

# Efficient pseudo-Bayesian inference on the volatility of a jump diffusion process

Ryan Martin

Department of Statistics  
North Carolina State University  
rgmarti3@ncsu.edu

Cheng Ouyang and Francois Domagni

Department of Mathematics, Statistics, and Computer Science  
University of Illinois at Chicago  
(couyang, fdomag2)@uic.edu

August 25, 2016

## Abstract

Jump diffusion processes are widely used to model asset prices over time, mainly for their ability to capture complex discontinuous behavior, but inference on the model parameters remains a challenge. Here our goal is Bayesian inference on the volatility of the diffusion part of the process based on discrete samples. Since volatility is our only interest, an unspecified model for the jump part of the process is desirable, but this causes serious problems for the familiar Bayesian approach. To avoid the difficulties caused by modeling the jumps, we consider a purposely misspecified model that ignores the jump part entirely. We work out precisely the asymptotic behavior of the posterior under the misspecified model, propose some simple modifications to correct for the effects of misspecification, and demonstrate that the modified pseudo-Bayesian inference on the volatility is efficient in the sense that its asymptotic variance equals the no-jumps model Cramér–Rao bound.

*Keywords and phrases:* Bernstein–von Mises theorem; Brownian motion; credible interval; Gibbs posterior; uncertainty quantification.

## 1 Introduction

Jump diffusion models have gained considerable attention in the last two decades, especially in finance and economics, where they are used to model asset prices as a function of time. An advantage of these models, over the classical Black–Scholes models (e.g. Musiela and Rutkowski 2005), based solely on a continuous Brownian motion, is their ability to accommodate the rapid—seemingly discontinuous—changes in asset prices often seen in applications. In fact, several authors have concluded that neither a purely-continuous nor

purely-jump model is sufficient for real applications (e.g., Aït-Sahalia and Jacod 2009, 2010; Barndorff-Nielsen and Shepard 2006; Podolskij 2006). More specifically, by comparing the observed behavior of at-the-money and out-of-the-money call option prices near expiration with their analogous theoretical behavior, Carr and Wu (2003) and Medvedev and Scaillet (2007) argued that both continuous and jump components are necessary to explain the implied volatility behavior of S&P500 index options. In this paper we consider a continuous-time process  $X = (X_t : t \in [0, T])$  over a fixed and finite time horizon  $[0, T]$  that can be decomposed as

$$X_t = \beta t + \theta^{1/2} W_t + J_t, \quad t \in [0, T], \quad (1)$$

where  $\beta t + \theta^{1/2} W_t$  is a continuous diffusion—with  $\beta$  the drift coefficient,  $\theta$  the volatility coefficient, and  $(W_t : t \in [0, T])$  a standard Brownian motion—and  $J = (J_t : t \in [0, T])$  is a pure jump process with finite activity and independent of  $W$ . We emphasize here that we only assume that, with probability 1, the jump process  $J$  has a finite number of jumps in  $[0, T]$  and that each jump is finite. For example, the results herein cover the case where  $J$  is a compound Poisson process. The quantity of interest here is the volatility coefficient  $\theta$ , a fundamentally important measure of uncertainty or risk (Musielka and Rutkowski 2005)). Our goal is to construct a Bayesian posterior distribution or, more generally, a data-dependent measure,  $\Pi_n$  on  $\mathbb{R}^+ := (0, \infty)$  that can be used to provide valid uncertainty quantification about the volatility coefficient.

If the entire process  $X$  were observable, then we could immediately identify the jumps and, by subtraction, this could be converted to a standard problem. However, here, as is typically the case in practice, the process  $X$  is not fully observable; in particular, we can only observe  $X$  at  $n$  fixed times  $0 < t_1 < t_2 < \dots < t_n < T$ . This is the context considered in, e.g., Aït-Sahalia and Jacod (2009) and Figueroa-López (2009). A consequence of observing the process only at discrete times is that the continuous and jump parts cannot be separated, so we are forced to deal with the jump part of the process. Bayesian uncertainty quantification is achieved by modeling all unknowns and using Bayes’s theorem to get a posterior distribution for inference on  $\theta$ . Unfortunately, the need to model the jump portion of the process causes a number of difficulties: in particular, for sufficiently flexible models for  $J$ , the likelihood function is analytically intractable; specifying priors and computing the posterior for the distribution of  $J$ , perhaps using ABC methods (e.g., Marin et al. 2012), may be non-trivial and expensive; and a misspecified model for  $J$  can bias the posterior inference on  $\theta$ . Since the distribution of  $J$  in (1) is a nuisance parameter, it is desirable to have an alternative “Bayesian-like” approach that gives a posterior distribution for the volatility while, at the same time, avoids those shortcomings of a proper Bayesian approach. This is our goal here.

Towards this, we consider a purposely misspecified model that completely ignores the jumps, basically treating the observations as if they arise from a simple diffusion model. This misspecified model is highly regular and easy to compute, so if this is not heavily influenced by misspecification, then perhaps it would suffice for valid uncertainty quantification on the volatility coefficient  $\theta$ . A special case of our Theorem 1 below says that the corresponding pseudo-Bayesian posterior is asymptotically normal but the misspecification has some undesirable effects, namely, the center is off-target and the spread is too large. To correct for the effects of misspecification, we consider two adjustments: a suitable scaling of the log-likelihood to correct the spread and a location shift. Both of

these adjustments rely on us having a suitable estimator of the quadratic variation of the jump process  $J$ . We then show, in Theorem 2, that the corresponding modified pseudo-Bayesian posterior is asymptotically normal, centered around a consistent estimator of the true volatility, with variance equal to the Cramér–Rao lower bound for optimal/ideal case when there are no jumps, i.e., the model is not misspecified. The particular modification is easy to implement and we present some simulation results to illustrate the validity of the corresponding pseudo-Bayesian posterior credible intervals in finite samples.

