3-2024

Optimal inference for spot regressions

Tim BOLLERSLEV
Duke University

Jia LI
Singapore Management University, jiali@smu.edu.sg

Yuexuan REN
Singapore Management University, yuexuan.ren.2020@phdecons.smu.edu.sg

Citation

BOLLERSLEV, Tim; LI, Jia; and REN, Yuexuan. Optimal inference for spot regressions. (2024). *American Economic Review*. 114, (3), 678-708.
Available at: https://ink.library.smu.edu.sg/soe_research/2645

Optimal Inference for Spot Regressions*

Tim Bollerslev† Jia Li‡ Yuexuan Ren§

May 27, 2023

Abstract

Betas from return regressions are commonly used to measure systematic financial market risks. "Good" beta measurements are essential for a range of empirical inquiries in finance and macroeconomics. We introduce a novel econometric framework for the nonparametric estimation of time-varying betas with high-frequency data. The "local Gaussian" property of the generic continuous-time benchmark model enables optimal "finite-sample" inference in a well-defined sense. It also affords more reliable inference in empirically realistic settings compared to conventional large-sample approaches. Two applications pertaining to the tracking performance of leveraged ETFs and an intraday event study illustrate the practical usefulness of the new procedures.

Keywords: beta, high-frequency data, optimal estimation, leveraged ETFs, event study.
JEL Codes: C14, C22, C58, G14.

*We would like to thank participants at various conferences and seminars for their useful suggestions. Jia Li's research is supported by Singapore Ministry of Education tier-1 grant (No. 22-SOE-SMU-016).

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

‡School of Economics, Singapore Management University, Singapore; e-mail: jiali@smu.edu.sg.

§School of Economics, Singapore Management University, Singapore; e-mail: yuexuan.ren.2020@phdecons.smu.edu.sg.

1 Introduction

Time series return regressions, in which the returns on some financial asset are regressed on some benchmark factor, are omnipresent in macroeconomics and finance. The regression coefficient estimates, or betas, derived from such regressions are commonly used in empirical applications as intrinsic measures for systematic market risks. The practical estimation of said regressions have traditionally been based on daily or lower-frequency monthly returns over multi-year horizons. However, macroeconomic and financial market risks are clearly time-varying (e.g., Engle (2004), Sims and Zha (2006)), and the assumption of constancy over extended timeframes implicit in the standard regression-based approach for estimating betas, has been called into question by numerous studies. As a case in point, a number of researchers have allowed the betas to depend parametrically on other variables (e.g., Ferson and Harvey (1996) and Gagliardini et al. (2016)), while others have inferred time-varying betas from nonlinear parametric models that explicitly allow for second-order dynamic dependencies (e.g., Bollerslev et al. (1988) and Braun et al. (1995)).

In response to these concerns, a more recent and rapidly growing literature, starting with Barndorff-Nielsen and Shephard (2004a) and Andersen et al. (2005, 2006), has sought to nonparametrically estimate so-called realized betas based on higher-frequency intraday data over shorter, typically daily, time intervals.[1] Related, a burgeoning literature in macroeconomics has argued for the usefulness of high-frequency-based identification procedures and related regressions to assess the effect of policy shocks in changing economic environments (e.g., Cochrane and Piazzesi (2002), Faust et al. (2004), Nakamura and Steinsson (2018)). Further pushing the estimation window to even finer within-day time-intervals, a series of newer studies (e.g., Bibinger and Reiß (2014), Bibinger et al. (2019), Andersen et al. (2021)) have argued that the betas may also vary non-trivially within the day, so that even daily estimation windows might be too coarse. Meanwhile, the assumption of a "large" number of observations over diminishing, or "local," intraday estimation windows underpinning the inference procedures in all of these studies can be difficult to realistically mimic in practice (see also the introductory discussion in Andersen and Bollerslev (2018)).

Hence, motivated by the need for more reliable estimation and inference procedures in high-frequency settings with time-varying betas, we propose a new fixed-k inference method for spot

[1] Precedent theory pertaining to the consistent filtering of spot betas has been developed by Nelson (1996).

return regressions that formally treats the window size k, or equivalently the number of observations underlying the nonparametric estimation, as a fixed constant. As such, our approach forces the econometrician to directly address the "small-sample" issue head-on without resorting to conventional large-sample asymptotic arguments, and the fictitious notion of an infinitely growing number of observations, or bandwidth sequence. Our approach builds on recent work by Bollerslev et al. (2021), and the observation that in a frictionless financial market without any arbitrage opportunities, the underlying latent continuous-time price process (formally modeled as an Itô semimartingale) should behave approximately as a scaled Brownian motion over "short" time intervals, and therefore locally also inherit the Gaussian property of a Brownian motion.[2]

This local Gaussian approximation in turn allows us to treat the nonparametric spot regression problem as if it were a finite-sample Gaussian linear regression. Correspondingly, our fixed-k theory naturally suggests the use of critical values based on the t-distribution, instead of the Gaussian-based critical values obtained from conventional large-k asymptotic theory. Importantly, going one step further, we show that the spot beta estimator and related t-test are both asymptotically optimal, in the sense of achieving uniformly minimum-variance among all unbiased estimators, and being uniformly most powerful among all unbiased tests, respectively. The former optimality claim, justifying our fixed-k least-squares estimator as the best unbiased estimator (BUE), in effect represents a new nonparametric "spot" version of the classic Gauss–Markov theorem.

To address concerns that the theoretically justified local Gaussianity underlying the fixed-k approach may not provide a sufficiently close approximation for a given "local" sample, we develop a new diagnostic test based on the spot skewness and kurtosis calculated from the high-frequency returns. Our Monte Carlo simulation study, geared to various empirically realistic settings, also further underscores the superior performance of the new fixed-k inference procedures compared to the conventional large-k methods. Notably, a simple-to-implement "robustified" version of the approach, which explicitly incorporates the aforementioned diagnostic test for local Gaussianity as a safe guard, maintains excellent size control across a wide range of estimation window sizes, including situations with rapidly changing volatility.

Underscoring the practical applicability and insights afforded by the new procedures, we present two separate empirical applications. In the first application, we investigate the tracking perfor-

[2]Shephard (2022) has also recently developed a new class of volatility estimators that exploit the quantile properties of Brownian increments, as an interesting alternative to the conventional high-frequency moment-based approach.

mance of two leveraged exchange-traded funds (ETFs), namely the popular TQQQ and SQQQ funds which aim to generate three and negative three times the return on the Nasdaq-100 stock market index, respectively. Using high-frequency intraday data spanning 2018 to 2022, our analysis uncovers significant systematic intraday and episodic deviations in the funds' estimated spot betas from the funds' stated tracking objectives, with especially notable deviations occurring during the COVID-19 pandemic in 2020. As such, our results call into question the efficacy of the two funds for practical hedging and risk management purposes during tumultuous times, exactly when they would be needed the most.

In our second application, we showcase the use of the new spot-regression-based inference in the context of an intraday event study, detailing the price impact of Microsoft's announcement of its new AI tool, Copilot, on March 16, 2023. By estimating the spot beta for Microsoft within a short pre-event estimation window, we are able to reliably assess the performance of Microsoft's stock in relation to a benchmark market model at a high intraday frequency. Doing so, our results reveal economically large and highly statistically significant abnormal positive returns during the live demonstration of the new AI system, with little evidence of any pre-announcement drift. At a broader level, this analysis clearly illustrates the rapid speed with which financial markets respond to certain economic news, and as such the practical usefulness of the new fixed-k inference tools for properly analyzing the impact of news shocks more generally.

The rest of the paper is organized as follows. Section 2 formally presents the new fixed-k spot regression inference procedure. Additional theoretical results, including the optimality of the spot estimators and a uniform functional inference theory, are provided in Section 3. This section is more technically involved, and it can be skipped by readers primarily interested in practical applications. Section 4 summarizes the results of our Monte Carlo simulation study. Our empirical applications related to the tracking performance of leveraged ETFs and the impact of Microsoft's AI announcement are presented in Section 5. Section 6 concludes the paper. All proofs are included in the Online Supplemental Appendix, which also provides additional theoretical results briefly mentioned in the main text, along with various empirical robustness checks.

2 Fixed-k Inference for Spot Regressions

Section 2.1 details the formal econometric setting. Section 2.2 establishes our new fixed-k inference theory for nonparametric spot regressions. Section 2.3 presents a diagnostic test that can be used to guide the choice of spot estimation window and further robustify the inference procedure. Section 2.4 describes a predictive inference procedure for intraday event studies. Below, we use $\xrightarrow{\mathbb{P}}$ and $\xrightarrow{d}$ to denote convergence in probability and convergence in distribution, respectively. All limits are for the number of observations $n \to \infty$, along with the sampling interval of the high-frequency data shrinking to zero $\Delta_n \to 0$, and the discretely sampled data series approaching their continuous-time population limit.

2.1 Formal setting and notation

We consider a bivariate jump-diffusion, or Itô semimartingale, price process $\boldsymbol{Z}_t = (X_t, Y_t)^\top$ defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ written as:

$$\boldsymbol{Z}_t = \boldsymbol{Z}_0 + \int_0^t \boldsymbol{b}_s ds + \int_0^t \boldsymbol{\sigma}_s d\boldsymbol{W}_s + \boldsymbol{J}_t, \tag{2.1}$$

where the drift process $\boldsymbol{b}$ and the stochastic volatility matrix process $\boldsymbol{\sigma}$ are both càdlàg adapted, $\boldsymbol{W} = (W_{1,t}, W_{2,t})_{t\geq 0}$ is a bivariate standard Brownian motion, and $\boldsymbol{J}$ denotes a pure-jump process with finite activity driven by a homogeneous Poisson random measure on $\mathbb{R}_+ \times \mathbb{R}$. For ease of notation, we will refer to the Brownian component of $\boldsymbol{Z}_t$ (i.e., $\int_0^t \boldsymbol{\sigma}_s d\boldsymbol{W}_s$) as $\boldsymbol{Z}_t^c = (X_t^c, Y_t^c)^\top$, and the individual elements of the spot covariance matrix process by:

$$\boldsymbol{c}_t \equiv \boldsymbol{\sigma}_t \boldsymbol{\sigma}_t^\top = \begin{pmatrix} c_{11,t} & c_{12,t} \\ c_{21,t} & c_{22,t} \end{pmatrix}. \tag{2.2}$$

The semimartingale model, inspired by the no-arbitrage principle (Delbaen and Schachermayer (1994)), serves as the cornerstone model in continuous-time finance and economics; comprehensive textbook treatments are available in Merton (1992), Duffie (2010), and Back (2017). In the broader realm of stochastic integration theory, the semimartingale process is also recognized as the most general "reasonable" stochastic differential possible, as detailed in Protter (2005). Due to its versatility and applicability, the semimartingale model, together with the infill asymptotic setting that we adopt here, have also emerged as the standard framework for analyzing high-frequency

data and financial asset prices in particular; see, e.g., the discussions in Jacod and Protter (2012) and Aït-Sahalia and Jacod (2014).[3]

We further specialize the semimartingale model in the form of the following regression representation:

$$\begin{aligned} dX_t^c &= v_t^{1/2} dW_{1,t}, \\ dY_t^c &= \beta_t dX_t^c + \varsigma_t^{1/2} dW_{2,t}. \end{aligned} \tag{2.3}$$

Our objective is to draw inferences about β_t, or the "spot beta" which represents the instantaneous regression coefficient that quantifies how Y moves together with X at a specific time t. The spot regression representation in (2.3) is equivalent to imposing a lower-triangular structure on the $\boldsymbol{\sigma}_t$ matrix:

$$\boldsymbol{\sigma}_t = \begin{pmatrix} v_t^{1/2} & 0 \\ \beta_t v_t^{1/2} & \varsigma_t^{1/2} \end{pmatrix}. \tag{2.4}$$

This structure effectively assigns the interpretation of the shock to the regressor process X_t^c to $dW_{1,t}$, and $dW_{2,t}$ as the idiosyncratic shock specific to Y_t^c. As a result, the spot covariance matrix can be expressed as:

$$\boldsymbol{c}_t = \begin{pmatrix} v_t & \beta_t v_t \\ \beta_t v_t & \beta_t^2 v_t + \varsigma_t \end{pmatrix}. \tag{2.5}$$

This representation in turn demonstrates how the spot beta β_t and the idiosyncratic variance ς_t can be directly identified from $\boldsymbol{c}_t$:

$$\beta_t = \frac{c_{12,t}}{c_{11,t}}, \qquad \varsigma_t = c_{22,t} - \frac{c_{12,t}^2}{c_{11,t}}. \tag{2.6}$$

The nonparametric "reduced-form" estimation for the spot covariance matrix $\boldsymbol{c}_t$ is well-known under the now standard infill asymptotic setting. To fix ideas, suppose that the price vector process $\boldsymbol{Z}$ is observed at discrete times $i\Delta_n$, for $i = 0, 1, \ldots, n$, within the fixed time interval $[0, T]$, where $T \equiv n \cdot \Delta_n$. Let the ith increment (i.e., return) of $\boldsymbol{Z}$ be denoted by

$$\Delta_i^n \boldsymbol{Z} \equiv \boldsymbol{Z}_{i\Delta_n} - \boldsymbol{Z}_{(i-1)\Delta_n}, \quad i \in \{1, \ldots, n\}.$$

[3]That said, it is important to acknowledge that the semimartingale model is clearly misspecified empirically at ultra-high frequencies due to a host of difficult to quantify market microstructure frictions; see, e.g., the discussions in Bandi and Russell (2008) and Diebold and Strasser (2013). Thus, in line with standard practice, we only treat the Itô semimartingale model as a reasonable approximation to reality for "not too finely" sampled data. As such, our approach should only be used with data sampled at a "moderate" high frequency, say on a time scale of minutes, as opposed to ultra-high-frequency data, say on a time scale of seconds.

