Jump Regressions*

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

November 2, 2014

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 optimal 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. 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 stability (in time and jump size) 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 Torben G. Andersen, Snehal Banerjee, Ravi Bansal, Tim Bollerslev, Frank X. Diebold, Peter R. Hansen, Andrew Patton, and Brian Weller for helpful discussions.

†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.[1] Bearing such non-diversifiable jump risk is significantly rewarded, as is evident for example from the expensiveness of short-maturity options written on the market index with strikes that are far from its current level.[2] Therefore, precise time series 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.

The goal of the current paper is to develop econometric tools for testing and efficiently estimating the relationship 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 $\Delta Y_\tau = Y_\tau - Y_{\tau-}$ and $\Delta Z_\tau = Z_\tau - Z_{\tau-}$, 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:

$$\Delta Y_\tau = \beta \Delta Z_\tau, \quad \text{for all } \tau \in \mathcal{T}, \tag{1.2}$$

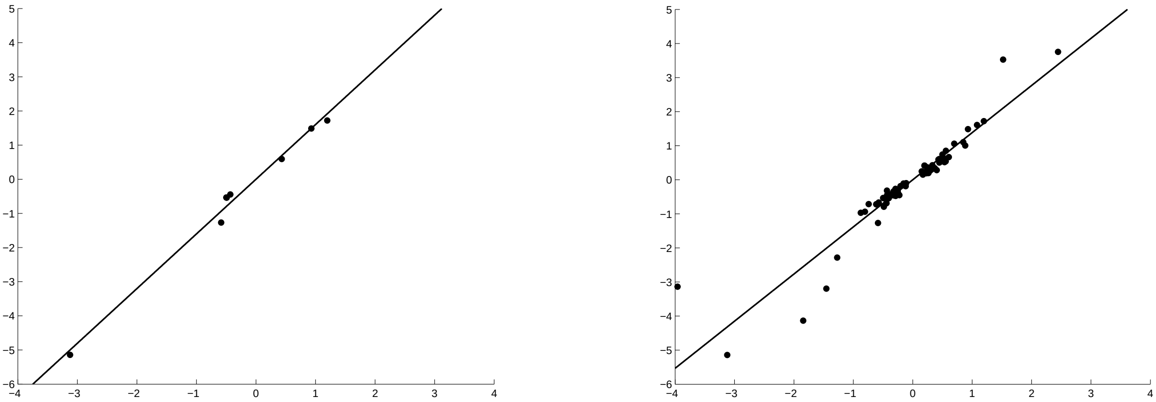
for some constant β. Equation (1.2) amounts to a constancy restriction on the spot jump beta. We can view (1.2) as a linear jump regression model, while noting the important fact that neither the jump time τ nor the jump sizes $(\Delta Y_\tau, \Delta Z_\tau)_{\tau\in\mathcal{T}}$ are directly observable from data sampled at discrete times.

A motivating empirical example of the jump regression is given in Figure 1. From log returns sampled at the 10-minute frequency, we select locally large jump returns of the S&P 500 exchange

[1] There is substantial earlier parametric evidence for jumps, as well as more recent strong nonparametric evidence based on high-frequency data and different jump tests; see, for example, Huang and Tauchen (2005), Barndorff-Nielsen and Shephard (2006), Jiang and Oomen (2008), Lee and Mykland (2008) and Aït-Sahalia and Jacod (2009).

[2] The presence of nontrivial market jump risk premium has been well documented in the empirical option pricing literature.

Figure 1: A Representative Illustration of Jump Regressions



Note: The horizontal axes are the jump returns of the S&P 500 ETF (SPY) while the vertical axes are the contemporaneous returns of the Financial Sector ETF (XLF) 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 6. In the left panel there are two observations with horizontal coordinates -0.4974 and -0.4920 that turn out to be visually indistinguishable in the plot.

traded fund (ETF), which is our proxy for the market, and plot them versus the contemporaneous Financial Sector ETF returns,[3] 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).[4] The fit for the six-year sample (right) appears reasonable, but it is less tight than the former, especially in the tails. Are these patterns statistically consistent with model (1.2)? On one hand, due to the very nature of jumps, jump regressions such as those displayed on Figure 1, 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.

Despite the apparent similarity between the classical linear regression and the jump regression, we emphasize some important distinctions at the outset. First, in the high-frequency setting with

[3]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. The underlying intuition is as follows: a return is considered "locally large", and is subsequently associated with a jump, if its magnitude exceeds a threshold slightly larger than 4 local standard deviations. The diffusive component of the price, which is locally mixed Gaussian, will generate such a large in magnitude 10-minute return about once every three years. Hence, the effect from misclassifying diffusive increments as ones containing jumps is quite negligible for the frequency and threshold used in our analysis.

[4]This good fit is consistent with Andersen, Bollerslev, Diebold, and Vega (2003) who find that in foreign exchange markets news announcements that generate sharp jump-like moves play a key role in price discovery. Therefore, we might expect to see the large moves embodied in a jump regression to reveal best the key economic relationships.

infill asymptotics, unknown parameters are identified from sample paths, instead of distributions, of the studied processes. For this reason, our method only requires some mild pathwise regularity without imposing any stationarity or weak-dependence assumptions. In particular, we make essentially no assumption on the stochastic volatility process; in the jump regression setting, this amounts to an arbitrary form of heteroskedasticity and, hence, leads to a nontrivial complication for efficient estimation. Second, the precision of our inference depends on the precision of recovering (certain functionals of) the price jumps from discretely sampled data. Since jumps are rare events, only a small subsample of observed returns, like those plotted on Figure 1, contain realized jumps that are directly related to model (1.2) and are informative for the inference of jump beta. Third, the $\sqrt{n}$-rate (n is the sample size) asymptotic mixed Gaussianity of our jump beta estimator is driven by the local approximate mixed Gaussianity of the diffusive component of asset returns.[5] The parametric convergence rate is achieved despite the fact that only a small number of observations are informative for jumps. This is unlike the classical econometric setting, as well as the high-frequency setting for volatility estimation, where the $\sqrt{n}$-rate asymptotic (mixed) normality is attained by a central limit theorem for the average of a large number of observations.

We now summarize our main theoretical results. 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

$$\sum_{\tau \in \mathcal{T}} \begin{pmatrix} \Delta Y_\tau^2 & \Delta Y_\tau \Delta Z_\tau \\ \Delta Z_\tau \Delta Y_\tau & \Delta Z_\tau^2 \end{pmatrix}. \tag{1.3}$$

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), 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

[5] By the martingale representation theorem, a continuous local martingale with absolutely continuous quadratic variation can be represented as a stochastic integral of a stochastic volatility process with respect to the Brownian motion. As a result, the diffusive component of the return is approximately mixed Gaussian with the mixing variable being the stochastic volatility.

scaled by (random) jumps and spot volatilities. Since this distribution is highly nonstandard, we further provide a simple simulation-based algorithm to compute the critical value for our test. The proposed test also has a natural interpretation via the realized jump correlation, that is, the correlation coefficient associated with the matrix (1.3). Indeed, the test rejects if and only if the squared realized jump correlation is sufficiently lower than one.

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 (common) assumptions, we derive a semiparametric lower efficiency bound for regular estimators of the jump beta. 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 (2014) study the efficient nonparametric estimation of general integrated volatility functionals, and Li, Todorov, and Tauchen (2014) 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. As well known, the econometric analysis of volatility and jumps require very distinct technical tools. Hence, our semiparametric efficiency analysis differs substantially from the aforementioned work.

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. This finding qualitatively resembles results in the classical Gaussian location model. However, an interesting and distinctive feature of our setting is that the number of these jumps and their locations are random, which means that the submodel that attains the lower efficiency bound, i.e., the least favorable submodel, depends on the realization of the underlying processes.

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 $n^{1/4}$; see, for example, the general results for statistics involving jumps and spot volatility in Jacod and Todorov (2010). The reason for the faster rate in our case is that the spot volatility estimates participate only in the weights in our jump beta estimator and their sampling variability is annihilated in the second-order asymptotics, as is typical for weighted estimators.[6]

To improve finite sample performance, we further derive a novel higher-order expansion for the optimally weighted estimator. This expansion clearly reveals the role of the spot covariance estimates in the efficient estimation of jump beta. 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. Using realistically calibrated numerical examples, we demonstrate that the proposed efficient estimator provides considerable efficiency gains over natural alternatives based on the ratio of the 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)). Indeed, for the frequency and jump size distribution as well as volatility variability as in the data sets included in our empirical analysis, we document reduction in the asymptotic standard deviation of the jump beta estimates of around 40%.

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, three well-known common stocks, and a gold ETF for the period 2007–2012. The premise of our empirical application is the well-known fact that standard market betas are strongly time-varying due to changing conditioning information; see, for example, the work of Barndorff-Nielsen and Shephard (2004a) using high frequency realized betas based on the realized covariance as well as Andersen, Bollerslev, Diebold, and Wu (2006) for a review. 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

[6] The reason for attaining the parametric convergence rate is thus very different for our estimator and the statistics of Jacod and Rosenbaum (2013) who similarly use block-based volatility estimates. For the statistics of Jacod and Rosenbaum (2013), the parametric convergence rate is obtained because the sampling errors of spot variance estimates are averaged out over an asymptotically increasing number of blocks.

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. Motivated by recent evidence on downside risk (see, e.g., Ang, Chen, and Xing (2006) and Lettau, Maggiori, and Weber (2014)), we further examine the possibility of piecewise linear jump regression function which allows for different positive and negative jump betas, and we find that in some cases this generalization can provide further improvement.

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. Sections 5 and 6 present results from a Monte Carlo study and an empirical application respectively. Section 7 concludes. All proofs are in the appendix.

2 Setting and modeling jump dependence

We start with introducing the formal setup for our analysis. Section 2.1 describes the conditions on the studied processes. Section 2.2 introduces the jump regression models. Section 2.3 presents an auxiliary result for the approximation of jumps. The following notations are 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 component-wise. 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 $|\mathcal{P}|$. For any random variable ξ, we use the standard shorthand notation $\{\xi \text{ satisfies some property}\}$ for $\{\omega \in \Omega : \xi(\omega) \text{ satisfies some property}\}$. 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 and the jump arrival times of its individual elements are disjoint, i.e., the jump components of its elements are independent of each other. The reason for this is that the asymptotics developed in the paper are determined solely by the behavior of the processes Y and Z around the jump times of Z. In the above mentioned multivariate setting the individual components of Z jump at different times, and therefore the asymptotic analysis will reduce to the bivariate setting that we consider henceforth.

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 (see, e.g., Jacod and Protter (2012), Section 2.1.4) with the form

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

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 positive definite 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 $\nu\left(dt, du\right) = dt \otimes \lambda\left(du\right)$ 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) $\nu\left([0,T] \times \mathbb{R}\right) < \infty$.*

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 the 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.[7]

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 $n \to \infty$.

[7] 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. This, however, comes at the price of losing some efficiency for the analysis of the "big" jumps.

2.2 Piecewise linear jump regression models

We proceed with introducing the jump regression models which concern the behavior of Y at the jump times of Z. 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. In a fully nonparametric setting, with the spot jump beta given by (1.1), we have

$$\Delta Y_\tau = \beta_\tau \Delta Z_\tau, \quad \tau \in \mathcal{T}. \tag{2.4}$$

We remark that β_τ is only defined, and identified, at the jump times of Z. We also note that Y can jump at times in $[0, T] \setminus \mathcal{T}$. We refer to these jumps as Y-specific jumps which, by definition, do not occur at the same times as the jumps of Z. Finally, we observe that β_τ can take a value of zero which will correspond to the situation in which Z jumps but Y does not.

In general the spot jump beta is a non-predictable function that depends on time as well as the jump size realization, which of course in general cannot be predicted from information prior to its arrival. This non-adaptiveness of β_τ is a unique characteristic of jump risk. Indeed, in a diffusive setting returns are locally Gaussian, and hence we have linear dependence between the increments of Y and Z over short time intervals. Therefore, the continuous beta process which measures the exposure of the diffusive risk of an asset towards the systematic diffusive risk, has trajectories that are predictable functions in time unlike those of the jump beta process.

Our interest in this paper is to develop inference techniques for situations where the spot jump beta remains constant over regions specified by time and jump size. Before introducing the formal setup, we consider a few motivating examples.

