NONPARAMETRIC LEVERAGE EFFECTS*

Federico M. Bandi† and Roberto Renò‡

First draft: October 16, 2008
This draft: October 1, 2009

Abstract

Vast empirical evidence points to the existence of a negative correlation, named "leverage effect," between shocks in volatility and shocks in 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 volatility, or both. Leverage is defined as a flexible function of the state of the firm, as summarized by the spot volatility 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 volatility levels.

*We thank Robert Taylor, the Co-Editors (Roberto Mariano, Zhijie Xiao, and Jun Yu), two anonymous referees, and the participants at the Festschrift in honor of Peter Phillips (Singapore, July 14-15, 2008) for very useful comments and discussions.

† *Carey Business School, Johns Hopkins University, 10 N. Charles Street, Baltimore, MD 21201, USA. E-mail: fbandi1@jhu.edu*

‡ *Dipartimento di Economia Politica, Università di Siena, Piazza San Francesco 7, 53100, Siena, Italy. E-mail: reno@unisi.it.*

1 Introduction

Shocks to returns have been found to be negatively correlated with shocks to volatility. Due to its traditional economic justification in the case of common stocks, this stylized fact has been termed "leverage effect." 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 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 next period 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 implied by negative shocks to prices) will increase stock volatility, whereas decreases in financial leverage (as implied by positive shocks to prices) will decrease stock volatility (Black, 1976). While this simple logic is appealing, the economics of leverage effects continues to remain somewhat controversial.

The empirical relevance of (alternative forms of) leverage effects has, however, been broadly established. The time-varying volatility literature has emphasized the significance of feedback effects between (innovations in) returns and volatility 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 volatility) is generally assumed to be a constant value. Indeed, the economic logic described earlier suggests, among other implications, that the magnitude of the leverage effect should be roughly independent of whether shocks to prices are positive or negative. Some recent work has, however, emphasized that there may be important asymmetries in the way in which volatility responds to prices changes. Figlewski and Wang (2000) and Yu (2008), for example, stress that, in the presence of positive shocks to prices, return volatility might not change (or even change positively) whereas negative shocks will likely lead to an increase in volatility, coherently with traditional - negative - leverage effects. Allowing for changing levels of correlation between innovations to prices and volatility may be empirically important. We argue that this time-variation is, indeed, coherent with basic finance principles and study an alternative, continuous-time, framework for doing so.

This paper provides a nonparametric treatment of leverage estimation in the context of a stochastic volatility, jump-diffusion, model with discontinuities in returns, volatility, 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 volatility as implied by the presence of risk-return trade-offs) and in the volatility equation (where the volatility drift and diffusion are generally modelled as functions of the volatility state). It is also economically meaningful in that it amounts to modelling the size of the correlation between shocks to returns and shocks to volatility as a function of the riskiness of the firm (as summarized by its spot volatility). Classical economic logic, as described above, implies that, for a fairly stable volatility of the firm value, higher return volatility should be positively correlated with higher financial leverage. Since times of high return volatility should be associated with relatively higher financial leverage, price changes of a certain size should have a larger, more negative, 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,

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

In other words, changes in the value of equity induce negative volatility changes whose magnitude depend on the volatility itself. In light of this discussion, we conjecture that times of higher return volatility (generally associated with higher financial leverage) are times in which shocks to returns are more negatively correlated with volatility changes. Our empirical work supports this conjecture. Importantly, we emphasize that we do not commit to a classical Modigliani-Miller economy. We solely use it here as a motivating example to stress that even classical economic principles imply time-variation in leverage. For instance, while we allow leverage to be a declining function of spot volatility (as implied by a Modigliani-Miller economy), we are general about its shape.

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 volatility processes. Several cases are considered: absence of jumps in either process, jumps in returns, jumps in volatility, independent jumps in returns and volatility, contemporaneous jumps (or co-jumps) in returns and volatility. We show that the fastest convergence rate to a (mixed) normal distribution arises in the absence of jumps in both returns and volatility. The presence of jumps in returns (without jumps in volatility) 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 volatility (without jumps in returns) is quite different in that consistent estimation of the volatility process' diffusion function 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. Interestingly, the addition of jumps in the return process (in a model now with independent jumps in returns and volatility) does not modify this result. Since, in the presence of jumps in volatility, the asymptotic variance of the leverage estimator is already completely induced by that of the spot variance's diffusion estimator, the addition of independent jumps in returns has now (contrary to the continuous volatility case with jumps in returns) no effect even in terms of decreased asymptotic efficiency. Finally, allowing for co-jumps in returns and volatility may yield inconsistency of the leverage estimator unless the jump sizes are independent and the jumps in returns are mean zero. In this case, we show that the limiting distribution is completely driven by the features of the price/variance discontinuities and discuss ways to re-establish consistency of the leverage estimator by virtue, for example, of appropriate asymptotic bias corrections. Our limit theory hinges on weaker conditions than stationarity. We solely assume recurrence[1] 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 kernel estimates of spot variance obtained by virtue of high-frequency asset price data and, when possible, allow for market microstructure noise as in Bandi and Renò (2008), BR henceforth. In particular, we show how 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 future data and are, therefore, for a broad-based US market index. We find important time-variation (as a function of spot volatility) in the correlation between price and volatility shocks. As conjectured, leverage increases (i.e., becomes more negative) with the volatility level. We find leverage values around -0.2 for low volatilities and about -0.5 for high volatility values.

BR (2008) have recently introduced a novel nonparametric approach to the estimation of stochastic volatility models with jumps (in returns and in volatility). The approach does not hinge on the filtering of the latent spot volatility process by virtue of simulation methods relying on low frequency return data (as is common in the parametric literature) but on the preliminary filtering on spot volatility using high-frequency return data. The resulting procedure is (semi-)nonparametric in nature in spite of the unobservability of spot volatility. From a methodological standpoint, the present paper relates to BR (2008) in that the price/volatility evolutions are jump-diffusive, the dynamics are estimated nonparametrically, and the preliminary spot volatility 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 BR (2008), (*1*) we offer economic justification for modelling leverage as a function of the latent spot volatility process, (*2*) we provide a complete limiting theory for nonparametric leverage estimators allowing for increasing layers of complications in the assumed model ranging from simple diffusive structures to structures with co-jumps and correlated jump sizes, (*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 nonparametric results with appropriately-defined (and theoretically justified) reduced-form parametric models.

The paper proceeds as follows. The next section provides *parametric* motivation for allowing leverage to be a function of the spot volatility process. This section is meant to introduce our approach in a more familiar (parametric) setting and show that the leverage dynamics are not a by-product of the use of nonparametric methods. Section 3 lays out our nonlinear, continuous-time, stochastic volatility model with (possibly correlated) jumps. In Section 4 through 8, we present the relevant limiting theory for an *observable* spot volatility process. Section 9 discusses "generalized leverage." Section 10 adapts the theory in Section 4 through 8 to the empirically-relevant case of an *estimated* spot volatility process. Section 11 provides empirical results pointing to the existence of a higher (more negative) leverage in the presence of higher volatility. Section 12 contains simulations. Section 13 offers further discussions (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

[1]For a review of nonparametric methods for continuous-time models under Harris recurrence we refer the interested reader to Bandi and Phillips (2009).

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 14 concludes. The proofs are in the Appendix.

2 Time-varying leverage: a parametric motivation

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[2] and write

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

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 iid noise. We infer "implied" leverage over the i^{th} interval ($\widehat{\rho}_i$) by the corresponding δ_i coefficient provided such a coefficient is rescaled appropriately. In fact, $\widehat{\delta}_i \approx \frac{\widehat{cov}\left(\sqrt{RV_{t:t-1}} - HAR_{t-1}, r_t | \eta_i \le \sqrt{RV_{t-1:t-2}} \le \eta_{i-1}\right)}{\widehat{var}\left(r_t | \eta_i \le \sqrt{RV_{t-1:t-2}} \le \eta_{i-1}\right)}$. Hence, $\widehat{\rho}_i \approx \widehat{\delta}_i\widehat{S}_i$, where the scaling factor $\widehat{S}_i$ is equal to $\frac{\widehat{std}\left(r_t | \eta_i \le \sqrt{RV_{t-1:t-2}} \le \eta_{i-1}\right)}{\widehat{std}\left(\sqrt{RV_{t:t-1}} - HAR_{t-1} | \eta_i \le \sqrt{RV_{t-1:t-2}} \le \eta_{i-1}\right)}$. In what follows, we set $N = 3$ and report three values of $\widehat{\rho}_i$ corresponding to alternative volatility levels.

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

$$\begin{aligned} d\log p_t &= \mu_t dt + \sigma_t dW_t^r, \\ d\sigma_t &= m_t dt + \Lambda dW_t^\sigma, \end{aligned}$$

where μ_t, m_t are adapted processes and $\{W^r, W^\sigma\}$ are correlated standard Brownian motions with $< dW_t^r, dW_t^\sigma > = \rho dt$. This system can be readily discretized as follows

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

or, equivalently,

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

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

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

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

In agreement with much recent empirical work, we may capture persistence (and model 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 $\delta_t = \frac{\Lambda\rho}{\sigma_t}$.[3] This last expression provides a more structural justification for estimating ρ parametrically by virtue of $\widehat{\delta}\widehat{S}$. The case $N \geq 1$, of course, leads to

$$\sigma_{t+1} \approx \sigma_t + m_t + \sum_{i=1}^{N} \delta_i \left(r_{t+1} - \mu_t\right) \mathbf{1}_{\{\eta_i \leq \sigma_t \leq \eta_{i-1}\}} + \varepsilon_{t+1}$$

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

Using the threshold bipower variation estimator to define $RV_{t-i:t-i-1}$ for all i (see Section 11 for details), 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 ($\widehat{\rho}_i \approx \widehat{\delta}_i\widehat{S}_i$) are -0.152, -0.187, and -0.273. In agreement with our discussion in the Introduction, this result is suggestive of an increasing leverage for higher volatility levels. 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 volatility level.