The nonparametric spot estimation is then based on high-frequency returns from a local window indexed by $i \in \mathcal{I}_{n,t} \equiv \{\lceil t/\Delta_n \rceil + 1, \ldots, \lceil t/\Delta_n \rceil + k_n\}$, where $\lceil \cdot \rceil$ denotes the ceiling function. Correspondingly, the spot estimator for $\boldsymbol{c}_t$ is simply constructed as

$$\hat{\boldsymbol{c}}_t \equiv \frac{1}{k_n \Delta_n} \sum_{i \in \mathcal{I}_{n,t}} (\Delta_i^n \boldsymbol{Z}) (\Delta_i^n \boldsymbol{Z})^\top , \tag{2.7}$$

with the corresponding estimators for the spot quantities of interest naturally defined by:

$$\hat{v}_t \equiv \hat{c}_{11,t}, \qquad \hat{\beta}_t \equiv \frac{\hat{c}_{12,t}}{\hat{c}_{11,t}}, \qquad \hat{\varsigma}_t \equiv \hat{c}_{22,t} - \frac{\hat{c}_{12,t}^2}{\hat{c}_{11,t}}. \tag{2.8}$$

Since Poisson jumps occur at a vanishing probability, all of these spot estimates are also formally robust to the presence of Poisson-type jumps.[4]

The estimator defined in (2.7) represents a natural multivariate extension of the univariate spot variance estimator considered by Foster and Nelson (1996); see also Comte and Renault (1998) and Kristensen (2010). It may also be interpreted as a localized version of the realized covariance matrix estimator and the corresponding realized beta proposed by Barndorff-Nielsen and Shephard (2004a), as they generate exactly the same estimate over the "short" time interval $[t, t + k_n\Delta_n]$. The window size k_n plays an analogous role to that of the bandwidth parameter in traditional kernel-based nonparametrics. Under the conventional spot estimation theory, $\hat{\boldsymbol{c}}_t$ consistently estimates $\boldsymbol{c}_t$, provided that the bandwidth sequence satisfies $k_n \to \infty$ and $k_n\Delta_n \to 0$. The $k_n \to \infty$ condition permits the use of a law of large numbers to establish consistency, while the $k_n\Delta_n \to 0$ condition ensures that the local estimation window "collapses" to the time point t so as to justify the "spot" interpretation. Under an additional "undersmoothing" condition (i.e., $k_n \to \infty$ "sufficiently slowly"), a feasible central limit theorem for $\hat{\boldsymbol{c}}_t$ may also be established (see, e.g., Theorem 13.3.3 in Jacod and Protter (2012)), which combined with the delta method implies

[4]This phenomenon is well understood in the literature on spot estimation; see, for example, Theorem 13.3.3 in Jacod and Protter (2012). To further improve the estimator's robustness against other types of jumps, it would be possible to rely on alternative multipower (Barndorff-Nielsen and Shephard (2004b) and Barndorff-Nielsen et al. (2006)) or truncation-type (Mancini (2001)) realized measures. Alternative estimators explicitly designed to accommodate "noisy" and irregularly spaced observations have also been developed (see, e.g., Hayashi and Yoshida (2005), Aït-Sahalia et al. (2010), and Mykland and Zhang (2009)). We purposely do not pursue any of these extensions here, so as to highlight the main novelty of our new approach with minimal technicalities, particularly as it pertains to the optimality analysis.

the following feasible asymptotic distributional approximations for the spot quantities of interest:

$$\frac{k_n^{1/2}\left(\hat{v}_t - v_t\right)}{\sqrt{2}\hat{v}_t} \xrightarrow{d} \mathcal{N}\left(0,1\right), \quad \frac{k_n^{1/2}(\hat{\beta}_t - \beta_t)}{\sqrt{\hat{\varsigma}_t/\hat{v}_t}} \xrightarrow{d} \mathcal{N}\left(0,1\right), \quad \frac{k_n^{1/2}\left(\hat{\varsigma}_t - \varsigma_t\right)}{\sqrt{2}\hat{\varsigma}_t} \xrightarrow{d} \mathcal{N}\left(0,1\right). \tag{2.9}$$

These conventional asymptotic-Gaussian-based approximations have long served as the leading approach for spot regression inference, the vast literature on the analysis of high-frequency financial data included.

That being said, it is evident that the asymptotic Gaussian approximations in (2.9) might work poorly when the number of observations used in the estimation is not "large enough" (i.e., for small values of k_n).[5] On the other hand, the choice of a longer estimation window (i.e., larger values of k_n) might induce nonparametric biases in the estimation stemming from potentially non-negligible temporal variation in the true latent continuous-time processes (as a case in point, the aforementioned recent study by Andersen et al. (2021) points to substantial intraday variation in the betas for many individual stocks). These considerations directly motivate our development of new and more reliable inference procedures intended to circumvent this delicate balancing act by explicitly treating the window size (i.e., k_n) as a fixed constant k. Importantly, this framework also allows for a more detailed "approximate finite-sample" decision-theoretic-based characterization of the optimal spot estimators.

2.2 Fixed-k inference for spot beta

The fixed-k asymptotic theory that we will develop for spot regressions is based on the simple idea that the Itô semimartingale model is *locally Gaussian*. Intuitively, within a narrow estimation window (i.e., $k\Delta_n \to 0$, as implied by $\Delta_n \to 0$ and a fixed k), the drift component is of a smaller order, jumps occur with vanishing probability, and the volatility process is nearly constant. As a result, the Itô semimartingale behaves approximately as a scaled Brownian motion, and as such exhibits the aforementioned local Gaussianity.

The local Gaussian approximation that we rely on is directly linked to the Gaussian nature of Brownian motion. However, that same approximation could in principle also be "micro-founded"

[5] When k_n is small, the sampling uncertainty is of similar order as the signal, and the asymptotic Gaussian approximation for the "reduced-form" estimator $\hat{\boldsymbol{c}}_t$ may be poor. Moreover, since β_t and ς_t are nonlinear transformations of the spot covariance matrix $\boldsymbol{c}_t$, the additional linear approximation underlying the delta method may further worsen the asymptotic approximation. A similar phenomenon has been studied extensively in the context of weak instrumental variables; see, for example, Staiger and Stock (1997) and Andrews et al. (2019).

in the sense of a functional central limit theorem type argument. Putting this notion into a broader historical perspective, Bachelier (1900) first suggested using Brownian motion to model asset prices, considering the continuous-time process as the scaling limit of a microscopic tick-by-tick random walk. The physical interpretation of the Gaussian property of Brownian motion was also discussed in Einstein's (1905) classic work. For additional discussion along these lines and more general results regarding semimartingales, see also Aït-Sahalia and Jacod (2020) and the many references therein.

Given the local Gaussianity mentioned above, the nonparametric spot regression described in (2.3) may similarly be approximated by a "limit" Gaussian linear model. Finite-sample results for this parametric problem can then be translated into the original nonparametric context. The requisite regularity conditions are gathered in the following set of assumptions, which are standard for nonparametric high-frequency econometrics; for a more detailed discussion of these conditions, see Jacod and Protter (2012), Jacod et al. (2021), Bollerslev et al. (2021), and Li et al. (2023).

Assumption 1. *Suppose that the process $\boldsymbol{Z}$ satisfies (2.1) and (2.4), and that there exists a sequence $(T_m)_{m\geq 1}$ of stopping times increasing to infinity and a sequence $(K_m)_{m\geq 1}$ of constants such that the following conditions hold for each $m \geq 1$: (i) $\|\boldsymbol{b}_t\|+\|\boldsymbol{\sigma}_t\|+v_t^{-1}+\varsigma_t^{-1}+F_t(\mathbb{R}\backslash\{0\}) \leq K_m$ for all $t \in [0, T_m]$, where F_t denotes the spot Lévy measure of $\boldsymbol{J}$; and (ii) for some constant $\kappa > 0$, $\mathbb{E}[\|\boldsymbol{\sigma}_{t\wedge T_m} - \boldsymbol{\sigma}_{s\wedge T_m}\|^2] \leq K_m|t-s|^{2\kappa}$ for all $t, s \in [0, T]$.*

Theorem 1, below, shows that under these assumptions the studentized spot beta estimator may be strongly approximated, or "coupled," by a student-t variable, defined by the following quantities:

$$\xi_{11} \equiv \sum_{i\in\mathcal{I}_{n,t}} \frac{(\Delta_i^n W_1)^2}{\Delta_n}, \quad \xi_{12} \equiv \sum_{i\in\mathcal{I}_{n,t}} \frac{(\Delta_i^n W_1)(\Delta_i^n W_2)}{\Delta_n}, \quad \xi_{22} \equiv \sum_{i\in\mathcal{I}_{n,t}} \frac{(\Delta_i^n W_2)^2}{\Delta_n}. \tag{2.10}$$

Theorem 1. *For $k \geq 2$, and under Assumption 1,*

$$\frac{\sqrt{k-1}\,(\hat{\beta}_t - \beta_t)}{\sqrt{\hat{\varsigma}_t/\hat{v}_t}} = \xi_\beta + o_p(1), \tag{2.11}$$

where $\xi_\beta \equiv \sqrt{k-1}\,\xi_{12}/\sqrt{\xi_{11}\xi_{22}-\xi_{12}^2}$ is t-distributed with $k-1$ degrees of freedom.

This infill asymptotic coupling result for the nonparametric Itô semimartingale model closely resembles a classical finite-sample result for Gaussian linear regressions, as in, e.g., Proposition 1.3

in Hayashi (2011).[6] The $o_p(1)$ term in (2.11) accounts for biases arising from the drift and jump components, as well as biases related to temporal variation in the stochastic volatility process.[7] In the limit model without these nonparametric "nuisances," the $o_p(1)$ term would be identically zero.

The fixed-k asymptotic theory can be readily applied to test hypotheses related to the spot beta. Specifically, to test the null hypothesis $H_0 : \beta_t = \beta^*$ against the alternative hypothesis $H_a : \beta_t \neq \beta^*$, one would reject the null at significance level α when:

$$\frac{\sqrt{k-1}\left|\hat{\beta}_t - \beta^*\right|}{\sqrt{\hat{\varsigma}_t/\hat{v}_t}} > z_{1-\alpha/2}, \tag{2.12}$$

where $z_{1-\alpha/2}$ denotes the $1-\alpha/2$ quantile of the t_{k-1} distribution. In the Supplemental Appendix, we further extend the coupling-based theory to prove that this test is also optimal, in the sense of being the asymptotically uniformly most powerful unbiased test. Correspondingly, the optimal fixed-k confidence interval (CI) at the $1-\alpha$ asymptotic level is naturally given by:

$$\mathrm{CI}_{1-\alpha} \equiv \left[\hat{\beta}_t - \frac{z_{1-\alpha/2}}{\sqrt{k-1}}\sqrt{\frac{\hat{\varsigma}_t}{\hat{v}_t}}, \hat{\beta}_t + \frac{z_{1-\alpha/2}}{\sqrt{k-1}}\sqrt{\frac{\hat{\varsigma}_t}{\hat{v}_t}}\right]. \tag{2.13}$$

With the window size k given, the proposed fixed-k inference procedure simply suggests that one may treat the problem as a "textbook" Gaussian linear regression and use the student-t distribution to compute critical values. In practical applications, of course, the window size k is not given but needs to be chosen somehow. This choice mirrors the familiar "bias-variance" trade-off omnipresent in nonparametric estimation. Making this choice "optimally" tends to be impractical, as it necessitates additional strong assumptions on both the data and the form of the loss function. This challenge, of course, is also not unique to the proposed fixed-k framework, but applies to the conventional large-k asymptotic framework as well. In that setting, a common practice has been to perform the inference for a range of "reasonable" choices of k as a way to check for robustness. To assist in determining which values of k may indeed be reasonable in the

[6] More generally, the spot covariance matrix estimator $\hat{\boldsymbol{c}}_t$ can be coupled by an $\mathcal{F}_t$-conditional Wishart variable. This result may be useful for making fixed-k inference for other covariance-related quantities, as long as they are pivotalizable. For instance, it can be shown that $\hat{\varsigma}_t/\varsigma_t$ admits a coupling by a scaled chi-squared variable with $k-1$ degrees of freedom.

