Journal of Econometrics ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Journal of Econometrics

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

Mixed-scale jump regressions with bootstrap inference☆

Jia Li [a], Viktor Todorov [b], George Tauchen [a,*], Rui Chen [a]

[a] Department of Economics, Duke University, Durham, NC 27708, United States
[b] Department of Finance, Kellogg School of Management, Northwestern University, Evanston, IL 60208, United States

ARTICLE INFO

Article history:
Available online xxxx

JEL classification:
C51
C52
G12

Keywords:
Bootstrap
High-frequency data
Jumps
Regression
Semimartingale
Specification test
Stochastic volatility

ABSTRACT

We develop an efficient mixed-scale estimator for jump regressions using high-frequency asset returns. A fine time scale is used to accurately identify the locations of large rare jumps in the explanatory variables such as the price of the market portfolio. A coarse scale is then used in the estimation in order to attenuate the effect of trading frictions in the dependent variable such as the prices of potentially less liquid assets. The proposed estimator has a non-standard asymptotic distribution that cannot be made asymptotically pivotal via studentization. We propose a novel bootstrap procedure for feasible inference and justify its asymptotic validity. We show that the bootstrap provides an automatic higher-order asymptotic approximation by accounting for the sampling variation in estimates of nuisance quantities that are used in efficient estimation. The Monte Carlo analysis indicates good finite-sample performance of the general specification test and confidence intervals based on the bootstrap. We apply the method to a high-frequency panel of Dow stock prices together with the market index defined by the S&P 500 index futures over the period 2007–2014. We document remarkable temporal stability in the way that stocks react to market jumps. However, this relationship for many of the stocks in the sample is significantly noisier and more unstable during sector-specific jump events.

1. Introduction

The availability of high-frequency data has led to new ways of estimating an asset's exposures to systematic risks such as the aggregate stock market return in the standard CAPM. The high-frequency estimation approach (Andersen et al., 2003; Barndorff-Nielsen and Shephard, 2004a; Andersen et al., 2006; Mykland and Zhang, 2009) uses realized variation measures to infer beta over a fixed period of time, usually a day or a month, and then tracks these estimates over non-overlapping sample periods. More recent practice is to conduct estimation using jump-robust measures of variation and covariation (Todorov and Bollerslev, 2010; Gobbi and Mancini, 2012). All of the above mentioned beta measures (with or without truncation) mainly pertain to the locally Gaussian diffusive moves in the market, because the large number of small diffusive moves are known to account for a major part of the market variation. Economically speaking, these small moves in part reflect the market's gradual price discovery process of distilling minor news on fundamentals from noise trading (Kyle, 1985) which can lead to a situation with low signal to noise ratio and temporal instability.[1] Li et al. (2017), on the other hand, suggest an opposite approach that mainly uses abrupt and locally large jump moves to generate an effective measure of beta.[2,3] Such moves are typically related to important market-wide shocks which include, but are not limited to, macro announcements, geopolitical events and natural disasters. [See Chapter 8 of Hasbrouck (2015) for more discussion.]

The use of large rare jumps in a regression setting requires new ways of thinking about regression and inference. On the one hand, in any given fixed span of time, there are only a finite number of jumps. This means that the number of informative observations

☆ We thank the guest editors and anonymous referees as well as Torben Andersen, Snehal Banerjee, Tim Bollerslev, Anna Cieslak, Silvia Gonçalves, Andrew Patton and Brian Weller for helpful discussions and comments. We also thank Rob Engle and other participants of the 6th French Econometrics Conference celebrating Professor Gourieroux's Contribution to Econometrics. Li's and Todorov's research have been partially supported by NSF grants SES-1326819 and SES-0957330, respectively.

* Corresponding author.
E-mail addresses: jl410@duke.edu (J. Li), v-todorov@northwestern.edu (V. Todorov), george.tauchen@duke.edu (G. Tauchen), rui.chen@duke.edu (R. Chen).

http://dx.doi.org/10.1016/j.jeconom.2017.08.017
0304-4076/

[1] Indeed, Kalnina (2013) and Reiss et al. (2015) document that spot betas remain constant only over very short periods of time, usually a week or, at best, a month.

[2] Jump betas have been first introduced in Todorov and Bollerslev (2010). Todorov and Bollerslev (2010) use higher order power variations to identify the jump betas from the high-frequency data. This approach, unlike Li et al. (2017), makes use of all of the high-frequency increments. Of course, the role of the increments without jumps vanishes asymptotically in the higher order power variations.

[3] Theoretically, the betas at jump and non-jump times do not need to coincide.

in a jump regression is finite and does not increase to infinity asymptotically.[4] Therefore, the common intuition underlying the law of large numbers does not apply here. On the other hand, we recognize that the jumps are of fixed size regardless of the sampling frequency, whereas the diffusive moves are on the order $\Delta_n^{1/2}$, where Δ_n is the sampling interval which goes to zero asymptotically. The diffusive moves in the vicinity of jumps can be viewed as measurement errors induced by discrete sampling, and they play the role of random disturbances in classical regressions. The magnitude of such measurement error shrink at the parametric rate with well-behaved asymptotic properties, which can be further used for studying the asymptotics of our estimators. In the same vein, the correct specification of a linear jump regression model amounts to a perfect fitting (i.e., $R^2 = 1$) of the dependent jumps in the continuous-time limit. This test can be carried out by examining whether the observed R^2 is statistically significantly below unity.

This paper develops a new mixed-scale strategy for jump regressions, which addresses a natural asymmetry between the explanatory and dependent variables seen in applications.[5] On the one hand, the explanatory variables are often returns of highly liquid assets such as market index futures. We sample these variables at a *fine* scale (1-minute in our application), which greatly improves the accuracy of jump detection. On the other hand, the dependent variables are typically returns of less liquid assets such as individual stocks, which are subject to a slower price discovery process for incorporating large bits of new information. Realistically speaking, due to the trading mechanisms on the exchanges, a jump typically cannot be observed instantly. Rather, it is often realized through a sequence of transactions. See Barndorff-Nielsen et al. (2009) for a discussion of what they term "gradual jumps." It is therefore prudent to sample the asset prices for the dependent variables at a *coarse* scale when estimating the jump regression model, at the cost of statistical efficiency. The mixed-scale approach provides a flexible way of using data that play distinct roles in the jump regression. The fact that the jump detection step and the jump regression step are performed under two (possibly) distinct scales also leads to novel asymptotic results (cf. Li et al. (2017)). In addition, we present all theory here in a multivariate setting so as to facilitate applications to multi-factor models of risk exposure.

We further extend the analysis of Li et al. (2017) by providing a refined inference for the mixed-scale jump regression which is beneficial, particularly when sampling at coarser frequencies. We first derive a higher-order asymptotic approximation for the jump regression estimates. This expansion accounts for the error in the volatility estimation around the jump times (which is of higher order). We then propose a bootstrap method which we show is asymptotically valid. The bootstrap provides a conceptually different alternative to the higher-order asymptotic expansion. The latter is based on direct higher-order asymptotic approximations while the current bootstrap method is aimed at approximating the finite sample distribution of the estimator using simulated data. Our motivation for using the bootstrap is that the asymptotic distribution of the estimator of jump beta is non-standard because volatility may co-jump at the jump times of the explanatory variable(s); see, for example, Jacod and Todorov (2010), Todorov and Tauchen (2011) and Bandi and Renó (2016). In fact, the limiting distribution of the estimator is not Gaussian even conditional on the underlying information set. The asymptotic covariance matrix alone is thereby insufficient for asymptotically valid inference; in particular, the conventional t-statistic is not asymptotically pivotal. We therefore propose a bootstrap method that is very simple to implement. The user only needs to repeatedly compute the estimator in a bootstrap sample that consists of small subsamples within local windows of the detected jump times. The bootstrap sample size is much smaller than the original sample size, resulting in a significant reduction in computational time. The same bootstrap sample can also be used to compute critical values for the specification test. The bootstrap procedure achieves a higher-order refinement over the asymptotic approximations to the usual order. Our bootstrap refinement is atypical because it does not concern Edgeworth expansions for asymptotically pivotal statistics; instead, here, the refinement accounts for the higher-order sampling variability in the weights of the efficient regression procedure. Monte Carlo evidence shows good finite-sample performance of the bootstrap.

The bootstrap has been first introduced to the high-frequency literature by Gonçalves and Meddahi (2009) in the context of estimating integrated volatility. Since we focus on the inference about jumps, which is well known to be very different from the inference about volatility, the proposed bootstrap method and the associated asymptotic theory deviate significantly from prior work. To the best of our knowledge, the current paper is the first to study the bootstrap inference for jumps using high-frequency data. Although the bootstrap method is presented in the context of jump regressions, it can be readily extended to many other contexts concerning jumps as well.

We apply the mixed-scale jump regression method to a high-frequency one-minute panel of Dow stock prices together with the S&P 500 E-mini futures price for the market index over the period 2007–2014. We start with concrete examples of how individual asset prices react, either promptly or gradually, to news events generating market jumps, so as to illustrate the empirical relevance of the mixed-scale approach. We further provide evidence that using a coarse scale of 3–5 min around jump times is sufficiently conservative in the jump regression step for these blue-chip stocks; our evidence also indicates that using the fine scale is still appropriate for tasks like estimating local volatility which depend on price increments away from the problematic intervals with gradual jumps. We then proceed to conduct stock-by-stock tests of the key hypothesis that $R^2 = 1$.[6] A striking finding is that by sampling the data on a slightly coarse scale in the regression step, the null hypothesis is rejected much less frequently. This reduction cannot be fully explained by pure statistical reasons. Instead, it reaffirms the usefulness of the mixed-scale approach in the testing context. Using the efficient estimator, we document how the market jump risk exposure varies across stocks and over time. Lastly, we study the sensitivity of various stocks to market risk at alternative jump times defined by sector-specific jumps in the nine industry ETFs for the S&P 500 composite index. For many of the stocks in our sample, we find the relationship between individual stocks and the market to be significantly noisier and more unstable at the sector-specific jump times than it is at the market-wide jump times.

The rest of the paper is organized as follows. Section 2 describes the econometric framework and Section 3 presents the main theorems. Section 4 contains the Monte Carlo evaluation and Section 5 shows the empirical results. Section 6 concludes. All proofs are given in Appendix.

[4] Even if the asset price process has infinitely active jumps, the number of jumps that have sizes greater than any fixed level remains finite.

[5] Our mixed-scale strategy is designed to improve the accuracy of jump detection for a subvector of a multivariate semimartingale process, so the goal here is to reduce the misclassification (i.e., jump or non-jump) error. This is fundamentally different from the multi-scale method of Zhang et al. (2005), which conducts a jackknife bias-correction using realized variances computed at subsamples with different frequencies in the estimation of integrated volatility.

[6] Earlier work by Roll (1987) has documented relatively low R^2-s of time series regressions of stocks' returns on their systematic risk exposures, even after excluding days with firm-specific news (and hence more idiosyncratic noise).

Please cite this article in press as: Li J., et al., Mixed-scale jump regressions with bootstrap inference. Journal of Econometrics (2017), http://dx.doi.org/10.1016/j.jeconom.2017.08.017.

2. The setting for mixed-scale jump regressions

We describe the formal high-frequency asymptotic setting in Section 2.1 and the mixed-scale jump regression setting in Section 2.2 . The following notation is used in the sequel. We denote the transpose of a matrix A by $A^\top$ and denote its (j,k) component by A^{jk}. All vectors are column vectors. For notational simplicity, we write (a,b) in place of $(a^\top, b^\top)^\top$. For two vectors a and b, we write $a \leq b$ if the inequality holds component-wise. The Euclidean norm of a linear space is denoted by $\|\cdot\|$. The cardinality of a (possibly random) set $\mathcal{P}$ is denoted by $|\mathcal{P}|$. The largest smaller integer function is $\lfloor\cdot\rfloor$. For two sequences of positive real numbers a_n and b_n, we write $a_n \asymp b_n$ if $b_n/c \leq a_n \leq cb_n$ for some constant $c \geq 1$ and all n. All limits are for $n \to \infty$. We use $\xrightarrow{\mathbb{P}}$ and $\xrightarrow{\mathcal{L}\text{-}s}$ to denote convergence in probability and stable convergence in law, respectively.

2.1. The formal setup

We proceed with the formal setup. Let Y and Z be defined on a filtered probability space $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P}\right)$ which take values in $\mathbb{R}$ and $\mathbb{R}^{d_z}$, respectively. Throughout the paper, all processes are assumed to be càdlàg (i.e., right continuous with left limit) adapted. Let $X \equiv (Y,Z)$ and $d \equiv d_z + 1$. The d-dimensional process X is observed at discrete times $i\Delta_n$, for $i \in \{0,\ldots,\lfloor T/\Delta_n\rfloor\}$, within the *fixed* time interval $[0,T]$, where the sampling interval $\Delta_n \to 0$ asymptotically. We denote the increments of X by

$$\Delta_i^n X \equiv X_{i\Delta_n} - X_{(i-1)\Delta_n}, \quad i \in \mathcal{I}_n \equiv \{1,\ldots,\lfloor T/\Delta_n\rfloor\}. \tag{2.1}$$

Our basic assumption is that X is a d-dimensional Itô semimartingale (see, e.g., Jacod and Protter (2012), Section 2.1.4) of the form

$$\begin{cases} X_t = X_t^c + J_t, \\ X_t^c = X_0 + \int_0^t b_s ds + \int_0^t \sigma_s dW_s \quad \text{(continuous component)}, \\ J_t = \int_0^t \int_{\mathbb{R}} \delta(s,u)\,\mu(ds,du) \quad \text{(jump component)}, \end{cases} \tag{2.2}$$

where the drift b_t takes value in $\mathbb{R}^d$; the volatility process σ_t takes value in $\mathcal{M}_d$, the space of d -dimensional positive definite matrices; W is a d-dimensional standard Brownian motion; $\delta(\cdot) \equiv (\delta_Y(\cdot), \delta_Z(\cdot)) : \Omega \times \mathbb{R}_+ \times \mathbb{R} \mapsto \mathbb{R}^d$ is a predictable function; μ is a Poisson random measure on $\mathbb{R}_+ \times \mathbb{R}$ with its compensator $\nu(dt,du) = dt \otimes \lambda(du)$ for some measure λ on $\mathbb{R}$. The jump of X at time t is denoted by

$$\Delta X_t \equiv X_t - X_{t-}, \quad \text{where} \quad X_{t-} \equiv \lim_{s\uparrow t} X_s. \tag{2.3}$$

We denote the spot covariance matrix of X at time t by $c_t \equiv \sigma_t\sigma_t^\top$. Our basic regularity condition for X is the following.

Assumption 1. (a) The process $(b_t)_{t\geq 0}$ is locally bounded; (b) c_t is nonsingular for $t \in [0,T]$ almost surely; (c) $\nu([0,T]\times\mathbb{R}) < \infty$.

The only nontrivial restriction in Assumption 1 is the assumption of finite-activity jumps in X. This assumption is used mainly to simplify our technical exposition because our empirical focus in this paper are the big jumps. Technically speaking, this means that we can drop Assumption 1(c) and focus on jumps with size bounded away from zero. Doing so automatically verifies the finite-activity assumption, but with very little effect on the empirical investigation in the current paper.

2.2. Mixed-Scale jump regressions

The jump regression is based on the following (population) relationship between the jumps of Y and Z:

$$\Delta Y_\tau = \beta^{*\top} g(\Delta Z_\tau), \quad \tau \in \mathcal{T}, \tag{2.4}$$

where $g(\cdot) : \mathbb{R}^{d_z} \mapsto \mathbb{R}^q$ is a deterministic function, τ is a jump time of the process Z, and $\mathcal{T}$ collects these jump times. We stress that the restriction (2.4) is only postulated at the jump times of Z. In particular, we allow Y to have idiosyncratic jumps, i.e., jumps that do not occur at the same times as those of Z. Therefore, in general (provided $g(\mathbf{0}) = \mathbf{0}$) we have

$$\Delta Y_t = \beta^{*\top} g(\Delta Z_t) + \Delta\epsilon_t, \quad \Delta Z_t \Delta\epsilon_t = \mathbf{0}, \quad t \in [0,T], \tag{2.5}$$

with ϵ_t capturing the idiosyncratic jump risk in the asset Y. We note that this type of model for the jump parts of assets naturally arises in economies in which the market-wide pricing kernel is specified as a function of systematic factors (containing jumps) and the cash flows of the assets contain in addition idiosyncratic jump shocks in the sense of Merton (1976). We refer to Li et al. (2017) for more discussion of our jump model.