Example 1 (Constant Beta)**.** *The simplest model restriction on (2.4) is to impose that* $\beta_\tau = \beta$ *for some constant* β *and all* $\tau \in \mathcal{T}$*, that is, a constant beta model. This constant beta restriction naturally arises in factor models, commonly used in asset pricing, in which* $J_{Y,t} = \beta J_{Z,t} + \varepsilon_t$ *with* ε_t *denoting the* Y*-specific jump process. We can represent the constant beta model equivalently as*

$$\Delta Y_t = \beta \Delta Z_t + \Delta\varepsilon_t, \quad \Delta Z_t \Delta\varepsilon_t = 0, \quad t \in [0, T]\,. \tag{2.5}$$

Example 2 (Temporal Breaks)**.** *Conditional asset pricing models allow for the exposure of assets to fundamental risks to change over time, see e.g., 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 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 in connection with the structural break models in time series (see Stock (1994) and references therein). More formally, let* $(\mathcal{S}_k)_{1\leq k\leq \bar{k}}$ *be a finite disjoint partition of* $[0,T]$*, which correspond 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, the model can be written as*

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

Example 3 (Spatial Breaks)**.** *An alternative way to generalize the constant beta model is to allow the spot jump beta 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:

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

Our jump regression setting incorporates the above examples as special cases. 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

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

That is, the spot jump beta is a constant for all jumps whose marks are in the region $\mathcal{D}$. Example 1 corresponds to $\mathcal{D} = [0,T]\times\mathbb{R}_*$, and Examples 2 and 3 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:

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

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

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

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 set $\mathcal{I}(\mathcal{D})$ collects the indices of sampling intervals that contain the jumps with marks in $\mathcal{D}$. Clearly, the set $\mathcal{I}(\mathcal{D})$ is random and unobservable. 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* $\partial\mathcal{D}$ *denotes 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 $((A_{j,n})_{1\le j\le N_n})_{n\ge1}$ of random elements, we write $(A_{j,n})_{1\le j\le N_n}\xrightarrow{\mathbb{P}}(A_j)_{1\le j\le N}$ if $\mathbb{P}(N_n=N)\longrightarrow1$ and $(A_{j,n})_{1\le j\le N}1_{\{N_n=N\}}\xrightarrow{\mathbb{P}}(A_j)_{1\le j\le N}$.

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

(a) $\mathbb{P}(\mathcal{I}_n(\mathcal{D})=\mathcal{I}(\mathcal{D}))\to1$;

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

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.[8]

3 Testing for constant jump beta

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, $|\mathcal{P}_\mathcal{D}| \geq 2$. This is fairly innocuous given the abundant evidence for the prevalence of jumps in high-frequency financial data. Formally, the testing problem is to decide in which of the following two sets the observed sample path falls:[9]

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

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 $\det[Q(\mathcal{D})] = 0$.

We now describe the test statistic. In view of Proposition 1, we construct a sample analogue estimator for $Q(\mathcal{D})$ as

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

[8] We remark some finite-sample considerations. Clearly, in finite samples, the estimate $\mathcal{I}_n(\mathcal{D})$ generally does not coincide with $\mathcal{I}(\mathcal{D})$, so scatter plots like Figure 1 are subject to classification errors. In the context of jump regressions, we recommend choosing the threshold v_n conservatively, that is, setting v_n large, so that $\mathcal{I}_n(\mathcal{D})$ is more likely to contain jumps (which model (2.10) is about) instead of large diffusive returns. This will mitigate the finite-sample bias resulting from fitting the jump regression model to diffusive price movements. The cost of doing so is that some small jumps may be discarded, which can potentially lead to efficiency loss in finite samples.

[9] 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; the reference of record is Jacod and Protter (2012), with references and discussions therein.

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[10] of $Q_n(\mathcal{D})$ is well understood from prior work. We have[11]

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

and by the delta-method,[12]

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

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 $\Omega_0(\mathcal{D})$,

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

Therefore, the standard convergence result (3.4) at the "usual" $\Delta_n^{-1/2}$ rate is not useful 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 $\det[Q_n(\mathcal{D})]$ 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.

[10]Stable convergence in law is stronger than the usual notion of weak convergence. It requires that the convergence holds jointly with any bounded random variable defined on the original probability space. Its importance for our problem stems from the fact that the limiting distributions of our estimators are characterized by $\mathcal{F}$-conditional laws and stable convergence allows for feasible inference using consistent estimators for their $\mathcal{F}$-conditional quantiles. See Jacod and Shiryaev (2003) for further details on stable convergence on filtered probability spaces.

[11]The claim can be proved by a straightforward adaptation of Theorem 13.1.1 in Jacod and Protter (2012). Technical details are omitted for brevity, but are available upon request.

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

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 $i \in \mathcal{I}'_n(\mathcal{D})$,

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

2 Compute

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

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.

Algorithm 1: Critical value of the constant jump beta test.

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[13]

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

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

$$\hat{c}_{n,i+} = \frac{1}{k_n\Delta_n}\sum_{j=1}^{k_n}(\Delta_{i+j}^n X)(\Delta_{i+j}^n X)^\top 1_{\{-v'_n \le \Delta_{i+j}^n X \le v'_n\}}, \tag{3.8}$$

$$\hat{c}_{n,i-} = \frac{1}{k_n\Delta_n}\sum_{j=0}^{k_n-1}(\Delta_{i-k_n+j}^n X)(\Delta_{i-k_n+j}^n X)^\top 1_{\{-v'_n \le \Delta_{i+j}^n X \le v'_n\}}. \tag{3.9}$$

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

$$\varsigma_p \equiv \left(-\frac{Q_{ZY}(\mathcal{D})}{Q_{ZZ}(\mathcal{D})}, 1\right)R_p, \quad p \ge 1. \tag{3.10}$$

[13]An asymptotically valid choice of v'_n is (v_n, v_n), where v_n is given by (2.12). In practice, it is useful to allow the truncation threshold to vary across assets.

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 $(\varsigma_p)_{p\geq 1}$ 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

$$\Delta_n^{-1} \det[Q_n(\mathcal{D})] \xrightarrow{\mathcal{L}\text{-}s} \zeta(\mathcal{D}) \equiv \left(\sum_{p\in\mathcal{P}_\mathcal{D}} \Delta Z_{\tau_p}^2\right)\left(\sum_{p\in\mathcal{P}_\mathcal{D}} \varsigma_p^2\right) - \left(\sum_{p\in\mathcal{P}_\mathcal{D}} \Delta Z_{\tau_p}\varsigma_p\right)^2. \tag{3.11}$$

(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 (3.6) and (3.11), 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.

In practice, the test above can be equivalently reported in terms of the *realized jump correlation coefficient* 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.

Finally, we remark that the test described in Theorem 1 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.

4 Efficient estimation for the 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 jump beta and show that this bound is achieved by our optimally weighted estimator. We finish the section with 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. *The function $(c_-,c_+,\beta)\mapsto w(c_-,c_+,\beta)$ 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

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

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 confidence interval (CI) as

$$\mathrm{CI}_n^\alpha = \left[\hat{\beta}_n(\mathcal{D},w) - \Delta_n^{1/2}cv_{n,\beta}^{\alpha/2}, \hat{\beta}_n(\mathcal{D},w) + \Delta_n^{1/2}cv_{n,\beta}^{\alpha/2}\right].$$

Algorithm 2: Confidence intervals for the jump beta.

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

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

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 (3.10))

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

It is easy to see from (3.2) and (3.10) 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. For brevity, we focus on two-sided symmetric CIs, while noting that other types of confidence sets can be constructed analogously. The asymptotic properties of the estimator $\hat{\beta}_n(\mathcal{D},w)$ and CI_n^α (see Algorithm 2) are described by Theorem 2 below.

Theorem 2. *Under Assumptions 1–3, the following statements 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:*

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

(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. We quantify this gain in empirically realistic settings in Section 5 below.

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 involve 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 their joint law does not depend on β.*

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

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

where the process $(\varepsilon_t)_{t\geq 0}$ denotes the Y-specific jumps. 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 (2014) 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 $(Z_t)_{t\geq 0}$ almost surely.*

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 $S \in \mathbf{S}$ 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 $\eta \in \mathbb{R}^{\dim(S)-1}$,

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

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 $(b, \sigma, J_Z, \varepsilon)$ 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.[14] 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.

[14] 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.

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 $t \geq 0$,

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

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

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

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 $(Z_t)_{0 \leq t \leq T}$.

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

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

(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.

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

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

where we use the superscript "a" to indicate adaptiveness, because $\bar{\Sigma}^{\mathrm{a}}_{\beta}$ 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

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

Comparing (4.21) and (4.22), it is clear that $\bar{\Sigma}^{\mathrm{a}}_{\beta} \le \Sigma^*$, where the equality holds if and only if the process $(\gamma_{1t})_{t\ge 0}$ 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 $\bar{\Sigma}^{\mathrm{a}}_{\beta}$ 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.[15]

4.3 Higher-order asymptotics for the optimally weighted estimator

We proceed with designing refined confidence sets of the optimally weighted estimator based on a higher-order asymptotic expansion. To motivate, we observe that while the optimally weighted estimator $\hat{\beta}_n(\mathcal{D}, w^*)$ depends on the spot covariance estimates $(\hat{c}_{n,i-}, \hat{c}_{n,i+})$, the sampling variability of the latter is not reflected in the asymptotic distribution described by Theorem 2. The reason is that these 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 optimally weighted estimator, so we proceed to derive its higher-order expansion. Based on this expansion, we provide a refinement to the confidence interval construction described in Algorithm 2.

We first need some additional regularity on the spot volatility σ, namely that it is an Itô semimartingale.

Assumption 6. *The process σ is an Itô semimartingale of the form*

$$\begin{aligned} vec(\sigma_t) &= vec(\sigma_0) + \int_0^t \tilde{b}_s ds + \int_0^t \tilde{\sigma}_s d\widetilde{W}_s + \int_0^t \int_{\mathbb{R}} \tilde{\delta}(s,u) 1_{\{\|\tilde{\delta}(s,u)\|>1\}} \tilde{\mu}(ds,du) \\ &\quad + \int_0^t \int_{\mathbb{R}} \tilde{\delta}(s,u) 1_{\{\|\tilde{\delta}(s,u)\|\leq 1\}} (\tilde{\mu} - \tilde{\nu})(ds,du), \end{aligned}$$

where the processes $\tilde{b}$ and $\tilde{\sigma}$ are locally bounded and take values respectively in $\mathbb{R}^4$ and $\mathbb{R}^{4\times 4}$, $\widetilde{W}$ is a 4-dimensional Brownian motion, $\tilde{\delta} : \Omega \times \mathbb{R}_+ \times \mathbb{R} \mapsto \mathbb{R}^4$ is a predictable function, $\tilde{\mu}$ is a Poisson random measure with compensator $\tilde{\nu}$ of the form $\tilde{\nu}(dt, du) = dt \otimes \tilde{\lambda}(du)$ for some σ-finite measure $\tilde{\lambda}$. Moreover, there exists a localizing sequence of stopping times $(T_m)_{m\geq 1}$ and $\tilde{\lambda}$-integrable functions $(\tilde{\Gamma}_m)_{m\geq 1}$, such that $\|\tilde{\delta}(\omega, t, u)\|^2 \wedge 1 \leq \tilde{\Gamma}(u)$ for all $\omega \in \Omega$, $t \leq T_m$ and $u \in \mathbb{R}$.

Assumption 6 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 6 allows for volatility jumps, and it does not restrict their activity

[15] Formal results for the adaptive estimation of beta under the condition $\beta_t^c = \beta_0$, $t \in [0, T]$, are available upon request.

and dependence with the price jumps. However, this assumption does rule out certain long-memory volatility models driven by the fractional Brownian motion (see Comte and Renault (1996)).

