Electronic Journal of Statistics
Vol. 15 (2021) 506–553
ISSN: 1935-7524
https://doi.org/10.1214/20-EJS1794

Bayesian inference on volatility in the presence of infinite jump activity and microstructure noise

Qi Wang, José E. Figueroa-López and Todd A. Kuffner

Department of Mathematics and Statistics,
Washington University in St. Louis
e-mail: qi.wang@wustl.edu; figueroa-lopez@wustl.edu; kuffner@wustl.edu

Abstract: Volatility estimation based on high-frequency data is important for accurate measurement and control of financial asset risks. A Lévy process with infinite jump activity and microstructure noise is considered one of the simplest models for financial data at high-frequency. Utilizing this model, we propose a "purposely misspecified" posterior of the volatility obtained by ignoring the the process' jump-component. The misspecified posterior is further corrected by a simple estimate of the location shift and re-scaling of the log likelihood. Our main result establishes a Bernstein-von Mises (BvM) theorem, which states that the proposed adjusted posterior is asymptotically Gaussian, centered at a consistent estimator, and with variance equal to the inverse of the Fisher information. In the absence of microstructure noise, our approach can be extended to make inferences for the integrated variance of general Itô semimartingales. Simulations are provided to demonstrate the accuracy of the resulting credible intervals, and the frequentist properties of the approximate Bayesian inference based on the adjusted posterior.

MSC2020 subject classifications: Primary 62M09; secondary 62F15.
Keywords and phrases: Bernstein-von Mises theorem, semiparametric and high-frequency inference, Itô semimartingales, microstructure noise.

Received September 2019.

1. Introduction

In the past decade, jumps have played an increasingly important role in asset price modeling. The necessity of jumps is supported by both empirical and realistic considerations such as (i) sudden and relatively large changes observed in real stock prices; (ii) the implied volatility smile phenomenon, which is more pronounced for short maturity options; and (iii) the proper management of risk [43, 9]. When jumps were first incorporated in the literature (e.g., Merton's model) the attention was centered on finite-jump activity models (i.e., those exhibiting finite jumps in finite time intervals). However, infinite-activity models are now considered more realistic as suggested by many studies based on real asset returns [4, 36, 42, 47, 46]. Here we consider a one-dimensional Lévy process $X = \{X_t\}_{t\geq 0}$ defined on some probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, P)$ over a fixed time horizon $t \in [0, T]$, which is a fundamental and widely-used tool to model

jump processes with infinite activity. Concretely,

$$X_t := \mu t + \theta^{1/2} W_t + J_t, \tag{1}$$

where $\mu \in \mathbb{R}$ and $\theta \in [0, \infty)$ are the drift and the variance parameters, respectively, $W = \{W_t\}_{t \geq 0}$ is a Wiener process, and $J = \{J_t\}_{t \geq 0}$ is an independent pure-jump Lévy process. In financial applications, X_t typically represents the log-return or log-price process $\log(S_t/S_0)$ of an asset with price process $\{S_t\}_{t \geq 0}$. In that case, the parameter $\sigma = \theta^{1/2}$ is called the volatility of the process and contributes to the total "variability" of the process X. Constant volatility can be generalized to a general Itô semimartingale (see § 7). Further details about the model and its components are given in § 2.

With improvements in computational power and the advent of electronic-based financial markets, intraday high-frequency data (every minute, second, or even nanosecond) has become widely available. While exploiting the convenience of massive data, analyses must also deal with market microstructure frictions (e.g., serial autocorrelation, price discreteness, and temporary demand-supply imbalance) caused by the nature of trading at high frequency. In an attempt to explain the nature of tick-by-tick data, [51] and [48] suggested the concept of microstructure noise, in which the observed transaction log-price Y_t at time t is a noisy measure of an underlying "efficient" log-price X_t:

$$Y_t = X_t + \varepsilon_t = \mu t + \theta^{1/2} W_t + J_t + \varepsilon_t. \tag{2}$$

Our purpose is to estimate the variance parameter θ based on high-frequency sampling observations $Y_{t_0}, Y_{t_1}, \ldots, Y_{t_n}$ $(0 = t_0 < \cdots < t_n = T)$ of the process over a fixed period of time $[0, T]$. From the perspective of frequentist point estimation, when there is no microstructure noise, [37] proposed a consistent estimator by eliminating those increments of the process, $\Delta_i Y := Y_{t_i} - Y_{t_{i-1}}$, which are larger in absolute value than a suitably chosen threshold. The asymptotic efficiency of the estimator with the restriction of a bounded variation jump process J is proved later in [10]. For a jump component of unbounded variation, there exists a rate efficient estimator introduced by [27] based on the empirical characteristic function. [33] also introduced a closely related estimator and established a central limit theorem for its estimator. When the microstructure noise is taken into account but jumps are not present, several estimators have been proposed. The two-scale estimator in [48] considered two different estimation scales of the process to estimate and eliminate the effect of the noise. It was generalized by [49] to achieve the optimal convergence rate $n^{-1/4}$. The preaveraging approach in [25] replaced the increments $\Delta_i Y$ with a weighted summation over a small window. The realized kernel (RV) method in [2] utilized the weighted realized autocovariances. A quasi-maximum likelihood estimator (QMLE) approach is proposed by [1, 18, 45]. The stochastic volatility is first misspecified as locally constant, independent with the previous state, or even constant, which allows construction of a quasi-likelihood and derivation of the maximum likelihood estimator (MLE). [11] showed that the estimator is robust to a microstructure noise following an MA(∞) process, and proposed a tuning-free procedure to

select the order of noise using the Akaike information criterion (AIC). Then, [8] achieved efficient asymptotic variance reduction for non-constant volatility by aggregating a local version of an RV estimator or QMLE estimator. [19] and [34] propose bootstrap methods to approximate the distribution of realized volatility. A second-order refinement is achieved over the limiting normal approximation. When both noise and jumps are present, [40, 41] introduced the modulated bipower variation estimator using the bipower variation of the weighted average of the increments. The estimator is consistent, but cannot achieve the efficient convergence rate $n^{-1/4}$, which represents the best rate that can be achieved for this estimation problem in the presence of noise and jumps. [7] proposed two quantile-based realized volatility estimators by employing empirical quantiles of the averaged returns. The estimators are both consistent and asymptotically efficient, but only applicable for processes with finite jumps. More recently, [30] combined the preaveraging method of [25] and the thresholding ideas of [37] to construct a consistent estimator of the integrated variance that is robust to both noise and infinite jump activity. The estimator proposed by [5] is the same as [30] under our settings. [5] considers the estimation of a functional of volatility under a general setting where the interaction of the noise and the underlying process is not merely additive. Consistency and efficiency are also proved. The details of this estimator are explained in § 6.

Whereas there are numerous frequentist estimators available, the development of an explicit and efficient Bayesian approach which can accommodate high-frequency data remains a largely open problem. Genuine Bayesian inference for a parameter of interest can only be based on the corresponding marginal posterior distribution, which is a conditional distribution on the parameter space of the interest parameter, and the conditioning is on the observed data. A fully Bayesian approach requires that the joint posterior of all parameters must be constructed based on the full likelihood function and a joint prior distribution over all the parameters. Then, integrating the joint posterior over the nuisance parameters (in our case, the parameters related to the jump component J and microstructure noise ε) yields a marginal posterior distribution for the parameter of interest. Since analytic derivation of the joint or marginal posterior is often intractable, Markov Chain Monte Carlo (MCMC) methods are typically used to sample from the joint posterior, and then numerical integration over the nuisance parameters is achieved by simply ignoring the corresponding MCMC output for those parameters. MCMC-based Bayesian methods have been applied to the volatility estimation problem by several previous authors. [6] and [13] used MCMC for a diffusion process augmented by a Poisson jump process. More recently, additional model complexity has been accommodated by taking infinite activity into consideration. [47] proposed an MCMC estimation method using both spot and option prices. Their jumps are assumed to follow either a variance gamma process or an α-stable process. [28] developed an automated sequential Monte Carlo algorithm by adding an additional re-sampling step for variance gamma jumps. [22] applied a slice sampling approach with a similar variance gamma assumption. [20] incorporated realized variation and realized power variation into an MCMC procedure, and analyzed a generalized variance

gamma process. [46] considered both returns and the Chicago Board of Options Exchange (CBOE) Volatility Index (VIX) to obtain the posterior for the jump part. The variance gamma process and normal inverse gamma process were considered.

Although the papers mentioned above considered Bayesian inference derived from the joint posterior, they all require strong assumptions about the structure of the jumps, which severely limit the practical value of these methods. Without these simplifying assumptions, it is quite challenging to write down the full likelihood function under the semi-parametric setting (1), which means that it is also difficult to obtain the full joint posterior without such assumptions. One of these assumptions is the choice of a particular specification of the jump process J, among many possible jump processes. However, empirical results in [36], [47], and [35] suggested that different jump assumptions lead to different estimation results for volatility. The posterior depends heavily on the structure of the jumps. Thus, sticking to just one jump type increases the possibility of misspecification and, therefore, can lead to inaccurate estimation and inference.

Moreover, specifying and calculating the distribution of the jump component may incur heavy computational costs, especially when working with high-frequency data. For this reason, nearly all of the aforementioned studies consider only daily returns data. Some authors, such as [28] and [20], did apply their methods to hourly data and 5-minute data, respectively. However, both studies fixed one of the parameters of the jump process as constant, in order to reduce the computational load.

The difficulties of deriving the posterior and the associated heavy computational costs are primarily caused by the jumps, which are only related to the nuisance parameters in the present context of inference for volatility. Our target of estimation, the variance or volatility, is not affected by the jumps, and is modeled by a simple Gaussian process, for which Bayesian inference can be more easily obtained. Based on this observation, one plausible idea to tackle the problem is to ignore the nuisance parameters in the nonparametric part of the process, replace the nuisance parameters in the parametric part by their consistent estimators, and construct a posterior only for the parameter of interest. The advantages of such an approach are that one need not specify a prior on the jump process, and it is not necessary to obtain samples from the full joint posterior. By contrast, we will directly obtain an approximation to the marginal posterior for the volatility, which we will show can be used for accurate Bayesian inference. This approach was recently used by [38]. They derived a 'purposely misspecified' posterior for a jump-diffusion model with constant volatility, finite jump activity and without microstructure noise, which targets the parameter of interest, the volatility, directly. Using a misspecified model on purpose, the inherent difficulty of specifying the likelihood function in a nonparametric model is tackled by omitting the complicated nuisance component of the model. The bias and the inaccurate variance caused by the misspecification are later corrected by applying a location shift and rescaling the likelihood using a Gibbs posterior. They showed that the adjusted posterior possesses good asymptotic properties, as guaranteed by a Bernstein-von Mises theorem.

In this paper, we study a 'purposely misspecified' posterior for the variance θ of the model (2), either with or without microstructure noise, which is a considerably more difficult and realistic setting in comparison to the finite jump activity model without microstructure noise that was studied by [38]. Our main result is a Bernstein-von Mises Theorem for the adjusted posterior for the volatility parameter, which shows that the proposed posterior is asymptotically normal and centered at a consistent estimator, and with variance shrinking at rates $n^{-1/2}$ and n^{-1}, respectively, depending on whether a microstructure noise is incorporated or not in the model.

The novel contributions of this paper can be summarized as follows. First, we allow the jump process to be any Lévy process, i.e., there is no parametric assumption about the nuisance component, and no assumption of finite jump activity. We also allow for an additive microstructure noise in the data. These relaxations of the stronger assumptions made in the existing literature help to alleviate inaccuracies introduced by model misinterpretation, and also avoid expensive computational costs. In fact, we also show that in the situations when the microstructure noise can be ignored (e.g., when working with medium-range frequencies), our approach can be extended to the estimation of the integrated variance of a general Itô semimartingale X. In particular, we allow stochastic volatility and a general pure-jump semimartingale component J.

It is important to remark that our proposed inference procedure is among the first Bayesian approaches that can accommodate truly high-frequency data; due to high computational costs and the lack of theoretical performance guarantees, most of the existing literature involves methods which are only applicable to low frequency data, such as daily observations. Finally, our results suggest that, under certain circumstances, misspecification on purpose can serve as a vehicle for accurate approximate Bayesian inference about low-dimensional interest parameters in complex models with possibly infinite-dimensional nuisance parameters.

The paper is organized as follows. A detailed description of the setting and model are provided in § 2. Differences between finite and infinite activity when deriving the 'purposely misspecified' posterior are highlighted in § 3. This analysis reveals the importance of proposing a modified version of the Bernstein-von Mises theorem, which is stated in § 4. The misspecified model is presented in § 5, and further extended in § 7. The main results are stated in § 5.2 and § 6. Simulation results given in § 8 illustrate the performance of our procedures. Discussion and concluding remarks are in § 9. The proofs and further technical details appear in the Appendix.

2. Model setup

As mentioned in the introduction, we consider a one-dimensional continuous-time process of the form (1), $X = \{X_t; t \in [0,T]\}$, defined on some probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, P)$. It consists of constant drift and diffusion coefficients $\mu \in \mathbb{R}$ and $\theta \in \mathbb{R}_+$, respectively, as well as a pure jump part $J = \{J_t\}_{t\geq 0}$. The

parameter space for θ, denoted as Θ, is assumed to be a bounded and open subset of $(0, +\infty)$ such that $0 \notin \bar{\Theta}$.

The jump component J is assumed to be a pure jump Lévy process, which is used in many fields of science. In mathematical finance, a Lévy process is widely recognized to provide a better fit to intraday returns than plain Brownian motion or even some stochastic volatility models. A comprehensive overview of the applications of Lévy processes can be found in [3] and [9]. A Lévy processes is defined as a càdlàg, real valued stochastic process which has independent and stationary increments, and is stochastically continuous. It is known that a Lévy process X takes the general form (1) with J defined as

$$\begin{aligned} J_t = J_{1t} + \tilde{J}_{2t}, \quad J_{1t} = \int_0^t \int_{|x|>1} x\, \mu(dx, ds), \\ \tilde{J}_{2t} = \int_0^t \int_{0<|x|\leq 1} x\, (\mu(dx, ds) - v(dx)ds), \end{aligned} \tag{3}$$

where μ is a Poisson random measure on $\mathbb{R}_+ \times \mathbb{R}\backslash\{0\}$ with mean measure $\nu(dx)dt$ such that $\int_{\mathbb{R}\backslash\{0\}}(|x|^2 \wedge 1)\nu(dx) < \infty$. This is the so-called Lévy-Itô decomposition of X and ν is called the Lévy measure of X.

We also consider the possibility that the observations of the process may be contaminated by random errors. Specifically, we assume that our observations take the form

$$Y_{t_j} = X_{t_j} + \varepsilon_{t_j}, \quad j = 0, \ldots, n, \tag{4}$$

with equally-spaced discrete times $0 = t_0 < t_1 < \ldots < t_n = T$ such that $t_j - t_{j-1} = \Delta_n = T/n$. To summarize, the data is assumed to be generated by the model (2)-(4) with true volatility value θ^*, which is the target to be estimated. The Lévy model with microstructure noise (2) is considered one of the simplest models for financial data at high-frequency (see [15] for an empirical assessment of the model and [39] for a survey of on parametric inference of Lévy models). Constant volatility is a strong assumption, but, as shown in §7, it can be relaxed to stochastic volatility when the microstructure noise can be ignored.

The process Y satisfies the following assumptions (see Remark 2.1 below for further comments about these):

Assumption (N).

1. The microstructure noise components, $\varepsilon = \{\varepsilon_{t_j}\}_{j=1}^n$, are independent and identically distributed (i.i.d.), and follow a $\mathcal{N}(0, \sigma_\varepsilon^2)$ distribution. In the Bayesian framework, we assume the i.i.d. property holds conditionally on the unknown parameter σ_ε.
2. The processes ε and X are independent.

Assumption (JD). The Blumenthal-Getoor index of J, defined by

$$\alpha = \inf\left\{ p > 0 : \int_{|x|\leq 1} |x|^p \nu(dx) < \infty \right\} \leq 2, \tag{5}$$

satisfies one of the following two scenarios:

1. $\alpha < 1$, or
2. $\alpha > 1$, and Lévy measure ν is such that

$$\left|\nu((x,\infty)) - x^{-\alpha}\right| \leq g(x), \quad \text{for } x \in (0,1],$$

where $g(x)$ is a decreasing function satisfying $\int_0^1 x^{r-1} g(x)dx < \infty$, for some $r < 1$.

Assumption (JF). The process J has a finite 16th moment. Equivalently, we have $\int_{|x|\geq 1} x^{16}\, \nu(dx) < \infty$.

Remark 2.1.

1. [21] suggested that the independence assumption for ε and X is reasonable for moderate intraday frequency (e.g. 1 minute). The i.i.d. normal assumption is used in §5.1 in order to give an explicit representation of the likelihood function, which will allow us to prove a local asymptotic normality (LAN) property. It may be possible to relax it as in the quasi-likelihood method of [1], but this is beyond the scope of this work.
2. For a Lévy process, the Blumenthal-Getoor index α controls the small jump activity of the process: it becomes larger as the small jumps are more persistent. The assumption that $\alpha < 1$ means that the paths of the process J are of bounded variation, almost surely. This assumption is widely used in the literature (see, e.g., [7], [10], [27], and [30]), and is used later in §6 to apply a central limit theorem (CLT) for the realized quadratic threshold estimator of the volatility. When $\alpha \geq 1$, [27] concluded that there is no CLT in general for such an estimator and its rate of convergence to the integrated variance is much slower than $n^{-1/2}$. The characteristic-function-based estimator [27] fills this gap and motivates scenario 2 of Assumption (JD). The restrictions on α with the Lévy measure therein are in fact inherited from [27] (see also [33]). The estimator is actually robust with $\alpha = 1$, but the asymptotic variance for the estimator when $\alpha = 1$ is not the same as when $\alpha \neq 1$ (see Remark 4 of [27]). Therefore, for consistency and simplicity, we omit the scenario when $\alpha = 1$.
3. In the absence of microstructure noise, we can accommodate a stochastic volatility model and much more general pure-jump semimartingales J (see § 7). We also don't require Assumption (JF).

For future reference, recall the following common notation for the increments and jumps of an arbitrary continuous-time càdlàg process $\{U_t\}_{t\geq 0}$:

$$\Delta_i U = \Delta_i^n U = U_{t_i} - U_{t_{i-1}}, \quad \Delta U_t = U_t - U_{t^-}.$$

3. Comparison with finite jump activity models

In this section, we present a motivating example using a simple finite jump activity model, in order to illustrate the usefulness of the approximate Bayesian inference obtained via purposeful misspecification. [38] proposed this approach,

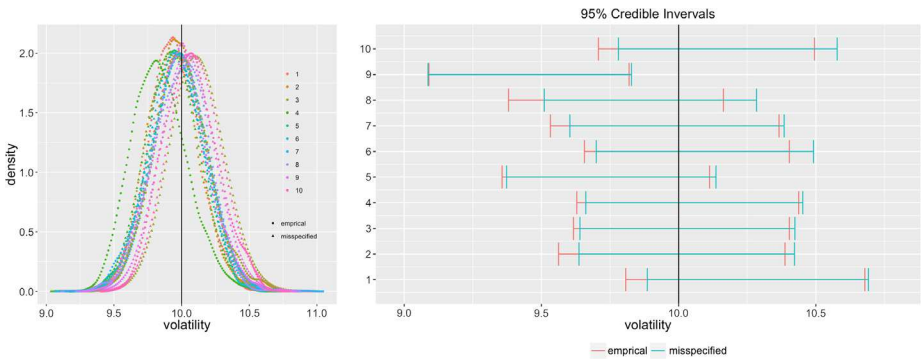
but did not make comparisons to the true marginal posterior for the volatility parameter. We then explain what issues will arise when considering the more complicated and realistic setting of infinite jump activity.

