The Annals of Statistics
2019, Vol. 47, No. 2, 1025–1060
https://doi.org/10.1214/18-AOS1703
© Institute of Mathematical Statistics, 2019

NONPARAMETRIC IMPLIED LÉVY DENSITIES[1]

BY LIKUAN QIN AND VIKTOR TODOROV

Northwestern University

This paper develops a nonparametric estimator for the Lévy density of an asset price, following an Itô semimartingale, implied by short-maturity options. The asymptotic setup is one in which the time to maturity of the available options decreases, the mesh of the available strike grid shrinks and the strike range expands. The estimation is based on aggregating the observed option data into nonparametric estimates of the conditional characteristic function of the return distribution, the derivatives of which allow to infer the Fourier transform of a known transform of the Lévy density in a way which is robust to the level of the unknown diffusive volatility of the asset price. The Lévy density estimate is then constructed via Fourier inversion. We derive an asymptotic bound for the integrated squared error of the estimator in the general case as well as its probability limit in the special Lévy case. We further show rate optimality of our Lévy density estimator in a minimax sense. An empirical application to market index options reveals relative stability of the left tail decay during high and low volatility periods.

1. Introduction. Option data provides a rich source of information to study risks in the economy and their pricing, and in particular tail events which are hard to measure from asset return data alone. Extracting information from option data, however, is challenging because option prices are determined by various sources of risk (e.g., jumps as well as shocks to stochastic volatility and jump intensity) which need to be explicitly modeled. Therefore, most of the existing work using option data relies on fully specified parametric models. This parametric based evidence, however, is subject to significant misspecification risk, the effects of which are rather unclear due to the highly nonlinear dependence of the option prices on the various sources of risk.

At the same time, recent developments on derivatives markets enable the development and practical implementation of nonparametric estimation techniques, particularly the ones for studying the jump part of the asset returns. More specifically, over the last five years the trading in options with very short time to expiration has increased significantly; see, for example, [1]. For example, for the S&P 500 market index, on each trading day there are now actively traded options with at most two days to expiration.

Received November 2017; revised February 2018.

[1]Supported in part by NSF Grant SES-1530748.

MSC2010 subject classifications. Primary 62G07, 62M05; secondary 60H10, 60J75.

Key words and phrases. Itô semimartingale, Fourier inversion, Lévy density, nonparametric density estimation, options, stochastic volatility.

In this paper, we develop nonparametric estimators for the Lévy density of asset returns from short-dated options. The Lévy density of the asset return process "summarizes" the information about the jumps and has been the main object of study of a large body of statistical work. In the finite activity jump case, the Lévy density can be viewed as the conditional probability of jump arrival of given size. When extracted from option data, the Lévy density contains information both for future expected jump risk as well as for the pricing of it (the Lévy density embedded in the option prices is under the so-called risk-neutral probability measure). Hence, this quantity is of major interest both from a theoretical and applied point of view.

Our nonparametric procedure can be described as follows. First, we aggregate the short-maturity option data into portfolios that provide model-free estimates of the conditional characteristic function of the asset return for different values of the characteristic exponent. This construction follows general results in [8] for replicating expected smooth transforms of the return distribution via portfolios of options. In a next step, we use the conditional characteristic function of the returns to back out the Lévy density. This step is based on the fact that over a short interval, the asset return is approximately like that of a Lévy process (process with i.i.d. increments) with the value of the stochastic volatility and jump intensity "frozen" at their values at the beginning of the interval. Hence, our problem reduces to the nonparametric estimation of the Lévy density of a Lévy process from estimates for the characteristic function of its increments. The main difficulty here is the separation of the volatility from the Lévy density. To achieve this separation, we use the fact that higher order derivatives (from third and above) of the characteristic exponent of a Lévy process are solely determined by the Fourier transform of a known transform of the Lévy density. Hence, the nonparametric estimation of the Lévy density can be done via Fourier inversion of an estimate of the third (or higher order) derivative of the characteristic exponent of the asset return. An alternative approach, which we analyze in the Supplementary Material [27], is to use the second derivative of the characteristic exponent. In this case, however, the diffusive spot volatility also plays a role and, therefore, we need to perform bias correction using a preliminary estimator for the latter.

We derive a bound on the order of magnitude of the integrated squared error in recovering the Lévy density (or a certain known transform of it to be precise) in an asymptotic setting of increasing number of options, with shrinking time to maturity and strike coverage that converges to the whole positive part of the real line. In the special Lévy case, we further derive the probability limit of the integrated squared error of the estimator, based on the third derivative of the characteristic function, and we show that it depends on the Lévy density of the asset price but not on its diffusion component. We further establish rate optimality of our Lévy density estimator in a minimax sense. We test the estimation procedure on simulated data and we apply it to infer stock market risk-neutral Lévy densities from option data on the S&P 500 market index.

The current paper is related to several strands of literature. First, Lévy-based approximations of short-dated options have been studied with various degrees of generality in earlier work; see, for example, [1, 6, 15–17] and [24], and the many references therein. The results of these papers are typically derived for a single option with a fixed strike while here we are interested in the approximation across the whole range of strikes that cover the positive real line. Unlike the current study, many of the above cited papers are not interested in the size of the approximation error or analyze it in somewhat restrictive settings (e.g., for part of the strike domain only and/or under stronger assumptions for the underlying Itô semimartingale). The asymptotic order of the approximation error depends on the strike of the option and, for the purposes of the analysis here, we need to assess its limiting behavior in a functional sense (in the strike). Second, there is a large literature on the nonparametric estimation of the Lévy density from discrete observations of a Lévy process; see, for example, [9, 10, 14, 18, 21, 22] and [25]. Some of these results are further extended to time-changed Lévy processes [3] and affine models [2]. The major difference between the current paper and this strand of work, from a statistical point of view, is that we use option data for the inference which results in a very different statistical setup. Third, most closely related to the current paper is a body of work that considers nonparametric Lévy density estimation in the context of exponential Lévy models from options with fixed maturity; see, for example, [4, 5, 11, 29, 30] and [31, 32]. The major differences between the current work and these papers are two. First, our method applies to the very general Itô semimartingale class of models which nests the exponential Lévy models but also allows for models with time-varying volatility and jump intensity. Second, there is a major difference in the asymptotic setup of the earlier work and the current study: in our case, unlike the previous work, the maturity of the options shrinks. This results in different methods of proofs and also different asymptotic behavior of the estimators: the short maturity of the options here helps the separation of volatility from jumps and we can thus achieve much faster rates of convergence (in probability) than what is feasible in the fixed maturity case. These differences are explained in more detail later in the text.

The rest of the paper is organized as follows. Section 2 describes the option observation scheme and in Section 3 we state our assumptions. The Lévy density estimator is given in Section 4 along with a bound on the asymptotic order of its integrated squared error. Sections 5 and 6 present the results from a Monte Carlo experiment and empirical application, respectively. Section 7 concludes. The proofs are given in Section 8. A Supplementary Material [27] contains additional theoretical results for the limit in probability of the integrated squared error of the estimator, a lower bound for the minmax risk of recovering Lévy density from short-dated noisy option data as well as an alternative estimator (to the one in the main text) based on the second derivative of the characteristic function and volatility debiasing.

2. Setup and notation. We define with X an asset price on a filtered probability space $(\Omega^{(0)}, \mathcal{F}^{(0)}, (\mathcal{F}_t^{(0)})_{t\geq 0}, \mathbb{P}^{(0)})$. In this paper, we will make nonparametric inference on the basis of derivatives written on X, that is, contracts whose payoffs are sole functions of the trajectory of X from inception till expiration. As known from finance theory (see, e.g., [12]), in the absence of arbitrage, the theoretical values of derivatives prices equal their expected future discounted payoffs under the so-called risk-neutral probability, which we henceforth denote with $\mathbb{Q}$. The latter deviates from the true probability because of the risk premia demanded by investors for bearing risk (it overweights bad scenarios and underweights the good scenarios) and is of major interest both for academic and practical applications. The log-price, $x_t = \ln X_t$, is an Itô semimartingale with the following dynamics under the risk-neutral probability measure:

$$x_t = \int_0^t a_s\, ds + \int_0^t \sigma_s\, dW_s + \int_0^t \int_{\mathbb{R}} x \widetilde{\mu}(ds, dx), \tag{2.1}$$

where W is a Brownian motion, μ is an integer-valued random measure on $\mathbb{R}_+ \times \mathbb{R}$, counting the jumps in x, with compensator $\nu_t(x)\, dt \otimes dx$ and $\widetilde{\mu}$ is the martingale measure associated with μ (W and ν_t are defined with respect to $\mathbb{Q}$). Our interest in this paper is the nonparametric estimation of $\nu_t(x)$ at fixed points in time from option data.

We assume that we have observations of option prices written on X at time t, which expire at $t + T$, for some $T > 0$. Since t will be fixed throughout, we will henceforth suppress the dependence on t in the notation of the option prices and other related quantities. Also, without loss of generality, and in order to further simplify notation, we will make the normalization $X_t = 1$. For simplicity, we will also assume that the dividend yield of X and the risk-free interest rate are both zero. With these normalizations, the theoretical values of the option prices we will use in our analysis are given by

$$O_T(k) = \begin{cases} \mathbb{E}_t^{\mathbb{Q}}\big(e^k - e^{x_{t+T}}\big)^+ & \text{if } k \leq \ln F_T, \\ \mathbb{E}_t^{\mathbb{Q}}\big(e^{x_{t+T}} - e^k\big)^+ & \text{if } k > \ln F_T, \end{cases} \tag{2.2}$$

where F_T is the price at time t of a forward contract which expires at time $t + T$, and $K \equiv e^k$ and k are the strike and log-strike, respectively, of the option. Given the simplifying assumption of zero dividend yield and zero interest rate and the normalization $X_t = 1$, we have $F_T \equiv 1$. $O_T(k)$ is the price of an out-of-the-money option (i.e., an option which will be worth zero if it were to expire today). This is a call contract (an option to buy the asset) if $k > 0$ and a put contract (an option to sell the asset) if $k \leq 0$.

Our data consists of out-of-the-money options at time t, expiring at $t + T$, and having log-strikes given by

$$\underline{k} \equiv k_1 < k_2 < \cdots < k_N \equiv \overline{k}, \tag{2.3}$$

with the corresponding strikes being

$$\underline{K} \equiv K_1 < K_2 < \cdots < K_N \equiv \overline{K}. \tag{2.4}$$

We denote the gaps between the log-strikes with $\Delta_i = k_i - k_{i-1}$, for $i = 2, \ldots, N$. We note that we do not assume an equidistant log-strike grid, that is, we allow for Δ_i to differ across i-s. The asymptotic theory developed below is of joint type, in which the time to maturity of the option T goes down to zero, the mesh of the log-strike grid $\sup_{i=2,\ldots,N} \Delta_i$ shrinks to zero and the log-strike limits $\underline{k}$ and $\overline{k}$ increase to infinity in absolute value.

Finally, as common in empirical asset pricing, we allow for observation error, that is, instead of observing $O_T(k_i)$, we observe

$$\widehat{O}_T(k_i) = O_T(k_i) + \varepsilon_i, \tag{2.5}$$

where the sequence of observation errors $\{\varepsilon_i\}_i$ is defined on a space $\Omega^{(1)} = \times_{k\in\mathbb{R}} \mathcal{A}_k$, for $\mathcal{A}_k = \mathbb{R}$. This space is equipped with the product Borel σ-field $\mathcal{F}^{(1)}$ and with transition probability $\mathbb{P}^{(1)}(\omega^{(0)}, d\omega^{(1)})$ from the original probability space $\Omega^{(0)}$—on which X is defined—to $\Omega^{(1)}$. We further define

$$\Omega = \Omega^{(0)} \times \Omega^{(1)}, \qquad \mathcal{F} = \mathcal{F}^{(0)} \times \mathcal{F}^{(1)},$$
$$\mathbb{P}(d\omega^{(0)}, d\omega^{(1)}) = \mathbb{P}^{(0)}(d\omega^{(0)})\mathbb{P}^{(1)}(\omega^{(0)}, d\omega^{(1)}).$$

3. Assumptions.

We proceed with our formal assumptions for the process x, the option observation scheme as well as the observation error. Below, for a generic function f, we will denote with f^* its Fourier transform, provided the latter is well defined.

A1. *The function $h_t = x^3 \nu_t(x)$ belongs to the class*

$$\mathcal{S}_r(C_t) = \left\{ f \in L^1(\mathbb{R}) \cap L^2(\mathbb{R}) : \int_{\mathbb{R}} |f^*(x)|^2 (1 + x^2)^r \, dx \le C_t \right\},$$

for some positive constant r and some positive and $\mathcal{F}_t$-adapted random variable C_t.

A2. *The process σ has the following dynamics under $\mathbb{Q}$:*

$$\sigma_t = \sigma_0 + \int_0^t b_s \, ds + \int_0^t \eta_s \, dW_s + \int_0^t \widetilde{\eta}_s \, d\widetilde{W}_s + \int_0^t \int_{\mathbb{R}} \delta^\sigma(s, u) \mu^\sigma(ds, du), \tag{3.1}$$

where $\widetilde{W}$ is a Brownian motion independent of W; μ^σ is a Poisson random measure on $\mathbb{R}_+ \times \mathbb{R}$ with compensator $\nu^\sigma(ds, du) = ds \otimes du$, having arbitrary dependence with the random measure μ; b, η and $\widetilde{\eta}$ are processes with càdlàg paths and $\delta^\sigma(s, u) : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}$ is left-continuous in its first argument.

A3. *With the notation of* A2, *there exists an* $\mathcal{F}_t$*-adapted random variable* $\overline{t} > t$ *such that for* $s \in [t, \overline{t}]$:

$$\mathbb{E}_t^{\mathbb{Q}}|a_s|^4 + \mathbb{E}_t^{\mathbb{Q}}|\sigma_s|^6 + \mathbb{E}_t^{\mathbb{Q}}(e^{4|x_s|}) + \mathbb{E}_t^{\mathbb{Q}}\left(\int_{\mathbb{R}}(e^{3|z|}-1)\nu_s(z)\,dz\right)^4 < C_t, \tag{3.2}$$

for some $\mathcal{F}_t$*-adapted random variable* C_t, *and in addition for some* $\iota > 0$

$$\mathbb{E}_t^{\mathbb{Q}}\left(\int_{\mathbb{R}}\left(\left|\delta^\sigma(s,z)\right|^4 \vee \left|\delta^\sigma(s,z)\right|\right)dz\right)^{1+\iota} \le C_t. \tag{3.3}$$

A4. *With the notation of* A2, *there exists an* $\mathcal{F}_t$*-adapted random variable* $\overline{t} > t$ *such that for* $s \in [t, \overline{t}]$:

$$\begin{aligned}&\mathbb{E}_t^{\mathbb{Q}}|a_s - a_t|^p + \mathbb{E}_t^{\mathbb{Q}}|\sigma_s - \sigma_t|^p + \mathbb{E}_t^{\mathbb{Q}}|\eta_s - \eta_t|^p + \mathbb{E}_t^{\mathbb{Q}}|\widetilde{\eta}_s - \widetilde{\eta}_t|^p \\ &\quad \le C_t|s-t| \qquad \forall p \in [2,4],\end{aligned} \tag{3.4}$$

and

$$\mathbb{E}_t^{\mathbb{Q}}\left(\int_{\mathbb{R}}(e^{z\vee 0}|z| \vee |z|^2)\left|\nu_s(z) - \nu_t(z)\right|dz\right)^p \le C_t|s-t| \qquad \forall p \in [2,3], \tag{3.5}$$

for some $\mathcal{F}_t$*-adapted random variable* C_t.

