Bias Reduction in Spot Volatility Estimation from Options

Viktor Todorov*and Yang Zhang†

December 2, 2021

Abstract

We consider the problem of nonparametric spot volatility estimation from options that is robust to time-variation in volatility and presence of jumps in the underlying asset price. Using a higher-order expansion of the characteristic function of the underlying price increment over shrinking time intervals and option-based estimates of the latter over two distinct horizons, we achieve asymptotic bias-reduction in spot volatility estimation, relative to existing methods, that is due to time-variation in volatility and presence of jumps. Further asymptotic improvement is achieved by de-biasing the volatility estimator using an estimate for the bias in it due to the small jumps in the price process. The gains from the newly-developed volatility estimation approach are illustrated on simulated data and in an empirical application.

Keywords: characteristic function, higher-order asymptotic expansion, jumps, options, volatility estimation.
JEL classification: C51, C52, G12.

*Kellogg School of Management, Northwestern University; e-mail: v-todorov@northwestern.edu.

†Evanston, IL 60201; e-mail: yang.zhang1@kellogg.northwestern.edu. Yang Zhang completed the majority of this work as a postdoc in the Finance Department, Kellogg School of Management, Northwestern University, and the rest on her personal time. The views expressed in the article are the author's own and not those of her current employer.

1 Introduction

Two well-documented salient features of financial data are stochastic volatility and jumps. Their presence in asset returns complicates significantly inference for asset pricing models when the inference is based on low frequency return data. The main reason for this is that stochastic volatility and jumps are hidden in asset returns, i.e., they cannot be estimated consistently from the low frequency returns and have to be treated as latent instead. One way to solve this latency problem is to use return data sampled at high frequency. The high frequency price record facilitates the construction of nonparametric volatility and jump estimators, which in turn can allow for inference for asset pricing models as if stochastic volatility and jumps are directly observable. Examples of jump-robust variance estimators[1] include the multipower variation of Barndorff-Nielsen and Shephard (2004, 2006) and their extensions proposed by Corsi et al. (2010) and Andersen et al. (2012), the truncated variance of Mancini (2001, 2009) and the empirical characteristic function based variance estimators of Jacod and Todorov (2014, 2018).[2]

The above jump-robust variance estimators have been designed with the goal of estimating the integrated diffusive variance over a fixed time interval. They can be readily extended to estimating the spot volatility at a given point in time by considering an asymptotically shrinking time window in the estimation as in Foster and Nelson (1996).[3] This, of course, leads to a drop in the rate of convergence of the estimators and makes spot volatility estimation from available high-frequency return data very imprecise. Options written on the asset provide an alternative source of information which can significantly reduce the noise in measuring spot volatility. The main challenge in option-based nonparametric estimation, however, is the bias in the volatility estimators due to the time-variation in volatility over the life of the option and the presence of jumps in the underlying asset dynamics. The goal of the current paper is to develop bias-reduction techniques for option-based volatility estimation. This can make the latter practically feasible in more general settings (in terms of option tenors) than previously considered.

Options provide a natural source of information for the volatility of the underlying process. In the classical option pricing model of Black and Scholes (1973), in which the asset price follows a geometric Brownian motion, the option price, normalized by the current value of the underlying

[1] In this paper, spot (diffusive) volatility refers to the instantaneous standard deviation of the continuous component of the (log) asset price while spot variance is the square of the spot volatility. Formal definitions are given in Section 2 below.

[2] A number of papers have extended the results of Barndorff-Nielsen and Shephard (2004, 2006) and Mancini (2001, 2009) in various directions, both by allowing for more general dynamics of the discretely observed process as well as by allowing for observation error (microstructure noise) and more general sampling schemes.

[3] Alternative spot volatility estimators in various settings have been considered by Bandi and Phillips (2003), Fan and Wang (2008), Kristensen (2010), Liu et al. (2018) and Bibinger et al. (2019), among others.

asset, is a known and nonlinear function of the volatility parameter (volatility is constant in the Black-Scholes model). Therefore, one can invert the observed option prices and recover the unobservable (from the point of the econometrician) volatility parameter. The resulting estimate is referred to as Black-Scholes Implied Volatility (BSIV) and is commonly used by market participants to quote option prices. In the context of the Black-Scholes model, BSIV provides, up to the measurement error in observed option prices, an exact estimate of the true volatility.

Unfortunately, the assumptions behind the Black-Scholes model are too strong and do not hold empirically. Mainly, volatility changes over time and assets are exposed to jump risk. This makes nonparametric identification of spot volatility from options rather challenging. However, if the time to maturity of the options is shrinking asymptotically to zero, then such spot volatility inference is possible. The asymptotic setting of observing options with shrinking time to maturity can be viewed as the counterpart of the asymptotic setup of high frequency sampling of the asset price in a fixed time interval. Importantly, there is a recent trend in options markets of exchanges offering options with very short times to maturity and there is an increase in the trading and liquidity of these contracts, see e.g., Andersen et al. (2017), making such an asymptotic setup of options with shrinking time to maturity of direct practical interest.

The main challenge of the option-based inference for the spot volatility is to design a suitable transformation of the options that can be a consistent estimator for it as the options' time to maturity shrinks. Options are conditional expectations of their terminal payoff, discounted at the risk-free rate, under the so-called risk-neutral probability measure. The latter is locally equivalent to the true probability measure and the difference between the two probability measures is due to the pricing of risk by investors. Importantly for the analysis here, the local equivalence of the statistical and risk-neutral probability measures implies that the diffusive volatility (the object of interest in this paper) is the same under the two measures. To design an option-based volatility estimator, therefore, we need to establish an asymptotic one-to-one map between conditional expectations of suitably chosen transforms of future returns and the current level of diffusive volatility, as the time interval of the return shrinks. Thus, the major challenge of using options to infer spot volatility is the bias contained in such estimators.

A natural candidate for spot volatility estimator is the BSIV of a short-dated at-the-money option, i.e., an option with strike equal to the current value of the asset price. Indeed, Medvedev and Scaillet (2007) (see also Medvedev and Scaillet (2010)) and Durrleman (2008), among others, show that this is a consistent estimator of spot volatility under weak conditions for the dynamics of the underlying process, and in particular in presence of jumps, when the option's tenor shrinks.

Unfortunately, as shown in Medvedev and Scaillet (2007), BSIV viewed as an estimator of spot volatility has nontrivial bias due to the time-variation in volatility and the presence of jumps in the asset price for times to maturities that are practically feasible.

Alternative nonparametric estimators of spot volatility from the short-dated options can be constructed by making use of the fact that options on a continuum set of strikes can uniquely identify the risk-neutral return distribution, see e.g., Breeden and Litzenberger (1978). In particular, following Carr and Madan (2001), portfolios of options can be used to infer conditional expectations under the risk-neutral probability measure of smooth transforms of the return over the life of the option. Todorov (2019) uses this result to propose a nonparametric estimator of spot volatility by constructing an estimator of the conditional characteristic function of the asset return over the life of the options and using the dominant role played by the diffusive volatility at asymptotically increasing values of the characteristic exponent.

As shown in Todorov (2019), the characteristic function based approach for volatility inference leads to a bias of much smaller asymptotic order of magnitude than that in the BSIV. In this paper, we improve on the estimator of Todorov (2019) by designing bias-reduction techniques for volatility inference. More specifically, we use a higher-order expansion of the characteristic function of the price increments to develop a volatility estimator that has significantly less bias due to time-varying volatility and presence of jumps in the asset dynamics. In particular, we utilize the fact that the leading term in the asymptotic expansion of the characteristic exponent of the price increment is linear in the length of the increment while the leading term from the volatility dynamics is proportional to the square of the increment's length. Therefore, by suitably differencing characteristic-function based volatility estimates formed from two different short option tenors, which is reminiscent to Jacknife bias-correction in statistics, we can minimize the impact of the time-varying volatility on the estimation. An additional benefit of this approach is that it leads asymptotically to a smaller bias due to jumps in the recovery of spot volatility.[4]

To further minimize the impact of jumps on the volatility estimation, we expand asymptotically the bias in the above volatility estimator due to the jumps. The leading term in this expansion is a power function of the characteristic exponent. This suggests an easy way to annihilate this bias. Mainly, by performing a nonlinear least squares estimation in characteristic exponent space

[4]Similar to the analysis here, one can consider bias-correction methods for BSIV, following higher-order expansions of BSIV, as in Medvedev and Scaillet (2007) and Figueroa-Lopez and Olafsson (2016). However, such expansions seem difficult to derive under general volatility dynamics and/or general specification for the jump part of the process (e.g., under presence of multiple volatility factors, general types of jumps in terms of jump activity and jump dynamics, etc). Moreover, the higher-order terms in such expansions (up to the order considered here) are more, which would imply more tenors needed for de-biasing, and de-biasing for the jumps does not appear easy (or even possible) to achieve up to the precision considered here.

in which the intercept is the square of the spot volatility estimator. The resulting estimator has a bias of significantly smaller asymptotic order than existing nonparametric option-based volatility estimates. This, in particular, implies better performance of the estimator for longer option tenors than previously feasible.

We evaluate the finite sample behavior of the developed inference methods on simulated data. The results from the Monte Carlo are in line with our theoretical findings. Mainly, our de-biasing procedures work well in accounting for the time-varying volatility and the presence of jumps in the underlying asset price on the spot volatility recovery from options in a wide range of Monte Carlo settings. Upon implementing the volatility estimators on S&P 500 index option data, we find that the bias-correction accounting for the volatility dynamics yields less downward-biased volatility estimator in high volatility regimes and the opposite in low volatility regimes. The bias-correction for the price jumps, on the other hand, results in a slightly lower level of estimated diffusive volatility, signifying the importance of small jumps in volatility inference. The de-biased option-based volatility estimator matches on average the level of return-based estimates of spot volatility in various volatility regimes but is significantly less noisy than them.

The rest of the paper is organized as follows. In Section 2, we introduce nonparametric estimation of spot volatility via characteristic functions. In Section 3 we propose bias-reduction techniques in volatility estimation using higher-order asymptotic expansion of the characteristic function of the price increment. Additional bias-reduction techniques, using inference for the small jumps, are developed in Section 4. These bias-reduction results are made feasible using option data in Section 5. Section 6 contains a Monte Carlo study and Section 7 an empirical application. Section 8 concludes. Technical assumptions and the proofs are given in Section 9.

2 Volatility Estimation via Characteristic Functions

The asset price process is denoted with X and the logarithm of it with x. The price process is defined on the sample space Ω, with the associated σ-algebra $\mathcal{F}$, and $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$ being the filtration. We will consider two probability measures, one being the true (statistical) one, denoted with $\mathbb{P}$, and the other one being the risk-neutral one, denoted with $\mathbb{Q}$. The latter, under the weak condition of arbitrage-free asset prices, is locally equivalent to the true one. The significance of $\mathbb{Q}$ stems from the fact that the discounted at the risk-free rate payoff process of any asset is a local martingale under $\mathbb{Q}$. We will use this result to connect the value of derivatives written on the the asset with the asset's dynamics (under $\mathbb{Q}$). More specifically, the dynamics of x under $\mathbb{Q}$ is given by

$$dx_t = \alpha_t dt + \sigma_t dW_t + \int_{\mathbb{R}} x\mu(dt, dx), \tag{1}$$

where W is a $\mathbb{Q}$ Brownian motion and μ is an integer-valued random measure on $\mathbb{R}_+ \times \mathbb{R}$ with $\mathbb{Q}$ jump compensator $dt \otimes \nu_t(x)dx$, for some measure ν_t satisfying $\int_{\mathbb{R}} |x|\nu_t(x)dx < \infty$. We note that local equivalence of $\mathbb{P}$ and $\mathbb{Q}$ implies that x obeys the same dynamics under $\mathbb{P}$ but with a different drift coefficient $\alpha_t^{\mathbb{P}}$ and a different jump compensator $dt \otimes \nu_t^{\mathbb{P}}(x)dx$. Importantly, however, the diffusion coefficient σ_t, which is the object of interest of this paper, is the same under the two probability measures. Therefore, whether expectation is taken under $\mathbb{P}$ or $\mathbb{Q}$ will make no difference as far as nonparametric inference about σ_t, from small-time increments of x, is concerned. For this reason, in the development of our inference procedures in this section and Sections 3 and 4, we will not specify whether expectation is taken under $\mathbb{P}$ or $\mathbb{Q}$, as the results henceforth will apply under both measures.

We refer to σ_t as spot volatility and to σ_t^2 as spot variance. Our spot volatility inference will be based on estimates of the conditional characteristic function of $x_{t+T} - x_t$ for asymptotically shrinking $T \downarrow 0$. Lets denote

$$\mathcal{L}_{t,T}(u) = \mathbb{E}_t \left(e^{iu(x_{t+T}-x_t)/\sqrt{T}} \right), \ u \in \mathbb{R}. \tag{2}$$

In Section 5 we will show how to estimate $\mathcal{L}_{t,T}(u)$ from options observed at time t. Our goal in this and the following two sections is to design efficient way of recovering σ_t from $\mathcal{L}_{t,T}(u)$. We begin with an asymptotic expansion of $\mathcal{L}_{t,T}(u)$ shown in Jacod and Todorov (2014) and Todorov (2019). In particular, under weak regularity conditions for x in (1), we have

$$\mathcal{L}_{t,T}(u) = \exp\left(-\frac{u^2}{2}\sigma_t^2 \right) + O_p\left(T^{1-r/2} \right), \text{ as } T \downarrow 0, \tag{3}$$

where the constant r is a bound on the jump activity. More specifically, r is such that the jump compensator satisfies

$$\int_{\mathbb{R}} (|x|^r \vee |x|^2)\nu_s(x)dx < \infty, \text{ for } s \text{ in a neighborhood of } t. \tag{4}$$

In the most general semimartingale case, r can take values in $[0, 2]$. In our dynamics for x above, as is common in most applications, we assume that jumps are of finite variation which implies $r \leq 1$. We note also that $r = 0$ corresponds to the most common case of finite activity jumps, whereas on each finite interval, x contains at most a finite number of jumps, almost surely.

Jacod and Todorov (2014) and Todorov (2019) used this result to propose the following estimate of the spot variance:[5]

$$V_{t,T}(u) = -\frac{2}{u^2} \log |\mathcal{L}_{t,T}(u)|. \tag{5}$$

[5]Another application of this result is to test against pure-jump dynamics, see Kong et al. (2015).

In Jacod and Todorov (2014), $\mathcal{L}_{t,T}(u)$ is replaced with its empirical counterpart from high-frequency return observations while Todorov (2019) uses estimate for it from the options data. In the latter case, the conditional characteristic function is computed under the risk-neutral probability measure while in the former case the empirical characteristic function is an estimate of the $\mathbb{P}$-conditional characteristic function of returns. Therefore, since the local equivalence of $\mathbb{P}$ and $\mathbb{Q}$ implies that the diffusion coefficient σ_t is the same under the two probability measures, the estimators of Jacod and Todorov (2014) and Todorov (2019) have the same estimand but use different sources of information.

Using the asymptotic expansion of $\mathcal{L}_{t,T}(u)$ in (3), we have

$$V_{t,T}(u) = \sigma_t^2 + O_p(T^{1-r/2}), \text{ as } T \downarrow 0. \tag{6}$$

The expansion in (3) and the corresponding result in (6) are based on "freezing" the spot characteristics of x (mainly, α_t, σ_t and $\nu_t(x)$) at their values at the beginning of the time interval and further ignoring the contribution of the jumps in x to the value of $\mathcal{L}_{t,T}(u)$. The latter is possible because of the asymptotically negligible role played by the jumps in the characteristic function of the increment of x for asymptotically increasing values of the characteristic exponent. In the next two sections, we will explicitly account for the effect of the time variation in the characteristics of x and the jumps in x on $\mathcal{L}_{t,T}(u)$. This, in turn, will allow us to improve on $V_{t,T}(u)$ as an estimator of σ_t^2.

3 Bias-Reduction in Volatility Estimation

To achieve improvements in the estimation of σ_t from the conditional characteristic function $\mathcal{L}_{t,T}(u)$, we will make use of a higher-order asymptotic expansion of $\mathcal{L}_{t,T}(u)$ stated in the next theorem.[6] In the statement of the theorem, for a random function $Z(u)$ and a positive-valued function f, $Z(u) = O_p^{lu}(f(T))$ means $\sup_{u\in\mathcal{U}} |Z(u)| = O_p(f(T))$ for any compact set $\mathcal{U} \in \mathbb{R}$.

Theorem 1. *If assumptions B1-r, B2-r and B3-r hold, then we have*

$$-\frac{2}{u^2}\log|\mathcal{L}_{t,T}(u)| = \sigma_t^2 + B_{t,T}(u) + \psi_t(u)T + O_p^{lu}(T^{3/2-r/2} \vee T^{3/2}|\log(T)|), \text{ for } T \downarrow 0, \tag{7}$$

[6] Bandi et al. (2021) also develop higher-order expansion for characteristic functions of returns over short-time intervals in a semiparametric setting, and use this expansion to estimate spot volatility from options. Their expansion result is derived under the assumption of diffusive volatility dynamics and constant jump intensity, and these restrictions on the return dynamics are utilized by Bandi et al. (2021) when constructing a spot volatility estimator (Bandi et al. (2021) do not provide the asymptotic order of the error term in the asymptotic expansion of the characteristic function and they do not show consistency and rate of convergence of their volatility estimator). In addition, the estimator of Bandi et al. (2021) relies on a parametric model for the jump component of the price. The results presented here, by contrast, work under general conditions for the underlying asset dynamics, allowing for volatility jumps, time-varying jump intensity and price-volatility cojumps, and they do not need one to model the jump distribution.

where $\psi_t(u)$ is some $\mathcal{F}_t$-adapted function of u, and

$$B_{t,T}(u) = \frac{2T}{u^2}\int_{\mathbb{R}}\left(1-\cos\left(ux/\sqrt{T}\right)\right)\nu_t(x)dx. \tag{8}$$

The term $\psi_t(u)$ in equation (7) is due to the time-variation in the drift coefficient, the stochastic volatility as well as some of its components over the interval $[t, t+T]$. On the other hand, $B_{t,T}(u)$ is the leading component in $\log|\mathcal{L}_{t,T}(u)|$ that is due to the jumps. It equals the real part of the characteristic exponent of the increment of the jumps in x when the intensity/jump compensator of the latter has been "frozen" at its value at the beginning of the interval.

From the higher-order asymptotic expansion result in (7), it is clear that

$$V_{t,T}(u) = \sigma_t^2 + B_{t,T}(u) + \psi_t(u)T + O_p(T^{3/2-r/2} \vee T^{3/2}|\log(T)|). \tag{9}$$

The bias term $\psi_t(u)T$ that is due to the time-variation in the spot characteristics of x is linear in T. This suggests an easy way to eliminate it from $V_{t,T}(u)$ by using characteristic functions over two different short time intervals $[t+T_1]$ and $[t+T_2]$, with

$$\tau = T_2/T_1 > 1. \tag{10}$$

More specifically, we introduce the following estimator

$$V_{t,T_1,T_2}(u) = \frac{1}{T_2-T_1}\left(T_2V_{t,T_1}(u) - T_1V_{t,T_2}(u)\right). \tag{11}$$

The construction of $V_{t,T_1,T_2}(u)$ is reminiscent of the classical Jacknife bias correction in statistics, see e.g., Efron and Tibshirani (1994). Using the asymptotic expansion result in (9), we have

$$V_{t,T_1,T_2}(u) = \sigma_t^2 + B_{t,T_1,T_2}(u) + O_p(T_1^{3/2-r/2} \vee T_1^{3/2}|\log(T_1)|), \text{ as } T_1 \downarrow 0, \tag{12}$$

where we use the shorthand notation

$$B_{t,T_1,T_2}(u) = \frac{T_2B_{t,T_1}(u) - T_1B_{t,T_2}(u)}{T_2-T_1}. \tag{13}$$

In the most general case, $B_{t,T_1,T_2}(u)$ is of the same order of magnitude as $B_{t,T}(u)$. However, under mild regularity type condition for the jump compensator ν_t, we can show that $B_{t,T_1,T_2}(u)$ is asymptotically smaller in magnitude than $B_{t,T}(u)$. This is done in the following theorem.

Theorem 2. *(a) Suppose that $0 < \int_{\mathbb{R}}\nu_t(x)dx < \infty$. Then, for $\tau > 1$ and $u \neq 0$, we have*

$$B_{t,T_1,T_2}(u)/B_{t,T_1}(u) \xrightarrow{\mathbb{P}} 0, \quad \text{as } T_1 \downarrow 0 \text{ and } T_2 = \tau T_1. \tag{14}$$

If in addition

$$\int_{\mathbb{R}} |\nu_t'(x)|dx < \infty, \tag{15}$$

where ν_t' denotes the first derivative of ν_t, then we have

$$B_{t,T_1,T_2}(u) = O_p\left(T_1^{3/2}\right) \text{ and } B_{t,T_1}(u) = O_p\left(T_1\right), \quad \text{as } T_1 \downarrow 0 \text{ and } T_2 = \tau T_1. \tag{16}$$

(b) Suppose

$$\nu_t(x) = \frac{c_t^- 1(x<0) + c_t^+ 1(x>0)}{|x|^{1+\beta_t}} + \overline{\nu}_t(x), \quad c_t^{\pm} \geq 0, \quad c_t^- + c_t^+ > 0, \quad \beta_t \in (0,2), \tag{17}$$

and

$$\int_{\mathbb{R}} |x|^{\widetilde{\beta}_t} |\widetilde{\nu}_t(x)|dx < \infty, \quad \text{for some } 0 \leq \widetilde{\beta}_t < \beta_t. \tag{18}$$

Then, for $\tau > 1$ and $u \neq 0$, we have

$$B_{t,T_1,T_2}(u)/B_{t,T_1}(u) \xrightarrow{\mathbb{P}} \frac{\tau - \tau^{1-\beta_t/2}}{\tau - 1} \in (0,1), \quad \text{as } T_1 \downarrow 0 \text{ and } T_2 = \tau T_1. \tag{19}$$

Part (a) of the above theorem shows that, in the case of finite activity jumps, $B_{t,T_1,T_2}(u)$ is of strictly smaller asymptotic order of magnitude than $B_{t,T_1}(u)$. That is, we have reduction in the asymptotic order of the bias due to the jumps in $V_{t,T_1,T_2}(u)$ relative to that in $V_{t,T}(u)$. The extent of the bias-reduction depends on the smoothness of ν_t. For example, under the additional smoothness assumption for $\nu_t(x)$ in (15), satisfied in most parametric specifications considered in prior work, we have

$$V_{t,T_1,T_2}(u) = \sigma_t^2 + O_p(T_1^{3/2}|\log(T_1)|) \text{ and } V_{t,T_1}(u) = \sigma_t^2 + O_p(T_1). \tag{20}$$

That is, the bias in $V_{t,T_1,T_2}(u)$ due to the jumps becomes of smaller asymptotic order while that in $V_{t,T_1}(u)$ remains unchanged.

