Adaptive Estimation of Continuous-Time Regression Models using High-Frequency Data*

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

January 11, 2016

Abstract

We derive the asymptotic efficiency bound for regular estimates of the slope coefficient in a linear continuous-time regression model for the continuous martingale parts of two Itô semimartingales observed on a fixed time interval with asymptotically shrinking mesh of the observation grid. We further construct an estimator from high-frequency data that achieves this efficiency bound and, indeed, is adaptive to the presence of infinite-dimensional nuisance components. The estimator is formed by taking optimal weighted average of local nonparametric volatility estimates that are constructed over blocks of high-frequency observations. To study the asymptotic behavior of the proposed estimator, we introduce a general spatial localization procedure which extends known results on the estimation of integrated volatility functionals to more general classes of functions of volatility. Empirically relevant numerical examples illustrate that the proposed efficient estimator provides nontrivial improvement over alternatives in the extant literature.

Keywords: adaptive estimation; beta; stochastic volatility; spot variance; semiparametric efficiency; high-frequency data.

*We would like to thank the Editor (Oliver Linton), the Associate Editor and three anonymous referees for many helpful comments and suggestions. Li's and Todorov's research have been partially supported by NSF grants SES-1326819 and SES-1530748, respectively.

†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

We study the problem of efficient inference for the slope coefficient β in a linear regression of the form

$$Y_t^c = \beta^\intercal Z_t^c + \varepsilon_t, \quad t \in [0,1], \tag{1.1}$$

where Y^c and Z^c are respectively the continuous local martingale components of two Itô semimartingales Y and Z, and ε is the "residual" process which is orthogonal in the martingale sense to Z. Our asymptotic setup is of infill type: the processes Y and Z are observed on a fixed time interval with mesh of the equidistant observation grid shrinking asymptotically to zero. The continuous-time linear regression (1.1) arises in many situations such as linear factor models where β captures the factor loadings. In financial applications estimators for β from high-frequency data have been considered by Bollerslev and Zhang (2003), Barndorff-Nielsen and Shephard (2004), Andersen et al. (2006), Mykland and Zhang (2006, 2009), Todorov and Bollerslev (2010), Gobbi and Mancini (2012), Patton and Verardo (2012), Kalnina (2012) among many others.

In this paper we are interested in deriving an asymptotic semiparametric efficiency bound for the estimation of the vector β from high frequency data, as well as constructing feasible efficient estimators that achieve it, in the presence of nuisance components that govern the law of $(Y, Z^\intercal)$. The nuisance components include the unobservable stochastic volatility processes of the two continuous local martingales Z^c and Y^c, i.e., their stochastic diffusion coefficients, as well as the jumps and the drift components of Z and Y. Our main result is the construction of an efficient estimator for β which is adaptive to the presence of these nuisance components; see Bickel (1982) and Bickel et al. (1998) for classical results on adaptive estimation.

We establish the adaptive estimation result by first showing the local asymptotic mixed normality (LAMN) property of a parametric submodel in which the only unknown parameter is β. A conditional convolution theorem (as in Jeganathan (1982)) allows us to derive a lower efficiency bound for estimating β among the class of regular estimators. More precisely, we show that the limit law of every regular estimator of β in the parametric setting can be expressed as a convolution of a centered mixed Gaussian distribution and another distribution. Importantly, the conditional covariance matrix of the above mixed Gaussian distribution depends only on the spot covariation matrix process of the continuous local martingale part of $(Y, Z^\intercal)$. Following Clément et al. (2013), to establish the above efficiency bound, we use sufficient conditions that restrict (nontrivially) the general semimartingale setup by ruling out jumps in $(Y, Z^\intercal)$ and embedding $(Y, Z^\intercal)$ in a certain conditional Markov setting.

In the second part of our analysis we construct a class of regular estimators. We establish the asymptotic properties of these estimators in a very general semiparametric setting, far more general than the one needed for deriving the asymptotic efficiency bound discussed in the previous paragraph. We show that, despite the presence of various infinite-dimensional nuisance components, the optimal estimator in this class achieves the aforementioned efficiency bound for estimating β and, hence, is adaptive. Similar to Jacod and Rosenbaum (2013), our proposed class of estimators builds on local volatility estimates for the multivariate process $(Y, Z^{\intercal})$ constructed using high-frequency data over blocks of asymptotically shrinking time span. Efficient estimation is then conducted by optimally weighting local estimates of β across time blocks. We note that this is feasible only due to the stronger form of stable convergence that we establish for our inference procedures. The efficient estimation problem can be viewed as a continuous-time analogue to the work of Robinson (1987) on efficient regression estimation in presence of heteroskedasticity of unknown form, but our high-frequency setting leads to distinct theoretical results.

The construction of our estimators, the optimal one in particular, involves transforming the local volatility estimates with weight functions that do not have polynomial growth, which thus violates a key condition used in prior work (Jacod and Protter (2012), Jacod and Rosenbaum (2013)). For this reason, we extend the limit results of Jacod and Rosenbaum (2013) to a more general class of volatility functionals. The key insight underlying the derivation of these results is the uniform convergence of our block volatility estimators toward the average volatility over the blocks, based on which a spatial localization procedure can be applied to extend the space of test functions which is considered in prior work. These results are of independent interest and allow one to carry out efficient (in the sense of Clément et al. (2013) and Renault et al. (2015)) estimation of integrated nonlinear transforms of volatility for univariate transforms such as $\log(\cdot)$, $\sqrt{\cdot}$, $\exp(\cdot)$ and multivariate transforms such as the correlation coefficient (Kalnina and Xiu (2014)), the idiosyncratic variance (Li et al. (2015)) and eigenvalues (Aït-Sahalia and Xiu (2015)). The results in Jacod and Rosenbaum (2013) do not apply in these cases.

By using optimal weighting, our proposed estimator is more efficient than an estimator with no weighting as studied by Jacod and Rosenbaum (2013). Of course, the results in Jacod and Rosenbaum (2013) are fully nonparametric while ours are semiparametric due to (1.1). The efficiency gain arises from the parametric restriction imposed by (1.1), which is essentially equivalent to assuming that β in (1.1) is constant instead of being a general stochastic process. That said, the semiparametric setup considered here appears in many economic and financial applications (and is empirically plausible over short samples; see, e.g. Reiss et al. (2015)). Our results provide a useful starting point for studying the semiparametrically efficient estimation in more complicated models.

We illustrate the efficiency gain of our estimator by comparing it in a bivariate setting with two alternatives that have been used before. The first is an estimator based on the results in Jacod and Rosenbaum (2013) discussed above, see also Mykland and Zhang (2009). The second is the estimator in Barndorff-Nielsen and Shephard (2004), Todorov and Bollerslev (2010) and Gobbi and Mancini (2012). The latter is simply based on expressing β as the ratio of the continuous covariation between Y and Z on the fixed interval and the continuous quadratic variation of Z, and forming consistent estimates of these two quantities. We show analytically that both alternative estimators are as efficient as ours only when certain transforms of the multivariate volatility process remain constant on the time interval. The efficiency gain of our estimator arises when such constancy condition on the volatility process fails. We show that these gains can be sizable in realistically calibrated situations. In an empirical application to a stock from the financial sector we illustrate the efficiency gains of the proposed estimator.

The rest of this paper is organized as follows. In Section 2 we introduce the formal setup and present an efficiency bound for estimating β. We present generic limit theorems for integrated volatility functionals in Section 3. Section 4 presents results on the adaptive estimation of beta. An empirical application is given in Section 5. Section 6 concludes. The appendix contains all proofs.

2 The regression model

In Section 2.1, we present the setting for the continuous-time regression model of interest. In Section 2.2, we establish an efficiency bound for estimating the model parameter under some additional assumptions.

We use the following notation throughout. For any matrix A, we denote its transpose by $A^{\intercal}$ and its (i,j) element by $[A]_{ij}$. The partial derivative of the function $A \mapsto g(A)$ with respect to $[A]_{ij}$ is denoted by $\partial_{ij} g(A)$, which has the same dimensionality as g; the second partial derivative $\partial^2_{jk,lm} g(A)$ with respect to $[A]_{jk}$ and $[A]_{lm}$ is understood similarly. We denote by $\mathcal{M}_d$ the collection of $d \times d$ positive semidefinite real-valued matrices. For $A \in \mathcal{M}_d$, $\lambda_{\min}(A)$ denotes the smallest eigenvalue of A. The d-dimensional identity matrix (resp. zero vector) is denoted by I_d (resp. 0_d). The $d \times d$ zero matrix is denoted by $\mathbf{0}_d$. For two sequences of positive real numbers $(a_n)_{n\geq 1}$ and (b_n), we write $a_n \asymp b_n$ if for some constant $c \geq 1$, $a_n/c \leq b_n \leq c a_n$. We use $\xrightarrow{\mathcal{L}}$ to denote convergence in law and use $\xrightarrow{\mathbb{P}}$ to denote convergence in probability. The Euclidean norm is denoted by $\|\cdot\|$.

2.1 The setting

The processes Z and Y in the linear regression (1.1) studied in this paper are assumed to take values in $\mathbb{R}^{d-1}$ and $\mathbb{R}$ (for $d \geq 2$), respectively, and we will denote $X = (Z^{\intercal}, Y)^{\intercal}$. With this notation, we will assume $(X_t)_{t\geq 0}$ to be an $\mathbb{R}^d$-valued semimartingale defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$. Throughout this paper, all processes are assumed to be càdlàg adapted. Our basic assumption is that X is an Itô semimartingale so it can be represented as (see, e.g., Jacod and Protter (2012), Section 2.1.4),

$$X_t = x_0 + \int_0^t b_s ds + \int_0^t \sigma_s dW_s + J_t, \tag{2.1}$$

where the drift b_t takes value in $\mathbb{R}^d$; the volatility process σ_t takes value in $\mathcal{M}_d$; W is a d-dimensional standard Brownian motion; J is a purely discontinuous process of the form

$$\begin{cases} J_t = \int_0^t \int_{\mathbb{R}} \delta(s,x) 1_{\{\|\delta(s,x)\|\leq 1\}} (\mu - \nu)(ds, dx) \\ \qquad + \int_0^t \int_{\mathbb{R}} \delta(s,x) 1_{\{\|\delta(s,x)\|>1\}} \mu(ds, dx), \end{cases} \tag{2.2}$$

and $\delta : \Omega \times \mathbb{R}_+ \times \mathbb{R} \mapsto \mathbb{R}^d$ is a predictable function, μ is a Poisson random measure on $\mathbb{R}_+ \times \mathbb{R}$ with its compensator $\nu(dt, dx) = dt \otimes \lambda(dx)$ for some σ-finite measure λ on $\mathbb{R}$.