We refer to the coefficient β^* as the *jump beta*, which is the parameter of interest in our econometric analysis. As in Li et al. (2017), we are mainly interested in the linear specification $g(\Delta Z_\tau) = \Delta Z_\tau$ because it turns out to deliver quite good fitting in practice. That being said, the general form (2.4) is also of economic interest. For example, with $g(\Delta Z_\tau) = \left(\Delta Z_\tau 1_{\{\Delta Z_\tau > 0\}}, \Delta Z_\tau 1_{\{\Delta Z_\tau < 0\}}\right)$, (2.4) conveniently allows for asymmetric response of Y with respect to positive and negative jumps in Z . Assumption 2, below, ensures the identification of the jump beta. It also imposes some mild smoothness condition on $g(\cdot)$ that facilitates the asymptotic analysis.

Assumption 2. (a) The matrix $\sum_{\tau\in\mathcal{T}} g(\Delta Z_\tau) g(\Delta Z_\tau)^\top$ is nonsingular almost surely.

(b) For each t, the measure defined by $A \mapsto \lambda(\{u : \delta_Z(t,u) \in A \setminus \{0\}\})$ is atomless. Moreover, $g(\cdot)$ is twice continuously differentiable almost everywhere.

In finite samples, neither the times nor the magnitudes of jumps are directly observable. Empirically, we need to use discretely sampled data $\Delta_i^n X = \left(\Delta_i^n Y, \Delta_i^n Z\right)$ to make statistical inference based on model (2.4). Since (2.4) only concerns the jump moves of the asset prices, it is conceptually natural to first select observed returns that contain jumps. We do so using a thresholding method (Mancini, 2001) as follows. We consider a sequence of thresholds $(u_n)_{n\geq 1} \subset \mathbb{R}^{d_z}$ such that

$$u_{j,n} \asymp \Delta_n^\varpi, \quad \text{for some } \varpi \in (0,1/2) \text{ and all } 1 \leq j \leq d_z.$$

We then collect the jump returns using

$$\mathcal{J}_n \equiv \mathcal{I}_n \setminus \left\{i : -u_n \leq \Delta_i^n Z \leq u_n\right\}. \tag{2.6}$$

Time-invariant choice for u_n, although asymptotically valid, leads to very bad results in practice as it does not account for the time-varying diffusive spot covariance matrix c_t. Hence, a sensible choice for u_n should take into account the variation of c_t in an adaptive, data-driven way. We refer to our application in Sections 4 and 5 for the details of such a way of constructing u_n using the bipower variation estimator of Barndorff-Nielsen and Shephard (2004b).

Under Assumption 1, it can be shown that $\mathcal{J}_n$ consistently locates the sampling intervals that contain jumps.[7] That is,

$$\mathbb{P}\left(\mathcal{J}_n = \mathcal{J}_n^*\right) \to 1, \quad \text{where} \quad \mathcal{J}_n^* \equiv \{i : \tau \in ((i-1)\Delta_n, i\Delta_n] \text{ for some } \tau \in \mathcal{T}\}. \tag{2.7}$$

[7] See, for example, Proposition 1 of Li et al. (2017).

Parallel to (2.4), the jump regression equation is then given by

$$\Delta_i^n Y = \beta^{*\top} g\left(\Delta_i^n Z\right) + \varepsilon_i^n, \quad i \in \mathcal{J}_n, \tag{2.8}$$

with the error term ε_i^n being implicitly defined by (2.8).

Despite the apparent similarity between the jump regression Eq. (2.8) and the classical regression, there are fundamental differences. We first observe that (2.8) only concerns a finite number of large jump returns even asymptotically (recall (2.7)). This means, the intuition underlying the classical law of large numbers and the central limit theorem does not apply here. The reason is that the finite number of error terms $\left(\varepsilon_i^n\right)_{i\in\mathcal{J}_n}$ would not "average out." However, we observe that these error terms are actually asymptotically small. Indeed, under (2.4), we have for each $i \in \mathcal{J}_n^*$,

$$\varepsilon_i^n = \Delta_i^n Y^c - \beta^{*\top}\left(g\left(\Delta Z_\tau + \Delta_i^n Z^c\right) - g\left(\Delta Z_\tau\right)\right),$$

where τ is the unique (which holds true at high frequency) jump time that occurs in $((i-1)\Delta_n, i\Delta_n]$. Since the diffusive moves $\left(\Delta_i^n Y^c, \Delta_i^n Z^c\right)$ are of order $O_p(\Delta_n^{1/2})$, so is ε_i^n. In addition, these small error terms have well-behaved asymptotic properties, which we use to derive the asymptotic property of our inference procedures.

In empirical work, the use of high-frequency data is confounded by various trading frictions that make the transaction price deviate from the efficient price, with the latter being typically the object of interest. The deviation of the observed from efficient (fundamental) price is commonly referred to as microstructure noise. Typical sources of microstructure noise are bid–ask bounces and rounding error. A standard assumption in the literature is to assume that the noise is centered at zero and it has some form of weak dependence across observation times. There is a large body of work dealing with microstructure noise of this type, see e.g., chapter 16 of Jacod and Protter (2012) and the many references therein. The earlier literature has "dealt" with the potential presence of noise by sampling sparsely, with the idea being that at the coarser frequencies the importance of the noise in relative terms is rather small and can be ignored. Subsequent work has developed formal statistical methods for dealing with the microstructure noise. Although the methods differ, they are all based on averaging the noise in some way. In other words, the existing methods all rely on weak dependence of the noise at observation times and apply law of large numbers to purge the high frequency based measures from it.

There is another type of microstructure noise, which stems from staleness and infrequent trading. Mainly, for assets which are not very liquid, the price can be relatively slow to react to news. In particular, when there is a big jump on the stock market, less liquid individual assets can be slow to react and adjust fully to the new (latent) efficient price level that corresponds to the new information. There can be various sources for this type of price staleness. One typical example is the presence of stale limit orders in the limit order book which get "hit" as the price is adjusting to the new equilibrium level. The staleness in prices causes a phenomenon referred to by Barndorff-Nielsen et al. (2009) as gradual jumps. Obviously, this type of noise is very difficult to deal with formally as by its very nature it has a lot of dependence across observation times and also it depends very strongly on the actual fundamental price. Hence, local averaging type procedures will not work for it. Also, it is clear that this type of trading friction has a rather nontrivial impact on the analysis of jumps since, by their very nature, jumps are rare events.

In this paper we are mainly concerned with the second type of noise, i.e., the one that is due to staleness.[8] To mitigate its impact, we will sample sparsely. The proper sampling scheme of course depends on the asset of interest as staleness and liquidity are asset specific. For example, in our applications, we take Y to be the price of a blue-chip stock and Z to be the price of a futures contract on a major market index, with Z expected to be more liquid than Y.

The difference in liquidity of the left- and the right-hand side assets around market jump times creates an interesting tradeoff in the choice of the sampling scheme. On the one hand, sampling at high frequency (e.g., 1 min) greatly increases the accuracy for jump detection and, hence, reduces jump-misclassification bias in finite samples. On the other hand, sampling at such frequency is unlikely to be conservative enough for mitigating the microstructure effect of gradual jumps in Y. Indeed, as we shall illustrate with concrete examples in Section 5, individual stocks may take longer time than the market to fully incorporate new information that leads to a visible jump in the market index. See Barndorff-Nielsen et al. (2009) for additional discussions on this type of gradual jumps.

We propose to break the tension between these two conflicting effects using a mixed-scale jump regression procedure: we maintain the jump detection (2.6) at the fine sampling scale Δ_n, but implement the jump regression at a (possibly) coarser scale $k\Delta_n$ for some $k \geq 1$. By doing so, we maintain high precision in detecting market jumps and reduce the concern of "breaking" gradual jumps. More precisely, we denote $\Delta_{i,k}^n X = (\Delta_{i,k}^n Y, \Delta_{i,k}^n Z)$, where

$$\Delta_{i,k}^n X = X_{(i-1+k)\Delta_n} - X_{(i-1)\Delta_n}.$$

The mixed-scale jump regression is then given by, with $\varepsilon_{i,k}^n$ implicitly defined below,

$$\Delta_{i,k}^n Y = \beta^{*\top} g\left(\Delta_{i,k}^n Z\right) + \varepsilon_{i,k}^n, \quad i \in \mathcal{J}_n. \tag{2.9}$$

Clearly, (2.8) is a special case of (2.9) with $k = 1$. The fact that the jump detection and the jump regression are performed at different sampling scales leads to notable differences between the inference procedures proposed below and those in the single-scale setting of Li et al. (2017) , mainly because of the presence of volatility-price co-jumps. We now turn to the details.

3. Asymptotic theory

3.1. The efficient estimation of jump beta

In this subsection, we describe a class of mixed-scale estimators for the jump beta and derive their asymptotic properties. In view of (2.9), a natural estimator of β^* is the mixed-scale ordinary least squares (OLS) estimator given by

$$\hat{\beta}_n \equiv \left(\sum_{i\in\mathcal{J}_n} g\left(\Delta_{i,k}^n Z\right) g\left(\Delta_{i,k}^n Z\right)^\top\right)^{-1}\left(\sum_{i\in\mathcal{J}_n} g\left(\Delta_{i,k}^n Z\right)\Delta_{i,k}^n Y\right).$$

However, since the error terms $(\varepsilon_{i,k}^n)_{i\in\mathcal{J}_n}$ can exhibit arbitrary heteroskedasticity due to both time-varying volatility and jump size, the mixed-scale OLS estimator is not efficient. Following Li et al. (2017), we consider efficient estimation using a semiparametric two-step weighted estimator.

To construct the weights, we first nonparametrically estimate the spot covariance matrices before and after each detected jump. To this end, we pick an integer sequence k_n of block sizes such that

$$k_n \to \infty \quad \text{and} \quad k_n\Delta_n \to 0. \tag{3.1}$$

We also pick a $\mathbb{R}^d$-valued sequence u_n' of truncation thresholds that satisfies

$$u_{j,n}' \asymp \Delta_n^{-\varpi}, \quad \text{for some } \varpi \in (0, 1/2) \text{ and all } 1 \leq j \leq d.$$

We then set the index set of the diffusive returns to be

$$\mathcal{C}_n = \left\{i \in \mathcal{I}_n : -u_n' \leq \Delta_i^n X \leq u_n'\right\}. \tag{3.2}$$

[8] For the frequencies we use in our empirical work, the first type of noise has relatively small impact, see Section 5 for further details.

For each $i \in \mathcal{J}_n$, we estimate the pre-jump and the post-jump spot covariance matrices using

$$\begin{cases} \hat{c}_{n,i-} \equiv \dfrac{\sum_{j=0}^{k_n-1}(\Delta^n_{i-k_n+j}X)(\Delta^n_{i-k_n+j}X)^\top 1_{\{i-k_n+j\in\mathcal{C}_n\}}}{\Delta_n \sum_{j=0}^{k_n-1} 1_{\{i-k_n+j\in\mathcal{C}_n\}}}, \\ \hat{c}_{n,i+} \equiv \dfrac{\sum_{j=0}^{k_n-1}(\Delta^n_{i+k+j}X)(\Delta^n_{i+k+j}X)^\top 1_{\{i+k+j\in\mathcal{C}_n\}}}{\Delta_n \sum_{j=0}^{k_n-1} 1_{\{i+k+j\in\mathcal{C}_n\}}}. \end{cases} \tag{3.3}$$

We note that these spot covariance estimates are constructed using returns sampled at the "fine" scale which, in our empirical analysis in Section 5, is set to be 1 minute. At such frequency, the microstructure noise due to bid–ask bounces and rounding has a negligible impact on volatility estimation for liquid stocks (we provide empirical evidence for that in the empirical section). That being said, one may also estimate spot volatilities at coarser sampling intervals to further guard against microstructure noise but at the cost of higher sampling variability in finite samples. This results in notational changes only in the theory that follows and we omit the details for brevity.

We consider a weight function $w : \mathcal{M}_d \times \mathcal{M}_d \times \mathbb{R}^{d_z} \times \mathbb{R}^q \mapsto (0, \infty)$ such that $w(c_-, c_+, z, \beta)$ is continuously differentiable at $\beta = \beta^*$, all $c_-, c_+ \in \mathcal{M}_d$ and almost every $z \in \mathbb{R}^{d_z}$. To simplify notation, we denote

$$\hat{w}_{n,i} = w\left(\hat{c}_{n,i-}, \hat{c}_{n,i+}, \Delta^n_{i,k}Z, \hat{\beta}_n\right).$$

The mixed-scaled WLS estimator is then given by

$$\hat{\beta}_n(w) \equiv \left(\sum_{i\in\mathcal{J}_n} \hat{w}_{n,i} g\left(\Delta^n_{i,k}Z\right) g\left(\Delta^n_{i,k}Z\right)^\top\right)^{-1} \times \left(\sum_{i\in\mathcal{J}_n} \hat{w}_{n,i} g\left(\Delta^n_{i,k}Z\right) \Delta^n_{i,k}Y\right). \tag{3.4}$$

In order to describe the asymptotic behavior of $\hat{\beta}_n(w)$, we introduce some auxiliary random variables. Let $(\tau_p)_{p\geq 1}$ denote the successive jump times of Z. We consider random variables $\left(\kappa_p, \xi_{p-}, \xi_{p+}\right)_{p\geq 1}$ that are mutually independent and are independent of $\mathcal{F}$ such that κ_p is uniformly distributed on $[0, 1]$ and the variables $\left(\xi_{p-}, \xi_{p+}\right)$ are d-dimensional standard normal. We then denote, for $p \geq 1$,

$$\begin{cases} \varsigma_p \equiv \left(1, -\beta^{*\top}\partial g\left(\Delta Z_{\tau_p}\right)\right)\left(\sqrt{\kappa_p}\sigma_{\tau_p-}\xi_{p-} \right. \\ \quad \left. + \sqrt{k-\kappa_p}\sigma_{\tau_p}\xi_{p+}\right), \\ w_p \equiv w\left(c_{\tau_p-}, c_{\tau_p}, \Delta Z_{\tau_p}, \beta^*\right). \end{cases} \tag{3.5}$$

Finally, we set

$$\Xi(w) \equiv \sum_{p\in\mathcal{P}} w_p g\left(\Delta Z_{\tau_p}\right) g\left(\Delta Z_{\tau_p}\right)^\top,$$

$$\Lambda(w) \equiv \sum_{p\in\mathcal{P}} w_p g\left(\Delta Z_{\tau_p}\right) \varsigma_p.$$

Theorem 1, below, describes the stable convergence in law of $\hat{\beta}_n(w)$.

Theorem 1. *Under Assumptions 1 and 2,* $\Delta_n^{-1/2}(\hat{\beta}_n(w) - \beta^*) \xrightarrow{\mathcal{L}\text{-}s} \Xi(w)^{-1}\Lambda(w)$.

