Journal of Financial Economics 119 (2016) 107–146

Contents lists available at ScienceDirect

Journal of Financial Economics

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

Price and volatility co-jumps ☆

F.M. Bandi [a,b,*], R. Renò [c]

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

ARTICLE INFO

Article history:
Received 13 December 2012
Received in revised form
23 May 2014
Accepted 5 August 2014
Available online 3 June 2015

JEL Classification:
G12
C14
C58

Keywords:
Stochastic volatility
Jumps in prices
Jumps in volatility
Co-jumps
Infinitesimal cross-moments
Return risk premia
Variance risk premia

ABSTRACT

The nature of the dependence between discontinuities in prices and contemporaneous discontinuities in volatility (co-jumps) has been reported by many as being elusive, in terms of sign, magnitude, and statistical significance. Using a novel identification strategy in continuous time relying on trade-level information for *spot variance* estimation, as well as *infinitesimal cross-moments*, we document that a sizeable proportion of discontinuous changes in prices are associated with strongly anti-correlated, contemporaneous, discontinuous changes in volatility. Assuming a possibly nonmonotonic pricing kernel, we illustrate the equilibrium implications of price and volatility co-jumps for return and variance risk premia.

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.

☆ We thank seminar participants at the 2012 WFA Annual Meetings (Las Vegas, NV), 2013 ICEEE (Genova, Italy), Bauer College of Business (University of Houston), Brown University, Carey Business School (Johns Hopkins University), CASS Business School (City University London), Edhec Business School, Georgetown University, Singapore Management University, Tolouse School of Economics, the Triangle Econometric Seminar, Wisconsin School of Business (University of Wisconsin), and University of Florence for useful discussions. We are very grateful to the Editor (Bill Schwert), two anonymous reviewers and Dobri Dobrev (the WFA discussant) for their comments and suggestions. Roberto Renò acknowledges financial support from The Castilla and León Council of Education and Science, Grant VA016A10-1.

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

http://dx.doi.org/10.1016/j.jfineco.2015.05.007
0304-405X/

Adding jumps in volatility provides a solution to this issue. As pointed out in Eraker, Johannes, and Polson (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, Cao, and Chen (1997), Bates (2000), and Pan (2002). Duffie, Pan, and Singleton (2000) and Eraker, Johannes, and Polson (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 likelihood of discontinuous movements occurring *jointly* and the nature of the correlation—or lack thereof—between the resulting jump sizes. Importantly, while the presence of co-jumps is broadly accepted (see, among others, Jacod and Todorov, 2010), the study of the sign and magnitude of their dependence has led to a fundamental disconnect between the strand of the literature that uses option data and the one that solely relies on price data.

The former is rather emphatic about the presence of *negatively correlated* price and volatility large comovements. Eraker (2004) finds a negative correlation between jump sizes *only when* augmenting return data with option data, the negative correlation between co-jump sizes being identified off of the implied volatility smirk.[1] Using high-frequency observations on the VIX and the S&P500 index, along with sample statistics developed in Jacod and Todorov (2009) and Todorov and Tauchen (2010), Todorov and Tauchen (2011) find evidence for a specification with a high likelihood of contemporaneous arrivals along with a lower likelihood of disjoint arrivals. Their reported sizes of the large price and volatility comovements are strongly anti-correlated.

The literature which does not employ option data is, on the other hand, drastically less clear-cut about the nature of the co-jump correlation, a strongly negative dependence not being a documented empirical feature in this case. Using tests developed in Jacod, Kluppelberg and Muller (2012a), Jacod, Kluppelberg and Muller (2012b) cannot reject the null of no correlation, except in a single case. Even in this single case, removal of the Flash Crash leads to insignificant dependence between discontinuous price and volatility changes. Insignificant correlation estimates are reported in Eraker, Johannes, and Polson (2003), Chernov, Gallant, Ghysels, and Tauchen (2003) (with a negative sign), and Eraker (2004) (with a positive sign). The state of this strand of the literature, and the issues related to the identification of co-jump correlation, are summarized effectively by Broadie, Chernov, and Johannes (2007) (see, e.g., their discussion on p. 1457), who assume co-jump independence.

Since pricing and risk management hinge crucially on the nature of the link between discontinuous price and volatility changes, it is important to understand whether an economically meaningful negative dependence, yielded by co-jump means of opposite sign and/or a negative co-jump size correlation, is indeed a stylized feature of sensible price formation mechanisms. If this were the case, the resulting *joint* directional and volatility jump risk would have to be properly managed. In this sense, the disconnect between findings with and without the information in option data is a concern. Is this disconnect mostly due to the impact of risk premia? In the case of the VIX, for instance, sudden implied volatility increases could be due to market fear (and increased risk premia), particularly in times of negative directional shocks (i.e., negative price jumps). This simple effect alone can give rise to a negative correlation between the volatility proxy (VIX) and market returns even in a situation where sudden, discontinuous changes in true volatility and market prices hardly covary with each other. Or, is the disconnect—to a first degree—due to the way in which volatility is filtered, option information providing the resolution that traditional volatility filters may otherwise lack? We provide evidence for this latter thesis.

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 argue that its accurate identification is important to evaluate fine-grain dynamics such as those reflected in volatility jumps and price and volatility co-jumps. In essence, the mixed findings in the literature can be the result of volatility filters which may not be granular enough as to shed light on rare co-variations. There are two key steps to the procedure we propose. The first step is the estimation of *spot volatility*. Spot volatility, shown in Section 5 to be crucial for the unbiased identification of both co-jump intensity and co-jump correlation, is estimated nonparametrically for every hour in our sample using intradaily price data. In the second step, volatility is utilized, along with price data, in the context of a novel moment-based procedure (dubbed Nonparametric Infinitesimal Method of Moments or NIMM) hinging on the notion of *infinitesimal cross-moment*. In essence, we evaluate the daily dynamics of returns and volatility by making use of the substantial informational content of intradaily price data for the purpose of spot volatility filtering.

We emphasize that the use of a spot, or local, notion of volatility is not immaterial. We show that volatility filters which use too large a bandwidth for identification (as is the case for daily realized variance or daily variances from likelihood procedures) may cause attenuation to zero in the estimation of the cross-moments. Being cross-moment information

[1] This finding is coherent with the calibration results in Duffie, Pan, and Singleton (2000) who, again, provide evidence of anti-correlated price and volatility co-jump sizes by fitting the typically found, pronounced smirk in option prices. Calibrations are also used in the more recent work of Cont and Kokholm (2011).

implicitly or explicitly, as in our case, used by any co-jump identification procedure, including likelihood methods, the mixed findings in the existing literature—as reported above—have the potential to be the outcome of volatility filters which lack the granularity needed to identify rare comovements and their dependence. When employing these filters, the use of option data in the identification of rare co-jumps may, therefore, be viewed as supplying the resolution which low-frequency price data, and the resulting filtered volatilities, could not provide. The presence of co-jumps and their negative dependence should, however, be revealed by price data only without potential contaminations imputable to risk premia. Using high-frequency price data for the purpose of local volatility filtering, and our proposed modeling and identification approach, we show that the co-jumps, along with negatively correlated co-jump sizes, are an important stylized feature of the dynamic evolution of daily market prices and volatility series, one that continuous-time models of price formation should embed. Price data *alone* forcefully uncover the negative dependence between large discontinuities in prices and volatilities, with no need for option data.

The paper employs S&P 500 futures high-frequency returns from April 21, 1982 to February 5, 2009. 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 fine-grain dynamics of the return and volatility process are such that some of the discontinuous components are fully idiosyncratic while some are common to the return/volatility processes. 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.

As stressed, the occurrence of co-jumps is also generally associated with price and volatility changes (i.e., co-jump sizes) of opposite sign. We estimate co-jump mean sizes with different signs, a price co-jump mean becoming more negative with increasing volatility levels, and a correlation between the size of the price jumps and the size of the volatility jumps very close to -1. These effects are important. Among other issues, they introduce 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), but the infrequent, large discontinuities also have negative conditional (on volatility) first cross-moments (a leverage effect induced by discontinuous comovements). The two combined effects lead to a form of "generalized leverage." Hence, there are two channels through which negative shocks to prices can increase variance and, therefore, induce 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").

Using a possibly nonmonotonic pricing kernel, i.e., a pricing kernel decreasing in prices and increasing in variance (as discussed by, e.g., Christoffersen, Heston, and Jacobs, 2013), we derive equilibrium pricing implications of *generalized leverage* and, therefore, of the presence of negatively correlated co-jumps. We show that generalized leverage delivers both a component of return risk premia and a component of variance risk premia. In other words, not only do the co-jumps lead to an additional source of skewness in stock returns, but they also modify—in equilibrium—the mean of the return distribution while contributing to a negative variance risk premium. Should the stochastic discount factor not depend on variance, and therefore be monotonic in prices, then generalized leverage, in general, and the negative correlation between price and volatility co-jumps, in particular, would not matter to generate return risk premia. They would, however, still play a role in determining variance risk premia.

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 successful parametric specifications in the literature to a nonparametric framework. Section 4 presents the notion of infinitesimal cross-moment and analyzes 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 discuss equilibrium pricing implications of co-jump presence both in terms of return risk premia and in terms of variance risk premia. 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 (conditional) standard deviation over the previous day.[2] Should the changes be induced by continuous Brownian shocks, in light of the approximate normality of Brownian motion over small time increments, they

[2] The econometric analysis of this paper, described in detail in Section 5 and Appendix B, employs "hourly" (over 65 minutes, to be accurate) 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 6.5-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 Fig. 6, were used without applying the smoother.

Table 1
Descriptive evidence on co-jumps. We use S&P500 intradaily futures prices from April 21, 1982 to February 5, 2009. The units are for daily returns expressed in percentage terms. Columns 2 and 3 report the largest daily returns and their *t*-statistics (obtained as the ratios of daily returns to daily volatilities). Columns 4 and 5 report the corresponding daily changes in logarithmic variance (estimated using intradaily data) and their *t*-statistics (obtained as the ratios of the daily logarithmic variance changes to their conditional volatility, as derived in Section 7).

Day	Return (%)	*t*-Stat	Log-variance change	*t*-Stat	Events
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

would be larger than two in absolute value with approximate probability equal to 5%. Unusually large (standardized) price changes would, therefore, signal jumps in the price process.[3]

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),[4] the Asian crisis (October 27, 1997), the Russian crisis (August 31, 1998), 9/11 (September 17, 2001),[5] 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 logarithmic variance changes are close-to-close. They are obtained by virtue of spot (hourly) variance estimates making use of intradaily 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 their corresponding conditional (on spot volatility) volatility[6] to rule out discontinuities which may be consistent with smooth Brownian dynamics. Two effects are striking. First, the logarithmic 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, these changes are, in general, of opposite sign with respect to the price changes. In other words, abnormally large, negative price discontinuities are often associated with abnormally large, positive variance discontinuities.

Fig. 1 reports the estimated probability density functions for both the co-jumps in price and the co-jumps in variance. It is interesting to observe the thickness of the tails of the distribution of the price co-jumps, signaling the need for a model specification with rich dynamics for the variance of the price co-jump sizes. We will return to this issue.

[3] This same logic is formally employed by Lee and Mykland (2008) to test for jumps using intradaily return data.

[4] On this day, the reported magnitude of the futures price jump in the table (−30.01%) is larger than that of the jump in spot prices (−22%). Harris (1989) studies the S&P 500 spot/futures basis on October 19, 1987.

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

[6] Computation of the volatility of the logarithmic variance changes 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 the logarithmic variance changes identified from our proposed stochastic volatility model in Section 7.

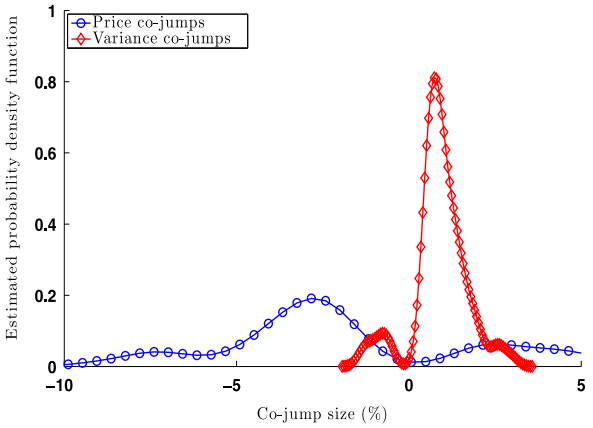

Fig. 1. Estimated probability density functions (pdfs) of the price and variance co-jumps. We use S&P500 intradaily futures prices from April 21, 1982 to February 5, 2009. The line with circles is the estimated pdf of the co-jumps in prices. The line with diamonds is the estimated pdf of the co-jumps in logarithmic variances.

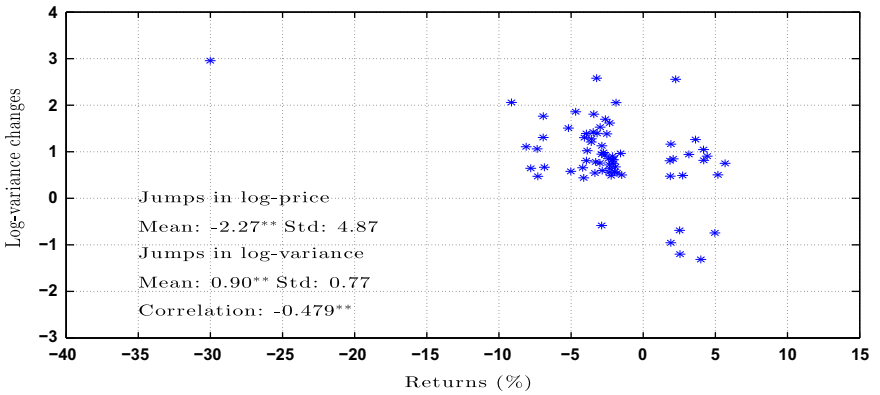

Fig. 2. Scatter plot of the significant (daily) co-jumps in prices and variances. We use S&P500 intradaily futures prices from April 21, 1982 to February 5, 2009. We report mean and standard deviation of the discontinuities in prices, mean and standard deviation of the discontinuities in logarithmic variances, and their correlation.

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 variance jump sizes, we report a scatter plot of the logarithmic variance jumps versus the price jumps. Instead of reporting the co-jumps associated with the 30 largest price changes, as in Table 1, we now report the co-jumps associated with price and logarithmic 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. For these jumps, plotted in Fig. 2, the mean and the standard deviation of the price discontinuities are estimated as being equal to -2.27% and 4.87%. The estimated values of the mean and standard deviation of the logarithmic variance jumps are 0.9 and 0.77. Thus, when contemporaneous jumps in price and variance occur, the variance jumps tend to be positive whereas the price jumps tend to be negative. In addition, the scatter plot provides compelling visual evidence about a strongly negative correlation between co-jump sizes (measured at a significant value of about -0.48). In other words, price jumps below their (negative) mean generally occur alongside variance spikes above their (positive) mean, thereby resulting in co-jumps with strongly negatively correlated jump sizes.[7]

By sorting with respect to the large price (and variance) 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 variance co-jumps in our sample. In agreement with this observation, our point estimates (below) will imply 8/9 daily co-jumps per annum.

In Fig. 3 we plot the large price discontinuities displayed in Fig. 2 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

[7] There is one main exception in our sample. During the aftermath of the 1987 crash, on October 21 1987, a percentage change of 16.11 in logarithmic variance was accompanied by a percentage (positive) change of 15.5 in the logarithmic prices. 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 subsequent estimates.

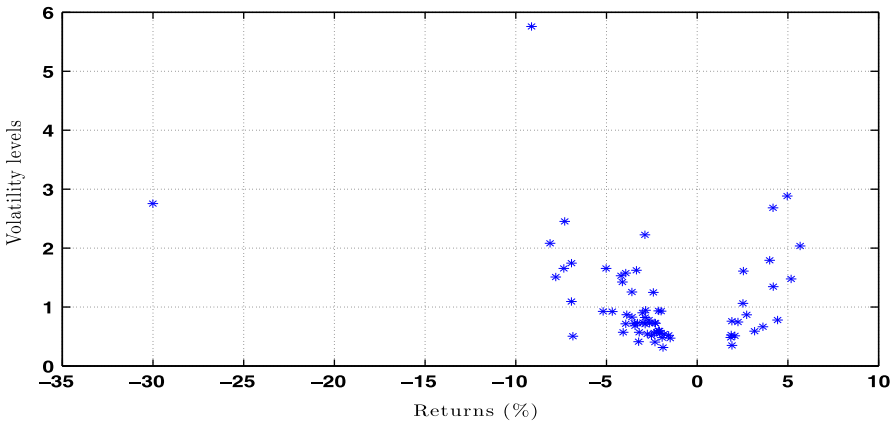

Fig. 3. Scatter plot of the significant daily price co-jumps and the volatility level (%, daily) at which they occur. We use S&P500 intradaily futures prices from April 21, 1982 to February 5, 2009.

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 variance. Consistent with the evidence reported here, the model will allow for correlation between the co-jump sizes. By permitting the price and variance co-jump sizes not to be mean zero, the proposed specification will also allow for a higher likelihood of negative price jumps and a higher likelihood of positive variance 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 structure for the price and variance dynamics, 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[8] 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

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

where $\xi(\cdot)$ is an increasingly monotonic function, $W=\{W^1,W^2\}$ is a bivariate standard Brownian motion vector, $J=\{J_r,J_\sigma,J_{r,\sigma}\}$ 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 solely such that a unique strong solution to the system exists. We do not require stationarity of the system, but only *recurrence* of the latent variable, i.e., variance, as in Bandi and Renò (2009). The model 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 volatility level. We, therefore, complement and extend models with affine volatility- dependent price jump intensities (Bates, 2000; Pan, 2002) as well as models with affine volatility-dependent variance jump intensities (Eraker, 2004).
3. (State-dependent jumps) The moments of the jump sizes are also permitted to be functions of the underlying volatility level. Assume $\xi(\cdot)=\log(\cdot)$, as in a logarithmic variance model, for instance. Let the sizes of the idiosyncratic jumps be normally distributed: $(c_{r,t}^J,c_{\sigma,t}^J)\sim\mathcal{N}(\boldsymbol{\mu}^J,\Sigma^J)$ with $\boldsymbol{\mu}^J=(\mu_{J,r},\mu_{J,\sigma})^\top$ and

$$\Sigma^J=\begin{pmatrix}\sigma_{J,r}^2 & \blacklozenge\\ 0 & \sigma_{J,\sigma}^2\end{pmatrix}. \tag{2}$$

[8] Alizadeh, Brandt, and Diebold (2002), Andersen, Benzoni, and Lund (2002), Bakshi, Cao, and Chen (1997), Bates (2000), Chernov and Ghysels (2000), Chernov, Gallant, Ghysels, and Tauchen (2003), Duffie, Pan, and Singleton (2000), Eraker (2004), Eraker, Johannes, and Polson (2003), Heston (1993), Jones (2003), Pan (2002), Broadie, Chernov, and Johannes (2007), among other influential contributions.

Similarly, let the sizes of the co-jumps be normally distributed: $(c^{JJ}_{r,t}, c^{JJ}_{\sigma,t}) \sim \mathcal{N}(\boldsymbol{\mu}^{JJ}, \Sigma^{JJ})$ with $\boldsymbol{\mu}^{JJ} = (\mu_{JJ,r}, \mu_{JJ,\sigma})^\top$ 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 price discontinuities in times of higher (lower) diffusive volatility, an effect which, as shown, is coherent with the properties of the data.

4. (Time-varying leverage) The correlation between continuous shocks to prices and continuous shocks to variance ($-1 \leq \rho(\cdot) \leq 1$) is, once more, a flexible function of the underlying spot volatility level. In a stochastic volatility model with independent jumps only, Bandi and Renò (2012) show that leverage is a decreasing function of the spot volatility level. In other words, leverage becomes more negative as volatility 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 variance. It therefore seems 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. Below we expand on this issue.
5. (Possibly nonlinear structures) We do not express the price and variance drifts, the variance-of-variance, and the intensities of the jumps as affine (or, more generally, parametric) functions of the underlying state, but operate nonparametrically. Robustness to potential misspecifications may, in fact, be beneficial. For example, there is evidence that the variance-of-variance ($\Lambda^2(\cdot)$) could be a 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 nonlinear 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. In what follows, guided by the nonparametric estimates, we are explicit about the types of nonlinearities which appear to be in our data.

We conclude this section with two observations. First, some recent work has emphasized the potential importance of a multifactor variance specification (see, among others, Chernov, Gallant, Ghysels, and Tauchen, 2003; Dobrev and Szerszen, 2010). In our context, such a model could be expressed as

$$\begin{aligned} d\log(p_t) &= \mu(\sigma_t)\,dt + \sigma_t\,dW_{r,t} + c^J_{r,t}\,dJ_r + c^{JJ}_{r,t}\,dJ_{r,\sigma} \\ d\xi(\sigma^2_t) &= dx_t + c^J_{\sigma,t}\,dJ_\sigma + c^{JJ}_{\sigma,t}\,dJ_{r,\sigma}, \end{aligned} \tag{4}$$

where x_t is a sum of continuous Ito semimartingales potentially correlated with $W_{r,t}$. While estimation of the system in Eq. (4) may require, for instance, simulation methods beyond our scopes, independent jump and co-jump identification would proceed just like in the present paper. As we will make clear below, our proposed identification method for the co-jumps hinges on the use of higher-order infinitesimal cross-moments and these cross-moments are not impacted, theoretically, by the diffusive dynamics. To this extent, in what follows, we will also report co-jump estimates obtained by employing higher-order moments *only*. These estimates will be statistically and economically very similar to those derived from the model in Eq. (1) and, importantly, will attest to the robustness of the reported co-jump dynamics to a potentially higher number of variance factors.

Second, the only parametric assumption we make is on the probability distribution of the jump sizes. We emphasize that any parametric distribution with a finite number of parameters could be employed. In addition, the explicit dependence of the size parameters on the level of volatility (as in point 3 above) provides unconditional features which go beyond the assumed distributional features of the jump sizes. For instance, when working with Gaussian sizes, as we do below and support empirically, these time-varying parameters will readily deliver unconditional skewness and tail fatness, thereby capturing the effects described in Fig. 1. A comparison of Fig. 1 with Fig. 9 (below) illustrates these ideas.

We now turn to the identification of the system by virtue of an infinitesimal moment-based procedure.

4. Infinitesimal cross-moments

Let prices and variances evolve dynamically as in Eq. (1). We work with a logarithmic variance specification ($\xi(\sigma^2_t) = \log(\sigma^2_t)$) and Gaussian jumps (Section 9 provides a specification analysis which supports a logarithmic transformation for variance). The moments of the jump sizes ($\boldsymbol{\mu}^J$, $\boldsymbol{\mu}^{JJ}$, Σ^J, and Σ^{JJ}, using the notation in point 3 above) are assumed to depend on the volatility level.