## 2 A misspecified model

Assume that we observe the continuous-time process  $(X_t)$  at  $n$  distinct time points  $t_1 < \dots < t_n$ , i.e., our observations are  $X_{t_1}, \dots, X_{t_n}$ ; for notational convenience later on, set  $t_0 = 0$  and  $X_0 \equiv 0$ . For notational simplicity, we will assume that the time points are equally spaced, so that each time difference  $t_i - t_{i-1}$  equals  $\Delta_n = Tn^{-1}$ ; the case of non-equally spaced sampling can be handled similarly. To avoid dealing directly with the jump component of the model (1), we consider a purposely misspecified model that ignores both the drift and the jump part, i.e., it assumes that the differences  $D_i = X_{t_i} - X_{t_{i-1}}$ ,  $i = 1, \dots, n$ , are iid  $N(0, \theta\Delta_n)$  for some  $\theta > 0$ . This misspecified model is easy to work with and has no nuisance parameters so, if it—or a simple modification thereof—also provides valid inference on the volatility, then it ought to be useful. The likelihood function for this misspecified model, up to proportionality constants, is given by

$$L_n(\theta) = \theta^{-n/2} \exp\left\{-\frac{1}{2\Delta_n\theta} \sum_{i=1}^n D_i^2\right\} = \theta^{-n/2} \exp\left\{-\frac{n\hat{\theta}_n}{2\theta}\right\}, \quad (2)$$

where

$$\hat{\theta}_n = (n\Delta_n)^{-1} \sum_{i=1}^n D_i^2 = T^{-1} \sum_{i=1}^n D_i^2,$$

is the maximum likelihood estimator. Just like in the familiar Bayes approach, we introduce a prior distribution  $\Pi$  for  $\theta$ , with density function  $\pi$ . Here we consider a generalization of the Bayesian setup by define the pseudo-posterior distribution as

$$\Pi_n(A) = \frac{\int_A L_n(\theta)^{1/\kappa_n} \pi(\theta) d\theta}{\int L_n(\theta)^{1/\kappa_n} \pi(\theta) d\theta}, \quad A \subseteq \mathbb{R}^+, \quad (3)$$

where  $\kappa_n$  is a suitable (possibly stochastic) sequence to be specified. The distribution  $\Pi_n$  in (3) is sometimes referred to as a “Gibbs posterior” (e.g., Bissiri et al. 2016; Grünwald and van Ommen 2016; Jiang and Tanner 2008; Syring and Martin 2015; Zhang 2006a,b) and  $\kappa_n$  is a “temperature” parameter; the case  $\kappa_n \equiv 1$  corresponds to the usual Bayes posterior. Unlike in the well-specified Bayesian setting, where posterior consistency is typical, our model being misspecified means that we cannot expect  $\Pi_n$  to converge to a point mass at the true volatility coefficient. Therefore, some correction will be needed to point the pseudo-posterior towards the true volatility coefficient, but first we need to understand how  $\Pi_n$  in (3) behaves without any intervention on our part.

A fundamental result in Bayesian asymptotics is the Bernstein–von Mises theorem, which states that, under certain regularity conditions, a suitably centered and scaled

version of the posterior will resemble a normal distribution, in the sense that the total variation distance between that normalized posterior and the normal distribution converges to zero in probability. This classical version is typically used in the case of a well-specified model, but recently there has been work on a version of the Bernstein–von Mises theorem for misspecified models. In particular, Kleijn and van der Vaart (2012), in their Theorem 2.1, give a Bernstein–von Mises theorem when the model is misspecified. Our result that follows is based on their approach.

Before stating the result, we need to introduce some notation. Let  $\mathbf{P}^*$  denote the distribution of the differences  $(D_1, \dots, D_n)$ , with  $D_i = X_{t_i} - X_{t_{i-1}}$ , under the jump diffusion model, and  $\mathbf{P}_J^*$  the corresponding conditional distribution, given the jump part  $J$  of the process (1). Also, let  $\beta^*$  and  $\theta^*$  denote the true drift and volatility coefficients, and define the expectation, conditional expectation, variance, and conditional variance as  $\mathbf{E}^*$ ,  $\mathbf{E}_J^*$ ,  $\mathbf{V}^*$ , and  $\mathbf{V}_J^*$ , respectively. We consider a “high-frequency” scenario (e.g., Aït-Sahalia and Jacod 2014), so  $n$  is large and it is safe to assume that, with probability 1, the time windows  $[t_{i-1}, t_i)$  contain at most one jump. Therefore, for almost all  $J$ , under  $\mathbf{P}_J^*$ , we have that  $(D_1, \dots, D_n)$  are independent,  $D_i \sim \mathbf{N}(\beta^* \Delta_n + \mu_i, \theta^* \Delta_n)$ , where

$$\mu_i = J_{t_i} - J_{t_{i-1}}, \quad i = 1, \dots, n. \quad (4)$$

For a given  $J$ , let  $\langle J \rangle = \sum_{i=1}^n \mu_i^2$  denote the quadratic variation of the jump process  $J$ , and let  $\{J\}$  denote the set of indices  $i$  such that the window  $[t_{i-1}, t_i)$  contains a jump, i.e.,  $\mu_i \neq 0$  if and only if  $i \in \{J\}$ . We assume that the process (1) has finite jump activity, so  $|\{J\}| \vee \langle J \rangle < \infty$  with  $\mathbf{P}^*$ -probability 1. We also assume that  $\kappa_n$  is a stochastic sequence and that there exists  $\kappa^\dagger$ , possibly depending on  $J$ , such that  $\kappa_n \rightarrow \kappa^\dagger$  in  $\mathbf{P}_J^*$ -probability for the given  $J$ . Finally, the point around which the pseudo-posterior will concentrate is

$$\theta^\dagger = \theta^* + T^{-1} \langle J \rangle.$$

Note that both  $\theta^\dagger$  and  $\kappa^\dagger$  are constants with respect to the conditional distribution  $\mathbf{P}_J^*$ .