A5. *The log-strike grid* $\{k_i\}_{i=1}^N$ *is* $\mathcal{F}_t^{(0)}$*-adapted and on a set with probability approaching one*, *we have*

$$\eta\overline{\Delta} \le \inf_{i=2,\ldots,N} \Delta_i \le \sup_{i=2,\ldots,N} \Delta_i \le \overline{\Delta}, \tag{3.6}$$

where $\eta \in (0,1)$ *is some positive constant and* $\overline{\Delta}$ *is a deterministic sequence with* $\overline{\Delta} \to 0$.

A6. *We have*: (1) $\mathbb{E}(\varepsilon_i|\mathcal{F}^{(0)}) = 0$, (2) $\mathbb{E}(|\varepsilon_i||\mathcal{F}^{(0)}) = O_T(k_i)\zeta_{t,i}$ *and* $\mathbb{E}(\varepsilon_i^2|\mathcal{F}^{(0)}) = O_T(k_i)^2\sigma_{t,i}^2$, *where* $\{\zeta_{t,i}\}_{i=1,\ldots,N}$ *and* $\{\sigma_{t,i}^2\}_{i=1,\ldots,N}$ *are sequences of* $\mathcal{F}_t$*-adapted random variables with* $\sup_{i=1,\ldots,N}\zeta_{t,i} = O_p(1)$ *and* $\sup_{i=1,\ldots,N}\sigma_{t,i}^2 = O_p(1)$, *and* (3) ε_i *and* ε_j *are* $\mathcal{F}^{(0)}$*-conditionally independent whenever* $i \ne j$.

We briefly discuss each of the assumptions. Assumption A1 is a standard assumption and the coefficient r controls the smoothness of the estimated function. As we will see in the next section, our method recovers $x^3\nu_t(x)$ and hence A1 is an assumption for the smoothness of this function. Assumption A2 assumes that the stochastic volatility process, σ, is an Itô semimartingale with jumps of finite variation. This assumption is satisfied in many applications. Importantly, A2 allows for general forms of dependence between the diffusion and jump components of x and σ. Our integrability assumptions are given in A3. We require existence of $\mathcal{F}_t$-conditional moments of the values of various processes evaluated at some, arbitrary close to t, time in the future. We note that A3 imposes the restriction that jumps in x are of finite variation. An extension to infinite variation jumps is possible but at the cost of much slower rates of convergence than the one we get here in Theorem 1 below. Assumption A4 is a "smoothness in expectation" assumption. This assumption will be satisfied if the corresponding processes are Itô semimartingales. Our assumption for the log-strike grid of the observed options is given in A5 and it allows, in particular, for nonequidistant sampling. Finally, A6 contains our assumption for the observation error. We assume that the observation error is centered at zero and that the errors are $\mathcal{F}^{(0)}$-conditionally independent. The latter assumption can be weakened to require only no correlation of the errors and certain products of them. It can be also weakened to allow for $\mathcal{F}^{(0)}$-conditional weak dependence between the errors. Assumption A6 allows for $\mathcal{F}^{(0)}$-conditional heteroskedasticity of the observation error. We note that A6 assumes that the $\mathcal{F}^{(0)}$-conditional variance of the error is of the same order of magnitude as the squared option price it is attached to. This is consistent with the relative observation error being of order $O_p(1)$.

4. Nonparametric Lévy density recovery. We now construct our nonparametric estimator of $\nu_t(x)$ and characterize its asymptotic properties. First, using results in [8] for spanning risk-neutral payoffs from portfolios of options, we have that

$$\mathbb{E}_t^{\mathbb{Q}}(e^{iux_{t+T}}) = e^{iu\ln F_T} - (u^2+iu)\int_{-\infty}^{\infty} e^{(iu-1)k} O_T(k)\,dk, \qquad u\in\mathbb{R}. \tag{4.1}$$

Therefore, using a Riemann sum approximation for the integral in the above expression, we have that

$$\hat{f}_T(u) = e^{iu\ln F_T} - (u^2+iu)\sum_{j=2}^{N} e^{(iu-1)k_{j-1}}\widehat{O}_T(k_{j-1})\Delta_j, \qquad u\in\mathbb{R}, \tag{4.2}$$

is a consistent estimate of $\mathbb{E}_t^{\mathbb{Q}}(e^{iux_{t+T}})$ under very general assumptions for the dynamics of the process x and provided $\underline{k}\downarrow -\infty$, $\overline{k}\uparrow +\infty$ and $\sup_i \Delta_i \to 0$. We note that similar option-spanning idea lies behind the construction of the popular option-implied volatility index VIX which is widely used as a measure of uncertainty and fear gauge. More specifically, the formula for the squared VIX, up to a higher order term, is given by $2\ln F_T - \frac{2}{i}\hat{f}_T'(0)$.

If T is small, then $x_{t+T} - x_t$ is approximately, $\mathcal{F}_t$-conditionally, the increment of a Lévy process with generating (or characteristic) triplet (a_t, σ_t^2, ν_t) (Definition 8.2 in [28]). Therefore, by applying Lévy–Khintchine formula (Theorem 8.1 in [28]) and assuming $\int_{\mathbb{R}}|x|\nu_t(dx) < \infty$ (which is implied by our assumptions), we have for $u\in\mathbb{R}$

$$\hat{f}_T(u) \approx \exp\biggl(iu\ln F_T + iuTa_t - \frac{u^2}{2}T\sigma_t^2 + T\int_{\mathbb{R}}(e^{iux}-1-iux)\nu_t(x)\,dx\biggr), \tag{4.3}$$

and in the proof we make the above statement formal. Recall that our goal is to recover ν_t without any knowledge of a_t and σ_t. To achieve this, we can consider derivatives of $\hat{f}_T(u)$. In particular, due to the approximation in (4.3), we have

$$\frac{\hat{f}_T^{(2)}(u)}{\hat{f}_T(u)} - \frac{\hat{f}_T^{(1)}(u)^2}{\hat{f}_T(u)^2} \approx -T\sigma_t^2 - T\int_{\mathbb{R}} x^2 e^{iux}\nu_t(x)\,dx, \tag{4.4}$$

where for a generic function $g(x)$ we denote with $g^{(p)}(x)$ its pth order derivative. Since the last integral above is the Fourier transform of $x^2\nu_t(x)$, we can recover the latter via Fourier inversion from the expression on the left-hand side of (4.4), provided we can estimate σ_t (a similar idea has been used by [26] in the context of high-frequency data with increasing time span for a Lévy process x). The estimation of σ_t, in turn, can be done either using the option data at hand or high-frequency data on x in a local neighborhood of the observation time of the options.

Yet another approach, which avoids estimation of σ_t, is to base the inference on a derivative of the expression on the left-hand side of (4.4). A similar idea has been used by [10] in the context of estimating the jump density of a Lévy process from high-frequency observations of the process with increasing time span. In particular, given the approximation in (4.3), we have

$$\frac{\hat{f}_T^{(3)}(u)}{\hat{f}_T(u)} - 3\frac{\hat{f}_T^{(2)}(u)\hat{f}_T^{(1)}(u)}{\hat{f}_T(u)^2} + 2\frac{\hat{f}_T^{(1)}(u)^3}{\hat{f}_T(u)^3} \approx -iT\int_{\mathbb{R}} x^3 e^{iux}\nu_t(x)\,dx \tag{4.5}$$

and, therefore,

$$h_t(x) = x^3\nu_t(x), \tag{4.6}$$

can be recovered from the expression on the left-hand side of (4.5) by Fourier inversion. One advantage of the estimation approach based on the expression on the left-hand side of (4.5) over the one based on the expression on the left-hand side of (4.4) is that it avoids inference for σ_t and debiasing. However, estimating $x^3\nu_t(x)$ puts less emphasis on fitting $\nu_t(x)$ around zero as compared to estimating $x^2\nu_t(x)$.

For brevity, here we will present the analysis based on (4.5) only and in the Supplementary Material [27] we present the results for the estimator based on (4.4). The nonparametric estimator using the approximation in (4.5) is constructed as follows. We first define for $u \in \mathbb{R}$,

$$\begin{aligned} \hat{h}_t^*(u) = {} & \frac{i}{T}\frac{\hat{f}_T^{(3)}(u)\hat{f}_T(u) - \hat{f}_T^{(2)}(u)\hat{f}_T^{(1)}(u)}{\hat{f}_T(u)^2} \\ & - \frac{2i}{T}\frac{\hat{f}_T^{(1)}(u)(\hat{f}_T^{(2)}(u)\hat{f}_T(u) - \hat{f}_T^{(1)}(u)^2)}{\hat{f}_T(u)^3}, \end{aligned} \tag{4.7}$$

which, given our discussion above, is an estimate of the Fourier transform of $h_t(x)$. Using $\hat{h}_t^*(u)$ our estimate for $h_t(x)$ is then given by

$$\hat{h}_t(x) = \frac{1}{2\pi}\int_{u\in[u_{N,l},u_{N,h}]} e^{-iux}\hat{h}_t^*(u)\,du, \tag{4.8}$$

where $-u_{N,l}$ and $u_{N,h}$ are deterministic sequences increasing to infinity as the the size of the option data grows. $\hat{h}_t(x)$ is truncated Fourier inverse of $\hat{h}_t^*(u)$, where the highest frequencies have been removed since they are less precisely recovered from the data.

A natural extension of the above estimator can be done by aggregating option data at a fixed number of times $t_1, t_2, \ldots, t_k$, and using the average characteristic function estimate over the observation times and its derivatives. In this case, we recover the time average $\frac{1}{k}\sum_{i=1}^{k} h_{t_i}(u)$. All the results derived below trivially carry over to such an extension of $\hat{h}_t$ and for simplicity of exposition we do not state the formal result for this extension.

The following theorem derives the order of magnitude (in probability) of the integrated squared error in recovering $h_t(x)$. In it, the notation $a_n \asymp b_n$ means that both sequences a_n/b_n and b_n/a_n are bounded.

THEOREM 1. *Suppose Assumptions A1–A6 hold and in addition* $\overline{\Delta} \asymp T^{\alpha}$, $\overline{K} \asymp T^{-\beta}$, $\underline{K} \asymp T^{\gamma}$, *for some* $\alpha > 0$, $\beta > 0$ *and* $\gamma > 0$ *and where* $\overline{\Delta}$ *is the mesh of the log-strike grid. Let* $-u_{N,l} \asymp u_{N,h} \asymp u_N$ *be such that*

$$u_N \to \infty \quad \textit{and} \quad u_N^2(T + \overline{\Delta})|\ln T|^2 \to 0. \tag{4.9}$$

Then we have

$$\int_{\mathbb{R}} \big(\hat{h}_t(x) - h_t(x)\big)^2\,dx = O_p\big(u_N^{-2r} \vee u_N^5 \Gamma_N\big), \tag{4.10}$$

where r *is the constant in Assumption A1 and*

$$\Gamma_N = \overline{\Delta} \vee T\big(\ln^8 T\big)\big(e^{-2\underline{k}}|\underline{k}|^6 \wedge T^{-1/3}\big) \vee e^{6\underline{k}}|\underline{k}|^6 \vee e^{-6\overline{k}}|\overline{k}|^6. \tag{4.11}$$

REMARK 4.1. The short-maturity options can be also used to recover nonparametrically σ_t. One possibility is to expand locally in T the option prices with k close to zero and then make use of the leading role played by σ_t in such an expansion. An alternative approach of recovering σ_t is to use $\frac{1}{T}\Re(\ln \hat{f}_T(u))$ for sufficiently large u where the signal for σ_t is strongest. In both cases, the options with log-strikes close to zero play a leading role in the estimation (which is unlike the case of the estimator $\hat{h}_t$). These options are much larger asymptotically (as $T \downarrow 0$) than the option prices with log-strikes away from zero. Therefore, the effect of the observation error on the estimation of σ_t will be in general bigger than on the recovery of h_t.

Condition (4.9) in the theorem is a relatively weak upper bound on the rate of growth of the sequence u_N which guarantees that $\hat{f}_T(u)$ is bounded from below in absolute value uniformly for $|u| \leq u_N$ with probability approaching one. This is needed as $\hat{h}_t^*(u)$ involves division by $\hat{f}_T(u)$. Note in this regard that the argument of the exponent on the right-hand side of (4.3) is $O_p(Tu^2)$ as $|u| \to \infty$ and $T \to 0$.

Equation (4.10) gives the order of magnitude in probability of the integrated squared error $\int_{\mathbb{R}}(\hat{h}_t(x) - h_t(x))^2\,dx$. Since $\hat{h}_t^*(u)$ involves division by $\hat{f}_T(u)$ and some of the bounds for the options hold only in probability, we characterize the behavior of the integrated squared error in probability and not in (conditional) expectation.

The terms in (4.10)–(4.11) determining the order of magnitude of the integrated squared error reflect the different sources of error in the estimation. As standard for nonparametric density estimation, we have a bias due to truncation of the high frequencies in the Fourier inversion in (4.8). Its contribution in the integrated squared error is $O_p(u_n^{-2r})$, where the parameter r captures the degree of smoothness of $h_t(x)$ (roughly the number of its derivatives). Not surprisingly, the higher r is, the smaller this bias is.

The remaining terms on the right-hand side of (4.10) reflect the various sources of error in approximating $\mathbb{E}_t^{\mathbb{Q}}(e^{iux_{t+T}})$ and its derivatives with the available option data. The first term in Γ_N in (4.11) is due to the measurement error in the option prices. Its order of magnitude is determined by the number of options used in the estimation as well as the order of magnitude of the option error variance, which by Assumption A6 is proportional to the option price that the error is added to. Since the option prices shrink asymptotically in magnitude as $T \downarrow 0$, so do the errors attached to them. We note in this regard that the option price $O_T(k)$ is of order $O_p(T)$ for values of k away from zero and is only of order $O_p(\sqrt{T})$ for k in the vicinity of zero (see Section 8 for the precise statements). As a result, the option error for log-strikes close to zero is asymptotically larger than the one for log-strikes that are large in absolute value. Nevertheless, since the loading on the options with k close to zero in our estimator is of asymptotically smaller order than on those with k away from zero, the asymptotic effect of the observation error on the estimation is determined by that in the observed options with k away from zero. We further point out that the effect of the option error on the recovery of h_t dominates the error due to the smoothness of the option price as a function of k. Recall in this regard that $\hat{f}_T(u)$ involves a Riemann sum approximation of the integral in (4.1).

The second term in Γ_N is due to error in the option price stemming from approximating $x_{t+T} - x_t$ by the increment of $\mathcal{F}_t$-conditionally Lévy process with characteristic triplet (a_t, σ_t^2, ν_t). Naturally, this error depends on the time to maturity T and not on the mesh of the observation grid of strikes $\overline{\Delta}$. Finally, the last two terms in Γ_N are due to the use of options with a finite span of strikes (recall that the integral in (4.1) involves integration with respect to k over $\mathbb{R}$ while the strike range used in the estimation is $[\underline{k}, \overline{k}]$). The size of this error is determined only by the order of magnitude of $\underline{k}$ and $\overline{k}$ and the tail decay of the (risk-neutral) conditional return distribution.

