Econometric Theory, **35**, 2019, 901–942.
doi:10.1017/S0266466618000373

INFERENCE FOR OPTION PANELS IN PURE-JUMP SETTINGS

TORBEN G. ANDERSEN
Northwestern University

NICOLA FUSARI
The Johns Hopkins University Carey Business School

VIKTOR TODOROV AND RASMUS T. VARNESKOV
Northwestern University

We develop parametric inference procedures for large panels of noisy option data in a setting, where the underlying process is of pure-jump type, i.e., evolves only through a sequence of jumps. The panel consists of options written on the underlying asset with a (different) set of strikes and maturities available across the observation times. We consider an asymptotic setting in which the cross-sectional dimension of the panel increases to infinity, while the time span remains fixed. The information set is augmented with high-frequency data on the underlying asset. Given a parametric specification for the risk-neutral asset return dynamics, the option prices are nonlinear functions of a time-invariant parameter vector and a time-varying latent state vector (or factors). Furthermore, no-arbitrage restrictions impose a direct link between some of the quantities that may be identified from the return and option data. These include the so-called jump activity index as well as the time-varying jump intensity. We propose penalized least squares estimation in which we minimize the L_2 distance between observed and model-implied options. In addition, we penalize for the deviation of the model-implied quantities from their model-free counterparts, obtained from the high-frequency returns. We derive the joint asymptotic distribution of the parameters, factor realizations and high-frequency measures, which is mixed Gaussian. The different components of the parameter and state vector exhibit different rates of convergence, depending on the relative (asymptotic) informativeness of the high-frequency return data and the option panel.

1. INTRODUCTION

Option data comprise a rich source of information about the volatility and jump risks of the underlying asset as well as their pricing. Over the last decade, both the amount of trading in existing option contracts and the number of newly marketed

Andersen and Varneskov gratefully acknowledge support from CREATES, Center for Research in Econometric Analysis of Time Series (DNRF78), funded by the Danish National Research Foundation. The work is partially supported by NSF Grant SES-1530748. We would like to thank the Editor (Peter C.B. Phillips), Co-Editor (Dennis Kristensen), and anonymous referees for many useful comments and suggestions. Address correspondence to Viktor Todorov, Department of Finance, Kellogg School of Management, Northwestern University, Evanston, IL 60208, USA; e-mail: v-todorov@northwestern.edu.

contracts have grown significantly. Nowadays, across several asset classes, there are active markets where a very large number of options written on the same underlying asset trade continuously throughout the trading hours. These options differ in terms of their tenor (maturity) and strike price. As a result, each derivative security offers unique information regarding the conditional risk-neutral distribution of the underlying asset. Moreover, for many assets, high-frequency price and quote data is readily available during trading hours and may aid in the estimation of the realized volatility of, and jump risks embedded in, the asset returns.

Taken together, the full trading record for each such asset can be overwhelming. The volume of tick data, reflecting every transaction and order book entry associated solely with the underlying asset, can be extremely large, amounting to hundreds of entries per second. Even so, the option data is typically far more challenging, as all quotes and transactions for hundreds of distinct options, differentiated by tenor and strike price, are recorded. Since option values invariably shift in response to movement in the underlying asset price, a large set of quotes is updated almost continuously. In addition, there is substantial heterogeneity in the option cross-section over time, as some contracts expire, others start trading after being introduced to the market, and a few simply fail to be quoted for some period, only to reenter later with nontrivial quote activity. Furthermore, the increasing liquidity and the availability of more, especially shorter, tenors in many option markets also generate significant low-frequency trends in the size of the cross-section, so the option panel is typically highly unbalanced.

Ideally, we should be looking for ways to exploit the high-frequency observations on both the option cross-section and the high-frequency return data to infer the evolving shape of the conditional term structure for the risk-neutral return distribution and monitor the evolution of the state variables driving the return dynamics. Given the current state of the option pricing literature, this goal remains elusive. However, significant progress is being made along the lines of formally developing inference tools that combine the high-frequency return information with lower frequency option data. The key is to rely on theory to identify the relevant statistics from the high-frequency return data that speak to the general distributional features of the risk-neutral distribution and to the concurrent value of variables that pertain to the state vector governing the future evolution of the return distribution.

Andersen et al. (2015) propose inference procedures for the parameters and factor realizations implied by a parametric model for the risk-neutral dynamics of the underlying asset based on an option panel with a fixed time span and a fixed set of option observation times, but an asymptotically expanding cross-sectional dimension. In this setting, the option-based information set is augmented by nonparametric estimates of the spot diffusive volatility constructed from high-frequency return data. Assuming that the high-frequency data is less informative than the option data (when combined with the parametric model) for recovery of the latent factor realizations, the system is estimated by penalized least squares, minimizing the L_2 distance between model-implied and observed option prices and further

https://doi.org/10.1017/S0266466618000373 Published online by Cambridge University Press

penalizing deviation between model-implied and nonparametric estimates of the spot diffusive volatility based on the high-frequency returns. This represents the first procedure to formally develop joint asymptotic distributional results for the parameters, state vector realizations, and the current value of the spot volatility, exploiting joint option and high-frequency return data.

In light of these observations, the aim of this article is twofold. First, we seek to relax the (fairly strict) assumption in Andersen et al. (2015) regarding the relative informativeness of the option and high-frequency return data about the parameters and factor realizations of the risk-neutral parametric model. Instead, we seek an approach that adapts to the quality of the two information sources and avoids the need for such a priori restrictions. Second, we wish to extend the framework by also including information from the high-frequency data about the jump component of the underlying asset returns.

We achieve these goals in a setting where the price process is of pure-jump type, i.e., in a model where the dynamics of the asset does not contain a diffusive component. In such models, the incessant small price fluctuations are instead captured through an infinite activity jump process, featuring the near continual arrival of minuscule return shocks. Models of pure-jump structure have been used previously by, e.g., Madan and Seneta (1990), Madan and Milne (1991), Barndorff-Nielsen and Shephard (2001), Carr, Geman, Madan, and Yor (2002, 2003). Moreover, nonparametric tests using high-frequency data in Todorov and Tauchen (2011b), Jing et al. (2011), and Hounyo and Varneskov (2017) find that some assets, e.g., the VIX index and several individual stocks, do not contain a diffusion.[1] The pure-jump setting readily emphasizes the jump features of the price process and complements the analysis in Andersen et al. (2015) by extending the inference technique to a setting void of diffusive components in the return dynamics. That said, the analysis in this article can be suitably adapted to situations in which the price contains a diffusion by replacing the high-frequency estimators of various quantities associated with the jumps, that we employ here, with ones that are robust to the presence of a diffusive component in the price process of the underlying asset.[2] The price to pay for robustness to the presence of a diffusion is a (much) slower rate of convergence of the high-frequency estimators relative to the ones used here.

Even in the absence of any parametric model for the actual return dynamics, high-frequency return data may be utilized as an additional source of information about the parametric model for its risk-neutral dynamics due to equivalence features of the statistical and risk-neutral probability measures implied by the no-arbitrage condition (a minimal assumption used in most theoretical and empirical

[1] Our approach thus deviates from the more common setting of capturing the "small" price moves via a diffusion component. For a recent study of the correlation between the diffusive part and the spot volatility, using high-frequency data, i.e., the so-called leverage effect, see Kalnina and Xu (2017). As noted in the text, however, there is growing nonparametric evidence that the price dynamics of some important assets lack a diffusion component.

[2] When the price process contains a diffusion, we may also incorporate diffusive spot volatility estimates in the estimation, in addition to the high-frequency measures for the jump part, as in Andersen et al. (2015).

asset pricing work). In a diffusive setting, the absence of arbitrage implies that the diffusion coefficient of the price process (spot volatility) is equivalent under the two probability measures, and this is exploited by Andersen et al. (2015). For the jumps in the model, no-arbitrage conditions are more complicated. For the "big" jumps, we have essentially no restrictions. This is intuitive, since no "big" jumps may materialize on a given path, even if they may occur with non-trivial probability. The equivalence of the statistical and risk-neutral probability measures does, however, impose a "similar" behavior of the "small" jumps under the two probabilities. In particular, the so-called jump activity index and the intensity of the "small" jumps should remain unchanged under an equivalent measure change. The jump activity index classifies jump processes according to the "vibrance" of their trajectories. For example, a jump activity index of less than one implies jumps of finite variation, while an activity index above one implies jumps of infinite variation. As such, the index has immediate implications for risk-measure estimation and interpretation as well as model specification. Consequently, the inference for jump activity using high-frequency data has received increasing attention in recent work, see, among others Woerner (2003, 2007), Ait-Sahalia and Jacod (2009), Todorov and Tauchen (2011a), Jing, Kong, and Liu (2011), Jing, Kong, and Liu (2012), Jing, Kong, Liu, and Mykland (2012), Bull (2016), Kong et al. (2015), Todorov (2015), Hounyo and Varneskov (2017), and Jacod and Todorov (2018).

Given the discussion above, we "summarize" the information in high-frequency return data about the risk-neutral parametric model for the underlying asset by estimates of the jump activity and the spot jump intensities at each option observation time. Specifically, we adopt the empirical characteristic function (ECF) approach of Todorov (2015) to estimate the jump activity, and we extend the analysis of the latter by providing a methodology to recover the spot jump intensities. We derive a central limit theorem (CLT) for our nonparametric high-frequency estimators, and we further show that it holds jointly with a corresponding limit theorem for the weighted sum of the option observation errors. This joint limit theory, in turn, allows us to characterize the limit distribution of an estimator that incorporates both high-frequency return data as well as option data. It is important to note that the analysis in this article may be readily adapted to alternative high-frequency jump activity and jump intensity estimators (e.g., ones that are robust to the presence of a diffusion in the dynamics of the price process), provided one can derive their asymptotic distribution.

The estimation of the risk-neutral model parameters and factor realizations is carried out via penalized least squares. In particular, we minimize the L_2 distance between observed and model-implied option prices and further penalize for deviations between the model-implied jump activity index and jump intensities and corresponding nonparametric estimates of these quantities based on high-frequency return data. The different parts of the parameter and state vectors may exhibit different rates of convergence depending on the relative information content (for our estimation purposes) of the return and option data. Importantly, however, the

https://doi.org/10.1017/S0266466618000373 Published online by Cambridge University Press

user does not need to take an a priori stand on this. That is, if the returns are more informative about, e.g., jump activity (in the sense of allowing for a faster rate of convergence), then our penalized least squares for this particular quantity asymptotically behaves as the nonparametric high-frequency estimator. The reverse holds true if the option data is more informative about the jump activity parameter—in which case our estimator, asymptotically, relies exclusively on the option data. In the boundary case where option and return data allow for estimators of jump activity with the same rate of convergence, we may assign the two information sources in the objective function optimal weights based on their relative precision. This feature is achieved by proposing a weighted penalized least squares extension of the described methodology, which has the added advantage of being free of tuning parameters.

The rest of the article is organized as follows. Section 2 introduces our formal model setup for the underlying asset and the associated option prices written on it. In Section 3, we discuss the observation scheme and the asymptotic setup for the inference. Section 4 presents our penalized least squares estimator and develops its associated asymptotic theory. In Section 5, we extend the results to a weighted penalized least squares, which provides efficiency gains. Section 6 reports the results from a Monte Carlo study of the newly developed inference method. Finally, Section 7 concludes. The formal statement of the assumptions and the proofs of the theorems are collected in Section A.

2. FRAMEWORK FOR PARAMETRIC PURE-JUMP MODELING OF OPTION PANELS

This section introduces a nonlinear parametric factor model for a panel of options written on an underlying asset, whose price is denoted by X. Specifically, the option prices are determined via a general parametric model of pure-jump type for the risk-neutral dynamics of X. In addition, we identify the characteristics of the underlying price process that are preserved under a change from the physical probability measure, $\mathbb{P}$, to the risk-neutral measure, $\mathbb{Q}$. These specific features of the return dynamics are invariant to any equivalent martingale measure transformation and provide fundamental restrictions on the joint dynamics of the statistical and risk-neutral distributions within all settings that retain the basic no-arbitrage assumption. These characteristics are important for the design of our inference procedures developed for parametric option pricing models in the subsequent sections, as we improve practical identification and enhance estimation efficiency by exploiting nonparametric estimates of the relevant quantities from high-frequency return data.

2.1. Pure-Jump Dynamics of Locally Stable Type

The dynamics for the price process X is defined on a filtered probability space $\left(\Omega^{(0)}, \mathcal{F}^{(0)}, (\mathcal{F}_t^{(0)})_{t\geq 0}, \mathbb{P}^{(0)}\right)$. Rather than imposing a parametric structure for X under $\mathbb{P}^{(0)}$, we merely assume that its $\mathbb{P}^{(0)}$–dynamics belongs to a general class of

https://doi.org/10.1017/S0266466618000373 Published online by Cambridge University Press

pure-jump models given by,

$$\frac{dX_t}{X_{t-}} = \alpha_t\, dt + \int_{\mathbb{R}} (e^x - 1)\, \tilde{\mu}^{\mathbb{P}}(dt, dx), \tag{1}$$

where the drift α_t is a process with càdlàg paths, and $\tilde{\mu}^{\mathbb{P}}(dt, dx) = \mu(dt, dx) - \nu^{\mathbb{P}}(dt, dx)$ is the martingale jump measure associated with the counting jump measure $\mu(dt, dx)$ and its compensator $\nu^{\mathbb{P}}(dt, dx)$. Specifically, $\nu^{\mathbb{P}}(dt, dx)$ is assumed to have the following structure,

$$\nu^{\mathbb{P}}(dt, dx) = \left(A_{t-}^{+}\, \nu_{+}^{\mathbb{P}}(x)\, \mathbf{1}_{\{x>0\}} + A_{t-}^{-}\, \nu_{-}^{\mathbb{P}}(x)\, \mathbf{1}_{\{x<0\}} \right) dt \otimes dx, \tag{2}$$

where the stochastic jump intensities for positive and negative jumps, A_t^+ and A_t^-, respectively, are processes with càdlàg paths, and the corresponding Lévy densities, $\nu_+^{\mathbb{P}}$ and $\nu_-^{\mathbb{P}}$, can be approximated around zero by the Lévy density of a stable process, that is,

$$\left| \nu_{\pm}^{\mathbb{P}}(x) - 1(x \gtrless 0) \frac{A_\beta}{|x|^{\beta+1}} \right| \leq \frac{C}{|x|^{\beta'+1}}, \quad A_\beta = \left(\frac{4\Gamma(2-\beta)|\cos(\beta\pi/2)|}{\beta(\beta-1)} \right)^{-1}, \tag{3}$$

with $|x| \leq x_0$, and $\beta' < \beta$ for some constants $C > 0$ and $x_0 > 0$. The coefficient β signifies the so-called jump activity, controlling the roughness of trajectories of X. That is, for every t,

$$\beta \equiv \inf\{p \geq 0 : \sum_{s \leq t} |\Delta X_s|^p < \infty\}, \qquad \text{almost surely.} \tag{4}$$

Remark 1. Obviously, the choice of A_β in (3) is simply a normalization. That is, if we have a specification for $\nu^{\mathbb{P}}$ as in (2) that satisfies (3), with A_β replaced with some other constant, then it also satisfies (2)–(3) with the constant A_β given in (3). The specific normalization in (3) using A_β simplifies the statement of the CLT for the estimators of β and $A_t^+ + A_t^-$, based on high-frequency return data, which we develop subsequently. This is clarified below.

We restrict attention to the case $1 < \beta < 2$, implying paths of infinite variation, which is found in earlier work to describe returns for a variety of assets well, see Todorov and Tauchen (2011b), Jing et al. (2011), Kong et al. (2015) and Hounyo and Varneskov (2017), among others.[3] The infinite variation mimics the equivalent property for diffusive representations of the return dynamics. In fact, as β approaches 2, the smaller jumps become more frequent, generating a smoother sample path, on average, thereby providing an increasingly better approximation

[3] We defer all formal assumptions to Section A.1. The restriction on β is only used for the high-frequency estimators. It can be relaxed at the expense of a more complicated exposition and a slower rate of convergence for the high-frequency estimators discussed below.

to a (scaled) Brownian motion. By letting $\beta < 2$, we facilitate direct analysis of the pure-jump alternative to diffusive specifications of the return dynamics. Moreover, the time-varying coefficients, A_t^+ and A_t^-, allow us to scale the increments of the jump innovations to generate volatility clustering, similarly to the scaling of Brownian increments in diffusive stochastic volatility models.

The specification (2)–(3) is very flexible and accommodates many parametric jump models used in empirical work. In particular, the "stable-like" restriction in equation (3) only applies for the behavior of the jump compensator around zero, leaving the behavior of "big" jumps virtually unrestricted. This assumption is obviously satisfied by the stable process, which has been used extensively for modeling a variety of economic phenomena, including asset return distributions, see, e.g., Mandelbrot (1961), Mandelbrot (1963), Fama (1963), and Fama and Roll (1968) for some early applications. It is also satisfied by, e.g., the CGMY model of Carr et al. (2002) as well as models in the class of tempered stable processes of Rosinski (2007), whose jumps may have much thinner tails than those of a stable process. Specifically, condition (3) is satisfied by $\nu_\pm^{\mathbb{P}}(x) = A_\beta \frac{e^{-\lambda_\pm |x|}}{|x|^{\beta+1}} 1_{\{x \gtrless 0\}}$, the popular CGMY model of Carr et al. (2002), adopted in many applications. In this model, the tail behavior of the jumps is controlled by the parameters $\lambda_\pm$, while the behavior of the "small" jumps is controlled by β.

The dynamics of the intensities A_t^+ and A_t^- is described by Assumption 1 in the Appendix. This assumption allows A_t^+ and A_t^- to be pure-jump Itô semimartingales, and it is general enough to allow for arbitrary dependence between the innovations in $A_t^\pm$ and X (i.e., we allow for a leverage effect). Similarly, it accommodates so-called self-excitation, where past jumps "feed" into the current jump intensity and thereby increase the probability of future arrivals of jumps. Assumption 1 rules out the presence of a diffusion in the dynamics of $A_t^\pm$, which is restrictive, but it, nevertheless, allows for a lot of models of stochastic volatility and jump intensity, e.g., the non-Gaussian Ornstein-Uhlenbeck processes of Barndorff-Nielsen and Shephard (2001). Note that Assumption 1 is only used to derive the asymptotic properties of the particular high-frequency estimators we adopt below. Alternative selections of high-frequency estimators may accommodate a diffusion in the dynamics of $A_t^\pm$.

Overall, our setup covers a general class of time-changed Lévy processes with absolutely continuous time-change, which is itself of the pure-jump type, see, e.g., Carr and Wu (2004), as well as any pure-jump model within the affine jump-diffusion class of Duffie, Pan, and Singleton (2000).

2.2. Parametric Pure-Jump Models for the Option Prices

We now specify the dynamics of X under the so-called risk-neutral measure which, in turn, enables us to determine the theoretical value of the options written on X. Assuming that arbitrage is absent, a risk-neutral probability measure, $\mathbb{Q}$, is guaranteed to exist, see, e.g., Section 6.K in Duffie (2001), and is locally equivalent to $\mathbb{P}^{(0)}$ (under some technical conditions). It transforms discounted asset

https://doi.org/10.1017/S0266466618000373 Published online by Cambridge University Press

prices into local martingales. Specifically, for X under $\mathbb{Q}$, we may write,

$$\frac{dX_t}{X_{t-}} = (r_t - q_t)\,dt + \int_{\mathbb{R}} (e^x - 1)\,\tilde{\mu}^{\mathbb{Q}}(dt,dx), \tag{5}$$

where r_t and q_t are the risk-free interest rate and dividend yield, respectively, and the martingale jump measure $\tilde{\mu}^{\mathbb{Q}}(dt,dx)$ is now defined with respect to the risk-neutral compensator, $\nu^{\mathbb{Q}}(dt,dx)$. As noted previously, in the absence of arbitrage, there are characteristics of the physical price process (1) that are preserved under the risk-neutral dynamics in (5). We identify these features below and utilize them explicitly when designing our estimation methodology.

Given the risk-neutral probability measure $\mathbb{Q}$, the theoretical value of European-style out-of-the-money (OTM) options written on X is given by the conditional expectation of their discounted terminal payoff,

$$O_{t,k,\tau} = \begin{cases} \mathbb{E}_t^{\mathbb{Q}}\left[e^{-\int_t^{t+\tau} r_s\,ds}\,(X_{t+\tau} - K)^+\right], & \text{if } K > F_{t,t+\tau}, \\ \mathbb{E}_t^{\mathbb{Q}}\left[e^{-\int_t^{t+\tau} r_s\,ds}\,(K - X_{t+\tau})^+\right], & \text{if } K \le F_{t,t+\tau}, \end{cases} \tag{6}$$

where τ and K are the tenor and strike price of the option, $F_{t,t+\tau}$ denotes the futures price of X at time t for the maturity date $t+\tau$, and we let $k = \ln(K/F_{t,t+\tau})$ denote the log-moneyness. We further define the Black-Scholes implied volatility (BSIV) corresponding to $O_{t,k,\tau}$ by $\kappa_{t,k,\tau}$, which represents a convenient monotone transformation often used to quote option prices in practice.

