Jump Regressions*

Jia Li,† Viktor Todorov,‡ and George Tauchen§

August 30, 2016

Abstract

We develop econometric tools for studying jump dependence of two processes from high-frequency observations on a fixed time interval. In this context, only segments of data around a few outlying observations are informative for the inference. We derive an asymptotically valid test for stability of a linear jump relation over regions of the jump size domain. The test has power against general forms of nonlinearity in the jump dependence as well as temporal instabilities. We further propose an efficient estimator for the linear jump regression model that is formed by optimally weighting the detected jumps with weights based on the diffusive volatility around the jump times. We derive the asymptotic limit of the estimator, a semiparametric lower efficiency bound for the linear jump regression, and show that our estimator attains the latter. The analysis covers both deterministic and random jump arrivals. A higher-order asymptotic expansion for the optimal estimator further allows for finite-sample refinements. In an empirical application, we use the developed inference techniques to test the temporal stability of market jump betas.

Keywords: efficient estimation, high-frequency data, jumps, LAMN, regression, semimartingale, specification test, stochastic volatility.

JEL classification: C51, C52, G12.

---

*We would like to thank a co-editor and three anonymous referees for their thoughtful comments, which helped greatly improve the paper. We would also like to thank Torben G. Andersen, Snehal Banerjee, Ravi Bansal, Tim Bollerslev, Tim Christensen, Frank X. Diebold, Prosper Dovonon, Rob Engle, Kirill Evdokimov, Eric Ghysels, Peter R. Hansen, Jean Jacod, Michael Jansson, Konrad Menzel, Ulrich Mueller, Per Mykland, Jose Luis Montiel Olea, Andrew Patton, Eric Renault, Marcel Rindisbacher, Norman Swanson, Brian Weller, as well as seminar participants and discussants at various conferences for helpful discussions and suggestions. Huidi Lin provided excellent research assistance. Li's work was partially funded by NSF grant SES-1326819. Todorov's work was partially funded by NSF grant SES-1530748.

†Department of Economics, Duke University, Durham, NC 27708; e-mail: jl410@duke.edu.

‡Department of Finance, Kellogg School of Management, Northwestern University, Evanston, IL 60208; e-mail: v-todorov@northwestern.edu.

§Department of Economics, Duke University, Durham, NC 27708; e-mail: george.tauchen@duke.edu.

1 Introduction

Aggregate market risks exhibit discontinuities (i.e., jumps) in their dynamics. Bearing such non-diversifiable jump risk is significantly rewarded, as is evident from the expensiveness of short-maturity options written on the market index with strikes that are far from its current level. Therefore, precise estimates of the comovement of jumps in asset prices with those of aggregate risk factors will play a key role in our understanding of the pricing of jump risk in the cross-section.

High-frequency data allows for robust nonparametric inference for jumps. In their pioneering work, Barndorff-Nielsen and Shephard (2004b, 2006), using realized multipower variation measures, and Mancini (2001, 2009), using threshold-based methods, have developed nonparametric tools for measuring jump variation from high-frequency data. The goal of the current paper is to extend this strand of literature by developing econometric tools for testing and efficiently estimating pathwise relationships between jumps of an asset price process $(Y_t)_{t\geq 0}$ and an aggregate risk factor $(Z_t)_{t\geq 0}$.

More specifically we study the relationship between ΔY_τ and ΔZ_τ for $\tau \in \mathcal{T}$, where $\Delta Y_t \equiv Y_t - Y_{t-}$ and $\Delta Z_t \equiv Z_t - Z_{t-}$ for any $t \geq 0$, and $\mathcal{T}$ is the collection of jump times of Z. The statistical inference is based on discrete observations of (Y, Z) sampled on an observation grid with asymptotically shrinking mesh. The ratio (henceforth referred to as the *spot jump beta*)

$$\beta_\tau \equiv \frac{\Delta Y_\tau}{\Delta Z_\tau}, \quad \tau \in \mathcal{T}, \tag{1.1}$$

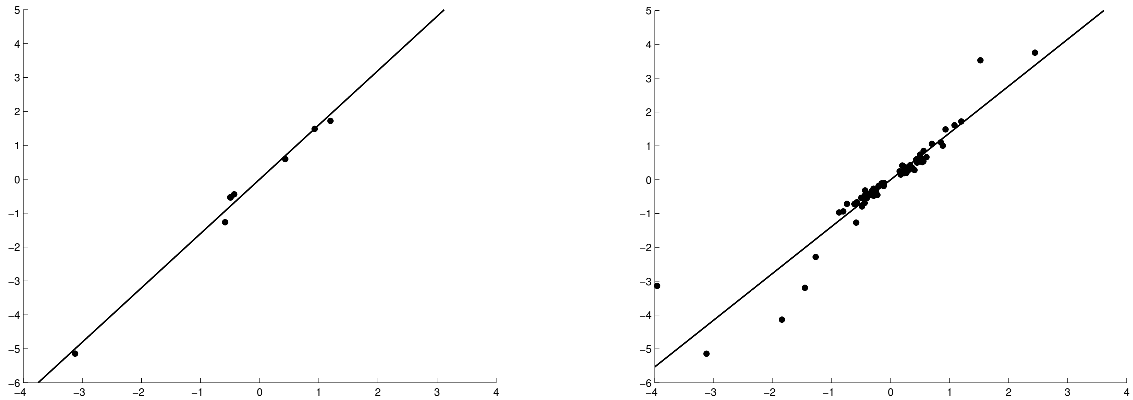
measures the co-movement of the jumps in the two processes. Without any model restriction, the spot jump beta is stochastic and varies across instances of jump events. However, in many cases such as factor models, which are used pervasively in asset pricing, the relationship between the jumps of Y and Z can be captured by a function which is known up to a finite-dimensional parameter. The most common is the linear function which leads to:

$$\Delta Y_t = \beta \Delta Z_t + \Delta \epsilon_t, \quad \Delta Z_t \Delta \epsilon_t = 0, \tag{1.2}$$

for some constant β, where ϵ captures the asset-specific jump risk as in the seminal work of Merton (1976). Equation (1.2) is akin to the usual regression in econometrics, except that the orthogonality condition is defined pathwise, i.e., ϵ has zero covariation with Z on the observed path (see next section for the formal definition of quadratic covariation). We can thus view (1.2) as a linear jump regression model, while noting the important fact that neither the jump time τ nor the jump sizes $(\Delta Y_\tau, \Delta Z_\tau)_{\tau\in\mathcal{T}}$ are directly observable from data sampled at discrete times.

A motivating empirical example of the jump regression is given in Figure 1. From 10-minute log returns, we select locally large (jump) returns of the S&P 500 exchange traded fund (ETF),

Figure 1: A Representative Illustration of Jump Regressions



Note: The horizontal axes are the jump returns of the S&P 500 ETF while the vertical axes are the contemporaneous returns of the Financial Sector ETF for data sampled at the 10-minute frequency in 2008 (left) and 2007-2012 (right), together with linear fits. The jump returns are selected according the thresholding procedure described in Section 5.

which is our proxy for the market, and plot them versus the contemporaneous Financial Sector ETF returns,[1] along with a linear fit based on model (1.2). We see that the simple linear jump regression model provides quite a good fit in the one-year subsample (left) and a less tight fit for the six-year sample (right). Are these patterns statistically consistent with model (1.2)? On one hand, due to the very nature of jumps, jump regressions are inevitably based on a few high-frequency observations. On the other hand, the signal-to-noise ratio of these observations is likely to be very high. If model (1.2) is true, how do we efficiently estimate the jump beta? The main contribution of this paper is to develop econometric tools that address the above questions.

In the first part of our analysis, we develop a specification test for the linear relationship (1.2) and its piecewise generalizations. The test is asymptotically consistent against all nonparametric fixed alternatives for which (1.2) is violated, for example, due to time variation in the jump beta and/or nonlinearity in the jump relationship (i.e., the dependence of the jump beta on the jump size). The test is based on the fact that the linear model (1.2) is equivalent to the singularity of the realized jump covariation matrix computed over the times of jumps of Z only.[2] Our test rejects the null hypothesis when the determinant of a sample analogue estimator of the jump covariation matrix is larger than a critical value. While the estimator for the jump covariation has a well-known central limit theorem at the usual parametric $\sqrt{n}$-rate, see e.g., Jacod (2008),

[1]The identification of the jump returns of the market portfolio is done using a standard adaptive thresholding technique, see e.g., Lee and Mykland (2008), that is described and rigorously justified in the main text below.

[2]This is an example of a reduced rank restriction, which arises naturally in several areas of statistics, econometrics, and several other disciplines as well; see Anderson (1951) and the more recent work of Anderson (2002), Gourieroux and Jasiak (2013) and Jacod and Podolskij (2013) for various such instances.

its determinant is asymptotically degenerate under the null hypothesis specified by (1.2) and no asymptotic theory has been developed for it to date. We thus consider higher-order asymptotics so as to characterize the non-degenerate asymptotic null distribution of the test statistic. The resultant null distribution can be represented as a quadratic form of mixed Gaussian variables scaled by (random) jumps and spot volatilities. Since this distribution is nonstandard, we further provide a simple simulation-based algorithm to compute the critical value for our test.

On the presumption of the linear model (1.2), for a given time interval and a range of the jump size, we further study the efficient estimation of the jump beta. Under certain assumptions, we derive a semiparametric lower efficiency bound for regular estimators of the jump beta.[3] Following Stein's insight (Stein (1956)) that the estimation in a semiparametric problem is no easier than in any parametric submodel, we compute the efficiency bound by first constructing a class of submodels. These submodels satisfy the local asymptotic mixed normality (LAMN) property with a random information matrix. For these submodels, we compute the worst-case Cramer-Rao information bound of estimating the jump beta. In addition, we show that this lower efficiency bound is actually sharp by constructing a semiparametric estimator which attains it. This direct approach reveals that the key nuisance component is the unknown heterogeneous jump sizes of Z and the least favorable submodel should fully account for their presence. In particular, the estimation of jump beta is generally not adaptive with respect to the sizes of jumps in Z.

The proposed efficient estimator is an optimally weighted linear estimator and has formal parallels to classical weighted least squares estimation in a linear heteroskedastic regression context. The optimal weights in the present setting are determined by nonparametric high-frequency estimates of the local volatility of the instantaneous residual term $Y - \beta Z$ at the jump times. The efficient estimator enjoys the parametric convergence rate $\sqrt{n}$, despite the presence of spot volatility estimates, which in general can be estimated at a convergence rate no faster than $n^{1/4}$.

