DEPARTMENT OF ECONOMICS
AND BUSINESS ECONOMICS
AARHUS UNIVERSITY

CREATES
Center for Research in Econometric Analysis of Time Series

Option Panels in Pure-Jump Settings

Torben G. Andersen, Nicola Fusari, Viktor Todorov and Rasmus T. Varneskov

CREATES Research Paper 2018-4

Department of Economics and Business Economics
Aarhus University
Fuglesangs Allé 4
DK-8210 Aarhus V
Denmark

Email: oekonomi@au.dk
Tel: +45 8716 5515

Option Panels in Pure-Jump Settings*

Torben G. Andersen[†] Nicola Fusari[‡] Viktor Todorov[§] Rasmus T. Varneskov[¶]

August 1, 2017

Abstract

We develop parametric inference procedures for large panels of noisy option data in the setting where the underlying process is of pure-jump type, i.e., evolve 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 observation times. We consider the asymptotic setting in which the cross-sectional dimension of the panel increases to infinity while its time span remains fixed. The information set is further 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 L_2 distance between observed and model-implied options and further penalize for the deviation of model-implied quantities from their model-free counterparts measured via 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 can exhibit different rates of convergence depending on the relative informativeness of the high-frequency return data and the option panel.

Keywords: Inference, Jump Activity, Large Data Sets, Nonlinear Factor Model, Options, Panel Data, Stable Convergence, Stochastic Jump Intensity.

JEL classification: C51, C52, G12.

---

*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.

[†]Department of Finance, Kellogg School of Management, Northwestern University, Evanston, IL 60208; NBER, Cambridge, MA; and CREATES, Aarhus, Denmark; e-mail: t-andersen@northwestern.edu.

[‡]The Johns Hopkins University Carey Business School, Baltimore, MD 21202; e-mail: nicola.fusari@jhu.edu.

[§]Department of Finance, Kellogg School of Management, Northwestern University, Evanston, IL 60208; e-mail: v-todorov@northwestern.edu.

[¶]Department of Finance, Kellogg School of Management, Northwestern University, Evanston, IL 60208; CREATES, Aarhus, Denmark; Multi Assets at Nordea Asset Management, Copenhagen, Denmark e-mail: rasmus.varneskov@kellogg.northwestern.edu.

1 Introduction

Option data provide rich source of information about volatility and jump risks and their pricing. The availability of option data has significantly increased over the last decade and nowdays for many assets there are a large number of options traded on them at any given point in time which differ in terms of their tenor and strike level. Each of these options provide unique source of information for the conditional risk-neutral distribution of the asset. At the same time, high-frequency return data is also readily available and can help in the estimation of the realized volatility and jump risks in the underlying asset.

In Andersen et al. (2015) we proposed inference for the parameters and factor realizations implied by a parametric model for the risk-neutral dynamics of the asset on the basis of an option panel with fixed time span and asymptotically increasing cross-sectional dimension. We further augmented the information set by the inclusion of high-frequency return data which in turn was used for constructing nonparametric estimates of the spot diffusive volatility. The estimation was then performed under the assumption that the high-frequency return data is less informative than the option data (when combined with the parametric model) for the recovery of the latent factor realizations. Under this assumption, we performed estimation via penalized least squares in which we minimized L_2 distance between model-implied and observed option prices and we further penalized for deviation between model-implied and nonparametric estimates of the spot diffusive volatility based on the high-frequency data.

The goal of the current paper is twofold. First, we want to relax the 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. Second, we would like to extend the analysis by including also information from the high-frequency data regarding the jump component of the asset. We achieve the above goals in the setting of a model of pure-jump type, i.e., model in which the dynamics of the asset does not contain a diffusive component. Models of pure-jump type have been used in prior work to describe the dynamics of various assets such as volatility indices and exchange rates.

The information in the high-frequency return data about the parametric model for its risk-neutral dynamics is due to the equivalence of the statistical and risk-neutral probability measures implied by a no-arbitrage condition (a minimal assumption that is used in most theoretical and empirical asset pricing work). For a diffusion, it implies that the diffusion coefficient of the price should remain the same under both probability measures. We utilized this condition in Andersen et al. (2015). For the jumps in the model, the no-arbitrage conditions are more complicated. For the "big" jumps, we have essentially no restrictions. This is intuitive as on a given path we might even not detect any "big" jumps. Equivalence of statistical and risk-neutral probability measures does impose, however, "similar" behavior under the two probability laws of the "small" jumps. In particular, the so-called jump activity should remain the same and the same should hold for the intensity of the "small" jumps. Jump activity is the infimum over the set of powers for which the power variation of the jumps remains finite (on a finite interval). It classifies the 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. The inference for jump activity from high-frequency data has received a lot of attention in recent work, see e.g., Ait-Sahalia and Jacod (2009), Jing et al. (2011), Jing et al. (2012), Todorov and Tauchen (2011), Todorov (2015) and Woerner (2003, 2007).

Given the above discussion, we "summarize" the information in the high-frequency return data about the risk-neutral parametric model for the underlying asset into estimates for the jump activity and the spot jump intensities at the option observation times. We adopt the approach of Todorov (2015) based on the empirical characteristic function and we further extend the analysis of that paper to the recovery of spot jump intensity (Todorov (2015) considers only jump activity estimation). 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 weighted sums of the option observation errors. This joint limit theory, in turn, allows us to characterize the limit distribution of an estimator that incorporates the high-frequency return data and the option data. We note that the analysis of the current paper can be easily extended to cover alternative high-frequency jump activity and intensity estimators (e.g., ones that are robust to the presence of a diffusion in the price dynamics) provided one can derive their asymptotic distribution.

The estimation of the parameters and the factor realizations of the risk-neutral parametric model is done via penalized least squares. In particular, we minimize L_2 distance between observed and model-implied option prices and we further penalize for deviations of model-implied jump activity and intensities estimates from their model-free counterparts based on the high-frequency return data. The different parts of the parameter and state vectors can exhibit different rates of convergence depending on the relative information content (for our estimation purposes) of the return and option data. Importantly, the user does not need to take an apriori stand on this. That is, if the returns are more informative about e.g. jump activity (in the sense of allowing for faster rates of convergence), then our penalized least squares for this quantity will behave asymptotically as the nonparametric high-frequency estimator. The reverse will be true when the option data is more informative for the jump activity parameter - our estimator will behave asymptotically as one build from the option data solely. In the boundary case where option and return data allow for estimators of jump activity with the same rate of convergence, we can further weigh optimally the two parts of the objective function (due to the return and option data) according to the noisiness of these two sources of information. This is achieved in a weighted penalized least squares extension of the above method which is free of tuning parameters.

The rest of the paper 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 present the observation scheme and the asymptotic setup for the inference. Section 4 presents our penalized least squares estimator and develops the associated asymptotic theory for it. Section 5 extends the results to the case of weighted penalized least squares. Section 6 concludes. Proofs of the theoretical results and assumptions are given in Section 7.

2 A Setup for Pure-Jump Modeling of the Option Panel

This section introduces a nonlinear parametric factor model for a panel of options written on an underlying asset whose price we denote with X. The option specification is based on a parametric model of pure-jump type for the risk-neutral dynamics of X. We further describe the characteristics of the underlying price process that are preserved under change from the physical probability measure, $\mathbb{P}$, to the risk-neutral measure, $\mathbb{Q}$. These measure-preserving characteristics will be utilized later on in our inference for the parametric model via nonparametric estimates for them from high-frequency return data.

