

Singapore Management University

Institutional Knowledge at Singapore Management University

Dissertations and Theses Collection (Open Access)

Dissertations and Theses

5-2024

Essays on high-frequency financial econometric

Qiyuan LI

Singapore Management University, qyli.2019@phdecons.smu.edu.sg

Follow this and additional works at: https://ink.library.smu.edu.sg/etd_coll



Part of the [Econometrics Commons](#)

Citation

LI, Qiyuan. Essays on high-frequency financial econometric. (2024). 1-180.

Available at: https://ink.library.smu.edu.sg/etd_coll/607

This PhD Dissertation is brought to you for free and open access by the Dissertations and Theses at Institutional Knowledge at Singapore Management University. It has been accepted for inclusion in Dissertations and Theses Collection (Open Access) by an authorized administrator of Institutional Knowledge at Singapore Management University. For more information, please email cherylds@smu.edu.sg.

ESSAYS ON HIGH-FREQUENCY FINANCIAL ECONOMETRICS

BY

QIYUAN LI

A DISSERTATION

IN

ECONOMICS

Presented to the Singapore Management University in Partial Fulfilment

of the Requirements for the Degree of PhD in Economics

2024

Supervisor of Dissertation

PhD in Economics, Programme Director

ESSAYS ON HIGH-FREQUENCY FINANCIAL ECONOMETRICS

by
Qiyuan Li

Submitted to the School of Economics in Partial Fulfilment of the
Requirements for the Degree of Doctor of Philosophy in Economics

Dissertation Committee:

Jia Li (Supervisor/Chair)
Lee Kong Chian Professor of Economics
Singapore Management University

Peter C. B. Phillips
Sterling Professor of Economics & Professor of Statistics
Yale University
University of Auckland
University of Southampton
Singapore Management University

Jun Yu
UMDF Chair Professor of Finance and Economics
University of Macau

Tim Bollerslev
Juanita and Clifton Kreps Professor of Economics
Duke University

Singapore Management University
2024

© Copyright (2024) Qiyuan Li

Abstract

This dissertation consists of three papers contributing to the theory of estimation and inference of high-frequency financial data.

In the second chapter, a general framework is introduced for optimal nonparametric spot volatility estimation based on intraday range data, comprised of the first, highest, lowest, and last prices over a given time interval. Employing a decision-theoretic approach together with a coupling-type argument, the form of the nonparametric estimator is directly tailored to the specific volatility measure of interest and the relevant loss function. The resulting new optimal estimators offer substantial efficiency gains compared to existing commonly used range-based procedures.

The third chapter extends the previous chapter to handle multiple candlesticks, proposing a computationally more efficient algorithm for spot volatility estimation. Additionally, an exact simulation scheme is introduced to address biases in Euler discretization, enabling precise risk comparison and further analysis involving extreme values of Brownian motions.

The fourth chapter addresses the uniform inference problem for high-frequency data that includes prices, volumes, and trading flows. Such data is modeled within a general state-space framework, where the latent state process is corresponding risk indicators, e.g., volatility, price jump, average order size, and arrival of events. The functional estimators are formed as a collection of localized estimates across different time points. Although the proposed estimators do not admit a functional central limit theorem, a Gaussian strong approximation, or coupling, is established under in-fill asymptotics to facilitate feasible inference. The proposed methodology is applied to distinguish the informative part from the Federal Open Market Committee speeches, and to analyze the impact of social media activities on cryptocurrency markets.

Table of Contents

1	Introduction	1
2	Optimal Nonparametric Range-Based Volatility Estimation	3
2.1	Introduction	3
2.2	Nonparametric Range-Based Volatility Estimation	7
2.2.1	Theoretical Setting and Decision-Theoretic Framework	7
2.2.2	Optimal Estimation for Spot Variance and Volatility	13
2.2.3	Risk Comparisons	18
2.2.4	Optimal Estimators with Multiple Candlesticks	21
2.3	An Empirical Illustration	25
2.4	Concluding Remarks	28
3	Optimal Spot Volatility Estimation Based on Multiple Candlesticks	30
3.1	Nonparametric Range-Based Volatility Estimation	30
3.1.1	Theoretical Setting and Background	30
3.1.2	AMRE Estimator with Multiple Candlesticks	32
3.1.3	Unbiased Evaluation of Estimation Risk	34
3.1.4	Risk Comparisons	37
3.1.5	Understanding the AMRE Estimators	44
3.2	An Empirical Illustration	47
3.3	Concluding Remarks	50
4	Uniform Inference for High-Frequency Data	51
4.1	Introduction	51
4.2	Theory	55
4.2.1	State-space Model for High-Frequency Data	55

4.2.2	Motivating Examples	57
4.2.3	Uniform Inference on Conditional Mean Process	61
4.2.4	Uniform Inference on Conditional Quantile Process	68
4.2.5	Application: Inference for Ranks	73
4.3	Monte Carlo Simulations	77
4.3.1	Data Generating Processes	77
4.3.2	The Results	80
4.4	Empirical Illustration	82
4.4.1	Detecting Information Flows during FOMC Speeches	82
4.4.2	Case Study	91
4.4.3	Impact of Twitter on Cryptocurrency Markets	93
4.5	Concluding Remarks	97
5	Conclusion	98
A	Technical Results for Chapter 2	100
A.1	Proofs of the Main Results	100
A.1.1	Proof of Theorem 2.1	100
A.1.2	Proof of Theorem 2.2	102
A.2	Additional Results	107
A.2.1	Optimal Estimation for Quarticity and Precision	107
A.2.2	Proof of Theorem A.1	110
A.2.3	Proof of Theorem A.2	110
A.2.4	Comparisons of Single Versus Multiple Candlesticks	111
A.2.5	Comparison Among Alternative Estimators	115
A.2.6	Monte Carlo Simulations	116
B	Technical Results for Chapter 3	124
B.1	Proofs of the Main Results	124
B.1.1	Proof of Theorem 3.2	124
B.1.2	Proof of Theorem 3.3	125
B.2	Additional Results	126
B.2.1	Best Subset Regression for Spot Variance Estimators	126
B.2.2	Volatility Estimation during Price Indices Release	128

C	Technical Results for Chapter 4	131
C.1	Proofs of the Main Results	131
C.1.1	Proof of Theorem 4.1	131
C.1.2	Proof of Theorem 4.2	138
C.1.3	Proof of Theorem 4.3	149
C.1.4	Proof of Corollary 4.1	156
C.2	Extension to Dependent Disturbance	156
	Bibliography	160

Acknowledgement

This dissertation represents the culmination of a challenging yet exciting journey, and it would not have been possible without the support and guidance of numerous individuals.

First and foremost, I am deeply indebted to my supervisor and committee chair, Professor Jia Li, whose expertise, dedication, patience, and encouragement have been invaluable to the completion of my PhD study. He is a personal role model to me, and his rigorous attitude toward research has significantly shaped my professional growth. I also wish to extend my heartfelt gratitude to my committee members, Professor Peter C. B. Phillips, Professor Jun Yu, and Professor Tim Bollerslev, for their valuable comments, constructive suggestions, and especially for their constant support during my job search process. I am incredibly fortunate and it is my greatest pleasure to have them on my committee.

My sincere thanks go to my colleagues and friends, especially Dennis Lim, Naiyuan Hu, and Xincheng Huang, for their companionship, stimulating discussions, and moral support. Their camaraderie has made this journey less isolating.

I owe my deepest gratitude to my family. To my parents, for their unconditional love and encouragement, their belief in me has been my guiding light. To my partner, Xinran Liu, for always being there during the highs and lows. Her patience, understanding, and unwavering support have been my anchor.

Lastly, to everyone who has been part of this journey, this accomplishment is as much yours as it is mine.

Chapter 1

Introduction

Over the past two decades, the analysis of high-frequency data has emerged as one of the most dynamic and popular fields in financial econometrics, driven by the growing availability of data from high-frequency trading. Such data offers a detailed depiction of price movements over narrow time windows, which naturally enables doing localized analysis. With a growing number of observations within a fixed time window, we gain more precise insights into underlying processes, without restrictions on their stationarity. Consequently, estimation and inference can be carried out in a nonparametric fashion. With more and more observation within a fixed time window, we can have a more precise information of underlying process regardless whether they are stationary or not. In contrast to the long-span (large T) asymptotics, this limiting notion is referred to as infill asymptotics. Analysis under infill asymptotics and corresponding data is made possible by the development of new mathematical methods and econometric tools.

One of the most central topics in high-frequency financial econometrics is volatility estimation and inference, a focus of the first two chapters. Unlike commonly used estimates in the literature, which rely solely on high-frequency returns, we make use of the information embedded in high-low ranges and so-called “candlestick charts,” which encompass open, high, low, and close prices over each observation interval. Such data is readily available on an intraday basis in many databases. We rely on ideas from decision theory to exploit the full information inherent in such candlesticks for volatility estimation. Specifically, in the second chapter, co-authored with Professor Tim Bollerslev and Professor Jia Li, we formally define op-

tinality of spot volatility estimator under a decision-theoretical framework, through the use of a “coupling” argument. Based on the definition, we derive closed-form expressions for optimal spot volatility estimators with a single candlestick. Our results show the superiority of the proposed estimators compared to conventional return-based estimators and other existing range-based estimators in the literature.

The third chapter, co-authored with Professor Tim Bollerslev, Professor Jia Li, and Assistant Professor Yifan Li, extends our analysis to multiple candlesticks. One of the primary challenges in computing optimal estimates with more than one candlestick is the absence of closed-form expressions in this scenario. Machine learning-based method proposed in the previous chapter suffers from the one-sided bias of the Euler discretization scheme when dealing with the supremum and infimum of a continuous Brownian path, which renders the training procedures computationally costly and less reliable. To address this problem, we introduce an exact simulation scheme capable of generating samples exactly following the same distribution of a continuous Brownian path. Additionally, we propose an alternative algorithm that outperforms the machine learning-based approach in both computational efficiency and precision.

In the fourth chapter, we turn to uniform inference, aimed at analyzing the global properties of the entire underlying process. Here, our focus extends beyond high-frequency price data to encompass other market indicators such as trading volumes and trading flows. Following the literature, we model these market indicators by a general state-space framework, where the latent state processes can be volatility, average order size, and arrival of events. The functional estimators are formed as the collection of localized estimates across different time points. The major challenge in conducting uniform inference arises from the fact that the proposed estimator does not admit a functional central limit theorem. To address this problem, we adopt a technique borrowed from the high-frequency statistics literature, namely, the strong approximation. The core concept is to find an alternative sequence with a known finite-sample distribution yet remains close enough to the original functional estimator under specific metric. Our results facilitate the construction of uniform confidence bands, and can be applied to tackle other econometric problems.

Chapter five concludes. Appendix contains all the proofs and additional results.

Chapter 2

Optimal Nonparametric Range-Based Volatility Estimation

2.1 Introduction

Most financial and macroeconomic time series exhibit time-varying volatility. Accurate assessments of said volatilities are important for financial decision making and the evaluation of economic policies alike. Accordingly, a large econometrics literature has emerged over the past several decades dedicated to the development of ever more reliable volatility estimation procedures. We add to this burgeoning literature by providing new optimal range-based volatility estimators.¹ We rely on a novel decision-theoretic approach together with a coupling-type asymptotic representation to explicitly tailor the form of the optimal estimator to the volatility measure of interest and relevant loss function. In so doing, we demonstrate non-trivial efficiency gains for the new optimal estimators compared to commonly used procedures.

Prompted by the increased availability of high-frequency intraday prices for a variety of financial assets and markets, most of the volatility estimation procedures proposed in the more recent literature have been nonparametric, built on the notion

¹Following the existing literature, we will refer to any estimator that exploits not only the information in the high and low prices over a given time interval, but also the first and last prices over the interval, as a “range-based” estimator. When there is no ambiguity, we will also frequently use the word “volatility” as a catchall for any scale measure, the variance included.

of ever finer sampled returns and corresponding infill asymptotic arguments (see, e.g., the introductory discussion in [Andersen and Bollerslev \(2018\)](#)). In a stylized theoretical setting, the use of finely sampled intraday returns naturally affords more accurate volatility estimates than the use of coarser, say daily, returns. Empirically, however, the presence of market microstructure “noise” presents formidable challenges to the direct use of ultra high-frequency returns, necessitating more advanced robust inference procedures and/or the use of “not-too-finely” sampled intraday returns (see, e.g., the discussion in [Jacod et al. \(2017\)](#) and [Li and Linton \(2022\)](#), along with the many additional references therein).

Meanwhile, pioneering work by [Parkinson \(1980\)](#) and [Garman and Klass \(1980\)](#), dating back almost half-a-century, first demonstrated the increased accuracy for daily variance estimation afforded by harnessing the richer information embedded in the daily high-low range and so-called “candlestick charts,” comprised of the open, high, low, and close prices over the day.² This type of daily data has long been freely available for a vast array of financial assets. It is now also readily available on an intraday basis.³ Importantly, and in parallel to the common use of “not-too-finely” sampled high-frequency intraday returns, intraday candlesticks sampled at “not-too-fine” a frequency offer a similar built-in robustness to market microstructure noise, and as such holds the promise of easy-to-implement improved volatility estimation.⁴ Yet, it remains an open question how to optimally exploit the full information inherent in such candlesticks for said estimation.

We rely on ideas from decision theory to provide a definitive answer to this question. Classical decision theory generally invokes specific parametric distributional assumptions to determine the optimal estimator that minimizes the specific risk. By contrast, our high-frequency framework adopts a nonparametric approach.

²Candlestick charts are also routinely used by finance practitioners in the formulation of technical trading strategies. The first such documented use of candlestick charting dates back to the 18th century and the Japanese rice trader Munehisa Homma; see, for example, [Nison \(2001\)](#) for an introduction to the main ideas.

³High-frequency candlestick data is provided by various online trading platforms (e.g. E-Trade, Robinhood), publicly available databases (e.g., Yahoo Finance), and commercial databases (e.g., Bloomberg, Tick Data, TAQ).

⁴Extending our ideas to range-based estimation with even finer sampled intraday candlesticks for which the noise cannot be ignored would be an interesting direction for future research. However, as discussed further below, the requisite task of pinning down the fine structure of the noise and the underlying economic mechanisms presents formidable challenges beyond our main research question.

We leverage the infill asymptotic “coupling” method recently developed by [Bollerslev et al. \(2021\)](#) to bridge the gap between our setting and the classical decision-theoretic approach. This enables us to derive unique optimal high-frequency range-based spot volatility estimators corresponding to particular loss functions (e.g., Quadratic or Stein) and volatility measures (e.g., σ_t , σ_t^2 , σ_t^4 , or σ_t^{-1}). For spot estimation based on a single candlestick, we derive closed-form analytical expressions for the optimal estimators. These estimators are non-standard, but straightforward to implement in practice. In cases involving multiple candlesticks, we provide semi-closed form solutions for the optimal estimators and illustrate how to employ machine learning tools to numerically compute the optimal estimation functions.

Our results are most closely related to the recent work of [Li et al. \(2022\)](#). Extending the original analysis in [Garman and Klass \(1980\)](#) based on the assumption of a continuous-time price process with constant volatility to a high-frequency nonparametric infill asymptotic setting, [Li et al. \(2022\)](#) propose a range-based estimator for the spot volatility that achieves the minimum asymptotic variance within the class of unbiased linear estimators. Their proposal may be regarded as the best linear unbiased estimator (BLUE) for spot volatility. While that analysis is informative, it is also incomplete, and by design much simpler than the present analysis. In particular, a priori restricting the functional form of the estimator to be linear simplifies the search for the “optimal” estimator to a search for the optimal set of weights, as opposed to a search for the risk-minimizing estimator in an infinite-dimensional functional space. Importantly, restricting the functional form also does not guarantee that the resulting “shape-constrained” optimal estimator is actually *the* optimal estimator.⁵ Indeed, as we demonstrate below, the “unconstrained” optimal nonparametric range-based spot volatility estimators derived here often provide nontrivial efficiency gains compared to existing procedures hitherto derived in the literature under various simplifying assumptions, the classical Garman–Klass estimator and the BLUE estimator of [Li et al. \(2022\)](#) included.

⁵Of course, seemingly ad hoc functional-form restrictions do not necessarily result in efficiency loss. For example, in Gaussian linear regression models, the ordinary least-squares estimator is also the uniformly minimum-variance unbiased estimator by the Lehmann–Scheffé theorem (see, e.g., [Shaffer \(1991\)](#)). That is, the BLUE estimator is also the best unbiased estimator (BUE); see also the related recent discussion pertaining to possibly non-Gaussian linear regression models in [Hansen \(2022\)](#) and [Pötscher and Preinerstorfer \(2022\)](#).

Further relating our work to the existing high-frequency literature on nonparametric volatility estimation, most of the prior theoretical work on optimal estimation of spot volatility (see, e.g., [Foster and Nelson \(1996\)](#), [Comte and Renault \(1998\)](#), [Kristensen \(2010\)](#), and Chapter 13 in [Jacod and Protter \(2012\)](#)) has primarily been concerned with rate optimality. However, that optimality criterion sheds little light on the estimators' actual finite-sample performance.⁶ Another strand of the literature has instead been concerned with the semiparametric efficient estimation of integrated volatility functionals (see, e.g., [Mykland and Zhang \(2009\)](#), [Jacod and Rosenbaum \(2013\)](#), [Renault et al. \(2017\)](#), and [Li and Liu \(2021\)](#)). The optimality concept typically adopted in that literature has been built on the convolution theorem and the related local asymptotic minimaxity results for locally asymptotically mixed normal (LAMN) models (see, e.g., [Le Cam \(1960\)](#), [Hájek \(1972\)](#), [Jeganathan \(1982, 1983\)](#)). By contrast, our coupling theory directly links the nonparametric range-based spot volatility estimation/decision problem with a *non*-Gaussian limit experiment. As a result, the functional form of our new optimal estimators generally depend on the loss function and are quite nonstandard, although straightforward to implement in practice.

The remainder of this paper is organized as follows. In Section 2, we start by outlining our nonparametric high-frequency setting and basic assumptions, followed by a discussion of our key coupling arguments. We then introduce the new optimal range-based spot volatility estimators and provide a characterization of their asymptotic properties. Section 2.3 illustrates the practical applicability of the new estimators, and shows the intraday candlestick-based spot volatility estimates for a market portfolio for each of the eight 2022 prescheduled Federal Open Market Committee (FOMC) announcement days. We conclude with a few suggestions for future research. All proofs are included in Appendix A.1, while additional theoretical and numerical results can be found in Appendix A.2.

⁶[Kristensen \(2010\)](#) does seek to characterize the optimal choice of the smoothing kernel. However, the underlying assumption that the volatility process has differentiable sample paths rules out all Brownian stochastic volatility models, as well as any model featuring volatility jumps.

2.2 Nonparametric Range-Based Volatility Estimation

2.2.1 Theoretical Setting and Decision-Theoretic Framework

The (log) price process P is assumed to follow an Itô semimartingale defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$ of the form

$$P_t = P_0 + \int_0^t b_s ds + \int_0^t \sigma_s dW_s + J_t, \quad 0 \leq t \leq T, \quad (2.1)$$

where the drift process b and the volatility process σ are both càdlàg adapted, W is a standard Brownian motion, and J is a pure-jump process driven by a Poisson random measure. We are interested in the optimal nonparametric estimation of the p th power of the spot volatility, σ_t^p , at some fixed time t under a standard infill asymptotic setting with the sampling interval $\Delta_n \rightarrow 0$. We will focus our discussion in the main text on cases with $p = 2$ (variance) and $p = 1$ (volatility). The same ideas may similarly be applied in the construction of optimal estimators for other powers p .⁷

The baseline Itô semimartingale in (2.1) is directly motivated by no-arbitrage arguments. However, it is well-known that the process is misspecified empirically at ultra high, or tick level, frequencies. In addition to a host of market microstructure frictions that “contaminate” the actually observed prices (see, e.g., [Diebold and Strasser \(2013\)](#) for a discussion of the underlying economic mechanisms), prices are also not truly recorded on a continuous-time scale. The most commonly used approach to circumvent these difficulties for the purpose of volatility estimation is to “down-sample” the available data, and rely on returns at “not-too-high” a frequency $1/\Delta_n$. The practical choice of Δ_n has typically been guided by the so-called volatility signature plot introduced by [Andersen et al. \(2000\)](#) (see also the discussion in [Hansen and Lunde \(2006\)](#), and the recent formalization of that approach in [Aït-Sahalia and Xiu \(2019\)](#)). The new estimation method proposed here is similarly intended to be used with “not-too-finely” sampled data. Put differently, acknowledging that the workhorse Itô semimartingale model is only meant as a plausible approximation over “coarser” time scales, effectively allows us to follow the common approach in the literature and remain agnostic about the fine structure of the

⁷Analogous results for $p = 4$ (quarticity) and $p = -1$ (precision) are presented in the Supplemental Appendix [A.2](#).

market microstructure noise.⁸

The existing high-frequency econometrics literature on nonparametric volatility estimation has primarily been focused on estimators formed using high-frequency returns; i.e., $P_{i\Delta_n} - P_{(i-1)\Delta_n}$. We augment the information in the high-frequency return by “looking inside” the Δ_n time-interval through the lens of high-frequency candlesticks. More specifically, denote the i th sampling interval by $\mathcal{T}_i = [(i-1)\Delta_n, i\Delta_n]$. The corresponding candlestick then provides information on the open, high, low, and close prices, formally defined by $P_{(i-1)\Delta_n}$, $\sup_{t \in \mathcal{T}_i} P_t$, $\inf_{t \in \mathcal{T}_i} P_t$, and $P_{i\Delta_n}$, respectively. This information may be summarized in the form of the three (normalized) returns

$$r_i \equiv \frac{P_{i\Delta_n} - P_{(i-1)\Delta_n}}{\sqrt{\Delta_n}}, \quad u_i \equiv \frac{\sup_{t \in \mathcal{T}_i} P_t - P_{(i-1)\Delta_n}}{\sqrt{\Delta_n}}, \quad l_i \equiv \frac{\inf_{t \in \mathcal{T}_i} P_t - P_{(i-1)\Delta_n}}{\sqrt{\Delta_n}}, \quad (2.2)$$

where r_i denotes the usual open-close return traditionally used for high-frequency-based volatility estimation, and u_i (resp. l_i) refers to the high (resp. low) return brought by the candlestick (to help fix ideas, see Figure 2.1). All range-based estimators may be expressed as functions of (r_i, u_i, l_i) . To facilitate our representation and subsequent discussion of the optimal estimators, it is convenient to also define the scaled range $w_i \equiv u_i - l_i$ (as also indicated in Figure 2.1), and a measure of asymmetry $a_i \equiv |u_i + l_i - r_i|$. The asymmetry measure quantifies the absolute difference between the lengths of the “wicks” above and below the rectangular box of the candlestick. The candlestick is symmetric if and only if $a_i = 0$.

To more clearly highlight the key novelty of our approach, we first focus on

⁸Alternatively, one could impose more explicit assumptions about the form of the noise, and the way in which the prices observed at ultra high frequencies differ from the efficient prices. However, it is far from obvious how the noise component should be modeled, plus the “right” choice is invariably asset and/or market specific. For instance, are the conditional moments of the noise constant or time-varying; does the noise exhibit conditional and/or unconditional serial dependence; should the noise be treated as “small” (i.e., local-to-zero) or “large;” is the noise correlated with the latent efficient price (see, e.g., Kalnina and Linton (2008), Zhang et al. (2005), Jacod et al. (2017), and Li and Linton (2022)). Further complicating matters, the broader econometrics literature on non-classical measurement errors (see, e.g., Schennach (2020)) also calls into question the “classical” additive separability and mean independence assumptions routinely invoked in the high-frequency econometrics literature, and instead suggests that the noise may be better accounted for using non-classical models (as in, e.g., Berkson (1950) and Hyslop and Imbens (2001)). Hence, while it is conceivable that the new approach developed here could be extended to allow for the use of ultra high-frequency data by explicating the “fine structure” of the noise, any associated theoretical efficiency claims would come with the perhaps even more challenging task of justifying the *additional* requisite assumptions.

Prototypical Candlesticks

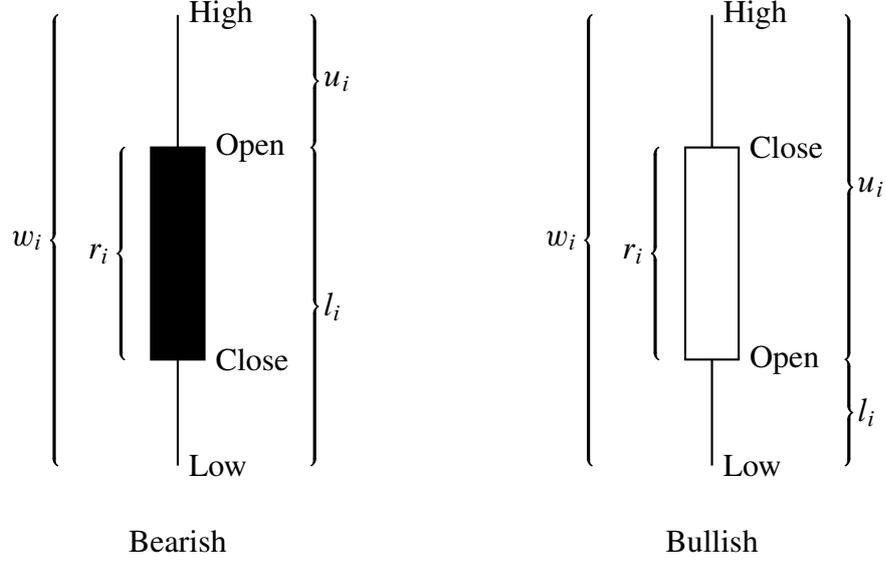


Figure 2.1: The figure shows two prototypical candlesticks, bearish and bullish, comprised of the open, high, low, and close prices. The corresponding return r_i , range w_i , high return u_i , and low return l_i , as defined in equation (2.2), are explicitly highlighted.

estimators based on a single high-frequency candlestick “neighboring” t in the sense that $|i \Delta_n - t| = o(1)$.⁹ Optimal estimation with multiple adjacent candlesticks is discussed in Section 2.2.4. Accordingly, we will express our estimators for σ_t^p generically as

$$S = f(r_i, u_i, l_i), \tag{2.3}$$

for some function $f(\cdot)$. Since spot volatility is fundamentally a “scale parameter,” we will restrict our attention to *scale-equivariant* estimators, requiring the function $f(\cdot)$ to be homogeneous of degree p , that is, $f(\lambda x) = \lambda^p f(x)$ for any $\lambda > 0$. We will further refer to the estimator as *regular* if $f(\cdot)$ is continuous (Lebesgue) almost everywhere. This regularity requirement seems rather innocuous. However, it ensures that any candidate estimator has a limit distribution that is also scale-equivariant.¹⁰ As shown in Theorem 2.1 below, it also proves sufficient to “couple” the original nonparametric estimation problem with a much simpler limit decision problem.

⁹Note, the index i generally also depends on n . We purposely suppress this dependence in our notation so as to avoid nested subscripts.

¹⁰This mirrors the notion of regularity in Gaussian shift limit experiments that requires the estimator to be asymptotically location-equivariant (see, e.g., Van der Vaart (1998)).

In the analysis of scale estimation problems, it is also standard to gauge the estimator's performance by a scale-invariant loss function. For any non-negative loss function $L(\cdot)$ this is readily achieved by considering the scaled loss $L(S/\sigma_t^P)$. Correspondingly, the *risk* of the estimator S may be succinctly expressed as

$$R(S; L) \equiv \mathbb{E}[L(S/\sigma_t^P)]. \quad (2.4)$$

It is impossible to obtain an optimal estimator that minimizes $R(S; L)$ under the general nonparametric model in (2.1). Intuitively, as the joint distribution of the data vector (r_i, u_i, l_i) is determined by the unknown joint law of the (b, σ, W, J) process this would essentially amount to an optimization problem with an infinite-dimensional nuisance parameter. Importantly, however, under mild regularity conditions on the price process, the multiplicative estimation error S/σ_t^P may be shown to be *asymptotically pivotal* for any regular scale-equivariant estimator. Consequently, the asymptotic loss and risk are both nuisance-free, simplifying the optimality analysis.

The following regularity condition, which is standard in the literature on non-parametric volatility estimation (see, e.g., [Jacod and Protter \(2012\)](#), [Jacod et al. \(2021\)](#), [Bollerslev et al. \(2021\)](#), and [Li et al. \(2022\)](#)) suffices for this pivotalization scheme to obtain.

Assumption 1. *Suppose that the price process P has the form in (2.1) 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 finite constants such that the following conditions hold for each $m \geq 1$: (i) for all $t \in [0, T_m]$, $|b_t| + |\sigma_t| + |\sigma_t|^{-1} + F_t(\mathbb{R} \setminus \{0\}) \leq K_m$, where F_t denotes the spot Lévy measure of J ; (ii) for some constant $\kappa > 0$, $\mathbb{E}[|\sigma_{t \wedge T_m} - \sigma_{s \wedge T_m}|^2] \leq K_m |t - s|^{2\kappa}$ for all $s, t \in [0, T]$.*

Assumption 1 necessitates that various processes are bounded by a finite constant K_m up to a stopping time T_m , without requiring the bound to hold over the entire sample span. This setup is commonly employed when applying localization, a standard technique in stochastic calculus used for extending limit theorems under weaker conditions. For a comprehensive discussion on its application in the analysis of high-frequency data, see, e.g., Section 4.4.1 in [Jacod and Protter \(2012\)](#). The parameter κ , defined as the Hölder continuity index for the volatility process σ

under the L_2 norm, pertains to the smoothness of σ . If the volatility is driven by a Brownian motion, κ is at most $1/2$, and the volatility path is non-differentiable everywhere. This setting differs from typical nonparametric problems, where unknown functions are often assumed to be differentiable of higher order. Values of $\kappa < 1/2$ also permits the volatility to exhibit “rough” paths, as emphasized by Gatheral et al. (2018) among others.

The following theorem stipulates a general asymptotic representation for any regular scale-equivariant estimator $S = f(r_i, u_i, l_i)$ allowed under these mild conditions. By linking the nonparametric estimation problem with that in a limit *non*-Gaussian experiment, the result differs notably from the Gaussian shift experiment commonly used in the analysis of semiparametric efficiency, the estimation of integrated volatility functionals included.

Theorem 2.1. *Under Assumption 1, any regular scale-equivariant estimator $S = f(r_i, u_i, l_i)$ with $|i \Delta_n - t| \rightarrow 0$ may be expressed as*

$$\frac{S}{\sigma_t^p} = f(\zeta_i) + o_p(1), \quad (2.5)$$

where $\zeta_i \equiv (\zeta_{i,r}, \zeta_{i,u}, \zeta_{i,l})$ and

$$\begin{aligned} \zeta_{i,r} &\equiv \frac{W_{i \Delta_n} - W_{(i-1) \Delta_n}}{\sqrt{\Delta_n}}, \\ \zeta_{i,u} &\equiv \frac{\sup_{s \in \mathcal{T}_i} (W_s - W_{(i-1) \Delta_n})}{\sqrt{\Delta_n}}, \\ \zeta_{i,l} &\equiv \frac{\inf_{s \in \mathcal{T}_i} (W_s - W_{(i-1) \Delta_n})}{\sqrt{\Delta_n}}. \end{aligned}$$

The theorem shows that the multiplicative estimation error in S may be decomposed into a nondegenerate leading term $f(\zeta_i)$ and an asymptotically negligible $o_p(1)$ term. The $o_p(1)$ term absorbs various nonparametric biases stemming from the drift, time-variation of volatility, and jumps. If the price was simply a scaled Brownian motion, this term would be identically equal to zero. Importantly, the distribution of the ζ_i random variable that determines the leading $f(\zeta_i)$ term is known in finite samples. To appreciate this point, let B denote a generic copy of the standard Brownian motion on the unit interval $[0, 1]$ with $B_0 = 0$. It then follows that

$$\zeta_i \stackrel{d}{=} \tilde{\zeta} \equiv \left(B_1, \sup_{t \in [0,1]} B_t, \inf_{t \in [0,1]} B_t \right). \quad (2.6)$$

Since this distribution, and by implication the $f(\zeta_i)$ term in (2.5), are both nuisance-free, the multiplicative estimation error S/σ_t^p is therefore also asymptotically pivotal.¹¹

If the loss function $L(\cdot)$ is continuous, Theorem 2.1 further implies an analogous coupling result for the estimation loss

$$L(S/\sigma_t^p) = L(f(\zeta_i)) + o_p(1). \quad (2.7)$$

Following the literature (e.g., Le Cam (1986b) and Van der Vaart (1998)), this naturally suggests defining the asymptotic risk of any regular scale-equivariant estimator as the expected value of the limit loss $L(f(\zeta_i))$. By (2.6) the asymptotic risk may thus be expressed as

$$\tilde{R}(S; L) \equiv \mathbb{E}[L(f(\tilde{\zeta}))]. \quad (2.8)$$

The distribution of $\tilde{\zeta}$ is known in finite sample and so $\tilde{R}(S; L)$ can be readily evaluated for any loss function $L(\cdot)$ and estimator $f(\cdot)$. We will refer to a regular scale-equivariant estimator S as *optimal*, or more precisely as an Asymptotic Minimum-Risk scale-Equivariant (AMRE) estimator, if it minimizes $\tilde{R}(S; L)$. Since this asymptotic risk does not depend on any unknown quantities in the nonparametric model in (2.1), this optimality concept is valid in a *uniform* sense. As such, it also implies asymptotic admissibility and minimaxity (within the class of regular scale-equivariant estimators). Consequently, any suboptimal estimator is necessarily asymptotically inadmissible.

Theorem 2.1 is based on the Itô semimartingale model (2.1), which as previously noted does not explicitly incorporate microstructure noise. However, the same result remains valid if the observed prices are affected by “small” noise. Specifically, if the magnitude of the noise is of order $o_p(\Delta_n^{1/2})$, the “noisy” observation of (r_i, u_i, l_i) deviates from their true value by $o_p(1)$, which, according to the continuous mapping theorem, implies that (2.5) also holds for the “noisy” estimator. Intuitively, the $o_p(\Delta_n^{1/2})$ rate requirement for the noise is more plausible when Δ_n is not “too-

¹¹This nuisance-free limit distribution also permits the construction of asymptotically valid confidence intervals for σ_t^p . For any $\alpha \in (0, 1)$, let L and U be constants such that $\mathbb{P}(L \leq 1/f(\tilde{\zeta}) \leq U) = 1 - \alpha$. Then $[LS, US]$ is a confidence interval for σ_t^p at asymptotic level $1 - \alpha$. The length of the interval is minimized by taking $[L, U]$ as the $1 - \alpha$ level highest density interval of the distribution of $1/f(\tilde{\zeta})$.

small,” consistent with the idea of not using “too-finely” sampled data, or coarse sampling.¹²

In order to construct an AMRE estimator, it is helpful to recognize that the asymptotic risk of a regular estimation function $f(\cdot)$ can be equivalently represented in terms of its finite-sample risk in a limit parametric model, where the (log) price process P is a simple scaled Brownian motion (i.e., $P_t = \sigma W_t$). This, in turn, facilitates the use of classical finite-sample theory for optimal equivariant estimation in determining the optimal estimation function and the AMRE estimator.¹³ The AMRE estimators presented in Section 2.2.2 and Section 2.2.4 below, as well as the additional estimators discussed in the Supplemental Appendix A.2, are developed using this approach.

2.2.2 Optimal Estimation for Spot Variance and Volatility

To streamline the presentation and more clearly highlight our main theoretical contributions, we will focus our discussion on the optimal estimation of the spot variance σ_t^2 and the spot volatility σ_t .¹⁴ We will restrict our attention to optimal estimators based on Stein’s loss and the standardized quadratic loss,

$$L_{\text{Stein}}(x) \equiv x - \log(x) - 1, \quad L_{\text{Quad}}(x) \equiv |x - 1|^2. \quad (2.9)$$

These specific loss functions arguably constitute the two most commonly used losses in practice. AMRE estimators for other, possibly case-specific, loss functions could be derived similarly.

To facilitate the representation of the optimal estimators, it is helpful to define the functions $G_q(\cdot)$ and $H_q(\cdot)$ for any integer $q \geq 0$ as,

$$G_q(x) \equiv \psi_q\left(\frac{1-x}{2}\right) + \psi_q\left(\frac{1+x}{2}\right) - \frac{x}{q+1} \left(\psi_{q+1}\left(\frac{1-x}{2}\right) - \psi_{q+1}\left(\frac{1+x}{2}\right) \right)$$

¹²In line with the existing empirical literature on high-frequency-based volatility estimation, we recommend adopting a $\Delta_n = 5$ -minute sampling scheme as the default choice in practice. Simulation results in the Supplemental Appendix A.2 also demonstrate that the noise, when calibrated to empirically realistic levels, has a negligible effect on the resulting 5-minute estimators. As the noise level becomes higher, the noise leads to larger positive bias in the volatility estimates, and the shrinkage estimator derived from minimizing the quadratic loss tends to outperform other estimators.

¹³According to Corollary 3.3.4 in Lehmann and Casella (1998), the solution to the functional minimization problem $\min_f \mathbb{E}[L(f(\tilde{\zeta}))]$ exists and is unique, provided that an equivariant estimator with finite risk exists and the function $x \mapsto L(e^x)$ is strictly convex and not monotone.

¹⁴Analogous derivations for the spot quarticity σ_t^4 and spot precision σ_t^{-1} are provided in the Supplemental Appendix A.2.

$$-\frac{1-x^2}{4(q+1)(q+2)} \left(\psi_{q+2} \left(\frac{1-x}{2} \right) + \psi_{q+2} \left(\frac{1+x}{2} \right) \right), \quad (2.10)$$

$$\begin{aligned} H_q(x) \equiv & \psi_q \left(1 - \frac{x}{2} \right) + \psi_q \left(\frac{x}{2} \right) - \frac{x}{q+1} \left(\psi_{q+1} \left(1 - \frac{x}{2} \right) - \psi_{q+1} \left(\frac{x}{2} \right) \right) \\ & + \frac{x^2}{4(q+1)(q+2)} \left(\psi_{q+2} \left(1 - \frac{x}{2} \right) + \psi_{q+2} \left(\frac{x}{2} \right) \right), \end{aligned} \quad (2.11)$$

where $\psi_q(x)$ denotes the polygamma function of order q , that is, the $(q+1)$ th-order derivative of the logarithm of the Gamma function. The $G_q(\cdot)$ and $H_q(\cdot)$ functions are both continuous almost everywhere, making them suitable for constructing regular estimators.¹⁵ Using these definitions, the subsequent theorem offers explicit analytical expressions for the AMRE estimators of the spot variance and spot volatility under each of the two loss functions.

Theorem 2.2. *Under the same setting as Theorem 2.1, we have*

(a) *the AMRE range-based estimator for σ_t^2 under Stein's loss is asymptotically unbiased and given by*

$$\hat{\sigma}_{\text{Stein}}^2 \equiv \frac{4w_i^2}{3} \cdot \frac{G_0(a_i/w_i) - H_0(|r_i|/w_i)}{G_2(a_i/w_i) - H_2(|r_i|/w_i)},$$

while the AMRE range-based estimator for σ_t^2 under standardized quadratic loss equals

$$\hat{\sigma}_{\text{Quad}}^2 \equiv \frac{32w_i^2}{5} \cdot \frac{G_2(a_i/w_i) - H_2(|r_i|/w_i)}{G_4(a_i/w_i) - H_4(|r_i|/w_i)},$$

(b) *the AMRE range-based estimator for σ_t under Stein's loss is asymptotically unbiased and given by*

$$\hat{\sigma}_{\text{Stein}} \equiv \frac{\sqrt{2\pi}}{3} w_i \cdot \frac{G_0(a_i/w_i) - H_0(|r_i|/w_i)}{H_1(|r_i|/w_i) - G_1(a_i/w_i)},$$

while the AMRE range-based estimator for σ_t under standardized quadratic loss equals

$$\hat{\sigma}_{\text{Quad}} \equiv 2\sqrt{\frac{2}{\pi}} w_i \cdot \frac{H_1(|r_i|/w_i) - G_1(a_i/w_i)}{G_2(a_i/w_i) - H_2(|r_i|/w_i)}.$$

COMMENT 1. The asymptotic unbiasedness of the $\hat{\sigma}_{\text{Stein}}^2$ and $\hat{\sigma}_{\text{Stein}}$ estimators is reminiscent of the classical finite-sample result that minimum-risk estimators under Stein's loss are guaranteed to be unbiased. As demonstrated by [Brown \(1968\)](#), Stein's loss is also the unique loss function (up to affine transformations) that sat-

¹⁵The almost everywhere continuity of the $G_q(\cdot)$ and $H_q(\cdot)$ functions follows from the fact that polygamma functions are formally meromorphic, meaning that they are analytic except for a discrete set of points.

isfies this property. Consequently, AMRE estimators under other loss functions are necessarily asymptotically biased.

COMMENT 2. The AMRE estimators depend solely on the shape of the candlestick, as summarized by the scaled range $w_i \equiv u_i - l_i$, the scaled absolute return $|r_i|$, and the a_i asymmetry measure. These shape-related features remain unaffected by a “color change” or an “upside-down flip” of the candlestick.¹⁶ Consequently, the optimal volatility estimators are also invariant to these “directional” transformations. This feature reduction is due to a sufficiency argument, as formalized by Lemma A.1 in the Appendix, which shows that the shape features $(w_i, |r_i|, a_i)$ are indeed sufficient statistics for σ in the limit model $P_t = \sigma W_t$. Recall that according to the Rao–Blackwell theorem, optimal estimators depend on data only through sufficient statistics.

The AMRE estimation functions defined in Theorem 2.2 are relatively complex. Clearly, it would have been challenging to accurately “intuit” these specific functional forms when searching for optimal estimators within a restricted class of functions. Since the AMRE estimators are uniquely determined (Corollary 3.3.4 in Lehmann and Casella (1998)), any ad hoc restrictions on the functional form used in the derivation of “shape-constrained” optimal estimators would therefore also generally result in suboptimal and, as previously mentioned, asymptotically inadmissible estimators.

For a more direct comparison, recall that Garman and Klass’s (1980) minimum-variance unbiased quadratic estimator for spot variance is given by¹⁷

$$\begin{aligned}\hat{\sigma}_{\text{GK}}^2 &\equiv 0.511(u_i - l_i)^2 - 0.019(r_i(u_i + l_i) - 2u_i l_i) - 0.383r_i^2 \\ &= 0.5015w_i^2 + 0.0095a_i^2 - 0.3925r_i^2,\end{aligned}$$

while the BLUE for spot volatility proposed by Li et al. (2022) is given by

$$\hat{\sigma}_{\text{BLUE}} \equiv 0.811w_i - 0.369|r_i|.$$

Meanwhile, approximating the functional forms of the AMRE estimators described

¹⁶More precisely, the color change corresponds to changing the sign of r_i and the upside-down flip amounts to swapping the upper and lower shadows of the candlestick.

¹⁷A simplified “practical” version of the Garman–Klass estimator, defined by $0.5w_i^2 - (2\log(2) - 1)r_i^2 \approx 0.5w_i^2 - 0.386r_i^2$, has also sometimes been used in empirical applications.

in Theorem 2.2 by cubic polynomials of the (maximal invariant) ratio statistics, $|r_i|/w_i$ and a_i/w_i , the spot variance estimator may be expressed as¹⁸

$$\begin{aligned}\hat{\sigma}_{\text{Stein}}^2 &\approx 0.5921w_i^2 - 0.2066|r_i|w_i - 0.1289a_i^2 - 0.5874r_i^2 - 0.0001\frac{a_i^3}{w_i} \\ &\quad + 0.0382\frac{|r_i|a_i^2}{w_i} - 0.0001\frac{r_i^2a_i}{w_i} + 0.3872\frac{|r_i|^3}{w_i}, \\ \hat{\sigma}_{\text{Quad}}^2 &\approx 0.4936w_i^2 - 0.0002a_iw_i - 0.2436|r_i|w_i - 0.1003a_i^2 + 0.0001|r_i|a_i \\ &\quad - 0.4316r_i^2 - 0.0006\frac{a_i^3}{w_i} + 0.0883\frac{|r_i|a_i^2}{w_i} - 0.0005\frac{r_i^2a_i}{w_i} + 0.3282\frac{|r_i|^3}{w_i},\end{aligned}$$

while the analogous approximations for the AMRE spot volatility estimators take the form

$$\begin{aligned}\hat{\sigma}_{\text{Stein}} &\approx 0.7859w_i - 0.1010|r_i| - 0.0888\frac{a_i^2}{w_i} - 0.4798\frac{r_i^2}{w_i} - 0.0178\frac{a_i^2|r_i|}{w_i^2} \\ &\quad + 0.2341\frac{|r_i|^3}{w_i^2}, \\ \hat{\sigma}_{\text{Quad}} &\approx 0.7526w_i - 0.1366|r_i| - 0.0846\frac{a_i^2}{w_i} - 0.0001\frac{a_i|r_i|}{w_i} - 0.4345\frac{r_i^2}{w_i} \\ &\quad + 0.0181\frac{a_i^2|r_i|}{w_i^2} - 0.0001\frac{a_i r_i^2}{w_i^2} + 0.2284\frac{|r_i|^3}{w_i^2}.\end{aligned}$$

While not exact, these cubic expansions formally highlight the differences between the AMRE estimators and the shape-constrained estimators, by explicating the former's dependence on additional higher-order nonlinear features.

To help more clearly visualize these differences, Figure 2.2 present the estimation functions for the three spot variance estimators: $\hat{\sigma}_{\text{Stein}}^2$, $\hat{\sigma}_{\text{Quad}}^2$, and $\hat{\sigma}_{\text{GK}}^2$. As the estimators are all scale-equivariant, we compare them without loss of generality under the scale normalization $w_i = 1$ (i.e., $|r_i|$ and a_i are interpreted in a relative sense). In the left panel of Figure 2.2, we further fix the asymmetry factor at $a_i = 0$, and plot the spot variance estimators as functions of the absolute return $|r_i|$. Looking at the two asymptotically unbiased estimators, $\hat{\sigma}_{\text{Stein}}^2$ and $\hat{\sigma}_{\text{GK}}^2$, the former is higher when $|r_i|$ is close to 0 or 1, and lower when $|r_i|$ takes medium values.¹⁹ Meanwhile, the estimation function associated with $\hat{\sigma}_{\text{Quad}}^2$ is systematically below

¹⁸The approximation for $\hat{\sigma}_{\text{Stein}}^2$ is constructed by projecting $(G_0(a_i/w_i) - H_0(|r_i|/w_i))/(G_2(a_i/w_i) - H_2(|r_i|/w_i))$ onto a cubic polynomial of the maximal invariant $(|r_i|/w_i, a_i/w_i)$ under the L_2 distance. The approximations for the other AMRE estimators are obtained similarly.

¹⁹As a point of reference, in the Brownian limit model, the interquartile range of $|r_i|/w_i$ spans 0.243 to 0.676, while the interdecile range covers 0.099 to 0.817.