We assume throughout that we have a valid parametric model for the risk-neutral law of X. Specifically, let $\boldsymbol{S}_t$ denote a $p \times 1$ vector of state variables, or factors, taking values in $\boldsymbol{\mathcal{S}} \subset \mathbb{R}^p$, and $\boldsymbol{\theta}_0$ be the (true) value of a parameter vector of dimension $q \times 1$. Furthermore, $A_t^+ \equiv \xi_1(\boldsymbol{S}_t, \boldsymbol{\theta}_0)$ and $A_t^- \equiv \xi_2(\boldsymbol{S}_t, \boldsymbol{\theta}_0)$, where $\xi_1(\cdot)$ and $\xi_2(\cdot)$ are known functions.[4] In addition, the risk-neutral jump compensator is parameterized via,

$$\nu^{\mathbb{Q}}(dt,dx) = \left(\xi_1(\boldsymbol{S}_t,\boldsymbol{\theta}_0)\,\nu_+^{\mathbb{Q}}(x)\,\mathbf{1}_{\{x>0\}} + \xi_2(\boldsymbol{S}_t,\boldsymbol{\theta}_0)\,\nu_-^{\mathbb{Q}}(x)\,\mathbf{1}_{\{x<0\}}\right) dt \otimes dx, \tag{7}$$

where $\nu_\pm^{\mathbb{Q}}(x) \equiv \nu_\pm^{\mathbb{Q}}(x,\boldsymbol{\theta}_0)$. It is important to note that, similarly to the spot volatility for Brownian semimartingales, the stochastic jump intensities, A_t^+ and A_t^- are characteristics that are preserved under the equivalent change of measure from $\mathbb{P}^{(0)}$ to $\mathbb{Q}$. Moreover, since the characterization of the jump activity in definition (4) applies for each sample path (almost surely), the jump activity index under $\mathbb{Q}$ is also given by β, because the null sets of $\mathbb{P}^{(0)}$ and $\mathbb{Q}$ coincide. Hence, we treat β as a fixed parameter that is part of the parameter vector $\boldsymbol{\theta}_0$. We denote the remaining $(q-1)$ elements by $\boldsymbol{\theta}_0^r$, as the parameters β and $\boldsymbol{\theta}_0^r$ play different roles in the econometric analysis below.

The density of the probability measure change is given by a stochastic exponential involving the ratio $\nu^{\mathbb{Q}}/\nu^{\mathbb{P}}$ which, to be well-defined, requires (see, e.g.,

[4] We also assume that r_t and q_t are known functions of $\boldsymbol{S}_t$ and $\boldsymbol{\theta}_0$.

Lemma III.5.17 in Jacod and Shiryaev, 2003),

$$\int_{x>0}\left(\sqrt{\nu_+^{\mathbb{Q}}(x)}-\sqrt{\nu_+^{\mathbb{P}}(x)}\right)^2 dx < \infty \quad \text{and} \quad \int_{x<0}\left(\sqrt{\nu_-^{\mathbb{Q}}(x)}-\sqrt{\nu_-^{\mathbb{P}}(x)}\right)^2 dx < \infty. \tag{8}$$

The above condition, along with $\nu^{\mathbb{P}} \sim \nu^{\mathbb{Q}}$, is necessary and sufficient for the equivalence of $\mathbb{P}^{(0)}$ and $\mathbb{Q}$ in the Lévy case (where the jump compensator and drift are time invariant), see, e.g., Theorem 33.1 of Sato (1999). It severely restricts the wedge between $\nu_\pm^{\mathbb{Q}}$ and $\nu_\pm^{\mathbb{P}}$ around zero. To illustrate the manifestation of this fundamental feature, we consider the CGMY specification for $\nu_\pm^{\mathbb{Q}}$ given by,

$$c_\pm \frac{e^{-\lambda_\pm |x|}}{|x|^{\alpha+1}}, \quad c_\pm > 0, \;\; \lambda_\pm > 0, \;\; \alpha < 2. \tag{9}$$

Now, if $\nu_\pm^{\mathbb{P}}$ is also generated by a CGMY model, but with possibly different parameters, then, given the restriction (3), the condition (8) implies,

$$\alpha = \beta \quad \text{and} \quad c_+ = c_- = A_\beta. \tag{10}$$

Note, in particular, that we have no restrictions for the parameters $\lambda_\pm$ governing the behavior of the jump compensator in the tails. By contrast, the parameters controlling the behavior of the jump compensator around zero are unchanged, when switching from $\mathbb{P}^{(0)}$ to $\mathbb{Q}$. This example illustrates that $\nu_t^{\mathbb{P}}$ and $\nu_t^{\mathbb{Q}}$ are "essentially identical" around zero, but can be very different away from zero.

Remark 2. Our specification of $\nu^{\mathbb{Q}}$ in (7) is slightly more restrictive than what local equivalence of $\mathbb{P}$ and $\mathbb{Q}$ in conjunction with (2) implies. Indeed, using Theorem III.5.34 in Jacod and Shiryaev (2003) for $\mathbb{Q} \ll \mathbb{P}$, we need $\int_0^t \int_{\mathbb{R}} (1-\sqrt{Y(s,x)})^2 \nu^{\mathbb{P}}(ds,dx) < \infty$ to hold $\mathbb{Q}$-a.s. for every $t \geq 0$ and where $Y(t,x)$ is defined via $\nu^{\mathbb{Q}}(dt,dx) = Y(t,x)\nu^{\mathbb{P}}(dt,dx)$. This implies that $\nu^{\mathbb{Q}}$ and $\nu^{\mathbb{P}}$ should only be the same for x around zero and when $\nu^{\mathbb{P}}$ explodes around zero (which is the case for our specification of $\nu^{\mathbb{P}}$ as it is of infinite activity). For the specification in (2), the intensities $A_t^\pm$ control both the "small" and "big" jumps, and our imposition of $A_t^\pm$ being the same under $\mathbb{P}$ and $\mathbb{Q}$ restricts the intensity of the "big" jumps under $\mathbb{Q}$ more than what no-arbitrage (and local equivalence of $\mathbb{P}$ and $\mathbb{Q}$) would imply. Indeed, no arbitrage implies essentially no restriction for the risk-neutral properties of the "big" jumps (which are of finite activity). All of the results that follow will continue to hold if one considers more general parametric specifications of $\nu^{\mathbb{Q}}$, which do not restrict the risk-neutral jump measure of $\mathbb{Q}$ for the "big" jumps. That said, common parametric specifications of the jump measure are of the form we assume for $\nu^{\mathbb{Q}}$ in (7), and this is the reason we work with it henceforth.

Under the parametric model, the BSIV may be written as a function $\kappa(k,\tau,\boldsymbol{Z}_t,\boldsymbol{\theta})$, with $\boldsymbol{Z}_t$ and $\boldsymbol{\theta}$ denoting particular values of the state and parameter vectors, respectively. We let the parameter vector take realizations on a compact subset $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^q$. In this setting, we may write $\kappa_{t,k,\tau} \equiv \kappa(k, \tau, \boldsymbol{S}_t, \boldsymbol{\theta}_0)$, implying that, conditional on the model parameters, option prices are functions of tenor, moneyness and the state vector, with the latter driving all the time variation in the option prices. The evolution of the state vector, $\boldsymbol{S}_t$, can be specified very generally. We only require it to be an $\mathcal{F}^{(0)}$-adapted stochastic process. The above option pricing framework complements the ones in Andersen et al. (2015) and Andersen, Fusari, Todorov, and Varneskov (2018) by allowing the underlying asset price, X, to obey a pure-jump specification. Hence, while the existing approaches accommodate general affine jump-diffusion representations, the current setting enables us to handle non-Gaussian pure-jump option pricing models, e.g., the finite moment log-stable model for the option surface in Carr and Wu (2003).

3. OBSERVATION SCHEME AND HIGH-FREQUENCY RETURN MEASURES

This section describes the observation scheme for the options and high-frequency return data. The latter is used to augment the option information set. Next, we introduce the nonparametric high-frequency based estimators of the jump activity and jump intensities that are preserved under equivalent measure changes. Finally, we summarize the asymptotic distribution for these estimators of the spot jump characteristics. These results are needed to develop the joint inference for the pure-jump risk-neutral parametric model based on the option and high-frequency data in Section 4.

3.1. Option Observation Scheme

The time span of the option panel is given by $[0, T]$ for some fixed and finite $T > 0$, and we assume observations are available from the option surface at the integer times $t = 1, \ldots, T$. For each observation date, the setting is similar to that in Andersen et al. (2015) and Andersen et al. (2018). Specifically, the option data cover a fairly wide range of strikes and tenors, k and τ, respectively. That is, for each t, we observe options $\{O_{t,k_j,\tau_j}\}_{j=1,\ldots,N_t}$, where N_t is a large integer and the index j runs across the full set of strike and tenor combinations. Moreover, the number of options for maturity τ is denoted by N_t^τ, so that, by definition, $N_t = \sum_\tau N_t^\tau$. We let N_t^τ and N_t be $\mathcal{F}_t^{(0)}$-adapted.

We allow for considerable heterogeneity in the available option panel over observation times t through, for example, variation over t in the available number of options, the observed strike-tenor combinations (k, τ), and, for given τ, the density, or clustering, of available strikes in the log-moneyness grid. In particular, we define the following asymptotic ratios $N_t^\tau / N_t \approx \pi_t^\tau$ and $N_t / N \approx \varsigma_t$, where π_t^τ and ς_t are positive-valued processes, and N is an unobserved number, representing the "average size of the cross-section".[5] Moreover, for each combination

[5] Again, all formal assumptions are deferred to Section A.1.

of t and τ, we let $\underline{k}(t,\tau)$ and $\bar{k}(t,\tau)$ denote the minimum and maximum log-moneyness, respectively, and define the $\mathcal{F}_t^{(0)}$–adapted grid of available strikes as,

$$\underline{k}(t,\tau) = k_{t,\tau}(1) < k_{t,\tau}(2) < \cdots < k_{t,\tau}(N_t^\tau) = \bar{k}(t,\tau), \quad \text{with} \quad \Delta_{t,\tau}(i) = k_{t,\tau}(i) - k_{t,\tau}(i-1),$$

for $i = 2, \ldots, N_t^\tau$. In analogy with in-fill asymptotics for high-frequency returns, our asymptotic scheme does not expand the strike coverage, but instead sequentially adds new strikes within $[\underline{k}(t,\tau), \bar{k}(t,\tau)]$, such that $\Delta_{t,\tau}(i) \xrightarrow{\mathbb{P}} 0$ as $N \to \infty$, while allowing the clustering of strike prices to differ across certain regions of the strike range. That is, we let $N_t^\tau \Delta_{t,\tau}(i) \approx \psi_{t,\tau}(k_{t,\tau}(i))$ for some positive valued process $\psi_{t,\tau}(k)$. This heterogenous setting accommodates, e.g., the relatively high density of available OTM put options "close to the money," in contrast to the more sparsely available deep OTM call options. These facets impact the precision of our inference for the state vector over time, and the quantities π_t^τ, ς_t and $\psi_{t,\tau}(k)$ appear explicitly in the asymptotic distribution theory, as detailed in Section A.

In addition, $\mathcal{T}_t$ denotes the tenors available at time t, and the vectors $\underline{\boldsymbol{k}}_t = \left(\underline{k}(t,\tau)\right)_{\tau \in \mathcal{T}_t}$ and $\overline{\boldsymbol{k}}_t = \left(\overline{k}(t,\tau)\right)_{\tau \in \mathcal{T}_t}$ indicate the lowest and highest log-moneyness across all the available tenors at time t. As described above, these quantities may vary randomly over time, thus accommodating any pronounced shifts in the characteristics of the observed option cross-section across the sample.

Next, we stipulate that the BSIVs are observed with error, that is,

$$\widehat{\kappa}_{t,k,\tau} = \kappa_{t,k,\tau} + \epsilon_{t,k,\tau}, \tag{11}$$

where the measurement errors are defined on a space $\Omega^{(1)} = \bigtimes_{t\in\mathbb{N}, k\in\mathbb{R}, \tau\in\Gamma} \mathcal{R}_{t,k,\tau}$, for $\mathcal{R}_{t,k,\tau} \subset \mathbb{R}$, with Γ denoting the set of all possible tenors. Moreover, $\Omega^{(1)}$ is equipped with a Borel σ-field $\mathcal{F}^{(1)}$ as well as a transition probability $\mathbb{P}^{(1)}(\omega^{(0)}, d\omega^{(w)})$ from the original probability space $\Omega^{(0)}$ to $\Omega^{(1)}$. Then, by defining the filtration on $\Omega^{(1)}$ via $\mathcal{F}_t^{(1)} = \sigma(\epsilon_{s,k,\tau} : s \le t)$, we may write the filtered probability space as $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\ge 0}, \mathbb{P})$, where $\Omega = \Omega^{(0)} \times \Omega^{(1)}$, $\mathcal{F} = \mathcal{F}^{(0)} \times \mathcal{F}^{(1)}$,

$$\mathcal{F}_t = \cap_{s>t} \mathcal{F}_s^{(0)} \times \mathcal{F}_s^{(1)}, \quad \text{and} \quad \mathbb{P}(d\omega^{(0)}, d\omega^{(1)}) = \mathbb{P}^{(0)}(d\omega^{(0)})\mathbb{P}^{(1)}(\omega^{(0)}, d\omega^{(1)}).$$

Processes defined on $\Omega^{(0)}$ and $\Omega^{(1)}$, respectively, such as X_t and $\epsilon_{t,k,\tau}$, may trivially be viewed as processes on Ω, and we assume that any local martingale and semimartingale properties are preserved on the extended space. This decomposition of the probability space may be motivated as follows. The option errors are defined on an auxiliary space $\Omega^{(1)}$, equipped with a "large" supporting product topology, since they may be associated with any strike, point in time and maturity. This space suffices because, at each point in time, only a countable number of errors appear in the estimation. Finally, since we want to accommodate dependence between $\epsilon_{t,k,\tau}$ and the underlying process X_t, we define the probability measure via a transition probability distribution from $\Omega^{(0)}$ to $\Omega^{(1)}$.

3.2. Inference for Jump Characteristics from High-Frequency Return Data

In addition to the option price panel, we utilize a second source of information for estimation, namely high-frequency data on the underlying asset X, to assist in the recovery of the state and parameter vectors (or parts of them). Specifically, we shall estimate the total jump intensity,

$$A_t = A_t^+ + A_t^-,$$

and the activity index, β, nonparametrically. To this end, we assume we have an equidistant high-frequency recording of X_t at times $0, 1/n, \dots, i/n, \dots, T$, so the increment size is $\Delta_n = 1/n$. Finally, we define the logarithmic price and return by $x_t = \log(X_t)$ and $\Delta_i^n x = x_{i/n} - x_{(i-1)/n}$.

3.2.1. Jump Activity Estimation. We compute the jump activity index, β, using the estimator in Todorov (2015), which is based on self-normalized statistics of the increments $\Delta_i^n x - \Delta_{i-1}^n x$, and their empirical characteristic function (ECF). The use of second-order differences alleviates the impact from the drift as well as the (possibly) asymmetric jump intensities. Moreover, the use of the ECF generates efficiency gains over corresponding power variation-based methods, see, e.g., Todorov (2015) and Remark 3 below.

To set the stage, let $1 < k_n < \lfloor nT/2 \rfloor$ be the block size. The first ingredient of the jump activity estimator is a local power variation estimate of the total jump intensity A_t,

$$\widehat{V}_i(p) = \frac{1}{k_n} \sum_{j=i-k_n}^{i-1} \left| \Delta_{2j}^n x - \Delta_{2j-1}^n x \right|^p, \qquad i = k_n + 1, \dots, \lfloor nT/2 \rfloor, \tag{12}$$

which is then used to scale the differenced increments in the construction of the ECF as,

$$\widehat{\mathcal{C}}(p,u) = \frac{1}{\lfloor nT/2 \rfloor - k_n} \sum_{i=k_n+1}^{\lfloor nT/2 \rfloor} \cos\left(u \frac{\Delta_{2i}^n x - \Delta_{2i-1}^n x}{(\widehat{V}_i(p))^{1/p}} \right), \quad u \in \mathbb{R}_+. \tag{13}$$

The above statistic differs slightly from its counterpart in Todorov (2015) by the summands in $\widehat{V}_i(p)$ and $\widehat{\mathcal{C}}(p,u)$ having nonoverlapping increments. This results in our jump activity estimator being slightly less efficient, as we have fewer summands in $\widehat{\mathcal{C}}(p,u)$ for a given data set. This modification, however, allows us to handle the more general setting, where the jump intensity around zero can be asymmetric, i.e., we may have $A_t^+ \neq A_t^-$.

The asymptotic properties of $\widehat{\mathcal{C}}(p,u)$ naturally depend on the properties of $\widehat{V}_i(p)$. In particular, consistency of the latter for the total intensity, A_t, requires $k_n \to \infty$. Similarly, $k_n/n \to 0$ is needed to avoid time-variation in A_t generating a bias. Moreover, Todorov (2015) shows that $k_n/\sqrt{n} \to 0$ suffices to ensure that

https://doi.org/10.1017/S0266466618000373 Published online by Cambridge University Press

the sampling error biases in $\widehat{V}_i(p)$ are sufficiently small, and that a bias-corrected ECF,

$$\widetilde{\mathcal{C}}(p,u,\beta) = \widehat{\mathcal{C}}(p,u) - \mathcal{B}_n(p,u,\beta), \tag{14}$$

accommodates a CLT.[6] Next, to fully utilize the advantages of a characteristic function-based approach, we estimate β in two steps. The first step consists of constructing a preliminary activity index estimate using the raw ECF,

$$\widehat{\beta}^{fs}(p,u,v) = \frac{\log\left(-\log\left(\widehat{\mathcal{C}}(p,u)\right)\right) - \log\left(-\log\left(\widehat{\mathcal{C}}(p,v)\right)\right)}{\log(u/v)}, \tag{15}$$

for some $u,v \in \mathbb{R}_+$ with $u \neq v$. Now, due to the asymptotic bias in $\widehat{\mathcal{C}}(p,u)$, induced by the sampling errors in $\widehat{V}_i(p)$, the rate of convergence of the estimator $\widehat{\beta}^{fs}(p,u,v)$ will be suboptimal. Specifically, we have $\widehat{\beta}^{fs}(p,u,v) - \beta = O_p(1/k_n)$, subject to certain regularity conditions on p and k_n. Hence, we follow Todorov (2015) and construct a second-step estimator based on the bias-corrected ECF as,

$$\widehat{\beta}(p,u,v) = \frac{\log\left(-\log\left(\widetilde{\mathcal{C}}(p,u,\widehat{\beta}^{fs})\right)\right) - \log\left(-\log\left(\widetilde{\mathcal{C}}(p,v,\widehat{\beta}^{fs})\right)\right)}{\log(u/v)}, \tag{16}$$

for $u,v \in \mathbb{R}_+$ with $u \neq v$, and where $\widehat{\beta}^{fs} \equiv \widehat{\beta}^{fs}(p,u,v)$ is used as short-hand notation.[7] Similarly, we often write $\widehat{\beta} = \widehat{\beta}(p,u,v)$ for brevity. As shown below, the estimator in equation (16) achieves an almost optimal speed of convergence of $1/\sqrt{n}$.

The asymptotic variance of $\widehat{\beta}(p,u,v)$ depends only on β and the pair (u,v), while, due to the self-normalization of the increments in $\widetilde{\mathcal{C}}(p,u,\beta)$, it is independent of the stochastic intensities $A_t^{\pm}$. The constants u and v can be chosen in such a way that the asymptotic limits of $\widetilde{\mathcal{C}}(p,u,\beta)$ and $\widetilde{\mathcal{C}}(p,v,\beta)$ are sufficiently removed from 0 for all possible values of β. We conjecture more efficient implementations of the estimator, in which u and v are selected adaptively based on a preliminary estimator of β, are feasible, but we do not consider such extensions here to avoid complicating the exposition.

3.2.2. Jump Intensity Estimation. This section provides a new nonparametric estimator of the total spot jump intensity. Unlike the jump activity, we allow the jump intensity to change over time. Given our option observation scheme, we need estimates for A_t at each $t = 1,\ldots,T$, a quantity for which no spot estimator has been developed previously. We construct such estimators using local blocks consisting of p_n differenced and nonoverlapping increments preceding the integer time points.

[6] The exact expression for $\mathcal{B}_n(p,u,\beta)$ is provided in Section A.2.

[7] Note that $\widehat{\beta}^{fs}$ is just one example of a first-stage estimator. Under suitable regularity conditions, we could also apply, e.g., power variation-based estimators such as those in Ait-Sahalia and Jacod (2009) or Todorov and Tauchen (2011a).

One candidate estimator of A_t is given by the local power variation $\widehat{V}_i(p)$ for an appropriate choice of i. However, as illustrated by Todorov (2015) in the context of analyzing the jump activity index, estimators based on the empirical characteristic function can provide nontrivial efficiency improvements. Consequently, we propose the following estimator,

$$\widehat{A}_t(u) \;=\; -\frac{1}{u^{\widehat{\beta}}}\log\left(\frac{1}{p_n}\sum_{i\in\mathbb{I}_t^n}\cos\left(u\,\Delta_n^{-1/\widehat{\beta}}(\Delta_{2i}^n x-\Delta_{2i-1}^n x)\right)\right),\qquad t=1,\ldots,T, \tag{17}$$

where $\mathbb{I}_t^n=\{\lfloor tn/2\rfloor-p_n+1,\ldots,\lfloor tn/2\rfloor\}$, and p_n is a deterministic sequence satisfying $p_n\to\infty$ and $p_n/n\to 0$. Note that, at the expense of a more complicated analysis, one may further generalize equation (17) to separately identify A_t^+ and A_t^-. We leave such an extension for future research.