3.1. An illustration through simulation

We first empirically compare the "purposely misspecified" posterior from [38] with a genuine marginal posterior derived from the full joint posterior. The goals of this comparison are to assess the accuracy of the former method in a situation where the full joint posterior and genuine marginal posterior are tractable, and also to motivate our proposed procedures. Model (1) is used with a compound Poisson jump process $J_t = \sum_{i=0}^{N_t} \xi_i$. Here, $N = \{N_t\}_{t\geq 0}$ is a Poisson process with rate λ, and $\{\xi_i\}_{i\geq 1}$ are i.i.d. random variables independent of N and W. We assume that $\{\xi_i\}_{i\geq 1}$ follow a uniform distribution $U(-1,1)$. This assumption enables us to derive a joint posterior and perform Gibbs sampling for the parameters $\Theta = (\mu, \theta, \lambda)$. The other parameter values are taken from [38]: $\lambda = 5$, $\mu = 1$, $\theta = 10$, $n = 5000$, and $T = 1$. In what follows, we approximate the Poisson process by a Bernoulli process; namely, N is assumed to be a point process such that $P[N_{t_i} - N_{t_{i-1}} = 1] = \lambda\Delta_n$ and $P[N_{t_i} - N_{t_{i-1}} = 0] = 1 - \lambda\Delta_n$. The joint posterior density based on the data $\boldsymbol{X}^{(n)} = (\Delta_1 X, \ldots, \Delta_n X)$ can then be written as

$$p(\Theta|\boldsymbol{X}^{(n)}) \propto \prod_{i=1}^{n} \frac{1}{\sqrt{\pi\theta\Delta_n}} \exp\left\{-\frac{(\Delta_i X - \mu\Delta_n)^2}{2\theta\Delta_n}\right\} (1-\lambda\Delta_n)p(\Theta)$$
$$+ \prod_{i=1}^{n} \int_{-1}^{1} \frac{1}{\sqrt{2\pi\theta\Delta_n}} \exp\left\{-\frac{(\Delta_i X - y - \mu\Delta_n)^2}{2\theta\Delta_n}\right\} dy \cdot \lambda\Delta_n p(\Theta).$$

The priors chosen for μ, θ, λ are a standard Gaussian distribution, an inverse gamma distribution, and a beta distribution, respectively. The posterior for θ is estimated by two methods: (i) Gibbs sampling from the full joint posterior, followed by numerical integration to yield the marginal posterior (i.e., we simply ignore the MCMC output for the nuisance parameters μ and λ); and (ii) a direct posterior for θ obtained by misspecification on purpose. We emphasize that the Gibbs sampling approach, which is exact modulo finite simulation error, is only available here because of the very strong assumptions made regarding the jump process. This method is not available for the more complicated and realistic settings we consider in this paper, whereas the second method works quite well under those settings (as shown later). The direct posterior proposed by [38] is an approximation using a misspecified model to directly obtain a posterior for θ without the need first to obtain the full joint posterior, and then marginalize. Figures 1a–1b compare the two approaches. Figure 1a shows the posteriors for 10 different simulations. The 'purposely misspecified' posterior typically resembles quite well the genuine marginal posterior obtained from Gibbs sampling from the joint posterior. Both posteriors are centered around the true volatility. The 95% highest posterior density intervals are shown in Figure 1b. The close



(a) Posteriors (b) Credible Intervals

FIG 1. ***Comparison with empirical posterior.*** *(a) Ten different processes distinguished by ten different colors are generated, and the corresponding posteriors are plotted. Each color has two distributions. The one formed by the triangle is the misspecified posterior, while the other represents the Gibbs sampling results. (b) The red lines represent the 95% highest posterior density (HPD) intervals calculated from the MCMC samples from the joint posterior. The blue lines are the 95% HPD intervals for the 'purposely misspecified' posterior.*

similarity of the direct posterior to the genuine marginal posterior, and also the corresponding credible intervals, demonstrates the accuracy of the 'purposely misspecified' posterior, and therefore, its validity for approximate Bayesian inference.

In general, it is quite complicated to perform fully Bayesian analysis for infinite jump activity models based on high-frequency data because the joint posterior is analytically intractable, and even MCMC-based procedures can be computationally demanding and may require additional assumptions about the underlying process in order to possess good properties. These additional assumptions can limit the flexibility of the analysis or lead to greater risk of misspecification. Moreover, when inference is only required for a low-dimensional interest parameter, it may be wasteful or cumbersome to construct the computationally-demanding full joint posterior. To perform MCMC sampling from the joint posterior, some studies (e.g., [36]) consider the unobserved jump increments $\Delta_i J$ as a latent parameter. However, with high-frequency data, this may cause numerical difficulties. Indeed, most of the previous Bayesian studies on high-frequency data that incorporate an infinite jump activity component in the model impose strong parametric restrictions in order to conduct MCMC sampling.

3.2. Theoretical challenges

[38] applied their purposely misspecified approach to the simpler model setting of an uncontaminated jump-diffusion model with constant volatility and finite jump activity. They first constructed a misspecified model by omitting the jump part J. Under this misspecified model, the resulting misspecified posterior was shown to be asymptotically normal conditionally on a given path of J. Since the

result works for all possible J, it can be generalized to a version that does not depend on J. Even though such an asymptotic normality does hold for a suitably centered and scaled misspecified posterior for the volatility, the misspecification of the model has the adverse effect of causing this misspecified posterior to center in the wrong place and to have an inefficient variance compared to the true marginal posterior obtained by marginalizing the full joint posterior over the drift and jump parts of the model. Therefore, [38] proposed to correct for the bias and inefficiency of the misspecified posterior by, respectively, shifting the center by an estimate of the bias, and rescaling the log likelihood using a properly chosen temperature parameter. Since the Bernstein-von Mises theorem involves convergence in total variation norm, and this norm is invariant with respect to location shifts, the resulting corrected posterior for volatility still admits a Bernstein-von Mises theorem but with a correct center and efficient variance equal to the Cramér-Rao lower bound.

In a model with infinite jump activity, we can similarly ignore the jump part and consider a misspecified model, but it is impossible to conclude an unconditional Bernstein-Von Mises theorem from the analogous result for the conditional posterior given a fixed path of the jump process J. The main reason is that for a jump process with infinite activity, the realized quadratic variation $[J]_n := \sum_{i=1}^{n} \Delta_i J^2$ does not converge to the quadratic variation $[J] := \sum_{0 \leq s < T} (\Delta J_s)^2$ for almost every path of J (i.e., a.s. convergence does not hold but merely convergence in probability). Almost sure convergence is necessary to prove local asymptotic normality of the likelihood and an optimal convergence rate for the posterior mean. The latter two conditions are required to apply a Bernstein-von Mises theorem under misspecification established in [31], which is the main tool behind the result of [38].

For a general semimartingale, it is well-known that $[J]_n$ does converge to $[J]$ in probability (cf., [26]). Furthermore, for Lévy processes, a rather good rate of convergence of $O_p(n^{-1/2})$ can be obtained (see Lemma A.2). We find that this weaker convergence is enough to prove the desired properties of the posterior by applying an unconditional version of the Bernstein-Von Mises theorem and skipping the intermediate results under the conditional probability measure given the jump part. In addition to the issues created by the presence of infinite jump activity, the parametric part is also affected by the presence of the noise ε. To deal with this, we treat the variance of the noise, σ_ε^2, as an additional nuisance parameter and prove a semiparametric type of Bernstein-von Mises theorem under misspecification. The adjusted posterior for volatility, and the associated Bernstein-von Mises theorem, must now include corrections for deliberately ignoring the presence of microstructure noise.

4. A semiparametric version of the misspecified BvM Theorem

As explained in the previous section, the misspecified Bernstein-von-Mises Theorem of [31] plays a crucial role in proving the asymptotic properties of the purposely misspecified posterior. To accommodate the more complicated setting of

our model, the result needs to be generalized to a semiparametric version, which is stated next.

Theorem 4.1. *Consider the space* $\Omega^{(n)} := \Omega_1^{(n)} \times \Omega_2 := \mathbb{R}^n \times D([0,\infty))$ *(D represents the Skorokhod space of all càdlàg* $\mathbb{R}$*-valued functions) and a collection of semiparametric models on* $\Omega^{(n)}$*,*

$$\left\{ P^{(n)}_{((\theta,\eta),\nu)} : (\theta,\eta) \in \Theta, \nu \in U \right\},$$

where Θ *is an open subset of* $\mathbb{R}\times\mathbb{R}^d$ *and* U *is an open subset of an infinite dimensional topological space* $\mathbb{F}$*. Let* $P_0^{(n)} := P^{(n)}_{((\theta^*,\eta^*),\nu^*)}$ *and let* $Z^{(n)} = (Z_1,\ldots,Z_n)$ *and* $\{Y_t\}_{t\geq 0}$ *be the canonical processes on* $\Omega^{(n)}$ *defined for* $\omega = (\omega_1,\omega_2) \in \Omega_1^{(n)} \times \Omega_2$ *as* $Z_i(\omega) = \omega_{1i}$ *and* $Y_t(\omega) = \omega_2(t)$*, respectively.*

Define $\Phi = g(\theta,\eta,Y_\cdot)$ *and* $\Phi^\dagger = g(\theta^*,\eta^*,Y_\cdot)$*, where* $Y_\cdot$ *denotes the sample path of* J *and* $g : \Theta \times D([0,\infty)) \to \Theta' \subset \mathbb{R}$ *is a known deterministic function. Our data consists of* $X^{(n)} := (X_1,\ldots,X_n) := T(Z^{(n)},Y_\cdot)$*, where the function* $T : \mathbb{R}^n \times D([0,\infty)] \to \mathbb{R}^n$ *is known.*

Suppose there are purposely misspecified models for $X^{(n)}$ *denoted as* $\tilde{P}_\vartheta(\cdot) := \tilde{P}(\cdot|\vartheta)$*,* $\vartheta \in \Theta'$*, which are distributions on* $\mathbb{R}^n$ *parameterized by* ϑ *with densities* $\tilde{p}_\vartheta$*. Let* Π *be a prior distribution with a density* π *that is continuous and positive on* Θ'*. Define the misspecified posterior distribution based on* Π *and* $\tilde{P}_\vartheta(\cdot)$ *as*

$$\Pi^n(\varphi \in B|X^{(n)}) = \frac{\int_B \tilde{p}_\varphi(X^{(n)})\pi(\varphi)\,d\varphi}{\int \tilde{p}_\xi(X^{(n)})\pi(\xi)\,d\xi}, \quad B \in \mathcal{B}(\Theta').$$

Assume $\{\tilde{P}_\vartheta, \vartheta \in \Theta'\}$ *satisfy a stochastic local asymptotic normality (LAN) condition relative to a given sequence* $\delta_n \to 0$ *as norming rate, i.e. there exist some random quantities* $\Delta_{n,}$ *and* V_n *such that for every compact set* $K \in \mathbb{R}$ *and* $\epsilon > 0$*,*

$$P_0^{(n)}\left(\sup_{h\in K}\left|\log\frac{\tilde{p}_{\Phi^\dagger+\delta_n h}}{\tilde{p}_{\Phi^\dagger}}(X^{(n)}) - V_{\Phi^\dagger}\Delta_{n,\Phi^\dagger}h - \frac{1}{2}V_{\Phi^\dagger}h^2\right| > \epsilon\right) \to 0, \quad \text{as } n\to\infty. \tag{6}$$

Also, for any sequence of constants $M_n \to \infty$*, the posterior* Π^n *is assumed to satisfy*

$$\Pi^n\left(|\varphi - \Phi^\dagger| > \delta_n M_n | X^{(n)}\right) \overset{P_0^{(n)}}{\to} 0, \quad n\to\infty. \tag{7}$$

Then, Π^n *converges to a sequence of normal distributions in total variation:*

$$P_0^{(n)}\left(\sup_B\left|\Pi^n\left((\varphi-\Phi^\dagger)/\delta_n \in B|X^{(n)}\right) - \mathcal{N}_{\Delta_{n,\Phi^\dagger},V_{\Phi^\dagger}^{-1}}(B)\right| > \epsilon\right)\to 0, \quad n\to\infty.$$

The proof of the above result follows the original proof in [31]. The main modifications are changing the almost sure convergence to convergence in probability, and adding a nuisance parameter which does not affect the proof. The nuisance

parameters, both in the parametric and nonparametric parts, can be omitted on purpose by using a misspecified model $\tilde{P}$. Thus, the theorem can be used with high flexibility. The theorem has two main assumptions: the LAN property (6), which defines the local smoothness of the model around a given point, and the posterior concentration property (7), which, in particular, determines the rate of convergence of the posterior distribution. Sufficient conditions for (7) can be found in Section 3 of [31]. Under these assumptions, we conclude that the posterior distribution of the parameter of interest can be approximated by a normal distribution. As the sample size grows, the posterior shrinks to a point which minimizes the Kullback-Leibler divergence within the model. It shares the same consistency property as the random quantity Δ_n in the LAN assumption. We will see later that Δ_n can be taken as the MLE of the misspecified model.

Remark 4.1. Condition (6) above is equivalent to

$$P_0^{(n)}\left[P_0^{(n)}\left(\sup_{h\in K}\left|\log\frac{\tilde{p}_{\Phi^\dagger+\delta_n h}}{\tilde{p}_{\Phi^\dagger}} - V_{\Phi^\dagger}\Delta_{n,\Phi^\dagger}h - \frac{1}{2}V_{\Phi^\dagger}h^2\right| > \zeta \;\middle|\; Y_\cdot\right) > \epsilon\right] \to 0,$$

for all $\zeta, \epsilon > 0$. This is weaker than the misspecified Bernstein-von-Mises Theorem in [31] when applying their theorem with $P_0^{(n)}(\cdot|\ Y_\cdot)$, which implies for almost all paths $Y_\cdot$,

$$P_0^{(n)}\left(\sup_{h\in K}\left|\log\frac{\tilde{p}_{\Phi^\dagger+\delta_n h}}{\tilde{p}_{\Phi^\dagger}} - V_{\Phi^\dagger}\Delta_{n,\Phi^\dagger}h - \frac{1}{2}V_{\Phi^\dagger}h^2\right| > \eta \;\middle|\; Y_\cdot\right) \to 0, \text{ for all } \epsilon > 0.$$

5. The misspecified model

Our methodology starts with a misspecified model ignoring the drift and the jump component. Namely, Y_t is assumed to follow

$$Y_{t_j} = X_{t_j} + \varepsilon_{t_j}, \quad \text{where} \quad X_t = \theta^{1/2}W_t. \tag{8}$$

This means that we first misinterpret the increments of the underlying process X as independent Gaussian variables, with mean zero, and variance θ.

Under the misspecified model (8), our target of estimation is still θ, but what it represents changes because of the misspecification. In the absence of jumps, θ measures the total variation of the underlying process X per unit time and, hence, it can be efficiently estimated by the scaled realized quadratic variation,

$$[X]_n := \frac{1}{T}\sum_{i=1}^{n}(\Delta_i X)^2, \tag{9}$$

which coincides with the MLE of the parameter θ in the underlying misspecified model $X_t = \theta^{1/2}W_t$. However, under the model $X_t = \theta^{1/2}W_t + J_t$, θ merely controls the variation of the continuous component and, in the infill limit, the realized quadratic variation (9) will aggregate both the true volatility, θ^*, and

the scaled variation introduced by the jump process J, $T^{-1}[J]$, recalling that $[J] = \sum_{s \leq T} (\Delta Y_s)^2$. Throughout, this total variation is denoted as

$$\theta^{\dagger} := \theta^{*} + T^{-1}[J], \tag{10}$$

which takes values on the random parameter domain

$$\Theta' := \left\{ \theta + T^{-1}[J]; \theta \in \Theta \right\}.$$

For any sample path of J, Θ' is an open set in $(0, +\infty)$, and $0 \notin \bar{\Theta}'$. Furthermore, there exists some deterministic constant $\delta_0 > 0$ such that $\Theta \subset (\delta_0, +\infty)$ and, hence, $\Theta' \subset (\delta_0, +\infty)$.

In § 5.1, we explicitly write the misspecified likelihood function and the corresponding MLE for θ under the model (8). Bayesian inference under this misspecified model is proposed in § 5.2. We will show that, given that the data Y is misinterpreted by the model (8), the posterior of θ can be approximated by a normal distribution. Further extensions are subsequently considered.

5.1. Misspecified likelihood function and MLE

Let us first note that, because of the presence of the noise ε, the increments $\Delta_j Y = Y_{t_j} - Y_{t_{j-1}}$, $j = 1, 2, \ldots, n$, are not independent. To deal with the dependence and write an explicit likelihood function, we follow [17] (see also [1]) and transform the observed data $\{\Delta_j Y\}_{j \leq n}$ into independent random variables $\{R_j\}_{j \leq n}$ via $\mathbf{R} = (P_n)(\Delta \mathbf{Y})$, where $\mathbf{R} = (R_1, \ldots, R_n)'$, $\Delta \mathbf{Y} = (\Delta_1 Y, \ldots, \Delta_n Y)'$, and P_n is a symmetric orthogonal matrix with entries

$$p_{ij}^{n} := \sqrt{\frac{2}{n+1}} \sin \frac{ij\pi}{n+1}, \quad i, j = 1, 2, \ldots, n.$$

[17] showed that, under the misspecified model (8), R_j is Gaussian distributed, with mean zero, and variance equal to

$$\lambda_j(\theta) := \lambda_j^n(\theta) := \theta \Delta_n + 2\sigma_{\varepsilon}^2 \left(1 - \cos \frac{j\pi}{n+1} \right), \quad j = 1, 2, \ldots, n.$$

For future reference, note that under the true model (2), the conditional distribution of R_j given J is

$$R_j | J \sim \mathcal{N} \left(\mu \Delta_n + \sum_{i=1}^{n} p_{ij} \Delta_i J, \lambda_j^n(\theta) \right). \tag{11}$$

Based on these Gaussian variables, the likelihood function of the parameters θ and σ_{ε}^2 given the data $\{\Delta_j Y\}_{j \leq n}$ can be explicitly written under the misspecified model. However, note that only θ is the parameter of interest, while σ_{ε}^2 is merely a nuisance parameter. Instead of writing the likelihood function based on $\lambda_j^n(\theta)$ and maximizing it over a two dimensional space, we replace the

nuisance parameter, σ_ε^2, with its consistent estimator $\hat{\sigma}_\varepsilon^2 = \frac{1}{2n}\sum_{j=1}^n \Delta_j Y^2$, and thereby obtain a pseudo-likelihood function for θ. The properties of $\hat{\sigma}_\varepsilon^2$ and the rationale of the replacement are further demonstrated in Lemmas B.3 and B.5. Then, it is natural to consider the following misspecified log likelihood function $\tilde{l}_n$ of θ given the data $\Delta Y_1, \Delta Y_2, \ldots, \Delta Y_n$:

$$\tilde{l}_n(\theta) = -\frac{1}{2}\sum_{i=1}^n \left\{ \log \lambda_j(\theta, \hat{\sigma}_\varepsilon^2) + \frac{R_j^2}{\lambda_j(\theta, \hat{\sigma}_\varepsilon^2)} \right\}, \tag{12}$$

where we set $\lambda_j(\theta, x) := \frac{\theta}{n} + 2x\left(1 - \cos\frac{j\pi}{n+1}\right)$. The corresponding MLE $\tilde{\theta}_n$ is the root of the score function

$$\dot{\tilde{l}}_n(\theta) = -\frac{1}{2n}\sum_{j=1}^n \left\{ \frac{1}{\lambda_j(\theta, \hat{\sigma}_\varepsilon^2)} - \frac{R_j^2}{\lambda_j(\theta, \hat{\sigma}_\varepsilon^2)^2} \right\}. \tag{13}$$

We further assume that the MLE $\tilde{\theta}_n$ is unique.

Remark 5.1. The misspecified likelihood function (12) can be simplified and directly applied to a model without microstructure noise (i.e., $Y = X$ in (2)) by taking $\sigma_\varepsilon^2 = 0$ and $\hat{\sigma}_\varepsilon^2 = 0$. Then,

$$\tilde{l}_n(\theta) = -\frac{1}{2}\sum_{j=1}^n \left\{ \log \theta\Delta_n + \frac{R_j^2}{\theta\Delta_n} \right\} = -\frac{1}{2}\sum_{j=1}^n \left\{ \log \theta\Delta_n + \frac{\Delta_j Y^2}{\theta\Delta_n} \right\}. \tag{14}$$

In this case, the MLE can be obtained in closed form as

$$\tilde{\theta}_n = \frac{1}{T}\sum_{i=1}^n R_i^2 = \frac{1}{T}\sum_{i=1}^n (\Delta_i Y)^2 = \frac{1}{T}\sum_{i=1}^n (\Delta_i X)^2. \tag{15}$$

Thus, the misspecified model is consistent with the one in [38] and, hence, the model with finite jump activity can be viewed as a special case of our results.

5.2. Bernstein-von Mises Theorems

We assume that the prior distribution Π for θ possesses a continuous and positive density π on $(\delta_0, +\infty)$. Denote P_* as the distribution of the process $\{Y_t\}_{t\geq 0}$ under the true model (2), and E_* as the corresponding expectation. Based on the prior Π and the likelihood function (12), we introduce the Gibbs posterior Π^n with temperature parameters κ_n ([50, 29]) as

$$\Pi_n(A) = \frac{\int_A \tilde{l}_n(\vartheta)^{1/\kappa_n}\pi(\vartheta)\,d\vartheta}{\int_\Theta \tilde{l}_n(\zeta)^{1/\kappa_n}\pi(\zeta)\,d\zeta}, \tag{16}$$

where A is a Borel set of $\mathbb{R}^+$. The Gibbs posterior increases the flexibility of the Bayesian procedure, which allows us to further correct for misspecification.