3 The continuous-time setting

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

$$\begin{aligned} r_{t,t+dt} &= d\log(p_t) = \mu_t dt + \sigma_t dW_t^r + dJ_t^r, & (4)\\ d\xi(\sigma_t^2) &= m_t dt + \Lambda(\sigma_t^2) dW_t^\sigma + dJ_t^\sigma, & (5) \end{aligned}$$

where $\{J_t^r, J_t^\sigma\}$ is a bi-dimensional compound Poisson process and, given the bi-variate standard Brownian motion $\left\{W_t^1, W_t^2\right\}$,

$$\{dW_t^r, dW_t^\sigma\} = \{\rho(\sigma_t^2)dW_t^1 + \sqrt{1-\rho^2(\sigma_t^2)}dW_t^2, dW_t^1\}$$

denotes increments of a bi-dimensional drift-less diffusion or, more explicitly in our framework, contemporaneous (continuous) shocks to returns and shocks to monotonic transformations $\xi(\cdot)$ of spot variance. Clearly, $< dW_t^r, dW_t^r > = < dW_t^s, dW_t^s > = dt$ and

$$< dW_t^r, dW_t^s > \; = \rho(\sigma_t^2)dt,$$

thereby implying that the function $\rho(\cdot)$ is a well-defined infinitesimal (conditional) correlation between continuous shocks to returns and continuous shocks to spot variance (if bounded between -1 and 1).

[3] 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.

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 jumps in variance, $\lambda_r(\sigma_t^2)$, the intensity of 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 a linear (in volatility) model. 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 the previous section.

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 recurrent strong solution of* (4)-(5) *exists.*

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 practise. 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 10 discusses the case of spot variance estimation by virtue of kernel methods applied to high-frequency price data. This section 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 allow for market microstructure noise in spot variance estimation, when possible.

Define the infinitesimal moments

$$\begin{aligned}\vartheta_{1,1}(\sigma^2) &= \lim_{\Delta\to 0}\frac{1}{\Delta}\mathbf{E}\left[(\log p_{t+\Delta} - \log p_t)\left(\xi(\sigma^2_{t+\Delta}) - \xi(\sigma_t^2)\right)\left|\sigma_t^2 = \sigma^2\right.\right],\\ \vartheta_j(\sigma^2) &= \lim_{\Delta\to 0}\frac{1}{\Delta}\mathbf{E}\left[\left(\xi(\sigma^2_{t+\Delta}) - \xi(\sigma_t^2)\right)^j\left|\sigma_t^2 = \sigma^2\right.\right], \qquad j = 1,2,...\end{aligned}$$

and the corresponding Nadaraya-Watson kernel estimators

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

$$\widehat{\vartheta}_j(\sigma^2) = \frac{\sum_{i=1}^{n-1}\mathbf{K}\left(\frac{\sigma^2_{iT/n}-\sigma^2}{h_{n,T}}\right)\left(\xi(\sigma^2_{(i+1)T/n}) - \xi(\sigma^2_{iT/n})\right)^j}{\Delta_{n,T}\sum_{i=1}^{n}\mathbf{K}\left(\frac{\sigma^2_{iT/n}-\sigma^2}{h_{n,T}}\right)} \qquad j = 1,2,..., \tag{7}$$

[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.

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.

Assumption 2. $\mathbf{K}(\cdot)$ *is a bounded, continuously-differentiable, symmetric, and nonnegative function whose derivative* $\mathbf{K}'(\cdot)$ *is absolutely integrable and bounded and for which* $\int \mathbf{K}(s)ds = 1$, $\mathbf{K}_1 = \int s^2\mathbf{K}(s)ds < \infty$, *and* $\mathbf{K}_2 = \int \mathbf{K}^2(s)ds < \infty$.

In what follows, the kernel estimators $\widehat{\vartheta}_{1,1}(\sigma^2)$ and $\widehat{\vartheta}_j(\sigma^2)$ (with $j = 1, \ldots$) will be employed to provide point-wise estimates of $\rho(\sigma^2)$ (and $\rho(\sigma)$, of course) in various scenarios allowing for jumps in prices, jumps in volatility, or both. We will resort to the classical notation $\xrightarrow{p}$, $\Rightarrow$ to denote 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.[5] 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

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

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, that σ^2_t spends in a small right neighborhood of x between time 0 and time T. Analogously, time can be measured in chronological units by defining

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

For a fixed T, and under assumptions, $\widehat{\overline{L}}_{\sigma^2}(T,x)$ is known to estimate the latter. Similarly, for an enlarging T and, again, under 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 stricter 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 $\widehat{\overline{L}}_{\sigma^2}(T,x)$.[6] For more discussions of 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 (2009).

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

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

where $S'(x)$ is the first derivative of the scale function, i.e.,

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

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

[6]Of course, $\widehat{\overline{L}}_{\sigma^2}(T,x)$ diverges at speed T in stationary models.

We now list conditions which the smoothing sequence $h_{n,T}$ and the chronological local time $\overline{L}_{\sigma^2}(T,\cdot)$ ought to satisfy (as $\Delta_{n,T} \to 0$ with $n, T \to \infty$) for the validity of the limiting results in the following sections.

Assumption 3. *Let $\overline{L}_{\sigma^2}(T,\cdot) \sim v(T)$, where $v(T)$ is a regularly-varying function at infinity. Let, also, $h_{n,T} \to 0$ with $\Delta_{n,T} \to 0$ for $n, T \to \infty$.*

3.1 $\frac{v(T)}{h_{n,T}}\left(\Delta_{n,T}\log\frac{1}{\Delta_{n,T}}\right)^{1/2} \to 0.$

3.2 $\frac{h_{n,T}^5 v(T)}{\Delta_{n,T}} \to C$, *where C is a suitable constant.*

3.3 $h_{n,T}v(T) \to \infty$

3.4 $h_{n,T}^5 v(T) \to C$, *where C is a suitable constant.*

We begin with a relevant benchmark case.

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

Write the kernel leverage estimator as

$$\widehat{\rho}(\sigma^2) = \frac{\widehat{\vartheta}_{1,1}(\sigma^2)}{\sigma\sqrt{\widehat{\vartheta}_2(\sigma^2)}}. \tag{8}$$

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 our object of interest. This result is shown in Theorem 1 below.[7]

Theorem 1. *Assume $J^r = J^\sigma = 0$. If Assumption 3.1 is satisfied, then $\widehat{\rho}(\sigma^2) \xrightarrow{p} \rho(\sigma^2)$. If Assumption 3.2 is also satisfied, then*

$$\sqrt{\frac{h_{n,T}\widehat{\overline{L}}_{\sigma^2}(T,\sigma^2)}{\Delta_{n,T}}}\left\{\widehat{\rho}(\sigma^2) - \rho(\sigma^2) - \widetilde{\Gamma}_\rho(\sigma^2)\right\} \Rightarrow \mathbf{N}\left(0, \mathbf{K}_2\left[1 - \frac{1}{2}\rho^2(\sigma^2)\right]\right), \tag{9}$$

where

$$\widetilde{\Gamma}_\rho(\sigma^2) = \frac{1}{\sigma\sqrt{\vartheta_2(\sigma^2)}}\Gamma_{\vartheta_{1,1}}(\sigma^2) - \frac{\vartheta_{1,1}(\sigma^2)}{2\sigma\sqrt{\vartheta_2^3(\sigma^2)}}\Gamma_{\vartheta_2}(\sigma^2). \tag{10}$$

Proof. *See Appendix A.*

Remark 1. In the absence of jumps in either volatility or returns, the leverage estimator converges at speed $\sqrt{\frac{h_{n,T}\widehat{\overline{L}}_{\sigma^2}(T,\sigma^2)}{\Delta_{n,T}}} \sim \sqrt{\frac{h_{n,T}v(T)}{\Delta_{n,T}}}$. 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

[7] 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 $\overline{T}$.

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. (9) is maximal (and 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 The discontinuous case: $J^r \neq 0,\ J^\sigma = 0$

Consider the same estimator as in Eq. (8) above. The case with jumps in returns is presented in Theorem 2.

Theorem 2. *Assume $J^r \neq 0$ and $J^\sigma = 0$. Under Assumption 3.1 and Assumption 3.2, we obtain*

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

Proof. *See Appendix A.*

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 ($\mathbf{E}[c_r^2]$).

6 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 $j = 3, 4, ...$) as proposed by BR (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

$$\widetilde{\rho}(\sigma^2) = \frac{\widehat{\vartheta}_{1,1}(\sigma^2)}{\sigma f(\widehat{\vartheta}_2, \widehat{\vartheta}_3, \widehat{\vartheta}_4, ...)(\sigma^2)}, \tag{11}$$

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

To lay out ideas, we turn to a specific identification scheme. Assume $\xi(\sigma^2) = \sigma^2$ and assume the variance jumps are exponentially distributed, i.e., $c_\sigma \sim \exp(\mu_\sigma)$. This specification is widely used in the parametric literature on volatility estimation (see, e.g., Eraker et al., 2003) and has been shown to perform very satisfactorily in a (semi-)nonparametric context (BR, 2008).[8] The proposed model implies

$$\begin{aligned}
\vartheta_2(\sigma^2) &= \Lambda^2(\sigma^2) + 2\mu_\sigma^2\lambda_\sigma(\sigma^2), \\
\vartheta_3(\sigma^2) &= 6\mu_\sigma^3\lambda_\sigma(\sigma^2), \\
\vartheta_4(\sigma^2) &= 24\mu_\sigma^4\lambda_\sigma(\sigma^2).
\end{aligned}$$

[8] Specifically, the nonparametric jump tests in BR (2008) provide statistical support for the presence of exponential jumps in variance.

Hence,

$$\widehat{\mu}_\sigma = \frac{1}{\overline{n}}\sum_{i=1}^{\overline{n}} \frac{\widehat{\vartheta}_4(\sigma^2_{i\overline{T}/\overline{n}})}{4\widehat{\vartheta}_3(\sigma^2_{i\overline{T}/\overline{n}})}, \tag{12}$$

$$\widehat{\lambda}_\sigma(\sigma^2) = \frac{\widehat{\vartheta}_4(\sigma^2)}{24\widehat{\mu}_\sigma^4},$$

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 μ_σ, $\lambda_\sigma(\sigma^2)$, and $\Lambda^2(\sigma^2)$, respectively.[9] These estimators can been shown to be consistent (with probability one) and (mixed) normally distributed (see, e.g., BR, 2008). Thus,

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

Importantly, alternative estimation schemes may be adopted by imposing, for instances, different distributional assumptions on c_σ. In these cases, our adopted methods can be adapted 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 Assumption 3.1 and Assumption 3.3 are satisfied, then $\widetilde{\rho}(\sigma^2) \xrightarrow{p} \rho(\sigma^2)$. If $\rho(\sigma^2) \neq 0$ and Assumption 3.4 is also satisfied, then*

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

with

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

Proof. *See Appendix A.*

Remark 4. The estimator $f(\widehat{\vartheta}_2,\widehat{\vartheta}_3,\widehat{\vartheta}_4,...)(\cdot)$ 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 $f(\widehat{\vartheta}_2,\widehat{\vartheta}_3,\widehat{\vartheta}_4,...)(\cdot)$ 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 $f(\widehat{\vartheta}_2,\widehat{\vartheta}_3,\widehat{\vartheta}_4,...)(\cdot)$ (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. (13) can be more explicitly expressed as $46\mathbf{K}_2\frac{\rho^2(\sigma^2)\lambda_\sigma(\sigma^2)}{\Lambda^4(\sigma^2)}\mu_\sigma^4$.

