PRICE AND VOLATILITY CO-JUMPS*

Federico M. Bandi* and Roberto Renò**
*Johns Hopkins University and Edhec-Risk and **Università di Siena

First draft: April 14, 2011
This draft: November 16, 2011

Abstract

A sizeable proportion of large, discontinuous, changes in asset prices are found to be associated with contemporaneous large, discontinuous, changes in volatility, i.e., co-jumps. When occurring jointly, the price jumps tend to be negative while the volatility jumps tend to be positive. Unusually large, negative, price discontinuities are often accompanied by unusually large, positive, volatility changes, as implied by a strong anti-correlation between the jump sizes. Further, the distribution of the price co-jump sizes is found to depend on the underlying volatility level and become more dispersed, as well as more negatively centered, as volatility increases. Finally, we show that the co-jumps yield an economically-meaningful portion of leverage, return skewness, and the implied volatility smirk. These effects are uncovered in the context of a flexible modeling approach (allowing, among other features, for independent as well as common jumps, volatility-dependent jump arrivals, and time-varying leverage) and a novel identification strategy relying on *infinitesimal cross-moments* and high-frequency price data.

Keywords: Stochastic volatility, jumps in prices, jumps in volatility, co-jumps, implied volatility smirk, infinitesimal cross-moments.

*We thank seminar participants at Wisconsin School of Business, University of Wisconsin, Bauer College of Business, University of Houston, and University of Florence for useful discussions.

1 Introduction

Understanding the dynamic properties of stock returns continues to be at the heart of research in finance. Strong consensus has emerged about the need to allow for discontinuities in asset prices, "return jumps", while incorporating stochastic time-variation in the standard deviation of the continuous shocks to returns, "stochastic volatility". Even though both features generate fat tails in the return distribution, they do not appear to be sufficient to yield the rapid increases in volatility which have been historically experienced.

Adding jumps in volatility, however, provides a solution to this issue. As pointed out in Eraker et al. (2003), jumps in returns and jumps in volatility serve different, but complementary, purposes. The former generate large, infrequently-observed, sudden movements, such as the October 1987 crash. These movements would require an unreasonably high level of volatility to occur. The latter lead to fast changes in the level of volatility and, due to volatility persistence, a lasting effect on the distribution of stock returns. Along with daily moves as large as 22%, during the 1987 crash, among other times of distress, the level of volatility more than doubled before slowly mean-reverting to lower levels. The inability of continuous-time return models with stochastic volatility and jumps only in returns to fit return/option data, and the need to incorporate jumps in volatility, has been recognized as early as in the work of, e.g., Bakshi et al. (1997), Bates (2000), and Pan (2002). Duffie et al. (2000) and Eraker et al. (2003) provide a thorough analysis of the effectiveness of discontinuities in returns and volatility by explicitly incorporating both components in their "double-jump" specifications.

A related problem has to do with the frequency of the return/volatility jumps and, in particular, the likelihood of discontinuous movements occurring *jointly*. Classical implementations impose either strict independence of the jump arrivals (Eraker et al., 2003) or contemporaneous arrivals with correlated jump sizes (Duffie et al., 2000, Eraker et al., 2003, and Eraker, 2004). The fine-grain dynamics of the return and volatility process may, however, be such that some of the discontinuous components are fully idiosyncratic while some may be common to the return/volatility processes. Using intra-daily observations on the VIX and the S&P500 index along with sample statistics developed in Todorov and Tauchen (2010) and Jacod and Todorov (2009). Todorov and Tauchen (2011), for instance, find evidence of a mixed specification with a high likelihood of contemporaneous arrivals along with a lower likelihood of disjoint arrivals.

The analysis of a flexible stochastic volatility model with co-jumps, as well as independent jumps, is the subject of this paper. Volatility is, of course, a latent variable. We identify it nonparametrically for every hour in our sample using intra-daily price data. We then apply functional methods to the obtained *hourly* volatility series, as well as to *hourly* returns, to identify *daily* return/volatility dynamics. In essence, we evaluate the *daily* evolution of returns and volatility by making use of the substantial informational content of *intra-daily* return data and

volatility series which are pre-filtered on the basis on *intra-daily* return data. Our interest in daily dynamics is analogous to that of Duffie et al. (2000) and Eraker et al. (2003), among others, who study the daily evolution of return and volatility series using parametric methods and alternative volatility filtering methods. To this extent, the emphasis on daily frequencies makes our findings comparable to those in the classical stochastic volatility literature. From a technical standpoint, however, we depart from the existing stochastic volatility literature in continuous time along two dimensions. First, with the exception of parametric assumptions imposed on the distribution of the jump sizes for identification, the methodology is nonparametric and, in this sense, robust to model mispecification. Second, volatility is not filtered from low-frequency return data, as generally done, but from high-frequency data as in much recent literature in other contexts (c.f., the discussions in Andersen et al., 2010). From a specification standpoint, we emphasize the statistical and economic importance of allowing for co-jumps, along with idiosyncratic jump components in returns and volatility, in the context of a flexible modelling approach for the evolution of prices and volatilities.

Using S&P 500 futures high-frequency returns from April 21, 1982, to February 5, 2009, we find strong evidence of return/volatility co-jumps. Many of the large co-jumps are associated with well-known events like Black Monday (October 19, 1987), the Asian crisis (October 27, 1997), the Russian crisis (August 31, 1998), and the Lehman-Brothers default (September 29, 2008) during the recent financial crisis. The number of estimated yearly co-jumps (8/9) is slightly larger than the number of idiosyncratic discontinuities in the price process (6/7, per annum). The estimated yearly number of idiosyncratic volatility jumps is about 13.

We find that the size distribution of the price co-jumps depends on the underlying volatility level. The dispersion and negative mean of this distribution increase with volatility, thereby leading to more negative and more erratic price co-jumps in times of high volatility. The occurrence of co-jumps is also generally associated with price/volatility changes (i.e., sizes of the jumps) of opposite sign. Large, negative shocks to prices occur along with sizeable volatility spikes. We estimate a functional correlation between the size of the price jumps and the size of the volatility jumps very close to -1.

This result differs from the parametric findings in Eraker et al. (2003) and Eraker (2004) who, in the context of a model with contemporaneous arrivals and correlated jump sizes, find a statistically insignificant correlation between the jump sizes. In contrast, we show that the explicit separation of the jumps into idiosyncratic and common components, along with an alternative volatility filtering method, lead to jump sizes that are strongly negatively correlated. This effect is important. Among other issues, it introduces a second, less conventional, source of leverage in the model. Not only are the small Brownian shocks affecting prices and volatility negatively correlated in the model (a traditional leverage effect), we find that the infrequently large discontinuities are also strongly negatively correlated. This feature implies that there are two channels through which

negative shocks to prices can increase variance and, therefore, lead to skewness in the conditional (on volatility) distribution of stock returns. Similarly, there are two channels through which the price and volatility dynamics will lead to downward-sloping implied volatility curves across strikes for short and medium maturity options (the ubiquitous "implied volatility smirk"). The paper discusses and illustrates both channels. Consistent with this observation, Eraker (2004) finds a negative correlation between jump sizes only when augmenting return data with option data. In his framework, the negative correlation between co-jump sizes is identified off of the implied volatility smirk. Using our proposed modelling and identification approach, as well as high-frequency return data for the purpose of volatility filtering, we find that the second source of skewness (i.e., a negatively correlated co-jump size) is, in fact, an important feature of the dynamic evolution of prices and volatility series. We show that price data alone may uncover this property with no need for a cross-section of options.

We proceed as follows. Section 2 provides economic evidence of large co-jumps. We show how the most sizeable jumps in returns in our sample are associated with contemporaneous jumps in volatility of opposite sign. In section 3 we present a flexible stochastic volatility model with idiosyncratic and common jumps. We show how the structure of the model we propose generalizes to a nonparametric framework successful parametric specifications in the literature. Section 4 presents the notion of *infinitesimal cross-moment* and discusses its usefulness for co-jump identification. Section 5 discusses a strategy for (cross-) moment-based identification of the model. Section 6 provides nonparametric empirical findings about the joint price and volatility dynamics. Section 7 turns to a parametric evaluation. In Section 8 we present implications for option pricing. The robustness of the model specification is tested in Section 9. Section 10 concludes. The asymptotic and finite sample properties of the proposed methods are discussed in two technical Appendices.

2 Large co-jumps: descriptive evidence

We go through our sample to identify unusually large price movements. In order to do so, we divide daily close-to-close price changes by the corresponding standard deviation over the previous day.[1] Should the changes be induced by continuous Brownian shocks, in light of the approximate normality of Brownian motion over small time increments, they would be larger than 2 in absolute value with approximate probability equal to 5%. Unusually large, standardized, price changes

[1]The econometric analysis of this paper, described in detail in Section 5 and Appendix B, employs hourly (spot) variance estimates. To obtain the daily estimates used in this section, we scale up the hourly estimates to a daily value and average them for every 6-hour trading day. We then apply an exponential smoother with a 40-day lag to reduce measurement error. The results would be qualitatively very similar if the previous days' variances, which are displayed in Figure 3, were used without applying the smoother.

would therefore signal jumps in the price process.[2]

Column 3 in Table 1 reports the 30 largest standardized (by volatility) daily price changes in our sample. Some of these are associated with well-known events, like Black Monday (October 19, 1987), the Asian crisis (October 27, 1997), the Russian crisis (August 31, 1998), 9/11 (September 17, 2001)[3] and the more recent Lehman-Brothers default (September 29, 2008) during the 2008/2009 financial crisis. Importantly, the vast majority of the largest, abnormal, price changes are negative (24 out of 30). In other words, when large price discontinuities occur, they tend to be the result of negative shocks.

Column 4 in Table 1 reports the corresponding variance changes. In agreement with our subsequent analysis, we report changes in logarithmic variance, rather than in actual variance. Consistent with the price changes, the variance changes are close-to-close. They are obtained by virtue of spot (hourly) volatility estimates making use of intra-daily price data in the last hour of the trading day before being scaled up to a daily value (see Section 5 and Appendix B for details). As earlier in the case of prices, we divide the changes in logarithmic variance by the corresponding standard deviation (volatility of log-variance)[4] to rule out changes which may be consistent with smooth Brownian dynamics. Two effects are striking. First, the log-variance changes associated with the largest standardized price changes have corresponding t-statistics which are, with very few exceptions, large and very statistically significant. Second, the volatility changes are, in general, of opposite sign with respect to the price changes. In other words, abnormally large price changes are associated with abnormally large volatility changes (co-jumps).

Because the sizes of the co-jumps may not have a zero mean, in order to gauge the extent of the correlation between price and volatility jump sizes, we report a scatter plot of the log-variance jumps versus the price jumps. Instead of reporting the co-jumps associated with the 30 largest price changes, as implied by Table 1, we now report the co-jumps associated with price and log-variance t-ratios jointly implying significance at the 0.5% level. The choice of this level is meant to highlight large price and variance discontinuities while, at the same time, preventing the insurgence of a sizeable number of false positives, something which is typical of jump tests. This procedure identifies 70 large daily jumps over 28 years which are plotted in Figure 1. For these jumps, the mean and standard deviation of the price discontinuities are estimated as being equal to -2.27% and 4.87%, whereas the estimated values of the mean and standard deviation of the log-variance jumps are 0.9 and 0.77. Thus, when contemporaneous jumps in price and volatility occur, the volatility jumps tend to be positive whereas the price jumps tend to be negative. In

[2]This same logic is formally employed by Lee and Mykland (2008) to test for jumps using intra-daily return data.

[3]The actual discontinuity is not on 9/11 since markets were closed for five days after 9/11.

[4]Computation of the volatility of log-variance requires a full-blown dynamic model for the price dynamics. In particular, for internal consistency, the model should account for diffusive as well as discontinuous shocks. Here, we employ the volatility of log-variance identified from our proposed stochastic volatility model in Section 7.

Day	Return (%)	t-stat	Log-variance change	t-stat	Description
03-Aug-1984	3.63	4.28	1.26	8.8	
18-Dec-1984	3.09	3.72	0.34	2.4	
08-Jan-1986	-3.30	-4.56	0.78	5.5	
11-Sep-1986	-5.19	-5.77	1.51	10.6	
16-Oct-1987	-7.35	-6.66	1.06	7.4	
19-Oct-1987	-30.01	-24.26	2.96	20.7	Black Monday
14-Apr-1988	-4.68	-4.27	1.86	13.0	Dollar plunge
17-Mar-1989	-2.75	-4.02	0.98	6.9	
13-Oct-1989	-6.85	-10.74	0.67	4.7	Friday 13th
12-Jan-1990	-3.43	-4.37	1.81	12.6	
22-Jan-1990	-3.47	-3.95	1.42	10.0	
17-Jan-1991	4.43	4.89	0.91	6.3	
21-Aug-1991	2.74	3.99	0.03	0.2	
15-Nov-1991	-4.08	-6.57	1.30	9.1	
16-Feb-1993	-2.52	-4.78	1.38	9.7	
04-Feb-1994	-2.33	-5.76	1.62	11.3	
08-Mar-1996	-3.94	-4.92	1.39	9.7	
05-Jul-1996	-2.36	-3.69	0.56	3.9	
27-Oct-1997	-7.80	-7.46	0.64	4.5	Asian Crisis
28-Oct-1997	5.68	5.09	0.75	5.3	
09-Jan-1998	-3.88	-4.08	1.02	7.2	
04-Aug-1998	-3.60	-3.76	1.21	8.5	
31-Aug-1998	-7.30	-5.41	0.47	3.3	Russian crisis
04-Jan-2000	-3.52	-3.99	-0.20	-1.4	
14-Apr-2000	-8.11	-4.90	1.11	7.7	Dot.com crash
03-Jan-2001	5.18	3.90	0.50	3.5	
17-Sep-2001	-5.02	-4.22	0.58	4.0	9/11
20-Jan-2006	-1.93	-3.64	0.67	4.7	
27-Feb-2007	-3.23	-7.07	2.58	18.0	Chinese Correction
29-Sep-2008	-6.93	-4.09	1.76	12.3	Lehman-Brothers default

Table 1: Descriptive evidence on co-jumps. The figures are for daily returns in percentage form.

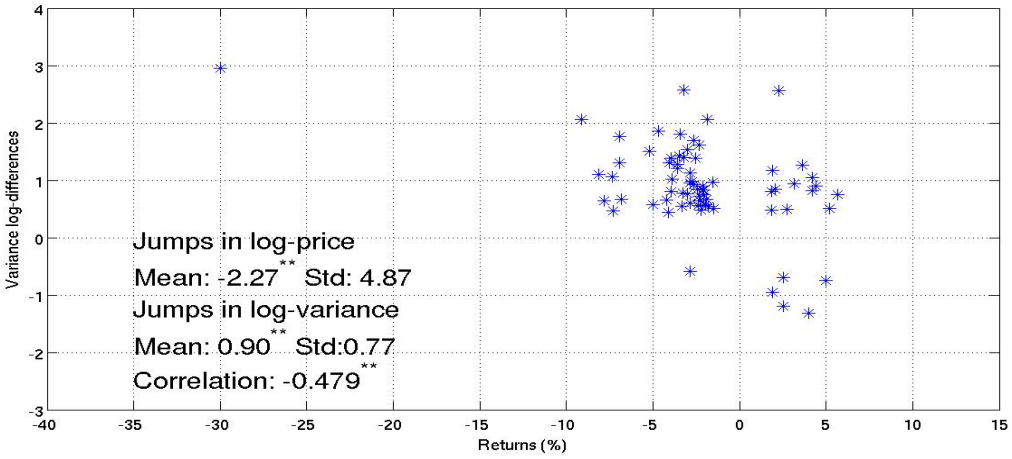

Figure 1: Scatter plot of the significant co-jumps in price and volatility.

addition, the scatter plot provides compelling visual evidence about a strongly negative correlation between co-jump sizes (measured at a significant value of about -0.5). In other words, price jumps below their (negative) mean generally occur alongside volatility spikes above their (positive) mean, thereby resulting in co-jumps with strongly negatively-correlated jump sizes.[5]

By sorting with respect to the large price (and volatility) discontinuities, because we only detect the most sizeable jumps, this evidence should be viewed as providing a lower bound on the number of price and volatility co-jumps in our sample. In agreement with this observation, our point estimates (below) will imply 8/9 daily co-jumps per annum.

In Figure 2 we plot the large price discontinuities displayed in Figure 1 with respect to the volatility level prevailing at the time when the jumps occur. This figure suggests that both the mean and the standard deviation of the sizes of the price co-jumps vary substantially with the volatility level. Specifically, we find that larger volatility levels are associated with co-jumps that are more negative, in expectation, and more variable. This result points to the need for rich, possibly nonlinear, dynamics for the price co-jump sizes.

Next, we present a model featuring joint and idiosyncratic discontinuities in prices and volatility. Consistent with the evidence reported here, the model will allow for correlation between the jump sizes. By permitting the price and volatility jump sizes not to be mean zero, the model will also allow for a higher likelihood of negative price jumps and a higher likelihood of positive volatility jumps. Further, we will let the mean and standard deviation of price co-jumps vary with the volatility level. In the context of a flexible specification for the price and volatility dynamics,

[5]There is one main exception in our sample. During the aftermath of the 1987 crash, on October 21st 1987, a percentage change of 16.11 in logarithmic volatility was accompanied by a percentage (positive) change of 15.5 in log-prices. The corresponding daily volatility was roughly 4.6% daily or 73% on a yearly basis. While we do not report this glaring outlier in the figure, the corresponding data point is used in all of our estimates.

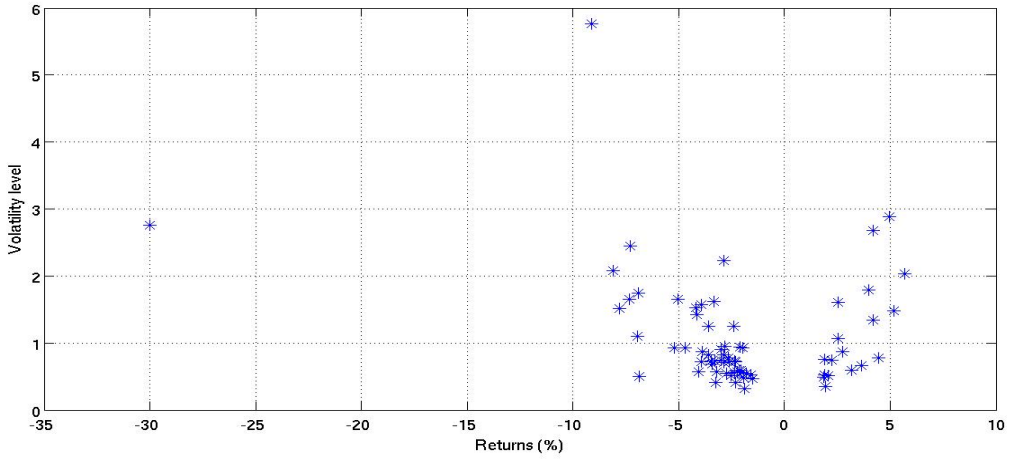

Figure 2: Scatter plot of the significant price co-jumps and the volatility level (%, daily) at which they occur.

we will find estimates of the features of the co-jump sizes, including their correlation, which will be qualitatively and quantitatively similar to those reported in this preliminary descriptive analysis.

3 A continuous-time stochastic volatility model with return/volatility co-jumps

We address the literature in continuous time[6] by casting the problem in the context of a classical jump-diffusion specification with stochastic volatility. Differently from the existing literature, however, with the sole exception of conditions on the probability distribution of the jump sizes, we do not impose any parametric assumption on the functions driving returns and volatility dynamics. The return/variance system we consider is

$$\begin{aligned} d(\log p_t) &= \mu(\sigma_t)dt + \sigma_t\left\{\rho(\sigma_t)dW_t^1 + \sqrt{1-\rho^2(\sigma_t)}dW_t^2\right\} + c_{r,t}^J dJ_r + c_{r,t}^{JJ} dJ_{r,\sigma} \\ d\xi(\sigma_t^2) &= m(\sigma_t)dt + \Lambda(\sigma_t)dW_t^1 + c_{\sigma,t}^J dJ_\sigma + c_{\sigma,t}^{JJ} dJ_{r,\sigma}, \end{aligned} \tag{1}$$

where $\xi(\cdot)$ is an increasingly monotonic function, $W = \left\{W^1, W^2\right\}$ is a bivariate standard Brownian motion vector, $J = \left\{J_r, J_\sigma, J_{r,\sigma^2}\right\}$ is an independent (of W) trivariate vector of mutually independent Poisson processes with intensities $\lambda_r(\sigma_t)$, $\lambda_\sigma(\sigma_t)$, and $\lambda_{r,\sigma}(\sigma_t)$, respectively. The functions $\mu(\cdot), m(\cdot), \Lambda(\cdot), \lambda_r(\cdot), \lambda_\sigma(\cdot), \lambda_{r,\sigma}(\cdot)$ and $\rho(\cdot)$ satisfy mild smoothness conditions and are

[6] Alizadeh et al. (2002); Andersen et al. (2002); Bakshi et al. (1997); Bates (2000); Chernov and Ghysels (2000); Chernov et al. (2003); Duffie et al. (2000); Eraker (2004); Eraker et al. (2003); Heston (1993); Jones (2003); Pan (2002), among others.

solely such that a unique, recurrent, strong solution to the system exists. The system has several features:

1. (Independent jumps and co-jumps) As mentioned, we allow for independent jumps J_r, J_σ as well as for common jumps $J_{r,\sigma}$.