The result in (4.10) is an upper bound for the estimation error in $\hat{h}_t$. However, in the Supplementary Material [27] we derive the limit in probability of the integrated squared error of the estimator in the specialized setting where the underlying process is Lévy. This result shows that the bound in (4.10) for the effect of the observation error on the estimation [first term in (4.11)] is sharp. The limit in probability result also shows that the precision in the estimation does not depend on the diffusive component of the underlying asset price. This is because our estimator puts more weight (in relative terms) to the options with k away from zero where the signal from the diffusion is relatively weak and is of higher asymptotic order.

The magnitude of the effect on the precision of $\hat{h}_t$ from the stochasticity and time-variation of the drift, volatility and Lévy density as well as from the finiteness of the strike range of the options used in the estimation (last three terms in Γ_n) is derived under rather general conditions for the process X. Under stronger assumptions for X, this bound on the effect of this approximation error in the estimation can be further improved upon. First, in the case when the Lévy measure has a finite support, then we do not need $\underline{k} \downarrow -\infty$ and $\overline{k} \uparrow +\infty$ but finite $[\underline{k}, \overline{k}]$, which covers the support of the Lévy measure, would suffice. In this case, the last three terms in (4.11) can be replaced with a term involving only T. Second, in the case when σ_t and $\nu_t(x)$ are time-varying but deterministic, our nonparametric estimation procedure recovers nonparametrically $\frac{1}{T}\int_t^{t+T} h_s(x)\,ds$, and the only source of error in this case stems from the approximation of the integral on the right-hand side of (4.1) by a finite set of options.

For a given degree of smoothness r of h_t, and provided the leading term in Γ_n is $\overline{\Delta}$ (due to the observation error), the asymptotic size of the integrated squared error of our estimator is $O_p(\overline{\Delta}^{2r/(2r+5)})$ whenever u_N is chosen optimally. This rate is achievable regardless of the presence of a diffusion component in the price and regardless of the presence of time-variation in σ_t and $\nu_t(x)$. In the Supplementary Material [27], we show that this is the best achievable rate for an estimator of h_t in a minimax sense in the specialized setting of X being pure-jump process of finite activity and when the observation errors are Gaussian with standard deviations proportional to the option prices.

We finish this section with a comparison of our method with existing nonparametric estimators of the Lévy measure from option data. Cont and Tankov [11] uses penalized least squares while [4, 5, 29, 30] and [31, 32] use spectral-based techniques for recovering the Lévy measure of exponential Lévy models from options with fixed maturity T. We will restrict comparison to the existing spectral-based estimators as their rates of convergence are explicitly derived. These estimators are based on measures of the Fourier transform of the option price as a function of the strike. That is, they are based on option-based estimates of $\mathbb{E}_t^{\mathbb{Q}}(e^{iux_{t+T}+x_{t+T}})$ which is very similar to our use of the characteristic function $\mathbb{E}_t^{\mathbb{Q}}(e^{iux_{t+T}})$. Belomestny and Reiss [4] and [29] work directly with $\mathbb{E}_t^{\mathbb{Q}}(e^{iux_{t+T}+x_{t+T}})$ in a setting of finite activity jumps while [30] and [31, 32] use derivatives of it and allow for more general jump specifications similar to our use of derivatives of $\hat{f}_T$.

There are two major differences between the existing option-based spectral methods and our work. First, the earlier work applies only to the class of exponential Lévy models while the developed method here is for a general Itô semimartingale. Second, the asymptotic setup in the current work is one in which $T \downarrow 0$ simultaneously with $\overline{\Delta} \downarrow 0$ while for the existing option-based methods T remains fixed. This results in a rather different asymptotic analysis as well as rates of convergence for the Lévy density estimate. Indeed, in a setting in which $T \downarrow 0$, $O_T(k) \downarrow 0$ but this does not happen uniformly in the strike domain. This fact has a nontrivial impact on the analysis because our estimator is a sum of options with different strikes. Furthermore, since the option error is proportional to the unobservable true option price, the magnitude of this error depends on T in a rather nontrivial way.

The decreasing maturity T of the options can be also utilized to separate volatility from jumps which has nontrivial asymptotic effect. In our analysis, similar to the previous spectral option-based estimators, we use increasing u_N but such that $u_N^2 T \to 0$. Since, in our case $T \downarrow 0$, we are effectively evaluating the characteristic function of the return around zero (which is unlike the fixed T case). This helps in the separation of volatility from jumps similar to the case of separating volatility from jumps from high-frequency data on x analyzed in [20]. This also allows to minimize the effect from the time-varying characteristics of the Itô semimartingale on the estimation. As a result, there is a difference in the rate of convergence in the fixed T and decreasing T cases. Focusing on the optimal rates for a given smoothness parameter r, and assuming the first term in Γ_n is leading, we have optimal rate of convergence of $\Delta_n^{-r/(2r+5)}$. This is exactly the same rate as for the estimator of [4] but only in the case when x is compound Poisson. When x contains a diffusion, on the other hand, the best possible rate for estimating the Lévy measure in the case T fixed is only logarithmic while our estimator in the setting $T \downarrow 0$ continues to converge at the fast rate $\Delta_n^{-r/(2r+5)}$.

5. Monte Carlo study. We next present results for the performance of our nonparametric procedure on simulated data from the following model for the risk-neutral dynamics of X:

$$X_t = X_0 + \int_0^t \sqrt{V_s}\, dW_s + \int_0^t \int_{\mathbb{R}} (e^x - 1)\widetilde{\mu}(ds, dx), \tag{5.1}$$

with W being a Brownian motion and V having the dynamics

$$dV_t = \kappa(\theta - V_t)\, dt + \sigma\sqrt{V_t}\rho\, dW_t + \sigma\sqrt{V_t}\sqrt{1-\rho^2}\, d\widetilde{W}_t, \tag{5.2}$$

where $\widetilde{W}$ is a Brownian motion orthogonal to W. The jump measure μ has a compensator $\nu_t(x)\,dt \otimes dx$, where

$$\nu_t(dx) = c_- V_t \frac{e^{-\lambda_-|x|}}{|x|^{1+\beta}}\,dx 1_{\{x<0\}} + c_+ V_t \frac{e^{-\lambda_+|x|}}{|x|^{1+\beta}}\,dx 1_{\{x>0\}}. \tag{5.3}$$

The specification in (5.1)–(5.3) belongs to the affine class of models of [13] commonly used in empirical option pricing work. Consistent with existing empirical evidence, the jumps have time-varying jump intensity. The jump size distribution is like the one of the tempered stable process of [7] which is found to provide good fit to option data. The parameter β controls the behavior of the jump measure around zero, with $\beta < 0$ corresponding to finite jump activity and $\beta \geq 0$ to infinite activity jump specifications. The parameters $\lambda_\pm$, on the other hand, control the behavior of the jump measure in the tails. We set the model parameters in a way that results in option prices similar to observed equity index option data. In particular, we set $\theta = 0.022$, $\kappa = 3.6124$, $\sigma = 0.2$ and $\rho = -0.5$ (our unit of time is one year). We set the jump tail parameters at $\lambda_- = 20$ and $\lambda_+ = 100$. We consider three cases for β. In all cases the parameters $c_\pm$ are set so that the total expected jump variation is equal to the expected diffusive variance, and further the ratio of negative to positive jump variation is 10 to 1. The parameter specifications are as follows. *Case* A: $\beta = -0.5$, $c_- = 1.2233 \times 10^3$ and $c_+ = 6.8387 \times 10^3$. *Case* B: $\beta = 0$, $c_- = 3.6364 \times 10^2$ and $c_+ = 9.0909 \times 10^2$. *Case* C: $\beta = 0.5$, $c_- = 0.9175 \times 10^2$ and $c_+ = 1.0258 \times 10^2$.

The strike grid, strike range and the total number of options per day are calibrated to match roughly the data we use in the empirical application. In particular, at time t, we set $\underline{k} = -8 \times \sigma_{\mathrm{ATM}}\sqrt{T} + \ln X_t$ and $\overline{k} = 2 \times \sigma_{\mathrm{ATM}}\sqrt{T} + \ln X_t$, where we denote with σ_{ATM} the Black–Scholes implied volatility of the at-the-money option. We further use $N = 60$ options and assume equidistant grid for the log-strike k. The option observation error is set to $\varepsilon_i = \frac{1}{2} Z_i \times O_T(k_i)\frac{\psi(k_i)}{Q_{0.995}}$, where $\{Z_i\}_i$ is a sequence of i.i.d. standard normal random variables, Q_α denotes the α-quantile of the standard normal and $\psi(k)$ is a function of the log-strike determined by running a nonparametric kernel regression on the data used in the empirical application of the relative option bid-ask spread (i.e., the bid-ask spread divided by the mid-quote) as a function of the volatility-adjusted log-strike $(k - \ln X_t)/(\sigma_{\mathrm{ATM}}\sqrt{T})$.

The estimation is done on the basis of one month of short-maturity option data (20 trading days). As in the real data, the options are observed during trading days at the time of market close. Similar to the available data, the time to maturity of the options in the estimation window changes in four cycles (each corresponding to one week) from 5 business days to expiration to 1 business day to expiration. Finally, in the Monte Carlo study, we consider three cases for the starting value of volatility: low, median and high, corresponding to 25th, 50th and 75th quantiles, respectively, of the unconditional distribution of V. For simplicity, assume that the statistical and risk-neutral probabilities for the volatility dynamics coincide.

The frequency cutoff parameter vector $(u_{N,l}, u_{N,h})$ is set according to the following simple rule. We first compute $\sum_{t=1}^{20} \hat{h}_t^*(u)$ for a wide range $[-50, 50]$ of u. We then set

$$u_{N,l} = \underset{u\in[-50,0)}{\operatorname{argmin}} \left|\sum_{t=1}^{20} \hat{h}_t^*(u)\right| \quad \text{and} \quad u_{N,h} = \underset{u\in(0,50]}{\operatorname{argmin}} \left|\sum_{t=1}^{20} \hat{h}_t^*(u)\right|.$$

The intuition behind this choice is that, while $\lim_{|u|\to\infty} |h_t^*(u)| = 0$ due to the smoothness of $h_t(u)$, $\hat{h}_t^*(u)$ will not shrink to zero for very large values of $|u|$, for a given N, due to the discreteness of the strike grid of the available options. The above choice of $u_{N,l}$ and $u_{N,h}$ picks the range of u which is roughly consistent with the asymptotic behavior of $h_t^*(u)$ in the tails.

On Figure 1, we illustrate the performance of our estimator on one simulated option data set for each of the cases and for the three different starting values of the volatility. Overall the recovery of the Lévy density for the negative jumps seems quite satisfactory while the estimate of the Lévy density for the positive jumps is relatively noisy. There are two explanations for this. First, the available strike range for out-of-the-money calls (whose value is determined predominantly by the positive jumps) is much smaller than for the puts. Second, the Lévy density for the positive jumps is typically quite small and as a result it becomes very steep around the origin (this is because the Lévy density explodes at zero for the infinite activity jump cases). This steep decay of the Lévy density for the positive jumps is hard to estimate precisely with the given mesh of the log-strike grid.

On the other hand, the Lévy density for the negative jump size is recovered well in all considered cases. In relative terms, the least precise results are obtained for the low volatility regime. This is to be expected since when volatility is rather low, the strike range is relatively narrow (remember our log-strike grid is set proportional to the current level of volatility). Nevertheless, the deviations from the true Lévy density even in the low volatility regime appear small.

The above observations are confirmed from the results of a Monte Carlo study which are summarized in Table 1. We report the integrated squared error as well as a measure of the variability of the function $h_t(x)$ over a range for the jump size of $[-0.3, 0.3]$. This range is rather wide. Indeed, in all considered cases we have $\int_{|x|\le 0.3} |x|\nu_t(x)\,dx / \int_{\mathbb{R}} |x|\nu_t(x)\,dx > 0.99$. From Table 1, we can notice that in all scenarios, the integrated squared error is small relative to the variation of $h_t(x)$. As suggested from the analysis of Figure 1 above, the low volatility regime is the most difficult for the recovery of the Lévy density.

Finally, we note that, similar to standard nonparametric kernel regressions, estimation is less precise at the edges of the support of the Lévy density, which here means for x around zero as well as for x approaching $\pm\infty$. In these regions, in finite samples we can even have negative estimates. This is of course not surprising and is mere reflection of the weak signal in the data about the Lévy density at zero

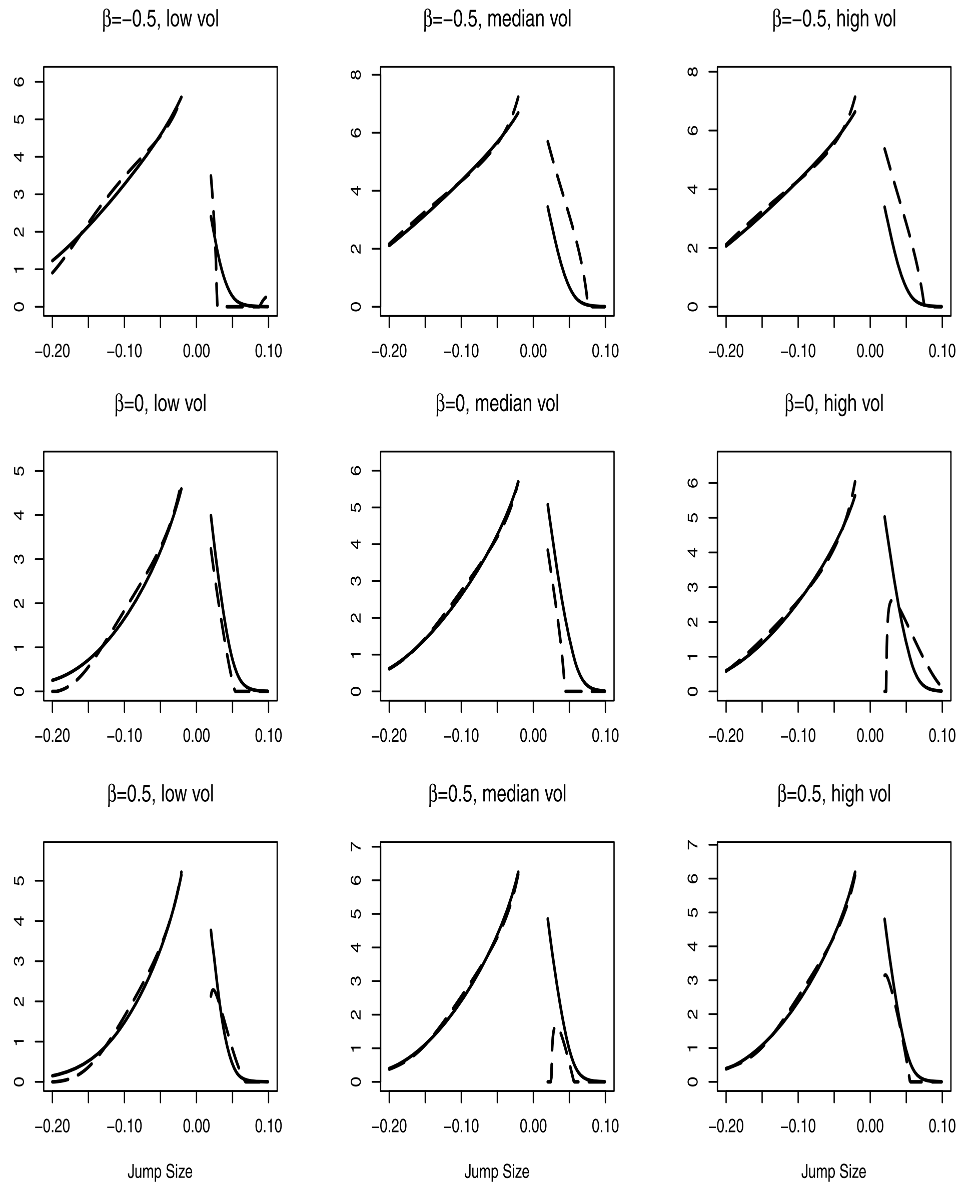

FIG. 1. *Estimated Lévy density in the Monte Carlo study. On each plot, we display* $\ln(1+\frac{1}{x^3}\hat{h}_t(x))$ (*dashed line*) *and* $\ln(1+\nu_t(x))$ (*solid line*) *on one simulated option data set.*

and infinity. Most accurate results for the recovery of $\nu_t(x)$ are obtained for values of x within the log-strike range $[\underline{k}, \overline{k}]$ and which in absolute value are slightly above zero (determined by the mesh of the strike grid relative to the underlying asset price).