[9] $\widehat{\mu}_\sigma$ is defined over a number of observations $\overline{n}$ growing to infinity over a fixed time span $\overline{T}$. This is simply done for technical reasons in order to simplify the limiting behavior 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 $\overline{T}$ 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 ($T \to \infty$). 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.

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

Consider the same estimator as in Eq. (11) above.

Theorem 4. *Assume $J^r \neq 0$, $J^\sigma \neq 0$, and $J^r \perp J^\sigma$. Under Assumptions 3.1, 3.3, and 3.4:*

$$\sqrt{h_{n,T}\widehat{\overline{L}}_{\sigma^2}(T,\sigma^2)}\left\{\widetilde{\rho}(\sigma^2)-\rho(\sigma^2)-\widetilde{\Gamma}_{\widetilde{\rho}}(\sigma^2)\right\}\Rightarrow \mathbf{N}\left(0,\mathbf{K}_2\frac{\rho^2(\sigma^2)}{4\Lambda^4(\sigma^2)}\lambda_\sigma(\sigma^2)\mathbf{E}\left(\left(c_\sigma^2-\frac{1}{12\mu_\sigma^2}c_\sigma^4\right)^2\right)\right) \quad (14)$$

with

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

Proof. *See Appendix A.*

Remark 6. Adding independent jumps in returns to the case of jumps in volatility does not modify the limiting distribution of $\widetilde{\rho}(\cdot)$ ((14) is the same as (13)). This is, of course, in contrast to the case where jumps in returns are added to the case of no jumps. 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 the denominator, $\sigma f(\widehat{\vartheta}_2,\widehat{\vartheta}_3,\widehat{\vartheta}_4,...)(\cdot)$.

8 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. (11).

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 3.3 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 Assumption 3.4 is also satisfied, then

$$\sqrt{h_{n,T}\widehat{\overline{L}}_{\sigma^2}(T,\sigma^2)}\left\{\widetilde{\rho}(\sigma^2)-\Xi(\sigma^2)-\widetilde{\Gamma}_{\widetilde{\rho}}(\sigma^2)\right\}\Rightarrow\mathbf{N}\left(0,\mathbf{K}_2 V_\Xi\right) \quad (15)$$

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

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

Proof. *See Appendix A.*

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, $f(\widehat{\vartheta}_2, \widehat{\vartheta}_3, \widehat{\vartheta}_4, ...)(\cdot)$, 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 $f(\widehat{\vartheta}_2, \widehat{\vartheta}_3, \widehat{\vartheta}_4, ...)(\cdot)$.

Corollary to Theorem 5 (A relevant example: 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:

$$\sqrt{h_{n,T}\widehat{\overline{L}}_{\sigma^2}(T,\sigma^2)}\left\{\widetilde{\rho}(\sigma^2) - \rho(\sigma^2) - \widetilde{\Gamma}_{\widetilde{\rho}}(\sigma^2)\right\} \Rightarrow \mathbf{N}\left(0, \mathbf{K}_2 \frac{\lambda_{r,\sigma}(\sigma^2)}{\sigma^2\Lambda^2(\sigma^2)}\left[2\sigma_r^2\mu_\sigma^2 + 46\frac{\sigma^2\rho^2(\sigma^2)}{\Lambda^2(\sigma^2)}\mu_\sigma^4\right]\right), \quad (16)$$

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

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

Remark 8. (Contemporaneous and non-contemporaneous jumps) The theorem solely 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).[10] In general, we could assume $J_r = J_r^* + J_r^{||}$ and $J_\sigma = J_\sigma^* + J_\sigma^{||}$, with $J_r^* \perp J_\sigma^*$, $J_r^* \perp J_r^{||}$, $J_r^* \perp J_\sigma^{||}$ and $J_\sigma^* \perp J_\sigma^{||}$, $J_\sigma^* \perp J_r^{||}$. More explicitly, we could assume that both processes comprise two components, $J_{r,\sigma}^*$ and $J_{r,\sigma}^{||}$, which are independent of each other and all others with the exception of $J_r^{||}$ and $J_\sigma^{||}$, which are dependent. Denote now by $c_{r,\sigma}^*$ and $\lambda_{r,\sigma}^*$ the jump sizes and intensities of the jumps of the independent components $J_{r,\sigma}^*$. Similarly, denote by $c_{r,\sigma}^{||}$ and $\lambda_{r,\sigma}^{||} = \lambda_r^{||} = \lambda_\sigma^{||}$ the jump sizes of the dependent components and the (common) intensity of the common shocks. The result in Eq. (15) continues to hold and may be re-written as follows:

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

where

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

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 ($c_{r,\sigma}^{||}$ and $\lambda_{r,\sigma}^{||}$) as well as on the size and intensity of the independent volatility jumps (c_σ^* and λ_σ^*).

[10]Their model SVIJ assumes independence of the jumps as in Section 6 above.

9 A notion of "generalized leverage"

The expression

$$\underbrace{\Xi(\sigma^2)}_{\text{generalized leverage}} = \underbrace{\rho(\sigma^2)}_{\text{continuous leverage}} + \underbrace{\frac{\lambda^{\|}_{r,\sigma}(\sigma^2)}{\sigma\Lambda(\sigma^2)}\mathbf{E}[c^{\|}_r c^{\|}_\sigma]}_{\text{co-jump leverage}} \tag{17}$$

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 in prices and "continuous" shocks in 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 in that one may standardize by 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 which one would employ if the jumps were assumed not to play a role. In this case, the expression would clarify the impact of the co-jump component on the continuous leverage estimates, as done in the previous section.

More generally, 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 to both. Disentangling the relative impact of alternative components is economically important. The following remarks provide further discussions on this issue. We however plan to elaborate on this issue, both theoretically and empirically, in future work.

Remark 9. (Co-jump identification) Methods have been put forward to identify the co-jumps. Gobbi and Mancini (2008), 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

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

where $r(\delta)$ is a function such that $\frac{\delta \log\left(\frac{1}{\delta}\right)}{r(\delta)} \to 0$ 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 ($\lambda^{\|}_{1,2}$) may be evaluated, possibly under an assumption of constancy, by computing the in-sample frequency of co-jumps. Similarly, the expected first cross-moment ($\mathbf{E}[c^{\|}_1 c^{\|}_2]$) can be consistently identified by virtue of sample averages of the co-jumps (under, of course, stationarity of the jump distribution). Finally, given $\widehat{\Lambda}^2(\cdot) = f(\widehat{\vartheta}_2, \widehat{\vartheta}_3, \widehat{\vartheta}_4, ...)(\cdot)$, the continuous leverage function $\rho(\sigma^2)$ may be estimated consistently by virtue of $\widetilde{\rho}(\sigma^2) - \frac{\widehat{\lambda}^{\|}_{r,\sigma}(\sigma^2)}{\sigma\sqrt{f(\widehat{\vartheta}_2,\widehat{\vartheta}_3,\widehat{\vartheta}_4)(\sigma^2)}}\widehat{\mathbf{E}}[c^{\|}_r c^{\|}_\sigma]$.

Remark 10. (More on co-jump identification) An alternative procedure is implied by the recent work of Bandi and Renò (2009) on cross-moment estimation. We present some preliminary ideas here. The procedure hinges on the nonparametric identification of the return/variance cross-moments of order p_1 and p_2, i.e.,

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

with $p_1 \geq p_2 \geq 0$. The intuition is as follows: the cross-moments of order higher than $1,1$ (i.e., so that $p_1 \geq p_2 \geq 1$ 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 $\lambda^{||}_{r,\sigma}$ and $\mathbf{E}[c^{||}_r c^{||}_\sigma]$. As an example, assume $\xi(\cdot) = \log(\cdot)$. Using the notation in Remark 8, express

$$\begin{aligned} J_r &= J^*_r + J^{||}_r = c^{||}_r \left(N^*_r + N^{||}_{r,\sigma} \right) \\ J_\sigma &= J^*_\sigma + J^{||}_\sigma = c^{||}_\sigma \left(N^*_\sigma + N^{||}_{r,\sigma} \right) \end{aligned}$$

where N^*_r, N^*_σ, and $N^{||}_{r,\sigma}$ are independent Poisson processes. Now write

$$\begin{pmatrix} c^{||}_r \\ c^{||}_\sigma \end{pmatrix} \sim \mathbf{N}(0, \Sigma_J) \text{ and } \Sigma_J = \begin{pmatrix} \sigma^2_{J,r} & \blacklozenge \\ \rho_J \sigma_{J,r} \sigma_{J,\sigma} & \sigma^2_{J,\sigma} \end{pmatrix},$$

where Σ_J may be a function of σ^2.[11] Then,

$$\left\{ \begin{array}{l} \vartheta_{1,0}(\cdot) = \mu(\cdot) \\ \vartheta_{2,0}(\cdot) = \cdot + (\lambda^*_r(\cdot) + \lambda^{||}_{r,\sigma}(\cdot))\sigma^2_{J,r} \\ \vartheta_{4,0}(\cdot) = 3(\lambda^*_r(\cdot) + \lambda^{||}_{r,\sigma}(\cdot))\sigma^4_{J,r} \\ \vartheta_{6,0}(\cdot) = 15(\lambda^*_r(\cdot) + \lambda^{||}_{r,\sigma}(\cdot))\sigma^6_{J,r} \end{array} \right\}$$

yield identification of $\mu(\cdot)$, $\sigma^2_{J,r}$, and $\left\{ \lambda^*_r(\cdot) + \lambda^{||}_{r,\sigma}(\cdot) \right\}$. Similarly,