The linear regression in (1.1) can be equivalently stated as

$$dY_t^c = \beta^\intercal dZ_t^c + d\varepsilon_t, \tag{2.3}$$

where $\beta \in \mathbb{R}^{d-1}$ is the parameter of interest and the disturbance process $(\varepsilon_t)_{t\geq 0}$ is a continuous local martingale that is orthogonal to $(Z_t^c)_{t\geq 0}$, i.e., $\langle \varepsilon, Z^c \rangle = 0$ identically, where $\langle \cdot, \cdot \rangle$ denotes the quadratic covariation. We can then represent model (2.1) as

$$\begin{cases} Z_t = z_0 + \int_0^t b_{Z,s} ds + \int_0^t \sigma_{Z,s} dW_{Z,s} + J_{Z,t} \\ Y_t = y_0 + \int_0^t b_{Y,s} ds + \beta^\intercal \int_0^t \sigma_{Z,s} dW_{Z,s} + \int_0^t \sigma_{\varepsilon,s} dW_{\varepsilon,s} + J_{Y,t}. \end{cases} \tag{2.4}$$

The processes in (2.1) and (2.4) are related by

$$b_t = \begin{pmatrix} b_{Z,t} \\ b_{Y,t} \end{pmatrix}, \quad J_t = \begin{pmatrix} J_{Z,t} \\ J_{Y,t} \end{pmatrix},$$
$$\sigma_t = \begin{pmatrix} \sigma_{Z,t} & 0_{d-1} \\ \beta^\intercal \sigma_{Z,t} & \sigma_{\varepsilon,t} \end{pmatrix}, \quad W_t = \begin{pmatrix} W_{Z,t} \\ W_{\varepsilon,t} \end{pmatrix}, \quad \langle W_Z, W_\varepsilon \rangle = 0_{d-1}.$$

Turning next to our statistical setting, we assume that the process X is observed at discrete times $t_i = i/n$, for $i = 0, 1, \ldots, n$, within the *fixed* time interval $[0, 1]$.[1] We denote the increments of X by

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

Asymptotically, the sampling interval goes to zero as $n \to \infty$. The data sequence $\left(X_{i/n}\right)_{0\leq i\leq n}$ is considered as a random variable taking values in a measurable space $(E_n, \mathcal{E}_n)$ with law P_n^X; below, we write $P_n^{X,\beta}$ to emphasize the dependence of P_n^X on β. In the continuous-time limit, it is well-known that the spot covariance matrix process $c_t \equiv \sigma_t \sigma_t^\intercal$ of X can be identified, which we further

[1] We leave for future work the analysis of the problem of efficient continuous-time regression estimation in the case when the data is contaminated by noise and the elements of X are asynchronously observed; see, for example, Hayashi and Yoshida (2011) and Bibinger et al. (2014) for ways to handle such complications in the data in the context of volatility estimation.

partition as

$$c_t = \begin{pmatrix} c_{ZZ,t} & c_{ZY,t} \\ c_{YZ,t} & c_{YY,t} \end{pmatrix} = \begin{pmatrix} \sigma_{Z,t}\sigma_{Z,t}^{\intercal} & \sigma_{Z,t}\sigma_{Z,t}^{\intercal}\beta \\ \beta^{\intercal}\sigma_{Z,t}\sigma_{Z,t}^{\intercal} & \beta^{\intercal}\sigma_{Z,t}\sigma_{Z,t}^{\intercal}\beta + \sigma_{\varepsilon,t}^2 \end{pmatrix}. \tag{2.5}$$

We observe that β can be identified as $\beta = c_{ZZ,t}^{-1} c_{ZY,t}$ provided that $c_{ZZ,t}$ is nonsingular. The spot variance of ε_t, that is, $c_{\varepsilon\varepsilon,t} \equiv \sigma_{\varepsilon,t}^2$ can be written as

$$c_{\varepsilon\varepsilon,t} = c_{YY,t} - c_{YZ,t} c_{ZZ,t}^{-1} c_{ZY,t}. \tag{2.6}$$

Equation (2.6) highlights the fact that the process $c_{\varepsilon\varepsilon,t}$ is a nonlinear transform of c_t. This definition is valid even when (2.3) is not imposed.

2.2 The efficiency bound for estimating β

We now characterize an efficiency bound for estimating β by using the conditional convolution theorem of Jeganathan (1982). A key part of the analysis is to establish the LAMN property in a parametric submodel. This task is done by applying the Malliavin calculus technique developed by Gobet (2001) and Clément et al. (2013), which is a useful device for studying the asymptotic behavior of the likelihood ratio. For this reason, we need some additional structure on (2.4) as in prior work:

$$\begin{cases} b_{Z,t} = b_Z(X_t), & b_{Y,t} = b_Y(X_t), \\ \sigma_{Z,t} = a_Z(X_t, F_t), & \sigma_{\varepsilon,t} = a_\varepsilon(X_t, F_t), \\ J_{Z,t} = 0_{d-1}, & J_{Y,t} = 0, \end{cases} \tag{2.7}$$

where $(F_t)_{t\geq 0}$ is a (possibly unobservable) $\mathbb{R}^q$-valued continuous Itô process and $b_Z(\cdot)$, $b_Y(\cdot)$, $a_Z(\cdot,\cdot)$ and $a_\varepsilon(\cdot,\cdot)$ are unknown smooth deterministic functions. In the basic case with $F_t = t$, X is a non-homogeneous diffusion process. More generally, the process F plays the role of latent stochastic volatility factors that drive the dynamics of X. Below, we consider F as a random element taking values in $\mathbb{C}$, the space of $\mathbb{R}^q$-valued continuous processes on $[0,1]$, and denote its law by P^F. The following regularity condition is used for the calculation of the efficiency bound.

Assumption L. We have (2.4) and (2.7) as well as the following.

(i) The functions $b_Z(\cdot)$, $b_Y(\cdot)$, $a_Z(\cdot,\cdot)$ and $a_\varepsilon(\cdot,\cdot)$ are three-times continuously differentiable with bounded derivatives.

(ii) There exists a constant $\underline{a} > 0$ such that for all $x \in \mathbb{R}^d$ and $f \in \mathbb{C}$, $(a_Z a_Z^{\intercal})(x,f) \geq \underline{a} I_{d-1}$ and $a_\varepsilon^2(x,f) \geq \underline{a}$.

(iii) The process $F = (F_t)_{t\geq 0}$ is continuous and is independent of W.

Assumption L(i) ensures regularities of the solution to the stochastic differential equation given by (2.4) and (2.7), and Assumption L(ii) further ensures the smoothness of the transition density of the process X. Assumption L(iii) effectively allows us to conduct analysis conditional on the process F and, hence, to treat the process X as Markovian.

The estimation of β is apparently complicated by the presence of various infinite-dimensional nuisance components, including unknown deterministic functions (i.e., $a_Z(\cdot)$, $a_\varepsilon(\cdot)$, $b_Z(\cdot)$, $b_Y(\cdot)$) and the process F. An important question in this semiparametric setting is whether the presence of these nuisance components affects the efficiency of estimating β.

To answer this question, we now compute the efficiency bound in a parametric submodel, in which the functions $b_Z(\cdot)$, $b_Y(\cdot)$, $a_Z(\cdot,\cdot)$ and $a_\varepsilon(\cdot,\cdot)$ are known and the process F is observed, leaving β as the only unknown parameter. In particular, the original data $(X_{i/n})_{0\leq i\leq n}$ are augmented by the observation of F. This model is of interest because an efficiency bound derived for this model holds *a fortiori* for any estimator based on $(X_{i/n})_{0\leq i\leq n}$.[2] Below, we denote the joint distribution of $((X_{i/n})_{0\leq i\leq n}, F)$ by $(P_n^\beta; \beta \in \mathbb{R}^{d-1})$. We show that this parametric submodel satisfies the LAMN property (Theorem 1), after recalling its definition (see Jeganathan (1982), Definition 1).

Definition 1 (LAMN)**.** *The sequence (P_n^β) satisfies the LAMN property at $\beta = \beta_0$ if there exists a sequence ζ_n of $\mathbb{R}^{d-1}$-valued random variables and a sequence Γ_n of $(d-1)\times(d-1)$ $P_n^{\beta_0}$-a.s. positive definite random matrices such that under the sequence $P_n^{\beta_0}$ of laws, for every $h \in \mathbb{R}^{d-1}$,*

$$\log \frac{dP_n^{\beta_0+n^{-1/2}h}}{dP_n^{\beta_0}} = h^\intercal \Gamma_n^{1/2}\zeta_n - \frac{1}{2}h^\intercal \Gamma_n h + o_p(1), \tag{2.8}$$

and

$$(\zeta_n, \Gamma_n) \xrightarrow{\mathcal{L}} (\zeta, \Gamma), \tag{2.9}$$

where Γ is an a.s. positive definite $(d-1)\times(d-1)$ matrix and ζ is a copy of the standard $(d-1)$-dimensional normal distribution independent of Γ.

We note that in the LAMN setting, the information matrix Γ is generally random, with Γ_n being its sample analogue. As a result, $\zeta_n \xrightarrow{\mathcal{L}} \zeta$ is not sufficient for the joint convergence (2.9). A sufficient condition for (2.9) is that ζ_n converges stably in law towards ζ and $\Gamma_n \xrightarrow{\mathbb{P}} \Gamma$ where Γ is $\mathcal{F}$-adapted.

Theorem 1. *Under Assumption L, (P_n^β) satisfies the LAMN property with*

$$\Gamma \equiv \int_0^1 c_{ZZ,s}/c_{\varepsilon\varepsilon,s}ds. \tag{2.10}$$

Theorem 1 is proved by using the Malliavin calculus technique of Gobet (2001) and Clément et al. (2013), which transforms the log-likelihood ratio into a sum of conditional expectations and, hence, facilitates the derivation of the weak convergence required by (2.9).[3] We remark that the

[2]We are grateful to a referee for suggesting this argument.

[3]We note that the proof of Clément et al. (2013) concerns the LAMN property regarding F, which is different from our focus here.

results below only rely on the LAMN property, so their scope is likely broader than the sufficient condition given in Assumption L.

We now turn to the efficiency bound for estimating β. We focus on an efficiency notion in the sense of the conditional convolution theorem (Jeganathan (1982), Theorem 3) for regular estimates. A sequence $(\hat{\beta}_n)_{n\geq 1}$ of estimates is *regular* at $\beta = \beta_0$ if for any $h \in \mathbb{R}^{d-1}$, under the sequence $P_n^{\beta_0+n^{-1/2}h}$ of laws,

$$\left(n^{1/2}(\hat{\beta}_n - \beta_0 - n^{-1/2}h), \Gamma_n\right) \xrightarrow{\mathcal{L}} \mathbb{L}_0, \tag{2.11}$$

where the limit law $\mathbb{L}_0$ does not depend on h. Compared with the usual notion of regularity under settings with Locally Asymptotic Normality (LAN), (2.11) not only requires the convergence of the standardized estimator, but also that it converges jointly with the information matrix.

Theorem 1 and the conditional convolution theorem yield the following convolution representation for the limit distribution of regular estimates.

Corollary 1. *Suppose that (P_n^β) satisfies the LAMN property and $(\hat{\beta}_n)_{n\geq 1}$ is regular at $\beta = \beta_0$. Then under the sequence $P_n^{\beta_0}$ of laws, the limit distribution of $n^{1/2}(\hat{\beta}_n - \beta_0)$, conditional on Γ, can be represented as a convolution between a mixed Gaussian distribution $\mathcal{MN}\left(0, \Gamma^{-1}\right)$ and another transition kernel.*

Corollary 1 provides a natural notion of efficiency for estimating β. A regular estimator $\hat{\beta}_n$ is efficient in the parametric submodel $(P_n^\beta : \beta \in \mathbb{R}^{d-1})$ if the limit distribution of $n^{1/2}(\hat{\beta} - \beta_0)$ is $\mathcal{MN}(0, \Gamma^{-1})$. A sequence $\hat{\beta}_n$ of regular estimates is *adaptive* in the presence of nuisance components if it is efficient in the parametric submodel $(P_n^\beta; \beta \in \mathbb{R}^{d-1})$. Generally speaking, an adaptive estimator may not exist; however, when it exists, it may be the best one can hope for in the semiparametric setting since it achieves the parametric efficiency bound as if the nuisance components were known. In Section 4, we construct an adaptive estimator for β, among a class of regular estimators.

We remark that the efficiency notion above concerns regular estimators. In applications, the regularity of an estimator needs to be verified. Lemma 1 below is useful for this purpose. Henceforth, the symbol $\xrightarrow{\mathcal{L}\text{-}s}$ indicates $\mathcal{F}$-stable convergence in law; see Section VIII.5c in Jacod and Shiryaev (2003) for the definition and basic properties of stable convergence.

Lemma 1. *Suppose (i) (P_n^β) satisfies the LAMN property at $\beta = \beta_0$; (ii) for an $\mathcal{F}$-measurable positive semidefinite matrix Σ,*

$$\begin{pmatrix} n^{1/2}(\hat{\beta}_n - \beta_0) \\ \Gamma_n^{1/2}\zeta_n \end{pmatrix} \xrightarrow{\mathcal{L}\text{-}s} \mathcal{MN}\left(0, \begin{pmatrix} \Sigma & I_{d-1} \\ I_{d-1} & \Gamma \end{pmatrix}\right). \tag{2.12}$$

Then $\hat{\beta}_n$ is regular.