To improve finite sample performance, we further derive a novel higher-order expansion for the optimally weighted estimator which clearly reveals the role of the spot covariance estimates in the estimation. Moreover, it allows us to design a simple finite-sample refinement (relative to the standard high-frequency asymptotics) for confidence sets of the jump beta. The efficient estimator provides considerable efficiency gains over natural alternatives based on the ratio of the

[3] The general theory of semiparametric efficient estimation has been developed for models admitting locally asymptotically normal (LAN) likelihood ratios, see e.g., Bickel, Klaassen, Ritov, and Wellner (1998) and references therein. By contrast, the infill asymptotic setting with high-frequency data is non-ergodic which renders the limiting distribution random, meaning that its variability depends on the realization of the underlying processes. In this nonstandard setting, Mykland and Zhang (2009), Jacod and Rosenbaum (2013), Clément, Delattre, and Gloter (2013) and Renault, Sarisoy, and Werker (2016) study the efficient nonparametric estimation of general integrated volatility functionals, and Li, Todorov, and Tauchen (2016a) study the adaptive estimation in a semiparametric regression model for the diffusive part of a multivariate semimartingale process. All this work focuses on the diffusive components of the asset prices, by either filtering out the price jumps or assuming them away. But the jumps are exactly the focus of the current paper and as well known, their econometric analysis is very different.

jump covariation between Y and Z to the jump variation of Z (Gobbi and Mancini (2012)) as well as ratios of higher order power variations (Todorov and Bollerslev (2010)).

In an empirical application of the proposed inference techniques, we study the market jump betas of the nine industry portfolios comprising the S&P 500 stock market index for the period 2007–2012. The premise of our empirical application is the well-known fact that standard market betas are strongly time-varying. The key empirical question that we address here is whether the previously documented temporal instability of market betas is constrained only to the regular non-jump moves or it is present for the jump moves as well. The pattern seen in Figure 1 is indeed representative for our empirical findings for all assets in our empirical analysis. While we find evidence for temporal instability over the whole six-year sample, market jump betas appear reasonably stable over periods as long as a year.

The rest of the paper is organized as follows. Section 2 presents the setting. The theory is developed in Section 3 for specification testing and in Section 4 for efficient estimation. Section 5 presents our empirical application. The online supplement to this paper contains (a) some additional theoretical results, (b) Monte Carlo and (c) all proofs.

2 Setting and modeling jump dependence

We start with introducing the formal setup for our analysis. The following notation is used throughout. We denote the transpose of a matrix A by $A^\top$. The adjoint matrix of a square matrix A is denoted $A^{\#}$. For two vectors a and b, we write $a \leq b$ if the inequality holds componentwise. The functions $vec(\cdot)$, $\det(\cdot)$ and $\mathrm{Tr}(\cdot)$ denote matrix vectorization, determinant and trace, respectively. The Euclidean norm of a linear space is denoted $\|\cdot\|$. We use $\mathbb{R}_*$ to denote the set of nonzero real numbers, that is, $\mathbb{R}_* \equiv \mathbb{R} \setminus \{0\}$. The cardinality of a (possibly random) set $\mathcal{P}$ is denoted $|\mathcal{P}|$. The largest smaller integer function is denoted by $\lfloor \cdot \rfloor$. 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 $\xrightarrow{\mathbb{P}}$, $\xrightarrow{\mathcal{L}}$ and $\xrightarrow{\mathcal{L}\text{-}s}$ to denote convergence in probability, convergence in law and stable convergence in law, respectively.

2.1 The underlying processes

The object of study of the paper is the dependence of the jumps in a univariate process Y on the jumps of another process Z. For simplicity of exposition, we will assume that Z is one-dimensional, but the results can be trivially generalized to settings where Z is multidimensional.

We proceed with the formal setup. Let Z and Y be defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$. Throughout the paper, all processes are assumed to be càdlàg adapted. We denote $X = (Z, Y)^\top$. Our basic assumption is that X is an Itô semimartingale with the form

$$X_t = x_0 + \int_0^t b_s ds + \int_0^t \sigma_s dW_s + J_t, \quad J_t = \sum_{s \leq t} \Delta X_s = \int_0^t \int_{\mathbb{R}} \delta(s, u) \mu(ds, du), \tag{2.1}$$

where the drift b_t takes value in $\mathbb{R}^2$; the volatility process σ_t takes value in $\mathcal{M}_2$, the space of 2×2 matrices; W is a 2-dimensional standard Brownian motion; $\delta = (\delta_Z, \delta_Y)^\top : \Omega \times \mathbb{R}_+ \times \mathbb{R} \mapsto \mathbb{R}^2$ 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 $\mathbb{R}$. Recall from the introduction, that the jump of X at time t is denoted by $\Delta X_t \equiv X_t - X_{t-}$, where $X_{t-} \equiv \lim_{s \uparrow t} X_s$. The spot covariance matrix of X at time t is denoted by $c_t \equiv \sigma_t \sigma_t^\top$, which we partition as

$$c_t = \begin{pmatrix} c_{ZZ,t} & c_{ZY,t} \\ c_{ZY,t} & c_{YY,t} \end{pmatrix}. \tag{2.2}$$

We also write $J_t = (J_{Z,t}, J_{Y,t})^\top$, so that J_Z and J_Y are the jump components of Z and Y, respectively. Our basic regularity condition for X is given by the following assumption.

Assumption 1. *(a) The process b is locally bounded; (b) c_t is nonsingular for $t \in [0, T]$; (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 for simplicity as our focus in this paper are "big" jumps, i.e., jumps that are not "sufficiently" close to zero. Alternatively, we can drop Assumption 1(c) and focus on jumps with sizes bounded away from zero.[4]

Turning to the sampling scheme, we assume that X is observed at discrete times $i\Delta_n$, for $0 \leq i \leq n \equiv \lfloor T/\Delta_n \rfloor$, within the *fixed* time interval $[0, T]$. The increments of X are denoted by

$$\Delta_i^n X \equiv X_{i\Delta_n} - X_{(i-1)\Delta_n}, \quad i = 1, \ldots, n. \tag{2.3}$$

Below, we consider an infill asymptotic setting, that is, $\Delta_n \to 0$ as $n \to \infty$.

2.2 Piecewise linear jump regression models

We proceed with the jump regression model which in the most general setting is given by

$$\Delta Y_t = f(\Delta Z_t) + \Delta \epsilon_t, \quad \Delta Z_t \Delta \epsilon_t = 0, \quad t \in [0, T], \tag{2.4}$$

where f is a piecewise linear function and ϵ captures Y-specific jumps. Similar to a standard regression model, we can equivalently define our jump regression model via the orthogonality

[4] Yet another strategy, that can allow for studying dependence in infinite activity jumps, is to use higher order powers in the statistics that we develop henceforth, see e.g., Todorov and Bollerslev (2010). This, however, comes at the price of losing some efficiency for the analysis of the "big" jumps.

condition

$$[\epsilon, Z] = 0, \quad \text{on} \quad [0, T], \tag{2.5}$$

where the quadratic covariation process $[\epsilon, Z]$ is given by $\sum_{s \leq \bullet} \Delta\epsilon_s \Delta Z_s$. Our use of the quadratic covariation to define the jump regression model parallels the realized regressions of Barndorff-Nielsen and Shephard (2004a) and Mykland and Zhang (2006), which concern processes with continuous paths (in which case the quadratic covariation is the integrated covariance).

The orthogonality condition in (2.5) can be equivalently written as

$$\int_0^t \int_{\mathbb{R}} [(\delta_Y(s,u) - f(\delta_z(s,u)))\delta_z(s,u)]\mu(ds, du) = 0, \quad \forall t \in [0, T].$$

Unlike the usual regression for which orthogonality of the error term and the explanatory variable is defined in terms of expectations, the orthogonality condition in our jump regression is defined with respect to the jump measure μ controlling the behavior of the jumps on the observed path.

The jump model in (2.4) (potentially extended to general f and multivariate Z) arises naturally in asset pricing models in which the economy-wide pricing kernel is a function of a low-dimensional systematic vector of factors and the cash flows of the asset depend on these factors driving the pricing kernel as well as on idiosyncratic shocks which can include idiosyncratic jump risk as proposed in the work of Merton (1976).

The leading case of our jump regression model, as already explained in the introduction, is the one in which $f(z) = \beta \cdot z$ for some constant parameter β. This case plays a central role in finance as it represents the only jump model for which the overall asset beta, defined as the ratio of the quadratic co-variation between Y and Z and the quadratic variation of Z (see, e.g., Barndorff-Nielsen and Shephard (2004a) and references therein) remains constant on the time interval $[0, T]$. Such temporal stability of the beta of an asset is important both for the validity of methods estimating beta over local windows of time (as is typically done in practice) but also plays a central role in judging whether conditional asset pricing models can potentially explain standard cross-sectional asset pricing puzzles (see, e.g., Lewellen and Nagel (2006)).[5]

The linear jump regression model can be generalized in the following two directions.

Example 1 (Temporal Breaks)**.** *Conditional asset pricing models allow for the exposure of assets to fundamental risks to change over time; see, for example, Hansen and Richard (1987). In our context, this implies that the jump beta can vary over time, but presumably not too erratically. A practically relevant model is to assume that the jump beta remains constant over fixed intervals*

[5] We note that if at the time $\tau \in \mathcal{T}$ of jump in the systematic factor Z, the jump ΔY_τ has an an asset-specific error term in addition to $f(\Delta Z_\tau)$, then one cannot identify β using infill asymptotics. This is because each of the finite number of jumps will contain a non-vanishing asset-specific error term which cannot be "averaged out" by sampling more frequently. In such a case, the ratio $\frac{\Delta Y_\tau}{\Delta Z_\tau}$ will be temporally unstable and its prediction from past observations limited. Such a scenario will imply limited temporal stability of overall asset betas (regardless of the estimation horizon), more generally, and we develop tests against such scenarios.

of time (e.g., months, quarters, years), an assumption which is often made in empirical asset pricing. We refer to such an extension of the constant beta model as a temporal structural break model. More formally, let $(\mathcal{S}_k)_{1\leq k\leq \bar{k}}$ be a finite disjoint partition of $[0,T]$, which corresponds to the horizon of $\bar{k}$ regimes. The structural break model amounts to imposing

$$\beta_\tau = \sum_{k=1}^{\bar{k}} \beta_{0,k} 1_{\{\tau\in\mathcal{S}_k\}}, \quad \tau \in \mathcal{T}, \tag{2.6}$$

where the constant $\beta_{0,k}$ is the jump beta during the time period $\mathcal{S}_k$. Equivalently,

$$\Delta Y_t = \sum_{k=1}^{\bar{k}} \beta_{0,k} \Delta Z_t 1_{\{t\in\mathcal{S}_k\}} + \Delta\epsilon_t, \quad \Delta Z_t \Delta\epsilon_t = 0, \quad t \in [0,T]\,. \tag{2.7}$$

