Variation and Efficiency of High-Frequency Betas*

Congshan Zhang,[†] *Jia Li,*[‡] *Viktor Todorov,*[§] *and George Tauchen*[¶]

May 20, 2020

Abstract

This paper studies the efficient estimation of betas from high-frequency return data on a fixed time interval. Under an assumption of equal diffusive and jump betas, we derive the semiparametric efficiency bound for estimating the common beta and develop an adaptive estimator that attains the efficiency bound. We further propose a Hausman type test for deciding whether the common beta assumption is true from the high-frequency data. In our empirical analysis we provide examples of stocks and time periods for which a common market beta assumption appears true and ones for which this is not the case. We further quantify empirically the gains from the efficient common beta estimation developed in the paper.

Keywords: adaptive estimation; beta; high frequency data; jump; semiparametric efficiency; volatility.

JEL Classification: C51, C52, G12.

*Todorov's research is partially supported by NSF grant SES-1530748. We would like to thank the Co-Editor (Xiaohong Chen) and anonymous referees for numerous constructive comments and suggestions.

[†]Department of Economics, Duke University, Durham, NC 27708; e-mail: congshan.zhang@duke.edu.

[‡]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

In this paper we study the problem of efficient estimation of asset betas from high-frequency return data on a fixed time interval. The estimation of factor loadings, or betas, is of central importance for practical risk management. This estimation constitutes also a critical step in evaluating the ability of asset pricing models to explain the cross-sectional behavior of asset prices.[1] If betas stay constant over a (short) period of time, then one can use high-frequency records of the factors and the asset prices within the time interval to estimate the betas. The goal of this paper is to derive efficient methods for doing so when betas associated with continuous and discontinuous moves in the factors are imposed to be the same, and to further develop tests for this assumption. More specifically, we are interested in the following continuous-time regression:

$$Y_t = B_t + \beta^c Z_t^c + \beta^d Z_t^d + \epsilon_t^c + \epsilon_t^d, \quad t \in [0, T], \tag{1}$$

where Y is the asset price, B is a continuous process of finite variation (i.e., the drift), Z^c and Z^d are the continuous and the discontinuous local martingale components of the discretely observed factor Z, and ϵ^c and ϵ^d are the residual continuous and discontinuous martingale components of Y which are orthogonal to Z^c and Z^d in a martingale sense, that is, their quadratic covariations with the factors are zero (this is formally defined in the main text).

There has been a lot of work on the estimation of betas using high-frequency data. Most of the existing work is about the estimation of the diffusive beta, β^c, with jumps either not allowed in the model setup or the inference being robust to their presence; see, for example, Bollerslev and Zhang (2003), Barndorff-Nielsen and Shephard (2004a), Andersen et al. (2006), Mykland and Zhang (2006, 2009), Todorov and Bollerslev (2010), Gobbi and Mancini (2012), Patton and Verardo (2012), Kalnina (2012) and Li et al. (2017a) among many others. In particular, Li et al. (2017a) propose an adaptive estimator that attains the semiparametric efficiency bound for recovering β^c. That is, the adaptive estimator attains the same asymptotic variance as that of the infeasible estimator with all nonparametric nuisance (e.g., the stochastic volatility of the factor process and the jumps of all processes) known. There is much less work on the estimation of the jump beta, β^d. Todorov and Bollerslev (2010) propose estimators based on higher-order power variation while Li et al. (2017b) consider semiparametrically efficient estimation based on estimates of the jump (co)variation by optimally weighting the information in the detected jumps in order to account for the

[1]Indeed, the error in recovering betas can carry over to the second stage cross-sectional regressions of the Fama-Macbeth procedures for estimating risk premia; see, for example, Shanken (1992), Jaganathan and Wang (1998), Kan and Zhang (1999), Gospodinov et al. (2009) and Kleibergen (2009).

heteroskedasticity present in the data. In contrast to the adaptive estimation result for the diffusive beta, the efficient estimation of jump beta is shown to be generally not adaptive with respect to the sizes of jumps in Z.

In the existing work to date, typically the discontinuous (resp. continuous) part of the asset price is treated as a nuisance component when estimating the continuous (resp. discontinuous) beta. In this paper, we study a "common beta" restriction $\beta^c \equiv \beta^d$ and explore the efficiency gains such an assumption may provide for making inference for asset betas. More specifically, our goal is to address three related theoretical questions that have not yet been studied in the aforementioned literature: (i) What is the semiparametric efficiency bound for beta estimation under the common beta restriction? (ii) Is this bound an adaptive one with respect to nonparametric nuisances? (iii) Is the efficiency bound attainable by a feasible estimator? In doing so, we aim to provide a more complete theoretical understanding on the issue of efficient beta estimation.

The theoretical contribution of this paper is to provide positive answers to the above questions. We establish the semiparametric efficiency bound for beta estimation under the common beta restriction, verify that it is in fact adaptive, and propose a (feasible) estimator that attains this efficiency bound. Specifically, to show the adaptiveness of the proposed estimator, we establish the local asymptotic mixed normality (LAMN) property in the infeasible model (with nonparametric nuisance known), which then allows us to derive the Cramer–Rao efficiency bound for estimating the common beta by using the convolution theorem. We then show that the bound can be attained by the proposed estimator. Hence, *a fortiori*, the efficiency bound is sharp and the proposed estimator is adaptive (consequently, semiparametrically efficient). This adaptiveness result in the common beta model is somewhat surprising because the semiparametrically efficient estimator of the jump beta is generally *not* adaptive to the sizes of jumps in Z. The interesting finding here is that adaptiveness is "recovered" under the common beta restriction.

The proposed adaptive estimator belongs to a class of weighted estimators for the (common) beta coefficient, and is constructed by optimally choosing the weight functions for both the continuous and jump returns, taking into consideration the heteroskedasticity in the data. Under general regularity conditions, we derive a feasible limit theory for this class of estimators. This class also nests various existing estimators by appropriately choosing the weight functions. For example, the "realized regression" beta estimator of Barndorff-Nielsen and Shephard (2004a) (see also Todorov and Bollerslev (2010) and Gobbi and Mancini (2012)), the block-based estimator of Mykland and Zhang (2009), the optimally weighted diffusive beta estimator, and the optimally weighted jump beta estimator in Li et al. (2017a,b) all fall into this framework as special cases.

As a by-product of the efficiency result, we propose a Hausman-type specification test (Hausman (1978)). The test compares two estimators for the common beta: the efficient adaptive estimator and an inefficient one from the existing literature. Under the common beta null hypothesis, the difference of these estimators is asymptotically centered mixed Gaussian with variance given by the difference between those of the inefficient and the efficient estimators. The test has power against alternatives under which the weighted integrated beta functionals associated with these two estimators are different. Note that under the alternative, the two estimators are no longer comparable, and they are consistent for their own integrated beta functionals.

We document satisfactory performance of the proposed estimation and inference techniques developed in the paper on simulated data from realistically calibrated models. In an empirical application, we illustrate the efficiency gain of the adaptive estimator relative to the commonly used estimator proposed by Barndorff-Nielsen and Shephard (2004a).

The rest of the paper is organized as follows. Section 2 presents the formal setup. Section 3 introduces our estimator and establishes the adaptive estimation result for the common beta. Section 4 contains a Monte Carlo study and Section 5 provides an empirical illustration. We conclude in Section 6. All proofs are given in Section 7. Additional robustness checks for our numerical results are provided in the supplemental appendix.

2 Setup

We start with introducing some notation that will be used throughout. We denote with $\xrightarrow{\mathbb{P}}$, $\xrightarrow{\mathcal{L}}$ and $\xrightarrow{\mathcal{L}\text{-}s}$ convergence in probability, convergence in law, and stable convergence in law, respectively. All limits are for $n \to \infty$, and the asymptotics is of infill type on a fixed time interval $[0, T]$.

The set of real numbers is $\mathbb{R}$ and $\mathbb{R}_* = \mathbb{R} \setminus \{0\}$. We denote by $\mathcal{M}_d$ the space of all $d \times d$ positive semidefinite matrices. The Euclidean norm of a finite-dimensional vector space is $\|\cdot\|$. The integer part of $x \in \mathbb{R}$ is $[x]$. We write $x \wedge y$ to denote the smaller number of x and y. For a matrix A, its transpose is denoted by $A^\top$, and its (j,k) element is A_{jk}, while $vec(\cdot)$ is the column vectorization operator. For matrix differentiation, if f is a generic differentiable function defined on $\mathcal{M}_d$, then $\partial_{jk} f(A) \equiv \partial f(A)/\partial A_{jk}$ and $\partial^2_{jk,lm} f(A) \equiv \partial^2 f(A)/\partial A_{jk} \partial A_{lm}$. Finally, we write $a_n \asymp b_n$ if for some constant $C \geq 1$, we have $a_n/C \leq b_n \leq C a_n$.

2.1 The underlying processes

We start with some regularity conditions for the processes Z and Y. These processes are defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$. We denote $X = (Z, Y)^\top$ and assume that X is an Itô semimartingale of the form

$$X_t = X_0 + \int_0^t b_s ds + \int_0^t \sigma_s dW_s + J_t, \quad J_t \equiv \sum_{s\leq t} \Delta X_s, \tag{2}$$

where b_t takes value in $\mathbb{R}^2$, the volatility process σ_t takes value in $\mathcal{M}_2$, and W is a 2-dimensional standard Brownian motion. 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 process $J_t = (J_{Z,t}, J_{Y,t})^\top$ can be alternatively written as $\int_0^t \int_{\mathbb{R}} \delta(\omega, s, u)\mu(ds, du)$, where $\delta : \Omega \times \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}^2$ is a predictable function and μ is a Poisson random measure on $\mathbb{R}_+ \times \mathbb{R}$ with its deterministic compensator $\nu(dt, du) = dt \otimes \lambda(du)$ for some σ-finite measure λ on $\mathbb{R}$. The spot covariance matrix of X at time t is defined by

$$c_t \equiv \sigma_t \sigma_t^\top = \begin{pmatrix} c_{ZZ,t} & c_{ZY,t} \\ c_{ZY,t} & c_{YY,t} \end{pmatrix}. \tag{3}$$

Assumption 1 *(a) The process b is locally bounded; (b) $\nu([0,T] \times \mathbb{R}) < \infty$; (c) c_t is non-singular for $t \in [0,T]$ and it is an Itô semimartingale of the form*

$$\begin{aligned} vec(c_t) &= vec(c_0) + \int_0^t \tilde{b}_s ds + \int_0^t \tilde{\sigma}_s d\widetilde{W}_s \\ &\quad + \int_0^t \int_{\mathbb{R}} \tilde{\delta}(s,z) 1_{\{\|\tilde{\delta}(s,z)\| \leq 1\}} (\tilde{\mu} - \tilde{\nu})(ds, dz) \\ &\quad + \int_0^t \int_{\mathbb{R}} \tilde{\delta}(s,z) 1_{\{\|\tilde{\delta}(s,z)\| > 1\}} \tilde{\mu}(ds, dz), \end{aligned} \tag{4}$$

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

Assumption 1 is quite standard for the analysis of high-frequency data. We note that part (b) requires the jumps in X to be of finite activity. In the context of the current paper, we extract jump information only through "large" enough jumps in our analysis on efficient estimation; we thus impose this assumption to simplify the exposition, although it may be further relaxed. Assumption 1(c) allows for the so-called "leverage effect," that is, correlation between W and $\widetilde{W}$. Moreover, we allow for volatility jumps and do not restrict their activity and dependence with the price jumps.

2.2 The common beta model

We now introduce formally the continuous-time regression model. Let Y be decomposed as in equation (1). First, the orthogonality conditions between the diffusive and jump components in Y can be formally stated as

$$\langle Z^c, \epsilon^c \rangle_t = 0, \quad [Z^d, \epsilon^d]_t = \sum_{s \le t} \Delta Z_s^d \Delta \epsilon_s^d = 0, \quad \text{for } t \in [0, T], \tag{5}$$

where $\langle \cdot, \cdot \rangle$ denotes the quadratic covariation of two continuous local martingales and $[\cdot, \cdot]$ denotes the quadratic covariation in general. Under the "common beta" assumption, namely $\beta^c = \beta^d$, we can write (1) more succinctly as

$$Y_t = B_t + \beta(Z_t^c + Z_t^d) + (\epsilon_t^c + \epsilon_t^d), \tag{6}$$

where the common beta β is the parameter of interest and the residual $\epsilon_t = \epsilon_t^c + \epsilon_t^d$ is a jump-diffusion process that is orthogonal to $(Z_t)_{t \ge 0}$, namely,

$$[Z, \epsilon]_t = 0, \quad \text{for } t \in [0, T]. \tag{7}$$

Equivalently, we can write the continuous-time regression model in (6) more explicitly as

$$\begin{cases} dZ_t = b_{Z,t}dt + \sqrt{c_{ZZ,t}}dW_{Z,t} + dJ_{Z,t}, \\ dY_t = b_{Y,t}dt + \beta\sqrt{c_{ZZ,t}}dW_{Z,t} + \sqrt{\varsigma_t}dW_{\epsilon,t} + \beta dJ_{Z,t} + dJ_{\epsilon,t}, \end{cases} \tag{8}$$

where W_Z and W_ϵ are univariate independent Brownian motions, J_ϵ is a pure jump process that contains the Y-specific jumps, and ς is the spot idiosyncratic variance of Y. It is easy to see that the processes in (2) and (8) are connected in the following way

$$b_t = \begin{pmatrix} b_{Z,t} \\ b_{Y,t} \end{pmatrix}, \quad \sigma_t = \begin{pmatrix} \sqrt{c_{ZZ,t}} & 0 \\ \beta\sqrt{c_{ZZ,t}} & \sqrt{\varsigma_t} \end{pmatrix}, \quad W_t = \begin{pmatrix} W_{Z,t} \\ W_{\epsilon,t} \end{pmatrix}, \quad J_t = \begin{pmatrix} J_{Z,t} \\ \beta J_{Z,t} + J_{\epsilon,t} \end{pmatrix}. \tag{9}$$

Moreover, a simple calculation implies $\beta = c_{ZY,t}/c_{ZZ,t}$, and the spot idiosyncratic variance ς_t is connected with c_t via

$$\varsigma_t \equiv c_{YY,t} - c_{ZY,t}^2/c_{ZZ,t}. \tag{10}$$

The common beta model in (6) can be trivially extended to the more general setting where Z is multidimensional. Such an extension is important from an empirical point of view in light of existing evidence for multiple risk factors in asset prices and it importantly can make an assumption of time-invariant beta weaker. In addition, the interval $[0, T]$ for which the common beta model is assumed to hold can be trivially replaced with any finite union of disjoint time intervals. We do not consider these extensions in the analysis that follows as they are trivial from a theoretical point of view and make notation somewhat cumbersome.