$$\left\{ \begin{array}{l} \vartheta_{0,1}(\cdot) = m(\cdot) \\ \vartheta_{0,2}(\cdot) = \Lambda^2(\cdot) + (\lambda^*_\sigma(\cdot) + \lambda^{||}_{r,\sigma}(\cdot))\sigma^2_{J,\sigma} \\ \vartheta_{0,4}(\cdot) = 3(\lambda^*_\sigma(\cdot) + \lambda^{||}_{r,\sigma}(\cdot))\sigma^4_{J,\sigma} \\ \vartheta_{0,6}(\cdot) = 15(\lambda^*_\sigma(\cdot) + \lambda^{||}_{r,\sigma}(\cdot))\sigma^6_{J,\sigma} \end{array} \right\}$$

identify $m(\cdot)$, $\Lambda(\cdot)$, $\sigma^2_{J,\sigma}$, and $\left\{ \lambda^*_\sigma(\cdot) + \lambda^{||}_{r,\sigma}(\cdot) \right\}$. Finally,

$$\left\{ \begin{array}{l} \vartheta_{1,1}(\cdot) = \rho(\cdot)\sqrt{\cdot}\Lambda(\cdot) + \lambda^{||}_{r,\sigma^2}(\cdot)\rho_J \sigma_{J,r}\sigma_{J,\sigma} \\ \vartheta_{2,2}(\cdot) = \lambda^{||}_{r,\sigma^2}(\cdot)\sigma^2_{J,r}\sigma^2_{J,\sigma}(1 + 2\rho^2_J) \\ \vartheta_{3,1}(\cdot) = 3\lambda^{||}_{r,\sigma^2}(\cdot)\rho_J \sigma^3_{J,r}\sigma_{J,\sigma} \\ \vartheta_{1,3}(\cdot) = 3\lambda^{||}_{r,\sigma^2}(\cdot)\rho_J \sigma_{J,r}\sigma^3_{J,\sigma} \end{array} \right\}$$

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

10 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 $j = 1, 2, 3, 4, ...$ one must replace $\sigma^2_{iT/n}$ with spot variance estimates $\widehat{\sigma}^2_{iT/n}$.

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

To this extent, assume availability of k (possibly not equi-spaced) high-frequency observations over each time interval $[i\Delta_{n,T}, i\Delta_{n,T} - \phi_{n,T}]$.[12] Assume these k observations are employed to estimate the *integrated variance* of the logarithmic price process over each interval (i.e., $\int_{i\Delta_{n,T}-\phi_{n,T}}^{i\Delta_{n,T}} \sigma_s^2 ds$) by virtue of a *generic* estimator $\widehat{V}_{iT/n}$. Now define $\widehat{\sigma}^2_{iT/n} = \frac{\widehat{V}_{iT/n}}{\phi_{n,T}}$ and let, asymptotically, $\phi_{n,T} \to 0$ as $k \to \infty$. BR (2008) have shown that a theory of consistent (and asymptotically mixed normal) *spot variance* estimation based on $\widehat{\sigma}^2_{iT/n}$ can be derived, for a large number of estimators $\widehat{V}_{iT/n}$ introduced in the literature, by controlling the rate at which $\phi_{n,T}$ vanishes as k goes off to infinity. Specifically, for these estimators, they have shown that, if $\phi^{\beta}_{n,T}k^{\alpha} \to \infty$, $\phi^{\beta}_{n,T}k^{\alpha}\left(\phi_{n,T}\log\left(\frac{1}{\phi_{n,T}}\right)\right)^{1/2} \to 0$, and under additional conditions which are specific to the $\widehat{V}_{iT/n}$ used,

$$\phi^{\beta}_{n,T}k^{\alpha}\left(\frac{\widehat{V}_{iT/n}}{\phi_{n,T}} - \sigma^2_{iT/n}\right) \Rightarrow \mathbf{MN}\left(0, a\left(\sigma^4_{iT/n}\right)^{\eta} + b\right), \tag{18}$$

where α, β, a, b, and η are parameters which are, again, specific to $\widehat{V}_{iT/n}$. For appropriate choices of these parameters, the above weak convergence result is satisfied, for example, by the classical realized variance estimator (Andersen et al., 2003, and Barndorff-Nielsen and Shephard, 2002). For alternative choices of the same parameters, as well as for appropriate choices of the number of subsamples/autocovariances, it is also satisfied by the two-scale estimator of Zhang et al. (2005) as well as by the family of flat-top symmetric kernels suggested by Barndorff-Nielsen et al. (2009). Both are, contrary to realized variance, robust to market microstructure noise, thereby leading to spot variance estimates which are also robust to noise. Appendix A in BR (2008) details the values of α, β, a, b, and η for a variety of spot variance estimators $\widehat{\sigma}^2_{iT/n} = \frac{\widehat{V}_{iT/n}}{\phi_{n,T}}$ with different robustness properties with respect to jumps in returns and market microstructure noise.[13]

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

Theorem 6. *Let $\widehat{\sigma}^2_{iT/n} = \frac{\widehat{V}_{iT/n}}{\phi_{n,T}}$ be a spot variance estimator for which Eq. (18) is satisfied. Let, also, $\phi_{n,T} \to 0$ and $k \to \infty$. If*

$$\frac{Tv(T)^{-1}\log(n)}{\Delta_{n,T}h_{n,T}k^{\alpha}\phi^{\beta}_{n,T}} + \frac{Tv(T)^{-1}}{\Delta_{n,T}h_{n,T}}\left(\phi_{n,T}\log\left(\frac{1}{\phi_{n,T}}\right)\right)^{1/2} \to 0$$

the consistency results in Theorems 1-5 hold when replacing $\sigma^2_{iT/n}$ with $\widehat{\sigma}^2_{iT/n}$. If

$$\frac{Tv(T)^{-1/2}\log(n)}{\Delta^{3/2}_{n,T}h^{1/2}_{n,T}k^{\alpha}\phi^{\beta}_{n,T}} + \frac{Tv(T)^{-1/2}}{\Delta^{3/2}_{n,T}h^{1/2}_{n,T}}\left(\phi_{n,T}\log\left(\frac{1}{\phi_{n,T}}\right)\right)^{1/2} \to 0$$

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

$$\frac{Tv(T)^{-1/2}\log(n)}{\Delta_{n,T}h^{1/2}_{n,T}k^{\alpha}\phi^{\beta}_{n,T}} + \frac{Tv(T)^{-1/2}}{\Delta_{n,T}h^{1/2}_{n,T}}\left(\phi_{n,T}\log\left(\frac{1}{\phi_{n,T}}\right)\right)^{1/2} \to 0$$

[12] Of course, the interval could also be $[i\Delta_{n,T} + \phi_{n,T}, i\Delta_{n,T}]$. Alternatively, it could be symmetric around $i\Delta_{n,T}$. In all cases, in fact, $\phi_{n,T} \to 0$. The interpretation of these alternative choices, and their relation with discrete-time approaches, are discussed in Subsection 13.2.

[13] For interesting, alternative approaches to spot variance estimation using realized variance and Fourier estimates, we refer the reader to the recent work of Malliavin and Mancino (2008) and Kristensen (2008).

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

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

11 Leverage estimates: the S&P 500 index

We apply the leverage estimator to S&P500 index futures. We employ high-frequency stock index future prices from January 1982 to February 2009 for a total of 6,675 days. As is customary in the literature, we focus on a broad-based US market index. Admittedly, however, classical economic logic behind leverage effects makes the analysis of firm-specific data compelling. We leave this analysis for future work.

Our estimated model is for $\xi(\cdot) = \sqrt{\cdot}$, the volatility process. We also express the quantities of interest as a function of spot volatility σ. To reduce the estimation error induced by daily spot volatility estimation, we use intra-daily observations interpolated on a 5-minute grid (80 intervals per day) to construct the volatility estimates ($k = 80$). These estimates are derived as a combination of the threshold methods advocated by Mancini (2007) and of bipower variation (Barndorff-Nielsen and Shephard, 2004), as proposed by Corsi et al. (2008), i.e., threshold bipower variation or $TBPV$.[14] We refer the interested reader to Corsi et al. (2008) for details on the estimator's construction. Importantly, Theorem 2.3 in Corsi et al. (2008) implies that $TBPV$ satisfies the result in Eq. (18) with $\alpha = 1/2$, $\beta = 0$, $a \simeq 2.6$, $b = 0$, and $\eta = 1$.[15] Thus, the spot volatility estimator is

$$\widehat{\sigma}_{iT/n} = \sqrt{\frac{TBPV_{iT/n:iT/n-\phi_{n,T}}}{\phi_{n,T}}}.$$

We consider our more general case of jumps both in returns and in volatility. The nonparametric estimates $\widehat{\vartheta}_{1,1}(\sigma)$ and $\widehat{\vartheta}_j(\sigma)$ with $j = 1, 2, 3, 4$ are implemented using $h_{n,T} = h_s \hat{s} n^{-\frac{1}{5}}$, where $\hat{s}$ is the standard deviation of the time-series of daily spot volatilities. Based on preliminary investigation, 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

$$\begin{aligned}
\vartheta_2(\sigma) &\approx \Lambda^2(\sigma) + 2\mu_\sigma^2 \lambda_\sigma(\sigma), \\
\vartheta_3(\sigma) &\approx 6\mu_\sigma^3 \lambda_\sigma(\sigma) + 3\vartheta_1(\sigma)\vartheta_2(\sigma)\Delta, \\
\vartheta_4(\sigma) &\approx 24\mu_\sigma^4 \lambda_\sigma(\sigma) + \left[3\left(\vartheta_2(\sigma)\right)^2 + 4\vartheta_1(\sigma)\vartheta_3(\sigma)\right]\Delta.
\end{aligned}$$

[14] The use of power variation (or solely threshold methods) would not change our results qualitatively. We opt for using $TBPV$ in light of its "double-blade" nature and superior robustness to jumps in finite samples (see Corsi et al., 2008, for further discussions).

[15] Contrary to threshold bipower variation, for the classical bipower variation estimator of Barndorff-Nielsen and Shephard (2004), Eq. (18) is not satisfied. However, the conditions in Theorem 6 are still valid with $\alpha = \beta = 1/2$.