[7] Theorem 1 remains valid with an estimation window centered at t, as the within-window variations in the spot quantities are of order $o_p(1)$.

fixed-k context, the next section introduces a new diagnostic test and discusses how it can be used to "robustify" the overall inference procedure.[8]

2.3 Diagnostic test for local Gaussianity and robustified inference

The fixed-k inference relies on the local Gaussianity of observed high-frequency returns. This feature is more likely to hold if the stochastic volatility process is nearly constant within the estimation window, which favors the choice of shorter window sizes. For our inferential purpose, it is therefore natural to determine whether k is in fact "small enough" by assessing the adequacy of the local Gaussian approximation. To do so, we suggest new diagnostic tests based on the corresponding "spot" skewness and kurtosis statistics. We present the idea for the regressor process X, but the same diagnostics are readily applicable for the Y process as well.

The spot skewness and kurtosis for the local sample $(\Delta_i^n X)_{i\in\mathcal{I}_{n,t}}$ are formally defined by:

$$\widehat{S}_t \equiv \frac{k^{-1}\sum_{i\in\mathcal{I}_{n,t}}\left(\Delta_i^n X - k^{-1}\sum_{i\in\mathcal{I}_{n,t}}\Delta_i^n X\right)^3}{\left(k^{-1}\sum_{i\in\mathcal{I}_{n,t}}\left(\Delta_i^n X - k^{-1}\sum_{i\in\mathcal{I}_{n,t}}\Delta_i^n X\right)^2\right)^{3/2}}, \tag{2.14}$$

$$\widehat{K}_t \equiv \frac{k^{-1}\sum_{i\in\mathcal{I}_{n,t}}\left(\Delta_i^n X - k^{-1}\sum_{i\in\mathcal{I}_{n,t}}\Delta_i^n X\right)^4}{\left(k^{-1}\sum_{i\in\mathcal{I}_{n,t}}\left(\Delta_i^n X - k^{-1}\sum_{i\in\mathcal{I}_{n,t}}\Delta_i^n X\right)^2\right)^2}. \tag{2.15}$$

These definitions closely resemble the classical skewness and kurtosis statistics frequently used to test for Gaussianity in applied work. However, in contrast to the conventional large-k asymptotic setting, where the skewness and kurtosis statistics are centered at 0 and 3, respectively, and diagnostic tests for normality may be constructed based on a standard central limit theorem together with the delta method (see, e.g., the often used tests by Jarque and Bera (1980)), in the fixed-k setting, the conventional large-sample limit theorems do not apply. Instead, we explicitly derive the infill asymptotic distributions for $\widehat{S}_t$ and $\widehat{K}_t$ using the coupling argument, thereby allowing for the determination of critical values to detect "abnormal" spot skewness and/or kurtosis for a given window size k. The following theorem formalizes the idea.

[8] Contrasting the fixed-k and conventional large-k approaches, if k is chosen to be numerically "large," the t_{k-1}-distribution implied by the fixed-k framework will, of course, be close to the standard normal distribution implied by the conventional framework. However, for k "small," the t_{k-1} and the normal distributions can differ substantially, and the large-k Gaussian approximation can only introduce additional distortions. Hence, we always recommend computing critical values using the t-distribution, regardless of whether the numerical value of k seems large or not.

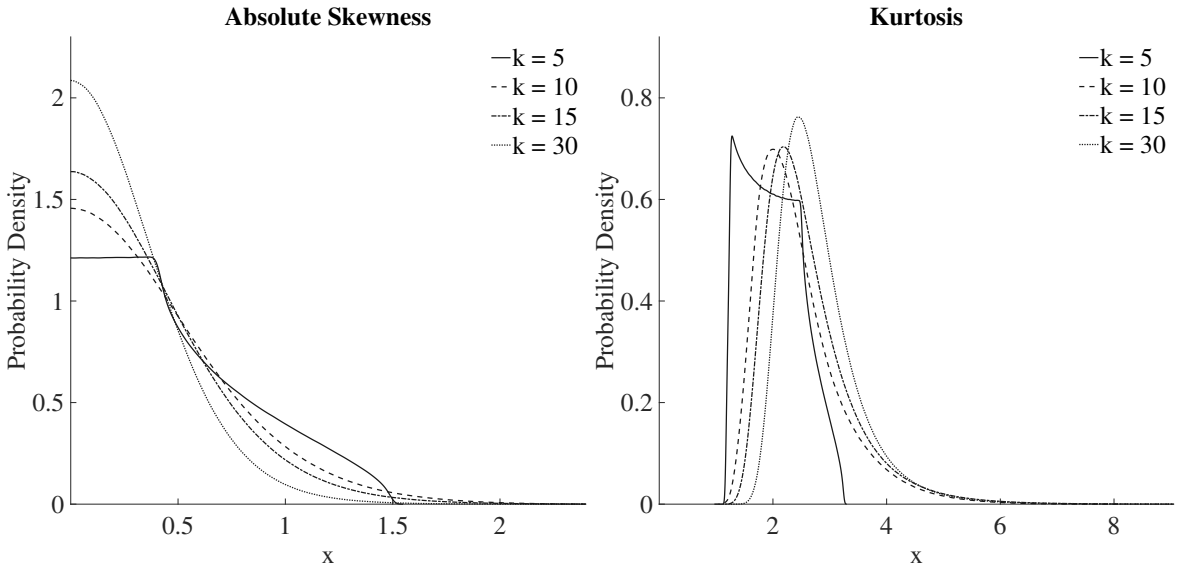

Figure 1: The figure shows the fixed-k asymptotic probability density functions for the absolute spot skewness and spot kurtosis for different values of k.

Theorem 2. *Under the same setting as Theorem 1, there exist i.i.d. standard normal random variables* $(\eta_i)_{i\in\mathcal{I}_{n,t}}$ *such that* $\widehat{S}_t = \widetilde{S}_t + o_p(1)$ *and* $\widehat{K}_t = \widetilde{K}_t + o_p(1)$*, where* $\widetilde{S}_t$ *and* $\widetilde{K}_t$ *are the sample skewness and kurtosis statistics of the coupling variables* $(\eta_i)_{i\in\mathcal{I}_{n,t}}$*, respectively.*

Theorem 2 shows that under the fixed-k asymptotic setting, the distributions of the spot skewness and kurtosis statistics formed using the high-frequency returns are given by the exact distributions of the same statistics formed using i.i.d. standard normal random variables. These distributions are highly nonstandard, but they can be easily computed via Monte Carlo simulations. To illustrate, Figure 1 plots the probability density functions of $|\widetilde{S}_t|$ and $\widetilde{K}_t$ for $k \in \{5, 10, 15, 30\}$.[9]

Under the Itô semimartingale model, these infill asymptotic approximations are theoretically valid for any fixed k. Throughout our analysis, we adhere to this standard model and do not interpret abnormal levels of skewness and/or kurtosis as evidence against this well-established benchmark model. Rather, we recognize that Theorem 2 provides "good" approximations in finite samples only to the extent that the local Gaussian approximation is in force, corresponding to

[9]For a finite sample with k observations, sample skewness and kurtosis statistics are bounded by $(k-2)/\sqrt{k-1}$ and $(k^2-3k+3)/(k-1)$, respectively; see Wilkins (1944) and Dalén (1987). As a result, when $k = 5$, the absolute skewness is bounded by 1.5 and the kurtosis is bounded by 3.25, which explains the bounded supports of the corresponding distributions seen in Figure 1.

k being "sufficiently small." Consistent with this reasoning, we view abnormal skewness and/or kurtosis as a sign that the local Gaussian approximation appears inadequate for a given local sample.

For this purpose, it is convenient to combine the spot skewness and kurtosis statistics into a single diagnostic test statistic

$$\widehat{D}_t \equiv \max\left\{A_k|\widehat{S}_t|, \widehat{K}_t\right\}, \tag{2.16}$$

where the scaling factor $A_k \equiv \text{median}(\widetilde{K}_t)/\text{median}(|\widetilde{S}_t|)$ ensures that the $A_k|\widehat{S}_t|$ and $\widehat{K}_t$ components have comparable magnitudes.[10] Theorem 2 and the continuous mapping theorem readily imply that $\widehat{D}_t = \widetilde{D}_t + o_p(1)$, where $\widetilde{D}_t = \max\{A_k|\widetilde{S}_t|, \widetilde{K}_t\}$. The distribution of the $\widetilde{D}_t$ coupling variable is known in finite samples, and we use its tail quantiles as critical values for the sample $\widehat{D}_t$ statistic. With a slight abuse of terminology, we will say that the diagnostic test "rejects local Gaussianity" for a given local sample at significance level δ, if $\widehat{D}_t$ exceeds the $1-\delta$ quantile of $\widetilde{D}_t$. Such a rejection by the diagnostic test in turn suggests that it would be inadvisable to proceed with the t-test for the spot beta estimate, as the underlying asymptotic approximation could be unreliable. The following algorithm formalizes this decision rule in the form of a "compound" test.

Algorithm 1: Robust t-test for $H_0 : \beta_t = \beta^*$ with diagnostic correction

Step 1. Conduct the $\widehat{D}_t$ diagnostic test for local Gaussianity at significance level δ. If the test rejects, stop the testing procedure without rejecting the $H_0 : \beta_t = \beta^*$ null hypothesis. Otherwise, proceed to step 2.

Step 2. Conduct the t-test for the $H_0 : \beta_t = \beta^*$ null hypothesis at significance level α as outlined in Section 2.2. Report the test decision accordingly. □

This diagnostic correction obviously exerts a larger influence the larger the value of δ. If a researcher fully trusts the asymptotic arguments presented in Section 2.2, and by extension the standard local-Gaussianity heuristic, there would be no need to employ the more conservative robustification, effectively fixing $\delta = 0$. Employing larger values of δ may in turn be viewed as expressing less confidence in the accuracy of the asymptotic theory for a given finite sample. In practice, we recommend employing δ at conventional significance levels, such as 1%, 5%, and

[10] Given that non-zero centered moments are strictly increasing in their orders, it is easy to show that $|\widehat{S}_t| < \widehat{K}_t$ whenever the statistics are well-defined. Hence, $\widehat{D}_t$ would trivially coincide with $\widehat{K}_t$ if A_k were simply set to unity.

10%.[11]

For any $\delta \in (0,1)$ this robust approach is inherently more conservative than the "uncorrected" t-test, as the compound procedure may be stopped in the first step by the local-Gaussianity diagnostic test. However, since the uncorrected t-test controls asymptotic size, the robust t-test does so as well. Importantly, by employing the diagnostic test as a "pre-test," it will not result in over-rejection, while providing better size control in finite samples compared to the uncorrected test. As the diagnostic test is likely to reject when the window size k becomes "too large," the robust approach will therefore also "automatically" become more conservative than the uncorrected t-test. Our simulation results, detailed in Section 4, further corroborate these observations and underscore the practical usefulness of this simple diagnostic correction.

2.4 Predictive inference for spot regressions

The discussion so far pertains to the "in-sample" analysis of spot beta. In many applications, however, the interest centers on predictive analysis. This section develops the requisite econometric tools for such analyses. For concreteness, we focus on the situation in which one relies on the spot beta estimate to form a prediction of Y over some future time period based on the future realization of X. This type of scenario naturally arises in high-frequency event studies, and applications where one is interested in assessing whether the intraday returns on some asset observed around the time of specific news arrivals appear "abnormal."

To set out the idea, let the high-frequency observations used for the spot regression be indexed by $\{1,\dots,k\}$. To emphasize the timing of the information, let $\hat{\beta}_{[1:k]}$ and $\hat{\varsigma}_{[1:k]}$ denote the resulting estimators for the spot beta and the idiosyncratic variance, respectively. Utilizing these estimates, we form the spot-regression-model-based prediction of the h-period return $Y_{(k+h)\Delta_n} - Y_{k\Delta_n}$, as $(X_{(k+h)\Delta_n} - X_{k\Delta_n})\hat{\beta}_{[1:k]}$. This prediction naturally serves as a benchmark for the "normal" level of return on Y to expect. Correspondingly, the ex-post cumulative abnormal return (CAR) of asset Y over the $[k\Delta_n,(k+h)\Delta_n]$ time period may be measured by:

$$\widehat{\mathrm{CAR}}_h = Y_{(k+h)\Delta_n} - Y_{k\Delta_n} - (X_{(k+h)\Delta_n} - X_{k\Delta_n}) \cdot \hat{\beta}_{[1:k]} \ . \tag{2.17}$$

[11] Examining the test results collectively across such a range may also be used to provide a more comprehensive set of statistical evidence to evaluate the hypothesis of interest. Should ambiguous conclusions arise, it may be interpreted as a reflection of the challenges associated with the specific setting.

To help assess the significance of this return, the standard error of $\Delta_n^{-1/2}\widehat{\mathrm{CAR}}_h$ may be estimated as:

$$\widehat{\mathrm{se}}_h = \sqrt{\left(h + \frac{(X_{(k+h)\Delta_n} - X_{k\Delta_n})^2}{\sum_{i=1}^k (\Delta_i^n X)^2}\right) \frac{k\hat{\varsigma}_{[1:k]}}{k-1}}\,, \tag{2.18}$$

which accounts for both the "in-sample" estimation error in the spot beta and the "out-of-sample" idiosyncratic shocks. Theorem 3 below further shows that the t-statistic for the CAR can be coupled by a t-distributed random variable.