2. (State-dependent intensities) The intensities of the jumps may be nonlinear functions of the underlying variance level. We therefore extend models with affine variance-dependent price jump intensities (Bates, 2000, and Pan, 2002) as well as models with affine variance-dependent volatility jump intensities (Eraker, 2004).

3. (State-dependent jumps) The moments of the jump sizes are permitted to be functions of the underlying variance level. Assume $\xi(\cdot) = \log(\cdot)$, as in a log-variance model, for instance. Let the sizes of the idiosyncratic jumps be normally distributed: $(c^J_{r,t}, c^J_{\sigma,t}) \sim \mathcal{N}(\mu^J, \Sigma^J)$ with $\mu^J = (\mu_{J,r}, \mu_{J,\sigma})^\intercal$ and
$$\Sigma^J = \begin{pmatrix} \sigma^2_{J,r} & \blacklozenge \\ 0 & \sigma^2_{J,\sigma} \end{pmatrix}. \tag{2}$$
Similarly, let the sizes of the co-jumps be normally distributed: $(c^{JJ}_{r,t}, c^{JJ}_{\sigma,t}) \sim \mathcal{N}(\mu^{JJ}, \Sigma^{JJ})$ with $\mu^{JJ} = (\mu_{JJ,r}, \mu_{JJ,\sigma})^\intercal$ and
$$\Sigma^{JJ} = \begin{pmatrix} \sigma^2_{JJ,r} & \blacklozenge \\ \rho_J \sigma_{JJ,r} \sigma_{JJ,\sigma} & \sigma^2_{JJ,\sigma} \end{pmatrix}. \tag{3}$$
We can allow all matrices to be a function of σ. This dependence may, for instance, lead to more (less) erratic discontinuities in times of higher (lower) diffusive variance, an effect which, as shown, is coherent with the properties of the data.

4. (Time-varying leverage) The correlation between shocks to prices and shocks to variance ($-1 \leq \rho(\cdot) \leq 1$) is also a flexible function of the underlying spot variance level. In a stochastic volatility model with independent jumps, Bandi and Renò (2011) show that leverage is a decreasing function of the spot variance level. In other words, leverage becomes more negative as the variance level increases. Here, we generalize their framework by employing the same flexible functional form for $\rho(\cdot)$ in the context of a stochastic volatility model with co-jumps (along with independent jumps). The addition of co-jumps is of course not immaterial for the purposes of leverage identification. Because $\rho(\cdot)$ is usually identified off of the covariance between shocks to prices and shocks to variance, the potential presence of co-jumps, which we accommodate explicitly, would make the estimates of $\rho(\cdot)$ contaminated by the (conditional) covariance between the discontinuous jump components in prices and in volatility. It is therefore important to distinguish between the component of the covariance between price changes and variance changes which is due to the continuous shocks to the

system ($\rho(\cdot)$) and the component due to discontinuous shocks, something which is done in this paper.

5. (Possibly non-affine structures) We do not express the price and volatility drifts, the variance-of-variance, and the intensities of the jumps as affine (or, more generally, parametric) function of the underlying state. While the non-affine nature of our problem prevents us from obtaining closed-form, or near closed-form, derivative prices, here we opt for robustness to potential model mispecifications. This robustness property is important. For example, there is evidence that the variance of variance ($\Lambda^2(\cdot)$) should be a more flexible, possibly CEV, function of the underlying spot variance (e.g., Jones, 2003). Similarly, it may be important to allow the variance process to mean-revert at different speeds depending on its level, as implied by a non-linear mean-reverting variance drift. Eraker (2004), for instance, emphasizes that such a specification could lead to the sharp declines in option implied volatilities following the high levels associated with the 1987 crash.

We now turn to identification of the system by virtue of a moment-based nonparametric procedure.

4 Infinitesimal cross-moments

Let prices and volatilities evolve dynamically as in Eq. (1). The jump sizes are expressed as in Eq. (2) and Eq. (3) with μ^J, μ^{JJ}, $\Sigma^J(\cdot)$, and $\Sigma^{JJ}(\cdot)$. Specifically, we work with a logarithmic variance specification ($\xi(\sigma_t^2) = \log(\sigma_t^2)$) and Gaussian jumps. The Gaussian jump moments also depend on the level of volatility. Section 8 provides a specification analysis which supports a logarithmic transformation for variance.

The key element of the identification method we propose is the generic *infinitesimal cross-moment* of order p_1, p_2 with $p_1 \geq p_2 \geq 0$, namely

$$\vartheta_{p_1,p_2}(\sigma) = \lim_{\Delta \to 0} \frac{1}{\Delta} \mathbf{E}\left[(\log p_{t+\Delta} - \log p_t)^{p_1} \left(\log(\sigma^2_{t+\Delta}) - \log(\sigma^2_t) \right)^{p_2} | \sigma_t = \sigma \right]. \tag{4}$$

By varying the orders p_1 and p_2, the cross-moments have useful interpretations in terms of the underlying functions of the system. Below, we provide the general expressions and a few illustrative examples. For notational convenience, we drop the dependence on σ and write ϑ_{p_1,p_2} instead of $\vartheta_{p_1,p_2}(\sigma)$. All moments and functions, however, should be interpreted as being dependent on the underlying spot volatility process.

We begin with the volatility moments. Write

$$\begin{aligned}
\vartheta_{0,1} &= m + \vartheta_{0,1}^{Jump}, && (5)\\
\vartheta_{0,2} &= \Lambda^2 + \vartheta_{0,2}^{Jump}, && (6)\\
\vartheta_{0,p_2} &= \vartheta_{0,p_2}^{Jump} \qquad p_2 \geq 3 && (7)
\end{aligned}$$

with

$$\vartheta_{0,p_2}^{Jump} = \lambda_{r,\sigma} \sum_{j=0}^{p_2} \left(\begin{array}{c} p_2 \\ j \end{array} \right) G_{0,j} \left(\sigma_{JJ,\sigma}\right)^j \left(\mu_{JJ,\sigma}\right)^{p_2-j} + \lambda_\sigma \sum_{j=0}^{p_2} \left(\begin{array}{c} p_2 \\ j \end{array} \right) G_{0,j} \left(\sigma_{J,\sigma}\right)^j \left(\mu_{J,\sigma}\right)^{p_2-j},$$

where $G_{0,0} = 1$ and, for $g, g_1, g_2 \geq 1$,

$$\begin{aligned}
G_{0,2g} &= (2g-1)!!,\\
G_{0,2g-1} &= 0,\\
G_{g_1,g_2} &= (g_1+g_2-1)\rho_J G_{g_1-1,g_2-1} + (g_1-1)(g_2-1)(1-\rho_J^2) G_{g_1-2,g_2-2}.
\end{aligned}$$

The expressions imply that

$$\vartheta_{0,1} = m + \lambda_{r,\sigma}\mu_{JJ,\sigma} + \lambda_\sigma \mu_{J,\sigma},$$

$$\vartheta_{0,2} = \Lambda^2 + \lambda_{r,\sigma}\left(\left(\mu_{JJ,\sigma}\right)^2 + \left(\sigma_{JJ,\sigma}\right)^2\right) + \lambda_\sigma \left(\left(\mu_{J,\sigma}\right)^2 + \left(\sigma_{J,\sigma}\right)^2\right),$$

$$\begin{aligned}
\vartheta_{0,3} &= \lambda_{r,\sigma}\left(\left(\mu_{JJ,\sigma}\right)^3 + 2\left(\mu_{JJ,\sigma}\right)\left(\sigma_{JJ,\sigma}\right)^2\right) + \lambda_\sigma \left(\left(\mu_{J,\sigma}\right)^3 + 2\left(\mu_{J,\sigma}\right)\left(\sigma_{J,\sigma}\right)^2\right),\\
\vartheta_{0,4} &= ...
\end{aligned}$$

In other words, *all* infinitesimal variance moments depend on the corresponding Gaussian moments of the idiosyncratic and common jumps. The first and the second infinitesimal moments also depend on the drift m and the variance of variance Λ^2, respectively. The price moments $\vartheta_{p_1,0}$ can be defined similarly. We now turn to the genuine cross-moments:

$$\vartheta_{1,1} = \rho\Lambda\sigma + \vartheta_{1,1}^{Jump} \tag{8}$$

and

$$\vartheta_{1+p_1,1+p_2} = \vartheta_{1+p_1,1+p_2}^{Jump} \qquad p_1 > 1 \text{ or } p_2 > 1 \tag{9}$$

with

$$\vartheta_{p_1,p_2}^{Jump} = \lambda_{r,\sigma} \sum_{j_1=0}^{p_1} \sum_{j_2=0}^{p_2} \left(\begin{array}{c} p_1 \\ j_1 \end{array} \right) \left(\begin{array}{c} p_2 \\ j_2 \end{array} \right) G_{j_1,j_2} \left(\sigma_{JJ,r}\right)^{j_1} \left(\sigma_{JJ,\sigma}\right)^{j_2} \left(\mu_{JJ,r}\right)^{p_1-j_1} \left(\mu_{JJ,\sigma}\right)^{p_2-j_2}.$$

The cross-moment expressions imply, for instance, that

$$\vartheta_{1,1} = \rho\Lambda\sigma + \lambda_{r,\sigma} \left(\rho_J \sigma_{JJ,r}\sigma_{JJ,\sigma} + \mu_{JJ,r}\mu_{JJ,\sigma}\right),$$

and

$$\begin{aligned} \vartheta_{2,2} &= \lambda_{r,\sigma}\{\left(\mu_{JJ,\sigma}\right)^2 \left(\mu_{JJ,r}\right)^2 + \left(\sigma_{JJ,\sigma}\right)^2 \left(\mu_{JJ,r}\right)^2 + \left(\mu_{JJ,\sigma}\right)^2 \left(\sigma_{JJ,r}\right)^2 \\ &\quad + \left(1+2\rho_J^2\right) \left(\sigma_{JJ,r}\right)^2 \left(\sigma_{JJ,\sigma}\right)^2 + 4\rho_J \mu_{JJ,r}\mu_{JJ,\sigma}\sigma_{JJ,r}\sigma_{JJ,\sigma}\}. \end{aligned}$$

Thus, the moment of order $1,1$ depends on the covariance between the continuous (Brownian) shocks to the system as well as on the conditional moment of order $1,1$ of the common jump sizes. This effect generates an additional source of leverage in the model. The higher (than $1,1$) cross-moments only depend on the corresponding cross-moments of the common jumps.

The existence of moments lends itself to (a form of) GMM estimation. In particular, given nonparametric estimates $\widehat{\vartheta}_{p_1,p_2}$, we may select an appropriate number of moment conditions, by varying p_1 and p_2, and identify all functions for every volatility level. This procedure effectively amounts to nonparametric GMM in that estimation is conducted for any chosen level in volatility range and results in nonparametric estimates. Similarly, by imposing a parametric structure on the relevant functions, we may use the estimated moments $\widehat{\vartheta}_{p_1,p_2}$ to identify parameters for the full system. In what follows, we use both strategies. In particular, we employ the unrestricted nonparametric functional estimates as useful inferential and descriptive tools to identify a rich system allowing for idiosyncratic, as well as common, discontinuous dynamics in both prices and volatility. We then turn to a parametric evaluation of the full system's dynamics.

5 Cross moment-based identification

The cross-moments are (infinitesimal) conditional expectations. We use sample analogues to conditional expectations based on classical kernel estimators to identify them. Consider a sample of T days and N intraday knots within each day. Assume availability of closing logarithmic prices $(\log p_{t,i})$ and spot volatility estimates $(\log \widehat{\sigma}_{t,i}^2)$ over each day $t = 1, \ldots, T$ and each knot $i = 1, \ldots, N$. The generic cross-moment estimator $\widehat{\vartheta}_{p_1,p_2}$ is defined as

$$\widehat{\vartheta}_{p_1,p_2}(\sigma) = \frac{\sum_{t=1}^{T_{days}-1} \sum_{i=1}^{N} \mathbf{K}\left(\frac{\widehat{\sigma}_{t,i}-\sigma}{h}\right) \left(\log p_{t+1,i} - \log p_{t,i}\right)^{p_1} \left(\log \widehat{\sigma}_{t+1,i}^2 - \log \widehat{\sigma}_{t,i}^2\right)^{p_2}}{\Delta \sum_{t=1}^{T_{days}} \sum_{i=1}^{N_{hours}} \mathbf{K}\left(\frac{\widehat{\sigma}_{t,i}-\sigma}{h}\right)}, \qquad (10)$$

where $\mathbf{K}(.)$ is a kernel function, with properties listed in Appendix A, whose role is to guarantee proper conditioning at σ for a small bandwidth h.

As emphasized, we use intra-daily observations for spot volatility estimation and, subsequently, for subsampling on the infinitesimal moments in Eq. (10). The use of intra-daily data, along with our identification procedure, separates us methodologically from the bulk of the stochastic volatility literature. However, importantly, since Δ is equal to 1 day, the resulting estimates are daily. This is done to make our empirical results comparable to those in the existing stochastic volatility literature.

In our implementation of the estimator in Eq. (10), we employ $N = 6$ knots in the interval 10.45am−3.45pm, each separated by one hour. Define one-minute logarithmic returns $r_{t,i,k} = \log p_{t,i,k} - \log p_{t,i,k-1}$, for $k = 1, \ldots, 60$, over each hour before a generic knot i (we start using price observations at 9.45am).[7] The spot variance estimates, for each day t and each knot i, are obtained by applying the jump-robust threshold bipower variation estimator

$$\widehat{\sigma}^2_{t,i} = \frac{60}{59 - n_j} \varsigma_1^{-2} \sum_{k=2}^{60} |r_{t,i,k}||r_{t,i,k-1}| I_{\{|r_{t,i,k}| \leq \theta_{t,i,k}\}} I_{\{|r_{t,i,k-1}| \leq \theta_{t,i,k-1}\}}, \tag{11}$$

where $\varsigma_1 \simeq 0.7979$, $\theta_{t,i,k}$ is a suitable threshold, and n_j is the number of returns whose absolute value is greater than the threshold $\theta_{t,i,k}$ (Corsi et al., 2010). Alternative spot volatility estimators may be used. However, any estimator used should identify spot (diffusive) variance and, hence, be robust to jumps in the price process. The estimator we employ eliminates jumps in two ways. Similarly to the classical bipower variation of Barndorff-Nielsen and Shephard (2004), discontinuities are annihilated asymptotically by the adjacent diffusive component. In addition, the jumps are discarded in finite samples, and asymptotically, when above a (vanishing, asymptotically) threshold $\theta_{t,i,k}$ (as in the threshold realized variance of Mancini, 2009).

The use of spot volatility estimates, as in Eq. (11), paired with the functional moment estimates, in Eq. (10), will be used for pointwise identification of all of the system's functions. Our approach relates to two existing approaches in the literature. Bandi and Renò (2009) first introduce nonparametric identification of discontinuous stochastic volatility models under the assumption of independent jumps. Bandi and Renò (2011) focus explicitly on nonparametric leverage estimation and, in agreement with Bandi and Renò (2009), only allow for independent jumps. Differently from our previous work on the subject, this paper allows for co-jumps and focuses on an empirical treatment which highlights the statistical and economic relevance of discontinuous joint changes in prices and volatility. In particular, the addition of co-jumps with non-zero jump sizes, which represents the substantive core of our treatment, leads to two substantial innovations for the purpose of identification. First, we recognize the crucial role played by the *infinitesimal*

[7] The one minute returns are constructed after pre-averaging all observed transaction prices within each minute. The pre-averaging is designed to reduce the impact of market microstructure noise (Jacod et al., 2009).

cross-moments for co-jump identification. With the exception of the cross-moment of order $1,1$, which is also affected by diffusive leverage, the cross-moments are zero when the discontinuities are idiosyncratic, as in our previous work, and, therefore, play no role in their identification. Second, when allowing for the more complex jump dynamics in this paper, closed-form estimation of the system's functions in terms of the estimated moments, as conducted in Bandi and Renò (2009, 2011) is, in general, not feasible. Importantly, however, the infinitesimal cross-moments introduced here lend themselves to an estimation method akin to pointwise GMM (Hansen, 1982), something which we discuss below.

Denote by $g_1(\sigma),\ldots,g_K(\sigma)$ the K functions driving the dynamics of the system. These are the estimation target. Consider a set of N cross-moments $\widehat{\vartheta}_{p_1,p_2}(\sigma)$ with $N \geq K$ for identification. The theoretical cross-moments $\vartheta_{p_1,p_2}(\sigma) = f_{p_1,p_2}(g_1(\sigma),\ldots,g_K(\sigma))$ are a mapping f_{p_1,p_2} from the functions $g_1(\sigma),\ldots,g_K(\sigma)$ to $\mathbb{R}$. For every value σ in the spot volatility range, the K vector of estimates $(\widehat{g}_1(\sigma),\ldots,\widehat{g}_K(\sigma))$ is defined as:

$$(\widehat{g}_1(\sigma),\ldots,\widehat{g}_K(\sigma)) = \underset{(g_1(\sigma),\ldots,g_K(\sigma))}{\arg\min} (\widehat{\vartheta}_{p_1,p_2}(\sigma) - \vartheta_{p_1,p_2}(\sigma))^\top W(\sigma)(\widehat{\vartheta}_{p_1,p_2}(\sigma) - \vartheta_{p_1,p_2}(\sigma)),$$

where $W(\sigma)$ is an $N\times N$ symmetrical and positive definite weighting matrix. We implement a two-stage procedure to optimize the choice of $W(\sigma)$. In the first stage, $W(\sigma)$ is the identity matrix. The first-stage estimates are used to compute V, the variance-covariance matrix of the cross-moments, i.e., the matrix with generic element $V_{(p_1,p_2),(p_3,p_4)} = \frac{f_{p_1+p_3,p_2+p_4}(\widehat{g}_1(\sigma),\ldots,\widehat{g}_K(\sigma))}{h(\sigma)\widehat{L}_T(\sigma)}$, where $h(\sigma)$ is a volatility-specific bandwidth and $\widehat{L}_T(\sigma)$ is an estimate of the volatility's occupation density. The estimation is then repeated with $\widehat{W}(\sigma) = V^{-1}$. The same procedure is implemented for different values of σ leading to the nonparametric estimates presented in the next section.

The functional estimates have the potential to provide important guidance about parametric specification. Importantly, given a parametrization of the system's functions, the procedure we propose can be adapted to also yield parametric estimates. Denote by η a vector of M parameters. Select a number G of knots $\sigma_1,\ldots,\sigma_G$, so that $N \times G \geq M$ for identification. Denote by $\widehat{\vartheta}_{p_1,p_2}$ the $N \times G$-vector of available estimated moments computed at the knots σ_i with $i = 1,...,G$ and by $\vartheta_{p_1,p_2}(\eta)$ the corresponding $N \times G$-vector of theoretical moments. The parametric estimates are now given by:

$$\widehat{\eta} \;=\; \underset{\eta}{\arg\min}(\widehat{\vartheta}_{p_1,p_2} - \vartheta_{p_1,p_2}(\eta))^\top W(\sigma)(\widehat{\vartheta}_{p_1,p_2} - \vartheta_{p_1,p_2}(\eta))$$

where $W(\sigma)$ is an $(N \times G) \times (N \times G)$ symmetrical and positive definite weighting matrix. To optimize the choice of $W(\sigma)$ and underweight relatively noisier moments, we evaluate the variance-covariance matrix between the N moments at the G knots (V) via simulation. We then employ $W = V^{-1}$.

Appendix A derives the consistency and asymptotic normality properties of $\widehat{\vartheta}_{p_1,p_2}$, for a known σ, under a suitable asymptotic sampling scheme. Appendix A also provides conditions on $\theta_{t,i,k}$,

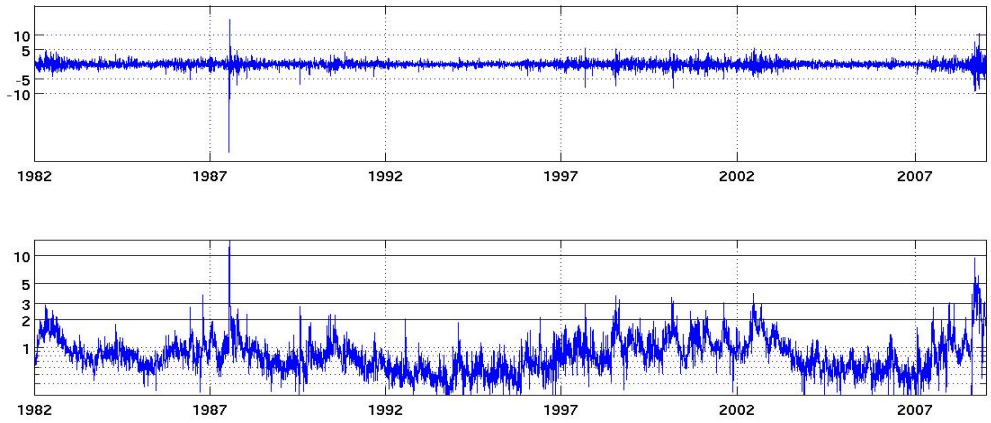

Figure 3: Top: Daily returns (%). Bottom: Filtered daily volatilities (in log scale) obtained by averaging hourly spot volatilities within each day (%).

among other choice variables, so that these properties are not affected by the estimation error necessarily introduced by the spot variance estimates obtained in Eq. (11) above. Appendix B discusses finite sample issues and issues of implementation.