Finally, we note that the optimality of the limit distribution $\mathcal{MN}\left(0, \Gamma^{-1}\right)$ is not limited to regular estimators. For example, it can be justified among all estimators under a local minimaxity criterion (see Jeganathan (1983), Theorem 4). This alternative notion of optimality is convenient to apply since it does not require verifying regularity. The optimality of the adaptive estimator constructed below can also be interpreted in this sense.

3 Limit theorems for integrated volatility functionals

In this section, we present some generic results for estimating integrated volatility functionals of the form

$$S(g) \equiv \int_0^1 g(c_s)ds,$$

where g is a continuous test function. With g properly chosen, estimators for $S(g)$ are building blocks for estimators of β that we consider later in Section 4.1.

Since the results about the estimation of $S(g)$ are of independent interest and can be used in many other applications, we present them in the general setting of (2.1) and (2.2). That is, in this section we do not impose the regression relationship (2.3) or the "diffusion-like" structure (2.7).

The estimation of $S(g)$ is done via the general block-based method of Jacod and Rosenbaum (2013), see also Jacod and Protter (2012). Here we extend the result of Jacod and Rosenbaum (2013) to a wider class of functions g that covers many practically relevant ones, including ones that are used for our estimator of Section 4.1, and for which the results of Jacod and Rosenbaum (2013) do not apply.

The regularity conditions on X needed for the results of this section are collected below.

Assumption A. Let $r \in [0, 2)$ be a constant. The process X is an Itô semimartingale given by (2.1) and (2.2). The processes b_t and σ_t are locally bounded. There are a sequence of nonnegative bounded λ-integrable functions D_m on $\mathbb{R}$ and a sequence of stopping times $(\tau_m)_{m\geq1}$ increasing to ∞, such that $\|\delta\left(\omega,t,z\right)\|^r \wedge 1 \leq D_m\left(z\right)$ for all (ω, t, z) with $t \leq \tau_m\left(\omega\right)$.

Assumption B. The process $(\sigma_t)_{t\geq0}$ is an Itô semimartingale with the form

$$\begin{aligned}\sigma_t &= \sigma_0 + \int_0^t \tilde{b}_s ds + \int_0^t \tilde{\sigma}_s dW_s + \int_0^t \tilde{\sigma}'_s dW'_s \\ &\quad + \int_0^t \int_{\mathbb{R}} \tilde{\delta}\left(s,z\right)\left(\mu-\nu\right)\left(ds,dz\right),\end{aligned}$$

where $\tilde{b}$, $\tilde{\sigma}$ and $\tilde{\sigma}'$ are locally bounded and adapted processes; W' is a Brownian motion orthogonal to W; $\tilde{\delta}\left(\cdot\right)$ is a predictable function. Moreover, there are a sequence of nonnegative bounded λ-integrable functions $\tilde{D}_m$ on $\mathbb{R}$ and a sequence of stopping times $(\tau_m)_{m\geq1}$ increasing to ∞, such that $\|\tilde{\delta}\left(\omega,t,z\right)\|^2 \wedge 1 \leq \tilde{D}_m\left(z\right)$ for all (ω, t, z) with $t \leq \tau_m\left(\omega\right)$.

We now describe the estimators for $S(g)$ and their asymptotic properties. Following Jacod and Protter (2012) and Jacod and Rosenbaum (2013), we conduct estimation by first forming an approximation of the volatility trajectory and then constructing an estimator of $S(g)$ from it. To this end, we pick a sequence k_n of integers and a real sequence v_n that satisfy the following assumption.

Assumption C. For some $\gamma \in (0,1)$ and $\varpi \in (0,1/2)$, $k_n \asymp n^{\gamma}$ and $v_n \asymp n^{-\varpi}$.

The local approximation for the spot covariance over the time interval $[ik_n/n,(i+1)k_n/n)$, $i \in \mathcal{I}_n \equiv \{0,\ldots,[n/k_n]-1\}$, is given by

$$\hat{c}_i^n \equiv \frac{n}{k_n} \sum_{j=1}^{k_n} \left(\Delta_{ik_n+j}^n X\right) \left(\Delta_{ik_n+j}^n X\right)^{\intercal} 1_{\{\|\Delta_{ik_n+j}^n X\| \le v_n\}}. \tag{3.1}$$

Here, k_n is the smoothing parameter for local estimation and v_n specifies the truncation threshold (see Mancini (2001)) for "eliminating" jumps in X. We refer to Section 4.3 for guidance on the choice of these tuning parameters in applications. If X is known to be continuous, the truncation is not needed. The nonparametric estimation of spot volatility can be dated back to Foster and Nelson (1996); see also Kristensen (2010) and references therein.

To guide intuition, we note that $\hat{c}_i^n$ provides a uniform (for $i \in \mathcal{I}_n$) approximation to the moving average of c_t within the local window $[ik_n/n,(i+1)k_n/n)$:

$$\bar{c}_i^n \equiv \frac{n}{k_n} \int_{ik_n/n}^{(i+1)k_n/n} c_s ds.$$

Formally, we have the following lemma, which is crucial for our subsequent analysis.

Lemma 2. *Suppose (i) Assumption A for some $r \in [0,2)$; (ii) Assumption C with $\gamma > r/2$ and $\varpi \in [(1-\gamma)/(2-r),1/2)$. Then*

$$\sup_{i \in \mathcal{I}_n} \|\hat{c}_i^n - \bar{c}_i^n\| = o_p(1). \tag{3.2}$$

Remark 1. *As an immediate consequence of Lemma 2, we observe that the sequence $(\hat{c}_i^n)_{i \in \mathcal{I}_n}$ also uniformly approximates $(c_{ik_n/n})_{i \in \mathcal{I}_n}$ up to an $o_p(1)$ term when the process $(c_t)_{t \ge 0}$ is continuous; but when $(c_t)_{t\ge 0}$ contain jumps, the latter uniform approximation does not hold.*

We are now ready to describe the estimator for $S(g)$. A natural candidate is the sample analogue estimator

$$\tilde{S}_n(g) \equiv \frac{k_n}{n} \sum_{i \in \mathcal{I}_n} g(\hat{c}_i^n). \tag{3.3}$$

However, as shown in Jacod and Rosenbaum (2013), $\tilde{S}_n(g)$ does not enjoy a central limit theorem due to high-order bias terms. We refer to $\tilde{S}_n(g)$ as the *uncorrected* estimator. Jacod and Rosenbaum

(2013) proposed a *bias-corrected* estimator given by

$$\hat{S}_n(g) \equiv \frac{k_n}{n} \sum_{i \in \mathcal{I}_n} \left(g(\hat{c}_i^n) - \frac{1}{k_n} \mathbb{B}g(\hat{c}_i^n) \right), \tag{3.4}$$

where

$$\mathbb{B}g(c) \equiv \frac{1}{2} \sum_{j,k,l,m=1}^{d} \partial^2_{jk,lm} g(c) \left([c]_{jl} [c]_{km} + [c]_{jm} [c]_{kl} \right), \quad c \in \mathcal{M}_d, \tag{3.5}$$

and derived a stable convergence in law for $\hat{S}_n(g)$:

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

where

$$V(g) \equiv \sum_{j,k,l,m=1}^{d} \int_0^1 \partial_{jk} g(c_s) \, \partial_{lm} g(c_s)^{\intercal} \left([c_s]_{jl} [c_s]_{km} + [c_s]_{jm} [c_s]_{kl} \right) ds. \tag{3.6}$$

The result in Jacod and Rosenbaum (2013) is based on a polynomial growth condition on the test function g: for some $p \geq 3$ and $K > 0$,

$$\left\| \partial^j g(c) \right\| \leq K(1 + \|c\|^{p-j}), \text{ all } j \in \{0, 1, 2, 3\} \text{ and } c \in \mathcal{M}_d. \tag{3.7}$$

However, for the development of adaptive estimators in Section 4.1, it is necessary to consider test functions that do not satisfy the condition in the above display. Therefore, we extend the theory of Jacod and Rosenbaum (2013) to all $\mathcal{C}^3$ test functions. To do so, we replace condition (3.7) with a mild condition (Assumptions K or K') on the pathwise regularity of the process $(c_t)_{t \geq 0}$.[4] Below, for a compact set $\mathcal{K} \subset \mathcal{M}_d$ and $\varepsilon > 0$, we denote the "ε-enlargement" about $\mathcal{K}$ by

$$\mathcal{K}^{\varepsilon} \equiv \{ M \in \mathcal{M}_d : \inf_{A \in \mathcal{K}} \|M - A\| < \varepsilon \}.$$

Assumption K (resp. K'). There exist a localizing sequence of stopping times $(\tau_m)_{m \geq 1}$ and a sequence of convex compact subsets $\mathcal{K}_m \subseteq \mathcal{M}_d$ such that $c_t \in \mathcal{K}_m$ for $t \leq \tau_m$ and g is continuous (resp. $\mathcal{C}^3$) on $\mathcal{K}_m^{\varepsilon}$ for some $\varepsilon > 0$.

Remark 2. *A simple but useful property of Assumption K (resp. K') is its stability under smooth transformations: if two functions g_1 and g_2 satisfy Assumption K (resp. K') for $(\tau_{1,m}, \mathcal{K}_{1,m})_{m \geq 1}$ and $(\tau_{2,m}, \mathcal{K}_{2,m})_{m \geq 1}$ respectively, and $g(\cdot) = \psi(g_1(\cdot), g_2(\cdot))$ for some continuous (resp. $\mathcal{C}^3$) function ψ, then g satisfies Assumption K (resp. K') for $\tau_m \equiv \tau_{1,m} \wedge \tau_{2,m}$ and $\mathcal{K}_m \equiv \mathcal{K}_{1,m} \cap \mathcal{K}_{2,m}$.*

To guide intuition, we note that the convexity requirement on $\mathcal{K}_m$ in Assumptions K and K'

[4] We note that Jacod and Rosenbaum (2013) do not need Assumption K or K' for their result. Hence, the theory in the current paper is complementary to theirs.

assures that the moving average $\bar{c}_i^n$ also belongs to $\mathcal{K}_m$ up to the stopping time τ_m. Then in view of the uniform approximation (3.2), the spot covariance estimates $(\hat{c}_i^n)_{i\in\mathcal{I}_n}$ fall in any fixed ε-enlargement of $\mathcal{K}_m$ with probability approaching one. Hence, only the smoothness of g on $\mathcal{K}_m^\varepsilon$ is relevant in the derivation of the limit theorems.

Assumptions K and K' are easily verified in specific settings. For example, if $g(c) = \log(c)$ or $\sqrt{c}$, it holds provided that the processes c_t and $1/c_t$ are locally bounded, with $\mathcal{K}_m$ being a compact interval on $(0,\infty)$. We also observe that these transformations do not satisfy (3.7) because their derivatives are explosive near zero.

The main result of this subsection is the following theorem. We denote by $\mathcal{C}_c$ (resp. $\mathcal{C}_c^3$) the collection of continuous (resp. $\mathcal{C}^3$) functions with compact support.

Theorem 2. *Under (3.2), the following statements hold.*
(a) If $\tilde{S}_n(h) \xrightarrow{\mathbb{P}} S(h)$ for all $h \in \mathcal{C}_c$, then $\tilde{S}_n(g) \xrightarrow{\mathbb{P}} S(g)$ for g satisfying Assumption K.
(b) If $n^{1/2}(\hat{S}_n(h) - S(h)) \xrightarrow{\mathcal{L}\text{-}s} \mathcal{MN}(0, V(h))$ for all $h \in \mathcal{C}_c^3$, then

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

for g satisfying Assumption K'.

Theorem 2 is essentially a (spatial) localization procedure which allows one to assume without loss of generality that the test functions are compactly supported and, hence, verify the polynomial growth condition (3.7). We can use Theorem 2 to extend known limit theorems to a broader class of test functions. In this direction, Theorems 3 and 4 extend the scope of Theorem 9.4.1 in Jacod and Protter (2012) and Theorem 3.2 in Jacod and Rosenbaum (2013), respectively.