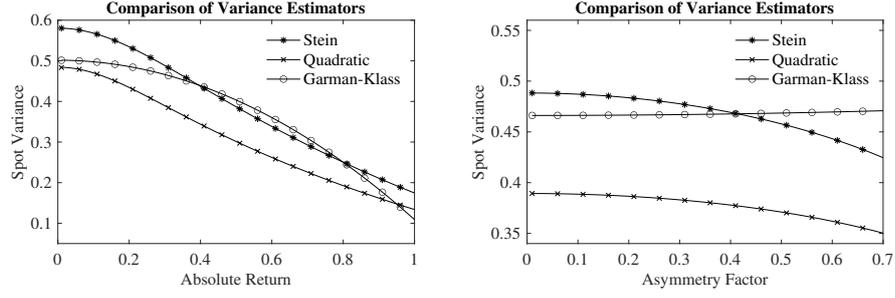


Figure 2.2: Alternative Range-Based Variance Estimators. The figure plots the AMRE estimators for the variance under Stein’s loss (Stein) and quadratic loss (Quadratic), together with the Garman–Klass estimator. The range w_i is normalized to unity. The left panel plots the spot variance estimator as a function of the absolute return $|r_i|$, with the asymmetry factor fixed at $a_i = 0$. The right panel plots the spot variance estimator as a function of the asymmetry factor a_i , with the absolute return fixed at $|r_i| = 0.3$.

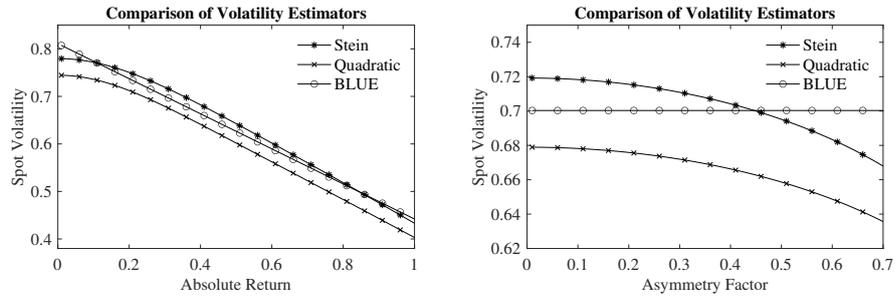


Figure 2.3: Alternative Range-Based Volatility Estimators. The figure plots the AMRE estimators for the volatility under Stein’s loss (Stein) and quadratic loss (Quadratic), together with the BLUE estimator proposed by Li et al. (2022). The range w_i is normalized to unity. The left panel plots the spot volatility estimator as a function of the absolute return $|r_i|$, with the asymmetry factor fixed at $a_i = 0$. The right panel plots the spot volatility estimator as a function of the asymmetry factor a_i , with the absolute return fixed at $|r_i| = 0.3$.

the estimation functions for the two unbiased estimators, indicating that the AMRE estimator under quadratic loss exhibits a certain “shrinkage” and therefore also is downward biased.

In the right panel of Figure 2.2, we fix $|r_i| = 0.3$ and plot the estimators as functions of the asymmetry factor a_i .²⁰ While $\hat{\sigma}_{\text{GK}}^2$ displays a slightly positive dependence on the asymmetry factor, the two AMRE estimators evidence a more pronounced negative dependence. Comparing the left and right panels further reveals that the absolute return has a greater impact on variance estimation than the asymmetry factor.

A similar comparison for the three spot volatility estimators, $\hat{\sigma}_{\text{Stein}}$, $\hat{\sigma}_{\text{Quad}}$, and $\hat{\sigma}_{\text{BLUE}}$, is provided in Figure 2.3. The overall patterns generally mirror those of Figure 2.2. Underscoring the difference between the two AMRE and the BLUE estimator of Li et al. (2022), which does not depend on a_i , the right panel clearly shows that the two optimal estimators both depend negatively, and nontrivially, on the asymmetry factor.

We turn next to a more direct assessment of how these differences in the functional forms of the estimators translate into asymptotic biases, variances, and differences in Stein and quadratic risks.

2.2.3 Risk Comparisons

We will focus our comparisons of the risks of the different estimators by considering the relative efficiency, defined as the ratio between the risk of the relevant AMRE estimator and the estimator under consideration. Table 2.1 reports the results for the three spot variance estimators: $\hat{\sigma}_{\text{Stein}}^2$, $\hat{\sigma}_{\text{Quad}}^2$, and $\hat{\sigma}_{\text{GK}}^2$. Since $\hat{\sigma}_{\text{BLUE}}$ is the BLUE for spot volatility, we also include $(\hat{\sigma}_{\text{BLUE}})^2$ as a fourth contender for estimating the spot variance. Table 2.2 presents the analogous results for the three spot volatility estimators: $\hat{\sigma}_{\text{Stein}}$, $\hat{\sigma}_{\text{Quad}}$, and $\hat{\sigma}_{\text{BLUE}}$. For comparison, we also include the transformed $(\hat{\sigma}_{\text{GK}}^2)^{1/2}$ spot volatility estimator.²¹

²⁰Since a_i measures the absolute difference between the lengths of the upper and lower shadows of the candlestick, it takes values in $[0, w_i - |r_i|]$.

²¹All of the numbers are computed numerically based on ten million Monte Carlo draws of a standard Brownian motion $(B_t)_{t \in [0,1]}$ (recall (2.6)) as the simulated sample averages corresponding to $\mathbb{E}[f(\tilde{\zeta})] - 1$, $\text{Var}(f(\tilde{\zeta}))$, and $\mathbb{E}[L(f(\tilde{\zeta}))]$, respectively.

Table 2.1: Asymptotic Properties of Spot Variance Estimators

Estimator	Bias	Variance	Relative Efficiency	
			Stein	Quadratic
$\hat{\sigma}_{\text{Stein}}^2$	0.000	0.259	1.000	0.803
$\hat{\sigma}_{\text{Quad}}^2$	-0.205	0.165	0.813	1.000
$\hat{\sigma}_{\text{GK}}^2$	0.000	0.270	0.968	0.770
$(\hat{\sigma}_{\text{BLUE}})^2$	0.062	0.295	0.976	0.698

Note: The table reports the asymptotic biases, variances, and relative efficiency under Stein's and quadratic risks for each of the spot variance estimators indicated in the first column.

Table 2.2: Asymptotic Properties of Spot Volatility Estimators

Estimator	Bias	Variance	Relative Efficiency	
			Stein	Quadratic
$\hat{\sigma}_{\text{Stein}}$	0.000	0.061	1.000	0.967
$\hat{\sigma}_{\text{Quad}}$	-0.058	0.055	0.909	1.000
$(\hat{\sigma}_{\text{GK}}^2)^{1/2}$	-0.030	0.060	0.938	0.952
$\hat{\sigma}_{\text{BLUE}}$	0.000	0.063	0.968	0.937

Note: The table reports the asymptotic biases, variances, and relative efficiency under Stein's and quadratic risks for each of the spot volatility estimators indicated in the first column.

Looking first at Table 2.1, the optimal $\hat{\sigma}_{\text{Quad}}^2$ estimator exhibits substantial downward asymptotic bias. This “shrinkage” feature is attributable to the fact that the quadratic loss assigns a heavier penalty to overestimation than underestimation, and as such the corresponding optimal estimator naturally sacrifices some downward bias in order to further reduce the variance. Indeed, the asymptotic variance of the $\hat{\sigma}_{\text{Quad}}^2$ estimator is notably lower than the corresponding numbers for all of the other estimators. Compared to the optimal estimator $\hat{\sigma}_{\text{Quad}}^2$, the relative efficiencies of the shape-constrained $\hat{\sigma}_{\text{GK}}^2$ and $(\hat{\sigma}_{\text{BLUE}})^2$ estimators equal 77.0% and 69.8% respectively.²²

Although the shape-constrained estimators clearly demonstrate suboptimal performance under quadratic loss, they exhibit “near-optimal” behavior under Stein’s loss. Specifically, the relative efficiencies of $\hat{\sigma}_{\text{GK}}^2$ and $(\hat{\sigma}_{\text{BLUE}})^2$ are 96.8% and 97.6%, respectively, when compared to the AMRE $\hat{\sigma}_{\text{Stein}}^2$. In other words, in scenarios where an economic agent’s loss function closely resembles Stein’s loss, the classical Garman–Klass and the BLUE estimators are both reasonable practical choices. Of course, since the AMRE estimator can also easily be calculated in practice using our explicit closed form solution, there is really no need to suffer any efficiency loss, however small it might be.

Turning to Table 2.2 and spot volatility estimation, the results again evidence notable bias for the $\hat{\sigma}_{\text{Quad}}$ estimator. In general, the efficiency gaps between the shape-constrained volatility estimators and the optimal estimator are smaller than for variance estimation. Intuitively, the optimal estimation of σ_t is “easier” than the optimal estimation of its nonlinear transform σ_t^2 .²³ As such, the Garman–Klass estimator and the simple linear estimator proposed by Li et al. (2022) turn out to perform quite well for spot volatility estimation under both quadratic and Stein’s loss functions, although both estimators, strictly speaking, are inadmissible.

²²Interestingly, $\hat{\sigma}_{\text{Stein}}^2$ exhibits lower quadratic risk than $\hat{\sigma}_{\text{GK}}^2$. Since both of these two estimators are asymptotically unbiased, this suggests that under quadratic loss the Garman–Klass estimator is asymptotically inadmissible, not only within the class of regular scale-equivariant estimators, but also within the subclass of asymptotically unbiased estimators.

²³Consistent with this intuition, Tables A.1 in the Supplemental Appendix A.2 shows that the efficiency gap between the AMRE estimators and the shape-constrained estimators is also larger for the quarticity σ_t^4 , an even “more nonlinear” transform of the volatility than the variance. For example, under quadratic loss, the relative efficiencies of $(\hat{\sigma}_{\text{GK}}^2)^2$ and $(\hat{\sigma}_{\text{BLUE}})^4$ are only 31.2% and 25.5%, respectively, in comparison with the AMRE estimator for σ_t^4 .

More broadly, these numerical comparisons also demonstrate that the relative asymptotic risks of alternative estimators, and in turn the design of optimal estimators, can depend quite strongly on the underlying loss function. This reflects the finite-sample nature of our coupling-based asymptotic analysis in a non-Gaussian limit experiment. By contrast, in the conventional “large sample” asymptotic setting with Gaussian shift limit experiments, different loss functions (as long as they are bowl-shaped) result in the same optimal estimators (see, e.g., Chapter 8 in [Van der Vaart \(1998\)](#)).

Acknowledging the practical challenge of precisizing the loss function in some applications, we observe an intriguing pattern for the risk comparisons in [Tables 2.1](#) and [2.2](#). In both tables, the AMRE estimators derived under Stein’s loss exhibit lower risks than the shape-constrained estimators, not only under Stein’s loss (which holds by construction), but also under quadratic loss.²⁴ Hence, for users who are uncertain about their specific loss function, we recommend employing $\hat{\sigma}_{\text{Stein}}^2$ and $\hat{\sigma}_{\text{Stein}}$ as “general purpose” estimators for spot variance and spot volatility estimation, respectively.

2.2.4 Optimal Estimators with Multiple Candlesticks

The estimators discussed above rely on a single candlestick for optimally estimating the spot volatility, or the spot variance. In this section we describe how to combine multiple adjacent candlesticks (over asymptotically shrinking time intervals) for optimally estimating σ_t^p .

To set out the notation, given a fixed integer $k \geq 1$, let

$$\mathbf{c}_i = (r_{i+j-1}, u_{i+j-1}, l_{i+j-1})_{1 \leq j \leq k},$$

collect the observed features of k successive candlesticks starting at the i th observation. Denote the corresponding estimator for σ_t^p that utilizes the k candlesticks by $f(\mathbf{c}_i)$. A direct extension of [Theorem 2.1](#) produces the following analogous coupling result for the k -candlestick setting.

Corollary 2.1. *Under [Assumption 1](#), given any fixed integer $k \geq 1$, any regular*

²⁴Additional results reported in the [Supplemental Appendix A.2](#) show that this phenomenon remains true for estimating the spot quarticity, σ_t^4 , and the spot precision, σ_t^{-1} .

scale-equivariant estimator $S = f(\mathbf{c}_i)$ with $|i \Delta_n - t| \rightarrow 0$ may be expressed as

$$\frac{S}{\sigma_t^p} = f(\zeta_i, \zeta_{i+1}, \dots, \zeta_{i+k-1}) + o_p(1),$$

where the variables ζ_{i+j} , $j = 1, \dots, k$, are defined as in Theorem 2.1.

Building on the same reasoning outlined in Section 2.2.2, we may therefore couple the original estimation problem with that in the Brownian limit experiment. Moreover, by a direct extension of the proof of Theorem 2.2, we can also derive semi-closed form expressions for the AMRE estimators that utilize k successive candlesticks. Concretely, the AMRE estimators under Stein's loss and standardized quadratic loss may be expressed as

$$\begin{aligned} \hat{\sigma}_{\text{Stein}}^p(k) &= w_i^p \cdot \frac{1}{F_{k,p} \left(\frac{|r_i|}{w_i}, \frac{a_i}{w_i}, \frac{w_{i+1}}{w_i}, \frac{|r_{i+1}|}{w_i}, \frac{a_{i+1}}{w_i}, \dots, \frac{w_{i+k-1}}{w_i}, \frac{|r_{i+k-1}|}{w_i}, \frac{a_{i+k-1}}{w_i} \right)}, \\ \hat{\sigma}_{\text{Quad}}^p(k) &= w_i^p \cdot \frac{F_{k,p} \left(\frac{|r_i|}{w_i}, \frac{a_i}{w_i}, \frac{w_{i+1}}{w_i}, \frac{|r_{i+1}|}{w_i}, \frac{a_{i+1}}{w_i}, \dots, \frac{w_{i+k-1}}{w_i}, \frac{|r_{i+k-1}|}{w_i}, \frac{a_{i+k-1}}{w_i} \right)}{F_{k,2p} \left(\frac{|r_i|}{w_i}, \frac{a_i}{w_i}, \frac{w_{i+1}}{w_i}, \frac{|r_{i+1}|}{w_i}, \frac{a_{i+1}}{w_i}, \dots, \frac{w_{i+k-1}}{w_i}, \frac{|r_{i+k-1}|}{w_i}, \frac{a_{i+k-1}}{w_i} \right)}. \end{aligned} \quad (2.12)$$

The function $F_{k,q} : \mathbb{R}^{3k-1} \rightarrow \mathbb{R}$ that enters these expressions for $q \in \{p, 2p\}$ is formally defined as a conditional expectation function:

$$\begin{aligned} F_{k,q} \left(\frac{|r_i|}{w_i}, \frac{a_i}{w_i}, \frac{w_{i+1}}{w_i}, \frac{|r_{i+1}|}{w_i}, \frac{a_{i+1}}{w_i}, \dots, \frac{w_{i+k-1}}{w_i}, \frac{|r_{i+k-1}|}{w_i}, \frac{a_{i+k-1}}{w_i} \right) \\ \equiv \mathbb{E} \left[\xi_{w,1}^q \left| \begin{array}{l} \xi_{w,j} = \frac{w_{i+j-1}}{w_i}, \xi_{r,j} = \frac{|r_{i+j-1}|}{w_i}, \xi_{a,j} = \frac{a_{i+j-1}}{w_i} \\ \text{for all } 1 \leq j \leq k \end{array} \right. \right], \end{aligned} \quad (2.13)$$

where $(\xi_{w,j}, \xi_{r,j}, \xi_{a,j})$, $j = 1, 2, \dots, k$, are independent copies of

$$\left(\sup_{t \in [0,1]} B_t - \inf_{t \in [0,1]} B_t, |B_1|, \left| \sup_{t \in [0,1]} B_t + \inf_{t \in [0,1]} B_t - B_1 \right| \right). \quad (2.14)$$

In parallel to the optimal estimators that rely on a single candlestick, the two AMRE estimators defined in (2.12) are also structurally similar. The w_i^p component, in particular, acts as a generic scale-equivariant estimator for σ_t^p , while the $F_{k,q}(\cdot)$ function depends on candlestick observations solely through the maximal invariant statistics, defined as the shape features $(w_{i+j-1}, |r_{i+j-1}|, a_{i+j-1})_{1 \leq j \leq k}$ normalized by w_i . The earlier closed form solutions for the single-candlestick case,

or $k = 1$, detailed in Theorem 2.2 were obtained by explicitly deriving the functional form of $F_{1,q}(\cdot)$. Regrettably, analytical solutions for $F_{k,q}(\cdot)$ for $k \geq 2$ are currently unattainable.

Nonetheless, the semi-closed form solutions in (2.12) still suggest a strategy for numerically computing the optimal estimation functions. In particular, since $F_{k,q}(\cdot)$ is defined as the conditional expectation of w_i^q given the maximal invariant statistics under the limit experiment, one may simulate the $(\xi_{w,j}, \xi_{r,j}, \xi_{a,j})_{1 \leq j \leq k}$ variables as i.i.d. copies of the Brownian functionals defined in (2.14) and then calculate the requisite conditional expectation function in (2.13) via Monte Carlo integration. This calculation, which formally entails the formation of a predictor that minimizes the mean squared error, may be conveniently implemented using popular machine learning tools such as neural networks, or random forests. We stress that for a given k and q , the function $F_{k,q}(\cdot)$ only needs to be computed once.

To illustrate the idea, consider the case with two candlesticks, or $k = 2$. Employing a neural network to compute the conditional expectation functions $F_{2,p}(\cdot)$ and $F_{2,2p}(\cdot)$ numerically, Tables 2.3 and 2.4 report the resulting asymptotic bias, variance, and relative efficiency for the AMRE estimators for estimating the spot variance and volatility, respectively, obtained by using these numerical approximations in place of the true unknown functions in (2.12).²⁵ The $k = 2$ versions of the shape-constrained $\hat{\sigma}_{\text{GK}}^2$ and $\hat{\sigma}_{\text{BLUE}}$ estimators, also included in the tables, are constructed as simple averages of their respective single-candlestick estimates, following the suggestion of Li et al. (2022).

The general results are qualitatively very similar to the ones for the single-candlestick estimators previously reported in Tables 2.1 and 2.2. The optimal estimators are notably more accurate, especially for estimating the spot variance under quadratic loss. At the same time, the “near optimality” property of the shape-constrained estimators under Stein’s loss observed for the single-candlestick case

²⁵More specifically, we rely on a logistic sigmoid activation function, and an architecture comprised of an input layer with 32 neurons, followed by two hidden layers with 16 and 8 neurons, respectively. We train the model based on five million random draws of $(\xi_{w,j}, \xi_{r,j}, \xi_{a,j})_{j=1,2}$, where the Brownian motion $(B_t)_{t \in [0,1]}$ is generated using an Euler scheme with a mesh size of 10^{-7} . Underscoring the accuracy of the approach, using the same procedures to calculate the functions for $k = 1$ results in numerical solutions that are practically indistinguishable from the closed form solutions detailed in Theorem 2.2.

Table 2.3: Asymptotic Properties of Spot Variance Estimators Based on Two Candlesticks

Estimator	Bias	Variance	Relative Efficiency	
			Stein	Quadratic
$\hat{\sigma}_{\text{Stein}}^2(2)$	0.000	0.128	1.000	0.891
$\hat{\sigma}_{\text{Quad}}^2(2)$	-0.103	0.103	0.833	1.000
$\hat{\sigma}_{\text{GK}}^2$	0.000	0.135	0.923	0.844
$(\hat{\sigma}_{\text{BLUE}})^2$	0.062	0.147	0.923	0.755

Note: The table reports the asymptotic biases, variances, and relative efficiency under Stein's and quadratic risks for each of the spot variance estimators indicated in the first column.

Table 2.4: Asymptotic Properties of Spot Volatility Estimators Based on Two Candlesticks

Estimator	Bias	Variance	Relative Efficiency	
			Stein	Quadratic
$\hat{\sigma}_{\text{Stein}}(2)$	0.000	0.030	1.000	0.966
$\hat{\sigma}_{\text{Quad}}(2)$	-0.025	0.029	0.939	1.000
$(\hat{\sigma}_{\text{GK}}^2)^{1/2}$	-0.030	0.030	0.940	0.935
$\hat{\sigma}_{\text{BLUE}}$	0.000	0.031	0.942	0.935

Note: The table reports the asymptotic biases, variances, and relative efficiency under Stein's and quadratic risks for each of the spot volatility estimators indicated in the first column.

does not appear to hold as well. For instance, the relative efficiency of the Garman–Klass variance estimator drops from 96.8% in the $k = 1$ case to 92.3% in the $k = 2$ case, while the relative efficiency of the BLUE volatility estimator drops from 96.8% to 94.2%. These findings further motivate the use of the AMRE estimators in practice.

Putting the results in Tables 2.3 and 2.4 further into perspective, it is, of course, not surprising that the spot estimators constructed by combining two candlesticks exhibit smaller theoretical asymptotic variances than their single-candlestick counterparts. At the same time, the temporal aggregation of multiple candlesticks can easily be harmful in practice, especially when the volatility fluctuates rapidly. In such situations, the limit experiment with constant volatility that formally underlies the theoretical asymptotic comparisons will likely not provide a good finite-sample guide either. Of course, this type of empirical scenario is precisely when spot estimation can be most useful and informative. The empirical application discussed in the next section further highlights these issues.

2.3 An Empirical Illustration

To demonstrate the practical applicability and insights afforded by the new optimal estimators, we present spot volatility estimates for a market portfolio on the eight 2022 prescheduled FOMC announcement days. Putting the results into perspective, at the start of the year U.S. inflation had already soared to its highest level since the 1980s. In response to this, the Federal Reserve indicated at its January 2022 meeting that it would soon be appropriate to raise the target range for the federal funds rate. Subsequently, the target rate was indeed increased by 25 basis points (bps) in March, followed by a more substantial 50 bps hike in May. The pace of rate increases further accelerated to 75 bps for the next four meetings, before moderating to a 50 bps rise at the final 2022 meeting in December. Each of these rate increases were detailed in a short formal release by the FOMC at exactly 14:00 EST, followed by additional comments and a press conference led by Federal Reserve Chairman Jerome Powell starting half-an-hour later.

It is well established that financial markets often reacts quite strongly to the

initial 14:00 FOMC announcement.²⁶ It is much less clear, however, what happens to market volatility at the exact time of the FOMC announcement, let alone in its immediate aftermath and during the subsequent press conference. We shed new light on this issue by utilizing intraday candlestick data for the S&P 500 index, in the form of the VOO exchange-traded fund (ETF) managed by the Vanguard Group, to estimate high-frequency spot volatilities. Guided by the simulation results discussed above, to mitigate the impact of microstructure noise, we employ 5-minute VOO candlesticks, sourced directly from Bloomberg. We focus our analyses on the 5-minute $\hat{\sigma}_{\text{Stein}}$ AMRE spot volatility estimator; comparisons with the other estimators discussed above are presented in the Supplemental Appendix A.2.²⁷

Figure 2.4 displays the resulting estimates, together with 90% level pointwise confidence intervals. To facilitate comparison across the different days, all of the plots are presented on a uniform percentage daily scale. As the figure shows, the market volatility generally spikes immediately following the initial FOMC announcement at 14:00.²⁸ The volatility then generally reverts towards a more “normal” level in the half-hour window between the initial release and the start of the press conference. By comparison, the volatility patterns observed during the press conference appear less systematic. In addition to reiterating key policy decisions, also summarized in the initial release, the press conference and the subsequent interaction with the media often provide additional forward guidance about future Fed policies, interspersed with comments about the general economic outlook as perceived by the Fed. The staggered information flow delivery throughout this process naturally manifest in event specific volatility spikes linked to the exact timing of the new information.

²⁶Andersen et al. (2007), Lee and Mykland (2008), Lee (2012), and Bollerslev et al. (2018), among others, have previously associated high-frequency jumps in asset prices with FOMC announcements. FOMC announcements have also been used as a powerful tool for the high-frequency identification of monetary policy shocks, as exemplified by Cochrane and Piazzesi (2002), Rigobon and Sack (2004), Bernanke and Kuttner (2005), and Nakamura and Steinsson (2018a), while Savor and Wilson (2014), Lucca and Moench (2015), Cieslak et al. (2019), and Ai and Bansal (2018) have emphasized the significance of an FOMC announcement risk premium and pre-announcement drifts.

²⁷The Supplemental Appendix A.2 also reports analogous results for the Dollar/Yen exchange rate.

²⁸Interestingly, for some of the days, most notably March 16, May 4, and June 15, the volatility actually increased slightly in advance of the official 14:00 announcement. Whether these “abnormal” pre-announcement increases can be attributed to information leaks during the Fed’s official blackout period may warrant further scrutiny.

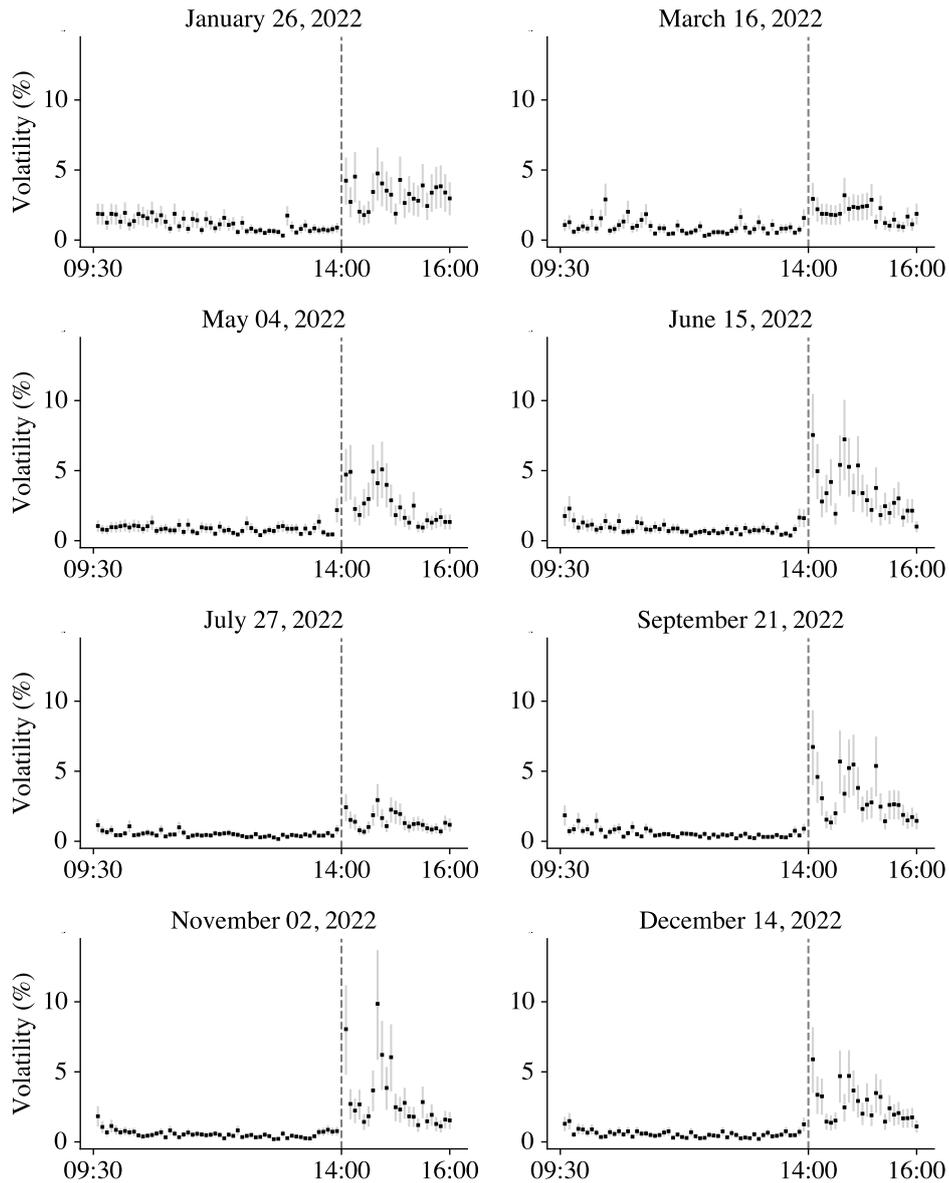


Figure 2.4: Spot Volatility Estimates for the VOO ETF on FOMC Announcement Days. The figure plots the $\hat{\sigma}_{\text{Stein}}$ spot volatility estimates based on 5-minute VOO ETF candlesticks, expressed in daily percentage units. Pointwise confidence intervals at the 90% level is calculated as detailed in footnote 11. The vertical lines included in each of the panels indicate the official 14:00 FOMC announcement times.

The November 2 announcement provides an interesting case in point. In line with the general pattern noted above, the spot volatility shows an initial burst at 14:00, followed by a gradual decline to a lower, albeit still elevated, level at 14:30. Then, concurrent with the start of Chairman Powell’s speech, the 14:30-14:35 volatility estimate increased moderately, reflecting the limited new information presented in the opening, relatively structured, portion of the speech. This modest uptick is then followed by a dramatic volatility spike, of even greater magnitude than the initial surge that accompanied the 14:00 announcement. This volatility spike coincided with the time at which Powell concluded his opening remarks and began the press conference by mentioning that the ultimate level of interest rates would be “*higher than previously expected.*”²⁹ Powell’s brief interaction with the media shortly thereafter further underscored the looming uncertainty surrounding the central bank’s monetary tightening agenda.³⁰ These comments on the likely trajectory of monetary policy offered crucial forward guidance, and their unexpected nature in effect amounted to a “policy shock” resulting in a sharp increase in market volatility at the time.

In sum, asset price volatilities often experience large changes over short time windows in response to the release of important new economic information. The new optimal high-frequency candlestick estimators developed here allows for meaningful estimation of such changes, which would otherwise be obscured by the use of longer estimation windows, in turn affording a more nuanced depiction and better understanding of the economic mechanisms at work.

2.4 Concluding Remarks

We develop a new class of optimal range-based nonparametric volatility estimators. The new optimal estimators are explicitly geared to the volatility object of interest and relevant loss function. They involve complex, yet closed-form and easy-to-evaluate, nonlinear functions of the range, the absolute return, and a mea-

²⁹A complete transcript of Powell’s statement is available at: <https://www.federalreserve.gov/monetarypolicy/fomcpresconf20221102.htm>.

³⁰The first two questions from the media, posed by Colby Smith of the *Financial Times* and Howard Schneider of Reuters, respectively, also both concerned the potential slowdown of future rate increases.

sure of asymmetry. The efficiency gains provided by the new estimators compared to currently used suboptimal range-based estimators rooted in ad hoc functional-form assumptions can be substantial.

Looking ahead, the same infill asymptotic decision-theoretic framework developed here, based on coupling the nonparametric volatility estimation problem with a finite-sample optimal estimation problem, could possibly be adapted to study other outstanding optimal nonparametric inference problems. High-frequency range-based estimators have also previously been used for the estimation of integrated volatility over non-trivial time intervals (e.g., [Christensen and Podolskij \(2007\)](#)). The new optimal estimators developed here could similarly be employed for that purpose, as well as the estimation of other volatility functionals. The integrated quarticity, in particular, has proven notoriously difficult to accurately estimate in practice, yet it plays a crucial role in assessing the estimation error of traditional realized volatility type estimators (e.g., [Barndorff-Nielsen and Shephard \(2002\)](#) and [Bollerslev et al. \(2016\)](#)). Prior empirical uses of range-based volatility estimators for modeling and forecasting time-varying volatility bounds (early contributions include [Gallant et al. \(1999\)](#) and [Alizadeh et al. \(2002\)](#)). The range-based estimators developed here may naturally be used in that context as well for obtaining more accurate inference. They could also help sharpen the inference in the recent and growing literature that relies on high-frequency identification through heteroskedasticity (following [Rigobon \(2003\)](#)), and volatilities being higher over short “treatment” windows following news events (e.g., [Nakamura and Steinsson \(2018a\)](#) and [Bollerslev et al. \(2018\)](#)). We leave further work on all of these theoretical and more empirically oriented issues for future research.

Chapter 3

Optimal Spot Volatility Estimation Based on Multiple Candlesticks

3.1 Nonparametric Range-Based Volatility Estimation

3.1.1 Theoretical Setting and Background

We begin by revisiting some established results in [Bollerslev et al. \(2024\)](#). We assume that the log-price process P is an Itô semimartingale defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$ expressed as follows:

$$P_t = P_0 + \int_0^t b_s ds + \int_0^t \sigma_s dW_s + J_t, \quad 0 \leq t \leq T.$$

where the drift b and the volatility σ are optional processes, W is a standard Brownian motion, and J_t is a pure jump process driven by a Poisson random measure. The focus of this paper is on the p th power transformation of spot volatility at some fixed time t , namely σ_t^p . Empirically relevant choices for p include $p = 1$ (spot volatility), $p = 2$ (the spot variance), and $p = -1$ (spot precision), etc.

We construct an estimator of σ_t^p based on k consecutive blocks of the form $I_n \equiv [s, s + k\Delta_n]$ such that $t \in I_n$, and we consider in-fill asymptotics where $\Delta_n \rightarrow 0$. On each block $I_n^{(i)} \equiv [s + (i-1)\Delta_n, s + i\Delta_n]$, we observe the high-frequency candlestick data comprised of the open, high, low, and close prices, defined as $P_{s+(i-1)\Delta_n}$, $\sup_{t \in I_n^{(i)}} P_t$, $\inf_{t \in I_n^{(i)}} P_t$, and $P_{s+i\Delta_n}$, respectively. Denoting $P_t^{(i)} \equiv P_t - P_{s+(i-1)\Delta_n}$, this information concisely summarized by three normalized re-

turns:

$$r_i \equiv \frac{P_{s+i\Delta_n}^{(i)}}{\sqrt{\Delta_n}}, \quad h_i \equiv \frac{\sup_{t \in I_n^{(i)}} P_t^{(i)}}{\sqrt{\Delta_n}}, \quad l_i \equiv \frac{\inf_{t \in I_n^{(i)}} P_t^{(i)}}{\sqrt{\Delta_n}}.$$

Based on the k candlestick observations $\mathbf{C}_k = (r_i, h_i, l_i)_{1 \leq i \leq k}$, the estimator of σ_t^p takes the form:

$$\hat{\sigma}_t^p = f_p(\mathbf{C}_k),$$

where $f_p(\cdot) : \mathbb{R}^{3k} \rightarrow \mathbb{R}_+$ is a *regular and scale-equivariant* function, in the sense that it is Lebesgue almost everywhere continuous and satisfies $f_p(\lambda \mathbf{x}) = \lambda^p f_p(\mathbf{x})$ for any constant $\lambda > 0$ and any vector $\mathbf{x} \in \mathbb{R}^{3k}$. As discussed in [Bollerslev et al. \(2024\)](#), the scale-equivariance restriction is natural since σ_t essentially represents a scale parameter. The regularity condition of $f_p(\cdot)$ further ensures that $\hat{\sigma}_t^p$ can be “coupled” with a limiting distribution that is also scale-equivariant. We shall call $\hat{\sigma}_t^p$ a regular scale-equivariant estimator if $f_p(\cdot)$ meets these criteria.

To derive the limiting distribution for $\hat{\sigma}_t^p$, we require the following mild technical assumption:

Assumption 2. *We assume that there exists an increasing and diverging sequence of stopping times $(T_m)_{m \geq 1}$ and a sequence of constants $(K_m)_{m \geq 1}$ such that the following hold for each $m \geq 1$: (1) $|b_t| + |\sigma_t| + |\sigma_t|^{-1} + F_t(\mathbb{R} \setminus \{0\}) \leq K_m$ for all $t \in [0, T_m]$, where F_t is the spot Lévy measure of J ; (2) for some constant $\kappa > 0$, $\mathbb{E}[|\sigma_{t \wedge T_m} - \sigma_{s \wedge T_m}|^2] \leq K_m |t - s|^{2\kappa}$.*

Under Assumption 2, we have the following coupling result for any fixed $k \geq 1$:

Theorem 3.1 ([Bollerslev et al. \(2024\)](#), Corollary 1). *Suppose Assumption 2 holds true. For any $k \geq 1$ and any regular scale-equivariant estimator $\hat{\sigma}_t^p$, it holds as $\Delta_n \rightarrow 0$ that*

$$\frac{\hat{\sigma}_t^p}{\sigma_t^p} = f_p(\boldsymbol{\zeta}_k) + o_p(1), \quad (3.1)$$

where $\boldsymbol{\zeta}_k \equiv (\zeta_{i,r}, \zeta_{i,h}, \zeta_{i,l})_{1 \leq i \leq k}$, in which

$$\zeta_{i,r} \equiv \frac{W_{s+i\Delta_n}^{(i)}}{\sqrt{\Delta_n}}, \quad \zeta_{i,h} \equiv \frac{\sup_{t \in I_n^{(i)}} W_t^{(i)}}{\sqrt{\Delta_n}}, \quad \zeta_{i,l} \equiv \frac{\inf_{t \in I_n^{(i)}} W_t^{(i)}}{\sqrt{\Delta_n}}, \quad (3.2)$$

with $W_t^{(i)} \equiv W_t - W_{s+(i-1)\Delta_n}$.

By the Brownian scaling law, we have $\xi_k \stackrel{d}{=} \tilde{\xi}_k \equiv (\tilde{\xi}_{i,r}, \tilde{\xi}_{i,h}, \tilde{\xi}_{i,l})_{1 \leq i \leq k}$, where

$$\tilde{\xi}_{i,r} \equiv \tilde{W}_1^{(i)}, \quad \tilde{\xi}_{i,h} \equiv \sup_{t \in [0,1]} \tilde{W}_t^{(i)}, \quad \tilde{\xi}_{i,l} \equiv \inf_{t \in [0,1]} \tilde{W}_t^{(i)},$$

and $(\tilde{W}^{(i)})_{1 \leq i \leq k}$ are k mutually independent standard Brownian motions. Therefore, Theorem 3.1 shows that for any regular scale-equivariant estimator $\hat{\sigma}_t^P$, the ratio $\hat{\sigma}_t^P / \sigma_t^P$ is coupled with the pivotal quantity $f_p(\xi_k)$ which does not depend on any nuisance parameters. In a decision-theoretic framework, this pivotal quantity allows us to evaluate the asymptotic estimation risk of $\hat{\sigma}_t^P$. Specifically, let $L(\cdot)$ be a strictly convex loss function, the asymptotic estimation risk is defined as the expected asymptotic loss, i.e.,

$$R(\hat{\sigma}_t^P, L) \equiv \mathbb{E} \left[\lim_{\Delta_n \rightarrow 0} L \left(\frac{\hat{\sigma}_t^P}{\sigma_t^P} \right) \right] = \mathbb{E} [L(f_p(\xi_k))], \quad (3.3)$$

where the equality follows from Theorem 3.1. As the distribution of $\tilde{\xi}_k$, and hence ξ_k , is well-established in the literature (see, e.g., Feller (1951), Bollerslev et al. (2024)), the value of $R(\hat{\sigma}_t^P, L)$ only depends on the choice of loss function $L(\cdot)$ and functional form of estimator $f_p(\cdot)$. Following the literature, we shall focus on Stein's loss ($L_s(\cdot)$) and the scaled quadratic loss ($L_q(\cdot)$) which are commonly employed for scale parameter estimation:

$$L_s(x) \equiv x - \log x - 1, \quad L_q(x) \equiv (x - 1)^2.$$

In the $k = 1$ case, Bollerslev et al. (2024) derive an analytical Asymptotic Minimum Risk Equivariant (AMRE) estimator of σ_t^P that minimizes $R(\hat{\sigma}_t^P, L_s)$ and $R(\hat{\sigma}_t^P, L_q)$. They also provide implicit forms for the AMRE estimator when $k > 1$ in (2.13)-(2.14), which, however, are nearly infeasible to implement in practice.¹

3.1.2 AMRE Estimator with Multiple Candlesticks

In this section, we derive an explicit form for the AMRE estimator of σ_t^P for any fixed $k \geq 1$ which is feasible to implement in practice. To this end, by a sufficient statistic argument, we transform the original candlestick data by $(r_i, h_i, l_i) \mapsto (|r_i|, w_i, a_i)$, where $w_i \equiv h_i - l_i$ is the range of the i th candlestick, and $a_i \equiv |h_i +$

¹For the $k = 2$ case, Bollerslev et al. (2024) propose to approximate the AMRE estimator by machine learning with extensive simulation. Nevertheless, we show in Section 3.1.3 that their simulation scheme introduces a discretization bias, failing to yield the exact estimator. Moreover, this method does not easily generalize to the $k > 2$ case due to the curse of dimensionality, and it does not exploit the symmetric form across different candlesticks in the functional form.

$|l_i - r_i|$ measures the degree of asymmetry of the candlestick. We rewrite $\mathbf{C}_k \equiv (|r_i|, w_i, a_i)_{1 \leq i \leq k}$ and define the function $M_{k,p}(\mathbf{C}_k)$ as

$$M_{k,p}(\mathbf{C}_k) \equiv \int_0^\infty v^{3k+p-1} \prod_{i=1}^k \tilde{g}(v|r_i|, vw_i, va_i) dv,$$

where

$$\tilde{g}(r, w, a) \equiv \sum_{m=-\infty}^{\infty} m^2 \phi''(2mw + r) - m(m+1) \phi''((2m+1)w - a), \quad (3.4)$$

and $\phi''(x) \equiv (x^2 - 1)\phi(x)$ is the second-order derivative of the standard Gaussian density $\phi(x) \equiv (2\pi)^{-1/2} e^{-x^2/2}$. We are now prepared to present our feasible AMRE estimator in the theorem below:

Theorem 3.2. *Under Assumption 2, for any fixed $k \geq (1 - 2p)/3 \vee 1$, the following estimators are AMRE under $L_s(\cdot)$ and $L_q(\cdot)$, respectively:*

$$\hat{\sigma}_s^{p,*}(k) \equiv \frac{M_{k,0}(\mathbf{C}_k)}{M_{k,p}(\mathbf{C}_k)}, \quad \hat{\sigma}_q^{p,*}(k) \equiv \frac{M_{k,p}(\mathbf{C}_k)}{M_{k,2p}(\mathbf{C}_k)}.$$

COMMENT. In essence, $\hat{\sigma}_s^{p,*}(k)$ and $\hat{\sigma}_q^{p,*}(k)$ are Pitman estimators of σ_t^p based on the observations \mathbf{C}_k (see, e.g., equation (3.19) in [Lehmann and Casella \(1998\)](#)), and the requirement $k \geq (1 - 2p)/3$ is necessary for $M_{k,2p}(\mathbf{C}_k)$ to converge when $p < 0$. To compute $\hat{\sigma}_s^{p,*}(k)$ and $\hat{\sigma}_q^{p,*}(k)$, one needs to evaluate $M_{k,p}(\mathbf{C}_k)$ twice with different choices of p , which amounts to the calculation of two improper integrals involving a finite product of the $\tilde{g}(\cdot)$ function. Initially, this integral may seem impossible to compute, as it entails integrating a product of k infinite sums over the entire positive real line. However, with a closer inspection of the $\tilde{g}(\cdot)$ function, its summands converge to zero rapidly as $|m| \rightarrow \infty$ and $v \rightarrow \infty$ at a rate of $e^{-(mv)^2/2}$. Therefore, one can safely truncate the order of the infinite sums and the integral limit to some large value with virtually no loss of working numerical precision. Consequently, the univariate integral can be computed numerically to *machine precision* in principle. MATLAB codes to compute the estimators are provided in the supplement of this paper.

Since the expressions for $\hat{\sigma}_s^{p,*}(k)$ and $\hat{\sigma}_q^{p,*}(k)$ can, in principle, be numerically computed for each k , they provide a numerical method to compute the analytical estimators in [Bollerslev et al. \(2024\)](#) with $k = 1$. This enables us to directly verify the performance of our numerical algorithm. Below, we denote the estimators com-

puted using analytical form of [Bollerslev et al. \(2024\)](#) as $\hat{\sigma}_s^{p,A}(1) \equiv \hat{\sigma}_s^{p,*}(1)$ and $\hat{\sigma}_q^{p,A}(1) \equiv \hat{\sigma}_q^{p,*}(1)$, respectively. For completeness, we present the general expressions for $\hat{\sigma}_s^{p,A}(1)$ and $\hat{\sigma}_q^{p,A}(1)$ here, which are numerically more stable than those in [Bollerslev et al. \(2024\)](#). Let $s \equiv |r|/w$ and $d \equiv 1 - a/w$ denote the two maximal invariants computed from the candlestick data $(|r|, w, a)$. Then, $\hat{\sigma}_s^{p,A}(1)$ and $\hat{\sigma}_q^{p,A}(1)$ take the form

$$\hat{\sigma}_s^{p,A}(1) = K_s(p)w^p \frac{G_0(d) - H_0(s)}{G_p(d) - H_p(s)}, \quad \hat{\sigma}_q^{p,A}(1) = K_q(p)w^p \frac{G_p(d) - H_p(s)}{G_{2p}(d) - H_{2p}(s)},$$

where $K_s(p)$ and $K_q(p)$ are two constants depending on p , and

$$\begin{aligned} G_m(x) &= \Psi_m^+(x) + \frac{(1-x)\Psi_{m+1}^-(x)}{m+1} - \frac{x(2-x)\Psi_{m+2}^+(x)}{4(m+1)(m+2)}, \\ H_m(x) &= \Psi_m^+(x) - \frac{x\Psi_{m+1}^-(x)}{m+1} + \frac{x^2\Psi_{m+2}^+(x)}{4(m+1)(m+2)}, \end{aligned} \quad (3.5)$$

here $\Psi_m^\pm(x) \equiv \psi_m(1-x/2) \pm \psi_m(1+x/2)$, where $\psi_m(x)$ is the m th order Polygamma function. The main difference between the above expressions and those in [Bollerslev et al. \(2024\)](#) lies in the functional form of $G_m(\cdot)$ and $H_m(\cdot)$. Specifically, let $\tilde{G}_m(\cdot)$ and $\tilde{H}_m(\cdot)$ denote the original functions in equation (2.10)-(2.11) of [Bollerslev et al. \(2024\)](#), then one can show that, by properties of the Polygamma function,

$$G_m(x) \equiv \tilde{G}_m(1-x), \quad H_m(x) \equiv \tilde{H}_m(x).$$

However, despite the above analytical equivalence, note that $\tilde{G}_m(1-x)$ contains the term $\psi_m(x/2)$, which diverges as $x \rightarrow 0$, leading to numerical instability and potential generation of negative volatility estimates. This issue is resolved by using the definitions in equation (3.5), as $\Psi_m^\pm(x)$ remains finite for all $x \in [0, 1]$.