Specifically, the type of misspecification we deliberately utilize is to assume a model with only the interest parameter, ignoring the high-dimensional nuisance parameter. This causes the posterior for volatility to contract too quickly (relative to the correct, high-dimensional parameter model), making the Bayes estimator (e.g., the posterior mean) superefficient. Rescaling the likelihood flattens out the likelihood and also the posterior, slowing down the contraction of the posterior. Choosing the temperature parameter optimally will make the posterior contract at the efficient rate established by frequentist asymptotic analysis. We assume that κ_n converges in probability to a random variable $\kappa^\dagger$ as $n \to \infty$ under the true measure P_*. Note that κ_n may be data-dependent, and therefore it is possible that the random variable $\kappa^\dagger$ also depends on the data under P_*.

Our main result states that, as the sample size n increases, the misspecified posterior based on prior Π and the misinterpreted data $\{\Delta Y_i\}$ will be approximately normal and centered at the MLE $\tilde{\theta}_n$ obtained from the misspecified likelihood (12) under the true measure P_*. The asymptotic variance is equal to the temperature parameter $\kappa^\dagger$ times the inverse of the Fisher information of the misspecified likelihood. We establish our results in two broad settings. The first result covers situations where the microstructure noise can be ignored. This is the case when, for instance, with intermediate frequencies such as 5-minute or daily observations. In that setting, our procedure achieves the standard n^{-1} convergence rate. The second result covers the more realistic setting where the microstructure noise is explicitly incorporated. This is needed when working with ultra-high frequencies, and comes at the cost of a slower $n^{-1/2}$ rate.

Theorem 5.1. *Suppose that the data $Y_{t_0}, \ldots, Y_{t_n}$ is generated according to (2)-(4) with $\varepsilon_t \equiv 0$. Then, the misspecified posterior defined in (16) with $\tilde{l}_n$ given as in (14) and $\kappa_n \xrightarrow{P_*} \kappa^\dagger$, for some positive r.v. $\kappa^\dagger$, can be approximated by a normal distribution in the sense that*

$$TV\left(\Pi_n,\ \mathcal{N}(\tilde{\theta}_n, 2\kappa^\dagger \theta^{\dagger 2} n^{-1})\right) \xrightarrow{P_*} 0, \qquad \text{as } n \to \infty,$$

where TV represents the total variation distance, $\tilde{\theta}_n$ is the MLE (15), and $\theta^\dagger$ is defined in (10).

Theorem 5.2. *Under the framework and assumptions (N), (JD)-1, and (JF) above, the misspecified posterior Π^n defined in (16) with $\tilde{l}_n$ given as in (12) and $\kappa_n \xrightarrow{P_*} \kappa^\dagger$, for some positive r.v. $\kappa^\dagger$, is such that*

$$TV\left(\Pi_n,\ \mathcal{N}(\tilde{\theta}_n, 8\kappa^\dagger \theta^{\dagger 3/2} \sigma_\varepsilon n^{-1/2})\right) \xrightarrow{P_*} 0, \qquad \text{as } n \to \infty,$$

where $\tilde{\theta}_n$ is the corresponding MLE (i.e., the root of the score function (13)) and $\theta^\dagger$ is defined in (10).

Proofs of these theorems are given in the Appendix.

Remark 5.2. It is worth noting that Theorem 5.1 holds without any restriction on the Blumenthal-Getoor index α. In fact, this result holds for a large class of

pure-jump semimartingales J and even quite general stochastic volatility models (see §7). The restrictions on α and ν stated in Assumption (JD) are only needed when correcting the posterior as shown in the following section.

6. Correcting for misspecification

The main conclusions of Theorems 5.1 and 5.2, namely, as $n \to \infty$,

$$TV\left(\Pi_n,\ \mathcal{N}(\tilde{\theta}_n, V_{asy} n^{2\beta})\right) \xrightarrow{P_*} 0,$$

where $V_{asy} n^{2\beta} = 8\kappa^\dagger \theta^{\dagger 3/2} \sigma_\varepsilon n^{-1/2} 1_{\sigma_\varepsilon \neq 0} + 2\kappa^\dagger \theta^{\dagger 2} n^{-1} 1_{\sigma_\varepsilon = 0}$, state that the misspecified posterior Π_n is approximately normally distributed, and centered at $\tilde{\theta}_n$, which is a biased estimator of θ^* in the presence of jumps. Furthermore, the asymptotic variance may not be the most efficient either since we ignored the drift and the jump components on purpose. To adjust the bias and variance, what we need is a consistent estimator for the true parameter θ^*, which admits a feasible central limit theorem with the right rate of convergence. In what follows, we will first propose a general correction procedure and the corresponding Bernstein-von Mises theorem for any estimator with these two properties. Concrete instances of these estimators for both the no-noise and the general cases are presented thereafter.

Suppose we have an estimator $\hat{\theta}_n$ of θ^* such that

$$\hat{\theta}_n \xrightarrow{P_*} \theta^*, \qquad n^{-\beta}(\hat{\theta}_n - \theta^*) \xrightarrow{\mathcal{L}} \mathcal{N}(0, V), \quad \text{as } n \to \infty, \tag{17}$$

where, in accordance with Theorems 5.1 and 5.2, the rate of convergence β is $-\frac{1}{4}$ when $\sigma_\varepsilon \neq 0$, and $-\frac{1}{2}$ when $\sigma_\varepsilon = 0$.

Our goal is to adjust the posterior so that it centers at $\hat{\theta}_n$ and matches the asymptotic variance of $\hat{\theta}_n$. For the center, we simply shift the posterior by the right amount, while for the asymptotic variance, we adjust the temperature parameter. Concretely, define the estimator

$$\widehat{[J]}_n := T(\tilde{\theta}_n - \hat{\theta}_n). \tag{18}$$

The notation $\widehat{[J]}_n$ comes from the fact that this is a consistent estimator for the quadratic variation of the jump component J, because, as shown in the Appendix (see (30) and (60)), $\tilde{\theta}_n$ converges to $\theta^\dagger = \theta^* + T^{-1}[J]$ and $\hat{\theta}_n$ is a consistent estimator of θ^* by construction. We will then adjust the location of the posterior by subtracting $T^{-1}\widehat{[J]}_n$ (this operation will necessarily center the posterior at $\tilde{\theta}_n - T^{-1}\widehat{[J]}_n = \hat{\theta}_n$). To adjust the variance, we adopt a sequence of temperature parameters and its limit of the form:

$$\kappa_n = \frac{\hat{V}_n}{\hat{V}_{asy,n}} \quad \text{and} \quad \kappa^\dagger = \frac{V}{V_{asy}}, \tag{19}$$

where $\hat{V}_n$ and $\hat{V}_{asy,n}$ are suitable consistent estimators of V and V_{asy}, respectively. The choice of these estimators will be specified below in § 6.1–§ 6.2.

Finally, we can define the adjusted misspecified posterior $\widetilde{\Pi}_n$ as one having the density function

$$\widetilde{\pi}_n(\vartheta) = \pi_n\left(\vartheta + T^{-1}\widehat{[J]}_n\right), \tag{20}$$

where π_n is the misspecified posterior obtained in Theorems 5.1 and 5.2 with κ_n and $\kappa^\dagger$ defined in (19). Asymptotic normality of the adjusted posterior is established by the following result.

Theorem 6.1. *With the same conditions as in Theorem 5.2 or Theorem 5.1 except for the temperature parameter κ_n defined as in (19), the adjusted posterior $\widetilde{\Pi}_n$ defined above can be approximated by a normal distribution in the sense that,*

$$TV\left(\widetilde{\Pi}_n,\ \mathcal{N}\left(\hat{\theta}_n, Vn^{2\beta}\right)\right) \xrightarrow{P_*} 0 \quad \text{as } n \to \infty. \tag{21}$$

A location shift in Theorem 5.1 or Theorem 5.2 with κ_n defined in (19) gives us the proof of Theorem 6.1.

This theorem illustrates that any type of $1-\alpha$ credible interval ($CI_{B,\alpha}$) of $\widetilde{\Pi}_n$ is asymptotically the same as a $1-\alpha$ confidence interval for θ^* based on $\mathcal{N}(\hat{\theta}_n, Vn^{2\beta})$. The upper and lower bounds of the $CI_{B,\alpha}$ can then be approximated by $\hat{\theta}_n \pm \sqrt{Vn^{2\beta}}z_{\alpha/2}$ as $n \to \infty$, where $z_{\alpha/2}$ is the $\alpha/2$ quantile of the standard normal distribution. Since $\hat{\theta}_n$ satisfies a central limit theorem with asymptotic variance V, we have that

$$P_*(\theta \in CI_{B,\alpha}) \approx P_*(\theta \in \hat{\theta}_n \pm \sqrt{Vn^{2\beta}}z_{\alpha/2}) = P_*(\hat{\theta}_n \in \theta \pm \sqrt{Vn^{2\beta}}z_{\alpha/2}) \approx 1-\alpha.$$

Therefore, the $1-\alpha$ credible interval has approximately the correct repeated sampling coverage under P_*, which indicates frequentist validity of the Bayesian inference based on the adjusted posterior.

6.1. Correction for a model without microstructure noise

When the variance σ_ε^2 of the noise is 0 and (JD)-1 holds, we can use the thresholded realized quadratic variation of [37],

$$\hat{\theta}_n = \frac{1}{T}\sum_{i=1}^{n} \Delta_i Y^2 1_{|\Delta_i Y| \le \eta_n}, \tag{22}$$

to correct the misspecified posterior. Above, η_n is a threshold proportional to n^{-w} for some suitable exponent w. Consistency of $\hat{\theta}_n$ is established in [37] for any $w \in (0, 1/2)$ when J consists of the superposition of a general finite-jump activity process and an independent Lévy process. [10] showed that $\hat{\theta}_n$ satisfies a central limit theorem (CLT) with asymptotic variance $2\theta^{*2}n^{-1}$ under Assumption (JD)-1 provided that $w \in \left(\frac{1}{4-2\alpha}, \frac{1}{2}\right)$.

Under Assumption (JD)-2, we can adopt the estimator proposed by [27] (see also [33] for a closely related estimator) based on the empirical characteristic function. The corresponding CLT is established in Theorem 5 of [27] with asymptotic variance $2\theta^{*2}n^{-1}$ and rate $n^{-1/2}$. The data is divided into k_n nonoverlapping blocks, each of length v_n. In addition, we also need a scaling sequence u_n. It is then assumed that k_n and u_n satisfy

$$k_n n^{-\frac{1}{2}} \to 0, \quad k_n n^{\epsilon-\frac{1}{2}} \to \infty, \quad u_n \to 0, \quad \sup_n \frac{k_n}{u_n^4\sqrt{n}} < \infty,$$

for any $\epsilon > 0$ as $n \to \infty$. The estimator is defined as

$$\begin{aligned}
\hat{\theta}_n &= \hat{C}(u_n)_n - \frac{(\hat{C}(\zeta u_n)_n - \hat{C}(u_n)_n)^2}{\hat{C}(\zeta^2 u_n)_n - 2\hat{C}(\zeta u_n)_n + \hat{C}(u_n)_n}, \\
\hat{C}(u)_n &= v_n \sum_{j=0}^{\lfloor T/v_n \rfloor - 1} \left(C_j(u) - \frac{2}{u^2(k_n-1)} \left(\sinh\left(u^2 C_j(u)/2\right)\right)^2 \right), \qquad (23) \\
C_j(u) &= \frac{2}{u^2} \log\left(L_j(u) \vee k_n^{-\frac{1}{2}} \right), \quad L_j(u) = \frac{1}{k_n} \sum_{l=1}^{k_n-1} \cos\left(\sqrt{n} u \Delta_{jk_n+l+1} X\right),
\end{aligned}$$

where the notation $\lfloor a \rfloor$ defines the largest integer that is smaller than a, and ζ can be taken as any fixed value larger than 1. In fact, according to Theorem 3 of [27], the intermediate statistic $\hat{C}(u)_n$ is a consistent and efficient estimator when $r < 1$. It is asymptotically equivalent to the threshold estimator (22), and therefore can be used under Assumption (JD)-1 instead of (22).

With the estimator $\hat{\theta}_n$ described in different scenarios as above, we apply Theorem 6.1 with $\beta = -1/2$, $V = 2\theta^{*2}$, and the temperature parameters

$$\kappa_n = \left(\frac{\hat{\theta}_n}{\tilde{\theta}_n}\right)^2 \quad \text{and} \quad \kappa^\dagger = \left(\frac{\theta^*}{\theta^\dagger}\right)^2. \qquad (24)$$

By Slutsky's Theorem, $\kappa_n \to \kappa^\dagger$ in P_*-probability. We then obtain the following.

Corollary 6.2. *Using the same conditions as in Theorem 5.1 except for the temperature parameter κ_n defined as in (24), and assuming (JD), the adjusted posterior $\widetilde{\Pi}^n$ with density (20) can be approximated by a normal distribution in the sense that,*

$$TV\left(\widetilde{\Pi}_n,\ \mathcal{N}(\hat{\theta}_n, 2\theta^{*2}n^{-1})\right) \xrightarrow{P_*} 0 \qquad \text{as } n \to \infty.$$

Remark 6.1. As we will show in §7 below, the result above also holds for stochastic volatility models and more general pure-jump processes J.

6.2. Correction for the general model

When the variance of the noise, σ_ε^2, is not zero (the noise is present), one possible solution is to adopt the estimator $\hat{\Sigma}_n$ proposed in [30] (see also [5]), which

combines the thresholding approach of [37] with the pre-averaging method of [25] (see also [26] for a detailed exposition of the theory). The pre-averaging method is used to mitigate the effect of the noise ε. Utilizing this method, we formulate several overlapping blocks of increments, and calculate proxies of the increments of the uncontaminated process X by taking the weighted average of the increments of Y within each block. Then, the estimator is defined as the sum of the squares of those new quasi-increments that are less than some threshold, and is further debiased using an estimator of the variance of the noise. This estimator meets our requirements, when we include both infinitely many jumps with bounded variation and normally distributed microstructure noise. For completeness, we describe the key aspects of this estimator below.

The estimator depends on two parameters: the length of the block k_n and the weight function g. The latter satisfies the following regularity conditions:

- g is continuous on $[0,1]$, piecewise C^1 with a piecewise Lipschitz derivative g', and
- $g(0) = g(1) = 0$, and $\bar{g} = \int_0^1 g^2(s)\,ds < \infty$.

One simple and common choice is $g(s) = s \wedge (1-s)$. Next, for some constant c, let $k_n = \lfloor cn^{1/2} \rfloor$, $c_1 = c\bar{g}$, $c_2 = \int_0^1 (g'(s))^2\,ds/c$, and also define

$$\begin{aligned}
\hat{\Sigma}_n &= c_1^{-1}\left(n^{-1/2}U(Y,g)_n - c_2\hat{\sigma}_\varepsilon^2\right),\\
U(Y,g)_n &= \sum_{i=1}^{n-k_n} \left(\Delta_{i,k_n}^n Y(g)\right)^2 1\{|\Delta_{i,k_n}^n Y(g)| \le u_n\},\\
\Delta_{i,k_n}^n Y(g) &= \sum_{j=1}^{k_n-1} g(j/k_n)\Delta_{i+j}Y,
\end{aligned}$$

where we recall that $\hat{\sigma}_\varepsilon^2 = \frac{1}{2n}\sum_{j=1}^n \Delta_j Y^2$ and the threshold u_n satisfies

$$u_n n^{w_1} \to 0,\ u_n n^{w_2} \to \infty, \quad \text{as } n \to \infty,$$

for some $0 \le w_1 < w_2 < 1/4$ and $w_1 > 1/(8-4\beta)$. The estimator is consistent and admits a central limit theorem. More specifically, by Theorems 1 and 3 in [30], (17) holds with $\hat{\theta}_n = \hat{\Sigma}_n$, $\beta = -1/4$, and

$$V = V_{noise} := \frac{c}{\bar{g}^2}\left[4\theta^2\Phi_{22} + \frac{2\theta\sigma_\varepsilon^2}{c^2}\Phi_{12} + \frac{\sigma_\varepsilon^4}{c^4}\Phi_{11}\right],$$

where $\Phi_{ij} = \int_0^1 \phi_i(x)\phi_j(x)\,dx$, $\phi_1 = \int_x^1 g'(y)g'(y-x)\,dy$, and $\phi_2(x) = \int_x^1 g(y)g(y-x)\,dy$.

The temperature parameters in (19) can be defined as

$$\kappa_n = \frac{\frac{c}{\bar{g}^2}(4\Phi_{22}\hat{\Sigma}_n + \frac{2\Phi_{12}}{c^2}\hat{\Sigma}_n^{1/2}\hat{\sigma}_\varepsilon^2 + \frac{\Phi_{11}}{c^4}\hat{\sigma}_\varepsilon^4)}{8\tilde{\theta}_n^{3/2}\hat{\sigma}_\varepsilon} \quad \text{and} \quad \kappa^\dagger = \frac{V_{noise}}{8\theta^{\dagger 3/2}\sigma_\varepsilon}. \tag{25}$$

The convergence of κ_n to $\kappa^\dagger$ can be established through the consistency of $\hat{\Sigma}_n$ and $\hat{\sigma}_\varepsilon^2$ for θ^* and σ_ε^2, respectively, as well as the property that when $X_n = O_{P_*}(1)$ and $Y_n \xrightarrow{P_*} 0$, then $X_n Y_n \xrightarrow{P_*} 0$.

Then, we have the following corollary of Theorem 6.1.

Corollary 6.3. *With the same conditions as in Theorem 5.2 and with the temperature parameter κ_n defined as in (25), the adjusted posterior $\widetilde{\Pi}_n$ defined above can be approximated by a normal distribution in the sense that,*

$$TV\left(\widetilde{\Pi}_n,\ \mathcal{N}\left(\hat{\Sigma}_n, \frac{c}{\bar{g}^2}\left[4\theta^2\Phi_{22} + \frac{2\theta\sigma_\varepsilon^2}{c^2}\Phi_{12} + \frac{\sigma_\varepsilon^4}{c^4}\Phi_{11}\right] n^{-1/2}\right)\right) \xrightarrow{P_*} 0, \quad \text{as } n \to \infty.$$

7. Extension to more general semimartingales without noise

Thus far, we have assumed constant parameters for both the drift and diffusion components and a Lévy process for the jump component J. In this section, we show that, in fact, when the microstructure noise can be ignored, the purposely misspecified posterior approach can also be applied to stochastic volatility models and more general jump processes J. As mentioned before, it is generally believed that the microstructure noise is relatively negligible when using medium range frequencies such as 5-minute or daily observations.

We consider the model

$$dX_t = \beta_t dt + \sigma_t dW_t + dJ_t, \text{ for } t \in [0, T], \tag{26}$$

where W is a Wiener process, J is a suitable pure-jump semimartingale, and $\beta = \{\beta_t\}_{t\geq 0}$ and $\sigma = \{\sigma_t\}_{t\geq 0}$ are càdlàg adapted processes. The parameter of interest is the scaled integrated variance

$$\theta^* = \frac{1}{T}\int_0^T \sigma_t^2\, dt. \tag{27}$$

We again use the misspecified model (8) for X with $\varepsilon = 0$. The corresponding log likelihood function would then be the same as in Remark 5.1 with MLE

$$\tilde{\theta}_n = \frac{1}{T}\sum_{i=1}^n \Delta_i X^2. \tag{28}$$

An analysis of the proof of Theorem 5.1 reveals that the key for the result therein is the CLT stated in Lemma A.2. Specifically, what is needed is that the misspecified MLE (28) converges to (27) at the rate $O_p(n^{-1/2})$ (see Eq. (31) in the proof). [24] (see Theorem 2.12 and Remark 2.13 therein) shows an analogous CLT to that of Lemma A.2 (with the same rate of convergence) under the more general setting (26) when σ and J are of the form:

$$\sigma_t = \sigma_0 + \int_0^t \tilde{b}_s ds + \int_0^t \tilde{\sigma}_s dW_s + \int_0^t \tilde{\sigma}'_s dW'_s$$

$$
\begin{aligned}
&+\int_0^t\int \tilde{\delta}(s,x)\mathbf{1}_{\{|\tilde{\delta}(s,x)|\le 1\}}(\mu(ds,dx)-\nu(ds,dx))\\
&+\int_0^t\int \tilde{\delta}(s,x)\mathbf{1}_{\{|\tilde{\delta}(s,x)|> 1\}}\mu(ds,dx),\\
J_t=&\int_0^t\int \delta(s,x)\mathbf{1}_{\{|\delta(s,x)|\le 1\}}(\mu(ds,dx)-\nu(ds,dx))\\
&+\int_0^t\int \delta(s,x)\mathbf{1}_{\{|\delta(s,x)|> 1\}}\mu(ds,dx),
\end{aligned}
$$

where W' is a Wiener process independent of W and μ is a Poisson random measure on $\mathbb{R}_+\times\mathbb{R}$ with predictable compensator $\nu(ds,dx)=dsdx$, independent of (W,W'). The coefficients of σ and J (including $\delta:\Omega\times\mathbb{R}_+\times\mathbb{R}\to\mathbb{R}\backslash\{0\}$ and $\tilde{\delta}:\Omega\times\mathbb{R}_+\times\mathbb{R}\to\mathbb{R}\backslash\{0\}$) are random processes satisfying standard conditions for the integrals therein to be well defined.