We now present the higher-order asymptotic expansion for the optimally weighted estimator. We need some additional notation for this. Let v denote the spot variance of the residual $Y - \beta_0 Z$. That is,

$$v_t \equiv (-\beta_0, 1)\, c_t\, (-\beta_0, 1)^\top . \tag{4.23}$$

Further, let $\left(\xi'_{p-}, \xi'_{p+}\right)_{p\geq 1}$ be a collection of mutually independent random variables which are also independent of $\mathcal{F}$ and $(\kappa_p, \xi_{p-}, \xi_{p+})_{p\geq 1}$, such that ξ'_{p-} and ξ'_{p+} are scalar standard normal variables. We set for $p \geq 1$,

$$\phi_p \equiv \frac{(-\beta_0, 1)\left(c_{\tau_p -} + c_{\tau_p}\right)(-\beta_0, 1)^\top}{2}, \quad F_p \equiv \frac{v_{\tau_p -}\xi'_{p-} + v_{\tau_p}\xi'_{p+}}{\sqrt{2}}. \tag{4.24}$$

To guide intuition, we note that the variable F_p captures the sampling variability for approximating the average residual volatility $(v_{\tau_p -} + v_{\tau_p})/2$. We further note that $\phi_p^{-1} = w^*(c_{\tau_p -}, c_{\tau_p}, \beta_0)$, so the limit variable $\zeta_\beta(\mathcal{D}, w^*)$ defined in (4.3) can be rewritten as

$$\zeta_\beta\left(\mathcal{D}, w^*\right) \equiv \frac{\sum_{p\in\mathcal{P}_\mathcal{D}} \Delta Z_{\tau_p} \varsigma_p/\phi_p}{\sum_{p\in\mathcal{P}_\mathcal{D}} \Delta Z_{\tau_p}^2/\phi_p}. \tag{4.25}$$

Theorem 4. *Let $k_n \asymp \Delta_n^{-a}$ for some constant $a \in (0, 1/2)$. Suppose Assumptions 1, 2 and 6 hold for $\varpi \in (a/4, 1/2)$. Then we have the following expansion for the optimally weighted estimator:*

$$\Delta_n^{-1/2}\left(\hat{\beta}_n\left(\mathcal{D}, w^*\right) - \beta_0\right) = \zeta^*_{n,\beta}\left(\mathcal{D}\right) + k_n^{-1/2} H^*_{n,\beta}\left(\mathcal{D}\right) + o_p(k_n^{-1/2}), \tag{4.26}$$

for some sequences of variables $\zeta^*_{n,\beta}(\mathcal{D})$ and $H^*_{n,\beta}(\mathcal{D})$ satisfying

$$\left(\zeta^*_{n,\beta}\left(\mathcal{D}\right), H^*_{n,\beta}\left(\mathcal{D}\right)\right) \xrightarrow{\mathcal{L}\text{-}s} \left(\zeta_\beta\left(\mathcal{D}, w^*\right), H^*_\beta\left(\mathcal{D}\right)\right), \tag{4.27}$$

where

$$H^*_\beta\left(\mathcal{D}\right) \equiv \frac{\left(\sum_{p\in\mathcal{P}} \frac{\Delta Z_{\tau_p}^2 F_p}{\phi_p^2}\right)\left(\sum_{p\in\mathcal{P}} \frac{\Delta Z_{\tau_p}\varsigma_p}{\phi_p}\right) - \left(\sum_{p\in\mathcal{P}} \frac{\Delta Z_{\tau_p}\varsigma_p F_p}{\phi_p^2}\right)\left(\sum_{p\in\mathcal{P}} \frac{\Delta Z_{\tau_p}^2}{\phi_p}\right)}{\left(\sum_{p\in\mathcal{P}} \frac{\Delta Z_{\tau_p}^2}{\phi_p}\right)^2}. \tag{4.28}$$

The leading term $\zeta^*_{n,\beta}(\mathcal{D})$ in (4.26) is what drives the convergence in Theorem 2. The higher-order term $k_n^{-1/2} H^*_{n,\beta}(\mathcal{D})$ is $O_p(k_n^{-1/2})$ and hence is asymptotically dominated by $\zeta^*_{n,\beta}(\mathcal{D})$. The limiting variable $H^*_\beta(\mathcal{D})$ involves both F_p and ς_p, which capture respectively the sampling variability that arise from the estimation of the spot covariance and the estimation of jumps.

1 Simulate $(\tilde{\varsigma}_{n,i})_{i\in\mathcal{I}'_n(\mathcal{D})}$ as in step 1 of Algorithm 1. Simulate $(\tilde{\xi}'^{n}_{i-},\tilde{\xi}'^{n}_{i+})_{i\in\mathcal{I}'_n(\mathcal{D})}$ consisting of independent copies of $\left(\xi'_{p-},\xi'_{p+}\right)$. Set

$$\tilde{F}_{n,i}\equiv\frac{\hat{\upsilon}_{n,i-}\tilde{\xi}'_{n,i-}+\hat{\upsilon}_{n,i+}\tilde{\xi}'_{n,i+}}{\sqrt{2}}.$$

2 Compute

$$\begin{aligned}&\tilde{\zeta}^*_{n,\beta}(\mathcal{D})\\&\equiv\left(\sum_{i\in\mathcal{I}'_n(\mathcal{D})}\frac{\Delta_i^n Z\,\tilde{\varsigma}_{n,i}}{\hat{\phi}_{n,i}}\right)\Big/\left(\sum_{i\in\mathcal{I}'_n(\mathcal{D})}\frac{\Delta_i^n Z^2}{\hat{\phi}_{n,i}}\right)\\&\quad+\frac{\left(\sum_{i\in\mathcal{I}'_n(\mathcal{D})}\frac{\Delta_i^n Z^2\tilde{F}_{n,i}}{\hat{\phi}^2_{n,i}}\right)\left(\sum_{i\in\mathcal{I}'_n(\mathcal{D})}\frac{\Delta_i^n Z\,\tilde{\varsigma}_{n,i}}{\hat{\phi}_{n,i}}\right)-\left(\sum_{i\in\mathcal{I}'_n(\mathcal{D})}\frac{\Delta_i^n Z\,\tilde{\varsigma}_{n,i}\,\tilde{F}_{n,i}}{\hat{\phi}^2_{n,i}}\right)\left(\sum_{i\in\mathcal{I}'_n(\mathcal{D})}\frac{\Delta_i^n Z^2}{\hat{\phi}_{n,i}}\right)}{k_n^{1/2}\left(\sum_{i\in\mathcal{I}'_n(\mathcal{D})}\frac{\Delta_i^n Z^2}{\hat{\phi}_{n,i}}\right)^2}.\end{aligned}$$

3 Generate a large number of Monte Carlo simulations in the first two steps and set $cv^{\alpha/2}_{n,\beta}$ 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 confidence interval (CI) as

$$\mathrm{CI}^{*\alpha}_n=\left[\hat{\beta}_n(\mathcal{D},w^*)-\Delta_n^{1/2}cv^{\alpha/2}_{n,\beta},\hat{\beta}_n(\mathcal{D},w^*)+\Delta_n^{1/2}cv^{\alpha/2}_{n,\beta}\right].$$

Algorithm 3: Refined confidence intervals for the jump beta.

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 4, 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 4 helps since if k_n is relatively small, the higher-order term $k_n^{-1/2}H^*_{n,\beta}(\mathcal{D})$ might have nontrivial finite sample effect.

For concreteness, we describe in Algorithm 3 a finite-sample correction for the CIs described in Algorithm 2, based on Theorem 4, where we set for $i\in\mathcal{I}'_n(\mathcal{D})$,

$$\hat{\phi}_{n,i}\equiv\frac{\left(-\tilde{\beta}_n,1\right)(\hat{c}_{n,i-}+\hat{c}_{n,i+})\left(-\tilde{\beta}_n,1\right)}{2},\quad\hat{\upsilon}_{n,i\pm}\equiv\left(-\tilde{\beta}_n,1\right)\hat{c}_{n,i\pm}\left(-\tilde{\beta}_n,1\right)^{\top}.\tag{4.29}$$

The proof of Theorem 2(c) can be easily adapted to show that $\mathrm{CI}^{*\alpha}_n$ defined in Algorithm 3 has asymptotic level $1-\alpha$, that is, $\mathbb{P}(\beta_0\in CI^{\alpha}_n)\to1-\alpha$; the details are omitted for brevity.

5 Numerical experiments

We now assess the efficiency gain provided by our efficient estimation procedure and we further examine the finite-sample performance of the asymptotic theory above in realistically calibrated simulations.

5.1 Relative efficiency of beta estimation

We start with gauging the efficiency gains of our efficient estimation procedure in empirical relevant scenarios. As seen from the asymptotic theory in Section 4, the sampling variability in the estimation of beta depends on the volatility processes of Y and Z, as well as the number and sizes of jumps. Therefore, to make the efficiency comparisons practically relevant, we calibrate the numerical environment using estimates of these quantities from our empirical application in Section 6. In particular, in the calculation of the asymptotic variances we will use the detected jumps in our empirical data sets and we will further set $c_t = \frac{1}{2}\left(\hat{c}_{n,i-} + \hat{c}_{n,i+}\right)$ for $t \in ((i-1)\Delta_n, i\Delta_n]$.

We conduct two efficiency comparisons. First, in order to have a general sense about how accurate the jump beta can be estimated, we compare the efficiency bound for estimating the jump beta, which is attained by our optimally weighted estimator, with that for estimating the continuous beta. Under the assumption that β_t^c (recall (4.15)) is a constant, Li, Todorov, and Tauchen (2014) show that the sharp lower efficiency bound for estimating the continuous beta is $(\int_0^T c_{ZZ,s}/v_s^c ds)^{-1}$. For the 13 assets studied in our empirical application, we find that estimating the continuous beta is 6 to 7 times more accurate, measured by the $\mathcal{F}$-conditional asymptotic standard deviation, than estimating the jump beta from the same data set. We note that this is in spite of the fact that the jump beta estimation is (effectively) based on 74 jump returns detected in the sample while the continuous beta is based on the remaining of the total $56,886$ high-frequency increments. The intuition is that, although the number of jump returns is small, these returns have much higher signal-to-noise ratio than their diffusive counterparts for the estimation of betas.[16]

Our second efficiency comparison concerns the role of the optimal weighting in the efficient estimation of jump beta. That is, we are interested in the efficiency gains from using the optimal weight function $w^*(\cdot)$ over the case of no weighting, corresponding to $w(\cdot) = 1$, which has been done in prior work such as Gobbi and Mancini (2012) and Todorov and Bollerslev (2010). The comparison is, again, implemented using estimates of jumps and volatility paths as explained above.

[16] Of course the above efficiency comparison is based on the premise that the continuous beta and the jump beta remain constant over the same time interval. Results from tests for temporal stability of continuous betas in Reiss, Todorov, and Tauchen (2014) indicate that continuous betas vary even over short periods such as months. This is unlike our empirical findings for the jump betas, reported in the next section, for which we find temporal stability over time spans of years.

Table 1: Relative Efficiency of Jump Beta Estimation

Asset	$\frac{a.s.e.\hat{\beta}_n}{a.s.e.\tilde{\beta}_n}$	**Asset**	$\frac{a.s.e.\hat{\beta}_n}{a.s.e.\tilde{\beta}_n}$	**Asset**	$\frac{a.s.e.\hat{\beta}_n}{a.s.e.\tilde{\beta}_n}$	**Asset**	$\frac{a.s.e.\hat{\beta}_n}{a.s.e.\tilde{\beta}_n}$
XLB	0.68	XLK	0.61	XLV	0.61	XOM	0.69
XLE	0.61	XLP	0.66	XLY	0.61	BAC	0.32
XLF	0.44	XLU	0.55	IBM	0.51	GLD	0.55
XLI	0.69						

Note: Calculations of the asymptotic standard error (a.s.e.) are based on detected jumps and volatility paths extracted from the empirical data set discussed in Section 6. The efficient estimator $\hat{\beta}_n$ and the unweighted estimator $\tilde{\beta}_n$ correspond to $\hat{\beta}_n([0,T]\times\mathbb{R}_*, w)$ with $w(\cdot) = w^*(\cdot)$ and $w(\cdot) = 1$, respectively.

Table 1 reports the relative efficiency of the unweighted estimator versus the efficient estimator. We see that the optimal weighting indeed provides nontrivial efficiency gains, with the minimal gain being 32% among all assets in our sample. Not surprisingly, these gains vary across assets and are bigger for those with more volatility variations over the time period. We note that optimal weighting is of particular relevance for the jump beta estimation: by their very nature, jumps are rare events and, hence, for estimating the jump beta we naturally pool information from distinct time periods which typically have very different volatility levels.

