Journal of Econometrics 169 (2012) 94–113

Contents lists available at SciVerse ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Time-varying leverage effects[☆]

Federico M. Bandi[a,b,*], Roberto Renò[c]

[a] Carey Business School, Johns Hopkins University, 100 International Drive, Baltimore, MD 21202, USA
[b] Edhec-Risk Institute, 393-400 Promenade des Anglais, BP 3116, 06202 Nice Cedex 3, France
[c] Dipartimento di Economia Politica, Università di Siena, Piazza San Francesco 7, 53100, Siena, Italy

ARTICLE INFO

Article history:
Available online 21 January 2012

ABSTRACT

Vast empirical evidence points to the existence of a negative correlation, named "leverage effect", between shocks to variance and shocks to returns. We provide a nonparametric theory of leverage estimation in the context of a continuous-time stochastic volatility model with jumps in returns, jumps in variance, or both. Leverage is defined as a flexible function of the state of the firm, as summarized by the spot variance level. We show that its point-wise functional estimates have asymptotic properties (in terms of rates of convergence, limiting biases, and limiting variances) which crucially depend on the likelihood of the individual jumps and co-jumps as well as on the features of the jump size distributions. Empirically, we find economically important time-variation in leverage with more negative values associated with higher variance levels.

1. Introduction

Shocks to returns have been found to be negatively correlated with shocks to variance. Due to its original economic justification in the case of common stocks (negative shocks to returns decrease the value of the firm's equity and increase the debt-to-equity ratio, i.e., financial leverage, thereby rendering equity riskier and more volatile), this stylized fact has been termed "leverage effect". While this classical logic derived in a Modigliani–Miller world is somewhat appealing, its empirical validity has been questioned. More generally, the economics of leverage effects continues to remain very controversial.

The empirical relevance of (alternative forms of) leverage effects has, however, been broadly established. The time-varying variance literature has emphasized the significance of feedback effects between returns and variance changes in a variety of parametric settings. Fundamental contributions in terms of modelling and pricing have been provided both in continuous time (Andersen et al., 2002, Bakshi et al., 1997 and Eraker et al., 2003, *inter alia*) and in discrete time (Engle and Ng, 1993, Glosten et al., 1993, Harvey and Shephard, 1996, Jacquier et al., 2004, and Yu, 2008, among others). Yu (2005) offers a review and an insightful assessment of the extant literature in discrete time.

The "leverage parameter" (i.e., the correlation between shocks to prices and shocks to variance) is generally assumed to be a constant value. Some recent work has, however, emphasized that there may be important asymmetries in the way in which variance responds to price changes. Figlewski and Wang (2000) and Yu (2008), for instance, stress that, in the presence of positive shocks to prices, return variance may not change (or even change positively) whereas negative shocks will likely lead to an increase in variance, coherently with traditional – negative – leverage effects. Allowing for changing levels of correlation between innovations to prices and innovations to variance may, therefore, be empirically important. We propose an alternative, continuous-time, framework to do so.

Specifically, we provide a nonparametric treatment of leverage estimation in the context of a stochastic volatility, jump-diffusion, model with discontinuities in returns, variance, or both. The model is (semi)nonparametrically specified. Parametric assumptions are solely imposed on the distributions of the jump sizes for identification. Importantly, we allow leverage to be a function of the state of the firm and, hence, time-varying. In our model the state of the firm is summarized by spot variance (or spot volatility). This approach is natural in continuous-time stochastic volatility models – and effectively extends them – since spot variance is used as a conditioning variable both in the return equation (where the return drift may depend on variance as implied by the presence

☆ We thank Robert Taylor, Matteo Bonato, Cecilia Mancini, the Co-Editors (Roberto Mariano, Zhijie Xiao, and Jun Yu) and two anonymous referees for their useful comments and suggestions. We are also grateful to the participants at the Festschrift in honour of Peter Phillips (Singapore, July 14-15, 2008), the 2010 Winter Meetings of the Econometric Society (Atlanta, January 3-5, 2010) and various other institutions for helpful discussions. Roberto Renò gratefully acknowledges financial support from The Castilla and León Council of Education and Science, grant VA016A10-1.

* Corresponding author at: Carey Business School, Johns Hopkins University, 100 International Drive, Baltimore, MD 21202, USA.
E-mail addresses: fbandi1@jhu.edu (F.M. Bandi), reno@unisi.it (R. Renò).

0304-4076/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2012.01.010

of risk-return trade-offs) and in the variance equation (where the variance drift and diffusion are generally modelled as functions of the variance state). It is also economically meaningful. Below we show that, if one were to take seriously all of the implications of the most classical justifications for leverage (as in a Modigliani–Miller economy, as mentioned above, for example), leverage should be a function of the underlying spot variance level in that the impact of negative shocks to returns on variance should depend on the debt-to-equity ratio but the latter should have a positive relation with the spot variance level itself.

From a theoretical standpoint, we show that the limiting features of our nonparametric leverage estimator crucially depend on the continuity properties of the price and variance processes. Several cases are considered: absence of jumps in either process, jumps in returns, jumps in variance, independent jumps in returns and variance, contemporaneous jumps (or co-jumps) in returns and variance. We show that the fastest convergence rate to a (mixed) normal distribution arises in the absence of jumps in variance. The presence of jumps in returns (without jumps in variance) does not affect the rate of convergence of the estimator as compared to the case with no jumps. However, it does affect asymptotic efficiency negatively by adding an additional term to the leverage estimator's limiting variance. The case of jumps in variance (either without jumps in returns or with independent return jumps) is quite different in that consistent estimation of the variance process' diffusion function (namely, the variance of variance) can only be conducted at a slower rate. This slower rate reduces the speed of convergence of the kernel leverage estimator. In particular, its limiting distribution is now driven by the asymptotic features of the spot variance's diffusion function estimator. Finally, allowing for co-jumps in returns and variance may yield inconsistency of the leverage estimator unless, for example, the jump sizes are uncorrelated and the jumps in returns are mean zero. We define a novel (generalized) notion of leverage in this case and show that the limiting distribution of the generalized leverage estimator is completely driven by the features of the price/variance discontinuities.

Importantly, our limit theory hinges on weaker conditions than stationarity. We solely assume recurrence of the spot variance process, thereby allowing for a considerable amount of variance persistence in any given sample. Finally, kernel estimation of leverage effects requires suitable filtering of the spot variance process. We do so by considering spot variance estimates obtained by virtue of high-frequency asset price data as in Bandi and Renò (2008), BR henceforth. In particular, we report conditions under which the estimation error induced by spot variance estimation can be made asymptotically negligible for the purpose of leverage estimation.

Our empirical findings hinge on S&P500 futures data and are, therefore, for a broad-based US market index. We find important time-variation (as a function of spot variance) in the correlation between price and variance shocks. As conjectured, leverage increases (i.e., becomes more negative) with the variance level. We confirm the robustness of this result by estimating parametric models with time-varying leverage (in Section 2) and by addressing important finite sample issues which may arise when estimating nonparametrically continuous-time stochastic volatility models with feedback effects (as induced by risk-return trade-offs) and estimated regressors (in Section 9).

Bandi and Renò (2008) have recently introduced a novel nonparametric approach to the estimation of stochastic volatility models with jumps (in returns and in variance). The approach does not hinge on the filtering of the latent spot variance process by virtue of simulation methods relying on low frequency return data (as is common in the parametric literature) but on the preliminary filtering of spot variance using high-frequency return data. The resulting procedure is (semi)nonparametric in nature in spite of the unobservability of spot variance. From a methodological standpoint, the present paper relates to Bandi and Renò (2008) in that the price/variance evolutions are jump-diffusive, the dynamics are estimated nonparametrically, and the preliminary spot variance estimates are derived from intra-daily return data. Our exclusive focus on leverage, along with the attention that leverage has been receiving in the literature, however, make this contribution of separate interest from our previous work. In the context of leverage evaluation, and differently from Bandi and Renò (2008), (1) we provide a complete limit theory for nonparametric leverage estimators allowing for increasing layers of complexity – and realism – in the assumed model ranging from simple diffusive structures to structures with co-jumps and correlated jump sizes, (2) we discuss related finite sample issues, (3) we introduce a broader notion of leverage (dubbed "generalized leverage") arising in models in which the jumps are common to the return and the variance process and their jump sizes are correlated (for which we also discuss identification methods), and (4) we implement extensive empirical work validating the time-varying nature of leverage with nonparametric estimates as well as with estimates derived from appropriately-defined (and theoretically justified) reduced-form parametric models.

The paper proceeds as follows. Section 2 provides parametric motivation for allowing leverage to be a flexible function of the spot variance process. This section is meant to introduce our approach in a more familiar setting and show that the leverage dynamics are not a by-product of the use of nonparametric methods. In Section 3, we illustrate how the most traditional economic justification for leverage effects does, in fact, imply dependence of leverage on spot variance and, hence, time-variation of leverage. Section 4 lays out a nonlinear, continuous-time, stochastic volatility model with (possibly correlated) jumps and time-varying leverage. In Section 5, we present the relevant limit theory for an *observable* spot variance process. Section 6 discusses "generalized leverage". Section 7 adapts the theory in Section 5 to the empirically-compelling case of an *estimated* spot variance process. Section 8 provides empirical results pointing to the existence of a higher (more negative) leverage in the presence of higher variance. In Section 9, we address important finite sample concerns and provide further support to changing leverage dynamics. Section 10 offers further perspectives (and directions) by returning to a parametric specification allowing for time-varying leverage. This section relates our approach to extant discrete-time approaches by analysing issues of timing in the estimation of leverage (in the form of contemporaneous versus lagged leverage). It also discusses the relative impact of alternative conditioning variables in the evaluation of time-varying leverage, namely spot volatility (as in this paper), returns (as in Figlewski and Wang, 2000, and Yu, 2008) and signed – by the contemporaneous return – volatility (a possible, alternative, state variable which borrows features from the former two). Section 11 concludes. The proofs are in the Appendix.

2. Time-varying leverage: a parametric model

We begin with a parametric approach allowing for time-varying leverage. In order to do so, we partition the volatility range into N non-overlapping intervals, using knots $0 = \eta_0 < \eta_1 < \cdots < \eta_{N-1} < \eta_N = \infty$,[1] and write

[1] In the empirical work, the intervals are chosen in such a way as to guarantee that they all contain the same number of observations.

Table 1
Parametric models with contemporaneous leverage.

Model: $\sqrt{RV_{t:t-1}} = HAR_{t-1} + r_t \sum_{i=1}^{N} \delta_i \mathbf{1}_{\{\eta_{i-1} \le X \le \eta_i\}} + \varepsilon_t$,					
Conditioning on volatility ($X = \sqrt{RV_{t-1:t-2}}$)					
i	Range	δ_i	Standard error	t-stat	Implied leverage
1	$0.250 < X < 0.536$	−0.226	0.077	−2.93	−0.152
2	$0.536 < X < 0.780$	−0.240	0.037	−6.55	−0.187
3	$0.780 < X < 1.700$	−0.350	0.046	−7.54	−0.273
Conditioning on returns ($X = r_{t-1}$)					
1	$-1.700 < X < -0.043$	−0.478	0.168	−2.85	−0.187
2	$-0.043 < X < 0.061$	−0.374	0.043	−8.59	−0.054
3	$0.061 < X < 1.700$	−0.358	0.063	−5.65	−0.164
Conditioning on signed volatility ($X = \sqrt{RV_{t-1:t-2}} \cdot \text{sign}(r_{t-1})$)					
1	$-1.700 < X < -0.525$	−0.317	0.046	−6.88	−0.155
2	$-0.525 < X < 0.531$	−0.221	0.074	−2.99	−0.143
3	$0.531 < X < 1.700$	−0.321	0.037	−8.76	−0.178

$$\sqrt{RV_{t:t-1}} = \underbrace{\alpha + \beta_1\sqrt{RV_{t-1:t-2}} + \beta_2\sqrt{RV_{t-1:t-6}} + \beta_3\sqrt{RV_{t-1:t-23}}}_{HAR_{t-1}\ \text{component}} + r_t \sum_{i=1}^{N} \delta_i \mathbf{1}_{\left\{\eta_{i-1} \le \sqrt{RV_{t-1:t-2}} \le \eta_i\right\}} + \varepsilon_t, \quad (1)$$

where $\sqrt{RV_{t:t-1}}$ is the square root of an appropriately-chosen realized variance measure between $t-1$ and t, $\sqrt{RV_{t-1:t-k}} = \sqrt{\frac{1}{k-1}\sum_{i=1}^{k-1} RV_{t-i:t-i-1}}$ for $k > 1$, $r_t = \log p_t - \log p_{t-1}$ and ε_t is a forecast error. We infer "implied" leverage over the ith interval ($\widehat{\rho}_i$) by the corresponding estimated coefficient $\widehat{\delta}_i$ provided such a coefficient is re-scaled appropriately. In fact, $\widehat{\delta}_i \approx \frac{\widehat{\text{cov}}\left(\sqrt{RV_{t:t-1}} - HAR_{t-1}, r_t | \eta_{i-1} \le \sqrt{RV_{t-1:t-2}} \le \eta_i\right)}{\widehat{\text{var}}\left(r_t | \eta_{i-1} \le \sqrt{RV_{t-1:t-2}} \le \eta_i\right)}$. Hence, $\widehat{\rho}_i \approx \widehat{\delta_i S_i}$, where the scaling factor $\widehat{S}_i$ is equal to $\frac{\widehat{\text{std}}\left(r_t | \eta_{i-1} \le \sqrt{RV_{t-1:t-2}} \le \eta_i\right)}{\widehat{\text{std}}\left(\sqrt{RV_{t:t-1}} - HAR_{t-1} | \eta_{i-1} \le \sqrt{RV_{t-1:t-2}} \le \eta_i\right)}$.

Importantly, this approach to (possibly) nonlinear, parametric, leverage estimation can be justified (more structurally) in the context of a continuous-time specification similar to the one which will represent (below) the substantive core of our work. To this extent, consider the following dynamics for the logarithmic price $\log p_t$:

$$d\log p_t = \mu_t dt + \sigma_t dW_t^r,$$
$$d\sigma_t = m_t dt + \Lambda dW_t^\sigma,$$

where μ_t, m_t are adapted processes, Λ is a constant, $\{W^r, W^\sigma\}$ are correlated standard Brownian motions with $\langle dW_t^r, dW_t^\sigma \rangle = \rho dt$, and ρ is a constant leverage parameter. This system can be readily discretized as follows:

$$\log p_{t+1} - \log p_t \approx \mu_t + \sigma_t \underbrace{(W_{t+1}^r - W_t^r)}_{u_{t+1}}$$
$$\sigma_{t+1} - \sigma_t \approx m_t + \Lambda \underbrace{(W_{t+1}^\sigma - W_t^\sigma)}_{v_{t+1}}$$

or, equivalently,

$$r_{t+1} \approx \mu_t + \sigma_t u_{t+1}$$
$$\sigma_{t+1} \approx \sigma_t + m_t + \Lambda\left(\rho u_{t+1} + \left(\sqrt{1-\rho^2}\right) w_{t+1}\right), \quad (2)$$

where u_t and w_t are uncorrelated shocks with zero mean and unit variance. Now, substituting u_{t+1} into Eq. (2), we arrive at

$$\sigma_{t+1} \approx \sigma_t + m_t + \underbrace{\left(\frac{\Lambda\rho}{\sigma_t}\right)}_{\delta_t} (r_{t+1} - \mu_t) + \Lambda\left(\sqrt{1-\rho^2}\right) w_{t+1}. \quad (3)$$

In agreement with much recent empirical work, we may capture persistence (and mean-reversion) in volatility by virtue of an *HAR* specification (Corsi, 2009). In other words, we may replace $\sigma_t + m_t$ with HAR_t. Set, also, the mean return equal to zero ($\mu_t = 0$). For $N = 1$, Eq. (3) is now fully consistent with Eq. (1) with $\varepsilon_t = \Lambda\left(\sqrt{1-\rho^2}\right) w_t$ and $\delta_t = \frac{\Lambda\rho}{\sigma_t}$.[2] This last expression, then, provides a more structural justification for estimating ρ parametrically by virtue of $\widehat{\delta S}$. The genuine time-varying ρ case with $N \ge 1$ may, of course, lead to

$$\sigma_{t+1} \approx \sigma_t + m_t + \sum_{i=1}^{N} \delta_i r_{t+1} \mathbf{1}_{\{\eta_{i-1} \le \sigma_t \le \eta_i\}} + \varepsilon_{t+1}$$

with $\varepsilon_{t+1} = \sum_{i=1}^{N} \Lambda\left(\sqrt{1-\rho_i^2}\right) w_{t+1}\mathbf{1}_{\{\eta_{i-1} \le \sigma_t \le \eta_i\}}$.

In what follows, we set $N = 3$ and report three values of $\widehat{\rho}_i$ corresponding to alternative volatility levels. Using the threshold bipower variation estimator (detailed in Section 7) to define $RV_{t-i:t-i-1}$ for all i, we find that the $\widehat{\delta}_i$ values are equal to −0.226, −0.240, and −0.350 with highly significant t-statistics equal to −2.93, −6.55, and −7.54 (see Table 1). The implied leverage estimates ($\widehat{\rho}_i \approx \widehat{\delta_i S_i}$) are −0.152, −0.187, and −0.273.[3] In agreement with our discussion in the Introduction, this result is suggestive of an increasing (i.e., more negative) leverage for higher volatility levels. In Section 4, we employ this evidence to motivate a (possibly) nonlinear, continuous-time model in which conditional leverage is allowed to be a function of the spot variance level.

3. Do classical economic principles imply time-varying leverage?

Classical finance principles à la Modigliani–Miller postulate that the fundamental asset of a corporation is the firm itself. Stocks and bonds are simply viewed as being alternative ways

[2] Interestingly, a continuous-time stochastic volatility model like that in Eq. (2) would imply time variation in δ (as a function of spot volatility) even if ρ is a constant value. Hence, if one takes the continuous-time model seriously, running a regression like that in Eq. (1) above is justified only if δ is assumed time-varying, as we do.

[3] We return to potential upward biases (i.e., biases towards zero) of these estimates in Section 9 (Section 9.2).

to divide up ownership. This said, changes in the firm's total value should translate into analogous changes in the value of the company's stock. Negative news decrease the firm value (and its stock price) and increase the debt-to-equity ratio (i.e., financial leverage). The increased debt-to-equity ratio leads to a larger stock return volatility for a given volatility of the total firm's value. In other words, the firm's stock volatility σ_S depends on the firm's total volatility σ_V and on the debt-to-equity, or leverage, value $\frac{D}{E}$. Specifically, $\sigma_S = \sigma_V\left(1+\frac{D}{E}\right)$ (see Christie, 1982, and Figlewski and Wang, 2000, among others, for discussions). If the firm's total volatility is fixed or relatively stable, time-variation in the firm's stock volatility will be induced by changing levels of leverage. In particular, increases in financial leverage (as possibly implied by negative shocks to prices) will increase stock volatility, whereas decreases in financial leverage (as possibly implied by positive shocks to prices) will decrease stock volatility (Black, 1976).

Note, however, that since times of high return volatility should be associated with relatively higher financial leverage, price changes of a certain size should have a larger effect on volatility changes when leverage is relatively higher (or, equivalently, when volatility is higher). To see this, return to the expression $\sigma_S = \sigma_V\left(1+\frac{D}{E}\right)$. Then,

$$\partial\sigma_S = -\sigma_V\frac{D}{E^2}\partial E = -\left(\frac{\sigma_S-\sigma_V}{E}\right)\partial E$$
$$\Rightarrow \frac{\partial\sigma_S}{\partial E} = -\left(\frac{\sigma_S-\sigma_V}{E}\right) < 0,$$

if $\frac{D}{E} > 0$. In other words, changes in the value of equity should induce opposite volatility changes whose magnitude depends on the volatility itself.

In the *constant* leverage case, the transmission mechanism (from low returns to high volatility) implied by a Modigliani–Miller economy has been questioned in empirical work. In agreement with this observation, we strongly emphasize that we do not commit to a Modigliani–Miller world. In no way our theoretical and empirical analysis, in fact, hinges on the validity of the transmission mechanism implied by this classical economy. Hence, the discussion in the present section should be viewed as illustrative. It is merely meant to stress that *even* the most traditional economic justification for leverage effects in the stochastic volatility literature implies time-variation in leverage as a function of spot volatility. Yet, this dependence, which appears to be in the data as shown in the previous section, has invariably been unaccounted for. This paper focuses on it.

4. A continuous-time stochastic volatility model with time-varying leverage

Assume a complete probability space $(\Omega, \Im, P, \{\Im_t\}_{t\geq 0})$. Consider the system

$$\begin{pmatrix} d\log p_t \\ d\xi(\sigma_t^2)\end{pmatrix} = \begin{pmatrix}\mu_t \\ m_t\end{pmatrix}dt + \begin{pmatrix}\sigma_t & 0 \\ 0 & \Lambda(\sigma_t^2)\end{pmatrix} \times \begin{pmatrix}\rho(\sigma_t^2) & \sqrt{1-\rho^2(\sigma_t^2)} \\ 1 & 0\end{pmatrix}\begin{pmatrix} dW_t^1 \\ dW_t^2\end{pmatrix} + \begin{pmatrix} dJ_t^r \\ dJ_t^\sigma\end{pmatrix} \quad (4)$$

where $\{J_t^r, J_t^\sigma\}$ and $\{W_t^1, W_t^2\}$ are a bidimensional compound Poisson jump process and a bivariate standard Brownian motion, respectively. Note that

$$\frac{\langle d\log p_t, d\xi(\sigma_t^2)\rangle}{\sqrt{\langle d\log p_t\rangle\langle d\xi(\sigma_t^2)\rangle}} = \rho(\sigma^2)$$

defines the infinitesimal (conditional) correlation between continuous shocks to returns and continuous shocks to (transformed, by $\xi(\cdot)$) variance.