As explained in §6, correcting the center and variance of the misspecified posterior Π_n requires an estimator $\hat{\theta}_n$ of θ^* enjoying a CLT with a rate of $n^{-1/2}$. For bounded variation jumps (Assumption (JD)-1), it turns out that the thresholded realized quadratic variation of [37], defined in (22), once again does the job. Specifically, [24] (see Theorems 2.4 and 2.11 therein) establishes a feasible CLT for (22) under the same framework as above.

For the extension to unbounded variation jumps (Assumption (JD)-2), more regularity conditions are required. We summarize them below.

Assumption (JI).

- When $\alpha\ge\frac{3}{2}$, we assume J is symmetric in the sense that J_t and $-J_t$ have the same law.
- We have a sequence τ_n of stopping times increasing to infinity, a sequence a_n of numbers, and a nonnegative Lebesgue-integrable function H on $\mathbb{R}$, such that the processes β, $\tilde{\sigma}$, and δ are càdlàg adapted, the coefficients δ, $\tilde{\delta}$ are predictable, the processes $\tilde{b}$, $\tilde{\sigma}'$ are progressively measurable, and

$$
\begin{aligned}
&t<\tau_n \quad\Rightarrow\quad |\tilde{\delta}(t,z)|^2\wedge 1\le a_nH(z),\\
&t<\tau_n,\quad V=\beta,\tilde{b},\tilde{\sigma},\tilde{\sigma}',\delta \quad\Rightarrow\quad |V_t|\le a_n,\\
&V=\beta,\tilde{\sigma},\delta \quad\Rightarrow\\
&\qquad |\mathbb{E}(V_{(t+s)\wedge\tau_n}-V_{t\wedge\tau_n}|\mathcal{F}_t)|+\mathbb{E}(|V_{(t+s)\wedge\tau_n}-V_{t\wedge\tau_n}|^2|\mathcal{F}_t)\le a_ns.
\end{aligned}
$$

Theorem 5 of [27] established a CLT of the characteristic-function-based estimator (23) with robustness under assumption (JI). Then, the misspecified posterior for the unbounded variational jumps can be corrected again by (23).

When the microstructure noise is taken into account, the extension is not as direct as for the no noise case, because after applying an orthonormal transformation to remove the autocovariance introduced by the noise, similar to that at the beginning of Section 5.1, the distribution of the transformed data does not

depend anymore only on the target parameter $\theta^* = T^{-1}\int_0^T \sigma_t^2\,dt$. Instead, the variance of each transformed data depends on a weighted sum of the 'volatility' of each increment. Analyzing the transformed data using the same procedure as before can only provide us an estimator of some value larger than the integrated volatility, but not the exact parameter θ^*.

8. Simulation

This section discusses the finite sample performance of the adjusted posterior defined in Theorem 6.1. We aim to show the plausibility of the limit (21) at a large sample size. This is demonstrated by comparing the empirical coverage probability of the credible interval derived from the adjusted posterior and the confidence interval from its corresponding asymptotic normal distribution in the theorem. We also aim to compare the "purposely misspecified" method with the frequentist central limit theorem (CLT) (17).

8.1. Infinite jump activity without noise

The jump component is set to be a variance gamma process

$$J_t = aG_t + bB_{G_t}, \tag{29}$$

where $a = -0.2$, $b = 0.2$, $\{G_t\}_{t\geq 0}$ is Gamma process such that $G_h \sim \Gamma(\Delta_n/c, c)$, with $c = 0.23$, and $\{B_t\}_{t\geq 0}$ is an Wiener process independent of the Wiener process W. For the drift and diffusion components, let $\mu = 0.1$ and $\theta = 0.3$. All parameter values are taken from [37]. For simplicity, we adopt the widely-used threshold $\eta_n = n^{-w}$, where $w \in (0, 0.5)$ and n is the sample size. This is a possible and conventional choice in terms of consistency and efficiency. The threshold η_n can also be calibrated using one of the iterative schemes proposed in [16]. These schemes were applied to the same model considered here (i.e., a Lévy process with variance Gamma jump component) and produced good results. For simplicity in what follows we fixed $\eta_n = n^{-w}$ with $w = 0.39$.

For the prior of θ, an inverse gamma distribution is assumed with shape and scale both equal to one. Since the temperature parameters do not affect conjugacy, the misspecified posterior and the adjusted posterior both follow inverse gamma distributions.

First, 5000 equally spaced observations are simulated based on the parameters defined above (sample size $n = 5000$). The adjusted posterior $\widetilde{\Pi}_n$ is constructed as in Corollary 6.2. The results are shown in Figure 2. The adjusted posterior for one possible sample path is plotted as the dashed line and compared with the corresponding asymptotic normal distribution $\mathcal{N}(\hat{\theta}_n, 2\theta^{*2}n^{-1})$ (the solid line). These two lines can hardly be distinguished from one another. Moreover, they are both roughly centered at the true volatility 0.3. This true volatility also lies between the dashed vertical lines which correspond to the 95% highest posterior density (HPD) interval of the adjusted posterior.

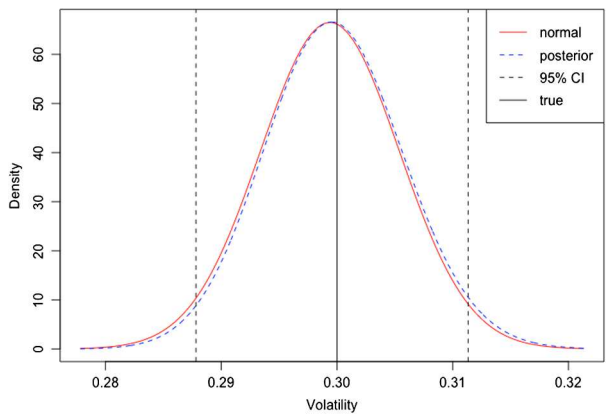


FIG 2. ***Comparison of adjusted posterior and asymptotic normal distribution for one sample path.*** *The solid line represents the asymptotic normal distribution in Corollary 6.2. The dashed line is the adjusted posterior. 95% HPD interval lies between the two black dashed lines.*

That the adjusted posterior is well-approximated by the asymptotic normal distribution is an illustration of Corollary 6.2. This indicates that the adjusted posterior will be centered at an efficient estimator with optimal variance when the sample size is large enough.

Next, we consider Bayesian point estimators associated with the adjusted direct posteriors for volatility. The biases of the means of two distributions defined in Corollary 6.2 are compared: the mean of the adjusted posterior $\widetilde{\Pi}_n$, and the mean of the asymptotic normal distribution $\hat{\theta}_n = \tilde{\theta}_n - T^{-1}\widehat{[J]}_n$, which is also the threshold estimator in [37]. We further consider the misspecified posterior adjusted by the latent realized quadratic variation of the jump component $T^{-1}[J]_n$ instead of $T^{-1}\widehat{[J]}_n$. The corresponding asymptotic normal distribution has mean $\hat{\theta}_n^* = \tilde{\theta}_n - T^{-1}[J]_n$. The analysis of these four point estimators is based on 1000 simulations. For each simulation, 5000 equally spaced observations are generated and used to calculate the biases.

The distribution of the biases is plotted in Figure 3. The solid line is formed by the biases of the threshold estimator, while the dashed line is formed by the biases of the mean of the adjusted posterior $\widetilde{\Pi}_n$. The dotted line represents the bias of $\hat{\theta}_n^*$. The bias of the mean of the adjusted posterior using the realized quadratic variation is represented by the dashed-dotted line.

The similarity of the solid and the dashed lines as well as the similarity of the dotted and the dashed-dotted lines suggest that the posterior mean and the mean of the asymptotic normal distribution have similar behavior in terms of their difference with the true volatility. The biases are relatively small since the volatility is 0.3 while most of the biases are within ± 0.01 of zero.

Remark 8.1. It may be possible to improve the accuracy of the adjusted posterior $\widetilde{\Pi}_n$ by using a better estimator of the quadratic variation of the jump

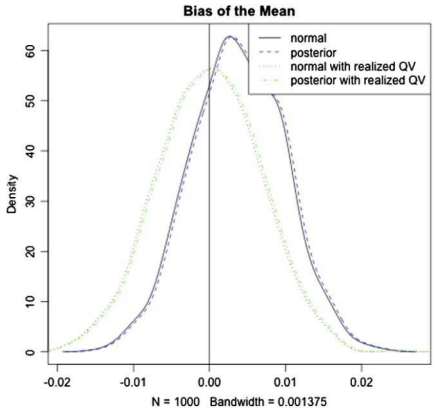


FIG 3. ***Bias of point estimators.*** *The biases of the mean of the asymptotic normal distribution in Corollary 6.2 form the solid lines. The dashed line is the distribution of the mean of the adjusted-posterior. The dotted and the dashed-dotted lines represent the distributions of the means of asymptotic normal distribution and posterior in Theorem 5.1 with location shift equal to the realized quadratic variation of the jumps.*

$[J]$ to correct the misspecified posterior Π_n defined in Theorem 5.1. While the dashed and the solid lines have higher probability for the positive values, the dotted and the dashed-dotted lines are more symmetric. This suggests that the right-skewed tendency of our posterior mean might be be due to poor estimates for the jump component.

We next consider frequentist properties of the posterior credible intervals obtained by, respectively, our proposed direct posterior and its large-sample normal approximation established by our theoretical results. Specifically, we study the frequentist accuracy of these credible intervals in the sense of nominal 95% credible intervals achieving the same repeated sampling coverage probability (0.95). This accuracy property is important because, while our Bernstein-von Mises theorems indicate large-sample frequentist validity of the posterior credible intervals (as $n \to \infty$), it is important to assess whether this validity property is approximately true for finite samples, and also if the posterior credible intervals tend to have frequentist coverage which is larger than the nominal level, which could indicate a lack of precision in the approximate Bayesian inference. In addition to assessing frequentist accuracy for these credible intervals, we also include a comparison of the frequentist coverage of nominal 95% frequentist confidence intervals derived using the CLT for the threshold estimator. For notational convenience, in this section, we use "CI" to represent both a credible interval and a confidence interval, with the meaning being clear from the context. For this simulation study, we increase the sample size n to 105,000, which is approximately the sample size corresponding to 5-minute observations during a 1-year time horizon. The empirical coverage probabilities of the 95% CIs based on 1000 repetitions are listed in Table 1. For each repetition, we simulate a sample path with 105,000 observations.

TABLE 1
Empirical coverage probabilities of Bayesian and frequentist interval estimates.

Empirical coverage probability	Distribution used to obtain 95% interval estimate
0.943	Asymptotic normal approximation $\mathcal{N}(\hat{\theta}_n, 2\theta^{*2}n^{-1})$ from BvM Theorem
0.944	HPD interval based on posterior $\tilde{\pi}_n(\theta) = \pi_n(\theta + \widehat{[J]}_n)$
0.940	Equal-tail credible interval based on posterior $\tilde{\pi}_n(\theta) = \pi_n(\theta + \widehat{[J]}_n)$
0.940	Frequentist CLT for threshold estimator and variance from [37]

In our simulations, the HPD interval has the highest coverage probability among all the CIs derived from various distributions defined above. However, all of the empirical coverage probabilities are slightly less than 0.95. Further studies are needed to ascertain whether this undercoverage phenomenon occurs in general, and to investigate the possible causes. Based on preliminary results, we conjecture that using a more refined frequentist estimator $\hat{\theta}_n$ to serve as the center of the posterior, or utilizing a less-misspecified model, may reduce the observed undercoverage. A complete analysis is beyond the scope of this paper.

Remark 8.2. Another issue which can affect the accuracy of the adjusted posterior or its asymptotic normal approximation is that the MLE is not actually approximating the posterior mean, and the Fisher information is not approximating the posterior variance. Asymptotically the effect of this non-Bayesian centering and scaling is negligible for finite-dimensional parameters in regular models, when considering errors of order smaller than the error incurred in the first-order normal approximation given by the Bernstein-von Mises Theorem (BvM). This is essentially because the BvM Theorem and asymptotic normality of the MLE are first-order approximations having the same order of approximation error. However, [32] show that centering the random parameter at the MLE instead of the posterior mean, and scaling by the Fisher information rather than the posterior standard deviation, can have substantial finite-sample effects on the properties of the centered and scaled posterior (e.g. the posterior cumulants), and can also make the first-order normal approximation given by the Bernstein-von Mises theorem less accurate in finite samples.

It is not necessary to use a conjugate prior as we have done in our simulations. Using certain non-informative priors (e.g. uniform) or an exponential distribution can be easily implemented in the simulations based on the same model. The results for these other priors are comparable with those we have reported for the inverse-gamma prior.

8.2. Lévy Model with microstructure noise

In order to illustrate the results in the presence of both infinitely many jumps and microstructure noise, we conduct simulations for the following model from

[30]:

$$X_t = W_t + J_t, \quad Y_{i/n} = X_{i/n} + \varepsilon_{i/n}, \quad \varepsilon_{i/n} \sim \mathcal{N}(0, 0.01^2),$$

for $i = 1, 2, \ldots, n$. The jump part J_t is a trimmed symmetric β-stable process with $\beta = 0.5$. The trimmed process means that after we simulate the increments of all the jumps, the largest 2% of them (ranked by absolute values) will be discarded to match the behavior of high-frequency tick-by-tick data. To allow a comparison, simulations are conducted based on exactly the same parameters described in the paper except one constant c, which determines the length of the preaveraging blocks according to $k_n = \lfloor c\Delta_n^{-1/2} \rfloor$. The choice of c is not clearly stated in [30]. Therefore, we choose the same $c = 1/3$ as in the original work [25]. The sample size is set as $n = 15,600$ and we have $\Delta_n = 1/7800$ as in [30].

For the adjusted posterior, the data is divided into two parts. The first half is used to evaluate the estimator $\hat{\Sigma}_n$ in [30], which is used in the prior, and the remainder is used to make inference. The prior is chosen to be a truncated normal distribution with lower boundary 0, centered at $\hat{\Sigma}_n$ obtained using the first half of the data, and standard deviation 0.06. We generate 25,000 MCMC samples and use the last 20,000 samples to construct the adjusted posterior (the first 5000 being discarded as burn-in samples). We repeat the experiment 1000 times, and hence we construct 1000 posteriors, each based on 20,000 MCMC samples.

To compare the point estimators, we compute the average bias and standard error of, respectively, the frequentist estimator $\hat{\Sigma}_n$, and the maximum a posteriori (MAP) point estimator defined as $\hat{\theta}_{MAP} = \arg\max_\theta \widetilde{\Pi}_n(\theta)$. Both the average bias and the standard errors are similar and small, providing some evidence of the accuracy of both the frequentist and Bayesian point estimators.

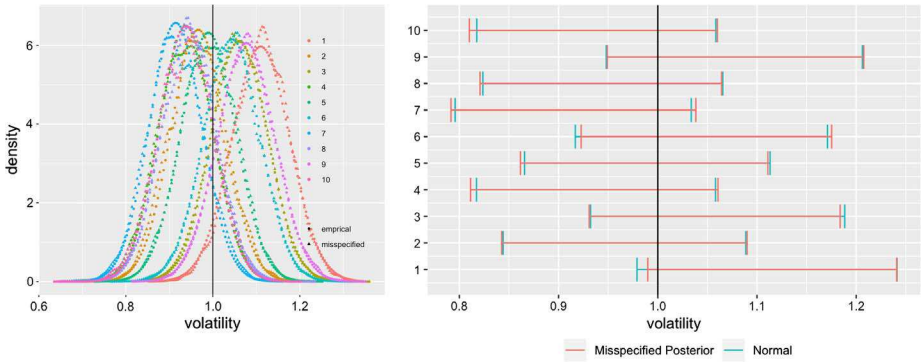
TABLE 2
Comparison of frequentist and Bayesian point estimators.

	Average Bias	Standard Error
$\hat{\Sigma}_n$	0.0131	0.0440
MAP	0.0110	0.0631

We also consider frequentist accuracy of interval estimates. The empirical coverage probability of the 95% credible interval of the adjusted posterior is slightly better than the that for the confidence interval derived from the CLT of the estimator $\hat{\Sigma}_n$.

TABLE 3
Empirical coverage probabilities of interval estimates for the model with noise.

Empirical Coverage Probability	Distribution used to obtain 95% interval estimate
0.945	CLT for the estimator $\hat{\Sigma}_n$
0.953	HPD interval based on the adjusted posterior
0.952	Asymptotic normal approximation from BvM Theorem



(a) Posteriors (b) Credible Intervals

FIG 4. ***Comparison with CLT for the model with noise.*** *(a) 10 different processes distinguished by 10 different colors are generated and the corresponding posterior are compared. Each color has two distributions. The one formed by little triangle is the misspecified posterior, while the other represents the Gibbs sampling results. (b) The red lines represent the 95% credible intervals calculated from the Gibbs sampling results of the joint posterior. The blue lines are the 95% HPD interval for the adjusted posterior.*

9. Conclusion

In this paper, we consider an infinite activity model with microstructure noise over a fixed time horizon. A "purposely misspecified" posterior is proposed for the volatility, the variation of the diffusion component. We prove that the posterior can be approximated by a normal distribution centered at a suitable estimator with the optimal variance. Simulation experiments illustrate the accuracy and frequentist validity of our proposed approximate Bayesian inference. Compared to [38], we generalize the feature of finitely many jumps to infinite jump activity, propose an extension to handle stochastic volatility and general Itô jump processes, and allow for microstructure noise.

Misspecification on purpose is an unusual idea, but we have shown that it can be an effective strategy to directly obtain a posterior on a parameter of interest in this complex model setting for which fully Bayesian inference is currently unavailable or intractable. Our proposal for Bayesian inference on volatility is the first procedure which can provide Bayesian inference on volatility with high-frequency data, while allowing infinite jump activity and microstructure noise. Moreover, our proposed direct posterior completely avoids the need to specify a prior on the complex nuisance component of the model, and does not require computationally-demanding construction of the full joint posterior to obtain the marginal posterior for volatility.

The recentering and rescaling procedure is also highly flexible. Any consistent and efficient estimator can be used as a correction. Furthermore, the variance can be adjusted in response to new information. For example, when the variance of the volatility θ^* is known, the temperature parameter can be set to achieve the optimal, efficient variance.

Appendix A: Proof of Theorem 5.1

Before we prove Theorem 5.1, we give some preliminary lemmas regarding the rate of convergence of the realized quadratic variation $[J]_n = \sum_{i=1}^n \Delta_i J^2 = \sum_{i=1}^n (J_{t_i} - J_{t_{i-1}})^2$ of a Lévy process J toward its quadratic variation $[J] = \sum_{s\le 1} \Delta J_s^2$. Without loss of generality, we assume $T = 1$ and let $\Delta_n = 1/n$.

The first result shows that, for bounded variation processes J, $[J]_n - [J] = O_P(n^{-1/2})$. The proof is classical and can be found in [44].

Lemma A.1. *Under Assumption (JD)-1,*

$$E_* \left| [J]_n - [J] \right| = O(n^{-1/2}).$$

The following lemma gives the rate of convergence of the realized quadratic variation of a general Lévy process with nonzero Brownian component ($\theta > 0$). The results is due to [23] (see Theorem 2.6 and Remark 5 therein). Related results for general semimartingales can be found in [24].

