATHAN R., AND MICHAEL S. JOHANNES (2014): “Bayesian modeling and forecasting of 24-hour high-frequency volatility,” *Journal of the American Statistical Association*, 109, 1368–1384.
- TAYLOR, STEPHEN J. (1982): “Financial returns modelled by the product of two stochastic processes : a study of daily sugar prices, 1961-79,” *Time series analysis : Theory and Practice*, 1, 203–226.
- WANG, XIAOHU, WEILIN XIAO, AND JUN YU (2023): “Modeling and forecasting realized volatility with the fractional orstein-uhlenbeck process,” *Journal of Econometrics*, 232, 389–415.
- WATANABE, TOSHIAKI, AND JOUCHI NAKAJIMA (2024): “High-frequency realized stochastic volatility model,” *Journal of Empirical Finance*, 79, 101559.

- ZHANG, LAN, PER A MYKLAND, AND YACINE AÏT-SAHALIA (2005): “A tale of two time scales,” *Journal of the American Statistical Association*, 100, 1394–1411.
- ZU, YANG, AND H. PETER BOSWIJK (2014): “Estimating spot volatility with high-frequency financial data,” *Journal of Econometrics*, 181, 117–135.

Appendix

A. MCMC Procedure for Stochastic Volatility Model with Jumps

Since we are doing Bayesian analysis using Gibbs Sampling, we collect local nonparametric estimation of volatility into **data** and treat all the remained components in the model as **parameters** in a generic sense.

$\{\ln(\hat{c}_{n,j})\}_{j=1}^{m_n}$: **data**, for instance, nonparametric estimation of spot volatility

$\{\ln(c_{n,j})\}_{j=1}^{m_n} := \{h_j\}_{j=1}^{m_n}$: **parameters**, latent spot volatility

$\{\phi, \mu, \sigma_e^2\}$: **parameters**, volatility parameters

b : **parameters**, intraday pattern parameters

$\{J_j\}_{j=1}^{m_n}$: **parameters**, jump component indicator

$\{\eta_j\}_{j=1}^{m_n}$: **parameters**, jump component magnitude

$\{\kappa, \mu_\eta, \sigma_\eta^2\}$: **parameters**, jump parameters

A.1. Sampling jump components and jump parameters

- Sampling $\{J_j\}_{j=1}^{m_n}$. This sampling step is conditional on data $\{\hat{c}_{n,j}\}_{j=1}^{m_n}$ and updated $\{\tilde{h}_j\}_{j=1}^{m_n}$, $\{\phi, \mu, \sigma_e^2\}$, $\{\eta_j\}_{j=1}^{m_n}$ and $\{\kappa, \mu_\eta, \sigma_\eta^2\}$. Define

$$\xi_j = \tilde{h}_{j+1} - \phi \tilde{h}_j = (1 - \phi) \mu + J_j \eta_j + e_j,$$

which implies that

$$\xi_j \mid J_j=1 = (1 - \phi) \mu + \eta_j + e_j,$$

$$\xi_j \mid J_j=0 = (1 - \phi) \mu + e_j.$$

Since η_j and e_j are assumed to be normally distributed independently with each other such that $\eta_j \sim \mathcal{N}(\mu_\eta, \sigma_\eta^2)$ and $e_j \sim \mathcal{N}(0, \sigma_e^2)$, this further implies that

$$\xi_j \mid J_j=1 \sim \mathcal{N}((1 - \phi) \mu + \mu_\eta, \sigma_\eta^2 + \sigma_e^2),$$

$$\xi_j \mid J_j=0 \sim \mathcal{N}((1 - \phi) \mu, \sigma_e^2).$$

With κ updated within Gibbs sampling loop, updating $\{J_j\}_{j=1}^{m_n}$ as follows

$$\mathbb{P}\{J_j = 1 \mid \text{rest}\} = \frac{\kappa f_N(\xi_j; (1 - \phi) \mu + \mu_\eta, \sigma_\eta^2 + \sigma_e^2)}{(1 - \kappa) f_N(\xi_j; (1 - \phi) \mu, \sigma_e^2) + \kappa f_N(\xi_j; (1 - \phi) \mu + \mu_\eta, \sigma_\eta^2 + \sigma_e^2)}.$$

- Sampling $\{\eta_j\}_{j=1}^{m_n}$. This sampling step is conditional on data $\{\hat{c}_{n,j}\}_{j=1}^{m_n}$, updated $\{\tilde{h}_j\}_{j=1}^{m_n}$, $\{\phi, \mu, \sigma_e^2\}$, $\{J_j\}_{j=1}^{m_n}$, and $\{\kappa, \mu_\eta, \sigma_\eta^2\}$. Given the dynamic transition system stated above, if $J_j = 1$, then

$$(\ln(\hat{c}_{n,j}) + \ln k) \mid \tilde{h}_j \sim \ln \chi_k^2 \text{ or GMM approximation}$$

$$\tilde{h}_{j+1} = \mu + \phi(\tilde{h}_j - \mu) + \eta_j + e_j$$

where η_j is regarded as parameter with prior distribution such that $\eta_j \sim \mathcal{N}(\mu_\eta, \sigma_\eta^2)$, therefore posterior distribution of η_j conditional on all the rest is based on the following joint likelihood function