**Theorem 1.** *Consider the pseudo-Bayesian posterior  $\Pi_n$  in (3) based on a prior  $\Pi$  and the misspecified model with likelihood in (2). If the prior density  $\pi$  is continuous and positive in a neighborhood of  $\theta^\dagger$ , then, for  $\mathbf{P}^*$ -almost all  $J$ , the posterior  $\Pi_n$  is asymptotically normal in the sense that*

$$d(\Pi_n, \mathbf{N}(\hat{\theta}_n, 2\kappa^\dagger \theta^{\dagger 2} n^{-1})) \rightarrow 0 \quad \text{in } \mathbf{P}_J^*\text{-probability as } n \rightarrow \infty,$$

where  $d$  is the total variation distance. The above conclusion also holds unconditionally, i.e., the above convergence is also in  $\mathbf{P}^*$ -probability.

*Proof.* See the Appendix. □

The theorem asserts that, for the “high-frequency” setting where  $n$  is large, if the data-generating process (1) has finite jump activity, then the pseudo-posterior will resemble a normal distribution centered around  $\hat{\theta}_n$ . Since  $\hat{\theta}_n$  converges to  $\theta^\dagger$  (see the proof of the theorem), it follows that the posterior will resemble a normal distribution centered at  $\theta^\dagger$ . This is different from the usual Bernstein–von Mises theorems found in the Bayesian literature in that the point around which the posterior concentrates depends on both parameters and a hidden portion of the data, namely,  $\langle J \rangle$ .

There are two seemingly undesirable consequences of misspecification. The first, as alluded to above, is that  $\Pi_n$  is biased in the sense that the point around which  $\Pi_n$  concentrates is  $\theta^\dagger$  instead of the true volatility coefficient  $\theta^*$ . The second is more subtle and concerns the spread of  $\Pi_n$ . Kleijn and van der Vaart (2012, Sec. 1) point out that the asymptotic variance in their Bernstein–von Mises theorem may not agree with that for  $\hat{\theta}_n$  based on M-estimation theory (e.g., van der Vaart 1998, Ch. 5). Indeed, the maximum likelihood estimator  $\hat{\theta}_n$  for the misspecified model can be viewed as an M-estimator and will, therefore, be asymptotically normal, with asymptotic variance given by the so-called “sandwich formula” which, in this case, gives

$$\mathbf{V}_J^*(\hat{\theta}_n) = \frac{2\theta^{\dagger 2}}{n} \left\{ 1 - \left( \frac{\langle J \rangle}{T\theta^\dagger} \right)^2 \right\} + O(n^{-2}).$$

This follows from the calculations leading up to (12) in the Appendix. Up to order  $n^{-1}$ , this closely resembles the asymptotic variance in Theorem 1; in particular, if  $\kappa^\dagger$  were equal to the term in braces above, then the two variance formulas agree. Note that the genuine Bayes posterior has  $\kappa^\dagger = 1$  and, therefore, will have asymptotic variance larger than that in the above display. Consequently, the pseudo-Bayesian posterior credible intervals would be too large, making the inference inefficient. Section 3 below describes how we can correct for these two undesirable consequences of misspecification.

### 3 Correcting for misspecification

As discussed above, there are two effects of the model misspecification on  $\Pi_n$ , both depending on the quadratic variation of the jump portion of the process. To deal with these effects, we will need a suitable estimator of the quadratic variation  $\langle J \rangle$ . Intuitively, those observed differences  $D_i$  which are of relatively large magnitude are likely due to jumps, so a reasonable estimator is

$$\widehat{\langle J \rangle}_n = \sum_{i=1}^n D_i^2 1(|D_i| > \eta_n), \quad (5)$$

where  $\eta_n$  is a sequence that vanishes sufficiently slow, and  $1(\cdot)$  is the indicator function. We claim that if  $\eta_n \propto n^{-\omega}$  for some  $\omega \in (0, \frac{1}{2})$ , then

$$\mathbf{E}_J^* |\widehat{\langle J \rangle}_n - \langle J \rangle| = O(n^{-1/2}), \quad n \rightarrow \infty, \quad (6)$$

for  $\mathbf{P}^*$ -almost all  $J$ . Results of this type for Lévy processes are available in the literature (e.g., Aït-Sahalia and Jacod 2014, Fact 3.7), but the proof of (6) given in the Appendix under only the finite jump activity assumption is relatively simple. With a suitable estimator in hand, now we are ready to address the effects of misspecification.

Towards constructing a modified pseudo-posterior for the volatility, we must consider the following question: what is the “correct/optimal” asymptotic variance in the normal approximation? Of course, the best possible inference obtains if the model is not misspecified, i.e., there are no jumps; this is equivalent to the case where the sample path of the process is fully observed since, in that case, the jumps are visible and can be

removed. An easy calculations reveals that, in this ideal case, the asymptotic variance is the Cramér–Rao bound,  $2\theta^{*2}n^{-1}$ . This optimal variance obtains in Theorem 1 if

$$\kappa^\dagger = \left(\frac{\theta^*}{\theta^\dagger}\right)^2 = \left(1 - \frac{\langle J \rangle}{T\theta^\dagger}\right)^2.$$