Theorem 3. *Suppose Assumptions A, C and K hold with $\gamma \in (r/2, 1)$ and $\varpi \in [(1-\gamma)/(2-r), 1/2)$. Then $\tilde{S}_n(g) \xrightarrow{\mathbb{P}} S(g)$.*

Theorem 4. *Suppose Assumptions A, B, C and K' hold with*

$$\frac{r}{2} \vee \frac{1}{3} < \gamma < \frac{1}{2}, \quad \frac{1-\gamma}{2-r} \le \varpi < \frac{1}{2}.$$

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

Finally, we remark that $\tilde{S}_n(g)$ and $\hat{S}_n(g)$ are formed using non-overlapping time blocks: $\hat{c}_i^n$ and $\hat{c}_j^n$ do not involve the same increment of X when $i \neq j$. Jacod and Protter (2012) and Jacod and Rosenbaum (2013) consider the use of overlapping time blocks and the latter shows that this alternative construction has no effect asymptotically.[5] Hence, we restrict attention only to the non-overlapping versions for brevity.

[5] That being said, estimators constructed using overlapping blocks and non-overlapping blocks may have different finite-sample performance; see, for example, Zu and Boswijk (2014) for simulation evidence concerning spot volatility estimation. A higher-order comparison between these approaches may be interesting for future research.

4 Adaptive estimation

In this section, using the limit theorems of the previous section, we develop adaptive estimation of β. Section 4.1 presents the adaptive estimator and its asymptotic properties. Section 4.2 provides an analytical efficiency comparison between the proposed estimator and some known alternatives. A numerical comparison and a Monte Carlo study of the different estimators are given in Section 4.3.

4.1 An adaptive estimator for β

We now return to the problem of estimating β, so (2.4) is in force. Consider the transform $g_b : \mathcal{M}_d \mapsto \mathbb{R}^{d-1}$ given by (recall (2.5))

$$g_b(c_t) \equiv c_{ZZ,t}^{-1} c_{ZY,t}. \tag{4.8}$$

Clearly, $g_b(c_t) = \beta$ identically under (2.4). It is natural to consider a class of estimators for β as the weighted average of $g_b(c_t)$ over $t \in [0,1]$. For a weight function $w : \mathcal{M}_d \mapsto \mathcal{M}_{d-1}$, we set

$$\hat{\beta}_n^w \equiv \hat{S}_n(w)^{-1} \hat{S}_n(w g_b). \tag{4.9}$$

By convention, $w(\cdot)$ takes values as symmetric matrices; but this is essentially not a restriction because the optimal weight function will be shown to be symmetric. We also assume the following.

Assumption W. The weight function w satisfies Assumption K' with g replaced by w.

The asymptotic behavior of $\hat{\beta}_n^w$ can be derived by using Theorem 4 and is summarized by Proposition 1 below, for which the following notations are needed. We set

$$\Xi\left(c_t\right) \equiv c_{ZZ,t}^{-1} c_{\varepsilon\varepsilon,t}.$$

The asymptotic covariance matrix of $\hat{\beta}_n^w$ is then given by

$$\Sigma\left(w\right) \equiv S(w)^{-1} S(w \Xi w) S(w)^{-1}.$$

We observe that $\Sigma(w)$ is minimized in the matrix sense at w^*, where

$$w^*(c_t) \equiv \Xi^{-1}(c_t) = c_{ZZ,t}/c_{\varepsilon\varepsilon,t}. \tag{4.10}$$

Indeed, it is easy to show that

$$\begin{aligned}\Sigma\left(w\right) - \Sigma\left(w^*\right) \ = \ & \int_0^1 \Big[\Big(S\left(w\right)^{-1} w(c_s) - S\left(w^*\right)^{-1} w^*(c_s)\Big) \\ & \qquad \Xi(c_s)\Big(w(c_s) S\left(w\right)^{-1} - w^*(c_s) S\left(w^*\right)^{-1}\Big)\Big]\, ds,\end{aligned}$$

which is positive semidefinite. We hence refer to w^* as the optimal weight function. For notational simplicity, we write $\hat{\beta}_n^*$ in place of $\hat{\beta}_n^{w^*}$.

Proposition 1. *Suppose*

(i) Assumptions A, B and C hold for $r \in [0, 1)$,

$$\frac{r}{2} \vee \frac{1}{3} < \gamma < \frac{1}{2}, \quad \frac{1-\gamma}{2-r} \leq \varpi < \frac{1}{2};$$

(ii) the process $\left(\lambda_{\min}^{-1}(c_t)\right)_{t \geq 0}$ is locally bounded.

Then the following statements hold.

(a) $n^{1/2}(\hat{\beta}_n^w - \beta_0) \xrightarrow{\mathcal{L}\text{-}s} \mathcal{MN}(0, \Sigma(w))$ for any weight function w satisfying Assumption W.

(b) $n^{1/2}(\hat{\beta}_n^* - \beta_0) \xrightarrow{\mathcal{L}\text{-}s} \mathcal{MN}\left(0, \Gamma^{-1}\right)$.

(c) $\tilde{S}_n(w^*)^{-1} \xrightarrow{\mathbb{P}} \Gamma^{-1}$.

Remark 3. *Under condition (ii) of Proposition 1, the test function g_b satisfies Assumption K' and w^* satisfies Assumption W. Indeed, let $(\tau_m)_{m \geq 1}$ be a sequence of stopping times increasing to $+\infty$ such that for constants $K_m > 0$, $\lambda_{\min}(c_t) \geq K_m^{-1}$ and $\|c_t\| \leq K_m$ on $\{t \leq \tau_m\}$. The set $\mathcal{K}_m = \{A \in \mathcal{M}_d : \lambda_{\min}(A) \geq K_m^{-1}, \|A\| \leq K_m\}$ is compact and convex. Observe that g_b and w^* are $\mathcal{C}^3$ on sufficiently small enlargement of $\mathcal{K}_m$.*

Proposition 1 shows that the $\mathcal{F}$-conditional asymptotic covariance matrix of $n^{1/2}(\hat{\beta}_n^* - \beta_0)$ is Γ^{-1}, the information bound in the submodel considered in Corollary 1. We stress that Proposition 1 is derived without assuming the additional structure (2.7) of Section 2.2 and is applicable under quite general settings. Moreover, Γ^{-1} can be consistently estimated by $\tilde{S}_n(w^*)^{-1}$ as shown by Proposition 1(c). Confidence sets for β can then be constructed using this estimator.

To further conclude that $\hat{\beta}_n^*$ is adaptive, it remains to verify that $\hat{\beta}_n^*$ is a regular estimator in the parametric submodel of Section 2.2. To this end, Theorem 5 below shows that the family $\hat{\beta}_n^w$ of estimators are regular.

Theorem 5. *Under Assumption L and conditions in Proposition 1, $\hat{\beta}_n^w$ is regular in the parametric submodel $(P_n^\beta; \beta \in \mathbb{R}^{d-1})$.*

Finally, we note some useful analytical results. Using the definitions (3.5) and (4.10), it is elementary (though somewhat tedious) to show that

$$\begin{cases} \mathbb{B}w^* = (d+1)w^*, \\ \mathbb{B}w^* g_b = (d+1)w^* g_b. \end{cases}$$

As a result, we observe the following relationship between bias-corrected and uncorrected estimators:

$$\begin{cases} \hat{S}_n(w^*) = (1 - (d+1)/k_n)\, \tilde{S}_n(w^*), \\ \hat{S}_n(w^* g_b) = (1 - (d+1)/k_n)\, \tilde{S}_n(w^* g_b). \end{cases} \tag{4.11}$$

Consequently,

$$\hat{\beta}_n^* = \tilde{S}_n(w^*)^{-1}\tilde{S}_n(w^* g_b).$$

That is, the adaptive estimator $\hat{\beta}_n^*$ is numerically equal to its uncorrected version

$$\tilde{S}_n(w^*)^{-1}\tilde{S}_n(w^* g_b) \tag{4.12}$$

formed using the uncorrected statistic $\tilde{S}_n$ (cf. (4.9)). Hence, in practice, one can conveniently compute the efficient estimator without implementing the bias correction, since the correction turns out to be automatic when the optimal weight function w^* is used.

We can draw a parallel between our optimally weighted $\widehat{\beta}_n^*$ and the efficient regression estimator of Robinson (1987) under heteroskedasticity of unknown form in the classical discrete setting. Indeed, in our setting without jumps and when we do not truncate the increments, $\widehat{\beta}_n^*$ is simply a weighted least squares estimator with weights being the inverses of the estimated variances of the residual increments over the blocks. This is analogous to the estimator of Robinson (1987). A key difference between our estimator and that of Robinson (1987) is in the estimates of the time-varying variance of the residuals.[6] This difference stems from the different asymptotic setups, infill in our case and long span in the case of Robinson (1987). In our infill asymptotic setup, we smooth locally in time the squared residual increments to form a consistent estimator of the residual variance. This is based on a very mild smoothness in expectation assumption for the volatility path (Assumption B) which is satisfied in most models of interest in economics. In the long span asymptotic setup, on the other hand, Robinson (1987) specifies the residual variance as an unknown function of the vector Z and performs smoothing of the squared residual increments as a function of Z.

4.2 Efficiency comparison

In this subsection, we provide some examples to examine the efficiency gain of $\hat{\beta}_n^*$ relative to alternative estimators. To simplify the discussion, we consider the case when β is a scalar (i.e., $d = 2$). We compare $\hat{\beta}_n^*$ with two estimators that have appeared in previous work:

$$\begin{cases} \hat{\beta}_n^1 = \hat{\beta}_n^{w_1}, & \text{where} \quad w_1(c_t) = c_{ZZ,t}, \\ \hat{\beta}_n^2 = \hat{\beta}_n^{w_2}, & \text{where} \quad w_2(c_t) = 1. \end{cases}$$

[6] Another key difference between our estimator and that of Robinson (1987) is that our regression is about the continuous martingale component of the vector X which is not directly observable due to the presence of drift and jumps in X. This necessitates, in particular, truncation to separate the jumps from the continuous part of the process, which has no obvious analogue in the discrete setting.

By Proposition 1(a), $n^{1/2}(\hat{\beta}_n^1 - \beta)$ and $n^{1/2}(\hat{\beta}_n^2 - \beta)$ are asymptotically mixed centered Gaussian with $\mathcal{F}$-conditional variance respectively given by

$$\begin{cases} \Sigma(w_1) = \int_0^1 c_{ZZ,s} c_{\varepsilon\varepsilon,s} ds \Big/ \left(\int_0^1 c_{ZZ,s} ds \right)^2, \\ \Sigma(w_2) = \int_0^1 \Xi(c_s)\, ds = \int_0^1 \frac{c_{\varepsilon\varepsilon,s}}{c_{ZZ,s}} ds. \end{cases}$$

The $\mathcal{F}$-conditional asymptotic variance of $\hat{\beta}_n^*$ is

$$\Sigma(w^*) = \left(\int_0^1 \Xi^{-1}(c_s)\, ds \right)^{-1} = \left(\int_0^1 \frac{c_{ZZ,s}}{c_{\varepsilon\varepsilon,s}} ds \right)^{-1}.$$

It is easy to compare these asymptotic variances using the Cauchy–Schwarz inequality. Indeed,

$$\sqrt{\frac{\Sigma(w_1)}{\Sigma(w^*)}} = \frac{\sqrt{\int_0^1 c_{ZZ,s} c_{\varepsilon\varepsilon,s} ds} \sqrt{\int_0^1 \frac{c_{ZZ,s}}{c_{\varepsilon\varepsilon,s}} ds}}{\int_0^1 \sqrt{c_{ZZ,s} c_{\varepsilon\varepsilon,s}} \sqrt{\frac{c_{ZZ,s}}{c_{\varepsilon\varepsilon,s}}} ds} \geq 1,$$

and the equality holds if and only if the process $c_{\varepsilon\varepsilon,t}$ is constant over time, that is, the residual process is homoskedastic. Furthermore, we have by Jensen's inequality $\Sigma(w_2) \geq \Sigma(w^*)$, which is also a simple restatement of the fact that the arithmetic average of the process $\Xi(c_t)$ is greater than its harmonic average. We have $\Sigma(w_2) = \Sigma(w^*)$ if and only if the process $\Xi(c_t) \equiv c_{\varepsilon\varepsilon,t}/c_{ZZ,t}$ is constant over time, that is, the volatility of the residual process perfectly comoves (in scale) with the volatility of the regressor. When $c_{\varepsilon\varepsilon,t}$ (resp. $c_{\varepsilon\varepsilon,t}/c_{ZZ,t}$) is time-varying, the efficiency gain of $\hat{\beta}_n^*$ relative to $\hat{\beta}_n^1$ (resp. $\hat{\beta}_n^2$) is strict.