3.2.3. Inference for Spot Jump Characteristics from High-Frequency Return Data. We need some additional notation to summarize the results regarding the asymptotic distribution of the nonparametric high-frequency estimators for the equivalent, measure invariant, spot jump features. First, we let $\widehat{A}_t\equiv\widehat{A}_t(u)$ and define the $T\times 1$ vectors $\widehat{\boldsymbol{A}}=(\widehat{A}_t)_{t=1}^T$ and $\boldsymbol{A}=(A_t)_{t=1}^T$. Next, we note that the convergence of the nonparametric estimators (after centering around their probability limits) is stable. This is denoted by $\xrightarrow{\mathcal{L}-s}$. Stable convergence is stronger than the usual notion of convergence and implies that the convergence holds jointly with any bounded random variable defined on the original probability space. This stronger form of convergence is critical for the derivation of the asymptotic distribution for the PLS estimator in Section 4.

THEOREM 1. *Suppose Assumption A.1 in Section A.1 holds. Moreover, let the power p as well as the sequences k_n and p_n in equations (12), (13), and (17) satisfy the following conditions,*

(*R1*) $p_n\asymp\sqrt{n}$,

(*R2*) $\frac{\beta\beta'}{2(\beta-\beta')}\vee\frac{\beta-1}{2}<p<\frac{\beta}{2}$,

(*R3*) $k_n\asymp n^{\varpi}$ *with* $\frac{p}{\beta}\vee\frac{1}{3}<\varpi<\frac{1}{2}$.

Then, it follows

$$\begin{pmatrix}\sqrt{nT} & 0\\ 0 & \sqrt{p_n}\end{pmatrix}\begin{pmatrix}\widehat{\beta}-\beta\\ \widehat{\boldsymbol{A}}-\boldsymbol{A}\end{pmatrix}\xrightarrow{\mathcal{L}-s}\begin{pmatrix}\Psi_\beta^{1/2} & \mathbf{0}_{1\times T}\\ \mathbf{0}_{T\times 1} & \boldsymbol{\Psi}_A^{1/2}\end{pmatrix}\times\begin{pmatrix}\Upsilon_\beta\\ \mathbf{Y}_A\end{pmatrix},$$

where the scalar Υ_β and the $T\times 1$ vector $\mathbf{Y}_A$ are standard Gaussian, defined on an extension of the original probability space, with each of them independent of each other as well as of $\mathcal{F}$. The scalar Ψ_β and the $T\times T$ matrix $\boldsymbol{\Psi}_A=\mathrm{diag}(\Psi_1,\ldots,\Psi_T)$ are defined in Section A.2.

Theorem 1 extends results from Todorov (2015) in two directions. First, we allow for asymmetry in the jump intensity around zero, i.e., we accommodate the setting $A_t^+ \neq A_t^-$. Second, in addition to estimating β, we consider estimates of the spot quantity A_t at each point in time t. Naturally, the rate of convergence of A_t is governed by the number of increments p_n used in its estimation, which is much smaller than the total number of high-frequency increments on the interval $[0, T]$ utilized in the estimation of β. Hence, as expected, $\widehat{\beta}$ converges at a faster rate than $\widehat{\boldsymbol{A}}$. Because of this feature, the use of $\widehat{\beta}$ in the construction of $\widehat{\boldsymbol{A}}$ has no effect on the limiting result in Theorem 1. Note that this is very different from the case where one aims to recover the integrated intensity, $\int_0^T A_s ds$. The asymptotic distribution of the latter is dominated by the use of $\widehat{\beta}$ in its construction and, as a result, this generates perfect asymptotic dependence between the integrated jump intensity estimator and the estimator of the jump activity. In our case, this asymptotic degeneracy is avoided by the slower rate of convergence of the jump intensity estimator. The choice of p_n in R1 is standard for estimation of spot quantities (e.g., spot diffusive volatility). It reflects a balance between the bias in the recovery of the spot jump intensity, caused by the time-variation in the latter, and the variance in its estimation.

The asymptotic distribution of $\widehat{\beta}$ is Gaussian with constant variance. This is expected as $\widetilde{\mathcal{C}}(p, u, \beta)$ is self-normalized, annihilating the effect from the time-variation in $A_t^{\pm}$ on its limiting distribution. On the other hand, the asymptotic distribution of $\widehat{\boldsymbol{A}}$ is mixed Gaussian, and hence the precision in the recovery of $\boldsymbol{A}$ depends on its random realization.

Conditions R2 and R3 are exactly as in Todorov (2015), determining the range of possible choices for the power and block size of the local power variation statistic used to normalize the differenced increments, which, in turn, are used to construct $\widehat{\beta}$. In general, it is sensible to select the block size parameter, ϖ, very close to $1/2$. For the power p, a feasible choice is setting it arbitrarily close to, yet above, $1/2$. In principle, given that the unknown parameter β appears in the restrictions R2 and R3, one may consider an adaptive choice for k_n and p. Such considerations are left for future work.

Remark 3. An alternative way to estimate β is to use realized power variations over two different time scales, see, e.g., Woerner (2003, 2007), Todorov and Tauchen (2011a), Jing et al. (2011) and Hounyo and Varneskov (2017). Given this estimator of β, one can construct an estimator of $\boldsymbol{A}$ based on (local) realized power variation computed over either one of the two time scales. However, as shown in Todorov (2015), methods based on the ECF, which we adopt here, offer nontrivial efficiency improvements over estimators based on power variations.

Remark 4. In the case where the price X contains a diffusion, the estimator $\widehat{\beta}$ will converge to 2 (which can be viewed as the "activity" of the diffusion). The estimator $\widehat{\boldsymbol{A}}$, in the presence of a diffusion, provides estimates of the diffusive spot volatility, which, importantly, is robust to jumps. Thus, suppose our model under $\mathbb{Q}$ is given by

https://doi.org/10.1017/S0266466618000373 Published online by Cambridge University Press

$$\frac{dX_t}{X_{t-}} = (r_t - \delta_t)dt + \sigma_t^{1/\beta} dS_t + dJ_t,$$

where S_t is a β-stable process, with $\beta = 2$ corresponding to the Brownian motion, and J_t is a "residual" jump process whose activity is dominated by that of S_t (e.g., J_t is of finite activity and controls the "big" jumps of X so that options written on X are finite-valued). In this case, $\widehat{\beta}$ will estimate the parameter β of S_t and, similarly, $\widehat{A}_t$ will estimate σ_t.

Remark 5. When X contains a diffusive component, one may use truncated power variations (where truncation is from below in order to minimize the effect of the diffusion in X), see, e.g., Ait-Sahalia and Jacod (2009), Jing et al. (2011), Jing et al. (2012), and Bull (2016) or empirical characteristic functions, see, e.g., Jacod and Todorov (2018), to estimate the jump activity as well as the jump intensity. In this setting, the fastest attainable rate of convergence for estimating β is reduced to $n^{\beta/4}$, and the corresponding estimate of $\boldsymbol{A}$ cannot achieve convergence faster than $n^{\beta/8}$ (under standard specifications for the dynamics of A_t).

Remark 6. Theorem 1 is a key building block for the derivation of the asymptotic distribution of our PLS estimator. If one exploits alternative estimators of β and $\boldsymbol{A}$, e.g., adapted to settings in which X may contain a diffusive component, then, in order to adapt the asymptotic analysis of Theorem 3 below, one simply needs to provide a CLT result equivalent to Theorem 1 for the chosen combination of nonparametric high-frequency estimators.

4. INFERENCE FOR PURE-JUMP MODELS FROM OPTION PANELS

The material in this section constitutes the core of our econometric analysis. We introduce a new penalized least squares (PLS) estimator for option panels associated with pure-jump parametric models for the underlying asset. We motivate the design of the estimator and develop the necessary asymptotic theory for feasible inference. The PLS estimator utilizes information from the high-frequency returns via the estimators for the jump activity and the jump intensity introduced in Section 3. In particular, the joint CLT for the nonparametric high-frequency jump estimators in Section 3.2.3 is an important ingredient in the derivation of the asymptotic distribution for our new PLS estimator.

4.1. Penalized Least Squares

In designing the PLS estimator for option price panels generated from pure-jump models, we use several key observations from Sections 2 and 3. First, given the signal-plus-noise decomposition of observed BSIVs in equation (11), it is natural to estimate the parameter, $\boldsymbol{\theta}_0$, and the latent factor realizations, $\boldsymbol{S} = \{\boldsymbol{S}_t\}_{t=1}^T$, via least squares. Second, as discussed in Section 2.2, the jump activity index, β, and the total spot jump intensity, $A_t = A_t^+ + A_t^-$, are preserved under change of measure from $\mathbb{P}$ to $\mathbb{Q}$. These quantities may be recovered nonparametrically from

https://doi.org/10.1017/S0266466618000373 Published online by Cambridge University Press

high-frequency return data with the estimators presented in Sections 3.2.1 and 3.2.2, and we shall utilize this additional source of information in the estimation.

Formally, we let $\boldsymbol{\theta}_0 = (\boldsymbol{\theta}_0^r, \beta)$ and $\boldsymbol{S}_t = (\boldsymbol{S}_t^r, A_t)$, $t = 1, \ldots, T$, denote decompositions of the latent parameter and state vector, respectively, and let $\boldsymbol{\theta} = (\boldsymbol{\theta}^r, \mathcal{B})$ and $\boldsymbol{Z}_t = (\boldsymbol{Z}_t^r, \mathcal{A}_t)$ be corresponding generic vectors. Then, by defining the $T \times p$ matrix of factor realizations as $\boldsymbol{Z} = \{\boldsymbol{Z}_t'\}_{t=1}^T$, we write the objective function, for some finite constants $\lambda_\beta \geq 0$ and $\lambda_A \geq 0$, as,

$$\mathcal{L}(\boldsymbol{Z}, \boldsymbol{\theta}) \equiv \sum_{t=1}^{T} \mathcal{L}_t(\boldsymbol{Z}_t, \boldsymbol{\theta}) + \lambda_\beta n T \left(\widehat{\beta} - \mathcal{B}\right)^2, \quad \text{with} \tag{18}$$

$$\mathcal{L}_t(\boldsymbol{Z}_t, \boldsymbol{\theta}) \equiv \left\{ \sum_{j=1}^{N_t} \left(\hat{\kappa}_{t,k_j,\tau_j} - \kappa(k_j, \tau_j, \boldsymbol{Z}_t, \boldsymbol{\theta})\right)^2 + \lambda_A\, p_n \left(\widehat{A}_t - \mathcal{A}_t\right)^2 \right\}.$$

The first part of the objective function is the L_2 distance between observed and model-implied option prices (quoted in BSIV). The second and third parts are penalization terms for the deviation of the model-implied jump activity index and jump intensities from direct, but noisy, nonparametric measures of them from high-frequency return data. These penalization terms aid identification and estimation of (parts of) the parameter and state vectors, which are obtained as follows,[8]

$$(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{S}}) = \underset{\boldsymbol{\theta} \in \boldsymbol{\Theta}, \boldsymbol{Z} \in \boldsymbol{\mathcal{S}}^T}{\operatorname{argmin}} \mathcal{L}(\boldsymbol{Z}, \boldsymbol{\theta}), \quad \boldsymbol{\mathcal{S}} \subset \mathbb{R}^p. \tag{19}$$

Our new estimator differs in several respects from the corresponding PLS estimators explored by Andersen et al. (2015) and Andersen et al. (2018). The latter exploit different asymptotic designs and, more fundamentally, they assume that the underlying price process contains a diffusion, i.e., a martingale component driven by a Brownian motion. As a result, for those estimators the penalization, at each option observation time, refers to deviations between the model-implied spot volatility and a nonparametric measure of spot volatility obtained from high-frequency return data. The analogue to the scaling of a Brownian motion with spot volatility in the pure-jump setting is the scaling of the martingale jump measure by the jump intensity A_t (or by A_t^+ and A_t^- separately for positive and negative jumps). By contrast, the inclusion of a penalty for the deviation between the model-implied and high-frequency return estimate for the jump activity index is unique to the pure jump setting. For jump-diffusive models, the activity index is two ($\beta \equiv 2$) by assumption, as the presence of the Wiener component is stipulated as an integral part of the model specification.[9] In our pure-jump scenario,

[8] The use of a noisy measure of the state vector (or a part of it) in the design of an estimator also bears resemblance with the FAVAR approach in Bernanke, Boivin, and Eliasz (2005), who augment a VAR of economic variables with a noisy estimate of a latent factor that is related to the variables in the system.

[9] The activity index is defined as the infimum over the set of powers for which the power variation is finite. When the price contains a nonvanishing diffusion component, the power variation for any power below 2 is infinite.

we assume $1 < \beta < 2$, but do not fix the index to any given value, so it becomes a key parameter that must be estimated from the option and high-frequency return sample. Hence, the added penalty term arises naturally from the restriction that this index also is invariant to equivalent martingale measure transformations. Another major difference to the earlier PLS estimators is that we avoid placing restrictions on the relative information content in high-frequency return and option data.[10] That is, we allow for arbitrary relations between N, n, and p_n. This enables the procedure to adapt (asymptotically) to the relative informativeness of the different data sources. Nevertheless, one should keep in mind that the option data, generally, is required in the estimation of the risk-neutral dynamics, since the high-frequency return data only aid in the estimation of those parts of the parameter and state vectors that are invariant across the two probability measures.

4.2. Consistency of the PLS Estimator

Exploiting Theorem 1, we may now establish the consistency of $\widehat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{S}} = (\widehat{\boldsymbol{S}}_t)_{t=1}^T$.

THEOREM 2. *Suppose the Assumptions A.1–A.5 in Section A.1 as well as R1–R3 of Theorem 1 hold. Then, for some $T \in \mathbb{N}$ and fixed $\lambda_\beta \geq 0$ and $\lambda_A \geq 0$, it follows that $(\widehat{\boldsymbol{\theta}}, (\widehat{\boldsymbol{S}}_t)_{t=1}^T)$ exists with probability approaching 1, and further that,*

$$\left\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right\| \xrightarrow{\mathbb{P}} 0, \qquad \left\|\widehat{\boldsymbol{S}}_t - \boldsymbol{S}_t\right\| \xrightarrow{\mathbb{P}} 0, \qquad t = 1, \ldots, T.$$

Theorem 2 shows that we can consistently recover the risk-neutral model parameters and the state vector under general conditions. As explained above, one major departure from the equivalent results in Andersen et al. (2015) and Andersen et al. (2018) arises from the inclusion of information from high-frequency data about *both* the parameter and state vectors in the estimation. Of course, if we set $\lambda_\beta = \lambda_A = 0$, we will not need Theorem 1 and may exclude the rate conditions R1–R3.

The critical condition needed for the above consistency result is the ability of the option cross-sections to identify uniquely the parameters of the model as well as the latent factor realizations at the times of observing the options. We have given a high-level identification condition for this in the Appendix, see Assumption 3. In our general setting, we cannot give more primitive conditions for identification and, as usual, the latter should be argued for on a case by case basis. Nevertheless, we can make the following general comments for identification of parameters and factors from cross-sections of options. In our infill asymptotic limit, we observe all options on the log-strike intervals $[\underline{k}(t,\tau), \overline{k}(t,\tau)]$ for each tenor τ and time point t. For identification of a risk-neutral model from these options, it suffices to show that we can achieve identification by matching the

[10] In comparison to Andersen et al. (2015) and Andersen et al. (2018), the scaling $1/N_t$ has been removed from the objective function in order to simplify the treatment of the (possibly) different rates of convergence of parts of the parameter and state vectors.

values for portfolios of options of the form $\int_{\underline{k}(t,\tau)}^{\overline{k}(t,\tau)} O_{t,k,\tau} f(k)dk$, for some known and smooth functions f. Using the spanning results of Carr and Madan (2001), these portfolios replicate (nonparametrically) risk-neutral moments of the returns, provided $[\underline{k}(t,\tau),\overline{k}(t,\tau)]$ cover the support of the return distribution over the interval $[t, t+\tau]$. In general, identification is easier to show in terms of risk-neutral moments of returns. For example, using the above-mentioned spanning results, the cross-section of options can recover the conditional characteristic function $\mathbb{E}_t^{\mathbb{Q}}(e^{iux_{t+\tau}})$, for $u \in \mathbb{R}$. Hence, for the identification condition needed to establish Theorem 2, it suffices to show that $\boldsymbol{\theta}_0$ and $\boldsymbol{S}_t$ uniquely identify the conditional characteristic function for the available tenors, which is equivalent to $\boldsymbol{\theta}_0$ and $\boldsymbol{S}_t$ uniquely identifying the conditional risk-neutral return distribution for the available tenors. Specifically, if the maturity of the shortest available tenor goes to zero asymptotically, then, as shown by Qin and Todorov (2018), we can identify the density of $\nu^{\mathbb{Q}}(dt,dx)$ from such short-dated options. In turn, longer dated options may be used to identify the additional parameters that control the risk-neutral dynamics of the latent factors. For example, suppose $\boldsymbol{S}_t = (A_t^+, A_t^-)'$ and, further, that $\boldsymbol{S}_t$ is a non-Gaussian Ornstein-Uhlenbeck process (see equation (25) in the Monte Carlo section below). In this case, given that the parameters controlling $\nu_\pm^{\mathbb{Q}}(x)$ and the level of $\boldsymbol{S}_t$ are identified from short-dated options, we can identify the parameters determining the dynamics of $\boldsymbol{S}_t$ from longer-dated options by utilizing portfolios that span the first and second conditional risk-neutral moments of the returns for these horizons.

4.3. Asymptotic Distribution of the PLS Estimator

The central limit theory for the parameters and the state vector realizations depends on the relative informativeness of the options and high-frequency data, respectively. To highlight this feature, let us, again, make the decompositions $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\theta}}^r, \widehat{\mathcal{B}})$ and $\widehat{\boldsymbol{S}} = (\widehat{\boldsymbol{S}}^r, \widehat{\mathcal{A}})$. Moreover, we define $\bar{n} = n \vee N$ and $\bar{p}_n = p_n \vee N$ as well as the scaling matrix,

$$\boldsymbol{W}_n \equiv \operatorname{diag}(\boldsymbol{W}^n_{\theta_0^r}, W^n_\beta, \boldsymbol{W}^n_{\boldsymbol{S}^r}, \boldsymbol{W}^n_A), \tag{20}$$

where $\boldsymbol{W}^n_{\theta_0^r} = \boldsymbol{\iota}_{q-1}/\sqrt{N}$, $W^n_\beta = 1/\sqrt{\bar{n}}$, $\boldsymbol{W}^n_{\boldsymbol{S}^r} = \boldsymbol{\iota}_{T(p-1)}/\sqrt{N}$, and $\boldsymbol{W}^n_A = \boldsymbol{\iota}_T/\sqrt{\bar{p}_n}$ contain information about the convergence rates of different parts of the parameter and state vectors, while $\boldsymbol{\iota}_d$ denotes a d-dimensional vector of ones. We may now state the limiting distribution result for our PLS estimator.

THEOREM 3. *Under Assumptions A.1–A.7 in Section A.1 as well as R1–R3 of Theorem 1 and for some fixed $\lambda_\beta \geq 0$ and $\lambda_A \geq 0$, we have*

$$\boldsymbol{W}_n^{-1}\begin{pmatrix}\widehat{\boldsymbol{\theta}}^r - \boldsymbol{\theta}_0^r\\ \widehat{\mathcal{B}} - \beta\\ \widehat{\boldsymbol{S}}^r - \boldsymbol{S}^r\\ \widehat{\mathcal{A}} - \boldsymbol{A}\end{pmatrix} \xrightarrow{\mathcal{L}-s} \boldsymbol{\mathcal{I}}^{-1}\boldsymbol{\Omega}^{1/2} \times \begin{pmatrix}\mathbf{E}_{\theta_0^r}\\ \mathrm{E}_\beta\\ \mathbf{E}_{S^r}\\ \mathbf{E}_A\end{pmatrix},$$

where E_β *and the* $(q-1)\times 1$*,* $T(p-1)\times 1$*, and* $T\times 1$ *vectors* $\mathbf{E}_{\theta_0^r}$*,* $\mathbf{E}_{S^r}$*, and* $\mathbf{E}_A$*, respectively, consist of standard Gaussian random variables defined on an extension of the original probability space, with each of them independent of the others as well as of the filtration* $\mathcal{F}$*. The Hessian and asymptotic covariance matrices,* $\mathcal{I}$ *and* $\mathbf{\Omega}$*, are defined in equations* (*A.7*) *and* (*A.8*) *of Section A.3.*

The limiting result in Theorem 3 shows different rates of convergence for the components of the PLS estimator. In particular, for the estimates of the components of the parameter and state vectors that we have no information on from the high-frequency returns, i.e., $\widehat{\boldsymbol{\theta}}^r$ and $\widehat{\boldsymbol{S}}^r$, the rate of convergence is simply $\sqrt{N}$ (recall from Section 3.1 that N denotes the average size of the option cross-section). On the other hand, the rate of convergence for the jump activity parameter, β, is determined by the faster of the $\sqrt{N}$ and $\sqrt{n}$ rates associated with utilizing the information from the parametric model as well as the option panel and the nonparametric estimator based on the high-frequency return data, respectively. In that regard, we note that the scaling of the penalization terms in the objective function in equation (18) plays an important role, ensuring that the latter have a negligible effect in the estimation, when the high-frequency data is less informative in relative terms than the option data (for the jump activity parameter), i.e., when $n \ll N$. In the opposite case, i.e., when $n \gg N$, the scaling of the penalty term in the objective function (corresponding to β) guarantees that the latter determines the asymptotic behavior of the jump activity estimator. In the borderline case $n \asymp N$, both the high-frequency return and option data contribute to the asymptotic variance of β, and this is reflected in their joint determination of the terms in $\mathcal{I}$ and $\mathbf{\Omega}$ that correspond to $\widehat{\mathcal{B}}$. Similar comments apply to the estimator of the jump intensity, $\widehat{\mathcal{A}}$. In this case, the relevant comparison is the convergence rate of $\sqrt{N}$, from utilizing the option data, versus the $\sqrt{p_n}$ rate, when using high-frequency data.

