Estimation of Continuous-Time Stochastic Volatility Models with Jumps using High-Frequency Data*

Viktor Todorov†

September 23, 2008

Abstract

This paper proposes a method of inference for general stochastic volatility models containing price jumps. The estimation is based on treating realized multipower variation statistics calculated from high-frequency data as their unobservable (fill-in) asymptotic limits. The paper provides easy-to-check conditions under which the error in estimation resulting from this approximation is $o_p(1)$ and additional ones under which it is $o_p(1/\sqrt{T})$, where T is the number of days in the sample. Extensive Monte Carlo shows that the proposed estimation method works well in finite samples, provided asymptotic approximations are used. The estimation technique is applied to the estimation of two semiparametric models.

Keywords: Continuous-time stochastic volatility models, jump processes, method-of-moments estimation, realized multipower variation.

JEL classification: C51, C52, G12.

*I am grateful to Tim Bollerslev, Ron Gallant, Han Hong, Jean Jacod, Mark Podolskij, Peter Robinson (the editor), George Tauchen, an associate editor and two anonymous referees for many valuable comments. I would also like to thank Jean Jacod for helpful discussions about the material in Section 5.

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

1 Introduction

The availability of reliable financial high-frequency data over the last two decades has allowed a closer analysis of the finer movements of financial asset prices. Numerous non-parametric tests have been proposed for the presence of jumps on the "observed" path of increasingly finer-sampled processes, see Ait-Sahalia and Jacod (2008b), Barndorff-Nielsen and Shephard (2004, 2006a), Huang and Tauchen (2005), Jiang and Oomen (2005), Lee and Mykland (2007), Mancini (2001). These studies have documented non-trivial jump component in asset prices. Moreover, recent results in Ait-Sahalia and Jacod (2008a) and Todorov and Tauchen (2008) suggest that the jumps in asset prices are more "active" than the compound Poisson jumps typically used in parametric specifications.

This non-parametric evidence reinforces the importance of estimating stochastic volatility models with general jump specification, which is the goal of the current paper. Our estimation is based on high-frequency data. The idea of our method is quite simple. We aggregate the high-frequency data (on daily level) into so-called realized measures (also known as realized multipower variations), which measure continuous and discontinuous price variation, and treat them as their unobservable fill-in asymptotic limits in estimation. This provides a general way of estimating stochastic volatility models since the realized measures are model-free. The convenience of our method stems from the fact that the inference is done directly on latent quantities of the price process, i.e. the integrated volatility and the sum of squared jumps. This simplifies the estimation process significantly, as it circumvents the need to integrate out unobservable quantities. For example, we can make inference for the jumps regardless of how complicated the model for the stochastic volatility is.

The use of realized measures in the estimation induces error, the size of which depends on how frequently we sample during the day. We analyze this error for three particular statistics: Realized Variance, Realized Bipower Variation and Realized Tripower Variation. The first of these statistics converges to the Quadratic Variation of the price process, as we sample more frequently. The (fill-in) asymptotic limit of the other two statistics is the continuous part of the Quadratic Variation, which we call Integrated Variance. We derive conditions under which the error in the estimation resulting from treating Realized Variation, Realized Bipower Variation and Realized Tripower Variation as their fill-in asymptotic limits converges in probability to zero. We provide also stronger conditions under which this error is $o_p(1/\sqrt{T})$, where T is the number of days in the sample. In this last case the inference based on realized multipower variation statistics is asymptotically equivalent to the infeasible inference based on their asymptotic limits. Our analysis in this paper is done in the context of a general GMM estimator, but it can be trivially extended to the case of M-estimators.

We apply our estimation strategy both on simulated and real high-frequency data. Our Monte Carlo study shows that the Realized Tripower Variation does a better job in measuring the Integrated Variance than the Realized Bipower Variation. Also, for sampling frequencies of 5 minutes, the Monte Carlo results indicate that the feasible and infeasible inference can be made equivalent when certain asymptotic corrections are included. In an empirical application, based on 5-minute $S\&P$ 500 futures data, we find strong (semiparametric) evidence for presence of jumps in the stochastic volatility, which are related with the price jumps.

There are several examples in the literature of estimating stochastic volatility models where certain realized measures are treated as their asymptotic limits. Bollerslev and Zhou (2002) estimate affine jump-diffusion stochastic volatility models treating the Realized Variance as the unobserved Quadratic Variation. Corradi and Distaso (2006) provide theoretical justification for the GMM estimator, used in Bollerslev and Zhou (2002), in the case of no price jumps. They also robustify the results to the case when price jumps are present, but only jumps of finite activity are considered. In such a case Corradi and Distaso (2006) use the Realized Bipower Variation as a proxy for the unobservable Integrated Variance. Also, if jumps are present, Corradi and Distaso (2006) are interested in the inference of parameters controlling the Integrated Variance only.

To compare our results with the results in the related literature we outline the main features of this paper. They are: (i) estimation of parameters controlling both the continuous and the discontinuous component of the price is considered, (ii) the joint behavior of the Realized Variance and the Realized Tripower Variation (Realized Bipower Variation respectively) is analyzed, (iii) price jumps can be infinitely active and can exhibit arbitrary time-variation (e.g. time-varying intensity), (iv) general GMM estimator is considered.

Finally, the results in this paper are related with the literature on testing for jumps using high-frequency data, cited at the beginning of the introduction. The main difference between those studies and the current work is that the non-parametric tests for jumps are about properties of the *observed path*, while the current paper relies on long-span (in addition to fill-in) asymptotics to make inference about (semi)parametric models containing jumps.

The rest of the paper is organized as follows. Section 2 contains the theoretical part. In Section 3 we conduct an extensive Monte Carlo analysis of the estimation technique proposed in the paper. Section 4 contains an application to S&P 500 futures data. Section 5 points out how the results in the paper can be extended to situations where microstructure noise is present and concludes. The proof of all results are contained in an Appendix.

2 Theoretical Results

This section contains the theoretical part of the paper. We start with specifying the class of stochastic volatility models for which our analysis apply and then we define the model-free realized measures that we use in the estimation. We finish with deriving the asymptotic properties of the estimation based on the realized measures.

2.1 Data Generating Process and Realized Measures

Our analysis in this paper applies to models for the logarithmic asset price $p(t)$ given by

$$p(t) = p(0) + \int_0^t b(s)ds + \int_0^t \sigma_1(s)dW(s) + \int_0^t \int_E \phi(s,x)\tilde{\mu}(ds,dx), \qquad (1)$$

where $W(t)$ is a standard Brownian motion; μ is a Poisson random measure with compensator $\nu(ds, dx) = dsG(dx)$, where $G : E \rightarrow \mathbb{R}_+$ and E is some measurable space; $\tilde{\mu} := \mu - \nu$; $\phi : \mathbb{R}_+ \times E \rightarrow \mathbb{R}$; the processes $b(t)$ and $\sigma_1(t)$ and the stochastic function $\phi(s, x)$ are arbitrary[1]. This specification of the price process nests most of the models used in the literature. In particular, the price jumps are allowed to have time-variation of arbitrary form (including time-varying intensity, see e.g. Theorem 14.80 in Jacod (1979)).

In this paper our unit of measurement will be one day. In the (unrealistic) case of continuous price record, we can directly "observe" the quadratic variation and its continuous and discontinuous components. The Quadratic Variation (hereafter abbreviated as QV) over day t is given by

$$[p, p]_{(t,t+1]} = \int_t^{t+1} \sigma_1^2(s)ds + \int_t^{t+1} \int_E \phi^2(s, x)\mu(ds, dx), \tag{2}$$

while its continuous part, which we refer to as Integrated Variance (hereafter abbreviated as IV), is

$$IV_{(t,t+1]} := \int_t^{t+1} \sigma_1^2(s)ds. \tag{3}$$

In the continuous record case, we can easily estimate the parameters controlling $\sigma_1^2(t)$ and the price jumps by making direct inference on QV and IV. In practice, we observe the price only at discrete times and therefore such estimation is infeasible. However, if we have high-frequency observations of the price during the day, we can make the estimation feasible by substituting QV and IV with *realized measures* that proxy them, which in turn are computed from the high-frequency data.

Formally our setting is as follows. On each day t we observe the price at times $t, t + \delta, t + 2\delta, ...$ for a total of $M := \lfloor 1/\delta \rfloor$ high-frequency returns. If $r_a(t) := p(t + a) - p(t)$ for some $a > 0$ denotes the return over $(t, t + a]$, then our *feasible* estimate of QV is the Realized Variance (hereafter RV) defined as

$$RV_\delta(t) := \sum_{i=1}^{M} r_\delta^2(t + (i - 1)\delta). \tag{4}$$

[1]We (implicitly) assume that the integrals in (1) are well defined. Assumptions A1-A3 of the next subsection guarantee that.

For estimating IV we construct (and later compare) two alternative estimators - Realized Bipower Variation (hereafter BV) and Realized Tripower Variation (hereafter TV). We define them as

$$TV_\delta(t) := \mu_{2/3}^{-3} \sum_{i=3}^{M} |r_\delta(t+(i-3)\delta)|^{2/3} |r_\delta(t+(i-2)\delta)|^{2/3} |r_\delta(t+(i-1)\delta)|^{2/3}, \quad (5)$$

$$BV_\delta(t) := \mu_1^{-1} \sum_{i=2}^{M} |r_\delta(t+(i-2)\delta)| |r_\delta(t+(i-1)\delta)|, \quad (6)$$

where $\mu_a = \mathbb{E}(|u|^a)$ for $u \sim \mathscr{N}(0,1)$. RV, BV and TV are also called realized multipower variations (see e.g. Barndorff-Nielsen et al. (2005)). The usefulness of these realized measures comes from the fact that, as we sample more frequently, RV is consistent for QV, while BV and TV are consistent for IV, and the rate of convergence for the three realized measures is $\sqrt{\delta}$. Importantly, this result is *model-free*, i.e. it does not depend on the particular parametric specification of (1).

2.2 Inference Based on Realized Measures

We proceed with the inference which is based on the realized measures introduced in the previous subsection. The idea is to estimate the parameters controlling the stochastic volatility and the price jumps by matching moments of QV and IV. To make the estimation feasible, we substitute QV and IV in the GMM estimation with the model-free measures RV, TV and BV, constructed from the high-frequency data on each day. In this subsection we derive conditions under which the error in measuring QV and IV does not influence asymptotically our results.

The statistical setup is as follows. The data generating process for the price p is given by (1) and we have high-frequency observations of the price for a total of T days with M high-frequency observations each day. We proceed with defining formally the parameters, the data and the moments used in the estimation.

Definitions: (1) *θ is a vector of parameters that controls $\sigma_1^2(t)$ and the price jumps in (1).*

(2) *For an arbitrary day t, $z(t)$ is a vector consisting of $IV(t)$, $QV(t)$ and a finite number of their lags, while $\widehat{z}(t)$ is constructed from $z(t)$ by substituting QV with RV*

and IV with TV or BV.

(3) $m(z,\theta)$ *is a "moment function" vector;* $m_T(\theta) := \frac{1}{T}\sum_{t=1}^{T} m(z(t),\theta)$ *and* $\widehat{m}_T(\theta) := \frac{1}{T}\sum_{t=1}^{T} m(\widehat{z}(t),\theta)$*;* W_T *is a weighting matrix constructed using* $\{z(t)\}_{t=1,..,T}$*, which converges in probability to some positive definite matrix;* $\widehat{W_T}$ *is constructed from* W_T *by replacing* $\{z(t)\}_{t=1,..,T}$ *with* $\{\widehat{z}(t)\}_{t=1,..,T}$.

Since our goal is to be as general as possible, and therefore we did not specify a parametric model for $\sigma_1^2(t)$ and the price jumps, here we define θ as some set of parameters which control these processes. The only (obvious) requirement we (will) impose is that the set of moments used in the estimation identify θ. We refer to the empirical part of the paper for examples of θ and the moment function $m(\cdot,\cdot)$ in (semi)parametric applications of our theoretical results. We proceed with the assumptions. In what follows for an arbitrary matrix A we denote $||A|| := \sqrt{\mathrm{Trace}(A'A)}$.

A1. (a) *The process* $\xi(t) := \left(\int_t^{t+1}\sigma_1^2(s)ds, \int_t^{t+1}\int_E \phi^2(s,x)\mu(ds,dx)\right)$ *is stationary, ergodic and such that CLT holds for the sequence* $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\xi(t)-\mathbb{E}(\xi(t)))$ *as* $T\to\infty$.

(b) $\exists$ *unique* θ_0 *inside the compact parameter set* Θ *that solves* $m_0(\theta)=0$ *for* $m_0(\theta) := \mathbb{E}(m(z(t),\theta))$*.* $m_T(\theta)$ *converges to* $m_0(\theta)$ *in probability uniformly on* Θ.

(c) $\nabla_\theta m(\cdot,\theta)$ *exists and is continuous in* θ*.* $\nabla_\theta m_T(\theta)$ *converges in probability uniformly on* Θ*.* $\mathbb{E}\left(\nabla_\theta m(z(t),\theta_0)\right)$ *is of full column rank.*

A2. *There exist some real-valued processes* $\sigma_2(s)$ *and* $h(x)$ *such that for every* $s\in\mathbb{R}$ *and* $x\in E$ *we have* $|\phi(s,x)|\le|\sigma_2(s)||h(x)|$ *and*

$$\alpha := \inf\left\{\gamma\ge 0 : \int_E 1_{\{|h(x)|\le 1\}}|h(x)|^\gamma G(dx)<\infty\right\} < 4/5. \tag{7}$$

A3. $b(t)$ *is stationary and* $\mathbb{E}\,|b(t)|^p<\infty$ *for every* $p>0$*;* $\sigma_1(t)$ *is stationary and* $\mathbb{E}\,|\sigma_1(t)|^p<\infty$ *for every* $p>0$*; for* $\sigma_2(t)$ *of A2 assume that it is stationary and* $\mathbb{E}\,|\sigma_2(t)|^p<\infty$ *for every* $p>0$*; for* $h(x)$ *of A2 assume* $\int_{(-\epsilon,\epsilon)^c} e^{\lambda|h(x)|}G(dx)<\infty$ *for some* $\epsilon>0$ *and* $\lambda>0$.

A4. $||m(z+y,\theta)-m(z,\theta)||\le||C(\theta)||||P(z+y)-P(z)||$ *for every* z *and* y *and some matrix valued functions* $C(\cdot)$ *and* $P(\cdot)$ *such that* $P(z)$ *has polynomial growth.*

A5. $||\nabla_\theta m(z+y,\theta)-\nabla_\theta m(z,\theta)||\le||C(\theta)||||P(z+y)-P(z)||$ *for every* z *and* y *and* *some matrix valued functions $C(\cdot)$ and $P(\cdot)$ such that $P(z)$ has polynomial growth.*

A6. $\nabla_z m(z, \theta_0)$ *exists for every z, is continuous in z, and has polynomial growth in z.*

Assumption A1 provides standard conditions, which guarantee that a GMM estimator based on the sample moment conditions, $m_T(\theta)$, is consistent and asymptotically normal (see e.g. Newey and McFadden (1994) and Wooldridge (1994)). In a particular parametric application, part (a) of A1 can be directly verified using for example Theorem VIII.3.79 in Jacod and Shiryaev (2003) or using mixing conditions (e.g. the ones in Masuda (2004) for the Lévy-driven CARMA processes that we estimate in the empirical part[2]). Assumptions A2-A6 are additional assumptions which guarantee that the measurement error $\widehat{z}(t) - z(t)$ does not influence asymptotically the parameter estimation. A2 puts a constraint on the "activity" of the jumps in the price, which is measured by the (generalized) Blumenthal and Getoor (1961) index[3]. This assumption guarantees that TV (and BV) can disentangle the discontinuous part of the quadratic variation "sufficiently" fast (as $\delta \to 0$). Assumption A3 contains the integrability conditions. Some of these conditions could be further relaxed for special cases of (1). We will make use of A3 in proving a uniform integrability result for the centered (and scaled) TV and BV in Lemma 2 in the Appendix. Finally, assumptions A4-A6 concern the moment function $m(\cdot, \cdot)$. They are satisfied if for example $m(\cdot, \cdot)$ is polynomial in z.

Now we are ready to state our main result about the inference based on the realized multipower variation statistics RV, TV and BV. It is given in the following Theorem.

Theorem 1 *Set* $\widehat{\theta}_{nf} := \underset{\theta \in \Theta}{argmin}\ m_T(\theta)' W_T m_T(\theta)$ *and* $\widehat{\theta}_f := \underset{\theta \in \Theta}{argmin}\ \widehat{m}_T(\theta)' \widehat{W}_T \widehat{m}_T(\theta)$.