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,

$$\vartheta_{p_1,p_2}(\sigma) = \lim_{\Delta \to 0} \frac{1}{\Delta} \mathrm{E}\left[(\log(p_{t+\Delta}) - \log(p_t))^{p_1} (\log(\sigma^2_{t+\Delta}) - \log(\sigma^2_t))^{p_2} | \sigma_t = \sigma \right]. \tag{5}$$

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}, for instance, 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 variance moments. Write

$$\vartheta_{0,1} = m + \vartheta^{Jump}_{0,1}, \tag{6}$$

$$\vartheta_{0,2} = \Lambda^2 + \vartheta^{Jump}_{0,2}, \tag{7}$$

$$\vartheta_{0,p_2} = \vartheta^{Jump}_{0,p_2}, \quad p_2 \geq 3 \tag{8}$$

with

$$\vartheta^{Jump}_{0,p_2} = \lambda_{r,\sigma} \sum_{j=0}^{p_2} \binom{p_2}{j} G_{0,j} (\sigma_{JJ,\sigma})^j (\mu_{JJ,\sigma})^{p_2 - j} + \lambda_\sigma \sum_{j=0}^{p_2} \binom{p_2}{j} G_{0,j} (\sigma_{J,\sigma})^j (\mu_{J,\sigma})^{p_2 - j},$$

where $G_{0,0} = 1$ and, for $g, g_1, g_2 \geq 1$, using the double factorial notation,

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

The expressions imply that

$$\vartheta_{0,1} = m + \lambda_{r,\sigma}\mu_{JJ,\sigma} + \lambda_\sigma \mu_{J,\sigma},$$
$$\vartheta_{0,2} = \Lambda^2 + \lambda_{r,\sigma}((\mu_{JJ,\sigma})^2 + (\sigma_{JJ,\sigma})^2) + \lambda_\sigma((\mu_{J,\sigma})^2 + (\sigma_{J,\sigma})^2),$$
$$\vartheta_{0,3} = \lambda_{r,\sigma}((\mu_{JJ,\sigma})^3 + 3(\mu_{JJ,\sigma})(\sigma_{JJ,\sigma})^2) + \lambda_\sigma((\mu_{J,\sigma})^3 + 3(\mu_{J,\sigma})(\sigma_{J,\sigma})^2),$$

and so forth. In other words, *all* infinitesimal variance moments depend on the corresponding 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:

$$\vartheta_{1,1} = \rho \Lambda \sigma + \vartheta^{Jump}_{1,1} \tag{9}$$

and

$$\vartheta_{1+p_1,1+p_2} = \vartheta^{Jump}_{1+p_1,1+p_2}, \quad p_1 > 1 \text{ or } p_2 > 1 \tag{10}$$

with

$$\vartheta^{Jump}_{p_1,p_2} = \lambda_{r,\sigma} \sum_{j_1=0}^{p_1} \sum_{j_2=0}^{p_2} \binom{p_1}{j_1} \binom{p_2}{j_2} G_{j_1,j_2} (\sigma_{JJ,r})^{j_1} (\sigma_{JJ,\sigma})^{j_2} (\mu_{JJ,r})^{p_1 - j_1} (\mu_{JJ,\sigma})^{p_2 - j_2}. \tag{11}$$

The cross-moment expressions imply, for instance, that

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

and

$$\vartheta_{2,2} = \lambda_{r,\sigma}\{(\mu_{JJ,\sigma})^2(\mu_{JJ,r})^2 + (\sigma_{JJ,\sigma})^2(\mu_{JJ,r})^2 + (\mu_{JJ,\sigma})^2(\sigma_{JJ,r})^2 + (1 + 2\rho_J^2)(\sigma_{JJ,r})^2(\sigma_{JJ,\sigma})^2 + 4\rho_J \mu_{JJ,r}\mu_{JJ,\sigma}\sigma_{JJ,r}\sigma_{JJ,\sigma}\}.$$

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, which we discuss below. The higher (than 1,1) cross-moments only depend on the corresponding cross-moments of the common jumps.

In essence, by setting $p_2 = 0$ and varying p_1 one can identify features specific to the price process. Conversely, by setting $p_1 = 0$ and varying p_2 one can identify features specific to the variance process. The genuine cross-moments with $p_1 \geq p_2 \geq 1$ are, however, required to identify anything that is common to the price and the variance process: Brownian leverage ρ, the intensity of the common jumps $\lambda_{r,\sigma}$, and the correlation between the co-jump sizes ρ_J.

Interestingly, sub-cases of the more general specification we consider may readily lead to *closed-form* formulations of the system's functions (μ, m, and so on) as "functions" of the infinitesimal cross-moments. Assume, for example, $c^J_{r,t} = c^{JJ}_{r,t}$,

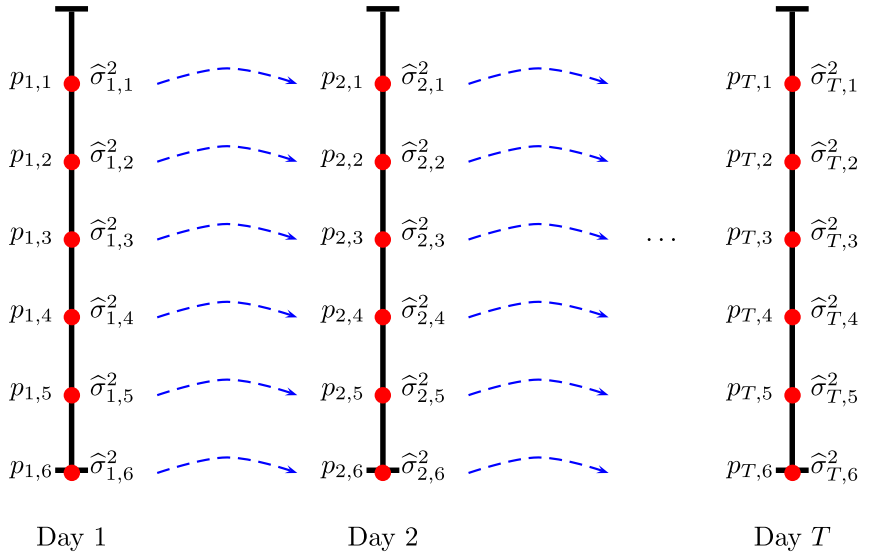

Fig. 4. Pictorial representation of the way in which we employ data in Eq. (12). The estimates $\widehat{\sigma}^2_{t,i}$ are computed with intradaily data between prices $p_{t,i-1}$ and $p_{t,i}$. The dashed arrows indicate the direction along which we compute logarithmic price and logarithmic variance differences. These differences are always daily.

$c^J_{\sigma,t} = c^{JJ}_{\sigma,t}$, and $\boldsymbol{\mu}^{JJ} = 0$ (same sizes across idiosyncratic and common jumps and zero means for all of the jump sizes). In this case, the moments $\vartheta_{p_1,0}$ with $p_1 = 1,2,4,6$, say, identify μ, $\sigma^2_{J,r}$, and $\lambda_r + \lambda_{r,\sigma}$; the moments ϑ_{0,p_2} with $p_2 = 1,2,4,6$, say, identify m, Λ, $\sigma^2_{J,\sigma}$, and $\lambda_\sigma + \lambda_{r,\sigma}$; and the moments $\vartheta_{1,1}$, $\vartheta_{2,2}$, and $\vartheta_{1,3}$, among other genuine cross-moments, identify ρ, $\lambda_{r,\sigma}$, and ρ_J. In fact, this is one of many, possible identification schemes resulting in closed-form dependencies between the model's functions and the employed infinitesimal cross-moments. For any volatility level, the model's functions would simply be the solution to a system of equations, those defining the stated infinitesimal cross-moments.

Based on the preliminary descriptive evidence provided earlier, it is important for us to differentiate between the sizes of the independent jumps and those of the co-jumps. It is also important to allow for jumps with nonzero means. Hence, we estimate the more general specification presented above and do not work with estimates of the system's functions expressed as closed-form functionals of the system's cross-moments. The cross-moments, however, lend themselves to (a form of) moment 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, dubbed Nonparametric Infinitesimal Method of Moments (NIMM), is applied to any chosen level in the 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 estimates as useful descriptive tools. Directed by the functional estimates, we then turn to a parametric evaluation of the full system's dynamics.

5. Continuous-time 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 equispaced times, defined as "knots," within each day. Assume availability of closing logarithmic prices ($\log(p_{t,i})$) and logarithmic spot variance estimates ($\log(\widehat{\sigma}^2_{t,i})$) over each day $t = 1,\ldots,T$ at each knot $i = 1,\ldots,N$. The generic cross-moment estimator $\widehat{\vartheta}_{p_1,p_2}(\sigma)$ is defined as

$$\widehat{\vartheta}_{p_1,p_2}(\sigma) = \frac{\sum_{t=1}^{T-1}\sum_{i=1}^{N}\mathbf{K}\left(\frac{\widehat{\sigma}_{t,i}-\sigma}{h}\right)(\log(p_{t+1,i})-\log(p_{t,i}))^{p_1}(\log(\widehat{\sigma}^2_{t+1,i})-\log(\widehat{\sigma}^2_{t,i}))^{p_2}}{\Delta\sum_{t=1}^{T}\sum_{i=1}^{N}\mathbf{K}\left(\frac{\widehat{\sigma}_{t,i}-\sigma}{h}\right)}, \tag{12}$$

where $\mathbf{K}(\cdot)$ is a kernel function, with properties listed in Appendix A, whose role is to guarantee proper conditioning at σ for a (vanishing, asymptotically) bandwidth h. In essence, then, the procedure is a two-step procedure. In the first step, spot variance is estimated. In the second step, kernel estimates of the cross-moments are put to work to identify the system's dynamics.

It is important to note that, in Eq. (12), we never "differentiate" variances and prices within the same day but always across subsequent days. Fig. 4 illustrates our sampling. In this sense, all of our moment estimates are daily ($\Delta = 1$).

In the following subsection, in addition to describing the estimation procedure for spot variance, we explain why it is beneficial to use N daily spot variance differences instead of a single difference ($N = 1$) of integrated daily variances.

5.1. Spot versus daily variance estimation

In the implementation of the estimator in Eq. (12), we use the full trading day (from 9:30 am to 4:00 pm) and employ $N=6$ knots (the first corresponding to 10:35 am, the sixth corresponding to 4:00 pm) separated by one hour and five minutes.

Define, now, one-minute logarithmic returns $r_{t,i,k}=\log(p_{t,i,k})-\log(p_{t,i,k-1})$, with $k=1,\ldots,65$, constructed over each hour and five minutes before a generic knot i on day t.[9] The spot variance estimates, for each knot i on each day t, are obtained by applying the jump-robust threshold bipower variation estimator to the previous 65 one-minute returns:

$$\widehat{\sigma}^2_{t,i}=\frac{65}{64-n_j}\varsigma_1^{-2}\sum_{k=2}^{65}|r_{t,i,k}||r_{t,i,k-1}|\mathbf{1}_{\{|r_{t,i,k}|\leq\theta_{t,i,k}\}}\mathbf{1}_{\{|r_{t,i,k-1}|\leq\theta_{t,i,k-1}\}}, \tag{13}$$

where $\mathbf{1}_{\{\cdot\}}$ is the indicator function, $\varsigma_1=0.7979$, $\theta_{t,i,k}$ is a suitable threshold, and n_j is the number of returns whose absolute value is greater than $\theta_{t,i,k}$ (Corsi, Pirino, and Renò, 2010). Even though, in order to span the trading day, the six knots are separated by 65 (rather than 60) minutes, we call these variance estimates "hourly" for conciseness. Alternative spot variance estimators, other than threshold bipower variation, may be used (see, e.g., the approach in Andersen, Dobrev, and Schaumburg, 2012). However, any chosen estimator 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).

Why local variance? The model and the moment estimates in Eq. (12) are written for a genuinely spot variance process. Even though traditional, the use of integrated daily estimates would likely induce attenuation effects in the estimation of the genuine cross-moments very much akin those determined by asynchronous effects in covariance estimation. Since, as explained earlier, the cross-moments are crucial to identify all aspects of the model which are common to the price and the variance series (the co-jumps, most importantly), we opt for a more local notion of variance than in the existing literature, one that is nicely afforded by our use of high-frequency data for identification. This said, the relative size of the measurement error in variance estimation increases with the degree of locality. In other words, variance measures integrated over a longer horizon are expected to be relatively less noisy than "less integrated" ones. When considered jointly, these observations point to a probable bias-variance trade-off in the identification of the cross-moments, and the final nonparametric estimates, as a function of the choice of locality. More local variance estimates should lead to smaller biases but less precise estimates.

The simulation evidence in Fig. 5 supports this intuition. We simulate 1,000 paths of the following model with co-jumps:

$$\begin{aligned} dX_t&=\sigma_t\,dW^1_t+c_X\,dJ_t,\\ d\log(\sigma^2_t)&=\eta\,dW^2_t+c_\sigma\,dJ_t, \end{aligned} \tag{14}$$

where W^1 and W^2 are independent Brownian motions, $\eta=0.6$, the intensity of the co-jumps λ is $50/252$, c_X and c_σ are normally distributed with zero mean and standard deviations 4 and 4η, respectively, and the co-jump size correlation ρ_J is -0.9. Each replication consists of 1,000 days. During each day, prices are sampled 360 times, mimicking one-minute returns over a six-hour trading day. Then, as a function of the number of intradaily knots, we back out the intensity of the co-jumps and the correlation of their jump sizes using the infinitesimal cross-moments of order 2,2 (for the intensity) and 3,1 (for ρ_J). In both cases, in order for us to simply focus on the impact of "frequency" on the final estimates, the parameters other than the intensity (when using the 2,2 cross-moment) or the correlation (when using the 3,1 cross-moment) are assumed to be fixed at their simulated value rather than being estimated. Estimation of all parameters would not modify the logic of the argument in this subsection.[10]

In Fig. 5, the circles represent the average intensity values (top panel) and co-jump correlations (bottom panel) obtained by virtue of the true integrated variances between knots (the higher the number of knots, the more local the spot variance values). The stars represent average values yielded by the estimated counterparts of the previous quantities (again, the higher the number of knots, the more local the spot variance estimates). The vertical bars are their 95% bounds across simulations. Finally, the crosses represent average values, across simulations, delivered by the true spot variances. In this case, the choice of the number of knots only translates into more, or fewer, terms being averaged in Eq. (12). It does not affect the degree of locality of variance since, at each set of knots, the true spot variances simulated from the model are used. The feasible case is the second case (corresponding to the stars). In practice, coherently with the representation in Fig. 4,

[9] 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, Li, Mykland, Podolskij, and Vetter, 2009).

[10] The simulation set-up in this section is designed to illustrate the impact of alternative variance estimates on the identification of the cross-moments and the model's functions. For this reason, it is purposely simple. When evaluating the accuracy of all of the model's functions, we simulate from our more general parametric specification in Section 7 (see Appendix C).

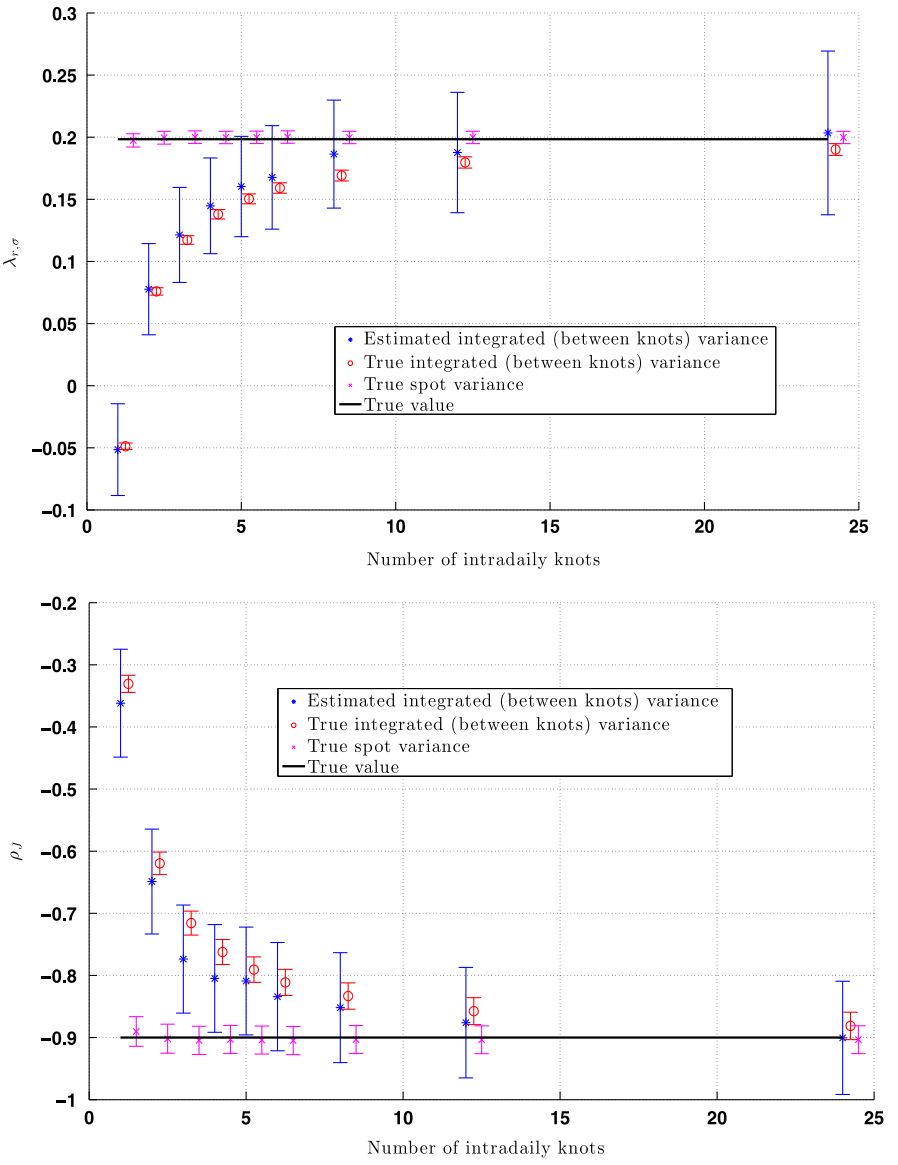

Fig. 5. The importance of the number of intradaily knots. We report the impact of the choice of the number of intradaily knots on the estimates of the co-jump intensity (top panel) and on the estimates of the co-jump correlation (bottom panel). The data is simulated using the model in Eq. (14), a simple specification yielding reasonable dynamics. In both figures, the black solid line is the true value, the exes are estimates based on the true (infeasible) spot variance, the circles are estimates based on the true (infeasible) integrated variance, and the stars are estimates based on the estimated integrated variance.

increasing the number of knots N translates into more local, than daily, variance estimates and a larger number of differences in the definition of the estimator in Eq. (12).

The use of daily ($N=1$) integrated variances, whether true or estimated, causes substantial attenuation in the estimated likelihood of co-jumps and their correlation. The use of more local estimates results in smaller biases but larger noise. Our choice of hourly variances ($N=6$) is, therefore, meant to balance the finite-sample MSEs of the resulting intensity and correlation estimates in a sensible way.

Existing estimation methods in the stochastic volatility literature, like likelihood methods, use the information in the cross-moments implicitly, something which we do explicitly. In addition, existing volatility filters are daily in nature. Coherently with this simulation evidence, it is the loss of granularity due to the use of integrated (daily) variance estimates which may have led to a body of ambiguous evidence about the presence of co-jumps, and the magnitude of their size correlation, when using price data alone. In the experiment above, we estimate an unreasonably negative co-jump intensity (top panel) and a correlation of -0.35 (bottom panel) when the true value is -0.9. In essence, overly integrated estimates

may drastically reduce the likelihood of co-jumps as well as the size of the estimated correlation between co-jump sizes, thereby justifying the uncorrelatedness usually found when employing price data recorded at daily frequencies (e.g., Eraker, Johannes, and Polson, 2003; Eraker, 2004). The use of more local variance estimates, afforded here by the availability and use of high-frequency price observations, appears to provide needed precision in the identification of infrequent, large co-variations.

Two observations are important. In a finite sample, the use of variance estimates which are more local in nature than daily integrated variance may be subject to intradaily U-shaped effects. To this extent, we recall that the estimator in Eq. (12) averages (scaled) differences of hourly variance measures across days (t is a daily index and, again, Δ is equal to one day). Since, as is typically found, the U-shape structure of intradaily variance effects is virtually unchanged between consecutive days in our sample, our final (daily) functional estimates are robust to this type of finite-sample effect. In addition, the spot variance estimates are scaled up to daily values so as to take into account the overnight returns. This is a typical adjustment to achieve robustness to biases due to the overnights.

We conclude this section by summarizing the logic behind our mixed frequency approach (minute-by-minute returns aggregated into 65-minute spot variance estimates in order to model daily dynamics).

Estimating the model entirely at high frequency would require the introduction, in the specification of the variance dynamics, of intraday effects superimposed on the assumed jump-diffusion structure. In order to fully capture high-frequency effects in the data, one would also have to carefully design specifications for market microstructure noise. By estimating the model at the one-day frequency after reducing the impact of noise, we may dispense with the need to model intradaily dynamics while still relying on accurate high-frequency estimates of spot variance. As a result, we focus on an horizon which is relevant for typical financial applications and provide estimates which can be directly compared to those in the existing stochastic volatility literature (for which the one-day frequency is typical).

The use of one-minute returns is needed to exploit the power of realized estimates in the evaluation of spot variance. We begin, in fact, with tick-by-tick data: consistent with pre-averaging, we aggregate returns over a one-minute interval solely to soften the impact of market microstructure noise.

The use of 65-minute variance estimates is motivated by a bias-variance trade-off, illustrated above. The lower the frequency, the larger the bias in estimating the quantities of interest (co-jumps intensity and correlation). The higher the frequency, the larger their variance. An intermediate choice was made to balance these two effects. While, theoretically, the optimal choice is necessarily a function of the unknown variance dynamics, our simulations show that splitting a 6.5-hour trading day into six intervals represents a sensible compromise between parsimony of the procedure and accuracy.

5.2. Cross-moment based identification: nonparametric infinitesimal method of moments (NIMM)

The spot variance estimates in Eq. (13), paired with the functional moment estimates in Eq. (12), are employed 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 study nonparametric identification of discontinuous stochastic volatility models under the assumption of independent jumps. Bandi and Renò (2012) 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 introduces co-jumps and focuses on an empirical treatment which highlights the statistical and economic relevance of discontinuous joint changes in prices and variance. In particular, the addition of co-jumps with nonzero 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 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, 2012) 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 next. The use of infinitesimal cross-moments to evaluate complex multivariate systems with discontinuities is, in our view, of general interest. We apply here to the evaluation of comovements in prices and variances. It could be employed to study a host of alternative problems, contagion, for instance.