In order to specify the vector $\{J_t^r, J_t^\sigma\}$, we define three intensity functions: $\lambda_\sigma(\sigma_t^2)$, the intensity of the jumps in variance, $\lambda_r(\sigma_t^2)$, the intensity of the jumps in returns, and $\lambda_{r,\sigma}(\sigma_t^2)$, the intensity of the co-jumps (we refer to Remark 8 for more details). The jump sizes of $\xi(\sigma_t^2)$ and $\log p_t$ are determined by the random variables c_r and c_σ, respectively. We allow for correlation in both jump times and jump sizes, but not between times and sizes.[4] We also assume independence between the jumps and the standard Brownian shocks W^1, W^2. The monotonic function $\xi(\cdot)$ in the variance process is introduced for generality. It is meant to allow for alternative specifications including the logarithmic model in, e.g., Jacquier et al. (1994), the linear (in variance) model proposed by, e.g., Duffie et al. (2000) and Eraker et al. (2003), and, possibly, a linear model in volatility. The object of econometric interest is the conditional leverage function $\rho(\cdot)$. Its dependence on spot variance (or spot volatility) generalizes to a nonparametric continuous-time framework the parametric specification used as a motivation in Section 2.

Assumption 1. The return and variance drifts μ_t and m_t are adapted stochastic processes. The functions $\Lambda(\cdot), \lambda_r(\cdot), \lambda_\sigma(\cdot), \lambda_{r,\sigma}(\cdot)$, and $\rho(\cdot)$ are at least twice continuously-differentiable Borel measurable functions of the Markov state. All objects are such that a unique and strong solution of Eq. (4) exists and the spot variance process is recurrent in a bounded set.[5]

We begin by assuming availability of $n+1$ observations on both $\log p_t$ and σ_t^2 in the time interval $[0, T]$. We denote by $\Delta_{n,T} = T/n$ the time distance between adjacent discretely-sampled observations. Our asymptotic design lets $\Delta_{n,T} \to 0$ with $n, T \to \infty$. The case of observability of σ_t^2 is, of course, unrealistic in practice. However, it is important in that it allows us to lay out the main ideas while avoiding the complications induced by spot variance estimation. Having made this point, we stress that Section 7 discusses the case of spot variance estimation by virtue of functional methods applied to high-frequency price data. Section 7 presents conditions which guarantee that the estimation error associated with the spot variance estimates is asymptotically negligible. These conditions take the nature of realistic intra-daily price formation mechanisms seriously and, coherently with Eq. (4), account for jumps in both returns and variance.

Define the infinitesimal moments

$$\vartheta_{1,1}(\sigma^2) = \lim_{\Delta\to 0}\frac{1}{\Delta}\mathbf{E}\left[(\log p_{t+\Delta}-\log p_t)\left(\xi(\sigma_{t+\Delta}^2)-\xi(\sigma_t^2)\right) \mid \sigma_t^2=\sigma^2\right],$$

$$\vartheta_j(\sigma^2) = \lim_{\Delta\to 0}\frac{1}{\Delta}\mathbf{E}\left[\left(\xi(\sigma_{t+\Delta}^2)-\xi(\sigma_t^2)\right)^j \mid \sigma_t^2=\sigma^2\right],\quad j=1,2,\ldots$$

and the corresponding Nadaraya–Watson kernel estimators given in Box I, where, as is traditional, $h_{n,T}$ denotes an asymptotically-vanishing window width and $\mathbf{K}(\cdot)$ is a kernel function. The function $\mathbf{K}(\cdot)$ satisfies the following assumption.

[4] It is hard to evaluate the empirical significance of the assumption of independence between times and sizes. While one could speculate about the economics of the problem, to the best of our knowledge this assumption has not been relaxed in empirical work on estimation of jump-diffusion stochastic volatility models. This said, we can allow for intensities of the jumps, as well as for moments of the jump size distributions, which depend on the underlying spot variance process.

[5] BR (2008) discusses how this assumption is without loss of generality. The theoretically interesting, but empirically vacuous, case in which boundedness is not satisfied is thoroughly studied in BR (2008). It solely translates into stronger, and unverifiable, discretization conditions (analogous to, e.g., condition 3.1 in Assumption 3).

$$\widehat{\vartheta}_{1,1}(\sigma^2) = \frac{\sum_{i=1}^{n-1} \mathbf{K}\left(\frac{\sigma^2_{iT/n}-\sigma^2}{h_{n,T}}\right)(\log p_{(i+1)T/n} - \log p_{iT/n})\left(\xi(\sigma^2_{(i+1)T/n}) - \xi(\sigma^2_{iT/n})\right)}{\Delta_{n,T}\sum_{i=1}^{n}\mathbf{K}\left(\frac{\sigma^2_{iT/n}-\sigma^2}{h_{n,T}}\right)}, \tag{5}$$

$$\widehat{\vartheta}_j(\sigma^2) = \frac{\sum_{i=1}^{n-1} \mathbf{K}\left(\frac{\sigma^2_{iT/n}-\sigma^2}{h_{n,T}}\right)\left(\xi(\sigma^2_{(i+1)T/n}) - \xi(\sigma^2_{iT/n})\right)^j}{\Delta_{n,T}\sum_{i=1}^{n}\mathbf{K}\left(\frac{\sigma^2_{iT/n}-\sigma^2}{h_{n,T}}\right)} \quad j = 1, 2, \ldots, \tag{6}$$

Box I.

Assumption 2. The function $\mathbf{K}(\cdot)$ is a nonnegative, bounded, twice continuously-differentiable, and symmetric kernel defined on a compact set S satisfying $\int_S \mathbf{K}(s)ds = 1$, $\mathbf{K}_2 = \int_S \mathbf{K}^2(s)ds < \infty$, and $\mathbf{K}_1 = \int_S s^2\mathbf{K}(s)ds < \infty$. The kernel's derivatives are absolutely integrable.

In what follows, the kernel estimators $\widehat{\vartheta}_{1,1}(\sigma^2)$ and $\widehat{\vartheta}_j(\sigma^2)$ (with $j = 1, 2, \ldots$) will be employed to provide point-wise estimates of $\rho(\sigma^2)$ (or $\rho(\sigma)$) in various scenarios allowing for jumps in prices, jumps in variance, or both. We will resort to the classical notation $\xrightarrow{a.s.}$, $\xrightarrow{p}$, $\Rightarrow$ to denote almost-sure convergence, convergence in probability, and weak convergence. The symbol $\Gamma_z(x)$ will be used to define $h^2_{n,T}\mathbf{K}_1\left[z'(x)\frac{s'(x)}{s(x)} + \frac{1}{2}z''(x)\right]$, where $s(dx)$ is the invariant measure of the spot variance process.[6] Finally, the notation $\widehat{\overline{L}}_{\sigma^2}(T,x) = \frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}\left(\frac{\sigma^2_{iT/n}-x}{h_{n,T}}\right)$ will denote kernel estimates of the chronological local time (at T and x) of the underlying spot variance process.

Since σ^2_t is a càdlàg semimartingale, its local time at T and x can be written as

$$L_{\sigma^2}(T,x) = \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\int_0^T \mathbf{1}_{[x,x+\varepsilon[}(\sigma^2_s)\left(\frac{\partial\xi^{-1}(\xi(\sigma^2_s))}{\partial\xi}\right)^2\Lambda^2(\sigma^2_s)ds \quad \text{a.s.}$$

The interpretation is standard. $L_{\sigma^2}(T,x)$ defines the amount of time, in information units or in units of the continuous component of the process' quadratic variation, which σ^2_t spends in a small right neighbourhood of x between time 0 and time T. Analogously, time can be measured in chronological units by defining

$$\overline{L}_{\sigma^2}(T,x) = \frac{1}{\left(\frac{\partial\xi^{-1}(\xi(x))}{\partial\xi}\right)^2\Lambda^2(x)}L_{\sigma^2}(T,x) \quad \text{a.s.}$$

For a fixed T, and under standard assumptions, $\widehat{\overline{L}}_{\sigma^2}(T,x)$ is known to estimate the latter. Similarly, for an enlarging T and, again, under standard assumptions, $\widehat{\overline{L}}_{\sigma^2}(T,x)$ has been shown to inherit the divergence properties of $\overline{L}_{\sigma^2}(T,x)$ (Bandi and Nguyen, 2003). As pointed out earlier, our asymptotic results will hinge on the recurrence of the variance process, rather than on the stronger assumption of stationarity—stationarity being a subcase of our more general framework. As a by-product of this generality, the look of our limiting results will be more explicit, than in the classical stationary framework, about what drives convergence of the (point-wise) functional moment estimates. The rates of convergence will, in fact, not depend on the (largely notional) divergence rate of the number of observations, as in the stationary case, but on the rate of divergence of the number of visits to a generic level x at which functional estimation is performed, as represented by $\widehat{\overline{L}}_{\sigma^2}(T,x)$.[7] For more discussions on the role of recurrence in continuous-time model estimation, and the importance of the notion of local time in this context, we refer the reader to the review article by Bandi and Phillips (2010).

We now list a set of conditions which the smoothing sequence $h_{n,T}$ and the chronological local time estimates $\widehat{\overline{L}}_{\sigma^2}(T,\cdot)$ will be required to satisfy (as $\Delta_{n,T}\to 0$ with $n, T\to\infty$) for the validity of the limiting results in the following section. The statements of the theorems will make clear which conditions will yield the corresponding results.

Assumption 3. Let $h_{n,T}\to 0$ with $\Delta_{n,T}\to 0$ and $n, T\to\infty$ *jointly*.

3.1 $\frac{\Delta_{n,T}}{h^2_{n,T}} \to 0$.

3.2 $\frac{\Delta_{n,T}\sqrt{\widehat{\overline{L}}_{\sigma^2}(T,x)}}{h^{3/2}_{n,T}} \xrightarrow{a.s.} 0$.

3.3 $\frac{\Delta_{n,T}\widehat{\overline{L}}_{\sigma^2}(T,x)}{h^3_{n,T}} \xrightarrow{a.s.} 0$.

3.4 $\frac{h_{n,T}\widehat{\overline{L}}_{\sigma^2}(T,x)}{\Delta_{n,T}} \xrightarrow{a.s.} \infty$.

3.5 $\frac{h^5_{n,T}\widehat{\overline{L}}_{\sigma^2}(T,x)}{\Delta_{n,T}} \xrightarrow{a.s.} C$, where C is a suitable constant.

3.6 $h_{n,T}\widehat{\overline{L}}_{\sigma^2}(T,x) \xrightarrow{a.s.} \infty$.

3.7 $h^5_{n,T}\widehat{\overline{L}}_{\sigma^2}(T,x) \xrightarrow{a.s.} C$, where C is a suitable constant.

We now turn to the asymptotics and evaluate increasing levels of generality of the model in Eq. (4).

5. Asymptotics

5.1. The continuous case: $J^r = 0, J^\sigma = 0$

Write the kernel leverage estimator as

$$\widehat{\rho}(\sigma^2) = \frac{\widehat{\vartheta}_{1,1}(\sigma^2)}{\sigma\sqrt{\widehat{\vartheta}_2(\sigma^2)}}. \tag{7}$$

[6] In the absence of variance jumps, this quantity corresponds to the speed measure of the variance process, namely

$$\mathbf{Sp}(dx) = \frac{2dx}{S'(x)\Lambda^2(x)}$$

where $\mathbf{Sc}'(x)$ is the first derivative of the scale function, i.e.,

$$\mathbf{Sc}(x) = \int_c^x \exp\left\{\int_c^y\left[-\frac{2m(s)}{\Lambda^2(s)}\right]ds\right\}dy,$$

and c is a generic constant in the range of the process.

[7] Of course, if $h_{n,T}T\to\infty$, $\widehat{\overline{L}}_{\sigma^2}(T,x)$ diverges almost surely at speed T in stationary models.

In the absence of discontinuities, we note that $\vartheta_{1,1}(\sigma^2) = \sigma\Lambda(\sigma^2)\rho(\sigma^2)$ and $\vartheta_2(\sigma^2) = \Lambda^2(\sigma^2)$. Hence, $\frac{\vartheta_{1,1}(\sigma^2)}{\sigma\sqrt{\vartheta_2(\sigma^2)}} = \rho(\sigma^2)$ and, of course, $\widehat{\rho}(\sigma^2)$ is expected to consistently estimate the object of interest. This result is shown in Theorem 1.[8]

Theorem 1. *Assume $J^r = J^\sigma = 0$. If Assumptions 3.1 and 3.4 are satisfied, then $\widehat{\rho}(\sigma^2) \xrightarrow{p} \rho(\sigma^2)$.[9] If Assumptions 3.3–3.5 are satisfied, then*

$$\sqrt{\frac{h_{n,T}\widehat{\bar{L}}_{\sigma^2}(T,\sigma^2)}{\Delta_{n,T}}}\left\{\widehat{\rho}(\sigma^2) - \rho(\sigma^2) - \widetilde{\Gamma}_\rho(\sigma^2)\right\}$$
$$\Rightarrow \mathbf{N}\left(0, \mathbf{K}_2\left[1 - \frac{1}{2}\rho^2(\sigma^2)\right]\right), \tag{8}$$

where

$$\widetilde{\Gamma}_\rho(\sigma^2) = \frac{1}{\sigma\sqrt{\vartheta_2(\sigma^2)}}\Gamma_{\vartheta_{1,1}}(\sigma^2) - \frac{\vartheta_{1,1}(\sigma^2)}{2\sigma\sqrt{\vartheta_2^3(\sigma^2)}}\Gamma_{\vartheta_2}(\sigma^2). \tag{9}$$

Remark 1. In the absence of jumps in either variance or returns, the leverage estimator converges at speed $\sqrt{\frac{h_{n,T}\widehat{\bar{L}}_{\sigma^2}(T,\sigma^2)}{\Delta_{n,T}}}$. In particular, both the numerator, $\widehat{\vartheta}_{1,1}(\sigma^2)$, and the denominator, $\sigma\sqrt{\widehat{\vartheta}_2(\sigma^2)}$, converge at this same velocity. The asymptotic distribution of $\widehat{\rho}(\sigma^2)$ is therefore a linear combination (with weights $\frac{1}{\sigma\sqrt{\vartheta_2(\sigma^2)}}$ and $-\frac{\vartheta_{1,1}(\sigma^2)}{2\sigma\sqrt{\vartheta_2^3(\sigma^2)}}$) of the limiting distributions of its components as evidenced by the resulting limiting bias $(\widetilde{\Gamma}_\rho(\sigma^2))$.

As is typical in semiparametric models, the rate of convergence of the leverage estimator (in this section and in the following sections) could be increased by averaging over evaluation points. This averaging would of course be natural, and beneficial, should one model leverage as a constant value, as in most existing literature, rather than as a general function, as in this paper.

Remark 2. The asymptotic variance in Eq. (8) is maximal (and is equal to $\mathbf{K}_2$) for $\rho(\cdot) = 0$. It tends to $\frac{1}{2}\mathbf{K}_2$ as either $\rho(\cdot) \to 1$ or $\rho(\cdot) \to -1$.

5.2. The discontinuous case: $J^r \neq 0, J^\sigma = 0$

Consider the same estimator as in Eq. (7) above. The case with jumps in returns is presented in Theorem 2.

Theorem 2. *Assume $J^r \neq 0$ and $J^\sigma = 0$. Under Assumptions 3.3–3.5, we obtain*

$$\sqrt{\frac{h_{n,T}\widehat{\bar{L}}_{\sigma^2}(T,\sigma^2)}{\Delta_{n,T}}}\left\{\widehat{\rho}(\sigma^2) - \rho(\sigma^2) - \widetilde{\Gamma}_\rho(\sigma^2)\right\}$$
$$\Rightarrow \mathbf{N}\left(0, \mathbf{K}_2\left[\left(1 - \frac{1}{2}\rho^2(\sigma^2)\right) + \frac{1}{2}\frac{\lambda_r(\sigma^2)\mathbf{E}[c_r^2]}{\sigma^2}\right]\right).$$

Remark 3. Allowing for jumps in returns only affects the (limiting) precision of the estimator. The asymptotic variance now contains an extra term $(\frac{1}{2}\frac{\lambda_r(\sigma^2)\mathbf{E}[c_r^2]}{\sigma^2})$ which, of course, depends on the frequency of the return jumps $(\lambda_r(\sigma^2))$ as well as on their size $(\mathbf{E}[c_r^2])$.

5.3. The discontinuous case: $J^r = 0, J^\sigma \neq 0$

When allowing for jumps in the variance process, $\widehat{\vartheta}_2(\sigma^2)$ estimates $\Lambda^2(\sigma^2)$ plus the conditional second moment of the jump component (i.e., $\lambda_\sigma(\sigma^2)\mathbf{E}(c_\sigma^2)$). In what follows, we show that $\lambda_\sigma(\sigma^2)\mathbf{E}(c_\sigma^2)$ can be identified, under appropriate parametric assumptions on the jump sizes, by virtue of the higher-order conditional moments (namely, $\widehat{\vartheta}_j(\sigma^2)$ with $j = 3, 4, \ldots$) as proposed by Bandi and Renò (2008) in the case of nonparametric stochastic volatility modelling (see, also, Bandi and Nguyen, 2003 and Johannes, 2004). The form of the kernel leverage estimator in this section will therefore be

$$\widetilde{\rho}(\sigma^2) = \frac{\widehat{\vartheta}_{1,1}(\sigma^2)}{\sigma\sqrt{f(\widehat{\vartheta}_2, \widehat{\vartheta}_3, \widehat{\vartheta}_4, \ldots)(\sigma^2)}}, \tag{10}$$

where $f(\widehat{\vartheta}_2, \widehat{\vartheta}_3, \widehat{\vartheta}_4, \ldots)(\cdot)$ is a specific function of the infinitesimal moments.

To lay out ideas, we turn to a specific identification scheme. Assume that $\xi(\sigma^2) = \sigma^2$ and that the variance jumps are exponentially distributed, i.e., $c_\sigma \sim \exp(\mu_\sigma)$. This specification is widely used in the parametric literature on variance estimation (see, e.g., Eraker et al., 2003) and has been shown to perform very satisfactorily with simulated data in a (semi)nonparametric context (Bandi and Renò, 2008). The proposed model implies

$$\vartheta_2(\sigma^2) = \Lambda^2(\sigma^2) + 2\mu_\sigma^2\lambda_\sigma(\sigma^2),$$
$$\vartheta_3(\sigma^2) = 6\mu_\sigma^3\lambda_\sigma(\sigma^2),$$
$$\vartheta_4(\sigma^2) = 24\mu_\sigma^4\lambda_\sigma(\sigma^2).$$

Hence,

$$\widehat{\mu}_\sigma = \frac{1}{\bar{n}}\sum_{i=1}^{\bar{n}}\frac{\widehat{\vartheta}_4(\sigma^2_{i\bar{T}/\bar{n}})}{4\widehat{\vartheta}_3(\sigma^2_{i\bar{T}/\bar{n}})}, \tag{11}$$
$$\widehat{\lambda}_\sigma(\sigma^2) = \frac{\widehat{\vartheta}_4(\sigma^2)}{24\widehat{\mu}_\sigma^4},$$

and

$$\widehat{\Lambda}^2(\sigma^2) = \widehat{\vartheta}_2(\sigma^2) - 2\widehat{\mu}_\sigma^2\widehat{\lambda}_{\sigma^2}(\sigma^2),$$

are (possible) kernel estimators of $\mu_\sigma, \lambda_\sigma(\sigma^2)$, and $\Lambda^2(\sigma^2)$, respectively.[10] These estimators have been shown to be consistent and (mixed) normally distributed (Bandi and Renò, 2008). Thus,

$$f(\widehat{\vartheta}_2, \widehat{\vartheta}_3, \widehat{\vartheta}_4, \ldots)(\sigma^2) = \widehat{\Lambda}^2(\sigma^2) = \widehat{\vartheta}_2(\sigma^2) - \frac{\widehat{\vartheta}_4(\sigma^2)}{12\widehat{\mu}_\sigma^2}$$
$$= \widehat{\vartheta}_2(\sigma^2) - \frac{\widehat{\vartheta}_4(\sigma^2)}{12\left(\frac{1}{\bar{n}}\sum_{i=1}^{\bar{n}}\frac{\widehat{\vartheta}_4(\sigma^2_{i\bar{T}/\bar{n}})}{4\widehat{\vartheta}_3(\sigma^2_{i\bar{T}/\bar{n}})}\right)^2}.$$

---

8 In agreement with Assumption 3 above, the results in Theorem 1 are stated for an enlarging time span. We do so to more clearly draw a comparison between this benchmark case and the cases with jumps (below) which, of course, require an enlarging span of data to identify sample path discontinuities. This said, in the no-jump case, consistency and weak convergence could be derived for a fixed span of data $\bar{T}$. For an alternative approach to leverage estimation in the fixed $\bar{T}$ case, we refer the reader to Mykland and Zhang (2009).

9 Similar methods and conditions lead to almost-sure convergence here and below. We do not emphasize this stronger mode of convergence because the use of estimated spot variances, which is needed for feasibility, would yield convergence in probability even starting from a stronger mode of convergence.

