Asymptotic Theory for Renewal Based High-Frequency Volatility Estimation*

Yifan Li† Ingmar Nolte‡ Sandra Nolte§

Lancaster University Lancaster University Lancaster University

This Version: June 21, 2018

Abstract

We propose a novel class of volatility estimators named the renewal based volatility estimators for high-frequency volatility estimation, which is constructed based on a renewal process in business time. We show the consistency and derive the asymptotic distribution of this class of estimators, and show that a parametric structure can lead to significant gains in the efficiency of volatility estimation compared to a pure non-parametric design. This class of estimators includes all the parametric and non-parametric estimators that are based on an absolute price change point process, e.g. Engle and Russell (1998), Gerhard and Hautsch (2002), Tse and Yang (2012) and Nolte, Taylor, and Zhao (2018). We examine the non-parametric duration (NPD) based volatility estimator proposed by Nolte, Taylor, and Zhao (2018), and show the properties of this estimator in the presence of drift, jump, time discretization, a general market microstructure (MMS) noise and price discretization noise. The main finding is that the NPD estimator is very robust to the presence of jumps but is generally biased due to time discretization and the MMS noise. Through simulations we show that the NPD estimator is more efficient than calendar time sampled RV-type estimators in the absence of MMS noise, but also that it is much more sensitive to the MMS noise than calendar time sampling methods. We propose an exponentially smoothed NPD estimator and show that it can significantly outperform commonly used calendar time bias corrected volatility estimators in terms of efficiency. Additionally, we propose a range-duration based renewal type volatility estimator that can outperform a general realized variance (RV) estimator under any sampling scheme.

JEL classification: C02, C13, C14, C51

Keywords: High-Frequency Volatility Estimation, Renewal Theory.

*We are grateful to Christian Gouriéroux, Per Mykland, Mark Podolskij, Stephen Taylor, Bezirgen Veliyev and Ye Zeng for their constructive comments. We thank the participants of the 11th International Conference on Computational and Financial Econometrics, 2017, the joint Finance-Econometrics workshop at CREATES, 2018, the 1st Quantitative Finance and Financial econometrics conference, 2018 and the 11th Annual Conference Society of Financial Econometrics, 2018 for their constructive comments. The usual disclaimer applies.

†Corresponding author: Lancaster University Management School, Bailrigg, Lancaster, LA1 4YX, UK. e-mail: Y.Li26@lancaster.ac.uk.

‡Lancaster University Management School Bailrigg, Lancaster LA1 4YX, UK. Phone +44 15245 92644, email: I.Nolte@lancaster.ac.uk

§Lancaster University Management School, Bailrigg, Lancaster, LA1 4YX, UK. Phone +44 152 459 3634, e-mail: S.Nolte@lancaster.ac.uk.

1 Introduction

Since the seminal paper by Engle and Russell (1998), a point process based high-frequency volatility estimator provides an important alternative to the Realized Volatility (RV)-type estimator as popularized by Andersen, Bollerslev, Diebold, and Labys (2001). The main argument supporting the point process based volatility estimator is its parametric structure and ability to provide intraday inference on local volatility, as opposed to an integrated volatility estimator from the RV estimator. The quality of volatility estimates from point process based estimators has be verified by Tse and Yang (2012) and Nolte, Taylor, and Zhao (2018). In these papers, Tse and Yang (2012) show that the volatility estimates from fitting an Autoregressive Conditional Duration (ACD) Engle and Russell (1998) to the absolute price change point process can outperform RV-type estimators under the assumption of various stochastic volatility models. With the same volatility estimator, Nolte, Taylor, and Zhao (2018) show that volatility estimates from the point process can provide better predictability compared to those from the RV and RV variants. Despite these promising results showing a clear advantage of the point process based volatility estimators over the RV-type estimators, its theoretical properties have not yet been established.

Closely linked to the parametric point process based volatility estimator, Andersen, Dobrev, and Schaumburg (2008) and Nolte, Taylor, and Zhao (2018) propose two different non-parametric volatility estimators that use the price duration, that is, the time for the cumulative price change to surpass a given threshold, as a measure of volatility. They demonstrate that the duration-based volatility estimator can easily outperform the RV-type estimator in ideal conditions with a smaller mean squared error (MSE). Much of the theoretical properties of these non-parametric estimators have been discussed in these papers respectively, but none of them generalize the properties of these non-parametric estimators to a setting where both time-varying volatility and a general market microstructure noise (MMS) are present. Moreover, the duration based approach suffers from a truncation bias, when the price change is not exactly the value of the threshold. Together with the market microstructure noise, the consistency and asymptotic behaviour of these non-parametric estimators are largely unknown, which greatly hinders their applications in empirical studies.

We propose a general class of volatility estimators that we will refer to as the Renewal Based Volatility (RBV) estimators, which provides a theoretical framework for the aforementioned point process based volatility estimators (with the exception of the estimators in Andersen, Dobrev, and Schaumburg (2008)). This class of volatility estimators are constructed based on a renewal process in business time, which is a time change that treats the integrated variance as a measure of time. Based on this renewal process and the fact that the counts of events are shared by both business time and calendar time, we can construct an estimator that estimates the time elapse in business time, which corresponds to the integrated variance in calendar time. As we do not require any knowledge about the dynamics of the volatility process, this estimator is a non-parametric estimator. Moreover, we show that, by specifying a dynamic structure on the observed point process in the calendar time and defining a link function that maps the durations in calendar time to its counterparts in business time, one can construct parametric RBV-class estimators that can achieve a higher efficiency than their non-parametric counterparts. This includes the parametric duration-based volatility estimator as in Engle and Russell (1998), Tse and Yang (2012) and Nolte, Taylor, and Zhao (2018), and the intensity-based volatility estimator Gerhard and Hautsch (2002); Li, Nolte, and Nolte (2018). We derive the asymptotic distribution of both the non-parametric and parametric RBV estimators, and show that they are both unbiased and consistent as long as one can construct a renewal process in business time. One desirable property of this class of estimators is that, the asymptotic variance is a function of the asymptotic mean, so one does not need to estimate the asymptotic variance separately to construct confidence bounds (such as the estimation of integrated quarticity in the RV framework, see

e.g. Barndorff-Nielsen and Shephard (2004)).

We examine Nolte, Taylor, and Zhao's (Nolte, Taylor, and Zhao, 2018) non-parametric duration-based volatility estimator (NPD) in the RBV framework as a complement to our theoretical discussion. We formalize the properties of the NPD estimator for a general semimartingale setting in the presence of jumps, time-varying volatility, irregular arrivals of observations, price discretization and a general MMS noise. Our finding suggests that, firstly, the NPD estimator is more robust to jumps than a realized bipower variation estimator. Secondly, although the NPD estimator has a smaller asymptotic variance than the calendar time RV-type estimators in the absence of noise, it is very sensitive to MMS noise. Consequently, the NPD estimator will be biased upwards more heavily compared to a calendar time RV-type estimators of similar sampling frequency, which significantly weakens its relative performance.

By correcting the biases for the NPD estimator and exploiting its smaller asymptotic variance, we propose to construct the NPD estimator on the exponentially smoothed price process, which we will refer to as the exponentially smoothed NPD estimator, denoted by NPD^z. In our simulation we show that, if the smoothing parameter is chosen optimally, the truncation bias due to time discretization can approximately offset the smoothed MMS noise bias at moderate to large sampling frequencies. At these sampling frequencies, the NPD^z estimator exhibits a significantly higher efficiency compared to the commonly used bias corrected calendar time sampling volatility estimators, including the Realized Kernel Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008), the pre-averaged RV and pre-averaged bipower variation Hautsch and Podolskij (2013) estimators. Additionally, we demonstrate that, although the optimal sampling frequency of the NPD^z estimator is much smaller than its calendar time competitors, its optimal efficiency is still better than the optimal performance from its competitors, which requires a much larger sampling frequency.

The main contributions of this paper are three-folded: Firstly, we develop a theoretical framework on which the asymptotic properties of the aforementioned point process based volatility estimators can be derived. Specifically, we show that, the duration-based volatility estimator is indeed superior to RV-type estimators in ideal conditions, and a parametric structure can lead to a substantial increase in the efficiency of volatility estimation. Secondly, we propose a range-duration based estimator that in theory is more efficient than any RV estimator under a stochastic sampling scheme discussed in Fukasawa (2010b) and Fukasawa and Rosenbaum (2012). However, the properties of this estimator in a more general setup is yet to be verified. Finally, we evaluate the theoretical properties of the non-parametric duration-based volatility estimator under a very general model. We propose the exponentially smoothed NPD estimator which shows a clear efficiency advantage over the commonly used bias corrected calendar time sampling volatility estimators.

The rest of the paper is structured as follows: Section 2 describes the general theory for the renewal process and renewal reward process. Section 3 and 4 introduces the renewal based volatility estimator and the parametric renewal based volatility estimator respectively. Section 5 gives some examples on both the non-parametric and parametric estimators that belong to the class of renewal based estimators. In Section 6, we examine the NPD estimator under a general semi-martingale in the presence of various market imperfections. We conduct a Monte Carlo simulation study in Section 7. Section 8 concludes.

2 Prerequisites: Renewal Theory

This section summarizes the related renewal theory used in constructing the renewal based volatility estimator. For a more comprehensive discussion, please refer to standard point process textbooks, e.g. Wolff (1989), Ross (1996), etc.

We start with the definition of a renewal process:

Definition 1. Renewal Process: *Let $\{D_i\}_{i=1,2,\cdots}$ be a sequence of positive i.i.d. random variables with $0 < \mu = \mathrm{E}[D_i] < \infty$ which represents the inter-event arrival time, and let t_i denote the arrival time of the i-th event (renewal epoch) given by:*

$$t_i = \sum_{j=1}^{i} D_j. \tag{1}$$

A renewal process $X(t)$ is defined as a random variable that counts the number of event arrivals in the interval $(0, t]$:

$$X(t) \equiv \sum_{i=1}^{\infty} \mathbb{1}_{\{t_i \leq t\}}. \tag{2}$$

A renewal process has the following asymptotic properties:

Theorem 1. Elemental Renewal Theorem: *Let $X = \{X(t)\}_{t\geq 0}$ be a renewal process with mean inter-arrival time $0 < \mu < \infty$ and renewal function $m(t) = \mathrm{E}[X(t)]$, then*

$$\lim_{t\to\infty} \frac{X(t)}{t} \overset{a.s.}{\to} \frac{1}{\mu}, \tag{3}$$

$$\lim_{t\to\infty} \frac{m(t)}{t} \to \frac{1}{\mu}. \tag{4}$$

Proof. See e.g. Feller (1941), Doob (1948) Theorem 3.3.4, Chapter 3 in Ross (1996). □

A seemingly trivial result from the above theorem is that for a given $0 < \mu < \infty$, $\lim_{t\to\infty} X(t) \to \infty$. The renewal function, $m(t) = \mathrm{E}[X(t)]$, has the following second order asymptotic expansion as $t \to \infty$:

Proposition 1. *Let $X(t)$ be a renewal process defined in Definition 1 with mean and variance of the inter-event arrival time denoted as $0 < \mu < \infty$ and $0 < \sigma^2 < \infty$ respectively. Let $m(t) = \mathrm{E}[X(t)]$ denote the renewal function. The process $m(t)$ has the following asymptotic expansion as $t \to \infty$:*

$$m(t) = \frac{t}{\mu} + \frac{\sigma^2}{2\mu^2} - 0.5 + o(1). \tag{5}$$

Proof. E.g. Corollary 3.4.7 in Ross (1996) □

It is useful to consider the distribution of time elapses since the last renewal epoch. This is known as the age process of a renewal process, formally defined as follows:

Definition 2. Age Process of A Renewal Process: *Let $X(t)$ denote a renewal process defined in Definition 1. The age process of a renewal process is defined as:*

$$A(t) = t - t_{X(t)} \tag{6}$$

The moments of $A(t)$ can be derived from the moments of the renewal process if they exist:

Theorem 2. *For an age process $A(t)$ defined in Definition 2, and let the n-th moments of the inter-epoch duration of the underlying be denoted by $\mathrm{E}[D_i^n] = \mu_n$. Provided that all μ_n exist, the moments of the age process $A(t)$ can be expressed as:*

$$\mathrm{E}[A^n(t)] = \frac{\mu_{n+1}}{(n+1)\mu} \tag{7}$$

Proof. See, e.g. Coleman (1982). □

We will also use the property of a renewal reward process, which is defined as follows:

Definition 3. Renewal Reward Process: *Let $X(t)$ denote a renewal process with i.i.d. inter-event duration $\{D_i\}_{i=1,2,\cdots}$ that has mean $\mu < \infty$ and variance $\sigma^2 < \infty$. Let $\{R_i\}_{i=1,2,\cdots}$ denote a sequence of real-valued i.i.d. random variables with mean $v < \infty$ and variance $\sigma_r^2 < \infty$ associated with each D_i. Then the renewal reward process is defined as:*

$$R(t) = \sum_{i=1}^{X(t)} R_i. \tag{8}$$

The expectation of this process, $r(t) = \mathrm{E}[R(t)]$, is defined as the reward function.

A renewal reward process has the following asymptotic properties:

Theorem 3. Renewal Reward Theorem: *For a renewal reward process defined in Definition 3, the following results hold:*

$$\lim_{t\to\infty} \frac{R(t)}{t} \overset{a.s.}{\to} \frac{v}{\mu}, \tag{9}$$

$$\lim_{t\to\infty} \frac{r(t)}{t} \to \frac{v}{\mu}. \tag{10}$$

Proof. E.g. Theorem 3.6.1, Chapter 3 in Ross (1996). □

3 Renewal Based Volatility Estimator

We are now in the position of constructing the renewal based volatility (RBV) estimator for financial price processes. We start with an assumption about the price process and the associated volatility process of interest:

Assumption 1. Price Processes: *Let the price process $\{P(t)\}_{t>0}$ be a stochastic process with an adapted, càdlàg and strictly positive integrated variance (IV) process defined by $IV(0,t) = \int_0^t \sigma_p^2(s)ds$ with $IV(0,t) \to \infty$ as $t \to \infty$. We define a time change $t \mapsto \tau(t)$ where $\tau(t) = IV(0,t)$ that converts the calendar time to the integrated variation time, which is also known as the business time. We assume that the time changed price process $\tilde{P}(\tau(t)) = P(t)$ is a Lévy process in business time.*

Note that we can reverse the time change $t \mapsto \tau(t)$ by using $t = \inf\{u \in \mathbb{R}^+ : IV(0,u) \geq \tau(t)\}$. It is clear that $\tau(t)$ is a stopping time for any t. Also, the time changed information set has the relationship $\mathcal{F}_t = \tilde{\mathcal{F}}_{\tau(t)}$. For a more rigorous discussion on the change of time method, please refer to Chapter 1 in Barndorff-Nielsen and Shiryaev (2010).

Assumption 1 may seem strict, but it is satisfied by a wide range of stochastic processes that are used in modelling financial price processes. We give two simple examples.

Example 1: Any continuous local martingale satisfies this assumption due to to following theorem.

Theorem 4. (Dambis-Dubin Schwarz): *Let $(M(t))_{t\geq 0}$ be a continuous $\mathcal{F}_t$-local martingale such that its quadratic variation $\langle M\rangle_\infty = +\infty$, then there exists a Brownian motion $(B(t))_{t\geq 0}$, such that for every $t \geq 0$, $M(t) = B(\langle M\rangle_t)$.*

Since the quadratic variation and integrated variance of $M(t)$ coincide, the resulting Lévy process in business time is a standard Wiener process. Note that Theorem 4 still holds when the stochastic volatility and the price process are correlated, which is known as the 'leverage effect' that is commonly observed in practice (see e.g. Bollerslev, Litvinova, and Tauchen (2006)).

Example 2: A (inhomogeneous) compounded Poisson process as in Oomen (2005) satisfies this assumption. The resulting Lévy process is a homogeneous compounded Poisson process. See Appendix A for details.[1]

The connection from the Lèvy process in business time and the renewal theory in the previous section is established by the following proposition:

Proposition 2. *Let $\{Y(t)\}_{t\geq 0}$ be a Lévy process on the filtered probability space $\{\Omega, \mathcal{F}, P\}$. Define a stopping time process that automatically renews once stopped as:*

$$t_i = \inf_{t\geq t_{i-1}} \{Y(t) \in \mathcal{S}(t_{i-1})\}, \tag{11}$$

in which $\mathcal{S}(t_i)$ is the stopping condition for t_i as a function of $\mathcal{F}_{t_i}$. If, for any i, j and $t > 0$, $\text{Prob}(Y(t_i + t) \in \mathcal{S}(t_i)) = \text{Prob}(Y(t_j + t) \in \mathcal{S}(t_j))$, then the sequence $\{t_i\}_{i=1,2,\cdots}$ corresponds to arrivals of a renewal process.

Proof. The condition $\text{Prob}(Y(t_i + t) \in \mathcal{S}(t_i)) = \text{Prob}(Y(t_j + t) \in \mathcal{S}(t_j))$ ensures that the stopping condition is equivalent to the paths of the Lévy process originating from all the possible starting points $t_i \in (0, \infty)$ regardless of when the previous event occurred. Then clearly the distribution of $t_i - t_{i-1}$ is i.i.d., which follows from the property of the Lévy process. As a result, $\{t_i\}_{i=1,2,\cdots}$ is by definition a renewal process. □

Consequently, when the price process $P(t)$ follows Assumption 1, we can obtain a Lèvy process $\tilde{P}(\tau(t))$ in business time. According to Proposition 2, we can construct a renewal process $\{\tau(t_i)\}_{i=1,2,\cdots}$ in business time by choosing an appropriate $\mathcal{S}(\tau(t_i))$ for each i. Effectively, we sample the price process at $\{t_i\}_{i=1,2,\cdots}$ in calendar time in such way that the business time counterpart $\{\tau(t_i)\}_{i=1,2,\cdots}$ is renewal. We therefore refer to this sampling scheme as renewal sampling:

Definition 4. Renewal Sampling: *For a price process $P(t)$ satisfying Assumption 1, a renewal sampling scheme samples $P(t)$ at $0 < t_1 < t_2 < \cdots$ where the arrivals in business time $\{\tau(t_i)\}_{i=1,2\cdots}$ is a renewal process in business time. Denote the unobservable renewal process in business time as $\tilde{X}(\tau(t)) = \sum_{i>0} \mathbb{1}_{\{\tau(t_i)\leq\tau(t)\}}$ and its observable calendar counterpart as $X(t) = \sum_{i>0} \mathbb{1}_{\{t_i\leq t\}}$.*

Note that the càdlàg property of the integrated variance guarantees that $X(t) = \tilde{X}(\tau(t))$. Using Proposition 2, we can construct $X(t)$ in calendar time if the stopping condition in calendar time is only a function of the paths of $P(t)$, but not a function of time. Heuristically, by observing the path of the price process in calendar time, we can decide where to 'stop' the price process and obtain a sample. If the condition in Proposition 2 for $\mathcal{S}(t_i)$ is satisfied, then the stopping times in business time is by construction a renewal process.

The central contribution of this paper is the following novel volatility estimator by sampling the price process $P(t)$ with a renewal sampling scheme:

[1] Relying on Theorem 4, we can account for the leverage effect if the latent price process follows a continuous local martingale. However, it is not clear if this claim is still valid in the case of this example, or other alternative specifications.

Definition 5. Renewal Based Volatility Estimator: *Let $\{P(t)\}_{t>0}$ be a price process that satisfies Assumption 1. Choose a $\mathcal{S}(t_i)$ according to Proposition 2, and apply renewal sampling on $\tilde{P}(\tau(t))$ to obtain the renewal sampling times $\{t_i\}_{i=1,2,\cdots}$ and the point process $X(t) = \sum_{i>0} \mathbb{1}_{\{t_i \leq t\}}$, which has a business time counterpart $\tilde{X}(\tau(t))$ that is a renewal process. Let $0 < \mu < \infty$ and $0 < \sigma^2 < \infty$ denote the first two moments of the inter-epoch duration in business time, then the Renewal Based Volatility estimator is defined by:*

$$RBV(0,t) = X(t)\mu. \tag{12}$$

The RBV estimator has the following asymptotic distribution:

Theorem 5. *The Renewal Based Volatility estimator as defined in Definition 5 has the following asymptotic distribution:*

$$\lim_{t\to\infty} \frac{RBV(0,t) - IV(0,t)}{\sqrt{X(t)}\sigma} \xrightarrow{d} \mathcal{N}(0,1) \tag{13}$$

Proof. See Appendix B. □

One remark on the RBV estimator is that we can compute standard errors of the estimator without estimating the integrated quarticity as in the RV literature, which implies less estimation bias for the standard errors and confidence bounds. Similar to the RV-type estimators, the RBV estimator does not require any parametric assumption on the IV process in calendar time. The obvious problem here is that μ is not explicitly specified, and is dependent on the assumption of $P(t)$ and the stopping condition $\mathcal{S}(t_i)$. In Section 5 we show that in some special cases μ is available in closed form. Also, the process $\tilde{P}(\tau(t))$ is usually very simple (for example, a standard Wiener process). In this case the moments of the renewal process can be simulated easily.

We would like to point out that the limiting distribution of the RBV estimator is obtained when μ is fixed and $t \to \infty$. This is known as the sprawl asymptotics which is typical in the context of point processes. However, this is different to the infill asymptotics usually applied in the RV context where we have a fixed time frame and let the sampling frequency go to infinity. To derive counterpart of Theorem 5 in the infill asymptotics setting, more assumptions are required for the asymptotic behaviour of $P(t)$ and $X(t)$, which is presented in Appendix C. For the rest of the paper we will confine ourselves to the sprawl asymptotics setting because it requires less assumptions about the price process, and corresponding asymptotic results for the infill case can always be derived with additional assumptions in a similar fashion as in Appendix C. The sprawl asymptotics is also more relevant to the parametric RBV estimator, which we elaborate below.

4 Parametric Renewal Based Volatility Estimator

The duration in business time $\tilde{D}_i$ is not directly observable, but we can observe its calendar time counterpart D_i. Using the fact that $\tilde{D}_i$ is i.i.d., the connection between D_i and the integrated variance process is that:

$$\int_{t_{i-1}}^{t_{i-1}+D_i} \sigma_p^2(s)ds = \tilde{D}_i. \tag{14}$$

If we can specify a parametric model $g(t|\mathcal{F}_t)$ that uses all the information available in such a way that the following variable is i.i.d:

$$R_i = \int_{t_{i-1}}^{t_{i-1}+D_i} g(s|\mathcal{F}_s)ds, \tag{15}$$

then we can use the quantity $\frac{R_i\mu}{\mathrm{E}[R_i]}$ as an estimator for $\tilde{D}_i$. Without any loss of generality we set $\mathrm{E}[R_i] = \mu$ to simplify notation. We will refer to this estimator as the parametric renewal based volatility (PRBV) estimator, formally defined as follows:

Definition 6. Parametric Renewal Based Volatility Estimator: *Let $\{P(t)\}_{t>0}$, $X(t)$, $\tilde{X}(\tau(t))$, μ and σ^2 be defined identically to Definition 5. Define a parametric model $g(t|\mathcal{F}_t)$ and an i.i.d. variable R_i that follows* (15) *with $0 < \mathrm{E}[R_i] = \mu < \infty$ and $0 < \mathrm{V}[R_i] = \sigma_r^2 < \infty$. Then the parametric renewal based volatility (PRBV) estimator is defined as:*

$$PRBV(0,t) = \sum_{i=1}^{X(t)} R_i = \int_0^t g(s|\mathcal{F}_s)ds. \tag{16}$$

Recall that the RBV estimator is already consistent, therefore for any i.i.d. R_i with finite moments, the $PRBV$ estimator will still be consistent. However, the randomness in R_i may introduce extra noise in the $PRBV$ estimator, unless there exists a substantial amount of positive correlation between R_i and $\tilde{D}_i$, which requires that $g(t|\mathcal{F}_t)$ is a good proxy of $\sigma_p^2(t)$ for all t. Thus, we can assess the efficiency of the $PRBV$ estimator by using the RBV estimator as a benchmark.

Conditioning on that we can observe the i.i.d. variable R_i, the asymptotic distribution of the $PRBV$ estimator can be derived analogously to the derivation of Theorem 5 noting that $R_i - \tilde{D}_i$ is a zero-mean i.i.d. variable, and the pair $\{R_i, \tilde{D}_i\}$ forms a renewal reward process.

Theorem 6. *The Parametric Renewal Based Volatility estimator as defined in Definition 6 has the following asymptotic distribution:*

$$\lim_{t\to\infty} \frac{PRBV(0,t) - IV(0,t)}{\sqrt{X(t)(\sigma^2 + \sigma_r^2 - 2\rho\sigma\sigma_r)}} \xrightarrow{d} \mathcal{N}(0,1) \tag{17}$$

where ρ is the correlation between R_i and $\tilde{D}_i$.

Proof. This follows similarly from the proof in Appendix B. Note that $\mathrm{V}[R_i - \tilde{D}_i] = \sigma^2 + \sigma_r^2 - 2\rho\sigma\sigma_r$. □

Note that the variance of $PRBV$ is zero when $R_i = \tilde{D}_i$ for all i, indicating that the $PRBV(0,t)$ can in theory be a perfect estimator for the integrated variance when R_i is known. The variance of the $PRBV$ estimator can be written as:

$$\mathrm{V}[PRBV(0,t)] = V[RBV(0,t)] + X(t)(\sigma_r^2 - 2\rho\sigma\sigma_r), \tag{18}$$

and as long as $\sigma_r^2 - 2\rho\sigma\sigma_r < 0$, that is, $\rho \in (\frac{\sigma_r}{2\sigma}, 1]$, the $PRBV$ estimator will always be more efficient than the RBV estimator. Obviously, the value of ρ is determined by the distance between $g(t|\mathcal{F}_t)$ and $\sigma_p^2(t)$, which is unfortunately model dependent.

We provide an example of $g(t|F_t)$ which allows us to examine ρ directly. Initially proposed by Gerhard and Hautsch (2002) derived from the instantaneous volatility estimator of Engle and Russell (1998), the conditional intensity process of $X(t)$ is used as a proxy of the instantaneous volatility. We define $g(t|F_t)$ as follows:

$$g(t|F_t) = \mu\lambda(t|\mathcal{F}_t), \tag{19}$$

where $\lambda(t|\mathcal{F}_t)$ is the $\mathcal{F}_t$-conditional intensity of the process $X(t)$ defined as:

$$\lambda(t|\mathcal{F}_t) \equiv \lim_{\Delta\downarrow 0} \frac{1}{\Delta} \mathrm{E}[X(t+\Delta) - X(t)|\mathcal{F}_t]. \tag{20}$$

The corresponding renewal reward variable R_i is then defined as:

$$R_i = \mu \int_{t_{i-1}}^{t_i} \lambda(s|\mathcal{F}_s)ds \equiv \mu\Lambda(t_{i-1}, t_i). \tag{21}$$

The i.i.d.-ness of R_i is guaranteed by the following theorem:

Theorem 7. Random Time Change Theorem (RTCT): *Let $X(t)$ be a simple point process adapted to a history $\mathcal{F}_t$ with bounded, strictly positive $\mathcal{F}_t$-conditional intensity $\lambda(t|\mathcal{F}_t)$ and $\mathcal{F}_t$-compensators $\Lambda(t) = \int_0^t \lambda(u|\mathcal{F}_u)du$ with $\Lambda(\infty) = \infty$ almost surely. Under the random time change $t \mapsto \Lambda(t)$, the transformed process*

$$\tilde{X}(t) = X(\Lambda^{-1}(t))$$

is a Poisson process with unit rate.