We identify the system through

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

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

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

and, of course,

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

with $\widehat{\vartheta}_{1,1}(\cdot)$ as defined in Eq. (6).[16] We apply a similar first-order correction to evaluate the confidence bands. These are obtained by using the limiting results in Section 7 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 $\widehat{\sigma}_{iT/n}$ $\forall i = 1, ..., n$.

The empirical findings are presented in Fig. 1 and in Fig. 2. Fig. 1 contains the S&P500 future volatility 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 very mild hump-shape for low volatilities, decreasing with the volatility level. These estimates vary between roughly -0.17 and -0.35 in the most populated volatility range, the value -0.6 being reached for high, seldomly seen, volatility levels. These results are consistent with economic logic, as laid out in the Introduction, and with the parametric evidence in Section 2. They are, once more, indicative of significant time-variation in the correlation between shocks to returns and shocks to volatility. For a dynamic assessment of this time variation, see Fig. 2.[17]

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 independent and the mean of the return jumps is equal to zero, as sometimes assumed in the literature - see the Corollary to Theorem 5). 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 9, the methods simply lead to the estimation of a broader notion of leverage. It is now of interest to separately identify (and evaluate the relative impact of) the different components of total leverage, dubbed "continuous leverage" and "co-jump leverage" earlier. Ideas on identification are laid out in

[16] Even though previous work on nonparametric jump-diffusion estimation has shown that jump identification off 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 bound to improve finite-sample inference by reducing discretization error. The performance of the resulting leverage estimator is evaluated below by simulation.

[17] The point estimate of $\widetilde{\mu}_\sigma$ is 0.3246.

Remark 10. This study is, however, beyond the scope of the current paper and better left for future work.

12 Simulations

Consider the discretized system

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

where $\{J_t^r, J_t^\sigma\}$ are independent Bernoulli random variables with constant intensities $\lambda^r\Delta$ and $\lambda^\sigma\Delta$, $\{\varepsilon_t^r, \varepsilon_t^\sigma\}$ are standard Gaussian random variables with correlation ρ, ξ_t^r is a mean zero Gaussian shock with standard deviation σ_ξ, ξ_t^σ is an exponential shock with mean μ_σ, and Δ is an interval corresponding to a day. Assume $\mu = 0.0506$, $\kappa = 0.025$, $\theta = 0.5585$, $\sigma_\upsilon = 0.09$, $\lambda^r = 0.0046$, $\lambda^\sigma = 0.0055$, $\sigma_\xi = 2.98$, $\mu_\sigma = 1.79$, and $\rho = -0.4$. These parameter values are virtually identical to those estimated by Eraker et al. (2003) and reported in their Table III for similar data. In agreement with our average leverage estimates, we lowered the ρ value to -0.4 (but leaving it unchanged to the Eraker et al's value of -0.5 would not affect our findings in any way). In light of the insignificance of the corresponding parameter in Eraker et al. (2003), we also do not allow for negative mean jumps in returns.

The system is simulated over 2,500 days. Fig. 4 reports the median leverage estimates (for each spot variance level) as well as 10^{th} and 90^{th} percentile bands across 1,000 simulations. For each simulated path, we estimate leverage using Eq. (11) in conjunction with the small-sample adjustments discussed in the previous section. Barring the presence of some upward bias for low spot variance levels, leverage is estimated rather accurately. Importantly for our purposes, we have no evidence that the reported nonlinear, decreasing behavior for increasing variance levels is a by-product of the use of nonparametric methods and their well-known boundary effects.

13 Further discussions

13.1 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. The consistency derives from the fact that the implied leverage function ρ_i captures the contemporaneous correlation between daily volatility 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 $\sqrt{\frac{TBPV_{iT/n:iT/n-\phi_{n,T}}}{\phi_{n,T}}} - \sqrt{\frac{TBPV_{(i-1)T/n:(i-1)T/n-\phi_{n,T}}}{\phi_{n,T}}}$ and $r_{iT/n} = \log p_{iT/n} - \log p_{(i-1)T/n}$ for a daily $\phi_{n,T}$.

While, in order to provide motivation, our treatment in Section 2 had to be consistent 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

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

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 $\widehat{\delta}_i \approx \frac{\widehat{cov}(\sqrt{RV_{t:t-1}}-HAR_{t-1}, r_{t-1}|\eta_i \le \sqrt{RV_{t-1:t-2}} \le \eta_{i-1})}{\widehat{var}(r_t|\eta_i \le \sqrt{RV_{t-1:t-2}} \le \eta_{i-1})}$, $\widehat{\rho}_i \approx \widehat{\delta}_i\widehat{S}_i$, and $\widehat{S}_i = \frac{\widehat{std}(r_{t-1}|\eta_i \le \sqrt{RV_{t-1:t-2}} \le \eta_{i-1})}{\widehat{std}\left(\sqrt{RV_{t:t-1}}-HAR_{t-1}|\eta_i \le \sqrt{RV_{t-1:t-2}} \le \eta_{i-1}\right)}$. This lagged definition of leverage may even appear more consistent with classical economic logic postulating that *past* changes in firm value (and in the value of the firm's stock) lead to *future* volatility changes of an opposite sign due to variations in the company's debt-to-equity ratio. For a daily $\phi_{n,T}$, the corresponding expression in our functional framework would of course be $\widetilde{\widehat{\rho}}^{lagged}(\sigma) = \frac{\widehat{\vartheta}_{1,1}^{lagged}(\sigma)}{\sqrt{\sigma\widetilde{\Lambda}^2(\sigma)}}$ with

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

Fig. 3 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 0.4 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. Interestingly, 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 $\widehat{\vartheta}_{1,1}^{lagged}(\sigma)$, $\sqrt{\frac{TBPV_{(i+1)T/n+\phi_{n,T}:(i+1)T/n}}{\phi_{n,T}}}$ can be viewed as an estimate of $\sigma_{(i+1)T/n}$ (for a shrinking $\phi_{n,T}$ "from the right"). Similarly, in $\widehat{\vartheta}_{1,1}(\sigma)$, $\sqrt{\frac{TBPV_{(i+1)T/n:(i+1)T/n-\phi_{n,T}}}{\phi_{n,T}}}$ can also be seen as an estimate of $\sigma_{(i+1)T/n}$ (for a shrinking $\phi_{n,T}$ "from the left"). 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 $\log p_{(i+1)T/n} - \log p_{iT/n}$.

Differently put, 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 $TBPV$) requires the choice of a kernel and a window width. In agreement with our previous discussion, the use of a left-kernel is more naturally comparable to a discrete-time model in which leverage is contemporaneous whereas the use of a right-kernel is more naturally comparable to a discrete-time model in which leverage is lagged. Needless to say, the kernel could be symmetric about iT/n but, as implied by our previous discussion, the use of symmetric kernels would complicate intuition as compared to discrete-time specifications.

In sum, we find that the effective use of a flat left-kernel (leading to a notion of contemporaneous leverage in the language of discrete-time models) yields a a more statistically significant and stronger negative correlation between shocks to volatility and shocks to returns than the use of a flat right-kernel (leading to a lagged notion of leverage, again, in the language of discrete-time models). In both cases, leverage becomes more negative with increasing volatility levels around the bulk of the data.

13.2 Alternative conditioning variables

Conditional leverage is modelled in this paper as a function of spot volatility. We argue that (*1*) this approach is natural, (*2*) it leads to an economically meaningful interpretation coherent with classical finance principles, and (*3*) it is technically more feasible, in continuous-time, than alternative conditioning methods.

First, this modelling approach is natural in continuous-time stochastic volatility models - and, effectively, extends them - since spot volatility is used as a conditioning variable both in the return equation (where the return drift may depend on volatility as implied by the presence of risk-return trade-offs) and in the volatility equation (where the volatility drift and diffusion are generally modelled as functions of the volatility state).

Second, our findings may be easily reconciled with classical economic logic as discussed in the Introduction. In our case, the *contemporaneous* (or *lagged,* depending on the employed kernel - see previous section) negative correlation between shocks to returns and shocks to volatility tends to be higher, the higher the volatility level. High volatility levels may have been induced by (a series of) negative past shocks to returns leading to higher debt-to-equity ratios and, consequently, stronger volatility changes for associated price changes. Conversely, leverage is lower (in absolute value) and less negative for lower volatility levels associated with smaller past shocks to returns (and relatively lower debt-to-equity ratios). In other words, volatility may serve as a proxy for the level of the debt-to-equity ratio (and, hence, the state of the firm). A higher (lower) debt-to-equity ratio associated with higher (lower) volatility levels translates into a higher (lower) correlation between shocks to returns and shocks to volatility: shocks to returns are more negatively correlated with shocks to volatility when risk is higher.

Finally, volatility is a "stock" variable. Alternative conditioning variables used in this literature (stock returns, for example - see Figlewski and Wang, 2000, and Yu, 2008) have a "flow" nature. This property makes conditioning on volatility 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).

This said, alternative approaches may be entertained, at least in discrete time. The parametric findings in Figlewski and Wang (2000) and Yu (2008) are notable in this area. When the stock value decreases, the debt-to-equity ratio increases, so does *future* volatility. Positive price changes yield reductions in the debt-to-equity ratio but *increase* future volatility, thereby leading to an asymmetry in the way in which volatility responds to price changes - conditionally on a small or positive return, the correlation between volatility 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 volatility).[18] 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 volatility 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 Table 1 and 2.

14 Conclusions

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

15 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

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

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

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

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

$$\frac{v(T)}{h_{n,T}}\left(\Delta_{n,T} \log \frac{1}{\Delta_{n,T}}\right)^{1/2} \to 0,$$

we have

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

[18] 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).

and

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

Moreover:

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

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

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 $\sigma^2_{s-} - \sigma^2_{iT/n}$.

Lemma A.2. *Consider the quantity*

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

If

$$\frac{v(T)}{h_{n,T}}\left(\Delta_{n,T}\log\frac{1}{\Delta_{n,T}}\right)^{1/2} \to 0$$

we have

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

and

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

Moreover:

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

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

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 $\sigma^2_{s-} - \sigma^2_{iT/n}$.