5.2 Monte Carlo

We proceed with assessing the performance of our inference techniques on simulated data from the following model

$$dZ_t = \sigma_t dL_t, \quad dY_t = \beta_t dZ_t + \sigma_t d\widetilde{L}_t, \quad d\sigma_t^2 = 0.03(1-\sigma_t^2)dt + 0.15\sigma_t dB_t, \tag{5.1}$$

where L and $\widetilde{L}$ are two independent Lévy processes with characteristic triplets $\left(0, 1, \frac{e^{-|x|}}{24}\right)$ and $\left(0, \frac{1}{\sqrt{2}}, \frac{e^{-|x|}}{96}\right)$ with respect to the zero truncation function; B is a Brownian motion independent of L and $\widetilde{L}$. This means that L and $\widetilde{L}$ are Brownian motions plus compound Poisson jumps, with jumps having double-exponential distribution. The frequency and jump size distributions are calibrated to mimic those in the real data that we are going to use. For the beta process we consider

$$\begin{cases} \beta_t = 1, \quad \text{for } t \in [0,T], & \text{under } H_0 \text{ (null hypothesis)}, \\ d\beta_t = 0.005(1-\beta_t)dt + 0.005\sqrt{\beta_t}d\widetilde{B}_t, & \text{under } H_a \text{ (alternative hypothesis)}, \end{cases} \tag{5.2}$$

Table 2: Monte Carlo Rejection Rates (%) of Tests for Constant Jump Beta

Case	Under H_0			Under H_a		
	Nominal Level			Nominal Level		
	10%	5%	1%	10%	5%	1%
$1/\Delta_n = 38$, $k_n = 19$	11.79	6.65	2.65	58.00	46.00	29.00
$1/\Delta_n = 38$, $k_n = 25$	11.09	6.32	2.52	57.00	47.00	29.00
$1/\Delta_n = 81$, $k_n = 27$	11.60	6.33	1.90	83.00	76.00	65.00
$1/\Delta_n = 81$, $k_n = 35$	10.94	6.14	2.11	83.00	75.00	64.00

where $\widetilde{B}$ is a Brownian motion independent of B, L and $\widetilde{L}$. The unconditional mean of β_t under the alternative hypothesis is 1 and the expected range of the process $(\beta_t)_{t\geq 0}$ over the interval of estimation is approximately 0.2.

We set $T = 1,500$ days (our unit of time is a trading day), and consider two sampling frequencies: $\Delta_n = 1/38$ which corresponds to sampling every 10 minutes in a 6.5 hours trading day, and $\Delta_n = 1/81$ which corresponds to sampling every 5 minutes. We experiment with two values of k_n for each of the sampling frequency in order to check the sensitivity of the inference techniques with respect to this tuning parameter. Finally, as is typical in truncation-based methods, we select the truncation threshold in the following data-driven way. For the increment $\Delta_i^n Z$ with $i = \lfloor (t-1)/\Delta_n \rfloor + 1, ..., \lfloor t/\Delta_n \rfloor$, we set

$$v_n = 4 \times \sqrt{BV_t} \times \Delta_n^{0.49}, \quad BV_t = \frac{\lfloor t/\Delta_n \rfloor}{\lfloor t/\Delta_n \rfloor - 1} \frac{\pi}{2} \sum_{\lfloor (t-1)/\Delta_n \rfloor + 2}^{\lfloor t/\Delta_n \rfloor} |\Delta_{i-1}^n Z||\Delta_i^n Z|. \tag{5.3}$$

Here, BV_t is the Bipower Variation of Barndorff-Nielsen and Shephard (2004b, 2006) which is a jump-robust estimator of volatility and importantly free of tuning parameters. For the construction of $\hat{c}_{n,i\pm}$ we include all increments for which both components are below a threshold set similarly as above but with 4 replaced by 3. There are 10,000 Monte Carlo trials.

In Table 2 we report the results from the Monte Carlo for the test of constant jump beta. As seen from the table, the test has good size properties with only mild overrejections. These overrejections decrease when the sampling frequency increases from 10 to 5 minutes. The test also has a reasonable power against the considered alternative which increases with the sampling frequency. In Table 3 we report the coverage probability for the refined CI of jump beta that is

Table 3: Monte Carlo Coverage Probability (%) of Confidence Intervals

Case	Nominal Level		
	90%	95%	99%
$1/\Delta_n = 38$, $k_n = 19$	88.00	93.43	98.21
$1/\Delta_n = 38$, $k_n = 25$	88.78	93.95	98.51
$1/\Delta_n = 81$, $k_n = 27$	88.60	94.18	98.52
$1/\Delta_n = 81$, $k_n = 35$	89.15	94.29	98.45

based on the efficient estimator and is described in Algorithm 3. The coverage probabilities are in general quite close to the nominal levels of the CIs. Not surprisingly, we see again improved performance at the higher sampling frequency. We also note that the coverage probability of the CIs is not very sensitive to the choice of the block size k_n. Overall, we find quite satisfactory finite-sample performance of our inference techniques for the jump betas, even for relatively sparse sampling of $1/\Delta_n = 38$.

6 Empirical application

The application concerns betas on market jumps for assets in three classes, with the market proxy (as discussed in the Introduction) being the ETF that tracks the S&P 500 index (ticker symbol: SPY). The first set of assets consists of the ETFs on the nine industry portfolios comprising the S&P 500 index. The industry portfolios are, with ticker symbols in parenthesis, as follows: materials (XLB), energy (XLE), financials (XLF), industry (XLI), technology (XLK), consumer staples (XLP), utilities (XLU), healthcare (XLV), and consumer discretionary (XLY). The second set consists of large-cap stocks: IBM (IBM), Exxon Mobil (XOM), and Bank of America (BAC). Finally, we also consider the precious metal asset gold ETF (GLD).

Data on each series are sampled at the 10-minute frequency over the period 2007–2012. The resultant data sets consists of $1,746$ 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, such as within-asset trading frictions (e.g., bid-ask bounce, rounding error, etc.) and asynchronous trading across assets.[17] We set the truncation threshold exactly as (5.3)

[17] There is now a vast literature on noise-robust estimation for integrated variance and covariance. Developing a noise-robust theory for the jump regressions is left for future research.

in the Monte Carlo, with further correction for the well-known deterministic diurnal pattern in volatility.[18] The block size is set to $k_n = 19$.

Figures 2 and 3 display the scatter plots of the detected jump increments of the various assets against those of the market index. These scatters are estimates, in the sense of Proposition 1, of their population counterparts. The figures also show the fit provided by the linear jump regression model (1.2) based on the optimally weighted estimator developed in Section 4.1. Perhaps surprisingly, the fit appears generally quite tight for equities and equity portfolios, 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. In contrast to equity-based assets, we observe no strong relationship between the jumps in the gold ETF and the market index, consistent with the fact that gold price has nontrivial exposure to other risk factors.

Table 4 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 time of the price jumps. Nevertheless, for most assets the volatility jump, at the time of the price jump, is quite small on average, suggesting our beta estimator is close to mixed Gaussian. The second to last column of Table 4, which reports the R^2s, confirms our observation, based on Figures 2 and 3, of the good fit provided by the linear jump regression, with the exception of gold. In general the R^2s are quite high - the worst fit in terms of R^2 is for gold, the financial ETF and the stock from the financial sector, BAC. Despite the apparently good fit, the formal test for constancy of the jump beta rejects the null for all but three of the assets in our sample at all conventional levels of the test; see the last column of Table 4 for p-values of these tests. The deviations from linearity observed in Figures 2 and 3 are thus in most cases strongly statistically significant.

Of course, as suggested by Figure 1 discussed in the Introduction, the linear 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 5. Allowing for beta to change over years improves the performance of the linear jump regression model for the industry portfolios. Indeed, the constant jump beta hypothesis is not rejected at the conventional

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

Figure 2: Scatter of Jumps: Industry Portfolios

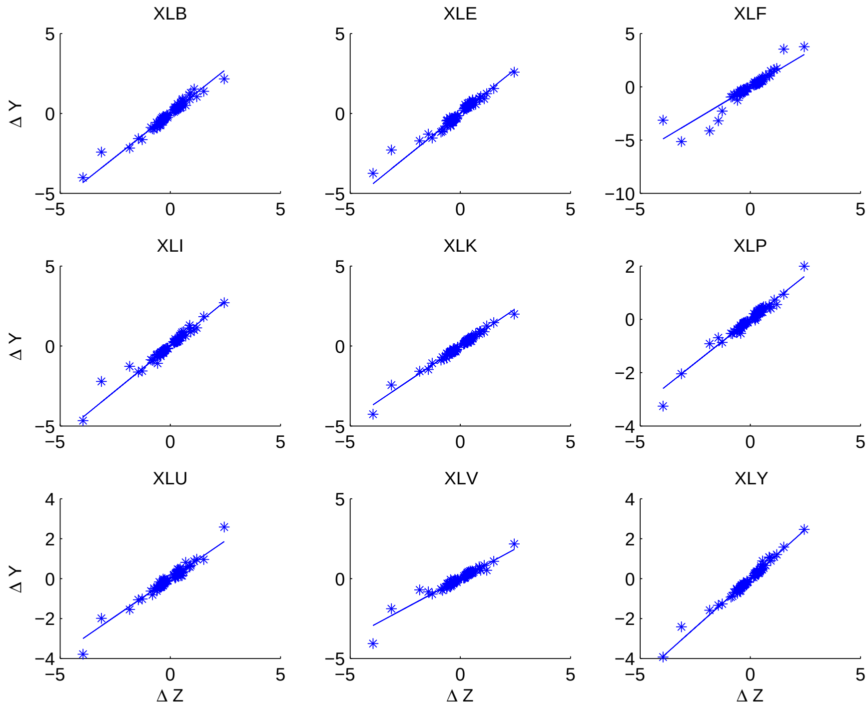


Figure 3: Scatter of Jumps: Stocks and the Gold ETF

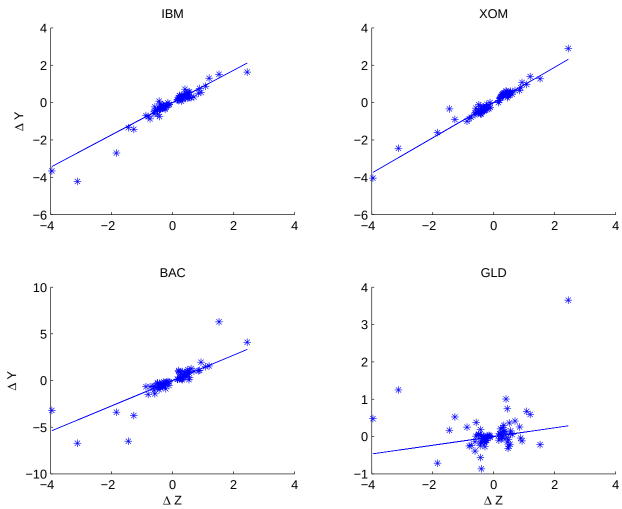


Table 4: Jump Betas and Tests for Constancy over the Full Sample

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
IBM	0.8647	[0.8172 0.9123]	0.6658	0.7187	0.9205	0.0000
XOM	0.9446	[0.8982 0.9910]	0.7005	0.7242	0.9437	0.0002
BAC	1.3582	[1.2617 1.4547]	1.6111	1.7155	0.7493	0.0186
GLD	0.1167	[0.0672 0.1661]	0.6020	0.6722	0.0434	0.0000

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.

1% significance level in the majority of cases. The same holds true, albeit to a far less extent, for the stocks.

The rejections of constant jump beta within a year seen in Table 4 could be due to temporal instability within a year, nonlinearity in the jump regression (recall jump betas can depend on the sign and size of jumps), or both. We thus further expand the above analysis by considering a piecewise linear specification where, for each asset and year, separate jump betas are estimated for the negative and positive market jumps. This is analogous to Ang, Chen, and Xing (2006) and Lettau, Maggiori, and Weber (2014), who in discrete-time settings investigate the pricing implications of separating upside and downside betas. Table 6 shows the summary results for the three stocks.[19] These results can reconcile and help understand the reasons for the rejections

[19] Since Table 5 shows that the constant jump beta model within each year works already quite well for the industry portfolios, we do not report analogous results for them. On the other hand, for gold, even after allowing for the jump beta to depend on the market jump size, we still detect rejection of such a piecewise linear jump regression in 2/3 of

Table 5: 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
IBM	0.001	0.030	0.099	0.009	0.006	0.013
XOM	0.000	0.007	0.007	0.831	0.303	0.001
BAC	0.299	0.166	0.000	0.278	0.343	0.087
GLD	0.047	0.001	0.000	0.000	0.000	0.000
	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.

by year and stock seen in Table 5. Starting with IBM, from Table 5, a linear jump relation is rejected at the 1% level for years 2007, 2010, and 2011. For 2007, Table 6 indicates the rejection is mostly due to instability of the linear relationship for negative market jumps, but also there is some difference in the estimated positive and negative jump betas as well. For 2010 and 2011, the jump betas appear quite stable within each region but differ across regions, suggesting the rejections for the entire year are due to nonlinearity in the jump regression. For XOM, from Table 5, a linear jump relation is rejected at 1% for years 2007, 2008, 2009, and 2012. From Table 6 it is seen that different positive and negative jump betas can account for all these years. In the first three years

the sign/year combinations. We do not report these results in order to save space.

the negative jump betas are lower than the positive ones, and for the rest of the sample we see the reverse. Further in 2009, there is in addition some instability in the negative jump beta.

