SCHOOL OF ECONOMICS AND MANAGEMENT
FACULTY OF SOCIAL SCIENCES
AARHUS UNIVERSITY

CREATES
Center for Research in Econometric Analysis of Time Series

CREATES Research Paper 2010-16

Estimation of Jump Tails

Tim Bollerslev and Viktor Todorov

School of Economics and Management
Aarhus University
Bartholins Allé 10, Building 1322, DK-8000 Aarhus C
Denmark

Estimation of Jump Tails*

Tim Bollerslev† and Viktor Todorov‡

This Version: April 14, 2010

Abstract

We propose a new and flexible non-parametric framework for estimating the jump tails of Itô semimartingale processes. The approach is based on a relatively simple-to-implement set of estimating equations associated with the compensator for the jump measure, or its "intensity", that only utilizes the weak assumption of regular variation in the jump tails, along with in-fill asymptotic arguments for uniquely identifying the "large" jumps from the data. The estimation allows for very general dynamic dependencies in the jump tails, and does not restrict the continuous part of the process and the temporal variation in the stochastic volatility. On implementing the new estimation procedure with actual high-frequency data for the S&P 500 aggregate market portfolio, we find strong evidence for richer and more complex dynamic dependencies in the jump tails than hitherto entertained in the literature.

Keywords: Extreme events, jumps, high-frequency data, jump tails, non-parametric estimation, stochastic volatility.

JEL classification: C13, C14, G10, G12.

*We would like to thank Yacine Ait-Sahalia, Torben Andersen, Richard Davis, Jean Jacod, Thomas Mikosch, Mark Podolskij, Eric Renault, George Tauchen as well as seminar participants at the Oberwolfach Workshop on "Challenges in Statistical Theory," and the "Financial Econometrics Conference" at Princeton University for useful comments. The research was partly funded by a grant from the NSF to the NBER. Bollerslev also acknowledges support from CREATES funded by the Danish National Research Foundation.

†Department of Economics, Duke University, Durham, NC 27708, and NBER and CREATES; e-mail: boller@duke.edu.

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

1 Introduction

The recent financial crises has spurred a renewed interest in the estimation of tail events. We add to the currently available tools for assessing tail behavior in financial markets by developing a new and flexible non-parametric framework for the estimation of the jump tails of Itô semi-martingales. These processes, which are ubiquitous in continuous-time economic modeling and modern asset pricing finance in particular, portray the dynamic evolution in form of a drift term and a combination of continuous and dis-continuous martingale increments, driven by separate stochastic volatility and jump compensators, respectively. While both of the martingale components can account for non-Gaussian behavior, the tails associated with the jumps manifest themselves very differently from a formal statistical perspective.[1] Exploiting these differences, we develop a new robust methodology for estimating the jump tails. The approach is based on a relatively simple-to-implement set of estimating equations associated with the compensator for the jump measure, or its intensity, that only utilize the weak assumption of regular variation in the jump tails, along with in-fill asymptotic arguments for uniquely identifying the "large" jumps from the data. Importantly, the procedure permits very general dynamic dependencies in the jump tails, and does not depend upon the behavior of the "small" jumps. Nor does it restrict the dynamic dependencies in the continuous part of the process and the form of the stochastic volatility.

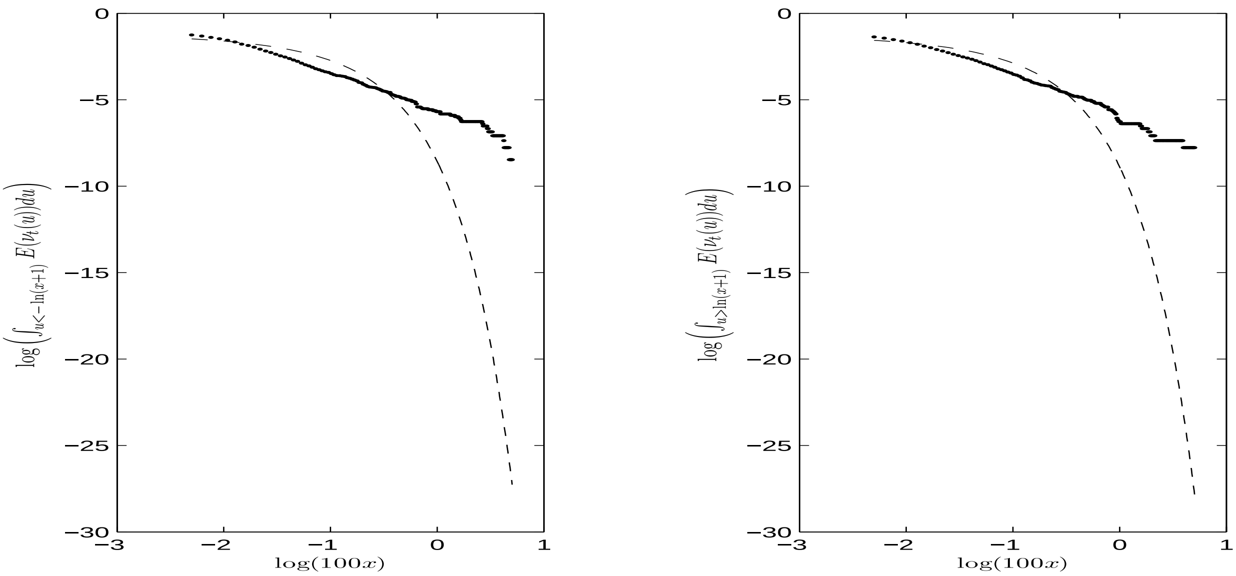
The existing empirical evidence pertaining to the behavior of jump tails in asset prices comes almost exclusively from tightly parameterized jump-diffusion models. In particular, following the seminal work of Merton (1976), most empirical studies to date have relied on relatively simple and tractable finite activity jump processes, with normally distributed jump sizes coupled with a constant jump intensity, or a jump intensity process that is affine in the diffusive stochastic variance. Although such a formulation is very convenient from an analytical perspective, anticipating our empirical findings, the data clearly suggest the existence of more complex dependencies and often larger jump tails that are formally outside this framework.

To illustrate this point, and the inability of the standard modeling framework to adequately describe the data, Figure 1 shows the unconditional empirical jump tails estimated directly from a sample of one-minute high-frequency futures data for the S&P 500 aggregate market portfolio spanning the period from January 1990 to December 2008.[2] In addition to the raw empirical

[1] The two types of risks are also very different from an economic perspective. Stochastic volatility in effect induces temporal variation in the investment opportunity set and a corresponding hedging component; see, e.g., Merton (1973). This additional risk may be spanned by an asset with payoff dependent on the stochastic volatility, e.g., an option. In contrast, the presence of jumps require a different derivative instrument for each possible jump size to completely span the jump risk. The seemingly high prices for close-to-expiration out-of-the-money puts observed in the options markets may also be seen as indirect evidence that investors demand a separate risk premium for jump tail events; see, e.g., Broadie et al. (2009).

[2] This same data also underlies our empirical illustration below, and we provide a more detailed description of the data in Section 6.

Figure 1: Empirical and Normal Jump Tails



Note: The dotted lines in the two separate panels report the left and right empirical jump tail intensities based on one-minute S&P 500 futures prices from 1990 to 2008. The dashed lines give the corresponding best fit by a Merton type model with normally distributed jump sizes. The results are reported on a double logarithmic scale.

jump intensities, we also include in the figure the jump tails implied by a model with normally distributed jump sizes estimated with the same high-frequency prices.[3] As the figure clearly shows, this now standard approach to jump modeling tends to overestimate the "medium-sized" jumps, while severely underestimating the likelihood of "large" jumps.

This points to a more fundamental problem with a fully parametric estimation of the jump tails. Parametric models generally link the behavior of the "small" and "large" jumps in a highly model-specific fashion.[4] Statistically, however, the "small" and "large" jumps are fundamentally different, and the requisite techniques for studying the relevant aspects of the jump compensator (e.g., the Lévy measure) reflect those differences. The behavior of the Lévy measure around 0 primarily captures the pathwise properties of the jump process; e.g., finite or infinite activity, finite or infinite variation, as well as the value of the Blumenthal-Getoor index. These features can only be reliably estimated using high-frequency data and corresponding fill-in asymptotic arguments; see, e.g., Aït-Sahalia and Jacod (2009b,c), Todorov and Tauchen (2010a), and Woerner (2003). On the other hand, the properties of the jump tails and the Lévy measure at infinity

[3]The parameter estimates are based on a simple method-of-moments type procedure. This estimation strategy corresponds directly to maximum likelihood when the jump intensity is constant, and it may be formalized more generally along the lines of the theoretical analysis in Todorov (2009).

[4]One example is the finite activity jump process with normally distributed jumps discussed above. Another is an α-stable process in which the *parameter* α determines both the Blumenthal-Getoor index of the jumps as well as the jump tail decay.

are strictly related to the underlying data generating process. These features therefore cannot be reliably estimated from a single realization over a fixed short time interval, but instead must be inferred using standard asymptotic arguments and the notion of an increasing sample over longer calendar time spans. Our estimation of the jump tails purposely avoids any link between "small" and "large" jumps by utilizing fill-in asymptotic arguments to isolate and directly work with the "large" jumps only, while at the same time relying on standard asymptotic arguments for reliably estimating the population characteristics.

By focusing directly on the jumps, our procedure works both for the case where the jump intensity is constant, i.e., pure Lévy type jumps, but importantly also in the practically more relevant case with time-varying jump tail intensities. Intuitively, while the jumps may cluster in time, the relative importance of differently sized jumps remains the same, leaving the ratios of the tail jump intensities constant across jump sizes. In contrast, if one were to base the inference on the price increments over *fixed* time-intervals, any clustering of the jumps would invariably impact the size of the tails of the increments and would have to be somehow accounted for in the estimation.[5]

The basic idea behind the estimation formally builds on the so-called peaks-over-threshold method together with the assumption of regular variation in the tails, as originally developed by Smith (1987) in the context of extreme-value theory with *i.i.d.* random variables. This basic minimal assumption implies, among other things, that the jump tail intensities obey a power-law for sufficiently large jump sizes.

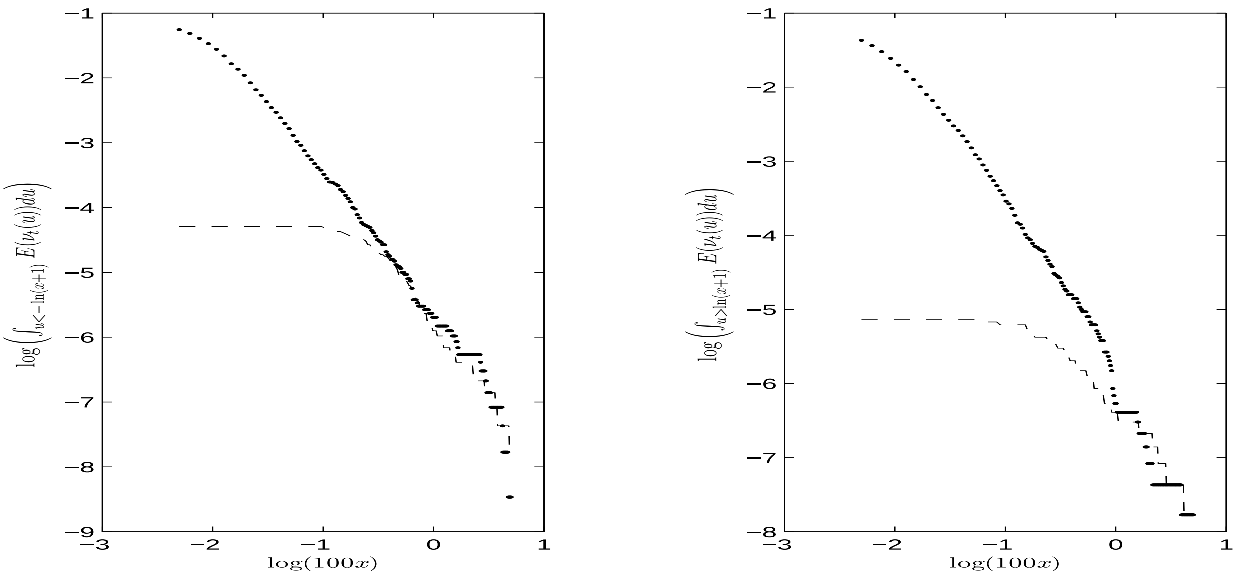
The importance of using high-frequency data for effectively identifying the "large" jumps and the jump tails is clearly illustrated by Figure 2, which compares the empirical jump tails for the S&P 500 market portfolio estimated with one- and ten-minutes returns, respectively. While the estimates coincide for the larger jump sizes, as they should, our ability to meaningfully extract the more moderate-sized jumps obviously become more limited at the ten-minute frequency. Intuitively, the coarser the sampling frequency, the more the continuous variation will obscure the identification of the jumps, and the greater cutoff values will need to be used in the jump-tail inference, in turn resulting in a loss of jump-observations and efficiency of the estimation.[6]

Figures 1 and 2 both corroborate the empirical validity of the assumed power-law decay underlying our asymptotic approximations. Importantly, however, our estimates of the jump tails go beyond the simple case of jumps with independent increments, i.e., Lévy type jumps, by explicitly incorporating dynamic dependencies in the jump tail intensities. Specifically, utilizing the assumption of regular variation in the tails, we show how appropriately rescaled and trans-

[5]Another advantage of working directly with the jumps is that our estimator does not depend upon the form of the stochastic volatility, and in particular is robust to the possibility of jumps in the volatility, as recently explored by, e.g., Jacod and Todorov (2010) and Todorov and Tauchen (2010b).

[6]Of course, the use of coarser daily frequency returns, as commonly done in the estimation of parametric jump-diffusion models, would even further exaggerate these same effects and handicap the detection of jumps.

Figure 2: Empirical Jump Tails and Sampling Frequency



Note: The two separate panels report the left and right empirical jump tail intensities based on one-minute (dotted line) and ten-minutes (dashed line) S&P 500 futures prices from 1990 to 2008. The results are reported on a double logarithmic scale.

formed versions of the tails of the jump compensators should be approximately equal to the cdf of the Generalized Pareto distribution, even for dynamically dependent jump tails. Going one step further, we show how this in turn implies that appropriately transformed - by the scores from a Generalized Pareto distribution - "large" jumps when integrated over time become approximate martingales, thus setting the stage for the construction of a moment type estimator for the jump tail parameters through the judicious choice of instruments.[7]

In practice, of course, our use of discretely sampled high-frequency data for inferring the "large" jumps invariably introduces a discretization error, the size of which is directly related to the mesh of the observation grid. In the last step of our theoretical analysis we provide formal conditions under which this error has no first-order asymptotic effect on the estimation. We further investigate the accuracy of these asymptotic based approximations through a series of Monte Carlo simulations, confirming the applicability of the feasible version of the new jump tail estimation procedure.

On actually implementing the estimators with the same high-frequency S&P 500 data underlying the average jump tail intensities depicted in the figures discussed above, we find strong evidence for temporal variation in the jump intensities and much richer and more complex dynamic dependencies in the resulting jump tails than hitherto entertained in the literature. As

[7]Even though our procedure is distinctly non-parametric in nature, it has the appealing feature that it corresponds directly to parametric maximum likelihood when the tail decay obeys an exact power-law.

such, our new econometric modeling framework developed in the paper has the potential of allowing for jump tail forecasting, and in turn can be used to provide a deeper economic understanding of the tail events of the types observed during the recent financial crises.

The rest of the paper is organized as follows. Section 2 introduces the basic notation and key assumptions. Section 3 describes the main idea behind the new estimation method and the relevant asymptotic results when continuous price records are available. Section 4 extends the analysis to the practically relevant situation of discretely sampled prices. The practical applicability of the new estimator is confirmed through a series of Monte Carlo simulations presented in Section 5. Section 6 discusses the empirical estimation results for the S&P 500 market index, and our findings related to the rich dynamic dependencies inherent in the jump tails of that portfolio. Section 7 concludes. All proofs are deferred to Section 8.

2 Setup and Assumptions

To set up the notation, let $p_t := \ln(P_t)$ denote the logarithmic price of a financial asset. Without loss of generality, we will assume that the log-price follows an Itô semimartingale defined on some filtered probability space, i.e.,

$$dp_t = \alpha_t dt + \sigma_t dW_t + \int_{\mathbb{R}} \kappa(x)\tilde{\mu}(dt, dx) + \int_{\mathbb{R}} \kappa'(x)\mu(dt, dx), \tag{2.1}$$

where α_t and σ_t are both locally bounded processes; W_t denotes a Brownian motion; μ is a one-dimensional measure on $[0, \infty) \times \mathbb{R}$ that counts the number of jumps of given size over a given interval of time; the compensator of the jump measure is denoted by $\nu_t(x)dxdt$, where $\tilde{\mu}(dt, dx) := \mu(dt, dx) - \nu_t(x)dxdt$ refers to the corresponding compensated measure; $\kappa(x)$ is a continuous function with bounded support equal to x around the origin, with $\kappa'(x) = x - \kappa(x)$. We will also assume throughout that $\nu_t(dx)$ is absolutely continuous with respect to Lebesgue measure, i.e., $\nu_t(dx) = \nu_t(x)dx$.[8] The main contribution of the paper is to provide a new, essentially model-free, robust framework for the estimation of the tail behavior of $\nu_t(x)$, leaving other aspects of the date generating process in equation (2.1), including the drift term α_t and the stochastic volatility σ_t, as well as the activity level of the jumps, unspecified.

As noted in the introduction, the existing evidence concerning the empirical features of $\nu_t(x)$ for large values of the jump sizes x come almost exclusively from tightly parameterized jump-diffusion models. In particular, following Merton (1976), most empirical studies to date have relied on relatively simple and tractable compound Poisson jump processes with normally distributed jump sizes. Under this specification the Lévy measure in equation (2.1) may be expressed

[8]With a slight abuse of notation we will henceforth refer to $\nu_t(x)$ as the Lévy measure instead of the correct technical term "intensity of the jump compensator with respect to the Lebesgue measure."

as $\nu_t(x) = \lambda_t e^{-(x-\mu)^2/(2\sigma^2)}(2\pi\sigma^2)^{-1/2}$, where λ_t denotes some predictable stochastic process intended to capture the time-varying probability of jump arrivals, typically postulated to be a linear function of the stochastic volatility σ_{t-}^2. While such a formulation is very convenient from an analytical perspective, Figure 1 above clearly shows that such a specification doesn't necessarily fit the tails very well.

As also noted in the introduction, another problem with fully parametric approaches to estimating the jump tails, is that they generally link the behavior of the "small" and the "large" jumps in a highly model-specific fashion. Statistically, however, the "small" and the "large" jumps are very different. The behavior of the Lévy measure around 0 captures mainly the pathwise properties of the jump process; e.g., finite or infinite activity, finite or infinite variation. These features are succinctly summarized by the generalized version of the Blumenthal-Getoor index recently proposed by Aït-Sahalia and Jacod (2009b),

$$\beta := \inf\left\{p : \int_0^T \int_{\mathbb{R}} (|x|^p \wedge 1)\mu(ds, dx) < \infty\right\}. \tag{2.2}$$

The index depends directly on the sample path of the jump process over $[0, T]$ and takes on values in $[0, 2]$, with more "vibrant" trajectories implying higher values. Importantly, however, the value of β is determined solely by the "small" jumps.

In contrast, the last term in equation (2.1) and the jump tails only depend on the "large" jumps.[9] Indeed, our basic minimal assumptions related to $\nu_t(x)$, as stated in A1 and A2 immediately below, only concern the behavior of the "large" jumps, and put no restrictions on the jump activity per se.

Assumption A1. *The jump compensator $\nu_t(x)$ satisfies,*

$$\nu_t(x) = (\varphi_t^+ 1_{\{x>0\}} + \varphi_t^- 1_{\{x<0\}})\nu(x), \tag{2.3}$$

where $\varphi_t^\pm$ are nonnegative-valued stochastic processes with càdlàg paths, and $\nu(x)$ is a positive measure on $\mathbb{R}$ with $\int_{\mathbb{R}}(|x|^2 \wedge 1)\nu(x)dx < \infty$.

Assumption A1 factors the dependence in the jump compensator on time (t) and jump size (x) into two separate functions. This implies that differently sized jumps will have the same dynamic properties.[10] All parametric jump specifications used to date, e.g., time-changed Lévy processes, satisfy this assumption. Still, the assumption is slightly stronger than what we actually need, and it would be possible to relax A1 to hold only for sufficiently large values of $|x|$. However, to avoid the unnecessary (and trivial) complications that arise in this situation, we will maintain A1 in its current form.

[9] Note that $\kappa'(x) = 0$ for x close to 0.

[10] Another implication is that the Blumenthal-Getoor index is deterministic and given by $\inf\left\{p : \int_{\mathbb{R}}(|x|^p \wedge 1)\nu(x)dx < \infty\right\}$.

Our interest center on the tail behavior of the Lévy density $\nu(x)$, which in turn determines the tail behavior of the jumps in the price process. Our next assumption concerns the variation in the tails of $\nu(x)$. Specifically, define the functions $\psi(x) := e^{|x|} - 1$, and

$$\psi^+(x) := \begin{cases} \psi(x) & x > 0, \\ 0 & x \leq 0 \end{cases} \qquad \psi^-(x) := \begin{cases} 0 & x > 0, \\ \psi(x) & x \leq 0. \end{cases}$$

Also, denote $\nu_\psi^+(y) = \frac{\nu(\ln(y+1))}{y+1}$ and $\nu_\psi^-(y) = \frac{\nu(-\ln(y+1))}{y+1}$ for $y \in (0, \infty)$. Then for every measurable set $\mathbf{A}$ in $(0, \infty)$, $\int_{(0,\infty)} 1_{\{x \in \mathbf{A}\}} \nu_\psi^+(x) dx = \int_{\mathbb{R}_+} 1_{\{e^x - 1 \in \mathbf{A}\}} \nu(x) dx$ and $\int_{(0,\infty)} 1_{\{x \in \mathbf{A}\}} \nu_\psi^-(x) dx = \int_{\mathbb{R}_-} 1_{\{e^{-x} - 1 \in \mathbf{A}\}} \nu(x) dx$. Moreover, denote the tails of the measures by $\overline{\nu}_\psi^\pm(x) := \int_x^\infty \nu_\psi^\pm(u) du$, for some $x > 0$. The function $\psi(x)$ maps the positive and negative jumps to $(0, \infty)$, with the Lévy densities for the transformed jumps given by the $\nu_\psi^\pm$ measures. Assumption A2 imposes regular variation for the tails of the latter.

Assumption A2.