Denote by $g_1(\sigma),\ldots,g_K(\sigma)$ the K functions driving the dynamics of the system (drifts, intensities, and so on). These are the estimation targets. Consider a set of S cross-moments $\widehat{\vartheta}_{p_1,p_2}(\sigma)$ with $S\geq K$ for identification. Each theoretical cross-moment $\vartheta_{p_1,p_2}(\sigma)=f_{p_1,p_2}(g_1(\sigma),\ldots,g_K(\sigma))$ is 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 $S\times S$ 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 asymptotic variance–covariance matrix of the cross-moments, i.e., the matrix with generic element

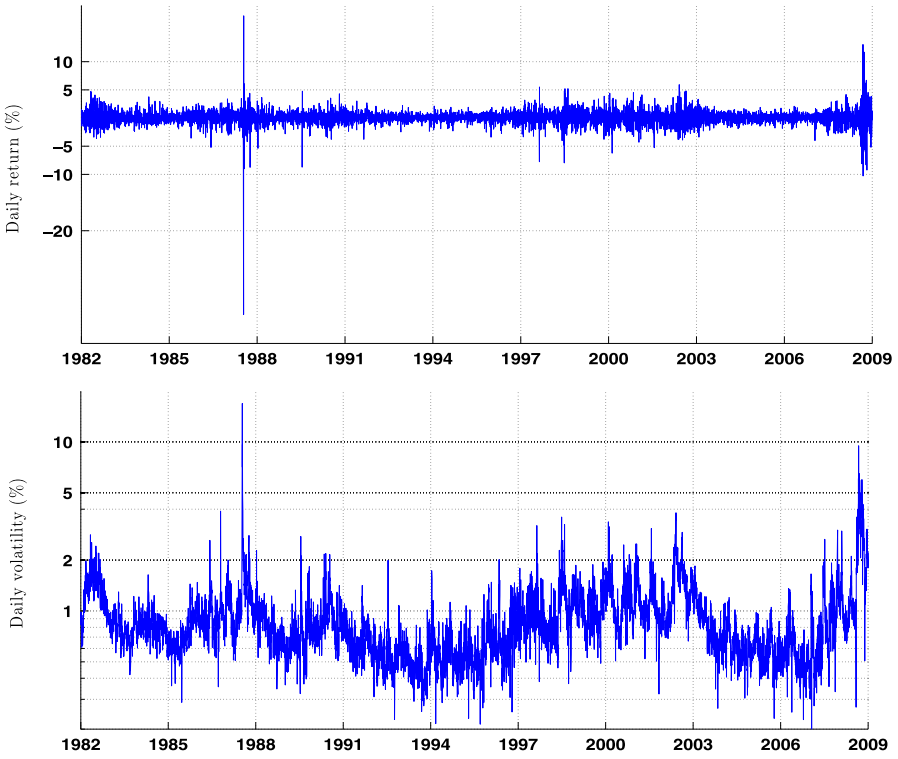

Fig. 6. Time series of daily returns and filtered daily volatilities. We use S&P500 intradaily futures prices from April 21, 1982 to February 5, 2009. Top graph: We report daily returns (%). Bottom graph: We report filtered (from intradaily prices) daily volatilities (in logarithmic scale, %) obtained by averaging hourly spot volatilities within each day.

$V_{(p_1,p_2),(p_3,p_4)}(\sigma) = f_{p_1+p_3,p_2+p_4}(\widehat{g}_1(\sigma),\ldots,\widehat{g}_K(\sigma))/h(\sigma)\widehat{L}_T(\sigma)$, where $h(\sigma)$ is a volatility-specific bandwidth and $\widehat{L}_T(\sigma)$ is an estimate of the volatility's occupation density.[11] The estimation is then repeated with $\widehat{W}(\sigma) = V^{-1}(\sigma)$. The same procedure is implemented for different values of σ leading to the nonparametric estimates presented in the next section.

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

$$\widehat{\boldsymbol{\eta}} = \arg\min_{\boldsymbol{\eta}} (\widehat{\vartheta}_{p_1,p_2} - \vartheta_{p_1,p_2}(\boldsymbol{\eta}))^{\top} W (\widehat{\vartheta}_{p_1,p_2} - \vartheta_{p_1,p_2}(\boldsymbol{\eta})),$$

where W is an $(S\times G)\times(S\times G)$ symmetrical and positive definite weighting matrix. To optimize the choice of W, and underweight relatively noisier moments, we evaluate the variance-covariance matrix (V) of the S moment estimates at the G volatility levels via simulation.[12] 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 $\theta_{t,i,k}$, 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. (13) above. Appendix C 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 April 21, 1982 to February 5, 2009, for a total of 6,748 trading days. The daily return data and the daily *filtered* (by virtue of high-frequency price data) logarithmic volatility series are reported in Fig. 6.

[11] For the structure of the asymptotic variance of the infinitesimal cross-moments used here, we refer the reader to Theorem 2 in Appendix A.

[12] The model is simulated by sampling six times intradaily. The infinitesimal moments are then computed by using the same "across days" averaging described in Fig. 4. As discussed in Section 5.1, this averaging leads to daily estimates while reducing the need for specifying intradaily dynamics.

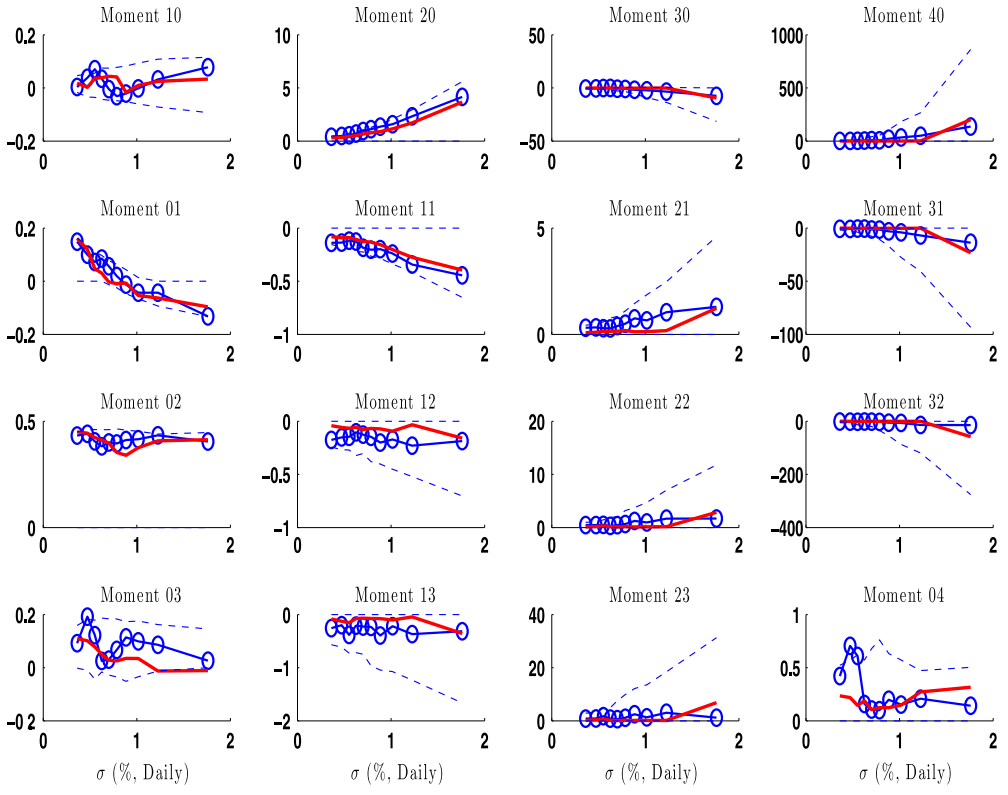

Fig. 7. The infinitesimal cross-moments. We report estimated infinitesimal cross-moments (solid lines with circles) with 95% confidence bands (dashed lines). The solid lines without circles are the implied *fitted* moments using functions, displayed in Fig. 8, obtained via pointwise NIMM. We use S&P500 intradaily futures prices from April 21, 1982 to February 5, 2009.

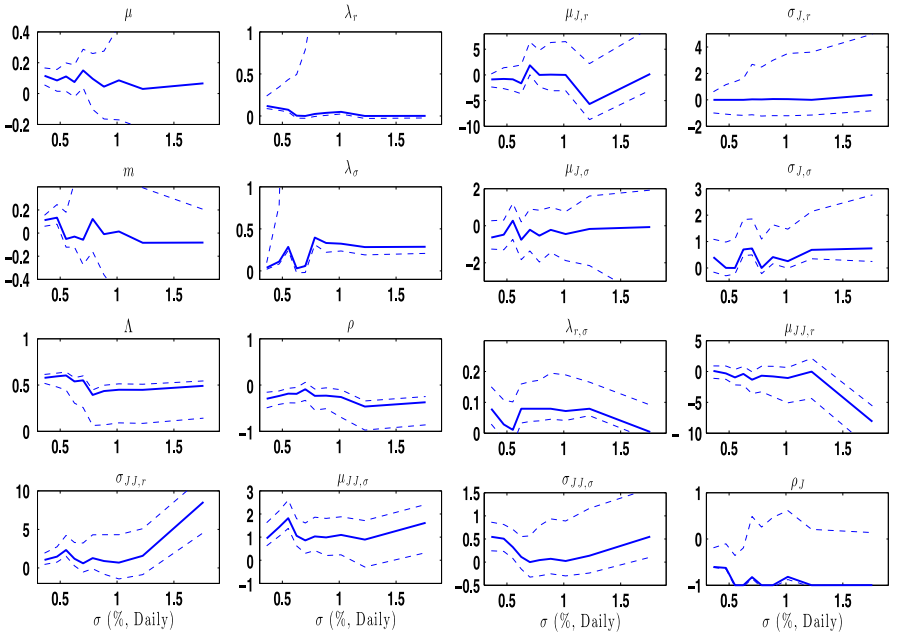

Fig. 8. The model's driving functions. We report estimated driving functions (solid lines) via NIMM with 95% confidence bands (dashed lines). We use S&P500 intradaily futures prices from April 21, 1982 to February 5, 2009.

Fig. 7 provides the estimated moments of order p_1, p_2 with $p_1+p_2 \leq 4$, plus the moments of order 2,3 and 3,2. These moments, which we employ for estimation, are sufficient for identification and are expected to be less noisy than their higher-order counterparts. The functions driving the price and variance dynamics in Eq. (1) are reported in Fig. 8. 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 (we return to this issue in Section 8). The inability to detect risk-return trade-offs with a (relatively) short sample period is a notorious phenomenon which we confirm here.[13] The intensity of the idiosyncratic jumps in prices λ_r is fairly stable, across volatility levels, and implies between roughly five and 15 jumps per year. As expected in this type of problem, the level of statistical uncertainty is substantial, particularly for 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 logarithmic variance. The volatility-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; 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 specification analysis in Section 9. The nonparametric inference presented here 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 variance jumps than independent price jumps, something which the 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 stable across median volatility levels (σ=0.8). When jumping jointly with prices, variance 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. 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, increasing with the volatility level. In sum, the estimated variance co-jump size distribution is fairly invariant to the underlying volatility level. The mean and the variance of the size distribution of the price co-jumps, on the other hand, become more negative and increase, respectively, as volatility increases. Importantly, the correlation between discontinuous price changes and discontinuous variance changes ρ_J is negative and very close to -1. As emphasized, the magnitude and sign of this correlation have been rather elusive in work that uses—like we do—price data alone (see, e.g., Broadie, Chernov, and Johannes, 2007). The economic and statistical strength of our reported finding, instead, clarifies that the negative co-jump correlation detected in earlier work employing option data (Eraker, 2004; Todorov and Tauchen, 2011) cannot solely depend on the negative covariance between price jumps and market risk premia. It appears to be a genuine feature of the market price process, one that effective derivative pricing and risk management should account for. We stress the importance of this additional source of return skewness in a specification which, differently from much of the literature, explicitly separates the jump components into idiosyncratic and common factors.[14]

We now turn to the most traditional source of skewness, Brownian leverage. As for other quantities, Fig. 8 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, 2012). In agreement with Bandi and Renò (2012), 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 (2012) appears to be robust to the presence of an additional source of skewness (further discussed below) induced by the joint occurrence of price and variance jumps.

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

[13] Andersen, Benzoni, and Lund (2002), Pan (2002), and Eraker, Johannes, and Polson (2003), among others, also find a statistically insignificant dependence of the price drift on volatility.

[14] This separation is deemed important by Duffie, Pan, and Singleton (2000) and defines their more general stochastic volatility model. Their more general model is, however, not estimated. Eraker, Johannes, and Polson (2003) and Eraker (2004) estimate specifications in which the jumps are either independent or perfectly dependent in terms of arrival times. In their framework, no allowance is made for both independent and common jumps.

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

$$
\begin{aligned}
&d\log(p_t) = \mu_r\, dt + \sigma_t\{\rho_t\, dW_t^1 + \sqrt{1-\rho_t^2}\, dW_t^2\} + c_{r,t}^{J}\, dJ_r + c_{r,t}^{JJ}\, dJ_{r,\sigma}\\
&d\log(\sigma_t^2) = (m_0 + m_1\log(\sigma_t^2))\, dt + \Lambda\, dW_t^1 + c_{\sigma,t}^{J}\, dJ_\sigma + c_{\sigma,t}^{JJ}\, dJ_{r,\sigma},\\
&\rho_t = \max(\min(\rho_0 + \rho_1\sigma_t, 1), -1),\\
&\{J_r, J_\sigma, J_{r,\sigma}\} \sim Poisson(\lambda_r, \lambda_\sigma, \lambda_{r,\sigma})\\
&c_{r,t}^{J} \sim \mathcal{N}(\mu_{J,r}, \sigma_{J,r}^2)\\
&c_{\sigma,t}^{J} \sim \mathcal{N}(\mu_{J,\sigma}, \sigma_{J,\sigma}^2)\\
&\begin{pmatrix} c_{r,t}^{JJ} \\ c_{\sigma,t}^{JJ} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_{JJ,r,0} + \mu_{JJ,r,1}\sigma_t \\ \mu_{JJ,\sigma} \end{pmatrix}, \quad \begin{pmatrix} (\sigma_{JJ,r,0} + \sigma_{JJ,r,1}\sigma_t^{\sigma_{JJ,r,2}})^2 & \blacklozenge \\ \rho_J(\sigma_{JJ,r,0} + \sigma_{JJ,r,1}\sigma_t^{\sigma_{JJ,r,2}})\sigma_{JJ,\sigma} & \sigma_{JJ,\sigma}^2 \end{pmatrix} \right).
\end{aligned}
\tag{15}
$$

The full system, which is readily viewed as a special case of the model in Eq. (1), has 21 parameters whose estimates are reported in column 5 of Table 2.[15] We estimate three relevant restricted models as well. The first restricted model sets $J_{r,\sigma}=0$ and has 12 parameters whose estimated values are reported under the title "no co-jumps" in column 3 of Table 2. The second specification sets $J_r=J_\sigma=0$ and has 15 parameters whose estimates are reported under the title "no independent jumps" in column 4 of Table 2. The third model is without idiosyncratic jumps ($J_r=J_\sigma=0$) and only uses higher-order moments ($p_1+p_2\geq 3$) for identification. For this latter model, which focuses on the co-jump part only and is theoretically robust to the specification of the diffusive part in Eq. (4), we report estimates in column 2 of Table 2. The statistical and economic similarity between these results and the results for the full system attest to the robustness of the co-jump dynamics to alternative (possibly multivariate) specifications for the diffusive variance.

We begin the discussion, again, with the price equation. Consistent with the nonparametric 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 $\mu_{r,1}$ close to zero, numerically and statistically. Hence, we only report estimates for a constant drift here. We return to a theoretical treatment of the compensation for variance risk in expected returns, and its often elusive detection in the data, in Section 8. 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 variance 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 variance. In particular, the parameter m_1 is equal to -0.0597 with a 95% confidence band equal to $(-0.075,\ -0.0381)$. The volatility of logarithmic variance Λ is precisely estimated at a constant value equal to about 0.558. The corresponding 95% confidence interval is (0.462, 0.583). The number of independent jumps in variance 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 $(-1.12, 0.17)$. 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 interval of (4.98, 22.45) 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 variance, 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 findings 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 (0.91, 1.51). The volatility of the common jumps in variance is estimated at a value of 0.1084 with a 95% confidence interval of (0, 0.409). To visualize the impact of these numbers, the size distributions of the price and variance co-jumps are represented in Fig. 9. 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. The implied unconditional distribution, also displayed in the figure, shows substantial kurtosis and some skewness, thereby departing from normality. As discussed, these effects are generated by means and variances that are functions of spot volatility. The effects are in line with the preliminary descriptive evidence in Fig. 1. Finally, importantly, the correlation between the common price jumps and the common variance jumps is estimated at a value which is very

[15] The confidence bands are obtained by simulation. The estimated parametric model in Table 2, column 5, is used to generate simulated samples as large as the original sample. The proposed estimation procedure is then applied to every simulated sample.

Table 2
Parametric estimates. In column 5 we report parametric estimates of the model in Eq. (15) with the corresponding 95% 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 reported in columns 3 and 4, respectively. In column 2 we display parametric co-jump estimates obtained by using only higher-order moments for identification. The parameters are relative to daily data, and imply returns expressed in percentage form. We use S&P500 intradaily futures prices from April 21, 1982 to February 5, 2009.

Parameter	Only co-jumps	No co-jumps	No independent jumps	Full model
μ_r	–	0.0423 (−0.0807,0.4736)	0.0631 (0.0530,0.1158)	0.0306 (−0.0302,0.0995)
ρ_0	–	−0.2280 (−0.3578,−0.1819)	−0.0977 (−0.1500,0.0222)	−0.0988 (−0.1824,0.0489)
ρ_1	–	−0.0874 (−0.1102,0.0283)	−0.1225 (−0.1921,−0.0560)	−0.1617 (−0.2594,−0.0832)
m_0	–	−0.0232 (−0.0395,−0.0019)	−0.0397 (−0.0865,−0.0241)	−0.0380 (−0.0910,0.0412)
m_1	–	−0.0704 (−0.0804,−0.0525)	−0.0576 (−0.0662,−0.0321)	−0.0597 (−0.0750,−0.0381)
Λ	–	0.6048 (0.5818,0.6128)	0.5950 (0.5674,0.6045)	0.5583 (0.4625,0.5830)
$\mu_{J,r}$	–	−0.1137 (−0.4050,0.1081)	–	1.3948 (−0.4353,2.9061)
$\mu_{JJ,r,0}$	−1.1333 (−2.5000,1.5944)	–	0.5210 (−0.4511,1.3991)	−0.0544 (−0.9019,1.1176)
$\mu_{JJ,r,1}$	−0.7022 (−4.0000,0.2566)	–	−1.8976 (−3.8597,−0.3171)	−1.0072 (−3.5152,0.1391)
$\sigma_{J,r}$	–	1.2715 (0.1553,1.8914)	–	0.6818 (0.0000,1.7859)
$\sigma_{JJ,r,0}$	0.7652 (0.0000,2.5921)	–	1.7428 (0.0000,2.0801)	0.6246 (0.0000,1.6674)
$\sigma_{JJ,r,1}$	0.8232 (0.0000,6.8987)	–	0.1718 (0.0000,3.1924)	2.2469 (0.9085,4.4756)
$\sigma_{JJ,r,2}$	3.4350 (0.0633,6.0732)	–	1.8828 (0.0000,4.8503)	1.0863 (0.5047,1.8921)
$\mu_{J,\sigma}$	–	0.3498 (−0.1487,0.8707)	–	−0.4497 (−1.1246,0.1742)
$\mu_{JJ,\sigma}$	1.2230 (0.2932,2.0986)	–	0.7816 (0.5484,0.8936)	1.4428 (0.9117,1.5126)
$\sigma_{J,\sigma}$	–	1.2575 (0.8452,1.5206)	–	0.7002 (0.0000,1.0035)
$\sigma_{JJ,\sigma}$	0.4479 (0.0476,1.2496)	–	0.4901 (0.1657,0.5299)	0.1084 (0.0000,0.4093)
ρ_J	−1.0000 (−1.0000,−0.3869)	–	−0.6416 (−1.0000,−0.7373)	−1.0000 (−1.0000,−0.2154)
λ_r	–	0.1033 (0.0023,3.2877)	–	0.0252 (0.0005,0.1515)
λ_σ	–	0.0279 (0.0117,0.0303)	–	0.0528 (0.0004,0.3352)
$\lambda_{r,\sigma}$	0.0297 (0.0000,0.0803)	–	0.0489 (0.0401,0.1081)	0.0339 (0.0198,0.0891)

clearly negative. The point estimate is at the lower boundary (−1) and the upper limit of the 95% confidence interval is −0.215.[16]

In sum, the estimates suggest that the price process is more likely than not to jump jointly with the variance process (the number of price co-jumps is larger than the number of independent price jumps). When moving jointly in a discontinuous fashion, variance shifts 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 Fig. 9). This is, again, consistent with the preliminary evidence in Section 2, Fig. 3. 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. Hence, the size distribution of the price co-jumps is fat-tailed and negatively

[16] 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. The measurement error will induce attenuation effects when evaluating the correlation between price and variance 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.

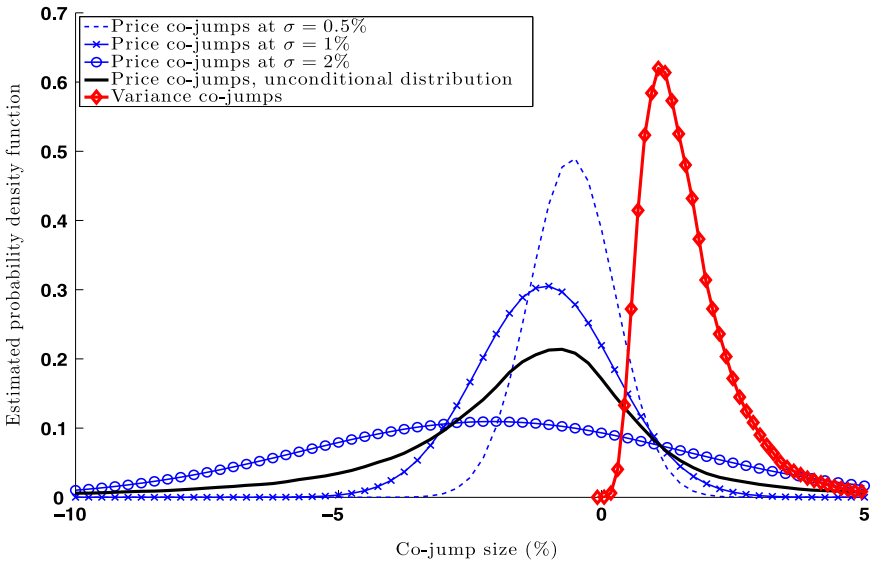

Fig. 9. The unconditional and conditional distributions of the price and variance co-jumps. We report unconditional and conditional distributions of the price and variance co-jumps implied by the parametric estimates in Table 2, column 5. The solid line with diamonds is the estimated unconditional distribution of the variance co-jumps. The dashed line, the solid line with exes and solid line with circles are the estimated distributions of the price co-jumps conditional on the volatility levels σ=0.5%, σ=1%, and σ=2%, respectively. The solid line is the estimated unconditional distribution of the price co-jumps. We use S&P500 intradaily futures prices from April 21, 1982 to February 5, 2009.

skewed. The size distribution of the variance co-jumps is, instead, positively skewed. The asymmetry is, in this case, the result of normally distributed jumps in logarithmic variance. Finally, when the variance 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.

Motivated, 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, Johannes, and Polson (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, Benzoni, and Lund (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 variance 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,

$$\rho_{total}(\sigma)=\frac{\vartheta_{1,1}}{\sigma\Lambda}=\rho+\frac{\lambda_{r,\sigma}(\rho_J\sigma_{JJ,r}\sigma_{JJ,\sigma}+\mu_{JJ,r}\mu_{JJ,\sigma})}{\sigma\Lambda}=\rho+\rho_{co-jumps}. \tag{16}$$

The standardization is given by spot volatility times the volatility of logarithmic variance and is designed to highlight ρ. Hence, $\rho_{co\text{-}jumps}$ is not, contrary to ρ, an actual correlation. Even though the number of yearly co-jumps $\lambda_{r,\sigma}$ may not be extremely high, the standardized covariance 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. It is, in fact, worth observing that the co-jump leverage derives from two sources: the product of the expected sizes ($\mu_{JJ,r}\mu_{JJ,\sigma}$) and the covariance between the co-jump sizes ($\rho_J\sigma_{JJ,r}\sigma_{JJ,\sigma}$). A negative $\mu_{JJ,r}$ and a positive $\mu_{JJ,\sigma}$, as in our data, would lead to a form of negative co-jump leverage even if ρ_J were zero (for example, in the extreme case of purely deterministic jumps).