BAC is an interesting case. We first note that inference for this stock is rather difficult because of the very high levels of idiosyncratic (with respect to the market) risk as evident from the relatively wide confidence intervals for the jump betas reported in Table 6. From Table 5, we observe that a test for a linear jump relation does not reject the null for all years except 2009. However, when we subdivide and perform the test for linear jump relation on the set of positive and negative jumps separately, we see that during some of the years there is some evidence for instability of the linear relationship. For example, in 2007, 2008 and 2010 there is some instability in the negative jump beta, and in 2011 in the positive jump beta. When pooling the set of positive and negative jumps, these instabilities become harder to detect. Finally, in year 2009 we note a very significant difference between positive and negative jump beta, with the latter being much higher than the former. This nonlinearity can explain the failure of the linear jump regression model for that year.

The preceding analysis illustrates that a linear jump regression model, potentially with separate betas for positive and negative market jumps, works well over periods of years in capturing the dependence between jumps in industry portfolios and equity-based assets on one hand and the market jumps on the other hand. The analysis here can be extended to a broader set of assets. It can be further expanded to include a larger set of systematic risk factors (in addition to the market portfolio). 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 6: Positive and Negative Jump Betas and Tests for Constancy over Years

Asset/Year	$\Delta Z < 0$			$\Delta Z > 0$		
IBM	$\widehat{\beta}$	95% CI	p-val	$\widehat{\beta}$	95% CI	p-val
2007	1.009	[0.848 1.170]	0.003	0.747	[0.595 0.899]	0.029
2008	1.309	[1.099 1.518]	0.025	0.722	[0.542 0.902]	0.029
2009	0.884	[0.728 1.040]	0.008	0.766	[0.669 0.863]	0.025
2010	0.849	[0.732 0.966]	0.728	0.892	[0.745 1.038]	0.526
2011	0.897	[0.660 1.133]	0.226	0.770	[0.601 0.939]	0.281
2012	0.837	[0.693 0.981]	0.660	0.847	[0.713 0.981]	0.000
XOM						
2007	0.881	[0.748 1.014]	0.193	1.033	[0.830 1.236]	0.478
2008	0.803	[0.627 0.978]	0.625	1.173	[0.908 1.438]	0.941
2009	0.578	[0.432 0.723]	0.000	0.943	[0.839 1.047]	0.012
2010	1.040	[0.915 1.165]	0.219	0.894	[0.692 1.097]	0.255
2011	1.028	[0.792 1.265]	0.569	0.788	[0.629 0.947]	0.489
2012	1.181	[1.030 1.332]	0.411	1.082	[0.952 1.212]	0.041
BAC						
2007	1.410	[1.218 1.602]	0.002	1.169	[1.009 1.330]	0.014
2008	1.649	[1.299 2.000]	0.004	1.894	[1.530 2.258]	0.151
2009	2.942	[2.273 3.611]	0.030	1.190	[0.912 1.467]	0.191
2010	0.974	[0.751 1.198]	0.005	1.194	[0.812 1.576]	0.624
2011	0.965	[0.588 1.341]	0.595	1.325	[0.983 1.666]	0.004
2012	1.468	[1.179 1.757]	0.674	2.056	[1.754 2.359]	0.012

Note: The left half of the table reports, for each stock and year, the estimated jump beta corresponding to the negative market jumps, the 95% CI, and the p-value for the null hypothesis of a constant linear jump regression model over that given year and jump size domain; the right side of the table reports the corresponding quantities when restricting to the set of positive market jumps.

7 Conclusion

We develop econometric tools for studying the dependence between jumps of two processes from high-frequency observations on a fixed time interval. We derive tests for deciding whether a linear relationship for jumps of two processes, implied by standard linear factor models, holds on a given time interval and for a given region of the jump size domain. We show that the test has power for detecting both nonlinearity in functional form and time-varying parameters. We further propose an efficient estimator for the jump beta and construct feasible confidence sets, as well as a finite-sample refinement of the latter based on a novel higher-order asymptotic expansion. The proposed efficient estimator is a weighted linear estimator, where the weights are constructed from block-based estimates of local stochastic volatility before and after the jump times, and fully account for an unrestricted form of heteroskedasticity. We derive a semiparametric efficiency bound for the estimation in the linear jump regression model and we show that our estimator achieves this efficiency bound. The asymptotic mixed Gaussianity of our estimator arises from the local approximate Gaussianity of the diffusive price increments, rather than from averaging out a large number of weakly dependent random disturbances. Therefore, despite the fact that only a small number of jump observations are informative for the jump beta, our estimator enjoys the parametric rate of convergence, with its asymptotic variance comparable with that for the estimation of the linear regression coefficient of diffusions observed at high-frequency.

In an empirical application, we document that market jump betas of financial assets remain stable over a period of one year, but find evidence for temporal variation for a longer time interval of six years. In some of the cases, the temporal stability is achievable only after allowing for different jump beta depending on the direction of market jumps. This evidence stands in contrast to that for diffusive (continuous) market betas which are time-varying over quite shorter time periods.

8 Appendix: Proofs

Throughout this appendix, we use K to denote a generic constant that may change from line to line; we sometimes emphasize the dependence of this constant on some parameter q by writing K_q. We use $0_{k\times q}$ to denote a $k\times q$ matrix of zeros and when $q=1$, we write 0_k for notational simplicity; 0_k is understood to be empty when $k=0$. For any sequence of variables $(\xi_{n,p})_{p\geq 1}$, the convergence $(\xi_{n,p})_{p\geq 1} \to (\xi_p)_{p\geq 1}$ is understood as $n\to\infty$ under the product topology. We write w.p.a.1 for "with probability approaching 1."

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

Assumption S1. We have Assumption 1. Moreover, the processes X_t, b_t and σ_t are bounded.

Proof of Proposition 1. (a) Since the jumps of Z have finite activity, we can assume without loss of generality that each interval $((i-1)\Delta_n, i\Delta_n]$ contains at most one jump; otherwise we can restrict our calculation to the w.p.a.1 set of sample paths on which this condition holds. We denote the continuous part of Z by Z^c, that is,

$$Z_t^c = Z_t - \sum_{s\leq t} \Delta Z_s, \quad t \geq 0. \tag{8.1}$$

Note that $\mathcal{I}_n(\mathcal{D})$ is the union of two disjoint sets $\mathcal{I}_{1n}(\mathcal{D})$ and $\mathcal{I}_{2n}(\mathcal{D})$ that are defined as

$$\mathcal{I}_{1n}(\mathcal{D}) = \mathcal{I}_n(\mathcal{D}) \cap \{i(p) : p \in \mathcal{P}\}, \quad \mathcal{I}_{2n}(\mathcal{D}) = \mathcal{I}_n(\mathcal{D}) \backslash \mathcal{I}_{1n}(\mathcal{D}). \tag{8.2}$$

It suffices to show that, w.p.a.1,

$$\mathcal{I}_{1n}(\mathcal{D}) = \mathcal{I}(\mathcal{D}), \quad \mathcal{I}_{2n}(\mathcal{D}) = \emptyset. \tag{8.3}$$

First consider $\mathcal{I}_{1n}(\mathcal{D})$. Since $v_n \to 0$, we have $|\Delta_{i(p)}^n Z| > v_n$ for all $p \in \mathcal{P}$, when n is large enough. Therefore,

$$\mathcal{I}_{1n}(\mathcal{D}) = \Big\{ i(p) : p \in \mathcal{P}, ((i(p)-1)\Delta_n, \Delta_{i(p)}^n Z) \in \mathcal{D} \Big\} \text{ w.p.a.1.} \tag{8.4}$$

Now, observe that

$$\sup_{p\in\mathcal{P}} \Big\| \Big((i(p)-1)\Delta_n, \Delta_{i(p)}^n Z \Big) - \big(\tau_p, \Delta Z_{\tau_p} \big) \Big\| \to 0 \quad \text{a.s.} \tag{8.5}$$

Indeed, almost surely,

$$\begin{aligned} \sup_{p\in\mathcal{P}} \Big\| \Big((i(p)-1)\Delta_n, \Delta_{i(p)}^n Z \Big) - \big(\tau_p, \Delta Z_{\tau_p} \big) \Big\| &= \sup_{p\in\mathcal{P}} \Big\| \Big((i(p)-1)\Delta_n - \tau_p, \Delta_{i(p)}^n Z^c \Big) \Big\| \\ &\leq \Delta_n + \sup_{s,t\leq T, |s-t|\leq \Delta_n} |Z_t^c - Z_s^c| \to 0. \end{aligned} \tag{8.6}$$

By Assumption 2, the marks $(\tau_p, \Delta Z_{\tau_p})_{p\in\mathcal{P}_\mathcal{D}}$ are contained in the interior of $\mathcal{D}$ a.s. Then, by (8.5), $((i(p)-1)\Delta_n, \Delta_{i(p)}^n Z)_{p\in\mathcal{P}_\mathcal{D}} \subseteq \mathcal{D}$ w.p.a.1. With the same argument but with $\mathcal{D}^c$ (i.e. the complement of $\mathcal{D}$) replacing $\mathcal{D}$, we deduce $((i(p)-1)\Delta_n, \Delta_{i(p)}^n Z)_{p\in\mathcal{P}\backslash\mathcal{P}_\mathcal{D}} \subseteq \mathcal{D}^c$ w.p.a.1. Therefore, the set on the right-hand side of (8.4) coincides with $\mathcal{I}(\mathcal{D})$ w.p.a.1. From here, the first claim of (8.3) readily follows.

It remains to show that $\mathcal{I}_{2n}(\mathcal{D})$ is empty w.p.a.1. Note that for $i \in \mathcal{I}_{2n}(\mathcal{D})$, $\Delta_i^n Z = \Delta_i^n Z^c$. Hence, for any $q > 2/(1-2\varpi)$,

$$\mathbb{P}(\mathcal{I}_{2n}(\mathcal{D}) \neq \emptyset) \leq \sum_{i=1}^{\lfloor T/\Delta_n \rfloor} \mathbb{P}(|\Delta_i^n Z^c| > v_n) \leq K_q \Delta_n^{-1} \frac{\Delta_n^{q/2}}{v_n^q} \to 0, \tag{8.7}$$

where the second inequality is by Markov's inequality and $\mathbb{E}|\Delta_i^n Z^c|^q \leq K_q \Delta_n^{q/2}$; the convergence is due to (2.12) and our choice of q. The proof of part (a) is now complete.

(b) By part (a), it suffices to show that

$$((i-1)\Delta_n, \Delta_i^n X)_{i \in \mathcal{I}(\mathcal{D})} - \left(\tau_p, \Delta X_{\tau_p}\right)_{p \in \mathcal{P}_\mathcal{D}} = o_p(1). \tag{8.8}$$

Observe that $((i-1)\Delta_n, \Delta_i^n X)_{i \in \mathcal{I}(\mathcal{D})}$ is simply $((i(p)-1)\Delta_n, \Delta_{i(p)}^n X)_{p \in \mathcal{P}_\mathcal{D}}$. We deduce the desired convergence via the same argument as that for (8.5). *Q.E.D.*

Proof of Theorem 1. (a) Let

$$\bar{\beta}(\mathcal{D}) \equiv \frac{Q_{ZY}(\mathcal{D})}{Q_{ZZ}(\mathcal{D})}. \tag{8.9}$$

For each $p \geq 1$, we set

$$R_{n,p} = \Delta_n^{-1/2}(\Delta_{i(p)}^n X - \Delta X_{\tau_p}) \quad \text{and} \quad \varsigma_{n,p} = (-\bar{\beta}(\mathcal{D}), 1) R_{n,p}. \tag{8.10}$$

With these notations, we have in restriction to $\Omega_0(\mathcal{D})$,

$$\Delta_{i(p)}^n Y = \beta_0 \Delta_{i(p)}^n Z + \Delta_n^{1/2} \varsigma_{n,p}. \tag{8.11}$$

By Proposition 4.4.10 in Jacod and Protter (2012), $(R_{n,p})_{p\geq 1} \xrightarrow{\mathcal{L}\text{-}s} (R_p)_{p \geq 1}$, where R_p is defined in (3.2). Consequently (recall the notation (3.10)),

$$(\varsigma_{n,p})_{p \geq 1} \xrightarrow{\mathcal{L}\text{-}s} (\varsigma_p)_{p \geq 1}. \tag{8.12}$$

By Proposition 1(a), w.p.a.1.,

$$\det[Q_n(\mathcal{D})] = \left(\sum_{p \in \mathcal{P}_\mathcal{D}} \Delta_{i(p)}^n Z^2\right)\left(\sum_{p \in \mathcal{P}_\mathcal{D}} \Delta_{i(p)}^n Y^2\right) - \left(\sum_{p \in \mathcal{P}_\mathcal{D}} \Delta_{i(p)}^n Z \Delta_{i(p)}^n Y\right)^2. \tag{8.13}$$

Plug (8.11) into (8.13). After some algebra, we deduce

$$\Delta_n^{-1} \det[Q_n(\mathcal{D})] = \left(\sum_{p \in \mathcal{P}_\mathcal{D}} \Delta_{i(p)}^n Z^2\right)\left(\sum_{p \in \mathcal{P}_\mathcal{D}} \varsigma_{n,p}^2\right) - \left(\sum_{p \in \mathcal{P}_\mathcal{D}} \Delta_{i(p)}^n Z\, \varsigma_{n,p}\right)^2. \tag{8.14}$$

Note that for each $p \geq 1$, $\Delta_{i(p)}^{n} Z \rightarrow \Delta Z_{\tau_p}$. Combining this convergence with (8.12), we use the property of stable convergence to derive the joint convergence

$$\left(\varsigma_{n,p}, \Delta_{i(p)}^{n} Z\right)_{p\geq 1} \xrightarrow{\mathcal{L}\text{-}s} \left(\varsigma_p, \Delta Z_{\tau_p}\right)_{p\geq 1}. \tag{8.15}$$

Since the set $\mathcal{P}_{\mathcal{D}}$ is a.s. finite, the assertion of part (a) follows from (8.14), (8.15) and the continuous mapping theorem.