Theorem 3. *For $k \geq 2$ and $h \geq 1$, and under Assumption 1, there exists a t_{k-1}-distributed random variable τ_h such that*

$$\frac{\Delta_n^{-1/2}\widehat{\mathrm{CAR}}_h}{\widehat{\mathrm{se}}_h} = \tau_h + o_p(1).$$

Theorem 3 readily suggests forming a predictive CI for the h-period CAR at the $1-\alpha$ level as

$$\mathrm{PCI}_{h,1-\alpha} = \left[-\Delta_n^{1/2}\, z_{1-\alpha/2}\, \widehat{\mathrm{se}}_h, \Delta_n^{1/2}\, z_{1-\alpha/2}\, \widehat{\mathrm{se}}_h\right], \tag{2.19}$$

where $z_{1-\alpha/2}$ denotes the $1-\alpha/2$ quantile of the t_{k-1} distribution. If the estimated CAR falls outside this predictive CI, the abnormal return may be deemed statistically significant at significance level α.

More concretely, for individual equities, large and seemingly abnormal returns typically arise from important firm-specific news releases, prompting the company's stock price to deviate from what is predicted by a benchmark pricing model. In this situation, the predictive CI at a given confidence level in essence depicts a range of probable "counterfactuals" that could have happened in the absence of the news. By design, the predictive CI depends on Y solely through the pre-event estimation window $\{1, \ldots, k\}$, ensuring that it is not "contaminated" by post-event information.

This method for determining "abnormal" returns closely resembles the widely used approach in the event study literature, as discussed in, e.g., Campbell et al. (1997). However, in contrast to the asymptotic Gaussian approximations traditionally employed in that literature, based on "large sample" arguments obtained by aggregating across assets, events, and time periods, the high-frequency method developed here achieves local Gaussianity by concentrating on shorter time windows, thereby allowing for the construction of formal statistical tests pertaining to the immediate impact of singular news events. Section 5.2 provides a concrete empirical illustration.

3 Optimality and Uniform Inference

This section collects some additional theoretical results and discussions. Section 3.1 clarifies a robustness property of the fixed-k method that is not afforded by the conventional large-k method. Section 3.2 demonstrates that the spot beta estimator is asymptotically optimal, achieving the minimum variance among all unbiased estimators, including potential nonlinear ones. Section 3.3 extends the pointwise inference for the spot beta to a uniform version for the entire spot beta process. The content in this section is somewhat technical in nature. Readers primarily interested in the finite-sample performance of the estimators or practical applications may choose to proceed directly to the Monte Carlo simulations in Section 4, or the empirical analysis in Section 5.

3.1 Robustness of the fixed-k method

The conventional large-k theory and the proposed fixed-k theory have their own distinct comparative advantages. On the one hand, the large-k theory implies that the spot beta estimator is consistent, while consistency can no longer be claimed under the fixed-k framework. On the other hand, for inference, the fixed-k theory attains a well-defined theoretical sense of robustness that is not afforded by the large-k theory.

To appreciate this added sense of robustness, it is useful to first recall the bias-variance characterization under the large-k setting with $k_n \to \infty$. Under the maintained assumption that the spot volatility matrix process $\boldsymbol{\sigma}_t$ is κ-Hölder continuous, the nonparametric bias of the spot estimator is of order $(k_n\Delta_n)^{\kappa}$ and its variance is of order k_n^{-1}. The rate-optimal choice of k_n is obtained by balancing the (squared) bias and variance such that $(k_n\Delta_n)^{2\kappa} \asymp k_n^{-1}$, corresponding to $k_n \asymp \Delta_n^{-2\kappa/(2\kappa+1)}$. In order to construct Gaussian-based CIs under the large-k framework, the standard approach is therefore to "undersmooth" by letting k_n grow slower than $\Delta_n^{-2\kappa/(2\kappa+1)}$, so that the nonparametric bias becomes asymptotically negligible. If the Hölder continuity index κ is known, that undersmoothing choice of k_n is feasible, at least in theory. However, if κ is unknown, the large-k Gaussian-based CI would not be robust due to this lack of knowledge of the appropriate smoothing parameter. To appreciate this, fix a large-k tuning scheme of the form $k_n \asymp \Delta_n^{-a}$ for some $a > 0$. Given this choice, there always exist data generating processes with sufficiently rough volatility paths (i.e., κ sufficiently small) such that $a > 2\kappa/(2\kappa+1)$. As such, the given k_n sequence would lead to "oversmoothing" for that collection of models. Correspondingly, the

resulting nonparametric bias will dominate the standard error, and the large-k CI will provide zero coverage asymptotically.

Meanwhile, the requisite undersmoothing is always in force under the fixed-k framework, no matter how small the continuity index κ might be (since $k \ll \Delta_n^{-2\kappa/(2\kappa+1)}$ for any $\kappa > 0$). Theoretically, the fixed-k inference is therefore robust with respect to "arbitrarily rough" volatility, whereas the large-k inference theory is not. In other words, in situations when the stochastic volatility path may lack smoothness, the fixed-k framework provides a built-in robustness not afforded by the conventional large-k approach.

These theoretical considerations aside, the fixed-k and large-k methods do suggest their own distinct approaches for conducting inference. In that regard, as discussed in Section 2.2, the fixed-k small-sample perspective arguably aligns more closely with the underlying finite-sample problem, and importantly as demonstrated by the simulations in Section 4 below, it also generally result in more accurate inference.

3.2 Optimality of the spot beta estimator

In this subsection, we demonstrate that the spot beta estimator, $\hat{\beta}_t$, is indeed optimal, as it is asymptotically minimum-variance among a broad class of asymptotically unbiased "regular" estimators.[12] Our approach can be viewed as a limits-of-experiments method specifically tailored to the infill high-frequency asymptotic setting. Analogous arguments have been successfully employed in other areas of econometrics, including Jansson (2008), Hirano and Porter (2009), and Andrews and Mikusheva (2022).

For ease of notation, denote the vectors of observed high-frequency returns in the local estimation window by:

$$\boldsymbol{r}_X \equiv \left(\Delta_i^n X/\sqrt{\Delta_n}\right)_{i\in\mathcal{I}_{n,t}}, \quad \boldsymbol{r}_Y \equiv \left(\Delta_i^n Y/\sqrt{\Delta_n}\right)_{i\in\mathcal{I}_{n,t}}.$$

A generic spot estimator for β_t may then be expressed as $f(\boldsymbol{r}_X, \boldsymbol{r}_Y)$ for some measurable function $f : \mathbb{R}^k \times \mathbb{R}^k \mapsto \mathbb{R}$. We will thus identify an estimator with the function f and refer to it as *regular* if f is continuous Lebesgue almost everywhere. This regularity requirement appears rather minimal. However, it is sufficient to ensure that $f(\boldsymbol{r}_X, \boldsymbol{r}_Y)$ can be coupled by an analogous estimator in the

[12] Utilizing the same approach, it can be demonstrated that the bias-corrected estimator $\hat{\varsigma}_t^* \equiv k\hat{\varsigma}_t/(k-1)$ is also optimal in the same sense for estimating the spot idiosyncratic variance, ς_t.

limit experiment.[13] The following lemma formalizes this assertion.

Lemma 1. *Under Assumption 1, any regular estimator $f(\boldsymbol{r}_X, \boldsymbol{r}_Y)$ may be expressed as*

$$f(\boldsymbol{r}_X, \boldsymbol{r}_Y) = f\big(v_t^{1/2}\boldsymbol{\eta}, \beta_t v_t^{1/2}\boldsymbol{\eta} + \varsigma_t^{1/2}\boldsymbol{\epsilon}\big) + o_p(1), \tag{3.20}$$

where $\boldsymbol{\eta}$ and $\boldsymbol{\epsilon}$ are independent k-dimensional standard Gaussian random vectors independent of the time-t information set $\mathcal{F}_t$.

Lemma 1 shows that any regular estimator $f(\boldsymbol{r}_X, \boldsymbol{r}_Y)$ may be decomposed into a generally nondegenerate leading term $f(v_t^{1/2}\boldsymbol{\eta}, \beta_t v_t^{1/2}\boldsymbol{\eta} + \varsigma_t^{1/2}\boldsymbol{\epsilon})$ and an asymptotically negligible $o_p(1)$ term. The leading coupling variable may be interpreted as an estimator constructed from the limit experiment with $v_t^{1/2}\boldsymbol{\eta}$ and $\beta_t v_t^{1/2}\boldsymbol{\eta} + \varsigma_t^{1/2}\boldsymbol{\epsilon}$ as the random observables. Since $(v_t, \beta_t, \varsigma_t)$ are $\mathcal{F}_t$-measurable and $(\boldsymbol{\eta}, \boldsymbol{\epsilon})$ is independent of $\mathcal{F}_t$, the limit experiment is de facto an $\mathcal{F}_t$-conditional Gaussian linear model.

Let $L(\cdot, \cdot)$ denote any nonnegative continuous loss function. By the continuous mapping theorem, Lemma 1 implies that $L(\beta_t, f(\boldsymbol{r}_X, \boldsymbol{r}_Y)) = L(\beta_t, f(v_t^{1/2}\boldsymbol{\eta}, \beta_t v_t^{1/2}\boldsymbol{\eta} + \varsigma_t^{1/2}\boldsymbol{\epsilon})) + o_p(1)$. The *asymptotic risk* of the $f(\boldsymbol{r}_X, \boldsymbol{r}_Y)$ estimator is naturally defined as the $\mathcal{F}_t$-conditional expectation of its limit loss,

$$R(f; v_t, \beta_t, \varsigma_t) \equiv \mathbb{E}\left[L\big(\beta_t, f\big(v_t^{1/2}\boldsymbol{\eta}, \beta_t v_t^{1/2}\boldsymbol{\eta} + \varsigma_t^{1/2}\boldsymbol{\epsilon}\big)\big) \,\middle|\, \mathcal{F}_t \right]. \tag{3.21}$$

Since $(\boldsymbol{\eta}, \boldsymbol{\epsilon})$ is independent of $\mathcal{F}_t$, the asymptotic risk depends on the conditioning information only through $(v_t, \beta_t, \varsigma_t)$. As a result, the asymptotic risk function $(v, \beta, \varsigma) \mapsto R(f; v, \beta, \varsigma)$ may readily be computed for any given estimator f and loss function L. However, the exact finite-sample risk $\mathbb{E}\left[L\left(\beta_t, f\left(\boldsymbol{r}_X, \boldsymbol{r}_Y\right)\right) | \mathcal{F}_t\right]$ is generally infeasible to compute due to the presence of various infinite-dimensional nuisance parameters that govern the conditional distribution of the observed data vectors, $\boldsymbol{r}_X$ and $\boldsymbol{r}_Y$.

In line with the previously defined asymptotic risk, the *asymptotic bias* of the $f(\boldsymbol{r}_X, \boldsymbol{r}_Y)$ estimator for β_t is naturally defined by:

$$B(f; v_t, \beta_t, \varsigma_t) \equiv \mathbb{E}\left[f\big(v_t^{1/2}\boldsymbol{\eta}, \beta_t v_t^{1/2}\boldsymbol{\eta} + \varsigma_t^{1/2}\boldsymbol{\epsilon}\big) \,\middle|\, \mathcal{F}_t \right] - \beta_t. \tag{3.22}$$

[13] In asymptotic statistics, as exemplified by Bickel et al. (1998), the notion of regularity involves not only specific asymptotic approximations for the estimator, typically in the form of weak convergence, but also demands the estimator to be equivariant, usually with respect to location shifts. However, in our analysis, we refrain from imposing equivariance restrictions and concentrate on asymptotically unbiased estimators.

We will focus on asymptotically unbiased regular estimators, that is estimators f for which the bias function $B(f; v, \beta, \varsigma)$ is identically equal to zero for all $v > 0$, $\beta \in \mathbb{R}$, and $\varsigma > 0$. An estimator is considered optimal, or the best unbiased estimator (BUE), if it achieves minimum asymptotic variance within that class of estimators. For the quadratic loss function $L(\cdot, \cdot)$, the optimal estimator f^* therefore satisfies[14]

$$R(f^*; v, \beta, \varsigma) \leq R(f; v, \beta, \varsigma), \quad \forall\, v > 0,\, \beta \in \mathbb{R},\, \varsigma > 0, \tag{3.23}$$

for every asymptotically unbiased regular estimator f. Following standard terminology, the optimal estimator defined in this way may be more precisely referred to as the *asymptotically uniformly minimum-variance unbiased estimator*.