This suggests we choose  $\kappa_n$  in (3) as

$$\kappa_n = \left(1 - \frac{\widehat{\langle J \rangle}_n}{T\widehat{\theta}_n}\right)^2, \quad (7)$$

so that, by (6),  $\kappa_n \rightarrow \kappa^\dagger$  in  $\mathbf{P}_J^*$ -probability for  $\mathbf{P}^*$ -almost all  $J$ . With this understanding, we define a “modified” pseudo-Bayesian posterior  $\widetilde{\Pi}_n$  as the distribution of  $\theta - T^{-1}\widehat{\langle J \rangle}_n$  when  $\theta$  is distributed as  $\Pi_n$  in (3), with  $\kappa_n$  as in (7). In other words, if  $\pi_n$  is the density function corresponding to  $\Pi_n$ , then  $\widetilde{\Pi}_n$  has density function

$$\widetilde{\pi}_n(\theta) = \pi_n(\theta + T^{-1}\widehat{\langle J \rangle}_n), \quad \theta \in \mathbb{R}^+. \quad (8)$$

Then we have the following Bernstein–von Mises theorem for  $\widetilde{\Pi}_n$ .

**Theorem 2.** *Under the same setup as in Theorem 1, for  $\mathbf{P}^*$ -almost all  $J$ , the modified pseudo-Bayesian posterior  $\widetilde{\Pi}_n$ , with  $\kappa_n$  as in (7), satisfies*

$$d(\widetilde{\Pi}_n, \mathbf{N}(\widehat{\theta}_n - T^{-1}\widehat{\langle J \rangle}_n, 2\theta^{*2}n^{-1})) \rightarrow 0 \quad \text{in } \mathbf{P}_J^*\text{-probability as } n \rightarrow \infty,$$

where  $d$  is the total variation distance. The above convergence is also in  $\mathbf{P}^*$ -probability.

*Proof.* Since the total variation distance is invariant to location shifts, we have that

$$d(\widetilde{\Pi}_n, \mathbf{N}(\widehat{\theta}_n - T^{-1}\widehat{\langle J \rangle}_n, 2\theta^{*2}n^{-1})) = d(\Pi_n, \mathbf{N}(\widehat{\theta}_n, 2\theta^{*2}n^{-1})).$$

That the right-hand side converges to 0 in  $\mathbf{P}_J^*$ -probability follows from Theorem 1 and the particular choice of  $\kappa_n$  in (7). The  $\mathbf{P}^*$  convergence is proved just like in Theorem 1.  $\square$

The first observation is that, since  $\widehat{\theta}_n - T^{-1}\widehat{\langle J \rangle}_n$  is a consistent estimator of  $\theta^*$ , the modified pseudo-posterior is concentrating around the true volatility coefficient, as desired. Furthermore, by our choice of the sequence  $\kappa_n$ , the asymptotic variance agrees with that achieved in the ideal case where there are no jumps present or, equivalently, when the sample path of the process is fully observed. The remaining question is if the posterior variance agrees with the variance of the center  $\widehat{\theta}_n - T^{-1}\widehat{\langle J \rangle}_n$  under  $\mathbf{P}^*$ . Proposition 1 in Ait-Sahalia and Jacod (2010) reveals that, in the present setting, under  $\mathbf{P}_J^*$ , the estimator  $\widehat{\theta}_n - T^{-1}\widehat{\langle J \rangle}_n$  satisfies a central limit theorem, with asymptotic variance  $2\theta^{*2}n^{-1}$ . Therefore, the credible intervals coming from the modified posterior  $\widetilde{\Pi}_n$  will be asymptotically valid under  $\mathbf{P}_J^*$ , i.e., the coverage probability of the  $100(1 - \alpha)\%$  credible intervals will converge to  $1 - \alpha$  for  $\mathbf{P}^*$ -almost all  $J$ . It follows immediately from the dominated convergence theorem that the coverage probability converges to  $1 - \alpha$  under  $\mathbf{P}^*$  as well. It turns out that the finite-sample performance depends on the choice of threshold  $\eta_n$  in (5) and we address this in Section 4.

## 4 Numerical results

An important question is how to choose the threshold  $\eta_n$ . The theory says that we need  $\eta_n = mn^{-\omega}$  for some  $m > 0$  and some  $\omega \in (0, \frac{1}{2})$  but, in finite samples,  $m$  and  $\omega$  are not independent parameters; that is, only the value of  $\eta_n$  matters, not the particular  $(m, \omega)$ . This point is discussed at length in Aït-Sahalia and Jacod (2014, Sec. 6.2.2), and they suggest one reasonable strategy for choosing  $\eta_n$ . We consider here a simpler approach based on outlier detection. That is, let  $Q$  denote the interquartile range of the observed increment magnitudes  $|D_1|, \dots, |D_n|$ ; this value is likely to be small since almost all of the increments correspond to the diffusion part of the process. Take  $\eta_n$  to be some value that lower-bounds the set of all  $|D_i|$ s that exceeds some cutoff, say,  $5Q$ . We make no claims that this approach is “optimal” in any sense, only that it is both simple and reasonable. A thorough investigation of this and the Aït-Sahalia and Jacod method for selecting the threshold  $\eta_n$  is beyond the scope of this paper.

For illustration, consider the model (1) with drift  $\beta^* = 1$ , volatility  $\theta^* = 10$ , and compound Poisson process jumps with a rate of  $\lambda = 5$  jumps per unit time and jumps sampled from the discrete uniform distribution on  $\{-\tau, \tau\}$  with  $\tau = 3$ . We simulate  $n = 5000$  equally spaced observations from this process. A plot of the observed sample path on the interval  $[0, 1]$  is shown in Figure 1(a) with the jump times highlighted by vertical lines. For the misspecified Bayes model, we consider a conjugate inverse gamma prior with shape  $a = 1$  and rate  $b = 1$ ; the presence of the temperature parameter  $\kappa_n$  does not affect conjugacy. We also fix  $\eta_n$  based on the interquartile range strategy described above. Figure 1(b) shows the pseudo-posterior density function (8), the corresponding 95% credible interval for the volatility, and the density function of the normal approximation in Theorem 2. The key observations are that the credible region contains the true volatility in this case and that the modified pseudo-Bayes posterior and the normal approximation are very similar.