10 $\widehat{\mu}_\sigma$ is defined over a number of observations $\bar{n}$ growing to infinity over a fixed time span $\bar{T}$. This is simply done for technical reasons in order to simplify the limiting behaviour of the sample averages in the nonstationary (but recurrent) case. From an applied standpoint, the restriction is hardly material in that one could always choose $\bar{T}$ as being very close to T. For asymptotic consistency, the kernel estimators $\widehat{\vartheta}_3(\cdot)$ and $\widehat{\vartheta}_4(\cdot)$ continue to be defined over an enlarging time span $(T \to \infty)$. Thus, we are simply averaging functionals of $\widehat{\vartheta}_3(\cdot)$ and $\widehat{\vartheta}_4(\cdot)$ over an infinite number of evaluation points for a fixed span of data.

Importantly, alternative estimation schemes may be adopted by imposing, for instance, different distributional assumptions on c_σ. In these cases, our methods can be modified accordingly.

In what follows, we assume the use of a slightly smaller bandwidth sequence to identify $\widehat{\vartheta}_4$ and $\widehat{\vartheta}_3$ for the purpose of $\widehat{\mu}_\sigma$ estimation. This choice will somewhat simplify the look of the limiting bias of $\widetilde{\rho}(\cdot)$ by preventing the insurgence of the asymptotic bias of $\widehat{\mu}_\sigma$.

Theorem 3. *Assume $J^r = 0$ and $J^\sigma \neq 0$. If Assumptions* 3.1 *and* 3.6 *are satisfied, then $\widetilde{\rho}(\sigma^2) \xrightarrow{p} \rho(\sigma^2)$. If $\rho(\sigma^2) \neq 0$ and Assumptions* 3.2, 3.6 *and* 3.7 *are also satisfied, then*

$$\sqrt{h_{n,T}\widehat{\overline{L}}_{\sigma^2}(T,\sigma^2)}\left\{\widetilde{\rho}(\sigma^2) - \rho(\sigma^2) - \widetilde{\Gamma}_{\widetilde{\rho}}(\sigma^2)\right\} \Rightarrow \mathbf{N}\left(0, \mathbf{K}_2 \frac{\rho^2(\sigma^2)}{4\Lambda^4(\sigma^2)}\lambda_\sigma(\sigma^2)\mathbf{E}\left(\left(c_\sigma^2 - \frac{1}{12\mu_\sigma^2}c_\sigma^4\right)^2\right)\right) \quad (12)$$

with

$$\widetilde{\Gamma}_{\widetilde{\rho}}(\sigma^2) = \frac{1}{\sigma\Lambda(\sigma^2)}\Gamma_{\vartheta_{1,1}}(\sigma^2) - \frac{\vartheta_{1,1}(\sigma^2)}{2\sigma\Lambda^3(\sigma^2)}\left(\Gamma_{\vartheta_2}(\sigma^2) - \frac{1}{12\mu_\sigma^2}\Gamma_{\vartheta_4}(\sigma^2)\right).$$

Remark 4. The estimator $f(\widehat{\vartheta}_2, \widehat{\vartheta}_3, \widehat{\vartheta}_4, \ldots)(\cdot)$ now converges to $\Lambda^2(\cdot)$ at a slower speed than $\widehat{\vartheta}_2(\cdot)$ for the case of no jumps ($\sqrt{h_{n,T}\widehat{\overline{L}}_{\sigma^2}(T,\sigma^2)}$ versus $\sqrt{\frac{h_{n,T}\widehat{\overline{L}}_{\sigma^2}(T,\sigma^2)}{\Delta_{n,T}}}$). Since $\widehat{\vartheta}_{1,1}(\cdot)$ continues to converge at speed $\sqrt{\frac{h_{n,T}\widehat{\overline{L}}_{\sigma^2}(T,\sigma^2)}{\Delta_{n,T}}}$, not only is the slower speed of convergence of $f(\widehat{\vartheta}_2, \widehat{\vartheta}_3, \widehat{\vartheta}_4, \ldots)(\cdot)$ driving the rate of convergence of $\widetilde{\rho}(\cdot)$ but, also, of course, the asymptotic variance of the leverage estimator is fully determined by the asymptotic variance of $f(\widehat{\vartheta}_2, \widehat{\vartheta}_3, \widehat{\vartheta}_4, \ldots)(\cdot)$ (times a term $\frac{\vartheta_{1,1}^2(\sigma^2)}{4\sigma^2\Lambda^6(\sigma^2)} = \frac{\rho^2(\sigma^2)}{4\Lambda^4(\sigma^2)}$ which readily derives from the delta method; see, e.g., Remark 1 above).

Remark 5. Under the assumed exponential jumps, the asymptotic variance in Eq. (12) can be more explicitly expressed as $46\mathbf{K}_2\frac{\rho^2(\sigma^2)\lambda_\sigma(\sigma^2)}{\Lambda^4(\sigma^2)}\mu_\sigma^4$.

5.4. The discontinuous case: $J^r \neq 0, J^\sigma \neq 0$, with independent jumps

Consider the same estimator as in Eq. (10) above.

Theorem 4. *Assume $J^r \neq 0, J^\sigma \neq 0$, and $J^r \perp J^\sigma$. Under Assumptions* 3.2, 3.6 *and* 3.7*:*

$$\sqrt{h_{n,T}\widehat{\overline{L}}_{\sigma^2}(T,\sigma^2)}\left\{\widetilde{\rho}(\sigma^2) - \rho(\sigma^2) - \widetilde{\Gamma}_{\widetilde{\rho}}(\sigma^2)\right\} \Rightarrow \mathbf{N}\left(0, \mathbf{K}_2 \frac{\rho^2(\sigma^2)}{4\Lambda^4(\sigma^2)}\lambda_\sigma(\sigma^2)\mathbf{E}\left(\left(c_\sigma^2 - \frac{1}{12\mu_\sigma^2}c_\sigma^4\right)^2\right)\right) \quad (13)$$

with

$$\widetilde{\Gamma}_{\widetilde{\rho}}(\sigma^2) = \frac{1}{\sigma\Lambda(\sigma^2)}\Gamma_{\vartheta_{1,1}}(\sigma^2) - \frac{\vartheta_{1,1}(\sigma^2)}{2\sigma\Lambda^3(\sigma^2)}\left(\Gamma_{\vartheta_2}(\sigma^2) - \frac{1}{12\mu_\sigma^2}\Gamma_{\vartheta_4}(\sigma^2)\right).$$

Remark 6. Adding independent jumps in returns to the case of jumps in variance does not modify the limiting distribution of $\widetilde{\rho}(\cdot)$ (Eq. (13) is the same as Eq. (12)). This is, of course, in contrast to the case where jumps in returns are added to the case of no jumps (as in Theorem 2). Here the addition of independent return jumps does not translate into efficiency losses, as implied by a higher asymptotic variance, since the limiting variance of the leverage estimator is, again, only driven by a slowly-converging quantity in the denominator of the estimator, i.e., $\sqrt{f(\widehat{\vartheta}_2, \widehat{\vartheta}_3, \widehat{\vartheta}_4, \ldots)(\cdot)}$.

5.5. The discontinuous case: $J^r \neq 0, J^\sigma \neq 0$, with correlated jumps

Finally, we allow for correlated jumps and, again, evaluate the estimator in Eq. (10).

Theorem 5. *Assume $J_r = J_r^* \neq 0, J_\sigma = J_\sigma^* \neq 0$, and the intensity of common shocks $\lambda_r(\sigma^2) = \lambda_\sigma(\sigma^2) = \lambda_{r,\sigma}(\sigma^2) \neq 0$. If Assumptions* 3.1 *and* 3.6 *are satisfied, then*

$$\widetilde{\rho}(\sigma^2) \xrightarrow{p} \Xi(\sigma^2) = \rho(\sigma^2) + \frac{\lambda_{r,\sigma}(\sigma^2)}{\sigma\Lambda(\sigma^2)}\mathbf{E}[c_r c_\sigma].$$

If Assumptions 3.2, 3.6 *and* 3.7 *are satisfied, then*

$$\sqrt{h_{n,T}\widehat{\overline{L}}_{\sigma^2}(T,\sigma^2)}\left\{\widetilde{\rho}(\sigma^2) - \Xi(\sigma^2) - \widetilde{\Gamma}_{\widetilde{\rho}}(\sigma^2)\right\} \Rightarrow \mathbf{N}\left(0, \mathbf{K}_2 V_\Xi\right) \quad (14)$$

with

$$V_\Xi = \frac{\lambda_{r,\sigma}(\sigma^2)}{\sigma^2\Lambda^2(\sigma^2)}\mathbf{E}\left[\left(c_r c_\sigma - \frac{\vartheta_{1,1}(\sigma^2)}{2\Lambda^2(\sigma^2)}\left(c_\sigma^2 - \frac{c_\sigma^4}{12\mu_\sigma^2}\right)\right)^2\right],$$

and

$$\widetilde{\Gamma}_{\widetilde{\rho}}(\sigma^2) = \frac{1}{\sigma\Lambda(\sigma^2)}\Gamma_{\vartheta_{1,1}}(\sigma^2) - \frac{\vartheta_{1,1}(\sigma^2)}{2\sigma\Lambda^3(\sigma^2)}\left(\Gamma_{\vartheta_2}(\sigma^2) - \frac{1}{12\mu_\sigma^2}\Gamma_{\vartheta_4}(\sigma^2)\right),$$

with

$$\vartheta_{1,1}(\sigma^2) = \sqrt{\sigma^2}\Lambda(\sigma^2)\rho(\sigma^2) + \lambda_{r,\sigma}(\sigma^2)\mathbf{E}[c_r c_\sigma].$$

Remark 7. In this case, the kernel leverage estimator is inconsistent for $\rho(\cdot)$. Since both the numerator, $\widehat{\vartheta}_{1,1}(\cdot)$, and the denominator, $f(\widehat{\vartheta}_2, \widehat{\vartheta}_3, \widehat{\vartheta}_4, \ldots)(\cdot)$, converge at the same rate, the limiting distribution of $\widetilde{\rho}(\cdot)$ is that of a linear combination of $\widehat{\vartheta}_{1,1}(\cdot)$ and $f(\widehat{\vartheta}_2, \widehat{\vartheta}_3, \widehat{\vartheta}_4, \ldots)(\cdot)$.

Corollary to Theorem 5 (*A Relevant Sub-Case: Independent Jump Sizes with Mean Zero Return Jumps*). *Under the assumptions of Theorem 5, if $c_r \perp c_\sigma$, $\mathbf{E}[c_r] = 0$, $\mathbf{E}[c_r^2] = \sigma_r^2$, and $c_\sigma \sim \exp(\mu_\sigma)$, then $\widetilde{\rho}(\sigma^2) \xrightarrow{p} \rho(\sigma^2)$ and consistency is preserved. In addition:*

$$\sqrt{h_{n,T}\widehat{\overline{L}}_{\sigma^2}(T,\sigma^2)}\left\{\widetilde{\rho}(\sigma^2) - \rho(\sigma^2) - \widetilde{\Gamma}_{\widetilde{\rho}}(\sigma^2)\right\} \Rightarrow \mathbf{N}\left(0, \mathbf{K}_2\frac{\lambda_{r,\sigma}(\sigma^2)}{\sigma^2\Lambda^2(\sigma^2)}\left[2\sigma_r^2\mu_\sigma^2 + 46\frac{\sigma^2\rho^2(\sigma^2)}{\Lambda^2(\sigma^2)}\mu_\sigma^4\right]\right), \quad (15)$$

with

$$\widetilde{\Gamma}_{\widetilde{\rho}}(\sigma^2) = \frac{1}{\sigma\Lambda(\sigma^2)}\Gamma_{\vartheta_{1,1}}(\sigma^2) - \frac{\vartheta_{1,1}(\sigma^2)}{2\sigma\Lambda^3(\sigma^2)}\left(\Gamma_{\vartheta_2}(\sigma^2) - \frac{1}{12\mu_\sigma^2}\Gamma_{\vartheta_4}(\sigma^2)\right),$$

and

$$\vartheta_{1,1}(\sigma^2) = \sqrt{\sigma^2}\Lambda(\sigma^2)\rho(\sigma^2). \quad \square$$

Remark 8 (*Contemporaneous and Non-Contemporaneous Jumps*). The theorem assumes contemporaneous jumps with an infinitesimal probability of co-jumps equal to $\lambda_{r,\sigma}(\sigma^2)dt$. This is a classical case of dependence in the parametric literature. It is considered, for example, in model SVCJ in Eraker et al. (2003).[11] In general, we could assume $J_r = J_r^* + J_r^{\parallel}$ and $J_\sigma = J_\sigma^* + J_\sigma^{\parallel}$, with $J_r^* \perp J_\sigma^*, J_r^* \perp J_r^{\parallel}, J_r^* \perp J_\sigma^{\parallel}$ and $J_\sigma^* \perp J_\sigma^{\parallel}, J_\sigma^* \perp J_r^{\parallel}$. More explicitly, we could assume that both processes comprise two components, $J_{r,\sigma}^*$ and $J_{r,\sigma}^{\parallel}$, which are independent of each other and all others with the exception of $J_r^{\parallel}$ and $J_\sigma^{\parallel}$, which are dependent. Denote now by $c_{r,\sigma}^*$ and $\lambda_{r,\sigma}^*$ the jump sizes and the intensities of the jumps of the independent components $J_{r,\sigma}^*$. Similarly, denote by $c_{r,\sigma}^{\parallel}$ and $\lambda_{r,\sigma}^{\parallel} = \lambda_r^{\parallel} = \lambda_\sigma^{\parallel}$ the jump sizes of the dependent components and the (common) intensity of the common shocks. The result in Eq. (14) continues to hold and may be re-written as follows:

$$\sqrt{h_{n,T}\widehat{\widetilde{L}}_{\sigma^2}(T,\sigma^2)}\left\{\widetilde{\rho}(\sigma^2) - \Xi(\sigma^2) - \widetilde{\Gamma}_{\widetilde{\rho}}(\sigma^2)\right\}$$
$$\Rightarrow \mathbf{N}\left(0, \mathbf{K}_2\left[\frac{\lambda_{r,\sigma}^{\parallel}(\sigma^2)}{\sigma^2\Lambda^2(\sigma^2)}\mathbf{E}\left[\left(c_r^{\parallel}c_\sigma^{\parallel} - \frac{\vartheta_{1,1}(\sigma^2)}{2\Lambda^2(\sigma^2)}\right.\right.\right.\right.$$
$$\left.\left.\times\left((c_\sigma^{\parallel})^2 - \frac{(c_\sigma^{\parallel})^4}{12\mu_\sigma^2}\right)\right)^2\right]$$
$$\left.\left.+\frac{\vartheta_{1,1}^2(\sigma^2)}{4\sigma^2\Lambda^6(\sigma^2)}\lambda_\sigma^*\mathbf{E}\left[\left((c_\sigma^*)^2 - \frac{(c_\sigma^*)^4}{12\mu_\sigma^4}\right)^2\right]\right]\right),$$

where

$$\Xi(\sigma^2) = \rho(\sigma^2) + \frac{\lambda_{r,\sigma}^{\parallel}(\sigma^2)}{\sigma\Lambda(\sigma^2)}\mathbf{E}[c_r^{\parallel}c_\sigma^{\parallel}].$$

In this case, the limiting value of $\widetilde{\rho}(\sigma^2)$ is the same as that in Theorem 5. The same is true for the convergence rate. However, the limiting variance depends now explicitly on the sizes and common intensity of the dependent jumps ($c_{r,\sigma}^{\parallel}$ and $\lambda_{r,\sigma}^{\parallel}$) as well as on the size and intensity of the independent variance jumps (c_σ^* and λ_σ^*).

6. A notion of "generalized leverage"

The expression

$$\underbrace{\Xi(\sigma^2)}_{\text{generalized leverage}} = \underbrace{\rho(\sigma^2)}_{\text{continuous leverage}} + \underbrace{\frac{\lambda_{r,\sigma}^{\parallel}(\sigma^2)}{\sigma\Lambda(\sigma^2)}\mathbf{E}[c_r^{\parallel}c_\sigma^{\parallel}]}_{\text{co-jump leverage}} \tag{16}$$

can be associated with a broader notion of leverage. Specifically, $\Xi(\sigma^2)$ is represented as the sum of the infinitesimal correlation between "continuous" shocks to prices and "continuous" shocks to spot variance (namely, the traditional leverage component) and a component arising from the presence of co-jumps. The latter is simply the (standardized) conditional covariance of the co-jumps.

The standardization (by $\sigma\Lambda(\sigma^2)$) may be viewed as arbitrary. One can, for example, standardize by (the square root of) the full conditional second moments of the price and variance processes rather than simply by their diffusive components. While ad-hoc, it is however a necessary standardization in order to isolate $\rho(\sigma^2)$. It is also somewhat natural in that it is the standardization that one would employ if the jumps were assumed not to play a role. In this case, the expression clarifies the impact of the co-jump component on the continuous leverage estimates, as done in the previous section.

In general, we can view $\Xi(\sigma^2)$ as an explicit representation of the fact that negative correlations between shocks to prices and shocks to variances may be imputed to a negative correlation between the "continuous" components of prices and variances, to a negative correlation between the joint "discontinuous" components of prices and variances, or both. Disentangling the relative impact of alternative components is economically important. The following remarks provide further discussions on this issue. We, however, elaborate on the topic, both theoretically and empirically, in Bandi and Renò (2011).

Remark 9 (*Co-Jump Identification*). Methods have been put forward to identify the co-jumps. Mancini and Gobbi (forthcoming), for example, suggest identifying the contemporaneous discontinuities of two generic jump-diffusion processes X_1 and X_2 by virtue of products of the type

$$\Delta X_1 \mathbf{1}_{\{(\Delta X_1)^2 \geq \theta(\Delta_{n,T})\}}\Delta X_2 \mathbf{1}_{\{(\Delta X_2)^2 \geq \theta(\Delta_{n,T})\}},$$

where $\theta(\delta)$ is a function such that $\frac{\delta\log\left(\frac{1}{\delta}\right)}{\theta(\delta)} \to 0$ when $\delta \to 0$. Asymptotically (for $\Delta_{n,T} \to 0$), the indicators eliminate variations which are smaller than a threshold. Since the threshold is modelled based on the modulus of continuity of Brownian notion, the variations being eliminated are of the Brownian type, thereby leading to identification of the contemporaneous Poisson jumps. Once a time-series of co-jumps is formed, the corresponding intensity ($\lambda_{1,2}^{\parallel}$) may be evaluated, possibly under an assumption of constancy, by computing the in-sample frequency of co-jumps. Similarly, the expected first cross-moment ($\mathbf{E}[c_1^{\parallel}c_2^{\parallel}]$) can be consistently identified by virtue of sample averages of the co-jumps (under, of course, stationarity of the jump distribution). Finally, given $\widehat{\Lambda}^2(\cdot) = f(\widehat{\vartheta}_2, \widehat{\vartheta}_3, \widehat{\vartheta}_4, \ldots)(\cdot)$, the continuous leverage function $\rho(\sigma^2)$ may be estimated consistently by virtue of $\widetilde{\rho}(\sigma^2) - \frac{\widehat{\lambda}_{r,\sigma}^{\parallel}(\sigma^2)}{\sigma\sqrt{f(\widehat{\vartheta}_2,\widehat{\vartheta}_3,\widehat{\vartheta}_4,\ldots)(\sigma^2)}}\widehat{\mathbf{E}}[c_r^{\parallel}c_\sigma^{\parallel}]$. For an alternative approach to co-jump testing, we refer the reader to Jacod and Todorov (2009).