Theorem 4. *Under Assumption 1, $\hat{\beta}_t$ is the asymptotically uniformly minimum-variance unbiased estimators for β_t.*

Although the optimality claim for the spot beta estimator is closely related to the Gauss–Markov theorem for standard linear regressions, there are several key differences. Firstly, while the limit experiment underlying the optimality of the spot beta estimator corresponds to a parametric linear regression model, our original econometric setting is inherently nonparametric, and as such it differs importantly from the classical Gauss–Markov setting. Secondly, while the classical Gauss–Markov theorem states that the least-squares estimator is the best linear unbiased estimator (BLUE), the linearity restriction is not necessary in the present context. In other words, when conducting spot regressions for Itô semimartingales, the "local" least-squares $\hat{\beta}_t$ estimator is asymptotically the BUE, not just the BLUE. In a recent paper, Hansen (2022) examines the equivalence between the notions of BLUE and BUE for conventional linear regressions, mainly through the lens of alternative definitions of unbiasedness. The $\hat{\beta}_t$ spot beta estimator achieves its asymptotic superior BUE status for a very different reason, namely the conditional Gaussianity embedded in the limit experiment (Lehmann and Scheffé, 1950).

Using a similar line of reasoning, it is also possible to demonstrate that the t-test for the spot beta, based on Theorem 1, is optimal in the sense that it is asymptotically uniformly most powerful among asymptotically unbiased tests. This additional theoretical result is further detailed in the Supplemental Appendix.

[14] Optimality results for other strictly convex loss functions may similarly be obtained by appealing to Theorem 2.1.11 in Lehmann and Casella (1998).

3.3 Uniform inference for the spot beta process

The fixed-k CI for the spot beta proposed in Section 2.2 pertains to the pointwise inference for β_t at a specific point in time. In this subsection, we develop a uniform functional inference theory for the path of the spot beta process $(\beta_t)_{t\in[0,T]}$ under the fixed-k framework.

Compared to the pointwise inference described in Theorem 1, the uniform inference theory is more challenging to establish for two reasons. First, it needs to be shown that the pointwise coupling result in Theorem 1 holds uniformly across $O(n)$ many local estimation windows, giving rise to a high-dimensional non-Gaussian coupling problem. Second, to conduct inference in this nonstandard setting, we also need to establish an anti-concentration inequality for the maximum absolute value of a large number of t-distributed coupling variables. To help focus our discussion on these more novel aspects of the theory, we purposely consider a simplified setting in which X and Y do not contain any jumps.[15]

Turning to the details, we begin by constructing our functional estimator for the spot beta process. To do so, we divide the n high-frequency returns into $m_n = n/k$ nonoverlapping blocks.[16] Let $\mathcal{I}_{n,j} \equiv \{(j-1)k+1,\ldots,jk\}$ collect the indices of the returns in the jth block. The spot covariance matrix estimator for the jth block is then given by:

$$\tilde{\boldsymbol{c}}_j \equiv \frac{1}{k\Delta_n}\sum_{i\in\mathcal{I}_{n,j}}(\Delta_i^n\boldsymbol{Z})(\Delta_i^n\boldsymbol{Z})^\top.$$

Analogous to equations (2.8) and (2.10), we further define

$$\tilde{v}_j \equiv \tilde{c}_{11,j},\quad \tilde{\beta}_j \equiv \frac{\tilde{c}_{12,j}}{\tilde{c}_{11,j}},\quad \tilde{\varsigma}_j \equiv \tilde{c}_{22,j} - \frac{\tilde{c}_{12,j}^2}{\tilde{c}_{11,j}},$$

and

$$\xi_{11,j} \equiv \sum_{i\in\mathcal{I}_{n,j}}\frac{(\Delta_i^n W_1)^2}{\Delta_n},\quad \xi_{12,j} \equiv \sum_{i\in\mathcal{I}_{n,j}}\frac{(\Delta_i^n W_1)(\Delta_i^n W_2)}{\Delta_n},\quad \xi_{22,j} \equiv \sum_{i\in\mathcal{I}_{n,j}}\frac{(\Delta_i^n W_2)^2}{\Delta_n}.$$

The collection of blockwise estimators $(\tilde{\beta}_j)_{1\le j\le m_n}$ naturally serves as a piecewise constant estimator for the path $(\beta_t)_{t\in[0,T]}$. With a slight abuse of notation, we identify $(\tilde{\beta}_j)_{1\le j\le m_n}$ with $(\tilde{\beta}_t)_{t\in[0,T]}$

[15] This is quite unrestrictive, as one could simply apply the standard truncation technique (Mancini, 2001) to consistently eliminate jumps in a preliminary step without affecting the asymptotics; see, e.g., Proposition 1 in Li et al. (2017) for a formal justification.

[16] For ease of exposition, we assume that n/k is an integer.

through $\tilde{\beta}_t \equiv \tilde{\beta}_j$ for $t \in [(j-1)\,k\Delta_n, jk\Delta_n)$ and $j \in \{1,\ldots,m_n\}$, and $\tilde{\beta}_T \equiv \tilde{\beta}_{m_n}$. The functional estimators $(\tilde{v}_t)_{t\in[0,T]}$ and $(\tilde{\varsigma}_t)_{t\in[0,T]}$ are defined analogously.

Theorem 1 can be easily modified to show the following pointwise coupling result for a given time point $t \in [(j-1)\,k\Delta_n, jk\Delta_n]$,

$$\frac{\sqrt{k-1}\,(\tilde{\beta}_t - \beta_t)}{\sqrt{\tilde{\varsigma}_t/\tilde{v}_t}} = \xi_{\beta,j} + o_p(1),$$

where $\xi_{\beta,j} \equiv \sqrt{k-1}\,\xi_{12,j}/\sqrt{\xi_{11,j}\xi_{22,j} - \xi_{12,j}^2}$ is t_{k-1}-distributed. The main challenge for developing a uniform inference theory is to show that this approximation holds uniformly across $t \in [0,T]$. That is, it pertains to all $m_n = O(n)$ estimation blocks simultaneously. To achieve this, we need to strengthen Assumption 1 as follows, where $\|\cdot\|_p$ denotes the L_p-norm.

Assumption 2. *Suppose that the process $\boldsymbol{Z}$ satisfies (2.1) and (2.4) with $\boldsymbol{J} = 0$, and that there exists a sequence $(T_m)_{m\geq 1}$ of stopping times increasing to infinity such that the following conditions hold for each $m \geq 1$: (i) there exists a constant K_m such that $\|\boldsymbol{b}_t\| + \|\boldsymbol{\sigma}_t\| + v_t^{-1} + \varsigma_t^{-1} \leq K_m$ for all $t \in [0, T_m]$; and (ii) for some constant $\kappa > 0$ and any $p \geq 2$, there exist constants $K_{m,p}$ such that $\| \sup_{s,t,|s-t|\leq h} \|\boldsymbol{\sigma}_{t\wedge T_m} - \boldsymbol{\sigma}_{s\wedge T_m}\| \|_p \leq K_{m,p} h^\kappa$ for any $h > 0$.*

Assumption 2 explicitly rules out jumps in the price process. As already noted, this is not overly restrictive as one could apply the standard truncation technique to consistently eliminate jumps. Meanwhile, condition (ii) in Assumption 2 is substantially more restrictive than condition (ii) in Assumption 1. In particular, while the previous milder condition only requires the κ-Hölder continuity to hold under the L_2-norm, the more stringent condition requires the volatility process to be κ-Hölder continuous under all L_p-norms. This stronger condition allows us to apply maximal inequalities under certain L_p-norms for the derivation of the requisite high-dimensional coupling. The following theorem formalizes the results.

Theorem 5. *Suppose that Assumption 2 holds and $k \geq \max\{13, 6/\kappa + 1\}$. Then,*

$$\sup_{t\in[0,T]} \left| \frac{\sqrt{k-1}\,(\tilde{\beta}_t - \beta_t)}{\sqrt{\tilde{\varsigma}_t/\tilde{v}_t}} \right| = \tau_n^* + o_p(1), \tag{3.24}$$

where $\tau_n^* \equiv \max_{1\leq j\leq m_n} |\xi_{\beta,j}|$ and $(\xi_{\beta,j})_{1\leq j\leq m_n}$ are i.i.d. t_{k-1}-distributed random variables. Moreover, with the $1-\alpha$ quantile of τ_n^* denoted by $z_{n,1-\alpha}^*$,

$$\mathbb{P}\left(\sup_{t\in[0,T]} \left| \frac{\sqrt{k-1}\,(\tilde{\beta}_t - \beta_t)}{\sqrt{\tilde{\varsigma}_t/\tilde{v}_t}} \right| > z_{n,1-\alpha}^* \right) \to \alpha. \tag{3.25}$$

The first part of Theorem 5 shows that the sup-t statistic associated with the spot beta process over the $[0,T]$ time interval can be coupled by τ_n^*, defined as the absolute maximum of m_n i.i.d. t_{k-1}-distributed random variables.[17] The second part of the theorem further shows that $z^*_{n,1-\alpha}$ is a valid critical value for conducting a functional sup-t test for the entire spot beta process. Since the distribution of the τ_n^* coupling variable is known in finite samples, its $1-\alpha$ quantile, $z^*_{n,1-\alpha}$, can be easily computed. Inverting the resulting sup-t test in turn yields a $1-\alpha$ level uniform confidence band for the spot beta process:

$$\text{CB}_{1-\alpha} \equiv \left[\tilde{\beta}_t - \frac{z^*_{n,1-\alpha}}{\sqrt{k-1}} \sqrt{\frac{\tilde{\varsigma}_t}{\tilde{v}_t}}, \tilde{\beta}_t + \frac{z^*_{n,1-\alpha}}{\sqrt{k-1}} \sqrt{\frac{\tilde{\varsigma}_t}{\tilde{v}_t}} \right], \quad t \in [0,T]\,. \tag{3.26}$$

This uniform confidence band obviously tends to be wider than the pointwise fixed-k CIs previously defined in (2.13). We will use these tools to draw across-time inference about the spot beta process in our empirical analysis in Section 5.1 below.

It is worth noting that the proof for (3.25) underlying the uniform confidence band in (3.26) not only relies on the uniform coupling given in (3.24), but also requires a so-called anti-concentration property for the coupling variable τ_n^* to ensure that its distribution does not concentrate "too much" at any given point. This property cannot be taken for granted, as τ_n^* is defined as the maximum of a growing dimensional random vector (i.e., $(\xi_{\beta,j})_{1\leq j\leq m_n}$) rather than a fixed random variable. Similar anti-concentration inequalities have also been widely used in the recent literature on high-dimensional inference. While existing work mainly pertains to Gaussian variables (see, e.g., Chernozhukov et al. (2015)), our analysis is concerned with t-distributed variables. Correspondingly, in our proof in the Supplemental Appendix, we establish that τ_n^* does indeed satisfy the requisite anti-concentration property. This result is to the best of our knowledge novel and of independent interest.

[17] Although the pointwise coupling result in Theorem 1 holds for any fixed $k \geq 2$, the uniform version given by Theorem 5 requires $k \geq 13$ or even larger when $\kappa < 1/2$. This requirement arises mainly because of the required uniform coupling of the t-statistic process, which involves the spot idiosyncratic variance estimator $\tilde{\varsigma}_t$ in the denominator. When $\tilde{\varsigma}_t$ is close to zero, the coupling error for the t-statistic will be magnified, and we need to "tame" this effect uniformly across the "many" estimation blocks. Bounding k from below fulfills this purpose because the estimation error in $\tilde{\varsigma}_t$ is approximately χ^2_{k-1} distributed. That noted, our sufficient condition may not be necessary, and it is possible that it could be further weakened.

4 Monte Carlo Simulations

Our Monte Carlo simulations are based on two different data generating processes (DGP). Under DGP I, the volatility v_t for the regressor process X_t is simulated using the setup originally proposed by Bollerslev and Todorov (2011). This same simulation design has also subsequently been used by a number of other studies (e.g., Bollerslev et al. (2021) and Li et al. (2023)). In particular, we rely on a two-factor model for $v_t = V_{1,t} + V_{2,t}$, where $V_{1,t}$ and $V_{2,t}$ are generated according to:

$$dV_{1,t} = 0.0128(0.4068 - V_{1,t})dt + 0.0954\sqrt{V_{1,t}}\left(\rho dW_{1,t} + \sqrt{1-\rho^2}dB_{1,t}\right),$$
$$dV_{2,t} = 0.6930(0.4068 - V_{2,t})dt + 0.7023\sqrt{V_{2,t}}\left(\rho dW_{1,t} + \sqrt{1-\rho^2}dB_{2,t}\right),$$

in which B_1 and B_2 represent independent standard Brownian motions, which are also independent of the bivariate Brownian motion $\boldsymbol{W}$ that drives the $\boldsymbol{Z}_t = (X_t, Y_t)^\top$ process (as defined below). The parameter $\rho = -0.7$ captures the well-documented negative correlation between price and volatility shocks. Interpreting the unit time interval as one day, the V_1 volatility factor exhibits high persistence with a half-life of 2.5 months, while the V_2 volatility factor is characterized by quick mean reversion with a half-life of only a day.

Our second DGP II is based on the following one-factor model for the v_t process:

$$dv_t = 18.0218(0.8136 - v_t)dt + 5.3153\sqrt{v_t}\left(\rho dW_{1,t} + \sqrt{1-\rho^2}dB_{1,t}\right),$$

where the parameters are calibrated such that the simulated rejection frequencies for the diagnostic test for local Gaussianity align with those for our empirical study discussed in the next section. Compared to DGP I, the volatility paths observed under DGP II tend to experience much more rapid changes, with a half-life of just 15 minutes. This higher level of volatility-of-volatility naturally complicates the local Gaussian approximation, and as such DGP II should present a more challenging scenario for nonparametric spot inference compared to DGP I.