To further investigate the finite-sample properties our proposed approach for inference on the volatility, we consider a simulation study. Using the same model as above, but varying the jump rate  $\lambda \in \{4, 8, 16, 32\}$ , the jump magnitude  $\tau \in \{1, 2, 4, 8\}$ , and the sample size  $n$ , we investigate the coverage probability of the modified pseudo-Bayesian credible intervals, based on the choice of threshold mentioned above. Figure 2 displays the empirical coverage probability, based on 5000 Monte Carlo simulations in each setting, summarized over the jump rate and standard deviation, for several values of  $n$ . This plot reveals that the choice of threshold based on the interquartile range performs reasonably well in this setting, giving coverage probabilities very close to the nominal 95% level.

## 5 Concluding remarks

In this paper we considered the construction of a pseudo-Bayes posterior for inference on the volatility in a complex jump diffusion model where the jump distribution is left unspecified. Our strategy here was to misspecify the model on purpose, to avoid dealing directly with a model for the jump portion of the process, and then to correct in some other way for the effects of misspecification. Of course, this “misspecification on purpose” strategy could be used in many other problems to provide valid uncertainty quantification for the parameters of interest without having to specify a complete model for the possibly

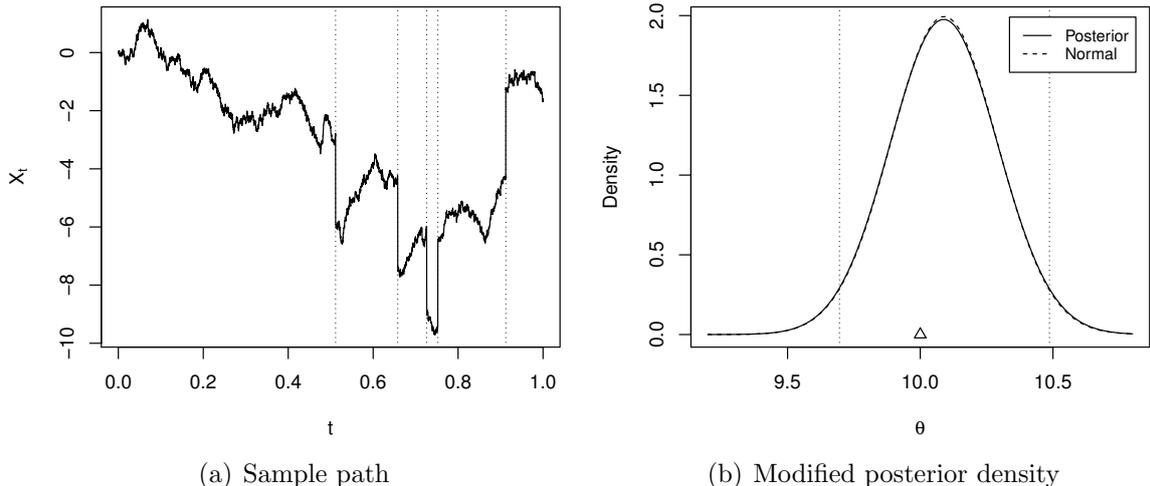


Figure 1: Plot of one sample path  $X$  (left) along with the corresponding density function for the modified pseudo-posterior (right); also shown in the right panel is the 95% posterior credible interval (vertical lines), the normal approximation in Theorem 2, and the true volatility coefficient ( $\Delta$ ).

very complex nuisance parameters, which is very attractive. Furthermore, when one has reliable prior information about the interest parameter, it is straightforward to incorporate this into the proposed pseudo-Bayesian analysis compared to a fully non-Bayesian approach, say, using M-estimation.

For this particular model, there are several extensions that one could consider. For example, if the drift parameter was also of interest, then, instead of ignoring  $\beta$  as we did here, it would be relatively straightforward to construct the same modified Bayesian posterior for the pair  $(\beta, \theta)$ . More interesting is the case where the volatility parameter is not a scalar constant but, instead, a function  $\theta_t$ . Certain functionals of  $\theta_t$ , in particular, the average volatility  $T^{-1} \int_0^T \theta_t dt$ , can be inferred directly using virtually the same techniques as presented here, but inference on the function itself would require more.

## Acknowledgments

The authors thank Nick Syring for some helpful comments on a previous version of the manuscript. A portion of this work was completed while the first author was with the Department of Mathematics, Statistics, and Computer Science, at the University of Illinois at Chicago.

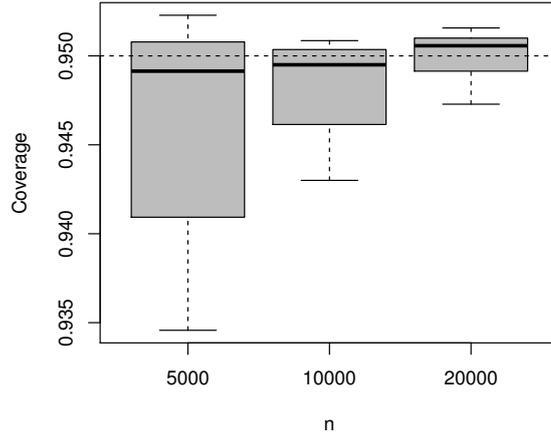


Figure 2: Summary of the empirical coverage probability of the modified pseudo-Bayesian 95% posterior credible intervals, over various several values of the jump rate and standard deviation (see the text), and for several values of  $n$ .