3.1.3 Unbiased Evaluation of Estimation Risk

For any regular scale-equivariant estimator $\hat{\sigma}_t^p$, its asymptotic risk under the loss function $L(\cdot)$ is given by $R(\hat{\sigma}_t^p, L)$. It is of crucial importance to compute $R(\hat{\sigma}_t^p, L)$, as it allows us to compare the relative performance amongst different estimators. From equation 3.3, we need to evaluate the expectation $\mathbb{E}[L(f_p(\xi_k))]$, which is unlikely to be in closed form due to the complex density of the Brownian functional ξ_k . Therefore, existing literature typically evaluates $\mathbb{E}[L(f_p(\xi_k))]$ by simulation. Note that $\tilde{\zeta}_{i,h}$ and $\tilde{\zeta}_{i,l}$ are extreme functionals of the continuous path of a Brownian motion, these random variables are typically drawn based on an Euler

discretization scheme. Specifically, consider the discretely observed i th Brownian motion $(W_t^{(i)})_{t \in \{0, 1/M, 2/M, \dots, 1\}}$ where $1/M$ is the size of the discretization grid, an approximated random draw of $\tilde{\zeta}_{i,h}$ and $\tilde{\zeta}_{i,l}$ can be computed as

$$\tilde{\zeta}_{i,h,M} \equiv \max_{t \in \{0, 1/M, 2/M, \dots, 1\}} W_t^{(i)}, \quad \tilde{\zeta}_{i,l,M} \equiv \min_{t \in \{0, 1/M, 2/M, \dots, 1\}} W_t^{(i)}.$$

Let $\tilde{\xi}_{k,M}^{(j)} \equiv (\tilde{\zeta}_{i,r}^{(j)}, \tilde{\zeta}_{i,h,M}^{(j)}, \tilde{\zeta}_{i,l,M}^{(j)})_{1 \leq i \leq k}$ denote the j th random draw of k candlesticks data, $\mathbb{E}[L(f_p(\tilde{\xi}_k))]$ is typically approximated by the Monte Carlo average of N random draws

$$\overline{R}_M^N(\hat{\sigma}_t^p, L) \equiv \frac{1}{N} \sum_{j=1}^N L(f_p(\tilde{\xi}_{k,M}^{(j)})).$$

For any fixed M , we have the standard estimate $\overline{R}_M^N(\hat{\sigma}_t^p, L) - \mathbb{E}[L(f_p(\tilde{\xi}_{k,M}))] = O_p(N^{-1/2})$ assuming a bounded variance of $L(f_p(\tilde{\xi}_{k,M}^{(j)}))$, so increasing N reduces the variance of the Monte Carlo average. Nevertheless, $\mathbb{E}[L(f_p(\tilde{\xi}_{k,M}))]$ is not exactly the estimation risk of $\hat{\sigma}_t^p$, as it still contains the discretization error. In detail, $\tilde{\zeta}_{i,h,M}$ (resp. $\tilde{\zeta}_{i,l,M}$) are strictly smaller (resp. larger) than the true value $\tilde{\zeta}_{i,h}$ (resp. $\tilde{\zeta}_{i,l}$) due to the fact that the supremum (resp. infimum) are taken over a discrete instead of a continuous grid. Proposition 3 in [Asmussen et al. \(1995\)](#) shows that, for any function $g(\cdot)$ with a bounded second-order derivative, it holds independently for every i that

$$\begin{aligned} \mathbb{E}[g(\tilde{\zeta}_{i,h,M})] &= \mathbb{E}[g(\tilde{\zeta}_{i,h})] - KM^{-1/2}\mathbb{E}[g'(\tilde{\zeta}_{i,h})] + o(M^{-1/2}), \\ \mathbb{E}[g(\tilde{\zeta}_{i,l,M})] &= \mathbb{E}[g(\tilde{\zeta}_{i,l})] + KM^{-1/2}\mathbb{E}[g'(\tilde{\zeta}_{i,l})] + o(M^{-1/2}), \end{aligned} \quad (3.6)$$

where $K \approx 0.5826$. Therefore, the discretization error persists even if $N \rightarrow \infty$, since it only diminishes at the rate $O(M^{-1/2})$ independent of N . This implies that both N and M need to diverge to improve the precision of $\overline{R}_M^N(\hat{\sigma}_t^p, L)$. For example, [Li et al. \(2022\)](#) uses $N = 10^8$ together with $M = 10^7$ to evaluate the asymptotic variance of the OK estimator, which amounts to 10^{15} total draw of random numbers hence is computationally very costly.

In light of equation (3.6), a bias correction method similar to [Wang et al. \(2011\)](#) is possible in this case to obtain a true asymptotic risk for a *correct* functional form of $\hat{\sigma}_t^p = f_p(\mathbf{C}_k)$. However, not only does the discretization bias prevent us from investigating the true asymptotic risk of an estimator, it can also interfere with

the functional bias in approximating $f_p(\cdot)$ through the machine learning method suggested in [Bollerslev et al. \(2024\)](#). To overcome this problem and improve the computational efficiency of evaluating $R(\hat{\sigma}_t^p, L)$, we develop an exact simulation scheme to numerically draw ξ_k from its analytical density. The sampling scheme is detailed in the following theorem.

Theorem 3.3. *Let r denote a random draw from $\mathcal{N}(0, 1)$. Draw u and v independently from $\mathcal{U}([0, 1])$ and set $h = (r + \sqrt{r - 2\log(1-u)})/2$. Given r and h , set l as the solution to $v = F(l | r, h)$ where*

$$F(l | r, h) = 1 - \sum_{m=-\infty}^{\infty} m \frac{\phi'(c - 2m(h-l))}{\phi'(2h-c)} - (m+1) \frac{\phi'(c - 2m(h-l) - 2l)}{\phi'(2h-c)},$$

in which $l \leq (r \wedge 0)$ and $\phi'(x) = -x\phi(x)$. It holds that $(r, h, l) \stackrel{d}{=} (\tilde{\xi}_{i,r}, \tilde{\xi}_{i,h}, \tilde{\xi}_{i,l})$.

COMMENT. To the best of our knowledge, we are among the first to propose an exact simulation scheme for the joint law of the terminal value, the supremum, and the infimum of a Brownian motion. In essence, we draw sequentially from the corresponding conditional distributions given the previously drawn random variables, guaranteeing exact simulation from the target joint distribution. The only complication here arises from numerically solving the nonlinear equation $v = F(l | r, h)$, which involves an infinite sum. Similar to $\tilde{g}(\cdot)$ in equation (3.4), the summands in $F(l | r, h)$ decay rapidly to zero, allowing us to truncate the sum at some large number. Furthermore, since $F(l | r, h)$ is a conditional cumulative distribution function, it monotonically increases in $l \in (-\infty, r \wedge 0]$ from 0 to 1. This ensures that the solution to $v = F(l | r, h)$ must be unique and thus can be efficiently solved to machine precision. MATLAB codes are provided in the supplement to perform the exact simulation.

As a direct consequence of Theorem 3.3, we can easily simulate N copies of $\tilde{\xi}_k$ and evaluate $\mathbb{E}[L(f_p(\tilde{\xi}_k))]$ using

$$\bar{R}^N(\hat{\sigma}_t^p, L) \equiv \frac{1}{N} \sum_{j=1}^N L(f_p(\tilde{\xi}_k^{(j)})), \quad (3.7)$$

which is an unbiased risk estimate that converges to $\mathbb{E}[L(f_p(\tilde{\xi}_k))]$ at the rate of $N^{-1/2}$. As $\bar{R}^N(\hat{\sigma}_t^p, L)$ is free from the discretization bias, it dominates $\bar{R}_M^N(\hat{\sigma}_t^p, L)$ for any M and should always be preferred. Henceforth, we shall use $\bar{R}^N(\hat{\sigma}_t^p, L)$ for

the risk comparison exercise in the next section.

3.1.4 Risk Comparisons

We proceed to investigate the asymptotical risks of the AMRE estimators across different choices k and compare their performances to some existing candlestick-based methods. To maintain brevity, we shall focus on the performances of spot volatility ($p = 1$) and spot variance ($p = 2$) estimators, which are the most commonly used choices.

We first characterize the properties of $\hat{\sigma}_s^{p,*}(k)$ and $\hat{\sigma}_q^{p,*}(k)$, namely, their asymptotic bias, variance, and the risks under both $L_s(\cdot)$ and $L_q(\cdot)$, where the asymptotic bias and variance of a generic estimator $\hat{\sigma}_t^p = f_p(\mathbf{C}_k)$ are defined as

$$\text{ABias} \equiv \mathbb{E}[f_p(\tilde{\xi}_k)] - 1, \quad \text{AVar} = \text{Var}(f_p(\tilde{\xi}_k)), \quad (3.8)$$

respectively. These quantities are simulated based on 10^6 random draws of $\tilde{\xi}_k$ using the exact sampling scheme of Theorem 3.3. To verify the reliability of our numerical scheme, we also report the same quantities computed using analytical expressions $\hat{\sigma}_s^{p,A}(1)$ and $\hat{\sigma}_q^{p,A}(1)$. These results are summarized in Table 3.1.

Several interesting findings can be drawn from Table 3.1. First, the properties of the analytical estimators are numerically indistinguishable from its numerical implementations up to at least four decimal places. This alignment validates the theoretical statements in Theorem 3.2 and underlies the reliability of our numerical algorithm. Second, the asymptotic bias and variance reported in Table 3.1 exhibit slight differences from those in Bollerslev et al. (2024), particularly noticeable when $k = 2$. For example, in Table 3 of Bollerslev et al. (2024), the asymptotic bias and variance of $\hat{\sigma}_q^{2,*}(2)$ are listed as -0.103 and 0.103 , respectively, contrasting with our results of -0.1113 and 0.0995 . These differences can be attributed to both the bias introduced by Euler discretization scheme, as well as the approximation error inherent in the machine learning approximation step for $k = 2$.

Analyzing the properties of AMRE estimators across different k , we observe that: (i) the biases of $\hat{\sigma}_s^{1,*}(k)$ and $\hat{\sigma}_s^{2,*}(k)$ are almost zero, which is in line with the fact that the AMRE estimator under Stein's loss is unbiased (see, e.g., Brown (1968)). Conversely, the biases of $\hat{\sigma}_q^{1,*}(k)$ and $\hat{\sigma}_q^{2,*}(k)$ are negative, as anticipated,

Table 3.1: Asymptotic Properties of AMRE Estimators Based on Multiple Candlesticks

$\hat{\sigma}_t^p$	ABias	AVar	$R(\hat{\sigma}_t^p, L_s)$	$R(\hat{\sigma}_t^p, L_q)$	$\hat{\sigma}_t^p$	ABias	AVar	$R(\hat{\sigma}_t^p, L_s)$	$R(\hat{\sigma}_t^p, L_q)$
$\hat{\sigma}_s^{1,A}(1)$	-0.0002	0.0622	0.0309	0.0622	$\hat{\sigma}_q^{1,A}(1)$	-0.0586	0.0551	0.0327	0.0585
$\hat{\sigma}_s^{1,*}(1)$	-0.0002	0.0622	0.0309	0.0622	$\hat{\sigma}_q^{1,*}(1)$	-0.0586	0.0551	0.0327	0.0585
$\hat{\sigma}_s^{1,*}(2)$	0.0001	0.0307	0.0153	0.0307	$\hat{\sigma}_q^{1,*}(2)$	-0.0296	0.0289	0.0157	0.0298
$\hat{\sigma}_s^{1,*}(3)$	0.0001	0.0203	0.0101	0.0203	$\hat{\sigma}_q^{1,*}(3)$	-0.0197	0.0195	0.0103	0.0198
$\hat{\sigma}_s^{1,*}(4)$	0.0001	0.0151	0.0075	0.0151	$\hat{\sigma}_q^{1,*}(4)$	-0.0148	0.0147	0.0076	0.0149
$\hat{\sigma}_s^{1,*}(5)$	0.0001	0.0120	0.0060	0.0120	$\hat{\sigma}_q^{1,*}(5)$	-0.0118	0.0118	0.0061	0.0119
$\hat{\sigma}_s^{1,*}(6)$	0.0000	0.0100	0.0050	0.0100	$\hat{\sigma}_q^{1,*}(6)$	-0.0099	0.0098	0.0050	0.0099
$\hat{\sigma}_s^{1,*}(7)$	0.0000	0.0086	0.0043	0.0086	$\hat{\sigma}_q^{1,*}(7)$	-0.0085	0.0084	0.0043	0.0085
$\hat{\sigma}_s^{1,*}(8)$	0.0000	0.0075	0.0037	0.0075	$\hat{\sigma}_q^{1,*}(8)$	-0.0074	0.0074	0.0038	0.0074
$\hat{\sigma}_s^{1,*}(9)$	-0.0001	0.0066	0.0033	0.0066	$\hat{\sigma}_q^{1,*}(9)$	-0.0066	0.0065	0.0033	0.0066
$\hat{\sigma}_s^{1,*}(10)$	-0.0001	0.0060	0.0030	0.0060	$\hat{\sigma}_q^{1,*}(10)$	-0.0060	0.0059	0.0030	0.0059
$\hat{\sigma}_s^{1,*}(11)$	0.0000	0.0054	0.0027	0.0054	$\hat{\sigma}_q^{1,*}(11)$	-0.0054	0.0054	0.0027	0.0054
$\hat{\sigma}_s^{1,*}(12)$	0.0000	0.0050	0.0025	0.0050	$\hat{\sigma}_q^{1,*}(12)$	-0.0050	0.0049	0.0025	0.0049
$\hat{\sigma}_s^{1,*}(13)$	0.0000	0.0046	0.0023	0.0046	$\hat{\sigma}_q^{1,*}(13)$	-0.0046	0.0045	0.0023	0.0046
$\hat{\sigma}_s^{1,*}(14)$	0.0000	0.0042	0.0021	0.0042	$\hat{\sigma}_q^{1,*}(14)$	-0.0042	0.0042	0.0021	0.0042
$\hat{\sigma}_s^{1,*}(15)$	0.0000	0.0040	0.0020	0.0040	$\hat{\sigma}_q^{1,*}(15)$	-0.0039	0.0039	0.0020	0.0039
$\hat{\sigma}_s^{1,*}(16)$	0.0000	0.0037	0.0019	0.0037	$\hat{\sigma}_q^{1,*}(16)$	-0.0037	0.0037	0.0019	0.0037
$\hat{\sigma}_s^{1,*}(17)$	0.0000	0.0035	0.0017	0.0035	$\hat{\sigma}_q^{1,*}(17)$	-0.0035	0.0035	0.0018	0.0035
$\hat{\sigma}_s^{1,*}(18)$	0.0000	0.0033	0.0016	0.0033	$\hat{\sigma}_q^{1,*}(18)$	-0.0032	0.0033	0.0017	0.0033
$\hat{\sigma}_s^{1,*}(19)$	0.0001	0.0031	0.0016	0.0031	$\hat{\sigma}_q^{1,*}(19)$	-0.0031	0.0031	0.0016	0.0031
$\hat{\sigma}_s^{1,*}(20)$	0.0001	0.0030	0.0015	0.0030	$\hat{\sigma}_q^{1,*}(20)$	-0.0029	0.0030	0.0015	0.0030
$\hat{\sigma}_t^p$	ABias	AVar	$R(\hat{\sigma}_t^p, L_s)$	$R(\hat{\sigma}_t^p, L_q)$	$\hat{\sigma}_t^p$	ABias	AVar	$R(\hat{\sigma}_t^p, L_s)$	$R(\hat{\sigma}_t^p, L_q)$
$\hat{\sigma}_s^{2,A}(1)$	-0.0003	0.2596	0.1221	0.2596	$\hat{\sigma}_q^{2,A}(1)$	-0.2055	0.1634	0.1471	0.2056
$\hat{\sigma}_s^{2,*}(1)$	-0.0003	0.2596	0.1221	0.2596	$\hat{\sigma}_q^{2,*}(1)$	-0.2055	0.1634	0.1471	0.2056
$\hat{\sigma}_s^{2,*}(2)$	0.0002	0.1263	0.0608	0.1263	$\hat{\sigma}_q^{2,*}(2)$	-0.1113	0.0995	0.0676	0.1119
$\hat{\sigma}_s^{2,*}(3)$	0.0003	0.0827	0.0402	0.0827	$\hat{\sigma}_q^{2,*}(3)$	-0.0758	0.0705	0.0433	0.0762
$\hat{\sigma}_s^{2,*}(4)$	0.0001	0.0613	0.0300	0.0613	$\hat{\sigma}_q^{2,*}(4)$	-0.0576	0.0544	0.0318	0.0577
$\hat{\sigma}_s^{2,*}(5)$	0.0001	0.0488	0.0240	0.0488	$\hat{\sigma}_q^{2,*}(5)$	-0.0463	0.0443	0.0251	0.0465
$\hat{\sigma}_s^{2,*}(6)$	0.0001	0.0404	0.0199	0.0404	$\hat{\sigma}_q^{2,*}(6)$	-0.0388	0.0373	0.0207	0.0388
$\hat{\sigma}_s^{2,*}(7)$	0.0000	0.0345	0.0171	0.0345	$\hat{\sigma}_q^{2,*}(7)$	-0.0333	0.0323	0.0176	0.0334
$\hat{\sigma}_s^{2,*}(8)$	0.0000	0.0301	0.0149	0.0301	$\hat{\sigma}_q^{2,*}(8)$	-0.0293	0.0284	0.0154	0.0293
$\hat{\sigma}_s^{2,*}(9)$	-0.0001	0.0267	0.0132	0.0267	$\hat{\sigma}_q^{2,*}(9)$	-0.0261	0.0253	0.0136	0.0260
$\hat{\sigma}_s^{2,*}(10)$	-0.0001	0.0240	0.0119	0.0240	$\hat{\sigma}_q^{2,*}(10)$	-0.0236	0.0229	0.0122	0.0234
$\hat{\sigma}_s^{2,*}(11)$	-0.0001	0.0218	0.0108	0.0218	$\hat{\sigma}_q^{2,*}(11)$	-0.0214	0.0209	0.0110	0.0213
$\hat{\sigma}_s^{2,*}(12)$	-0.0001	0.0199	0.0099	0.0199	$\hat{\sigma}_q^{2,*}(12)$	-0.0196	0.0192	0.0101	0.0195
$\hat{\sigma}_s^{2,*}(13)$	-0.0001	0.0184	0.0091	0.0184	$\hat{\sigma}_q^{2,*}(13)$	-0.0181	0.0177	0.0093	0.0180
$\hat{\sigma}_s^{2,*}(14)$	0.0000	0.0171	0.0085	0.0171	$\hat{\sigma}_q^{2,*}(14)$	-0.0168	0.0165	0.0086	0.0168
$\hat{\sigma}_s^{2,*}(15)$	0.0000	0.0159	0.0079	0.0159	$\hat{\sigma}_q^{2,*}(15)$	-0.0157	0.0154	0.0080	0.0157
$\hat{\sigma}_s^{2,*}(16)$	0.0000	0.0149	0.0074	0.0149	$\hat{\sigma}_q^{2,*}(16)$	-0.0147	0.0145	0.0075	0.0147
$\hat{\sigma}_s^{2,*}(17)$	0.0000	0.0140	0.0070	0.0140	$\hat{\sigma}_q^{2,*}(17)$	-0.0138	0.0137	0.0071	0.0138
$\hat{\sigma}_s^{2,*}(18)$	0.0001	0.0133	0.0066	0.0133	$\hat{\sigma}_q^{2,*}(18)$	-0.0130	0.0129	0.0067	0.0131
$\hat{\sigma}_s^{2,*}(19)$	0.0001	0.0125	0.0062	0.0125	$\hat{\sigma}_q^{2,*}(19)$	-0.0123	0.0122	0.0063	0.0124
$\hat{\sigma}_s^{2,*}(20)$	0.0001	0.0119	0.0059	0.0119	$\hat{\sigma}_q^{2,*}(20)$	-0.0117	0.0116	0.0060	0.0118

Note: For each estimator, the table presents the asymptotic bias (ABias) and the asymptotic variance (AVar) defined in equation (3.8), and the asymptotic risks under both Stein's loss $R(\hat{\sigma}_t^p, L_s)$ and the quadratic loss $R(\hat{\sigma}_t^p, L_q)$. Each number is computed from 10^6 simulated candlestick datasets using the exact simulation scheme in Theorem 3.3.

given the quadratic loss's inclination to penalize overestimation more severely than underestimation. (ii) The biases of $\hat{\sigma}_q^{1,*}(k)$ and $\hat{\sigma}_q^{2,*}(k)$, along with the variances and estimation risks of all estimators, diminish as k increases, approximately at a rate of k^{-1} . This is not surprising, since both estimators have the same limiting distribution as MLE estimator when k goes to infinity, as corroborated in the discussion below Theorem 8.3 in [Lehmann and Casella \(1998\)](#). (iii) For each k , the risk of $\hat{\sigma}_s^{P,*}(k)$ is superior to that of $\hat{\sigma}_q^{P,*}(k)$ under $L_s(\cdot)$, while the inverse holds true when assessing risk under $L_q(\cdot)$. This is consistent with the design that the AMRE estimator under a given loss function stands as the unique choice achieves the smallest asymptotic risk, see Corollary 3.3.4 of [Lehmann and Casella \(1998\)](#).

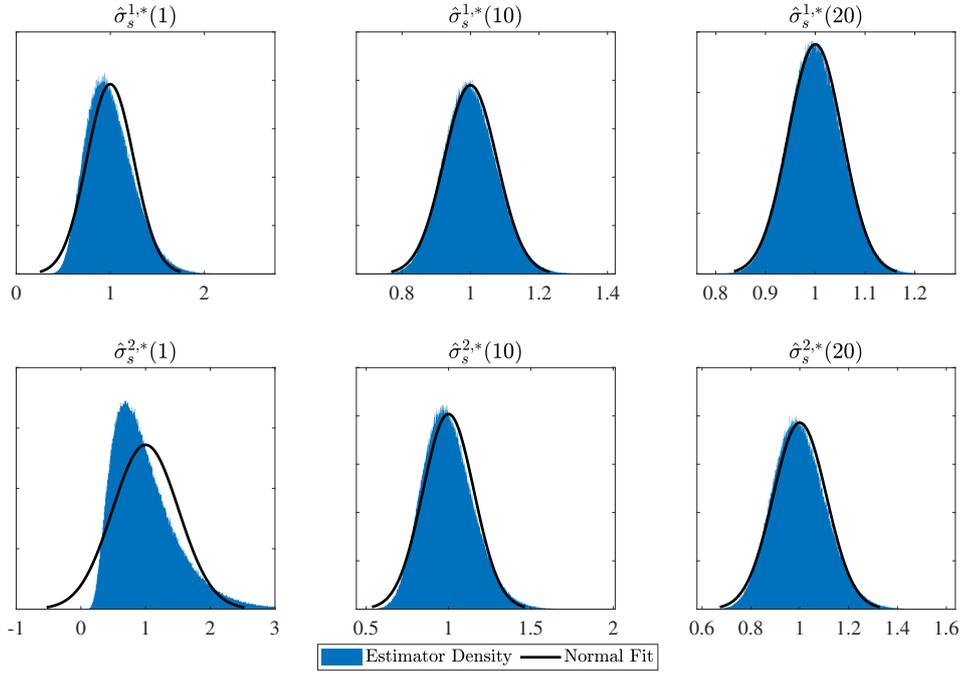
We proceed to examine the finite sample distributions of the AMRE estimators. For $k \in \{1, 10, 20\}$, we plot the histogram of the AMRE estimators along side with a normal fit in [Figure 3.1](#). The figure shows that the finite sample distributions of the AMRE estimators are generally skewed to the right for small k , where those under $L_q(\cdot)$ tend to exhibit thinner right tails compared to those under $L_s(\cdot)$, due to the heavier penalty of overestimation under the quadratic loss. As anticipated, the finite-sample distributions of all estimators become approximately normal when k increases. This provides an empirically convenient way to construct confidence intervals for the AMRE estimators using a large k (say, e.g., $k \geq 20$).

We conclude this section by comparing the AMRE estimators with some commonly used candlestick-based estimators in the literature, including the Best Linear Unbiased Estimator (BLUE) proposed in [Li et al. \(2022\)](#), and the Garman–Klass (GK) best quadratic unbiased estimator proposed in [Garman and Klass \(1980\)](#). As the GK estimator is not designed to use multiple candlesticks, we shall combine these estimator in a simple average fashion when k candlesticks are available. Specifically:

$$\begin{aligned}\hat{\sigma}_{BLUE}^1(k) &\equiv \frac{1}{k} \sum_{i=1}^k \hat{\sigma}_{BLUE,i}, & \hat{\sigma}_{BLUE,i} &\equiv 0.811w_i - 0.369|r_i|, \\ \hat{\sigma}_{GK}^2(k) &\equiv \frac{1}{k} \sum_{i=1}^k \hat{\sigma}_{GK,i}^2, & \hat{\sigma}_{GK,i}^2 &\equiv 0.5015w_i^2 + 0.0095a_i^2 - 0.3925r_i^2.\end{aligned}$$

Using the same construction, we can also consider a simple linear combination of

Panel I: AMRE estimators under the Stein's loss



Panel II: AMRE estimators under the quadratic loss

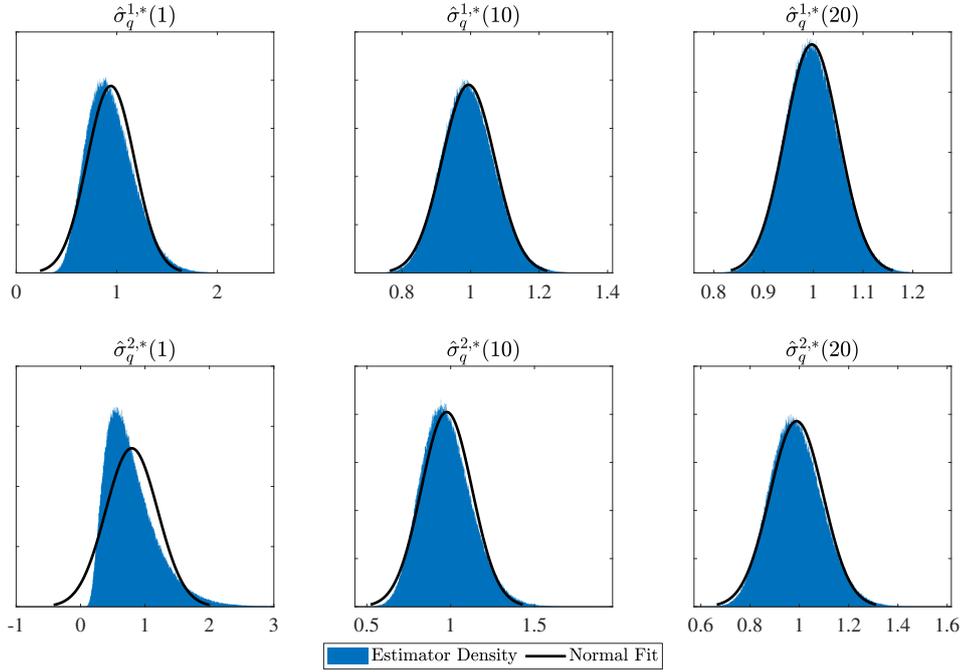


Figure 3.1: Finite sample distributions with fitted normal density curves for the AMRE estimators based on multiple candlesticks. The figure shows histograms of the estimators along with normal density plots generated using the sample mean and variance of the estimators. These distributions are based on 10^6 simulated candlestick datasets using the exact simulation scheme in Theorem 3.3.

the optimal estimator proposed in [Bollerslev et al. \(2024\)](#):

$$\hat{\sigma}_s^{p,A}(k) \equiv \frac{1}{k} \sum_{i=1}^k \hat{\sigma}_{s,i}^{p,A}, \quad \hat{\sigma}_q^{p,A}(k) \equiv \frac{1}{k} \sum_{i=1}^k \hat{\sigma}_{q,i}^{p,A},$$

where $\hat{\sigma}_{s,i}^{p,A}$ and $\hat{\sigma}_{q,i}^{p,A}$ are the AMRE estimators under $L_s(\cdot)$ and $L_q(\cdot)$ based on single candlestick (r_i, w_i, a_i) , respectively. To estimate power transforms of spot volatility/variance using a spot variance/volatility estimator, it is natural to apply the inverse transform to the estimators. For instance, we can use $(\hat{\sigma}_{BLUE}^1(k))^2$ and $(\hat{\sigma}_s^{1,A}(k))^2$ to estimate σ_t^2 , or $\sqrt{\hat{\sigma}_{GK}^2(k)}$ and $\sqrt{\hat{\sigma}_s^{2,A}(k)}$ to estimate σ_t .

To sum up, for each k and p , we consider ten different estimators, where the first five are intended to estimate the target power of volatility without additional transforms:

$$\begin{aligned} \hat{\sigma}_t^1 \in & \left\{ \hat{\sigma}_s^{1,*}(k), \hat{\sigma}_q^{1,*}(k), \hat{\sigma}_s^{1,A}(k), \hat{\sigma}_q^{1,A}(k), \hat{\sigma}_{BLUE}^1(k), \right. \\ & \left. \sqrt{\hat{\sigma}_s^{2,*}(k)}, \sqrt{\hat{\sigma}_q^{2,*}(k)}, \sqrt{\hat{\sigma}_s^{2,A}(k)}, \sqrt{\hat{\sigma}_q^{2,A}(k)}, \sqrt{\hat{\sigma}_{GK}^2(k)} \right\}, \\ \hat{\sigma}_t^2 \in & \left\{ \hat{\sigma}_s^{2,*}(k), \hat{\sigma}_q^{2,*}(k), \hat{\sigma}_s^{2,A}(k), \hat{\sigma}_q^{2,A}(k), \hat{\sigma}_{GK}^2(k), \right. \\ & \left. (\hat{\sigma}_s^{1,*}(k))^2, (\hat{\sigma}_q^{2,*}(k))^2, (\hat{\sigma}_s^{1,A}(k))^2, (\hat{\sigma}_q^{1,A}(k))^2, (\hat{\sigma}_{BLUE}^1(k))^2 \right\}. \end{aligned}$$

For each choice of k , we compute the relative efficiency of the above estimators under both $L_s(\cdot)$ and $L_q(\cdot)$ w.r.t. the AMRE estimators. For example, for the estimation of σ_t^p under $L_s(\cdot)$, the relative efficiency of $\hat{\sigma}_t^p$ is

$$R(\hat{\sigma}_s^{p,*}(k), L_s) / R(\hat{\sigma}_t^p, L_s), \quad (3.9)$$

which is less or equal to unity by construction. We present these results in [Table 3.2](#) and [3.3](#).

The results in [Table 3.2](#) and [3.3](#) carry several significant implications regarding the selection of estimators when multiple candlesticks are available. First, for estimators under the Stein's loss, one naturally expects that the unbiased estimators, such as $\hat{\sigma}_s^{p,A}(k)$, $\hat{\sigma}_{BLUE}^1(k)$, and $\hat{\sigma}_{GK}^2(k)$, should exhibit higher relative efficiency than those biased estimators. However, aside from $k = 1$, it appears that the power-transformed AMRE estimator $\sqrt{\hat{\sigma}_s^{2,*}(k)}$ is nearly as efficient as the AMRE estimator for $p = 1$. Similarly, $(\hat{\sigma}_s^{1,*}(k))^2$ is nearly as efficient as $\hat{\sigma}_s^{2,*}(k)$ for $p = 2$. The simple averaged minimum variance optimal estimators $\hat{\sigma}_{BLUE}^1(k)$ and $\hat{\sigma}_{GK}^2(k)$ are dominated by the average of AMRE estimates of single candlesticks $\hat{\sigma}_s^{1,A}(k)$

Table 3.2: Relative Efficiency of Candlestick-based Estimators of Spot Volatility

k	$\hat{\sigma}_q^{1,*}(k)$	$\hat{\sigma}_s^{1,A}(k)$	$\hat{\sigma}_q^{1,A}(k)$	$\hat{\sigma}_{BLUE}^1(k)$	$\sqrt{\hat{\sigma}_s^{2,*}(k)}$	$\sqrt{\hat{\sigma}_q^{2,*}(k)}$	$\sqrt{\hat{\sigma}_s^{2,A}(k)}$	$\sqrt{\hat{\sigma}_q^{2,A}(k)}$	$\sqrt{\hat{\sigma}_{GK}^2(k)}$
Panel I: Relative efficiency under the Stein's loss									
1	0.9440	1.0000	0.9440	0.9908	0.9851	0.7509	0.9851	0.7509	0.9613
2	0.9708	0.9827	0.8819	0.9755	0.9925	0.8479	0.9542	0.6349	0.9274
3	0.9805	0.9743	0.8322	0.9677	0.9951	0.8915	0.9415	0.5546	0.9131
4	0.9852	0.9691	0.7892	0.9627	0.9963	0.9157	0.9342	0.4935	0.9054
5	0.9882	0.9659	0.7517	0.9596	0.9970	0.9313	0.9301	0.4454	0.9009
6	0.9901	0.9634	0.7177	0.9574	0.9975	0.9419	0.9268	0.4057	0.8976
7	0.9915	0.9615	0.6869	0.9557	0.9979	0.9497	0.9245	0.3727	0.8949
8	0.9925	0.9600	0.6589	0.9542	0.9981	0.9556	0.9227	0.3448	0.8928
9	0.9933	0.9589	0.6329	0.9532	0.9983	0.9602	0.9215	0.3205	0.8912
10	0.9939	0.9580	0.6090	0.9523	0.9984	0.9639	0.9201	0.2996	0.8896
11	0.9945	0.9571	0.5872	0.9514	0.9986	0.9671	0.9190	0.2813	0.8884
12	0.9950	0.9565	0.5670	0.9508	0.9987	0.9699	0.9182	0.2652	0.8877
13	0.9954	0.9559	0.5481	0.9501	0.9988	0.9721	0.9175	0.2508	0.8868
14	0.9957	0.9554	0.5305	0.9497	0.9989	0.9741	0.9170	0.2379	0.8863
15	0.9960	0.9550	0.5143	0.9493	0.9990	0.9759	0.9166	0.2265	0.8857
16	0.9963	0.9549	0.4989	0.9491	0.9991	0.9774	0.9163	0.2160	0.8856
17	0.9965	0.9543	0.4844	0.9486	0.9991	0.9788	0.9158	0.2065	0.8850
18	0.9968	0.9540	0.4707	0.9482	0.9992	0.9800	0.9154	0.1977	0.8845
19	0.9970	0.9536	0.4577	0.9478	0.9993	0.9811	0.9150	0.1896	0.8841
20	0.9971	0.9535	0.4454	0.9478	0.9993	0.9821	0.9148	0.1822	0.8838
k	$\hat{\sigma}_s^{2,*}(k)$	$\hat{\sigma}_s^{2,A}(k)$	$\hat{\sigma}_q^{2,A}(k)$	$\hat{\sigma}_{GK}^2(k)$	$(\hat{\sigma}_s^{1,*}(k))^2$	$(\hat{\sigma}_q^{1,*}(k))^2$	$(\hat{\sigma}_s^{1,A}(k))^2$	$(\hat{\sigma}_q^{1,A}(k))^2$	$(\hat{\sigma}_{BLUE}^1(k))^2$
Panel II: Relative efficiency under the quadratic loss									
1	0.9408	0.9408	1.0000	0.9357	0.9851	0.9024	0.9851	0.9024	0.9593
2	0.9695	0.9551	0.9601	0.9503	0.9923	0.9423	0.9580	0.7550	0.9298
3	0.9796	0.9567	0.9116	0.9519	0.9948	0.9594	0.9453	0.6521	0.9158
4	0.9848	0.9564	0.8657	0.9515	0.9962	0.9687	0.9376	0.5749	0.9080
5	0.9878	0.9560	0.8243	0.9510	0.9969	0.9747	0.9332	0.5149	0.9033
6	0.9899	0.9553	0.7863	0.9504	0.9975	0.9786	0.9296	0.4661	0.8998
7	0.9914	0.9547	0.7517	0.9500	0.9978	0.9815	0.9270	0.4260	0.8969
8	0.9926	0.9542	0.7200	0.9494	0.9982	0.9837	0.9251	0.3924	0.8947
9	0.9935	0.9538	0.6906	0.9490	0.9984	0.9854	0.9237	0.3634	0.8930
10	0.9942	0.9535	0.6637	0.9487	0.9986	0.9867	0.9223	0.3386	0.8914
11	0.9947	0.9530	0.6390	0.9482	0.9987	0.9879	0.9210	0.3170	0.8900
12	0.9951	0.9528	0.6162	0.9479	0.9988	0.9890	0.9200	0.2981	0.8892
13	0.9955	0.9525	0.5949	0.9475	0.9989	0.9898	0.9191	0.2812	0.8882
14	0.9958	0.9523	0.5751	0.9473	0.9990	0.9905	0.9186	0.2663	0.8876
15	0.9960	0.9521	0.5568	0.9471	0.9990	0.9912	0.9180	0.2530	0.8868
16	0.9963	0.9521	0.5396	0.9471	0.9991	0.9918	0.9176	0.2409	0.8866
17	0.9965	0.9517	0.5234	0.9466	0.9991	0.9923	0.9170	0.2300	0.8860
18	0.9966	0.9514	0.5081	0.9463	0.9991	0.9928	0.9165	0.2199	0.8854
19	0.9968	0.9512	0.4935	0.9461	0.9992	0.9932	0.9160	0.2106	0.8849
20	0.9969	0.9512	0.4799	0.9462	0.9992	0.9936	0.9158	0.2022	0.8845

Note: For both the Stein's loss and the quadratic loss, the numbers in the table reports the relative efficiency of each estimator computed based on equation (3.9) using 10^6 simulated candlestick data using the exact simulation scheme in Theorem 3.3.

Table 3.3: Relative Efficiency of Candlestick-based Estimators of Spot Variance

k	$\hat{\sigma}_q^{2,*}(k)$	$\hat{\sigma}_s^{2,A}(k)$	$\hat{\sigma}_q^{2,A}(k)$	$\hat{\sigma}_{GK}^2(k)$	$(\hat{\sigma}_s^{1,*}(k))^2$	$(\hat{\sigma}_q^{1,*}(k))^2$	$(\hat{\sigma}_s^{1,A}(k))^2$	$(\hat{\sigma}_q^{1,A}(k))^2$	$(\hat{\sigma}_{BLUE}^1(k))^2$
Panel I: Relative efficiency under the Stein's loss									
1	0.8301	1.0000	0.8301	0.9749	0.9850	0.9851	0.9850	0.9851	0.9778
2	0.8989	0.9634	0.6943	0.9357	0.9922	0.9926	0.9763	0.9259	0.9703
3	0.9286	0.9482	0.6021	0.9191	0.9948	0.9951	0.9704	0.8742	0.9647
4	0.9449	0.9395	0.5329	0.9102	0.9961	0.9963	0.9664	0.8286	0.9608
5	0.9553	0.9344	0.4789	0.9048	0.9969	0.9971	0.9638	0.7885	0.9582
6	0.9623	0.9305	0.4347	0.9009	0.9975	0.9975	0.9617	0.7521	0.9563
7	0.9674	0.9277	0.3984	0.8978	0.9978	0.9979	0.9602	0.7193	0.9549
8	0.9712	0.9256	0.3677	0.8955	0.9981	0.9981	0.9589	0.6893	0.9536
9	0.9742	0.9241	0.3412	0.8936	0.9984	0.9983	0.9579	0.6615	0.9527
10	0.9766	0.9226	0.3183	0.8918	0.9986	0.9984	0.9572	0.6361	0.9519
11	0.9787	0.9213	0.2985	0.8904	0.9987	0.9986	0.9563	0.6128	0.9510
12	0.9805	0.9203	0.2810	0.8895	0.9988	0.9987	0.9558	0.5913	0.9505
13	0.9820	0.9194	0.2654	0.8885	0.9989	0.9988	0.9553	0.5712	0.9499
14	0.9833	0.9188	0.2515	0.8879	0.9990	0.9989	0.9549	0.5524	0.9495
15	0.9845	0.9182	0.2392	0.8871	0.9990	0.9990	0.9545	0.5352	0.9492
16	0.9854	0.9178	0.2280	0.8869	0.9991	0.9991	0.9544	0.5189	0.9490
17	0.9863	0.9172	0.2178	0.8863	0.9991	0.9991	0.9538	0.5036	0.9484
18	0.9872	0.9167	0.2084	0.8857	0.9991	0.9992	0.9535	0.4891	0.9480
19	0.9879	0.9162	0.1997	0.8852	0.9992	0.9993	0.9531	0.4753	0.9477
20	0.9885	0.9160	0.1918	0.8848	0.9992	0.9993	0.9531	0.4623	0.9477
k	$\hat{\sigma}_s^{2,*}(k)$	$\hat{\sigma}_s^{2,A}(k)$	$\hat{\sigma}_q^{2,A}(k)$	$\hat{\sigma}_{GK}^2(k)$	$(\hat{\sigma}_s^{1,*}(k))^2$	$(\hat{\sigma}_q^{1,*}(k))^2$	$(\hat{\sigma}_s^{1,A}(k))^2$	$(\hat{\sigma}_q^{1,A}(k))^2$	$(\hat{\sigma}_{BLUE}^1(k))^2$
Panel II: Relative efficiency under the quadratic loss									
1	0.7920	0.7920	1.0000	0.7650	0.6912	0.8818	0.6912	0.8818	0.6909
2	0.8856	0.8586	0.9018	0.8293	0.8267	0.9349	0.8169	0.9823	0.8153
3	0.9220	0.8794	0.7892	0.8492	0.8807	0.9557	0.8624	0.9778	0.8601
4	0.9412	0.8888	0.6959	0.8585	0.9096	0.9667	0.8854	0.9484	0.8825
5	0.9527	0.8945	0.6213	0.8639	0.9271	0.9732	0.8990	0.9128	0.8958
6	0.9606	0.8976	0.5600	0.8672	0.9391	0.9777	0.9077	0.8756	0.9045
7	0.9663	0.8998	0.5097	0.8692	0.9478	0.9809	0.9140	0.8397	0.9108
8	0.9706	0.9015	0.4676	0.8707	0.9543	0.9834	0.9186	0.8055	0.9152
9	0.9740	0.9028	0.4315	0.8718	0.9595	0.9854	0.9223	0.7732	0.9188
10	0.9767	0.9037	0.4007	0.8725	0.9637	0.9869	0.9253	0.7432	0.9218
11	0.9787	0.9042	0.3739	0.8729	0.9669	0.9880	0.9274	0.7152	0.9237
12	0.9805	0.9047	0.3506	0.8735	0.9696	0.9890	0.9293	0.6892	0.9255
13	0.9820	0.9049	0.3299	0.8736	0.9719	0.9899	0.9307	0.6648	0.9269
14	0.9832	0.9054	0.3116	0.8741	0.9739	0.9906	0.9321	0.6421	0.9282
15	0.9843	0.9056	0.2955	0.8741	0.9755	0.9911	0.9331	0.6212	0.9292
16	0.9852	0.9060	0.2809	0.8747	0.9770	0.9917	0.9343	0.6014	0.9304
17	0.9860	0.9060	0.2676	0.8748	0.9783	0.9921	0.9349	0.5828	0.9309
18	0.9867	0.9061	0.2555	0.8748	0.9794	0.9925	0.9356	0.5651	0.9315
19	0.9874	0.9061	0.2443	0.8748	0.9804	0.9928	0.9361	0.5484	0.9320
20	0.9880	0.9064	0.2342	0.8749	0.9813	0.9932	0.9369	0.5328	0.9329

Note: For both the Stein's loss and the quadratic loss, the numbers in the table reports the relative efficiency of each estimator computed based on equation (3.9) using 10^6 simulated candlestick data using the exact simulation scheme in Theorem 3.3.

and $\hat{\sigma}_s^{2,A}(k)$ due to the functional form restriction of the former estimators.

3.1.5 Understanding the AMRE Estimators

Given the rather obscure forms of the AMRE estimators given in Theorem 3.2, in this section, we attempt to elucidate the expressions of AMRE estimators with the observed candlestick features, which provides a more transparent interpretation of the AMRE estimators through the lens of traditional technical analysis. Specifically, we delve into various features of the candlestick data \mathbf{C}_k that might elucidate the behavior of the AMRE estimators. Since the AMRE estimators in Theorem 3.2 remain indifferent to the ordering of the candlesticks, it must depend symmetrically on all individual candlestick features, or equivalently, the average features among all candlesticks. As shown in the analysis below, in the presence of multiple candlesticks, the AMRE estimator is also expected to depend on the heterogeneity of the features.

Based on the above discussion, we propose to summarize the average size and heterogeneity of features by the power transformation of their sample means and standard deviations. This transformation aims to capture some nonlinear dependence structure that might exist. In detail, the features are

$$(x_j)_{1 \leq j \leq 6} \equiv \{\bar{w}, \overline{|r|}, \bar{a}, v(w), v(|r|), v(a)\},$$

where for a random sample z we define $\bar{z} \equiv k^{-1} \sum_{i=1}^k z_i$, and $v(z) \equiv \sqrt{k^{-1} \sum_{i=1}^k (z_i - \bar{z})^2}$.

Following the polynomial design of the BLUE estimator and the Garman–Klass estimator, we propose to approximate the AMRE estimator as a polynomial of the observed candlestick features²

$$X(p) \equiv \left\{ \prod_{j=1}^6 x_j^{\alpha_j} : \alpha_j \in \{0, \dots, p\} \text{ such that } \sum_{j=1}^6 \alpha_j = p \right\},$$

which sheds light on the source of precision loss for the polynomial-based estimators. To select the q most important features from a pool of possible features, we

²We use these polynomial-type features for the following reasons. Firstly, for estimating variance, the setting nests those features such as the sample average of squared range by noting that $\overline{w^2} = \bar{w}^2 + v(w)^2$. Additionally, as per Theorem 3.2, note that $\hat{\sigma}_s^{p,*}(k) \hat{\sigma}_q^{p,*}(k) \equiv \hat{\sigma}_s^{2p,*}(k)$, implying that the product of the AMRE estimator of σ_t^p under Stein's and quadratic loss equals to the AMRE estimator of σ_t^{2p} under Stein's loss. Using polynomial-type features preserves a similar structure for approximated estimators.

employ the best subset regression approach. Specifically, given a vector $y_{N \times 1}$ of N simulated AMRE estimators and a matrix of simulated features $\mathbf{X}_{N \times |X(p)|}$, the q most important features out of $|X(p)|$ total features can be selected by solving the following best subset regression problem:

$$\min_{\beta \in \mathbb{R}^{|X(p)|}} \|y - \mathbf{X}\beta\|_2^2, \quad \text{subject to } \|\beta\|_0 \leq q,$$

where β is the $|X(p)|$ -by-1 vector of coefficients for the $|X(p)|$ features, $\|\cdot\|_p$ is the ℓ_p -norm of a vector, and when $p = 0$, $\|\cdot\|_0$ simply counts the number of non-zero elements of the vector. Although the best subset regression is known to be an NP-hard problem due to the discrete and vast dimension of the space of possible subsets [Welch \(1982\)](#), recent advances in mixed-integer optimization enable efficient solutions to such problems, following procedures proposed in [Bertsimas et al. \(2016\)](#). Below, we focus on the results of approximating optimal volatility estimators, with the approximation of optimal variance estimators provided in the online appendix.

Table 3.4 shows how different features are sequentially selected in the best subset regression of optimal volatility estimators. Notably, the selected features remain consistent across different loss functions. Additionally, the standard error of the range $v(w)$ emerges as the third most important feature and even outweighs the asymmetry level \bar{a} . For instance, consider the case when $k = 5$. The approximation using two features is close to the BLUE estimator proposed in [Li et al. \(2022\)](#):

$$\hat{\sigma}_{(s)}^{1,*}(5) \approx 0.809\bar{w} - 0.365\overline{|r|}.$$

However, upon increasing the number of features to three, the approximation includes an additional term not considered in the BLUE estimator:

$$\hat{\sigma}_{(s)}^{1,*}(5) \approx 0.824\bar{w} - 0.346\overline{|r|} - 0.089v(w).$$

This additional inclusion arises due to the absence of a dimension-reduced complete sufficient statistic for the joint distribution of multiple candlesticks. Consequently, the dispersion among different candlesticks contributes to optimal estimation, which has not been taken into account by the approach of averaging single candlestick estimations as in [Garman and Klass \(1980\)](#) and [Li et al. \(2022\)](#). As a result, the relative efficiency of those estimators decreases as k grows, given that this part of information is continuously discarded in their construction.

Table 3.4: Best Subset Regression Result of $\hat{\sigma}_{(\cdot)}^{1,*}(k)$

q	\bar{w}	\bar{a}	$\overline{ r }$	$\nu(w)$	$\nu(a)$	$\nu(r)$	R.Eff.
Panel I: AMRE under Stein's loss							
1	19	0	0	0	0	0	67.7%
2	19	0	19	0	0	0	95.4%
3	19	0	19	19	0	0	96.8%
4	19	19	19	19	0	0	97.1%
5	19	19	19	19	0	19	97.2%
6	19	19	19	19	19	19	97.2%
Panel II: AMRE under Quadratic loss							
1	19	0	0	0	0	0	67.6%
2	19	0	19	0	0	0	95.6%
3	19	0	19	19	0	0	96.9%
4	19	19	19	19	0	0	97.2%
5	19	19	19	19	0	19	97.2%
6	19	19	19	19	19	19	97.2%

Note: The table reports the total number of selections for $k \in \{2, \dots, 20\}$ of each feature, along with the average relative efficiency w.r.t. the exact AMRE estimators for $q \in \{1, \dots, 6\}$.