Meanwhile, armed with the two different DGPs for the v_t process, the full volatility matrix $\boldsymbol{\sigma}_t$ defined in (2.4) is determined by

$$\beta_t = 1 + 0.25\sin(t)^2, \qquad \varsigma_t = \left(1.5 + 0.25\sin(t)^2\right)v_t.$$

under both DGPs. The formula for β_t implies that it changes between 1.0 and 1.25 over the course of roughly one-and-a-half day. The expression for ς_t implies that the idiosyncratic variance comoves

Table 1: Diagnostic Tests for Local Gaussianity

	DGP I			DGP II		
k	1%	5%	10%	1%	5%	10%
5	0.010	0.050	0.100	0.010	0.052	0.104
10	0.010	0.050	0.100	0.016	0.066	0.124
15	0.010	0.051	0.101	0.023	0.083	0.144
30	0.010	0.051	0.102	0.044	0.123	0.193
60	0.011	0.053	0.104	0.081	0.180	0.260

Note: The table presents the finite-sample rejection rates of the $\widehat{D}_t$ diagnostic test for local Gaussianity for the DGP I and II data generating processes, for various window sizes k, at the 1%, 5%, and 10% significance levels. The rejection rates are calculated using one million Monte Carlo replications.

with the variance of X_t, with the ratio ς_t/v_t varying between 1.5 and 1.75. Finally, completing both DGPs, we generate the bivariate $\boldsymbol{Z}_t$ process according to $d\boldsymbol{Z}_t = \boldsymbol{\sigma}_t d\boldsymbol{W}_t$.

We simulate all of the "continuous-time" processes using an Euler scheme on a one-second mesh. However, the actually observed X and Y processes used for the inference are only sampled at a coarser one-minute, or $\Delta_n = 1/390$, time interval. This again is motivated by applications involving high-frequency financial data for actively traded assets, where as discussed above a one-minute sampling scheme is commonly employed as a convenient way to guard against complicated and difficult-to-specify high-frequency market microstructure noise. It also directly mirrors the sampling frequency of the data employed in the empirical applications discussed in Section 5 below.

We begin our simulation analysis by examining the accuracy of the local Gaussian approximation across various window sizes, by implementing the diagnostic test based on $\widehat{D}_t$ (recall (2.16)) for $k \in \{5, 10, 15, 30, 60\}$, and significance levels 1%, 5%, and 10%. Finite-sample rejection rates under the different settings are displayed in Table 1. Focusing on the left panel and the results obtained under DGP I, it is evident that the rejection rates all closely align with their respective nominal levels. As such, this supports the prediction from the asymptotic theory, as formalized in Theorem 2, and suggests that the local Gaussian approximation works quite well under DGP

I for a broad range of window sizes. The results for DGP II, reported in the right panel, reveals a markedly different picture. With a narrow window size of $k = 5$, the rejection rates are again all close to their nominal levels. However, as the window size k increases, the number of rejections steadily increase, underscoring the relevance of properly assessing the accuracy of the local Gaussian approximations underlying the small-sample fixed-k framework.

Table 2 in turn presents the finite-sample rejection rates for different t-tests pertaining to the spot betas. All of the tests are conducted at the 10% nominal significance level. The table is divided into two panels corresponding to each of the two DGPs. Each panel contains results for different estimation window sizes k, and the rejection rates obtained for five different versions of the spot beta tests: the conventional large-k asymptotic-Gaussian method based on (2.9), the uncorrected fixed-k method based on (2.12), and the fixed-k method "robustified" with 1%, 5%, or 10% level diagnostic correction, as outlined in Algorithm 1.

Looking first at the results for DGP I, reported in the top panel, we observe that the conventional asymptotic-Gaussian-based test exhibits significant over-rejections for all window sizes, particularly when k is small. By comparison, the tests based on the fixed-k critical values all demonstrate notably better size control. The rejection rates for the uncorrected test are generally very close to the 10% nominal level, while the robust t-tests tend to be slightly undersized.

Turning to the bottom panel, and the results for the more challenging DGP II, the conventional asymptotic-Gaussian-based test continues to display considerable size distortion. The fixed-k test without diagnostic correction maintains reasonable size control for $k = 5$ and 10, but it starts to exhibit nontrivial over-rejections as k increases beyond that. This aligns with the previous observations from Table 1, and the finding that the accuracy of the local Gaussian approximation deteriorates for larger values of k. Meanwhile, as seen from the last three columns of the table, this size distortion may be effectively addressed by employing the simple "robustified" version of the tests outlined in Algorithm 1. The 5% diagnostic correction appears to work especially well, but a 10% correction also achieves very good, albeit slightly conservative, size control across all values of k.

In sum, the simulation results demonstrate that the fixed-k method offers more reliable inference than the conventional large-k asymptotic-Gaussian-based method routinely employed in the literature, especially when implemented over relatively narrow estimation windows. We turn next to a pair of applications that naturally call for such empirical design.

Table 2: Finite-Sample Rejection Rates for Spot Beta t-tests

		Fixed-k with Diagnostic Correction			
k	Gaussian	Uncorrected	1%	5%	10%
		DGP I			
5	0.215	0.100	0.098	0.091	0.083
10	0.154	0.100	0.099	0.091	0.082
15	0.134	0.100	0.098	0.090	0.082
30	0.118	0.101	0.099	0.091	0.082
60	0.122	0.113	0.111	0.102	0.092
		DGP II			
5	0.225	0.107	0.104	0.096	0.087
10	0.172	0.116	0.112	0.101	0.089
15	0.162	0.125	0.118	0.103	0.090
30	0.159	0.140	0.126	0.106	0.089
60	0.178	0.168	0.141	0.113	0.092

Note: The table reports the finite-sample rejection rates for spot beta t-tests under the null hypothesis. All of the tests are conducted at the 10% nominal significance level. The tests are based on the conventional asymptotic-Gaussian method (Gaussian), the fixed-k method without diagnostic correction (Uncorrected), and the fixed-k method with 1%, 5%, or 10% level diagnostic correction, as described in Algorithm 1. The rejection rates are calculated using one million Monte Carlo replications.

5 Empirical Applications

To highlight the practical applicability and usefulness of the new inference procedures, we consider two separate empirical applications. Section 5.1 examines the tracking performance of leveraged ETFs as measured by their spot betas, while Section 5.2 showcases the use of the predictive spot inference in the context of an intraday event study. We base both of the studies on high-frequency data sampled at one-minute frequency, sourced directly from Tick Data. As discussed above, following the extant literature, we deliberately rely on a "coarse" one-minute sampling frequency to mitigate the impact of difficult-to-specify market microstructure noise. The dataset for the ETF spans the period from January 2, 2018, to December 31, 2022, with stock market holidays and half-trading days excluded following standard practice. To ensure robustness against price jumps, we further apply a five-standard-deviation truncation rule, with the truncation threshold calibrated using the bipower variation estimator of Barndorff-Nielsen and Shephard (2004b) along with a correction for diurnal patterns in the volatility following Li et al. (2017).[18]

5.1 Are leveraged ETFs properly leveraged?

The ETF market has witnessed tremendous growth over the past two decades in both size and the variety of individual ETFs tailored to unique investment goals. Among the most notable innovations are leveraged ETFs, which have considerably expanded the range of investment strategies available to investors. These financial instruments aim to magnify the returns of an underlying asset using leverage, thereby offering the potential for greater short-term gains or more effective hedging. Initially, this burgeoning industry concentrated on market indices and sectors, but in recent years, it has extended to include many individual stocks as well. Many of these innovative financial instruments also involve the active management and trading of various derivative contracts.

It remains a fundamental question whether these new funds actually live up to their stated objectives and achieve their intended leverage? To investigate this issue, we analyze the relationship between three ETFs directly linked to the Nasdaq-100 stock market index. The Nasdaq-100 includes 100 of the largest companies listed on the Nasdaq Stock Exchange, and it serves as one of

[18] As shown in a previous working paper version, our main empirical findings remain largely unaffected with respect to alternative ways of applying this theoretically motivated jump-removal procedure.

the primary barometers for the tech sector. The index itself is not directly tradable, but the QQQ index fund passively tracks it as closely as possible, and that fund has now also become one of the world's largest ETFs in terms of assets under management. The other two ETFs that we examine are designed to replicate leveraged positions in the Nasdaq-100 index: the TQQQ fund aims to generate three times the return on the index, while the inverse SQQQ fund seeks to produce three times the return on a short position in the index. Both of these funds are now also extensively used for investment and risk management purposes.[19]

However, it is important to note that both the TQQQ and SQQQ only aim to track their specific multiples of the Nasdaq-100 index over a single day, and as such the funds are primarily recommended to be used over short intraday time frames.[20] Thus, rather than questioning whether the two leveraged funds produce the desired multiples of the return on the QQQ fund over longer multi-day investment horizons, the empirically more relevant question is whether the fund's intraday tracking performance is in line with the stated objectives. Properly addressing this question naturally calls for the use of the optimal spot inference procedures. Specifically, treating QQQ as the regressor process X and TQQQ or SQQQ as the dependent process Y, we want to assess if the intraday spot betas for the funds do indeed equal their respective target values of 3 and -3.

To do so, we begin by assessing the adequacy of the local Gaussian approximation using the skewness-kurtosis $\widehat{D}_t$ diagnostic statistic in (2.16) for $k \in \{5, 10, 15, 30, 60\}$, corresponding to estimation windows ranging from five minutes to one hour. Table 3 presents the proportion of estimation windows for which local Gaussianity is rejected at the 1%, 5%, or 10% significance level, for each of the three ETFs, over the full 2018–2022 sample period. For $k = 5$, the empirical rejection rates are close to the nominal level, indicating that the local Gaussian approximation appears adequate for such short windows. Meanwhile, as expected, the diagnostic test tends to reject more often for larger values of k, with the local Gaussian approximation appearing quite inadequate for $k = 30$ and 60. Correspondingly, we will focus our main empirical investigation

[19] For example, an investor with a long position in several technology stocks may buy the SQQQ ETF to help neutralize her systematic risk exposure to the tech sector. But, if the SQQQ fund deviates from its stated objective of returning -3 times the QQQ, that investor may find herself exposed to unintended systematic risks.

[20] Quoting from ProShares' (the group that manages the funds) descriptions, they explicitly caution that: "due to the compounding of daily returns, holding periods of greater than one day can result in returns that are significantly different than the target return, and ProShares' returns over periods other than one day will likely differ in amount and possibly direction from the target return for the same period."

Table 3: Diagnostic Tests for Local Gaussianity of Nasdaq 100 ETFs

	QQQ			TQQQ			SQQQ		
k	1%	5%	10%	1%	5%	10%	1%	5%	10%
5	0.011	0.054	0.106	0.011	0.054	0.105	0.015	0.055	0.105
10	0.017	0.069	0.126	0.017	0.069	0.126	0.015	0.065	0.120
15	0.025	0.085	0.146	0.025	0.084	0.146	0.022	0.078	0.137
30	0.047	0.124	0.193	0.047	0.124	0.193	0.038	0.110	0.177
60	0.081	0.177	0.225	0.082	0.179	0.256	0.065	0.154	0.228

Note: The table presents the average rejection rates of the $\widehat{D}_t$ diagnostic test across all estimation blocks over the 2018–2022 sample period, for different window sizes k. The diagnostic test counts a rejection for local Gaussianity at the 1%, 5%, or 10% significance level when $\widehat{D}_t$ statistic exceeds the 99%, 95%, or 90% quantile of $\widetilde{D}_t$, respectively.

and discussion below on $k = 15$, or 15-minute spot-estimation windows, with additional supportive robustness checks for $k = 5$ and 10 provided in the Supplemental Appendix.

Turning to the evaluation of the intraday tracking performance of the two leveraged funds, we begin by examining the frequency at which the null hypothesis that their spot betas equal their target values of 3 and -3, respectively, are rejected. We conduct the t-test at a 10% significance level, and summarize the results by calculating the average rejection rates across all within-day 15-minute estimation windows for each of the 60 months in our sample. To address potential size distortions due to "finite-sample breakdowns" of local Gaussianity, we also employ the robust t-tests outlined in Algorithm 1 in Section 2.3, with the diagnostic tests performed at 1%, 5%, and 10% significance levels.