## A Proofs

### A.1 Proof of Theorem 1

Without loss of generality, we will assume in the proof that  $T = 1$  and  $\Delta_n = n^{-1}$ . To prove the Bernstein–von Mises theorem, follow the approach described in Theorem 2.1 of Kleijn and van der Vaart (2012). Our first objective is to show that

$$\mathbf{E}_J^*[\Pi_n(\{\theta : |\theta - \theta^\dagger| > M_n n^{-1/2}\})] \rightarrow 0, \quad n \rightarrow \infty,$$

for any sequence of constants  $M_n \rightarrow \infty$ . To establish this, we need only to study the posterior mean and variance. That is, if  $\mathbf{E}_{\Pi_n}$  denotes expectation with respect to the pseudo-posterior, then, by Markov’s inequality, we have

$$\Pi_n(\{\theta : |\theta - \theta^\dagger| > M_n n^{-1/2}\}) \leq n M_n^{-2} \mathbf{E}_{\Pi_n}(\theta - \theta^\dagger)^2. \quad (9)$$

To show that the expectation of the left-hand side in the above display vanishes, it suffices to show that

$$\mathbf{E}_J^*\{\mathbf{E}_{\Pi_n}(\theta - \theta^\dagger)^2\} = O(n^{-1}). \quad (10)$$

Towards this, we will use the Laplace approximation which says that, for suitable functions  $g$ , the posterior mean of  $g(\theta)$  is

$$g(\hat{\theta}_n)\{1 + O(n^{-1})\}, \quad n \rightarrow \infty.$$

Therefore, in our case, if we apply the above to  $g(\theta) = (\theta - \theta^\dagger)^2$ , then we have

$$\mathbf{E}_{\Pi_n}(\theta - \theta^\dagger)^2 = (\hat{\theta} - \theta^\dagger)^2\{1 + O(n^{-1})\}.$$

Since the log-likelihood function for our misspecified model has a unique maximum  $\hat{\theta}_n$  in the interior of the  $\theta$ -space and satisfies  $(\log L_n)''(\hat{\theta}_n) < 0$ , the big-oh term above is uniform

in observations, i.e., the  $O(n^{-1})$  term in the above display is a function of  $(D_1, \dots, D_n)$  that can be uniformly bounded by a constant times  $n^{-1}$  and, in particular, the scaling by  $\kappa_n^{-1}$  does not affect this conclusion. Therefore, to get (10), it suffices to show that

$$\mathbf{E}_J^*(\hat{\theta}_n - \theta^\dagger)^2 = O(n^{-1}) \quad (11)$$

Towards showing (11), we recall that  $\hat{\theta}_n = \sum_{i=1}^n D_i^2$  where, under  $\mathbf{P}_J^*$ ,  $(D_1, \dots, D_n)$  are independent with

$$D_i \sim \mathbf{N}(\beta^* \Delta_n + \mu_i, \theta^* \Delta), \quad i = 1, \dots, n,$$

and  $\mu_i$  are defined in (4). It follows that  $\hat{\theta}_n$  has the same distribution as  $\theta^* \Delta_n$  times a non-central chi-square random variable,  $Y$ , with  $n$  degrees of freedom and non-centrality parameter  $\lambda = (\theta^* \Delta_n)^{-1} \sum_{i=1}^n (\beta^* \Delta_n + \mu_i)^2$ . In particular,

$$\mathbf{E}(Y) = n + \lambda \quad \text{and} \quad \mathbf{V}(Y) = 2(n + 2\lambda).$$

If we let  $\mathbf{V}_J^*$  denote variance with respect to  $\mathbf{P}_J^*$ , then we have that

$$\begin{aligned} \mathbf{E}_J^*(\hat{\theta}_n - \theta^\dagger)^2 &= \mathbf{V}_J^*(\hat{\theta}_n) + \{\mathbf{E}_J^*(\hat{\theta}_n) - \theta^\dagger\}^2 \\ &= (\theta^* \Delta_n)^2 \mathbf{V}(Y) + \{(\theta^* \Delta_n) \mathbf{E}(Y) - \theta^\dagger\}^2. \end{aligned}$$

Plugging in the formulas for the mean and variance of  $Y$  and simplifying, gives

$$\mathbf{E}_J^*(\hat{\theta}_n - \theta^\dagger)^2 = 2\theta^* \theta^\dagger n^{-1} + O(n^{-2}). \quad (12)$$

This is clearly  $O(n^{-1})$ , so we have established (11). Note that this derivation depends on  $J$  only through  $\{|J|\}$  and  $\langle J \rangle$ . Since (11) implies (10), we have proved the claimed pseudo-posterior concentration rate result.

Next, we need to demonstrate that the model is suitably regular. More specifically, Theorem 2.1 in Kleijn and van der Vaart (2012) require that the model satisfies a certain local asymptotic normality property, i.e., the log-likelihood ratio has a quadratic approximation locally around the specified  $\theta^\dagger$ . Since the misspecified model is so nice, it is a straightforward exercise to show that

$$\left| \frac{1}{\kappa_n} \log \frac{L_n(\theta^\dagger + n^{-1/2}h)}{L_n(\theta^\dagger)} - V_{\theta^\dagger} Z_n(\theta^\dagger) h + \frac{1}{2} V_{\theta^\dagger} h^2 \right| = o_{\mathbf{P}_J^*}(1),$$

where  $V_{\theta^\dagger} = (2\kappa^\dagger \theta^{\dagger 2})^{-1}$  and  $Z_n(\theta^\dagger) = n^{1/2}(\hat{\theta}_n - \theta^\dagger)$ . The above display holds uniformly on compact subsets of  $h$ , and it follows from (11) that  $Z_n(\theta^\dagger)$  is bounded in  $\mathbf{P}_J^*$ -probability. Therefore, the assertion in Theorem 1 follows from Kleijn and van der Vaart's.

