Realized Semicovariances

This version: February 6, 2019

Tim Bollerslev[a,*], Jia Li[b], Andrew J. Patton[b], Rogier Quaedvlieg[c]

[a]*Department of Economics, Duke University, NBER and CREATES*
[b]*Department of Economics, Duke University*
[c]*Erasmus School of Economics, Erasmus University Rotterdam*

Abstract

We propose a new decomposition of the realized covariance matrix into components based on the signs of the underlying high-frequency returns. Under an asymptotic setting in which the sampling interval goes to zero, we derive the asymptotic properties of the resulting *realized semicovariance* measures. The first-order asymptotic results highlight how the concordant components and the mixed-sign component load differently on economic information concerning stochastic correlation and jumps. The second-order asymptotics, taking the form of a novel non-central limit theorem, further reveals the fine structure underlying the concordant semicovariances, as manifest in the form of co-drifting and dynamic "leverage" type effects. In line with this anatomy, we empirically document distinct dynamic dependencies in the different realized semicovariance components based on data for a large cross-section of individual stocks. We further show that the accuracy of portfolio return variance forecasts may be significantly improved by using the realized semicovariance matrices to "look inside" the realized covariance matrices for signs of direction.

Keywords: High-frequency data; realized variances; semicovariances; co-jumps; volatility forecasting.
JEL: C22, C51, C53, C58

[☆]We would like to thank conference and seminar participants at Aarhus, Boston, Ca'Foscari Venice, Chicago, Cologne, Gerzensee, ITAM, Konstanz, Lancaster, Padova, Pennsylvania, PUC Rio, QUT, Rutgers, Toulouse for helpful comments and suggestions. We would also like to thank Bingzhi Zhao for kindly providing us with the cleaned high-frequency data underlying our empirical investigations. Quaedvlieg was financially supported by the Netherlands Organisation for Scientific Research (NWO) Grant 451-17-009. This paper subsumes the 2017 paper "Realized Semicovariances: Looking for Signs of Direction Inside the Covariance Matrix" by Bollerslev, Patton, and Quaedvlieg.

[*]Corresponding author: Department of Economics, Duke University, 213 Social Sciences Building, Box 90097, Durham, NC 27708-0097, United States. Email: boller@duke.edu. A supplemental appendix to the paper is available at http://www.econ.duke.edu/∼boller/research.html.

1. Introduction

The covariance matrix of asset returns arguably constitutes the most crucial input for asset pricing, portfolio and risk management decisions. Correspondingly, there is a substantial literature devoted to the estimation, modeling, and prediction of covariance matrices dating back more than half-a-century (e.g., Kendall (1953), Elton and Gruber (1973), and Bauwens, Laurent, and Rombouts (2006)). Meanwhile, a large and rapidly growing recent literature has forcefully advocated for the use of high-frequency intraday data for a more reliable estimation of lower-frequency realized return covariance matrices (e.g., Andersen, Bollerlsev, Diebold, and Labys (2003), Barndorff-Nielsen and Shephard (2004), and Barndorff-Nielsen, Hansen, Lunde, and Shephard (2011)).

Set against this background, we propose a new decomposition of the realized covariance matrix into three *realized semicovariance matrix* components dictated by the signs of the underlying high-frequency returns. The realized semicovariance matrices may be seen as a high-frequency multivariate extension of the semivariances first proposed in the finance literature several decades ago (e.g., Markowitz (1959), Mao (1970), Hogan and Warren (1972, 1974), and Fishburn (1977)). They also naturally extend the high-frequency based realized semivariances proposed more recently (Barndorff-Nielsen, Kinnebrock, and Shephard (2010)) to a multivariate context.

To fix ideas, let $X_t = (X_{1,t}, \ldots, X_{d,t})^\top$ denote a d-dimensional log-price process, sampled on a regular time grid $\{i\Delta_n : 0 \leq i \leq [T/\Delta_n]\}$ over some fixed time span $T > 0$. Let the ith return of X be denoted by $\Delta_i^n X \equiv X_{i\Delta_n} - X_{(i-1)\Delta_n}$. The *realized covariance matrix* (Barndorff-Nielsen and Shephard (2004)) is then defined as:

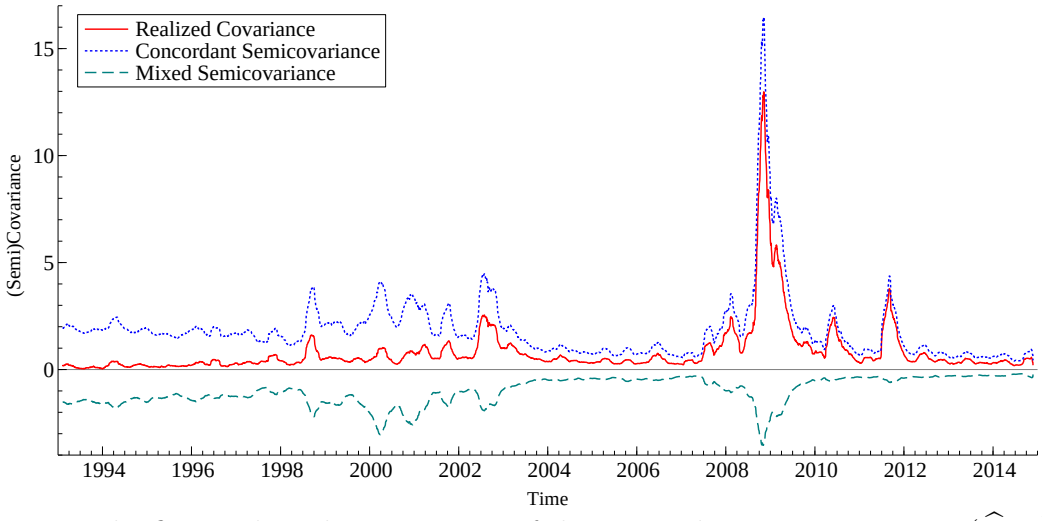
$$\widehat{C} \equiv \sum_{i=1}^{[T/\Delta_n]} (\Delta_i^n X)\,(\Delta_i^n X)^\top . \tag{1.1}$$

If we let $p(x) \equiv \max\{x, 0\}$ and $n(x) \equiv \min\{x, 0\}$ denote the component-wise positive and negative elements of the real vector x, the corresponding "positive," "negative," and "mixed" *realized semicovariance matrices* are then simply defined as:

$$\begin{aligned} \widehat{P} &\equiv \sum_{i=1}^{[T/\Delta_n]} p(\Delta_i^n X)\, p(\Delta_i^n X)^\top , \quad \widehat{N} \equiv \sum_{i=1}^{[T/\Delta_n]} n(\Delta_i^n X)\, n(\Delta_i^n X)^\top , \\ \widehat{M} &\equiv \sum_{i=1}^{[T/\Delta_n]} \left(p(\Delta_i^n X)\, n(\Delta_i^n X)^\top + n(\Delta_i^n X)\, p(\Delta_i^n X)^\top \right). \end{aligned} \tag{1.2}$$

Note that $\widehat{C} = \widehat{P} + \widehat{N} + \widehat{M}$ for any sampling frequency Δ_n. The concordant realized semicovariance matrices, $\widehat{P}$ and $\widehat{N}$, are defined as sums of vector outer-products and thus are positive semidefinite. By contrast, the mixed semicovariance matrix, $\widehat{M}$, has diagonal elements that are identically zero, and thus is necessarily indefinite.

Figure 1: Decomposition of the Realized Covariance



Note: The figure plots the time series of the concordant semicovariance ($\widehat{P}+\widehat{N}$), the mixed semicovariance ($\widehat{M}$) and the realized covariance ($\widehat{C}$). Each series is constructed as the 25-day moving average of the corresponding daily realized semicovariance measures averaged across 500 randomly selected pairs of S&P 500 stocks over the 1993–2014 period. The dataset is described in Section 4.

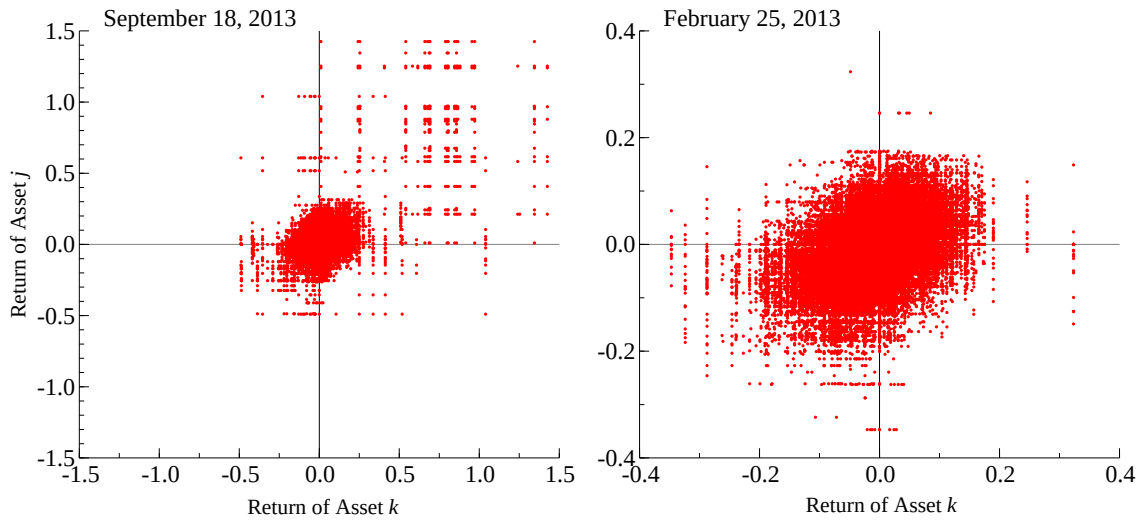
As an initial empirical illustration of the different dynamic dependencies and information conveyed by the realized semicovariances, Figure 1 plots the daily realized covariance averaged across 500 randomly-selected pairs of S&P 500 stocks, together with its concordant ($\widehat{P}+\widehat{N}$) and mixed ($\widehat{M}$) semicovariance components.[1] The mixed component is, of course, always negative, while the concordant component is always positive. The two components are typically fairly similar in absolute magnitude during "normal" time periods. In periods of high volatility, however, the concordant component increases substantially more than the mixed component declines, in line with the widely held belief that during periods of financial market stress correlations and tail dependencies among most financial assets tend to increase. As such, the (total) realized covariance is largely determined by the concordant realized semicovariance components in these "crisis" periods.

To help understand these empirical features, consider a simple setting in which the vector log-price process X_t is generated by a driftless Brownian motion with unit volatility and constant correlation ρ.[2] By the law of large numbers, the probability limits (as

[1] More precisely, we compute for each day $\overline{C} \equiv (1/500)\sum_{j\neq k}\widehat{C}_{jk}$, where the sum is over 500 randomly selected asset pairs, and define $\overline{P}$, $\overline{N}$ and $\overline{M}$ similarly. More detailed descriptions of the data and the procedures used in calculating the realized semicovariances are provided in Section 4 below. To avoid cluttering the figure, we sum $\overline{P}$ and $\overline{N}$ into the single concordant component, and smooth the daily measures using a 25-day moving average.

[2] Although stylized, this simple model captures the central force in the first-order asymptotic behavior of the semicovariance estimators in the no-jump setting. Theorem 1 provides a more general asymptotic result for Itô semimartingales.

Figure 2: Signed Return-Pairs for DJIA Stocks.



Note: The figure shows a scatter plot of the one-minute returns of each pair of the 30 Dow Jones Industrial Average stocks on two days in 2013.

$\Delta_n \to 0$) of the (j,k) off-diagonal elements of the realized semicovariance matrices are then given by

$$\operatorname{plim} \widehat{P}_{jk} = \operatorname{plim} \widehat{N}_{jk} = \psi(\rho), \quad \operatorname{plim} \widehat{M}_{jk} = -2\psi(-\rho), \tag{1.3}$$

where

$$\psi(\rho) = (2\pi)^{-1} \left(\rho \arccos(-\rho) + \sqrt{1-\rho^2} \right), \tag{1.4}$$

corresponding to $\mathbb{E}[Z_1 Z_2 1_{\{Z_1<0, Z_2<0\}}]$ for (Z_1, Z_2) bivariate standard normally distributed with correlation ρ. As these expressions illustrate, the relative contribution of the concordant and mixed semicovariance components to the (total) covariance depends crucially on the value of ρ. Indeed, as ρ increases to 1, the limiting value of the concordant component $\widehat{P} + \widehat{N}$ approaches one while the mixed component $\widehat{M}$ approaches zero, and vice versa when ρ decreases to -1. This, of course, is also consistent with the empirical observation from Figure 1 that the concordant component accounts for most of the covariance in periods of market stress, which are generally believed to be accompanied by increased positive correlations.

This simple diffusive setting highlights the potentially different information conveyed by the concordant and the mixed semicovariance components. It does not, however, reveal any differences between the $\widehat{P}$ and $\widehat{N}$ components as they have the same limits in this stylized setting. This is at odds with the intuition that these signed measures ought to carry distinct economic information as a result of the types of "news" that arrive on different days. By way of illustration, consider the high-frequency returns for the 30 Dow Jones Industrial Average (DJIA) stocks on the two different days presented in Figure

2.[3] On September 18, 2013, shown in the left panel, the Federal Reserve announced that it would *not* taper its asset-purchasing program, in contrast to what the market had been anticipating; individual stocks responded abruptly with positive jumps at the announcement time, resulting in much larger estimates of $\widehat{P}$ than $\widehat{N}$. By contrast, the right panel shows the returns on February 25, 2013, when the DJIA drifted down by 1.5% over the course of the day amid concerns, according to market anecdotes, about the political uncertainty in Italy, in turn resulting in a much larger estimates of $\widehat{N}$ than $\widehat{P}$.[4] Hence, the empirical estimates of $\widehat{P}$ and $\widehat{N}$ can indeed be very different depending on the "directional" content of the news and the corresponding information processing process, whether it manifests in the form of price jumps and/or apparent price drifts. As such, the difference $\widehat{P}-\widehat{N}$, which we refer to as the *concordant semicovariance differential* (CSD), is likely to carry additional useful information.

Motivated by these empirical observations, in Section 2 we derive both the first- and the second-order asymptotics for the positive and the negative semicovariance estimators in a general Itô semimartingale setting, focusing particularly on a deeper understanding of their information content. The limit theory identifies three distinct channels through which $\widehat{P}$ and $\widehat{N}$ may differ: directional co-jumps, a type of "co-drifting," and a specific form of dynamic leverage effect. While the "co-jump" channel manifests straightforwardly in the first-order asymptotics, the latter two channels take the form of second-order bias terms in a non-central limit theorem. These bias terms are specific to our analysis of the semicovariances, and methodologically speaking sets our asymptotic analysis apart from the usual high-frequency econometric analysis, in which central limit theorems are generally applied for the purpose of conducting statistical inference (see, e.g., Aït-Sahalia and Jacod (2014)). By contrast, the main purpose of our higher-order asymptotic results is to further "dissect" the semicovariance estimators, thereby allowing for additional theoretical and empirical insights by appropriately comparing and contrasting the relevant terms.

More specifically, in line with the different intraday price behavior evident for the two days depicted in Figure 2, we rely on a standard truncation technique (Mancini (2001)) to obtain two CSD estimators, corresponding to separate jump and diffusive "signals," respectively. For the former, we establish a feasible central limit theorem that may be used to construct formal statistical inference. For the latter diffusive component (which is related to price drift and a form of leverage effect and, hence, much more complicated), we provide a standard error estimator that quantifies its sampling variability in a well-defined sense, and which, under more restrictive regularity conditions, also results in an asymptotically valid and unbiased test.

Implementing the new inference procedures with high-frequency data for the 30 DJIA

[3]Further details on the underlying data are provided in Section 3.

[4]Source: https://money.cnn.com/2013/02/25/investing/stocks-markets/index.html.

stocks over a nine-year period reveals strong statistical evidence for significant differences in the $\widehat{P}$ and $\widehat{N}$ semicovariance components on many different days. Consistent with economic intuition, we find that large differences in the jump semicovariance components are typically associated with "sharp" public news announcements (e.g., FOMC announcements). Meanwhile, large differences in the diffusive semicovariance components are typically associated with more difficult-to-interpret news, which manifests in the form of common price drifts within the day.[5] In addition to the more detailed discussion of such event days, Section 3 further documents that days with significantly different $\widehat{P}$ and $\widehat{N}$ are associated with subsequent different dynamic dependencies both across and within the three realized semicovariance components.

This naturally suggests that decomposing the realized covariance matrix into its semicovariance components may be useful for volatility forecasting. In an effort to corroborate this conjecture, we analyze a large cross-section of stocks comprised of all of the S&P 500 constituents spanning more than two decades. We show that the out-of-sample forecasts of return variances for portfolios comprised of up to one hundred stocks may indeed be significantly improved by "looking inside" the covariance matrix through the lens of the new semicovariance measures. Moreover, the gains from doing so increase with the number of stocks included in the portfolio, although in line with the gains from naïve portfolio diversification the relative gains appear to plateau at around 30-40 stocks in the portfolio. Further dissecting the forecasting gains, we find that the models that incorporate the additional information that resides in the realized semicovariances generally respond faster to new information compared with standard models that only use realized variances or realized semivariances (see, e.g., Corsi (2009) and Patton and Sheppard (2015)). Interestingly, while the erratic nature of volatility during the financial crisis leads most existing volatility forecasting models to *reduce* the weight on recent observations, the new semicovariance-based models developed here actually *increase* the weight, primarily due to an increase in the short-run importance of the negative semicovariance component.