Theorem 1 shows that $\hat{\beta}_n(w)$ is a $\Delta_n^{-1/2}$-consistent estimator of the jump beta, with $\mathcal{F}$-conditional asymptotic covariance matrix given by

$$\Sigma(w) \equiv \Xi(w)^{-1}\left(\sum_{p\in\mathcal{P}} w_p^2 \mathbb{E}\left[\varsigma_p^2|\mathcal{F}\right] g\left(\Delta Z_{\tau_p}\right) g\left(\Delta Z_{\tau_p}\right)^\top\right)\Xi(w)^{-1},$$

where

$$\mathbb{E}\left[\varsigma_p^2|\mathcal{F}\right] = \left(1, -\beta^{*\top}\partial g\left(\Delta Z_{\tau_p}\right)\right)\left(\frac{1}{2}c_{\tau_p-} + \left(k - \frac{1}{2}\right)c_{\tau_p}\right) \times \left(1, -\beta^{*\top}\partial g\left(\Delta Z_{\tau_p}\right)\right)^\top.$$

It is easy to see that $\Sigma(w)$ can be minimized using the weight function

$$w(c_-, c_+, z, \beta) \equiv \frac{1}{\left(1, -\beta^\top\partial g(z)\right)\left(\frac{1}{2}c_- + \left(k - \frac{1}{2}\right)c_+\right)\left(1, -\beta^\top\partial g(z)\right)^\top}.$$

We refer to the associated estimator as the optimally weighted estimator. By construction, it is more efficient than an unweighted estimator. Moreover, Li et al. (2017) establish the semiparametric efficiency bound for estimating the jump beta in the case without volatility-price cojumps. In this case, the optimally weighted estimator defined above attains the efficiency bound computed for the coarsely sampled data. The reason for using the coarser frequency in the analysis of the semiparametric efficiency of the jump beta estimation is that the limiting distribution of the jump regression coefficient is determined only by the high-frequency increments containing the jumps. However, recall that these increments are aggregated to a coarser scale in order to guard against the gradual jump phenomenon. In this regard, we should stress that the frequency used for jump detection as well as for the estimation of volatility has no bearing on the efficiency statement. The reason is that the error coming from the jump detection as well as volatility measurement is of higher order in the jump regression.

3.2. Higher-order refinement and bootstrap inference

We now develop refined inference for the jump regression estimate of β^*. We first derive a higher-order asymptotic result and then propose a bootstrap procedure which we show achieves the asymptotic refinement.

To motivate, we observe that while the weighted estimator $\hat{\beta}_n(w)$ depends on the spot covariance estimates $\left(\hat{c}_{n,i-}, \hat{c}_{n,i+}\right)$, the sampling variability of the latter is not reflected in the asymptotic distribution described by Theorem 1 . The reason is that the local volatility estimates enter only the weights and their sampling errors are annihilated in the second-order asymptotics. In finite samples, the sampling variability of the spot covariance estimates may still have some effect, because the latter enjoy only a nonparametric convergence rate. To account for such effects, we need a refined characterization of the asymptotic behavior of the weighted estimator which we now provide. For the analysis here we need the following additional assumption for the volatility process.

Assumption 3. The process σ_t is also an Itô semimartingale of the form

$$\sigma_t = \sigma_0 + \int_0^t \tilde{b}_s ds + \int_0^t \tilde{\sigma}_s dW_s + \int_0^t \int_{\mathbb{R}} \tilde{\delta}(s,u) 1_{\{\|\tilde{\delta}(s,u)\|>1\}} \mu(ds, du) + \int_0^t \int_{\mathbb{R}} \tilde{\delta}(s,u) 1_{\{\|\tilde{\delta}(s,u)\|\leq 1\}} (\mu - \nu)(ds, du),$$

where the processes $\tilde{b}_t$ and $\tilde{\sigma}_t$ are locally bounded and for a sequence of stopping times $(T_m)_{m\geq 1}$ increasing to infinity and a sequence $(\tilde{J}_m)_{m\geq 1}$ of λ-integrable bounded functions, $\|\tilde{\delta}(t,u)\|^2 \wedge 1 \leq \tilde{J}_m(u)$ for all $t \leq T_m$ and $u \in \mathbb{R}$.

Assumption 3 is needed for characterizing the stable convergence of the spot covariance estimates. This assumption is fairly unrestrictive and is satisfied by many models in finance. In particular, it allows for "leverage effect," that is, the Brownian motions W and $\widetilde{W}$ can be correlated. Moreover, Assumption 3 allows for volatility jumps, and it does not restrict their activity and dependence with the price jumps. However, this assumption does rule out certain long-memory volatility models driven by a fractional Brownian motion (see Comte and Renault (1996)).

We also need some additional notation. We consider $d \times d$ random matrices $\left(\zeta_{p-}, \zeta_{p+}\right)_{p\geq 1}$ which, conditional on $\mathcal{F}$, are centered Gaussian, mutually independent and independent of $\left(\kappa_p, \xi_{p-}, \xi_{p+}\right)_{p\geq 1}$, with conditional covariances given by

$$\begin{cases} \mathbb{E}[\zeta_{p-}^{jk}\zeta_{p-}^{lm}|\mathcal{F}] = c_{\tau_p-}^{jl}c_{\tau_p-}^{km} + c_{\tau_p-}^{jm}c_{\tau_p-}^{kl}, \\ \mathbb{E}[\zeta_{p+}^{jk}\zeta_{p+}^{lm}|\mathcal{F}] = c_{\tau_p}^{jl}c_{\tau_p}^{km} + c_{\tau_p}^{jm}c_{\tau_p}^{kl}, \end{cases} \quad 1 \leq j, k, l, m \leq d.$$

We denote the first differential of w by $dw\left(c_-, c_+, z, b\right) = \dot{w}(c_-, c_+, z, b; dc_-, dc_+, dz, db)$ and then set, for $p \geq 1$,

$$\tilde{w}_p \equiv \dot{w}\left(c_{\tau_p-}, c_{\tau_p}, \Delta Z_{\tau_p}, \beta^*; \zeta_{p-}, \zeta_{p+}, 0, 0\right).$$

Finally, for notational simplicity, we set

$$\begin{cases} \Xi\left(w\right) \equiv \sum_{p\in\mathcal{P}} w_p g\left(\Delta Z_{\tau_p}\right) g\left(\Delta Z_{\tau_p}\right)^\top, \\ \Lambda\left(w\right) \equiv \sum_{p\in\mathcal{P}} w_p g\left(\Delta Z_{\tau_p}\right) \varsigma_p, \\ \tilde{\Xi}\left(w\right) \equiv \sum_{p\in\mathcal{P}} \tilde{w}_p g\left(\Delta Z_{\tau_p}\right) g\left(\Delta Z_{\tau_p}\right)^\top, \\ \tilde{\Lambda}\left(w\right) \equiv \sum_{p\in\mathcal{P}} \tilde{w}_p g\left(\Delta Z_{\tau_p}\right) \varsigma_p. \end{cases}$$

The higher-order asymptotic expansion for $\Delta_n^{-1/2}(\hat{\beta}_n\left(w\right) - \beta^*)$ is given in the following theorem.

Theorem 2. *Suppose Assumptions 1–3, and $k_n \asymp \Delta_n^{-a}$ for some $a \in (0, 1/2)$.*

(a) *We can decompose*

$$\Delta_n^{-1/2}\left(\hat{\beta}_n\left(w\right) - \beta^*\right) = \mathcal{L}_n\left(w\right) + k_n^{-1/2}\mathcal{H}_n\left(w\right) + o_p(k_n^{-1/2}), \quad (3.6)$$

such that

$$\left(\mathcal{L}_n\left(w\right), \mathcal{H}_n\left(w\right)\right) \xrightarrow{\mathcal{L}\text{-}s} \left(\mathcal{L}\left(w\right), \mathcal{H}\left(w\right)\right),$$

where

$$\begin{cases} \mathcal{L}\left(w\right) \equiv \Xi(w)^{-1}\Lambda\left(w\right), \\ \mathcal{H}\left(w\right) \equiv \Xi(w)^{-1}\tilde{\Lambda}\left(w\right) - \Xi(w)^{-1}\tilde{\Xi}\left(w\right)\Xi(w)^{-1}\Lambda\left(w\right). \end{cases}$$

(b) *If, in addition, there are no price-volatility cojumps and W is independent of (σ, J), then* $\sup_x|\mathbb{P}(\mathcal{L}_n\left(w\right) \leq x|\sigma, J) - \mathbb{P}(\mathcal{L}\left(w\right) \leq x|\sigma, J)| = O_p(\Delta_n^{1/2})$.

The leading term $\mathcal{L}_n\left(w\right)$ in (3.6) is what drives the convergence in Theorem 1. The higher-order term $k_n^{-1/2}\mathcal{H}_n\left(w\right)$ is $O_p(k_n^{-1/2})$ and hence is asymptotically dominated by $\mathcal{L}_n\left(w\right)$. The limiting variable $\mathcal{H}_n\left(w\right)$ involves not only $\left(\varsigma_p\right)_{p\geq 1}$ but also $\left(\tilde{w}_p\right)_{p\geq 1}$, where the latter sequence captures the sampling variability in the weights due to the spot variance estimates. Part (b) of Theorem 2 further shows that the conditional law of the leading term converges at a (fast) parametric rate under the uniform metric.

Because of the higher-order asymptotic effect played by $\hat{c}_{n,i\pm}$ in the efficient beta estimation, the user has a lot of freedom in setting the block size k_n. Indeed, as seen from Theorem 2, we need only $k_n \asymp \Delta_n^{-a}$ with a in the wide range of $(0, 1/2)$. This is unlike the block-based volatility estimators, see e.g., Jacod and Rosenbaum (2013), where one has significantly less freedom in choosing k_n. Having the refined asymptotic result in Theorem 2 helps since if k_n is relatively small, the higher-order term $k_n^{-1/2}\mathcal{H}_n\left(w\right)$ might have nontrivial finite sample effect.

We now introduce a bootstrap algorithm and show that (see Theorem 3 below) it provides the higher-order approximation described in Theorem 2. With a mild adjustment, the same bootstrap sample can also be used to compute critical values for the specification test developed in Section 3.3. The bootstrap was first introduced to the high-frequency setting by Gonçalves and Meddahi (2009) and Dovonon et al. (2013) for making inference for integrated variance and covariance matrices; also see Hounyo (2013) and Dovonon et al. (2014). We apply here the bootstrap to make inference for jumps, which is therefore very different from prior work that concerns volatility inference.[9]

Algorithm 1 (*Bootstrapping $\hat{\beta}_n(w)$*). *Step 1.* In each bootstrap sample, we generate a d -dimensional standard Brownian motion W^* and random times $\left(\tau_i^*\right)_{i\in\mathcal{J}_n}$ which are mutually independent and independent of the data, such that each τ_i^* is drawn uniformly from $[(i-1)\Delta_n, i\Delta_n]$.[10] Set the diffusive return for each $i \in \mathcal{J}_n$ as

$$\begin{aligned} \Delta_{i,k}^n X^{*c} \equiv \begin{pmatrix} \Delta_{i,k}^n Y^{*c} \\ \Delta_{i,k}^n Z^{*c} \end{pmatrix} &= \hat{c}_{n,i-}^{1/2}(W_{\tau_i^*}^* - W_{(i-1)\Delta_n}^*) \\ &+ \hat{c}_{n,i+}^{1/2}(W_{(i-1+k)\Delta_n}^* - W_{\tau_i^*}^*). \end{aligned} \quad (3.7)$$

Step 2. Set $\Delta_{i,k}^n Z^* = \Delta_{i,k}^n Z + \Delta_{i,k}^n Z^{*c}$ and $\Delta_{i,k}^n Y^* = \hat{\beta}_n(w)^\top g(\Delta_{i,k}^n Z) + \Delta_{i,k}^n Y^{*c}$ for $i \in \mathcal{J}_n$. Compute $\hat{\beta}_n^*$ as the OLS estimator by regressing $\Delta_{i,k}^n Y^*$ on $g(\Delta_{i,k}^n Z^*)$ in the subsample $i \in \mathcal{J}_n$.
Step 3. For each $i \in \mathcal{J}_n$, set

$$\begin{aligned} &\Delta_{i-k_n+j}^n X^{*c} = \hat{c}_{n,i-}^{1/2}\Delta_{i-k_n+j}^n W^*, \quad \Delta_{i+k+j}^n X^{*c} = \hat{c}_{n,i+}^{1/2}\Delta_{i+k+j}^n W^*, \\ &0 \leq j \leq k_n - 1, \end{aligned} \quad (3.8)$$

and compute $(\hat{c}_{n,i-}^*, \hat{c}_{n,i+}^*)$ as

$$\begin{cases} \hat{c}_{n,i-}^* \equiv \dfrac{1}{k_n\Delta_n}\sum_{j=0}^{k_n-1}\left(\Delta_{i-k_n+j}^n X^{*c}\right)\left(\Delta_{i-k_n+j}^n X^{*c}\right)^\top, \\ \hat{c}_{n,i+}^* \equiv \dfrac{1}{k_n\Delta_n}\sum_{j=0}^{k_n-1}\left(\Delta_{i+k+j}^n X^{*c}\right)\left(\Delta_{i+k+j}^n X^{*c}\right)^\top. \end{cases} \quad (3.9)$$

Step 4. Compute $\hat{\beta}_n^*\left(w\right)$ in the bootstrap sample using (3.4) with $\left(\Delta_{i,k}^n Y, \Delta_{i,k}^n Z, \hat{w}_{n,i}\right)_{i\in\mathcal{J}_n}$ replaced by $(\Delta_{i,k}^n Y^*, \Delta_{i,k}^n Z^*, \hat{w}_{n,i}^*)_{i\in\mathcal{J}_n}$, where $\hat{w}_{n,i}^* \equiv w(\hat{c}_{n,i-}^*, \hat{c}_{n,i+}^*, \Delta_{i,k}^n Z^*, \hat{\beta}_n^*)$.

In summary, Algorithm 1 suggests computing $\hat{\beta}_n^*\left(w\right)$ in the same way as $\hat{\beta}_n(w)$ using the bootstrap sample. One exception is that the computation of the spot covariances (see (3.9)) does not require truncation, because we only use the diffusive returns in the bootstrap. It is important to observe that the spot covariance matrices and the weights are also resampled so as to capture their sampling variability in the higher-order asymptotics.

Theorem 3, below, describes the convergence in probability of the $\mathcal{F}$-conditional law of the bootstrap estimator $\hat{\beta}_n^*\left(w\right)$. For a sequence of random variables A_n, we write $A_n \xrightarrow{\mathcal{L}|\mathcal{F}} A$ if the $\mathcal{F}$-conditional law of A_n converges in probability to that of A under any metric for the weak convergence of probability measures.[11]

[9] Dovonon et al. (2014) consider an application of the bootstrap for approximating the null asymptotic distribution of jump tests, which mainly concerns the jump-robust inference for the integrated variance, rather than the jump process itself.

[10] We note that the Gaussian increments of W^* are only needed within two-sided k_n-windows around the jump returns. This fact is useful for reducing the computational cost in practice.

[11] We note that $A_n \xrightarrow{\mathcal{L}|\mathcal{F}} A$ amounts to the convergence of $\mathcal{F}$-conditional law in a weak sense, namely the convergence is in probability for measure-valued random

Theorem 3. *Suppose the same conditions as in Theorem 2. Then we can decompose*

$$\Delta_n^{-1/2}\left(\hat{\beta}_n^*(w)-\hat{\beta}_n(w)\right)=\mathcal{L}_n^*(w)+k_n^{-1/2}\mathcal{H}_n^*(w)+o_p(k_n^{-1/2}),\tag{3.10}$$

such that

$$\left(\mathcal{L}_n^*(w),\mathcal{H}_n^*(w)\right)\xrightarrow{\mathcal{L}|\mathcal{F}}(\mathcal{L}(w),\mathcal{H}(w)),$$

where $(\mathcal{L}(w),\mathcal{H}(w))$ *are defined as in Theorem 2.*

Theorem 3 justifies using the $\mathcal{F}$-conditional distribution of $\Delta_n^{-1/2}(\hat{\beta}_n^*(w)-\hat{\beta}_n(w))$ to approximate the $\mathcal{F}$-conditional limiting distribution of $\Delta_n^{-1/2}(\hat{\beta}_n(w)-\beta^*)$. Importantly, the approximation not only captures the leading term $\mathcal{L}(w)$, but also the higher-order term $k_n^{-1/2}\mathcal{H}(w)$.[12] We further note that both $\mathcal{L}(w)$ and $\mathcal{H}(w)$ are $\mathcal{F}$-conditionally symmetric. Therefore, the basic bootstrap and the percentile bootstrap (see, e.g., Davison and Hinkley (1997)) can both be used for constructing bootstrap confidence intervals.