TABLE 1
Monte Carlo Results. The MISE is the average of $\int_{|x|\le 0.3}(\hat{h}_t(x) - h_t(x))^2\,dx$ *over* 1000 *replications*

Model case	**Vol regime**	$\int_{|x|\le 0.3} h_t^2(x)\,dx$	**MISE**
A	Low	3.13×10^{-4}	7.65×10^{-6}
A	Median	5.93×10^{-4}	1.06×10^{-5}
A	High	1.04×10^{-3}	1.61×10^{-5}
B	Low	7.11×10^{-6}	6.86×10^{-7}
B	Median	1.36×10^{-5}	6.86×10^{-7}
B	High	2.34×10^{-5}	6.14×10^{-7}
C	Low	4.68×10^{-6}	2.44×10^{-7}
C	Median	8.82×10^{-6}	2.62×10^{-7}
C	High	1.52×10^{-5}	2.84×10^{-7}

Overall, the results from the Monte Carlo study suggest satisfactory performance of the Lévy density recovery in empirically realistic settings.

6. Empirical application. We now apply our nonparametric Lévy density estimation method to data on short-maturity options written on the S&P 500 market index. With the introduction of the weekly options (i.e., options that expire on a weekly basis), the availability of short-dated options have increased significantly, see the evidence in [1]. As a result, for each trading day we have options written on the S&P 500 index that expire on the Friday of the week. Therefore, the time to maturity of the closest to expiration available options ranges from one business day to five business days, resulting in an average time to maturity of our option data of 2.5 business days. Our data covers the period 2014–2015 but for brevity we present only results for April 2015 and September 2015. These are two very different periods in terms of market behavior: the first is a very calm period with low market volatility and the second is a more turbulent one generated in part by heightened global economic uncertainty.

We use mid-quotes at market close from OptionMetrics and remove strikes with zero bids for the traded out-of-the-money options. The average number of options per day in our sample is 64 and the average log-strike range is $[-9.14 \times \sigma_{\text{ATM}}\sqrt{T}, 3.27 \times \sigma_{\text{ATM}}\sqrt{T}]$, which are similar to the corresponding numbers we used in our Monte Carlo study. Finally, we set the value of the frequency cutoff vector $(u_{N,l}, u_{N,h})$ used in the Fourier inversion exactly as in the simulation study.

The results from the estimation are presented in Figure 2. We can make several observations. First, in both periods, our estimates for the Lévy density of the positive jumps are very close to zero. This is consistent with earlier empirical evidence

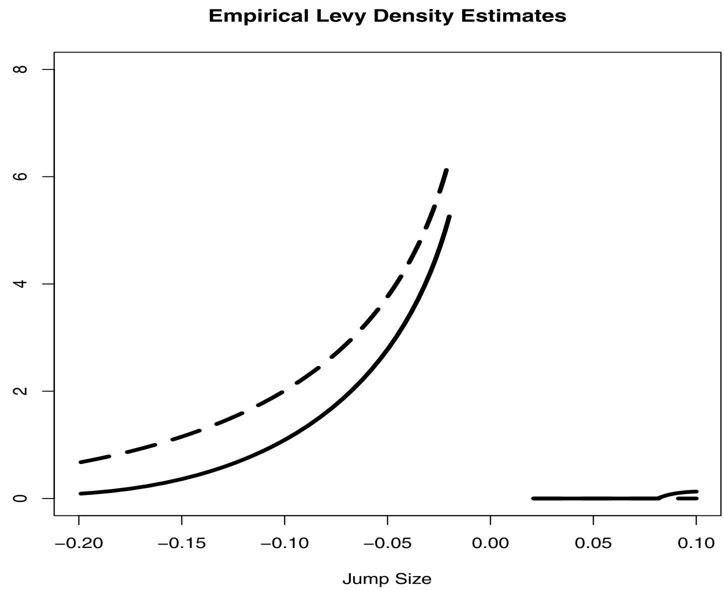

FIG. 2. *Empirical estimates of Lévy density. The lines represent estimate of* $\ln(1+\nu_t(x))$ *for April* 2015 (*solid line*) *and September* 2015 (*dashed line*).

and is also in line with our Monte Carlo study where we found difficulty in accurately recovering the shape of the Lévy density of the positive jumps due to its small magnitude (the authors of [1] report difficulty in its recovery even in a parametric setting). Second, we recover a Lévy density for the negative jumps which is monotone in the jump size. This is like the parametric models we used in the Monte Carlo study but is unlike a Gaussian model for the jump distribution (with negative mean) commonly used in finance since the seminal work of [23]. Third, even though the two considered periods are very different in terms of volatility, the estimated shape of the left tail appears quite stable and the change in the left jump tail from April 2015 to September 2015 can be instead explained by a level shift.

7. Conclusion. In this paper, we develop a nonparametric estimator for the Lévy density of an asset price from noisy observations of short-dated options written on it. We derive a (sharp) bound for the asymptotic order of the integrated squared error of the estimator and we show its rate optimality in the current asymptotic setting. The nonparametric Lévy density estimator can be used as an important diagnostic tool for building and testing parametric models as well as for the construction of nonparametric measures of (risk-neutral) jump risk. Existing model specifications for asset returns differ both in terms of dynamics of the stochastic volatility and jump intensity as well as the jump distribution (Lévy density). The current estimator applies to settings with time-varying Itô semimartingale characteristics and can help the parametric modeling in a robust way.

8. Proofs. For ease of exposition, in the proofs we will set $u_N = -u_{N,l} = u_{N,h}$, with the more general asymmetric cutoff case being shown in exactly the same way.

Furthermore, in the proofs we will denote with C_t a finite-valued $\mathcal{F}_t$-adapted random variable which might change from line to line. If the variable depends on some parameter q, then we will use the notation $C_t(q)$.

8.1. *Decompositions and notation.* The jump part of the process x_t can be represented as an integral with respect to Poisson random measure. In particular, using the so-called Grigelionis representation of the jump part of a semimartingale (Theorem 2.1.2 of [19]), upon suitably extending the probability space, we can write

$$\int_0^t \int_{\mathbb{R}} \widetilde{\mu}(ds, dx) \equiv \int_0^t \int_E \delta^x(s, z)\widetilde{\mu}^x(ds, dz), \tag{8.1}$$

where $\mu^x(ds, dz)$ is a Poisson measure on $\mathbb{R}_+ \times E$ with compensator $dt \otimes \lambda(dz)$, for some σ-finite measure on E, $\widetilde{\mu}^x$ is the martingale counterpart of μ^x, and δ^x is a predictable and $\mathbb{R}$-valued function on $\Omega \times \mathbb{R}_+ \times E$ such that $\nu_t(z)\,dz$ is the image of the measure λ under the map $z \to \delta^x(t, z)$ on the set $\{z : \delta^x(\omega, t, z) \neq 0\}$.

There are different choices for E, λ and the function δ^x. For the analysis here, it will be convenient to use $E = \mathbb{R}_+ \times \mathbb{R}$, λ to be the Lebesgue measure, and $\delta^x(t, z) = z_2 1_{\{z_1 \leq \nu_t(z_2)\}}$ for $z = (z_1, z_2)$.

We proceed with introducing some notation that will be used throughout the proofs. By noting that $x_t = 0$, we can split x_s into

$$\begin{aligned} x_s^c &= \int_t^s a_u\,du + \int_t^s \sigma_u\,dW_u, \\ x_s^d &= \int_t^s \int_E \delta^x(u, z)\widetilde{\mu}^x(du, dz), \qquad s \geq t. \end{aligned} \tag{8.2}$$

We now introduce two approximations for x_s. The first is $\widetilde{x}_s = \widetilde{x}_s^c + \widetilde{x}_s^d$, where for $s \geq t$:

$$\widetilde{x}_s^c = a_t(s - t) + \sigma_t(W_s - W_t), \qquad \widetilde{x}_s^d = \int_t^s \int_E \delta^x(t, z)\widetilde{\mu}^x(du, dz). \tag{8.3}$$

The second approximation is given by $\overline{x}_s = \overline{x}_s^c + \overline{x}_s^d$, where for $s \geq t$:

$$\begin{aligned} \overline{x}_s^c &= a_t(s - t) + \int_t^s \overline{\sigma}_u\,dW_u, \qquad \overline{x}_s^d = \widetilde{x}_s^d, \\ \overline{\sigma}_s &= \sigma_t + \eta_t(W_s - W_t) + \widetilde{\eta}_t(\widetilde{W}_s - \widetilde{W}_t). \end{aligned} \tag{8.4}$$

The option prices at time t associated with terminal value $\widetilde{x}_{t+T}$ are denoted with $\widetilde{O}_T(k)$ and the ones with terminal value of $\overline{x}_{t+T}$ are denoted with $\overline{O}_T(k)$.

8.2. *Proof of Theorem 1*. We set

$$f_T(u) = e^{iu \ln F_T} - (u^2 + iu) \int_{-\infty}^{+\infty} e^{(iu-1)k} \widetilde{O}_T(k)\, dk. \tag{8.5}$$

From [8], we have

$$f_T(u) \equiv \mathbb{E}_t^{\mathbb{Q}}(e^{iu \ln \widetilde{X}_{t+T}}), \tag{8.6}$$

and, therefore, since $\ln \widetilde{X}_s$ is an $\mathcal{F}_t$-conditional Lévy process for $s \geq t$, by applying Lévy–Khintchine formula (Theorem 8.1 in [28]), we can further write

$$f_T(u) = \exp\left(iua_t T - \frac{u^2}{2} T\sigma_t^2 + T \int_{\mathbb{R}} (e^{iux} - 1 - iux)\nu_t(x)\, dx\right), \tag{8.7}$$

where we used the normalization $\ln \widetilde{X}_t = 1$. From here, if we denote

$$h_t^*(u) = \int_{\mathbb{R}} e^{iux} h_t(x)\, dx, \tag{8.8}$$

then we observe that $h_t^*(u)$ coincides with the second derivative in u of the function $(i/T) f_T^{(1)}(u)/f_T(u)$ and $\hat{h}_t^*(u)$ is its sample analogue. Applying the Plancherel's identity, we can write

$$\begin{aligned} &\int_{\mathbb{R}} (\hat{h}_t(x) - h_t(x))^2\, dx \\ &\quad = \underbrace{\frac{1}{2\pi} \int_{|u| \leq u_N} |\hat{h}_t^*(u) - h_t^*(u)|^2\, du}_{IV} + \underbrace{\frac{1}{2\pi} \int_{|u| > u_N} |h_t^*(u)|^2\, du}_{IB}. \end{aligned} \tag{8.9}$$

By Assumption A1, for the bias due to the truncation of the higher frequencies, we have $IB = O(u_N^{-2r})$. For the analysis of IV, we will suitably decompose the difference $\hat{h}_t^*(u) - h_t^*(u)$ and we will analyze the terms in the decomposition separately. To this end, we split

$$\hat{f}_T(u) - f_T(u) = \hat{f}_{T,1}(u) + \hat{f}_{T,2}(u) + \hat{f}_{T,3}(u) + \hat{f}_{T,4}(u), \tag{8.10}$$

where $\hat{f}_{T,k}(u) = -(u^2 + iu)\overline{f}_{T,k}(u)$, for $k = 1, 2, 3, 4$, with

$$\overline{f}_{T,1}(u) = \sum_{j=2}^{N} e^{(iu-1)k_{j-1}} \varepsilon_{j-1} \Delta_j,$$

$$\overline{f}_{T,2}(u) = \sum_{j=2}^{N} e^{(iu-1)k_{j-1}} (O_T(k_{j-1}) - \widetilde{O}_T(k_{j-1}))\Delta_j,$$

$$\overline{f}_{T,3}(u) = \sum_{j=2}^{N} \int_{k_{j-1}}^{k_j} (e^{(iu-1)k_{j-1}} \widetilde{O}_T(k_{j-1}) - e^{(iu-1)k} \widetilde{O}_T(k))\, dk,$$

$$\overline{f}_{T,4}(u) = -\int_{-\infty}^{k_1} e^{(iu-1)k} \widetilde{O}_T(k)\, dk - \int_{k_N}^{+\infty} e^{(iu-1)k} \widetilde{O}_T(k)\, dk.$$

We decompose in a similar way $\hat{f}_T^{(j)}(u) - f_T^{(j)}(u) = \sum_{k=1}^4 \hat{f}_{T,k}^{(j)}(u)$, for $j = 1, 2, 3$.