Example 2 (Spatial Breaks)**.** *An alternative generalization is to allow the slope coefficient to depend on the jump size of Z, but in a time-invariant manner. In other words, Y reacts differently to jumps in Z depending on the size of the latter. The simplest model is to allow the jump beta to be different depending on the sign of ΔZ, leading to the notion of up-side and down-side jump betas. The latter can be viewed as continuous-time analogues of the downside betas of Ang, Chen, and Xing (2006) and Lettau, Maggiori, and Weber (2014) which are based on discrete (large) returns. More generally, let $(\mathcal{S}_k)_{1\leq k\leq \bar{k}}$ be a finite disjoint partition of $\mathbb{R}$. We set*

$$\beta_\tau = \sum_{k=1}^{\bar{k}} \beta_{0,k} 1_{\{\Delta Z_\tau\in\mathcal{S}_k\}}, \quad \tau \in \mathcal{T}. \tag{2.8}$$

This corresponds to a piece-wise linear model:

$$\Delta Y_t = \sum_{k=1}^{\bar{k}} \beta_{0,k} \Delta Z_t 1_{\{\Delta Z_t\in\mathcal{S}_k\}} + \Delta\epsilon_t \quad \Delta Z_t \Delta\epsilon_t = 0, \quad t \in [0,T]\,. \tag{2.9}$$

We now introduce some notation for the jump regression model that we will use henceforth. Let $(\tau_p)_{p\geq 1}$ be the successive jump times of the process Z. We define two random sets $\mathcal{P} = \{p \geq 1 : \tau_p \leq T\}$ and $\mathcal{T} = \{\tau_p : p \in \mathcal{P}\}$, which collect respectively the indices of the jump times on $[0,T]$ and the jump times themselves. Since Z has finite-activity jumps, these sets are finite almost surely. Below, we refer to a Borel measurable subset $\mathcal{D} \subseteq [0,T]\times\mathbb{R}_*$ as a (temporal-spatial) *region*. For the jump of Z that occurs at stopping time $\tau \in \mathcal{T}$, we call $(\tau, \Delta Z_\tau)$ its *mark*. For each region $\mathcal{D}$, we set $\mathcal{P}_\mathcal{D} \equiv \{p \in \mathcal{P} : (\tau_p, \Delta Z_{\tau_p}) \in \mathcal{D}\}$; the random set $\mathcal{P}_\mathcal{D}$ collects the indices of jumps whose marks fall in the region $\mathcal{D}$. The jump regression is a model of the form (2.4) with f given by

$$f(\Delta Z_{\tau_p}) = \beta \Delta Z_{\tau_p}, \quad \text{for some constant } \beta \in \mathbb{R} \text{ and all } p \in \mathcal{P}_\mathcal{D}. \tag{2.10}$$

That is, the spot jump beta is a constant for all jumps whose marks are in the region $\mathcal{D}$. The linear regression model (1.2) corresponds to $\mathcal{D} = [0,T]\times\mathbb{R}_*$, and Examples 1 and 2 concern regions of the form $\mathcal{D}_k = \mathcal{S}_k \times \mathbb{R}_*$ and $[0, T] \times \mathcal{S}_k$, respectively. Below, the jump covariation matrix on the region $\mathcal{D}$ is given by

$$Q(\mathcal{D}) = \begin{pmatrix} Q_{ZZ}(\mathcal{D}) & Q_{ZY}(\mathcal{D}) \\ Q_{ZY}(\mathcal{D}) & Q_{YY}(\mathcal{D}) \end{pmatrix} \equiv \sum_{p \in \mathcal{P}_\mathcal{D}} \Delta X_{\tau_p} \Delta X_{\tau_p}^\top. \tag{2.11}$$

2.3 Inference for the jump marks

We finish this section with an auxiliary result concerning the approximation of the jump marks of the process X using discretely sampled data. This result provides guidance for the theory for the jump regressions developed below. It also gives a theoretical justification for scatter plots like Figure 1. In order to disentangle jumps from the diffusive component of asset returns, we choose a sequence v_n of truncation threshold values which satisfy the following condition:

$$v_n \asymp \Delta_n^\varpi \text{ for some constant } \varpi \in (0, 1/2). \tag{2.12}$$

Time-invariant choice for v_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 v_n should take into account the variation of c_t in an adaptive, data-driven way. We refer to supplemental appendix B for the details of such a way of constructing v_n using the bipower variation estimator (Barndorff-Nielsen and Shephard (2004b)).

For each $p \in \mathcal{P}$, $i(p)$ is the unique random index i such that $\tau_p \in ((i-1)\Delta_n, i\Delta_n]$. We set

$$\begin{aligned} \mathcal{I}_n(\mathcal{D}) &\equiv \{i : 1 \leq i \leq n, ((i-1)\Delta_n, \Delta_i^n Z) \in \mathcal{D}, |\Delta_i^n Z| > v_n\}, \\ \mathcal{I}(\mathcal{D}) &\equiv \{i(p) : p \in \mathcal{P}_\mathcal{D}\}. \end{aligned} \tag{2.13}$$

The set-valued statistic $\mathcal{I}_n(\mathcal{D})$ collects the indices of returns whose "marks" $((i-1)\Delta_n, \Delta_i^n Z)$ are in the region $\mathcal{D}$, where the truncation criterion $|\Delta_i^n Z| > v_n$ eliminates diffusive returns asymptotically. The random and unobservable set $\mathcal{I}(\mathcal{D})$ collects the indices of sampling intervals that contain the jumps with marks in $\mathcal{D}$. We also impose the following mild regularity condition on $\mathcal{D}$, which amounts to requiring that the jump marks of Z almost surely do not fall on the boundary of $\mathcal{D}$.

Assumption 2. $\nu(\{(s,u) \in [0,T] \times \mathbb{R} : (s, \delta_Z(s,u)) \in \partial\mathcal{D}\}) = 0$, *where* $\partial\mathcal{D}$ *is the boundary of* $\mathcal{D}$.

Below, we use the following definition for the convergence of random vectors with possibly different length: for a sequence N_n of random integers and a sequence $((A_{j,n})_{1\leq j\leq N_n})_{n\geq 1}$ of random elements, we write $(A_{j,n})_{1\leq j\leq N_n} \xrightarrow{\mathbb{P}} (A_j)_{1\leq j\leq N}$ if $\mathbb{P}(N_n = N) \longrightarrow 1$ and $(A_{j,n})_{1\leq j\leq N} 1_{\{N_n = N\}} \xrightarrow{\mathbb{P}} (A_j)_{1\leq j\leq N}$.

Proposition 1 (Approximation of Jump Marks)**.** *Under Assumptions 1 and 2,*

(a) $\mathbb{P}(\mathcal{I}_n(\mathcal{D}) = \mathcal{I}(\mathcal{D})) \to 1$;

(b) $((i-1)\Delta_n, \Delta_i^n X)_{i \in \mathcal{I}_n(\mathcal{D})} \xrightarrow{\mathbb{P}} (\tau_p, \Delta X_{\tau_p})_{p \in \mathcal{P}_\mathcal{D}}$.

Proposition 1(a) shows that the set $\mathcal{I}_n(\mathcal{D})$ coincides with $\mathcal{I}(\mathcal{D})$ with probability approaching one. In this sense, $\mathcal{I}_n(\mathcal{D})$ consistently locates the discrete time intervals that contain jumps with marks in the region $\mathcal{D}$. A by-product of this result is that $|\mathcal{I}_n(\mathcal{D})|$ is a consistent (integer-valued) estimator of the number of jumps with marks in $\mathcal{D}$. Proposition 1(b) further shows that the jump marks of interest, that is $\left(\tau_p, \Delta X_{\tau_p}\right)_{p\in\mathcal{P}_\mathcal{D}}$, can be consistently estimated by the collection of time-return pairs $((i-1)\Delta_n, \Delta_i^n X)_{i\in\mathcal{I}_n(\mathcal{D})}$.

Proposition 1 has a useful implication for data visualization in empirical work. Indeed, the collections $((i-1)\Delta_n, \Delta_i^n X)_{i\in\mathcal{I}_n(\mathcal{D})}$ and $\left(\tau_p, \Delta X_{\tau_p}\right)_{p\in\mathcal{P}_\mathcal{D}}$ can be visualized as scatter plots on $[0,T]\times\mathbb{R}^2$, or its low-dimensional projections like Figure 1. Proposition 1(b) thus provides a sense in which the graph of the former consistently estimates that of the latter.

3 Testing for constant jump beta

3.1 The specification test

We start our theoretical analysis of the jump dependence with testing the hypothesis of constant jump beta on a fixed region $\mathcal{D}$. We shall consider the nondegenerate case where Z has at least two jumps with marks in $\mathcal{D}$, that is, $|\mathcal{P}_\mathcal{D}|\geq 2$. Formally, the testing problem is to decide in which of the following two sets the observed sample path falls:

$$\begin{cases} \Omega_0(\mathcal{D}) \equiv \{\omega\in\Omega : \text{condition (2.10) holds for some } \beta_0(\omega) \text{ on path } \omega\} \cap \{|\mathcal{P}_\mathcal{D}|\geq 2\}, \\ \Omega_a(\mathcal{D}) \equiv \{\omega\in\Omega : \text{condition (2.10) does not hold on path } \omega\} \cap \{|\mathcal{P}_\mathcal{D}|\geq 2\}. \end{cases} \tag{3.1}$$

By the Cauchy–Schwarz inequality, it is easy to see that condition (2.10) is equivalent to the singularity of the positive semidefinite matrix $Q(\mathcal{D})$. Hence, a test for constant jump beta can be carried out via a one-sided test for $\det[Q(\mathcal{D})]=0$.