Part (b) of Theorem 2 considers the case of infinite activity jumps under the assumption that their jump compensator is "locally stable", i.e., that the jump intensity around zero behaves like that of a time-changed stable process, see e.g., Barndorff-Nielsen and Shiryaev (2010), but is otherwise unrestricted. Most parametric applications that consider jumps of infinite activity satisfy such a "local stable" assumption, e.g., the CGMY process of Carr et al. (2002) and time-changes of it considered in Carr et al. (2003). In the case of "locally stable" jumps, $B_{t,T_1,T_2}(u)$ and $B_{t,T_1}(u)$ are of the same asymptotic order of magnitude. Nevertheless, in this case, as in case (a) of the theorem, we also have that asymptotically $B_{t,T_1,T_2}(u)$ is strictly smaller in magnitude than $B_{t,T_1}(u)$. The bias reduction becomes bigger for smaller values of β_t.

In Figure 1, we illustrate the reduction in bias due to the jumps in volatility estimation for the parametric models used in our Monte Carlo experiment in Section 6. The jump specification J1

corresponds to finite activity jumps satisfying the smoothness condition in (15). Consequently, the bias $B_{t,T_1,T_2}(u)$ is much smaller than $B_{t,T_1}(u)$ and the ratio of the two converges to zero for large values of u. As a result, the bias-reduction in volatility estimation offered by $V_{t,T_1,T_2}(u)$ is large. Next, specification J2 is again that of finite activity jumps but now the jump intensity is less smooth and does not satisfy the condition in (15). The bias-reduction for this case, therefore, is smaller than for case J1 and $B_{t,T_1,T_2}(u)/B_{t,T_1}(u)$ converges to zero as $u \to \infty$ at a slower rate. Specification J3 corresponds to infinite activity jumps which are not locally stable. Although Theorem 2 does not cover this specification (this specification will be covered by our analysis in the next section however), one can show in this case that $B_{t,T_1,T_2}(u)/B_{t,T_1}(u)$ should converge to zero but at a slower rate than a power of T_1. The evidence from Figure 1 is consistent with that. Finally, specification J4 corresponds to part (b) of Theorem 2 with $\beta_t = 0.5$. Here, we see reduction in the bias in the volatility estimation, but as implied by Theorem 2, $B_{t,T_1,T_2}(u)/B_{t,T_1}(u)$ converges to a constant (that is less than one in absolute value) for large values of u, i.e., there is a limit in the bias reduction offered by $V_{t,T_1,T_2}(u)$ in this case for the effect of price jumps on the volatility estimation.

4 Further Bias-Reduction in Volatility Estimation

Without additional structure on the jumps, it seems impossible to reduce further the bias in the spot volatility estimation. However, an assumption for local power behavior of the jump compensator around the origin can allow us to expand the bias term $B_{t,T_1,T_2}(u)$ and find an estimator for it. More specifically, we will consider the following additional assumption for the jumps:

A. *Suppose that the density of the jump compensator satisfies*

$$\begin{aligned}\nu_t(x) = c_t^+ \frac{1}{|x|^{\beta+1}} 1_{\{x>0\}} + \overline{c}_t^+ \frac{1}{|x|^{\beta}} 1_{\{x>0,\, \beta>0\}} \\ + c_t^- \frac{1}{|x|^{\beta+1}} 1_{\{x<0\}} + \overline{c}_t^- \frac{1}{|x|^{\beta}} 1_{\{x<0,\, \beta>0\}} + \overline{\nu}_t(x),\end{aligned} \tag{21}$$

for some $\beta \in (-1, 1)$, $c_t^{\pm} \geq 0$ *with* $c_t^- + c_t^+ > 0$, $\overline{c}_t^{\pm} \in \mathbb{R}$, *and where* $\lim_{x \to \pm\infty} \overline{\nu}_t(x) = 0$ *and the following holds*

$$\int_{\mathbb{R}} (|\overline{\nu}_t(x)| + |\overline{\nu}'_t(x)|) dx < \infty, \tag{22}$$

for $\overline{\nu}'_t(x)$ *denoting the derivative of* $\overline{\nu}_t(x)$ *with respect to* x.

Assumption A can be viewed as a semiparametric model for the jump compensator because it imposes parametric structure for the leading term of $\nu_t(x)$ near the origin. The parameter β in (21) determines the value of the so-called (spot) Blumenthal-Getoor index of the jumps defined as

$$\inf\{p \geq 0 : \int_{\mathbb{R}} (|x|^p \wedge 1) \nu_t(x) dx < \infty\}, \tag{23}$$

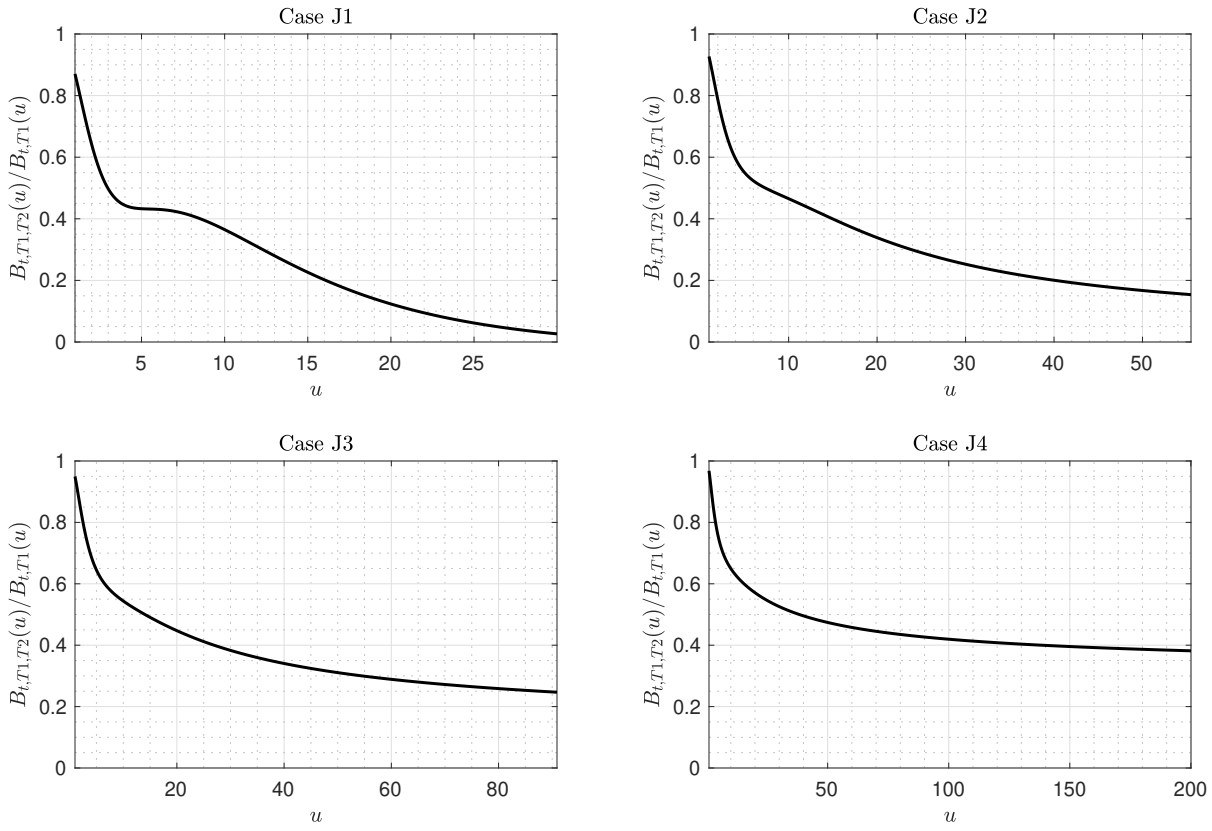

Figure 1: **The ratio $B_{t,T_1,T_2}(u)/B_{t,T_1}(u)$ in Parametric Models**. The model specifications J1-J4 are given in Section 6. The value of σ_t^2 is set to its unconditional mean of 0.02. The maximum value of u on each of the plots is the minimum value of u for which $B_{t,T_1}(u)$ falls below $0.01 \times \sigma_t^2$.

with the spot Blumenthal-Getoor index of the jumps in x being equal to $\beta \vee 0$ under assumption A. This assumption is quite general and many parametric models such as the time-changed CGMY processes of Carr et al. (2003), satisfy it. Essentially all past work on inference for the behavior of "small" jumps from high-frequency data assumes a structure for the jumps similar to that in assumption A, see e.g., Jacod and Protter (2012) and references therein.

In the case $\beta > 0$, assumption A implies "local stability" of the jumps that we considered in part (b) of Theorem 2. In this case, the "small" jumps of x behave like those of a time-changed stable process and assumption A further imposes restriction on the deviation of ν_t from that of a time-changed stable process. Assumption A will hold, for example, when $x^{\beta+1}\nu_t(x)$ is analytic in a neighborhood of zero. The case $\beta = 0$ corresponds to the case of infinite activity jumps with Blumenthal-Getoor index of zero and can be considered as the borderline case between jumps of finite and infinite activity. In this case the jump compensator around zero behaves like the

difference of two time-changed Gamma processes. Finally, the case $\beta < 0$ corresponds to jumps of finite activity but for which the jump compensator is still exploding near the origin. In this regard, we note that jumps satisfying assumption A for some $\beta \in (-1, 0)$ will not satisfy the smoothness condition in (15). The explosion of the jump compensator around zero will make $B_{t,T_1,T_2}(u)$ of bigger asymptotic order than $T_1^{3/2}$ in that case.

With the help of assumption A, we can make an asymptotic expansion of $B_{t,T_1,T_2}(u)$ for $T_1 \downarrow 0$. The formal result is given in the following theorem.

Theorem 3. *If assumption A holds, then we have*

$$B_{t,T_1,T_2}(u) = \overline{B}_{t,T_1,T_2}(u) + O_p^{lu}\left(T_1^{1+\left(\frac{1}{2}-\frac{\beta}{2}\right)\wedge\frac{1}{2}}\right), \tag{24}$$

$$\overline{B}_{t,T_1,T_2}(u) = \frac{2\tau^{1-\beta/2}}{\tau-1}(c_t^+ + c_t^-)T_1^{1-\beta/2}u^{\beta-2}\Gamma(1-\beta)\cos\left(\frac{\pi\beta}{2}\right)g(\sqrt{\tau},\beta), \tag{25}$$

as $T_1 \downarrow 0$ and $T_2 = \tau T_1$, for some $\tau > 1$, and where for $\beta \in (-1, 1)$ and $\zeta > 1$, we denote

$$g(\zeta,\beta) = \begin{cases} \frac{(\zeta)^\beta - 1}{\beta}, & \text{if } \beta \in (-1,1)\setminus\{0\}, \\ \log(\zeta), & \text{if } \beta = 0. \end{cases} \tag{26}$$

In Figure 2, we compare the bias $B_{t,T_1,T_2}(u)$ with its leading term $\overline{B}_{t,T_1,T_2}(u)$ for the parametric models used in our Monte Carlo experiment in Section 6. We omit case J1 as the jumps in that specification do not satisfy assumption A. In all remaining three cases, $B_{t,T_1,T_2}(u)$ and $\overline{B}_{t,T_1,T_2}(u)$ differ significantly for small values of u. This is not surprising because for these values of u, the "big" jumps play a role and the expansion in Theorem 3 is driven by small jumps. For large values of u, $B_{t,T_1,T_2}(u)$ and $\overline{B}_{t,T_1,T_2}(u)$ become much closer, consistent with the asymptotic expansion result of Theorem 3. We can see differences in the quality of the approximation across the different specifications: it is worst for case J2 and it is best for case J4. Case J2 corresponds to $\beta = -0.5$ and case J4 to $\beta = 0.5$, for β being the parameter in assumption A. The reason for this difference is that the size of $\overline{B}_{t,T_1,T_2}(u)$ relative to the residual term in (24) is smallest (asymptotically) for smaller values of β. As we will see later, however, the worse performance of the approximation for case J2 will not have an adverse effect on our ability to purge the effect of jumps on the volatility estimation because for lower values of β, $B_{t,T_1,T_2}(u)$ is also smaller.

We can use the result of Theorem 3 to estimate the leading term of $B_{t,T_1,T_2}(u)$ and remove it from the variance estimator $V_{t,T_1,T_2}(u)$. This will lead to reduction in the asymptotic order of the bias in the latter even in the case when jumps in x are of infinite activity. More specifically,

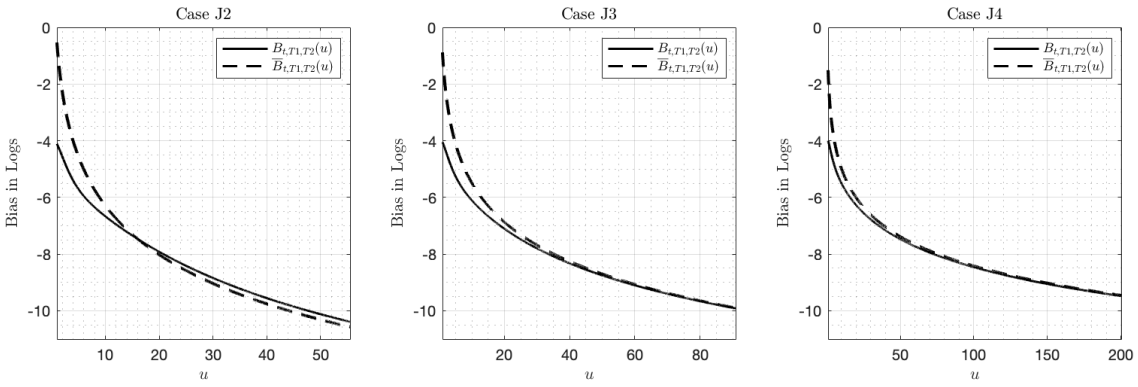

Figure 2: $B_{t,T_1,T_2}(u)$ **and** $\overline{B}_{t,T_1,T_2}(u)$ **in Parametric Models**. The model specifications J2-J4 are given in Section 6. The value of σ_t^2 is set to its unconditional mean of 0.02. The maximum value of u on each of the plots is the minimum value of u for which $B_{t,T_1}(u)$ falls below $0.01 \times \sigma_t^2$.

from the higher-order asymptotic expansion of the conditional characteristic function in (7) and Theorem 3, we have for arbitrary small $\iota > 0$:

$$V_{t,T_1,T_2}(u) = \sigma_t^2 + \psi_{t,T_1} u^{\beta-2} + O_p\left(T_1^{\left(\frac{3}{2}-\frac{\beta}{2}-\iota\right)\wedge\frac{3}{2}} |\log(T_1)|\right), \quad \text{as } T_1 \downarrow 0 \text{ and } T_2 = \tau T_1, \tag{27}$$

with ψ_{t,T_1} being an $\mathcal{F}_t$-adapted random variable which does not depend u but depends on T_1, β and τ. We can use this expansion to estimate jointly σ_t^2 and the "nuisance parameters" β and ψ_{t,T_1} by nonlinear least squares estimation which we now describe.

Since the estimation of the activity parameter β can be rather difficult and because it is typically assumed constant over time, we will consider estimation of this parameter from data on multiple days. Towards this end, we define with $\mathbb{T}$ a finite set of numbers taking values in $\mathbb{R}_{>0}$ and we use the shorthand notation $V_{\mathbb{T}}(u) = \sum_{t\in\mathbb{T}} V_{t,T_1,T_2}(u)$. We next introduce the following vector

$$\underline{u} = (u_1, ..., u_K) \in \mathbb{R}^K_{>0}, \ K \geq 3 \text{ and } u_i = \kappa u_{i-1}, \text{ for } i = 2, ..., K \text{ and some } \kappa > 1. \tag{28}$$

Our estimator of β is given by

$$\beta_{\mathbb{T}}(\underline{u}) = \underset{x\in[-1,1]}{\operatorname{argmin}} \sum_{i=1}^{K} \left(V_{\mathbb{T}}(u_i) - V_{\mathbb{T}}(\underline{u}; x) - \psi_{\mathbb{T}}(\underline{u}; x) u_i^{x-2}\right)^2, \tag{29}$$

where

$$V_{\mathbb{T}}(\underline{u}; x) = \frac{\sum_{i=1}^K u_i^{2x-4} \sum_{i=1}^K V_{\mathbb{T}}(u_i) - \sum_{i=1}^K u_i^{x-2} \sum_{i=1}^K V_{\mathbb{T}}(u_i) u_i^{x-2}}{K \sum_{i=1}^K u_i^{2x-4} - \left(\sum_{i=1}^K u_i^{x-2}\right)^2}, \tag{30}$$

$$\psi_{\mathbb{T}}(\underline{u}; x) = \frac{K \sum_{i=1}^K V_{\mathbb{T}}(u_i) u_i^{x-2} - \sum_{i=1}^K u_i^{x-2} \sum_{i=1}^K V_{\mathbb{T}}(u_i)}{K \sum_{i=1}^K u_i^{2x-4} - \left(\sum_{i=1}^K u_i^{x-2}\right)^2}. \tag{31}$$

Since the function that is minimized is continuously differentiable in x for $x \in [-1,1]$, the minimum exists. If this minimum is achieved by several values of x, then $\beta_{\mathbb{T}}(\underline{u})$ denotes the smallest of them.

The bias-corrected estimate of σ_t^2 is given by $V_t(\underline{u}, \beta_{\mathbb{T}}(\underline{u}))$ while $\beta_{\mathbb{T}}(\underline{u})$ is an estimate of β. We note that $\psi_{\mathbb{T}}(\underline{u}; \beta_{\mathbb{T}}(\underline{u}))$ is a drifting parameter estimate of ψ_{t,T_1} in (27), which recall is asymptotically shrinking. This complicates the analysis of $V_t(\underline{u}, \beta_{\mathbb{T}}(\underline{u}))$ and makes it quite different from standard nonlinear least squares estimators. We start with establishing the properties of $\beta_{\mathbb{T}}(\underline{u})$.

Theorem 4. *If assumptions A, B1-r, B2-r and B3-r hold, for some $\tau > 1$ and $\underline{u}$ given in (28), we have*

$$\beta_{\mathbb{T}}(\underline{u}) - \beta = O_p\left(T_1^{\frac{1}{2} \wedge \left(\frac{\beta+1}{2}\right)}\right), \text{ as } T_1 \downarrow 0 \text{ and } T_2 = \tau T_1. \tag{32}$$

We note that the rate of convergence of the estimator $\beta_{\mathbb{T}}(\underline{u})$ is slow, particularly for lower values of β. This is natural as separation of diffusion from small jumps, which determine the jump activity index, is particularly difficult for lower values of the latter. However, we are not interested in $\beta_{\mathbb{T}}(\underline{u})$ per se but rather in the estimator of the spot diffusive volatility. In the case when the recovery of $\beta_{\mathbb{T}}(\underline{u})$ is poor in relative terms, i.e., for lower values of β, the bias term in $V_{t,T_1,T_2}(u)$ due to the jumps is also asymptotically small. As a result, $V_t(\underline{u}, \beta_{\mathbb{T}}(\underline{u}))$ improves on $V_{t,T_1,T_2}(u)$, regardless of the value of β, as we show in the following theorem.

Theorem 5. *If assumptions A, B1-r, B2-r and B3-r hold, then for some arbitrary small $\iota > 0$, $\tau > 1$ and $\underline{u}$ given in (28), we have*

$$V_t(\underline{u}, \beta_{\mathbb{T}}(\underline{u})) = \sigma_t^2 + O_p\left(T_1^{\left(\frac{3}{2} - \frac{\beta}{2} - \iota\right) \wedge \frac{3}{2}} |\log(T_1)|\right), \text{ as } T_1 \downarrow 0 \text{ and } T_2 = \tau T_1. \tag{33}$$

In addition, if $\int_{\mathbb{R}} \nu_t(x)dx < \infty$ and $\int_{\mathbb{R}} |\nu_t'(x)|dx < \infty$, we have

$$V_t(\underline{u}, \beta_{\mathbb{T}}(\underline{u})) = \sigma_t^2 + O_p\left(T_1^{\frac{3}{2}} |\log(T_1)|\right), \text{ as } T_1 \downarrow 0 \text{ and } T_2 = \tau T_1. \tag{34}$$

The first result of the theorem, shows that when assumption A holds, then $V_t(\underline{u}, \beta_{\mathbb{T}}(\underline{u}))$ improves on $V_{t,T_1,T_2}(u)$ by removing the leading term in it that is due to the jumps in x. This leads to a much smaller approximation error of $V_t(\underline{u}, \beta_{\mathbb{T}}(\underline{u}))$ as an estimator of σ_t^2. Importantly, however, this improvement does not come at the cost of bad behavior (at least asymptotically) of $V_t(\underline{u}, \beta_{\mathbb{T}}(\underline{u}))$ when assumption A does not hold. More specifically, the result in (34) shows that if jumps are of finite activity and assumption A does not hold, the de-biasing of the jumps in $V_t(\underline{u}, \beta_{\mathbb{T}}(\underline{u}))$ has no asymptoticaly detrimental effect.