The forecasting gains obtained through the use of the realized semicovariances are naturally linked to the early work on parametric asymmetric volatility models (e.g., Kroner and Ng (1998) and Cappiello, Engle, and Sheppard (2006)). The new realized semicovariance measures themselves and our CSD-based tests, in particular, are also closely related to other tests for asymmetric dependencies that have previously been proposed in the literature (e.g., Longin and Solnik (2001), Ang and Chen (2002) and Hong, Tu, and Zhou (2007)). They are also related to existing empirical work on the correlations between asset returns in "bear" versus "bull" markets, and notions of asymmetric tail dependencies (e.g., Patton (2004), Poon, Rockinger, and Tawn (2004), and Tjøsthem and Hufthammer (2013)), along with more recent work on high-frequency based co-skewness

[5]The two days plotted in Figure 2 also correspond to these two scenarios, and are indeed detected by using the new inference method, as further discussed in Section 3 below.

and co-kurtosis measures (e.g., Neuberger (2012) and Amaya, Christoffersen, Jacobs, and Vasquez (2015)), as well as recent work on jumps and co-jumps (e.g., Das and Uppal (2004), Bollerslev, Law, and Tauchen (2008), Lee and Mykland (2008), Mancini and Gobbi (2012), Jacod and Todorov (2009), Aït-Sahalia and Xiu (2016) and Li, Todorov, and Tauchen (2017b)). In contrast to all of these existing studies, however, we retain the covariance matrix as the summary measure of dependence, and instead use information from signed high-frequency returns to "look inside" the realized covariance matrix as a way to reveal additional information about the inherent dependencies, both dynamically and cross-sectionally at a given point in time.

The rest of the paper is organized as follows. Section 2 presents the first- and second-order asymptotic properties of the realized semicovariances. Section 3 discusses our empirical findings related to the implementation of the semicovariance-based tests. Our results pertaining to the use of the realized semicovariances in the construction of improved volatility forecasts are discussed in Section 4. Section 5 concludes. Technical regularity conditions and proofs are deferred to the appendix. Additional robustness checks and extensions are available in a supplemental appendix.

2. Asymptotic properties of realized semicovariances

This section presents the asymptotic properties of the realized semicovariances. Sections 2.1 and 2.2 present the first- and the second-order asymptotics, respectively. Section 2.3 describes feasible inference methods. Below, for a matrix A, we denote its (j,k) element by A_{jk} and its transpose by $A^\top$. Convergence in probability and stable convergence in law are denoted by $\xrightarrow{\mathbb{P}}$ and $\xrightarrow{\mathcal{L}\text{-}s}$, respectively. All limits are for the sample frequency $\Delta_n \to 0$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

2.1. First-order asymptotic properties

Suppose that the log-price vector X_t is an Itô semimartingale of the form

$$X_t = X_0 + \int_0^t b_s ds + \int_0^t \sigma_s dW_s + J_t, \tag{2.1}$$

where b is the $\mathbb{R}^d$-valued drift process, W is a d-dimensional standard Brownian motion, σ is the $d \times d$ dimensional stochastic volatility matrix and J is a finitely active pure-jump process. We denote the spot covariance matrix of X by $c_t \equiv \sigma_t \sigma_t^\top$ and further set $v_{j,t} \equiv \sqrt{c_{jj,t}}$ and $\rho_{jk,t} \equiv c_{jk,t}/v_{j,t}v_{k,t}$. That is, $v_{j,t}$ and $\rho_{jk,t}$ denote the spot volatility of asset j and the spot correlation coefficient between assets j and k, respectively. We explicitly allow for so-called "leverage effect" (i.e., dependence between changes in the price and changes in volatility), stochastic volatility of volatility, volatility jumps and price-volatility co-jumps.

We begin by characterizing the first-order limiting behavior of the realized semicovariance estimators defined by equation (1.2) in the introduction. Let ΔX_s denote the price jump occurring at time s, if a jump occurred, and set it to zero if no jump occurred at time s. Further define

$$\begin{aligned} P^{\dagger} &\equiv \sum_{s \leq T} p(\Delta X_s) p(\Delta X_s)^{\top}, \\ N^{\dagger} &\equiv \sum_{s \leq T} n(\Delta X_s) n(\Delta X_s)^{\top}, \\ M^{\dagger} &\equiv \sum_{s \leq T} \left(p(\Delta X_s) n(\Delta X_s)^{\top} + n(\Delta X_s) p(\Delta X_s)^{\top} \right). \end{aligned}$$

These measures characterize the discontinuous parts of the semicovariance measures, as formally spelled out in the following theorem.

Theorem 1 *Under Assumption 1 in the appendix,* $(\widehat{P}, \widehat{N}, \widehat{M}) \xrightarrow{\mathbb{P}} (P, N, M)$*, where* P*,* N *and* M *are* $d \times d$ *matrices with their* (j,k) *elements given by*

$$\begin{aligned} P_{jk} &\equiv \int_0^T v_{j,s} v_{k,s} \psi(\rho_{jk,s}) ds + P^{\dagger}_{jk}, \\ N_{jk} &\equiv \int_0^T v_{j,s} v_{k,s} \psi(\rho_{jk,s}) ds + N^{\dagger}_{jk}, \\ M_{jk} &\equiv -2 \int_0^T v_{j,s} v_{k,s} \psi(-\rho_{jk,s}) ds + M^{\dagger}_{jk}, \end{aligned}$$

and $\psi(\cdot)$ *is defined in equation (1.4).*

It follows from Theorem 1 that each of the realized semicovariances contains both diffusive and jump covariation components. Importantly, the limiting variables P and N share exactly the same diffusive component, but their jump components differ. In particular,

$$\widehat{P} - \widehat{N} \xrightarrow{\mathbb{P}} P - N = P^{\dagger} - N^{\dagger}.$$

That is, the first-order asymptotic behavior of the concordant semicovariance differential (CSD) is fully characterized by the "directional co-jumps." Consequently, in line with the stylized model in equation (1.2) discussed in the introduction, Theorem 1 cannot distinguish the information conveyed by $\widehat{P}$ and $\widehat{N}$ in periods when there are no jumps. Hence, in order to reveal the differential information inherent in the realized measures more generally, we turn next to a more refined second-order asymptotic analysis.

2.2. Second-order asymptotic properties

Since the main theoretical lessons about the second-order asymptotic behavior of the concordant realized semicovariance components can be readily learnt in a bivariate

setting, we set $d = 2$ and focus on the analysis of $\widehat{P}_{12}$ and $\widehat{N}_{12}$ throughout this subsection.[6] Correspondingly, we also write ρ_t in place of $\rho_{12,t}$ for simplicity.

We need to impose some additional structure on the volatility dynamics. In particular, we will assume that the stochastic volatility σ_t is also an Itô semimartingale of the form

$$\sigma_t = \sigma_0 + \int_0^t \tilde{b}_s ds + \int_0^t \tilde{\sigma}_s dW_s + \widetilde{M}_t + \sum_{s \leq t} \Delta\sigma_s 1_{\{\|\Delta\sigma_s\| > \underline{\sigma}\}}, \tag{2.2}$$

where $\tilde{b}$ is the drift, $\tilde{\sigma}$ is the $d \times d \times d$ tensor-valued "volatility of volatility" process, $\widetilde{M}$ is a local martingale that is orthogonal to the Brownian motion W.[7] We note that the local martingale $\widetilde{M}$ may contain both an orthogonal (w.r.t., W) Brownian component and compensated "small" volatility jumps. Meanwhile, the term $\sum_{s \leq t} \Delta\sigma_s 1_{\{\|\Delta\sigma_s\| > \underline{\sigma}\}}$ collects the "big" volatility jumps (with an arbitrary but fixed threshold $\underline{\sigma} > 0$), which may occur during major news announcements (see, e.g., Bollerslev, Li, and Xue (2018)). Further regularity conditions on the σ process are collected in the appendix.

Theorem 2, below, describes the $\mathcal{F}$-stable convergence in law of the normalized statistic $\Delta_n^{-1/2}(\widehat{P}_{12} - P_{12}, \widehat{N}_{12} - N_{12})$. The limit variable turns out to be fairly complicated, but it may be succinctly expressed as

$$B^{(1)} + B^{(2)} + \zeta + \tilde{\zeta} + \xi, \tag{2.3}$$

where $B^{(1)}$ and $B^{(2)}$ are bias terms, and $(\zeta, \tilde{\zeta}, \xi)$ capture sampling variabilities that arise from various sources. Before presenting the actual limit theorem, we begin by briefly describing each of these separate terms.

Bias components due to price drift, $B^{(1)}$. The first type of bias is related to the price drift, which is defined for the semicovariance estimators $\widehat{P}_{12}$ and $\widehat{N}_{12}$ respectively as

$$B^{(1)} = \begin{pmatrix} B_P^{(1)} \\ B_N^{(1)} \end{pmatrix} \equiv \begin{pmatrix} \frac{1}{2\sqrt{2\pi}} \int_0^T \left(\frac{b_{1,s}}{v_{1,s}} + \frac{b_{2,s}}{v_{2,s}} \right) v_{1,s} v_{2,s} \left(1 + \rho_s\right) ds \\ -\frac{1}{2\sqrt{2\pi}} \int_0^T \left(\frac{b_{1,s}}{v_{1,s}} + \frac{b_{2,s}}{v_{2,s}} \right) v_{1,s} v_{2,s} \left(1 + \rho_s\right) ds \end{pmatrix}. \tag{2.4}$$

We note that $B_P^{(1)} = -B_N^{(1)}$, so the effects of price drift on $\widehat{P}$ and $\widehat{N}$ will always be of opposite signs. Other things equal, this bias term is proportional to the average spot "Sharpe ratio" of the two assets

$$\frac{1}{2} \left(\frac{b_{1,s}}{v_{1,s}} + \frac{b_{2,s}}{v_{2,s}} \right). \tag{2.5}$$

[6] Our theory can be easily extended to include the mixed component $\widehat{M}$ and to a joint setting across different asset pairs. These extensions, however, necessitate more complicated notation (and discussion). Since our empirical study only involves pairwise analysis, we purposely refrain from presenting these most general results to avoid unnecessary technicality. Details for the more general setting are available on request.

[7] By convention, the (j, k) element of the stochastic integral $\int_0^t \tilde{\sigma}_s dW_s$ equals $\sum_{l=1}^d \int_0^t \tilde{\sigma}_{jkl,s} dW_{l,s}$.

In particular, the $B^{(1)}$ term tends to be more pronounced when the two assets drift in the same direction, akin to a "*co-drift*" type phenomenon.

Bias components due to continuous price-volatility covariation, $B^{(2)}$. The second bias term stems from the fact that the volatility matrix process σ_t may be partially driven by the Brownian motion W (i.e., $\tilde{\sigma} \neq 0$), corresponding to a "dynamic leverage" type effect. To more precisely describe this bias term for $\widehat{P}$, define $f_1(x) \equiv 1_{\{x_1 \geq 0\}} \max\{x_2, 0\}$ and $f_2(x) \equiv \max\{x_1, 0\} 1_{\{x_2 \geq 0\}}$ and then set, for any 2×2 matrix A,

$$F_j(A) \equiv \mathbb{E}\left[f_j(AW_1) \int_0^1 W_s dW_s^\top \right], \quad j = 1, 2.$$

The bias term in $\widehat{P}$ due to the common price-volatility Brownian dependence may then be expressed as

$$B_P^{(2)} \equiv \sum_{j=1}^{2} \int_0^T \text{Trace}\left[\tilde{\sigma}_{j,s} F_j(\sigma_s) \right] ds, \tag{2.6}$$

where $\tilde{\sigma}_{j,s}$ denotes the 2×2 matrix $[\tilde{\sigma}_{jkl,s}]_{1 \leq k,l \leq 2}$. The bias term $B_N^{(2)}$ for $\widehat{N}$ may be defined similarly, and it can be shown that $B_P^{(2)} = -B_N^{(2)}$.[8]

Diffusive sampling error spanned by price risk, ζ. The third component captures the sampling variability in $p(\Delta_i^n X_1)\, p(\Delta_i^n X_2)$ and $n(\Delta_i^n X_1)\, n(\Delta_i^n X_2)$ that is spanned by the Brownian price shock $\sigma_t dW_t$. Formally, $\zeta = (\zeta_P, \zeta_N)^\top$ where

$$\zeta_P = -\zeta_N \equiv \int_0^T \left(c_s^{-1} \gamma_s \right)^\top (\sigma_s dW_s), \tag{2.7}$$

and the γ_t process is defined as[9]

$$\gamma_t \equiv \frac{(1+\rho_t)^2 v_{1,t} v_{2,t}}{2\sqrt{2\pi}} \begin{pmatrix} v_{1,t} \\ v_{2,t} \end{pmatrix}. \tag{2.8}$$

Note that the quadratic covariation matrix of the local martingale ζ equals $\int_0^T \Gamma_s ds$, where

$$\Gamma_t \equiv \begin{pmatrix} \gamma_t^\top c_t^{-1} \gamma_t & -\gamma_t^\top c_t^{-1} \gamma_t \\ -\gamma_t^\top c_t^{-1} \gamma_t & \gamma_t^\top c_t^{-1} \gamma_t \end{pmatrix}. \tag{2.9}$$

Diffusive sampling error orthogonal to price risk, $\tilde{\zeta}$. While ζ defined above captures the diffusive risk in the semicovariances spanned by the Brownian shocks to the price

[8] More precisely, if we set $g_1(x) \equiv 1_{\{x_1 \leq 0\}} \min\{x_2, 0\}$, $g_2(x) \equiv \min\{x_1, 0\} 1_{\{x_2 \leq 0\}}$, and $G_j(A) \equiv \mathbb{E}[g_j(AW_1) \int_0^1 W_s dW_s^\top]$, the $B_N^{(2)}$ bias term is then given by $B_N^{(2)} \equiv \sum_{j=1}^2 \int_0^T \text{Trace}[\tilde{\sigma}_{j,s} G_j(\sigma_s)]\, ds$. The assertion that $B_P^{(2)} = -B_N^{(2)}$ follows from $f_j(-x) = -g_j(x)$ for $j = 1, 2$.

[9] Further, $\gamma_{(i-1)\Delta_n}$ is computed as $\mathbb{E}[f(\sigma_{(i-1)\Delta_n} \Delta_i^n W / \Delta_n^{1/2}) \sigma_{(i-1)\Delta_n} \Delta_i^n W / \Delta_n^{1/2} | \mathcal{F}_{(i-1)\Delta_n}]$, where $f(x) = p(x_1)p(x_2)$. Hence, $c_{(i-1)\Delta_n}^{-1} \gamma_{(i-1)\Delta_n}$ corresponds to the population regression coefficient obtained from regressing $f(\sigma_{(i-1)\Delta_n} \Delta_i^n W / \Delta_n^{1/2})$ on the Brownian shock $\sigma_{(i-1)\Delta_n} \Delta_i^n W / \Delta_n^{1/2}$.

process, $\tilde{\zeta}$ captures the diffusive risk component orthogonal to those stocks. This limit variable may be represented by its $\mathcal{F}$-conditional distribution as

$$\tilde{\zeta} = \begin{pmatrix} \tilde{\zeta}_P \\ \tilde{\zeta}_N \end{pmatrix} = \int_0^T \bar{\gamma}_s^{1/2} d\widetilde{W}_s, \tag{2.10}$$

where $\widetilde{W}$ is a 2-dimensional standard Brownian motion that is independent of the σ-field $\mathcal{F}$, and the $\bar{\gamma}$ process is defined by

$$\bar{\gamma}_t \equiv \overline{\Gamma}_t - \Gamma_t, \tag{2.11}$$

where

$$\overline{\Gamma}_t \equiv v_{1,t}^2 v_{2,t}^2 \begin{pmatrix} \Psi(\rho_t) - \psi(\rho_t)^2 & -\psi(\rho_t)^2 \\ -\psi(\rho_t)^2 & \Psi(\rho_t) - \psi(\rho_t)^2 \end{pmatrix}, \tag{2.12}$$

and

$$\Psi(\rho) \equiv \frac{3\rho\sqrt{1-\rho^2} + (1+2\rho^2)\arccos(-\rho)}{2\pi}, \tag{2.13}$$

with $\Psi(\rho)$ corresponding to $\mathbb{E}\left[Z_1^2 Z_2^2 1_{\{Z_1<0, Z_2<0\}}\right]$ for (Z_1, Z_2) standard normally distributed with correlation ρ.