2.1 Pure-Jump Dynamics of Locally Stable Type

We start with describing the dynamics of the underlying price process, X, which is defined on a filtered probability space $\big(\Omega^{(0)}, \mathcal{F}^{(0)}, (\mathcal{F}_t^{(0)})_{t\geq 0}, \mathbb{P}^{(0)}\big)$. We will not assume a parametric model for X under $\mathbb{P}^{(0)}$ but instead we will only assume that its $\mathbb{P}^{(0)}$ dynamics belongs in a general class of models of pure-jump type given by,

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

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 jump martingale measure associated with the counting jump measure $\mu(dt, dx)$ and its jump compensator $\nu^{\mathbb{P}}(dt, dx)$. We assume $\nu^{\mathbb{P}}(dt, dx)$ has the following structure

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

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,

$$\left|\nu^{\mathbb{P}}_{\pm}(x) - A_\beta |x|^{-\beta-1}\right| \leq C|x|^{-\beta'-1}, \quad A_\beta = \left(\frac{4\Gamma(2-\beta)|\cos(\beta\pi/2)|}{\beta(\beta-1)}\right)^{-1}, \quad |x| \leq x_0, \ \ \beta' < \beta, \tag{3}$$

for some constants $C > 0$ and $x_0 > 0$. The coefficient β coincides with the so-called jump activity which controls the roughness of trajectories of X. That is, we have for every t:

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

We will restrict attention to the empirically realistic case of $1 < \beta < 2$ which implies paths of infinite variation.[1]

The jump specification in (2)-(3) is very flexible and accommodates many parametric jump specifications used in empirical work. The "stable like" restriction in (3) is only for the behavior of the jump compensator around zero and we leave the behavior of the latter for the big jumps essentially unrestricted. Condition (3) is needed for deriving CLT for the estimators of β and $A^+_t + A^-_t$ from high-frequency return data on X that we propose below. This assumption will be satisfied by the CGMY model of Carr et al. (2002) as well as models in the class of tempered stable processes of Rosinski (2007), whose jumps can have, in particular, much thinner tails than those of the stable process. We further note that we allow in our setup for asymmetry in the jump compensator by having (possibly) different A^+_t and A^-_t. We also allow for time-varying jump intensities through the time variation of A^+_t and A^-_t. Thus we can nest in our setup Lévy-driven SDE's such as the COGARCH model of Klüppelberg et al. (2004), the pure-jump CGMY model with stochastic volatility in Carr et al. (2003), and more generally models of pure-jump type within the affine jump-diffusion class of Duffie et al. (2000).

2.2 Parametric model for the Option Prices

We turn next to specifying the dynamics of X under the so-called risk-neutral measure which in turn will allow us to determine the theoretical value of the option prices 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 prices into local martingales. Specifically, for X under $\mathbb{Q}$, we may write,

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

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, $\tilde{\nu}^{\mathbb{Q}}(dt, dx)$.

[1] We defer all formal assumptions to Section 7.1.

It is important to note that, in the absence of arbitrage, there are characteristics of the physical price process (1), which are preserved under the risk-neutral dynamics in (5). We shall detail such equivalences and utilize them when designing our estimation methodology below.

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 $\mathbb{Q}$ 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 variance (BSIV) corresponding to $O_{t,k,\tau}$ by $\kappa_{t,k,\tau}$, which represents a convenient monotone transformation that is often used to quote option prices in practice.

We shall further assume a 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 $q \times 1$. Then, we assume $A_t^+ \equiv \xi_1(\boldsymbol{S}_t, \boldsymbol{\theta}_0)$ and $A_t^- \equiv \xi_2(\boldsymbol{S}_t, \boldsymbol{\theta}_0)$, where $\xi_1(\,\cdot\,)$ and $\xi_2(\,\cdot\,)$ are known functions that are invariant to the parameter $\boldsymbol{\theta}_0$.[2] In addition, the risk-neutral jump compensator is parametrized via

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

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 $\mathbb{P}$ to $\mathbb{Q}$. Similarly, the jump activity index under $\mathbb{Q}$ is given by β because there is agreement between the null sets of $\mathbb{P}$ and $\mathbb{Q}$, see the definition of jump activity in (4). Therefore, we assume that β is part of the parameter vector $\boldsymbol{\theta}_0$. We define the remaining $(q-1)\times 1$ elements by $\boldsymbol{\theta}_0^r$. The parameters β and $\boldsymbol{\theta}_0^r$ will play different roles in the econometric analysis below.

The density of the probability measure change is given by a stochastic exponential involving the ratio $\nu^{\mathbb{Q}}/\nu^{\mathbb{P}}$ for which to be well defined we need (see e.g., 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 equivalence of $\mathbb{P}$ and $\mathbb{Q}$ in the Lévy case (where the jump compensator and the drift are time invariant), see e.g., Theorem 33.1

[2]This is almost universally satisfied in empirical applications. In addition, r_t and q_t should be known functions of $\boldsymbol{S}_t$.

of Sato (1999). It severely limits the wedge between $\nu_{\pm}^{\mathbb{Q}}$ and $\nu_{\pm}^{\mathbb{P}}$ around zero. To better illustrate this, we consider the CGMY specification for $\nu_{\pm}^{\mathbb{Q}}$ given by

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

If $\nu_{\pm}^{\mathbb{P}}$ is also from a CGMY model, but with possibly different parameters, then given our restriction in (3), (8) is equivalent to

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

Note, in particular, that we have no restriction for the parameters $\lambda_{\pm}$ which govern the behavior of the jump compensator in the tails. By contrast the parameters that control the behavior of the jump compensator around zero are unchanged when switching from $\mathbb{P}$ to $\mathbb{Q}$. This example illustrates that $\nu_t^{\mathbb{P}}$ and $\nu_t^{\mathbb{Q}}$ are approximately the same around zero and can be very different outside zero.

Under the parametric model, we may write the BSIV 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, driving all time variation in the option prices, and which we only require to be an $\mathcal{F}^{(0)}$-adapted stochastic process. The above option pricing framework extends the corresponding one in Andersen et al. (2015) and Andersen et al. (2017) by allowing the underlying asset price, X, to obey pure-jump process specification. Hence, in addition to accommodating affine jump-diffusions, we also allow for the 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

We continue next with describing the observation scheme for the option prices and the high-frequency return data used to augment the option information set. We also introduce the nonparametric estimators from the high-frequency data of the jump activity and jump intensities, which are quantities that are preserved under an equivalent measure change.

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 to have observations 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. (2017). Specifically, we suppose that the options data cover a wide range of strike prices and tenors (k and τ, respectively). That is, for each t we observe options $\{O_{t,k_j,\tau_j}\}_{j=1,\dots,N_t}$, where N_t is some large integer and the index j is running across the full set of strike and tenor combinations. Moreover, the number of options for maturity τ is denoted by N_t^τ such that by definition $N_t = \sum_\tau N_t^\tau$, and 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 time through, for example, a varying number of options at a given time t, the available strike-tenor combinations (k, τ), and, for a given τ, the density, or clustering, of available strikes in the log-moneyness grid. In particular, we suppose to have 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 with N being an unobserved number, representing the "average size of the cross-section".[3] Moreover, for each combination 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 $i = 2, \dots, N_t^\tau$. In analogy with infill asymptotics for high-frequency observations, our asymptotic scheme sequentially adds new strikes within $[\underline{k}(t,\tau), \bar{k}(t,\tau)]$ such that $\Delta_{t,\tau}(i) \stackrel{\mathbb{P}}{\to} 0$ 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 that are "mildly" out-of-the-money, in contrast to the more sparsely available deep OTM call options. These facets will impact the precision of the 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 below.

In addition, $\mathcal{T}_t$ denotes the tenors available at time t, and the vectors $\underline{\boldsymbol{k}}_t = (\underline{k}(t,\tau))_{\tau \in \mathcal{T}_t}$ and $\overline{\boldsymbol{k}}_t = (\overline{k}(t,\tau))_{\tau \in \mathcal{T}_t}$ indicate the lowest and highest log-moneyness at time t across the available tenors. As described above, these quantities may vary over time and be random, thus capturing and summarizing the pronounced shifts in the characteristics of the observed option cross-section that may occur over 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 $\Omega^{(1)} = \times_{t \in \mathbb{N}, k \in \mathbb{R}, \tau \in \Gamma} \mathcal{R}_{t,k,\tau}$, for $\mathcal{R}_{t,k,\tau} \in \mathbb{R}$, 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

[3] Again, we defer all formal assumptions to Section 7.1.

as $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ 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 that are defined on $\Omega^{(0)}$ and $\Omega^{(1)}$ such as X_t and $\epsilon_{t,k,\tau}$, respectively, 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 $\Omega^{(1)}$, which is equipped with a "large" product topology to support them, as they may be associated with any strike, point in time and maturity. This space suffices since, 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 High-Frequency Data and Nonparametric Measures

In addition to the panel of options prices, we utilize a second source of information in our estimation, namely high-frequency data on the underlying asset X, to assist in the recovery of (a part of) the state and parameter vectors. 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 to have an equidistant high-frequency recording of X_t at times $0, 1/n, \dots, i/n, \dots, T$, and we denote the increment size with $\Delta_n = 1/n$. Finally, we define the logarithmic asset price by $x_t = \log(X_t)$ and the associated log-return by $\Delta_i^n x = x_{i/n} - x_{(i-1)/n}$.