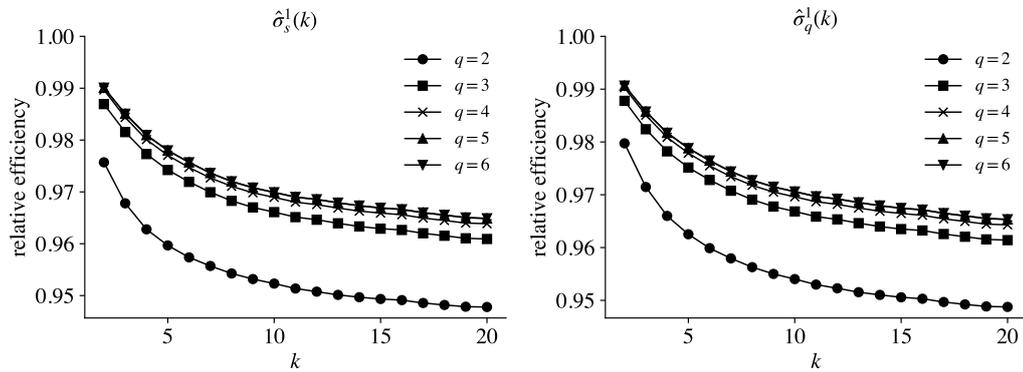


Figure 3.2: Relative efficiency of polynomial-based approximation for optimal volatility estimators under Stein's loss (left) and quadratic loss (right), with $q \in \{2, \dots, 6\}$ and $k \in \{2, \dots, 20\}$. The selection of features and determination of corresponding coefficients are computed using best subset regression.

Figure 3.2 shows the pattern of relative efficiency of polynomial-based approximation for optimal volatility estimators. With a fixed number of features q , increases, the relative efficiency decreases. As mentioned previously, this is due to the absence of a dimension-reduced complete sufficient statistic. Therefore, larger k naturally implies a greater loss of information, resulting in an estimator further from the optimal ones that utilize all available information. However, this decrease become gradual in the case of large k , suggesting that one may trade some efficiency for faster computation and easier-to-understand estimators. Moreover, the figure indicates that the marginal improvement of adding an additional feature becomes nearly negligible as q becomes larger, say, greater than 3. For instance, for the optimal estimator under Stein's loss with $k = 5$, the relative efficiency levels of approximation using $q = 1, 2, 3, 4$ features are 53.08%, 95.70%, 97.31%, 97.61%, respectively. Based on this observation, a three-feature polynomial-based approximation seems to strike a balance, preserving a reasonable level of efficiency while remaining practically useful and computationally convenient.

3.2 An Empirical Illustration

Consumer Price Index (CPI), Producer Price Index (PPI), and Personal Consumption Expenditures (PCE) are fundamental economic indicators used to measure inflation in an economy. CPI measures the average change over time in the prices paid by urban consumers for a predetermined basket of consumer goods and services. In contrast, PPI monitors changes in selling prices received by domestic producers. PCE, on the other hand, has a broader perspective, including not only goods and services bought by households but also those purchased by nonprofits, governments, and businesses. The primary distinction among them lies in their scope and intended audience.

These indicators are announced regularly by U.S. government agencies such as the Bureau of Labor Statistics and the Bureau of Economic Analysis.³ The Federal Reserve closely monitors price indices as part of its mandate to maintain stable prices and maximum employment. If inflation appears to be above the target level,

³Detailed schedules and release can be found on, e.g., <https://www.bls.gov/bls/newsrels.htm>.

the Fed may decide to increase interest rates to cool down the economy and prevent out-of-control inflation. Conversely, if inflation falls significantly below the target level, the economy shows signs of sluggishness, the Fed may reduce interest rates to stimulate borrowing and spending. Therefore, these inflation measures play a pivotal role in shaping the Fed's monetary policy decisions, including changes in interest rates and government purchases, hence can be used to predict potential monetary shocks (Romer and Romer (2004), Rigobon and Sack (2008), Miranda-Agrippino and Ricco (2021)). Consequently, the release of these price indices often triggers significant market reactions owing to their implications for future monetary policy decisions.

Normally, CPI is released first, followed by PPI, while PCE is released several days later due to its more comprehensive measurements. These releases occur at 8:30 a.m., one hour before the market opens. To analyze how the market reacts to these releases, we use the high-frequency price data of E-mini S&P500 future continuous contract obtained from Tick Data.⁴ We estimate spot volatility using five 1-minute data points within a window of 1.5 hours before and after the release, the prices right after the release are excluded to mitigate the effects of potential price jumps.

Figure 3.3 shows the estimated volatility for the last four releases of price indices in 2023. The left columns are the AMRE estimates under Stein's loss, where the right columns are the conventional return-based estimates. As shown in the left column, there is a consistent pattern of volatility for all releases. Firstly, volatility increases after the release, then slowly reverts back within 60 minutes. Then, another increase occurs due to the market opens. Meanwhile, the impact of the CPI release is larger than that of the other two indices, despite PCE being the Fed's choice for target inflation rate and should be the direct source for predicting monetary shocks. This is because CPI is released first, and these price indices are highly correlated. Therefore, the new CPI level can be used to update the prediction of PPI and PCE. Such a regular pattern, however, is hard to discern from the conventional return-based estimates by looking at the right column, as the estimation appears to be too noisy. This comparison illustrates the added precision afforded by the

⁴See <https://www.tickdata.com/product/historical-futures-data/>.

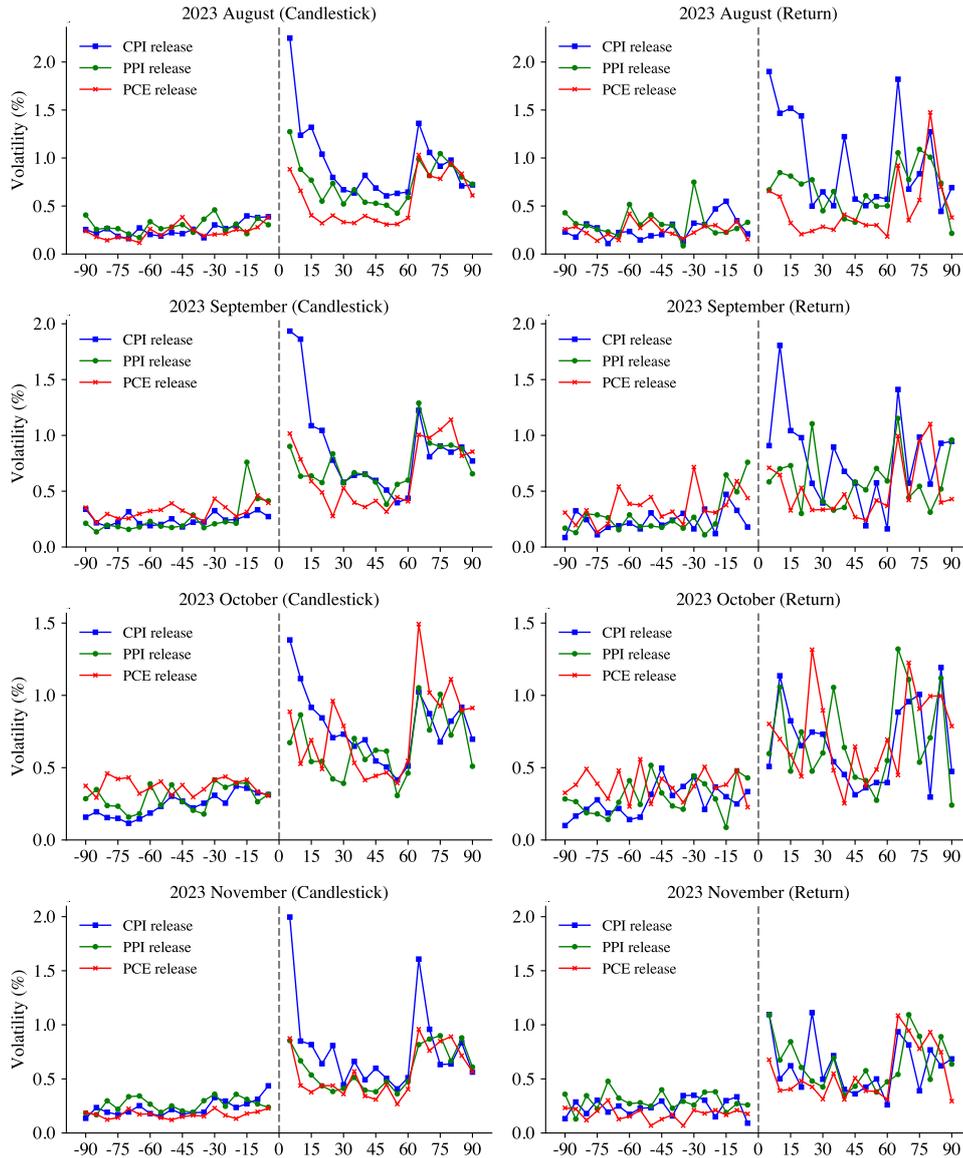


Figure 3.3: Spot volatility estimation of E-mini S&P500 future within a 1.5 hours window before and after price indices releases. The estimation is based on five consecutive observation intervals sampled at a 1-minute frequency. The left (resp. right) column are AMRE estimates under Stein’s loss (resp. conventional return-based estimates). Prices immediately following the releases are excluded to mitigate the effects of potential price jumps.

candlestick estimator vis-à-vis the traditional return-based volatility estimator, and highlights the superiority of using candlesticks in identifying high-frequency market activities over short windows (Nakamura and Steinsson (2018a) and Bollerslev et al. (2018)).

3.3 Concluding Remarks

We study optimal spot volatility estimators based on multiple candlesticks. In contrast to the machine learning-based method proposed in Bollerslev et al. (2024), we first propose an alternative computational algorithm that is both less computationally intensive and more accurate. Additionally, we introduce an exact simulation scheme that overcomes the one-sided bias issue inherent in Euler discretization schemes when dealing with supremum and infimum. This exact simulation scheme enables more precise risk comparison and facilitates further analysis involving extreme values of Brownian motions. Using these tools, the comparison of optimal estimators with existing ones like those in Garman and Klass (1980) and Li et al. (2022) reveals that the relative relationship between different candlesticks plays a crucial role in delivering more efficient estimations. This aspect has not been adequately considered in existing estimators. These findings could also provide insights into deriving more accurate estimations of candlestick-based integrated volatility functionals, a topic we leave for future research.

Chapter 4

Uniform Inference for High-Frequency Data

4.1 Introduction

As high-frequency financial data becomes increasingly accessible, development of inference methods tailored for such data emerges as a trending topic. In particular, inference for volatility or jumps using high-frequency prices has been extensively studied (see, e.g., [Jacod and Protter \(2012\)](#), [Aït-Sahalia and Jacod \(2014\)](#)). However, the workhorse model for price data used by most researchers, an Itô semi-martingale plus noise, is evidently unsuitable for other market indicators, such as volumes and trading flows.^{1,2} To accommodate a broader range of high-frequency data, [Li and Xiu \(2016\)](#) proposed a continuous-time state-space model, in which the observed data approximately equals a general transformation of the state process and some random disturbance. Special cases include price, volume, and trading

¹In contrast to prices, volumes and trading flows are discrete-valued and may not exhibit long-memory properties. Consequently, they cannot be accommodated within the conventional Itô semi-martingale model.

²Other market indicators deserve their own analysis. It is widely acknowledged that price and volumes may carry different aspects of information. For example, [Kandel and Pearson \(1995\)](#) argued that certain news may not significantly alter investors' average opinion, resulting in small price movement. However, it can lead to a substantial divergence of opinions among investors, prompting disagreements and subsequent trading activity between these disagreeing investors. Consequently, a surge in volume is observed but no significant change in price. Conversely, [Kyle \(1985\)](#) demonstrated that in the presence of asymmetric information, even a small trade can trigger a notable price change if the market impact is significant. In a nutshell, price reflects the average opinion (first moment), while trading activity also reflects the dispersion of opinion (second moment).

flow, with the corresponding states being volatility, average order size, and trading intensity. Without specific restrictions on the state dynamics and the functional form of transformations, this framework exhibits great versatility to accommodate various model specifications, such as the Poisson volume-volatility model (Andersen (1996)) and Cox trading flow model (Christensen and Kolokolov (2023)).

In this paper, we adopt the general state-space framework proposed by Li and Xiu (2016). Our emphasis lies in the *uniform inference*, which speaks to global properties of the entire state process. Specifically, we develop functional estimators and associated inference procedures for distributional features of the transformed state process. These functional estimators are constructed by collecting all localized estimates across different time points. The major challenge in uniform inference stems from the asymptotic independence of estimation errors between distinct time points. Consequently, the functional estimators do not admit a functional central limit theorem. Recent literature shed light on such non-Donsker problems, highlighting the use of strong approximation, or coupling (see, e.g., Chernozhukov et al. (2013), Belloni et al. (2015), and Li and Liao (2020)). Building on this insight, our contribution in this paper is to establish a Gaussian coupling theory for functional estimators of both the conditional mean process (Theorem 4.1) and the conditional quantile process (Theorem 4.2). These results are formulated within the general state-space model aligned with various high-frequency data, accommodating dependencies and nonstationarity in both state processes and observations.

A large body of literature has evolved around estimations of volatility using high-frequency returns, a specific case of our general state-space model. In particular, the nonparametric estimation of the stochastic volatility at some fixed time point, referred to as spot estimation (see, e.g., Foster and Nelson (1996) and Comte and Renault (1998)), and the semiparametric estimation of integrated volatility functionals (see, e.g., Andersen et al. (2003), Barndorff-Nielsen and Shephard (2004), Mykland and Zhang (2009)) have been extensively explored.^{3,4} However, the uniform

³These problems are closely related to each other, in the sense that nonparametric spot volatility estimators can be used to construct semiparametrically efficient estimators of integrated volatility functionals (see, e.g., Jacod and Rosenbaum (2013), Li et al. (2017), and Renault et al. (2017)).

⁴Another problem, which is orthogonal to the nonparametric setting, involves estimating parameters within specified volatility dynamics. Such models are suggested by Nelson (1990) and Heston (1993), with associated estimation methods proposed in Harvey et al. (1994), Andersen and Sørensen

inference for the entire volatility process is an emerging concern, as recently explored by [Jacod et al. \(2021\)](#) and [Bollerslev et al. \(2021\)](#). In line with this literature, spot estimation for the state process under general state-space model is developed in [Bollerslev et al. \(2018\)](#). Setting against this background, the strong approximation result regarding conditional mean process in this paper can be contextualized as an extension of [Jacod et al. \(2021\)](#) to a more general state-space setting.⁵

Meanwhile, inference concerning quantiles is relatively underexplored in high-frequency literature. In a recent paper, [Shephard \(2022\)](#) introduced an estimator of integrated variance based on in-fill medians. The use of quantiles becomes notably significant when returns display heavy tails, a common feature observed in cryptocurrency markets (see, e.g., [Kolokolov \(2022\)](#)). Our Gaussian strong approximation for conditional quantile process is derived, in part, by a novel uniform Bahadur representation for all in-fill quantiles (Lemma [C.1](#)). While such representation has been established for i.i.d. data ([Bahadur \(1966\)](#), [Ghosh \(1971\)](#)) and weakly dependent stationary data ([Hesse \(1990\)](#), [Wu \(2005\)](#)), our observations are nonstationary and may exhibit strong dependencies due to the persistence within the state process. Notably, as a special case, our results can be applied to capture volatile level of Lévy-driven price. To the best of our knowledge, this is the first paper that contributes to the uniform inference of these processes.

The established strong approximation results have broader applications in tackling other econometric problems. As a byproduct, we provide an application involving constructing confidence sets for ranks of spot values of the studied process, which is typically useful in determining arrivals of certain events. Specifically, we leverage insights from [Mogstad et al. \(2023\)](#), reframing the construction as a multiple hypotheses testing problem. Notably, our strong approximation results aid in determining the valid critical value required for this purpose. The paper is also related to prior studies in Gaussian coupling, such as [Chernozhukov et al. \(2013\)](#), [Belloni et al. \(2015\)](#), [Li and Liao \(2020\)](#), our work stands out due to its emphasis on a nonstationary time series setting, a departure from the high-dimensional context typically explored in previous studies.

(1996), [Durbin and Koopman \(1997\)](#), and [Knight and Yu \(2002\)](#), among others.

⁵An extension under fixed- k framework akin to [Bollerslev et al. \(2021\)](#) is feasible with additional information about the transformation and the distribution of random disturbances.

As a concrete empirical illustration of the proposed methodology, we conduct a sentence-by-sentence study to discern the informative part of the Federal Open Market Committee (FOMC) press conference speeches. In light of the more accurate volatility estimations, recent observations by [Bollerslev et al. \(2024\)](#) highlight that press conferences sometimes trigger more pronounced market impact than the initial release of FOMC statements. We employ the uniform inference procedure to analyze trading intensity processes, aiming to identify additional information arrivals during these press conferences. Our comparison of results to stand-alone textual analysis reveals that the latter tends to inappropriately smooth out information flow. In view of the growing attention towards generative AI tools and large language models, primarily leveraging in-context learning for tasks, our method serves as a complement, enabling the deployment of supervised learning for higher accuracy. Additionally, we provide another empirical application to highlight the importance of employing quantiles in addressing specific problems. Due to the heavy-tailedness of Bitcoin returns, realized variances computed in the usual way become diverging, rendering the detection of abnormal returns invalid. Comparing to the outcomes of mean-based t -test in [Ante \(2023\)](#), results using quantile-based measurements of volatile levels indicate more substantial price impact over an extended time window following social media activities.

The rest of the paper is organized as the following. We present the theory in Section 4.2. In Section 4.3, a Monte-Carlo experiment analysis is conducted. Two empirical studies are presented in Section 4.4, where the proposed inference methodology is applied to discern information flows during the FOMC press conference speeches, and to analyze the price impact of Elon Musk’s twitter on Bitcoin. Section 4.5 concludes. The appendix contains all the proofs.

Notation. We use $|\cdot|$ to denote the absolute value of a real scalar or the cardinality of a set, $\|\cdot\|$ to denote the vector ℓ_2 -norm. For any $p \geq 1$, $\|\cdot\|_{L_p}$ denotes the L_p -norm for random variables. We use $\mathcal{L}(\cdot)$ to denote the law of random objects, use $\mathbb{1}\{\cdot\}$ to denote the indicator function. For two real numbers a and b , we write $\min\{a, b\}$ as $a \wedge b$ and $\max\{a, b\}$ as $a \vee b$. For two real sequences a_n and b_n , we write $a_n \asymp b_n$ if $a_n/C \leq b_n \leq Ca_n$ for some finite constant $C \geq 1$.

4.2 Theory

In Section 4.2.1, we introduce the state-space model employed in our research. In Section 4.2.2, three running examples are provided to illustrate adaptability of our framework for modeling different market indicators. Sections 4.2.3 and 4.2.4 present constructions and strong approximation results for the functional estimators of both conditional mean processes and conditional quantile processes, respectively. Section 4.2.5 provides an application in constructing confidence sets for ranks of spot values of the investigated process.

4.2.1 State-space Model for High-Frequency Data

We observe a data sequence $(Y_{i\Delta_n})$ at some regular sampled times where $1 \leq i \leq n \equiv \lfloor T/\Delta_n \rfloor$, within a *fixed* time span $[0, T]$. In what follows, we consider *in-fill* asymptotics, i.e., $\Delta_n \rightarrow 0$. It is assumed that the data is generated based on the following state-space model

$$Y_{i\Delta_n} = \mathcal{Y}(\zeta_{i\Delta_n}, \varepsilon_{n,i}) + R_{n,i}, \quad \text{for } 1 \leq i \leq n, \quad (4.1)$$

where $(\zeta_t)_{t \in [0, T]}$ is a càdlàg state process which takes value in an open set \mathcal{Z} and is defined on some filtered probability space satisfying the usual conditions, denoted as $(\Omega^{(0)}, \mathcal{F}^{(0)}, (\mathcal{F}_t^{(0)})_{t \geq 0}, \mathbb{P}^{(0)})$. The function $\mathcal{Y}(\cdot, \cdot)$ represents a deterministic noisy transform of the *current* state $\zeta_{i\Delta_n}$ through a random disturbance $\varepsilon_{n,i}$ which takes value in some Polish space \mathcal{D} . Additionally, $R_{n,i}$ denotes a residual term, which is defined on an extended probability space that will be elaborated upon later.⁶ This residual term can be considered uniformly negligible in comparison with the dominant term, as per the requirement provided in subsequent sections.

We will make the assumption that the random disturbance $(\varepsilon_{n,i})_{1 \leq i \leq n}$ is a $\mathcal{F}^{(0)}$ -conditionally independently and identically distributed (i.i.d.) sequence.⁷ This is not a necessary condition, as the framework presented here can be extended to accommodate conditionally stationary and weakly dependent disturbances by employ-

⁶The incorporation of residual term is first proposed in Bugni et al. (2023), and is assumed to be zero in Li and Xiu (2016) and Bollerslev et al. (2018).

⁷There is no loss of generality to impose independence between disturbances and state processes here. One can always select an appropriate normalization of the representation to let $\mathcal{Y}(\cdot, \cdot)$ account for the dependence structure such that $\varepsilon_{n,i}$ is independent from $\zeta_{i\Delta_n}$.

ing methodologies developed in [Zhang and Cheng \(2014\)](#), [Li and Liao \(2020\)](#), and [Cattaneo et al. \(2022\)](#). However, it is worth mentioning that, in many empirical scenarios illustrated in the examples provided in Section 4.2.2, the disturbance exhibits conditional independence. Hence, in order to avoid unnecessary technical complexities, our primary focus lies on conditional independent disturbances, whereas the extension to dependent case will be discussed in the Appendix. In order to formally describe the framework, we introduce another probability space denoted as $(\Omega^{(1)}, \mathcal{F}^{(1)}, \mathbb{P}^{(1)})$ endowed with an i.i.d. sequence $(\varepsilon_{n,i})_{1 \leq i \leq n}$ with its marginal distribution denoted by \mathbb{P}_ε . Additionally, we denote

$$\Omega \equiv \Omega^{(0)} \times \Omega^{(1)}, \quad \mathcal{F} \equiv \mathcal{F}^{(0)} \otimes \mathcal{F}^{(1)}, \quad \mathcal{F}_t \equiv \bigcap_{s>t} \mathcal{F}_s^{(0)} \otimes \sigma(\varepsilon_s : s \leq t), \quad \mathbb{P} \equiv \mathbb{P}^{(0)} \otimes \mathbb{P}^{(1)}.$$

In this context, processes defined in each individual space, whether $\Omega^{(0)}$ or $\Omega^{(1)}$, can be extended in the usual way to product space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$, which serves as the probability space underlying our analysis.

We highlight that the seeming Markovian assumption that observation Y_t solely relies on current state ζ_t through the function $\mathcal{Y}(\cdot, \cdot)$ is not overly restrictive owing to the inclusion of additional residual term $R_{n,i}$. Although, from an intuitive standpoint, Y_t could potentially depend on historical states. Given that state processes exhibit sufficient smoothness, information encapsulated in the difference between past state and current state could be effectively captured within the residual term. For example, when the observation Y_t depends on a local window of historical states $(\zeta_s)_{s \in [t-h, t]}$ through some noisy functional, this approximation holds when (i) the functional has a bounded partial Fréchet derivative with respect to $(\zeta_s)_{s \in [t-h, t]}$; (ii) the state process ζ is smooth enough in a proper sense, e.g., $\sup_{s,r \in [t-h, t]} \|\zeta_s - \zeta_r\| = O_p(h)$; and (iii) window size is shrinking, i.e., $h = o(1)$. In the meantime, this additional residual term can also absorb the dependence of observations on some nuisance process when its effect is negligible. Consequently, the incorporation of residual $R_{n,i}$ renders our framework to an essentially “approximately Markovian” setting, which is more general comparing with simpler Markov state-space models employed in [Li and Xiu \(2016\)](#) and [Bollerslev et al. \(2018\)](#).

4.2.2 Motivating Examples

To facilitate a better understanding of broad implications of the general model (4.1), it is beneficial to outline a discussion using some empirically relevant running examples. In this section, we provide three motivating examples, showing how commonly used financial econometric models align with our state-space framework.

EXAMPLE 1 (LOCATION-SCALE MODEL). First, consider a simple model with an additive structure

$$Y_{i\Delta_n} = \mu_{i\Delta_n} + \sigma_{i\Delta_n}\varepsilon_{n,i}, \quad \text{for } 1 \leq i \leq n.$$

In this model, μ_t represents the local mean at time t and σ captures potential heteroskedasticity in time. This additive structure directly fits in model (4.1) by setting

$$\zeta_t = (\mu_t, \sigma_t), \quad \mathcal{Y}((\mu, \sigma), \varepsilon) = \mu + \sigma\varepsilon, \quad R_{n,i} = 0.$$

Note that this elementary model has found applications in various important contexts, as we do not need to specify dynamics of state processes. For example if $Y_{i\Delta_n}$ is the observed price of some derivative contract, then $\mu_{i\Delta_n}$ represents the efficient price and $\sigma_{i\Delta_n}\varepsilon_{n,i}$ could be the pricing error. [Liu and Tang \(2013\)](#) employ this additive state-space model to devise an expectation-maximization algorithm tailored for estimating integrated volatility matrices, particularly when asset prices are observed with microstructure noise. In their model, $Y_{i\Delta_n}$ is observed price, $\mu_{i\Delta_n}$ is the associated latent efficient price and is assumed to have a VAR dynamics, $\sigma_{i\Delta_n}\varepsilon_{n,i}$ is a microstructure noise component where $\sigma_{i\Delta_n}$ captures time-varying heterogeneity in the magnitude of noise. [Bugni et al. \(2023\)](#) also used this additive state-space model to describe trading volume processes, where $\mu_{i\Delta_n}$ is the local mean of volume, and $\sigma_{i\Delta_n}$ captures time-varying heterogeneity in order size. A particularly fitting application of this additive state-space model emerges when the observation is, in itself, a spot estimation of state process. This specification aligns closely with the fixed- k estimation framework introduced in [Bollerslev et al. \(2021\)](#). Specifically, let $\log(\hat{\sigma}_{n,i})$ be the logarithm of fixed- k estimator for spot variance at time $i\Delta_n$, and $\log(\sigma_{n,i})$ be the logarithm of true value. [Bollerslev et al. \(2021\)](#) proved that $\log(\hat{\sigma}_{n,i}) = \log(\sigma_{n,i}) + \varepsilon_{n,i} + o_{pu}(1)$ where $\varepsilon_{n,i}$ follows a scaled log chi-square distribution with degree of freedom k . Based on this formulation, such

additive state-space model is adaptable to various volatility dynamics, for example Hull–White log-normal short-term stochastic volatility. \square

EXAMPLE 2 (LÉVY-DRIVEN ASSET RETURNS). The proposed state-space model can be applied to characterize a wide range of price dynamics studied in the high-frequency financial econometrics literature. Consider the log price which has a drift component and a jump-diffusion component driven by a *Lévy martingale* L , i.e., log price process P_t takes the following form

$$P_t = \int_0^t \mu_s ds + \int_0^t \sigma_s dL_s, \quad \text{for } t \in [0, T],$$

where μ is the drift process, σ is the stochastic volatility process, L is a stable process with Blumenthal–Gettoor index $\beta \in (0, 2]$ and is assumed to be independent with σ .^{8,9} The extension to general stable Lévy process is motivated by empirical evidence that jump index of cryptocurrency prices (see, e.g., [Kolokolov \(2022\)](#)) is strictly smaller than 2, i.e., price is driven by a pure jump process. We treat the value of β as known, then the normalized squared return $Y_{i\Delta_n} = \Delta_n^{-2/\beta} (P_{(i+1)\Delta_n} - P_{i\Delta_n})^2$ over each observation window $(i\Delta_n, (i+1)\Delta_n]$ can be written as

$$Y_{i\Delta_n} = \Delta_n^{-2/\beta} \left(\int_{i\Delta_n}^{(i+1)\Delta_n} \mu_s ds + \int_{i\Delta_n}^{(i+1)\Delta_n} \sigma_s dL_s \right)^2.$$

In light of the property of stable processes, scaled Lévy increments $\Delta_n^{-1/\beta} (L_{(i+1)\Delta_n} - L_{i\Delta_n})$ are i.i.d. across $1 \leq i \leq n$ and have a non-degenerate distribution. Therefore, upon expanding above display and collecting dominant terms, the normalized squared return can be rewritten in the form of model (4.1) by setting

$$\begin{aligned} \zeta_t &= \sigma_t, \quad \varepsilon_{n,i} = \Delta_n^{-1/\beta} (L_{(i+1)\Delta_n} - L_{i\Delta_n}), \quad \mathcal{Y}(\sigma, \varepsilon) = (\sigma \varepsilon)^2, \\ R_{n,i} &= \Delta_n^{-2/\beta} \left(\int_{i\Delta_n}^{(i+1)\Delta_n} \mu_s ds + \int_{i\Delta_n}^{(i+1)\Delta_n} (\sigma_s - \sigma_{i\Delta_n}) dL_s \right)^2 \end{aligned}$$

⁸Note that for a stable process, Blumenthal–Gettoor index and stability index agree. A general stable process has a characteristic triple $(0, c, F)$ where $F(dx) = 0$ if $\beta = 2$, i.e., L is a scaled Brownian motion $\sqrt{c}W$, or $c = 0$ and $F(dx) = a\beta/|x|^{1+\beta} dx$ for some positive constant $a > 0$ if $\beta \in (0, 2)$. In particular, if $\beta = 1$, L is a Cauchy process. Also note that for positive constant K , KL remains a stable process, along with σ/K , generates the same price process. Therefore, to avoid non-identification issues between σ and the “scale” of L , we make additional restriction that $c = 1$ if $\beta = 2$ and $a = 1/\pi$ if $\beta \in (0, 2)$.

⁹The independence assumption between L and σ rules out the interaction between price and volatility, i.e., the so-called “leverage” effect. Note that in this explicit configuration, the transformation has a multiplicative structure, hence it is easy to separate volatility and Lévy increments. That being said, the independence requirement can be dropped here, for the case when L is a Brownian motion, see [Jacod et al. \(2021\)](#).

$$\begin{aligned}
& + 2\Delta_n^{-2/\beta} \left(\int_{i\Delta_n}^{(i+1)\Delta_n} \mu_s ds + \int_{i\Delta_n}^{(i+1)\Delta_n} (\sigma_s - \sigma_{i\Delta_n}) dL_s \right) \\
& \times \sigma_{i\Delta_n} (L_{(i+1)\Delta_n} - L_{i\Delta_n}).
\end{aligned}$$

Distinct with preceding examples, here we encounter the presence of a non-zero residual term $R_{n,i}$. This inclusion stresses the notion that even though $Y_{i\Delta_n}$ may not adhere strictly to Markovian properties with respect to the filtration engendered by current volatility $\sigma_{i\Delta_n}$ and remains dependent on the ancillary drift process μ , it may still conform to an “approximate Markovian” characterization involving only the current volatility. As discussed in subsequent sections 4.2.3 and 4.2.4, the residual term can be proved to be uniformly negligible providing processes μ and σ satisfying some fairly weak regularity conditions. \square

EXAMPLE 3 (COX TRADING FLOWS). Consider the number of trades during time $[0, t]$, as denoted by N_t . It is cogent to model trading flows as a Cox process—or referred to as doubly stochastic Poisson process—which was originally introduced by Cox (1955) for modeling the neps over fibrous threads, i.e., conditional on the process μ , $(N_t)_{t \in [0, T]}$ behaves as an inhomogeneous Poisson process with an intensity function $(\mu_t)_{t \in [0, T]}$. Let $Y_{i\Delta_n} = N_{(i+1)\Delta_n} - N_{i\Delta_n}$ denote number of transactions during each observation window $(i\Delta_n, (i+1)\Delta_n]$. According to the sparseness property of Poisson process (see, e.g., Section 5.4.1 in Ross (1995)), we have (i) $\mathbb{P}(Y_{i\Delta_n} \geq 2 | \mu) = o(\Delta_n)$ and (ii) $\mathbb{P}(Y_{i\Delta_n} = 1 | \mu) = \Delta_n \mu_{i\Delta_n} + o(\Delta_n)$. This naturally suggest a compelling approximation of $Y_{i\Delta_n}$ by a mixed Bernoulli random variable with parameter $\Delta_n \mu_{i\Delta_n}$.¹⁰ Consequently, there exists a sequence of independent, uniformly distributed variables $(\varepsilon_{n,i})_{1 \leq i \leq n}$ on $[0, 1]$ which are also independent of process μ such that

$$Y_{i\Delta_n} = \mathbb{1}\{\varepsilon_{n,i} < \Delta_n \mu_{i\Delta_n}\} + R_{n,i}, \quad \text{for } 1 \leq i \leq n,$$

where the residual takes value in $\{-1\} \cup \mathbb{N}$ and satisfies $\mathbb{P}(R_{n,i} \neq 0 | \mu) = o(\Delta_n)$ according to property (i) and (ii).¹¹ Note that increments over disjoint intervals can

¹⁰The approximation has been explored from a different direction as well, see, e.g., Section 1.6 of Karr (1991) where they discuss the optimal approximation of a Bernoulli point process by a Poisson process.

¹¹In some cases, this approximation holds in a stronger sense. Specifically, let $N_t^n \equiv \sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \mathbb{1}\{\varepsilon_{n,i} < \Delta_n \mu_{i\Delta_n}\}$ denote the partial sum process of these Bernoulli random variables. Under some strong regularity conditions on the intensity function, Theorem 2 in Ruzankin (2004) implies $\|\mathcal{L}(N) - \mathcal{L}(N^n)\|_{\text{TV}} \leq K \Delta_n \sup_{t \in [0, T]} |\mu_t^2|$, where $\|\cdot\|_{\text{TV}}$ denotes the total variation norm of

be in general dependent in a Cox process through the μ_t part, as contrasted with the postulated independence in conventional Poisson processes. Above display shows the increment of trading flow process can be expressed in the form of model (4.1) by setting

$$\zeta_t = \Delta_n \mu_t, \quad \mathcal{Y}(\zeta, \varepsilon) = \mathbb{1}\{\varepsilon < \zeta\}, \quad \varepsilon_{n,i} \sim \text{Uniform}(0, 1).$$

We stress the importance of analyzing trading flow process for following reasons. In the Trade and Quote (TAQ) database, each trade is recorded with a precision of nanoseconds (10^{-9} seconds).¹² Consequently, our mixed Bernoulli approximation exactly matches with empirical data: a binary sequence is observed indicating whether a trade has transpired within each preceding nanosecond window. Comparing with volume processes which are noisier, as they may also oscillate due to unobserved trader-specific heterogeneity; and price processes which are often contaminated by microstructure noises, trading flows allow to be analyzed at a much higher frequency and are more closely related to information flows. That being said, as a compliment to the price movement, which contains consensual decisions and viewpoints of market participants, trading frequency also reflects the speed at which market participants react to and incorporate new information into their idiosyncratic trading strategies. As discussed in [Du and Zhu \(2017\)](#), a surge in trading intensity usually indicates higher level of information flow and potentially reflects real-time changes in market sentiment or news announcements that influence trading activity. \square

Aforementioned examples show the general state-space model (4.1) can be cast to model various market indicators such as high-frequency volumes, returns, and trading flows. In the following sections, we will construct functional estimators and associated inference procedure for conditional mean process and conditional quantile process of transformed states, and provide further practical implementation details of these examples.

measures. This result aligns with the asymptotic equivalence of statistical experiments in Le Cam's sense, see, e.g., [Le Cam \(1986a\)](#) and [Le Cam and Yang \(2000\)](#), whereas the statistical equivalence between estimating Poisson intensity with a Gaussian shift model is of more theoretical importance, see, e.g., [Grama and Nussbaum \(1998\)](#) and [Genon-Catalot et al. \(2002\)](#).

¹²Timestamps in TAQ database have evolved over time. For Consolidated Tape Association (CTA) trade and quote feeds, the accuracy of timestamps is milliseconds (10^{-3} seconds) since October 2003; microseconds (10^{-6} seconds) since August 3, 2015; nanoseconds since September 18, 2017.

4.2.3 Uniform Inference on Conditional Mean Process

Although our primary interest lies in unobservable states, we do not target on estimating the state process per se, we estimate instead some specific distributional features of transformed state process. Following [Li and Xiu \(2016\)](#) and [Bollerslev et al. \(2018\)](#), in this section, we focus on estimating the instantaneous conditional mean process g .¹³ Formally, we define

$$g_t \equiv \int_{\mathcal{D}} \mathcal{Y}(\zeta_t, \varepsilon) \mathbb{P}_\varepsilon(d\varepsilon), \quad \text{for } t \in [0, T].$$

Note that conditional mean processes may not always be well-defined, especially when the disturbance exhibits heavy tails. As a supplementary measure, we discuss estimation and inference of conditional quantile processes in the next section, which always exist. The precise implications of these processes, along with the identification procedure of state process ζ from them, intrinsically depend on specific properties of transformation $\mathcal{Y}(\cdot, \cdot)$ and the distribution \mathbb{P}_ε . These aspects should be analyzed on a meticulous case-by-case basis.

In preparation for a deep dive into the estimation procedure, we first introduce some additional notations concerning a block sampling scheme which is particularly useful in uniform inference for high-frequency data. This scheme divides the observation window into distinct, manageable blocks, facilitating the construction of local estimates, and paving the way for localized analysis. Formally, we divide the sample into m_n nonoverlapping blocks by partitioning the whole index set $\{1, \dots, n\} = \cup_{j=1}^{m_n} \mathcal{I}_{n,j}$, where $\mathcal{I}_{n,j}$ denote the set of $k_{n,j}$ consecutive indices contained in the j th block. Specifically, we define $\iota(i, j) \equiv \min \mathcal{I}_{n,j} + i - 1$ as the i th index in the j th block, and $\tau(i, j) \equiv \iota(i, j) \Delta_n$ as the associated time. In particular, we set $\tau(1, m_n + 1) \equiv T$. Consequently, we have $\mathcal{I}_{n,j} \equiv \{\iota(i, j) : 1 \leq i \leq k_{n,j}\}$, which spans time interval $\mathcal{T}_{n,j} \equiv [\tau(1, j), \tau(1, j + 1))$ for $1 \leq j \leq m_n$.

Given that g_t is simply the conditional mean of transformed state ζ_t , it naturally suggests forming an estimator by taking local average within the block which contains time t , while keeping block size shrinking. To fix ideas, we first consider conducting spot inference on g_t at some given time point t . Then there exists a

¹³In particular, inference regarding the integrated conditional mean process and spot conditional mean process has been studied in [Li and Xiu \(2016\)](#) and [Bollerslev et al. \(2018\)](#), respectively.

block j such that $t \in \mathcal{T}_{n,j}$, define \hat{g}_t as the local average of observations $Y_{i\Delta_n}$ over this block $\hat{g}_{n,t} \equiv k_{n,j}^{-1} \sum_{i \in \mathcal{I}_{n,j}} Y_{i\Delta_n}$. Theorem 1 in [Bollerslev et al. \(2018\)](#) shows that when $R_{n,i} = 0$, under fairly weak conditions on the local smoothness of ζ and bounded second conditional moments of $\mathcal{Y}(\cdot, \varepsilon)$, as $k_{n,j} \rightarrow \infty$ and $k_{n,j} \Delta_n \rightarrow 0$,

$$\sqrt{k_{n,j}}(\hat{g}_{n,t} - g_t) \xrightarrow{\mathcal{L}\text{-s}} \mathcal{MN}(0, \sigma_t^2), \quad (4.2)$$

where $\sigma_t^2 \equiv \int_{\mathcal{D}} \mathcal{Y}(\zeta_t, \varepsilon)^2 \mathbb{P}_\varepsilon(d\varepsilon) - \left(\int_{\mathcal{D}} \mathcal{Y}(\zeta_t, \varepsilon) \mathbb{P}_\varepsilon(d\varepsilon)\right)^2$ denotes conditional variance, $\xrightarrow{\mathcal{L}\text{-s}}$ denotes stable convergence in law, and \mathcal{MN} denotes mixed Gaussian distribution. The choice of block size corresponds to the trade-off between utilizing enough data to form an asymptotically Gaussian estimate and ensuring this estimate not to suffer from the bias due to local dynamics of state process. Consequently, noting that $\hat{\sigma}_{n,t}^2 \equiv k_{n,j}^{-1} \sum_{i \in \mathcal{I}_{n,j}} Y_{i\Delta_n}^2 - \left(k_{n,j}^{-1} \sum_{i \in \mathcal{I}_{n,j}} Y_{i\Delta_n}\right)^2$ is a consistent estimator of conditional variance σ_t^2 , we have the feasible central limit theorem

$$\frac{\sqrt{k_{n,j}}(\hat{g}_{n,t} - g_t)}{\hat{\sigma}_{n,t}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1).$$

Therefore, with $z_{1-\alpha/2}$ denoting the $(1 - \alpha/2)$ quantile of a standard Gaussian distribution, let

$$\widehat{C}_{n,t}^\pm(\alpha) \equiv \hat{g}_{n,t} \pm z_{1-\alpha/2} \times k_{n,j}^{-1/2} \hat{\sigma}_{n,t}, \quad (4.3)$$

then $\widehat{C}_{n,t}(\alpha) \equiv [\widehat{C}_{n,t}^-(\alpha), \widehat{C}_{n,t}^+(\alpha)]$ is an asymptotic $(1 - \alpha)$ confidence interval of g_t , i.e.,

$$\mathbb{P}(g_t \in \widehat{C}_{n,t}(\alpha)) \rightarrow 1 - \alpha, \quad \text{for every } t \in [0, T].$$

Above results can be easily extended to the case when $R_{n,i} \neq 0$ yet remains uniformly negligible, and furthermore, joint convergence of \hat{g}_n , on a *finite* set of time points $\{t_1, \dots, t_\ell\} \subset [0, T]$. By a classic Bonferroni approach, the hyperrectangle $C_{n,t_1}^\pm(\alpha/\ell) \times \dots \times C_{n,t_\ell}^\pm(\alpha/\ell)$ serves as a valid confidence set for vector $(g_{t_1}, \dots, g_{t_\ell})$. However, difficulty arises in extending this to estimation of the entire process g on a *continuum* set of indices, which is primarily due to the absence of functional central limit theorems. To better illustrate this limitation, we define blockwise estimator for the j th block similar as before

$$\hat{g}_{n,j} \equiv \frac{1}{k_{n,j}} \sum_{i \in \mathcal{I}_{n,j}} Y_{i\Delta_n} = \frac{1}{k_{n,j}} \sum_{i=1}^{k_{n,j}} Y_{\tau(i,j)}, \quad \text{for } 1 \leq j \leq m_n.$$

Given block size $k_{n,j}\Delta_n$ keeps shrinking, the block scheme becomes ever finer. Therefore, we can form a functional estimator for the entire process $(g_t)_{t \in [0, T]}$ as a collection of all blockwise estimates $(\hat{g}_{n,j})_{1 \leq j \leq m_n}$. Namely, we set

$$\hat{g}_{n,t} \equiv \hat{g}_{n,j}, \quad \text{for } t \in \mathcal{T}_{n,j} \text{ and } 1 \leq j \leq m_n.$$

Note that the blocks are non-overlapping, estimation errors within different blocks are asymptotically independent. Consequently, pointwise central limit theorem (4.2) shows that process of spot estimators have a path structure similar to a Gaussian white noise, hence is not asymptotically equicontinuous in probability on $[0, T]$ (see, e.g., Section 1.5 in [van der Vaart and Wellner \(1996\)](#)). The uniform inference problem based on this type of functional estimators is non-Donsker in nature. That being said, such non-Donsker problems that commonly arise from uniform inference in nonparametric settings, can be addressed using strong approximation of the functional estimators by variables with known finite-sample distributions, see, e.g., [Chernozhukov et al. \(2013\)](#) for the independent data and [Li and Liao \(2020\)](#) for generalization to time series data.¹⁴ To help fix ideas, we define the *sup-t statistic* as

$$\hat{T}_n^* \equiv \sup_{t \in [0, T]} |\hat{T}_{n,t}|, \quad \text{where } \hat{T}_{n,t} \equiv \frac{\sqrt{k_{n,j}}(\hat{g}_{n,t} - g_t)}{\hat{\sigma}_{n,t}} \text{ for } t \in \mathcal{T}_{n,j} \text{ and } 1 \leq j \leq m_n,$$

where $\hat{\sigma}_{n,t} \equiv \hat{\sigma}_{n,j}$ for $t \in \mathcal{T}_{n,j}$ and $1 \leq j \leq m_n$, and $\hat{\sigma}_{n,j}^2 \equiv k_{n,j}^{-1} \sum_{i \in \mathcal{I}_{n,j}} Y_{i\Delta_n}^2 - (k_{n,j}^{-1} \sum_{i \in \mathcal{I}_{n,j}} Y_{i\Delta_n})^2$. Theorem 4.1 below, shows the sup-t statistic can be strongly approximated, or coupled, by maximum of a growing dimensional folded Gaussian variables, whose distribution is well-understood in finite sample. First, we introduce some regularity conditions.

Assumption 3. *The observation process $(Y_{i\Delta_n})_{1 \leq i \leq n}$ is given by (4.1). There exist a sequence $(T_m)_{m \geq 1}$ of stopping times increasing to infinity, a sequence of compact subsets $(\mathcal{K}_m)_{m \geq 1}$ of \mathcal{Z} , and a sequence $(K_m)_{m \geq 1}$ of positive constants such that for each $m \geq 1$ such that:*

(i) $\zeta_{t \wedge T_m}$ takes value in \mathcal{K}_m ; for all $s, t \in \mathcal{T}_{n,j}$ where $1 \leq j \leq m_n$, and for each $p > 0$, $\mathbb{E}[\|\zeta_{t \wedge T_m} - \zeta_{s \wedge T_m}\|^p] \leq K_{m,p}|t - s|^{p/2}$ for some constant $K_{m,p}$;

(ii) for all $z, z' \in \mathcal{K}_m$ with $z \neq z'$, $\text{Var}(\mathcal{Y}(z, \varepsilon))^{-1} + \|\mathcal{Y}(z, \varepsilon) - \mathcal{Y}(z', \varepsilon)\|_{L_2} / \|z -$

¹⁴A Yurinskii-type coupling for the entire t -statistic process does not hold in general case, unless the state process is very smooth or the transformations take special forms (e.g., [Jacod et al. \(2021\)](#)).

$z' \| \leq K_m$;

(iii) for all $x > 0$ and $z \in \mathcal{K}_m$, $\mathbb{P}_\varepsilon(|\mathcal{Y}(z, \varepsilon)| \geq x) \leq K_m \exp\{-(x/K_m)^{1/\eta}\}$ for some $\eta > 0$;

(iv) $\max_{1 \leq i \leq n} |R_{n,i}| = o_p(\Delta_n^r)$ for some $r > 0$.