Jump-induced sampling error, ξ. The price jumps also induce sampling errors. Let $\mathcal{T}_j$ for $j \in \{1,2\}$ denote the collection of jump times of $(X_{j,t})_{t\in[0,T]}$, with the corresponding "signed" subsets denoted by

$$\mathcal{T}_{j+} \equiv \{\tau \in \mathcal{T}_j : \Delta X_{j,\tau} > 0\}, \quad \mathcal{T}_{j-} \equiv \{\tau \in \mathcal{T}_j : \Delta X_{j,\tau} < 0\}.$$

For each $\tau \in \mathcal{T}_1 \cup \mathcal{T}_2$ associate the variables $(\kappa_\tau, \tilde{\xi}_{\tau-}, \tilde{\xi}_{\tau+})$ that are, conditionally on $\mathcal{F}$, mutually independent with the following conditional distributions: $\kappa_\tau \sim \text{Uniform}[0,1]$, $\tilde{\xi}_{\tau-} \sim \mathcal{MN}(0, c_{\tau-})$, and $\tilde{\xi}_{\tau+} \sim \mathcal{MN}(0, c_\tau)$. Further define $\tilde{\eta}_\tau = (\tilde{\eta}_{1,\tau}, \tilde{\eta}_{2,\tau})^\top \equiv \sqrt{\kappa_\tau}\tilde{\xi}_{\tau-} + \sqrt{1-\kappa_\tau}\tilde{\xi}_{\tau+}$.[10] The limiting variable $\xi = (\xi_P, \xi_N)^\top$ may then be expressed as

$$\begin{aligned} \xi_P \equiv & \sum_{\tau \in \mathcal{T}_{1+} \cap \mathcal{T}_{2+}} (\Delta X_{1,\tau}\tilde{\eta}_{2,\tau} + \Delta X_{2,\tau}\tilde{\eta}_{1,\tau}) \\ & + \sum_{\tau \in \mathcal{T}_{1+} \setminus \mathcal{T}_2} \Delta X_{1,\tau} p(\tilde{\eta}_{2,\tau}) + \sum_{\tau \in \mathcal{T}_{2+} \setminus \mathcal{T}_1} \Delta X_{2,\tau} p(\tilde{\eta}_{1,\tau}), \end{aligned}$$

[10] It is instructive to recall the intuition for these limiting variables. The uniform variable κ_τ captures the indeterminacy of the jump time within a discrete sampling interval, while $\sqrt{\kappa_\tau}\tilde{\xi}_{\tau-}$ and $\sqrt{1-\kappa_\tau}\tilde{\xi}_{\tau+}$ capture the distribution of the Brownian increment before and after the jump time, respectively. The variable $\tilde{\eta}_\tau$ in turn represents the limiting behavior of the Brownian sampling error around the jump time τ.

$$\begin{aligned}\xi_N \;\equiv\; & \sum_{\tau\in\mathcal{T}_{1-}\cap\mathcal{T}_{2-}} (\Delta X_{1,\tau}\tilde{\eta}_{2,\tau} + \Delta X_{2,\tau}\tilde{\eta}_{1,\tau}) \\ & + \sum_{\tau\in\mathcal{T}_{1-}\setminus\mathcal{T}_2} \Delta X_{1,\tau} n\left(\tilde{\eta}_{2,\tau}\right) + \sum_{\tau\in\mathcal{T}_{2-}\setminus\mathcal{T}_1} \Delta X_{2,\tau} n\left(\tilde{\eta}_{1,\tau}\right).\end{aligned}$$

Note that the first component in ξ_P (resp. ξ_N) concerns the times when both assets have positive (resp. negative) jumps, while the other two terms are active when one asset jumps upwards (resp. downwards) and the other asset does not jump. Interestingly, the latter terms involve the half-truncated doubly mixed Gaussian variable $p(\tilde{\eta}_\tau)$ (resp. $n\left(\tilde{\eta}_\tau\right)$). To the best of our knowledge, this type of limiting distribution is new to the literature.

With the definitions above, we are now ready to state the stable convergence in law of the realized semicovariances.

Theorem 2 *Under Assumption 2 in the appendix,*

$$\Delta_n^{-1/2}\begin{pmatrix}\widehat{P}_{12}-P_{12}\\ \widehat{N}_{12}-N_{12}\end{pmatrix} \xrightarrow{\mathcal{L}\text{-}s} \begin{pmatrix}B_P^{(1)}\\ B_N^{(1)}\end{pmatrix} + \begin{pmatrix}B_P^{(2)}\\ B_N^{(2)}\end{pmatrix} + \begin{pmatrix}\zeta_P\\ \zeta_N\end{pmatrix} + \begin{pmatrix}\tilde{\zeta}_P\\ \tilde{\zeta}_N\end{pmatrix} + \begin{pmatrix}\xi_P\\ \xi_N\end{pmatrix}.$$

Theorem 2 depicts a non-central limit theorem for the positive and negative realized semicovariances, where $(B^{(1)}, B^{(2)})$ represent bias terms, while $(\zeta, \tilde{\zeta}, \xi)$ stem from various sources of "sampling errors." The latter sampling error terms are all formed as (local) martingales and have zero mean under mild integrability conditions.

However, the presence of the bias terms means that Theorem 2 is not directly suitable for the construction of confidence intervals for the (P, N) estimand; but that is *not* our goal either. Instead, the main insight derived from Theorem 2 is to reveal the differential second-order behavior of the realized semicovariances $\widehat{P}$ and $\widehat{N}$, about which the first-order asymptotics in Theorem 1 remains entirely silent in the absence of jumps. Indeed, while Theorem 1 states that $\widehat{P}$ and $\widehat{N}$ have the same limit in the no-jump case, Theorem 2 clarifies that they actually load on higher-order "signals" $B^{(1)}$ and $B^{(2)}$ in the exact opposite way (i.e., $B_P = -B_N$). From a theoretical perspective, this therefore explains why $\widehat{P}$ and $\widehat{N}$ may behave differently. From an empirical perspective, it helps guide our understanding of the actual $\widehat{P}$ and $\widehat{N}$ estimates discussed in Section 3, and the use of these measures in the construction of improved volatility forecasts discussed in Section 4.

2.3. Tests based on concordant semicovariance differentials

Even though Theorem 2 does not allow for the construction of standard confidence intervals, it is still possible to develop feasible inference methods for the difference $\widehat{P}-\widehat{N}$, that is, the CSD. In particular, the asymptotic theory in the previous subsection reveals three types of signals underlying the CSD: a directional co-jump effect (i.e., $P^\dagger - N^\dagger$),

a co-drifting effect (i.e., $B_P^{(1)} - B_N^{(1)}$), and a dynamic leverage effect (i.e., $B_P^{(2)} - B_N^{(2)}$). Empirically, it is of great interest to separate the variation due to jumps from that due to the diffusive price moves. Below, we use the standard truncation method (see, e.g., Mancini (2001, 2009)) to achieve such a separation. As in Section 2.2, we consider a bivariate setting, or $d = 2$, and focus on the inference for $\widehat{P}_{12} - \widehat{N}_{12}$.

The truncation method involves a sequence $u_n \in \mathbb{R}_+^2$ of truncation thresholds satisfying $u_{j,n} \asymp \Delta_n^{\varpi}$ for some $\varpi \in (0, 1/2)$ and $j \in \{1, 2\}$. Under our maintained assumption of finite activity jumps, it can be shown that the index set

$$\widehat{\mathcal{I}} \equiv \{i : -u_n \leq \Delta_i^n X \leq u_n \text{ does not hold}\}$$

locates the jumps of the vector log-price process X with probability approaching one.[11] As a result, the diffusive and jump returns may be separated, allowing for the separate estimation of the *diffusive* components of the semicovariances using the "small" (non-jump) returns:

$$\widehat{P}^{\star} \equiv \sum_{i \notin \widehat{\mathcal{I}}} p(\Delta_i^n X) p(\Delta_i^n X)^{\top}, \qquad \widehat{N}^{\star} \equiv \sum_{i \notin \widehat{\mathcal{I}}} n(\Delta_i^n X) n(\Delta_i^n X)^{\top},$$

and the *jump* components of the semicovariances using the "jump" returns:

$$\widehat{P}^{\dagger} \equiv \sum_{i \in \widehat{\mathcal{I}}} p(\Delta_i^n X) p(\Delta_i^n X)^{\top}, \qquad \widehat{N}^{\dagger} \equiv \sum_{i \in \widehat{\mathcal{I}}} n(\Delta_i^n X) n(\Delta_i^n X)^{\top}.$$

Given the aforementioned jump detection result, the asymptotic property of these truncated estimators can be established as a straightforward extension of Theorem 2, stated as follows.

Proposition 1 *Under Assumption 2 in the appendix, the following convergences hold jointly*

$$\Delta_n^{-1/2} \begin{pmatrix} \widehat{P}_{12}^{\dagger} - P_{12}^{\dagger} \\ \widehat{N}_{12}^{\dagger} - N_{12}^{\dagger} \end{pmatrix} \xrightarrow{\mathcal{L}\text{-}s} \xi, \quad \Delta_n^{-1/2} \begin{pmatrix} \widehat{P}_{12}^{\star} - P_{12}^{\star} \\ \widehat{N}_{12}^{\star} - N_{12}^{\star} \end{pmatrix} \xrightarrow{\mathcal{L}\text{-}s} B^{(1)} + B^{(2)} + \zeta + \tilde{\zeta},$$

where $P_{12}^{\star} = N_{12}^{\star} \equiv \int_0^T v_{1,s} v_{2,s} \psi(\rho_s) ds$.

Proposition 1 allows for the construction of feasible inference for each of the two separate CSD components, $\widehat{P}_{12}^{\dagger} - \widehat{N}_{12}^{\dagger}$ and $\widehat{P}_{12}^{\star} - \widehat{N}_{12}^{\star}$, respectively. We will refer to these as the jump (resp. diffusive) concordant semicovariance differential, or JCSD (resp. DCSD) for short.

[11] See Proposition 1 of Li, Todorov, and Tauchen (2017b). The inequality $-u_n \leq \Delta_i^n X \leq u_n$ is interpreted element-by-element.

We start with a discussion of how to implement the JCSD test, which is the simpler of the two as it admits a central limit theorem. In particular, it follows immediately from Proposition 1,

$$\Delta_n^{-1/2}\left(\widehat{P}_{12}^{\dagger}-\widehat{N}_{12}^{\dagger}-\left(P_{12}^{\dagger}-N_{12}^{\dagger}\right)\right)\xrightarrow{\mathcal{L}\text{-}s}\xi_P-\xi_N,$$

where, as discussed above, ξ_P and ξ_N are centered doubly-mixed Gaussian variables. Hence, to consistently estimate the distribution of the limiting variable, we first need to estimate the spot covariance matrix before and after each detected jump time. In order to do so, we choose an integer sequence k_n of local windows that satisfies $k_n\to\infty$ and $k_n\Delta_n\to 0$, and set, for each i,

$$\hat{c}_{i-}\equiv\frac{1}{k_n\Delta_n}\sum_{l=1}^{k_n}\left(\Delta_{i-l}^n X\right)\left(\Delta_{i-l}^n X\right)^{\top}1_{\left\{-u_n\le\Delta_{i-l}^n X\le u_n\right\}},$$
$$\hat{c}_{i+}\equiv\frac{1}{k_n\Delta_n}\sum_{l=1}^{k_n}\left(\Delta_{i+l}^n X\right)\left(\Delta_{i+l}^n X\right)^{\top}1_{\left\{-u_n\le\Delta_{i+l}^n X\le u_n\right\}}.$$

Algorithm 1 describes the requisite steps for implementing the resulting JCSD test for the null hypothesis $P_{12}^{\dagger}=N_{12}^{\dagger}$, that is, equal directional jump covariation.

Algorithm 1 (JCSD Test).
Step 1. Draw random variables $(\kappa_i^*,\tilde{\xi}_{i-}^*,\tilde{\xi}_{i+}^*)$ that are mutually independent such that $\kappa_i^*\sim$ Uniform$[0,1]$ and $\tilde{\xi}_{i\pm}^*\sim\mathcal{MN}(0,\hat{c}_{i\pm})$. Set $\tilde{\eta}_i^*=(\tilde{\eta}_{i,1}^*,\tilde{\eta}_{i,2}^*)=\sqrt{\kappa_i^*}\tilde{\xi}_{i-}^*+\sqrt{1-\kappa_i^*}\tilde{\xi}_{i+}^*$.
Step 2. Let $\Delta_i^n X_j^*\equiv\Delta_i^n X_j 1_{\{|\Delta_i^n X_j|>u_{j,n}\}}$ for $j\in\{1,2\}$ and set

$$\xi_P^*=\Delta_n^{-1/2}\sum_{i\in\widehat{\mathcal{I}}}\left(p\left(\Delta_i^n X_1^*+\Delta_n^{1/2}\tilde{\eta}_{i,1}^*\right)p\left(\Delta_i^n X_2^*+\Delta_n^{1/2}\tilde{\eta}_{i,2}^*\right)-p\left(\Delta_i^n X_1^*\right)p\left(\Delta_i^n X_2^*\right)\right),$$
$$\xi_N^*=\Delta_n^{-1/2}\sum_{i\in\widehat{\mathcal{I}}}\left(n\left(\Delta_i^n X_1^*+\Delta_n^{1/2}\tilde{\eta}_{i,1}^*\right)n\left(\Delta_i^n X_2^*+\Delta_n^{1/2}\tilde{\eta}_{i,2}^*\right)-n\left(\Delta_i^n X_1^*\right)n\left(\Delta_i^n X_2^*\right)\right).$$

Step 3. Repeat steps 1–2 many times. Compute the $1-\alpha$ (resp. α) quantile of $\xi_P^*-\xi_N^*$ as the critical value of $\Delta_n^{-1/2}(\widehat{P}_{12}^{\dagger}-\widehat{N}_{12}^{\dagger})$ for the null hypothesis $P_{12}^{\dagger}=N_{12}^{\dagger}$ in favor of the one-sided alternative $P_{12}^{\dagger}>N_{12}^{\dagger}$ (resp. $P_{12}^{\dagger}<N_{12}^{\dagger}$) at significance level α. □

Algorithm 1 may be seen as a parametric bootstrap that exploits the approximate (parametric) doubly-mixed Gaussian distribution of the detected jump returns given the estimated spot covariances, with ξ_P^* and ξ_N^* being the bootstrap analogue of the original normalized estimators. While this type of simulation-based inference is often used in the study of jumps, a non-standard feature of Algorithm 1 is its use of the truncated return $\Delta_i^n X_j^*=\Delta_i^n X_j 1_{\{|\Delta_i^n X_j|>u_{j,n}\}}$, which shrinks the detected diffusive returns to zero. This shrinkage is needed in situations where exactly one asset jumps at time τ, so that the sampling variability contributed by the other (no-jump) asset, say j, is given by the half-truncated doubly-mixed Gaussian variable like $p\left(\tilde{\eta}_{\tau,j}\right)$. This distribution may in turn

be mimicked by $\Delta_n^{-1/2} p(\Delta_i^n X_j^* + \Delta_n^{1/2} \tilde{\eta}_{i,j}^*) = p(\tilde{\eta}_{i,j}^*)$, which differs from the "un-shrunk" variable $\Delta_n^{-1/2} p(\Delta_i^n X_j + \Delta_n^{1/2} \tilde{\eta}_{i,j}^*)$.

Proposition 2 *Under Assumption 2 in the appendix, the conditional distribution of $\xi_P^* - \xi_N^*$ given the data converges in probability to the $\mathcal{F}$-conditional distribution of $\xi_P - \xi_N$ under the uniform metric. Consequently, the test described in Algorithm 1 has asymptotic level α under the null $\{P_{12}^\dagger = N_{12}^\dagger\}$ and asymptotic power of one under one-sided alternatives.*

We turn next to the conduct of feasible inference using the DCSD statistic $\widehat{P}_{12}^\star - \widehat{N}_{12}^\star$. This involves some additional non-standard theoretical subtlety. Proposition 1 implies that

$$\Delta_n^{-1/2} \left(\widehat{P}_{12}^\star - \widehat{N}_{12}^\star \right) - D_1 - D_2 \xrightarrow{\mathcal{L}\text{-}s} \zeta_P - \zeta_N + \tilde{\zeta}_P - \tilde{\zeta}_N, \tag{2.14}$$

where $D_1 \equiv B_P^{(1)} - B_N^{(1)}$ and $D_2 \equiv B_P^{(2)} - B_N^{(2)}$ capture co-drift and dynamic leverage effects, respectively. The first-order limiting variables $P_{12}^\star$ and $N_{12}^\star$ exactly cancel with each other in the $\widehat{P}_{12}^\star - \widehat{N}_{12}^\star$ difference. Consequently, as revealed by the convergence in (2.14), the remaining "signal" carried by the DCSD is given by the higher-order term $D_1 + D_2$, which is comparable in magnitude with the statistical noise term $\zeta_P - \zeta_N + \tilde{\zeta}_P - \tilde{\zeta}_N$ (defined as a local martingale). Since the signal-to-noise ratio does not diverge to infinity even in large samples, the resulting test is generally not consistent.

A further non-standard complication related to (2.14) stems from the fact that the limiting variable $\zeta_P - \zeta_N + \tilde{\zeta}_P - \tilde{\zeta}_N$ is generally not mixed Gaussian. Specifically, while $\tilde{\zeta}_P - \tilde{\zeta}_N$ is $\mathcal{F}$-conditional Gaussian, the remaining part

$$\zeta_P - \zeta_N = 2 \int_0^T \left(c_s^{-1} \gamma_s \right)^\top (\sigma_s dW_s)$$

is generally not mixed Gaussian unless, of course, the stochastic volatility σ is independent of the Brownian motion W that drives the diffusive price moves.[12]

Although these non-standard features of the limit theory prevent us from conducting formal tests in the most general setting, it is nevertheless possible to assess the sampling variability of $\widehat{P}_{12}^\star - \widehat{N}_{12}^\star$ in a well-defined way. Indeed, it follows from (2.7) and (2.10) that the quadratic variation of the continuous local martingale $\zeta_P - \zeta_N + \tilde{\zeta}_P - \tilde{\zeta}_N$ is given by

$$\Sigma^\star \equiv 2 \int_0^T v_{1,s}^2 v_{2,s}^2 \Psi(\rho_s) \, ds.$$

Therefore, $\sqrt{\Sigma^\star}$ may be naturally used as the "standard error" for gauging the sampling

[12] This complication arises from the fact that the functions $x \mapsto p(x_1) p(x_2)$ and $x \mapsto n(x_1) n(x_2)$ are not globally even. Consequently, the local covariance between $p(\Delta_i^n X_1) p(\Delta_i^n X_2)$ (or $n(\Delta_i^n X_1) n(\Delta_i^n X_2)$) and the Brownian increment $\Delta_i^n W$ is non-zero, resulting in $\gamma_t \neq 0$ in general.

variability of the centered variable $\Delta_n^{-1/2}(\widehat{P}_{12}^{\star} - \widehat{N}_{12}^{\star}) - (D_1 + D_2)$. The $\Sigma^{\star}$ variable is defined as an integrated functional of the spot covariance matrix and may therefore be consistently estimated using a nonparametric "plug-in" estimator $\widehat{\Sigma}^{\star}$.[13] These results suggest a DCSD-based test described in the following algorithm.