Since $p_n/n \to 0$ by condition R1 in Theorem 1, if $N \gg n$, then $N \gg p_n$. Hence, if the option data is more efficient for estimation of β, it is also more efficient for recovery of $\boldsymbol{A}$. In this case, all components of $\widehat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{S}}$ converge at the rate $\sqrt{N}$. By contrast, if $p_n \gg N$, then $n \gg N$, so the high-frequency data is more informative about both the jump activity and intensity, each component of the partitioned parameter vector and state vector realization will converge at different rates.

Importantly, our PLS estimators of β and $\boldsymbol{A}$ automatically adapt to the situation at hand. When the high-frequency data is more informative than the option data ($p_n \gg N$ or $n \gg N$), then the PLS estimator for these quantities is asymptotically equivalent to their nonparametric high-frequency measures. On the other hand, when the option data (together with the parametric model) carries more information than the high-frequency return data about either β or $\boldsymbol{A}$ ($N \gg n$ or $N \gg p_n$), then the corresponding PLS estimator behaves as if only the option data is used for the estimation of this quantity. Consequently, the user does not need to take an a priori stand on whether the option or the high-frequency data is more informative about β or $\boldsymbol{A}$, which is very convenient from a practical point of view. In the boundary cases of either $N \asymp n$ or $N \asymp p_n$, both the option and high-frequency return data contribute to the estimation of (parts of) the parameters and state vectors. In this case, one may choose λ_β and λ_A in a way that accounts for the difference in the variance of the option errors and the asymptotic variances of the high-frequency estimators. This generates further gains in efficiency and renders the PLS estimator free of tuning parameters (other than those needed for the construction of the nonparametric high-frequency estimators). We present the details of such adaptive choices for λ_β and λ_A in the next section.

In a typical application, the state vector includes separate intensities A_t^+ and A_t^-, which, in turn, may be determined by additional factors, in analogy with multi-factor stochastic volatility models. In this case, if $p_n \gg N$, then A_t^+ and A_t^- will each be estimated at the slower rate $\sqrt{N}$, and their joint distribution will be degenerate. Their sum, however, $A_t = A_t^+ + A_t^-$ is estimated at the faster rate $\sqrt{p_n}$. In our statement of Theorem 3, we reparametrize the state vector through separating A_t in a manner so as to avoid degeneracy of the limiting distribution. This enables one to characterize the limiting distribution of arbitrary transformations of the state vector. This situation is similar to other econometric settings, where the convergence rates of components within a joint system may differ, e.g., inference for regressions with integrated processes, see, e.g., Park and Phillips (1988, 1989), Phillips (1988), and Sims, Stock, and Watson (1990).

Finally, the asymptotic distribution of both the parameters and the state vector is generally mixed Gaussian. That is, the matrices $\mathcal{I}$ and $\mathbf{\Omega}$ are likely random. This is due to the mixed-Gaussian distribution for the estimates of $\boldsymbol{A}$ from the high-frequency data as well as the conditional heteroskedasticity in the option observation error. Since the convergence in Theorem 3 is stable, this, however, does not constitute a major practical difficulty. All that is needed for feasible inference based on the limit result in Theorem 3 is consistent estimators for $\mathcal{I}$ and $\mathbf{\Omega}$, which are easy to construct directly from least squares procedures; see Section A.10 for the details.

5. WEIGHTED PENALIZED LEAST SQUARES

The definition of the PLS estimator in Section 4.1 involves the penalty weights λ_β and λ_A. We now propose suitable selection procedures for these values, period-by-period, that generate efficiency improvements. Moreover, we discuss how to weight the elements of the L_2 part of the objective function in a manner analogous to classical weighted least squares. We label the combination of such weighting with the suitable selection of the λ_β and λ_A the weighted PLS (WPLS) estimator.

First, let $\hat{\Psi}_\beta$ and $\hat{\Psi}_t$, $t = 1, \ldots, T$, be plug-in estimators of Ψ_β and Ψ_t, respectively, where we recall that $\mathbf{\Psi}_A = \text{diag}(\Psi_1, \ldots, \Psi_T)$, and further note that the plug-in estimators are defined explicitly in Section A.4. We then readily obtain that $\hat{\Psi}_\beta \xrightarrow{\mathbb{P}} \Psi_\beta$ and $\hat{\Psi}_t \xrightarrow{\mathbb{P}} \Psi_t$ from Theorem 3 in conjunction with the continuous mapping theorem. Now, since the size of the $\mathcal{F}$-conditional variance of the errors stemming from the two penalization terms generally are unknown a priori, we propose to standardize their contribution to the objective function through estimates of the $\mathcal{F}$-conditional asymptotic variances of the nonparametric estimators from high-frequency data, provided by Theorem 1. This will imply that their respective contributions to the objective function are similar in scale.

Next, concerning the optimal weighting of the elements in the option part of $\mathcal{L}_t(\boldsymbol{Z}_t, \boldsymbol{\theta})$, we ideally would like to standardize these by an estimate of the $\mathcal{F}$-conditional variance of the BSIV observation errors in equation (11), defined by $\phi_{t,k,\tau}$ in Assumption A.6 of Section A.1. However, despite such a procedure being feasible, we simplify the analysis and assign identical weights to all options on a given day. Although this approach neglects potential heteroskedasticity in the strike and tenor dimensions of the option panel, it still generates nontrivial efficiency improvements due to pronounced heteroskedasticity in the $\mathcal{F}$-conditional option error variances over time. Moreover, it is sufficient to ensure that all components of the (weighted) objective function are of comparable scale. Formally, we use,

$$\hat{\phi}_t = \frac{1}{N_t} \sum_{j=1}^{N_t} \left(\hat{\kappa}_{t,k_j,\tau_j} - \kappa(k_j, \tau_j, \widehat{\boldsymbol{S}}_t, \widehat{\boldsymbol{\theta}})\right)^2, \qquad t = 1, \ldots, T, \tag{21}$$

where $\widehat{\boldsymbol{S}}_t$ and $\widehat{\boldsymbol{\theta}}$ are based on first-stage PLS estimation.[11] As one would expect, $\hat{\phi}_t$ is a consistent estimator of the cross-sectional average of $\phi_{t,k,\tau}$, which is generally random, at a given point in time.

Now, using $\hat{\phi}_t$, $\hat{\Psi}_\beta$, and $\hat{\Psi}_t$, we define the WPLS objective function as,

$$\mathcal{L}^w(\boldsymbol{Z}, \boldsymbol{\theta}) \equiv \sum_{t=1}^{T} \mathcal{L}_t^w(\boldsymbol{Z}_t, \boldsymbol{\theta}) + nT \frac{(\widehat{\beta} - \mathcal{B})^2}{w(\hat{\Psi}_\beta)}, \qquad \text{with} \tag{22}$$

$$\mathcal{L}_t^w(\boldsymbol{Z}_t, \boldsymbol{\theta}) \equiv \left\{ \sum_{j=1}^{N_t} \frac{\left(\hat{\kappa}_{t,k_j,\tau_j} - \kappa(k_j, \tau_j, \boldsymbol{Z}_t, \boldsymbol{\theta})\right)^2}{w(\hat{\phi}_t)} + p_n \frac{\left(\widehat{A}_t - \mathcal{A}_t\right)^2}{w(\hat{\Psi}_t)} \right\},$$

where the function $w(x) \geq \epsilon$, for some $\epsilon > 0$, is a twice differentiable function on $\mathbb{R}_+$ with bounded first and second derivatives. Smooth approximations of $x \vee \epsilon$ are examples of such functions. Ideally, we would like to choose $w(x) = x$, but we rule this case out when developing our general distribution theory for WPLS to avoid imposing boundedness from below on $\phi_{t,k,\tau}$ as well as on the asymptotic variances Ψ_β and Ψ_t. Nonetheless, we consider this scenario in a corollary below, which results in a simplification of the expression for the limiting distribution.

Given the objective function in equation (22), the WPLS estimator is defined as,

$$(\widehat{\boldsymbol{\theta}}^w, \widehat{\boldsymbol{S}}^w) = \underset{\boldsymbol{\theta} \in \boldsymbol{\Theta}, \boldsymbol{Z} \in \boldsymbol{\mathcal{S}}^T}{\operatorname{argmin}} \; \mathcal{L}^w(\boldsymbol{Z}, \boldsymbol{\theta}), \qquad \boldsymbol{\mathcal{S}} \subset \mathbb{R}^p. \tag{23}$$

[11] A natural candidate is the estimator without penalization, i.e., one that is purely option based, as this circumvents the issue of choosing the relative weight of the penalty terms.

The procedure of weighting the first part of the criterion function by the size of the average errors at each observation time for the option panel, using equation (21), is reminiscent of the approach in Andersen et al. (2018). As such, it should provide similar benefits in terms of efficiency gains. By contrast, the importance of additionally using $\hat{\Psi}_\beta$ and $\hat{\Psi}_t$ are much larger in our pure-jump setting. This follows from the fact that the "regularization" devices naturally are of a different scale in the pure-jump setting. For the diffusive case, the noisy spot variance measure is automatically scaled sensibly, as the options (quoted in BSIV), and therefore, all parts of the PLS objective function in this case are in terms of "return variance measures." However, this is not true for equation (18), where the three components reflect return variances, their jump activity index and their jump intensities. Hence, the use of the weighted objective function (23) will generate a more stable numerical estimation procedure, in addition to providing asymptotic efficiency gains. Finally, we emphasize that the weighting in equation (22) is only feasible due to our stable central limit theory in Theorem 3, allowing for estimation and utilization of weights that are asymptotically random.

We are now in position to state our asymptotic distribution result.

THEOREM 4. *Suppose the conditions of Theorem 3 hold. Moreover, let* $\widehat{\boldsymbol{\theta}}^w = (\widehat{\boldsymbol{\theta}}_r^w, \widehat{\mathcal{B}}^w)$ *and* $\widehat{\boldsymbol{S}}^w = (\widehat{\boldsymbol{S}}_r^w, \widehat{\mathcal{A}}^w)$ *denote the WPLS estimators of* $\boldsymbol{\theta}_0 = (\boldsymbol{\theta}_0^r, \beta)$ *and* $\boldsymbol{S} = (\boldsymbol{S}^r, \boldsymbol{A})$*, respectively, then a convergence result similar to that in Theorem 3 holds as long as* $\mathcal{I}$ *and* $\boldsymbol{\Omega}$ *are replaced with* $\mathcal{I}^w$ *and* $\boldsymbol{\Omega}^w$*, which are defined in equation* (*A.11*) *of Section A.4.*

The special case where Ψ_β, Ψ_t and $\phi_{t,k,\tau}$ are bounded (uniformly) from below is given in the following corollary.

COROLLARY 1. *Suppose the conditions of Theorem 4 hold and, in addition, that the following lower bounds are satisfied,* $\Psi_\beta > \epsilon$, $\inf_{t\in 1,\dots,T} \Psi_t > \epsilon$, *and,*

$$\inf_{t\in 1,\dots,T} \inf_{\tau\in\mathcal{T}_t} \inf_{k\in[\underline{k}(t,\tau),\bar{k}(t,\tau)]} \phi_{t,k,\tau} > \epsilon, \quad \text{for some finite } \epsilon > 0, \quad \text{with} \quad \phi_{t,k,\tau} = \phi_t.$$

Finally, letting $w(x) = x$*, it then follows that,*

$$\boldsymbol{W}_n^{-1} \begin{pmatrix} \widehat{\boldsymbol{\theta}}_t^w - \boldsymbol{\theta}_0^r \\ \widehat{\mathcal{B}}^w - \beta \\ \widehat{\boldsymbol{S}}_r^w - \boldsymbol{S}^r \\ \widehat{\mathcal{A}}^w - \boldsymbol{A} \end{pmatrix} \xrightarrow{\mathcal{L}-s} (\mathcal{I}^w)^{-1/2} \times \begin{pmatrix} \mathbf{E}_{\theta_0^r} \\ \mathrm{E}_\beta \\ \mathbf{E}_{S^r} \\ \mathbf{E}_A \end{pmatrix},$$

where E_β *and the* $(q-1)\times 1$, $T(p-1)\times 1$, *and* $T\times 1$ *vectors* $\mathbf{E}_{\theta_0^r}$, $\mathbf{E}_{S^r}$, *and* $\mathbf{E}_A$, *respectively, consist of standard Gaussian random variables defined on an extension of the original probability space, independent of each other as well as of* $\mathcal{F}$, *and* $\mathcal{I}^w$ *is defined in equation* (*A.11*) *of Section A.4.*

6. MONTE CARLO STUDY

We next assess the performance of the PLS-based inference procedures in finite samples. To this end, we set up a Monte Carlo study and simulate a parametric model according to equations (1) and (5), with $\alpha_t = r_t = q_t = 0$ and,

$$\nu^{\mathbb{P}}(dt,dx) = \nu^{\mathbb{Q}}(dt,dx) = A_t A_\beta \frac{e^{-\lambda|x|}}{|x|^{\beta+1}} dt \otimes dx, \tag{24}$$

where A_β is the function of β, given in (3), and,

$$dA_t = -\kappa A_t dt + dL_t, \tag{25}$$

with L_t being an Inverse Gaussian process, independent of the jump measure μ, and having parameters c_L and μ_L. Recall that the Inverse Gaussian process is a Lévy process and its characteristic function is given by $\mathbb{E}(e^{iuL_t}) = \exp\left(-2c_L t\sqrt{\pi}(\sqrt{\mu_L - iu} - \sqrt{\mu_L})\right)$. The specification for A_t in (25) is a non-Gaussian Ornstein-Uhlenbeck process, similar to the ones used in Barndorff-Nielsen and Shephard (2001) for modeling volatility. Hence, using the notation of the theoretical section, we have $\boldsymbol{\theta}^r = (\lambda, \kappa, c_L, \mu_L)$ and $\boldsymbol{Z}_t^r = \varnothing$. Moreover, the true values of the parameters are set to $\beta = 1.5$ and $\boldsymbol{\theta}_0^r = (15, 3, 1.415, 20)$ (quoted in annualized terms). This corresponds to having an annual average variance of x of 0.16^2 and the half-life of a shock to A_t being approximately two months.

We sample a cross-section of option prices at the end of each week over a period of two months. This amounts to 8 cross-sections, each of which consists of $N = 120$ option prices with 4 tenors. For each tenor, we have 30 options on an equidistant log-strike grid covering $[-4\sigma_t^{ATM}\sqrt{\tau}, 4\sigma_t^{ATM}\sqrt{\tau}]$, where σ_t^{ATM} denotes the ATM BSIV at time t. The option error is specified as $\epsilon_j = 0.02 \times \kappa_{t_j,k_j} \times z_j$, where $\{z_j\}_{j=1}^{N_t}$ is a sequence of i.i.d. standard normal random variables, implying that average absolute relative error (in terms of BSIV) is approximately 2%. This option sampling setup mimics available option data, see, e.g., Andersen et al. (2015) and Andersen et al. (2018).

The second source of information being utilized in the estimation of the model is high-frequency return data. In particular, we let $n = 300$, corresponding approximately to sampling the stock price every 5 minutes during a 24-hour trading day. The local window for estimating the power variation in the construction of the high-frequency jump activity estimator is set to $k_n = 100$. Finally, for the construction of the jump activity estimator, we follow Todorov (2015) and set the power to $p = 0.51$ as well as the arguments of the characteristic function to $u = 0.3$ and $v = 1.0$.

For simplicity, we refrain from using high-frequency estimates of $\widehat{\boldsymbol{A}}$ in the estimation nor do we consider optimal weighting. Instead, we set the penalization parameter λ_β to the ratio of the average option variance across the strikes and tenors used in the estimation and the asymptotic variance of the high-frequency

TABLE 1. Monte Carlo results

Panel A

	Coverage rate of two-sided confidence interval					
	No penalization			Penalization		
Parameter	**99%**	**95%**	**90%**	**99%**	**95%**	**90%**
λ	98.90	94.30	88.70	98.70	94.10	87.10
k	98.10	94.90	90.50	98.10	95.30	90.30
c_L	97.70	94.30	90.10	98.00	95.30	90.40
μ_L	97.20	93.50	89.80	96.80	94.00	90.00
β	99.20	95.60	91.30	99.00	95.30	90.30
A_t	99.10	96.05	90.54	98.71	95.03	90.45

Panel B

		No penalization		Penalization	
Parameter	True	Bias	RMSE	Bias	RMSE
λ	15.000	0.0280	0.3783	0.0005	0.3361
k	3.000	0.0179	0.4938	0.0217	0.4915
c_L	1.415	0.0185	0.2386	0.0137	0.2294
μ_L	20.000	0.2013	2.7528	0.2688	2.7514
β	1.500	−0.0012	0.0154	0.0000	0.0133
A_t		0.0015	0.0136	0.0003	0.0116

Notes: Monte Carlo results are based on 1,000 draws.

estimator $\widehat{\beta}$, with the averages determined via simulation.[12] The results from the Monte Carlo exercise are presented in Table 1. From panel A of the table, we see that the empirical coverage rates for standard two-sided confidence intervals of each parameter as well as the jump intensity realizations are very close to their nominal levels. Panel B of Table 1 further reveals that all parameters and the jump intensity realizations are recovered without any significant biases. This applies to estimation both with and without penalization. A comparison of the root mean squared error (RMSE) for parameter and jump intensity realizations across the two estimation methods quantifies the gains from incorporating information from the high-frequency return data. Not surprisingly, the biggest efficiency gain is for the recovery of the jump activity parameter β and the jump intensity realizations, for which the reduction in the RMSE is around 13%. However, the more efficient recovery of β also generate a "spillover" effect for the other parameters, most notably for the parameter λ, controling the behavior of the "big" jumps, which becomes easier to disentangle from β.

Overall, the Monte Carlo study documents good finite sample performance of the proposed inference procedures.

[12] This choice of λ_β will correspond to the optimal weight in Corollary 1 if ϕ_t did not depend on time.

https://doi.org/10.1017/S0266466618000373 Published online by Cambridge University Press

7. CONCLUSION

In this article, we develop inference techniques for noisy option panels with a fixed time span and an asymptotically increasing cross-sectional dimension in which the option prices are generated from a parametric model for the risk-neutral dynamics of the underlying asset that is of pure-jump type. The option-based information set utilized in the estimation is augmented by high-frequency return data, covering the time span of the option panel. The return data is used to construct nonparametric measures of the jump activity parameter as well as the vector of jump intensity realizations at the integer times, where the cross-sections of the option panel are observed. Estimation of the risk-neutral parameters and the state vector realizations of the model is carried out via penalized least squares, minimizing the L_2 distance between observed and model-implied option prices while penalizing deviations of the model-implied jump activity and jump intensities from nonparametric estimates of them based on the high-frequency data. The distribution theory for estimates of different parts of the parameter and state vectors differs depending on the relative informativeness of the high-frequency return data (through the nonparametric jump measures) and the option data (via the parametric model). Importantly, our PLS estimator adapts to the situation at hand without any need for a priori assessment of what data source is more efficient for estimation. In addition, while the asymptotic distributions may appear complex, involving mixed-Gaussian limiting distributions and stable convergence, the application of our theory for practical inference is relatively straightforward, involving only quantities that arise naturally from estimation through nonlinear least squares.

The results complement corresponding inference techniques for noisy option panels developed in Andersen et al. (2015) and Andersen et al. (2018) for the case where the asset returns are governed by a jump-diffusion. These procedures impose the restriction that the vector of diffusive spot volatility realizations is invariant across the risk-neutral and statistical measure. This property is replaced by the analogous restriction that the jump intensity at each observation time is identical across the two measures in the pure jump case. The additional constraint that the jump activity index is identical across the two measures has no parallel in the diffusive scenario. The imposition of these no-arbitrage conditions as an integral part of a formal inference procedure for the risk-neutral dynamics and the state vector realizations from noisy option and high-frequency return observations is novel.

Finally, we give an extension of the PLS estimator involving weighting of the individual terms in the objective function by their asymptotic variances. This WPLS estimator provides additional robustness and efficiency, which is likely more critical in the pure-jump setting than for the jump-diffusive models explored through similar techniques previously in the literature.