5 Feasible Volatility Estimation from Short-Dated Options

In the previous two sections, we considered estimation of σ_t^2 assuming $\mathcal{L}_{t,T}(u)$ is known. In practice, this is not the case, of course, and hence the above variance estimators are not feasible. In this section, we will make the inference procedures feasible by replacing $\mathcal{L}_{t,T}(u)$ with an estimate for it from options written on the asset. More specifically, following Carr and Madan (2001), we have

$$\mathcal{L}_{t,T}(u) = 1 - \left(\frac{u^2}{T} + i\frac{u}{\sqrt{T}}\right) e^{-x_t} \int_{\mathbb{R}} e^{(iu/\sqrt{T}-1)(k-x_t)} O_{t,T}(k)dk, \tag{35}$$

where $O_{t,T}(k)$ denotes the price at time t of a European-style out-of-the-money option price, expiring at time $t+T$ with strike e^k, and whose underlying asset price at time t is x_t. We recall that $O_{t,T}(k)$ is the minimum of the put and call option prices with strike e^k. In practice, we do not observe options on a continuum of (log)-strikes but rather on a discrete grid, which we denote with

$$k_{t,T}(1) < ... < k_{t,T}(N_{t,T}), \tag{36}$$

and the gap between log-strikes is $\Delta_{t,T}(j) = k_{t,T}(j) - k_{t,T}(j-1)$. Option prices are observed with error, i.e., we observe

$$\widehat{O}_{t,T}(k_{t,T}(j)) = O_{t,T}(k_{t,T}(j)) + \epsilon_{t,T}(k_{t,T}(j)), \tag{37}$$

where the errors $\epsilon_{t,T}(k_{t,T}(j))$ are defined on a space $\Omega^{(1)} = \mathbb{R}^{\mathbb{R}} \times \mathbb{R}^{\mathbb{R}}$ which is equipped with the product Borel σ-field $\mathcal{F}^{(1)}$, and transition probability $\mathbb{P}^{(1)}(\omega^{(0)}, d\omega^{(1)})$ from the probability space $\Omega^{(0)}$, on which X is defined, to $\Omega^{(1)}$. We further define,

$$\Omega \,=\, \Omega^{(0)} \times \Omega^{(1)}, \quad \mathcal{F} \,=\, \mathcal{F}^{(0)} \times \mathcal{F}^{(1)},$$

and

$$\mathbb{P}(d\omega^{(0)}, d\omega^{(1)}) \,=\, \mathbb{P}^{(0)}(d\omega^{(0)})\, \mathbb{P}^{(1)}(\omega^{(0)}, d\omega^{(1)})\,.$$

Using the available options, a Riemann sum approximation of the integral in (35) leads to the following feasible counterpart of $\mathcal{L}_{t,T}(u)$:

$$\widehat{\mathcal{L}}_{t,T}(u) = 1 - \left(\frac{u^2}{T} + i\frac{u}{\sqrt{T}}\right) e^{-x_t} \sum_{j=2}^{N_{t,T}} e^{(iu/\sqrt{T}-1)(k_{t,T}(j-1)-x_t)} \widehat{O}_{t,T}(k_{t,T}(j-1))\Delta_{t,T}(j),\ u \in \mathbb{R}. \tag{38}$$

We denote the real and imaginary parts of $\mathcal{L}_{t,T}(u)$ with $\mathcal{R}_{t,T}(u)$ and $\mathcal{I}_{t,T}(u)$, and those of $\widehat{\mathcal{L}}_{t,T}(u)$ with $\widehat{\mathcal{R}}_{t,T}(u)$ and $\widehat{\mathcal{I}}_{t,T}(u)$.

The convergence in distribution of the centered $\widehat{\mathcal{L}}_{t,T}(u)$ that we present next holds $\mathcal{F}^{(0)}$-conditionally. This is denoted by $\xrightarrow{\mathcal{L}|\mathcal{F}^{(0)}}$ and formally means convergence in probability of the

conditional probability laws when the latter are considered as random variables taking values in the space of probability measures equipped with the weak topology, see e.g., VIII.5.26 of Jacod and Shiryaev (2003).

For stating the next theorem, we introduce the following additional notation

$$\widetilde{\Phi}(k) = f(k) + |k|\Phi(-|k|), \quad k \in \mathbb{R}, \tag{39}$$

where f and Φ are the pdf and cdf, respectively, of a standard normal random variable. We also use the shorthand notation $\underline{K} = \max_{t\in\mathbb{T},\, l=1,2} e^{k_{t,T_l}(1)}$ and $\overline{K} = \min_{t\in\mathbb{T},\, l=1,2} e^{k_{t,T_l}(N_{t,T_l})}$. Finally, Δ is a reference "average" log-strike gap, formally defined in assumption C2.

Theorem 6. *Suppose assumptions C1-C3 hold for $t \in \mathbb{T}$, with $\mathbb{T}$ being a finite set whose elements are in $\mathbb{R}_{>0}$. Let $T_1 \downarrow 0$, $T_2 = \tau T_1$ for some $\tau > 1$, $\Delta \asymp T_1^{\alpha}$, $\underline{K} \asymp T_1^{\beta}$, $\overline{K} \asymp T_1^{\gamma}$, for $\beta, \gamma > 0$ and $\alpha > \frac{1}{2}$. Then, for any compact $\mathcal{U} \in \mathbb{R}$, we have*

$$\frac{T_l^{1/4}}{\sqrt{\Delta}} \begin{pmatrix} \widehat{\mathcal{I}}_{t,T_l}(u) - \mathcal{I}_{t,T_l}(u) \\ \widehat{\mathcal{R}}_{t,T_l}(u) - \mathcal{R}_{t,T_l}(u) \end{pmatrix} \xrightarrow{\mathcal{L}|\mathcal{F}^{(0)}} \begin{pmatrix} Z_{t,l,I}(u) \\ Z_{t,l,R}(u) \end{pmatrix}, \; l = 1, 2, \tag{40}$$

uniformly in $u \in \mathcal{U}$, and where $\mathcal{F}^{(0)}$-conditionally $(Z_{t,1,I}(u)\; Z_{t,1,R}(u))$ and $(Z_{t,2,l}(u)\; Z_{t,2,R}(u))$ are independent of each other, and across $t \in \mathbb{T}$, and are centered Gaussian processes with

$$\begin{aligned} &\begin{pmatrix} \mathbb{E}\left(Z_{t,l,I}(u)Z_{t,l,I}(v)\middle|\mathcal{F}^{(0)}\right) & \mathbb{E}\left(Z_{t,l,I}(u)Z_{t,l,R}(v)\middle|\mathcal{F}^{(0)}\right) \\ \mathbb{E}\left(Z_{t,l,R}(u)Z_{t,l,I}(v)\middle|\mathcal{F}^{(0)}\right) & \mathbb{E}\left(Z_{t,l,R}(u)Z_{t,l,R}(v)\middle|\mathcal{F}^{(0)}\right) \end{pmatrix} \\ &= \Sigma_{t,l}(u,v) := \begin{pmatrix} \Sigma_{t,l,I}(u,v) & 0 \\ 0 & \Sigma_{t,l,R}(u,v) \end{pmatrix}, \end{aligned} \tag{41}$$

and

$$\Sigma_{t,l,I}(u,v) = \sigma_t^3 \psi_{t,l}(0)\zeta_{t,l}^2(0) \int_{\mathbb{R}} \sin(\sigma_t^2 uk)\sin(\sigma_t^2 vk)\widetilde{\Phi}^2(k)dk, \tag{42}$$

$$\Sigma_{t,l,R}(u,v) = \sigma_t^3 \psi_{t,l}(0)\zeta_{t,l}^2(0) \int_{\mathbb{R}} \cos(\sigma_t^2 uk)\cos(\sigma_t^2 vk)\widetilde{\Phi}^2(k)dk, \tag{43}$$

where the functions $\psi_{t,l}$ and $\zeta_{t,l}$ appear in assumption C3.

We define next the variance estimators formed on the basis of $\widehat{\mathcal{L}}_{t,T}(u)$. The feasible counterparts of $V_{t,T}(u)$ and $V_{t,T_1,T_2}(u)$ are given by

$$\widehat{V}_{t,T}(u) = -\frac{2}{u^2}\log|\widehat{\mathcal{L}}_{t,T}(u)| \text{ and } \widehat{V}_{t,T_1,T_2}(u) = \frac{1}{T_2 - T_1}\left(T_2\widehat{V}_{t,T_1}(u) - T_1\widehat{V}_{t,T_2}(u)\right), \; u \in \mathbb{R}. \tag{44}$$

We further define $\widehat{V}_{\mathbb{T}}(u) = \sum_{t\in\mathbb{T}} \widehat{V}_{t,T_1,T_2}(u)$, and from here $\widehat{\beta}_{\mathbb{T}}(\underline{u})$ and $\widehat{V}_{\mathbb{T}}(\underline{u}, x)$ are the counterparts of $\beta_{\mathbb{T}}(\underline{u})$ and $V_{\mathbb{T}}(\underline{u}, x)$ in which $V_{\mathbb{T}}(u)$ is replaced by $\widehat{V}_{\mathbb{T}}(u)$. In the next corollary we derive the rate of convergence of the feasible variance estimators.

Corollary 1. *Suppose assumptions B1-r, B2-r, B3-r and C1-C3 hold for $t \in \mathbb{T}$, with $\mathbb{T}$ being a finite set whose elements are in $\mathbb{R}_{>0}$. Let $T_1 \downarrow 0$, $T_2 = \tau T_1$ for some $\tau > 1$, $\Delta \asymp T_1^{\alpha}$, $\underline{K} \asymp T_1^{\beta}$, $\overline{K} \asymp T_1^{\gamma}$, for $\beta, \gamma > 0$ and $\alpha > \frac{1}{2}$. Fix $u \neq 0$ and define $\underline{u}$ as in (28).*

(a) If $0 < \int_{\mathbb{R}} \nu_t(x)dx < \infty$ and $\int_{\mathbb{R}} |\nu_t'(x)|dx < \infty$, we have

$$\widehat{V}_{t,T_1,T_2}(u) = \sigma_t^2 + O_p\left(T_1^{\frac{3}{2}} |\log(T_1)| \bigvee \frac{\sqrt{\Delta}}{T_1^{1/4}}\right), \tag{45}$$

$$\widehat{V}_t(\underline{u}, \widehat{\beta}_{\mathbb{T}}(\underline{u})) = \sigma_t^2 + O_p\left(T_1^{\frac{3}{2}} |\log(T_1)| \bigvee \frac{\sqrt{\Delta}}{T_1^{1/4}}\right). \tag{46}$$

(b) If assumption A holds, we have

$$\widehat{V}_{t,T_1,T_2}(u) = \sigma_t^2 + O_p\left(T_1^{1-\frac{\beta}{2}} \bigvee \frac{\sqrt{\Delta}}{T_1^{1/4}}\right), \tag{47}$$

$$\widehat{V}_t(\underline{u}, \widehat{\beta}_{\mathbb{T}}(\underline{u})) = \sigma_t^2 + O_p\left(T_1^{\left(\frac{3}{2}-\frac{\beta}{2}-\iota\right) \wedge \frac{3}{2}} \bigvee \frac{\sqrt{\Delta}}{T_1^{1/4}}\right), \tag{48}$$

for some arbitrary small $\iota > 0$.

We note that from Theorem 6, we can also derive a CLT for the feasible variance estimators but for brevity we do not state these results.

Remark 1. *The CLT of Theorem 6 is derived under the assumption of $\mathcal{F}^{(0)}$-conditional independence of the option observation errors, see assumption C3. This assumption can be weakened to requiring only weak $\mathcal{F}^{(0)}$-conditional spatial dependence in the observation error. In this case, the variance of the limiting distribution in (40) will contain a long-run variance component incorporating the effect from the spatial dependence in the observation error. We do not consider such an extension here as the empirical evidence in Andersen et al. (2021) indicates that the spatial dependence in the option observation error of S&P 500 index short-dated options, used in our empirical analysis, is very small.*

6 Monte Carlo Study

In this section, we evaluate the performance of the various feasible variance estimators introduced above on simulated data.

6.1 Setup

We use the following model for the underlying asset dynamics, under the risk-neutral probability, to generate the true option prices:

$$\frac{dX_t}{X_{t-}} = \sqrt{V_t} dW_t + \int_{\mathbb{R}} (e^x - 1)\, \mu(dt, dx), \tag{49}$$

$$dV_t = \kappa_v(\theta_v - V_t)dt + \sigma_v \sqrt{V_t} dB_t, \tag{50}$$

where W_t and B_t are $\mathbb{Q}$ Brownian motions with $\text{corr}(dW_t, dB_t) = \rho dt$, and μ is an integer-valued random measure with $\mathbb{Q}$ compensator $dt \otimes \nu_t(dx)$, for ν_t given by

$$\nu_t(dx) = V_t \times \nu^{ts}(x)dx,\ \nu^{ts}(x) = c_- \frac{e^{-\lambda_- |x|}}{|x|^{1+\mathfrak{b}}} 1_{\{x<0\}} + c_+ \frac{e^{-\lambda_+ |x|}}{|x|^{1+\mathfrak{b}}} 1_{\{x>0\}}. \tag{51}$$

In the above specification for X, the stochastic variance is modeled as a square-root diffusion process like in the popular Heston model (Heston (1993)). The price jumps have intensity that is affine in the level of diffusive variance like in Bates (2000) and Pan (2002), and subsequent empirical option pricing work. Our jump specification is a time-changed CGMY process, see Barndorff-Nielsen and Shiryaev (2010), with the time-change being the integrated diffusive variance. The CGMY process is a general Lévy jump process, with separate parameters controlling the tails (λ_- and λ_+) and the behavior of the small jumps ($\mathfrak{b}$).

We consider several parameter settings for the above model which can help us understand the role of various biases involved in the option-based volatility estimation as well as the effectiveness of the proposed bias-reduction techniques. The parameter values for all cases are given in Table 1. In the first five model specifications, we exclude the price jumps and vary the parameters capturing the volatility dynamics. In all considered cases, we set $\theta_v = 0.02$, which is close to the average value of market variance. In scenario D1, we set the mean reversion parameter κ_v so that the half-life of a shock to variance is one month. For this case, we set the volatility of volatility parameter σ_v to a relatively low value of 0.2 (recall the Feller condition $\sigma_v^2 < 2\kappa_v\theta_v$ that implies that zero is a reflecting barrier for the square-root diffusion process) and the leverage coefficient to $\rho = -0.5$. In scenario D2, we keep all parameters, except for ρ, the same as in scenario D1, and we decrease ρ to -0.9 (i.e., we consider stronger leverage effect). In Scenario D3, we increase the speed of mean reversion in variance, with κ_v now implying half-life of a variance shock of only one week and we adjust σ_v so that the coefficient of variation of V (which is given by $\sigma_v/\sqrt{2\kappa_v\theta_v}$) remains approximately the same as in case D1. Scenario D4 is the same as scenario D1 but with the volatility of volatility parameter increased to 0.5. Finally, in scenario D5, we consider the combined effect of increasing mean reversion in variance, higher leverage and higher volatility of volatility, relative to case D1. As

such, this case is the most challenging for the asymptotic expansion of the conditional characteristic function and the associated variance estimators.

Table 1: **Parameter Setting for the Monte Carlo**

Case	Variance Parameters				Jump Parameters				
	θ_v	κ_v	σ_v	ρ	$\mathfrak{b}$	λ_-	λ_+	c_-	c_+
D-1	0.02	8.3	0.2	-0.5	–	–	–	0	0
D-2	0.02	8.3	0.2	-0.9	–	–	–	0	0
D-3	0.02	34.9	0.4	-0.5	–	–	–	0	0
D-4	0.02	8.3	0.5	-0.5	–	–	–	0	0
D-5	0.02	34.9	1.0	-0.9	–	–	–	0	0
J-1	0.02	8.3	0.2	-0.5	-1.5	20	100	9.69×10^3	3.01×10^5
J-2	0.02	8.3	0.2	-0.5	-0.5	20	100	1.21×10^3	7.53×10^3
J-3	0.02	8.3	0.2	-0.5	0.0	20	100	3.6×10^2	10×10^2
J-4	0.02	8.3	0.2	-0.5	0.5	20	100	0.91×10^2	1.13×10^2

Next, in order to study the bias in the variance estimation due to the price jumps, we consider the model specifications J-1, J-2, J-3 and J-4. In all of them, we set the variance parameters to their values in case D-1. We consider four cases for the parameter $\mathfrak{b}$ that determines the jump activity: -1.5, -0.5, 0.0 and 0.5. The first two correspond to jumps of finite activity, the third to the variance gamma process of Madan and Seneta (1990), and the last one to a difference of two inverse Gaussian processes. For case J-1, we have $\int_{\mathbb{R}} |\nu_t'(x)|dx < \infty$. Therefore, in this case de-biasing for the jumps like in Section 4 is not needed but remember from Corollary 1 that such de-biasing procedure should do no harm in this case. We will assess this in the Monte Carlo. Cases J-3 and J-4 are ones with jumps of infinite activity (i.e., ones whose paths contain infinite number of small jumps over finite time intervals). In all considered cases, we set $\lambda_- = 20$ and $\lambda_+ = 100$. This choice implies tail decays of out-of-the-money puts and calls similar to those of observed options written on the S&P 500 index, see e.g., Andersen et al. (2015). Finally, for a given $\mathfrak{b}$, $\lambda_\pm$, we set $c_\pm$ according to

$$c_- = 0.9 \times \frac{\lambda_-^{2-\mathfrak{b}}}{\Gamma(2-\mathfrak{b})} \text{ and } c_+ = 0.1 \times \frac{\lambda_+^{2-\mathfrak{b}}}{\Gamma(2-\mathfrak{b})},$$

which implies that spot jump variation is equal to spot diffusive variance, and further that 90% of the jump variation is due to negative jumps. This separation of the risk-neutral variation into

diffusive and one due to positive and negative jumps is similar to that implied from parametric models fitted to observed S&P 500 index options, see e.g., Andersen et al. (2015).

We turn next to the option observation scheme. Observed options are given by

$$\widehat{O}_{t,T}(k_{t,T}(j)) = O_{t,T}(k_{t,T}(j))(1 + 0.03 \times z_{t,T}(j)),\ j = 1, ..., N_{t,T}, \tag{52}$$

where $\{z_{t,T}(j)\}_{j=1}^{N_{t,T}}$ are sequences of i.i.d. standard normal variables which are independent of each other. The size of the observation error is calibrated to match roughly bid-ask spreads of index option data. The initial level of the underlying stock price is drawn from a uniform distribution on the interval $[1997.5,\ 2002.5]$, and for each pair (t, T), the strike grid is equidistant with gaps between strikes of 5. The strike closest to the money is set to 2000 and the strikes below and above 2000 are extended in both directions by increments of 5 until the true out-of-the-money option price falls below 0.075. This specification of the strike grid mimics that of available of S&P 500 index options. The current value of variance is set to 10-th, 50-th or 90-th quantile of its marginal distribution, and we also experiment with more extreme starting values of V_t in the case of the jump-diffusion models. Finally, in all simulation scenarios, we consider three tenors: $T_1 = 3/252$, $T_2 = 5/252$ and $T_3 = 10/252$, which correspond to 3, 5 and 10 business days to expiration, respectively.

6.2 Bias-Reduction in Variance Estimation due to the Volatility Dynamics

We start with investigating the bias in option-based variance estimation due to the volatility dynamics in the underlying asset. For this, we will compare the performance of the original characteristic function based variance estimator $\widehat{V}_{t,T}(u)$ with the bias-corrected $\widehat{V}_{t,T_1,T_2}(u)$. The comparison will be done for the scenarios D1-D5 in which the underlying asset does not contain jumps and any bias in $V_{t,T}(u)$ is due to the volatility dynamics. For implementing the variance estimators, we need to set the value of the characteristic exponent. We do this in the following data-driven way:

$$\widehat{u}_{t,T} = \inf\{u \geq 0 : |\widehat{\mathcal{L}}_{t,T}(u)| \leq 0.3\} \wedge \text{argmin}_{u\in[0,\overline{u}_{t,T}]}|\widehat{\mathcal{L}}_{t,T}(u)|, \tag{53}$$

where $\overline{u}_{t,T} = \sqrt{-2\log(0.05)}/\widehat{\sigma}_{t,ATM}$ and $\widehat{\sigma}_{t,ATM}$ is the at-the-money Black-Scholes implied volatility at time t for the shortest available maturity on that day. The idea behind the above specification of $\widehat{u}_{t,T}$ is the following. We want to pick the characteristic exponent as high as possible so that the contribution of jumps in $\widehat{V}_{t,T}(u)$ and $\widehat{V}_{t,T_1,T_2}(u)$ is minimized. At the same, very high values of u will lead to noisy variance estimators, and in addition, higher order terms in the expansion in (7) (recall that the result in (7) is for bounded and fixed u) as well as higher order terms in the Riemann sum approximation of the integral in $\mathcal{L}_{t,T}(u)$ will start playing a role. For this reason, our data-driven $\widehat{u}_{t,T}$ is an estimate of the smallest positive u for which $\operatorname{plim}_{T\to 0}\widehat{\mathcal{L}}_{t,T}(u)$ reaches some

small value, which we take here to be 0.3. We note in this regard that $\text{argmin}_{u\in[0,\overline{u}_{t,T}]}|\widehat{\mathcal{L}}_{t,T}(u)|$ is above $\inf\{u \geq 0 : |\widehat{\mathcal{L}}_{t,T}(u)| \leq 0.3\}$, with probability approaching one, and is used only as a guard against finite-sample distortions.