Proof of Lemma A.1. We prove the lemma for $\Psi(\sigma^2)_1$. The case $\Psi(\sigma^2)_0$ follows analogously. Let T be fixed and define:

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

where the $u_{iT/n,(i+1)T/n}s$ are square-integrable martingale difference sequences. We immediately have

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

and, by virtue of Ito's Lemma on $\left(\log p_{s-} - \log p_{iT/n}\right)^2$:

$$\begin{aligned}&\sum_{i=1}^{n-1}\mathbf{E}[u^2_{iT/n,(i+1)T/n}|\Im_{iT/n}] \\ &\overset{p}{\underset{\Delta_{n,T}\to 0}{\to}} \frac{1}{2}\frac{1}{h_{n,T}}\int_0^T \mathbf{K}^2\left(\frac{\sigma^2_{s-} - \sigma^2}{h_{n,T}}\right)\left(\sigma^2_{s-} + \lambda_r(\sigma^2_{s-})\mathbf{E}[c_r^2]\right)\lambda_\sigma(\sigma^2_{s-})\mathbf{E}[g^2]ds \\ =\ & \widetilde{\mathbf{V}}_T.\end{aligned}$$

Now write

$$\begin{aligned}
& \sum_{i=1}^{n-1} \mathbf{E}\left[u_{iT/n,(i+1)T/n}^2 \mathbf{1}_{(|u_{iT/n,(i+1)T/n}|>\epsilon)} | \Im_{iT/n}\right] \\
= & \sum_{i=1}^{n-1} \mathbf{E}\left[u_{iT/n,(i+1)T/n}^2 | \Im_{iT/n}\right] - \sum_{i=1}^{n-1} \mathbf{E}\left[u_{iT/n,(i+1)T/n}^2 \mathbf{1}_{(|u_{iT/n,(i+1)T/n}|\le\epsilon)} | \Im_{iT/n}\right] \\
= & \widetilde{\mathbf{V}}_T - \sum_{i=1}^{n-1} \mathbf{E}\left[u_{iT/n,(i+1)T/n}^2 \mathbf{1}_{\left(\left|\mathbf{K}\left(\frac{\sigma_{iT/n}^2-\sigma^2}{h_{n,T}}\right)\int_{iT/n}^{(i+1)T/n}\left(\log p_{s-}-\log p_{iT/n}\right)\int_Z g(z)\bar{\nu}_\sigma(ds,dz)\right|\le\epsilon\sqrt{h_{n,T}\Delta_{n,T}}\right)} | \Im_{iT/n}\right],
\end{aligned} \tag{21}$$

but $\mathbf{K}\left(\frac{\sigma_{iT/n}^2-\sigma^2}{h_{n,T}}\right)\int_{iT/n}^{(i+1)T/n}\left(\log p_{s-}-\log p_{iT/n}\right)\int_Z g(z)\bar{\nu}_\sigma(ds,dz) = O_p(\Delta_{n,T})$. Hence, the indicator converges in probability to 1 and, given boundedness of $\widetilde{\mathbf{V}}_T$, Eq. (21) converges in probability to 0 (as $\Delta_{n,T}\to 0$). This is a conditional Lindeberg condition. Using Theorem VIII.3.33 in Jacod and Shiryaev (2002), we conclude that, for each T, $\sqrt{\frac{h_{n,T}}{\Delta_{n,T}}}\Psi^{num} \Rightarrow W\left(\widetilde{\mathbf{V}}_T\right)$ and W is an independent Brownian motion. This implies that

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

uniformly in T. By the Quotient limit theorem (see, e.g., Revuz and Yor, 1998, Theorem 3.12) we now have that, as $T\to\infty$ (with $h_{n,T}\to 0$),

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

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 that of Lemma A.1.

Proof of Theorem 1. Under the assumptions of the theorem, Bandi and Phillips (2003) prove that $\widehat{\vartheta}_2(\sigma^2)\to\vartheta_2(\sigma^2)=\Lambda^2(\sigma^2)$ with probability one and

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

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

$$\begin{aligned}
& \widehat{\vartheta}_{1,1} \\
= & \widetilde{\mathbf{K}}\left(\int_{iT/n}^{(i+1)T/n}\left(\sigma_{s-}^2-\sigma_{iT/n}^2\right)\mu_{s-}ds\right) + \widetilde{\mathbf{K}}\left(\int_{iT/n}^{(i+1)T/n}\left(\sigma_{s-}^2-\sigma_{iT/n}^2\right)\sigma_{s-}dW_s^r\right) \\
& +\widetilde{\mathbf{K}}\left(\int_{iT/n}^{(i+1)T/n}\left(\log(p_{s-})-\log\left(p_{iT/n}\right)\right)m_{s-}ds\right) + \widetilde{\mathbf{K}}\left(\int_{iT/n}^{(i+1)T/n}\left(\log(p_{s-})-\log\left(p_{iT/n}\right)\right)\Lambda(\sigma_{s-}^2)dW_s^\sigma\right) \\
& +\widetilde{\mathbf{K}}\left(\int_{iT/n}^{(i+1)T/n}\rho(\sigma_{s-}^2)\sigma_{s-}\Lambda(\sigma_{s-}^2)ds\right) \\
: & = \widehat{\vartheta}_{1,1}^c.
\end{aligned} \tag{23}$$

Using Lemma A.2. we get:

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

and

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

The asymptotic covariance between

$$\widetilde{\mathbf{K}}\left(\int_{iT/n}^{(i+1)T/n}\left(\sigma_{s-}^2-\sigma_{iT/n}^2\right)\sigma_{s-}dW_s^r\right)$$

and

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

is equal to

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

Hence,

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

since, by the Quotient limit theorem,

$$\widetilde{\mathbf{K}}\left(\int_{iT/n}^{(i+1)T/n}\rho(\sigma_{s-}^2)\sigma_{s-}\Lambda(\sigma_{s-}^2)ds\right)-\vartheta_{1,1}(\sigma^2)=\Gamma_{\vartheta_{1,1}}(\sigma^2)+o_p(h_{n,T}^2).$$

In the same way,

$$\begin{aligned}&\widehat{\vartheta}_2(\sigma^2)\\=\;&\widetilde{\mathbf{K}}\left(2\int_{iT/n}^{(i+1)T/n}\left(\sigma_{s-}^2-\sigma_{iT/n}^2\right)m_{s-}ds\right)+\widetilde{\mathbf{K}}\left(2\int_{iT/n}^{(i+1)T/n}\left(\sigma_{s-}^2-\sigma_{iT/n}^2\right)\Lambda(\sigma_{s-}^2)dW_s^{\sigma}\right)\\&+\widetilde{\mathbf{K}}\left(\int_{iT/n}^{(i+1)T/n}\Lambda^2(\sigma^2)ds\right).\end{aligned}$$

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

$$Asycov(\widehat{\vartheta}_{1,1},\widehat{\vartheta}_2)=\mathbf{K}_2\frac{\Delta_{n,T}}{h_{n,T}\widehat{\overline{L}}_{\sigma^2}(T,\sigma^2)}\left(2\Lambda^3(\sigma^2)\sigma\rho(\sigma^2)\right).$$

Finally, by the delta method:

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

Hence,

$$\begin{aligned}&Asyvar\left(\sqrt{\frac{h_{n,T}\widehat{\overline{L}}_{\sigma^2}(T,\sigma^2)}{\Delta_{n,T}}}\widehat{\rho}(\sigma^2)\right)\\=\;&\left[\frac{1}{\sigma^2\Lambda^2(\sigma)}Asyvar\left(\widehat{\vartheta}_{1,1}(\sigma^2)\right)-\frac{\vartheta_{1,1}}{\sigma^2\Lambda^4(\sigma^2)}Asycov\left(\widehat{\vartheta}_{1,1}(\sigma^2),\widehat{\vartheta}_2(\sigma^2)\right)+\frac{\vartheta_{1,1}^2}{4\sigma^2\Lambda^6(\sigma^2)}Asyvar\left(\widehat{\vartheta}_2(\sigma^2)\right)\right]\\=\;&\frac{1}{\sigma^2\Lambda^2(\sigma)}\left[Asyvar\left(\widehat{\vartheta}_{1,1}(\sigma^2)\right)-\frac{\sigma\rho(\sigma^2)}{\Lambda(\sigma^2)}Asycov\left(\widehat{\vartheta}_{1,1}(\sigma^2),\widehat{\vartheta}_2(\sigma^2)\right)+\left(\frac{\sigma\rho(\sigma^2)}{2\Lambda(\sigma^2)}\right)^2Asyvar\left(\widehat{\vartheta}_2(\sigma^2)\right)\right]\\=\;&\frac{1}{\sigma^2\Lambda^2(\sigma)}\left[\Lambda^2(\sigma^2)\sigma^2(1+\rho^2(\sigma^2))-\frac{\sigma\rho(\sigma^2)}{\Lambda(\sigma^2)}(2\Lambda^3(\sigma^2)\sigma\rho(\sigma^2))+\left(\frac{\sigma\rho(\sigma^2)}{2\Lambda(\sigma^2)}\right)^2 2\Lambda^4(\sigma^2)\right]\\=\;&\mathbf{K}_2\left(1-\frac{1}{2}\rho^2(\sigma^2)\right).\end{aligned}$$

As for the asymptotic bias, clearly

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

■

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

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

where $\widehat{\vartheta}^c_{1,1}$ is defined in Eq. (23). 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.,

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

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

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

■

Proof of Theorem 3. BR (2008) show that $\widehat{\Lambda}^2(\sigma^2) \xrightarrow{p} \Lambda^2(\sigma^2)$ and

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

Ito'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 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. ■