Remark 10 (*More on Co-Jump Identification*). A different procedure for co-jump identification is implied by the recent work of Bandi and Renò (2011) on *cross-moment* estimation. Here, we summarize some ideas. The procedure hinges on the nonparametric identification of the return/variance cross-moments of order p_1 and p_2, i.e.,

$$\vartheta_{p_1,p_2}(\sigma^2) = \lim_{\Delta\to 0}\frac{1}{\Delta}\mathbf{E}\Big[(\log p_{t+\Delta} - \log p_t)^{p_1}$$
$$\times\left(\xi(\sigma_{t+\Delta}^2) - \xi(\sigma_t^2)\right)^{p_2}|\sigma_t^2 = \sigma^2\Big],$$

with $p_1 \geq p_2 \geq 0$. The intuition is as follows: the cross-moments of order higher than 1, 1 (i.e., $p_1 \geq p_2 \geq 1$ and $p_1 > p_2$ if $p_2 = 1$ or $p_2 > p_1$ if $p_1 = 1$) depend solely on the features of the co-jumps and may therefore be used to identify $\lambda_{r,\sigma}^{\parallel}$ and $\mathbf{E}[c_r^{\parallel}c_\sigma^{\parallel}]$. As an example, assume $\xi(\cdot) = \log(\cdot)$. Using the notation in Remark 8, express

$$J_r = J_r^* + J_r^{\parallel} = c_r^{\parallel}\left(N_r^* + N_{r,\sigma}^{\parallel}\right)$$
$$J_\sigma = J_\sigma^* + J_\sigma^{\parallel} = c_\sigma^{\parallel}\left(N_\sigma^* + N_{r,\sigma}^{\parallel}\right)$$

where N_r^*, N_σ^*, and $N_{r,\sigma}^{\parallel}$ are independent Poisson processes. Now write

$$\begin{pmatrix} c_r^{\parallel} \\ c_\sigma^{\parallel}\end{pmatrix} \sim \mathbf{N}(0, \Sigma_J) \quad\text{and}\quad \Sigma_J = \begin{pmatrix}\sigma_{J,r}^2 & \blacklozenge \\ \rho_J\sigma_{J,r}\sigma_{J,\sigma} & \sigma_{J,\sigma}^2\end{pmatrix},$$

[11] Their model SVIJ assumes independence of the jumps as in Section 5.4 above.

where Σ_J may be a function of σ^2.[12] Then,

$$\begin{Bmatrix} \vartheta_{1,0}(\cdot) = \mu(\cdot) \\ \vartheta_{2,0}(\cdot) = \cdot + (\lambda_r^*(\cdot) + \lambda_{r,\sigma}^{\parallel}(\cdot))\sigma_{J,r}^2 \\ \vartheta_{4,0}(\cdot) = 3(\lambda_r^*(\cdot) + \lambda_{r,\sigma}^{\parallel}(\cdot))\sigma_{J,r}^4 \\ \vartheta_{6,0}(\cdot) = 15(\lambda_r^*(\cdot) + \lambda_{r,\sigma}^{\parallel}(\cdot))\sigma_{J,r}^6 \end{Bmatrix}$$

yield identification of $\mu(\cdot)$, $\sigma_{J,r}^2$, and $\left\{\lambda_r^*(\cdot) + \lambda_{r,\sigma}^{\parallel}(\cdot)\right\}$. Similarly,

$$\begin{Bmatrix} \vartheta_{0,1}(\cdot) = m(\cdot) \\ \vartheta_{0,2}(\cdot) = \Lambda^2(\cdot) + (\lambda_\sigma^*(\cdot) + \lambda_{r,\sigma}^{\parallel}(\cdot))\sigma_{J,\sigma}^2 \\ \vartheta_{0,4}(\cdot) = 3(\lambda_\sigma^*(\cdot) + \lambda_{r,\sigma}^{\parallel}(\cdot))\sigma_{J,\sigma}^4 \\ \vartheta_{0,6}(\cdot) = 15(\lambda_\sigma^*(\cdot) + \lambda_{r,\sigma}^{\parallel}(\cdot))\sigma_{J,\sigma}^6 \end{Bmatrix}$$

identify $m(\cdot)$, $\Lambda(\cdot)$, $\sigma_{J,\sigma}^2$, and $\left\{\lambda_\sigma^*(\cdot) + \lambda_{r,\sigma}^{\parallel}(\cdot)\right\}$. Finally, for instance,

$$\begin{Bmatrix} \vartheta_{1,1}(\cdot) = \rho(\cdot)\sqrt{\cdot}\Lambda(\cdot) + \lambda_{r,\sigma^2}^{\parallel}(\cdot)\rho_J\sigma_{J,r}\sigma_{J,\sigma} \\ \vartheta_{2,2}(\cdot) = \lambda_{r,\sigma^2}^{\parallel}(\cdot)\sigma_{J,r}^2\sigma_{J,\sigma}^2(1+2\rho_J^2) \\ \vartheta_{3,1}(\cdot) = 3\lambda_{r,\sigma^2}^{\parallel}(\cdot)\rho_J\sigma_{J,r}^3\sigma_{J,\sigma} \\ \vartheta_{1,3}(\cdot) = 3\lambda_{r,\sigma^2}^{\parallel}(\cdot)\rho_J\sigma_{J,r}\sigma_{J,\sigma}^3 \end{Bmatrix}$$

identify $\rho(\cdot)$, ρ_J, and $\lambda_{r,\sigma^2}^{\parallel}(\cdot)$. Naturally, even in this logarithmic model with Gaussian jumps, alternative identification schemes may be entertained.

7. Allowing for spot variance estimation

The spot variance process is latent. Thus, when implementing $\widehat{\vartheta}_{1,1}(\sigma^2)$ and $\widehat{\vartheta}_j(\sigma^2)$ with $j = 1, 2, 3, 4, \ldots$ one must replace $\sigma_{iT/n}^2$ with estimates $\widehat{\sigma}_{iT/n}^2$ for $\forall i = 1, \ldots, n$.

To this extent, assume availability of k equispaced high-frequency observations over each time interval $[i\Delta_{n,T} - \phi, i\Delta_{n,T}]$.[13] Assume these k observations are employed to estimate the integrated variance of the logarithmic price process over each interval (i.e., $\int_{i\Delta_{n,T}-\phi}^{i\Delta_{n,T}} \sigma_s^2 ds$) by virtue of the threshold bipower variation estimator

$$TBPV_{iT/n} = \varsigma_1^{-2} \sum_{j=2}^{k} \left(\left|\Delta_{j-1}\log p\right| \left|\Delta_j \log p\right|\right) \times \mathbf{1}_{\left\{\left|\Delta_{j-1}\log p\right|^2 \le \theta_{j-1}\right\}} \mathbf{1}_{\left\{\left|\Delta_j \log p\right|^2 \le \theta_j\right\}}, \tag{17}$$

where $\varsigma_1 = \sqrt{2/\pi} \simeq 0.7979$ (Corsi et al., 2010). This estimator has a "double-sword" feature in eliminating jumps, and evaluating the diffusive integrated variance $\int_{i\Delta_{n,T}-\phi}^{i\Delta_{n,T}} \sigma_s^2 ds$, which makes it particularly suitable for our purposes: discontinuities are annihilated asymptotically by the adjacent diffusive component (similar to the classical bipower variation of Barndorff-Nielsen and Shephard, 2004) and are discarded in finite samples (and asymptotically) when above a threshold $\theta_j \to 0$ (as in the threshold realized variance of Mancini, 2009).

Now define $\widehat{\sigma}_{iT/n}^2 = \frac{TBPV_{iT/n}}{\phi}$ and let $\phi \to 0$ as $k \to \infty$, asymptotically. We may show the following.

Theorem 6. *Assume $\theta_t = \xi_t \Theta(\frac{\phi}{k})$, where $\Theta(\frac{\phi}{k})$ is a real function satisfying $\Theta\left(\frac{\phi}{k}\right) \xrightarrow[\phi\to 0, k\to\infty]{} 0$ and $\frac{1}{\Theta\left(\frac{\phi}{k}\right)}\left(\frac{\phi}{k}\log\left(\frac{\phi}{k}\right)\right) \xrightarrow[\phi\to 0, k\to\infty]{} 0$ and ξ_t is an a.s. bounded process with a strictly positive lower bound.*

(i) *If $\sqrt{k\phi} \to 0$, then*

$$\sqrt{k}\left(\frac{TBPV_{iT/n}}{\phi} - \sigma_{iT/n}^2\right) \Rightarrow \mathbf{MN}\left(0, c_2\sigma_{iT/n}^4\right),$$

with $c_2 \approx 2.61$.

(ii) *We also have*

$$\max_{1\le i\le n}\left|\frac{TBPV_{iT/n}}{\phi} - \sigma_{iT/n}^2\right| = O_p\left(\sqrt{\phi}\right) + O_p\left(\sqrt{\frac{\log(n)}{k}}\right).$$

If feasibility is restored by employing $\widehat{\sigma}_{iT/n}^2 = \frac{TBPV_{iT/n}}{\phi}$ in place of the unobservable $\sigma_{iT/n}^2$, $\forall i = 1, \ldots, n$, the resulting estimation error must be controlled by relating the limiting properties of n, T, and $\Delta_{n,T}$ to those of ϕ, k, and θ. The following theorem does so.

Theorem 7. *Let $\widehat{\sigma}_{iT/n}^2 = \frac{TBPV_{iT/n}}{\phi}$ be defined as in Eq. (17). Let, also, $\phi \to 0$ and $k \to \infty$. Assume $\theta_t = \xi_t\Theta(\frac{\phi}{k})$, where $\Theta(\frac{\phi}{k})$ is a real function satisfying $\Theta\left(\frac{\phi}{k}\right) \xrightarrow[\phi\to 0, k\to\infty]{} 0$ and $\frac{1}{\Theta\left(\frac{\phi}{k}\right)}\left(\frac{\phi}{k}\log\left(\frac{\phi}{k}\right)\right) \xrightarrow[\phi\to 0, k\to\infty]{} 0$ and ξ_t is an a.s. bounded process with a strictly positive lower bound. Write $\mathbf{\Psi}_{n,k,\phi} = \sqrt{\frac{\log(n)}{k}} + \sqrt{\phi}$. If*

$$\frac{\mathbf{\Psi}_{n,k,\phi}}{\Delta_{n,T}^{1/2}} \to 0$$

the consistency results in Theorems 1 and 2 hold when replacing $\sigma_{iT/n}^2$ with $\widehat{\sigma}_{iT/n}^2$. If

$$\sqrt{\frac{h_{n,T}\widehat{\overline{L}}_{\sigma^2}(T, \sigma^2)}{\Delta_{n,T}}}\frac{\mathbf{\Psi}_{n,k,\phi}}{\Delta_{n,T}^{1/2}} \to 0$$

the weak convergence results in Theorems 1 and 2 hold when replacing $\sigma_{iT/n}^2$ with $\widehat{\sigma}_{iT/n}^2$. If

$$\frac{\mathbf{\Psi}_{n,k,\phi}}{\Delta_{n,T}^{1/2}h_{n,T}} \to 0$$

the consistency results in Theorems 3–5 hold when replacing $\sigma_{iT/n}^2$ with $\widehat{\sigma}_{iT/n}^2$. If

$$\sqrt{h_{n,T}\widehat{\overline{L}}_{\sigma^2}(T, \sigma^2)}\frac{\mathbf{\Psi}_{n,k,\phi}}{\Delta_{n,T}^{1/2}h_{n,T}} \to 0$$

the weak convergence results in Theorems 3–5 hold when replacing $\sigma_{iT/n}^2$ with $\widehat{\sigma}_{iT/n}^2$.

If the return/variance dynamics are evaluated by virtue of infinitesimal conditional moment estimates based on n *daily* observations, as in this paper, the conditions in Theorem 7 require availability of a sufficiently large number k of *intra-daily* observations for the purpose of estimating spot variance for each day $i = 1, \ldots, n$ and suitably reducing the resulting measurement error. Bandi and Renò (2008) provide more discussions. We now turn to the empirical work.

8. Leverage estimates: the S&P500 index

We apply the leverage estimator to S&P500 index futures data. We employ high-frequency index futures prices from January 1982

[12] This can be easily seen. If the moments of the size distribution are treated as constant parameters, one would have to average functionals of the high-order moments across evaluation points to gain efficiency (see, e.g., Eq. (11) above). If the moments of the size distribution are not treated as parameters, but rather as general functions of the state variable, then no averaging would take place.

[13] Of course, the interval could also be $[i\Delta_{n,T}, i\Delta_{n,T} + \phi]$. Alternatively, it could be symmetric around $i\Delta_{n,T}$. In all cases, in fact, $\phi \to 0$. The interpretation of these alternative choices, and their relation with discrete-time approaches, are discussed in Section 10.2.

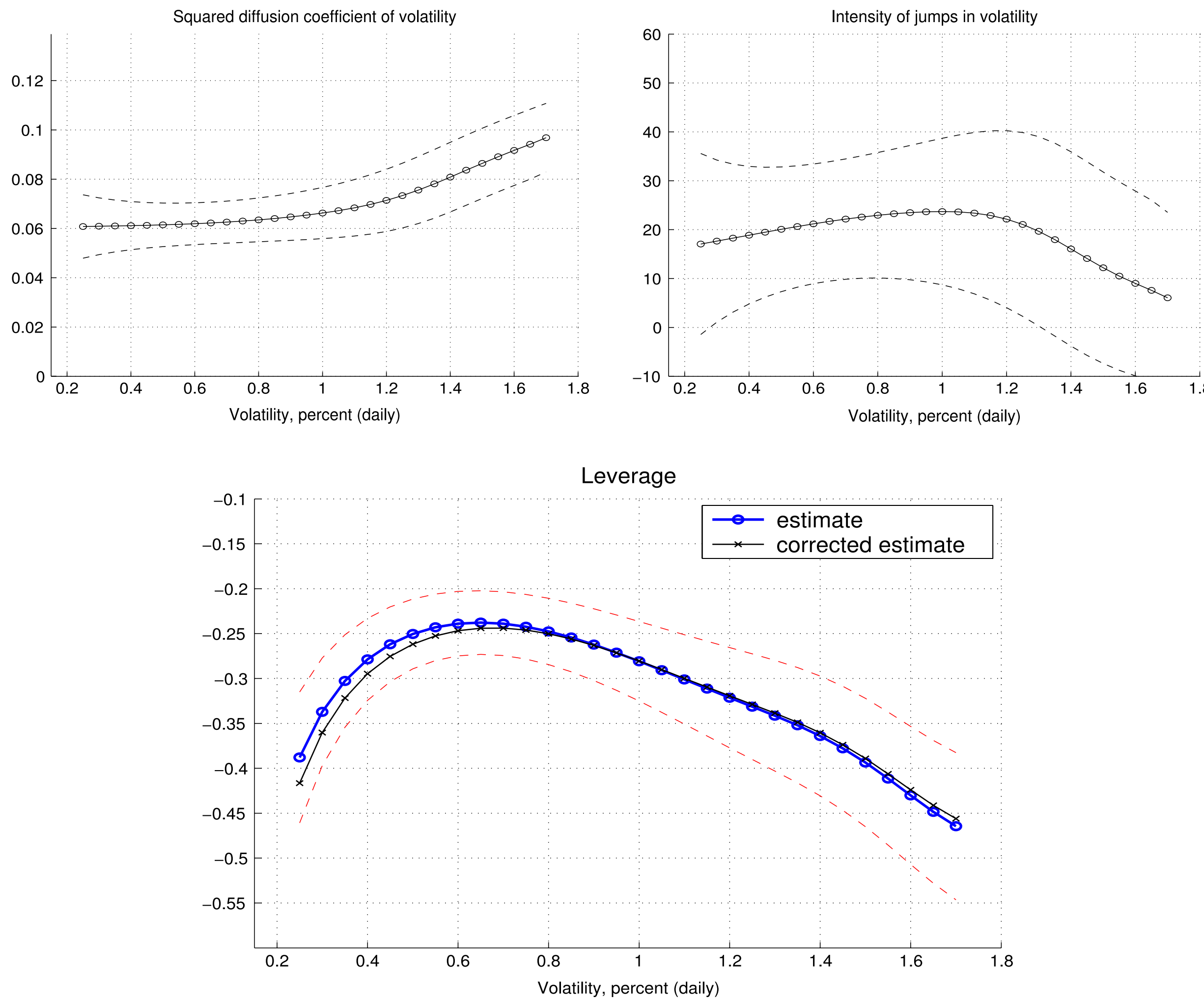

Fig. 1. Functional estimates of $\Lambda^2(\sigma)$ (top left), $\lambda_\sigma(\sigma)$ (top right), and $\rho(\sigma)$ (bottom) for the spot volatility process of the S&P500 index futures. The leverage function is also estimated including the small-sample correction described in the text. The jump intensity is in terms of the number of jumps per year.

to February 2009 for a total of 6675 days. As is customary in the literature, we focus on a broad-based US market index.[14]

Our estimated model is for $\xi(\cdot) = \sqrt{\cdot}$, the volatility process. We also express the moments of interest as a function of spot volatility σ (rather than as a function of spot variance σ^2). The threshold bipower variation estimates are constructed using intra-daily observations over a 6.5-h period interpolated on a 5-min grid (80 intervals per day yielding $k = 80$). The threshold is chosen as in Corsi et al. (2010). Also, we set ϕ equal to a day. Daily estimates are obtained by multiplying the 6.5-h estimates by a factor equal to the ratio between the sum of the squared daily returns and the sum of the 6.5-h variance estimates over the period. We return to the issue of choosing ϕ in practice in Section 9.2.

We consider our more general (and empirically-warranted) case of jumps in both returns and volatility. The nonparametric estimates $\widehat{\vartheta}_{1,1}(\sigma)$ and $\widehat{\vartheta}_j(\sigma)$ with $j = 1, 2, 3, 4$ are implemented using $h_{n,T} = h_s \hat{s} n^{-1/5}$, where $\hat{s}$ is the standard deviation of the time-series of daily spot volatilities. Based on preliminary investigations, we set $h_s = 2$ for $\widehat{\vartheta}_{1,1}(\sigma)$ and $\widehat{\vartheta}_2(\sigma)$ and use $h_s = 4$ for $\widehat{\vartheta}_1(\sigma)$, $\widehat{\vartheta}_3(\sigma)$, and $\widehat{\vartheta}_4(\sigma)$. Nonparametric identification is conducted by virtue of a first-order correction in $\Delta_{n,T}$. This correction is immaterial asymptotically but has the potential to improve finite-sample performance, particularly when evaluating the intensity of the volatility jumps (which, of course, plays a role in the denominator of $\widetilde{\rho}(\sigma)$). Specifically, assuming exponential jumps in volatility with parameter μ_σ, as done earlier, we obtain

$$\vartheta_2(\sigma) \approx \Lambda^2(\sigma) + 2\mu_\sigma^2 \lambda_\sigma(\sigma),$$
$$\vartheta_3(\sigma) \approx 6\mu_\sigma^3 \lambda_\sigma(\sigma) + 3\vartheta_1(\sigma)\vartheta_2(\sigma)\Delta,$$
$$\vartheta_4(\sigma) \approx 24\mu_\sigma^4 \lambda_\sigma(\sigma) + 4\vartheta_1(\sigma)\vartheta_3(\sigma)\Delta + 3\left(\vartheta_2(\sigma)\right)^2 \Delta.$$

We identify the system through

$$\overline{\mu}_\sigma = \frac{1}{4\overline{n}} \sum_{i=1}^{\overline{n}} \left(\frac{\widehat{\vartheta}_4(\widehat{\sigma}_{i\overline{T}/\overline{n}}) - 4\Delta_{n,T}\widehat{\vartheta}_1(\widehat{\sigma}_{i\overline{T}/\overline{n}})\widehat{\vartheta}_3(\widehat{\sigma}_{i\overline{T}/\overline{n}}) - 3\Delta_{n,T}\left(\widehat{\vartheta}_2(\widehat{\sigma}_{i\overline{T}/\overline{n}})\right)^2}{\widehat{\vartheta}_3(\widehat{\sigma}_{i\overline{T}/\overline{n}}) - 3\Delta_{n,T}\widehat{\vartheta}_1(\widehat{\sigma}_{i\overline{T}/\overline{n}})\widehat{\vartheta}_2(\widehat{\sigma}_{i\overline{T}/\overline{n}})} \right),$$

$$\overline{\lambda}_\sigma(\sigma) = \frac{\widehat{\vartheta}_4(\sigma) - 4\Delta_{n,T}\widehat{\vartheta}_1(\sigma)\widehat{\vartheta}_3(\sigma) - 3\Delta_{n,T}\left(\widehat{\vartheta}_2(\sigma)\right)^2}{24\overline{\mu}_\sigma^4},$$

$$\overline{\Lambda}^2(\sigma) = \widehat{\vartheta}_2(\sigma) - 2\overline{\mu}_\sigma^2 \overline{\lambda}_\sigma(\sigma),$$

and, of course,

$$\overline{\rho}(\sigma) = \frac{\widehat{\vartheta}_{1,1}(\sigma)}{\sqrt{\sigma \overline{\Lambda}^2(\sigma)}},$$

[14] Admittedly, classical economic logic behind leverage effects makes the analysis of firm-specific data important. We leave this analysis for future work.

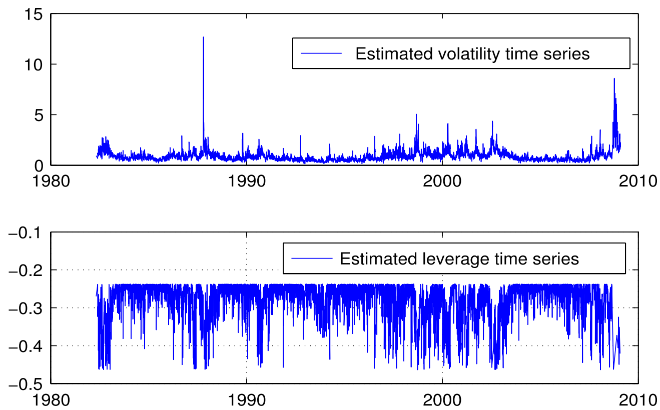

Fig. 2. Estimated time series of spot volatility (top) and leverage (bottom) for the S&P500 index futures.

with $\widehat{\vartheta}_{1,1}(\cdot)$ as defined in Eq. (5).[15] We apply a similar first-order correction to evaluate the confidence bands. These are obtained by using the limiting results in Section 5 for the case with independent return/volatility jumps. Finally, when estimating μ_σ we weigh the addend by virtue of the estimated local time at $\widehat{\sigma}_{iT/n}$ $\forall i = 1, \ldots, n$.

The empirical findings are presented in Figs. 1 and 2. Fig. 1 contains the S&P500 future variance of volatility function, the intensity of the jumps in volatility (expressed in terms of the number of yearly jumps), and the leverage estimates. In all cases, spot volatility is expressed (on the horizontal axis) in daily percentage terms. In agreement with much empirical work in which volatility is filtered from low-frequency (daily) stock returns (see, e.g., Eraker et al., 2003), the volatility of volatility is found to be increasing. The point estimates of the number of yearly jumps are centered around 20 and are statistically significant. The leverage estimates are, as expected, negative and, barring a hump-shape for low volatilities, decreasing with the volatility level. These estimates vary between roughly −0.24 and −0.35 in the most populated volatility range, the value −0.47 being reached for high, seldomly seen, volatility levels. These findings are consistent with the parametric evidence in Section 2. They are, once more, indicative of time-variation in the correlation between shocks to returns and shocks to volatility. For a dynamic assessment of this time variation, see Fig. 2. We evaluate the potential for upward biases (i.e., biases towards zero) of the leverage estimates in Section 9 (Section 9.2).