With the above choice of the characteristic exponent, we compare the performance of the feasible estimators $\widehat{V}_{t,T_1}(\widehat{u}_{t,T_1})$ and $\widehat{V}_{t,T_1,T_2}(\widehat{u}_{t,T_1})$. The results from the Monte Carlo are reported in Table 2. We can draw several conclusions from the reported results. First, in the baseline scenario D1, the original estimator $\widehat{V}_{t,T_1}(\widehat{u}_{t,T_1})$ performs very well and so does the two-tenor estimator $\widehat{V}_{t,T_1,T_2}(\widehat{u}_{t,T_1})$. The biases in all of these estimators are quite small in relative terms. The performance of the estimators in scenario D2 looks very similar to that in scenario D1. That is, increasing the leverage effect close to its maximum possible value seems not to affect adversely the estimators. On the other hand, the results for scenario D3, in which the mean-reversion parameter κ_v is increased, indicate that this feature of the model does affect the variance recovery. Due to the higher mean reversion of variance, an upward/downward bias appears in $\widehat{V}_{t,T_1}(\widehat{u}_{t,T_1})$ when current variance is below/above its mean. These biases are reduced when using the two-tenor variance estimator $\widehat{V}_{t,T_1,T_2}(\widehat{u}_{t,T_1})$ in the high and low volatility regimes when these effects are the strongest. The results for scenario D4, in which the volatility of volatility parameter is increased, reveal that this feature of the volatility dynamics tends to generate downward bias in $\widehat{V}_{t,T_1}(\widehat{u}_{t,T_1})$. This bias is offset by the variance mean reversion effect in the low volatility regime. As a result, the use of $\widehat{V}_{t,T_1,T_2}(\widehat{u}_{t,T_1})$ reduces the bias in $\widehat{V}_{t,T_1}(\widehat{u}_{t,T_1})$ in the high and medium volatility regimes only.

Overall, the bias in the two-tenor variance estimators is small in relative terms in all scenarios D1-D4. In the high volatility regime, where the bias in $\widehat{V}_{t,T_1}(\widehat{u}_{t,T_1})$ due to the volatility of volatility and mean reversion in variance is of the same sign, the use of the two-tenor estimator $\widehat{V}_{t,T_1,T_2}(\widehat{u}_{t,T_1})$ always helps reduce this bias. In the low (and in some cases median) volatility regime, in which the biases due to volatility of volatility and variance mean reversion can partially cancel out, $\widehat{V}_{t,T_1}(\widehat{u}_{t,T_1})$ can have smaller bias than $\widehat{V}_{t,T_1,T_2}(\widehat{u}_{t,T_1})$. That said, the relative bias in $\widehat{V}_{t,T_1,T_2}(\widehat{u}_{t,T_1})$ in these cases remains very small (less than 6.1% when $T_1 = 3/252$ and $T_2 = 5/252$).[7]

We turn next to the results for scenario D5 in which we increase leverage effect, mean reversion in variance and volatility of volatility, relative to the baseline case D1. As such this scenario is somewhat unrealistic but it is nevertheless useful to study the bias in volatility estimation. Now, for all starting values of variance, $\widehat{V}_{t,T_1}(\widehat{u}_{t,T_1})$ has nontrivial downward bias. With the exception of the lowest volatility regime, $\widehat{V}_{t,T_1,T_2}(\widehat{u}_{t,T_1})$ reduces this bias, and in the high volatility regime this reduction is by a factor of two. Nevertheless, $\widehat{V}_{t,T_1,T_2}(\widehat{u}_{t,T_1})$ is also downward biased which

[7]Some of these small differences can be also due to the use of finite number of options on discrete strike grid as well as the numerical error in computing the option prices from the model.

means that higher order terms in the characteristic function expansion, beyond those considered for constructing $\widehat{V}_{t,T_1,T_2}(\widehat{u}_{t,T_1})$, seem to have some effect on the estimation.

Finally, comparing $\widehat{V}_{t,T_1,T_2}(\widehat{u}_{t,T_1})$, $\widehat{V}_{t,T_1,T_3}(\widehat{u}_{t,T_1})$ and $\widehat{V}_{t,T_2,T_3}(\widehat{u}_{t,T_2})$, we can see that the three estimators have similar biases (with the exception of scenarios D4 and D5 for medium and high levels of current variance, for which $\widehat{V}_{t,T_1,T_2}(\widehat{u}_{t,T_1})$ has the smallest bias). They differ, however, in terms of their precision, with $\widehat{V}_{t,T_1,T_3}(\widehat{u}_{t,T_1})$ being the least noisy of the three estimators. This is to be expected as $\log(\mathcal{L}_{t,T}(u))$ is approximately constant in T, the option errors are proportional to the true option prices, and T_3/T_1 is higher than T_2/T_1 and T_3/T_2.

6.3 Bias-Reduction in Variance Estimation due to the Price Jumps

We next study the bias in the variance estimation due to the presence of jumps as well as the performance of the techniques of Section 4 aimed at removing this bias. The tuning parameters of the variance estimator $\widehat{V}_t(\underline{u}, \widehat{\beta}_{\mathbb{T}}(\underline{u}))$ are set as follows. First, $\mathbb{T}$ contains one element which is t, i.e., the current point in time. For $\underline{u}$, used in the computation of $\widehat{\beta}_{\mathbb{T}}(\underline{u})$, we use the vector $\underline{\widehat{u}}_{\mathbb{T},T_1}$ which we define in the following way. It is of size $K = 20$, and the logarithms of its elements form equidistant grid. The first and last element of $\underline{\widehat{u}}_{\mathbb{T},T_1}$ are given by

$$\underline{\widehat{u}}_{\mathbb{T},T_1}(1) = \inf\{u \geq 0 : |\widehat{\mathcal{L}}_{\mathbb{T},T_1}(u)| \leq 0.8\} \wedge \operatorname{argmin}_{u\in[0,\overline{u}_{\mathbb{T},T_1}]} |\widehat{\mathcal{L}}_{\mathbb{T},T_1}(u)|, \tag{54}$$

$$\underline{\widehat{u}}_{\mathbb{T},T_1}(K) = \inf\{u \geq 0 : |\widehat{\mathcal{L}}_{\mathbb{T},T_1}(u)| \leq 0.3\} \wedge \operatorname{argmin}_{u\in[0,\overline{u}_{\mathbb{T},T_1}]} |\widehat{\mathcal{L}}_{\mathbb{T},T_1}(u)|, \tag{55}$$

where $\overline{u}_{\mathbb{T},T_1} = \sqrt{-2\log(0.05)}/\widehat{\sigma}_{\mathbb{T},ATM}$, and use the notation $\widehat{\mathcal{L}}_{\mathbb{T},T_1}(u) = \frac{1}{|\mathbb{T}|}\sum_{t\in\mathbb{T}} \widehat{\mathcal{L}}_{t,T_1}(u)$ as well as $\widehat{\sigma}^2_{\mathbb{T},ATM} = \frac{1}{|\mathbb{T}|}\sum_{t\in\mathbb{T}} \widehat{\sigma}^2_{t,ATM}$.

With the above choice of tuning parameters, we compare the performance of the feasible estimators $\widehat{V}_{t,T_1}(\widehat{u}_{t,T_1})$, $\widehat{V}_{t,T_1,T_2}(\widehat{u}_{t,T_1})$ and $\widehat{V}_t(\underline{\widehat{u}}_{t,T_1}, \widehat{\beta}_{\mathbb{T}}(\underline{\widehat{u}}_{\mathbb{T},T_1}))$. The results from the Monte Carlo are reported in Table 3. Overall, once jumps are added in the underlying asset price dynamics, all negative biases in $\widehat{V}_{t,T_1}(\widehat{u}_{t,T_1})$ reported in the previous section become positive. This is because the presence of jumps always generates an upward bias in variance estimation. This positive bias in the variance recovery tends to dominate any possible negative bias that is due to the volatility dynamics. Not surprisingly, the upward bias in $\widehat{V}_{t,T_1}(\widehat{u}_{t,T_1})$ is smallest in scenario J1 and it gradually increases when going from case J1 to case J4. The reason for this is that the $\mathfrak{b}$ parameter, controlling the jump activity, increases from -1.5 in case J1 to 0.5 in case J4.[8] Higher jump activity implies higher concentration of small jumps, which are harder to disentangle from a diffusion, and this leads to higher bias in variance estimation. Consistent with our theoretical results, in all considered cases, $\widehat{V}_{t,T_1,T_2}(\widehat{u}_{t,T_1})$ reduces the positive bias due to the jumps in the variance estimation.

[8]The parameters in J1-J4 are set so that $\int_{\mathbb{R}} x^2\nu_t(x)dx$ remains the same across the different jump scenarios.

Table 2: Monte Carlo Results for the Diffusive Models

Estimator	Bias	STD	RMSE	Bias	STD	RMSE	Bias	STD	RMSE
Panel A: Case D1									
	$V_t = 0.0118$			$V_t = 0.0192$			$V_t = 0.0293$		
$\widehat{V}_{t,T_1}$	0.0002	0.0003	0.0004	-0.0002	0.0005	0.0005	-0.0004	0.0006	0.0008
$\widehat{V}_{t,T_2}$	0.0005	0.0003	0.0006	-0.0001	0.0004	0.0004	-0.0010	0.0005	0.0011
$\widehat{V}_{t,T_3}$	0.0008	0.0003	0.0009	-0.0003	0.0003	0.0005	-0.0018	0.0005	0.0019
$\widehat{V}_{t,T_1,T_2}$	-0.0002	0.0009	0.0009	-0.0004	0.0013	0.0013	0.0004	0.0017	0.0018
$\widehat{V}_{t,T_1,T_3}$	-0.0001	0.0005	0.0005	-0.0001	0.0007	0.0007	0.0002	0.0009	0.0009
$\widehat{V}_{t,T_2,T_3}$	0.0001	0.0006	0.0006	0.0002	0.0009	0.0009	-0.0002	0.0012	0.0012
Panel B: Case D2									
	$V_t = 0.0118$			$V_t = 0.0192$			$V_t = 0.0293$		
$\widehat{V}_{t,T_1}$	0.0001	0.0003	0.0003	-0.0004	0.0005	0.0006	-0.0004	0.0006	0.0007
$\widehat{V}_{t,T_2}$	0.0004	0.0003	0.0005	-0.0003	0.0004	0.0005	-0.0011	0.0005	0.0012
$\widehat{V}_{t,T_3}$	0.0006	0.0003	0.0006	-0.0006	0.0003	0.0007	-0.0021	0.0005	0.0021
$\widehat{V}_{t,T_1,T_2}$	-0.0003	0.0009	0.0010	-0.0004	0.0013	0.0014	0.0006	0.0017	0.0018
$\widehat{V}_{t,T_1,T_3}$	-0.0001	0.0005	0.0005	-0.0003	0.0007	0.0007	0.0003	0.0009	0.0009
$\widehat{V}_{t,T_2,T_3}$	0.0002	0.0006	0.0006	-0.0001	0.0009	0.0009	-0.0002	0.0012	0.0012
Panel C: Case D3									
	$V_t = 0.0120$			$V_t = 0.0192$			$V_t = 0.0290$		
$\widehat{V}_{t,T_1}$	0.0010	0.0004	0.0010	-0.0004	0.0005	0.0006	-0.0022	0.0006	0.0023
$\widehat{V}_{t,T_2}$	0.0018	0.0003	0.0018	-0.0003	0.0004	0.0005	-0.0032	0.0005	0.0032
$\widehat{V}_{t,T_3}$	0.0029	0.0003	0.0029	-0.0005	0.0003	0.0006	-0.0051	0.0004	0.0051
$\widehat{V}_{t,T_1,T_2}$	-0.0003	0.0010	0.0010	-0.0006	0.0013	0.0014	-0.0008	0.0017	0.0019
$\widehat{V}_{t,T_1,T_3}$	0.0002	0.0005	0.0005	-0.0004	0.0007	0.0008	-0.0010	0.0009	0.0014
$\widehat{V}_{t,T_2,T_3}$	0.0007	0.0007	0.0010	-0.0001	0.0009	0.0009	-0.0014	0.0011	0.0018
Panel D: Case D4									
	$V_t = 0.0033$			$V_t = 0.0153$			$V_t = 0.0429$		
$\widehat{V}_{t,T_1}$	0.0000	0.0002	0.0002	-0.0004	0.0004	0.0006	-0.0019	0.0008	0.0021
$\widehat{V}_{t,T_2}$	0.0002	0.0001	0.0002	-0.0008	0.0004	0.0009	-0.0032	0.0007	0.0033
$\widehat{V}_{t,T_3}$	0.0009	0.0001	0.0009	-0.0016	0.0003	0.0016	-0.0059	0.0006	0.0059
$\widehat{V}_{t,T_1,T_2}$	-0.0002	0.0004	0.0004	0.0002	0.0011	0.0011	-0.0001	0.0023	0.0023
$\widehat{V}_{t,T_1,T_3}$	-0.0003	0.0002	0.0004	0.0000	0.0006	0.0006	-0.0004	0.0012	0.0012
$\widehat{V}_{t,T_2,T_3}$	-0.0004	0.0003	0.0005	-0.0002	0.0007	0.0007	-0.0008	0.0015	0.0017
Panel E: Case D5									
	$V_t = 0.0036$			$V_t = 0.0155$			$V_t = 0.0424$		
$\widehat{V}_{t,T_1}$	0.0001	0.0003	0.0003	-0.0037	0.0004	0.0038	-0.0093	0.0008	0.0094
$\widehat{V}_{t,T_2}$	0.0005	0.0003	0.0006	-0.0047	0.0004	0.0047	-0.0133	0.0007	0.0133
$\widehat{V}_{t,T_3}$	0.0020	0.0003	0.0020	-0.0055	0.0004	0.0055	-0.0199	0.0006	0.0199
$\widehat{V}_{t,T_1,T_2}$	-0.0003	0.0005	0.0006	-0.0027	0.0010	0.0029	-0.0043	0.0020	0.0047
$\widehat{V}_{t,T_1,T_3}$	-0.0004	0.0003	0.0005	-0.0032	0.0006	0.0033	-0.0057	0.0011	0.0058
$\widehat{V}_{t,T_2,T_3}$	-0.0004	0.0005	0.0006	-0.0042	0.0007	0.0043	-0.0082	0.0013	0.0083

Note: We use the shorthand notation $\widehat{V}_{t,T}$ for $\widehat{V}_{t,T}(\widehat{u}_{t,T})$ and $\widehat{V}_{t,T_1,T_2}$ for $\widehat{V}_{t,T_1,T_2}(\widehat{u}_{t,T_1})$. Reported results are based on $5,000$ Monte Carlo replications. STD stands for standard deviation and RMSE for root mean squared error. The top row of each panel reports the true value of spot variance.

As a result, the bias in $\widehat{V}_{t,T_1,T_2}(\widehat{u}_{t,T_1})$ in the finite activity jump cases J1 and J2 become minimal in relative terms (they are of relative size of up to 6.5% in the case $T_1 = 3/252$ and $T_2 = 5/252$).

Table 3: Monte Carlo Results for the Jump-Diffusion Models, Part I

	$V_t = 0.0118$			$V_t = 0.0192$			$V_t = 0.0293$		
Estimator	Bias	STD	RMSE	Bias	STD	RMSE	Bias	STD	RMSE
Panel A: Case J1									
$\widehat{V}_{t,T_1}$	0.0006	0.0004	0.0007	0.0009	0.0007	0.0011	0.0014	0.0008	0.0016
$\widehat{V}_{t,T_2}$	0.0010	0.0004	0.0011	0.0014	0.0005	0.0015	0.0017	0.0008	0.0019
$\widehat{V}_{t,T_3}$	0.0018	0.0003	0.0019	0.0021	0.0005	0.0021	0.0029	0.0007	0.0030
$\widehat{V}_{t,T_1,T_2}$	0.0001	0.0010	0.0010	0.0000	0.0019	0.0019	0.0011	0.0021	0.0024
$\widehat{V}_{t,T_1,T_3}$	0.0001	0.0006	0.0006	0.0004	0.0010	0.0010	0.0009	0.0011	0.0014
$\widehat{V}_{t,T_2,T_3}$	0.0002	0.0009	0.0009	0.0008	0.0010	0.0013	0.0006	0.0016	0.0017
$\widehat{V}^{\beta}_{t,T_1,T_2}$	0.0000	0.0013	0.0013	-0.0005	0.0020	0.0021	-0.0010	0.0030	0.0031
$\widehat{V}^{\beta}_{t,T_1,T_3}$	-0.0001	0.0007	0.0007	-0.0002	0.0013	0.0013	-0.0005	0.0020	0.0020
$\widehat{V}^{\beta}_{t,T_2,T_3}$	-0.0001	0.0010	0.0010	0.0000	0.0016	0.0016	-0.0002	0.0025	0.0025
Panel B: Case J2									
$\widehat{V}_{t,T_1}$	0.0012	0.0004	0.0012	0.0019	0.0006	0.0020	0.0039	0.0010	0.0040
$\widehat{V}_{t,T_2}$	0.0018	0.0004	0.0018	0.0029	0.0006	0.0030	0.0052	0.0008	0.0053
$\widehat{V}_{t,T_3}$	0.0034	0.0004	0.0034	0.0044	0.0005	0.0045	0.0067	0.0007	0.0067
$\widehat{V}_{t,T_1,T_2}$	0.0003	0.0011	0.0012	0.0006	0.0016	0.0017	0.0019	0.0026	0.0033
$\widehat{V}_{t,T_1,T_3}$	0.0004	0.0006	0.0007	0.0009	0.0009	0.0013	0.0028	0.0014	0.0031
$\widehat{V}_{t,T_2,T_3}$	0.0004	0.0008	0.0009	0.0016	0.0013	0.0021	0.0040	0.0016	0.0043
$\widehat{V}^{\beta}_{t,T_1,T_2}$	-0.0008	0.0015	0.0017	-0.0004	0.0025	0.0025	-0.0001	0.0037	0.0037
$\widehat{V}^{\beta}_{t,T_1,T_3}$	-0.0005	0.0010	0.0011	-0.0002	0.0017	0.0017	-0.0009	0.0022	0.0024
$\widehat{V}^{\beta}_{t,T_2,T_3}$	0.0000	0.0013	0.0013	-0.0003	0.0020	0.0020	-0.0019	0.0021	0.0028
Panel C: Case J3									
$\widehat{V}_{t,T_1}$	0.0017	0.0005	0.0018	0.0034	0.0007	0.0035	0.0068	0.0009	0.0068
$\widehat{V}_{t,T_2}$	0.0027	0.0005	0.0028	0.0046	0.0006	0.0047	0.0078	0.0009	0.0078
$\widehat{V}_{t,T_3}$	0.0044	0.0004	0.0044	0.0063	0.0006	0.0064	0.0102	0.0008	0.0102
$\widehat{V}_{t,T_1,T_2}$	0.0004	0.0012	0.0013	0.0017	0.0019	0.0026	0.0054	0.0024	0.0059
$\widehat{V}_{t,T_1,T_3}$	0.0007	0.0007	0.0010	0.0023	0.0010	0.0025	0.0056	0.0013	0.0057
$\widehat{V}_{t,T_2,T_3}$	0.0012	0.0010	0.0016	0.0032	0.0012	0.0034	0.0058	0.0017	0.0060
$\widehat{V}^{\beta}_{t,T_1,T_2}$	-0.0001	0.0018	0.0018	-0.0002	0.0026	0.0026	0.0000	0.0031	0.0031
$\widehat{V}^{\beta}_{t,T_1,T_3}$	-0.0001	0.0012	0.0012	-0.0007	0.0013	0.0014	-0.0003	0.0014	0.0014
$\widehat{V}^{\beta}_{t,T_2,T_3}$	-0.0003	0.0013	0.0013	-0.0007	0.0014	0.0015	-0.0003	0.0020	0.0020
Panel D: Case J4									
$\widehat{V}_{t,T_1}$	0.0033	0.0005	0.0033	0.0057	0.0007	0.0058	0.0101	0.0010	0.0102
$\widehat{V}_{t,T_2}$	0.0042	0.0004	0.0042	0.0069	0.0007	0.0070	0.0122	0.0009	0.0122
$\widehat{V}_{t,T_3}$	0.0062	0.0004	0.0062	0.0094	0.0006	0.0094	0.0146	0.0008	0.0146
$\widehat{V}_{t,T_1,T_2}$	0.0020	0.0013	0.0024	0.0041	0.0018	0.0045	0.0076	0.0026	0.0080
$\widehat{V}_{t,T_1,T_3}$	0.0022	0.0007	0.0023	0.0044	0.0009	0.0045	0.0085	0.0014	0.0086
$\widehat{V}_{t,T_2,T_3}$	0.0024	0.0009	0.0026	0.0049	0.0013	0.0050	0.0101	0.0019	0.0102
$\widehat{V}^{\beta}_{t,T_1,T_2}$	0.0001	0.0015	0.0015	0.0006	0.0021	0.0022	0.0026	0.0031	0.0040
$\widehat{V}^{\beta}_{t,T_1,T_3}$	0.0001	0.0007	0.0008	0.0009	0.0011	0.0014	0.0033	0.0016	0.0037
$\widehat{V}^{\beta}_{t,T_2,T_3}$	0.0002	0.0010	0.0011	0.0014	0.0015	0.0021	0.0047	0.0021	0.0051

Note: We use the shorthand notation of Table 2 as well as $\widehat{V}^{\beta}_{t,T_1,T_2}$ for $\widehat{V}_t(\underline{\widehat{u}}_{t,T_1}, \widehat{\beta}_{\mathbb{T}}(\underline{\widehat{u}}_{t,T_1}))$. Reported results are based on $5,000$ Monte Carlo replications. STD stands for standard deviation and RMSE for root mean squared error. The top row reports the true value of spot variance.