(a) Suppose **A1-A4** *hold. Then for $T \to \infty$ and $\delta \to 0$ we have*

$$\widehat{\theta}_f \xrightarrow{p} \theta_0. \tag{8}$$

(b) Suppose **A1-A6** *hold. Then if $T \to \infty$, $\delta \to 0$ and $T\delta^{1-\epsilon} \to 0$ for some $\epsilon > 0$*

[2] See also Masuda (2007) (and references therein) for bounds of mixing coefficients of general jump-diffusion processes.

[3] This assumption allows for jumps of infinite activity, but restricts them to be of finite variation.

we have

$$\sqrt{T}\left(\widehat{\theta}_f - \theta_0\right) \xrightarrow{d} \mathcal{N}(\mathbf{0}, Avar(\widehat{\theta}_{nf})). \tag{9}$$

The estimator $\widehat{\theta}_{nf}$, defined in Theorem 1, is not feasible because it is based on (the unobservable) QV and IV. Instead, the econometrician can plug in the GMM the feasible but noisy measures of QV and IV, RV and BV (or TV), and estimate $\widehat{\theta}_f$. Theorem 1 provides conditions under which we do as good as $\widehat{\theta}_{nf}$ in estimation (asymptotically), by using the feasible $\widehat{\theta}_f$. For the consistency of $\widehat{\theta}_f$ we do not need a condition for the relative speed at which $T \to \infty$ and $\delta \to 0$. However, for the asymptotic normality result in part (b) of the Theorem we need to know the relative speed at which $T \to \infty$ and $\delta \to 0$. Intuitively, we need the number of high-frequency observations in each of the days in the sample to go to zero slightly faster than the time span T goes to infinity. We can slightly strengthen the result in part (b) of Theorem 1 by requiring the weaker assumption $T\delta \to 0$ for the following special case of (1)

$$dp(t) = b(t)dt + \sqrt{V^{\perp}(t) + V^{\parallel}(t)}dW(t) + \int_{\mathbb{R}_0^n} h(\mathbf{x})\tilde{\mu}(dt, d\mathbf{x}), \tag{10}$$

$$V^{\parallel}(t) = \int_{-\infty}^{t}\int_{\mathbb{R}_0^n} g(t-s)k(\mathbf{x})\mu(ds, d\mathbf{x}) \text{ and } V^{\perp} \perp \mu \ (V^{\perp} \text{ is independent from } \mu), \tag{11}$$

where μ is a n-dimensional Poisson random measure with compensator $\nu(ds, dx) = dsG(dx)$, $G : \mathbb{R}_0^n \to \mathbb{R}_+$; $h : \mathbb{R}_0^n \to \mathbb{R}_0$, $k : \mathbb{R}_0^n \to \mathbb{R}_+$; $g : \mathbb{R}_+ \to \mathbb{R}_+$. This specification is still quite general and can nest many popular models in the literature. For this model, A3 can be given in slightly more primitive form stated as assumption A3$'$.

A3$'$. *For the model (10)-(11) assume $b(t)$ is stationary and $\mathbb{E}\,|b(t)|^p < \infty$; $\int_0^\infty |g(s)|^p ds < \infty$ for every $p > 0$; $\int_{\mathbb{R}_0^n} 1_{|h(\mathbf{x})|>\epsilon} e^{\lambda|h(\mathbf{x})|} G(d\mathbf{x}) < \infty$ and $\int_{\mathbb{R}_0^n} 1_{k(\mathbf{x})>\epsilon} e^{\lambda k(\mathbf{x})} G(d\mathbf{x}) < \infty$ for some $\epsilon > 0$ and $\lambda > 0$; $g(\cdot)$ is bounded around zero; $V^{\perp}(t)$ is stationary and $\mathbb{E}\left|V^{\perp}(t)\right|^p < \infty$ for every $p > 0$.*

The next Corollary strengthens the result about the asymptotic equivalence of $\widehat{\theta}_f$ and $\widehat{\theta}_{nf}$ of Theorem 1 when the data generating process for $p(t)$ is given by (10)-(11).

Corollary 1 *Suppose the data generating process for p is the model (10)-(11) and assume that* **A1-A2**, A3$'$ *and* **A4-A6** *hold. In addition, suppose that $z(t)$ does not* *include $BV(t)$ or its lags. Then, if $T \to \infty$, $\delta \to 0$ and $T\delta \to 0$, we have the result (9) in Theorem 1.*

Note that even in this special case we need to exclude BV, however TV is present. This suggests that TV will perform better in estimating IV, a fact that will be later confirmed in the Monte Carlo. We mention that even the condition $T\delta \to 0$ can be weakened but under some restrictive assumptions, e.g. when we use only RV and a specific set of moment conditions (see for example Corradi and Distaso (2006)).

Remark. *In the analysis here, we (implicitly) assumed that the moment conditions can be computed directly. However, in some applications this might not be the case, because moments of QV and IV are not known in closed-form. In such scenarios we will need to compute $m_T(\theta)$, and hence $\widehat{m}_T(\theta)$, via simulation. All of our preceding results can be extended to such situations as in Corradi and Distaso (2006)*[4].

3 Monte Carlo Study

For empirical applications it is important to know whether, for feasible sampling frequencies, QV and IV are reasonably approximated by their realized counterparts. Our goal in this section is to compare the finite sample performance of a GMM estimator based on RV and BV or TV with the infeasible one based on QV and IV.

3.1 Monte Carlo Setup

We work with the following stochastic volatility model

$$dp(t) = \sigma_1(t)dW(t) + \int_{\mathbb{R}_0} k_1 x \tilde{\mu}_1(dt, dx),\ \sigma_1^2(t) = \int_{-\infty}^{t} \int_{\mathbb{R}_0} e^{-\rho(t-s)} k_2 x \mu_2(ds, dx), \quad (12)$$

where k_1 and k_2 are constants and μ_1 and μ_2 are independent Poisson random measures. This is the Non-Gaussian OU model of Barndorff-Nielsen and Shep-

[4]Corradi and Distaso (2006) analyze only the case when IV is used in the estimation. However, note that QV-IV is just a sum of squared jumps, i.e. a finite variation jump process whose moments we are interested in evaluating at a fixed point. Hence, for example, the case of Lévy price jumps is absolutely trivial, since in this case the daily jumps are i.i.d. For certain time-nonhomogeneous jump specifications, where Euler discretisation is needed, the results in Protter and Talay (1997), Jacod (2004) and Bruti-Liberati and Platen (2007) can be applied to prove the convergence of the simulated moment conditions to the true ones exactly as the proof of Theorem 2 of Corradi and Distaso (2006).

hard (2001). The compensators for the price and variance jumps are respectively $\nu_1(dt, dx) = dtG_1(dx)$ and $\nu_2(dt, dx) = dtG_2(dx)$ where

$$G_1(dx) = c_p \frac{e^{-\frac{x^2}{2\sigma_p^2}}}{\sqrt{2\pi}\sigma_p} dx \quad \text{or} \quad G_1(dx) = c_p \frac{e^{-\lambda_p|x|}}{|x|^{1+\alpha_p}} dx, \quad c_p > 0, \lambda_p > 0, \alpha_p \in [0, 2), \tag{13}$$

$$G_2(dx) = c_v \frac{e^{-\lambda_v x}}{x^{1+\alpha_v}} I(x > 0) dx, c_v > 0, \lambda_v > 0, \alpha_v \in [0, 1). \tag{14}$$

The first compensator in (13) is of compound Poisson process with normally distributed mean-zero jumps, while the second one is of a symmetric tempered stable process (Rosiński (2007)). G_2 is the compensator of a tempered stable subordinator.

We note that the compound Poisson process has Blumenthal-Getoor index of zero, while the Blumenthal-Getoor index of the tempered stable process is equal to the parameter α_p. In the Monte Carlo study we look at the following values for α_p: 0, 0.1 and 0.5. Previous work (e.g. Bollerslev and Zhou (2002), Corradi and Distaso (2006)) has considered only the case of compound Poisson price jumps, thus we use this simulation scenario as a benchmark to compare with the more general (and active) jump specifications that our estimation technique encompasses.

In all cases we set $\rho = 0.07$, corresponding to half-life of a shock to the stochastic variance of approximately 10 days. The parameters of the jumps in the price and in the variance are chosen so that the variance of the continuous price component is always 1.0, while that of the price jumps is always 0.1. The resulting contribution of the price jumps in the total price variation is within the range of values found in high-frequency financial data. The parameter values in all simulation scenarios are given in Appendix F.

Finally, we work with numbers of intraday observations which are practically feasible: $M = 100$ and $M = 300$. The two choices correspond to sampling approximately every 5 minutes in a 6.5 and 24 hour trading days. The number of days in each Monte Carlo draw is set to 3000 days which corresponds to approximately 12 years of daily data and the number of Monte Carlo replications is 1000.

3.2 Details on the Estimators

The jumps in the price and the variance of the simulated model (12) are quite richly parameterized. This is important for the Monte Carlo as it gives us the flexibility to investigate different scenarios. Estimation of all the parameters of the price and variance jumps will require an efficient estimator which takes into account the particular parametric specification of the jumps. Here we adopt a simpler approach. Instead of estimating all the parameters of the price and variance jumps we estimate only cumulants of these processes. Such an approach is used also in Barndorff-Nielsen and Shephard (2006b) in the context of quasi-maximum likelihood. This estimator is convenient for comparing different semiparametric classes of stochastic volatility models with minimal distributional assumptions, as illustrated later in the empirical part. The parameters which are estimated are

$$\rho, \quad \frac{1}{\rho}\int_{\mathbb{R}_0} k_2 x G_2(dx), \quad \int_{\mathbb{R}_0} k_1^2 x^2 G_1(dx), \quad \sqrt{\int_{\mathbb{R}_0} k_2^2 x^2 G_2(dx)}, \quad \sqrt{\int_{\mathbb{R}_0} k_1^4 x^4 G_1(dx)}.$$

The second and the third parameters correspond to the means of the stochastic variance, $\sigma_1^2(t)$, and the squared price jumps respectively. The last two parameters are the standard deviations of the variance jumps and the squared price jumps respectively. Altogether we have 5 parameters. We use the following moments for their estimation: (1) mean and variance of IV, (2) mean and variance of QV and (3) autocorrelation of IV for lags 1, 3 and 6. Thus, we have 7 moment conditions resulting in 2 overidentifying restrictions. We use optimal weighting matrix in the GMM. For its calculation we compute a HAC estimator with Parzen kernel and lag-length of 80.

We compare the performance of 5 different estimators. The first one is the infeasible one, based on QV and IV. For the other estimators QV is replaced with RV and IV with BV or TV, for the two different cases of number of intraday observations $M = 100$ and $M = 300$. We note that the case of using RV and BV is similar to the estimation in Corradi and Distaso (2006)[5].

[5]However, Corradi and Distaso (2006) consider only estimation of parameters controlling $\sigma_1^2(t)$ and only in the context of finite activity price jumps (as in the first simulation scenario here).

3.3 Monte Carlo Results

The results from the Monte Carlo are reported in Tables 1-8 and Figures 1-4. We summarize them as follows.

Estimation based on QV and IV. The results for this case are reported in Panels A in Tables 1-8. As expected, the hardest parameters to estimate are the persistence parameter ρ and the standard deviations of the variance and squared price jumps. For the standard deviation parameters the mean and the median are slightly different indicating departure from normality. These two parameter estimates are also slightly downward biased.

Estimation based on RV and BV. We start with the case when the number of intraday observations is $M = 100$. Panels B in Tables 1-8 report the Monte Carlo results for this estimator. The first thing to note is that almost all parameter estimates have systematic biases. The only exception is the estimate of ρ. The parameter estimate of the mean of $\sigma_1^2(t)$ is significantly upward biased, while the estimate of the variance of the price jumps is substantially downward biased. Thus, the contribution of the price jumps in the total price variation is underestimated. Also, the estimate of the standard deviation of the variance jumps is significantly upward biased. On the other hand, the estimate of the standard deviation of the squared price jumps is downward biased. The RMSE of the parameter estimates are significantly higher than the corresponding ones for the infeasible estimator. The J test indicates significant misspecification. The situation improves when the number of intraday observations increases from $M = 100$ to $M = 300$. The Monte Carlo results when $M = 300$ are reported in Panels C of Tables 1-8. In the case of $M = 300$ the bias in the estimates is reduced, the only exception being the estimate of ρ where a downward bias appears[6]. Also, the RMSE-s of the estimates are substantially reduced, though still higher (in most cases) than the corresponding ones for the infeasible estimator. Finally, the values of the J test are significantly smaller (as compared with the case $M = 100$), but the test still signals model misspecificarion.

Estimation based on RV and TV. Panels D and E in Tables 1-8 report the results

[6] However, the RMSE of ρ for $M = 300$ is still lower than for the case $M = 100$.

from the Monte Carlo when RV and TV are used in the estimation for $M = 100$ and $M = 300$ respectively. Comparing the estimator based on RV and BV with the estimator based on RV and TV we see that, for all different cases, the estimator based on RV and TV does a much better job. First, when RV and TV are used in the GMM, the estimate of the mean of $\sigma_1^2(t)$ is still upward biased but the bias is smaller. Second, the estimate of the variance of the price jumps when RV and TV are used is practically unbiased, while when RV and BV are used this estimate is significantly downward biased. Third, the downward bias in the estimate of the standard deviation of the squared price jumps is reduced when BV is replaced with TV (and even disappears in the compound Poisson case). Finally, the estimates of these three parameters have smaller RMSE-s. The other two parameters (ρ and the standard deviation of the variance jumps) have comparable performance in the cases when either TV or BV is used to substitute for the unobservable IV. Interestingly, TV has superior performance over BV in estimation even when the price jumps are finitely active with low frequency of arrival, see Tables 1-2 [7]. Increasing the sampling frequency has a similar effect on the properties of the estimates as when RV and BV are used. Finally in both cases, $M = 100$ and $M = 300$, when RV and TV are used in the GMM the J test signals model misspecification. The values of the J test in the case of $M = 300$ are smaller than in the case $M = 100$, but still relatively high.

Overall, we conclude that among the feasible estimators the best performing are the ones in which TV is used as an estimate of the infeasible IV. Therefore, in the further analysis we focus only on estimators based on RV and TV [8]. To improve the performance of the feasible estimators we use the following CLT result for TV and RV (which holds for the general model (1) under the assumptions of the previous section), see Barndorff-Nielsen et al. (2006) and Jacod (2006)

[7] When no price jumps on a given day, BV is more efficient (asymptotically) than TV. In the Monte Carlo scenario with compound Poisson jumps, we have on average only 300 out of the 3000 days in the sample with jumps.

[8] Another reason not to consider BV is that the presence of price jumps affects a CLT for it derived in the continuous price case; see Barndorff-Nielsen et al. (2006).

$$\delta^{-1/2}\left(TV_\delta(t) - \int_t^{t+1} \sigma_1^2(u)du\right) \xrightarrow{law} \sqrt{2}\int_t^{t+1} \sigma_1^2(u)d\overline{W}_1(u) + \sqrt{1.0613}\int_t^{t+1} \sigma_1^2(u)d\overline{W}_2(u),$$
$$\delta^{-1/2}\left(RV_\delta(t) - \int_t^{t+1} \sigma_1^2(s)ds - \int_t^{t+1}\int_E \phi(s,x)\mu(ds,dx)\right) \xrightarrow{law} L_1(t) + L_2(t), \tag{15}$$
$$L_1(t) = \int_t^{t+1} \sqrt{2}\sigma_1^2(u)d\overline{W}_1(u) \quad \text{and} \quad L_2(t) = \sum_{t<s\le t+1} 2\Delta p_s\left(\sqrt{\kappa_s}u_s\sigma_1(s-) + \sqrt{1-\kappa_s}u_s'\sigma_1(s)\right),$$

where $\overline{W} = (\overline{W}_1, \overline{W}_2)$ is a standard Wiener process, $\kappa_s \sim U[0,1]$, $u_s \sim \mathscr{N}(0,1)$, $u_s' \sim \mathscr{N}(0,1)$. Furthermore the process $\overline{W}$ and the sequences (κ_s), (u_s), (u_s') are independent of each other, are defined on an extension of the original probability space and are independent from the process p. The above CLT results suggest using the following approximations for the variance of TV and RV in the estimation of (12)

$$\mathrm{Var}\left(TV_\delta(t)\right) \approx \mathrm{Var}\left(IV(t)\right) + \frac{3.0613}{M}\mathbb{E}\left(\int_t^{t+1} \sigma_1^4(u)du\right), \tag{16}$$

$$\mathrm{Var}\left(RV_\delta(t)\right) \approx \mathrm{Var}\left(QV(t)\right) + \frac{2}{M}\mathbb{E}\left(\int_t^{t+1} \sigma_1^4(u)du\right) + \frac{1}{M}\int_{\mathbb{R}_0} k_1^2 x^2 G_1(dx)\mathbb{E}\left(\int_t^{t+1} \sigma_1^2(s)ds\right). \tag{17}$$

That is, in addition to the variance of IV and RV respectively, we add correction terms of order $O(\delta)$ in approximating the variance of TV and RV. These correction terms will have no asymptotic effect in the estimation (provided the condition $T\delta^{1-\epsilon} \to 0$ in Theorem 1 holds), but can lead to finite sample improvements. We note that the CLT results suggest no correction terms for the mean and autocovariance of TV and RV, i.e. we still use the corresponding moments of their infeasible limits. The estimation results with the corrections in the variance of RV and TV are reported in Panels F and G of Tables 1-8. We can see that all biases in the parameter estimates are practically eliminated. Moreover, the RMSE-s for all parameter estimates are comparable with the ones of the infeasible estimation. Importantly, this holds true both for $M = 100$ and $M = 300$. Turning to the J tests, we see that they do not signal model misspecification anymore. When $M = 100$ there is very slight overrejection, but it disappears for $M = 300$.