Algorithm 2 (DCSD Test).
Step 1. Define $\hat{v}_{1,i}$, $\hat{v}_{2,i}$ and $\hat{\rho}_i$ implicitly by decomposing the spot covariance matrix estimator $\hat{c}_{i+}$ as

$$\hat{c}_{i+} = \begin{pmatrix} \hat{v}_{1,i}^2 & \hat{\rho}_i \hat{v}_{1,i} \hat{v}_{2,i} \\ \hat{\rho}_i \hat{v}_{1,i} \hat{v}_{2,i} & \hat{v}_{2,i}^2 \end{pmatrix},$$

and set

$$\widehat{\Sigma}^{\star} \equiv \frac{2\Delta_n}{[T/\Delta_n] - k_n + 1} \sum_{i=0}^{[T/\Delta_n]-k_n} \hat{v}_{1,i}^2 \hat{v}_{2,i}^2 \Psi(\hat{\rho}_i). \tag{2.15}$$

Step 2. Use the $1-\alpha$ quantile of a standard normal distribution z_α as the critical value for the t-statistic $\Delta_n^{-1/2}(\widehat{P}_{12}^{\star} - \widehat{N}_{12}^{\star})/\sqrt{\widehat{\Sigma}^{\star}}$ for one-sided tests of the null hypothesis $D_1 + D_2 = 0$.

The DCSD test described in Algorithm 2 should be properly interpreted. In particular, due to the aforementioned lack of mixed Gaussianity for the limiting variable under the most general conditions, the t-statistic $\Delta_n^{-1/2}(\widehat{P}_{12}^{\star} - \widehat{N}_{12}^{\star})/\sqrt{\widehat{\Sigma}^{\star}}$ is generally not asymptotically standard normally distributed. For this reason, the asymptotic level of the proposed one-sided test is not guaranteed to be α. Instead, the t-statistic may be interpreted as a well-defined "signal-to-noise" measure, as opposed to a formal statistical test. That being said, mixed Gaussianity can be restored under the additional assumption that the volatility process σ is independent of the Brownian motion W. In this case, the limiting variable $\zeta_P - \zeta_N + \tilde{\zeta}_P - \tilde{\zeta}_N$ is, conditional on the σ-field generated by the σ process, centered Gaussian with conditional variance $\Sigma^{\star}$. Consequently, the t-statistic $\Delta_n^{-1/2}(\widehat{P}_{12}^{\star} - \widehat{N}_{12}^{\star})/\sqrt{\widehat{\Sigma}^{\star}}$ converges in σ-conditional law to a standard Normal distribution. Correspondingly, in that situation, it then also readily follows that the DCSD test described in Algorithm 2 has asymptotic level α under the null and is asymptotically unbiased, with strictly nontrivial power under the alternative.

[13] Theorem 9.4.1 in Jacod and Protter (2012) shows the consistency of a class of "plug-in" estimators for integrated volatility functionals. However, their theory requires the test function to have "polynomial growth" in the spot covariance matrix, which cannot be verified for the spot correlation. This restriction is relaxed in the extension of Li and Xiu (2016) and Li, Todorov, and Tauchen (2017a). Under our maintained assumptions, the latter theory can be directly invoked to show $\widehat{\Sigma}^{\star} \xrightarrow{\mathbb{P}} \Sigma^{\star}$.

3. Empirical semicovariance tests

We begin our empirical investigations by looking at the realized semicovariances for the 30 Dow Jones Industrial Average (DJIA) stocks.[14] Our estimation is based on one-minute returns obtained from the Trades and Quotes (TAQ) database, spanning the period from January 2006 to December 2014, for a total of 2,265 trading days. Our choice of a relatively high one-minute sampling frequency and a fairly short recent sample for this part of our analysis is dictated by the need to reliably estimate the spot covariance matrix used for implementing the tests described in the previous section.[15] To help more clearly pinpoint within-day market-wide price moves and economic events associated with the significance of the tests, we further exclude the returns for the first half-hour of the trading day in the calculation of the tests.

3.1. Economic news and differences in semicovariances

We start by calculating the concordant JCSD and DCSD semicovariance-based tests for all of the 435 unique DJIA stock pairs and 2,265 days in the sample, resulting in close to one million test statistics for each of the two tests.[16] We rely on one-sided versions of the tests and a 5% significance level. The average rejection rates for both of the tests far exceed this nominal level: the JCSD test rejects in favor of positive (resp. negative) co-jumps for about 30% (resp. 32%) of the pairs, while the DCSD test suggests a significantly positive (resp. negative) difference for 13% (resp. 10%) of all stock pairs.[17]

To help shed additional light on these test results, Table 1 lists the days with the most rejections for each of the two tests for each of the nine years in the sample. In addition to the date and the rejection frequencies, we also include a short description of the most important economic events that occurred on each of these days. As Panel A shows, all but one of the days with the most rejections for the JCSD test are associated with FOMC statements and/or changes in the federal funds rate, the only exception being a major geopolitical event in 2014. This finding is consistent with the prior literature that links high-frequency-detected jumps in individual assets with public news announcements (e.g., Andersen, Bollerslev, and Diebold (2007), Lee and Mykland (2008) and Lee (2012)). It is also in line with the literature on testing for co-jumps and the argument that those jumps

[14] We use the DJIA composition as of September 23, 2013, which remained unchanged until the end of our sample period.

[15] The choice of a one-minute sampling frequency also mirrors that of Li, Todorov, Tauchen, and Chen (2017), among others, in their estimation of spot covariances; see **?**, Hansen and Lunde (2006) and Jacod, Li, and Zheng (2017) for further discussion of market microstructure effects.

[16] We rely on the dynamic threshold advocated by Bollerslev and Todorov (2011a,b) based on three times the trailing bipower variation, adjusted for the intraday periodicity in the volatility. We set the local window $k_n = 45$.

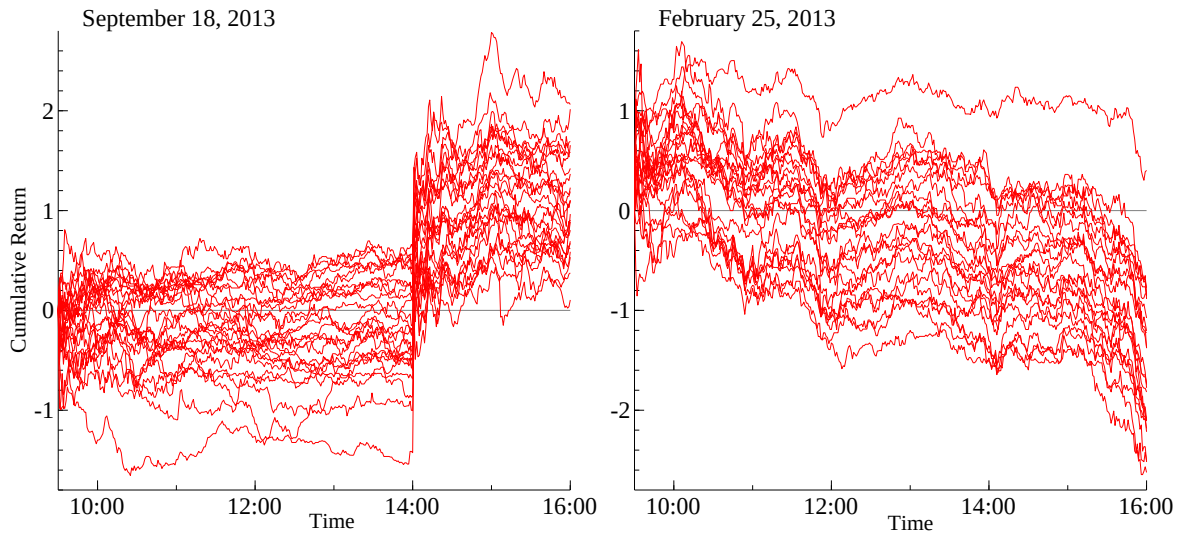
[17] We do not intend to make a formal statistical statement jointly across all pairs. Instead, we view the rejection frequencies as simple summary statistics of the pairwise test results.

Table 1: Top Rejection Days by Year

Year	Date	Direction	%	Headline Event
				Panel A. JCSD Test
2006	June 29	+	100	Fed raises short-term rate by a quarter-percentage point.
2007	September 18	+	99	Fed cuts short-term rate by a half-percentage point.
2008	December 16	+	100	Fed cuts short-term rate by a quarter-percentage point.
2009	March 18	+	100	Fed announces it will buy up to $300 billion in long-term Treasuries.
2010	August 10	+	98	Fed announces it will continue Quantitative Easing.
2011	September 22	+	100	Fed announces Operation Twist.
2012	September 13	+	100	Fed announces it will continue buying Mortgage Backed Securities.
2013	September 18	+	98	Fed announces it will sustain the asset buying program.
2014	August 5	−	97	Russian troops are reported lining on the borders of Ukraine.
				Panel B. DCSD Test
2006	July 19	+	57	Bernanke explains to the Senate Banking Committee how the Fed sees the economic slowdown.
2007	August 29	+	58	Bernanke writes letter to senator that Fed is monitoring and ready to step in if necessary.
2008	January 2	−	67	Markets react to poor manufacturing, housing and credit news.
2009	March 23	+	90	Obama administration announces its plan to buy $1 trillion in bad bank assets.
2010	July 7	+	76	EU reveals its first list of stress test banks.
2011	June 1	−	83	Moody's cuts Greece's bond rating by three notches.
2012	June 21	−	84	Rumors of Moody's downgrade for global banks.
2013	February 25	−	83	Political uncertainty surrounding Italian elections.
2014	February 3	−	85	Janet Yellen sworn in as the new Fed chair.

Note: The table reports the top rejection dates by year for the daily semicovariance-based tests across all 435 DJIA stocks-pairs. The first column gives the date, the second gives the direction in which the rejections occurred, and the third provides the fraction of pairs of stocks for which the test rejects at the 5% level in that direction. The final column summarizes headline economic news events for the different days. The top Panel A reports the results for the jump CSD test, while the bottom Panel B is based on the diffusive CSD test.

Figure 3: DJIA Cumulative Returns on Representative Event Days



Note: The figure plots the cumulative return of the 30 Dow Jones Industrial Average stocks on two of the event dates associated with market-wide jump CSD and diffusive CSD events from Table 1.

are naturally associated with economy-wide news that affect all assets (e.g., Bollerslev, Law, and Tauchen (2008) and Lahaye, Laurent, and Neely (2011)).

In contrast to the "sharp" economic events associated with the JCSD test, Panel B shows that the days with the most rejections for the DCSD test are typically associated with "softer" and more difficult-to-interpret information. Kyle-type equilibrium microstructure models (Kyle (1985)) can be used to establish a more formal economic link between the "soft" information and price drift (which drives the DCSD test through the co-drift effect). In these models, informed traders trade strategically with liquidity traders to maximize their profit, and they do so patiently for the sake of managing the market marker's belief. As shown more formally by Back (1992), informed traders' optimal order flow is smooth (i.e., differentiable) in time, which in turn determines the drift of the equilibrium price. In a general setting with stochastic liquidity, Collin-Dufresne and Fos (2016) further show that, in equilibrium, the price drift exhibits mean reversion towards the asset's true value, where the mean reversion is strong (weak) when the short-term liquidity is high (low) relative to the long-term liquidity. This equilibrium theory suggests that, other things equal, the price drift is greater in magnitude when there is higher level of mispricing and/or the revelation of the private information is more imminent. The latter effect may manifest in the form of "soft" information gathered through news articles.

To more clearly illustrate the distinct price dynamics on these "sharp" and "soft" event days identified by the two different tests, Figure 3 plots the cumulative returns throughout the day for each of the 30 DJIA stocks for two representative days selected

Table 2: Conditional First-Order Autocorrelations

		Conditioning Information			
	Unconditional	$\widehat{P}^{\star}_{jk} \gg \widehat{N}^{\star}_{jk}$	$\widehat{P}^{\star}_{jk} \ll \widehat{N}^{\star}_{jk}$	$\widehat{P}^{\dagger}_{jk} \gg \widehat{N}^{\dagger}_{jk}$	$\widehat{P}^{\dagger}_{jk} \ll \widehat{N}^{\dagger}_{jk}$
$\widehat{C}_{jk}$	0.641	0.732	0.757	0.642	0.687
$\widehat{P}_{jk}$	0.510	0.769	0.574	0.536	0.560
$\widehat{N}_{jk}$	0.658	0.608	0.773	0.659	0.704
$\widehat{M}_{jk}$	0.407	0.447	0.451	0.444	0.410

Note: The table reports the autocorrelations between the realized measures at time $t+1$ and t, conditional on the rejection of CSD-based tests at time t. The first column presents the unconditional autocorrelations, while columns two and three (resp. four and five) report the autocorrelations conditional on the DCSD (resp. JCSD) test, where we use the $\gg$ and $\ll$ symbols to denote statistically significant inequalities.

from Table 1: September 18, 2013 and February 25, 2013.[18] For the jump event detected by the JCSD test (left panel), all stocks experienced a large positive shock at 2pm when the FOMC meeting statement was released, stating that the Fed would sustain its asset-purchasing program. This announcement led to an immediate one-off average return of more than 1% for all stocks, whilst the prices appeared relatively stable before, and after, that statement release. By contrast, for the diffusive event detected by the DCSD test (right panel), we observe slow and steadily decreasing price paths throughout the day for all of the stocks. The total daily return is large, with the median daily return around negative 2%, but no "extreme" returns occurred for any of the stocks during the course of that day.

3.2. Dynamic dependencies and differences in semicovariances

The significant differences in the realized semicovariance components highlighted in the previous subsection naturally raises the question of whether they also lead to different dynamic dependencies in the different components across days. To shed light on this question, we compute, for each pair of stocks, the first-order autocorrelation between the realized semicovariance measures on day $t+1$ and day t, conditional on the DCSD or JCSD tests being significant on day t. The resulting conditional autocorrelations averaged across all of the stock pairs are reported in columns two through five of Table 2.

Looking first at the unconditional autocorrelations reported in the first column, we see that on average $\widehat{P}$ and $\widehat{M}$ both have lower persistence than the realized covariance $\widehat{C}$, while $\widehat{N}$ appears to be more strongly persistent. This effectively represents the multivariate equivalent of the univariate result of Patton and Sheppard (2015) that negative realized semivariances tend to be more persistent than their positive counterparts.

[18] Figure 2 discussed in the introduction is also based on these same two days. The Supplemental Appendix S1 presents analogous plots for all of the event days listed in Table 1.

Meanwhile, conditioning on the outcome of the CSD-based tests leads to notably different autocorrelations. In particular, $\widehat{C}$ generally appears to be more persistent after a diffusive event detected by the DCSD test. The effect of co-jumps and a rejection by the JCSD test is smaller and depends on the direction of the rejection: the persistence is slightly higher after negative co-jumps. This again may be seen as a multivariate extension of the effect of signed jump variation presented in Patton and Sheppard (2015).

The impact of the CSD-based tests on the realized semicovariance dynamics is somewhat more complex. The persistence of $\widehat{P}$, in particular, is higher following a positive DCSD event on the previous day, with a much smaller increase after a negative DCSD event. On the other hand, the persistence of $\widehat{N}$ actually drops after a positive DCSD event, but increases after a negative DCSD event. This is consistent with the idea that certain types of "soft" news is processed over multiple days. By comparison, the dynamic dependency after a JCSD jump event also shows a different pattern: the persistence of both the positive and the negative realized semicovariances only increases after negative co-jumps, while the effect of positive co-jumps are negligible.

4. Forecasting with realized semicovariances

The results discussed in the previous section highlight the additional information and economic insights afforded by realized semicovariances beyond those from standard realized covariances. The results also point to the existence of different dynamic dependencies, both conditionally and unconditionally, in the realized semicovariance components. In this section we further explore these empirical differences from the perspective of forecasting future variances and covariances.

To allow for the construction of larger dimensional portfolios, we expand our previous sample of 30 DJIA stocks to include all of the S&P 500 constituent stocks. We also consider a longer sample period from January 1993 to December 2014, for a total of 5,541 trading days. In order to reliably estimate models for covariances and semicovariances, we include only stocks with at least 2,000 daily observations, resulting in a total of 749 unique stocks. Most of these stocks are not as actively traded as the DJIA stocks, especially during the earlier part of the sample. Correspondingly, since we only require consistent estimates for this part of our analysis, we rely on a coarser 15-minute sampling scheme to construct the realized measures. Finally, similar to most existing work on volatility forecasting (e.g., Hansen, Huang, and Shek (2012), and Noureldin, Shephard, and Sheppard (2012)), we focus on the intra-daily period excluding the overnight returns.[19]

[19]The supplemental appendix contains empirical results that include the overnight returns. All of our main empirical findings remain qualitatively unaltered.