3 Adaptive estimation of the common beta

3.1 Asymptotic properties of weighted estimators

In this subsection, we introduce a class of weighted estimators for the common beta and establish their asymptotic properties. The issue of efficiency and adaptiveness will be addressed in Section 3.2 below. We denote the true value of the common beta as β_0 and use β as a generic reference for it.

Our inference is done on the basis of high-frequency data on the fixed interval $[0, T]$. More specifically, the vector X is discretely observed at times $i\Delta_n$, for $1 \leq i \leq [T/\Delta_n]$ with a discretization mesh $\Delta_n \equiv 1/n$ going to 0 asymptotically. The increments of X are denoted by

$$\Delta_i^n X \equiv X_{i\Delta_n} - X_{(i-1)\Delta_n}, \text{ for } i = 1, 2, ..., [T/\Delta_n].$$

Since we are going to exploit the information about β from both the diffusive and jump moves, but in distinct ways, we first need to identify high-frequency intervals that contain jumps of Z. We collect the jump times of Z in the set $\mathcal{T} \equiv \{\tau_p : p \in \mathcal{P}\}$, where $\mathcal{P} \equiv \{p \geq 1 : \tau_p \leq T\}$ denotes the indices of the jumps. Both sets are finite almost surely since Z has finite-activity jumps. To identify jumps from the high-frequency data, we use the (standard) truncation method based on a real sequence $(u_n)_{n\geq 1}$ of truncation thresholds, which satisfies the following assumption:

Assumption 2 *For some $\varpi \in (0, 1/2)$, $u_n \asymp \Delta_n^{\varpi}$.*

For each $p \in \mathcal{P}$, we denote by $i(p)$ the unique random index i such that $\tau_p \in ((i-1)\Delta_n, i\Delta_n]$. Let $\mathcal{I}_n \equiv \{i : 1 \leq i \leq [T/\Delta_n], |\Delta_i^n Z| > u_n\}$ and $\mathcal{I} \equiv \{i(p) : p \in \mathcal{P}\}$. Proposition 1 of Li et al. (2017b) shows that $\mathcal{I}_n = \mathcal{I}$ with probability approaching one. That is, in the limit, we can use $\mathcal{I}_n$ to consistently disentangle jumps from the continuous moves using this truncation methodology.[2]

For the inference on the common beta, we also need to estimate the path of the spot covariance matrix of X. To do this, we pick an integer sequence $(k_n)_{n\geq 1}$ of local window sizes, which eventually goes to infinity, and a sequence of $\mathbb{R}^2$-valued truncation thresholds $(u_n')_{n\geq 1}$.[3] These two sequences satisfy the following condition:

[2]Since this result holds with probability approaching one, the detection error has no asymptotic effect on the subsequent asymptotic analysis.

[3]The truncation sequences u_n and u_n' satisfy the same assumptions. Theoretically speaking, one could set the components of u_n' simply to be u_n. By distinguishing these two sequences, we allow the truncation

Assumption 3 *For some* $\rho \in (0,1)$*, and* $\varpi \in (0, 1/2)$*,* $k_n \asymp \Delta_n^{-\rho}$ *and* $u'_n \asymp \Delta_n^{\varpi}$*.*

The spot covariance matrix at time $i\Delta_n$ can then be recovered locally using returns over the time interval $[i\Delta_n, (i+k_n)\,\Delta_n]$, for $0 \leq i \leq [T/\Delta_n] - k_n$, through a truncated variation estimator due to Mancini (2001):

$$\hat{c}_{n,i} \equiv \frac{1}{k_n \Delta_n} \sum_{j=1}^{k_n} \Delta_{i+j}^n X \Delta_{i+j}^n X^\top 1_{\{-u'_n \leq \Delta_{i+j}^n X \leq u'_n\}}. \tag{11}$$

When a jump is detected, we also define a pre-jump covariance estimator in a similar way:

$$\hat{c}_{n,i-} \equiv \frac{1}{k_n \Delta_n} \sum_{j=0}^{k_n - 1} \Delta_{i-k_n+j}^n X \Delta_{i-k_n+j}^n X^\top 1_{\{-u'_n \leq \Delta_{i-k_n+j}^n X \leq u'_n\}}, \text{ for } i \in \mathcal{I}'_n, \tag{12}$$

where $\mathcal{I}'_n \equiv \{i \in \mathcal{I}_n : k_n + 1 \leq i \leq [T/k_n] - k_n\}$. We note that $\mathcal{I}'_n$ differs from the index set $\mathcal{I}_n$ only due to its exclusion of the boundary terms that are needed for the spot estimation. But this difference is asymptotically negligible for the subsequent analysis.

We are now ready to construct a class of weighted estimators for the common beta. It turns out that, in order to obtain the (efficient) adaptive estimator, we need to weight diffusive and jump returns in distinct ways. Hence, we consider two weight functions $v : \mathcal{M}_2 \mapsto (0, \infty)$ and $w : \mathcal{M}_2 \times \mathcal{M}_2 \times \mathbb{R} \mapsto (0, \infty)$ for the diffusive and jump parts, respectively, such that the following assumption holds:

Assumption 4 *(a) There is a sequence* $(T_m)_{m \geq 1}$ *of stopping times increasing to infinity and a sequence of convex compact subsets* $\mathcal{K}_m \subset \mathcal{M}_2$ *such that* $c_t \in \mathcal{K}_m$ *for* $t \leq T_m$ *and* $v(\cdot)$ *is three-time continuously differentiable on an* ε*-enlargement about* $\mathcal{K}_m$ *for some* $\varepsilon > 0$*; (b)* $w(\cdot)$ *is continuous at* (c_-, c, β_0) *for any* $c_-, c \in \mathcal{M}_2$*.*[4]

Assumption 4(a) is used to relax the polynomial growth condition on test functions imposed by Jacod and Rosenbaum (2013) for estimating integrated volatility functionals, and is easy to verify in typical applications. Assumption 4(b) is needed for using a continuous mapping argument.

For any twice continuously differentiable function $g(\cdot)$, we set $S(g) \equiv \int_0^T g(c_s)ds$ and define its bias-corrected estimator as

$$\hat{S}_n(g) \equiv \Delta_n \sum_{i=0}^{[T/\Delta_n]-k_n} \left(g(\hat{c}_{n,i}) - \frac{1}{k_n} \mathbb{B} g(\hat{c}_{n,i}) \right) \tag{13}$$

used in spot volatility estimation to be different from that for detecting jumps, and (possibly) different across assets.

[4]The ε-enlargement of $\mathcal{K}_m$ is defined as the collection of points whose Euclidean distance from the set $\mathcal{K}_m$ is less than ε.

where the function $\mathbb{B}g(\cdot)$ in the bias-correction term is defined as

$$\mathbb{B}g(c) \equiv \frac{1}{2} \sum_{j,k,l,m=1}^{2} \partial^2_{jk,lm} g(c)(c_{jl}c_{km} + c_{jm}c_{kl}). \tag{14}$$

This bias-correction term is needed for correcting a nonlinearity bias in the estimation of integrated volatility functionals as in, for example, Jacod and Rosenbaum (2013).

We consider a class of weighted estimators for the common beta constructed as

$$\hat{\beta}_n(v,w) \equiv \frac{\hat{S}_n\left(vg_b\right) + \sum_{i\in\mathcal{I}'_n} w\left(\hat{c}_{n,i-}, \hat{c}_{n,i}, \breve{\beta}_n\right)\Delta_i^n Z \Delta_i^n Y}{\hat{S}_n\left(v\right) + \sum_{i\in\mathcal{I}'_n} w\left(\hat{c}_{n,i-}, \hat{c}_{n,i}, \breve{\beta}_n\right)\left(\Delta_i^n Z\right)^2}, \tag{15}$$

where the function $g_b(c) \equiv c_{ZY}/c_{ZZ}$ and $\breve{\beta}_n \equiv \sum_{i\in\mathcal{I}'_n} \Delta_i^n Z \Delta_i^n Y / \sum_{i\in\mathcal{I}'_n} \left(\Delta_i^n Z\right)^2$ is a pilot estimator of β using only jump returns. The estimator $\hat{\beta}_n$ essentially combines the weighted diffusive beta of Li et al. (2017a)

$$\hat{\beta}_n^c\left(v\right) \equiv \frac{\hat{S}_n\left(vg_b\right)}{\hat{S}_n\left(v\right)}, \tag{16}$$

and the weighted jump beta of Li et al. (2017b)

$$\hat{\beta}_n^J\left(w\right) \equiv \frac{\sum_{i\in\mathcal{I}'_n} w\left(\hat{c}_{n,i-}, \hat{c}_{n,i}, \breve{\beta}_n\right)\Delta_i^n Z \Delta_i^n Y}{\sum_{i\in\mathcal{I}'_n} w\left(\hat{c}_{n,i-}, \hat{c}_{n,i}, \breve{\beta}_n\right)\left(\Delta_i^n Z\right)^2}. \tag{17}$$

While optimally weighted $\hat{\beta}_n^c$ and $\hat{\beta}_n^J$ have been shown by these prior work to be semiparametrically efficient for the diffusive beta and the jump beta, respectively, it is intuitively clear that they cannot be efficient when the common beta assumption holds, in that neither of them exploits the full information from *all* observed data.[5]

We now describe precisely the optimally weighted estimator for the common beta, which is actually adaptive as shown in Section 3.2 below. This adaptive estimator is given by

$$\hat{\beta}_n^\star \equiv \frac{\left(1 - \frac{3}{k_n}\right)\Delta_n \sum_{i=0}^{[T/\Delta_n]-k_n} \frac{\hat{c}_{ZY,n,i}}{\hat{\varsigma}_{n,i}} + \sum_{i\in\mathcal{I}'_n} \frac{\Delta_i^n Z \Delta_i^n Y}{\hat{\varsigma}_{n,i}}}{\left(1 - \frac{3}{k_n}\right)\Delta_n \sum_{i=0}^{[T/\Delta_n]-k_n} \frac{\hat{c}_{ZZ,n,i}}{\hat{\varsigma}_{n,i}} + \sum_{i\in\mathcal{I}'_n} \frac{\left(\Delta_i^n Z\right)^2}{\hat{\varsigma}_{n,i}}}, \tag{18}$$

[5]Although the number of jump returns is, asymptotically speaking, "much smaller" than the number of diffusive returns, the information content of the former is not negligible relative to the latter; this is because the jump returns carry a "much higher" signal-to-noise ratio compared to the diffusive ones at high frequency.

where $\hat{\varsigma}_{n,i} \equiv \hat{c}_{YY,n,i} - \hat{c}^2_{ZY,n,i}/\hat{c}_{ZZ,n,i}$ is the spot estimator for the diffusive idiosyncratic variance $\varsigma_{i\Delta_n}$ of Y (recall the definition (10)). Note that $\hat{\beta}^\star_n$ is a special case of $\hat{\beta}_n$, corresponding to weight functions[6]

$$v^\star(c) = \frac{c_{ZZ}}{c_{YY} - c^2_{ZY}/c_{ZZ}}, \quad w^\star(c_-, c, \beta) = \frac{1}{c_{YY} - c^2_{ZY}/c_{ZZ}}.$$

Theorem 1, below, characterizes the asymptotic properties of the general weighted estimator $\hat{\beta}_n$ and shows the optimality of the adaptive estimator $\hat{\beta}^\star_n$ within this class. To simplify notation, we set

$$Q_{n,ZZ}(w) \equiv \sum_{i \in \mathcal{I}'_n} w(\hat{c}_{n,i-}, \hat{c}_{n,i}, \breve{\beta}_n) \left(\Delta^n_i Z\right)^2, \quad Q_{ZZ}(w) \equiv \sum_{\tau \in \mathcal{T}} w(c_{\tau-}, c_\tau, \beta_0) \left(\Delta Z_\tau\right)^2.$$

We also assume the following.

Assumption 5 *ς and Z do not jump at the same times almost surely.*

Assumption 5 says that the idiosyncratic variance ς of the individual stock price Y does not co-jump with the factor Z. Although the asymptotic distribution of $\hat{\beta}_n$ can be derived without this assumption, we impose it nonetheless so that the limiting distribution of $\hat{\beta}_n$ is $\mathcal{F}$-conditionally mixed Gaussian; otherwise, the limiting distribution would be "doubly" mixed Gaussian with an extra layer of mixing due to the indeterminancy of the exact jump time within the corresponding sampling interval. We note that this assumption does not rule out the co-jump between the prices and volatilities of Y and Z, as it only concerns the idiosyncratic variance of Y.