Assumption 3 imposes some regularity conditions on the state process, the transformation of random disturbance, and the residual term, which allow for essentially unrestricted nonstationary state process and heavy-tailed disturbance. Specifically, condition (i) requires state process to be locally taken value in compact set. Condition (i) also imposes the smoothness of state process *within* each block. Namely, it requires state process to be 1/2-Hölder continuous under the L_p -norm for any positive p . This condition is stronger than that needed for conducting pointwise inference, see [Bollerslev et al. \(2018\)](#). It holds if the state process is a continuous Itô semimartingale or long-memory process within each block, and it also allows state process to have jumps on the boundary time points between blocks. Condition (ii) requires the variance of $\mathcal{Y}(z, \varepsilon)$ to be locally bounded away from zero, and the random mapping $z \mapsto \mathcal{Y}(z, \varepsilon)$ to be Lipschitz on compact set \mathcal{K}_m under the L_2 norm, which is a minor restriction and can be easily verified for aforementioned examples. Condition (iii) requires transformed disturbance to have a sub-Weibull tail with parameter $\eta > 0$, which is a generalization of sub-Gaussian and sub-Exponential families to potentially heavier-tailed distributions including Exponential distribution and Poisson distribution, see [Vladimirova et al. \(2020\)](#) and [Kuchibhotla and Chakraborty \(2022\)](#) for a detailed discussion of sub-Weibull tails. This condition holds for any $\eta \geq 1/2$ (resp. $\eta \geq 1$) if $\mathcal{Y}(z, \varepsilon)$ has sub-Gaussian (resp. sub-Exponential) tail, and can be verified even for the disturbance arises from machine learning models, see [Hayou et al. \(2019\)](#) for a proof under deep neural networks. We highlight that condition (iii) also ensures the existence of conditional mean process. Condition (iv) is a high-level condition which requires residual term to be uniformly negligible in the sense that it shrinks at a polynomial rate uniformly for all $1 \leq i \leq n$.

Before state the strong approximation result of sup- t statistic, we provide some additional implementation details by revisiting three examples outlined in the preceding section. Discussion of implementation details primarily aims to shed light

on the interplay between conditional mean process and state process, together with a validation of Assumption 3 (especially condition iv), under those specific models.

EXAMPLE 1 (LOCATION-SCALE MODEL, CONTINUED). In the simple location-scale model with additive structure, suppose that disturbance is centered. Then by definition, the conditional mean process inherently translates into local mean process, i.e., $g_t = \mu_t$ for all $t \in [0, T]$. Consequently, the first state process μ can be directly identified from g . Assumption 3(i) is satisfied if $(\mu_t, \sigma_t)_{t \in \mathcal{T}_{n,j}}$ is a two dimensional continuous Itô semimartingale or long-memory process within each block. Suppose in addition that \mathbb{P}_ε has a sub-Weibull tail, Assumption 3(iii) is met. This, combined with σ maintaining bounded away from zero, leads to the fulfillment of Assumption 3(ii). Recall that in this example residual terms $R_{n,i} = 0$ for all $1 \leq i \leq n$, Assumption 3(iv) trivially holds for any $r > 0$. \square

EXAMPLE 2 (LÉVY-DRIVEN ASSET RETURNS, CONTINUED). Recall the characteristic triple of stable Lévy process described in footnote 8, conditional mean process is coherently well-defined only when $\beta = 2$, i.e., L is a Brownian motion.¹⁵ Therefore, subsequent discussion in this section is confined to the case where $\beta = 2$, scenarios regarding $\beta \in (0, 2)$ will be addressed in Section 4.2.4. Assumption 3(i) is satisfied if the volatility $(\sigma_t)_{t \in \mathcal{T}_{n,j}}$ is a continuous Itô semimartingale or long-memory process within each block, which is congruent with most popular stochastic volatility models. Note that in this example, the disturbance is a sequence of i.i.d. standard Gaussian variables, indicating the transformed disturbance $(\sigma\varepsilon)^2$ follows a scaled $\chi^2(1)$ distribution. As a result, the conditional mean process translates into variance process, i.e., $g_t = \sigma_t^2$ for all $t \in [0, T]$. Also, Assumption 3(iii) holds for any $\eta \geq 1$, Assumption 3(ii) is satisfied provided that volatility is bounded away from zero. Suppose in addition that the drift process μ is locally bounded, by a combined use of the Burkholder–Davis–Gundy inequality, the Hölder inequality, and a maximal inequality, we can deduce for all $p \geq 1$,

$$\mathbb{E} \left[\max_{1 \leq i \leq n} |R_{n,i}|^p \right] \leq \Delta_n^{-1} \mathbb{E} \left[\sup_{|t-s| \leq \Delta_n} |\sigma_t - \sigma_s|^{2p} \right] \leq K_p \Delta_n^{p-1},$$

confirming that Assumption 3(iv) holds for any $0 < r < 1$. \square

¹⁵The instantaneous conditional mean diverges at a rate of $\Delta_n^{2/\beta-1}$ by the definition of Blumenthal–Gatoor index.

EXAMPLE 3 (COX TRADING FLOWS, CONTINUED). In the context of Cox trading flow model, recall that state process is $\Delta_n \mu_t$. Assumption 3(i) and (ii) hence requires the scaled intensity $\Delta_n \mu_t$ to be 1/2-Hölder continuous within each block, and more critically, to be both bounded from above and away from zero,¹⁶ which alludes to the “high traffic” assumption, as introduced in Kingman (1961). As a complement elaboration, Christensen and Kolokolov (2023) provides an alternative justification for this assumption by modeling trading flow as a sum of n independent copies of Cox processes with conditional intensity $\Delta_n \mu_t$. This “heavy traffic” assumption is a natural precursor for econometric analysis of high-frequency financial data, in the sense that a Cox process endowed with “high traffic” intensity can generate the class of valid stochastic sampling schemes studied in Hayashi et al. (2011).¹⁷ Note that the transformation takes binary values, Assumption 3(iii) is automatically satisfied for any $\eta > 0$. For the residual term, recall $\mathbb{P}(R_{n,i} \neq 0 | \mu) = o(\Delta_n)$, by the law of iterated expectation we have for any $r > 0$,

$$\mathbb{P}\left(\max_{1 \leq i \leq n} |R_{n,i}| > \Delta_n^r\right) \leq \sum_{i=1}^n \mathbb{P}(R_{n,i} \neq 0) = n o(\Delta_n) = o(1),$$

confirming that Assumption 3(iv) also holds for any $r > 0$. \square

We are now ready to formally state our strong approximation result for sup- t statistics.

Theorem 4.1. *Suppose that (i) Assumption 3 is satisfied; (ii) $k_{n,j} \asymp \Delta_n^{-\rho}$ uniformly for all $1 \leq j \leq m_n$ such that $\rho \in (0, 2r \wedge 1/2)$. Let $(Z_1, Z_2, \dots, Z_{m_n})^\top$ be a standard Gaussian random vector in \mathbb{R}^{m_n} . Then for some positive constant ϵ ,*

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P}(\widehat{T}_n^* \leq x) - \mathbb{P}\left(\max_{1 \leq j \leq m_n} |Z_j| \leq x\right) \right| \leq K \Delta_n^\epsilon.$$

COMMENT 1. Theorem 4.1 shows the sup- t statistic can be strongly approximated by maximum of a increasing dimensional folded standard Gaussian random variables, in the sense that their Kolmogorov–Smirnov distance shrinks to zero at a polynomial rate. A similar result holds under the Kantorovich–Monge–Rubinstein met-

¹⁶This is not surprising since the intensity of a Poisson process is not consistently estimable over a fixed time window, not even in the homogeneous case (see, e.g., Brillinger (1975), Karr (1991), and Helmers and Zitikis (1999)).

¹⁷We assume additionally the trading flow process to have refractoriness, see, e.g., Citi et al. (2014).

ric.¹⁸ In that case, there exist sequences on a common probability space $\widehat{T}'_n \stackrel{\mathcal{L}}{=} \widehat{T}_n^*$ and $Z'_n \stackrel{\mathcal{L}}{=} \max_{1 \leq j \leq m_n} |Z_j|$ such that $\widehat{T}'_n = Z'_n + o_p(1)$. However, here it is not straightforward that convergence under the Kantorovich–Monge–Rubinstein metric implies convergence under the Kolmogorov–Smirnov metric, since the density of Z'_n is unbounded.¹⁹ Consequently, due to the particular usefulness in making inference, Theorem 4.1 and other strong approximation results in this paper, are presented under the Kolmogorov–Smirnov distance.

COMMENT 2. We emphasize that distribution of coupling variable $\max_{1 \leq j \leq m_n} |Z_j|$ is known in finite sample, which renders Theorem 4.1 particularly useful for inferential purposes. Formally, given any $\alpha \in (0, 1/2)$, let $cv_n(\alpha) \equiv \inf\{x \in \mathbb{R} : \mathbb{P}(\max_{1 \leq j \leq m_n} |Z_j| \leq x) \geq 1 - \alpha\}$ denote the $(1 - \alpha)$ quantile of $\max_{1 \leq j \leq m_n} |Z_j|$, which can be easily computed for any m_n .²⁰ Then Theorem 4.1 implies $|\mathbb{P}(\widehat{T}_n^* \leq cv_n(\alpha)) - \mathbb{P}(\max_{1 \leq j \leq m_n} |Z_j| \leq cv_n(\alpha))| \leq K \Delta_n^\epsilon$. Consequently, let

$$\widehat{B}_{n,t}^\pm(\alpha) \equiv \widehat{g}_{n,t} \pm cv_n(\alpha) \times k_{n,j}^{-1/2} \widehat{\sigma}_{n,t}, \quad \text{for all } t \in \mathcal{T}_{n,j}, \text{ and } 1 \leq j \leq m_n, \quad (4.4)$$

then $\widehat{B}_{n,t}(\alpha) \equiv [\widehat{B}_{n,t}^-(\alpha), \widehat{B}_{n,t}^+(\alpha)]$ constitutes an asymptotic $(1 - \alpha)$ confidence band for the entire process $(g_t)_{t \in [0, T]}$, i.e.,

$$\mathbb{P}(g_t \in \widehat{B}_{n,t}(\alpha) \text{ for all } t \in [0, T]) = \mathbb{P}(\widehat{T}_n^* \leq cv_n(\alpha)) \rightarrow 1 - \alpha.$$

Observing that the uniform confidence band (4.4) is generally wider than pointwise confidence intervals (4.3), this difference magnifies as the number of blocks m_n becomes larger. To better illustrate the intuition behind this difference, we present a simple comparative visualization for uniform confidence bands and pointwise confidence intervals under different numbers of blocks in Figure 4.1. Given that total number of observations is typically fixed in application, the number of blocks is intrinsically determined by the block size. Consequently, m_n stands inversely pro-

¹⁸The Kantorovich–Monge–Rubinstein metric between two measures \mathbb{P}_1 and \mathbb{P}_2 is defined as $\sup\{|\int f d\mathbb{P}_1 - \int f d\mathbb{P}_2| : \|f\|_{\text{Lip}} \leq 1\}$, Theorem 2 in Szulga (1983) shows it is equivalent to the Wasserstein 1-metric $\inf\{\mathbb{E}\|X - Y\| : \mathcal{L}(X) = \mathbb{P}_1, \mathcal{L}(Y) = \mathbb{P}_2\}$.

¹⁹The density of $\max_{1 \leq j \leq m_n} |Z_j|$ is given by $f(x) \equiv 2m_n(2\Phi(x) - 1)^{m_n-1} \phi(x) \mathbb{1}\{x \geq 0\}$, where $\phi(\cdot)$ and $\Phi(\cdot)$ denote the density and distribution functions of standard Gaussian distribution, respectively. Note that the Mills ratio $(1 - \Phi(x))/\phi(x) \rightarrow 1/x$, by verifying a Von Mises type condition and applying Corollary 1.7 in Resnick (2008), we can show $f(x) \simeq 4\sqrt{\log m_n}/e$ as $x \rightarrow \sqrt{2 \log m_n} + (2 \log 2 - \log \log m_n - \log(4\pi))/\sqrt{8 \log m_n}$ and $m_n \rightarrow \infty$.

²⁰For instance, use one-line command `fsolve(@(x)(2*normcdf(x)-1).^m-(1-alpha), log(m))` in MATLAB.

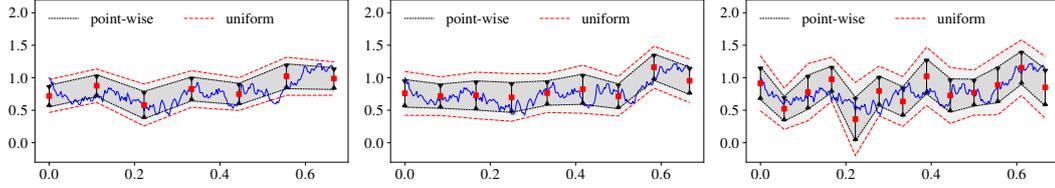


Figure 4.1: Comparison of Confidence Bands under Different Numbers of Blocks. In each panel, we mark spot estimates in red squares, 90% pointwise confidence interval in black vertical segments, 90% uniform confidence band in red dashed lines, and the true process in blue lines. The pointwise confidence band is constructed by connecting each confidence interval computed using (4.3), the uniform confidence band is computed using (4.4). Three panels from left to right show results for the case where m_n equals 6, 8, and 12, respectively, corresponding to the tuning sequence k_n being 40, 30, and 20.

portional to k_n . When the number of blocks is small, each block becomes wide, leading to a large time variation effect which undermines the coverage of pointwise confidence interval. In contrast, when the number of blocks is large, probability of committing type I error across distinct blocks accumulates. Such accumulating errors are not accommodated for in pointwise confidence intervals.

4.2.4 Uniform Inference on Conditional Quantile Process

As we mentioned in the previous section, if the disturbance exhibits exceedingly heavy tails, instantaneous conditional mean process is not well-defined. This section pivots to explore an alternative method of analyzing these heavy-tailed models, centering on instantaneous conditional quantile of the transformed state as a supplemental measure. In contrast to conditional mean process, the conditional quantile process remains well-defined, regardless of the nature of \mathbb{P}_ε .²¹ To be precise, for some pre-determined level $\chi \in (0, 1)$, we define the conditional quantile process as a version of càdlàg inverse of conditional distribution function of $\mathcal{Y}(\zeta_t, x)$, i.e.,

$$q_t(\chi) \equiv \inf\{x \in \mathbb{R} : \mathbb{P}_\varepsilon(\mathcal{Y}(\zeta_t, \varepsilon) \leq x) \geq \chi\}, \quad \text{for } t \in [0, T].$$

The analysis of quantile has developed rapidly since the foundational [Koenker and Bassett Jr \(1978\)](#). It has been highlighted that quantile is the unique solution of minimizing expected loss utilizing the check function $u_\chi(y) \equiv y(\chi - \mathbb{1}\{y < 0\})$. Based on this insight, it is natural to define an estimator through the sample analogue, which also offers a heuristic method of deriving asymptotic behaviors through the

²¹Sample quantiles has other applications, see, e.g., [Coeurjolly \(2008\)](#) for estimating the Hurst parameter of fractional Brownian motion using a convex combination of sample quantiles.

monotonicity of first order conditions, see, e.g., Section 3.2 in [Koenker \(2005\)](#). Alternatively, although essentially equivalent in most cases, some statisticians opt to define quantile estimators directly through its corresponding order statistics. Here, its asymptotic properties and optimalities are extensively explored via the elegant Bahadur representation. In the pioneered paper, [Bahadur \(1966\)](#) first established almost sure bound of representing the difference between population quantile and corresponding order statistics as a sample average of some i.i.d. auxiliary variables. [Ghosh \(1971\)](#) provided a simple proof for a weaker but sufficiently useful bound. The result has been extended to nonparametric quantile regression by [Chaudhuri \(1991\)](#), and to weakly dependent stationary data by [Hesse \(1990\)](#) and [Wu \(2005\)](#).

We adopt the idea from classic statistic methodology to define each spot estimator as local “in-fill order statistic” of observations inside the shrinking block, instead of through the convention of minimization problem. Namely, within each block, we reindex the sequence $(Y_{i\Delta_n})_{i \in \mathcal{I}_{n,j}}$ in the non-decreasing order and denoted as $Y_{1,j}^o \leq \dots \leq Y_{k_{n,j},j}^o$. The spot estimator for conditional quantile, in this scheme, is defined as $[k_{n,j}\chi]$ -order statistic.²² Analogous to the previous section, we form a functional estimator as the collection of all blockwise estimates

$$\hat{q}_{n,j}(\chi) \equiv Y_{[k_{n,j}\chi],j}^o, \quad \hat{q}_{n,t}(\chi) \equiv \hat{q}_{n,j}(\chi) \text{ for } t \in \mathcal{T}_{n,j} \text{ and } 1 \leq j \leq m_n.$$

Although observations from model (4.1) are neither independent nor stationary, in the appendix we show that a uniform Bahadur representation holds for all blockwise in-fill χ -sample quantiles given some regularity conditions (Lemma C.1), which forms the bedrock for deriving strong approximation results for the functional conditional quantile process estimator. To the best of our knowledge, this is the first paper to consider uniform (over time) inference of quantile process under in-fill setting. We first introduce some regularity conditions.

Assumption 4. *The observation process $(Y_{i\Delta_n})_{1 \leq i \leq n}$ is given by (4.1). There exists a sequence $(T_m)_{m \geq 1}$ of stopping times increasing to infinity, a sequence of compact subsets $(\mathcal{K}_m)_{m \geq 1}$ of \mathcal{Z} , and a sequence $(K_m)_{m \geq 1}$ of positive constants such that:*

- (i) $\xi_{t \wedge T_m}$ takes value in \mathcal{K}_m ; for all $s, t \in \mathcal{T}_{n,j}$ where $1 \leq j \leq m_n$, and for each

²²Note that the results presented in this section hold for all $\ell_{n,j}$ -order statistics with $\ell_{n,j} - k_{n,j}\chi = o(k_{n,j}^{1/2} \log k_{n,j})$. We focus on $[k_{n,j}\chi]$ -order statistic to avoid unnecessary complexity.

$p > 0$, $\mathbb{E}[\|\zeta_{t \wedge T_m} - \zeta_{s \wedge T_m}\|^p] \leq K_{m,p}|t - s|^{p/2}$ for some constant $K_{m,p}$;

(ii) for each $x \in \mathbb{R}$, for all $z, z' \in \mathcal{K}_m$, $|F(z, x) - F(z', x)| \vee |\partial_x F(z, x) - \partial_x F(z', x)| \leq K_m \|z - z'\|$ where $F(\cdot, x) \equiv \mathbb{P}_\varepsilon(\mathcal{Y}(\cdot, \varepsilon) \leq x)$;

(iii) for each $t \in [0, T_m]$ and x in some neighborhood of $q_t(\chi)$, $f_t(x) + f_t(x)^{-1} + |\partial_x f_t(x)| < K_m$ where $f_t(\cdot) \equiv \partial_{(\cdot)} F(\zeta_t, \cdot)$;

(iv) $\max_{1 \leq i \leq n} |R_{n,i}| = o_p(\Delta_n^r)$ for some $r > 0$.

Condition (i) remains the same as in Assumption 3, i.e., it requires state process to be locally taken value in compact set and 1/2-Hölder continuous under the L_p -norm for any positive p . Likewise, it is satisfied if the state process is a continuous Itô semimartingale or long-memory process within each block and does not exclude jumps on the boundary time points between blocks. Condition (ii) necessitates that, for a given value of x , the function $F(\cdot, x)$ and its derivative $\partial_x F(\cdot, x)$ to be Lipschitz over the set \mathcal{K}_m . This condition can be verified if $F(\cdot, \cdot) \in C^{2,1}(\mathcal{K}_m, \mathbb{R})$. Condition (iii) is a local requirement that conditional density function at true state ζ_t evaluated at a neighborhood of the quantile is positive and not too concentrate around that point, which holds if $f_t(\cdot)$ is continuous and has no point mass.²³ Condition (iv) is the same high-level requirement as in Assumption 3, which requires residual terms to shrink uniformly at a polynomial rate.

EXAMPLE 2 (LÉVY-DRIVEN ASSET RETURNS, CONTINUED). Recent advances in high-frequency financial data analysis have accentuated the significance of inference using sample order statistics.²⁴ Specifically, in a special case when $\beta = 2$ and choosing $\chi = 1/2$, Shephard (2022) consider estimating integrated volatility over $[0, T]$ through the normalized sum of “in-fill median” in each block. Asymptotic properties of this estimator are derived via the monotonicity of first order condition of minimization problems in the spirit of Koenker and Bassett Jr (1978). Although integrated volatility estimators constructed using median are asymptotically less efficient than realized variance in the Brownian motion case, it remains robust to ab-

²³Observe that this requirement excludes the case where random disturbances are discretely distributed. This is not surprising since even the classic Bahadur representation for i.i.d. data requires absolute continuity of the distribution. Analysis of sample quantiles for discretely distributed data deserves its own research.

²⁴The use of extreme order statistics, although beyond the scope of this paper as we assume $\chi \in (0, 1)$, has been utilized in estimating volatility even earlier, see, e.g., Garman and Klass (1980), Parkinson (1980).

normal returns which often arise when the price contains jumps. As a complement to Shephard (2022), in this example, our focus is on uniform inference for the entire volatility process even in the case when $\beta < 2$, a setting wherein conditional mean process becomes not well-defined and Assumption 3(iii) no longer holds. Consequently, return-based estimation procedure becomes invalid. Nevertheless, recall the state-space formation of Lévy-driven returns, it is evident that for all $t \in [0, T]$,

$$q_t(\chi) = \sigma_t^2 Q(L, \chi),$$

where $Q(L, \chi)$ denote the χ -quantile of $\varepsilon_{n,i} = \Delta_n^{-2/\beta} (L_{(i+1)\Delta_n} - L_{i\Delta_n})^2$, hence is free of nuisance. This proportional structure between $q(\chi)$ and σ suggests that conditional quantile process can serve as a feasible proxy for volatility. Note that formally defining the volatility process in a heuristic way via quadratic variation of continuous part is impossible in this case,²⁵ whereas interquantile range effectively captures the volatile level of price. Although for the cases $\beta \neq 1$, closed-form densities of $\varepsilon_{n,i}$ is almost never known, we do have explicit closed-form characteristic functions. This facilitates the numerical computation of $Q(L, \chi)$ and validation of condition (ii) and (iii) in Assumption 4, see, e.g., Zolotarev (1986).²⁶ Moreover, noting that $[L]_t = \sum_{s \leq t} |\Delta L_s|^2 < \infty$ almost surely for any $t > 0$, a similar argument as in the previous section yields that condition (iv) remains valid for all $0 < r < 1$. \square

Analogous to Theorem 4.1, we present Theorem 4.2 below, which states the strong approximation result for our functional quantile estimator using the Kolmogorov–Smirnov metric.

Theorem 4.2. *Suppose that (i) Assumption 4 is satisfied; (ii) $k_{n,j} \asymp \Delta_n^{-\rho}$ uniformly for all $1 \leq j \leq m_n$ such that $\rho \in (0, 2r \wedge 1/2)$. Let $(Z_1, Z_2, \dots, Z_{m_n})^\top \sim \mathcal{MN}(0, \text{diag}\{v_1^2, \dots, v_{m_n}^2\})$ be a mixed Gaussian random vector in \mathbb{R}^{m_n} such that $v_j^2 \equiv \chi(1 - \chi) / f_{\tau(1,j)}(q_{\tau(1,j)}(\chi))^2$. Then for any $\chi \in (0, 1)$, for some positive con-*

²⁵Namely, the quadratic variation of continuous part of P is zero when $\beta < 2$.

²⁶Note that semi-closed-form expressions of densities of stable distributions are available, for example in the form of an one-dimensional integral or a convergent infinite series. Various numerical computation procedures and associated error bounds are discussed in Ament and O’Neil (2018).

start ϵ ,

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P} \left(\max_{1 \leq j \leq m_n} \sup_{t \in \mathcal{T}_{n,j}} \sqrt{k_{n,j}} |\hat{q}_{n,t}(\chi) - q_t(\chi)| \leq x \right) - \mathbb{P} \left(\max_{1 \leq j \leq m_n} |Z_j| \leq x \right) \right| \leq K \Delta_n^\epsilon.$$

COMMENT. In contrast to Theorem 4.1, the coupling variable $\max_{1 \leq j \leq m_n} |Z_j|$ here is not pivotal as the variance matrix remains unknown, which is not surprising in quantile-related inference. This problem can be addressed, since the density function $f_t(\cdot)$ is nonparametrically estimable. Alternatively, a practically more convenient choice is to employ the bootstrap method to get an asymptotically valid critical value, as justified by Zuo (2015) who derived a Bahadur representation for empirical bootstrap quantiles. We stress that in certain scenarios, the distribution can indeed be pivotalized. For instance the multiplicative transformation (see Example 2) where conditional quantile estimation is extremely useful, we have for all $t \in [0, T]$ that

$$f_t(q_t(\chi)) = \frac{1}{\sigma_t^2} \bar{f} \left(\frac{q_t(\chi)}{\sigma_t} \right) = \frac{Q(L, \chi)^2 \bar{f}(Q(L, \chi))}{q_t(\chi)^2},$$

where $\bar{f}(\cdot)$ denotes the density of $\Delta_n^{-2/\beta} (L_{(i+1)\Delta_n} - L_{i\Delta_n})^2$ which is free of nuisance hence can be computed numerically. Let

$$\hat{v}_{n,j}^2 \equiv \frac{\chi(1-\chi)Q(L, \chi)^2 \bar{f}(Q(L, \chi))^2}{\hat{q}_{n,j}(\chi)^2}, \quad \text{for all } 1 \leq j \leq m_n.$$

Given that $\bar{f}(\cdot)$ is Lipschitz in the neighborhood of $Q(L, \chi)$ by Assumption 6(iii), Theorem 4.2 then implies that

$$\max_{1 \leq j \leq m_n} |\hat{v}_{n,j}^2 - v_j^2| = O_p(\Delta_n^{\rho/2} \log(\Delta_n^{-1})^{1/2}).$$

Consequently, let $cv_n(\alpha)$ be defined identically as in (4.4), denote

$$\widehat{B}_{n,t}^{\pm}(\alpha) \equiv (\hat{q}_{n,t}(\chi) \pm cv_n(\alpha) \times k_{n,j}^{-1/2} \hat{v}_{n,j}) / Q(L, \chi), \quad \text{for all } t \in \mathcal{T}_{n,j}, \text{ and } 1 \leq j \leq m_n, \quad (4.5)$$

Then $\widehat{B}'_{n,t}(\alpha) \equiv [\widehat{B}_{n,t}^-(\alpha), \widehat{B}_{n,t}^+(\alpha)]$ constitutes an asymptotic $(1 - \alpha)$ confidence band for the entire variance process $(\sigma_t^2)_{t \in [0, T]}$, i.e.,

$$\mathbb{P}(\sigma_t^2 \in \widehat{B}'_{n,t}(\alpha) \text{ for all } t \in [0, T]) \rightarrow 1 - \alpha.$$

4.2.5 Application: Inference for Ranks

The strong approximation results established in this paper can be used to tackle other econometric problems. As a byproduct, we discuss the problem of doing inference for *ranks* in this section. Namely, given a path of certain stochastic process, rankings of the values at a set of time points are often of great interest. Notably, such interest stems when the process indicates some time-varying signals, while quantifying these signals is challenging hence we are interested instead in their relative magnitudes. These rankings illuminate which segments of the process possess comparatively higher signal level in relation to others. For instance, vigors of trading intensities can shed light on the real-time information level that affects the market (see, e.g., [Du and Zhu \(2017\)](#)).

Usually, the realized path is unobservable. Thus, rankings are invariably deduced using functional estimators instead of the true process. Such procedure inevitably introduces uncertainties, necessitating careful considerations before drawing definitive conclusions regarding rankings of the true process. To illustrate this inherent uncertainty, consider a simple example where $\sqrt{k_n}(\hat{g}_{t_i} - g_{t_i}) \sim \mathcal{N}(0, 1)$ for $i \in \{1, 2\}$, then we have $\mathbb{P}(\hat{g}_{t_1} > \hat{g}_{t_2} \mid g_{t_1} < g_{t_2}) = 1 - \Phi(\sqrt{k_n}(g_{t_2} - g_{t_1})/2)$, i.e., in finite samples, there is a nonzero probability that estimated rankings do not coincide with their true rankings. While the probability of such misranking tends to zero with a increasing number of observations, it conversely accumulates with a increasing number of candidates under comparison.

In a recent paper, [Mogstad et al. \(2023\)](#) provided a comprehensive framework for inferring ranks via the introduction of confidence sets for ranks. This methodology is congruent with the problem at hand. Given a designated set of inspected time points, observe that the length of blocks shrinks to zero. Consequently, as Δ_n becoming small enough, each time point in that set falls exactly in one distinct block. Therefore, we may assume without loss of generality that the set of inspected time points takes the form of $\{t_1, \dots, t_{m_n}\}$ where $t_j \in \mathcal{T}_{n,j}$ for all $1 \leq j \leq m_n$. To give a detailed illustration, we focus on the case investigating conditional mean process $(g_t)_{t \in [0, T]}$. Analogous results can be formulated for conditional quantile process via uniform Bahadur representation and [Theorem 4.2](#). To avoid double subscripts,

with a slight abuse of notation, we denote $g_{n,j} \equiv g_{t_j}$ for $1 \leq j \leq m_n$. Following [Mogstad et al. \(2023\)](#), we define ranks of $(g_{n,j})_{1 \leq j \leq m_n}$ and the entire rank vector as

$$\text{Rank}_n(j) \equiv 1 + \sum_{j'=1}^{m_n} \mathbb{1}\{g_{n,j'} > g_{n,j}\} \quad \text{and} \quad \text{Rank}_n \equiv (\text{Rank}_n(1), \dots, \text{Rank}_n(m_n))^\top.$$

Then a joint $(1 - \alpha)$ confidence set for ranks at all time points is defined as a random set $\widehat{\text{Rank}}_n \subset \mathbb{R}^{m_n}$ such that

$$\liminf_{\Delta_n \rightarrow 0} \mathbb{P}(\text{Rank}_n \in \widehat{\text{Rank}}_n) \geq 1 - \alpha.$$

Let $\mathcal{S}_n^{\text{all}} \equiv \{(j, j') : 1 \leq j, j' \leq m_n \text{ and } j \neq j'\}$ denote the set of all paired indices. Based on the insight of Theorem 3.4 in [Mogstad et al. \(2023\)](#), the confidence level of a joint confidence set for all ranks is bounded below by one minus the *familywise error rate*, denoted as FWER_n , for testing following family of multiple one-sided hypotheses

$$H_{j,j'} : g_{n,j} \leq g_{n,j'} \quad \text{against} \quad K_{j,j'} : g_{n,j} > g_{n,j'}, \quad \text{where } (j, j') \in \mathcal{S}_n^{\text{all}}. \quad (4.6)$$

According to which null hypotheses hold true, we can partition all paired indices into two subsets $\mathcal{S}_n^{\text{all},-} \equiv \{(j, j') \in \mathcal{S}_n^{\text{all}} : g_{n,j} \leq g_{n,j'}\}$, $\mathcal{S}_n^{\text{all},+} \equiv \{(j, j') \in \mathcal{S}_n^{\text{all}} : g_{n,j} \geq g_{n,j'}\}$. We also denote the set of rejected hypotheses as $\text{Rej}_n^-(j) \equiv \{(j, j') \in \mathcal{S}_n^{\text{all}} : H_{j',j} \text{ is rejected}\}$ and $\text{Rej}_n^+(j) \equiv \{(j, j') \in \mathcal{S}_n^{\text{all}} : H_{j,j'} \text{ is rejected}\}$. Moreover, define $\text{Rej}_n^\pm \equiv \bigcup_{j=1}^{m_n} \text{Rej}_n^\pm(j)$. Then the familywise error rate for testing family (4.6) can be formally expressed as

$$\begin{aligned} \text{FWER}_n &\equiv \mathbb{P}(\text{reject at least one true hypothesis } H_{j,j'}) \\ &= \mathbb{P}(\mathcal{S}_n^{\text{all},-} \cap \text{Rej}_n^+ \neq \emptyset \text{ or } \mathcal{S}_n^{\text{all},+} \cap \text{Rej}_n^- \neq \emptyset). \end{aligned}$$

Our goal is to find a valid test such that $\limsup_{\Delta_n \rightarrow 0} \mathbb{P}(\text{FWER}_n) \leq \alpha$. We will describe the detailed testing procedure in the rest of this section. Before presenting the procedure, we highlight that our setting here differs from that of [Mogstad et al. \(2023\)](#) in two aspects. Firstly, note that [Mogstad et al. \(2023\)](#) focus on the rankings across different populations, which implies their rankings are deterministic. On the contrary, we consider ranks that defined for a single realized path of the investigated process at different time points. Consequently, rankings Rank_n hence the partition $\mathcal{S}_n^{\text{all},\pm}$ are both random in nature. Secondly, we allow the number of evaluated time

points to diverge as $\Delta_n \rightarrow 0$ at a rate identical to number of blocks m_n , contrasting with the case in [Mogstad et al. \(2023\)](#) where the total number of populations remains fixed.

For the sake of notational simplicity, we assume for the moment that $k_{n,j} = k_n$ for $1 \leq j \leq m_n$, i.e., we partition observations into blocks with equal length. For each elementary null hypothesis $H_{j,j'}$ where $(j, j') \in \mathcal{S}_n^{\text{all}}$, we construct tests statistic concerning the difference $\hat{g}_{n,j} - \hat{g}_{n,j'}$. Denote the corresponding variance estimator as $\hat{\zeta}_n(j, j')^2 \equiv \hat{\sigma}_{n,j}^2 + \hat{\sigma}_{n,j'}^2$. Then we reject $H_{j,j'}$ whenever the associated t -statistic

$$\hat{d}_n(j, j') \equiv \frac{\sqrt{k_n}(\hat{g}_{n,j} - \hat{g}_{n,j'})}{\hat{\zeta}_n(j, j')},$$

is sufficiently large, say, exceeds some carefully selected threshold. To determine the proper value of critical value that controls FWER_n , we define the sup- t statistics as $\hat{D}_n \equiv \max_{(j,j') \in \mathcal{S}_n^{\text{all}}} \hat{d}_n(j, j')$.²⁷ A direct application of [Theorem 4.1](#) indicates a similar strong approximation result holding for \hat{D}_n . Nonetheless, additional difficulty arises since the distribution of coupling variable becomes more complicated. This stems from the fact that covariance matrix becomes non-identity since off-diagonal components can be non-zero given that $\mathcal{S}_n^{\text{all}}$ contains pairs with coinciding indices. In light of this, we propose an employment of a Gaussian multiplier bootstrap technique to determine the requisite confidence value. Namely, we generate i.i.d. standard Gaussian variables $(e_i)_{1 \leq i \leq k_n}$ independent of $(Y_{i\Delta_n})_{1 \leq i \leq n}$. Denote

$$\hat{g}_{n,j}^B \equiv \frac{1}{k_n} \sum_{i=1}^{k_n} e_i (Y_{\tau(i,j)} - \hat{g}_{n,j}).$$

Repeat this step to generate a large number of bootstrap sample of $(\hat{g}_{n,j}^B)_{1 \leq j \leq m_n}$. Then we can compute the conditional $(1 - \alpha)$ quantile of the maximum of studentized bootstrap statistics via

$$cv_n^B(\alpha, \mathcal{S}_n^{\text{all}}) \equiv \inf \left\{ x \in \mathbb{R} : \mathbb{P} \left(\max_{(j,j') \in \mathcal{S}_n^{\text{all}}} \frac{\sqrt{k_n}(\hat{g}_{n,j}^B - \hat{g}_{n,j'}^B)}{\hat{\zeta}_n(j, j')} \leq x \mid (Y_{i\Delta_n})_{1 \leq i \leq n} \right) \geq 1 - \alpha \right\}, \quad (4.7)$$

²⁷Existing literature offers alternative test statistic formulations. For example [Bai et al. \(2019\)](#) suggest using $\hat{D}'_n \equiv \max_{(j,j') \in \mathcal{S}_n^{\text{all}}} \hat{d}_n(j, j') \vee 0$, which leads to a better power if many elementary nulls $H_{j,j'}$ are violated simultaneously. On the contrary, our emphasis is on detecting deviations when at least one $H_{j,j'}$ is violated too much. Observing that [Theorem 2.1\(i\)](#) and [2.2\(ii\)](#) in [Lehmann et al. \(2005\)](#) indicate the impossibility of maximizing power across both cases even when total number of nulls are limited to be 2, we use \hat{D}_n instead of \hat{D}'_n here.

The following theorem provides validity of this Gaussian multiplier bootstrap procedure.

Theorem 4.3. *Suppose that (i) Assumption 3 is satisfied; (ii) $k_n \asymp \Delta_n^{-\rho}$ such that $\rho \in (0, 2r \wedge 1/2)$, then for some positive ϵ ,*

(i) $\mathbb{P}(\widehat{D}_n > cv_n^B(\alpha, \mathcal{S}_n^{\text{all}})) \leq \alpha + K\Delta_n^\epsilon$ if $\max_{(j,j') \in \mathcal{S}_n^{\text{all}}}(g_{n,j} - g_{n,j'}) \leq 0$. In addition, $|\mathbb{P}(\widehat{D}_n > cv_n^B(\alpha, \mathcal{S}_n^{\text{all}})) - \alpha| \leq K\Delta_n^\epsilon$ if $g_{n,j} - g_{n,j'} = 0$ for all $(j, j') \in \mathcal{S}_n^{\text{all}}$;

(ii) $\mathbb{P}(\widehat{D}_n > cv_n^B(\alpha, \mathcal{S}_n^{\text{all}})) \geq 1 - K\Delta_n^\epsilon$ if $\max_{(j,j') \in \mathcal{S}_n^{\text{all}}}(g_{n,j} - g_{n,j'}) \geq \Upsilon$ for some positive Υ .

COMMENT 1. Theorem 4.3 ensures the test $\hat{\phi}_n \equiv \mathbb{1}\{\widehat{D}_n > cv_n^B(\alpha, \mathcal{S}_n^{\text{all}})\}$ achieves asymptotic size control in detecting whether at least one of alternative $K_{j,j'}$ holds where $(j, j') \in \mathcal{S}_n^{\text{all}}$. Based on this result, we can show the test

$$\hat{\phi}_n(j, j') \equiv \mathbb{1}\{\hat{d}_n(j, j') > cv_n^B(\alpha, \mathcal{S}_n^{\text{all}})\},$$

provides a strong control of the familywise error rate, in the sense that $\mathbb{P}(\text{FWER}_n) \leq \alpha + K\Delta_n^\epsilon$. Furthermore, the theorem also shows proposed test is consistent against any (non-local) alternatives. Lemma 5.1 in Chernozhukov et al. (2019) indicates, under a simplified case where ζ is constant within each blocks and $R_{n,i} = 0$, no test can be uniformly consistent against all local alternatives with $\max_{(j,j') \in \mathcal{S}_n^{\text{all}}}(g_{n,j} - g_{n,j'}) = o(\Delta_n^{\rho/2} \log(\Delta_n^{-1})^{1/2})$.

COMMENT 2. The test $\hat{\phi}_n(j, j')$ proposed above is a straightforward one-step procedure that controls the familywise error rate, which could be conservative in application with finite sample. In the appendix, we prove that Theorem 4.3 remains valid even when $\mathcal{S}_n^{\text{all}}$ in formulations of \widehat{D}_n and $cv_n^B(\alpha, \mathcal{S}_n^{\text{all}})$ are replaced by any arbitrary subset $\mathcal{S}_n \subseteq \mathcal{S}_n^{\text{all}}$ with $|\mathcal{S}_n| \geq 3$. This stronger result facilitates the incorporation of a stepdown improvement akin to those provided in Romano and Wolf (2005). We summarize the ultimate testing procedure in the following steps contained in Algorithm 1.

The corollary below shows the validity of confidence sets generated by this stepdown procedure.

Corollary 4.1. *Under the same setting as Theorem 4.3. For $1 \leq j \leq m_n$, let*

$$\widehat{\text{Rank}}_n(j) \equiv \{|\text{Rej}_n^-(j)| + 1, \dots, m_n - |\text{Rej}_n^+(j)|\},$$

Algorithm 1 Stepdown Procedure

Step 1. Set $\mathcal{S}^{(0)} = \mathcal{S}_n^{\text{all}}$ and $i = 0$.

Step 2. Compute the critical value $cv_n^{(i)} = cv_n^B(\alpha, \mathcal{S}^{(i)})$ using (4.7).

Step 3. For all $(j, j') \in \mathcal{S}^{(i)}$, reject $H_{j,j'}$ according to $\hat{\phi}_n^{(i)}(j, j') = \mathbb{1}\{\hat{d}_n(j, j') > cv_n^{(i)}\}$. For $1 \leq j \leq m_n$, form $\text{Rej}_n^{(i),-}(j)$ and $\text{Rej}_n^{(i),+}(j)$ by the sets of nulls $H_{j,\cdot}$ and $H_{\cdot,j}$ rejected in this step, respectively. Let $\text{Rej}_n^{(i),\pm} = \bigcup_{j=1}^{m_n} \text{Rej}_n^{(i),\pm}(j)$.

If $|\text{Rej}_n^{(i),-}| = |\text{Rej}_n^{(i),+}| = 0$, form $\text{Rej}_n^{\pm}(j) = \bigcup_{\ell=0}^i \text{Rej}_n^{(\ell),\pm}(j)$, then stop.

Else, set $\mathcal{S}^{(i+1)} = \mathcal{S}^{(i)} \setminus \{(j, j') : (j, j') \in \text{Rej}_n^{(i),-} \cup \text{Rej}_n^{(i),+}\}$, $i \leftarrow i + 1$, return to Step 2.

where $\text{Rej}_n^{\pm}(j)$ is computed according to Algorithm 1. Then $\widehat{\text{Rank}}_n \equiv \prod_{j=1}^{m_n} \widehat{\text{Rank}}_n(j)$ constitutes a joint $(1 - \alpha)$ confidence set for ranks of process $(g_t)_{t \in [0, T]}$ at all evaluated time points.

4.3 Monte Carlo Simulations

4.3.1 Data Generating Processes

We conduct a Monte Carlo experiment to evaluate the performance of proposed inference procedures. Our simulation is anchored in the setting of motivating examples mentioned in Section 4.2.2. In each example, parameters used in data generating processes (DGP) and sampling schemes are selected to closely resemble the real data encountered in empirical application.

We first consider the location-scale model discussed in Example 1. Specifically, we focus on the following two data generating processes:

$$\text{DGP 1: } Y_{i\Delta_n} = \mu_{i\Delta_n} + \varepsilon_{n,i}, \text{ where } \varepsilon_{n,i} \sim_{i.i.d.} \mathcal{N}(0, 1),$$

$$\text{DGP 2: } Y_{i\Delta_n} = \mu_{i\Delta_n} + \sigma_{i\Delta_n} \varepsilon_{n,i}, \text{ where } \varepsilon_{n,i} \sim_{i.i.d.} t(3).$$

DGP 1 and 2 align with the conventional additive state-space model, wherein the state process of interest is $(\mu_t)_{t \in [0, T]}$ and will be estimated through conditional mean process analyzed in Section 4.2.3. Notably, in DGP 1, the random disturbance is assumed to follow an i.i.d. standard Gaussian distribution, so that each spot estimator retains its Gaussianity even when the number of observations in each block is small. In contrast, DGP 2 introduces both heteroskedasticity in time and non-Gaussian disturbance. Regarding the Lévy-driven returns discussed in Exam-

ple 2, we simulate price processes with Blumenthal–Gettoor index $\beta \in \{2, 1.5, 1\}$, which correspond to instances of Cauchy process C , a general Lévy process L , and a Brownian motion W . Specifically, we focus on the following three data generating processes:

$$\begin{aligned} \text{DGP 3 : } \quad Y_{i\Delta_n} &= \Delta_n^{-1} \left(\int_{i\Delta_n}^{(i+1)\Delta_n} \mu_s ds + \int_{i\Delta_n}^{(i+1)\Delta_n} \sigma_s dW_s \right)^2, \\ \text{DGP 4 : } \quad Y_{i\Delta_n} &= \Delta_n^{-4/3} \left(\int_{i\Delta_n}^{(i+1)\Delta_n} \mu_s ds + \int_{i\Delta_n}^{(i+1)\Delta_n} \sigma_s dL_s \right)^2, \\ \text{DGP 5 : } \quad Y_{i\Delta_n} &= \Delta_n^{-2} \left(\int_{i\Delta_n}^{(i+1)\Delta_n} \mu_s ds + \int_{i\Delta_n}^{(i+1)\Delta_n} \sigma_s dC_s \right)^2. \end{aligned}$$

In forming these processes, we adopt a truncation technique analogous to the one employed in Bugni et al. (2023) for stable distributions such that the normalized increment takes value in $[-30, 30]$ to avoid unrealistic price paths. The state process of interest is variance process $(\sigma_t^2)_{t \in [0, T]}$, which is estimated through conditional mean process for DGP 3, or through conditional median process (i.e. $\chi = 1/2$) analyzed in Section 4.2.4 for DGP 4 and 5. Additionally, we focus on DGP 6 which serves as a representative illustration of Cox trading flow process discussed in Example 3:

$$\text{DGP 6 : } \quad Y_{i\Delta_n} = N_{(i+1)\Delta_n} - N_{i\Delta_n},$$

where $(N_t)_{t \in [0, T]}$ is a Cox process with intensity $(\mu_t)_{t \in [0, T]}$.

The state process of interest is the normalized intensity $(\mu_t)_{t \in [0, T]}$, which will be estimated through conditional mean process.

Recall that we have two auxiliary processes μ and σ which serve as state processes in our specified DGPs. In alignment with the conventional setting in existing literature, see, e.g., Jacod et al. (2017) and Li and Linton (2022), we assume μ and $c \equiv \sigma^2$ to follow these Ornstein–Uhlenbeck-type processes

$$\begin{aligned} d\mu_t &= \rho(\bar{\mu}_t - \mu_t)dt + \zeta dB_t, \\ dc_t &= \kappa(\alpha_t - c_t)dt + \gamma \sqrt{c_t} dB'_t, \end{aligned}$$

where B and B' are two independent Brownian motions. Following empirical results calibrated in the literature, we choose two parameter configurations summarized in Table 4.1. Setting (a) is more conservative comparing with setting (b),

Table 4.1: Parameter Specification for the Simulation Study

Setting	$\bar{\mu}_t$	ρ	ζ	α_t	κ	γ
(a)	1.2	8/252	1.25/252	0.04/252	5/252	0.05/252
(b)	$1.2h(t)$	4/252	2.5/252	$0.04/252h(t)$	4/252	0.1/252

Note: The table displays parameter configurations used in the simulation study. All parameters are in their daily value as the fixed time span $T = 1$ has been normalized to one trading day. Here $h(t) \equiv 1 + 0.1 \cos(2\pi t)$ is a U-shaped function to mimic the diurnal feature.

in the sense that μ is stationary, and c follows a Cox–Ingersoll–Ross (CIR) model which has been extensively utilized to capture the volatility dynamics, see, e.g., [Cox et al. \(1985\)](#) and [Heston \(1993\)](#). The parameters are chosen in accordance to [Li and Linton \(2022\)](#). Setting (b) differs from the previous configuration in two aspects. First, the mean processes $\bar{\mu}$ and α are time variant and exhibit systematic moves in time, which the literature identifies as diurnal features. Namely, a nearly U-shaped pattern has been documented for both intraday trading volume and volatility in real data, see [Ito \(2013\)](#), [Christensen et al. \(2018\)](#), and [Andersen et al. \(2019\)](#).²⁸ Moreover, state processes under setting (b) are more volatile than those under the previous configuration, attributable to smaller mean reverting parameters and larger variance magnitude. In summary, we have six types of DGPs in conjunction with two sets of parameter configurations. The combination yields $6 \times 2 = 12$ different DGPs for examination. For notation clarity, we use DGP 1(a) to indicate DGP 1 equipped with parameter setting (a), and similarly for other combinations.