To summarize, refined inference for the jump regression coefficients can be done either by the use of the higher order asymptotic result in Theorem 2 or the bootstrap procedure based on the result of Theorem 3. We provide no asymptotic justification for preferring one over the other. In both methods the higher order asymptotic effect from the estimation of volatility around the jump times is accounted for. In addition, both methods ignore errors in the jump regression which are of even higher order (than the error due to the estimation of volatility), like the errors in detecting the locations of the jump times as well as the error due to the time variation in the volatility in the local blocks around the jump times. The difference between the inference based on the higher order asymptotic result and the bootstrap method is that the latter is based on mimicking the finite sample distribution of the regression estimator assuming jumps are located correctly and the volatility does not vary over the local blocks. The inference based on the higher order asymptotic result, on the other hand, is based on asymptotic expansion of the regression estimator in the above simplified setting (i.e., when assuming jumps are located correctly and volatility is constant over the local windows around the jump times). In that sense, the difference between the two methods of refined inference for the jump regression is similar to the difference between inference based on asymptotic theory and bootstrap in classical settings, see e.g., Section 2 of Horowitz (2001).[13] Finally, on the practical side the bootstrap is conceptually simple to grasp in the sense that the econometrician only needs to repeatedly compute the mixed-scale OLS or WLS estimator in the bootstrap samples.

3.3. Specification testing and its bootstrap implementation

We proceed with a specification test for (2.4), which generalizes the test of Li et al. (2017) to a multivariate mixed-scale setting. Since (2.4) is no longer assumed to be correct, we introduce the pseudo-true parameter

$$\bar{\beta}\equiv\left(\sum_{\tau\in\mathcal{T}}g(\Delta Z_\tau)\,g(\Delta Z_\tau)^\top\right)^{-1}\left(\sum_{\tau\in\mathcal{T}}g(\Delta Z_\tau)\,\Delta Y_\tau\right).$$

Clearly, $\bar{\beta}$ coincides with β^* whenever ((2.4)) is correctly specified, but $\bar{\beta}$ remains well-defined even under misspecification. Formally, the testing problem is to decide in which of the following two sets the observed sample path falls[14]:

$$\begin{cases}\Omega_0\equiv\left\{\Delta Y_\tau=\bar{\beta}^\top g(\Delta Z_\tau)\ \text{for all}\ \tau\in\mathcal{T}\right\}\cap\{|\mathcal{P}|>q\},\\ \quad\text{(Null Hypothesis)}\\ \Omega_a\equiv\left\{\Delta Y_\tau\neq\bar{\beta}^\top g(\Delta Z_\tau)\ \text{some}\ \tau\in\mathcal{T}\right\}\cap\{|\mathcal{P}|>q\},\\ \quad\text{(Alternative Hypothesis).}\end{cases}\tag{3.11}$$

We note that the event $\{|\mathcal{P}|>q\}$ rules out the degenerate situation where (2.4) holds trivially (recall that q is the dimension of $g(\cdot)$). Like in the classical setting, this condition says that β^* is overidentified, so that a specification test is possible.

We carry out the test by examining whether the sum of squared residuals (SSR) of a linear regression is "close enough" to zero. The SSR statistic is given by

$$SSR_n\equiv\sum_{i\in\mathcal{J}_n}\left(\Delta_{i,k}^nY-g\left(\Delta_{i,k}^nZ\right)^\top\hat{\beta}_n\right)^2.\tag{3.12}$$

We reject the null hypothesis that (2.4) is correctly specified at significance level $\alpha\in(0,1)$ if SSR_n is greater than a critical value cv_n^α that is described in Algorithm 2 below. In practice, it may be useful to report the test in terms of the R^2 of the regression (2.9), that is,

$$R_n^2\equiv1-\frac{SSR_n}{\sum_{i\in\mathcal{J}_n}\Delta_{i,k}^nY^2}.$$

We reject the null hypothesis when $1-R_n^2$ is greater than $cv_n^\alpha/\sum_{i\in\mathcal{J}_n}(\Delta_{i,k}^nY)^2$.

Algorithm 2 (*Bootstrapping Critical Values for the Specification Test*)**.**
Step 1. Generate $(\Delta_{i,k}^nX^{*c})_{i\in\mathcal{J}_n}$ as in step 1 of Algorithm 1.
Step 2. Set $\Delta_{i,k}^nZ^*=\Delta_{i,k}^nZ+\Delta_{i,k}^nZ^{*c}$ and $\Delta_{i,k}^nY^*=\hat{\beta}_n^\top g(\Delta_{i,k}^nZ)+\Delta_{i,k}^nY^{*c}$ for $i\in\mathcal{J}_n$.
Step 3. Set cv_n^α to be the $(1-\alpha)$-quantile of SSR_n^* of the bootstrap sample, where SSR_n^* is the SSR obtained from regressing $\Delta_{i,k}^nY^*$ on $g(\Delta_{i,k}^nZ^*)$. □

Theorem 4. *Under Assumptions 1 and 2, the following statements hold.*

(a) *In restriction to* Ω_0, $\Delta_n^{-1}SSR_n$ *converges stably in law to*

$$\sum_{p\in\mathcal{P}}\varsigma_p^2-\left(\sum_{p\in\mathcal{P}}g\left(\Delta Z_{\tau_p}\right)\varsigma_p\right)^\top\left(\sum_{p\in\mathcal{P}}g\left(\Delta Z_{\tau_p}\right)g\left(\Delta Z_{\tau_p}\right)^\top\right)^{-1}\times\left(\sum_{p\in\mathcal{P}}g\left(\Delta Z_{\tau_p}\right)\varsigma_p\right).$$

In restriction to Ω_a, SSR_n *converges in probability to*

$$\sum_{p\in\mathcal{P}}\Delta Y_{\tau_p}^2-\left(\sum_{p\in\mathcal{P}}g\left(\Delta Z_{\tau_p}\right)\Delta Y_{\tau_p}\right)^\top\left(\sum_{p\in\mathcal{P}}g\left(\Delta Z_{\tau_p}\right)g\left(\Delta Z_{\tau_p}\right)^\top\right)^{-1}\times\left(\sum_{p\in\mathcal{P}}g\left(\Delta Z_{\tau_p}\right)\Delta Y_{\tau_p}\right).$$

elements. This convergence is weaker than the almost sure convergence of the $\mathcal{F}$-conditional law of A_n towards that of A, but is stronger than the stable convergence in law.

12 It is useful to note that the spot volatility estimates $\hat{c}_{n,i\pm}$ in Algorithm 1 can be taken differently from those used in the estimation of $\hat{\beta}_n(w)$. In particular, if these spot volatility estimates attain the optimal $\Delta_n^{-1/4}$ rate, then it can be shown that the $\mathcal{F}$-conditional distribution $\mathcal{L}_n^*(w)$ converges to that of $\mathcal{L}(w)$ under the uniform metric with rate $\Delta_n^{-1/4}$.

13 The type of refinement offered by the bootstrap is nevertheless nonstandard and theoretically interesting because our bootstrap is not applied to an asymptotically pivotal statistic, see Section 3.2 of Horowitz (2001) for a review of standard results on the asymptotic refinement of the bootstrap for asymptotically pivotal statistics. Instead, here, the refinement accounts for a higher-order sampling variability from the nonparametrically constructed weights (due to spot covariance estimation) that are used for efficient estimation.

14 Specifying hypotheses in terms of random events is unlike the classical setting of hypothesis testing (e.g., Lehmann and Romano (2005)), but is standard in the study of high frequency data; see Jacod and Protter (2012), and references and discussions therein.

(b) The test associated with the critical region $\{SSR_n > cv_n^\alpha\}$ *has asymptotic level* α *under the null hypothesis and asymptotic power one under the alternative hypothesis, that is,*

$$\mathbb{P}\left(SSR_n > cv_n^\alpha | \Omega_0\right) \to \alpha, \quad \mathbb{P}\left(SSR_n > cv_n^\alpha | \Omega_a\right) \to 1.$$

4. Simulation results

We now examine the asymptotic theory above in simulations that mimic our empirical setting in Section 5. We set the sample span $T = 1$ year, or equivalently, 250 trading days. Each day contains $m = 400$ high-frequency returns, roughly corresponding to 1-minute sampling. Each Monte Carlo sample contains $n = 100{,}000$ returns, which are expressed in annualized percentage terms. We set the fine scale $\Delta_n = 1/n$ and implement the mixed-scale jump regression at the coarse scale $k\Delta_n$, for $k = 3$, 5 and 10. While our main focus is on results with mixed scales, we also report results for $k = 1$ as a benchmark. There are 2000 Monte Carlo trials.

We consider a data generating process that allows for important features such as leverage effect and price-volatility co-jumps. For independent Brownian motions $W_{1,t}$, $W_{2,t}$, $B_{1,t}$ and $B_{2,t}$, we set

$$\begin{cases} d\log\left(V_{1,t}^*\right) = -\lambda_N \mu_F dt + 0.5\left(dB_{1,t} + J_{V,t} dN_t\right), \\ \quad V_{1,0}^* = \bar{V}_1, \\ \log\left(V_{2,t}^*\right) = \log\left(\bar{V}_2 - \beta_C^2 \bar{V}_1\right) + B_{2,t}, \\ V_{1,t} = TOD_t V_{1,t}^*, \quad V_{2,t} = TOD_t V_{2,t}^*, \\ dZ_t = \sqrt{V_{1,t}}\left(\rho dB_{1,t} + \sqrt{1-\rho^2} dW_{1,t}\right) + \varphi_{Z,t} dN_t, \\ dY_t = \beta_C \sqrt{V_{1,t}}\left(\rho dB_{1,t} + \sqrt{1-\rho^2} dW_{1,t}\right) \\ \quad + \sqrt{V_{2,t}} dW_{2,t} + \varphi_{Y,t} dN_t, \end{cases} \tag{4.1}$$

where TOD_t is a daily periodic function that captures the time-of-day effect in volatility.[15] The jump regression relationship is given by

$$\varphi_{Y,t} = \beta_J \varphi_{Z,t}, \tag{4.2}$$

and the parameters are, in annualized terms,

$$\begin{cases} \bar{V}_1 = 18^2, \quad \bar{V}_2 = 26^2, \quad \rho = -0.7, \quad \beta_C = 0.89, \quad \beta_J = 1, \\ J_{V,t} \overset{i.i.d.}{\sim} \text{Exponential}\left(\mu_F\right), \quad \mu_F = 0.1, \\ \varphi_{Z,t} | V_{1,t} \overset{i.i.d.}{\sim} \mathcal{N}\left(0, \dfrac{\phi^2 V_{1,t}}{n}\right), \quad \phi = 7.5, 10, \text{or } 12.5, \\ N_t \text{ is a Poisson process with intensity } \lambda_N = 20. \end{cases} \tag{4.3}$$

These parameters are calibrated to match some key features of the data used in Section 5. In particular, the signal-to-noise ratio parameter ϕ controls the relative size of price jumps with respect to that of the (local) 1-minute diffusive returns, which is about 10.3 in our empirical sample. The jump intensity λ_N is calibrated so that the average number of detected jumps in the simulation is close to what we observe in the data, which is about 10.6 jumps per year. When $\phi = 7.5$, 10 and 12.5, the average number of detected jumps in the simulation are 8.7, 11.9 and 14.2, respectively.

In order to examine the power of the specification test, we also implement the test under the following alternative model:

$$\varphi_{Y,t} = \beta_J \varphi_{Z,t} - \frac{\gamma}{\phi\sqrt{\bar{V}_1/n}} \varphi_{Z,t}^2 1_{\{\varphi_{Z,t}<0\}},$$

where the normalization via the average jump size $\phi\sqrt{\bar{V}_1/n}$ makes the interpretation of the parameter γ comparable across simulations. We note that the correctly specified model (4.2) corresponds to $\gamma = 0$. We generate misspecified models by setting $\gamma = 1$ or 2.

Tuning parameters are chosen as follows. We set $k_n = 60$, corresponding to a one-hour local window for spot covariance estimation. For each trading day $t \in \{1, \ldots, 250\}$, the truncation thresholds for Z are chosen adaptively as

$$u_{n,t} = 7m^{-0.49}\sqrt{BV_t}, \quad u'_{n,t} = 4m^{-0.49}\sqrt{BV_t}.$$

Here, BV_t is a slightly modified version of the bipower variation estimator of Barndorff-Nielsen and Shephard (2004b):

$$BV_t \equiv \frac{m}{m-4} \sum_i \left|\Delta_i^n Z\right| \left|\Delta_{i+1}^n Z\right|,$$

where the sum is over all returns in day t but with the largest 3 summands excluded.[16] The truncation threshold for Y is computed similarly. Finally, we use the procedure detailed in the supplemental material of Todorov and Tauchen (2012) to adjust for the time-of-day effect.

In Table 1, we report the finite-sample rejection rates of the specification test described in Theorem 4. Under the null hypothesis (i.e., $\gamma = 0$), we see that the rejection rates are fairly close to the nominal levels of the test across various jump sizes and mixed scales. Under the alternative model (i.e., $\gamma = 1$ or 2), the rejection rates are well above the nominal level. Not surprisingly, the finite-sample power decreases as we use coarser scale (i.e. larger k), but it is interesting to note that the drop of power from $k = 1$ to $k = 3$ is not severe. As ϕ and γ increase, the rejection rates approach one.

In Table 2, we report some summary statistics for the mixed-scale OLS and WLS estimators of the jump beta. We see that the WLS estimator is always more accurate than the OLS estimator as measured by the root mean squared error (RMSE). Moreover, the coverage rates of confidence intervals (CI) constructed using Algorithm 1 and the percentile bootstrap are generally very close to the nominal levels, regardless of the sampling scale and the jump size. Coverage results based on the basic bootstrap are very similar to the percentile bootstrap and, hence, are omitted for brevity.

5. Empirical application

We use the developed tools to study the systematic jump risk in the stocks comprising the Dow Jones Industrial Average Index as of December 2014, except Visa Inc. (V) is replaced by Bank of America (BAC) to make a balanced panel covering the period January 3, 2007 to December 12, 2014. The proxy for the market is the front-month E-mini S&P 500 index futures contract, which is among the most liquid instruments in the world.[17] In some of our analysis we also make use of the ETFs on the nine industry portfolios comprising the S&P 500 index. We remove market holidays and half trading days. We also remove the two "Flash Crashes" (May 6, 2010 and April 23, 2013) because the dramatic market fluctuations in these days are known to be due to market malfunctioning. The resultant sample contains 1982 trading days. The intraday observations are sampled at 1-minute frequency from 9:35 to 15:55 EST; the prices at the first and the last 5 min are discarded to guard against possible adverse microstructure effects at market open and close. Finally, the truncation and the window size for the local volatility estimation are set as in the Monte Carlo.

[15] The time-of-day effect is calibrated using the data in our empirical study by averaging across days for each fixed sampling interval within a day.

[16] In empirical applications, there may be large consecutive returns with similar magnitude but opposite signs (i.e., bouncebacks). The bipower variation estimator is sensitive to such issues. Removing the largest three summands is a simple but effective finite-sample robustification in this respect.

[17] Hasbrouck (2003) estimates that 90% of U.S. equity price formation takes place in the E-mini market futures market.

Table 1
Monte Carlo rejection rates of specification tests.