Lemma A.2. *Let*

$$QV_t = \sum_{i=1}^{[t/\Delta_n]} (\Delta_i^n X)^2, \quad t \ge 0,$$

be the realized quadratic variation of the process X defined in (1) with $\theta > 0$ (i.e., X is a Lévy process). Then,

$$\frac{1}{\sqrt{\Delta_n}} (QV_t - IV_t) \xrightarrow{\mathcal{D}} U_t, \tag{30}$$

where $IV_t = \theta t + \sum_{s\le t} (\Delta J_s)^2$ and

$$U_t = \sqrt{2\theta} W_t' + 2\sqrt{\theta} \int_0^t \int_{\mathbb{R}\setminus\{0\}} x Z_s \mu(ds, dx),$$

with W' being a Wiener process independent of W and $\{Z_s\}_{s\ge 0}$ being i.i.d. $\mathcal{N}(0,1)$ variables, independent of W and W'.

Proof of Theorem 5.1. We apply Theorem 4.1 with

$$Z_j = \Delta_j Y - \Delta_j J = \theta^{1/2} \Delta_j W, \quad Y_t = J_t, \quad \eta = 0, \quad P_0 = P_*,$$
$$\Phi = \theta + [J], \quad \Phi^\dagger = \theta^* + [J] = \theta^\dagger, \quad \delta_n = n^{-1/2}, \quad \log \tilde{p}(Y^{(n)} | \vartheta) \propto \tilde{l}_n(\vartheta)^{1/\kappa_n}.$$

We shall prove next that the two condition (6) and (7) are satisfied.

We start with the condition (7), which requires that, for every sequence of constants $M_n \to \infty$,

$$\Pi_n \left(|\vartheta - \theta^\dagger| > n^{-1/2} M_n \middle| Y^{(n)} \right) \xrightarrow{P_*} 0, \quad n \to \infty.$$

Using Markov's inequality,

$$\Pi_n\left(|\vartheta-\theta^\dagger|>n^{-1/2}M_n\middle|Y^{(n)}\right)\le nM_n^{-2}E_{\Pi_n}(\vartheta-\theta^\dagger)^2.$$

Since $\tilde{\theta}_n$ is the unique MLE, we could approximate the right hand side expectation by the Laplace approximation ([12]):

$$E_{\Pi_n}(\vartheta-\theta^\dagger)^2\approx(\tilde{\theta}_n-\theta^\dagger)^2\{1+O(n^{-1})\}.$$

Since $M_n\to\infty$, for condition (7), it suffices to show that, for $n\to\infty$,

$$|\tilde{\theta}_n-\theta^\dagger|=O_{P_*}(n^{-1/2}). \tag{31}$$

Since $\tilde{\theta}_n$ is the realized quadratic variation of the Lévy process X, this directly follows from Lemma A.2. Second, the models should satisfy the stochastic local asymptotic normality (LAN) condition (6). That means that, for every $\epsilon>0$,

$$P_*\left(\sup_{h\in K}\left|\frac{1}{\kappa_n}\log\frac{\tilde{l}_n(\theta^\dagger+\delta_n h)}{\tilde{l}_n(\theta^\dagger)}-V_{\theta^\dagger}\Delta_{n,\theta^\dagger}h-\frac{1}{2}V_{\theta^\dagger}h^2\right|>\epsilon\right)=o(1).$$

Let $V_{\theta^\dagger}=(2\kappa^\dagger\theta^{\dagger 2})^{-1}$ and $\Delta_{n,\theta^\dagger}=n^{1/2}\left(\tilde{\theta}_n-\theta^\dagger\right)$. Using Taylor expansion to approximate the log likelihood and plugging in the first and second derivatives, the left hand side of the LAN condition can be written as

$$P_*\left(\sup_{h\in K}\left|-\frac{h}{2\kappa^\dagger\theta^{\dagger 2}}\left(1-\frac{\kappa^\dagger}{\kappa_n}\right)\sqrt{n}\left(\tilde{\theta}_n-\theta^\dagger\right)+\frac{h^2}{4\kappa^\dagger\theta^{\dagger 2}}\left[\frac{2\kappa^\dagger}{\theta^\dagger\kappa_n}\sqrt{n}\left(\tilde{\theta}_n-\theta^\dagger\right)+\left(1-\frac{\kappa^\dagger}{\kappa_n}\right)\right]\right|>\epsilon\right).$$

Noted that $P_*(|\kappa_n-\kappa^\dagger|>\delta)\to 0$ when $n\to\infty$ for arbitrary $\delta>0$ and, from the result obtained in Lemma A.2, the LAN condition holds. □

Appendix B: Proof of Theorem 5.2

As in the proof of Theorem 5.1, we apply Theorem 4.1 with

$$Z_j=\Delta_jY-\Delta_jJ=\theta^{1/2}\Delta_jW+\varepsilon_{t_j}-\varepsilon_{t_{j-1}},\quad Y_t=J_t,\quad \eta=\sigma_\varepsilon^2,\quad P_0=P_*,$$
$$\Phi=\theta+[J],\quad \Phi^\dagger=\theta^*+[J]=\theta^\dagger,\quad \delta_n=n^{-1/4},\quad \log\tilde{p}(Y^{(n)}|\vartheta)\propto\tilde{l}_n(\vartheta)^{1/\kappa_n}.$$

As before, there are two conditions that need to be satisfied. The first is the LAN property (6), which will be proved in § B.4. The second condition is (7). By applying the same Markov inequality and Laplace approximation as in the proof of Theorem 5.1, we can conclude that a sufficient condition for (7) is

$$|\tilde{\theta}_n-\theta^\dagger|=O_{P_*}(n^{-1/4}),$$

which will be proved in § B.3. Before we give some preliminary lemmas in § B.1.

The following notations are often used throughout the proof:

1. $a_n \lesssim b_n$ indicates that there exists a constant C such that $|a_n| \leq C|b_n|$ for every n large enough. If $a_n \lesssim b_n$ and $b_n \lesssim a_n$, then we write $a_n \asymp b_n$.
2. To simpify the notation, in what follows, we use p_{ij} to represent p_{ij}^n, and λ_j to represent λ_j^n.

B.1. Preliminary lemmas

Without loss of generality, we assume $E_*[\Delta_j J] = 0$ (otherwise, the drift μ can be redefined to $\mu + E_*[\Delta_j J]/\Delta_n$. We start we collecting some useful properties of the orthogonal matrix P_n defined in § 2.1 (the proof can be found in [44]):

Lemma B.1. *We have the following relationships:*

$$\sum_{j=1}^n p_{ij}^2 = \sum_{i=1}^n p_{ij}^2 = 1, \quad \sum_{j=1}^n p_{ij}p_{kj} = 0, \quad i \neq k, \tag{32}$$

$$\sum_{j=1}^n p_{ij}^4 = \frac{3}{2(n+1)}, \quad \sum_{j=1}^n p_{ij}^2 p_{kj}^2 = \frac{1}{n+1}, \quad \left|\sum_{i\neq k\neq u} p_{is}p_{it}p_{ks}p_{ut}\right| = O(1), \tag{33}$$

$$\left|\sum_{i\neq k} p_{is}p_{it}p_{ks}p_{kt}\right| = \frac{1}{n+1}, \quad \left(\sum_{i=1}^n p_{ij}\right)^2 \lesssim \frac{n}{j}. \tag{34}$$

The limiting behavior of the moments of the jump increments will be frequently used later in the proof. We summarize it in the following lemma.

Lemma B.2. *Under Assumptions (N)–(JF), for $k = 1, \ldots, 16$, and $i = 1, \ldots, n$,*

$$\text{(i) } E_*\left[|\Delta_i J|^k\right] = O(n^{-1}), \quad \text{(ii) } E_*\left[\left(\frac{\mu}{n} + \sum_{j=1}^n p_{ij}\Delta_j J\right)^k\right] = O(n^{-1}) \tag{35}$$

$$\text{(iii) } E_*\left[\left(\frac{\mu}{n} + \sum_{j=1}^n p_{ij}\Delta_j J\right)^4\right] = O(n^{-2}). \tag{36}$$

The first statement directly follows from Theorem 4.3 of [14]. The proofs of the second and the third statements can be found in [44].

Recall that σ_ε^2 is the variance of the noise, which can be estimated using $\hat{\sigma}_\varepsilon^2 = \frac{1}{2n}\sum_{i=1}^n \Delta_i Y^2$. The following result states some needed properties of $\hat{\sigma}_\varepsilon$.

Lemma B.3. *Under assumption (JF), we have*

$$P_*\left(\left|\hat{\sigma}_\varepsilon^2 - \sigma_\varepsilon^2\right| \geq \frac{\sigma_\varepsilon^2}{2}\right) = O(n^{-4}), \quad \text{and} \quad E_*|\hat{\sigma}_\varepsilon^2 - \sigma_\varepsilon^2| = O(n^{-1/2}).$$

Proof. Let $H = \sqrt{n}\left(\hat{\sigma}_\varepsilon^2 - \sigma_\varepsilon^2\right)$. When H has finite 8th moment, by Markov's inequality, we have

$$P_*\left(\left|\hat{\sigma}_\varepsilon^2-\sigma_\varepsilon^2\right|\geq\frac{\sigma_\varepsilon^2}{2}\right)=P_*\left(n^4\left|\hat{\sigma}_\varepsilon^2-\sigma_\varepsilon^2\right|^8\geq n^4\frac{\sigma_\varepsilon^{16}}{2^8}\right)\leq\frac{2^8E_*\left[H^8\right]}{n^4\sigma_\varepsilon^{16}}=O(n^{-4}),$$

$$E_*\left[\left|\hat{\sigma}_\varepsilon^2-\sigma_\varepsilon^2\right|^2\right]=E_*\left[H^2\right]/n=O(n^{-1}).$$

Then, it suffices to demonstrate that H has finite 8th moment. We prove this by analyzing the moment generating function (m.g.f.) of H conditioned on J, and showing that the 8th derivative of that m.g.f. is finite at zero.

Because P_n is an orthogonal matrix, we have $\hat{\sigma}_\varepsilon^2=\frac{1}{2n}\sum_{j=1}^n(\Delta_jY)^2=\frac{1}{2n}\sum_{j=1}^nR_j^2$. Then, H can be written as $\frac{1}{2\sqrt{n}}\sum_{j=1}^nR_j^2-\sqrt{n}\sigma_\varepsilon^2$. Using (11) and the m.g.f. of a noncentral χ^2-distribution, the m.g.f. of H conditioned on J is

$$M_H(t)=e^{-\sqrt{n}\sigma_\varepsilon^2t}\prod_{i=1}^n\exp\left(\frac{\alpha_it_i}{1-2t_i}\right)(1-2t_i)^{-\frac{1}{2}},\quad\text{where } t_i=\frac{\lambda_it}{2\sqrt{n}}.$$

It is not hard to see that its kth-derivative $M^{(k)}(t):=\frac{d^k}{dt^k}M_H(t)$ takes the form:

$$M^{(8)}(t)=-\sqrt{n}\sigma_\varepsilon^2M^{(7)}+$$
$$\sum_{k=0}^7C_k\left[\sum_{i=1}^n\frac{k+1}{(1-2t_i)^{k+2}}\frac{\alpha_i\lambda_i^{k+1}}{(2\sqrt{n})^{k+1}}+\sum_{i=1}^n\frac{1}{(1-2t_i)^{k+1}}\frac{\lambda_i^{k+1}}{(2\sqrt{n})^{k+1}}\right]M^{(7-k)},$$

for some constants C_k. Let $H_k=\sum_{i=1}^n\left(\frac{\mu}{n}+\sum_{j=1}^np_{ij}\Delta_jJ\right)^2\lambda_i^k$ and $\Lambda^k=\sum_{i=1}^n\lambda_i^k$. Defined $c_{n,k}=M^{(k)}(0)2^kn^{k/2}$. As it turns out $c_{n,8}$ can be expressed as

$$c_{n,8}=(\theta^*+H_0)^{m_1}\prod_{k=2}^8\left((k+1)H_k+\Lambda^k\right)^{m_k}\lesssim(\theta^*+H_0)^{m_1}\prod_{k=2}^8\left(H_0+\Lambda^k\right)^{m_k}$$
$$=:\bar{c}_{n,8},$$

where $\sum_{k=1}^8km_k=8$, $m_k\in\mathbb{N}$. By (35)-(36), and following a similar procedure as that for proving (35), for any $1\leq k\leq 8$,

$$E_*[H_0^k]=E_*\Big[\Big(\sum_{i=1}^n\Big(\frac{\mu}{n}+\sum_{j=1}^np_{ij}\Delta_jJ\Big)^2\Big)^k\Big]=O(1).$$

Then,

$$E_*[H_k]\leq E_*\Big[\sum_{i=1}^n\Big(\frac{\mu}{n}+\sum_{j=1}^np_{ij}\Delta_jJ\Big)^2\left(\frac{\theta^*}{n}+4\sigma_\varepsilon^2\right)^k\Big]\lesssim E_*[H_0]=O(1),$$

$$\Lambda^k\leq\sum_{j=1}^n\left(\theta^*+2\sigma_\varepsilon^2\right)^k=O(n).$$

Thus, all the terms in $\bar{c}_{n,8}$ have expectations of $O(n^4)$. The rate n^4 can be only achieved when $m_2=4, m_j=0, j=1,3,4,\ldots,8$. Thus, $E_*[c_{n,8}]=O(n^4)$. This implies that H has 8th finite moment and completes the proof. □

The following Lemma will be used later to prove the asymptotic properties of the log likelihood function (12), and its derivative.

Lemma B.4. *For any fixed constants $a > 0, b > 0$, and $p \geq 1$,*

$$\sum_{j=1}^{n} \frac{1}{\left(a + 2bn(1 - \cos\frac{j\pi}{n+1})\right)^p} \asymp n^{\frac{1}{2}}.$$

Proof. For the lower bound, note that since $\sin x \leq x$,

$$\begin{aligned}\sum_{j=1}^{n} \frac{1}{\left(a + 2bn(1 - \cos\frac{j\pi}{n+1})\right)^p} &\geq \sum_{j=1}^{\sqrt{n}} \frac{1}{\left(a + 4bn(\sin\frac{j\pi}{2(n+1)})^2\right)^p} \\ &\geq \sum_{j=1}^{\sqrt{n}} \frac{1}{\left(a + 4bn\frac{j^2\pi^2}{4(n+1)^2}\right)^p} \overset{j\leq\sqrt{n}}{\geq} \sum_{j=1}^{\sqrt{n}} \frac{1}{(a + b\pi^2)^p} \gtrsim \sqrt{n}.\end{aligned}$$

For the upper bound, we divide the summation into two parts. For $j \leq \sqrt{n}$,

$$\sum_{j=1}^{\sqrt{n}} \frac{1}{\left(a + 2bn(1 - \cos\frac{j\pi}{n+1})\right)^p} \leq \sum_{j=1}^{\sqrt{n}} \frac{1}{a^p} \lesssim \sqrt{n}.$$

For $j > \sqrt{n}$, since $\sin x > \frac{2}{\pi}x$ for $0 < x < \frac{\pi}{2}$,

$$n\left(1 - \cos\frac{j\pi}{n+1}\right) = 2n\sin^2\frac{j\pi}{2(n+1)} \geq 2\frac{j^2 n}{(n+1)^2} > \frac{j^2}{n+1}.$$

Then,

$$\sum_{j=\sqrt{n}}^{n} \frac{1}{\left(a + 2bn(1 - \cos\frac{j\pi}{n+1})\right)^p} \leq \sum_{j=\sqrt{n}}^{n} \left(\frac{n+1}{2bj^2}\right)^p \leq \int_{\sqrt{n}-1}^{n} \left(\frac{n+1}{2bx^2}\right)^p dx \lesssim \sqrt{n}.$$

□

Recall that $\lambda_j(\theta) = \frac{\theta}{n} + 2\sigma_\varepsilon^2\left(1 - \cos\frac{j\pi}{n+1}\right)$. Applying Lemma B.4 with $a = \theta$ and $b = \sigma_\varepsilon^2$, and since $0 \notin \bar{\Theta}'$, we get

$$\sup_{\theta\in\bar{\Theta}'} \sum_{j=1}^{n} \frac{1}{n^p \lambda_j^p(\theta)} = O(n^{1/2}), \quad p = 2, 3, \ldots \tag{37}$$

B.2. Likelihood functions

In this section, we introduce several properties of the misspecified likelihood function $\tilde{l}_n$ defined in (12). In the misspecified model (8), when the variance of the noise is assumed to be known, the likelihood function of θ is given by

$$l_n(\theta) = -\frac{1}{2}\sum_{i=1}^{n}\left\{\log\lambda_j(\theta, \sigma_\varepsilon^2) + \frac{R_j^2}{\lambda_j(\theta, \sigma_\varepsilon^2)}\right\}, \tag{38}$$

where recall that $\lambda_j(\theta) := \lambda_j(\theta, \sigma_\varepsilon^2) := \frac{\theta}{n} + 2\sigma_\varepsilon^2\left(1 - \cos\frac{j\pi}{n+1}\right)$. In what follows, we denote the corresponding first and second derivatives of l_n and $\tilde{l}_n$ with respect to θ as $\dot{l}_n$, $\dot{\tilde{l}}_n$, $\ddot{l}_n$, and $\ddot{\tilde{l}}_n$, respectively.

The moments of the variable R_j are frequently used below. We summarize them here. From (11) and (35)-(36), the moments of R_j are such that

$$E_* R_j^2 = \lambda_j(\theta^*) + O(n^{-1}), \tag{39}$$

$$E_* R_j^4 \lesssim n^{-2} + n^{-1}\lambda_j(\theta^*) + \lambda_j^2(\theta^*) \lesssim \lambda_j^2(\theta^*) + O(n^{-2}). \tag{40}$$

The last inequality holds because $n^{-1} \le \lambda_j(\theta^*)/\theta^*$.

The following result aims to control the difference between l and $\tilde{l}$ and their corresponding derivatives. Recall that l uses the true variance σ_ε^2, while $\tilde{l}$ adopts $\hat{\sigma}_\varepsilon^2$ to replace σ_ε^2.