(b) By a standard localization argument (see Section 4.4.1 of Jacod and Protter (2012)), we assume that Assumption S1 holds without loss of generality. Since $\mathcal{P}_{\mathcal{D}}$ is a.s. finite, we can also assume that $|\mathcal{P}_{\mathcal{D}}| \leq M$ for some constant $M > 0$ for the purpose of proving convergence in probability; otherwise, we can fix some large M to make $\mathbb{P}\left(|\mathcal{P}_{\mathcal{D}}| > M\right)$ arbitrarily small and restrict the calculation below on the set $\{|\mathcal{P}_{\mathcal{D}}| \leq M\}$.

By Theorem 9.3.2 in Jacod and Protter (2012), we have,

$$\hat{c}_{n,i(p)-} \xrightarrow{\mathbb{P}} c_{\tau_p -}, \quad \hat{c}_{n,i(p)+} \xrightarrow{\mathbb{P}} c_{\tau_p}, \quad \text{all} \quad 1 \leq p \leq M. \tag{8.16}$$

By Proposition 1(b),

$$Q_n\left(\mathcal{D}\right) \xrightarrow{\mathbb{P}} Q\left(\mathcal{D}\right), \tag{8.17}$$

which further implies (with $\tilde{\beta}_n \equiv Q_{ZY,n}\left(\mathcal{D}\right)/Q_{ZZ,n}\left(\mathcal{D}\right)$)

$$\tilde{\beta}_n \xrightarrow{\mathbb{P}} \bar{\beta}\left(\mathcal{D}\right). \tag{8.18}$$

Furthermore, by essentially the same argument as in the proof of Proposition 1(a), we deduce

$$\mathcal{I}_n'\left(\mathcal{D}\right) = \mathcal{I}\left(\mathcal{D}\right) \quad \text{w.p.a.1.} \tag{8.19}$$

Therefore,

$$\tilde{\zeta}_n\left(\mathcal{D}\right) = \left(\sum_{p\in\mathcal{P}_{\mathcal{D}}} \Delta_{i(p)}^{n} Z^2\right)\left(\sum_{p\in\mathcal{P}_{\mathcal{D}}} \tilde{\varsigma}_{n,i(p)}^2\right) - \left(\sum_{p\in\mathcal{P}_{\mathcal{D}}} \Delta_{i(p)}^{n} Z\, \varsigma_{n,i(p)}\right)^2 \quad \text{w.p.a.1.} \tag{8.20}$$

Fix any subsequence $\mathbb{N}_1 \subseteq \mathbb{N}$. By (8.16) and (8.18), we can extract a further subsequence $\mathbb{N}_2 \subseteq \mathbb{N}_1$, such that along $\mathbb{N}_2$,

$$\left(\left(\hat{c}_{n,i(p)-}, \hat{c}_{n,i(p)+}\right)_{1\leq p\leq M}, \tilde{\beta}_n\right) \rightarrow \left(\left(c_{\tau_p -}, c_{\tau_p}\right)_{1\leq p\leq M}, \bar{\beta}\left(\mathcal{D}\right)\right) \tag{8.21}$$

on some set $\tilde{\Omega}$ with $\mathbb{P}(\tilde{\Omega}) = 1$. Then, for each $\omega \in \tilde{\Omega}$ fixed, the transition kernel of $\tilde{\zeta}_n\left(\mathcal{D}\right)$ given $\mathcal{F}$ converges weakly to the $\mathcal{F}$-conditional law of $\zeta\left(\mathcal{D}\right)$. Moreover, observe that the $\mathcal{F}$-conditional law of the variables $(\varsigma_p)_{1\leq p\leq M}$ does not have atoms and has full support on $\mathbb{R}^M$. Therefore, the

$\mathcal{F}$-conditional distribution function of $\zeta(\mathcal{D})$ is continuous and strictly increasing. By Lemma 21.2 in van der Vaart (1998), we deduce that on each path $\omega \in \tilde{\Omega}$, along the subsequence $\mathbb{N}_2$, $cv_n^\alpha \to cv^\alpha$, where cv^α is the $\mathcal{F}$-conditional $(1-\alpha)$-quantile of $\zeta(\mathcal{D})$. Since the subsequence $\mathbb{N}_1$ is arbitrarily chosen, we further deduce that $cv_n^\alpha \xrightarrow{\mathbb{P}} cv^\alpha$ by the subsequence characterization of convergence in probability. The proof for part (b) is now complete.

(c) By part (a) and part (b), as well as the property of stable convergence, we have

$$(\Delta_n^{-1}\det[Q_n(\mathcal{D})], cv_n^\alpha, 1_{\Omega_0(\mathcal{D})}) \xrightarrow{\mathcal{L}\text{-}s} (\zeta(\mathcal{D}), cv^\alpha, 1_{\Omega_0(\mathcal{D})}). \tag{8.22}$$

In particular,

$$\mathbb{P}\left(\{\Delta_n^{-1}\det[Q_n(\mathcal{D})] > cv_n^\alpha\} \cap \Omega_0(\mathcal{D})\right) \to \mathbb{P}\left(\{\zeta(\mathcal{D}) > cv^\alpha\} \cap \Omega_0(\mathcal{D})\right). \tag{8.23}$$

Since $\mathbb{P}(\zeta(\mathcal{D}) > cv^\alpha|\mathcal{F}) = \alpha$ and $\Omega_0(\mathcal{D}) \in \mathcal{F}$, the right-hand side of (8.23) equals to $\alpha\mathbb{P}(\Omega_0(\mathcal{D}))$. The first assertion of part (c) then follows from (8.23). To show the second assertion of part (c), we first observe that (8.17) implies $\det[Q_n(\mathcal{D})] \xrightarrow{\mathbb{P}} \det[Q(\mathcal{D})]$. In restriction to $\Omega_a(\mathcal{D})$, $\det[Q(\mathcal{D})] > 0$ and, hence, $\Delta_n^{-1}\det[Q_n(\mathcal{D})]$ diverges to $+\infty$ in probability. Part (b) implies that cv_n^α is tight in restriction to $\Omega_a(\mathcal{D})$. Consequently, $\mathbb{P}\left(\Delta_n^{-1}\det[Q_n(\mathcal{D})] > cv_n^\alpha|\Omega_a(\mathcal{D})\right) \to 1$ as asserted. *Q.E.D.*

Proof of Theorem 2. (a) Observe that

$$Q_{ZY,n}(\mathcal{D},w) - \beta_0 Q_{ZZ,n}(\mathcal{D},w) = \sum_{i\in\mathcal{I}'_n(\mathcal{D})} w\left(\hat{c}_{i-}^n, \hat{c}_{i+}^n, \tilde{\beta}_n\right)\Delta_i^n Z\left(\Delta_i^n Y - \beta_0\Delta_i^n Z\right). \tag{8.24}$$

Recall the notation $\varsigma_{n,p}$ from (8.10). By (8.19), we further deduce that, w.p.a.1,

$$\Delta_n^{-1/2}\left(Q_{ZY,n}(\mathcal{D},w) - \beta_0 Q_{ZZ,n}(\mathcal{D},w)\right) = \sum_{p\in\mathcal{P}_\mathcal{D}} w\left(\hat{c}_{n,i(p)-}, \hat{c}_{n,i(p)+}, \tilde{\beta}_n\right)\Delta_{i(p)}^n Z\,\varsigma_{n,p}. \tag{8.25}$$

By (8.16), (8.18) and Assumption 3,

$$w\left(\hat{c}_{i(p)-}^n, \hat{c}_{i(p)+}^n, \tilde{\beta}_n\right) \xrightarrow{\mathbb{P}} w\left(c_{\tau_p-}, c_{\tau_p}, \beta_0\right), \quad p \geq 1. \tag{8.26}$$

Since $\mathcal{P}_\mathcal{D}$ is a.s. finite, we use properties of stable convergence to deduce from (8.12) and (8.26) that

$$\Delta_n^{-1/2}\left(Q_{ZY,n}(\mathcal{D},w) - \beta_0 Q_{ZZ,n}(\mathcal{D},w)\right) \xrightarrow{\mathcal{L}\text{-}s} \sum_{p\in\mathcal{P}_\mathcal{D}} w\left(c_{\tau_p-}, c_{\tau_p}, \beta_0\right)\Delta Z_{\tau_p}\varsigma_p. \tag{8.27}$$

Note that

$$\Delta_n^{-1/2}(\hat{\beta}_n(\mathcal{D},w) - \beta_0) = \frac{\Delta_n^{-1/2}\left(Q_{ZY,n}(\mathcal{D},w) - \beta_0 Q_{ZZ,n}(\mathcal{D},w)\right)}{Q_{ZZ,n}(\mathcal{D},w)}. \tag{8.28}$$

By (8.19),

$$Q_n(\mathcal{D},w)=\sum_{p\in\mathcal{P}_\mathcal{D}} w\left(\hat{c}^n_{i(p)-},\hat{c}^n_{i(p)+},\tilde{\beta}_n\right)\Delta^n_{i(p)}X\Delta^n_{i(p)}X^\top. \tag{8.29}$$

By $\Delta^n_{i(p)}X\to\Delta X_{\tau_p}$ and (8.26), we deduce

$$Q_n(\mathcal{D},w)\xrightarrow{\mathbb{P}}\sum_{p\in\mathcal{P}_\mathcal{D}} w\left(c_{\tau_p-},c_{\tau_p},\beta_0\right)\Delta X_{\tau_p}\Delta X_{\tau_p}^\top. \tag{8.30}$$

The first assertion of part (a), that is, $\Delta_n^{-1/2}(\hat{\beta}_n(\mathcal{D},w)-\beta_0)\xrightarrow{\mathcal{L}\text{-}s}\zeta_\beta(\mathcal{D},w)$ readily follows from (8.27), (8.28) and (8.30).

Turning to the second assertion of part (a), we first observe that when c_t does not jump at the same time as Z_t, each ς_p is $\mathcal{F}$-conditionally centered Gaussian; moreover, the variables $(\varsigma_p)_{p\geq1}$ are $\mathcal{F}$-conditionally independent. Therefore, the limiting variable $\zeta_\beta(\mathcal{D})$ is centered Gaussian conditional on $\mathcal{F}$, with conditional variance given by $\Sigma(\mathcal{D},w)$. This finishes the proof of the second assertion.