		$\gamma = 0$			$\gamma = 1$			$\gamma = 2$		
		1%	5%	10%	1%	5%	10%	1%	5%	10%
$\phi = 7.5$	$k = 1$	1.2	4.3	9.8	90.1	93.5	95.2	96.6	97.7	98.5
	$k = 3$	1.5	5.3	9.7	75.9	83.5	87.0	92.3	94.5	95.8
	$k = 5$	1.3	5.0	10.2	64.3	74.6	79.8	87.9	91.5	93.9
	$k = 10$	1.3	5.1	9.7	46.1	59.9	67.3	75.3	82.8	87.2
$\phi = 10$	$k = 1$	0.9	5.1	10.0	96.0	97.9	98.6	99.1	99.4	99.4
	$k = 3$	1.5	5.5	9.8	86.9	91.8	93.7	97.5	98.4	98.8
	$k = 5$	1.3	5.9	10.9	80.1	87.0	89.8	94.8	96.9	97.7
	$k = 10$	1.5	6.3	10.4	64.1	75.3	80.3	87.1	92.3	94.3
$\phi = 12.5$	$k = 1$	1.3	5.5	10.4	98.2	99.1	99.3	99.2	99.4	99.6
	$k = 3$	1.4	5.6	10.4	93.3	95.5	96.8	97.8	98.3	98.8
	$k = 5$	1.6	5.7	10.1	87.6	92.1	94.3	96.4	97.5	98.0
	$k = 10$	1.4	5.8	10.6	74.2	83.6	86.7	91.6	94.4	96.1

Note: We report the Monte Carlo rejection rates of the specification test at significance level 1%, 5% and 10%. We report results under the null hypothesis ($\gamma = 0$) and the alternative hypothesis ($\gamma = 1, 2$) for various mixed scales ($k = 1$, 3, 5 and 10) as well as various relative jump sizes ($\phi = 7.5$, 10 and 12.5). The inference is based on 1000 bootstrap draws. Each experiment has 2000 Monte Carlo trials.

Table 2
Summary of estimation and coverage results.

		Mixed-Scale OLS				Mixed-Scale WLS			
		RMSE	99% CI	95% CI	90% CI	RMSE	99% CI	95% CI	90% CI
$\phi = 7.5$	$k = 1$	0.063	98.5	93.7	88.1	0.057	98.9	94.4	88.7
	$k = 3$	0.111	98.2	92.9	88.4	0.101	98.3	92.7	88.8
	$k = 5$	0.143	98.6	94.6	88.7	0.129	98.5	94.0	88.7
	$k = 10$	0.194	98.5	94.9	89.3	0.182	98.6	93.9	88.9
$\phi = 10$	$k = 1$	0.045	98.7	94.1	88.3	0.039	98.5	94.6	89.2
	$k = 3$	0.075	98.7	94.2	88.8	0.065	98.5	94.6	89.5
	$k = 5$	0.097	98.0	93.5	88.6	0.084	99.0	95.0	88.4
	$k = 10$	0.131	98.9	95.1	91.0	0.116	98.9	95.6	90.4
$\phi = 12.5$	$k = 1$	0.034	98.8	94.6	88.6	0.029	98.9	94.8	89.3
	$k = 3$	0.062	98.7	93.0	87.9	0.051	98.7	94.8	89.2
	$k = 5$	0.080	98.2	94.0	88.2	0.067	98.6	94.0	89.0
	$k = 10$	0.110	98.5	93.0	88.2	0.095	98.8	93.6	88.3

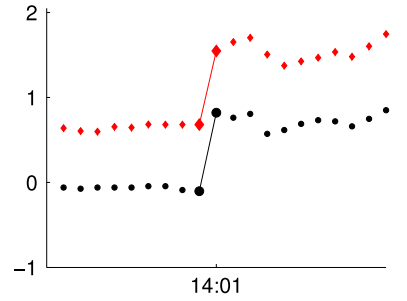
Note: We report the root mean squared error (RMSE) and the Monte Carlo coverage rates of confidence intervals (CI) at levels 99%, 95% and 90%. We report results for various mixed scales ($k = 1$, 3, 5 and 10) and relative jump sizes ($\phi = 7.5$, 10 and 12.5) for both mixed-scale OLS and WLS estimators. The CIs are constructed using Algorithm 1 and the percentile bootstrap based on 1000 bootstrap draws. Each experiment has 2000 Monte Carlo trials.

For this choice of the truncation (corresponding to a move slightly higher than 7 standard deviations), we detect a total of 85 market jumps in our sample.

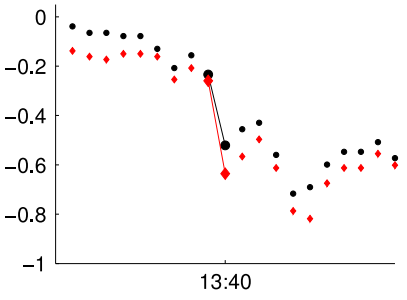
To gauge the importance of microstructure noise that is weakly dependent on time, we compare the average value of realized volatility at 1-minute (our sampling frequency) and the coarser sampling frequency of 5-minutes. In presence of weakly dependent noise, the realized volatility should be higher for the higher sampling frequency due to this type of noise. For our data set, the median value of the ratio of 1-minute realized volatility over 5-minute realized volatility is 1.08. This indicates a relatively modest impact of the noise at the frequency we use here for the local volatility estimation around the jump times.

We start our empirical analysis with illustrating how stocks react to market jumps using four representative market jumps in our sample (two positive and two negative). In Fig. 1, we plot the prices of the market and a set of selected stocks before and after the market jump event. The top left panel shows the behavior around the market jump on September 18, 2013 which was associated with the (positive) surprise by the Fed of not tapering its QE programs. In this case, both the market and the BA stock reacted within the same minute and fully adjusted to their new higher levels. A similar example, but in the opposite direction, is illustrated on the top right panel of the figure. This panel plots the market and AXP prices around the market jump on August 5, 2014. There were growing fears on this day associated with the impact of geopolitical risks on the economy along with concerns among investors that the Fed might raise interest rates sooner than expected in the wake of signs that the economy is gaining strength. In this example, like the previous one, both the market and the stock adjust to their lower level within the minute. Our third example of a market jump is of October 1, 2008 in the midst of the recent financial crisis. In this case, the CVX stock appeared to take more than one minute to fully incorporate the positive market jump, a seemingly delayed reaction which could be driven by market microstructure issues (e.g., stale limit orders). Another example of this type is the reaction of the WMT stock to the market jump on February 23, 2010 which is displayed on the bottom right panel of Fig. 1. This market jump was associated with a surprisingly weak consumer confidence index reflecting the pessimism among investors for the strength of the economic recovery. While the market reacted within the minute of the release of the negative consumer confidence data, the WMT stock took at least 2 min to fully incorporate the bad news.

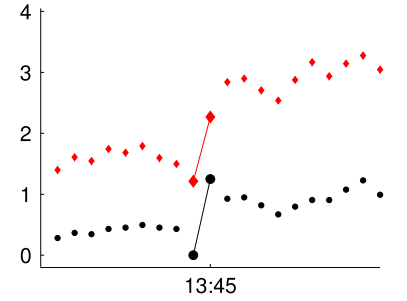
Overall, the above four examples suggest that, in general, the stocks in our sample react quickly to the news triggering the market jumps. However, in some instances staleness and lower liquidity can confound the reaction of stocks to the market jumps. These market microstructure related issues, however, seem to be fairly short-lived. To verify that this is indeed the case, in Table 3, we report the jump beta estimates for all the stocks in the sample using aggregation of 3 and 5 minutes for the beta estimation



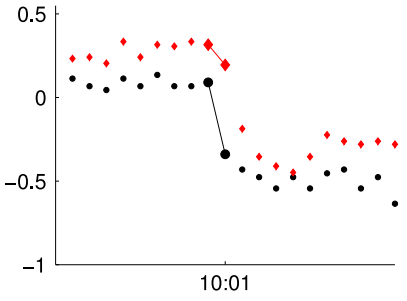
(a) BA: Sep/18/2013.



(b) AXP: Aug/05/2014.



(c) CVX: Oct/01/2008.



(d) WMT: Feb/23/2010.

Fig. 1. Market jump events and stock reaction. Circle and diamond dots correspond to cumulative (over the day) 1-minute log-returns on market and stock respectively. The connected dots correspond to the minute interval in which a market jump is detected.

(and using the whole sample). In the absence of confounding market microstructure effects, the two beta estimates should not be statistically different from each other. The results of the table show that this is largely the case. Indeed, the two beta estimates are fairly close with the median difference between the 5-minute and 3-minute estimates being only 0.01. The largest difference of 0.14 in our data set is for the DD stock, and this difference is only marginally statistically significant. Given this evidence, for the results that follow we will focus attention on the beta estimates based on three minute aggregation of returns following the market jump.

Table 3
Full sample WLS beta estimates.

Ticker	β $k=3$	95% CI	β $k=5$	95% CI
AXP	1.15	[1.08; 1.20]	1.17	[1.09; 1.22]
BA	0.99	[0.92; 1.03]	1.02	[0.94; 1.07]
BAC	1.36	[1.27; 1.43]	1.36	[1.25; 1.44]
CAT	1.06	[0.99; 1.11]	1.08	[1.00; 1.13]
CSCO	0.84	[0.77; 0.90]	0.89	[0.82; 0.97]
CVX	0.99	[0.94; 1.03]	0.98	[0.92; 1.02]
DD	1.09	[1.03; 1.14]	1.23	[1.15; 1.27]
DIS	0.97	[0.92; 1.01]	0.98	[0.91; 1.02]
GE	1.16	[1.09; 1.21]	1.17	[1.08; 1.23]
GS	1.20	[1.12; 1.25]	1.21	[1.11; 1.27]
HD	1.05	[0.98; 1.09]	1.07	[0.99; 1.12]
IBM	0.81	[0.76; 0.84]	0.81	[0.75; 0.84]
INTC	0.88	[0.81; 0.94]	0.93	[0.85; 0.99]
JNJ	0.67	[0.62; 0.70]	0.67	[0.62; 0.70]
JPM	1.31	[1.24; 1.37]	1.29	[1.20; 1.34]
KO	0.70	[0.65; 0.74]	0.66	[0.60; 0.70]
MCD	0.51	[0.47; 0.54]	0.50	[0.46; 0.54]
MMM	1.00	[0.95; 1.03]	1.04	[0.97; 1.07]
MRK	0.94	[0.87; 0.97]	0.91	[0.84; 0.95]
MSFT	0.81	[0.75; 0.85]	0.81	[0.75; 0.87]
NKE	0.78	[0.72; 0.83]	0.82	[0.75; 0.87]
PFE	0.87	[0.80; 0.92]	0.88	[0.80; 0.94]
PG	0.68	[0.63; 0.71]	0.65	[0.59; 0.68]
T	0.84	[0.78; 0.88]	0.82	[0.75; 0.86]
TRV	0.94	[0.88; 0.97]	0.86	[0.79; 0.90]
UNH	0.86	[0.80; 0.91]	0.92	[0.84; 0.97]
UTX	1.02	[0.96; 1.05]	1.04	[0.97; 1.08]
VZ	0.72	[0.67; 0.76]	0.71	[0.65; 0.76]
WMT	0.64	[0.59; 0.67]	0.62	[0.57; 0.66]
XOM	0.98	[0.93; 1.02]	0.97	[0.91; 1.01]

Note: We report the efficient k-mixed-scale ($k = 3$ or 5) WLS estimates and their 95% confidence intervals (CI) of the 30 Dow stocks over the full sample. The CIs are from the percentile bootstrap using 1000 draws.

In Fig. 2, we present scatter plots of stock jumps versus market jumps along with the fit implied by a constant market jump beta model for the whole sample. Overall, we see a very good fit. Most of the jump observations are fairly close to the fit implied by the constant jump beta model. Nevertheless, for some of the stocks, particularly those in the financial sector (bottom row), we see somewhat non-trivial deviations from the linear jump regression model. Of course, this can be merely due to the temporal variation in betas. In terms of levels, the jump betas of the stocks in the banking industry are systematically above one, while those of the consumer and healthcare sectors like MCD, WMT, JNJ and PG, are significantly below one.

Given the overwhelming prior evidence on time variation in market betas, we next present results from testing for constancy of market jump betas over periods of years. The results are reported in the top panel of Table 4. We conduct the test over different aggregation frequencies ranging from one minute (no aggregation) to ten minutes. Naturally, more aggregation leads to diminishing power of detecting the variation in jump betas. This is consistent with our Monte Carlo results reported in the previous section. However, the drop of rejection rates going from one to three minutes evident from Table 4 is too big to be solely explained by the statistical effect of losing power when aggregating returns for the jump beta estimation. Instead, the relatively high rejection rates of the test for one minute aggregation are likely due to market microstructure effects like the ones illustrated on the bottom panels of Fig. 1. At the three minute aggregation level, the rejection rates of the test are relatively low except for years 2007, 2008 and 2013. Some of these rejections can be still due to microstructure issues. However, some of the rejections probably reflect genuine variation of market jump betas, particularly during the period of the recent global financial crisis.

Please cite this article in press as: Li J., et al., Mixed-scale jump regressions with bootstrap inference. Journal of Econometrics (2017), http://dx.doi.org/10.1016/j.jeconom.2017.08.017.

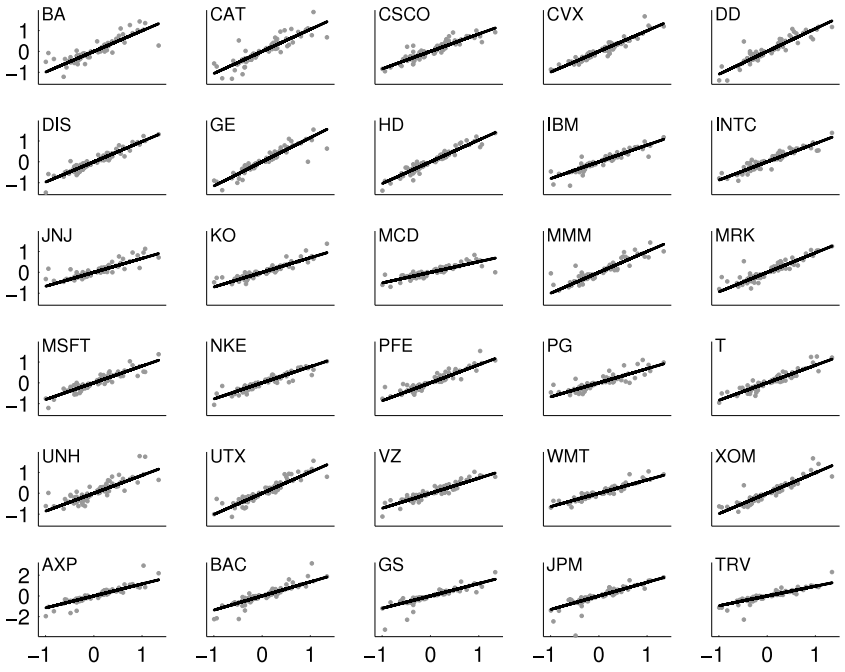


Fig. 2. Scatter of stock versus market returns at market jump times of the full sample.

Table 4
Specification testing results for 30 DJIA stocks.

	2007	2008	2009	2010	2011	2012	2013	2014
	Cross-Sectional rejection rate							
$k = 1$	0.80	0.77	0.77	0.50	0.50	0.93	0.70	0.77
$k = 3$	0.50	0.43	0.07	0.17	0.10	0.40	0.33	0.17
$k = 5$	0.27	0.03	0.00	0.10	0.17	0.40	0.10	0.10
$k = 10$	0.17	0.10	0.10	0.00	0.00	0.17	0.07	0.03
	Cross-Sectional median of R^2							
$k = 1$	0.90	0.90	0.93	0.93	0.97	0.90	0.96	0.91
$k = 3$	0.86	0.80	0.87	0.92	0.95	0.88	0.95	0.86
$k = 5$	0.83	0.75	0.87	0.93	0.91	0.88	0.97	0.85
$k = 10$	0.82	0.81	0.93	0.86	0.87	0.82	0.97	0.81

Note: On the top panel, we report the cross-sectional rejection rate of the specification test at 1% significance level for the k-mixed samples year-by-year. On the bottom panel, we report the cross-sectional median of the R^2s of the 30 stocks for the k-mixed samples.