$$f_N(\eta_j; \mu_\eta, \sigma_\eta^2) \cdot f_N(\tilde{h}_{j+1} - \phi\tilde{h}_j - (1 - \phi)\mu; \eta_j, \sigma_e^2) \cdot f_{\ln \chi_k^2}(\ln(\hat{c}_{n,j}); \tilde{h}_j),$$

that is

$$\frac{1}{\sqrt{2\pi}\sigma_\eta} \exp\left\{-\frac{(\eta_j - \mu_\eta)^2}{2\sigma_\eta^2}\right\} \times \frac{1}{\sqrt{2\pi}\sigma_e} \exp\left\{-\frac{(\tilde{h}_{j+1} - \phi\tilde{h}_j - (1 - \phi)\mu - \eta_j)^2}{2\sigma_e^2}\right\} \times f_{\ln \chi_k^2}(\ln(\hat{c}_{n,j}); \tilde{h}_j).$$

This further implies that

$$\begin{aligned}
\eta_j \mid \text{rest} &\propto \exp \left\{ -\frac{\eta_j^2 - 2\eta_j\mu_\eta + \mu_\eta^2}{2\sigma_\eta^2} - \frac{\left(\tilde{h}_{j+1} - \phi\tilde{h}_j - (1-\phi)\mu\right)^2 - 2\left(\tilde{h}_{j+1} - \phi\tilde{h}_j - (1-\phi)\mu\right)\eta_j + \eta_j^2}{2\sigma_e^2} \right\} \\
&\propto \exp \left\{ -\frac{1}{2} \underbrace{\left(\frac{1}{\sigma_\eta^2} + \frac{1}{\sigma_e^2}\right)}_{\text{coef_a}} \eta_j^2 + \underbrace{\left(\frac{\mu_\eta}{\sigma_\eta^2} + \frac{\tilde{h}_{j+1} - \phi\tilde{h}_j - (1-\phi)\mu}{\sigma_e^2}\right)}_{\text{coef_b}} \eta_j \right\} \\
&\propto \exp \left\{ -\frac{1}{2(\sqrt{1/\text{coef_a}})^2} \left(\eta_j - \frac{\text{coef_b}}{\text{coef_a}} \right)^2 \right\}.
\end{aligned}$$

Therefore posterior distribution of $\eta_j \mid \text{rest}$ is normally distributed with mean equal to

$$\left(\frac{1}{\sigma_\eta^2} + \frac{1}{\sigma_e^2} \right)^{-1} \left(\frac{\mu_\eta}{\sigma_\eta^2} + \frac{\tilde{h}_{j+1} - \phi\tilde{h}_j - (1-\phi)\mu}{\sigma_e^2} \right),$$

and variance equal to

$$\left(\frac{1}{\sigma_\eta^2} + \frac{1}{\sigma_e^2} \right)^{-1}.$$

- Sampling κ . This sampling step is conditional on data $\{\hat{c}_{n,j}\}_{j=1}^{m_n}$, updated $\{\tilde{h}_j\}_{j=1}^{m_n}$, $\{\phi, \mu, \sigma_e^2\}$, $\{J_j\}_{j=1}^{m_n}$, $\{\eta_j\}_{j=1}^{m_n}$ and $\{\kappa, \mu_\eta, \sigma_\eta^2\}$. Since $\{J_j\}_{j=1}^{m_n}$ follows Bernoulli distribution, $\mathcal{Bern}(\kappa)$, and by specification we as econometricians have prior knowledge on κ such that κ follows Beta distribution $\mathcal{B}(\alpha_\kappa, \beta_\kappa)$, then κ is sampled from posterior distribution conditional on $\{J_j\}_{j=1}^{m_n}$. Actually, $\{J_j\}_{j=1}^{m_n}$ as binary random variables, realizations of the sum of $\{J_j\}_{j=1}^{m_n}$ lie in between 0 and m_n and we denote it as k . Given that

$$L(k \mid \kappa) := \mathbb{P} \left(\sum_{j=1}^{m_n} J_j = k \mid m_n, \kappa \right) = \binom{m_n}{k} \kappa^k (1-\kappa)^{m_n-k},$$

and the probability density function of Beta distribution ¹

$$\pi(\kappa \mid \alpha_\kappa, \beta_\kappa) = \frac{\kappa^{\alpha_\kappa-1} (1-\kappa)^{\beta_\kappa-1}}{B(\alpha_\kappa, \beta_\kappa)},$$

then the joint likelihood of k and κ is given as follows

$$L(k \mid \kappa) \cdot \pi(\kappa \mid \alpha_\kappa, \beta_\kappa).$$

By integrating out κ from this joint likelihood as follows yields

$$\begin{aligned} f(k \mid m_n, \alpha_\kappa, \beta_\kappa) &= \int_0^1 L(k \mid \kappa) \cdot \pi(\kappa \mid \alpha_\kappa, \beta_\kappa) d\kappa \\ &= \binom{m_n}{k} \frac{1}{B(\alpha_\kappa, \beta_\kappa)} \int_0^1 \kappa^{k+\alpha_\kappa-1} (1-\kappa)^{m_n-k+\beta_\kappa-1} d\kappa \\ &= \binom{m_n}{k} \frac{B(k+\alpha_\kappa, m_n-k+\beta_\kappa)}{B(\alpha_\kappa, \beta_\kappa)} \\ &= \binom{m_n}{k} \frac{B\left(\sum_{j=1}^{m_n} J_j + \alpha_\kappa, m_n - \sum_{j=1}^{m_n} J_j + \beta_\kappa\right)}{B(\alpha_\kappa, \beta_\kappa)}. \end{aligned}$$

Given the property of Beta function and Gamma function, it is easy to show that conditional on realizations of $\{J_j\}_{j=1}^{m_n}$, posterior distribution of κ follows Beta distribution as $\mathcal{B}\left(\sum_{j=1}^{m_n} J_j + \alpha_\kappa, m_n - \sum_{j=1}^{m_n} J_j + \beta_\kappa\right)$.