Proof. See the proof in Theorem 7.4I in Daley and Vere-Jones (2003), Brown and Nair (1988) and Bowsher (2007). □

Theorem 7 suggests that $\Lambda(t_{i-1}, t_i) \sim i.i.d.exp(1)$, so we have $R_i \sim i.i.d.exp(\mu^{-1})$. The mean and variance of R_i are $v = \mu$ and $\sigma_r^2 = \mu^2$ respectively. We derive the following important proposition that characterizes the relationship between the conditional intensity processes in calendar time and business time:

Proposition 3. *Let $X(t)$ be a simple point process with conditional intensity process $\lambda(t|\mathcal{F}_t)$, and let $t \mapsto \tau(t)$ be a change of time from calendar time to business time. The conditional intensity process $\tilde{\lambda}(\tau(t)|\tilde{\mathcal{F}}_{\tau(t)})$ of the time-changed point process $\tilde{X}(\tau(t))$ follows:*

$$\tilde{\lambda}(\tau(t)|\tilde{\mathcal{F}}_{\tau(t)})\sigma_p^2(t) = \lambda(t|\mathcal{F}_t). \tag{22}$$

Proof. See Appendix D. □

Proposition 3 has some very powerful implications that provide theoretical foundations for intensity and duration based volatility estimation.

Corollary 1. *In Definition 6 with $g(t|\mathcal{F}_t) = \mu\lambda(t|\mathcal{F}_t)$, the rank correlation between R_i and $\tilde{D}_i$ is 1. Additionally, if $\tilde{D}_i$ is i.i.d. exponentially distributed, then $g(t|\mathcal{F}_t) = \sigma_p^2(t)$.*

Proof. See Appendix E. □

Corollary 1 suggests that, firstly, the $PRBV$ estimator is likely to perform very well due to the monotonic non-linear relationship between R_i and $\tilde{D}_i$. Secondly, the optimal renewal sampling scheme for $g(t|F_t) = \mu\lambda(t|\mathcal{F}_t)$ is a homogeneous Poisson sampling scheme in business time. In this case, the conditional intensity of $X(t)$ in calendar time is proportional to the spot volatility, so that the conditional intensity is a perfect estimator of instantaneous volatility for all t. However, the assumption that $\tilde{D}_i$ is i.i.d. exponentially distributed requires further assumptions on the price process (e.g. the compounded Poisosn process in Oomen (2006)), which is not desirable.

Alternatively, we can also correct for the discrepancy between the density of $\tilde{D}_i$ and R_i:

Corollary 2. *In Definition 6 with $g(t|F_t) = \mu\lambda(t|\mathcal{F}_t)$, let $F_{\tilde{D}}(x)$ and $F_{\tilde{D}}^{-1}(x)$ denote the CDF of $\tilde{D}_i$ and its inverse correspondingly. The following relationship holds for all i:*

$$\tilde{D}_i = F_{\tilde{D}}^{-1}(1 - \exp(-R_i/\mu)). \tag{23}$$

Proof. This is straightforward from the proof in Appendix E. □

The expression $F_{\tilde{D}}^{-1}(1-\exp(-R_i/\mu))$ is effectively an exponential inverse probability integral transformation of $\tilde{D}_i$, which is a perfect estimator for the volatility between the two points $(t_{i-1}, t_i]$. However, this is a weaker result compared to Corollary 1 because in general $g(t|\mathcal{F}_t) \neq \sigma_p^2(t)$. In this case, inference based on R_i only reflects $\tilde{D}_i$ in expectation with $\mathrm{E}[R_i] = \mathrm{E}[\tilde{D}_i]$ with a rank correlation of 1, and $g(t|\mathcal{F}_t)$ does not estimate the actual spot volatility. From Proposition 3, one should use $g^*(t|\mathcal{F}_t) = \lambda(t|\mathcal{F}_t)/\tilde{\lambda}(\tau(t)|\mathcal{F}_t)$ to correctly estimate the spot volatility.

From above, it is clear that regardless of the value of μ, if we know the true conditional intensity process in calendar time, we in principle know the underlying integrated variance process. Therefore, in practice one does not need to sample at ultra high-frequency to improve the precision of the volatility estimates, which is the common approach in the RV literature. Instead, one only needs to append the estimation window of the econometric model of the conditional intensity process to obtain a more precise estimate of the conditional intensity, which in turn leads to a more precise estimate of $\tilde{D}_i$ for each i. This is in stark contrast with the RV-type estimators which relies heavily on the availability of data within the volatility estimation window, and justifies our sprawl asymptotics setting. We stress that this is a very important property of the $PRBV$ estimator that validate the intraday volatility estimates as in Engle and Russell (1998) and Tse and Yang (2012), and also renders the $PRBV$ estimator advantageous over the RV-type estimator in the situation where the availability of data is limited either because of a short time span or liquidity of the security.

To summarize our findings on the $PRBV$ estimators, we have shown that, it is possible to construct a $PRBV$ estimator as in Corollary 2 that always has zero variance if R_i is known. However, these properties are unlikely to hold in practice as we do not observe $g(t|\mathcal{F}_t)$ and have to use a model to estimate $\hat{g}(t|\mathcal{F}_t)$ and $\hat{R}_i$ instead. This will inevitably introduce estimation noise in the model, even if the specification of $g(t|\mathcal{F}_t)$ is correct. As this is more related to the properties of the econometric model used for the observed point process that deserves individual investigations, we will leave it for future research.

4.1 End-of-Sample Bias

In practice, we do not have data of infinite length, and the sample has to stop somewhere. Therefore, there will be a small End-of-Sample (EoS) bias for the renewal process when the last renewal epoch is before the end of the sample. The correction of this bias can be obtained from the second order asymptotic expansion of the renewal function as in Proposition 1. The EoS bias correction is:

$$EoS = 0.5\mu - \frac{\sigma^2}{2\mu}. \tag{24}$$

Therefore the bias correction is smaller than 0.5μ, and can even be negative when $\sigma^2 > \mu^2$. In theory one should always add this bias correction to the RBV and $PRBV$ estimator. Nevertheless, when $\frac{\sigma^2}{\mu} \to 0$ as $\mu \to 0$, we have $EoS \to 0$. In this case, when one selects a small μ to construct the RBV estimator, the EoS bias becomes negligible.

5 Some Examples

We give some concrete examples of RBV and $PRBV$ in this section and summarize their properties. Assume the efficient log-price follows a semi-martingale of the following form:

$$dP(t) = \alpha(t)dt + \sigma_p(t)dW(t), \tag{25}$$

where $\alpha(t)$ is a continuous $\mathcal{F}_t$-predictable process and $\sigma(t)$ is assumed to be càdlàg and strictly positive with $\int_0^t \sigma^2(s)ds \to \infty$ when $t \to \infty$. For now, we assume $\alpha(t) = 0$ and no discontinuities in the diffusion process for simplicity, and will discuss the effect of the drift term and jumps in the next section. The quantity of interest here is the integrated variance of the process over an interval $(0, T)$:

$$IV(0,T) = \int_0^T \sigma_p^2(s)ds. \tag{26}$$

Example 1: The first example of an RBV estimator, which will also be examined in detail in later sections, is the non-parametric duration-based (NPD) volatility estimator proposed by Nolte, Taylor, and Zhao (2018). We start by defining the absolute price change point process, firstly introduced by Engle and Russell (1998):

Definition 7. The Absolute Price Change Point Process: *The absolute price change point process* $\{t_i^{(\delta)}\}_{i=0,1,\cdots}$ *for an observed price process* $P(t)$ *and a given price change threshold* δ *is constructed as follows:*

1. *Set* $t_0^{(\delta)} = 0$ *and choose a threshold* δ*.*
2. *For* $i = 1, 2, \cdots$*, compute the first exit time,* $t_i^{(\delta)}$*, of* $P(t_{i-1}^{(\delta)})$ *through the double barrier* $[P(t_{i-1}^{(\delta)}) - \delta, P(t_{i-1}^{(\delta)}) + \delta]$ *as:*

$$t_i^{(\delta)} = \inf_{t > t_{i-1}^{(\delta)}} \{|P(t) - P(t_{i-1}^{(\delta)})| \geq \delta\}.$$

Iterate until the sample is depleted.

The arrivals of $t_i^{(\delta)}$ are referred to as price events. In the RBV framework, we can write $\mathcal{S}^{(\delta)}(t_i^{(\delta)}) = \{P(t_i^{(\delta)}) - \delta, P(t_i^{(\delta)}) + \delta\}$ and clearly it satisfies the condition in Proposition 2. Define the time change as $\tau(t) = \int_0^t \sigma^2(s)ds = IV(0,t)$, and $P(\tau(t))$ is a standard Brownian motion by Theorem 4. As a result from Theorem 2, under business time, $\{\tau(t_i^{(\delta)})\}_{i=1,2,\cdots}$ forms a renewal process, denoted by $X^{(\delta)}(\tau(t))$.

Let $D_i^{(\delta)} = t_i^{(\delta)} - t_{i-1}^{(\delta)}$ and $\tilde{D}_i^{(\delta)} = \tau(t_i^{(\delta)}) - \tau(t_{i-1}^{(\delta)})$ denote the duration under calendar time and business time respectively. Note that $\tilde{D}_i^{(\delta)}$ is the stopping time for a Wiener process (starting at zero) to exit a symmetric interval $[-\delta, \delta]$. We can retrieve its moments from its moment generating function (see Table 1 in Andersen, Dobrev, and Schaumburg (2008)). The first three moments are:

$$\mathrm{E}[\tilde{D}_i^{(\delta)}] = \delta^2, \quad \mathrm{E}[(\tilde{D}_i^{(\delta)})^2] = \frac{5}{3}\delta^4, \quad \mathrm{E}[(\tilde{D}_i^{(\delta)})^3] = \frac{61}{15}\delta^6. \tag{27}$$

The NPD estimator in Nolte, Taylor, and Zhao (2018) is of the following form:

$$NPD(0,t) = X^{(\delta)}(t)\delta^2 = X^{(\delta)}(t)\mu(\delta). \tag{28}$$

Note we use the notation $\mu(\delta)$ and $\sigma^2(\delta)$ to denote the mean and variance of the price duration in business time for some δ. Therefore it is clear that the NPD estimator belongs to the class of RBV estimators. The asymptotic distribution of the NPD estimator can be derived easily from (13):

$$\lim_{t\to\infty} \frac{NPD(0,t) - IV(0,t)}{\sqrt{\frac{2}{3}X^{(\delta)}(t)\delta^4}} \xrightarrow{d} \mathcal{N}(0,1) \tag{29}$$

Using the asymptotic relationship $\delta^2 = \frac{IV(0,t)}{X^{(\delta)}(t)}$, we see that $\mathrm{V}[NPD(0,t)] \to \frac{2IV(0,t)^2}{3X^{(\delta)}(t)}$. This suggests that, given a common sampling frequency, on average the NPD estimator will be more than six times as efficient as the RV sampled in calendar time, exactly six times as efficient as the RV sampled in business time, and more efficient than the RV under tick time sampling due to that $IV(0,t)^2 \leq IQ(0,t)$ from Jensen's inequality Fukasawa (2010a).

The efficiency gain from the RV estimator is not surprising. Since the NPD estimator uses information in the path of the prices, it effectively uses more data than the RV estimator under the same sampling frequency. Additionally, as discussed in Section C.1, the NPD estimator is both an RBV estimator and a renewal RV estimator. It achieves the optimal efficiency for the renewal RV estimators due to the fact that the kurtosis of the return is 1.

Example 2: Inspired by Christensen and Podolskij (2007) and Andersen, Dobrev, and Schaumburg (2008) and following the idea of the NPD estimator, we can also construct a range duration-based RBV-type volatility estimator. Let r denote a fixed range size, then the following sequence of stopping times forms a renewal process in business time:

$$t_i^{(r)} = \inf_{t > t_{i-1}^{(\delta)}} \{P(t) \in \mathcal{S}^{(r)}(t_i^{(r)})\}, \tag{30}$$

where $\mathcal{S}^{(r)}(t_i^{(r)}) = \{P(t) : \sup_{t_i^{(r)} < s < t} P(s) - \inf_{t_i^{(r)} < s < t} P(s) \geq r\}$. Similar to the NPD estimator, let $X^{(r)}(\tau(t))$ denote the renewal process under business time. The first three moments of $\tilde{D}_i^{(r)} = \tau(t_i^{(r)}) - \tau(t_{i-1}^{(r)})$ is as follows Andersen, Dobrev, and Schaumburg (2008)):

$$\mathrm{E}[\tilde{D}_i^{(r)}] = \frac{1}{2}r^2, \quad \mathrm{E}[(\tilde{D}_i^{(r)})^2] = \frac{1}{3}r^4, \quad \mathrm{E}[(\tilde{D}_i^{(r)})^3] = \frac{17}{60}r^6, \tag{31}$$

and the non-parametric range duration-based volatility (NPR) estimator is simply:

$$NPR(0,t) = \frac{1}{2}X^{(r)}r^2, \tag{32}$$

which has the following asymptotic distribution as $t \to \infty$:

$$\lim_{t\to\infty} \frac{NPR(0,t) - IV(0,t)}{\sqrt{\frac{1}{12}X^{(r)}(t)r^4}} \xrightarrow{d} \mathcal{N}(0,1) \tag{33}$$

Using the asymptotic relationship $r^2 = \frac{2IV(0,t)}{X^{(r)}(t)}$, we have $\mathrm{V}[NPR(0,t)] = \frac{IV^2(0,t)}{3X^{(r)}(t)}$. So the NPR estimator is twice as efficient as the NPD estimator for a common sampling frequency.

The efficiency gain of the range-based estimators compared to the RV-based estimators has been addressed by Christensen and Podolskij (2007) and Andersen, Dobrev, and Schaumburg (2008), as price ranges exploit both the supremum and infimum of the price process, which can measure volatility more precisely than using price changes. We would like to stress that the asymptotic variance of the NPR estimator is smaller than the asymptotic variance of a general RV estimator under any sampling scheme Fukasawa (2010b); Fukasawa and Rosenbaum (2012). With this NPR example, it is clear that the RBV-class of estimators are in essence different from the RV-type estimators.

Example 3: The parametric duration (intensity) based volatility estimator, initially proposed by Engle and Russell (1998) and further developed by Gerhard and Hautsch (2002), Tse and Yang (2012), Nolte, Taylor, and Zhao (2018) and Li, Nolte, and Nolte (2018) is an example of a $PRBV$ estimator. Specifically, it specifies the dynamics of $D_i^{(\delta)}$ with a fully parametric model (for example, the Autoregressive Conditional Duration model

by Engle and Russell (1998)), and defines

$$g^{(\delta)}(t|F_t) = \mu\lambda^{(\delta)}(t|\mathcal{F}_t) = \delta^2\lambda^{(\delta)}(t|\mathcal{F}_t), \tag{34}$$

in which $\lambda^{(\delta)}(t|\mathcal{F}_t)$ is the conditional intensity process of $X^{(\delta)}(t)$ defined in (20). Gerhard and Hautsch (2002) propose an instantaneous volatility estimator defined as $InsV^{(\delta)}(t) = g^{(\delta)}(t|F_t)$, and an estimator of the IV between the arrival of two price events can be constructed as follows:

$$R_i^{(\delta)} = \int_{t_{i-1}^{(\delta)}}^{t_i^{(\delta)}} g(s|\mathcal{F}_s)ds = \delta^2\Lambda_i^{(\delta)} \sim i.i.d.\ exp(\delta^{-2}), \tag{35}$$

with $\mathrm{E}[R_i^{(\delta)}] = \delta^2$ and $\mathrm{V}[R_i^{(\delta)}] = \delta^4$. As this quantity is i.i.d. from Theorem 7, the parametric duration (intensity) based (PD) estimator of the following form:

$$PD(0,t) = \sum_{i=1}^{X^{(\delta)}} R_i^{(\delta)}, \tag{36}$$

is by definition a $PRBV$-class estimator. The asymptotic properties of the $PRBV$ estimator discussed in Theorem (6) and Proposition 3 can be applied directly to derive the asymptotic distribution of the PD estimator:

$$\lim_{t\to\infty} \frac{PD(0,t) - IV(0,t)}{\sqrt{C \cdot X^{(\delta)(t)}\delta^4}} \xrightarrow{d} \mathcal{N}(0,1), \tag{37}$$

in which C is a constant which cannot be solved analytically. From Proposition 3, since $\tilde{D}_i^{(\delta)}$ can be easily simulated based on a Wiener process, we can simulate the constant C easily. Details of this simulation can be found in Appendix F. Based on 1000000 replications, we found that $C \approx 0.047$. Therefore, the asymptotic variance of the PD estimator is roughly one-twentieth of the NPD counterpart. It shows that, if the parametric model of $\lambda^{(\delta)}(t|\mathcal{F}_t)$ is well-specified, then there can be a substantial efficiency gain from the parametric estimation.

Based on the NPR estimator, we can construct a parametric range (PR) based volatility estimator by defining the renewal variable $R_i^{(r)}$ as:

$$R_i^{(r)} = 0.5r^2 \int_{t_{i-1}^{(r)}}^{t_i^{(r)}} \lambda^{(r)}(s|\mathcal{F}_s)ds. \tag{38}$$

The PR estimator is defined analogously to the PD estimator as:

$$PR(0,t) = \sum_{i=1}^{X^{(r)}} R_i^{(r)} = R^{(r)}(t), \tag{39}$$

From the simulation in Appendix F and the property of the $PRBV$ estimator, we can derive the asymptotic distribution of PD given $R_i^{(r)}$:

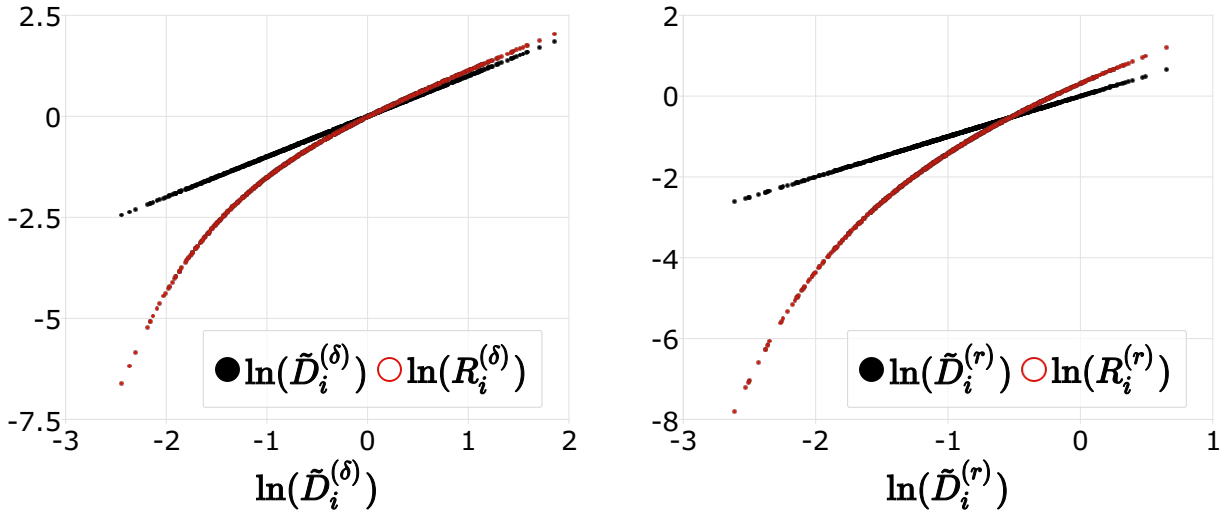
$$\lim_{t\to\infty} \frac{PR(0,t) - IV(0,t)}{\sqrt{C \cdot X^{(r)(t)}r^4}} \xrightarrow{d} \mathcal{N}(0,1), \tag{40}$$

where $C \approx 0.024$. Actually the PR estimator has a larger asymptotic variance than the PD estimator if we control for the same sampling frequency (when $r^2 = 2\delta^2$). This is due to the fact that the density $\tilde{D}_i^{(r)}$ deviates further from the exponential distribution.

We would like to stress that, although the PD and PR estimators for the integrated variance are unbiased and consistent in the sense of expectation, using $R_i^{(\cdot)}$ as an estimator of $\tilde{D}_i^{(\cdot)}$ will introduce a non-zero error due to the discrepancy between $R_i^{(\cdot)}$ and $\tilde{D}_i^{(\cdot)}$. We plot the simulated $\ln(\tilde{D}_i^{(\cdot)})$ against $\ln(R_i^{(\cdot)})$ in Figure 1. The

figure suggests that, as the discrepancy between $R_i^{(r)}$ and $\tilde{D}_i^{(r)}$ is larger than that of $R_i^{(\delta)}$ and $\tilde{D}_i^{(\delta)}$, the PR estimator will be less efficient compared to the PD estimator. Also, based on the simulated $\tilde{D}_i^{(\cdot)}$, one can correct this discrepancy by the method in Corollary 2. After the correction, both estimators will have zero variance conditioning on the knowledge of $R_i^{(\cdot)}$.

Figure 1: Discrepancy between the density of $R_i^{(\cdot)}$ and $\tilde{D}_i^{(\cdot)}$



Note: $N = 1000000$. Descriptive statistics of $\{\tilde{D}^{(\cdot)}\}_{i=1:N}$ and $\{R_i^{(\cdot)}\}_{i=1:N}$ can be seen in Table 3.

The results above suggest that, if the parametric model to estimate $R_i^{(\cdot)}$ performs equally well, then the two parametric estimators will have equal performance. This is in contrast to the efficiency difference between the NPD and NPR estimators as the NPD estimator is half as efficient as the NPR estimator. Because the variance of the reward variable $R_i^{(\cdot)}$ offsets the variance of $\tilde{D}_i^{(\cdot)}$ completely, the advantage of a lower variance for $\tilde{D}_i^{(r)}$ for the NPR estimator disappears. However, the PR estimator might be still preferred over the PD estimator because in a finite sample, one can obtain a larger sample size with range-based renewal sampling, resulting in more precise estimates for $R^{(\cdot)}$. Finally, as a result from Corollary 1, the instantaneous volatility estimator proposed in Gerhard and Hautsch (2002) does not hold for all t if the price process is assumed to follow (25), and can only serve as a proxy for the instantaneous volatility.

6 The Non-Parametric Duration-Based Volatility Estimator Under Market Frictions

This section discusses the theoretical properties of the NPD volatility estimator defined in (28) in the presence of drift, jumps, time discretization, market microstructure noise, and rounding effect.

6.1 Drift Effect

This section aims to clarify that the drift will not bias our estimator asymptotically. As the drift effect is very small in empirical high frequency applications, we will follow the approach by Barndorff-Nielsen and Shephard (2002) and discuss the drift effect in this section and assume it to be zero in other sections.

Firstly, as discussed in previous section, the NPD estimator is also a renewal RV estimator, and the quadratic variation theory can be applied. Therefore under the assumption of (25) and as $\delta \to 0$, the drift term will not bias our estimator. Moreover, we would like to note that there always exists a probability measure where the price process does not possess a drift by the use of Girsanov-Maruyama transformation. As the volatility remains unchanged after the change of measure and the NPD estimator can also be constructed on that probability measure, the presence of a drift is not a main concern in this paper.

6.2 Jump Effect

This section discusses the possible effect of jumps on the NPD estimator. The NPD estimator is by construction very robust to large jumps, as pointed out by Andersen, Dobrev, and Schaumburg (2008), Tse and Yang (2012) and Nolte, Taylor, and Zhao (2018), because of its truncation feature. For simplicity, we consider the following diffusion process with jumps:

$$P(t) = P(0) + \int_0^t \sigma_p(s) dW(s) + \sum_{j=1}^{J(t)} L_j, \tag{41}$$

where $J(t)$ is a counting process independent of $W(t)$, and L_j is the size of the j-th jump. We assume that $|L_j| \gg \delta$, so that each arrival of jump will almost surely trigger a price event. For a simple RV-type estimator under any sampling scheme, the $IV(0,t)$ estimates will be positively biased and include the jump variations. Let us consider the point process $\tilde{X}^{(\delta)}(\tau(t))$ under business time with $\tau(t) = IV(0,t)$. Denote the number of jumps in the duration $\tilde{D}_i^{(\delta)}$ by J_i, then conditional on that there is no jumps in the duration, the conditional mean and variance are δ^2 and $\frac{2}{3}\delta^4$ respectively. For the durations that contain a jump, it will always end the duration with the jump.

We can split the durations in business time by whether they contain a jump. For the durations that do not contain a jump, we have $\mathrm{E}[\tilde{D}_i^{(\delta)}|J_i = 0] = \delta^2$. For the durations that contain a jump, we interpret the jump as a random inspection time to a duration in business time, and the renewal process is immediately renewed when it is inspected. The time travelled on the business clock till the inspection time but before the actual price event would have occurred is therefore the length of the duration in business time. The density of the duration that contains a jump then can be interpreted as the age process of the renewal process defined in Definition 2. According to Theorem 2, we have:

$$\mathrm{E}[\tilde{D}_i^{(\delta)}|J_i = 1] = \frac{\mu^2(\delta) + \sigma^2(\delta)}{2\mu(\delta)} = \frac{5}{6}\delta^2 < \delta^2 \tag{42}$$

This suggests that each jump will on average shorten the distance travelled on the business clock by $\frac{1}{6}\delta^2$. For a total of $\tilde{X}^{(\delta)}(\tau(t))$ events in which $J(t)$ of them are jump-induced, the expected business time elapse is therefore:

$$IV(0,t) = \tilde{X}^{(\delta)}(\tau(t))\delta^2 - \frac{1}{6}\mathrm{E}[J(t)]\delta^2. \tag{43}$$

It is then clear that the bias introduced by a jump is just $\frac{1}{6}\mathrm{E}[J(t)]\delta^2$, which goes to zero as $\delta \to 0$ given a fixed number of jumps. It is interesting to see that the NPD estimator is less affected by jumps when $\delta \to 0$, in contrast to an RV estimator which is not robust to jumps at all regardless of the sampling scheme. As empirically price jumps are found to be very infrequent (on average less than one per week as documented in Andersen, Bollerslev, and Dobrev (2007) and Lee and Hannig (2010)), we can safely conclude that the estimator is robust to jumps in the limit and will ignore the jump component in the analysis hereafter. Note that following the same notion, the NPR estimator is also very robust to jumps.