In view of Proposition 1, we construct a sample analogue estimator for $Q(\mathcal{D})$ as

$$Q_n(\mathcal{D}) = \begin{pmatrix} Q_{ZZ,n}(\mathcal{D}) & Q_{ZY,n}(\mathcal{D}) \\ Q_{ZY,n}(\mathcal{D}) & Q_{YY,n}(\mathcal{D}) \end{pmatrix} = \sum_{i\in\mathcal{I}_n(\mathcal{D})} \Delta_i^n X \Delta_i^n X^\top.$$

The determinant of $Q(\mathcal{D})$ can then be estimated by $\det[Q_n(\mathcal{D})]$. At significance level $\alpha\in(0,1)$, our test rejects the null hypothesis of constant jump beta if $\det[Q_n(\mathcal{D})] > cv_n^\alpha$ for some sequence cv_n^α of critical values. Before specifying the critical value cv_n^α, we first discuss the asymptotic behavior of $\det[Q_n(\mathcal{D})]$, for which we need some notation. Let $(\kappa_p, \xi_{p-}, \xi_{p+})_{p\geq 1}$ be a collection of mutually independent random variables which are also independent of $\mathcal{F}$, such that κ_p is uniformly distributed on the unit interval and both ξ_{p-} and ξ_{p+} are bivariate standard normal variables. For each $p\geq 1$, we define a 2-dimensional vector R_p as

$$R_p \equiv \sqrt{\kappa_p}\sigma_{\tau_p -}\xi_{p-} + \sqrt{1-\kappa_p}\sigma_{\tau_p}\xi_{p+}. \tag{3.2}$$

The stable convergence in law of $Q_n(\mathcal{D})$ follows from a straightforward adaptation of Theorem 13.1.1 in Jacod and Protter (2012). In particular, we have

$$\Delta_n^{-1/2}\left(Q_n\left(\mathcal{D}\right)-Q\left(\mathcal{D}\right)\right)\overset{\mathcal{L}\text{-}s}{\longrightarrow}\sum_{p\in\mathcal{P}_{\mathcal{D}}}\left(\Delta X_{\tau_p}R_p^{\top}+R_p\Delta X_{\tau_p}^{\top}\right), \tag{3.3}$$

and by the delta-method,[6]

$$\Delta_n^{-1/2}\left(\det\left[Q_n\left(\mathcal{D}\right)\right]-\det\left[Q\left(\mathcal{D}\right)\right]\right)\overset{\mathcal{L}\text{-}s}{\longrightarrow}2\mathrm{Tr}\left[Q\left(\mathcal{D}\right)^{\#}\sum_{p\in\mathcal{P}_{\mathcal{D}}}\Delta X_{\tau_p}R_p^{\top}\right]. \tag{3.4}$$

However, it is important to note that the limiting variable in the above convergence is degenerate under the null hypothesis of constant jump beta. Indeed, in restriction to $\Omega_0\left(\mathcal{D}\right)$,

$$Q\left(\mathcal{D}\right)^{\#}\sum_{p\in\mathcal{P}_{\mathcal{D}}}\Delta X_{\tau_p}R_p^{\top}=Q_{ZZ}\left(\mathcal{D}\right)\begin{pmatrix}\beta_0^2 & -\beta_0\\ -\beta_0 & 1\end{pmatrix}\begin{pmatrix}1\\ \beta_0\end{pmatrix}\sum_{p\in\mathcal{P}_{\mathcal{D}}}\Delta Z_{\tau_p}R_p^{\top}=0. \tag{3.5}$$

Therefore, the standard convergence result (3.4) at the "usual" $\Delta_n^{-1/2}$ rate is not enough for characterizing the asymptotic null distribution of our test statistic.

The novel technical component underlying our testing result is to characterize the nondegenerate asymptotic null distribution of $\det\left[Q_n\left(\mathcal{D}\right)\right]$ at a faster rate Δ_n^{-1}, as detailed in Theorem 1 below. The limiting distribution is characterized by an $\mathcal{F}$-conditional law and the critical value cv_n^{α} should consistently estimate its conditional $(1-\alpha)$-quantile. Since the null asymptotic distribution is highly nonstandard, its conditional quantiles cannot be written in closed form. Nevertheless, the critical values can be easily determined via simulation which we now explain.

To construct the critical value cv_n^{α}, we need to approximate the spot covariance matrix around each jump time. To this end, we pick a sequence k_n of integers such that

$$k_n\to\infty \quad \text{and} \quad k_n\Delta_n\to 0. \tag{3.6}$$

We also pick a $\mathbb{R}^2$-valued sequence v_n' of truncation threshold that satisfies

$$\left\|v_n'\right\|\asymp\Delta_n^{\varpi}\text{ for some constant }\varpi\in(0,1/2). \tag{3.7}$$

Let $\mathcal{I}_n'\left(\mathcal{D}\right)=\{i\in\mathcal{I}_n\left(\mathcal{D}\right):k_n+1\leq i\leq\lfloor T/\Delta_n\rfloor-k_n\}$. For each $i\in\mathcal{I}_n'\left(\mathcal{D}\right)$, we approximate the pre-jump and the post-jump spot covariance matrices respectively by

$$\hat{c}_{n,i+} = \frac{1}{k_n\Delta_n}\sum_{j=1}^{k_n}(\Delta_{i+j}^nX)(\Delta_{i+j}^nX)^{\top}1_{\left\{-v_n'\leq\Delta_{i+j}^nX\leq v_n'\right\}}, \tag{3.8}$$

$$\hat{c}_{n,i-} = \frac{1}{k_n\Delta_n}\sum_{j=0}^{k_n-1}(\Delta_{i-k_n+j}^nX)(\Delta_{i-k_n+j}^nX)^{\top}1_{\left\{-v_n'\leq\Delta_{i+j}^nX\leq v_n'\right\}}. \tag{3.9}$$

[6]Recall that, for a matrix A, the differential of $\det(A)$ is $\mathrm{Tr}\left[A^{\#}dA\right]$, where $A^{\#}$ is the adjoint matrix of A.

Algorithm 1, below, describes how to compute the critical value cv_n^α.

Algorithm 1. (1) Simulate a collection of variables $\left(\tilde{\kappa}_i, \tilde{\xi}_{i-}, \tilde{\xi}_{i+}\right)_{i\in\mathcal{I}'_n(\mathcal{D})}$ consisting of independent copies of $(\kappa_p, \xi_{p-}, \xi_{p+})$. Set for $i \in \mathcal{I}'_n(\mathcal{D})$,

$$\begin{cases} \tilde{R}_{n,i} = \sqrt{\tilde{\kappa}_i}\hat{c}_{n,i-}^{1/2}\tilde{\xi}_{i-} + \sqrt{1-\tilde{\kappa}_i}\hat{c}_{n,i+}^{1/2}\tilde{\xi}_{i+}, \\ \tilde{\varsigma}_{n,i} = \left(-\dfrac{Q_{ZY,n}(\mathcal{D})}{Q_{ZZ,n}(\mathcal{D})}, 1\right)\tilde{R}_{n,i}. \end{cases} \tag{3.10}$$

(2) Compute

$$\tilde{\zeta}_n(\mathcal{D}) \equiv \left(\sum_{i\in\mathcal{I}'_n(\mathcal{D})} \Delta_i^n Z^2\right)\left(\sum_{i\in\mathcal{I}'_n(\mathcal{D})} \tilde{\varsigma}_{n,i}^2\right) - \left(\sum_{i\in\mathcal{I}'_n(\mathcal{D})} \Delta_i^n Z\tilde{\varsigma}_{n,i}\right)^2. \tag{3.11}$$

(3) Generate a large number of Monte Carlo simulations according to step 1 and step 2, and then set cv_n^α as the $(1-\alpha)$-quantile of $\tilde{\zeta}_n(\mathcal{D})$ in the Monte Carlo sample. □

Theorem 1, below, provides the asymptotic justification for the proposed test. To state it, we use the following additional notation: recall R_p from (3.2) and set

$$\varsigma_p \equiv \left(-\frac{Q_{ZY}(\mathcal{D})}{Q_{ZZ}(\mathcal{D})}, 1\right)R_p, \quad p \geq 1. \tag{3.12}$$

Note that, in restriction to $\Omega_0(\mathcal{D})$, we have $\varsigma_p \equiv (-\beta_0, 1)R_p$. It is useful to note that $(\varsigma_p)_{p\geq1}$ are $\mathcal{F}$-conditionally independent. Moreover, each ς_p is a mixture of two $\mathcal{F}$-conditionally Gaussian random variables with possibly distinct conditional variances. The variable ς_p becomes $\mathcal{F}$-conditionally Gaussian when $\Delta c_{\tau_p} = 0$.

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

(a) In restriction to $\Omega_0(\mathcal{D})$, we have

$$\Delta_n^{-1}\det[Q_n(\mathcal{D})] \xrightarrow{\mathcal{L}\text{-}s} \zeta(\mathcal{D}) \equiv \left(\sum_{p\in\mathcal{P}_\mathcal{D}} \Delta Z_{\tau_p}^2\right)\left(\sum_{p\in\mathcal{P}_\mathcal{D}} \varsigma_p^2\right) - \left(\sum_{p\in\mathcal{P}_\mathcal{D}} \Delta Z_{\tau_p}\varsigma_p\right)^2. \tag{3.13}$$

(b) In restriction to $\Omega_0(\mathcal{D}) \cup \Omega_a(\mathcal{D})$, the sequence cv_n^α of variables defined in Algorithm 1 converges in probability to the $\mathcal{F}$-conditional $(1-\alpha)$-quantile of $\zeta(\mathcal{D})$.

(c) The test defined by the critical region $\{\Delta_n^{-1}\det[Q_n(\mathcal{D})] > cv_n^\alpha\}$ has asymptotic size α under the null and asymptotic power one under the alternative, that is,

$$\mathbb{P}\left(\Delta_n^{-1}\det[Q_n(\mathcal{D})] > cv_n^\alpha | \Omega_0(\mathcal{D})\right) \longrightarrow \alpha, \quad \mathbb{P}\left(\Delta_n^{-1}\det[Q_n(\mathcal{D})] > cv_n^\alpha | \Omega_a(\mathcal{D})\right) \longrightarrow 1.$$

Part (a) of Theorem 1 describes the stable convergence of the test statistic $\det[Q_n(\mathcal{D})]$ under the null hypothesis, which occurs at the Δ_n^{-1} convergence rate. The limiting variable $\zeta(\mathcal{D})$ is quadratic in the variables ς_p, which, conditional on $\mathcal{F}$, are mutually independent mixed Gaussian variables.

Comparing (3.11) and (3.13), it is easy to see that $\tilde{\zeta}_n(\mathcal{D})$ is designed to mimic the limiting variable $\zeta(\mathcal{D})$. Part (b) shows that the quantile of the former consistently estimates that of the latter. We note that part (b) holds under both the null and the alternative. Part (c) shows that the proposed test has valid size control and is consistent against general fixed alternatives.

Our test can be equivalently reported in terms of the *realized jump correlation* defined as

$$\rho_n(\mathcal{D}) \equiv \frac{Q_{ZY,n}(\mathcal{D})}{\sqrt{Q_{ZZ,n}(\mathcal{D})\, Q_{YY,n}(\mathcal{D})}}.$$

Observe that

$$\frac{\det[Q_n(\mathcal{D})]}{Q_{ZZ,n}(\mathcal{D})\, Q_{YY,n}(\mathcal{D})} = 1 - \rho_n^2(\mathcal{D}).$$

Therefore, the test rejects the null hypothesis of constant jump beta when $\rho_n^2(\mathcal{D})$ is sufficiently lower than 1, with the critical value for their difference being $\Delta_n cv_n^\alpha / Q_{ZZ,n}(\mathcal{D})\, Q_{YY,n}(\mathcal{D})$. Since the jump correlation coefficient is scale-invariant, its value is easier to interpret and compare across studies than the determinant. For this reason we recommend reporting the test in terms of the jump correlation coefficient in empirical work.