The results of the article should be of direct use for estimating continuous-time stochastic volatility models of the pure-jump type and for studying the associated risk premiums. Local equivalence between the statistical and risk-neutral probability measures imposes restrictions between the parameters and state variables that contain important information about both risks and risk premiums. Unlike prior work, we incorporate this information directly into our procedure, which leads to easy-to-implement techniques that optimally combines return and option data for efficient inference.

REFERENCES

Ait-Sahalia, Y. & J. Jacod (2009) Estimating the degree of activity of jumps in high frequency financial data. *Annals of Statistics* 37, 2202–2244.

Andersen, T.G., N. Fusari, & V. Todorov (2015) Parametric inference and dynamic state recovery from option panels. *Econometrica* 83, 1081–1145.

Andersen, T.G., N. Fusari, V. Todorov, & R.T. Varneskov (2018) Unified inference for nonlinear factor models from panels with fixed and large time span. *Journal of Econometrics*, forthcoming.

Barndorff-Nielsen, O.E. & N. Shephard (2001) Non-gaussian ornstein-uhlenbeck-based models and some of their applications in financial economics. *Journal of the Royal Statistical Society*: *Series B* 63, 167–241.

Bernanke, B., J. Boivin, & P. Eliasz (2005) Measuring the effects of monetary policy: A Factor-Augmented Vector Autoregressive (FAVAR) Approach. *Quarterly Journal of Economics* 120, 387–422.

Bull, A. (2016) Near-optimal estimation of jump activity in semimartingales. *Annals of Statistics* 44, 58–86.

Carr, P., H. Geman, D. Madan, & M. Yor (2002) The fine structure of asset returns: An empirical investigation. *Journal of Business* 75, 305–332.

Carr, P., H. Geman, D.B. Madan, & M. Yor (2003) Stochastic volatility for Lévy processes. *Journal of Business* 75, 345–382.

Carr, P. & D. Madan (2001) Optimal positioning in derivative securities. *Quantitiative Finance* 1, 19–37.

Carr, P. & L. Wu (2003) The finite moment log stable process and option pricing. *Journal of Finance* LVIII, 753–778.

Carr, P. & L. Wu (2004) Time-changed Lévy processes and option pricing. *Journal of Financial Economics* 17, 113–141.

Duffie, D. (2001) *Dynamic Asset Pricing Theory*, 3rd ed. Princeton University Press.

Duffie, D., J. Pan, & K. Singleton (2000) Transform analysis and asset pricing for affine jump-diffusions. *Econometrica* 68, 1343–1376.

Fama, E. (1963) Mandelbrot and the stable paretian hypothesis. *Journal of Business* 36, 420–429.

Fama, E. & R. Roll (1968) Some properties of symmetric stable distributions. *Journal of the American Statistical Association* 63, 817–836.

Filipović, D. (2001) A general characterization of one factor affine term structure models. *Finance Stochastic* 5, 389–412.

Hounyo, U. & R.T. Varneskov (2017) A local stable bootstrap for power variations of pure-jump semimartingales and activity index estimation. *Journal of Econometrics* 198, 10–28.

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

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

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

Jing, B., X. Kong, & Z. Liu (2011) Estimating the jump activity index under noisy observations using high-frequency data. *Journal of the American Statistical Association* 106, 558–568.

Jing, B., X. Kong, & Z. Liu (2012) Modeling high-frequency financial data by pure jump processes. *Annals of Statistics* 40, 759–784.
Jing, B., X. Kong, Z. Liu, & P. Mykland (2012) On the jump activity index of semimartingales. *Journal of Econometrics* 166, 213–223.
Kalnina, I. & D. Xu (2017) Nonparametric estimation of the leverage effect: A trade-off between robustness and efficiency. *Journal of the American Statistical Association* 112, 384–396.
Kong, X., Z. Liu, & B. Jing (2015) Testing for pure-jump processes for high-frequency data. *Annals of Statistics* 43, 847–877.
Madan, D. & F. Milne (1991) Option pricing with VG martingale components. *Mathematical Finance* 1, 39–56.
Madan, D. & E. Seneta (1990) The variance gamma (VG) model for share market returns. *Journal of Business* 63, 511–524.
Mandelbrot, B. (1961) Stable paretian random functions and the multiplicative variation of income. *Econometrica* 29, 517–543.
Mandelbrot, B. (1963) The variation of certain speculative prices. *Journal of Business* 36, 394–419.
Park, J.Y. & P.C.B. Phillips (1988) Statistical inference in regressions with integrated processes: Part 1. *Econometric Theory* 4, 468–497.
Park, J.Y. & P.C.B. Phillips (1989) Statistical inference in regressions with integrated processes: Part 2. *Econometric Theory* 5, 95–132.
Phillips, P.C.B. (1988) Weak convergence of sample covariance matrices to stochastic integrals via martingale approximations. *Econometric Theory* 4, 528–533.
Qin, L. & V. Todorov (2018) Nonparametric implied Lévy densities. *Annals of Statistics*, forthcoming.
Rosinski, J. (2007) Tempering stable processes. *Stochastic Processes and Applications* 117, 677–707.
Sato, K. (1999) *Lévy Processes and Infinitely Divisible Distributions*. Cambridge University Press.
Sims, C., J. Stock, & M. Watson (1990) Inference in linear time series models with some unit roots. *Econometrica* 58, 113–144.
Todorov, V. (2015) Jump activity estimation for pure-jump semimartingales via self-normalized statistics. *Annals of Statistics* 43, 1831–1864.
Todorov, V. & G. Tauchen (2011a) Limit theorems for power variations of pure-jump processes with application to activity estimation. *The Annals of Applied Probability* 21, 546–588.
Todorov, V. & G. Tauchen (2011b) Volatility jumps. *Journal of Business and Economic Statistics* 29, 356–371.
Todorov, V. & G. Tauchen (2012) Realized laplace transforms for pure-jump semimartingales. *Annals of Statistics* 40, 1233–1262.
Woerner, J. (2003) Variational sums and power variation: A unifying approach to model selection and estimation in semimartingale models. *Statistics and Decisions* 21, 47–68.
Woerner, J. (2007) Inference in Lévy-type stochastic volatility models. *Advances in Applied Probability* 39, 531–549.

APPENDIX

This section states the formal assumptions for the theoretical analysis and provides proofs of the asymptotic results. Furthermore, we outline how to feasibly implement the inference procedures as well as give details on the computation of the option prices in the Monte Carlo study.

Before proceeding, let us introduce some convenient notation. We adopt the shorthand notation, $\widehat{\kappa}_{t,k_j,\tau_j} \equiv \widehat{\kappa}_{t,j}$, $\epsilon_{t,k_j,\tau_j} \equiv \epsilon_{t,j}$, and $\kappa(k_j,\tau_j,\boldsymbol{Z},\boldsymbol{\theta}) \equiv \kappa_j(\boldsymbol{Z},\boldsymbol{\theta})$. The Hadamard product is indicated by $\circ$; and the matrix norm used throughout is the Frobenius (or Euclidean) norm which, for an $m \times n$ dimensional matrix $\boldsymbol{A}$, may be written as $||\boldsymbol{A}|| = \sqrt{\sum_{i,j} a_{i,j}^2} = \sqrt{\mathrm{Tr}(\boldsymbol{A}\boldsymbol{A}')}$. Moreover, K denotes a generic constant, which may take different values in different places, and we signify conditional expectations by $\mathbb{E}_i^n(\cdot) \equiv \mathbb{E}(\cdot|\mathcal{F}_{i\Delta_n})$. Note that (stochastic) orders sometimes refer to scalars, vectors, and sometimes to matrices; we refrain from making distinctions among these. Finally, let $(E, \mathcal{E})$ denote an auxiliary measure space on the original filtered probability space $\big(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P}\big)$.

A.1. Assumptions

Assumption A.1 (Price process)**.** The price process of the underlying asset X_t satisfies the conditions (1)–(3) of Section 2.1. Moreover, letting $q_t = \{\alpha_t, A_t^+, A_t^-\}$, these processes obey,

$$q_t = q_0 + \int_0^t b_s^q \, ds + \int_0^t \int_E \kappa(\delta^q(s,x)) \, \underline{\tilde{\vartheta}}(ds,dx) + \int_E \kappa'(\delta^q(s,x)) \underline{\vartheta}(ds,dx), \quad \textbf{(A.1)}$$

where $\kappa(x) = x$ is the usual truncation function, for which $\kappa(-x) = -\kappa(x)$ and $\kappa'(x) = x - \kappa(x)$. The process (A.1) and its remaining components satisfy the following:

(i) $|q_t|^{-1}$ and $|q_{t-}|^{-1}$ are strictly positive;

(ii) $\underline{\tilde{\vartheta}}$ is the associated martingale measure of $\underline{\vartheta}$, which is a Poisson measure on $\mathbb{R}_+ \times E$, having arbitrary dependence with the jump measure μ, equipped with compensator $dt \otimes \underline{\lambda}(dx)$ for some σ-finite measures $\underline{\lambda}$ on E;

(iii) let $\gamma_k(x)$ be a deterministic function on $\mathbb{R}$ with $\int_{\mathbb{R}} (|\gamma_k(x)|^{r+\iota} \wedge 1) \, \underline{\lambda}(dx) < \infty$ for some arbitrarily small $\iota > 0$ and some $0 \leq r \leq \beta$, and furthermore let T_k be a sequence of stopping times increasing to $+\infty$, then $\delta^q(t,x)$ is assumed to be predictable, left-continuous with right limits in t, and with $|\delta^q(t,x)| \leq \gamma_k(x)$ for all $t \leq T_k$;

(iv) b_t^q is an Itô semimartingale having dynamics as specified in equation (A.1) with coefficients satisfying conditions analogous to conditions (ii) and (iii) above.

Assumption A.2 (Sampling scheme)**.** As $N \to \infty$, $p_n \to \infty$, and $n \to \infty$ with $p_n/n \to 0$, as well as with $\bar{n} = n \vee N$ and $\overline{p}_n = p_n \vee N$, we have for each $t = 1, \ldots, T$ and each maturity $\tau \in \mathcal{T}_t$ that,

(i) $N_t^\tau / N_t \xrightarrow{\mathbb{P}} \pi_t^\tau$ and $N_t/N \xrightarrow{\mathbb{P}} \varrho_t$ where π_t^τ and ϱ_t are adapted to $\mathcal{F}_t^{(0)}$ with $\inf_{t\in[1,T], \tau\in\mathcal{T}_t} \pi_t^\tau > 0$ and $\sup_{t\in[1,T], \tau\in\mathcal{T}_t} \pi_t^\tau < \infty$ as well as $\inf_{t\in[1,T]} \varrho_t > 0$ and $\sup_{t\in[1,T]} \varrho_t < \infty$.

(ii) For the grids of strike prices, let $i_k = \min\{i \geq 2 : k_{t,\tau}(i) \geq k\}$, then uniformly for each $k \in [\underline{k}(t,\tau), \bar{k}(t,\tau)]$, we have $N_t^\tau \Delta_{t,\tau}(i_k) \xrightarrow{\mathbb{P}} \psi_{t,\tau}(k)$, where $\psi_{t,\tau}(k)$ is some $\mathcal{F}_t^{(0)}$-adapted process with,

$$\inf_{t\in[1,T],\, \tau\in\mathcal{T}_t,\, k\in[\underline{k}(t,\tau), \bar{k}(t,\tau)]} \psi_{t,\tau}(k) > 0, \quad \text{and} \quad \sup_{t\in[1,T],\, \tau\in\mathcal{T}_t,\, k\in[\underline{k}(t,\tau), \bar{k}(t,\tau)]} \psi_{t,\tau}(k) < \infty.$$

(iii) Finally, we have the following finite relative limits for N, p_n, n, $\bar{n}$, and $\overline{p}_n$,

$$\frac{N}{\bar{n}} \to \varpi_1 \geq 0, \qquad \frac{n}{\bar{n}} \to \varpi_2 \geq 0, \qquad \frac{N}{\overline{p}_n} \to \zeta_1 \geq 0, \qquad \text{and} \qquad \frac{p_n}{\overline{p}_n} \to \zeta_2 \geq 0.$$

https://doi.org/10.1017/S0266466618000373 Published online by Cambridge University Press

Assumption A.3 (Identification)**.** For every $\epsilon > 0$ and $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, we have, almost surely, for N sufficiently large,

$$\inf_{(\bigcap_{t=1}^{T}\{\|\boldsymbol{Z}_t - \boldsymbol{S}_t\|\}\cap\{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\le\epsilon\})^c} \sum_{t=1}^{T}\sum_{j=1}^{N_t} \frac{\left(\kappa(k_j,\tau_j,\boldsymbol{S}_t,\boldsymbol{\theta}_0) - \kappa(k_j,\tau_j,\boldsymbol{Z}_t,\boldsymbol{\theta})\right)^2}{N_t} > 0.$$

Assumption A.4 (Differentiability)**.** The function $\kappa(\tau,k,\boldsymbol{Z},\boldsymbol{\theta})$ is twice continuously differentiable in its arguments.

Assumption A.5 (Observation error: Consistency)**.** For every $\epsilon > 0$, $t = 1,\ldots,T$, and any positive-valued $\mathcal{F}_T^{(0)}$-adapted process $\zeta_t(k,\tau)$ on the product space $\mathbb{R}\times\mathcal{T}_t$, which is continuous in its first argument, we have for $N \to \infty$ and $\boldsymbol{\theta} \in \boldsymbol{\Theta}$,

$$\sup_{\{\|\boldsymbol{Z}_t - \boldsymbol{S}_t\|>\epsilon\}\cup\{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|>\epsilon\}} \frac{\sum_{j=1}^{N_t}\zeta_t(k,\tau)\left(\kappa(k_j,\tau_j,\boldsymbol{S}_t,\boldsymbol{\theta}_0) - \kappa(k_j,\tau_j,\boldsymbol{Z}_t,\boldsymbol{\theta})\right)\epsilon_{t,k_j,\tau_j}}{\sum_{j=1}^{N_t}\left(\kappa(k_j,\tau_j,\boldsymbol{S}_t,\boldsymbol{\theta}_0) - \kappa(k_j,\tau_j,\boldsymbol{Z}_t,\boldsymbol{\theta})\right)^2} \xrightarrow{\mathbb{P}} 0.$$

Assumption A.6 (Observation error: Central limit theory)**.** For the error process, $\epsilon_{t,k,\tau}$, it follows:

(i) $\mathbb{E}(\epsilon_{t,k,\tau}|\mathcal{F}^{(0)}) = 0$,

(ii) $\mathbb{E}(\epsilon_{t,k,\tau}^2|\mathcal{F}^{(0)}) = \phi_{t,k,\tau}$, with $\phi_{t,k,\tau}$ being a continuous function in its second argument,

(iii) $\epsilon_{t,k,\tau}$ and $\epsilon_{t',k',\tau'}$ are independent conditional on $\mathcal{F}^{(0)}$, whenever $(t,k,\tau) \neq (t',k',\tau')$,

(iv) $\mathbb{E}(|\epsilon_{t,k,\tau}|^4|\mathcal{F}^{(0)}) < \infty$, almost surely.

Assumption A.7 (Invertibility of the Hessian Matrix)**.** The following matrix is positive definite almost surely:

$$\sum_{\tau}\pi_t^{\tau}\int_{\underline{k}(t,\tau)}^{\overline{k}(t,\tau)}\frac{1}{\psi_{t,\tau}(k)}\begin{pmatrix}\nabla_{\boldsymbol{\theta}}\kappa(k,\tau,\boldsymbol{S}_t,\boldsymbol{\theta}_0)\nabla_{\boldsymbol{\theta}'}\kappa(k,\tau,\boldsymbol{S}_t,\boldsymbol{\theta}_0) & \nabla_{\boldsymbol{\theta}}\kappa(k,\tau,\boldsymbol{S}_t,\boldsymbol{\theta}_0)\nabla_{\boldsymbol{Z}'}\kappa(k,\tau,\boldsymbol{S}_t,\boldsymbol{\theta}_0)\\ \nabla_{\boldsymbol{Z}}\kappa(k,\tau,\boldsymbol{S}_t,\boldsymbol{\theta}_0)\nabla_{\boldsymbol{\theta}'}\kappa(k,\tau,\boldsymbol{S}_t,\boldsymbol{\theta}_0) & \nabla_{\boldsymbol{Z}}\kappa(k,\tau,\boldsymbol{S}_t,\boldsymbol{\theta}_0)\nabla_{\boldsymbol{Z}'}\kappa(k,\tau,\boldsymbol{S}_t,\boldsymbol{\theta}_0)\end{pmatrix}dk.$$

These assumptions are similar to those in Andersen et al. (2015) and Todorov (2015) for the option panel and price process, respectively. The main departure is Assumption A.2(iii), which is needed to accommodate a central limit theorem with different rates of convergence for different parts of the parameter and state vector. Its impact is detailed in Section 4.3.

A.2. Definitions for the High-Frequency Estimators

This section provides additional details for the activity index and jump intensity estimators, both for their definitions and for developing their joint asymptotic theory.

Exact expression for $\mathcal{B}_n(p,u,\beta)$. First, let S_β be a β-stable random variable with characteristic function $\mathbb{E}(e^{iuS_\beta}) = \exp(-|u|^\beta)$ and denote $\mu_{p,\beta} = (\mathbb{E}|S_\beta|^p)^{\beta/p}$. With this notation, we set,

$$\varsigma(p,u,\beta) = \left(\cos\left(\frac{uS_\beta}{\mu_{p,\beta}^{1/\beta}}\right) - \mathcal{C}(p,u,\beta), \quad \frac{|S_\beta|^p}{\mu_{p,\beta}^{p/\beta}} - 1\right)', \quad u \in \mathbb{R}_+,$$

where the standardized characteristic function $\mathcal{C}(p,u,\beta)$ is defined as,

$$\mathcal{C}(p,u,\beta) = e^{-C_{p,\beta}u^{\beta}}, \qquad \text{with} \quad C_{p,\beta} = \left[\frac{2^p\,\Gamma((1+p)/2)\,\Gamma(1-p/\beta)}{\sqrt{\pi}\,\Gamma(1-p/2)}\right]^{-\beta/p}, \tag{A.2}$$

and $\Gamma(\cdot)$ being the gamma function. Next, for $u,v \in \mathbb{R}_+$, we then let,

$$\zeta(p,u,v,\beta) = \mathbb{E}\left(\varsigma(p,u,\beta)\varsigma(p,v,\beta)'\right),$$
$$G(p,u,\beta) = \frac{\beta}{p}e^{-C_{p,\beta}u^{\beta}}C_{p,\beta}u^{\beta}, \quad H(p,u,\beta) = G(p,u,\beta)\left(\frac{\beta}{p}C_{p,\beta}u^{\beta} - \frac{\beta}{p} - 1\right).$$

Finally, we may write the bias-correction $\mathcal{B}_n(p,u,\beta)$ as,

$$\mathcal{B}_n(p,u,\beta) = H(p,u,\beta)\zeta^{(2,2)}(p,u,u,\beta)/(2k_n). \tag{A.3}$$

Exact expressions for Ψ_β and Ψ_t. Using the definitions above, we may readily define the asymptotic variances for the nonparametric high-frequency measures in Theorem 1 as,

$$\Psi_\beta = \frac{2}{\log^2(u/v)}\left[\frac{\zeta^{(1,1)}(p,u,u,\beta)}{\log^2(\mathcal{C}(p,u,\beta))\mathcal{C}^2(p,u,\beta)} + \frac{\zeta^{(1,1)}(p,v,v,\beta)}{\log^2(\mathcal{C}(p,v,\beta))\mathcal{C}^2(p,v,\beta)} - 2\frac{\zeta^{(1,1)}(p,u,v,\beta)}{\log(\mathcal{C}(p,u,\beta))\mathcal{C}(p,u,\beta)\log(\mathcal{C}(p,v,\beta))\mathcal{C}(p,v,\beta)}\right], \tag{A.4}$$

$$\Psi_t = \frac{e^{2A_t u^{\beta}}}{u^{2\beta}}\left(\frac{1+e^{-2^{\beta}A_t u^{\beta}}}{2} - e^{-2A_t u^{\beta}}\right), \quad t = 1,\ldots,T. \tag{A.5}$$

A.3. Definitions for the Hessian and Asymptotic Variance

This section defines the empirical and limiting Hessian matrices, which are used in the proof and statement of Theorem 3. The definition of asymptotic covariance matrix in Theorem 3 is also given.