6.3 A More Realistic Model

Real data does not follow the model specified in (25), as it possesses various type of market imperfections, including irregularly spaced observations, market microstructure noise, price discretization, etc. It greatly complicates the analysis of the theoretical properties of the NPD estimator as the properties of the RBV-class estimators may not apply in some cases. In this section we attempt to derive some asymptotic results for the NPD estimator under a general setting with random arrival times of observations and a very general structure of market microstructure noise.

Our strategy here is to add features to the pure diffusion model in (25). We firstly define the latent efficient log-price process as:

$$P^*(t) = P^*(0) + \int_0^t \sigma_p(t) dW(t). \tag{44}$$

To account for the random arrival of observations, we define a sequence of random arrival times of the tick changes[2] (or revisions for quote data) $0 = t_0 < t_1 < t_2 \cdots$, and assume that the process $P^*(t)$ is only observed at these random arrival times. The sequence $\{t_j\}$ and the arrival times in business time $\{\tau(t_j)\}$ with $\tau(t) = \int_0^t \sigma_p^2(s) ds$ are natural stopping times.

At each t_j, the observed process $P(t)$ is measured with noise V_j, commonly referred to as the MMS noise:

$$P_j = P_j^* + V_j \tag{45}$$

Whenever no confusion is caused, we suppress the notation of $P(t)$ as a function of calendar time and use $P_j = P(t_j)$ to denote the j-th observed price. We build our assumptions of the MMS noise based on the noise assumptions in Zhang (2006), Bandi and Russell (2008) and Aït-Sahalia, Mykland, and Zhang (2011):

Assumption 2. The Market Microstructure Noise: *The MMS noise component V_j in* (45) *is assumed to possess the following properties:*

1. *V_j is strictly stationary with mean 0 and density $f_V(\cdot)$.*
2. *All moments of V_j exist and are finite.*
3. *V_j is ρ-mixing.*
4. *$V_j \perp\!\!\!\perp P_j^*$.*

Note that conditions (2) and (3) can be replaced by other mixing conditions given that a corresponding version of the central limit theorem is available. We exclude the case where the noise is correlated with the efficient price movement as argued by Hansen and Lunde (2006). This is a common assumption in the existing literature mentioned above, and to a large extent simplifies our analysis.

The literature suggest that the trade durations in calendar time $d_j = t_j - t_{j-1}$ have seasonality patterns, are very persistent and are correlated with the volatility of the efficient price (e.g. Easley and O'Hara (1992), Chen, Diebold, and Schorfheide (2013)). However, we are more interested in the properties of the trade durations in business time denoted by $\tilde{d}_j = \tau(t_j) - \tau(t_{j-1})$, which are more relevant to our analysis. Since $P^*(\tau(t))$

[2]Note that the NPD estimator will always sample data in tick time, and we only consider the arrival of tick changes as the flat trades are irrelevant in the discussion.

is a standard Wiener process, by the martingale stopping theorem, P_j^* is a martingale, and the martingale difference sequence (MDS) $r_j^* = P_j^* - P_{j-1}^*$ is mixture normally distributed:

$$r_j^* \sim \mathcal{MN}(0, \tilde{d}_j). \tag{46}$$

It is therefore clear that the property of $\tilde{d}_j$ is embedded in the property of the tick returns of the efficient price r_j^*. We make the following assumption on the MDS process r_j^*:

Assumption 3. Tick Return of the Efficient Price Process: *The tick return of the efficient price process r_j^* is strongly mixing and strictly stationary with finite moments.*

The purpose of Assumption 3 here is to facilitate the CLT for the MDS sequence, which will be used to derive the asymptotic properties of the NPD estimator.

The NPD estimator is constructed by Definition 7 and (28) on the observed price process P_j with the following form:

$$NPD(0,t) = X^{(\delta)}(t)\delta^2, \tag{47}$$

where $X^{(\delta)}(t) = \sum_{i=1}^{\infty} \mathbb{1}_{\{t_i^{(\delta)} \le t\}}$ and $t_i^{(\delta)}$ is the arrival time of the i-th price event.

Deviations from the continuous martingale setting results in a biased NPD estimator, since the mean duration in business time is not δ^2 any more, and the point process in business time ceases to be renewal due to the existence of MMS noise. Fortunately the mixing assumptions for the MMS noise and the tick returns ensure that when δ is large enough, $\tilde{d}_j$ can be regarded as the stopping time from a Wiener process due to the functional central limit theorem via martingale approximation in e.g. Gordin and Peligrad (2011). In the following section we analyse the bias of the NPD estimator in detail based on our assumptions of the price process above, and show that the bias diminishes as δ increases.

6.4 Bias of the NPD estimator

To derive the bias of the NPD estimator in the presence of MMS noise and time discretization, we start from the renewal RV estimator based on $X^{(\delta)}(t)$:

$$RV^{(\delta)}(0,t) = \sum_{i=1}^{X^{(\delta)}(t)} (r_i^{(\delta)})^2 \tag{48}$$

Since the NPD estimator simply truncates $(r_i^{(\delta)})^2$ to δ^2, we must have that $(r_i^{(\delta)})^2 \ge \delta^2$. Intuitively, if there is no MMS noise, $RV^{(\delta)}$ would be unbiased, and the difference between the two estimators is the bias of the NPD estimator caused by time discretization. We will therefore use $Bias_{TD}^{(\delta)}(0,t) = NPD(0,t) - RV^{(\delta)}(0,t)$ to denote the time discretization bias of the NPD estimator, which is always negative.

In the presence of MMS noise, let us decompose $r_i^{(\delta)}$ as:

$$r_i^{(\delta)} = r_i^{(*,\delta)} + V_i^{(\delta)} - V_{i-1}^{(\delta)}, \tag{49}$$

in which $r_i^{(*,\delta)} = P^*(t_i^{(\delta)}) - P^*(t_{i-1}^{(\delta)})$ denotes the return of the efficient price. A well-established result (see e.g. Hansen and Lunde (2006), Bandi and Russell (2008)) of the RV estimator under autocorrelated noise is that:

$$\mathrm{E}[RV^{(\delta)}(0,t)] = IV(0,t) + \sum_{i=1}^{X^{(\delta)}(t)} \mathrm{E}[(V_i^{(\delta)} - V_{i-1}^{(\delta)})^2] \tag{50}$$

And therefore:

$$\mathrm{E}[NPD(0,t)] = IV(0,t) + \mathrm{E}[Bias^{(\delta)}_{MMS}(0,t)] + \mathrm{E}[Bias^{(\delta)}_{TD}(0,t)], \tag{51}$$

in which $Bias^{(\delta)}_{MMS}(0,t) = \sum_{i=1}^{X^{(\delta)}(t)}(V_i^{(\delta)} - V_{i-1}^{(\delta)})^2$ is the bias induced by the market microstructure noise, which is strictly positive. The above results suggest that the NPD estimator is generally biased with two source of bias: the truncation bias introduced by time discretization and the MMS noise bias.

From Assumption 2 we know that $V_i^{(\delta)}$ is asymptotically independent of $V_{i-1}^{(\delta)}$ if the number of observations within a duration tends to infinity. We therefore have:

$$\lim_{\delta\to\infty} \mathrm{E}[Bias^{(\delta)}_{MMS}(0,t)] \to 2\,\mathrm{E}[X^{(\delta)}(t)]\,\mathrm{V}[V_j], \tag{52}$$

which corresponds to the bias of the i.i.d. MMS noise discussed in Zhang, Mykland, and Aït-Sahalia (2005). As a result, $Bias^{(\delta)}_{MMS}$ decreases as we sample more sparsely, similar to the RV estimator.

For the TD bias, we are unable to derive an explicit expression in the general case. We show in Appendix G that an approximated version of $Bias^{(\delta)}_{TD}$ converges to zero in the absence of MMS noise with a rate of δ^{-1}. If we believe that $Bias^{(\delta)}_{TD}$ is of the order δ^{-1} in the general case, then $Bias^{(\delta)}_{TD}$ decays much slower than $Bias^{(\delta)}_{MMS}$ as $Bias^{(\delta)}_{MMS}$ is of the order δ^{-2}. Also, $Bias^{(\delta)}_{TD}$ will always bias the NPD estimator towards zero when $\delta \to 0$, which is due to the fact that $X^{\delta}(t)$ is capped at the number of observed tick changes. To give a graphical illustration of the bias of the NPD estimator under the two sources of biases, we simulate a simple price model and analyse the bias of the NPD estimator by adding the features to the price model. The results are present in Appendix H.

The discussion above also suggests that, the NPD estimator will be less biased compared to the $RV^{(\delta)}$ estimator if $Bias^{(\delta)}_{MMS}$ dominates $Bias^{(\delta)}_{TD}$, but will perform worse than the $RV^{(\delta)}$ estimator if there is no MMS noise at all. Interestingly, we may find cases where $Bias^{(\delta)}_{MMS}$ approximately offsets $Bias^{(\delta)}_{TD}$ when δ is large (as in Figure 13 in Appendix H for example). In this case the NPD estimator will have a bias close to zero and thus very efficient, although the $Bias^{(\delta)}_{MMS}$ is not zero. This suggests a potential bias correction technique if one can 'adjust' $Bias^{(\delta)}_{MMS}$ or $Bias^{(\delta)}_{TD}$ in a way that the two biases approximately cancels as $\delta \to \infty$. We will exploit this property in Section 6.6 to construct bias corrected NPD estimators.

6.5 Price Discretization

The observed price in practice is not continuously distributed, due to price discretization. The minimum allowed quote change is known as the tick size, which is typically 1 cent for securities in the US market that are traded above \$1. This is known as the round-off error discussed in the RV literature (see e.g. Delattre and Jacod (1997), Li and Mykland (2015) and the reference therein). We show, that this noise will also have a very special impact on the NPD estimator with simulation evidence in Appendix H.

We write the discretized return as $\mathring{r}_j = h_\epsilon(P_j) - h_\epsilon(P_{j-1})$, where $h_\epsilon(x)$ is a rounding function for the log price P_j. If we compare the discretized return $\mathring{r}_j$ and the return r_j without discretization, we have the following expression:

$$\begin{aligned} \mathring{r}_j &= r_j + \Xi_j - \Xi_{j-1} \\ \Xi_j &= h_\epsilon(P_j) - P_j, \end{aligned} \tag{53}$$

and Ξ_j is thus the price discretization error. Depending on the assumed rounding function $h_\epsilon(x)$, the theoretical property of Ξ_j will differ. To simplify our analysis, we use the rounding function: $h_\epsilon(x) = \epsilon\mathrm{nint}(\frac{x}{\epsilon})$ and $\mathrm{nint}(x)$

returns the nearest integer of x. This basically assumes that the price discretization is in the log scale, which is reasonable if the price level is roughly constant within the time period. We can interpret ϵ as the log tick size. The distribution of Ξ_j is then roughly identical for all j but can be potentially autocorrelated if ϵ is large, and we also have $\Xi_j \in (-0.5\epsilon, 0.5\epsilon)$. Thus we can regard Ξ_j as another noise term in the price process similar to the MMS noise component V_j and incorporate this in $Bias^{(\delta)}_{MMS}$.

The price discretization has a more profound impact on the sampling scheme. Specifically, when one takes δ to be between $(x\epsilon, (x+1)\epsilon]$ for some integer x, the resulting sampling scheme $X^{(\delta)}(t)$ will be exactly the same due to the discreteness in $\mathring{r}_j$. As a result, choosing multiple δ in the range $(x\epsilon, (x+1)\epsilon]$ does not effectively change the asymptotic property of $X^{(\delta)}(t)$. An implication of this is that one can influence the level of the truncation bias for a fixed sampling scheme $X^{(\delta)}(t)$ by choosing a δ within the range $((x-1)\epsilon, x\epsilon]$ for some integer x. When $Bias^{(\delta)}_{TD}$ dominates, we should always choose $\delta = x\epsilon$ to minimize $Bias^{(\delta)}_{TD}$. When $Bias^{(\delta)}_{MMS}$ is large, one can choose $\delta \to (x-1)\epsilon$ to inflate $Bias^{(\delta)}_{TD}$ and counterbalance the positive $Bias^{(\delta)}_{MMS}$. As is shown in Figure 14 in Appendix H, there can be a δ in the range of $((x-1)\epsilon, x\epsilon]$ that corrects the bias of the NPD estimator completely. However, this requires the knowledge of $Bias^{(\delta)}_{TD}$ at any δ, which can be very difficult to estimate empirically.

6.6 A Possible Bias Correction Method for the NPD Estimator

In this section we propose a bias correction method for the NPD estimator, and compare the performance of this bias correction method in a simulation study in Section 7 against some commonly used calendar time volatility estimators.

Inspired by the pre-averaging estimator in Jacod, Li, Mykland, Podolskij, and Vetter (2009), we propose to smooth the transaction price before constructing the NPD estimator. In detail, instead of constructing the NPD estimator based on the observed discrete price $h_\epsilon(P_j)$, we construct a smoothed price process Z_j, and construct the NPD estimator based on Z_j instead. We choose a simple exponential smoothing structure for the process Z_j:

$$\begin{aligned} Z_1 &= h_\epsilon(P_1) \\ Z_j &= (1-\gamma)Z_{j-1} + \gamma h_\epsilon(P_j), \quad \gamma \in [0,1] \end{aligned} \tag{54}$$

where γ is a smoothing factor. Clearly when $\gamma = 1$, $Z_j = h_\epsilon(P_j)$ so the process is not smoothed, and when $\gamma = 0$, $Z_j = h_\epsilon(P_1)$ for all j. Intuitively, the variation of the noise is diminished by this exponential smoothing to some extent, thus the NPD estimator constructed on Z_j is less affected by MMS noise. We will denote the exponentially smoothed NPD estimator constructed on Z_j as NPD^z.

The exponentially smoothed price process Z_j, is still contaminated by noise, albeit the magnitude of noise is reduced by the smoothing. Intuitively, the larger the γ, the larger the impact of MMS noise on the NPD^z estimator. Thus the exponential smoothing provides a way to alter the impact of MMS noise on the NPD^z estimator. As is shown in the previous section, if we can choose a γ so that $Bias^{(\delta)}_{MMS} + Bias^{(\delta)}_{TD}$ is approximately zero for some moderate to large δ, we can greatly improve the performance of the NPD estimator.

The price smoothing approach has two additional advantages over the original NPD estimator: firstly, it is a natural solution to the price discretization, and the sampling frequency will change more smoothly with respect to δ. Secondly, empirical data contains a very large amount of flat trades which will be completely ignored by the price change point process. By smoothing the price process, we can sample the data at every

transaction instead of every tick change, which greatly increases the maximum sampling frequency.

Nevertheless, in this paper we do not provide an analytical solution to choose γ optimally, as the truncation bias is not available in closed form. In practice, we can choose γ by benchmarking the NPD^z estimator on some unbiased volatility estimator and minimize the MSE, as documented in Hautsch and Podolskij (2013). Moreover, the estimator will be less robust to jumps compared to the NPD estimator, simply because the exponential smoothing distributes a jump to all previous transactions, which will have a larger chance to be absorbed into a price event. This is however not a significant problem if the jumps are assumed to be large and rare, so that the smoothed jumps still trigger price events as they occur.

7 Simulation Study

7.1 Simulation Design

We conduct a simulation study to demonstrate the properties of the price duration based volatility estimators (NPD and $RV^{(\delta)}$) and compare their performance to existing calendar time methods. We list all volatility estimators considered in this paper in Table 1.

Table 1: List of all volatility estimators considered in the simulation study

Acronym	Description	Type	MMS	Jump
NPD	See Section 5	δ	N	Y
$RV^{(\delta)}$	Renewal RV	δ	N	N
NPD^z	See Section 6.6	δ	N	Y
RV	Realized Variance	CTS	N	N
RBip	Realized Bipower Variation	CTS	N	Y
RK	Realized Kernel	CTS	Y	N
PRV	Pre-averaged Realized Variance	CTS	Y	N
PBip	Pre-averaged Bipwer Variation	CTS	Y	Y

Note: The column Type shows the type of sampling schemes: δ stands for the δ-associated price change point process sampling and CTS refers to calendar time sampling. The column MMS describes whether the estimator is robust in the presence of MMS noise, and the column Jump shows the robustness to jumps for the volatility estimators.

We consider a one-factor stochastic volatility (1FSV) model[3] with jumps to simulate the efficient price process, a model commonly used in this literature (see e.g. Huang and Tauchen (2005), Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008), etc.). The log-efficient price is specified as:

$$\begin{aligned} dP^*(t) &= \mu dt + \sigma_p(t) dW(t) + dJ(t), \quad \sigma_p(t) = \exp(\beta_0 + \beta_1 \tau_t) s_p(t) \\ d\tau(t) &= \alpha \tau(t) dt + dB(t), \quad corr(dW(t), dB(t)) = \varphi, \end{aligned} \tag{55}$$

in which $J(t) = \sum_{i=0}^{N_J(t)} J_i$ is a pure jump process. We assume that $N_J(t)$ is a homogeneous Poisson process with rate λ_J, and J_i is i.i.d. normal with zero mean and variance σ_J^2. Note that we augment the original 1FSV model by a time deterministic function $s_p(t)$ to accommodate the well-documented L or U-shaped pattern of intraday volatility. In the simulation study we set $t \in [0, 1]$ to represent fractions of time from a trading day from 9:30 to 16:00, and the process $\tau(t)$ is initialized by a random draw from its unconditional distribution. The

[3]We will use the subscript $1FSV_J$ to denote a 1FSV model with jump.

function $s_p(t)$ in our simulation study is specified as:

$$s_p(t) = \frac{1}{a_1 t + a_2} - \frac{\ln(\frac{a_1}{a_2} + 1)}{a_1} + 1, \quad a_1 > 0, a_2 > 0. \tag{56}$$

This function has the property that $\int_0^1 s(t)dt = 1$. When a_1 and a_2 are properly chosen, the function can produce a L-shaped pattern.

We build the MMS noise component in transaction time instead of calendar time. Specifically, we assume that the point process of transaction arrivals (or quote updates), denoted by $N(t)$, follows a inhomogeneous Poisson process where the intensity function $\lambda(t)$ is specified as a time-deterministic function to mimic the empirical U-shaped pattern of transaction arrivals. We specify the intensity function as:

$$\lambda(t) = \frac{1}{\Delta t}\Big(b_1(b_2 t - 1)^{b_3} - \frac{b_1(b_2 - 1)^{b_3+1} + 1}{b_2(b_3 + 1)} + \lambda_0\Big), \tag{57}$$

in which $b_1 > 0$, $b_2 > 0$ and $b_3 = 2, 4, \cdots$, λ_0 is the baseline arrival rate of transactions. The quantity Δt is the discretization step size of the simulation. The expected number of transactions in the interval $(0, 1)$ is therefore $\mathrm{E}[N(t)] = \int_0^1 \lambda(t)dt = \frac{\lambda_0}{\Delta t}$.

Let t_j denote the j-th arrival of transaction, and $P_j^* = P^*(t_j)$ denote the efficient price at the j-th transaction. Empirically we cannot observe P_j^* due to the presence of MMS noise, and the following decomposition is frequently used in the literature:

$$P_j = P_j^* + V_j, \tag{58}$$

in which V_j is a MMS noise term satisfying Assumption 2, and P_j is the log-price process measured with error. We assume that V_j follows an Gaussian AR(1) process specified as follows:

$$V_j = \rho V_{j-1} + v_j, \quad v_j \sim \mathcal{N}(0, (1 - \rho^2)\sigma_v^2) \tag{59}$$

For the sake of stationarity we require that $|\rho| < 1$. The unconditional variance of the noise is therefore $\mathrm{V}[V_j] = \sigma_v^2$.

Empirically, the transaction returns contain a large amount of flat trades where the transaction price do not move at all. For example, in Liesenfeld, Nolte, and Pohlmeier (2006), the proportion of flat trades for two stocks traded in NYSE is over 60%. Jacod, Li, and Zheng (2017) reports an over 70% of flat trades in the transaction data from Citigroup. For the mid-quote data the proportion of flat price changes will be even larger, as the best quotes can remain constant even when the transaction price moves. As a result, the empirical transaction returns are typically found to have excess kurtosis due to the amount of flat trades that cannot be reasonably explained by the normal assumption. To account for this effect, we follow the approach of Griffin and Oomen (2008) and assume that the tick change of price process is governed by a first order Markov chain. Let S_j be a stationary and recurrent two-state first order Markov chain with transitional parameters $P(S_{j+1} = n|S_j = m) = p_{mn}$ where $m, n \in \{0, 1\}$. We rewrite (58) as:

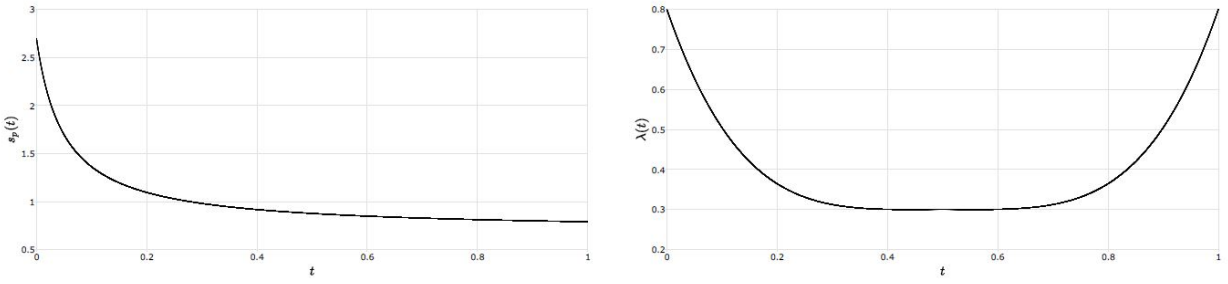
$$P_j = \begin{cases} P_j^* + V_j, & S_j = 1 \\ P_{j-1}, & S_j = 0 \end{cases} \tag{60}$$

Therefore, when $S_j = 1$, the observed price change is updated by the rounded efficient price process plus noise, and remains constant whenever $S_j = 0$.

The observed log-price process, $h_\epsilon(P_j)$ is specified as follows:

$$h_\epsilon(P_j) = \left(\epsilon \mathrm{nint}\Big(\frac{P_j}{\epsilon}\Big)\right), \tag{61}$$

Figure 2: Simulated diurnal pattern of intraday volatility and transaction arrival rate



Note: $s_p(t)$ is specified in (56) with $a_1 = 10$, $a_2 = 0.5$. $\lambda(t)$ is specified in (57) with $b_1 = 0.5$, $b_2 = 2$, $b_3 = 4$, $\lambda_0 = 0.4$ and $\Delta t = 1$. t is the fraction of time in a trading day.

in which nint(x) returns the nearest integer of a real number x. Note that the rounding will introduce additional flat trades to the observed price process when the price change is rounded to zero. We set $\epsilon = \ln(P_0+0.01)-\ln(P_0)$ to represent the log tick size.

The parameters for the 1FSV model we use are: $\mu = 0$, $\beta_0 = -4.3711$, $\beta_1 = 0.05934$, $\alpha = -0.011$, $\varphi = -0.3$, $\lambda_J = 2$, $a_1 = 10$, $a_2 = 0.5$, and $\sigma_J = 0.01$. The unconditional mean of the annualized daily volatility is roughly 27%, and the expected jump variation is about 0.0002 per day. The transaction and tick arrival parameters are set as $b_1 = 0.5$, $b_2 = 2$, $b_3 = 4$, $\lambda_0 = 0.4$, $p_{11} = 0.6$ and $p_{22} = 0.8$. We set the Euler step size of the simulation to be $\Delta t = \frac{1}{23400}$, so that the expected number of transactions within a trading day is 9360. The diurnal patterns of intraday volatility and the arrivals of transactions are plotted in Figure 2. From the figure we can clearly see that the intraday volatility has a L-shaped pattern and the arrivals of observations possess a U-shaped pattern. An example of a simulate price path of the 1FSV_J is presented in Figure 4.

For the MMS noise parameters, we set $\rho = -0.5$ and $\sigma_v^2 = \omega IV$, where ω is the noise-to-signal ratio. Empirically ω is found to be quite small (typically smaller than 0.1% as documented in Hansen and Lunde (2006)). We therefore choose $\omega = 0.005, 0.001$ and 0.0002 to represent high, moderate and low noise scenarios. The resulting $\sigma_v^2 \approx 0.00115^2, 0.00052^2$ and 0.00023^2 respectively. The expected number of flat trades implied by the Markov chain is about 67% of the total transactions. The actual amount of flat trades in $h_\epsilon(P_j)$ depends on the initial price $P(0)$, as the rounding error is smaller when $P(0)$ is large, and vice versa. We set $P(0) = 20$, and the resulting proportion of flat trades is approximately 70%. We plot a histogram and the correlogram for the simulated price change $h_\epsilon(P_j) - h_\epsilon(P_{j-1})$ in Figure 3 with $J(t) = 0$ and $\Delta t = \frac{1}{23400}$ for the moderate noise case. It is clear that the observed price change is leptokurtic with a sample kurtosis of approximately 15. This closely resembles the empirical density of the price changes as in Liesenfeld, Nolte, and Pohlmeier (2006). The autocorrelation for price changes suggests that the price changes follow an ARMA-type process with negative first order autocorrelation, which is consistent with the findings in e.g. Oomen (2006).