4 Empirical Application

This section applies the general result of Section 2 to the estimation of two semi-parametric stochastic volatility models. We first describe the data used in the study and specify the set of moments used in the estimation. Following that we define the models that are estimated and briefly discuss the estimation results.

4.1 Data and Moments used in the Estimation

For the empirical application we use continuously-compounded five-minute returns on the S&P 500 futures contract. The data spans the period from January 2, 1990, till November 29, 2002. The total number of days in the data set is 3256, each of which consists of 80 five-minute continuously compounded returns covering the day trading session from 9:30am till 4:15am. For each day we calculate RV and TV, using the high-frequency returns over that day, as given in equations (4) and (5).

We estimate the models using the technique of Section 2. In addition to RV and TV, we compute from the high-frequency data on each day the Realized Forth Variation (hereafter abbreviated as FV), which we define as

$$FV_\delta(t) = \sum_{i=1}^{M} r_\delta^4(t + (i-1)\delta).$$

This statistic is useful since, for δ close to zero, it is capturing the sum of the price jumps raised to the power four. The moment conditions used in the estimation of the models are: mean and variance of IV; autocorrelation of IV; mean and variance of QV; mean of FV. The infeasible QV and IV are replaced with RV and TV. For the autocorrelation in IV, we use lags 1, 3 and 6 as well as the average autocorrrelation for lags $11-20$, $21-30$ and $31-40$. Altogether we end up with 11 moment conditions. To improve the performance, as discussed in Section 3, we use the asymptotic approximations in (16) and (17)[9]. For the two models we are going to estimate, the moments used in the estimation are available in closed-form and can be found in Todorov (2008).

[9]For the approximation of RV we have an additional term reflecting the dependence of the price and variance jumps. This term is easy to derive for the different models estimated here using the CLT result in (15).

4.2 Model Specification

In this subsection we define the two models that we estimate in the paper. Both of the models are nested in the general stochastic volatility model defined in (1) as well as in its special case given in (10)-(11). Common feature for the estimated models is that the jumps in the price are time-homogeneous. The two models differ in their specification of $\sigma_1^2(t)$.

Two-factor Affine Jump-Diffusion Model. This model falls in the widely-used affine jump-diffusion models of Duffie et al. (2000). Distinctive feature of the model is that the stochastic variance of the continuous martingale is diffusive, i.e. it has no jumps. The model dynamics is given with the following equations

$$dp(t) = \bar{b}dt + \sqrt{V_1(t) + V_2(t)}dW(t) + \int_{\mathbb{R}_0^n} h(\mathbf{x})\tilde{\mu}(dt, d\mathbf{x}), \tag{18}$$

$$dV_i(t) = \kappa_i(\overline{V}_i - V_i(t))dt + \sigma_{iv}\sqrt{V_i(t)}dB_i(t), \qquad i = 1, 2, \tag{19}$$

where $\bar{b}$ is some constant and $B_1(t)$ and $B_2(t)$ are two independent standard Brownian motions (they might be correlated with $W(t)$). $h : \mathbb{R}_0^n \to \mathbb{R}_0$ and μ is homogeneous Poisson measure with compensator $\nu(ds, \mathbf{x}) = dsG(d\mathbf{x})$, where $G : \mathbb{R}_0^n \to \mathbb{R}_+$. To avoid identification problems in the estimation, we make the following reparametrization. We set $\zeta := \overline{V}_1 + \overline{V}_2$ and $\overline{\sigma}_i := \sigma_{iv}\sqrt{\frac{\overline{V}_i}{2\kappa_i}}$ for $i = 1, 2$. $\overline{\sigma}_i^2$ is the variance of $V_i(t)$. Therefore, the stochastic volatility parameters we estimate are ζ and $\kappa_i, \overline{\sigma}_i$ for $i = 1, 2$.

The jumps in the price are Lévy and for the estimation we do not specify them parametrically. Instead, as in the Monte Carlo study, we model the second and the forth cumulants which are needed for the GMM-type estimator used here (with moments specified in the previous subsection), i.e. we estimate as free the nonnegative parameters $\int_{\mathbb{R}_0^n} h^2(\mathbf{x})G(d\mathbf{x}) \geq 0$ and $\int_{\mathbb{R}_0^n} h^4(\mathbf{x})G(d\mathbf{x}) \geq 0$.

CARMA(2,1)-Jump-Driven SV Model[10]. This model was analyzed in Brockwell (2001a) and Todorov and Tauchen (2006) among others. It has the distinctive feature that the stochastic variance, $\sigma_1^2(t)$, is solely driven by jumps. The CARMA(2,1) kernel

[10]CARMA stands for continuous autoregressive moving-average.

(for definition and properties of CARMA models see Brockwell (2001b)) implies the same autocorrelation of IV as in the two-factor affine jump-diffusion model. The model dynamics is given by

$$dp(t) = \bar{b}dt + \sigma_1(t)dW(t) + \int_{\mathbb{R}_0^n} h(\mathbf{x})\tilde{\mu}(dt, d\mathbf{x}), \tag{20}$$

$$\sigma_1^2(t) = \int_{-\infty}^{t} \int_{\mathbb{R}_0^n} g(t-s)k(\mathbf{x})\mu(ds, d\mathbf{x}),\ g(u) = \frac{b_0+\rho_1}{\rho_1-\rho_2}e^{\rho_1 u} + \frac{b_0+\rho_2}{\rho_2-\rho_1}e^{\rho_2 u},\ u \geq 0, \tag{21}$$

where $\bar{b}$ is some constant; $h : \mathbb{R}_0^n \to \mathbb{R}_0$, $k : \mathbb{R}_0^n \to \mathbb{R}_+$, μ is a homogeneous Poisson measure with compensator $\nu(ds, d\mathbf{x}) = dsG(d\mathbf{x})$. $g(\cdot)$ in equation (21) is the CARMA(2,1) kernel and hence the name of the model.

The jumps in the price and the variance in the stochastic volatility model are Lévy jumps. In the estimation of the CARMA(2,1)-jump-driven stochastic volatility model we do not parametrize the Poisson measure μ, neither we specify the functions $h(\cdot)$ and $k(\cdot)$ which determine the jump size. Instead, we estimate only the cumulants associated with the jumps in the price and the variance which are needed for the estimation of the model by the GMM estimator outlined in the previous subsection. These cumulants are $\int_{\mathbb{R}_0^n} h^2(\mathbf{x})G(d\mathbf{x})$, $\int_{\mathbb{R}_0^n} h^4(\mathbf{x})G(d\mathbf{x})$, $\int_{\mathbb{R}_0^n} k(\mathbf{x})G(d\mathbf{x})$, $\int_{\mathbb{R}_0^n} k^2(\mathbf{x})G(d\mathbf{x})$ and $\int_{\mathbb{R}_0^n} h^2(\mathbf{x})k(\mathbf{x})G(d\mathbf{x})$. The cumulants are treated as free (nonnegative) parameters in the estimation. However, to guarantee that there exists a Poisson measure μ and functions $h(\cdot)$ and $k(\cdot)$ such that the corresponding cumulants are equal to the estimated parameters, we restrict $\int_{\mathbb{R}_0^n} h^2(\mathbf{x})k(\mathbf{x})G(d\mathbf{x}) \leq \sqrt{\int_{\mathbb{R}_0^n} k^2(\mathbf{x})G(d\mathbf{x}) \int_{\mathbb{R}_0^n} h^4(\mathbf{x})G(d\mathbf{x})}$.

4.3 Estimation Results

The results from the estimation of the two models are reported in Table 9. Here we provide discussion of these results. First, one-factor restrictions of the current models are strongly rejected. Second, when $\sigma_1^2(t)$ is modelled with a square-root process, the fit is relatively bad. The reason for this is that the square-root process cannot generate enough volatility in IV to match the observed one. While this fact does not necessarily imply that $\sigma_1^2(t)$ contains jumps (since our conclusion is solely based on distributional properties of IV), it does indicate that models with jumps in the stochastic variance are more plausible and will provide a better fit. This is exactly the case here as we can see comparing Panel A and Panel B of Table 9.

We analyze closer the best performing model, i.e. the CARMA(2,1)-jump-driven SV model. First, one of the estimated autoregressive roots corresponds to a persistent factor in the volatility (with half-life of 18 days) while the other one corresponds to a quickly mean-reverting factor (with half-life of half day). These results are in line with many other studies estimating two-factor type stochastic volatility models. Second, the results show that the price has a non-trivial jump component. The price jumps contribute around 15% in the total price variation, which is in line with the non-parametric results of Barndorff-Nielsen and Shephard (2006a), Andersen et al. (2007) and Huang and Tauchen (2005). We proceed with analysis of the dependence structure of the jumps. The value of $\int_{\mathbb{R}_0^n} h^2(\mathbf{x})k(\mathbf{x})G(d\mathbf{x})$ indicates whether there is a relationship between the jumps in the stochastic variance, $\sigma_1^2(t)$, and the price jumps. If the value of the last integral is equal to zero this will imply independence, while positive value corresponds to dependence. The t-statistic associated with $\int_{\mathbb{R}_0^n} h^2(\mathbf{x})k(\mathbf{x})G(d\mathbf{x})$ is 3.87, which is well above the critical value of 1.65. This indicates dependence between price and variance jumps.

The simplest way of modeling the dependence between the jumps is to set the jumps in the variance proportional to the jumps in the price, i.e. to set the jump functions as $k(\mathbf{x}) \propto h(\mathbf{x})$. However, such modeling of the dependence is too restrictive since it implies that price jumps are of the same sign. Another way to model the jump dependence is to set the jumps in the variance proportional to the jumps in the price, i.e. to set $k(\mathbf{x}) \propto h^2(\mathbf{x})$. Such modeling has analogy with the GARCH models in discrete time and was analyzed in Todorov (2008). This dependence structure does not restrict the sign of the price jumps. The hypothesis that the variance jumps are proportional to the squared price jumps has a testable implication. Mainly,

$$\sqrt{\int_{\mathbb{R}_0^n} k^2(\mathbf{x})G(d\mathbf{x}) \int_{\mathbb{R}_0^n} h^4(\mathbf{x})G(d\mathbf{x})} \equiv \int_{\mathbb{R}_0^n} h^2(\mathbf{x})k(\mathbf{x})G(d\mathbf{x}).$$

The three integrals in the above equality are estimated as free parameters and therefore we can test whether they satisfy the above restriction. We note that this equality implies perfect correlation between the squared price jumps and the variance jumps. The t-statistic associated with the restriction is 2.45 with corresponding p-value of 0.0071. This implies that the hypothesis that the variance jumps are proportional to the squared price jumps can be rejected. Thus, we can conclude that neither of the two extreme cases of dependence hold. That is, the squared price jumps and the variance jumps seem to be dependent, but this dependence is not perfect. Therefore we need richer parametrization of the dependence between the jumps.

5 Concluding Remarks and Extension to the Case of Noisy Observations

This paper proposes and implements inference for continuous-time stochastic volatility models, based on matching moments of realized measures of continuous and discontinuous QV, constructed from high-frequency data. We derive conditions under which the substitution of QV with RV and IV with BV or TV does not influence asymptotically the GMM estimation. We conclude with a brief discussion on how to extend the results in this paper to situations where the data is observed with noise.

This is well-known to be the case for financial data sampled at very high frequencies. The so-called microstructure noise at such frequencies renders our realized measures RV, BV and TV, and hence the estimation based on them, inconsistent. One way to "deal" with microstructure noise is to ignore it and sample not so frequently, e.g. at 5 minutes, the idea being that at such frequencies the impact of the microstructure noise is negligible. This approach has been followed in most studies that deal with jumps using high-frequency data and we also adopted it in our empirical application. However, this leads to discarding a lot of (potentially) useful data. If we want to apply our results to very high-frequencies we need realized measures of QV and IV that are robust to microstructure noise. Formally, instead of observing the "efficient" log-price $p(t)$ we now assume that we observe $\widetilde{p}(t) = p(t) + \varepsilon(t)$, where $\varepsilon(t)$ is i.i.d. noise with finite moments[11]. *When $p(t)$ contains no jumps*, estimators

[11]For $p(t)$ we continue to assume that A1-A6 hold. Also, some of the results in this section can be further generalized to non-i.i.d. noise specifications, see Jacod et al. (2007).

of QV robust to noise have been derived already, see Barndorff-Nielsen et al. (2006), Zhang et al. (2005) and Zhang (2006). These estimators can be used in estimation of models containing no jumps, see e.g. Corradi and Distaso (2006). The challenge lies in the case when we (can) have jumps in the price and further the price is contaminated by microstructure noise. In this case, we propose to substitute $RV(t)$, $BV(t)$ and $TV(t)$ with the following robust to noise realized measures

$$\widetilde{RV}_\delta(t) := \frac{\sqrt{\delta}}{\psi_2 \tau} \sum_{i=0}^{M-k_M+1} |\bar{r}_\delta(t+i\delta)|^2 - \frac{\psi_1 \delta RV_\delta(t)}{2\tau^2\psi_2}, \tag{22}$$

$$\widetilde{TV}_\delta(t) := \frac{\sqrt{\delta}}{\psi_2 \tau \mu_{2/3}^3} \sum_{i=0}^{M-3k_M+1} |\bar{r}_\delta(t+i\delta)|^{2/3} |\bar{r}_\delta(t+(i+k_M)\delta)|^{2/3} |\bar{r}_\delta(t+(i+2k_M)\delta)|^{2/3} - \frac{\psi_1 \delta RV_\delta(t)}{2\tau^2\psi_2\mu_{2/3}^3}, \tag{23}$$

$$\widetilde{BV}_\delta(t) := \frac{\sqrt{\delta}}{\psi_2 \tau \mu_1^2} \sum_{i=0}^{M-2k_M+1} |\bar{r}_\delta(t+i\delta)||\bar{r}_\delta(t+(i+k_M)\delta)| - \frac{\psi_1 \delta RV_\delta(t)}{2\tau^2\psi_2\mu_1^2}, \tag{24}$$

$$\bar{r}_\delta(t+i\delta) := \sum_{j=1}^{k_M-1} \eta\left(\frac{j}{k_M}\right) r_\delta(t+(i+j)\delta), \quad k_M := \lfloor \tau/\sqrt{\delta} \rfloor, \quad \text{for some } \tau > 0, \tag{25}$$

$$\psi_1 = \int_0^1 (\eta'(s))^2 ds, \quad \psi_2 = \int_0^1 (\eta(s))^2 ds, \tag{26}$$

where $\eta : [0,1] \to R$ is C^1 with piecewise Lipschitz derivative and further $\eta(0) = \eta(1) = 0$ and $\int_0^1 \eta^2(u) > 0$ [12]. $\widetilde{RV}(t)$ has been proposed in Jacod et al. (2007) as an estimator of $QV(t)$ (but in the continuous case), building on work by Podolskij and Vetter (2007). The estimators $\widetilde{BV}(t)$ and $\widetilde{TV}(t)$ follow from a generalization, as in Jacod et al. (2007), of the pre-averaged multipower variation estimators of Podolskij and Vetter (2007) [13]. The intuition behind these estimators is very simple. We substitute the high-frequency increments in computing RV, TV and BV with local averages constructed from the adjacent *future* high-frequency increments. The length of the local averaging is determined by k_M which increases at a certain rate as we sample more frequently. The second terms in (22)-(24) correct for the bias created by the averaging of the noise.

[12] An example of η used in Jacod et al. (2007) is $\eta(s) = s \wedge (1-s)$.

[13] The generalization is in the weighting of the increments in (25), which leads to efficiency gains.

These estimators are consistent for QV and IV in the simultaneous presence of jumps and noise, and therefore estimation based on them will be consistent (i.e. Theorem 1, part (a) will hold). By extending the analysis of Podolskij and Vetter (2007)[14], it should be possible to prove also that $\widetilde{BV}_\delta(t)$ and $\widetilde{TV}_\delta(t)$ converge to $IV(t)$ at rate $\delta^{1/4}$. The rate of convergence of $\widetilde{RV}_\delta(t)$ (in the case of jumps) has not been established yet. Thus, the asymptotic equivalence result of Theorem 1, part (b) should remain valid, at least when only $\widetilde{TV}_\delta(t)$ (or $\widetilde{BV}_\delta(t)$) is used in the estimation, provided the rate condition $T\delta^{1-\epsilon}$ is replaced by the stronger one $T\delta^{1/4-\epsilon}$ for some $\epsilon > 0$. We leave for future work the complete analysis of the estimation based on the robust measures in (22)-(24).