(a) $\overline{\nu}_\psi^\pm(x)$ *are regularly varying at infinity functions; i.e.,* $\overline{\nu}_\psi^\pm(x) = x^{-\alpha^\pm} L^\pm(x)$, *where* $\alpha^\pm > 0$, *and* $L^\pm(x)$ *are slowly varying at infinity.*[11]

(b) $L^\pm(x)$ *satisfy the condition* $L^\pm(tx)/L^\pm(x) = 1 + O(\tau^\pm(x))$ *as* $x \uparrow \infty$ *for* $t > 0$, *where* $\tau^\pm(x) > 0$, $\tau^\pm(x) \to 0$ *as* $x \uparrow \infty$, *and* $\tau^\pm(x)$ *are nonincreasing.*

Assumption A2 is key to our analysis and several comments are in order. First, the close to linear behavior of the empirical jump tail estimates for the "large" jumps depicted in Figure 1 is directly in line with A2(a). Second, A2(a) rules out Lévy measures with light tails, i.e., Merton-type jumps, whose tails belong in the maximum domain of attraction of the Gumbel distribution; see e.g., Embrechts et al. (2001).[12] Third, the decay of the tail measures $\overline{\nu}_\psi^\pm(x)$ is directly linked to the fat-tailedness of the transformed jumps $\psi(\Delta p_t)$. In particular, the integrability of $\int_t^{t+a} \int_{\mathbb{R}} |\psi(x)|^p \mu(ds, dx)$ depends on whether $p \gtrless \alpha^\pm$. A2(a) therefore implies that all powers of the jumps in the logarithmic price exist. Alternatively, one could assume that A2(a) holds for $\overline{\nu}^\pm(x)$ instead of $\overline{\nu}_\psi^\pm(x)$, where $\overline{\nu}^+(x) = \int_x^\infty \nu(u) du$ and $\overline{\nu}^-(x) = \int_{-\infty}^{-x} \nu(u) du$ for $x > 0$. Or equivalently, that the continuously-compounded returns $\ln\left(\frac{p_t}{p_{t-}}\right)$, instead of the "discrete" returns $\frac{P_t - P_{t-}}{P_{t-}}$, should be modeled with Lévy densities with power decay in their tails. We think the former is less appealing from an economic perspective.[13] Note also that in

[11] A function $L(x)$ is said to be slowly varying at infinity if $\lim_{x \to \infty} \frac{L(tx)}{L(x)} = 1$ for every $t > 0$.

[12] Although the estimation method developed below could be adopted to cover this case as well, parts of the proof would require slightly different techniques, but since this arguably isn't empirically relevant, we do not consider it here.

[13] Assuming a heavy-tailed distribution for the continuously-compounded returns would imply an infinite conditional variance for the price level, which in turn can result in infinite option prices, and as conjectured by Merton

contrast to the tail behavior, the behavior of the Lévy density around 0 is not affected by the transformation $\psi(x)$, since $\psi(x) \sim x$ for $|x| \to 0$, so that the value of the Blumenthal-Getoor index remains unaltered.

The second part A2(b) of the assumption is taken directly from Smith (1987); see also Goldie and Smith (1987). It essentially limits the deviation of the tail measures $\overline{\nu}_{\psi}^{\pm}(x)$ from the power law. We will use this assumption in determining the rate of convergence and establishing asymptotic normality of the estimates for the jump-tail probabilities.

Our next assumption imposes minimal stationarity and integrability conditions on $\varphi_t^{\pm}$. This assumption is needed to ensure that the standard long span asymptotics works in conjunction with the other assumptions for consistently inferring the jump tails.

Assumption A3. *$\varphi_s^{\pm}$ are stationary processes with $\mathbb{E}\,|\varphi_s^{\pm}|^{1+\iota} < K$, for some $K > 0$ and $\iota > 0$,*

Our final assumption restricts $\varphi_s^{\pm}$ to be an Itô semimartingale. It also imposes some weak additional integrability conditions on the stochastic processes that appear in the definition of p_t. We need this assumption in the empirically realistic situation when the price process is only observed at discrete points in time.

Assumption A4.

(a) $\varphi_t^{\pm}$ are Itô semimartingales satisfying,

$$\begin{aligned}\varphi_t^{\pm} =& \varphi_0^{\pm} + \int_0^t \alpha_u^{\pm\prime} du + \int_0^t \sigma_u^{\pm\prime} dW_u + \int_0^t \sigma_u^{\pm\prime\prime} dB_u + \int_0^t \int_{\mathbb{R}^2} \kappa(\delta^{\pm}(u-,\mathbf{x}))\tilde{\mu}'(ds, d\mathbf{x}) \\ &+ \int_0^t \int_{\mathbb{R}^2} \kappa'(\delta^{\pm}(u-,\mathbf{x})), \mu'(ds, d\mathbf{x})\end{aligned} \tag{2.4}$$

where B_t is a Brownian motion orthogonal to W_t, the processes $\alpha_t^{\pm\prime}$, $\sigma_t^{\pm\prime}$ and $\sigma_t^{\pm\prime\prime}$, and the functions $\delta^{\pm}$ in their first argument, all have càdlàg paths, and μ' is a Poisson measure on $\mathbb{R}^2$ with independent marginals, the first of which counts the jumps, with compensator $\nu_t(x_1)dx_1 \otimes \nu'(x_2)dx_2$, for $\nu'(\cdot)$ a valid Lévy density.

(b) *For every $p > 0$ and every $t > 0$,*

$$\mathbb{E}\left|\int_0^t (|\alpha_s| + \sigma_s^2 + |\alpha_s^{\pm\prime}| + (\sigma_s^{\pm\prime})^2 + (\sigma_s^{\pm\prime\prime})^2) ds + \int_0^t \int_{\mathbb{R}^2} (\delta^{\pm}(s-,\mathbf{x}))^2 \mu'(ds, d\mathbf{x})\right|^p < K_p, \tag{2.5}$$

where $K_p > 0$.

(1976) might result in infinite equilibrium interest rates. In practice, of course, it is impossible to differentiate whether assumption A2(a) holds for $\overline{\nu}_{\psi}^{\pm}(x)$ or $\overline{\nu}^{\pm}(x)$, as the difference between $\ln\left(\frac{p_t}{p_{t-}}\right)$ and $\frac{P_t - P_{t-}}{P_{t-}}$ is very small for the size of jumps that we observe.

Assumption A4(a) is very weak. It is easily satisfied for virtually all parametric jump specifications used in the literature to date, including the most commonly applied affine jump-diffusions. The assumption also allows for so-called self-exciting jump processes in which $\varphi_t^{\pm}$ depend directly on the jump measure μ, as in, e.g., Todorov (2010).

This completes our discussion of the basic setup and assumptions underlying the new jump tail estimation procedures. We begin in the next section with a discussion of the infeasible case in which continuously recorded prices are available for the estimation. This obviously facilitates the estimation, as it allows us to perfectly separate the continuous from the discontinuous price moves. We subsequently extend the analysis in Section 4 to the empirically realistic case when prices are only observed at discrete points in time.

3 Estimation of Jump Tails: Continuous Price Records

The basic idea behind our estimation scheme builds on the three assumptions in A1-A3 and the relevant extreme value theory type approximations for appropriately transformed versions of the jump tails. The common approach for assessing tail behavior in extreme value theory rely on discretely sampled prices, or returns, and a corresponding estimate of the tail index; see, e.g., Embrechts et al. (2001) and the references therein. Importantly, however, we are after the tail behavior of the jump measure μ itself, as opposed to that of the discrete returns.[14]

In general, there is not a direct link between the tails of the discrete returns and the Lévy measure of the price process. For one, time-varying volatility in the continuous part of the price process, as determined by σ_t, invariably impacts the tails of the discrete returns.[15] Secondly, temporal dependencies in the jump intensity itself, i.e., the dependence of $\nu_t(x)$ on t, also affects these tails. While, it would be possible to circumvent the first problem in the continuous-record case by looking only at the jump increments, any time-variation in the jump intensity would still blur the link between the tails of the latter and the tails of $\nu(x)$ in the decomposition in A1.

Instead, we base our inference directly on the jumps, or in the case of discretely observed prices estimates thereof, and a set of moment conditions for the jump intensity $\nu_t(x)$ derived from assumptions A1 and A2. Using the fact that the random jump measure μ differs from its compensator by a martingale, we translate the moment conditions for $\nu_t(x)$ to a set of moment conditions for μ that involve the actual in-sample jumps. To conserve space, we focus our discussion on the estimation of the right tail only; the estimation of the left tail proceeds completely

[14] In the case of a Lévy process, it is well known that, under fairly general conditions, the tail of discrete increments would be proportional to the tail of its Lévy measure; see e.g., Rosinski and Samorodnitsky (1993), Theorem 2.1. Intuitively, since the continuous part of any Lévy increment is normal, its contribution becomes negligible sufficiently "deep" in the tail.

[15] Following Leadbetter et al. (1983) and Leadbetter and Rootzen (1988), the extremes of two sequences of discrete returns with the same marginal law, one with and the other without any temporal dependencies, is generally different.

analogous.

We begin by approximating the distribution of $1 - \frac{\overline{\nu}_\psi^+(y)}{\overline{\nu}_\psi^+(x)}$ for $y \geq x > 0$, using A1 and A2. We then rely on the scores from this approximating distribution to define a set of feasible estimating equations based on the observed "large" jumps. This idea originates in the so-called peaks-over-thresholds method for estimating the tails, and the tail decay, of *i.i.d.* random variables developed by Smith (1987); see also the references therein.

Specifically, as formally shown in the Appendix, it follows from assumption A2 that

$$1 - \frac{\overline{\nu}_\psi^+(u+x)}{\overline{\nu}_\psi^+(x)} \overset{\text{appr}}{\sim} G(u;\sigma^+,\xi^+) = 1 - \left(1 + \xi^+ u/\sigma^+\right)^{-1/\xi^+}, \quad \xi^+ \neq 0,\ \sigma^+ > 0, \tag{3.1}$$

where $u > 0$, $x > 0$ is some "large" value, $G(u;\sigma^+,\xi^+)$ denotes the cdf of a Generalized Pareto distribution with parameters $\sigma^+ = \frac{x}{\alpha^+}$ and $\xi^+ = \frac{1}{\alpha^+}$, and the tail decay parameter α^+ determined by A2(a). Let the scores associated with the log-likelihood function of the generalized Pareto distribution be denoted by,

$$\begin{aligned} \phi_1^+(u,\xi^+,\sigma^+) &= \tfrac{1}{\sigma^+}\left(\tfrac{1}{\xi^+} - \left(1+\tfrac{1}{\xi^+}\right)\left(1+\tfrac{\xi^+ u}{\sigma^+}\right)^{-1}\right), \\ \phi_2^+(u,\xi^+,\sigma^+) &= \tfrac{1}{(\xi^+)^2}\log\left(1+\tfrac{\xi^+ u}{\sigma^+}\right) - \tfrac{1}{\xi^+}\left(1+\tfrac{1}{\xi^+}\right) + \tfrac{1}{\xi^+}\left(1+\tfrac{1}{\xi^+}\right)\left(1+\tfrac{\xi^+ u}{\sigma^+}\right)^{-1}. \end{aligned} \tag{3.2}$$

where $i = 1, 2$ refer to the derivative with respect to σ^+ and ξ^+, respectively.

The idea is then to pick a "large" threshold tr_T, and fit the scores to the jumps above this threshold. Doing so results in the following set of moment conditions involving the realized "large" jumps,

$$g_T(\theta, tr_T) = \frac{1}{M_T^+}\sum_{t=1}^{T-1}\begin{pmatrix} \int_t^{t+1}\int_{\mathbb{R}} \phi_1^+(\psi(x) - tr_T, \theta^{(1)}, \theta^{(2)}) 1_{\{\psi^+(x) > tr_T\}}\mu(ds,dx) \\ \int_t^{t+1}\int_{\mathbb{R}} \phi_2^+(\psi(x) - tr_T, \theta^{(1)}, \theta^{(2)}) 1_{\{\psi^+(x) > tr_T\}}\mu(ds,dx) \end{pmatrix}, \tag{3.3}$$

where θ denotes the 2×1 vector of unknown parameters, and

$$M_T^+ = \sum_{t=1}^{T-1}\int_t^{t+1}\int_{\mathbb{R}} 1_{\{\psi^+(x) > tr_T\}}\mu(ds,dx), \tag{3.4}$$

counts the number of positive in-sample jumps that are above the threshold tr_T when transformed by $\psi(\cdot)$. In theory, of course, tr_T will have to grow to infinity with the sample size T. Denote the true parameter values implicitly defined by the moment conditions $\theta_T^0 = \left(\xi^+, \sigma_T^+\right)'$, where $\sigma_T^+ = \frac{tr_T}{\alpha^+}$ increases with the sample size T. We then have the following theorem.[16]

[16] Instead of the approximation in (3.1), we could have used the score based on the approximation $\frac{\overline{\nu}_\psi^+(u+x)}{\overline{\nu}_\psi^+(x)} \overset{\text{appr}}{\sim} \left(1 + \frac{u}{x}\right)^{-1/\xi^+}$, $\xi^+ \neq 0$, obtained upon substitution of the true value of $\sigma^+ = \frac{x}{\alpha^+}$ in (3.1). This approach would involve only a single parameter, and it could be seen as an analogue of Hill (1975)'s estimator in the jump tail setting. However, such an estimator would not be scale free, and the analysis in Smith (1987) also suggests that it would be less robust than the estimator in Theorem 1.

Theorem 1 *For the process p_t defined in equation (2.1), assume that assumptions A1-A3 hold. Let the sequence of truncation levels satisfy*

$$tr_T \to \infty, \quad T\overline{\nu}_\psi^\pm(tr_T) \to \infty, \quad \text{and} \quad \sqrt{T\overline{\nu}_\psi^\pm(tr_T)}\tau^\pm(tr_T) \to 0, \tag{3.5}$$

where $\tau^+(\cdot)$ is defined in A2(b). Then with probability approaching one, $g_T(\theta, tr_T) = 0$ has a solution $\widehat{\theta}_T := \left(\widehat{\xi}^+, \widehat{\sigma}_T^+\right)'$, which satisfies

$$\sqrt{M_T^+}\begin{pmatrix} \widehat{\xi}^+ - \xi^+ \\ \widehat{\sigma}^+/\sigma_T^+ - 1 \end{pmatrix} \xrightarrow{\mathcal{L}} \sqrt{c}Z, \tag{3.6}$$

where Z is a standard multivariate normal, and $\sqrt{c}$ is an arbitrary square root of the non-random square matrix c.

Unlike conventional asymptotic theory, the scaling factor for the difference between the estimated and true parameters that control the tails is given by the random number M_T^+. Of course, $M_T^+/(T\overline{\nu}_\psi^+(tr_T)) \xrightarrow{\mathbb{P}} 1$, and $T\overline{\nu}_\psi^+(tr_T)$ is non-random. However, since $\overline{\nu}_\psi^+(tr_T)$ converges to 0, the rate of convergence is in general slower than the standard $\sqrt{T}$ rate. In particular, it follows from the conditions for the truncation level in (3.5), that the larger the deviations of the tail from the power-law decay, i.e., the slower the rate at which $\tau^+(x)$ goes to zero as $x \uparrow \infty$, the slower the rate of convergence of the estimator. Intuitively, the further are the tails from the eventual power-law decay, the larger the required truncation level, which in turn slows down the rate of convergence as fewer observations are employed in the estimation.

To further appreciate this result, suppose that $\tau^+(x) = |x|^{-k}$ for some $k > 0$. In this situation, the required rate condition in (3.5) stipulates that $tr_T = O\left(T^{\frac{1}{\alpha^+ + 2k}}\right)$, so that for $k \to \infty$, i.e., $L^+(x)$ in A2 converging to unity and diminishing deviations from the power-law, it is possible to get arbitrarily close to the standard parametric $\sqrt{T}$ rate of convergence for optimally chosen tr_T.

In practice, of course, we do not know a-priori the form of the slowly-varying function $L^+(x)$ that dictates the optimal choice of the truncation level, and we are faced with a tradeoff in terms of robustness versus efficiency in the estimation. A low value of tr_T would entail the use of more observations, i.e., more jumps, and hence a more efficient estimator. On the other hand, by choosing tr_T too small, we run the risk of larger deviations from the eventual power-law tail decay and non-robustness of the estimation. We will explore these tradeoffs more fully in the Monte Carlo simulations reported in Section 5 below.

Importantly, the estimating equations in (3.3) correctly identify the tail behavior of $\nu(x)$, even in the presence of time-varying jump intensities. Intuitively, temporal dependence in the jump intensity does not affect the distribution of the "large" jumps, and as such the presence of more jumps in certain periods does not systematically bias the estimator. In contrast, any estimator based on the jump increments over fixed intervals of time, e.g., days, would invariably be affected by jump clustering and a failure to properly account for that effect would result in biased tail index estimates.

Even though Theorem 1 allows for jump clustering, it doesn't fully exploit the dynamic structure of the jump tails implied by assumptions A1-A2, relying merely on the unconditional approximations, $\int_{\mathbb{R}} \phi_i^+(\psi(x) - tr_T, \theta_T^{0(1)}, \theta_T^{0(2)})\nu(x)dx \approx 0,\ i = 1, 2$. Going one step further, it follows from the formal proofs in the Appendix that for $t, s \geq 0$,[17]

$$\begin{aligned}\mathbb{E}_t\left(\int_t^{t+s}\int_{\mathbb{R}} \phi_i^+(\psi(x) - tr_T, \theta^{(1)}, \theta^{(2)})\mu(du, dx)\right) &= \mathbb{E}_t\left(\int_t^{t+s}\int_{\mathbb{R}} \phi_i^+(\psi(x) - tr_T, \theta^{(1)}, \theta^{(2)})\tilde{\mu}(du, dx)\right) \\ &+ \int_{\mathbb{R}} \phi_i^+(\psi(x) - tr_T, \theta^{(1)}, \theta^{(2)})\nu(x)dx\mathbb{E}_t\left(\int_t^{t+1} \varphi_u^+ du\right) \\ &= \int_{\mathbb{R}} \phi_i^+(\psi(x) - tr_T, \theta^{(1)}, \theta^{(2)})\nu(x)dx\mathbb{E}_t\left(\int_t^{t+s} \varphi_u^+ du\right) \approx 0,\end{aligned}$$

where we have used the shorthand notation $\mathbb{E}_t(\cdot) = \mathbb{E}(\cdot|\mathcal{F}_t)$. In particular, for any x_t adapted to $\mathcal{F}_t$,

$$\mathbb{E}\left(x_t \int_t^{t+1}\int_{\mathbb{R}} \phi_i^+(\psi(x) - tr_T, \theta^{(1)}, \theta^{(2)})\mu(ds, dx)\right) \approx 0. \tag{3.7}$$

This in turn suggests the following extension of Theorem 1.

Theorem 2 *For the process p_t defined in equation (2.1), assume that assumptions A1-A3 hold. Let the sequence of truncation levels satisfy condition (3.5) of Theorem 1. Define the vector of moment conditions*

$$g_T(\theta, tr_T) = \frac{1}{M_T^+}\sum_{t=1}^{T-1} \mathbf{x}_t \otimes \begin{pmatrix} \int_t^{t+1}\int_{\mathbb{R}} \phi_1^+(\psi(x) - tr_T, \theta^{(1)}, \theta^{(2)})1_{\{\psi^+(x) > tr_T\}}\mu(ds, dx) \\ \int_t^{t+1}\int_{\mathbb{R}} \phi_2^+(\psi(x) - tr_T, \theta^{(1)}, \theta^{(2)})1_{\{\psi^+(x) > tr_T\}}\mu(ds, dx)\end{pmatrix}, \tag{3.8}$$

where $\mathbf{x}_t$ is a $q \times 1$ $\mathcal{F}_t$-adapted stationary vector process, satisfying $\mathbb{E}||\mathbf{x}_t||^{2+\iota} < \infty$ for some $\iota > 0$ and such that a law of large numbers holds for $\frac{1}{T}\sum_{t=1}^{T-1} \mathbf{x}_t \int_t^{t+1} \varphi_s^+ ds$ and $\frac{1}{T}\sum_{t=1}^{T-1} \mathbf{x}_t\mathbf{x}_t' \int_t^{t+1} \varphi_s^+ ds$ with $\mathbb{E}\left(\mathbf{x}_t \int_t^{t+1} \varphi_s^+ ds\right) \neq \mathbf{0}$. Further, let $\widehat{W}_T$ denote a sequence of positive semidefinite $2q \times 2q$ matrices, such that $\widehat{W}_T \overset{\mathbb{P}}{\to} W$, where W is a $2q \times 2q$ positive definite matrix. Denote $\widehat{\theta}_T = \text{argmin}_{\theta \in \Theta_T^l} g_T(\theta, tr_T)'\widehat{W}_T g_T(\theta, tr_T)$, where $\Theta_T^l = \left\{\theta : \alpha_l \theta_T^{0(i)} \leq \theta^{(i)} \leq \alpha_h \theta_T^{0(i)}, i = 1, 2\right\}$ for some constants $0 < \alpha_l < 1 < \alpha_h$. Then $\widehat{\theta}_T$ exists with probability approaching one, and

$$\sqrt{M_T^+}\begin{pmatrix} \widehat{\xi}^+ - \xi^+ \\ \widehat{\sigma}^+/\sigma_T^+ - 1\end{pmatrix} \overset{\mathcal{L}}{\to} \sqrt{c}Z, \tag{3.9}$$

where Z is a standard multivariate normal, and $\sqrt{c}$ is an arbitrary square root of the non-random square matrix c.

[17] Recall that the counting jump measure μ is not a martingale, but that its compensation version, $\tilde{\mu}(ds, dx) = \mu(ds, dx) - \nu_t(x)dxdt$, is.

Aside from being properly adapted to $\mathcal{F}_t$, Theorem 2 poses no restrictions on the $\{x_t\}$ process. Given additional assumptions about the underlying model structure, it might be theoretically possible to solve for the optimal set of instrument(s) following, e.g., Hansen (1985), and in certain settings, i.e., affine specifications, also practically feasible, see, e.g., Nagel and Singleton (2010). However, doing so in general is very challenging, and we will not pursue that issue further here. Instead, we will discuss ways in which to expand on the basic moment conditions in the theorem in an effort to provide more detailed information about the dynamic tail dependencies.

Thus far our focus has centered on recovering the tail properties of $\nu(x)$. In most practical applications, however, one would be interested in the tails of $\nu_t(x)$. Building on the decomposition for $\nu_t(x)$ in assumption A1 into its time-varying components $\varphi_t^{\pm}$, it follows that for the "large" jumps, the difference

$$\int_0^t \int_{\mathbb{R}} \phi(s-,x)\mu(ds,dx) - \int_0^t \int_{\mathbb{R}} \phi(s-,x)\varphi_s^{+} ds \nu(x) dx,$$

must be a martingale for any function $\phi(s,x)$ with càdlàg paths, and $\phi(s,x) = 0$ for $x < K$, where $K > 0$ denotes some constant. In parallel to the discussion above, this in turn allows for the construction of a set of unconditional estimating equations through the appropriate choice of instrument(s) $\mathbf{x}_t$.