The extension of these results to the unconditional distribution,  $\mathbf{P}^*$ , is also straightforward. Based on the finite jump activity assumption, all that we demonstrated above holds with  $\mathbf{P}^*$ -probability 1. In particular, we have that, for any  $\varepsilon > 0$ ,

$$\mathbf{P}_J^* \{d(\Pi_n, \mathbf{N}(\hat{\theta}_n, 2\kappa^\dagger \theta^{\dagger 2} n^{-1}) > \varepsilon\} \rightarrow 0, \quad \text{for } \mathbf{P}^*\text{-almost all } J.$$

Since this sequence is bounded and converges almost surely, it follows from the dominated convergence theorem that

$$\mathbf{P}^* \{d(\Pi_n, \mathbf{N}(\hat{\theta}_n, 2\kappa^\dagger \theta^{\dagger 2} n^{-1}) > \varepsilon\} = \mathbf{E}^* [\mathbf{P}_J^* \{d(\Pi_n, \mathbf{N}(\hat{\theta}_n, 2\kappa^\dagger \theta^{\dagger 2} n^{-1}) > \varepsilon\}] \rightarrow 0,$$

i.e.,  $d(\Pi_n, \mathbf{N}(\hat{\theta}_n, 2\kappa^\dagger \theta^{\dagger 2} n^{-1}) \rightarrow 0$  in  $\mathbf{P}^*$ -probability.

## A.2 Proof of the claim in Equation (6)

For  $\mathbf{P}^*$ -almost all  $J$ , we have that  $(D_1, \dots, D_n)$  are independent under  $\mathbf{P}_J^*$  with

$$D_i = \mu_i + Z_i \sim \mathbf{N}(\beta^* \Delta_n + \mu_i, \theta^* \Delta_n), \quad i = 1, \dots, n$$

where  $\mu_i$  are given in (4) and  $\mu_i \neq 0$  only for those indices  $i \in \{J\}$ . To prove the claim in Equation (6), we split the indices to those that contain a jump (in  $\{J\}$ ) and those that do not (in  $\{J\}^c$ ). Then we get

$$\begin{aligned} \widehat{\langle J \rangle}_n - \langle J \rangle &= \sum_{i=1}^n D_i^2 \mathbf{1}(|D_i| > \eta_n) - \langle J \rangle \\ &= \sum_{i \notin \{J\}} Z_i^2 \mathbf{1}(|Z_i| > \eta_n) + \sum_{i \in \{J\}} Z_i^2 \mathbf{1}(|Z_i + J_i| > \eta_n) \\ &\quad + 2 \sum_{i \in \{J\}} \mu_i Z_i \mathbf{1}(|Z_i + \mu_i| > \eta_n) + \sum_{i \in \{J\}} \mu_i^2 \mathbf{1}(|Z_i + \mu_i| \leq \eta_n). \end{aligned}$$

Take absolute value of both sides, apply the triangle inequality, and then take expectation. This yields the inequality

$$\begin{aligned} \mathbf{E}_J^* |\widehat{\langle J \rangle}_n - \langle J \rangle| &\leq \sum_{i \notin \{J\}} \mathbf{E}_J^* \{Z_i^2 \mathbf{1}(|Z_i| > \eta_n)\} + \sum_{i \in \{J\}} \mathbf{E}_J^* \{Z_i^2 \mathbf{1}(|Z_i + \mu_i| > \eta_n)\} \\ &\quad + 2 \sum_{i \in \{J\}} \mathbf{E}_J^* \{|\mu_i Z_i| \mathbf{1}(|Z_i + \mu_i| > \eta_n)\} + \sum_{i \in \{J\}} \mathbf{E}_J^* \{\mu_i^2 \mathbf{1}(|Z_i + \mu_i| \leq \eta_n)\}. \end{aligned}$$

We will proceed by showing, one by one, that each of the four terms in the upper bound above is  $O(n^{-1/2})$ . First, note that

$$Z_i \sim \mathbf{N}(\beta^* \Delta_n, \theta^* \Delta_n), \quad i = 1, \dots, n,$$

are iid and hence its fourth moment is bounded by a constant independent of  $n$ . Then we have, by the Cauchy–Schwartz inequality

$$\begin{aligned} \sum_{i \notin \{J\}} \mathbf{E}_J^* \{Z_i^2 \mathbf{1}(|Z_i| > \eta_n)\} &\lesssim (\mathbf{E}_J^* |Z_1|^4)^{1/2} \sum_{i \notin \{J\}} (\mathbf{P}(|Z_i| > \eta_n))^{1/2} \\ &\lesssim |\{J\}^c| \mathbf{P}(|Z_i| > \eta_n)^{1/2}. \end{aligned}$$

It is clear that  $|\{J\}^c|$  is of order  $n$ . So we only need to find a good bound for the tail probability. Assume, for the moment, that  $\beta^* > 0$ . Using the usual normal tail probability bounds, we get

$$\mathbf{P}(|Z_i| > \eta_n) \lesssim \frac{\Delta_n^{1/2}}{\eta_n - \beta^* \Delta_n} e^{-(\eta_n - \beta^* \Delta_n)^2 / \Delta_n} \lesssim \frac{\Delta_n^{1/2}}{\eta_n} e^{-\eta_n^2 / \Delta_n}.$$