We also remark that our test can be easily extended to test the joint null hypothesis that (2.10) holds on each of a finite number of disjoint regions $(\mathcal{D}_k)_{1\leq k\leq \bar{k}}$, with possibly different betas across regions. To avoid repetition, we only sketch the procedure here. Among many possible choices, one can employ a "sup" test using the test statistic $\Delta_n^{-1}\max_{1\leq k\leq \bar{k}} \det[Q_n(\mathcal{D})]$. By a trivial extension of Theorem 1(a), it can be shown that in restriction to the (joint) null hypothesis $\cap_{1\leq k\leq \bar{k}}\Omega_0(\mathcal{D}_k)$,

$$\left(\Delta_n^{-1}\det[Q_n(\mathcal{D})]\right)_{1\leq k\leq \bar{k}} \xrightarrow{\mathcal{L}\text{-}s} \left(\zeta(\mathcal{D}_k)\right)_{1\leq k\leq \bar{k}},$$

and, hence, the asymptotic null distribution of the test statistic is $\max_{1\leq k\leq \bar{k}} \zeta(\mathcal{D}_k)$. The critical value at significance level α can be obtained by computing the $(1-\alpha)$-quantile of $\max_{1\leq k\leq \bar{k}} \tilde{\zeta}_n(\mathcal{D}_k)$ via simulation.

Finally, we note that an interesting direction for future research is to test (3.1) using other types of statistics. An important example is the Wald statistic, which in a more standard context has been studied by Gourieroux and Jasiak (2013) and further extended by Dufour, Renault, and Zinde-Walsh (2013) in cases with general singularities. In the current setting, the Wald test can be obtained by studentizing the test statistic $\det[Q_n(\mathcal{D})]$ with a preliminary estimator of its scale in the limit (see (3.4)). Analyzing the null asymptotic distribution of the Wald test is more complicated because it also depends on that of the normalizing factor, which in turn involves nonparametric estimators of jumps and the spot volatilities. The full analysis is left to future research.

3.2 Inference when some jumps arrive at deterministic times

In supplemental appendix A we extend the above result to a setting where a subset of jump times can be identified using prior information. Examples of such jump events are the ones caused by pre-scheduled macro announcements (Andersen, Bollerslev, Diebold, and Vega (2003)). In a liquid market, one may expect that price jumps "immediately" after the announcement, so that the announcement time can be used to locate the jump time. Pre-scheduled announcement times are deterministic, which, technically speaking, are excluded from model (2.1) that features random jump arrivals. That being said, inference procedures in this paper can be extended straightforwardly to accommodate fixed jump times. We use the specification test above as an example for a detailed illustration. Modifications to inference procedures in later sections are essentially the same and, hence, will be omitted for brevity.

4 Efficient estimation of jump beta

We continue with the efficient estimation of jump beta under a constant beta model. We first derive an optimally weighted estimator and its asymptotic properties. We then compute the semiparametric efficiency bound for estimating the jump beta and show that this bound is achieved by our optimally weighted estimator. In supplemental appendix A we derive a higher-order expansion for the estimator and use it to construct refined confidence sets for jump betas.

4.1 The optimally weighted estimator

In this subsection, we fix a region $\mathcal{D}$, on which we suppose the constant beta condition (2.10) holds for some true value β_0. Clearly, in order to identify β_0, it is necessary that Z has at least one jump with mark in $\mathcal{D}$. The results below hence are in restriction to the set $\{|\mathcal{P}_{\mathcal{D}}| \geq 1\}$.

We propose a class of estimators of the constant jump beta formed by using weighted jump covariations. To this end, we consider weight functions $w : \mathcal{M}_2 \times \mathcal{M}_2 \times \mathbb{R} \mapsto (0, \infty)$ that satisfy Assumption 3 below.

Assumption 3. *$(c_-, c_+, \beta) \mapsto w(c_-, c_+, \beta)$ is continuous at (c_-, c_+, β_0) for any c_-, $c_+ \in \mathcal{M}_2$.*

With any weight function w, we associate a weighted estimator of the jump beta defined as

$$\hat{\beta}_n\left(\mathcal{D}, w\right) = \frac{\sum_{i \in \mathcal{I}'_n(\mathcal{D})} w(\hat{c}_{n,i-}, \hat{c}_{n,i+}, \tilde{\beta}_n) \Delta_i^n Z \Delta_i^n Y}{\sum_{i \in \mathcal{I}'_n(\mathcal{D})} w(\hat{c}_{n,i-}, \hat{c}_{n,i+}, \tilde{\beta}_n) \left(\Delta_i^n Z\right)^2}, \tag{4.1}$$

where $\tilde{\beta}_n$ is a consistent preliminary estimator for β_0. For concreteness, below, we fix

$$\tilde{\beta}_n \equiv \frac{Q_{ZY,n}\left(\mathcal{D}\right)}{Q_{ZZ,n}\left(\mathcal{D}\right)}, \tag{4.2}$$

which corresponds to no weighting. In Theorem 2 below, we describe the central limit theorem for the weighted estimator $\hat{\beta}_n(\mathcal{D}, w)$. The limiting variable takes the form (recall ς_p from (3.12))

$$\zeta_\beta(\mathcal{D}, w) \equiv \frac{\sum_{p\in\mathcal{P}_\mathcal{D}} w\left(c_{\tau_p-}, c_{\tau_p}, \beta_0\right) \Delta Z_{\tau_p}\varsigma_p}{\sum_{p\in\mathcal{P}_\mathcal{D}} w\left(c_{\tau_p-}, c_{\tau_p}, \beta_0\right) \Delta Z_{\tau_p}^2}. \tag{4.3}$$

It is easy to see from (3.2) and (3.12) that, conditional on $\mathcal{F}$, the limiting variable $\zeta_\beta(\mathcal{D}, w)$ has zero mean with variance

$$\Sigma(\mathcal{D}, w) \equiv \frac{\sum_{p\in\mathcal{P}_\mathcal{D}} w\left(c_{\tau_p-}, c_{\tau_p}, \beta_0\right)^2 \Delta Z_{\tau_p}^2 (-\beta_0, 1)\left(c_{\tau_p-} + c_{\tau_p}\right)(-\beta_0, 1)^\top}{2\left(\sum_{p\in\mathcal{P}_\mathcal{D}} w\left(c_{\tau_p-}, c_{\tau_p}, \beta_0\right) \Delta Z_{\tau_p}^2\right)^2}. \tag{4.4}$$

From here, we shall also show (Theorem 2(b)) that the optimal weight function, in the sense of minimizing the $\mathcal{F}$-conditional asymptotic variance $\Sigma(\mathcal{D}, w)$ among all weight functions, is

$$w^*(\hat{c}_{n,i-}, \hat{c}_{n,i+}, \tilde{\beta}_n) = \frac{2}{(-\tilde{\beta}_n, 1)(\hat{c}_{n,i-} + \hat{c}_{n,i+})(-\tilde{\beta}_n, 1)^\top}. \tag{4.5}$$

Since the variables ς_p are generally not conditionally Gaussian, nor is $\zeta_\beta(\mathcal{D}, w)$. Consequently, a consistent estimator for the conditional asymptotic variance $\Sigma(\mathcal{D}, w)$ is not sufficient for constructing confidence intervals (CI). Instead, we construct CIs for β_0 by approximating the conditional law of $\zeta_\beta(\mathcal{D}, w)$ as described by Algorithm 2 below. For brevity, we focus on two-sided symmetric CIs, while noting that other types of confidence sets can be constructed analogously.

Algorithm 2. (1) Simulate $(\tilde{\varsigma}_{n,i})_{i\in\mathcal{I}'_n(\mathcal{D})}$ as in step 1 of Algorithm 1.

(2) Compute

$$\tilde{\zeta}_{n,\beta}(\mathcal{D}, w) \equiv \frac{\sum_{i\in\mathcal{I}'_n(\mathcal{D})} w\left(\hat{c}_{n,i-}, \hat{c}_{n,i+}, \tilde{\beta}_n\right) \Delta_i^n Z \tilde{\varsigma}_{n,i}}{\sum_{i\in\mathcal{I}'_n(\mathcal{D})} w\left(\hat{c}_{n,i-}, \hat{c}_{n,i+}, \tilde{\beta}_n\right) (\Delta_i^n Z)^2}.$$

(3) Generate a large number of Monte Carlo simulations in the first two steps and set $cv_{n,\beta}^{\alpha/2}$ as the $(1-\alpha/2)$-quantile of $\tilde{\zeta}_{n,\beta}(\mathcal{D}, w)$ in the Monte Carlo sample. Set the $1-\alpha$ level two-sided symmetric CI as $\mathrm{CI}_n^\alpha = [\hat{\beta}_n(\mathcal{D}, w) - \Delta_n^{1/2} cv_{n,\beta}^{\alpha/2}, \hat{\beta}_n(\mathcal{D}, w) + \Delta_n^{1/2} cv_{n,\beta}^{\alpha/2}]$. □

The asymptotic properties of the estimator $\hat{\beta}_n(\mathcal{D}, w)$ and the confidence interval CI_n^α are described by Theorem 2 below.

Theorem 2. *Under Assumptions 1, 2 and 3, the following hold in restriction to* $\{|\mathcal{P}_\mathcal{D}| \geq 1\}$.

(a) We have $\Delta_n^{-1/2}\left(\hat{\beta}_n(\mathcal{D}, w) - \beta_0\right) \xrightarrow{\mathcal{L}\text{-}s} \zeta_\beta(\mathcal{D}, w)$. *If, in addition, the process* $(c_t)_{t\geq 0}$ *does not jump at the same time as* $(Z_t)_{t\geq 0}$, *then the limiting distribution is mixed Gaussian:*

$$\Delta_n^{-1/2}\left(\hat{\beta}_n(\mathcal{D}, w) - \beta_0\right) \xrightarrow{\mathcal{L}\text{-}s} \mathcal{MN}(0, \Sigma(\mathcal{D}, w)). \tag{4.6}$$

(b) $\Sigma(\mathcal{D}, w^*) \leq \Sigma(\mathcal{D}, w)$ *for any weight function* w.