4.1. Vector autoregressions for realized semicovariances

The differential dynamic dependencies of the average realized semicovariance components for the S&P 500 stocks evident in the previously discussed Figure 1, underscored by the first-order sample autocorrelations for the more limited sample of DJIA stocks shown in Table 2, suggest that more accurate volatility and covariance forecasts may be obtained by separately modeling the realized semicovariance components, $\widehat{P}$, $\widehat{N}$, and $\widehat{M}$, that make up the realized covariance.

In order to investigate this conjecture and better characterize the dynamic dependencies inherent in the semicovariance components, we estimate a vector version of the popular HAR model of Corsi (2009), in which each of the elements in the realized semicovariance matrix is allowed to depend on its own daily, weekly, and monthly lags, as well as the lags of the other realized semicovariance components. Specifically, for each pair of assets (j,k) we estimate the following three-dimensional vector autoregression:

$$\begin{pmatrix} \widehat{P}_{jk,t} \\ \widehat{N}_{jk,t} \\ \widehat{M}_{jk,t} \end{pmatrix} = \begin{pmatrix} \phi_{jk,P} \\ \phi_{jk,N} \\ \phi_{jk,M} \end{pmatrix} + \Phi_{jk,Day} \begin{pmatrix} \widehat{P}_{jk,t-1} \\ \widehat{N}_{jk,t-1} \\ \widehat{M}_{jk,t-1} \end{pmatrix} + \Phi_{jk,Week} \begin{pmatrix} \widehat{P}_{jk,t-2:t-5} \\ \widehat{N}_{jk,t-2:t-5} \\ \widehat{M}_{jk,t-2:t-5} \end{pmatrix} + \Phi_{jk,Month} \begin{pmatrix} \widehat{P}_{jk,t-6:t-22} \\ \widehat{N}_{jk,t-6:t-22} \\ \widehat{M}_{jk,t-6:t-22} \end{pmatrix} + \begin{pmatrix} \epsilon^{P}_{jk,t} \\ \epsilon^{N}_{jk,t} \\ \epsilon^{M}_{jk,t} \end{pmatrix}, \tag{4.1}$$

where $\widehat{P}_{t-l:t-k} \equiv \frac{1}{k-l+1} \sum_{s=l}^{k} \widehat{P}_{t-s}$, with the other components defined analogously.[20]

The first three columns of Table 3 report the resulting parameter estimates averaged across 500 randomly selected (j,k) pairs of stocks. The table reveals a clear block structure in the coefficients of this general specification. Most notably, the dynamic dependencies in $\widehat{P}$ and $\widehat{N}$ are almost exclusively driven by the lagged $\widehat{N}$ terms, while the dynamic behavior of the mixed $\widehat{M}$ elements is primarily determined by their own lags, with the monthly lag receiving the largest weight.

The last two columns of Table 3 report the parameter estimates from regressing the realized covariances $\widehat{C}$ on the lagged realized semicovariances and the lagged covariances.[21] The model with individual semicovariances clearly reveals the most important components: the three lags of $\widehat{N}$ and the monthly lag of $\widehat{M}$ constitute the main drivers of the realized covariance $\widehat{C}$. Interestingly, the models based on the semicovariances also put a greater weight on more recent information compared to the standard HAR model

[20]To simplify the interpretation of the estimates, we define the weekly variables to exclude the daily lag and the monthly variables to similarly exclude the daily and weekly lags. This, of course, does not affect the overall fit of the model.

[21]Note that due to the linear nature of the HAR model and the fact that realized semicovariances sum exactly to the realized covariance, each coefficient in the fourth column is simply the sum of the corresponding coefficients in the first three columns.

Table 3: Semicovariance HAR Estimates

	$\widehat{P}_{jk,t}$	$\widehat{N}_{jk,t}$	$\widehat{M}_{jk,t}$	$\widehat{C}_{jk,t}$	
$\widehat{P}_{jk,t-1}$	0.038*	0.050*	-0.035*	0.052*	
$\widehat{P}_{jk,t-2:t-5}$	0.004	0.057	-0.002	0.059	
$\widehat{P}_{jk,t-6:t-22}$	-0.074	0.023	0.099	0.048	
$\widehat{N}_{jk,t-1}$	0.248**	0.192**	-0.096**	0.344**	
$\widehat{N}_{jk,t-2:t-5}$	0.312**	0.250**	-0.090*	0.472**	
$\widehat{N}_{jk,t-6:t-22}$	0.349**	0.206*	-0.021	0.534**	
$\widehat{M}_{jk,t-1}$	-0.075*	-0.072*	0.141**	-0.006	
$\widehat{M}_{jk,t-2:t-5}$	-0.044	-0.049	0.209**	0.116	
$\widehat{M}_{jk,t-6:t-22}$	0.028	-0.020	0.409**	0.417**	
$\widehat{C}_{jk,t-1}$					0.184**
$\widehat{C}_{jk,t-2:t-5}$					0.305**
$\widehat{C}_{jk,t-6:t-22}$					0.304**
R^2	0.397	0.376	0.354	0.313	0.284
R^2_{adj}	0.395	0.374	0.352	0.311	0.283

Note: The table reports the average parameter estimates for the vector HAR model in (4.1) averaged across 500 randomly selected pairs of stocks. The first three columns report results for the unrestricted models. The fourth column reports the estimates from a model that restricts the rows of $\Phi_{jk,Day}$, $\Phi_{jk,Week}$ and $\Phi_{jk,Month}$ to be the same, corresponding to a model for $\widehat{C}_{j,kt}$, whilst the final column reports the results of a standard HAR model on $\widehat{C}_{jk,t}$. ** and * signify that the estimates for that coefficient are significant at the 5% level for 75% and 50% of the randomly selected pairs of stocks, respectively.

reported in the last column: normalizing each of the explanatory variables by their sample means, the semicovariance-based HAR models effectively put a weight of 0.339 on lagged daily information, while the final column shows that a standard HAR model on average puts a weight of only 0.184 on the daily lag, implying a more muted reaction to new information. These differences are naturally associated with an improved fit of the semicovariance-based models, as shown by the R^2s in the bottom two rows. In the next section we investigate whether this improved in-sample fit is accompanied by a similar improvement in out-of-sample forecast performance for models that utilize the realized semicovariances.

4.2. Portfolio volatility forecasting

The realized variance of a portfolio obviously depends on the realized semicovariances of the assets included in the portfolio. In particular, utilizing the result that $\widehat{C} = \widehat{P} +$

$\widehat{N} + \widehat{M}$, the realized variance of a portfolio with portfolio weights w may be expressed as

$$\begin{aligned}\widehat{RV}^w &\equiv w^\top \widehat{C} w \\ &= w^\top \widehat{P} w + w^\top \widehat{N} w + w^\top \widehat{M} w \\ &\equiv \widehat{P}^w + \widehat{N}^w + \widehat{M}^w, \end{aligned} \tag{4.2}$$

where we use the superscript w to indicate the relevant (scalar-valued) portfolio quantities. These portfolio semicovariance measures are distinctly different from the portfolio semivariances of Barndorff-Nielsen, Kinnebrock, and Shephard (2010), which only depend on the high-frequency returns of the portfolio. The portfolio semicovariances, on the other hand, depend on the high-frequency returns for *all* of the individual assets included in the portfolio, and cannot be computed using only returns on the portfolio itself.

To explore whether portfolio semicovariances convey useful information beyond familiar realized variances and semivariances, we extend the HAR model of Corsi (2009) to allow the forecasts to depend on each of the realized portfolio semicovariance components. Accordingly, the one-day-ahead forecast for the portfolio return variance is constructed from

$$\begin{aligned} RV^w_{t+1|t} \;=\; & \phi_0 + \phi_{Day,P}\widehat{P}^w_t + \phi_{Week,P}\widehat{P}^w_{t-1:t-4} + \phi_{Month,P}\widehat{P}^w_{t-5:t-21} \\ & + \phi_{Day,N}\widehat{N}^w_t + \phi_{Week,N}\widehat{N}^w_{t-1:t-4} + \phi_{Month,N}\widehat{N}^w_{t-5:t-21} \\ & + \phi_{Day,M}\widehat{M}^w_t + \phi_{Week,M}\widehat{M}^w_{t-1:t-4} + \phi_{Month,M}\widehat{M}^w_{t-5:t-21}. \end{aligned} \tag{4.3}$$

We will refer to this model as the SemiCovariance HAR (SCHAR) model. The general SCHAR model in (4.3) is obviously quite richly parameterized. Hence, motivated by the results in Table 3, we also consider a restricted version, in which we only include the daily, weekly and monthly lags of $\widehat{N}^w$, and the monthly lag of $\widehat{M}^w$. We will refer to this specification as the restricted SCHAR model, or SCHAR-r for short.

If the parameters associated with the lagged realized semicovariance component all coincide (i.e., $\phi_{j,P} = \phi_{j,N} = \phi_{j,M} = \phi_j$ for $j \in \{Day, Week, Month\}$), the SCHAR model trivially reduces to the basic HAR model and the corresponding forecasting scheme

$$RV^w_{t+1|t} = \phi_0 + \phi_{Day}\widehat{RV}^w_t + \phi_{Week}\widehat{RV}^w_{t-1:t-4} + \phi_{Month}\widehat{RV}^w_{t-5:t-21}. \tag{4.4}$$

This simple and easy-to-implement model has arguably emerged as the benchmark for judging alternative realized volatility-based forecasting procedures in the literature.

In addition to this commonly-used benchmark, we also consider the forecasts from the Semivariance HAR (SHAR) model of Patton and Sheppard (2015). This model uses

the portfolio realized semivariances defined as

$$\widehat{PSV} = \sum_{i=1}^{[T/\Delta_n]} p\left(w^\top \Delta_i^n X\right)^2, \qquad \widehat{NSV} = \sum_{i=1}^{[T/\Delta_n]} n\left(w^\top \Delta_i^n X\right)^2,$$

to decompose the daily realized variance into positive and negative semivariances, resulting in the forecasting scheme:

$$RV^w_{t+1|t} = \phi_0 + \phi_{Day,+}\widehat{PSV}_t + \phi_{Day,-}\widehat{NSV}_t + \phi_{Week}\widehat{RV}^w_{t-1:t-4} + \phi_{Month}\widehat{RV}^w_{t-5:t-21}. \quad (4.5)$$

This model has been found to perform particularly well from the perspective of portfolio variance forecasting, performing on par with or better than the forecasts from other HAR-style models, and as such constitutes another particularly challenging benchmark.

We consider equally-weighted portfolios comprised of $d = 10$ ("small") and 100 ("large") stocks randomly selected from the full set of 749 individual stocks. In each case, we ensure that the selected stocks contain an overlap of at least 1,100 daily observations. We then construct rolling out-of-sample forecasts based on each of the different models, with model parameters re-estimated daily using the most recent 1,000 daily observations. We rely on the commonly used mean-square-error (MSE) and QLIKE loss functions to evaluate the performance of the forecasts vis-a-vis the actual portfolio realized variances $\widehat{RV}^w_{t+1}$.[22] Table 4 reports the resulting losses averaged across 500 randomly selected portfolios. In addition, for each of the 500 portfolios, we also compute the ratio of each model's average loss relative to the benchmark HAR model, and report the average of these ratios over all of the random samples.[23]

Consistent with Patton and Sheppard (2015), the SHAR-based forecasts that utilize the portfolio realized semivariances do result in fairly large relative gains vis-a-vis the benchmark HAR-based forecasts, especially for the large dimensional portfolios. The performance of the unrestricted SCHAR model, however, is mixed: it has smaller MSE loss than the HAR forecasts, but underperforms that same benchmark under QLIKE loss. This finding is hardly surprising. There is ample evidence in the forecasting literature emphasizing the importance of parsimony (see, e.g., Zellner (1992)), and the results in Table 3 clearly suggest that the unrestricted SCHAR model is "over-parameterized," and as such is likely to perform poorly in a forecasting context. Indeed, looking at the results for the restricted SCHAR-r model guided by the estimates in Table 3, we see that the forecasts from this model unambiguously outperform those from the other models; the 13.8% improvement in terms of predictive accuracy (measured by MSE) for the large

[22]The MSE and QLIKE loss functions may both be formally justified for the purpose of volatility model forecast evaluation based on the use of imperfect ex-post volatility proxies; see Patton (2011).

[23]The supplemental appendix reports corroborative evidence from a series of additional robustness checks.

Table 4: Performance Comparison for Portfolio Variance Forecasts

Model	MSE		QLIKE	
	Average	Ratio	Average	Ratio
Panel A. Small Portfolio Case (d =10)				
HAR	1.849	1.000	0.141	1.000
SHAR	1.671	0.966	0.139	0.986
SCHAR	1.643	0.955	0.210	1.318
SCHAR-r	1.567	0.908	0.139	0.979
Panel B. Large Portfolio Case (d =100)				
HAR	0.048	1.000	0.119	1.000
SHAR	0.045	0.935	0.115	0.957
SCHAR	0.045	0.976	0.236	1.495
SCHAR-r	0.041	0.862	0.111	0.925

Note: The table reports the loss for forecasting the portfolio variance for portfolios of size $d = 10$ and 100 for each of the different forecasting models. The reported numbers are based on 500 randomly selected portfolios. The Average column provides the average loss over time and all portfolios. The Ratio column gives the time-average ratio of losses across all sets of portfolios relative to the HAR model.

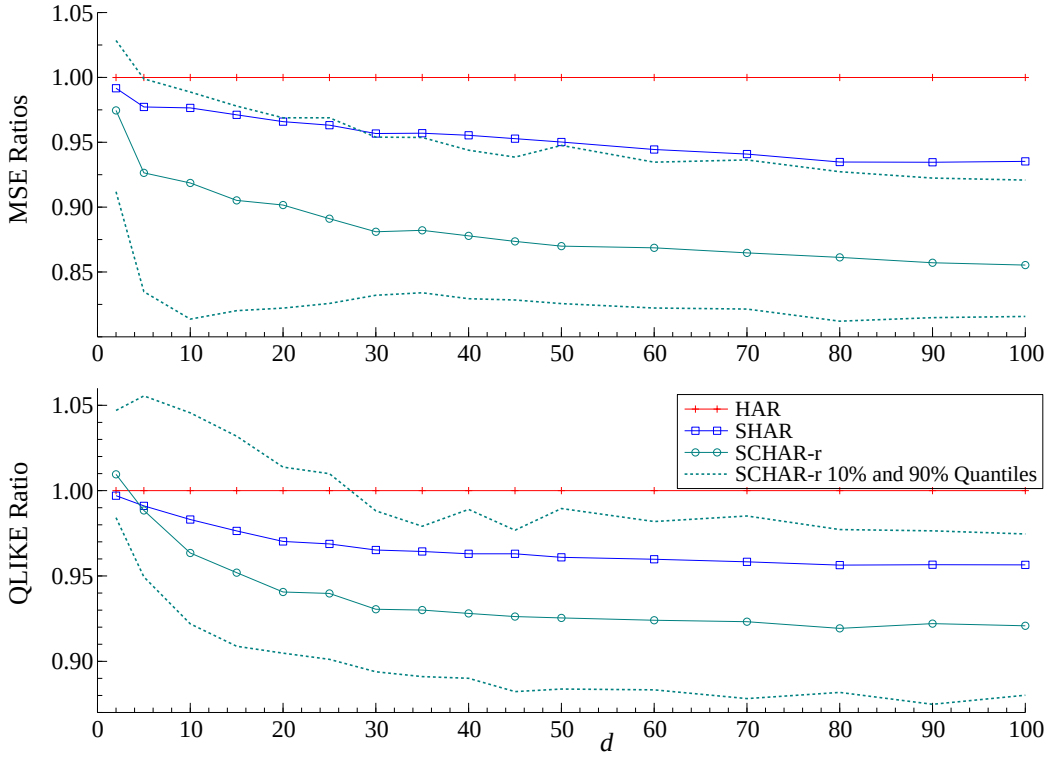
portfolio case is particularly impressive. As such, this clearly shows the benefit of utilizing the additional information inherent in the realized semicovariances, compared to both the HAR- and SHAR-based forecasts that only rely on the portfolio realized (semi)variances.

These gains in forecast accuracy are not unique to our specific choice of the two portfolio dimensions highlighted in Table 4. Figure 4 plots the median loss ratios for the HAR, SHAR and SCHAR-r models for all values of d ranging from 2 to 100, together with the 10% and 90% quantiles for the SCHAR-r forecasts computed across the 500 random portfolio-formations. As the figure shows, the median loss ratios for the SCHAR-r model are systematically below those of the other two models, the only exception being the QLIKE loss for $d = 2$. Also, the gains from using the information in the realized semicovariances accrue relatively quickly as the number of stocks in the portfolio increases, and for the QLIKE loss appear to reach somewhat of a plateau for $d \approx 40$ stocks.