Proof of Theorem 4. See BR (2008), Theorem 7. ■

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,

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

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

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

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

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

we have

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

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

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

thereby leading to the stated result.■

Proof of Theorem 6. Denote by $\widetilde{\vartheta}_{1,1}(\sigma^2)$ and by $\widetilde{\vartheta}_j(\sigma^2)$ with $j=2,3,4$, the moment estimators constructed using estimated spot variance in place of the true, unknown spot variance. Using the method of proof of Theorem 2 of BR (2008), we can show that

$$\widetilde{\vartheta}_{1,1}(\sigma^2)=\widehat{\vartheta}_{1,1}(\sigma^2)+O_p\left(\frac{Tv(T)^{-1}\log(n)}{\Delta_{n,T}h_{n,T}k^\alpha\phi_{n,T}^\beta}+\frac{Tv(T)^{-1}}{\Delta_{n,T}h_{n,T}}\left(\phi_{n,T}\log\left(\frac{1}{\phi_{n,T}}\right)\right)^{1/2}\right)$$

and

$$\widetilde{\vartheta}_j(\sigma^2)=\widehat{\vartheta}_j(\sigma^2)+O_p\left(\frac{Tv(T)^{-1}\log(n)}{\Delta_{n,T}h_{n,T}k^\alpha\phi_{n,T}^\beta}+\frac{Tv(T)^{-1}}{\Delta_{n,T}h_{n,T}}\left(\phi_{n,T}\log\left(\frac{1}{\phi_{n,T}}\right)\right)^{1/2}\right)$$

with $j=2,3,4$. Therefore,

$$\frac{Tv(T)^{-1}\log(n)}{\Delta_{n,T}h_{n,T}k^\alpha\phi_{n,T}^\beta}+\frac{Tv(T)^{-1}}{\Delta_{n,T}h_{n,T}}\left(\phi_{n,T}\log\left(\frac{1}{\phi_{n,T}}\right)\right)^{1/2}\to 0$$

is sufficient for consistency of the feasible leverage estimator. For weak convergence, this condition ought to be strenghtened. The relevant condition is

$$\frac{Tv(T)^{-1/2}\log(n)}{\Delta_{n,T}^{3/2}h_{n,T}^{1/2}k^\alpha\phi_{n,T}^\beta}+\frac{Tv(T)^{-1/2}}{\Delta_{n,T}^{3/2}h_{n,T}^{1/2}}\left(\phi_{n,T}\log\left(\frac{1}{\phi_{n,T}}\right)\right)^{1/2}\to 0,$$

for Theorem 1 and 2 (i.e., when the convergence rate is $\sqrt{\frac{h_{n,T}v(T)}{\Delta_{n,T}}}$). It is

$$\frac{Tv(T)^{-1/2}\log(n)}{\Delta_{n,T}h_{n,T}^{1/2}k^\alpha\phi_{n,T}^\beta}+\frac{Tv(T)^{-1/2}}{\Delta_{n,T}h_{n,T}^{1/2}}\left(\phi_{n,T}\log\left(\frac{1}{\phi_{n,T}}\right)\right)^{1/2}\to 0,$$

for Theorems 3, 4, and 5 (i.e., when the convergence rate is $\sqrt{h_{n,T}v(T)}$). ■

References

[1] Andersen, T.G., L. Benzoni, and J. Lund (2002). Towards an empirical foundation for continuous-time equity return models. *Journal of Finance* 57, 1047-1091.

[2] Andersen, T.G., T. Bollerslev, F.X. Diebold, and P. Labys (2003). Modeling and forecasting realized volatility. *Econometrica* 71, 579-625.

[3] Bakshi, G., C. Cao, and Z. Chen (1997). Empirical performance of alternative option pricing models. *Journal of Finance* 52, 2003-2049.

[4] Bandi, F.M. and T. Nguyen (2003). On the functional estimation of jump-diffusion models. *Journal of Econometrics* 116, 293-328.

[5] Bandi, F.M. and P.C.B. Phillips (2003). Fully nonparametric estimation of scalar diffusion models. *Econometrica* 71, 241-283.

[6] Bandi, F.M. and P.C.B. Phillips (2009). Nonstationary continuous-time models. In *Handbook of Financial Econometrics* (L.P. Hansen and Y. Ait-Sahalia, Eds.) Elsevier Science, New York, forthcoming.

[7] Bandi, F. M. and R. Renò (2008). Nonparametric Stochastic Volatility. Available at SSRN: http://ssrn.com/abstract=1158438.

[8] Bandi, F. M. and R. Renò (2009). Infinitesimal cross-moments and return/volatility co-jumps. *Working paper.*

[9] Barndorff-Nielsen, O.E., P. Hansen, A. Lunde and N. Shephard (2009). Designing realized kernels to measure ex-post variation of equity prices in the presence of noise. *Econometrica*, forthcoming.

[10] Barndorff-Nielsen, O.E., and N. Shephard (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society, Series B 64*, 253-280.

[11] Barndorff-Nielsen, O.E., and N. Shephard (2004). Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics* 2, 1-37.

[12] Black, F. (1976). Studies of stock market volatility changes. Proceedings of the American Statistical Association, Business and Economic Statistics section, 177-181.

[13] Christie, A. (1982). The stochastic behaviour of common stock variances: value, leverage, and interest rate effects. *Journal of Financial Economics* 3, 407-431.

[14] Cont, R. and P. Tankov (2004). *Financial Modelling with Jump Processes.* Chapman & Hall.

[15] Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics* 7, 174-196.

[16] Corsi, F., D. Pirino, and R. Renò (2008). Threshold bipower variation and the impact of jumps on volatility forecasting. *Working paper.*

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

[18] Engle, R. and V. Ng (1993). Measuring and testing the impact of news in volatility. *Journal of Finance* 43, 1749-1778.

[19] Eraker, B., M. Johannes, and N. Polson (2003). The impact of jumps in volatility and returns. *Journal of Finance* 12, 69-1300.

[20] Figlewski, S., and X. Wang (2000). Is the "leverage effect" a leverage effect? *Working paper.*

[21] Glosten, L.R., R. Jagannathan, and D. Runkle (1993). Relationship between the expected value and the volatility of the nominal excess return on stocks. Journal of Finance 48, 1779-1802.

[22] Gobbi, F. and C. Mancini (2008). Identifying the covariation between the diffusion parts and the co-jumps given discrete observations. *Working paper.*

[23] Harvey, A.C. and N. Shephard (1996). The estimation of an asymmetric stochastic volatility model for asset returns. *Journal of Business and Economic Statistics* 14, 429-434.

[24] Jacod, J. and A.N. Shiryaev (2002). Limit theorems for stochastic processes. Springer.

[25] Jacquier, E., N. Polson, and P. Rossi (1994). Bayesian analysis of stochastic volatility models. *Journal of Business and Economic Statistics* 12, 371-389.

[26] Jacquier, E., N. Polson, and P. Rossi (2004). Bayesian analysis of stochastic volatility models with fat tails and correlated errors. *Journal of Econometrics* 122, 185-212.

[27] Johannes, M. (2004). The statistical and economic role of jumps in continuous-time interest rate models. *Journal of Finance,* 227-260.

[28] Kristensen, D. (2008). Nonparametric filtering of the realized spot volatility: a kernel-based approach. *Econometric Theory,* forthcoming.

[29] Malliavin, P. and M.E. Mancino (2008). A Fourier transform method for nonparametric estimation of volatility. *Annals of Statistics,* forthcoming.

[30] Mancini, C. (2007). Nonparametric threshold estimation for models with stochastic diffusion coefficients and jumps. *Scandinavian Journal of Statistics*, forthcoming.

[31] Revuz, D. and M. Yor (1998). *Continuous Martingales and Brownian Motion.* Springer-Verlag.

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

[33] Yu, J. (2005). On leverage in a stochastic volatility model. *Journal of Econometrics* 127, 165-178.

[34] Yu, J. (2008). A semiparametric stochastic volatility model. *Working paper.*

[35] Zhang, L., P. Mykland, and Y. Aït-Sahalia (2005). A tale of two time scales: determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association* 100, 1394-1411.

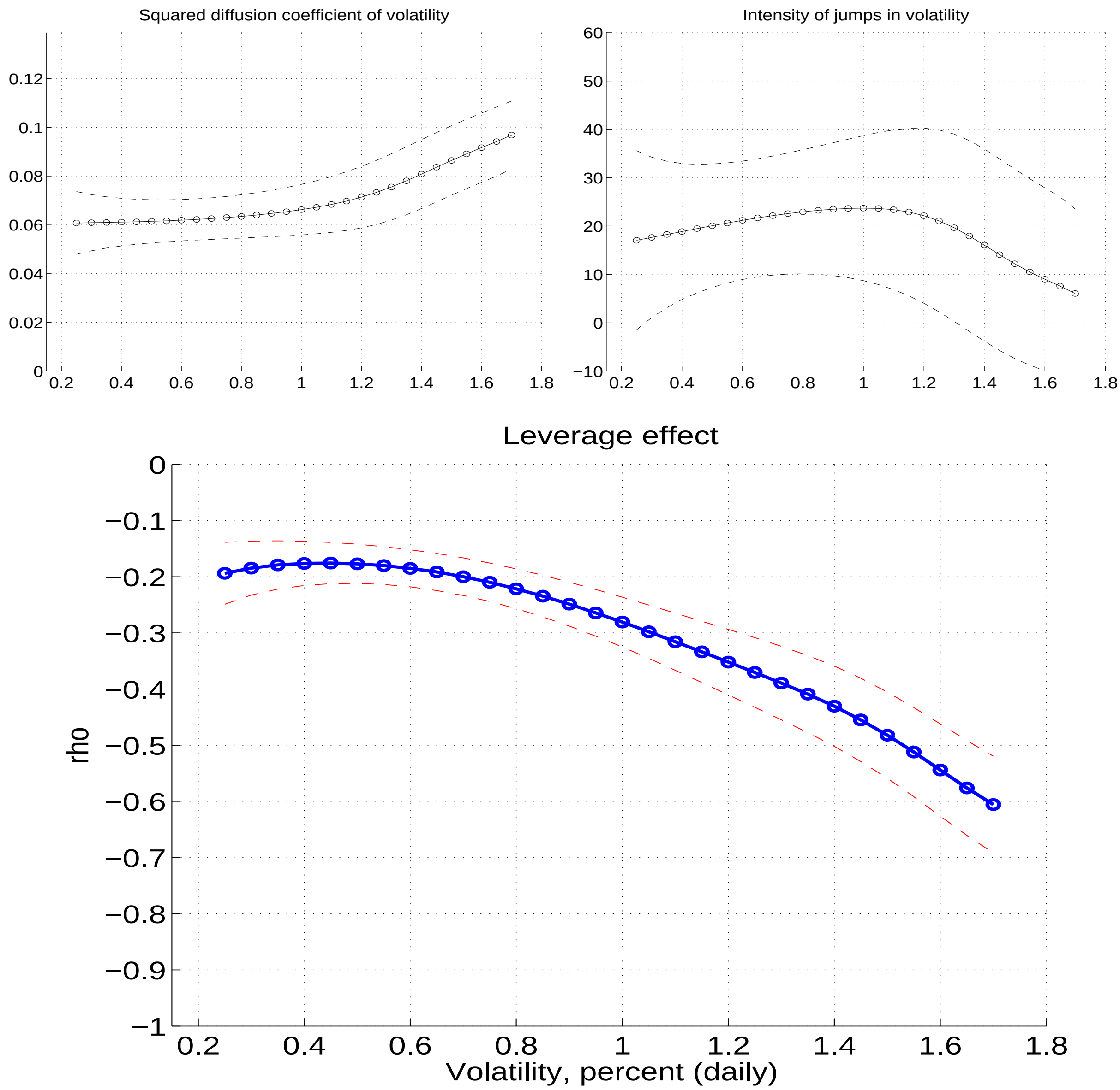

Figure 1: Functional estimates of $\Lambda^2(\sigma)$ (top left), $\lambda_\sigma(\sigma)$ (top right), and $\rho(\sigma)$ (bottom) for the spot volatility process of the S&P500 index futures. With the exception of the jump intensity which is in terms of the number of jumps per year, all estimates are daily and in percentage form.