For the observation scheme, we normalize $T = 1$ trading day, and consider two sampling frequency, $\Delta_n \in \{1/390, 1/23400\}$, which correspond to 1-minute and 1-second data, respectively. We stress that 1-second sampling frequency is not practically feasible for DGP 3-5 to hold in reality, wherein the observed price in such high-frequency is contaminated by the so-called microstructure noise, see, e.g., the discussion in [Zhang et al. \(2005\)](#). Empirical evidence such as a signature plot of the realized volatility in relation to sampling frequency shows that noise component overshadows when sampling scheme is “too fine,” typically less than 1 minute.

²⁸The rationale from economic theory concerning these observed intraday pattern is provided in [Admati and Pfleiderer \(1988\)](#) and [Hong and Wang \(2000\)](#), among others.

Therefore, for DGP 3-5 we exclusively consider 1-minute data, in which the effect of noise is inconsequential with respect to returns of efficient price. Conversely, given our application of DGP 6 in empirical illustrations wherein trading flow data is recorded at an ultra-high-frequency and where approximations could falter with coarser sampling frequency, we exclusively consider 1-second data for DGP 6. The selection of tuning parameter $k_{n,j}$ is described as follows. We partition observations into equal-sized blocks, i.e. $k_{n,j} = k_n$ for all $1 \leq j \leq m_n$. For 1-minute data, we adopt $k_n \in \{20, 30, 40\}$, representing blocks of $\{20, 30, 40\}$ minutes, respectively. The corresponding number of blocks is $m_n \in \{19, 13, 9\}$. For 1-second data, we adopt $k_n \in \{300, 600, 1200\}$, representing blocks of $\{5, 10, 20\}$ minutes, respectively. The corresponding number of blocks is $m_n \in \{78, 39, 19\}$. All the “continuous-time processes” are simulated using a Euler scheme with mesh size being 10^{-4} minute. The simulation is based on 10000 Monte Carlo draws. We examine the coverage rate of 90% confidence bands constructed in accordance with (4.4) and (4.5) for conditional mean processes and conditional median processes, respectively.

4.3.2 The Results

Table 4.2 shows the coverage rate of confidence bands (4.4) and (4.5) under our specified DGPs. In the case where $\Delta_n = 1/390$, i.e. data is observed every one minute, not surprisingly, proposed confidence bands perform bad when the number of observation in each block is small, say $k_n = 20$, especially for DGP 2(a) and 2(b). This is particularly due to the poor approximation of Gaussian distribution for spot estimators in small sample. As k_n becomes larger, coverage rates elevate remarkably. For instance, when $k_n = 40$, coverage rates are above 80% for all DGPs, with the exception of 2(a) and 2(b). In the meantime, there is a considerable increment in time-variation effects of state processes within each block as block size expands. Notably, coverage rates for DGPs equipped with parameter setting (b) are generally lower than the same DGPs equipped with parameter setting (a) when k_n becomes larger. Intriguingly, coverage rates under DGP 5(a)-5(b) are higher than those under 3(a) and 4(b), suggesting that the employment of conditional quantile processes is particularly efficient when driving processes of price markedly deviate

Table 4.2: Coverage Rate of Uniform Confidence Band

DGP	$\Delta_n = 1/390$			$\Delta_n = 1/23400$		
	$k_n = 20$	$k_n = 30$	$k_n = 40$	$k_n = 300$	$k_n = 600$	$k_n = 1200$
1(a)	0.7253	0.8257	0.8113	0.8907	0.8933	0.8937
1(b)	0.7166	0.8254	0.8058	0.8824	0.8834	0.8841
2(a)	0.6271	0.7339	0.7212	0.8115	0.8654	0.8829
2(b)	0.6223	0.7303	0.7191	0.7996	0.8580	0.8792
3(a)	0.7268	0.8311	0.8308	—	—	—
3(b)	0.7295	0.8339	0.8290	—	—	—
4(a)	0.7744	0.8147	0.8304	—	—	—
4(b)	0.7858	0.8044	0.8282	—	—	—
5(a)	0.8823	0.8916	0.8949	—	—	—
5(b)	0.8809	0.8912	0.8915	—	—	—
6(a)	—	—	—	0.8585	0.8868	0.8905
6(b)	—	—	—	0.8628	0.8782	0.8890

Note: The table reports the coverage rates of a 90%-level confidence band computed according to (4.4) for DGP 1(a)-4(b), DGP 7(a), and 7(b), according to (4.5) for DGP 5(a)-6(b). Column 2-4 correspond to 1-minute data, column 5-7 correspond to 1-second data. Note that some results are omitted with dash signs (—), which indicates the sampling frequency is not practically appropriate for certain models to hold true in real observed data.

from Brownian motions. For a higher sampling frequency, $\Delta_n = 1/23400$, where data is observed every one second, coverage rates are above 85% for all DGPs when $k_n \geq 600$. Drawing a parallel between results for DGP 1(a)-1(b) under column 1 and 7, both scenarios have a block length of 20 minutes and same number of blocks, i.e., time-variation effects are same. There is a substantial improvement in convergence rate from $\Delta_n = 1/390$ to $\Delta_n = 1/23400$. Recall the Gaussian nature of disturbance terms, each spot estimator maintains its Gaussianity in finite samples, hence the only difference lies in sampling frequency. A similar comparison for 2(a)-2(b) indicates pointwise approximation errors and time variation effects can be controlled simultaneously by adapting a finer sampling scheme.

In summary, above simulation results show that proposed confidence bands aptly cover true processes across all data generating processes aligned with an appropriate sampling frequency. Although under certain DGPs they appear to have poor performance when the number of observations in each block is insufficient, this problem can be effectively addressed by adapting a larger block size with a finer sampling scheme. These simulation results stress that the proposed inference method remains robust in contexts analogous to market settings. Moreover, in order to achieve better performance of proposed inference procedures, one should employ the highest justifiable sampling frequency and choose block sizes carefully in a suitable range to mitigate time variation effects in state processes.

4.4 Empirical Illustration

4.4.1 Detecting Information Flows during FOMC Speeches

The Federal Open Market Committee (FOMC) announcement, accompanied by the subsequent press conference held by chair of the Federal Reserve, currently Jerome Powell, plays a pivotal role in disseminating Fed decisions and conveying information pertinent to future financial policy. On each pre-scheduled date and time, Fed issues an official statement that summarizes the committee's assessment of U.S. economy, its policy decisions, and the rationale behind those decisions. In particular, the statement provides insights into committee's outlook on inflation, employment, and other economic indicators. The release of this official document

usually has a significant market impact, see, e.g., [Cochrane and Piazzesi \(2002\)](#), [Rigobon and Sack \(2004\)](#), [Bernanke and Kuttner \(2005\)](#), and [Nakamura and Steinsson \(2018b\)](#). In addition, [Savor and Wilson \(2014\)](#), [Lucca and Moench \(2015\)](#), and [Bollerslev et al. \(2021\)](#) also found evidence of pre-announcement effects of the initial release. On the other hand, with more accurate volatility estimation, [Bollerslev et al. \(2024\)](#) found that announcements of new policy decision may not cause the most substantial shocks during FOMC days, especially when corresponding policy changes are well anticipated by the market.²⁹ In that case, information embedded with forward guidance, which can be used to forecast future financial policies, tends to have a more pronounced market impact.

In conjunction with FOMC statements, Fed holds a press conference which usually starts 30 minutes after the initial release and lasts about 60 minutes. The press conference provides an opportunity for Powell to elaborate on FOMC's decision-making process, provide additional context, and address questions from media. It allows for a more in-depth discussion of committee's views on the economy and financial policy. During press conferences, Powell inevitably reveals some (possibly subtle) forward guidance, more precisely, information about the expected path of monetary policy in the future. Such information may include hints about potential changes in interest rates, the balance sheet, or other policy tools. The aim is to offer transparency and help market participants anticipate Fed's future actions.

Pinpointing the exact sentences in press conferences that provide additional information regarding forward guidance, however, is a challenging task. Since each sentence in the press conference is typically spoken within a few seconds, this rapid succession of sentences and limited time span of each sentence makes it difficult to isolate their individual impact on market volatility. Namely, analyzing volatility changes at second level requires examining ultra-high-frequency data, such as tick-by-tick price. That being said, ultra-high-frequency price data is often subject to microstructure noise, which distorts the identification of precise volatility patterns, see, e.g., [Zhang et al. \(2005\)](#). To mitigate the impact of noise on volatility analysis, existing procedures such as [Barndorff-Nielsen et al. \(2008\)](#), [Jacod et al. \(2009\)](#), and

²⁹For instance, market predicted probabilities of changes to the Fed rate and monetary policy are reported on FedWatch website (<https://www.cmegroup.com/markets/interest-rates/cme-fedwatch-tool.html>), which is provided by CME Group and updated at a daily frequency.

Kristensen (2010) often use increasing number of return observations, hence have to employ wider estimation windows. This, however, makes it more involved to detect specific volatility patterns within seconds.

Utilizing textual analysis on the conference scripts is another approach to studying FOMC press conferences. With developing natural language processing (NLP) methods, textual analysis algorithms have found prevalent application in economics and finance, as discussed in Gentzkow et al. (2019), Ke et al. (2019), Engle et al. (2020), Loughran and McDonald (2020), and Cheng et al. (2021). Nonetheless, in the formal announcing scenario like FOMC meetings, conventional NLP methods based on experiences might exhibit considerable inaccuracies. To better understand this possible limitation of stand-alone textual analysis, we deploy an algorithm to score each sentence by the level of forward guidance it carries. The assessment of forward guidance levels is based on a combination of factors such as the presence of specific trigger keywords and phrases that are commonly associated with forward guidance, the clarity of future policy intentions, and the level of details provided about future actions. To this end, we use Generative Pre-trained Transformer (ChatGPT) 3.5,³⁰ an expansive language model pioneered by OpenAI, to extract features that could be essential signals indicating a high level of forward guidance.³¹ Below is a brief overview of features the algorithm takes into account:

Trigger Keywords and Phrases: Certain keywords and phrases are strong indicators of forward guidance, including words that refer to future actions, intentions, or plans, such as “expect,” “anticipate,” “will be appropriate,” “likely,” “plan,” and so on.

Level of Detail: Sentences that provide specific details about future policy actions are more informative, including the announcement of specific interest rate changes, plans for balance sheet reduction, or discussions about future meetings.

Clarity and Directness: Sentences that clearly state the course of future monetary policy are given higher scores. The more direct and unambiguous the statement

³⁰ChatGPT 3.5 was trained with data up to September 2021, hence has no knowledge beyond that cutoff. This ensures that extracted features are intrinsically rooted in the in-context learning procedure, without “sneak peek” at contemporaneous market activities. Even so, the same analysis performed with ChatGPT 4 yields a similar result.

³¹Recently, Hansen and Kazinnik (2023) showed GPT models deliver a considerable improvement in determining sentences in FOMC statements as “dovish” or “hawkish”, over other commonly used classification methods.

is, the more likely it is to be a clear form of forward guidance.

Contextual Analysis: The overall context of each sentence and how it fits within the whole speech matters. This includes patterns and consistency in the language used to convey future policy intentions.

Quantitative and Qualitative Aspects: Both quantitative aspects (e.g., specific percentages or values) and qualitative aspects (e.g., intentions, expectations) are assessed.

Comparative Analysis: The comparison of each sentence with other sentences within the speech is considered to obtain a relative ranking of strength in forward guidance. This takes into account the range of guidance provided throughout the speech.

For illustrative purposes, we present the following two sentences extracted from May 4, 2022 speech, offering contrasting levels of forward guidance based on above features.

Against the backdrop of the rapidly evolving economic environment, our policy has been adapting, and it will continue to do so.

14:34:15-14:34:23

Assuming that economic and financial conditions evolve in line with expectations, there is a broad sense on the Committee that additional 50-basis-point increases should be on the table at the next couple of meetings.

14:34:50-14:35:04

The algorithm then computes a weighted averaged scores of aforementioned aspects. Note that this algorithm is designed to identify potential forward guidance purely based on linguistic patterns and context, where scores are indicative rather than definitive. The assessment also accounts for variations in language and communication styles, so it may represent a nuanced interpretation of forward guidance strength in the given context. Based on this algorithm, we can partition each speech into five groups, indicates the possible level of forward guidance contained in each sentence:

Very Low	No forward guidance or very limited forward guidance
Low	General mention of current economic situation, no clear future policy intentions
Medium	Some specific indications about future policy intentions, but not very clear
High	Clear and specific forward guidance about future policy intentions
Critical	Very strong and specific forward guidance about future policy intentions

We apply the above textual analysis procedure to eight press conference speeches on the FOMC announcement days last year. The proportion of sentences marked as “very low,” “low,” “medium,” “high,” and “critical” information level are 8.4%, 10.3%, 43.7%, 37.4%, 0.2%, respectively. This indicates that there are about 80% of speeches has been designated to carry medium or high level of information. Drawing on the Mixture of Distribution Hypothesis (MDH), see, e.g., [Tauchen and Pitts \(1983\)](#), [Harris \(1987\)](#), and [Andersen \(1996\)](#), market indicators can effectively serve as proxies for corresponding information flows. In particular, as shown in [Du and Zhu \(2017\)](#), higher trading frequency indicates a higher information level. Therefore, in order to gain a direct insight on the accuracy of this procedure, we mark relative information level and estimated trading intensities in the same timeline, to conduct a visual comparison. For trading flows, we use nanosecond-level data of S&P 500 ETF (ticker: SPY), downloaded from Trade and Quote (TAQ) database. We estimate second-level trading intensities during each FOMC press conference speech, i.e. $\Delta_n = 1/(2.34 \times 10^{13})$, $k_n = 10^9$ so that $k_n \Delta_n = 1$ sec corresponds to one-second block. In [Figure 4.2](#), we plot estimated trading intensities during the press conference speeches, and colored each horizontal line in the gradient spectrum such that sentences with lowest information level (i.e., labeled “very low”) tend to be transparent green, where sentences with highest information level (i.e., labeled “critical”) tend to be red. As [Figure 4.2](#) shows, there are large amount of informative sentences following by barely no intensity variation, indicating the market has

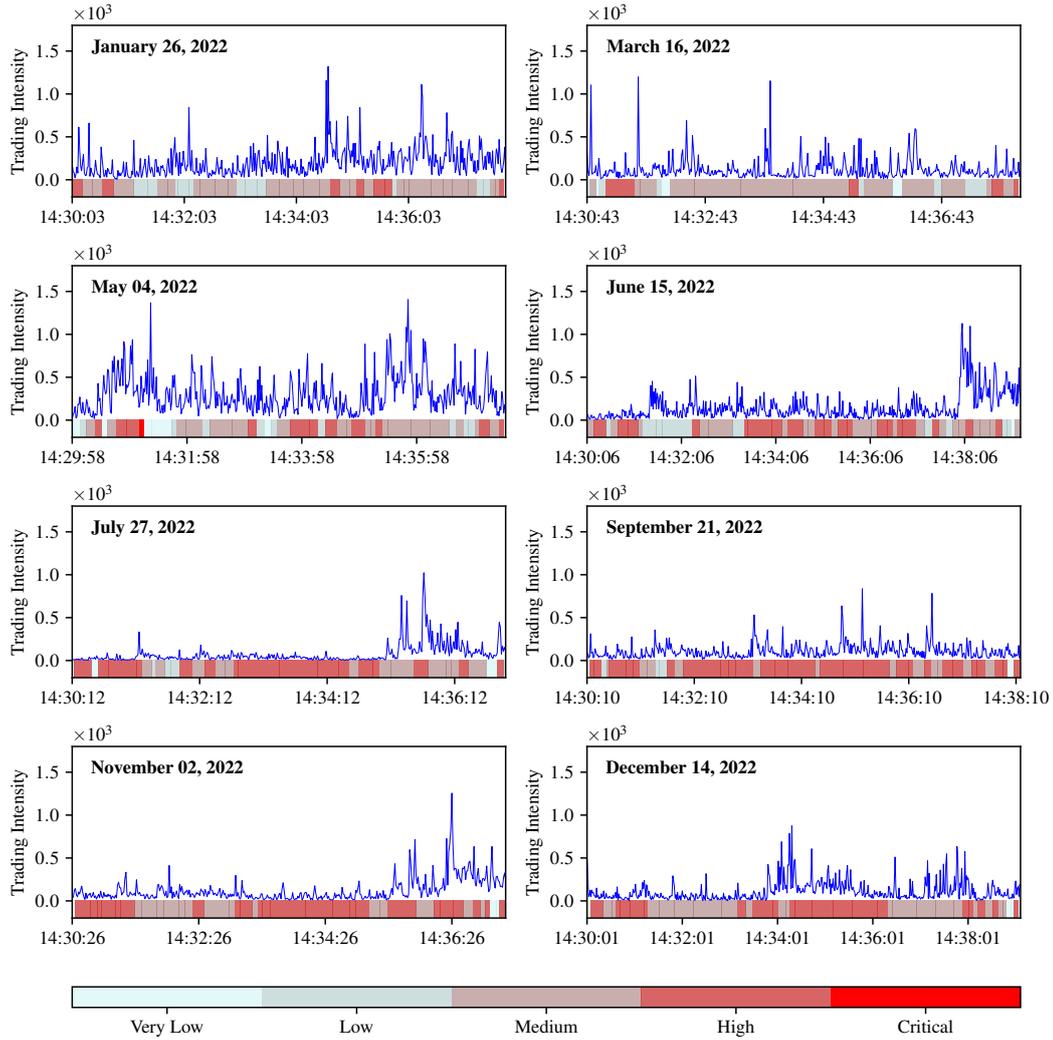


Figure 4.2: Trading Intensities and Relative Information Levels during FOMC Press Conference Speeches. The figure plots one-second trading intensities during eight FOMC press conference speeches in 2022. The horizontal axis is colored according to the relative information level embedded in potential forward guidance contained in each sentence, which is computed using the algorithm described in this section. The color bar is shown at the bottom, and is determined by $RGB\alpha = (s/5, 1 - s/5, 1 - s/5, (s/5)^{1.25})$ where s denotes the information level in the scale of 1 to 5, with 1 being “very low,” 5 being “critical.”

no reactions to them.

Next, we delve deeper into the textual analysis outcomes, exploring trading intensities across categorized groups. Considering potential reactive latency between information arrivals and correspondent trading actions, we shift observation windows to the right, spanning lags as $\{0, 1, \dots, 19\}$ seconds. Figure 4.3 illustrates the dispersion of trading intensities across different groups for various lags, together with medians and means with each group. We further conduct Welch’s t -tests to determine if sentences identified with a higher information level truly exhibit an elevated trading intensity. The results indicate that, even under the best case (i.e., a 14 seconds lag), where the group labeled “critical” has significantly higher intensity than other groups, we cannot conclusively negate the possibility of no significant distinctions among all other four groups.

The main inherent challenge of pure textual analysis approaches stems from the carefully crafted nature of speech scripts and potential overlaps between successive press conferences. The language used in FOMC press conference scripts is often meticulously chosen to avoid causing sudden market shocks. Consequently, detecting specific keywords or phrases that could potentially trigger market reactions may not yield significant insights, given the scripts are designed to convey information while maintaining stability and avoiding unnecessary shocks. Moreover, press conference speeches tend to have recurring themes and structures, resulting in similarities between successive scripts, as visually shown in Figure 4.4. Namely, we characterize the speech at time t_i as a set A_{t_i} of individual sentences, and gauge similarities by computing Jaccard similarity coefficients (Jaccard (1912)) between these sets,³²

$$S(A_{t_1}, \dots, A_{t_n}) \equiv \frac{|\bigcap_{i=1}^n A_{t_i}|}{|\bigcup_{i=1}^n A_{t_i}|}.$$

The repetition of certain phrases or topics have two-sided effects. Obviously, it will diminish their impact on market expectations over time. On a flip side, a nuance in language of these topics could result in a considerable market effect. Textual analysis techniques that focus solely on keyword detection might identify familiar

³²Alternatively, one can use Szymkiewicz–Simpson coefficient $S'(A_{t_1}, \dots, A_{t_n}) \equiv |\bigcap_{i=1}^n A_{t_i}| / \min_{1 \leq i \leq n} |A_{t_i}|$, the results are similar given that the lengths of speeches under consideration do not exhibit significant difference.

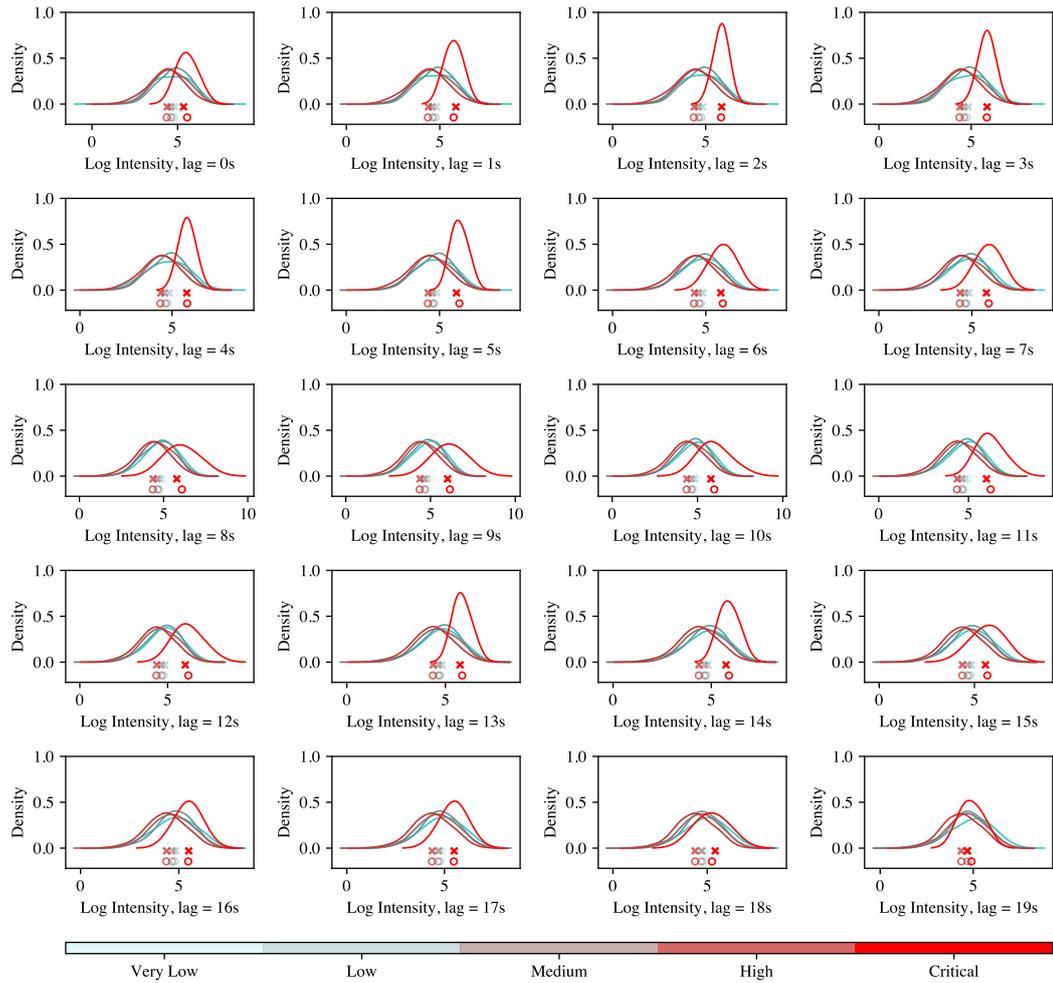


Figure 4.3: Distribution of Intensity with Different Information Levels. The figure plots the kernel density estimation of trading intensities with different relative information level embedded in the potential forward guidance contained in each sentence, which is determined using the algorithm described in this section. In each panel, we shift the window by several seconds to take account the effect of market reaction time between information arrivals and tradings. The color of each line follows the same rule as in Figure 4.2, the median and the mean of each group are marked in \times and \circ sign, respectively.

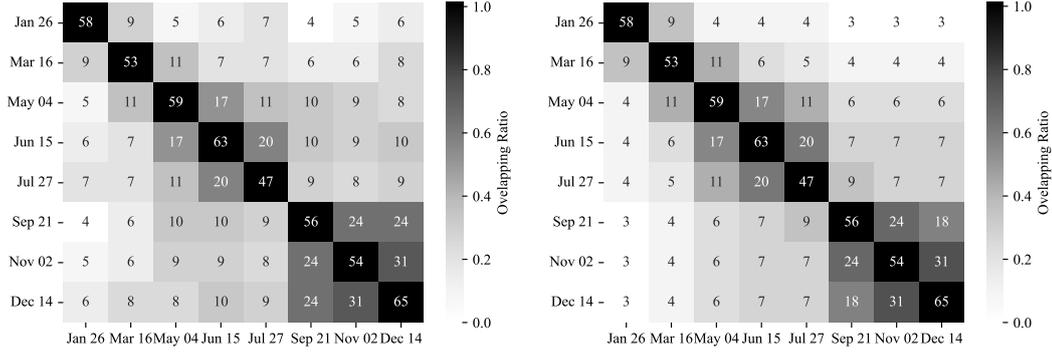


Figure 4.4: Similarity of FOMC Press Conference Speeches. The figure plots the overlapping ratio between different speeches. The overlapping ratio is defined as $\log S_{i,j}^{\text{type}}$, where $\text{type} \in \{\text{pw}, \text{cm}\}$. In the left panel, $S_{i,j}^{\text{pw}}$ is the pairwise Jaccard similarity index, defined as the number of pairwise overlapping sentences between speeches at date t_i and t_j divided by the total number of sentences. In the right panel, $S_{i,j}^{\text{cm}}$ is the cumulative Jaccard similarity index, defined as the number of cumulative overlapping sentences between speeches within $\{t_i, \dots, t_j\}$ divided by the total number of sentences. Exact numbers of pairwise and cumulative overlapping sentences are displayed in each square.

terms without considering market’s prior knowledge of their significance, hence tend to overestimate the market impact of those sentences.

To establish a reference for the “true” information level predicated on actual market reactions, we partition speeches in accordance with estimated trading intensities. Specifically, on each day, we conduct the joint testing procedure proposed in Section 4.2.5, and construct a 90% confidence set for ranks of all second-level intensities. Based on this results, we can partition each speech into groups $G \in \{1, \dots, \bar{G}\}$ via the following algorithm: First, we permute indices such that $\hat{g}_{\pi(1)} \leq \hat{g}_{\pi(2)} \leq \dots \leq \hat{g}_{\pi(m_n)}$. Starting from $\pi(1)$, which initiates the first group $G = 1$, if $\widehat{\text{Rank}}_n(\pi(j+1)) \cap \widehat{\text{Rank}}_n(\pi(1)) \neq \emptyset$, then $\pi(j+1)$ belongs to the same group as $\pi(1)$; otherwise, $\pi(j+1)$ initiates a new group $G \leftarrow G + 1$. Repeat until the last second $j = m_n$. In Figure 4.5, we present a heatbar of speeches according to the trading intensity and color it in the same way such that groups with lowest intensity tends to be transparent light green, groups with highest intensity tends to be red. The resulting pieces marked as “very low,” “low,” “medium,” “high,” and “critical” information level are 51.1%, 34.6%, 11.6%, 2.3%, 0.4%, respectively. Comparing with the outcomes given by pure textual analysis, around 80% of these speeches actually impart minimal information, as evidenced by low trading intensities. Most of them are repeated sentences across consecutive speeches, which theoretically,

should not disseminate any novel information after their debut. Meanwhile, on the contrary, we detect more sentences that are markedly informative.

In conclusion, the comparison result suggests stand-alone NLP methods overstates the information level of individual sentence, and in the meantime fails to accurately identify the most informative parts, indicating that NLP methods tend to smooth out true information flows. This is driven by the *in-context learning* nature of our task, i.e., no “training sample” is provided. Therefore, the classification is solely based on ChatGPT’s pre-existing knowledge, hence the intrinsic Bayes classifier method gives mediocre scores to most sentences based on its inherent prior, which is improper for analyzing these scripts. On the other hand, our intensity-based analysis based on proposed uniform inference procedure offers a compliment to NLP methods. One can refine textual analysis procedures by deploying a *supervised learning*, i.e., utilize the intensity-level-labeled text as training samples in order to obtain a more accurate classification.³³ The detailed implementation of this supervised learning procedure is beyond the scope of this paper and deserves dedicated study.

4.4.2 Case Study

Next, we conduct a case study to better illustrate preceding findings, opting for specific sentences from these speeches that stand out as high level of information about forward guidance and followed with considerable intensity spikes. The first sentence is a shift in tone about longer-term inflation expectations that presents a double twist, first mentioned in the September conference:

[A] *Despite elevated inflation, longer-term inflation expectations appear to remain well anchored, as reflected in a broad range of surveys of households, businesses, and forecasters as well as measures from financial markets. But that is not grounds for complacency; the longer the current bout of high inflation continues, the greater the chance that expectations of higher inflation will become entrenched.*

³³See, e.g., Table 4 in Hansen and Kazinnik (2023), where the mean-absolute-error of fine-tuned model (supervised learning) is nearly half of that of zero-shot model (in-context learning) in classifying the policy stances of Fed speeches.

The first twist offers an optimistic note: even though the prevailing inflation remains not fully controlled, there exists empirical evidence suggesting that longer-term inflation is effectively anchored. After that, a second twist makes additional comments that this situation is not yet ripe for complacency, rendering the entire statement more balanced. Top panel of Figure 4.5 illustrates there are two succeeding trading intensity spikes a few seconds after these twist indications. The second sentence of interest sounds more assertive and supports the second twist of sentence [A], which is also first mentioned during the September conference:

[B] *The historical record cautions strongly against prematurely loosening policy.*

Another intensity spike is observed several seconds after sentence [B]. Interestingly, aforementioned sentences [A] and [B] recur in both November and December conferences. On the contrary, these repetitions do not elicit similar intensity spikes. In fact, the bottom panel of Figure 4.5 indicates an overall absence of significant trading spikes during the December conference. This observation aligns with the result shown in Figure 4.4 that approximately half of the December speech mirrors exact content from preceding conferences. This coincides with the intuition that new information occurs only when it is introduced for the first time. After this immediate reaction, market quickly accepts it and subsequent repetitions of the same sentence are lack of novelty.

During the September conference, inquiries emerged concerning Fed's consideration of variable lags in inflation. This stemmed from the apprehensions that reported inflation was not accurately reflecting real-time economic conditions, and that the prevailing interest rate was overly elevated. In response to these concerns, Fed incorporates specific remarks about such lags in both the official statement and press conference speech:

[C] *That's why we say in our statement that in determining the pace of future increases in the target range, we will take into account the cumulative tightening of monetary policy and the lags with which monetary policy affects economic activity and inflation.*

As shown in the middle panel of Figure 4.5, there is also a considerable intensity spike shortly after sentence [C]. In the same speech, upon mentioning short-term appropriateness of decelerating the pace of rate hikes as it is near a level sufficiently restrictive to realign inflation with the 2 percent target, Powell acknowledged the uncertainty about that specific interest rate level and concludes with:

[D] *Even so, we still have some ways to go, and incoming data since our last meeting suggest that the ultimate level of interest rates will be higher than previously expected.*

Above sentence [D], although not definitive, is followed by a substantial intensity shock, as shown in the middle panel of Figure 4.5. Given projections released in the September meeting, market anticipation was an additional 75bps increase in November, followed by a deceleration in December. The shock stems from the revelation that incoming data after September might imply a trajectory towards a higher level than market initially expects.

4.4.3 Impact of Twitter on Cryptocurrency Markets

We provide another empirical application to highlight the importance of employing quantiles in addressing specific problems. As an active participant in cryptocurrency market,³⁴ the impact of Elon Musk’s tweets on cryptocurrency market has been extensively examined, see, e.g., Shen et al. (2019), Tandon et al. (2021), and Ante (2023). Notably, while these studies reveal substantial effects of tweets on the trading volumes of various cryptocurrencies, price effects are statistically significant only in the case of Dogecoin-related tweets, with barely no considerable impact on Bitcoin. Recent evidence in Kolokolov (2022) shed light on this phenomenon, showing that estimated jump activity index of Bitcoin is strictly less than 2, i.e., Bitcoin price is driven primarily by a pure jump process. Consequently, realized variances computed in the usual way becomes diverging,³⁵ and the detection of abnormal returns, as well as associated t -tests, would be invalid.³⁶

³⁴Namely, Tesla invested \$1.5 billion in Bitcoin during the first quarter of 2021, as indicated in the annual report of Tesla, Inc., U.S. Securities and Exchange Commission (https://www.sec.gov/Archives/edgar/data/1318605/000156459021004599/tsla-10k_20201231.htm).

³⁵Recall the second moment of normalized Lévy increments has an order of $\Delta_n^{2/\beta-1}$.

³⁶A robust t -test (see, e.g., Ibragimov and Müller (2010)) could be implemented in this case, but that requires estimating β first, making it a sequential inference procedure.

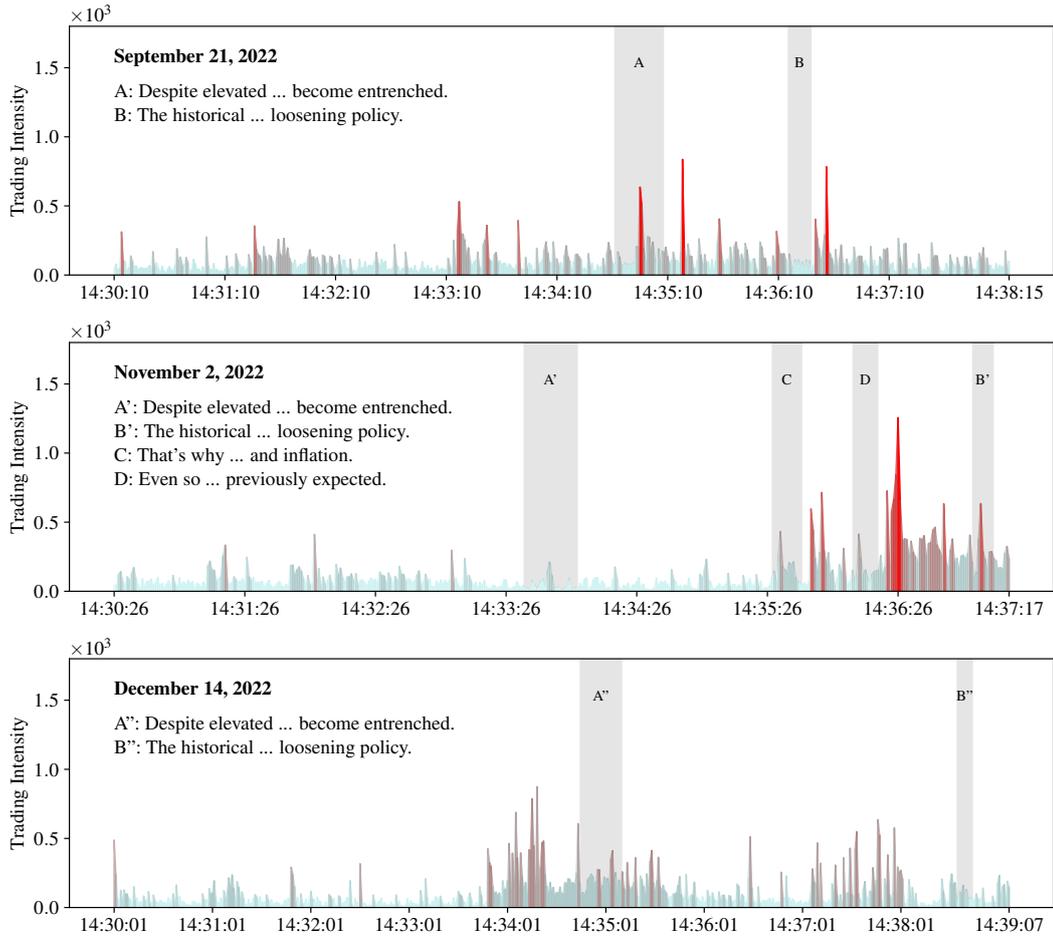


Figure 4.5: Trading Intensity during FOMC Press Conference Speeches. The figure shows the heatbar of estimated trading intensities during FOMC press conference speeches on September 21, November 2, and December 14 in 2022, arranged from the top panel to the bottom panel, respectively. On each of these dates, a 90% confidence set of joint ranks is constructed using Algorithm 1 proposed in section 4.2.5. Further, each speech was partitioned into groups using the strategy described in this section. The heatbar is colored according to group structure by the rule $RGB\alpha = (G/\bar{G}, 1 - G/\bar{G}, 1 - G/\bar{G}, (G/\bar{G})^{1.25})$ so that the color of each group remains the same as in Figure 4.2. The duration of target sentences are shaded light gray in each panel, where primes in the label indicate repetitions.

As discussed in Section 4.2.4, a feasible measurement for price volatile level can be constructed using quantile. To better illustrate this point further, we conduct an event study employing the same set of tweets investigated by studied in Ante (2023). These tweets, posted by Elon Musk between January 2020 and July 2021, are either directly or indirectly related to Bitcoin. For each event, we estimate the blockwise level of volatile V_j of (log) BTC/USD prices in the same day. We consider two proxies for this volatile level: $V_{j,1} \equiv q_j(0.5)$, representing the median, and $V_{j,2} \equiv q_j(0.75) - q_j(0.25)$, representing the interquartile range. To assess the price impact, we jointly test whether price volatile level in the block immediately following the tweet significantly deviates from those in other blocks. Formally, the null hypotheses and associated alternatives are defined as

$$H_j^{(i)} : V_{j^*,i} = V_{j,i} \quad \text{against} \quad K_j^{(i)} : V_{j^*,i} \neq V_{j,i},$$

where $i \in \{1, 2\}$, $1 \leq j \leq m_n$ with $j \neq j^*$, and j^* indexes the first block starting at the time when the tweet is posted. The length of each block was selected to be one and two hours, corresponding to $m_n = 24$ and 12, receptively. The test is performed using pairwise t -type statistics similar to the method outlined in Section 4.2, and the critical value is computed using bootstrap.

Table 4.3 presents the test statistics along with their corresponding significance levels. Comparing to the results obtained from the conventional mean-based t -test as presented in Ante (2023), we find evidence that a larger number of events exhibit a significant impact on the Bitcoin price. Namely, within a 2-hour horizon, twelve out of the fourteen tweets yield a significant price impact, in contrast to only four that can be identified using the t -test based on abnormal returns. As mentioned before, this disparity can be attributed to the potential divergence in return variance, rendering the conventional t -test invalid. Meanwhile, we stress that the result remains robust when considering different proxies for measuring the volatile level, highlighting the significance of our quantile-based inference procedure.

Table 4.3: Event Study Results for BTC/USD Price

No.	Time & Date	Tweet	1 Hour ($m_n = 24$)			2 Hours ($m_n = 12$)		
			t -stat.	Med.	IQR	t -stat.	Med.	IQR
1	07:53 Jan 10, 2020	Bitcoin is not my safe word	-0.88	1.81	0.01	-0.78	8.82***	6.02***
2	09:21 Dec 20, 2020	Bitcoin is my safe word	-0.46	0.20	1.10	-1.18	1.43	1.49
3	09:22 Jan 29, 2021	In retrospect, it was inevitable (<i>Twitter bio change</i>)	1.94*	1.24	2.98**	1.76	9.87***	4.68***
4	08:18 Feb 10, 2021	This is true power haha (<i>picture about Bitcoin</i>)	-0.51	0.80	1.14	0.36	0.92	1.58
5	00:42 Feb 21, 2021	Cryptocurrency explained (<i>link to a video</i>)	1.62	1.89	1.04	2.16*	6.13***	2.50*
6	18:50 Mar 02, 2021	Scammers & crypto should get a room	0.43	0.99	2.45	0.56	11.04***	7.48***
7	19:58 Mar 12, 2021	BTC (Bitcoin) is an anagram of TBC (The Boring Company)	-0.95	0.29	0.15	-1.18	6.26***	11.72***
8	08:02 Mar 24, 2021	You can now buy a Tesla with Bitcoin	1.17	0.46	0.14	1.63	4.55***	3.07**
9	00:06 May 13, 2021	Tesla & Bitcoin (<i>picture about suspending Bitcoin</i>)	-0.91	3.03**	3.11**	-1.84*	2.95**	4.14***
10	11:54 May 13, 2021	Energy usage trend over past few months (<i>picture for Bitcoin</i>)	-0.12	0.98	3.87***	0.46	6.45***	4.30***
11	16:42 May 19, 2021	Tesla has [diamond] [hands]	1.83*	2.12	1.31	2.69**	2.35*	6.90***
12	21:42 May 24, 2021	Spoke with North American Bitcoin miners	0.74	0.97	0.49	0.42	3.23**	3.60***
13	03:07 Jun 04, 2021	#Bitcoin [brokenheart] (<i>picture of a couple's conversation</i>)	-1.50	0.39	1.42	-1.58	3.46***	5.88***
14	04:10 Jun 25, 2021	How many Bitcoin maxis does it take to screw in a lightbulb?	-0.21	0.57	0.96	0.28	4.75***	3.06**

Note: The table includes mean-based t -statistics of abnormal returns in 1- and 2-hour window after 14 Bitcoin-related tweets by Elon Musk studied in [Ante \(2023\)](#). For each window, associated statistic of testing whether there is significant change in volatile level of price with the rest windows in the same trading day, computed using median squared return $q_j(0.5)$ (resp. interquartile range $q_j(0.75) - q_j(0.25)$) is reported in the second (resp. third) column, where *, **, *** indicate significance at 10%, 5% and 1% level.

4.5 Concluding Remarks

We introduce a valid methodology for conducting inference on a general continuous-time state-space model over a fixed time span. Through the inclusion of a residual term, we allow the model to be “approximately Markovian.” Notably, this model accommodates Lévy-driven returns and Cox trading flow processes. We allow for undefined dynamics in state processes, and propose uniform inference procedure for both entire conditional mean processes and entire conditional quantile processes of transformed states.

To construct functional estimators for the investigated processes, we gather all spot estimates with the local block size that shrinks to zero. The challenge of conducting uniform inference for these functional estimators arises from their non-Donsker nature. To address this, we establish Gaussian strong approximation, enabling valid uniform inference. These results can also be applied to tackle other econometric problems, such as constructing confidence sets for the ranks of spot values of studied processes.

We apply the proposed inference procedure to analyze trading flow processes and detect informative sentences from the FOMC press conference speeches. Our method allows for a comparison of trading intensity at a one-second level, enabling precise identification of speech segments containing valuable information. This inference procedure complements existing methodologies, such as volatility-based detection mechanisms and traditional textual analysis tools. Additionally, we apply this procedure to assess the impact of Elon Musk’s tweets on cryptocurrency markets, a scenario where mean-based tests might falter due to heavy-tailed returns. Results obtained through quantile-based measurements of volatility levels indicate a substantial price impact over an extended time window following tweet postings.

Chapter 5

Conclusion

This study contributes to the estimation and inference theory for high-frequency financial data. We have explored two types of inference procedures: spot and uniform inference. Our proposed methodologies are versatile, applicable not only to price data but also to other market indicators, rendering them to be practically useful. Extensive Monte Carlo experiments have been conducted to show the efficacy of the proposed estimators and tests. In empirical applications, these methods have been employed to address various issues, showing their ability of detecting real-time market activities and predicting the occurrence of new or information shocks.

In future research, there could be some interesting extension of the current research. Firstly, given the demonstrated superiority of candlestick data over returns in reflecting price volatility levels, there is potential to explore the use of such candlestick data for analyzing covariance between different assets. While practitioners commonly employ such tools under the banner of technical analysis, its foundation remains largely based on empirical observation rather than theoretical justification. Secondly, armed with established spot estimators proposed in this dissertation, one can employ them to conduct a uniform inference, and a more efficient inference of integrated volatility functionals. Finally, note that the conventional limiting experiments theory developed by Le Cam is not tailored for random parameters, whereas in many financial models, parameters such as volatility are inherently stochastic processes. While not directly applicable, statistical decision theory could be invoked in a different way—through a “coupling” argument, which is feasible for numer-

ous high-frequency models. Hence, the decision-theoretic approach used in the first two papers could find application in other contexts within or even beyond financial econometrics.

Appendix A

Technical Results for Chapter 2

A.1 Proofs of the Main Results

A.1.1 Proof of Theorem 2.1

In this proof, we focus on a specific time point, denoted as t , and examine i_n such that $i_n \Delta_n = t + o(1)$. To simplify our notation, we write i instead of i_n . We employ a generic constant $K > 0$, which may vary throughout the proof. Relying on a standard localization technique, we can strengthen Assumption 1 by assuming that the boundedness conditions hold uniformly over the whole sample. For more details on the localization method, refer to Section 4.4.1 in [Jacod and Protter \(2012\)](#).

Under Assumption 1(i), the probability of the interval \mathcal{T}_i containing at least one price jump is $O(\Delta_n)$. Consequently, price jumps occur in \mathcal{T}_i with asymptotically negligible probability. As our analysis focuses on this particular interval, we can assume without loss of generality that there are no jumps.

Denote $r'_i \equiv \sigma_{(i-1)\Delta_n} \zeta_{i,r}$, $u'_i \equiv \sigma_{(i-1)\Delta_n} \zeta_{i,u}$, and $l'_i \equiv \sigma_{(i-1)\Delta_n} \zeta_{i,l}$. Since there is no jump within the \mathcal{T}_i interval, we can rewrite (r_i, u_i, l_i) as

$$\begin{aligned} r_i &= \Delta_n^{-1/2} \left(\int_{(i-1)\Delta_n}^{i\Delta_n} b_s ds + \int_{(i-1)\Delta_n}^{i\Delta_n} \sigma_s dW_s \right), \\ u_i &= \Delta_n^{-1/2} \sup_{s \in \mathcal{T}_i} \left(\int_{(i-1)\Delta_n}^s b_u du + \int_{(i-1)\Delta_n}^s \sigma_u dW_u \right), \\ l_i &= \Delta_n^{-1/2} \inf_{s \in \mathcal{T}_i} \left(\int_{(i-1)\Delta_n}^s b_u du + \int_{(i-1)\Delta_n}^s \sigma_u dW_u \right). \end{aligned}$$

Under Assumption 1(i), it is easy to see that

$$\left| \int_{(i-1)\Delta_n}^{i\Delta_n} b_s ds \right| \leq \int_{(i-1)\Delta_n}^{i\Delta_n} |b_s| ds = O_p(\Delta_n). \quad (\text{A.1})$$

Moreover, by the Burkholder–David–Gundy inequality and Assumption 1(ii), we have

$$\begin{aligned} \mathbb{E} \left[\sup_{s \in \mathcal{T}_i} \left| \int_{(i-1)\Delta_n}^s (\sigma_u - \sigma_{(i-1)\Delta_n}) dW_u \right|^2 \right] &\leq K \mathbb{E} \left[\int_{(i-1)\Delta_n}^{i\Delta_n} |\sigma_u - \sigma_{(i-1)\Delta_n}|^2 du \right] \\ &\leq K \Delta_n^{1+2\kappa}, \end{aligned}$$

and hence,

$$\sup_{s \in \mathcal{T}_i} \left| \int_{(i-1)\Delta_n}^s (\sigma_u - \sigma_{(i-1)\Delta_n}) dW_u \right| = O_p(\Delta_n^{1/2+\kappa}). \quad (\text{A.2})$$

By the triangle inequality, (A.1), and (A.2),

$$\begin{aligned} |r_i - r'_i| &\leq \Delta_n^{-1/2} \left| \int_{(i-1)\Delta_n}^{i\Delta_n} b_s ds \right| + \Delta_n^{-1/2} \left| \int_{(i-1)\Delta_n}^{i\Delta_n} (\sigma_s - \sigma_{(i-1)\Delta_n}) dW_s \right| \\ &= O_p(\Delta_n^{(1/2)\wedge\kappa}). \end{aligned} \quad (\text{A.3})$$

In addition, we note that

$$\begin{aligned} |u_i - u'_i| &= \Delta_n^{-1/2} \left| \sup_{s \in \mathcal{T}_i} \left(\int_{(i-1)\Delta_n}^s b_u du + \int_{(i-1)\Delta_n}^s \sigma_u dW_u \right) \right. \\ &\quad \left. - \sigma_{(i-1)\Delta_n} \sup_{s \in \mathcal{T}_i} (W_s - W_{(i-1)\Delta_n}) \right| \\ &\leq \Delta_n^{-1/2} \left(\int_{(i-1)\Delta_n}^{i\Delta_n} |b_u| du + \sup_{s \in \mathcal{T}_i} \left| \int_{(i-1)\Delta_n}^s (\sigma_u - \sigma_{(i-1)\Delta_n}) dW_u \right| \right) \\ &= O_p(\Delta_n^{(1/2)\wedge\kappa}), \end{aligned} \quad (\text{A.4})$$

where the last line follows from (A.1) and (A.2). Similarly, we can derive

$$|l_i - l'_i| = O_p(\Delta_n^{(1/2)\wedge\kappa}). \quad (\text{A.5})$$

Since $f(\cdot)$ is continuous a.e., the estimates from (A.3)–(A.5) imply that

$$f(r_i, u_i, l_i) = f(r'_i, u'_i, l'_i) + o_p(1).$$

Since σ is bounded away from zero under Assumption 1(i), we further have

$$\frac{f(r_i, u_i, l_i)}{\sigma_{(i-1)\Delta_n}^p} = \frac{f(r'_i, u'_i, l'_i)}{\sigma_{(i-1)\Delta_n}^p} + o_p(1) = f(\zeta_i) + o_p(1), \quad (\text{A.6})$$

where the second equality follows from the homogeneity of $f(\cdot)$ and the definition of ζ_i . By Assumption 1(ii), $|\sigma_t - \sigma_{(i-1)\Delta_n}| = O_p(|t - i\Delta_n|^\kappa) = o_p(1)$ as $|i\Delta_n - t| \rightarrow 0$, which together with (A.6) implies the assertion of Theorem 2.1. $Q.E.D.$

A.1.2 Proof of Theorem 2.2

To prove Theorem 2.2, we first prove two lemmas. Lemma A.1 shows the sufficiency of the shape features for volatility estimation under the limit model. Lemma A.2 derives a closed-form expression for the conditional expectation of certain Brownian functionals.