So, if  $\eta_n \propto n^{-\omega}$ , for  $\omega \in (0, \frac{1}{2})$ , then the upper bound for the above tail probability is  $o(n^{-k})$  for any positive integer  $k$ . Hence it follows easily

$$\sum_{i \notin \{J\}} \mathbf{E}_J^* \{Z_i^2 \mathbf{1}(|Z_i| > \eta_n)\} = o(n^{-1/2});$$

the same conclusion can be reached if  $\beta^* < 0$ . Next,

$$\sum_{i \in \{J\}} \mathbf{E}_J^* \{Z_i^2 \mathbf{1}(|Z_i + \mu_i| > \eta_n)\} \leq |\{J\}| \{\theta^* \Delta_n + (\beta^* \Delta_n)^2\} = O(n^{-1})$$

and, similarly, using Cauchy–Schwartz,

$$\sum_{i \in \{J\}} \mathbf{E}_J^* \{|\mu_i Z_i| \mathbf{1}(|Z_i + \mu_i| > \eta_n)\} \lesssim \{\theta^* \Delta_n + (\beta^* \Delta_n)^2\}^{1/2} \sum_{i \in \{J\}} |\mu_i| = O(n^{-1/2}).$$

For the last term, we need to bound  $\mathbf{P}_J^*(|Z_i + \mu_i| \leq \eta_n)$ . Again, without loss of generality, if we assume that  $\beta^* > 0$  and  $\mu_i > 0$ , then we get

$$\mathbf{P}_J^*(|Z_i + \mu_i| \leq \eta_n) \leq \mathbf{P}\{\mathbf{N}(0, 1) > \Delta_n^{-1/2}(\mu_i + \beta^* \Delta_n - \eta_n)\},$$

which, using the normal tail probability bound again, is bounded by

$$\frac{\Delta_n^{1/2}}{\mu_i + \beta^* \Delta_n - \eta_n} e^{-(\mu_i + \beta^* \Delta_n - \eta_n)^2 / \Delta_n}.$$

Since  $\mu_i > 0$  is a fixed constant, the above quantity vanishes exponentially fast, so

$$\sum_{i \in \{J\}} \mu_i^2 \mathbf{P}(|Z_i + \mu_i| \leq \eta_n) \leq o(n^{-1/2}) \sum_{i \in \{J\}} \mu_i^2.$$

All four terms have been shown to be  $O(n^{-1/2})$ , completing the proof of (6). Finally, note that the result holds for all  $J$  such that  $|\{J\}|$  and  $\langle J \rangle = \sum_{i \in \{J\}} \mu_i^2$  are finite. Since this is a  $\mathbf{P}^*$ -probability 1 event, (6) holds for  $\mathbf{P}^*$ -almost all  $J$ .

## References

- Aït-Sahalia, Y. and Jacod, J. (2009). Testing for jumps in a discretely observed process. *Ann. Statist.*, 37(1):184–222.
- Aït-Sahalia, Y. and Jacod, J. (2010). Is Brownian motion necessary to model high-frequency data? *Ann. Statist.*, 38(5):3093–3128.
- Aït-Sahalia, Y. and Jacod, J. (2014). *High-Frequency Financial Econometrics*. Princeton University Press.
- Barndorff-Nielsen, O. and Shepard, N. (2006). Econometrics of testing for jumps in financial economics using bipower variation. *J. Financ. Economet.*
- Bissiri, P., Holmes, C., and Walker, S. G. (2016). A general framework for updating belief distributions. *J. Roy. Statist. Soc. Ser. B*, to appear; [arXiv:1306.6430](https://arxiv.org/abs/1306.6430).
- Carr, P. and Wu, L. (2003). What type of process underlies options? a simple robust test. *J. Finance*, 58(6):2581–2610.

- Figuroa-López, J. E. (2009). Nonparametric estimation of Lévy models based on discrete-sampling. In *Optimality*, volume 57 of *IMS Lecture Notes Monogr. Ser.*, pages 117–146. Inst. Math. Statist., Beachwood, OH.
- Grünwald, P. and van Ommen, T. (2016). Inconsistency of bayesian inference for misspecified linear models and a proposal for repairing it. Unpublished manuscript, [arXiv:1412:3730](https://arxiv.org/abs/1412.3730).
- Jiang, W. and Tanner, M. A. (2008). Gibbs posterior for variable selection in high-dimensional classification and data mining. *Ann. Statist.*, 36(5):2207–2231.
- Kleijn, B. J. K. and van der Vaart, A. W. (2012). The Bernstein-Von-Mises theorem under misspecification. *Electron. J. Stat.*, 6:354–381.
- Marin, J.-M., Pudlo, P., Robert, C. P., and Ryder, R. J. (2012). Approximate Bayesian computational methods. *Stat. Comput.*, 22(6):1167–1180.
- Medvedev, A. and Scaillet, O. (2007). Approximation and calibration of short-term implied volatilities under jump-diffusion stochastic volatility. *Rev. Financ. Stud.*, 20(2):427–459.
- Musiela, M. and Rutkowski, M. (2005). *Martingale Methods in Financial Modelling*, volume 36 of *Stochastic Modelling and Applied Probability*. Springer-Verlag, Berlin, second edition.
- Podolskij, M. (2006). *New theory on estimation of integrated volatility with applications*. PhD thesis, Ruhr-Universität Bochum.
- Syring, N. and Martin, R. (2015). Scaling the Gibbs posterior credible regions. Unpublished manuscript, [arXiv:1509.00922](https://arxiv.org/abs/1509.00922).
- van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press, Cambridge.
- Zhang, T. (2006a). From  $\epsilon$ -entropy to KL-entropy: analysis of minimum information complexity density estimation. *Ann. Statist.*, 34(5):2180–2210.
- Zhang, T. (2006b). Information theoretical upper and lower bounds for statistical estimation. *IEEE Trans. Inform. Theory*, 52(4):1307–1321.