(b) For notational simplicity, we denote

$$A_p=\frac{(-\beta_0,1)\left(c_{\tau_p-}+c_{\tau_p}\right)(-\beta_0,1)^\top}{2\Delta Z^2_{\tau_p}},\quad B_p=w\left(c_{\tau_p-},c_{\tau_p},\beta_0\right)\Delta Z^2_{\tau_p}.$$

Then we can rewrite $\Sigma(\mathcal{D},w)$ and $\Sigma(\mathcal{D},w^*)$ as

$$\Sigma(\mathcal{D},w)=\frac{\sum_{p\in\mathcal{P}_\mathcal{D}}B_p^2A_p}{\left(\sum_{p\in\mathcal{P}_\mathcal{D}}B_p\right)^2},\quad \Sigma(\mathcal{D},w^*)=\left(\sum_{p\in\mathcal{P}_\mathcal{D}}A_p^{-1}\right)^{-1}.$$

The assertion of part (b) is then proved by observing

$$\sqrt{\frac{\Sigma(\mathcal{D},w)}{\Sigma(\mathcal{D},w^*)}}=\frac{\sqrt{\sum_{p\in\mathcal{P}_\mathcal{D}}B_p^2A_p}\sqrt{\sum_{p\in\mathcal{P}_\mathcal{D}}A_p^{-1}}}{\sum_{p\in\mathcal{P}_\mathcal{D}}B_p}\geq1,$$

where the inequality is by the Cauchy-Schwarz inequality.

(c) By (8.19) and (8.26), as well as $\Delta^n_{i(p)}Z\to\Delta Z_{\tau_p}$, we deduce that the $\mathcal{F}$-conditional law of $\tilde{\zeta}_{n,\beta}(\mathcal{D},w)$ converges in probability to that of $\zeta_\beta(\mathcal{D},w)$ under any metric for weak convergence. From here, by using an argument similar to that in the proof of Theorem 1(b), we further deduce that

$$cv^{\alpha/2}_{n,\beta}\xrightarrow{\mathbb{P}}cv^{\alpha/2}_\beta, \tag{8.31}$$

where $cv^{\alpha/2}_\beta$ denotes the $(1-\alpha/2)$-quantile of the $\mathcal{F}$-conditional law of $\zeta_\beta(\mathcal{D},w)$. It is easy to see that the $\mathcal{F}$-conditional law of $\zeta_\beta(\mathcal{D},w)$ is symmetric. The assertion of part (c) then follows from part (a) and (8.31). *Q.E.D.*

Proof of Theorem 3. (a) Fix $S \in \mathbf{S}$ and let $m = \dim(S) - 1$. We consider a sequence of subsets Ω_n defined by

$$\Omega_n = \left\{ \begin{array}{c} \text{For every } 1 \leq i \leq \lfloor T/\Delta_n \rfloor \text{, if } ((i-1)\Delta_n, i\Delta_n] \text{ contains} \\ \text{some jump of } Z\text{, then this interval is contained in } (S_{j-1}, S_j] \\ \text{for some } 1 \leq j \leq m \text{ and it contains exactly one jump of } Z. \end{array} \right\}$$

Under Assumption 1, the process Z has finitely active jumps without any fixed time of discontinuity. Hence, $\mathbb{P}(\Omega_n) \to 1$, so we can restrict our calculation below on Ω_n without loss of generality.

Below, we write $h = (h_0, \ldots h_m)^\top$ and denote the log likelihood ratio by

$$L_n(h) = \log \frac{dP^n_{\theta_0 + \Delta_n^{1/2} h}}{dP^n_{\theta_0}}.$$

For each $i \geq 1$, we set $h(n,i) = h_j$, where j is the unique integer in $\{1, \ldots, m\}$ such that $i\Delta_n \in (S_{j-1}, S_j]$. On the set Ω_n, with $\theta = \theta_0 + \Delta_n^{1/2} h$, we have

$$\Delta_i^n X = \int_{(i-1)\Delta_n}^{i\Delta_n} b_s ds + \int_{(i-1)\Delta_n}^{i\Delta_n} \sigma_s dW_s + \begin{pmatrix} (1 + \Delta_n^{1/2} h(n,i)) \Delta_i^n J_Z \\ \left(\beta_0 + \Delta_n^{1/2} h_0\right) (1 + \Delta_n^{1/2} h(n,i)) \Delta_i^n J_Z + \Delta_i^n \varepsilon \end{pmatrix}.$$

To simplify notations, we denote for each $i \geq 1$,

$$\begin{aligned} x_{n,i} &\equiv \Delta_n^{-1/2} \int_{(i-1)\Delta_n}^{i\Delta_n} \sigma_s dW_s, \\ \bar{b}_{n,i} &\equiv \int_{(i-1)\Delta_n}^{i\Delta_n} b_s ds, \quad \bar{c}_{n,i} \equiv \Delta_n^{-1} \int_{(i-1)\Delta_n}^{i\Delta_n} c_s ds, \\ J_{n,i} &\equiv \begin{pmatrix} \Delta_i^n J_Z \\ \beta_0 \Delta_i^n J_Z + \Delta_i^n \varepsilon \end{pmatrix}, \quad d_{n,i} \equiv \begin{pmatrix} h(n,i) \\ h_0 + \beta_0 h(n,i) + \Delta_n^{1/2} h_0 h(n,i) \end{pmatrix}. \end{aligned}$$

Note that under Assumption 4, $(x_{n,i})_{i \geq 1}$ are independent conditional on $(b_t, \sigma_t, J_{Z,t}, \varepsilon_t)_{t \geq 0}$ and each $x_{n,i}$ is distributed as $\mathcal{N}(0, \bar{c}_{n,i})$. With these notations, we can write the log likelihood ratio explicitly as

$$L_n(h) = \sum_{i=1}^{\lfloor T/\Delta_n \rfloor} \Delta_i^n J_Z d_{n,i}^\top \bar{c}_{n,i}^{-1} x_{n,i} - \frac{1}{2} \sum_{i=1}^{\lfloor T/\Delta_n \rfloor} \Delta_i^n J_Z^2 d_{n,i}^\top \bar{c}_{n,i}^{-1} d_{n,i}. \tag{8.32}$$

Note that on Ω_n, $\Delta_i^n J_Z \neq 0$ only if $((i-1)\Delta_n, i\Delta_n]$ contains one (and only one) jump of Z. Therefore,

$$L_n(h) = \sum_{p \in \mathcal{P}} \Delta Z_{\tau_p} d_{n,i(p)}^\top \bar{c}_{n,i(p)}^{-1} x_{n,i(p)} - \frac{1}{2} \sum_{p \in \mathcal{P}} \Delta Z_{\tau_p}^2 d_{n,i(p)}^\top \bar{c}_{n,i(p)}^{-1} d_{n,i(p)}. \tag{8.33}$$

By Proposition 4.4.10 in Jacod and Protter (2012), $\left(x_{n,i(p)}\right)_{p\geq 1} \xrightarrow{\mathcal{L}\text{-}s} (R_p)_{p\geq 1}$. Under Assumption 5, the variables $(R_p)_{p\geq 1}$ are $\mathcal{F}$-conditionally independent, where the $\mathcal{F}$-conditional law of R_p is $\mathcal{N}\left(0, c_{\tau_p}\right)$; moreover, $\bar{c}_{n,i(p)} \to c_{\tau_p}$ a.s. for each $p \geq 1$. Further note that for each $p \geq 1$,

$$d_{n,i(p)} \longrightarrow D_p h. \tag{8.34}$$

where the matrix D_p is defined as

$$D_p \equiv \begin{pmatrix} 0 & 0_{j-1}^{\intercal} & 1 & 0_{m-j}^{\intercal} \\ 1 & 0_{j-1}^{\intercal} & \beta_0 & 0_{m-j}^{\intercal} \end{pmatrix} \quad \text{for } j \text{ such that } \tau_p \in (S_{j-1}, S_j]. \tag{8.35}$$

Since $\mathcal{P}$ is a.s. finite, we deduce (4.9) from (8.33) and (8.34), that is,

$$L_n(h) = h^\top \Gamma_n^{1/2} \psi_n - \frac{1}{2} h^\top \Gamma_n h + o_p(1), \tag{8.36}$$

where

$$\Gamma_n \equiv \sum_{p\in\mathcal{P}} \Delta Z_{\tau_p}^2 D_p^\top \bar{c}_{n,i(p)}^{-1} D_p, \quad \psi_n = \Gamma_n^{-1/2} \sum_{p\in\mathcal{P}} \Delta Z_{\tau_p} D_p^\top \bar{c}_{n,i(p)}^{-1} x_{n,i(p)}. \tag{8.37}$$

In addition, (4.10) follows with

$$\Gamma \equiv \sum_{p\in\mathcal{P}} \Delta Z_{\tau_p}^2 D_p^\top c_{\tau_p}^{-1} D_p, \quad \psi \equiv \Gamma^{-1/2} \sum_{p\in\mathcal{P}} \Delta Z_{\tau_p} D_p^\top c_{\tau_p}^{-1} R_p. \tag{8.38}$$

It is easy to verify that Γ defined in (8.38) equals to $\Gamma(S)$ defined by (4.17). To see, we make the following explicit calculation using (8.35),

$$D_p^\top c_{\tau_p}^{-1} D_p = \begin{pmatrix} \frac{1}{v_{\tau_p}^c} & 0_{j-1}^\top & \frac{\beta_0 - \beta_{\tau_p}^c}{v_{\tau_p}^c} & 0_{m-j}^\top \\ 0_{j-1} & 0_{(j-1)\times(j-1)} & 0_{j-1} & 0_{(j-1)\times(m-j)} \\ \frac{\beta_0 - \beta_{\tau_p}^c}{v_{\tau_p}^c} & 0_{j-1}^\top & \frac{\left(\beta_0 - \beta_{\tau_p}^c\right)^2}{v_{\tau_p}^c} + \frac{1}{c_{ZZ,\tau_p}} & 0_{m-j}^\top \\ 0_{m-j} & 0_{(m-j)\times(j-1)} & 0_{m-j} & 0_{(m-j)\times(m-j)} \end{pmatrix}. \tag{8.39}$$

Finally, we note that conditional on $\mathcal{F}$, ψ has a standard normal distribution and, hence, is independent of $\mathcal{F}$. The proof for the LAMN property is now complete.

From the proof of Theorem 3 of Jeganathan (1982), we see that the convolution theorem can be applied in restriction to the set $\Omega(S) \equiv \{\Gamma(S) \text{ is nonsingular}\}$. The information bound for estimating β, that is, the first diagonal element of $\Gamma(S)^{-1}$, can then be easily computed by using the inversion formula for partitioned matrices.

(b) Since the jumps of Z have finite activity, on each sample path $\omega \in \Omega$ there exists some $S^*(\omega) \in \mathbf{S}$ that shatters its jumps. That is, each interval $(S^*_{j-1}(\omega), S^*_j(\omega)]$ contains at exactly one jump time of Z. We can then evaluate $\bar{\Sigma}_\beta(\cdot)$ at S^* on each sample path and obtain

$$\bar{\Sigma}_\beta(S^*) = \left(\sum_{s \le T} \left(\frac{\Delta Z_s^2}{v_s^c} - \frac{\gamma_{1s}^2}{\gamma_{2s}} \right) \right)^{-1}. \tag{8.40}$$

Plugging the definitions of γ_{1s} and γ_{2s} (see (4.16)) into (8.40), we can verify that

$$\bar{\Sigma}_\beta(S^*) = \left(\sum_{s \le T} \frac{\Delta Z_s^2}{c_{YY,s} - 2\beta_0 c_{ZY,s} + \beta_0^2 c_{ZZ,s}} \right)^{-1}. \tag{8.41}$$

Recall that we fix $\mathcal{D} = [0,T] \times \mathbb{R}_*$ and $\Sigma^* \equiv \Sigma(\mathcal{D}, w^*)$, with the latter given by (4.8). Under Assumption 5, we see $\bar{\Sigma}_\beta(S^*) = \Sigma^*$.