Finally, we note that the three estimators $\hat{\beta}_n^*$, $\hat{\beta}_n^1$ and $\hat{\beta}_n^2$ have natural nonparametric interpretations when the regression relationship (2.3) is not imposed. In the general setting of (2.1) and (2.2), we can still use Theorems 3 and 4 to derive the asymptotics of these estimators. In particular, we have

$$\hat{\beta}_n^* \xrightarrow{\mathbb{P}} \frac{\int_0^1 \frac{c_{ZY,s}}{c_{\epsilon\epsilon,s}} ds}{\int_0^1 \frac{c_{ZZ,s}}{c_{\epsilon\epsilon,s}} ds}, \quad \hat{\beta}_n^1 \xrightarrow{\mathbb{P}} \frac{\int_0^1 c_{ZY,s} ds}{\int_0^1 c_{ZZ,s} ds}, \quad \hat{\beta}_n^2 \xrightarrow{\mathbb{P}} \int_0^1 \frac{c_{ZY,s}}{c_{ZZ,s}} ds.$$

We remark that $\hat{\beta}_n^1$ is, up to negligible boundary terms, a truncated version (in the sense of Mancini (2001)) of the "realized regression" estimator considered by Barndorff-Nielsen and Shephard (2004). Indeed, the asymptotic variance $\Sigma(w_1)$ coincides with that in Proposition 1 of Barndorff-Nielsen and Shephard (2004). The estimator $\hat{\beta}_n^2$ is simply $\hat{S}_n(g_b)$; hence, it is closely related to the integrated volatility estimators of Jacod and Rosenbaum (2013), although the test function g_b does not satisfy the polynomial growth condition (3.7) there.

4.3 Numerical examples

We now provide some numerical illustrations for the comparisons made in Section 4.2 using the following bivariate model for the continuous martingale components of Y and Z:

$$\begin{aligned} dY_t^c &= dZ_t^c + d\varepsilon_t, \quad dZ_t^c = \sigma_{Z,t} dW_{Z,t}, \quad d\varepsilon_t = \sigma_{\varepsilon,t} dW_{\varepsilon,t}, \\ \sigma_{Z,t} &= e^{V_{Z,t}/2-1/2}, \quad dV_{Z,t} = -0.1 V_{Z,t} dt + dL_{Z,t}, \\ \sigma_{\varepsilon,t} &= e^{V_{\varepsilon,t}/2-1/2}, \quad dV_{\varepsilon,t} = -0.1 V_{\varepsilon,t} dt + dL_{\varepsilon,t}, \end{aligned} \tag{4.13}$$

where $L_{Z,t}$ and $L_{\varepsilon,t}$ are two independent Lévy martingales uniquely defined by the marginal laws of $V_{Z,t}$ and $V_{\varepsilon,t}$ respectively, which in turn have self-decomposable distributions (see Theorem 17.4 of Sato (1999)), both with characteristic triplet (Definition 8.2 of Sato (1999)) of $(0, 1, \nu)$ for $\nu(dx) = \frac{4.837 e^{-3|x|}}{|x|^{1+0.5}} 1_{\{x>0\}} dx$ with respect to the identity truncation function. The volatility processes in (4.13) are exponentials of Lévy-driven Ornstein-Uhlenbeck processes. These volatility specifications are quite general, and in particular they allow for both diffusive and jump shocks in the two volatility processes. Such volatility models have been found to fit well financial asset price data; see, e.g., Todorov et al. (2014). We refer to Todorov et al. (2014) for analysis of the roles of the different parameters in these models. We choose the volatility model parameters such that $\mathbb{E}(e^{V_{Z,t}-1}) = \mathbb{E}(e^{V_{\varepsilon,t}-1}) = 1$ (our unit of time is a trading day and we measure returns in percentage) and the persistence of a shock in each of the volatility processes has a half-life of approximately 7 days. If Z corresponds to the market portfolio, then the above mean of volatility is close to the one observed in real data. On the other hand, the level of the idiosyncratic volatility risk will depend on the particular application, with the above choice being somewhat on the conservative side.

In Table 1, we report the relative efficiency of the three estimators $\widehat{\beta}_n^*$, $\widehat{\beta}_n^1$ and $\widehat{\beta}_n^2$ in the context of model (4.13) for various time spans of the regression, ranging from one week (equal to 5 business days) to 4 weeks (equal to 20 business days). We report the quantiles of the ratios $\Sigma(w_1)/\Sigma(w^*)$ and $\Sigma(w_2)/\Sigma(w^*)$ over 5000 Monte Carlo replications from the volatility specification in (4.13). We remind the reader that these ratios are random quantities, since the relative efficiency depends on the realization of the volatility processes. As seen from the table, the proposed estimator $\widehat{\beta}_n^*$ provides nontrivial improvements over the alternatives $\widehat{\beta}_n^1$ and $\widehat{\beta}_n^2$. The efficiency gain of $\widehat{\beta}_n^*$ is more significant when compared with $\widehat{\beta}_n^2$ and it also increases (with respect to both $\widehat{\beta}_n^1$ and $\widehat{\beta}_n^2$) with the length of the estimation horizon. The latter is fairly intuitive as longer estimation horizon allows for greater variations in the paths of $\sigma_{Z,t}$ and $\sigma_{\varepsilon,t}$ which in turn increases the efficiency gains from accounting for the heteroskedasticity in the data.

We next study the finite sample behavior of the three alternative estimates of β in a sampling setting that mimics that of a typical financial application such as the one performed in the next section. We continue to use the model (4.13), with $Y = Y^c$ and $Z = Z^c$ for simplicity. The sampling scheme in the Monte Carlo is as follows. We consider estimation horizons from one to four weeks,

Table 1: Relative efficiency of alternative β estimates in model (4.13)

Horizon	**Ratio** $\Sigma(w_1)/\Sigma(w^*)$			**Ratio** $\Sigma(w_2)/\Sigma(w^*)$		
	$Q_{0.25}$	$Q_{0.50}$	$Q_{0.75}$	$Q_{0.25}$	$Q_{0.50}$	$Q_{0.75}$
1 week	1.1172	1.2224	1.4670	1.2574	1.4677	1.9147
2 weeks	1.2157	1.4017	1.8091	1.5556	2.0204	3.1819
3 weeks	1.3118	1.5693	2.1526	1.8630	2.6589	4.5297
4 weeks	1.4015	1.7131	2.4620	2.2002	3.3555	6.0021

Note: The value of Q_α stands for the α quantile. Quantiles are computed based on 5000 *Monte Carlo replications.*

exactly as for the comparisons of the asymptotic variances in Table 1. With unit time being a trading day, we consider $\Delta_n = 1/40$ and $\Delta_n = 1/80$, corresponding approximately to 10-minute and 5-minute, respectively, sampling in a typical trading day.

The estimators of β depend on the two tuning parameters k_n and v_n, so we next explain how we set them. The requirement for the asymptotic order of k_n in Theorem 4 is relatively weak (at least when the jump activity r is not very high) which is suggestive that the estimation procedure is not very sensitive to the choice of k_n. In any application, the choice of k_n needs to be guided by the amount of data in hand. The requirement for k_n is to be large enough to expect a reasonable amount of averaging out of local volatility estimation error, but at the same time not so large as to generate significant biases. With this in mind, we set $k_n = 20$ for $\Delta_n = 1/40$ and $k_n = 30$ for $\Delta_n = 1/80$.

Turning next to the choice of the truncation parameter, we perform truncation componentwise.[7] Following existing work on truncation-based estimators, for a generic univariate process S, we set v_n as

$$v_n = 3\Delta_n^{0.49}\sqrt{BV_t^n}, \quad \text{for } i = (t-1)n+1, ..., tn \text{ and } t = 1, 2, ..., \tag{4.14}$$

where BV_t is the Bipower Variation of Barndorff-Nielsen and Shephard (2006) on day t which consistently estimates the integrated volatility on that day, and is defined as

$$BV_t^n = \frac{\pi}{2}\frac{n}{n-1}\sum_{i=(t-1)n+2}^{tn} |\Delta_i^n S||\Delta_{i-1}^n S|. \tag{4.15}$$

Intuitively, this choice of v_n "classifies" a high-frequency increment as one with a jump if it is above three standard deviations in absolute value. Importantly, this choice of v_n takes into account the time-varying volatility, as what is a "big" or a "small" move for the diffusive component of S depends on the current level of the diffusive volatility.

The results from the Monte Carlo are summarized in Table 2. The three alternative β estimates

[7]This is a slight departure from the way $\widehat{c}_i^n$ is defined but that difference is asymptotically negligible.

Table 2: Finite sample performance of alternative β estimates in model (4.13)

Horizon		$\widehat{\beta}_n^*$			$\widehat{\beta}_n^1$			$\widehat{\beta}_n^2$	
	$Q_{0.25}$	$Q_{0.50}$	$Q_{0.75}$	$Q_{0.25}$	$Q_{0.50}$	$Q_{0.75}$	$Q_{0.25}$	$Q_{0.50}$	$Q_{0.75}$
				Case $\Delta_n = 1/40,\ k_n = 20$					
1 week	0.9456	0.9927	1.0325	0.9338	0.9873	1.0309	0.9282	0.9884	1.0381
2 weeks	0.9588	0.9926	1.0164	0.9475	0.9873	1.0177	0.9399	0.9869	1.0253
3 weeks	0.9668	0.9927	1.0114	0.9550	0.9877	1.0133	0.9445	0.9868	1.0206
4 weeks	0.9711	0.9927	1.0085	0.9585	0.9874	1.0094	0.9473	0.9847	1.0174
				Case $\Delta_n = 1/80,\ k_n = 30$					
1 week	0.9574	0.9935	1.0188	0.9506	0.9894	1.0178	0.9468	0.9896	1.0229
2 weeks	0.9687	0.9932	1.0102	0.9587	0.9885	1.0104	0.9534	0.9878	1.0150
3 weeks	0.9741	0.9933	1.0061	0.9642	0.9880	1.0062	0.9567	0.9864	1.0111
4 weeks	0.9774	0.9937	1.0038	0.9659	0.9882	1.0032	0.9596	0.9863	1.0085

Note: The true value of β is 1. The value of Q_α stands for the α quantile. Quantiles are computed based on 5000 *Monte Carlo replications.*

are slightly downward biased, with the bias being the smallest for our optimal estimator $\widehat{\beta}_n^*$ and shrinking (in all but one case) with the increase of the sampling frequency. Consistent with the asymptotic comparisons reported in Table 1, the finite sample precision of $\widehat{\beta}_n^*$ is highest among the three estimators.

Overall the results in Tables 1 and 2 confirm the efficiency gains of our new estimation procedure in realistic settings.

5 Empirical example

We apply the efficient procedure developed in Section 4 to high frequency data on U.S. Bancorp (USB) with the S&P 500 Index ETF (SPY) used as the market. In terms of capitalization, U.S. Bancorp ranks near the top of the stocks comprising the S&P sector ETF for Financials (XLF), and the results reported here are representative of financial stocks. The span of the data set is 2007–2014, and we consider three sampling frequencies: 3-min, 5-min, and 10-min. These frequencies allow us to circumvent all the issues associated with microstructure noise and thereby focus on issues pertinent to the paper.