Figure 2 displays the monthly average rejection rates for both the uncorrected t-test and the robustified tests based on various degrees of diagnostic corrections. Since all of the tests are conducted at the 10% nominal level, we would expect average rejection rates not to exceed 10% if the funds' betas are indeed consistent with their target values. However, as the figure shows, the empirical rejection rates surpass 10% for most of the months in the sample, and quite dramatically so for certain months. Also, while the diagnostic correction reduces the rejection rates as expected, it does not materially affect the main findings. It is noteworthy that, compared

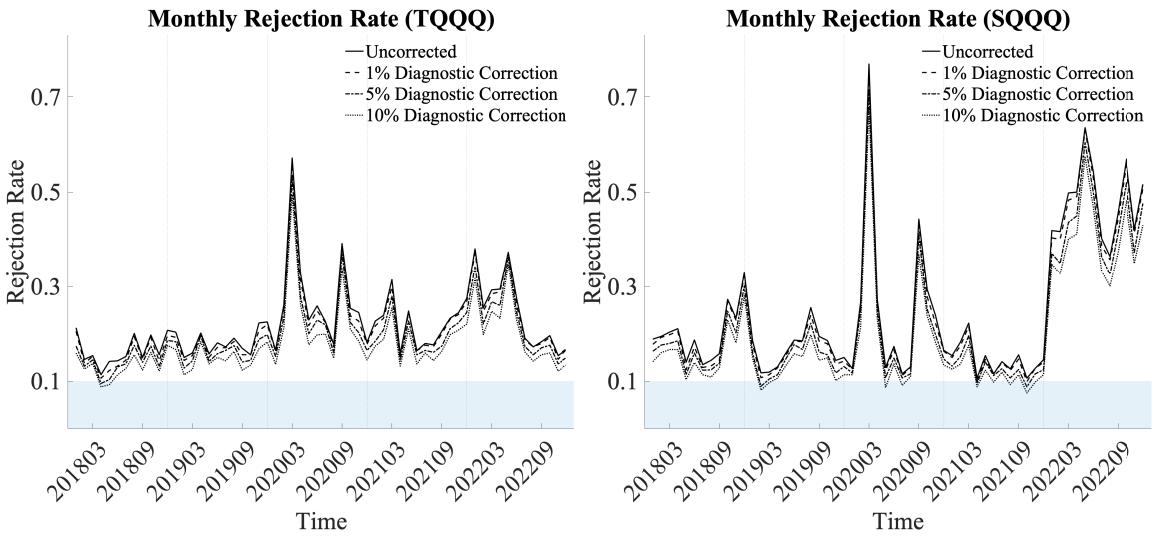

Figure 2: The figure presents average monthly rejection rates for the fixed-k t-tests for spot beta, including the uncorrected test, and the three robustified tests with diagnostic correction at the $\delta = 1\%$, 5%, and 10% levels. All of the tests are evaluated at the 10% significance level. The left panel reports the results for the null hypothesis that the spot betas for the TQQQ with respect to the QQQ equal 3. The right panel tests that the betas for the SQQQ with respect to the QQQ equal -3. The averages are computed across all 15-minute estimation windows for each of the individual months in the 2018–2022 sample.

to TQQQ, the inverse SQQQ was much more frequently "off target" during 2022, highlighting the unique challenges involved in properly managing short-ETFs during certain time periods and economic conditions.[21]

Meanwhile, the month of March 2020 clearly stands out as the month with the most rejections for both funds. That month, of course, is also coincident with the initial outbreak of the COVID-19 pandemic in the U.S., and it witnessed unusually high levels of stock market volatility. In an effort to curb some of that volatility, the U.S. stock market was subjected to four separate (and rare) trading halts, or "circuit breakers," on March 9, 12, 16, and 18, 2020. Thus, while the deteriorating performance of the two leveraged funds around that time is not necessarily surprising, it is exactly during such extreme market events that the funds may be the most useful for risk management

[21] During that time period, the Federal Reserve raised its target range for the federal funds rate from $[0, 0.25\%]$ to $[4.25\%, 4.5\%]$, arguably resulting in a bear market, especially for the tech sector.

purposes.

In an effort to gain further insight into the funds' performance during this tumultuous time, Figure 3 plots the estimated spot betas and the associated 90% fixed-k CIs, as described in (2.13), for all of the 15-minute estimation windows during the two weeks from March 9 to March 20, 2020. In addition, each of the different panels in the figure also reports the p-value for the joint, or uniform, hypothesis that the spot beta process for the entire day is in line with its target value based on the functional test described in Theorem 5.

Looking first at the top panel of Figure 3 and the results for the TQQQ fund, it is noteworthy that on the four circuit-breaker days (i.e., March 9, 12, 16, and 18, as indicated by bold font in the figure), the estimated spot betas are almost all significantly higher than the target value of 3. In other words, the fund was actually "over-leveraged" on those specific days. By contrast, for the four days immediately following the circuit-breaker days, the spot betas for the TQQQ fund are almost all significantly below the target value, indicating that the fund was "under-leveraged" on those days. The p-values for the functional tests pertaining to the betas over the entire day further corroborate the significance of these general conclusions.

The results for the SQQQ fund, displayed in the bottom panel, evidence similar and, if anything, even more pronounced systematic deviations. However, unlike the findings of excessively high betas for the TQQQ fund on the four circuit-breaker days, the estimated spot betas for the SQQQ fund are all numerically too low compared to the target value of -3. Put differently, the inverse leveraged fund was "under-leveraged" on those days and did not provide sufficient negative exposure to the Nasdaq-100 index. The functional tests again reinforce these same qualitative findings. As noted above, such systematically "wrong" intraday leverage levels can be quite detrimental from a practical risk management perspective. Yet, it would be impossible to accurately tell whether these deviations are actually statistically significant, or merely due to chance, without the new spot inference procedures developed here.

5.2 An intraday event study: Microsoft's AI moment

Our second empirical application pertains to the way in which financial markets respond to new information. Specifically, we seek to characterize the abnormal return on Microsoft (MSFT) stock around the time of Microsoft's widely publicized "The Future of Work With AI" event on March 16, 2023.

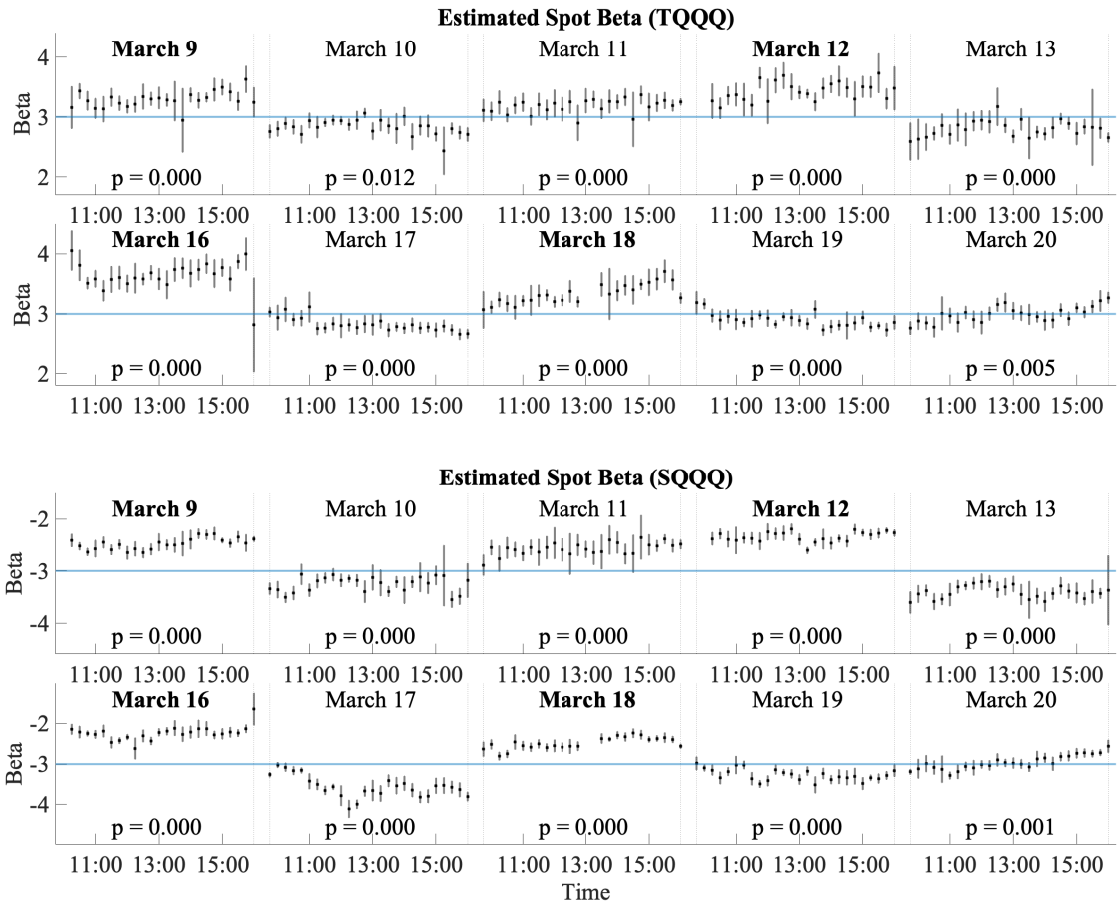

Figure 3: The figure presents the spot betas for the TQQQ (top) and SQQQ (bottom) funds with respect to QQQ estimated over 15-minute windows, together with the corresponding 90% confidence intervals for the two weeks from March 9 to March 20, 2020. The p-values reported in each of the panels refer to the uniform, or functional, hypotheses that the spot beta processes for the different days are in line with their target values. The four days on which the market circuit-breaker went into effect are highlighted in bold font.

The event commenced at 11:00 AM, with the introduction of Microsoft's new AI tool named *Copilot* designed to integrate large language models into its suite of productivity software, and it lasted until approximately 11:36 AM. Meanwhile, to prevent any "pre-event drift" from confounding the results, we impose a 15-minute "buffer" between the official start time of the event and the estimation window for the spot beta used for calculating the abnormal returns.[22] Relatedly, even though the event commenced at 11:00 AM, the new information was only revealed gradually, with the actual demonstration of the new AI tool, which arguably represented the most novel aspect of the event, starting more than 10 minutes into the presentation.

Accordingly, following the discussion in Section 2.4, we estimate the CARs for the log return on MSFT (Y) with respect to the market portfolio (X), as proxied by the SPY ETF for the S&P 500 index, over h-minute horizons after 10:45 AM as:

$$\widehat{\mathrm{CAR}}_h = Y_{10:45+h} - Y_{10:45} - (X_{10:45+h} - X_{10:45}) \cdot \hat{\beta}_{[10:30,10:45]},$$

with h spanning up until Noon.[23] To assess the statistical significance of the resulting CARs, we compute the associated 90% level predictive CIs, or PCI_h's as defined in (2.19). As previously discussed, these are naturally interpreted as the probable counterfactuals in scenarios without any significant news.

Figure 4 presents the results, along with various annotated news segments, as per Microsoft's official announcement. The predictive CIs obviously tend to widen as the horizon h increases, commensurate with increased uncertainty for predicting the returns over longer horizons. Interestingly, there appears to be a hint of a moderate positive drift in the CARs during the 10:45-11:00 pre-event window, although the estimates remain within the bounds of the predictive CI. The estimated CARs also stay relatively "flat" and within the CI bounds for the first 15 minutes following the official start of the event, further corroborating the lack of any information leakage.[24]

[22]The existence of pre-event drifts, possibly associated with the leakage of information, has been well-documented in the event study literature; see, e.g., the early analysis in Rendelman et al. (1982) pertaining to earnings announcements. Perhaps more surprisingly, Lucca and Moench (2015) has also recently found evidence for a FOMC pre-announcement drift.

[23]The Supplemental Appendix provides additional robustness checks for $k = 5$ and 10, yielding very similar, albeit slightly noisier, results compared to the ones for $k = 15$ discussed below. Very similar results are also obtained using the more tech-heavy QQQ ETF in place of the SPY as the regressor X.

[24]During the first 15 minutes of the event, presenters primarily reiterated the company's general vision regarding the adoption of AI technologies. Given that similar information was already known since the debut of ChatGPT-4

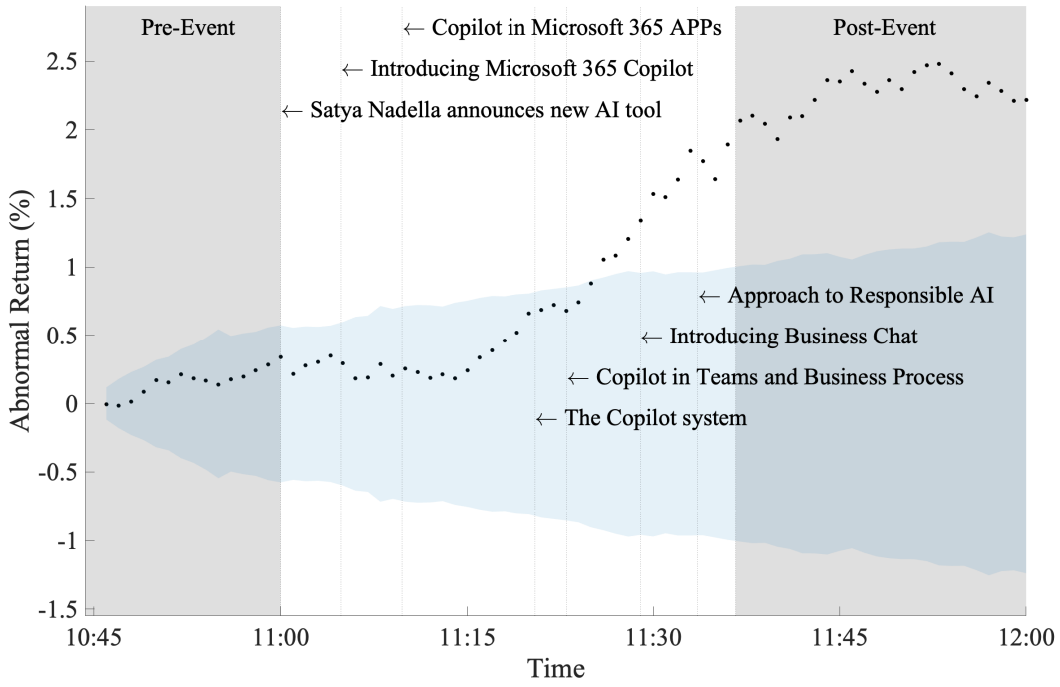

Figure 4: The figure shows the minute-by-minute cumulative abnormal returns on Microsoft stock from 10:45 AM to Noon on March 16, 2023 (dots), together with 90% level predictive confidence intervals (shaded area). The spot regression used in the calculation of the abnormal returns is estimated with returns over the 15-minute window spanning 10:30–10:45 AM. The timelines for specific announcements during "The Future of Work With AI" event are explicitly highlighted.