From the results in Table 3, we can see that the estimator $\widehat{V}_t(\underline{\widehat{u}}_{t,T_1}, \widehat{\beta}_{\mathbb{T}}(\underline{\widehat{u}}_{\mathbb{T},T_1}))$, in which we bias-

correct for the effect of jumps, tends to work very well. In particular, in case J1 for which the debiasing for the presence of jumps is not needed because of their low activity, $\widehat{V}_t(\underline{\widehat{u}}_{t,T_1}, \widehat{\beta}_{\mathbb{T}}(\underline{\widehat{u}}_{\mathbb{T},T_1}))$ exhibits only very small negative bias and its volatility is only slightly higher than that of $\widehat{V}_{t,T_1,T_2}(\widehat{u}_{t,T_1})$. On the other hand, $\widehat{V}_t(\underline{\widehat{u}}_{t,T_1}, \widehat{\beta}_{\mathbb{T}}(\underline{\widehat{u}}_{\mathbb{T},T_1}))$ removes most of the positive bias in the variance estimation when jumps are of infinite activity, i.e., in cases J3 and J4. Indeed, when $T_1 = 3/252$ and $T_2 = 5/252$, the bias in $\widehat{V}_t(\underline{\widehat{u}}_{t,T_1}, \widehat{\beta}_{\mathbb{T}}(\underline{\widehat{u}}_{\mathbb{T},T_1}))$ does not exceed 8.9% in relative terms. This is to be contrasted with the relative bias up to 41.6% in $\widehat{V}_{t,T}(\widehat{u}_{t,T})$ (for $T = T_1$ or $T = T_2$) in scenario J4 when current variance is at its highest level. Thus, the bias-correction for the jumps seems to work well in finite samples.

Table 4: Monte Carlo Results for the Jump-Diffusion Models, Part II

	$V_t = 0.0025$			$V_t = 0.0750$			$V_t = 0.0025$			$V_t = 0.0750$		
Estimator	Bias	STD	RMSE	Bias	STD	RMSE	Bias	STD	RMSE	Bias	STD	RMSE
	Case J1						**Case J2**					
$\widehat{V}_{t,T_1}$	0.0009	0.0002	0.0009	0.0085	0.0019	0.0087	0.0006	0.0001	0.0006	0.0174	0.0020	0.0175
$\widehat{V}_{t,T_2}$	0.0011	0.0002	0.0011	0.0100	0.0016	0.0101	0.0012	0.0002	0.0012	0.0218	0.0018	0.0219
$\widehat{V}_{t,T_3}$	0.0024	0.0002	0.0024	0.0127	0.0015	0.0128	0.0027	0.0002	0.0027	0.0267	0.0015	0.0268
$\widehat{V}_{t,T_1,T_2}$	0.0006	0.0004	0.0008	0.0065	0.0049	0.0081	-0.0002	0.0004	0.0005	0.0119	0.0050	0.0129
$\widehat{V}_{t,T_1,T_3}$	0.0003	0.0002	0.0004	0.0070	0.0026	0.0075	-0.0001	0.0002	0.0003	0.0141	0.0027	0.0143
$\widehat{V}_{t,T_2,T_3}$	-0.0001	0.0004	0.0004	0.0077	0.0032	0.0084	-0.0001	0.0004	0.0004	0.0177	0.0036	0.0181
$\widehat{V}^{\beta}_{t,T_1,T_2}$	0.0005	0.0003	0.0006	0.0006	0.0073	0.0073	-0.0002	0.0005	0.0005	-0.0052	0.0060	0.0080
$\widehat{V}^{\beta}_{t,T_1,T_3}$	0.0003	0.0002	0.0003	-0.0033	0.0049	0.0059	-0.0002	0.0003	0.0003	-0.0034	0.0031	0.0046
$\widehat{V}^{\beta}_{t,T_2,T_3}$	-0.0002	0.0004	0.0004	-0.0103	0.0044	0.0112	-0.0001	0.0004	0.0004	0.0002	0.0039	0.0039
	Case J3						**Case J4**					
$\widehat{V}_{t,T_1}$	0.0010	0.0002	0.0010	0.0258	0.0021	0.0259	0.0012	0.0002	0.0012	0.0374	0.0021	0.0375
$\widehat{V}_{t,T_2}$	0.0015	0.0002	0.0015	0.0312	0.0019	0.0312	0.0016	0.0002	0.0016	0.0414	0.0018	0.0415
$\widehat{V}_{t,T_3}$	0.0029	0.0002	0.0030	0.0344	0.0016	0.0345	0.0036	0.0002	0.0036	0.0417	0.0015	0.0418
$\widehat{V}_{t,T_1,T_2}$	0.0004	0.0005	0.0006	0.0190	0.0055	0.0198	0.0007	0.0005	0.0008	0.0320	0.0056	0.0325
$\widehat{V}_{t,T_1,T_3}$	0.0003	0.0003	0.0004	0.0226	0.0030	0.0228	0.0005	0.0002	0.0005	0.0357	0.0030	0.0358
$\widehat{V}_{t,T_2,T_3}$	0.0002	0.0004	0.0004	0.0283	0.0037	0.0285	0.0002	0.0004	0.0004	0.0411	0.0037	0.0413
$\widehat{V}^{\beta}_{t,T_1,T_2}$	0.0003	0.0004	0.0005	0.0031	0.0063	0.0070	0.0006	0.0004	0.0008	0.0169	0.0067	0.0182
$\widehat{V}^{\beta}_{t,T_1,T_3}$	0.0002	0.0002	0.0003	0.0067	0.0033	0.0074	0.0004	0.0003	0.0005	0.0211	0.0034	0.0214
$\widehat{V}^{\beta}_{t,T_2,T_3}$	0.0000	0.0004	0.0004	0.0119	0.0042	0.0126	0.0000	0.0005	0.0005	0.0272	0.0043	0.0275

Note: We use the shorthand notation of Table 2 as well as $\widehat{V}^{\beta}_{t,T_1,T_2}$ for $\widehat{V}_t(\underline{\widehat{u}}_{t,T_1}, \widehat{\beta}_{\mathbb{T}}(\underline{\widehat{u}}_{t,T_1}))$. Reported results are based on 5,000 Monte Carlo replications. STD stands for standard deviation and RMSE for root mean squared error. The top row reports the true value of spot variance.

We finish this section with presenting results for the case when the starting values of the variance are very low and very high, mainly $V_t = 0.0025$ and $V_t = 0.0750$. These values match roughly the 10-th and 90-th quantiles of the spot variance estimated in our empirical application but they correspond to more extreme quantiles for the unconditional distribution of V_t in our Monte Carlo experiments. The results from this experiment are reported in Table 4. Now the biases in $\widehat{V}_{t,T}(\widehat{u}_{t,T})$ due to mean reversion in volatility are much stronger because the starting values of V_t are further

away from its unconditional mean. As a result, $\widehat{V}_{t,T}(\widehat{u}_{t,T})$ is significantly upward biased in the low volatility regime for all considered values of T. Overall, even in these more extreme settings, the bias-correction techniques proposed in the paper appear to work well.

6.4 Sensitivity to Strike Range in Variance Recovery

In the experiments so far, the cheapest out-of-the-money puts and calls for any pair (t, T) have prices of around 0.075. This matches the S&P 500 Index option data set, used in our empirical application, for most of the day-maturity pairs.[9] Nevertheless, it is important to investigate the sensitivity of the volatility option recovery to the width of the strike range of available options.

Before presenting the results from the experiments in this section, we make a few general comments in this regard. First, as shown in Todorov (2019), $\widehat{V}_{t,T}(u)$ remains consistent even if the log-strike range is asymptotically finite. This, of course, carries over to the other bias-corrected variance estimators proposed in this paper. The reason for this is that the value of $\widehat{V}_{t,T}(u)$ is derived from options with strikes in the vicinity of the current stock price, see e.g., Figure 1 in Todorov and Zhang (2021) for an empirical illustration of this. Second, if there are a lot of missing option observations for very small and/or very large strikes, then one can do a tail extension based on extreme value theory to approximate the prices for the missing deep out-of-the-money option prices as done in Todorov and Zhang (2021) for individual stock options. Third, the size of the error from the truncation of the strike range in the computation of $\widehat{\mathcal{L}}_{t,T}(u)$ naturally depends on how fat the return distribution is, which in our short-tenor case is governed by how fat the jump tails are (if jumps are present of course).

We will now illustrate the sensitivity of the variance recovery to the strike range using one of our Monte Carlo cases, mainly case J3. For this experiment, we will assume no option observation error in order to better highlight the role of truncation of the available strike range. In Figure 3, we plot the bias in $\widehat{V}_{t,T}(\widehat{u}_{t,T})$ as a function of the degree of truncation of the strike range, with the latter measured as a percentage of the maximum of the deepest available out-of-the-money call and put over the most expensive available option (i.e., the one closest-to-the-money). The range of strike truncation is calibrated to the data of our empirical application. As seen from the figure, the sensitivity $\widehat{V}_{t,T}(\widehat{u}_{t,T})$ to strike truncation is not very big. For example, the relative bias in $\widehat{V}_{t,T}(\widehat{u}_{t,T})$ for the strike cutoff corresponding to the 75-th empirical quantile of this quantity for our data is at most around 3% and for the 90-th empirical quantile of the strike range cutoff, the biggest change in the relative bias of $\widehat{V}_{t,T}(\widehat{u}_{t,T})$ is up to 5%. For this reason, just like for the calculation of the

[9] The minimum tick for the S&P 500 option quotes is 0.05, and since we exclude zero bid quotes, the minimum mid-quote for option prices in our empirical analysis is 0.075.

VIX volatility index, we do not do tail adjustments to $\widehat{\mathcal{L}}_{t,T}(u)$.[10]

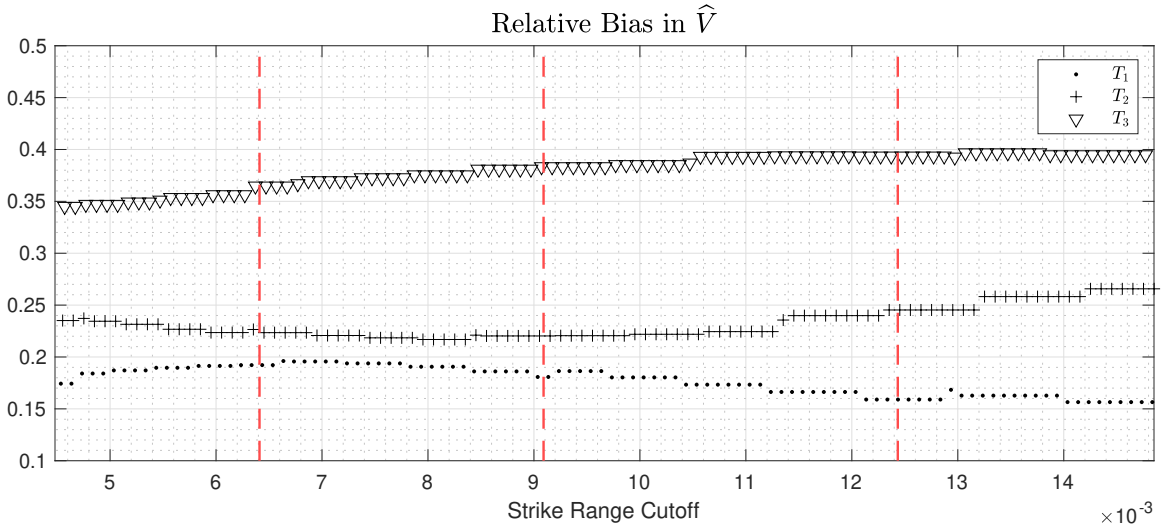

Figure 3: **Sensitivity of $\widehat{V}_{t,T}(u)$ to Strike Range**. The ratio on the x-axis is the maximum of the deepest out-of-the-money put and call divided by the most expensive out-of-the-money option for (t,T). The vertical lines from left to right correspond to the 50-th, 75-th and 90-th empirical quantiles of this quantity for the S&P 500 Index Option data used in the empirical application. The y-axis is the relative bias $\widehat{V}_{t,T}(\widehat{u}_{t,T})/V_t - 1$. The option data is generated from model J3 for medium level of current variance.

7 Empirical Application

We apply the newly-developed variance estimators to options written on the S&P 500 market index. Our sample period is from January 2, 2010 until December 31, 2020. The data is obtained from OptionMetrics. It consists of closing best bid and best ask quotes. For each day in the sample, we keep the two shortest available tenors with time-to-maturity of at least 2 business days and $T_2/T_1 \geq 1.5$. We consider only tenors for which there are at least 5 out-of-the-money options with different strikes and non-zero bid quotes. The moneyness is determined by the forward rate, which in turn is computed by using put-call parity for three distinct strikes with the smallest gap between call and put mid-quotes. The gap between the strikes of observed options is typically five, and when this is not the case, we conduct linear interpolation in BSIV space to create option prices on equidistant strike grid. Finally, we remove from the analysis, the dates just before Brexit vote, 2016 US election and the 2017 April IMF conference, which were days with large pre-scheduled

[10] If we use the tail extension, discussed above, in computing $\widehat{\mathcal{L}}_{t,T}(u)$, the biases reported in Figure 3 essentially disappear. These results are available upon request.

jump events.[11] We further remove from the analysis May 6, 2010, which contained a flash-crash event, and the following trading day. With these filters in place, the median days to maturity T_1 and T_2 in our sample are 3 and 7 business days, respectively. From the option data, we compute the combined variance estimator $\widehat{V}_t = \frac{1}{2}\left(\widehat{V}_{t,T_1}(\widehat{u}_{t,T_1}) + \widehat{V}_{t,T_2}(\widehat{u}_{t,T_2})\right)$ as well as the bias-corrected $\widehat{V}_{t,T_1,T_2}(\widehat{u}_{t,T_1})$ and $\widehat{V}_t(\underline{\widehat{u}}_{t,T_1}, \widehat{\beta}_{\mathbb{T}}(\underline{\widehat{u}}_{t,T_1}))$. The tuning parameters of these estimators are set exactly as in the Monte Carlo. We also report average at-the-money BSIV computed from the two tenors, denoted with AIV_t (and reported in annualized variance units).

We compliment the option data with one-minute price records of the SPY exchange traded fund (ETF), which is the ETF on the S&P 500 index. We consider observations during the regular trading hours from 9.30 am till 4.00 pm EST. In order to avoid potential adverse effects in the opening minutes of trading, we start sampling at 9.35 am on each day. We also exclude from the analysis days with more than 25% of zero returns (which occur typically around holidays). A unit interval from market close on the previous day to market close on the current day, $[t-1, t]$, consists of overnight period $[t-1, t-1+\kappa]$ and a trading interval $[t-1+\kappa, t]$, for some $\kappa \in [0,1]$. The number of increments over the trading interval are n, and are denoted with

$$\Delta^n_{t,i} x = x_{t-1+\kappa+i(1-\kappa)/n} - x_{t-1+\kappa+(i-1)(1-\kappa)/n}, \quad t = 1, 2, ..., \ i = 1, ..., n. \tag{56}$$

From the return data, we construct two alternative estimates of the spot variance at market close. The first is a local version of the truncated variance estimator of Mancini (2001), defined as

$$TV_t = \frac{n}{k_n} \sum_{i=n-k_n+1}^{n} (\Delta^n_{t,i} x)^2 1_{\{|\Delta^n_{t,i} x| \leq 3\sqrt{RV_t \wedge BV_t}/\sqrt{n}\}}, \tag{57}$$

where k_n is an integer in $[1, n]$, which here we set to $k_n = 60$ (corresponding to one hour window), and

$$RV_t = \frac{n}{n-k_n} \sum_{i=1}^{n-k_n} (\Delta^n_{t,i} x)^2, \ BV_t = \frac{n}{n-k_n} \frac{\pi}{2} \sum_{i=2}^{n-k_n} |\Delta^n_{t,i-1} x||\Delta^n_{t,i} x|, \tag{58}$$

with the first quantity in (58) being the realized variance over the part of the trading day not including the local window before market close and the second one being the jump-robust bipower variation estimator over the same time window. The use of time-varying threshold in defining TV_t is done in order to account for the time-varying diffusive variance. Naturally, this estimator is going to be sensitive to the choice of threshold level. We therefore compute an alternative to it based on the empirical characteristic function of the returns, which has a built-in truncation for the jumps

[11] Jumps with known arrival times change the short-term asymptotics for the return process, see Todorov (2020).

in the increments. It is defined as

$$LV_t = -\frac{2}{\widehat{u}_t^2}\log\left|\frac{1}{k_n}\sum_{i=n-k_n+1}^{n}\exp(i\widehat{u}_t\sqrt{n}\Delta_{t,i}^n x)\right|,\quad \widehat{u}_t = \frac{\sqrt{-2\log(0.3)}}{\sqrt{RV_t \wedge BV_t}}. \tag{59}$$

LV_t was introduced in Jacod and Todorov (2014) (its integrated over time version to be precise), see also Liu et al. (2018), and is the return-counterpart of $\widehat{V}_{t,T}(\widehat{u}_{t,T})$ in which we use the empirical characteristic function of the high-frequency returns. Note also that our choice of the characteristic exponent $\widehat{u}_t$ for LV_t corresponds directly to the choice of $\widehat{u}_{t,T}$ for $\widehat{V}_{t,T}(\widehat{u}_{t,T})$. We annualize TV_t and LV_t by using an overnight adjustment factor, computed for each calendar year in the sample, by simply comparing average realized variance over the trading day with the sample variance of the overnight return.

Finally, we also extract from the high-frequency data a measure of the relative size of the infinite activity jump component of the price. In particular, we estimate

$$IJA_t = IV_t(\overline{u}/2)/IV_t(\overline{u}) - 1,\ IV_t(u) = -\frac{2}{u^2}\sum_{j=1}^{\lfloor n/k_n\rfloor}\log\left|\frac{1}{k_n}\sum_{i=(j-1)k_n+1}^{jk_n}\exp(iu\sqrt{n}\Delta_{t,i}^n x)\right|, \tag{60}$$

where $\overline{u} = \sqrt{-2\log(0.3)}/\sqrt{\overline{BV}}$ and $\overline{BV}$ is the sample average of the daily BV. Following Jacod and Todorov (2014), $IV_t(\overline{u}/2) - IV_t(\overline{u})$ is, up to a constant, an estimate of the daily integrated counterpart of $\overline{B}_{t,T_1,T_2}$ with $\tau = \sqrt{2}$ and T_1 corresponding to the length of the high-frequency increment (for $\beta > 0$). The leading component of $IV_t(\overline{u})$ is the daily integrated diffusive variance, so IJA_t can be viewed as an estimate of the importance of the infinite jump activity component of the price in variance estimation. Higher value of IJA_t corresponds to higher contribution of the infinite jump activity component of the price to the total return variation. The estimate IJA_t is noisy but it is nevertheless useful in studying the effect of the infinite activity jumps on the variance recovery.[12]

In Table 5, we report summary statistics for the various volatility estimators introduced above. Panel A of the table contains unconditional moments of the estimators. We can see from the reported results that $\sqrt{\widehat{V}_t}$ has a slightly lower mean than the bias-corrected $\sqrt{\widehat{V}_{t,T_1,T_2}(\widehat{u}_{t,T_1})}$. Looking at the quantiles of the two estimators, we can note that this is largely due to the higher level of $\sqrt{\widehat{V}_{t,T_1,T_2}(\widehat{u}_{t,T_1})}$ in high volatility regimes. Recall from our theoretical analysis and the Monte Carlo experiment that $\widehat{V}_{t,T_1,T_2}(\widehat{u}_{t,T_1})$ corrects for an upward or downward bias in variance estimation due to the volatility dynamics (depending on whether current variance is below or above

[12]Estimating the jump activity index, on the other hand, is much harder because it requires separating the scale from the shape of the jump measure near zero. As a result, reasonable precision for feasible frequencies (like the one used here) can be achieved for estimation windows of at least several months, see e.g., the Monte Carlo evidence in Jacod and Todorov (2014). Note, however, that for our purposes a proxy for $\overline{B}_{t,T_1,T_2}$ is only needed.

its unconditional level) as well as for an upward bias due to price jumps. Our empirical results indicate that the former effect is negative and dominates in the high volatility regime when the term structure of option implied volatility is downward sloping. Comparing $\sqrt{\widehat{V}_{t,T_1,T_2}(\widehat{u}_{t,T_1})}$ with $\sqrt{\widehat{V}_t(\underline{\widehat{u}}_{t,T_1}, \widehat{\beta}_{\mathbb{T}}(\underline{\widehat{u}}_{t,T_1}))}$, we see that the latter is smaller than the former, with the difference being a bias-correction for the effect of small price jumps on the volatility recovery. Nevertheless, we note that the difference between the two volatility estimators is relatively small. Finally, and not surprisingly, $\sqrt{AIV_t}$ is significantly higher than all other option-based volatility measures. This is primarily because of the bias in it that is due to price jumps.