3.2.1 Jump Activity Estimation

We compute the activity index, β, using the estimator in Todorov (2015), which is based on self-normalized statistics of 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 of the drift and the asymmetry of the jump intensity. Further, the use of the ECF generates efficiency gains over power variation-based methods, see e.g., Todorov (2015).

To set the stage, let $1 < k_n < \lfloor nT/2 \rfloor$ be the block size, then 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) due to the fact that unlike that paper, here the summands in $\widehat{V}_i(p)$ and $\widehat{\mathcal{C}}(p,u)$ use only non-overleaping increments. This results in some loss of efficiency of our jump activity estimator relative to the one of Todorov (2015) (as we have fewer summands in $\widehat{\mathcal{C}}(p,u)$ for a given data set). This change, however, allows us to handle the more general setting in which the intensity around zero can be asymmetric, i.e., we can have $A_t^+ \neq A_t^-$.

The asymptotic properties of $\widehat{\mathcal{C}}(p,u)$ will naturally depend on the properties of $\widehat{V}_i(p)$. In particular, for consistency of the latter for the total intensity, A_t, we require $k_n \to \infty$, and $k_n/n \to 0$ is similarly needed to avoid letting time-variation in A_t generate a bias. Moreover, to develop central limit theory for the ECF, Todorov (2015) shows that $k_n/\sqrt{n} \to 0$ suffices to ensure that the sampling error biases in $\widehat{V}_i(p)$ are sufficiently small, and that a bias-correction

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

where the exact expression for $\mathcal{B}_n(p,u,\beta)$ is provided in Section 7.2, can have a CLT. 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.[4] We will show below that the estimator (16) achieves the almost optimal rate of convergence, i.e., its rate is $O_p(1/\sqrt{n})$.

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

[4]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) and Todorov and Tauchen (2011).

3.2.2 Jump Intensity Estimation

We continue next with constructing nonparametric estimators for the total jump intensity. Unlike the jump activity, the jump intensity changes in general over time. Given our option observation scheme, we will need estimates for A_t at the integer times $t = 1, ..., T$ only. We will construct such estimators using local blocks consisting of p_n differenced and non-overlapping increments that precede the integer times.

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 in Todorov (2015) (in the context of jump activity), estimators based on the empirical characteristic function can provide nontrivial efficiency improvements. For this reason we consider the following

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

where $\mathbb{I}_t^n = \{\lfloor tn/2\rfloor - p_n + 1, ..., \lfloor tn/2\rfloor\}$ and p_n is a deterministic sequence satisfying $p_n \to \infty$ and $p_n/n \to 0$. We note that at the cost of more complicated analysis, we can further extend the above jump intensity estimator to separately identify A_t^+ and A_t^-. We leave such an extension for future work.

4 Inference for Pure-Jump Models from Option Panels

We now proceed with the core of our econometric analysis. We introduce a new penalized least squares (PLS) estimator for option panels generated from 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 and as an intermediary step in the analysis of the asymptotic distribution of the PLS estimator, we derive a joint CLT for the nonparametric high-frequency estimators.

4.1 Penalized Least Squares Estimator

In designing our new PLS estimator for option panels from pure-jump models, we use three key relations from Sections 2 and 3. First, given the signal-plus-noise decomposition of observed BSIVs in (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, and as discussed in Section 2.2 above, the jump activity index, β, and the total spot jump intensity $A_t = A_t^+ + A_t^-$ are preserved under change of measure from $\mathbb{P}^{(0)}$ to $\mathbb{Q}$. Moreover, these quantities may be recovered nonparametrically from high-frequency return data with the estimators presented in Sections 3.2.1 and 3.2.2, and we 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

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

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

using $\widehat{\beta} \equiv \widehat{\beta}(p, u, v)$ and $\widehat{A}_t \equiv \widehat{A}_t(u)$ as shorthand notation. 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:

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

Whereas both Andersen et al. (2015) and Andersen et al. (2017) consider PLS for option panels in different asymptotic settings, they both assume that the underlying price process contains a diffusion, i.e., a martingale component driven by a Brownian motion, and they penalize deviation of model-implied spot volatility from a nonparametric measure of it from high-frequency return data.[5] The major differences of the PLS estimator in the current paper from the one in the above-cited papers are two.[6] First, we incorporate in the estimation (via the penalization terms) information in the high-frequency data about parts of the parameter and state vectors which are due to the implications of the equivalence of the $\mathbb{P}$ and $\mathbb{Q}$ measures regarding the jump distribution. Intuitively, the second penalization term in (18) can be viewed as the jump counterpart of the penalization for divergence between mode-implied and high-frequency spot volatility estimates in the diffusion context. The first penalization term in (18) regarding the jump activity has no diffusive analogue though. The second

[5]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 et al. (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.

[6]In comparison to Andersen et al. (2015) and Andersen et al. (2017), 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 the parts of the parameter and state vectors.

difference between the current work and Andersen et al. (2015) and Andersen et al. (2017) is that unlike these papers, we do not put any restrictions on the relative information content of the high-frequency return and option data sets. That is, we allow for arbitrary relations between N, n and k_n. Of course, one should keep in mind that the option data is in general always needed in the estimation as the high-frequency return data can only help in the estimation of part of the parameter and state vectors.

Before proceeding to the asymptotic theory for the PLS estimator, we need to develop CLT for the joint behavior of our nonparametric estimators of β and A_t, $t = 1, \dots, T$ from the high-frequency data.

4.2 Preliminary High-Frequency Asymptotics

To state the result for the asymptotic distribution of the nonparametric high-frequency estimators, we need some additional notation. Define the $T \times 1$ vectors $\widehat{\boldsymbol{A}} = (\widehat{A}_t)_{t=1}^T$, $\boldsymbol{A} = (A_t)_{t=1}^T$ and the $T \times T$ matrix $\boldsymbol{\Psi}_A = \text{diag}(\Psi_1, \dots, \Psi_T)$. Note that both Ψ_β and $\boldsymbol{\Psi}_A$ are defined in Section 7.2. Finally, 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 will play an important role for deriving the asymptotic distribution of the PLS estimator later on.

Theorem 1. *Suppose Assumption 1 holds. Moreover, let the power p as well as the sequences k_n and p_n in* (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

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

where Y_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* $\mathcal{F}$.

Theorem 1 extends results in Todorov (2015) in two directions. First, unlike Todorov (2015), here we allow for asymmetry in the jump intensity around zero, i.e., we accommodate the setting in which

$A_t^+ \neq A_t^-$. Second, in addition to estimating β, here we consider also estimates of A_t at various points in time. 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, the use of $\widehat{\beta}$ in the construction of $\widehat{\boldsymbol{A}}$ has no effect on the limiting result in Theorem 1. We note that this is very different from the case in which one tries to recover the integrated intensity $\int_0^T A_s ds$. The asymptotic distribution of the latter will be driven by the use of $\widehat{\beta}$ in its construction, and hence this will generate perfect asymptotic dependence between the integrated jump intensity estimator and the estimator of the jump activity. In our case such asymptotic degeneracy is avoided by the slower rate of convergence of the spot jump intensity estimator. The choice of p_n in R1 is standard for estimation of spot quantities (e.g., estimation of the spot diffusive volatility) and balances bias in the recovery of the spot jump intensity due to 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 to be expected as $\widetilde{\mathcal{C}}(p, u, \beta)$ contains self-normalization which annihilates the effect from the time-variation in $A_t^\pm$ on the limiting distribution of $\widehat{\beta}$. On the other hand, the asymptotic distribution of $\widehat{\boldsymbol{A}}$ is mixed Gaussian and the precision of estimating $\boldsymbol{A}$ depends on its random realization.

Conditions R2 and R3 are exactly as in Todorov (2015) and determine the range of possible choices for the power and the block size of the local power variation used in the normalization of the differenced increments which in turn are used in the construction of $\widehat{\beta}$. In general, a good choice for the block size is to pick ϖ as close as possible to 1/2. For the power p, a possible choice is to set it arbitrary close and above 1/2. In principle, given the participation of the unknown β in the restrictions in R2 and R3, one can consider an adaptive choice for k_n and p. We leave these considerations for future work.

Finally, we point out that Theorem 1 is a key building block in the derivation of the asymptotic distribution of our PLS estimator. If one is to use an alternative estimator of β and $\boldsymbol{A}$, e.g., an estimator adapted to settings in which X might contain a diffusion, then in order to adopt the asymptotic analysis of Theorem 3 below, all that is needed is an equivalent to Theorem 1 for the alternative nonparametric high-frequency estimator.