Lemma A.1 (Sufficient Statistic of Candlestick Data). *Under the limit model $P_t = \sigma W_t$, the collection of shape features $(|r_i|, w_i, a_i)$ is a sufficient statistic for σ given the observation (r_i, u_i, l_i) .*

PROOF OF LEMMA A.1. Recall that $(B_t)_{t \in [0,1]}$ is a standard Brownian motion with $B_0 = 0$. Let $g(r, u, l)$ denote the probability density function of

$$\left(B_1, \sup_{0 \leq t \leq 1} B_t, \inf_{0 \leq t \leq 1} B_t \right).$$

The density of (r_i, u_i, l_i) is then given by the function

$$(r, u, l) \mapsto \frac{1}{\sigma^3} g\left(\frac{r}{\sigma}, \frac{u}{\sigma}, \frac{l}{\sigma}\right). \quad (\text{A.7})$$

By equation (1.15.8) in Borodin and Salminen (2002),

$$\begin{aligned} & \mathbb{P}\left(l < \inf_{0 \leq t \leq 1} B_t, \sup_{0 \leq t \leq 1} B_t < u, B_1 \in dr\right) \\ &= \frac{1}{\sqrt{2\pi}} \sum_{k=-\infty}^{\infty} \left\{ \exp\left(-\frac{(2k(u-l)+r)^2}{2}\right) - \exp\left(-\frac{(2k(u-l)+r-2l)^2}{2}\right) \right\} dr. \end{aligned} \quad (\text{A.8})$$

The function $g(r, u, l)$ is thus proportional to $\sum_{k=-\infty}^{\infty} (A_k(r, u, l) - B_k(r, u, l))$, where

$$\begin{aligned} A_k(r, u, l) &\equiv k^2 \{(2k(u-l)+r)^2 - 1\} \exp\left(-\frac{(2k(u-l)+r)^2}{2}\right), \\ B_k(r, u, l) &\equiv k(k+1) \{(2k(u-l)+r-2l)^2 - 1\} \exp\left(-\frac{(2k(u-l)+r-2l)^2}{2}\right). \end{aligned}$$

By a change of variable via $w = u - l$ and $d = u + l - r$, we may identify these functions with

$$\begin{aligned} \tilde{A}_k(r, w, d) &\equiv k^2 \{(2kw+r)^2 - 1\} \exp\left(-\frac{(2kw+r)^2}{2}\right), \\ \tilde{B}_k(r, w, d) &\equiv k(k+1) \{(2kw+w-d)^2 - 1\} \exp\left(-\frac{(2kw+w-d)^2}{2}\right). \end{aligned}$$

Note that for each $k \geq 0$, we can verify that $\tilde{A}_k(-r, w, d) = \tilde{A}_{-k}(r, w, d)$ and $\tilde{B}_k(r, w, -d) = \tilde{B}_{-k-1}(r, w, d)$. Thus, $g(r, u, l)$ depends on (r, u, l) only through $(|r|, w, |d|)$. The assertion of the lemma then follows from the Fisher–Neyman factorization theorem. *Q.E.D.*

Lemma A.2 (Analytical Conditional Expectation). *Let B be a standard Brownian motion on the unit interval with $B_0 = 0$ and*

$$\xi_1 \equiv \sup_t B_t - \inf_t B_t, \quad \xi_2 \equiv \frac{|\sup_t B_t + \inf_t B_t - B_1|}{\sup_t B_t - \inf_t B_t}, \quad \xi_3 \equiv \frac{|B_1|}{\sup_t B_t - \inf_t B_t},$$

where \sup_t and \inf_t are taken over $[0, 1]$. Then for each integer $q \geq 1$ we have

$$\mathbb{E}[\xi_1^q | \xi_2, \xi_3] = (-1)^q \frac{(q+2)}{\sqrt{2^q \pi} q!} \Gamma\left(\frac{q+3}{2}\right) \frac{G_q(\xi_2) - H_q(\xi_3)}{G_0(\xi_2) - H_0(\xi_3)},$$

where $G_q(\cdot)$ and $H_q(\cdot)$ are defined as in (2.10) and (2.11).

PROOF OF LEMMA A.2. Let $g_\xi(\cdot)$ denote the joint density of (ξ_1, ξ_2, ξ_3) . The conditional expectation of interest can then be written as

$$\mathbb{E}[\xi_1^q | \xi_2 = y, \xi_3 = z] = \frac{\int_0^\infty x^q g_\xi(x, y, z) dx}{\int_0^\infty g_\xi(x, y, z) dx}.$$

The main task is to calculate the numerator $\int_0^\infty x^q g_\xi(x, y, z) dx$ for $q \geq 1$ and the denominator $\int_0^\infty g_\xi(x, y, z) dx$. (The calculation for the latter is not a special case of the former by simply setting $p = 0$, as it requires a slightly more refined technical argument due to the lack of convergence of certain series.)

We first calculate $\int_0^\infty x^q g_\xi(x, y, z) dx$. From (A.8) and the definition of (ξ_1, ξ_2, ξ_3) , we obtain $g_\xi(x, y, z) = 4\sqrt{2/\pi} \sum_{k=-\infty}^\infty (C_k(x, z) - D_k(x, y))$, where

$$C_k(x, z) \equiv k^2 x^2 ((2k+z)^2 x^2 - 1) \exp\left(-\frac{(2k+z)^2 x^2}{2}\right),$$

$$D_k(x, y) \equiv k(1+k)x^2 ((2k+1-y)^2 x^2 - 1) \exp\left(-\frac{(2k+1-y)^2 x^2}{2}\right).$$

Since $z \in [0, 1]$, for $q \geq 1$, by a direct integration, we have

$$\int_0^\infty \sum_{k=-\infty}^\infty x^q C_k(x, z) dx = M_q \cdot \sum_{k=-\infty}^\infty \frac{k^2}{|\frac{z}{2} + k|^{q+3}}, \quad (\text{A.9})$$

where we denote $M_r \equiv 2^{-(r+5)/2} (r+2) \Gamma((r+3)/2)$ for any $r \geq 0$. (Note that the convergence of the above series requires the integer $q \geq 1$.) We now express (A.9)

using polygamma functions. By (5.15.1) in [Olver et al. \(2010\)](#), for $r \geq 1$ we have

$$\psi_r\left(\frac{z}{2}\right) = \sum_{k=0}^{\infty} \frac{(-1)^{r+1} r!}{\left(\frac{z}{2} + k\right)^{r+1}}. \quad (\text{A.10})$$

Note that when $k \geq 0$, the summand in (A.9) may be rewritten in the form of the summand in (A.10) as

$$\begin{aligned} \frac{k^2}{\left(\frac{z}{2} + k\right)^{q+3}} &= (-1)^{q+1} \left(\frac{1}{q!} \cdot \frac{(-1)^{q+1} q!}{\left(\frac{z}{2} + k\right)^{q+1}} + \frac{1}{(q+1)!} \cdot z \frac{(-1)^{q+2} (q+1)!}{\left(\frac{z}{2} + k\right)^{q+2}} \right. \\ &\quad \left. + \frac{1}{4(q+2)!} \cdot z^2 \frac{(-1)^{q+3} (q+2)!}{\left(\frac{z}{2} + k\right)^{q+3}} \right). \end{aligned} \quad (\text{A.11})$$

Combining (A.9)–(A.11) yields

$$\begin{aligned} &\sum_{k=0}^{\infty} \int_0^{\infty} x^q C_k(x, z) dx \\ &= \frac{(-1)^{q+1} M_q}{q!} \left(\psi_q\left(\frac{z}{2}\right) + \frac{1}{q+1} z \psi_{q+1}\left(\frac{z}{2}\right) + \frac{1}{4(q+1)(q+2)} z^2 \psi_{q+2}\left(\frac{z}{2}\right) \right). \end{aligned}$$

The summation in (A.9) over $k < 0$ can be rewritten, with a change of variable $m = -k - 1$, as

$$\sum_{k=-\infty}^{-1} \frac{k^2}{\left(-\frac{z}{2} - k\right)^{q+3}} = \sum_{m=0}^{\infty} \frac{(m+1)^2}{\left(1 - \frac{z}{2} + m\right)^{q+3}}.$$

Using an argument similar to (A.11), we also have

$$\begin{aligned} &\sum_{k=-\infty}^{-1} \int_0^{\infty} x^q C_k(x, z) dx \\ &= \frac{(-1)^{q+1} M_q}{q!} \left(\psi_q\left(1 - \frac{z}{2}\right) - \frac{1}{q+1} z \psi_{q+1}\left(1 - \frac{z}{2}\right) \right. \\ &\quad \left. + \frac{1}{4(q+1)(q+2)} z^2 \psi_{q+2}\left(1 - \frac{z}{2}\right) \right). \end{aligned}$$

Combining the above results for $k \geq 0$ and $k < 0$ and recalling the definition of $H_q(\cdot)$, we obtain

$$\int_0^{\infty} \sum_{k=-\infty}^{\infty} x^q C_k(x, z) dx = \frac{(-1)^{q+1} M_q}{q!} H_q(z). \quad (\text{A.12})$$

By a similar argument leading to (A.12), we can also show that

$$\int_0^{\infty} \sum_{k=-\infty}^{\infty} x^q D_k(x, y) dx = \frac{(-1)^{q+1} M_q}{q!} G_q(y).$$

Hence, for $q \geq 1$,

$$\int_0^{\infty} x^q g_{\xi}(x, y, z) dx = 4 \sqrt{\frac{2}{\pi}} \frac{(-1)^q M_q}{q!} (G_q(y) - H_q(z)). \quad (\text{A.13})$$

For the denominator $\int_0^\infty g_\xi(x, y, z)dx$, by a direct integration, we have

$$\int_0^\infty \sum_{k=-\infty}^{\infty} (C_k(x, z) - D_k(x, y))dx = \frac{\sqrt{2\pi}}{8} \sum_{k=-\infty}^{\infty} \left(\frac{k^2}{|\frac{z}{2} + k|^3} - \frac{k(k+1)}{|\frac{1-y}{2} + k|^3} \right). \quad (\text{A.14})$$

By (5.7.6) in [Olver et al. \(2010\)](#), we obtain

$$\psi_0\left(\frac{1-y}{2}\right) - \psi_0\left(\frac{z}{2}\right) = \sum_{k=0}^{\infty} \left(\frac{1}{\frac{z}{2} + k} - \frac{1}{\frac{1-y}{2} + k} \right). \quad (\text{A.15})$$

Note that when $k \geq 0$, the summand in (A.14) may be rewritten in the form of the summand in (A.10) and (A.15) as

$$\begin{aligned} \frac{k^2}{(\frac{z}{2} + k)^3} - \frac{k(k+1)}{(\frac{1-y}{2} + k)^3} &= \left(\frac{1}{\frac{z}{2} + k} - \frac{1}{\frac{1-y}{2} + k} \right) - z \frac{1}{(\frac{z}{2} + k)^2} - y \frac{1}{(\frac{1-y}{2} + k)^2} \\ &\quad - \frac{1}{8} z^2 \frac{-2}{(\frac{z}{2} + k)^3} - \frac{1}{8} (1-y)^2 \frac{-2}{(\frac{1-y}{2} + k)^3}. \end{aligned} \quad (\text{A.16})$$

Combining (A.10) and (A.14)–(A.16) yields

$$\begin{aligned} &\sum_{k=0}^{\infty} \int_0^\infty (C_k(x, z) - D_k(x, z))dx \\ &= \frac{\sqrt{2\pi}}{8} \left(\psi_0\left(\frac{1-y}{2}\right) - \psi_0\left(\frac{z}{2}\right) - z \psi_1\left(\frac{z}{2}\right) - y \psi_1\left(\frac{1-y}{2}\right) \right. \\ &\quad \left. - \frac{1}{8} z^2 \psi_2\left(\frac{z}{2}\right) - \frac{1}{8} (1-y)^2 \psi_2\left(\frac{1-y}{2}\right) \right). \end{aligned}$$

The summation in (A.14) over $k < 0$ can be rewritten, with a change of variable $m = -k - 1$, as

$$\sum_{k=-\infty}^{-1} \left(\frac{k^2}{(-\frac{z}{2} - k)^3} - \frac{k(k+1)}{(-\frac{1-y}{2} - k)^3} \right) = \sum_{m=0}^{\infty} \left(\frac{(m+1)^2}{(1-\frac{z}{2} + m)^3} - \frac{m(m+1)}{(\frac{1+y}{2} + m)^3} \right).$$

Using an argument similar to (A.16), we also have

$$\begin{aligned} &\sum_{k=-\infty}^{-1} \int_0^\infty (C_k(x, z) - D_k(x, z))dx \\ &= \frac{\sqrt{2\pi}}{8} \left(\psi_0\left(\frac{1+y}{2}\right) - \psi_0\left(1-\frac{z}{2}\right) + z \psi_1\left(1-\frac{z}{2}\right) + y \psi_1\left(\frac{1+y}{2}\right) \right. \\ &\quad \left. - \frac{1}{8} z^2 \psi_2\left(1-\frac{z}{2}\right) - \frac{1}{8} (1-y)^2 \psi_2\left(\frac{1+y}{2}\right) \right). \end{aligned}$$

Combining the above results for $k \geq 0$ and $k < 0$ and recalling the definition of $G_0(\cdot)$, $H_0(\cdot)$, and M_0 , we obtain

$$\int_0^\infty g_\xi(x, y, z)dx = 4\sqrt{\frac{2}{\pi}} M_0(G_0(y) - H_0(z)). \quad (\text{A.17})$$

The assertion of the lemma then readily follows from (A.13), (A.17), and the fact that

$$\frac{M_q}{M_0} = \frac{(q+2)}{\sqrt{2^q \pi}} \Gamma\left(\frac{q+3}{2}\right). \quad Q.E.D.$$

PROOF OF THEOREM 2.2. We first consider the case with Stein's loss. Recall that the asymptotic risk $\mathbb{E}[L(f(\tilde{\xi}))]$ equals the finite-sample risk of the estimator $f(r_i, u_i, l_i)$ under the limit model $P_t = \sigma W_t$. Therefore, minimizing the asymptotic risk is equivalent to finding the minimum-risk scale-equivariant estimator for σ under the limit model. By Lemma A.1 and the Rao–Blackwell theorem, this optimal estimator only depends on the shape features $(|r_i|, w_i, a_i)$. Note that $(w_i, a_i/w_i, |r_i|/w_i)$ has the same distribution as (ξ_1, ξ_2, ξ_3) defined in Lemma A.2. With an appeal to Corollary 3.3.8 in Lehmann and Casella (1998), the minimum-risk scale-equivariant estimation function under the limit problem is given by

$$w_i^p / \mathbb{E}[\xi_1^p \mid \xi_2 = a_i/w_i, \xi_3 = |r_i|/w_i].$$

For estimating spot variance, taking $p = 2$ and applying Lemma A.2 with $q = 2$, we may rewrite this function in closed form as

$$\frac{4w_i^2}{3} \cdot \frac{G_0(a_i/w_i) - H_0(|r_i|/w_i)}{G_2(a_i/w_i) - H_2(|r_i|/w_i)}.$$

Recalling the meromorphic property of the polygamma functions, we see that this estimation function is continuous almost everywhere. This estimator is thus regular, and so, is also the AMRE estimator under the original nonparametric model as asserted in Theorem 2.2.

The proof for the quadratic loss is similar, except that we now apply (3.3.18) in Lehmann and Casella (1998) and Lemma A.2 above with $p = 2$, $q = 2$ and 4 to show that the optimal estimation function of σ^2 under the limit model is

$$w_i^2 \cdot \frac{\mathbb{E}[\xi_1^2 \mid \xi_2 = a_i/w_i, \xi_3 = |r_i|/w_i]}{\mathbb{E}[\xi_1^4 \mid \xi_2 = a_i/w_i, \xi_3 = |r_i|/w_i]} = \frac{32w_i^2}{5} \cdot \frac{G_2(a_i/w_i) - H_2(|r_i|/w_i)}{G_4(a_i/w_i) - H_4(|r_i|/w_i)}.$$

This estimator is also regular and thus is the AMRE estimator under the original nonparametric model as asserted.

The derivation of AMRE estimators of spot volatility follows the same lines of arguments, except for taking $p = 1$, $q = 1$ and 2. Q.E.D.

A.2 Additional Results

A.2.1 Optimal Estimation for Quarticity and Precision

In Theorem A.1 and Theorem A.2, we provide closed-form expressions for the AMRE estimators of the spot quarticity, denoted as σ_t^4 , and the spot precision, denoted as σ_t^{-1} , respectively. However, due to a technical integrability issue, we have only derived the closed-form solution for the AMRE spot precision estimator under Stein's loss.

Theorem A.1. *Under the same setting as Theorem 1, the AMRE range-based estimator for the spot quarticity σ_t^4 under Stein's loss is asymptotically unbiased and is given by*

$$\hat{\sigma}_{\text{Stein}}^4 \equiv \frac{128w_i^4}{15} \cdot \frac{G_0(a_i/w_i) - H_0(|r_i|/w_i)}{G_4(a_i/w_i) - H_4(|r_i|/w_i)},$$

while the AMRE range-based estimator for σ_t^4 under standardized quadratic loss equals

$$\hat{\sigma}_{\text{Quad}}^4 \equiv 256w_i^4 \cdot \frac{G_4(a_i/w_i) - H_4(|r_i|/w_i)}{G_8(a_i/w_i) - H_8(|r_i|/w_i)}.$$

Theorem A.2. *Under the same setting as Theorem 1, the AMRE range-based estimator for the spot precision σ_t^{-1} under Stein's loss is asymptotically unbiased and is given by*

$$\hat{\sigma}_{\text{Stein}}^{-1} \equiv \sqrt{\frac{\pi}{2}} w_i^{-1} \cdot \frac{G_0(a_i/w_i) - H_0(|r_i|/w_i)}{G_{-1}(a_i/w_i) - H_{-1}(|r_i|/w_i)},$$

where $G_0(\cdot)$ and $H_0(\cdot)$ are defined in (2.10) and (2.11) in the main text and

$$G_{-1}(x) \equiv x \left(\psi_0\left(\frac{1-x}{2}\right) - \psi_0\left(\frac{1+x}{2}\right) \right) + \frac{1-x^2}{4} \left(\psi_1\left(\frac{1-x}{2}\right) + \psi_1\left(\frac{1+x}{2}\right) \right),$$

$$H_{-1}(x) \equiv x \left(\psi_0\left(1 - \frac{x}{2}\right) - \psi_0\left(\frac{x}{2}\right) \right) - \frac{x^2}{4} \left(\psi_1\left(1 - \frac{x}{2}\right) + \psi_1\left(\frac{x}{2}\right) \right).$$

In parallel to Tables 1 and 2 for the spot variance and spot volatility estimators in the main text, Tables A.1 and A.2 below display the asymptotic bias, variance, and relative efficiency of various estimators for spot quarticity and precision, respectively. The results are qualitatively similar to the ones presented in the main text.

We once again observe that Stein-AMRE estimators demonstrate lower asymptotic risk compared to shape-constrained estimators, not only under Stein's loss

but also under the quadratic loss. Interestingly, the relative efficiencies of shape-constrained estimators are considerably lower when estimating quarticity and precision as opposed to volatility estimation. For instance, the relative efficiency of $(\hat{\sigma}_{\text{BLUE}})^4$ as a spot quarticity estimator is a mere 25.5% under the quadratic loss function. These risk comparisons further emphasize the efficiency advantages of employing AMRE estimators for a wider variety of volatility estimation tasks.

Table A.1: Asymptotic Risk Properties of Alternative Estimators for Spot Quaticity

Estimator	Bias	Variance	Relative Efficiency	
			Stein	Quadratic
$\hat{\sigma}_{\text{Stein}}^4$	0.000	1.305	1.000	0.544
$\hat{\sigma}_{\text{Quad}}^4$	-0.564	0.392	0.535	1.000
$(\hat{\sigma}_{\text{GK}}^2)^2$	0.271	2.206	0.921	0.312
$(\hat{\sigma}_{\text{BLUE}})^4$	0.423	2.608	0.874	0.255

Note: The table reports the asymptotic biases, variances, and relative efficiency under Stein's and quadratic risks for each of the regular estimators indicated in the first column.

Table A.2: Asymptotic Risk Properties of Alternative Estimators for Spot Precision

Estimator	Bias	Variance	Relative Efficiency	
			Stein	Quadratic
$\hat{\sigma}_{\text{Stein}}^{-1}$	0.000	0.066	1.000	1.000
$(\hat{\sigma}_{\text{GK}}^2)^{-1/2}$	0.099	0.081	0.838	0.725
$\hat{\sigma}_{\text{BLUE}}^{-1}$	0.065	0.076	0.912	0.825

Note: The table reports the asymptotic biases, variances, and relative efficiency under Stein's and quadratic risks for each of the regular estimators indicated in the first column. Since the quadratic-AMRE estimator is unavailable, the relative efficiencies of the spot precision estimators are calculated with respect to the Stein-AMRE estimator under both Stein's loss and the quadratic loss.

A.2.2 Proof of Theorem A.1

The proof is similar to that of Theorem 2 in the main text, and so, we only emphasize the main difference. Recall the definition of (ξ_1, ξ_2, ξ_3) from Lemma A2 in the main text. The minimum-risk scale-equivariant estimation function for volatility under the limit problem is given by

$$w_i^4 \cdot \frac{1}{\mathbb{E}[\xi_1^4 | \xi_2 = a_1/w_i, \xi_3 = |r_i|/w_i]} \quad \text{and} \quad w_i^4 \cdot \frac{\mathbb{E}[\xi_1^4 | \xi_2 = a_1/w_i, \xi_3 = |r_i|/w_i]}{\mathbb{E}[\xi_1^8 | \xi_2 = a_1/w_i, \xi_3 = |r_i|/w_i]},$$

for Stein's loss and quadratic loss, respectively. The asserted closed-form expressions are obtained by applying Lemma A2 with $q = 4$ and 8 . *Q.E.D.*

A.2.3 Proof of Theorem A.2

Under Stein's loss, the minimum-risk scale-equivariant estimation function for the precision in the limit problem is given by $w_i^{-1}/\mathbb{E}[\xi_1^{-1} | \xi_2 = a_1/w_i, \xi_3 = |r_i|/w_i]$. The remaining task is to derive its closed-form expression, which requires a more refined argument than those used in the analysis for optimal variance and/or volatility estimation.

Let $g_\xi(\cdot)$ denote the joint density of (ξ_1, ξ_2, ξ_3) . Note that

$$\mathbb{E}[\xi_1^{-1} | \xi_2 = y, \xi_3 = z] = \frac{\int_0^\infty x^{-1} g_\xi(x, y, z) dx}{\int_0^\infty g_\xi(x, y, z) dx}, \quad (\text{A.18})$$

and the denominator $\int_0^\infty g_\xi(x, y, z) dx = G_0(y) - H_0(z)$ as implied by Lemma A2. It remains to compute $\int_0^\infty x^{-1} g_\xi(x, y, z) dx$. Recall that

$$g_\xi(x, y, z) = 4\sqrt{\frac{2}{\pi}} \sum_{k=-\infty}^{\infty} (C_k(x, z) - D_k(x, y)),$$

where

$$\begin{aligned} C_k(x, z) &\equiv k^2 x^2 ((2k+z)^2 x^2 - 1) \exp\left(-\frac{(2k+z)^2 x^2}{2}\right), \\ D_k(x, y) &\equiv k(1+k)x^2 ((2k+1-y)^2 x^2 - 1) \exp\left(-\frac{(2k+1-y)^2 x^2}{2}\right). \end{aligned}$$

By a direct integration, we have

$$\begin{aligned} \int_0^\infty x^{-1} C_k(x, z) dx &= \frac{k^2}{(2k+z)^2}, \\ \int_0^\infty x^{-1} D_k(x, y) dx &= \frac{k(k+1)}{(2k+1-y)^2}. \end{aligned}$$

Therefore, the numerator in (A.18) can be rewritten as

$$\begin{aligned}
& 4\sqrt{\frac{\pi}{2}} \int_0^\infty x^{-1} \sum_{k=-\infty}^{\infty} (C_k(x, z) - D_k(x, y)) dx \\
&= 4\sqrt{\frac{\pi}{2}} \sum_{k=-\infty}^{\infty} \left(\frac{k^2}{(2k+z)^2} - \frac{k(k+1)}{(2k+1-y)^2} \right) \\
&= \sqrt{\frac{\pi}{2}} \sum_{k=0}^{\infty} \left(\frac{k^2}{(k+\frac{z}{2})^2} + \frac{(k+1)^2}{(k+1-\frac{z}{2})^2} - \frac{k(k+1)}{(k+\frac{1-y}{2})^2} - \frac{k(k+1)}{(k+\frac{1+y}{2})^2} \right).
\end{aligned} \tag{A.19}$$

By (5.7.6) and (5.15.1) in Olver et al. (2010), we have

$$\psi_0(z_1) - \psi_0(z_2) = \sum_{k=0}^{\infty} \left(\frac{1}{z_2+k} - \frac{1}{z_1+k} \right), \tag{A.20}$$

$$\psi_1(z_1) + \psi_1(z_2) = \sum_{k=0}^{\infty} \left(\frac{1}{(z_1+k)^2} + \frac{1}{(z_2+k)^2} \right). \tag{A.21}$$

Note that, the summand in (A.19) may be rewritten as

$$\begin{aligned}
& \frac{k^2}{(k+\frac{z}{2})^2} + \frac{(k+1)^2}{(k+1-\frac{z}{2})^2} - \frac{k(k+1)}{(k+\frac{1-y}{2})^2} - \frac{k(k+1)}{(k+\frac{1+y}{2})^2} \\
&= y \left(\frac{1}{\frac{1+y}{2}+k} - \frac{1}{\frac{1-y}{2}+k} \right) + \frac{1}{4} \cdot (1-y^2) \left(\frac{1}{(\frac{1+y}{2}+k)^2} + \frac{1}{(\frac{1-y}{2}+k)^2} \right) \\
&\quad - z \left(\frac{1}{\frac{z}{2}+k} - \frac{1}{1-\frac{z}{2}+k} \right) + \frac{1}{4} \cdot z^2 \left(\frac{1}{(\frac{z}{2}+k)^2} + \frac{1}{(1-\frac{z}{2}+k)^2} \right).
\end{aligned} \tag{A.22}$$

Combining (A.19)–(A.22) yields

$$4\sqrt{\frac{\pi}{2}} \int_0^\infty x^{-1} \sum_{k=-\infty}^{\infty} (C_k(x, z) - D_k(x, y)) dx = \sqrt{\frac{\pi}{2}} (G_{-1}(y) - H_{-1}(z)).$$

This completes the derivation of the closed-form expression for (A.18). Since that function is continuous almost everywhere (recall that polygamma functions are meromorphic), the estimator is indeed regular, and so, defines the AMRE estimator under the original nonparametric model as asserted in Theorem A.2. *Q.E.D.*

A.2.4 Comparisons of Single Versus Multiple Candlesticks

The spot volatility estimates for the VOO ETF plotted in Figure 2.4 in the main text are formed using the single-candlestick estimator $\hat{\sigma}_{\text{Stein}}$ described in Theorem 2.2. To illustrate the practical use of the AMRE k -candlestick estimators detailed in Section 2.2.4, we also compute the AMRE spot volatility estimator $\hat{\sigma}_{\text{Stein}}(k)$ for

$k = 2$. Figure A.1 visually compares the single-candlestick estimates with the two-candlestick estimates.

Unsurprisingly, the two-candlestick estimates essentially manifest as “averages” of the corresponding pair of single-candlestick estimates. As discussed in more detail in the main text, such aggregation is suitable when the underlying volatility does not change “too much,” but it can be problematic when volatility moves rapidly. This, of course, is exactly what happens after many FOMC announcements, as highlighted, for example, by the November 2, 2022 announcement.

Figure A.2 presents similar results for the Dollar/Yen exchange rate. The comparisons between the single and two-candlestick estimators evidence the same general features as the comparisons of the $k = 1$ and $k = 2$ estimates for the VOO ETF in Figure A.1. Meanwhile, comparing the estimates across Figure A.1 and Figure A.2, there are also notable differences between the way in which the spot volatility of the VOO ETF and the Dollar/Yen exchange rate vary around the time of the FOMC announcements, reflecting the differential response of the equity and currency markets to monetary policy shocks.

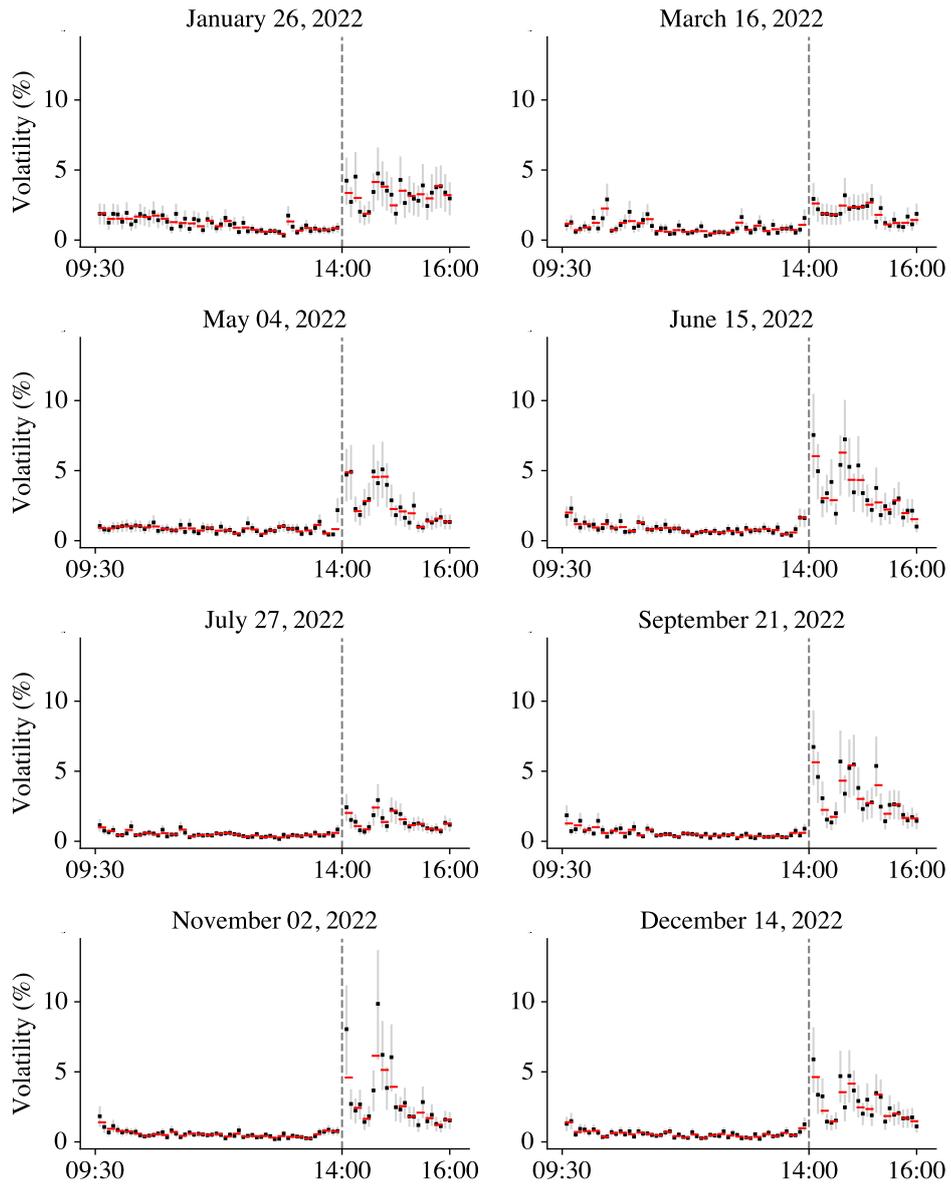


Figure A.1: Comparison of Single-Candlestick and Two-Candlestick Estimates. The figure plots the estimates of the AMRE spot volatility estimator $\hat{\sigma}_{\text{Stein}}$ constructed using one candlestick (dot) and two candlesticks (dash), expressed in daily percentage terms and calculated using individual 5-minute frequency candlesticks of the VOO ETF. Pointwise confidence intervals at the 90% level for the single-candlestick estimates are also plotted.

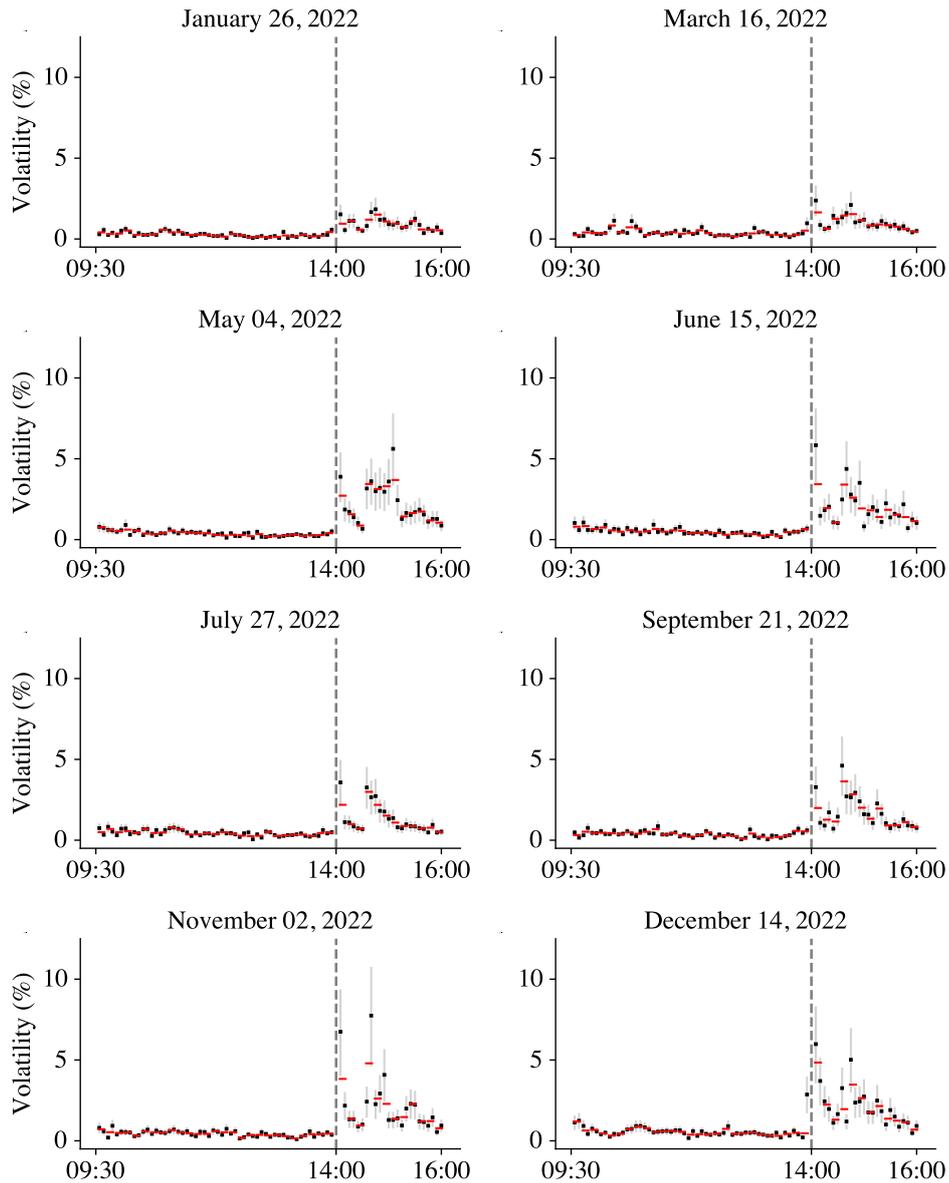


Figure A.2: Comparison of Single-Candlestick and Two-Candlestick Estimates. The figure plots the estimates of the AMRE spot volatility estimator $\hat{\sigma}_{\text{Stein}}$ constructed using one candlestick (dot) and two candlesticks (dash), expressed in daily percentage terms and calculated using individual 5-minute frequency candlesticks of the Dollar/Yen exchange rate. Pointwise confidence intervals at the 90% level for the single-candlestick estimates are also plotted.

A.2.5 Comparison Among Alternative Estimators

The empirical estimates in the main text, as well as those in Section A.2.4 above, are based on AMRE estimators for the spot volatility σ_t derived under Stein’s loss. The use of alternative estimators would yield different estimates. It is helpful to understand the magnitude of these differences.

To this end, we report summary statistics for the relative discrepancy between alternative estimators of σ_t^p for $p \in \{1, 2, 4\}$. For each σ_t^p , we consider four estimators: the AMRE estimators $\hat{\sigma}_{\text{Stein}}^p$ and $\hat{\sigma}_{\text{Quad}}^p$, a transformed version of the Garman–Klass variance estimator $(\hat{\sigma}_{\text{GK}}^2)^{p/2}$, and a transformed version of Li et al.’s (2022) volatility estimator $(\hat{\sigma}_{\text{BLUE}})^p$. Table A.3 presents the results for alternative single-candlestick estimators with respect to the AMRE estimator $\hat{\sigma}_{\text{Stein}}^p$ (Panel A) or $\hat{\sigma}_{\text{Quad}}^p$ (Panel B). Specifically, when comparing $(\hat{\sigma}_{\text{GK}}^2)^{p/2}$ with $\hat{\sigma}_{\text{Stein}}^p$ in Panel A, we calculate the relative discrepancy measure as

$$\sqrt{\text{sample average of } \left| \frac{(\hat{\sigma}_{\text{GK}}^2)^{p/2} - \hat{\sigma}_{\text{Stein}}^p}{\hat{\sigma}_{\text{Stein}}^p} \right|^2},$$

where the sample average is computed across all spot estimates for the eight FOMC announcement days in 2022. We do the calculations separately for the VOO ETF and the Dollar/Yen exchange rate. Similar results for estimators based on two candlesticks are reported in Table A.4.

The main findings may be summarized as follows. Firstly, we note that alternative estimates for σ_t^p exhibit larger differences when p is larger. For example, for volatility estimation (i.e., $p = 1$), the linear $\hat{\sigma}_{\text{BLUE}}$ estimator differs from the Stein-AMRE estimator $\hat{\sigma}_{\text{Stein}}$ by 2.5% and 4.4% for the VOO ETF and the Dollar/Yen, respectively. The corresponding relative discrepancy measures increase to 8.1% and 12.2% for variance estimation (i.e., $p = 2$), and further increase to 44.5% and 53.7% for quarticity estimation (i.e., $p = 4$).

Secondly, comparing the discrepancy numbers between Panel A and Panel B shows that the shape-constrained estimators, $(\hat{\sigma}_{\text{GK}}^2)^{p/2}$ and $(\hat{\sigma}_{\text{BLUE}})^p$, are more distinct from $\hat{\sigma}_{\text{Quad}}^p$ than $\hat{\sigma}_{\text{Stein}}^p$. This is also consistent with Figures 2 and 3 in the main text, which show the alternative estimation functions. This suggests that under the quadratic loss criterion, suboptimal estimators can be quite different from

the AMRE estimator. For example, $\hat{\sigma}_{\text{BLUE}}$ differs from $\hat{\sigma}_{\text{Quad}}$ by 6.6% (resp. 8.1%) for VOO (resp. Dollar/Yen), while the relative differences increase to more than 30% for variance estimation. For the more extreme case of estimating quarticity σ_t^4 , the AMRE estimator $\hat{\sigma}_{\text{Quad}}^4$ is generally much smaller (due to shrinkage) than any of the other estimators, as evidenced by the large discrepancy measures seen in the $p = 4$ columns in Panel B.

Finally, we observe that the relative discrepancies between alternative estimators are generally larger in the two-candlestick case than the single-candlestick case. As a case in point, the $\hat{\sigma}_{\text{BLUE}}$ and $\hat{\sigma}_{\text{Stein}}$ volatility estimators based on a single-candlestick display a high degree of similarity, with discrepancy summary statistics of 2.5% for the VOO and 4.4% for the Dollar/Yen, respectively, while the two-candlestick versions exhibit more significant differences, with discrepancy summary statistics equal to 4.6% for the VOO and 12.1% for the Dollar/Yen.

In summary, the comparisons in Tables A.3 and A.4 highlight potentially large empirical differences among the different estimators. The extent of the discrepancy depends, among other factors, on the estimand (i.e., σ_t^p), the loss function employed in deriving the AMRE estimator, and the number of candlesticks utilized in the estimation. It is, of course, also data dependent, as evidenced by the differences in the summary statistics for the VOO ETF and the Dollar/Yen exchange rate for the same set of estimators and times.

A.2.6 Monte Carlo Simulations

We evaluate the finite-sample performance of the proposed estimation methods through a Monte Carlo experiment. The data generating process for the price process is defined as follows

$$\begin{aligned} dP_t &= \sigma_t dW_t, \quad \sigma_t^2 = V_{1,t} + V_{2,t}, \\ dV_{1,t} &= 0.0128(0.4068 - V_{1,t})dt + 0.0954\sqrt{V_{1,t}}(\rho dW_t + \sqrt{1-\rho^2}dB_{1,t}), \\ dV_{2,t} &= 0.6930(0.4068 - V_{2,t})dt + 0.7023\sqrt{V_{2,t}}(\rho dW_t + \sqrt{1-\rho^2}dB_{2,t}), \end{aligned}$$

where W , B_1 , and B_2 denote independent standard Brownian motions, $\rho = -0.7$, $V_{1,0} = V_{2,0} = 0.5$, so that $\sigma_0 = 1$. We simulate the ‘‘continuous-time processes’’ using a Euler scheme with mesh size being 10^{-4} minute. The candlesticks utilized

in the calculations are constructed on 5-minute intervals, a common default choice in applied work, which we also adopt in our empirical study. The estimand σ_t^p with $p \in \{1, 2\}$ is sampled at a random point within each 5-minute estimation window. All numerical results reported below are based on 10,000 Monte Carlo replications.

To evaluate the potential distorting effects of market microstructure noise, we also consider a “noisy setting,” in which the observed price, denoted by Y_t , is generated as

$$Y_t = P_t + \varepsilon_t,$$

where the ε_t noise terms are i.i.d. $\mathcal{N}(0, \zeta^2)$. We examine values of $\zeta \in \{5, 10, 20, 40, 80, 160\} \times 10^{-4}$. The value $\zeta = 5 \times 10^{-4}$ is in line with the empirically realistic simulation settings of [Da and Xiu \(2021\)](#) and [Li and Linton \(2022\)](#). Hence, in addition to this “representative” noise level, our experiment also involves much larger noise levels to help illuminate the effect of sampling “too finely” relative to the magnitude of the noise.

Tables [A.5–A.11](#) present the finite-sample biases, relative efficiencies under Stein’s loss and the quadratic loss, and coverage rates of 90% confidence intervals (as detailed in footnote [11](#) of the main text) for various estimators. The finite-sample properties of these estimators under the no-noise scenario, displayed in Table [A.5](#), align with the asymptotic theory. The AMRE estimators exhibit the lowest risks, and the coverage rates for all confidence intervals are close to the nominal level.

As evidenced by Table [A.6](#), the results are effectively the same at an empirically realistic noise level of $\zeta = 5 \times 10^{-4}$. This observation aligns with the “conventional wisdom” that in typical applications, microstructure noise has a negligible impact on volatility estimates based on a 5-minute sampling frequency.

Looking across Tables [A.7](#) to [A.11](#), the noise level is progressively doubled to highlight the potential distortion effects of noise. As the noise level increases, all estimators experience a growing upward bias. The quadratic-AMRE estimator, $\hat{\sigma}_{\text{Quad}}^p$, is the only shrinkage estimator considered, and its inherent downward bias offers a unique advantage in counteracting the upward bias caused by the noise. When the noise level is not too high, the bias of the quadratic-AMRE estimator tends towards zero, and it eventually becomes the least biased estimator at very high noise

levels. Consequently, $\hat{\sigma}_{\text{Quad}}^P$ achieves the lowest finite-sample quadratic risk across all the different settings. Furthermore, when the noise level is sufficiently high, $\hat{\sigma}_{\text{Quad}}^P$ surpasses $\hat{\sigma}_{\text{Stein}}^P$ and emerges as the minimum risk estimator in finite samples, even under Stein's loss.

Looking at the finite-sample coverage rates of the confidence intervals, it is interesting to note that as the noise level increases, the confidence intervals tend to cover the true value more frequently than the nominal level. This indicates that noise-induced distortion does not necessarily result in under-coverage. However, over-coverage is not consistently observed either. Indeed, at the highest noise level (Table A.11), the confidence intervals all display severe under-coverage.

In summary, our simulation results support the widely-held belief that for empirically realistically calibrated noise levels, 5-minute coarse sampling effectively mitigates the detrimental impacts of the noise. Meanwhile, in more extreme situations with very high noise levels, the quadratic-AMRE estimator demonstrates superior finite-sample performance, primarily because its shrinkage properties allow it to partially counteract the upward bias induced by excessive noise.