- Sampling σ_η^2 and μ_η . Given the prior structure $(\mu_\eta, \sigma_\eta^2) \sim \mathcal{NIG}(\tilde{\mu}_\eta, \lambda_{\sigma_\eta}, \alpha_{\sigma_\eta}, \beta_{\sigma_\eta})$ and for a given μ_η and σ_η^2 , $\eta_j \stackrel{i.i.d.}{\sim} \mathcal{N}(\mu_\eta, \sigma_\eta^2)$. Then, $\mu_\eta, \sigma_\eta^2 \mid \{\eta_j\}_{j=1}^{m_n}, \{J_j\}_{j=1}^{m_n}$ follows

¹ $B(\alpha_\kappa, \beta_\kappa)$ denotes beta function defined as the Euler integral of the first kind,

$$B(\alpha_\kappa, \beta_\kappa) = \int_0^1 t^{\alpha_\kappa-1} (1-t)^{\beta_\kappa-1} dt.$$

normal-inverse-gamma distribution $\mathcal{NIG}(\tilde{\mu}_\eta^*, \lambda_{\sigma_\eta}^*, \alpha_{\sigma_\eta}^*, \beta_{\sigma_\eta}^*)$ with

$$\begin{aligned}\tilde{\mu}_\eta^* &= \left(\sum_{j=1}^{m_n} J_j \eta_j + \tilde{\mu}_\eta \lambda_{\sigma_\eta} \right) / \left(\lambda_{\sigma_\eta} + \sum_{j=1}^{m_n} J_j \right) \\ \lambda_{\sigma_\eta}^* &= \lambda_{\sigma_\eta} + \sum_{j=1}^{m_n} J_j \\ \alpha_{\sigma_\eta}^* &= \sum_{j=1}^{m_n} J_j / 2 + \alpha_{\sigma_\eta} \\ \beta_{\sigma_\eta}^* &= \left[- \left(\sum_{j=1}^{m_n} \eta_j + \tilde{\mu}_\eta \lambda_{\sigma_\eta} \right)^2 / \left(\lambda_{\sigma_\eta} + \sum_{j=1}^{m_n} J_j \right) + \sum_{j=1}^{m_n} (J_j \eta_j)^2 + \lambda_{\sigma_\eta} \tilde{\mu}_\eta^2 \right] / 2 + \beta_{\sigma_\eta}.\end{aligned}$$

- Sampling b . Since we have assumed that b follows a prior distribution such that $b \sim \mathcal{TN}(\tilde{\mu}_b, \tilde{\sigma}_b^2, \alpha_b, \beta_b)$ with $\mathcal{TN}$ denoting a truncated normal distribution, we can show that the posterior sampling of b follows a conjugate scheme conditional on data $\{\hat{c}_{n,j}\}_{j=1}^{m_n}$, latent variables and other parameters.

A.2. Sampling volatility parameters and latent spot volatilities

- Sampling $\{\tilde{h}_j\}_{j=1}^{m_n}$. This sampling step is conditional on data $\{\ln(\hat{c}_{n,j})\}_{j=1}^{m_n}$ and updated $\{\phi, \mu, \sigma_e^2\}$, $\{J_j\}_{j=1}^{m_n}$, $\{\eta_j\}_{j=1}^{m_n}$, and $\{\kappa, \mu_\eta, \sigma_\eta^2\}$. Given the approximate linear Gaussian state-space model, Kalman filter and forward filtering and backward smoothing (FFBS) algorithm (Giordani, Pitt, and Kohn, 2011) applies for sampling $\{\tilde{h}_j\}_{j=1}^{m_n}$ in this setting.
- Sampling μ . This sampling step is conditional on data $\{\ln(\hat{c}_{n,j})\}_{j=1}^{m_n}$, updated $\{\phi, \sigma_e^2\}$, $\{\tilde{h}_j\}_{j=1}^{m_n}$, $\{J_j\}_{j=1}^{m_n}$, $\{\eta_j\}_{j=1}^{m_n}$ and $\{\kappa, \mu_\eta, \sigma_\eta^2\}$. Conditional on updated $\{J_j\}_{j=1}^{m_n}$, $\{\eta_j\}_{j=1}^{m_n}$, and $\{\tilde{h}_j\}_{j=1}^{m_n}$ updated within Gibbs loop, sampling μ is straightforward as $e_j = \tilde{h}_{j+1} - \phi \tilde{h}_j - (1 - \phi)\mu - J_j \eta_j$ is normally distributed with zero mean and variance equal to

σ_e^2 . Recall the functional form of joint normal density function as follows

$$\begin{aligned} & \prod_{j=1}^{m_n} f_N(\xi_j - J_j \eta_j; (1 - \phi)\mu, \sigma_e^2) \\ &= \prod_{j=1}^{m_n} \frac{1}{\sqrt{2\pi}\sigma_e} \exp \left\{ -\frac{[\xi_j - J_j \eta_j - (1 - \phi)\mu]^2}{2\sigma_e^2} \right\}. \end{aligned}$$