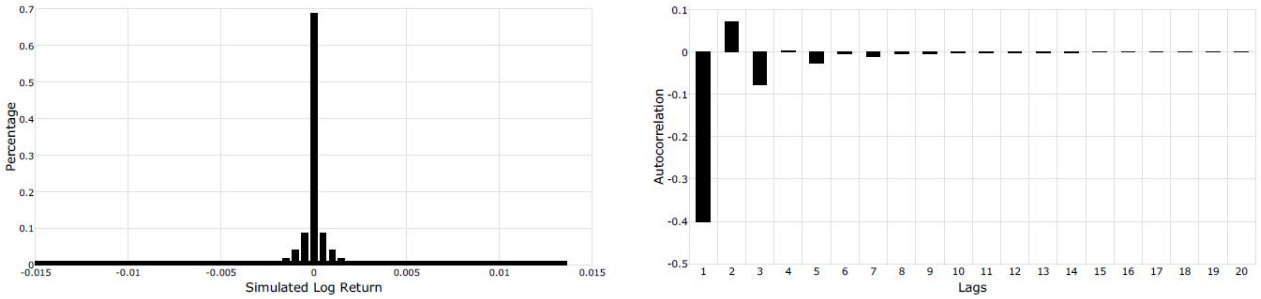
We use the bias, the mean squared error (MSE) and the QLIKE measure to compare the performance among estimators. For the true integrated variance $IV(0,t)$ and an estimate of IV denoted by $\widehat{IV}(0,t)$, the three measures are defined as follows:

$$Bias(\widehat{IV}(0,t)) = \mathrm{E}[\widehat{IV}(0,t) - IV(0,t)], \tag{62}$$

$$MSE(\widehat{IV}(0,t)) = \mathrm{E}[(\widehat{IV}(0,t) - IV(0,t))^2], \tag{63}$$

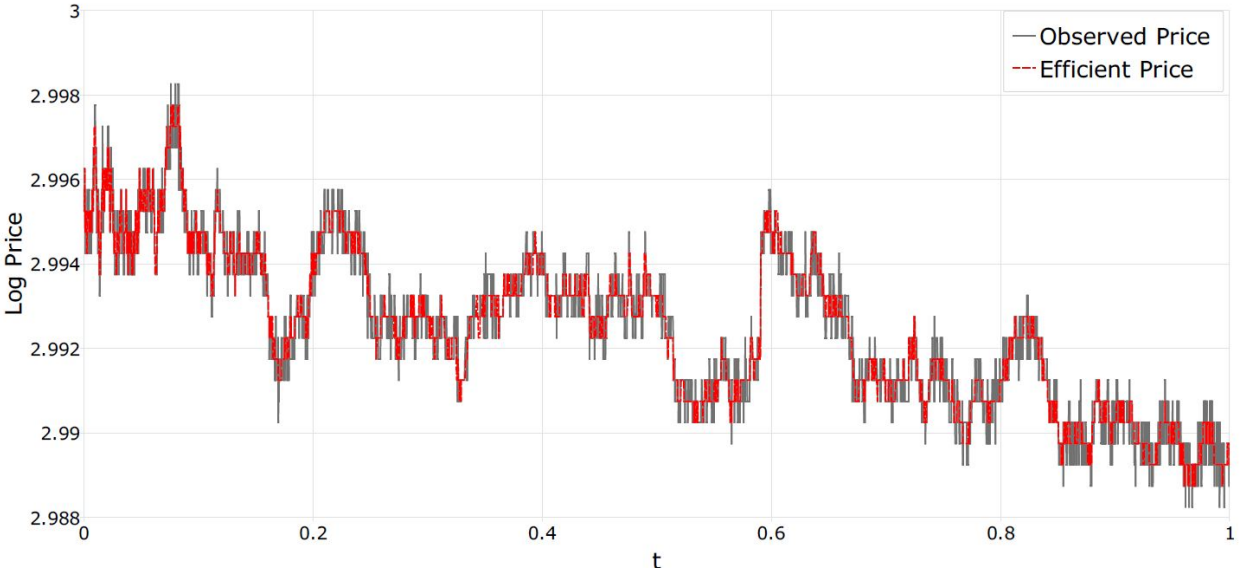
$$QLIKE(\widehat{IV}(0,t)) = \mathrm{E}\left[\frac{\widehat{IV}(0,t)}{IV(0,t)} - \ln\frac{\widehat{IV}(0,t)}{IV(0,t)} - 1\right]. \tag{64}$$

Figure 3: Histogram and correlogram for the simulated price change with moderate level of MMS noise and no jumps



Note: The histogram is on the left side and the correlogram is on the right.. The histogram is constructed based on simulated price changes for 10000 days without jumps. The correlogram is constructed based on the average autocorrelation for 10000 days without jumps. $\Delta t = \frac{1}{23400}$ in the simulation. The noise-to-signal ratio is set to be $\omega = 0.001$.

Figure 4: An example of simulated price path of the IFSV_J model with moderate level of noise



Note: t is the fraction of time in a trading day. $\Delta t = \frac{1}{23400}$ in the simulation. The noise-to-signal ratio is set to be $\omega = 0.001$.

7.2 1FSV Model Without Noise and Price Discretization

Firstly, we would like to show that price duration based volatility estimators, namely NPD and $RV^{(\delta)}$ are indeed superior to calendar time RV and realized bipower (RBip) estimators when we can observe $P^*(t)$ in continuous time without noise or price discretization.[4] We simulate 10000 replications of $P^*(t)$ for $t \in (0,1)$ with and without jump. We construct the NPD and $RV^{(\delta)}$ estimators on a grid of δs with $\delta = x\delta_0$, in which $\delta_0 = 0.1\epsilon$ and $x \in \mathbb{Z}^+$. The calendar time sampled (CTS) RV and RBip estimators are constructed based on the average sampling frequency of the NPD estimator for each x, so that CTS estimators will have a fixed sampling frequency that is comparable to that of the NPD estimator.[5] The Bias, MSE and QLIKE of the four estimators are plotted in Figure 5 against the log sampling frequency.

From the plots on the first column in Figure 5, we can see a strong negative bias for the NPD estimator at the maximum frequency when δ is small due to time discretization in the simulation. In theory the NPD estimator should converge to the integrated variance as δ decreases, but in simulation whenever we use a discrete approximation to the continuous efficient price process, the truncation bias will affect the performance of the NPD estimator when δ is small. Since $RV^{(\delta)}$ is unaffected by this truncation, it converges to the theoretical quadratic variation as δ decreases.

Comparing the efficiency of $RV^{(\delta)}$ with RV and RBip in the absence of jump, we can see clearly that $RV^{(\delta)}$ is indeed superior to RV and RBip at any sampling frequency considered in this simulation, as discussed in Section C.1. NPD and $RV^{(\delta)}$ have similar efficiency when δ is large, but the performance of NPD deteriorates as δ shrinks and the truncation bias becomes larger. However, even in the presence of truncation bias the NPD estimator is still more efficient than CTS estimators for sampling frequencies less than roughly 140 per day. When the jumps are present, $RV^{(\delta)}$ and RV are not robust to jumps and their efficiency drops sharply. We also see that the NPD estimator is more robust to jumps compared to RBip estimator as the jump variation for the NPD estimator is of a much smaller magnitude. Consequently, the efficiency advantage of the NPD estimator is even larger in the presence of jumps.

7.3 Full 1FSV Model: Primal Volatility Estimators

We proceed to add irregular transaction arrivals, price discretization and MMS noise to the 1FSV model, and compare the performance of price duration based volatility estimators NPD and $RV^{(\delta)}$ to the calendar time estimators RV and RBip. Note that these estimators are all 'primal' estimators without any correction for MMS bias. The average (log) sampling frequency for the NPD estimator is presented in Figure 6 for the 1FSV and 1FSV_J model.

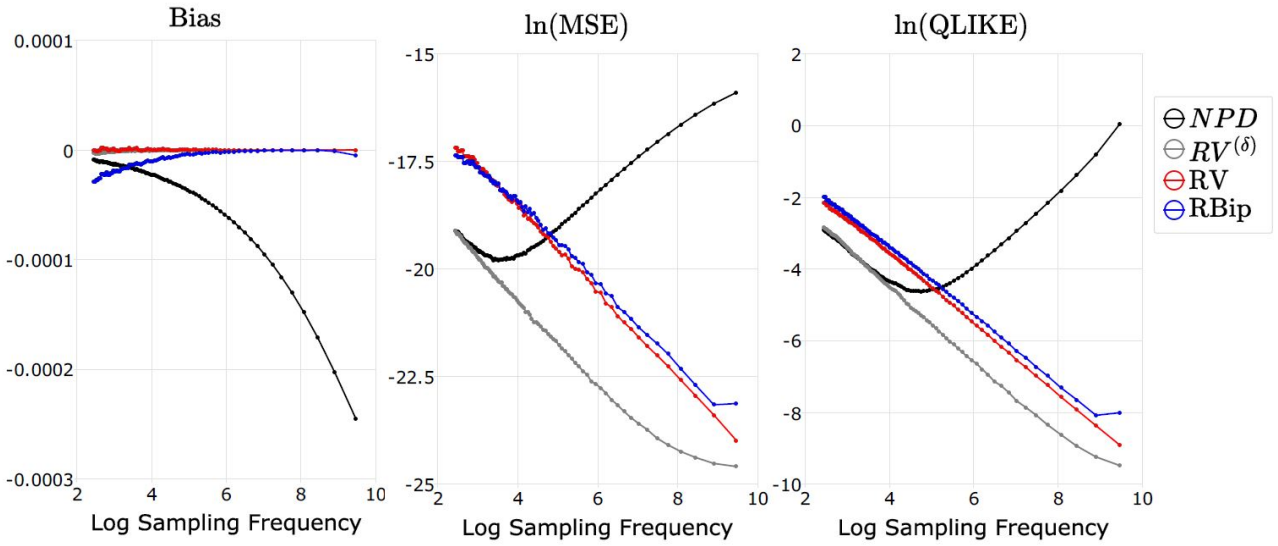
From Figure 6 we see that the sampling frequency of NPD estimators always decreases at multiples of $10\delta_0 = \epsilon$ due to price discretization. The sampling frequency ranges from roughly 3000 (exp(8)) which is the average number of tick returns per day, to roughly 7 (exp(2)) for all three levels of noise. The presence of jumps does not have a large impact on the average sampling frequency for small δs as expected, and will increase the sampling

[4] Technically, when $P^*(t)$ is observed in continuous time, NPD and $RV^{(\delta)}$ coincide in the absence of jump. Due to that we use an Euler method to approximate the continuous time $P^*(t)$, NPD will be different from $RV^{(\delta)}$ even in the absence of jump as a result of time discretization.

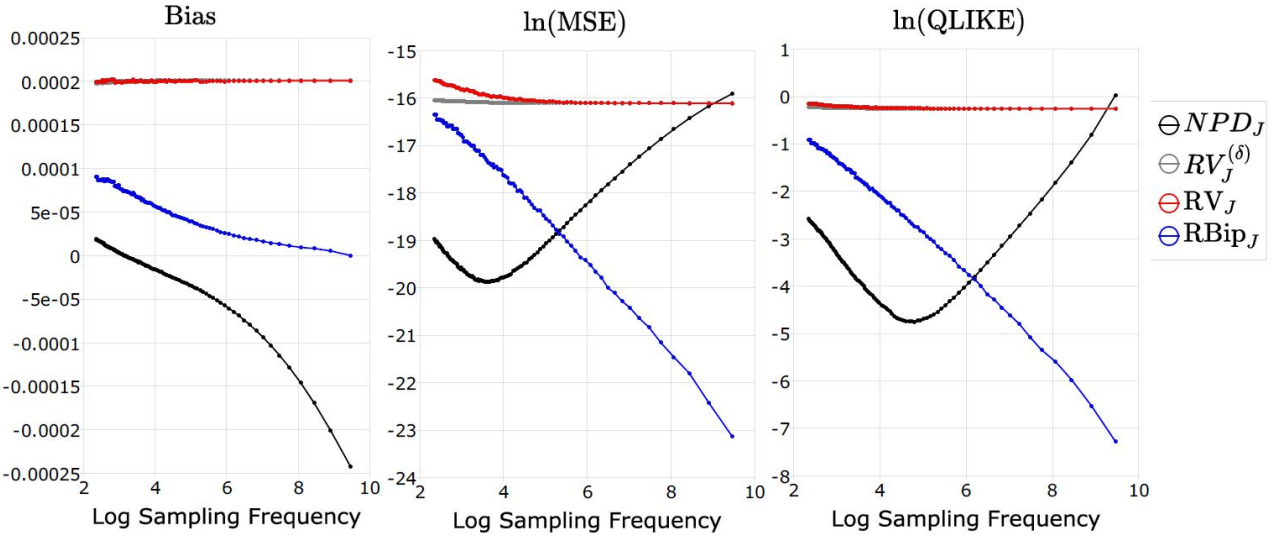
[5] Note that it is not always possible to construct a NPD estimator from a sample if the maximum range of the price is smaller than the threshold. A similar issue arises when constructing kernel and pre-averaging estimators as they are not guaranteed to be positive. The computation of Bias, MSE and QLIKE is only based on valid volatility estimates and ignores all invalid volatility estimates.

Figure 5: Simulated Bias, MSE and QLIKE for daily volatility estimates obtained from NPD, $RV^{(\delta)}$, RV and RBip for 1FSV model without noise and price discretization

Panel 1: 1FSV

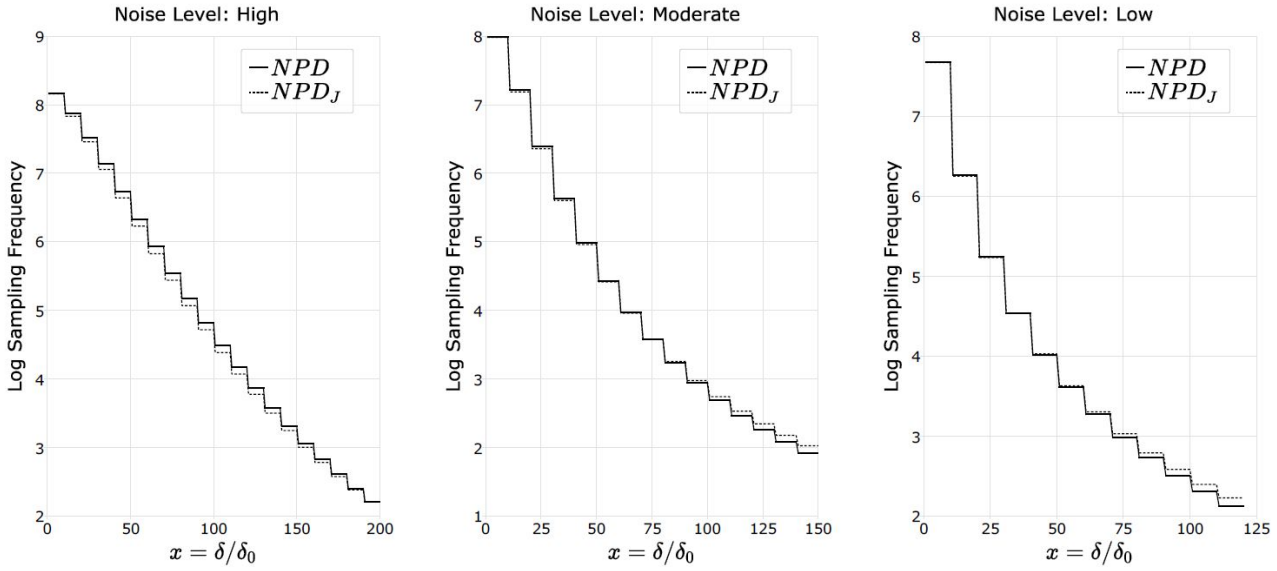


Panel 2: 1FSV$_J$



Note: The results are based on 10000 replications of the 1FSV model with and without jumps. The x-axis denotes the average log sampling frequency for a given δ for NPD and $RV^{(\delta)}$, or the log sampling frequency of the equidistant intraday return per day for RV and RBip. The subscript J represents an estimator constructed on the 1FSV model with jumps. The Euler discretization step $\Delta t = \frac{1}{23400}$.

Figure 6: Average sampling frequency of the NPD estimator for the 1FSV and 1FSV$_\text{J}$ models



Note: For the high, moderate and low levels of noise, the δ ranges from δ_0 to $200\delta_0$, $150\delta_0$ and $120\delta_0$ correspondingly. The step size is set to be $\delta_0 = 0.1\epsilon$, with $\epsilon = \ln(20.01) - \ln(20)$. For each δ, we compute the average sampling frequency by averaging the number of price durations over 10000 Monte Carlo draws of 1FSV and 1FSV$_\text{J}$ model. The noise-to-signal ratios for the high, moderate and low levels of noise are $\omega = 0.005$, 0.001 and 0.0002 respectively.

frequency slightly when δ is large. Similar to the previous case, we use the average sampling frequency of the NPD estimator to construct the calendar time RV and RBip estimators for each δ. The performance of these estimators under moderate noise can be viewed in Figure 7, and results for the high and low levels of noise cases can be found in Figure 17 and 18 in Appendix K.

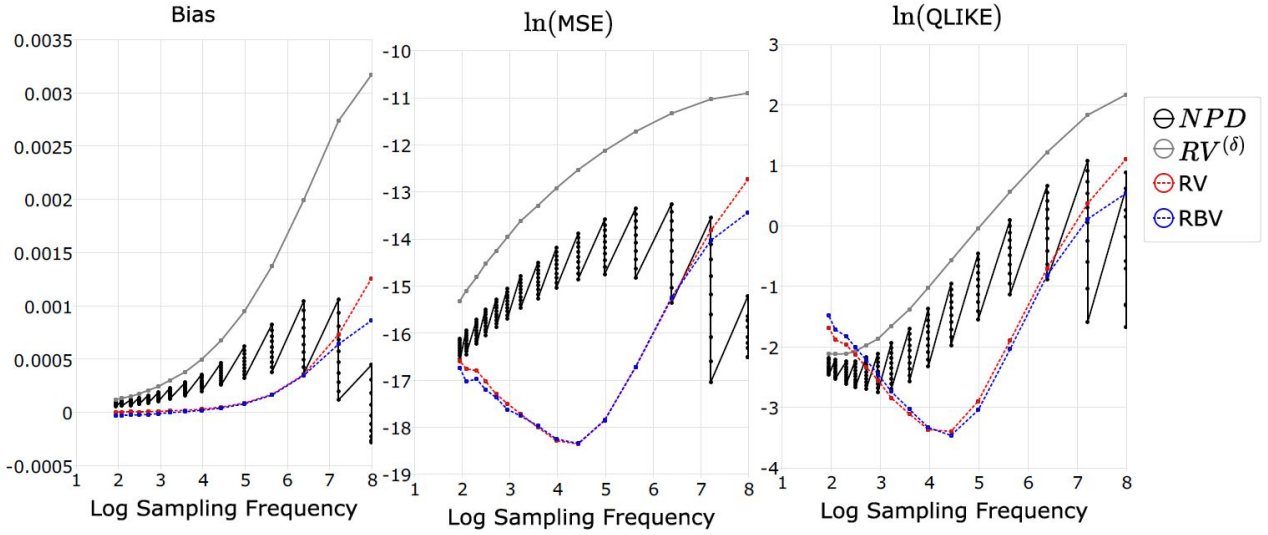
From Figure 7 we can observe that, due to the price discretization, for $\delta \in ((x-1)\epsilon, x\epsilon]$ the sampling scheme does not change. As a result, there will be multiple volatility estimates from the NPD for a given sampling frequency as δ changes within the range $((x-1)\epsilon, x\epsilon]$. It is clear that the $RV^{(\delta)}$ is the worst estimator among all 4 estimators that has a significantly larger bias and is not robust to jumps at all. Although NPD performs better than $RV^{(\delta)}$, the efficiency advantage of NPD over RV and RBip is greatly weakened by the MMS noise bias as calendar time estimators outperforms the NPD estimator for a very large range of δ. For RV and RBip, we see that the optimal sampling frequency is around $exp(4.4)$, which corresponds to a sampling frequency of 84 per day. It is evident that RBip has the overall best performance when sampled optimally due to its smallest MSE and QLIKE and its robustness to jumps. Note that the optimal sampling frequency is close to the theoretical optimal sampling frequency as proposed by Bandi and Russell (2008): $(2\omega)^{-2/3} \approx 63$.

The inferior performance of price duration based estimators to the calendar time estimators is due to that the price duration returns have a much more pronounced autocorrelation structure than the calendar time returns with the same sampling frequency. We plot the average correlogram for the calendar time returns and price duration returns sampled at RV's optimal sampling frequency in Figure 8.

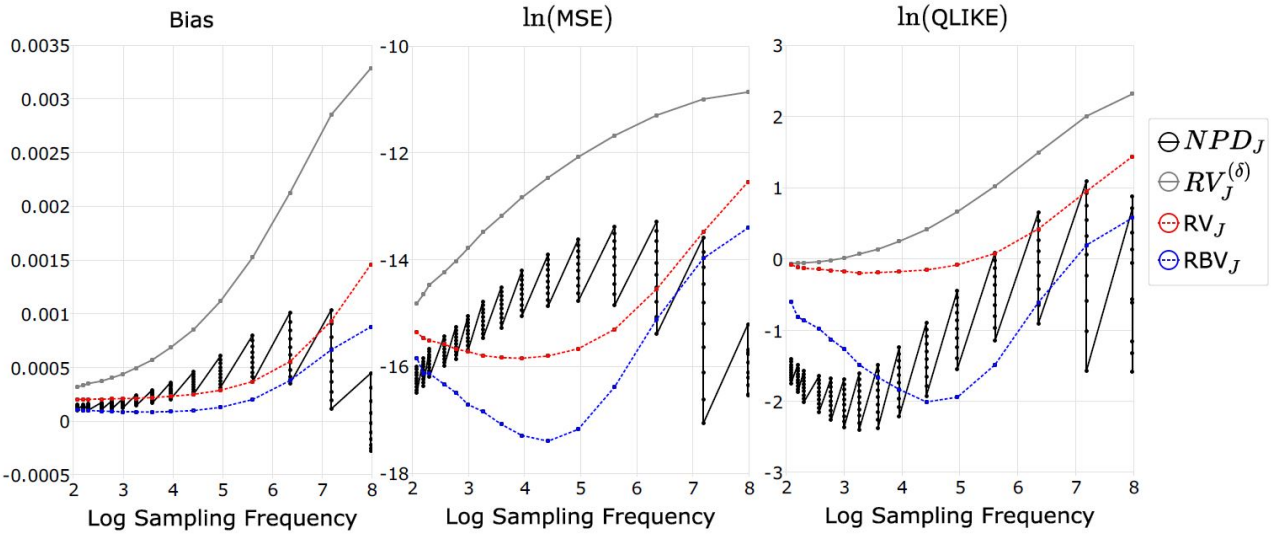
Figure 8 shows a MA(1) dependence structure for the calendar time returns, and an ARMA-type dependence structure for the price duration returns that clearly has a higher magnitude. This suggests that the MMS noise under calendar time sampling can be regarded as i.i.d. when we sample sparsely, thus the calendar time estimators are much less affected by the MMS noise. For the renewal based estimators, we see that the $RV^{(\delta)}$

Figure 7: Simulated Bias, MSE and QLIKE for daily volatility estimates obtained from NPD, $RV^{(\delta)}$, RV and RBip for 1FSV model with moderate level of MMS noise

Panel 1: 1FSV No Jump

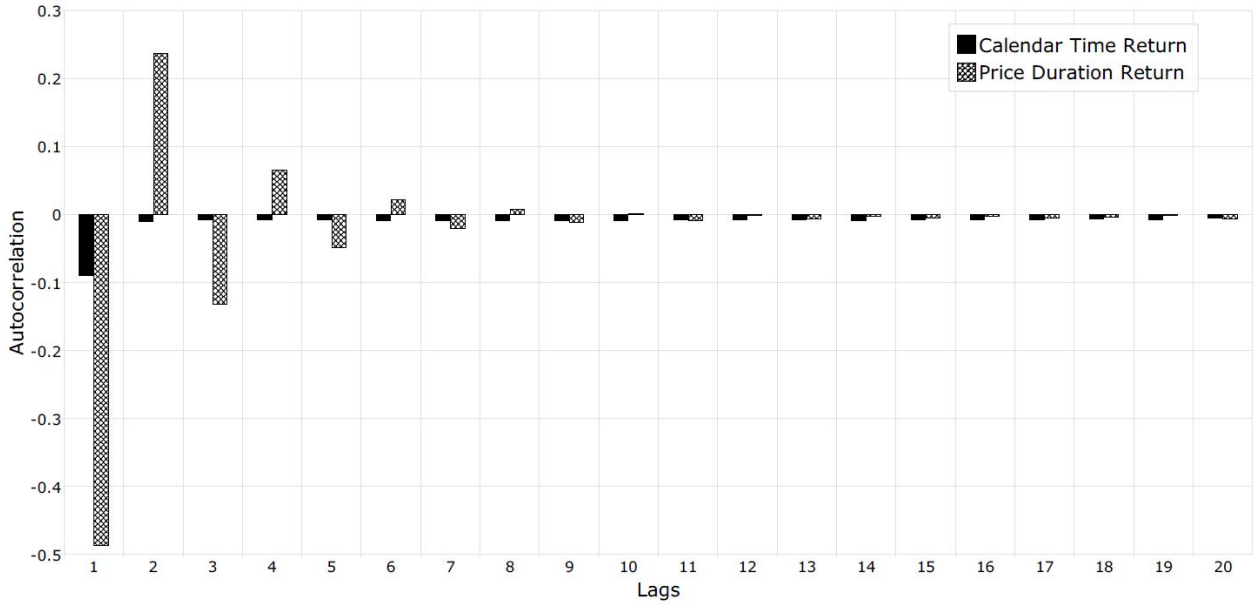


Panel 2: $1FSV_J$



Note: The results are based on 10000 replications of the 1FSV model with and without jumps. The x-axis denotes the average log sampling frequency for a given δ for NPD and $RV^{(\delta)}$, or the log sampling frequency of the equidistant intraday return per day for RV and RBip. The truncation threshold δ ranges from $\delta_1 50$ to δ_0 with a step size of $\delta_0 = 0.1\epsilon$, with $\epsilon = \ln(20.01) - \ln(20)$. The subscript J represents an estimator constructed on the 1FSV model with jumps. The noise-to-signal ratio is set to be $\omega = 0.001$.

Figure 8: Average correlogram of calendar time returns and price duration returns



Note: The results are based on averaging the first 20 autocorrelations of calendar time and price duration returns from 10000 replications of the 1FSV model with moderate noise. The sampling frequency for the calendar time return is 84 per day. The corresponding threshold of price duration is $\delta = 51\delta_0$.

performs the worst due to the dependence in the noise structure, and NPD performs better simply because the truncation bias mitigates part of the MMS noise bias. More importantly, the performance of NPD is more sensitive to the size of MMS noise than calendar time methods when the sampling frequency is on the same level. The sensitivity to the size of noise for the NPD estimator can also be seen from Figure 17 and 18 in Appendix K. In the low noise case NPD performs significantly better than the calendar time methods with smaller MSE and QLIKE if the sampling frequency is smaller than 84, similar to the no noise case. This advantage quickly diminishes as the size of the MMS noise increases, and in the large noise case the performance of NPD is completely dominated by the CTS methods for any sampling frequency smaller than 1000.