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 Fig. 10) that ρ is decreasing with the volatility level while $\rho_{co\text{-}jumps}$ is nearly constant. Fig. 10 plots the percentage of 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, Fan, and Li, 2013; Bandi and Renò, 2012). Specifically, referring back to Eq. (16), compelling statistical arguments have been made about the need to account for noise in the estimation of the variance-of-variance Λ^2 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, to the belief that Brownian leverage may, in fact, be

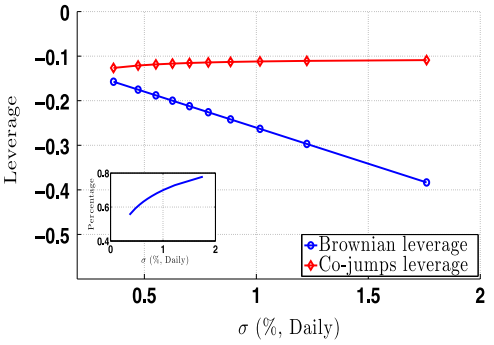

Fig. 10. Leverage effect. We report leverage, as a function of volatility, decomposed into its Brownian part (solid line with circles) and the part due to co-jumps (solid line with diamonds), as implied by the parametric estimates in Table 2, column 5. In the inset, we plot the percentage of ρ_{total} (see text) explained by the Brownian correlation. We use S&P500 intradaily futures prices from April 21, 1982 to February 5, 2009.

larger—i.e., more negative—than previously recognized. We address these statistical concerns explicitly by relying on *spot* (hourly) variance measures as well as by reducing the residual measurement error in the spot variance measures explicitly (see Appendix B). Monte Carlo simulations from the parametric model in Table 2 (Fig. C1, Appendix) show that our small-sample adjustments are reliable in generating unbiased estimates 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, offers further, more direct, evidence. Following our small-sample adjustments, this evidence attests to minimal distortions in estimating both ρ_0 and ρ_1 (Fig. C2, Appendix, for a visual representation).

This paper, however, 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 variance changes is due to common jumps, the presence of anti-correlated co-jumps will, as was shown by virtue of the decomposition illustrated in Fig. 10, lead to smaller Brownian correlations.

To conclude this section, we now ask 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, columns 3 and 4, respectively). Both specifications, implying the presence of independent or common jumps *only*, represent important models in the literature (see, e.g., Eraker, Johannes, and Polson, 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 variance 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 seven yearly variance 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 (-0.64). 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. The pricing of co-jumps: economic implications

We placed particular emphasis on the substantial negative dependence between price and variance co-jumps. We now turn to some of its economic implications. Specifically, we show that not only do negatively correlated co-jumps add needed flexibility to the empirical dynamics of the price process, but they also represent a valuable economic channel leading to broader notions of return and variance risk premia.

We assume that the stock price follows the model in Eq. (1). We also assume that there exists a pricing kernel M_t taking the form

$$M_t = M_0 \left(\frac{p_t}{p_0}\right)^{\phi} \exp^{\int_0^t \delta(\sigma_s)\, ds + \psi(\xi(\sigma_t^2) - \xi(\sigma_0^2))}, \tag{17}$$

where ϕ and ψ are parameters controlling aversion to price risk and variance risk, respectively, while $\delta(\cdot)$ is a function controlling time preferences. A similar pricing kernel has been described by the recent work of Christoffersen, Heston, and Jacobs (2013) as being able to reconcile a variety of empirical facts about option prices, from the implied volatility puzzle (leading to a negative price of variance risk and implied volatilities that are higher than the observed, realized volatilities) to the difficulty in explaining out-of-the money option prices. The assumed pricing kernel is monotonically decreasing in

prices when $\phi<0$ (keeping variance constant) and monotonically increasing in variance when $\psi>0$ (keeping prices constant). However, since high prices are often associated with high variance, the pricing kernel may have a U-shape when projected onto prices. We will later return to the nonmonotonicity of the pricing kernel.

The following proposition provides a closed-form representation of return and variance risk premia in the presence of the nonparametric price dynamics in Eq. (1) and the pricing kernel in Eq. (17). The proposition clarifies the role of leverage (whether induced by continuous shocks or co-jumps) and uncertainty (whether price or variance uncertainty) in determining risk compensations. Below, we use the notation $\mathrm{E}_{\sigma_t}[\cdot]$ to signify $\mathrm{E}[\cdot|\sigma_t]$.

Proposition 8.1. Denote by r the risk-free rate and write the return risk premium as

$$\tilde{\mu}(\sigma_t)=\mu(\sigma_t)-r+\tfrac{1}{2}\sigma_t^2.$$

Given Eq. (17), absence of arbitrage implies

$$\begin{aligned}\delta(\sigma_t)=&-r(1+\phi)-\phi\left(\tilde{\mu}(\sigma_t)-\tfrac{1}{2}\sigma_t^2\right)-\psi m(\sigma_t)-\tfrac{1}{2}\left(\phi^2\sigma_t^2+\psi^2\Lambda^2(\sigma_t)+2\phi\psi\rho(\sigma_t)\sigma_t\Lambda(\sigma_t)\right)\\&-\lambda_{r,\sigma}(\sigma_t)\mathrm{E}_{\sigma_t}[\exp^{\phi c^{JJ}_{r,t}+\psi c^{JJ}_{\sigma,t}}-1]-\lambda_r(\sigma_t)\mathrm{E}_{\sigma_t}[\exp^{\phi c^{J}_{r,t}}-1]-\lambda_\sigma(\sigma_t)\mathrm{E}_{\sigma_t}[\exp^{\psi c^{J}_{\sigma,t}}-1],\end{aligned}\tag{18}$$

and

$$\tilde{\mu}(\sigma_t)=-\phi\sigma_t^2-\psi\rho(\sigma_t)\sigma_t\Lambda(\sigma_t)-\lambda_{r,\sigma}(\sigma_t)\mathrm{E}_{\sigma_t}[\exp^{\phi c^{JJ}_{r,t}+\psi c^{JJ}_{\sigma,t}}(\exp^{c^{JJ}_{r,t}}-1)]-\lambda_r(\sigma_t)\mathrm{E}_{\sigma_t}[\exp^{\phi c^{J}_{r,t}}(\exp^{c^{J}_{r,t}}-1)].\tag{19}$$

Further, assume that the price process' risk-neutral dynamics are given by

$$\begin{aligned}d(\log p_t)&=\left(r-\tfrac{1}{2}\sigma_t^2-\lambda^*_{r,\sigma}(\sigma_t)\mathrm{E}_{\sigma_t}\left[c^{JJ*}_{r,t}\right]-\lambda^*_r(\sigma_t)\mathrm{E}_{\sigma_t}\left[c^{J*}_{r,t}\right]\right)dt+\sigma_t\{\rho(\sigma_t)\,dW^{1*}_t+\sqrt{1-\rho^2(\sigma_t)}\,dW^{2*}_t\}+c^{JJ*}_{r,t}\,dJ^*_{r,\sigma}+c^{J*}_{r,t}\,dJ^*_r\\d\xi(\sigma_t^2)&=m^*(\sigma_t)\,dt+\Lambda(\sigma_t)\,dW^{1*}_t+c^{JJ*}_{\sigma,t}\,dJ^*_{r,\sigma}+c^{J*}_{\sigma,t}\,dJ^*_\sigma,\end{aligned}$$

where $J^*=\{J^*_r,J^*_\sigma,J^*_{r,\sigma}\}$ is an independent (of the bivariate vector $W^*=\{W^{1*}_t,W^{2*}_t\}$) trivariate vector of mutually independent Poisson jumps with risk-neutral intensities $\lambda^*_r(\sigma_t)$, $\lambda^*_\sigma(\sigma_t)$, and $\lambda^*_{r,\sigma}(\sigma_t)$, respectively. The risk-neutral jump sizes are $c^{J*}_{r,t}$, $c^{J*}_{\sigma,t}$, $c^{JJ*}_{r,t}$, and $c^{JJ*}_{\sigma,t}$. We have

$$m^*(\sigma_t)-m(\sigma_t)=\phi\rho(\sigma_t)\sigma_t\Lambda(\sigma_t)+\psi\Lambda^2(\sigma_t)\tag{20}$$

and, for all $u\in\mathbb{R}$ and all attainable σ_t values,

$$\lambda^*_{r,\sigma}(\sigma_t)=\lambda_{r,\sigma}(\sigma_t)\mathrm{E}_{\sigma_t}\left[e^{\phi c^{JJ}_{r,t}+\psi c^{JJ}_{\sigma,t}}\right],\quad \mathrm{E}_{\sigma_t}\left[e^{iuc^{JJ*}_{\sigma,t}}\right]=\frac{\mathrm{E}_{\sigma_t}[e^{iuc^{JJ}_{\sigma,t}}(e^{\phi c^{JJ}_{r,t}+\psi c^{JJ}_{\sigma,t}})]}{\mathrm{E}_{\sigma_t}[e^{\phi c^{JJ}_{r,t}+\psi c^{JJ}_{\sigma,t}}]},\tag{21}$$

$$\mathrm{E}_{\sigma_t}\left[e^{iuc^{JJ*}_{r,t}}\right]=\frac{\mathrm{E}_{\sigma_t}[e^{iuc^{JJ}_{r,t}}(e^{\phi c^{JJ}_{r,t}+\psi c^{JJ}_{\sigma,t}})]}{\mathrm{E}_{\sigma_t}[e^{\phi c^{JJ}_{r,t}+\psi c^{JJ}_{\sigma,t}}]},\tag{22}$$

$$\lambda^*_r(\sigma_t)=\lambda_r(\sigma_t)\mathrm{E}_{\sigma_t}\left[e^{\phi c^{J}_{r,t}}\right],\quad \mathrm{E}_{\sigma_t}\left[e^{iuc^{J*}_{r,t}}\right]=\frac{\mathrm{E}_{\sigma_t}[e^{iuc^{J}_{r,t}}(e^{\phi c^{J}_{r,t}})]}{\mathrm{E}_{\sigma_t}[e^{\phi c^{J}_{r,t}}]},\tag{23}$$

$$\lambda^*_\sigma(\sigma_t)=\lambda_\sigma(\sigma_t)\mathrm{E}_{\sigma_t}\left[e^{\psi c^{J}_{\sigma,t}}\right],\quad \mathrm{E}_{\sigma_t}\left[e^{iuc^{J*}_{\sigma,t}}\right]=\frac{\mathrm{E}_{\sigma_t}[e^{iuc^{J}_{\sigma,t}}(e^{\psi c^{J}_{\sigma,t}})]}{\mathrm{E}_{\sigma_t}[e^{\psi c^{J}_{\sigma,t}}]}.\tag{24}$$

Proposition 8.1 outlines the restrictions on the coefficients $\delta(\cdot)$, ϕ, and ψ induced by absence of arbitrage along with the risk-neutral transformations compatible with the pricing kernel in Eq. (17).

Eq. (18) describes time preferences. It is convenient to interpret it in light of the Black-Scholes model for which $\sigma_t=\sigma$ and $\mu(\sigma)=\overline{\mu}-\sigma^2/2$ are constants and there are no jumps. In this case, it is known that the pricing kernel takes the form

$$M^{BS}_t=M_0\exp^{-(r+\lambda^2/2)t-\lambda W_t}$$

with $\lambda=(\overline{\mu}-r)/\sigma$, thereby yielding

$$M^{BS}_t=M_0\left(\frac{p_t}{p_0}\right)^{-\lambda/\sigma}\exp^{(-r(1-\lambda/\sigma)+\lambda^2/2-\lambda\sigma/2)t}.$$

Thus, in the Black-Scholes model,

$$\phi_{BS}=-\left(\frac{\overline{\mu}-r}{\sigma^2}\right)=-\frac{1}{2}-\left(\frac{\mu-r}{\sigma^2}\right),\tag{25}$$

$\psi_{BS}=0$, and

$$\delta_{BS}=-r(1+\phi_{BS})+\tfrac{1}{2}\phi^2_{BS}\sigma^2+\tfrac{1}{2}\phi_{BS}\sigma^2,$$

which represents a simplified version of Eq. (18) when taking into account that, in this case,

$$\tilde{\mu}(\sigma) = \overline{\mu} - r = -\phi_{BS}\sigma^2. \tag{26}$$

Our more general specification includes an additional adjustment for the stochastic variance mean ($-\psi m(\sigma_t)$), extra components generated by the volatility of stochastic variance and leverage effects ($\psi^2\Lambda^2(\sigma_t)+2\phi\psi\rho(\sigma_t)\sigma_t\Lambda(\sigma_t)$), as well as adjustments due to the means of the three jump components.[17]

Eq. (19) describes the return risk premium $\tilde{\mu}(\sigma_t)$. We now provide a "beta" representation. To this extent, we note that

$$-\frac{1}{dt}\mathrm{E}_{\sigma_t}\left[\frac{dM_t^c}{M_t}\frac{dp_t^c}{p_t}\right] = \underbrace{-\phi\sigma_t^2}_{-(1/dt)\mathrm{E}_{\sigma_t}[(dM_t^{c,p}/M_t)dp_t^c/p_t]} \underbrace{-\psi\rho(\sigma_t)\Lambda(\sigma_t)\sigma_t}_{-(1/dt)\mathrm{E}_{\sigma_t}[(dM_t^{c,\xi(\sigma^2)}/M_t)dp_t^c/p_t]}, \tag{27}$$

where $dM_t^{c,p}/M_t$ denotes the price shock component of the evolution of the stochastic discount factor, whereas $dM_t^{c,\xi(\sigma^2)}/M_t$ denotes the corresponding variance shock component. Since, given $\phi<0$, positive price shocks are associated with lower stochastic discounting and negative price shocks are associated with higher stochastic discounting, the contribution of the first term to the risk premium ($-\phi\sigma_{t^2}$) is positive. This term is classical (cf., Eq. (26)). Given $\psi>0$, the term $-\psi\rho(\sigma_t)\Lambda(\sigma_t)\sigma_t$ also contributes positively to the return risk premium provided $\rho(\sigma_t)<0$, a well-known stylized fact further illustrated above. The logic is as follows: positive price shocks are associated with low variance (given $\rho(\sigma_t)<0$) and low discounting (if $\psi>0$). Similarly, negative price shocks lead to high variance and high discounting. Here, in agreement with the documented relation between leverage and volatility (Bandi and Renò, 2012), we also let leverage be a function of the underlying spot volatility. Importantly, however, $\rho(\sigma_t)$ only captures one component of leverage, i.e., diffusive leverage, the second component being associated with the co-jumps. We now turn to the co-jumps and their contribution to the return risk premium as formalized in Eq. (19).

Writing $\mu^{JJ} = -\lambda_{r,\sigma}(\sigma_t)\mathrm{E}_{\sigma_t}[\exp^{\phi c_{r,t}^{JJ}+\psi c_{\sigma,t}^{JJ}}(\exp^{c_{r,t}^{JJ}}-1)]$, we have

$$\begin{aligned}\mu^{JJ} &= -\lambda_{r,\sigma}(\sigma_t)\mathrm{E}_{\sigma_t}\left[\left(\exp^{\phi c_{r,t}^{JJ}+\psi c_{\sigma,t}^{JJ}}+1-1\right)\left(\exp^{c_{r,t}^{JJ}}-1\right)\right]\\ &= -\lambda_{r,\sigma}(\sigma_t)\mathrm{E}_{\sigma_t}\left[\left(\exp^{\phi c_{r,t}^{JJ}+\psi c_{\sigma,t}^{JJ}}-1\right)\left(\exp^{c_{r,t}^{JJ}}-1\right)\right]-\lambda_{r,\sigma}(\sigma_t)\mathrm{E}_{\sigma_t}\left[\exp^{c_{r,t}^{JJ}}-1\right]\\ &= \underbrace{-\frac{1}{dt}\mathrm{E}_{\sigma_t}\left[\frac{dM_t^{JJ}}{M_t}\frac{dp_t^{JJ}}{p_t}\right]}_{co-jump\ risk\ premium}\underbrace{-\frac{1}{dt}\mathrm{E}_{\sigma_t}\left[\frac{dp_t^{JJ}}{p_t}\right]}_{compensation},\end{aligned}$$

where M_t^{JJ} and p^{JJ} denote the co-jump part of the pricing kernel and the price process, respectively. The same considerations apply to the idiosyncratic jump part.

To gain intuition about the determinants of the return risk premium induced by co-jumps, let us approximate it using a first-order expansion (for small jump sizes):

$$-\lambda_{r,\sigma}(\sigma_t)\mathrm{E}_{\sigma_t}[(\exp^{\phi c_{r,t}^{JJ}+\psi c_{\sigma,t}^{JJ}}-1)(\exp^{c_{r,t}^{JJ}}-1)] \approx -\phi\lambda_{r,\sigma}(\sigma_t)\mathrm{E}_{\sigma_t}[(c_{r,t}^{JJ})^2]-\psi\lambda_{r,\sigma}(\sigma_t)\mathrm{E}_{\sigma_t}[c_{r,t}^{JJ}c_{\sigma,t}^{JJ}],$$

where the first term represents the positive (since $\phi<0$) contribution of the price co-jump variance and the second term represents the positive (since $\psi>0$) contribution of the co-jump leverage (which was found to be negative), thereby leading to a *positive* contribution of the co-jumps to the overall return risk premium. We note that, consistent with the leverage effect in Eq. (16), the risk premium due to co-jumps depends on $\lambda_{r,\sigma}(\sigma_t)\mathrm{E}_{\sigma_t}[c_{r,t}^{JJ}c_{\sigma,t}^{JJ}]$ instead of the co-jump covariance. This effect is intuitive, since deterministic jump sizes could also lead to leverage and a risk compensation.

In the co-jump case, large negative price shocks are associated with high discounting through lower prices (since $\phi<0$) and higher variance (since $\psi>0$). Hence, there are two "approximate"[18] channels through which the co-jumps enter return risk premia: prices, via ϕ, and variance, via ψ. The first effect is the co-jump analogue of $-\phi\sigma_t^2$, the second effect is the co-jump analogue of $-\psi\rho(\sigma_t)\Lambda(\sigma_t)\sigma_t$.

Remark 1 (*The impact of co-jumps on the return risk premium*). In sum, generalized leverage, as given by diffusive leverage and leverage induced by the co-jumps, contributes to expected returns and return risk premia provided $\psi>0$. Hence, not only do the co-jumps lead to an additional source of skewness in stock returns, but they also modify—in equilibrium—the mean of the return distribution by virtue of their corresponding risk compensation.

Analogous considerations apply to the variance risk premium which is composed of a continuous drift part and a portion due to jumps. Specifically, the total variance risk premium is given by

$$(m^*(\sigma_t)+\lambda_{r,\sigma}^*(\sigma_t)\mathrm{E}_{\sigma_t}[c_{\sigma,t}^{JJ*}]+\lambda_{\sigma}^*(\sigma_t)\mathrm{E}_{\sigma_t}[c_{\sigma,t}^{J*}])-(m(\sigma_t)+\lambda_{r,\sigma}(\sigma_t)\mathrm{E}_{\sigma_t}[c_{\sigma,t}^{JJ}]+\lambda_{\sigma}(\sigma_t)\mathrm{E}_{\sigma_t}[c_{\sigma,t}^{J}]). \tag{28}$$

[17] The mapping between the Black-Scholes model and our specification is clear when taking into account that the drift in Eq. (1) can be expressed as follows: $\mu(\sigma_t)=\overline{\mu}(\sigma_t)-\frac{1}{2}\sigma_t^2$. This implies that $\tilde{\mu}(\sigma_t)$ is, as stated in Proposition 8.1, a return risk premium.

[18] The qualifier "approximate" is used because of the first-order expansion used to justify the statement.

Eqs. (20) (through 24) illustrate its components. Using the properties of the measure change provided by M_t, we can write

$$\underbrace{\phi\rho(\sigma_t)\sigma_t\Lambda(\sigma_t)+\psi\Lambda^2(\sigma_t)}_{m^*(\sigma_t)-m(\sigma_t)}+\underbrace{\lambda_{r,\sigma}(\sigma_t)\mathrm{E}_{\sigma_t}[(\exp^{\phi c^{JJ}_{r,t}+\psi c^{JJ}_{\sigma,t}}-1)c^{JJ}_{\sigma,t}]}_{\lambda^*_{r,\sigma}(\sigma_t)\mathrm{E}_{\sigma_t}[c^{JJ*}_{\sigma,t}]-\lambda_{r,\sigma}(\sigma_t)\mathrm{E}_{\sigma_t}[c^{JJ}_{\sigma,t}]}+\underbrace{\lambda_{\sigma}(\sigma_t)\mathrm{E}_{\sigma_t}[(\exp^{\psi c^{J}_{\sigma,t}}-1)c^{J}_{\sigma,t}]}_{\lambda^*_{\sigma}(\sigma_t)\mathrm{E}_{\sigma_t}[c^{J*}_{\sigma,t}]-\lambda_{\sigma}(\sigma_t)\mathrm{E}_{\sigma_t}[c^{J}_{\sigma,t}]}.$$

In fact,

$$\mathrm{E}_{\sigma_t}[dJ^*_{r,\sigma}]=\mathrm{E}_{\sigma_t}[dJ_{r,\sigma}\exp^{\phi c^{JJ}_{r,t}+\psi c^{JJ}_{\sigma,t}}]=\lambda_{r,\sigma}(\sigma_t)\mathrm{E}_{\sigma_t}[c^{JJ}_{\sigma,t}\exp^{\phi c^{JJ}_{r,t}+\psi c^{JJ}_{\sigma,t}}] \tag{29}$$

$$\overset{implies}{\Rightarrow}\lambda^*_{r,\sigma}(\sigma_t)\mathrm{E}_{\sigma_t}[c^{JJ*}_{\sigma,t}]-\lambda_{r,\sigma}(\sigma_t)\mathrm{E}_{\sigma_t}[c^{JJ}_{\sigma,t}]=\lambda_{r,\sigma}(\sigma_t)\mathrm{E}_{\sigma_t}[c^{JJ}_{\sigma,t}(\exp^{\phi c^{JJ}_{r,t}+\psi c^{JJ}_{\sigma,t}}-1)] \tag{30}$$

and

$$\mathrm{E}_{\sigma_t}[dJ^*_{\sigma}]=\mathrm{E}_{\sigma_t}[dJ_{\sigma}\exp^{\psi c^{J}_{\sigma,t}}]=\lambda_{\sigma}(\sigma_t)\mathrm{E}_{\sigma_t}[c^{J}_{\sigma,t}\exp^{\psi c^{J}_{\sigma,t}}] \tag{31}$$

$$\overset{implies}{\Rightarrow}\lambda^*_{\sigma}(\sigma_t)\mathrm{E}_{\sigma_t}[c^{J*}_{\sigma,t}]-\lambda_{\sigma}(\sigma_t)\mathrm{E}_{\sigma_t}[c^{J}_{\sigma,t}]=\lambda_{\sigma}(\sigma_t)\mathrm{E}_{\sigma_t}[c^{J}_{\sigma,t}(\exp^{\psi c^{J}_{\sigma,t}}-1)]. \tag{32}$$

The term $m^*(\sigma_t)-m(\sigma_t)$ can be interpreted similarly to Eq. (27) since

$$\frac{1}{dt}\mathrm{E}_{\sigma_t}\left[\frac{dM^{c,\xi(\sigma^2)}_t}{M_t}d\xi(\sigma^2_t)\right]=\psi\Lambda^2(\sigma_t)\quad\text{and}\quad\frac{1}{dt}\mathrm{E}_{\sigma_t}\left[\frac{dM^{c,p}_t}{M_t}d\xi(\sigma^2_t)\right]=\phi\rho(\sigma_t)\sigma_t\Lambda(\sigma_t).$$

In other words, price uncertainty enters the variance risk premium through leverage while variance uncertainty enters the variance risk premium through the variance-of-variance. Analogously,

$$\lambda^*_{r,\sigma}(\sigma_t)\mathrm{E}_{\sigma_t}\left[c^{JJ*}_{\sigma,t}\right]-\lambda_{r,\sigma}(\sigma_t)\mathrm{E}_{\sigma_t}\left[c^{JJ}_{\sigma,t}\right]=\frac{1}{dt}\mathrm{E}_{\sigma_t}\left[\frac{dM^{JJ}_t}{M_t}d\xi^{JJ}(\sigma^2_t)\right],$$

and

$$\lambda^*_{\sigma}(\sigma_t)\mathrm{E}_{\sigma_t}\left[c^{J*}_{\sigma,t}\right]-\lambda_{\sigma}(\sigma_t)\mathrm{E}_{\sigma_t}\left[c^{J}_{\sigma,t}\right]=\frac{1}{dt}\mathrm{E}_{\sigma_t}\left[\frac{dM^{J}_t}{M_t}d\xi^{J}(\sigma^2_t)\right].$$

With $\psi>0$, large idiosyncratic variance shocks require a *negative* variance premium (since $(1/dt)\mathrm{E}_{\sigma_t}[(dM^J_t/M_t)\,d\xi^J(\sigma^2_t)]>0$). We also expect the variance co-jumps to yield a negative variance premium if the correlation between the price co-jumps and the variance co-jumps is negative. Using, again, a first-order expansion (for small jump sizes), we have

$$\mathrm{E}_{\sigma_t}[c^{JJ}_{\sigma,t}(\exp^{\phi c^{JJ}_{r,t}+\psi c^{JJ}_{\sigma,t}}-1)]\approx\phi\mathrm{E}_{\sigma_t}[c^{JJ}_{r,t}c^{JJ}_{\sigma,t}]+\psi\mathrm{E}_{\sigma_t}[(c^{JJ}_{\sigma,t})^2]>0,$$

leading to a *positive* contribution of the co-jumps to the variance risk premium when $\phi<0$, $\mathrm{E}_{\sigma_t}[c^{JJ}_{r,t}c^{JJ}_{\sigma,t}]<0$ and $\psi>0$.