As we have specified the prior $\pi(\mu) \sim \mathcal{N}(\tilde{\mu}, \tilde{\sigma}^2)$, then the joint likelihood is given as follows

$$\begin{aligned} & f_N(\mu; \tilde{\mu}, \tilde{\sigma}^2) \cdot \prod_{j=1}^{m_n-1} f_N(\xi_j - J_j \eta_j; (1 - \phi)\mu, \sigma_e^2) \\ &= \frac{1}{\sqrt{2\pi}\tilde{\sigma}} \left[\frac{1}{\sqrt{2\pi}\sigma_e} \right]^{m_n-1} \exp \left\{ -\frac{1}{2\tilde{\sigma}^2}(\mu - \tilde{\mu})^2 - \frac{1}{2\sigma_e^2} \sum_{j=1}^{m_n-1} [\xi_j - J_j \eta_j - (1 - \phi)\mu]^2 \right\} \\ &\propto \exp \left\{ -\frac{1}{2\tilde{\sigma}^2}(\mu - \tilde{\mu})^2 - \frac{1}{2\sigma_e^2} \sum_{j=1}^{m_n-1} [\xi_j - J_j \eta_j - (1 - \phi)\mu]^2 \right\}. \end{aligned}$$

Again by applying the method of completing squares, we can show that conjugate posterior distribution of μ is Gaussian distribution with mean equal to

$$\left[\frac{1}{\tilde{\sigma}^2} + \frac{(m_n - 1)(1 - \phi)^2}{\sigma_e^2} \right]^{-1} \left[\frac{\tilde{\mu}}{\tilde{\sigma}^2} + \frac{1 - \phi}{\sigma_e^2} \sum_{j=1}^{m_n-1} (\xi_j - J_j \eta_j) \right],$$

and variance equal to

$$\left[\frac{1}{\tilde{\sigma}^2} + \frac{(m_n - 1)(1 - \phi)^2}{\sigma_e^2} \right]^{-1}.$$

- Sampling σ_e^2 . This sampling step is conditional on data $\{\hat{c}_{n,j}\}_{j=1}^{m_n}$ and updated $\{\phi, \mu\}$, $\{\tilde{h}_j\}_{j=1}^{m_n}$, $\{J_j\}_{j=1}^{m_n}$, $\{\eta_j\}_{j=1}^{m_n}$ and $\{\kappa, \mu_\eta, \sigma_\eta^2\}$. This sampling step is based on the following joint likelihood function assuming that σ_e^2 has inverse-gamma prior equipped

with shape parameter α_{σ_e} and scale parameter β_{σ_e} ,

$$\begin{aligned} \sigma_e^2 \mid \text{rest} &\propto (\sigma_e^2)^{-\alpha_{\sigma_e}-1} \exp \left\{ -\frac{\beta_{\sigma_e}}{\sigma_e^2} \right\} \times (\sigma_e^2)^{-(m_n-1)/2} \times \\ &\exp \left\{ -\frac{1}{2\sigma_e^2} \sum_{j=2}^{m_n} [\xi_j - (1-\phi)\mu - J_j\eta_j]^2 \right\} \\ &\propto (\sigma_e^2)^{-\alpha_{\sigma_e}-\frac{m_n-1}{2}-1} \exp \left\{ -\left(\beta_{\sigma_e} + \frac{1}{2} \sum_{j=1}^{m_n-1} [\xi_j - (1-\phi)\mu - J_j\eta_j]^2 \right) / \sigma_e^2 \right\} \end{aligned} \quad (\text{A.1})$$

which suggest that $\sigma_e^2 \mid \text{rest} \sim \mathcal{IG} \left(\alpha_{\sigma_e} + \frac{m_n-1}{2}, \beta_{\sigma_e} + \frac{1}{2} \sum_{j=1}^{m_n-1} [\xi_j - (1-\phi)\mu - J_j\eta_j]^2 \right)$.

- Sampling ϕ . This sampling step is conditional on data $\{\hat{c}_{n,j}\}_{j=1}^{m_n}$ and updated $\{\mu, \sigma_e^2\}$, $\{\tilde{h}_j\}_{j=1}^{m_n}$, $\{J_j\}_{j=1}^{m_n}$, $\{\eta_j\}_{j=1}^{m_n}$ and $\{\kappa, \mu_\eta, \sigma_\eta^2\}$. If prior imposed on ϕ is

$$\pi(\phi) \propto \left\{ \frac{(1+\phi)}{2} \right\}^{\alpha_\phi-1} \left\{ \frac{(1-\phi)}{2} \right\}^{\beta_\phi-1},$$

then full conditional density of ϕ is proportional to

$$\pi(\phi) \prod_{j=1}^{m_n} \frac{1}{\sqrt{2\pi\sigma_e}} \exp \left\{ -\frac{[\xi_j - J_j\eta_j - (1-\phi)\mu]^2}{2\sigma_e^2} \right\}.$$