In general, of course, the resulting moments will depend on the exact specification of the $\varphi_t^{\pm}$ processes. To illustrate, we next consider the special case in in which the time-varying part of the right jump intensity is assumed to be an affine function of the spot variance, i.e., $\varphi_t^{+} = k_0^{+} + k_1^{+}\sigma_t^2$. This same basic assumption also underlies our empirical illustration in Section 6 below.

3.1 Affine Jump Intensities

The assumption that the temporal dependencies in the jump intensities are affine in the spot volatility nests virtually all parametric jump-diffusion models hitherto considered in the literature, including the affine jump-diffusion class of models popularized by Duffie et al. (2000). Importantly, however, by making no parametric assumptions about the volatility process itself, the semi-parametric setup is much more flexible, allowing for the possibility of so-called self-exciting jumps and models in which σ_t depends on the jump measure μ.

The maintained assumption of continuous price records underlying all of the results in this section and our ability to perfectly identify the "large" jumps, similarly allows us to perfectly infer the integrated variation $\int_{t-1}^{t} \sigma_s^2 ds$. As discussed further below, the integrated variation is also relatively easy to accurately estimate empirically in a non-parametric fashion. This in turn suggests using that measure to help identify the dependence of φ_t^{+} on σ_t^2. The following corollary extends the results above to cover this situation.

Corollary 1 *For the process p_t defined in equation (2.1), assume that assumptions A1-A3 hold, and that $\varphi_t^+ = k_0^+ + k_1^+\sigma_t^2$. Denote $\theta = (\xi^+, \sigma^+, k_0^+\overline{\nu}_\psi^+(tr_T), k_1^+\overline{\nu}_\psi^+(tr_T))$, and define the vector of moment conditions,*

$$g_T(\theta, tr_T) = \frac{1}{M_T^+}\sum_{t=1}^{T-1} \mathbf{x}_t \otimes \begin{pmatrix} \int_t^{t+1}\int_{\mathbb{R}} \phi_1^+(\psi(x) - tr_T, \theta^{(1)}, \theta^{(2)}) 1_{\{\psi^+(x)>tr_T\}}\mu(ds,dx) \\ \int_t^{t+1}\int_{\mathbb{R}} \phi_2^+(\psi(x) - tr_T, \theta^{(1)}, \theta^{(2)}) 1_{\{\psi^+(x)>tr_T\}}\mu(ds,dx) \\ \int_t^{t+1}\int_{\psi(x)>tr_T}\mu(ds,dx) - \theta^{(3)} - \theta^{(4)}\int_t^{t+1}\sigma_s^2 ds \end{pmatrix}, \quad (3.10)$$

with $\mathbf{x}_t = \begin{pmatrix} 1 & \int_{t-1}^t \sigma_s^2 ds\end{pmatrix}'$. Assume that the growth condition for tr_T in equation (3.5) is satisfied and that a law of large numbers holds for $\frac{1}{T}\sum_{t=1}^{T-1}\int_{t-1}^t \sigma_s^2 ds \int_t^{t+1}\sigma_s^2 ds$ and $\frac{1}{T}\sum_{t=1}^{T-1}\left(\int_{t-1}^t \sigma_s^2 ds\right)^2 \int_t^{t+1}\sigma_s^2 ds$ with $\mathbb{E}|\sigma_t|^{6+\iota} < \infty$ for some $\iota > 0$. Finally, let $\widehat{W}_T$ denote a sequence of positive semidefinite 6×6 matrices, such that $\widehat{W}_T \xrightarrow{\mathbb{P}} W$, where W is a 6×6 positive definite matrix, and denote $\widehat{\theta}_T = argmin_{\theta\in\Theta_T^l} g_T(\theta, tr_T)'\widehat{W}_T g_T(\theta, tr_T)$.

(a) The estimator $\widehat{\theta}_T$ then exists with probability approaching one, and

$$\sqrt{M_T^+}\begin{pmatrix} \widehat{\xi}^+ - \xi^+ \\ \widehat{\sigma}^+/\sigma_T^+ - 1 \\ \widehat{k_0^+\overline{\nu}_\psi^+(tr_T)}/\overline{\nu}_\psi^+(tr_T) - k_0^+ \\ \widehat{k_1^+\overline{\nu}_\psi^+(tr_T)}/\overline{\nu}_\psi^+(tr_T) - k_1^+ \end{pmatrix} \xrightarrow{\mathcal{L}} \sqrt{c}Z,$$

where Z is a standard multivariate normal, and $\sqrt{c}$ is an arbitrary square root of the non-random square matrix c.

(b) Let $z_T = \eta tr_T$ for some constant $\eta \geq 1$, and denote

$$\widehat{k_i^+\overline{\nu}_\psi^+(z_T)} = \widehat{k_i^+\overline{\nu}_\psi^+(tr_T)}\left(1 + \frac{\widehat{\xi}^+}{\widehat{\sigma}^+}(z_T - tr_T)\right)^{-1/\widehat{\xi}^+}, \quad i = 0, 1. \quad (3.11)$$

Then

$$\sqrt{M_T^+}\begin{pmatrix} \widehat{k_0^+\overline{\nu}_\psi^+(z_T)}/\overline{\nu}_\psi^+(z_T) - k_0^+ \\ \widehat{k_1^+\overline{\nu}_\psi^+(z_T)}/\overline{\nu}_\psi^+(z_T) - k_1^+ \end{pmatrix} \xrightarrow{\mathcal{L}} \sqrt{d}Z, \quad (3.12)$$

where Z refers to the relevant part of the standard normal vector from (a), and d is a square matrix of constants.

The last two moment conditions serve to disentangle the constant and time-varying parts; i.e., $k_0^+\overline{\nu}_\psi^+(tr_T)$ and $k_1^+\overline{\nu}_\psi^+(tr_T)$, respectively. They may be interpreted as linear projections of the counts of "large" jumps on a constant and the integrated variation over the previous period. As such, these two moment conditions only require that the affine structure holds for the "large" jumps.

In the special case when σ_t^2 is an affine jump-diffusion, the optimal instrument x_t will also be an affine function of σ_t^2. Given the high degree of persistence in financial market volatility, our choice of the integrated variation as an instrument should therefore be very close to the infeasible optimal instrument-choice in this case. However, as previously noted, empirically it is much easier to reliably estimate the integrated, as opposed to the spot, volatility, rendering empirical equivalents to the conditions in Corollary 1 easier to mimic with actual high-frequency data.

Part (b) of the corollary shows how the estimation framework may be extended to meaningfully characterize the behavior of the jump-tails at levels for which we (invariably) have few in-sample observations. These, of course, are also the levels of interest in many risk management situations involving extreme value-at-risk type quantities. We further illustrate this important new dimension of our result in the empirical application discussion in Section 6 below.

Our formal analysis up until now has been based on the assumption of continuously recorded prices. We next discuss how this empirically unrealistic assumption may be relaxed by only having prices observed at discrete points in time.

4 Estimation of Jump Tails: Discretely Sampled Prices

The results in the previous section relied on our ability to directly identify the jumps in a continuously observed realization of process. The theoretical notion of continuous price records is, of course, practically infeasible. Instead, we will now assume that over each unit time interval $[t, t+1]$, the price process p_t is "only" observed at the discrete points in time $t, t+\Delta_n, ..., t+n\Delta_n$, for some $\Delta_n > 0$. We will refer to $n = [1/\Delta_n]$ as the number of high-frequency price observation over the "day."

In order to adapt the same basic estimation strategies to the case of actual high-frequency data, we will assume that the length of the sampling interval goes to zero, i.e., $\Delta_n \to 0$.[18] This will allow us at least in some limiting sense to identify the "large" jumps. We then proceed with the construction of feasible estimates of the same integrals with respect to the jump measures and corresponding moment conditions analyzed above, say $\widehat{g}_T(\theta, tr_T)$. These high-frequency based estimates will, of course, contain discretization errors, but we will show that under appropriate conditions, the errors shrink to zero and do not affect the estimates.

To facilitate the exposition, we will use the shorthand notation $\Delta_i^{n,t} p := p_{t+i\Delta_n} - p_{t+(i-1)\Delta_n}$ for $i = 1, ..., n$. Our estimates for the integrals $\int_t^{t+1} \int_{\psi^+(x)>tr_T} \phi_i^+(\psi(x) - tr_T, \theta^{(1)}, \theta^{(2)})\mu(ds, dx)$,

[18] The assumption of equally spaced observations is not critical, but the assumption that the largest mesh size goes to zero, or $\Delta_n \to 0$ in the case of equidistant observations, is.

$i = 1, 2$, may then be expressed as,

$$\sum_{j=1}^{n} \phi_i^+(\psi(\Delta_j^{n,t} p) - tr_T, \theta^{(1)}, \theta^{(2)}) \qquad i = 1, 2.$$

These estimators rely on the fact that the score functions $\phi_i^+(\psi(x) - tr_T, \theta^{(j)}, \theta^{(j+1)})$ are zero for values of $|x|$ in a neighborhood of zero. Using the modulus of continuity of càdlàg functions, all, but the high-frequency intervals containing the "large" jumps, can be made arbitrary small uniformly over a given fixed time-interval, and those increments therefore won't matter in the estimation of the integrals. This argument, of course, is only pathwise, and in our analysis the time span T will also increase as Δ_n goes to zero. This requires different arguments in the formal proof, but the intuition remains the same.[19]

Altogether this implies that the feasible estimation with discretely sampled high-frequency prices will be subject to three distinct types of errors: (1) sample error associated with the empirical processes employed in the moment vector, controlled by the span of the data T, (2) approximation error for the jump tail, controlled by the truncation size tr_T, and (3) discretization error from "filtering" the jumps from the high-frequency data, controlled by the length of the high-frequency interval Δ_n, or equivalently the number of high-frequency observations per unit time-interval n. The following theorem provides the requisite rate conditions on the relative speed with which T, tr_T and n need to increase, in order to ensure that the feasible estimation is equivalent to the infeasible procedures discussed in the previous section.

Theorem 3 *For the process p_t defined in equation (2.1) sampled at times $0, \Delta_n, ..., n\Delta_n, ..., t, t+\Delta_n, ..., t + n\Delta_n, ...,$ assume that assumptions A1-A4 hold, with $\nu(x)$ nondecreasing for x sufficiently large. If in the moment vector $g_T(\theta, tr_T)$ of Theorem 2, $\int_t^{t+1} \int_{\psi^+(x)>tr_T} \phi_i^+(\psi(x) - tr_T, \theta^{(1)}, \theta^{(2)})\mu(ds, dx)$ is replaced by $\sum_{j=1}^n \phi_i^+(\psi(\Delta_j^{n,t} p) - tr_T, \theta^{(1)}, \theta^{(2)})$ for $i = 1, 2$ and $t = 0, ..., T-1$, then the conclusions of that theorem continue to hold, provided that*

$$\sqrt{M_T^+ \Delta_n^{1-\iota}} \left(1 \bigvee \frac{\sqrt{\Delta_n} \log(tr_T)}{\overline{\nu}_\psi^+(tr_T)} \right) \to 0, \tag{4.1}$$

where $\iota > 0$ is arbitrary small.

As before, since the score functions are only based on the "large" jumps, or the feasible estimates thereof, the condition in (4.1) puts no restrictions on the behavior of the "small" jumps.

To better understand the rate condition in (4.1) behind the asymptotic equivalence result, it is instructive to consider the case in which $\tau^+(x) = |x|^{-k}$ for some $k > 0$. In that situation

[19] An additional complication arises from the fact that our integrands with respect to μ are discontinuous at the point x for which $\psi^+(x) = tr_T$, and this point of discontinuity changes with the time span. At the point of discontinuity, however, $\nu_t(x)$ is absolutely continuous, at least asymptotically for increasing values of $|x|$.

Theorem 2 dictates the optimal truncation level $tr_T = T^{\frac{1}{\alpha+2k}+\iota}$, which translates into the rate condition $T^{\frac{2k\wedge(\alpha^++k)}{\alpha^++2k}}\Delta_n^{1+\iota} \to 0$. Recall that $k \to \infty$ implies ever diminishing deviations from the power-decay law for the jump tails, in turn allowing for the use of lower truncation levels. That is, $k \to \infty$ may be interpreted as the "parametric limit case" of our estimation, with the relative rate condition implied by the theorem given by $T\Delta_n^{1+\iota} \to 0$. But, that condition is essentially equivalent to the corresponding condition for the estimation of diffusion processes with discretely sampled data; see e.g., Prakasa Rao (1988). Conversely, when the tail decay doesn't perfectly adhere to a power law, i.e., for finite k, we need to resort to higher truncation levels and larger sized jumps, weakening the rate condition in (4.1) somewhat relative to the fully parametric limit case.

The result in Theorem 3 is general and pertains to any discretely sampled Itô semimartingale process. We next discuss how to make the estimation for the special case of affine jump intensities, previously analyzed in Section 3.1 for the continuous record case, practically feasible.

4.1 Affine Jump Intensities

Given the feasible estimates for the integrals with respect to the jump measures discussed above, the primary obstacle in implementing the estimator in Corollary 1 relates to the integrated variation $\int_t^{t+1}\sigma_s^2 ds$. We will base our estimates for this quantity on the so-called Truncated Variation (TV) measure originally proposed by Mancini (2009); see also Jacod (2008),[20]

$$TV_t^n = \sum_{j=1}^{n}\left(\Delta_j^{n,t}p\right)^2 1_{\{|\Delta_j^{n,t}p|\le\alpha\Delta_n^{\varpi}\}}, \quad \alpha>0, \varpi\in\left(0,\frac{1}{2}\right). \tag{4.2}$$

As the formula shows, the truncated variation is simply constructed by summing the "continuous" squared price increments obtained by purging the price process of jumps, i.e., all of the price increments above the threshold $\alpha\Delta_n^{\varpi}$. Asymptotically, of course, $\Delta_n \to 0$ so that the threshold $\Delta_n^{\varpi} \downarrow 0$.[21] The following corollary provides the feasible analogue to Corollary 1 based on the TV estimator.

Corollary 2 *For the process p_t defined in equation (2.1) sampled at times $0, \Delta_n, ..., n\Delta_n, ..., t, t+\Delta_n, ..., t+n\Delta_n, ...,$ assume that assumptions A1-A4 hold, with $\nu(x)$ nondecreasing for x sufficiently large. If in the moment vector $g_T(\theta, tr_T)$ of Corollary 1, $\int_t^{t+1}\int_{\psi^+(x)>tr_T}\phi_i^+(\psi(x)-tr_T, \theta^{(1)}, \theta^{(2)})\mu(ds,dx)$ is replaced by $\sum_{j=1}^n \phi_i^+(\psi(\Delta_j^{n,t}p)-tr_T, \theta^{(1)}, \theta^{(2)})$ for $i = 1,2$ and $t =$*

[20] Alternatively, we could have used the multipower variation estimators developed by Barndorff-Nielsen and Shephard (2004, 2006).

[21] To be consistent, in our numerical implementations of the integrated jump measures, we similarly truncated the price increments from below by $\alpha\Delta_n^{\varpi}$. This obviously doesn't change anything asymptotically, as all of the estimators are based on the "large" jumps, and $\Delta_n^{\varpi} \downarrow 0$.

$0,...,T-1$, and $\int_{t-1}^{t}\sigma_s^2 ds$ is replaced by TV_t^n defined in (4.2) for $t=1,...,T$, then the conclusions of Corollary 1 continue to hold true, provided condition (4.1) of Theorem 3 is satisfied, and in addition

$$\sqrt{M_T^+\Delta_n^{1-\iota}}\Delta_n^{(2-\beta)\varpi-1}\to 0, \tag{4.3}$$

where $\iota>0$ denotes an arbitrary small constant.

In contrast to the general rate condition given by equation (4.1) in Theorem 3, the condition in (4.3) does depend on the behavior of the "small" jumps, as manifest by the presence of the Blumenthal-Getoor index β. This additional requirement arises from the need to control the size of the discretization error in estimating the integrated variation. Intuitively, the more active the jumps, the more difficult it is to separate the continuous and the jump components of the price process, and in turn the more difficult it is to estimate the integrated variation.

Meanwhile, in the numerical implementations reported on below, we systematically fix the tuning parameter ϖ to be very close to its upper bound of $\frac{1}{2}$. Hence, for values of the Blumenthal-Getoor index less than 1, i.e., jumps of finite variation, the condition in (4.3) will be automatically satisfied by (4.1).

The feasible results in Theorem 3 and Corollary 2 are, of course, still based on asymptotic approximations. To gauge the accuracy of these approximations and the practical applicability of the new jump tail estimation procedures, we next present the results from a series of Monte Carlo simulations.

5 Monte Carlo Simulations

The Monte Carlo simulation is designed to mimic the actual data analyzed in the next section. To facilitate the interpretation of the results, we will calibrate all of model parameters so that the unit time interval corresponds to a day. We will assume that the continuous spot volatility process is determined by the popular affine, also sometimes called Heston, model,

$$d\sigma_t^2 = 0.0128(0.8136-\sigma_t^2)dt + 0.0954\sigma_t dB_t, \tag{5.1}$$

where B_t is a Brownian motion independent from W_t and the particular parameter values were taken from the estimation results reported in Eraker et al. (2003).

The Lévy measure $\nu_t(x)$ for the jumps in the log-price process satisfies assumption A1 with $\varphi_t^\pm = k_0^\pm + k_1^\pm\sigma_t^2$, and Lévy density,

$$\nu(x) = \left\{c_0\frac{e^{|x|}}{(e^{|x|}-1)^{\beta_0+1}} + c_1\frac{e^{|x|}}{(e^{|x|}-1)^{\beta_1+1}}\right\}1_{\{|x|\geq 0.4\}}. \tag{5.2}$$

This density represents a mixture of two measures with tail decay parameters β_0 and β_1, respectively. In all of the simulations $\beta_0 < \beta_1$, so that the tail decay of the simulated price jumps is always determined by β_0.

Table 1: Jump Parameters

Case	Parameters					
	β_0	c_0	β_1	c_1	$k_0^+ = k_0^-$	$k_1^+ = k_1^-$
T1	2.0	0.0077	6.0	1.4746×10^{-4}	0.5	0.6146
T2	3.0	0.0046	9.0	1.4156×10^{-5}	0.5	0.6146
T3	4.0	0.0025	12.0	1.2080×10^{-6}	0.5	0.6146

Note: The table reports the values of the jump parameters used in the Monte Carlo simulations. All of the values are reported in units of daily continuously-compounded percentage returns.

We experimented with several different jump parameter configurations, the details of which are given in Table 1. The values of β_0 were chosen to cover the range of values for the tail decay for financial returns typically reported in the literature; see e.g., Embrechts et al. (2001) and the references therein. We set β_1 to be three times that of β_0. The value of β_1 essentially controls the behavior of the residual functions $L^{\pm}$ in A2(b). The two scale parameters c_0 and c_1 were chosen to satisfy the following two criteria. First, we restrict the "daily" $\overline{\nu}(|x| > 0.4) = 0.06$, where the jumps x are measured in percentages. This value approximately matches our estimate for the actual financial data reported in the next section. Second, we fix the proportion of $\overline{\nu}(x > 0.4)$ due to the second measure in (5.2) to be 20%. The values of $k_0^{\pm}$ and $k_1^{\pm}$ were chosen to ensure that the time-varying and the time-homogenous part of the jump measures are equally important, i.e., $k_0^{\pm} = k_1^{\pm}\mathbb{E}(\sigma_t^2)$.[22] Lastly, the sampling frequency $n = 400$ and time span of the data $T = 5,000$, corresponding to roughly 20 years of one-minute intraday prices over a 6.5 hours trading day, were both chosen to match the data used in the actual empirical estimation.

Our estimates of the truncated variation in (4.2) were based on $\varpi = 0.49$ and α equal to $4 \times \sqrt{BV_t \wedge RV_t}$, where BV_t denotes the bipower variation of Barndorff-Nielsen and Shephard (2004, 2006) and RV_t stands for the realized variance, both calculated over that particular "day." To gauge the sensitivity of the estimation results to the choice of truncation level, we report the results for three different values of tr_T, corresponding to jump tails equal to 0.030, 0.015, and 0.010, respectively. In parallel to the theoretical analysis, we focus on the right tail only.

All of the simulations were based on a total of 1,000 replications. To facilitate the interpretation of the results, we focus on three distinct aspects of the new estimation procedure, namely

[22] Altogether the parameters imply that the contribution of jumps to the total quadratic price variation is in the range of $5 - 15\%$. This is directly in line with the recent non-parametric empirical evidence reported in, e.g., Huang and Tauchen (2005), Andersen et al. (2007), and Aït-Sahalia and Jacod (2009a).

Table 2: Monte Carlo Simulation Results

Case	True Value	Truncation Level					
		$\overline{\nu}_{\psi}^{+}(tr_T) = 0.030$		$\overline{\nu}_{\psi}^{+}(tr_T) = 0.015$		$\overline{\nu}_{\psi}^{+}(tr_T) = 0.010$	
		Median	IQR	Median	IQR	Median	IQR
Tail Index $1/\widehat{\xi}^{+}$							
T1	2.0	2.092	[1.778 2.592]	2.003	[1.625 2.565]	2.118	[1.584 3.041]
T2	3.0	3.731	[2.875 5.425]	3.057	[2.260 4.473]	3.254	[2.275 5.421]
T3	4.0	6.551	[4.479 14.67]	4.445	[3.163 7.999]	4.237	[2.822 9.775]
Homogeneous Proportion $k_0^{+}/\left(k_0^{+} + k_1^{+}\mathbb{E}(\sigma_t^2)\right)$							
T1	0.5	0.558	[0.426 0.677]	0.474	[0.309 0.637]	0.474	[0.286 0.687]
T2	0.5	0.564	[0.441 0.698]	0.479	[0.316 0.642]	0.434	[0.222 1.386]
T3	0.5	0.572	[0.439 0.705]	0.479	[0.313 0.629]	0.397	[0.197 0.596]
Tail Precision $\widehat{\overline{\nu}_{\psi}^{+}(2.0)}/\overline{\nu}_{\psi}^{+}(2.0)$							
T1	1.0	0.913	[0.674 1.160]	0.927	[0.692 1.184]	0.919	[0.681 1.177]
T2	1.0	0.791	[0.349 1.230]	0.907	[0.426 1.462]	0.821	[0.395 1.384]
T3	1.0	0.351	[0.056 1.059]	0.682	[0.174 1.742]	0.702	[0.139 1.899]

Note: The table reports the median tail estimates based on the estimating equations in Section 4 and the corresponding interquartile range (IQR) obtained across a total of 1,000 replications for each of the different models defined in Table 1.

its ability to accurately assess the tail decay, the time-varying proportion of the tails, and the extreme tail behavior. For each of the relevant statistics, we report in Table 2 the median values obtained across all of the simulations, together with the interquartile range (IQR).