Interestingly, when size of the noise is large, one may choose a very small δ in such a way that the truncation bias exactly offsets the MMS noise bias, which explains why the NPD estimator has better performance when δ is small. However, even if we can reliably choose such a δ, the performance of this NPD estimator is still inferior to an optimally sampled CTS estimator. Moreover, it is difficult to choose a δ that can maximize MSE or QLIKE for a $\delta \in ((x-1)\epsilon, x\epsilon]$. If the goal is to choose an estimator that has a smaller MSE or QLIKE, then for the NPD estimator one needs to choose a large δ that are less affected by the truncation bias, and hopes that the MMS bias does not outweigh the smaller asymptotic variance of renewal sampling. As a result, CTS primal estimators are preferred over the NPD estimator due to that the optimal sampling frequency already has closed form approximations (See e.g. Bandi and Russell (2008) and Hansen and Lunde (2006)) and their optimal performance dominate the NPD estimator in the presence of moderate to high level of MMS noise,

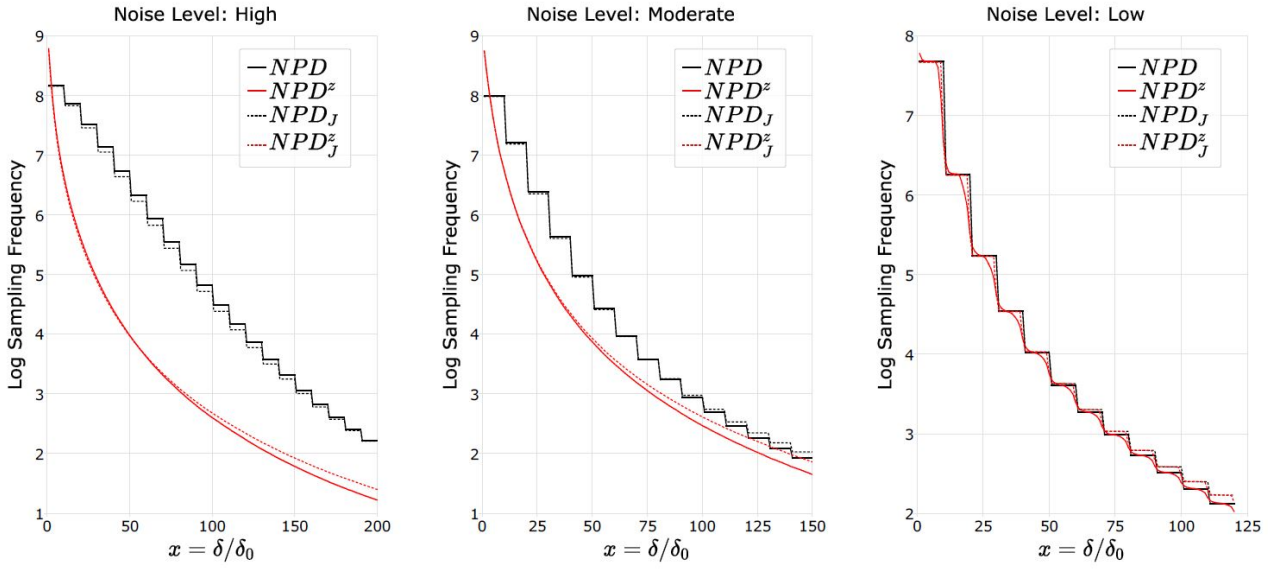
7.4 Full 1FSV Model: Bias Corrected Estimators

The discussion above suggests that, to fully exploit the smaller asymptotic variance of the price duration based estimators, it is necessary to mitigate impact of the MMS noise bias for the NPD estimator. To this end, we compare the performance of the exponentially smoothed NPD^z estimator to calendar time bias corrected

methods, namely RK, PRV and PBip estimators, which are state-of-the-art calendar time volatility estimators that are known to be highly efficient and robust to MMS noise (also robust to jumps for PBip). Similar to the previous comparison, we compare the NPD^z estimator to the calendar time rivals with the same average sampling frequency.

The choice of tuning parameters for theses estimators are non-trivial, as they have a very large impact on the performances of these estimators. Our aim here is to compare the optimal performance of all these estimators, therefore we will use optimized tuning parameters assuming they are known in advance. For the RK estimator the optimal choice of the bandwidth is provided in Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008), but there is no analytical solution to the optimal tuning parameters for NPD^z, PRV and PBip estimators. We therefore choose the tuning parameters for NPD^z, PRV and PBip by a grid search method that minimizes the simulated MSE of the estimators. Details of tuning parameter selection and implementation of all estimators considered is presented in Appendix J.

Figure 9: Average sampling frequency of the NPD and NPD^z estimator for the 1FSV and 1FSV$_\text{J}$ models under optimal γ

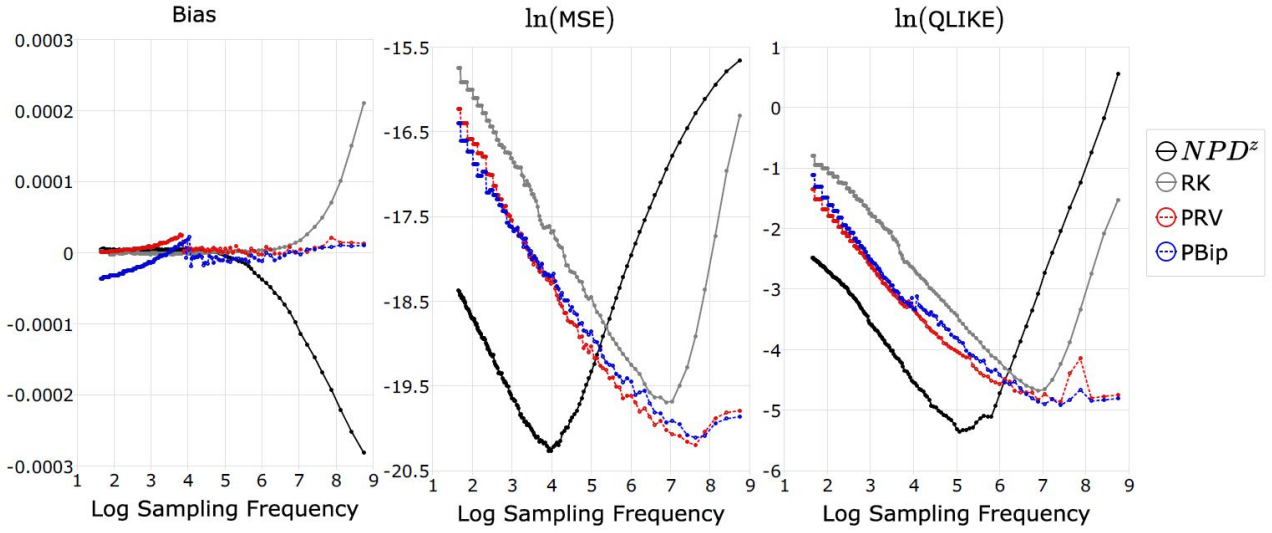


Note: For the high, moderate and low levels of noise, the δ ranges from δ_0 to $200\delta_0$, $150\delta_0$ and $120\delta_0$ correspondingly. The step size is set to be $\delta_0 = 0.1\epsilon$, with $\epsilon = \ln(20.01) - \ln(20)$. For each δ, we compute the average sampling frequency by averaging the number of price durations over 10000 Monte Carlo draws of 1FSV and 1FSV$_\text{J}$ model. The noise-to-signal ratio is set to be $\omega = 0.001$. See Appendix J for the values of the tuning parameter γ under different levels of noise.

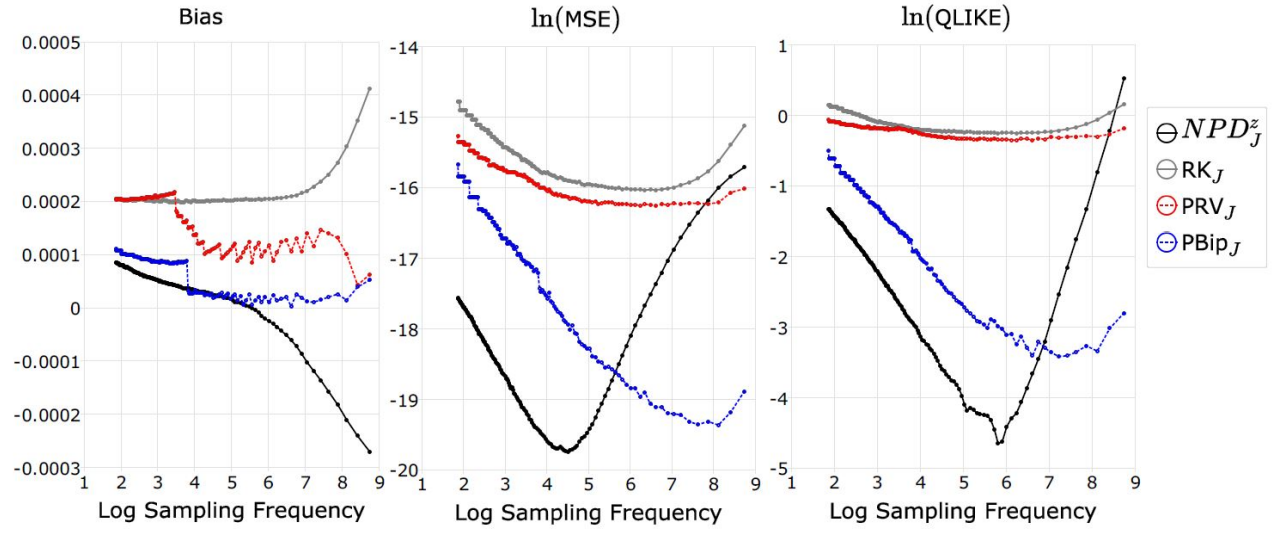
Figure 9 shows the average sampling frequency of the NPD^z estimator under optimal γ compared to that of the NPD estimator. It is clear that as the level of noise increases, the sampling frequency of the NPD^z estimator deviates from that of the NPD estimator. As the impact of noise is alleviated by the smoothing, it is expected that the sampling frequency for the NPD^z estimator is smaller than that of the NPD estimator to reduce the positive MMS bias. It is also interesting to see that the sampling frequency of NPD^z can exceed the average number of ticks in a day as smoothing removes all the flat trades. The sampling frequency is also a smoother function of δ due to exponential smoothing. Finally note that in the low level of noise case, we can still observe a step-shaped sampling frequency curve for the NPD^z, as the optimal γs are very close to 1. This suggests that smoothing does not improve the MSE of the NPD estimator in the low level of noise case, similar to the optimal θs for the pre-averaged estimators for sparsely sampled returns.

Figure 10: Simulated Bias, MSE and QLIKE for daily volatility estimates obtained from NPD^z, RK, PRV and PBip for 1FSV model with moderate level of MMS noise

Panel 1: 1FSV No Jump



Panel 2: $1FSV_J$



Note: The results are based on 10000 replications of the 1FSV model with and without jumps. The x-axis denotes the average log sampling frequency for a given δ for the NPD^z model, or the log sampling frequency of the equidistant intraday return per day for RK, PRV and PBip. The truncation threshold δ ranges from $\delta_1 50$ to δ_0 with a step size of $\delta_0 = 0.1\epsilon$, with $\epsilon = \ln(20.01) - \ln(20)$. The subscript J represents an estimator constructed on the 1FSV model with jumps. The noise-to-signal ratio is set to be $\omega = 0.001$.

We plot the Bias, MSE and QLIKE of NPD^z, RK, PRV and PBip for the moderate level of noise case in Figure 10, and the other two cases can be viewed in Figure 19 and 20 in Appendix K. In Figure 10, we see that the bias correction leads to a significant improvement in the MSE and QLIKE of volatility estimates compared to the primal estimators, especially at larger sampling frequencies. The RK estimator performs worse than the pre-averaged estimators because the window length H is optimized for convergence rate instead of MSE, thus the MMS noise bias is not fully corrected for at high sampling frequencies. For PRV and PBip, we see that these two estimators are very robust to MMS noise in the absence of jump. The performance of PRV is however affected by jumps as θ^* is optimized to minimize the MSE which uses the actual IV instead of QV. As a result, optimal PRV_J underestimates the QV so that it is less biased.

Comparing the MSE and QLIKE for the four estimators in Figure 10 we can see that, the NPD^z estimator has a clear advantage of efficiency at any sampling frequency smaller than approximately $\exp(5) \approx 150$. The NPD^z is biased towards zero when the sampling frequency is large due to the truncation bias. Similar to a NPD estimator, the truncation bias diminishes as δ increases. As is discussed in Section 6.4, the optimal γ shrinks the MMS bias in a way that it approximately offsets the truncation bias when one samples relatively sparsely. As a result, the smaller asymptotic variance of the RBV-class estimators leads to a more efficient NPD^z estimator compared to its calendar time rivals for a moderate to small sampling frequency. Interestingly, the smoothed price process Z_j itself is not noise free, so constructing RV-type estimators based on Z_j is still inferior to the NPD^z estimator.

We provide a comprehensive comparison of the optimal MSEs of all volatility estimators considered under various model settings in Table 2. A similar comparison of optimal QLIKEs can be found in Table 4 in Appendix K. From Tables 2 and 4, we see that despite a much smaller optimal sampling frequency of the NPD^z estimator compared to the pre-averaged estimators, its optimal MSE and QLIKE still outperform those of the pre-averaged estimators. Moreover, the exponential smoothing to some extent preserves the robustness to jumps of the NPD estimator, and it is evident that the efficiency advantage of the NPD^z estimator over the calendar time competitors is more pronounced in the presence of jumps.

From the discussion above we can conclude that NPD^z has the overall best MSE and QLIKE which is also very robust to jumps. Its performance is closely followed by the pre-averaging estimators PRV and PBip in the absence of noise. It is suboptimal to use a very high sampling frequency for the NPD^z estimator due to the truncation error, but the NPD^z estimator under a sparse sampling frequency can still beat the pre-averaged estimators that uses much more observations in terms of efficiency. Also note that the optimal MSE and QLIKE for the NPD^z is even lower than the optimal MSE and QLIKE of NPD in the absence of noise. This is because the smoothed MMS noise bias serves as a bias correction to the truncation bias, which reduces the bias of the NPD estimator without greatly affecting its variance.

8 Concluding Remarks

This paper proposes the class of renewal based volatility estimator for high frequency volatility estimation, and develops its asymptotic theory of the estimator based on renewal theory. The renewal based volatility estimator differs from RV-type estimators as it does not require an equidistant deterministic sampling grid and does not rely on computing squared returns. Our theory opens up a wide range of possibilities to construct alternative volatility estimators such as range duration-based RBV-type estimators with more efficiency compared to RV-

Table 2: Comparison of the optimal MSEs for all volatility estimators in Table 1 for the 1FSV and 1FSV$_{\text{J}}$ models with low, moderate and high levels of noise

Estimator	NPD	$RV^{(\delta)}$	RV	RBip	NPD^{z}	RK	PRV	PBip
1FSV model with low level of noise								
Optimal log MSE	-20.9544	-18.1988	-19.3120	-19.3286	**-20.9757**	-20.2790	-20.7407	-20.8603
δ/δ_0	34	101	21	21	35	6	5	1
Sampling Freq.	93	10	189	189	93	2160	2160	2160
1FSV model with moderate level of noise								
Optimal log MSE	-17.0324	-15.3145	-18.3553	-18.3411	**-20.2637**	-19.6876	-20.1954	-20.1100
δ/δ_0	11	141	51	51	47	9	5	5
Sampling Freq.	1356	7	84	84	146	2955	2955	2955
1FSV model with high level of noise								
Optimal log MSE	-16.5537	-11.6995	-17.2449	-17.3405	**-20.1142**	-18.6052	-19.4332	-19.4418
δ/δ_0	6	191	151	151	49	12	7	6
Sampling Freq.	3529	9	21	21	841	2618	3529	3529
1FSV$_{\text{J}}$ model with low level of noise								
Optimal log MSE	-21.0187	-15.8231	-15.9375	-18.3767	**-21.0195**	-16.0064	-16.2491	-20.0609
δ/δ_0	34	111	31	21	34	10	17	8
Sampling Freq.	93	9	93	187	93	2142	517	2142
1FSV$_{\text{J}}$ model with moderate level of noise								
Optimal log MSE	-17.0643	-14.8169	-15.8456	-17.3982	**-19.7464**	-15.9891	-16.2552	-19.3642
δ/δ_0	11	141	61	51	37	11	11	3
Sampling Freq.	1322	8	52	83	272	1322	1322	2929
1FSV$_{\text{J}}$ model with high level of noise								
Optimal log MSE	-16.5834	-11.7460	-15.5769	-16.4333	**-19.6157**	-15.8562	-16.1501	-18.5307
δ/δ_0	6	191	151	151	39	19	7	8
Sampling Freq.	3491	9	20	20	1159	2524	3491	3491

Note: Optimal log MSE for an estimator is the smallest log MSE among all the sampling frequencies considered. The smallest value is highlighted in bold. The entries for the rows δ/δ_0 represents the value of the threshold as multiples of $\delta_0 = 0.1\epsilon$, with $\epsilon = \ln(20.01) - \ln(20)$. The sampling frequency is the average sampling frequency at the optimal δs for NPD, $RV^{(\delta)}$ and NPD^{z}, and is the calendar time sampling frequency for RV, RBip, RK, PRV and PBip.

type estimators, while providing consistency and asymptotic distribution for the entire class of renewal based volatility estimators. Moreover, the stochastic sampling duration in calendar time is allowed to be parametrized, which can potentially lead to a significant efficiency gain compared to non-parametric renewal based volatility estimators.

Using the theory of renewal based volatility estimators, we prove theoretically the consistency and provide the asymptotic distribution for the point process based volatility estimator as in Engle and Russell (1998), Gerhard and Hautsch (2002), Tse and Yang (2012), Nolte, Taylor, and Zhao (2018) and Li, Nolte, and Nolte (2018) under a continuous martingale setting. We examine Nolte, Taylor, and Zhao's (Nolte, Taylor, and Zhao, 2018) NPD estimator in detail, showing its robustness to drifts and jumps, and establishing its bias structure under MMS noise, time discretization and price discretization. In our simulation study we show that: (1) it is suboptimal to choose a very small δ due to truncation bias. (2) When the MMS noise level is small, the NPD estimator is more efficient than the calendar time estimators. (3) The NPD estimator in general is more robust to jumps than the RBip estimator. (4) The NPD estimator is much more sensitive to the level of noise compared to the calendar time methods. (5) Exponentially smoothing the contaminated price process can yield an approximately unbiased NPD^z estimator that provides high efficiency compared to optimized RK and pre-averaged estimators while preserving the robustness to jumps.

This paper has several limitations that provide rooms for future research. Firstly, the idea of a range duration-based volatility estimator can be further developed as it is showing some very promising properties under the pure diffusion assumption. Different from the realized range estimator proposed by Christensen and Podolskij (2007), the normalizing coefficient for the NPR estimator is just 0.5, and the asymptotic properties follow directly from our theory. However, the properties of this estimator under various noise structures are yet to be verified, but it is promising that its properties can be analysed following the same approach for the NPD estimator presented in this paper. Secondly, the properties of the $PRBV$ estimator require further analysis, as we assume that the renewal reward process R_i is known. Therefore it is also helpful to examine the impact of estimation noise of R_i on the efficiency of the $PRBV$ estimator. Finally, theoretical properties of the NPD^z estimator and a data-driven method to select the optimal smoothing parameter γ are also worth separate investigation.

References

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

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2000): "Great Realizations," *Risk*, 13(3), 105–108.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2001): "The Distribution of Realized Exchange Rate Volatility," *Journal of the American Statistical Association*, 96(453), 42–55.

Andersen, T. G., T. Bollerslev, and D. Dobrev (2007): "No-arbitrage semi-martingale restrictions for continuous-time volatility models subject to leverage effects, jumps and i.i.d. noise: Theory and testable distributional implications," *Journal of Econometrics*, 138(1), 125–180.

Andersen, T. G., D. Dobrev, and E. Schaumburg (2008): "Duration-Based Volatility Estimation," *Working Paper, Northwestern University*, (March).

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

Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard (2008): "Designing Realized Kernels to Measure the ex post Variation of Equity Prices in the Presence of Noise," *Econometrica*, 76(6), 1481–1536.

Barndorff-Nielsen, O. E., and N. Shephard (2002): "Econometric analysis of realized volatility and its use in estimating stochastic volatility models," *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 64(2), 253–280.

——— (2004): "Power and Bipower Variation with Stochastic Volatility and Jumps," *Journal of Financial Econometrics*, 2, 1–48.

Barndorff-Nielsen, O. E., and A. Shiryaev (2010): *Change of Time and Change of Measure*. World Scientific Publishing Company, Singapore.

Billingsley, P. (2009): *Convergence of Probability Measures*. Wiley, New York.

Bollerslev, T., J. Litvinova, and G. Tauchen (2006): "Leverage and volatility feedback effects in high-frequency data," *Journal of Financial Econometrics*, 4(3), 353–384.

Bowsher, C. G. (2007): "Modelling security market events in continuous time: Intensity based, multivariate point process models," *Journal of Econometrics*, 141(2002), 876–912.

Brown, T. C., and M. G. Nair (1988): "A Simple Proof of the Multivariate Random Time Change Theorem for Point Processes," *Journal of Applied Probability*, 25(1), 210–214.

Chen, F., F. X. Diebold, and F. Schorfheide (2013): "A Markov-switching multifractal inter-trade duration model, with application to US equities," *Journal of Econometrics*, 177(2), 320–342.

Christensen, K., and M. Podolskij (2007): "Realized range-based estimation of integrated variance," *Journal of Econometrics*, 141(2), 323–349.

Coleman, R. (1982): "The Moments of Recurrence Time," *European Journal of Operational Research*, 9(2), 181–183.

Daley, D. J., and D. Vere-Jones (2003): *An introduction to the theory of point processes*, vol. I. Springer Science & Business Media.

Delattre, S., and J. Jacod (1997): "A central limit theorem for normalized functions of increments of a diffusion process, in the presence of round-off errors," *Bernoulli*, 3(1), 1–28.

Doob, J. L. (1948): "Renewal Theory From the Point of View of the Theory of Probability," *Transactions of the Americal Mathematical Society*, 63(3), 422–438.

Easley, D., and M. O'Hara (1992): "Time and the Process of Security Price Adjustment," *The Journal of Finance*, 47(2), 577–605.

Engle, R. F., and J. R. Russell (1998): "Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data," *Econometrica*, 66, 1127–1162.

Feller, W. (1941): "On the Integral Equation of Renewal Theory," *Annals of Mathematical Statistics*, 12, 243–267.

Fukasawa, M. (2010a): "Central limit theorem for the realized volatility based on tick time sampling," *Finance and Stochastics*, 14(2), 209–233.

——— (2010b): "Realized volatility with stochastic sampling," *Stochastic Processes and their Applications*, 120(6), 829–852.

Fukasawa, M., and M. Rosenbaum (2012): "Central limit theorems for realized volatility under hitting times of an irregular grid," *Stochastic Processes and their Applications*, 122(12), 3901–3920.

Gerhard, F., and N. Hautsch (2002): "Volatility estimation on the basis of price intensities," *Journal of Empirical Finance*, 9, 57–89.

Gordin, M., and M. Peligrad (2011): "On the functional central limit theorem via martingale approximation," *Bernoulli*, 17(1), 424–440.

Griffin, J. E., and R. C. A. Oomen (2008): "Sampling Returns for Realized Variance Calculations: Tick Time or Transaction Time?," *Econometric Reviews*, 27(1-3), 230–253.

Gut, A. (2012): "Anscombe's Theorem 60 Years Later," *Sequential Analysis*, 31(3), 368–396.

Hansen, P. R., and A. Lunde (2006): "Realized Variance and Market Microstructure Noise," *Journal of Business & Economic Statistics*, 24(2), 127–161.

Häusler, E., and H. Luschgy (2015): *Stable Convergence and Stable Limit Theorems*. Springer International Publishing, Switzerland.

Hautsch, N., and M. Podolskij (2013): "Preaveraging-Based Estimation of Quadratic Variation in the Presence of Noise and Jumps: Theory, Implementation, and Empirical Evidence," *Journal of Business and Economic Statistics*, 31(2), 165–183.

Huang, X., and G. Tauchen (2005): "The relative contribution of jumps to total price variance," *Journal of Financial Econometrics*, 3(4), 456–499.

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

Jacod, J., Y. Li, and X. Zheng (2017): "Statical Property of Market Microstructure Noise," *Econometrica*, 8(4), 1133–1174.

Lee, S. S., and J. Hannig (2010): "Detecting jumps from Lévy jump diffusion processes," *Journal of Financial Economics*, 96(2), 271–290.

Li, Y., and P. A. Mykland (2015): "Rounding errors and volatility estimation," *Journal of Financial Econometrics*, 13(2), 478–504.

Li, Y., I. Nolte, and S. Nolte (2018): "High-Frequency Volatility Estimation and the Relative Importance of Market Microstructure Effects : An Autoregressive Conditional Intensity Approach," *Lancaster University Management School Working Paper*.

Liesenfeld, R., I. Nolte, and W. Pohlmeier (2006): "Modelling financial transaction price movements: A dynamic integer count data model," *Empirical Economics*, 30(4), 795–825.

Lotov, I. V. (1996): "On Some Boundary Crossing Problems for Gaussian Random Walks," *The Annals of Probability*, 24(4), 2154–2171.

Nolte, I., S. Taylor, and X. Zhao (2018): "More Accurate Volatility Estimation and Forecasts Using Price Durations," *Working Paper, Lancaster University Management School*.

Oomen, R. C. A. (2005): "Properties of bias-corrected realized variance under alternative sampling schemes," *Journal of Financial Econometrics*, 3(4), 555–577.

Oomen, R. C. A. (2006): "Properties of Realized Variance Under Alternative Sampling Schemes," *Journal of Business & Economic Statistics*, 24(2), 219–237.

Peligrad, M. (1986): "Recent advances in the central limit theorem and its weak invariance principle for mixing sequences of random variables (a survey)," in *Dependence in Probability and Statistics: A Survey of Recent Results*, ed. by E. Eberlein, and M. S. Taqqu. Birkhäuser, Boston.

Ross, S. M. (1996): *Stochastic Processes*. John Wiley & Sons, Inc., New York.

Tse, Y.-k., and T. T. Yang (2012): "Estimation of High-Frequency Volatility: An Autoregressive Conditional Duration Approach," *Journal of Business & Economic Statistics*, 30(4), 533–545.

Wolff, R. W. (1989): *Stochastic Modeling and the Theory of Queues*. Prentice Hall, Englewood Cliffs NJ.

Zhang, L. (2006): "Efficient Estimation of Stochastic Volatility Using Noisy Observations: A Multi-Scale Approach," *Bernoulli*, 12(6), 1019–1043.