Table 5: Summary Statistics for S&P 500 Index Volatility Measures

Panel A: Unconditional Moments

Volatility Estimator	Mean	STD	Q25	Median	Q75
$\sqrt{\widehat{V}_t}$	0.1194	0.0722	0.0774	0.1027	0.1401
$\sqrt{\widehat{V}_{t,T_1,T_2}}$	0.1239	0.0854	0.0765	0.1034	0.1441
$\sqrt{\widehat{V}^{\beta}_{t,T_1,T_2}}$	0.1161	0.0811	0.0711	0.0961	0.1362
$\sqrt{AIV_t}$	0.1468	0.0868	0.0957	0.1266	0.1706
$\sqrt{TV_t}$	0.1120	0.0865	0.0633	0.0890	0.1322
$\sqrt{LV_t}$	0.1186	0.0991	0.0631	0.0910	0.1393

Panel B: Conditional Means (Level of Volatility)

Volatility Regime	$\sqrt{\widehat{V}_t}$	$\sqrt{\widehat{V}_{t,T_1,T_2}}$	$\sqrt{\widehat{V}^{\beta}_{t,T_1,T_2}}$	$\sqrt{TV_t}$	$\sqrt{LV_t}$
Low ($AIV_t < Q_{0.1}(AIV_t)$)	0.0518	0.0491	0.0445	0.0455	0.0436
Medium ($Q_{0.25}(AIV_t) < AIV_t < Q_{0.75}(AIV_t)$)	0.1041	0.1055	0.0988	0.0944	0.0978
High ($AIV_t > Q_{0.9}(AIV_t)$)	0.2709	0.3003	0.2812	0.2834	0.3176

Panel C: Conditional Means (Jump Contribution)

Jump Contribution	$\sqrt{\widehat{V}_t}$	$\sqrt{\widehat{V}_{t,T_1,T_2}}$	$\sqrt{\widehat{V}^{\beta}_{t,T_1,T_2}}$	$\sqrt{TV_t}$	$\sqrt{LV_t}$
Low ($IJA_t < Q_{0.1}(IJA_t)$)	0.1161	0.1207	0.1133	0.1068	0.1130
Medium ($Q_{0.25}(IJA_t) < IJA_t < Q_{0.75}(IJA_t)$)	0.1032	0.1049	0.0984	0.0910	0.0948
High ($IJA_t > Q_{0.9}(IJA_t)$)	0.2313	0.2563	0.2390	0.2608	0.2880

Note: $Q_\alpha(X)$ stands for the α-quantile of the random variable X. Notation as in Table 3. All volatility numbers are reported annualized.

Turning next to the two return-based volatility estimators, we see from Panel A of Table 5 that they have very similar unconditional quantiles, with $\sqrt{LV_t}$ having slightly higher mean than $\sqrt{TV_t}$. Comparing the moments of the return-based volatility measures with those of the option-based $\sqrt{\widehat{V}_t}$, $\sqrt{\widehat{V}_{t,T_1,T_2}(\widehat{u}_{t,T_1})}$ and $\sqrt{\widehat{V}_t(\underline{\widehat{u}}_{t,T_1}, \widehat{\beta}_{\mathbb{T}}(\underline{\widehat{u}}_{t,T_1}))}$, we can see that they are in general close.

The closest option-based estimator in terms of unconditional median to the return-based ones is $\sqrt{\widehat{V}_t(\underline{\widehat{u}}_{t,T_1}, \widehat{\beta}_{\mathbb{T}}(\underline{\widehat{u}}_{t,T_1}))}$.

To better assess the differences between the volatility estimators, in Panel B of Table 5, we report their means in periods of high, low and medium volatility, where the volatility regimes are determined on the basis of the value of AIV. In the low volatility regime, $\widehat{V}_t$ is highest of the three compared option-based estimators because of the upward bias in it due to the price jumps and the volatility dynamics (option implied volatility term structure in this case is upward sloping). The opposite happens in the high volatility regime. Mainly, $\widehat{V}_t$ is the lowest of the three compared option-based estimators. The reason for this is that the bias in $\widehat{V}_t$ due to the volatility dynamics is now negative and this is corrected for in the other two option-based volatility estimates. Overall, in all three regimes, $\widehat{V}_t(\underline{\widehat{u}}_{t,T_1}, \widehat{\beta}_{\mathbb{T}}(\underline{\widehat{u}}_{t,T_1}))$ appears closest to the return-based TV_t and LV_t. We note some discrepancy in the high volatility regime between TV_t and LV_t, which is probably due to the difficulty in disentangling volatility from jumps from asset returns in such environments.

The highest improvement of $\widehat{V}_t(\underline{\widehat{u}}_{t,T_1}, \widehat{\beta}_{\mathbb{T}}(\underline{\widehat{u}}_{t,T_1}))$ over $\widehat{V}_t$, in relative terms, appears in the low and medium volatility regimes while in the Monte Carlo scenarios J1-J4 we see that the improvement across volatility regimes appears more uniform. The difference is likely due to partial cancelation of biases in $\widehat{V}_t$ and the fact that the comparison here is with the return-based variance estimates which can also contain biases.[13]

We next study the impact of the jump activity component of the price on the volatility recovery using the indicator IJA_t. From the results reported in Panel C of Table 5, we can see that for small and moderate values of IJA_t, the return-based variance estimates and the option-based ones (with the exception of AIV) are close, with $\widehat{V}_t(\underline{\widehat{u}}_{t,T_1}, \widehat{\beta}_{\mathbb{T}}(\underline{\widehat{u}}_{t,T_1}))$ performing best among the latter.[14] Interestingly, high values of IJA_t tend to occur when volatility level is high. In the high IJA_t regime, the gap between the option-based volatility proxies and the return-based ones is somewhat bigger (in relative terms). This is consistent with higher bias in the option-based volatility recovery in this case, as documented in the Monte Carlo. Note, however, that in the Monte Carlo we compare our estimates with the true volatility while here we compare the option-based estimates with the return-based ones. It is very likely, consistent with prior empirical evidence, that the return-based

[13] As discussed already, biases in $\widehat{V}_t$ depend on various features of the return process. Parametric models, such as the ones in the exponentially affine class used in most empirical work (and in the Monte Carlo here), can restrict the variation of these biases as a function of volatility. For example, the jump distribution is commonly assumed constant in existing parametric models while nonparametric evidence in Bollerslev and Todorov (2014) suggests that it can vary over time.

[14] We note that, when conditioning on the value of IJA_t, $\sqrt{\widehat{V}_t}$ and $\sqrt{\widehat{V}_t(\underline{\widehat{u}}_{t,T_1}, \widehat{\beta}_{\mathbb{T}}(\underline{\widehat{u}}_{t,T_1}))}$ have similar means. This is likely due to biases of different direction in $\widehat{V}_t$ canceling out when averaging across the days in the different regimes of IJA_t.

volatility estimates are also upward-biased when IJA_t is very high.[15] We leave comparisons of jump activity estimation from asset returns and options for future work.

In Figure 4, we plot $\sqrt{\widehat{V}_t(\widehat{\underline{u}}_{t,T_1}, \widehat{\beta}_{\mathbb{T}}(\widehat{\underline{u}}_{t,T_1}))}$ and $\sqrt{LV_t}$. Consistent with the evidence presented in Table 5, we note that the two volatility measures are very close, with no apparent systematic deviations between the two. This is true even during the very turbulent period of March 2020 at the start of the pandemic crisis. Figure 4 also reveals that the return-based LV_t is significantly noisier than its option counterpart $\widehat{V}_t(\widehat{\underline{u}}_{t,T_1}, \widehat{\beta}_{\mathbb{T}}(\widehat{\underline{u}}_{t,T_1}))$. This highlights the gains offered by option data for measuring spot volatility.

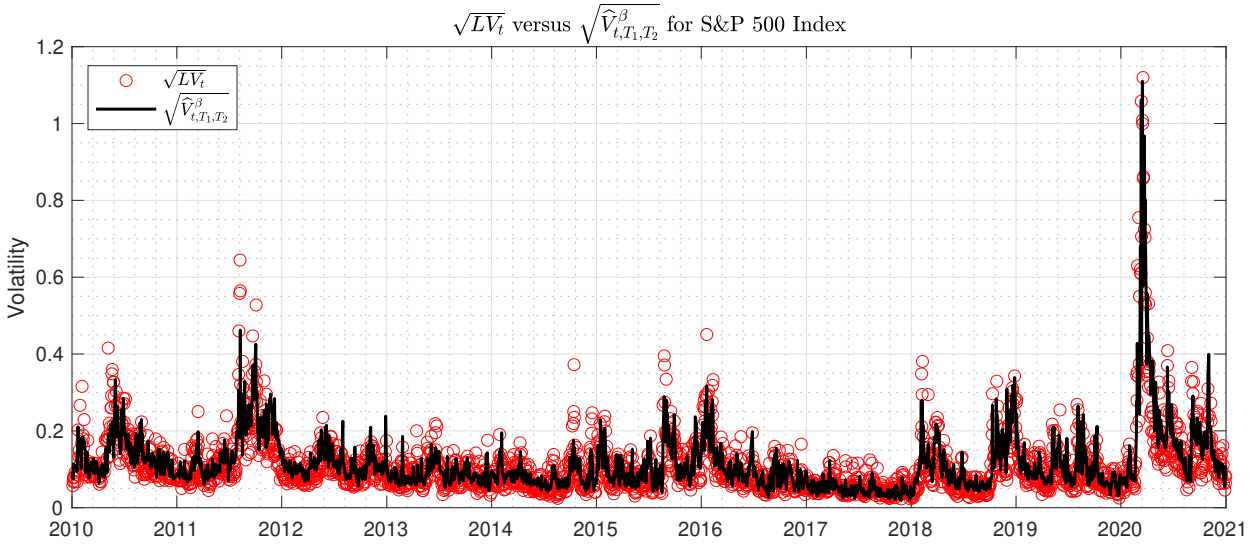

Figure 4: **S&P 500 Index Volatility Measures**. Notation as in Table 3.

To further understand the properties of potential biases present in the option-based volatility estimators, we next study the time series behavior of the difference between the option-based volatility estimates and a return-based one, which we take to be LV_t. If bias in volatility estimation is negligible, we should have approximately $\log(\widehat{OV}_t) = \log(V_t) + \epsilon_t$ and $\log(LV_t) = \log(V_t) + \zeta_t$, where $\widehat{OV}_t$ is one of the option-based volatility estimates and the two errors satisfy $\mathbb{E}_{t-1}(\zeta_t) = 0$ and $\mathbb{E}(\epsilon_t) = 0$. If option observation errors are $\mathcal{F}^{(0)}$-conditionally independent across time, then we have also the stronger condition $\mathbb{E}_{t-1}(\epsilon_t) = 0$. The above results follow from the CLT in Theorem 6 and a CLT for LV_t derived in Jacod and Todorov (2014). Thus, time series dependence in the difference between an option-based volatility estimate and a return-based one is evidence of either non-negligible (time-varying) biases in the volatility proxies and/or option observation errors with

[15] For example, the threshold in TV_t is automatically elevated when volatility is high allowing for a lot small of jumps to "escape" truncation.

time-series persistence. To study this, in Figure 5, we plot the difference in logs between the option-based volatility estimates and LV_t. The log-transformation of the volatility series is done in order to minimize the impact of outliers and because the error term in $\log(LV_t)$ is approximately i.i.d. (this follows from the CLT for LV_t). As seen from the figure, the differences are approximately homoskedastic and neither of them exhibit long persistent shifts. Nevertheless, some short-term deviations from the long-run means of the series do appear, e.g., in the beginning of 2010 or towards the end of 2019.

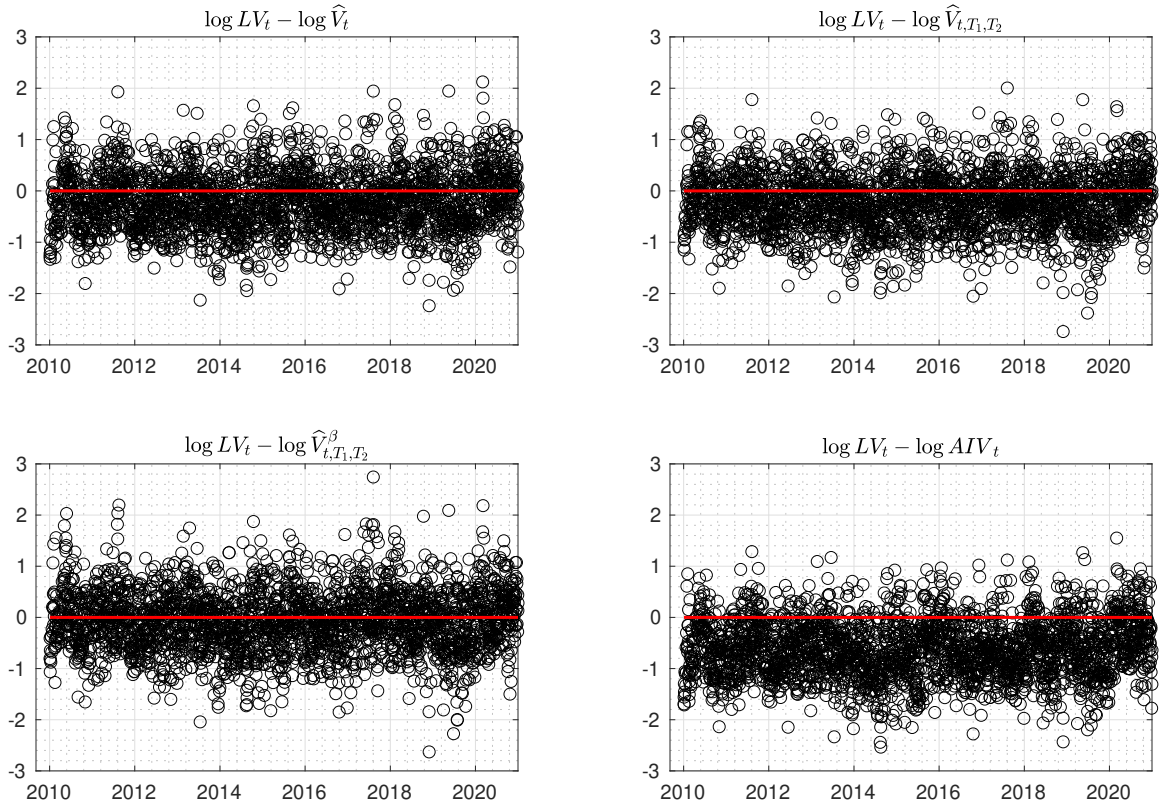

Figure 5: **Daily Difference between Option and Return Based Volatility Measures.** The red horizontal line on each plot is at zero. Notation as in Table 3.

The autocorrelations of the daily residual series, displayed in Figure 6, confirm the above observations. Mainly, mild positive autocorrelations are present in all series. They are strongest for the gap $\log(LV_t) - \log(AIV_t)$, for which autocorrelations are statistically significant even at lag 22 (time is measured in business days). Importantly, for the bias-corrected $\widehat{V}_{t,T_1,T_2}(\widehat{u}_{t,T_1})$ and $\widehat{V}_t(\underline{\widehat{u}}_{t,T_1}, \widehat{\beta}_{\mathbb{T}}(\underline{\widehat{u}}_{t,T_1}))$, the differences from LV_t exhibit least time series dependence, with autocorrelations past lag 10 being very small and in most cases insignificant. This is consistent with less bias

in these two option-based volatility estimates.

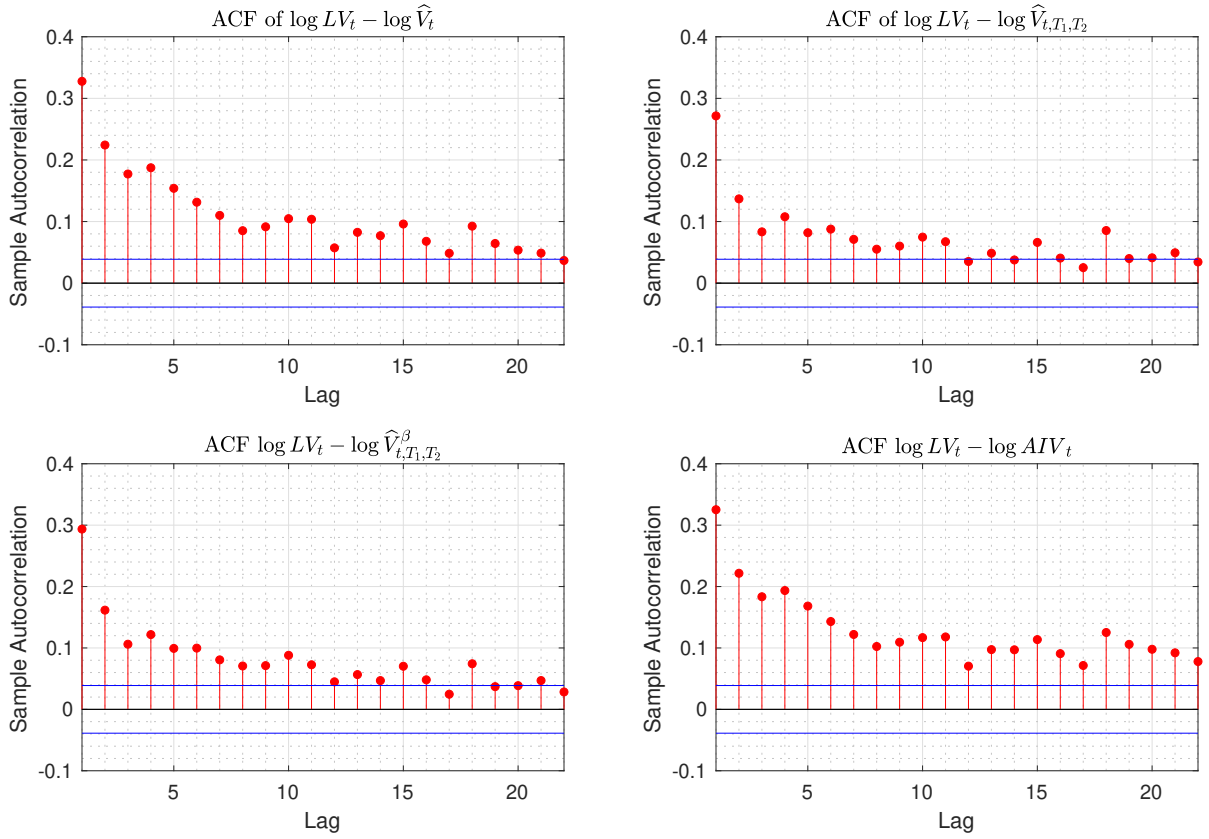

Figure 6: **Autocorrelation of Daily Gap between Option and Return Based Volatility Measures.** Blue horizontal lines on each plot correspond to standard errors under i.i.d. assumption for the series. Notation as in Table 3.

Of course, the above evidence can be also explained with option observation errors which have some time series dependence. To separate such an explanation for the autocorrelation patterns observed in Figure 6 from the presence of bias in the volatility proxies, we next study the dependence of the volatility gap on past volatility. More specifically, we run linear regressions of $\log(\widehat{OV}_t) - \log(LV_t)$ on a constant and past $\log(LV_t)$, where $\widehat{OV}_t$ is one of the option-based volatility estimators. The results are summarized in Figure 7. If bias in the volatility proxies is present and it depends on volatility, then one would expect that past volatility should be able to predict future gaps between the option and return based volatility proxies. On the other hand, since the return-based LV_t is adapted to $\mathcal{F}^{(0)}$, $\mathcal{F}^{(0)}$-conditional time series dependence in the observation errors will typically not generate predictability of that gap in the volatility proxies. To see this, suppose that the option observation errors are of the form $\epsilon_{t,T_l}(k_{t,T_l}(j)) = \zeta_{t,l}(k_{t,T_l}(j) - x_t)\overline{\epsilon}_{t,l,j}O_{t,T_l}(k_{t,T_l}(j))$, for

$j = 1, ..., N_{t,T_l}$, $l = 1, 2$ and where $\zeta_{t,l}(k_{t,T_l}(j) - x_t)$ is $\mathcal{F}^{(0)}$-adapted, $\overline{\epsilon}_{t,l,j}$ is $\mathcal{F}^{(1)}$-adapted and independent from $\mathcal{F}^{(0)}$. In this case, even if $\overline{\epsilon}_{t,l,j}$ has time series dependence (which implies time series dependence in the option observation errors), we still have $\mathbb{E}(\epsilon_{t,T_l}(k_{t,T_l}(j))|\mathcal{F}_s^{(0)}) = 0$, for $s < t$.

The reported results in Figure 7 reveal nontrivial differences between the volatility proxies in terms of the predictive regressions. Mainly, for AIV, the gap to LV is strongly predictable using past volatility. For all considered lags, the t-statistic for the coefficient in front of the past $\log(LV_t)$ is highly statistically significant and the R^2 of the predictive regression is nontrivial. Both these quantities, decrease monotonically when going from AIV_t to $\widehat{V}_t$, $\widehat{V}_{t,T_1,T_2}(\widehat{u}_{t,T_1})$ and $\widehat{V}_t(\underline{\widehat{u}}_{t,T_1}, \widehat{\beta}_{\mathbb{T}}(\underline{\widehat{u}}_{t,T_1}))$, in that order. In particular, for $\widehat{V}_t(\underline{\widehat{u}}_{t,T_1}, \widehat{\beta}_{\mathbb{T}}(\underline{\widehat{u}}_{t,T_1}))$, the highest R^2 is slightly above 2% only and past volatility has no statistical significance in the predictive regression past lag 4.

Overall, the empirical analysis illustrates that the de-biasing procedures introduced in the paper work well and the resulting option-based volatility estimator can accurately measure the spot volatility. This should allow for a more precise and robust study of the properties of the volatility process (e.g., roughness of the volatility path, presence of diffusion in its dynamics, etc), which relies on accurate spot volatility measurement, than previously possible. We leave such analysis for future work.

8 Conclusion

In this paper we consider methods for removing the bias in nonparametric spot volatility estimation from short-dated options that is due to time-varying volatility and presence of jumps in the underlying asset price. The bias-reduction techniques build on characteristic function based volatility estimates formed from short-dated options with different tenors. We show that by suitably differencing the volatility estimates from options with different times to maturity, we can reduce asymptotically the bias in volatility estimation due to the time-variation in volatility and the presence of jumps. Applying additional nonlinear least squares techniques in characteristic exponent space leads to further reduction in bias in spot volatility estimation that is due to price jumps. The theoretical results are illustrated on simulated and real data.