Empirical Hessian matrix. For generic values of $\boldsymbol{S}$ and $\boldsymbol{\theta}_0$, $\boldsymbol{Z}$ and $\boldsymbol{\theta}$, respectively, define the Hessian,

$$\boldsymbol{H}(\boldsymbol{Z},\boldsymbol{\theta}) \equiv \begin{pmatrix} \boldsymbol{H}_{\theta_0^r}(\boldsymbol{Z},\boldsymbol{\theta}) & \boldsymbol{H}_{\theta_0^r\beta}(\boldsymbol{Z},\boldsymbol{\theta}) & \boldsymbol{H}_{\theta_0^r S^r}(\boldsymbol{Z},\boldsymbol{\theta}) & \boldsymbol{H}_{\theta_0^r A}(\boldsymbol{Z},\boldsymbol{\theta}) \\ \boldsymbol{H}_{\theta_0^r\beta}(\boldsymbol{Z},\boldsymbol{\theta})' & \boldsymbol{H}_{\beta}(\boldsymbol{Z},\boldsymbol{\theta}) & \boldsymbol{H}_{\beta S^r}(\boldsymbol{Z},\boldsymbol{\theta}) & \boldsymbol{H}_{\beta A}(\boldsymbol{Z},\boldsymbol{\theta}) \\ \boldsymbol{H}_{\theta_0^r S^r}(\boldsymbol{Z},\boldsymbol{\theta})' & \boldsymbol{H}_{\beta S^r}(\boldsymbol{Z},\boldsymbol{\theta})' & \boldsymbol{H}_{S^r}(\boldsymbol{Z},\boldsymbol{\theta}) & \boldsymbol{H}_{S^r A}(\boldsymbol{Z},\boldsymbol{\theta}) \\ \boldsymbol{H}_{\theta_0^r A}(\boldsymbol{Z},\boldsymbol{\theta})' & \boldsymbol{H}_{\beta A}(\boldsymbol{Z},\boldsymbol{\theta})' & \boldsymbol{H}_{S^r A}(\boldsymbol{Z},\boldsymbol{\theta})' & \boldsymbol{H}_{A}(\boldsymbol{Z},\boldsymbol{\theta}) \end{pmatrix}, \tag{A.6}$$

whose elements along the diagonal, that is, the $(q-1)\times(q-1)$ matrix $\boldsymbol{H}_{\theta_0^r}(\boldsymbol{Z},\boldsymbol{\theta})$, the scalar $H_\beta(\boldsymbol{Z},\boldsymbol{\theta})$, the $T(p-1)\times T(p-1)$ matrix $\boldsymbol{H}_{S^r}(\boldsymbol{Z},\boldsymbol{\theta})$, and the $T\times T$ matrix $\boldsymbol{H}_A(\boldsymbol{Z},\boldsymbol{\theta})$, are defined as $\boldsymbol{H}_{S^r}(\boldsymbol{Z},\boldsymbol{\theta}) \equiv \operatorname{diag}\left(\boldsymbol{H}_{S_1^r}(\boldsymbol{Z}_1,\boldsymbol{\theta}),\ldots,\boldsymbol{H}_{S_T^r}(\boldsymbol{Z}_T,\boldsymbol{\theta})\right)$,

https://doi.org/10.1017/S0266466618000373 Published online by Cambridge University Press

$\boldsymbol{H}_A(\boldsymbol{Z},\boldsymbol{\theta}) \equiv \operatorname{diag}\left(H_{A_1}(\boldsymbol{Z}_1,\boldsymbol{\theta}),\ldots,H_{A_T}(\boldsymbol{Z}_T,\boldsymbol{\theta})\right)$, and with,

$$\begin{aligned}
\boldsymbol{H}_{\theta_0^r}(\boldsymbol{Z},\boldsymbol{\theta}) &\equiv \sum_{t=1}^{T}\sum_{j=1}^{N_t} \nabla_{\theta_0^r}\kappa_j(\boldsymbol{Z}_t,\boldsymbol{\theta})\,\nabla_{\theta_0^r}\kappa_j(\boldsymbol{Z}_t,\boldsymbol{\theta})',\\
H_\beta(\boldsymbol{Z},\boldsymbol{\theta}) &\equiv \sum_{t=1}^{T}\sum_{j=1}^{N_t} \nabla_\beta\kappa_j(\boldsymbol{Z}_t,\boldsymbol{\theta})\,\nabla_\beta\kappa_j(\boldsymbol{Z}_t,\boldsymbol{\theta})' + \lambda_\beta\, n\, T,\\
\boldsymbol{H}_{S_t^r}(\boldsymbol{Z}_t,\boldsymbol{\theta}) &\equiv \sum_{j=1}^{N_t} \nabla_{S^r}\kappa_j(\boldsymbol{Z}_t,\boldsymbol{\theta})\,\nabla_{S^r}\kappa_j(\boldsymbol{Z}_t,\boldsymbol{\theta})',\\
H_{A_t}(\boldsymbol{Z}_t,\boldsymbol{\theta}) &\equiv \sum_{j=1}^{N_t} \nabla_A\kappa_j(\boldsymbol{Z}_t,\boldsymbol{\theta})\,\nabla_A\kappa_j(\boldsymbol{Z}_t,\boldsymbol{\theta})' + \lambda_A\, p_n,
\end{aligned}$$

for $t = 1,\ldots,T$. The remaining elements of the $(q+Tp)\times(q+Tp)$ Hessian matrix (A.6) have the same generic structure as the explicated diagonal elements and are, thus, defined analogously.

Limiting Hessian matrix. The limiting Hessian matrix has the same block-wise structure as equation (A.6) and may be written,

$$\mathcal{I} = \mathcal{L}_1 \circ \mathcal{M} + \mathcal{L}_2 \circ \boldsymbol{\Gamma}, \tag{A.7}$$

where the first scaled matrix in the decomposition, $\mathcal{L}_1 \circ \mathcal{M}$, is defined as,

$$\mathcal{L}_1 \circ \mathcal{M} \equiv \begin{pmatrix} \mathcal{M}_{\theta_0^r} & \sqrt{\varpi_1}\mathcal{M}_{\theta_0^r\beta} & \mathcal{M}_{\theta_0^r S^r} & \sqrt{\zeta_1}\mathcal{M}_{\theta_0^r A}\\ \sqrt{\varpi_1}\mathcal{M}'_{\theta_0^r\beta} & \varpi_1\mathcal{M}_\beta & \sqrt{\varpi_1}\mathcal{M}_{\beta S^r} & \sqrt{\varpi_1\zeta_1}\mathcal{M}_{\beta A}\\ \mathcal{M}'_{\theta_0^r S^r} & \sqrt{\varpi_1}\mathcal{M}'_{\beta S^r} & \mathcal{M}_{S^r} & \sqrt{\zeta_1}\mathcal{M}_{S^r A}\\ \sqrt{\zeta_1}\mathcal{M}'_{\theta_0^r A} & \sqrt{\varpi_1\zeta_1}\mathcal{M}'_{\beta A} & \sqrt{\zeta_1}\mathcal{M}'_{S^r A} & \zeta_1\mathcal{M}_A \end{pmatrix}$$

where, e.g., the $(q-1)\times T$ matrix $\mathcal{M}_{\theta_0^r A} = \left(\mathcal{M}_{\theta_0^r A_1},\ldots,\mathcal{M}_{\theta_0^r A_T}\right)$ has column vectors,

$$\mathcal{M}_{\theta_0^r A_t} \equiv \varrho_t \sum_\tau \pi_t^\tau \int_{\underline{k}(t,\tau)}^{\bar{k}(t,\tau)} \frac{1}{\psi_{t,\tau}(k)} \nabla_{\theta_0^r}\kappa(k,\tau,\boldsymbol{S}_t,\boldsymbol{\theta}_0)\,\nabla_A\kappa(k,\tau,\boldsymbol{S}_t,\boldsymbol{\theta}_0)'\,dk,$$

for $t = 1,\ldots,T$. The remaining elements of $\mathcal{M}$ are defined similarly, the only change being the respective gradient arguments. The second term in the decomposition (A.7), $\mathcal{L}_2 \circ \boldsymbol{\Gamma}$, is given by

$$\mathcal{L}_2 \circ \boldsymbol{\Gamma} \equiv \operatorname{diag}\left(\boldsymbol{0}_{(q-1)\times 1},\, \varpi_2\lambda_\beta T,\, \boldsymbol{0}_{T(p-1)\times 1},\, \zeta_2\lambda_A\boldsymbol{\iota}_T\right),$$

with, again, $\boldsymbol{0}_d$ and $\boldsymbol{\iota}_d$ being d-dimensional vectors of zeros and ones, respectively.

Limiting covariance matrix. The $(q+Tp)\times(q+Tp)$ limiting covariance may be decomposed, similarly to equation (A.7), as,

$$\boldsymbol{\Omega} = \mathcal{L}_1 \circ \mathcal{C} + \mathcal{L}_2 \circ \boldsymbol{\Gamma} \circ \boldsymbol{\Psi}, \tag{A.8}$$

where, as above, the first scaled matrix in the decomposition, $\mathcal{L}_1 \circ \mathcal{C}$, is defined as,

$$\mathcal{L}_1 \circ \mathcal{C} \equiv \begin{pmatrix} \mathcal{C}_{\theta_0^r} & \sqrt{\varpi_1}\mathcal{C}_{\theta_0^r\beta} & \mathcal{C}_{\theta_0^r S^r} & \sqrt{\zeta_1}\mathcal{C}_{\theta_0^r A} \\ \sqrt{\varpi_1}\mathcal{C}_{\theta_0^r\beta}' & \varpi_1 \mathcal{C}_\beta & \sqrt{\varpi_1}\mathcal{C}_{\beta S^r} & \sqrt{\varpi_1\zeta_1}\mathcal{C}_{\beta A} \\ \mathcal{C}_{\theta_0^r S^r}' & \sqrt{\varpi_1}\mathcal{C}_{\beta S^r}' & \mathcal{C}_{S^r} & \sqrt{\zeta_1}\mathcal{C}_{S^r A} \\ \sqrt{\zeta_1}\mathcal{C}_{\theta_0^r A}' & \sqrt{\varpi_1\zeta_1}\mathcal{C}_{\beta A}' & \sqrt{\zeta_1}\mathcal{C}_{S^r A}' & \zeta_1 \mathcal{C}_A \end{pmatrix}$$

where, equivalently, the $(q-1) \times T$ matrix $\mathcal{C}_{\theta_0^r A} = \left(\mathcal{C}_{\theta_0^r A_1}, \ldots, \mathcal{C}_{\theta_0^r A_T}\right)$ has column vectors,

$$\mathcal{C}_{\theta_0^r A_t} \equiv \varrho_t \sum_\tau \pi_t^\tau \int_{\underline{k}(t,\tau)}^{\bar{k}(t,\tau)} \frac{\phi_{t,k,\tau}}{\psi_{t,\tau}(k)} \nabla_{\theta_0^r}\kappa(k,\tau,\boldsymbol{S}_t,\boldsymbol{\theta}_0)\, \nabla_A \kappa(k,\tau,\boldsymbol{S}_t,\boldsymbol{\theta}_0)'\, dk,$$

for $t = 1, \ldots, T$, and the remaining elements of $\mathcal{L}_1 \circ \mathcal{C}$ are defined similarly. The additional term in the second part of the decomposition (A.8), $\boldsymbol{\Psi}$, is given by,

$$\boldsymbol{\Psi} \equiv \operatorname{diag}\left(\mathbf{0}_{(q-1)\times 1}, \lambda_\beta \Psi_\beta, \mathbf{0}_{T(p-1)\times 1}, \lambda_A \Psi_1, \ldots, \lambda_A \Psi_T\right),$$

where Ψ_β and Ψ_t, for $t = 1, \ldots, T$, are defined as in Theorem 1.

A.4. Definitions for WPLS Estimation

This section defines the plug-in estimators for the WPLS objective function in equation (22). Moreover, it provides the limiting asymptotic variances for the WPLS estimator in Theorem 4 and Corollary 1.

Expressions for $\hat{\Psi}_\beta$ and $\hat{\Psi}_t$. Letting $\widehat{\mathcal{B}}$ and $\widehat{\mathcal{A}}_t$, $t = 1, \ldots, T$, be first-stage PLS estimates of β and A_t, respectively, then we define the plug-in estimators $\hat{\Psi}_\beta$ and $\hat{\Psi}_A$ as,

$$\hat{\Psi}_\beta = \frac{2}{\log^2(u/v)}\left[\frac{\zeta^{(1,1)}(p,u,u,\widehat{\mathcal{B}})}{\log^2(\mathcal{C}(p,u,\widehat{\mathcal{B}}))\mathcal{C}^2(p,u,\widehat{\mathcal{B}})} + \frac{\zeta^{(1,1)}(p,v,v,\widehat{\mathcal{B}})}{\log^2(\mathcal{C}(p,v,\widehat{\mathcal{B}}))\mathcal{C}^2(p,v,\widehat{\mathcal{B}})} \right.$$
$$\left. -2\frac{\zeta^{(1,1)}(p,u,v,\widehat{\mathcal{B}})}{\log(\mathcal{C}(p,u,\widehat{\mathcal{B}}))\mathcal{C}(p,u,\widehat{\mathcal{B}})\log(\mathcal{C}(p,v,\widehat{\mathcal{B}}))\mathcal{C}(p,v,\widehat{\mathcal{B}})}\right], \quad \textbf{(A.9)}$$

$$\hat{\Psi}_t = \frac{e^{2\widehat{\mathcal{A}}_t u^{\widehat{\mathcal{B}}}}}{u^{2\widehat{\mathcal{B}}}}\left(\frac{1+e^{-2^{\widehat{\mathcal{B}}}\widehat{\mathcal{A}}_t u^{\widehat{\mathcal{B}}}}}{2} - e^{-2\widehat{\mathcal{A}}_t u^{\widehat{\mathcal{B}}}}\right), \quad t = 1, \ldots, T, \quad \textbf{(A.10)}$$

whose consistency for Ψ_β and Ψ_t follows by Theorem 3 and the continuous mapping theorem.

Limiting Covariance for WPLS. The limiting Hessian and covariance matrices for the WPLS estimator have the same block-wise structure as for the PLS in (A.7) and (A.8) and may be written as,

$$\mathcal{I}^w = \mathcal{L}_1 \circ \mathcal{M}^w + \mathcal{L}_2 \circ \boldsymbol{\Gamma}^w, \qquad \boldsymbol{\Omega}^w = \mathcal{L}_1 \circ \mathcal{C}^w + \mathcal{L}_2 \circ \boldsymbol{\Gamma}^w \circ \boldsymbol{\Psi}^w, \quad \textbf{(A.11)}$$

respectively. First, for the Hessian, $\mathcal{I}^w$, whose first scaled matrix, $\mathcal{L}_1 \circ \mathcal{M}^w$, is defined as,

$$\mathcal{L}_1 \circ \mathcal{M}^w \equiv \begin{pmatrix} \mathcal{M}_{\theta_0^r}^w & \sqrt{\varpi_1}\mathcal{M}_{\theta_0^r\beta}^w & \mathcal{M}_{\theta_0^r S^r}^w & \sqrt{\zeta_1}\mathcal{M}_{\theta_0^r A}^w \\ \sqrt{\varpi_1}(\mathcal{M}_{\theta_0^r\beta}^w)' & \varpi_1 \mathcal{M}_\beta^w & \sqrt{\varpi_1}\mathcal{M}_{\beta S^r}^w & \sqrt{\varpi_1\zeta_1}\mathcal{M}_{\beta A}^w \\ (\mathcal{M}_{\theta_0^r S^r}^w)' & \sqrt{\varpi_1}(\mathcal{M}_{\beta S^r}^w)' & \mathcal{M}_{S^r}^w & \sqrt{\zeta_1}\mathcal{M}_{S^r A}^w \\ \sqrt{\zeta_1}(\mathcal{M}_{\theta_0^r A}^w)' & \sqrt{\varpi_1\zeta_1}(\mathcal{M}_{\beta A}^w)' & \sqrt{\zeta_1}(\mathcal{M}_{S^r A}^w)' & \zeta_1 \mathcal{M}_A^w \end{pmatrix}$$

https://doi.org/10.1017/S0266466618000373 Published online by Cambridge University Press

where, e.g., the $(q-1)\times T$ matrix $\mathcal{M}^w_{\theta_0^r A} = (\mathcal{M}^w_{\theta_0^r A_1}, \ldots, \mathcal{M}^w_{\theta_0^r A_T})$ has column vectors that are defined by $\mathcal{M}^w_{\theta_0^r A_t} = \mathcal{M}_{\theta_0^r A_t}/w(\phi_t)$ for $t = 1, \ldots, T$. The remaining elements of $\mathcal{M}^w$ are similarly adjusted versions of the corresponding element in $\mathcal{M}$ using the weight $1/w(\phi_t)$ at each point in time. The second term in the decomposition of $\mathcal{I}^w$, that is, $\mathcal{L}_2 \circ \boldsymbol{\Gamma}^w$, is given by,

$$\mathcal{L}_2 \circ \boldsymbol{\Gamma}^w \equiv \operatorname{diag}\left(\mathbf{0}_{(q-1)\times 1}, \frac{\varpi_2 T}{w(\Psi_\beta)}, \mathbf{0}_{T(p-1)\times 1}, \frac{\zeta_2}{w(\Psi_1)}, \ldots, \frac{\zeta_2}{w(\Psi_T)}\right).$$

Next, for the covariance matrix, $\boldsymbol{\Omega}^w$, the first part in its decomposition, $\mathcal{L}_1 \circ \mathcal{C}^w$, is defined as,

$$\mathcal{L}_1 \circ \mathcal{C}^w \equiv \begin{pmatrix} \mathcal{C}^w_{\theta_0^r} & \sqrt{\varpi_1}\mathcal{C}^w_{\theta_0^r \beta} & \mathcal{C}^w_{\theta_0^r S^r} & \sqrt{\zeta_1}\mathcal{C}^w_{\theta_0^r A} \\ \sqrt{\varpi_1}(\mathcal{C}^w_{\theta_0^r \beta})' & \varpi_1 \mathcal{C}^w_\beta & \sqrt{\varpi_1}\mathcal{C}^w_{\beta S^r} & \sqrt{\varpi_1 \zeta_1}\mathcal{C}^w_{\beta A} \\ (\mathcal{C}^w_{\theta_0^r S^r})' & \sqrt{\varpi_1}(\mathcal{C}^w_{\beta S^r})' & \mathcal{C}^w_{S^r} & \sqrt{\zeta_1}\mathcal{C}^w_{S^r A} \\ \sqrt{\zeta_1}(\mathcal{C}^w_{\theta_0^r A})' & \sqrt{\varpi_1 \zeta_1}(\mathcal{C}^w_{\beta A})' & \sqrt{\zeta_1}(\mathcal{C}^w_{S^r A})' & \zeta_1 \mathcal{C}^w_A \end{pmatrix}$$

where, similarly, the $(q-1)\times T$ matrix $\mathcal{C}^w_{\theta_0^r A} = (\mathcal{C}^w_{\theta_0^r A_1}, \ldots, \mathcal{C}^w_{\theta_0^r A_T})$ has column vectors that are adjusted to account for the weighting as $\mathcal{C}^w_{\theta_0^r A_t} = \mathcal{C}_{\theta_0^r A_1}/w(\phi_t)^2$ for $t = 1, \ldots, T$. The remaining elements of the first part $\mathcal{L}_1 \circ \mathcal{C}^w$ are defined analogously using scaling with $1/w(\phi_t)^2$. The additional term in the second part of the decomposed WPLS covariance matrix in (A.11), $\boldsymbol{\Psi}^w$, is given by,

$$\boldsymbol{\Psi}^w \equiv \operatorname{diag}\left(\mathbf{0}_{(q-1)\times 1}, \frac{\Psi_\beta}{w(\Psi_\beta)}, \mathbf{0}_{T(p-1)\times 1}, \frac{\Psi_1}{w(\Psi_1)}, \ldots, \frac{\Psi_T}{w(\Psi_T)}\right).$$

A.5. Auxiliary Results

LEMMA A.1. *Under the conditions for Theorem 3,*

$$\frac{1}{\sqrt{N}} \begin{pmatrix} \sum_{t=1}^T \sum_{j=1}^{N_t} \nabla_{\theta_0^r} \kappa(k_j, \tau_j, \boldsymbol{S}_t, \boldsymbol{\theta}_0)\epsilon_{t,k_j,\tau_j} \\ \sum_{t=1}^T \sum_{j=1}^{N_t} \nabla_\beta \kappa(k_j, \tau_j, \boldsymbol{S}_t, \boldsymbol{\theta}_0)\epsilon_{t,k_j,\tau_j} \\ \sum_{j=1}^{N_1} \nabla_{S^r} \kappa(k_j, \tau_j, \boldsymbol{S}_1, \boldsymbol{\theta}_0)\epsilon_{1,k_j,\tau_j} \\ \vdots \\ \sum_{j=1}^{N_T} \nabla_{S^r} \kappa(k_j, \tau_j, \boldsymbol{S}_T, \boldsymbol{\theta}_0)\epsilon_{T,k_j,\tau_j} \\ \sum_{j=1}^{N_1} \nabla_A \kappa(k_j, \tau_j, \boldsymbol{S}_1, \boldsymbol{\theta}_0)\epsilon_{1,k_j,\tau_j} \\ \vdots \\ \sum_{j=1}^{N_T} \nabla_A \kappa(k_j, \tau_j, \boldsymbol{S}_T, \boldsymbol{\theta}_0)\epsilon_{T,k_j,\tau_j} \end{pmatrix} \xrightarrow{\mathcal{L}-s} \mathcal{C}^{1/2} \times \begin{pmatrix} \mathbf{E}_{\theta_0^r} \\ \tilde{\mathrm{E}}_\beta \\ \mathbf{E}_{S^r} \\ \tilde{\mathbf{E}}_A \end{pmatrix}$$

where $\mathbf{E}_{\theta_0^r}$ and $\mathbf{E}_A$ are defined in Theorem 3, $\tilde{\mathrm{E}}_\beta$ and the $T\times 1$ vector $\tilde{\mathbf{E}}_A$ contain standard Gaussian random variables, which are independent of each other and of the filtratation $\mathcal{F}$, and the asymptotic covariance matrix, $\mathcal{C}$, is defined through the Hadamard product in equation (A.8).