Following the discussion in Section 4.3, we use $k_n = 32$ as the base value for 3-min data (128 returns per day), which corresponds to four daily sub-periods, early morning, late morning, early afternoon, and late afternoon; we use $k_n = 25$ as the base value for 5-min data (77 returns per day), i.e., morning, mid-day, and afternoon, and $k_n = 19$ for 10-min data (38 returns per day) for

Table 3: Summary of $\hat{\beta}_n^*$ estimates for USB on SPY

	Empirical Quantiles					
	$k_n =$ base values			$k_n =$ modified values		
Frequency	$Q_{0.25}$	$Q_{0.50}$	$Q_{0.75}$	$Q_{0.25}$	$Q_{0.50}$	$Q_{0.75}$
3-min	0.882	1.065	1.197	0.875	1.049	1.177
5-min	0.890	1.068	1.204	0.897	1.067	1.204
10-min	0.888	1.067	1.252	0.864	1.068	1.259

Notes: We estimate β for each calendar month during 2007–2014. *The base settings of k_n are* 32, 25, 19 *for sampling frequencies 3-min, 5-min, and 10-min, and the modified settings of k_n are* 27, 20, 14 *for the three sampling frequencies. The value of Q_α stands for the α quantile. Quantiles are computed over* 96 *monthly observations on $\hat{\beta}_n^*$ characterized in Proposition 1* above.

morning and afternoon. We set the truncation parameter v_n exactly as in Section 4.3.[8] Inspection of plots (not shown) revealed considerable month-to-month variation in β over the sample period, which is to be expected given the turbulent times for financial stocks. At the same time, plots and previous experience suggested that β is reasonably constant over a month, so we conduct estimation monthly over the 96 calendar months of the data set.

Our two points of focus here are robustness and efficiency. Table 3 shows the quantiles of the $\hat{\beta}_n^*$ of Proposition 1 computed over months. From the table it is seen that for the base values of k_n the median values of the estimator are quite insensitive to the sampling frequency, while as to be expected the interquantile ranges are slightly wider at coarser sampling frequencies. The table also shows quantiles of the efficient estimator for modified values of k_n, and little sensitivity is evident. As regards efficiency, the following display shows the quantiles (over 96 months) of the ratios of the estimated asymptotic standard errors of the (inefficient) estimators $\hat{\beta}_n^1$ and $\hat{\beta}_n^2$ discussed above to those of $\hat{\beta}_n^*$:

Ratio $\widehat{\Sigma}(w_1)/\widehat{\Sigma}(w^*)$			**Ratio** $\widehat{\Sigma}(w_2)/\widehat{\Sigma}(w^*)$		
$Q_{0.25}$	$Q_{0.50}$	$Q_{0.75}$	$Q_{0.25}$	$Q_{0.50}$	$Q_{0.75}$
1.5163	1.6571	1.8207	1.4960	1.5962	1.7650

The above empirically-implied efficiency ratios are in line with those expected from the Monte Carlo work reported in Table 1. The empirical results taken together with the theory and Monte Carlo suggest that the new estimator of this paper should prove useful for applied financial econometrics.

[8]We also perform a time-of-day adjustment to the truncation level as in Todorov and Tauchen (2012) to account for the well-known diurnal pattern in volatility.

6 Conclusion

In a LAMN setting, we derive the semiparametric efficiency bound for estimating the slope coefficient in a linear regression model for continuous-time Itô semimartingales sampled at asymptotically increasing frequencies. We construct an adaptive estimator which achieves this efficiency bound. This estimator is in closed form and easy to compute. We illustrate analytically and numerically the efficiency gain of the proposed efficient estimator relative to existing alternatives. To analyze the asymptotic behavior of the proposed estimator, we derive a general limit theory for the estimation of integrated volatility transforms, which extends known results in the literature (Jacod and Rosenbaum (2013)) to a larger class of volatility functionals.

7 Appendix: Proofs

This section contains all proofs. Below, K denotes a generic constant that may change from line to line but does not depend on $i \in \mathcal{I}_n$ or $t \in [0,1]$; we sometimes write K_u to indicate the dependence of this constant on some parameter u. For any matrix A, we denote its ith row (resp. column) by $[A]_{i\cdot}$ (resp. $[A]_{\cdot i}$) and its trace by $\mathrm{Tr}(A)$. For any vector a, we denote its ith element by $a^{(i)}$. For $p \geq 1$, $\|\cdot\|_p$ denotes the L_p-norm.

7.1 Proofs in Section 2

Proof of Theorem 1. Let $P_n^{\beta,f}$ be the conditional distribution of $(X_{i/n})_{0\leq i\leq n}$ given $F = f$. Since the distribution of F does not depend on β, it suffices to derive the LAMN property for the conditional law $P_n^{\beta,f}$. Therefore, we shall conditional on $F = f$ below.

The proof follows similar steps as in Section A.1 of Clément et al. (2013), which use the Malliavin calculus technique developed in Gobet (2001). We only sketch the main steps, but with the difference detailed.

Denote

$$b(x) = \begin{pmatrix} b_Z(x) \\ b_Y(x) \end{pmatrix}, \quad a(\beta,x,f) = \begin{pmatrix} a_Z(x,f) & 0_{d-1} \\ \beta^\intercal a_Z(x,f) & a_\varepsilon(x,f) \end{pmatrix}. \tag{7.1}$$

For $1 \leq j \leq d-1$, let $\dot{a}_j(\beta,x,f)$ denote the derivative of $a(\beta,x,f)$ with respect to $\beta^{(j)}$. Evidently,

$$\dot{a}_j(\beta,x,f) = \begin{pmatrix} \mathbf{0}_{d-1} & 0_{d-1} \\ [a_Z(x,f)]_{j\cdot} & 0 \end{pmatrix}. \tag{7.2}$$

For $h \in \mathbb{R}^{d-1}$, we denote the log-likelihood ratio by

$$L_n(h) = \log dP_n^{\beta_0+h/\sqrt{n}}/dP_n^{\beta_0}. \tag{7.3}$$

We now define Γ_n and ζ_n in (2.8). For $0 \leq i \leq n-1$ and $1 \leq j, l \leq d-1$, we evaluate the functions a^{-1}, $\dot{a}_j$ and $\dot{a}_l$ at $(\beta_0, X_{i/n}, f_{i/n})$ and set

$$\begin{aligned} [\Gamma_n]_{jl} &\equiv \frac{1}{n}\sum_{i=0}^{n-1} \mathrm{Tr}\left[\left(a^{-1}\dot{a}_l\right)\left(a^{-1}\dot{a}_j\right)^\intercal + \left(a^{-1}\dot{a}_j\right)\left(a^{-1}\dot{a}_l\right)\right], \\ \tilde{\zeta}^{(j)}_{n,i+1} &\equiv \Delta^n_{i+1}W^\intercal\left(a^{-1}\dot{a}_j\right)\Delta^n_{i+1}W - n^{-1}\mathrm{Tr}\left(a^{-1}\dot{a}_j\right), \\ \tilde{\zeta}^{(j)}_n &\equiv n^{1/2}\sum_{i=0}^{n-1}\tilde{\zeta}^{(j)}_{n,i+1}, \end{aligned}$$

and

$$\zeta_n \equiv \Gamma_n^{-1/2}\tilde{\zeta}_n. \tag{7.4}$$

It is easy to see that $(\tilde{\zeta}_{n,i}, \mathcal{F}_{i/n})$ forms an array of martingale differences. Moreover, following the same steps that lead to eq. (54) in Clément et al. (2013) (by setting $h(\cdot)$ there to be a constant, but with an extension to allow for multivariate h), we deduce

$$\left\{ \begin{aligned} L_n(h) = h^\intercal\tilde{\zeta}_n - \frac{1}{2}\sum_{i=0}^{n-1}\sum_{j,l=1}^{d} h^{(j)}h^{(l)}\Delta^n_{i+1}W^\intercal\Big[&\left(a^{-1}\dot{a}_l\right)\left(a^{-1}\dot{a}_j\right)^\intercal \\ &+ \left(a^{-1}\dot{a}_j\right)\left(a^{-1}\dot{a}_l\right)\Big]\Delta^n_{i+1}W + o_p(1). \end{aligned} \right. \tag{7.5}$$

We now simplify these expressions by computing $a^{-1}\dot{a}_j$ explicitly. Some straightforward algebra yields

$$\left(a^{-1}\dot{a}_j\right)(\beta_0, X_{i/n}, f_{i/n}) = \begin{pmatrix} \mathbf{0}_{d-1} & 0_{d-1} \\ \left[\sigma_{Z,i/n}\right]_{j\cdot}/\sigma_{\varepsilon,i/n} & 0 \end{pmatrix}.$$

Hence,

$$\tilde{\zeta}^{(j)}_{n,i+1} = \frac{1}{\sigma_{\varepsilon,i/n}}\Delta^n_{i+1}W_\varepsilon\left[\sigma_{Z,i/n}\right]_{j\cdot}\Delta^n_{i+1}W_Z. \tag{7.6}$$

From here, we observe

$$\begin{gathered} \mathbb{E}\left[\tilde{\zeta}^{(j)}_{n,i+1}\tilde{\zeta}^{(l)}_{n,i+1}\middle|\mathcal{F}_{i/n}\right] = n^{-2}\left[c_{ZZ,i/n}\right]_{jl}/c_{\varepsilon\varepsilon,i/n}, \\ \mathbb{E}\left[\tilde{\zeta}^{(j)}_{n,i+1}\Delta^n_{i+1}W\middle|\mathcal{F}_{i/n}\right] = 0_d, \quad \mathbb{E}\left[(\tilde{\zeta}^{(j)}_{n,i+1})^4\middle|\mathcal{F}_{i/n}\right] \leq Kn^{-4}. \end{gathered}$$

By Theorem IX.7.28 of Jacod and Shiryaev (2003), we derive (recall that $\Gamma \equiv \int_0^1 c_{ZZ,s}/c_{\varepsilon\varepsilon,s}ds$)

$$\tilde{\zeta}_n \xrightarrow{\mathcal{L}\text{-}s} \mathcal{MN}\left(0_{d-1}, \Gamma\right). \tag{7.7}$$

For the second term on the right-hand side of (7.5), we first observe

$$\begin{aligned}&\mathbb{E}\left[\Delta_{i+1}^n W^\intercal\left[\left(a^{-1}\dot{a}_l\right)\left(a^{-1}\dot{a}_j\right)^\intercal+\left(a^{-1}\dot{a}_j\right)\left(a^{-1}\dot{a}_l\right)\right]\Delta_{i+1}^n W|\mathcal{F}_{i/n}\right]\\&\quad=\frac{1}{n}\mathrm{Tr}\left[\left(a^{-1}\dot{a}_l\right)\left(a^{-1}\dot{a}_j\right)^\intercal+\left(a^{-1}\dot{a}_j\right)\left(a^{-1}\dot{a}_l\right)\right]=\frac{\left[c_{ZZ,i/n}\right]_{jl}}{nc_{\varepsilon\varepsilon,i/n}}.\end{aligned}$$

By using a law of large numbers, we deduce that

$$\text{the second term on the right-hand side of (7.5)} \;=-h^\intercal\Gamma h/2+o_p(1). \tag{7.8}$$

Moreover, by a Riemann approximation, we also observe

$$\Gamma_n=\Gamma+o_p(1). \tag{7.9}$$

By the properties of stable convergence, (7.7), (7.8) and (7.9) hold jointly in the usual sense of weak convergence. The assertion of the theorem then follows from (7.4) and (7.5). □

Proof of Lemma 1. By the properties of stable convergence, (2.12) also holds jointly with $\Gamma_n \xrightarrow{\mathbb{P}} \Gamma$. Recall the notation $L_n(h)$ from (7.3). Under the conditions of Lemma 1, along the sequence $P_n^{\beta_0}$,

$$\begin{pmatrix} n^{1/2}(\hat{\beta}_n-\beta_0)\\ L_n(h)\end{pmatrix}\xrightarrow{\mathcal{L}\text{-}s}\mathcal{MN}\left(\begin{pmatrix}0\\ -\frac{1}{2}h^\intercal\Gamma h\end{pmatrix},\begin{pmatrix}\Sigma & h\\ h^\intercal & h^\intercal\Gamma h\end{pmatrix}\right).$$

By Le Cam's third lemma, we have, along the sequence $P_n^{\beta_0+n^{-1/2}h}$,

$$n^{1/2}(\hat{\beta}_n-\beta_0)\xrightarrow{\mathcal{L}\text{-}s}\mathcal{MN}\left(h,\Sigma\right).$$

Therefore, along the sequence $P_n^{\beta_0+n^{-1/2}h}$,

$$n^{1/2}(\hat{\beta}_n-\beta_0-n^{-1/2}h)\xrightarrow{\mathcal{L}\text{-}s}\mathcal{MN}\left(0,\Sigma\right). \tag{7.10}$$

Furthermore, the LAMN property implies contiguity between $P_n^{\beta_0+n^{-1/2}h}$ and $P_n^{\beta_0}$ (Jeganathan (1982), Proposition 1). Hence, Γ_n converges to Γ under $P_n^{\beta_0+n^{-1/2}h}$ too. By the properties of stable convergence, (7.10) holds jointly with the convergence of Γ_n towards Γ, where the limit distribution does not depend on h. Hence, $\hat{\beta}_n$ is regular. □

7.2 Proofs in Section 4

By a standard localization argument (Jacod and Protter (2012), Section 4.4.1), we can replace Assumption A with the following stronger version without loss of generality.