(c) The sequence CI_n^α *described in Algorithm 2 has asymptotic level* $1-\alpha$, *that is,*

$$\mathbb{P}(\beta_0 \in CI_n^\alpha) \to 1-\alpha. \tag{4.7}$$

Part (a) shows the central limit theorem for the estimator $\hat{\beta}_n(\mathcal{D}, w)$ at the parametric rate $\Delta_n^{-1/2}$. It is interesting to note that the two building blocks of $\hat{\beta}_n(\mathcal{D}, w)$, i.e., $Q_{ZY,n}(\mathcal{D}, w)$ and $Q_{ZZ,n}(\mathcal{D}, w)$, converge only at a slower rate. Indeed, their sampling error is driven by that in $(\hat{c}_{n,i-}, \hat{c}_{n,i+})$, the optimal convergence rate of which is $\Delta_n^{-1/4}$; see Theorem 3.2 of Jacod and Todorov (2010). Part (a) also shows that $\hat{\beta}_n(\mathcal{D}, w)$ has an $\mathcal{F}$-conditionally Gaussian asymptotic distribution in the absence of price-volatility co-jumps. Part (b) shows that $w^*(\cdot)$ minimizes the $\mathcal{F}$-conditional asymptotic variance. Part (c) shows that CI_n^α is asymptotically valid.

We refer to the estimator associated with the optimal weight function, i.e., $\hat{\beta}_n(\mathcal{D}, w^*)$, as the *optimally weighted estimator*. The corresponding $\mathcal{F}$-conditional asymptotic variance is

$$\Sigma(\mathcal{D}, w^*) = \left(\sum_{p \in \mathcal{P}_\mathcal{D}} \frac{2\Delta Z_{\tau_p}^2}{(-\beta_0, 1)\left(c_{\tau_p-} + c_{\tau_p}\right)(-\beta_0, 1)^\top} \right)^{-1}. \tag{4.8}$$

It is instructive to illustrate the efficiency gain of the optimally weighted estimator with respect to the unweighted estimator $\tilde{\beta}_n(\mathcal{D})$. Up to asymptotically negligible boundary terms, the latter is equivalent to $\hat{\beta}_n(\mathcal{D}, w_1)$ for $w_1(\cdot) = 1$ identically. Using the Cauchy–Schwarz inequality, it can be shown that $\Sigma(\mathcal{D}, w^*) \leq \Sigma(\mathcal{D}, w_1)$ and the equality holds if and only if the variables $(-\beta_0, 1)\left(c_{\tau_p-} + c_{\tau_p}\right)(-\beta_0, 1)^\top$ are constant across $p \in \mathcal{P}_\mathcal{D}$, that is, under a homoskedasticity-type condition. When this condition is violated, the efficiency gain of the optimally weighted estimator relative to the unweighted estimator is strict.

In a follow-up work in Li, Todorov, and Tauchen (2016b), we have extended the OLS-type estimator $\tilde{\beta}_n(\mathcal{D})$ to a whole class of M-estimators which in particular allows for robust type estimation of β via quantile regressions. The extension of Li, Todorov, and Tauchen (2016b), however, does not allow for weighting of the different observations which is the source of the efficiency gains of the optimally weighted estimator derived in Theorem 2.

4.2 The semiparametric efficiency of the optimally weighted estimator

In the previous subsection, we constructed the optimally weighted estimator as the most efficient estimator within a class of weighted estimators. We now compute the semiparametric efficiency bound for estimating the jump beta under some additional simplifications on the data generating process; see Assumptions 4 and 5 below. We stress from the outset that these assumptions are only needed for this subsection. We further show that the optimally weighted estimator attains this efficiency bound and, hence, is semiparametrically efficient. To simplify the discussion, we fix $\mathcal{D} = [0, T] \times \mathbb{R}_*$ throughout this subsection, while noting that the extension to multiple regions only involves notational complications.

We note that the current setting is very nonstandard in comparison with the classical setting for studying semiparametric efficiency (see, e.g., Bickel, Klaassen, Ritov, and Wellner (1998)), which mainly concerns independent and identically distributed data. By contrast, the current setting is non-ergodic, where asymptotic distributions are characterized as $\mathcal{F}$-conditional laws which depend on the realized values of the stochastic volatility and the jump processes. Since these processes are time-varying, an essentially arbitrary form of data heterogeneity needs to be accommodated. In view of these nonstandard features, it appears necessary to develop the semiparametric efficiency bound for estimating the jump beta from first principles. Our approach relies on the specific structure of the problem at hand but should be a useful start for a more general theory in the spirit of Bickel, Klaassen, Ritov, and Wellner (1998).

Our approach is outlined as follows. We first construct a class of parametric submodels which pass through the true model. We show that these submodels satisfy the LAMN property. Unlike the LAN setting, the information matrix in the LAMN setting is random. By results in Jeganathan (1982, 1983), the inverse of the random information matrix provides an information bound for estimating β. We then compute a lower efficiency bound as the supremum of the Cramer-Rao bound for estimating β over this class of submodels. Since the class of submodels under consideration do not exhaust all possible smooth parametric submodels, it is possible that this supremum is lower than the semiparametric efficiency bound. We rule out this possibility by verifying that this lower efficiency bound is sharp. Indeed, the optimally weighted estimator attains this bound. From here, we conclude that the optimally weighted estimator is semiparametrically efficient. The key to our approach is the construction of a class of submodels that contains, in a well-defined sense, the least favorable submodel.

We now proceed with the details. Below, we denote by P_θ^n the joint distribution of the data sequence $(\Delta_i^n X)_{1\leq i\leq n}$, in a parametric model with an unknown parameter $\theta \in \mathbb{R}^{d_\theta}$. The sequence (P_θ^n) is said to satisfy the LAMN property at $\theta = \theta_0$ if there exist a sequence Γ_n of $d_\theta \times d_\theta$ a.s. positive semidefinite matrices and a sequence ψ_n of d_θ-vectors, such that, for any $h \in \mathbb{R}^{d_\theta}$,

$$\log \frac{dP^n_{\theta_0+\Delta_n^{1/2}h}}{dP^n_{\theta_0}} = h^\top \Gamma_n^{1/2}\psi_n - \frac{1}{2}h^\top \Gamma_n h + o_p(1), \tag{4.9}$$

and

$$(\psi_n, \Gamma_n) \xrightarrow{\mathcal{L}} (\psi, \Gamma), \tag{4.10}$$

where the information matrix Γ is a $d_\theta \times d_\theta$ positive semidefinite $\mathcal{F}$-measurable random matrix and ψ is a d_θ-dimensional standard normal variable independent of Γ.

In order to establish the asymptotic behavior of the log likelihood ratio, we maintain the following assumption in this subsection.

Assumption 4. *We have Assumption 1 and the processes $(b_t)_{t\geq 0}$, $(\sigma_t)_{t\geq 0}$ and $(J_t)_{t\geq 0}$ are independent of $(W_t)_{t\geq 0}$, and the joint law of $(b, \sigma, J_Z, \epsilon)$ does not depend on β.*

Assumption 4 allows for a closed-form expression for the likelihood ratio. Since the law of $(b, \sigma, J_Z, \epsilon)$ does not depend on β, it does not determine the likelihood ratio. Moreover, conditional on these processes, the returns $(\Delta_i^n X)_{1 \leq i \leq n}$ are independent with (non-identical) marginal distribution

$$\Delta_i^n X | b, \sigma, J \sim \mathcal{N}\left(\int_{(i-1)\Delta_n}^{i\Delta_n} b_s ds + \begin{pmatrix} \Delta_i^n J_Z \\ \beta \Delta_i^n J_Z + \Delta_i^n \epsilon \end{pmatrix}, \int_{(i-1)\Delta_n}^{i\Delta_n} c_s ds \right). \tag{4.11}$$

Assumption 4 greatly simplifies our analysis because, otherwise, the closed-form expression for transition densities are unavailable for general stochastic differential equations. We stress that we only need this sufficient condition for the analysis of semiparametric efficiency, while the testing and estimation results in Sections 3 and 4 are valid under settings that are far more general. Finally, we remark that independence-type assumption have also been used by Reiß (2011) and Renault, Sarisoy, and Werker (2016) in the study of efficient estimation of integrated volatility functionals.

In order to ensure that the estimators are asymptotically $\mathcal{F}$-conditionally Gaussian, we restrict our analysis to the case without price-volatility co-jumps (recall Theorem 2(a)).

Assumption 5. *The process $(c_t)_{t \geq 0}$ does not jump at the same time as the process $(Z_t)_{t \geq 0}$ a.s.*

We now proceed to constructing a class of parametric submodels which pass through the original model. We do so by perturbing multiplicatively the jump process J_Z by a step function with known, but arbitrary, break points and unknown step sizes. Each set of break points corresponds to a submodel in which the unknown step sizes play the role of nuisance parameters for the estimation of β. More precisely, we denote the collection of break points by

$$\mathbf{S} \equiv \left\{ S = (S_j)_{0 \leq j \leq m} : 0 = S_0 < \cdots < S_m = T,\ m \geq 1 \right\}. \tag{4.12}$$

Note that each vector $S \in \mathbf{S}$ specifies $\dim(S) - 1$ steps with the form $(S_{j-1}, S_j]$. With any $S \in \mathbf{S}$, we associate the following parametric model: for some unknown parameter $\eta \in \mathbb{R}^{\dim(S)-1}$,

$$dX_t = b_t dt + \sigma_t dW_t + \begin{pmatrix} \eta_j dJ_{Z,t} \\ \beta \eta_j dJ_{Z,t} + d\epsilon_t \end{pmatrix}, \quad \text{for} \quad t \in (S_{j-1}, S_j], \quad 1 \leq j \leq \dim(S) - 1. \tag{4.13}$$

We denote the law of $(\Delta_i^n X)_{1 \leq i \leq n}$ under this model by P_θ^n, where $\theta = (\beta, \eta)$. Below, it is useful to emphasize the dependence of P_θ^n on S by writing $P_\theta^n(S)$. The parametric submodel $(P_\theta^n(S) : \theta \in \mathbb{R}^{\dim(S)})$ is formed by treating $\theta = (\beta, \eta)$ as the unknown parameter and treating the vector S and the law of $(b, \sigma, J_Z, \epsilon)$ as known. Clearly, each submodel passes through the true model at $\theta_0 = \left(\beta_0, \eta_0^\top\right)^\top$, where η_0 is a vector of 1's.