4.3 Consistency of PLS Estimator

We may now use Theorem 1 to establish consistency of $\widehat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{S}} = (\widehat{\boldsymbol{S}}_t)_{t=1}^T$. The formal result is given in the following theorem.

Theorem 2. *Under Assumptions 1-5 as well as R1-R3 in Theorem 1, then for some $T \in \mathbb{N}$, it follows* *that* $(\widehat{\boldsymbol{\theta}}, (\widehat{\boldsymbol{S}}_t)_{t=1}^T)$ *exists with probability approaching 1, and further that*

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

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. (2017) 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.

4.4 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 the 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, and with $\boldsymbol{\iota}_d$ denoting a d-dimensional vector of ones. Using this notation, we may now state the limiting distribution result for our PLS estimator.

Theorem 3. *Under Assumptions 1-6 as well as R1-R3 in Theorem 1, 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 each other as well as of the filtration* $\mathcal{F}$*, and where the Hessian and asymptotic covariance matrices,* $\boldsymbol{\mathcal{I}}$ *and* $\boldsymbol{\Omega}$*, are defined in* (30) *and* (31) *of Section 7.3.*

The limiting result in Theorem 3 exhibits different rates of convergence for the various components of the PLS estimator. In particular, for the estimates of the components of the parameter and state vectors for which we have no information from the high-frequency return data, 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 $\sqrt{N}$ rate associated with utilizing the information from the parametric model and the option panel and the $\sqrt{n}$ rate associated with the nonparametric estimator based on the high-frequency return data. In that regard, we note that the scaling of the penalization terms in the objective function in (18) plays an important role as it ensures 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, i.e., when $n << N$. In the opposite case, i.e., when $n >> N$, the scaling of the penalty term in the objective function ensures that it is the latter that determines the asymptotic behavior of the jump activity estimator. In the border case when $n \asymp N$, both the high-frequency return data and the option data contribute to the asymptotic variance of β and this is reflected in their joint determination of the terms in $\boldsymbol{\mathcal{I}}$ and $\boldsymbol{\Omega}$ that correspond to $\widehat{\boldsymbol{\mathcal{B}}}$. Similar analysis applies for the estimator of the jump intensity $\widehat{\boldsymbol{\mathcal{A}}}$. In this case, the relevant comparison is $\sqrt{N}$ convergence from utilizing the option data versus $\sqrt{p_n}$ rate of convergence from using the high-frequency data. Since $p_n/n \to 0$ by R1 in Theorem 1, then if $N >> n$ we also have $N >> k_n$. That is, if option data is more efficient for estimation of β it is also more efficient for the recovery of $\boldsymbol{A}$. In this case all components of $\widehat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{S}}$ converge at the common rate of $\sqrt{N}$. Similarly, if $k_n >> N$, then we also have $n >> N$. That is, in this case the high-frequency data is more informative both for the jump activity and the jump intensity. In this case, each of the components in the partitioning of the parameter vector and the vector of the state realizations converges at different rates.

Overall, our PLS estimators of β and $\boldsymbol{A}$ adapts to the situation at hand. When the high-frequency data is more informative than the option data ($k_n >> N$ or $n >> N$), then the PLS estimator for these quantities is asymptotically equivalent to their nonparametric high-frequency counterparts. On the other hand, when the option data (together with the parametric model) carries more information than the high-frequency return data for either β or $\boldsymbol{A}$ ($N >> n$ or $N >> k_n$), then the corresponding PLS estimator behaves as if only the option data is used for the estimation of this quantity. Importantly, the user does not need to take a stand apriori on whether the option or the high-frequency data is more informative for β or $\boldsymbol{A}$. This is very convenient from a practical point of view. In the boundary cases of either $N \asymp n$ or $N \asymp k_n$, both the option and high-frequency return data contribute to the estimation of (some of) the parameters and state vectors. In this case, one can choose λ_β and λ_A in a way that takes into account the difference in the variance of the option error and the high-frequency estimators' asymptotic variances. This provides further gains of efficiency and makes 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 an adaptive choice for λ_β and λ_A in the next section.

In a typical application, the state vector will consist of separate states A_t^+ and A_t^- (and they in turn can be further determined by additional factors like in the multi-factor stochastic volatility models). In this case, if we have $k_n >> N$, then A_t^+ and A_t^- will be each estimated at the slower rate of $\sqrt{N}$ and their joint distribution will be degenerate. Their sum, however, $A_t = A_t^+ + A_t^-$ will be estimated at the faster rate of $\sqrt{k_n}$. In our statement of Theorem 3, we reparametrize the state vector in a way that we avoid degeneracy of the limiting distribution and this allows one to characterize the limiting distribution of arbitrary transformations of the state vector. The situation here is similar to other econometric setups, e.g., the case of inference for autoregressive processes around deterministic time trends.

Finally, the asymptotic distribution of both the parameters and the state vector is in general mixed Gaussian. That is, the matrixes $\boldsymbol{\mathcal{I}}$ and $\boldsymbol{\Omega}$ are in general random. This reflects on one hand the mixed Gaussian distribution for the estimates of $\boldsymbol{A}$ from the high-frequency return data and the conditional heteroskedasticity in the option observation error on the other hand. Since the convergence in Theorem 3 is stable, this, however, does not constitute a major difficulty. All that is needed for feasible inference on the basis of the limit result in Theorem 3 is consistent estimators for $\boldsymbol{\mathcal{I}}$ and $\boldsymbol{\Omega}$ which are easy to construct. For brevity we omit the details.

5 Weighted PLS Estimation

Two key components of the PLS estimator are the constants λ_β and λ_A, and we will 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 manner that is analogous to classical weighted least squares. We call 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, which are defined in Section 7.4, then $\hat{\Psi}_\beta \xrightarrow{\mathbb{P}} \Psi_\beta$ and $\hat{\Psi}_t \xrightarrow{\mathbb{P}} \Psi_t$ readily follows by Theorem 3 in conjunction with the continuous mapping theorem. Now, since the $\mathcal{F}$-conditional variance of the errors due to the two penalization terms generally have unknown form, we propose to standardize their contribution to the objective function with estimates of the $\mathcal{F}$-conditional asymptotic variances of the nonparametric estimators from high-frequency data, which are provided by Theorem 1, thereby insuring that their respective contributions to the objective function are similar in scale.

Next, for optimal weighting of the elements in the L_2 part of $\mathcal{L}_t\left(\boldsymbol{Z}_t, \boldsymbol{\theta}\right)$, we would ideally like to standardize these using an estimate of the $\mathcal{F}$-conditional variance of the BSIV observation errors in (11), defined by $\phi_{t,k,\tau}$ in Assumption 6 of Section 7.1. However, despite such a procedure being feasible, we simplify the analysis and assign identical weights to all options on a given day. Although, this procedure neglects potential heteroskedasticity in the strike and tenor dimensions of the option panel, it may still generate non-trivial efficiency improvements due to pronounced heteroskedasticity in the $\mathcal{F}$-conditional option error variance over time. Moreover, it ensures that all components of the (weighted) objective function are on comparable scales. Formally, we use

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

where $\widehat{\boldsymbol{S}}_t$ and $\widehat{\boldsymbol{\theta}}$ are based on first-stage PLS estimation. Not surprisingly, $\hat{\phi}_t$ 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

$$\begin{aligned} \mathcal{L}^w(\boldsymbol{Z}, \boldsymbol{\theta}) &\equiv \sum_{t=1}^{T} \mathcal{L}_t^w(\boldsymbol{Z}_t, \boldsymbol{\theta}) + nT \frac{\left(\widehat{\beta} - \mathcal{B}\right)^2}{w(\hat{\Psi}_\beta)}, \quad \text{with} \\ \mathcal{L}_t^w(\boldsymbol{Z}_t, \boldsymbol{\theta}) &\equiv \left\{ \sum_{j=1}^{N_t} \frac{\left(\tilde{\kappa}_{t,k_j,\tau_j} - \kappa(k_j, \tau_j, \boldsymbol{Z}_t, \boldsymbol{\theta})\right)^2}{w(\hat{\phi}_t)} + p_n \frac{\left(\widehat{A}_t - \mathcal{A}_t\right)^2}{w(\hat{\Psi}_t)} \right\}, \end{aligned} \tag{22}$$

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. However, we consider such a situation as a corollary below.