Recall that

$$\begin{aligned} &\exp \left\{ -\frac{\sum_{j=1}^{m_n} [\tilde{h}_{j+1} - \mu - J_j\eta_j - \phi(\tilde{h}_j - \mu)]^2}{2\sigma_e^2} \right\} \\ &\propto \exp \left\{ -\frac{\sum_{j=1}^{m_n} (\tilde{h}_j - \mu)^2}{2\sigma_e^2} \phi^2 + \frac{\sum_{j=1}^{m_n} (\tilde{h}_{j+1} - \mu - J_j\eta_j)(\tilde{h}_j - \mu)}{\sigma_e^2} \phi \right\} \\ &\propto \exp \left\{ -\frac{1}{2V_\phi} (\phi - \hat{\phi})^2 \right\}, \end{aligned}$$

where

$$V_\phi = \frac{\sigma_e^2}{\sum_{j=1}^{m_n} (\tilde{h}_j - \mu)^2},$$

and

$$\hat{\phi} = \frac{\sum_{j=1}^{m_n} (\tilde{h}_{j+1} - \mu - J_j \eta_j) (\tilde{h}_j - \mu)}{\sum_{j=1}^{m_n} (\tilde{h}_j - \mu)^2}.$$

This suggests that we can use $\mathcal{N}(\hat{\phi}, V_\phi)$ as proposal distribution to construct Metropolis-Hastings algorithm for sampling ϕ . Specifically, within the loop of Gibbs Sampling, given the current value of $\phi^{(q-1)}$ at the q -th iteration, sampling ϕ' from $\mathcal{N}(\hat{\phi}, V_\phi)$. Since the associated acceptance-rejection ratio is constructed as

$$\min \left\{ 1, \underbrace{\frac{\pi(\phi') f\left(\left\{\tilde{h}_j\right\}_{j=2}^{m_n} \mid \left\{J_j\right\}_{j=2}^{m_n}, \left\{\eta_j\right\}_{j=2}^{m_n}, \left\{\phi', \mu, \sigma_e^2\right\}, \left\{\kappa, \mu_\eta, \sigma_\eta^2\right\}\right)}{\pi(\phi^{(q-1)}) f\left(\left\{\tilde{h}_j\right\}_{j=2}^{m_n} \mid \left\{J_j\right\}_{j=2}^{m_n}, \left\{\eta_j\right\}_{j=2}^{m_n}, \left\{\phi^{(q-1)}, \mu, \sigma_e^2\right\}, \left\{\kappa, \mu_\eta, \sigma_\eta^2\right\}\right)}}_{\text{part I}} \times \underbrace{\frac{f_N(\phi^{(q-1)}; \hat{\phi}, V_\phi)}{f_N(\phi'; \hat{\phi}, V_\phi)}}_{\text{part II}} \right\}$$

then $f_N(\phi^{(q-1)}; \hat{\phi}, V_\phi)$ and $f_N(\phi'; \hat{\phi}, V_\phi)$ would be cancelled with $\exp\left\{-\frac{1}{2V_\phi}(\phi^{(q-1)} - \hat{\phi})^2\right\}$ contained in the denominator of part I and $\exp\left\{-\frac{1}{2V_\phi}(\phi' - \hat{\phi})^2\right\}$ contained in the numerator of part I respectively. This suggests that the proposed value ϕ' is accepted as $\phi^{(q)}$ with probability equal to $\min\{1, \exp\{g(\phi') - g(\phi^{(q-1)})\}\}$ where $g(\phi) = \log \pi(\phi)$. If the proposed value is rejected, set $\phi^{(q)}$ equal to $\phi^{(q-1)}$.

B. Particle Filter for Calculating Log-likelihood

We follow Doucet and Johansen (2009), Malik and Pitt (2011), and Stroud and Johannes (2014) by constructing a particle filter algorithm to approximate the log-likelihood function used for calculating DIC, model log-marginal likelihood, and the marginal predictive distribution forecasting. Our particle filter algorithm is summarized as follows. Let $\{\ln(\hat{c}_{n,1:j})\}$

denote the local nonparametric estimation of spot volatility from $\mathcal{T}_{n,1}$ to $\mathcal{T}_{n,j}$. We denote the latent state variables as $\tilde{h}_j = h_j + \mu$. For a fixed value at $\boldsymbol{\theta}$, the likelihood function can be approximated using auxiliary particle filter (APF) as follows

$$\mathcal{L} \left(\{\ln(\hat{c}_{n,j})\}_{j=1}^{m_n} \mid \boldsymbol{\theta} \right) = \prod_{j=1}^{m_n} \left(\frac{1}{N} \sum_{i=1}^N \pi_j^{(i)} \right) \left(\frac{1}{N} \sum_{i=1}^N w_j^{(i)} \right), \quad (\text{B.1})$$

where N denotes the number of particles used in APF. The generic steps for obtaining $\{\pi_j^{(i)}\}$ and $\{w_j^{(i)}\}$ are summarized as follows:

1. Start with sample $\tilde{h}_{j-1}^{(i)} \sim p \left(\tilde{h}_{j-1} \mid \ln(\hat{c}_{n,1:j-1}) \right)$. For the initial state $\tilde{h}_0^{(i)}$, we sample $\tilde{h}_0^{(i)}$ from initial stationary distribution following mixture normal such that

$$\tilde{h}_0^{(i)} \sim \mathcal{N} \left(\mu + \frac{\kappa \mu_\eta}{1 - \phi}, \frac{\kappa (\sigma_e^2 + \sigma_\eta^2 + \mu_\eta^2) + (1 - \kappa) \sigma_e^2 - \kappa^2 \mu_\eta^2}{1 - \phi^2} \right),$$

by noting that

$$e_j + J_j \eta_j \sim \kappa \mathcal{N}(\mu_\eta, \sigma_e^2 + \sigma_\eta^2) + (1 - \kappa) \mathcal{N}(0, \sigma_e^2),$$

and

$$\mathbb{E}(e_j + \eta_j) = \kappa \mu_\eta,$$

$$\text{Var}(e_j + \eta_j) = \kappa (\sigma_e^2 + \sigma_\eta^2 + \mu_\eta^2) + (1 - \kappa) \sigma_e^2 - \kappa^2 \mu_\eta^2.$$

Accordingly,

$$\begin{aligned} \mathbb{E}(\tilde{h}_0) &= \mu + \frac{\kappa \mu_\eta}{1 - \phi}, \\ \text{Var}(\tilde{h}_0) &= \frac{\kappa (\sigma_e^2 + \sigma_\eta^2 + \mu_\eta^2) + (1 - \kappa) \sigma_e^2 - \kappa^2 \mu_\eta^2}{1 - \phi^2}. \end{aligned}$$

2. Compute $\pi_j^{(i)} \propto p \left(\ln(\hat{c}_{n,j}) \mid \hat{\tilde{h}}_j^{(i)} \right)$, where

$$\hat{\tilde{h}}_j^{(i)} = \mathbb{E} \left(\tilde{h}_j \mid \tilde{h}_{j-1}^{(i)}, \ln(\hat{c}_{n,1:j-1}) \right).$$

More specifically, note that

$$\begin{aligned}
\hat{h}_j^{(i)} &= \mathbb{E} \left(\tilde{h}_j \mid \tilde{h}_{j-1}^{(i)}, \ln(\hat{c}_{n,1:j-1}) \right) \\
&= \phi \tilde{h}_{j-1}^{(i)} + (1 - \phi) \mu + \mathbb{E} [J_{j-1} \eta_{j-1}] + \mathbb{E} [e_{j-1}] \\
&= \phi \tilde{h}_{j-1}^{(i)} + (1 - \phi) \mu + \kappa \mu_\eta.
\end{aligned}$$

3. With the results calculated in step 2, we then generate

$$\mathbf{k}^i \sim \mathcal{M} \left(\pi_j^{(1)}, \dots, \pi_j^{(N)} \right)$$

where $\mathcal{M}(\cdot)$ refers to the generic multinomial distribution specified by $\left\{ \pi_j^{(i)} \right\}$.

4. Generate $\tilde{h}_j^{(i)} \sim p \left(\tilde{h}_j \mid \tilde{h}_{j-1}^{(\mathbf{k}^i)}, \ln(\hat{c}_{n,1:j-1}) \right)$. This is a mixture normal distribution such that if $J_{j-1} = 1$, $\mathcal{N} \left((1 - \phi) \mu + \mu_\eta, \sigma_\eta^2 + \sigma_e^2 \right)$; if $J_{j-1} = 0$, $\mathcal{N} \left((1 - \phi) \mu, \sigma_e^2 \right)$. Therefore,

$$\begin{aligned}
&\tilde{h}_j^{(i)} \mid \tilde{h}_{j-1}^{(\mathbf{k}^i)}, \ln(\hat{c}_{n,1:j-1}) \\
&\sim \begin{cases} \mathcal{N} \left(\phi \tilde{h}_{j-1} + (1 - \phi) \mu + \rho \sigma_e \varepsilon_{j-1} + \mu_\eta, \sigma_\eta^2 + \sigma_e^2 (1 - \rho^2) \right) & \text{if } J_{j-1} = 1, \\ \mathcal{N} \left(\phi \tilde{h}_{j-1} + \rho \sigma_e \varepsilon_{j-1} + (1 - \phi) \mu, \sigma_e^2 (1 - \rho^2) \right) & \text{if } J_{j-1} = 0. \end{cases}
\end{aligned}$$

Hence,

$$\begin{aligned}
\tilde{h}_j^{(i)} \mid \tilde{h}_{j-1}^{(\mathbf{k}^i)}, \ln(\hat{c}_{n,1:j-1}) &\sim \kappa \mathcal{N} \left(\phi \tilde{h}_{j-1} + (1 - \phi) \mu + \rho \sigma_e \varepsilon_{j-1} + \mu_\eta, \sigma_\eta^2 + \sigma_e^2 (1 - \rho^2) \right) \\
&\quad + (1 - \kappa) \mathcal{N} \left(\phi \tilde{h}_{j-1} + \rho \sigma_e \varepsilon_{j-1} + (1 - \phi) \mu, \sigma_e^2 (1 - \rho^2) \right).
\end{aligned}$$

5. Compute $w_j^{(i)} \propto p \left(\ln(\hat{c}_{n,j}) \mid \tilde{h}_j^{(i)} \right) / \pi_j^{(\mathbf{k}^i)}$.

6. Generate

$$\mathbf{v}^i \sim \mathcal{M} \left(w_j^{(1)}, \dots, w_j^{(N)} \right)$$

and set $\tilde{h}_j^{(i)} = \tilde{h}_j^{(\mathbf{v}^i)}$.