In this case, large positive variance shocks are associated with high discounting through higher variance (since $\psi>0$) and lower prices (since $\phi<0$). As in the case of the price equation, there are "approximately" two channels through which the co-jumps enter variance risk premia: variances, via ψ, and prices, via ϕ. The first effect is the co-jump analogue of $\psi\Lambda^2(\sigma_t)$, the second effect is the co-jump analogue of $\phi\rho(\sigma_t)\sigma_t\Lambda(\sigma_t)$.

Eqs. (21) through (24) imply that the risk-neutral intensities are expected to be larger than the corresponding intensities under the natural measure, thereby determining a larger number of (idiosyncratic and common) jumps under the change of measure. Finally, Eqs. (21) through (24) also provide the risk-neutral characteristic functions of the sizes of all jumps.

Remark 2 (*The impact of co-jumps on the variance risk premium*). Generalized leverage, as given by diffusive leverage and leverage induced by the co-jumps, contributes to variance risk premia provided $\phi<0$. Hence, the co-jumps lead to an additional source of skewness in stock returns and modify—in equilibrium—the mean of the return distribution by virtue of their corresponding risk compensation (see Remark 1 above) while also inducing a variance risk premium.

Remark 3 (*Co-jumps and nonmonotonic pricing kernels*). Should the stochastic discount factor not depend on variance, and therefore be monotonic in prices, the negative correlation between price and variance co-jumps would not matter to determine return risk premia. It would, however, still matter to determine variance risk premia.

Our modeling assumptions, along with the specification of the pricing kernel in Eq. (17), allow the identification of the risk premia parameters ϕ and ψ once a meaningful restriction on the observed interest rate is imposed. We note that Eq. (18) implies that r is, in general, a function of volatility ($r(\sigma)$). We, therefore, fit Eq. (19) by replacing, on the left-hand side, $r(\sigma)$ with an estimate obtained by a nonparametric regression of the T-bill 3-month rate (divided by 252 to obtain daily units) on daily volatilities, and by plugging in nonparametric estimates of all other functions, as illustrated in Fig. 8. The jump

moments are computed as

$$\mathrm{E}_\sigma[\exp^{\phi c^{JJ}_{r,t}+\psi c^{JJ}_{\sigma,t}}(\exp^{c^{JJ}_{r,t}}-1)]=\exp^{\phi\mu_{JJ,r}(\sigma)+\psi\mu_{JJ,\sigma}(\sigma)+(1/2)\phi^2\sigma^2_{JJ,r}(\sigma)+(1/2)\psi^2\sigma^2_{JJ,\sigma}(\sigma)}\times(\exp^{\mu_{JJ,r}(\sigma)+(1/2)\sigma^2_{JJ,r}(\sigma)(1+2\phi)+\psi\rho_J(\sigma)\sigma_{JJ,r}(\sigma)\sigma_{JJ,\sigma}(\sigma)}-1),$$

and

$$\mathrm{E}_\sigma[\exp^{\phi c^{J}_{r,t}}(\exp^{c^{J}_{r,t}}-1)]=\exp^{\phi\mu_{J,r}(\sigma)+(1/2)\phi^2\sigma^2_{J,r}(\sigma)}(\exp^{\mu_{J,r}(\sigma)+(1/2)\sigma^2_{J,r}(\sigma)(1+2\phi)}-1).$$

The resulting nonlinear system is now overidentified since two parameters (ϕ and ψ) should satisfy Eq. (19) for all volatility levels. Estimation by least-squares on the same volatility levels used for the nonparametric moments yields $\widehat{\phi}=-1.7489$ (with a standard error of 3.0866) and $\widehat{\psi}=0.057143$ (with a standard error of 0.1745). Fig. 11 reports the estimated risk premium, along with the contribution (implied by these parameters) to the overall risk premium of the continuous component ($-\phi\sigma_t^2-\psi\rho(\sigma_t)\sigma_t\Lambda(\sigma_t)$), of the co-jump component, and of the idiosyncratic jump component.

Identifying risk premia over a (relatively) short time period is well-known to be an extremely low signal-to-noise problem. In spite of their large statistical uncertainty, the reported parameter estimates yield an increasing (in variance) return premium and appear economically meaningful. Both observations are evidenced by Fig. 11. We may therefore provide a quantitative assessment of the potential economic impact of co-jumps on risk compensations. Our point estimates imply that the co-jumps explain roughly 25% of the overall return risk premium. Idiosyncratic jumps explain only about 4% of it. The contribution provided by the continuous component amounts to the remaining 71%, 47% being attributable to the variance component $-\phi\sigma_t^2$ and 24% being attributable to the leverage component $-\psi\rho(\sigma_t)\sigma_t\Lambda(\sigma_t)$. In essence, roughly a quarter of the implied return risk premium (given the assumed model specification and our point estimates) can be imputed to contemporaneous price and variance discontinuities. This is an economically sizeable proportion.

We expect the joint use of the return risk premium in Eq. (19) and the variance risk premium in Eq. (28), for which we also offer a closed-form expression under Remark 1, to lead to a reduction in the statistical uncertainty associated with the estimates of ϕ and ψ. The estimation of the risk-neutral variance drift $m^*(\sigma_t)$, however, requires, e.g., option data. Identification of the bivariate system using option data, and a complete exploration of its economic implications, are beyond the scopes of the present paper and are, therefore, left for future work.

8.1. Nonmonotonic pricing kernels

While classical models postulate a pricing kernel that is monotonically decreasing with prices, some empirical work points to lack of monotonicity. In a recent paper, for example, Christoffersen, Heston, and Jacobs (2013) propose a version of Eq. (17) as a way to explain important stylized facts in derivative pricing. They show that $\psi>0$ implies a U-shaped structure in prices. The intuition goes as follows. Variance is high in correspondence with low or high asset prices. When prices are low and variance is high, the discount factor is also high (given $\phi<0$ and $\psi>0$). When prices are high and variance is high, the discount factor is relatively higher than it would be if variance were not in it. In fact, higher prices lower it (through ϕ), whereas higher variance increases it (through ψ). It is the effect of variance at high prices which induces nonmonotonicity. We can now build on this intuition to understand the impact of (1) the presence of co-jumps and (2) the presence of negatively correlated co-jumps. Adding co-jumps gives the model a chance of relatively higher variance at low prices and relatively lower variance at high prices (due to leverage induced by the co-jumps), thereby leading to a rotation of the U-shaped stochastic discount factor. This rotation should be larger, the larger the correlation between the co-jump sizes.

We confirm this intuition using our estimated nonparametric model. To this extent, the price process is simulated using the estimated specification in Section 7 without independent jumps. The parameters are those in Table 2, column 4. The only exception is ρ_J which is set equal to either zero or -1.

We employ the pricing kernel in Eq. (17). Consistent with our previous discussion, the parameter values ϕ and ψ are set equal to their estimated values. For every simulation (indexed by j), the pricing model delivers a value $M_t(p_{j,t},\sigma_{j,t})$ at maturity. In Fig. 12, we display the average value of the pricing kernel for each price level (and, in the inset, the ratios with respect to the case without co-jumps). This average is computed as a nonparametric moment. We report

$$\widehat{M}_t(p)=\frac{\sum_{j=1}^{\#sim}\mathbf{K}\left(\frac{p_{j,t}-p}{\zeta}\right)M_t\left(p_{j,t},\sigma_{j,t}\right)}{\sum_{j=1}^{\#sim}\mathbf{K}\left(\frac{p_{j,t}-p}{\zeta}\right)},$$

where ζ is Silverman's bandwidth and *#sim* denotes the number of simulations.

As expected, the pricing kernel in the absence of co-jumps, shown in the top panel of Fig. 12, is U-shaped. The inclusion of co-jumps rotates it. The extent of this effect depends on the magnitude of the correlation between the co-jump sizes, a higher correlation being associated with a stronger effect. Interestingly, in light of this rotation, the higher the negative correlation between the co-jumps, the less nonmonotonic the pricing kernel. In essence, then, while a higher ψ value leads to stronger nonmonotonicity (as shown below), other aspects of the model equal, accounting for anti-correlated co-jumps will soften nonmonotonicity somewhat.

To further illustrate this point, we also show in the bottom panel of Fig. 12 the pricing kernel corresponding to a higher value of ψ, with ϕ left unchanged. The new value of ψ is set equal to its point estimate plus one standard deviation. Hence, it is compatible with our data given the reported statistical uncertainty. As expected, a larger ψ implies a higher variance risk

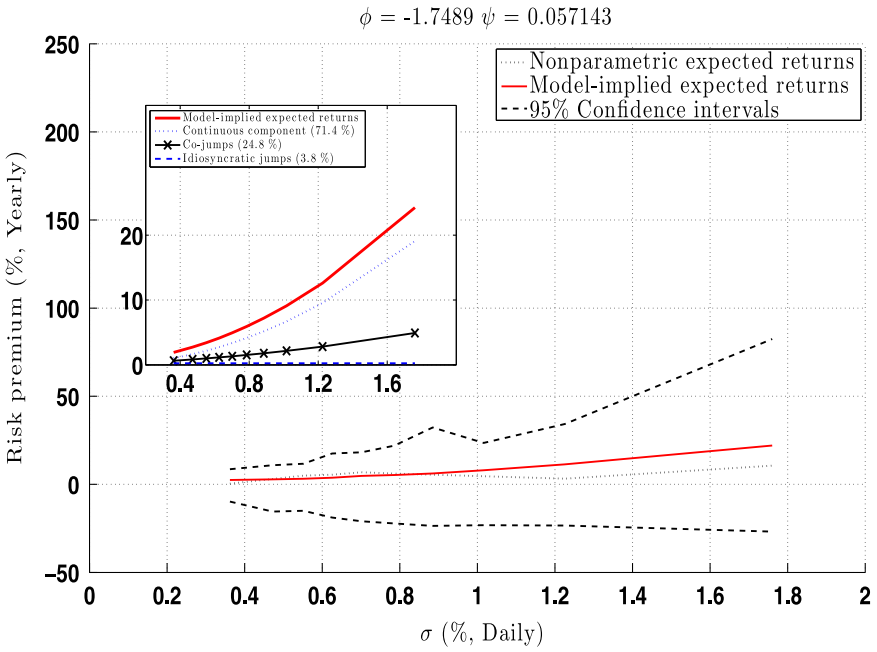

Fig. 11. Return risk premium. We report the return risk premium yielded by the nonparametric estimates (dotted line) along with the model-implied risk premium (solid line). The confidence intervals, displayed as dashed lines, are obtained by simulation. The inset shows the contribution to the overall return risk premium of various components, the dashed line being the contribution of the idiosyncratic jumps, the line with exes being the contribution of the co-jumps, and the dotted line being the contribution of the continuous component. The reported figures in the inset are weighted averages of the contributions, across volatility levels, with weights proportional to the estimated occupation density. We use S&P500 intradaily futures prices from April 21, 1982 to February 5, 2009.

premium and, thus, a more convex pricing kernel. Again, the inclusion of the co-jumps rotates the pricing kernel, with a less nonmonotonic pricing kernel associated with a higher negative correlation between the jump sizes.

This discussion is meant to illustrate the main effects, for specific sensible values of ϕ and ψ, but is necessarily qualitative in nature. A joint quantitative study of return and variance risk premia may lead to an alternative parametrization of the stochastic discount factor for the purpose of effective derivative pricing while, possibly, preserving nonmonotonicity and its empirical appeal.

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 given to parametric specifications in the presence of models with latent variables, such as 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 variance to the identification of the system's parametric structure.

We dispense with this link and identify daily variance, *before* inference on the dynamics begins, by virtue of intradaily price data. In this sense, our filtered variance series is especially useful to select a function $\xi(\cdot)$ and, hence, a variance transformation, which conforms with the assumed jump-diffusion structure and can, therefore, be modeled 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, $f_\lambda(\cdot) = \frac{\lambda - 1}{\lambda}$ 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 variance drift and the volatility of variance function associated with the same transformation.

Consider a specification without jumps. Because the system's shocks are assumed to be Brownian, the (standardized) residuals $\varepsilon_{t,t+\Delta}$ should be locally Gaussian. The most suitable choice of λ is, therefore, the one which guarantees Gaussianity of the residuals. In a model with symmetric Poisson jumps, the residuals should still have zero skewness and only inflated kurtosis. The excess kurtosis should be an increasing function of the number of jumps and the standard deviation of the jump size. To this extent, Fig. 13 reports the skewness and kurtosis of the variance residuals for different values of the parameter λ. A value of λ equal to zero yields zero skewness and the smallest magnitude for kurtosis. Consistent with a model with some

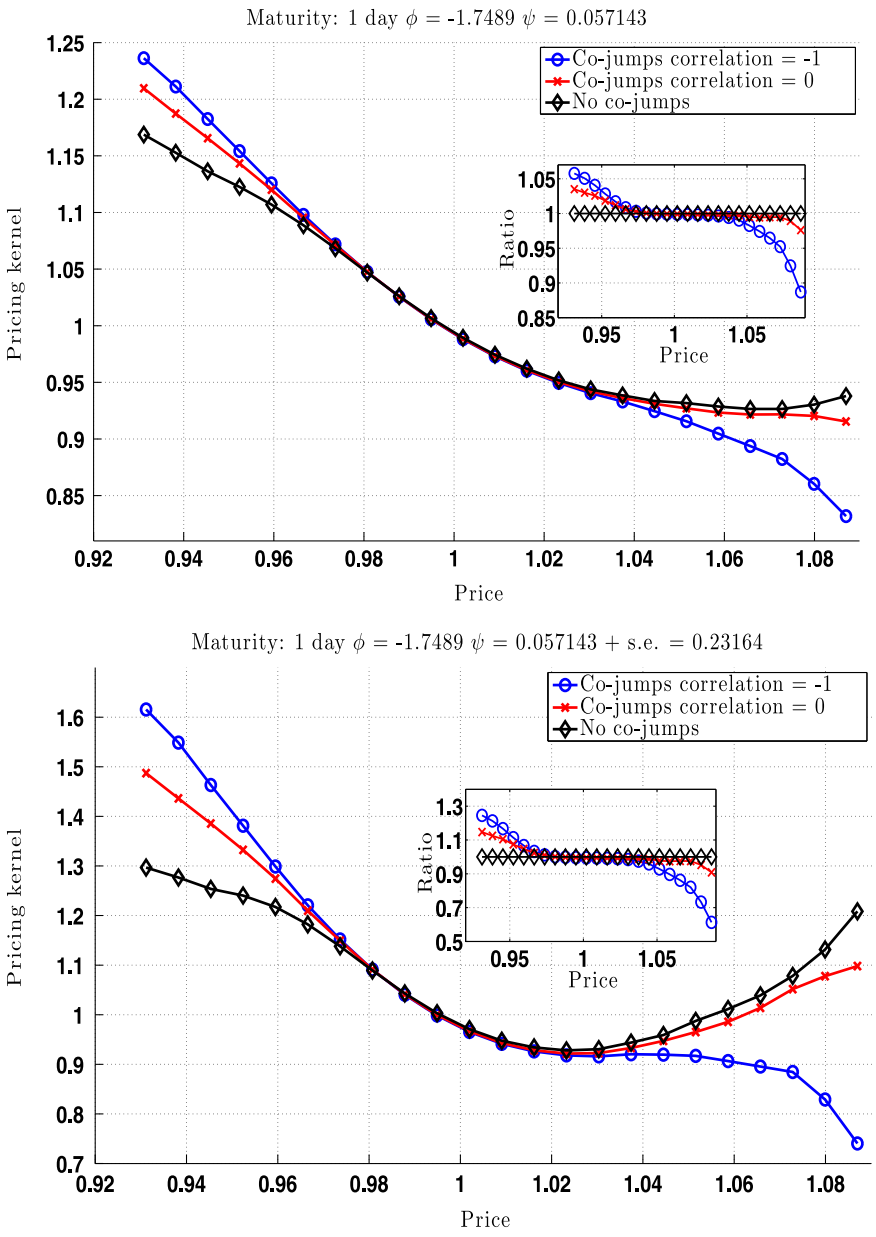

Fig. 12. Pricing kernel. Top panel: We report the pricing kernel based on estimated parameters ϕ and ψ. Bottom panel: We report the pricing kernel based on the estimated parameter ϕ and a value of ψ equal to its estimate plus the estimate's standard error. In the inset, we report pricing kernel ratios with respect to the case without co-jumps. The figures are computed using data simulated based on the estimated parametric model in Table 2. In both figures, and in the insets, the line with circles corresponds to the case with a correlation between co-jumps equal to -1, the line with crosses to the case with a correlation between co-jumps equal to 0, and the line with diamonds to the case without co-jumps. We use S&P500 intradaily futures prices from April 21, 1982 to February 5, 2009 to estimate ϕ and ψ.

discontinuities, this magnitude is slightly larger than in the Gaussian case. Importantly, $\lambda=0$ coincides with the logarithmic transformation which we adopt in this paper.[19] A value equal to one is, instead, associated with an affine variance model. This classical specification does not appear to be supported by our data. The second panel in Fig. 6 shows the filtered logarithmic variances. It is, again, 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 logarithmic variances appear to have a rather

[19] Papers on stochastic volatility estimation in discrete time which employ a logarithmic transformation are Chib, Nardari, and Shephard (2002), Harvey and Shephard (1996), Jacquier, Polson, and Rossi (1994, 2004), and Yu (2005, 2012), among others. For a recent discussion in continuous time, and some related literature, see, e.g., Dobrev and Szerszen (2010).

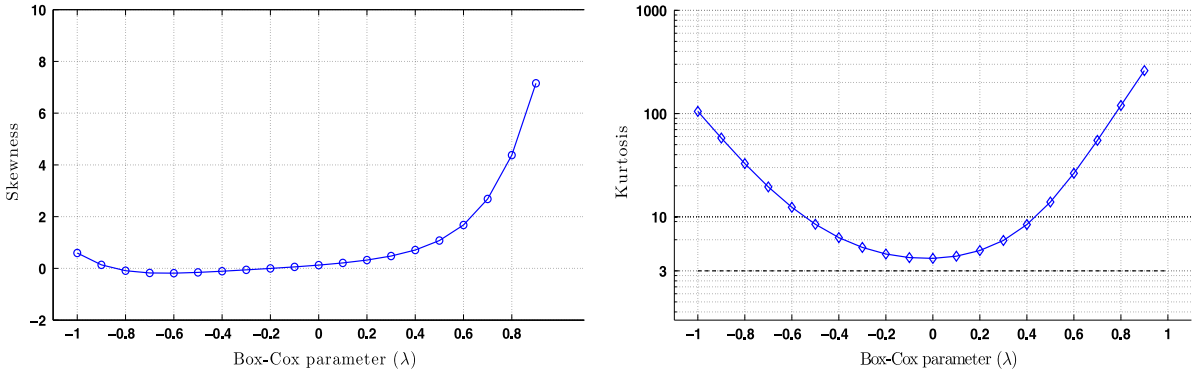

Fig. 13. Box-Cox specification test. Left panel: We report the skewness of the residuals of the estimated nonparametric model as a function of the Box-Cox transformation parameter. Right panel: We report the kurtosis of the residuals of the estimated nonparametric model as a function of the Box-Cox transformation parameter. We use S&P500 intradaily futures prices from April 21, 1982 to February 5, 2009.

constant variance $\Lambda^2(\cdot)$ in our sample. Similarly, the variance jumps seem to have a small standard deviation $\sigma_{J,\sigma}$. Both observations are consistent with the nonparametric and parametric analysis in Sections 6 and 7.

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. We then compare, in Fig. 14, the estimated parametric functions implied by the model to the estimated empirical moments. We find that the model-implied estimates are close to the "free" estimates, thereby offering visual evidence of satisfactory model specification.

We now provide a formal analysis. The (nonparametric and parametric) estimation of the model in Sections 6 and 7 yields, 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 $\langle \vartheta_{p_1,p_2} \rangle$ denotes the estimated moment of order (p_1, p_2) averaged over the estimated volatility density, tested against $\mathcal{H}_1: \langle \vartheta_{p_1,p_2} \rangle \neq 0$. When p_1 and p_2 are both ≥ 1, and at least one of them is strictly larger than one, theory dictates that these higher-order cross-moments should all be directly proportional to $\lambda_{r,\sigma}$ (see Eq. (10)). Specifically, $\langle \vartheta_{p_1,p_2} \rangle = 0$ if $\lambda_{r,\sigma} = 0$. Conversely, a zero moment ($\langle \vartheta_{p_1,p_2} \rangle = 0$) 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 the 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 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, robust to the likely overrated precision which would characterize asymptotic inference and, therefore, conservative.

10. Final remarks

Negative price changes often occur along with positive spikes in the VIX. Just to mention a couple of events, between August 5, 2011 (a Friday) and August 8, 2011 (a Monday), the VIX rose from 32 to 48. The associated, negative S&P 500 change on Monday, August 8, was about 6%. Similarly, on August 17, 2011, the VIX rose from 31 to 42. The corresponding market price change was about -4.5%. These occurrences, along with an array of other similar events over the years, are often invoked as evidence of a clear association between negative market shocks and sudden, positive shifts in the level of volatility. Because it is an important barometer of market sentiment, however, the VIX is a complicated object capturing changes in "fundamental volatility" as well as "market fear," through risk premia.[20]

Interestingly, at daily frequencies, 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. We show, in this paper, that 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 the granular dynamics of rare events, like volatility jumps and co-jumps.

[20] In the framework we propose in Section 8, variance risk premia reacting to sudden price jumps can be captured by risk preference parameters (ϕ and ψ) changing with the price discontinuities.