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

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

Before proceeding to deriving its asymptotic distribution, it is important to highlight two features of the WPLS estimator. First, whereas the weighting by (21) is similar to the corresponding procedure in Andersen et al. (2017), the importance of additionally using $\hat{\Psi}_\beta$ and $\hat{\Psi}_t$ are much larger in our pure-jump setting. This follows from applying different "regularization" devices. The use of noisy spot variance measures, as in the former, are on a similar scale as the L_2 component, even without re-weighting the observations, since they both reflect "return variance measures". However, this is not the case for (18), whose three components reflect return variances, their jump activity index and their jump intensities. Hence, the use of (23) will result in a more stable estimation procedure (numerically), in addition to providing asymptotic efficiency gains. Second, it is important to note that the weighting in (22) is only feasible due to our stable central limit theory in Theorem 3, allowing for estimation and utilization of weights that asymptotically random. We now present the distribution theory.

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* $\boldsymbol{\mathcal{I}}$ *and* $\boldsymbol{\Omega}$ *are replaced with* $\boldsymbol{\mathcal{I}}^w$ *and* $\boldsymbol{\Omega}^w$*, which are defined in* (34) *of Section 7.4.*

Corollary 1. *Suppose the conditions of* (4) *hold and, in addition, that* $\Psi_\beta > \epsilon$, $\inf_{t\in 1,\dots,T} \Psi_t > \epsilon$*, and*

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

Finally, let $w(x) = x$*, then we have*

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

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 where* $\boldsymbol{\mathcal{I}}$ *is defined in* (34) *of Section 7.4*

6 Conclusion

In this paper we develop inference techniques for noisy option panels with fixed time span and asymptotically increasing cross-sectional dimension in which the option data is generated from a parametric model for the risk-neutral dynamics of the underlying asset that is of pure-jump type. The information set used in the estimation is further 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 and the vector of jump intensity realizations at the integer times of observing the cross-sections of options in the panel. Estimation of the parameters and the state vector realizations of the model is done via penalized least squares in which we minimize L_2 distance between observed and model-implied option prices and we further penalize for deviations of jump activity and jump intensity estimates from their nonparametric counterparts based on the high-frequency return data. The estimates for different components of the parameter and state vectors differ depending on the relative informativeness of the high-frequency return data (via the nonparametric measures developed from it) and the option data (via the parametric model), and our PLS estimator adapts to the situation at hand.

7 Proofs

This section states the formal assumptions for the theoretical analysis as well as provides proofs of the asymptotic results in the paper. Before proceeding, however, we introduce some notation. Specifically, we adopt the shorthand notation $\widehat{\kappa}_{t,k_j,\tau_j} \equiv \widehat{\kappa}_{t,j}$, $\epsilon_{t,k_j,\tau_j} \equiv \epsilon_{t,j}$, $\kappa(k_j, \tau_j, \boldsymbol{Z}, \boldsymbol{\theta}) \equiv \kappa_j(\boldsymbol{Z}, \boldsymbol{\theta})$. The Hadamard product is indicated by $\circ$; and the matrix norm being used 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, denote by K a generic constant, which may take different values in different places, and we signify conditional expectations by $\mathbb{E}_i^n(\,\cdot\,) \equiv \mathbb{E}(\,\cdot\,|\mathcal{F}_{i\Delta_n})$. Finally, note that (stochastic) orders sometimes refer to scalars, vectors, and sometimes to matrices. We refrain from making distinctions.

7.1 Assumptions

Before proceeding to the assumptions, let $(E, \mathcal{E})$ denote an auxiliary measure space on the original filtered probability space $\big(\Omega^{(0)}, \mathcal{F}^{(0)}, (\mathcal{F}_t^{(0)})_{t\geq 0}, \mathbb{P}^{(0)}\big)$.

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

$$q_t = q_0 + \int_0^t b_s^q ds + \int_0^t \int_E \kappa(\delta^q(s,x))\underline{\tilde{\vartheta}}(ds,dx) + \int_E \kappa'(\delta^q(s,x))\underline{\vartheta}(ds,dx) \tag{24}$$

where $\kappa(x) = x$ is the usual truncation function, for which $\kappa(-x) = -\kappa(x)$ and $\kappa'(x) = x - \kappa(x)$. The process (24) and its remaining components satisfy:

(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 $\mu^{\mathbb{P}}$, 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 in (24) with coefficients satisfying conditions analogous to (ii) and (iii) above.

Assumption 2 (Sampling scheme). *As $N \to \infty$, $k_n \to \infty$, and $n \to \infty$ with $k_n/n \to 0$, as well as with $\bar{n} = n \vee N$ and $\bar{k}_n = k_n \vee N$, 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*

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

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

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

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

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

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

Assumption 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* $\boldsymbol{\theta} \in \boldsymbol{\Theta}$*,*

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

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

(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.*

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 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 4.4.

7.2 Definitions for the High-frequency Estimators

This section provides additional details for the activity index and jump intensity estimators, both for their definitions and in 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

$$\mathcal{C}(p,u,\beta) = e^{-C_{p,\beta}u^\beta}, \quad \text{with} \quad C_{p,\beta} = \left[\frac{2^p\Gamma((1+p)/2)\Gamma(1-p/\beta)}{\sqrt{\pi}\Gamma(1-p/2)}\right]^{-\beta/p}, \tag{25}$$

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

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

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

$$\mathcal{B}_n(p,u,\beta) = H(p,u,\beta)\boldsymbol{\zeta}^{(2,2)}(p,u,u,\beta)/(2k_n). \tag{26}$$

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

$$\begin{aligned}
\Psi_\beta = \frac{2}{\log^2(u/v)}\Bigg[&\frac{\boldsymbol{\zeta}^{(1,1)}(p,u,u,\beta)}{\log^2(\mathcal{C}(p,u,\beta))\mathcal{C}^2(p,u,\beta)} + \frac{\boldsymbol{\zeta}^{(1,1)}(p,v,v,\beta)}{\log^2(\mathcal{C}(p,v,\beta))\mathcal{C}^2(p,v,\beta)} \\
&- 2\frac{\boldsymbol{\zeta}^{(1,1)}(p,u,v,\beta)}{\log(\mathcal{C}(p,u,\beta))\mathcal{C}(p,u,\beta)\log(\mathcal{C}(p,v,\beta))\mathcal{C}(p,v,\beta)}\Bigg],
\end{aligned} \tag{27}$$

$$\Psi_t = \frac{e^{2A_{t-}u^\beta}}{u^{2\beta}}\left(\frac{1+e^{-2^\beta A_{t-}u^\beta}}{2} - e^{-2A_{t-}u^\beta}\right), \quad t = 1, ..., T. \tag{28}$$

7.3 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,

$$\boldsymbol{H}(\boldsymbol{Z},\boldsymbol{\theta}) \equiv \begin{pmatrix} \boldsymbol{H}_{\theta_0^r}(\boldsymbol{Z},\boldsymbol{\theta}) & \boldsymbol{H}_{\theta_0^r\beta}(\boldsymbol{Z},\boldsymbol{\theta}) & \boldsymbol{H}_{\theta_0^r S^r}(\boldsymbol{Z},\boldsymbol{\theta}) & \boldsymbol{H}_{\theta_0^r A}(\boldsymbol{Z},\boldsymbol{\theta}) \\ \boldsymbol{H}_{\theta_0^r\beta}(\boldsymbol{Z},\boldsymbol{\theta})' & H_{\beta}(\boldsymbol{Z},\boldsymbol{\theta}) & \boldsymbol{H}_{\beta S^r}(\boldsymbol{Z},\boldsymbol{\theta}) & \boldsymbol{H}_{\beta A}(\boldsymbol{Z},\boldsymbol{\theta}) \\ \boldsymbol{H}_{\theta_0^r S^r}(\boldsymbol{Z},\boldsymbol{\theta})' & \boldsymbol{H}_{\beta S^r}(\boldsymbol{Z},\boldsymbol{\theta})' & \boldsymbol{H}_{S^r}(\boldsymbol{Z},\boldsymbol{\theta}) & \boldsymbol{H}_{S^r A}(\boldsymbol{Z},\boldsymbol{\theta}) \\ \boldsymbol{H}_{\theta_0^r A}(\boldsymbol{Z},\boldsymbol{\theta})' & \boldsymbol{H}_{\beta A}(\boldsymbol{Z},\boldsymbol{\theta})' & \boldsymbol{H}_{S^r A}(\boldsymbol{Z},\boldsymbol{\theta})' & \boldsymbol{H}_{A}(\boldsymbol{Z},\boldsymbol{\theta}) \end{pmatrix}, \tag{29}$$