Appendices

A Notation and some preliminary estimates

First, we introduce some notation to be used later in the proofs. The return over the interval $(t, t+\delta]$ can be broken into $r_\delta(t) = r_\delta^d(t) + r_\delta^c(t) + r_\delta^j(t)$, where $r_\delta^d(t) := \int_t^{t+\delta} b(s)ds$, $r_\delta^c(t) := \int_t^{t+\delta} \sigma_1(s)dW(s)$ and $r_\delta^j(t) := \int_t^{t+\delta}\int_E \phi(s,x)\tilde{\mu}(ds,dx)$. Denote with $X(t)$ the continuous martingale $X(t) := \int_0^t \sigma_1(s)dW(s)$ and with $Y(t)$ the discontinuous martingale $Y(t) := \int_0^t\int_E \phi(u,x)\tilde{\mu}(du,dx)$. In all proofs we will denote with K constant that does not depend on δ and which can change from line to line. The following bounds are easy to derive using our assumptions for arbitrary $p \geq 0$.

$$\mathbb{E}|r_\delta^d(t)|^p \leq K\delta^p, \quad \mathbb{E}|r_\delta^c(t)|^p \leq K\delta^{\frac{p}{2}}, \quad \mathbb{E}|r_\delta^j(t)|^p \leq K\delta^{\min\{p,1\}}, \mathbb{E}|r_\delta(t)|^p \leq K\delta^{\min\{\frac{p}{2},1\}}. \tag{A.1}$$

We can obtain stronger bounds for the jump part if we remove the compensator (recall (7)), i.e. if we make the decomposition $c(t) := r_\delta^d(t) + r_\delta^c(t) - \int_t^{t+\delta}\int_E \phi(s,x)dsG(dx)$ and $d(t) := r_\delta^j(t) + \int_t^{t+\delta}\int_E \phi(s,x)dsG(dx)$. Then we have

$$\mathbb{E}\left(|d(t)|^p\right) \leq K\delta^{\min\{\frac{p}{\alpha+\epsilon},1\}}, \quad \text{for } \forall\epsilon > 0 \tag{A.2}$$

[14]Their proof is for the continuous case, but generalization to the case of price jumps can be done easily using the same approach as for the proof of Lemma 1 and 2 of this paper.

where α is the index defined in (7) and we set $\frac{p}{0} = \infty$. Using similar arguments as for the proof of (A.1)-(A.2) and Hölder's inequality we can extend (A.1)-(A.2) as follows. Suppose that the intervals $(t_1, t_1+\delta], ..., (t_n, t_n+\delta]$ are arbitrary, but non-overlapping. Let $p = \max\{i : k_i \leq 2\}$, then we have

$$\mathbb{E}\left(|r_\delta(t_1)|^{k_1}...|r_\delta(t_n)|^{k_n}\right) \leq K\delta^\tau, \qquad \tau = \frac{k_1 + k_2 + ...k_p}{2} + n - p - \epsilon, \quad \text{for } \forall \epsilon > 0. \tag{A.3}$$

and the constant K does not depend on $t_1, ..., t_n$. Further, if for $\forall i$ $k_i \leq 2$, then the above holds with $\epsilon = 0$. Similar to the result in (A.3) we have

$$\mathbb{E}\left(c(t_1)^{p_1}...c(t_k)^{p_k} d(s_1)^{q_1}...d(s_n)^{q_n}\right) \leq K\delta^\tau, \tag{A.4}$$

where $\tau = \frac{p_1+...+p_k}{2} + \min\{\frac{q_1}{\alpha}, 1\} + ... + \min\{\frac{q_n}{\alpha}, 1\} - \epsilon$, for $\forall \epsilon > 0$.

B Lemma 1

Lemma 1 *In the stochastic volatility model (1) assume A2 and A3 hold. Then for every $t \geq 0$ and every $p > 0$ we have*

$$\sup_\delta \mathbb{E}\left(RV_\delta(t)\right)^p < \infty, \quad \sup_\delta \mathbb{E}\left(TV_\delta(t)\right)^p < \infty, \quad \sup_\delta \mathbb{E}\left(BV_\delta(t)\right)^p < \infty. \tag{B.1}$$

Proof. It is sufficient to establish the result for $p \geq 1$ and integer. So we assume from now on that this is the case. We start with the Realized Variance. We have

$$\left(RV_\delta(t)\right)^p \leq K\left(RV_\delta^d(t)\right)^p + K\left(RV_\delta^c(t)\right)^p + K\left(RV_\delta^j(t)\right)^p, \tag{B.2}$$

where $RV_\delta^d(t) := \sum_{i=1}^M |r_\delta^d(t+(i-1)\delta)|^2$, $RV_\delta^c(t) := \sum_{i=1}^M |r_\delta^c(t+(i-1)\delta)|^2$ and $RV_\delta^j(t) := \sum_{i=1}^M |r_\delta^j(t+(i-1)\delta)|^2$.

We prove the uniform integrability of each of the three terms on the right hand side of the above inequality. For the first two terms this is straightforward and follows directly from the preliminary results. We show uniform integrability of $\left(RV_\delta^j(t)\right)^p$, for which it suffices to show that $\left(RV_\delta^j(t) - \int_t^{t+1}\int_E \phi^2(s,x)\mu(ds,dx)\right)^p$ is uniformly integrable. Using Ito's formula we have

$$RV_\delta^j(t) - \int_t^{t+1}\int_E \phi^2(u,x)\mu(du,dx) = \sum_{i=1}^M y_\delta(t+(i-1)\delta),$$

where $y_\delta(t) = 2\int_t^{t+\delta}\int_E (Y(u-) - Y(t))\phi(u,x)\tilde{\mu}(du,dx)$. We start with providing a bound for the p-th absolute moment of $y_\delta(0)$. Using the integrability assumption for the process $\sigma_2(u)$ and with a successive application of the Burkholder-Davis-Gundy inequality for $p \geq 1$ we get

$$\mathbb{E}\left|\int_0^\delta \int_E 2Y(u-)\phi(u,x)\tilde{\mu}(du,dx)\right|^p \leq K\int_0^\delta \mathbb{E}\left(|Y(u)|^p|\sigma_2(u)|^p\right)du \int_E |h(x)|^p G(dx).$$

Applying Hölder's inequality we have $\mathbb{E}\left(|Y(u)|^p|\sigma_2(u)|^p\right) \leq Ku^{\frac{1}{1+\epsilon}}$, and therefore $\mathbb{E}\left(y_\delta(0)\right)^p \leq K\delta^{2-\epsilon}$ for $\forall p \geq 1$ and $\forall \epsilon > 0$. This can be further generalized. First, as above for $p \geq 1$ using Burholder-Davis-Gundy inequality, we can write

$$\mathbb{E}_t\,|y_\delta(t)|^p \leq K\mathbb{E}_t\left(\int_t^{t+\delta} |Y(u-) - Y(t)|^p|\sigma_2(u-)|^p du\right)\int_E |h(x)|^p G(dx),$$

where $\mathbb{E}_t$ is a shorthand for $\mathbb{E}\left(\cdot|\mathscr{F}_t\right)$. Therefore, using law of iterated expectations and Hölder's inequality, we have

$$\mathbb{E}\left(|y_\delta(t_1)|^{p_1}...|y_\delta(t_k)|^{p_k}\right) \leq K\delta^{k+1-\epsilon}, \quad \text{for } \forall\epsilon > 0, \tag{B.3}$$

where $p_i \geq 1$ and $t_i > t_{i-1} + \delta$ for $i = 1,...,k$. With this we are ready to prove the uniform integrability of $(RV_\delta^j(t))^p$. For a positive integer p, $\left(\sum_{i=1}^M y_\delta(t+(i-1)\delta)\right)^p$ has a typical term of the form $y_\delta(t_1)^{p_1}...y_\delta(t_k)^{p_k}$, where $1 \leq k \leq p$; $p_1, p_2, ..., p_k$ are positive distinct integers such that $p_1 + p_2 + ... + p_k = p$; $t_1, ..., t_k$ are distinct and all in the interval $(t, t+1]$. For given $p_1, p_2, ..., p_k$ there are $\binom{M}{k}$ of these elements. Therefore it is sufficient to show that $\mathbb{E}\left(y_\delta(t_1)^{p_1}...y_\delta(t_k)^{p_k}\right) = O(\delta^k)$. But this is an easy consequence of the result in (B.3). Hence we proved that for $\forall p \geq 1$ we have

$$\sup_\delta \mathbb{E}\left(RV_\delta^j(t) - \int_t^{t+1}\int_E \phi^2(u,x)\mu(du,dx)\right)^p < \infty,$$

and from here we get the uniform integrability result in (B.1).

We continue with proving the uniform integrability of $(TV_\delta)^p$. It has a typical term of the form $Ka_{i_1}^{p_1}a_{i_2}^{p_2}...a_{i_k}^{p_k}$, where $1 \leq k \leq p$; $p_1, p_2, ..., p_k$ are positive distinct integers such that $p_1 + p_2 + ... + p_k = p$; i_1,...,i_k are distinct integers between 1 and M; $a_i = |r_\delta(t+(i-2)\delta)|^{\frac{2}{3}}|r_\delta(t+(i-1)\delta)|^{\frac{2}{3}}|r_\delta(t+i\delta)|^{\frac{2}{3}}$ for $i \geq 2$ (a_1 and a_2 are set to zero). To prove the uniform integrability of $(TV_\delta)^p$ it suffices to show that $\mathbb{E}\left(a_{i_1}^{p_1}a_{i_2}^{p_2}...a_{i_k}^{p_k}\right) =$

$O(\delta^k)$ in the case when there is no overlapping of returns in the terms a_i and that $\mathbb{E}\left(a_{i_1}^{p_1} a_{i_2}^{p_2} ... a_{i_k}^{p_k}\right) = O(\delta^{k-\epsilon})$ for some $\epsilon > 0$ when there is some overlapping of the returns in the terms a_i. First, when there is no overlapping of returns in the terms a_i we have $\mathbb{E}\left(a_{i_1}^{p_1} a_{i_2}^{p_2} ... a_{i_k}^{p_k}\right) = O(\delta^d)$, where $d = 3\min\{\frac{p_1}{3}, 1-\epsilon\} + ... + 3\min\{\frac{p_k}{3}, 1-\epsilon\}$ for some $\epsilon > 0$. It is clear that $d \geq p \geq k$.

We establish the result when there is overlapping of returns in the a_i terms. It is sufficient to show this when there is maximum overlapping of returns, i.e. $t_{i_1}, ..., t_{i_k}$ being consecutive. Without loss of generality we set $i_1 = 3$. In this case we can write

$$\begin{aligned} a_{i_1}^{p_1} a_{i_2}^{p_2} ... a_{i_k}^{p_k} \quad = \quad & |r_\delta(t+\delta)|^{\frac{2}{3}p_1} |r_\delta(t+2\delta)|^{\frac{2}{3}(p_1+p_2)} |r_\delta(t+3\delta)|^{\frac{2}{3}(p_1+p_2+p_3)} |r_\delta(t+4\delta)|^{\frac{2}{3}(p_2+p_3+p_4)} ... \\ & |r_\delta(t+k\delta)|^{\frac{2}{3}(p_{k-2}+p_{k-1}+p_k)} |r_\delta(t+(k+1)\delta)|^{\frac{2}{3}(p_{k-1}+p_k)} |r_\delta(t+(k+2)\delta)|^{\frac{2}{3}p_k}, \end{aligned}$$

and since $p_i \geq 1$ we have $\frac{2}{3}(p_{i-2} + p_{i-1} + p_i) \geq 2$ for $i \geq 2$. Therefore using the result in (A.3) we have $\mathbb{E}\left(a_{i_1}^{p_1} a_{i_2}^{p_2} ... a_{i_k}^{p_k}\right) = O(\delta^d)$, where $d = \min\{\frac{p_1}{3}, 1\} + \min\{\frac{p_1+p_2}{3}, 1\} + k - 2 + \min\{\frac{p_{k-1}+p_k}{3}, 1\} + \min\{\frac{p_k}{3}, 1\} - \epsilon$ for every $\epsilon > 0$ and therefore $d \geq k - \epsilon$ and this proves the second claim in (B.1).

We prove the uniform integrability result for $(BV_\delta(t))^p$ in the same way. The typical term of $(BV_\delta(t))^p$ is $K a_{i_1}^{p_1} a_{i_2}^{p_2} ... a_{i_k}^{p_k}$, where $1 \leq k \leq p$; $p_1, p_2, ..., p_k$ are positive distinct integers such that $p_1 + p_2 + ... + p_k = p$; i_1,...,i_k are distinct integers between 1 and M; $a_i = |r_\delta(t + (i-1)\delta)||r_\delta(t + i\delta)|$ for $i \geq 1$ (a_1 is set to zero). When there is no overlapping in the returns of the terms a_i (i.e. no consecutive i_j) we have $\mathbb{E}\left(a_{i_1}^{p_1} a_{i_2}^{p_2} ... a_{i_k}^{p_k}\right) = O(\delta^d)$, where $d = 2\min\{\frac{p_1}{2}, 1-\epsilon\} + ... + 2\min\{\frac{p_k}{2}, 1-\epsilon\}$ for some $\epsilon > 0$. It is clear that $d \geq p \geq k$.

In the case of overlapping returns in the a_i terms we look at the worst case when $t_{i_1}, ..., t_{i_k}$ are consecutive. Without loss of generality we set $i_1 = 2$. In this case we can write

$$a_{i_1}^{p_1} a_{i_2}^{p_2} ... a_{i_k}^{p_k} = |r_\delta(t+\delta)|^{p_1} |r_\delta(t+2\delta)|^{(p_1+p_2)} |r_\delta(t+3\delta)|^{(p_2+p_3)} ... |r_\delta(t+k\delta)|^{(p_{k-1}+p_k)} |r_\delta(t+(k+1)\delta)|^{p_k},$$

and since $p_i \geq 1$ we have $p_{i-1} + p_i \geq 2$ for $i \geq 1$. Therefore using the result in (A.3) we have $\mathbb{E}\left(a_{i_1}^{p_1} a_{i_2}^{p_2} ... a_{i_k}^{p_k}\right) = O(\delta^d)$, where $d = \min\{\frac{p_1}{2}, 1\} + k - 1 + \min\{\frac{p_k}{2}, 1\} - \epsilon$ for every $\epsilon > 0$ and therefore $d \geq k - \epsilon$. This proves the last claim in (B.1). $\square$

C Lemma 2

Lemma 2 *In the stochastic volatility model (1) assume A2 and A3 hold. Then for every $t \geq 0$ and $\epsilon > 0$ we have*

$$\sup_{\delta} \delta^{-1+\epsilon}\mathbb{E}\left(RV_\delta(t) - \int_t^{t+1}\sigma_1^2(s)ds - \int_t^{t+1}\int_E \phi^2(s,x)\mu(ds,dx)\right)^2 < \infty, \quad \text{(C.1)}$$

$$\sup_{\delta} \delta^{-1}\mathbb{E}\left(TV_\delta(t) - \int_t^{t+1}\sigma_1^2(s)ds\right)^2 < \infty, \quad \sup_{\delta} \delta^{-1+\epsilon}\mathbb{E}\left(BV_\delta(t) - \int_t^{t+1}\sigma_1^2(s)ds\right)^2 < \infty. \quad \text{(C.2)}$$

Proof. We start with the result for RV. We provide an upper bound for the expression in (C.1). We use the following inequality

$$\begin{aligned}
&\mathbb{E}\left(RV_\delta(t) - \int_t^{t+1}\sigma_1^2(s)ds - \int_t^{t+1}\int_E \phi^2(s,x)\mu(ds,dx)\right)^2 \\
&\leq 6\mathbb{E}\left(RV_\delta^c(t) - \int_t^{t+1}\sigma_1^2(s)ds\right)^2 + 6\mathbb{E}\left(RV_\delta^j(t) - \int_t^{t+1}\int_E \phi^2(s,x)\mu(ds,dx)\right)^2 \\
&+24\mathbb{E}\left(\sum_{i=1}^M |r_\delta^c(t+(i-1)\delta)||r_\delta^j(t+(i-1)\delta)|\right)^2 + 24\mathbb{E}\left(\sum_{i=1}^M |r_\delta^j(t+(i-1)\delta)||r_\delta^d(t+(i-1)\delta)|\right)^2 \\
&+6\mathbb{E}\left(RV_\delta^d(t)\right)^2 + 24\mathbb{E}\left(\sum_{i=1}^M |r_\delta^c(t+(i-1)\delta)||r_\delta^d(t+(i-1)\delta)|\right)^2. \quad \text{(C.3)}
\end{aligned}$$

We proceed with bounding each of the terms on the right hand side of (C.3). Using our preliminary estimates, it is easy to show that the last four terms on the right hand side of the above inequality are bounded by $K\delta^{1-\epsilon}$ for $\forall\epsilon > 0$. For the first term on the right hand side of (C.3) easy transformations give

$$\mathbb{E}\left(RV_\delta^c(t) - \int_t^{t+1}\sigma_1^2(s)ds\right)^2 = M\mathbb{E}\left(|r_\delta^c(0)|^2 - \int_0^\delta \sigma_1^2(s)ds\right)^2 = 4M\int_0^\delta \mathbb{E}(X^2(s)\sigma_1^2(s))ds \leq K\delta.$$

Similarly, for the second term on the right hand side of (C.3) we have

$$\mathbb{E}\left(RV_\delta^j(t) - \int_t^{t+1}\int_E \phi^2(s,x)\mu(ds,dx)\right)^2 = M\mathbb{E}\left(y_\delta(0)\right)^2 \leq K\delta^{1-\epsilon}, \quad \text{for } \forall\epsilon > 0,$$

and from here the uniform integrability result in (C.1) follows.