C. Additional Plots

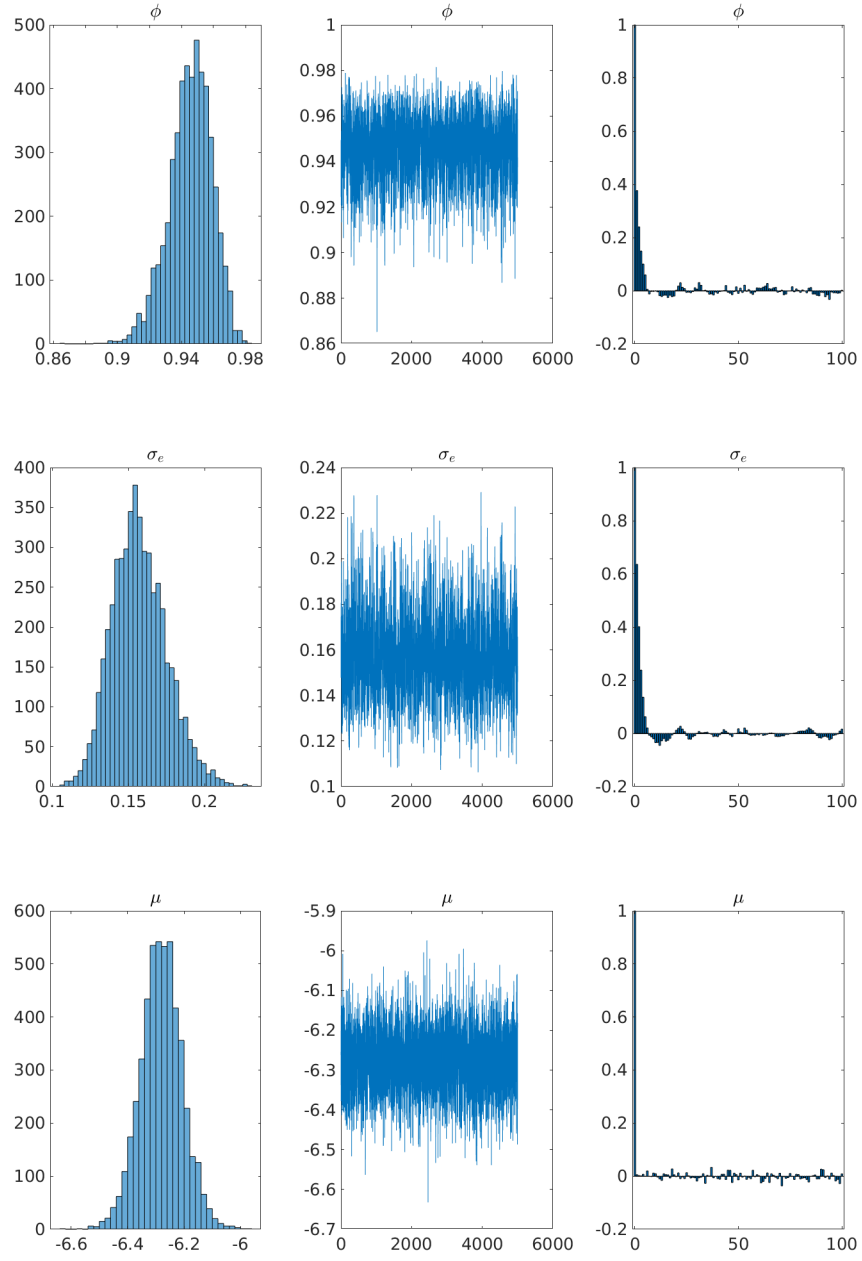


Figure C.1: Posterior densities, MCMC samples, and autocorrelation functions for **DGP**