To further gauge the performance of the year-by-year linear jump regression model, the second panel of Table 4 reports the R^2 of the model fit at the market jump events. As seen from the table, the R^2 numbers are generally very high. For example, the time series average of R^2 at one- and three-minute aggregation are 0.93 and 0.89 respectively. As expected from theory, when increasing aggregation from one minute to ten minutes, the R^2 drops because the volatility of the diffusive aggregated increments around the jumps increases. Nevertheless, we see that with the exception of year 2008, the loss of R^2 going from one-minute to three-minute aggregation is quite moderate. Comparing the two panels of Table 4, we notice that there is no direct correspondence between the rejection rates and the magnitude of the R^2 of the linear jump regression. For example, focusing at the three-minute aggregation results, we can see that year 2007 is associated with the highest rejection rate of the linear jump regression model and yet has relatively high R^2. On the other hand, year 2008 is associated with high rejection rate and has the lowest R^2 in the sample (using again the three-minute aggregation results). This difference can be explained with the different magnitude of the volatility around the jump event: it is relatively higher in 2008 than in 2007 and, as a result, the inference in the latter is sharper than the former.

To better assess the time variation in market jump betas, we plot next yearly jump betas in Fig. 3. There are some clearly distinguishable time-series patterns evident from the figure. For example, the market jump betas of stocks in the financial sector, such as AXP, BAC and JPM, increase in the first two years in our sample and gradually decrease afterwards. On the other hand, stocks such as INTC and WMT exhibit very little time variation.

The analysis so far has been based at the market jump times. We next investigate how stocks react to other systematic jump events. In particular, we focus attention on jump events in the nine industries comprising the S&P 500 index (our proxy for the market index) which are not detected as market jump events. In a market jump model in which the jumps in stocks are of two types, idiosyncratic and market, aggregate portfolios, such as the nine industry portfolios, should contain only jumps at the times when the market jumps (as the idiosyncratic jumps get diversified away). However, some systematic jump events can have much bigger impact on a particular industry sector than on the market as a whole and, hence, the magnitude of an industry jump can be much bigger than that of the market co-jump. In such instances, given our discrete setting and high truncation level, we can fail to detect such jump events on the market level but still find them in a particular industry sector portfolio. Hence, we label jump events in

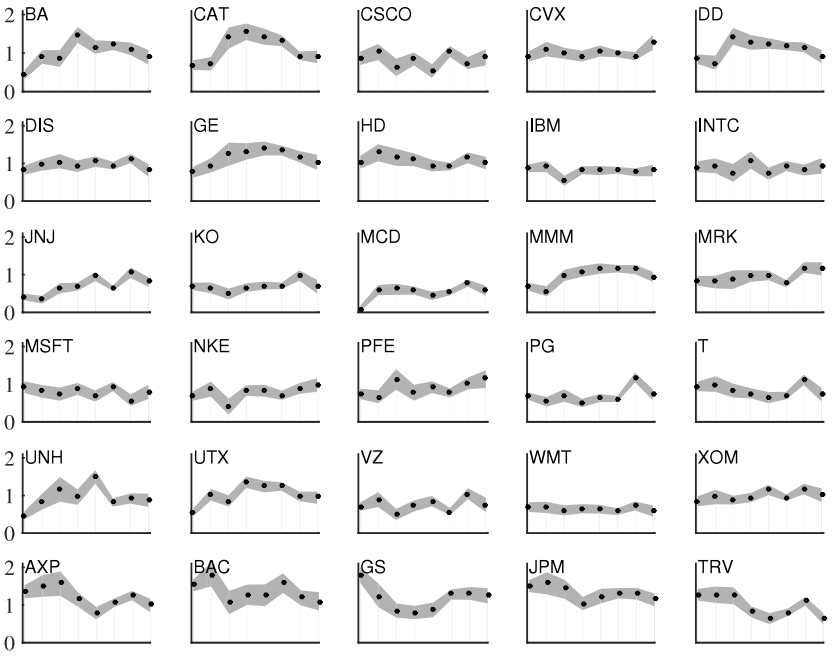


Fig. 3. Time series of yearly jump betas, 2007–2014. The dots correspond to the yearly WLS beta estimates and the shaded areas correspond to the associated 95% confidence intervals.

Table 5
R^2 of the market factor for two types of jumps.

R^2 of Market-wide jumps					
AXP	0.81	HD	0.92	NKE	0.84
BA	0.75	IBM	0.84	PFE	0.84
BAC	0.81	INTC	0.87	PG	0.74
CAT	0.77	JNJ	0.73	T	0.87
CSCO	0.82	JPM	0.70	TRV	0.78
CVX	0.90	KO	0.83	UNH	0.71
DD	0.84	MCD	0.72	UTX	0.85
DIS	0.89	MMM	0.81	VZ	0.85
GE	0.84	MRK	0.81	WMT	0.87
GS	0.72	MSFT	0.84	XOM	0.85
R^2 of Sector-specific jumps					
AXP	0.45	HD	0.35	NKE	0.39
BA	0.33	IBM	0.22	PFE	0.46
BAC	0.75	INTC	0.48	PG	0.61
CAT	0.46	JNJ	0.50	T	0.53
CSCO	0.72	JPM	0.83	TRV	0.42
CVX	0.29	KO	0.31	UNH	0.21
DD	0.32	MCD	0.68	UTX	0.52
DIS	0.49	MMM	0.62	VZ	0.58
GE	0.46	MRK	0.32	WMT	0.29
GS	0.45	MSFT	0.78	XOM	0.33

an industry sector, which are not detected as market jump times, as sector-specific jumps. These jumps have relatively much bigger importance for the particular sector than for the market.

To study the reaction of stocks to sector-specific jump events, we first associate with each of the stocks in our analysis the industry sector it belongs to.[18] In Table 5 we report the R^2 for a linear jump regression model of the stock jump against the market jump at the sector-specific jump events for each of the stocks based on the whole sample. For comparison we also report the corresponding R^2 for the linear market jump model at the market jump times. The results present a rather mixed picture for the performance of the linear market jump model at the sector-specific jump events. For some stocks such as BAC, JPM, MCD and MSFT, the performance of the linear regression at the sector-specific jump events in terms of R^2 is comparable to its performance at the market jump events. However, for stocks like CVX, IBM, WMT and XOM, the R^2 of the regression at the sector-specific jump times is very low. Some of the loss of fit when comparing the performance of the linear jump model at market-wide jump events and sector-specific jump events can be due to the "signal" being smaller, that is, the market jump size at the sector-specific jump events being smaller in absolute value. This, however, cannot be the sole explanation, since as explained above, for some of the stocks in our sample the drop in R^2 is quite small. Another reason for the worsening fit at the sector-specific jump events can be due to larger "noise", i.e., the diffusive volatility around the sector-specific jump events can be much bigger than around market-wide jump events for some of the stocks. Yet a third reason can be that the linear market jump model does not hold at the sector-specific jump events.

To get further insight into the performance of the linear market jump model at the sector-specific jump events, we display in Fig. 4 scatter plots of the stock jumps against the market jumps at the sector-specific jump events for four representative (in terms of R^2) stocks. As seen from the figure, the performance of the model for IBM is very good with the observations being very close to the linear fit. On the other hand, for GE we notice that the jump observations are much more dispersed around the linear fit. This is suggestive of larger diffusive volatility around the sector-specific jump events for GE which consequently lowers the R^2 of the regression. Similar reasoning can explain the low R^2 for XOM. For this stock, however, we can also notice a few outliers in the lower left corner of the plot which are indicative of model failure, i.e., that the market jumps cannot solely explain the XOM jumps at the sector-specific jump events. Finally, the fit for WMT is fairly poor with no strong association between the stock and market jumps at the sector-specific jump events. This is in sharp contrast with the performance of the linear jump market model for this stock at the market jump events.

Overall, we can conclude that for some stocks the linear market jump model continues to work well at the sector-specific jump

[18] The stocks in our study are all part of the S&P 500 index during the sample period. We, therefore, use the industry classification that is used to split the stocks in the S&P 500 index into nine industry portfolio ETFs.

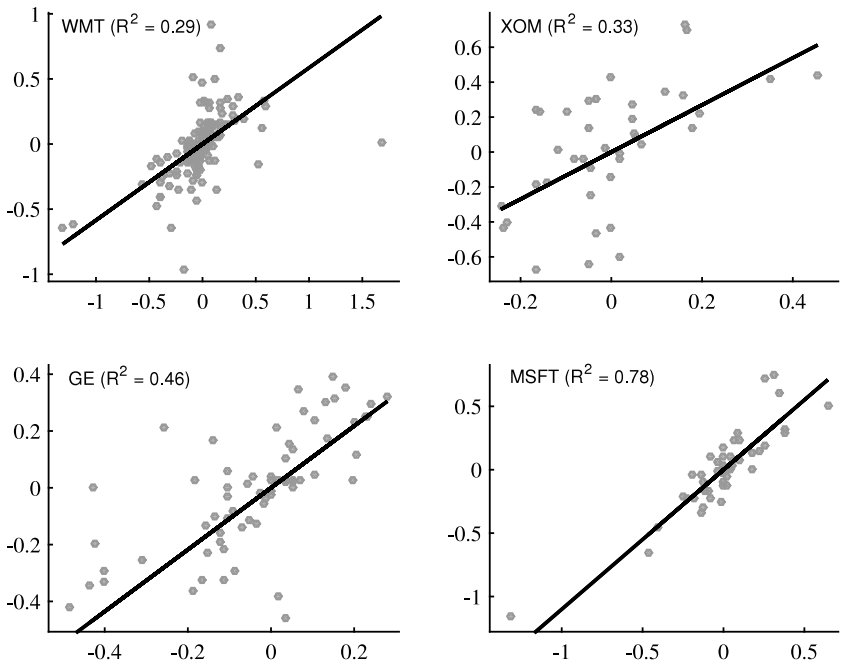


Fig. 4. Scatter of stock versus market returns at sector-specific jump times.

events. For many of the stocks, however, this is also associated with increased diffusive volatility around the sector-specific jump events which makes inference for the market jump beta at these events much noisier when compared with inference conducted at the market-wide jump events. Finally, for some of the stocks, the linear market jump model fails to account for behavior of the stock market jumps at the sector-specific jump events and other factors are probably needed.

6. Conclusion

We propose a new mixed-scale jump regression framework for studying deterministic dependencies among jumps in a multivariate setting. A fine time scale is used to identify with high accuracy the times of large rare jumps in the explanatory variable(s). A coarser scale is then used to conduct the estimation in order to attenuate the effects of trading friction noise. We derive the asymptotic properties of an efficient estimator of the jump regression coefficients and a test for its specification. The limiting distributions of the estimator and the test statistic are non-standard, but a simple bootstrap method is shown to be valid for feasible inference. We further show that the bootstrap provides a higher-order refinement that accounts for the sampling variation in spot covariance estimates which are used to construct the efficient estimator. In a realistically calibrated Monte Carlo setting, which features leverage effects and price-volatility co-jumps, we report good size and power properties of the general specification test and good coverage properties of the confidence intervals.

The empirical application employs a 1-minute panel of Dow stock prices together with the front-month E-mini S&P 500 stock market index futures over the period 2007–2014. The 1-minute market index is used to locate jump times, and subsequent 3-minute sampling around the jump times is used to conduct the jump regression. We find a strong relationship between market jumps and stock price moves at market jump times. The market jump betas exhibit remarkable temporal stability and the jump regressions have very high observed R^2s. On the other hand, for many of the stocks in the sample, the relationship between stock and market jumps at sector-specific jump times is significantly noisier, and temporally more unstable, than the tight relationship seen at market jump times.

Appendix. Proofs

We now prove the theorems in the main text. Throughout this appendix, we use K to denote a generic positive constant that may change from line to line; we sometimes emphasize the dependence of this constant on some parameter q by writing K_q. Below, the convergence $(\xi_{n,p})_{p\geq 1} \to (\xi_p)_{p\geq 1}$, as $n \to \infty$, is understood to be under the product topology. We write w.p.a.1 for "with probability approaching one." Recall that $(\tau_p)_{p\geq 1}$ are the successive jump times of Z and $\mathcal{P} = \{p \geq 1 : \tau_p \in [0, T]\}$. We use $i(p)$ to denote the unique integer such that $\tau_p \in ((i-1)\Delta_n, i\Delta_n]$.

By a standard localization procedure (see Section 4.4.1 of Jacod and Protter (2012)), we can respectively strengthen Assumptions 1 and 3 to the following stronger versions without loss of generality.

Assumption 4. We have Assumption 1. The processes X_t, b_t and σ_t are bounded.

Assumption 5. We have Assumption 3. The processes $\tilde{b}_t$ and $\tilde{\sigma}_t$ are bounded. Moreover, there exists some bounded λ-integrable function $\tilde{J}$ such that $\|\tilde{\delta}(t,u)\|^2 \leq \tilde{J}(u)$ for all $t \in [0,T]$ and $u \in \mathbb{R}$.

A.1. Proof of Theorem 1

Proof of Theorem 1. By Proposition 1 of Li et al. (2017),

$$\mathbb{P}\left(\mathcal{J}_n = \bar{\mathcal{C}}_n = \mathcal{J}_n^*\right) \to 1, \tag{A.1}$$

where $\bar{\mathcal{C}}_n$ is the complement of $\mathcal{C}_n$. Since the jumps of X have finite activity, differences between distinct indices in $\mathcal{J}_n$ can be bounded below by $2k_n$ w.p.a.1 and, hence,

$$\begin{cases} \hat{c}_{n,i(p)-} = \dfrac{1}{k_n\Delta_n}\displaystyle\sum_{j=0}^{k_n-1}(\Delta^n_{i(p)-k_n+j}X)(\Delta^n_{i(p)-k_n+j}X)^\top 1_{\{i(p)-k_n+j\in\mathcal{C}_n\}}, \\ \hat{c}_{n,i(p)+} = \dfrac{1}{k_n\Delta_n}\displaystyle\sum_{j=0}^{k_n-1}(\Delta^n_{i(p)+k+j}X)(\Delta^n_{i(p)+k+j}X)^\top 1_{\{i(p)+k+j\in\mathcal{C}_n\}}. \end{cases}$$

Then, by Theorem 9.3.2 in Jacod and Protter (2012) and $\Delta^n_{i(p),k}Z \longrightarrow \Delta Z_{\tau_p}$, we derive

$$\left(\hat{c}^n_{i(p)-}, \hat{c}^n_{i(p)+}, \Delta^n_{i(p),k}Z\right) \xrightarrow{\mathbb{P}} \left(c_{\tau_p-}, c_{\tau_p}, \Delta Z_{\tau_p}\right).$$

By (A.1), we see that the following holds w.p.a.1,

$$\begin{aligned}\hat{\beta}_n(w) - \beta^* &= \left(\sum_{p\in\mathcal{P}} \hat{w}_{n,i(p)} g(\Delta^n_{i(p),k}Z) g(\Delta^n_{i(p),k}Z)^\top\right)^{-1} \\ &\times \left(\sum_{p\in\mathcal{P}} \hat{w}_{n,i(p)} g(\Delta^n_{i(p),k}Z) \varepsilon^n_{i(p),k}\right).\end{aligned} \tag{A.2}$$

By Assumption 2, $g(\cdot)$ is continuously differentiable at ΔZ_{τ_p} almost surely. Then, noting that $\Delta^n_{i(p),k}Z^c = O_p(\Delta_n^{1/2})$, we use a second-order Taylor expansion to deduce

$$\begin{aligned}\Delta_n^{-1/2}\varepsilon^n_{i(p),k} &= \Delta_n^{-1/2}\left(\Delta^n_{i(p),k}Y^c - \beta^{*\top}\left(g(\Delta Z_{\tau_p} + \Delta^n_{i(p),k}Z^c)\right.\right. \\ &\quad \left.\left. - g(\Delta Z_{\tau_p})\right)\right) \\ &= \Delta_n^{-1/2}\left(\Delta^n_{i(p),k}Y^c - \beta^{*\top}\partial g\left(\Delta Z_{\tau_p}\right)\Delta^n_{i(p),k}Z^c\right) \\ &\quad + O_p(\Delta_n^{1/2}).\end{aligned} \tag{A.3}$$