https://doi.org/10.1017/S0266466618000373 Published online by Cambridge University Press

Proof. Follows by the same arguments as Lemma 1 in Andersen et al. (2015). ■

LEMMA A.2. *Under the conditions for Theorem 3, the convergence in Lemma A.1 and Theorem 1 holds jointly, and further, the vectors* $(\mathbf{E}'_{\theta_0^r}, \tilde{\mathrm{E}}_\beta, \mathbf{E}'_{S^r}, \tilde{\mathbf{E}}'_A)'$ *and* $(\mathrm{Y}_\beta, \mathbf{Y}'_A)'$ *are independent.*

Proof. Follows by the same arguments as Lemma 3 in Andersen et al. (2015). ■

A.6. Proof of Theorem 1

First, it is more convenient to work with the dynamics of $x = \log(X)$ throughout the proof, which by an application of Itô lemma (under $\mathbb{P}$), is given by,

$$dx_t = \alpha'_t\, dt + \int_{\mathbb{R}} x\, \widetilde{\mu}^{\mathbb{P}}(dt, dx). \tag{A.12}$$

Next, for our analysis, it is easier to work with an alternative representation of x where integration is defined with respect to a Poisson measure. To this end, we set,

$$\overline{\nu}_+^{\mathbb{P}}(x) = A_\beta |x|^{-\beta-1} + \max\{\nu_+^{\mathbb{P}}(x) - A_\beta |x|^{-\beta-1}, 0\}, \quad \text{for } x > 0, \tag{A.13}$$

and $\overline{\nu}_-^{\mathbb{P}}(x)$ is defined analogously. Using the Grigelionis representation (Theorem 2.1.2, Jacod and Protter, 2012), and upon suitably extending the probability space, we can represent the dynamics of x under $\mathbb{P}$ as,

$$\begin{aligned} dx_t = {} & \alpha'_t\, dt + \int_{\mathbb{R}_+ \times \mathbb{R}_+ \times [0,1] \times \mathbb{R}} 1(u \le A_{t-}^+,\, x > 0)\, 1(z \le \nu_+^{\mathbb{P}}(x)/\overline{\nu}_+^{\mathbb{P}}(x))\, x\, \widetilde{\underline{\mu}}(dt, du, dz, dx) \\ & + \int_{\mathbb{R}_+ \times \mathbb{R}_+ \times [0,1] \times \mathbb{R}} 1(u \le A_{t-}^-,\, x < 0)\, 1(z \le \nu_-^{\mathbb{P}}(x)/\overline{\nu}_-^{\mathbb{P}}(x))\, x\, \widetilde{\underline{\mu}}(dt, du, dz, dx), \end{aligned} \tag{A.14}$$

where $\underline{\mu}$ is an integer-valued random measure on $\mathbb{R}_+ \times \mathbb{R}_+ \times [0,1] \times \mathbb{R}$ with compensator defined by $dt \otimes du \otimes dz \otimes (\overline{\nu}_-^{\mathbb{P}}(x) 1_{\{x<0\}} + \overline{\nu}_+^{\mathbb{P}}(x) 1_{\{x>0\}})\, dx$. Noting that $\beta > 1$, we may then write,

$$\begin{aligned} dx_t = {} & \alpha''_t\, dt + \int_{\mathbb{R}_+ \times \mathbb{R}_+ \times [0,1] \times \mathbb{R}} 1(u \le A_{t-}^+,\, x > 0)\, 1(z \le A_\beta |x|^{-\beta-1}/\overline{\nu}_+^{\mathbb{P}}(x))\, x\, \widetilde{\underline{\mu}}(dt, du, dz, dx) \\ & + \int_{\mathbb{R}_+ \times \mathbb{R}_+ \times [0,1] \times \mathbb{R}} 1(u \le A_{t-}^-,\, x < 0)\, 1(z \le A_\beta |x|^{-\beta-1}/\overline{\nu}_-^{\mathbb{P}}(x))\, x\, \widetilde{\underline{\mu}}(dt, du, dz, dx) + dY_t, \end{aligned} \tag{A.15}$$

where α'' is a drift term, which is a weighted sum of α and $A^\pm$, and Y is a "residual" process satisfying Assumption A in Todorov (2015). Importantly, note that the two jump martingales in equation (A.15) have jump compensators $A_{t-}^+ \frac{A_\beta}{|x|^{\beta+1}} 1_{\{x>0\}}$ and $A_{t-}^- \frac{A_\beta}{|x|^{\beta+1}} 1_{\{x<0\}}$, respectively. These correspond to time-changed stable processes and, as a result, we can finally write,

$$dx_t = \alpha''_t\, dt + |A_{t-}^+|^{1/\beta} dS_t^+ + |A_{t-}^-|^{1/\beta} dS_t^- + dY_t, \tag{A.16}$$

where S^+ and S^- are independent stable processes with Lévy densities $\frac{A_\beta}{|x|^{\beta+1}} 1_{x>0}$ and $\frac{A_\beta}{|x|^{\beta+1}} 1_{x<0}$, respectively, and with zero drifts. This representation of x is used in what follows.

We start with $\widehat{\beta}-\beta$, where we can follow the same steps provided for the corresponding proof in Todorov (2015). Note that the setup in Todorov (2015) is more restrictive, assuming $A_t^- = A_t^+$. However, due the differencing of the increments of x in the construction of our statistic as well as the fact that the summands do not overlap (in the sense that they use different increments of x), the difference between the models here and in Todorov (2015) is irrelevant. Hence, we have,

$$\widehat{\beta}-\beta = \sum_{i=k_n+1}^{\lfloor nT \rfloor/2} \chi_i^n + o_P(\sqrt{\Delta_n}), \tag{A.17}$$

where we set,

$$\chi_i^n = \frac{1}{\log(u/v)} \frac{1}{\lfloor nT \rfloor/2 - k_n} \times \left[\frac{\cos(u\Delta_n^{-1/\beta} \mu_{p,\beta}^{-1/\beta} S_i^n) - \mathcal{C}(p,u,\beta)}{\log(\mathcal{C}(p,u,\beta))\mathcal{C}(p,u,\beta)} - \frac{\cos(v\Delta_n^{-1/\beta} \mu_{p,\beta}^{-1/\beta} S_i^n) - \mathcal{C}(p,v,\beta)}{\log(\mathcal{C}(p,v,\beta))\mathcal{C}(p,v,\beta)} \right], \tag{A.18}$$

with $\mu_{p,\beta}$ and $\mathcal{C}(p,v,\beta)$ given in Section A.2, and,

$$S_i^n = \frac{|A^+_{(i-2)\Delta_n -}|^{1/\beta}(\Delta_i^n S^+ - \Delta_{i-1}^n S^+) + |A^-_{(i-2)\Delta_n -}|^{1/\beta}(\Delta_i^n S^- - \Delta_{i-1}^n S^-)}{|A_{(i-2)\Delta_n -}|^{1/\beta}}. \tag{A.19}$$

Next, we turn to the difference $\widehat{\boldsymbol{A}} - \boldsymbol{A}$. First, using $\widehat{\beta}-\beta = O_P(\sqrt{\Delta_n})$ as well as the fact that $\mathbb{E}|\Delta_i^n x| \le K\Delta_n^{1/\beta-\iota}$ for some arbitrary small $\iota > 0$ (after appropriate localization), the following bound holds for each $t = 1, \ldots, T$,

$$\frac{1}{p_n} \sum_{i \in \mathbb{I}_t^n} \left[\cos\left(u\Delta_n^{-1/\widehat{\beta}} (\Delta_{2i}^n x - \Delta_{2i-1}^n x) \right) - \cos\left(u\Delta_n^{-1/\beta} (\Delta_{2i}^n x - \Delta_{2i-1}^n x) \right) \right] = O_P\left(\sqrt{\Delta_n^{1-\iota}} \right), \quad \forall \iota > 0. \tag{A.20}$$

Now, using Assumption A.1 for the residual jump component in equation (A.15), Y, as well as for the dynamics of the drift term in Assumption A.1, and the restriction for β' in the theorem, we have,

$$\frac{1}{p_n} \sum_{i \in \mathbb{I}_t^n} \left[\cos\left(u\Delta_n^{-1/\beta} (\Delta_{2i}^n x - \Delta_{2i-1}^n x) \right) - \cos\left(u\Delta_n^{-1/\beta} A_{(i-2)\Delta_n}^{1/\beta} S_i^n \right) \right] = o_P(1/\sqrt{p_n}). \tag{A.21}$$

Moreover, by the dynamics of the processes $A^\pm$ in Assumption A.1, it follows that,

$$\frac{1}{p_n} \sum_{i \in \mathbb{I}_t^n} e^{-A_{(i-2)\Delta_n} u^\beta} - e^{-A_t u^\beta} = O_P\left((p_n \Delta_n)^{1/\beta - \iota} \right), \quad \forall \iota > 0, \tag{A.22}$$

https://doi.org/10.1017/S0266466618000373 Published online by Cambridge University Press

$$e^{-A_{2\Delta_n(\lfloor nt/2\rfloor-p_n)}u^\beta}-e^{-A_tu^\beta} = O_p\left((p_n\Delta_n)^{1/\beta-\iota}\right), \quad \forall\iota>0. \tag{A.23}$$

Finally, using the uncorrelatedness of the summands below, we readily have,

$$\frac{1}{p_n}\sum_{i\in\mathbb{I}_t^n}\left[\cos\left(u\Delta_n^{-1/\beta}A_{(i-2)\Delta_n}^{1/\beta}S_i^n\right)-e^{-A_{(i-2)\Delta_n}u^\beta}\right] = O_p(1/\sqrt{p_n}). \tag{A.24}$$

By combining the above results and using a Taylor expansion, it follows that,

$$\widehat{A}_t-A_t = \sum_{i\in\mathbb{I}_t^n}\chi_{t,i}^n+o_p(1/\sqrt{p_n}), \quad t=1,\ldots,T, \tag{A.25}$$

where we denote,

$$\chi_{t,i}^n = \begin{cases}-\frac{e^{A_{2\Delta_n(\lfloor nt/2\rfloor-p_n)}u^\beta}}{u^\beta}\frac{1}{p_n}\left[\cos\left(u\Delta_n^{-1/\beta}A_{(i-2)\Delta_n-}^{1/\beta}S_i^n\right)-e^{-A_{(i-2)\Delta_n}u^\beta}\right], & \text{if } i\in\mathbb{I}_t^n,\\ 0, \quad \text{otherwise.}\end{cases} \tag{A.26}$$

Therefore, what remains to be proved is, that the vector $\sum_{i=k_n+1}^{\lfloor nT\rfloor/2}\left(\sqrt{nT}\chi_i^n,\ \sqrt{p_n}\left(\chi_{t,i}^n\right)_{t=1}^T\right)$ converges to the limit in the theorem (without loss of generality, we can, and do, assume $n>k_n+p_n$). First, direct calculations as well as our assumption for the dynamics of $A_t^\pm$ imply,

$$\mathbb{E}_{2i-2}^n(\chi_i^n)=0, \quad \mathbb{E}_{2i-2}^n(\chi_{t,i}^n)=0, \tag{A.27}$$

$$nT\sum_{i=k_n+1}^{\lfloor nT\rfloor/2}\mathbb{E}_{2i-2}^n(\chi_i^n)^2 = \frac{nT}{\lfloor nT\rfloor/2-k_n}\Psi_\beta, \quad p_n\sum_{i=k_n+1}^{\lfloor nT\rfloor/2}\mathbb{E}_{2i-2}^n(\chi_{t,i}^n)^2 = \Psi_t+o_p\left(\sqrt{p_n\Delta_n}\right), \tag{A.28}$$

$$\sqrt{nT}\sqrt{p_n}\sum_{i=k_n+1}^{\lfloor nT\rfloor/2}\mathbb{E}_{2i-2}^n\left(\chi_i^n\chi_{t,i}^n\right) = O_p(\sqrt{p_n/n}), \quad p_n\sum_{i=k_n+1}^{\lfloor nT\rfloor/2}\mathbb{E}_{2i-2}^n\left(\chi_{s,i}^n\chi_{t,i}^n\right)=0, \ s\neq t, \tag{A.29}$$

$$n^2\sum_{i=k_n+1}^{\lfloor nT\rfloor/2}\mathbb{E}_{2i-2}^n(\chi_i^n)^4 = O_p(1/n), \quad p_n^2\sum_{i=k_n+1}^{\lfloor nT\rfloor/2}\mathbb{E}_{2i-2}^n(\chi_{t,i}^n)^4 = O_p(1/p_n). \tag{A.30}$$

In addition, using the proof of Theorem 1 in Todorov and Tauchen (2012), we have,

$$\sqrt{n}\sum_{i=k_n+1}^{\lfloor nT\rfloor/2}\mathbb{E}_{2i-2}^n[\chi_i^n(M_{2i\Delta_n}-M_{(2i-2)\Delta_n})]=o_p(1), \quad \sqrt{p_n}\sum_{i=k_n+1}^{\lfloor nT\rfloor/2}\mathbb{E}_{2i-2}^n[\chi_{t,i}^n(M_{2i\Delta_n}-M_{(2i-2)\Delta_n})]=o_p(1), \tag{A.31}$$

for any bounded martingale M defined on the original probability space. Hence, by combining the above results, we may apply Theorem IX.7.28 of Jacod and Shiryaev (2003) to conclude that the sequence $\sum_{i=k_n+1}^{\lfloor nT\rfloor/2}\left(\sqrt{nT}\chi_i^n,\ \sqrt{p_n}\left(\chi_{t,i}^n\right)_{t=1}^T\right)$ converges to the limit in the theorem.

A.7. Proof of Theorem 2

The consistency result follows by applying Theorem 1 in conjunction with the same arguments provided to establish consistency in Theorem 1 of Andersen et al. (2015).

A.8. Proof of Theorem 3

By utilizing the consistency result in Theorem 2 as well as differentiability of the implied volatility function, we have that $\widehat{\boldsymbol{\theta}}^r$, $\widehat{\mathcal{B}}$, $\{\widehat{\boldsymbol{S}}_t^r\}_{t=1,\ldots,T}$ and $\{\widehat{\mathcal{A}}_t\}_{t=1,\ldots,T}$ with probability approaching one, solve,

$$\begin{cases} \sum_{t=1}^{T}\sum_{j=1}^{N_t}\left(\widehat{\kappa}_{t,j}-\kappa_j(\widehat{\boldsymbol{S}}_t,\widehat{\boldsymbol{\theta}})\right)\nabla_{\theta_0^r}\kappa_j(\widehat{\boldsymbol{S}}_t,\widehat{\boldsymbol{\theta}}) \;=\; \boldsymbol{0}, \\ \sum_{t=1}^{T}\sum_{j=1}^{N_t}\left(\widehat{\kappa}_{t,j}-\kappa_j(\widehat{\boldsymbol{S}}_t,\widehat{\boldsymbol{\theta}})\right)\nabla_{\beta}\kappa_j(\widehat{\boldsymbol{S}}_t,\widehat{\boldsymbol{\theta}})+\lambda_\beta\, n\, T\left(\widehat{\beta}-\widehat{\mathcal{B}}\right) \;=\; 0, \\ \sum_{j=1}^{N_1}\left(\widehat{\kappa}_{1,j}-\kappa_j(\widehat{\boldsymbol{S}}_1,\widehat{\boldsymbol{\theta}})\right)\nabla_{S^r}\kappa_j(\widehat{\boldsymbol{S}}_1,\widehat{\boldsymbol{\theta}}) \;=\; \boldsymbol{0}, \\ \vdots \\ \sum_{j=1}^{N_T}\left(\widehat{\kappa}_{T,j}-\kappa_j(\widehat{\boldsymbol{S}}_T,\widehat{\boldsymbol{\theta}})\right)\nabla_{S^r}\kappa_j(\widehat{\boldsymbol{S}}_T,\widehat{\boldsymbol{\theta}}) \;=\; \boldsymbol{0}, \\ \sum_{j=1}^{N_1}\left(\widehat{\kappa}_{1,j}-\kappa_j(\widehat{\boldsymbol{S}}_1,\widehat{\boldsymbol{\theta}})\right)\nabla_{A}\kappa_j(\widehat{\boldsymbol{S}}_1,\widehat{\boldsymbol{\theta}})+\lambda_A\, p_n\left(\widehat{A}_1-\widehat{\mathcal{A}}_1\right)=0, \\ \vdots \\ \sum_{j=1}^{N_T}\left(\widehat{\kappa}_{T,j}-\kappa_j(\widehat{\boldsymbol{S}}_T,\widehat{\boldsymbol{\theta}})\right)\nabla_{A}\kappa_j(\widehat{\boldsymbol{S}}_T,\widehat{\boldsymbol{\theta}})+\lambda_A\, p_n\left(\widehat{A}_T-\widehat{\mathcal{A}}_T\right)=0. \end{cases} \tag{A.32}$$

Next, by a first-order Taylor expansion for (A.32), the mean-value theorem and Assumption A.2,

$$(\boldsymbol{W}_n\tilde{\boldsymbol{H}}\boldsymbol{W}_n)\boldsymbol{W}_n^{-1}\begin{pmatrix}\widehat{\boldsymbol{\theta}}_\beta-\boldsymbol{\theta}_\beta\\ \widehat{\mathcal{B}}-\beta\\ \widehat{\boldsymbol{S}}-\boldsymbol{S}\\ \widehat{\mathcal{A}}-\boldsymbol{A}\end{pmatrix} \;=\; \boldsymbol{W}_n\begin{pmatrix}\boldsymbol{\mathcal{S}}_{\boldsymbol{\theta}_0^r}\\ \mathcal{S}_\beta\\ \boldsymbol{\mathcal{S}}_{S^r\cdot}\\ \boldsymbol{\mathcal{S}}_A\end{pmatrix}+o_p(\boldsymbol{W}_n), \tag{A.33}$$

where the $(q+Tp)\times(q+Tp)$ Hessian matrix $\tilde{\boldsymbol{H}}\equiv\boldsymbol{H}(\tilde{\boldsymbol{S}},\tilde{\boldsymbol{\theta}})$ is defined by equation (A.6) for some intermediate values of the state vectors $\tilde{\boldsymbol{S}}\in[\widehat{\boldsymbol{S}},\boldsymbol{S}]$ and parameters $\tilde{\boldsymbol{\theta}}\in[\widehat{\boldsymbol{\theta}},\boldsymbol{\theta}_0]$, and with score functions given as,

$$\boldsymbol{\mathcal{S}}_{\theta_0^r}\;\equiv\;\sum_{t=1}^{T}\sum_{j=1}^{N_t}\epsilon_{t,j}\,\nabla_{\theta_0^r}\kappa_j(\boldsymbol{S}_t,\boldsymbol{\theta}_0),\qquad \mathcal{S}_\beta\;\equiv\;\sum_{t=1}^{T}\sum_{j=1}^{N_t}\epsilon_{t,j}\,\nabla_\beta\kappa_j(\boldsymbol{S}_t,\boldsymbol{\theta}_0)+\lambda_\beta\, n\, T\,(\widehat{\beta}-\beta),$$

$$\boldsymbol{\mathcal{S}}_{S^r}\;\equiv\;(\boldsymbol{\mathcal{S}}'_{S_1^r},\ldots,\boldsymbol{\mathcal{S}}'_{S_T^r})',\qquad\text{with}\quad \boldsymbol{\mathcal{S}}_{S_t^r}\;\equiv\;\sum_{j=1}^{N_t}\epsilon_{t,j}\,\nabla_{S^r}\kappa_j(\boldsymbol{S}_t,\boldsymbol{\theta}_0),\qquad\text{and}$$

$$\boldsymbol{\mathcal{S}}_A\;\equiv\;(\mathcal{S}_{A_1},\ldots,\mathcal{S}_{A_T})',\qquad\text{with}\quad \mathcal{S}_{A_t}\;\equiv\;\sum_{j=1}^{N_t}\epsilon_{t,j}\,\nabla_A\kappa_j(\boldsymbol{S}_t,\boldsymbol{\theta}_0)+\lambda_A\, p_n\,(\widehat{A}_t-A_t).$$