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

Appendices

A The Time Changed Compounded Poisson Process

Following Oomen (2006), the price process $\{P(t)\}_{t>0}$ is specified as:

$$P(t) = \sum_{i=1}^{N(t)} r_i, \quad r_i \sim i.i.d.\, \mathcal{N}(0, \sigma^2), \tag{65}$$

where $N(t)$ is a inhomogeneous Poisson process with time-varying intensity function $\lambda(t|\mathcal{F}_t) = \mathrm{E}[N(t)]$. The integrated variance of this process is defined by:

$$IV(0,t) = \sigma^2 \int_0^t \lambda(s|\mathcal{F}_s)ds. \tag{66}$$

It then follows directly from Theorem 7 that the time changed counting process $\tilde{N}(\tau(t)) = N(t)$ where $\tau(t) = IV(0,t)$ follows a homogeneous Poisson process with constant intensity σ^{-2}. Since r_i is i.i.d., $\tilde{P}(\tau(t)) = P(t)$ is by definition a Lèvy process.

B Proof to Theorem 5

We start by listing some important facts about the renewal process $\tilde{X}(\tau(t))$ and the inter-epoch durations $\tilde{D}_i$ in business time. Firstly, from Assumption 1, we have $\lim_{t\to\infty} \tau(t) \to \infty$, so the two limiting conditions $t \to \infty$ and $\tau(t) \to \infty$ can be used interchangeably. Next, from Theorem 1 we have:

$$\lim_{t\to\infty} \frac{\tilde{X}(\tau(t))}{\tau(t)} \xrightarrow{a.s.} \frac{1}{\mu} \tag{67}$$

Since $0 < \mu < \infty$, by applying Theorem 2.2 in Gut (2012) we see that:

$$\lim_{t\to\infty} \sum_{i=1}^{\tilde{X}(\tau(t))} \tilde{D}_i \xrightarrow{a.s.} \tau(t) = IV(0,t) \tag{68}$$

Now from Theorem 2.3 in Gut (2012) we have:

$$\lim_{t\to\infty} \frac{\sum_{i=1}^{\tilde{X}(\tau(t))}(\tilde{D}_i - \mu)}{\sigma\sqrt{\tilde{X}(\tau(t))}} \xrightarrow{d} \mathcal{N}(0,1) \tag{69}$$

Applying (68) and substituting $X(t) = \tilde{X}(\tau(t))$ and $RBV(0,t) = \tilde{X}(\tau(t))\mu$ into the above equation yield the desired result.

C Asymptotic Properties of the RBV Estimator under Infill Asymptotics and Comparison to the RV Estimator

The reason why the sprawl asymptotics is preferred in the renewal literature is that usually we assume the data generating parameters μ and σ^2 to be fixed, and we estimate these parameters by an infinitely long sample. In our case, we can actually change μ arbitrarily by altering the stopping criteria $\mathcal{S}(t_i)$. In this section we use the superscript (μ) to distinguish between the renewal sampling schemes with different μ.

We consider the asymptotic properties of the *RBV* estimator defined in Definition 5 under a fixed sampling period $(0,T)$. We assume that the price process follows the assumptions in Assumption 1. The quantity of interest is therefore $IV(0,T) = \tau(T) = \int_0^T \sigma_P^2(s)ds$, which is a random variable. The durations in business time $\{\tilde{D}_i^{(\mu)}\}_{i=1:X(T)}$, are still i.i.d., so the point process $\tilde{X}^{(\mu)}(\tau(t))$ is still a renewal process. We can think of the quantity $\tilde{X}^{(\mu)}(\tau(T))$ as the counts of renewal epochs when the renewal process is stopped randomly at time $\tau(T)$.

To derive the counterpart result of Theorem 5 under infill asymptotics, we require the following additional assumptions on $P(t)$ and $X^{(\mu)}(t)$:

Assumption 4. *For a fixed time period* $(0,T)$*:*

1. *(Continuity of the price process) We assume that* $\lim_{\mu\to 0} X^{(\mu)}(T) \to \infty$.
2. *(Convergence of the age density) We assume that* $\lim_{\mu\to 0} \frac{\mu_{n+1}}{(n+1)\mu} \to 0$ *for all* $n = 1, 2, \ldots$*, where* μ_n *is the* n*-th moment of* $\tilde{D}_i^{(\mu)}$.

Assumption 4.1 ensures that by sampling with an infinitesimally small μ in business time, the renewal sampling frequency goes to infinity. This implies that the price process must contain a diffusion component or a jump component with infinite activity. A direct consequence of Assumption 4.2 is that:

$$\lim_{\mu\to 0} \sum_{i=1}^{\tilde{X}^{(\mu)}(\tau(T))} D_i^{(\mu)} \xrightarrow{a.s.} \tau(T). \tag{70}$$

This is due to the fact that the age process defined in Definition 2 converges uniformly to a point mass at zero as μ approaches zero, so the arrival time of the last epoch $\tau(t_{X(T)})$ converges almost surely to a random variable $\tau(T)$. Assumption 4.3 will be used in the derivation of the asymptotic results.

The asymptotic result of the *RBV* estimator in the infill asymptotics case is derived by a direct application of Corollary 6.4 in Häusler and Luschgy (2015). Since μ can be chosen arbitrarily, we choose $\mu(n) = n^{-1}$ with $n = 1, 2, \ldots$, so that $\mu(n) \to 0$ is equivalent to $n \to \infty$. We then construct the following random variable:

$$Z_{ni} = \frac{D_i^{(\mu(n))} - \mu(n)}{\mu(n)^{-0.5}\sigma(n)}. \tag{71}$$

Note that Z_{ni} is a square integrable martingale difference array w.r.t. its natural filtration $\mathcal{F}_{nk}$. Additional technical assumptions are required for Corollary 6.4 in Häusler and Luschgy (2015) to hold:

Assumption 5. *Technical assumptions for the stable convergence of the RBV estimator:*

1. *(Measurability)* $X^{(\mu(n))}(\tau(T))$ *is a finite stopping time w.r.t.* $\mathcal{F}_{nk}$.
2. *(Conditional Lindeberg's condition):*

$$\lim_{n\to\infty} \sum_{i=1}^{\tilde{X}^{(\mu(n))}(\tau(T))} \mathrm{E}[Z_{ni}^2 \mathbb{1}_{\{|Z_{nk}\geq\varepsilon\}} | \mathcal{F}_{n,i-1}] \xrightarrow{p} 0 \tag{72}$$

for every $\varepsilon > 0$.

From Assumption 4.2 we also see that:

$$\lim_{n\to\infty} \sum_{i=1}^{\tilde{X}^{(\mu(n))}(\tau(T))} \mathrm{E}[Z_{ni}^2 | \mathcal{F}_{n,i-1}] \xrightarrow{a.s.} \tau(T), \tag{73}$$

and the measurability of $X^{(\mu(n))}(\tau(T))$ ensures that $\tau(T)$ is also measurable w.r.t. $\mathcal{F}_{nk}$. Therefore we can apply Corollary 6.4 in Häusler and Luschgy (2015) to Z_{ni}, which yields:

$$\lim_{n\to\infty} \sum_{i=1}^{\tilde{X}^{(\mu(n))}(\tau(T))} Z_{ni} \xrightarrow{s.t.} \sqrt{\tau(T)}\mathcal{N}(0,1), \tag{74}$$

where s.t. refers to stable convergence in law. (74) leads to the following asymptotic distribution of RBV:

$$\lim_{\mu\to 0} \frac{RBV(0,T) - IV(0,T)}{\sqrt{IV(0,T)\mu^{-1}\sigma^2}} \xrightarrow{d} \mathcal{N}(0,1), \tag{75}$$

and note that $\lim_{\mu\to 0} X(T) \xrightarrow{a.s.} \sqrt{IV(0,T)\mu^{-1}}$. Thus, similar asymptotic results to Theorem 5 also holds under the setting of infill asymptotics, with the expense of additional assumptions in Assumption 4 and 5.

C.1 Relationship to the RV Estimator

The infill asymptotics results for the RBV estimator can be linked naturally to the RV estimator, as we can interpret the renewal sampling scheme as a stochastic sampling scheme for the RV estimator. We start with the assumption that $P(t)$ is a continuous local martingale to which Theorem 4 can be applied. For a given μ, let us denote the renewal sampling scheme as $X^{(\mu)}(t)$, the sampling times as $\{t_i^{(\mu)}\}_{i=1,2,\cdots}$ and the inter-event return as $r_i^{(\mu)} = P(t_i^{(\mu)}) - P(t_{i-1}^{(\mu)})$. We define the renewal RV and the RBV estimator as

$$\begin{aligned} RV^{(\mu)}(0,T) &= \sum_{i=1}^{X^{(\mu)}(T)} (r_i^{(\mu)})^2, \\ RBV^{(\mu)}(0,T) &= X^{(\mu)}(T)\mu. \end{aligned} \tag{76}$$

From the theory of quadratic variation and (75) we know that both estimators are consistent, and converge to $IV(0,T)$. Specially, for the $RV^{(\mu)}$ estimator, due to the i.i.d-ness of the inter-event arrival time in business time denoted by $\tilde{D}_i^{(\mu)} = \tau(t_i^{(\mu)}) - \tau(t_{i-1}^{(\mu)})$, $r_i^{(\mu)}$ is also i.i.d. From the martingale property of the Wiener process we have:

$$\mathrm{E}[r^{(\mu)}] = 0, \mathrm{E}[(r_i^{(\mu)})^2|\tilde{D}_i] = \tilde{D}_i, \mathrm{E}[(r_i^{(\mu)})^2] = \mu. \tag{77}$$

This suggests that a natural and consistent estimator of μ is just the sample moment of the squared return, $\hat{\mu} = \frac{1}{X^{(\mu)}(T)} \sum_{i=1}^{X^{(\mu)}(T)} (r_i^{(\mu)})^2$. Obviously, by using $\hat{\mu}$ instead of μ, the $RBV^{(\mu)}$ estimator coincides with the $RV^{(\mu)}$ estimator. The cost of using $\hat{\mu}$ in the $RBV^{(\mu)}$ estimator is then a larger asymptotic variance. Using Corollary 3.11 in Fukasawa (2010b) and Assumption 4, we see that as $\mu \to 0$:

$$\mathrm{V}[RV^{(\mu)}(0,T)] \to \frac{2}{3} \sum_{i=1}^{X^{(\mu)}(T)} (r_i^{(\mu)})^4. \tag{78}$$

When the unconditional kurtosis $\kappa^{(\mu)}$ of $r_i^{(\mu)}$ exists, the above asymptotic variance converges to $\frac{2}{3}X^{(\mu)}(T)\kappa^{(\mu)}\mu^2$, which is due to the i.i.d.-ness of $r_i^{(\mu)}$.

The asymptotic variance $\frac{2}{3}X^{(\mu)}(T)\kappa^{(\mu)}\mu^2$ has some very interesting implications. Firstly, if $\kappa = 3$ and $r_i^{(\mu)}$ is normally distributed, we have $\mathrm{V}[RV^{(\mu)}](0,T) \to 2IV(0,T)^2/X^{(\mu)}(T)$, which is identical to the asymptotic variance of the RV estimator sampled in business time (e.g. Hansen and Lunde (2006), Oomen (2006)). The business time RV can indeed be considered as a RBV estimator with a constant duration in business time. Moreover, if we can sample $r_i^{(\mu)}$ by setting $\kappa^{(\mu)} = 1$, then the asymptotic variance of the $RV^{(\mu)}$ estimator can be minimized, and is equal to $2IV(0,T)^2/3X^{(\mu)}(T)$. This implies that the optimal renewal RV estimator must have $r_i^{(\mu)}$ following a two-point distribution. We will show later in Section 5 that, the non-parametric duration-based volatility estimator in Nolte, Taylor, and Zhao (2018) is both a RBV-class estimator and an optimal renewal RV estimator.

D Proof of Corollary 3

To prove the corollary, we use the Doob-Meyer decomposition of a point process. In detail, any $\mathcal{F}_t$-adapted point process $X(t)$ is a submartingale, and for a submartingale, the following decomposition is unique:

$$X(t) = \Lambda(t) + M(t), \tag{79}$$

where $\Lambda(t)$ is a $\mathcal{F}_t$-predictable increasing process called the compensator of $X(t)$, and $M(t)$ is a $\mathcal{F}_t$ martingale. The compensator process and the intensity process is linked via the following relationship:

$$\Lambda(t) = \int_0^t \lambda(t|\mathcal{F}_t)dt \tag{80}$$

Therefore, to prove the corollary, we firstly show that under business time, $\tilde{\Lambda}(\tau(t)) = \Lambda(t)$ is the compensator of the process $\tilde{X}(\tau(t))$. Note that under business time, we have the following decomposition for $\tilde{X}(\tau(t))$:

$$\tilde{X}(\tau(t)) = \breve{\Lambda}(\tau(t)) + \breve{M}(\tau(t)), \tag{81}$$

in which $\breve{\Lambda}(\tau(t))$ is the compensator of $\tilde{X}(\tau(t))$ and $\breve{M}(\tau(t))$ is a martingale in business time. Moreover, if we change $X(t)$ from calendar time to business time, we have that:

$$\tilde{X}(\tau(t)) = \tilde{\Lambda}(\tau(t)) + \tilde{M}(\tau(t)). \tag{82}$$

Importantly, $\tilde{M}(\tau(t))$ is also a martingale due to the fact that the time change preserves the martingale property according to the optional stopping theorem. Then from the uniqueness of the Doob-Meyer decomposition we see that for all t, $\tilde{M}(\tau(t)) = \breve{M}(\tau(t))$ and therefore $\breve{\Lambda}(\tau(t)) = \tilde{\Lambda}(\tau(t))$.

By the definition of conditional intensity we see that:

$$\tilde{\Lambda}(\tau(t)) = \int_{\tau(0)}^{\tau(t)} \tilde{\lambda}(\tau(s)|\mathcal{F}_s)d\tau(s) = \int_0^t \lambda(s|\mathcal{F}_s)ds, = \Lambda(t) \tag{83}$$

and it is therefore clear that $\tilde{\lambda}(\tau(t)|\mathcal{F}_t)$ is the conditional intensity process of $\tilde{X}(\tau(t))$. Now, since the above equation holds for an arbitrary t, it must also hold that:

$$\tilde{\lambda}(\tau(t)|\mathcal{F}_t)d\tau(t) = \lambda(t|\mathcal{F}_s)dt. \tag{84}$$

Substituting $d\tau(t) = \sigma_p^2(t)dt$ into the above equation yields the desired result.

E Proof of Corollary 1

To prove the first part of the corollary, we only need to show that R_i is a monotonically increasing function of $\tilde{D}_i$ in the sense that for any $\tilde{D}_i > \tilde{D}_j$, $R_i > R_j$.

From the proof of Proposition 3, we can write R_i in terms of $\tilde{\lambda}(\tau(t)|\mathcal{F}_t)$:

$$R_i = \mu \int_{t_{i-1}}^{t_{i-1}+D_i} \lambda(s|\mathcal{F}_s ds = \mu \int_{\tau(t_{i-1})}^{\tau(t_{i-1})+\tilde{D}_i} \tilde{\lambda}(\tau(s)|\mathcal{F}_{\tau(t)})d\tau(s). \tag{85}$$

Note that by the definition of conditional intensity and due to that the process $\tilde{X}(\tau(t))$ is renewal, we have $\tilde{\lambda}(\tau(t_i)+s|\mathcal{F}_{\tau(t_i)}) = h_{\tilde{D}}(s)$, where $h_{\tilde{D}}(s)$ is the hazard function of the renewal process $\tilde{X}(\tau(t))$ defined by:

$$h_{\tilde{D}}(s) = -\frac{\partial \ln(1-F_{\tilde{D}}(s))}{\partial s}, \tag{86}$$

in which $F_{\tilde{D}}(s)$ is the CDF of $\tilde{D}_i$. The cumulative hazard function $H_{\tilde{D}}(x)$ is defined as:

$$H_{\tilde{D}}(s) = \int_0^s h_{\tilde{D}}(u)du = -ln(1-F_{\tilde{D}}(s)). \tag{87}$$

The equivalence between $h_{\tilde{D}}(s)$ and $\tilde{\lambda}(\tau(t_i)+s|\mathcal{F}_{\tau(t_i)})$ suggests the following relationship which holds true for all t:

$$\int_{\tau(t_{i-1})}^{\tau(t_{i-1}+t)} \tilde{\lambda}(\tau(s)|\mathcal{F}_{\tau(t)})d\tau(s) = \int_{t_{i-1}}^{t_{i-1}+t} \lambda(s|\mathcal{F}_s)ds = -\ln(1-F_{\tilde{D}}(\tau(t_{i-1}+t)-\tau(t_{i-1})). \tag{88}$$

Taking $t = D_i$ and substitute into the equation above:

$$R_i = -\mu \ln(1-F_{\tilde{D}}(\tilde{D}_i)). \tag{89}$$

Note that the term $-\ln(1-F_{\tilde{D}}(\tilde{D}_i))$ in the above equation is the exponential inverse probability integral transform of $\tilde{D}_i$ which follows an i.i.d. unit exponential distribution. This is consistent with the result that R_i is i.i.d. exponential. More importantly, $-\ln(1-F_{\tilde{D}}(\tilde{D}_i))$ is a monotonically increasing function of $\tilde{D}_i$, which completes the proof of the first part.

To prove the second part of the corollary, we note that when $\tilde{D}_i$ is i.i.d. exponential with mean μ and variance μ^2, $\tilde{\lambda}(\tau(t)|\mathcal{F}_t) = \mu^{-1}$. Apply Proposition 3 and observe that $\sigma_p^2(t) = \mu\lambda(t|\mathcal{F}_t) = g(t|\mathcal{F}_t)$ as desired.

F Simulation of $\rho^{(\delta)}$ and $\rho^{(r)}$ for the PD and PR estimators

To simulate $\rho^{(\delta)}$ and $\rho^{(r)}$, we firstly simulate a standard Wiener process. Let $\Delta W_i \sim \mathcal{N}(0,\Delta)$, and the (discrete) Wiener process is simulated as:

$$W_j = \sum_{k=1}^{\infty} \Delta W_k. \tag{90}$$

In the simulation we set $\Delta = 10^{-5}$. The stopping times $\{\tilde{D}_i^{(\delta)}\}_{i=1:N}$ and $\{\tilde{D}_i^{(r)}\}i = 1:N$ are then constructed by setting $\delta = r = 1$ based on this Wiener process as follows:

$$\begin{aligned} \tilde{D}_i^{(\delta)} &= \frac{1}{\Delta} \inf_{j>i-1}\{W_j : |W_j - W_{i-1}| \geq 1\}, \\ \tilde{D}_i^{(r)} &= \frac{1}{\Delta} \inf_{j>i-1}\{W_j : \sup_{i-1<s<j}(W_s) - \inf_{i-1<s<j}(W_s) \geq 1\}. \end{aligned} \tag{91}$$

We choose $N = 1000000$. Note that there will be a small truncation bias due to the discreteness of the simulated Wiener process. This will cause the simulated $\tilde{D}_i^{(\delta)}$ and $\tilde{D}_i^{(r)}$ to be biased upward slightly, and the bias vanishes as $\Delta \downarrow 0$. This will not have a significant impact as long as Δ is relatively small compared to δ or r. Based on the simulated $\tilde{D}_i^{(\delta)}$ and $\tilde{D}_i^{(r)}$, we can construct $R_i^{(\delta)}$ and $R_i^{(r)}$ as:

$$R_i^{(\cdot)} = -\,\mathrm{E}[\tilde{D}^{(\cdot)}]\ln(1-\hat{F}_{\tilde{D}^{(\cdot)}}(\tilde{D}_i^{(\cdot)})), \tag{92}$$

in which $\hat{F}_{\tilde{D}^{(\cdot)}}(x)$ is the empirical CDF of $\tilde{D}^{(\cdot)}$. We do not use the theoretical CDF because it is not available in closed form. The correlation $\rho^{(\cdot)}$ is then computed based on $\{\tilde{D}^{(\cdot)}\}_{i=1:N}$ and $\{R_i^{(\cdot)}\}_{i=1:N}$. We plot the simulated moments for $\{\tilde{D}^{(\cdot)}\}_{i=1:N}$ and $\{R_i^{(\cdot)}\}_{i=1:N}$ and the simulated $\rho^{(\cdot)}$ in Table 3. $\mathrm{V}[\tilde{D}_i^{(\cdot)} - R_i^{(\cdot)}]$ of an arbitrary δ or r can be easily obtained by scaling the corresponding variables.

Table 3: Simulated moments for $\tilde{D}^{(\cdot)}$ and $R_i^{(\cdot)}$ and the simulated $\rho^{(\cdot)}$

	$\mathrm{E}[\tilde{D}^{(\delta)}]$	$\mathrm{V}[\tilde{D}^{(\delta)}]$	$\mathrm{E}[R_i^{(\delta)}]$	$\mathrm{V}[R_i^{(\delta)}]$	$\rho^{(\delta)}$	$\mathrm{V}[\tilde{D}_i^{(\delta)} - R_i^{(\delta)}]$
Simulated	1.0033	0.6707	1.0000	0.9999	0.9998	0.0330
Theoretical	1	0.6667	1	1	-	0.0340
	$\mathrm{E}[\tilde{D}^{(r)}]$	$\mathrm{V}[\tilde{D}^{(r)}]$	$\mathrm{E}[R_i^{(r)}]$	$\mathrm{V}[R_i^{(r)}]$	$\rho^{(r)}$	$\mathrm{V}[\tilde{D}_i^{(r)} - R_i^{(r)}]$
Simulated	0.5036	0.0844	0.5000	0.2500	0.9915	0.0463
Theoretical	0.5	0.0833	0.5	0.25	-	0.0471

Note: $\delta = r = 1$. $N = 1000000$. Theoretical values of the simulated moments can be found in (27) and (31). Note that for theoretical moments of $\mathrm{V}[\tilde{D}_i^{(\cdot)} - R_i^{(\cdot)}]$ we plug in the simulated $\rho^{(\cdot)}$ in the relationship: $\mathrm{V}[\tilde{D}_i^{(\cdot)}] + \mathrm{V}[R_i^{(\cdot)}] - 2\rho^{(\cdot)}\sqrt{\mathrm{V}[\tilde{D}_i^{(\cdot)}]\,\mathrm{V}[R_i^{(\cdot)}]}$.

G An Approximated Time Discretization Bias

Throughout this section we assume $r_j = r_j^*$, that is, the MMS noise is absent in the price process. We start by decomposing $r_i^{(\delta)}$ as:

$$r_i^{(\delta)} = \sum_{j=j_{i-1}^{(\delta)}}^{j_{i-1}^{(\delta)}+M_i^{(\delta)}} r_j, \tag{93}$$

where $j_i^{(\delta)}$ is the observation index of $t_i^{(\delta)}$, and $M_i^{(\delta)}$ is the number of observations in the i-th price duration (excluding the starting point). We see that since we assume r_j to be strongly mixing and strictly stationary with finite moments, from the central limit theorem (e.g. Peligrad (1986), Billingsley (2009)) it holds that:

$$\lim_{N\to\infty} \sum_{n=1}^{N} r_j \sim \mathcal{N}(0, NV[r_j]), \tag{94}$$

where $V[r_j] = \mathrm{E}[\tilde{d}_j]$ is the unconditional variance of the tick return. Now consider N_i being a sufficiently large random variable, so that $\sum_{n=1}^{N_i} r_j$ is approximately mixture normal. The absolute price change point process truncates this random variable $\sum_{n=1}^{N_i} r_j$ whenever $|\sum_{n=1}^{N_i} r_j| \geq \delta$, and the distribution of $r_i^{(\delta)}$ becomes very complicated.

To provide an approximated result, we treat the sequence r_j as i.i.d. normal variates with variance $\mathrm{V}[r_j]$. Let $S_i = \sum_{j=1}^{N_i} r_j$ denote the partial sum of the returns till step N_i, the process S_i is then a Gaussian random walk. For a truncation threshold δ, we use the joint distribution $\{N_i^{(\delta)}, S_{N_i^{(\delta)}}\}$ to approximate $\{M_i^{(\delta)}, r_i^{(\delta)}\}$. The asymptotic expansion of $\mathrm{E}[S^2_{N_i^{(\delta)}}]$ as $\delta \to \infty$ is given in Lotov (1996):

$$\mathrm{E}[S^2_{N_i^{(\delta)}}] = \delta^2 + 2\delta\sqrt{\mathrm{V}[r_j]}\mathcal{K} + \mathrm{V}[r_j]\mathcal{K}^2 + \frac{1}{4} + o(1), \tag{95}$$

where $\mathcal{K} \approx 0.58258087$ is defined through: $\mathcal{K} = \frac{1}{\sqrt{2\pi}} \lim_{n\to\infty}[2\sqrt{n} - \sum_{m=1}^{n} m^{-1}]$. From Wald's identity we also have that: $\mathrm{E}[S^2_{N_i^{(\delta)}}] = \mathrm{V}[r_j]\,\mathrm{E}[N_i^{(\delta)}]$.

In this Gaussian random walk setting, $S_{N_i^{(\delta)}}$ can be interpreted as the return for the i-th price duration. As a result, the expected TD bias is just: $TD^{(\delta)}(0,t) = \sum_{i=1}^{X(t)} S^2_{N_i^{(\delta)}} - X(t)\delta^2$. Apply Wald's identity once again, we have $\mathrm{E}[TD^{(\delta)}(0,t)] \to \mathrm{E}[X(t)](\mathrm{E}[S^2_{N_i^{(\delta)}}] - \delta^2)$ in the limit. Also, $\mathrm{E}[X(t)]$ in the limit converges to $\frac{IV(0,t)}{\mathrm{E}[S^2_{N_i^{(\delta)}}]}$ which

is from the property of renewal processes. The approximated $\widetilde{TD}$ bias is therefore:

$$\begin{aligned}\lim_{\delta\to\infty}\widetilde{TD}^{(\delta)}(0,t) &\to \frac{IV(0,t)}{\mathrm{E}[S^2_{N_i^{(\delta)}}]}\,\mathrm{E}[S^2_{N_i^{(\delta)}}-\delta^2]\\ &= IV(0,t)(1-\frac{1}{1+O(\delta^{-1})})\end{aligned} \tag{96}$$

which converges to zero as $\delta\to\infty$ with a rate of δ^{-1}.