6 Price and variance dynamics: functional estimates

As mentioned, we employ all transactions on S&P 500 futures from from April 21, 1982, to February 5, 2009, for a total of 6,748 trading days. The daily return data and daily *filtered* (by virtue of high-frequency price data) logarithmic volatility series are reported in Figure 3. Figure 4 provides the estimated moments of order p_1, p_2 above the diagonal of a matrix obtained by varying p_1 and p_2 between 0 and 4. In addition, the figure reports the moments of order $2, 3$ and $3, 2$. The upper diagonal moments and the moments of order $2, 3$ and $3, 2$ are those used for estimation. The functions driving the price and volatility dynamics in Eq. (1) are reported in Figure 5. All functions are plotted with respect to daily volatility expressed as a percentage. In terms of interpretation, a value of 1 on the horizontal axis corresponds to a yearly volatility of about $\sqrt{252}\% = 15.87\%$.

We begin by commenting on the price process. The price drift μ is insignificantly different from zero across volatility states. The inability to detect risk-return trade-offs using daily return data is a notorious phenomenon which we confirm here.[8] The intensity of the idiosyncratic jumps in prices λ_r is fairly stable, across volatility levels, and implies between roughly 5 and 15 jumps per year in terms of point estimates. As expected in this type of problems, the level of statistical

[8] Andersen et al. (2002), Pan (2002) and Eraker et al. (2004) also find a statistically insignificant dependence of the price drift on spot variance.

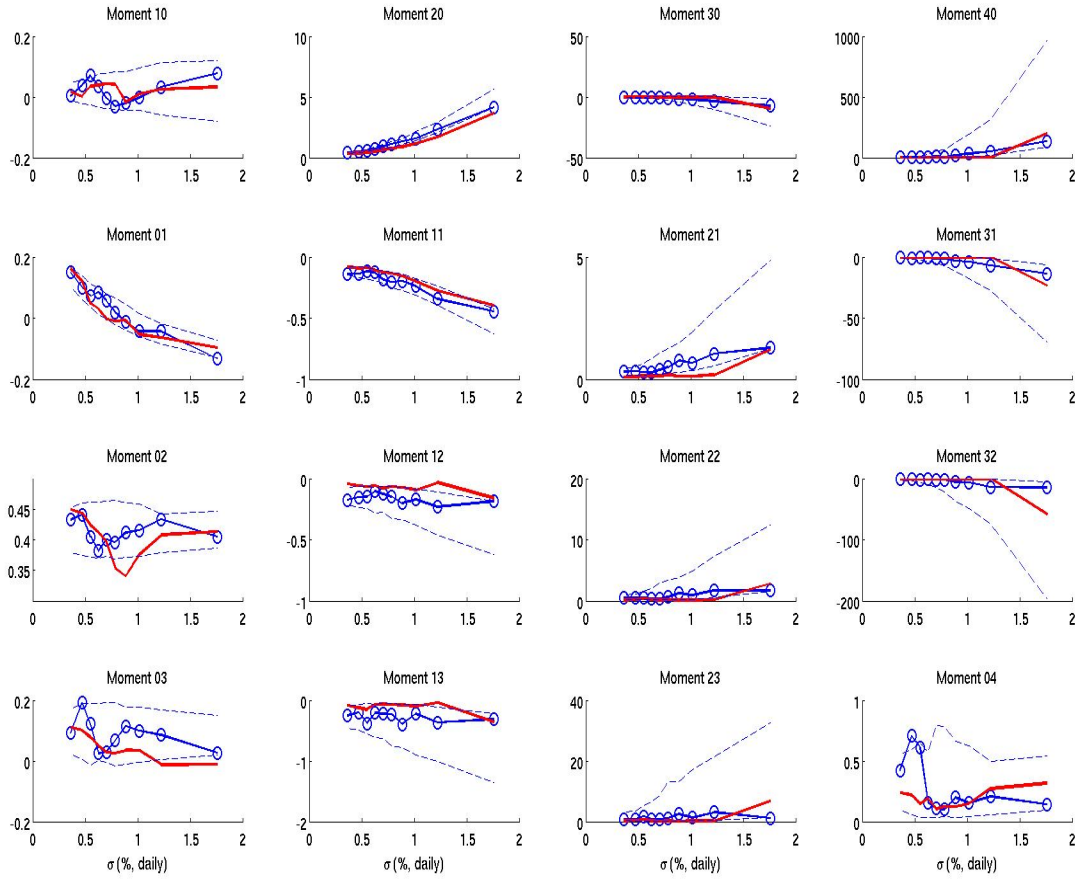

Figure 4: Estimated infinitesimal cross-moments (solid circled lines) with 95% confidence bands (dashed lines). The solid lines are implied *fitted* moments using functions, displayed in Figure 5, obtained via pointwise GMM.

uncertainty is substantial, particularly for relatively larger, seldom visited, volatility levels. The parametric estimates will refine this analysis. The mean of the idiosyncratic price jumps $\mu_{J,r}$ is indistinguishable from zero. The corresponding standard deviation $\sigma_{J,r}$ has a very slight tendency to increase with the volatility level, thereby suggesting the possibility of increasing jump sizes in times of higher uncertainty. This finding will be more evident in the case of the price co-jumps.

We now turn to stochastic variance. As expected, the variance drift m implies mean-reversion in log variance. The variance of variance Λ is relatively flat across volatility states. For instance, it does not display the CEV shape that has been emphasized as being important in some existing work (see, e.g., Chacko and Viceira, 2003, and Jones, 2003). This is, however, an unsurprising outcome of the logarithmic transformation of variance which we adopt in the present paper and justify in the subsequent specification analysis. The nonparametric inference presented here

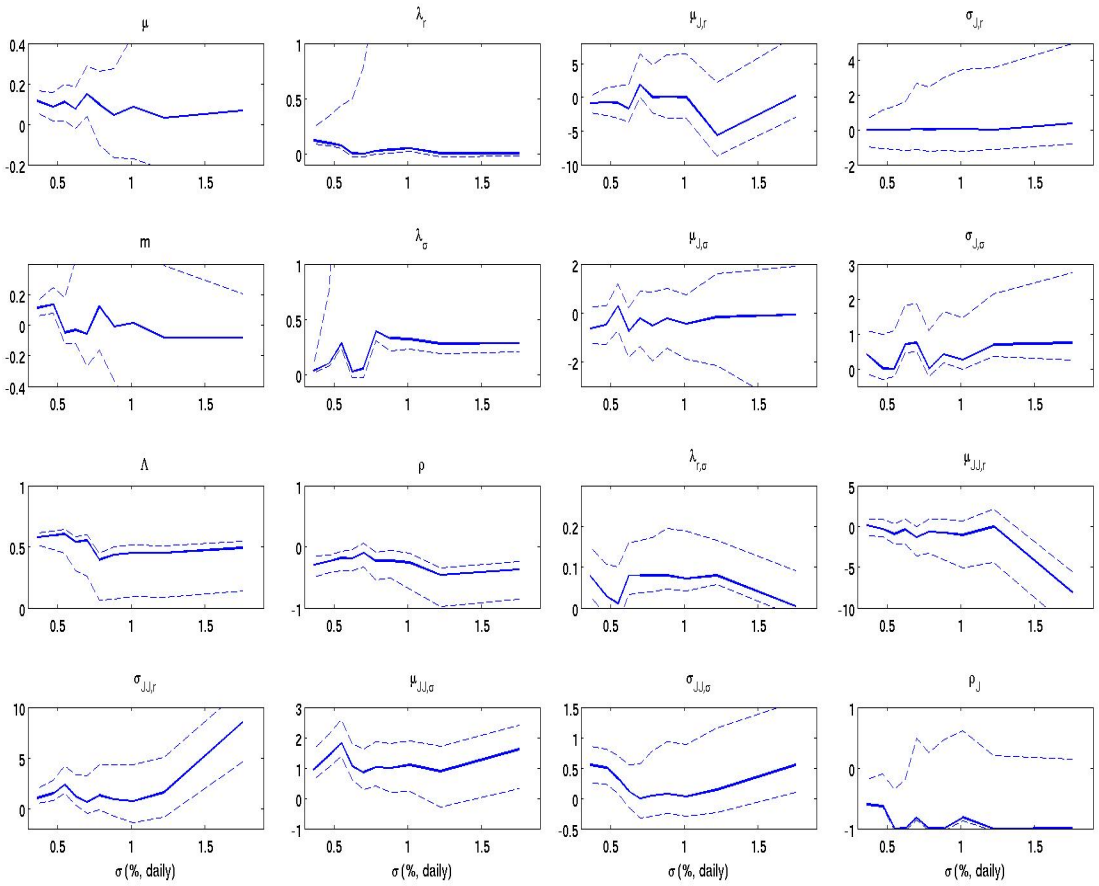

Figure 5: Estimated functions (solid lines) with 95% confidence bands (dashed lines).

yields point estimates for the intensity of the variance jumps λ_σ which vary considerably across volatility states while also carrying a substantial degree of statistical uncertainty. The functional λ_σ estimates suggest a larger number of independent volatility jumps than independent price jumps, something which our parametric estimates in the next section will confirm. The mean of the idiosyncratic variance jumps is close to zero. The size of the independent variance jumps is relatively stable across volatility levels.

We now focus on the joint dynamics. The number of co-jumps implied by $\lambda_{r,\sigma}$ is larger than zero and fairly stable across volatility levels, mainly around the bulk of the data ($\sigma = 0.8\%$). When jumping jointly with prices, volatility has a higher probability to increase than to decrease as indicated by a mean of the corresponding jump size distribution $\mu_{JJ,\sigma}$ which is significant and positive. In contrast, the mean of the common jumps in prices $\mu_{JJ,r}$ is negative. Also, more negative price co-jumps are associated with higher volatility levels. As in the idiosyncratic jump case, the standard deviation of the sizes of the common volatility jumps is relatively stable

across volatility levels. The standard deviation of the sizes of the common price jumps $\sigma_{JJ,r}$ is instead strongly increasing with the volatility level. In sum, the estimated volatility co-jump size distribution is fairly invariant to the underlying volatility level. In contrast, the mean and the variance of the size distribution of the price co-jumps become more negative and increase, respectively, as volatility increases. Importantly, the correlation between discontinuous price changes and discontinuous volatility changes ρ_J is negative and very close to -1.

Our evidence about the strong negative correlation between price and volatility jumps is consistent with that in Todorov and Tauchen (2011). Using high-frequency data on the VIX and the S&P 500, they find a high likelihood of contemporaneous jump arrivals, opposite signs for the jumps sizes, and numerous jumps in volatility (attributed, in their framework, to the pure jump nature of the volatility process). Their results hinge on suitable descriptive statistics for high-frequency data, and corresponding tests, developed in Todorov and Tauchen (2010) and Jacod and Todorov (2009). In contrast, our evidence derives from the identification of the price and variance dynamics in the context of a specification allowing for a rich array of discontinuities. As said, we focus on daily series and use high-frequency price observations only for the purpose of filtering the daily volatility estimates used to estimate the model. Differently from Todorov and Tauchen (2011), our analysis is conducted using price data only, rather than data on options and the underlying.

Papers which estimate the price and volatility dynamics in the context of a "doubly" jump-diffusion model and filter volatility using only price data are, in a sense, methodologically more comparable to our approach (Duffie et al., 2000, Eraker et al., 2003, and Eraker, 2004). Unless option data are employed (Eraker, 2004), these papers do not find a negative correlation between co-jumps. In contrast, we emphasize the importance of this additional source of return skewness in a model which, differently from existing applied work, explicitly separates the jump components into idiosyncratic and common factors.[9]

We now turn to the most traditional source of skewness, Brownian leverage. As for other quantities, Figure 5 provides estimates of leverage as a function of spot volatility. This dependence has been emphasized as being important in a specification with independent jumps (Bandi and Renò, 2011). In agreement with Bandi and Renò (2011), but in the context of a more flexible model specification allowing for co-jumps, we find a decrease in the leverage estimates, i.e., values which are more negative associated with higher volatility levels, particularly around the center of the observed volatility range. Thus, the time-varying (with volatility) structure of ρ emphasized by Bandi and Renò (2011) appears to be robust to the presence of an additional source of skewness

[9]This separation is deemed important by Duffie et al. (2000) and defines their more general stochastic volatility model. The more general model is however not estimated. Eraker et al. (2003) and Eraker (2004) estimate models in which the jumps are either independent or they are perfectly dependent in terms of arrival times. In other words, no allowance is made for independent jumps and common jumps, as in our case.

(further discussed below) induced by the joint occurrence of price and volatility jumps.

To summarize, the descriptive, nonparametric, analysis presented here suggests that logarithmic prices and logarithmic variances are likely to feature both independent and common jumps. The independent jumps are roughly mean zero. The common jumps have positive means, if in volatility, and negative means, if in prices. In the joint case, the price size distribution expands and re-centers around more negative values as volatility increases. The size distribution of the volatility co-jumps is, instead, fairly stable. The sizes of the price and volatility co-jumps are strongly anti-correlated.

7 Price and variance dynamics: parametric estimates

We now turn to a parametric assessment. The parametric model is motivated by the evidence in Section 2 and the previous nonparametric analysis. Its specification is:

$$
\begin{aligned}
d\log p_t &= \mu_r dt + \sigma_t \left\{ \rho_t dW_t^1 + \sqrt{1-\rho_t^2} dW_t^2 \right\} + c_{r,t}^J dJ_r + c_{r,t}^{JJ} dJ_{r,\sigma} \\
d\log(\sigma_t^2) &= \left(m_0 + m_1 \log(\sigma_t^2) \right) dt + \Lambda dW_t^1 + c_{\sigma,t}^J dJ_\sigma + c_{\sigma,t}^{JJ} dJ_{r,\sigma}, \\
\rho_t &= \max(\min(\rho_0 + \rho_1 \sigma_t, 1), -1), \\
\{J_r, J_\sigma, J_{r,\sigma}\} &\sim Poisson(\lambda_r, \lambda_\sigma, \lambda_{r,\sigma}) \\
c_{r,t}^J &\sim \mathcal{N}\left(\mu_{J,r}, \sigma_{J,r}^2\right) \\
c_{\sigma,t}^J &\sim \mathcal{N}\left(\mu_{J,\sigma}, \sigma_{J,\sigma}^2\right) \\
\begin{pmatrix} c_{r,t}^{JJ} \\ c_{\sigma,t}^{JJ} \end{pmatrix} &\sim \mathcal{N}\left(\begin{pmatrix} \mu_{JJ,r,0} + \mu_{JJ,r,0}\sigma_t \\ \mu_{JJ,\sigma} \end{pmatrix}, \right. \\
&\qquad \left. \begin{pmatrix} \left(\sigma_{JJ,r,0} + \sigma_{JJ,r,1}\sigma_t^{\sigma_{JJ,r,2}}\right)^2 & \rho_J \left(\sigma_{JJ,r,0} + \sigma_{JJ,r,1}\sigma_t^{\sigma_{JJ,r,2}}\right)\sigma_{JJ,\sigma} \\ \blacklozenge & \sigma_{JJ,\sigma}^2 \end{pmatrix} \right).
\end{aligned}
\tag{12}
$$

The system, which is readily viewed as a special case of the model in Eq. (1), has 21 parameters whose estimates are reported in Column 4 of Table 2.[10] We estimate two relevant restricted models as well. The first model sets $J_{r,\sigma} = 0$ and has 12 parameters whose estimated values are

[10]The confidence bands are obtained by simulation. The estimated parametric model in Table 2, Column 4 is used to generate simulated samples as large as the original sample. The procedure is then applied to every simulated sample.

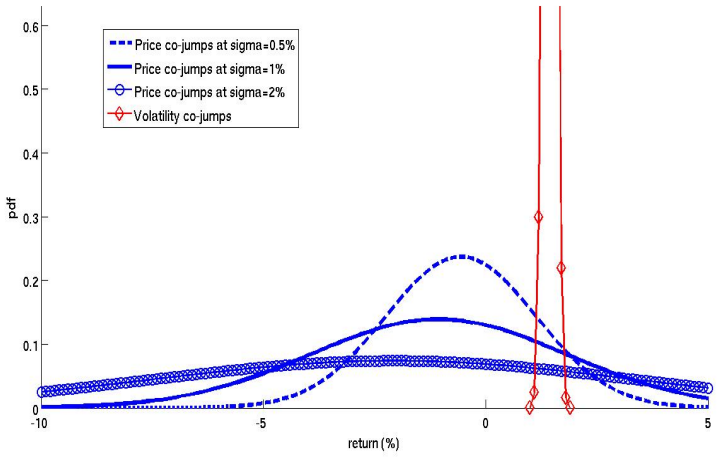

Figure 6: The distributions of the price and volatility co-jumps implied by the parametric estimates in Table 2, Column 4.

reported under the title "no co-jumps" in Column 2 of Table 2. The second model sets $J_r = J_\sigma = 0$ and has 15 parameters whose estimates are reported under the title "no independent jumps" in Column 3 of Table 2.

We begin the discussion, again, with the price equation. Consistent with the functional evidence reported earlier, preliminary estimation of a linear function for the price drift μ, namely $\mu_{r,0} + \mu_{r,1}\sigma$, resulted in an estimated slope coefficient $\mu_{r,1}$ very close to 0, numerically and statistically. Hence, we only report estimates for a constant drift. The extremely weak compensation for variance risk, when evaluated at the daily level, has been confirmed by several empirical studies, including Eraker et al. (2003). Idiosyncratic jumps in prices do not play a relevant role. Even though, when inferred from an estimated constant value for λ_r, the number of idiosyncratic price jumps is found to be around 6/7 per year (0.0252×252), these jumps are fairly small. The mean of the idiosyncratic price jumps is positive (1.39%, in terms of point estimate) but insignificantly different from zero. Their standard deviation is only 0.68%.

Before discussing co-jumps, we turn to the volatility equation. We fit a linear drift, namely, $m = m_0 + m_1 \log(\sigma^2)$, and find very accurately estimated linear mean-reversion in logarithmic volatility. In particular, the parameter m_1 is equal to -0.0597 with a 95% band equal to $[-0.0736, -0.0355]$. The volatility of volatility Λ is precisely estimated at a constant value equal to about 0.56. The corresponding 95% confidence interval is $[0.40, 0.58]$. The number of independent jumps in volatility is estimated at an annual value of about 13 (0.0528×252). The mean of the independent variance jumps is, again, statistically zero (point estimate equal to -0.45 with a 95% confidence band equal to $[-1.06, 0.18]$). The corresponding standard deviation is about 0.7.

The estimated number of co-jumps per year is 8.54 (0.0339×252) with a 95% confidence

parameter	no cojumps	no independent jumps	with cojumps	
μ_r	0.0423	0.0631	0.0306	$(0.0000, 0.1112)$
ρ_0	-0.2280	-0.0977	-0.0988	$(-0.2046, 0.0267)$
ρ_1	-0.0874	-0.1225	-0.1617	$(-0.2839, -0.0850)$
m_0	-0.0232	-0.0397	-0.0380	$(-0.0827, 0.1798)$
m_1	-0.0704	-0.0576	-0.0597	$(-0.0736, -0.0355)$
Λ	0.6048	0.5950	0.5583	$(0.3999, 0.5836)$
$\mu_{J,r}$	-0.1137	$-$	1.3948	$(-0.7035, 2.9739)$
$\mu_{JJ,r,0}$	$-$	0.5210	-0.0544	$(-1.0728, 1.0830)$
$\mu_{JJ,r,1}$	$-$	-1.8976	-1.0072	$(-4.0391, 0.0274)$
$\sigma_{J,r}$	1.2715	$-$	0.6818	$(0.0000, 2.0108)$
$\sigma_{JJ,r,0}$	$-$	1.7428	0.6246	$(0.0000, 1.9913)$
$\sigma_{JJ,r,1}$	$-$	0.1718	2.2469	$(0.7727, 5.0776)$
$\sigma_{JJ,r,2}$	$-$	1.8828	1.0863	$(0.5601, 2.1114)$
$\mu_{J,\sigma}$	0.3498	$-$	-0.4497	$(-1.0624, 0.1792)$
$\mu_{JJ,\sigma}$	$-$	0.7816	1.4428	$(0.9435, 1.5448)$
$\sigma_{J,\sigma}$	1.2575	$-$	0.7002	$(0.0002, 1.1645)$
$\sigma_{JJ,\sigma}$	$-$	0.4901	0.1084	$(0.0074, 0.5011)$
ρ_J	$-$	-0.6416	-1.0000	$(-1.0000, -0.1483)$
λ_r	0.1033	$-$	0.0252	$(0.0045, 0.4548)$
λ_σ	0.0279	$-$	0.0528	$(0.0103, 0.8920)$
$\lambda_{r,\sigma}$	$-$	0.0489	0.0339	$(0.0211, 0.0934)$

Table 2: In Column 4 we display parametric estimates of the model in Eq. (12) with the corresponding confidence intervals. Estimates of the same model with the restriction $J_{r,\sigma} = 0$ (no co-jumps) and $J_r = J_\sigma = 0$ (no independent jumps) are in Column 2 and 3, respectively. The parameters are relative to daily data, and imply returns expressed in percentage form.

interval of $[5.31, 23.53]$ per annum. This number is consistent with the empirical evidence provided in Section 2. Further statistical evidence on the presence of significant co-jumps in the data is provided in Section 9. When jumping jointly with volatility, the price process displays a tendency to jump downwards. This tendency increases with the volatility level as implied by the point estimates of $\mu_{JJ,r,0}$ (-0.05%) and $\mu_{JJ,r,1}$ (-1.01). Guided by the nonparametric evidence in the previous section and the preliminary analysis in Section 2, we fit a nonlinear structure to the standard deviation of the common price jumps, namely $\sigma_{J,r,0} + \sigma_{J,r,1}\sigma^{\sigma_{J,r,2}}$. All parameters are positive and, with the exception of the constant term $\sigma_{J,r,0}$, statistically significant, thereby suggesting an increase in the size of the common price jumps along with increases in volatility. Variance has a tendency to jump upward. The mean of the common jumps in variance is estimated at a value equal to 1.44. The 95% confidence band for this parameter is $[0.94, 1.54]$. The volatility of the common jumps in variance is estimated at a value of 0.1084 with a 95% confidence interval of $[0.007, 0.50]$. To visualize the impact of these numbers, the size distributions of the price and volatility co-jumps are plotted in Figure 6. The distribution of the size of the price co-jumps is plotted as a function of three representative volatility levels (namely, 0.5%, 1%, and 2%). As volatility increases, this distribution expands and re-centers around more negative values. Finally, importantly, the correlation between the common price jumps and the common volatility jumps is estimated at a value which is very clearly negative. The point estimate is at the lower boundary (-1) and the upper limit of the 95% confidence interval is -0.1483.[11]

In sum, the estimates suggest that the price process is more likely than not to jump jointly with the volatility process (the number of price co-jumps is larger than the number of independent price jumps). When moving jointly in a discontinuous fashion, volatility tends to shift upward, whereas prices have a tendency to move downward. This said, the model estimates imply a substantial likelihood of positive price co-jumps, both at high and at low volatility levels (see Figure 6). This is, again, consistent with the preliminary evidence in Section 2, Figure 2. Large positive price co-jumps are less frequent than negative price co-jumps, but they are in the data. Their mean and variability also increase with the underlying volatility level, something which the parametric model readily delivers. Finally, when the volatility jumps are above their positive mean, the price jumps tend to be below their negative mean, as implied by a strongly negative correlation between the common jump sizes.