We now derive bounds for $\hat{f}_{T,k}^{(j)}(u)$ using the auxiliary lemmas in Section 8.3. First, from Lemma 8, we have

$$\sup_{u\in\mathbb{R}}\left[\frac{1}{(1+|u|)^2}\sum_{j=0}^{3}\left|\hat{f}_{T,1}^{(j)}(u)\right|\right] = O_p\big(T|\ln T|\big) \tag{8.11}$$

and

$$\int_{-u_N}^{u_N}\left|\hat{f}_{T,1}(u)\right|^2 du = O_p\big(u_N^5 T^{3/2}\overline{\Delta}\big),$$
$$\int_{-u_N}^{u_N}\left|\hat{f}_{T,1}^{(1)}(u)\right|^2 du = O_p\big(u_N^5 T^2\overline{\Delta} \vee u_n^3 T^{3/2}\overline{\Delta}\big), \tag{8.12}$$

$$\int_{-u_N}^{u_N}\left|\hat{f}_{T,1}^{(2)}(u)\right|^2 du = O_p\big(u_N^5 T^2\overline{\Delta} \vee u_n T^{3/2}\overline{\Delta}\big),$$
$$\int_{-u_N}^{u_N}\left|\hat{f}_{T,1}^{(3)}(u)\right|^2 du = O_p\big(u_N^5 T^2\overline{\Delta}\big). \tag{8.13}$$

In addition, taking into account $u_N^2 T \to 0$, we also have for $|u| \le u_N$

$$\left|f_T^{(1)}(u)\right| \le C\sqrt{T}, \qquad \left|f_T^{(2)}(u)\right| + \left|f_T^{(3)}(u)\right| \le CT. \tag{8.14}$$

Next, using Assumption A5 for the observation grid of strike prices, we have for $j = 0, 1, 2, 3$

$$\begin{aligned}
\sum_{i:|e^{k_{i-1}}-1|\le\sqrt{T}} \Delta_i &\le \int_{\ln(1-\sqrt{T})}^{\ln(1+\sqrt{T})} dk + 2\overline{\Delta} \\
&= O_p(\sqrt{T}) \qquad \text{if } \alpha > 1/2, \\
\sum_{i:|e^{k_{i-1}}-1|\in(\sqrt{T},0.5\vee\sqrt{T}]} \frac{\Delta_i}{|k_{i-1}|} &\le \int_{\ln(1-\sqrt{T})\wedge\ln(0.5)}^{\ln(1-\sqrt{T})+\overline{\Delta}} \frac{dk}{|k|} + \int_{\ln(1+\sqrt{T})-\overline{\Delta}}^{\ln(1+\sqrt{T})\vee\ln(1.5)} \frac{dk}{|k|} \\
&= O_p\big(\ln(1/T)\big) \qquad \text{if } \alpha > 1/2,
\end{aligned}$$

and similarly

$$\sum_{i:|e^{k_{i-1}}-1|\le 2\overline{\Delta}} \Delta_i = O_p(\overline{\Delta}),$$
$$\sum_{i:|e^{k_{i-1}}-1|\in(2\overline{\Delta},0.5]} \frac{\Delta_i}{|k_{i-1}|} = O_p\big(\ln(1/T)\big) \qquad \text{if } \alpha \le 1/2,$$

as well as

$$\sum_{i:|e^{k_{i-1}}-1|>0.5\vee\sqrt{T}} k_{i-1}^j 1_{\{k_{i-1}>0\}}\Delta_i$$
$$\leq \int_{\ln(1+\sqrt{T})\vee\ln(1.5)}^{\overline{k}+\overline{\Delta}} k^j\,dk = O_p(\overline{k}^{j+1}),$$
$$\sum_{i:|e^{k_{i-1}}-1|>0.5\vee\sqrt{T}} e^{-k_{i-1}}|k_{i-1}|^j 1_{\{k_{i-1}\in(\frac{1}{6}\ln T,0)\}}\Delta_i$$
$$\leq \int_{(\ln(T^{1/6})\vee\underline{k})-\overline{\Delta}}^{\ln(1-\sqrt{T})\wedge\ln(0.5)} e^{-k}|k|^j\,dk$$
$$= O_p\big(e^{-\underline{k}}|\underline{k}|^j \wedge T^{-\frac{1}{6}}|\ln T|^j\big),$$

and if further $\underline{k} < \frac{1}{6}\ln T$ we have

$$\sum_{i:|e^{k_{i-1}}-1|>0.5\vee\sqrt{T}} e^{2k_{i-1}}|k_{i-1}|^j 1_{\{k_{i-1}\in[\underline{k},\frac{1}{6}\ln T]\}}\Delta_i$$
$$\leq \int_{\underline{k}}^{\ln(T^{1/6})+\overline{\Delta}} e^{2k}|k|^j\,dk = O_p\big(T^{\frac{1}{3}}|\ln T|^j\big).$$

For $\underline{k} < \frac{1}{6}\ln T$, we have

$$\begin{aligned}\big|\overline{f}_{T,2}(u)\big| \leq &\sum_{i:k_{i-1}\in[\overline{k},\ln(T^{1/6}\wedge(1-0.5\vee\sqrt{T}))]} e^{-k_{i-1}}\big(O_T(k_{i-1})+\widetilde{O}_T(k_{i-1})\big)\Delta_i\\ &+\sum_{i:k_{i-1}>\ln(T^{1/6}\wedge(1-0.5\vee\sqrt{T}))} e^{-k_{i-1}}\big|O_T(k_{i-1})-\widetilde{O}_T(k_{i-1})\big|\Delta_i,\end{aligned}$$

and using Lemma 2 for the first summand on the right-hand side of the above inequality and Lemmas 3–4 for the other one, we have for $\underline{k} < \frac{1}{6}\ln T$

$$\begin{aligned}\big|\overline{f}_{T,2}(u)\big| \leq C_t T &\sum_{i:k_{i-1}\in[\overline{k},\ln(T^{1/6}\wedge(1-0.5\vee\sqrt{T}))]} e^{2k_{i-1}}\Delta_i + C_t|\ln T|\\ &\times\sum_{i:k_{i-1}>\ln(T^{1/6}\wedge(1-0.5\vee\sqrt{T}))} e^{-k_{i-1}}\bigg(T^{3/2}\vee\bigg(\frac{T^{3/2}}{|e^{k_{i-1}}-1|}\wedge T\bigg)\bigg)\Delta_i.\end{aligned}$$

Similarly, for $\underline{k} \geq \frac{1}{6}\ln T$, by application of Lemmas 3–4, we get

$$\big|\overline{f}_{T,2}(u)\big| \leq C_t|\ln T|\sum_{i=1}^N e^{-k_{i-1}}\bigg(T^{3/2}\vee\bigg(\frac{T^{3/2}}{|e^{k_{i-1}}-1|}\wedge T\bigg)\bigg)\Delta_i.$$

These bounds can be easily extended to ones for $|\hat{f}_{T,2}(u)|$ by using $|\hat{f}_{T,2}(u)| \leq 2(u^2 \vee 1)|\overline{f}_{T,2}(u)|$, and analogous bounds can be derived also for $|\hat{f}^{(j)}_{T,2}(u)|$. From here, by taking into account the orders of magnitude derived above for various sums over functions of k_i, Δ_i and T, we have

$$\begin{aligned} &|\hat{f}^{(j)}_{T,2}(u)| \\ &\quad \leq C_t\big(|u| \vee 1\big)^2\big[T^{\frac{3}{2}}|\ln T|^{j+1}\big(\big(|\underline{k}|^j e^{-\underline{k}} \wedge T^{-\frac{1}{6}}\big) + \overline{k}^{j+1}\big) + |\ln T|T\overline{\Delta}\big]. \end{aligned} \tag{8.15}$$

We turn next to $\hat{f}^{(j)}_{T,3}(u)$. We first split $\overline{f}_{T,3}(u)$ into

$$\overline{f}^{(a)}_{T,3}(u) = \sum_{j=2}^{N} \int_{k_{j-1}}^{k_j} \big(e^{(iu-1)k_{j-1}} - e^{(iu-1)k}\big)\widetilde{O}_T(k_{j-1})\,dk,$$

$$\overline{f}^{(b)}_{T,3}(u) = \sum_{j=2}^{N} \int_{k_{j-1}}^{k_j} e^{(iu-1)k}\big(\widetilde{O}_T(k_{j-1}) - \widetilde{O}_T(k)\big)\,dk,$$

and we make similar separations of $\overline{f}^{(j)}_{T,3}(u)$, for $j = 1, 2, 3$. For $\overline{f}^{(a,j)}_{T,3}(u)$, we make use of the following algebraic inequalities:

$$\begin{aligned} &\big|e^{(iu-1)k_{j-1}} - e^{(iu-1)k}\big| \\ &\qquad \leq \big|e^{-k_{j-1}} - e^{-k}\big| + e^{-k_{j-1}}\big|e^{iuk_{j-1}} - e^{iuk}\big| \\ &\qquad \leq C\big(|u\overline{\Delta}| \wedge 1\big)e^{-k_{j-1}}, \qquad k \in [k_{j-1}, k_j], j = 2, \dots, N, \\ &\big|k^p - k^p_{j-1}\big| \\ &\qquad \leq C\overline{\Delta}\big(\overline{\Delta}^{p-1} \vee |k_{j-1}|^{p-1}\big), \qquad p = 1, 2, 3, k \in [k_{j-1}, k_j], j = 2, \dots, N. \end{aligned}$$

We then split the summation into three parts: the first part consists of the summands for which the intervals of integration over k are over regions for which $|e^k - 1| > 1/2$, the second part consists of summands for which the intervals of integration over k are with $|e^k - 1| \leq (\sqrt{T} \vee 2\overline{\Delta})$, and the third part consists of the rest of the summands (see Lemma 5). The regions do not overlap for $T \leq 1/2$ which can be assumed without loss of generality as T shrinks asymptotically. Using Lemma 6 for the first part of the summation (provided T is sufficiently small so that $k_{l,t} > \ln\frac{1}{2}$ and $k_{h,t} < \ln\frac{3}{2}$) and Lemma 5 for the other two parts, we then have

$$\big|\hat{f}^{(a,j)}_{T,3}(u)\big| \leq C_t\big(|u| \vee 1\big)^2\big(|u\overline{\Delta}| \wedge 1\big)T|\ln T|, \qquad j = 0, 1, 2, 3. \tag{8.16}$$

For $\overline{f}^{(b,j)}_{T,3}(u)$, we first consider the case $\alpha > 1/2$. Then we split the summation into three parts: the first part consists of the summands for which the intervals of integration over k are over regions for which k is above 1 in absolute value, the second part consists of summands for which the intervals of integration over k

are with $|k| \le \sqrt{T}$, and the third part consists of the rest of the summands (see Lemma 7). Then, using Lemma 7 and the condition in (4.9), we have for $j = 0, 1, 2, 3$ and $\alpha > 1/2$

$$(8.17)\quad \left|\hat{f}_{T,3}^{(b,j)}(u)\right| \le C_t\left[\left(|u| \vee 1\right)^2\left(\ln|\underline{k}| + \ln\overline{k}\right)\overline{\Delta}T + \alpha_{T,\overline{\Delta}}^{(j)}(u)\right], \qquad |u| \le u_N,$$

where

$$\alpha_{T,\overline{\Delta}}^{(0)}(u) = \left(|u| \vee 1\right)^2 |\ln T| \overline{\Delta}\sqrt{T},$$
$$\alpha_{T,\overline{\Delta}}^{(1)}(u) = \left(|u| \vee 1\right) |\ln T| \overline{\Delta}\sqrt{T},$$
$$\alpha_{T,\overline{\Delta}}^{(2)}(u) = |\ln T| \overline{\Delta}\sqrt{T},$$
$$\alpha_{T,\overline{\Delta}}^{(3)}(u) = |\ln T| \overline{\Delta} T |\ln T|.$$

For the case $\alpha \le 1/2$, we make similar decomposition as for the case $\alpha > 1/2$, except that the second part now consists of summands for which the intervals of integration over k are with $|k| \le 2\overline{\Delta}$. For this part, we apply Lemma 5 while for the other two we use Lemma 7 (we use the second bound in Lemma 7 for the third part). This leads to the same bound as above also to hold for $|\hat{f}_{T,3}^{(b,j)}(u)|$ when $\alpha \le 1/2$.

Finally, applying Lemma 6, provided $\underline{k} < k_{l,t}$ and $\overline{k} > k_{h,t}$ (which will eventually happen as $\underline{k} \downarrow -\infty$ and $\overline{k} \uparrow \infty$ and $k_{l,t}$ and $k_{h,t}$ are $\mathcal{F}_t$-adapted constants that do not change as we add more option data), we have

$$(8.18)\qquad \left|\hat{f}_{T,4}^{(j)}(u)\right| \le C_t\left(|u| \vee 1\right)^2\left(|\underline{k}|^j e^{3\underline{k}} + \overline{k}^j e^{-3\overline{k}}\right)T, \qquad j = 0, 1, 2, 3.$$

Since $u_N^2 T \to 0$ [due to (4.9)] and $\int_{\mathbb{R}} |z|^2 \nu_t(z)\,dz$ is a finite $\mathcal{F}_t$-adapted variable due to Assumption A3, from (8.7), we then have that for sufficiently high value of N, there exists a constant $\varepsilon > 0$ such that we have $\inf_{|u| \le u_N} |f_T(u)| > \varepsilon$. Similarly, using the bounds in (8.11)–(8.18) and the condition in (4.9), we have that

$$(8.19)\qquad \sup_{|u| \le u_N} \sum_{j=0}^{3}\sum_{i=1}^{4} \left|\hat{f}_{T,i}^{(j)}(u)\right| = O_p\left(u_N^2 |\ln T|\left(T \vee \sqrt{T}\overline{\Delta}\right)\right),$$

and, therefore, since $u_N^2 T|\ln T| \to 0$ and $u_N^2 \overline{\Delta} T^{1/2} |\ln T| \to 0$ [due to (4.9)], the probability of the event $\sup_{|u| \le u_N} |\hat{f}_T(u) - f_T(u)| \ge \frac{\varepsilon}{2}$ converges to 0. Therefore,

$$(8.20)\qquad \mathbb{P}(\Omega_n) \to 0 \qquad \text{for } \Omega_n = \left\{\omega : \inf_{|u| \le u_N} \left|\hat{f}_T(u)\right| < \frac{\varepsilon}{2}\right\}.$$

Using the above two results, $u_N^2 T|\ln T| \to 0$ and $u_N^2 \overline{\Delta} T^{1/2}|\ln T| \to 0$, we have on the set Ω_n^c

$$
\begin{aligned}
T\big|\hat{h}_t^*(u) - h^*(u)\big| \\
&\le C_t\big|\hat{f}_T^{(3)}(u) - f_T^{(3)}(u)\big| + C_t\big|f_T^{(3)}(u)\big|\big|\hat{f}_T(u) - f_T(u)\big| \\
&\quad + C_t\big|f_T^{(1)}(u)\big|\big|\hat{f}_T^{(2)}(u) - f_T^{(2)}(u)\big| \\
&\quad + C_t\big|\hat{f}_T^{(1)}(u) - f_T^{(1)}(u)\big|\big|\hat{f}_T^{(2)}(u) - f_T^{(2)}(u)\big| \\
&\quad + C_t\big(\big|f_T^{(1)}(u) f_T^{(2)}(u)\big| + \big|f_T^{(1)}(u)\big|^3\big)\big|\hat{f}_T(u) - f_T(u)\big| \\
&\quad + C_t\big|\hat{f}_T^{(1)}(u) - f_T^{(1)}(u)\big|^3 \\
&\quad + C_t\big|f_T^{(1)}(u)\big|^2\big|\hat{f}_T^{(1)}(u) - f_T^{(1)}(u)\big|.
\end{aligned}
\tag{8.21}
$$

Using this inequality and the bounds in (8.11)–(8.18), we have altogether

$$
\int_{|u|\le u_N} \big(\hat{h}_t^*(u) - h_t^*(u)\big)^2\,du = O_p(A_1 + A_2 + A_3 + A_4),
$$

where we denote

$$
\begin{aligned}
A_1 &= u_N^5 \overline{\Delta}, \\
A_2 &= u_N^5 T \ln^8 T\big(|\underline{k}|^6 e^{-2\underline{k}} \wedge T^{-1/3} + \overline{k}^8\big), \\
A_3 &= u_N^7 \ln^2 T \overline{\Delta}^2 + u_N^5 \overline{\Delta}^2\big(\ln^2|\underline{k}| + \ln^2 \overline{k}\big), \\
A_4 &= u_N^5\big(|\underline{k}|^6 e^{6\underline{k}} + |\overline{k}|^6 e^{-6\overline{k}}\big).
\end{aligned}
$$

From here, the result of the theorem follows by taking into account the restriction on u_N given in (4.9).