We continue with showing the first uniform integrability result in (C.2). We make use of the following inequality. For arbitrary real numbers $a_1, a_2, ...a_n,\ b_1, b_2, ..., b_n$ and $p_1, p_2, ..., p_n$ such that $0 < p_i < 1$ for $i = 1, ..., n$ we have

$$\left| |a_1 + b_1|^{p_1} ... |a_n + b_n|^{p_n} - |a_1|^{p_1} ... |a_n|^{p_n} \right| \leq \sum |x_1|^{p_1} ... |x_n|^{p_n} - |a_1|^{p_1} ... |a_n|^{p_n}, \quad \text{where } x_i = \{a_i, b_i\},$$

and the summation in the first term on right hand side of the above inequality is over $x_i = \{a_i, b_i\}$ for $i = 1, ..., n$. Using this inequality we can write

$$\begin{aligned}
&\delta^{-1}\mathbb{E}\left(TV_\delta(t) - \int_t^{t+1} \sigma_1^2(s)ds\right)^2 \\
&\leq K\delta^{-1}\mathbb{E}\left(\sum_{i=3}^{M-1} |c(t+(i-2)\delta)|^{\frac{2}{3}} |c(t+(i-1)\delta)|^{\frac{2}{3}} |c(t+i\delta)|^{\frac{2}{3}} - \int_t^{t+1} \sigma_1^2(s)ds\right)^2 + A,
\end{aligned}$$

where A is a sum of (expectation of) terms similar to the summation inside the expectation on the right hand side of the above inequality, but where at least one of the terms in the products is $d(\cdot)$. From the results in Barndorff-Nielsen et al. (2005) follows that the first term on the right hand side of the above inequality is bounded by a constant. For A we can use the inequalities in the preliminary section to show that $A \leq K\delta^{\frac{4-5\alpha}{3\alpha} \wedge \frac{1}{3} - \epsilon}$. This proves the first uniform integrability result in (C.2). The proof of the second result in (C.2) is exactly the same and is therefore skipped. □

D Proof of Theorem 1

(a) For the consistency of $\widehat{\theta}_f$ we need to show that $\sup_{\theta\in\Theta} ||\widehat{m}_T(\theta) - m_0(\theta)|| \xrightarrow{p} 0$. Using the triangle inequality we have

$$||\widehat{m}_T(\theta) - m_0(\theta)|| \leq ||m_T(\theta) - m_0(\theta)|| + ||\widehat{m}_T(\theta) - m_T(\theta)||.$$

In view of A1, we obviously need to show only that the second term on the right hand side of the above inequality converges uniformly in probability to 0. Using A4, we have

$$||\widehat{m}_T(\theta) - m_T(\theta)|| \leq \frac{1}{T}\sum_{t=1}^{T} ||m(\widehat{z}(t), \theta) - m(z(t), \theta)|| \leq ||C(\theta)||\frac{1}{T}\sum_{t=1}^{T} ||P(\widehat{z}(t)) - P(z(t))||.$$

To proceed we use the fact that $\widehat{z}(t) \xrightarrow{p} z(t)$ for every t (Jacod and Shiryaev (2003), Barndorff-Nielsen et al. (2006)). Then using the uniform integrability result of Lemma 1 we can conclude that for every t and $r > 0$: $\mathbb{E}\left(||\widehat{z}(t) - z(t)||^r\right) \to 0$. Since $P(z)$ has at most polynomial growth in z, the uniform integrability of $||\widehat{z}(t)||^r$ for arbitrary $r > 0$ and $t > 0$ implies the uniform integrability of $||P(\widehat{z}(t))||$ and its powers for arbitrary $t > 0$. Also, since $P(z)$ is continuous in z, by continuous mapping $P(\widehat{z}(t)) \xrightarrow{p} P(z(t))$. These two facts combined yield

$$\mathbb{E}\left(||P(\widehat{z}(t)) - P(z(t))||\right) \to 0, \qquad \text{for } \forall t > 0.$$

Therefore, using the stationarity of $z(t)$ and $\widehat{z}(t)$ we have

$$\mathbb{E}\left(\sup_{\theta\in\Theta} ||\widehat{m}_T(\theta) - m_T(\theta)||\right) \leq \sup_{\theta\in\Theta} ||C(\theta)||\mathbb{E}\left(||P(\widehat{z}(t)) - P(z(t))||\right).$$

Therefore, $\widehat{m}_T(\theta)$ converges uniformly to $m_0(\theta)$ in a neighborhood of θ_0, and from here the result in part (a) of Theorem 1 follows.

(b) Using A1

$$\sqrt{T}\left(\widehat{\theta}_f - \theta_0\right) = \left(\nabla_\theta \widehat{m}_T(\widehat{\theta}_f)'\widehat{W}_T \nabla_\theta \widehat{m}_T(\tilde{\theta})\right)^{-1} \nabla_\theta \widehat{m}_T(\widehat{\theta}_f)'\widehat{W}_T \left(\sqrt{T}\widehat{m}_T(\theta_0)\right),$$

where $\tilde{\theta}$ is between θ_0 and $\widehat{\theta}_f$. Now, from A1 we know that a CLT for $m_T(\theta_0)$ holds and that $\nabla_\theta m_T(\theta)$ converges in probability uniformly around θ_0. Therefore, part (b) of the Theorem will follow if we show

$$\sup_{\theta\in\Theta} ||\nabla_\theta \widehat{m}_T(\theta) - \nabla_\theta m_T(\theta)|| \xrightarrow{p} 0 \quad \text{and} \quad \sqrt{T}||\widehat{m}_T(\theta_0) - m_T(\theta_0)|| \xrightarrow{p} 0. \tag{D.1}$$

For the first result in (D.1) we use A5 and we have

$$\begin{aligned} ||\nabla_\theta \widehat{m}_T(\theta) - \nabla_\theta m_T(\theta)|| &\leq \frac{1}{T}\sum_{t=1}^{T} \left|\left|\nabla_\theta m(\widehat{z}(t), \theta) - \nabla_\theta m(z(t), \theta)\right|\right| \\ &\leq ||C(\theta)||\frac{1}{T}\sum_{t=1}^{T} ||P(\widehat{z}(t)) - P(z(t))||. \end{aligned}$$

As in part (a), we use Lemma 1 and we have $\mathbb{E}\left(||P(\widehat{z}(t)) - P(z(t))||\right) \to 0$ for $\forall t > 0$. Therefore, using the stationarity of $z(t)$ and $\widehat{z}(t)$

$$\mathbb{E}\left(\sup_{\theta\in\Theta} ||\nabla_\theta \widehat{m}_T(\theta) - \nabla_\theta m_T(\theta)||\right) \leq \sup_{\theta\in\Theta} ||C(\theta)||\mathbb{E}\left(||P(\widehat{z}(t)) - P(z(t))||\right),$$

and from here the first result in (D.1) follows. We prove the second convergence in probability result in (D.1) by showing again convergence in L^1.

$$\begin{aligned}\sqrt{T}\mathbb{E}\left(||\widehat{m}_T(\theta_0) - m_T(\theta_0)||\right) &\leq \frac{\sqrt{T}}{T}\mathbb{E}\left(\sum_{t=1}^{T}\left|\left|m(\widehat{z}(t),\theta_0) - m(z(t),\theta_0)\right|\right|\right) \\ &\leq \sqrt{T\delta^{1-\epsilon}}\mathbb{E}\left(\delta^{-1/2+\epsilon/2}\left|\left|m(\widehat{z}(t),\theta_0) - m(z(t),\theta_0)\right|\right|\right),\end{aligned}$$

where $\epsilon > 0$ is such that $T\delta^{1-\epsilon} \to 0$. Expanding $m(\widehat{z}(t),\theta_0)$ around $z(t)$ and using Cauchy-Schwartz inequalities

$$\mathbb{E}\left(\delta^{-1/2+\epsilon/2}\left|\left|m(\widehat{z}(t),\theta_0) - m(z(t),\theta_0)\right|\right|\right) \leq \sqrt{\mathbb{E}\left(||\nabla_z m(\tilde{z}(t),\theta_0)||^2\right)}\sqrt{\mathbb{E}\left(\delta^{-1+\epsilon}||\widehat{z}(t) - z(t)||^2\right)},$$

where $\tilde{z}(t)$ is between $\widehat{z}(t)$ and $z(t)$. Since $\widehat{z}(t) \xrightarrow{p} z(t)$ and $||\tilde{z}(t) - z(t)||^r \leq ||\widehat{z}(t) - z(t)||^r$ for $r > 0$, the uniform integrability of $||\widehat{z}(t)||^r$ implies uniform integrability of $||\tilde{z}(t)||^r$. In addition, using A6, $\nabla_z m(z,\theta)$ has polynomial growth in z and therefore $||\nabla_z m(\tilde{z}(t),\theta_0)||^2$ is uniformly integrable. From continuous mapping also $||\nabla_z m(\tilde{z}(t),\theta_0)||^2 \xrightarrow{p} ||\nabla_z m(z(t),\theta_0)||^2$. Therefore $\mathbb{E}\left(||\nabla_z m(\tilde{z}(t),\theta_0)||^2\right) \to \mathbb{E}\left(||\nabla_z m(z(t),\theta_0)||^2\right)$. Using the uniform integrability result of Lemma 2, we have $\sup_\delta \mathbb{E}\left(\delta^{-1+\epsilon}||\widehat{z}(t) - z(t)||^2\right) < K$, for some constant K. Therefore, since $T\delta^{1-\epsilon} \to 0$, we finally have $\sqrt{T}\mathbb{E}||\widehat{m}_T(\theta_0) - m_T(\theta_0)|| \to 0$, and hence the second convergence result in (D.1). $\square$

E Proof of Corollary 1

Under A3$'$ we have $\int_{\mathbb{R}_0^n} 1_{(k(\mathbf{x})>1)}|k(\mathbf{x})|^p G(d\mathbf{x}) < \infty$ and $\int_{-\infty}^{t} |g(t-s)|^p ds < \infty$ for any $p > 0$. Therefore, using Theorem 3.3 in Rajput and Rosiński (1989), we have $\mathbb{E}|V^{\|}(t)|^p < \infty$. This implies that, under the conditions in A3$'$, we have integrability of all powers of $\sigma_1^2(t)$. The proof of the Corollary is exactly the same as the proof of Theorem 1, part(b). We need only to establish the following uniform integrability result for RV, which strengthens the result (C.1) in Lema 2

$$\sup_\delta \delta^{-1}\mathbb{E}\left(RV_\delta(t) - \int_t^{t+1}\sigma_1^2(s)ds - \int_t^{t+1}\int_{\mathbb{R}_0^n} h^2(\mathbf{x})\mu(ds,d\mathbf{x})\right)^2 < \infty \qquad \text{(E.1)}$$

From the proof of (C.1) in Lemma 2, it is clear that we will be done if we can show

$$\delta^{-1}\mathbb{E}\left(RV_\delta^j(t) - \int_t^{t+1}\int_{\mathbb{R}_0^n} h^2(\mathbf{x})\mu(ds,d\mathbf{x})\right)^2 \leq K \qquad \text{(E.2)}$$

$$\delta^{-1}\mathbb{E}\left(\sum_{i=1}^{M}|r_\delta^j(t+(i-1)\delta)||r_\delta^c(t+(i-1)\delta)|\right)^2 \leq K. \tag{E.3}$$

(E.2) follows from

$$\delta^{-1}\mathbb{E}\left(\sum_{i=1}^{M}|r_\delta^j(t+(i-1)\delta)|^2 - \int_t^{t+1}\int_{\mathbb{R}_0^n} h^2(\mathbf{x})\mu(ds,d\mathbf{x})\right)^2 = 2\left(\int_{\mathbb{R}_0^n} h^2(\mathbf{x})G(d\mathbf{x})\right)^2.$$

Therefore we are left with showing (E.3). First, note that $V^{\|}(t)$ is adapted to the filtration generated by the homogeneous Poisson measure μ, while the Brownian motion in the price, $W(t)$, as well as $V^{\perp}(t)$ are independent from it. To show the result (E.3) we need to verify for $t \neq s$

$$\mathbb{E}\left(r_\delta^c(t)r_\delta^j(t)\right)^2 = O(\delta^2) \quad \text{and} \quad \mathbb{E}\left(|r_\delta^c(s)||r_\delta^j(s)||r_\delta^c(t)||r_\delta^j(t)|\right) = O(\delta^3). \tag{E.4}$$

In what follows for any $t > 0$, u and a we set

$$H_a(t,u) = \begin{cases} \int_t^{t+a} g(z-u)dz & \text{if } u < t \\ \int_u^{t+a} g(z-u)dz & \text{if } t \leq u < t+a. \end{cases}$$

Then, the first result in (E.4) follows from

$$\mathbb{E}\left(r_\delta^c(t)r_\delta^j(t)\right)^2 \leq K\int_t^{t+\delta} H_\delta(t,u)du \int_{\mathbb{R}_0^n} h^2(\mathbf{x})k(\mathbf{x})G(d\mathbf{x}) \leq K\delta^2.$$

We proceed with showing the second result in (E.4). For $t > s$, by conditioning on the filtration generated by μ up to time $t+\delta$ we can write

$$\begin{aligned}\mathbb{E}\left|r_\delta^c(s)r_\delta^j(s)r_\delta^c(t)r_\delta^j(t)\right| \leq\ & \mathbb{E}\Bigg|\sqrt{K_1\delta + \int_{-\infty}^{s+\delta}\int_{\mathbb{R}_0^n} H_\delta(s,u)k(\mathbf{x})\mu(du,d\mathbf{x})}\int_s^{s+\delta}\int_{\mathbb{R}_0^n} h(\mathbf{x})\tilde{\mu}(du,d\mathbf{x}) \\ & \times\sqrt{K_2\delta + \int_{-\infty}^{t+\delta}\int_{\mathbb{R}_0^n} H_\delta(t,u)k(\mathbf{x})\mu(du,d\mathbf{x})}\int_t^{t+\delta}\int_{\mathbb{R}_0^n} h(\mathbf{x})\tilde{\mu}(du,d\mathbf{x})\Bigg|,\end{aligned}$$

where K_1 and K_2 are some constants. To proceed further we can split each of the integrals on the right hand side of the above inequality into integrals over the disjoint intervals $(-\infty, s]$, $(s, s+\delta]$, $(s+\delta, t]$ and $(t, t+\delta]$. Thus we can bound $\mathbb{E}\left|r_\delta^c(s)r_\delta^j(s)r_\delta^c(t)r_\delta^j(t)\right|$ by a sum of terms, each of which is a product of deterministic integrals with respect to μ (or square root of them) over either exactly the same interval or non-overlapping interval. Using the time-homogeneity property of the Poisson measure μ, the expectation of these products will be equal to the product of the expectations of the terms defined over the same interval. Therefore, we need to check the order of magnitude only of the terms in the products with overlapping interval. Applying Cauchy-Schwartz inequality we have

$$\mathbb{E}\left|\sqrt{\int_t^{t+\delta}\int_{\mathbb{R}_0^n} H_\delta(t,u)\mu(du,d\mathbf{x})}\int_t^{t+\delta}\int_{\mathbb{R}_0^n} h(\mathbf{x})\tilde{\mu}(du,d\mathbf{x})\right| \leq K\delta^{3/2},$$

$$\mathbb{E}\left|\sqrt{\int_s^{s+\delta}\int_{\mathbb{R}_0^n} H_\delta(t,u)\mu(du,d\mathbf{x})}\int_s^{s+\delta}\int_{\mathbb{R}_0^n} h(\mathbf{x})\tilde{\mu}(du,d\mathbf{x})\right| \leq K\delta^{3/2},$$

$$\mathbb{E}\left|\sqrt{\int_s^{s+\delta}\int_{\mathbb{R}_0^n} H_\delta(t,u)\mu(du,d\mathbf{x})}\sqrt{\int_s^{s+\delta}\int_{\mathbb{R}_0^n} H_\delta(s,u)\mu(du,d\mathbf{x})}\int_s^{s+\delta}\int_{\mathbb{R}_0^n} h(\mathbf{x})\tilde{\mu}(du,d\mathbf{x})\right| \leq K\delta^2,$$

$$\mathbb{E}\left(\sqrt{\int_{-\infty}^{s}\int_{\mathbb{R}_0^n} H_\delta(t,u)k(\mathbf{x})\mu(du,d\mathbf{x})\int_{-\infty}^{s}\int_{\mathbb{R}_0^n} H_\delta(s,u)k(\mathbf{x})\mu(du,d\mathbf{x})}\right) \leq K\delta,$$

$$\mathbb{E}\left(\sqrt{\int_{s+\delta}^{t}\int_{\mathbb{R}_0^n} H_\delta(t,u)k(\mathbf{x})\mu(du,d\mathbf{x})}\right) \leq K\delta^{1/2}.$$

Combining these results we verify (E.4) and from here we readily have (E.3). □

F Parameter Settings for the Monte Carlo Study

In all cases the parameters governing $\sigma_1^2(t)$ are kept the same with values: $\rho = 0.07$, $k_2 = 0.0520$, $c_v = 0.1$, $\lambda_v = 0.0173$ and $\alpha_v = 0.5$. The price jump parameters were specified as follows

- **compound Poisson case:** $k_1 = 1.0$, $c_p = 0.1$ and $\sigma_p = 1.0$,
- **case** $\alpha_p = 0.0$**:** $k_1 = 0.0106$, $c_p = 0.1$ and $\lambda_p = 0.015$,
- **case** $\alpha_p = 0.1$**:** $k_1 = 0.0119$, $c_p = 0.125$ and $\lambda_p = 0.015$,
- **case** $\alpha_p = 0.5$**:** $k_1 = 0.0161$, $c_p = 0.4$ and $\lambda_p = 0.015$.