By a straightforward adaptation of Proposition 4.4.10 of Jacod and Protter (2012), we have

$$\Delta_n^{-1/2}\left(\Delta^n_{i(p),k}X^c\right)_{p\geq 1} \xrightarrow{\mathcal{L}\text{-}s} \left(\sqrt{\kappa_p}\sigma_{\tau_p-}\xi_{p-} + \sqrt{k-\kappa_p}\sigma_{\tau_p}\xi_{p+}\right)_{p\geq 1} \tag{A.4}$$

By (A.3) and (A.4),

$$\Delta_n^{-1/2}\left(\varepsilon^n_{i(p),k}\right)_{p\geq 1} \xrightarrow{\mathcal{L}\text{-}s} \left(\varsigma_p\right)_{p\geq 1}. \tag{A.5}$$

From (A.2) and (A.5), it is easy to see that $\hat{\beta}_n \xrightarrow{\mathbb{P}} \beta^*$. Hence, $\hat{w}_{n,i(p)} \xrightarrow{\mathbb{P}} w_p$ for each $p \geq 1$. From here, as well as (A.2) and (A.5), we derive

$$\Delta_n^{-1/2}\left(\hat{\beta}_n(w) - \beta^*\right) \xrightarrow{\mathcal{L}\text{-}s} \Xi(w)^{-1}\Lambda(w)$$

as asserted. □

A.2. Proof of Theorem 2

Proof of Theorem 2. (a) By Theorem 13.3.3 in Jacod and Protter (2012), we have

$$\left(k_n^{1/2}(\hat{c}_{n,i(p)-} - c_{\tau_p-}), k_n^{1/2}(\hat{c}_{n,i(p)+} - c_{\tau_p})\right)_{p\geq 1} \xrightarrow{\mathcal{L}\text{-}s} \left(\zeta_{p-}, \zeta_{p+}\right)_{p\geq 1}. \tag{A.6}$$

This convergence indeed holds jointly with (A.5) because $\Delta^n_{i,k}X^c$ and $(\hat{c}_{n,i-}, \hat{c}_{n,i+})$ involve non-overlapping returns. Let

$$\tilde{\varepsilon}^n_{i(p),k} \equiv \Delta^n_{i(p),k}Y^c - \beta^{*\top}\partial g\left(\Delta Z_{\tau_p}\right)\Delta^n_{i(p),k}Z^c.$$

From (A.3), we see that

$$\Delta_n^{-1/2}\varepsilon^n_{i(p),k} - \Delta_n^{-1/2}\tilde{\varepsilon}^n_{i(p),k} = O_p(\Delta_n^{1/2}) = o_p(k_n^{-1/2}).$$

Recall that $\dot{w}(\cdot)$ is the first differential of $w(\cdot)$, that is, as $\left(c'_-, c'_+, z', b'\right) \to 0$,

$$\begin{aligned}&w\left(c_- + c'_-, c_+ + c'_+, z + z', b + b'\right) - w(c_-, c_+, z, b) \\ &\quad = \dot{w}\left(c_-, c_+, z, b; c'_-, c'_+, z', b'\right) \\ &\qquad + o\left(\left\|c'_-\right\| + \left\|c'_+\right\| + \left\|z'\right\| + \left\|b'\right\|\right).\end{aligned}$$

Recall that $\tilde{w}_p \equiv \dot{w}\left(c_{\tau_p-}, c_{\tau_p+}, \Delta Z_{\tau_p}, \beta^*; \zeta_{p-}, \zeta_{p+}, 0, 0\right)$. By the delta method, we derive from (A.5) and (A.6), $\Delta^n_{i(p),k}Z = \Delta Z_{\tau_p} + O_p(\Delta_n^{1/2})$ and $\hat{\beta}_n = \beta^* + O_p(\Delta_n^{1/2})$ that

$$\left(k_n^{1/2}\left(\hat{w}_{n,i(p)} - w_p\right), \Delta_n^{-1/2}\tilde{\varepsilon}^n_{i(p),k}\right)_{p\geq 1} \xrightarrow{\mathcal{L}\text{-}s} \left(\tilde{w}_p, \varsigma_p\right)_{p\geq 1}. \tag{A.7}$$

We now note that, w.p.a.1,

$$\begin{aligned}&\Delta_n^{-1/2}\left(\hat{\beta}_n(w) - \beta^*\right) \\ &= \left(\sum_{p\in\mathcal{P}} \hat{w}_{n,i(p)} g(\Delta^n_{i(p),k}Z) g(\Delta^n_{i(p),k}Z)^\top\right)^{-1} \\ &\quad \times \left(\sum_{p\in\mathcal{P}} \hat{w}_{n,i(p)} g(\Delta^n_{i(p),k}Z)\Delta_n^{-1/2}\varepsilon^n_{i(p),k}\right) \\ &= \left(\Xi(w) + \sum_{p\in\mathcal{P}}\left(\hat{w}_{n,i(p)} - w_p\right) g\left(\Delta Z_{\tau_p}\right) g\left(\Delta Z_{\tau_p}\right)^\top \right. \\ &\quad \left. + o_p(k_n^{-1/2})\right)^{-1} \\ &\quad \times \left(\sum_{p\in\mathcal{P}} w_p g\left(\Delta Z_{\tau_p}\right)\Delta_n^{-1/2}\tilde{\varepsilon}^n_{i(p),k}\right. \\ &\quad \left. + \sum_{p\in\mathcal{P}}\left(\hat{w}_{n,i(p)} - w_p\right) g\left(\Delta Z_{\tau_p}\right)\Delta_n^{-1/2}\tilde{\varepsilon}^n_{i(p),k} + o_p(k_n^{-1/2})\right) \\ &= \mathcal{L}_n(w) + k_n^{-1/2}\mathcal{H}_n(w) + o_p(k_n^{-1/2}),\end{aligned}$$

where

$$\begin{cases}\mathcal{L}_n(w) \equiv \Xi(w)^{-1}\sum_{p\in\mathcal{P}} w_p g\left(\Delta Z_{\tau_p}\right)\Delta_n^{-1/2}\tilde{\varepsilon}^n_{i(p),k}, \\ \mathcal{H}_n(w) \equiv \Xi(w)^{-1}\sum_{p\in\mathcal{P}} k_n^{1/2}\left(\hat{w}_{n,i(p)}\right. \\ \quad \left. - w_p\right) g\left(\Delta Z_{\tau_p}\right)\Delta_n^{-1/2}\tilde{\varepsilon}^n_{i(p),k} \\ \quad - \Xi(w)^{-1}\left(\sum_{p\in\mathcal{P}} k_n^{1/2}\left(\hat{w}_{n,i(p)} - w_p\right) g\left(\Delta Z_{\tau_p}\right) g\left(\Delta Z_{\tau_p}\right)^\top\right) \\ \quad \times \Xi(w)^{-1} \\ \quad \times \left(\sum_{p\in\mathcal{P}} w_p g\left(\Delta Z_{\tau_p}\right)\Delta_n^{-1/2}\tilde{\varepsilon}^n_{i(p),k}\right).\end{cases} \tag{A.8}$$

From (A.7), we have $\left(\mathcal{L}_n(w), \mathcal{H}_n(w)\right) \xrightarrow{\mathcal{L}\text{-}s} \left(\mathcal{L}(w), \mathcal{H}(w)\right)$ as asserted.

(b) Under the independence assumption, conditionally on (σ, J), $\Delta_n^{-1/2}\tilde{\varepsilon}^n_{i(p),k}$ are independent centered Gaussian with conditional variance $(1, -\beta^{*\top}\partial g\left(\Delta Z_{\tau_p}\right))\bar{c}_{n,p}(1, -\beta^{*\top}\partial g\left(\Delta Z_{\tau_p}\right))^\top$, where

$$\bar{c}_{n,p} \equiv \frac{1}{k\Delta_n}\int_{(i(p)-1)\Delta_n}^{(i(p)-1+k)\Delta_n} c_s ds.$$

Since there is no price-volatility cojump, conditionally on (σ, J), the ς_p variables are independent centered Gaussian with conditional variance $(1, -\beta^{*\top}\partial g\left(\Delta Z_{\tau_p}\right))c_{\tau_p}(1, -\beta^{*\top}\partial g\left(\Delta Z_{\tau_p}\right))^\top$. Therefore,

$$\begin{aligned}&\sup_x \left|\mathbb{P}(\mathcal{L}_n(w) \leq x|\sigma, J) - \mathbb{P}(\mathcal{L}(w) \leq x|\sigma, J)\right| \\ &\quad \leq O_p(1)\sum_{p\in\mathcal{P}}\left\|\bar{c}_{n,p} - c_{\tau_p}\right\|.\end{aligned}$$

Since the process c is an Itô semimartingale, the majorant side of the above display is $O_p(\Delta_n^{1/2})$. □

A.3. Proof of Theorem 3

Proof of Theorem 3. We consider the following set of sample paths

$$\Omega_n \equiv \{\mathcal{J}_n = \bar{\mathcal{C}}_n = \mathcal{J}_n^*\} \cap \{|i-j| > 2k_n \text{ for any } i, j \in \mathcal{J}_n \text{ with } i \neq j\}.$$

From (A.1), it is easy to see that $\mathbb{P}(\Omega_n) \to 1$. Hence, we can restrict the calculation below on the sample paths in Ω_n without loss of generality.

Since $\hat{c}_{n,i(p)\pm} = O_p(1)$,

$$\begin{cases} \mathbb{E}\left[\, \| \Delta^n_{i(p),k} X^{*c} \|^2 \,\middle|\, \mathcal{F} \right] = O_p(\Delta_n), \\ \mathbb{E}\left[\, \| g(\Delta^n_{i(p),k} Z^*) - g(\Delta^n_{i(p),k} Z) \|^2 \,\middle|\, \mathcal{F} \right] = O_p(\Delta_n). \end{cases} \tag{A.9}$$

Since $\hat{\beta}_n - \hat{\beta}_n(w) = O_p(\Delta_n^{1/2})$ and $k_n \Delta_n \to 0$, we further have

$$k_n^{1/2} \left(\Delta^n_{i(p),k} Y^* - \hat{\beta}_n^\top g(\Delta^n_{i(p),k} Z^*) \right) = o_p(1). \tag{A.10}$$

Note that, w.p.a.1,

$$k_n^{1/2} \left(\hat{\beta}^*_n - \hat{\beta}_n \right) = \left(\sum_{p \in \mathcal{P}} g(\Delta^n_{i(p),k} Z^*) g(\Delta^n_{i(p),k} Z^*)^\top \right)^{-1} \times \left(\sum_{p \in \mathcal{P}} g(\Delta^n_{i(p),k} Z^*) k_n^{1/2} \left(\Delta^n_{i(p),k} Y^* - \hat{\beta}_n^\top g(\Delta^n_{i(p),k} Z^*) \right) \right).$$

From (A.10), we deduce

$$k_n^{1/2} \left(\hat{\beta}^*_n - \hat{\beta}_n \right) = o_p(1). \tag{A.11}$$

Next, we observe from (3.9) that $\hat{c}^*_{n,i(p)\pm} - \hat{c}_{n,i(p)\pm}$ are averages of $\mathcal{F}$-conditionally independent mean-zero variables with stochastically bounded fourth $\mathcal{F}$-conditional moments. We further note that

$$\begin{aligned} &\mathbb{E}\left[\left(\hat{c}^{*jk}_{n,i(p)\pm} - \hat{c}^{jk}_{n,i(p)\pm} \right) \left(\hat{c}^{*lm}_{n,i(p)\pm} - \hat{c}^{lm}_{n,i(p)\pm} \right) \middle| \mathcal{F} \right] \\ &\quad = \frac{1}{k_n} \left(\hat{c}^{jl}_{n,i(p)\pm} \hat{c}^{km}_{n,i(p)\pm} + \hat{c}^{jm}_{n,i(p)\pm} \hat{c}^{kl}_{n,i(p)\pm} \right), \\ &\mathbb{E}\left[\left(\hat{c}^{*jk}_{n,i(p)\pm} - \hat{c}^{jk}_{n,i(p)\pm} \right) \left(\hat{c}^{*lm}_{n,i(p)\mp} - \hat{c}^{lm}_{n,i(p)\mp} \right) \middle| \mathcal{F} \right] = 0. \end{aligned}$$

Upon using a subsequence characterization of convergence in probability and applying Lindeberg's central limit theorem under the $\mathcal{F}$-conditional probability, we derive

$$k_n^{1/2} \left(\hat{c}^*_{n,i(p)-} - \hat{c}_{n,i(p)-}, \hat{c}^*_{n,i(p)+} - \hat{c}_{n,i(p)+} \right)_{p \geq 1} \xrightarrow{\mathcal{L}|\mathcal{F}} \left(\zeta_{p-}, \zeta_{p+} \right)_{p \geq 1}. \tag{A.12}$$

By (A.9), (A.11) and (A.12), we use the delta method to deduce

$$k_n^{1/2} \left(\hat{w}^*_{n,i(p)} - \hat{w}_{n,i(p)} \right)_{p \geq 1} \xrightarrow{\mathcal{L}|\mathcal{F}} \left(\tilde{w}_p \right)_{p \geq 1}. \tag{A.13}$$

We now set

$$\begin{aligned} \varepsilon^{n*}_{i,k} &= \Delta^n_{i,k} Y^{*c} - \hat{\beta}_n(w)^\top \left(g\left(\Delta^n_{i,k} Z^* \right) - g\left(\Delta^n_{i,k} Z \right) \right), \\ \tilde{\varepsilon}^{n*}_{i,k} &= \Delta^n_{i,k} Y^{*c} - \hat{\beta}_n(w)^\top \partial g\left(\Delta^n_{i,k} Z \right) \Delta^n_{i,k} Z^{*c}. \end{aligned}$$

Note that $\varepsilon^{n*}_{i,k} - \tilde{\varepsilon}^{n*}_{i,k} = O_p(\Delta_n)$. Recall the $\mathcal{F}$-conditional distribution of $\Delta^n_{i,k} X^{*c}$ from Algorithm 1. Since $\hat{\beta}_n(w) \xrightarrow{\mathbb{P}} \beta^*$ and $\partial g(\Delta^n_{i(p),k} Z) \xrightarrow{\mathbb{P}} \partial g\left(\Delta Z_{\tau_p} \right)$, we further deduce

$$\Delta_n^{-1/2} \left(\tilde{\varepsilon}^{n*}_{i(p),k} \right)_{p \geq 1} \xrightarrow{\mathcal{L}|\mathcal{F}} \left(\varsigma_p \right)_{p \geq 1}. \tag{A.14}$$

Note that, in restriction to Ω_n, $\hat{c}^*_{n,i(p)\pm}$ and $\tilde{\varepsilon}^{n*}_{i(p),k}$ are $\mathcal{F}$-conditionally independent because they do not involve overlapping increments of W^*. Hence, (A.13) and (A.14) hold jointly, yielding

$$\left(\Delta_n^{-1/2} \tilde{\varepsilon}^{n*}_{i(p),k}, k_n^{1/2} \left(\hat{w}^*_{n,i(p)} - \hat{w}_{n,i(p)} \right) \right)_{p \geq 1} \xrightarrow{\mathcal{L}|\mathcal{F}} \left(\varsigma_p, \tilde{w}_p \right)_{p \geq 1}. \tag{A.15}$$

For notational simplicity, we denote

$$\begin{aligned} \Xi_n &\equiv \sum_{p \in \mathcal{P}} \hat{w}_{n,i(p)} g(\Delta^n_{i(p),k} Z) g(\Delta^n_{i(p),k} Z)^\top, \\ \Xi^*_n &\equiv \sum_{p \in \mathcal{P}} \hat{w}^*_{n,i(p)} g(\Delta^n_{i(p),k} Z^*) g(\Delta^n_{i(p),k} Z^*)^\top, \\ \Lambda^*_n &\equiv \Delta_n^{-1/2} \sum_{p \in \mathcal{P}} \hat{w}^*_{n,i(p)} g(\Delta^n_{i(p),k} Z^*) \varepsilon^{n*}_{i(p),k}. \end{aligned}$$

Note that, w.p.a.1,

$$\Delta_n^{-1/2} \left(\hat{\beta}^*_n(w) - \hat{\beta}_n(w) \right) = \Xi^{*-1}_n \Lambda^*_n. \tag{A.16}$$

By $\hat{w}^*_{n,i(p)} = O_p(1)$ and (A.9), we have

$$\Xi^*_n = \Xi_n + \sum_{p \in \mathcal{P}} \left(\hat{w}^*_{n,i(p)} - \hat{w}_{n,i(p)} \right) g(\Delta^n_{i(p),k} Z) g(\Delta^n_{i(p),k} Z)^\top + O_p(\Delta_n^{1/2}).$$

Therefore,

$$\Xi^{*-1}_n = \Xi^{-1}_n - \Xi^{-1}_n \tilde{\Xi}^*_n \Xi^{-1}_n + o_p(k_n^{-1/2}), \tag{A.17}$$

where $\tilde{\Xi}^*_n \equiv \sum_{p \in \mathcal{P}} (\hat{w}^*_{n,i(p)} - \hat{w}_{n,i(p)}) g(\Delta^n_{i(p),k} Z) g(\Delta^n_{i(p),k} Z)^\top$.