Theorem 1 *Suppose Assumptions 1–5 hold and $k_n \asymp \Delta_n^{-\rho}$ such that $\rho \in (1/3, 1/2)$ and $(1-\rho)/2 < \varpi < 1/2$. Then,*

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

$$\Sigma(v,w) \equiv \frac{\int_0^T v(c_s)^2 \frac{\varsigma_s}{c_{ZZ,s}} ds + \sum_{\tau \in \mathcal{T}} w(c_{\tau-}, c_\tau, \beta_0)^2 \left(\Delta Z_\tau\right)^2 \varsigma_\tau}{\left(S(v) + Q_{ZZ}(w)\right)^2};$$

(b) $\Delta_n^{-1/2}(\hat{\beta}^\star_n - \beta_0) \xrightarrow{\mathcal{L}\text{-}s} \mathcal{MN}\left(0, V^\star\right)$ where

$$V^\star \equiv \left(\int_0^T \frac{d[Z,Z]_s}{\varsigma_s}\right)^{-1};$$

moreover, $V^\star \leq \Sigma(v,w)$ and the equality is attained if and only if $v(c) = \varphi c_{ZZ}/\varsigma$ and $w(c_\tau) = \varphi/\varsigma_\tau$ for all $\tau \in \mathcal{T}$ and some constant $\varphi > 0$;

(c) $V^\star$ can be consistently estimated by $\hat{V}^\star_n \equiv (\hat{S}_n(v^\star) + Q_{n,ZZ}(w^\star))^{-1}$.

[6]We note that the $3/k_n$ factor in the definition of $\hat{\beta}^\star_n$ corresponds to the bias-correction term $\mathbb{B}g\left(\cdot\right)/k_n$, and is obtained from direct calculation.

COMMENTS. (i) Part (a) of Theorem 1 establishes the asymptotic mixed normality of $\hat{\beta}_n$. This result can be regarded as a generalization of both the limit theory for the diffusive beta estimator presented in Proposition 1 of Li et al. (2017a) and that for the jump beta estimator shown in Theorem 2 of Li et al. (2017b).

(ii) More importantly, part (b) shows that $\hat{\beta}_n^\star$ attains the smallest asymptotic variance among all weighted estimators. In particular, the asymptotic variance $V^\star$ is strictly smaller than both the semiparametric efficiency bound for estimating the constant jump beta in Li et al. (2017b) and the adaptive bound for estimating the constant diffusive beta in Li et al. (2017a). From here, we see clearly that exploiting both the diffusive and the jump dependence will provide an efficiency gain. In Section 3.2, below, we show that $V^\star$ is indeed the adaptive efficiency bound.

(iii) Part (c) shows that $\hat{V}_n^\star$ is a consistent estimator for $V^\star$, which can be used to make feasible inference.

(iv) In general, if the common beta assumption does not hold, then the estimator $\hat{\beta}_n^\star$ converges to a weighted mixture of the diffusive and jump betas; see Lemma 3 in the appendix for details. □

3.2 Efficiency bound and adaptive estimation

In this subsection, we show that the optimally weighted estimator $\hat{\beta}_n^\star$ is indeed adaptive in the presence of the nonparametric nuisance components $(b, c_{ZZ}, \varsigma, J_Z, \epsilon)$.[7] We proceed as follows. We first derive the efficiency bound for estimating β in a parametric submodel of (6) in which the only unknown parameter is β, whereas the nonparametric nuisance components $(b, c_{ZZ}, \varsigma, J_Z, \epsilon)$ are observed, which can be equivalently thought as augmenting the original data with the observation of these nuisance processes. We prove the LAMN property for this submodel and establish the efficiency bound by invoking the conditional convolution theorem (Jeganathan (1982, 1983)). The resulting efficiency bound coincides with the $\mathcal{F}$-conditional asymptotic variance of $\hat{\beta}_n^\star$. Since the estimator $\hat{\beta}_n^\star$ only depends on the original data (instead of the augmented data), we conclude that it attains the adaptive efficiency bound *a fortiori*.

It is instructive to recall the LAMN property. Compared with the commonly used local asymptotic normality (LAN) property, the LAMN property is more general because it allows the information matrix to be random. In the sequel, we use P_β^n to denote the joint distribution of the data sequence $(\Delta_i^n X)_{1\leq i\leq[T/\Delta_n]}$ in a parametric model with an unknown

[7]We recall that an estimator is adaptive in the presence of a nuisance component if the estimator attains the efficiency bound in any submodel in which the nuisance is known. See Section 2.4 of Bickel et al. (1998) for additional background on adaptive estimation and, in particular, Definition 1 there.

parameter β. We say that the sequence (P_β^n) satisfies the LAMN property at $\beta = \beta_0$ if there exist a sequence Γ_n of nonnegative random variables and a sequence ψ_n of random variables, such that for any $h \in \mathbb{R}$,

$$\log \frac{dP^n_{\beta_0 + \Delta_n^{1/2} h}}{dP^n_{\beta_0}} = h\Gamma_n^{1/2}\psi_n - \frac{1}{2}\Gamma_n h^2 + o_p(1),$$

and

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

where the information Γ is a positive $\mathcal{F}$-measurable random variable and ψ is a standard normal variable independent of Γ.

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

Assumption 6 *We have Assumption 1 and the processes $(b_t)_{t\geq 0}$, $(\sigma_t)_{t\geq 0}$ and $(J_t)_{t\geq 0}$ are independent of $(W_t)_{t\geq 0}$, and the joint law of $(b, c_{ZZ}, \varsigma, J_Z, \epsilon)$ does not depend on β.*

Note that Assumption 6 is *only* needed in this subsection for getting a closed-form expression for the likelihood ratio in the derivation of the efficiency bound for estimating β. Our estimation and inference methods work under far more general settings. Theorem 2, below, shows that the aforementioned submodel satisfies the LAMN property and characterizes the information for the estimation of β.

Theorem 2 *Under Assumptions 5 and 6, the sequence $(P_\beta^n : \beta \in \mathbb{R})$ satisfies the LAMN property at $\beta = \beta_0$ with information $\int_0^T (1/\varsigma_s)\, d[Z,Z]_s$, where β_0 is the true value of β.*

COMMENTS. (i) Theorem 2 reveals that the "local" information bound for estimating beta is $(1/\varsigma_s)\, d[Z,Z]_s$, which is exactly the local signal-to-noise ratio between the instantaneous local quadratic variation of Z and the idiosyncratic variance ς.

(ii) The information in Theorem 2 can be decomposed as

$$\begin{aligned}\int_0^T \frac{d[Z,Z]_s}{\varsigma_s} &= \int_0^T \frac{d\langle Z^c, Z^c\rangle_s}{\varsigma_s} + \int_0^T \frac{d\left[Z^d, Z^d\right]_s}{\varsigma_s} \\ &= \int_0^T \frac{c_{ZZ,s}}{\varsigma_s} ds + \sum_{\tau \in \mathcal{T}} \frac{(\Delta Z_\tau)^2}{\varsigma_\tau}.\end{aligned}$$

The two terms on the right-hand side of the above display are exactly the information bounds for separately estimating the diffusive beta and the jump beta; see Theorem 1 in Li et al. (2017a) and Theorem 2 in Li et al. (2017b).[8] □

[8]Note that if the common beta assumption does not hold, the idiosyncratic variance ς appearing in the information bound of the continuous beta and that of the jump beta would be defined differently using the corresponding betas.

From Theorem 2 and the conditional convolution theorem (see Jeganathan (1982, 1983)), we deduce that the efficiency bound for estimating β_0 in the adaptive case is given by the inverse of the information, that is,

$$\left(\int_0^T \frac{d[Z,Z]_s}{\varsigma_s}\right)^{-1}.$$

This efficiency bound coincides exactly with the $\mathcal{F}$-conditional asymptotic variance of $\hat{\beta}_n^\star$, and hence, $\hat{\beta}_n^\star$ is an (efficient) adaptive estimator as claimed above.

Finally, we provide some intuition on why the common beta can be estimated adaptively. From Li et al. (2017a) and Li et al. (2017b), we know that the diffusive beta admits adaptive estimation, while the jump beta is generally not adaptive with respect to the presence of unknown jump size of Z. To see how the common beta restriction resolves the latter issue, we consider a stripped-down scenario in which Z has only one jump at time $\tau \in ((i-1)\,\Delta_n, i\Delta_n]$ for some integer i. In this case the jump beta estimator is simply $\Delta_i^n Y/\Delta_i^n Z$, which has the following expansion

$$\begin{aligned} \Delta_n^{-1/2}\left(\frac{\Delta_i^n Y}{\Delta_i^n Z} - \beta^d\right) &\approx \frac{\Delta_n^{-1/2}\left(\Delta_i^n Y - \beta^d \Delta_i^n Z\right)}{\Delta Z_\tau} \\ &= \frac{\Delta_n^{-1/2}\left(\Delta_i^n Y^c - \beta^d \Delta_i^n Z^c\right)}{\Delta Z_\tau} \end{aligned}$$

where we recall that Y^c and Z^c are the diffusive parts of Y and Z. Meanwhile, the jump size ΔZ_τ is estimated by $\Delta_i^n Z$, for which we have

$$\Delta_n^{-1/2}\left(\Delta_i^n Z - \Delta Z_\tau\right) \approx \Delta_n^{-1/2}\Delta_i^n Z^c.$$

Whether the estimation of β^d is adaptive to the presence of the unknown ΔZ_τ depends on whether their "residuals" are orthogonal. This does not hold in general. But, if $\beta^d = \beta^c$, then $\Delta_i^n Y^c - \beta^d \Delta_i^n Z^c = \Delta_i^n Y^c - \beta^c \Delta_i^n Z^c$ is orthogonal to $\Delta_i^n Z^c$ due to the diffusive orthogonality condition. In this case, the estimation of Z's jumps does not have an asymptotic effect on the estimation of the jump beta. From Li et al. (2017a), we also know that it does not affect the estimation of the diffusive beta, either. We thus have the adaptive estimation result for the common beta.

3.3 A Hausman test for common beta

The theory above says that the $\hat{\beta}_n^\star$ estimator is efficient in two senses: Theorem 1 shows that it is the most efficient estimator among all weighted estimators under general conditions, and Theorem 2 shows that it is semiparametrically efficient provided that the submodel satisfies the LAMN property. As recognized by Hausman (1978), a useful by-product of efficient estimation is to test the underlying model restriction by examining whether the efficient estimator and an inefficient estimator are statistically different. The Hausman test has a very convenient feature: as a consequence of efficiency, the asymptotic variance of the difference between the two estimators is the difference of their asymptotic variances. In the present context, we can test the common beta hypothesis by comparing the efficient estimator and other inefficient ones.

We now illustrate how to implement the Hausman test in an example. For concreteness, we take the inefficient estimator as:

$$\hat{\beta}_n^{BNS} = \frac{\Delta_n \sum_{i=0}^{[T/\Delta_n]-k_n} \hat{c}_{ZY,n,i} + \sum_{i\in\mathcal{I}'_n} \Delta_i^n Z \Delta_i^n Y}{\Delta_n \sum_{i=0}^{[T/\Delta_n]-k_n} \hat{c}_{ZZ,n,i} + \sum_{i\in\mathcal{I}'_n} (\Delta_i^n Z)^2}. \tag{19}$$

This estimator is an interesting benchmark because it is essentially the same as the realized regression estimator proposed by Barndorff-Nielsen and Shephard (2004a).[9] Like the efficient estimator, this estimator relies on both diffusive and jump returns and is a special case of the weighted estimator $\hat{\beta}_n(\cdot,\cdot)$ with weight functions $v^{BNS}(c) = c_{ZZ}$ and $w^{BNS}(\cdot) = 1$.

Under the common beta null hypothesis, the diffusive and jump betas equal to the same constant β_0. Consequently, both $\hat{\beta}_n^{BNS}$ and $\hat{\beta}_n^{\star}$ are consistent estimators of β_0 and the latter is efficient. To precisely state the alternative hypothesis for the Hausman test, we consider a general data generating process with time-varying diffusive and jump betas (denoted by β_t^c and β_t^d respectively) with the form:

$$Y_t = B_t + \beta_t^c Z_t^c + \beta_t^d Z_t^d + \epsilon_t^c + \epsilon_t^d \tag{20}$$

such that the orthogonality condition (5) holds, which amounts to setting $\beta_t^c = c_{YZ,t}/c_{ZZ,t}$ for all t and $\beta_\tau^d = \Delta Y_\tau / \Delta Z_\tau$ for each jump time τ of Z. In this general case, $\hat{\beta}_n^{BNS}$ and $\hat{\beta}_n^{\star}$

[9] The estimator of Barndorff-Nielsen and Shephard (2004a) is defined as the ratio between the realized versions of $[Y,Z]_T$ and $[Z,Z]_T$, where we remind the reader that $[\cdot,\cdot]$ denotes the quadratic covariation operator. In the general setting with time-varying beta, $\hat{\beta}_n^{BNS}$ converges to exactly the same limit $[Y,Z]_T/[Z,Z]_T$ in probability. We note that Barndorff-Nielsen and Shephard (2004a) originally consider a setting without jumps, and in that case, the difference between their estimator and $\hat{\beta}_n^{BNS}$ defined in (19) arises only from boundary effects, which are asymptotically negligible also on the second order. In the more general case with jumps, the asymptotic distribution of Barndorff-Nielsen and Shephard's original estimator can be derived using Theorem 5.4.2 in Jacod and Protter (2012), which provides the stable convergence in law for the quadratic covariation matrix. In that case, $\hat{\beta}_n^{BNS}$ still has the same limiting distribution as Barndorff-Nielsen and Shephard's original estimator when Y does not contain idiosyncratic jumps. When Y contains idiosyncratic jumps, Barndorff-Nielsen and Shephard's original estimator is less efficient than $\hat{\beta}_n^{BNS}$, because the former involves an additional source of sampling variability (which is $\mathcal{F}$-conditionally independent of the other limiting components) stemming from the interaction between the idiosyncratic jumps of Y and the contemporaneous Brownian shocks in Z.