1

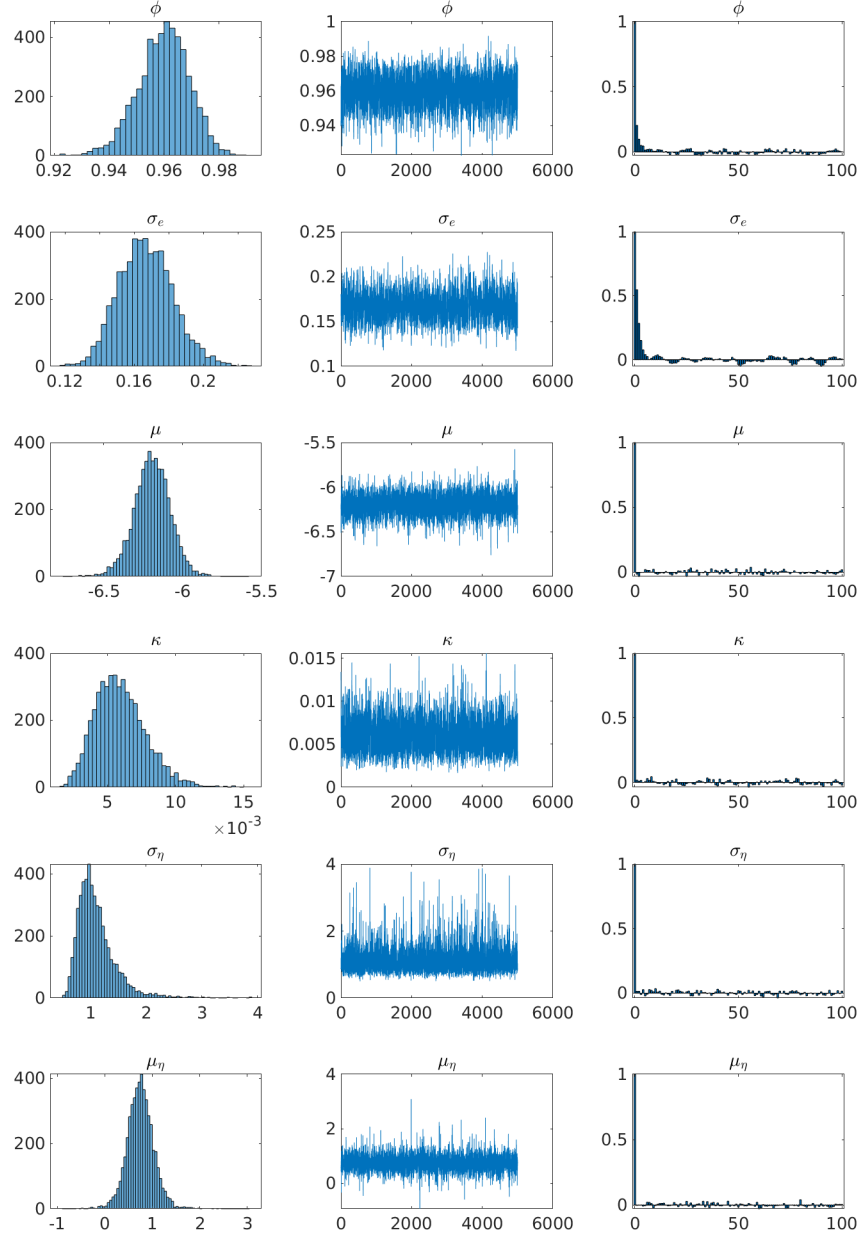


Figure C.2: Posterior densities, MCMC samples, and autocorrelation functions for **DGP**

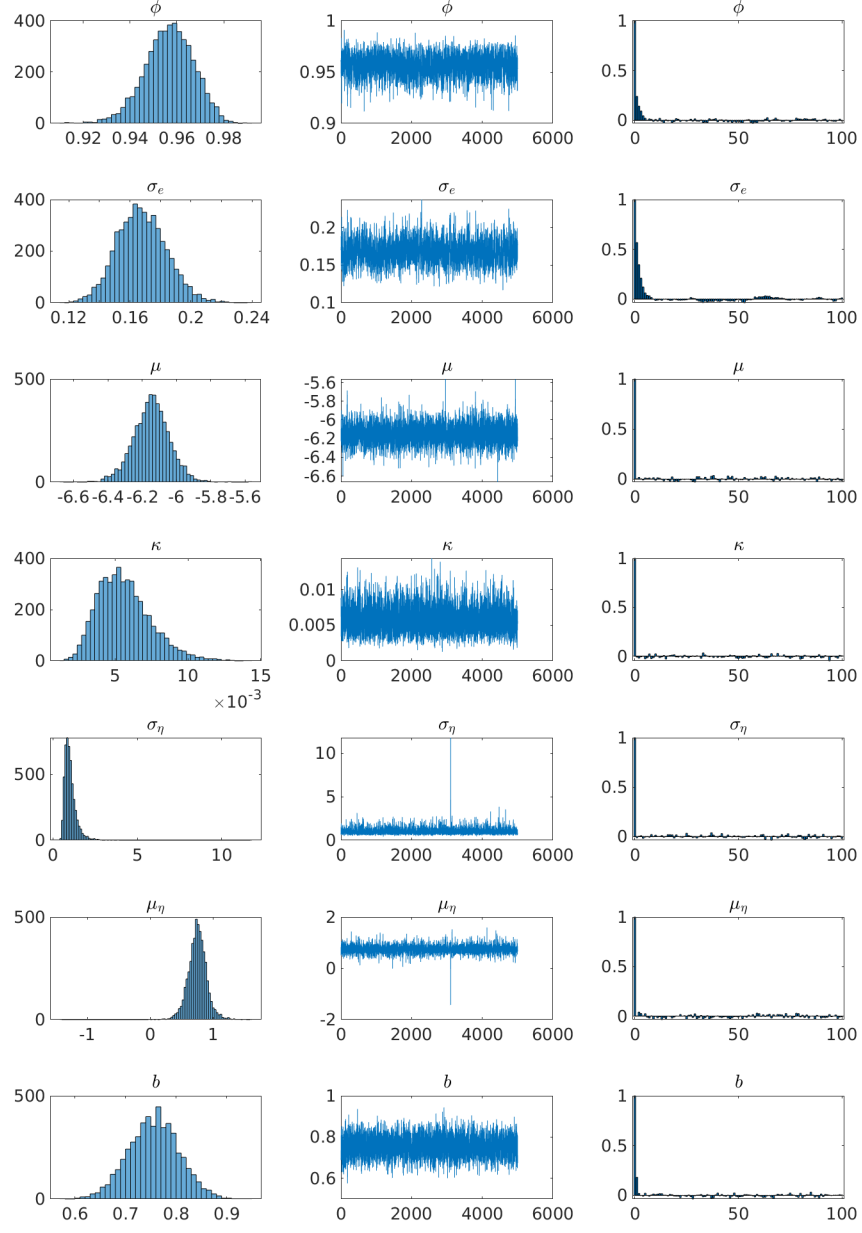


Figure C.3: Posterior densities, MCMC samples, and autocorrelation functions for **DGP**