We, of course, emphasize that, in light of Theorem 5, the reported leverage estimates are theoretically consistent for $\rho(\sigma)$ only in the absence of co-jumps (or if, in the presence of co-jumps, the jump sizes are uncorrelated and the mean of the return jumps is equal to zero, as sometimes assumed in the literature). In spite of the impact of co-jumps on estimating $\rho(\cdot)$, however, the presence of co-jumps does not invalidate the empirical relevance of our methods. As emphasized in Section 6, the methods simply lead to the estimation of a broader notion of leverage. We leave the issue of separately identifying (and evaluating the relative impact of) the different components of total leverage, dubbed "continuous leverage" and "co-jump leverage" earlier, for future work. Ideas on identification were laid out in Remark 10.

9. Finite sample issues

9.1. Discretization: the impact of risk-return trade-offs and volatility feedback effects

Feedback effects generated by unaccounted risk-return trade-offs may imply time-variation of the leverage estimates as a function of spot volatility. To see this, return to the model in Eq. (2) (Section 2). Assume $\mu_t = 0$, as yielded by a commonly-employed martingale difference assumption on the return process. Then, of course,

$$\frac{\mathrm{cov}_t(r_{t+1}, \Delta\sigma_t)}{\mathrm{std}_t(r_{t+1})\mathrm{std}_t(\Delta\sigma_t)} \underset{\text{under } \mu_t=0}{=} \frac{E_t[r_{t+1}\Delta\sigma_t]}{\mathrm{std}_t(r_{t+1})\mathrm{std}_t(\Delta\sigma_t)}.$$

If, however, the true model is, instead, so that $\mu_t = \gamma\sigma_t$, i.e., there is a risk-return trade-off, and spot volatility is mean-reverting as given by $m_t = \beta_0(\beta_1 - \sigma_t)$, for instance, then

$$\frac{E_t[r_{t+1}\Delta\sigma_t]}{\mathrm{std}_t(r_{t+1})\mathrm{std}_t(\Delta\sigma_t)} \underset{\text{if, instead, } \mu_t=\gamma\sigma_t}{=} \frac{\gamma\beta_0\beta_1}{\varLambda} - \frac{\gamma\beta_0\sigma_t}{\varLambda} + \rho. \quad (18)$$

In this case, the empirical leverage should vary (become more negative) with the volatility level because of an unaccounted risk-return trade-off and volatility mean reversion. Clearly, this is not the type of dependence on spot volatility which we have been emphasizing in that our focus has been on ρ, the genuine correlation between shocks to prices and shocks to volatility.

Importantly, however, this effect may have implications for our analysis. Our estimator is, in fact, a sample counterpart to $\frac{E_t[r_{t+1}\Delta\sigma_t]}{\mathrm{std}_t(r_{t+1})\mathrm{std}_t(\Delta\sigma_t)}$. We are of course not erroneously setting $\mu_t = 0$. More simply, the drift components in returns and volatility are asymptotically zero in our continuous-time framework since $\Delta_{n,T} \rightarrow 0$. Hence, asymptotically, our estimator is, as shown, consistent for $\rho(\cdot)$. In finite samples, however, $\Delta_{n,T}$ is fixed (albeit small with daily data), and the estimates may be contaminated by a term (like term $\frac{\gamma\beta_0\beta_1}{\varLambda} - \frac{\gamma\beta_0\sigma_t}{\varLambda}$ in Eq. (18)) which is decreasing with the volatility level, given mean reversion in volatility, thereby potentially inducing spurious (negative) dependence on volatility.

We account for this possibility by recomputing our leverage estimator after subtracting the infinitesimal means $(\mu(\sigma_t)\Delta$ and $(m(\sigma_t) - \mu_\sigma\lambda_\sigma(\sigma_t))\,\Delta)$ from the price changes and the spot volatility changes. Fig. 1 contains the "corrected" leverage estimates. As shown, consistent with asymptotic arguments, the impact of the correction is minimal, thereby giving support to our reported genuine dependence between $\rho(\cdot)$ and σ.

9.2. Measurement error: estimating spot volatility

While we use the language of continuous-time modelling, the model is estimated using daily return data and daily estimates of spot volatility (obtained by virtue of high-frequency price data, as discussed in Section 7). This is, of course, standard in the continuous-time pricing literature (see, e.g., Eraker et al., 2003, and the references therein). What differentiates us from the existing literature is the way in which we filter spot volatility (nonparametrically, using intra-daily data, rather than by simulation methods). In finite samples, however, any notion of spot volatility is effectively a notion of integrated volatility in that ϕ is fixed. Setting ϕ equal to a day, as in our case, is consistent with common practice in the stochastic volatility literature and makes our analysis comparable to existing results.

Having made this point, it is important to evaluate the possible impact of different levels of spot variance measurement error

[15] Even though the previous work on nonparametric jump-diffusion estimation has shown that jump identification by virtue of higher-order moments is empirically feasible (Bandi and Nguyen, 2003 and Johannes, 2004, for instance), estimating higher-order moments is known to be cumbersome. The proposed bias-corrections are meant to improve finite-sample inference by reducing discretization error.

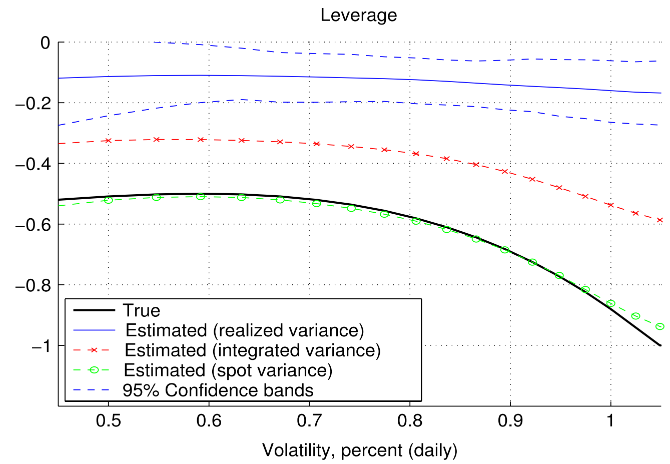

Fig. 3. Estimated leverage on simulated paths. Confidence bands are reported only for the case of estimated integrated variance.

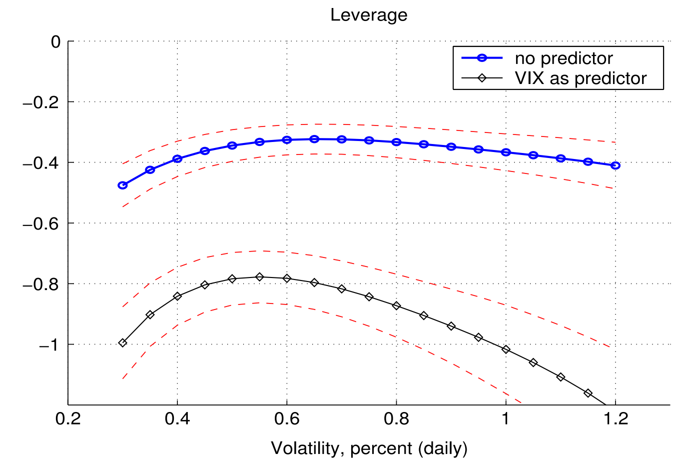

Fig. 4. Leverage estimates for the 1990–2009 period obtained with a threshold bipower variation and with the combination of threshold bipower variation and the VIX used as a predictor.

(inclusive of aggregation issues, for different ϕ levels) on the resulting leverage estimates. In particular, we expect genuine measurement error in spot variance to have a larger (positive) effect on the variance of variance than on the conditional covariance between returns and spot variance changes (since the error is likely to be fairly uncorrelated with the return process). On the other hand, aggregation (as implied by a non-vanishing ϕ in a finite sample) might lead to Epps-type effect in the estimated conditional covariance. Both effects will lead to an attenuation of the leverage estimates.

We investigate this issue by simulation. Consider the discretized system:

$$r_{t,t+\Delta} = \mu\Delta + \sigma_t\sqrt{\Delta}\varepsilon_t^r,$$
$$\sigma_{t+\Delta}^2 - \sigma_t^2 = \kappa(\theta - \sigma_t^2)\Delta + \sigma_\upsilon\sigma_t\sqrt{\Delta}\varepsilon_t^\sigma,$$

where $\{\varepsilon_t^r, \varepsilon_t^\sigma\}$ are standard Gaussian random variables with correlation $\rho(\sigma^2) = \max(-0.9(\sigma^2 - 0.35)^2 - 0.5, -1)$. Assume $\mu = 0.0506, \kappa = 0.025, \theta = 0.75$, and $\sigma_\upsilon = 0.0896$, as in Eraker et al. (2003). The system is simulated over 2500 days. Fig. 3 reports mean leverage estimates across 1000 simulations associated with the true spot variance, the true integrated variance, and realized variance (a classical estimate of integrated variance in the absence of jumps). For each simulated path, we estimate leverage using Eq. (10) in conjunction with the small-sample adjustments discussed in the previous section. Going from spot variance, to daily integrated variance, to estimated daily integrated variance, attenuates the leverage estimates. Importantly for our purposes, the nonlinear shape at high spot variance levels is also attenuated. This is due to the fact that the biases are positively correlated with the variance of the measurement error and this is a positive function of spot variance.

The simulations suggest that the methods proposed in this paper are effective in identifying true leverage with accurate spot variance estimates. The true leverage level is, however, likely lower than estimated on discretized variance estimates, a point made by Aït-Sahalia et al. (2010). Aggregation and integrated variance estimation error, in fact, will bias the leverage estimates towards zero. The extent of these biases is of course a function of the true model, of the level of aggregation, and of the precision of the integrated variance estimates for each level of aggregation. While these simulations should therefore be solely viewed as illustrative, they suggest that the nonconstant pattern found with data is likely genuine. In effect, the finite sample biases are higher at higher variance levels. Thus, if anything, the reported nonconstant leverage shape estimated from data (in Sections 2 and 8) may just be somewhat attenuated.

It is now of interest to study methods which will reduce potential biases. The use of more localized spot variance estimates requires care. It increases the relative size of the measurement error variance since integrated variance notions are more easily estimated than instantaneous notions. In agreement with the simulations reported above, however, the conditions for a vanishing asymptotic measurement error in Theorem 7 provide orders which may be consistent with the empirical usefulness of a higher degree of localization than a daily ϕ (see BR, 2008, for further discussions). Thus, the careful use of more localized spot variance estimates is a promising direction for future research. This direction is pursued by Bandi and Renò (2011). Alternatively, one may "instrument" the daily variance estimates by employing variables which are highly correlated with them but are uncorrelated with measurement error. Consider the regression $\widehat{\sigma}_t = \alpha + \beta\widetilde{\sigma}_t + \eta_t$, where $\widetilde{\sigma}_t$ is the VIX. In the presence of an approximate linear relation between integrated variance and $\widetilde{\sigma}_t$, and since η is uncorrelated with $\widetilde{\sigma}$, the use of the fitted values $\widehat{\alpha} + \widehat{\beta}\widetilde{\sigma}_t$ should lead to consistent estimates of integrated variance. These "relatively less noisy" estimates of integrated variance will translate into lower variance of variance estimates and more negative leverage estimates. Consistent with this observation, for the shorter sample over which VIX values are available (January 2, 1990–February 28, 2009), Fig. 4 reports much lower leverage estimates when regressing the variance estimates on the VIX. Importantly, the VIX-based estimates display a very clear nonlinear pattern which is somewhat attenuated in the case of the original estimates.[16] This effect is consistent with our simulations: higher estimation error increases the leverage level, and attenuates potential nonlinearities, precisely in regions where the spot variance-related biases are more prevalent. These results are suggestive of leverage dynamics which are time-varying, as emphasized in this paper, and more negative than generally found in the literature. The relationship between the VIX and the price volatility is, however, a complicated object which depends on the true model as well as on the features of the volatility risk premia. We view these findings as being promising but requiring a more complete exploration which is better left for future work.

[16] As pointed out, the difference between the threshold bipower variation estimates in this figure and those in Fig. 1 depends on the use of a different sample (a sample over which the VIX values are available).

10. Further discussions: the link to discrete-time modelling

10.1. Issues of timing: contemporaneous vs. lagged leverage

The discrete-time model introduced in Section 2 is clearly consistent with the nonparametric model which is the subject of this paper. Consistency derives from the fact that the implied leverage function ρ_i captures the contemporaneous correlation between daily volatility changes and daily returns (i.e., $\sqrt{TBPV_{t:t-1}} - HAR_{t-1}$ and $r_t = \log p_t - \log p_{t-1}$) for changing values of η_i. Equivalently, our continuous-time leverage estimates – which, by necessity, have to rely on discretizations – capture the contemporaneous correlation between $\sqrt{\frac{TBPV_{iT/n:iT/n-\phi}}{\phi}} - \sqrt{\frac{TBPV_{(i-1)T/n:(i-1)T/n-\phi}}{\phi}}$ and $r_{iT/n} = \log p_{iT/n} - \log p_{(i-1)T/n}$ for a daily ϕ.

While, in order to provide motivation, our treatment in Section 2 had to be coherent with our continuous-time specification (in its discretized form, of course), it is admittedly somewhat different from classical specifications in discrete time. It is, in fact, generally the case that, in reduced-form discrete-time models, leverage is measured as a lagged correlation. To this extent, consider now the model

$$\sqrt{RV_{t:t-1}} = \underbrace{\alpha + \beta_1\sqrt{RV_{t-1:t-2}} + \beta_2\sqrt{RV_{t-1:t-6}} + \beta_3\sqrt{RV_{t-1:t-23}}}_{HAR_{t-1}\text{ component}} + r_{t-1}\sum_{i=1}^{N}\delta_i \mathbf{1}_{\left\{\eta_{i-1}\le\sqrt{RV_{t-1:t-2}}\le\eta_i\right\}} + \varepsilon_t, \tag{19}$$

where the variables have the same interpretation as earlier but the regression is on lagged returns r_{t-1} rather than on contemporaneous returns r_t. We now find that $\widehat{\delta}_i \approx \frac{\widehat{\text{cov}}\left(\sqrt{RV_{t:t-1}}-HAR_{t-1}, r_{t-1}|\eta_{i-1}\le\sqrt{RV_{t-1:t-2}}\le\eta_i\right)}{\widehat{\text{var}}\left(r_{t-1}|\eta_{i-1}\le\sqrt{RV_{t-1:t-2}}\le\eta_i\right)}$, $\widehat{\rho}_i \approx \widehat{\delta}_i\widehat{S}_i$, and $\widehat{S}_i = \frac{\widehat{\text{std}}\left(r_{t-1}|\eta_{i-1}\le\sqrt{RV_{t-1:t-2}}\le\eta_i\right)}{\widehat{\text{std}}\left(\sqrt{RV_{t:t-1}}-HAR_{t-1}|\eta_{i-1}\le\sqrt{RV_{t-1:t-2}}\le\eta_i\right)}$. For a daily ϕ, the corresponding expression in our functional framework would of course be $\widetilde{\widetilde{\rho}}^{\text{lagged}}(\sigma) = \frac{\widehat{\vartheta}^{\text{lagged}}_{1,1}(\sigma)}{\sqrt{\sigma\widetilde{\Lambda}^2(\sigma)}}$ with Eq. (20) given in Box II.

Fig. 5 reports functional estimates of *lagged* leverage as implied by Eq. (20). Table 2 provides estimates of the implied *lagged* leverage given the parametric model in Eq. (19). We notice that lagging reduces the size of leverage as compared to the contemporaneous case. This is apparent in both contexts. Around the bulk of the data (namely, for daily volatility levels between 0.5 and 1.1) both the parametric and the nonparametric model imply increasingly negative leverage for higher volatility levels. The nonparametric case yields a bit more action at the boundaries of the volatility range.

The issue of timing is clearly important conceptually and empirically. However, the need to effectively discretize continuous-time stochastic volatility models for the purpose of their functional estimation may render both lagged and contemporaneous leverage compatible with a continuous-time specification in which leverage is, as is natural in continuous-time modelling, a contemporaneous notion (see the motivating model in Section 2). This can be easily gauged by noticing that, in $\widehat{\vartheta}^{\text{lagged}}_{1,1}(\sigma)$, $\sqrt{\frac{TBPV_{(i+1)T/n+\phi:(i+1)T/n}}{\phi}}$ can be viewed as an estimate of $\sigma_{(i+1)T/n}$ (for a shrinking ϕ "from the left"). Similarly, in $\widehat{\vartheta}_{1,1}(\sigma)$, $\sqrt{\frac{TBPV_{(i+1)T/n:(i+1)T/n-\phi}}{\phi}}$ can also be seen

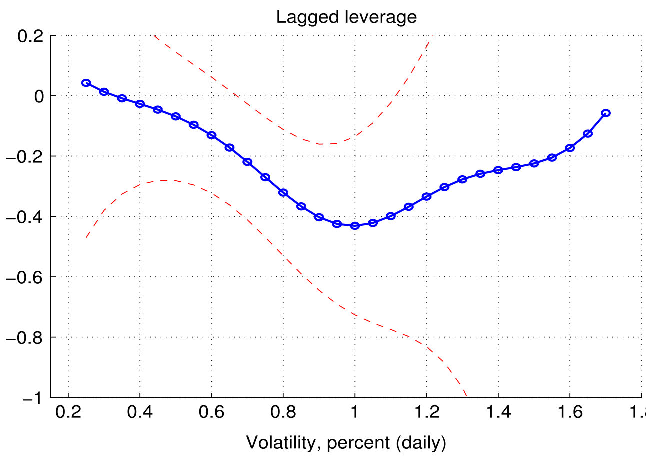

Fig. 5. Functional estimates of *lagged* leverage for the S&P500 index futures.

as an estimate of $\sigma_{(i+1)T/n}$ (for a shrinking ϕ "from the right").[17] In both cases, the resulting leverage estimator may be thought of as estimating the instantaneous correlation between $\sigma_{(i+1)T/n} - \sigma_{iT/n}$ and $\log p_{(i+1)T/n} - \log p_{iT/n}$.[18]

Differently put, again, the fundamental empirical issue is that spot volatility is an instantaneous notion. Filtering it by virtue of high-frequency kernel estimates (as is the case for threshold bipower variation) requires the choice of a window width. The use of a left-window is more naturally comparable to a discrete-time model in which leverage is contemporaneous whereas the use of a right-window is more naturally comparable to a discrete-time model in which leverage is lagged. Needless to say, the window could be symmetric about iT/n but, as implicit in our previous discussion, the use of symmetric windows would complicate intuition as compared to discrete-time specifications.

In sum, we find that the effective use of a flat left-window (leading to a notion of contemporaneous leverage in the language of discrete-time models) yields a more statistically significant and stronger negative correlation between shocks to volatility changes and shocks to prices than the use of a flat right-window (leading to a lagged notion of leverage, again, in the language of discrete-time models). While the estimated leverage *level* is subject to the caveats made in Section 9.2, in both cases leverage becomes more negative with increasing volatility levels around the bulk of the data.

10.2. Alternative conditioning variables

Conditional leverage is modelled in this paper as a function of spot variance. This modelling approach is economically meaningful. It is also natural in continuous-time stochastic volatility models – and, effectively, extends them – since spot variance is used as a conditioning variable both in the return equation (where the return drift may depend on variance as implied by the presence of risk-return trade-offs) and in the variance equation (where the variance drift and diffusion are generally modelled as functions of the variance state). Finally, the approach is technically more feasible, in continuous-time, than alternative conditioning methods. Variance has, in fact, a "stock" nature in continuous-time models. Alternative conditioning variables used in this literature – stock

17 For any $\omega \in \Omega$, if there is a jump at $(i+1)T/n$, on the other hand, the former will lead to $\sigma_{(i+1)T/n-}$, whereas the latter will lead to $\sigma_{(i+1)T/n}$.

18 One important difference is, however, that $\sqrt{\frac{TBPV_{(i+1)T/n:(i+1)T/n-\phi}}{\phi}}$ only uses data before $(i+1)T/n$ to estimate $\sigma_{(i+1)T/n}$, whereas the alternative uses data after $(i+1)T/n$. The former makes conditioning statements possible.

$$\widehat{\vartheta}_{1,1}^{\text{lagged}}(\sigma) = \frac{\sum_{i=1}^{n-1} \mathbf{K}\left(\frac{\sqrt{\frac{TBPV_{iT/n+\phi:iT/n}}{\phi_{n,T}}}-\sigma}{h_{n,T}}\right)(\log p_{(i+1)T/n} - \log p_{iT/n})\left(\sqrt{\frac{TBPV_{(i+1)T/n+\phi:(i+1)T/n}}{\phi_{n,T}}} - \sqrt{\frac{TBPV_{iT/n+\phi:iT/n}}{\phi_{n,T}}}\right)}{\Delta_{n,T}\sum_{i=1}^{n} \mathbf{K}\left(\frac{\sqrt{\frac{TBPV_{iT/n+\phi:iT/n}}{\phi}}-\sigma}{h_{n,T}}\right)}. \quad (20)$$

Box II.

Table 2
Parametric models with lagged leverage.

Model: $\sqrt{RV_{t:t-1}} = HAR_{t-1} + r_{t-1}\sum_{i=1}^{N}\delta_i \mathbf{1}_{\{\eta_{i-1}\le X\le\eta_i\}} + \varepsilon_t$,					
Conditioning on volatility ($X = \sqrt{RV_{t-1:t-2}}$)					
i	Range	δ_i	Standard error	t-stat	Implied leverage
1	$0.250 < X < 0.536$	−0.163	0.110	−1.48	−0.110
2	$0.536 < X < 0.780$	−0.096	0.035	−2.75	−0.075
3	$0.780 < X < 1.700$	−0.136	0.044	−3.06	−0.105
Conditioning on returns ($X = r_{t-1}$)					
1	$-1.700 < X < -0.043$	−0.716	0.114	−6.28	−0.280
2	$-0.043 < X < 0.061$	−0.018	0.132	−0.14	−0.003
3	$0.061 < X < 1.700$	0.158	0.072	2.21	0.072
Conditioning on signed volatility ($X = \sqrt{RV_{t-1:t-2}} \cdot \text{sign}(r_{t-1})$)					
1	$-1.700 < X < -0.525$	−0.338	0.059	−5.70	−0.165
2	$-0.525 < X < 0.531$	−0.129	0.121	−1.07	−0.084
3	$0.531 < X < 1.700$	0.083	0.060	1.40	0.046

returns, for example – have a "flow" nature. This property makes conditioning on variance feasible. When working in continuous-time, conditioning on returns would not be possible (unless the model is specified for increments in returns – which is unusual – rather than for increments in prices).