converge to different integrated functionals of the β_t^c and β_t^d processes. Specifically, Lemma 3 in the appendix shows that

$$\begin{aligned}
\hat{\beta}_n^{BNS} &\xrightarrow{\mathbb{P}} \bar{\beta}^{BNS} \equiv \left(\int_0^T d\langle Z^c, Z^c\rangle_s \beta_s^c + \sum_{\tau\in\mathcal{T}} (\Delta Z_\tau)^2 \beta_\tau^d\right) \Big/ \int_0^T d\,[Z,Z]_s\,, \\
\hat{\beta}_n^{\star} &\xrightarrow{\mathbb{P}} \bar{\beta}^{\star} \equiv \left(\int_0^T \frac{d\langle Z^c, Z^c\rangle_s \beta_s^c}{\varsigma_s} + \sum_{\tau\in\mathcal{T}} \frac{(\Delta Z_\tau)^2 \beta_\tau^d}{\varsigma_\tau}\right) \Big/ \int_0^T \frac{d\,[Z,Z]_s}{\varsigma_s},
\end{aligned} \tag{21}$$

where we intentionally represent the limits as integrated weighted beta functionals for the ease of comparison.[10] From here, it is easy to see that the Hausman test will have power against alternatives under which $\bar{\beta}^{BNS} \neq \bar{\beta}^{\star}$. We can thus formally represent the null and alternative hypotheses for the Hausman test respectively using the following events:

$$\Omega_0 \equiv \{\beta_t^c = \beta_t^d = \beta_0 \text{ for some } \beta_0 \in \mathbb{R} \text{ and all } t \in [0,T]\}, \quad \Omega_a \equiv \{\bar{\beta}^{BNS} \neq \bar{\beta}^{\star}\}.$$

We carry out the test by examining whether the difference $\hat{\beta}_n^{\star} - \hat{\beta}_n^{BNS}$ is statistically different from zero. Due to the efficiency of $\hat{\beta}_n^{\star}$, the asymptotic variance of $\Delta_n^{-1/2}(\hat{\beta}_n^{\star} - \hat{\beta}_n^{BNS})$ is simply $\Xi \equiv V^{BNS} - V^{\star}$ where

$$V^{BNS} \equiv \frac{\int_0^T c_{ZZ,s}\varsigma_s ds + \sum_{\tau\in\mathcal{T}} (\Delta Z_\tau)^2 \varsigma_\tau}{\left(\int_0^T c_{ZZ,s} ds + \sum_{\tau\in\mathcal{T}} (\Delta Z_\tau)^2\right)^2}. \tag{22}$$

We can consistently estimate Ξ using a sample analogue estimator $\hat{\Xi}_n = \hat{V}_n^{BNS} - \hat{V}_n^{\star}$, where

$$\hat{V}_n^{BNS} \equiv \frac{\hat{S}_n(g_\Xi) + \sum_{i\in\mathcal{I}_n'} (\Delta_i^n Z)^2 \hat{\varsigma}_{n,i}}{\left(\hat{S}_n(v^{BNS}) + Q_{n,ZZ}(w^{BNS})\right)^2}, \text{ for } g_\Xi(c) = c_{ZZ}\left(c_{YY} - c_{ZY}^2/c_{ZZ}\right).$$

The t-statistic for the Hausman test is then defined as

$$\hat{T}_n \equiv \frac{\Delta_n^{-1/2}\left(\hat{\beta}_n^{\star} - \hat{\beta}_n^{BNS}\right)}{\sqrt{\hat{\Xi}_n}}.$$

Proposition 1, below, describes the asymptotic properties of the Hausman test, where we use z_q to denote the q-quantile of the standard normal distribution.

Proposition 1 *Suppose that the conditions in Theorem 1 hold and $\Xi > 0$ almost surely. Then, (a) in restriction to Ω_0, $\hat{T}_n$ converges stably in law to a standard normal distribution; (b) at significance level $\alpha \in (0,1)$, the test with critical region $\{|\hat{T}_n| > z_{1-\alpha/2}\}$ has asymptotic level α under the null hypothesis and asymptotic power one under the alternative, that is, $\mathbb{P}(C_n|\Omega_0) \to \alpha$ and $\mathbb{P}(C_n|\Omega_a) \to 1$.*

[10] The limit of $\hat{\beta}_n^{BNS}$ can be written more concisely as $\bar{\beta}^{BNS} = [Y,Z]_T \,/\, [Z,Z]_T$.

In applications, we recommend using this Hausman test as a post-estimation diagnostic tool to evaluate the plausibility of the common beta assumption. We stress that the test should not be used as a pre-test for choosing between $\hat{\beta}_n^{\star}$ and $\hat{\beta}_n^{BNS}$ (or any other beta estimators). In fact, even if one is willing to do so, it is not clear what would be the choice when the test rejects, because in that scenario, the two estimators simply converge to two different weighted integrated beta functionals and are generally not comparable.[11]

3.4 Practical considerations

We finish with some remarks regarding the use of the theoretical results derived above. The theoretical contribution of the paper is to provide the most efficient estimator for empirical researchers who are willing to make the assumption of common beta (as is often implicitly done in empirical work). In other words, we characterize a new point on the frontier of the "assumption versus efficiency" trade-off. This trade-off is the most relevant for beta estimation within short estimation periods, for which the common beta hypothesis is more plausible. Our numerical work below is designed to illustrate this point more concretely.

Our practical recommendation is as follows. If the empirical researcher is willing to assume that beta is constant for all returns (i.e., common beta) within a certain period (say, e.g., a week), then we unambiguously recommend the proposed estimator on the ground of statistical efficiency. On the other hand, if the empiricist is concerned with the issue of heterogeneous beta within the time window, we still recommend reporting the proposed estimator in the analysis, accompanied with a careful interpretation. In this situation, different beta estimators converge to different weighted integrated functionals of the time-varying beta process, and it is generally difficult to argue why one weighted functional should be more relevant than the other. The potential advantage of using the proposed estimator is that it is likely to be statistically more accurate (i.e., with smaller standard error) if the common beta hypothesis "almost" holds in the data. A "free" diagnostic tool after the efficient estimation is the Hausman test, which can also be used to assess the extent to which heterogeneous

[11]By analogy, we note the Hausman test in the current context should be used in the same way as the over-identification test in the classical GMM. The GMM problem share three fundamental similarities with what we study here. First, if the model is correctly specified, any weighting matrix produces a consistent estimator, but the optimally weighted GMM is efficient. Second, if the moment equality model is misspecified, then GMM estimators with different weighting matrices will converge to different pseudo-true parameters, and are generally not comparable. Third, the over-identification test is a by-product of the efficient GMM estimation; it is used to examine whether the moment restrictions are correctly specified, but not for deciding whether the optimal weighting matrix should be used in the first place.

beta is a concern in the empiricist's specific application.

4 Monte Carlo study

We continue with evaluating our inference procedures developed above on simulated data.

4.1 The data generating process

The unit of time is one business day, consisting of 6.5 trading hours. We simulate the factor process Z according to

$$dZ_t = \sigma_t dL_t,$$

where L is a Lévy process with characteristic triplet $(0, 1, \nu_1)$ and $\nu_1(dx) = (1/16)e^{-|x|}dx$. Following Bollerslev and Todorov (2011), we simulate the stochastic volatility σ_t process from a two-factor affine diffusion model, that is, $\sigma_t^2 = V_{1,t} + V_{2,t}$, where

$$\begin{aligned} dV_{1,t} &= 0.0116\,(0.5 - V_{1,t})\,dt + 0.1023\sqrt{V_{1,t}}dW_{1,t}, \\ dV_{2,t} &= 0.6930\,(0.5 - V_{2,t})\,dt + 0.7909\sqrt{V_{2,t}}dW_{2,t}, \end{aligned}$$

and (W_1, W_2) is a two-dimensional Brownian motion that is independent of L. The parameters chosen in this model imply that V_1 is a highly persistent process with half-life of 60 days, and V_2 is a fast mean-reverting process with half-life of 1 day. The residual process is simulated using

$$d\epsilon_t = \sigma_t d\widetilde{L}_t,$$

where $\widetilde{L}$ is another Lévy process, independent of (L, W_1, W_2), with characteristic triplet $(0, 1/\sqrt{2}, \nu_2)$ and $\nu_2(dx) = (1/90)e^{-|x|}dx$. Finally, we simulate Y as

$$dY_t = dZ_t + d\epsilon_t. \tag{23}$$

under which the common beta restriction holds with $\beta_0 = 1$.

In this specification, the continuous parts in the driving Lévy processes L and $\widetilde{L}$ are Brownian diffusions with constant volatility and the jump parts are compound Poisson processes with jump size following double-exponential distributions. In particular, the Z process has jump activity of one jump every 8 days. The frequency and size of jumps are calibrated to match roughly those observed in real data. All subsequent calculations are based on 1,000 Monte Carlo trials.

4.2 Efficiency comparison

We first demonstrate numerically the efficiency gain of the efficient $\hat{\beta}_n^\star$ estimator relative to other estimators in our Monte Carlo experiment. In this exercise, we aim to illustrate when the efficiency gain is large/small, and provide some intuitive explanation for it based on our theory. Recall that $V^\star$ denotes the conditional asymptotic variance of the efficient estimator $\hat{\beta}_n^\star$.

We compare the performance of $\hat{\beta}_n^\star$ with that of the following four benchmarks: the adaptive diffusive beta estimator $\hat{\beta}_n^c$ of Li et al. (2017a), the optimal jump beta estimator $\hat{\beta}_n^J$ of Li et al. (2017b), the realized regression estimator $\hat{\beta}_n^{BNS}$ and its truncated version $\hat{\beta}_n^{BNS,c}$. We note that the truncated BNS estimator $\hat{\beta}_n^{BNS,c}$ only relies on diffusive returns, and it is included here for completeness because Barndorff-Nielsen and Shephard (2004a) originally consider a setting without jumps (but the estimator includes all returns). Recall that $\hat{\beta}_n^{BNS}$ is defined by (19) and the other three estimators are defined as

$$\hat{\beta}_n^c \equiv \frac{\sum_i \hat{c}_{ZY,n,i}/\hat{\varsigma}_{n,i}}{\sum_i \hat{c}_{ZZ,n,i}/\hat{\varsigma}_{n,i}}, \quad \hat{\beta}_n^J \equiv \frac{\sum_{i\in\mathcal{I}'_n} \Delta_i^n Z \Delta_i^n Y/\hat{\varsigma}_{n,i}}{\sum_{i\in\mathcal{I}'_n} (\Delta_i^n Z)^2/\hat{\varsigma}_{n,i}}, \quad \hat{\beta}_n^{BNS,c} \equiv \frac{\sum_i \hat{c}_{ZY,n,i}}{\sum_i \hat{c}_{ZZ,n,i}}.$$

Since these estimators are all special cases of the weighted estimator $\hat{\beta}_n(v,w)$, we can compute their $\mathcal{F}$-conditional asymptotic variances using part (a) of Theorem 1. We denote these quantities as V^c, V^J, V^{BNS} and $V^{BNS,c}$, respectively. We can measure the relative efficiency of the estimators using the ratio of their asymptotic standard deviations with respect to that of the efficient estimator $\hat{\beta}_n^\star$. For example, a larger $\sqrt{V^{BNS}}/\sqrt{V^\star}$ ratio indicates a higher efficiency gain for the efficient estimator $\hat{\beta}_n^\star$ relative to $\hat{\beta}_n^{BNS}$. We remind the reader that the conditional asymptotic variances are random in the present non-ergodic setting. As a result, the values of the corresponding relative efficiency measures vary across Monte Carlo trials, depending on the realized paths of the stochastic volatility and the jump processes.

Before diving into the numerical results, we briefly summarize predictions from the asymptotic theory to guide intuition. From our theoretical results, the relative efficiency measures should be greater than 1 on all sample paths. However, the magnitude of the efficiency gain varies among different estimators and scenarios, as explained below:

- *Case* $\hat{\beta}_n^c$ *versus* $\hat{\beta}_n^\star$: Compared with $\hat{\beta}_n^c$, $\hat{\beta}_n^\star$ gains efficiency by exploiting the additional information from jump returns. Other things being equal, we expect to see higher efficiency gain (i.e., larger value of$\sqrt{V^c}/\sqrt{V^\star}$) when there are more jumps per unit of time (i.e., higher jump intensity). In the boundary case with no jumps, these estimators are asymptotically identical.

- *Case* $\hat{\beta}_n^J$ *versus* $\hat{\beta}_n^\star$: The $\hat{\beta}_n^\star$ estimator gains efficiency over the jump beta estimator $\hat{\beta}_n^J$ by incorporating diffusive returns. When jumps arrive more frequently, the *relative* contribution of using the diffusive returns is smaller, yielding less efficiency gain. In the calculation of efficiency gains, the effect of jumps on $\hat{\beta}_n^J$ is thus opposite to their effect on $\hat{\beta}_n^c$.

- *Case* $\hat{\beta}_n^{BNS}$ *versus* $\hat{\beta}_n^\star$: The comparison between $\hat{\beta}_n^{BNS}$ and $\hat{\beta}_n^\star$ is slightly more subtle, because both estimators make use of all jump and diffusive returns. In this case, the efficiency gain stems from the fact that $\hat{\beta}_n^\star$ employs an optimal weighting scheme that is inversely proportional to the stochastic spot idiosyncratic variance ς_t, to which $\hat{\beta}_n^{BNS}$ is insensitive. Hence, using $\hat{\beta}_n^\star$ will result in a larger efficiency gain if the idiosyncratic variance process ς_t exhibits more temporal heterogeneity in its sample path. In our simulation design, this effect is captured by both volatility factors, with the degree of variation in volatility depending on the distance of the current values of the volatility factors from their unconditional means, the speeds of mean reversion as well as the volatility of volatility parameters. In particular, at the longer horizon of $T = 10$ days, the idiosyncratic variance process tends to vary more, and hence, leads to more efficiency gain from using efficient weighting.