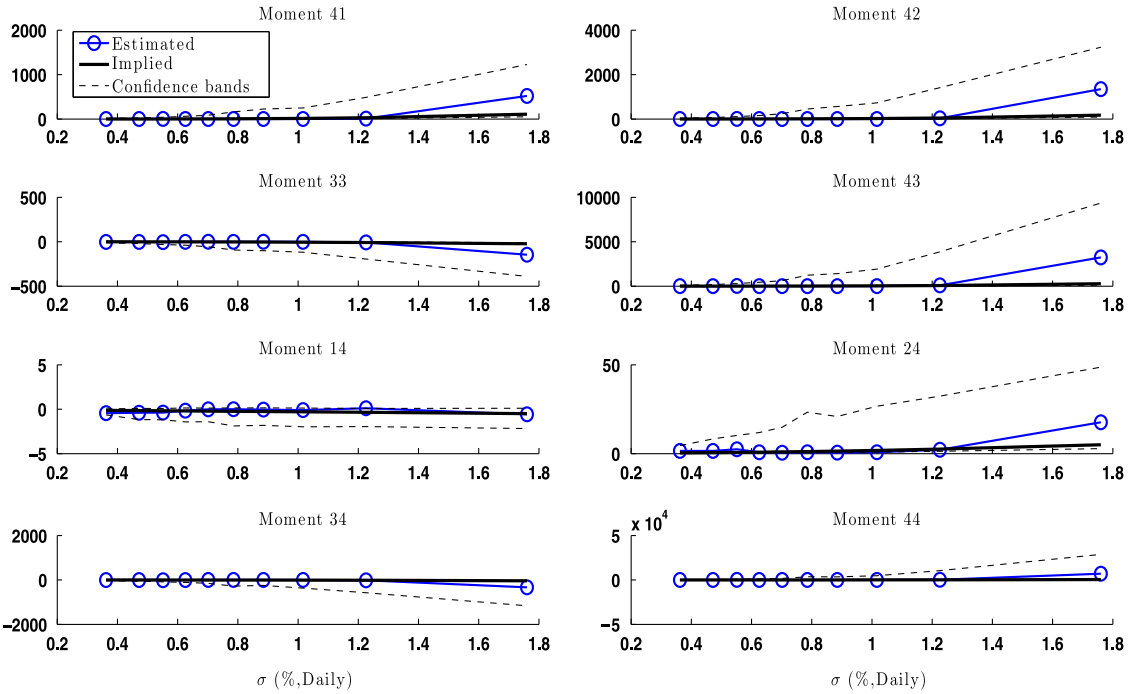

Fig. 14. Tests of overidentifying restrictions. We report graphical representations of tests of overidentifying restrictions based on estimated moments via NIMM. The lines with circles correspond to the estimated moments. The solid lines are the moments implied by the parametric estimates in Table 2. The dashed lines are 95% confidence bands. We use S&P500 intradaily futures prices from April 21, 1982 to February 5, 2009.

Table 3
Co-jump tests. We report moment-based tests for the presence of co-jumps along with corresponding *p*-values. We use S&P500 intradaily futures prices from April 21, 1982 to February 5, 2009.

Test	Value	*P*-value
J-test	239.4	0.20%
$\langle \vartheta_{2,2} \rangle$	0.2862	0.10%
$\langle \vartheta_{1,2} \rangle$	−0.0681	0.00%
$\langle \vartheta_{2,1} \rangle$	0.2669	0.00%
$\langle \vartheta_{1,3} \rangle$	−0.1091	0.00%
$\langle \vartheta_{3,1} \rangle$	−0.9861	0.20%
$\langle \vartheta_{2,3} \rangle$	0.4846	0.10%
$\langle \vartheta_{3,2} \rangle$	−2.3185	0.10%
$\langle \vartheta_{3,3} \rangle$	−6.2580	0.10%

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 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 use of 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 volatility is filtered effectively and a sufficiently rich specification is adopted, without the need for the, arguably economically confounding (due to risk premia), information contained in traded or synthetic options.

Using a possibly nonmonotonic pricing kernel, we derive equilibrium implications of co-jump presence for both return and variance risk premia. We show, again using price data alone, that the impact of co-jumps on the implied (from our adopted specification) return risk premium can be economically sizeable. A more comprehensive empirical investigation of the economic implications of co-jumps on both risk premia requires data on nonlinear derivatives and is a topic left for a future study.

Appendix A. Asymptotic properties

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

$$dX_t = \mu_X(Y_t)\,dt + \sigma_X(Y_t)\{\rho(Y_t)\,dW_t^1 + \sqrt{1-\rho^2(Y_t)}\,dW_t^2\} + c_X\,dJ_t^X + d_X\,dJ_t,$$
$$dY_t = \mu_Y(Y_t)\,dt + \sigma_Y(Y_t)\,dW_t^1 + c_Y\,dJ_t^Y + d_Y\,dJ_t, \tag{A.1}$$

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 unit variance 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 $\{X_{\Delta_{n,T}}, Y_{\Delta_{n,T}}\}, \{X_{2\Delta_{n,T}}, Y_{2\Delta_{n,T}}\}, \ldots, \{X_{n\Delta_{n,T}}, Y_{n\Delta_{n,T}}\}$, 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 $\Delta_{n,T} \to 0$, jointly. We estimate

$$\vartheta_{p_1,p_2}(y) = \lim_{\varepsilon \to 0} \frac{\mathrm{E}[(X_{t+\varepsilon} - X_t)^{p_1}(Y_{t+\varepsilon} - Y_t)^{p_2} | Y_t = y]}{\varepsilon}, \tag{A.2}$$

by using classical Nadaraya-Watson kernel estimates, namely,

$$\widehat{\vartheta}_{p_1,p_2}(y) = \frac{\sum_{i=1}^{n-1} \mathbf{K}\left(\frac{Y_{iT/n} - y}{h_{n,T}}\right)(X_{(i+1)T/n} - X_{iT/n})^{p_1}(Y_{(i+1)T/n} - Y_{iT/n})^{p_2}}{\Delta_{n,T}\sum_{i=1}^{n} \mathbf{K}\left(\frac{Y_{iT/n} - y}{h_{n,T}}\right)}. \tag{A.3}$$

Define

$$\widehat{L}_{n,T}(y) = \frac{\Delta_{n,T}}{h_{n,T}} \sum_{i=1}^{n} \mathbf{K}\left(\frac{Y_{iT/n} - y}{h_{n,T}}\right) = \frac{\Delta_{n,T}}{h_{n,T}} \sum_{i=1}^{n} \mathbf{K}_{i,n,T}(y),$$

i.e., the empirical occupation density of the Y process.

The following proofs provide details which are specific to inference for *infinitesimal cross-moments* depending on uniformly-bounded functions. For more information about the general method of proof in this type of models, including the case of unbounded driving functions, 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 $(X_t, Y_t) = (\log(p_t), \sigma_t^2)$ and provides limiting conditions for a vanishing measurement error in the variance estimates $\widehat{\sigma}_t^2$. Finally, we note that, when the jump sizes depend on the specific level of Y (or σ below), the notation E[.] should be interpreted as E[$\cdot|y$] (or E[$\cdot|\sigma$]). We are concise about this notation here to avoid clutter but are explicit in the main text.

The symbols $\xrightarrow{p}$, $\overset{p}{\sim}$, $\Rightarrow$, $\overset{d}{=}$ will be used to denote "convergence in probability", "of the same probability order as", "convergence in distribution", and "distributional equivalence", respectively.

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

Theorem 1 (*Consistency*). *If $n, T \to \infty$ and $\Delta_{n,T} = T/n \to 0$ so that $h_{n,T}\widehat{L}_{n,T}(y) \xrightarrow{p} \infty$ and $\Delta_{n,T}/h_{n,T}^2 \to 0$, 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 $h_{n,T}\widehat{L}_{n,T}(y) \xrightarrow{p} \infty$ and $\Delta_{n,T}\sqrt{\widehat{L}_{n,T}(y)}/h_{n,T}^{3/2} \xrightarrow{p} 0$. If*

$$h_{n,T}^5 \widehat{L}_{n,T}(y) = \mathrm{O_p}(1),$$

then

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

with

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

where $s(dy) = s(y)\,dy$ *is the invariant measure of the Y process.*

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

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

with $n_1 = 0, 1, \ldots$ *and* $n_2 = 0, 1, \ldots$. *Assume* $n, T \to \infty$ *with* $\Delta_{n,T} \to 0$. *If* $\Delta_{n,T}/h_{n,T}^2 \to 0$, *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. Using results akin to the ratio-limit theorem (Revuz and Yor, 1994), as in Bandi and Renò (2009), and handling discretization the way they do, we have

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

where $\mathbf{K}_s(y) = \mathbf{K}((Y_s - y)/h_{n,T})$ since $\Delta_{n,T}/h_{n,T}^2 \to 0$. An application of Burkholder-Davis-Gundy (BDG) inequality now yields

$$\chi_{n_1,n_2}(y) \overset{p}{\sim} \mathrm{O_p}(\Delta_{n,T}^{1/2}) \chi_{0,0}(y) \xrightarrow{p} 0.$$

where the order term derives from the presence of Poisson jumps.□

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

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

and

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

where $\bar{\nu}(ds, dz)$ *is a compensated Poisson measure, with* $n_1 = 0, 1, \ldots$ *and* $n_2 = 0, 1, \ldots$. *Assume* $n, T \to \infty$ *with* $\Delta_{n,T} \to 0$. *If* $\Delta_{n,T}/h_{n,T}^2 \to 0$ *and* $h_{n,T} \widehat{L}_{n,T}(y) \xrightarrow{p} \infty$, *then*

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

and

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

Proof. 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 $\Delta_{n,T}/h_{n,T}^2 \to 0$ and $h_{n,T} \widehat{L}_{n,T}(y) \xrightarrow{p} \infty$,

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

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

By applying BDG inequality to the conditional variance of $\Psi_{n_1,n_2}(y)$ and $\Xi_{n_1,n_2}(y)$ we obtain

$$\Psi_{n_1,n_2}(y) \overset{p}{\sim} \mathrm{O_p}(\Delta_{n,T}^{1/4}) \Psi_{0,0}(y) \xrightarrow{p} 0,$$

and

$$\Xi_{n_1,n_2}(y) \overset{p}{\sim} \mathrm{O_p}(\Delta_{n,T}^{1/4}) \Xi_{0,0}(y) \xrightarrow{p} 0.\square$$

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

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

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

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

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

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

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

Proof. Without loss of generality, let $n_1\geq n_2\geq 1$ and consider $\Xi_{n_1,n_2}(y)$ first. Write

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

We have $(1/\sqrt{\Delta_{n,T}})\sum_{i=1}^{n-1}\mathrm{E}[u_i|\Im_{iT/n}]=0$. Ito's Lemma now yields

$$\begin{aligned}&(X_{s-}-X_{iT/n})^{n_1}(Y_{s-}-Y_{iT/n})^{n_2}\\&=\int_{iT/n}^{s}n_1(X_{u-}-X_{iT/n})^{n_1-1}(Y_{u-}-Y_{iT/n})^{n_2}\mu_X\,du\\&+\int_{iT/n}^{s}n_2(X_{u-}-X_{iT/n})^{n_1}(Y_{u-}-Y_{iT/n})^{n_2-1}\mu_Y\,du\\&+\int_{iT/n}^{s}n_1(X_{u-}-X_{iT/n})^{n_1-1}(Y_{u-}-Y_{iT/n})^{n_2}\sigma_X\,dW_u^X\\&+\int_{iT/n}^{s}n_2(X_{u-}-X_{iT/n})^{n_1}(Y_{u-}-Y_{iT/n})^{n_2-1}\sigma_Y\,dW_u^Y\\&+\int_{iT/n}^{s}\frac{1}{2}n_1(n_1-1)(X_{u-}-X_{iT/n})^{n_1-2}(Y_{u-}-Y_{iT/n})^{n_2}\sigma_X^2\,du\\&+\int_{iT/n}^{s}\frac{1}{2}n_2(n_2-1)(X_{u-}-X_{iT/n})^{n_1}(Y_{u-}-Y_{iT/n})^{n_2-2}\sigma_Y^2\,du\\&+\int_{iT/n}^{s}n_1n_2(X_{u-}-X_{iT/n})^{n_1-1}(Y_{u-}-Y_{iT/n})^{n_2-1}\rho\sigma_X\sigma_Y\,du\\&+\sum_{\Delta X_u\neq 0\text{ or }\Delta Y_u\neq 0}[(X_{u-}+\Delta X_u-X_{iT/n})^{n_1}(Y_{u-}+\Delta Y_u-Y_{iT/n})^{n_2}\\&\quad-(X_{u-}-X_{iT/n})^{n_1}(Y_{u-}-Y_{iT/n})^{n_2}].\end{aligned}\tag{A.4}$$

Thus, the higher-order term in the expression above is given by $\sum(\Delta X_u)^{n_1}(\Delta Y_u)^{n_2}$. Using arguments in Bandi and Renò (2009), we obtain

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

and, for all $\epsilon>0$,

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

as $\Delta_{n,T}\to 0$, uniformly in T and $h_{n,T}$ in $\mathcal{H}(\varepsilon)=\{\Delta_{n,T}^{1/2}/\varepsilon<h_{n,T}<\varepsilon\}$ for an arbitrarily small $\varepsilon>0$. Theorem VIII 3.33 in Jacod and

Shiryaev (2003) allows us to conclude that $\Xi^{num}_{n_1,n_2}(y) \underset{\Delta_{n,T}\to 0}{\Rightarrow} W_{\mathbf{U}^2}$, where W is a Brownian motion independent of $\mathbf{U}^2$, and

$$\sqrt{\frac{h_{n,T}\widehat{L}_{n,T}(y)}{\Delta_{n,T}}}\Xi_{n_1,n_2}(y)=\frac{\Xi^{num}_{n_1,n_2}(y)}{\sqrt{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}_{i,n,T}(y)}}\overset{d}{=}W_{\mathbf{U}^2/\int_0^T(1/h_{n,T})\mathbf{K}_s(y)\,ds},$$

uniformly in T and $h_{n,T}$ in $\mathcal{H}(\varepsilon)$. Analogous arguments to the ratio-limit theorem (Bandi and Renò, 2009) now give

$$\frac{\mathbf{U}^2}{\int_0^T\frac{1}{h_{n,T}}\mathbf{K}_s(y)\,ds}\overset{T\to\infty,h_{n,T}\to 0}{\overset{p}{\to}}\frac{1}{2}\mathbf{K}_2\mathrm{E}\left[g_2^2(y,z)\right]\lambda_{XY}(y)\mathrm{E}\left[d_X^{2n_1}d_Y^{2n_2}\right].$$

Theorem 4.1 in van Zanten (2000) implies the desired result. The case $n_1\geq 0$ with $n_2=0$ can be treated in a similar way.

Next, consider $\Psi_{n_1,n_2}(y)$ and let

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

Compute its quadratic variation, i.e.,

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

We have

$$\left[\Psi^{num}_{n_1,n_2}(y)\right]\overset{p}{\to}\frac{1}{2}\frac{1}{h_{n,T}}\int_0^T\mathbf{K}_s^2(y)g_1^2(Y_s)\lambda_{XY}(Y_s)\mathrm{E}\left[d_X^{2n_1}d_Y^{2n_2}\right]ds$$

uniformly in T and $h_{n,T}$ in $\mathcal{H}(\varepsilon)$. Reasoning in the same way as for $\Xi_{n_1,n_2}(y)$, we obtain

$$\sqrt{\frac{h_{n,T}\widehat{L}_{n,T}(y)}{\Delta_{n,T}}}\Psi_{n_1,n_2}(y)\Rightarrow\mathcal{N}\left(0,\frac{1}{2}\mathbf{K}_2g_1^2(y)\vartheta_{2n_1,2n_2}(y)\right),$$

if $h_{n,T}\widehat{L}_{n,T}(y)\overset{p}{\to}\infty$ and $\Delta_{n,T}/h_{n,T}^2\to 0$. Clearly, for $n_1\geq 1$ and $n_2=0$,

$$\sqrt{\frac{h_{n,T}\widehat{L}_{n,T}(y)}{\Delta_{n,T}}}\Psi_{n_1,0}(y)\Rightarrow\mathcal{N}\left(0,\frac{1}{2}\mathbf{K}_2g_1^2(y)\vartheta_{2n_1,0}(y)\right),$$

if $h_{n,T}\widehat{L}_{n,T}(y)\overset{p}{\to}\infty$ and $\Delta_{n,T}/h_{n,T}^2\to 0$. The case $n_1=n_2=0$ is now obvious.□

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

$$\begin{aligned}(X_{(i+1)T/n}-X_{iT/n})^{p_1}&=p_1\int_{iT/n}^{(i+1)T/n}(X_{s-}-X_{iT/n})^{p_1-1}\mu_X\,ds+p_1\int_{iT/n}^{(i+1)T/n}(X_{s-}-X_{iT/n})^{p_1-1}\sigma_X\,dW_s^X\\&+\frac{1}{2}p_1(p_1-1)\int_{iT/n}^{(i+1)T/n}(X_{s-}-X_{iT/n})^{p_1-2}\sigma_X^2\,ds+\sum_{\Delta X_s\neq 0}\left[(X_{s-}+\Delta X_s-X_{iT/n})^{p_1}-(X_{s-}-X_{iT/n})^{p_1}\right].\end{aligned}\tag{A.5}$$

Now, write

$$\sum_{\Delta X_s\neq 0}\left[(X_{s-}+\Delta X_s-X_{iT/n})^{p_1}-(X_{s-}-X_{iT/n})^{p_1}\right]=\sum_{k=0}^{p_1-1}\binom{p_1}{k}\int_{iT/n}^{(i+1)T/n}\int_Z(X_{s-}-X_{iT/n})^k(c_X^{p_1-k}\nu_X(ds,dz)+d_X^{p_1-k}\nu_{XY}(ds,dz)).$$

De-meaning the random jump measures and denoting, as earlier, by $\overline{\nu}_X$ and $\overline{\nu}_{XY}$ the corresponding compensated measures, we write $\overline{\nu}_X(ds,dz)=\nu_X(ds,dz)-\lambda_X\mathrm{E}[c_X^{p_1-k}]\,ds$ and $\overline{\nu}_{XY}(ds,dz)=\nu_{XY}(ds,dz)-\lambda_{XY}\mathrm{E}[d_X^{p_1-k}]\,ds$. Thus,

$$\begin{aligned}\widehat{\vartheta}_{p_1,0}(y)=&\frac{\sum_{i=1}^{n-1}p_1\mathbf{K}_{i,n,T}(y)\int_{iT/n}^{(i+1)T/n}(X_{s-}-X_{iT/n})^{p_1-1}\mu_X\,ds}{\Delta_{n,T}\sum_{i=1}^{n}\mathbf{K}_{i,n,T}(y)}\\&+\frac{\sum_{i=1}^{n-1}p_1\mathbf{K}_{i,n,T}(y)\int_{iT/n}^{(i+1)T/n}(X_{s-}-X_{iT/n})^{p_1-1}\sigma_X\,dW_s^X}{\Delta_{n,T}\sum_{i=1}^{n}\mathbf{K}_{i,n,T}(y)}\\&+\frac{\sum_{i=1}^{n-1}\frac{1}{2}p_1(p_1-1)\mathbf{K}_{i,n,T}(y)\int_{iT/n}^{(i+1)T/n}(X_{s-}-X_{iT/n})^{p_1-2}\sigma_X^2\,ds}{\Delta_{n,T}\sum_{i=1}^{n}\mathbf{K}_{i,n,T}(y)}\\&+\frac{\sum_{i=1}^{n-1}\mathbf{K}_{i,n,T}(y)\sum_{k=0}^{p_1-1}\binom{p_1}{k}\int_{iT/n}^{(i+1)T/n}(X_{s-}-X_{iT/n})^k\int_Z(c_X^{p_1-k}\overline{\nu}_X(ds,dz)+d_X^{p_1-k}\overline{\nu}_{XY}(ds,dz))}{\Delta_{n,T}\sum_{i=1}^{n}\mathbf{K}_{i,n,T}(y)}\end{aligned}$$

$$+\frac{\sum_{i=1}^{n-1}\mathbf{K}_{i,n,T}(y)\sum_{k=0}^{p_1-1}\binom{p_1}{k}\int_{iT/n}^{(i+1)T/n}(X_{s-}-X_{iT/n})^k(\lambda_X(Y_s)\mathrm{E}[c_X^{p_1-k}]+\lambda_{XY}(Y_s)\mathrm{E}[d_X^{p_1-k}])\,ds}{\Delta_{n,T}\sum_{i=1}^{n}\mathbf{K}_{i,n,T}(y)}$$
$$=\mathbf{R}_1+\mathbf{R}_2+\mathbf{R}_3+\mathbf{R}_4+\mathbf{A}_5.$$

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

$$\mathbf{A}_5=\frac{\sum_{i=1}^{n-1}\mathbf{K}_{i,n,T}(y)\int_{iT/n}^{(i+1)T/n}(\lambda_X(Y_s)\mathrm{E}[c_X^{p_1}]+\lambda_{XY}(Y_s)\mathrm{E}[d_X^{p_1}])\,ds}{\Delta_{n,T}\sum_{i=1}^{n}\mathbf{K}_{i,n,T}(y)}$$
$$+\frac{\sum_{i=1}^{n-1}\mathbf{K}_{i,n,T}(y)\sum_{k=1}^{p_1-1}\binom{p_1}{k}\int_{iT/n}^{(i+1)T/n}(X_{s-}-X_{iT/n})^k(\lambda_X(Y_s)\mathrm{E}[c_X^{p_1-k}]+\lambda_{XY}(Y_s)\mathrm{E}[d_X^{p_1-k}])\,ds}{\Delta_{n,T}\sum_{i=1}^{n}\mathbf{K}_{i,n,T}(y)}$$
$$=\mathbf{A}_5^1+\mathbf{A}_5^2.$$

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. The case $p_1=1$, $p_2=0$ and, given Eq. (A.5), the case $p_1=2$, $p_2=0$ are obtained accordingly. In the latter situation, in fact, $\mathbf{R}_3\xrightarrow{p}\sigma_X^2(y)$. Consider, now, the case $p_1>p_2\geq 1$. Use Ito's Lemma as in Eq. (A.4) and rewrite the last term in that expression as

$$\sum_{\Delta X_s\neq 0 \text{ or } \Delta Y_s\neq 0}[(X_{s-}+\Delta X_s-X_{iT/n})^{p_1}(Y_{s-}+\Delta Y_s-Y_{iT/n})^{p_2}-(X_{s-}-X_{iT/n})^{p_1}(Y_{s-}-Y_{iT/n})^{p_2}]$$
$$=\sum_{\Delta X_s\neq 0 \text{ or } \Delta Y_s\neq 0}\left(\sum_{k=0}^{p_1-1}\binom{p_1}{k}(X_{s-}-X_{iT/n})^k(\Delta X_s)^{p_1-k}\right)\left(\sum_{k=0}^{p_2-1}\binom{p_2}{k}(Y_{s-}-Y_{iT/n})^k(\Delta Y_s)^{p_2-k}\right)-\sum_{\Delta X_s\neq 0 \text{or} \Delta Y_s\neq 0}(X_{s-}-X_{iT/n})^{p_1}(Y_{s-}-Y_{iT/n})^{p_2}$$
$$=\sum_{\Delta X_s\Delta Y_s\neq 0}(\Delta X_s)^{p_1}(\Delta Y_s)^{p_2}+\mathbf{B}_{i,n,T}$$
$$=\int_{iT/n}^{(i+1)T/n}\int_Z d_X^{p_1}d_Y^{p_2}\nu_{XY}(ds,dz)+\mathbf{B}_{i,n,T}$$