The gains in forecast accuracy obtained from using the realized semicovariance measures are not restricted to the one-day forecast horizon analyzed in Table 4 and Figure 4 either. To illustrate, we report in Table 5 the results for 5-day and 22-day ahead forecasts, corresponding to one week and one month, respectively. For simplicity we rely on "direct" forecasts constructed by simply replacing the left-hand-side variable in the estimation of the different models with the realized portfolio variance over the relevant horizon.[24] As

[24]The construction of "iterated" forecasts for the SHAR and SCHAR models would necessitate ad-

Figure 4: Median Loss Ratios



Note: The graph plots the median loss ratios as a function of the number of stocks in the portfolio, d. The ratio is calculated as the average loss of the models divided by the average loss of of the standard HAR, based on 500 random samples of d-stock portfolios.

the table shows, if anything the forecast improvements obtained by the SCHAR-r model vis-a-vis the standard HAR model appear even larger over longer forecast horizons; as a case in point the 18.4% relative improvement (measured by MSE) for the large portfolio at the monthly horizon is especially impressive.

To help further understand the source of these improvements it is instructive to represent the SHAR and SCHAR models as HAR models with time-varying parameters. To illustrate the idea, consider the forecasts from a SCHAR-type model based only on the lagged daily realized semicovariances,

$$\begin{aligned} RV^{w}_{t+1|t} &= \phi_0 + \phi_{Day,P}\widehat{P}^{w}_{t} + \phi_{Day,N}\widehat{N}^{w}_{t} + \phi_{Day,M}\widehat{M}^{w}_{t} \\ &= \phi_0 + \left(\phi_{Day,P}\frac{\widehat{P}^{w}_{t}}{\widehat{RV}^{w}_{t}} + \phi_{Day,N}\frac{\widehat{N}^{w}_{t}}{\widehat{RV}^{w}_{t}} + \phi_{Day,M}\frac{\widehat{M}^{w}_{t}}{\widehat{RV}^{w}_{t}} \right) \widehat{RV}^{w}_{t} \\ &\equiv \phi_0 + \phi_{Day,t}\widehat{RV}^{w}_{t} . \end{aligned}$$

ditional modeling assumptions about the realized semi(co)variances; for further discussion of "direct" versus "iterated" volatility forecasts see, e.g., Bollerslev, Hood, Huss, and Pedersen (2018) and the references therein.

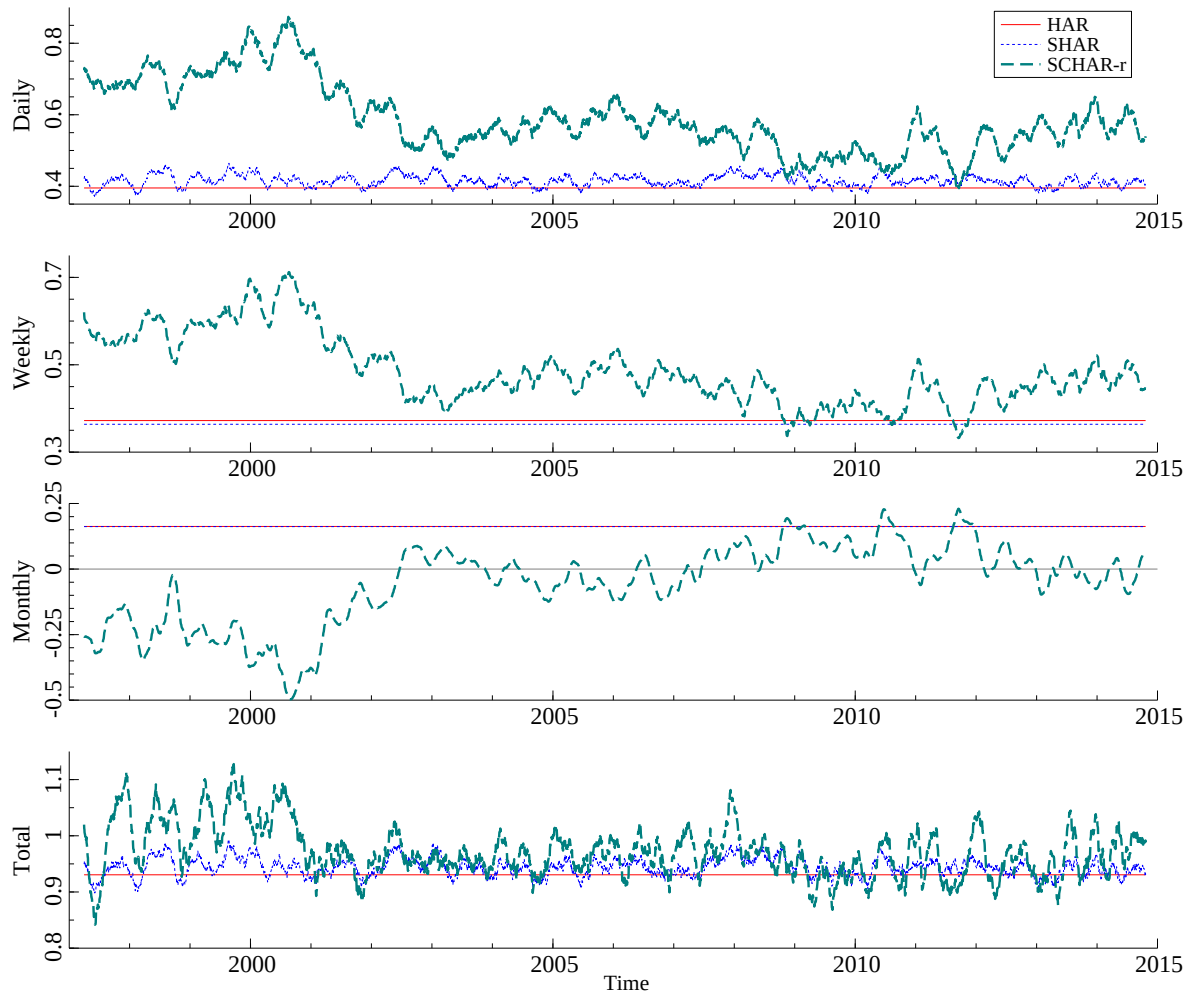
Table 5: Long-Horizon Results

Model	MSE		QLIKE	
	Average	Ratio	Average	Ratio
Panel A. Weekly Forecasts, Small Portfolio Case (d =10)				
HAR	1.314	1.000	0.086	1.000
SHAR	1.222	0.958	0.083	0.967
SCHAR	1.075	0.900	0.151	1.596
SCHAR-r	1.062	0.874	0.085	0.974
Panel B. Weekly Forecasts, Large Portfolio Case (d =100)				
HAR	0.028	1.000	0.070	1.000
SHAR	0.026	0.917	0.065	0.936
SCHAR	0.024	0.881	0.068	0.981
SCHAR-r	0.023	0.823	0.062	0.884
Panel C. Monthly Forecasts, Small Portfolio Case (d =10)				
HAR	1.407	1.000	0.108	1.000
SHAR	1.372	0.982	0.106	0.982
SCHAR	1.121	0.906	0.160	1.339
SCHAR-r	1.135	0.867	0.111	1.000
Panel D. Monthly Forecasts, Large Portfolio Case (d =100)				
HAR	0.031	1.000	0.071	1.000
SHAR	0.030	0.954	0.068	0.956
SCHAR	0.028	0.926	0.060	0.859
SCHAR-r	0.025	0.816	0.058	0.837

Note: The table reports the loss for forecasting the portfolio variance for portfolios of size $d = 10$ and 100 for each of the different forecasting models, for weekly and monthly horizons. The reported numbers are based on 500 randomly selected portfolios. The Average column provides the average loss over time and all portfolios. The Ratio column gives the time-average ratio of losses across all sets of portfolios relative to the HAR model.

As these equations show, even though the $\phi_{Day,P}$, $\phi_{Day,N}$ and $\phi_{Day,M}$ parameters used in the formulation of the model are all constant, the model may alternatively be interpreted as a first-order autoregression for $\widehat{RV}^{w}_{t+1}$ with a time-varying autoregressive parameter $\phi_{Day,t}$. This idea obviously generalizes to the SHAR and the more elaborate SCHAR-r forecasting models used in our empirical analysis, in which the parameters associated with the weekly and monthly lags in the implied HAR-type representations would be time-varying as well.

Figure 5: Implied HAR Parameters



Note: The figure plots the implied HAR-type parameters for predicting the variance of a ten-stock portfolio, averaged across 500 randomly selected portfolios. The models are estimated over the full sample period.

In order to explicitly quantify these effects, Figure 5 plots the daily, weekly and monthly time-varying HAR parameters implied by the SHAR and SCHAR-r models for $d = 10$, averaged across 500 randomly selected ten-stock portfolios.[25] In addition to the implied daily, weekly and monthly parameters, the last panel reports their sum as a measure of the overall persistence of the different models.

The daily, weekly, and monthly parameters for the HAR model are by definition all constant, with an average implied persistence of around 0.94. By comparison, the implied

[25]In contrast to the out-of-sample forecast results in Table 4, which are based on a rolling estimation scheme, Figure 5 plots the implied parameter estimates obtained over the full sample period. Requiring observations to be available over the full sample reduces the number of stocks to 121. To avoid "contaminating" the results by a few influential outliers, we exclude any portfolios for which the maximum $\widehat{P}^w/\widehat{RV}^w$, $\widehat{N}^w/\widehat{RV}^w$ and $\widehat{M}^w/\widehat{RV}^w$ ratios exceed ten, and further smooth the implied parameters using a centered 50-day moving average.

daily parameter estimates for the SHAR model vary slightly above the constant daily HAR parameter, while the constant weekly parameter for the SHAR model is slightly below that of the HAR model. As such, the overall persistence of the SHAR-based forecasts are generally very close to that of the standard HAR-based forecasts, which explains why the two models perform fairly similar over the longer monthly forecast horizon analyzed in Table 5.

By contrast, the implied time-varying daily and weekly HAR parameters for the SCHAR-r model both far exceed those of the standard HAR model, especially over the earlier part of the sample. On the other hand, the implied time-varying monthly parameters for the SCHAR-r model are typically less than the monthly parameters for the standard HAR and SHAR models. Meanwhile, the sum of the three implied parameters for the SCHAR-r model are typically greater than for the other two models.[26] Thus, not only do the superior SCHAR-based forecasts respond more quickly to new information, the forecasts are typically also more persistent and slower to mean-revert than the benchmark HAR- and SHAR-based forecasts. This explains why the SCHAR-r model performs well not just over the short one-day forecast horizon, but also over the longer monthly forecast horizons. It also highlights the usefulness of the richer information residing in the new realized semicovariance measures for volatility forecasting purposes more generally.

5. Conclusion

We propose a new decomposition of the realized covariance matrix based on the signs of the underlying high-frequency returns into separate positive, negative and mixed-sign realized semicovariance components. Under a standard infill asymptotic setting for continuous-time Itô semimartingales, we derive the first- and second-order asymptotic properties of these new realized semicovariance measures. The asymptotic theory, taking the form of a non-central limit theorem, reveals the differential information carried by each of the realized semicovariance components, related to stochastic correlation, signed co-jumps, and notions of co-drifting and dynamic leverage effects. Using high-frequency data for a large cross-section of U.S. equities, we demonstrate how the asymptotic theory may be used in understanding key features of the realized semicovariance components associated with different economic events. Consistent with the theory, we further document distinctly different dynamic dependencies in the different realized semicovariance components. These differences in turn translate into markedly superior forecast performance for models that utilize the realized semicovariance measures.

[26] Even though the sum of the autoregressive coefficients for the SCHAR-r model occasionally exceeds unity, this does not necessarily imply non-stationarity, as the temporal variation in the realized semicovariance measures may induce stationarity; see Conley, Hansen, Luttmer, and Scheinkman (1997) and Nielsen and Rahbek (2014) for a discussion of volatility-induced stationarity.

References

Aït-Sahalia, Y., and J. Jacod (2014): *High-Frequency Financial Econometrics*. Princeton University Press, Princeton.

Aït-Sahalia, Y., and D. Xiu (2016): "Increased Correlation Among Asset Classes: Are Volatility or Jumps to Blame, or Both?," *Journal of Econometrics*, 194(2), 205–219.

Amaya, D., P. Christoffersen, K. Jacobs, and A. Vasquez (2015): "Does Realized Skewness Predict the Cross-Section of Equity Returns?," *Journal of Financial Economics*, 118(1), 135–167.

Andersen, T. G., T. Bollerlsev, F. X. Diebold, and P. Labys (2003): "Modeling and Forecasting Realized Volatility," *Econometrica*, 71(2), 579–625.

Andersen, T. G., T. Bollerslev, and F. X. Diebold (2007): "Roughing It Up: Including Jump Components in the Measurement, Modeling, and Forecasting of Return Volatility," *Review of Economics and Statistics*, 89(4), 701–720.

Ang, A., and J. Chen (2002): "Asymmetric Correlations of Equity Portfolios," *Journal of Financial Economics*, 63(3), 443–494.

Back, K. (1992): "Insider Trading in Continuous Time," *Review of Financial Studies*, 5(3), 387–409.

Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard (2011): "Multivariate Realised Kernels: Consistent Positive Semi-Definite Estimators of the Covariation of Equity Prices with Noise and Non-Synchronous Trading," *Journal of Econometrics*, 162(2), 149–169.

Barndorff-Nielsen, O. E., S. Kinnebrock, and N. Shephard (2010): "Measuring Downside Risk: Realised Semivariance," *In "Volatility and Time Series Econometrics: Essays in Honor of Robert F. Engle" (Edited by T. Bollerslev, J. Russell and M. Watson), Oxford University Press*, pp. 117–136.

Barndorff-Nielsen, O. E., and N. Shephard (2004): "Econometric Analysis of Realized Covariation: High-Frequency based Covariance, Regression, and Correlation in Financial Economics," *Econometrica*, 72(3), 885–925.

Bauwens, L., S. Laurent, and J. V. K. Rombouts (2006): "Multivariate GARCH Models: A Survey," *Journal of Applied Econometrics*, 21(1), 79–109.

Bollerslev, T., B. Hood, J. Huss, and L. H. Pedersen (2018): "Risk Everywhere: Modeling and Managing Volatility," *Review of Financial Studies*, 31(7), 2729–2773.

Bollerslev, T., T. Law, and G. Tauchen (2008): "Risk, Jumps and Diversification," *Journal of Econometrics*, 144(1), 234–256.

Bollerslev, T., J. Li, and Y. Xue (2018): "Volume, Volatility, and Public News Announcements," *Review of Economic Studies*, 85(4), 2005–2041.

Bollerslev, T., and V. Todorov (2011a): "Estimation of Jump Tails," *Econometrica*, 79(6), 1727–1783.

rica, 79(6), 1727–1783.

——— (2011b): "Tails, Fears, and Risk Premia," *Journal of Finance*, 66(6), 2165–2211.

Cappiello, L., R. Engle, and K. Sheppard (2006): "Asymmetric Dynamics in the Correlations of Global Equity and Bond Returns," *Journal of Financial Econometrics*, 4(4), 537–572.

Conley, T., L. Hansen, E. Luttmer, and J. Scheinkman (1997): "Short-Term Interest Rates as Subordinated Diffusions," *Review of Financial Studies*, 10(3), 525–577.

Corsi, F. (2009): "A Simple Approximate Long-Memory Model of Realized Volatility," *Journal of Financial Econometrics*, 7(2), 174–196.

Das, S. R., and R. Uppal (2004): "Systemic Risk and International Portfolio Choice," *Journal of Finance*, 59(6), 2809–2834.

Elton, E. J., and M. J. Gruber (1973): "Estimating the Dependence Structure of Share Prices – Implications for Portfolio Selection," *Journal of Finance*, 28(5), 1203–1232.

Fishburn, P. C. (1977): "Mean-Risk Analysis with Risk Associated with Below-Target Returns," *American Economic Review*, 67(2), 116–126.

Hansen, P. R., Z. Huang, and H. H. Shek (2012): "Realized GARCH: A Joint Model for Returns and Realized Measures of Volatility," *Journal of Applied Econometrics*, 27(6), 877–906.

Hansen, P. R., and A. Lunde (2006): "Realized Variance and Market Microstructure Noise," *Journal of Business and Economic Statistics*, 24(2), 127–161.

Hogan, W. W., and J. M. Warren (1972): "Computation of the Efficient Boundary in the E-S Portfolio Selection Model," *Journal of Financial and Quantitative Analysis*, 7(4), 1881–1896.

——— (1974): "Toward the Development of an Equilibrium Capital-Market Model Based on Semivariance," *Journal of Financial and Quantitative Analysis*, 9(1), 1–11.

Hong, Y., J. Tu, and G. Zhou (2007): "Asymmetries in Stock Returns: Statistical Tests and Economic Interpretation," *Review of Financial Studies*, 20(5), 1547–1581.

Jacod, J., Y. Li, and X. Zheng (2017): "Statistical Properties of Microstructure Noise," *Econometrica*, 85(4), 1133–1174.

Jacod, J., and P. Protter (2012): *Discretization of Processes*. Springer Verlag.

Jacod, J., and V. Todorov (2009): "Testing for Common Arrivals of Jumps for Discretely Observed Multidimensional Processes," *Annals of Statistics*, 37(4), 1792–1838.

Kendall, M. G. (1953): "The Analysis of Economic Time-Series – Part I: Prices," *Journal of the Royal Statistical Society, Series A*, 116(1), 11–34.

Kroner, K. F., and V. K. Ng (1998): "Modeling Asymmetric Comovements of Asset Returns," *Review of Financial Studies*, 11(4), 817–844.