Alternative approaches may, however, be entertained in discrete time. The parametric findings in Figlewski and Wang (2000) and Yu (2008) are notable in this area. They highlight an asymmetry in the way in which variance responds to price changes—conditionally on a small or positive return, the correlation between variance and price changes is small or positive, it is negative when conditioning on a negative return. In Table 2, we confirm their results, again, in a parametric framework. Positive or negative returns today have a different impact on the sign of *lagged* leverage (namely on the correlation between current returns and future variance).[19] Exploring nonparametrically (in *discrete time*, given the infeasibility of return conditioning in continuous time) their alternative approach is an interesting topic for future work. So is exploring the conditioning power of signed – by contemporaneous returns – volatility. This approach, of course, borrows features from both variance and return conditioning and, again, is only feasible in discrete-time nonparametric or parametric models of leverage. For completeness, the empirical potential of this alternative method in parametric models for contemporaneous and lagged leverage is reported in Tables 1 and 2.

11. Conclusions

We adopt a flexible nonparametric specification in a family of discontinuous stochastic volatility models in order to provide a framework to better understand the nature of the correlation between price and variance shocks. We show that kernel estimates of leverage effects have asymptotic sampling distributions which crucially depend on the features of objects which are fundamentally hard to pin down, namely the probability and size distribution of the individual and joint discontinuities in the price and volatility sample paths. We discuss the nature of this dependence and its implications, while providing tools for the feasible identification of *time-varying* leverage effects under mild parametric structures and weak recurrence assumptions. Our empirical work shows that, for index futures, stronger leverage effects are associated with higher variance regimes. This novel finding is suggestive of the importance of time-varying dynamics in the relation between shocks to prices and shocks to variance.

Appendix. Proofs

We consider $\xi(x) = x$ for brevity and, when not differently indicated, $c_\sigma \sim \exp(\mu_\sigma)$. Alternative specifications for $\xi(\cdot)$ and c_σ may be treated similarly. The notation $\widetilde{\mathbf{K}}(A_i)$ denotes

$$\widetilde{\mathbf{K}}(A_i) = \frac{\frac{1}{h_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}\left(\frac{\sigma^2_{iT/n}-\sigma^2}{h_{n,T}}\right)A_i}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}\left(\frac{\sigma^2_{iT/n}-\sigma^2}{h_{n,T}}\right)}.$$

We also write $\left\{dW_t^r, dW_t^\sigma\right\} = \{\rho(\sigma_t^2)dW_t^1 + \sqrt{1-\rho^2(\sigma_t^2)}\, dW_t^2, dW_t^1\}$ in what follows.

Lemma A.1. *Given a Borel measurable function $g(\cdot)$, consider the quantity*

$$\Psi(\sigma^2)_j = \widetilde{\mathbf{K}}\left(\int_{iT/n}^{(i+1)T/n}\left(\log p_{s-} - \log p_{iT/n}\right)^j \int_Z g(z)\bar{\nu}_\sigma(ds,dz)\right),$$

[19] Both in Figlewski and Wang (2000) and Yu (2008) leverage is measured in its *lagged* version, as described in the previous subsection. When conditioning on the previous period's return, the sign of contemporaneous leverage may be different. We find it to always be negative and highly statistically significant (see Table 1).

where $\bar{\nu}_\sigma$ is the compensated measure of J^σ. If

$$\frac{\Delta_{n,T}}{h_{n,T}^2} \to 0,$$

we have

$$\Psi(\sigma^2)_0 = O_p\left(\sqrt{\frac{1}{h_{n,T}\widehat{\widehat{L}}_{\sigma^2}(T,\sigma^2)}}\right),$$

and

$$\Psi(\sigma^2)_1 = O_p\left(\sqrt{\frac{\Delta_{n,T}}{h_{n,T}\widehat{\widehat{L}}_{\sigma^2}(T,\sigma^2)}}\right).$$

Moreover,

$$\sqrt{h_{n,T}\widehat{\widehat{L}}_{\sigma^2}(T,\sigma^2)}\Psi(\sigma^2)_0 \Rightarrow \mathbf{N}\left(0, \mathbf{K}_2\lambda_\sigma(\sigma^2)\mathbf{E}[g^2]\right),$$

and

$$\sqrt{\frac{h_{n,T}\widehat{\widehat{L}}_{\sigma^2}(T,\sigma^2)}{\Delta_{n,T}}}\Psi(\sigma^2)_1 \Rightarrow \mathbf{N}\left(0, \frac{1}{2}\mathbf{K}_2\left[\lambda_r(\sigma^2)\mathbf{E}[c_r^2]+\sigma^2\right]\lambda_\sigma(\sigma^2)\mathbf{E}[g^2]\right).$$

Remark to Lemma A.1. Similar results hold if we replace $\bar{\nu}_\sigma$ with $\bar{\nu}_r$ (the compensated measure of J^r) or $\bar{\nu}_{r,\sigma}$ (the compensated measure of the contemporaneous jumps between r and σ^2) and, of course, if we replace $\log p_{s-} - \log p_{iT/n}$ with $\sigma_{s-}^2 - \sigma_{iT/n}^2$. □

Lemma A.2. *Consider the quantity*

$$\Phi(\sigma^2)_j = \widetilde{\mathbf{K}}\left(\int_{iT/n}^{(i+1)T/n}\left(\log p_{s-} - \log p_{iT/n}\right)^j \Lambda(\sigma_s^2)dW_s^\sigma\right).$$

If

$$\frac{\Delta_{n,T}}{h_{n,T}^2} \to 0,$$

we have

$$\Phi(\sigma^2)_0 = O_p\left(\sqrt{\frac{1}{h_{n,T}\widehat{\widehat{L}}_{\sigma^2}(T,\sigma^2)}}\right),$$

and

$$\Phi(\sigma^2)_1 = O_p\left(\sqrt{\frac{\Delta_{n,T}}{h_{n,T}\widehat{\widehat{L}}_{\sigma^2}(T,\sigma^2)}}\right).$$

Moreover,

$$\sqrt{h_{n,T}\widehat{\widehat{L}}_{\sigma^2}(T,\sigma^2)}\Phi(\sigma^2)_0 \Rightarrow \mathbf{N}\left(0, \mathbf{K}_2\Lambda^2(\sigma^2)\right),$$

and

$$\sqrt{\frac{h_{n,T}\widehat{\widehat{L}}_{\sigma^2}(T,\sigma^2)}{\Delta_{n,T}}}\Phi(\sigma^2)_1 \Rightarrow \mathbf{N}\left(0, \frac{1}{2}\mathbf{K}_2\left[\lambda_r(\sigma^2)\mathbf{E}[c_r^2]+\sigma^2\right]\Lambda^2(\sigma^2)\right).$$

Remark to Lemma A.2. Similar results hold if we replace W^σ with W^r and, of course, if we replace $\log p_{s-} - \log p_{iT/n}$ with $\sigma_{s-}^2 - \sigma_{iT/n}^2$. □

Proof of Lemma A.1. We prove the lemma for $\Psi(\sigma^2)_1$. The case $\Psi(\sigma^2)_0$ follows analogously. After standardizing the numerator of $\Psi(\sigma^2)_1$, define

$$\sqrt{\frac{h_{n,T}}{\Delta_{n,T}}}\Psi^{\text{num}} := \frac{1}{\sqrt{h_{n,T}\Delta_{n,T}}}\sum_{i=1}^{n-1}\mathbf{K}\left(\frac{\sigma_{iT/n}^2-\sigma^2}{h_{n,T}}\right) \times \int_{iT/n}^{(i+1)T/n}\left(\log p_{s-} - \log p_{iT/n}\right)\int_Z g(z)\bar{\nu}_\sigma(ds,dz) := \sum_{i=1}^{n-1} u_{iT/n,(i+1)T/n},$$

where $u_{iT/n,(i+1)T/n}$ is a martingale difference array. We immediately have

$$\sum_{i=1}^{n-1}\mathbf{E}[u_{iT/n,(i+1)T/n}|\Im_{iT/n}] = 0 \quad \text{a.s.}$$

and, by virtue of Itô's Lemma on $\left(\log p_{s-} - \log p_{iT/n}\right)^2$, following Bandi and Renò (2008),

$$\mathbf{V}_T = \sum_{i=1}^{n-1}\mathbf{E}[u_{iT/n,(i+1)T/n}^2|\Im_{iT/n}] \xrightarrow[\Delta_{n,T}\to 0]{p} \frac{1}{2}\frac{1}{h_{n,T}}\int_0^T \mathbf{K}^2\left(\frac{\sigma_s^2-\sigma^2}{h_{n,T}}\right) \times \left(\sigma_s^2+\lambda_r(\sigma_s^2)\mathbf{E}[c_r^2]\right)\lambda_\sigma(\sigma_s^2)\mathbf{E}[g^2]ds = \widetilde{\mathbf{V}}_T,$$

uniformly in T and $h_{n,T}$ in $\mathcal{H}(\varepsilon) = \left\{\frac{\Delta_{n,T}^{1/2}}{\varepsilon} < h_{n,T} < \varepsilon\right\}$ for a small $\varepsilon > 0$. Now, write Eq. (21) as given in Box III and notice that $\mathbf{K}\left(\frac{\sigma_{iT/n}^2-\sigma^2}{h_{n,T}}\right)\int_{iT/n}^{(i+1)T/n}\left(\log p_{s-} - \log p_{iT/n}\right) \times \int_Z g(z)\bar{\nu}_\sigma(ds,dz) = O_p(\Delta_{n,T})$. Hence, uniformly over $\mathcal{H}(\varepsilon)$, the indicator converges to 1 and Eq. (21) converges in probability to 0 (as $\Delta_{n,T} \to 0$). Using Theorem 3.33 in Jacod and Shiryaev (2003, page 478), we conclude that, uniformly in T and $h_{n,T}$ in $\mathcal{H}(\varepsilon)$, $\sqrt{\frac{h_{n,T}}{\Delta_{n,T}}}\Psi^{\text{num}} \Rightarrow W\left(\widetilde{\mathbf{V}}_T\right)$ and W is an independent Brownian motion. This implies that

$$\sqrt{\frac{h_{n,T}\widehat{\widehat{L}}_{\sigma^2}(T,\sigma^2)}{\Delta_{n,T}}}\Psi(\sigma^2)_1 = \frac{\sqrt{\frac{h_{n,T}}{\Delta_{n,T}}}\Psi^{\text{num}}}{\sqrt{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}\left(\frac{\sigma_{iT/n}^2-\sigma^2}{h_{n,T}}\right)}} \Rightarrow W\left(\frac{\widetilde{\mathbf{V}}_T}{\frac{1}{h_{n,T}}\int_0^T\mathbf{K}\left(\frac{\sigma_s^2-\sigma^2}{h_{n,T}}\right)ds}\right),$$

given Assumption 3.1, as in Bandi and Renò (2008). By the Quotient Limit Theorem (see, e.g., Revuz and Yor, 1998, Theorem 3.12) we now have that, as $T \to \infty$ with $h_{n,T} \to 0$ in $\mathcal{H}(\varepsilon)$,

$$\frac{\widetilde{\mathbf{V}}_T}{\frac{1}{h_{n,T}}\int_0^T\mathbf{K}\left(\frac{\sigma_s^2-\sigma^2}{h_{n,T}}\right)ds} \xrightarrow{\text{a.s.}} \frac{1}{2}\mathbf{K}_2\left(\sigma^2+\lambda_r(\sigma^2)\mathbf{E}[c_r^2]\right)\lambda_\sigma(\sigma^2)\mathbf{E}[g^2]$$

which, using Skorohod embedding arguments as in Theorem 4.1 in Van Zanten (2000), for example, gives the desired result. □

Proof of Lemma A.2. The proof follows the same lines as those of Lemma A.1. □

$$\sum_{i=1}^{n-1} \mathbf{E}\left[u^2_{iT/n,(i+1)T/n}\mathbf{1}_{(|u_{iT/n,(i+1)T/n}|>\epsilon)}|\Im_{iT/n}\right] = \sum_{i=1}^{n-1} \mathbf{E}\left[u^2_{iT/n,(i+1)T/n}|\Im_{iT/n}\right] - \sum_{i=1}^{n-1} \mathbf{E}\left[u^2_{iT/n,(i+1)T/n}\mathbf{1}_{(|u_{iT/n,(i+1)T/n}|\leq\epsilon)}|\Im_{iT/n}\right]$$

$$= \mathbf{V}_T - \sum_{i=1}^{n-1} \mathbf{E}\left[u^2_{iT/n,(i+1)T/n}\mathbf{1}_{\left(\left|\mathbf{K}\left(\frac{\sigma^2_{iT/n}-\sigma^2}{h_{n,T}}\right)\int_{iT/n}^{(i+1)T/n}(\log p_{s-}-\log p_{iT/n})\int_Z g(z)\bar{\nu}_\sigma(ds,dz)\right|\leq\epsilon\sqrt{h_{n,T}\Delta_{n,T}}\right)}|\Im_{iT/n}\right], \quad (21)$$

Box III.

Proof of Theorem 1. Under the assumptions of the theorem, the methods in Bandi and Phillips (2003) yield $\widehat{\vartheta}_2(\sigma^2) \xrightarrow{p} \vartheta_2(\sigma^2) = \Lambda^2(\sigma^2)$ and

$$\sqrt{\frac{h_{n,T}\widehat{\bar{L}}_{\sigma^2}(T,\sigma^2)}{\Delta_{n,T}}}\left(\widehat{\vartheta}_2(\sigma^2) - \Lambda^2(\sigma^2) - \Gamma_{\vartheta_2}(\sigma^2)\right) \Rightarrow \mathbf{N}\left(0, 2\mathbf{K}_2\Lambda^4(\sigma^2)\right). \quad (22)$$

Now consider $\widehat{\vartheta}_{1,1}$. Itô's Lemma gives

$$\begin{aligned}\widehat{\vartheta}_{1,1} = {} & \widetilde{\mathbf{K}}\left(\int_{iT/n}^{(i+1)T/n}\left(\sigma^2_{s-} - \sigma^2_{iT/n}\right)\mu_s ds\right) \\ & + \widetilde{\mathbf{K}}\left(\int_{iT/n}^{(i+1)T/n}\left(\sigma^2_{s-} - \sigma^2_{iT/n}\right)\sigma_s dW^r_s\right) \\ & + \widetilde{\mathbf{K}}\left(\int_{iT/n}^{(i+1)T/n}\left(\log p_{s-} - \log p_{iT/n}\right)m_s ds\right) \\ & + \widetilde{\mathbf{K}}\left(\int_{iT/n}^{(i+1)T/n}\left(\log p_{s-} - \log p_{iT/n}\right)\Lambda(\sigma^2_s)dW^\sigma_s\right) \\ & + \widetilde{\mathbf{K}}\left(\int_{iT/n}^{(i+1)T/n}\rho(\sigma^2_s)\sigma_s\Lambda(\sigma^2_s)ds\right) \\ := {} & \widehat{\vartheta}^c_{1,1}. \end{aligned} \quad (23)$$

Using Lemma A.2, we now obtain

$$\sqrt{\frac{h_{n,T}\widehat{\bar{L}}_{\sigma^2}(T,\sigma^2)}{\Delta_{n,T}}}\left(\widetilde{\mathbf{K}}\left(\int_{iT/n}^{(i+1)T/n}\left(\sigma^2_{s-} - \sigma^2_{iT/n}\right)\sigma_s dW^r_s\right)\right) \Rightarrow \mathbf{N}\left(0, \frac{1}{2}\mathbf{K}_2\Lambda^2(\sigma^2)\sigma^2\right)$$

and

$$\sqrt{\frac{h_{n,T}\widehat{\bar{L}}_{\sigma^2}(T,\sigma^2)}{\Delta_{n,T}}} \times \left(\widetilde{\mathbf{K}}\left(\int_{iT/n}^{(i+1)T/n}\left(\log p_{s-} - \log p_{iT/n}\right)\Lambda(\sigma^2_s)dW^\sigma_s\right)\right) \Rightarrow \mathbf{N}\left(0, \frac{1}{2}\mathbf{K}_2\Lambda^2(\sigma^2)\sigma^2\right).$$

The asymptotic covariance between

$$\widetilde{\mathbf{K}}\left(\int_{iT/n}^{(i+1)T/n}\left(\sigma^2_{s-} - \sigma^2_{iT/n}\right)\sigma_s dW^r_s\right)$$

and

$$\widetilde{\mathbf{K}}\left(\int_{iT/n}^{(i+1)T/n}\left(\log p_{s-} - \log p_{iT/n}\right)\Lambda(\sigma^2_s)dW^\sigma_s\right)$$

is equal to

$$\frac{1}{2}\mathbf{K}_2\Lambda^2(\sigma^2)\sigma^2\rho^2(\sigma^2)\left(\frac{\Delta_{n,T}}{\widehat{\bar{L}}_{\sigma^2}(T,\sigma^2)h_{n,T}}\right).$$

Hence,

$$\sqrt{\frac{h_{n,T}\widehat{\bar{L}}_{\sigma^2}(T,\sigma^2)}{\Delta_{n,T}}}\left(\widehat{\vartheta}_{1,1} - \rho(\sigma^2)\sigma\Lambda(\sigma^2) - \Gamma_{\vartheta_{1,1}}(\sigma^2)\right) \Rightarrow \mathbf{N}\left(0, \mathbf{K}_2\Lambda^2(\sigma^2)\sigma^2(1+\rho^2(\sigma^2))\right)$$

since, by the Quotient Limit Theorem,

$$\widetilde{\mathbf{K}}\left(\int_{iT/n}^{(i+1)T/n}\rho(\sigma^2_s)\sigma_s\Lambda(\sigma^2_s)ds\right) - \vartheta_{1,1}(\sigma^2) = \Gamma_{\vartheta_{1,1}}(\sigma^2) + o(h^2_{n,T}).$$

In the same way,

$$\begin{aligned}\widehat{\vartheta}_2(\sigma^2) = {} & \widetilde{\mathbf{K}}\left(2\int_{iT/n}^{(i+1)T/n}\left(\sigma^2_{s-} - \sigma^2_{iT/n}\right)m_s ds\right) \\ & + \widetilde{\mathbf{K}}\left(2\int_{iT/n}^{(i+1)T/n}\left(\sigma^2_{s-} - \sigma^2_{iT/n}\right)\Lambda(\sigma^2_s)dW^\sigma_s\right) \\ & + \widetilde{\mathbf{K}}\left(\int_{iT/n}^{(i+1)T/n}\Lambda^2(\sigma^2)ds\right).\end{aligned}$$

Hence, the asymptotic covariance between $\widehat{\vartheta}_{1,1}$ and $\widehat{\vartheta}_2$ is given by

$$\mathbf{K}_2 2\Lambda^3(\sigma^2)\sigma\rho(\sigma^2)\left(\frac{\Delta_{n,T}}{h_{n,T}\widehat{\bar{L}}_{\sigma^2}(T,\sigma^2)}\right).$$

Finally, by the delta method

$$\begin{aligned}&\sqrt{\frac{h_{n,T}\widehat{\bar{L}}_{\sigma^2}(T,\sigma^2)}{\Delta_{n,T}}}\left\{\widehat{\rho}(\sigma^2) - \rho(\sigma^2)\right\} \\ &\overset{d}{\sim} \sqrt{\frac{h_{n,T}\widehat{\bar{L}}_{\sigma^2}(T,\sigma^2)}{\Delta_{n,T}}}\left\{\frac{1}{\sigma\Lambda(\sigma^2)}\left\{\widehat{\vartheta}_{1,1}(\sigma^2) - \vartheta_{1,1}(\sigma^2)\right\} \right.\\ &\left. \quad - \frac{\vartheta_{1,1}}{2\sigma\Lambda^3(\sigma^2)}\left\{\widehat{\vartheta}_2(\sigma^2) - \vartheta_2(\sigma^2)\right\}\right\}.\end{aligned} \quad (24)$$