- *Case* $\hat{\beta}_n^{BNS,c}$ *versus* $\hat{\beta}_n^\star$: The $\hat{\beta}_n^{BNS,c}$ estimator only uses diffusive returns (like $\hat{\beta}_n^c$) and employs sub-optimal weighting (like $\hat{\beta}_n^{BNS}$). This leads to two sources of efficiency gains as explained in the two cases related to $\hat{\beta}_n^c$ and $\hat{\beta}_n^{BNS}$, respectively.

We now proceed to the numerical results. As is evident from the discussion above, when no jump occurs, the jump beta cannot be identified and the $\hat{\beta}_n^c$ estimator would be identical to $\hat{\beta}_n^\star$. To avoid this uninteresting degenerate situation, we report the relative efficiency for paths on which Z has exactly one jump over the $[0, T]$ horizon. In doing so, we can directly examine the effect of *realized* jump intensity by varying T from 2 to 10 days (with the $T = 2$ case corresponding to the high-intensity scenario). As explained above, having more jumps per unit of time will lead to more (resp. less) efficiency gain in $\hat{\beta}_n^\star$ relative to the $\hat{\beta}_n^c$ (resp. $\hat{\beta}_n^J$) estimator, which would have the effect as shortening the sample span.

Table 1 reports the empirical quantiles of the (random) relative efficiency measure. Overall, the observed pattern is exactly as discussed above. The relative efficiency measures are all greater than unity, which confirms the efficiency gain offered by the efficient estimator. More specifically, we observe from the first column that the efficiency gain relative to the diffusive beta estimator $\hat{\beta}_n^c$ has a median of 12%, and it reaches a nontrivial 43% at the 75% quantile when the horizon is $T = 2$. This shows that in a short estimation window, optimally including even one jump return (if it occurs) in the beta estimation can improve the estimation precision by a notable amount. As T increases, the realized jump intensity drops (as we restrict the calculation to paths with exactly one jump in the estimation window), and the efficiency gain becomes smaller as expected.

Looking at the second column of Table 1, we see substantial efficiency gain relative to the jump beta estimator $\hat{\beta}_n^J$. This is not surprising because the jump information is very scarce in this numerical experiment. At longer horizons, the efficient gain becomes even higher; this pattern is opposite to what we see in the first column, which is consistent with the intuition discussed above.

Finally, we turn attention to the two versions of the BNS estimators. The $\hat{\beta}_n^{BNS}$ estimator, which pools information from both diffusive and jump returns, has impressive performance, albeit it is still less efficient than $\hat{\beta}_n^{\star}$. In terms of median, the efficiency gain grows from 2% in the 2-day case to 6% in the 10-day case.[12] While the efficiency gain apparently grows with the sample horizon, we remind the reader that the driving factor here is not the length of the time span of the sample per se. Instead, the efficiency gain stems from the temporal heterogeneity in the stochastic idiosyncratic variance process, which tends to increase with T in this numerical experiment. We finally note that the truncated version $\hat{\beta}_n^{BNS,c}$ is much less efficient.

All in all, the numerical comparison in Table 1 provides a more concrete illustration of our efficiency theory. Obviously, this specific numerical experiment cannot possibly exhaust all the scenarios that one may encounter in empirical work. That being said, the aforementioned intuition—backed by our asymptotic theory—should be useful more broadly for understanding efficiency issues in high-frequency beta estimation.

4.3 Finite-sample performance of inference methods

In this subsection, we examine the finite-sample performance of the proposed inference methods. This analysis includes two parts: the first concerns the asymptotic mixed Gaussian approximation for the efficient estimator $\hat{\beta}_n^{\star}$ (see Theorem 1), and the second is for the size and power properties of the Hausman test discussed in Section 3.3. We fix the sample span at $T = 5$ days.[13]

We consider two sampling schemes with $\Delta_n = 1$ minute and 5 minutes, respectively. In addition, we employ a range of local window sizes (i.e., k_n) to examine the robustness

[12] In our empirical example below, the efficiency gain from using the optimal estimator instead of $\hat{\beta}_n^{BNS}$ is much larger. This suggests more high-frequency idiosyncratic volatility moves than those implied by our model in the Monte Carlo.

[13] We have also conducted the same experiments for $T = 2$ and 10 days, yielding very similar results. For brevity, we collect these additional simulation results in the supplemental appendix.

Table 1: Relative Efficiency of Alternative Beta Estimators in Monte Carlo.

Horizon	Quantile	$\sqrt{V^c}/\sqrt{V^\star}$	$\sqrt{V^J}/\sqrt{V^\star}$	$\sqrt{V^{BNS}}/\sqrt{V^\star}$	$\sqrt{V^{BNS,c}}/\sqrt{V^\star}$
	25%	1.02	1.40	1.01	1.06
2 days	50%	1.12	2.21	1.02	1.25
	75%	1.43	4.70	1.03	1.47
	25%	1.01	1.91	1.03	1.06
5 days	50%	1.06	3.26	1.04	1.12
	75%	1.17	7.63	1.06	1.23
	25%	1.01	2.42	1.04	1.07
10 days	50%	1.03	4.47	1.06	1.11
	75%	1.10	9.80	1.09	1.19

Note: The table reports the Monte Carlo quantiles of the relative efficiency measures of $\hat{\beta}_n^c$, $\hat{\beta}_n^J$, $\hat{\beta}_n^{BNS}$, and $\hat{\beta}_n^{BNS,c}$, respectively in four columns. For each estimator, the relative efficiency measure is defined as the ratio of its conditional asymptotic standard deviation with respect to that of the efficient estimator $\hat{\beta}_n^\star$. The relative efficiency measure is random, and we summarize its distributional features using quantiles at 25%, 50%, and 75% levels.

of the inference method with respect to this tuning parameter. Specifically, we consider $k_n \in \{70, 85, 100\}$ for 1-minute sampling and $k_n \in \{20, 24, 28\}$ for 5-minute sampling. Note that in each case, the largest value of k_n is about 40% larger than the smallest, providing a reasonably wide range for checking robustness.

We also need to set the truncation threshold u_n that satisfies $u_n \asymp \Delta_n^\varpi$ (recall Theorem 1). While the formal asymptotic theory is valid with $u_n = C\Delta_n^\varpi$ for any constant $C > 0$, we set u_n more specifically at

$$u_n = 3.5 \times \sqrt{\gamma_i BV_t} \times \Delta_n^{0.49}, \tag{24}$$

which roughly corresponds to a 3.5-standard-deviation rule with the local standard deviation being approximated by $\sqrt{\gamma_i BV_t}$. Here, BV_t is the bipower variation (see Barndorff-Nielsen and Shephard (2004b)) measuring the average volatility on day t and is given by

$$BV_t \equiv \frac{\pi}{2} \frac{n}{n-1} \sum_{i=(t-1)n+2}^{tn} |\Delta_{i-1}^n Z||\Delta_i^n Z|,$$

Table 2: Monte Carlo Coverage Rates of Efficient Confidence Intervals

	$\Delta_n = 1$ minute			$\Delta_n = 5$ minutes		
Level	$k_n = 70$	$k_n = 85$	$k_n = 100$	$k_n = 20$	$k_n = 24$	$k_n = 28$
90%	87.60	87.90	88.40	86.60	86.90	87.20
95%	93.20	93.10	93.00	92.40	93.20	93.10
99%	98.30	98.50	98.80	98.00	98.10	98.00

Note: The table reports the Monte Carlo coverage rates (%) of confidence intervals associated with the efficient adaptive estimator $\hat{\beta}_n^\star$. The confidence intervals are two-sided symmetric and are constructed based on Theorem 1 at nominal levels 90%, 95%, and 99%. Two sampling frequencies (i.e., Δ_n) and a range of local window sizes (i.e., k_n) are considered. The sample span is fixed at $T = 5$ days. There are 1,000 Monte Carlo trials.

and γ_i is a time-of-day adjustment for the intraday seasonality pattern of volatility given by $\gamma_i \equiv b_i/((1/n)\sum_{j=1}^n b_j)$, with $b_i \equiv T^{-1}\sum_{t=1}^T |\Delta^n_{(t-1)n+i-1}Z||\Delta^n_{(t-1)n+i}Z|$ for $i = 2, ..., n$ and $b_1 = b_2$, for observation i in day t. We stress that the specific choice of the truncation threshold u_n in (24) is motivated only by finite-sample considerations, but has no asymptotic effect on the beta estimators, provided that the scaling factor $\sqrt{\gamma_i BV_t}$ is of constant asymptotic order.[14] Additional robustness checks regarding different choices of truncation threshold are reported in the supplemental appendix.

Table 2 reports the convergence rates of two-sided symmetric confidence intervals (CI) constructed using the efficient $\hat{\beta}_n^\star$ estimator based on Theorem 1 at nominal levels 90%, 95%, and 99%. Overall, we see that the finite-sample converge rates are close to the corresponding nominal levels, albeit with slight under-coverage. The results are robust to perturbations in the local window size.

Next, we examine the size and power properties of the Hausman test described in Section 3.3. We recall that the null hypothesis of common beta is imposed by equation (23) with common beta $\beta_0 = 1$. To examine the power property of the test, we consider two

[14] This statement can be formalized as follows. Suppose that $\max_i \gamma_i BV_t$ is of sharp order $O_p(1)$. If γ_i is generated as $f(i\Delta_n)$ for a continuous positive function $f(\cdot)$ on $[0, 1]$, then this condition is satisfied because $f(\cdot)$ is uniformly bounded on the unit interval and the bipower variation BV_t is of sharp $O_p(1)$ order. Recall from Theorem 1 that $\varpi \in ((1-\rho)/2, 1/2)$. Hence, there exists ϖ_- and ϖ_+ in $((1-\rho)/2, 1/2)$ such that $\varpi_- < \varpi < \varpi_+$. Then with probability approaching 1, u_n defined in (24) falls in between $\Delta_n^{\varpi_-}$ and $\Delta_n^{\varpi_+}$, which is sufficient for the truncation technique to work in the proofs. Also see Li et al. (2013) for a similar analysis using data-driven truncation.

alternatives. The first is given by

$$dY_t = dZ_t^c + 0.75 dJ_{Z,t} + d\epsilon_t.$$

This alternative violates the common beta restriction by imposing distinct diffusive and jump betas that take values 1 and 0.75, respectively. The second alternative features time-varying diffusive beta and is given by:

$$dY_t = \beta_t^c dZ_t^c + dJ_{Z,t} + d\epsilon_t, \quad \text{where} \quad \beta_t^c = 1 + 0.3 \times \sin(0.6t).$$

Under this alternative, the diffusive beta β_t^c is time-varying whereas the jump beta is fixed at 1. In both cases, $\bar{\beta}^\star \neq \bar{\beta}^{BNS}$ holds almost surely, and we expect the Hausman test to have asymptotic power one.

Table 3 reports the rejection rate of Hausman test under the common beta null hypothesis and the aforementioned two alternatives in panels A, B, and C, respectively. From panel A, we see that the test's null rejection rates are close to the corresponding nominal significance levels. The size control for the coarser 5-minute sampling is slightly worse than the 1-minute case. Panels B and C of the table show that the test has reasonable power against these alternatives. As expected, the power is higher in the larger sample with $\Delta_n = 1$ minute. Overall, these results are supportive of our asymptotic theory regarding the Hausman test and the efficient estimator.

5 Empirical illustration

We now illustrate the performance of the proposed efficient estimator $\hat{\beta}_n^\star$ in an empirical example. We consider beta estimation with relatively short horizons, that is, $T \in \{2, 5, 10\}$ days, as studied in the numerical experiments. This setting is ideal for illustrating the usefulness of the efficient estimator for two reasons. The first one is theoretical: the common beta assumption is more plausible within shorter samples, so that we are close to the theoretical "laboratory" setting, providing a more direct link between the empirical findings and theoretical insights. The other reason is practical: short estimation horizon represents a data-scarce environment, and this is exactly the situation in which an empirical researcher values statistical efficiency the most. Short estimation windows are often seen in event studies.

In this exercise, we estimate market betas of three large-cap stocks: JPMorgan (JPM), Walmart (WMT), and Johnson & Johnson (JNJ). Our proxy for the market portfolio is

Table 3: Finite-sample Size and Power Properties of Hausman Test

	$\Delta_n = 1$ min			$\Delta_n = 5$ min		
Level	$k_n = 70$	$k_n = 85$	$k_n = 100$	$k_n = 20$	$k_n = 24$	$k_n = 28$
Panel A: Null Hypothesis						
1%	1.30	1.20	1.10	1.40	1.30	1.40
5%	4.80	5.10	5.30	5.60	4.60	4.90
10%	10.00	10.50	10.50	11.00	9.70	9.10
Panel B: Alternative with Unequal Diffusive and Jump Betas						
1%	56.00	55.10	55.60	22.90	20.90	20.60
5%	69.60	64.60	67.40	36.40	36.50	33.00
10%	73.50	73.40	73.00	43.90	44.50	42.20
Panel C: Alternative with Time-varying Beta						
1%	50.60	52.40	52.70	23.90	24.70	25.50
5%	62.00	62.80	62.50	38.30	39.20	39.70
10%	66.90	67.50	67.80	47.00	47.10	48.90

Note: The table reports the Monte Carlo rejection rates of the Hausman test described in Section 3.3. Panel A reports the size property, and Panels B and C report powers under two alternative hypotheses. Two sampling frequencies (i.e., Δ_n) and a range of local window sizes (i.e., k_n) are considered. The sample span is fixed at $T = 5$. There are 1,000 Monte Carlo trials.