Meanwhile, at approximately 11:15 AM there is a marked change in the CARs, and when judged by the predictive CIs, the CARs observed at the conclusion of the event at around 11:36 AM are all highly significant. This noticeable change in the CARs coincides with the presenter stating that "*the real power of Copilot is in how it unleashes creativity at work*," proceeding to showcase a series of practical business applications. The scope and potential impact of these applications apparently went far beyond what the market expected, resulting in a highly significant abnormal return of close to 2% over just 15 minutes, corresponding to a roughly $45 billion increase in the market value for Microsoft stock.

Putting these findings further into perspective, the Microsoft event occurred only two days after

a few days earlier, it is not surprising that the corresponding CAR is essentially "flat."

OpenAI's launch of ChatGPT-4 on March 14, 2023. That same time period also coincided with other generative AI technologies garnering substantial attention. Such clustering of many different closely related news events would pose a serious challenge to the application of conventional low-frequency event study methodology, which attempts to achieve asymptotic Gaussianity and assess statistical significance by pooling across time and news events. Doing so in the present context would obviously confound different pre- and post-event windows. Instead, by relying on local Gaussianity over a short event window, the new high-frequency-based fixed-k approach developed here allows for meaningful inference concerning singular news events without the need for any pooling or aggregation across days or events.

6 Concluding Remarks

The conventional approach for nonparametric inference on spot regressions relies on asymptotic-Gaussian-based approximations. The quality of such approximations can be poor when the estimators are constructed from relatively few observations over narrow estimation windows, as often done in practice to accommodate potentially rapid fluctuations in the true betas and the other latent state processes. The new optimal estimation theory and fixed-k inference framework developed here explicitly recognizes this disconnect between conventional asymptotic-Gaussian-based inference procedures and the small sample sizes typically utilized in practice. Specifically, by exploiting the local Gaussianity of the underlying workhorse continuous-time model, the fixed-k framework effectively converts the original nonparametric spot regression problem into a finite-sample Gaussian linear regression problem. This transformation in turn allows for the use of classical finite-sample inference techniques and decision-theoretic arguments to establish optimality for both estimation and testing. Importantly, it also allows for much sharper inference in empirically realistic settings.

Moving forward, the fixed-k approach offers a new avenue for possibly incorporating additional insights from classical statistics into the high-frequency data setting. In particular, since the limit model is fully parametric, a Bayesian approach could be adopted. Such an approach may be particularly beneficial in the context of spot estimation, as it inherently presents a data-scarce environment making any prior information especially valuable. At a broader level, the strategy of obtaining an approximate parametric model through localization, and the subsequent use of finite-sample analysis and inference, is not exclusive to the continuous-time high-frequency data

setting either. For example, if the observed data distribution is mixture-Gaussian, and the mixing variable remains nearly constant within specific subsamples, one may similarly localize the problem, permitting the use of Gaussian-based optimal statistical decision theory from more conventional settings.[25] We leave further explorations of these ideas for future research.

References

Aït-Sahalia, Y., J. Fan, and D. Xiu (2010): "High-Frequency Covariance Estimates With Noisy and Asynchronous Financial Data," *Journal of the American Statistical Association*, 108, 1504–1517.

Aït-Sahalia, Y. and J. Jacod (2014): *High-Frequency Financial Econometrics*, Princeton University Press.

Aït-Sahalia, Y. and J. Jacod (2020): "From Tick Data to Semimartingales," *The Annals of Applied Probability*, 30, 2740 – 2768.

Andersen, T. G. and T. Bollerslev (2018): *Volatility*, Edward Elgar Publishing.

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

Andersen, T. G., T. Bollerslev, F. X. Diebold, and G. G. Wu (2005): "A Framework for Exploring the Macroeconomic Determinants of Systematic Risk," *American Economic Review*, 95, 398–404.

Andersen, T. G., M. Thyrsgaard, and V. Todorov (2021): "Recalcitrant Betas: Iintraday Variation in the Cross-Sectional Dispersion of Systematic Risk," *Quantitative Economics*, 12, 637–682.

Andrews, I. and A. Mikusheva (2022): "Optimal Decision Rules for Weak GMM," *Econometrica*, 90, 715–748.

Andrews, I., J. H. Stock, and L. Sun (2019): "Weak Instruments in Instrumental Variables Regression: Theory and Practice," *Annual Review of Economics*, 11, 727–753.

Arkhangelsky, D., S. Athey, D. A. Hirshberg, G. W. Imbens, and S. Wager (2021): "Synthetic Difference-in-Differences," *American Economic Review*, 111, 4088–4118.

[25] As a case in point, this strategy could potentially be useful in the context of synthetic control problems (as in e.g., Arkhangelsky et al. (2021)), where error terms are assumed to be Gaussian.

Bachelier, L. (1900): "Théorie de la spéculation," *Annales scientifiques de l'École Normale Supérieure*, 3e série, 17, 21–86.

Back, K. E. (2017): *Asset Pricing and Portfolio Choice Theory*, Oxford University Press.

Bandi, F. M. and J. R. Russell (2008): "Microstructure Noise, Realized Variance, and Optimal Sampling," *Review of Economic Studies*, 75, 339–369.

Barndorff-Nielsen, O., N. Shephard, and M. Winkel (2006): "Limit Theorems for Multipower Variation in the Presence of Jumps in Financial Econometrics," *Stochastic Processes and Their Applications*, 116, 796–806.

Barndorff-Nielsen, O. E. and N. Shephard (2004a): "Econometric Analysis of Realized Covariation: High Frequency Based Covariance, Regression, and Correlation in Financial Economics." *Econometrica*, 72, 885 – 925.

——— (2004b): "Power and Bipower Variation With Stochastic Volatility and Jumps," *Journal of Financial Econometrics*, 2, 1–37.

Bibinger, M., N. Hautsch, P. Malec, and M. Reiss (2019): "Estimating the Spot Covariation of Asset Prices - Statistical Theory and Empirical Evidence," *Journal of Business & Economic Statistics*, 37, 419–435.

Bibinger, M. and M. Reiss (2014): "Spectral Estimation of Covalitility From Noisy Observations Using Local Weights," *Scandinavian Journal of Statistics*, 41, 23–50.

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., R. F. Engle, and J. M. Wooldridge (1988): "A Capital Asset Pricing Model With Time-Varying Covariances," *Journal of Political Economy*, 96, 116–131.

Bollerslev, T., J. Li, and Z. Liao (2021): "Fixed-k Inference for Volatility," *Quantitative Economics*, 12, 1053–1084.

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

Braun, P. A., D. B. Nelson, and A. M. Sunier (1995): "Good News, Bad News, Volatility, and Betas," *Journal of Finance*, 50, 1575–1603.

Campbell, J. Y., A. W. Lo, and A. MacKinlay (1997): *The Econometrics of Financial Markets*, Princeton University Press.

Chernozhukov, V., D. Chetverikov, and K. Kato (2015): "Comparison and Anti-Concentration Bounds for Maxima of Gaussian Random Vectors," *Probability Theory and Related Fields*, 162, 47–70.

Cochrane, J. H. and M. Piazzesi (2002): "The Fed and Interest Rates - A High-Frequency Identification," *American Economic Review*, 92, 90–95.

Comte, F. and E. Renault (1998): "Long Memory in Continuous-Time Stochastic Volatility Models," *Mathematical Finance*, 8, 291–323.

Dalén, J. (1987): "Algebraic Bounds on Standardized Sample Moments," *Statistics & Probability Letters*, 5, 329–331.

Delbaen, F. and W. Schachermayer (1994): "A general version of the fundamental theorem of asset pricing," *Mathematische annalen*, 300, 463–520.

Diebold, F. X. and G. Strasser (2013): "On the Correlation Structure of Microstructure Noise: A Financial Economic Approach," *Review of Economic Studies*, 80, 1304–1337.

Duffie, D. (2010): *Dynamic Asset Pricing Theory*, Princeton University Press.

Einstein, A. (1905): "Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen," *Annalen der Physik*, 322, 549–560.

Engle, R. F. (2004): "Risk and Volatility: Econometric Models and Financial Practice," *American Economic Review*, 94, 405–420.

Faust, J., E. Swanson, and J. Wright (2004): "Identifying VARs Based on High Frequency Futures Data," *Journal of Monetary Economics*, 51, 1107–1131.

Ferson, W. E. and C. R. Harvey (1996): "Conditioning Variables and the Cross Section of Stock Returns," *Journal of Finance*, 54, 1325–1360.

Foster, D. P. and D. B. Nelson (1996): "Continuous Record Asymptotics for Rolling Sample Variance Estimators," *Econometrica*, 64, 139–174.

Gagliardini, P., E. Ossola, and O. Scaillet (2016): "Time-Varying Risk Premium in Large Cross-Sectional Equity Data Sets," *Econometrica*, 84, 985–1046.

Hansen, B. E. (2022): "A Modern Gauss–Markov Theorem," *Econometrica*, 90, 1283–1294.

Hayashi, F. (2011): *Econometrics*, Princeton University Press.

Hayashi, T. and N. Yoshida (2005): "On Covariance Estimation of Non-Synchronously Observed Diffusion Processes," *Bernoulli*, 11, 359–379.

Hirano, K. and J. R. Porter (2009): "Asymptotics for Statistical Treatment Rules," *Econometrica*, 77, 1683–1701.

Jacod, J., J. Li, and Z. Liao (2021): "Volatility Coupling," *The Annals of Statistics*, 49, 1982–1998.

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

Jansson, M. (2008): "Semiparametric Power Envelopes for Tests of the Unit Root Hypothesis," *Econometrica*, 76, 1103–1142.

Jarque, C. M. and A. K. Bera (1980): "Efficient tests for normality, homoscedasticity and serial independence of regression residuals," *Economics Letters*, 6, 255–259.

Kristensen, D. (2010): "Nonparametric Filtering of the Realized Spot Volatility: A Kernel-Based Approach," *Econometric Theory*, 26, 60–93.

Lehmann, E. L. and G. Casella (1998): *Theory of Point Estimation*, Springer Texts in Statistics.

Lehmann, E. L. and H. Scheffé (1950): "Completeness, Similar Regions, and Unbiased Estimation: Part I," *Sankhya: The Indian Journal of Statistics (1933-1960)*, 10, 305–340.

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

Li, J., D. Wang, and Q. Zhang (2023): "Reading the Candlesticks: An OK Estimator for Volatility," *Review of Economics and Statistics*, forthcoming.

Lucca, D. O. and E. Moench (2015): "The Pre-FOMC Announcement Drift," *Journal of Finance*, 70, 329–371.

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

Merton, R. (1992): *Continuous Time Finance*, Basil Blackwell.

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

Nakamura, E. and J. Steinsson (2018): "High-Frequency Identification of Monetary Non-Neutrality: The Information Effect," *Quarterly Journal of Economics*, 133, 1283–1330.

Nelson, D. B. (1996): "Asymptotic Filtering Theory for Multivariate ARCH Models," *Journal of Econometrics*, 71, 1–47.

Protter, P. E. (2005): *Stochastic Integration and Differential Equations*, Springer Berlin Heidelberg.

Rendelman, R. J., C. P. Jones, and H. A. Latané (1982): "Empirical Anomalies Based on Unexpected Earnings and the Importnace of Risk Adjustmemnts," *Journal of Financial Economics*, 10, 269–287.

Shephard, N. (2022): "Some Properties of the Sample Median of an In-Fill Sequence of Events With an Application to High Frequency Financial Econometrics," Tech. rep., Harvard University.

Sims, C. A. and T. Zha (2006): "Where There Regime Switches in US Monetary Policy?" *American Economic Review*, 96, 54–81.

Staiger, D. and J. H. Stock (1997): "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 65, 557–586.

Wilkins, J. E. (1944): "A Note on Skewness and Kurtosis," *The Annals of Mathematical Statistics*, 15, 333 – 335.