Before stating the formal results, we provide some heuristics to guide intuition concerning the submodels constructed above. To focus on the main idea, we discuss a simple case where both the drift b and the Y-specific jump ϵ are absent, so (4.11) becomes a bivariate Gaussian experiment

$$\Delta_i^n X | b, \sigma, J \sim \mathcal{N}\left(\begin{pmatrix} \Delta_i^n J_Z \\ \beta \Delta_i^n J_Z \end{pmatrix}, \int_{(i-1)\Delta_n}^{i\Delta_n} c_s ds \right). \tag{4.14}$$

It is intuitively clear that the observation $\Delta_i^n X$ contains information for β only when the process Z has a jump during the interval $((i-1)\Delta_n, i\Delta_n]$. The size of each jump of Z can be considered as a nuisance parameter for the estimation of β.[7] Analogous to standard Gaussian location-scale experiments, the estimation of β is not adaptive to the jump size (i.e. location); this is unlike the local covariance matrix $\int_{(i-1)\Delta_n}^{i\Delta_n} c_s ds$, to which the estimation of β is adaptive. Furthermore, it is crucial to treat all jump sizes as separate nuisance parameters because jump sizes are time-varying. Constructing a submodel which captures the heterogeneity in jump sizes would be straightforward in the ideal (but counterfactual) scenario where there are a fixed number of jumps at fixed times. Indeed, any submodel (4.13) would suffice provided that each interval $(S_{j-1}, S_j]$ contains at most one jump time, so that the size of each jump is assigned a nuisance parameter. That being said, the complication here is that both the number of jumps (which is finite but unbounded) and the jump times are actually random. This means, any fixed submodel cannot fully capture the heterogeneity in jump sizes. Therefore, it is important to consider a "sufficiently rich" class of submodels, in the sense that, on every sample path, we can find some submodels in this class that play the role of the least favorable model.

As shown in Theorem 3 below, the parametric submodel $(P_\theta^n(S) : \theta \in \mathbb{R}^{\dim(S)})$ satisfies the LAMN property for each $S \in \mathbf{S}$. To describe the information matrix in each submodel, we need some notation. We define the continuous beta and the spot idiosyncratic variance respectively as

$$\beta_t^c \equiv \frac{c_{ZY,t}}{c_{ZZ,t}} \quad \text{and} \quad v_t^c \equiv c_{YY,t} - \frac{c_{ZY,t}^2}{c_{ZZ,t}}. \tag{4.15}$$

We then set for $t \geq 0$,

$$\begin{cases} \gamma_{1t} = \dfrac{\Delta Z_t^2}{v_t^c} \left(\beta_0 - \beta_t^c\right), \\ \gamma_{2t} = \Delta Z_t^2 \left(\dfrac{(\beta_0 - \beta_t^c)^2}{v_t^c} + \dfrac{1}{c_{ZZ,t}} \right). \end{cases} \tag{4.16}$$

[7]Referring to the jump size as a nuisance "parameter" may be nonstandard, because the jump size is itself an random variable. Note that in the continuous-time limit (i.e., the "population"), the jump process is identified pathwise.

The information matrix for $P_\theta^n(S)$ at $\theta = \theta_0$ is given by

$$\Gamma(S) = \begin{pmatrix} \sum_{s \le T} \frac{\Delta Z_s^2}{v_s^c} & \sum_{S_0 < s \le S_1} \gamma_{1s} & \cdots & \sum_{S_{m-1} < s \le S_m} \gamma_{1s} \\ \sum_{S_0 < s \le S_1} \gamma_{1s} & \sum_{S_0 < s \le S_1} \gamma_{2s} & & \mathbf{0} \\ \vdots & & \ddots & \\ \sum_{S_{m-1} < s \le S_m} \gamma_{1s} & \mathbf{0} & & \sum_{S_{m-1} < s \le S_m} \gamma_{2s} \end{pmatrix}. \tag{4.17}$$

We note that the nonsingularity of the (nonrandom) information matrix is typically imposed for a regular parametric submodel in the LAN setting; see, for example, Definition 1 in Section 2.1 of Bickel, Klaassen, Ritov, and Wellner (1998). In the LAMN setting, the information matrix is random, so this type of regularity generally depends on the realization. Therefore, for each $S \in \mathbf{S}$, we consider the set

$$\Omega(S) \equiv \{\Gamma(S) \text{ is nonsingular}\}. \tag{4.18}$$

By applying the conditional convolution theorem (Theorem 3, Jeganathan (1982)) in restriction to $\Omega(S)$, the efficiency bound for estimating θ is given by $\Gamma(S)^{-1}$.

We now present the main theorem of this subsection. Below, a nonrandom vector $S \in \mathbf{S}$ is said to *shatter* the jump times on a sample path if each time interval $(S_{j-1}, S_j]$ contains exactly one jump. It is useful to note that $\Gamma(S)$ is nonsingular whenever S shatters the jumps of $(Z_t)_{0 \le t \le T}$.

Theorem 3. *Under Assumptions 4 and 5, the following statements hold.*

(a) For each $S \in \mathbf{S}$, the sequence $(P_\theta^n(S) : \theta \in \mathbb{R}^{\dim(S)})$ satisfies the LAMN property at $\theta = \theta_0$ with information matrix $\Gamma(S)$. In restriction to $\Omega(S)$, the information bound for estimating β, that is, the first diagonal element of $\Gamma(S)^{-1}$, has the form

$$\bar{\Sigma}_\beta(S) = \left(\sum_{s \le T} \frac{\Delta Z_s^2}{v_s^c} - \sum_{j=1}^{\dim(S)-1} \frac{\left(\sum_{S_{j-1} < s \le S_j} \gamma_{1s}\right)^2}{\sum_{S_{j-1} < s \le S_j} \gamma_{2s}} \right)^{-1}. \tag{4.19}$$

(b) We have

$$\sup_{S \in \mathbf{S}} \bar{\Sigma}_\beta(S) 1_{\Omega(S)} = \Sigma^*, \tag{4.20}$$

where Σ^* is given by (4.8) with $\mathcal{D} = [0, T] \times \mathbb{R}_*$. Moreover, on each sample path, the supremum is attained by any S that shatters the jump times of the process $(Z_t)_{0 \le t \le T}$.

The key message of Theorem 3 is part (b), which shows that the lower efficiency bound (i.e., $\sup_{S \in \mathbf{S}} \bar{\Sigma}_\beta(S) 1_{\Omega(S)}$) for estimating β among the aforementioned class of submodels is attained by the optimally weighted estimator. We remind the reader that the asymptotic property of the optimally weighted estimator (Theorem 2) is valid in a general setting without imposing the parametric submodel. In other words, the lower efficiency bound derived for these submodels is sharp and the optimally weighted estimator is semiparametrically efficient. Part (b) also shows that the lower efficiency bound is attained by submodels with a sufficiently rich and properly located set of break points (collected by S) which can shatter the realized jump times. In this sense, the least favorable submodel is implicitly chosen in a "random" manner in the sense that it depends on the realization of jump times. A similar phenomenon also arises in the efficient estimation of volatility functionals (see Renault, Sarisoy, and Werker (2016)).

Part (a) of Theorem 3 confirms the intuition that the estimation of β is generally not adaptive to the (unobservable) jumps of Z. Indeed, we see from (4.17) that in the absence of the nuisance parameter η, the Cramer-Rao bound for estimating β is

$$\bar{\Sigma}_{\beta}^{\mathrm{a}} \equiv \left(\sum_{s \leq T} \frac{\Delta Z_s^2}{v_s^c} \right)^{-1}, \tag{4.21}$$

where we use the superscript "a" to indicate adaptiveness, because $\bar{\Sigma}_{\beta}^{\mathrm{a}}$ is the information bound for estimating β in the parametric model where the only unknown parameter is β. From Theorem 3, we also see that Σ^* can be written as

$$\Sigma^* = \left(\sum_{s \leq T} \left(\frac{\Delta Z_s^2}{v_s^c} - \frac{\gamma_{1s}^2}{\gamma_{2s}} \right) \right)^{-1}. \tag{4.22}$$

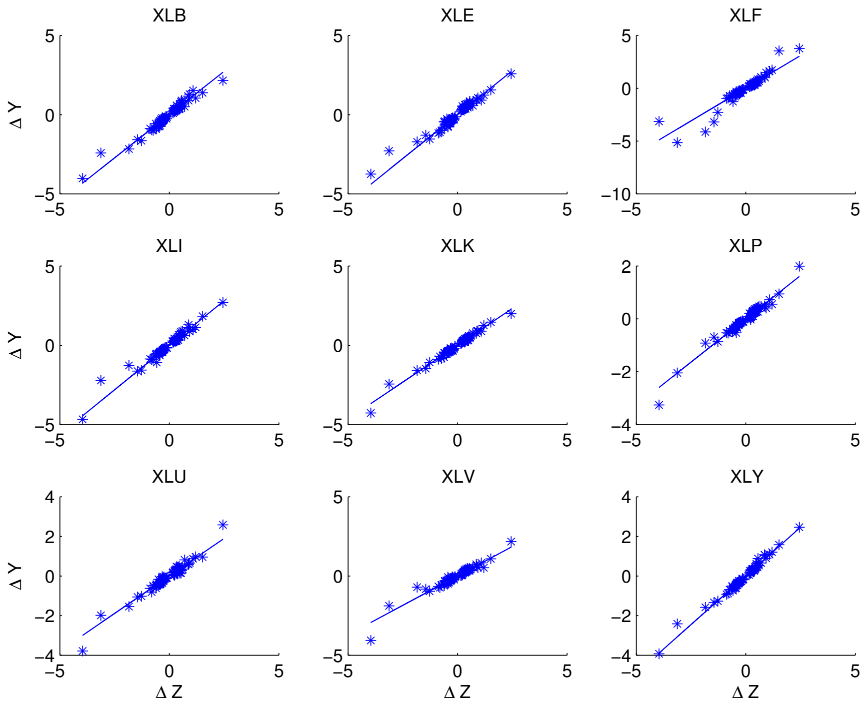
Comparing (4.21) and (4.22), it is clear that $\bar{\Sigma}_{\beta}^{\mathrm{a}} \leq \Sigma^*$, where the equality holds if and only if the process $(\gamma_{1t})_{t \geq 0}$ is identically zero over $[0, T]$. Observe that the latter condition amounts to saying that $\beta_t^c = \beta_0$ whenever $\Delta Z_t \neq 0$. In other words, Σ^* coincides with the adaptive bound $\bar{\Sigma}_{\beta}^{\mathrm{a}}$ only when the continuous beta is equal to the constant jump beta at all jump times of Z. From a practical point of view, this condition appears to be rather peculiar. A stronger, but arguably more natural, restriction is to assume that the continuous beta process β^c coincides with the constant jump beta over the entire time span $[0, T]$. But this additional restriction can be exploited to improve the semiparametric efficiency bound for estimating the common (i.e., continuous and jump) beta. It can be shown that under this stronger assumption, adaptive estimation for the common beta can be achieved.[8]

5 Empirical application

The application concerns betas on market jumps, with the market proxy being the ETF that tracks the S&P 500 index (ticker symbol: SPY). The assets we study are the ETFs on the nine industry portfolios comprising the S&P 500 index: materials (XLB), energy (XLE), financials

[8]Formal results for the adaptive estimation of beta under the condition $\beta_t^c = \beta_0$, $t \in [0, T]$, are presented in supplement appendix A.