Now, we have

$$\widehat{\vartheta}_{p_1,p_2}(y)=\frac{\sum_{i=1}^{n-1}\mathbf{K}_{i,n,T}(y)\int_{iT/n}^{(i+1)T/n}\int_Z d_X^{p_1}d_Y^{p_2}\nu_{XY}(ds,dz)}{\Delta_{n,T}\sum_{i=1}^{n}\mathbf{K}_{i,n,T}(y)}$$
$$+\frac{\sum_{i=1}^{n-1}\mathbf{K}_{i,n,T}(y)\int_{iT/n}^{(i+1)T/n}p_1(X_{s-}-X_{iT/n})^{p_1-1}(Y_{s-}-Y_{iT/n})^{p_2}\mu_X\,ds}{\Delta_{n,T}\sum_{i=1}^{n}\mathbf{K}_{i,n,T}(y)}$$
$$+\frac{\sum_{i=1}^{n-1}\mathbf{K}_{i,n,T}(y)\int_{iT/n}^{(i+1)T/n}p_2(X_{s-}-X_{iT/n})^{p_1}(Y_{s-}-Y_{iT/n})^{p_2-1}\mu_Y\,ds}{\Delta_{n,T}\sum_{i=1}^{n}\mathbf{K}_{i,n,T}(y)}$$
$$+\frac{\sum_{i=1}^{n-1}\mathbf{K}_{i,n,T}(y)\int_{iT/n}^{(i+1)T/n}p_1(X_{s-}-X_{iT/n})^{p_1-1}(Y_{s-}-Y_{iT/n})^{p_2}\sigma_X\,dW_s^X}{\Delta_{n,T}\sum_{i=1}^{n}\mathbf{K}_{i,n,T}(y)}$$
$$+\frac{\sum_{i=1}^{n-1}\mathbf{K}_{i,n,T}(y)\int_{iT/n}^{(i+1)T/n}p_2(X_{s-}-X_{iT/n})^{p_1}(Y_{s-}-Y_{iT/n})^{p_2-1}\sigma_Y\,dW_s^Y}{\Delta_{n,T}\sum_{i=1}^{n-1}\mathbf{K}_{i,n,T}(y)}$$
$$+\frac{\sum_{i=1}^{n-1}\mathbf{K}_{i,n,T}(y)\int_{iT/n}^{(i+1)T/n}\frac{1}{2}p_1(p_1-1)(X_{s-}-X_{iT/n})^{p_1-2}(Y_{s-}-Y_{iT/n})^{p_2}\sigma_X^2\,ds}{\Delta_{n,T}\sum_{i=1}^{n}\mathbf{K}_{i,n,T}(y)}$$
$$+\frac{\sum_{i=1}^{n-1}\mathbf{K}_{i,n,T}(y)\int_{iT/n}^{(i+1)T/n}\frac{1}{2}p_2(p_2-1)(X_{s-}-X_{iT/n})^{p_1}(Y_{s-}-Y_{iT/n})^{p_2-2}\sigma_Y^2\,ds}{\Delta_{n,T}\sum_{i=1}^{n}\mathbf{K}_{i,n,T}(y)}$$
$$+\frac{\sum_{i=1}^{n-1}\mathbf{K}_{i,n,T}(y)\int_{iT/n}^{(i+1)T/n}p_1p_2(X_{s-}-X_{iT/n})^{p_1-1}(Y_{s-}-Y_{iT/n})^{p_2-1}\rho\sigma_X\sigma_Y\,ds}{\Delta_{n,T}\sum_{i=1}^{n}\mathbf{K}_{i,n,T}(y)}$$
$$+\frac{\sum_{i=1}^{n-1}\mathbf{K}_{i,n,T}(y)\mathbf{B}_{i,n,T}}{\Delta_{n,T}\sum_{i=1}^{n}\mathbf{K}_{i,n,T}(y)}$$
$$=\frac{\sum_{i=1}^{n-1}\mathbf{K}_{i,n,T}(y)\int_{iT/n}^{(i+1)T/n}\int_Z d_X^{p_1}(z)\,d_Y^{p_2}(z)\nu_{XY}(ds,dz)}{\Delta_{n,T}\sum_{i=1}^{n}\mathbf{K}_{i,n,T}(y)}$$
$$+\mathbf{R}_1+\mathbf{R}_2+\mathbf{R}_3+\mathbf{R}_4+\mathbf{R}_5+\mathbf{R}_6+\mathbf{R}_7+\mathbf{R}_8. \tag{A.6}$$

Finally, we 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

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

Given Eq. (A.6), the case $p_2 = p_1 = 1$ is, again, obvious since

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

Proof of Theorem 2. Consider, without loss of generality, the case $p_1 > p_2 \geq 1$. All other cases follow analogously. Use Ito's Lemma as in Eq. (A.4) and, again, rewrite the jump term as

$$\begin{aligned} &\sum_{\Delta X_s \neq 0 \text{ or } \Delta Y_s \neq 0} [(X_{s-} + \Delta X_s - X_{iT/n})^{p_1} (Y_{s-} + \Delta Y_s - Y_{iT/n})^{p_2} - (X_{s-} - X_{iT/n})^{p_1} (Y_{s-} - Y_{iT/n})^{p_2}] \\ &= \int_{iT/n}^{(i+1)T/n} \int_Z d_X^{p_1} d_Y^{p_2} \nu_{XY}(ds, dz) + \mathbf{B}_{i,n,T}, \end{aligned}$$

where $\mathbf{B}_{i,n,T} = o_p(\int_{iT/n}^{(i+1)T/n} \int_Z d_X^{p_1} d_Y^{p_2} \nu_{XY}(ds, dz))$. Then, as in the proof of Theorem 1, after compensating the random measure ν_{XY}, let

$$\begin{aligned} &\widehat{\vartheta}_{p_1,p_2}(y) - \frac{\sum_{i=1}^{n-1} \mathbf{K}_{i,n,T}(y) \int_{iT/n}^{(i+1)T/n} \lambda_{XY}(Y_s) \mathrm{E}[d_X^{p_1} d_Y^{p_2}]\, ds}{\Delta_{n,T} \sum_{i=1}^{n} \mathbf{K}_{i,n,T}(y)} \\ &= \mathbf{A} + \mathbf{R}_1 + \mathbf{R}_2 + \mathbf{R}_3 + \mathbf{R}_4 + \mathbf{R}_5 + \mathbf{R}_6 + \mathbf{R}_7 + \mathbf{R}_8, \end{aligned} \tag{A.7}$$

where

$$\mathbf{A} = \frac{\sum_{i=1}^{n-1} \mathbf{K}_{i,n,T}(y) \int_{iT/n}^{(i+1)T/n} \int_Z d_X^{p_1} d_Y^{p_2} \overline{\nu}_{XY}(ds, dz)}{\Delta_{n,T} \sum_{i=1}^{n} \mathbf{K}_{i,n,T}(y)}.$$

Now, using Lemma A.3, we have that

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

Naturally, **A** is the dominating term in the sum. Finally, we note that the left-hand side of Eq. (A.7) can be expressed as

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

with

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

This completes the proof. □

A.1. Replacing spot variance with its estimated value

Rewrite the estimator in Eq. (13) as $\widehat{\sigma}_{t,i}^2 = TBPV_{t,i}/\phi$ with

$$TBPV_{t,i} = \frac{k}{(k-1) - n_j} \varsigma_1^{-2} \sum_{j=2}^{k} \left| r_{t,i,j} \right| \left| r_{t,i,j-1} \right| \mathbf{1}_{\{|r_{t,i,j}| \leq \theta_{t,i,j}\}} \mathbf{1}_{\{|r_{t,i,k-1}| \leq \theta_{t,i,j-1}\}},$$

where *TBPV* stands for "Threshold Bipower Variation" and all symbols were defined in the main text. In our case, ϕ is equal to one hour and five minutes and $k = 65$. 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. Let $\phi \to 0$ with $k \to \infty$. Assume $\theta_{t,i} = \xi_{t,i} \Theta(\phi/k)$, where $\Theta(\phi/k)$ is a real function satisfying

$$\Theta\left(\frac{\phi}{k}\right)_{\phi \to 0, k \to \infty} \to 0 \quad \text{and} \quad \frac{1}{\Theta\left(\frac{\phi}{k}\right)} \left(\frac{\phi}{k} \log\left(\frac{\phi}{k}\right) \right)_{\phi \to 0, k \to \infty} \to 0$$

and $\xi_{t,i}$ is an almost-surely bounded process with a strictly positive lower bound. Write $\omega_{n,k,\phi} = \sqrt{\frac{\log(n)\, \log(nk)}{k}} + \phi n + \sqrt{\phi \log(n)}$.

Consider $\widehat{\vartheta}_{p_1,p_2}(.)$ *with* $(p_1,p_2)=(1,0)$ *or* $(0,1)$. *If*

$$\frac{\omega_{n,k,\phi}}{\Delta_{n,T}}\to 0,$$

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

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

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

$$\sqrt{h_{n,T}\widehat{L}_{n,T}(\sigma^2)}\frac{\omega_{n,k,\phi}}{\Delta_{n,T}}\xrightarrow{p}0,$$

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

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

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

Proof of Theorem 3. We work with $X_t=\log(p_t)$ and $Y_t=\sigma_t^2$. The situation $Y_t=\xi(\sigma_t^2)$ can be treated similarly. In light of the results in Bandi and Renò (2009, 2012) we only have to consider the case $p_1\geq 1$, $p_2\geq 1$, with at least one of them being strictly larger than one. Write

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

where $\mathbf{K}_i=\mathbf{K}((\sigma^2_{iT/n}-\sigma^2)/h_{n,T})$ and $\widehat{\mathbf{K}}_i=\mathbf{K}((\widehat{\sigma}^2_{iT/n}-\sigma^2)/h_{n,T})$ (for conciseness, we drop the subscripts n and T used earlier for the kernels). In the presence of jumps, we have

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

and

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

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

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

Write $\mathbf{M}_{n,T,k,\phi}=\max_{1\leq i\leq n}|\widehat{\sigma}^2_{iT/n}-\sigma^2_{iT/n}|=\mathrm{O_p}(\omega_{n,k,\phi})$, as shown in Bandi and Renò (2009, 2012). Define $\gamma_{n,T,k,\phi}=\omega_{n,k,\phi}/h_{n,T}$.

We have

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

Now, for suitable constants $c_{j_1j_2j_3}$,

$$(\log(p_{(i+1)T/n})-\log(p_{iT/n}))^{p_1}(\widehat{\sigma}^2_{(i+1)T/n}-\sigma^2_{(i+1)T/n}+\sigma^2_{(i+1)T/n}-\sigma^2_{iT/n}+\sigma^2_{iT/n}-\widehat{\sigma}^2_{iT/n})^{p_2}$$

$$= \sum_{j_1+j_2+j_3=p_2} c_{j_1 j_2 j_3} (\log(p_{(i+1)T/n}) - \log(p_{iT/n}))^{p_1} (\sigma^2_{iT/n} - \sigma^2_{iT/n})^{j_3} (\widehat{\sigma}^2_{(i+1)T/n} - \sigma^2_{(i+1)T/n})^{j_1} (\widehat{\sigma}^2_{iT/n} - \sigma^2_{iT/n})^{j_2}$$
$$= \mathrm{O_p}(\Delta_{n,T}^{1/2}).$$

Thus, using the mean-value theorem,

$$\mathbf{a} = \mathrm{O_p}\left(\Delta_{n,T}^{1/2}\right) \frac{1 \le i \le n \max \left| \frac{\widehat{\sigma}^2_{i\Delta_{n,T}} - \sigma^2_{i\Delta_{n,T}}}{h_{n,T}} \right| \sum_{i=1}^{n-1} \left| \mathbf{K}'\left(\frac{\sigma^2_{iT/n} + \mathrm{O_p}(\mathbf{M}_{n,T,k,\phi}) - x}{h_{n,T}} \right) \right|}{\Delta_{n,T} \sum_{i=1}^{n} \widehat{\mathbf{K}}_i}$$
$$= \mathrm{O_p}\left(\frac{\mathbf{M}_{n,T,k,\phi} \Delta_{n,T}^{1/2}}{\Delta_{n,T} h_{n,T}} \right) \frac{\frac{1}{h_{n,T}} \int_0^T \left| \mathbf{K}'\left(\frac{\sigma_s^2 - x}{h_{n,T}} \right) \right| ds}{\frac{1}{h_{n,T}} \int_0^T \mathbf{K}\left(\frac{\sigma_s^2 - x}{h_{n,T}} \right) ds} + \mathrm{O_p}\left(\frac{\Delta_{n,T}}{h_{n,T}^2} \right)}{1 + \mathrm{O_p}\left(\frac{\Delta_{n,T}}{h_{n,T}^2} \right) + \mathrm{O_p}\left(\gamma_{n,T,k,\phi} \right)}$$
$$= \mathrm{O_p}\left(\frac{\mathbf{M}_{n,T,k,\phi}}{\Delta_{n,T}^{1/2} h_{n,T}} \right),$$

if $\Delta_{n,T}/h^2_{n,T} \to 0$ and $\gamma_{n,T,k,\phi} \to 0$. We now turn to $\mathbf{b}$:

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

We can prove that $\xi_{iT/n} = (\log(p_{(i+1)T/n}) - \log(p_{iT/n}))^{p_1} \ [(\widehat{\sigma}^2_{(i+1)T/n} - \widehat{\sigma}^2_{iT/n})^{p_2} - (\sigma^2_{(i+1)T/n} - \sigma^2_{iT/n})^{p_2}]$ is $\mathrm{O_p}(\mathbf{M}_{n,T,k,\phi} \Delta_{n,T}^{1/2})$. 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

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

We can now decompose $\xi_{iT/n}$ into its terms, the first corresponding to $k=1$:

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

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

$$(\widehat{\sigma}^2_{(i+1)T/n} - \widehat{\sigma}^2_{iT/n} - \sigma^2_{(i+1)T/n} + \sigma^2_{iT/n}) (\widehat{\sigma}^2_{(i+1)T/n} - \widehat{\sigma}^2_{iT/n})^{p_2 - k} \times \underbrace{(\log(p_{(i+1)T/n}) - \log(p_{iT/n}))^{p_1} (\sigma^2_{(i+1)T/n} - \sigma^2_{iT/n})^{k-1}}_{\mathrm{O_p}(\Delta_{n,T}^{1/2})}$$
$$= \mathrm{O_p}(\mathbf{M}_{n,T,k,\phi} \Delta_{n,T}^{1/2}) (\widehat{\sigma}^2_{(i+1)T/n} - \sigma^2_{(i+1)T/n} - \widehat{\sigma}^2_{iT/n} + \sigma^2_{iT/n} + \sigma^2_{(i+1)T/n} - \sigma^2_{iT/n})^{p_2 - k}$$
$$= o_p(\mathbf{M}_{n,T,k,\phi} \Delta_{n,T}^{1/2}).$$

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

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

Thus,

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

if $\Delta_{n,T}/h^2_{n,T} \to 0$ and $\gamma_{n,T,k,\phi} \to 0$. Finally,

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

$$= \mathrm{O_p}\left(\frac{\mathbf{M}_{n,T,k,\phi}}{h_{n,T}}\right),$$

if $\Delta_{n,T}/h_{n,T}^2 \to 0$ and $\gamma_{n,T,k,\phi} \to 0$. Now, notice that term **c** is of higher order than term **b** since

$$\frac{\mathbf{M}_{n,T,k,\phi}}{h_{n,T}} = \frac{\mathbf{M}_{n,T,k,\phi}}{\Delta_{n,T}^{1/2}}\frac{\Delta_{n,T}^{1/2}}{h_{n,T}} = \frac{\mathbf{M}_{n,T,k,\phi}}{\Delta_{n,T}^{1/2}}\mathrm{o}(1)$$

and, of course, both **c** and **b** are of higher order than **a**. Therefore,

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

which completes the proof.□

Appendix B. Economic implications

Proof of Proposition 8.1. All results, and their implications, could be proved by using the properties of the change of measure, given by the assumed stochastic discount factor, as in the main text (see, e.g., Eq. (29)). Here we follow a different, but equivalent, strategy. As in Christoffersen, Heston, and Jacobs (2013), we show the results by using pricing restrictions.

From Eq. (17), we have

$$\log M_t - \log M_0 = \phi(\log(p_t) - \log(p_0)) + \int_0^t \delta(\sigma_s)\, ds + \psi(\xi(\sigma_t^2) - \xi(\sigma_0^2))$$

and

$$\begin{aligned} d\log M_t &= \phi d\log(p_t) + \delta(\sigma_t)\, dt + \psi\, d\xi(\sigma_t^2) \\ &= [\phi\mu(\sigma_t) + \delta(\sigma_t) + \psi m(\sigma_t)]\, dt + [\phi\sigma_t\rho(\sigma_t) + \psi\Lambda(\sigma_t)]\, dW_t^1 \\ &\quad + [\phi\sigma_t\sqrt{(1-\rho^2(\sigma_t))}]\, dW_t^2 + (\phi c_{r,t}^{JJ} + \psi c_{\sigma,t}^{JJ})\, dJ_{r,\sigma} + \phi c_{r,t}^{J}\, dJ_r + \psi c_{\sigma,t}^{J}\, dJ_\sigma. \end{aligned}$$

Using Ito's Lemma and denoting by dM_t^c the continuous part of dM_t, we obtain

$$\begin{aligned} dM_t &= M_t d\log(M_t^c) + \tfrac{1}{2} M_t\left[\phi^2\sigma_t^2 + \psi^2\Lambda^2(\sigma_t) + 2\phi\psi\rho(\sigma_t)\sigma_t\Lambda(\sigma_t)\right] dt \\ &\quad + M_t(\exp^{\phi c_{r,t}^{JJ} + \psi c_{\sigma,t}^{JJ}} - 1) dJ_{r,\sigma} + M_t(\exp^{\phi c_{r,t}^{J}} - 1) dJ_r + M_t(\exp^{\psi c_{\sigma,t}^{J}} - 1)\, dJ_\sigma. \end{aligned}$$

Now, define $B_t = B_0 \exp^{rt}$ and $Y_{1,t} = B_t M_t$, so that

$$dY_{1,t} = Y_{1,t}\left(\frac{dM_t}{M_t} + r\, dt\right).$$

To guarantee absence of arbitrage, $Y_{1,t}$ has to be a martingale. To this extent, setting $(1/dt)\mathrm{E}_t[dY_{1,t}] = 0$, we obtain Eq. (18).

Similarly, consider $Y_{2,t} = p_t M_t$. By Ito's Lemma, we get

$$\begin{aligned} \frac{dY_{2,t}}{Y_{2,t}} &= \frac{dM_t^c}{M_t} + \frac{dp_t^c}{p_t} + \left(\phi\sigma_t^2 + \psi\rho(\sigma_t)\Lambda(\sigma_t)\sigma_t\right) dt \\ &\quad + (\exp^{(\phi+1)c_{r,t}^{JJ} + \psi c_{\sigma,t}^{JJ}} - 1) dJ_{r,\sigma} + (\exp^{(\phi+1)c_{r,t}^{J}} - 1)\, dJ_r + (\exp^{\psi c_{\sigma,t}^{J}} - 1)\, dJ_\sigma. \end{aligned}$$

Again, $Y_{2,t}$ should be a martingale. Hence, using $\mathrm{E}_t[dM_t/M_t] = -r\, dt$, we obtain Eq. (19).

Now, consider $Y_{3,t} = M_t\mathcal{D}(t, p_t, \xi(\sigma_t^2))$, where $\mathcal{D}(t, p_t, \xi(\sigma_t^2))$ is—as in Christoffersen, Heston, and Jacobs (2013)—the value at time t of a traded asset with payoff at $T > t$ equal to $\mathcal{D}(p_T)$. By Ito's Lemma,

$$\begin{aligned} d\mathcal{D}_t &= \frac{\partial\mathcal{D}}{\partial t}dt + \frac{\partial\mathcal{D}}{\partial p}dp^c + \frac{\partial\mathcal{D}}{\partial\xi(\sigma^2)}\, d\xi^c(\sigma_t^2) + \frac{\partial^2\mathcal{D}}{\partial p\partial\xi(\sigma^2)}\sigma_t\Lambda(\sigma_t)\rho(\sigma_t)\, dt + \frac{1}{2}\frac{\partial^2\mathcal{D}}{\partial p}\sigma_t^2 p_t^2\, dt + \frac{1}{2}\frac{\partial^2\mathcal{D}}{\partial\xi(\sigma^2)}\Lambda^2(\sigma_t)\, dt + \Big[\mathcal{D}\left(t, p_t \exp^{c_{r,t}^{JJ}}, \xi(\sigma_t^2) + c_{\sigma,t}^{JJ}\right) \\ &\quad - \mathcal{D}(t, p_t, \xi(\sigma_t^2))\Big]\, dJ_{r,\sigma} + [\mathcal{D}(t, p_t \exp^{c_{r,t}^{J}}, \xi(\sigma_t^2)) - \mathcal{D}(t, p_t, \xi(\sigma_t^2))]\, dJ_r + [\mathcal{D}(t, p_t, \xi(\sigma_t^2) + c_{\sigma,t}^{J}) - \mathcal{D}(t, p_t, \xi(\sigma_t^2))]\, dJ_\sigma \end{aligned}$$

and

$$\frac{dY_{3,t}}{Y_{3,t}} = \frac{dM_t^c}{M_t} + \frac{d\mathcal{D}_t^c}{\mathcal{D}_t} + \frac{1}{\mathcal{D}_t}\left(\frac{\partial\mathcal{D}}{\partial p}\left(\phi\sigma_t^2 + \psi\rho(\sigma_t)\Lambda(\sigma_t)\sigma_t\right) + \frac{\partial\mathcal{D}}{\partial\xi(\sigma^2)}\left(\phi\sigma_t\Lambda(\sigma_t)\rho(\sigma_t) + \psi\Lambda^2(\sigma_t)\right)\right) dt + c_{Y_{3,t}}^{r,\sigma}\, dJ_{r,\sigma} + c_{Y_{3,t}}^{r}\, dJ_r + c_{Y_{3,t}}^{\sigma}\, dJ_\sigma$$

with

$$c_{Y_{3,t}}^{r,\sigma} = \frac{\mathcal{D}(t, p_t \exp^{c_{r,t}^{JJ}}, \xi(\sigma_t^2) + c_{\sigma,t}^{JJ}) \exp^{\phi c_{r,t}^{JJ} + \psi c_{\sigma,t}^{JJ}} - \mathcal{D}(t, p_t, \xi(\sigma_t^2))}{\mathcal{D}(t, p_t, \xi(\sigma_t^2))},$$

$$c^r_{Y_{3,t}} = \frac{\mathcal{D}(t,p_t\exp^{c^J_{r,t}},\xi(\sigma_t^2))\exp^{\phi c^J_{r,t}} - \mathcal{D}(t,p_t,\xi(\sigma_t^2))}{\mathcal{D}(t,p_t,\xi(\sigma_t^2))},$$

$$c^\sigma_{Y_{3,t}} = \frac{\mathcal{D}(t,p_t,\xi(\sigma_t^2)+c^J_{\sigma,t})\exp^{\psi c^J_{\sigma,t}} - \mathcal{D}(t,p_t,\xi(\sigma_t^2))}{\mathcal{D}(t,p_t,\xi(\sigma_t^2))}.$$