We decompose

$$\begin{aligned} \Lambda^*_n = &\sum_{p \in \mathcal{P}} \hat{w}_{n,i(p)} g(\Delta^n_{i(p),k} Z^*) \Delta_n^{-1/2} \tilde{\varepsilon}^{n*}_{i(p),k} \\ &+ \sum_{p \in \mathcal{P}} \left(\hat{w}^*_{n,i(p)} - \hat{w}_{n,i(p)} \right) g(\Delta^n_{i(p),k} Z^*) \Delta_n^{-1/2} \tilde{\varepsilon}^{n*}_{i(p),k} \\ &+ O_p(\Delta_n^{1/2}). \end{aligned} \tag{A.18}$$

Note that $\tilde{\Xi}^*_n$ and the second term on the right-hand side of (A.18) are both $O_p(k_n^{-1/2})$. Therefore, from (A.16), (A.17) and (A.18), we obtain the decomposition

$$\Delta_n^{-1/2} \left(\hat{\beta}^*_n(w) - \hat{\beta}_n(w) \right) = \mathcal{L}^*_n(w) + k_n^{-1/2} \mathcal{H}^*_n(w) + o_p(k_n^{-1/2}),$$

where

$$\begin{aligned} \mathcal{L}^*_n(w) &\equiv \Xi^{-1}_n \sum_{p \in \mathcal{P}} \hat{w}_{n,i(p)} g(\Delta^n_{i(p),k} Z^*) \Delta_n^{-1/2} \tilde{\varepsilon}^{n*}_{i(p),k}, \\ \mathcal{H}^*_n(w) &\equiv \Xi^{-1}_n \sum_{p \in \mathcal{P}} k_n^{1/2} \left(\hat{w}^*_{n,i(p)} - \hat{w}_{n,i(p)} \right) g(\Delta^n_{i(p),k} Z^*) \Delta_n^{-1/2} \tilde{\varepsilon}^{n*}_{i(p),k} \\ &\quad - \Xi^{-1}_n \tilde{\Xi}^*_n \Xi^{-1}_n \left(\sum_{p \in \mathcal{P}} \hat{w}_{n,i(p)} g(\Delta^n_{i(p),k} Z^*) \Delta_n^{-1/2} \tilde{\varepsilon}^{n*}_{i(p),k} \right). \end{aligned}$$

From (A.15), it is easy to deduce $\left(\mathcal{L}^*_n(w), \mathcal{H}^*_n(w) \right) \xrightarrow{\mathcal{L}|\mathcal{F}} \left(\mathcal{L}(w), \mathcal{H}(w) \right)$ as asserted. □

A.4. Proof of Theorem 4

Proof of Theorem 4. (a) Since $\Delta^n_{i(p),k} X \xrightarrow{\mathbb{P}} \Delta X_{\tau_p}$, it is easy to see from (2.7) that

$$SSR_n \xrightarrow{\mathbb{P}} \sum_{p \in \mathcal{P}} \Delta Y^2_{\tau_p} - \left(\sum_{p \in \mathcal{P}} g\left(\Delta Z_{\tau_p} \right) \Delta Y_{\tau_p} \right)^\top \times \left(\sum_{p \in \mathcal{P}} g\left(\Delta Z_{\tau_p} \right) g\left(\Delta Z_{\tau_p} \right)^\top \right)^{-1} \left(\sum_{p \in \mathcal{P}} g\left(\Delta Z_{\tau_p} \right) \Delta Y_{\tau_p} \right).$$

In particular, the second assertion of part (a) holds. It remains to show the first assertion of part (a). It is easy to see that

$$SSR_n = \sum_{i \in \mathcal{J}_n} \Delta^n_{i,k} Y^2 - \left(\sum_{i \in \mathcal{J}_n} \Delta^n_{i,k} Y g\left(\Delta^n_{i,k} Z \right) \right)^\top \times \left(\sum_{i \in \mathcal{J}_n} g\left(\Delta^n_{i,k} Z \right) g\left(\Delta^n_{i,k} Z \right)^\top \right)^{-1} \left(\sum_{i \in \mathcal{J}_n} \Delta^n_{i,k} Y g\left(\Delta^n_{i,k} Z \right) \right). \tag{A.19}$$

Plug (2.9) into (A.19). After some elementary algebra, we derive

$$SSR_n = \sum_{i\in\mathcal{J}_n}\left(\varepsilon^n_{i,k}\right)^2 - \left(\sum_{i\in\mathcal{J}_n}\varepsilon^n_{i,k} g\left(\Delta^n_{i,k}Z\right)\right)^\top \times\left(\sum_{i\in\mathcal{J}_n} g\left(\Delta^n_{i,k}Z\right) g\left(\Delta^n_{i,k}Z\right)^\top\right)^{-1}\left(\sum_{i\in\mathcal{J}_n}\varepsilon^n_{i,k} g\left(\Delta^n_{i,k}Z\right)\right). \tag{A.20}$$

By (2.7), (A.20) holds w.p.a.1 with $\mathcal{J}_n$ replaced by $\mathcal{J}^*_n$. By (A.5) and $g(\Delta^n_{i(p),k}Z) \xrightarrow{\mathbb{P}} g\left(\Delta Z_{\tau_p}\right)$, we derive the first assertion of part (a).

(b) Similar to (A.19), we have

$$SSR^*_n = \sum_{i\in\mathcal{J}_n}\Delta^n_{i,k}Y^{*2} - \left(\sum_{i\in\mathcal{J}_n}\Delta^n_{i,k}Y^* g\left(\Delta^n_{i,k}Z^*\right)\right)^\top \times\left(\sum_{i\in\mathcal{J}_n} g\left(\Delta^n_{i,k}Z^*\right) g\left(\Delta^n_{i,k}Z^*\right)^\top\right)^{-1}\left(\sum_{i\in\mathcal{J}_n}\Delta^n_{i,k}Y^* g\left(\Delta^n_{i,k}Z^*\right)\right).$$

We now set

$$\varepsilon^{n*}_{i,k} \equiv \Delta^n_{i,k}Y^* - \hat{\beta}^\top_n g\left(\Delta^n_{i,k}Z^*\right).$$

After some elementary algebra, we can rewrite

$$SSR^*_n = \sum_{i\in\mathcal{J}_n}\left(\varepsilon^{n*}_{i,k}\right)^2 - \left(\sum_{i\in\mathcal{J}_n}\varepsilon^{n*}_{i,k} g\left(\Delta^n_{i,k}Z^*\right)\right)^\top \times\left(\sum_{i\in\mathcal{J}_n} g\left(\Delta^n_{i,k}Z^*\right) g\left(\Delta^n_{i,k}Z^*\right)^\top\right)^{-1}\left(\sum_{i\in\mathcal{J}_n}\varepsilon^{n*}_{i,k} g\left(\Delta^n_{i,k}Z^*\right)\right). \tag{A.21}$$

We further observe $\varepsilon^{n*}_{i,k} = \Delta^n_{i,k}Y^{*c} - \hat{\beta}^\top_n(g(\Delta^n_{i,k}Z^*) - g(\Delta^n_{i,k}Z))$. It is easy to show that $\hat{\beta}_n \xrightarrow{\mathbb{P}} \bar{\beta}$. From the construction of $\Delta^n_{i,k}X^{*c}$ described in Algorithm 1, as well as the fact that $\hat{c}_{n,i(p)\pm} \xrightarrow{\mathbb{P}} c_{\tau_p\pm}$, we deduce

$$\Delta_n^{-1/2}\left(\varepsilon^{n*}_{i(p),k}\right)_{p\geq 1} \xrightarrow{\mathcal{L}|\mathcal{F}} \left(\bar{\zeta}_p\right)_{p\geq 1}, \tag{A.22}$$

where $\bar{\zeta}_p \equiv \left(1, -\bar{\beta}^\top \partial g\left(\Delta Z_{\tau_p}\right)\right)\left(\sqrt{\kappa_p}\sigma_{\tau_p-}\xi_{p-} + \sqrt{k-\kappa_p}\sigma_{\tau_p}\xi_{p+}\right)$. By (A.21) and (A.22), we deduce

$$\Delta_n^{-1}SSR^*_n \xrightarrow{\mathcal{L}|\mathcal{F}} \sum_{p\in\mathcal{P}}\bar{\zeta}^2_p - \left(\sum_{p\in\mathcal{P}} g\left(\Delta Z_{\tau_p}\right)\bar{\zeta}_p\right)^\top \times\left(\sum_{p\in\mathcal{P}} g\left(\Delta Z_{\tau_p}\right) g\left(\Delta Z_{\tau_p}\right)^\top\right)^{-1}\left(\sum_{p\in\mathcal{P}} g\left(\Delta Z_{\tau_p}\right)\bar{\zeta}_p\right). \tag{A.23}$$

In restriction to Ω_0, the limiting distribution characterized by (A.23) coincides with the limiting distribution of $\Delta_n^{-1}SSR_n$. We further note that, conditionally on $\mathcal{F}$, this limiting distribution is atomless with positive density on $(0,\infty)$. Therefore, the $\mathcal{F}$-conditional $(1-\alpha)$ -quantile of this limiting distribution can be consistently estimated by $\Delta_n^{-1}cv^\alpha_n$. From here and part (a) of this theorem, it readily follows that $\mathbb{P}\left(SSR_n > cv^\alpha_n|\Omega_0\right) \to \alpha$.

In restriction to Ω_a, we see from part (a) that the probability limit of SSR_n is strictly positive. Moreover, the sequence $\Delta_n^{-1}cv^\alpha_n$ is still tight because of (A.23). The assertion $\mathbb{P}\left(SSR_n > cv^\alpha_n|\Omega_a\right) \to 1$ readily follows. □

References

Andersen, T.G., Bollerslev, T., Diebold, F.X., Vega, C., 2003. Micro effects of macro announcements: real-time price discovery in foreign exchange. Amer. Econ. Rev. 93 (1), 251–277.

Andersen, T.G., Bollerslev, T., Diebold, F.X., Wu, G., 2006. Realized beta: persistence and predictability. In: Advances in Econometrics: Econometric Analysis of Financial and Economic Time Series, vol. 20 (Part 2), Emerald Group Publishing Limited, pp. 1–39.

Bandi, F., Renó, R., 2016. Price and volatility co-jumps. J. Financ. Econ. 119, 107–146.

Barndorff-Nielsen, O., Hansen, P., Lunde, A., Shephard, N., 2009. Realized kernels in practice: trades and quotes. Econom. J. 12 (3), C1–C32.

Barndorff-Nielsen, O.E., Shephard, N., 2004a. Econometric analysis of realized covariation: high frequency based covariance, regression, and correlation in financial economics. Econometrica 72 (3), 885–925.

Barndorff-Nielsen, O.E., Shephard, N., 2004b. Power and bipower variation with stochastic volatility and jumps. J. Financ. Econom. 2, 1–37.

Comte, F., Renault, E., 1996. Long memory continuous time models. J. Econometrics 73, 101–149.

Davison, A.C., Hinkley, D.V., 1997. Bootstrap methods and their application. In: Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press.

Dovonon, P., Gonçalves, S., Meddahi, N., 2013. Bootstrapping realized multivariate volatility measures. J. Econometrics 172 (1), 49–65.

Dovonon, P., Hounyo, U., Gonçalves, S., Meddahi, N., 2014. Bootstrapping High Frequency Jump Tests. Toulouse School of Economics.

Gobbi, F., Mancini, C., 2012. Identifying the Brownian covariation from the co-jumps given discrete observations. Econometric Theory 28, 249–273.

Gonçalves, S., Meddahi, N., 2009. Bootstrapping realized volatility. Econometrica 77, 283–306.

Hasbrouck, J., 2003. Intraday price formation in U.S. equity index markets. J. Finance 58 (6), 2375–2399.

Hasbrouck, J., 2015. Securities Trading: Procedures and Principles. New York University.

Horowitz, J.L., 2001. The bootstrap. In: Handbook of Econometrics, vol. 5, Elsevier.

Hounyo, U., 2013. Bootstrapping Realized Volatility and Realized Beta under a Local Gaussianity Assumption. University of Oxford.

Jacod, J., Protter, P., 2012. Discretization of Processes. Springer.

Jacod, J., Rosenbaum, M., 2013. Quarticity and other functionals of volatility: efficient estimation. Ann. Statist. 41, 1462–1484.

Jacod, J., Todorov, V., 2010. Do price and volatility jump together? Ann. Appl. Probab. 20, 1425–1469.

Kalnina, I., 2013. Nonparametric Tests of Time Variation in Betas. University of Montreal.

Kyle, A.S., 1985. Continuous auctions and insider trading. Econometrica 53 (6), 1315–1335.

Lehmann, E.L., Romano, J.P., 2005. Testing Statistical Hypothesis. Springer.

Li, J., Todorov, V., Tauchen, G., 2017. Jump regressions. Econometrica 85, 173–195.

Mancini, C., 2001. Disentangling the jumps of the diffusion in a geometric jumping Brownian motion. Giornale Dell'Istituto Italiano Degli Attuari LXIV, 19–47.

Merton, R.C., 1976. Option pricing when underlying stock returns are discontinuous. J. Financ. Econ. 3 (1/2), 125–144.

Mykland, P., Zhang, L., 2009. Inference for continuous semimartingales observed at high frequency. Econometrica 77, 1403–1445.

Reiss, M., Todorov, V., Tauchen, G., 2015. Nonparametric test for a constant beta between ito semimartingales based on high-frequency data. Stochastic Process. Appl. 125, 2955–2988.

Roll, R., 1987. R^2. J. Finance 43, 541–566.

Todorov, V., Bollerslev, T., 2010. Jumps and Betas: A new framework for disentangling and estimating systematic risks. J. Econometrics 157, 220–235.

Todorov, V., Tauchen, G., 2011. Volatility Jumps. J. Bus. Econom. Statist. 29, 356–371.

Todorov, V., Tauchen, G., 2012. The realized Laplace transform of volatility. Economtrica 80, 1105–1127.

Zhang, L., Mykland, P.A., Aït-Sahalia, Y., 2005. A tale of two time scales: determining integrated volatility with noisy high-frequency data. J. Amer. Statist. Assoc. 100, 1394–1411.

Please cite this article in press as: Li J., et al., Mixed-scale jump regressions with bootstrap inference. Journal of Econometrics (2017), http://dx.doi.org/10.1016/j.jeconom.2017.08.017.