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 $\boldsymbol{H}_{S^r}(\boldsymbol{Z},\boldsymbol{\theta}) \equiv \operatorname{diag}\left(\boldsymbol{H}_{S_1^r}(\boldsymbol{Z}_1,\boldsymbol{\theta}),\ldots,\boldsymbol{H}_{S_T^r}(\boldsymbol{Z}_T,\boldsymbol{\theta})\right)$, $\boldsymbol{H}_A(\boldsymbol{Z},\boldsymbol{\theta}) \equiv \operatorname{diag}\left(H_{A_1}(\boldsymbol{Z}_1,\boldsymbol{\theta}),\ldots,H_{A_T}(\boldsymbol{Z}_T,\boldsymbol{\theta})\right)$, and with

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

for $t = 1,\ldots,T$. The remaining elements of the $(q+Tp)\times(q+Tp)$ Hessian matrix (29) 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 (29) and may be written

$$\boldsymbol{\mathcal{I}} = \boldsymbol{\mathcal{L}}_1 \circ \boldsymbol{\mathcal{M}} + \boldsymbol{\mathcal{L}}_2 \circ \boldsymbol{\Lambda}, \tag{30}$$

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

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

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

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

for $t = 1, \ldots, T$. The remaining elements of $\boldsymbol{\mathcal{M}}$ are defined similarly, the only change being the respective gradient arguments. The second term in the decomposition (30), $\boldsymbol{\mathcal{L}}_2 \circ \boldsymbol{\Lambda}$, is given by

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

with, again, $\mathbf{0}_d$ 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 (30), as

$$\boldsymbol{\Omega} = \boldsymbol{\mathcal{L}}_1 \circ \boldsymbol{\mathcal{C}} + \boldsymbol{\mathcal{L}}_2 \circ \boldsymbol{\Lambda} \circ \boldsymbol{\Psi}, \tag{31}$$

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

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

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

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

for $t = 1, \ldots, T$, and the remaining elements of $\boldsymbol{\mathcal{L}}_1 \circ \boldsymbol{\mathcal{C}}$ are defined similarly. The additional term in the second part of the decomposition (31), $\boldsymbol{\Psi}$, is given by

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

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

7.4 Definitions for WPLS Estimation

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

Expressions for $\hat{\Psi}_\beta$ and $\hat{\Psi}_A$. Letting $\widehat{\mathcal{B}}$ and $\widehat{\mathcal{A}}_t$, $t = 1, \ldots, T$ by 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

$$\begin{aligned} \hat{\Psi}_\beta = \frac{2}{\log^2(u/v)} \Bigg[& \frac{\boldsymbol{\zeta}^{(1,1)}(p, u, u, \widehat{\mathcal{B}})}{\log^2(\mathcal{C}(p, u, \widehat{\mathcal{B}}))\mathcal{C}^2(p, u, \widehat{\mathcal{B}})} + \frac{\boldsymbol{\zeta}^{(1,1)}(p, v, v, \widehat{\mathcal{B}})}{\log^2(\mathcal{C}(p, v, \widehat{\mathcal{B}}))\mathcal{C}^2(p, v, \widehat{\mathcal{B}})} \\ & - 2\frac{\boldsymbol{\zeta}^{(1,1)}(p, u, v, \widehat{\mathcal{B}})}{\log(\mathcal{C}(p, u, \widehat{\mathcal{B}}))\mathcal{C}(p, u, \widehat{\mathcal{B}}) \log(\mathcal{C}(p, v, \widehat{\mathcal{B}}))\mathcal{C}(p, v, \widehat{\mathcal{B}})} \Bigg], \end{aligned} \tag{32}$$

$$\hat{\Psi}_t = \frac{e^{2\widehat{\mathcal{A}}_t u^{\widehat{\mathcal{B}}}}}{u^{2\widehat{\mathcal{B}}}} \left(\frac{1 + e^{-2^{\widehat{\mathcal{B}}}\widehat{\mathcal{A}}_t u^{\widehat{\mathcal{B}}}}}{2} - e^{-2\widehat{\mathcal{A}}_t u^{\widehat{\mathcal{B}}}} \right), \quad t = 1, ..., T, \tag{33}$$

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 (30) and (31) and may be written as

$$\boldsymbol{\mathcal{I}}^w = \boldsymbol{\mathcal{L}}_1 \circ \boldsymbol{\mathcal{M}}^w + \boldsymbol{\mathcal{L}}_2 \circ \boldsymbol{\Lambda}^w, \qquad \boldsymbol{\Omega}^w = \boldsymbol{\mathcal{L}}_1 \circ \boldsymbol{\mathcal{C}}^w + \boldsymbol{\mathcal{L}}_2 \circ \boldsymbol{\Lambda}^w \circ \boldsymbol{\Psi}^w, \tag{34}$$

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

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

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

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

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

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

where, similarly, the $(q-1) \times T$ matrix $\boldsymbol{\mathcal{C}}^w_{\theta_0^r A} = \left(\boldsymbol{\mathcal{C}}^w_{\theta_0^r A_1}, \ldots, \boldsymbol{\mathcal{C}}^w_{\theta_0^r A_T}\right)$ has column vectors that are adjusted to accounting for the weighting as $\boldsymbol{\mathcal{C}}^w_{\theta_0^r A_t} = \boldsymbol{\mathcal{C}}_{\theta_0^r A_1}/w(\phi_t)^2$ for $t = 1, \ldots, T$. The remaining elements of the first part $\boldsymbol{\mathcal{L}}_1 \circ \boldsymbol{\mathcal{C}}^w$ 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 (34), $\boldsymbol{\Psi}^w$, is given by

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

7.5 Auxiliary Results

Lemma 1. *Under the conditions for Theorem 3,*

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

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, $\boldsymbol{\mathcal{C}}$, is defined through the Hadamard product in (31).

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

Lemma 2. *Under the conditions for Theorem 3, then the convergence in Lemma 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). □

7.6 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

$$dx_t=\alpha'_t dt+\int_{\mathbb{R}} x\widetilde{\mu}^{\mathbb{P}}(dt,dx). \tag{35}$$

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

$$\overline{\nu}_+^{\mathbb{P}}(x)=A_\beta|x|^{-\beta-1}+\max\{\nu_+^{\mathbb{P}}(x)-A_\beta|x|^{-\beta-1},0\},\quad \text{for } x>0, \tag{36}$$

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

$\mathbb{P}$ as

$$dx_t = \alpha'_t dt + \int_{\mathbb{R}_+\times\mathbb{R}_+\times[0,1]\times\mathbb{R}} 1(u \le A^+_{t-}, x > 0)1(z \le \nu^{\mathbb{P}}_+(x)/\overline{\nu}^{\mathbb{P}}_+(x))x\widetilde{\underline{\mu}}(dt, du, dz, dx)$$
$$+ \int_{\mathbb{R}_+\times\mathbb{R}_+\times[0,1]\times\mathbb{R}} 1(u \le A^-_{t-}, x < 0)1(z \le \nu^{\mathbb{P}}_-(x)/\overline{\nu}^{\mathbb{P}}_-(x))x\widetilde{\underline{\mu}}(dt, du, dz, dx), \tag{37}$$

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

$$dx_t = \alpha''_t dt + \int_{\mathbb{R}_+\times\mathbb{R}_+\times[0,1]\times\mathbb{R}} 1(u \le A^+_{t-}, x > 0)1(z \le A_\beta|x|^{-\beta-1}/\overline{\nu}^{\mathbb{P}}_+(x))x\widetilde{\underline{\mu}}(dt, du, dz, dx)$$
$$+ \int_{\mathbb{R}_+\times\mathbb{R}_+\times[0,1]\times\mathbb{R}} 1(u \le A^-_{t-}, x < 0)1(z \le A_\beta|x|^{-\beta-1}/\overline{\nu}^{\mathbb{P}}_-(x))x\widetilde{\underline{\mu}}(dt, du, dz, dx) + dY_t, \tag{38}$$

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 (38) 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

$$dx_t = \alpha''_t dt + |A^+_{t-}|^{1/\beta}dS^+_t + |A^-_{t-}|^{1/\beta}dS^-_t + dY_t, \tag{39}$$