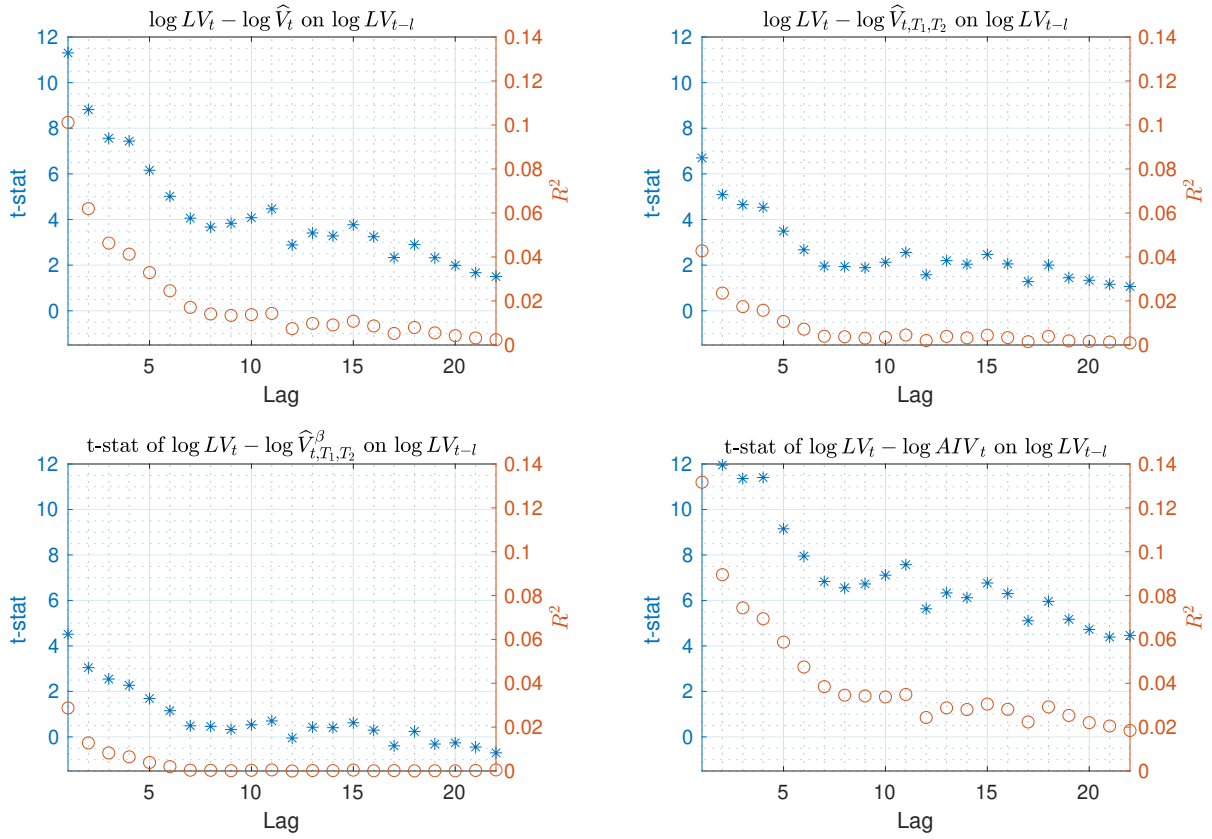

Figure 7: **Gap between Option and Return Based Volatility Measures and Past Volatility**. The plots display t-statistics with $*$ (left axis) and R^2 with o (right axis) from a regression of $\log(LV_t) - \log(\widehat{OV}_t)$ on a constant and lagged $\log(LV_t)$, for $\widehat{OV}_t$ being one of the option-based volatility proxies. Standard errors are based on Newey-West HAC long-run covariance estimator with lag length of 22. Notation as in Table 3.

9 Assumptions and Proofs

9.1 Assumptions

The process x is defined on a filtered probability space $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{Q}\right)$. The volatility process obeys the following dynamics

$$\sigma_t = \sigma_0 + \int_0^t b_s ds + \int_0^t \eta_s dW_s + \int_0^t \widetilde{\eta}_s d\widetilde{W}_s + \sum_{s\leq t} \Delta\sigma_s, \tag{61}$$

where $\widetilde{W}$ is a Brownian motion independent from W. The dynamics of x, given in (1) and (61), is under the probability measure $\mathbb{Q}$, and the jumps in x and σ are given by

$$\sum_{s\leq t} \Delta x_s = \int_0^t \int_{\mathbb{R}} \delta_x(s,z)\mu(ds,dz), \quad \sum_{s\leq t} \Delta\sigma_s = \int_0^t \int_{\mathbb{R}} \delta_\sigma(s,z)\mu(ds,dz), \tag{62}$$

where μ is a Poisson measure on $\mathbb{R}_+ \times \mathbb{R}$ with compensator $\lambda(ds, dz) = ds \otimes dz$, and $\delta_x : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}$ and $\delta_y : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}$ are two predictable functions. Just like x and σ, the processes α, η and $\widetilde{\eta}$ will be general Itô semimartingales with the following dynamics

$$\begin{aligned} \alpha_t = \alpha_0 &+ \int_0^t b_s^\alpha ds + \int_0^t \sigma_s^\alpha dW_s + \int_0^t \widetilde{\sigma}_s^\alpha d\widetilde{W}_s + \int_0^t \overline{\sigma}_s^\alpha d\overline{W}_s \\ &+ \int_0^t \int_{\mathbb{R}} \delta_\alpha(s,z)\mu(ds,dz), \end{aligned} \tag{63}$$

$$\begin{aligned} \eta_t = \eta_0 &+ \int_0^t b_s^\eta ds + \int_0^t \sigma_s^\eta dW_s + \int_0^t \widetilde{\sigma}_s^\eta d\widetilde{W}_s + \int_0^t \overline{\sigma}_s^\eta d\overline{W}_s \\ &+ \int_0^t \int_{\mathbb{R}} \delta_\eta(s,z)\mu(ds,dz), \end{aligned} \tag{64}$$

$$\begin{aligned} \widetilde{\eta}_t = \widetilde{\eta}_0 &+ \int_0^t b_s^{\widetilde{\eta}} ds + \int_0^t \sigma_s^{\widetilde{\eta}} dW_s + \int_0^t \widetilde{\sigma}_s^{\widetilde{\eta}} d\widetilde{W}_s + \int_0^t \overline{\sigma}_s^{\widetilde{\eta}} d\overline{W}_s \\ &+ \int_0^t \int_{\mathbb{R}} \delta_{\widetilde{\eta}}(s,z)\mu(ds,dz), \end{aligned} \tag{65}$$

where b^α, b^η, $b^{\widetilde{\eta}}$, σ^α, $\widetilde{\sigma}^\alpha$, $\overline{\sigma}^\alpha$, σ^η, $\widetilde{\sigma}^\eta$, $\overline{\sigma}^\eta$, $\sigma^{\widetilde{\eta}}$, $\widetilde{\sigma}^{\widetilde{\eta}}$ and $\overline{\sigma}^{\widetilde{\eta}}$ are processes with càdlàg paths, $\overline{\sigma}_t^\alpha$, $\overline{\sigma}_t^\eta$ and $\overline{\sigma}_t^{\widetilde{\eta}}$ are 1×3 vectors and the rest of the above processes are one-dimensional, $\overline{W}$ is a 3-dimensional Brownian motion orthogonal to W and $\widetilde{W}$, and $\delta_\alpha : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}$, $\delta_\eta : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}$ and $\delta_{\widetilde{\eta}} : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}$ are predictable functions.

We now state our assumptions for the dynamics of x which we need for the asymptotic analysis. Since later we will use different probability measures, to avoid confusion in the statements that follow, we will denote with $\mathbb{E}^{\mathbb{Q}}$ the expectation under $\mathbb{Q}$ and similarly $\mathbb{E}_t^{\mathbb{Q}}$ will be the $\mathcal{F}_t$-conditional expectation under $\mathbb{Q}$.

B1-r. *There exist $\mathcal{F}_t$-adapted random variables $C_t > 0$ and $\overline{t} > t$ such that for $s \in [t, \overline{t}]$:*

$$\mathbb{E}_t^{\mathbb{Q}} |z_s|^2 < C_t, \tag{66}$$

for z being each one of the processes b, b^α, b^η, $b^{\widetilde{\eta}}$, σ^α, $\widetilde{\sigma}^\alpha$, $\overline{\sigma}^\alpha$, σ^η, $\widetilde{\sigma}^\eta$, $\overline{\sigma}^\eta$, $\sigma^{\widetilde{\eta}}$, $\widetilde{\sigma}^{\widetilde{\eta}}$ and $\overline{\sigma}^{\widetilde{\eta}}$. In addition, for some $r \in [0,1]$, we have

$$\mathbb{E}_t^{\mathbb{Q}} \left(\int_{\mathbb{R}} |\delta_x(s,z)|^r dz \right) + \mathbb{E}_t^{\mathbb{Q}} \left(\int_{\mathbb{R}} (|\delta_\sigma(s,z)| \vee |\delta_\sigma(s,z)|^4) dz \right)^2 < C_t, \tag{67}$$

and further for some $\iota > 0$

$$\begin{aligned} &\mathbb{E}_t^{\mathbb{Q}} \left(\int_{\mathbb{R}} (|\delta_\eta(s,z)| \vee |\delta_\eta(s,z)|^{1+\iota}) dz \right)^2 \\ &\quad + \mathbb{E}_t^{\mathbb{Q}} \left(\int_{\mathbb{R}} (|\delta_{\widetilde{\eta}}(s,z)| \vee |\delta_{\widetilde{\eta}}(s,z)|^{1+\iota}) dz \right)^2 < C_t. \end{aligned} \tag{68}$$

B2-r. *There exist $\mathcal{F}_t$-adapted random variables $C_t > 0$ and $\bar{t} > t$ such that for $u, s \in [t, \bar{t}]$:*

$$\mathbb{E}_t^{\mathbb{Q}} |z_s - z_u|^2 \leq C_t |s - u|, \tag{69}$$

for z being each one of the processes: b, σ^α, $\widetilde{\sigma}^\alpha$, $\overline{\sigma}^\alpha$, σ^η, $\widetilde{\sigma}^\eta$, $\overline{\sigma}^\eta$, $\sigma^{\widetilde{\eta}}$, $\widetilde{\sigma}^{\widetilde{\eta}}$ and $\overline{\sigma}^{\widetilde{\eta}}$. Furthermore, for some $r \in [0, 1]$ and $\iota > 0$, we have

$$\mathbb{E}_t^{\mathbb{Q}} \left(\int_{\mathbb{R}} |\delta_x(s, z) - \delta_x(u, z)|^r ds dz \right) \leq C_t \sqrt{|s - u|}, \tag{70}$$

$$\mathbb{E}_t^{\mathbb{Q}} \left(\int_{\mathbb{R}} \left(|\delta_\sigma(s, z) - \delta_\sigma(u, z)| \vee |\delta_\sigma(s, z) - \delta_\sigma(u, z)|^{1+\iota} \right) ds dz \right)^2 \leq C_t |s - u|. \tag{71}$$

B3-r. *We have*

$$1_{\{x \neq 0\}} \nu_t(dx, dy) = \overline{\nu}_t(x, y) dx dy + 1_{\{y = g(x),\, x \neq 0\}} \widetilde{\nu}_t(x) dx, \tag{72}$$

where $\overline{\nu}_t$ is a predictable $\mathbb{R}_+$-valued function on $\mathbb{R}^2$, $\widetilde{\nu}_t(x)$ is a predictable $\mathbb{R}_+$-valued function on $\mathbb{R}$, and g is $\mathbb{R}$-valued function on $\mathbb{R}$. We have the following conditions

$$\int_{\mathbb{R}} \int_{\mathbb{R}} \left(|y| \, |\overline{\nu}_t(x + h, y) - \overline{\nu}_t(x, y)| \right) dx dy \leq C_t |h|^{1-r}, \tag{73}$$

$$\int_{\mathbb{R}} |g(x + h) \widetilde{\nu}_t(x + h) - g(x) \widetilde{\nu}_t(x)| \, dx \leq C_t |h|^{1-r}, \tag{74}$$

$$|g(x + h) - g(x)| \leq \overline{g}(x) h^{1-r}, \text{ for some } \overline{g}(x) \text{ with } \int_{\mathbb{R}} |\overline{g}(x) g(x) \widetilde{\nu}_t(x)| dx \leq C_t, \tag{75}$$

for some $\mathcal{F}_t$-adapted random variable C_t and some $r \in [0, 1]$.

C1. *There exist $\mathcal{F}_t^{(0)}$-adapted random variables C_t and $\bar{t} > t$ such that for $s \in [t, \bar{t}]$:*

$$\mathbb{E}_t^{\mathbb{Q}} |\alpha_s|^4 + \mathbb{E}_0^{\mathbb{Q}} |\sigma_s|^6 + \mathbb{E}_0^{\mathbb{Q}} (e^{4|x_s|}) + \mathbb{E}_t^{\mathbb{Q}} \left(\int_{\mathbb{R}^2} (e^{3|x|} - 1) \nu_s(dx, dy) \right)^4 < C_t. \tag{76}$$

C2. *The log-strike grids $\{k_{t,T_l}(j)\}_{j=1}^{N_l}$, for $l = 1, 2$, are $\mathcal{F}^{(t)}$-adapted and we have*

$$c_t \Delta \leq k_{l,j} - k_{l,j-1} \leq C_t \Delta, \quad l = 1, 2, \text{ as } \Delta \downarrow 0, \tag{77}$$

where Δ is a deterministic sequence, and $c_t > 0$ and $C_t < \infty$ are $\mathcal{F}^{(0)}$-adapted random variables. In addition, for some arbitrary small $\zeta > 0$:

$$\sup_{j : |k_{t,T_l}(j) - x_t| < \zeta} \left| \frac{k_{t,T_l}(j) - k_{t,T_l}(j-1)}{\Delta} - \psi_{t,l}(k_{t,T_l}(j-1) - x_t) \right| \xrightarrow{\mathbb{P}} 0, \quad l = 1, 2, \text{ as } \Delta \downarrow 0, \tag{78}$$

where $\psi_{t,T_l}(k)$ are $\mathcal{F}^{(0)}$-adapted functions which are continuous in k at 0.

C3. *We have* $\epsilon_{t,T_l}(k_{t,T_l}(j)) = \zeta_{t,l}(k_{t,T_l}(j) - x_t)\overline{\epsilon}_{t,l,j}O_{t,T_l}(k_{t,T_l}(j))$ *for* $l = 1, 2$*, where for* k *in a neighborhood of zero, we have* $|\zeta_{t,l}(k) - \xi_{t,l}(0)| \leq C_t|k|^{\iota}$*, for some* $\iota > 0$ *and* $C_t < \infty$ *being an* $\mathcal{F}^{(0)}$*-adapted random variable. The two sequences* $\{\overline{\epsilon}_{t,1,j}\}_{j=1}^{N_1}$ *and* $\{\overline{\epsilon}_{t,2,j}\}_{j=1}^{N_2}$ *are defined on* $\mathcal{F}^{(1)}$*, are i.i.d. and independent of each other and of* $\mathcal{F}^{(0)}$*. We further have* $\mathbb{E}(\overline{\epsilon}_{t,l,j}|\mathcal{F}^{(0)}) = 0$*,* $\mathbb{E}((\overline{\epsilon}_{t,l,j})^2|\mathcal{F}^{(0)}) = 1$ *and* $\mathbb{E}(|\overline{\epsilon}_{t,l,j}|^{\kappa}|\mathcal{F}^{(0)}) < \infty$*, for some* $\kappa \geq 4$ *and* $l = 1, 2$*.*

9.2 Proof of Theorem 1

The proof of the theorem follows from Theorem 1 in Todorov (2021), provided we can show that under assumption we have

$$\int_0^1 \int_{\mathbb{R}^2} e^{\frac{iux}{\sqrt{T}}} \left(e^{-u^2 y \sigma_t s - \frac{u^2}{2} y^2 s} - 1\right) 1_{\{x \neq 0\}} \nu_t(dx, dy) ds = O_p\left(T^{\frac{1-r}{2}}\right). \tag{79}$$

We use the following notation for some $\sigma_t \neq 0$ and $u \neq 0$

$$h_t(u, y) = \begin{cases} 2\frac{\int_0^1 \left(e^{-u^2 y \sigma_t s - \frac{u^2}{2} y^2 s} - 1\right) ds}{u^2 \sigma_t y}, \text{ if } y \neq 0, \\ 1, \text{ if } y = 0. \end{cases} \tag{80}$$

It is easy to see that, for a given $\sigma_t \neq 0$ and $u \neq 0$, $h_t(u, y)$ is bounded and differentiable in y with $\nabla_y h_t(u, y)$ bounded. The boundedness, together with (73), implies

$$\int_{\mathbb{R}} \int_{\mathbb{R}} \left(|h_t(u, y)||y| \, |\overline{\nu}_t(x + h, y) - \overline{\nu}_t(x, y)|\right) dxdy \leq C_t(u)|h|^{1-r}. \tag{81}$$

Next, using the differentiability of $h_t(u, y)$ in y and the fact that $\nabla_y h_t(u, y)$ is bounded, for given $\sigma_t \neq 0$ and $u \neq 0$, as well as (75), we have

$$\int_{\mathbb{R}} |(h_t(u, g(x + h)) - h_t(u, g(x))g(x)\widetilde{\nu}_t(x)| \, dx \leq C_t(u)|h|^{1-r}. \tag{82}$$

Using the boundedness of $h_t(u, y)$ for given $\sigma_t \neq 0$ and $u \neq 0$ and (74), we have

$$\int_{\mathbb{R}} |h_t(u, g(x + h))(g(x + h)\widetilde{\nu}_t(x + h) - g(x)\widetilde{\nu}_t(x))| \, dx \leq C_t(u)|h|^{1-r}. \tag{83}$$

The above two bounds imply altogether

$$\int_{\mathbb{R}} |h_t(u, g(x + h))g(x + h)\widetilde{\nu}_t(x + h) - h_t(u, g(x))g(x)\widetilde{\nu}_t(x)| \, dx \leq C_t(u)|h|^{1-r}. \tag{84}$$

The claim to be proved in (79) then follows from the fact that for an arbitrary function $\zeta : \mathbb{R}^2 \to \mathbb{R}$, we have for a fixed $v \in \mathbb{R}_+$

$$\int_{\mathbb{R}} |\zeta(x + h, v) - \zeta(x, v)|dx \leq C(v)|h|^{1-r} \implies \int_{\mathbb{R}} e^{iuvx}\zeta(x, v)dx = O(|u|^{-(1-r)}), \text{ as } |u| \to \infty, \tag{85}$$

for $C(v)$ being some positive and finite-valued function of v. To establish the above result, we note first that by a change of variable of integration, we have

$$\int_{\mathbb{R}} e^{iuvx}\zeta\left(x+\frac{\pi}{uv},v\right)dx = e^{-i\pi}\int_{\mathbb{R}} e^{iuvx}\zeta(x,v)dx = -\int_{\mathbb{R}} e^{iuvx}\zeta(x,v)dx, \text{ for } u,v\neq 0. \tag{86}$$

Therefore,

$$\int_{\mathbb{R}} e^{iux}\left(\zeta\left(x+\frac{\pi}{uv},v\right)-\zeta(x,v)\right)dx = -2\int_{\mathbb{R}} e^{iuvx}\zeta(x,v)dx. \tag{87}$$

Using the Hölder condition for the function ζ, we have

$$\left|\int_{\mathbb{R}} e^{iuvx}\zeta(x,v)dx\right| \leq \int_{\mathbb{R}}\left|\zeta\left(x+\frac{\pi}{uv},v\right)-\zeta(x,v)\right|dx \leq C(v)|u|^{-(1-r)}, \tag{88}$$

and hence the result to be proved.

9.3 Proof of Theorem 2

Part (a) follows by application of Riemann-Lebesgue lemma, see e.g., Proposition 2.2.17 in Grafakos (2008).

We now show part (b). First, if we denote

$$\nu_t^{\mathrm{st}}(x) = \frac{c_t^- 1(x<0)+c_t^+ 1(x>0)}{|x|^{1+\beta_t}}, \tag{89}$$

then, by using Lemma 14.11 in Sato (1999), we have

$$\int_{\mathbb{R}}(1-\cos(ux))\nu_t^{\mathrm{st}}(x)dx = (c_t^-+c_t^+)\chi(\beta_t)u^{\beta_t}, \tag{90}$$

where

$$\chi(\beta) = \begin{cases} \frac{1}{\beta}\Gamma(1-\beta)\cos\left(\frac{\beta\pi}{2}\right), & \text{if } \beta\neq 1, \\ \frac{\pi}{2}, & \text{if } \beta = 1. \end{cases} \tag{91}$$

Second, using our assumption for the residual measure $\widetilde{\nu}_t$, we have

$$\int_{\mathbb{R}}(1-\cos(ux))|\widetilde{\nu}_t(x)|dx \leq 2^{1-\widetilde{\beta}_t}|u|^{\widetilde{\beta}_t}\int_{\mathbb{R}}|x|^{\widetilde{\beta}_t}|\widetilde{\nu}_t(x)|dx. \tag{92}$$

Combining these two results, we can write

$$B_{t,T}(u) = 2T^{1-\beta_t/2}(c_t^-+c_t^+)\chi(\beta_t)u^{\beta_t-2}+o_p(T^{1-\beta_t/2}), \quad \text{as } T\downarrow 0. \tag{93}$$

From here, part (b) of the theorem follows directly.