The $o_p(\boldsymbol{W}_n)$ term in equation (A.33) comes from (higher-order) Taylor expansion effects of the gradient as well as second-order derivatives of the form, e.g., $(\widehat{\mathcal{B}}-\beta)\sum_{t=1}^{T}\sum_{j=1}^{N_t}\epsilon_{t,j}\nabla_{\beta\beta}\kappa_j(\boldsymbol{S}_t,\boldsymbol{\theta}_0)$, which are both asymptotically negligible in the present setting, since T is fixed, see, e.g., the equivalent expansion in Section 8.3.2 of Andersen

https://doi.org/10.1017/S0266466618000373 Published online by Cambridge University Press

et al. (2018). Now, since $\tilde{\boldsymbol{\theta}} \xrightarrow{\mathbb{P}} \boldsymbol{\theta}_0$ and $\tilde{\boldsymbol{S}}_t \xrightarrow{\mathbb{P}} \boldsymbol{S}_t$ for $t = 1, \ldots, T$, uniformly, by Theorem 2, and we have that the mesh of the log-moneyness grid $N_t^\tau \Delta_{t,\tau}(i_k) \xrightarrow{\mathbb{P}} \psi_{t,\tau}(k)$ uniformly on the interval $(\underline{k}(t,\tau), \bar{k}(t,\tau))$, in addition to,

$$\frac{N}{\bar{n}} \to \varpi_1, \qquad \frac{n}{\bar{n}} \to \varpi_2, \qquad \frac{N}{\overline{p}_n} \to \zeta_1, \qquad \frac{p_n}{\overline{p}_n} \to \zeta_2, \qquad \frac{p_n}{n} \to 0,$$

by Assumption A.2 as well as the function $\kappa(k, \tau, \boldsymbol{Z}, \boldsymbol{\theta})$ being second-order differentiable in their arguments by Assumption A.4 for any finite $\boldsymbol{Z}$ and $\boldsymbol{\theta}$, we may combine results to establish convergence for the Hessian matrix,

$$\boldsymbol{W}_n \tilde{\boldsymbol{H}} \boldsymbol{W}_n \xrightarrow{\mathbb{P}} \boldsymbol{\mathcal{I}}, \tag{A.34}$$

locally uniformly in $\boldsymbol{Z}$ and $\boldsymbol{\theta}$, where the $(q+Tp)\times(q+Tp)$ limiting matrix $\boldsymbol{\mathcal{I}}$ is defined in equation (A.7). To see this, note that we may write the elements along the diagonal as,

$$\frac{1}{N}\boldsymbol{H}_{\theta_0^r}(\tilde{\boldsymbol{Z}}, \tilde{\boldsymbol{\theta}}) = \sum_{t=1}^{T} \frac{N_t}{N} \frac{1}{N_t} \sum_{j=1}^{N_t} \nabla_{\theta_0^r} \kappa_j(\tilde{\boldsymbol{Z}}_t, \tilde{\boldsymbol{\theta}}) \nabla_{\theta_0^r} \kappa_j(\tilde{\boldsymbol{Z}}_t, \tilde{\boldsymbol{\theta}})' \xrightarrow{\mathbb{P}} \boldsymbol{\mathcal{M}}_{\theta_0^r},$$

$$\frac{1}{\bar{n}} H_\beta(\tilde{\boldsymbol{Z}}, \tilde{\boldsymbol{\theta}}) = \frac{N}{\bar{n}} \sum_{t=1}^{T} \frac{N_t}{N} \frac{1}{N_t} \sum_{j=1}^{N_t} \nabla_\beta \kappa_j(\tilde{\boldsymbol{Z}}_t, \tilde{\boldsymbol{\theta}}) \nabla_\beta \kappa_j(\tilde{\boldsymbol{Z}}_t, \tilde{\boldsymbol{\theta}})' + \lambda_\beta \frac{n}{\bar{n}} T \xrightarrow{\mathbb{P}} \varpi_1 \mathcal{M}_\beta + \varpi_2 \lambda_\beta T,$$

$$\frac{1}{N}\boldsymbol{H}_{S_t^r}(\tilde{\boldsymbol{Z}}_t, \tilde{\boldsymbol{\theta}}) = \frac{N_t}{N} \frac{1}{N_t} \sum_{j=1}^{N_t} \nabla_{S_t^r} \kappa_j(\tilde{\boldsymbol{Z}}_t, \tilde{\boldsymbol{\theta}}) \nabla_{S_t^r} \kappa_j(\tilde{\boldsymbol{Z}}_t, \tilde{\boldsymbol{\theta}})' \xrightarrow{\mathbb{P}} \boldsymbol{\mathcal{M}}_{S_t^r},$$

$$\frac{1}{\overline{p}_n} H_{A_t}(\tilde{\boldsymbol{Z}}, \tilde{\boldsymbol{\theta}}) = \frac{N}{\overline{p}_n} \frac{N_t}{N} \frac{1}{N_t} \sum_{j=1}^{N_t} \nabla_A \kappa_j(\boldsymbol{Z}_t, \boldsymbol{\theta}) \nabla_A \kappa_j(\boldsymbol{Z}_t, \boldsymbol{\theta})' + \lambda_A \frac{p_n}{\overline{p}_n} \xrightarrow{\mathbb{P}} \zeta_1 \mathcal{M}_{A_t} + \zeta_2 \lambda_A,$$

for $t = 1, \ldots, T$. As equivalent probability limits for the off-diagonal elements follow similarly, the asymptotic distribution result in Theorem 3 is established by using equation (A.34) in conjunction with Lemmas A.1–A.2 and Theorem 1 for (A.33), the continuous mapping theorem and Slutsky's theorem, as well as the invertibility of $\boldsymbol{\mathcal{I}}$ implied by Assumption 7.

A.9. Proof of Theorem 4

First, by Theorems 2 and 3, we have the bounds $\|\widehat{\boldsymbol{\theta}}^r - \boldsymbol{\theta}_0^r\| \le O_p(1/\sqrt{N})$, $\|\widehat{\mathcal{B}} - \beta\| \le O_p(1/\sqrt{\bar{n}})$, $\|\widehat{\boldsymbol{S}}^r - \boldsymbol{S}^r\| \le O_p(1/\sqrt{N})$ and $\|\widehat{\mathcal{A}} - \boldsymbol{A}\| \le O_p(1/\sqrt{\overline{p}_n})$. Next, make the decomposition,

$$\hat{\phi}_t - \frac{1}{N_t}\sum_{j=1}^{N_t} \phi_{j,t} = \hat{\phi}_t^{(1)} + \hat{\phi}_t^{(2)} + \hat{\phi}_t^{(3)}, \qquad \text{with} \quad \hat{\phi}_t^{(1)} = \frac{1}{N_t}\sum_{j=1}^{N_t} (\epsilon_{j,t}^2 - \phi_{j,t}),$$

$$\hat{\phi}_t^{(2)} = \frac{1}{N_t}\sum_{j=1}^{N_t} \epsilon_{j,t}\left(\kappa_{j,t}(\boldsymbol{S}_t, \boldsymbol{\theta}_0) - \kappa_{j,t}(\widehat{\boldsymbol{S}}_t, \widehat{\boldsymbol{\theta}})\right), \quad \hat{\phi}_t^{(3)} = \frac{1}{N_t}\sum_{j=1}^{N_t} \left(\kappa_{j,t}(\boldsymbol{S}_t, \boldsymbol{\theta}_0) - \kappa_{j,t}(\widehat{\boldsymbol{S}}_t, \widehat{\boldsymbol{\theta}})\right)^2,$$

where $\phi_{j,t} = \phi_{t,k_j\tau_j}$ is used as shorthand notation. Hence, by applying the above consistency bounds in conjunction with Assumption A.6, we have $|\hat{\phi}_t^{(2)}| + |\hat{\phi}_t^{(3)}| \le O_p(N^{\iota-1})$ for some arbitrarily small $\iota > 0$. Together with $\hat{\Psi}_\beta \xrightarrow{\mathbb{P}} \Psi_\beta$ and $\hat{\Psi}_t \xrightarrow{\mathbb{P}} \Psi_t$ by Theorem 3 and the continuous mapping theorem, we can use exactly the same arguments as provided for Theorem 5 in Andersen et al. (2018) in conjunction with WPLS equivalents to the expansions (A.32) and (A.33) as well as Theorem 1 to establish the result.

A.10. Feasible Inference

For better finite sample performance, we propose feasible inference that contains higher-order adjustment terms relative to the limit result in Theorem 1. These terms account for the use of $\widehat{\beta}$ in the construction of $\widehat{\boldsymbol{A}}$ as well as the nonlinear transformation of the empirical characteristic function used in the design of the estimator $\widehat{\boldsymbol{A}}$. Specifically, we first denote by $\widehat{\Psi}_\beta$ and $\widehat{\boldsymbol{\Psi}}_{\boldsymbol{A}}$ the estimates of Ψ_β and $\boldsymbol{\Psi}_{\boldsymbol{A}}$ constructed by plugging in $\widehat{\beta}$ and $\widehat{\boldsymbol{A}}$. Then, we set

$$\widehat{\boldsymbol{\Psi}}^{\mathrm{Adj}} = \begin{pmatrix} \frac{\widehat{\Psi}_\beta}{nT} & -\left(\frac{\widehat{\boldsymbol{A}}'\log(n)}{\widehat{\beta}} + \widehat{\boldsymbol{A}}'\log(u)\right)\frac{\widehat{\Psi}_\beta}{nT} \\ -\left(\frac{\widehat{\boldsymbol{A}}\log(n)}{\widehat{\beta}} + \widehat{\boldsymbol{A}}\log(u)\right)\frac{\widehat{\Psi}_\beta}{nT} & \frac{\widehat{\boldsymbol{\Psi}}_A}{p_n} + \left(\frac{\widehat{\boldsymbol{A}}\log(n)}{\widehat{\beta}} + \widehat{\boldsymbol{A}}\log(u)\right)\left(\frac{\widehat{\boldsymbol{A}}\log(n)}{\widehat{\beta}} + \widehat{\boldsymbol{A}}\log(u)\right)'\frac{\widehat{\Psi}_\beta}{nT} \end{pmatrix},$$
$$\equiv \begin{pmatrix} \widehat{\Psi}_\beta^{\mathrm{Adj}} & \widehat{\boldsymbol{\Psi}}_{\beta A}^{\mathrm{Adj}} \\ \widehat{\boldsymbol{\Psi}}_{A\beta}^{\mathrm{Adj}} & \widehat{\boldsymbol{\Psi}}_{A}^{\mathrm{Adj}} \end{pmatrix}, \tag{A.35}$$

and a $T \times 1$ vector $\mathcal{B}_{\boldsymbol{A}}$ with t-th element given by

$$\frac{e^{2\widehat{A}_t u^{\widehat{\beta}}}}{2}\left(\frac{\widehat{\boldsymbol{\Psi}}_A(t,t)}{p_n} + \frac{\widehat{A}_t^2}{\widehat{\beta}^2}\log^2 n\frac{\widehat{\Psi}_\beta}{nT}\right). \tag{A.36}$$

With this notation, we have the following feasible version of Theorem 1:

$$\left(\widehat{\boldsymbol{\Psi}}^{\mathrm{Adj}}\right)^{-1/2}\begin{pmatrix} \widehat{\beta}-\beta \\ \widehat{\boldsymbol{A}}+\mathcal{B}_{\boldsymbol{A}}-\boldsymbol{A} \end{pmatrix} \xrightarrow{\mathcal{L}_s} \begin{pmatrix} \mathrm{Y}_\beta \\ \mathbf{Y}_A \end{pmatrix}, \tag{A.37}$$

where $\left(\mathrm{Y}_\beta \;\; \mathbf{Y}_A'\right)$ is a vector of standard normals. The above higher-order expansion result accounts for the dependence between the estimation error in the jump intensities across the different days and with that from the recovery of β. These dependencies stem from using $\widehat{\beta}$ in the construction of the estimator $\widehat{\boldsymbol{A}}$. Furthermore, the asymptotic bias correction of $\widehat{\boldsymbol{A}}$ is due to the nonlinear transformation of the empirical characteristic function used when forming $\widehat{\boldsymbol{A}}$. The derivation of the above result follows trivially from applying the properties of stable convergence and the limit result in Theorem 1.

With the above convergence, the asymptotic result in Theorem 3 can be made feasible by using plug-in estimators of the Hessian and asymptotic covariance matrices. Specifically, using the notation in (A.6), the estimates $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\theta}}^r, \widehat{\mathcal{B}})$ and $\widehat{\boldsymbol{S}} = (\widehat{\boldsymbol{S}}^r, \widehat{\boldsymbol{A}})$ may be used to form

$$\widehat{\mathcal{I}} = \boldsymbol{W}_n \boldsymbol{H}(\widehat{\boldsymbol{S}}, \widehat{\boldsymbol{\theta}}) \boldsymbol{W}_n, \tag{A.38}$$

which, by continuous mapping theorem is a consistent estimate of $\mathcal{I}$. Next, for the estimation of the asymptotic covariance matrix, $\boldsymbol{\Omega}$, let us first define our estimate for the option observation error,

$$\hat{\epsilon}_{t,j} = \hat{\kappa}_{t,j} - \kappa_j(\widehat{\boldsymbol{S}}_t, \widehat{\boldsymbol{\theta}}), \tag{A.39}$$

then we can estimate the first component of (A.8) as $\boldsymbol{W}_n \widehat{\mathcal{L}_1 \circ \mathcal{C}} \boldsymbol{W}_n$, with

$$\widehat{\mathcal{L}_1 \circ \mathcal{C}} \equiv \begin{pmatrix} \widehat{\mathcal{C}}_{\theta_0^r} & \widehat{\mathcal{C}}_{\theta_0^r\beta} & \widehat{\mathcal{C}}_{\theta_0^r S^r} & \widehat{\mathcal{C}}_{\theta_0^r A} \\ \widehat{\mathcal{C}}_{\theta_0^r\beta}' & \widehat{\mathcal{C}}_{\beta} & \widehat{\mathcal{C}}_{\beta S^r} & \widehat{\mathcal{C}}_{\beta A} \\ \widehat{\mathcal{C}}_{\theta_0^r S^r}' & \widehat{\mathcal{C}}_{\beta S^r}' & \widehat{\mathcal{C}}_{S^r} & \widehat{\mathcal{C}}_{S^r A} \\ \widehat{\mathcal{C}}_{\theta_0^r A}' & \widehat{\mathcal{C}}_{\beta A}' & \widehat{\mathcal{C}}_{S^r A}' & \widehat{\mathcal{C}}_{A} \end{pmatrix}, \tag{A.40}$$

whose elements along the diagonal, that is, the $(q-1)\times(q-1)$ matrix $\widehat{\mathcal{C}}_{\theta_0^r}$, the scalar $\widehat{\mathcal{C}}_\beta$, the $T(p-1)\times T(p-1)$ matrix $\widehat{\mathcal{C}}_{S^r}$, and the $T\times T$ matrix $\widehat{\mathcal{C}}_A$, are defined as $\widehat{\mathcal{C}}_{S^r}\equiv \operatorname{diag}(\widehat{\mathcal{C}}_{S_1^r},\ldots,\widehat{\mathcal{C}}_{S_T^r})$, $\widehat{\mathcal{C}}_A\equiv\operatorname{diag}(\widehat{\mathcal{C}}_{A_1},\ldots,\widehat{\mathcal{C}}_{A_T})$, and with,

$$\widehat{\mathcal{C}}_{\theta_0^r} \equiv \sum_{t=1}^{T}\sum_{j=1}^{N_t}\hat{\epsilon}_{t,j}^2\,\nabla_{\theta_0^r}\kappa_j(\widehat{\boldsymbol{S}}_t,\widehat{\boldsymbol{\theta}})\,\nabla_{\theta_0^r}\kappa_j(\widehat{\boldsymbol{S}}_t,\widehat{\boldsymbol{\theta}})', \quad \widehat{\mathcal{C}}_{\beta} \equiv \sum_{t=1}^{T}\sum_{j=1}^{N_t}\hat{\epsilon}_{t,j}^2\,\nabla_{\beta}\kappa_j(\widehat{\boldsymbol{S}}_t,\widehat{\boldsymbol{\theta}})\,\nabla_{\beta}\kappa_j(\widehat{\boldsymbol{S}}_t,\widehat{\boldsymbol{\theta}})',$$

$$\widehat{\mathcal{C}}_{S_t^r} \equiv \sum_{j=1}^{N_t}\hat{\epsilon}_{t,j}^2\,\nabla_{S^r}\kappa_j(\widehat{\boldsymbol{S}}_t,\widehat{\boldsymbol{\theta}})\,\nabla_{S^r}\kappa_j(\widehat{\boldsymbol{S}}_t,\widehat{\boldsymbol{\theta}})', \quad \widehat{\mathcal{C}}_{A_t} \equiv \sum_{j=1}^{N_t}\nabla_{A}\kappa_j(\widehat{\boldsymbol{S}}_t,\widehat{\boldsymbol{\theta}})\,\nabla_{A}\kappa_j(\widehat{\boldsymbol{S}}_t,\widehat{\boldsymbol{\theta}})',$$

for $t=1,\ldots,T$. The remaining elements of the $(q+Tp)\times(q+Tp)$ covariance matrix in (A.40) have the same generic structure as the explicated diagonal elements and are, thus, defined analogously.

The second component of (A.8) may be estimated as $\boldsymbol{W}_n\,\mathcal{L}_2\widehat{\circ\Lambda\circ}\boldsymbol{\Psi}\,\boldsymbol{W}_n$, where

$$\mathcal{L}_2\widehat{\circ\Lambda\circ}\boldsymbol{\Psi} \equiv \begin{pmatrix} \mathbf{0}_{(q-1)\times(q-1)} & \mathbf{0}_{(q-1)\times 1} & \mathbf{0}_{(q-1)\times T(p-1)} & \mathbf{0}_{(q-1)\times T} \\ \mathbf{0}_{1\times(q-1)} & \lambda_\beta^2(nT)^2\widehat{\Psi}_\beta^{\mathrm{Adj}} & \mathbf{0}_{1\times T(p-1)} & \lambda_\beta\lambda_A(nT)p_n\widehat{\boldsymbol{\Psi}}_{\beta A}^{\mathrm{Adj}} \\ \mathbf{0}_{T(p-1)\times(q-1)} & \mathbf{0}_{T(p-1)\times 1} & \mathbf{0}_{T(p-1)\times T(p-1)} & \mathbf{0}_{T(p-1)\times T} \\ \mathbf{0}_{T\times(q-1)} & \lambda_\beta\lambda_A(nT)p_n\widehat{\boldsymbol{\Psi}}_{A\beta}^{\mathrm{Adj}} & \mathbf{0}_{T\times T(p-1)} & \lambda_A^2p_n^2\widehat{\boldsymbol{\Psi}}_A^{\mathrm{Adj}} \end{pmatrix}. \tag{A.41}$$

Altogether, we have

$$\widehat{\boldsymbol{\Omega}} = \boldsymbol{W}_n\left(\widehat{\mathcal{L}_1\circ\mathcal{C}} + \mathcal{L}_2\widehat{\circ\Lambda\circ}\boldsymbol{\Psi}\right)\boldsymbol{W}_n,$$

which, by using similar arguments as for the proof of Theorem 3 in Andersen et al. (2015) and the feasible limit result in (A.37) above, is consistent for $\boldsymbol{\Omega}$. Hence, using $\widehat{\boldsymbol{\Omega}}$ and $\widehat{\mathcal{I}}$, we can draw feasible inference on the basis of Theorem 3 in conjunction with the continuous mapping theorem and Slutsky's theorem. The feasible version of Theorem 4 is designed in an analogous way.

A.11. Parametric Option Price Computations in the Monte Carlo Study

To compute the option prices for the parametric model in (24)–(25), we solve for the conditional characteristic function of X_t and then apply Fourier inversion techniques. Specifically, the model in (24)–(25) may be written as a time-changed Lévy process:

$$X_t = Y_{T_t}, \quad \text{with} \quad T_t = \int_0^t A_s ds \quad \text{and} \quad dA_t = -\kappa A_t dt + dL_t. \tag{A.42}$$

Here, T_t is usually referred to as the business clock and A_t represents the corresponding activity rate. In our model specification, and since Y_t and A_t are independent, the conditional characteristic function of $x_{t+\tau}=\ln(X_{t+\tau})$ (with $\tau>0$), $\phi_x(u)$ (with $u\in\mathbb{C}$), is equal to

$$\phi_x(u) = \mathbb{E}_t[e^{ux_{t+\tau}}] = \mathbb{E}_t[e^{\tilde{\Psi}_y(u)T_{t+\tau}}],$$

https://doi.org/10.1017/S0266466618000373 Published online by Cambridge University Press

where $\tilde{\Psi}_y(u) = \Psi_y(u) - \Psi_y(1)$, $\Psi_y(u)$ being the characteristic exponent of the Lévy process $y_t = \ln(Y_t)$, given by:

$$\Psi_y(u) = A_\beta \Gamma(-\beta)\lambda^\beta \left\{\left(1-\frac{u}{\lambda}\right)^\beta + \left(1+\frac{u}{\lambda}\right)^\beta - 2\right\},$$

where A_β is the function of β defined in equation (3). As shown in Carr and Wu (2004) and Filipović (2001), $\phi_x(u)$ is an exponentially affine function in the current value of the activity rate A_t and the log-price x_t,

$$\phi_x(u) = e^{c(\tau)+b(\tau)A_t+ux_t},$$

where $c(t)$ and $b(t)$ are the solutions to the following ordinary differential equations:

$$b'(t) = \Psi_Y(u) - kb(t), \quad c'(t) = \int_{\mathbb{R}_0^+} \left(1-e^{-zb(t)}\right) m(z)dz,$$

with boundary conditions $c(0) = 0$ and $b(0) = 0$ and with $m(z)$ being the Lévy density of the inverse Gaussian process:

$$m(z) = \frac{c_L e^{-\mu_L z}}{z^{3/2}}.$$

In this setting, the unconditional expected value of A_t is $\mathbb{E}[A_t] = c_L\sqrt{\pi}/(k\sqrt{\mu_L})$, while the unconditional annualized variance of the log-return process equals $2\mathbb{E}[A_t] \times A_\beta\Gamma(2-\beta)\lambda^{\beta-2}$.

https://doi.org/10.1017/S0266466618000373 Published online by Cambridge University Press