Kyle, A. S. (1985): "Continuous Auctions and Insider Trading," *Econometrica*, pp. 1315–1335.

Lahaye, J., S. Laurent, and C. J. Neely (2011): "Jumps, Cojumps and Macro Announcements," *Journal of Applied Econometrics*, 26(6), 893–921.

Lee, S. S. (2012): "Jumps and Information Flow in Financial Markets," *Review of Financial Studies*, 25(2), 439–479.

Lee, S. S., and P. A. Mykland (2008): "Jumps in Financial Markets: A New Nonparametric Test and Jump Dynamics," *Review of Financial Studies*, 21(6), 2535–2563.

Li, J., V. Todorov, and G. Tauchen (2017a): "Adaptive Estimation of Continuous-Time Regression Models Using High-Frequency Data," *Journal of Econometrics*, 200(1), 36–47.

——— (2017b): "Jump Regressions," *Econometrica*, 85(1), 173–195.

Li, J., V. Todorov, G. Tauchen, and R. Chen (2017): "Mixed-Scale Jump Regressions with Bootstrap Inference," *Journal of Econometrics*, 201(2), 417–432.

Li, J., and D. Xiu (2016): "Generalized Method of Integrated Moments for High-Frequency Data," *Econometrica*, 84(4), 1613–1633.

Longin, F., and B. Solnik (2001): "Extreme Correlation of International Equity Markets," *Journal of Finance*, 56(2), 649–676.

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

——— (2009): "Non-Parametric Threshold Estimation for Models with Stochastic Diffusion Coefficient and Jumps," *Scandinavian Journal of Statistics*, 36(2), 270–296.

Mancini, C., and F. Gobbi (2012): "Identifying the Brownian Covariation from the Co-Jumps given Discrete Observations," *Econometric Theory*, 28(2), 249–273.

Mao, J. C. T. (1970): "Survey of Capital Budgeting: Theory and Practice," *Journal of Finance*, 25(2), 349–360.

Markowitz, H. M. (1959): *Portfolio Selection*. Wiley.

Neuberger, A. (2012): "Realized Skewness," *Review of Financial Studies*, 25(11), 3423–3455.

Nielsen, H. B., and A. Rahbek (2014): "Unit Root Vector Autoregression with Volatility Induced Stationarity," *Journal of Empirical Finance*, 29, 144–167.

Noureldin, D., N. Shephard, and K. Sheppard (2012): "Multivariate High-Frequency-Based Volatility (HEAVY) Models," *Journal of Applied Econometrics*, 27(6), 907–933.

Patton, A. J. (2004): "On the Out-of-Sample Importance of Skewness and Asymmetric Dependence for Asset Allocation," *Journal of Financial Econometrics*, 2(1), 130–168.

——— (2011): "Volatility Forecast Comparison using Imperfect Volatility Proxies," *Journal of Econometrics*, 160(1), 246–256.

Patton, A. J., and K. Sheppard (2015): "Good Volatility, Bad Volatility: Signed

Jumps and the Persistence of Volatility," *Review of Economics and Statistics*, 97(3), 683–697.

Poon, S.-H., M. Rockinger, and J. Tawn (2004): "Extreme Value Dependence in Financial Markets: Diagnostics, Models, and Financial Implications.," *Review of Financial Studies*, 17(2), 581–610.

Tjøsthem, D., and K. O. Hufthammer (2013): "Local Gaussian Correlation: A New Measure of Dependence," *Journal of Econometrics*, 172(1), 33–48.

Zellner, A. (1992): "Presidential Address: Statistics, Science and Public Policy," *Journal of the American Statistical Association*, 87(417), 1–6.

Appendix A. Regularity conditions

In this appendix, we describe the regularity conditions that are needed for the asymptotic theory presented in the main text.

Assumption 1 *The process X is an Itô semimartingale defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ of the form (2.1) with*

$$J_t = \int_0^t \int_{\mathbb{R}} \delta(s,u)\, \mu(ds,du),$$

where the process b is locally bounded, the process σ is càdlàg and takes value in $\mathbb{R}^{d\otimes d}$, δ is a predictable function and μ is a Poisson random measure defined on $\mathbb{R}_+ \times \mathbb{R}$ with compensator $\nu(dt,du) = dt \otimes \lambda(du)$ for some finite measure λ on $\mathbb{R}$.

Assumption 2 *We have Assumption 1. Moreover, the process σ has the form (2.2) such that (i) σ_t is non-singular almost surely for all t; (ii) $\tilde{b}$ is locally bounded; (iii) $\tilde{\sigma}$ is $d \times d \times d$ càdlàg adapted process; (iv) the process $\widetilde{M}$ is a local martingale that is orthogonal to W with $\|\Delta \widetilde{M}_t\| \leq \underline{\sigma}$ for some constant $\underline{\sigma} > 0$ and its predictable quadratic covariation process has the form $\langle \widetilde{M}, \widetilde{M} \rangle_t = \int_0^t \tilde{q}_s ds$ for some locally bounded process $\tilde{q}$; (v) the compensator of the pure-jump process $\sum_{s \leq t} \Delta\sigma_s 1_{\{\|\Delta\sigma_s\| > \underline{\sigma}\}}$ has the form $\int_0^t q_s ds$ for some locally bounded process q.*

Overall, these assumptions are fairly mild and quite standard in the analysis of high-frequency data. In particular, they allow for price and volatility jumps and co-jumps, as well as the so-called leverage effect. The only notable restriction is the finite activity of the price jumps. As in Li, Todorov, and Tauchen (2017b), we purposely impose this condition because, in the current paper, the empirical interest vis-a-vis jumps mainly concerns "large" market-wide co-jumps, which occur relatively infrequently (and thus aligned with the finite-activity condition). Relaxing this condition would greatly complicate our technical exposition, without leading to any change in the actual numerical implementation.

Appendix B. Proofs

Appendix B.1. Notation and preliminary results

We begin by defining some notation. Recall that for $j \in \{1, \ldots, d\}$, $\mathcal{T}_j = \{\tau : \Delta X_{j,\tau} \neq 0\}$ collects the jump times of asset j, and

$$\mathcal{T}_{j+} \equiv \{\tau \in \mathcal{T} : \Delta X_{j,\tau} > 0\}, \quad \mathcal{T}_{j-} \equiv \{\tau \in \mathcal{T} : \Delta X_{j,\tau} < 0\}.$$

That is, $\mathcal{T}_{j+}$ and $\mathcal{T}_{j-}$ collect the times at which asset j has positive and negative jumps, respectively. For each jump time τ, we denote by $i(\tau)$ the unique random integer i such

that $(i-1)\Delta_n < \tau \leq i\Delta_n$. We then set

$$\mathcal{I}_j \equiv \{i(\tau) : \tau \in \mathcal{T}_j\} \text{ and } \mathcal{I}_{j\pm} \equiv \{i(\tau) : \tau \in \mathcal{T}_{j\pm}\}. \tag{B.1}$$

Recall that $p(x) \equiv \max\{x, 0\}$ and $n(x) \equiv \min\{x, 0\}$. To simplify notation, we denote, for $x, y \in \mathbb{R}$,

$$f(x,y) \equiv p(x)\,p(y), \quad g(x,y) \equiv n(x)\,n(y), \quad m(x,y) = p(x)\,n(y).$$

Note that f and g are both continuously differentiable except on $\{(x,y) : x = 0 \text{ or } y = 0\}$, with gradients

$$\partial f(x,y) = \begin{pmatrix} 1_{\{x\geq 0\}} \max\{y, 0\} \\ \max\{x, 0\}\, 1_{\{y\geq 0\}} \end{pmatrix}, \quad \partial g(x,y) = \begin{pmatrix} 1_{\{x\leq 0\}} \min\{y, 0\} \\ \min\{x, 0\}\, 1_{\{y\leq 0\}} \end{pmatrix}.$$

For generic functions h, h_1 and h_2 defined on $\mathbb{R}^d$, and an invertible $d \times d$ matrix a such that $c = aa^\top$, we define the following quantities:

$$\begin{cases} R_c(h) \equiv E_U[h(aU)], \quad \gamma_c(h) \equiv E_U[h(aU)(aU)], \\ \hat{\gamma}_a(h) \equiv E_U[h(aU)U], \\ H_j \equiv h_j(aU) - R_c(h) - \left(c^{-1}\gamma_c(h_j)\right)^\top aU, \quad j = 1, 2, \\ \bar{\gamma}_c(h_1, h_2) \equiv Cov_U(H_1, H_2), \\ \Gamma_c(h_1, h_2) \equiv Cov_U\left(\left(c^{-1}\gamma_c(h_1)\right)^\top aU, \left(c^{-1}\gamma_c(h_2)\right)^\top aU\right), \\ \overline{\Gamma}_c(h_1, h_2) \equiv Cov_U(h_1(aU), h_2(aU)), \end{cases} \tag{B.2}$$

where U is a generic d-dimensional standard normal variable, and E_U and Cov_U are the expectation and covariance operators with respect to U, respectively. We note that, except for $\hat{\gamma}_a(h)$, the expected values in (B.2) depend on a only through $c = aa^\top$, which explains our notation $R_c(\cdot)$, $\gamma_c(\cdot)$, etc.

We need to establish a few identities among these functionals. Observe that the variable H_j is the residual obtained from projecting the demeaned variable $h_j(aU) - E_U[h_j(aU)]$ onto aU, with $c^{-1}\gamma_c(h_j)$ the corresponding projection coefficient. Since a is nonsingular, this residual may alternatively be written as

$$H_j = h_j(aU) - E_U[h_j(aU)] - \hat{\gamma}_a(h_j)^\top U.$$

Hence, the functional $\bar{\gamma}_c(h_1, h_2)$ can be rewritten as

$$\begin{aligned} \bar{\gamma}_c(h_1, h_2) = E_U\left[\left(h_1(aU) - \hat{\gamma}_a(h_1)^\top U\right)\left(h_2(aU) - \hat{\gamma}_a(h_2)^\top U\right)\right] \\ - E_U[h_1(aU)]\, E_U[h_2(aU)]. \end{aligned} \tag{B.3}$$

We further note that $\overline{\Gamma}_c(h_1, h_2)$ computes the covariance of $h_1(aU) - E_U[h_1(aU)]$ and $h_2(aU) - E_U[h_2(aU)]$, while $\Gamma_c(h_1, h_2)$ computes the covariance of their projections. By a decomposition of covariance, we then deduce

$$\bar{\gamma}_c(h_1, h_2) = \overline{\Gamma}_c(h_1, h_2) - \Gamma_c(h_1, h_2). \tag{B.4}$$

In addition,

$$\Gamma_c(h_1, h_2) = \gamma_c(h_1)^\top c^{-1} \gamma_c(h_2). \tag{B.5}$$

Our analysis for the semicovariances relies on some explicit calculations of the expectations in (B.2). Lemma 1, below, provides the details for c taking the form

$$c = \begin{pmatrix} v_1^2 & \rho v_1 v_2 \\ \rho v_1 v_2 & v_2^2 \end{pmatrix}. \tag{B.6}$$

The proof is done by direct integration, and is omitted for brevity.

Lemma 1 *The following statements hold when the matrix c has the form (B.6):*
(a) $R_c(f) = R_c(g) = v_1 v_2 \psi(\rho)$ *and* $R_c(m) = -v_1 v_2 \psi(-\rho)$*;*
(b) $R_c(\partial f) = -R_c(\partial g) = \left(2\sqrt{2\pi}\right)^{-1}(1+\rho)(v_2, v_1)^\top$*;*
(c) $\gamma_c(f) = -\gamma_c(g) = \left(2\sqrt{2\pi}\right)^{-1}(1+\rho)^2 v_1 v_2 (v_1, v_2)^\top$*;*
(d) $\overline{\Gamma}_c(f, f) = \overline{\Gamma}_c(g, g) = v_1^2 v_2^2(\Psi(\rho) - \psi(\rho)^2)$ *and* $\overline{\Gamma}_c(f, g) = -v_1^2 v_2^2 \psi(\rho)^2$.

Appendix B.2. Proofs

Proof of Theorem 1. Let X' denote the continuous part of X, that is, $X'_t \equiv \int_0^t b_s ds + \int_0^t \sigma_s dW_s$. We define $\widehat{P}'$ in the same way as $\widehat{P}$, but with X replaced by X'. It follows that

$$\widehat{P}_{jk} = \widehat{P}'_{jk} + \sum_{i \in \mathcal{I}_j \cup \mathcal{I}_k} p(\Delta_i^n X_j)\, p(\Delta_i^n X_k) - \sum_{i \in \mathcal{I}_j \cup \mathcal{I}_k} p\left(\Delta_i^n X'_j\right) p\left(\Delta_i^n X'_k\right).$$

By Theorem 3.4.1(b) in Jacod and Protter (2012) and Lemma 1(a),

$$\widehat{P}'_{jk} \xrightarrow{\mathbb{P}} \int_0^T v_{j,s} v_{k,s} \psi(\rho_{jk,s}) ds. \tag{B.7}$$

We also note that

$$\sum_{i \in \mathcal{I}_j \cup \mathcal{I}_k} p(\Delta_i^n X_j)\, p(\Delta_i^n X_k) = \sum_{\tau \in \mathcal{T}_j \cup \mathcal{T}_k} p\left(\Delta_{i(\tau)}^n X_j\right) p\left(\Delta_{i(\tau)}^n X_k\right),$$

and $\Delta_{i(\tau)}^n X \to \Delta X_\tau$ pathwise. Hence,

$$\sum_{i \in \mathcal{I}_j \cup \mathcal{I}_k} p(\Delta_i^n X_j)\, p(\Delta_i^n X_k) \xrightarrow{\mathbb{P}} P_{jk}^\dagger. \tag{B.8}$$

Finally, by a standard estimate for continuous Itô semimartingales and the fact that $\mathcal{I}_j \cup \mathcal{I}_k$ is almost surely finite, we deduce that $\sum_{i \in \mathcal{I}_j \cup \mathcal{I}_k} p\left(\Delta_i^n X_j'\right) p\left(\Delta_i^n X_k'\right) = O_p(\Delta_n)$. This estimate, combined with (B.7) and (B.8), implies the asserted convergence $\widehat{P}_{jk} \xrightarrow{\mathbb{P}} P_{jk}$. The proof for $\widehat{N}$ and $\widehat{M}$ follows essentially the same argument. *Q.E.D.*

Proof of Theorem 2. Step 1. We begin by outlining the basic steps of the proof. First, we decompose

$$\Delta_n^{-1/2}\left(\widehat{P}_{12} - P_{12}\right) = \sum_{j=1}^{5} \widetilde{P}^{(j)}, \quad \Delta_n^{-1/2}\left(\widehat{N}_{12} - N_{12}\right) = \sum_{j=1}^{5} \widetilde{N}^{(j)},$$

where

$$\begin{cases} \widetilde{P}^{(1)} \equiv \Delta_n^{-1/2}\left(\displaystyle\sum_{i \in \mathcal{I}_{1+} \cap \mathcal{I}_{2+}} p\left(\Delta_i^n X_1\right) p\left(\Delta_i^n X_2\right) - P_{12}^{\dagger}\right), \\ \widetilde{P}^{(2)} \equiv \Delta_n^{-1/2} \displaystyle\sum_{i \in \mathcal{I}_{1+} \setminus \mathcal{I}_2} p\left(\Delta_i^n X_1\right) p\left(\Delta_i^n X_2\right), \\ \widetilde{P}^{(3)} \equiv \Delta_n^{-1/2} \displaystyle\sum_{i \in \mathcal{I}_{2+} \setminus \mathcal{I}_1} p\left(\Delta_i^n X_1\right) p\left(\Delta_i^n X_2\right), \\ \widetilde{P}^{(4)} \equiv \Delta_n^{-1/2} \displaystyle\sum_{i \in \mathcal{I}_{1-} \cup \mathcal{I}_{2-}} p\left(\Delta_i^n X_1\right) p\left(\Delta_i^n X_2\right), \\ \widetilde{P}^{(5)} \equiv \Delta_n^{-1/2}\left(\displaystyle\sum_{i \notin \mathcal{I}_1 \cup \mathcal{I}_2} p\left(\Delta_i^n X_1\right) p\left(\Delta_i^n X_2\right) - P_{12}^{\star}\right), \end{cases}$$

and

$$\begin{cases} \widetilde{N}^{(1)} \equiv \Delta_n^{-1/2}\left(\displaystyle\sum_{i \in \mathcal{I}_{1-} \cap \mathcal{I}_{2-}} n\left(\Delta_i^n X_1\right) n\left(\Delta_i^n X_2\right) - N_{12}^{\dagger}\right), \\ \widetilde{N}^{(2)} \equiv \Delta_n^{-1/2} \displaystyle\sum_{i \in \mathcal{I}_{1-} \setminus \mathcal{I}_2} n\left(\Delta_i^n X_1\right) n\left(\Delta_i^n X_2\right), \\ \widetilde{N}^{(3)} \equiv \Delta_n^{-1/2} \displaystyle\sum_{i \in \mathcal{I}_{2-} \setminus \mathcal{I}_1} n\left(\Delta_i^n X_1\right) n\left(\Delta_i^n X_2\right), \\ \widetilde{N}^{(4)} \equiv \Delta_n^{-1/2} \displaystyle\sum_{i \in \mathcal{I}_{1+} \cup \mathcal{I}_{2+}} n\left(\Delta_i^n X_1\right) n\left(\Delta_i^n X_2\right), \\ \widetilde{N}^{(5)} \equiv \Delta_n^{-1/2}\left(\displaystyle\sum_{i \notin \mathcal{I}_1 \cup \mathcal{I}_2} n\left(\Delta_i^n X_1\right) n\left(\Delta_i^n X_2\right) - N_{12}^{\star}\right). \end{cases}$$

To prove the assertion of Theorem 2, it suffices to show the following joint stable convergence in law

$$\left(\sum_{j=1}^{4} \widetilde{P}^{(j)}, \sum_{j=1}^{4} \widetilde{N}^{(j)}\right)^{\top} \xrightarrow{\mathcal{L}\text{-}s} \xi, \tag{B.9}$$

$$\left(\widetilde{P}^{(5)}, \widetilde{N}^{(5)}\right)^{\top} \xrightarrow{\mathcal{L}\text{-}s} B^{(1)} + B^{(2)} + \zeta + \tilde{\zeta}. \tag{B.10}$$

In steps 2 and 3, below, we prove these claims in turn. We note that, by a standard argument, it is easy to show that the convergences in (B.9) and (B.10) hold jointly with $\mathcal{F}$-conditionally independent limits (as they involve non-overlapping Brownian increments). Therefore, the remaining task is to show (B.9) and (B.10).