8.3. *Auxiliary results.*

LEMMA 1. *Suppose Assumptions A2–A4 hold. Then there exist $\mathcal{F}_t$-adapted $\overline{t} > t$ and C_t such that for $s \in [t, \overline{t} \wedge (t+T)]$, we have*

$$
\mathbb{E}_t^{\mathbb{Q}}\big(e^{x_s} - 1\big)^2 + \mathbb{E}_t^{\mathbb{Q}}\big(e^{-x_s} - 1\big)^2 + \mathbb{E}_t^{\mathbb{Q}}\big(e^{\overline{x}_s} - 1\big)^2 + \mathbb{E}_t^{\mathbb{Q}}\big(e^{\widetilde{x}_s} - 1\big)^2 \le C_t T, \tag{8.22}
$$

$$
\mathbb{E}_t^{\mathbb{Q}}\big|e^{\widetilde{x}_s^d} - 1\big| + \mathbb{E}_t^{\mathbb{Q}}\big|e^{\widetilde{x}_s^d} - 1\big|^2 + \mathbb{E}_t^{\mathbb{Q}}\big|e^{\widetilde{x}_s^d} - 1\big|^3 + \mathbb{E}_t^{\mathbb{Q}}\big|e^{-\widetilde{x}_s^d} - 1\big|^3 \le C_t T, \tag{8.23}
$$

$$
\mathbb{E}_t^{\mathbb{Q}}\big|e^{\overline{x}_s - \widetilde{x}_s} - 1\big|^4 \le C_t T^4. \tag{8.24}
$$

PROOF. We start with the first term in (8.22). By application of Itô's formula,

$$e^{x_s} - 1 = \int_t^s e^{x_u} a_u \, du + \int_t^s e^{x_u} \sigma_u \, dW_u + \frac{1}{2} \int_t^s e^{x_u} \sigma_u^2 \, du + \int_t^s \int_E e^{x_{u-}} \big(e^{\delta^x(u,z)} - 1\big) \widetilde{\mu}^x(du, dz) + \int_t^s \int_E e^{x_{u-}} \big(e^{\delta^x(u,z)} - 1 - \delta^x(u,z)\big) \nu^x(du, dz),$$

where we made use of the fact that $x_t = 0$. From here, applying Hölder's inequality and using the integrability assumptions for x_s and a_s, σ_s and $\nu_s(z)$ in Assumption A3, we get the result in (8.22) for $\mathbb{E}_t^{\mathbb{Q}}(e^{x_s} - 1)^2$. The result for $\mathbb{E}_t^{\mathbb{Q}}(e^{-x_s} - 1)^2$ is showed in exactly the same way.

We continue with (8.23). Using Itô's lemma and the fact that $\widetilde{x}_t^d = 0$, we have

$$e^{\widetilde{x}_s^d} - 1 = \int_t^s e^{\widetilde{x}_{u-}^d} \big(e^{\delta^x(t,z)} - 1\big) \widetilde{\mu}^x(du, dz) + \int_t^s e^{\widetilde{x}_{u-}^d} \big(e^{\delta^x(t,z)} - 1 - \delta^x(t,z)\big) \nu^x(du, dz). \tag{8.25}$$

We can then apply the Burkholder–Davis–Gundy inequality and make use of the fact that $\sup_{s\in[t,t+T]} \mathbb{E}_t^{\mathbb{Q}}(e^{3\widetilde{x}_s^d}) < \infty$ due to our assumption for the Lévy measure in Assumption A3 and Theorem 25.17(iii) of [28]. From here, the result in (8.23) concerning $e^{\widetilde{x}_s^d} - 1$ follows. The part of (8.23) about $e^{-\widetilde{x}_s^d} - 1$ is shown in exactly the same way by making use of the fact that we have $\sup_{s\in[t,t+T]} \mathbb{E}_t^{\mathbb{Q}}(e^{-3\widetilde{x}_s^d}) < \infty$ due to our assumption on the Lévy measure in Assumption A3 and Theorem 25.17(iii) of [28].

Turning next to the bound in (8.24), by application of Itô's formula, we can write for $s \in [t, t+T]$

$$\overline{x}_s - \widetilde{x}_s = \frac{\eta_t}{2}(W_s - W_t)^2 - (s-t)\frac{\eta_t}{2} + \widetilde{\eta}_t \int_t^s (\widetilde{W}_u - \widetilde{W}_t) \, dW_u. \tag{8.26}$$

Then for every p such that $p\eta_t T < 1$, we have $\mathbb{E}_t^{\mathbb{Q}}(e^{\frac{p\eta_t}{2}(W_s - W_t)^2}) < \infty$. In addition, for any $p > 0$ using successive conditioning and the Jensen's inequality, we have

$$\mathbb{E}_t^{\mathbb{Q}}\big(e^{p\widetilde{\eta}_t \int_t^s (\widetilde{W}_u - \widetilde{W}_t) \, dW_u}\big) = \mathbb{E}_t^{\mathbb{Q}}\big(e^{\frac{p^2\widetilde{\eta}_t^2}{2} \int_t^s (W_u - W_t)^2 \, du}\big) \le \frac{1}{s-t} \mathbb{E}_t^{\mathbb{Q}}\bigg(\int_t^s e^{\frac{p^2\widetilde{\eta}_t^2}{2}(s-t)(W_u - W_t)^2} \, du\bigg).$$

The latter integral is finite as soon as $p|\widetilde{\eta}_t|T < 1$. Combining the above bounds and using a first-order Taylor series expansion, we have for T sufficiently small (depending on the values of η_t and $\widetilde{\eta}_t$) the bound in (8.24).

We are left with the bounds in (8.22) that involve $\overline{x}_s$ and $\widetilde{x}_s$. The result involving $\widetilde{x}_s$ follows by applying the algebraic inequality $|xy-1| \le |x-1| + |y-1| + |x-1||y-1|$ for $x, y \in \mathbb{R}$, the bounds in (8.23), the $\mathcal{F}_t$-conditional independence of $\widetilde{x}_s^c$ and $\widetilde{x}_s^d$ for $s \ge t$, as well as the bound $\mathbb{E}_t^{\mathbb{Q}}(e^{\widetilde{x}_s^c} - 1)^2 \le C_t T$ for $s \in [t, t+T]$ which follows by direct evaluation of exponential moments of a normal random variable.

We finish with the bound in (8.22) for $\overline{x}_s$. First, we can write $e^{\overline{x}_s} = e^{\overline{x}_s - \widetilde{x}_s} e^{\widetilde{x}_s^c} e^{\widetilde{x}_s^d}$, use the $\mathcal{F}_t$-conditional independence of $\widetilde{x}_s^d$ from $\overline{x}_s - \widetilde{x}_s$ and $\widetilde{x}_s^c$, the bounds in (8.23) and (8.24), the bound $\mathbb{E}_t^{\mathbb{Q}}(e^{p\widetilde{x}_s}) \le C_t(p)$ for every finite p and $s \ge t$, as well as Hölder's inequality to conclude $\mathbb{E}_t^{\mathbb{Q}}(e^{3\overline{x}_s}) \le C_t$ for $s \in [t, \overline{t} \wedge (t+T)]$. From here, we can apply Itô's formula for $e^{\overline{x}_s} - 1$, use the above result, the integrability assumptions of A3 as well as the Burkholder–Davis–Gundy inequality to get the result. □

LEMMA 2. *Suppose Assumptions A2–A4 hold. There exist $\mathcal{F}_t$-adapted random variables $\overline{t} > t$ and $C_t > 0$ that do not depend on k and T, such that for $T < \overline{t} - t$ we have*

$$O_T(k) \le C_t T \begin{cases} \dfrac{e^{2k}}{e^{-k} - 1} & \text{if } k < 0, \\ \dfrac{1}{e^k - 1} & \text{if } k > 0. \end{cases} \tag{8.27}$$

PROOF. For $k < 0$, we use (8.22) and the following algebraic inequality:

$$(e^k - e^x)^+ \le e^{2k}\big((e^{-x} - 1) - (e^{-k} - 1)\big)^+ \le e^{2k} \frac{|e^{-x} - 1|^2}{|e^{-k} - 1|}, \qquad k < 0, x \in \mathbb{R}.$$

For $k > 0$, we use (8.22) and the following algebraic inequality

$$(e^x - e^k)^+ \le \frac{(e^x - 1)^2}{e^k - 1}, \qquad k > 0, x \in \mathbb{R}.$$ □

LEMMA 3. *Suppose Assumptions A2–A4 hold. There exist $\mathcal{F}_t$-adapted random variables $\overline{t} > t$ and $C_t > 0$ that do not depend on k and T, such that for $T < \overline{t} - t$ we have*

$$\big|O_T(k) - \overline{O}_T(k)\big| \le C_t |\ln T| T^{3/2}. \tag{8.28}$$

PROOF. Throughout the proof, we will assume $T < \overline{t} - t$, where $\overline{t}$ is defined in the statement of the lemma. First, given the definitions of $O_T(k)$ and $\overline{O}_T(k)$, we have

$$\big|O_T(k) - \overline{O}_T(k)\big| \le \mathbb{E}_t^{\mathbb{Q}} \big|e^{x_{t+T}} - e^{\overline{x}_{t+T}}\big|. \tag{8.29}$$

Further, we make use of the following algebraic inequality for $x, y \in \mathbb{R}$ and $\varepsilon > 0$:

$$\begin{aligned}
|e^y - e^x| &\le |e^y - e^x| 1_{\{|x-y|>\varepsilon\}} + e^{x+\varepsilon}|x-y| \\
&\le |e^y - 1| 1_{\{|x-y|>\varepsilon\}} + |e^x - 1| 1_{\{|x-y|>\varepsilon\}} + e^{x+\varepsilon}|x-y| \\
&\le \frac{|e^y-1||x-y|}{\varepsilon} + (\varepsilon^{-1} + e^{\varepsilon})|e^x - 1||x-y| + e^{\varepsilon}|x-y|.
\end{aligned} \tag{8.30}$$

We apply the above inequality with $y = x_{t+T}$ and $x = \overline{x}_{t+T}$ and some $\varepsilon > 0$. To proceed further, we make use of bounds for $\mathbb{E}_t^{\mathbb{Q}}|x_{t+T} - \overline{x}_{t+T}|^p$, $\mathbb{E}_t^{\mathbb{Q}}(e^{x_{t+T}} - 1)^2$ and $\mathbb{E}_t^{\mathbb{Q}}(e^{\overline{x}_{t+T}} - 1)^2$ for powers $p \in [1, 2]$. First, by applying the Burkholder–Davis–Gundy inequality and using our assumption for the processes η and $\widetilde{\eta}$ in A4, we have

$$\mathbb{E}_t^{\mathbb{Q}}\left|\int_t^s (\eta_u - \eta_t)\,dW_u + \int_t^s (\widetilde{\eta}_u - \widetilde{\eta}_t)\,d\widetilde{W}_u\right|^2 \le C_t T^2 \qquad \text{for } s \in [t, t+T].$$

This bound, another application of the Burkholder–Davis–Gundy inequality and our assumption for the process a in A4 and for δ^σ in A3, leads to the following inequality for $p \in [1, 2]$ and $s \in [t, t+T]$:

$$\mathbb{E}_t^{\mathbb{Q}}\left|x_s^c - \overline{x}_s^c - \int_t^s \int_t^u \int_{\mathbb{R}} \delta^\sigma(v, z)\widetilde{\mu}^\sigma(dv, dz)\,dW_u\right|^p \le C_t(p) T^{\frac{3p}{2}}. \tag{8.31}$$

Next, using integration by parts, we can write

$$\begin{aligned}
&\int_t^s \int_t^u \int_{\mathbb{R}} \delta^\sigma(v, z)\widetilde{\mu}^\sigma(dv, dz)\,dW_u \\
&\quad = (W_s - W_t)\int_t^s \int_{\mathbb{R}} \delta^\sigma(u, z)\widetilde{\mu}^\sigma(du, dz) \\
&\qquad - \int_t^s \int_{\mathbb{R}} (W_u - W_t)\delta^\sigma(u, z)\widetilde{\mu}^\sigma(du, dz),
\end{aligned}$$

and we can further split

$$\begin{aligned}
&|W_s - W_t|\left|\int_t^s \int_{\mathbb{R}} \delta^\sigma(u, z)\widetilde{\mu}^\sigma(du, dz)\right| \\
&\quad \le \sqrt{|s-t|}\,\big|\ln|s-t|\big|\left|\int_t^s \int_{\mathbb{R}} \delta^\sigma(u, z)\widetilde{\mu}^\sigma(du, dz)\right| \\
&\qquad + |W_s - W_t|\left|\int_t^s \int_{\mathbb{R}} \delta^\sigma(u, z)\widetilde{\mu}^\sigma(du, dz)\right| 1_{\left\{\frac{|W_s - W_t|}{\sqrt{|s-t|}|\ln|s-t||} > 1\right\}}.
\end{aligned}$$

In addition, by the applying the Burkholder–Davis–Gundy inequality and the algebraic inequality $|\sum_i x_i|^\alpha \le \sum_i |x_i|^\alpha$ for a sequence of reals $\{x_i\}_{i \ge 1}$ and some

$\alpha \in (0, 1]$, we have

$$\mathbb{E}_t^{\mathbb{Q}}\left|\int_t^s \int_{\mathbb{R}} (W_u - W_t)^q \delta^\sigma(u,z)\widetilde{\mu}^\sigma(du,dz)\right|^p \leq C_t(p)\mathbb{E}_t^{\mathbb{Q}}\left|\int_t^s |W_u - W_t|^{qp} \int_{\mathbb{R}} \left|\delta^\sigma(u,z)\right|^p dz\,du\right|,$$

for $p \in [1,2]$ and $q = 0, 1$. Successive application of the Burkholder–Davis–Gundy inequality and the integrability conditions for δ^σ yield

$$\mathbb{E}_t^{\mathbb{Q}}\left|\int_t^s \int_{\mathbb{R}} \delta^\sigma(u,z)\widetilde{\mu}^\sigma(du,dz)\right|^p \leq C_t(p)|s-t|, \qquad p \in [1,3].$$

Combining the above results and using Hölder's inequality, we have for $s \in [t, t+T]$ and $p \in [1,2]$

$$\mathbb{E}_t^{\mathbb{Q}}\left|\int_t^s \int_t^u \int_{\mathbb{R}} \delta^\sigma(v,z)\widetilde{\mu}^\sigma(dv,dz)\,dW_u\right|^p \leq C_t(p)|\ln T|^p T^{1+\frac{p}{2}}. \tag{8.32}$$

Next, applying the Burkholder–Davis–Gundy inequality and the algebraic inequality $|\sum_i x_i|^q \leq \sum_i |x_i|^q$, for a sequence of reals $\{x_i\}_{i\geq 1}$ and $q \in (0,1]$, and recalling the definition of δ^x yields

$$\begin{aligned}\mathbb{E}_t^{\mathbb{Q}}|x_s^d - \overline{x}_s^d|^p &\leq C_t(p)\mathbb{E}_t^{\mathbb{Q}}\left|\int_t^s \int_E \left|\delta^x(u,z) - \delta^x(t,z)\right|^p \nu^x(du,dz)\right| \\ &\leq C_t(p)\mathbb{E}_t^{\mathbb{Q}}\left|\int_t^s \int_{\mathbb{R}} |z|^p \left|\nu_s(z) - \nu_t(z)\right| dz\right|, \qquad p \in [1,2].\end{aligned} \tag{8.33}$$