Table 4: Relative Efficiency Comparison

	JPM			JNJ			WMT		
Horizon	$Q_{0.25}$	$Q_{0.50}$	$Q_{0.75}$	$Q_{0.25}$	$Q_{0.50}$	$Q_{0.75}$	$Q_{0.25}$	$Q_{0.50}$	$Q_{0.75}$
2 days	1.11	1.16	1.24	1.10	1.15	1.23	1.11	1.16	1.24
5 days	1.16	1.21	1.26	1.13	1.19	1.24	1.15	1.20	1.26
10 days	1.19	1.23	1.30	1.16	1.20	1.25	1.18	1.21	1.29

Note: This table compares the relative efficiency between the efficient estimator $\hat{\beta}_n^\star$ and realized regression estimator $\hat{\beta}_n^{BNS}$. The relative efficiency measure is defined as the ratio of the estimated standard errors of these estimators, that is, $\sqrt{\hat{V}_n^{BNS}}/\sqrt{\hat{V}_n^\star}$. For each horizon $T \in \{2, 5, 10\}$, we divide the entire 8-year sample into non-overlapping blocks with length T, and compute the relative efficiency measure separately on each block. Their values are summarized via their empirical quantiles (denoted Q_q) computed across all estimations blocks at level $q = 0.25$, 0.5, and 0.75.

the S&P 500 ETF (SPY).[15] High-frequency transaction price data is obtained from the Trade and Quote (TAQ) database from 2007 to 2014. We sample the prices at the 5-minute frequency, which results in 78 price observations per asset on each day. Since the assets under consideration are quite liquid, the 5-minute sparse sampling is typically considered coarse enough to sufficiently mitigate complications related to various microstructure issues. As in the simulations, we set the local window to $k_n = 24$.[16] The truncation threshold is also calculated in the same way as in the simulation study.

In the present literature, arguably the most commonly used beta estimator in the high-frequency setting is the realized regression estimator of Barndorff-Nielsen and Shephard (2004a). It is thus of great empirical interest to examine whether—and to which extent—we can improve the estimation accuracy by using the efficient estimator relative to this influential benchmark.

Table 4 compares the relative efficiency of $\hat{\beta}_n^\star$ and $\hat{\beta}_n^{BNS}$ in our sample. We measure the relative efficiency using the ratio of the estimated standard errors of these estimators, that

[15] JPMorgan is major bank in the US with a relatively high beta. Walmart and Johnson & Johnson are representatives of the consumers and health care sectors, and they have much less sensitivity towards the market index.

[16] Additional robustness checks regarding the choices of the local window size k_n and the truncation threshold u_n are collected in the supplemental appendix and show similar results.

is,

$$\text{Relative Efficiency} = \sqrt{\hat{V}_n^{BNS}}/\sqrt{\hat{V}_n^{\star}}.$$

This is analogous to the asymptotic relative efficiency measure used in Table 1, but with the asymptotic variances replaced with their consistent estimators.[17] For each given horizon T, we divide the entire 8-year sample into non-overlapping estimation periods with length T. We compute the relative efficiency measure separately for each estimation period, and then summarize these numbers using empirical quantiles across all estimation periods in Table 4. Each stock is treated on its own. From the table, we see clearly that the efficient estimator provides sizable efficiency gain relative the benchmark $\hat{\beta}_n^{BNS}$. Specifically, the efficiency gain is roughly 16% for the short 2-day estimation window on the median. The gain becomes larger as T increases. Recall that the latter effect is also seen in the numerical example, for which the theoretical intuition is that the stochastic idiosyncratic variance process tends to vary more in longer samples.

The efficiency gain documented in Table 4 should be connected with our efficient estimation theory with care. Unlike the "laboratory" setting in the numerical example, where the common beta assumption is imposed by design, the econometrician cannot know with certainty whether this assumption holds or not in the data (any statistical test is still subject to type 1 and type 2 errors). The findings thus need to be interpreted differently in two scenarios. The first is when the common beta assumption holds. In this scenario, both $\hat{\beta}_n^{\star}$ and $\hat{\beta}_n^{BNS}$ (as well as other beta estimators) converge to the true beta and our efficient estimation theory provides a definitive explanation for why the former should always have a smaller asymptotic variance than the latter. In the alternative (more complicated) scenario with heterogeneous beta, these estimators instead converge to their associated integrated beta functionals (recall equation (21)). The asymptotic variance of $\hat{\beta}_n^{\star}$ is no longer guaranteed to be smaller than that of $\hat{\beta}_n^{BNS}$, and hence, our theory cannot fully explain why $\hat{V}_n^{BNS} > \hat{V}_n^{\star}$ as we see in the empirical table. The theory nevertheless provides a partial answer: to the extent that the common beta hypothesis "almost" captures the more complicated reality—particularly in short samples—our theory on efficiency can "approximately" account for the efficiency gain shown in Table 4.

The Hausman test described in Section 3.3 provides a more direct statistical answer on whether the common beta hypothesis holds. In Table 5, we report the proportion of T-day estimation periods for which the Hausman test rejects the common beta hypothesis at the 5% significance level. We clarify that these tests are implement separately, and we report the

[17]We remind the reader that a higher ratio indicates a larger efficiency gain from using the efficient estimator $\hat{\beta}_n^{\star}$.

Table 5: Rejection Rates of the Hausman test.

Horizon	JPM	JNJ	WMT
2 days	8.17%	6.35%	5.24%
5 days	13.89%	12.63%	12.12%
10 days	32.32%	18.18%	17.17%

Note: This table reports the rejection rates of the Hausman test across all estimation blocks. The test is implemented by comparing $\hat{\beta}_n^{\star}$ with $\hat{\beta}_n^{BNS}$.

rejection proportion as a summary statistic of the individual testing results, with no intention to make formal statement on the joint testing.[18] From the table, we see that the common beta hypothesis is rejected for slightly more than 5% of all 2-day estimation periods. Since the test is implemented at the 5% significance level, we expect to see 5% rejections even if the null hypothesis is true, so these empirical rejection rates should be considered to be low. Obviously, it would be naive to use this finding to argue that the common beta hypothesis actually "holds." After all, as in any scientific model, such an assumption is meant to be used as an approximation of the complicated reality. That said, our finding does suggests that such an approximation is empirically plausible, particularly at short horizons. As discussed in the previous paragraph, this is consistent with the large efficiency gain delivered by the efficient $\hat{\beta}_n^{\star}$ estimator reported in Table 4. Finally, we also note that the rejection rates increase, but remain moderate, when increasing the horizon from 2 to 5 days. When $T = 10$, we see that the common beta hypothesis is rejected considerably more often, especially for the financial firm (JPM). JPM's high rejection rate may be explained by the fact that the bank can change its exposure to various sectors/firms in the economy more rapidly by trading in the financial market. Since the betas of different sectors/firms are generally different, the bank's beta can vary a lot very quickly due to the changes in its asset holdings, even if the betas of underlying assets do not change individually.

In summary, this empirical example shows that the $\hat{\beta}_n^{\star}$ estimator has smaller standard errors than the benchmark estimator $\hat{\beta}_n^{BNS}$. Our asymptotic theory provides a definitive explanation for this finding if one is willing to assume common beta. Otherwise, the theory

[18] A formal joint test can be implemented as follows. Under the null hypothesis of common beta, the Hausman t-statistics for individual (non-overlapping) estimation blocks are asymptotically independent standard normal. One can construct a sup-t statistic as the maximum of the absolute value of individual t-statistics. The critical value is given by the quantile of $\max_j \xi_j$ for i.i.d. standard normal variables ξ_j, with the maximum taken over all blocks (indexed by j).

offers a partial explanation by treating the common beta hypothesis as an approximation of a more complicated reality. The Hausman test can be conveniently implemented as a by-product after the efficient estimation. Empirically, the test suggests that the common beta hypothesis is plausible for short samples.

6 Concluding remarks

In this paper, we study the common beta assumption—that is, the beta is constant for all returns (diffusive returns and jumps)—in the context of efficient beta estimation. We propose an optimally weighted estimator that fully exploits the information content of this model restriction. We show that the proposed estimator is the most efficient among a large class of weighted estimators under general conditions. Moreover, under additional regularities that ensures the LAMN property, we establish the semiparametric efficiency bound for estimating the common beta, and show that the proposed estimator is semiparametrically efficient. Interestingly, this efficient estimator is actually adaptive with respect to the presence of nonparametric nuisances such as the stochastic volatility of both the factor process and the idiosyncratic component, and the jumps in the factor process. To our knowledge, the efficiency bound, the adaptiveness result, and the efficient estimator are all new results to the literature. As a by-product of the efficient estimator, we propose a Hausman test for the common beta null hypothesis. We demonstrate the efficiency gain of the proposed estimator both in a numerical experiment and in an empirical example.

7 Appendix: Proofs

We first introduce some notation that will be used in the proofs. 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 $R_p \equiv \sqrt{\kappa_p}\sigma_{\tau_p-}\xi_{p-} + \sqrt{1-\kappa_p}\sigma_{\tau_p}\xi_{p+}$ and $\rho_p \equiv (-\beta_0, 1)\, R_p$. If Assumption 5 holds, ρ_p has an $\mathcal{F}$-conditionally centered mixed Gaussian distribution with variance ς_{τ_p}.

We also need a preliminary lemma, which is a straightforward extension of Theorem 4 in Li et al. (2017a), except that we consider estimators for integrated volatility functionals using overlapping windows.

Lemma 1 *Suppose (i) Assumptions 1 and 3 hold; (ii) there exist a sequence* $(T_m)_{m\geq 1}$ *of stopping times increasing to infinity, a sequence of convex compact subsets* $\mathcal{K}_m \subset \mathcal{M}_2$ *such* *that $c_t \in \mathcal{K}_m$ for $t \leq T_m$ and $g(\cdot)$ is a three-time continuously differentiable function on an ε-enlargement of $\mathcal{K}_m$ for some $\varepsilon > 0$; (iii)*

$$\frac{1}{3} < \rho < \frac{1}{2}, \ \frac{1-\rho}{2} \leq \varpi < \frac{1}{2}.$$

Then, $\Delta_n^{-1/2}(\hat{S}_n(g) - S(g)) \xrightarrow{\mathcal{L}\text{-}s} \mathcal{MN}(0, V(g))$, *where*

$$V(g) \equiv \sum_{j,k,l,m=1} \int_0^T \partial_{jk} g(c_s) \partial_{lm} g(c_s)^\top (c_{s,jl} c_{s,km} + c_{s,jm} c_{s,kl})\, ds.$$

Proof of Lemma 1. By Lemma 2 of Li and Xiu (2016), $\sup_{0 \leq i \leq [T/\Delta_n] - k_n} \|\hat{c}_{n,i} - \bar{c}_{n,i}\| = o_p(1)$, where $\bar{c}_{n,i} \equiv (k_n \Delta_n)^{-1} \int_{i\Delta_n}^{i\Delta_n + k_n \Delta_n} c_s ds$. This uniform approximation result allows us to use the spatial localization technique as in the proof of Theorem 2 in Li et al. (2017a). In particular, we can assume that $g(\cdot)$ has bounded derivatives up to the third order without loss of generality. The assertion of the theorem then follows from Theorem 3.2 in Jacod and Rosenbaum (2013). *Q.E.D.*

7.1 Proof of Theorem 1.

Proof of Theorem 1. (a) We set $Q_{n,ZY}(w) \equiv \sum_{i \in \mathcal{I}'_n} w(\hat{c}_{n,i-}, \hat{c}_{n,i}, \breve{\beta}_n) \Delta_i^n Z \Delta_i^n Y$, so that we can rewrite

$$\hat{\beta}_n(v, w) = \frac{\hat{S}_n(v g_b) + Q_{n,ZY}(w)}{\hat{S}_n(v) + Q_{n,ZZ}(w)}.$$

Hence,

$$\Delta_n^{-1/2}\left(\hat{\beta}_n(v,w) - \beta_0\right) = \frac{\Delta_n^{-1/2}\left(\hat{S}_n(v g_b) - \beta_0 \hat{S}_n(v)\right) + \Delta_n^{-1/2}\left(Q_{n,ZY}(w) - \beta_0 Q_{n,ZZ}(w)\right)}{\hat{S}_n(v) + Q_{n,ZZ}(w)}.$$

Define $g(c) \equiv v(c) g_b(c) - \beta_0 v(c)$ and note that $S(g) = 0$. Applying Lemma 1 and simplifying the form of asymptotic variance with direct calculation, we deduce

$$\Delta_n^{-1/2} \hat{S}_n(g) \xrightarrow{\mathcal{L}\text{-}s} \mathcal{MN}\left(0, \int_0^T v(c_s)^2 \frac{\varsigma_s}{c_{ZZ,s}} ds\right). \tag{25}$$

In addition, since $\mathcal{I}'_n = \mathcal{I}$ with probability approaching one, we have

$$\Delta_n^{-1/2}\left(Q_{n,ZY}(w) - \beta_0 Q_{n,ZZ}(w)\right) = \sum_{p \in \mathcal{P}} w(\hat{c}_{n,i(p)-}, \hat{c}_{n,i(p)}, \breve{\beta}_n) \Delta_{i(p)}^n Z \rho_{n,p},$$

where $\rho_{n,p} \equiv (-\beta_0, 1) R_{n,p}$ and $R_{n,p} \equiv \Delta_n^{-1/2}(\Delta_{i(p)}^n X - \Delta X_{\tau_p})$. Theorem 9.3.2 in Jacod and Protter (2012) implies that $\hat{c}_{n,i(p)-} = c_{\tau_p -} + o_p(1)$ and $\hat{c}_{n,i(p)} = c_{\tau_p} + o_p(1)$ for each $p \geq 1$.