9.4 Proof of Theorem 3

We begin with an auxiliary lemma.

Lemma 1. *For $\zeta > 1$, we have*

$$\int_0^\infty \frac{e^{iux} - e^{iu\zeta x}}{x^{\beta+1}} dx = u^\beta \Gamma(1-\beta) e^{-\frac{i\pi\beta}{2}} g(\zeta, \beta), \quad \beta \in (-1, 1) \text{ and } u \in \mathbb{R}_+. \tag{94}$$

Proof of Lemma 1. For the case $\beta \in (-1, 0)$, direct calculation yields (note that the integral below is not absolutely convergent in this case):

$$\int_0^\infty \frac{e^{iux}}{x^{\beta+1}} dx = -u^\beta \frac{\Gamma(1-\beta)}{\beta} e^{-\frac{i\pi\beta}{2}}. \tag{95}$$

For the case $\beta \in (0, 1)$, we have from Lemma 14.11 in Sato (1999),

$$\int_0^\infty \frac{(e^{iux} - 1)}{x^{\beta+1}} dx = -u^\beta \frac{\Gamma(1-\beta)}{\beta} e^{-\frac{i\pi\beta}{2}}. \tag{96}$$

We are therefore left with the case $\beta = 0$. First, for the sine integral, we have $\int_0^\infty \frac{\sin(x)}{x} dx = \frac{\pi}{2}$, see e.g., Section 5.2 in Abramowitz and Stegun (1970). Therefore, we need to compute only $\int_0^\infty \left(\frac{\cos(x) - \cos(\zeta x)}{x} \right) dx$. For the cosine integral, we can write

$$\begin{aligned} \int_1^\infty \frac{\cos(z)}{z} dz &= -\gamma - \log(1) + \int_0^1 \frac{1 - \cos(z)}{z} dz, \\ \int_1^\infty \frac{\cos(\zeta z)}{z} dz &= -\gamma - \log(\zeta) + \int_0^1 \frac{1 - \cos(\zeta z)}{z} dz, \end{aligned} \tag{97}$$

where γ is the Euler–Mascheroni constant, see e.g., Section 5.2 in Abramowitz and Stegun (1970). Therefore,

$$\int_1^\infty \left(\frac{\cos(z) - \cos(\zeta z)}{z} \right) dz = \log(\zeta) - \int_0^1 \frac{\cos(z) - \cos(\zeta z)}{z} dz. \tag{98}$$

From here, the result in the case $\beta = 0$ follows. □

Integration by parts and assumption A yields

$$\int_{\mathbb{R}} e^{iux} \overline{\nu}(x) dx = -\int_{\mathbb{R}} \frac{e^{iux}}{iu} \overline{\nu}'(x) dx. \tag{99}$$

Therefore

$$\int_{\mathbb{R}} e^{iux/\sqrt{T}} \overline{\nu}(x) dx = O_p^{\text{lu}}(\sqrt{T}), \tag{100}$$

for u taking values outside of zero. From here, the result of the theorem follows by application of Lemma 1. □

9.5 Proof of Theorem 4

In order to simplify notation, we present the proof in the case when the set $\mathbb{T}$ consists of t only. We also use the following simplifying notation

$$\psi_{t,T_1} = \frac{2\tau^{1+\beta/2}}{\tau-1}(c_t^+ + c_t^-)T_1^{1-\beta/2}\Gamma(1-\beta)\cos\left(\frac{\pi\beta}{2}\right)g(\sqrt{\tau},\beta), \tag{101}$$

where g is the function appearing in the statement of Theorem 3.

We make the following decomposition

$$V_t(\underline{u};x) = V_t + \psi_{t,T_1}\frac{\sum_{i=1}^K u_i^{2x-4}\sum_{i=1}^K u_i^{\beta-2} - \sum_{i=1}^K u_i^{x-2}\sum_{i=1}^K u_i^{x-2}u_i^{\beta-2}}{K\sum_{i=1}^K u_i^{2x-4} - \left(\sum_{i=1}^K u_i^{x-2}\right)^2} + \epsilon_t(x,V), \tag{102}$$

$$\psi_t(\underline{u};x) = \psi_{t,T_1}\frac{K\sum_{i=1}^K u_i^{x-2}u_i^{\beta-2} - \sum_{i=1}^K u_i^{x-2}\sum_{i=1}^K u_i^{\beta-2}}{K\sum_{i=1}^K u_i^{2x-4} - \left(\sum_{i=1}^K u_i^{x-2}\right)^2} + \epsilon_t(x,\psi), \tag{103}$$

where $\epsilon_t(x,V)$ and $\epsilon_t(x,\psi)$ satisfy

$$\sup_{x\in[-1,1]}(|\epsilon_t(x,V)| + |\epsilon_t(x,\psi)|) = O_p\left(T_1^{\left(\frac{3}{2}-\frac{\beta}{2}-\iota\right)\wedge\frac{3}{2}}|\log(T_1)|\right), \tag{104}$$

which follows from Theorem 3 and the fact that $\inf_{x\in[-1,1]}\left(K\sum_{i=1}^K u_i^{2x-4} - \left(\sum_{i=1}^K u_i^{x-2}\right)^2\right) > 0$ (using strict inequality in means and the fact that $K\sum_{i=1}^K u_i^{2x-4} - \left(\sum_{i=1}^K u_i^{x-2}\right)^2$ is a continuous function of x). We next introduce the following notation:

$$\begin{aligned} Z_{i,t,T_1}(x) = &-\psi_{t,T_1}\frac{\sum_{i=1}^K u_i^{2x-4}\sum_{i=1}^K u_i^{\beta-2} - \sum_{i=1}^K u_i^{x-2}\sum_{i=1}^K u_i^{x-2}u_i^{\beta-2}}{K\sum_{i=1}^K u_i^{2x-4} - \left(\sum_{i=1}^K u_i^{x-2}\right)^2} \\ &+ \psi_{t,T_1}u_i^{\beta-2} - \psi_{t,T_1}u_i^{x-2}\frac{K\sum_{i=1}^K u_i^{x-2}u_i^{\beta-2} - \sum_{i=1}^K u_i^{x-2}\sum_{i=1}^K u_i^{\beta-2}}{K\sum_{i=1}^K u_i^{2x-4} - \left(\sum_{i=1}^K u_i^{x-2}\right)^2}, \end{aligned} \tag{105}$$

$$\epsilon_{i,t,T_1}(x) = V_{t,T_1,T_2}(u_i) - V_t - \psi_{t,T_1}u_i^{\beta-2} - \epsilon_t(x,V) - \epsilon_t(x,\psi)u_i^{x-2}. \tag{106}$$

With this notation, we have

$$\beta_t(\underline{u}) = \operatorname*{argmin}_{x\in[-1,1]}\sum_{i=1}^K\left[Z_{i,t,T_1}(x)^2 + 2Z_{i,t,T_1}(x)\epsilon_{i,t,T_1}(x) + \epsilon_{i,t,T_1}(x)^2\right]. \tag{107}$$

For $\sum_{i=1}^K Z_{i,t,T_1}(x)^2$, we have

$$\sum_{i=1}^K Z_{i,t,T_1}(x)^2 \geq \psi_{t,T_1}^2\sum_{i=1}^K\left(u_i^{\beta-2} - f_1(x) - f_2(x)u_i^{x-2}\right)^2, \tag{108}$$

where $f_1(x)$ and $f_2(x)$ are the intercept and the slope coefficient in a linear regression of $u_i^{\beta-2}$ on a constant and u_i^{x-2}, for $i = 1, ..., K$. It is easy to check that

$$Z_{i,t,T_1}(\beta) = 0, \quad i = 1, ..., K. \tag{109}$$

For $x \neq \beta$, using the properties of the vector $\underline{u}$, we have

$$\frac{u_i^{\beta-2} - u_{i+1}^{\beta-2}}{u_{i+1}^{\beta-2} - u_{i+2}^{\beta-2}} \neq \frac{u_i^{x-2} - u_{i+1}^{x-2}}{u_{i+1}^{x-2} - u_{i+2}^{x-2}}, \ i = 1, ..., K-2. \tag{110}$$

Therefore, $\sum_{i=1}^{K} \left(u_i^{\beta-2} - f_1(x) - f_2(x)u_i^{x-2}\right)^2$ is a strictly positive number for $x \neq \beta$,

$$x \neq \beta \implies \frac{1}{\psi_{t,T_1}^2} \sum_{i=1}^{K} Z_{i,t,T_1}(x)^2 > 0. \tag{111}$$

From here, since $\sum_{i=1}^{K} Z_{i,t,T_1}(x)^2$ is a continuous function of x, for every $\epsilon > 0$, there exists arbitrary small $\delta > 0$, such that

$$|x - \beta| > \epsilon \implies \sum_{i=1}^{K} Z_{i,t,T_1}(x)^2 > \delta\psi_{t,T_1}^2. \tag{112}$$

Next, from Theorem 3 and the result above for $\epsilon_t(x, V)$ and $\epsilon_t(x, \psi)$, we have

$$\begin{aligned} \sup_{x\in[-1,1]} |Z_{i,t,T_1}(x)\epsilon_{i,t,T_1}| &\leq M\psi_{t,T_1} \sup_{x\in[-1,1]} |\epsilon_{i,t,T_1}| \\ &= O_p\left(T_1^{\left(\frac{3}{2}-\frac{\beta}{2}-\iota\right)\wedge\frac{3}{2}} |\log(T_1)|\right), \end{aligned} \tag{113}$$

for some positive constant $M > 0$. Now, since $\beta > -1$, we have that $T_1^{\left(\frac{3}{2}-\frac{\beta}{2}-\iota\right)\wedge\frac{3}{2}} |\log(T_1)| \to 0$. Finally, using the bound derived above for $\epsilon_t(x, V)$ and $\epsilon_t(x, \psi)$, we have

$$\sum_{i=1}^{K} \epsilon_{i,t,T_1}(x)^2 = O_p\left(T_1^{(3-\beta-2\iota)\wedge 3} |\log(T_1)|^2\right). \tag{114}$$

Combining the above three results, we have the consistency of $\beta_t(\underline{u})$.

We turn next to deriving the order of magnitude of $\beta_t(\underline{u})$. Due to the established consistency and since $\beta \in (-1, 1)$, with probability approaching 1, $\beta_t(\underline{u})$ solves the first-order condition of the optimization problem. A first-order Taylor expansion then leads to

$$H_{t,T_1}(\widetilde{x})(\beta_t(\underline{u}) - \beta) = S_{t,T_1}(\beta), \tag{115}$$

where $H_{t,T_1}(x)$ and $S_{t,T_1}(x)$ are some continuous functions of x, and $\widetilde{x}$ is a number between x and β. We have

$$H_{t,T_1}(\widetilde{x}) \xrightarrow{\mathbb{P}} \sum_{i=1}^{K} R(u_i, \underline{u}, \beta)^2, \tag{116}$$

where

$$R(u,\underline{u},\beta) = \frac{\sum_{i=1}^K u_i^{2\beta-4}\log(u_i)\sum_{i=1}^K u_i^{\beta-2} - \sum_{i=1}^K u_i^{2\beta-4}\sum_{i=1}^K u_i^{\beta-2}\log(u_i)}{K\sum_{i=1}^K u_i^{2\beta-4} - \left(\sum_{i=1}^K u_i^{\beta-2}\right)^2} + u^{\beta-2}\log(u) - u^{\beta-2}\frac{K\sum_{i=1}^K u_i^{2\beta-4}\log(u_i) - \sum_{i=1}^K u_i^{\beta-2}\sum_{i=1}^K u_i^{\beta-2}\log(u_i)}{K\sum_{i=1}^K u_i^{2\beta-4} - \left(\sum_{i=1}^K u_i^{\beta-2}\right)^2}, \tag{117}$$

and

$$S_{t,T_1}(\beta) = O_p\left(T_1^{\left(\frac{3}{2}-\frac{\beta}{2}-\iota\right)\wedge\frac{3}{2}}|\log(T_1)|\right). \tag{118}$$

Now, we have

$$\sum_{i=1}^K R(u_i,\underline{u},\beta)^2 \geq \sum_{i=1}^K \left(u_i^{\beta-2}\log(u_i) - a_0(\underline{u}) - a_1(\underline{u})u_i^{\beta-2}\right)^2, \tag{119}$$

where $a_0(\underline{u})$ and $a_1(\underline{u})$ are the intercept and slope, respectively, in a linear regression of $u_i^{\beta-2}\log(u_i)$ on a constant and $u_i^{\beta-2}$, for $i = 1, ..., K$. Given the properties of $\underline{u}$, we have

$$\frac{u_i^{\beta-2}\log(u_i) - u_{i+1}^{\beta-2}\log(u_{i+1})}{u_i^{\beta-2} - u_{i+1}^{\beta-2}} \neq \frac{u_{i+1}^{\beta-2}\log(u_{i+1}) - u_{i+2}^{\beta-2}\log(u_{i+2})}{u_{i+1}^{\beta-2} - u_{i+2}^{\beta-2}},\ i = 1, ..., K-2. \tag{120}$$

Therefore,

$$\beta \in [-1,1] \implies \sum_{i=1}^K R(u_i,\underline{u},\beta)^2 > 0. \tag{121}$$

Combining the above two results, we have the claim in the theorem.

9.6 Proof of Theorem 5

In the case when assumption A holds, the result of the theorem follows by making use of the decomposition of $V_t(\underline{u};\beta_{\mathbb{T}}(\underline{u}))$ in (102), the bound in (104) as well as the result of Theorem 4.

In the case when $\int_{\mathbb{R}}\nu_t(x)dx < \infty$ and $\int_{\mathbb{R}}|\nu_t'(x)|dx < \infty$, the result follows from part (a) of Theorem 2.

9.7 Proof of Theorem 6

The theorem follows from Theorem 2 in Todorov (2021).

9.8 Proof of Corollary 1

The results in the corollary concerning $\widehat{V}_{t,T_1,T_2}(u)$ follow directly from Theorems 2 and 6. The proof regarding $\widehat{V}_t(\underline{u};\widehat{\beta}_{\mathbb{T}}(\underline{u}))$ parallels the proof of Theorems 4 and 5. In particular, in the decompositions

in (102) and (103), we have now additional terms due to the option observation error, denoted with $\widehat{\epsilon}_t(x,V)$ and $\widehat{\epsilon}_t(x,\psi)$, respectively. From Theorem 6, we have for them

$$\sup_{x\in[-1,1]}(|\widehat{\epsilon}_t(x,V)|+|\widehat{\epsilon}_t(x,\psi)|)=O_p\left(\frac{\sqrt{\Delta}}{T_1^{1/4}}\right). \tag{122}$$

Similar additional term appears in $S_{t,T_1}(\beta)$ (with the same error bound as above). From here, the proof of the results in the corollary regarding $\widehat{V}_t(\underline{u};\widehat{\beta}_{\mathbb{T}}(\underline{u}))$ proceed exactly as the proof of Theorems 2 and 6.

Acknowledgements

We would like to thank Serena Ng (the Editor), the Associate Editor and two anonymous referees for many useful comments and suggestions. We also thank Nicola Fusari for providing the option pricing codes used in the Monte Carlo section.

References

Abramowitz, M. and I. A. Stegun (1970). *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, Volume 55. US Government printing office.

Andersen, T. G., D. Dobrev, and E. Schaumburg (2012). Jump-robust volatility estimation using nearest neighbor truncation. *Journal of Econometrics 169*(1), 75–93.

Andersen, T. G., N. Fusari, and V. Todorov (2015). The Risk Premia Embedded in Index Options. *Journal of Financial Economics 117*, 558–584.

Andersen, T. G., N. Fusari, and V. Todorov (2017). The Pricing of Short-Term Market Risk: Evidence from Weekly Options. *Journal of Finance 72*, 1335–1386.

Andersen, T. G., N. Fusari, V. Todorov, and R. T. Varneskov (2021). Spatial dependence in option observation errors. *Econometric Theory 37*(2), 205–247.

Bandi, F., N. Fusari, and R. Reno (2021). Structural stochastic volatility. Working paper.

Bandi, F. M. and P. C. Phillips (2003). Fully nonparametric estimation of scalar diffusion models. *Econometrica 71*(1), 241–283.

Barndorff-Nielsen, O. and N. Shephard (2004). Power and Bipower Variation with Stochastic Volatility and Jumps. *Journal of Financial Econometrics 2*, 1–37.

Barndorff-Nielsen, O. and N. Shephard (2006). Econometrics of Testing for Jumps in Financial Economics using Bipower Variation. *Journal of Financial Econometrics 4*, 1–30.

Barndorff-Nielsen, O. and A. Shiryaev (2010). *Change of Time and Change of Measure*. World Scientific.

Bates, D. S. (2000). Post-'87 crash fears in the s&p 500 futures option market. *Journal of econometrics 94*(1-2), 181–238.

Bibinger, M., N. Hautsch, P. Malec, and M. Reiss (2019). Estimating the spot covariation of asset prices—statistical theory and empirical evidence. *Journal of Business & Economic Statistics 37*(3), 419–435.

Black, F. and M. Scholes (1973). The Pricing of Options and Corporate Liabilities. *Journal of Political Economy 81*, 637–654.

Bollerslev, T. and V. Todorov (2014). Time-varying jump tails. *Journal of Econometrics 183*(2), 168–180.

Breeden, D. T. and R. H. Litzenberger (1978). Prices of state-contingent claims implicit in option prices. *Journal of business 51*, 621–651.

Carr, P., H. Geman, D. Madan, and M. Yor (2002). The Fine Structure of Asset Returns: An Empirical Investigation. *Journal of Business 75*, 305–332.

Carr, P., H. Geman, D. Madan, and M. Yor (2003). Stochastic Volatility for Lévy Processes. *Mathematical Finance 13*, 345–382.

Carr, P. and D. Madan (2001). Optimal Positioning in Derivative Securities. *Quantitiative Finance 1*, 19–37.

Corsi, F., D. Pirino, and R. Reno (2010). Threshold bipower variation and the impact of jumps on volatility forecasting. *Journal of Econometrics 159*(2), 276–288.

Durrleman, V. (2008). Convergence of at-the-money implied volatilities to the spot volatility. *Journal of Applied Probability 45*(2), 542–550.

Efron, B. and R. J. Tibshirani (1994). *An introduction to the bootstrap*. CRC press.

Fan, J. and Y. Wang (2008). Spot volatility estimation for high-frequency data. *Statistics and its Interface 1*(2), 279–288.

Figueroa-Lopez, J. and S. Olafsson (2016). Short-time Expansions for Close-to-the-money Options under a Lévy Jump Model with Stochastic Volatility. *Finance and Stochastics 20*, 219–265.

Foster, D. and D. Nelson (1996). Continuous Record Asymptotics for Rolling Sample Variance Estimators. *Econometrica 64*, 139–174.

Grafakos, L. (2008). *Classical Fourier Analysis*, Volume 2. Springer.

Heston, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The review of financial studies 6*(2), 327–343.

Jacod, J. and P. Protter (2012). *Discretization of Processes*. Berlin: Springer-Verlag.

Jacod, J. and A. Shiryaev (2003). *Limit Theorems For Stochastic Processes* (2nd ed.). Berlin: Springer-Verlag.

Jacod, J. and V. Todorov (2014). Efficient Estimation of Integrated Volatility in Presence of Infinite Variation Jumps. *Annals of Statistics 42*, 1029–1069.

Jacod, J. and V. Todorov (2018). Limit theorems for integrated local empirical characteristic exponents from noisy high-frequency data with application to volatility and jump activity estimation. *The Annals of Applied Probability 28*(1), 511–576.

Kong, X.-B., Z. Liu, and B.-Y. Jing (2015). Testing for Pure-Jump Processes for High-Frequency Data. *Annals of Statistics 43*(2), 847–877.

Kristensen, D. (2010). Nonparametric filtering of the realized spot volatility: A kernel-based approach. *Econometric Theory 26*, 60–93.

Liu, Q., Y. Liu, and Z. Liu (2018). Estimating spot volatility in the presence of infinite variation jumps. *Stochastic Processes and their Applications 128*(6), 1958–1987.

Madan, D. B. and E. Seneta (1990). The variance gamma (vg) model for share market returns. *Journal of business 63*, 511–524.

Mancini, C. (2001). Disentangling the Jumps of the Diffusion in a Geometric Jumping Brownian Motion. *Giornale dell'Istituto Italiano degli Attuari LXIV*, 19–47.

Mancini, C. (2009). Non-parametric Threshold Estimation for Models with Stochastic Diffusion Coefficient and Jumps. *Scandinavian Journal of Statistics 36*, 270–296.

Medvedev, A. and O. Scaillet (2007). Approximation and calibration of short-term implied volatilities under jump-diffusion stochastic volatility. *The Review of Financial Studies 20*(2), 427–459.

Medvedev, A. and O. Scaillet (2010). Pricing american options under stochastic volatility and stochastic interest rates. *Journal of Financial Economics 98*(1), 145–159.

Pan, J. (2002). The jump-risk premia implicit in options: Evidence from an integrated time-series study. *Journal of financial economics 63*(1), 3–50.

Sato, K. (1999). *Lévy Processes and Infinitely Divisible Distributions*. Cambridge, UK: Cambridge University Press.

Todorov, V. (2019). Nonparametric spot volatility from options. *The Annals of Applied Probability 29*(6), 3590–3636.

Todorov, V. (2020). Testing and inference for fixed times of discontinuity in semimartingales. *Bernoulli 26*(4), 2907–2948.

Todorov, V. (2021). Higher-order small time asymptotic expansion of Itô semimartingale characteristic function with application to estimation of leverage from options. *Stochastic Processes and their Applications 142*, 671–705.

Todorov, V. and Y. Zhang (2021). Information gains from using short-dated options for measuring and forecasting volatility. *Journal of Applied Econometrics*.