Therefore, using the assumption for $\nu_t(z)$ in A4, we have

$$\mathbb{E}_t^{\mathbb{Q}}|x_s^d - \overline{x}_s^d|^p \leq C_t(p)T^{\frac{3}{2}} \qquad \forall s \in [t, t+T], p \in [1,2]. \tag{8.34}$$

We next denote $\xi_{s,1} = \int_t^s e^{x_u}\sigma_u\,dW_u$ and $\xi_{s,2} = \int_t^s \int_E e^{x_{u-}}(e^{\delta^x(u,z)} - 1)\widetilde{\mu}^x(du,dz)$ for $s \geq t$. By using the Cauchy–Schwarz inequality, the bound in (8.34), the same steps as for the proof of (8.22) involving $\mathbb{E}_t^{\mathbb{Q}}(e^{x_s} - 1)^2$, we get

$$\mathbb{E}_t^{\mathbb{Q}}\left(\left|e^{x_s} - 1 - \xi_{s,1} - \xi_{s,2}\right|\left|x_s^d - \overline{x}_s^d\right|\right) \leq C_t T^{\frac{7}{4}}.$$

Next, we can further split (recall that $x_t = 0$)

$$\xi_{s,1} = \int_t^s \left(e^{x_u}\sigma_u - \sigma_t\right) dW_u + \int_t^s \sigma_t\,dW_u \equiv \xi_{s,1}^a + \xi_{s,1}^b,$$

and analyze separately $\xi_{s,1}^a(x_s^d - \overline{x}_s^d)$ and $\xi_{s,1}^b(x_s^d - \overline{x}_s^d)$. For $\xi_{s,1}^b(x_s^d - \overline{x}_s^d)$, we can apply Hölder's inequality and (8.34). For $\xi_{s,1}^a(x_s^d - \overline{x}_s^d)$, we can apply Hölder's inequality and use the bounds in (8.34) and (8.22) as well as Assumptions A3 and A4. Altogether

$$\mathbb{E}_t^{\mathbb{Q}}\left|\xi_{s,1}\left(x_s^d - \overline{x}_s^d\right)\right| \leq C_t T^{\frac{3}{2}} \qquad \forall s \in [t, t+T].$$

Next, using integration by parts,

$$\xi_{s,2}(x_s^d - \overline{x}_s^d) = \int_t^s \int_E e^{x_{u-}}(e^{\delta^x(u,z)} - 1)(\delta^x(u,z) - \delta^x(t,z))\mu(du,dz) + \int_t^s \int_E e^{x_{u-}}(e^{\delta^x(u,z)} - 1)(x_{u-}^d - \overline{x}_{u-}^d)\widetilde{\mu}(du,dz) + \int_t^s \int_E (\delta^x(u,z) - \delta^x(t,z))\xi_{u-,2}\widetilde{\mu}(du,dz).$$

Using the bound in (8.22), the Cauchy–Schwarz inequality as well as Assumptions A3 and A4, we have

$$\mathbb{E}_t^{\mathbb{Q}}\left(\int_t^s \int_E e^{x_{u-}}|e^{\delta^x(u,z)} - 1||\delta^x(u,z) - \delta^x(t,z)|\mu(du,dz)\right) \le \mathbb{E}_t^{\mathbb{Q}}\left(\int_t^s e^{x_u}\int_{\mathbb{R}} e^z|z||\nu_u(z) - \nu_t(z)|\,du\right) \le C_t T^{\frac{3}{2}}.$$

We have a similar bound for the other two terms in the decomposition of $\xi_{s,2}(x_s^d - \overline{x}_s^d)$ by using Hölder's inequality, the bound in (8.34) as well as Assumptions A3 and A4. Thus, altogether

$$\mathbb{E}_t^{\mathbb{Q}}|\xi_{s,2}(x_s^d - \overline{x}_s^d)| \le C_t T^{\frac{3}{2}}.$$

Combining the above bounds, we have

$$\mathbb{E}_t^{\mathbb{Q}}|(e^{x_s} - 1)(x_s^d - \overline{x}_s^d)| \le C_t T^{\frac{3}{2}}. \tag{8.35}$$

Exactly the same analysis leads to

$$\mathbb{E}_t^{\mathbb{Q}}|(e^{\overline{x}_s} - 1)(x_s^d - \overline{x}_s^d)| \le C_t T^{\frac{3}{2}}. \tag{8.36}$$

Combining the bounds in (8.31), (8.32), (8.34), (8.35) and (8.36) with the one in (8.22), and using the Cauchy–Schwarz inequality and (8.30), we get the bound of the lemma to be proved. □

LEMMA 4. *Suppose Assumptions A2–A4 hold. There exist $\mathcal{F}_t$-adapted random variables $\overline{t} > t$ and $C_t > 0$ that do not depend on k and T, such that for $T < \overline{t} - t$ we have*

$$|\overline{O}_T(k) - \widetilde{O}_T(k)| \le C_t\left(T^{\frac{3}{2}} \vee \left(\frac{T^{3/2}}{|e^k - 1|} \wedge T\right)\right). \tag{8.37}$$

PROOF. Throughout the proof, we will assume $T < \overline{t} - t$, where $\overline{t}$ is defined in the statement of the lemma. For X, Y being reals and K being a nonnegative number, we have the following algebraic inequality:

$$|(Y - K)^+ - (X - K)^+| \le |Y - X|(1_{\{|Y-X| > K/2\}} + 1_{\{X \ge K/2\}}),$$

and similarly for X, Y being reals and K being a nonpositive number, we have

$$\left|(K-Y)^+ - (K-X)^+\right| \leq |Y-X|(1_{\{|Y-X|>-K/2\}} + 1_{\{X\leq K/2\}}).$$

Applying these inequalities [note that $\widetilde{O}_T(k) = \mathbb{E}_t^{\mathbb{Q}}(e^{\widetilde{x}_{t+T}} - e^k)^+$ if $e^k > 1$ and $\widetilde{O}_T(k) = \mathbb{E}_t^{\mathbb{Q}}(e^k - e^{\widetilde{x}_{t+T}})^+$ if $e^k \leq 1$], we have

$$\begin{aligned}&\left|\overline{O}_T(k) - \widetilde{O}_T(k)\right| \\ &\quad \leq \mathbb{E}_t^{\mathbb{Q}}\left[\left|e^{\overline{x}_{t+T}} - e^{\widetilde{x}_{t+T}}\right|(1_{\{|e^{\overline{x}_{t+T}} - e^{\widetilde{x}_{t+T}}|>|e^k-1|/2\}} + 1_{\{|e^{\widetilde{x}_{t+T}}-1|\geq|e^k-1|/2\}})\right].\end{aligned}$$

Then, if $\sqrt{T} > |e^k - 1|$, we can bound

$$\left|\overline{O}_T(k) - \widetilde{O}_T(k)\right| \leq \mathbb{E}_t^{\mathbb{Q}}\left|e^{\overline{x}_{t+T}} - e^{\widetilde{x}_{t+T}}\right|,$$

and from here the result in (8.37) in the case $\sqrt{T} > |e^k - 1|$ follows from the results in (8.23) and (8.24), the Cauchy–Schwarz inequality as well as the $\mathcal{F}_t$-conditional independence of $\widetilde{x}_{t+T}^d$ from $\overline{x}_{t+T} - \widetilde{x}_{t+T}$.

If $\sqrt{T} \leq |e^k - 1|$, first using the algebraic inequality $|xy - 1| \leq |x - 1| + |y - 1| + |x - 1||y - 1|$, for some reals x and y, we can bound

$$\begin{aligned}&\left|e^{\overline{x}_{t+T}} - e^{\widetilde{x}_{t+T}}\right| \\ &\quad \leq \left|e^{\overline{x}_{t+T}-\widetilde{x}_{t+T}} - 1\right|\left(\left|e^{\widetilde{x}_{t+T}^c} - 1\right| + \left|e^{\widetilde{x}_{t+T}^d} - 1\right| + \left|e^{\widetilde{x}_{t+T}^c} - 1\right|\left|e^{\widetilde{x}_{t+T}^d} - 1\right| + 1\right).\end{aligned}$$

From here, if $\sqrt{T} \leq |e^k - 1|$, we have

$$\begin{aligned}\left|\overline{O}_T(k) - \widetilde{O}_T(k)\right| \leq{}& \frac{2}{|e^k-1|}\mathbb{E}_t^{\mathbb{Q}}\left[\left|e^{\overline{x}_{t+T}-\widetilde{x}_{t+T}} - 1\right|^2 e^{\widetilde{x}_{t+T}}\right] \\ &+ \left(2 + \frac{2}{|e^k-1|}\right) \\ &\times \mathbb{E}_t^{\mathbb{Q}}\left[\left|e^{\overline{x}_{t+T}-\widetilde{x}_{t+T}} - 1\right|\left(\left|e^{\widetilde{x}_{t+T}^c} - 1\right| + \left|e^{\widetilde{x}_{t+T}^d} - 1\right| \right.\right. \\ &\left.\left.+ \left|e^{\widetilde{x}_{t+T}^c} - 1\right|\left|e^{\widetilde{x}_{t+T}^d} - 1\right|\right)\right].\end{aligned} \tag{8.38}$$

To proceed further, we first note that, since $\mathbb{E}_t^{\mathbb{Q}}(e^{p\widetilde{x}_{t+T}^c}) < \infty$ for every finite p, by using a first-order Taylor series expansion, we have

$$\mathbb{E}_t^{\mathbb{Q}}\left|e^{\widetilde{x}_{t+T}^c} - 1\right|^p \leq C_t(p)T^{p/2} \qquad \forall p \geq 1.$$

Combining the above bound with the ones in (8.23) and (8.24) as well as (8.38), and using the Cauchy–Schwarz inequality as well as the fact that $\widetilde{x}_{t+T}^d$ is $\mathcal{F}_t$-conditionally independent from $\widetilde{x}_{t+T}^c$ and $\overline{x}_{t+T} - \widetilde{x}_{t+T}$, we get the result of the lemma in the case $\sqrt{T} \leq |e^k - 1|$. □

LEMMA 5. *Suppose Assumptions A2–A4 hold. There exist $\mathcal{F}_t$-adapted random variables $\bar{t} > t$ and $C_t > 0$ that do not depend on k and T, such that for $T < \bar{t} - t$ we have*

$$\widetilde{O}_T(k) \le C_t\left(\sqrt{T} \wedge \frac{T}{|e^k - 1|}\right). \tag{8.39}$$

PROOF. We look only at the case $k > 0$, with the case $k \le 0$ being proven in an analogous way. For $k > 0$, we have

$$\begin{aligned}\widetilde{O}_T(k) &\le \mathbb{E}_t^{\mathbb{Q}}\big[(e^{\widetilde{x}_{t+T}} - 1)1_{\{e^{\widetilde{x}_{t+T}}-1>e^k-1\}}\big] \\ &\le \mathbb{E}_t^{\mathbb{Q}}\left(\left|e^{\widetilde{x}_{t+T}} - 1\right|\left(\frac{|e^{\widetilde{x}_{t+T}} - 1|}{e^k - 1} \wedge 1\right)\right) \\ &\le \frac{\mathbb{E}_t^{\mathbb{Q}}|e^{\widetilde{x}_{t+T}} - 1|^2}{e^k - 1} \wedge \mathbb{E}_t^{\mathbb{Q}}\left|e^{\widetilde{x}_{t+T}} - 1\right|.\end{aligned}$$

From here, the result to be proved follows by making use of $\mathbb{E}_t^{\mathbb{Q}}|e^{\widetilde{x}_{t+T}} - 1|^2 \le C_t T$ and $\mathbb{E}_t^{\mathbb{Q}}|e^{\widetilde{x}_{t+T}} - 1| \le C_t\sqrt{T}$. The first of these two inequalities is shown in Lemma 1 and the second one follows from the first one and an application of Jensen's inequality. □

LEMMA 6. *Suppose Assumptions A2–A4 hold and denote $k_{l,t} = -\sigma_t \times \sqrt{T}|\ln T|$ and $k_{h,t} = \sigma_t\sqrt{T}|\ln T|$. Then there exist $\mathcal{F}_t$-adapted random variables $C_t > 0$ and $\bar{t} > t$ that do not depend on k and T, such that for $T < \bar{t} - t$, we have*

$$\begin{cases} k < k_{l,t} \Longrightarrow \widetilde{O}_T(k) \le C_t \dfrac{e^{2k}}{(e^{-k+k_{l,t}} - 1)^2} T, \\ k > k_{h,t} \Longrightarrow \widetilde{O}_T(k) \le C_t \dfrac{1}{(e^{k-k_{h,t}} - 1)^2} T. \end{cases} \tag{8.40}$$

PROOF. Throughout the proof, we will assume $T < \bar{t} - t$, where $\bar{t}$ is defined in the statement of the lemma. We introduce the set $\mathcal{C}_t = \{\omega : |\widetilde{x}^c_{t+T}| \le \sigma_t\sqrt{T}|\ln T|\}$. Applying Chebyshev's and Hölder's inequalities and using (8.23) as well as the $\mathcal{F}_t$-conditional independence of $\widetilde{x}^c_{t+T}$ and $\widetilde{x}^d_{t+T}$, we have

$$\mathbb{E}_t^{\mathbb{Q}}\big[(e^{\widetilde{x}_{t+T}} - e^k)^+ 1_{\{\mathcal{C}_t^c\}}\big] \le 2e^{-2k}\mathbb{E}_t^{\mathbb{Q}}\big[e^{3\widetilde{x}_{t+T}} 1_{\{\mathcal{C}_t^c\}}\big] \le C_t e^{-2k} T. \tag{8.41}$$

Next, taking into account the definition of the set $\mathcal{C}_t$, we have

$$\mathbb{E}_t^{\mathbb{Q}}\big[(e^{\widetilde{x}_{t+T}} - e^k)^+ 1_{\{\mathcal{C}_t\}}\big] \le e^{\sigma_t\sqrt{T}|\ln T|}\mathbb{E}_t^{\mathbb{Q}}\big(e^{\widetilde{x}^d_{t+T}} - e^{k-\sigma_t\sqrt{T}|\ln T|}\big)^+.$$

For $k > k_{h,t}$, we have $e^{k-\sigma_t\sqrt{T}|\ln T|} > 1$ and, therefore, by an application of Chebyshev's inequality and the preceding inequality, we have

$$k > k_{h,t} \Longrightarrow \mathbb{E}_t^{\mathbb{Q}}\big[(e^{\widetilde{x}_{t+T}} - e^k)^+ 1_{\{\mathcal{C}_t\}}\big] \le \frac{C_t}{(e^{k-k_{h,t}} - 1)^2}\mathbb{E}_t^{\mathbb{Q}}\left|e^{\widetilde{x}^d_{t+T}} - 1\right|^3.$$

We can then use (8.23) to get altogether

$$k > k_{h,t} \Longrightarrow \mathbb{E}_t^{\mathbb{Q}}\big[\big(e^{\widetilde{x}_{t+T}} - e^k\big)^+ 1_{\{\mathcal{C}_t\}}\big] \le \frac{C_t}{(e^{k-k_{h,t}} - 1)^2} T. \tag{8.42}$$

From here, the second bound in (8.40) easily follows. For the first bound, exactly as before, using (8.23), we first have

$$\mathbb{E}_t^{\mathbb{Q}}\big[\big(e^k - e^{\widetilde{x}_{t+T}}\big)^+ 1_{\{\mathcal{C}_t^c\}}\big] \le 2e^{4k}\mathbb{E}_t^{\mathbb{Q}}\big[\big(e^{-3\widetilde{x}_{t+T}}\big) 1_{\{\mathcal{C}_t^c\}}\big] \le C_t e^{4k} T. \tag{8.43}$$