Proposition 1(b) in Li et al. (2017b) implies $\Delta^n_{i(p)} Z = \Delta Z_{\tau_p} + o_p(1)$. Since $\breve{\beta}_n = \beta_0 + o_p(1)$, we can apply the continuous mapping theorem to deduce

$$w(\hat{c}_{n,i(p)-}, \hat{c}_{n,i(p)}, \breve{\beta}_n) \xrightarrow{\mathbb{P}} w\left(c_{\tau_p -}, c_{\tau_p}, \beta_0\right).$$

By Proposition 4.4.10 in Jacod and Protter (2012), $(R_{n,p})_{p\geq 1}$ converges stably in law to $(R_p)_{p\geq 1}$ and, hence, $(\rho_{n,p})_{p\geq 1}$ converges stably in law to $(\rho_p)_{p\geq 1}$. Note that the variables $(\rho_p)_{p\geq 1}$ are $\mathcal{F}$-conditionally independent and, under Assumption 5, are $\mathcal{F}$-conditionally centered mixed Gaussian with variance ς_{τ_p}. Since the set $\mathcal{P}$ is finite almost surely, we have

$$\Delta_n^{-1/2}\left(Q_{n,ZY}(w) - \beta_0 Q_{n,ZZ}(w)\right) \xrightarrow{\mathcal{L}\text{-}s} \mathcal{MN}\left(0, \sum_{p\in\mathcal{P}} w(c_{\tau_p -}, c_{\tau_p}, \beta_0)^2 \left(\Delta Z_{\tau_p}\right)^2 \varsigma_{\tau_p}\right). \quad (26)$$

By a standard argument, we can show that the convergences in (25) and (26) hold jointly with $\mathcal{F}$-conditionally independent limits. Moreover, it is easy to show that $\hat{S}_n(v) = S(v) + o_p(1)$ and $Q_{n,ZZ}(w) = Q_{ZZ}(w) + o_p(1)$. Hence,

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

where

$$\Sigma(v,w) = \frac{\int_0^T v(c_s)^2 \frac{\varsigma_s}{c_{ZZ,s}} ds + \sum_{\tau\in\mathcal{T}} w(c_{\tau-}, c_\tau, \beta_0)^2 \left(\Delta Z_\tau\right)^2 \varsigma_\tau}{\left(S(v) + Q_{ZZ}(w)\right)^2}.$$

(b) Since $\hat{\beta}^\star_n$ is a special case of $\hat{\beta}_n(v,w)$, the asserted convergence follows readily from part (a). It remains to solve the minimization of $\Sigma(v,w)$. We consider two generic weight functions v and w. Below, we write $v_s = v(c_s)$ and $w_\tau = w(c_{\tau-}, c_\tau, \beta_0)$ for notational simplicity. Observe that

$$\begin{aligned}
\frac{\Sigma(v,w)}{V^\star} &= \frac{\left(\int_0^T \frac{v_s^2 \varsigma_s}{c_{ZZ,s}} ds + \sum_{\tau\in\mathcal{T}} w_\tau^2 \left(\Delta Z_\tau\right)^2 \varsigma_\tau\right)\left(\int_0^T \frac{c_{ZZ,s}}{\varsigma_s} ds + \sum_{\tau\in\mathcal{T}} \frac{(\Delta Z_\tau)^2}{\varsigma_\tau}\right)}{\left(S(v) + Q_{ZZ}(w)\right)^2} \\
&\geq \frac{\left(\int_0^T \frac{v_s^2 \varsigma_s}{c_{ZZ,s}} ds\right)\left(\int_0^T \frac{c_{ZZ,s}}{\varsigma_s} ds\right) + \left(\sum_{\tau\in\mathcal{T}} w_\tau^2 \left(\Delta Z_\tau\right)^2 \varsigma_\tau\right)\left(\sum_{\tau\in\mathcal{T}} \frac{(\Delta Z_\tau)^2}{\varsigma_\tau}\right)}{\left(S(v) + Q_{ZZ}(w)\right)^2} \\
&\quad + \frac{2\sqrt{\left(\int_0^T \frac{v_s^2 \varsigma_s}{c_{ZZ,s}} ds\right)\left(\int_0^T \frac{c_{ZZ,s}}{\varsigma_s} ds\right)\left(\sum_{\tau\in\mathcal{T}} w_\tau^2 \left(\Delta Z_\tau\right)^2 \varsigma_\tau\right)\left(\sum_{\tau\in\mathcal{T}} \frac{(\Delta Z_\tau)^2}{\varsigma_\tau}\right)}}{\left(S(v) + Q_{ZZ}(w)\right)^2} \\
&\geq \frac{\left(\int_0^T v_s ds\right)^2 + \left(\sum_{\tau\in\mathcal{T}} w_\tau \left(\Delta Z_\tau\right)^2\right)^2 + 2\left(\int_0^T v_s ds\right)\left(\sum_{\tau\in\mathcal{T}} w_\tau \left(\Delta Z_\tau\right)^2\right)}{\left(S(v) + Q_{ZZ}(w)\right)^2} \\
&= 1
\end{aligned}$$

where the first inequality is due to a simple quadratic inequality and the equality is attained if and only if

$$\left(\int_0^T \frac{v_s^2 \varsigma_s}{c_{ZZ,s}} ds\right) \left(\sum_{\tau \in \mathcal{T}} \frac{(\Delta Z_\tau)^2}{\varsigma_\tau}\right) = \left(\int_0^T \frac{c_{ZZ,s}}{\varsigma_s} ds\right) \left(\sum_{\tau \in \mathcal{T}} w_\tau^2 (\Delta Z_\tau)^2 \varsigma_\tau\right);$$

the second inequality is due to the Cauchy–Schwarz inequality and the equality holds if and only if $v(c)$ is a multiple of c_{ZZ}/ς and $w(c_\tau)$ is a multiple of $1/\varsigma_\tau$. Taking these estimates together, we deduce that $V^\star \leq \Sigma(v,w)$ in general and the equality holds if and only if $v(c) = \varphi c_{ZZ}/\varsigma$ and $w(c_\tau) = \varphi/\varsigma_\tau$ for some constant $\varphi > 0$ (recall that v and w are strictly positive functions).

(c) By Theorem 3 of Li et al. (2017a), $\hat{S}_n(v^\star) = \int_0^T (c_{ZZ,s}/\varsigma_s) ds + o_p(1)$. A similar argument to part (a) yields $Q_{n,ZZ}(w^\star) = Q_{ZZ}(w^\star) + o_p(1)$. The assertion of part (c) readily follows from the convergence results. *Q.E.D.*

7.2 Proof of Theorem 2.

Proof of Theorem 2. We consider a sequence Ω_n of events defined by

$$\Omega_n = \left\{ \begin{array}{l} \text{For every } 1 \leq i \leq [T/\Delta_n], \ \ ((i-1)\Delta_n, i\Delta_n] \\ \text{contains at most one jump of } Z. \end{array} \right\}.$$

Under the maintained assumptions, the process Z has finitely active jumps. Hence, $\mathbb{P}(\Omega_n) \to 1$ and we can restrict our calculation below on Ω_n without loss of generality.

We denote the log likelihood ratio by

$$L_n(h) \equiv \log \frac{dP^n_{\beta_0 + \Delta_n^{1/2} h}}{dP^n_{\beta_0}}, \quad h \in \mathbb{R}.$$

Let $\mathcal{G}$ denote the σ-field generated by the processes $(b, c_{ZZ}, \varsigma, J_Z, \epsilon)$. Given the maintained assumptions, we see that, under the law P^n_β, the observed returns $(\Delta_i^n X)_{i \geq 0}$ are independently normally distributed conditional on $\mathcal{G}$. Using this fact, we can obtain an explicit expression for $L_n(h)$. For notational simplicity, we denote

$$z_{n,i} \equiv \Delta_n^{-1/2} \int_{(i-1)\Delta_n}^{i\Delta_n} \sqrt{c_{ZZ,s}} dW_{Z,s}, \quad y_{n,i} \equiv \Delta_n^{-1/2} \int_{(i-1)\Delta_n}^{i\Delta_n} \sqrt{\varsigma_s} dW_{\epsilon,s},$$
$$\bar{c}_{n,i} \equiv \Delta_n^{-1} \int_{(i-1)\Delta_n}^{i\Delta_n} c_{ZZ,s} ds, \quad \bar{\varsigma}_{n,i} \equiv \Delta_n^{-1} \int_{(i-1)\Delta_n}^{i\Delta_n} \varsigma_s ds.$$

Some straightforward (although somewhat cumbersome) algebra yields

$$L_n(h) = h\tilde{\psi}_n - \frac{h^2}{2}\Gamma_n, \tag{27}$$

where

$$\tilde{\psi}_n \equiv \sum_{i=1}^{[T/\Delta_n]} \frac{y_{n,i}\left(\Delta_i^n J_Z + \Delta_n^{1/2} z_{n,i}\right)}{\bar{\varsigma}_{n,i}}, \quad \Gamma_n \equiv \sum_{i=1}^{[T/\Delta_n]} \frac{\left(\Delta_i^n J_Z + \Delta_n^{1/2} z_{n,i}\right)^2}{\bar{\varsigma}_{n,i}}.$$

It remains to analyze the asymptotic properties of $\tilde{\psi}_n$ and Γ_n. We decompose $\tilde{\psi}_n = \tilde{\psi}'_n + \tilde{\psi}''_n$, where $\tilde{\psi}'_n$ and $\tilde{\psi}''_n$ are sums over the subset $\{i : \Delta_i^n J_Z \neq 0\}$ and its complement, respectively. Similarly, we decompose $\Gamma_n = \Gamma'_n + \Gamma''_n$.

We now proceed to derive the joint convergence in law of $(\tilde{\psi}'_n, \tilde{\psi}''_n)$ under the $\mathcal{G}$-conditional probability. We note that $\Delta_i^n J_Z$ and $\bar{\varsigma}_{n,i}$ are $\mathcal{G}$-measurable, and $(z_{n,i}, y_{n,i})$ are $\mathcal{G}$-conditionally independent with conditional distributions given by

$$z_{n,i}|\mathcal{G} \sim \mathcal{MN}\left(0, \bar{c}_{n,i}\right), \quad y_{n,i}|\mathcal{G} \sim \mathcal{MN}\left(0, \bar{\varsigma}_{n,i}\right).$$

In particular, $\tilde{\psi}'_n$ and $\tilde{\psi}''_n$ are $\mathcal{G}$-conditionally independent, so it is enough to derive the marginal convergence of each sequence. Since the jumps are finitely active, it is easy to see that

$$\tilde{\psi}'_n = \sum_{i:\Delta_i^n J_Z \neq 0} \frac{y_{n,i}\Delta_i^n J_Z}{\bar{\varsigma}_{n,i}} + o_p(1), \quad \tilde{\psi}''_n = \Delta_n^{1/2} \sum_{i=1}^{[T/\Delta_n]} \frac{y_{n,i} z_{n,i}}{\bar{\varsigma}_{n,i}} + o_p(1).$$

By applying the Lindeberg–Lévy central limit theorem under the $\mathcal{G}$-conditional probability, we deduce the following $\mathcal{G}$-conditional convergence in law

$$\tilde{\psi}'_n \xrightarrow{\mathcal{L}} \mathcal{MN}\left(0, \sum_{\tau\in\mathcal{T}} \frac{\Delta Z_\tau^2}{\varsigma_\tau}\right), \quad \tilde{\psi}''_n \xrightarrow{\mathcal{L}} \mathcal{MN}\left(0, \int_0^T \frac{c_{ZZ,s}}{\varsigma_s} ds\right).$$

From here, we deduce the following convergence under the $\mathcal{G}$-conditional probability,

$$\tilde{\psi}_n \xrightarrow{\mathcal{L}} \tilde{\psi} \sim \mathcal{MN}\left(0, \int_0^T \frac{d[Z,Z]_s}{\varsigma_s}\right). \tag{28}$$

Similarly, we can derive the convergence in probability for Γ_n:

$$\Gamma_n \xrightarrow{\mathbb{P}} \Gamma \equiv \int_0^T \frac{d[Z,Z]_s}{\varsigma_s}. \tag{29}$$

Since Γ_n is $\mathcal{G}$-measurable, (28) and (29) imply that $(\tilde{\psi}_n, \Gamma_n)$ converges in law to $(\tilde{\psi}, \Gamma)$. From here, the assertion of the theorem readily follows (recall (27)). *Q.E.D.*

7.3 Proof of Proposition 1.

The proof of Proposition 1 relies on two auxiliary lemmas. Lemma 2 describes the joint convergence for a generic weighted common beta estimator and the adaptive estimator. In particular, this lemma shows clearly that $\hat{\beta}_n(v,w) - \hat{\beta}_n^\star$ is asymptotically orthogonal to $\hat{\beta}_n^\star$. Lemma 3 establishes the probability limit of $\hat{\beta}_n^\star$ under the alternative hypothesis (i.e., the common beta assumption does not hold).

Lemma 2 *Suppose that the conditions in Theorem 1 hold. Then, in restriction to Ω_0, we have*

$$\Delta_n^{-1/2}\left(\hat{\beta}_n(v,w) - \beta_0, \hat{\beta}_n^\star - \beta_0\right) \xrightarrow{\mathcal{L}\text{-}s} \mathcal{MN}\left(0, \begin{pmatrix} \Sigma(v,w) & V^\star \\ V^\star & V^\star \end{pmatrix}\right). \tag{30}$$

Proof of Lemma 2. Recall that $\hat{S}_n(v) = S(v) + o_p(1)$ and $Q_{n,ZZ}(w) = Q_{ZZ}(w) + o_p(1)$. Hence, the left-hand side of (30) has the asymptotic representation $\zeta_{n,1} + \zeta_{n,2} + o_p(1)$, where

$$\begin{aligned} \zeta_{n,1} &\equiv \left(\frac{\Delta_n^{-1/2}\hat{S}_n(vg_b - \beta_0 v)}{S(v) + Q_{ZZ}(w)}, \frac{\Delta_n^{-1/2}\hat{S}_n(v^\star g_b - \beta_0 v^\star)}{S(v^\star) + Q_{ZZ}(w^\star)}\right)^\top, \\ \zeta_{n,2} &\equiv \left(\frac{\Delta_n^{-1/2}(Q_{n,ZY}(w) - \beta_0 Q_{n,ZZ}(w))}{S(v) + Q_{ZZ}(w)}, \frac{\Delta_n^{-1/2}(Q_{n,ZY}(w^\star) - \beta_0 Q_{n,ZZ}(w^\star))}{S(v^\star) + Q_{ZZ}(w^\star)}\right)^\top. \end{aligned}$$

Next, we proceed to derive the joint stable convergence in law for $\zeta_{n,1}$ and $\zeta_{n,2}$. Since the central limit theorem for $\zeta_{n,2}$ is driven completely by the Brownian motion in finitely many jump intervals, a usual argument implies that $\zeta_{n,1}$ and $\zeta_{n,2}$ are asymptotically $\mathcal{F}$-conditionally independent. Hence, it suffices to derive the marginal convergence of $\zeta_{n,1}$ and $\zeta_{n,2}$.