Assumption SA. We have Assumption A. The processes $(b_t)_{t\geq 0}$ and $(\sigma_t)_{t\geq 0}$ are bounded. Moreover, for a λ-integrable function D, $\|\delta(\omega,t,z)\|^r \leq D(z)$ for all $\omega\in\Omega$, $t\in[0,1]$ and $z\in\mathbb{R}$.

Proof of Lemma 2. By localization, we assume that Assumption SA holds. By a polarization argument, we can assume that X is $\mathbb{R}$-valued without loss of generality. We denote by X' the continuous part of X, that is,

$$X'_t = x_0 + \int_0^t b_s ds + \int_0^t \sigma_s dW_s. \tag{7.11}$$

We then set

$$\hat{c}_i'^n \equiv \frac{n}{k_n}\sum_{j=1}^{k_n}\left(\Delta_{ik_n+j}^n X'\right)^2.$$

By (4.8) of Jacod and Rosenbaum (2013), there exists a sequence a_n of constants such that $a_n \to 0$ and

$$\left\|\hat{c}_i^n - \hat{c}_i'^n\right\|_1 \leq K a_n n^{-(2-r)\varpi}.$$

By using a maximal inequality, we then derive

$$\begin{aligned}\left\|\sup_{i\in\mathcal{I}_n}\left|\hat{c}_i^n - \hat{c}_i'^n\right|\right\|_1 &\leq K a_n\,(n/k_n)\,n^{-(2-r)\varpi} \\ &\leq K a_n n^{1-\gamma-(2-r)\varpi} \to 0,\end{aligned} \tag{7.12}$$

where the convergence follows from condition (ii).

By Itô's formula,

$$\hat{c}_i'^n - \bar{c}_i^n = \frac{2n}{k_n}\sum_{j=1}^{k_n}\int_{(ik_n+j-1)/n}^{(ik_n+j)/n}\left(X'_s - X'_{(ik_n+j-1)/n}\right)\left(b_s ds + \sigma_s dW_s\right).$$

From here, standard estimates for continuous Itô semimartingales yield, for every $\tilde{q}\geq 1$,

$$\left\|\hat{c}_i'^n - \bar{c}_i^n\right\|_{\tilde{q}} \leq K_{\tilde{q}} k_n^{-1/2}.$$

By using a maximal inequality and picking $\tilde{q} > 2(1-\gamma)/\gamma$, we deduce

$$\left\|\sup_{i\in\mathcal{I}_n}\left|\hat{c}_i'^n - \bar{c}_i^n\right|\right\|_{\tilde{q}} \leq K_{\tilde{q}}\,(n/k_n)^{1/\tilde{q}}\,k_n^{-1/2} \to 0. \tag{7.13}$$

The assertion of the lemma then follows from (7.12) and (7.13). $\square$

Proof of Theorem 2. We only prove part (b) while noting that the proof for part (a) is similar and slightly simpler. Let τ_m and $\mathcal{K}_m$ be given as in Assumption K'. We denote the closure of a set A by $\overline{A}$. By Assumption K', $c_t \in \mathcal{K}_m$ for all $t\in[0,\tau_m]$. By a standard localization argument with respect to the localizing sequence $(\tau_m)_{m\geq 1}$, we can assume that $c_t \in \mathcal{K}$ for some convex compact set $\mathcal{K}$ and all $t \in [0,1]$ without loss of generality. Moreover, there exists some $\varepsilon > 0$ such that g is $\mathcal{C}^3$ on $\mathcal{K}^\varepsilon$. Fix an arbitrary constant $\eta \in (0, \varepsilon)$. Observe that $\overline{\mathcal{K}^\eta}$ is a compact subset of $\mathcal{K}^\varepsilon$. Let $\psi : \mathcal{M}_d \mapsto [0,1]$ be a compactly supported $\mathcal{C}^\infty$ function such that $\psi(c) = 1$ when $c \in \overline{\mathcal{K}^\eta}$ and $\psi(c) = 0$ when $c \notin \mathcal{K}^\varepsilon$; the existence of such ψ is due to the $\mathcal{C}^\infty$ Urysohn lemma (Folland (1999), Theorem 8.18). We then set $h = \psi g$, so $h \in \mathcal{C}_c^3$. By assumption of Theorem 2, $n^{-1/2}(\hat{S}_n(h) - S(h)) \xrightarrow{\mathcal{L}\text{-}s} \mathcal{MN}(0, V(h))$.

Consider a sequence Ω_n of events given by $\Omega_n \equiv \{\hat{c}_i^n \in \mathcal{K}^\eta, \text{ all } i \in \mathcal{I}_n\}$. Since $\mathcal{K}$ is convex, $\{\bar{c}_i^n : i \in \mathcal{I}_n\} \subseteq \mathcal{K}$. Under the assumption $\sup_{i\in\mathcal{I}_n} \|\hat{c}_i^n - \bar{c}_i^n\| = o_p(1)$, we deduce that $\mathbb{P}(\Omega_n) \to 1$. Note that $\partial^j g(c) = \partial^j h(c)$ for all $j = 0, 1, 2$ and $c \in \mathcal{K}^\eta$. Therefore, $n^{-1/2}(\hat{S}_n(h) - S(h))$ and $V(h)$ respectively coincide with $n^{-1/2}(\hat{S}_n(g) - S(g))$ and $V(g)$ on Ω_n. Hence, in restriction to Ω_n, $n^{-1/2}(\hat{S}_n(g) - S(g)) \xrightarrow{\mathcal{L}\text{-}s} \mathcal{MN}(0, V(g))$. Since $\mathbb{P}(\Omega_n) \to 1$, the assertion of the theorem readily follows. $\square$

Proof of Theorem 3. By Lemma 2, (3.2) holds. Then by Theorem 2(a), we can assume that g has compact support without loss of generality. The setting is now the same as Theorem 9.4.1(b) in Jacod and Protter (2012). The proof of Jacod and Protter (2012) can be easily adapted to the current setting with non-overlapping time blocks, and yields $\tilde{S}_n(g) \xrightarrow{\mathbb{P}} S(g)$. $\square$

Proof of Theorem 4. By Lemma 2 and Theorem 2(b), we can assume that $g \in \mathcal{C}_c^3$ without loss of generality. We can then use Theorem 3.2 of Jacod and Rosenbaum (2013) to deduce the asserted convergence. Since g is compactly supported, the condition of Jacod and Rosenbaum (2013) can be weakened as $\varpi \in [\frac{1}{2(2-r)}, \frac{1}{2})$ after a straightforward adaptation of their proof. This condition on ϖ is implied by the condition in Theorem 4. $\square$

Proof of Proposition 1. Observe that the mapping $g \mapsto \hat{S}_n(g)$ is linear. Hence,

$$n^{1/2}\left(\hat{\beta}_n^w - \beta_0\right) = \hat{S}_n(w)^{-1} n^{1/2} \hat{S}_n(g), \quad \text{where} \quad g \equiv w g_b - w\beta_0. \tag{7.14}$$

Under (2.4), $S(g) = 0$. As noted in Remark 3, g_b satisfies Assumption K'. Since w satisfies Assumption W, g satisfies Assumption K'; see Remark 2. By Theorem 4,

$$n^{1/2}\left(\hat{\beta}_n^w - \beta_0\right) \xrightarrow{\mathcal{L}\text{-}s} \mathcal{MN}\left(0, S(w)^{-1} V(g) S(w)^{-1}\right).$$

Some straightforward (but somewhat tedious) calculation using (3.6) yields $V(g) = S(w\Xi w)$. From here, the assertion in part (a) readily follows.

As noted in Remark 3, w^* satisfies Assumption W under condition (ii). Further observe that $\Sigma(w^*) = S(w^*)^{-1} = \Gamma^{-1}$. Part (b) then follows from part (a).

Finally, note that Assumption K is satisfied with w^* in place of g. The assertion of part (c) then follows directly from Theorem 3. $\square$

Proof of Theorem 5. Step 1. In this step, we recall some known results from Jacod and Rosenbaum (2013) and derive some preliminary estimates. Let h be a $\mathbb{R}$-valued $\mathcal{C}^3$ function with compact support. Let ∂h denote the $d \times d$ matrix such that $[\partial h]_{jk} = \partial_{jk} h$. It is shown in Jacod and Rosenbaum (2013) that (see p. 1482)

$$\begin{cases} n^{1/2}\left(\hat{S}_n(h) - S(h)\right) = \\ \qquad n^{1/2} \sum_{i=0}^{[n/k_n]-1} \sum_{u=1}^{k_n} \sum_{l,m=1}^{d} \left[\partial h\left(c_{ik_n/n}\right)\right]_{lm} \left[\alpha_{ik_n+u}^n\right]_{lm} + o_p(1), \end{cases} \tag{7.15}$$

where, for X' defined by (7.11),

$$\alpha_i^n \equiv \left(\Delta_i^n X'\right)\left(\Delta_i^n X'\right)^{\intercal} - n^{-1} c_{(i-1)/n}.$$

Denote

$$\alpha_i'^n \equiv \mathbb{E}\left[\alpha_i^n | \mathcal{F}_{(i-1)/n}\right], \quad \alpha_i''^n \equiv \alpha_i^n - \alpha_i'^n.$$

By (4.10) in Jacod and Rosenbaum (2013), for any $p \geq 0$,

$$\mathbb{E}[\, \|\alpha_i^n\|^p \, |\mathcal{F}_{(i-1)/n}] \leq K n^{-p}, \quad \left\|\alpha_i'^n\right\| \leq K n^{-3/2}. \tag{7.16}$$

Note that for any $1 \leq u \leq k_n$,

$$\mathbb{E}\left\|\partial h\left(c_{ik_n/n}\right) - \partial h\left(c_{(ik_n+u-1)/n}\right)\right\|^2 \leq K k_n/n. \tag{7.17}$$

Then by the second inequality of (7.16), it is easy to see

$$\begin{cases} n^{1/2} \sum_{i=0}^{[n/k_n]-1} \sum_{u=1}^{k_n} \left[\partial h\left(c_{ik_n/n}\right) - \partial h\left(c_{(ik_n+u-1)/n}\right)\right]_{lm} \left[\alpha_{ik_n+u}'^n\right]_{lm} \\ \qquad = O_p(\sqrt{k_n/n}). \end{cases} \tag{7.18}$$