For the simulation of the jumps in the price and in the variance we use the series representation method (see Rosiński (2001)).

Table 1: Monte Carlo Results for the case of compound Poisson price jumps

Parameter	True value	Mean	Median	RMSE	Asy./Small sample s.e.
Panel A. GMM with QV and IV					
ρ	0.0700	0.0726	0.0729	0.0124	0.8693
$\int_{\mathbb{R}_0} k_2 x G_2(dx)/\rho$	1.0000	0.9828	0.9803	0.0885	0.9304
$\int_{\mathbb{R}_0} k_1^2 x^2 G_1(dx)$	0.1000	0.0996	0.0992	0.0099	1.0205
$\sqrt{\int_{\mathbb{R}_0} k_2^2 x^2 G_2(dx)}$	0.3240	0.3093	0.3032	0.0578	0.9375
$\sqrt{\int_{\mathbb{R}_0} k_1^4 x^4 G_1(dx)}$	0.5477	0.5429	0.5430	0.0552	1.0145
Panel B. GMM with RV and BV from high-frequency data for $M = 100$					
ρ	0.0700	0.0662	0.0614	0.0413	0.2553
$\int_{\mathbb{R}_0} k_2 x G_2(dx)/\rho$	1.0000	1.1244	1.1215	0.1814	0.6118
$\int_{\mathbb{R}_0} k_1^2 x^2 G_1(dx)$	0.1000	0.0883	0.0882	0.0174	0.7752
$\sqrt{\int_{\mathbb{R}_0} k_2^2 x^2 G_2(dx)}$	0.3240	0.4046	0.3918	0.1796	0.3265
$\sqrt{\int_{\mathbb{R}_0} k_1^4 x^4 G_1(dx)}$	0.5477	0.4955	0.5005	0.1176	0.5294
Panel C. GMM with RV and BV from high-frequency data for $M = 300$					
ρ	0.0700	0.0492	0.0472	0.0329	0.4125
$\int_{\mathbb{R}_0} k_2 x G_2(dx)/\rho$	1.0000	1.0654	1.0640	0.1230	0.7758
$\int_{\mathbb{R}_0} k_1^2 x^2 G_1(dx)$	0.1000	0.0903	0.0899	0.0144	0.9435
$\sqrt{\int_{\mathbb{R}_0} k_2^2 x^2 G_2(dx)}$	0.3240	0.2972	0.2944	0.1092	0.4953
$\sqrt{\int_{\mathbb{R}_0} k_1^4 x^4 G_1(dx)}$	0.5477	0.5200	0.5174	0.0748	0.8029
Panel D. GMM with RV and TV from high-frequency data for $M = 100$					
ρ	0.0700	0.0685	0.0630	0.0463	0.2272
$\int_{\mathbb{R}_0} k_2 x G_2(dx)/\rho$	1.0000	1.0848	1.0812	0.1615	0.5873
$\int_{\mathbb{R}_0} k_1^2 x^2 G_1(dx)$	0.1000	0.1098	0.1101	0.0175	0.6897
$\sqrt{\int_{\mathbb{R}_0} k_2^2 x^2 G_2(dx)}$	0.3240	0.3973	0.3865	0.1849	0.3086
$\sqrt{\int_{\mathbb{R}_0} k_1^4 x^4 G_1(dx)}$	0.5477	0.5463	0.5472	0.1072	0.5210

*N*ote: For the last column, the theoretical standard errors were computed using the true parameter values and sample averages from simulated series of IV and QV with length $600,000$ days.

Table 2: Monte Carlo Results for the case of compound Poisson price jumps

Parameter	True value	Mean	Median	RMSE	Asy./Small sample s.e.
Panel E. GMM with RV and TV from high-frequency data for $M = 300$					
ρ	0.0700	0.0485	0.0467	0.0344	0.3910
$\int_{\mathbb{R}_0} k_2 x G_2(dx)/\rho$	1.0000	1.0469	1.0474	0.1158	0.7626
$\int_{\mathbb{R}_0} k_1^2 x^2 G_1(dx)$	0.1000	0.1009	0.1003	0.0114	0.8850
$\sqrt{\int_{\mathbb{R}_0} k_2^2 x^2 G_2(dx)}$	0.3240	0.2902	0.2866	0.1148	0.4777
$\sqrt{\int_{\mathbb{R}_0} k_1^4 x^4 G_1(dx)}$	0.5477	0.5410	0.5378	0.0721	0.7771
Panel F. GMM with RV and TV from high-frequency data for $M = 100$ **and asymptotic refinement**					
ρ	0.0700	0.0699	0.0695	0.0121	0.8693
$\int_{\mathbb{R}_0} k_2 x G_2(dx)/\rho$	1.0000	0.9636	0.9602	0.0949	0.9219
$\int_{\mathbb{R}_0} k_1^2 x^2 G_1(dx)$	0.1000	0.1091	0.1085	0.0134	1.0205
$\sqrt{\int_{\mathbb{R}_0} k_2^2 x^2 G_2(dx)}$	0.3240	0.2900	0.2859	0.0689	0.8749
$\sqrt{\int_{\mathbb{R}_0} k_1^4 x^4 G_1(dx)}$	0.5477	0.5581	0.5544	0.0573	0.9911
Panel G. GMM with RV and TV from high-frequency data for $M = 300$ **and asymptotic refinement**					
ρ	0.0700	0.0710	0.0704	0.0129	0.8218
$\int_{\mathbb{R}_0} k_2 x G_2(dx)/\rho$	1.0000	0.9788	0.9761	0.0891	0.9326
$\int_{\mathbb{R}_0} k_1^2 x^2 G_1(dx)$	0.1000	0.1010	0.1008	0.0098	1.0310
$\sqrt{\int_{\mathbb{R}_0} k_2^2 x^2 G_2(dx)}$	0.3240	0.3024	0.2985	0.0623	0.8974
$\sqrt{\int_{\mathbb{R}_0} k_1^4 x^4 G_1(dx)}$	0.5477	0.5478	0.5466	0.0555	1.0054

Table 3: Monte Carlo Results for the case $\alpha_p = 0.0$

Parameter	True value	Mean	Median	RMSE	Asy./Small sample s.e.
Panel A. GMM with QV and IV					
ρ	0.0700	0.0730	0.0730	0.0127	0.8552
$\int_{\mathbb{R}_0} k_2 x G_2(dx)/\rho$	1.0000	0.9890	0.9817	0.0885	0.9198
$\int_{\mathbb{R}_0} k_1^2 x^2 G_1(dx)$	0.1000	0.1004	0.1000	0.0101	1.0272
$\sqrt{\int_{\mathbb{R}_0} k_2^2 x^2 G_2(dx)}$	0.3240	0.3112	0.3030	0.0580	0.9259
$\sqrt{\int_{\mathbb{R}_0} k_1^4 x^4 G_1(dx)}$	0.5477	0.5385	0.5196	0.1150	1.2744
Panel B. GMM with RV and BV from high-frequency data for $M = 100$					
ρ	0.0700	0.0687	0.0638	0.0441	0.2385
$\int_{\mathbb{R}_0} k_2 x G_2(dx)/\rho$	1.0000	1.1393	1.1369	0.1918	0.6123
$\int_{\mathbb{R}_0} k_1^2 x^2 G_1(dx)$	0.1000	0.0817	0.0809	0.0223	0.8169
$\sqrt{\int_{\mathbb{R}_0} k_2^2 x^2 G_2(dx)}$	0.3240	0.4074	0.3975	0.1817	0.3245
$\sqrt{\int_{\mathbb{R}_0} k_1^4 x^4 G_1(dx)}$	0.5477	0.4548	0.4484	0.1926	0.8657
Panel C. GMM with RV and BV from high-frequency data for $M = 300$					
ρ	0.0700	0.0477	0.0469	0.0331	0.4311
$\int_{\mathbb{R}_0} k_2 x G_2(dx)/\rho$	1.0000	1.0777	1.0724	0.1328	0.7492
$\int_{\mathbb{R}_0} k_1^2 x^2 G_1(dx)$	0.1000	0.0853	0.0852	0.0179	1.0073
$\sqrt{\int_{\mathbb{R}_0} k_2^2 x^2 G_2(dx)}$	0.3240	0.2932	0.2946	0.1075	0.5088
$\sqrt{\int_{\mathbb{R}_0} k_1^4 x^4 G_1(dx)}$	0.5477	0.4951	0.4862	0.1340	1.1843
Panel D. GMM with RV and TV from high-frequency data for $M = 100$					
ρ	0.0700	0.0732	0.0645	0.0514	0.2050
$\int_{\mathbb{R}_0} k_2 x G_2(dx)/\rho$	1.0000	1.1011	1.0968	0.1687	0.5982
$\int_{\mathbb{R}_0} k_1^2 x^2 G_1(dx)$	0.1000	0.1036	0.1027	0.0150	0.7106
$\sqrt{\int_{\mathbb{R}_0} k_2^2 x^2 G_2(dx)}$	0.3240	0.4118	0.4013	0.1980	0.2952
$\sqrt{\int_{\mathbb{R}_0} k_1^4 x^4 G_1(dx)}$	0.5477	0.5243	0.5079	0.1671	0.8830

Table 4: Monte Carlo Results for the case $\alpha_p = 0.0$

Parameter	True value	Mean	Median	RMSE	Asy./Small sample s.e.
Panel E. GMM with RV and TV from high-frequency data for $M = 300$					
ρ	0.0700	0.0474	0.0460	0.349	0.3954
$\int_{\mathbb{R}_0} k_2 x G_2(dx)/\rho$	1.0000	1.0604	1.0562	0.1249	0.7382
$\int_{\mathbb{R}_0} k_1^2 x^2 G_1(dx)$	0.1000	0.0970	0.0969	0.0120	0.8944
$\sqrt{\int_{\mathbb{R}_0} k_2^2 x^2 G_2(dx)}$	0.3240	0.2892	0.2871	0.1163	0.4721
$\sqrt{\int_{\mathbb{R}_0} k_1^4 x^4 G_1(dx)}$	0.5477	0.5280	0.5106	0.1325	1.1148
Panel F. GMM with RV and TV from high-frequency data for $M = 100$ **and asymptotic refinement**					
ρ	0.0700	0.0713	0.0707	0.0106	0.8283
$\int_{\mathbb{R}_0} k_2 x G_2(dx)/\rho$	1.0000	0.9771	0.9727	0.0943	0.8836
$\int_{\mathbb{R}_0} k_1^2 x^2 G_1(dx)$	0.1000	0.1033	0.1029	0.0106	1.0272
$\sqrt{\int_{\mathbb{R}_0} k_2^2 x^2 G_2(dx)}$	0.3240	0.2956	0.2910	0.0680	0.8480
$\sqrt{\int_{\mathbb{R}_0} k_1^4 x^4 G_1(dx)}$	0.5477	0.5527	0.5349	0.1131	1.2924
Panel G. GMM with RV and TV from high-frequency data for $M = 300$ **and asymptotic refinement**					
ρ	0.0700	0.0716	0.0713	0.0134	0.7909
$\int_{\mathbb{R}_0} k_2 x G_2(dx)/\rho$	1.0000	0.9879	0.9836	0.0894	0.9115
$\int_{\mathbb{R}_0} k_1^2 x^2 G_1(dx)$	0.1000	0.0974	0.0969	0.0103	1.0375
$\sqrt{\int_{\mathbb{R}_0} k_2^2 x^2 G_2(dx)}$	0.3240	0.3043	0.2982	0.0645	0.8521
$\sqrt{\int_{\mathbb{R}_0} k_1^4 x^4 G_1(dx)}$	0.5477	0.5410	0.5275	0.1139	1.2844

Table 5: Monte Carlo Results for the case $\alpha_p = 0.1$

Parameter	True value	Mean	Median	RMSE	Asy./Small sample s.e.
Panel A. GMM with QV and IV					
ρ	0.0700	0.0722	0.0727	0.0124	0.8622
$\int_{\mathbb{R}_0} k_2 x G_2(dx)/\rho$	1.0000	0.9879	0.9869	0.0866	0.9413
$\int_{\mathbb{R}_0} k_1^2 x^2 G_1(dx)$	0.1000	0.1003	0.0994	0.0108	0.9446
$\sqrt{\int_{\mathbb{R}_0} k_2^2 x^2 G_2(dx)}$	0.3240	0.3108	0.3050	0.0591	0.9098
$\sqrt{\int_{\mathbb{R}_0} k_1^4 x^4 G_1(dx)}$	0.5906	0.5753	0.5517	0.1287	0.8293
Panel B. GMM with RV and BV from high-frequency data for $M = 100$					
ρ	0.0700	0.0707	0.0632	0.0446	0.2358
$\int_{\mathbb{R}_0} k_2 x G_2(dx)/\rho$	1.0000	1.1461	1.1379	0.1992	0.5960
$\int_{\mathbb{R}_0} k_1^2 x^2 G_1(dx)$	0.1000	0.0815	0.0803	0.0226	0.7788
$\sqrt{\int_{\mathbb{R}_0} k_2^2 x^2 G_2(dx)}$	0.3240	0.4182	0.4052	0.1888	0.3203
$\sqrt{\int_{\mathbb{R}_0} k_1^4 x^4 G_1(dx)}$	0.5906	0.4836	0.4780	0.2086	0.5918
Panel C. GMM with RV and BV from high-frequency data for $M = 300$					
ρ	0.0700	0.0476	0.0461	0.0339	0.4141
$\int_{\mathbb{R}_0} k_2 x G_2(dx)/\rho$	1.0000	1.0764	1.0714	0.1304	0.7640
$\int_{\mathbb{R}_0} k_1^2 x^2 G_1(dx)$	0.1000	0.0854	0.0847	0.0182	0.9274
$\sqrt{\int_{\mathbb{R}_0} k_2^2 x^2 G_2(dx)}$	0.3240	0.2927	0.2935	0.1098	0.4982
$\sqrt{\int_{\mathbb{R}_0} k_1^4 x^4 G_1(dx)}$	0.5906	0.5264	0.5142	0.1563	0.7438
Panel D. GMM with RV and TV from high-frequency data for $M = 100$					
ρ	0.0700	0.0743	0.0653	0.0508	0.2079
$\int_{\mathbb{R}_0} k_2 x G_2(dx)/\rho$	1.0000	1.1111	1.1059	0.1783	0.5793
$\int_{\mathbb{R}_0} k_1^2 x^2 G_1(dx)$	0.1000	0.1041	0.1022	0.0159	0.6668
$\sqrt{\int_{\mathbb{R}_0} k_2^2 x^2 G_2(dx)}$	0.3240	0.4221	0.4088	0.2048	0.2915
$\sqrt{\int_{\mathbb{R}_0} k_1^4 x^4 G_1(dx)}$	0.5906	0.5620	0.5409	0.1838	0.5836