Lemma B.5. *Let l and $\tilde{l}$ be given as in (12) and (38), respectively. If assumptions (N) and (JF) hold true, then for any integer $k \ge 1$,*

$$\sup_{\theta\in\Theta'}\left|\frac{d^k l_n(\theta)}{d\theta^k} - \frac{d^k \tilde{l}_n(\theta)}{d\theta^k}\right| = O_{P_*}(1). \tag{41}$$

Proof. The expressions inside the absolute values in (41) can be expressed as

$$\sum_{j=1}^n \left[g_j(\hat{\sigma}_\varepsilon^2) - g_j(\sigma_\varepsilon^2)\right] + \sum_{j=1}^n nR_j^2\left[h_j(\hat{\sigma}_\varepsilon^2) - h_j(\sigma_\varepsilon^2)\right] =: \sum_{j=1}^n \alpha_j + \sum_{j=1}^n \beta_j =: G_n + H_n,$$

where

$$g_j(x) = \frac{(-1)^k(k-1)!}{2n^k\lambda_j^k(\theta, x)}, \quad h_j(x) = \frac{(-1)^{k+1}k!}{2n^{k+1}\lambda_j^{k+1}(\theta, x)}.$$

We first derive an upper bound for the first and second derivative of g and h. Fixed a $\tau > 0$, and consider all $x \ge \tau$. Noted that $2n\left(1 - \cos\frac{j\pi}{n+1}\right) \le n\lambda_j(\theta, x)/x \le n\lambda_j(\theta, x)/\tau$, we have

$$\begin{aligned}\sum_{j=1}^n \left|\frac{\partial^m}{\partial x^m}\frac{1}{n^k\lambda_j^k(\theta, x)}\right| &= \sum_{j=1}^n \frac{2^m n^m (1-\cos\frac{j\pi}{n+1})^m}{(n\lambda_j(\theta, x))^{m+k}} \\ &\le \frac{1}{\tau^m}\sum_{j=1}^n \frac{1}{(n\lambda_j(\theta, x))^k} \overset{n\lambda_j(\theta,x)\ge\theta}{\le} \frac{1}{\tau^m\theta^{k-1}}\sum_{j=1}^n \frac{1}{n\lambda_j(\theta, \tau)}.\end{aligned}$$

By (37),

$$\sup_{\theta\in\Theta'}\sup_{x>\tau}\sum_{j=1}^n |g_j'(x)| = O(n^{1/2}), \quad \sup_{\theta\in\Theta'}\sup_{x>\tau}\sum_{j=1}^n |g_j''(x)| = O(n^{1/2}), \tag{42}$$

$$|h_j'(\tau)| \lesssim \frac{1}{n^2\lambda_j^2(\theta, \tau)}, \quad \sup_{x>\tau}|h_j''(x)| \lesssim \frac{1}{n^2\lambda_j^2(\theta, \tau)}. \tag{43}$$

Next, let $\tau = 0$. Noted that $\lambda_j(\zeta, x) \geq \lambda_j(\theta, x) \geq \frac{\theta}{n}$, and

$$\frac{\partial^2}{\partial x^2} \frac{1}{n^k \lambda_j^k(\theta, x)} = \frac{2^2 n^2 (1 - \cos \frac{j\pi}{n+1})^2}{n^{2+k} \lambda_j^{2+k}(\theta, x)} \lesssim \frac{n^2}{\theta^{2+k}} = O(n^2),$$

we have

$$\sup_{\theta \in \Theta} \sup_{x \geq 0} |g_j''(x)| = O(n^2), \quad \sup_{\theta \in \Theta} \sup_{x \geq 0} |h_j''(x)| = O(n^2). \tag{44}$$

Now, we are ready to prove the boundedness of G_n and H_n uniformly for $\theta \in \Theta'$. Let $\mathcal{D}_n$ be the interval between $\hat{\sigma}_\varepsilon^2$ and σ_ε^2. By Taylor expansion, $\alpha_j = g_j(\hat{\sigma}_\varepsilon^2) - g_j(\sigma_\varepsilon^2)$ can be bounded:

$$\sum_{j=1}^n |\alpha_j| \leq |\hat{\sigma}_\varepsilon^2 - \sigma_\varepsilon^2| \sum_{j=1}^n |g_j'(\sigma_\varepsilon^2)| + \frac{1}{2} \sum_{j=1}^n \sup_{x \in \mathcal{D}_n} |g_j''(x)| \cdot |\hat{\sigma}_\varepsilon^2 - \sigma_\varepsilon^2|^2$$

The first term is $O_{P_*}(1)$ by Lemma B.3, and (42) with $\tau = \sigma_\varepsilon^2/2$. For the second term, we break up $\mathcal{D}_n$ into two sub-regions: $\{x > \sigma_\varepsilon^2/2\} \cap \mathcal{D}_n$ and $\{x \leq \sigma_\varepsilon^2/2\} \cap \mathcal{D}_n$. Then, $\sum_{j=1}^n \sup_{x \in \mathcal{D}_n} |g_j''(x)| \cdot |\hat{\sigma}_\varepsilon^2 - \sigma_\varepsilon^2|^2$, can be bounded by

$$\sum_{j=1}^n \sup_{x > \sigma_\varepsilon^2/2} |g_j''(x)| \cdot |\hat{\sigma}_\varepsilon^2 - \sigma_\varepsilon^2|^2 + \sum_{j=1}^n \sup_{x \leq \sigma_\varepsilon^2/2} |g_j''(x)| \cdot |\hat{\sigma}_\varepsilon^2 - \sigma_\varepsilon^2|^2 1_{\left\{ \frac{\sigma_\varepsilon^2}{2} \leq |\hat{\sigma}_\varepsilon^2 - \sigma_\varepsilon^2| \leq \sigma_\varepsilon^2 \right\}}.$$

In the first sub-region, x is bounded away from zero, so we can still apply Lemma B.3, and (42) with $\tau = \sigma_\varepsilon^2/2$, and obtain $O_{P_*}(1)$. For the second sub-region, $x \leq \sigma_\varepsilon^2/2$, which means x lies far away from the true variance, σ_ε^2. If this sub-region is not empty, then, $0 \leq \hat{\sigma}_\varepsilon^2 \leq \sigma_\varepsilon^2/2$. This is covered by the region $\frac{\sigma_\varepsilon^2}{2} \leq |\hat{\sigma}_\varepsilon^2 - \sigma_\varepsilon^2| \leq \sigma_\varepsilon^2$. Thus, we can bound $|\hat{\sigma}_\varepsilon^2 - \sigma_\varepsilon^2|^2$ using σ_ε^4. By (44),

$$\begin{aligned} &E_* \Big[\sum_{j=1}^n \sup_{x \leq \sigma_\varepsilon^2/2} |g_j''(x)| \cdot |\hat{\sigma}_\varepsilon^2 - \sigma_\varepsilon^2|^2 1_{\left\{ \frac{\sigma_\varepsilon^2}{2} \leq |\hat{\sigma}_\varepsilon^2 - \sigma_\varepsilon^2| \leq \sigma_\varepsilon^2 \right\}} \Big] \\ &\quad \leq E_* \Big[\sum_{j=1}^n n^2 2 \sigma_\varepsilon^4 1_{\left\{ |\hat{\sigma}_\varepsilon^2 - \sigma_\varepsilon^2| \geq \frac{\sigma_\varepsilon^2}{2} \right\}} \Big] = 2n^3 \sigma_\varepsilon^4 P_* \left(|\hat{\sigma}_\varepsilon^2 - \sigma_\varepsilon^2| \geq \frac{\sigma_\varepsilon^2}{2} \right) = O(n^{-1}). \end{aligned}$$

The last equality is based on Lemma B.3. Therefore, $G_n = O_{P_*}(1)$. For H_n, similarly, using Taylor expansion,

$$\sum_{j=1}^n |\beta_j| \leq \sum_{j=1}^n |h_j'(\sigma_\varepsilon^2)| \cdot |n R_j^2 (\hat{\sigma}_\varepsilon^2 - \sigma_\varepsilon^2)| + \frac{1}{2} \sum_{j=1}^n \sup_{x \in \mathcal{D}_n} |h_j''(x)| \cdot |n R_j^2 (\hat{\sigma}_\varepsilon^2 - \sigma_\varepsilon^2)^2|. \tag{45}$$

For the first term of (45), recalling that $\hat{\sigma}_\varepsilon^2 = \frac{1}{2n} \sum_{i=1}^n R_i^2$, in order to obtain the expectation of $R_j^2 (\hat{\sigma}_\varepsilon^2 - \sigma_\varepsilon^2)$, we take expectation of R_j^2 and $\hat{\sigma}_\varepsilon^2 - \sigma_\varepsilon^2$ seperately, subtract the term $\left(E_* R_j^2\right)^2 / (2n)$, and then add back the term $E_* R_j^4 / (2n)$:

$$E_* \left| n R_j^2 (\hat{\sigma}_\varepsilon^2 - \sigma_\varepsilon^2) \right| \leq n E_* R_j^2 \cdot E_* \left| \hat{\sigma}_\varepsilon^2 - \sigma_\varepsilon^2 \right| + n E_* R_j^2 \cdot \frac{1}{2n} E_* R_j^2 + \frac{1}{2} E_* R_j^4.$$

All the three terms above can be bounded applying Lemma B.3, (39), and (40):

$$nE_*R_j^2 \cdot E_* \left|\hat{\sigma}_\varepsilon^2 - \sigma_\varepsilon^2\right| \lesssim n(\lambda_j(\theta^*) + n^{-1})n^{-1/2} \lesssim n^{\frac{1}{2}}\lambda_j(\theta^*) + n^{-\frac{1}{2}},$$
$$nE_*R_j^2 \cdot \frac{1}{2n}E_*R_j^2 \lesssim (\lambda_j(\theta^*) + n^{-1})^2 \lesssim \lambda_j(\theta^*) + n^{-2},$$
$$E_*R_j^4 = \lambda_j^2(\theta^*) + n^{-2} \lesssim \lambda_j(\theta^*) + n^{-2}.$$

The last inequality is because $\lambda_j(\theta) \le \theta + 2\sigma_\varepsilon^2$. We then have $E_*\left|nR_j^2(\hat{\sigma}_\varepsilon^2 - \sigma_\varepsilon^2)\right| \lesssim n^{\frac{1}{2}}\lambda_j(\theta^*) + n^{-\frac{1}{2}}$. By (43),

$$E_*\sum_{j=1}^n |nR_j^2(\hat{\sigma}_\varepsilon^2 - \sigma_\varepsilon^2)||h_j'(\sigma_\varepsilon^2)| \lesssim \sum_{j=1}^n \left(\frac{n^{\frac{1}{2}}}{n^2\lambda_j(\theta)} + \frac{n^{-\frac{1}{2}}}{n^2\lambda_j^2(\theta)}\right) \overset{(37)}{=} O(1).$$

The expectation of the first term of (45) is bounded.

Similarly to the proof of the second term of $G_n = \sum_{i=1}^n |\alpha_i|$, we derive the limiting behavior of the second term of (45) by dividing $\mathcal{D}_n$ into two sub-regions:

$$E_*\sum_{j=1}^n \sup_{x\in\mathcal{D}_n} |h_j''(x)| \cdot 2n|R_j^2(\hat{\sigma}_\varepsilon^2 - \sigma_\varepsilon^2)^2| \le E_*\sum_{j=1}^n \sup_{x>\sigma_\varepsilon^2/2} |h_j''(x)| \cdot 2nR_j^2(\hat{\sigma}_\varepsilon^2 - \sigma_\varepsilon^2)^2$$
$$+ E_*\sum_{j=1}^n \sup_{x\le\sigma_\varepsilon^2/2} |h_j''(x)| \cdot 2nR_j^2(\hat{\sigma}_\varepsilon^2 - \sigma_\varepsilon^2)^2 1_{\left\{\frac{\sigma_\varepsilon^2}{2} \le |\hat{\sigma}_\varepsilon^2 - \sigma_\varepsilon^2| \le \sigma_\varepsilon^2\right\}}.$$

By Lemma B.3 and (40), for $j \in \{1, 2, \ldots, n\}$,

$$2E_*R_j^2(\hat{\sigma}_\varepsilon^2 - \sigma_\varepsilon^2)^2 \le E_*R_j^4 + E_*(\hat{\sigma}_\varepsilon^2 - \sigma_\varepsilon^2)^4 \lesssim \lambda_j^2(\theta^*) + n^{-2}.$$

Then, using (43),

$$E_*\sum_{j=1}^n \sup_{x>\sigma_\varepsilon^2/2} |h_j''(x)|2nR_j^2(\hat{\sigma}_\varepsilon^2 - \sigma_\varepsilon^2)^2 \lesssim \sum_{j=1}^n \frac{1}{n} + \frac{1}{n}\sum_{j=1}^n \frac{1}{n^2\lambda_j^2(\theta)} \overset{(37)}{=} O(1).$$

Noted that when $\frac{\sigma_\varepsilon^2}{2} \le |\hat{\sigma}_\varepsilon^2 - \sigma_\varepsilon^2| \le \sigma_\varepsilon^2$, we have $(\hat{\sigma}_\varepsilon^2 - \sigma_\varepsilon^2)^2 \le \sigma_\varepsilon^4$ and $\sum_{j=1}^n R_j^2 = 2n\hat{\sigma}_\varepsilon^2 \le 4n\sigma_\varepsilon^2$. Then,

$$E_*\sum_{j=1}^n \sup_{x\le\sigma_\varepsilon^2/2} |h_j''(x)| \cdot 2nR_j^2(\hat{\sigma}_\varepsilon^2 - \sigma_\varepsilon^2)^2 1_{\left\{\frac{\sigma_\varepsilon^2}{2} \le |\hat{\sigma}_\varepsilon^2 - \sigma_\varepsilon^2| \le \sigma_\varepsilon^2\right\}}$$
$$\overset{(44)}{\le} E_*\sum_{j=1}^n n^2 \cdot 2nR_j^2(\hat{\sigma}_\varepsilon^2 - \sigma_\varepsilon^2)^2 1_{\left\{\frac{\sigma_\varepsilon^2}{2} \le |\hat{\sigma}_\varepsilon^2 - \sigma_\varepsilon^2| \le \sigma_\varepsilon^2\right\}}$$
$$\le 8n^4\sigma_\varepsilon^6 P_*\left(|\hat{\sigma}_\varepsilon^2 - \sigma_\varepsilon^2| \ge \frac{\sigma_\varepsilon^2}{2}\right) \overset{\text{Lemma B.3}}{=} O(1).$$

Thus, $H_n = O_{P_*}(1)$. Because equations (37), (42) and (44) all hold uniformly for $\theta \in \Theta'$, we have $\sup_{\theta\in\Theta'} G_n = O_{P_*}(1)$ and $\sup_{\theta\in\Theta'} H_n = O_{P_*}(1)$. □

The following lemma states a property of the jump component that will be used in Lemma B.7 below. The proof is classical and hence, it is omitted.

Lemma B.6. *Let g and h be known deterministic functions such that the expectation below is finite. Then, for any $a, b, c, d \in \{1, 2, \ldots, n\}$, $a < b < c < d$,*

$$E_* \left[g(\Delta_a J, \Delta_b J, \Delta_c J, \Delta_d J) h([J]) \right] = E_* \left[g(\Delta_1 J, \Delta_2 J, \Delta_3 J, \Delta_4 J) h([J]) \right].$$

The following result will be needed in Lemmas B.8 and B.9.

Lemma B.7. *Recalling that $\theta^\dagger = \theta^* + [J]$ and $\lambda_j(\theta) = \frac{\theta}{n} + 2\sigma_\varepsilon^2 \left(1 - \cos \frac{j\pi}{n+1}\right)$, for $p = 2, 3$, under the assumptions (N), (JD)-1, and (JF), we have*

$$\left| \sum_{j=1}^n \frac{n\lambda_j(\theta^\dagger) - nR_j^2}{n^p \lambda_j^p(\theta^\dagger)} \right| = O_{P_*}(n^{1/4}).$$

Proof. Denote $E_J(\cdot) = E_*(\cdot|J)$ and recall (11). Our first step is to take the conditional expectation of the expanded square given J:

$$\begin{aligned} E_J \left(\sum_{j=1}^n \frac{n\lambda_j(\theta^\dagger) - nR_j^2}{n^p \lambda_j^p(\theta^\dagger)} \right)^2 &= \sum_{j=1}^n \frac{E_J \left(n\lambda_j(\theta^\dagger) - nR_j^2 \right)^2}{n^{2p} \lambda_j^{2p}(\theta^\dagger)} \\ &\quad + E_J \sum_{j \neq k} \frac{n\lambda_j(\theta^\dagger) - nR_j^2}{n^p \lambda_j^p(\theta^\dagger)} \frac{n\lambda_k(\theta^\dagger) - nR_k^2}{n^p \lambda_k^p(\theta^\dagger)}. \end{aligned} \tag{46}$$

We compare (46) with the following equation:

$$\begin{aligned} \left(\sum_{j=1}^n \frac{n\lambda_j(\theta^\dagger) - nE_J R_j^2}{n^p \lambda_j^p(\theta^\dagger)} \right)^2 &= \sum_{j=1}^n \frac{\left(n\lambda_j(\theta^\dagger) - nE_J R_j^2 \right)^2}{n^{2p} \lambda_j^{2p}(\theta^\dagger)} \\ &\quad + \sum_{j \neq k} \frac{n\lambda_j(\theta^\dagger) - nE_J R_j^2}{n^p \lambda_j^p(\theta^\dagger)} \frac{n\lambda_k(\theta^\dagger) - nE_J R_k^2}{n^p \lambda_k^p(\theta^\dagger)}. \end{aligned} \tag{47}$$

By the mutually independence of R_js, the second term of the right-hand side of (46) is equal to the second term of (47). Then, the absolute value of the difference between the left-hand sides of (46) and (47) is

$$\left| \sum_{j=1}^n \frac{n^2 E_J R_j^4 - n^2 \left(E_J R_j^2 \right)^2}{n^{2p} \lambda_j^{2p}(\theta^\dagger)} \right| \overset{(39)-(40)}{\lesssim} \sum_{j=1}^n \frac{n^2(\lambda_j^2(\theta) + n^{-2}) + n^2 \left(\lambda_j(\theta) + n^{-1} \right)^2}{n^{2p} \lambda_j^{2p}(\theta^\dagger)},$$

whose square root is $O(n^{1/4})$ because of (37). Then, to prove the result, it suffices to show

$$B_n := \left| \sum_{j=1}^n \frac{n\lambda_j(\theta^\dagger) - nE_J R_j^2}{n^p \lambda_j^p(\theta^\dagger)} \right| = O_{P_*}(n^{1/4}).$$

By expanding $E_J R_j^2$, note that

$$B_n \le \left| \sum_{j=1}^{n} \frac{[J] - \sum_{i=1}^{n} (\Delta_i J)^2}{n^p \lambda_j^p(\theta^\dagger)} \right| + \left| \sum_{i=1}^{n} (\Delta_i J)^2 \sum_{j=1}^{n} \frac{1 - np_{ij}^2}{n^p \lambda_j^p(\theta^\dagger)} \right| + \left| \sum_{j=1}^{n} \frac{\frac{\mu^2}{n}}{n^p \lambda_j^p(\theta^\dagger)} \right| + \left| \sum_{j=1}^{n} \frac{n \sum_{i \neq k} p_{ij} p_{kj} \Delta_i J \Delta_k J}{n^p \lambda_j^p(\theta^\dagger)} \right| + \left| \sum_{j=1}^{n} \frac{2\mu \sum_{i=1}^{n} p_{ij} \Delta_i J}{n^p \lambda_j^p(\theta^\dagger)} \right|. \tag{48}$$

By Lemma A.1, the first term is such that:

$$\left| \sum_{j=1}^{n} \frac{[J] - \sum_{i=1}^{n} (\Delta_i J)^2}{n^p \lambda_j^p(\theta^\dagger)} \right| \le O_{P_*}(n^{-1/2}) \sum_{j=1}^{n} \frac{1}{n^p \lambda_j^p(\theta^*)} \overset{(37)}{=} O_{P_*}(1).$$

Similarly, for the third term of (48) can be shown to be $O_{P_*}(n^{-1/2})$. For the fifth term of (48), by (35) and $|p_{ij}| \le \sqrt{2(n+1)^{-1}}$,

$$E_* \left| \sum_{j=1}^{n} \frac{2\mu \sum_{i=1}^{n} p_{ij} \Delta_i J}{n^p \lambda_j^p(\theta^\dagger)} \right| \le E_* \sum_{j=1}^{n} \frac{2\mu \sum_{i=1}^{n} |p_{ij}| \, |\Delta_i J|}{n^p \lambda_j^p(\theta^\dagger)} \lesssim \sum_{j=1}^{n} \frac{\sum_{i=1}^{n} \sqrt{2(n+1)^{-1}} E_* |\Delta_j J|}{n^p \lambda_j^p(\theta^*)} \lesssim \sum_{j=1}^{n} \frac{n^{-1/2}}{n^p \lambda_j^p(\theta^*)},$$

which is bounded by (37). For the second term of (48), note that

$$\left(\sum_{i=1}^{n} (\Delta_i J)^2 \sum_{j=1}^{n} \frac{1 - np_{ij}^2}{n^p \lambda_j^p(\theta^\dagger)} \right)^2 = \sum_{i=1}^{n} (\Delta_i J)^4 \left(\sum_{j=1}^{n} \frac{1 - np_{ij}^2}{n^p \lambda_j^p(\theta^\dagger)} \right)^2 + \sum_{i \neq k} (\Delta_i J)^2 (\Delta_k J)^2 \left(\sum_{j=1}^{n} \frac{1 - np_{ij}^2}{n^p \lambda_j^p(\theta^\dagger)} \right) \left(\sum_{j=1}^{n} \frac{1 - np_{kj}^2}{n^p \lambda_j^p(\theta^\dagger)} \right). \tag{49}$$

Next, applying Lemma B.6 with $g(x, y, u, v) = x^4$ and $h(x) = \left(\sum_{j=1}^{n} \frac{1 - np_{ij}^2}{n^p \lambda_j^p(\theta^* + x)} \right)^2$, the expectation of the first term of (49) is

$$E_* \left[\sum_{i=1}^{n} (\Delta_i J)^4 \left(\sum_{j=1}^{n} \frac{1 - np_{ij}^2}{n^p \lambda_j^p(\theta^\dagger)} \right)^2 \right] = E_* \left[(\Delta_1 J)^4 \sum_{i=1}^{n} \left(\sum_{j=1}^{n} \frac{1 - np_{ij}^2}{n^p \lambda_j^p(\theta^\dagger)} \right)^2 \right].$$

By (35), to achieve the convergence rate $n^{1/2}$, we need to prove

$$\operatorname*{ess\,sup}_{\omega \in \Omega} \sum_{i=1}^{n} \left(\sum_{j=1}^{n} \frac{1 - np_{ij}^2}{n^p \lambda_j^p(\theta^\dagger)} \right)^2 = O(n^{3/2}). \tag{50}$$

In fact, we can expand the square of the left-hand side, and then interchange the summations. The expression $\sum_{i=1}^n \left(\sum_{j=1}^n \frac{1-np_{ij}^2}{n^p\lambda_j^p(\theta^\dagger)}\right)^2$ can be written as

$$\sum_{i=1}^n\sum_{j=1}^n \frac{1-2np_{ij}^2+n^2p_{ij}^4}{n^{2p}\lambda_j^{2p}(\theta^\dagger)} + \sum_{i=1}^n\sum_{j\neq k} \frac{1-np_{ij}^2-np_{ik}^2+n^2p_{ij}^2p_{ik}^2}{n^{2p}\lambda_j^p(\theta^\dagger)\lambda_k^p(\theta^\dagger)}$$
$$= \sum_{j=1}^n \frac{\sum_{i=1}^n(1-2np_{ij}^2+n^2p_{ij}^4)}{n^{2p}\lambda_j^{2p}(\theta^\dagger)} + \sum_{j\neq k} \frac{\sum_{i=1}^n(1-np_{ij}^2-np_{ik}^2+n^2p_{ij}^2p_{ik}^2)}{n^{2p}\lambda_j^p(\theta^\dagger)\lambda_k^p(\theta^\dagger)}.$$

By the orthogonality of matrix P_n, we have $\sum_{i=1}^n p_{ij}^2 = 1$, $\sum_{i=1}^n p_{ij}^4 = O(n^{-1})$, and $\sum_{i=1}^n p_{ij}^2p_{ik}^2 = \frac{1}{n+1}$ (see Appendix D in [44] for the detailed derivations). Then, since $\theta^* \le \theta^\dagger$,

$$\sum_{j=1}^n \frac{\left|\sum_{i=1}^n(1-2np_{ij}^2+n^2p_{ij}^4)\right|}{n^{2p}\lambda_j^{2p}(\theta^\dagger)} \le \sum_{j=1}^n \frac{n^2\sum_{i=1}^n p_{ij}^4}{n^{2p}\lambda_j^{2p}(\theta^*)} \lesssim \sum_{j=1}^n \frac{n}{n^{2p}\lambda_j^{2p}(\theta^*)} \overset{(37)}{=} O(n^{\frac{3}{2}}),$$
$$\sum_{j\neq k} \frac{\left|\sum_{i=1}^n(1-np_{ij}^2-np_{ik}^2+n^2p_{ij}^2p_{ik}^2)\right|}{n^{2p}\lambda_j^p(\theta^\dagger)\lambda_k^p(\theta^\dagger)} \le \Big(\sum_{j=1}^n \frac{1}{n^p\lambda_j^p(\theta^*)}\Big)^2 \overset{(37)}{=} O(n).$$

These imply (50), and thus the first term of (49) is $O_{P_*}(n^{1/2})$.