H Determinants of the Bias of the NPD Model

We provide a simple example to illustrate the bias of the NPD estimator as a function of δ via simulation. Based on (44), we assume that the arrival of $\tau(t_j)$ in business time follows a homogeneous Gamma process with intensity measure $v(x)=\frac{\gamma}{xe^{\lambda x}}$. The inter-observation durations in business time are then i.i.d. Gamma distributed: $\tilde{d}_j\sim\Gamma(\gamma,\lambda)$. Let z_j^* denote an i.i.d. standard normal variable, we have:

$$r_j^*=z_j^*\sqrt{\tilde{d}_j}\sim\mathcal{N}(0,\tilde{d}_j). \tag{97}$$

so the tick return is unconditionally mixture normal. This simple structure allows for a leptokurtic distribution of r_j^* with the following sample moments: $\mathrm{V}[r_j^*]=\gamma\lambda$, $K[r_j^*]=3+\frac{3}{\gamma}$. For the noise term V_j, we assume that it follows an AR(1) process:

$$V_j=\rho V_{j-1}+v_j,\quad v_j\sim\mathcal{N}\Big(0,\frac{\sigma_v^2}{1-\rho^2}\Big). \tag{98}$$

To ensure that V_j complies with Assumption 2, we further require that $|\rho|<1$ and $v_j\perp\!\!\!\perp r_j^*$. The unconditional variance V_j is therefore σ_v^2 for any $\rho\in(-1,1)$. The tick return r_j is therefore conditionally normally distributed with an ARMA-type autoregressive structure. We will refer to this model as the Gamma subordinated transaction (GST) model. Some moment conditions for r_j are summarized in Appendix I.

To illustrate the asymptotic properties of the NPD estimator in this setting, we construct the $X^{(\delta)}(t)$ process for various parameter settings and a range of δ based on the simulated P_i. We then compare the simulated $\mu(\delta)$ with δ^2, which describes the bias of the NPD estimator. To show this difference graphically, we plot the volatility signature plot Andersen, Bollerslev, Diebold, and Labys (2000) of the NPD estimator for a theoretical interval using the asymptotic property of $X^{(\delta)}(t)$. The volatility signature plot is constructed by plotting $\mathrm{E}[NPD(0,t)]=\frac{IV(0,t)}{\mu(\delta)}$ against δ for some finite $IV(0,t)$, and comparing it to the true integrated variance. The mean duration in business time can be simulated by collecting the number of transactions $M_i^{(\delta)}$ required to trigger the i-th price duration, and the mean duration $\mu(\delta)$ can be obtained as:

$$\mu(\delta)=\frac{\gamma\lambda}{N}\sum_{i=1}^{k}M_i^{(\delta)}, \tag{99}$$

in which k is the size of the simulation. Alternatively, it can be simulated by the renewal RV estimator (with a larger simulation error) as:

$$\mu(\delta)=\frac{1}{N}\sum_{i=1}^{k}(r_i^{(\delta)})^2. \tag{100}$$

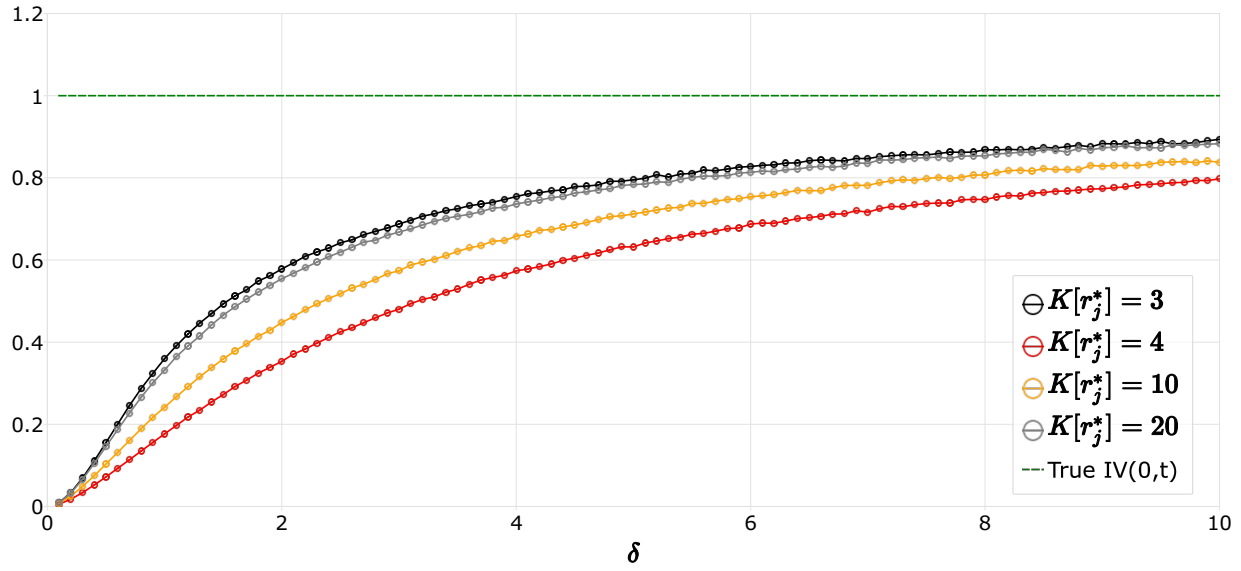
For each δ, we choose $k=100000$. All the parameters of the GST model are set for illustrative purposes only.

The first case we examine is the case where $V_i=0$ and $V[r_j^*]=1$. We set the kurtosis $K[r_j^*]$ to be 20, 10, 4 and 3 to examine the effect of an excess kurtosis on the NPD estimator.[6] The volatility signature plots

[6]For the first three cases, the corresponding parameter values for (γ, λ) are ($\frac{3}{47}$, $\frac{47}{3}$), ($\frac{3}{17}$, $\frac{17}{3}$) and (3, $\frac{1}{3}$) respectively. When $K[r_j^*]=3$, $\tilde{d}_j=1$ for all j so that r_j^* is i.i.d. normal.

of the NPD estimator under these parameter settings are presented in Figure 11. In the simulation we set $IV(0,t) = 1$ with δ ranging from 0.1 to 10 with a step size of 0.1. From Figure 11 we see that, as discussed in the previous section, the NPD estimator is downwardly biased in the absence of MMS noise due to $Bias_{TD}^{(\delta)}$, which is a function of $V[r_j^*]$ and kurtosis. Generally, holding the variance constant, r_j^* with heavier tails will have a larger truncation bias on average, as is shown in Figure 11. It is also clear that the bias decays slowly as δ increases, which corroborates our result in G.

Figure 11: Simulated volatility signature plot for the NPD estimator on the GST model with no MMS noise

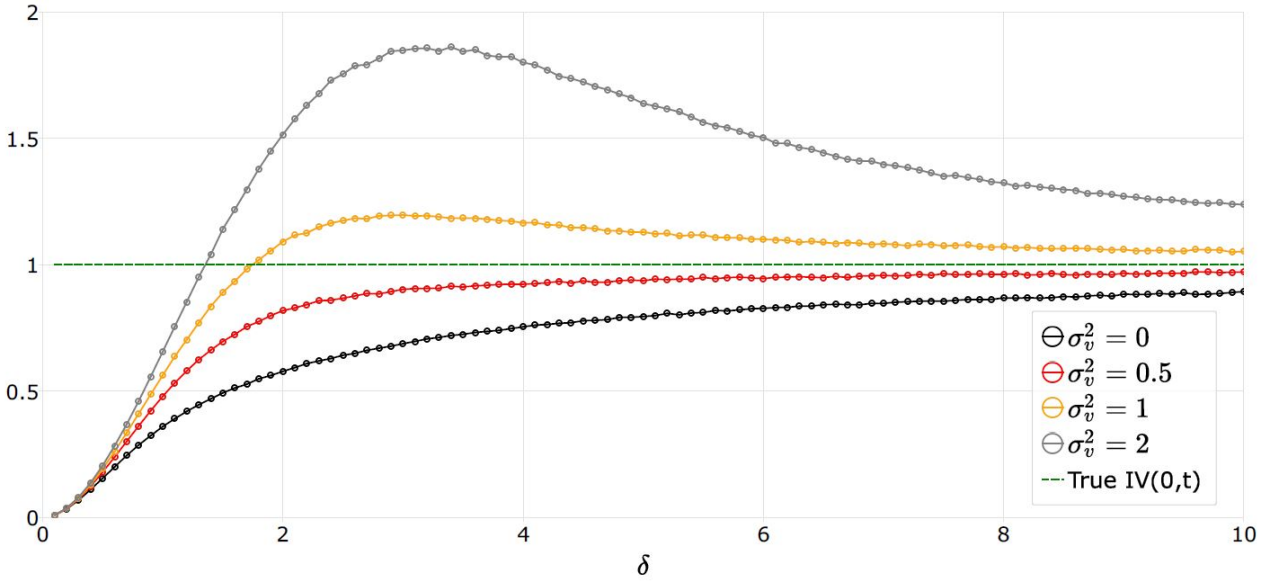


Note: We simulate $\mu(\delta)$ based on (99) for $\delta \in [0.1, 10]$ with a step size of 0.1. For each δ, the volatility signature curve is computed by $IV(0,t)\frac{\delta^2}{\mu(\delta)}$ where $IV(0,t) = 1$. Each circle represents a volatility estimate from the NPD estimator computed at the value of δ.

In the next case, we examine the effect of various sizes of i.i.d. MMS noise by choosing $\rho = 0$ with $\sigma_v^2 \in \{0, 0.5, 1, 2\}$. We will use the same parameter settings from the previous case with $K[r_j^*] = 4$ for illustration, as the effect of kurtosis is similar for both cases. The volatility signature plots are presented in Figure 12. The figure corroborates our previous discussion on the truncation bias and the MMS noise bias. From the graph, we see that when the size of the noise is small ($\sigma_v^2 \leq 0.5$), the MMS noise bias is smaller than the truncation bias and the volatility signature curve converges from below. When the size of the MMS bias is large enough to compensate for the truncation bias, the volatility signature curve has a hump shape and converges from above. This result is consistent with Figures 2, and 3 in Nolte, Taylor, and Zhao (2018), which document a similar curve with a different setting. Also, in the case where the MMS bias is large enough, we see that the volatility signature curve intersects the true $IV(0,t)$ at some finite δ so the NPD estimator is unbiased. Unfortunately, we are unable to derive an analytical form for this particular NPD estimator as we cannot estimate the amount of $Bias_{TD}^{(\delta)}$.

Figure 13 shows the case with AR(1) MMS noise. In the simulation we use the settings from the previous case with $\sigma_v^2 = 0.5$ and $\rho \in \{-0.9, -0.5, 0, 0.5, 0.9\}$, so that the unconditional variance of the noise remains unchanged. The figure shows that, negative autocorrelation inflates $Bias_{MMS}^{(\delta)}$ when δ is small, and affects the shape of the volatility signature plot. We can clearly see a hump-shaped volatility signature curve for $\rho = -0.9$. The effect of negatively correlation decays as δ increases, and the volatility signature curves converge to the i.i.d. noise case. The impact of positively correlated noise is more persistent and has less of an impact on the NPD estimator. However, in the positively correlated noise case the volatility signature curve deviates from

Figure 12: Simulated volatility signature plot for the NPD estimator on the GST model with i.i.d. MMS noise



Note: We simulate $\mu(\delta)$ based on (99) for $\delta \in [0.1, 10]$ with a step size of 0.1. For each δ, the volatility signature curve is computed by $IV(0,t)\frac{\delta^2}{\mu(\delta)}$ where $IV(0,t) = 1$. The GST model parameters are $\gamma = 3$ and $\lambda = 1/3$ for all four cases. Each circle represents a volatility estimate from the NPD estimator computed at the value of δ.

the i.i.d. case as ρ increases.

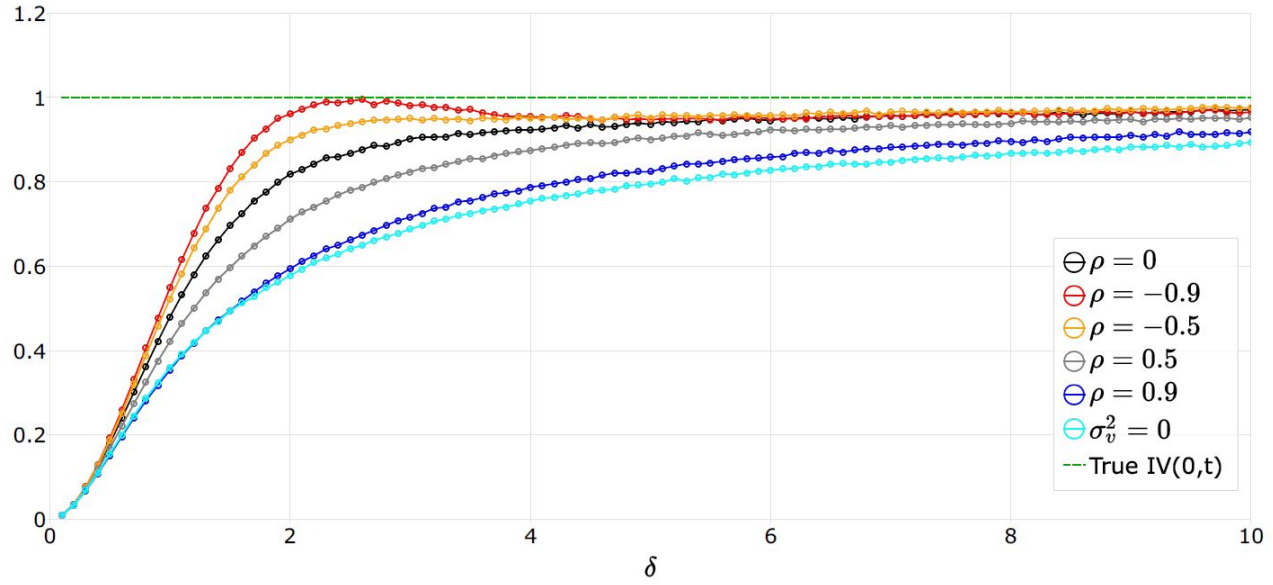
We also include price discretization in the example. As discussed in Section 6.5, we assume that r_j is discrete with the support $\{\cdots, -2\epsilon, -\epsilon, 0, \epsilon, 2\epsilon, \cdots\}$. We construct the same volatility signature plots for $\epsilon \in [0, 0.1, 0.5, 1]$, and construct the discrete log-price process $h_\epsilon(P_j)$, where $h_\epsilon(x) = \epsilon \text{nint}(\frac{x}{\epsilon})$ and $\text{nint}(x)$ is the nearest integer function. A slight complication arises in this situation. As the NPD estimator always samples in tick time, all the zero entries in r_j are completely disregarded. We choose the parameter settings for P_j from the previous AR(1) noise case with $\rho = -0.5$ and examine the effect of different levels of ϵ on the NPD estimator. The volatility signature plots in this case are presented in Figure 14.

Figure 14 reveals some very interesting features of the NPD estimator under price discretization. Comparing the case with $\epsilon = 0$ and $\epsilon = 0.1$, we see that the bias increases slightly as a result of the price discretization. When $\epsilon = 0.5$ or 1, the volatility signature curves have a zigzag pattern. As discussed in Section 6.5, this is due to the invariant sampling scheme for $\delta \in ((n-1)\epsilon, n\epsilon]$, so that $\mu(\delta)$ is also constant within the range. As a result, the NPD volatility estimates for $\delta \in (n-1)\epsilon, n\epsilon]$ will become a quadratic function of δ peaking at every $n\epsilon$. By sampling at $n\epsilon$, we obtain the volatility signature curve that has the least truncation bias, and this bias can be artificially increased by letting $\delta \downarrow (n-1)\epsilon$ without changing the properties of the sampling scheme. Therefore, if the magnitude of the MMS noise is large enough, one may be able to obtain solutions of δ^* for multiple n, represented by the multiple intersections between the volatility signature curves and the true IV for $\epsilon \geq 0.5$.

In the last case, we examine the effect of jumps and price discretization on the NPD estimator and assume that $V_i = 0$ for simplicity. The discrete price process with jumps is specified as follows:

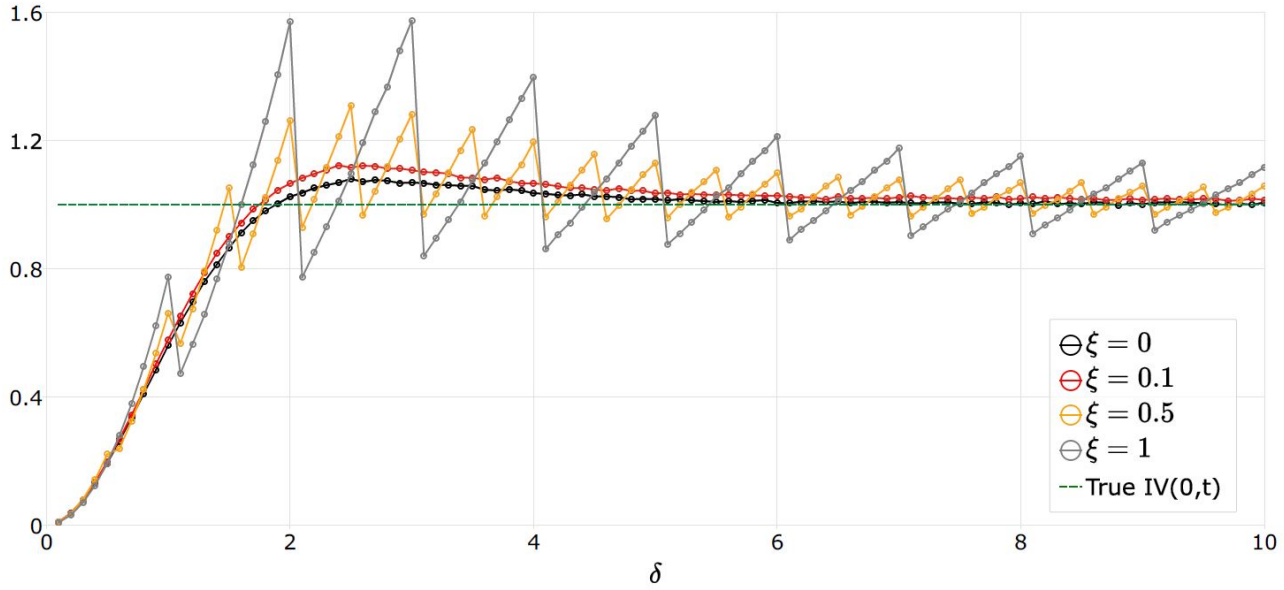
$$\begin{aligned} \tilde{P}_j &= h_\epsilon(P_j^*) + L \cdot L_j \cdot J_j \\ L_j &\sim i.i.d. Bernoulli(p), J_j \sim i.i.d. Rademacher \end{aligned} \tag{101}$$

Figure 13: Simulated volatility signature plot for the NPD estimator on the GST model with AR(1) MMS noise



Note: We simulate $\mu(\delta)$ based on (99) for $\delta \in [0.1, 10]$ with a step size of 0.1. For each δ, the volatility signature curve is computed by $IV(0,t)\frac{\delta^2}{\mu(\delta)}$ where $IV(0,t) = 1$. The GST model parameters are $\gamma = 3$, $\lambda = 1/3$ and $\sigma_v^2 = 0.5$ for all five cases. Each circle represents a volatility estimate from the NPD estimator computed at the value of δ.

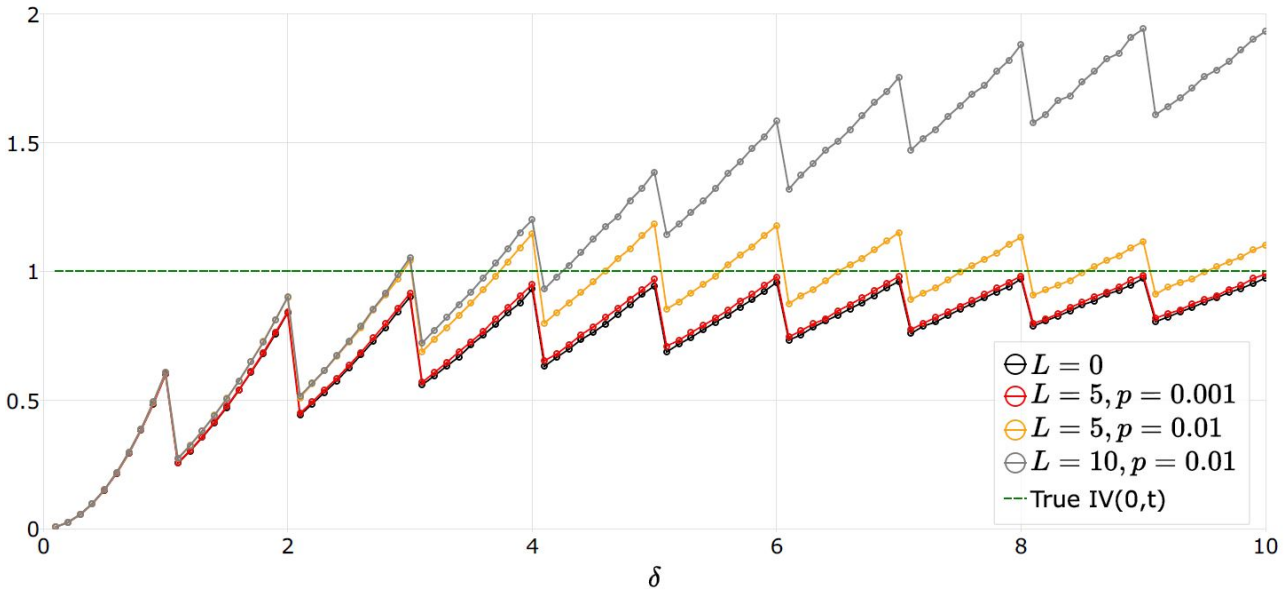
Figure 14: Simulated volatility signature plot for the NPD estimator on the GST model with AR(1) MMS noise and price discretization



Note: We simulate $\mu(\delta)$ based on (99) for $\delta \in [0.1, 10]$ with a step size of 0.1. For each δ, the volatility signature curve is computed by $IV(0,t)\frac{\delta^2}{\mu(\delta)}$ where $IV(0,t) = 1$. The GST model parameters are $\gamma = 3$, $\lambda = 1/3$, $\rho = -0.5$ and $\sigma_v^2 = 0.5$ for all five cases. Each circle represents a volatility estimate from the NPD estimator computed at the value of δ.

In this simple setting, L is the size of each jump which is assumed to be a constant, L_j is a Bernoulli draw on each arrival of transaction representing the arrivals of jumps, and J_j determines the direction of the jump. We plot the simulated volatility signature plot in this case in Figure 15. From the figure, we see that both L and p influences the bias of the NPD estimator. We see that when δ is very small, the four curves coincide, which proves our previous theoretical result on the jump effect. As the jump size and jump intensity increase, the NPD estimator absorbs more jump variation and are also affected. In the extreme case with $L = 10$, we see that the NPD estimator diverges from the true IV. However, the jump intensity used here (one per 100 transaction) is highly unlikely in reality (as opposed to less than one per week as documented in Andersen, Bollerslev, and Dobrev (2007) and Lee and Hannig (2010)).

Figure 15: Simulated volatility signature plot for the NPD estimator on the GST model with price discretization and jumps



Note: We simulate $\mu(\delta)$ based on (99) for $\delta \in [0.1, 10]$ with a step size of 0.1. For each δ, the volatility signature curve is computed by $IV(0,t)\frac{\delta^2}{\mu(\delta)}$ where $IV(0,t) = 1$. The GST model parameters are $\gamma = 3$, $\lambda = 1/3$ and $V_j = 0$ for all four cases. Jumps are specified as (101). Each circle represents a volatility estimate from the NPD estimator computed at the value of δ.

I Moment conditions for r_j of the GST model

$$\mathrm{E}[r_j] = 0, \tag{102}$$

$$\mathrm{V}[r_j] = \gamma\lambda + \frac{2\sigma_v^2}{1+\rho}, \tag{103}$$

$$\mathrm{E}[r_j r_{j-k}] = \frac{\rho-1}{\rho+1}\rho^{k-1}\sigma_v^2, \tag{104}$$

$$\mathrm{E}[r_j^4] = 3\gamma\lambda^2(1+\gamma) + \frac{12\gamma\lambda\sigma_v^2}{1+\rho} + \frac{12\sigma_v^4}{(1+\rho)^2}. \tag{105}$$

J Implementation details for RK, NPD^z, PRV and PBip estimators

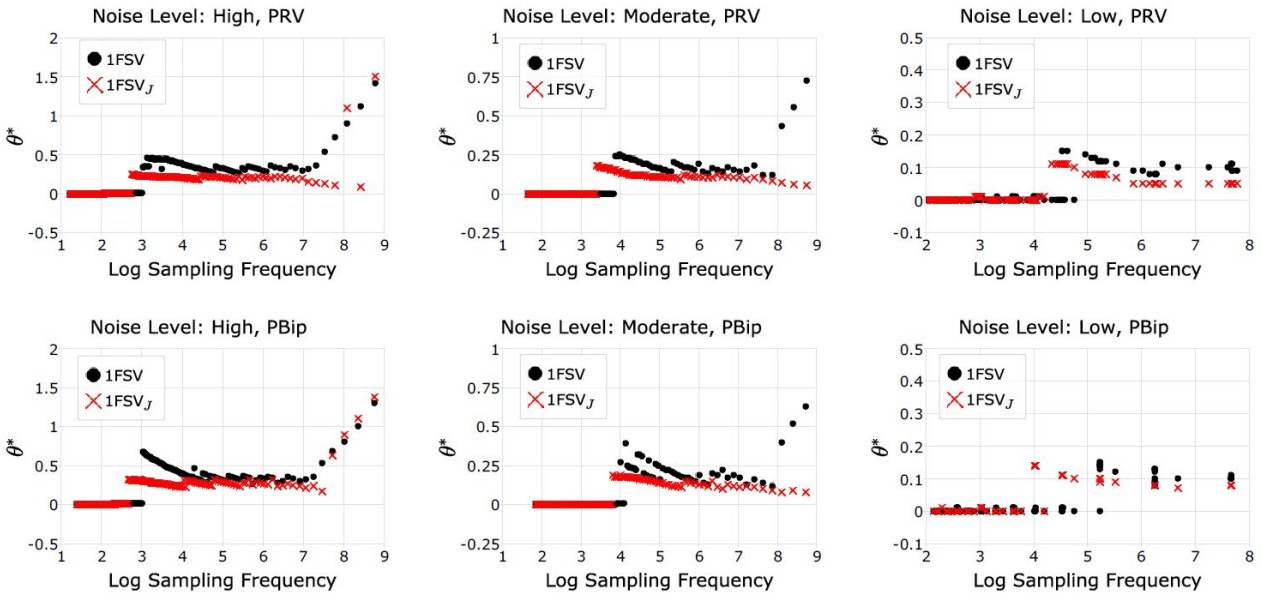
For the RK estimator, we use a Tukey-Hannings$_2$ kernel, with the optimal bandwidth $H = 5.74\xi N^{0.5}$, in which $\xi = \frac{\sigma_v^2}{\sqrt{\int_0^1 \sigma_p^4(s)ds}}$ and N is the sampling frequency, as given by Barndorff-Nielsen, Hansen, Lunde, and Shephard

(2008). In the simulation, we use the true value of σ_v^2 and $\int_0^1 \sigma_p^4(s)$ for each random draw of the 1FSV model to construct the RK estimator. The RK estimator is then constructed based on the calendar time sampled returns with the sampling frequency given by the average sampling frequency of the NPD^z estimator (or the NPD_J^z estimator in the presence of jump) for each $\delta = x\delta_0$. An example of the sampling frequencies for the moderate level of noise case is shown in Figure 9.