Step 2. This step proves the claim in (B.9). Let $\mathcal{T} \equiv \mathcal{T}_1 \cup \mathcal{T}_2$ collect all jump times of the bivariate process X. We set, for each $\tau \in \mathcal{T}$,

$$\hat{\eta}_\tau = (\hat{\eta}_{1,\tau}, \hat{\eta}_{2,\tau})^\top \equiv \Delta_n^{-1/2} \left(\Delta_{i(\tau)}^n X - \Delta X_\tau \right).$$

By Proposition 4.4.10 in Jacod and Protter (2012),

$$(\hat{\eta}_\tau)_{\tau \in \mathcal{T}} \xrightarrow{\mathcal{L}\text{-}s} (\tilde{\eta}_\tau)_{\tau \in \mathcal{T}}, \tag{B.11}$$

where $\tilde{\eta}_\tau = (\tilde{\eta}_{1,\tau}, \tilde{\eta}_{2,\tau})^\top \equiv \sqrt{\kappa_\tau}\tilde{\xi}_{\tau-} + \sqrt{1-\kappa_\tau}\tilde{\xi}_{\tau+}$ (recall the definitions in Section 2.2).

From (B.11), it is easy to derive the following representations uniformly for all $\tau \in \mathcal{T}$ (note that $\mathcal{T}$ is finite almost surely): for $j \in \{1, 2\}$,

$$p\left(\Delta_{i(\tau)}^n X_j\right) = \begin{cases} \Delta X_{j,\tau} + \Delta_n^{1/2} \hat{\eta}_{j,\tau} + o_p(\Delta_n^{1/2}), & \text{if } \tau \in \mathcal{T}_{j+}, \\ \Delta_n^{1/2} p\left(\hat{\eta}_{j,\tau}\right) + o_p(\Delta_n^{1/2}), & \text{if } \tau \in \mathcal{T} \setminus \mathcal{T}_j, \\ o_p(\Delta_n^{1/2}), & \text{if } \tau \in \mathcal{T}_{j-}, \end{cases} \tag{B.12}$$

and

$$n\left(\Delta_{i(\tau)}^n X_j\right) = \begin{cases} o_p(\Delta_n^{1/2}), & \text{if } \tau \in \mathcal{T}_{j+}, \\ \Delta_n^{1/2} n\left(\hat{\eta}_{j,\tau}\right) + o_p(\Delta_n^{1/2}), & \text{if } \tau \in \mathcal{T} \setminus \mathcal{T}_j, \\ \Delta X_{j,\tau} + \Delta_n^{1/2} \hat{\eta}_{j,\tau} + o_p(\Delta_n^{1/2}), & \text{if } \tau \in \mathcal{T}_{j-}. \end{cases} \tag{B.13}$$

Using these representations, we further deduce

$$\begin{cases} \widetilde{P}^{(1)} = \displaystyle\sum_{\tau \in \mathcal{T}_{1+} \cap \mathcal{T}_{2+}} (\Delta X_{1,\tau} \hat{\eta}_{2,\tau} + \Delta X_{2,\tau} \hat{\eta}_{1,\tau}) + o_p(1), \\ \widetilde{P}^{(2)} = \displaystyle\sum_{\tau \in \mathcal{T}_{1+} \setminus \mathcal{T}_2} \Delta X_{1,\tau} p\left(\hat{\eta}_{2,\tau}\right) + o_p(1), \\ \widetilde{P}^{(3)} = \displaystyle\sum_{\tau \in \mathcal{T}_{2+} \setminus \mathcal{T}_1} \Delta X_{2,\tau} p\left(\hat{\eta}_{1,\tau}\right) + o_p(1), \\ \widetilde{P}^{(4)} = o_p(1), \end{cases}$$

and

$$\begin{cases} \widetilde{N}^{(1)} = \displaystyle\sum_{\tau \in \mathcal{T}_{1-} \cap \mathcal{T}_{2-}} (\Delta X_{1,\tau} \hat{\eta}_{2,\tau} + \Delta X_{2,\tau} \hat{\eta}_{1,\tau}) + o_p(1), \\ \widetilde{N}^{(2)} = \displaystyle\sum_{\tau \in \mathcal{T}_{1-} \setminus \mathcal{T}_2} \Delta X_{1,\tau} n\left(\hat{\eta}_{2,\tau}\right) + o_p(1), \\ \widetilde{N}^{(3)} = \displaystyle\sum_{\tau \in \mathcal{T}_{2-} \setminus \mathcal{T}_1} \Delta X_{2,\tau} n\left(\hat{\eta}_{1,\tau}\right) + o_p(1), \\ \widetilde{N}^{(4)} = o_p(1). \end{cases}$$

These representations, together with the convergence (B.11), imply (B.9).

Step 3. This step proves the claim in (B.10). Let X' denote the continuous part of X, that is,

$$X'_t \equiv \int_0^t b_s ds + \int_0^t \sigma_s dW_s.$$

We define $\widehat{P}'$ and $\widehat{N}'$ in the same way as $\widehat{P}$ and $\widehat{N}$, but with X replaced by X'. It is easy to see that $\widehat{P}' - \sum_{i \notin \mathcal{I}_1 \cup \mathcal{I}_2} p\left(\Delta_i^n X_1\right) p\left(\Delta_i^n X_2\right)$ and $\widehat{N}' - \sum_{i \notin \mathcal{I}_1 \cup \mathcal{I}_2} n\left(\Delta_i^n X_1\right) n\left(\Delta_i^n X_2\right)$ are $O_p(\Delta_n)$. Hence, it suffices to prove the claim with $\widetilde{P}^{(5)}$ and $\widetilde{N}^{(5)}$ replaced, respectively, by the following variables:

$$\widetilde{P}' \equiv \Delta_n^{-1/2}\left(\widehat{P}'_{12} - P^\star_{12}\right), \quad \widetilde{N}' \equiv \Delta_n^{-1/2}\left(\widehat{N}'_{12} - N^\star_{12}\right).$$

In order to derive the stable convergence in law of $(\widetilde{P}', \widetilde{N}')$, we apply Theorem 5.3.5 in Jacod and Protter (2012) to the test function $x \mapsto (p(x_1)p\left(x_2\right), n\left(x_1\right) n\left(x_2\right))$. To do so, it is enough to verify that the limiting variables $B^{(1)}$, $B^{(2)}$, ζ and $\tilde{\zeta}$ coincide with Jacod and Protter's $\overline{A}$, $\overline{A}'$, $\overline{U}$ and $\overline{U}'$ variables. Turning to the details, we first note that, by Lemma 1(a), we can rewrite

$$P^\star_{12} = \int_0^T R_{c_s}\left(f\right) ds, \quad N^\star_{12} = \int_0^T R_{c_s}\left(g\right) ds. \tag{B.14}$$

By Lemma 1(b), we can rewrite $B_P^{(1)}$ and $B_N^{(1)}$ as

$$B_P^{(1)} = \int_0^T b_s^\top R_{c_s}\left(\partial f\right) ds, \quad B_N^{(1)} = \int_0^T b_s^\top R_{c_s}\left(\partial g\right) ds. \tag{B.15}$$

Given (2.6), no further rewriting is needed for the $B^{(2)}$ term.

By Lemma 1(c), we can rewrite

$$\gamma_t = \gamma_{c_t}\left(f\right) \text{ and } -\gamma_t = \gamma_{c_t}\left(g\right). \tag{B.16}$$

Hence, $\gamma_t = \sigma_t \hat{\gamma}_{\sigma_t}\left(f\right) = -\sigma_t \hat{\gamma}_{\sigma_t}\left(g\right)$ and, by (2.7),

$$\zeta_P = \int_0^T \hat{\gamma}_{\sigma_s}\left(f\right)^\top dW_s, \quad \zeta_N = \int_0^T \hat{\gamma}_{\sigma_s}\left(g\right)^\top dW_s. \tag{B.17}$$

By (B.16) and (B.5), we see that Γ_t defined in (2.9) can be written as

$$\Gamma_t = \begin{pmatrix} \Gamma_{c_t}\left(f,f\right) & \Gamma_{c_t}\left(f,g\right) \\ \Gamma_{c_t}\left(g,f\right) & \Gamma_{c_t}\left(g,g\right) \end{pmatrix}. \tag{B.18}$$

By Lemma 1(d), we see that $\overline{\Gamma}_t$ defined in (2.12) can be expressed equivalently as

$$\overline{\Gamma}_t = \begin{pmatrix} \overline{\Gamma}_{c_t}(f,f) & \overline{\Gamma}_{c_t}(f,g) \\ \overline{\Gamma}_{c_t}(g,f) & \overline{\Gamma}_{c_t}(g,g) \end{pmatrix}. \tag{B.19}$$

Recall from (2.11) that $\bar{\gamma}_t \equiv \overline{\Gamma}_t - \Gamma_t$. By (B.4), (B.18) and (B.19), it follows that

$$\bar{\gamma}_t = \begin{pmatrix} \bar{\gamma}_{c_t}(f,f) & \bar{\gamma}_{c_t}(f,g) \\ \bar{\gamma}_{c_t}(g,f) & \bar{\gamma}_{c_t}(g,g) \end{pmatrix}.$$

In view of (B.3), we verify that the local quadratic variation of the $\tilde{\zeta}$ term coincide with that defined in (5.2.4) of Jacod and Protter (2012).

We are now ready to apply Theorem 5.3.5 in Jacod and Protter (2012) to finish the proof of (B.10) and, hence, the assertion of Theorem 2. *Q.E.D.*

Proof of Proposition 1. By Proposition 1 in Li, Todorov, and Tauchen (2017b), the set $\widehat{\mathcal{I}}$ coincides with $\mathcal{I}_1 \cup \mathcal{I}_2$ with probability approaching one and, in restriction to such events,

$$\begin{aligned} \widehat{P}^{\star}_{12} &= \sum_{i \notin \mathcal{I}_1 \cup \mathcal{I}_2} p(\Delta_i^n X_1)\, p(\Delta_i^n X_2), \quad \widehat{P}^{\dagger}_{12} = \sum_{i \in \mathcal{I}_1 \cup \mathcal{I}_2} p(\Delta_i^n X_1)\, p(\Delta_i^n X_2), \\ \widehat{N}^{\star}_{12} &= \sum_{i \notin \mathcal{I}_1 \cup \mathcal{I}_2} n(\Delta_i^n X_1)\, n(\Delta_i^n X_2), \quad \widehat{N}^{\dagger}_{12} = \sum_{i \in \mathcal{I}_1 \cup \mathcal{I}_2} n(\Delta_i^n X_1)\, n(\Delta_i^n X_2). \end{aligned}$$

The assertion of Proposition 1 then follows from (B.9) and (B.10) in the proof of Theorem 2. *Q.E.D.*

Proof of Proposition 2. Let $\mathcal{T} = \mathcal{T}_1 \cup \mathcal{T}_2$. For notational simplicity, we denote, for each subset $\mathcal{S} \subseteq \mathcal{T}$,

$$\begin{aligned} \xi_P^*(\mathcal{S}) \equiv \Delta_n^{-1/2} \sum_{\tau \in \mathcal{S}} \big(& p\left(\Delta^n_{i(\tau)} X_1^* + \Delta_n^{1/2} \tilde{\eta}^*_{i(\tau),1}\right) p\left(\Delta^n_{i(\tau)} X_2^* + \Delta_n^{1/2} \tilde{\eta}^*_{i(\tau),2}\right) \\ & - p\left(\Delta^n_{i(\tau)} X_1^*\right) p\left(\Delta^n_{i(\tau)} X_2^*\right) \big). \end{aligned} \tag{B.20}$$

By Proposition 1 in Li, Todorov, and Tauchen (2017b), $\widehat{\mathcal{I}}$ coincides with $\mathcal{I}_1 \cup \mathcal{I}_2$ with probability approaching one. Hence, we can restrict our calculations to such events without loss of generality. In particular, we can write ξ_P^* as $\xi_P^*(\mathcal{T})$ using the notation (B.20). We can then decompose ξ_P^* as

$$\xi_P^* = \xi_P^*(\mathcal{T}_{1+} \cap \mathcal{T}_{2+}) + \xi_P^*(\mathcal{T}_{1+} \setminus \mathcal{T}_2) + \xi_P^*(\mathcal{T}_{2+} \setminus \mathcal{T}_1) + \xi_P^*(\mathcal{T}_{1-} \cup \mathcal{T}_{2-}).$$

We note that $\Delta^n_{i(\tau)} X_j^* = 0$ for all $\tau \in \mathcal{T} \setminus \mathcal{T}_j$ with probability approaching one. It is then

easy to deduce that

$$p\left(\Delta^n_{i(\tau)}X^*_j + \Delta^{1/2}_n \tilde{\eta}^*_{i(\tau),j}\right) = \begin{cases} p(\Delta^n_{i(\tau)}X^*_j) + \Delta^{1/2}_n \tilde{\eta}^*_{i(\tau),j} + o_p(\Delta^{1/2}_n), & \text{if } \tau \in \mathcal{T}_{j+}, \\ \Delta^{1/2}_n p(\tilde{\eta}^*_{i(\tau),j}) + o_p(\Delta^{1/2}_n), & \text{if } \tau \in \mathcal{T} \setminus \mathcal{T}_j, \\ o_p(\Delta^{1/2}_n), & \text{if } \tau \in \mathcal{T}_{j-}. \end{cases}$$

From there, we readily deduce that

$$\begin{cases} \xi^*_P\left(\mathcal{T}_{1+} \cap \mathcal{T}_{2+}\right) = \displaystyle\sum_{\tau \in \mathcal{T}_{1+} \cap \mathcal{T}_{2+}} \left(p(\Delta^n_{i(\tau)}X^*_1)\tilde{\eta}^*_{i(\tau),2} + p(\Delta^n_{i(\tau)}X^*_2)\tilde{\eta}^*_{i(\tau),1}\right) + o_p(1), \\ \xi^*_P\left(\mathcal{T}_{1+} \setminus \mathcal{T}_2\right) = \displaystyle\sum_{\tau \in \mathcal{T}_{1+} \setminus \mathcal{T}_2} p(\Delta^n_{i(\tau)}X^*_1) p\left(\tilde{\eta}^*_{i(\tau),2}\right) + o_p(1), \\ \xi^*_P\left(\mathcal{T}_{2+} \setminus \mathcal{T}_1\right) = \displaystyle\sum_{\tau \in \mathcal{T}_{2+} \setminus \mathcal{T}_1} p(\Delta^n_{i(\tau)}X^*_2) p\left(\tilde{\eta}^*_{i(\tau),1}\right) + o_p(1), \\ \xi^*_P\left(\mathcal{T}_{1-} \cup \mathcal{T}_{2-}\right) = o_p(1). \end{cases} \tag{B.21}$$

By a standard result for spot covariance estimation (see, e.g., Theorem 9.3.2 in Jacod and Protter (2012)), we have that $\hat{c}_{i(\tau)\pm} \xrightarrow{\mathbb{P}} c_{\tau\pm}$. Consequently,

$$\left(\tilde{\eta}^*_{i(\tau)}\right)_{\tau \in \mathcal{T}} \xrightarrow{\mathcal{L}|\mathcal{F}} \left(\tilde{\eta}_\tau\right)_{\tau \in \mathcal{T}},$$

where $\xrightarrow{\mathcal{L}|\mathcal{F}}$ denotes the convergence in probability of $\mathcal{F}$-conditional laws under the uniform metric. In addition, we note that $p(\Delta^n_{i(\tau)}X^*_j) \xrightarrow{\mathbb{P}} \Delta X_{j,\tau}$ for all $\tau \in \mathcal{T}_{j+}$. Combining these convergence results with (B.21), we deduce that $\xi^*_P \xrightarrow{\mathcal{L}|\mathcal{F}} \xi_P$. Evidently, the same argument can be used to show that $\xi^*_N \xrightarrow{\mathcal{L}|\mathcal{F}} \xi_N$, jointly with $\xi^*_P \xrightarrow{\mathcal{L}|\mathcal{F}} \xi_P$, which implies the first assertion of Proposition 2 (i.e., $\xi^*_P - \xi^*_N \xrightarrow{\mathcal{L}|\mathcal{F}} \xi_P - \xi_N$). The size and power properties of the test readily follow from here. *Q.E.D.*