By equating $(1/dt)\mathrm{E}[dY_{3,t}/Y_{3,t}]$ to zero, we obtain

$$\begin{aligned}
&\Big(-r-\lambda_{r,\sigma}(\sigma_t)\Big[\exp^{\phi c^{JJ}_{r,t}+\psi c^{JJ}_{\sigma,t}}-1\Big]-\lambda_r(\sigma_t)\Big[\exp^{\phi c^J_{r,t}}-1\Big]-\lambda_\sigma(\sigma_t)\Big[\exp^{\psi c^J_{\sigma,t}}-1\Big]\Big)\mathcal{D}\big(t,p_t,\xi(\sigma_t^2)\big)\\
&\quad+\frac{\partial\mathcal{D}}{\partial t}+\frac{\partial\mathcal{D}}{\partial p}\big(r+\tilde\mu(\sigma^2)\big)+\frac{\partial\mathcal{D}}{\partial\xi(\sigma^2)}m(\sigma_t)+\frac{\partial^2\mathcal{D}}{\partial p\partial\xi(\sigma^2)}\sigma_t\Lambda(\sigma_t)\rho(\sigma_t)+\frac{1}{2}\frac{\partial^2\mathcal{D}}{\partial p}\sigma_t^2p_t^2+\frac{1}{2}\frac{\partial^2\mathcal{D}}{\partial\xi(\sigma^2)}\Lambda^2(\sigma_t)\\
&\quad+\frac{\partial\mathcal{D}}{\partial p}\big(\phi\sigma_t^2+\psi\sigma_t\Lambda(\sigma_t)\rho(\sigma_t)\big)+\frac{\partial\mathcal{D}}{\partial\xi(\sigma^2)}\big(\phi\sigma_t\Lambda(\sigma_t)\rho(\sigma_t)+\psi\Lambda^2(\sigma_t)\big)\\
&\quad+\lambda_{r,\sigma}(\sigma_t)\mathrm{E}[\mathcal{D}(t,p_t\exp^{c^{JJ}_{r,t}},\xi(\sigma_t^2)+c^{JJ}_{\sigma,t})\exp^{\phi c^{JJ}_{r,t}+\psi c^{JJ}_{\sigma,t}}-\mathcal{D}(t,p_t,\xi(\sigma_t^2))]+\lambda_r(\sigma_t)\mathrm{E}[\mathcal{D}(t,p_t\exp^{c^J_{r,t}},\xi(\sigma_t^2))-\mathcal{D}(t,p_t,\xi(\sigma_t^2))]\\
&\quad+\lambda_\sigma(\sigma_t)\mathrm{E}[\mathcal{D}(t,p_t,\xi(\sigma_t^2)+c^J_{\sigma,t})-\mathcal{D}(t,p_t,\xi(\sigma_t^2))]=0,
\end{aligned}$$

and, simplifying, we have

$$\begin{aligned}
&-r\mathcal{D}\big(t,p_t,\xi(\sigma_t^2)\big)+\frac{\partial\mathcal{D}}{\partial t}+\frac{\partial\mathcal{D}}{\partial p}\big(r+\tilde\mu(\sigma_t^2)\big)+\frac{\partial\mathcal{D}}{\partial\xi(\sigma^2)}m(\sigma_t)+\frac{\partial^2\mathcal{D}}{\partial p\partial\xi(\sigma^2)}\sigma_t\Lambda(\sigma_t)\rho(\sigma_t)+\frac{1}{2}\frac{\partial^2\mathcal{D}}{\partial p}\sigma_t^2p_t^2+\frac{1}{2}\frac{\partial^2\mathcal{D}}{\partial\xi(\sigma^2)}\Lambda^2(\sigma_t)\\
&\quad+\frac{\partial\mathcal{D}}{\partial p}\big(\phi\sigma_t^2+\psi\sigma_t\Lambda(\sigma_t)\rho(\sigma_t)\big)+\frac{\partial\mathcal{D}}{\partial\xi(\sigma^2)}\big(\phi\sigma_t\Lambda(\sigma_t)\rho(\sigma_t)+\psi\Lambda^2(\sigma_t)\big)\\
&\quad+\lambda_{r,\sigma}(\sigma_t)\mathrm{E}[\mathcal{D}(t,p_t\exp^{c^{JJ}_{r,t}},\xi(\sigma_t^2)+c^{JJ}_{\sigma,t})\exp^{\phi c^{JJ}_{r,t}+\psi c^{JJ}_{\sigma,t}}-\mathcal{D}(t,p_t,\xi(\sigma_t^2))]+\lambda_r(\sigma_t)\mathrm{E}[\mathcal{D}(t,p_t\exp^{c^J_{r,t}},\xi(\sigma_t^2))-\mathcal{D}(t,p_t,\xi(\sigma_t^2))]\\
&\quad+\lambda_\sigma(\sigma_t)\mathrm{E}[\mathcal{D}(t,p_t,\xi(\sigma_t^2)+c^J_{\sigma,t})-\mathcal{D}(t,p_t,\xi(\sigma_t^2))]=0.
\end{aligned}\tag{B.1}$$

Now, for every traded asset with price $\mathcal{D}(t,p,\xi(\sigma^2))$, we also have, slightly abusing notation and using (σ_t,p_t) in place of (σ,p),

$$\begin{aligned}
&-r\mathcal{D}\big(t,p_t,\xi(\sigma_t^2)\big)+\frac{\partial\mathcal{D}}{\partial t}+\frac{\partial\mathcal{D}}{\partial p}\Big(r-\lambda^*_{r,\sigma}(\sigma_t)\mathrm{E}\Big(\exp^{c^{JJ*}_{r,t}}-1\Big)-\lambda^*_r(\sigma_t)\mathrm{E}\Big(\exp^{c^{J*}_{r,t}}-1\Big)\Big)\\
&\quad+\frac{\partial\mathcal{D}}{\partial\xi(\sigma^2)}m^*(\sigma_t)+\frac{1}{2}\frac{\partial^2\mathcal{D}}{\partial p}\sigma_t^2p_t^2+\frac{1}{2}\frac{\partial^2\mathcal{D}}{\partial\xi(\sigma^2)}\Lambda^2(\sigma_t)+\frac{\partial^2\mathcal{D}}{\partial p\partial\xi(\sigma^2)}\sigma_t\Lambda(\sigma_t)\rho(\sigma_t)\\
&\quad+\lambda^*_{r,\sigma}(\sigma_t)\mathrm{E}[(\mathcal{D}(t,p_t\exp^{c^{JJ*}_{r,t}},\xi(\sigma_t^2)+c^{JJ*}_{\sigma,t})-\mathcal{D}(t,p_t,\xi(\sigma_t^2)))]+\lambda^*_r(\sigma_t)\mathrm{E}[(\mathcal{D}(t,p_t\exp^{c^{J*}_{r,t}},\xi(\sigma_t^2))-\mathcal{D}(t,p_t,\xi(\sigma_t^2)))]\\
&\quad+\lambda^*_\sigma(\sigma_t)\mathrm{E}[(\mathcal{D}(t,p_t,\xi(\sigma_t^2)+c^{J*}_{\sigma,t})-\mathcal{D}(t,p_t,\xi(\sigma_t^2)))]=0.
\end{aligned}\tag{B.2}$$

If we now equate Eq. (B.1) with Eq. (B.2), we obtain

$$\lambda_{r,\sigma}(\sigma_t)\mathrm{E}[\exp^{\phi c^{JJ}_{r,t}+\psi c^{JJ}_{\sigma,t}}(\mathcal{D}(t,p_t\exp^{c^{JJ}_{r,t}},\xi(\sigma_t^2)+c^{JJ}_{\sigma,t})-\mathcal{D}(t,p_t,\xi(\sigma_t^2)))]=\lambda^*_{r,\sigma}(\sigma_t)\mathrm{E}[(\mathcal{D}(t,p_t\exp^{c^{JJ*}_{r,t}},\xi(\sigma_t^2)+c^{JJ*}_{\sigma,t})-\mathcal{D}(t,p_t,\xi(\sigma_t^2)))],\tag{B.3}$$

$$\lambda_r(\sigma_t)\mathrm{E}[\exp^{\phi c^{JJ}_{r,t}}(\mathcal{D}(t,p_t\exp^{c^J_{r,t}},\xi(\sigma_t^2))-\mathcal{D}(t,p_t,\xi(\sigma_t^2)))]=\lambda^*_r(\sigma_t)\mathrm{E}[(\mathcal{D}(t,p_t\exp^{c^{J*}_{r,t}},\xi(\sigma_t^2))-\mathcal{D}(t,p_t,\xi(\sigma_t^2)))],\tag{B.4}$$

$$\lambda_\sigma(\sigma_t)\mathrm{E}[\exp^{\psi c^{JJ}_{\sigma,t}}(\mathcal{D}(t,p_t,\xi(\sigma_t^2)+c^J_{\sigma,t})-\mathcal{D}(t,p_t,\xi(\sigma_t^2)))]=\lambda^*_\sigma(\sigma_t)\mathrm{E}[(\mathcal{D}(t,p_t,\xi(\sigma_t^2)+c^{J*}_{\sigma,t})-\mathcal{D}(t,p_t,\xi(\sigma_t^2)))],\tag{B.5}$$

as well as Eqs. (19) and (20).□

Appendix C. Finite sample issues

C.1. Measuring spot variance with microstructure noise and jumps

The spot variance measures are obtained as follows. Each day, we split the day into six consecutive slots of 65 min (the end-points are called "knots" in the main text), starting at 9.30 am. Denote by $\sigma^2_{t,i}$ the spot variance estimator corresponding to day t, knot i, $i=1,\ldots,6$. To estimate $\sigma^2_{t,i}$, we first compute logarithmic prices $\log(p_{t,i,k})$ every minute, with $k=0,\ldots,65$, by pre-averaging all observed transaction prices inside each minute. Doing so softens the impact of microstructure noise by reducing the contaminations handled by pre-averaging (see Jacod, Li, Mykland, Podolskij, and Vetter, 2009, for further discussions). Next, we compute one-minute logarithmic returns $r_{t,i,k}=\log(p_{t,i,k})-\log(p_{t,i,k-1})$ for $k=1,\ldots,65$. At each knot and for each day, the spot variance estimates are obtained by applying the jump-robust threshold bipower variation estimator in Eq. (13) to the previous 65 one-minute returns. These estimates are then scaled by 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 variances. Finally, we also record the last logarithmic price $\log(p_{t,i})$ in each hour.

C.2. Infinitesimal cross-moment estimation with estimated variances

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

$$\tilde{\hat{\vartheta}}_{p_1,p_2}(\sigma)=\frac{\sum_{t=1}^{T-1}\sum_{i=1}^{N}\mathbf{K}\left(\frac{\tilde{\hat{\sigma}}_{t,i}-\sigma}{h}\right)(\log(p_{t+1,i})-\log(p_{t,i}))^{p_1}(\log(\hat{\sigma}^2_{t+1,i})-\log(\hat{\sigma}^2_{t,i}))^{p_2}}{\Delta\sum_{t=1}^{T}\sum_{i=1}^{N}\mathbf{K}\left(\frac{\tilde{\hat{\sigma}}_{t,i}-\sigma}{h}\right)},\quad p_2=0,1$$

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

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

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

with $N=6$ and $c=2.61/64$ 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,

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

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

C.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:

$$\tilde{\hat{\vartheta}}^{\Delta}_{p_1,p_2}=\tilde{\hat{\vartheta}}_{p_1,p_2}-\frac{\Delta}{2}\sum_{j_1=0}^{p_1}\sum_{j_2=0}^{p_2}\binom{p_1}{j_1}\binom{p_2}{j_2}\tilde{\hat{\vartheta}}^{\Delta}_{j_1,j_2}\tilde{\hat{\vartheta}}^{\Delta}_{p_1-j_1,p_2-j_2}$$

with $\tilde{\hat{\vartheta}}^{\Delta}_{0,0}=0$.

C.4. Simulations

The reliability of the estimated infinitesimal cross-moments, and their finite-sample adjustments, is evaluated by simulation. We simulate the bivariate system in Eq. (15), described in Section 7, with the parameters in column 5, 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 $\hat{\sigma}^2_t=\sigma^2_t+\sqrt{2.61/64}\sigma^2_t\varepsilon_t$, where ε_t is a standard normal draw and $\sqrt{2.61/64}\sigma^2_t$ is the (asymptotic) standard deviation of threshold bipower variation estimator computed with 65 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 our asymptotics, h smoothes more in regions with sparse observations near the boundaries of the range of

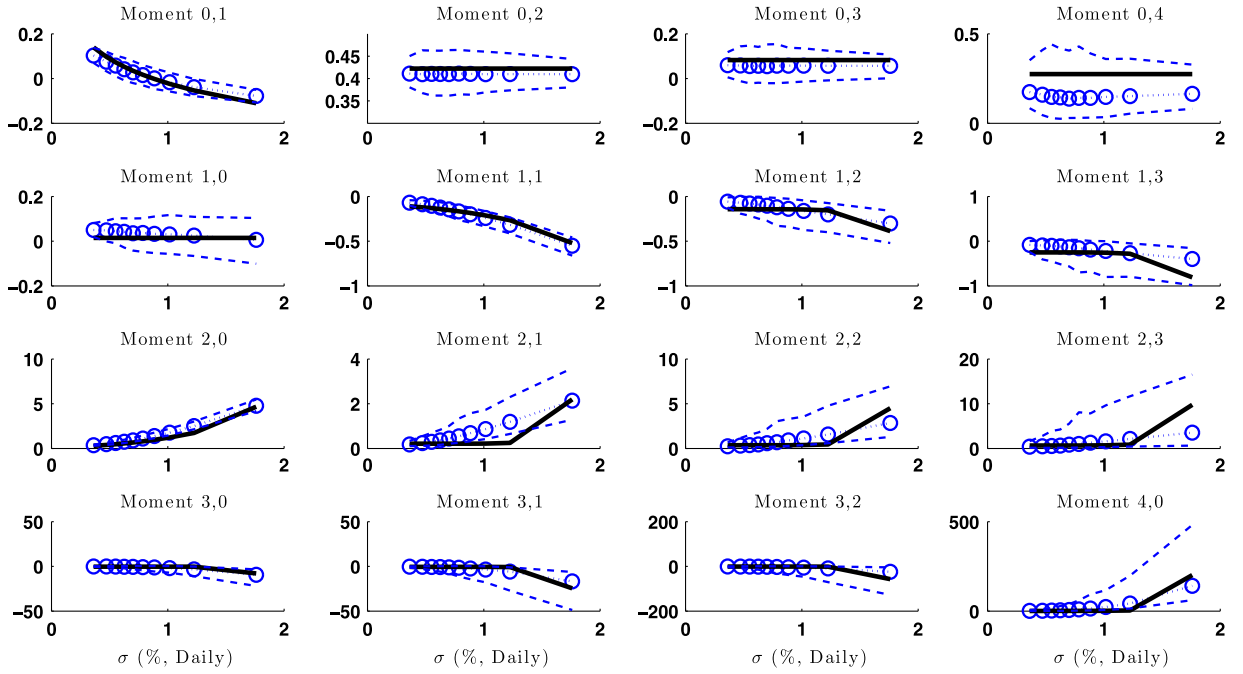

Fig. C1. Monte Carlo. We report simulation results for the estimated moments. The dotted lines with circles are median estimates across simulations. The dashed lines are 95% confidence bands. The solid lines are the true moments.

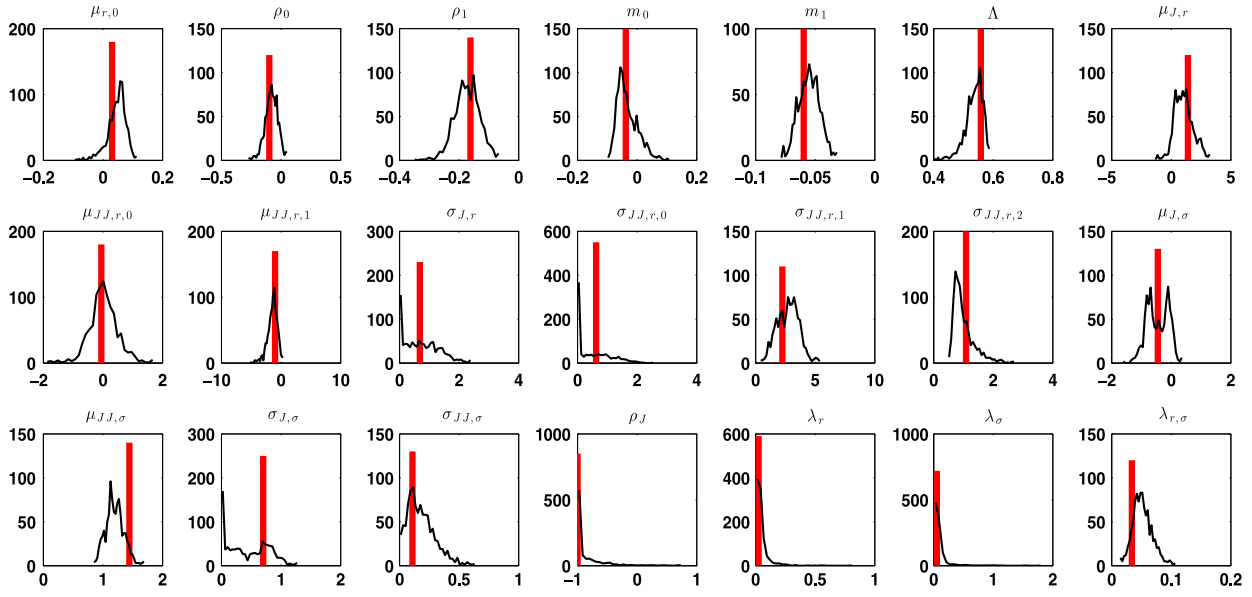

Fig. C2. Monte Carlo. We report simulation results for the parametric model in Table 2. The vertical bars are the true parameters. The solid lines are histograms of the estimated parameters on simulations.

the estimated variance series. We implement 1,500 simulations. Fig. C1 presents the generated moments, together with the estimated median and 95% confidence bands on the replications. The figure shows that the small-sample procedures we employ are able to deliver accurate estimates of the infinitesimal cross-moments. Fig. C2 displays the estimated parameters (on simulations) of the parametric model in Section 7 with the parameters in Table 2, column 5. Estimation of the parameters via NIMM is satisfactory and not affected by meaningful biases.

References

Aït-Sahalia, Y., Fan, J., Li, Y., 2013. The leverage effect puzzle: disentangling sources of bias at high frequency. Journal of Financial Economics 109, 224–249.
Alizadeh, S., Brandt, M., Diebold, F., 2002. High and low frequency exchange rate volatility dynamics: range-based estimation of stochastic volatility models. Journal of Finance 57, 1047–1092.
Andersen, T., Benzoni, L., 2011. Stochastic volatility. In: Meyers, R.A. (Ed.), Complex Systems in Finance and Econometrics, Springer, New York, pp. 694–726.
Andersen, T., Benzoni, L., Lund, J., 2002. An empirical investigation of continuous-time equity return models. Journal of Finance 57, 1239–1284.
Andersen, T., Dobrev, D., Schaumburg, E., 2012. Jump robust volatility estimation using nearest neighbor truncation. Journal of Econometrics 169 (1), 75–93.
Bakshi, G., Cao, C., Chen, Z., 1997. Empirical performance of alternative option pricing models. Journal of Finance 52, 2003–2049.
Bandi, F., Renò, R., 2009. Nonparametric stochastic volatility. Unpublished paper. Johns Hopkins University and University of Siena.
Bandi, F., Renò, R., 2012. Time-varying leverage effects. Journal of Econometrics 169 (1), 94–113.
Barndorff-Nielsen, O.E., Shephard, N., 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.
Broadie, M., Chernov, M., Johannes, M., 2007. Model specification and risk premia: evidence from futures options. Journal of Finance 62 (3), 1453–1490.
Chacko, G., Viceira, L., 2003. Spectral GMM estimation of continuous-time processes. Journal of Econometrics 116 (1–2), 259–292.
Chernov, M., Gallant, R., Ghysels, E., Tauchen, G., 2003. Alternative models for stock price dynamics. Journal of Econometrics 116 (1), 225–258.
Chernov, M., Ghysels, E., 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., Nardari, F., Shephard, N., 2002. Markov chain Monte Carlo methods for stochastic volatility models. Journal of Econometrics 108 (2), 281–316.
Christoffersen, P., Heston, S., Jacobs, K., 2013. Capturing option anomalies with a variance-dependent pricing kernel. Review of Financial Studies 16 (8), 1963–2006.
Cont, R., Kokholm, T., 2011. A consistent pricing model for index options and volatility derivatives. Mathematical Finance 29, 1–27.
Corsi, F., Pirino, D., Renò, R., 2010. Threshold bipower variation and the impact of jumps on volatility forecasting. Journal of Econometrics 159, 276–288.
Dobrev, D., Szerszen, P.J., 2010. The information content of high-frequency data for estimating equity return models and forecasting risk. International Finance Discussion Papers 1005. Board of Governors of the Federal Reserve System (U.S.).
Duffie, D., Pan, J., Singleton, K., 2000. Transform analysis and asset pricing for affine jump-diffusions. Econometrica 68, 1343–1376.
Eraker, B., 2004. Do stock prices and volatility jump? reconciling evidence from spot and option prices. Journal of Finance 59 (3), 1367–1404.
Eraker, B., Johannes, M., Polson, N., 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.
Harris, L., 1989. The October 1987 S&P 500 stock-futures basis. Journal of Finance 44 (1), 77–99.
Harvey, A., Shephard, N., 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., Kluppelberg, C., Muller, G., 2012a. Functional relationships between price and volatility jumps and their consequences for discretely observed data. Journal of Applied Probability 49, 901–914.
Jacod, J., Kluppelberg, C., Muller, G., 2012b. Testing for non-correlation between price and volatility jumps. Unpublished paper. Université Pierre et Marie Curie Paris VI and Technische Univesität München.
Jacod, J., Li, Y., Mykland, P., Podolskij, M., Vetter, M., 2009. Microstructure noise in the continuous case: the pre-averaging approach. Stochastic Processes and Their Applications 119 (7), 2249–2276.
Jacod, J., Shiryaev, A.N., 2003. Limit Theorems for Stochastic Processes. Springer, Berlin.
Jacod, J., Todorov, V., 2009. Testing for common arrivals of jumps for discretely observed multidimensional processes. Annals of Statistics 37, 1792–1838.
Jacod, J., Todorov, V., 2010. Do price and volatility jump together? Annals of Applied Probability 20, 1425–1469.
Jacquier, E., Polson, N., Rossi, P., 1994. Bayesian analysis of stochastic volatility models. Journal of Business and Economic Statistics 12, 371–389.
Jacquier, E., Polson, N., Rossi, P., 2004. Bayesian analysis of stochastic volatility models with fat-tails and correlated errors. Journal of Econometrics 122 (1), 185–212.
Jones, C., 2003. Nonlinear mean reversion in the short-term interest rate. Review of Financial Studies 16, 765–791.
Lee, S., Mykland, P., 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: Hansen, L.P., Aït-Sahalia, Y. (Eds.), Handbook of Financial Econometrics, North-Holland, Amsterdam, pp. 691–766.
Revuz, D., Yor, M., 1994. Continuous Martingales and Brownian Motion. Springer-Verlag, Berlin.
Todorov, V., Tauchen, G., 2010. Activity signature functions for high-frequency data analysis. Journal of Econometrics 54, 125–138.
Todorov, V., Tauchen, G., 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., 2012. A semiparametric stochastic volatility model. Journal of Econometrics 167, 473–482.