For the second component of (49), applying Lemma B.6 with $g(x,y,u,v) = x^2y^2$ and $h(x) = \left(\sum_{j=1}^n \frac{1-np_{ij}^2}{n^p\lambda_j^p(\theta+x)}\right)\left(\sum_{j=1}^n \frac{1-np_{kj}^2}{n^p\lambda_j^p(\theta+x)}\right)$, the expectation of (49) can be written as

$$E_*\Big[\sum_{i\neq k}(\Delta_iJ)^2(\Delta_jJ)^2\Big(\sum_{j=1}^n \frac{1-np_{ij}^2}{n^p\lambda_j^p(\theta^\dagger)}\Big)\Big(\sum_{j=1}^n \frac{1-np_{kj}^2}{n^p\lambda_j^p(\theta^\dagger)}\Big)\Big]$$
$$= E_*\Big\{(\Delta_1J)^2(\Delta_2J)^2\Big[\Big(\sum_{i=1}^n\sum_{j=1}^n \frac{1-np_{ij}^2}{n^p\lambda_j^p(\theta^\dagger)}\Big)^2 - \sum_{i=1}^n\Big(\sum_{j=1}^n \frac{1-np_{ij}^2}{n^p\lambda_j^p(\theta^\dagger)}\Big)^2\Big]\Big\}$$

By (32), i.e. $\sum_{i=1}^n p_{ij}^2 = 1$, $\sum_{i=1}^n\sum_{j=1}^n(1-np_{ij}^2)/n^p\lambda_j^p(\theta^\dagger) = 0$, combined with (50) and (35), the second component (49) is $O_P(n^{1/2})$. This finishes the proof to bound the second term of (48). To analyze the fourth term of (48), consider

$$\Big(\sum_{j=1}^n \frac{n\sum_{i\neq k}p_{ij}p_{kj}\Delta_iJ\Delta_kJ}{n^p\lambda_j^p(\theta^\dagger)}\Big)^2 = \sum_{j=1}^n \frac{n^2\Sigma^2(j)}{n^{2p}\lambda_j^{2p}(\theta^\dagger)} + \sum_{s\neq t} \frac{n^2\Sigma(s)\Sigma(t)}{n^{2p}\lambda_s^p(\theta^\dagger)\lambda_t^p(\theta^\dagger)}, \quad (51)$$

where $\Sigma(j) = \sum_{i\neq k}p_{ij}p_{kj}\Delta_iJ\Delta_kJ$. We will show that the two components of the right-hand side of (51) are both $O_{P_*}(n^{1/2})$. For the the first, note that, since $E_*\Delta_iJ = 0$, the expectation of the numerator is bounded since

$$E_*n^2\Sigma^2(j) = n^2\sum_{i\neq k}p_{ij}^2p_{kj}^2E_*[(\Delta_iJ)^2]E_*[(\Delta_kJ)^2] \overset{(35)}{\lesssim} n^2\cdot n^{-1}n^{-1}\left(\sum_{i=1}^n p_{ij}^2\right)^2.$$

Then, for the first component of (51), we have

$$E_*\sum_{j=1}^{n}\frac{n^2\Sigma^2(j)}{n^{2p}\lambda_j^{2p}(\theta^\dagger)}\le E_*\sum_{j=1}^{n}\frac{n^2\Sigma^2(j)}{n^{2p}\lambda_j^{2p}(\theta^*)}=\sum_{j=1}^{n}\frac{n^2E_*\Sigma^2(j)}{n^{2p}\lambda_j^{2p}(\theta^*)}\lesssim\sum_{j=1}^{n}\frac{1}{n^{2p}\lambda_j^{2p}(\theta^*)}.$$

The last term is $O(n^{1/2})$ by (37). For the second component of (51), we first note that expansion $\Sigma(s)\Sigma(t)$ results in three different types of terms:

$$\begin{aligned}
&p_{is}p_{ks}p_{ut}p_{vt}\Delta_i J\Delta_k J\Delta_u J\Delta_v J, && \text{for } i<k<u<v,\\
&p_{is}p_{ks}p_{it}p_{ut}(\Delta_i J)^2\Delta_k J\Delta_u J, && \text{for } i<k<u, v=i,\\
&p_{is}p_{ks}p_{it}p_{kt}(\Delta_i J)^2(\Delta_k J)^2, && \text{for } i<k, u=i, v=k.
\end{aligned}$$

For these different terms, we apply Lemma B.6 with corresponding g and h functions. Then, the expectation of the second component of (51) can be upper bounded, in absolute value, by the summation over s and t from 1 to n ($s\neq t$) of the following three types of terms:

$$\begin{aligned}
&E_*|\Delta_1 J\Delta_2 J\Delta_3 J\Delta_4 J|\frac{\left|\sum_{i\neq k\neq u\neq v}p_{is}p_{ks}p_{ut}p_{vt}\right|}{n^{2p}\lambda_s^p(\theta^*)\lambda_t^p(\theta^*)},\\
&2E_*|(\Delta_1 J)^2\Delta_2 J\Delta_3 J|\frac{\left|\sum_{i\neq k\neq u}p_{is}p_{it}p_{ks}p_{ut}\right|}{n^{2p}\lambda_s^p(\theta^*)\lambda_t^p(\theta^*)}\overset{(35)}{\lesssim}\frac{n^{-3}}{n^{2p}\lambda_s^p(\theta^*)\lambda_t^p(\theta^*)}, \qquad (52)\\
&E_*|(\Delta_1 J)^2(\Delta_2 J)^2|\frac{\left|\sum_{i\neq k}p_{is}p_{it}p_{ks}p_{kt}\right|}{n^{2p}\lambda_s^p(\theta^*)\lambda_t^p(\theta^*)}\overset{(35)}{\lesssim}\frac{n^{-3}}{n^{2p}\lambda_s^p(\theta^*)\lambda_t^p(\theta^*)}. \qquad (53)
\end{aligned}$$

where we used that $\left|\sum_{i\neq k\neq u}p_{is}p_{it}p_{ks}p_{ut}\right|=O(1)$ and $\left|\sum_{i\neq k}p_{is}p_{it}p_{ks}p_{kt}\right|=\frac{1}{n+1}$ (by Lemma B.1). For (52) and (53), we take summation over s,t from 1 to n, and then multiply it by n^2. The resulting expressions are both bounded by (37). Combined with (35), to show that the second term of (51) is $O(n^{1/2})$, it suffices to show that

$$\sum_{s=1}^{n}\sum_{t=1}^{n}\frac{\left|\sum_{i\neq k\neq u\neq v}p_{is}p_{ks}p_{ut}p_{vt}\right|}{n^{2p}\lambda_s^p(\theta^*)\lambda_t^p(\theta^*)}=O(n^{1/2+4-2})=O(n^{5/2}). \qquad (54)$$

Indeed, let us start by noting that, since $\sum_{i=1}^{n}p_{ij}^2=1$,

$$\begin{aligned}
\left(\sum_i p_{is}\right)^2\left(\sum_u p_{ut}\right)^2&=\sum_{u\neq v}(p_{ut}p_{vt}+p_{us}p_{vs})+\sum_{u\neq v}p_{us}p_{vs}p_{ut}p_{vt}+\\
&\quad+4\sum_{i\neq u\neq v}p_{is}p_{us}p_{ut}p_{vt}+\sum_{i\neq k\neq u\neq v}p_{is}p_{ks}p_{ut}p_{vt}+1.
\end{aligned}$$

The second term and the third term are both $O(1)$. Then,

$$\left|\sum_{i\neq k\neq u\neq v}p_{is}p_{ks}p_{ut}p_{vt}\right|\lesssim\left(\sum_i p_{is}\right)^2\left(\sum_u p_{ut}\right)^2+1+2\left|\sum_{u\neq v}p_{ut}p_{vt}+\sum_{u\neq v}p_{us}p_{vs}\right|.$$

By (34), and $\lambda_j(\theta) \geq \theta/n$, we have

$$\sum_{s=1}^{n} \frac{\left(\sum_i p_{is}\right)^2}{n^p \lambda_s^p(\theta^*)} \lesssim \sum_{s=1}^{n} \frac{n/s^2}{n^p(\theta/n)^p} \lesssim n + \sum_{s=2}^{n} \frac{n}{s^2} \leq n + \int_1^n \frac{n}{s^2}\, ds = O(n).$$

Then,

$$\sum_{s=1}^{n}\sum_{t=1}^{n} \frac{\left(\sum_i p_{is}\right)^2 \left(\sum_u p_{ut}\right)^2}{n^{2p}\lambda_s^p(\theta^*)\lambda_t^p(\theta^*)} = \left[\sum_{s=1}^{n} \frac{\left(\sum_i p_{is}\right)^2}{n^p \lambda_s^p(\theta^*)}\right]^2 = O(n^2).$$

Since $|p_{ut}| = O(n^{-1/2})$,

$$\sum_{s=1}^{n}\sum_{t=1}^{n} \frac{\left|\sum_{u\neq v} p_{ut}p_{vt}\right|}{n^{2p}\lambda_s^p(\theta^*)\lambda_t^p(\theta^*)} \lesssim \sum_{s=1}^{n}\sum_{t=1}^{n} \frac{n^2 \cdot \frac{1}{n}}{n^{2p}\lambda_s^p(\theta^*)\lambda_t^p(\theta^*)} = n\left(\sum_{s=1}^{n} \frac{1}{n^p \lambda_s^p(\theta^*)}\right)^2,$$

which is $O(n^2)$ by (37). Then, we proved (54), and thus, the third term of (48) is $O_{P_*}(n^{1/2})$. This completes the proof of the Lemma. □

In the following two lemmas, we establish some needed asymptotic properties of the misspecified likelihood function under all the assumptions in § 2.

Lemma B.8. *Let* $\Gamma(\theta,\zeta) := \frac{(\sqrt{\theta}-\sqrt{\zeta})^2}{2\sqrt{\zeta}\sigma_\varepsilon}$, *recalling that* σ_ε *is the standard deviation of the noise. Then,*

$$\sup_{\zeta\in\Theta'} |n^{-\frac{1}{2}}(\tilde{l}_n(\theta^\dagger) - \tilde{l}_n(\zeta)) - \Gamma(\theta^\dagger,\zeta)| = o_{P_*}(1), \quad \left|n^{-\frac{1}{4}}\dot{\tilde{l}}_n(\theta^\dagger)\right| = O_{P_*}(1).$$

Proof. For any $\zeta \in \Theta'$, there exist a $\zeta^* \in \Theta$ such that $\zeta = \zeta^* + [J]$ and, noting that $\theta^\dagger = \theta^* + [J]$, we have $\zeta - \theta^\dagger = \zeta^* - \theta^*$. Hence, by the boundedness of Θ, for the first statement, it is enough to prove that, for any fixed $\delta > 0$,

$$\sup_{\zeta\in\Theta':|\zeta-\theta^\dagger|<\delta} |n^{-\frac{1}{2}}(\tilde{l}_n(\theta^\dagger) - \tilde{l}_n(\zeta)) - \Gamma(\theta^\dagger,\zeta)| \xrightarrow{P_*} 0.$$

By Lemma B.5,

$$\sup_{\theta,\zeta\in\Theta':\theta\neq\zeta} \left|\frac{l_n(\theta)-l_n(\zeta)}{\theta-\zeta} - \frac{\tilde{l}_n(\theta)-\tilde{l}_n(\zeta)}{\theta-\zeta}\right| \leq \sup_{\theta\in\Theta'} \left|\frac{dl_n(\theta)}{d\theta} - \frac{d\tilde{l}_n(\theta)}{d\theta}\right| = O_{P_*}(1). \tag{55}$$

Then, for every $\zeta \in \Theta'$ such that $|\zeta - \theta^\dagger| < \delta$,

$$n^{-\frac{1}{2}}(\tilde{l}_n(\zeta) - \tilde{l}_n(\theta^\dagger)) = n^{-\frac{1}{2}}(l_n(\zeta) - l_n(\theta^\dagger)) + O_{P_*}(n^{-\frac{1}{2}}).$$

Hence, it suffices to show the statement for $l_n(\zeta) - l_n(\theta^\dagger)$. Next, recalling (38), the difference of the log likelihoods can be rewritten as

$$n^{-\frac{1}{2}}(l_n(\theta^\dagger) - l_n(\zeta)) = n^{-\frac{1}{2}} \sum_{j=1}^{n} \left[\frac{\lambda_j(\theta^\dagger)}{\lambda_j(\zeta)} - 1 - \log\frac{\lambda_j(\theta^\dagger)}{\lambda_j(\zeta)}\right]$$

$$+ n^{-\frac{1}{2}} \sum_{j=1}^{n} \frac{nR_j^2 - n\lambda_j(\theta^\dagger)}{n^2\lambda_j^2(\theta^\dagger)} \frac{(\theta^\dagger - \zeta)\lambda_j(\theta^\dagger)}{\lambda_j(\zeta)}.$$

The second term is $o_{P_*}(1)$ uniformly for ζ such that $|\zeta - \theta^\dagger| < \delta$, which follows from Lemma B.7 with $p = 2$, and

$$\frac{\lambda_j(\theta^\dagger)}{\lambda_j(\zeta)} = \frac{\theta^\dagger + 2n\sigma_\varepsilon^2\left(1 - \cos\frac{j\pi}{n+1}\right)}{\zeta + 2n\sigma_\varepsilon^2\left(1 - \cos\frac{j\pi}{n+1}\right)} \leq \frac{\theta^\dagger + 2n\sigma_\varepsilon^2\left(1 - \cos\frac{j\pi}{n+1}\right)}{\delta_0 + 2n\sigma_\varepsilon^2\left(1 - \cos\frac{j\pi}{n+1}\right)} \leq \frac{\theta^\dagger}{\delta_0}, \quad (56)$$

where recall that $\delta_0 > 0$ is such that $\Theta' \subset (\delta_0, \infty)$. Then, it remains to prove the following convergence:

$$\sup_{\zeta\in\Theta':|\zeta-\theta^\dagger|\leq\delta} \left| n^{-\frac{1}{2}} \sum_{j=1}^{n} \left[\frac{\lambda_j(\theta^\dagger)}{\lambda_j(\zeta)} - 1 - \log\frac{\lambda_j(\theta^\dagger)}{\lambda_j(\zeta)} \right] - \Gamma(\theta^\dagger, \zeta) \right| \xrightarrow{P_*} 0. \quad (57)$$

Based on (4.5) and (6.6) in [18] with $a = \theta^\dagger$ and $b = \zeta$, the uniform convergence holds almost surely for $\zeta \in [1/C, C]$, where C is some constant. The result can be generalized to $\zeta \in [\delta_0, \theta^\dagger + \delta]$ when we use convergence in probability instead of almost surely convergence, because $\theta^\dagger$ can be treated as a constant under measure P_* (see more details in [44]). Hence, the first asymptotics of the Lemma can be demonstrated. For the second, again, by Lemma B.5,

$$\left|n^{-\frac{1}{4}}\dot{\tilde{l}}_n(\theta^\dagger)\right| \leq \left|n^{-\frac{1}{4}}(\dot{\tilde{l}}_n(\theta^\dagger) - \dot{l}_n(\theta^\dagger))\right| + \left|n^{-\frac{1}{4}}\dot{l}_n(\theta^\dagger)\right| = \left|n^{-\frac{1}{4}}\dot{l}_n(\theta^\dagger)\right| + o_{P_*}(1).$$

Then, it suffices to prove the boundedness of $\left|n^{-\frac{1}{4}}\dot{l}_n(\theta^\dagger)\right|$, which follows directly from Lemma B.7 with $p = 2$ since $\dot{l}_n(\theta^\dagger) = -2^{-1}\sum_{i=1}^{n}(\lambda_j(\theta^\dagger) - R_j^2)/n\lambda_j^2(\theta^\dagger)$. This completes the proof. □

Lemma B.9. *Let $I(\theta) = 1/(8\theta^{3/2}\sigma_\varepsilon)$. Then, under all the assumptions in § 2,*

$$n^{-\frac{1}{2}}\ddot{\tilde{l}}_n(\theta^\dagger) + I(\theta^\dagger) \xrightarrow{P_*} 0,$$

and, for any sequence of nonnegative random variables $\{\eta_n\}$ such that $\eta_n \xrightarrow{P_*} 0$,

$$\sup_{\zeta,\zeta'\in\Theta':|\zeta-\zeta'|\leq\eta_n} n^{-\frac{1}{2}} \left|\ddot{\tilde{l}}_n(\zeta) - \ddot{\tilde{l}}_n(\zeta')\right| \xrightarrow{P_*} 0, \text{ as } n \to \infty.$$

Proof. By Lemma B.5, it suffices to prove the statements only for l_n. For the first statement, we split $\ddot{l}_n$ into two components:

$$n^{-\frac{1}{2}}\ddot{l}_n(\theta^\dagger) = -2n^{-\frac{1}{2}}\sum_{j=1}^{n}\frac{nR_j^2 - n\lambda_j(\theta^\dagger)}{n^3\lambda_j^3(\theta^\dagger)} - n^{-\frac{1}{2}}\sum_{j=1}^{n}\frac{1}{n^2\lambda_j^2(\theta^\dagger)}.$$

The first component is $O_{P_*}(n^{-1/4})$ by Lemma B.7 with $p = 3$. The second component converges to $-I(\theta)$, which results from (3.1)-(3.2) in [17]. This concludes the proof of the first assertion. For the second statement, without loss of generality, let $\zeta \leq \zeta'$ and note that

$$\left|\frac{1}{\lambda_j^2(\zeta)} - \frac{1}{\lambda_j^2(\zeta')}\right| = |\zeta' - \zeta| \left(\frac{1}{n\lambda_j(\zeta)\lambda_j^2(\zeta')} + \frac{1}{n\lambda_j^2(\zeta)\lambda_j(\zeta')}\right) \overset{\zeta \leq \zeta'}{\leq} 2|\zeta' - \zeta| \frac{1}{n\lambda_j^3(\zeta)},$$

$$\left|\frac{1}{\lambda_j^3(\zeta)} - \frac{1}{\lambda_j^3(\zeta')}\right| \leq 3|\zeta' - \zeta| \frac{1}{n\lambda_j^4(\zeta)}.$$

Recalling that $\ddot{\tilde{l}}_n(\zeta) = \sum_{i=1}^n (2n^2\lambda_j^2(\zeta))^{-1} - \sum_{i=1}^n R_j^2/(n^2\lambda_j^3(\zeta))$, we have

$$\left|\ddot{\tilde{l}}_n(\zeta) - \ddot{\tilde{l}}_n(\zeta')\right| \leq 3|\zeta' - \zeta| \left(\sum_{j=1}^n \frac{1}{n^3\lambda_j^3(\zeta)} + \sum_{j=1}^n \frac{nR_j^2}{n^4\lambda_j^4(\zeta)}\right).$$

Then, because $\Theta' \subset (\delta_0, \infty)$,

$$\sup_{|\zeta - \zeta'| \leq \eta_n} n^{-\frac{1}{2}} \left|\ddot{\tilde{l}}_n(\zeta) - \ddot{\tilde{l}}_n(\zeta')\right| \leq \eta_n n^{-\frac{1}{2}} \left(\sum_{j=1}^n \frac{1}{n^3\lambda_j^3(\delta_0)} + \sum_{j=1}^n \frac{nR_j^2}{n^4\lambda_j^4(\delta_0)}\right). \quad (58)$$

By (39),

$$E_* \left[n^{-\frac{1}{2}} \sum_{j=1}^n \frac{nR_j^2}{n^4\lambda_j^4(\delta_0)}\right] \lesssim n^{-\frac{1}{2}} \sum_{j=1}^n \left\{\frac{1}{n^3\lambda_j^3(\delta_0)} + \frac{1}{n^4\lambda_j^4(\delta_0)}\right\} \overset{(37)}{=} O(1).$$

Then,

$$n^{-\frac{1}{2}} \sum_{j=1}^n \frac{nR_j^2}{n^4\lambda_j^4(\delta_0)} = O_{P_*}(1). \quad (59)$$

Applying Slutsky's theorem, (59) and (37) to (58), the lemma can be proved. □

B.3. MLE and its convergence rate

In this section, we prove that

$$\left|\tilde{\theta}_n - \theta^\dagger\right| = O_{P_*}(n^{-1/4}), \quad (60)$$

which, as explained at the beginning of Appendix B, implies the condition (7) of Theorem 4.1.