[11] This negative correlation is larger, in absolute value, than that reported in the descriptive analysis in Section 2. This is due to the fact the preliminary descriptive evidence in Section 2 is based on variance measures which are, by definition, estimates of the true variances and, therefore, affected by some measurement error. This measurement error will induce attenuation effects when evaluating the correlation between price and volatility jump sizes. Thus, on the one hand, we should view the reported correlation in Section 2 as being conservative and, likely, too small in absolute value. On the other hand, the finite sample adjustments leading to the estimates in this section are designed to lead to accurate identification. Appendix B provides details on these adjustments and confirms their effectiveness by simulation.

Guided, again, by the nonparametric estimates, we fit a linear function to leverage, namely $\rho_0 + \rho_1\sigma$. We find that leverage decreases with the volatility level ($\widehat{\rho}_0 = -0.10$ and $\widehat{\rho}_1 = -0.16$). For volatilities between 0.5% and 2%, estimated leverage varies between -0.18 and -0.42. These values are lower, in absolute value, than those reported in the literature. Eraker et al. (2003), for instance, find a leverage value equal to -0.4838 for their specification with contemporaneously-arriving, but empirically uncorrelated in their framework, jumps. Similarly, in a model with the same logarithmic specification used here, but with no jumps in variance (neither idiosyncratic nor contemporaneous), Andersen et al. (2002) estimate a leverage parameter equal to -0.61. This outcome is, however, unsurprising. The presence of *anti-correlated* common jumps is expected to reduce the size of classical leverage. Leverage is generally identified off of the covariance between price changes and volatility changes. In the model we study, this covariance is not just driven by Brownian correlation but also by the correlation between discontinuous shocks to the system. In fact,

$$\rho_{total} = \frac{\vartheta_{1,1}}{\sigma\Lambda} = \rho + \frac{\lambda_{r,\sigma}\left(\rho_J\sigma_{JJ,r}\sigma_{JJ,\sigma} + \mu_{JJ,r}\mu_{JJ,\sigma}\right)}{\sigma\Lambda} = \rho + \rho_{co-jumps}. \tag{13}$$

Even though the number of yearly co-jumps $\lambda_{r,\sigma}$ may not be extremely high, the correlation between the jump sizes is negative and large in absolute value. The means of the price/variance jump sizes are also large and of opposite sign. For a volatility level of $\sigma = 1\%$, for instance, we have $\rho = -0.26$, $\rho_{co-jumps} = -0.11$ and $\rho_{total} = -0.37$. Thus, the contribution of the jumps is found to be substantial. The parametric estimates also indicate (see Figure 7) that ρ is decreasing with the volatility level while $\rho_{co-jumps}$ is nearly constant. Figure 7 also plots the percentage of the correlation ρ_{total} attributable to the Brownian part computed using the parametric estimates in Table 2. The figure shows that this percentage is roughly 60/70% around the center of the volatility range (0.8%). The value varies somewhat with the volatility level, becoming increasingly important as spot volatility increases.

The size of Brownian leverage, and its implications for the skewness of stock returns, has drawn attention in the recent literature (see, e.g., Aït-Sahalia et al., 2011, and Bandi and Renò, 2011). Specifically, referring back to Eq. (13), compelling statistical arguments have been made about the need to account for noise in the estimation of the variance of variance Λ as well as downward biases in the estimation of the price and variance covariance $\vartheta_{1,1}$. Both effects have been shown to lead to attenuation effects in the estimation of leverage and, consequently, the belief that Brownian leverage may, in fact, be larger - i.e., more negative - than previously recognized. This paper addresses these important statistical concerns explicitly by relying on *spot* (hourly) volatility measures as well as by reducing the residual measurement error in the spot variance measures explicitly (see Section 12.4. in Appendix B). Monte Carlo simulations from the parametric model in Table 2 (Figure 11, Appendix B) show that our small sample adjustments are reliable in

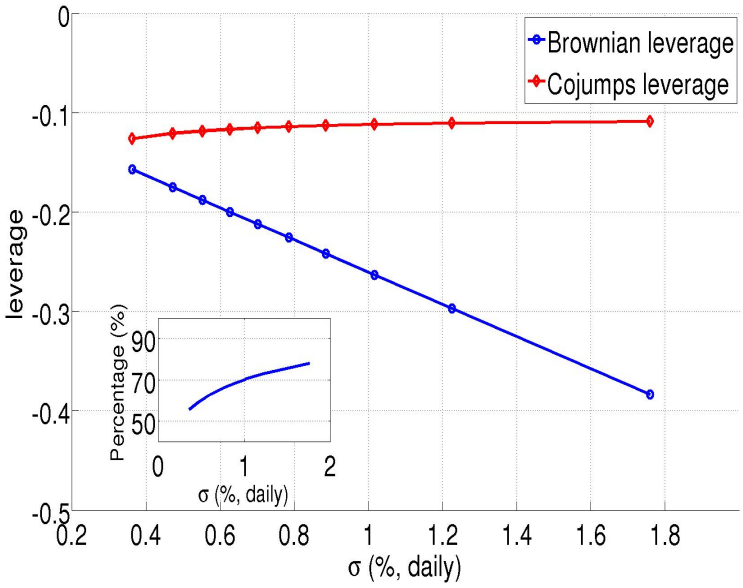

Figure 7: The leverage effect, as a function of volatility, decomposed into its Brownian part and the part due to co-jumps, as implied by the parametric estimates in Table 2, Column 4. In the inset we plot the percentage of covariance explained by Brownian correlation.

providing unbiased estimate of the infinitesimal cross-moments, including the moment $\vartheta_{1,1}$, which is crucial for the identification of leverage effects. Monte Carlo analysis on the estimated leverage parameters, and all other parameters of the systems, provides further, more direct, evidence. Following our small sample adjustments, this analysis provides evidence of minimal distortions in estimating both ρ_0 and ρ_1 (Figure 12, Appendix B, for a visual representation).

In sum, this paper emphasizes an economic argument which may run counter to the idea that true Brownian leverage may be more negative than previously believed. Because a portion of the covariance between price and volatility changes is due to common jumps, the presence of anti-correlated co-jumps will, as was shown by virtue of the decomposition illustrated in Figure 7, lead to smaller Brownian correlations.

Having made these points, it is now worth asking what would happen to the model estimates should one impose the restrictions that (1) the intensity of the common jumps is equal to zero and (2) the intensity of the independent jumps is equal to zero (Table 2, Column 2 and Column 3, respectively). Both specifications, implying presence of independent or common jumps *only*, represent important models in the literature (see, e.g., Eraker et al., 2003), the mixed case discussed in this paper combining aspects of the two. In the first case, i.e., independent jumps only, all functions and parameters are qualitatively and quantitatively similar to the unrestricted model with two interesting exceptions. First, the correlation between Brownian shocks has now

a more familiar magnitude. For volatilities between 0.5% and 2%, the estimated leverage values vary between -0.27 and -0.40. The absence of co-jumps justifies an increased Brownian leverage. Second, some discontinuous changes in prices and volatility are, in terms of their number, now attributed by the model to the independent jumps. In terms of point estimates, we find roughly 26 yearly price jumps (with small sizes) and 7 yearly volatility jumps. In the second case, i.e., perfectly correlated jump arrivals or co-jumps only, we expect not accounting for independent jumps to possibly attenuate the negative correlation between the common jump sizes. This would happen if some of the independent jumps were attributed to common discontinuous variation. Consistent with this observation, we find attenuation in the jump size correlation ($-.64$). This attenuation is, however, not accompanied by drastic changes in Brownian leverage since it is compensated by an increase in the likelihood of co-jumping.

8 Implications for option pricing

Stochastic volatility determines excess (as compared to the Gaussian case) kurtosis in the conditional (on volatility) distribution of returns. The excess kurtosis gives rise to symmetrically higher implied volatilities for strikes away from the current prices, i.e., for levels of moneyness away from the at-the-money level (volatility smiles). Similarly, the negative correlation between continuous (Brownian) shocks to returns and shocks to volatility leads to skewness in the conditional distribution of stocks returns and implied volatility surfaces that are convex and downward sloping (volatility smirks). Both effects are stronger for short and medium maturity options than for long maturity options for which conditional returns are closer to normal (see, e.g., Das and Sundaram, 1999, for a lucid discussion).

The presence of co-jumps, and the negative correlation between the co-jump sizes, yields an additional source of skewness in the conditional distribution of stock returns. As said, in a model with co-jumps, the standardized conditional covariance between shocks to prices and shocks to variance depends on the correlation between continuous shocks ρ and the standardized conditional covariance between discontinuous shocks $\frac{\lambda_{r,\sigma}(\rho_J \sigma_{JJ,r} \sigma_{JJ,\sigma} + \mu_{JJ,r}\mu_{JJ,\sigma})}{\sigma\Lambda}$. If both quantities are negative, there are two sources of skewness in the model. Because ρ and ρ_J are, as documented, negative, and of meaningful magnitude in absolute value, the model we propose features this double effect.

To appreciate the relative influence of ρ_J, Figure 8 provides a representation of the impact of the co-jump correlation on the implied volatility smirk. We consider 4 maturities: 5 , 15, 30, and 120 days. The price process is simulated using the model estimated in Section 7 with no independent jumps. The parameters are those in Table 2, Column 3. The only exceptions are ρ_J (set to either 0 or -1) and a drift properly-adjusted for the presence of co-jumps so as to ensure a risk-neutral drift equal to $r = 3\%$ (i.e., $\mu = r - \lambda_{r,\sigma}\mu_{JJ,r}$). We report implied volatility surfaces

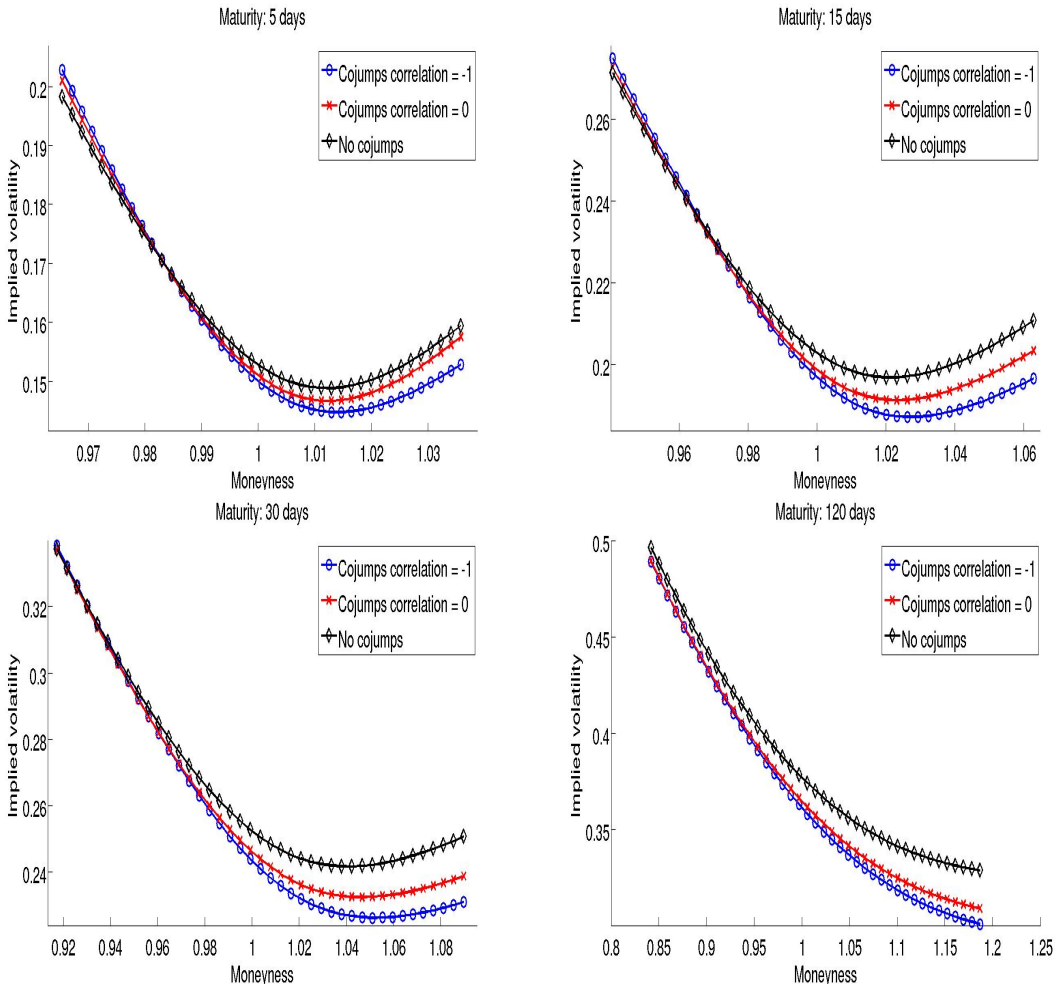

Figure 8: Implied volatility smirks with and without co-jumps.

as a function of moneyness. The surfaces correspond to the case with co-jumps and a correlation between jump sizes equal to -1, the case with co-jumps and a correlation between jump sizes equal to 0, and the case without co-jumps. When considering the case without co-jumps, we introduce independent jumps with the same intensity and size distribution as for the co-jump (only) case. This is to make the obtained volatility surfaces comparable in level.

For all maturities, given our parameters, introduction of (uncorrelated) co-jumps as well as increases in the correlation of the common jump sizes (from 0 to -1) lead to a rotation of the implied volatility smirk. This effect results in higher prices and higher implied volatilities, than in the case of uncorrelated common jump sizes, for in-the-money calls. Conversely, it results in lower implied volatilities for out-of-the-money calls. In essence, then, the presence of co-jumps, especially with a strong (negative) correlation between the common jump sizes, has the potential to add a considerable degree of flexibility to the pricing of options.

It is useful to emphasize that this outcome is not obtained, in our framework, by implying the degree of correlation between the common jumps on the basis of option data. In a calibration study designed to match the model's implied prices and the actual observed prices, Duffie et al. (2000) find a superior fit of the implied volatility smirk when calibrating a more negative correlation between jump sizes. They conclude that, based on their option data, the price and volatility jump sizes should be "nearly perfectly anti-correlated". Similarly, Eraker (2004) finds a statistically significant correlation between the jump sizes only when employing option data in addition to return data. In this sense our evidence may be viewed as complementing the existing option-based evidence. In a model that allows for independent and joint discontinuities, we show that anti-correlated amplitudes of the jumps are a fundamental property of prices and volatilities. The use of flexible identification methods, and a volatility filter relying on high-frequency data, are sufficient to reveal this property with no need for option prices.

9 Specification analysis

Much emphasis has been placed on affine structures and their usefulness in deriving near closed-form prices for a wide array of securities (see, e.g., Piazzesi, 2010, for a review). More generally, emphasis has been placed on parametric specifications in the presence of models with latent variables, such a stochastic volatility. In these models, the estimation of the parameters and the filtering of the latent states are a joint problem, one that has been successfully undertaken in several recent papers (see, for a recent discussion, Andersen and Benzoni, 2011). By its very nature, however, this joint problem is bound to link the filtering of volatility to the identification of the system's parametric structure.

We dispense with this link and identify daily volatility, *before* inference on the dynamics begins, by virtue of intra-daily price data. In this sense, our filtered volatility series is especially useful to select a function $\xi(.)$ and, hence, a variance transformation, which conforms with the assumed jump-diffusion structure and can therefore be modelled as a jump-diffusion process. Write

$$\varepsilon_{t,t+\Delta} = \frac{f_\lambda(\sigma^2_{t+\Delta}) - f_\lambda(\sigma^2_t) - m_\lambda(\sigma_t)\Delta}{\Lambda_\lambda(\sigma_t)\sqrt{\Delta}},$$

where $f_\lambda(\cdot)$ is a Box-Cox transformation, namely $f_\lambda(\cdot) = \frac{\cdot^\lambda - 1}{\lambda}$ for $\lambda \neq 0$ and $f_\lambda(\cdot) = \log(\cdot)$ for $\lambda = 0$. The functions $m_\lambda(\sigma_t)$ and $\Lambda_\lambda(\sigma_t)$ are the volatility drift and the volatility of volatility function associated with the same transformation.

Consider a specification without jumps. Because the system's shocks are assumed to be Brownian shocks, the (standardized) residuals $\varepsilon_{t,t+\Delta}$ should be locally Gaussian. The most suitable choice of λ is, therefore, the choice which guarantees Gaussianity of the residuals. In a model with symmetric Poisson jumps, the residual should still have zero skewness and only inflated kurtosis.

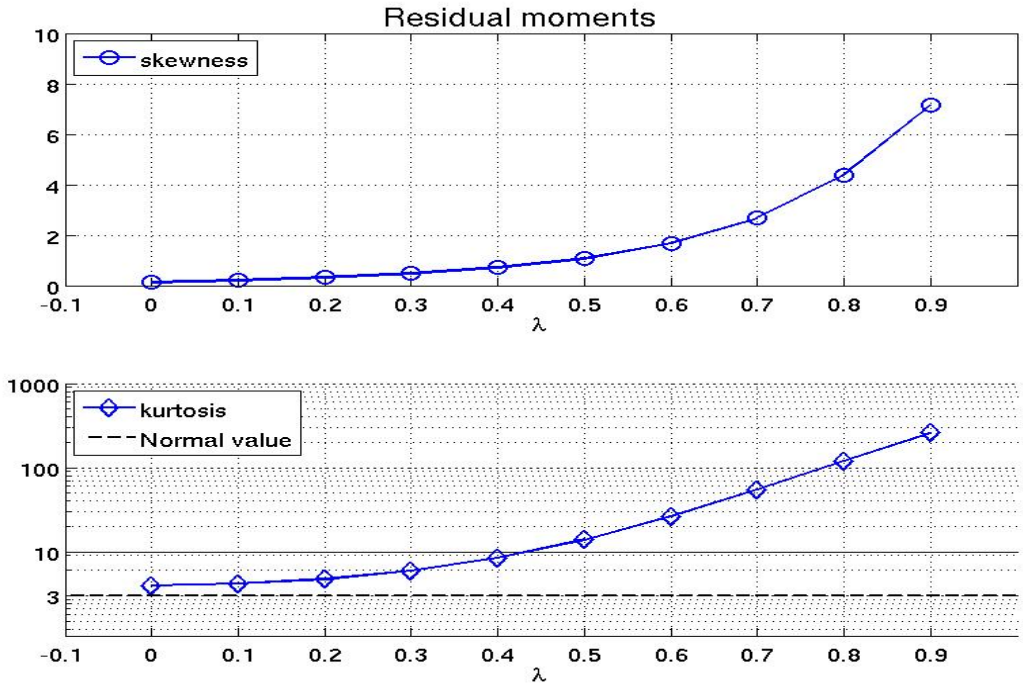

Figure 9: Specification test.

The excess kurtosis should be an increasing function of the number of jumps and the standard deviation of the jump size. To this extent, Figure 9 reports the skeweness and kurtosis of the volatility residuals for different values of the parameter λ. A value of λ equal to 0 yields zero skewness and, consistent with a model with some discontinuities, mildly inflated kurtosis. This value coincides with the logarithmic transformation which we adopt in this paper.[12] The second panel in Figure 3 shows the filtered logarithmic variances. It is visually clear that the daily logarithmic variances conform rather well with a model in which the Brownian shocks occur along with discontinuous shocks. In addition, the Brownian shocks appear to have a rather constant variance $\Lambda^2(\cdot)$ in our sample. Similarly, the volatility jumps seem to have a rather small standard deviation $\sigma^2_{J,\sigma}$. Both observations are consistent with the nonparametric and parametric analysis in Sections 5 and 6.