The true tail decay for all of the three models is determined by the value of β_0. Our non-parametric estimate for the tail decay is given by the inverse of $\widehat{\xi}^{+}$. The results reported in the first panel of the table show that the new estimation procedure generally permits fairly accurate estimation of the tail decay. The choice of truncation level does matter, however. On the one hand, choosing a low truncation level, results in the use of more observations, and hence everything else equal, reduces the sampling error. On the other hand, choosing too low a truncation level increases the deviation from the power-law decay and the error associated with the presence of the slowly varying function $L^{+}(x)$ in assumption A2. Too low a truncation level also renders the impact of the discretization error, and the ability to separate jumps from continuous moves, relatively more important.

Turning to the second panel, we report the estimates for $k_0^{+}/\left(k_0^{+} + k_1^{+}\mathbb{E}(\sigma_t^2)\right)$. This ratio in effect summarizes the estimation procedure's ability to disentangle the time-varying from the time-homogenous parts of the jump tails. The results indicate the same tradeoff in terms of the choice of truncation level: the use of lower truncation levels reduces sampling error, but at the same time increases the impact of the discretization error. This also helps explain the slight downward and upward biases observed in the first and third columns, respectively, and point to the middle truncation level of 0.015 as the preferred choice.

A distinct advantage of the new estimation procedure is that it allows us to meaningfully extrapolate the behavior of the jump tails to "extreme" levels for which inference based on historical sample averages is bound to be unreliable. To illustrate this important point, the third panel in the table reports the estimates for the jump tail intensities for jump sizes in excess of 2%, a very "large" value in typical financial data. To allow for a direct comparison across the different models, we report the estimates relative to their true values; i.e., $\widehat{\overline{\nu}_{\psi}^{+}(2.0)}/\overline{\nu}_{\psi}^{+}(2.0)$. Further, corroborating the accuracy of the underlying approximations, most of the estimated ratios are indeed quite close to unity. Of course, the same bias-variance type tradeoff as before pertains to the choice of truncation level, again pointing to the middle value as the most reliable.

All-in-all, the simulation results clearly indicate that the new estimation procedure works well, and that it gives rise to reasonable accurate estimates of the jump tail features of interest in practical applications. To further illustrate the applicability, we turn next to an empirical application involving actual high-frequency data for the S&P 500 aggregate market portfolio.

6 S&P 500 Jump Tails

Our estimates for the aggregate market jump tails are based on high-frequency intraday data for the S&P 500 futures contract spanning the period from January 1, 1990 to December 31, 2008. The theory underlying the new estimator builds on the idea of increasingly finer sampled observations over fixed time intervals, or $\Delta_n \to 0$. In practice, of course, market microstructure frictions prevent us from sampling too finely, while at the same time maintaining the basic Itô semimartingale assumption in equation (2.1); see, e.g., the discussion in Andersen et al. (2001), Zhang et al. (2005), and Barndorff-Nielsen et al. (2008). In lieu of this tradeoff, we choose to sample the prices at a one-minute frequency, resulting in a total of 400 observations per day for each of the $4,750$ trading days in the sample.[23]

Turning to the results, Table 3 reports the parameter estimates based on the assumption of affine in σ_t^2 time-varying jump intensities, without otherwise restricting the volatility dynamics, following the practical implementation strategy in Corollary 2.[24] The validity of the underlying

[23]The one-minute returns are approximately serially uncorrelated, with first and second order autocorrelation coefficients equal to -0.0016 and 0.0015, respectively. We also experimented with the use of coarser five- and ten-minutes sampling, resulting in very similar, albeit somewhat less precise, estimates for the tail decay parameters to the ones for the one-minute returns discussed below; see also Figure 2 in the introduction.

[24]Guided by the simulation results in the previous section, we set the truncation level at $\overline{\nu}_{\psi}(tr_T) = 0.03$, or

Table 3: S&P Jump Tail Estimates

Parameter	Estimate	St.dev.	Parameter	Estimate	St.dev.
	Left Tail			Right Tail	
ξ^-	0.2664	0.1153	ξ^+	0.2059	0.1301
σ_T^-	0.2566	0.0536	σ_T^+	0.2435	0.0487
$k_0^- \overline{\nu}_\psi(tr_T)$	-0.0004	0.0057	$k_0^+ \overline{\nu}_\psi(tr_T)$	0.0023	0.0052
$k_1^- \overline{\nu}_\psi(tr_T)$	0.0161	0.0065	$k_1^+ \overline{\nu}_\psi(tr_T)$	0.0129	0.0057
J-test	4.1606		J-test	1.8256	

Note: The table reports the estimates for the jump tail parameters based on one-minute S&P 500 futures prices from January 1, 1990 to December 31, 2008. The estimates rely on the moment conditions in Corollary 1 and the practical implementation thereof in Corollary 2. The truncation level is set at $tr_T = 0.5124$, corresponding to $\overline{\nu}_\psi^\pm(tr_T) = 0.015$ for each of the tails. The J-test involves two over-identifying restrictions.

modeling assumptions is corroborated by the J-tests for the two over-identifying moment restrictions reported in the last row of the table. Consistent with the idea of a power law decay, the estimates for $\xi^\pm$ are both statistically different from zero. Maybe somewhat surprisingly, the pairwise estimates for the left and right tail parameters are generally fairly close, implying that the tails are approximately symmetric. Importantly, the results also point to the existence of strong dynamic tail dependencies. Indeed, it appears that the tail jump intensities are almost exclusively determined by the time-varying parts of ν_t.

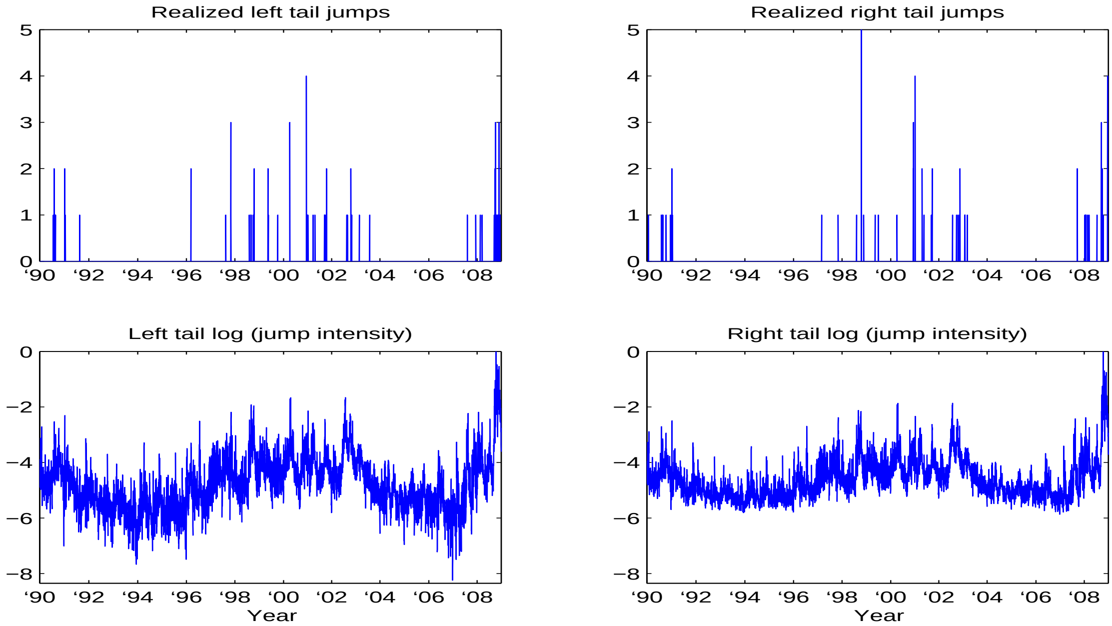
In order to more clearly illustrate these dynamic dependencies, we plot in Figure 3 the actual in-sample "large" jump realizations, together with the estimated jump tail intensities; i.e., $\overline{\nu}_t^\pm(x)$.[25] It is evident that the "large" jumps tend to cluster in time, with most of the realizations during the early 1990-91 part of the sample, the 1999-2002 time period associated with the Russian default, LTCM debacle, and the burst of the "tech bubble", as well as the recent 2008 financial crises. These tendencies for the jumps to cluster in time is also directly manifest in the estimated jump intensities depicted in the two lower panels in the figure. Reported on a relative logarithmic scale, the estimates imply large variations in the jump intensities, with tenfold changes within a few years not at all uncommon.

Rather than focussing on the estimated jump intensities, from a risk management perspective

approximately 0.015 for each of the tails, corresponding to $tr_T = 0.5124$ in percentage terms. Similarly, we set α equal to $4 \times \sqrt{BV_t \wedge RV_t}$ and $\varpi = 0.49$ in our calculation of TV_t^n, with the "large" jumps based on $\psi^{-1}(tr_T) \wedge \alpha\Delta_n^\varpi$. In addition, we adjust for the well-known diurnal pattern, by scaling α with an estimate of the time-of-day volatility as in Bollerslev and Todorov (2009).

[25]The estimate for the spot variance used in the calculation of the jump intensities depicted in the figure is based on the summation of the previous 200 truncated from above by $\alpha\Delta_n^\varpi$ squared one-minute returns; Jacod and Todorov (2010) provide a formal justification for this estimator.

Figure 3: Tail Jump Events



Note: The two top panels show the daily realized "large" jumps in the one-minute S&P 500 futures prices from January 1, 1990 to December 31, 2008, based on a truncation level of $tr_T = 0.5124$, or $\overline{\nu}_\psi(tr_T) = 0.03$. The two bottom panels show the estimated logarithmic time-varying jump tail intensities $\overline{\nu}_t^{\pm}(x)$.

it is often more informative to consider the likely size of a jump. In particular, keeping the jump intensity constant, the intrinsic time-dependence in the jump sizes may be formally revealed through,

$$
\begin{aligned}
q_{t,\alpha}^{-} &= \sup\{x < 0 : \overline{\nu}_t^{-}(x) \le \alpha\}, \qquad \overline{\nu}_t^{-}(x) = \int_{-\infty}^{x} \nu_t(z)dz, \\
q_{t,\alpha}^{+} &= \inf\{x > 0 : \overline{\nu}_t^{+}(x) \le \alpha\}, \qquad \overline{\nu}_t^{+}(x) = \int_{x}^{\infty} \nu_t(z)dz,
\end{aligned}
\tag{6.3}
$$

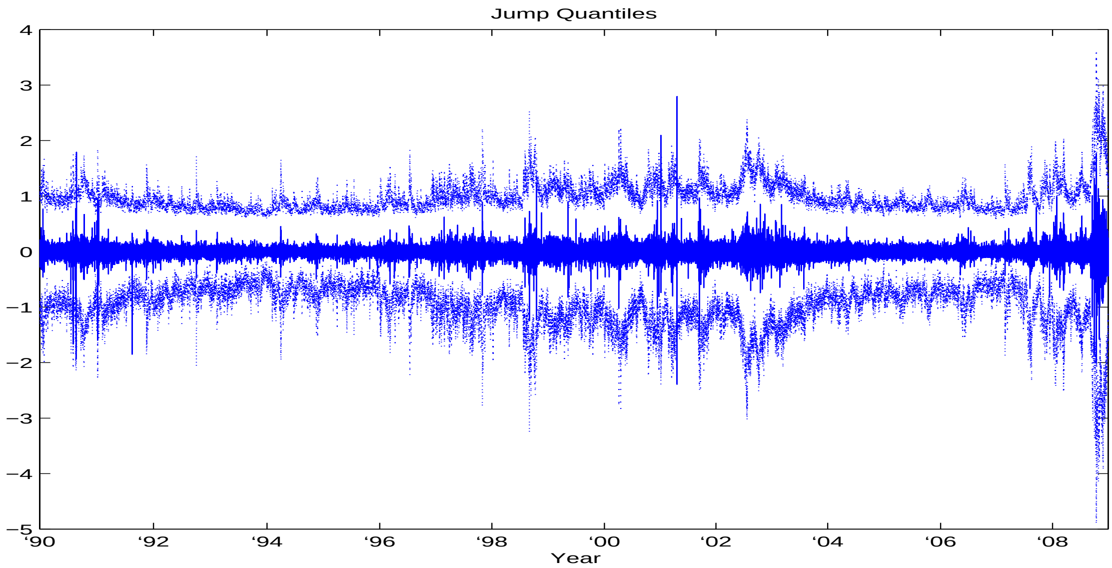
which define the time-varying jump sizes corresponding to the time t jump intensity of $\alpha > 0$ for negative and positive jumps exceeding those values. The $q_{t,\alpha}^{\pm}$ may also be interpreted as the inverse of the maps $x \to \overline{\nu}_t^{\pm}(x)$, and we will refer to them correspondingly as the "jump quantiles." Such quantities would generally be very difficult to accurately estimate empirically. However, the key approximation in (3.1), together with the assumption of affine jump intensities underlying our jump tail estimation, permits us to readily evaluate the jump quantiles in a non-parametric fashion. Specifically, for the right tail we have the following approximation,

$$
\widehat{q}_{t,\alpha}^{+} = \psi^{-1}\left\{ tr_T + \left[\left(\frac{k_0^{+}\widehat{\overline{\nu}_\psi^{+}(tr_T)} + k_1^{+}\widehat{\overline{\nu}_\psi^{+}(tr_T)}\widehat{\sigma}_t^2}{\alpha} \right)^{\widehat{\xi}^{+}} - 1 \right] \frac{\widehat{\sigma}^{+}}{\widehat{\xi}^{+}} \right\}, \tag{6.4}
$$

where $\widehat{\sigma}_t^2$ denotes a consistent estimator for the spot volatility, as discussed above, and the left tail estimator $\widehat{q_{t,\alpha}^-}$ may be defined analogously.[26]

Figure 4 shows the resulting estimated jump sizes corresponding to one positive, respectively negative, jump larger, respectively smaller, than that value every two calendar year; i.e., one jump of that *absolute size* per calendar year. The estimates again reveal surprisingly close to symmetric tail behavior, albeit slightly larger variations in the negative jump quantiles due to the slightly larger estimated value for $k_1^- \overline{\nu}_\psi(tr_T)$. The figure also shows that the size of the "large" jumps vary quite dramatically over time, with jumps in excess of one percent highly unlikely for most of the sample, while such jumps are fairly common during the recent financial crises.

Figure 4: Jump Quantiles



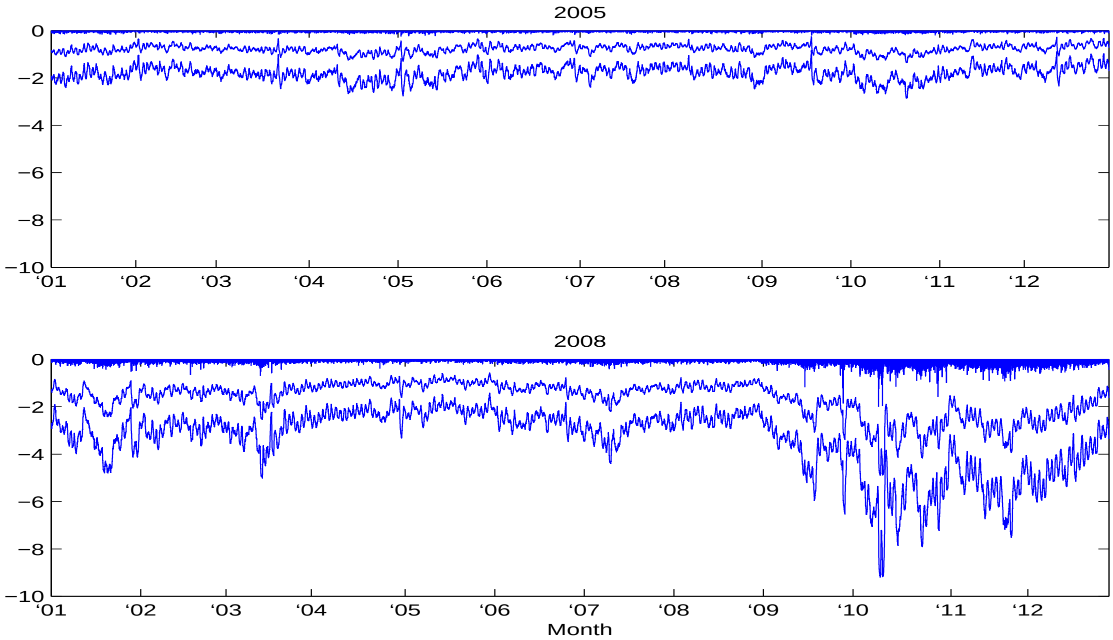
Note: The figure shows the one-minute S&P 500 futures returns from January 1, 1990 to December 31, 2008, together with the estimated jump sizes corresponding to a jump intensity of one positive, respectively negative, jump every two calendar years, as formally defined by the "jump quantiles" $q_{t,\alpha}^{\pm}$.

To further highlight these important dependencies, we plot in Figure 5 the estimated left jump quantiles for 2005, a relatively quiet year, together with the quantiles for 2008. In addition to the two-year quantiles shown in the previous figure, we also include the extreme jump sizes corresponding to a negative jump every twenty years; i.e., once in the sample. These latter extreme quantiles would be impossible to meaningfully estimate by extrapolating from standard parametric procedures and coarser frequency, e.g., daily data. Looking at the figure, 2005 was obviously an "easy" year from a risk management perspective. The two and twenty year jump quantiles are both approximately constant, and hover around less than negative one and two

[26] To formally justify these estimators for $q_{t,\alpha}^{\pm}$ we need $\alpha_T \propto tr_T$.

percent, respectively. In sharp contrast, the jump quantiles for 2008 vary quite dramatically throughout the year, reaching their peak in October in the aftermath of the Lehman bankruptcy and the government TARP bailout program, gradually stabilizing towards the end of the year.

Figure 5: Left Tail Jump Quantiles



Note: The figure shows the negative one-minute S&P 500 futures returns for 2005 (top panel) and 2008 (bottom panel), together with the estimated left tail "jump quantiles" corresponding to a jump intensity of one negative jump every two calendar years and one negative jump every twentieth calendar years, respectively, as formally defined by $q^{\pm}_{t,\alpha}$.

7 Conclusion

The availability of high-frequency intraday asset prices has spurred a large and rapidly growing literature. This paper further expands on our ability to extract useful information about important economic phenomena from this new rich source of data through the development of a flexible non-parametric estimation procedure for the jump tails. The method allows for very general dynamic dependencies in the tails and imposes essentially no restrictions on the continuous part of the price process. The basic idea is based on the assumption of regular variation in the jump tails, and how that assumption translates into certain functionals of the "large" jumps being approximate martingales. We confirm the reliability of the new estimation procedure through a series of Monte Carlo simulation experiments, and illustrate its applicability with actual high-frequency data for the S&P 500 market portfolio.

Looking ahead, the new estimation framework should be of use in many situations of practical import. In particular, the most important and difficult to manage financial market risks are invariably associated with tail events. Hence, the ability to more accurately measure and possibly forecast the jump tails, holds the promise of improved risk management techniques better geared toward controlling large risks, leaving aside the smaller approximately "continuous" price moves. By enhancing our understanding of the type of economic "news" that induce large price moves, or tail events, empirical implementations of the new estimation procedure could also help shed new light on the fundamental linkages between asset markets and the real economy.

The lack of investor confidence and fear of tail events are often singled out as one of the main culprits behind the massive losses in market values in the advent of the Fall 2008 financial crises, and the idea that rare disasters may help explain apparent mis-pricing has spurred a rapidly growing recent literature. The arguments put forth in that literature often hinge on probabilities of severe events that exceed those materialized in sample, or probabilities calibrated to reflect a much broader set of assets and/or countries; e.g., Barro (2006) and Gabaix (2010). Instead, as discussed in Bollerslev and Todorov (2009), the new econometric procedures developed here hold the promise of reliably estimating the likely occurrence of tail events based on actually observed high-frequency data, without having to resort to "peso" type explanations or having to impose tight restrictions on the volatility dynamics.

8 Proofs

8.1 Proofs of Theorem 1 and Corollary 1

Follow from the proof of Theorem 2 below. $\square$

8.2 Proof of Theorem 2

For notational convenience, denote

$$\tilde{g}(\theta, z_t, tr_T) = g(\theta, z_t, tr_T) \bullet \begin{pmatrix} \theta^{(2)} \\ 1 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} \widetilde{\phi}_1^+(u,\xi,\sigma) \\ \widetilde{\phi}_2^+(u,\xi,\sigma) \end{pmatrix} = \begin{pmatrix} \phi_1^+(u,\xi,\sigma)\sigma \\ \phi_2^+(u,\xi,\sigma) \end{pmatrix},$$

$$G(\theta, z_t, tr_T) = \left(G^{(ij)}(\theta, z_t, tr_T)\right)_{i=1,\dots,2q,\ j=1,2}, \quad G^{(ij)}(\theta, z_t, tr_T) = \frac{\partial \tilde{g}^{(i)}}{\partial \theta^{(j)}}(\theta, z_t, tr_T),$$

with $\tilde{g}_T(\theta, tr_T)$ and $\tilde{g}_T(\theta, tr_T)$ defined similar to $g_T(\theta, tr_T)$ from $\tilde{g}(\theta, z_t, tr_T)$. Further, set

$$H_i(\theta, z_t, tr_T) = \left(H_i^{(kl)}(\theta, z_t, tr_T)\right)_{k,l=1,2}, \quad H_i^{(kl)}(\theta, z_t, tr_T) = \frac{\partial \tilde{g}^{(i)}}{\partial \theta^{(k)} \partial \theta^{(k)}}(\theta, z_t, tr_T).$$

We start by showing some preliminary results, which we will make use of later in the proof. First, by a change of variable it follows that for any function $\phi(u)$,

$$\begin{aligned}\int_{\mathbb{R}} \phi(\psi^+(x) - tr_T)1_{\{\psi^+(x)>tr_T\}}\nu(x)dx &= \int_0^\infty \phi(u)\nu_\psi^+(tr_T+u)du \\ &= \overline{\nu}_\psi^+(tr_T)\int_0^\infty \phi(u)\left(1-\frac{\overline{\nu}_\psi^+(u+tr_T)}{\overline{\nu}_\psi^+(tr_T)}\right)' du.\end{aligned} \tag{8.1}$$

Next, using assumption A2 on the slowly varying function $L^+(x)$, integration by parts, and using the results of Goldie and Smith (1987) for slowly varying functions with residuals (see also Smith (1987), Proposition 3.1), we have for some $\beta > 0$ and $r > 0$,s

$$\int_0^\infty \left(1+\beta\frac{u}{tr_T}\right)^{-r}\left(1-\frac{\overline{\nu}_\psi^+(u+tr_T)}{\overline{\nu}_\psi^+(tr_T)}\right)' du = \kappa(\beta,r,\alpha^+) + K\tau^+(\beta tr_T) + o(\tau^+(\beta tr_T)), \tag{8.2}$$

where K denotes some constant, and the function κ is continuous in its first argument with $\kappa(1,r,\alpha^+) = \frac{\alpha^+}{\alpha^++r}$. Similarly, for $\beta > 0$ and an integer s,

$$\int_0^\infty \left(-\ln\left(1+\beta\frac{u}{tr_T}\right)\right)^s\left(1-\frac{\overline{\nu}_\psi^+(u+tr_T)}{\overline{\nu}_\psi^+(tr_T)}\right)' du = \widetilde{\kappa}(\beta,s,\alpha^+) + K\tau^+(\beta tr_T) + o(\tau^+(\beta tr_T)), \tag{8.3}$$

where K denotes some constant (generally different from the constant in the previous equation), and the function $\widetilde{\kappa}$ is continuous in its first argument with $\widetilde{\kappa}(1,s,\alpha^+) = (-\alpha^+)^{-s}\Gamma(s+1)$.