Observe that $\partial h\left(c_{ik_n/n}\right) - \partial h\left(c_{(ik_n+u-1)/n}\right)$ is $\mathcal{F}_{(ik_n+u-1)/n}$-measurable and $\left(\alpha_i''^n, \mathcal{F}_{i/n}\right)$ is an array of martingale differences. We then use the first inequality of (7.16) to derive

$$\begin{aligned} &\mathbb{E}\left[\left(n^{1/2} \sum_{i=0}^{[n/k_n]-1} \sum_{u=1}^{k_n} \left[\partial h\left(c_{ik_n/n}\right) - \partial h\left(c_{(ik_n+u-1)/n}\right)\right]_{lm} \left[\alpha_{ik_n+u}''^n\right]_{lm}\right)^2\right] \\ &\quad = n \sum_{i=0}^{[n/k_n]-1} \sum_{u=1}^{k_n} \mathbb{E}\left[\left(\left[\partial h\left(c_{ik_n/n}\right) - \partial h\left(c_{(ik_n+u-1)/n}\right)\right]_{lm} \left[\alpha_{ik_n+u}''^n\right]_{lm}\right)^2\right] \\ &\quad \leq K k_n/n. \end{aligned} \tag{7.19}$$

Since $k_n/n \to 0$, the terms in (7.18) and (7.19) are $o_p(1)$. Hence, (7.15) further implies

$$\begin{aligned}
& n^{1/2}\left(\hat{S}_n(h) - S(h)\right) \\
&= n^{1/2} \sum_{i=0}^{[n/k_n]k_n - 1} \sum_{l,m=1}^{d} \left[\partial h\left(c_{i/n}\right)\right]_{lm} \left[\alpha_{i+1}^n\right]_{lm} + o_p(1) \\
&= n^{1/2} \sum_{i=0}^{n-1} \sum_{l,m=1}^{d} \left[\partial h\left(c_{i/n}\right)\right]_{lm} \left[\alpha_{i+1}^n\right]_{lm} + o_p(1), \\
&= n^{1/2} \sum_{i=0}^{n-1} \mathrm{Tr}\left[\partial h\left(c_{i/n}\right) \alpha_{i+1}^n\right] + o_p(1) \\
&= n^{1/2} \sum_{i=0}^{n-1} \left(\left(\Delta_{i+1}^n X'\right)^\intercal \partial h\left(c_{i/n}\right) \Delta_{i+1}^n X' - n^{-1}\mathrm{Tr}\left[\partial h\left(c_{i/n}\right) c_{i/n}\right]\right) \\
&\quad + o_p(1),
\end{aligned}$$

where the second equality is obtained by using (7.16) and $k_n/n^{1/2} \to 0$. By routine manipulation using Itô calculus, we can further approximate $\Delta_{i+1}^n X'$ with $\sigma_{i/n}\Delta_{i+1}^n W$, yielding

$$\left\{\begin{aligned}
& n^{1/2}\left(\hat{S}_n(h) - S(h)\right) \\
& \qquad = n^{1/2} \sum_{i=0}^{n-1} \left(\xi(h)_{n,i+1} - \mathbb{E}\left[\xi(h)_{n,i+1} \,|\mathcal{F}_{i/n}\right]\right) + o_p(1),
\end{aligned}\right. \tag{7.20}$$

where

$$\left\{\begin{aligned}
& \xi(h)_{n,i+1} \equiv \Delta_{i+1}^n W^\intercal \sigma_{i/n}^\intercal \partial h\left(c_{i/n}\right) \sigma_{i/n} \Delta_{i+1}^n W, \\
& \mathbb{E}\left[\xi(h)_{n,i+1} \,|\mathcal{F}_{i/n}\right] = n^{-1}\mathrm{Tr}\left[\partial h\left(c_{i/n}\right) c_{i/n}\right].
\end{aligned}\right.$$

Step 2. In this step, we derive an asymptotic linear representation for $n^{1/2}(\hat{\beta}_n^w - \beta_0)$. Let $g(\cdot) \equiv w(\cdot)\left(g_b(\cdot) - \beta_0\right)$. Following the argument as in the proof of Theorem 2, we can find a $\mathcal{C}_c^3$ function h, a compact set $\mathcal{K}$ and $\eta > 0$ such that (i) $\{c_t : t \in [0,1]\} \subseteq \mathcal{K}$ (ii) $\{\hat{c}_i^n : i \in \mathcal{I}_n\} \subseteq \mathcal{K}^\eta$ with probability approaching one (iii) $h(\cdot)$ coincides with $g(\cdot)$ on $\mathcal{K}^\eta$.

Since h is $\mathcal{C}^3$ with compact support, by applying (7.20) to each component of h (recall that the kth component is denoted by $h^{(k)}$), we obtain the same representation but now with $\xi(h)_{n,i+1}$ being a $(d-1)$-vector with its kth element given by

$$\xi(h)_{n,i+1}^{(k)} = \Delta_{i+1}^n W^\intercal \sigma_{i/n}^\intercal \partial h^{(k)}\left(c_{i/n}\right) \sigma_{i/n} \Delta_{i+1}^n W. \tag{7.21}$$

Since $h(\cdot)$ and $g(\cdot)$ coincide on $\mathcal{K}^\eta$ and $S(g) = 0$, we also have

$$n^{1/2}\hat{S}_n(g) = n^{1/2} \sum_{i=0}^{n-1} \left(\xi(g)_{n,i+1} - \mathbb{E}\left[\xi(g)_{n,i+1} \,|\mathcal{F}_{i/n}\right]\right) + o_p(1). \tag{7.22}$$

Now, we compute $\partial g^{(k)}(c_{i/n})$ explicitly. Note that, under (2.4), $g_b(c_{i/n}) = \beta_0$. Hence, for $1 \leq l, m \leq d$, $\partial_{lm} g^{(k)}\left(c_{i/n}\right) = \left[w(c_{i/n})\right]_{k\cdot} \partial_{lm} g_b\left(c_{i/n}\right)$. Some elementary calculation yields

$$\partial g^{(k)}\left(c_{i/n}\right) = \begin{pmatrix} -\left[c_{ZZ,i/n}^{-1} w\left(c_{i/n}\right)\right]_{\cdot k} \beta_0^\intercal & \left[c_{ZZ}^{-1} w\left(c_{i/n}\right)\right]_{\cdot k} \\ 0_{d-1}^\intercal & 0 \end{pmatrix}.$$

Plugging this into (7.21), we deduce

$$\xi\left(g\right)_{n,i+1}^{(k)} = \Delta_{i+1}^n W_Z^\intercal \sigma_{Z,i/n}^\intercal \left[c_{ZZ,i/n}^{-1} w\left(c_{i/n}\right)\right]_{\cdot k} \sigma_{\varepsilon,i/n} \Delta_{i+1}^n W_\varepsilon. \tag{7.23}$$

It is then easy to see that $\mathbb{E}[\xi\left(g\right)_{n,i+1} | \mathcal{F}_{i/n}] = 0_{d-1}$. From (7.14) and (7.22), we deduce

$$n^{1/2}(\hat{\beta}_n^w - \beta_0) = S(w)^{-1} n^{1/2} \sum_{i=0}^{n-1} \xi\left(g\right)_{n,i+1} + o_p(1). \tag{7.24}$$

Step 3. We complete the proof of Theorem 5 by using Lemma 1. It suffices to show (2.12). The marginal stable convergence of $\Gamma_n^{1/2}\zeta_n$ and $n^{1/2}(\hat{\beta}_n^w - \beta_0)$ has been shown in Theorem 1 and Proposition 1. The joint convergence can be similarly derived by using Theorem IX.7.28 of Jacod and Shiryaev (2003). The additional step needed here is to verify that the asymptotic covariance between $\Gamma_n^{1/2}\zeta_n$ and $n^{1/2}(\hat{\beta}_n^w - \beta_0)$ is I_{d-1}. In view of (7.24), it suffices to verify that the asymptotic covariance between $\Gamma_n^{1/2}\zeta_n$ and $n^{1/2}\sum_{i=0}^{n-1} \xi\left(g\right)_{n,i+1}$ is $S\left(w\right)$. Recall from the proof of Theorem 1 that

$$\begin{aligned} \Gamma_n^{1/2}\zeta_n &= \tilde{\zeta}_n = n^{1/2} \sum_{i=0}^{n-1} \tilde{\zeta}_{n,i+1}, \\ \tilde{\zeta}_{n,i+1}^{(j)} &= \frac{1}{\sigma_{\varepsilon,i/n}} \Delta_{i+1}^n W_\varepsilon \left[\sigma_{Z,i/n}\right]_{j\cdot} \Delta_{i+1}^n W_Z. \end{aligned}$$

Then, from (7.23), for $1 \leq j, k \leq d-1$,

$$\begin{aligned} & n \sum_{i=0}^{n-1} \mathbb{E}\left[\tilde{\zeta}_{n,i+1}^{(j)} \xi\left(g\right)_{n,i+1}^{(k)} \middle| \mathcal{F}_{i/n} \right] \\ = & \frac{1}{n} \sum_{i=0}^{n-1} \left[\sigma_{Z,i/n}\right]_{j\cdot} \sigma_{Z,i/n}^\intercal \left[c_{ZZ,i/n}^{-1} w\left(c_{i/n}\right)\right]_{\cdot k} \\ = & \frac{1}{n} \sum_{i=0}^{n-1} \left[w\left(c_{i/n}\right)\right]_{jk} \to \left[S\left(w\right)\right]_{jk}, \end{aligned}$$

as wanted. The proof is now complete. $\square$

References

Aït-Sahalia, Y. and D. Xiu (2015). Principal Component Analysis of High Frequency Data. Technical report, Princeton University.

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 (2004). Econometric Analysis of Realized Covariation: High Frequency Based Covariance, Regression, and Correlation in Financial Economics. *Econometrica 72*(3), pp. 885–925.

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

Bibinger, M., N. Hautsch, P. Malec, and M. Reiss (2014). Estimating the Quadratic Covariation Matrix from Noisy Observations: Local Method of Moments and Efficiency. *Annals of Statistics 42*, 1312–1346.

Bickel, P. J. (1982). On adaptive estimation. *The Annals of Statistics 10*(3), 647–671.

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

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

Folland, G. B. (1999). *Real Analysis: Modern Techniques and Their Applications* (2nd ed.). Wiley.

Foster, D. and D. B. Nelson (1996). Continuous record asymptotics for rolling sample variance estimators. *Econometrica 64*, 139–174.

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

Gobet, E. (2001). Local Asymptotic Mixed Normality Property for Elliptic Diffusion: A Malliavin Calculus Approach. *Bernoulli 7*(6), pp. 899–912.

Hayashi, T. and N. Yoshida (2011). Nonsynchronous Covariation Process and Limit Theorems. *Stochastic Processes and Their Applications 121*, 2416–2454.

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

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

Jacod, J. and A. N. Shiryaev (2003). *Limit Theorems for Stochastic Processes* (Second ed.). New York: Springer-Verlag.

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

Jeganathan, P. (1983). Some Asymptotic Properties of Risk Functions when the Limit of the Experiment is Mixed Normal. *Sankhya: The Indian Journal of Statistics, Series A (1961-2002) 45*(1), pp. 66–87.

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

Kalnina, I. and D. Xiu (2014). Nonparametric Estimation of the Leverage Effect using Information from Derivatives Markets. Technical report, University of Chicago.

Kristensen, D. (2010). Nonparametric filtering of the realized spot volatility: A kernel-based approach. *Econometric Theory 26*(1), pp. 60–93.

Li, J., V. Todorov, and G. Tauchen (2015). Inference Theory for Volatility Functional Dependencies. Technical report, Duke University and Northwestern University.

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

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.

Reiss, M., V. Todorov, and G. Tauchen (2015). Nonparametric Test for a Constant Beta between Ito Semimartingales based on High-Frequency Data. *Stochastic Processes and their Applications 125*, 2955–2988.

Renault, E., C. Sarisoy, and B. J. Werker (2015). Efficient estimation of integrated volatility and related processes. *Econometric Theory, Forthcoming.*

Robinson, P. M. (1987). Asymptotically efficient estimation in the presence of heteroskedasticity of unknown form. *Econometrica 55*(4), 875–891.

Sato, K. (1999). *Lévy Processes and Infinitely Divisible Distributions.* Cambridge, UK: Cambridge University Press.

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.

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

Todorov, V., G. Tauchen, and I. Grynkiv (2014). Volatility Activity: Specification and Estimation. *Journal of Econometrics 178*, 180–193.

Zu, Y. and P. Boswijk (2014). Estimating spot volatility with high-frequency financial data. *Journal of Econometrics 181*, 117–135.