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

$$\widehat{\beta} - \beta = \sum_{i=k_n+1}^{\lfloor nT\rfloor/2} \chi^n_i + o_p(\sqrt{\Delta_n}), \tag{40}$$

where we set

$$\chi^n_i = \frac{1}{\log(u/v)}\frac{1}{\lfloor nT\rfloor/2 - k_n}$$
$$\times \left[\frac{\cos(u\Delta_n^{-1/\beta}\mu_{p,\beta}^{-1/\beta}S^n_i) - \mathcal{C}(p, u, \beta)}{\log(\mathcal{C}(p, u, \beta))\mathcal{C}(p, u, \beta)} - \frac{\cos(v\Delta_n^{-1/\beta}\mu_{p,\beta}^{-1/\beta}S^n_i) - \mathcal{C}(p, v, \beta)}{\log(\mathcal{C}(p, v, \beta))\mathcal{C}(p, v, \beta)}\right], \tag{41}$$

with $\mu_{p,\beta}$ and $\mathcal{C}(p, v, \beta)$ given in Section 7.2, and

$$S^n_i = \frac{|A^+_{(i-2)\Delta_n-}|^{1/\beta}(\Delta^n_i S^+ - \Delta^n_{i-1}S^+) + |A^-_{(i-2)\Delta_n-}|^{1/\beta}(\Delta^n_i S^- - \Delta^n_{i-1}S^-)}{|A_{(i-2)\Delta_n-}|^{1/\beta}}. \tag{42}$$

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 $\mathbb{E}|\Delta_i x| \le K\Delta_n^{1/\beta-\iota}$ for some arbitrary small $\iota > 0$ (after appropriate localization), the following bound holds for each $t = 1, ..., T$:

$$\frac{1}{p_n}\sum_{i\in\mathbb{I}_t^n}\left[\cos\left(u\Delta_n^{-1/\widehat{\beta}}(\Delta_{2i}^n x - \Delta_{2i-1}^n x)\right) - \cos\left(u\Delta_n^{-1/\beta}(\Delta_{2i}^n x - \Delta_{2i-1}^n x)\right)\right] = O_p\left(\sqrt{\Delta_n^{1-\iota}}\right), \quad \forall \iota > 0. \tag{43}$$

Now, using Assumption 1 for the residual jump component in (38), Y, as well as for the dynamics of the drift term in Assumption 1 and the restriction for β' in the theorem, we have

$$\frac{1}{p_n}\sum_{i\in\mathbb{I}_t^n}\left[\cos\left(u\Delta_n^{-1/\beta}(\Delta_{2i}^n x - \Delta_{2i-1}^n x)\right) - \cos\left(u\Delta_n^{-1/\beta}A_{(i-2)\Delta_n-}^{1/\beta}S_i^n\right)\right] = o_p(1/\sqrt{p_n}). \tag{44}$$

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

$$\frac{1}{p_n}\sum_{i\in\mathbb{I}_t^n} e^{-A_{(i-2)\Delta_n-}u^\beta} - e^{-A_t u^\beta} = O_p\left((p_n\Delta_n)^{1/\beta-\iota}\right), \quad \forall \iota > 0, \tag{45}$$

$$e^{-A_{2\Delta_n(\lfloor nt/2\rfloor - p_n)-}u^\beta} - e^{-A_t u^\beta} = O_p\left((p_n\Delta_n)^{1/\beta-\iota}\right), \quad \forall \iota > 0. \tag{46}$$

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

$$\frac{1}{p_n}\sum_{i\in\mathbb{I}_t^n}\left[\cos\left(u\Delta_n^{-1/\beta}A_{(i-2)\Delta_n-}^{1/\beta}S_i^n\right) - e^{-A_{(i-2)\Delta_n-}u^\beta}\right] = O_p(1/\sqrt{p_n}). \tag{47}$$

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

$$\widehat{A}_t - A_t = \sum_{i\in\mathbb{I}_t^n}\chi_{t,i}^n + o_p(1/\sqrt{p_n}), \quad t = 1, ..., T, \tag{48}$$

where we denote

$$\chi_{t,i}^n = \begin{cases} -\frac{e^{A_{2\Delta_n(\lfloor nt/2\rfloor - p_n)-}u^\beta}}{u^\beta}\frac{1}{p_n}\left[\cos\left(u\Delta_n^{-1/\beta}A_{(i-2)\Delta_n-}^{1/\beta}S_i^n\right) - e^{-A_{(i-2)\Delta_n-}u^\beta}\right], & \text{if } i \in \mathbb{I}_t^n, \\ 0, \quad \text{otherwise.} \end{cases} \tag{49}$$

Therefore, what remains to be proved is that the vector $\sum_{i=k_n+1}^{\lfloor nT\rfloor/2}\big(\sqrt{nT}\chi_i^n,\ \sqrt{p_n}(\chi_{t,i}^n)_{t=1}^T\big)$ 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

$$\mathbb{E}_{2i-2}^n(\chi_i^n) = 0, \quad \mathbb{E}_{2i-2}^n(\chi_{t,i}^n) = 0, \tag{50}$$

$$nT\sum_{i=k_n+1}^{\lfloor nT\rfloor/2}\mathbb{E}_{2i-2}^n(\chi_i^n)^2 = \frac{nT}{\lfloor nT\rfloor/2 - k_n}\Psi_\beta, \quad p_n\sum_{i=k_n+1}^{\lfloor nT\rfloor/2}\mathbb{E}_{2i-2}^n(\chi_{t,i}^n)^2 = \Psi_t + o_p\left(\sqrt{p_n\Delta_n}\right), \tag{51}$$

$$\sqrt{nT}\sqrt{p_n}\sum_{i=k_n+1}^{\lfloor nT\rfloor/2}\mathbb{E}_{2i-2}^n\left(\chi_i^n\chi_{t,i}^n\right)=O_p(\sqrt{p_n/n}),\quad p_n\sum_{i=k_n+1}^{\lfloor nT\rfloor/2}\mathbb{E}_{2i-2}^n\left(\chi_{s,i}^n\chi_{t,i}^n\right)=0,\quad s\neq t, \tag{52}$$

$$n^2\sum_{i=k_n+1}^{\lfloor nT\rfloor/2}\mathbb{E}_{2i-2}^n(\chi_i^n)^4=O_p(1/n),\quad p_n^2\sum_{i=k_n+1}^{\lfloor nT\rfloor/2}\mathbb{E}_{2i-2}^n(\chi_{t,i}^n)^4=O_p(1/p_n). \tag{53}$$

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

$$\sqrt{n}\sum_{i=k_n+1}^{\lfloor nT\rfloor/2}\mathbb{E}_{2i-2}^n[\chi_i^n(M_{2i\Delta_n}-M_{(2i-2)\Delta_n})]=o_p(1),\quad \sqrt{p_n}\sum_{i=k_n+1}^{\lfloor nT\rfloor/2}\mathbb{E}_{2i-2}^n[\chi_{t,i}^n(M_{2i\Delta_n}-M_{(2i-2)\Delta_n})]=o_p(1), \tag{54}$$