Table 6: Monte Carlo Results for the case $\alpha_p = 0.1$

Parameter	True value	Mean	Median	RMSE	Asy./Small sample s.e.
Panel E. GMM with RV and TV from high-frequency data for $M = 300$					
ρ	0.0700	0.0469	0.0436	0.0360	0.3811
$\int_{\mathbb{R}_0} k_2 x G_2(dx)/\rho$	1.0000	1.0582	1.0568	0.1224	0.7499
$\int_{\mathbb{R}_0} k_1^2 x^2 G_1(dx)$	0.1000	0.0971	0.0963	0.0122	0.8573
$\sqrt{\int_{\mathbb{R}_0} k_2^2 x^2 G_2(dx)}$	0.3240	0.2861	0.2840	0.1190	0.4646
$\sqrt{\int_{\mathbb{R}_0} k_1^4 x^4 G_1(dx)}$	0.5906	0.5650	0.5392	0.1474	0.7299
Panel F. GMM with RV and TV from high-frequency data for $M = 100$ and asymptotic refinement					
ρ	0.0700	0.0702	0.0698	0.0119	0.8839
$\int_{\mathbb{R}_0} k_2 x G_2(dx)/\rho$	1.0000	0.9738	0.9732	0.0929	0.9064
$\int_{\mathbb{R}_0} k_1^2 x^2 G_1(dx)$	0.1000	0.1035	0.1025	0.0115	0.9359
$\sqrt{\int_{\mathbb{R}_0} k_2^2 x^2 G_2(dx)}$	0.3240	0.2920	0.2901	0.0673	0.8852
$\sqrt{\int_{\mathbb{R}_0} k_1^4 x^4 G_1(dx)}$	0.5906	0.5889	0.5664	0.1283	0.8261
Panel G. GMM with RV and TV from high-frequency data for $M = 300$ and asymptotic refinement					
ρ	0.0700	0.0707	0.0702	0.0124	0.8483
$\int_{\mathbb{R}_0} k_2 x G_2(dx)/\rho$	1.0000	0.9860	0.9838	0.0884	0.9251
$\int_{\mathbb{R}_0} k_1^2 x^2 G_1(dx)$	0.1000	0.0975	0.0966	0.0124	0.9624
$\sqrt{\int_{\mathbb{R}_0} k_2^2 x^2 G_2(dx)}$	0.3240	0.3021	0.2977	0.0617	0.9082
$\sqrt{\int_{\mathbb{R}_0} k_1^4 x^4 G_1(dx)}$	0.5906	0.5775	0.5534	0.1286	0.8287

Table 7: Monte Carlo Results for the case $\alpha_p = 0.5$

Parameter	True value	Mean	Median	RMSE	Asy./Small sample s.e.
Panel A. GMM with QV and IV					
ρ	0.0700	0.0735	0.0735	0.0127	0.8622
$\int_{\mathbb{R}_0} k_2 x G_2(dx)/\rho$	1.0000	0.9925	0.9913	0.0866	0.9358
$\int_{\mathbb{R}_0} k_1^2 x^2 G_1(dx)$	0.1000	0.0999	0.0992	0.0119	0.9825
$\sqrt{\int_{\mathbb{R}_0} k_2^2 x^2 G_2(dx)}$	0.3240	0.3129	0.3068	0.0571	0.9358
$\sqrt{\int_{\mathbb{R}_0} k_1^4 x^4 G_1(dx)}$	0.6572	0.6292	0.5972	0.1840	1.2780
Panel B. GMM with RV and BV from high-frequency data for $M = 100$					
ρ	0.0700	0.0674	0.0616	0.0423	0.2493
$\int_{\mathbb{R}_0} k_2 x G_2(dx)/\rho$	1.0000	1.1443	1.1361	0.1928	0.6324
$\int_{\mathbb{R}_0} k_1^2 x^2 G_1(dx)$	0.1000	0.0797	0.0783	0.0249	0.8063
$\sqrt{\int_{\mathbb{R}_0} k_2^2 x^2 G_2(dx)}$	0.3240	0.4076	0.3971	0.1823	0.3235
$\sqrt{\int_{\mathbb{R}_0} k_1^4 x^4 G_1(dx)}$	0.6572	0.5120	0.4979	0.2707	1.0168
Panel C. GMM with RV and BV from high-frequency data for $M = 300$					
ρ	0.0700	0.0488	0.0469	0.0326	0.4241
$\int_{\mathbb{R}_0} k_2 x G_2(dx)/\rho$	1.0000	1.0818	1.0764	0.1295	0.8044
$\int_{\mathbb{R}_0} k_1^2 x^2 G_1(dx)$	0.1000	0.0834	0.0824	0.0206	0.9583
$\sqrt{\int_{\mathbb{R}_0} k_2^2 x^2 G_2(dx)}$	0.3240	0.2975	0.2931	0.1092	0.4949
$\sqrt{\int_{\mathbb{R}_0} k_1^4 x^4 G_1(dx)}$	0.6572	0.5720	0.5467	0.2015	1.2717
Panel D. GMM with RV and TV from high-frequency data for $M = 100$					
ρ	0.0700	0.0715	0.0637	0.0489	0.2151
$\int_{\mathbb{R}_0} k_2 x G_2(dx)/\rho$	1.0000	1.1099	1.1064	0.1716	0.6132
$\int_{\mathbb{R}_0} k_1^2 x^2 G_1(dx)$	0.1000	0.1032	0.1014	0.0174	0.6837
$\sqrt{\int_{\mathbb{R}_0} k_2^2 x^2 G_2(dx)}$	0.3240	0.4137	0.4009	0.1982	0.2964
$\sqrt{\int_{\mathbb{R}_0} k_1^4 x^4 G_1(dx)}$	0.6572	0.6161	0.5911	0.2362	0.9989

Table 8: Monte Carlo Results for the case $\alpha_p = 0.5$

Parameter	True value	Mean	Median	RMSE	Asy./Small sample s.e.
Panel E. GMM with RV and TV from high-frequency data for $M = 300$					
ρ	0.0700	0.0488	0.0464	0.0346	0.3853
$\int_{\mathbb{R}_0} k_2 x G_2(dx)/\rho$	1.0000	1.0641	1.0616	0.1222	0.7765
$\int_{\mathbb{R}_0} k_1^2 x^2 G_1(dx)$	0.1000	0.0955	0.0943	0.0143	0.8661
$\sqrt{\int_{\mathbb{R}_0} k_2^2 x^2 G_2(dx)}$	0.3240	0.2936	0.2889	0.1197	0.4529
$\sqrt{\int_{\mathbb{R}_0} k_1^4 x^4 G_1(dx)}$	0.6572	0.6168	0.5857	0.2043	1.1605
Panel F. GMM with RV and TV from high-frequency data for $M = 100$ **and asymptotic refinement**					
ρ	0.0700	0.0706	0.0695	0.0124	0.8483
$\int_{\mathbb{R}_0} k_2 x G_2(dx)/\rho$	1.0000	0.9821	0.9773	0.0891	0.9261
$\int_{\mathbb{R}_0} k_1^2 x^2 G_1(dx)$	0.1000	0.1022	0.1014	0.0124	0.9825
$\sqrt{\int_{\mathbb{R}_0} k_2^2 x^2 G_2(dx)}$	0.3240	0.2953	0.2922	0.0660	0.8822
$\sqrt{\int_{\mathbb{R}_0} k_1^4 x^4 G_1(dx)}$	0.6572	0.6434	0.6112	0.1810	1.2872
Panel G. GMM with RV and TV from high-frequency data for $M = 300$ **and asymptotic refinement**					
ρ	0.0700	0.0715	0.0711	0.0129	0.8218
$\int_{\mathbb{R}_0} k_2 x G_2(dx)/\rho$	1.0000	0.9940	0.9920	0.0854	0.9479
$\int_{\mathbb{R}_0} k_1^2 x^2 G_1(dx)$	0.1000	0.0959	0.0951	0.0125	0.9908
$\sqrt{\int_{\mathbb{R}_0} k_2^2 x^2 G_2(dx)}$	0.3240	0.3047	0.2996	0.0615	0.8974
$\sqrt{\int_{\mathbb{R}_0} k_1^4 x^4 G_1(dx)}$	0.6572	0.6324	0.6004	0.1852	1.2662

Table 9: Estimation Results

Parameter	Estimate	Parameter	Estimate
Panel A. Affine Jump-Diffusion SV Model			
κ_1	1.5961 (0.4654)	ζ	0.6058 (0.0507)
$\overline{\sigma}_1$	0.3399 (0.2894)	$\int_{\mathbb{R}_0^2} h^2(\mathbf{x})G(d\mathbf{x})$	0.1135 (0.0085)
κ_2	0.0021 (0.0093)	$\int_{\mathbb{R}_0^2} h^4(\mathbf{x})G(d\mathbf{x})$	0.0001 (0.0054)
$\overline{\sigma}_2$	0.2647 (0.2040)		
GMM test of overidentifying restrictions			40.5980
d.o.f			(4)
p-value			0.0000
Panel B. CARMA Jump-Driven SV Model			
ζ	0.7990 (0.0512)	$\int_{\mathbb{R}_0^2} k^2(\mathbf{x})G(d\mathbf{x})$	2.3925 (0.4780)
b_0	0.2350 (0.0458)	$\int_{\mathbb{R}_0^2} h^2(\mathbf{x})G(d\mathbf{x})$	0.1513 (0.0159)
$-\rho_1$	0.0397 (0.0097)	$\int_{\mathbb{R}_0^2} h^4(\mathbf{x})G(d\mathbf{x})$	0.2718 (0.1297)
$-\rho_2$	1.5476 (0.2301)	$\int_{\mathbb{R}_0^2} h^2(\mathbf{x})k(\mathbf{x})G(d\mathbf{x})$	0.5386 (0.1392)
GMM test of overidentifying restrictions			0.9732
d.o.f			(3)
p-value			0.8077

Note: In the estimation: (1) we set $\zeta := \overline{V}_1 + \overline{V}_2$ and $\overline{\sigma}_i := \sigma_{iv}\sqrt{\frac{\overline{V}_i}{2\kappa_i}}$ for $i = 1, 2$ and impose the stationarity conditions $\overline{\sigma}_1 + \overline{\sigma}_2 < \zeta$ and $\kappa_i > 0$ for $i = 1, 2$ for the Affine Jump-Diffusion Model and (2) we set $\zeta := \frac{b_0}{\rho_1\rho_2}\int_{\mathbb{R}_0^2} h(\mathbf{x})G(d\mathbf{x})$ and impose the stationarity conditions (see Todorov and Tauchen (2006)) $b_0 > -\max\{\rho_1, \rho_2\}$ and $\rho_i < 0$ for $i = 1, 2$ for the CARMA Jump-Driven SV model. The model is estimated using GMM-type estimator with moment conditions specified in Section 4. The asymptotic variance-covariance matrix, used for calculating the optimal weighting matrix, is estimated using Parzen weights with a lag length of 80. Standard errors for the parameter estimates are reported in parentheses.

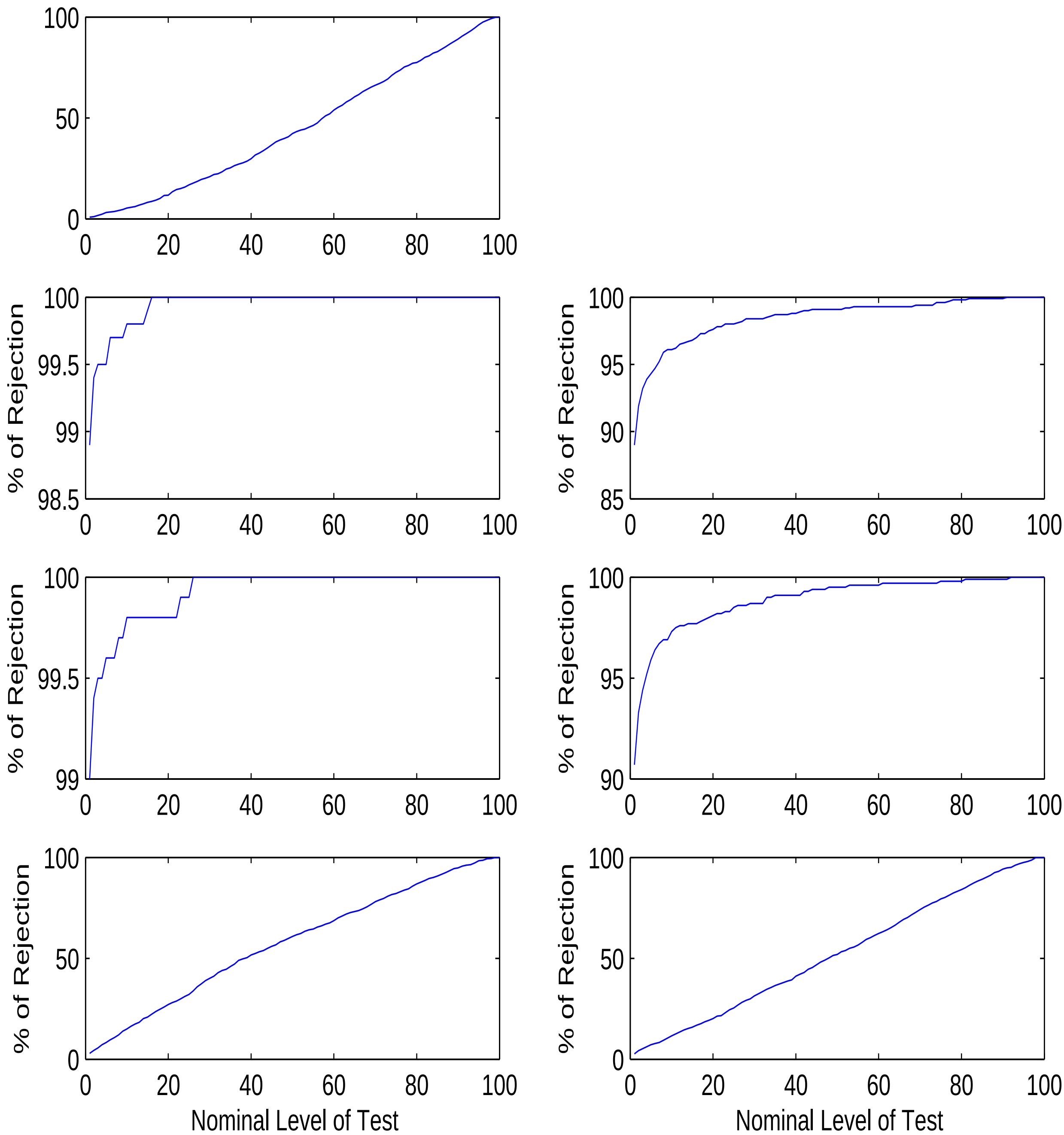

Figure 1: Percentage of Rejection versus Nominal Level of GMM Test of overidentifying restriction for the Monte Carlo study, case compound Poisson price jumps. First row - estimation with QV and IV; second row - estimation with RV and BV; third row - estimation with RV and TV; forth row - estimation with RV and TV and asymptotic refinement. For the second, third and forth row the left side corresponds to the case $M = 100$ and the right side to the case $M = 300$.