The proof proceeds in two steps by first showing consistency and then asymptotic normality of the estimator.

Part 1. Consistency. From the definition of the random measure μ, it follows readily that

$$\frac{M_T^+}{T\overline{\nu}_\psi^+(tr_T)} \xrightarrow{\mathbb{P}} \mathbb{E}(\varphi_t^+). \tag{8.4}$$

Now, by a standard law of large numbers, for any fixed $\xi \in (0,\infty)$ and $\beta \in (0,\infty)$ with $\sigma = \xi tr_T/\beta$, we have

$$\begin{cases} \frac{1}{M_T^+}\int_0^T\int_{\psi^+(x)>tr_T}\frac{1}{1+\xi(\psi^+(x)-tr_T)/\sigma}\mu(ds,dx) \xrightarrow{\mathbb{P}} \kappa(\beta,1,\alpha^+), \\ \frac{1}{M_T^+}\int_0^T\int_{\psi^+(x)>tr_T}\log(1+\xi(\psi^+(x)-tr_T)/\sigma)\mu(ds,dx) \xrightarrow{\mathbb{P}} \widetilde{\kappa}(\beta,1,\alpha^+). \end{cases}$$

Moreover, since $\log(1+x)$ and $1/(1+x)$ are monotone in x, the above convergence can be trivially extended to uniform over the sets $\xi \in [0, K_\xi]$ and $\beta \in (K_\beta,\infty]$ for any $K_\xi > 0$, $K_\beta > 0$. Then, for a local neighborhood of the true parameter value we have

$$\sup_{\theta\in\Theta_T^l} ||\tilde{g}_T(\theta,tr_T) - g(\theta,tr_T)|| \xrightarrow{\mathbb{P}} 0,$$

where

$$g(\theta, tr_T) = \begin{pmatrix} \frac{1}{\xi} - \left(1 + \frac{1}{\xi}\right) \kappa\left(\beta, 1, \alpha^+\right) \\ -\frac{1}{\xi^2}\widetilde{\kappa}\left(\beta, 1, \alpha^+\right) - \frac{1}{\xi}\left(1 + \frac{1}{\xi}\right)\left(1 - \kappa\left(\beta, 1, \alpha^+\right)\right) \end{pmatrix}.$$

Thus, $g(\theta) = \mathbf{0}$ for $\theta = \theta_T^0$. Further, the derivative of $g(\theta, tr_T)$ with respect to the parameter θ, when evaluated at the true value is nonsingular (properly taking into account the fact that the true value of σ_T^+ grows with the time span by multiplying the two derivatives involving differentiation with respect to σ by tr_T). Therefore, $g(\theta) = \mathbf{0}$ is solved uniquely by $\theta = \theta_T^0$ in a local neighborhood.

Part 2. Asymptotic Normality. Let $\theta = \theta_T^0 + \tilde{\mathbf{r}}$, where

$$\tilde{\mathbf{r}} = \sqrt{T\overline{\nu}_\psi(tr_T)\mathbb{E}(\varphi_s^+)}\mathbf{r} \bullet \begin{pmatrix} \frac{tr_T}{\alpha^+ T\overline{\nu}_\psi^+(tr_T)} \\ \frac{1}{T\overline{\nu}_\psi^+(tr_T)} \end{pmatrix},$$

for some $\mathbf{r} \in \mathbb{R}^2$. Then, using a second-order Taylor expansion for this value of the vector θ we can write

$$\sqrt{M_T^+}\widetilde{g}_T(\theta, tr_T) = \sqrt{M_T^+}\widetilde{g}_T(\theta_T^0, tr_T) + \sqrt{M_T^+}G_T(\theta_T^0, tr_T)\tilde{\mathbf{r}} + R_T(\mathbf{r}),$$

$$R_T(\mathbf{r}) = \frac{1}{2}\frac{1}{\sqrt{M_T^+}}\sum_{t=1}^{T-1}\begin{pmatrix} \tilde{\mathbf{r}}' H_1(\tilde{\theta}, z_t, tr_T)\tilde{\mathbf{r}} \\ \vdots \\ \tilde{\mathbf{r}}' H_{2q}(\tilde{\theta}, z_t, tr_T)\tilde{\mathbf{r}} \end{pmatrix},$$

where $\tilde{\theta}$ denotes some value between θ and θ_T^0 and $G_T(\theta, tr_T)$, $H_1(\theta, z_t, tr_T)$,...,$H_{2q}(\theta, z_t, tr_T)$ denote the corresponding first and second derivatives.

Our goal will be to show that for $\theta = \theta_T^0 + \tilde{\mathbf{r}}$, $\sqrt{M_T^+}\widetilde{g}_T(\theta, tr_T)$ converges uniformly in $\mathbf{r}$ to $\mathbf{dZ} + \Pi\mathbf{r}$ for some non-random matrices $\mathbf{d}$ and Π (with Π of full column rank) and a standard normal vector $\mathbf{Z}$. By the continuity of the limit process (in $\mathbf{r}$), this establishes the result in (3.9). The proof proceeds in several steps.

Step 1. We prove $\frac{1}{\sqrt{M_T^+}}\sum_{t=1}^{T-1}\tilde{g}(\theta_T^0, z_t, tr_T) \xrightarrow{\mathcal{L}} \mathbf{dZ}$, where $\mathbf{d}$ is a matrix of constants and $\mathbf{Z}$ is the standard normal random vector of the theorem.

We first decompose the moment vector $\tilde{g}(\theta_T^0, z_t, tr_T)$ into two components. Let $\tilde{g}_1(\theta, z_t, tr_T)$ denote the vector

$$\mathbf{x}_t \otimes \begin{pmatrix} \int_t^{t+1}\int_{\psi^+(x)>tr_T}\widetilde{\phi}_1^+(\psi(x) - tr_T, \theta^{(1)}, \theta^{(2)})\tilde{\mu}(ds, dx) \\ \int_t^{t+1}\int_{\psi^+(x)>tr_T}\widetilde{\phi}_2^+(\psi(x) - tr_T, \theta^{(1)}, \theta^{(2)})\tilde{\mu}(ds, dx) \end{pmatrix},$$

and $\tilde{g}_2(\theta, z_t, tr_T)$ the vector

$$\mathbf{x}_t \otimes \begin{pmatrix} \int_t^{t+1}\int_{\psi^+(x)>tr_T}\widetilde{\phi}_1^+(\psi(x) - tr_T, \theta^{(1)}, \theta^{(2)})\nu(ds, dx) \\ \int_t^{t+1}\int_{\psi^+(x)>tr_T}\widetilde{\phi}_2^+(\psi(x) - tr_T, \theta^{(1)}, \theta^{(2)})\nu(ds, dx) \end{pmatrix},$$

so that by definition $\tilde{g}(\theta_T^0, z_t, tr_T) = \tilde{g}_1(\theta_T^0, z_t, tr_T) + \tilde{g}_2(\theta_T^0, z_t, tr_T)$.

We first prove that $\frac{1}{\sqrt{M_T^+}} \sum_{t=1}^{T-1} \tilde{g}_2(\theta_T^0, z_t, tr_T) \xrightarrow{\mathbb{P}} 0$. Using the fact that for the true parameter value, θ_T^0, we have $\frac{\theta^{(2)}}{\theta^{(1)}} = tr_T$, together with the definition of the score functions ϕ_1^+ and ϕ_2^+ in (3.2), and the results in (8.1)-(8.3), the two elements of $\widetilde{g}^{(2)}(\theta_T^0, z_t, tr_T)$ may be expressed as

$$C\overline{\nu}_\psi^+(tr_T)\left(\tau^+(tr_T) + o(\tau^+(tr_T))\right)\int_t^{t+1} \varphi_s^+ ds,$$

for some constant C which differ for each of the two elements. Then, since $\sqrt{T\overline{\nu}_\psi^+(tr_T)}\tau^+(tr_T) \rightarrow 0$ and assumption A3 implies that the process φ_t^+ is stationary and integrable, it follows that

$$\frac{1}{\sqrt{T\overline{\nu}_\psi^+(tr_T)}}\sum_{t=1}^{T-1} \mathbb{E}||\widetilde{g}_2(\theta_T^0, z_t, tr_T)|| \xrightarrow{\mathbb{P}} 0.$$

Then, combining this result with (8.4), we get $\frac{1}{\sqrt{M_T^+}} \sum_{t=1}^{T-1} \tilde{g}_2(\theta_T^0, z_t, tr_T) \xrightarrow{\mathbb{P}} 0$.

We are left with showing that $\frac{1}{\sqrt{M_T^+}} \sum_{t=1}^{T-1} \tilde{g}_1(\theta_T^0, z_t, tr_T) \xrightarrow{\mathcal{L}} \mathbf{dZ}$. This convergence will follow from a Central Limit Theorem for a triangular array; see, e.g., Jacod and Shiryaev (2003), Theorem VIII.2.29. It suffices to prove that

$$\begin{cases} \frac{1}{M_T^+} \sum_{t=1}^{T-1} \mathbb{E}_t\tilde{g}_1(\theta_T^0, z_t, tr_T) \xrightarrow{\mathbb{P}} 0, \\ \frac{1}{M_T^+} \sum_{t=1}^{T-1} \mathbb{E}_t\tilde{g}_1(\theta_T^0, z_t, tr_T)\tilde{g}_1(\theta_T^0, z_t, tr_T)' \xrightarrow{\mathbb{P}} \mathbf{dd}', \\ \frac{1}{(M_T^+)^{1+\alpha/2}} \sum_{t=1}^{T-1} \mathbb{E}_t||\tilde{g}_1^{(i)}(\theta_T^0, z_t, tr_T)||^{2+\alpha} \xrightarrow{\mathbb{P}} 0, \quad \text{for some } \alpha > 0 \text{ and } i = 1, 2. \end{cases} \tag{8.5}$$

The first condition in (8.5) is trivially satisfied, as $\{\tilde{g}_1(\theta_T^0, z_t, tr_T)\}_{t=1,2,...}$ is a martingale difference sequence. To show the second convergence in (8.5), note that for $i, j = 1, ..., 2q$,

$$\begin{aligned} &\mathbb{E}_t\left(\tilde{g}_1^{(i)}(\theta_T^0, z_t, tr_T)\tilde{g}_1^{(j)}(\theta_T^0, z_t, tr_T)\right) \\ &= x_t^{(i-\lfloor(i-1)/q\rfloor q)} x_t^{(j-\lfloor(j-1)/q\rfloor q)}\mathbb{E}_t\left[\int_t^{t+1}\int_{\psi^+(x)>tr_T} \zeta_1(x)\tilde{\mu}(ds, dx)\int_t^{t+1}\int_{\psi^+(x)>tr_T} \zeta_2(x)\tilde{\mu}(ds, dx)\right] \\ &= x_t^{(i-\lfloor(i-1)/q\rfloor q)} x_t^{(j-\lfloor(j-1)/q\rfloor q)}\int_{\psi^+(x)>tr_T} \zeta_1(x)\zeta_2(x)\nu(x)dx\mathbb{E}_t\int_t^{t+1} \varphi_s^+ ds, \end{aligned}$$

where $\zeta_1(x)$ and $\zeta_2(x)$ are one of the functions that appear as integrands of $\widetilde{\mu}$ in the definition of $\tilde{g}_1(\theta_T^0, z_t, tr_T)$, and the second equality follows from Itô 's lemma. Now, using the results in (8.1)-(8.3), we can write

$$\int_{\psi^+(x)>tr_T} \zeta_1(x)\zeta_2(x)\nu(x)dx = \overline{\nu}_\psi^+(tr_T)\left(K + \tau^+(tr_T) + o(\tau^+(tr_T))\right),$$

where the constant K depend on the true parameter vector θ_T^0, but do not depend on T. Also, by assumption A3 the process φ_t^+ is stationary and integrable and by the assumption of the theorem we get

$$\frac{1}{T}\sum_{t=1}^{T-1}\mathbf{x}_t\mathbf{x}_t'\mathbb{E}_t\left(\int_t^{t+1}\varphi_s^+ ds\right) \overset{\mathbb{P}}{\to} \mathbb{E}\left(\mathbf{x}_t\mathbf{x}_t'\int_t^{t+1}\varphi_s^+ ds\right),$$

Then in parallel to Step 1, we can use $M_T^+ \sim T\overline{\nu}_\psi^+(tr_T)$ to show the second part of (8.5).

To prove the third part of (8.5), let $\alpha \leq 2$ such that $\mathbb{E}|\varphi_s^+|^{1+\alpha} < \infty$. The existence of α is guaranteed by assumption A3. Then, using the Burkholder-Davis-Gundy inequality,

$$\mathbb{E}_t\left[\int_t^{t+1}\int_{\psi^+(x)>tr_T}\zeta(x)\widetilde{\mu}(ds,dx)\right]^{2+\alpha} \leq \mathbb{E}_t\left(\int_t^{t+1}\int_{\psi^+(x)>tr_T}\zeta^2(x)\mu(ds,dx)\right)^{1+\alpha/2},$$

where ζ is one of the functions that appear as integrands of $\widetilde{\mu}$ in the definition of $\tilde{g}_1(\theta_T^0, z_t, tr_T)$. Further,

$$\begin{aligned}\mathbb{E}_t\left(\int_t^{t+1}\int_{\psi^+(x)>tr_T}\zeta^2(x)\mu(ds,dx)\right)^{1+\alpha/2} &\leq K\mathbb{E}_t\left|\int_t^{t+1}\int_{\psi^+(x)>tr_T}\zeta^2(x)\widetilde{\mu}(ds,dx)\right|^{1+\alpha/2} \\ &+ K\left(\int_{\psi^+(x)>tr_T}\zeta^2(x)\nu(x)dx\right)^{1+\alpha/2}\mathbb{E}_t\left(\int_t^{t+1}\varphi_s^+ ds\right)^{1+\alpha/2},\end{aligned}$$

for $K > 0$ some constant. For the first term on the right hand side of the above inequality, applying the inequality $(\sum_i |a_i|)^p \leq \sum_i |a_i|^p$ for $0 < p \leq 1$, and the fact that $\alpha \leq 2$, together with the definition of the jump compensator, we have

$$\mathbb{E}_t\left|\int_t^{t+1}\int_{\psi^+(x)>tr_T}\zeta^2(x)\widetilde{\mu}(ds,dx)\right|^{1+\alpha/2} \leq K\int_t^{t+1}\int_{\psi^+(x)>tr_T}\zeta^{2+\alpha}(x)\nu(x)dx\mathbb{E}_t\left(\int_t^{t+1}\varphi_s^+ ds\right).$$

The third result of (8.5) now follows directly. full column-rank matrix of constants depending on the true parameter vector θ_T^0. We denote with $G_1(\theta, z_t, tr_T)$ the $2q \times 2$ matrix with the following elements for $i = 1, ..., 2q$,

$$G_1^{(ij)}(\theta, z_t, tr_T) = x_t^{(i-\lfloor(i-1)/q\rfloor q)}\begin{cases}\frac{tr_T}{\alpha^+}\int_t^{t+1}\int_{\psi^+(x)>tr_T}\frac{\partial\widetilde{\phi}_{i_q}^+}{\partial\theta^{(1)}}(\psi(x)-tr_T,\theta^{(1)},\theta^{(2)})\widetilde{\mu}(ds,dx), & j=1,\\ \int_t^{t+1}\int_{\psi^+(x)>tr_T}\frac{\partial\widetilde{\phi}_{i_q}^+}{\partial\theta^{(2)}}(\psi(x)-tr_T,\theta^{(1)},\theta^{(2)})\widetilde{\mu}(ds,dx), & j=2,\end{cases}$$

where $i_q = 1$ for $i = 1, ..., q$ and $i_q = 2$ for $i = q+1, ..., 2q$. As in the previous step, it is possible to show

$$\frac{1}{\sqrt{M_T^+}}\sum_{t=1}^{T-1}G_1(\theta_T^0, z_t, tr_T) \overset{\mathcal{L}}{\to} \mathbf{bZ},$$

where $\mathbf{b}$ is some vector of constants, and $\mathbf{Z}$ is a standard normal vector, and therefore

$$\frac{1}{\sqrt{M_T^+}}\sum_{t=1}^{T-1}\tilde{G}_1(\theta_T^0, z_t, tr_T) \overset{\mathbb{P}}{\to} 0,$$

for $\tilde{G}_1(\theta_T^0, z_t, tr_T) = \frac{1}{\sqrt{T\overline{\nu}_\psi^+(tr_T)\mathbb{E}(\varphi_s^+)}} G_1(\theta_T^0, z_t, tr_T)$.

Next, define the $G_2(\theta, z_t, tr_T)$ matrix such that $G_2(\theta, z_t, tr_T)\mathbf{r} = G(\theta, z_t, tr_T)\tilde{\mathbf{r}} - \tilde{G}_1(\theta, z_t, tr_T)\mathbf{r}$ for every $\mathbf{r}$. We are then left with proving $\frac{1}{\sqrt{M_T^+}} \sum_{t=1}^{T-1} G_2(\theta_T^0, z_t, tr_T) \xrightarrow{\mathbb{P}} \Pi$. Using the results in (8.1)-(8.3) it is possible to show that the above matrix sequence is equal to

$$\mathbf{K}_T \otimes \left(\frac{1}{T} \sum_{t=1}^{T-1} \mathbf{x}_t \int_t^{t+1} \varphi_s^+ ds \right),$$

where $\mathbf{K}_T$ is a 2×2 matrix with elements of the form $\mathbf{K}_T^{(ij)} = K^{(ij)} + \tau^+(tr_T) + o(\tau^+(tr_T))$ for $i, j = 1, 2$ with the constants $K^{(ij)}$ depending on the value of the true parameter vector θ_T^0 and such that the matrix $\{K^{(ij)}\}_{i,j=1,2}$ is non-singular. Since $\sqrt{T\overline{\nu}_\psi^+(tr_T)}\tau^+(tr_T) \to 0$, using the assumption in the theorem for $\frac{1}{T}\sum_{t=1}^{T-1} x_t^{(i)} \int_t^{t+1} \varphi_s^+ ds$, it follows that it converges in probability. Then note that Π is of full column rank because of our assumption $\mathbb{E}\left(\mathbf{x}_t \int_t^{t+1} \varphi_s^+ ds\right) \neq \mathbf{0}$ and this proves the claim of this step.

Step 3. To complete the proof, we show that $\sup_{\mathbf{r}:\theta\in\Theta_T^l} ||R_T(\mathbf{r})|| \xrightarrow{\mathbb{P}} 0$. Using the condition $\sqrt{T\overline{\nu}_\psi^+(tr_T)}\tau^+(tr_T) \to 0$, the elements of the matrix $R_T(\mathbf{r})$ may be expressed as,

$$K_T \times Z_T(\theta),$$

where $K_T \xrightarrow{\mathbb{P}} 0$, which does not depend on θ, and

$$Z_T(\theta) = \frac{1}{M_T^+} \sum_{t=1}^{T-1} x_t^{(i)} \int_t^{t+1} \int_{\psi^+(x)>tr_T} \phi(\psi^+(x) - tr_T, \theta)\mu(ds, dx),$$

for $i = 1, ..., q$, and the function $\phi(\cdot, \theta)$ involves second derivatives of ϕ_1 and ϕ_2 (with respect to the parameter). We may therefore bound $Z_T(\theta) \leq N_T$ on the set Θ_T^l, where $N_T \xrightarrow{\mathbb{P}} K$ for some constant K. $\square$

8.3 Proofs of Theorem 3 and Corollary 2

We start by establishing several preliminary lemmas. In what follows we use the notation $\mathbb{E}_i^n$ for $\mathbb{E}(\cdot|\mathcal{F}_{i\Delta_n})$ and $\mathbb{P}_i^n$ for $\mathbb{P}(\cdot|\mathcal{F}_{i\Delta_n})$. Also, we will use C to denote a positive constant that might change from line to line.

Lemma 1 *Suppose we observe the process p_t at the discrete times $0, \Delta_n, 2\Delta_n, ..., [T/\Delta_n]$, and assume that A1 and A4 hold. Then for some $\alpha > 0$ and $\varpi \in \left(0, \frac{1}{2}\right)$, we have*

$$\frac{\sqrt{N_T}}{T} \left(\sum_{i=1}^{[T/\Delta_n]} (\Delta_i^n p)^2 1_{\{|\Delta_i^n p| \leq \alpha\Delta_n^\varpi\}} - \int_0^T \sigma_s^2 ds \right) \xrightarrow{\mathbb{P}} 0, \quad \text{as } \Delta_n \downarrow 0,\ T \uparrow \infty, \tag{8.6}$$

where $\Delta_i^n p = p_{i\Delta_n} - p_{(i-1)\Delta_n}$, and N_T is some deterministic sequence of T increasing to infinity with the property that $\sqrt{N_T}\Delta_n^{(2-\beta)\varpi} \to 0$.