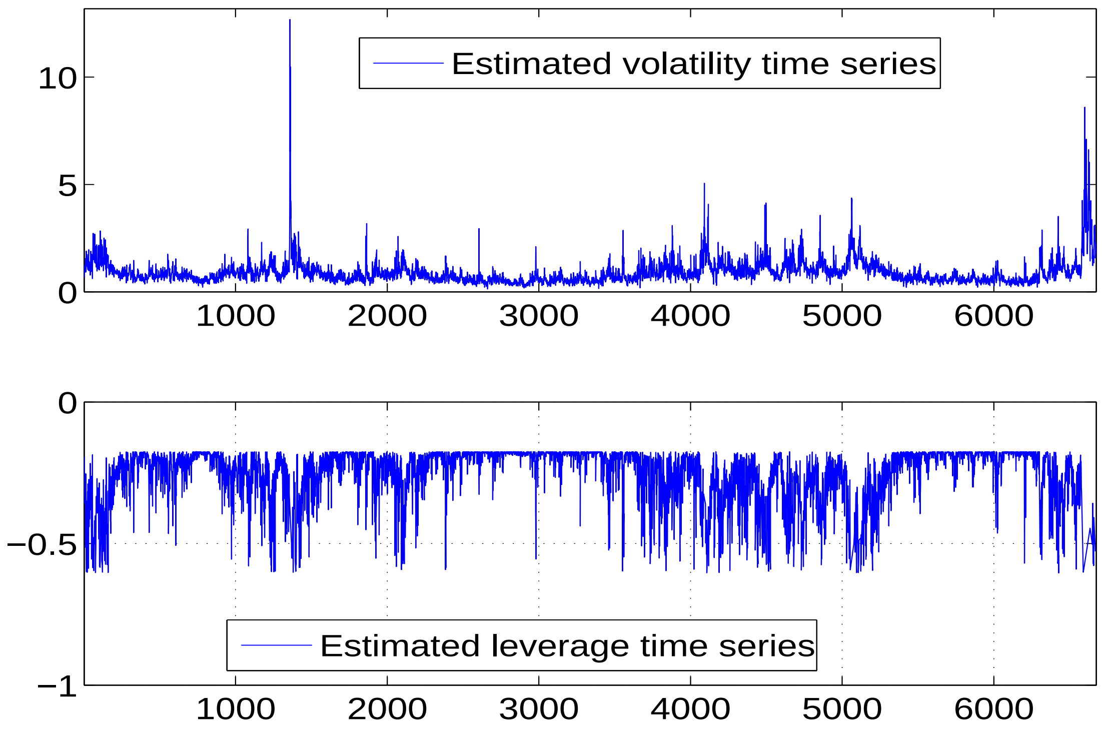

Figure 2: Estimated time series of spot volatility (top) and leverage (bottom) for the S&P500 index futures.

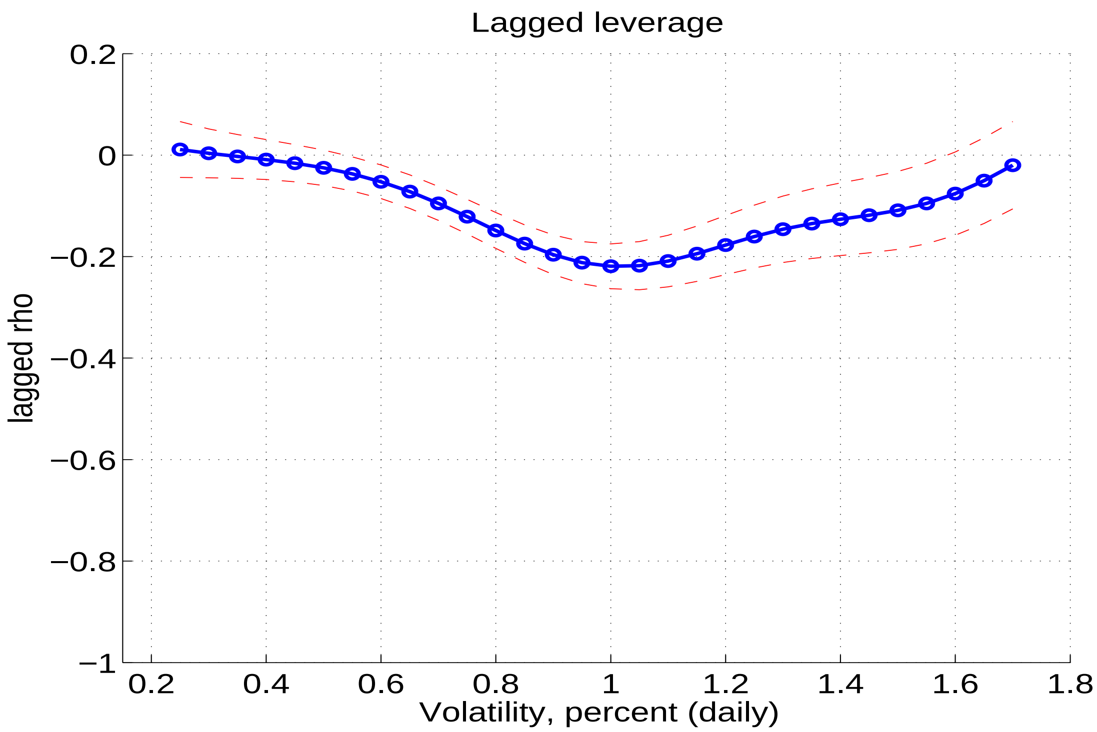

Figure 3: Functional estimates of *lagged* leverage for the S&P500 index futures.

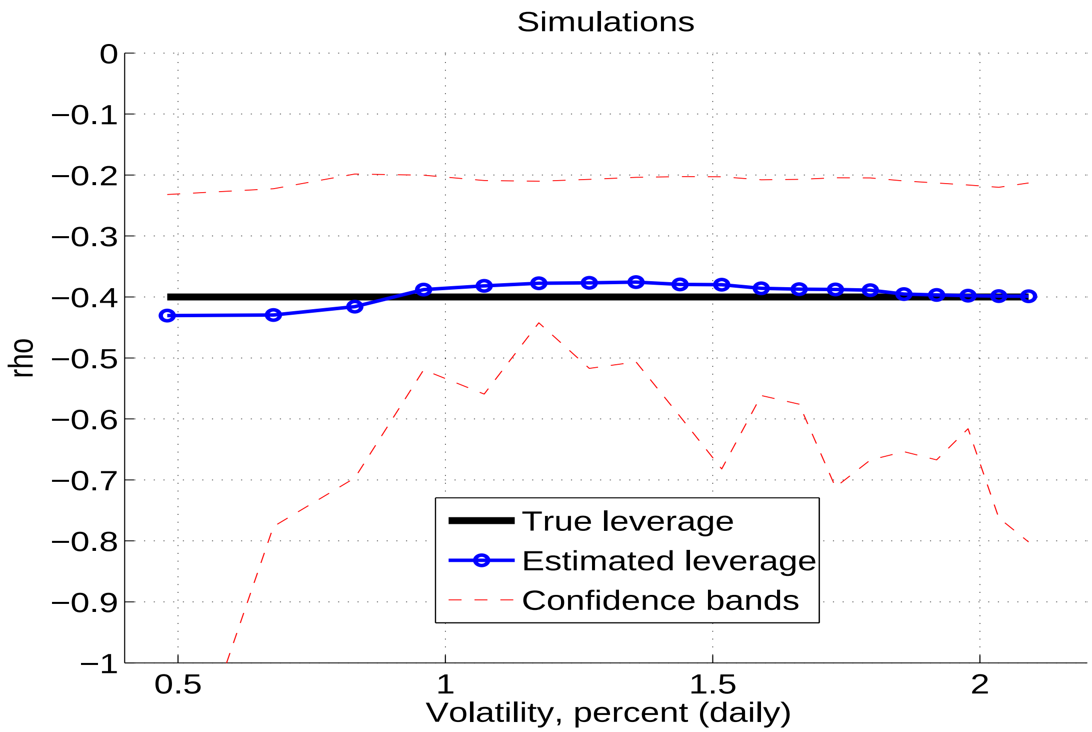

Figure 4: Estimated leverage on simulated paths. The table reports the 10th, 50th, and 90th percentile of the leverage estimates across 1,000 simulated paths.

Model:

$$\sqrt{RV_{t:t-1}} = HAR_{t-1} + r_t \sum_{i=1}^{N} \delta_i \mathbf{1}_{\{\eta_i \le X \le \eta_{i-1}\}} + \varepsilon_t,$$

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 ($X = r_{t-1}$)

i	Range	δ_i	standard error	t-stat	implied leverage
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 ($X = \sqrt{RV_{t-1:t-2}} \cdot sign(r_{t-1})$)

i	Range	δ_i	standard error	t-stat	implied leverage
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

Table 1: Parametric models with contemporaneous leverage

Model:

$$\sqrt{RV_{t:t-1}} = HAR_{t-1} + r_{t-1} \sum_{i=1}^{N} \delta_i \mathbf{1}_{\{\eta_i \leq X \leq \eta_{i-1}\}} + \varepsilon_t,$$

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 ($X = r_{t-1}$)

i	Range	δ_i	standard error	t-stat	implied leverage
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 ($X = \sqrt{RV_{t-1:t-2}} \cdot sign(r_{t-1})$)

i	Range	δ_i	standard error	t-stat	implied leverage
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

Table 2: Parametric models with lagged leverage