We now turn to specification testing on the functional estimates. In particular, we focus on co-jumps. The use of moment conditions for estimation nicely lends itself to tests of over-identifying restrictions. We begin with a visual assessment. We employ infinitesimal cross-moments which were not used for estimation, i.e., the moments below the diagonal in a matrix $p_1,\ p_2 \in (0, 4)$ with the exception of moment $3, 2$ and moment $2, 3$. We then compare the estimated parametric

[12]Papers on stochastic volatility estimation in discrete time which employ a logarithmic transformation are Chib et al. (2002); Harvey and Shephard (1996); Jacquier et al. (1994, 2004), and Yu (2005, 2010) among others.

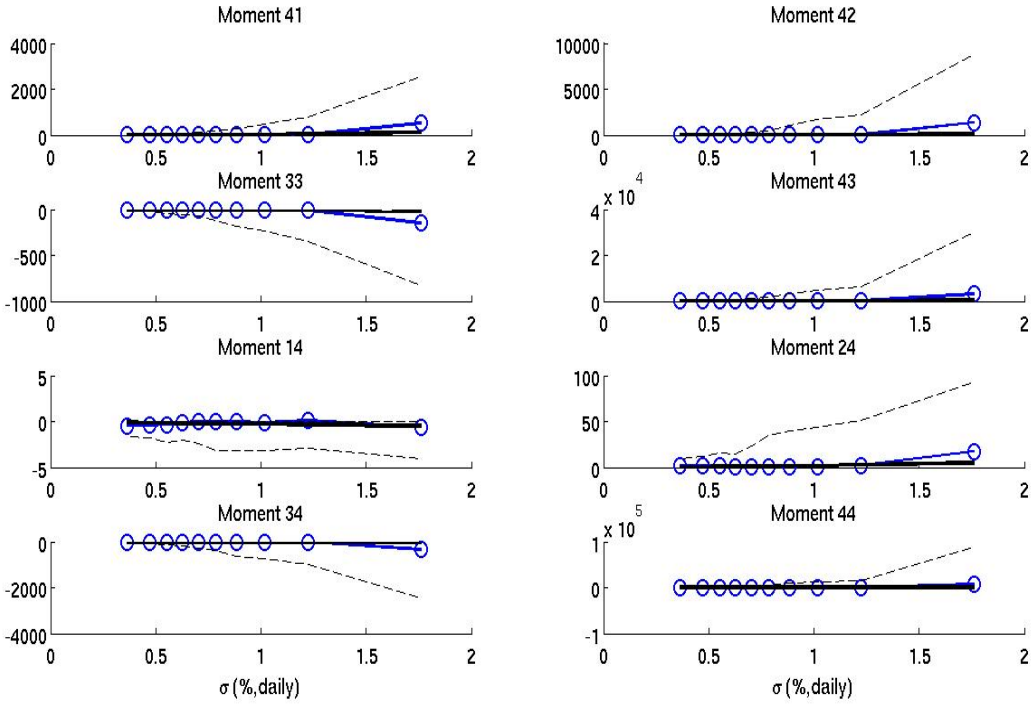

Figure 10: Overidentifying restrictions.

forms implied by the model to the estimated empirical moments. We find that the model-implied estimates are close to the "free" estimates, thereby providing visual evidence of satisfactory model specification.

We now turn to a more formal analysis. The (nonparametric and parametric) estimation of the model in Section 6 and 7 provides, as a straightforward by-product, a series of tests for the presence of co-jumps in our sample (the alternative hypothesis $\mathcal{H}_1 : \lambda_{r,\sigma} > 0$ being tested against the null of absence of co-jumps $\mathcal{H}_0 : \lambda_{r,\sigma} = 0$). We begin with $\mathcal{H}_0 :< \vartheta_{p_1,p_2} >= 0$ ($p_1 \geq 1$ and $p_2 \geq 1$ with $(p_1, p_2) \neq (1, 1)$), where $< \vartheta_{p_1,p_2} >$ denotes the estimated moment of order (p_1, p_2) averaged over the estimated volatility density, tested against $\mathcal{H}_1 :< \vartheta_{p_1,p_2} >\neq 0$. When p_1 and p_2 are both ≥ 1, and at least one of them is strictly larger than 1, theory dictates that these higher-order cross-moments should all be directly proportional to $\lambda_{r,\sigma}$. Specifically, $< \vartheta_{p_1,p_2} >= 0$ if $\lambda_{r,\sigma} = 0$. Conversely, a zero moment ($< \vartheta_{p_1,p_2} >= 0$) implies $\lambda_{r,\sigma} = 0$, if $\rho_J \neq 0$, since $\sigma_{JJ,r} > 0$ and $\sigma_{JJ,\sigma} > 0$. Somewhat more explicitly, we also employ a GMM-type $J-$test directly on the restriction $\lambda_{r,\sigma} = 0$. The distribution, and resulting size, of all the proposed tests under the null is evaluated using simulations from the parametric model in Table 2 after setting $\lambda_{r,\sigma} = 0$. Table 3 provides the test values along with their simulated p-values. All tests yield a clear rejection of the null of absence of co-jumps at the 1% level.

One final observation is in order. While the use of infinitesimal cross-moments for co-jump

test	value	p-value
J-test	239.4	0.20%
$< \vartheta_{2,2} >$	0.2862	0.10%
$< \vartheta_{1,2} >$	-0.0681	0.00%
$< \vartheta_{2,1} >$	0.2669	0.00%
$< \vartheta_{1,3} >$	-0.1091	0.00%
$< \vartheta_{3,1} >$	-0.9861	0.20%
$< \vartheta_{2,3} >$	0.4846	0.10%
$< \vartheta_{3,2} >$	-2.3185	0.10%
$< \vartheta_{3,3} >$	-6.2580	0.10%

Table 3: Tests for the presence of co-jumps and corresponding p-values.

identification is justified asymptotically in Appendix A, statistical uncertainty in the functional estimates, in the parametric estimates, and in the above tests is evaluated throughout this work by simulating the model using the estimated parameter values in Table 2 before running the described procedures for every simulated data set. In this sense, our inference is finite sample in nature, conservative, and unaffected by the overrated precision which would characterize asymptotic inference.

10 Final remarks

Market participants, as well as more casual observers, would believe that negative market shocks are associated with sudden, positive, changes in the level of volatility. This belief, confirmed by several recent financial crisis including the rather wild market gyrations of August 2011, originates from the frequent observation of the parallel, but opposite in sign, behavior of major stock indices and volatility indices, like the VIX. Negative price changes often occur along with positive spikes in the VIX. Just to mention a couple of recent events, between August 5th 2011 (a Friday) and August 8th 2011 (a Monday) the VIX rose from 32 to 48. The associated, negative, S&P 500 change on Monday, August 8th, was about 6%. Similarly, on August 17th the VIX rose from 31 to 42. The corresponding market price change was about -4.5%. Because it is an important barometer of market sentiment, the VIX is a complicated object capturing market fear, through risk-premia, as well as changes in fundamental values.

Interestingly, attempts to relate volatility measures unaffected by risk premia to sudden price changes have been largely unsuccessful. This outcome may be the result of pricing models which do not explicitly allow for independent price and volatility jumps along with co-jumps and, therefore, excessively constrain the jump dynamics. It may also be the result of volatility filtering methods relying on low-frequency return data, as well as on the parametric structure of the model, which may not be capable of yielding enough resolution, in terms of final estimates, so as to thoroughly identify rare events, like volatility jumps and co-jumps.

Combining jump-robust, high-frequency, spot variance estimates with a novel identification procedure based on infinitesimal cross-moments, we provide empirical evidence about the likelihood of price and volatility jumps occurring jointly. When co-jumping, price and volatility move in opposite directions, negative and positive respectively, and are strongly negatively correlated. We emphasize that this result does not hinge on sudden changes in risk premia associated with market downturns (as possibly yielded by the VIX) or implied volatility smirks (as given by cross-sectional option prices). Said differently, the effect is solely revealed by the dynamic properties of stock prices, once a sufficiently rich specification is adopted, without the need for the, arguably economically confounding, information contained in traded or synthetic derivatives.

11 Appendix A: Asymptotic properties

Consider a bivariate jump-diffusion process with compound Poisson jumps expressed as follows:

$$\begin{aligned} dX_t &= \mu_X(Y_t)dt + \sigma_X(Y_t)\left\{\rho(Y_t)dW_t^1 + \sqrt{1-\rho^2(Y_t)}dW_t^2\right\} + c_X dJ_t^X + d_X dJ_t, \\ dY_t &= \mu_Y(Y_t)dt + \sigma_Y(Y_t)dW_t^1 + c_Y dJ_t^Y + d_Y dJ_t, \end{aligned} \tag{14}$$

where $dW^X = \rho(Y_t)dW_t^1 + \sqrt{1-\rho^2(Y_t)}dW_t^2$, $dW^Y = dW^1$ are correlated diffusion processes with correlation coefficient $\rho(Y_t)$ and dJ^X, dJ^Y, and dJ are independent (of the Brownian motions W^1 and W^2, as well as each other) Poisson processes with intensities $\lambda_X(Y_t), \lambda_Y(Y_t)$,and $\lambda_{XY}(Y_t)$, respectively. The functions are such that a strong solution to the system exists. The system does not have to be stationary. It is sufficient for the state variable $\{Y_t, t \geq 0\}$ to be Harris recurrent, something which we assume.

We observe the skeleton $\left\{X_{\Delta_{n,T}}, Y_{\Delta_{n,T}}\right\}, \left\{X_{2\Delta_{n,T}}, Y_{2\Delta_{n,T}}\right\}, ..., \left\{X_{n\Delta_{n,T}}, Y_{n\Delta_{n,T}}\right\}$, i.e., n equally-spaced observations sampled at intervals $\Delta_{n,T} = T/n$. The asymptotic design is such that $n \to \infty$, $T \to \infty$, and $\Delta_{n,T} \to 0$. We estimate

$$\vartheta_{p_1,p_2}(y) = \lim_{\varepsilon \to 0} \frac{\mathrm{E}[(X_{t+\varepsilon} - X_t)^{p_1}(Y_{t+\varepsilon} - Y_t)^{p_2} | Y_t = y]}{\varepsilon}, \tag{15}$$

by using classical Nadaraya-Watson kernel estimates, namely

$$\widehat{\vartheta}_{p_1,p_2}(y) = \frac{\sum_{i=1}^{n-1} \mathbf{K}\left(\frac{Y_{iT/n}-y}{h_{n,T}}\right)\left(X_{(i+1)T/n} - X_{iT/n}\right)^{p_1}\left(Y_{(i+1)T/n} - Y_{iT/n}\right)^{p_2}}{\Delta_{n,T}\sum_{i=1}^{n} \mathbf{K}\left(\frac{Y_{iT/n}-y}{h_{n,T}}\right)}. \tag{16}$$

Define $\widehat{L}_{n,T}(y) = \frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n} \mathbf{K}\left(\frac{Y_{iT/n}-y}{h_{n,T}}\right) = \frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n} \mathbf{K}_{i,n,T}(y)$, the empirical occupation density of the Y process.

The following proofs provide details which are specific to inference for *infinitesimal cross-moments*. For more details on the general method of proof in this type of models, we refer the interested reader to the treatment in Bandi and Renò (2009). The notation in this Appendix is purposely for a general system of observable processes $\{(X_t, Y_t), t \geq 0\}$. The next subsection specializes the analysis to the case $(X_t, Y_t) = (\log p_t, \widehat{\sigma}_t^2)$ while providing limiting conditions for a vanishing measurement error in the variance estimates $\widehat{\sigma}_t^2$.

Assumption. *The function $\mathbf{K}(x)$ is a nonnegative, bounded, continuous, and symmetric kernel defined on a compact set* S *satisfying $\int_{\mathrm{S}} \mathbf{K}(s)ds = 1$, $\mathbf{K}_2 = \int_{\mathrm{S}} \mathbf{K}^2(s)ds < \infty$, and $\mathbf{K}_1 = \int_{\mathrm{S}} s^2\mathbf{K}(s)ds < \infty$. The kernel's first and second derivative are absolutely integrable.*

Theorem 1. **(Consistency.)** *If $n, T \to \infty$ and $\Delta_{n,T} = T/n \to 0$ so that $h_{n,T}\widehat{L}_{n,T}(y) \xrightarrow{a.s.} \infty$ and $\frac{\Delta_{n,T}}{h_{n,T}^2} \to 0$, then*

$$\widehat{\vartheta}_{p_1,0}(y) \xrightarrow{p} \begin{cases} \mu_X(y) + \lambda_X(y)\mathrm{E}[c_X] + \lambda_{XY}(y)\mathrm{E}[d_X] & p_1 = 1 \\ \sigma_X^2(y) + \lambda_X(y)\mathrm{E}[c_X^2] + \lambda_{XY}(y)\mathrm{E}[d_X^2] & p_1 = 2 \\ \lambda_X(y)\mathrm{E}[c_X^{p_1}] + \lambda_{XY}(y)\mathrm{E}[d_X^{p_1}] & p_1 \geq 3 \end{cases},$$

$$\widehat{\vartheta}_{1,1}(y) \xrightarrow{p} \rho(y)\sigma_X(y)\sigma_Y(y) + \lambda_{XY}(y)\mathrm{E}[d_X d_Y],$$

and, without loss of generality, for $p_1 \geq p_2 \geq 1$ (with $p_1 > p_2$ if $p_2 = 1$),

$$\widehat{\vartheta}_{p_1,p_2}(y) \xrightarrow{p} \lambda_{XY}(y)\mathrm{E}[d_X^{p_1} d_Y^{p_2}].$$

Theorem 2. **(Weak convergence.)** *Let $n, T \to \infty$ and $\Delta_{n,T} = T/n \to 0$ so that $h_{n,T}\widehat{L}_{n,T}(y) \xrightarrow{a.s.} \infty$ and $\frac{\Delta_{n,T}\sqrt{\widehat{L}_{n,T}(y)}}{h_{n,T}^{3/2}} \xrightarrow{a.s.} 0$. If*

$$h_{n,T}^5 \widehat{L}_{n,T}(y) = O_{a.s.}(1),$$

then

$$\sqrt{h_{n,T}\widehat{L}_{n,T}(y)}\left\{\widehat{\vartheta}_{p_1,p_2}(y) - \vartheta_{p_1,p_2}(y) - \Gamma_{\vartheta_{p_1,p_2}}(y)\right\} \Rightarrow \mathbf{N}(0, \mathbf{K}_2 \vartheta_{2p_1,2p_2}(y)),$$

with

$$\Gamma_{\vartheta_{p_1,p_2}} = h_{n,T}^2 \mathbf{K}_1 \left(\frac{\partial \vartheta_{p_1,p_2}(y)}{\partial y} \frac{\frac{\partial s(y)}{\partial y}}{s(y)} + \frac{1}{2} \frac{\partial^2 \vartheta_{p_1,p_2}(y)}{\partial^2 y} \right),$$

where $s(\mathrm{d}x) = s(x)\mathrm{d}x$ is the invariant measure of the Y process.

Lemma A.1. *Given a Borel measurable function $g(y) : \mathcal{R} \to \mathcal{R}$, write*

$$\chi_{n_1,n_2}(y) = \frac{\sum_{i=1}^{n-1} \mathbf{K}_{i,n,T}(y) \int_{iT/n}^{(i+1)T/n} \left(X_{s-} - X_{iT/n}\right)^{n_1} \left(Y_{s-} - Y_{iT/n}\right)^{n_2} g(Y_s) ds}{\Delta_{n,T} \sum_{i=1}^{n} \mathbf{K}_{i,n,T}(y)}$$

with $n_1 = 0, 1, \ldots$ and $n_2 = 0, 1, \ldots$. Assume $n, T \to \infty$ with $\Delta_{n,T} \to 0$. If $\frac{\Delta_{n,T}}{h_{n,T}^2} \to 0$, then

$$\chi_{0,0}(y) \xrightarrow{p} g(y),$$

and, if $n_1 \neq 0$ or $n_2 \neq 0$,

$$\chi_{n_1,n_2}(y) \xrightarrow{p} 0.$$

Proof of Lemma A.1. Using the ratio-limit theorem(Revuz and Yor, 1994) as in Bandi and Renò (2009), and handling discretization the way they do, we have

$$\chi_{0,0}(y) = \frac{\frac{\int_0^T \mathbf{K}_s(y) g(Y_s)\mathrm{d}s}{\int_0^T \mathbf{K}_s(y)\mathrm{d}s} + O_p\left(\frac{\Delta_{n,T}}{h_{n,T}^2}\right)}{\frac{\int_0^T \mathbf{K}_s(y)\mathrm{d}s}{\int_0^T \mathbf{K}_s(y)\mathrm{d}s} + O_p\left(\frac{\Delta_{n,T}}{h_{n,T}^2}\right)} \xrightarrow{p} g(y),$$

if $\frac{\Delta_{n,T}}{h_{n,T}^2} \to 0$. And, immediately,

$$\chi_{n_1,n_2}(y) = O_p\left(\Delta_{n,T}^{1/2}\right)^{n_1+n_2} \chi_{0,0}(y) \xrightarrow{p} 0.$$

Lemma A.2. *Given two Borel measurable functions $g_1(y) : \mathcal{R} \to \mathcal{R}$ and $g_2(y, z) : \mathcal{R}^2 \to \mathcal{R}$, write*

$$\Psi_{n_1,n_2}(y) = \frac{\sum_{i=1}^{n-1} \mathbf{K}_{i,n,T}(y) \int_{iT/n}^{(i+1)T/n} \left(X_{s-} - X_{iT/n}\right)^{n_1} \left(Y_{s-} - Y_{iT/n}\right)^{n_2} g_1(Y_s)\mathrm{d}W_s}{\Delta_{n,T} \sum_{i=1}^{n} \mathbf{K}_{i,n,T}(y)},$$

and

$$\Xi_{n_1,n_2}(y) = \frac{\sum_{i=1}^{n-1} \mathbf{K}_{i,n,T}(y) \int_{iT/n}^{(i+1)T/n} \left(X_{s-} - X_{iT/n}\right)^{n_1} \left(Y_{s-} - Y_{iT/n}\right)^{n_2} \int_Z g_2(Y_s, z)\bar{\nu}(\mathrm{d}s, \mathrm{d}z)}{\Delta_{n,T} \sum_{i=1}^{n} \mathbf{K}_{i,n,T}(y)},$$

with $n_1 = 0, 1, \ldots$ *and* $n_2 = 0, 1, \ldots$. *Assume* $n, T \to \infty$ *with* $\Delta_{n,T} \to 0$. *If* $\frac{\Delta_{n,T}}{h_{n,T}^2} \to 0$ *and* $h_{n,T}\widehat{L}_{n,T}(y) \xrightarrow{a.s.} \infty$, *then*

$$\Psi_{n_1,n_2}(y) \xrightarrow{p} 0,$$

and

$$\Xi_{n_1,n_2}(y) \xrightarrow{p} 0.$$

Proof of Lemma A.2. Both $\Psi_{0,0}(y)$ and $\Xi_{0,0}(y)$ are appropriately re-scaled (by $\widehat{L}_{n,T}(y)$) sums of martingale difference sequences. Following Bandi and Renò (2009), if $\frac{\Delta_{n,T}}{h_{n,T}^2} \to 0$ and $h_{n,T}\widehat{L}_{n,T}(y) \xrightarrow{a.s.} \infty$,

$$\Psi_{0,0}(y) \xrightarrow{p} 0,$$

$$\Xi_{0,0}(y) \xrightarrow{p} 0.$$

Thus, immediately,

$$\Psi_{n_1,n_2}(y) = O_p\left(\Delta_{n,T}^{1/2}\right)^{n_1+n_2} \Psi_{0,0}(y) \xrightarrow{p} 0,$$

and

$$\Xi_{n_1,n_2}(y) = O_p\left(\Delta_{n,T}^{1/2}\right)^{n_1+n_2} \Xi_{0,0}(y) \xrightarrow{p} 0.$$

Lemma A.3. *Consider* Ψ_{n_1,n_2} *and* Ξ_{n_1,n_2} *as defined in Lemma A.2. Let* $n, T \to \infty$ *so that* $\Delta_{n,T} = T/n \to 0$. *If* $h_{n,T}\widehat{L}_{n,T}(y) \xrightarrow{a.s.} \infty$ *and* $\frac{\Delta_{n,T}}{h_{n,T}^2} \to 0$, *we have*

$$\sqrt{h_{n,T}\widehat{L}_{n,T}(y)}\Psi_{0,0} \Rightarrow \mathbf{N}\left(0, \mathbf{K}_2 g_1^2(y)\right).$$

For $n_1 \geq 0$, *if* $h_{n,T}\widehat{L}_{n,T}(y) \xrightarrow{a.s.} \infty$ *and* $\frac{\Delta_{n,T}}{h_{n,T}^2} \to 0$,

$$\sqrt{\frac{h_{n,T}\widehat{L}_{n,T}(y)}{\Delta_{n,T}}}\Psi_{n_1,0} \Rightarrow \mathbf{N}\left(0, \frac{1}{2}\mathbf{K}_2 g_1^2(y)\vartheta_{2n_1,0}(y)\right).$$

For $n_1 \geq n_2 \geq 1$, *if* $h_{n,T}\widehat{L}_{n,T}(y) \xrightarrow{a.s.} \infty$ *and* $\frac{\Delta_{n,T}}{h_{n,T}^2} \to 0$,

$$\sqrt{\frac{h_{n,T}\widehat{L}_{n,T}(y)}{\Delta_{n,T}}}\Psi_{n_1,n_2} \Rightarrow \mathbf{N}\left(0, \frac{1}{2}\mathbf{K}_2 g_1^2(y)\vartheta_{2n_1,2n_2}(y)\right).$$

Similar expressions apply to Ξ_{n_1,n_2} *with* $\mathrm{E}[g_2^2(y,z)]$ *replacing* $g_1^2(y)$.