Proof: We have

$$\left((\Delta_i^n p)^2 1_{\{|\Delta_i^n p|\le \alpha\Delta_n^{\varpi}\}} - \int_{(i-1)\Delta_n}^{i\Delta_n} \sigma_s^2 ds\right) = a_i^1 - a_i^2 + a_i^3 + a_i^4 + a_i^5 - a_i^6 + a_i^7,$$

$$a_i^1 = \left((\Delta_i^n Z)^2 - \int_{(i-1)\Delta_n}^{i\Delta_n} \sigma_s^2 ds\right), \quad a_i^2 = (\Delta_i^n Z)^2 1_{\{|\Delta_i^n p|>\alpha\Delta_n^{\varpi}\}},$$

$$a_i^3 = (\Delta_i^n Y)^2 1_{\{|\Delta_i^n p|\le\alpha\Delta_n^{\varpi}\}}, \quad a_i^4 = 2\Delta_i^n \hat{Z}\Delta_i^n \widetilde{Y},$$

$$a_i^5 = 2\Delta_i^n \hat{Z}\Delta_i^n \widehat{Y}, \quad a_i^6 = 2\Delta_i^n \tilde{Z}\Delta_i^n Y 1_{\{|\Delta_i^n p|>\alpha\Delta_n^{\varpi}\}}, \quad a_i^7 = 2\Delta_i^n \tilde{Z}\Delta_i^n Y 1_{\{|\Delta_i^n p|\le\alpha\Delta_n^{\varpi}\}},$$

$$Z_t = \int_0^t \sigma_s dW_s + \int_0^t \tilde{\alpha}_s ds, \quad \tilde{\alpha}_t = \begin{cases} \int_0^t \alpha_s ds, & \text{if } \beta \ge 1, \\ \int_0^t \alpha_s ds - \int_0^t \int_{\mathbb{R}} \kappa(x) ds \nu_s(dx), & \text{if } \beta < 1, \end{cases}$$

$$\tilde{Z}_i^n = \int_{(i-1)\Delta_n}^{i\Delta_n} (\sigma_s - \sigma_{(i-1)\Delta_n}) dW_s + \int_{(i-1)\Delta_n}^{i\Delta_n} \tilde{\alpha}_s ds, \quad \hat{Z}_i^n = \sigma_{(i-1)\Delta_n}\Delta_i^n W,$$

$$Y_t = \begin{cases} \int_0^t \int_{\mathbb{R}} \kappa(x)\tilde{\mu}(ds,dx) + \int_0^t \int_{\mathbb{R}} \kappa'(x)\mu(ds,dx), & \text{if } \beta \ge 1, \\ \int_0^t \int_{\mathbb{R}} x\mu(ds,dx), & \text{if } \beta < 1, \end{cases}$$

$$\widehat{Y}_t = \begin{cases} \int_0^t \int_{\mathbb{R}} \kappa'(x)\nu(ds,dx), & \text{if } \beta \ge 1, \\ \int_0^t \int_{\mathbb{R}} x\nu(ds,dx), & \text{if } \beta < 1, \end{cases}, \qquad \tilde{Y}_t = Y_t - \hat{Y}_t.$$

We will need an equivalent (in distribution) decomposition of the process Y. We define it on an independent copy of the original probability space with the only difference being that the jump measure is defined from a homogeneous Poisson measure via thinning, i.e.,

$$Y_t^{(1)} = \begin{cases} \int_0^t \int_{\mathbb{R}^+} \int_{\mathbb{R}} \kappa(x) (1_{\{x<0,\ u<\varphi^-_{(i-1)\Delta_n-}\}} + 1_{\{x>0,\ u<\varphi^+_{(i-1)\Delta_n-}\}}) \underline{\tilde{\mu}}(ds,du,dx) \\ + \int_0^t \int_{\mathbb{R}^+} \int_{\mathbb{R}} \kappa'(x) (1_{\{x<0,\ u<\varphi^-_{(i-1)\Delta_n-}\}} + 1_{\{x>0, u<\varphi^+_{(i-1)\Delta_n-}\}}) \underline{\mu}(ds,du,dx), & \text{if } \beta \ge 1, \\ \int_0^t \int_{\mathbb{R}^+} \int_{\mathbb{R}} x (1_{\{x<0,\ u<\varphi^-_{(i-1)\Delta_n-}\}} + 1_{\{x>0,\ u<\varphi^+_{(i-1)\Delta_n-}\}}) \underline{\mu}(ds,du,dx), & \text{if } \beta < 1, \end{cases}$$

$$Y_t^{(2)} = \begin{cases} \int_0^t \int_{\mathbb{R}^+} \int_{\mathbb{R}} \kappa(x) \left(1_{\{x<0,\ u<\varphi^-_{s-}\}} - 1_{\{x<0,\ u<\varphi^-_{(i-1)\Delta_n-}\}}\right) \underline{\tilde{\mu}}(ds,du,dx) \\ + \int_0^t \int_{\mathbb{R}^+} \int_{\mathbb{R}} \kappa(x) \left(1_{\{x>0,\ u<\varphi^+_{s-}\}} - 1_{\{x>0,\ u<\varphi^+_{(i-1)\Delta_n-}\}}\right) \underline{\tilde{\mu}}(ds,du,dx) \\ + \int_0^t \int_{\mathbb{R}^+} \int_{\mathbb{R}} \kappa'(x) \left(1_{\{x<0,\ u<\varphi^-_{s-}\}} - 1_{\{x<0,\ u<\varphi^-_{(i-1)\Delta_n-}\}}\right) \underline{\mu}(ds,du,dx), \\ + \int_0^t \int_{\mathbb{R}^+} \int_{\mathbb{R}} \kappa'(x) \left(1_{\{x>0,\ u<\varphi^+_{s-}\}} - 1_{\{x>0,\ u<\varphi^+_{(i-1)\Delta_n-}\}}\right) \underline{\mu}(ds,du,dx), & \text{if } \beta \ge 1, \\ \int_0^t \int_{\mathbb{R}^+} \int_{\mathbb{R}} x \left(1_{\{x<0,\ u<\varphi^-_{s-}\}} - 1_{\{x<0,\ u<\varphi^-_{(i-1)\Delta_n-}\}}\right) \underline{\mu}(ds,du,dx) \\ + \int_0^t \int_{\mathbb{R}^+} \int_{\mathbb{R}} x \left(1_{\{x>0,\ u<\varphi^+_{s-}\}} - 1_{\{x>0,\ u<\varphi^+_{(i-1)\Delta_n-}\}}\right) \underline{\mu}(ds,du,dx), & \text{if } \beta < 1, \end{cases}$$

where $\underline{\mu}$ is a Poisson measure with compensator $ds \otimes du \otimes \nu(x)dx$.

The rest of the proof consists in showing the asymptotic negligibility of the scaled sums of the terms a_i^j for $j = 1, ..., 7$ and their subcomponents. We will use either convergence in L_1 or L_2 for proving this.

We start with the term a_i^1. Application of Itô lemma gives

$$a_i^1 = \tilde{a}_i^1 + \hat{a}_i^1,$$

$$\tilde{a}_i^1 = 2\int_{(i-1)\Delta_n}^{i\Delta_n} Z_s^n \tilde{\alpha}_s ds, \quad \hat{a}_i^1 = 2\int_{(i-1)\Delta_n}^{i\Delta_n} Z_s^n \sigma_s dW_s, \quad Z_s^n = Z_s - Z_{(i-1)\Delta_n}.$$

Further decomposing $\tilde{a}_i^1 = \tilde{a}_i^1(1) + \tilde{a}_i^1(2) + \tilde{a}_i^1(3)$,

$$\begin{cases} \tilde{a}_i^1(1) = 2\int_{(i-1)\Delta_n}^{i\Delta_n} \tilde{\alpha}_{(i-1)\Delta_n} \int_{(i-1)\Delta_n}^{s} \sigma_u dW_u ds, \\ \tilde{a}_i^1(2) = 2\int_{(i-1)\Delta_n}^{i\Delta_n} \tilde{\alpha}_{(i-1)\Delta_n} \int_{(i-1)\Delta_n}^{s} \tilde{\alpha}_u dW_u ds, \\ \tilde{a}_i^1(3) = 2\int_{(i-1)\Delta_n}^{i\Delta_n} (\tilde{\alpha}_s - \tilde{\alpha}_{(i-1)\Delta_n}) Z_s^n ds, \end{cases}$$

we have for $q \geq 2$

$$\mathbb{E}_{i-1}^n \tilde{a}_i^1(1) = 0, \quad \mathbb{E}|\tilde{a}_i^1(1)|^q \leq C\Delta_n^{3q/2},$$

while for $q \geq 1$

$$\mathbb{E}|\tilde{a}_i^1(2)|^q \leq C\Delta_n^{2q}, \quad \mathbb{E}|\tilde{a}_i^1(3)|^q \leq C\Delta_n^{3q/2+q/2\wedge 1}.$$

Also, for $q \geq 2$ using Doob's inequality and the Cauchy-Schwartz inequality,

$$\mathbb{E}|\hat{a}_i^1|^q \leq C\mathbb{E}\left(\int_{(i-1)\Delta_n}^{i\Delta_n} (Z_s^n\sigma_s)^2 ds\right)^{q/2} \leq C\Delta_n^{q/2-1}\int_{(i-1)\Delta_n}^{i\Delta_n} \mathbb{E}|Z_s^n\sigma_s|^q ds \leq C\Delta_n^q.$$

Using Holder's inequality for any $1 \leq q < p/2$,

$$\mathbb{E}|a_i^2|^q \leq (\mathbb{E}|\Delta_i^n Z|^p)^{2q/p} (\mathbb{P}(|\Delta_i^n p| \geq \alpha\Delta_n^{\varpi}))^{1-2q/p} \leq C\Delta_n^{q+(1-2q/p)(1-\beta\omega)-\epsilon}, \quad \epsilon > 0.$$

We proceed with a_i^3 and the following decomposition,

$$|a_i^3| \leq |\overline{a}_i^3| + |\tilde{a}_i^3| + |\hat{a}_i^3|, \quad \overline{a}_i^3 = (\Delta_i^n Y)^2 1_{\{|\Delta_i^n Y| \leq 1.5\alpha\Delta_n^{\varpi}\}},$$

$$\tilde{a}_i^3 = (\Delta_i^n Y)^2 1_{\{|\tilde{Z}_i^n| \geq 0.25\alpha\Delta_n^{\varpi}\}}, \quad \hat{a}_i^3 = (\Delta_i^n Y)^2 1_{\{|\hat{Z}_i^n| \geq 0.25\alpha\Delta_n^{\varpi}\}}.$$

Then, for any $\epsilon > 0$ and $q \geq \beta/2$,

$$\mathbb{E}|\overline{a}_i^3|^q \leq C\Delta_n^{(2q-\beta-\epsilon)\varpi}\mathbb{E}|\Delta_i^n Y|^{\beta+\epsilon} \leq C\Delta_n^{(2q-\beta)\varpi+1-\epsilon}.$$

For $\tilde{a}_i^3$ and $q < p/2$, we can write

$$\begin{aligned} \mathbb{E}|\tilde{a}_i^3|^q &\leq C\mathbb{E}\left(|\Delta_i^n\tilde{Y}|^{2q} 1_{\{|\tilde{Z}_i^n| \geq 0.25\alpha\Delta_n^{\varpi}\}}\right) \\ &\leq C\left(\mathbb{E}|\Delta_i^n\tilde{Y}|^p\right)^{2q/p}\left(\mathbb{P}\left(|\tilde{Z}_i^n| \geq 0.25\alpha\Delta_n^{\varpi}\right)\right)^{1-2q/p} \leq C\Delta_n^{1+(\frac{1}{2}-\varpi)(p-2q)}. \end{aligned}$$

Further decomposing $\hat{a}_i^3$,

$$\mathbb{E}|\hat{a}_i^3|^q \leq C\left(\mathbb{E}|\hat{a}_i^3(1)|^q + \mathbb{E}|\hat{a}_i^3(2)|^q\right),$$
$$\hat{a}_i^3(1) = \left(\Delta_i^n Y^{(1)}\right)^2 1_{\{|\hat{Z}_i^n|\geq 0.25\alpha\Delta_n^{\varpi}\}}, \quad \hat{a}_i^3(2) = \left(\Delta_i^n Y^{(2)}\right)^2 1_{\{|\hat{Z}_i^n|\geq 0.25\alpha\Delta_n^{\varpi}\}},$$

where $\hat{Z}_i^n$ is defined on the probability space of $Y_t^{(1)}$, and $Y_t^{(2)}$ as the same process on the original probability space, where for simplicity we have kept the same notation for this process. Then, for $1 \leq q < p/2$ we have

$$\mathbb{E}|\hat{a}_i^3(1)|^q \leq \mathbb{E}\left(\mathbb{E}_{i-1}^n|\Delta_i^n Y^{(1)}|^{2q}\mathbb{E}_{i-1}^n\left(1_{\{|\hat{Z}_i^n|\geq 0.25\alpha\Delta_n^{\varpi}\}}\right)\right) \leq C\Delta_n^{1+(\frac{1}{2}-\varpi)(p-2)}.$$

Using Holder's inequality, it follows that for every $x > 1$,

$$\mathbb{E}|\hat{a}_i^3(2)|^q \leq \left(\mathbb{E}|\Delta_i^n Y^{(2)}|^{2qx}\right)^{1/x}\left(\mathbb{P}\left(|\hat{Z}_i^n| \geq 0.25\alpha\Delta_n^{\varpi}\right)\right)^{1-1/x} \leq C\Delta_n^{\frac{3}{2}\frac{1}{x}+(1-\frac{1}{x})(\frac{1}{2}-\varpi)p}.$$

Next, note that $\mathbb{E}_{i-1}^n a_i^4 = 0$, and for $2 \leq q < p$

$$\mathbb{E}|a_i^4|^q \leq C\left(\mathbb{E}|\Delta_i^n Z|^p\right)^{q/p}\left(\mathbb{E}|\Delta_i^n\tilde{Y}|^{\frac{qp}{p-q}}\right)^{1-q/p} \leq C\Delta_n^{1+q(1/2-1/p)}.$$

We can decompose a_i^5 as follows

$$a_i^5 = \tilde{a}_i^5 + \hat{a}_i^5,\ \tilde{a}_i^5 = \Delta_n(C^+\varphi_{(i-1)\Delta_n-}^+ + C^-\varphi_{(i-1)\Delta_n-}^-)\Delta_i^n\hat{Z},$$
$$\hat{a}_i^5 = \Delta_i^n\hat{Z}\int_{(i-1)\Delta_n}^{i\Delta_n}(C^+(\varphi_s^+ - \varphi_{(i-1)\Delta_n-}^+) + C^-(\varphi_s^- - \varphi_{(i-1)\Delta_n-}^-))ds,$$

where C^+ and C^- are some constants. Thus for $q \geq 2$

$$\mathbb{E}_{i-1}^n\tilde{a}_i^5 = 0, \quad \mathbb{E}|\tilde{a}_i^5|^q \leq C\Delta_n^{3q/2}.$$

Application of Holder's inequality for $q \geq 1$ and arbitrary small $\epsilon > 0$ implies that

$$\mathbb{E}|\hat{a}_i^5|^q \leq C\Delta_n^{3q/2+q/2\wedge 1-\epsilon}.$$

We turn now to a_i^6, which may be bounded as

$$|a_i^6| \leq |a_i^{6a}| + |a_i^{6b}| + |a_i^{6c}|, \quad a_i^{6a} = 2\hat{Z}_i^n\Delta_i^n Y 1_{\{|\tilde{Z}_i^n|>0.25\alpha\Delta_n^{\varpi}\}},$$
$$a_i^{6b} = 2\hat{Z}_i^n\Delta_i^n Y 1_{\{|\Delta_i^n Y|>0.5\alpha\Delta_n^{\varpi}\}}, \quad a_i^{6c} = 2\hat{Z}_i^n\Delta_i^n Y 1_{\{|\hat{Z}_i^n|>0.25\alpha\Delta_n^{\varpi}\}}.$$

For a_i^{6a} and $q < p$, we can apply Holder's inequality twice together with the fact that moments of all powers of the normal distribution exist, to conclude that some sufficiently small $\epsilon > 0$,

$$\begin{aligned}\mathbb{E}|a_i^{6a}|^q &\leq C\left(\mathbb{E}|\Delta_i^n Y|^p\right)^{q/p}\left(\mathbb{E}\left(|\sigma_{(i-1)\Delta_n}|^{\frac{qp}{p-q}}|\Delta_i^n W|^{\frac{qp}{p-q}}1_{\{|\tilde{Z}_i^n|>0.25\alpha\Delta_n^{\varpi}\}}\right)\right)^{1-q/p}\\ &\leq C\Delta_n^{\frac{q}{p}+q/2}\left(\mathbb{E}\left(|\sigma_{(i-1)\Delta_n}|^{(1+\epsilon)\frac{qp}{p-q}}1_{\{|\tilde{Z}_i^n|>0.25\alpha\Delta_n^{\varpi}\}}\right)\right)^{\frac{1-q/p}{1+\epsilon}}\\ &\leq C\Delta_n^{\frac{q}{p}+q/2}\left(\mathbb{E}\left(|\sigma_{(i-1)\Delta_n}|^{(1+\epsilon)\frac{qp}{p-q}}\mathbb{P}_{i-1}^n\left(|\tilde{Z}_i^n| > 0.25\alpha\Delta_n^{\varpi}\right)\right)\right)^{\frac{1-q/p}{1+\epsilon}}.\\ &\leq C\Delta_n^{\frac{q}{p}+q/2}\Delta_n^{\frac{1-q/p}{1+\epsilon}+(\frac{1}{2}-\varpi)\frac{(p-q)-(1+\epsilon)q}{(1+\epsilon)}}.\end{aligned}$$

Next we have

$$\mathbb{E}|a_i^{6b}|^q \leq C\left(\mathbb{E}|a_i^{6b}(1)|^q + \mathbb{E}|a_i^{6b}(2)|^q + \mathbb{E}|a_i^{6b}(3)|^q\right), \quad a_i^{6b}(1) = 2\hat{Z}_i^n \Delta_i^n Y^{(1)} 1_{\{|\Delta_i^n Y^{(1)}|>0.25\alpha\Delta_n^{\varpi}\}},$$
$$a_i^{6b}(2) = 2\hat{Z}_i^n \Delta_i^n Y^{(1)} 1_{\{|\Delta_i^n Y^{(2)}|>0.25\alpha\Delta_n^{\varpi}\}}, \quad a_i^{6b}(3) = 2\hat{Z}_i^n \Delta_i^n Y^{(2)} 1_{\{|\Delta_i^n Y|>0.5\alpha\Delta_n^{\varpi}\}}.$$

Then, $\mathbb{E}_{i-1}^n a_i^{6b}(1) = 0$ and for some arbitrary small $\epsilon > 0$,

$$\mathbb{E}|a_i^{6b}(1)|^q \leq C\mathbb{E}\left(\mathbb{E}_{i-1}^n|\hat{Z}_i^n|^q \mathbb{E}_{i-1}^n|\Delta_i^n Y^{(1)}|^q 1_{\{|\Delta_i^n Y^{(1)}|>0.25\alpha\Delta_n^{\varpi}\}}\right) \leq C\Delta_n^{1+q/2-((\beta-q)\vee 0)\varpi-\epsilon}.$$

For $p > q \vee \beta$ and any $\epsilon > 0$,

$$\begin{aligned}\mathbb{E}|a_i^{6b}(2)|^q &\leq C\left(\mathbb{E}|\hat{Z}_i^n \Delta_i^n Y^{(1)}|^p\right)^{q/p}\left(\mathbb{E}\left(1_{\{|\Delta_i^n Y^{(2)}|>0.25\alpha\Delta_n^{\varpi}\}}\right)\right)^{1-q/p} \\ &\leq C\Delta_n^{\frac{q}{2}+\frac{q}{p}+(1-\frac{q}{p})(\frac{3}{2}-\beta\varpi)-\epsilon}.\end{aligned}$$

By Holder's inequality for any $\epsilon > 0$ and $q < p$,

$$\mathbb{E}|a_i^{6b}(3)|^q \leq C\left(\mathbb{E}|\hat{Z}_i^n|^p\right)^{q/p}\left(\mathbb{E}|\Delta_i^n Y^{(2)}|^{qp/(p-q)}\right)^{1-q/p} \leq C\Delta_n^{\frac{q}{2}+\frac{3}{2}\left(\frac{1}{\beta}\wedge\frac{p-q}{pq}\right)q-\epsilon}.$$

Next, note that

$$\mathbb{E}|a_i^{6c}|^q \leq C\left(\mathbb{E}|a_i^{6c}(1)|^q + \mathbb{E}|a_i^{6c}(2)|^q\right),$$
$$a_i^{6c}(1) = 2\hat{Z}_i^n \Delta_i^n Y^{(1)} 1_{\{|\hat{Z}_i^n|>0.25\alpha\Delta_n^{\varpi}\}}, \quad a_i^{6c}(2) = 2\hat{Z}_i^n \Delta_i^n Y^{(2)} 1_{\{|\hat{Z}_i^n|>0.25\alpha\Delta_n^{\varpi}\}}.$$

Also, for $q < p$

$$\mathbb{E}|a_i^{6c}(1)|^q \leq C\mathbb{E}\left(\mathbb{E}_{i-1}^n\left(|\hat{Z}_i^n|^q 1_{\{|\hat{Z}_i^n|>0.25\alpha\Delta_n^{\varpi}\}}\right)\mathbb{E}_{i-1}^n|\Delta_i^n Y^{(1)}|^q\right) \leq C\Delta_n^{\frac{q}{2}+\frac{q}{\beta}\wedge 1+(\frac{1}{2}-\varpi)(p-q)}.$$

For $a_i^{6c}(2)$, we can proceed the same way as for $\hat{a}_i^{6b}(3)$ and get that for any $\epsilon > 0$ and $q < p$,

$$\mathbb{E}|a_i^{6c}(2)|^q \leq C\left(\mathbb{E}|\hat{Z}_i^n|^p\right)^{q/p}\left(\mathbb{E}|\Delta_i^n Y^{(2)}|^{qp/(p-q)}\right)^{1-q/p} \leq C\Delta_n^{\frac{q}{2}+\frac{3}{2}\left(\frac{1}{\beta}\wedge\frac{p-q}{pq}\right)q-\epsilon}.$$

Turning to a_i^7, we have for Δ_n small enough using the integrability conditions on α_s and σ_s, and the Cauchy-Schwartz inequality,

$$\mathbb{E}|a_i^7|^q \leq C\mathbb{E}\left(|\Delta_i^n \tilde{Z}\Delta_i^n Y|^q 1_{\{|\Delta_i^n Y|\leq\alpha\Delta_n^{\varpi}\}}\right) \leq \Delta_n^{1+1/(2q)+(2q-\beta)\varpi/2}.$$

$\square$

Lemma 2 *Suppose we observe the process p_t at the discrete times $0, \Delta_n, ..., n\Delta_n, ..., t, t+\Delta_n, ..., t+n\Delta_n, ...$ Assume that A1-A4 hold, and let either $X_t = \int_t^{t+1}\sigma_s^2 ds$ or $X_t = \int_t^{t+1}\int_{\psi^+(x)>tr_T}\tilde{\phi}_i(\psi(x)-tr_T, \theta^{(1)}, \theta^{(2)})\mu(ds, dx)$ for $i = 1, 2$ and $\theta \in \Theta_T^l$. Then for some $\alpha > 0$ and $\varpi \in \left(0, \frac{1}{2}\right)$,*

$$\frac{\sqrt{N_T}}{T}\sum_{t=1}^{T-1}X_t\left(TV_{t-1}^n - \int_{t-1}^t \sigma_s^2 ds\right) \xrightarrow{\mathbb{P}} 0, \tag{8.7}$$

provided that $\sqrt{N_T}\Delta_n^{(2-\beta)\varpi} \to 0$.