For the NPD^z estimator, we choose the optimal smoothing parameter by computing the MSE of the NPD^z estimator based on 10000 random draws of the 1FSV and 1FSV$_\text{J}$ models for a grid of $\delta = x\delta_0$, and choose the γ that minimizes the MSE of the NPD^z estimator for some δ. The resulting optimized γ^*s for the 1FSV model with high, moderate and low levels of noise are 0.1, 0.30 and 0.92 respectively. For the 1FSV$_\text{J}$ model, the corresponding γ^*s are: 0.1, 0.31 and 0.99. The γ^*s for the 1FSV model are very close to those of the 1FSV$_\text{J}$ model with the same noise level. We also see that the smoothing parameter is reversely related to the size of noise as expected. The NPD^z estimator is then constructed on a grid of $\delta = x\delta_0$ on the smoothed price process Z_j.

For the PRV and PBip estimators, we need to determine the tuning parameter θ that controls for the window width of pre-averaging (see e.g. Jacod, Li, Mykland, Podolskij, and Vetter (2009) or Hautsch and Podolskij (2013)), and σ_v^2 to correct for the pre-averaged MMS bias. As the optimal value of θ varies with the sampling frequency according to Hautsch and Podolskij (2013), we optimize θ for each sampling frequency used in order to obtain optimized performance for the PRV and PBip estimators at each sampling frequency. In detail we use a grid of $\theta \in [0, 2]$ to construct both estimators and to choose an optimal θ^* that minimizes the MSE of the estimator at each sampling frequency. Note that when $\theta^* = 0$, we use RV and RBip instead. We plot the optimal θ^*s of PRV and PBip for both the 1FSV and 1FSV$_\text{J}$ models under three different levels of noise in Figure 16:

Figure 16: Optimal θs of PRV and PBip estimators



Note: The results are based on 10000 replications of the 1FSV and 1FSV$_\text{J}$ models. For each black dot, the x-axis shows the log sampling frequency used to construct PRV and PBip estimators and the y-axis represents the optimized value for θ. For each sampling frequency, θ^* is computed by a grid search method for $\theta \in [0, 2]$ that minimizes the simulation MSE.

Figure 16 shows that the optimal θs indeed vary with the sampling frequency. Generally, a much larger θ

is required for the highest sampling frequency, and for the sampling frequency within $exp(4)$ to $exp(7)$, θ is very stable. When the sampling frequency decreases further, θ quickly drop to zero, as the simple RV and RBip estimators have better efficiency when the impact of MMS noise is small. The presence of jump seems to decrease the optimal θ^* slightly, but the optimal θ^*s have a similar pattern with or without jumps.

K Additional Tables and Figures

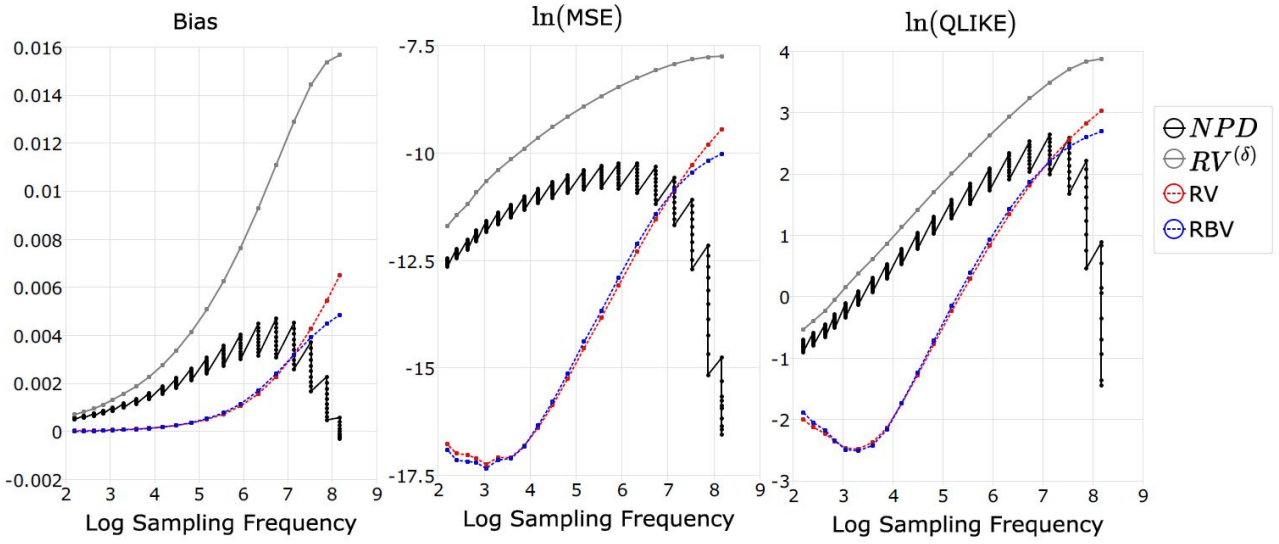
Table 4: Comparison of the optimal QLIKEs for all volatility estimators in Table 1 for the 1FSV and 1FSV$_\text{J}$ models with low, moderate and high levels of noise

Estimator	NPD	$RV^{(\delta)}$	RV	RBip	NPD^z	RK	PRV	PBip
1FSV model with low level of noise								
Optimal log QLIKE	-5.3841	-3.1229	-4.1165	-4.3328	-5.4041	-5.1137	-4.8457	**-5.5014**
δ/δ_0	24	61	21	21	25	9	11	8
Sampling Freq.	189	26	189	189	189	2160	524	2160
1FSV model with moderate level of noise								
Optimal log QLIKE	-2.7444	-2.1163	-3.3983	-3.4633	**-5.3551**	-4.6682	-4.8649	-4.9156
δ/δ_0	91	131	51	51	27	9	6	6
Sampling Freq.	19	8	84	84	594	2955	2955	2955
1FSV model with high level of noise								
Optimal log QLIKE	-1.4389	-0.5337	-2.4789	-2.5070	**-5.1677**	-3.5907	-4.4197	-4.3951
δ/δ_0	5	191	141	141	30	14	6	6
Sampling Freq.	3529	9	27	27	1845	2618	3529	3529
1FSV$_\text{J}$ model with low level of noise								
Optimal log QLIKE	-5.4757	-0.1705	-0.2015	-2.8499	**-5.4833**	-0.2122	-0.3157	-4.0977
δ/δ_0	24	111	51	11	24	14	17	1
Sampling Freq.	187	9	38	517	187	517	517	2142
1FSV$_\text{J}$ model with moderate level of noise								
Optimal log QLIKE	-2.3990	-0.0655	-0.2012	-2.0102	**-4.6502**	-0.1925	-0.3497	-3.4153
δ/δ_0	81	141	81	51	18	18	15	7
Sampling Freq.	26	8	26	83	1322	1322	1322	2929
1FSV$_\text{J}$ model with high level of noise								
Optimal log QLIKE	-1.3215	0.4063	-0.0663	-1.1812	**-4.0293**	-0.0827	-0.1327	-2.7092
δ/δ_0	4	191	161	131	20	19	7	7
Sampling Freq.	3491	9	16	33	2524	2524	3491	3491

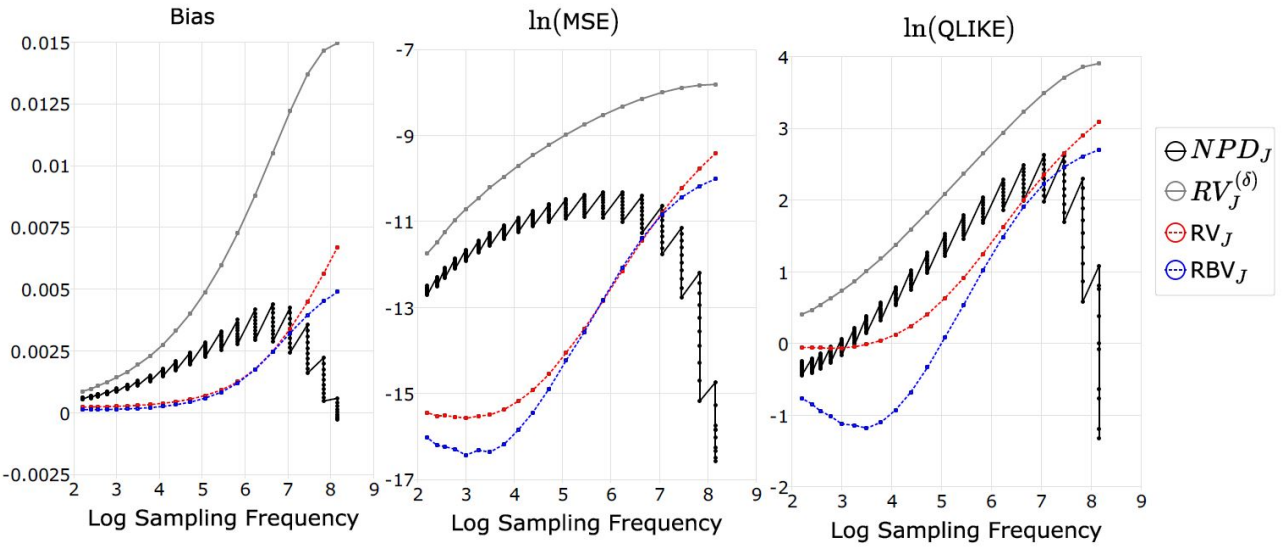
Note: Optimal log QLIKE for an estimator is the smallest log QLIKE among all the sampling frequencies considered. The smallest value is highlighted in bold. The entries for the rows $\delta = x\delta_0$ represents the value of the threshold as multiples of $\delta_0 = 0.1\epsilon$, with $\epsilon = \ln(20.01) - \ln(20)$. The sampling frequency is the average sampling frequency at the optimal δs for NPD, $RV^{(\delta)}$ and NPD^z, and is the calendar time sampling frequency for RV, RBip, RK, PRV and PBip.

Figure 17: Simulated Bias, MSE and QLIKE for daily volatility estimates obtained from NPD, $RV^{(\delta)}$, RV and RBip for 1FSV model with high level of MMS noise

Panel 1: 1FSV No Jump



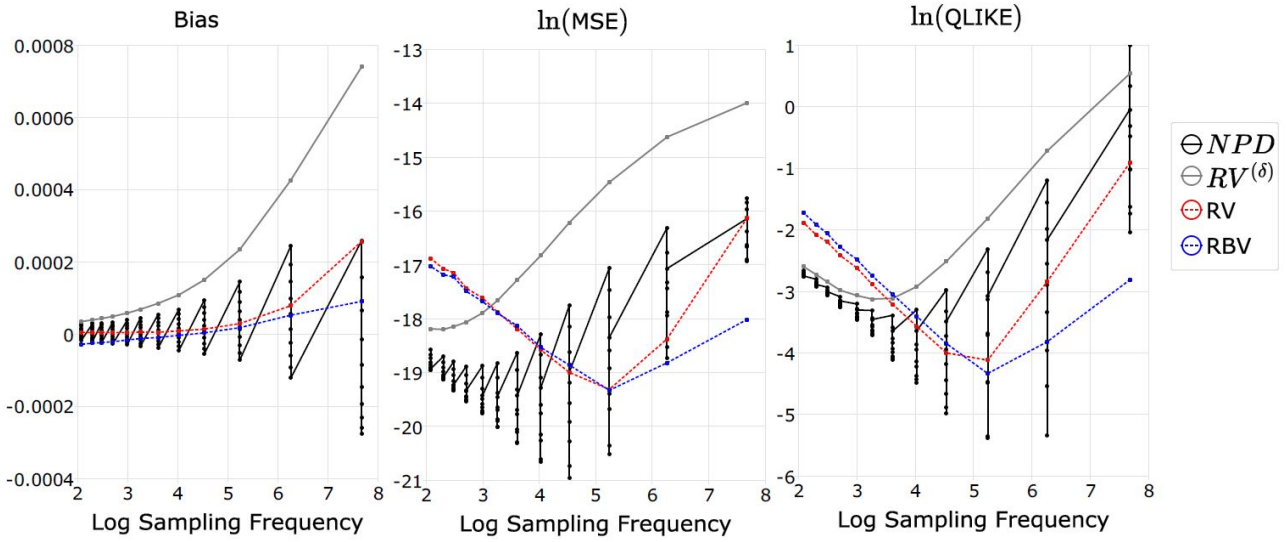
Panel 2: $1FSV_J$



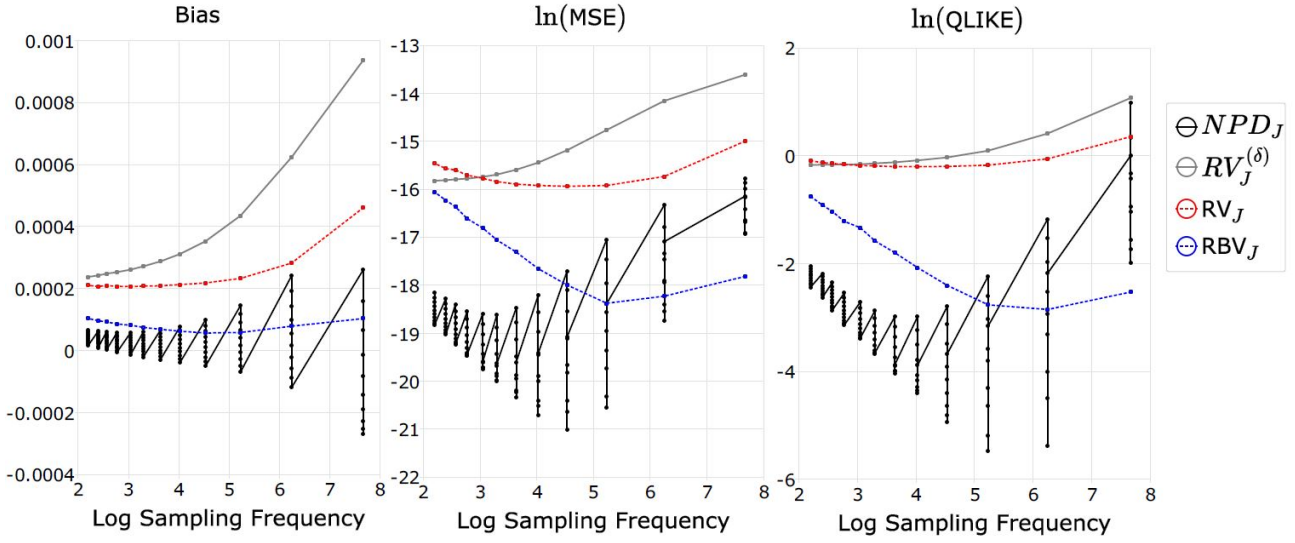
Note: The results are based on 10000 replications of the 1FSV model with and without jumps. The x-axis denotes the average log sampling frequency for a given δ for NPD and $RV^{(\delta)}$, or the log sampling frequency of the equidistant intraday return per day for RV and RBip. The truncation threshold δ ranges from $200\delta_0$ to δ_0 with a step size of $\delta_0 = 0.1\epsilon$, with $\epsilon = \ln(20.01) - \ln(20)$. The subscript J represents an estimator constructed on the 1FSV model with jumps. The noise-to-signal ratio is set to be $\omega = 0.005$.

Figure 18: Simulated Bias, MSE and QLIKE for daily volatility estimates obtained from NPD, $RV^{(\delta)}$, RV and RBip for 1FSV model with low level of MMS noise

Panel 1: 1FSV No Jump



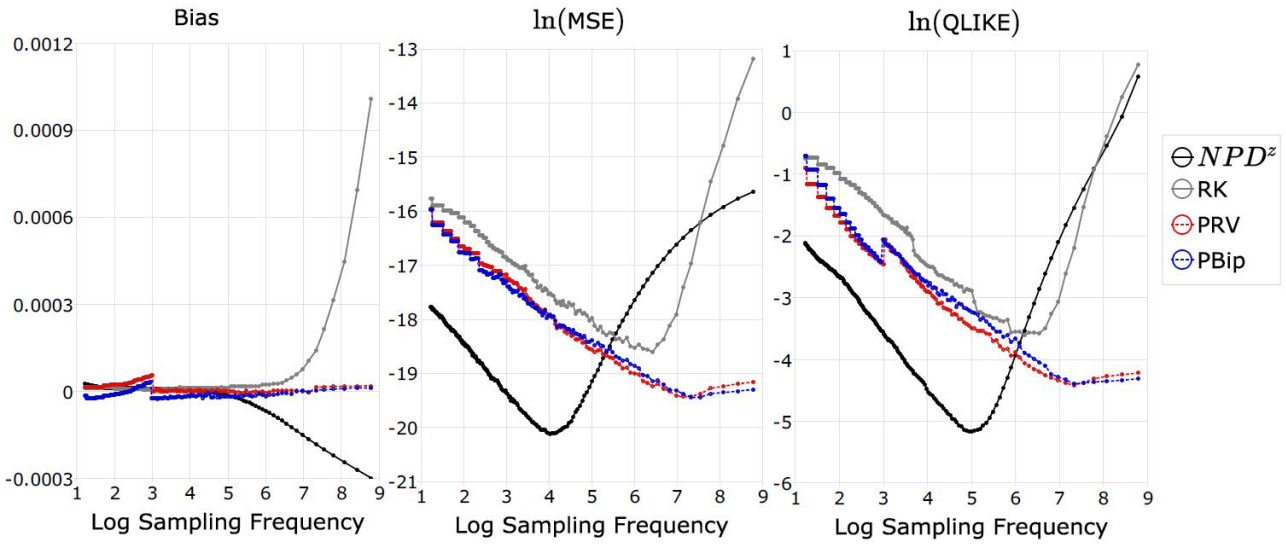
Panel 2: $1FSV_J$



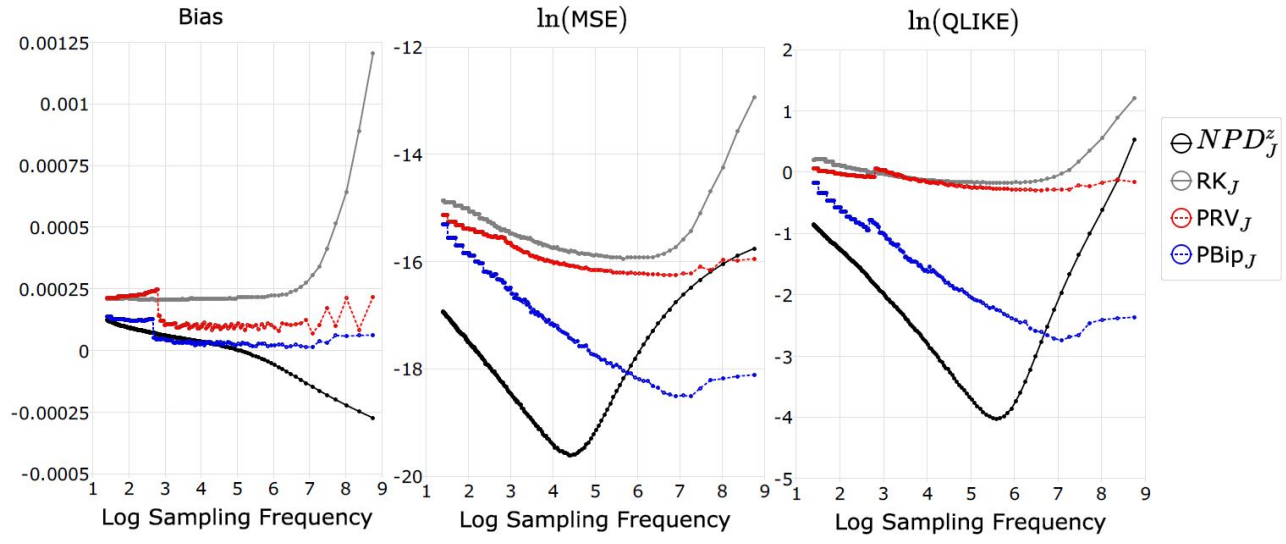
Note: The results are based on 10000 replications of the 1FSV model with and without jumps. The x-axis denotes the average log sampling frequency for a given δ for NPD and $RV^{(\delta)}$, or the log sampling frequency of the equidistant intraday return per day for RV and RBip. The truncation threshold δ ranges from $\delta_1 20$ to δ_0 with a step size of $\delta_0 = 0.1\epsilon$, with $\epsilon = \ln(20.01) - \ln(20)$. The subscript J represents an estimator constructed on the 1FSV model with jumps. The noise-to-signal ratio is set to be $\omega = 0.0002$.

Figure 19: Simulated Bias, MSE and QLIKE for daily volatility estimates obtained from NPD^z, RK, PRV and PBip for 1FSV model with high level of MMS noise

Panel 1: 1FSV No Jump



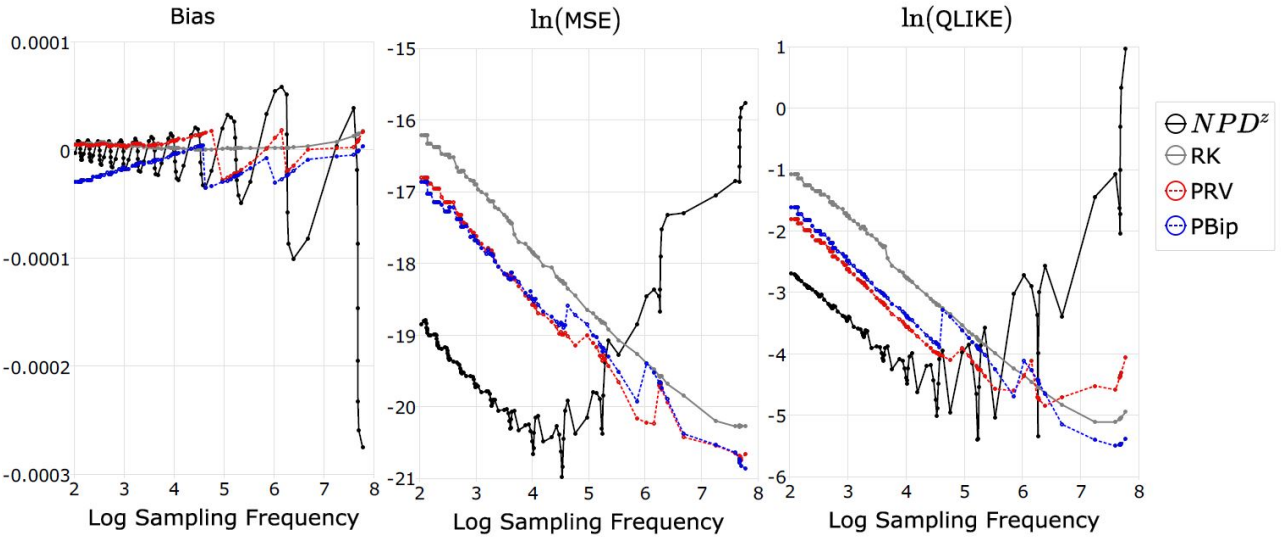
Panel 2: $1FSV_J$



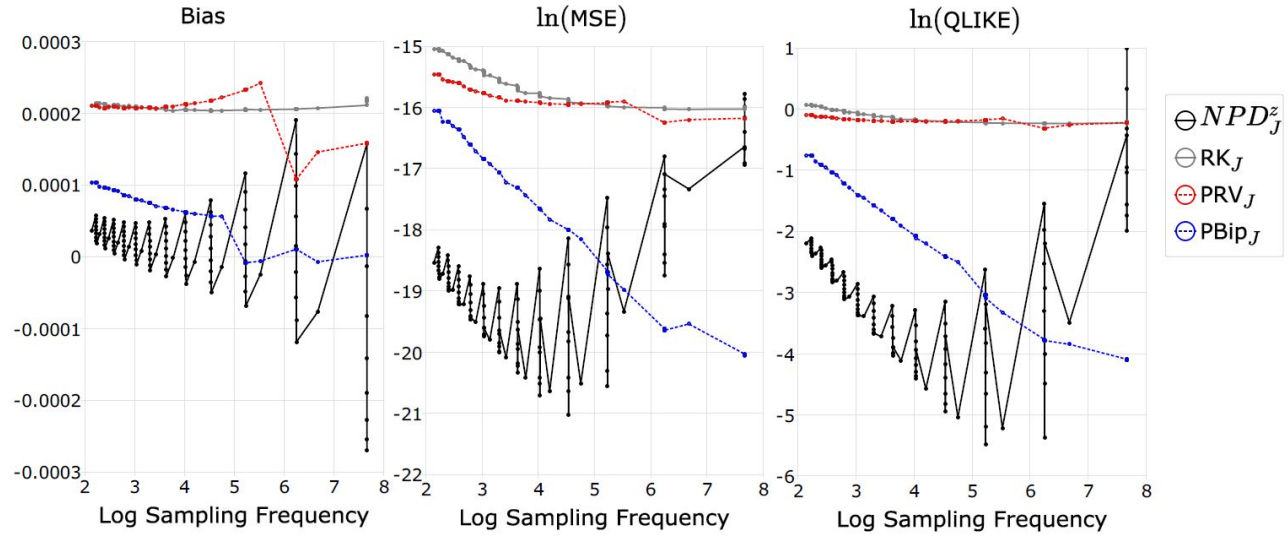
Note: The results are based on 10000 replications of the 1FSV model with and without jumps. The x-axis denotes the average log sampling frequency for a given δ for the NPD^z model, or the log sampling frequency of the equidistant intraday return per day for RK, PRV and PBip. The truncation threshold δ ranges from $\delta_1 50$ to δ_0 with a step size of $\delta_0 = 0.1\epsilon$, with $\epsilon = \ln(20.01) - \ln(20)$. The subscript J represents an estimator constructed on the 1FSV model with jumps. The noise-to-signal ratio is set to be $\omega = 0.005$.

Figure 20: Simulated Bias, MSE and QLIKE for daily volatility estimates obtained from NPD^z, RK, PRV and PBip for 1FSV model with low level of MMS noise

Panel 1: 1FSV No Jump



Panel 2: $1FSV_J$



Note: The results are based on 10000 replications of the 1FSV model with and without jumps. The x-axis denotes the average log sampling frequency for a given δ for the NPD^z model, or the log sampling frequency of the equidistant intraday return per day for RK, PRV and PBip. The truncation threshold δ ranges from $\delta_1 50$ to δ_0 with a step size of $\delta_0 = 0.1\epsilon$, with $\epsilon = \ln(20.01) - \ln(20)$. The subscript J represents an estimator constructed on the 1FSV model with jumps. The noise-to-signal ratio is set to be $\omega = 0.0002$.