We further have

$$\begin{aligned}\mathbb{E}_t^{\mathbb{Q}}\big[\big(e^k - e^{\widetilde{x}_{t+T}}\big)^+ 1_{\{\mathcal{C}_t\}}\big] &\le e^{\sigma_t\sqrt{T}|\ln T|}\mathbb{E}_t^{\mathbb{Q}}\big(e^{k+\sigma_t\sqrt{T}|\ln T|} - e^{\widetilde{x}^d_{t+T}}\big)^+ \\ &\le e^{2k+3\sigma_t\sqrt{T}|\ln T|}\mathbb{E}_t^{\mathbb{Q}}\big(e^{-\widetilde{x}^d_{t+T}} - e^{-k-\sigma_t\sqrt{T}|\ln T|}\big)^+.\end{aligned}$$

From here, by using (8.23), we get

$$\begin{aligned}k < k_{l,t} \Longrightarrow \mathbb{E}_t^{\mathbb{Q}}\big[\big(e^k - e^{\widetilde{x}_{t+T}}\big)^+ 1_{\{\mathcal{C}_t\}}\big] &\le \frac{2e^{2k+3\sigma_t\sqrt{T}\ln T|}}{(e^{-k+k_{l,t}} - 1)^2}\mathbb{E}_t^{\mathbb{Q}}\big|e^{-\widetilde{x}^d_{t+T}} - 1\big|^3 \\ &\le C_t \frac{e^{2k}}{(e^{-k+k_{l,t}} - 1)^2} T.\end{aligned} \tag{8.44}$$

Combining the above bounds, we have the first result in (8.40). □

LEMMA 7. *Suppose Assumptions A2–A4 hold. For $k_1 < k_2 < 0$ or $k_1 > k_2 > 0$, we have*

$$\begin{aligned}&\big|\widetilde{O}_T(k_1) - \widetilde{O}_T(k_2)\big| \\ &\quad \le C_t\bigg[\bigg(\frac{T}{k_2^2} \wedge 1\bigg)1_{\{|k_2|\le 1\}} + \frac{T}{k_2^4}1_{\{|k_2|>1\}}\bigg]\big|e^{k_1} - e^{k_2}\big|,\end{aligned} \tag{8.45}$$

and further

$$\big|\widetilde{O}_T(k_1) - \widetilde{O}_T(k_2)\big| \le C_t\bigg(\frac{T}{|k_2|} + e^{-\frac{k_2^2}{12\sigma_t^2 T}}\bigg)\big|e^{k_1} - e^{k_2}\big| \qquad \text{if } |k_2| \le 1, \tag{8.46}$$

where C_t is an $\mathcal{F}_t$-adapted random variable that does not depend on k_1, k_2 and T.

PROOF. We have the following algebraic inequalities:

$$\begin{aligned}\big|(X - K_1)^+ - (X - K_2)^+\big| &\le |K_1 - K_2| 1_{\{X > K_2\}} \qquad \forall X, K_1 \ge K_2, \\ \big|(K_1 - X)^+ - (K_2 - X)^+\big| &\le |K_1 - K_2| 1_{\{X < K_2\}} \qquad \forall X, K_1 \le K_2.\end{aligned}$$

Therefore, to prove the claim it suffices to evaluate for any $k > 0$ the probability $\mathbb{Q}_t(|\widetilde{x}_{t+T}| > k)$. Using Chebychev's and the Burkholder–Davis–Gundy inequalities, and upon noting that $\widetilde{x}_t = 0$ and taking into account the integrability conditions on the processes a_t, σ_t and ν_t, we have

$$\begin{cases} \mathbb{Q}_t(|\widetilde{x}^d_{t+T} + a_t T| > k) \le C_t \dfrac{T}{k} & \text{if } k \in (0, 1], \\ \mathbb{Q}_t(|\widetilde{x}_{t+T}| > k) \le C_t \dfrac{T}{k^4} & \text{if } k > 1. \end{cases}$$

We can further write

$$\mathbb{Q}_t(\sigma_t|W_{t+T} - W_t| > k) = \mathbb{Q}_t(e^{|W_{t+T}-W_t|^2/(3T)} > e^{k^2/(3\sigma_t^2 T)}) \le Ce^{-k^2/(3\sigma_t^2 T)} \qquad \text{for } k > 0.$$

From here, the first result of the lemma follows by applying the first bound with $k = |k_2|$ and the second one follows by applying both inequalities with $k = |k_2|/2$. □

LEMMA 8. *Suppose Assumptions A2–A6 hold and in addition* $\overline{\Delta} \asymp T^{\alpha}$, $\overline{K} \asymp T^{-\beta}$, $\underline{K} \asymp T^{\gamma}$ *for some* $\alpha > 0$, $\beta > 0$ *and* $\gamma > 0$. *We then have*

$$\sup_{u\in\mathbb{R}} \left| \sum_{j=2}^{N} e^{(iu-1)k_{j-1}} k_{j-1}^p \varepsilon_{j-1} \Delta_j \right| = O_p(T|\ln T|), \qquad p = 0, 1, 2, 3, \tag{8.47}$$

$$\sup_{u\in\mathbb{R}} \mathbb{E}\left(\left| \sum_{j=2}^{N} e^{(iu-1)k_{j-1}} \varepsilon_{j-1} \Delta_j \right|^2 \middle| \mathcal{F}^{(0)} \right) = O_p(T^{3/2}\overline{\Delta}), \tag{8.48}$$

$$\sup_{u\in\mathbb{R}} \mathbb{E}\left(\left| \sum_{j=2}^{N} e^{(iu-1)k_{j-1}} k_{j-1}^p \varepsilon_{j-1} \Delta_j \right|^2 \middle| \mathcal{F}^{(0)} \right) = O_p(T^2\overline{\Delta}), \qquad p = 1, 2, 3. \tag{8.49}$$

PROOF. We start with (8.47). We have

$$\sup_{u\in\mathbb{R}} \left| \sum_{j=2}^{N} e^{(iu-1)k_{j-1}} k_{j-1}^p \varepsilon_{j-1} \Delta_j \right| \le \sum_{j=2}^{N} e^{-k_{j-1}} |k_{j-1}|^p |\varepsilon_{j-1}| \Delta_j.$$

Furthermore, using Lemmas 2–6 and provided $T < \overline{t} - t$ for some $\mathcal{F}_t$-adapted $\overline{t} > t$, we have

$$\mathbb{E}\left(\sum_{j=2}^{N} |k_{j-1}|^p |\varepsilon_{j-1}| \Delta_j \middle| \mathcal{F}^{(0)} \right) \le C_t T |\ln T|.$$

Combining the above two results, we get (8.47). We turn next to (8.48)–(8.49). Using the assumption for the $\mathcal{F}^{(0)}$-conditional variance of the errors ε_j in Assumption A6 as well as Assumption A5 for the mesh of the log-strike grid, we have

$$\mathbb{E}\left(\left|\sum_{j=2}^{N} e^{(iu-1)k_{j-1}} k_{j-1}^p \varepsilon_{j-1}\Delta_j\right|^2 \middle| \mathcal{F}^{(0)}\right) \leq C_t \sup_{i=1,\dots,N} \sigma_{t,i}^2 \sum_{j=1}^{N-1} k_j^{2p} e^{-2k_j} O_T(k_j)^2 \overline{\Delta}^2.$$

Therefore, since $\sup_{i=1,\dots,N} \sigma_{t,i}^2$ is $O_p(1)$ and using the results of Lemmas 2–6, provided $T < \overline{t} - t$ for some $\mathcal{F}_t$-adapted $\overline{t} > t$, we have

$$\mathbb{E}\left(\left|\sum_{j=2}^{N} e^{(iu-1)k_{j-1}} \varepsilon_{j-1}\Delta_j\right|^2 \middle| \mathcal{F}^{(0)}\right) \leq C_t T^{3/2}\overline{\Delta},$$

$$\mathbb{E}\left(\left|\sum_{j=2}^{N} e^{(iu-1)k_{j-1}} k_{j-1}^p \varepsilon_{j-1}\Delta_j\right|^2 \middle| \mathcal{F}^{(0)}\right) \leq C_t T^2\overline{\Delta}, \qquad p = 1, 2, 3.$$

From here, the bounds in (8.48)–(8.49) follow. □

Acknowledgments. We would like to thank the Editor, Associate Editor and two referees for many helpful comments and suggestions which have led to significant improvements. We also thank Nicola Fusari for providing the option pricing codes used in the Monte Carlo section.

SUPPLEMENTARY MATERIAL

Supplement to "Nonparametric implied Lévy densities" (DOI: 10.1214/18-AOS1703SUPP; .pdf). The supplement contains the following items: (1) limit results for the integrated squared error of the nonparametric estimator, (2) lower bounds for the minimax risk of recovering Lévy density from noisy option data with heteroskedastic Gaussian observation errors, and (3) alternative Lévy density estimator based on the second derivatives of the characteristic function of the asset return estimated from the option data.

REFERENCES

[1] ANDERSEN, T., FUSARI, N. and TODOROV, V. (2017). The pricing of short-term market risk: Evidence from weekly options. *J. Finance* **72** 1335–1386.

[2] BELOMESTNY, D. (2011). Spectral estimation of the Lévy density in partially observed affine models. *Stochastic Process. Appl.* **121** 1217–1244. MR2794974

[3] BELOMESTNY, D. (2011). Statistical inference for time-changed Lévy processes via composite characteristic function estimation. *Ann. Statist.* **39** 2205–2242. MR2893866

[4] BELOMESTNY, D. and REISS, M. (2006). Spectral calibration of exponential Lévy models. *Finance Stoch*. **10** 449–474. MR2276314

[5] BELOMESTNY, D. and REISS, M. (2015). *Lévy Matters IV: Estimation for Discretely Observed Lévy Processes*. *Lecture Notes in Mathematics* **2128**. Springer, Berlin.

[6] BENTATA, A. and CONT, R. (2012). Short-time asymptotics for marginal distributions of semimartingales. Working paper.

[7] CARR, P., GEMAN, H., MADAN, D. and YOR, M. (2002). The fine structure of asset returns: An empirical investigation. *J. Bus*. **75** 305–332.

[8] CARR, P. and MADAN, D. (2001). Optimal positioning in derivative securities. *Quant. Finance* **1** 19–37. MR1810014

[9] COMTE, F. and GENON-CATALOT, V. (2009). Nonparametric estimation for pure jump Lévy processes based on high frequency data. *Stochastic Process. Appl*. **119** 4088–4123. MR2565560

[10] COMTE, F. and GENON-CATALOT, V. (2011). Estimation for Lévy processes from high frequency data within a long time interval. *Ann. Statist*. **39** 803–837. MR2816339

[11] CONT, R. and TANKOV, P. (2004). Nonparametric calibration of jump-diffusion option pricing models. *J. Comput. Finance* **7** 1–49.

[12] DUFFIE, D. (2001). *Dynamic Asset Pricing Theory*, 3rd ed. Princeton Univ. Press, Princeton, NJ.

[13] DUFFIE, D., PAN, J. and SINGLETON, K. (2000). Transform analysis and asset pricing for affine jump-diffusions. *Econometrica* **68** 1343–1376. MR1793362

[14] FIGUEROA-LÓPEZ, J. E. (2009). Nonparametric estimation of time-changed Lévy models under high-frequency data. *Adv. in Appl. Probab.* **41** 1161–1188. MR2663241

[15] FIGUEROA-LÓPEZ, J. E., GONG, R. and HOUDRÉ, C. (2012). Small-time expansions of the distributions, densities, and option prices of stochastic volatility models with Lévy jumps. *Stochastic Process. Appl*. **122** 1808–1839. MR2914773

[16] FIGUEROA-LÓPEZ, J. E. and ÓLAFSSON, S. (2016). Short-term asymptotics for the implied volatility skew under a stochastic volatility model with Lévy jumps. *Finance Stoch*. **20** 973–1020. MR3551858

[17] FIGUEROA-LÓPEZ, J. E. and ÓLAFSSON, S. (2016). Short-time expansions for close-to-the-money options under a Lévy jump model with stochastic volatility. *Finance Stoch*. **20** 219–265. MR3441292

[18] GUGUSHVILI, S. (2009). Nonparametric estimation of the characteristic triplet of a discretely observed Lévy process. *J. Nonparametr. Stat*. **21** 321–343. MR2530929

[19] JACOD, J. and PROTTER, P. (2012). *Discretization of Processes*. *Stochastic Modelling and Applied Probability* **67**. Springer, Heidelberg. MR2859096

[20] JACOD, J. and TODOROV, V. (2014). Efficient estimation of integrated volatility in presence of infinite variation jumps. *Ann. Statist*. **42** 1029–1069. MR3210995

[21] JONGBLOED, G., VAN DER MEULEN, F. H. and VAN DER VAART, A. W. (2005). Nonparametric inference for Lévy-driven Ornstein–Uhlenbeck processes. *Bernoulli* **11** 759–791. MR2172840

[22] KAPPUS, J. and REISS, M. (2010). Estimation of the characteristics of a Lévy process observed at arbitrary frequency. *Stat. Neerl*. **64** 314–328. MR2683463

[23] MERTON, R. (1976). Option pricing when underlying asset returns are discontinuous. *J. Financ. Econ*. **3** 125–144.

[24] MUHLE-KARBE, J. and NUTZ, M. (2011). Small-time asymptotics of option prices and first absolute moments. *J. Appl. Probab.* **48** 1003–1020. MR2896664

[25] NEUMANN, M. H. and REISS, M. (2009). Nonparametric estimation for Lévy processes from low-frequency observations. *Bernoulli* **15** 223–248. MR2546805

[26] NICKL, R., REISS, M., SÖHL, J. and TRABS, M. (2016). High-frequency Donsker theorems for Lévy measures. *Probab. Theory Related Fields* **164** 61–108. MR3449386

[27] QIN, L. and TODOROV, V. (2019). Supplement to "Nonparametric implied Lévy densities." DOI:10.1214/18-AOS1703SUPP.

[28] SATO, K. (1999). *Lévy Processes and Infinitely Divisible Distributions*. *Cambridge Studies in Advanced Mathematics* **68**. Cambridge Univ. Press, Cambridge. MR1739520

[29] SÖHL, J. (2014). Confidence sets in nonparametric calibration of exponential Lévy models. *Finance Stoch*. **18** 617–649. MR3232018

[30] SÖHL, J. and TRABS, M. (2014). Option calibration of exponential Lévy models: Confidence intervals and empirical results. *J. Comput. Finance* **18** 91–119.

[31] TRABS, M. (2014). Calibration of self-decomposable Lévy models. *Bernoulli* **20** 109–140. MR3160575

[32] TRABS, M. (2015). Quantile estimation for Lévy measures. *Stochastic Process. Appl*. **125** 3484–3521. MR3357617

DEPARTMENT OF INDUSTRIAL ENGINEERING
NORTHWESTERN UNIVERSITY
EVANSTON, ILLINOIS 60208
USA
E-MAIL: likuanqin2012@u.northwestern.edu

DEPARTMENT OF FINANCE
NORTHWESTERN UNIVERSITY
EVANSTON, ILLINOIS 60208
USA
E-MAIL: v-todorov@northwestern.edu