Proof: We make the same decomposition of the difference

$$\left(\left(\Delta_i^{n,t-1} p\right)^2 1_{\{|\Delta_i^{n,t-1} p| \le \alpha \Delta_n^{\varpi}\}} - \int_{t-1+(i-1)\Delta_n}^{t-1+i\Delta_n} \sigma_s^2 ds \right)$$

as in Lemma 1. We denote the corresponding components in this decomposition for the high-frequency interval $[t+(i-1)\Delta_n, t+i\Delta_n]$ by $a_{t,i}^j$, with the components denoted analogously. Then, using Holder's inequality for some $a > 1$,

$$\frac{1}{T-1}\sum_{t=1}^{T-1} X_t L_t \le \left(\frac{1}{T-1}\sum_{t=1}^{T-1} |X_t|^a\right)^{1/a} \left(\frac{1}{T-1}\sum_{t=1}^{T-1} |L_t|^{a/(a-1)}\right)^{1-1/a},$$

where $L_t = \sum_{i=1}^n a_{t,i}^j$, or the identical sum over their subcomponents. For the terms involving $\tilde{a}_{t,i}^1(1)$, $\hat{a}_{t,i}^1$, $a_{t,i}^4$, $\tilde{a}_{t,i}^5$ and $a_{t,i}^{6b}(1)$, we can use $a = 2$, and show convergence to zero of $\frac{N_T}{T}\sum_{t=1}^T |L_t|^2$. The latter follows from the bounds on the second moments of $\hat{a}_{t,i}^1$ and $a_{t,i}^4$ derived in the previous Lemma, and the fact that these terms form martingale difference sequences.

For the rest of the terms we can set set a arbitrarily large. $\frac{1}{T-1}\sum_{t=1}^{T-1} |X_t|^a$ will be bounded in L_1 by the integrability assumptions on the processes σ_s and $\varphi_s^{\pm}$ in A4. Asymptotic negligibility of $\frac{N_T^{0.5a/(a-1)}}{T-1}\sum_{t=1}^{T-1} |L_t|^{a/(a-1)}$ follows from the basic inequality $\left(\sum_{i=1}^N |a_i|\right)^q \le N^{q-1}\sum_{i=1}^N |a_i|^q$ for any $q > 1$, and the bounds derived in the previous Lemma, when we pick a sufficiently high. $\square$

Lemma 3 *Suppose we observe the process p_t at the discrete times $0, \Delta_n, 2\Delta_n, ..., [T/\Delta_n]$. Let $f_T(x)$ be a function in x (changing with T) with the following properties for a given deterministic sequence $tr_T > 1$ (which depends only on the time span T):*

(a) $f_T(x) = 0$ for $x < \log(tr_T)$,

(b) $|f_T(x)| \le C(\log(tr_T) \vee x)$ and $|f_T'(x)| \le C$ for $x \ge \log(tr_T)$,

(c) $|f_T(\log(tr_T) + \delta)| \le C_\delta$ for $\delta \ge 0$, where $C_\delta > 0$ is a constant that depends on δ,

and $f_T'(x)$ denotes the right derivative for $x = \log(tr_T)$. Then under assumption A1, with $\nu(x)$ nondecreasing for x sufficiently large, and assumption A4, we have

$$\frac{1}{\sqrt{N_T}}\left(\sum_{i=1}^{[T/\Delta_n]} f_T(\Delta_i^n p) - \sum_{s\le T} f_T(\Delta p_s)\right) \xrightarrow{\mathbb{P}} 0, \quad \text{as } \Delta_n \downarrow 0,\ T \uparrow \infty, \tag{8.8}$$

for $N_T = T\overline{\nu}^+(\log(tr_T))$ with $\overline{\nu}^+(z) = \int_z^\infty x\nu(x)dx$, provided that

$$\sqrt{N_T \Delta_n^{1-\epsilon}}\left(1 \bigvee \frac{\sqrt{\Delta_n}\log(tr_T)}{\overline{\nu}^+(\log(tr_T))}\right) \to 0, \tag{8.9}$$

where $\epsilon > 0$ is arbitrary small.

Proof: For a constant $K > 0$, we denote

$$p_t(K) = \int_0^t \alpha_s ds + \int_0^t \sigma_s dW_s + \int_0^t \int_{|x|<K} \kappa(x)\tilde{\mu}(ds, dx) + \int_0^t \int_{|x|<K} \kappa'(x)\mu(ds, dx).$$

The proof goes through several steps.

Step 1. We start by showing that for any $s < t$ and $K \to \infty$,

$$\mathbb{P}\left(\sup_{s\le u\le t} \varphi_u^- + \sup_{s\le u\le t} \varphi_u^+ \ge K\right) \le K^{-q}, \quad \forall q > 0.$$

Using the assumed dynamics for $\varphi_u^\pm$, we can write $\varphi_u^\pm - \varphi_s^\pm = \int_s^u \alpha_v^{\pm\prime} dv + \int_s^u \sigma_v^{\pm\prime} dW_v + \int_s^u \sigma_v^{\pm\prime\prime} dB_v + \int_s^u \int_{\mathbb{R}^2} \kappa(\delta^\pm(v-, \mathbf{x}))\tilde{\mu}'(ds, d\mathbf{x}) + \int_s^u \int_{\mathbb{R}^2} \kappa'(\delta^\pm(v-, \mathbf{x}))\mu'(ds, d\mathbf{x})$. Then, using the basic inequality $|\varphi_u^\pm| \le |\varphi_s^\pm| + |\varphi_u^\pm - \varphi_s^\pm|$, the Burkholder-Davis-Gundy inequality, Chebychev's inequality, and finally the integrability assumption on the process $\varphi_t^\pm$, it is possible to derive the result of this step.

Step 2. We next show that for some constants $K_0, K_1 > 0$ and $\alpha < 1/\beta$,

$$\mathbb{P}\left(\int_{(i-1)\Delta_n}^{i\Delta_n} \int_{|x|\ge K_0\Delta_n^\alpha} \mu(ds, dx) \ge K_1\right) \le C\Delta_n^{(1-\alpha\beta-\iota)\lfloor K_1\rfloor}, \quad \forall \iota > 0.$$

We start by introducing the following sets that we will rely on later in the proof,

$$R_n = \{x : |x| \ge K_0\Delta_n^\alpha\}, \; S_n = \{x : |x| \le K_0\Delta_n^\alpha\}, \; T_n = \{x : K_0\Delta_n^\alpha \le |x| \le K_0\}.$$

Using the representation of the jumps with the homogenous measure $\underline{\mu}$ introduced in Lemma 1, we may write

$$\begin{aligned}
&\mathbb{P}\left(\int_{(i-1)\Delta_n}^{i\Delta_n} \int_{|x|\ge K_0\Delta_n^\alpha} \mu(ds, dx) \ge K_1\right) \\
&= \mathbb{P}\left(\int_{(i-1)\Delta_n}^{i\Delta_n} \int_{\mathbb{R}^+} \int_{x\in R_n} \left(1_{\{x<0,\ u<\varphi_{s-}^-\}} + 1_{\{x>0,\ u<\varphi_{s-}^+\}}\right) \underline{\mu}(ds, du, dx) \ge K_1\right) \\
&\le \mathbb{P}\left(\int_{(i-1)\Delta_n}^{i\Delta_n} \int_{\mathbb{R}^+} \int_{x\in R_n} 1_{\{u<\Delta_n^{-\epsilon}\}} \underline{\mu}(ds, du, dx) \ge K_1\right) \\
&\quad + \mathbb{P}\left(\sup_{s\in[(i-1)\Delta_n, i\Delta_n]} \varphi_s^- \ge \Delta_n^{-\epsilon}\right) + \mathbb{P}\left(\sup_{s\in[(i-1)\Delta_n, i\Delta_n]} \varphi_s^+ \ge \Delta_n^{-\epsilon}\right),
\end{aligned}$$

where $\epsilon > 0$ is arbitrary small. For the second probability we can apply the result of Step 1. For the first probability, we can use the fact that

$$\int_{(i-1)\Delta_n}^{i\Delta_n} \int_0^{\Delta_n^{-\epsilon}} \int_{x\in R_n} \underline{\mu}(ds, du, dx)$$

has a Poisson distribution with intensity $\Delta_n^{1-\epsilon} \int_{|x|\ge K_0\Delta_n^\alpha} \nu(x)dx$. Therefore,

$$\begin{aligned}
\mathbb{P}\left(\int_{(i-1)\Delta_n}^{i\Delta_n} \int_0^{\Delta_n^{-\epsilon}} \int_{x\in R_n} \underline{\mu}(ds, du, d\mathbf{x}) \ge K_1\right) &\le C\Delta_n^{(1-\epsilon)\lfloor K_1\rfloor} \left(\int_{|x|\ge K_0\Delta_n^\alpha} \nu(x)dx\right)^{\lfloor K_1\rfloor} \\
&\le C\Delta_n^{(1-\beta\alpha-2\epsilon)\lfloor K_1\rfloor},
\end{aligned}$$

where for the last inequality made use of the fact that $\int_{\mathbb{R}}(|x|^{\beta+\iota}\wedge 1)\nu(x)dx < \infty$ for $\iota > 0$ arbitrary small, in particular $\iota < \epsilon$.

Step 3. For $\beta < 1$ and for every $\iota > 0$, we have

$$\mathbb{P}\left(\Big|\int_{(i-1)\Delta_n}^{i\Delta_n}\int_{|x|\le K_0}\kappa(x)\tilde{\mu}(ds,dx)+\int_{(i-1)\Delta_n}^{i\Delta_n}\int_{|x|\le K_0}\kappa'(x)\mu(ds,dx)\Big|\ge K_1\right)\le C\Delta_n^{(1-\iota)\lfloor K_1/K_0\rfloor}.$$

This result follows from the fact that $\int_{(i-1)\Delta_n}^{i\Delta_n}\int_{\mathbb{R}}\kappa(x)\nu_s(x)dxds < \infty$ for $\beta < 1$, together with the following sequence of bounds,

$$\begin{aligned}
&\mathbb{P}\left(\Big|\int_{(i-1)\Delta_n}^{i\Delta_n}\int_{|x|\le K_0}\kappa(x)\tilde{\mu}(ds,dx)+\int_{(i-1)\Delta_n}^{i\Delta_n}\int_{|x|\le K_0}\kappa'(x)\mu(ds,dx)\Big|\ge K_1\right)\\
&\le \mathbb{P}\left(\Big|\int_{(i-1)\Delta_n}^{i\Delta_n}\int_{|x|\le K_0}x\mu(ds,dx)\Big|\ge K_1-C\int_{(i-1)\Delta_n}^{i\Delta_n}(\varphi_s^-+\varphi_s^+)ds\right)\\
&\le \mathbb{P}\left(\Big|\int_{(i-1)\Delta_n}^{i\Delta_n}\int_{|x|\le K_0}x\mu(ds,dx)\Big|\ge K_1-C\Delta_n^{1-\epsilon}\right)+\mathbb{P}\left(\sup_{s\in[(i-1)\Delta_n,i\Delta_n]}\varphi_s^-\ge\Delta_n^{-\epsilon}\right)\\
&\quad+\mathbb{P}\left(\sup_{s\in[(i-1)\Delta_n,i\Delta_n]}\varphi_s^+\ge\Delta_n^{-\epsilon}\right)\\
&\le \mathbb{P}\left(\Big|\int_{(i-1)\Delta_n}^{i\Delta_n}\int_{|x|\le K_0}\mu(ds,dx)\Big|\ge\frac{K_1-C\Delta_n^{1-\epsilon}}{K_0}\right)+C\Delta_n^q,
\end{aligned}$$

for arbitrary small $\epsilon > 0$, and arbitrary large $q > 0$ (recall C denotes some positive constant). From here we can apply the result of Step 2 to get the final result of this step.

Step 4. For $\beta \ge 1$ we prove

$$\mathbb{P}\left(\Big|\int_{(i-1)\Delta_n}^{i\Delta_n}\int_{|x|\le K_0}\kappa(x)\tilde{\mu}(ds,dx)+\int_{(i-1)\Delta_n}^{i\Delta_n}\int_{|x|\le K_0}\kappa'(x)\mu(ds,dx)\Big|\ge K_1\right)\le C\Delta_n^{\lfloor K_1/K_0\rfloor-\iota},$$

for $\iota > 0$ sufficiently small. Again, relying on the representation of the jumps by the homogenous Poisson measure $\underline{\mu}$ on an extended space with the extra dimension used for thinning, we have for Δ_n sufficiently small ($\kappa'(x)$ is zero for $|x|$ in some neighborhood of 0) and $0 < \alpha < 1/\beta$,

$$\begin{aligned}
&\mathbb{P}\left(\Big|\int_{(i-1)\Delta_n}^{i\Delta_n}\int_{|x|\le K_0}\kappa(x)\tilde{\mu}(ds,dx)+\int_{(i-1)\Delta_n}^{i\Delta_n}\int_{|x|\le K_0}\kappa'(x)\mu(ds,dx)\Big|\ge K_1\right)\\
&\le \mathbb{P}\Bigg(\Big|\int_{(i-1)\Delta_n}^{i\Delta_n}\int_{\mathbb{R}^+}\int_{x\in T_n}\kappa(x)\left(1_{\{x<0,\ u<\varphi_{s-}^-\}}+1_{\{x>0,\ u<\varphi_{s-}^+\}}\right)\underline{\tilde{\mu}}(ds,du,dx)\\
&\qquad+\int_{(i-1)\Delta_n}^{i\Delta_n}\int_{\mathbb{R}^+}\int_{x\in T_n}\kappa'(x)\left(1_{\{x<0,\ u<\varphi_{s-}^-\}}+1_{\{x>0,\ u<\varphi_{s-}^+\}}\right)\underline{\mu}(ds,du,dx)\Big|\ge\rho K_1\Bigg)\\
&\quad+\mathbb{P}\left(\Big|\int_{(i-1)\Delta_n}^{i\Delta_n}\int_{\mathbb{R}^+}\int_{x\in S_n}\kappa(x)\left(1_{\{x<0,\ u<\varphi_{s-}^-\}}+1_{\{x>0,\ u<\varphi_{s-}^+\}}\right)\underline{\tilde{\mu}}(ds,du,dx)\Big|\ge(1-\rho)K_1\right),
\end{aligned}$$

for any $\rho \in (0,1)$. For the second probability successive applications of the Burkholder-Davis-Gundy inequality together with the fact that $\int_{|x|\le K_0\Delta_n^\alpha}|\kappa(x)|^q\nu(x)dx \le C\Delta_n^{(q-\beta)\alpha-\iota}$ for $q > \beta$

and $\iota > 0$ arbitrary small, imply that for any $q > \beta$,

$$\mathbb{P}\left(\left| \int_{(i-1)\Delta_n}^{i\Delta_n} \int_{\mathbb{R}^+} \int_{x \in S_n} \kappa(x) \left(1_{\{x<0,\ u<\varphi_{s-}^-\}} + 1_{\{x>0,\ u<\varphi_{s-}^+\}}\right) \underline{\tilde{\mu}}(ds, du, dx)\right| \geq (1-\rho)K_1\right)$$
$$\leq C\Delta_n \int_{|x| \leq K_0 \Delta_n^\alpha} |\kappa(x)|^q \nu(x) dx + C\Delta_n^{q\alpha + (1-\beta\alpha)z - \iota} \leq C\Delta_n^{(q-\beta)\alpha - \iota},$$

for $\iota > 0$ arbitrary small, and some $z \in [0, 1)$. For the first probability, following Step 3 we can split the integral with respect to the compensated measure into two parts, and then for Δ_n small conclude that

$$\mathbb{P}\Bigg(\Bigg| \int_{(i-1)\Delta_n}^{i\Delta_n} \int_{\mathbb{R}^+} \int_{x \in T_n} \kappa(x) \left(1_{\{x<0,\ u<\varphi_{s-}^-\}} + 1_{\{x>0,\ u<\varphi_{s-}^+\}}\right) \underline{\tilde{\mu}}(ds, du, dx)$$
$$+ \int_{(i-1)\Delta_n}^{i\Delta_n} \int_{\mathbb{R}^+} \int_{x \in T_n} \kappa'(x) \left(1_{\{x<0,\ u<\varphi_{s-}^-\}} + 1_{\{x>0,\ u<\varphi_{s-}^+\}}\right) \underline{\mu}(ds, du, dx)\Bigg| \geq \rho K_1\Bigg)$$
$$\leq \mathbb{P}\Bigg(\Bigg| \int_{(i-1)\Delta_n}^{i\Delta_n} \int_{\mathbb{R}^+} \int_{x \in T_n} x \left(1_{\{x<0,\ u<\varphi_{s-}^-\}} + 1_{\{x>0,\ u<\varphi_{s-}^+\}}\right) \underline{\mu}(ds, du, dx)\Bigg|$$
$$\geq \rho K_1 - C\Delta_n^{\alpha(1-\beta-\iota)} \int_{(i-1)\Delta_n}^{i\Delta_n} (\varphi_s^- + \varphi_s^+) ds\Bigg)$$
$$\leq \mathbb{P}\left(\left| \int_{(i-1)\Delta_n}^{i\Delta_n} \int_0^{\Delta_n^{-\epsilon}} \int_{x \in T_n} x \underline{\mu}(ds, du, dx)\right| \geq \rho K_1 - C\Delta_n^{1-\alpha(\beta-1+\iota)-\epsilon}\right)$$
$$+ \mathbb{P}\left(\sup_{s \in [(i-1)\Delta_n, i\Delta_n]} \varphi_s^- \geq \Delta_n^{-\epsilon}\right) + \mathbb{P}\left(\sup_{s \in [(i-1)\Delta_n, i\Delta_n]} \varphi_s^+ \geq \Delta_n^{-\epsilon}\right)$$
$$\leq C\mathbb{P}\left(\int_{(i-1)\Delta_n}^{i\Delta_n} \int_0^{\Delta_n^{-\epsilon}} \int_{x \in R_n} \underline{\mu}(ds, du, dx) \geq \frac{\rho K_1 - C\Delta_n^{1-\alpha(\beta-1+\iota)-\epsilon}}{K_0}\right)$$
$$\leq C\Delta_n^{(1-\alpha\beta)\lfloor \rho K_1/K_0 \rfloor},$$

where $\iota > 0$ is arbitrary small, and we made use of the result of Step 1 and Step 2. The final part of the proof for this step then follows by applying the above two inequalities with ρ sufficiently close to 1, and α close to 0.

Step 5. Using the integrability assumptions on the processes α and σ, we have

$$\mathbb{P}\left(\left| \int_{(i-1)\Delta_n}^{i\Delta_n} \alpha_s ds + \int_{(i-1)\Delta_n}^{i\Delta_n} \sigma_s dW_s\right| \geq K_1\right) \leq C\Delta_n^q, \quad \forall q > 0.$$

Step 6. Combining the results of steps 1-5, it follows that for any $\lfloor K_1/K_0 \rfloor > q$,

$$\mathbb{P}\left(|\Delta_i^n p(K_0)| > K_1\right) \leq C\Delta_n^{q-\iota},$$

for some arbitrary small $\iota > 0$.

Step 7. For some $\delta \in (0, 1)$ with $\delta > 1/tr_T$, let $K > 0$ be such that $K < |\log(\delta)|/3 \wedge \log(\delta tr_T)$, then

$$\mathbb{E}\left(\sum_{i=1}^{[T/\Delta_n]} \left| f_T(\Delta_i^n p) - \sum_{(i-1)\Delta_n \leq s \leq i\Delta_n} f_T(\Delta p_s)\right| 1_{\{A_i\}}\right) \leq CT \log(tr_T) \Delta_n^{1-\epsilon},$$

where $A_i = \left\{\omega : \int_{(i-1)\Delta_n}^{i\Delta_n} \int_{|x|\geq K} \mu(ds, dx) \geq 2\right\}$ and $\iota > 0$ is arbitrary small. To prove this result we first use the fact that

$$\begin{cases} f_T(\Delta_i^n p) \leq C(\log(tr_T) \vee |\Delta_i^n p|), \\ \mathbb{E}\left(\sum_{(i-1)\Delta_n \leq s \leq i\Delta_n} |f_T(\Delta p_s)|\right) \leq C \log(tr_T)\Delta_n \overline{\nu}^+(\log(tr_T)). \end{cases}$$

The result then follows readily from an application of Holder's inequality and the result of Step 2.