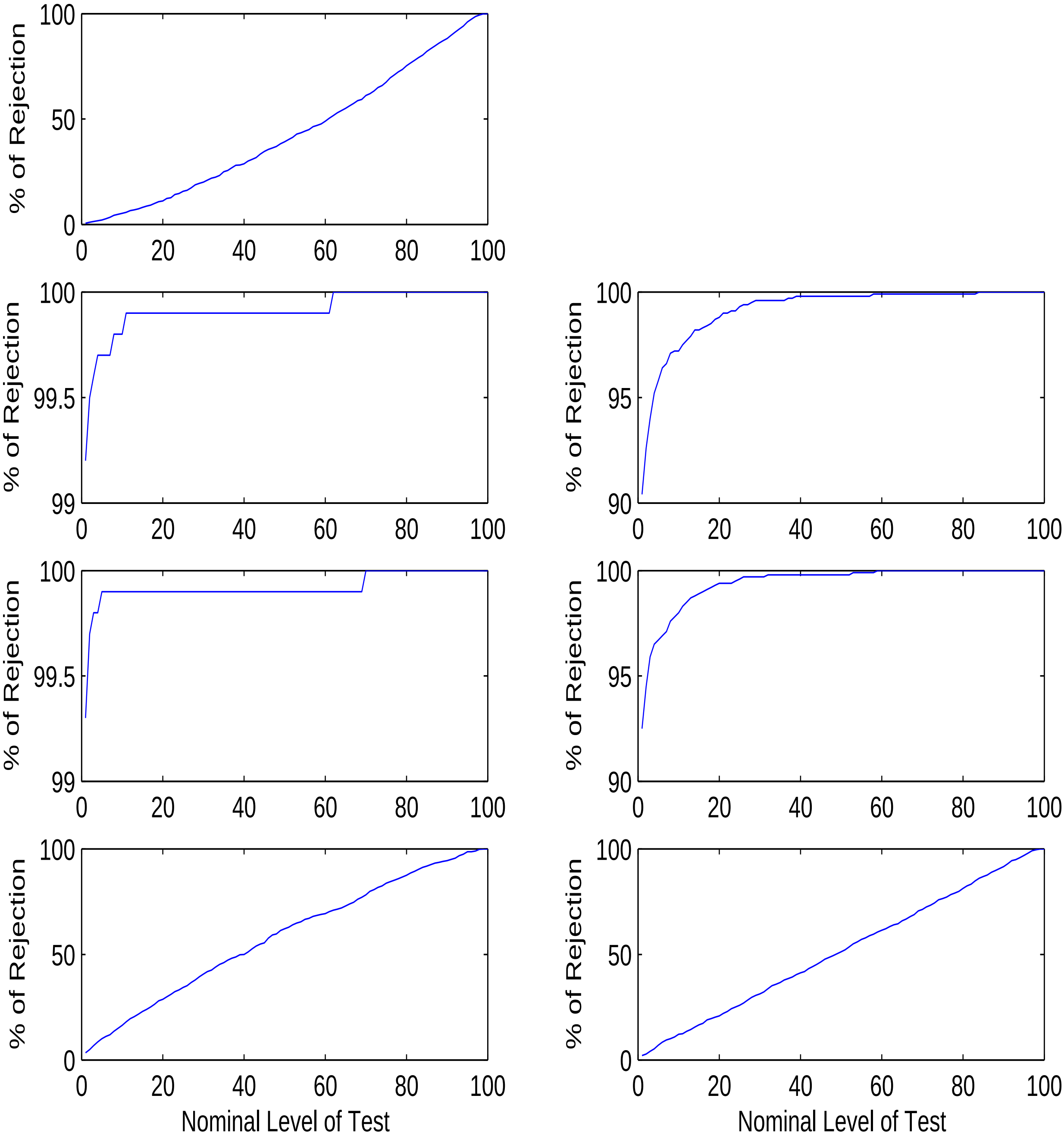

Figure 2: Percentage of Rejection versus Nominal Level of GMM Test of overidentifying restriction for the Monte Carlo study, case $\alpha_p = 0.0$. First row - estimation with QV and IV; second row - estimation with RV and BV; third row - estimation with RV and TV; forth row - estimation with RV and TV and asymptotic refinement. For the second, third and forth row the left side corresponds to the case $M = 100$ and the right side to the case $M = 300$.

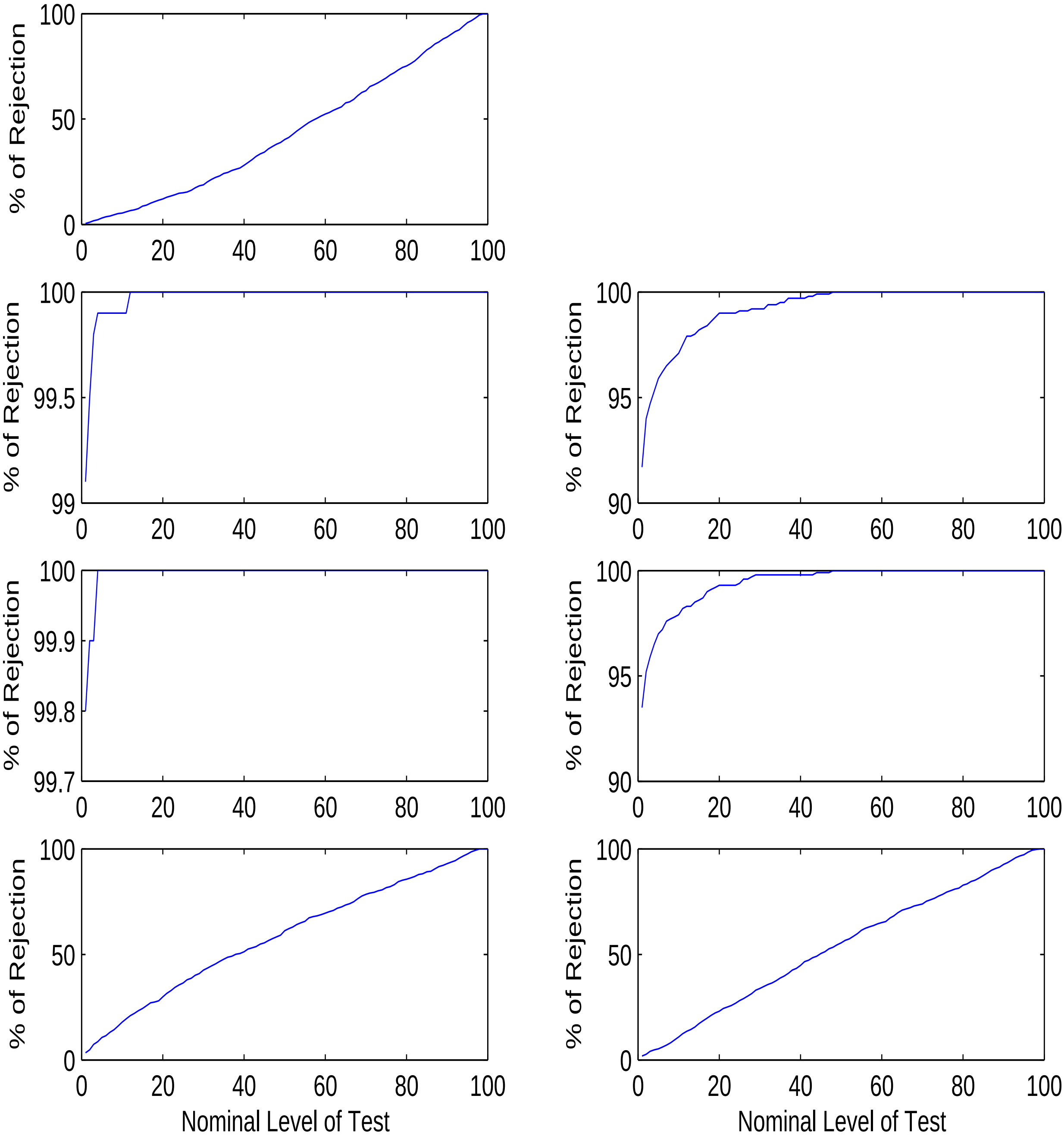

Figure 3: Percentage of Rejection versus Nominal Level of GMM Test of overidentifying restriction for the Monte Carlo study, case $\alpha_p = 0.1$. First row - estimation with QV and IV; second row - estimation with RV and BV; third row - estimation with RV and TV; forth row - estimation with RV and TV and asymptotic refinement. For the second, third and forth row the left side corresponds to the case $M = 100$ and the right side to the case $M = 300$.

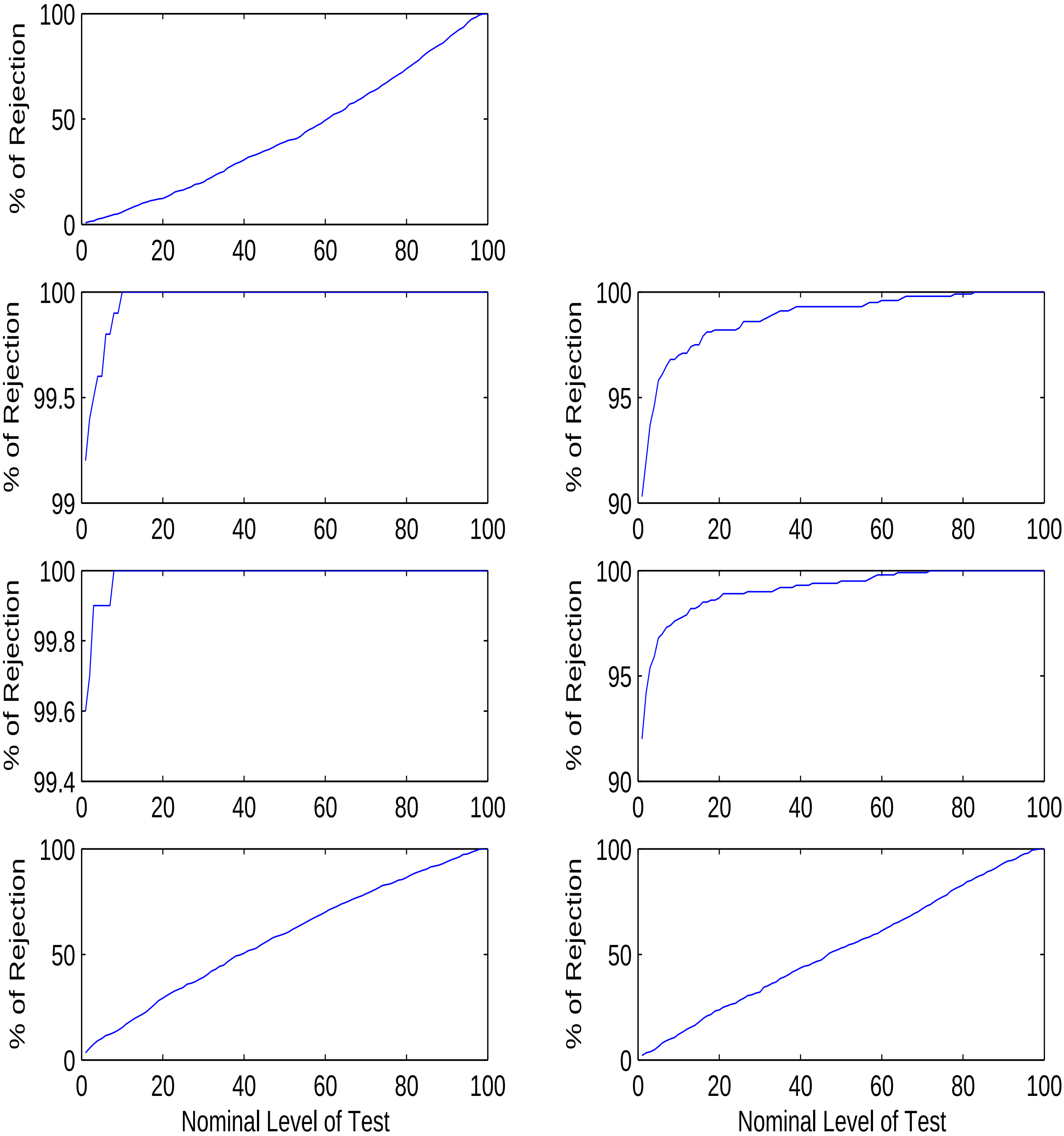

Figure 4: Percentage of Rejection versus Nominal Level of GMM Test of overidentifying restriction for the Monte Carlo study, case $\alpha_p = 0.5$. First row - estimation with QV and IV; second row - estimation with RV and BV; third row - estimation with RV and TV; forth row - estimation with RV and TV and asymptotic refinement. For the second, third and forth row the left side corresponds to the case $M = 100$ and the right side to the case $M = 300$.

References

Ait-Sahalia, Y. and J. Jacod (2008a). Estimating the Degree of Activity of Jumps in High Frequency Financial Data. *Annals of Statistics*, forthcoming.

Ait-Sahalia, Y. and J. Jacod (2008b). Testing for Jumps in a Discretely Observed Process. *Annals of Statistics, forthcoming*.

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*.

Barndorff-Nielsen, O. E., S. Graversen, J. Jacod, M. Podolskij, and N. Shephard (2005). A Central Limit Theorem for Realised Power and Bipower Variations of Continuous Semimartingales. In Y. Kabanov and R. Lipster (Eds.), *From Stochastic Analysis to Mathematical Finance, Festschrift for Albert Shiryaev*. Springer.

Barndorff-Nielsen, O. E., P. Hansen, A. Lunde, and N. Shephard (2006). Designing Realized Kernels to Measure the Ex-post Variation of Equity Prices in the Presence of Noise. Working paper, Nuffield College, Oxford.

Barndorff-Nielsen, O. E. and N. Shephard (2001). Non- Gaussian Ornstein-Uhlenbeck-based Models and Some of Their Applicaions in Financial Economics. *Journal of the Royal Statistical Society: Series B 63*, 167–241.

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 (2006a). Econometrics of Testing for Jumps in Financial Economics using Bipower Variation. *Journal of Financial Econometrics 4*, 1–30.

Barndorff-Nielsen, O. E. and N. Shephard (2006b). Impact of Jumps on Returns and Realised Variances: Econometric Analysis of Time-Deformed Lévy Processes. *Journal of Econometrics 131*, 217–252.

Barndorff-Nielsen, O. E., N. Shephard, and M. Winkel (2006). Limit Theorems for Multipower Variation in the Presence of Jumps in Financial Econometrics. *Stochastic Processes and Their Applications 116*, 796–806.

Blumenthal, R. and R. Getoor (1961). Sample Functions of Stochastic Processes with Independent Increments. *Journal of Math. Mech. 10*, 493–516.

Bollerslev, T. and H. Zhou (2002). Estimating Stochastic Volatility Diffusion using Conditional Moments of Integrated Volatility. *Journal of Econometrics 109*, 33–65.

Brockwell, P. (2001a). Lévy -Driven CARMA Processes. *Ann.Inst.Statist.Math 53*, 113–124.

Brockwell, P. (2001b). Continuous-Time ARMA Processes. In D. Shanbhag and C. Rao (Eds.), *Handbook of Statistics*, Volume 19. North-Holland.

Bruti-Liberati, N. and E. Platen (2007). Strong Approximations of Stochastic Differential Equations with Jumps. *Journal of Computational and Applied Mathematics 205*, 982–1001.

Corradi, V. and W. Distaso (2006). Semiparametric Comparision of Stochastic Volatility Models Using Realized Measures. *Review of Economic Studies 73*, 635–667.

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

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

Jacod, J. (1979). *Calcul Stochastique et Problèmes de Martingales*. Lecture notes in Mathemtatics 714. Berlin Heidelberg New York: Springer-Verlag.

Jacod, J. (2004). The Euler Scheme for Lévy-Driven Stochastic Differential Equations: Limit theorems. *Annals of Probability 32*, 1830–1872.

Jacod, J. (2006). Asymptotic Properties of Power Variations and Associated Functionals of Semimartingales. *Stochastic Processes and their Applications, forthcoming.*

Jacod, J., Y. Li, P. Mykland, M. Podolskij, and M. Vettter (2007). Microstructure Noise in the Continuous Case: The Pre-Averaging Approach. Working paper.

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

Jiang, G. and R. Oomen (2005). A New Test for Jumps in Asset Prices. Working paper.

Lee, S. and P. Mykland (2007). Jumps in Financial Markets: A New Nonparametric Test and Jump Dynamics. *Review of Financial Studies 20*, forthcoming.

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

Masuda, H. (2004). On Multidimensional Ornstein-Uhlenbeck Processes Driven by a General Lévy Process. *Bernoulli 10*, 97–120.

Masuda, H. (2007). Ergodicity and Exponential β-mixing Bounds for Multidimensional Diffusions with Jumps. *Stochastic Processes and their Applications 117*, 35–56.

Newey, W. and D. McFadden (1994). Large Sample Estimation and Hypothesis Testing. In R. Engle and D. McFadden (Eds.), *Handbook of Econometrics*, Volume 4, pp. 2113–2241. Amsterdam: North-Holland.

Podolskij, M. and M. Vetter (2007). Estimation of Volatility Functionals in the Simultaneuous Presence of Microstructure Noise and Jumps. Working paper.

Protter, P. and D. Talay (1997). The Euler Scheme for Lévy-Driven Stochastic Differential Equations. *Annals of Probability 25*, 393–423.

Rajput, B. and J. Rosiński (1989). Spectral Representation of Infinitely Divisible Processes Vectors. *Probability Theory and Related Fields 82*, 451–487.

Rosiński, J. (2001). Series Representations of Lévy Processes from the Perspective of Point Processes. In O. E. Barndorff-Nielsen, T. Mikosch, and S. Resnick (Eds.), *Lévy Processes-Theory and Applications*. Boston: Birkhäser.

Rosiński, J. (2007). Tempering Stable Processes. *Stochastic Processes and their Applications 117*, 677–707.

Todorov, V. (2008). Econometric Analysis of Jump-Driven Stochastic Volatility Models. *Journal of Econometrics*, forthcoming.

Todorov, V. and G. Tauchen (2006). Simulation Methods for Lévy -Driven CARMA Stochastic Volatility Models. *Journal of Business and Economic Statistics 24*, 455–469.

Todorov, V. and G. Tauchen (2008). Activity Signature Functions for High-Frequency Data Analysis. Working paper.

Wooldridge, J. (1994). Estimation and Inference for Dependent Processes. In R. Engle and D. McFadden (Eds.), *Handbook of Econometrics*, Volume 4, pp. 2639–2738. Amsterdam: North-Holland.

Zhang, L. (2006). Efficient Estimation of Stochastic Volatility using Noisy Observations: A Multi-scale Approach. *Bernoulli 12*, 1019–1043.

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