Proof of Lemma A.3. Without loss of generality, write $n_1 \geq n_2 \geq 1$ and consider Ξ_{n_1,n_2}. Write

$$\begin{aligned}
\Xi_{n_1,n_2}^{num}(y) &:= \frac{1}{\sqrt{h_{n,T}\Delta_{n,T}}} \sum_{i=1}^{n-1} \mathbf{K}_{i,n,T}(y) \int_{iT/n}^{(i+1)T/n} \left(X_{s-} - X_{iT/n}\right)^{n_1} \left(Y_{s-} - Y_{iT/n}\right)^{n_2} \int_Z g_2(Y_s)\bar{\nu}(\mathrm{d}s, \mathrm{d}z) \\
&:= \frac{1}{\sqrt{\Delta_{n,T}}} \sum_{i=1}^{n-1} u_i.
\end{aligned}$$

We immediately have $\frac{1}{\sqrt{\Delta_{n,T}}}\sum_{i=1}^{n-1}\mathrm{E}[u_i|\Im_{iT/n}]=0$. Using arguments in Bandi and Renò (2009), we obtain

$$\begin{aligned}\frac{1}{\Delta_{n,T}}\sum_{i=1}^{n-1}\mathrm{E}[u_i^2|\Im_{iT/n}] &= \frac{1}{2}\int_0^T \mathbf{K}_s^2(y)\mathrm{E}[g_2^2]\lambda_{XY}(Y_s)\mathrm{E}[d_X^{2n_1}d_Y^{2n_2}]\mathrm{d}s+\frac{\Delta_{n,T}}{h_{n,T}^2}O_p\left(\frac{1}{h_{n,T}}\int_0^T\mathbf{K}\left(\frac{Y_s-x}{h_{n,T}}\right)\mathrm{d}s\right)\\ &= \mathbf{U}_{n,T}^2,\end{aligned}$$

and

$$\frac{1}{\Delta_{n,T}}\sum_{i=1}^{n-1}\mathrm{E}\left[u_i^2\mathbf{1}(|u_i|>\epsilon)|\Im_{iT/n}\right]\xrightarrow{p}0.$$

Theorem VIII. 3.33 in Jacod and Shiryaev (2003) allows us to conclude that, uniformly in T and $h_{n,T}$ in $\mathcal{H}(\varepsilon)=\left\{\frac{\Delta_{n,T}^{1/2}}{\varepsilon}<h_{n,T}<\varepsilon\right\}$ for a small $\varepsilon>0$, $\Xi_{n_1,n_2}^{num}(y)\underset{\Delta_{n,T}\to 0}{\Rightarrow}W_{\mathbf{U}^2}$. Hence,

$$\sqrt{\frac{h_{n,T}\widehat{L}_{n,T}(y)}{\Delta_{n,T}}}\Xi_{n_1,n_2}(y)=\frac{\Xi_{n_1,n_2}^{num}(y)}{\sqrt{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^n\mathbf{K}_{i,n,T}(y)}}\Rightarrow W_{\frac{\mathbf{U}_{n,T}^2}{\int_0^T K_s(y)\mathrm{d}s}},$$

if $\frac{\Delta_{n,T}}{h_{n,T}^2}\to 0$. The ratio-limit theorem now yields

$$\frac{\mathbf{U}_{n,T}^2}{\int_0^T\mathbf{K}_s(y)\mathrm{d}s}\underset{T\to\infty,h_{n,T}\to 0,\Delta_{n,T}\to 0}{\xrightarrow{p}}\frac{1}{2}\mathbf{K}_2\mathrm{E}[g_2^2(y)]\lambda_{X,Y}(y)\mathrm{E}[d_X^{2n_1}d_Y^{2n_2}].$$

Theorem 4.1 in van Zanten (2000) implies the desired result. The case $n_2=0$ can be treated in a similar way. Next, consider Ψ_{n_1,n_2} and write,

$$\Psi_{n_1,n_2}^{num}(y)=\frac{1}{\sqrt{h_{n,T}\Delta_{n,T}}}\sum_{i=1}^{n-1}\mathbf{K}_{i,n,T}(y)\int_{iT/n}^{(i+1)T/n}\left(X_{s-}-X_{iT/n}\right)^{n_1}\left(Y_{s-}-Y_{iT/n}\right)^{n_2}g_1(Y_s)\mathrm{d}W_s.$$

Compute now its quadratic variation, i.e.,

$$[\Psi_{n_1,n_2}^{num}(y)]=\frac{1}{h_{n,T}\Delta_{n,T}}\sum_{i=1}^{n-1}\left(\mathbf{K}_{i,n,T}(y)\right)^2\int_{iT/n}^{(i+1)T/n}\left(X_{s-}-X_{iT/n}\right)^{2n_1}\left(Y_{s-}-Y_{iT/n}\right)^{2n_2}g_1^2(Y_s)\mathrm{d}s.$$

Using Ito's lemma, we obtain

$$
\begin{aligned}
& \left(X_{s-}-X_{iT/n}\right)^{2n_1}\left(Y_{s-}-Y_{iT/n}\right)^{2n_2} \\
= & \int_{iT/n}^{s} 2n_1\left(X_{u-}-X_{iT/n}\right)^{2n_1-1}\left(Y_{u-}-Y_{iT/n}\right)^{2n_2}\mu_{X,s}\mathrm{d}u \\
+ & \int_{iT/n}^{s} 2n_2\left(X_{u-}-X_{iT/n}\right)^{2n_1}\left(Y_{u-}-Y_{iT/n}\right)^{2n_2-1}\mu_{Y,s}\mathrm{d}u \\
+ & \int_{iT/n}^{s} 2n_1\left(X_{u-}-X_{iT/n}\right)^{2n_1-1}\left(Y_{u-}-Y_{iT/n}\right)^{2n_2}\sigma_{X,s}\mathrm{d}W_u^X \\
+ & \int_{iT/n}^{s} 2n_2\left(X_{u-}-X_{iT/n}\right)^{2n_1}\left(Y_{u-}-Y_{iT/n}\right)^{2n_2-1}\sigma_{Y,s}\mathrm{d}W_u^Y \\
+ & \int_{iT/n}^{s} \frac{1}{2}2n_1(2n_1-1)\left(X_{u-}-X_{iT/n}\right)^{2n_1-2}\left(Y_{u-}-Y_{iT/n}\right)^{2n_2}\sigma_{X,s}^2\mathrm{d}u \\
+ & \int_{iT/n}^{s} \frac{1}{2}2n_2(2n_2-1)\left(X_{u-}-X_{iT/n}\right)^{2n_1}\left(Y_{u-}-Y_{iT/n}\right)^{2n_2-2}\sigma_{Y,s}^2\mathrm{d}u \\
+ & \int_{iT/n}^{s} 2n_1 2n_2\left(X_{u-}-X_{iT/n}\right)^{2n_1-1}\left(Y_{u-}-Y_{iT/n}\right)^{2n_2-1}\rho_s\sigma_{X,s}\sigma_{Y,s}\mathrm{d}u \\
+ & \sum_{\Delta X_u\neq 0 \text{ or } \Delta Y_u\neq 0}\Big[(X_{u-}+\Delta X_u-X_{iT/n})^{2n_1}(Y_{u-}+\Delta Y_u-Y_{iT/n})^{2n_2} \\
& -\left(X_{u-}-X_{iT/n}\right)^{2n_1}\left(Y_{u-}-Y_{iT/n}\right)^{2n_2}\Big]. \qquad (17)
\end{aligned}
$$

If $n_1 \geq n_2 \geq 1$, the higher-order term above is given by $\sum(\Delta X_u)^{2n_1}(\Delta Y_u)^{2n_2}$. Thus,

$$
[\Psi_{n_1,n_2}^{num}(y)] \overset{a.s.}{\to} \frac{1}{2}\frac{1}{h_{n,T}}\int_0^T \mathbf{K}_s^2(y)g_1^2(Y_s)\lambda_{XY}(Y_s)\mathrm{E}\left(d_X^{2n_1}(z)d_Y^{2n_2}(z)\right)\mathrm{d}s
$$

Reasoning in the same way as for Ξ_{n_1,n_2}, we have

$$
\sqrt{\frac{h_{n,T}\widehat{L}_{n,T}(y)}{\Delta_{n,T}}}\Psi_{n_1,n_2} \Rightarrow \mathbf{N}\left(0,\frac{1}{2}\mathbf{K}_2 g_1^2(y)\vartheta_{2n_1,2n_2}\right),
$$

if $h_{n,T}\widehat{L}_{n,T}(y) \overset{a.s.}{\to} \infty$ and $\frac{\Delta_{n,T}}{h_{n,T}^2} \to 0$. Clearly, for $n_1 \geq 0$,

$$
\sqrt{\frac{h_{n,T}\widehat{L}_{n,T}(y)}{\Delta_{n,T}}}\Psi_{n_1,0} \Rightarrow \mathbf{N}\left(0,\frac{1}{2}\mathbf{K}_2 g_1^2(y)\vartheta_{2n_1,0}\right),
$$

if $h_{n,T}\widehat{L}_{n,T}(y) \overset{a.s.}{\to} \infty$ and $\frac{\Delta_{n,T}}{h_{n,T}^2} \to 0$.

Proof of Theorem 1. Consider the $p_2 = 0$, $p_1 \geq 3$ case. Ito's Lemma gives

$$
\begin{aligned}
(X_{(i+1)T/n}-X_{iT/n})^{p_1} = & \; p_1\int_{iT/n}^{(i+1)T/n}\left(X_{s-}-X_{iT/n}\right)^{p_1-1}\mu_{X,s}\mathrm{d}s \\
& + p_1\int_{iT/n}^{(i+1)T/n}\left(X_{s-}-X_{iT/n}\right)^{p_1-1}\sigma_{X,s}\mathrm{d}W_s^X \\
& + \frac{1}{2}p_1(p_1-1)\int_{iT/n}^{(i+1)T/n}\left(X_{s-}-X_{iT/n}\right)^{p_1-2}\sigma_{X,s}^2\mathrm{d}s \\
& + \sum_{\Delta X_s\neq 0}\left[\left(X_{s-}+\Delta X_s-X_{iT/n}\right)^{p_1}-\left(X_{s-}-X_{iT/n}\right)^{p_1}\right]. \qquad (18)
\end{aligned}
$$

Now, write

$$\begin{aligned}
&\sum_{\Delta X_s\neq 0}\left[\left(X_{s-}+\Delta X_s-X_{iT/n}\right)^{p_1}-\left(X_{s-}-X_{iT/n}\right)^{p_1}\right]\\
&=\sum_{k=0}^{p_1-1}\left(\begin{array}{c}p_1\\k\end{array}\right)\int_{iT/n}^{(i+1)T/n}\int_Z\left(X_{s-}-X_{iT/n}\right)^k\left(c_X^{p_1-k}(z)\nu_X(\mathrm{d}s,\mathrm{d}z)+d_X^{p_1-k}(z)\nu_{XY}(\mathrm{d}s,\mathrm{d}z)\right).
\end{aligned}$$

Compensating the random jump measures and denoting by $\bar{\nu}_X$ and $\bar{\nu}_{XY}$ the corresponding compensated random jump measures, we write $\bar{\nu}_X(\mathrm{d}s,\mathrm{d}y)=\nu_X(\mathrm{d}s,\mathrm{d}y)-\lambda_{X,s}\mathrm{E}[c_X^{p_1-k}]\mathrm{d}s$ and $\bar{\nu}_{XY}(\mathrm{d}s,\mathrm{d}y)=\nu_{XY}(\mathrm{d}s,\mathrm{d}y)-\lambda_{XY,s}\mathrm{E}[d_X^{p_1-k}]\mathrm{d}s$. Thus,

$$\begin{aligned}
\widehat{\vartheta}_{p_1,0}(y) &= \frac{\sum_{i=1}^{n-1}p_1\mathbf{K}_{i,n,T}(y)\int_{iT/n}^{(i+1)T/n}\left(X_{s-}-X_{iT/n}\right)^{p_1-1}\mu_{X,s}\mathrm{d}s}{\Delta_{n,T}\sum_{i=1}^{n}\mathbf{K}_{i,n,T}(y)}\\
&+\frac{\sum_{i=1}^{n-1}p_1\mathbf{K}_{i,n,T}(y)\int_{iT/n}^{(i+1)T/n}\left(X_{s-}-X_{iT/n}\right)^{p_1-1}\sigma_{X,s}\mathrm{d}W_s^X}{\Delta_{n,T}\sum_{i=1}^{n}\mathbf{K}_{i,n,T}(y)}\\
&+\frac{\sum_{i=1}^{n-1}\frac{1}{2}p_1(p_1-1)\mathbf{K}_{i,n,T}(y)\int_{iT/n}^{(i+1)T/n}\left(X_{s-}-X_{iT/n}\right)^{p_1-2}\sigma_{X,s}^2\mathrm{d}s}{\Delta_{n,T}\sum_{i=1}^{n}\mathbf{K}_{i,n,T}(y)}\\
&+\frac{\sum_{i=1}^{n-1}\mathbf{K}_{i,n,T}(y)\sum_{k=0}^{p_1-1}\left(\begin{array}{c}p_1\\k\end{array}\right)\int_{iT/n}^{(i+1)T/n}\left(X_{s-}-X_{iT/n}\right)^k\int_Y\left(c_X^{p_1-k}(y)\bar{\nu}_X(\mathrm{d}s,\mathrm{d}y)+d_X^{p_1-k}(y)\bar{\nu}_{XY}(\mathrm{d}s,\mathrm{d}y)\right)}{\Delta_{n,T}\sum_{i=1}^{n}\mathbf{K}_{i,n,T}(y)}\\
&+\frac{\sum_{i=1}^{n-1}\mathbf{K}_{i,n,T}(y)\sum_{k=0}^{p_1-1}\left(\begin{array}{c}p_1\\k\end{array}\right)\int_{iT/n}^{(i+1)T/n}\left(X_{s-}-X_{iT/n}\right)^k\left(\lambda_{X,s}\mathrm{E}[c_X^{p_1-k}]+\lambda_{XY,s}\mathrm{E}[d_X^{p_1-k}]\right)ds}{\Delta_{n,T}\sum_{i=1}^{n}\mathbf{K}_{i,n,T}(y)}\\
&=\mathbf{R}_1+\mathbf{R}_2+\mathbf{R}_3+\mathbf{R}_4+\mathbf{A}_5.
\end{aligned}$$

The terms $\mathbf{R}_1$ through $\mathbf{R}_4$ converge to zero in probability given Lemma A.1. and Lemma A.2. Now write $\mathbf{A}_5$ as follows

$$\begin{aligned}
\mathbf{A}_5 &= \frac{\sum_{i=1}^{n-1}\int_{iT/n}^{(i+1)T/n}\mathbf{K}_{i,n,T}(y)\left(\lambda_{X,s}\mathbf{E}[c_X^{p_1}]+\lambda_{XY,s}\mathbf{E}[d_X^{p_1}]\right)\mathrm{d}s}{\Delta_{n,T}\sum_{i=1}^{n}\mathbf{K}_{i,n,T}(y)}\\
&+\frac{\sum_{i=1}^{n-1}\mathbf{K}_{i,n,T}(y)\sum_{k=1}^{p_1-1}\left(\begin{array}{c}p_1\\k\end{array}\right)\int_{iT/n}^{(i+1)T/n}\left(X_{s-}-X_{iT/n}\right)^k\left(\lambda_{X,s}\mathrm{E}[c_X^{p_1-k}]+\lambda_{XY,s}\mathrm{E}[d_X^{p_1-k}]\right)\mathrm{d}s}{\Delta_{n,T}\sum_{i=1}^{n}\mathbf{K}_{i,n,T}(y)}\\
&=\mathbf{A}_5^1+\mathbf{A}_5^2.
\end{aligned}$$

Lemma A.1. gives

$$\mathbf{A}_5^1 \xrightarrow{p} \lambda_X(y)\mathrm{E}[c_X^{p_1}] + \lambda_{XY}(y)\mathrm{E}[d_X^{p_1}],$$
$$\mathbf{A}_5^2 \xrightarrow{p} 0,$$

which proves the stated result. Given Eq. (18) the cases $p_1 = 1$ and $p_1 = 2$ are obvious. Consider the case $p_2 > p_1 \geq 1$, without loss of generality. Using Ito's lemma as in Eq. (17) and rewriting the last term as

$$\begin{aligned}
& \sum_{\Delta X_s \neq 0 \ or \ \Delta Y_s \neq 0} \left[(X_{s-} + \Delta X_s - X_{iT/n})^{p_1} (Y_{s-} + \Delta Y_s - Y_{iT/n})^{p_2} - \left(X_{s-} - X_{iT/n}\right)^{p_1} \left(Y_{s-} - Y_{iT/n}\right)^{p_2} \right] \\
= & \sum_{\Delta X_s \neq 0 \ or \ \Delta Y_s \neq 0} \left(\sum_{k=0}^{p_1-1} \left(\begin{array}{c} p_1 \\ k \end{array} \right) \left(X_{s-} - X_{iT/n}\right)^k (\Delta X_s)^{p_1-k} \right) \left(\sum_{k=0}^{p_2-1} \left(\begin{array}{c} p_2 \\ k \end{array} \right) \left(Y_{s-} - Y_{iT/n}\right)^k (\Delta Y_s)^{p_2-k} \right) \\
& - \sum_{\Delta X_s \neq 0 \ or \ \Delta Y_s \neq 0} \left(X_{s-} - X_{iT/n}\right)^{p_1} \left(Y_{s-} - Y_{iT/n}\right)^{p_2} \\
= & \sum_{\Delta X_s \Delta Y_s \neq 0} (\Delta X_s)^{p_1} (\Delta Y_s)^{p_2} + \sum_{\Delta X_s \neq 0 \ or \ \Delta Y_s \neq 0} \left(\left(X_{s-} - X_{iT/n}\right) f_{1,i} + \left(Y_{s-} - Y_{iT/n}\right) f_{2,i} \right) \\
= & \int_{iT/n}^{(i+1)T/n} \int_Y d_X^{p_1}(z) d_Y^{p_2}(z) \nu_{XY}(\mathrm{d}z, \mathrm{d}s) + \int_{iT/n}^{(i+1)T/n} \left(X_{s-} - X_{iT/n}\right) f_{1,s} + \left(Y_{s-} - Y_{iT/n}\right) f_{2,s} \mathrm{d}s
\end{aligned}$$

for suitable $f_{1,s}, f_{2,s}$ functions. Now, write

$$\begin{aligned}
\widehat{\vartheta}_{p_1,p_2}(y) = & \frac{\sum_{i=1}^{n-1} \mathbf{K}_{i,n,T}(y) \int_{iT/n}^{(i+1)T/n} \int_Y d_X^{p_1}(y) d_Y^{p_2}(y) \nu_{XY}(\mathrm{d}y, \mathrm{d}s)}{\Delta_{n,T} \sum_{i=1}^{n} \mathbf{K}_{i,n,T}(y)} \\
& + \frac{\sum_{i=1}^{n-1} \mathbf{K}_{i,n,T}(y) \int_{iT/n}^{(i+1)T/n} p_1 \left(X_{s-} - X_{iT/n}\right)^{p_1-1} \left(Y_{s-} - Y_{iT/n}\right)^{p_2} \mu_{X,s} \mathrm{d}s}{\Delta_{n,T} \sum_{i=1}^{n} \mathbf{K}_{i,n,T}(y)} \\
& + \frac{\sum_{i=1}^{n-1} \mathbf{K}_{i,n,T}(y) \int_{iT/n}^{(i+1)T/n} p_2 \left(X_{s-} - X_{iT/n}\right)^{p_1} \left(Y_{s-} - Y_{iT/n}\right)^{p_2-1} \mu_{Y,s} \mathrm{d}s}{\Delta_{n,T} \sum_{i=1}^{n} \mathbf{K}_{i,n,T}(y)} \\
& + \frac{\sum_{i=1}^{n-1} \mathbf{K}_{i,n,T}(y) \int_{iT/n}^{(i+1)T/n} p_1 \left(X_{s-} - X_{iT/n}\right)^{p_1-1} \left(Y_{s-} - Y_{iT/n}\right)^{p_2} \sigma_{X,s} \mathrm{d}W_s^X}{\Delta_{n,T} \sum_{i=1}^{n} \mathbf{K}_{i,n,T}(y)} \\
& + \frac{\sum_{i=1}^{n-1} \mathbf{K}_{i,n,T}(y) \int_{iT/n}^{(i+1)T/n} p_2 \left(X_{s-} - X_{iT/n}\right)^{p_1} \left(Y_{s-} - Y_{iT/n}\right)^{p_2-1} \sigma_{Y,s} \mathrm{d}W_s^Y}{\Delta_{n,T} \sum_{i=1}^{n-1} \mathbf{K}_{i,n,T}(y)} \\
& + \frac{\sum_{i=1}^{n-1} \mathbf{K}_{i,n,T}(y) \int_{iT/n}^{(i+1)T/n} \frac{1}{2} p_1(p_1-1) \left(X_{s-} - X_{iT/n}\right)^{p_1-2} \left(Y_{s-} - Y_{iT/n}\right)^{p_2} \sigma_{X,s}^2 \mathrm{d}s}{\Delta_{n,T} \sum_{i=1}^{n} \mathbf{K}_{i,n,T}(y)}
\end{aligned} \tag{19}$$

$$
\begin{aligned}
&+ \frac{\sum_{i=1}^{n-1} \mathbf{K}_{i,n,T}(y) \int_{iT/n}^{(i+1)T/n} \frac{1}{2} p_2(p_2-1) \left(X_{s-} - X_{iT/n}\right)^{p_1} \left(Y_{s-} - Y_{iT/n}\right)^{p_2-2} \sigma_{Y,s}^2 \mathrm{d}s}{\Delta_{n,T} \sum_{i=1}^{n} \mathbf{K}_{i,n,T}(y)} \\
&+ \frac{\sum_{i=1}^{n-1} \mathbf{K}_{i,n,T}(y) \int_{iT/n}^{(i+1)T/n} p_1 p_2 \left(X_{s-} - X_{iT/n}\right)^{p_1-1} \left(Y_{s-} - Y_{iT/n}\right)^{p_2-1} \rho_s \sigma_{X,s} \sigma_{Y,s} \mathrm{d}s}{\Delta_{n,T} \sum_{i=1}^{n} \mathbf{K}_{i,n,T}(y)} \\
&+ \frac{\sum_{i=1}^{n-1} \mathbf{K}_{i,n,T}(y) \int_{iT/n}^{(i+1)T/n} (\left(X_{s-} - X_{iT/n}\right) f_{1,i} + \left(Y_{s-} - Y_{iT/n}\right) f_{2,i}) \mathrm{d}s}{\Delta_{n,T} \sum_{i=1}^{n} \mathbf{K}_{i,n,T}(y)} \\
&= \frac{\sum_{i=1}^{n-1} \mathbf{K}_{i,n,T}(y) \int_{iT/n}^{(i+1)T/n} \int_Y d_X^{p_1}(y) d_Y^{p_2}(z) \nu_{XY}(\mathrm{d}z, \mathrm{d}s)}{\Delta_{n,T} \sum_{i=1}^{n} \mathbf{K}_{i,n,T}(y)} \\
&\quad + \mathbf{R}_1 + \mathbf{R}_2 + \mathbf{R}_3 + \mathbf{R}_4 + \mathbf{R}_5 + \mathbf{R}_6 + \mathbf{R}_7 + \mathbf{R}_8. \qquad (20)
\end{aligned}
$$