Step 8. For the same choice of δ and K used in Step 7, denote the set $B_i = \{\omega : |\Delta_i^n p(K)| \geq |\log(\delta)|\}$. Then as in Step 7, using the result of Step 6, we get

$$\mathbb{E}\left(\sum_{i=1}^{[T/\Delta_n]} \left| f_T(\Delta_i^n p) - \sum_{(i-1)\Delta_n \leq s \leq i\Delta_n} f_T(\Delta p_s) \right| 1_{\{A_i^c,\ B_i\}}\right) \leq CT \log(tr_T)\Delta_n.$$

Step 9. Denote the sets $C_i = \{\omega : \exists s \in [(i-1)\Delta_n, i\Delta_n] : \Delta p_s \geq \log(tr_T)\}$ and $D_i = \{\omega : \Delta_i^n p \geq \log(tr_T)\}$. We will show that

$$\mathbb{E}\left(\sum_{i=1}^{[T/\Delta_n]} \left| f_T(\Delta_i^n p) - \sum_{(i-1)\Delta_n \leq s \leq i\Delta_n} f_T(\Delta p_s) \right| 1_{\{A_i^c,\ B_i^c,\ C_i^c,\ D_i\}}\right) \\ \leq C\Delta_n^{1/2-\iota} T (\overline{\nu}^+(\log(tr_T)))^{1-\iota},$$

for some arbitrary small $\iota > 0$. On the set $A_i^c \cap B_i^c \cap C_i^c \cap D_i$, there is exactly one jump of size above K in absolute value, and its size must be in the interval $[\log(\delta tr_T), \log(tr_T)]$ (recall the restriction $\delta > 1/tr_T$). Therefore, using the fact that $|f_T(x) - f_T(\log(tr_T))| \leq C|x - \log(tr_T)|$ for $x \geq \log(tr_T)$ (by first-order Taylor series expansion) and that $f_T(\log(tr_T)) \leq C$, we have

$$\begin{aligned} &\left| f_T(\Delta_i^n p) - \sum_{(i-1)\Delta_n \leq s \leq i\Delta_n} f_T(\Delta p_s) \right| 1_{\{A_i^c,\ B_i^c,\ C_i^c,\ D_i\}} \\ &\leq C|\Delta_i^n p(K)| \int_{(i-1)\Delta_n}^{i\Delta_n} \int_{\mathbb{R}} 1_{\{x \in [\log(\delta tr_T), \log(tr_T)]\}} \mu(ds, dx) \\ &+ C \int_{(i-1)\Delta_n}^{i\Delta_n} \int_{\mathbb{R}} 1\left(|\Delta_i^n p(K)| \geq \log(tr_T) - x,\ x \in [\log(\delta tr_T), \log(tr_T)]\right) \mu(ds, dx). \end{aligned} \tag{8.10}$$

Note that in the last integral, the integrand is not adapted but this does not matter as the integral with respect to μ is defined in the usual Riemann-Stieltjes sense; i.e. not in a stochastic sense. Now using the representation of the jumps with respect to $\underline{\mu}$, we have for the first term on the right-hand side of (8.10),

$$\begin{aligned}
&\mathbb{E}\left(|\Delta_i^n p(K)|\int_{(i-1)\Delta_n}^{i\Delta_n}\int_{\mathbb{R}} 1_{\{x\in[\log(\delta tr_T),\log(tr_T)]\}}\mu(ds,dx)\right)\\
&\leq \mathbb{E}\left(\left|\int_{(i-1)\Delta_n}^{i\Delta_n}\alpha_s ds+\int_{(i-1)\Delta_n}^{i\Delta_n}\sigma_s dW_s\right|\int_{(i-1)\Delta_n}^{i\Delta_n}\int_{\mathbb{R}} 1_{\{x\in[\log(\delta tr_T),\log(tr_T)]\}}\mu(ds,dx)\right)\\
&+\mathbb{E}\Bigg(\left|\int_{(i-1)\Delta_n}^{i\Delta_n}\int_{\mathbb{R}^+}\int_{|x|\leq K}\kappa(x)\left(1_{\{x<0,\ u<\varphi_{s-}^-\}}+1_{\{x>0,\ u<\varphi_{s-}^+\}}\right)\underline{\tilde{\mu}}(ds,du,dx)\right|\\
&\qquad\qquad\qquad\times\int_{(i-1)\Delta_n}^{i\Delta_n}\int_0^{\Delta_n^{-\epsilon}}\int_{x\in[\log(\delta tr_T),\log(tr_T)]}\underline{\mu}(ds,du,d\mathbf{x})\Bigg)\\
&+\mathbb{E}\left(\left|\int_{(i-1)\Delta_n}^{i\Delta_n}\int_0^{\Delta_n^{-\epsilon}}\int_{|x|\leq K}\kappa'(x)\underline{\mu}(ds,du,dx)\right|\int_{(i-1)\Delta_n}^{i\Delta_n}\int_0^{\Delta_n^{-\epsilon}}\int_{x\in[\log(\delta tr_T),\log(tr_T)]}\underline{\mu}(ds,du,dx)\right)\\
&+C\Delta_n^q\leq C\Delta_n^{3/2-\rho}\overline{\nu}^+(\log(\delta tr_T))^{1-\rho},
\end{aligned}\tag{8.11}$$

for some $\epsilon > 0$, $\rho > 0$ arbitrary small, and arbitrary $q > 0$. For the first expectation on the right hand side we have used Holder's inequality, for the second we have conditioned first on the filtration generated by $\underline{\mu}(\mathbb{R}^+, \mathbb{R}^+, x \in [\log(\delta tr_T), \log(tr_T)])$ and then applied the Burkholder-Davis-Gundy inequality and Holder's inequality, while for the last term we have used the independence of the filtration generated by $\underline{\mu}(\mathbb{R}^+, \mathbb{R}^+, x \in [\log(\delta tr_T), \log(tr_T)])$ from that of $\underline{\mu}(\mathbb{R}^+, \mathbb{R}^+, |\mathbf{x}| \leq K)$ (note that K is less than $\log(\delta tr_T)$). Now, for the second term on the right hand-side of (8.10)

$$\begin{aligned}
&\int_{(i-1)\Delta_n}^{i\Delta_n}\int_{\mathbb{R}} 1\left(|\Delta_i^n p(K)|\geq \log(tr_T)-x,\ x\in[\log(\delta tr_T),\log(tr_T)]\right)\mu(ds,dx)\\
&\quad\leq C|\Delta_i^n p(K)|^{1-\rho}\int_{(i-1)\Delta_n}^{i\Delta_n}\int_{\mathbb{R}}\frac{1}{|\log(tr_T)-x|^{1-\rho}}1\left(x\in[\log(\delta tr_T),\log(tr_T)]\right)\mu(ds,dx),
\end{aligned}$$

for arbitrary small $\rho > 0$. From here we can proceed exactly as in (8.11) upon using the following bound for any $1 \leq \alpha < 1/(1-\rho)$,

$$\begin{aligned}
&\mathbb{E}\left(\int_{(i-1)\Delta_n}^{i\Delta_n}\int_{\mathbb{R}}\frac{1}{|\log(tr_T)-x|^{1-\rho}}1\left(x\in[\log(\delta tr_T),\log(tr_T)]\right)\mu(ds,dx)\right)^{\alpha}\\
&\quad\leq C\Delta_n\int_{x\in[\log(\delta tr_T),\log(tr_T)]}\frac{1}{|\log(tr_T)-x|^{\alpha(1-\rho)}}\nu(x)dx\\
&\qquad+C\Delta_n^{\alpha}\left(\int_{x\in[\log(\delta tr_T),\log(tr_T)]}\frac{1}{|\log(tr_T)-x|^{1-\rho}}\nu(x)dx\right)^{\alpha},\\
&\quad\leq C\Delta_n\overline{\nu}^+(\log(\delta tr_T)),
\end{aligned}$$

where for the first inequality we made use of the Burkholder-Davis-Gundy inequality, and for the second the restriction that $\alpha < 1/(1-\rho)$ together with the fact that $\nu(x)$ is non-increasing in the tails.

Step 10. In this step we show

$$\mathbb{E}\left(\sum_{i=1}^{[T/\Delta_n]}\left|f_T(\Delta_i^n p) - \sum_{(i-1)\Delta_n\le s\le i\Delta_n} f_T(\Delta p_s)\right| 1_{\{A_i^c,\ B_i^c,\ C_i,\ D_i^c\}}\right)$$
$$\le C\Delta_n^{1/2-\iota}T(\overline{\nu}^+(\log(tr_T)))^{1-\iota},$$

for some arbitrary small $\iota > 0$. On the set $A_i^c \cap B_i^c \cap C_i \cap D_i^c$, there is exactly one jump of size above K in absolute value and its size must be in the interval $[\log(tr_T), \log(tr_T) - \log(\delta)]$. Then, using the fact that $|f_T(x)| \le C$ for $x \in [\log(tr_T), \log(tr_T) - \log(\delta)]$ (C depends on δ), we have

$$\left|f_T(\Delta_i^n p) - \sum_{(i-1)\Delta_n\le s\le i\Delta_n} f_T(\Delta p_s)\right| 1_{\{A_i^c,\ B_i^c,\ C_i,\ D_i^c\}}$$
$$\le C\int_{(i-1)\Delta_n}^{i\Delta_n}\int_{\mathbb{R}} 1\left(|\Delta_i^n p(K)| \ge x - \log(tr_T),\ x \in [\log(tr_T), \log(tr_T) - \log(\delta)]\right)\mu(ds, dx).$$

From here the result follows exactly as in Step 9.

Step 11. In the final step we show

$$\mathbb{E}\left(\sum_{i=1}^{[T/\Delta_n]}\left|f_T(\Delta_i^n p) - \sum_{(i-1)\Delta_n\le s\le i\Delta_n} f_T(\Delta p_s)\right| 1_{\{A_i^c,\ B_i^c,\ C_i,\ D_i\}}\right)$$
$$\le CT\Delta_n^{1/2-\iota}\left(\overline{\nu}^+(\log(tr_T))\right)^{1-\iota},$$

for some $\iota > 0$. On the set $A_i^c \cap B_i^c \cap C_i \cap D_i$, we have only one jump of p above $\log(tr_T)$. Therefore, since the function f_T is differentiable for values of the argument exceeding $\log(tr_T)$, a first-order Taylor expansion together with the boundedness of the first derivative of f_T yields

$$\left|f_T(\Delta_i^n p) - \sum_{(i-1)\Delta_n\le s\le i\Delta_n} f_T(\Delta p_s)\right| 1_{\{A_i^c,\ B_i^c,\ C_i,\ D_i\}} \le C|\Delta_i^n p(\log(tr_T)|1_{\{A_i^c,\ B_i^c,\ C_i,\ D_i\}}.$$

To continue further we introduce the following two sets,

$$R_T = \{x : |x| \ge \log(tr_T)\} \quad \text{and} \quad S_T = \{x : |x| \le \log(tr_T)\}.$$

Using the alternative representation of the jumps with respect to $\underline{\mu}$, we have

$$\mathbb{E}\left(\left|f_T(\Delta_i^n p) - \sum_{(i-1)\Delta_n\le s\le i\Delta_n} f_T(\Delta p_s)\right| 1_{\{A_i^c,\ B_i^c,\ C_i,\ D_i\}}\right)$$
$$\le C\mathbb{E}\left(1_{\{A_i^c,\ B_i^c,\ C_i,\ D_i\}}|\Delta_i^n p(\log(tr_T)|\int_{(i-1)\Delta_n}^{i\Delta_n}\int_{u<\varphi_{s-}^+}\int_{x\in R_T}\underline{\mu}(ds, du, dx)\right)$$
$$\le C\mathbb{E}\left(|\Delta_i^n p(\log(tr_T)|\int_{(i-1)\Delta_n}^{i\Delta_n}\int_{u<\Delta_n^{-\epsilon}}\int_{x\in R_T}\underline{\mu}(ds, du, dx)\right)$$
$$+ C\mathbb{E}\left(|\Delta_i^n p(\log(tr_T)|1_{\{\sup_{s\in[(i-1)\Delta_n, i\Delta_n]}\varphi_s^+ \ge \Delta_n^{-\epsilon}\}}\right).$$

Furthermore,

$$\mathbb{E}\left(|\Delta_i^n p(\log(tr_T)| \int_{(i-1)\Delta_n}^{i\Delta_n}\int_{u<\Delta_n^{-\epsilon}}\int_{x\in R_T}\underline{\mu}(ds,du,d\mathbf{x})\right) \leq I + II,$$

$$I = C\mathbb{E}\Bigg[\left(\int_{(i-1)\Delta_n}^{i\Delta_n}(|\alpha_s| + \varphi_{s-}^{+} + \varphi_{s-}^{-})ds + \int_{(i-1)\Delta_n}^{i\Delta_n}\sigma_s dW_s\right)\int_{(i-1)\Delta_n}^{i\Delta_n}\int_{u<\Delta_n^{-\epsilon}}\int_{x\in R_T}\underline{\mu}(ds,du,dx)\Bigg],$$

$$II = C\mathbb{E}\Bigg[\left|\int_{(i-1)\Delta_n}^{i\Delta_n}\int_{\mathbb{R}^+}\int_{\mathbf{x}\in S_T} x\left(1_{\{x<0,\ u<\varphi_{s-}^{-}\}} + 1_{\{x>0,\ u<\varphi_{s-}^{+}\}}\right)\underline{\tilde{\mu}}(ds,du,dx)\right| \times \int_{(i-1)\Delta_n}^{i\Delta_n}\int_{u<\Delta_n^{-\epsilon}}\int_{x\in R_T}\underline{\mu}(ds,du,d\mathbf{x})\Bigg].$$

By Holder's inequality and the integrability conditions,

$$I \leq C\Delta_n^{3/2-\iota}\left(\overline{\nu}^{+}(\log(tr_T))\right)^{1-\iota},$$

for some arbitrary small $\iota > 0$. For the term II, we can condition on the filtration generated by the measure $\underline{\mu}(\mathbb{R}^+, \mathbb{R}^+, x \in R_T)$, denoted with $\mathscr{F}^*$, and the fact that the homogenous measure on disjoint sets creates independent filtration, see e.g. Sato (1999), to get by an application of the Burkholder-Davis-Gundy inequality,

$$\begin{aligned} II \leq & C\mathbb{E}\Bigg[\sqrt{\int_{(i-1)\Delta_n}^{i\Delta_n}\mathbb{E}((\varphi_s^{+} + \varphi_s^{-})|\mathscr{F}^*)ds}\int_{(i-1)\Delta_n}^{i\Delta_n}\int_{u<\Delta_n^{-\epsilon}}\int_{x\in R_T}\underline{\mu}(ds,du,dx)\Bigg] \\ \leq & C\Delta_n^{1+1/2-2\epsilon}\overline{\nu}^{+}(\log(tr_T)) + II', \\ II' = & C\mathbb{E}\left(\sqrt{\mathbb{E}\left(\int_{(i-1)\Delta_n}^{i\Delta_n}(\varphi_s^{+} + \varphi_s^{-})ds\left(\int_{(i-1)\Delta_n}^{i\Delta_n}\int_{u<\Delta_n^{-\epsilon}}\int_{x\in R_T}\underline{\mu}(ds,du,dx)\right)^2\Bigg|\mathscr{F}^*\right)} \times 1_{\{\sup_{s\in[(i-1)\Delta_n, i\Delta_n]}\varphi_s^{+} \geq \Delta_n^{-\epsilon}\}}\right). \end{aligned}$$

An application of the Cauchy-Schwartz inequality then implies,

$$II' \leq C\Delta_n^{q/2}\sqrt{\mathbb{E}\left(\int_{(i-1)\Delta_n}^{i\Delta_n}(\varphi_s^{+} + \varphi_s^{-})ds\left(\int_{(i-1)\Delta_n}^{i\Delta_n}\int_{u<\Delta_n^{-\epsilon}}\int_{x\in R_T}\underline{\mu}(ds,du,dx)\right)^2\right)} \leq C\Delta_n^{q/2},$$

for any $q > 0$.

Step 12. Combining the results of Steps 7-11, we get (8.8) provided condition (8.9) holds. □

Lemma 4 *Suppose we observe the process p_t at the discrete times $0, \Delta_n, ..., n\Delta_n, ..., t, t+\Delta_n, ..., t+n\Delta_n, ...$, and assume that assumptions A1, with $\nu(x)$ nondecreasing for x sufficiently large, and A4 hold. Then, for the function f_T defined in Lemma 3,*

$$\frac{1}{\sqrt{N_T}}\sum_{t=1}^{T-1}\left(\sum_{i=1}^{n} f_T(\Delta_i^{n,t}p) - \sum_{s=t}^{t+1} f_T(\Delta p_s)\right) TV_{t-1}^n \xrightarrow{\mathbb{P}} 0, \quad \text{as } \Delta_n \downarrow 0,\ T \uparrow \infty, \tag{8.12}$$

provided condition (8.9) holds.

Proof: We can proceed exactly as in the proof of Lemma 2, using the result in Lemma 3. For this we only need that $\mathbb{E}|TV_{t-1}^n|^p < \infty$ for arbitrary $p > 0$. But, this follows from the fact that by successive conditioning and application of Holder's inequality, $\mathbb{E}\left(|\Delta_{i_1}^{n,t-1}p|^{q_1}...|\Delta_{i_k}^{n,t-1}p|^{q_k}\right) \leq C\Delta_n^{k-\epsilon}$, for k an integer, $\epsilon > 0$ arbitrary small, i_j for $j = 1, ..., k$, and $q_j \geq 2$ for $j = 1, ..., k$. $\square$

Proofs of Theorem 3 and Corollary 2. The proofs will follow from the proof of Theorem 1 if we can show

$$\sup_{\theta \in \Theta_T^l} \sqrt{M_T^+}||\hat{g}_T(\theta, tr_T) - \tilde{g}_T(\theta, tr_T)|| \xrightarrow{\mathbb{P}} 0,$$

where $\hat{g}_T(\theta, tr_T)$ is defined from $\tilde{g}_T(\theta, tr_T)$ by substituting $\int_t^{t+1}\int_{\psi^+(x)>tr_T}\phi_i^+(\psi(x) - tr_T, \theta^{(1)}, \theta^{(2)})\mu(ds, dx)$ with $\sum_{j=1}^n \phi_i^+(\psi(\Delta_j^{n,t}p) - tr_T, \theta^{(1)}, \theta^{(2)})$ for $i = 1, 2$ and $t = 0, ..., T-1$, and in the case of Corollary 1 $\int_{t-1}^t \sigma_s^2 ds$ is also replaced by TV_t^n for $t = 1, ..., T$. But, this result follows directly from Lemma 2 and Lemma 4, as the conditions on the function f_T are satisfied by our score functions φ_i^+ on the set $\theta \in \Theta_T^l$. $\square$

References

Aït-Sahalia, Y. and J. Jacod (2009a). Analyzing the Spectrum of Asset Returns: Jump and Volatility Components in High-Frequency Data. Working paper, Princeton University and Université de Paris-6.

Aït-Sahalia, Y. and J. Jacod (2009b). Estimating the Degree of Activity of Jumps in High Frequency Financial Data. *Annals of Statistics 37*, 2202–2244.

Aït-Sahalia, Y. and J. Jacod (2009c). Testing Whether Jumps have Finite or Infinite Activity. Working paper, Princeton University and Université de Paris-6.

Andersen, T., T. Bollerslev, and F. Diebold (2007). Roughing It Up: Including Jump Components in the Measurement, Modeling, and Forecasting of Return Volatility. *Review of Economics and Statistcis 89*, 701–720.

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

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

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

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

Barro, R. J. (2006). Rare Disasters and Asset Markets in the Twentieth Century. *Quarterly Journal of Economics 121*, 823–866.

Bollerslev, T. and V. Todorov (2009). Tails, Fears and Risk Premia. Working paper, Duke and Northwestern Universities.

Broadie, M., M. Chernov, and M. Johannes (2009). Understanding Index Options Returns. *Review of Financial Studies, forthcoming*.

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

Embrechts, P., C. Kluppelberg, and T. Mikosch (2001). *Modelling Extremal Events* (3rd ed.). Berlin: Springer-Verlag.

Eraker, B., M. Johannes, and N. Polson (2003). The Impact of Jumps in Volatility and Returns. *Journal of Finance 58*, 1269–1300.

Gabaix, X. (2010). Variable Rare Disasters: An Exactly Solved Framework for Ten Puzzles in Macro-Finance. Working paper, New York University.

Goldie, C. and R. Smith (1987). Slow Variation with Remainder: A Survey of the Theory and its Applications. *The Quarterly Journal of Mathematics 38*, 45–71.

Hansen, L. (1985). A Method for Calculating Bounds on the Asymptotic Covariance Matrices of Generalized Method of Moments Estimators. *Journal of Econometrics 30*, 203–238.

Hill, B. (1975). A Simple General Approach to Inference about the Tail of Distribution. *Annals of Statistics 12*, 1163–1174.

Huang, X. and G. Tauchen (2005). The Relative Contributions of Jumps to Total Variance. *Journal of Financial Econometrics 3*, 456–499.

Jacod, J. (2008). Asymptotic Properties of Power Variations and Associated Functionals of Semimartingales. *Stochastic Processes and their Applications 118*, 517–559.

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

Jacod, J. and V. Todorov (2010). Do Price and Volatility Jump Together? *Annals of Applied Probability, forthcoming*.

Leadbetter, M., G. Lindgren, and H. Rootzen (1983). *Extremes and Related Properties of Random Sequencies and Series*. New York: Springer-Verlag.

Leadbetter, M. and H. Rootzen (1988). Extremal Theory for Stochastic Processes. *Annals of Probability 16*, 431–478.

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

Merton, R. (1973). An Intertemporal Capital Asset Pricing Model. *Econometrica 41*, 867–887.

Merton, R. (1976). Option Pricing when Underlying Asset Returns are Discontinuous. *Journal of Financial Economics 3*, 125–144.

Nagel, S. and K. Singleton (2010). Estimation and Evaluation of Conditional Asset Pricing Models. *Journal of Finance, forthcoming*.

Prakasa Rao, B. (1988). Statistical Inference from Sampled Data for Stochastic Processes. In *Contemporary Mathematics*, Volume 80. Amer. Math. Soc., Providence.

Rosinski, J. and G. Samorodnitsky (1993). Distributions of Subadditive Functionals of Sample Paths of Infinite Divisible Processes. *Annals of Probability 21*, 996–1014.

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

Smith, R. (1987). Estimating Tails of Probability Distributions. *Annals of Statistics 15*, 1174–1207.

Todorov, V. (2009). Estimation of Coninuous-time Stochastic Volatility Models with Jumps using High-Frequency Data. *Journal of Econometrics 148*, 131–148.

Todorov, V. (2010). Variance Risk Premia Dynamics: The Role of Jumps. *Review of Financial Studies 23*, 345–383.

Todorov, V. and G. Tauchen (2010a). Activity Signature Functions for High-Frequency Data Analysis. *Journal of Econometrics 154*, 125–138.

Todorov, V. and G. Tauchen (2010b). Volatility Jumps. *Journal of Business and Economic Statistics, forthcoming*.

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.

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

Research Papers 2010

CREATES
Center for Research in Econometric Analysis of Time Series

2010-03: Jesper Rangvid, Maik Schmeling and Andreas Schrimpf: Cash Flow-Predictability: Still Going Strong

2010-04: Helle Bunzel and Walter Enders: The Taylor Rule and "Opportunistic" Monetary Policy

2010-05: Martin M. Andreasen: Non-linear DSGE Models and The Optimized Particle Filter

2010-06: Søren Johansen and Bent Nielsen: Discussion of The Forward Search: Theory and Data Analysis by Anthony C. Atkinson, Marco Riani, and Andrea Ceroli

2010-07: Giuseppe Cavaliere, Anders Rahbek and A.M.Robert Taylor: Bootstrap Sequential Determination of the Co-integration Rank in VAR Models

2010-08: Peter R. Hansen and Asger Lunde: Estimating the Persistence and the Autocorrelation Function of a Time Series that is Measured with Error

2010-09: Tom Engsted, Thomas Q. Pedersen and Carsten Tanggaard: Pitfalls in VAR based return decompositions: A clarification

2010-10: Torben G. Andersen and Luca Benzoni: Stochastic Volatility

2010-11: Torben B. Rasmussen: Affine Bond Pricing with a Mixture Distribution for Interest Rate Time-Series Dynamics

2010-12: Martin M. Andreasen and Bent Jesper Christensen: The SR Approach: a new Estimation Method for Non-Linear and Non-Gaussian Dynamic Term Structure Models

2010-13: Peter Reinhard Hansen, Zhuo (Albert) Huang and Howard Howan Shek: Realized GARCH: Complete Model of Returns and Realized Measures of Volatility

2010-14: Bent Jesper Christensen and Michel van der Wel: An Asset Pricing Approach to Testing General Term Structure Models including Heath-Jarrow-Morton Specifications and Affine Subclasses

2010-15: Nektarios Aslanidis and Charlotte Christiansen: Smooth Transition Patterns in the Realized Stock Bond Correlation

2010-16: Tim Bollerslev and Viktor Todorov: Estimation of Jump Tails