We set $g(\cdot) = (v(\cdot)g_b(\cdot) - \beta_0 v(\cdot), v^\star(\cdot)g_b(\cdot) - \beta_0 v^\star(\cdot))^\top$. In restriction to Ω_0, $S(g) = 0$. By Lemma 1,

$$\Delta_n^{-1/2}\hat{S}_n(g) \xrightarrow{\mathcal{L}\text{-}s} \mathcal{MN}(0, V(g)),$$

where the asymptotic covariance matrix is given by (after some straightforward algebra)

$$V(g) = \begin{pmatrix} \int_0^T v(c_s)^2 \frac{\varsigma_s}{c_{ZZ,s}} ds & \int_0^T v(c_s) ds \\ \int_0^T v(c_s) ds & \int_0^T \frac{c_{ZZ,s}}{\varsigma_s} ds \end{pmatrix}.$$

From here, we deduce

$$\zeta_{n,1} \xrightarrow{\mathcal{L}\text{-}s} \mathcal{MN}(0, \Sigma_1), \tag{31}$$

where

$$\Sigma_1 = \begin{pmatrix} \frac{\int_0^T v(c_s)^2 \frac{\varsigma_s}{c_{ZZ,s}} ds}{(S(v)+Q_{ZZ}(w))^2} & \frac{S(v)}{(S(v^\star)+Q_{ZZ}(w^\star))(S(v)+Q_{ZZ}(w))} \\ \frac{S(v)}{(S(v^\star)+Q_{ZZ}(w^\star))(S(v)+Q_{ZZ}(w))} & \frac{\int_0^T \frac{c_{ZZ,s}}{\varsigma_s} ds}{(S(v^\star)+Q_{ZZ}(w^\star))^2} \end{pmatrix}.$$

Turning to $\zeta_{n,2}$, we note that with probability approaching one,

$$\zeta_{n,2} = \sum_{p \in \mathcal{P}} \begin{pmatrix} \frac{w\left(\hat{c}_{n,i(p)-}, \hat{c}_{n,i(p)}, \breve{\beta}_n\right)}{S(v)+Q_{ZZ}(w)} \\ \frac{1/\hat{\varsigma}_{n,i(p)}}{S(v^\star)+Q_{ZZ}(w^\star)} \end{pmatrix} \Delta_{i(p)}^n Z \rho_{n,p}.$$

Recall from the proof of Theorem 1 that $w(\hat{c}_{i(p)-}^n, \hat{c}_{i(p)}^n, \breve{\beta}_n) \xrightarrow{\mathbb{P}} w(c_{\tau_p -}, c_{\tau_p}, \beta_0)$ and $(\rho_{n,p})_{p \geq 1}$ converges stably in law to $(\rho_p)_{p \geq 1}$, where $(\rho_p)_{p \geq 1}$ are $\mathcal{F}$-conditionally independent. Since $\mathcal{P}$ is finite almost surely, we deduce from the property of stable convergence in law that

$$\zeta_{n,2} \xrightarrow{\mathcal{L}\text{-}s} \sum_{p \in \mathcal{P}} \begin{pmatrix} \frac{w\left(c_{\tau_p -}, c_{\tau_p}, \beta_0\right)}{S(v)+Q_{ZZ}(w)} \\ \frac{1/\varsigma_{\tau_p}}{S(v^\star)+Q_{ZZ}(w^\star)} \end{pmatrix} \Delta Z_{\tau_p} \rho_p.$$

Under Assumption 5, ρ_p is conditionally centered mixed Gaussian with variance ς_{τ_p} for each $p \in \mathcal{P}$. Hence,

$$\zeta_{n,2} \xrightarrow{\mathcal{L}\text{-}s} \mathcal{MN}\left(0, \Sigma_2\right), \tag{32}$$

where

$$\Sigma_2 \equiv \begin{pmatrix} \frac{\sum_{p \in \mathcal{P}} w\left(c_{\tau_p -}, c_{\tau_p}, \beta_0\right)^2 \left(\Delta Z_{\tau_p}\right)^2 \varsigma_{\tau_p}}{(S(v)+Q_{ZZ}(w))^2} & \frac{Q_{ZZ}(w)}{(S(v^\star)+Q_{ZZ}(w^\star))(S(v)+Q_{ZZ}(w))} \\ \frac{Q_{ZZ}(w)}{(S(v^\star)+Q_{ZZ}(w^\star))(S(v)+Q_{ZZ}(w))} & \frac{\sum_{p \in \mathcal{P}} \left(\Delta Z_{\tau_p}\right)^2 / \varsigma_{\tau_p}}{(S(v^\star)+Q_{ZZ}(w^\star))^2} \end{pmatrix}.$$

Finally, we note that $V^\star = \left(S\left(v^\star\right) + Q_{ZZ}\left(w^\star\right)\right)^{-1}$. The assertion of the lemma readily follows from (31) and (32). *Q.E.D.*

Lemma 3 *Suppose that the conditions in Theorem 1 hold. Then, in the general case without imposing the common beta hypothesis,*

$$\hat{\beta}_n^\star \xrightarrow{\mathbb{P}} \bar{\beta}^\star \equiv \left(\int_0^T \frac{d\langle Z^c, Z^c \rangle_s}{\varsigma_s} \beta_s^c + \sum_{\tau \in \mathcal{T}} \frac{\left(\Delta Z_\tau\right)^2}{\varsigma_\tau} \beta_\tau^d \right) \Big/ \int_0^T \frac{d\left[Z, Z\right]_s}{\varsigma_s},$$
$$\hat{\beta}_n^{BNS} \xrightarrow{\mathbb{P}} \bar{\beta}^{BNS} \equiv \left(\int_0^T d\langle Z^c, Z^c \rangle_s \beta_s^c + \sum_{\tau \in \mathcal{T}} \left(\Delta Z_\tau\right)^2 \beta_\tau^d \right) \Big/ \int_0^T d\left[Z, Z\right]_s = \frac{[Y, Z]_T}{[Z, Z]_T},$$

where $\beta_s^c = c_{ZY,s}/c_{ZZ,s}$ *and* $\beta_\tau^d \equiv \Delta Y_\tau / \Delta Z_\tau$, $\tau \in \mathcal{T}$.

Proof of Lemma 3. For each $p \geq 1$, let $\beta_{n,i(p)} \equiv \Delta_{i(p)}^n Y / \Delta_{i(p)}^n Z$. With probability approaching one, we have

$$\hat{\beta}_n^{\star} = \frac{\hat{S}_n(v^{\star} g_b) + \sum_{p \in \mathcal{P}} \beta_{n,i(p)} \left(\Delta_{i(p)}^n Z \right)^2 / \hat{\varsigma}_{n,i(p)}}{\hat{S}_n(v^{\star}) + Q_{n,ZZ}(w^{\star})}.$$

Note that the following convergence holds on the entire space

$$\begin{array}{l} \hat{S}_n(v^{\star} g_b) \xrightarrow{\mathbb{P}} S\left(v^{\star} g_b\right), \quad \hat{S}_n(v^{\star}) \xrightarrow{\mathbb{P}} S\left(v^{\star}\right), \quad \hat{\varsigma}_{n,i(p)} \xrightarrow{\mathbb{P}} \varsigma_{\tau_p}, \\ Q_{n,ZZ}(w^{\star}) \xrightarrow{\mathbb{P}} \sum_{\tau \in \mathcal{T}} \left(\Delta Z_{\tau}\right)^2 / \varsigma_{\tau}, \quad \Delta_{i(p)}^n Z \xrightarrow{\mathbb{P}} \Delta Z_{\tau_p}, \quad \beta_{n,i(p)} \xrightarrow{\mathbb{P}} \beta_{\tau_p}. \end{array}$$

In addition, note that $\beta_s^c = g_b\left(c_s\right)$. The first assertion readily follows from these convergence results. The second convergence follows from almost the same argument. *Q.E.D.*

Proof of Proposition 1. (a) Note that $\hat{\beta}_n^{BNS}$ is a special case of $\hat{\beta}_n\left(v, w\right)$ with $v\left(c\right) = c_{ZZ}$ and $w\left(\cdot\right) = 1$. Applying Lemma 2, we get

$$\Delta_n^{-1/2} \left(\hat{\beta}_n^{BNS} - \beta_0, \hat{\beta}_n^{\star} - \beta_0 \right) \xrightarrow{\mathcal{L}\text{-}s} \mathcal{MN}\left(0, \begin{pmatrix} V^{BNS} & V^{\star} \\ V^{\star} & V^{\star} \end{pmatrix} \right),$$

where V^{BNS} is defined in (22). This further implies that (recalling $\Xi = V^{BNS} - V^{\star}$),

$$\Delta_n^{-1/2} \left(\hat{\beta}_n^{\star} - \hat{\beta}_n^{BNS} \right) \xrightarrow{\mathcal{L}\text{-}s} \mathcal{MN}\left(0, \Xi\right).$$

Similar to the proof of Theorem 1(c), we can show that

$$\hat{V}_n^{BNS} \xrightarrow{\mathbb{P}} V^{BNS},$$

and hence, $\hat{\Xi}_n \xrightarrow{\mathbb{P}} \Xi$. The assertion of part (a) then follows from these convergence results.

(b) The size property is implied by part (a). We note that $\hat{\Xi}_n \xrightarrow{\mathbb{P}} \Xi$ holds over the whole sample space. By Lemma 3, the $|\hat{T}_n|$ diverges in probability to $+\infty$ whenever $\bar{\beta}^{BNS} \neq \bar{\beta}^{\star}$. The power statement then readily follows. *Q.E.D.*

References

Andersen, T., T. Bollerslev, F. Diebold, and G. Wu (2006). Realized Beta: Persistence and Predictability. In T. Fomby and D. Terrell (Eds.), *Advances in Econometrics: Econometric Analysis of Economic and Financial Time Series*, Volume 20. Elsevier Science.

Barndorff-Nielsen, O. and N. Shephard (2004a). Econometric Analysis of Realized Covariation: High Frequency Based Covariance, Regression, and Correlation in Financial Economics. *Econometrica 72*(3), pp. 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), 1–37.

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.

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

Bollerslev, T. and B. Zhang (2003). Measuring and Modeling Systematic Risk in Factor Pricing Models using High-frequency Data. *Journal of Empirical Finance 10*, 533–558.

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

Gospodinov, N., R. Kan, and C. Robotti (2009). Misspecification-Robust Inference in Linear Asset Pricing Models with Irrelevant Risk Factors. *Review of Financial Studies 27*, 2139–2170.

Hausman, J. A. (1978). Specification Tests in Econometrics. *Econometrica*, 1251–1271.

Jacod, J. and P. Protter (2012). *Discretization of Processes*, Volume 67. Springer, Heidelberg.

Jacod, J. and M. Rosenbaum (2013). Quarticity and Other Functionals of Volatility: Efficient Estimation. *The Annals of Statistics 41*(3), 1462–1484.

Jaganathan, R. and Z. Wang (1998). An Asymptotic Theory for Estimating Beta-Pricing Models using Cross-Sectional Regression. *Journal of Finance 53*, 1285–1309.

Jeganathan, P. (1982). On the Asymptotic Theory of Estimation when the Limit of the Log-Likelihood Ratios is Mixed Normal. *Sankhyā: The Indian Journal of Statistics, Series A*, 173–212.

Jeganathan, P. (1983). Some Asymptotic Properties of Risk Functions when the Limit of the Experiment is Mixed Normal. *Sankhyā: The Indian Journal of Statistics, Series A*, 66–87.

Kalnina, I. (2012). Nonparametric Tests of Time Variation in Betas. Technical report, Université de Montréal.

Kan, R. and C. Zhang (1999). Two-Pass Tests of Asset Pricing Models with Useless Factors. *Journal of Finance 54*, 204–235.

Kleibergen, F. (2009). Tests of Risk Premia in Linear Factor Models. *Journal of Econometrics 149*, 149–173.

Li, J., V. Todorov, and G. Tauchen (2013). Volatility Occupation Times. *Annals of Statistics 41*, 1865–1891.

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

Li, J., V. Todorov, and G. Tauchen (2017b). Jump Regressions. *Econometrica 85*(1), 173–195.

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

Mancini, C. (2001). Disentangling the Jumps of the Diffusion in a Geometric Jumping Brownian Motion. *Giornale dell'Istituto Italiano degli Attuari 64*(19-47), 44.

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

Mykland, P. and L. Zhang (2009). Inference for Continuous Semimartingales Observed at High Frequency. *Econometrica 77*, 1403–1445.

Patton, A. and M. Verardo (2012). Does Beta Move with News? Firm-specific Information Flows and Learning about Profitability. *Review of Financial Studies 25*, 2789–2839.

Shanken, J. (1992). On the Estimation of Beta Pricing Models. *Review of Financial Studies 5*, 1–33.

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