Hence,

$$\begin{aligned}&\mathrm{Asyvar}\left(\sqrt{\frac{h_{n,T}\widehat{\bar{L}}_{\sigma^2}(T,\sigma^2)}{\Delta_{n,T}}}\left\{\widehat{\rho}(\sigma^2) - \rho(\sigma^2)\right\}\right) \\ &= \frac{h_{n,T}\widehat{\bar{L}}_{\sigma^2}(T,\sigma^2)}{\Delta_{n,T}}\left[\frac{1}{\sigma^2\Lambda^2(\sigma)}\mathrm{Asyvar}\left(\widehat{\vartheta}_{1,1}(\sigma^2)\right) \right.\\ &\quad - \frac{\vartheta_{1,1}}{\sigma^2\Lambda^4(\sigma^2)}\mathrm{Asycov}\left(\widehat{\vartheta}_{1,1}(\sigma^2), \widehat{\vartheta}_2(\sigma^2)\right) \\ &\left.\quad + \frac{\vartheta^2_{1,1}}{4\sigma^2\Lambda^6(\sigma^2)}\mathrm{Asyvar}\left(\widehat{\vartheta}_2(\sigma^2)\right)\right]\end{aligned}$$

$$= \left(\frac{h_{n,T}\widehat{\overline{L}}_{\sigma^2}(T,\sigma^2)}{\Delta_{n,T}}\right)\frac{1}{\sigma^2\Lambda^2(\sigma)}\Bigg[\text{Asyvar}\left(\widehat{\vartheta}_{1,1}(\sigma^2)\right) - \frac{\sigma\rho(\sigma^2)}{\Lambda(\sigma^2)}\text{Asycov}\left(\widehat{\vartheta}_{1,1}(\sigma^2),\widehat{\vartheta}_2(\sigma^2)\right) + \left(\frac{\sigma\rho(\sigma^2)}{2\Lambda(\sigma^2)}\right)^2\text{Asyvar}\left(\widehat{\vartheta}_2(\sigma^2)\right)\Bigg]$$

$$= \frac{1}{\sigma^2\Lambda^2(\sigma)}\Bigg[\Lambda^2(\sigma^2)\sigma^2(1+\rho^2(\sigma^2)) - \frac{\sigma\rho(\sigma^2)}{\Lambda(\sigma^2)}(2\Lambda^3(\sigma^2)\sigma\rho(\sigma^2)) + \left(\frac{\sigma\rho(\sigma^2)}{2\Lambda(\sigma^2)}\right)^2 2\Lambda^4(\sigma^2)\Bigg]$$

$$= \mathbf{K}_2\left(1-\frac{1}{2}\rho^2(\sigma^2)\right).$$

As for the asymptotic bias,

$$\widetilde{\Gamma}_\rho(\sigma^2) = \frac{1}{\sigma\sqrt{\vartheta_2(\sigma^2)}}\Gamma_{\vartheta_{1,1}}(\sigma^2) - \frac{\vartheta_{1,1}(\sigma^2)}{2\sigma\sqrt{\vartheta_2^3(\sigma^2)}}\Gamma_{\vartheta_2}(\sigma^2). \quad \square$$

Proof of Theorem 2. Since $\underset{\sim}{J}^\sigma = 0$, the speed of convergence and asymptotic distribution of $\widehat{\vartheta}_2$ do not change. On the other hand, Itô's Lemma now implies that

$$\widehat{\vartheta}_{1,1} = \widehat{\vartheta}^c_{1,1} + \widetilde{\mathbf{K}}\left(\sum_{s\in[iT/n,(i+1)T/n[}\left[(\sigma^2_{s-}-\sigma^2_{iT/n})\Delta\log(p_s)\right]\right),$$

where $\widehat{\vartheta}^c_{1,1}$ is defined in Eq. (23) above. The extra term does not affect consistency (nor does it contribute to the asymptotic bias) and does not change the speed of convergence. However, the limiting variance of $\widehat{\vartheta}_{1,1}$ (and $\widehat{\rho}(\sigma^2)$) does change. Notice, in fact, that by Lemma A.1, after compensating the measure $\nu_r(\cdot,\cdot)$,

$$\sqrt{\frac{h_{n,T}\widehat{\overline{L}}_{\sigma^2}(T,\sigma^2)}{\Delta_{n,T}}}\widetilde{\mathbf{K}}\left(\int_{iT/n}^{(i+1)T/n}\left(\sigma^2_{s-}-\sigma^2_{iT/n}\right)\int_Z c_r\bar{\nu}_r(ds,dz)\right) \Rightarrow \mathbf{N}\left(0,\frac{1}{2}\mathbf{K}_2\Lambda^2(\sigma^2)\lambda_r(\sigma^2)\mathbf{E}[c_r^2]\right).$$

This implies that, by the delta method as in Eq. (24),

$$\text{Asyvar}\left(\sqrt{\frac{h_{n,T}\widehat{\overline{L}}_{\sigma^2}(T,\sigma^2)}{\Delta_{n,T}}}\left\{\widehat{\rho}(\sigma^2)-\rho(\sigma^2)\right\}\right) = \mathbf{K}_2\left[\left(1-\frac{1}{2}\rho^2(\sigma^2)\right)+\frac{1}{2}\frac{\lambda_r(\sigma^2)\mathbf{E}[c_r^2]}{\sigma^2}\right]. \quad \square$$

Proof of Theorem 3. Bandi and Renò (2008) have shown that $\widehat{\Lambda}^2(\sigma^2) \xrightarrow{p} \Lambda^2(\sigma^2)$, if $\frac{\Delta_{n,T}}{h^2_{n,T}} \to 0$ and $h_{n,T}\widehat{\overline{L}}_{\sigma^2}(T,\sigma^2) \xrightarrow{\text{a.s.}} \infty$, and

$$\sqrt{h_{n,T}\widehat{\overline{L}}_{\sigma^2}(T,\sigma^2)}\left\{\widehat{\Lambda}^2(\sigma^2)-\Lambda^2(\sigma^2)\right\} \Rightarrow \mathbf{N}\left(0,\mathbf{K}_2\lambda_\sigma(\sigma^2)\mathbf{E}\left(\left((c_\sigma)^2-\frac{1}{12\mu^2_\sigma}(c_\sigma)^4\right)^2\right)\right), \tag{25}$$

if $\frac{\Delta_{n,T}\sqrt{\widehat{\overline{L}}_{\sigma^2}(T,\sigma^2)}}{h^{3/2}_{n,T}} \xrightarrow{\text{a.s.}} 0$, $h_{n,T}\widehat{\overline{L}}_{\sigma^2}(T,\sigma^2) \xrightarrow{\text{a.s.}} \infty$, and $h^5_{n,T}\widehat{\overline{L}}_{\sigma^2}(T,\sigma^2) \xrightarrow{\text{a.s.}} 0$. Itô's Lemma now implies

$$\widehat{\vartheta}_{1,1} = \widehat{\vartheta}^c_{1,1} + \widetilde{\mathbf{K}}\left(\sum_{s\in[iT/n,(i+1)T/n[}\left[\left(\log p_{s-}-\log p_{iT/n}\right)\Delta\sigma^2_s\right]\right)$$

which gives, as earlier, $\widehat{\vartheta}_{1,1} = O_p\left(\sqrt{\frac{\Delta_{n,T}}{h_{n,T}\widehat{\overline{L}}_{\sigma^2}(T,\sigma^2)}}\right)$. Hence, the rate of convergence of $\widehat{\Lambda}^2(\sigma^2)$ is slower and therefore dominating. In other words, from Eq. (24), the limiting variance of $\widehat{\rho}(\sigma^2)$ solely depends on

$$\sqrt{h_{n,T}\widehat{\overline{L}}_{\sigma^2}(T,\sigma^2)}\frac{\vartheta_{1,1}}{2\sigma\Lambda^3(\sigma^2)}\left\{\widehat{\Lambda}^2(\sigma^2)-\Lambda^2(\sigma^2)\right\} = \sqrt{h_{n,T}\widehat{\overline{L}}_{\sigma^2}(T,\sigma^2)}\frac{\rho(\sigma^2)}{2\Lambda^2(\sigma^2)}\left\{\widehat{\Lambda}^2(\sigma^2)-\Lambda^2(\sigma^2)\right\},$$

thereby giving the stated result. $\square$

Proof of Theorem 4. See Bandi and Renò (2008), Theorem 7. $\square$

Proof of Theorem 5. The result in Eq. (25) still holds. Using Itô's Lemma in the presence of contemporaneous jumps gives

$$\widehat{\vartheta}_{1,1} = \widehat{\vartheta}^c_{1,1} + \widetilde{\mathbf{K}}\left(\sum_{s\in[iT/n,(i+1)T/n[}\left[\Delta\log p_s\Delta\sigma^2_s + \left(\log p_{s-}-\log p_{iT/n}\right)\Delta\sigma^2_s + \Delta\log p_s(\sigma^2_{s-}-\sigma^2_{iT/n})\right]\right).$$

From Lemma A.1. we now obtain that the contemporaneous jump part is $O_p\left(\sqrt{\frac{1}{h_{n,T}\widehat{\overline{L}}_{\sigma^2}(T,\sigma^2)}}\right)$ while all other terms are $O_p\left(\sqrt{\frac{\Delta_{n,T}}{h_{n,T}\widehat{\overline{L}}_{\sigma^2}(T,\sigma^2)}}\right)$. Thus, immediately,

$$\text{Asyvar}(\widehat{\vartheta}_{1,1}(\sigma^2)) = \mathbf{K}_2\lambda_{r,\sigma}(\sigma^2)\mathbf{E}[c_r^2c_\sigma^2]\left(\frac{1}{h_{n,T}\widehat{\overline{L}}_{\sigma^2}(T,\sigma^2)}\right).$$

Similarly, since the lower-order term in $\widehat{\vartheta}_2(\sigma^2)$ is

$$\widetilde{\mathbf{K}}\left(\sum_{s\in[iT/n,(i+1)T/n[}(\Delta\sigma^2_s)^2\right)$$

and in $\widehat{\vartheta}_4(\sigma^2)$ is

$$\widetilde{\mathbf{K}}\left(\sum_{s\in[iT/n,(i+1)T/n[}(\Delta\sigma^2_s)^4\right),$$

we have, since $\widehat{\mu}$ converges at a faster rate,

$$\text{Asycov}(\widehat{\vartheta}_{1,1}(\sigma^2),\widehat{\Lambda}^2(\sigma^2)) = \text{Asycov}\left(\widehat{\vartheta}_{1,1}(\sigma^2),\widehat{\vartheta}_2(\sigma^2)-\frac{\widehat{\vartheta}_4(\sigma^2)}{12\widehat{\mu}^2_\sigma}\right) = \mathbf{K}_2\lambda_{r,\sigma}(\sigma^2)\left(\mathbf{E}[c_rc_\sigma^3]-\frac{\mathbf{E}[c_rc_\sigma^5]}{12\mu^2_\sigma}\right)\left(\frac{1}{h_{n,T}\widehat{\overline{L}}_{\sigma^2}(T,\sigma^2)}\right).$$

Due to the presence of co-jumps, compensation of the object $\sum_{s\in[iT/n,(i+1)T/n[}\left[\Delta\log(p_s)\Delta\sigma^2_s\right]$ requires subtraction of $\int_{iT/n}^{(i+1)T/n}\lambda_{r,\sigma}(\sigma^2_s)\mathbf{E}[c_rc_\sigma]ds$. This term contributes to the probability limit of $\widehat{\vartheta}_{1,1}(\sigma^2)$ which now is $\rho(\sigma^2)\sigma\Lambda(\sigma^2)+\lambda_{r,\sigma}(\sigma^2)\mathbf{E}[c_rc_\sigma]$. In consequence, $\widetilde{\rho}(\sigma^2) \xrightarrow{p} \rho(\sigma^2)+\frac{\lambda_{r,\sigma}(\sigma^2)\mathbf{E}[c_rc_\sigma]}{\sigma\Lambda(\sigma^2)}$. Finally, the delta method yields

$$\mathrm{Asycov}(\{\widetilde{\rho}(\sigma^2)-\rho(\sigma^2)\}) = \frac{\mathbf{K}_2\frac{\lambda_{r,\sigma}(\sigma^2)}{\sigma^2\Lambda^2(\sigma^2)}}{h_{n,T}\widehat{\overline{L}}_{\sigma^2}(T,\sigma^2)}\left(\mathbf{E}[c_r^2c_\sigma^2] - \frac{\vartheta_{1,1}(\sigma^2)}{\Lambda^2(\sigma^2)}2\mathbf{E}\left[c_rc_\sigma\left(c_\sigma^2-\frac{c_\sigma^4}{12\mu_\sigma^2}\right)\right] + \left(\frac{\vartheta_{1,1}(\sigma^2)}{2\Lambda^2(\sigma^2)}\right)^2\mathbf{E}\left[\left(c_\sigma^2-\frac{c_\sigma^4}{12\mu_\sigma^2}\right)^2\right]\right)$$

$$= \frac{\mathbf{K}_2\frac{\lambda_{r,\sigma}(\sigma^2)}{\sigma^2\Lambda^2(\sigma^2)}}{h_{n,T}\widehat{\overline{L}}_{\sigma^2}(T,\sigma^2)}\mathbf{E}\left[\left(c_rc_\sigma-\frac{\vartheta_{1,1}(\sigma^2)}{2\Lambda^2(\sigma^2)}\left(c_\sigma^2-\frac{c_\sigma^4}{12\mu_\sigma^2}\right)\right)^2\right],$$

thereby leading to the stated result. □

Proof of Theorem 6. Without loss of generality, consider the interval $[i\Delta_{n,T}, i\Delta_{n,T}+\phi]$. Write the price process as $\log p = \log\widetilde{p} + J$, where $\log\widetilde{p}$ is the continuous price component and J is the jump component. Let $\theta_t = \xi_t\Theta\left(\frac{\phi}{k}\right)$, where $\Theta\left(\frac{\phi}{k}\right)$ is a real function satisfying $\Theta\left(\frac{\phi}{k}\right) \xrightarrow[\phi\to 0,k\to\infty]{} 0$ and $\frac{1}{\Theta\left(\frac{\phi}{k}\right)}\left(\frac{\phi}{k}\log\left(\frac{\phi}{k}\right)\right) \xrightarrow[\phi\to 0,k\to\infty]{} 0$, and ξ_t is an a.s. bounded process with a strictly positive lower bound. Then, the difference between spot threshold bipower variation applied to $\log p$ $(\widehat{\sigma}^2_{iT/n})$ and spot bipower variation applied to $\log\widetilde{p}$ $(\widetilde{\sigma}^2_{iT/n})$, where $\widetilde{\sigma}^2_{iT/n} = \left(\frac{\varsigma_1^{-2}}{\phi}\right)\sum_{j=2}^k\left(|\Delta_{j-1}\log\widetilde{p}|\ |\Delta_j\log\widetilde{p}|\right)$, is

$$\max_{1\le i\le n-1}\left|\widehat{\sigma}^2_{iT/n}-\widetilde{\sigma}^2_{iT/n}\right| \overset{p}{\sim} \max_{1\le i\le n-1}\left|\left(\frac{\varsigma_1^{-2}}{\phi}\right)\sum_{\left\{|\Delta_{j-1}\log p|^2>\theta_{j-1}\right\}\left\{|\Delta_j\log p|^2>\theta_j\right\}} \times\left(|\Delta_{j-1}\log\widetilde{p}|\ |\Delta_j\log\widetilde{p}|\right)\right| \overset{p}{\sim}\left(\frac{1}{k}\right)\max_{1\le i\le n-1}|N_i| \overset{p}{\sim}\left(\frac{\phi}{k}\right),$$

since N_i, the number of finite-activity Poisson jumps, is of probability order ϕ. For notational convenience, let $\phi < \Delta_{n,T}$.[20] Write $\widetilde{p}_j = \widetilde{p}_{t_j}$, where $t_j = \left[\frac{j}{k}\right]\frac{T}{n} + \left(\frac{j}{k}-\left[\frac{j-1}{k}\right]\right)\phi$, for $j = 1,\ldots,nk$. We have

$$\max_{1\le i\le n-1}\left|\widetilde{\sigma}^2_{iT/n}-\sigma^2_{iT/n}\right| \le \max_{1\le i\le n-1}\left|\widetilde{\sigma}^2_{iT/n} - \left(\frac{\varsigma_1^{-2}}{\phi}\right)\sum_{j=2}^{nk-1}\mathbf{1}_{\left\{i<\frac{j-1}{k}<i+1\right\}}\times\mathbf{E}\left(\left(\left|\int_{t_{j-1}}^{t_j}\sigma_v dW_v^r\right|\left|\int_{t_j}^{t_{j+1}}\sigma_v dW_v^r\right|\right)|\mathcal{F}_{j-1}\right)\right|$$
$$+\max_{1\le i\le n-1}\left|\left(\frac{\varsigma_1^{-2}}{\phi}\right)\sum_{j=2}^{nk-1}\mathbf{1}_{\left\{i<\frac{j-1}{k}<i+1\right\}}\mathbf{E}\left(\left(\left|\int_{t_{j-1}}^{t_j}\sigma_v dW_v^r\right| \times\left|\int_{t_j}^{t_{j+1}}\sigma_v dW_v^r\right|\right)|\mathcal{F}_{j-1}\right) - \frac{1}{\phi}\int_{iT/n}^{iT/n+\phi}\sigma_s^2ds\right|$$
$$+\max_{1\le i\le n-1}\left|\frac{1}{\phi}\int_{iT/n}^{iT/n+\phi}\sigma_s^2ds-\sigma^2_{iT/n}\right|.$$

The first term is dominated by the martingale component

$$\left(\frac{\varsigma_1^{-2}}{\phi}\right)\sum_{j=2}^{nk-1}\mathbf{1}_{\left\{i<\frac{j-1}{k}<i+1\right\}}\left(\left|\int_{t_{j-1}}^{t_j}\sigma_v dW_v^r\right|\left|\int_{t_j}^{t_{j+1}}\sigma_v dW_v^r\right| - \mathbf{E}\left(\left(\left|\int_{t_{j-1}}^{t_j}\sigma_v dW_v^r\right|\left|\int_{t_j}^{t_{j+1}}\sigma_v dW_v^r\right|\right)|\mathcal{F}_{j-1}\right)\right).$$

Using a maximal inequality for martingales as in BR (2008), the term is of order $O_p\left(\sqrt{\frac{\log n}{k}}\right)$. The second term, by adaptation of the method of proof of Theorem 3 in Barndorff-Nielsen et al. (2006) as it applies to A.4 on page 702, is of higher order than the first term. The last term is of probability order $\phi^{1/2}$. This proves the stated result. □

Proof of Theorem 7. We begin with the case of no-jumps, either in prices or in variance, as in Theorem 1. Write

$$\widehat{\mathbf{K}}_i = \mathbf{K}\left(\frac{\widehat{\sigma}^2_{i\Delta_{n,T}}-\sigma^2}{h_{n,T}}\right),\qquad \mathbf{K}_i = \mathbf{K}\left(\frac{\sigma^2_{i\Delta_{n,T}}-\sigma^2}{h_{n,T}}\right),$$

and $\mathbf{M}_{n,k,T} = \max_{1\le i\le n}\left|(\widehat{\sigma}^2_{iT/n}-\sigma^2_{iT/n})\right|$. We denote the infinitesimal moments obtained by virtue of estimated spot variances by $\widetilde{\vartheta}_j(\sigma^2)$ instead of $\widehat{\vartheta}_j(\sigma^2)$. The same notation is used for the feasible infinitesimal covariance estimates. We have

$$\widetilde{\vartheta}_2(\sigma^2) - \frac{\sum_{i=1}^{n-1}\mathbf{K}_i\left[\xi(\sigma^2_{(i+1)T/n})-\xi(\sigma^2_{iT/n})\right]^2}{\Delta_{n,T}\sum_{i=1}^n\mathbf{K}_i}$$
$$= \underbrace{\frac{\sum_{i=1}^{n-1}\widehat{\mathbf{K}}_i\left[\xi(\widehat{\sigma}^2_{(i+1)T/n})-\xi(\widehat{\sigma}^2_{iT/n})\right]^2}{\Delta_{n,T}\sum_{i=1}^n\widehat{\mathbf{K}}_i} - \frac{\sum_{i=1}^{n-1}\mathbf{K}_i\left[\xi(\widehat{\sigma}^2_{(i+1)T/n})-\xi(\widehat{\sigma}^2_{iT/n})\right]^2}{\Delta_{n,T}\sum_{i=1}^n\widehat{\mathbf{K}}_i}}_{(a)}$$
$$+\underbrace{\frac{\sum_{i=1}^{n-1}\mathbf{K}_i\left[\xi(\widehat{\sigma}^2_{(i+1)T/n})-\xi(\widehat{\sigma}^2_{iT/n})\right]^2}{\Delta_{n,T}\sum_{i=1}^n\widehat{\mathbf{K}}_i} - \frac{\sum_{i=1}^{n-1}\mathbf{K}_i\left[\xi(\sigma^2_{(i+1)T/n})-\xi(\sigma^2_{iT/n})\right]^2}{\Delta_{n,T}\sum_{i=1}^n\widehat{\mathbf{K}}_i}}_{(b)}$$
$$+\underbrace{\frac{\sum_{i=1}^{n-1}\mathbf{K}_i\left[\xi(\sigma^2_{(i+1)T/n})-\xi(\sigma^2_{iT/n})\right]^2}{\Delta_{n,T}\sum_{i=1}^n\widehat{\mathbf{K}}_i} - \frac{\sum_{i=1}^{n-1}\mathbf{K}_i\left[\xi(\sigma^2_{(i+1)T/n})-\xi(\sigma^2_{iT/n})\right]^2}{\Delta_{n,T}\sum_{i=1}^n\mathbf{K}_i}}_{(c)}.$$

By the mean-value theorem,

$$\max_{1\le i\le n}\left|\xi(\widehat{\sigma}^2_{iT/n})-\xi(\sigma^2_{iT/n})\right| = O_p\left(\mathbf{M}_{n,k,T}\right).$$

By the properties of Brownian motion,

$$\max_{1\le i\le n}\left|\left\{\xi(\sigma^2_{(i+1)T/n})-\xi(\sigma^2_{iT/n})\right\}^2\right| = O_p\left(\Delta_{n,T}\right).$$

[20] When turning to the estimation of the dynamics, this condition is also implied by Theorem 7.