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. □

7.7 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). □

7.8 Proof of Theorem 3

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

$$\begin{cases}
\sum_{t=1}^T\sum_{j=1}^{N_t}\left(\widehat{\kappa}_{t,j}-\kappa_j(\widehat{\boldsymbol{S}}_t,\widehat{\boldsymbol{\theta}})\right)\nabla_{\theta_0^r}\kappa_j(\widehat{\boldsymbol{S}}_t,\widehat{\boldsymbol{\theta}})=\mathbf{0}\\
\sum_{t=1}^T\sum_{j=1}^{N_t}\left(\widehat{\kappa}_{t,j}-\kappa_j(\widehat{\boldsymbol{S}}_t,\widehat{\boldsymbol{\theta}})\right)\nabla_{\beta}\kappa_j(\widehat{\boldsymbol{S}}_t,\widehat{\boldsymbol{\theta}})+\lambda_\beta nT\left(\widehat{\beta}-\widehat{\mathcal{B}}\right)=0,\\
\sum_{j=1}^{N_1}\left(\widehat{\kappa}_{1,j}-\kappa_j(\widehat{\boldsymbol{S}}_1,\widehat{\boldsymbol{\theta}})\right)\nabla_{S^r}\kappa_j(\widehat{\boldsymbol{S}}_1,\widehat{\boldsymbol{\theta}})=\mathbf{0},\\
\quad\vdots\\
\sum_{j=1}^{N_T}\left(\widehat{\kappa}_{T,j}-\kappa_j(\widehat{\boldsymbol{S}}_T,\widehat{\boldsymbol{\theta}})\right)\nabla_{S^r}\kappa_j(\widehat{\boldsymbol{S}}_T,\widehat{\boldsymbol{\theta}})=\mathbf{0},\\
\sum_{j=1}^{N_1}\left(\widehat{\kappa}_{1,j}-\kappa_j(\widehat{\boldsymbol{S}}_1,\widehat{\boldsymbol{\theta}})\right)\nabla_{A}\kappa_j(\widehat{\boldsymbol{S}}_1,\widehat{\boldsymbol{\theta}})+\lambda_A p_n\left(\widehat{A}_1-\widehat{\mathcal{A}}_1\right)=0,\\
\quad\vdots\\
\sum_{j=1}^{N_T}\left(\widehat{\kappa}_{T,j}-\kappa_j(\widehat{\boldsymbol{S}}_T,\widehat{\boldsymbol{\theta}})\right)\nabla_{A}\kappa_j(\widehat{\boldsymbol{S}}_T,\widehat{\boldsymbol{\theta}})+\lambda_A p_n\left(\widehat{A}_T-\widehat{\mathcal{A}}_T\right)=0.
\end{cases} \tag{55}$$

Next, by a first-order Taylor expansion for (55), the mean-value theorem and Assumption 2,

$$(\boldsymbol{W}_n\tilde{\boldsymbol{H}}\boldsymbol{W}_n)\boldsymbol{W}_n^{-1}\begin{pmatrix}\widehat{\boldsymbol{\theta}}_\beta-\boldsymbol{\theta}_\beta\\ \widehat{\mathcal{B}}-\beta\\ \widehat{\boldsymbol{S}}-\boldsymbol{S}\\ \widehat{\mathcal{A}}-\boldsymbol{A}\end{pmatrix}=\boldsymbol{W}_n\begin{pmatrix}\boldsymbol{\mathcal{S}}_{\boldsymbol{\theta}_0^r}\\ \mathcal{S}_\beta\\ \boldsymbol{\mathcal{S}}_{S^r}.\\ \boldsymbol{\mathcal{S}}_A\end{pmatrix}+o_p(\boldsymbol{W}_n), \tag{56}$$

where the $(q+Tp)\times(q+Tp)$ Hessian matrix $\tilde{\boldsymbol{H}} \equiv \boldsymbol{H}(\tilde{\boldsymbol{S}},\tilde{\boldsymbol{\theta}})$ is defined by (29) 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

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

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

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

The $o_p(\boldsymbol{W}_n)$ term in (56) 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 et al. (2017). 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

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

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

$$\boldsymbol{W}_n\tilde{\boldsymbol{H}}\boldsymbol{W}_n \xrightarrow{\mathbb{P}} \boldsymbol{\mathcal{I}}, \tag{57}$$

locally uniformly in $\boldsymbol{Z}$ and $\boldsymbol{\theta}$, where the $(q+Tp)\times(q+Tp)$ limiting matrix $\boldsymbol{\mathcal{I}}$ is defined in (30). To see this, note that we may write the elements along the diagonal as:

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

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

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

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

for $t=1,\ldots,T$. As equivalent probability limits for the off-diagonal elements follow similarly, the asymptotic distribution result is established by using (57) in conjunction with Lemmas 1-2 and Theorem 1 for (56), the continuous mapping theorem and Slutsky's theorem. □

7.9 Proof of Theorem 4

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

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

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

where $\phi_{j,t} = \phi_{t,k_j\tau_j}$ is used as shorthand notation. Hence, by applying above consistency bounds in conjunction with Assumption 6, we have $|\hat{\phi}_t^{(2)}| + |\hat{\phi}_t^{(3)}| \leq O_p(N^{\iota-1})$ 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. (2017) in conjunction with WPLS equivalents to the expansions (55) and (56) as well as Theorem 1 to establish the result. □

References

Ait-Sahalia, Y. and 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, and V. Todorov (2015). Parametric Inference and Dynamic State Recovery from Option Panels. *Econometrica 83*, 1081–1145.

Andersen, T. G., N. Fusari, V. Todorov, and R. T. Varneskov (2017). Unified inference for nonlinear factor models from panels with fixed and large time span. Unpublished manuscript, Northwestern University.

Bernanke, B., J. Boivin, and P. Eliasz (2005). Measuring the Effects of Monetary Policy: A Factor-Augmented Vector Autoregressive (FAVAR) Approach. *Quarterly Journal of Economics 120*, 387–422.

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

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

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

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

Duffie, D., J. Pan, and K. Singleton (2000). Transform Analysis and Asset Pricing for Affine Jump-Diffusions. *Econometrica 68*, 1343–1376.

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

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

Jing, B., X. Kong, and 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, and P. Mykland (2012). On the Jump Activity Index of Semimartingales. *Journal of Econometrics 166*, 213–223.

Klüppelberg, C., A. Lindner, and R. Maller (2004). A Continuous Time GARCH Process Driven by a Lévy Process: Stationarity and Second Order Behavior. *Journal of Applied Probability 41*, 601–622.

Rosinski, J. (2007). Tempering stable processes. *Stochastic Processes and Applications 117*, 677–707.

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

Todorov, V. (2015). Jump activity estimation for pure-jump semimartingales via self-normalized statistics. *Annals of Statistics 43*, 1831–1864.

Todorov, V. and G. Tauchen (2011). Limit theorems for power variations of pure-jump processes with application to activity estimation. *The Annals of Applied Probability 21*, 546–588.

Todorov, V. and 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.

Research Papers 2018

CREATES
Center for Research in Econometric Analysis of Time Series

2017-26: Mikkel Bennedsen, Asger Lunde and Mikko S. Pakkanen: Decoupling the short- and long-term behavior of stochastic volatility

2017-27: Martin M. Andreasen, Jens H.E. Christensen and Simon Riddell: The TIPS Liquidity Premium

2017-28: Annastiina Silvennoinen and Timo Teräsvirta: Consistency and asymptotic normality of maximum likelihood estimators of a multiplicative time-varying smooth transition correlation GARCH model

2017-29: Cristina Amado, Annastiina Silvennoinen and Timo Teräsvirta: Modelling and forecasting WIG20 daily returns

2017-30: Kim Christensen, Ulrich Hounyo and Mark Podolskij: Is the diurnal pattern sufficient to explain the intraday variation in volatility? A nonparametric assessment

2017-31: Martin M. Andreasen, Jens H.E. Christensen and Glenn D. Rudebusch: Term Structure Analysis with Big Data

2017-32: Timo Teräsvirta: Nonlinear models in macroeconometrics

2017-33: Isabel Casas, Eva Ferreira and Susan Orbe: Time-varying coefficient estimation in SURE models. Application to portfolio management

2017-34: Hossein Asgharian, Charlotte Christiansen, Ai Jun Hou and Weining Wang: Long- and Short-Run Components of Factor Betas: Implications for Equity Pricing

2017-35: Juan Carlos Parra-Alvarez, Olaf Posch and Mu-Chun Wang: Identification and estimation of heterogeneous agent models: A likelihood approach

2017-36: Andrés González, Timo Teräsvirta, Dick van Dijk and Yukai Yang: Panel Smooth Transition Regression Models

2017-37: Søren Johansen and Morten Ørregaard Nielsen: Testing the CVAR in the fractional CVAR model

2017-38: Nektarios Aslanidis and Charlotte Christiansen: Flight to Safety from European Stock Markets

2017-39: Tommaso Proietti, Niels Haldrup and Oskar Knapik: Spikes and memory in (Nord Pool) electricity price spot prices

2018-01: Emilio Zanetti Chini: Forecaster's utility and forecasts coherence

2018-02: Torben G. Andersen, Nicola Fusari and Viktor Todorov: The Pricing of Tail Risk and the Equity Premium: Evidence from International Option Markets

2018-03: Torben G. Andersen, Nicola Fusari, Viktor Todorov and Rasmus T. Varneskov: Unified Inference for Nonlinear Factor Models from Panels with Fixed and Large Time Span

2018-04: Torben G. Andersen, Nicola Fusari, Viktor Todorov and Rasmus T. Varneskov: Option Panels in Pure-Jump Settings