Finally, we compensate the random measure ν_{XY} and use Lemmas A.1. and A.2. to obtain

$$\mathbf{R}_1, \mathbf{R}_2, \mathbf{R}_3, \mathbf{R}_4, \mathbf{R}_5, \mathbf{R}_6, \mathbf{R}_7, \mathbf{R}_8 \xrightarrow{p} 0,$$

and

$$\frac{\sum_{i=1}^{n-1} \mathbf{K}_{i,n,T}(y) \int_{iT/n}^{(i+1)T/n} \int_Y d_X^{p_1}(z) d_Y^{p_2}(z) \nu_{XY}(\mathrm{d}z, \mathrm{d}s)}{\Delta_{n,T} \sum_{i=1}^{n} \mathbf{K}_{i,n,T}(y)} \xrightarrow{p} \lambda_{XY}(y) \mathrm{E}[d_X^{p_1} d_Y^{p_2}].$$

Given Eq. (19), the case $p_2 = p_1 = 1$ is obvious since

$$\mathbf{R}_7 \xrightarrow{p} \rho(y) \sigma_X(y) \sigma_Y(y).$$

Proof of Theorem 2. Consider the case $p_1 > p_1 \geq 1$. Use Ito's lemma as in Eq. (17) and re-write the jump term as

$$
\begin{aligned}
&\sum_{\Delta X_s \neq 0 \text{ or } \Delta Y_s \neq 0} \left[(X_{s-} + \Delta X_s - X_{iT/n})^{p_1} (Y_{s-} + \Delta Y_s - Y_{iT/n})^{p_2} - \left(X_{s-} - X_{iT/n}\right)^{p_1} \left(Y_{s-} - Y_{iT/n}\right)^{p_2} \right] \\
&= \int_{iT/n}^{(i+1)T/n} \int_Z d_{X,s}^{p_1}(z) d_{Y,s}^{p_2}(z) \nu_{XY}(\mathrm{d}s, \mathrm{d}z) + \int_{iT/n}^{(i+1)T/n} f_{3,i} \mathrm{d}s,
\end{aligned}
$$

for a suitable $f_{3,i} = o_p\left(\int_{iT/n}^{(i+1)T/n} \int_Z d_{X,s}^{p_1}(z) d_{Y,s}^{p_2}(z) \nu_{XY}(\mathrm{d}s, \mathrm{d}z)\right)$. Then, as in the consistency proof, after compensating the random measure ν_{XY}, write

$$
\begin{aligned}
&\widehat{\vartheta}_{p_1,p_2}(y) - \underbrace{\frac{\sum_{i=1}^{n-1} \mathbf{K}_{i,n,T}(y) \int_{iT/n}^{(i+1)T/n} \int_Z \mathrm{E}(d_X^{p_1}(z) d_Y^{p_2}(z)) \lambda_{XY,s} \mathrm{d}s \mathrm{d}z}{\Delta_{n,T} \sum_{i=1}^{n} \mathbf{K}_{i,n,T}(y)}}_{\mathbf{B}} \\
&:= \mathbf{A} + \mathbf{R}_1 + \mathbf{R}_2 + \mathbf{R}_3 + \mathbf{R}_4 + \mathbf{R}_5 + \mathbf{R}_6 + \mathbf{R}_7 + \mathbf{R}_8', \qquad (21)
\end{aligned}
$$

where

$$\mathbf{A} = \frac{\sum_{i=1}^{n-1} \mathbf{K}_{i,n,T}(y) \int_{iT/n}^{(i+1)T/n} \int_Z d_{X,s}^{p_1}(z) d_{Y,s}^{p_2}(z) \bar{\nu}_{XY}(\mathrm{d}s, \mathrm{d}z)}{\Delta_{n,T} \sum_{i=1}^{n} \mathbf{K}_{i,n,T}(y)},$$

and

$$\mathbf{R}_8' = \frac{\sum_{i=1}^{n-1} \mathbf{K}_{i,n,T}(y) \int_{iT/n}^{(i+1)T/n} f_{3,i}(s) ds}{\Delta_{n,T} \sum_{i=1}^{n} \mathbf{K}_{i,n,T}(y)}.$$

Now, using Lemma A.3., we have that

$$\sqrt{h_{n,T} \widehat{L}_{n,T}(y)} \mathbf{A} \Rightarrow \mathbf{N}\left(0, \mathbf{K}_2 \lambda_{XY}(y) \mathrm{E}[d_X^{2p_1} d_Y^{2p_2}]\right).$$

Naturally, $\mathbf{A}$ is the dominating term in the sum. Finally, note that the right-hand side of Eq. (21) can be expressed as

$$\widehat{\vartheta}_{p_1,p_2}(y) - \vartheta_{p_1,p_2}(y) - \underbrace{\left(\frac{\sum_{i=1}^{n-1} \mathbf{K}_{i,n,T}(y) \int_{iT/n}^{(i+1)T/n} \vartheta_{p_1,p_2}(Y_s) \mathrm{d}s}{\Delta_{n,T} \sum_{i=1}^{n} \mathbf{K}_{i,n,T}(y)} - \vartheta_{p_1,p_2}(y) \right)}_{\mathbf{B}'},$$

with

$$\mathbf{B}' = h_{n,T}^2 \left(\int_{\mathrm{S}} s^2 \mathbf{K}(s) ds \right) \left(\frac{\partial \vartheta_{p_1,p_2}(y)}{\partial x_i} \frac{\frac{\partial s(y)}{\partial x_i}}{s(y)} + \frac{1}{2} \frac{\partial^2 \vartheta_{p_1,p_2}(y)}{\partial y} \right) + O_p\left(\frac{\Delta_{n,T}}{h_{n,T}^2} \right).$$

This completes the proof.

11.1 Replacing spot variance with its estimated value

Rewrite the estimator in Eq. (11) as $\widehat{\sigma}_{t,i}^2 = \frac{TBPV_{t,i}}{\phi}$ with

$$TBPV_{t,i} = \varsigma_1^{-2} \sum_{j=2}^{k} |r_{t,i,j}||r_{t,i,j-1}| I_{\{|r_{t,i,j}| \leq \theta_{t,i,j}\}} I_{\{|r_{t,i,k-1}| \leq \theta_{t,i,j-1}\}}.$$

In our case ϕ is equal to 1 hour and $k = 60$ minutes. We show that, under assumptions, the spot variance estimator is consistent for spot variance and the resulting measurement error can be made asymptotically negligible.

Theorem 3. *We let $\phi \to 0$ with $k \to \infty$. Assume $\theta_{t,i} = \xi_{t,i} \Theta(\frac{\phi}{k})$, where $\Theta(\frac{\phi}{k})$ is a real function satisfying $\Theta\left(\frac{\phi}{k}\right) \underset{\phi \to 0,\ k \to \infty}{\to} 0$ and $\frac{1}{\Theta\left(\frac{\phi}{k}\right)} \left(\frac{\phi}{k} \log\left(\frac{\phi}{k}\right)\right) \underset{\phi \to 0,\ k \to \infty}{\to} 0$ and $\xi_{t,i}$ is an a.s. bounded process with a strictly positive lower bound. Write $\mathbf{\Psi}_{n,k,\phi} = \sqrt{\frac{\log(n)}{k}} + \sqrt{\phi}$. Consider $\widehat{\vartheta}_{p_1,p_2}(.)$ with $(p_1, p_2) = (1, 0)$ or $(0, 1)$. If*

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

the consistency result in Theorem 1 holds when replacing $\sigma_{iT/n}^2$ with $\widehat{\sigma}_{iT/n}^2$. For any other combination of (p_1, p_2), if

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

the consistency result in Theorem 1 holds when replacing $\sigma^2_{iT/n}$ *with* $\widehat{\sigma}^2_{iT/n}$. *Assume* $(p_1, p_2) = (1, 0)$ or $(0, 1)$, *if*

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

where $\widehat{L}_{n,T}(\sigma^2)$ *is the estimated occupation density of spot variance process, the weak convergence results in Theorem 2 holds when replacing* $\sigma^2_{iT/n}$ *with* $\widehat{\sigma}^2_{iT/n}$. *For any other combination of* (p_1, p_2), *if*

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

the weak convergence results in Theorem 2 holds when replacing $\sigma^2_{iT/n}$ *with* $\widehat{\sigma}^2_{iT/n}$.

Proof of Theorem 3. In what follows, $X = \log(p)$ and $Y = \widehat{\sigma}^2$. The transformation $\log(\widehat{\sigma}^2)$ can be treated similarly. Given the results in Bandi and Renò (2009, 2011), we only have to consider the case in which $p_1 \geq 1$ and $p_2 \geq 1$, with at least one being strictly larger than 1. Write

$$\begin{aligned}
&\widehat{\vartheta}_{p_1,p_2}(x) - \frac{\sum_{i=1}^{n-1} \mathbf{K}_i \left(\log(p_{(i+1)T/n}) - \log(p_{iT/n})\right)^{p_1} \left(\sigma^2_{(i+1)T/n} - \sigma^2_{iT/n}\right)^{p_2}}{\Delta_{n,T}\sum_{i=1}^{n}\mathbf{K}_i} \\
= &\frac{\sum_{i=1}^{n-1} \widehat{\mathbf{K}}_i \left(\log(p_{(i+1)T/n}) - \log(p_{iT/n})\right)^{p_1} (\widehat{\sigma}^2_{(i+1)T/n} - \widehat{\sigma}^2_{iT/n})^{p_2}}{\Delta_{n,T}\sum_{i=1}^{n}\widehat{\mathbf{K}}_i} - \frac{\sum_{i=1}^{n-1} \mathbf{K}_i \left(\log(p_{(i+1)T/n}) - \log(p_{iT/n})\right)^{p_1} (\widehat{\sigma}^2_{(i+1)T/n} - \widehat{\sigma}^2_{iT/n})^{p_2}}{\Delta_{n,T}\sum_{i=1}^{n}\widehat{\mathbf{K}}_i} \\
+ &\frac{\sum_{i=1}^{n-1} \mathbf{K}_i \left(\log(p_{(i+1)T/n}) - \log(p_{iT/n})\right)^{p_1} (\widehat{\sigma}^2_{(i+1)T/n} - \widehat{\sigma}^2_{iT/n})^{p_2}}{\Delta_{n,T}\sum_{i=1}^{n}\widehat{\mathbf{K}}_i} - \frac{\sum_{i=1}^{n-1} \mathbf{K}_i \left(\log(p_{(i+1)T/n}) - \log(p_{iT/n})\right)^{p_1} (\sigma^2_{(i+1)T/n} - \sigma^2_{iT/n})^{p_2}}{\Delta_{n,T}\sum_{i=1}^{n}\widehat{\mathbf{K}}_i} \\
+ &\frac{\sum_{i=1}^{n-1} \mathbf{K}_i \left(\log(p_{(i+1)T/n}) - \log(p_{iT/n})\right)^{p_1} (\sigma^2_{(i+1)T/n} - \sigma^2_{iT/n})^{p_2}}{\Delta_{n,T}\sum_{i=1}^{n}\widehat{\mathbf{K}}_i} - \frac{\sum_{i=1}^{n-1} \mathbf{K}_i \left(\log(p_{(i+1)T/n}) - \log(p_{iT/n})\right)^{p_1} (\sigma^2_{(i+1)T/n} - \sigma^2_{iT/n})^{p_2}}{\Delta_{n,T}\sum_{i=1}^{n}\mathbf{K}_i} \\
= &(a) + (b) + (c).
\end{aligned}$$

In the presence of jumps, we have

$$\left(\log(p_{(i+1)T/n}) - \log(p_{iT/n})\right)^{p_1} = O_p(\Delta_{n,T}^{1/2}),$$

and

$$(\sigma^2_{(i+1)T/n} - \sigma^2_{iT/n})^{p_2} = O_p(\Delta_{n,T}^{1/2}).$$

In the presence of co-jumps, a simple application of Ito's lemma yields

$$\left(\log(p_{(i+1)T/n}) - \log(p_{iT/n})\right)^{p_1} (\sigma^2_{(i+1)T/n} - \sigma^2_{iT/n})^{p_2} = O_p(\Delta_{n,T}^{1/2}).$$

Also write $\mathbf{M}_{n,k,T} = \max_{1\leq i\leq n}\left|\widehat{\sigma}^2_{iT/n} - \sigma^2_{iT/n}\right| = O_p(\mathbf{\Psi}_{n,k,\phi})$ as shown in Bandi and Renò (2009, 2011). Thus

$$(a) = \frac{\sum_{i=1}^{n-1}\left(\widehat{\mathbf{K}}_i - \mathbf{K}_i\right)\left(\log(p_{(i+1)T/n}) - \log(p_{iT/n})\right)^{p_1} (\widehat{\sigma}^2_{(i+1)T/n} - \sigma^2_{(i+1)T/n} + \sigma^2_{(i+1)T/n} - \sigma^2_{iT/n} + \sigma^2_{iT/n} - \widehat{\sigma}^2_{iT/n})^{p_2}}{\Delta_{n,T}\sum_{i=1}^{n}\widehat{\mathbf{K}}_i}.$$

Now, for suitable constants $c_{i,j,k}$,

$$\begin{aligned}
&\left(\log(p_{(i+1)T/n}) - \log(p_{iT/n})\right)^{p_1} (\widehat{\sigma}^2_{(i+1)T/n} - \sigma^2_{(i+1)T/n} + \sigma^2_{(i+1)T/n} - \sigma^2_{iT/n} + \sigma^2_{iT/n} - \widehat{\sigma}^2_{iT/n})^{p_2} \\
= &\sum_{i+j+k=p_2} c_{i,j,k} \left(\log(p_{(i+1)T/n}) - \log(p_{iT/n})\right)^{p_1} \left(\sigma^2_{iT/n} - \sigma^2_{iT/n}\right)^k \left(\widehat{\sigma}^2_{(i+1)T/n} - \sigma^2_{(i+1)T/n}\right)^i \left(\widehat{\sigma}^2_{iT/n} - \sigma^2_{iT/n}\right)^j \\
= &O_p(\Delta^{1/2}_{n,T}).
\end{aligned}$$

Thus, using the mean-value theorem,

$$\begin{aligned}
(a) &= O_p(\Delta^{1/2}_{n,T}) \frac{\max_{1\le i\le n} \left| \frac{\widehat{\sigma}^2_{i\Delta_{n,T}} - \sigma^2_{i\Delta_{n,T}}}{h_{n,T}} \right| \sum_{i=1}^{n-1} \left| \mathbf{K}'\left(\frac{\sigma^2_{iT/n} + O_p(\mathbf{M}_{n,k,T}) - x}{h_{n,T}} \right) \right|}{\Delta_{n,T} \sum_{i=1}^{n} \widehat{\mathbf{K}}_i} \\
&= O_p\left(\frac{\mathbf{M}_{n,k,T}\Delta^{1/2}_{n,T}}{\Delta_{n,T} h_{n,T}} \right) \frac{\frac{\frac{1}{h_{n,T}} \int_0^T \left| \mathbf{K}'\left(\frac{\sigma^2_s - x}{h_{n,T}} \right) \right| \mathrm{d}s}{\frac{1}{h_{n,T}} \int_0^T \mathbf{K}\left(\frac{\sigma^2_s - x}{h_{n,T}} \right) \mathrm{d}s} + O_p\left(\frac{\Delta_{n,T}}{h^2_{n,T}} \right)}{1 + O_p\left(\frac{\Delta_{n,T}}{h^2_{n,T}} \right) + O_p(g(n,T,k,\phi))} \\
&= O_p\left(\frac{\mathbf{M}_{n,k,T}}{\Delta^{1/2}_{n,T} h_{n,T}} \right)
\end{aligned}$$

if $\frac{\Delta_{n,T}}{h^2_{n,T}} \to 0$ and $g(n,T,k,\phi) \to 0$. We now turn to (b).

$$(b) = \frac{\sum_{i=1}^{n-1} \mathbf{K}_i \left(\log(p_{(i+1)T/n}) - \log(p_{iT/n})\right)^{p_1} \left[(\widehat{\sigma}^2_{(i+1)T/n} - \widehat{\sigma}^2_{iT/n})^{p_2} - (\sigma^2_{(i+1)T/n} - \sigma^2_{iT/n})^{p_2} \right]}{\Delta_{n,T} \sum_{i=1}^{n} \widehat{\mathbf{K}}_i}.$$

We can prove that $\xi_i := \left(\log(p_{(i+1)T/n}) - \log(p_{iT/n})\right)^{p_1} \left[(\widehat{\sigma}^2_{(i+1)T/n} - \widehat{\sigma}^2_{iT/n})^{p_2} - (\sigma^2_{(i+1)T/n} - \sigma^2_{iT/n})^{p_2} \right]$ is $O_p(\mathbf{M}_{n,k,T}\Delta^{1/2}_{n,T})$. This is straightforward if $p_2 = 1$. If $p_2 > 1$, using $a^p - b^p = (a-b)\sum_{k=1}^{p} a^{p-k} b^{k-1}$, we have

$$\begin{aligned}
&(\widehat{\sigma}^2_{(i+1)T/n} - \widehat{\sigma}^2_{iT/n})^{p_2} - (\sigma^2_{(i+1)T/n} - \sigma^2_{iT/n})^{p_2} \\
= &(\widehat{\sigma}^2_{(i+1)T/n} - \widehat{\sigma}^2_{iT/n} - \sigma^2_{(i+1)T/n} + \sigma^2_{iT/n}) \sum_{k=1}^{p_2} (\widehat{\sigma}^2_{(i+1)T/n} - \widehat{\sigma}^2_{iT/n})^{p_2-k} (\sigma^2_{(i+1)T/n} - \sigma^2_{iT/n})^{k-1}.
\end{aligned}$$

We can now decompose ξ_i into its terms, the first corresponding to $k = 1$:

$$\begin{aligned}
&\left(\log(p_{(i+1)T/n}) - \log(p_{iT/n})\right)^{p_1} (\widehat{\sigma}^2_{(i+1)T/n} - \widehat{\sigma}^2_{iT/n} - \sigma^2_{(i+1)T/n} + \sigma^2_{iT/n})(\widehat{\sigma}^2_{(i+1)T/n} - \widehat{\sigma}^2_{iT/n})^{p_2-1} \\
= &O_p(\mathbf{M}_{n,k,T}) \left(\log(p_{(i+1)T/n}) - \log(p_{iT/n})\right)^{p_1} (\widehat{\sigma}^2_{(i+1)T/n} - \sigma^2_{(i+1)T/n} - \widehat{\sigma}^2_{iT/n} + \sigma^2_{iT/n} + \sigma^2_{(i+1)T/n} - \sigma^2_{iT/n})^{p_2-1} \\
= &o_p(\mathbf{M}_{n,k,T}\Delta^{1/2}_{n,T}).
\end{aligned}$$

For the terms with $2 \le k \le p_2 - 1$, we may write:

$$\begin{aligned}
&(\widehat{\sigma}^2_{(i+1)T/n} - \widehat{\sigma}^2_{iT/n} - \sigma^2_{(i+1)T/n} + \sigma^2_{iT/n})(\widehat{\sigma}^2_{(i+1)T/n} - \widehat{\sigma}^2_{iT/n})^{p_2-k} \underbrace{\left(\log(p_{(i+1)T/n}) - \log(p_{iT/n})\right)^{p_1} (\sigma^2_{(i+1)T/n} - \sigma^2_{iT/n})^{k-1}}_{O_p(\Delta^{1/2}_{n,T})} \\
= &O_p(\mathbf{M}_{n,k,T}\Delta^{1/2}_{n,T})(\widehat{\sigma}^2_{(i+1)T/n} - \sigma^2_{(i+1)T/n} - \widehat{\sigma}^2_{iT/n} + \sigma^2_{iT/n} + \sigma^2_{(i+1)T/n} - \sigma^2_{iT/n})^{p_2-k} \\
= &o_p(\mathbf{M}_{n,k,T}\Delta^{1/2}_{n,T}).
\end{aligned}$$