Since $\tilde{\theta}_n$ is the maximum of the misspecified log likelihood function $\tilde{l}_n$ defined in (12), we then have

$$-\dot{\tilde{l}}_n(\theta^\dagger) = (\tilde{\theta}_n - \theta^\dagger)\ddot{\tilde{l}}_n(\theta^\dagger) + (\tilde{\theta}_n - \theta^\dagger) \int_0^1 \left[\ddot{\tilde{l}}_n(\theta^\dagger + w(\tilde{\theta}_n - \theta^\dagger)) - \ddot{\tilde{l}}_n(\theta^\dagger)\right] dw.$$

Rearranging the terms, we obtain the following equation:

$$n^{\frac{1}{4}}\left|\tilde{\theta}_n-\theta^\dagger\right|=\frac{|n^{-\frac{1}{4}}\dot{\tilde{l}}_n(\theta^\dagger)|}{\left|n^{-\frac{1}{2}}\ddot{\tilde{l}}_n(\theta^\dagger)+\int_0^1 n^{-\frac{1}{2}}\left[\ddot{\tilde{l}}_n(\theta^\dagger+w(\tilde{\theta}_n-\theta^\dagger))-\ddot{\tilde{l}}_n(\theta^\dagger)\right]\,dw\right|}. \quad (61)$$

Next, we apply Theorem 1 of [45] with

$$Q_n(x)=-n^{-1/2}\left(\tilde{l}_n(\theta^\dagger)-\tilde{l}_n(x+[J])\right),\ \bar{Q}_n(x)=-\Gamma\left(\theta^\dagger,x+[J]\right),$$

and $\hat{\theta}_n=\tilde{\theta}_n-[J]$, to conclude that $\tilde{\theta}_n-[J]$ is a consistent estimator of θ^*. This implies $\tilde{\theta}_n-\theta^\dagger\xrightarrow{P_*}0$. Indeed, Lemma B.8 and the definition of Γ yield that the conditions are satisfied since the maximum of $-\Gamma(\theta^\dagger,x+[J])$ is 0 when $x=\theta^\dagger-[J]=\theta^*$. The just stated consistency combined with Lemma B.9 (applied with $\eta_n=|\tilde{\theta}_n-\theta^\dagger|$) implies that the denominator of the right-hand side of (61) converges to some constant value. The numerator is $O_{P_*}(1)$ by Lemma B.8. Then, the convergence rate of $\tilde{\theta}_n$ to $\theta^\dagger$ is $O_{P_*}(n^{-\frac{1}{4}})$, as claimed.

B.4. Local Asymptotic Normality (LAN)

The following LAN property is the condition (6) required in Theorem 4.1. For notational simplicity, in this section, we write $\tilde{l}_n(\theta)$ as $\tilde{l}_\theta$.

Theorem B.10. *Recall that we assumed that $\kappa_n\to\kappa^\dagger$ in P_*-probability. Assume that $\kappa^\dagger$ is bounded away from zero and infinity in P_*-probability, and (N), (JD)-1, and (JF) hold true. Then, for every compact set $K\subset\mathbb{R}$, $\tilde{l}$ satisfies*

$$\sup_{h\in K}\left|\frac{1}{\kappa_n}\left(\tilde{l}_{\theta^\dagger+n^{-1/4}h}-\tilde{l}_{\theta^\dagger}\right)-\frac{1}{\kappa^\dagger}hn^{\frac{1}{4}}I(\theta^\dagger)(\tilde{\theta}_n-\theta^\dagger)+\frac{1}{\kappa^\dagger}\frac{1}{2}h^2I(\theta^\dagger)\right|=o_{P_*}(1), \quad (62)$$

where $I(\theta)=\frac{1}{8\theta^{3/2}\sigma_\varepsilon}$.

Proof. Rewrite the the left-hand side expression in (62) as

$$\begin{aligned}&\left|\frac{1}{\kappa_n}\left(\tilde{l}_{\theta^\dagger+n^{-1/4}h}-\tilde{l}_{\theta^\dagger}\right)-\frac{1}{\kappa^\dagger}hn^{\frac{1}{4}}I(\theta^\dagger)(\tilde{\theta}_n-\theta^\dagger)+\frac{1}{\kappa^\dagger}\frac{1}{2}h^2I(\theta^\dagger)\right|\\&\quad\le\left|\left(\frac{1}{\kappa_n}-\frac{1}{\kappa^\dagger}\right)\left(hn^{\frac{1}{4}}I(\theta^\dagger)(\tilde{\theta}_n-\theta^\dagger)-\frac{1}{2}h^2I(\theta^\dagger)\right)\right|\\&\qquad+\frac{1}{\kappa_n}\left|\tilde{l}_{\theta^\dagger+n^{-1/4}h}-\tilde{l}_{\theta^\dagger}-hn^{\frac{1}{4}}I(\theta^\dagger)(\tilde{\theta}_n-\theta^\dagger)+\frac{1}{2}h^2I(\theta^\dagger)\right|.\end{aligned}$$

By (60) (i.e. $\left|\tilde{\theta}_n-\theta^\dagger\right|=O_{P_*}(n^{-1/4})$), the fact that $\kappa_n\to\kappa^\dagger$ in P_*-probability, and Slutsky's Theorem, we only need to prove

$$\sup_{h\in K}\left|\tilde{l}_{\theta^\dagger+n^{-1/4}h}-\tilde{l}_{\theta^\dagger}-hn^{\frac{1}{4}}I(\theta^\dagger)(\tilde{\theta}_n-\theta^\dagger)+\frac{1}{2}h^2I(\theta^\dagger)\right|=o_{P_*}(1).$$

Let us start by writing

$$\tilde{l}_{\theta^\dagger + n^{-1/4}h} - \tilde{l}_{\theta^\dagger} = hn^{-\frac{1}{4}}\dot{\tilde{l}}_{\theta^\dagger} + \frac{1}{2}h^2 n^{-\frac{1}{2}}\ddot{\tilde{l}}_{\theta^\dagger} + r_n. \tag{63}$$

For the first term of (63), since $\dot{\tilde{l}}$ is continuous and differentiable w.r.t. θ, there exists $\ddot{\theta}_n$ lying on the segment which connects $\theta^\dagger$ and $\tilde{\theta}_n$, such that

$$n^{-\frac{1}{4}}\dot{\tilde{l}}_{\theta^\dagger} = n^{-\frac{1}{4}}\dot{\tilde{l}}_{\theta^\dagger} - n^{-\frac{1}{4}}\dot{\tilde{l}}_{\tilde{\theta}_n} = n^{-\frac{1}{2}}\ddot{\tilde{l}}_{\ddot{\theta}_n} \cdot n^{\frac{1}{4}}(\theta^\dagger - \tilde{\theta}_n). \tag{64}$$

Because $\tilde{\theta}_n$ converges to $\theta^\dagger$ in probability and $\ddot{\theta}_n$ lies on the segment joining $\tilde{\theta}_n$ and $\theta^\dagger$, we conclude that $\ddot{\theta}_n$ converges to $\theta^\dagger$ in probability. Then, applying Lemma B.9 with $\zeta = \ddot{\theta}_n$, $\zeta' = \theta^\dagger$, and $\eta_n = |\tilde{\theta}_n - \theta^\dagger|$,

$$n^{-\frac{1}{2}}\ddot{\tilde{l}}_{\ddot{\theta}_n} + I(\theta^\dagger) = \left[n^{-\frac{1}{2}}\ddot{\tilde{l}}_{\ddot{\theta}_n} - n^{-\frac{1}{2}}\ddot{\tilde{l}}_{\theta^\dagger}\right] + \left[n^{-\frac{1}{2}}\ddot{\tilde{l}}_{\theta^\dagger} + I\left(\theta^\dagger\right)\right] = o_{P_*}(1). \tag{65}$$

Thus, combining (64) and (65), the first term of the right-hand side of (63) can be written as,

$$hn^{-\frac{1}{4}}\dot{\tilde{l}}_{\theta^\dagger} = -hn^{\frac{1}{4}} I\left(\theta^\dagger\right)\left(\theta^\dagger - \tilde{\theta}_n\right) + o_{P_*}(1).$$

The second term of (63) converges to $-\frac{1}{2}h^2 I(\theta^\dagger)$ by Lemma B.9. For the reminder term of (63), r_n, note first that by Lemma B.5, for all $\theta \in \Theta'$,

$$\left|\frac{d^3\tilde{l}_\theta}{d\theta^3}\right| \le \left|\frac{d^3\tilde{l}_\theta}{d\theta^3} - \frac{d^3 l_\theta}{d\theta^3}\right| + \left|\frac{d^3 l_\theta}{d\theta^3}\right| = O_{P_*}(1) + \frac{(-1)^k}{2}\sum_{i=1}^{n}\left|\frac{(k-1)!}{n^k\lambda_i^k(\theta)} - \frac{k! n R_i^2}{n^{k+1}\lambda_i^{k+1}(\theta)}\right|.$$

Applying (37) and (39), the last term is $O_{P_*}(\sqrt{n})$. The proof is the same as in (59). We use $E_* n R_j^2 = n\lambda_j(\theta^*) + O(1)$ and cancel out one $n\lambda_j(\theta)$ from the denominator (up to some constant) by (56), and then use (37) to obtain the rate $n^{1/2}$. Then,

$$|r_n| \le \frac{1}{6}h^3 n^{-\frac{3}{4}} \sup_{\theta \in [\theta^\dagger, \theta^\dagger + n^{-1/4}h]}\left|\frac{d^3\tilde{l}_\theta}{d\theta^3}\right| = o_{P_*}(1).$$

Combining all three terms, we can obtain the LAN property of $\tilde{l}$. □

Acknowledgements

The second author's research was supported in part by the NSF Grants: DMS-1561141 and DMS-1613016. The third author's research was supported by NSF award DMS-1811936. The authors are grateful to the Associate Editor and two anonymous referees for their multiple corrections and helpful suggestions that help to significantly improve the original manuscript.

References

[1] Aït-Sahalia, Y., Mykland, P. A., and Zhang, L. (2005). How often to sample a continuous-time process in the presence of market microstructure noise. *The Review of Financial Studies*. **18**(2) 351–416.

[2] Barndorff-Nielsen, O., Hansen, P., Lunde, A. and Shephard, N. (2008). Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise. *Stochastic Processes and their Applications*. **76**(6) 1481–1536. MR2468558

[3] Barndorff-Nielsen, O. E., Mikosch, T. and Resnick, S. (2001). Lévy processes: Theory and applications. Birkhauser. MR1833689

[4] Carr, P. and Geman, H. and Madan, D. and Yor, M. (2002). The fine structure of asset returns: an empirical investigation. *The Journal of Business*. **75**(2) 305–332.

[5] Chen, R. Y. (2018). Inference for volatility functionals of multivariate Itô semimartingales observed with jump and noise. *Working Paper available at arXiv:1810.04725*.

[6] Christopher, S. J. (1999). Bayesian estimation of continuous-time finance models. Working Paper. *Rochester University*. Working Paper.

[7] Christensen, K., Oomen, R. and Podolskij, M. (2010). Realised quantile-based estimation of the integrated variance. *Journal of Econometrics*. **159**(1) 74–98. MR2720845

[8] Clinet, S., and Potiron, Y. (2018). Efficient asymptotic variance reduction when estimating volatility in high frequency data. *Journal of Econometrics*. **206**(1) 103–142. MR3840785

[9] Cont, R. and Tankov, P. (2004). Financial modelling with Jump Processes. *Chapman & Hall/CRC Finance*. Math. Ser., Chapman & Hall/CRC, Boca Raton, FL. MR2042661

[10] Cont, R. and Mancini, C. (2011). Nonparametric tests for pathwise properties of semimartingales. *Bernoulli*. **17**(2) 781–813. MR2787615

[11] Da, R. and Xiu, D. (2017). When moving-average models meet high-frequency data: Uniform inference on volatility. *Working paper available on Dacheng Xiu's website*

[12] de Bruijn, N. G. (1961). Asymptotic methods in analysis. North Holland, Amsterdam. MR0177247

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

[14] Figueroa-López, J. E. (2008). Small-time moment asymptotics for Lévy processes. *Statistics & Probability Letters*. **78**(18) 3355–3365. MR2479503

[15] Figueroa-López, J. E. and Lee, K. (2017). Estimation of a noisy subordinated Brownian Motion via two-scale power variations. *Journal of Statistical Planning and Inference*. **189** 16–37. MR3659847

[16] Figueroa-López, J. E. and Mancini, C. (2019). Optimum thresholding using mean and conditional mean square error. *Journal of Econometrics*. **208**(1) 179–210. MR3906967

[17] Gloter, A. and Jacod, J. (2001). Diffusions with measurement errors. I.

Local asymptotic normality. *ESAIM: Probability and Statistics*. **5** 225–242. MR1875672

[18] Gloter, A. and Jacod, J. (2001). Diffusions with measurement errors. II. Optimal estimators. *ESAIM: Probability and Statistics*. **5** 243–260. MR1875673

[19] Gonçalves, S. and Meddahi, N. (2009). Bootstrapping realized volatility. *Econometrica*. **77** 283–306. MR2477851

[20] Griffin, J. E. (2016). Flexibly Modelling Volatility and Jumps Using Realised and Bi-Power Variation. University of Kent, School of Mathematics, Statistics and Actuarial Science.

[21] Hansen, P. and Lunde, A. (2006). Realized variance and market microstructure noise. *Journal of Business & Economic Statistics*. **24**(2) 127–161. MR2234447

[22] He, C. and Wang, Z. (2014). Efficient estimation of stochastic diffusion models with leverage effects and Lévy jumps. *Journal of Information & Computational Science*. **11**(2) 367–382.

[23] Jacod, J. (2007). Asymptotic properties of power variations of Lévy processes. *ESAIM: Probability and Statistics*. **11** 173–196. MR2320815

[24] Jacod, J. (2008). Asymptotic properties of realized power variations and related functionals of semimartingales. *Stochastic Processes and Their Applications*. **118** 517–559. MR2394762

[25] Jacod, J., Li, Y., Mykland, P., Podolskij, M. and Vetter, M. (2009). Microstructure noise in the continuous case: The pre-averaging approach. *Stochastic Processes and their Applications*. **119**(7) 2249–2276. MR2531091

[26] Jacod, J. and Protter, P. E. (2012). Discretization of processes. *Stochastic Modelling and Applied Probability*. **67**. MR2859096

[27] Jacod, J. and Todorov, V. (2014). Efficient estimation of integrated volatility in presence of infinite variation jumps. *The Annals of Statistics*. **42**(3) 1029–1069. MR3210995

[28] Jasra, A., Stephens, D., Doucet, A., and Tsagaris, T. (2011). Inference for Lévy-driven stochastic volatility models via adaptive sequential Monte Carlo. *Scandinavian Journal Of Statistics*. **38**(1) 1–22. MR2760137

[29] Jiang, W. and Tanner, M. (2008). Gibbs posterior for variable selection in high-dimensional classification and data mining. *The Annals of Statistics*. **36**(10) 2207–2231. MR2458185

[30] Jing, B., Liu, Z. and Kong, X. (2014). On the estimation of integrated volatility with jumps and microstructure noise. *Journal of Business & Economic Statistics*. **32**(3) 457–467. MR3238598

[31] Kleijn, B. J. K. and van der Vaart, A. W. (2012). The Bernstein-von Mises theorem under misspecification. *Electronic Journal of Statistics*. **6** 354–381. MR2988412

[32] Kolassa, J. E. and Kuffner, T. A. (2020). On the validity of the formal Edgeworth expansion for posterior densities. *Ann. Statist.*, to appear. MR4134781

[33] Kong, X., Liu, Z. and Jing, B. (2015). Testing for pure-jump processes

for high-frequency data. *Ann. Statist.* **4**(2) 847–877. MR3325712
[34] Kong, X., Xu, S. and Zhou, W. (2018). Bootstrapping volatility functionals: a local and nonparametric perspective. *Biometrika*. **105**(2) 463–469. MR3804415
[35] Kou, S., Yu, C. and Zhong, H. (2017). Jumps in equity index returns before and during the recent financial crisis: a Bayesian analysis. *Management Science*. **63**(4) 988–1010.
[36] Li, H., Wells, M. and Yu, C. (2008). A Bayesian analysis of return dynamics with Lévy jumps. *The Review of Financial Studies*. **21**(5) 2345–2378.
[37] Mancini, C. (2009). Non-parametric threshold estimation for models with stochastic diffusion coefficient and jumps. *Scandinavian Journal of Statistics*. **36** 270–296. MR2528985
[38] Martin, R., Ouyang, C. and Domagni, F. (2018). 'Purposely misspecified' posterior inference on the volatility of a jump diffusion process. *Statistics & Probability Letters*. **134** 106–113. MR3758588
[39] Masuda, H. (2015). Parametric estimation of Lévy processes. In: Lévy Matters IV, pages 179–286. Springer. MR3364258
[40] Podolskij, M. and Vetter, M. (2009). Bipower-type estimation in a noisy diffusion setting. *Stochastic Processes and their Applications*. **119**(9) 2803–2831. MR2554029
[41] Podolskij, M. and Vetter, M. (2009). Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps. *Bernoulli*. **15**(3) 634–658. MR2555193
[42] Szerszen, P. (2009). Bayesian analysis of stochastic volatility models with Lévy jumps: Application to risk analysis. *Board of Governors of the Federal Reserve System*. 2009-40.
[43] Tankov, P. (2007). Lévy processes in finance and risk management. *Wilmott Magazine*. Sept-Oct, 89–97.
[44] Wang, Q. (2020). Bayesian posterior inference and LAN for Lévy processes under high-frequency sampling. *PhD thesis, Washington University in St. Louis*. MR4094311
[45] Xiu, D. (2010). Quasi-maximum likelihood estimation of volatility with high frequency data. *Journal of Econometrics*. **159**(1) 235–250. MR2720855
[46] Yang, H. and Kanniainen, J. (2017). Jump and volatility dynamics for the S&P 500: evidence for infinite-activity jumps with non-affine volatility dynamics from stock and option markets. *Review of Finance*. **21**(2) 811–844.
[47] Yu, C., Li, H. and Wells, M. (2011). MCMC estimation of Lévy jump models using stock and option prices. *Mathematical Finance*. **21**(3) 383–422. MR2830427
[48] Zhang, L., Mykland, P. and Aït-Sahalia, Y. (2005). A tale of two time scales: determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association*. **100**(472) 1394–1411. MR2236450
[49] Zhang, L. (2006). Efficient estimation of stochastic volatility using

noisy observations: a multi-scale approach. *Bernoulli*. **12**(6) 1019–1043. MR2274854

[50] Zhang, T. (2006). Information-theoretic upper and lower bounds for statistical estimation. *IEEE Transactions on Information Theory*. **52**(4) 1307–1321. MR2241190

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