It remains to verify that $\bar{\Sigma}_\beta(S^*) \ge \bar{\Sigma}_\beta(S)$ for all $S \in \mathbf{S}$. By the Cauchy–Schwarz inequality,

$$\frac{\left(\sum_{S_{j-1} < s \le S_j} \gamma_{1s}\right)^2}{\sum_{S_{j-1} < s \le S_j} \gamma_{2s}} \le \sum_{S_{j-1} < s \le S_j} \frac{\gamma_{1s}^2}{\gamma_{2s}}. \tag{8.42}$$

From (4.19), (8.40) and (8.42), $\bar{\Sigma}_\beta(S^*) \ge \bar{\Sigma}_\beta(S)$ readily follows. *Q.E.D.*

Proof of Theorem 4. We complement the notations in (4.29) with

$$\tilde{v}_{n,i\pm} \equiv (-\beta_0, 1)\, \hat{c}_{n,i\pm} (-\beta_0, 1)^\top. \tag{8.43}$$

Observe that v_t (recall (4.23)) is the spot covariance matrix of the process $Y - \beta_0 Z$. Then, by applying Theorem 13.3.3(c) of Jacod and Protter (2012) to the process $Y - \beta_0 Z$, we deduce that

$$k_n^{1/2} \left(\tilde{v}_{n,i(p)-} - v_{\tau_p -}, \tilde{v}_{n,i(p)+} - v_{\tau_p} \right)_{p \ge 1} \xrightarrow{\mathcal{L}\text{-}s} \left(\sqrt{2} v_{\tau_p -} \xi'_{p-}, \sqrt{2} v_{\tau_p} \xi'_{p+} \right)_{p \ge 1}. \tag{8.44}$$

Recall the notations in (4.24) and (4.29). For each $p \ge 1$, we can decompose

$$\hat{\phi}_{n,i(p)} = \phi_p + k_n^{-1/2} F_{n,p} + G_{n,p}, \tag{8.45}$$

where

$$\begin{cases} F_{n,p} \equiv k_n^{1/2} \left(\left(\tilde{v}_{n,i(p)-} + \tilde{v}_{n,i(p)+} \right)/2 - \phi_p \right), \\ G_{n,p} \equiv \hat{\phi}_{n,p} - \left(\tilde{v}_{n,i(p)-} + \tilde{v}_{n,i(p)+} \right)/2. \end{cases} \tag{8.46}$$

From (8.44), it follows that

$$(F_{n,p})_{p \ge 1} \xrightarrow{\mathcal{L}\text{-}s} (F_p)_{p \ge 1}. \tag{8.47}$$

We also see from Theorem 2(a) that $\tilde{\beta}_n - \beta_0 = O_p(\Delta_n^{1/2}) = o_p(k_n^{-1/2})$, so we further deduce

$$G_{n,p} = o_p(k_n^{-1/2}). \tag{8.48}$$

We now turn to the estimator $\hat{\beta}_n(\mathcal{D}, w^*)$. By (8.19), we have

$$\Delta_n^{-1/2}\left(\hat{\beta}_n(\mathcal{D}, w^*) - \beta_0\right) = \frac{\Delta_n^{-1/2}\sum_{p\in\mathcal{P}} \Delta_{i(p)}^n Z\left(\Delta_{i(p)}^n Y - \beta_0 \Delta_{i(p)}^n Z\right)/\hat{\phi}_{n,i(p)}}{\sum_{p\in\mathcal{P}} \Delta_{i(p)}^n Z^2/\hat{\phi}_{n,i(p)}} \quad \text{w.p.a.1.} \tag{8.49}$$

Recall the notations $R_{n,p}$ and $\varsigma_{n,p}$ from (8.10) and write $R_{n,p} = (R_{Z,n,p}, R_{Y,n,p})^\top$. We can rewrite (8.49) as

$$\Delta_n^{-1/2}\left(\hat{\beta}_n(\mathcal{D}, w^*) - \beta_0\right) = \frac{\sum_{p\in\mathcal{P}}\left(\Delta Z_{\tau_p} + \Delta_n^{1/2} R_{Z,n,p}\right)\varsigma_{n,p}/\hat{\phi}_{n,i(p)}}{\sum_{p\in\mathcal{P}}\left(\Delta Z_{\tau_p} + \Delta_n^{1/2} R_{Z,n,p}\right)^2/\hat{\phi}_{n,i(p)}} \quad \text{w.p.a.1.} \tag{8.50}$$

Next, we derive expansions for the numerator and the denominator of the right-hand side of (8.50) separately. Observe that the numerator satisfies

$$\begin{aligned}
&\sum_{p\in\mathcal{P}} \frac{\left(\Delta Z_{\tau_p} + \Delta_n^{1/2} R_{Z,n,p}\right)\varsigma_{n,p}}{\hat{\phi}_{n,i(p)}} - \sum_{p\in\mathcal{P}} \frac{\Delta Z_{\tau_p}\varsigma_{n,p}}{\phi_p} \\
&\quad = \sum_{p\in\mathcal{P}} \frac{\left(\Delta Z_{\tau_p} + \Delta_n^{1/2} R_{Z,n,p}\right)\varsigma_{n,p}\,\phi_p - \Delta Z_{\tau_p}\varsigma_{n,p}\left(\phi_p + k_n^{-1/2} F_{n,p} + G_{n,p}\right)}{\hat{\phi}_{n,i(p)}\phi_p} \\
&\quad = -k_n^{-1/2}\sum_{p\in\mathcal{P}} \frac{\Delta Z_{\tau_p}\varsigma_{n,p} F_{n,p}}{\hat{\phi}_{n,i(p)}\phi_p} + \sum_{p\in\mathcal{P}} \frac{\Delta_n^{1/2} R_{Z,n,p}\varsigma_{n,p}\,\phi_p - \Delta Z_{\tau_p}\varsigma_{n,p} G_{n,p}}{\hat{\phi}_{n,i(p)}\phi_p} \\
&\quad = -k_n^{-1/2}\sum_{p\in\mathcal{P}} \frac{\Delta Z_{\tau_p}\varsigma_{n,p} F_{n,p}}{\phi_p^2} + o_p(k_n^{-1/2}),
\end{aligned} \tag{8.51}$$

where the first equality is obtained by using (8.45); the second equality is obvious; the third equality follows from $R_{Z,n,p} = O_p(1)$, $\varsigma_{n,p} = O_p(1)$, $\hat{\phi}_{n,i(p)} - \phi_p = o_p(1)$ and (8.48). Similarly, the denominator of the right-hand side of (8.50) satisfies

$$\begin{aligned}
&\sum_{p\in\mathcal{P}} \frac{\left(\Delta Z_{\tau_p} + \Delta_n^{1/2} R_{Z,n,p}\right)^2}{\hat{\phi}_{n,i(p)}} - \sum_{p\in\mathcal{P}} \frac{\Delta Z_{\tau_p}^2}{\phi_p} \\
&= \sum_{p\in\mathcal{P}} \frac{\left(\Delta Z_{\tau_p} + \Delta_n^{1/2} R_{Z,n,p}\right)^2 \phi_p - \Delta Z_{\tau_p}^2 \left(\phi_p + k_n^{-1/2} F_{n,p} + G_{n,p}\right)}{\hat{\phi}_{n,i(p)}\phi_p} \\
&= -k_n^{-1/2} \sum_{p\in\mathcal{P}} \frac{\Delta Z_{\tau_p}^2 F_{n,p}}{\hat{\phi}_{n,i(p)}\phi_p} \\
&\quad + \sum_{p\in\mathcal{P}} \frac{\left(2\Delta_n^{1/2}\Delta Z_{\tau_p} R_{Z,n,p} + \Delta_n R_{Z,n,p}^2\right)\phi_p - \Delta Z_{\tau_p}^2 G_{n,p}}{\hat{\phi}_{n,i(p)}\phi_p} \\
&= -k_n^{-1/2} \sum_{p\in\mathcal{P}} \frac{\Delta Z_{\tau_p}^2 F_{n,p}}{\phi_p^2} + o_p(k_n^{-1/2}).
\end{aligned} \tag{8.52}$$

Finally, we plug the expansions (8.51) and (8.52) into (8.50) and deduce, w.p.a.1,

$$\begin{aligned}
&\Delta_n^{-1/2}\left(\hat{\beta}_n\left(\mathcal{D}, w^*\right) - \beta_0\right) \\
&= \frac{\sum_{p\in\mathcal{P}} \Delta Z_{\tau_p} \varsigma_{n,p}/\phi_p - k_n^{-1/2} \sum_{p\in\mathcal{P}} \Delta Z_{\tau_p} \varsigma_{n,p} F_{n,p}/\phi_p^2 + o_p(k_n^{-1/2})}{\sum_{p\in\mathcal{P}} \Delta Z_{\tau_p}^2/\phi_p - k_n^{-1/2} \sum_{p\in\mathcal{P}} \Delta Z_{\tau_p}^2 F_{n,p}/\phi_p^2 + o_p(k_n^{-1/2})} \\
&= \zeta_{n,\beta}^*\left(\mathcal{D}\right) + k_n^{-1/2} H_{n,\beta}^*\left(\mathcal{D}\right) + o_p(k_n^{-1/2}),
\end{aligned} \tag{8.53}$$

where

$$\begin{cases}
\zeta_{n,\beta}^*\left(\mathcal{D}\right) \equiv \left(\sum_{p\in\mathcal{P}} \frac{\Delta Z_{\tau_p} \varsigma_{n,p}}{\phi_p}\right) \Big/ \left(\sum_{p\in\mathcal{P}} \frac{\Delta Z_{\tau_p}^2}{\phi_p}\right), \\
H_{n,\beta}^*\left(\mathcal{D}\right) \equiv \frac{\left(\sum_{p\in\mathcal{P}} \frac{\Delta Z_{\tau_p}^2 F_{n,p}}{\phi_p^2}\right)\left(\sum_{p\in\mathcal{P}} \frac{\Delta Z_{\tau_p} \varsigma_{n,p}}{\phi_p}\right) - \left(\sum_{p\in\mathcal{P}} \frac{\Delta Z_{\tau_p} \varsigma_{n,p} F_{n,p}}{\phi_p^2}\right)\left(\sum_{p\in\mathcal{P}} \frac{\Delta Z_{\tau_p}^2}{\phi_p}\right)}{\left(\sum_{p\in\mathcal{P}} \frac{\Delta Z_{\tau_p}^2}{\phi_p}\right)^2}.
\end{cases} \tag{8.54}$$

We now observe that the estimators $\hat{c}_{n,i-}$ and $\hat{c}_{n,i+}$ do not involve the increment $\Delta_i^n X$. From here, it is easy to see that the convergence in (8.12) and (8.47) hold jointly with $\mathcal{F}$-conditionally independent limits, that is,

$$\left(\varsigma_{n,p}, F_{n,p}\right)_{p\geq 1} \xrightarrow{\mathcal{L}\text{-}s} \left(\varsigma_p, F_p\right)_{p\geq 1}. \tag{8.55}$$

By properties of stable convergence, we deduce

$$\left(\zeta_{n,\beta}^*\left(\mathcal{D}\right), H_{n,\beta}^*\left(\mathcal{D}\right)\right) \xrightarrow{\mathcal{L}\text{-}s} \left(\zeta_{\beta}^*\left(\mathcal{D}\right), H_{\beta}^*\left(\mathcal{D}\right)\right). \tag{8.56}$$

This finishes the proof. *Q.E.D.*

References

Aït-Sahalia, Y., and J. Jacod (2009): "Testing for Jumps in a Discretely Observed Process," *Annals of Statistics*, 37, 184–222.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and C. Vega (2003): "Micro Effects of Macro Announcements: Real-Time Price Discovery in Foreign Exchange," *The American Economic Review*, 93(1), 251 – 277.

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

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 Rconomics 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.

Comte, F., and E. Renault (1996): "Long memory continuous time models," *Journal of Econometrics*, 73, 101–149.

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

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

Huang, X., and G. Tauchen (2005): "The Relative Contributions of Jumps to Total Variance," *Journal of Financial Econometrics*, 3, 456–499.

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

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 A. N. Shiryaev (2003): *Limit Theorems for Stochastic Processes*. Springer-Verlag, New York, second edn.

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), pp. 66–87.

JIANG, G. J., AND R. C. OOMEN (2008): "Testing for Jumps when Asset Prices are Observed with Noise - A "Swap Variance" Approach," *Journal of Econometrics*, 144, 352–370.

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.

LEHMANN, E. L., AND J. P. ROMANO (2005): *Testing Statistical Hypothesis*. Springer.

LETTAU, M., M. MAGGIORI, AND M. WEBER (2014): "Conditional Risk Premia in Currency Markets and Other Asset Classes.," Forthcoming in the Journal of Financial Economics.

LI, J., V. TODOROV, AND G. TAUCHEN (2014): "Adaptive Estimation of Continuous-Time Regression Models using High-Frequency Data," Discussion paper, Duke Univeristy.

MYKLAND, P., AND L. ZHANG (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.

REISS, M., V. TODOROV, AND G. TAUCHEN (2014): "Nonparametric Test for a Constant Beta between Ito Semimartingales based on High-Frequency Data," Discussion paper, Northwestern University and Humboldt University.

RENAULT, E., C. SARISOY, AND B. J. WERKER (2014): "Efficient Estimation of Integrated Volatility and Related Processes," Discussion paper, Brown University.

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.

STOCK, J. H. (1994): "Unit Roots, Structural Breaks and Trends," in *Handbook of Econometrics, volume 4*, ed. by R. Engle, and D. McFadden, pp. 2739–2841. Elsevier, Amsterdam.

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.

VAN DER VAART, A. W. (1998): *Asymptotic Statistics*. Cambridge University Press.