The term $k = p_2$ is the dominating one since

$$\left(\log(p_{(i+1)T/n}) - \log(p_{iT/n})\right)^{p_1} (\sigma^2_{(i+1)T/n} - \sigma^2_{iT/n})^{p_2 - 1} (\widehat{\sigma}^2_{(i+1)T/n} - \widehat{\sigma}^2_{iT/n} - \sigma^2_{(i+1)T/n} + \sigma^2_{iT/n}) = O_p(\mathbf{M}_{n,k,T} \Delta^{1/2}_{n,T}).$$

Thus,

$$(b) = O_p\left(\frac{\mathbf{M}_{n,k,T}\Delta^{1/2}_{n,T}}{\Delta_{n,T}}\right) \frac{\Delta_{n,T} \sum_{i=1}^{n-1} \mathbf{K}_i}{\Delta_{n,T} \sum_{i=1}^{n} \widehat{\mathbf{K}}_i} = O_p\left(\frac{\mathbf{M}_{n,k,T}}{\Delta^{1/2}_{n,T}}\right),$$

if $\frac{\Delta_{n,T}}{h^2_{n,T}} \to 0$ and $g(n,T,k,\phi) \to 0$. Finally,

$$\begin{aligned}
(c) &= \frac{\sum_{i=1}^{n-1} \mathbf{K}_i \left(\log(p_{(i+1)T/n}) - \log(p_{iT/n})\right)^{p_1} (\sigma^2_{(i+1)T/n} - \sigma^2_{iT/n})^{p_2}}{\Delta_{n,T} \sum_{i=1}^{n} \mathbf{K}_i} \times \frac{\Delta_{n,T}\left(\sum_{i=1}^{n} \widehat{\mathbf{K}}_i - \sum_{i=1}^{n} \mathbf{K}_i\right)}{\Delta_{n,T} \sum_{i=1}^{n} \widehat{\mathbf{K}}_i} \\
&= O_p\left(\frac{\max_{1 \le i \le n} \left| \frac{\widehat{\sigma}^2_{i\Delta_{n,T}} - \sigma^2_{i\Delta_{n,T}}}{h_{n,T}} \right| \frac{\Delta_{n,T}}{h_{n,T}} \sum_{i=1}^{n} \left| \mathbf{K}'\left(\frac{\sigma^2_{i\Delta_{n,T}} - x}{h_{n,T}}\right) \right|}{\frac{\Delta_{n,T}}{h_{n,T}} \sum_{i=1}^{n} \widehat{\mathbf{K}}_i}\right) \\
&= O_p\left(\frac{\mathbf{M}_{n,k,T}}{h_{n,T}}\right),
\end{aligned}$$

if $\frac{\Delta_{n,T}}{h^2_{n,T}} \to 0$ and $g(n,T,k,\phi) \to 0$. Now, notice that (c) is of higher order than (b) since $\frac{\mathbf{M}_{n,k,T}}{h_{n,T}} = \frac{\mathbf{M}_{n,k,T}}{\Delta^{1/2}_{n,T}} \frac{\Delta^{1/2}_{n,T}}{h_{n,T}} = \frac{\mathbf{M}_{n,k,T}}{\Delta^{1/2}_{n,T}} o(1)$. Thus

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

which completes the proof.

12 Appendix B: Finite Sample Issues

12.1 Measuring spot volatility with microstructure noise and jumps

The spot variance measures are obtained as follows. Each day, we split the day into 6 consecutive hours, starting at 9.45am. Denote by $\sigma^2_{t,i}$ the spot variance estimator corresponding to day t, hour i, $i = 1, \ldots, 6$. To estimate $\sigma^2_{t,i}$, we first compute logarithmic prices $\log p_{t,i,k}$ every minute, with $k = 0, \ldots, 60$, by averaging all observed transaction prices inside each minute. This pre-averaging should make the impact of microstructure noise negligible (Jacod et al., 2009). Next, we compute one-minute logarithmic returns $r_{t,i,k} = \log p_{t,i,k} - \log p_{t,i,k-1}$ for $k = 1, \ldots, 60$. The spot variance estimates are obtained by applying the jump-robust threshold bipower variation estimator in Eq. (11). These estimates are then scaled with a constant scale factor so as to guarantee that the average of the hourly spot variance estimates equals the threshold bipower variation estimated on daily close-to-close returns. Only variance estimates with a sufficiently high number of transactions inside an hour are retained. The exact filter is: the number of transactions in an hour should be no less than one third of the average number of transactions in the last month. We also eliminate spot variance estimates lower than 0.01. In our sample, this simple filter retains 98.88% of the estimated volatilities. Finally, we also record the last price $\log p_{t,i}$ in each hour.

12.2 Infinitesimal cross-moment estimation with measured volatilities

To estimate the moments, we first compute an exponential moving average of estimated volatilities $\widetilde{\widehat{\sigma}}^2_{t,i}$ with 20 lags. These smoother volatilities are only used inside the kernels. The exact moment expressions are:

$$\widetilde{\widehat{\vartheta}}_{p_1,p_2}(\sigma) = \frac{\sum_{t=1}^{T-1}\sum_{i=1}^{N_{hours}} \mathbf{K}\left(\frac{\widetilde{\widehat{\sigma}}_{t,i}-\sigma}{h}\right)(\log p_{t+1,i} - \log p_{t,i})^{p_1}\left(\log \widehat{\sigma}^2_{t+1,i} - \log \widehat{\sigma}^2_{t,i}\right)^{p_2}}{\Delta \sum_{t=1}^{T}\sum_{i=1}^{N_{hours}} \mathbf{K}\left(\frac{\widetilde{\widehat{\sigma}}_{t,i}-\sigma}{h}\right)}, \qquad p_2 = 0, 1$$

$$\widetilde{\widehat{\vartheta}}_{p_1,2}(\sigma) = \frac{\sum_{t=1}^{T-1}\sum_{i=1}^{N_{hours}} \mathbf{K}\left(\frac{\widetilde{\widehat{\sigma}}_{t,i}-\sigma}{h}\right)(\log p_{t+1,i} - \log p_{t,i})^{p_1}\left[\left(\log \widehat{\sigma}^2_{t+1,i} - \log \widehat{\sigma}^2_{t,i}\right)^2 - 2c\right]}{\Delta \sum_{t=1}^{T}\sum_{i=1}^{N_{hours}} \mathbf{K}\left(\frac{\widetilde{\widehat{\sigma}}_{t,i}-\sigma}{h}\right)},$$

$$\widetilde{\widehat{\vartheta}}_{p_1,3}(\sigma) = \frac{\sum_{t=1}^{T-1}\sum_{i=1}^{N_{hours}} \mathbf{K}\left(\frac{\widetilde{\widehat{\sigma}}_{t,i}-\sigma}{h}\right)(\log p_{t+1,i} - \log p_{t,i})^{p_1}\left[\left(\log \widehat{\sigma}^2_{t+1,i} - \log \widehat{\sigma}^2_{t,i}\right)^3 - 6c\left(\log \widehat{\sigma}^2_{t+1,i} - \log \widehat{\sigma}^2_{t,i}\right)\right]}{\Delta \sum_{t=1}^{T}\sum_{i=1}^{N_{hours}} \mathbf{K}\left(\frac{\widetilde{\widehat{\sigma}}_{t,i}-\sigma}{h}\right)},$$

$$\widetilde{\widehat{\vartheta}}_{p_1,4}(\sigma) = \frac{\sum_{t=1}^{T-1}\sum_{i=1}^{N_{hours}} \mathbf{K}\left(\frac{\widetilde{\widehat{\sigma}}_{t,i}-\sigma}{h}\right)(\log p_{t+1,i} - \log p_{t,i})^{p_1}\left[\left(\log \widehat{\sigma}^2_{t+1,i} - \log \widehat{\sigma}^2_{t,i}\right)^4 - 12c\left(\log \widehat{\sigma}^2_{t+1,i} - \log \widehat{\sigma}^2_{t,i}\right)^2 - 12c^2\right]}{\Delta \sum_{t=1}^{T}\sum_{i=1}^{N_{hours}} \mathbf{K}\left(\frac{\widetilde{\widehat{\sigma}}_{t,i}-\sigma}{h}\right)},$$

with $N_{hours} = 6$ and $c = 2.61/59$ in our case. The finite-sample adjustments for $p_2 = 2, 3, 4$ are designed to correct for measurement error in the spot variance estimates. The adjustments are computed using the

asymptotic results which hold for our assumed spot variance estimates, namely

$$\widehat{\sigma}^2_{t,i} = \sigma^2_{t,i} + \varepsilon_{t,i},$$

where $\varepsilon_{t,i}$ is normally distributed with mean zero and variance $c\sigma^4_{t,i}$ (Bandi and Renò, 2009, and Corsi et al., 2010). To account for the log-transformation, the delta method has been employed. This is accurate since, in our sample, $\sqrt{c} \ll 1$.

12.3 Infinitesimal cross-moment estimation for a fixed Δ

The infinitesimal nature of the moments is so that finite sample contaminations may remain in the estimates if Δ is not small enough. To this extent, we implement exact first-order corrections (in Δ) for the estimates. These corrections are immaterial asymptotically and useful in finite samples. The recursive formula to compute them is:

$$\widetilde{\widehat{\vartheta}}^{\Delta}_{p_1,p_2} = \widetilde{\widehat{\vartheta}}_{p_1,p_2} - \frac{\Delta}{2}\sum_{j_1=0}^{p_1}\sum_{j_2=0}^{p_2}\left(\begin{array}{c} p_1 \\ j_1 \end{array}\right)\left(\begin{array}{c} p_2 \\ j_2 \end{array}\right)\widetilde{\widehat{\vartheta}}_{j_1,j_2}\widetilde{\widehat{\vartheta}}_{p_1-j_1,p_2-j_2}$$

with $\widehat{\widehat{\vartheta}}_{0,0} = 0$.

12.4 Simulations

The reliability of the estimated infinitesimal cross-moments, and their finite-sample adjustments, are evaluated by simulation. We first simulate the bivariate system in Eq. (12), described in Section 7, with the parameters in Column 4, Table 2. We use a first-order Euler scheme. Initially, σ^2_t does not contain any estimation error. The estimation error is added using the (asymptotic) formula $\hat{\sigma}^2_t = \sigma^2_t + \sqrt{2.61/59}\sigma^2_t\varepsilon_t$, where ε_t is a standard normal draw and $\sqrt{2.61/59}\sigma^2_t$ is the (asymptotic) standard deviation of threshold bipower variation estimator computed with 60 one-minute returns. With the resulting time series, we estimate the infinitesimal cross-moments using the small-sample corrections described above. The choice of the conditioning points, kernel function, and bandwidth(s) are the same as those used to compute the moments for parametric and nonparametric estimation in the main text. Specifically, we use a Gaussian kernel, the set

$$\sigma^2 = \{0.1311, 0.2227, 0.3041, 0.3913, 0.4922, 0.6169, 0.7829, 1.0328, 1.5018, 3.0984\}$$

for the conditioning points (these are the midpoints of bins which contain the same number of estimated volatilities), and the bandwidth vector (corresponding to each conditioning point)

$$h = \{0.4165, 0.1811, 0.1374, 0.1173, 0.1126, 0.1131, 0.1239, 0.1496, 0.2134, 0.4397\}.$$

Consistent with the asymptotics, h smoothes more in regions with sparse observations near the boundaries of the range of the estimated volatility series. We implement $1,500$ simulations. Figure 11 presents the generated moments, together with the estimated median and 95% confidence bands on the replications. The figure shows that our small-sample procedures are able to deliver accurate estimates of the infinitesimal cross-moments. Figure 12 displays the estimated parameters on simulations of the parametric model in Section 7 with the parameters in Table 2, Column 4. Estimation of the parameters via GMM is satisfactory and not affected by meaningful biases.

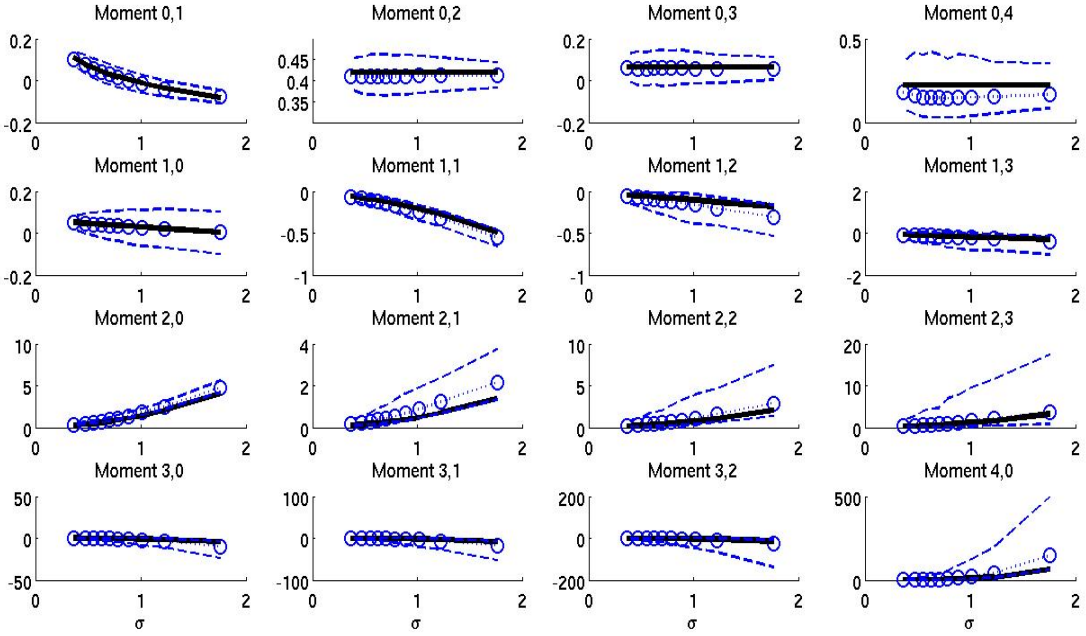

Figure 11: Simulation results for the estimated moments. Dotted lines with circles: median estimates. Dashed lines: 95% confidence bands. Solid lines: generated moments.

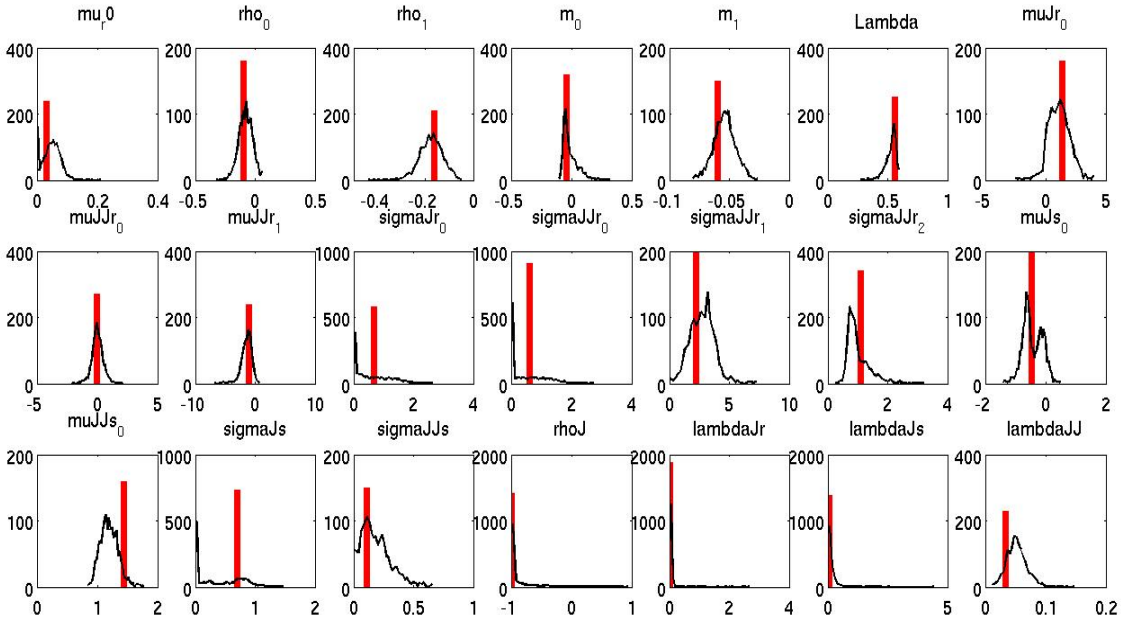

Figure 12: Simulation results for the parametric model. Vertical bars: Generated parameters. Solid lines: Histograms of the estimated parameters on simulations.

References

Aït-Sahalia, Y., J. Fan, and Y. Li (2011). The leverage effect puzzle: Disentangling sources of bias at high frequency. Working paper.

Alizadeh, S., M. Brandt, and F. Diebold (2002). High and low frequency exchange rate volatility dynamics: Range-based estimation of stochastic volatility models. *Journal of Finance 57*, 1047–1092.

Andersen, T. and L. Benzoni (2011). Stochastic volatility. In *Complex Systems in Finance and Econometrics*, pp. 694–726. Springer.

Andersen, T., L. Benzoni, and J. Lund (2002). An empirical investigation of continuous-time equity return models. *Journal of Finance 57*, 1239–1284.

Andersen, T., T. Bollerslev, and F. X. Diebold (2010). Parametric and nonparametric volatility measurement. In L. P. Hansen and Y. Ait-Sahalia (Eds.), *Handbook of Financial Econometrics*. Amsterdam: North-Holland.

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

Bandi, F. and R. Renò (2009). Nonparametric stochastic volatility. Working Paper.

Bandi, F. and R. Renò (2011). Time-varying leverage effects. *Journal of Econometrics*. Forthcoming.

Barndorff-Nielsen, O. E. and N. Shephard (2004). Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics 2*, 1–48.

Bates, D. (2000). Post-'87 crash fears in the S&P 500 futures option market. *Journal of Econometrics 94*, 181–238.

Chacko, G. and L. Viceira (2003). Spectral GMM estimation of continuous-time processes. *Journal of Econometrics 116*(1-2), 259–292.

Chernov, M., R. Gallant, E. Ghysels, and G. Tauchen (2003). Alternative models for stock price dynamics. *Journal of Econometrics 116*(1), 225–258.

Chernov, M. and E. Ghysels (2000). A study towards a unified approach to the joint estimation of objective and risk neutral measures for the purpose of options valuation. *Journal of Financial Economics 56*, 407–458.

Chib, S., F. Nardari, and N. Shephard (2002). Markov chain monte carlo methods for stochastic volatility models. *Journal of Econometrics 108*(2), 281–316.

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

Das, S. and R. Sundaram (1999). Of smiles and smirks: A term structure perspective. *Journal of Financial and Quantitative Analysis 34*(2), 211–239.

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

Eraker, B. (2004). Do Stock Prices and Volatility Jump? Reconciling Evidence from Spot and Option Prices. *The Journal of Finance 59*(3), 1367–1404.

Eraker, B., M. Johannes, and N. Polson (2003). The impact of jumps in volatility and returns. *Journal of Finance 58*, 1269–1300.

Hansen, L. (1982). Large sample properties of generalized method of moments estimators. *Econometrica 59*(4), 1029–1054.

Harvey, A. and N. Shephard (1996). Estimation of an asymmetric stochastic volatility model for asset returns. *Journal of Business & Economic Statistics 14*(4), 429–434.

Heston, S. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of financial studies 6*, 327–343.

Jacod, J., Y. Li, P. Mykland, M. Podolskij, and M. Vetter (2009). Microstructure noise in the continuous case: the pre-averaging approach. *Stochastic Processes and their Applications 119*(7), 2249–2276.

Jacod, J. and A. N. Shiryaev (2003). *Limit Theorems for Stochastic Processes*. Springer.

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

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

Jacquier, E., N. Polson, and P. Rossi (2004). Bayesian analysis of stochastic volatility models with fat-tails and correlated errors. *Journal of Econometrics 122*(1), 185–212.

Jones, C. (2003). Nonlinear mean reversion in the short-term interest rate. *Review of Financial Studies 16*, 765–791.

Lee, S. and P. Mykland (2008). Jumps in financial markets: A new nonparametric test and jump dynamics. *Review of Financial studies 21*(6), 2535.

Mancini, C. (2009). Non-parametric threshold estimation for models with stochastic diffusion coefficient and jumps. *Scandinavian Journal of Statistics 36*(2), 270–296.

Pan, J. (2002). The jump-risk premia implicit in options: Evidence from an integrated time series study. *Journal of Financial Economics 63*, 3–50.

Piazzesi, M. (2010). Affine term structure models. In L. P. Hansen and Y. Ait-Sahalia (Eds.), *Handbook of Financial Econometrics*. Amsterdam: North-Holland.

Revuz, D. and M. Yor (1994). *Continuous Martingales and Brownian Motion*. Springer.

Todorov, V. and G. Tauchen (2010). Activity signature functions for high-frequency data analysis. *Journal of Econometrics 54*, 125–138.

Todorov, V. and G. Tauchen (2011). Volatility jumps. *Journal of Business and Economic Statistics 29*, 356–371.

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

Yu, J. (2005). On leverage in a stochastic volatility model. *Journal of Econometrics 127*(2), 165–178.

Yu, J. (2010). A semiparametric stochastic volatility model. *Journal of Econometrics*. Forthcoming.