Figure 2: Scatter of Jumps: Industry Portfolios



(XLF), industry (XLI), technology (XLK), consumer staples (XLP), utilities (XLU), healthcare (XLV), and consumer discretionary (XLY).

Data on each series are sampled at the 10-minute frequency over the period 2007–2012, resulting in $1,746$ days of 38 within-day returns (log-price increments). By using 10-minute sampling on liquid assets we essentially eliminate the impact of biases due to various microstructure effects. We set the truncation threshold exactly as in the Monte Carlo (presented in supplemental appendix B), with further correction for the well-known deterministic diurnal pattern in volatility.[9] The block size is set to $k_n = 19$.

Figures 2 displays the scatter plots of the detected jump increments of the various assets against those of the market index. The figure also shows the fit provided by the linear jump regression model (1.2) based on the optimally weighted estimator. Perhaps surprisingly, the fit appears generally quite tight, despite the tail nature of jumps and the fact that the sample spans both tranquil and turbulent market environments. That noted, there is lack of fit on the left tail for certain assets. Table 1 reports summary statistics for the linear jump beta regressions over the full sample. As seen from the table, the confidence intervals for beta are relatively tight which further confirms the high precision with which we can estimate jump betas. It is also interesting to note that the average volatility of the residual $Y - \beta Z$ at the jump times of Z is higher than its value at the times immediately preceding the jumps. This provides evidence for volatility jumps at the

[9]We use the procedure detailed in the supplemental material of Todorov and Tauchen (2012).

Table 1: Jump Betas and Tests for Constancy over the Full Sample

Asset	$\widehat{\beta}$	95% CI	$\widehat{\sigma}_{\tau_{p-}}$	$\widehat{\sigma}_{\tau_p}$	R^2	p-val
XLB	1.0920	[1.0525 1.1315]	0.5775	0.5928	0.9614	0.0000
XLE	1.1093	[1.0669 1.1518]	0.7025	0.7283	0.9592	0.0111
XLF	1.2378	[1.1829 1.2926]	0.6515	0.7124	0.8875	0.0000
XLI	1.1225	[1.0918 1.1533]	0.3580	0.3989	0.9548	0.0000
XLK	0.9295	[0.9032 0.9559]	0.3753	0.3956	0.9800	0.0004
XLP	0.6546	[0.6270 0.6823]	0.3916	0.3735	0.9633	0.0003
XLU	0.7574	[0.7146 0.8001]	0.5706	0.6276	0.9534	0.0201
XLV	0.7425	[0.7120 0.7730]	0.3721	0.3991	0.9305	0.0000
XLY	0.9829	[0.9555 1.0102]	0.3949	0.4066	0.9821	0.0012

Note: The columns show the estimated jump beta, the 95% confidence interval (CI), the average level of volatility of $Y - \beta Z$ pre- and post-market jump, R^2 of the regression, and the p-values for the null hypothesis of a constant linear jump regression model for the period 2007–2012.

time of the price jumps. The second to last column of Table 1, which reports the R^2s, confirms our observation of the good fit provided by the linear jump regression. Despite the apparently good fit, the formal test for constancy of the jump beta rejects the null for all but two of the assets in our sample at all conventional levels of the test; see the last column of Table 1 for p-values of these tests. The deviations from linearity observed in Figure 2 are thus in most cases strongly statistically significant.

Of course, as suggested by Figure 1 discussed in the introduction, the jump regression fits can probably be further stabilized when the regressions are run over a shorter period such as one year. This is consistent with the conditional asset pricing models in which betas change over time (see, e.g., Hansen and Richard (1987)). We hence perform the jump regressions year by year, with results from the tests for the constant linear specification reported in Table 2. Allowing for beta to change over years improves the performance of the jump regression. Indeed, the constant jump beta hypothesis is not rejected at the conventional 1% significance level in the majority of cases.

The preceding analysis illustrates that a linear jump regression model works well over periods of years in capturing the dependence between jumps in industry portfolios on one hand and the market jumps on the other hand. The analysis here can be extended to allow for different betas depending on the sign and size of the market jump. It can be further expanded to include a larger set of systematic risk factors (in addition to the market portfolio) and a larger set of test assets. Overall, the tools developed in the paper should prove useful in studying jump dependence which is a key building block in the analysis of pricing of jump risk in the cross-section.

Table 2: Tests for Constant Jump Beta over Years

Asset	Year					
	2007	2008	2009	2010	2011	2012
XLB	0.030	0.016	0.015	0.085	0.007	0.049
XLE	0.539	0.421	0.090	0.064	0.426	0.027
XLF	0.000	0.133	0.019	0.009	0.029	0.071
XLI	0.002	0.000	0.597	0.006	0.033	0.000
XLK	0.001	0.058	0.280	0.004	0.261	0.077
XLP	0.043	0.015	0.008	0.009	0.343	0.002
XLU	0.533	0.782	0.047	0.209	0.061	0.022
XLV	0.000	0.027	0.004	0.022	0.260	0.000
XLY	0.022	0.038	0.020	0.291	0.173	0.267
	Number of jumps within year					
	15	8	9	12	10	20

Note: The table reports p-values of the test for constant linear jump regression model for every asset and every year in the sample.

References

Andersen, T. G., T. Bollerslev, F. X. Diebold, and C. Vega (2003): "Micro Effects of Macro Announcements: Real-Time Price Discovery in Foreign Exchange," *American Economic Review*, 93, 38–62.

Anderson, T. W. (1951): "Linear Restrictions on Regression Coefficients for Multivariate Normal Distributions," *The Annals of Mathematical Statistics*, 22(3), 327–351.

——— (2002): "Reduced Rank Regression in Cointegrated Models," *Journal of Econometrics*, 106(2), 203 – 216.

Ang, A., J. Chen, and Y. Xing (2006): "Downside Risk," *Review of Financial Studies*, 19(4), 1191 – 1239.

Barndorff-Nielsen, O. E., and N. Shephard (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., and N. Shephard (2004b): "Power and Bipower Variation with Stochastic Volatility and Jumps," *Journal of Financial Econometrics*, 2, 1–37.

Barndorff-Nielsen, O. E., and N. Shephard (2006): "Econometrics of Testing for Jumps in Financial Economics Using Bipower Variation," *Journal of Financial Econometrics*, 4, 1–30.

Bickel, P. J., C. A. J. Klaassen, Y. Ritov, and J. A. Wellner (1998): *Efficient and Adapative Estimation for Semiparametric Models.* New York: Springer-Verlag.

Clément, E., S. Delattre, and A. Gloter (2013): "An Infinite Dimensional Convolution Theorem with Applications to the Efficient Estimation of the Integrated Volatility," *Stochastic Processes and their Applications*, 123, 2500–2521.

Dufour, J.-M., E. Renault, and V. Zinde-Walsh (2013): "Wald Tests when Restrictions are Locally Singular," Discussion paper, Brown University.

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

Gourieroux, C., and J. Jasiak (2013): *Size Distortion in the Analysis of Volatility and Covolatility Effects*.pp. 91 – 118. Advances in Intelligent Systems and Computing, vol. 200. New York and Heidelberg: Springer, CEPREMAP and U Toronto.

Hansen, L. P., and S. F. Richard (1987): "The Role of Conditioning Information in Deducing Testable Restrictions Implied by Dynamic Asset Pricing Models.," *Econometrica*, 55(3), 587 – 613.

Jacod, J. (2008): "Asymptotic Properties of Power Variations and Associated Functionals of Semimartingales," *Stochastic Processes and their Applications*, 118, 517–559.

Jacod, J., and M. Podolskij (2013): "A Test for the Rank of the Volatility Process: the Random Perturbation Approach," *Annals of Statistics*, 41, 2391–2427.

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

Jacod, J., and M. Rosenbaum (2013): "Quarticity and Other Functionals of Volatility: Efficient Estimation," *Annals of Statistics*, 41, 1462–1484.

Jacod, J., and V. Todorov (2010): "Do Price and Volatility Jump Together?," *Annals of Applied Probability*, 20, 1425–1469.

Jeganathan, P. (1982): "On the Asymptotic Theory of Estimation when the Limit of the Log-likelihood is Mixed Normal," *Sankhya, Ser. A*, 44, 173–212.

——— (1983): "Some Asymptotic Properties of Risk Functions When the Limit of the Experiment Is Mixed Normal," *Sankhya: The Indian Journal of Statistics, Series A (1961-2002)*, 45(1), 66–87.

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

Lettau, M., M. Maggiori, and M. Weber (2014): "Conditional Risk Premia in Currency Markets and Other Asset Classes," *Journal of Financial Economics*, 114, 197–225.

Lewellen, J., and S. Nagel (2006): "The Conditional CAPM Does Not Explain Asset-Pricing Anomalies," *Journal of Financial Economics*, 82(2), 289 – 314.

Li, J., V. Todorov, and G. Tauchen (2016a): "Adaptive Estimation of Continuous-Time Regression Models using High-Frequency Data," Discussion paper, Duke Univeriesty.

——— (2016b): "Robust Jump Regressions," *Journal of the American Statistical Association, forthcoming.*

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

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

Merton, R. C. (1976): "Option Pricing When Underlying Stock Returns Are Discontinuous," *Journal of Financial Economics*, 3, 125 – 144.

Mykland, P., and L. Zhang (2006): “ANOVA for Diffusions and Ito Processes,” *Annals of Statistics*, 34, 1931–1963.

——— (2009): “Inference for Continuous Semimartingales Observed at High Frequency,” *Econometrica*, 77, 1403–1445.

Reiss, M. (2011): “Asymptotic Equivalence for Inference on the Volatility from Noisy Observations,” *The Annals of Statistics*, 39(2), pp. 772–802.

Renault, E., C. Sarisoy, and B. J. Werker (2016): “Efficient Estimation of Integrated Volatility and Related Processes,” *Econometric Theory, Forthcoming.*

Stein, C. (1956): “Efficient Nonparametric Testing and Estimation,” in *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, pp. 187–195. University of California Press, Berkeley, Calif.

Todorov, V., and T. Bollerslev (2010): “Jumps and Betas: A New Framework for Disentangling and Estimating Systematic Risks,” *Journal of Econometrics*, 157, 220–235.

Todorov, V., and G. Tauchen (2012): “The Realized Laplace Transform of Volatility,” *Economtrica*, 80, 1105–1127.