Thus, we have

$$(a) = \frac{\sum_{i=1}^{n-1}\left(\widehat{\mathbf{K}}_i - \mathbf{K}_i\right)\left[\xi(\widehat{\sigma}^2_{(i+1)T/n}) - \xi(\widehat{\sigma}^2_{iT/n})\right]^2}{\Delta_{n,T}\sum_{i=1}^{n}\widehat{\mathbf{K}}_i}$$

$$= \frac{O_p\left(\frac{\mathbf{M}_{n,k,T}}{h_{n,T}}\right)\sum_{i=1}^{n-1}\mathbf{K}'\left(\frac{\sigma^2_{iT/n}+O_p(\mathbf{M}_{n,k,T})-\sigma^2}{h_{n,T}}\right)\left[\xi(\widehat{\sigma}^2_{(i+1)T/n}) - \xi(\widehat{\sigma}^2_{iT/n})\right]^2}{\Delta_{n,T}\sum_{i=1}^{n}\widehat{\mathbf{K}}_i}.$$

Note that

$$\begin{aligned}
&\max_{1\le i\le n-1}\left[\xi(\widehat{\sigma}^2_{(i+1)T/n}) - \xi(\widehat{\sigma}^2_{iT/n})\right]^2 \\
&= \max_{1\le i\le n-1}\left[\xi(\widehat{\sigma}^2_{(i+1)T/n}) - \xi(\sigma^2_{(i+1)T/n}) + \xi(\sigma^2_{iT/n})\right. \\
&\quad \left. -\ \xi(\widehat{\sigma}^2_{iT/n}) + \xi(\sigma^2_{(i+1)T/n}) - \xi(\sigma^2_{iT/n})\right]^2 \\
&= \max_{1\le i\le n-1}\left|\left[\xi(\widehat{\sigma}^2_{(i+1)T/n}) - \xi(\sigma^2_{(i+1)T/n}) + \xi(\sigma^2_{iT/n})\right.\right. \\
&\quad \left.\left. -\ \xi(\widehat{\sigma}^2_{iT/n}) + \xi(\sigma^2_{(i+1)T/n}) - \xi(\sigma^2_{iT/n})\right]^2\right| \\
&= O_p\left(\mathbf{M}^2_{n,k,T} + \Delta_{n,T} + \mathbf{M}_{n,k,T}\Delta^{1/2}_{n,T}\right). \qquad (26)
\end{aligned}$$

Using the same method of proof as in Bandi and Renò (2008), the bound becomes as in Box IV.

We now turn to (b). Write

$$(b) = \frac{\sum_{i=1}^{n-1}\mathbf{K}_i\left[\left(\xi(\widehat{\sigma}^2_{(i+1)T/n}) - \xi(\widehat{\sigma}^2_{iT/n})\right)^2 - \left(\xi(\sigma^2_{(i+1)T/n}) - \xi(\sigma^2_{iT/n})\right)^2\right]}{\Delta_{n,T}\sum_{i=1}^{n-1}\widehat{\mathbf{K}}_i},$$

and

$$\begin{aligned}
&\max_{1\le i\le n-1}\left|\left(\xi(\widehat{\sigma}^2_{(i+1)T/n}) - \xi(\widehat{\sigma}^2_{iT/n})\right)^2 - \left(\xi(\sigma^2_{(i+1)T/n}) - \xi(\sigma^2_{iT/n})\right)^2\right| \\
&= \max_{1\le i\le n-1}\left|\left(\left(\xi(\widehat{\sigma}^2_{(i+1)T/n}) - \xi(\widehat{\sigma}^2_{iT/n})\right) - \left(\xi(\sigma^2_{(i+1)T/n})\right.\right.\right. \\
&\quad \left.\left. -\ \xi(\sigma^2_{iT/n})\right)\right)\left(\left(\xi(\widehat{\sigma}^2_{(i+1)T/n}) - \xi(\widehat{\sigma}^2_{iT/n})\right)\right. \\
&\quad \left.\left. +\ \left(\xi(\sigma^2_{(i+1)T/n}) - \xi(\sigma^2_{iT/n})\right)\right)\right| \\
&= \max_{1\le i\le n-1}\left|\left(\left(\xi(\widehat{\sigma}^2_{(i+1)T/n}) - \xi(\sigma^2_{(i+1)T/n})\right)\right.\right. \\
&\quad \left. -\ \left(\xi(\widehat{\sigma}^2_{iT/n}) - \xi(\sigma^2_{iT/n})\right)\right) \\
&\quad \times\ \left(\left(\xi(\widehat{\sigma}^2_{(i+1)T/n}) - \xi(\sigma^2_{(i+1)T/n})\right) - \left(\xi(\widehat{\sigma}^2_{iT/n}) - \xi(\sigma^2_{iT/n})\right)\right. \\
&\quad \left.\left. +\ 2\left(\xi(\sigma^2_{(i+1)T/n}) - \xi(\sigma^2_{iT/n})\right)\right)\right| \\
&= O_p(\mathbf{M}^2_{n,k,T} + \mathbf{M}_{n,k,T}\Delta^{1/2}_{n,T}).
\end{aligned}$$

Thus,

$$(b) = \frac{\frac{O_p\left(\mathbf{M}^2_{n,k,T}+\mathbf{M}_{n,k,T}\Delta^{1/2}_{n,T}\right)}{\Delta_{n,T}}\left(\frac{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}_i}{\frac{1}{h_{n,T}}\int_0^T\mathbf{K}\left(\frac{\sigma^2_s-\sigma^2}{h_{n,T}}\right)ds}\right)}{1+O_p\left(\frac{\Delta_{n,T}}{h^2_{n,T}}\right)+O_p\left(\frac{\mathbf{M}_{n,k,T}}{h_{n,T}}\right)}$$

$$= O_p\left(\frac{\mathbf{M}^2_{n,k,T}+\mathbf{M}_{n,k,T}\Delta^{1/2}_{n,T}}{\Delta_{n,T}}\right).$$

Finally, for (c) we have

$$\begin{aligned}
(c) &= \frac{\sum_{i=1}^{n-1}\mathbf{K}_i\left[\xi(\sigma^2_{(i+1)T/n}) - \xi(\sigma^2_{iT/n})\right]^2}{\Delta_{n,T}\sum_{i=1}^{n}\mathbf{K}_i} \times \frac{\Delta_{n,T}\left(\sum_{i=1}^{n-1}\widehat{\mathbf{K}}_i - \sum_{i=1}^{n-1}\mathbf{K}_i\right)}{\Delta_{n,T}\sum_{i=1}^{n}\widehat{\mathbf{K}}_i} \\
&= O_p(1) \\
&\quad \times O_p\left(\frac{\max_{1\le i\le n}\left|\frac{\widehat{\sigma}^2_{i\Delta_{n,T}}-\sigma^2_{i\Delta_{n,T}}}{h_{n,T}}\right|\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n-1}\left|\mathbf{K}'\left(\frac{\sigma^2_{i\Delta_{n,T}}+O_p(\mathbf{M}_{n,k,T})-\sigma^2}{h_{n,T}}\right)\right|}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\widehat{\mathbf{K}}_i}\right) \\
&= O_p(1)O_p\left(\frac{\mathbf{M}_{n,k,T}}{h_{n,T}}\right).
\end{aligned}$$

Now,

$$\begin{aligned}
(a)+(b)+(c) &= O_p\left(\frac{\mathbf{M}^3_{n,k,T}}{\Delta_{n,T}h_{n,T}} + \frac{\mathbf{M}_{n,k,T}}{h_{n,T}} + \frac{\mathbf{M}^2_{n,k,T}}{\Delta^{1/2}_{n,T}h_{n,T}}\right) \\
&\quad + O_p\left(\frac{\mathbf{M}^2_{n,k,T}}{\Delta_{n,T}} + \frac{\mathbf{M}_{n,k,T}}{\Delta^{1/2}_{n,T}}\right) + O_p\left(\frac{\mathbf{M}_{n,k,T}}{h_{n,T}}\right).
\end{aligned}$$

Now notice that, since $\frac{\Delta^{1/2}_{n,T}}{h_{n,T}} \to 0$, $\frac{\mathbf{M}_{n,k,T}}{h_{n,T}} = \frac{\mathbf{M}_{n,k,T}}{\Delta^{1/2}_{n,T}}\frac{\Delta^{1/2}_{n,T}}{h_{n,T}} = o_p\left(\frac{\mathbf{M}_{n,k,T}}{\Delta^{1/2}_{n,T}}\right)$. Similarly, $\frac{\mathbf{M}^2_{n,k,T}}{\Delta^{1/2}_{n,T}h_{n,T}} = \frac{\mathbf{M}_{n,k,T}}{\Delta^{1/2}_{n,T}}\frac{\mathbf{M}_{n,k,T}}{h_{n,T}} = \frac{\mathbf{M}_{n,k,T}}{\Delta^{1/2}_{n,T}}\frac{\mathbf{M}_{n,k,T}}{\Delta^{1/2}_{n,T}}\frac{\Delta^{1/2}_{n,T}}{h_{n,T}} = o_p\left(\frac{\mathbf{M}^2_{n,k,T}}{\Delta_{n,T}}\right)$. In addition,

$$\begin{aligned}
\frac{\mathbf{M}^3_{n,k,T}}{\Delta_{n,T}h_{n,T}} &= \frac{\mathbf{M}^2_{n,k,T}\mathbf{M}_{n,k,T}}{\Delta_{n,T}h_{n,T}} \\
&= \frac{\mathbf{M}^2_{n,k,T}}{\Delta_{n,T}}\frac{\mathbf{M}_{n,k,T}}{h_{n,T}} = \frac{\mathbf{M}^2_{n,k,T}}{\Delta_{n,T}}\frac{\mathbf{M}_{n,k,T}}{\Delta^{1/2}_{n,T}}\frac{\Delta^{1/2}_{n,T}}{h_{n,T}} \\
&= o_p\left(\frac{\mathbf{M}^3_{n,k,T}}{\Delta^{3/2}_{n,T}}\right). \qquad (27)
\end{aligned}$$

Thus,

$$(a)+(b)+(c) = O_p\left(\frac{\mathbf{M}_{n,k,T}}{\Delta^{1/2}_{n,T}}\right). \qquad (28)$$

Bandi and Renò (2008) show that, for the case of jumps in volatility, if $j = 1$

$$\widetilde{\vartheta}_j(\sigma^2) = \widehat{\vartheta}_j(\sigma^2) + O_p\left(\frac{\mathbf{M}_{n,k,T}}{\Delta_{n,T}}\right) \qquad (29)$$

and, if $j > 1$,

$$\widetilde{\vartheta}_j(\sigma^2) = \widehat{\vartheta}_j(\sigma^2) + O_p\left(\frac{\mathbf{M}_{n,k,T}}{\Delta^{1/2}_{n,T}h_{n,T}}\right), \qquad (30)$$

which is a milder order than $O_p\left(\frac{\mathbf{M}_{n,k,T}}{\Delta_{n,T}}\right)$ since $\frac{\Delta^{1/2}_{n,T}}{h_{n,T}} \to 0$. As for the covariance term, following Bandi and Renò (2008, 2011), it is easy to show that

$$\widetilde{\vartheta}_{1,1}(\sigma^2) - \widehat{\vartheta}_{1,1}(\sigma^2) = O_p\left(\frac{\mathbf{M}_{n,k,T}}{\Delta^{1/2}_{n,T}}\right), \qquad (31)$$

if the jumps are in prices or in variance only (or absent). However, we have $\widetilde{\vartheta}_{1,1}(\sigma^2) - \widehat{\vartheta}_{1,1}(\sigma^2) = O_p\left(\frac{\mathbf{M}_{n,k,T}}{\Delta^{1/2}_{n,T}h_{n,T}}\right)$ if the price and

$$(a) = \frac{O_p\left(\frac{\mathbf{M}_{n,k,T}}{h_{n,T}}\right) O_p\left(\mathbf{M}_{n,k,T}^2 + \Delta_{n,T} + \mathbf{M}_{n,k,T}\Delta_{n,T}^{1/2}\right) \frac{1}{h_{n,T}} \sum_{i=1}^{n-1} \mathbf{K}'\left(\frac{\sigma_{iT/n}^2 + O_p(\mathbf{M}_{n,k,T}) - \sigma^2}{h_{n,T}}\right)}{\frac{\Delta_{n,T}}{h_{n,T}} \sum_{i=1}^{n} \widehat{\mathbf{K}}_i}$$

$$= \frac{O_p\left(\frac{\mathbf{M}_{n,k,T}}{\Delta_{n,T} h_{n,T}}\right) O_p\left(\mathbf{M}_{n,k,T}^2 + \Delta_{n,T} + \mathbf{M}_{n,k,T}\Delta_{n,T}^{1/2}\right) \left(\frac{\frac{1}{h_{n,T}} \int_0^T \left|\mathbf{K}'\left(\frac{\sigma_s^2 - \sigma^2}{h_{n,T}}\right)\right| ds}{\frac{1}{h_{n,T}} \int_0^T \mathbf{K}\left(\frac{\sigma_s^2 - \sigma^2}{h_{n,T}}\right) ds} + O_p\left(\frac{\Delta_{n,T}}{h_{n,T}^2}\right)\right)}{1 + O_p\left(\frac{\Delta_{n,T}}{h_{n,T}^2}\right) + O_p\left(\frac{\mathbf{M}_{n,k,T}}{h_{n,T}}\right)}$$

$$= O_p\left(\frac{\mathbf{M}_{n,k,T}}{\Delta_{n,T} h_{n,T}}\left(\mathbf{M}_{n,k,T}^2 + \Delta_{n,T} + \mathbf{M}_{n,k,T}\Delta_{n,T}^{1/2}\right)\right).$$

Box IV.

variance jumps are joint. Now consider Theorem 1. The delta method in Eq. (24), combined with Eqs. (28) and (31) gives

$$\begin{aligned} &\frac{1}{\sigma\Lambda(\sigma^2)}\left\{\widetilde{\vartheta}_{1,1}(\sigma^2) - \vartheta_{1,1}(\sigma^2)\right\} \\ &\quad - \frac{\vartheta_{1,1}}{2\sigma\Lambda^3(\sigma^2)}\left\{\widetilde{\vartheta}_2(\sigma^2) - \vartheta_2(\sigma^2)\right\} \\ &= \frac{1}{\sigma\Lambda(\sigma^2)}\left\{\widehat{\vartheta}_{1,1}(\sigma^2) - \vartheta_{1,1}(\sigma^2)\right\} \\ &\quad - \frac{\vartheta_{1,1}}{2\sigma\Lambda^3(\sigma^2)}\left\{\widehat{\vartheta}_2(\sigma^2) - \vartheta_2(\sigma^2)\right\} + O_p\left(\frac{\mathbf{M}_{n,k,T}}{\Delta_{n,T}^{1/2}}\right), \end{aligned}$$

which leads to the desired result. The presence of only jumps in returns (Theorem 2) is analogous. We now turn to Theorem 3, jumps in volatility only. Using Eq. (30), notice that

$$\widetilde{\Lambda}^2(\sigma^2) - \Lambda^2(\sigma^2) = \widehat{\Lambda}^2(\sigma^2) - \Lambda^2(\sigma^2) + O_p\left(\frac{\mathbf{M}_{n,k,T}}{\Delta_{n,T}^{1/2} h_{n,T}}\right),$$

since

$$\begin{aligned} &\frac{1}{\overline{n}}\sum_{i=1}^{\overline{n}} \frac{\widetilde{\vartheta}_4(\sigma_{i\overline{T}/\overline{n}}^2)}{4\widetilde{\vartheta}_3(\sigma_{i\overline{T}/\overline{n}}^2)} - \mu_\sigma \overset{d}{\sim} \frac{1}{\overline{n}}\sum_{i=1}^{\overline{n}} \left(\frac{\left(\widetilde{\vartheta}_4(\sigma_{i\overline{T}/\overline{n}}^2) - \vartheta_4(\sigma_{i\overline{T}/\overline{n}}^2)\right)}{4\vartheta_3(\sigma_{i\overline{T}/\overline{n}}^2)}\right. \\ &\quad \left. - \frac{\vartheta_4(\sigma_{i\overline{T}/\overline{n}}^2)\left(\widetilde{\vartheta}_3(\sigma_{i\overline{T}/\overline{n}}^2) - \vartheta_3(\sigma_{i\overline{T}/\overline{n}}^2)\right)}{4\left(\vartheta_3(\sigma_{i\overline{T}/\overline{n}}^2)\right)^2}\right) \\ &\overset{d}{\sim} \frac{1}{\overline{n}}\sum_{i=1}^{\overline{n}} \left(\frac{\left(\widehat{\vartheta}_4(\sigma_{i\overline{T}/\overline{n}}^2) - \vartheta_4(\sigma_{i\overline{T}/\overline{n}}^2)\right)}{4\vartheta_3(\sigma_{i\overline{T}/\overline{n}}^2)}\right. \\ &\quad \left. - \frac{\vartheta_4(\sigma_{i\overline{T}/\overline{n}}^2)\left(\widehat{\vartheta}_3(\sigma_{i\overline{T}/\overline{n}}^2) - \vartheta_3(\sigma_{i\overline{T}/\overline{n}}^2)\right)}{4\left(\vartheta_3(\sigma_{i\overline{T}/\overline{n}}^2)\right)^2}\right) + O_p\left(\frac{\mathbf{M}_{n,k,T}}{\Delta_{n,T}^{1/2} h_{n,T}}\right). \end{aligned}$$

This argument leads to the desired result. Similar expressions apply in the case of Theorems 4 and 5. □

References

Aït-Sahalia, Y., Fan, J., Li, Y., 2010. Studying the leverage effect based on high-frequency data. Working Paper.

Andersen, T.G., Benzoni, L., Lund, J., 2002. Towards an empirical foundation for continuous-time equity return models. Journal of Finance 57, 1047–1091.

Bakshi, G., Cao, C., Chen, Z., 1997. Empirical performance of alternative option pricing models. Journal of Finance 52, 2003–2049.

Bandi, F.M., Nguyen, T., 2003. On the functional estimation of jump-diffusion models. Journal of Econometrics 116, 293–328.

Bandi, F.M., Phillips, P.C.B., 2003. Fully nonparametric estimation of scalar diffusion models. Econometrica 71, 241–283.

Bandi, F.M., Phillips, P.C.B., 2010. Nonstationary continuous-time models. In: Hansen, L.P., Aït-Sahalia, Y. (Eds.), Handbook of Financial Econometrics. Elsevier Science, New York.

Bandi, F.M., Renò, R., 2008. Nonparametric stochastic volatility. Available at SSRN: http://ssrn.com/abstract=1158438.

Bandi, F.M., Renò, R., 2011. Price and volatility co-jumps. Working Paper.

Barndorff-Nielsen, O.E., Graversen, S.E., Jacod, J., Shephard, N., 2006. Limit theorems for bipower variation in financial econometrics. Econometric Theory 22, 677–719.

Barndorff-Nielsen, O.E., Shephard, N., 2004. Power and bipower variation with stochastic volatility and jumps. Journal of Financial Econometrics 2, 1–37.

Black, F., 1976. Studies of stock market volatility changes. In: Proceedings of the American Statistical Association, Business and Economic Statistics Section, pp. 177–181.

Christie, A., 1982. The stochastic behaviour of common stock variances: value, leverage, and interest rate effects. Journal of Financial Economics 3, 407–431.

Corsi, F., 2009. A simple approximate long-memory model of realized volatility. Journal of Financial Econometrics 7, 174–196.

Corsi, F., Pirino, D., Renò, R., 2010. Threshold bipower variation and the impact of jumps on volatility forecasting. Journal of Econometrics 159 (2), 276–288.

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

Engle, R., Ng, V., 1993. Measuring and testing the impact of news in volatility. Journal of Finance 43, 1749–1778.

Eraker, B., Johannes, M., Polson, N., 2003. The impact of jumps in volatility and returns. Journal of Finance 12, 69–1300.

Figlewski, S., Wang, X., 2000. Is the "leverage effect" a leverage effect? Working Paper.

Glosten, L.R., Jagannathan, R., Runkle, D., 1993. Relationship between the expected value and the volatility of the nominal excess return on stocks. Journal of Finance 48, 1779–1802.

Harvey, A.C., Shephard, N., 1996. The estimation of an asymmetric stochastic volatility model for asset returns. Journal of Business and Economic Statistics 14, 429–434.

Jacod, J., Shiryaev, A.N., 2003. Limit Theorems for Stochastic Processes. Springer.

Jacod, J., Todorov, V., 2009. Testing for common arrivals of jumps for discretely observed multidimensional processes. Annals of Statistics 37, 1792–1838.

Jacquier, E., Polson, N., Rossi, P., 1994. Bayesian analysis of stochastic volatility models. Journal of Business and Economic Statistics 12, 371–389.

Jacquier, E., Polson, N., Rossi, P., 2004. Bayesian analysis of stochastic volatility models with fat tails and correlated errors. Journal of Econometrics 122, 185–212.

Johannes, M., 2004. The statistical and economic role of jumps in continuous-time interest rate models. Journal of Finance 227–260.

Mancini, C., 2009. Nonparametric threshold estimation for models with stochastic diffusion coefficients and jumps. Scandinavian Journal of Statistics 36, 279–296.

Mancini, C., Gobbi, F., 2012. Identifying the Brownian covariation from the co-jumps given discrete observations. Econometric Theory (forthcoming).

Mykland, P.A., Zhang, L., 2009. Inference for continuous semimartingales observed at high frequency. Econometrica 77, 1403–1455.

Revuz, D., Yor, M., 1998. Continuous Martingales and Brownian Motion. Springer-Verlag.

Van Zanten, J.H., 2000. A multivariate central limit theorem for continuous local martingales. Statistics and Probability Letters 50 (3), 229–235.

Yu, J., 2005. On leverage in a stochastic volatility model. Journal of Econometrics 127, 165–178.

Yu, J., 2008. A semiparametric stochastic volatility model. Working Paper.