Table A.3: Relative Discrepancies for Alternative One-Candlestick Estimators

	VOO			Dollar/Yen		
	$p = 1$	$p = 2$	$p = 4$	$p = 1$	$p = 2$	$p = 4$
<i>Panel A: Comparison versus the Stein-AMRE Estimator</i>						
$\hat{\sigma}_{\text{Quad}}^p$	0.058	0.210	0.587	0.057	0.211	0.593
$(\hat{\sigma}_{\text{GK}}^2)^{p/2}$	0.059	0.094	0.361	0.089	0.142	0.389
$(\hat{\sigma}_{\text{BLUE}})^p$	0.025	0.081	0.445	0.044	0.122	0.537
<i>Panel B: Comparison versus the Quadratic-AMRE Estimator</i>						
$\hat{\sigma}_{\text{Stein}}^p$	0.062	0.268	1.480	0.061	0.270	1.561
$(\hat{\sigma}_{\text{GK}}^2)^{p/2}$	0.056	0.297	2.145	0.074	0.399	2.216
$(\hat{\sigma}_{\text{BLUE}})^p$	0.066	0.347	2.558	0.081	0.379	3.032

Note: The table reports summary statistics for the relative differences of alternative single-candlestick estimators of σ_t^p with respect to the AMRE estimators. In Panel A and Panel B, the benchmark AMRE estimators are set as $\hat{\sigma}_{\text{Stein}}^p$ and $\hat{\sigma}_{\text{Quad}}^p$, respectively. The reported values are the root mean squared relative differences, calculated across all spot estimates during the eight FOMC announcement days in 2022.

Table A.4: Relative Discrepancies for Alternative Two-Candlestick Estimators

	VOO			Dollar/Yen		
	$p = 1$	$p = 2$	$p = 4$	$p = 1$	$p = 2$	$p = 4$
<i>Panel A: Comparison versus the Stein-AMRE Estimator</i>						
$\hat{\sigma}_{\text{Quad}}^p$	0.034	0.122	3.072	0.043	0.134	3.170
$(\hat{\sigma}_{\text{GK}}^2)^{p/2}$	0.081	0.224	1.733	0.177	0.882	1.981
$(\hat{\sigma}_{\text{BLUE}})^p$	0.046	0.130	0.816	0.121	0.534	1.105
<i>Panel B: Comparison versus the Quadratic-AMRE Estimator</i>						
$\hat{\sigma}_{\text{Stein}}^p$	0.035	0.151	3.526	0.046	0.177	4.194
$(\hat{\sigma}_{\text{GK}}^2)^{p/2}$	0.105	0.479	9.313	0.216	1.542	9.664
$(\hat{\sigma}_{\text{BLUE}})^p$	0.071	0.331	6.853	0.156	0.968	4.995

Note: The table reports summary statistics for the relative differences of alternative two-candlestick estimators of σ_t^p with respect to the AMRE estimators. In Panel A and Panel B, the benchmark AMRE estimators are set as $\hat{\sigma}_{\text{Stein}}^p(2)$ and $\hat{\sigma}_{\text{Quad}}^p(2)$, respectively. The reported values are the root mean squared relative differences, calculated across all spot estimates during the eight FOMC announcement days in 2022.

Table A.5: Finite-Sample Properties of Alternative Estimators: No Noise

Estimator	Spot Variance ($p = 2$)				Spot Volatility ($p = 1$)			
	Bias	Rel. Eff.			Bias	Rel. Eff.		
		Stein	Quad.	C.R.		Stein	Quad.	C.R.
$\hat{\sigma}_{\text{Stein}}^p$	-0.007	1.000	0.896	0.899	-0.005	1.000	0.975	0.899
$\hat{\sigma}_{\text{Quad}}^p$	-0.211	0.824	1.000	0.898	-0.063	0.937	1.000	0.899
$(\hat{\sigma}_{\text{BLUE}})^p$	0.050	0.984	0.841	0.898	-0.007	0.993	0.974	0.898
$(\hat{\sigma}_{\text{GK}}^2)^{p/2}$	0.006	0.984	0.872	0.903	-0.029	0.971	0.981	0.902

Note: The table reports the finite-sample relative biases, relative efficiencies under Stein's and quadratic loss functions, and coverage rates of 90% confidence intervals for alternative estimators of the spot variance (left) and spot volatility (right). The noise level is set as $\zeta = 0$.

Table A.6: Finite-Sample Properties of Alternative Estimators: Noise Level $\zeta = 5 \times 10^{-4}$

Estimator	Spot Variance ($p = 2$)				Spot Volatility ($p = 1$)			
	Bias	Rel. Eff.			Bias	Rel. Eff.		
		Stein	Quad.	C.R.		Stein	Quad.	C.R.
$\hat{\sigma}_{\text{Stein}}^p$	-0.004	1.000	0.887	0.903	0.001	1.000	0.969	0.902
$\hat{\sigma}_{\text{Quad}}^p$	-0.202	0.836	1.000	0.904	-0.057	0.944	1.000	0.903
$(\hat{\sigma}_{\text{BLUE}})^p$	0.062	0.980	0.832	0.903	-0.001	0.993	0.969	0.902
$(\hat{\sigma}_{\text{GK}}^2)^{p/2}$	0.018	0.982	0.862	0.907	-0.023	0.974	0.979	0.906

Note: The table reports the finite-sample relative biases, relative efficiencies under Stein's and quadratic loss functions, and coverage rates of 90% confidence intervals for alternative estimators of the spot variance (left) and spot volatility (right). The noise level is set as $\zeta = 5 \times 10^{-4}$.

Table A.7: Finite-Sample Properties of Alternative Estimators: Noise Level $\zeta = 1 \times 10^{-3}$

Estimator	Spot Variance ($p = 2$)				Spot Volatility ($p = 1$)			
	Bias	Rel. Eff.			Bias	Rel. Eff.		
		Stein	Quad.	C.R.		Stein	Quad.	C.R.
$\hat{\sigma}_{\text{Stein}}^p$	0.028	1.000	0.866	0.911	0.016	1.000	0.955	0.913
$\hat{\sigma}_{\text{Quad}}^p$	-0.182	0.864	1.000	0.913	-0.043	0.970	1.000	0.913
$(\hat{\sigma}_{\text{BLUE}})^p$	0.087	0.968	0.810	0.910	0.014	0.990	0.955	0.912
$(\hat{\sigma}_{\text{GK}}^2)^{p/2}$	0.043	0.976	0.840	0.914	-0.008	0.983	0.970	0.915

Note: The table reports the finite-sample relative biases, relative efficiencies under Stein's and quadratic loss functions, and coverage rates of 90% confidence intervals for alternative estimators of the spot variance (left) and spot volatility (right). The noise level is set as $\zeta = 1 \times 10^{-3}$.

Table A.8: Finite-Sample Properties of Alternative Estimators: Noise Level $\zeta = 2 \times 10^{-3}$

Estimator	Spot Variance ($p = 2$)				Spot Volatility ($p = 1$)			
	Bias	Rel. Eff.			Bias	Rel. Eff.		
		Stein	Quad.	C.R.		Stein	Quad.	C.R.
$\hat{\sigma}_{\text{Stein}}^p$	0.089	1.000	0.820	0.934	0.043	1.000	0.930	0.926
$\hat{\sigma}_{\text{Quad}}^p$	-0.134	0.949	1.000	0.935	-0.017	1.026	1.000	0.925
$(\hat{\sigma}_{\text{BLUE}})^p$	0.148	0.942	0.762	0.933	0.040	0.996	0.932	0.925
$(\hat{\sigma}_{\text{GK}}^2)^{p/2}$	0.103	0.969	0.794	0.936	0.018	1.008	0.955	0.927

Note: The table reports the finite-sample relative biases, relative efficiencies under Stein's and quadratic loss functions, and coverage rates of 90% confidence intervals for alternative estimators of the spot variance (left) and spot volatility (right). The noise level is set as $\zeta = 2 \times 10^{-3}$.

Table A.9: Finite-Sample Properties of Alternative Estimators: Noise Level $\zeta = 4 \times 10^{-3}$

Estimator	Spot Variance ($p = 2$)				Spot Volatility ($p = 1$)			
	Bias	Rel. Eff.			Bias	Rel. Eff.		
		Stein	Quad.	C.R.		Stein	Quad.	C.R.
$\hat{\sigma}_{\text{Stein}}^p$	0.244	1.000	0.730	0.962	0.119	1.000	0.873	0.935
$\hat{\sigma}_{\text{Quad}}^p$	-0.009	1.260	1.000	0.963	0.055	1.197	1.000	0.933
$(\hat{\sigma}_{\text{BLUE}})^p$	0.306	0.894	0.672	0.962	0.113	1.011	0.881	0.935
$(\hat{\sigma}_{\text{GK}}^2)^{p/2}$	0.255	0.962	0.709	0.963	0.090	1.073	0.919	0.935

Note: The table reports the finite-sample relative biases, relative efficiencies under Stein's and quadratic loss functions, and coverage rates of 90% confidence intervals for alternative estimators of the spot variance (left) and spot volatility (right). The noise level is set as $\zeta = 4 \times 10^{-3}$.

Table A.10: Finite-Sample Properties of Alternative Estimators: Noise Level $\zeta = 8 \times 10^{-3}$

Estimator	Spot Variance ($p = 2$)				Spot Volatility ($p = 1$)			
	Bias	Rel. Eff.			Bias	Rel. Eff.		
		Stein	Quad.	C.R.		Stein	Quad.	C.R.
$\hat{\sigma}_{\text{Stein}}^p$	0.633	1.000	0.660	0.935	0.289	1.000	0.836	0.865
$\hat{\sigma}_{\text{Quad}}^p$	0.302	1.926	1.000	0.936	0.216	1.377	1.000	0.860
$(\hat{\sigma}_{\text{BLUE}})^p$	0.703	0.883	0.612	0.942	0.279	1.041	0.855	0.871
$(\hat{\sigma}_{\text{GK}}^2)^{p/2}$	0.631	0.986	0.651	0.937	0.249	1.168	0.909	0.868

Note: The table reports the finite-sample relative biases, relative efficiencies under Stein's and quadratic loss functions, and coverage rates of 90% confidence intervals for alternative estimators of the spot variance (left) and spot volatility (right). The noise level is set as $\zeta = 8 \times 10^{-3}$.

Table A.11: Finite-Sample Properties of Alternative Estimators: Noise Level $\zeta = 1.6 \times 10^{-2}$

Estimator	Spot Variance ($p = 2$)				Spot Volatility ($p = 1$)			
	Bias	Rel. Eff.			Bias	Rel. Eff.		
		Stein	Quad.	C.R.		Stein	Quad.	C.R.
$\hat{\sigma}_{\text{Stein}}^p$	1.677	1.000	0.708	0.598	0.665	1.000	0.872	0.401
$\hat{\sigma}_{\text{Quad}}^p$	1.146	1.721	1.000	0.590	0.573	1.268	1.000	0.381
$(\hat{\sigma}_{\text{BLUE}})^p$	1.772	0.924	0.673	0.633	0.646	1.046	0.895	0.424
$(\hat{\sigma}_{\text{GK}}^2)^{p/2}$	1.612	1.052	0.727	0.653	0.595	1.184	0.959	0.454

Note: The table reports the finite-sample relative biases, relative efficiencies under Stein's and quadratic loss functions, and coverage rates of 90% confidence intervals for alternative estimators of the spot variance (left) and spot volatility (right). The noise level is set as $\zeta = 1.6 \times 10^{-2}$.

Appendix B

Technical Results for Chapter 3

B.1 Proofs of the Main Results

B.1.1 Proof of Theorem 3.2

The proof is based on the following lemma.

Lemma B.1. *Let $X = (X_1, \dots, X_m)$ be nonnegative random variables with density $f_X(x_1, \dots, x_m)$, for $1 \leq i \leq m$ denote $Z_i = X_i/X_1$ and let $Z \equiv (Z_i)_{1 \leq i \leq m}$. Then we have for $p \geq 1$*

$$\mathbb{E}[X_1^p | Z] = x_1^p \cdot \frac{\int_0^\infty v^{m+p-1} f_X(x_1 v, \dots, x_m v) dv}{\int_0^\infty v^{m-1} f_X(x_1 v, \dots, x_m v) dv}.$$

To prove the above result, note that the conditional density of X_1 given Z is

$$f_{X_1|Z}(x_1 | z) = \frac{f_{X_1, Z}(x_1, z)}{\int_0^\infty f_{X_1, Z}(t, z) dt}.$$

Also note that the Jacobian $|dx/d(x, z)| = x_1^{m-1}$, hence

$$f_{X_1, Z}(x_1, z) = f_X(x_1, z_2 x_1, \dots, z_m x_1) x_1^{m-1}.$$

Combining above results, the conditional density of X_1 is then

$$f_{X_1|Z}(x_1 | z) = \frac{x_1^{m-1} f_X(x_1, z_2 x_1, \dots, z_m x_1)}{\int_0^\infty t^{m-1} f_X(t, z_2 t, \dots, z_m t) dt}.$$

Therefore, the conditional expectation of X_1^p can be written as

$$\mathbb{E}[X_1^p | Z] = \int_0^\infty x_1^p \cdot \frac{x_1^{m-1} f_X(x_1, z_2 x_1, \dots, z_m x_1)}{\int_0^\infty t^{m-1} f_X(t, z_2 t, \dots, z_m t) dt} dx_1$$

$$\begin{aligned}
&= \frac{\int_0^\infty x_1^{m+p-1} f_X(x_1, z_2 x_1, \dots, z_m x_1) dx_1}{\int_0^\infty t^{m-1} f_X(t, z_2 t, \dots, z_m t) dt} \\
&= \frac{\int_0^\infty t^{m+p-1} f_X(t, z_2 t, \dots, z_m t) dt}{\int_0^\infty t^{m-1} f_X(t, z_2 t, \dots, z_m t) dt} \\
&= \frac{\int_0^\infty (x_1 v)^{m+p-1} f_X(x_1 v, z_2 x_1 v, \dots, z_m x_1 v) x_1 dv}{\int_0^\infty (x_1 v)^{m-1} f_X(x_1 v, z_2 x_1 v, \dots, z_m x_1 v) x_1 dv} \\
&= x_1^p \cdot \frac{\int_0^\infty v^{m+p-1} f_X(x_1 v, x_2 v, \dots, x_m v) x_1 dv}{\int_0^\infty v^{m-1} f_X(x_1 v, x_2 v, \dots, x_m v) x_1 dv},
\end{aligned}$$

which completes the proof.

Taking $m = 3k$ and $(X_1, \dots, X_{3k}) = \mathbf{C}_k$, the theorem then follows the same procedure as Theorem 2 in [Bollerslev et al. \(2024\)](#) by applying the above formula to Equation (2.12) of [Bollerslev et al. \(2024\)](#). *Q.E.D.*

B.1.2 Proof of Theorem 3.3

To prove $(r, h, l) \stackrel{d}{=} (\tilde{\zeta}_{i,r}, \tilde{\zeta}_{i,h}, \tilde{\zeta}_{i,l})$, it suffices to show that:

$$r \stackrel{d}{=} \tilde{\zeta}_{i,r}, \quad h | r \stackrel{d}{=} \tilde{\zeta}_{i,h} | \tilde{\zeta}_{i,r}, \quad l | (h, r) \stackrel{d}{=} \tilde{\zeta}_{i,l} | (\tilde{\zeta}_{i,r}, \tilde{\zeta}_{i,h}). \quad (\text{B.1})$$

The first relation is obvious. For the second relation, we start with the joint law of $(\tilde{\zeta}_{i,h}, \tilde{\zeta}_{i,r})$ which can be found in, e.g., [Shepp \(1979\)](#):

$$f_{\tilde{\zeta}_{i,h}, \tilde{\zeta}_{i,r}}(h, r) = \frac{2(2h-c)}{\sqrt{2\pi}} \exp\left\{-\frac{(2h-c)^2}{2}\right\} = -2\phi'(2h-c),$$

where $h \geq (c \vee 0)$. The density of $\tilde{\zeta}_{i,h}$ conditional on $\tilde{\zeta}_{i,r}$ is thus

$$f_{\tilde{\zeta}_{i,h} | \tilde{\zeta}_{i,r}}(h | r) = \frac{-\phi'(2h-c)}{\phi(c)} = 2(2h-c) \exp\{-2h(h-c)\}.$$

Direct integration reveals that

$$F_{\tilde{\zeta}_{i,h} | \tilde{\zeta}_{i,r}}(h | r) = \int_{c \vee 0}^h f_{\tilde{\zeta}_{i,h} | \tilde{\zeta}_{i,r}}(s | r) ds = 1 - \exp\{2(c-h)h\}.$$

From which one can directly calculate its inverse function

$$F_{\tilde{\zeta}_{i,h} | \tilde{\zeta}_{i,r}}^{-1}(u | r) = \frac{1}{2} \left(r + \sqrt{r - 2 \log(1-u)} \right).$$

By the probability integral transform, the second relation in Equation (B.1) is then immediate. For the last relation, we derive the density of $\tilde{\zeta}_{i,l} | (\tilde{\zeta}_{i,r}, \tilde{\zeta}_{i,h})$:

$$f_{\tilde{\zeta}_{i,l} | (\tilde{\zeta}_{i,r}, \tilde{\zeta}_{i,h})}(l | r, h)$$

$$\begin{aligned}
&= \frac{f_{\tilde{\xi}_{i,r}, \tilde{\xi}_{i,h}, \tilde{\xi}_{i,h}}(r, h, l)}{f_{\tilde{\xi}_{i,h}, \tilde{\xi}_{i,r}}(h, r)} \\
&= -2 \sum_{m=-\infty}^{\infty} m^2 \frac{\phi''(c - 2m(h-l))}{\phi'(2h-c)} - m(m+1) \frac{\phi''(c - 2kh + 2(k-1)l)}{\phi'(2h-c)},
\end{aligned}$$

where $l \leq r \wedge 0$ and the analytical form of $f_{\tilde{\xi}_{i,r}, \tilde{\xi}_{i,h}, \tilde{\xi}_{i,h}}(r, h, l)$ can be found in, e.g., [Feller \(1951\)](#). It is now straightforward to verify that $F'(l | r, h) = f_{\tilde{\xi}_{i,l} | (\tilde{\xi}_{i,r}, \tilde{\xi}_{i,h})}(l | r, h)$, and one can also check that $F(-\infty | r, h) = 0$ and $F(r \wedge 0 | r, h) = 1$. In other words, $F(l | r, h)$ is the cumulative distribution function of $\tilde{\xi}_{i,l}$ conditional on $\tilde{\xi}_{i,r}$ and $\tilde{\xi}_{i,h}$. The solution $v = F(l | r, h)$ is thus an implicit probability integral transform, and the third relation in Equation (B.1) readily follows, which completes the proof. *Q.E.D.*

B.2 Additional Results

B.2.1 Best Subset Regression for Spot Variance Estimators

This section contains best subset regression for AMRE spot variance estimators under Stein's loss and Quadratic loss. We adopt the polynomial design as described in Section 3.1. Table B.1 shows how different features are sequentially selected in this analysis. In contrast to the case of spot volatility estimators, the selected features are slightly different across different loss functions and across different values of k . Notably, the product of mean and standard error of the range $\bar{w}v(w)$ emerges as one of the most important features and outweighs the asymmetry level \bar{a} . Similar to the case of spot volatility estimators, the approximation using two features is in line with to the BQUE estimator proposed in [Garman and Klass \(1980\)](#), for instance when $k = 5$, we have

$$\hat{\sigma}_{(s)}^{2,*}(5) \approx 0.559\overline{w^2} - 0.399\overline{r^2}.$$

Figure B.1 shows the pattern of relative efficiency of polynomial-based approximation for optimal variance estimators. With a fixed number of features q , increases, the relative efficiency decreases. As mentioned previously, this is due to the absence of a dimension-reduced complete sufficient statistic. Therefore, larger k naturally implies a greater loss of information, resulting in an estimator further from the optimal ones that utilize all available information. Similar to the case of spot volatility estimators, this decrease become gradual in the case of large k , suggesting

Table B.1: Best Subset Regression Result of $\hat{\sigma}_{(c)}^{2,*}(k)$

Panel I: AMRE under Stein's loss											
q	\bar{w}^2	$ \bar{r} ^2$	$\nu(w)^2$	$\bar{w}\nu(w)$	$\bar{w}\nu(a)$	$\bar{a}\nu(w)$	$\bar{a}\nu(a)$	$\bar{a}\nu(r)$	$ \bar{r} \nu(w)$	$ \bar{r} \nu(a)$	R.Eff.
1	19	0	0	0	0	0	0	0	0	0	67.2%
2	19	19	0	0	0	0	0	0	0	0	94.8%
3	19	19	0	0	0	15	0	4	0	0	96.3%
4	19	19	1	18	0	0	0	19	0	0	96.7%
5	19	19	16	19	0	0	0	19	3	0	96.8%
6	19	19	4	18	16	0	3	19	0	16	96.9%

Panel II: AMRE under Quadratic loss											
q	\bar{w}^2	\bar{a}^2	$ \bar{r} ^2$	$\nu(w)^2$	$\bar{w}\nu(w)$	$\bar{w}\nu(a)$	$\bar{w}\nu(r)$	$\bar{a}\nu(w)$	$\bar{a}\nu(r)$	$ \bar{r} \nu(a)$	R.Eff.
1	19	0	0	0	0	0	0	0	0	0	67.5%
2	19	0	19	0	0	0	0	0	0	0	95.1%
3	19	0	19	0	0	0	0	12	7	0	96.5%
4	19	0	19	1	18	0	0	0	19	0	96.8%
5	19	2	19	15	15	2	0	2	19	2	97.0%
6	19	0	19	1	8	19	10	10	9	19	97.1%

Note: The table reports the total number of selection for $k \in \{2, \dots, 20\}$ of each feature, and the average relative efficiency w.r.t. the analytical AMRE estimators for $q \in \{1, \dots, 6\}$.

that one may trade some efficiency for faster computation and easier-to-understand estimators. Moreover, the figure indicates that the marginal improvement of adding an additional feature becomes nearly negligible as q becomes larger, say, greater than 3. For instance, for the optimal estimator under Stein’s loss with $k = 5$, the relative efficiency levels of approximation using $q = 1, 2, 3, 4$ features are 53.92%, 95.44%, 97.12%, 97.66%, respectively.

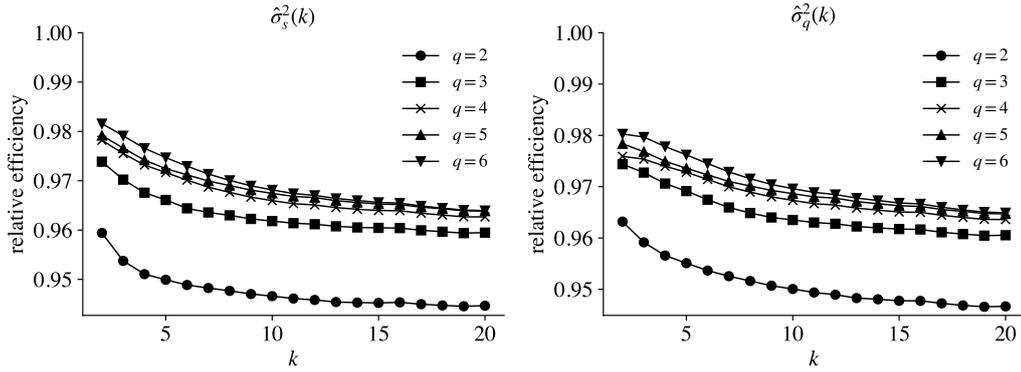


Figure B.1: Relative efficiency of polynomial-based approximation for optimal variance estimators under Stein’s loss (left) and quadratic loss (right), with $q \in \{2, \dots, 6\}$ and $k \in \{2, \dots, 20\}$. The selection of features and determination of corresponding coefficients are computed using best subset regression.

B.2.2 Volatility Estimation during Price Indices Release

Figure B.3 and B.2 show Stein-AMRE spot volatility estimates of E-mini future continuous contract for price indices releases in the years 2023 and 2022, respectively. Similar patterns can be found in these figures as well: volatility initially increases after release, gradually reverts within a 60-minute window, and then undergoes another uptick coinciding with market opening. Moreover, upon closer examination of individual release days, one can find some discernible pre-release volatility movements (see, e.g., observed in PCE of September 2022, CPI of December 2022, and PPI of May 2023). Note that in accordance with government regulations, dissemination of materials related to price index releases is embargoed until 8:30 a.m. on the scheduled date. The existence of pre-release movements suggests a potential leakage of information.

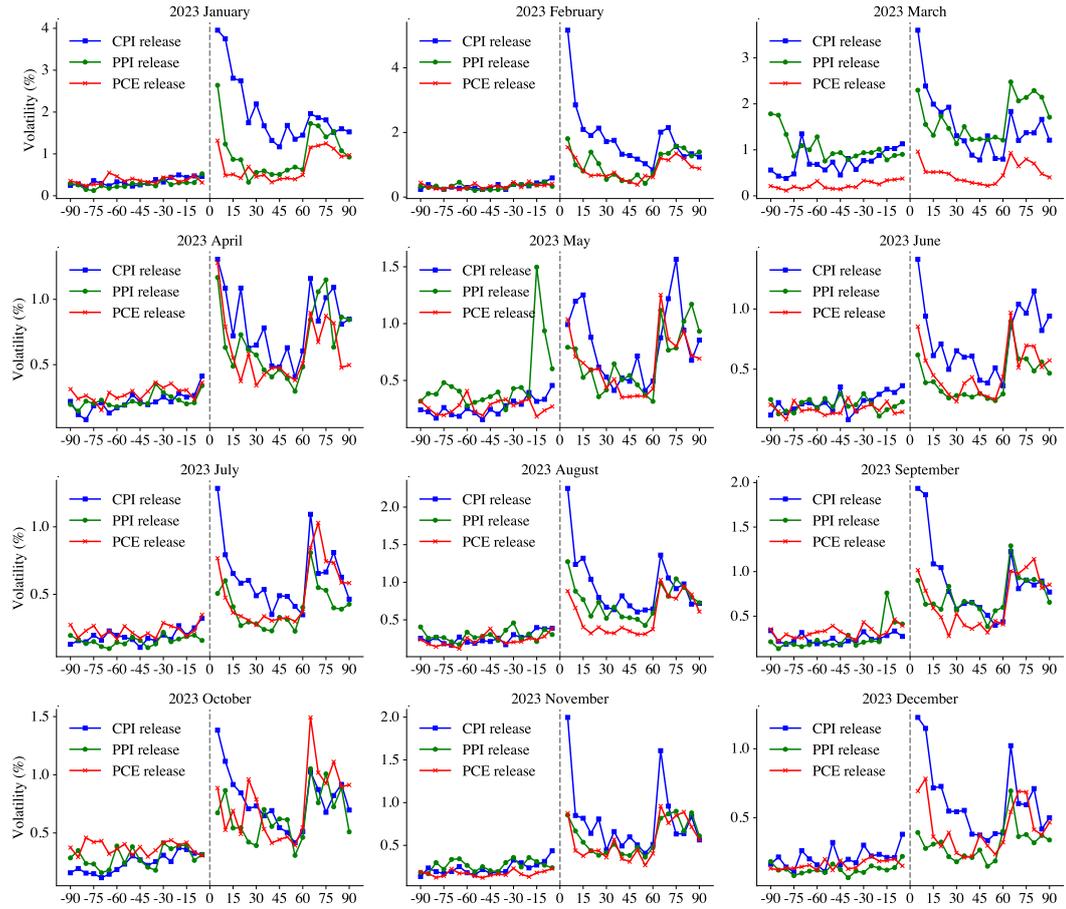


Figure B.2: Stein-AMRE spot volatility estimates of E-mini S&P500 future within a 1.5 hours window before and after price indices releases in 2023. The estimation is based on five consecutive observation intervals sampled at a 1-minute frequency. Prices immediately following the releases are excluded to mitigate the effects of potential price jumps.

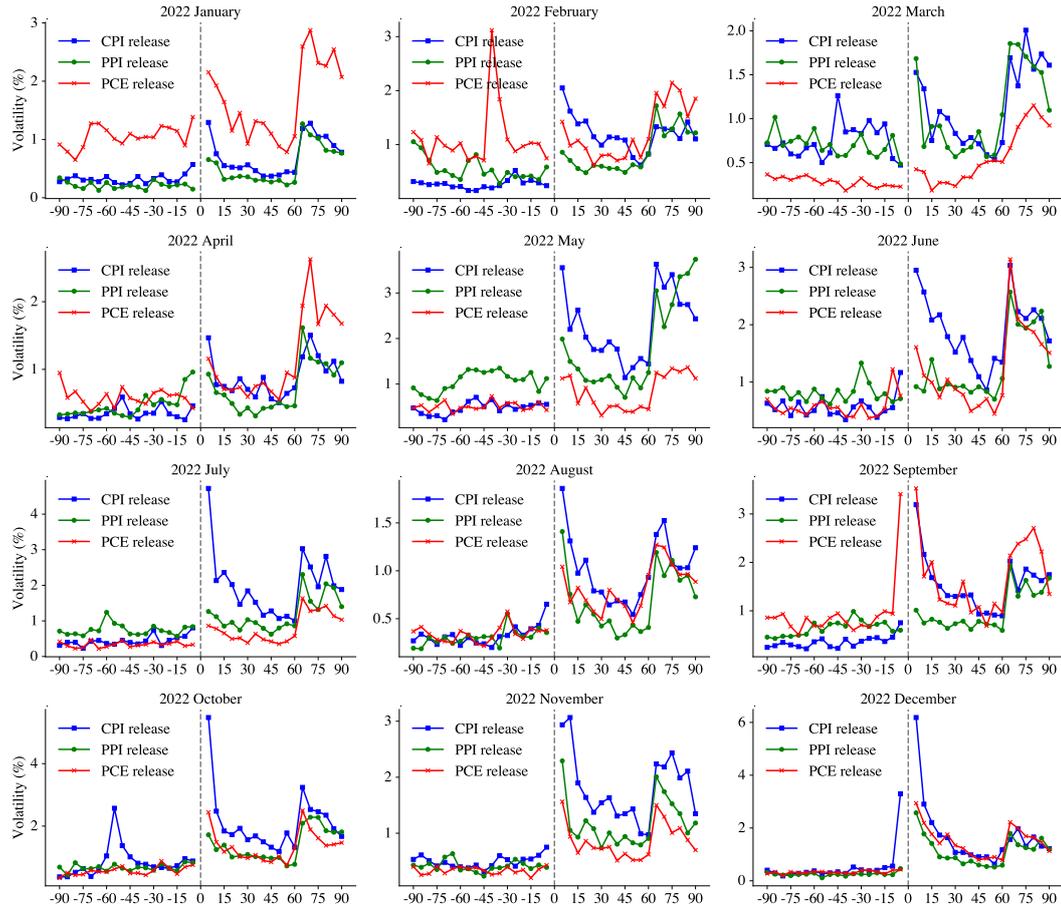


Figure B.3: Stein-AMRE spot volatility estimates of E-mini S&P500 future within a 1.5 hours window before and after price indices releases in 2022. The estimation is based on five consecutive observation intervals sampled at a 1-minute frequency. Prices immediately following the releases are excluded to mitigate the effects of potential price jumps.

Appendix C

Technical Results for Chapter 4

C.1 Proofs of the Main Results

Throughout the proofs, we use K and K' to denote some positive constants that may change from line to line, and write K_p to emphasize its dependence on some parameter p . In order to make a distinction, we use M to denote some positive constant defined in the context which is hold fixed across lines. For notation simplicity, we denote $L_n \equiv \log(\Delta_n^{-1})$.

C.1.1 Proof of Theorem 4.1

By a standard localization procedure (see, e.g., Section 4.4.1 in [Jacod and Protter \(2012\)](#) for a detailed discussion of localization procedure), we can strengthen Assumption 3 by assuming $T_1 = \infty$, $\mathcal{K}_m = \mathcal{K}$, and $K_m = K$ for some fixed compact set \mathcal{K} and constant $K > 0$. That is, it suffices to prove the results under Assumption 5.

Assumption 5. *There exist a positive constant K , and a compact subset $\mathcal{K} \subset \mathcal{Z}$ such that: (i) ζ takes value in \mathcal{K} ; for all $s, t \in \mathcal{T}_{n,j}$ where $1 \leq j \leq m_n$, and for each $p > 0$, $\mathbb{E}[\|\zeta_t - \zeta_s\|^p] \leq K_p |t - s|^{p/2}$ for some constant K_p ; (ii) for all $z, z' \in \mathcal{K}$ with $z \neq z'$, $\text{Var}(\mathcal{Y}(z, \varepsilon))^{-1} + \|\mathcal{Y}(z, \varepsilon) - \mathcal{Y}(z', \varepsilon)\|_{L_2} / \|z - z'\| \leq K$; (iii) for all $x > 0$ and $z \in \mathcal{K}$, $\mathbb{P}_\varepsilon(|\mathcal{Y}(z, \varepsilon)| \geq x) \leq K \exp\{-(x/K)^{1/\eta}\}$ for some $\eta > 0$; (iv) $\max_{1 \leq i \leq n} |R_{n,i}| = o_p(\Delta_n^r)$ for some $r > 0$.*

Consequently, we have ζ globally takes values in the compact set \mathcal{K} and is $1/2$ -Hölder continuous under the L_p norm within each block. Denote $G_p(\cdot) \equiv \int_{\mathcal{D}} \mathcal{Y}(\cdot, \varepsilon)^p \mathbb{P}_\varepsilon(d\varepsilon)$, we have for all $z \in \mathcal{K}$, $\text{Var}(\mathcal{Y}(z, \varepsilon)) = G_2(z) - G_1^2(z)$ is bounded away from zero. Note that by Theorem 2.1 in [Vladimirova et al. \(2020\)](#), Assumption 5(iii) implies for all $p \geq 1$, $G_p(z)$ is bounded from above by K_p uniformly over $z \in \mathcal{K}$, and by a maximal inequality (see, e.g., Lemma 2.2.2 in [van der Vaart and Wellner \(1996\)](#)),

$$\sup_{z \in \mathcal{K}} \left\| \max_{1 \leq j \leq m_n} \mathcal{Y}(z, \varepsilon_j) \right\|_{L_p} \leq K_p (\log m_n)^\eta \leq K_p L_n^\eta. \quad (\text{C.1})$$

We prove the validity of the assertion in the theorem for all positive ε satisfying

$$\varepsilon < \frac{\rho}{6} \wedge \left(\frac{1}{6} - \frac{\rho}{3} \right) \wedge \left(\frac{r}{3} - \frac{\rho}{6} \right).$$

Note that such values of ε exist due to the assumption that $\rho \in (0, 2r \wedge 1/2)$. Correspondingly, we fix some positive γ constant satisfying

$$2\varepsilon < \gamma < \left(\frac{1}{2} - \rho - \varepsilon \right) \wedge \left(r - \frac{\rho}{2} - \varepsilon \right),$$

which is possible given the requirement of ε . To facilitate our analysis, we introduce some additional notations. For $1 \leq j \leq m_n$ and $1 \leq i \leq k_{n,j}$, denote

$$\begin{aligned} \tilde{Y}_{i,j} &\equiv \mathcal{Y}(\zeta_{\tau(i,j)}, \varepsilon_{n,t(i,j)}) - g_{\tau(i,j)}, \\ \sigma_{n,j}^2 &\equiv \frac{1}{k_{n,j}} \sum_{i=1}^{k_{n,j}} (G_2(\zeta_{\tau(i,j)}) - g_{\tau(i,j)}^2) \end{aligned}$$

Note that by the above construction, the variables $\tilde{Y}_{i,j}$ are $\mathcal{F}^{(0)}$ -conditionally independent across different values of i and j , with zero mean and conditional variance given by $G_2(\zeta_{\tau(i,j)}) - g_{\tau(i,j)}^2$. Furthermore, we define the infeasible sup- t statistic as

$$\tilde{T}_n^* \equiv \max_{1 \leq j \leq m_n} \left| \frac{1}{\sqrt{k_{n,j}}} \sum_{i=1}^{k_{n,j}} \frac{\tilde{Y}_{i,j}}{\sigma_{n,j}} \right|.$$

The proof is divided into three parts. In Step 1, we establish that \hat{T}_n^* can be strongly approximated by \tilde{T}_n^* in the following sense:

$$\mathbb{P}(|\hat{T}_n^* - \tilde{T}_n^*| > \delta_n) \leq K \Delta_n^\epsilon, \quad (\text{C.2})$$

for some real sequence satisfying $\delta_n \rightarrow 0$ and $\delta_n \sqrt{L_n} \leq K \Delta_n^\epsilon$. In Step 2, we con-

struct $(Z_j)_{1 \leq j \leq m_n}$ and prove the validity of the following inequality for \widetilde{T}_n^* :

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P}(\widetilde{T}_n^* \leq x) - \mathbb{P}\left(\max_{1 \leq j \leq m_n} |Z_j| \leq x\right) \right| \leq K \Delta_n^\epsilon. \quad (\text{C.3})$$

Step 3 concludes the proof by establishing the asserted statement.

STEP 1. Note that we can rewrite

$$\widehat{T}_n^* = \max_{1 \leq j \leq m_n} \sup_{t \in \mathcal{T}_{n,j}} \left| \frac{\sqrt{k_{n,j}}(\widehat{g}_{n,j} - g_t)}{\widehat{\sigma}_{n,j}} \right|.$$

By simple algebra we can verify that $|(a-b)/c - a/d| \leq |d/c - 1| \times |(a-b)/d| + |b/d|$. Recall equation (4.1), the proof of this step thus relies on the following decomposition

$$|\widehat{T}_n^* - \widetilde{T}_n^*| \leq \max_{1 \leq j \leq m_n} \left| \frac{\sigma_{n,j}}{\widehat{\sigma}_{n,j}} - 1 \right| \times \max_{1 \leq j \leq m_n} \left| \frac{1}{\sqrt{k_{n,j}}} \sum_{i=1}^{k_{n,j}} \frac{\widetilde{Y}_{i,j}}{\sigma_{n,j}} \right| + \max_{1 \leq j \leq m_n} |\mathfrak{A}_{n,j}|, \quad (\text{C.4})$$

where for $1 \leq j \leq m_n$, $\mathfrak{A}_{n,j} \equiv \mathfrak{A}_{n,j}^{(I)} + \mathfrak{A}_{n,j}^{(II)}$ with

$$\begin{aligned} \mathfrak{A}_{n,j}^{(I)} &\equiv \frac{1}{\sqrt{k_{n,j}}} \sum_{i=1}^{k_{n,j}} \frac{R_{n,t(i,j)}}{\sigma_{n,j}}, \\ \mathfrak{A}_{n,j}^{(II)} &\equiv \sup_{t \in \mathcal{T}_{n,j}} \frac{1}{\sqrt{k_{n,j}}} \sum_{i=1}^{k_{n,j}} \frac{g_{\tau(i,j)} - g_t}{\sigma_{n,j}}. \end{aligned}$$

Note that by Assumption 5(ii) and the definition of $\sigma_{n,j}$, we have $1/K \leq \sigma_{n,j} \leq K$ for all $1 \leq j \leq m_n$. Then Assumption 5(iv), together with $k_{n,j} \asymp \Delta_n^{-\rho}$, implies that

$$\max_{1 \leq j \leq m_n} |\mathfrak{A}_{n,j}^{(I)}| \leq K \Delta_n^{-\rho/2} \max_{1 \leq i \leq n} |R_{n,i}| = o_p(\Delta_n^{r-\rho/2}) = o_p(\Delta_n^{\epsilon+\gamma}). \quad (\text{C.5})$$

Note that Assumption 5(ii) implies function $G_1(\cdot)$ is Lipschitz since by the triangle inequality and the Hölder inequality $|G_1(z) - G_1(z')| \leq \|\mathcal{Y}(z, \varepsilon) - \mathcal{Y}(z', \varepsilon)\|_{L_2}$.

Also note that $m_n \asymp \Delta_n^{\rho-1}$ by $k_{n,j} \asymp \Delta_n^{-\rho}$, applying a maximal inequality, we have

$$\left\| \max_{1 \leq j \leq m_n} \mathfrak{A}_{n,j}^{(II)} \right\|_{L_p} \leq K_p m_n^{1/p} \max_{1 \leq j \leq m_n} \sqrt{k_{n,j}} (k_{n,j} \Delta_n)^{1/2} \leq K_p \Delta_n^{(\rho-1)/p+1/2-\rho}. \quad (\text{C.6})$$

Taking $p > (1-\rho)/(1/2-\rho-\epsilon-\gamma)$, the right-hand side becomes $o(\Delta_n^{\epsilon+\gamma})$. Then combining (C.5) and (C.6), it follows the triangle inequality and the Hölder inequality that

$$\max_{1 \leq j \leq m_n} |\mathfrak{A}_{n,j}| \leq \max_{1 \leq j \leq m_n} |\mathfrak{A}_{n,j}^{(I)}| + \max_{1 \leq j \leq m_n} |\mathfrak{A}_{n,j}^{(II)}| = o_p(\Delta_n^{\epsilon+\gamma}). \quad (\text{C.7})$$

For $1 \leq j \leq m_n$ and $1 \leq i \leq k_{n,j}$, denote

$$\tilde{\sigma}_{n,j}^2 \equiv \frac{1}{k_{n,j}} \sum_{i=1}^{k_{n,j}} \tilde{Y}_{i,j}^2 - \left(\frac{1}{k_{n,j}} \sum_{i=1}^{k_{n,j}} \tilde{Y}_{i,j} \right)^2.$$

Equation (C.5) and (C.6) also yield $\max_{1 \leq j \leq m_n} |\hat{\sigma}_{n,j} - \tilde{\sigma}_{n,j}| = o_p(\Delta_n^{\epsilon+\gamma})$. Recall $\sigma_{n,j}$ is bounded below by $1/K$ uniformly for all $1 \leq j \leq m_n$, by the triangle inequality, this implies

$$\begin{aligned} \max_{1 \leq j \leq m_n} \left| \frac{\hat{\sigma}_{n,j}}{\sigma_{n,j}} - 1 \right| &\leq \max_{1 \leq j \leq m_n} \left| \frac{\tilde{\sigma}_{n,j}}{\sigma_{n,j}} - 1 \right| + K \max_{1 \leq j \leq m_n} |\hat{\sigma}_{n,j} - \tilde{\sigma}_{n,j}| \\ &\leq \max_{1 \leq j \leq m_n} \left| \frac{\tilde{\sigma}_{n,j}}{\sigma_{n,j}} - 1 \right| + o_p(\Delta_n^{\epsilon+\gamma}). \end{aligned} \quad (\text{C.8})$$

Let $\bar{k}_n \equiv \max_{1 \leq j \leq m_n} k_{n,j}$, then $\bar{k}_n \asymp \Delta_n^{-\rho}$ and $1/K \leq \bar{k}_n/k_{n,j} \leq K$ uniformly for all $1 \leq j \leq m_n$. For each $1 \leq i \leq \bar{k}_n$ and $1 \leq j \leq m_n$, define $\tilde{U}_{i,j}$ and $v_{i,j}$ as follows:

$$\begin{aligned} \tilde{U}_{i,j} &\equiv \sqrt{\frac{\bar{k}_n}{k_{n,j}}} \frac{\tilde{Y}_{i,j}}{\sigma_{n,j}} \mathbb{1}\{1 \leq i \leq k_{n,j}\}, \\ v_{i,j} &\equiv \frac{\bar{k}_n (G_2(\zeta_{\tau(i,j)}) - g_{\tau(i,j)}^2)}{k_{n,j} \sigma_{n,j}^2} \mathbb{1}\{1 \leq i \leq k_{n,j}\}. \end{aligned}$$

By construction the variables $\tilde{U}_{i,j}$ remain $\mathcal{F}^{(0)}$ -conditionally independent across different values of $1 \leq i \leq \bar{k}_n$ and $1 \leq j \leq m_n$ with zero mean and conditional variance $v_{i,j}$. Note that

$$\begin{aligned} \frac{\tilde{\sigma}_{n,j}^2}{\sigma_{n,j}^2} - 1 &= \left(\frac{1}{k_{n,j}} \sum_{i=1}^{k_{n,j}} \frac{\tilde{Y}_{i,j}^2}{\sigma_{n,j}^2} - 1 \right) - \left(\frac{1}{k_{n,j}} \sum_{i=1}^{k_{n,j}} \frac{\tilde{Y}_{i,j}}{\sigma_{n,j}} \right)^2 \\ &= \left(\frac{1}{\bar{k}_n} \sum_{i=1}^{\bar{k}_n} \tilde{U}_{i,j}^2 - 1 \right) - \left(\frac{1}{\bar{k}_n} \sum_{i=1}^{\bar{k}_n} \tilde{U}_{i,j} \right)^2. \end{aligned}$$

Also note that by simple algebra we can verify that for positive a , $|\sqrt{a} - 1| = |a - 1|/(\sqrt{a} + 1) \leq |a - 1|$, then we can deduce

$$\begin{aligned} \mathbb{P} \left(\max_{1 \leq j \leq m_n} \left| \frac{\tilde{\sigma}_{n,j}}{\sigma_{n,j}} - 1 \right| > x \mid \mathcal{F}^{(0)} \right) &\leq \mathbb{P} \left(\max_{1 \leq j \leq m_n} \left| \frac{1}{\bar{k}_n} \sum_{i=1}^{\bar{k}_n} \tilde{U}_{i,j}^2 - 1 \right| > \frac{x}{2} \mid \mathcal{F}^{(0)} \right) \\ &\quad + \mathbb{P} \left(\max_{1 \leq j \leq m_n} \left| \frac{1}{\bar{k}_n} \sum_{i=1}^{\bar{k}_n} \tilde{U}_{i,j} \right| > \sqrt{\frac{x}{2}} \mid \mathcal{F}^{(0)} \right). \end{aligned} \quad (\text{C.9})$$

For the first term, noting that by Assumption 5(ii), we have

$$\max_{1 \leq j \leq m_n} \sum_{i=1}^{k_{n,j}} \mathbb{E}[\tilde{U}_{i,j}^4 | \mathcal{F}^{(0)}] \leq K \max_{1 \leq j \leq m_n} \sum_{i=1}^{k_{n,j}} G_4(\zeta_{\tau(i,j)}) \leq K \Delta_n^{-\rho}.$$

By (C.1), we can further deduce for each $1 \leq i \leq \bar{k}_n$,

$$\mathbb{E} \left[\max_{1 \leq j \leq m_n} \tilde{U}_{i,j}^4 | \mathcal{F}^{(0)} \right] \leq K \sup_{z \in \mathcal{K}} \mathbb{E} \left[\max_{1 \leq j \leq m_n} \mathcal{Y}(z, \varepsilon_{n,t(i,j)})^4 \right] \leq K L_n^{4\eta}. \quad (\text{C.10})$$

Then by a maximal inequality, we obtain

$$\mathbb{E} \left[\max_{1 \leq i \leq \bar{k}_n} \max_{1 \leq j \leq m_n} \tilde{U}_{i,j}^4 | \mathcal{F}^{(0)} \right] \leq K \Delta_n^{-\rho} L_n^{4\eta}.$$

Observing that by the definition of $v_{i,j}$ and $\sigma_{n,j}$, we can verify

$$\frac{1}{\bar{k}_n} \sum_{i=1}^{\bar{k}_n} \mathbb{E}[\tilde{U}_{i,j}^2 | \mathcal{F}^{(0)}] = \frac{1}{\bar{k}_n} \sum_{i=1}^{\bar{k}_n} v_{i,j} = \frac{\sigma_{n,j}}{\sigma_{n,j}} = 1.$$

Then by Lemma 8 in Chernozhukov et al. (2015), we obtain

$$\mathbb{E} \left[\max_{1 \leq j \leq m_n} \left| \frac{1}{\bar{k}_n} \sum_{i=1}^{\bar{k}_n} \tilde{U}_{i,j}^2 - 1 \right| \middle| \mathcal{F}^{(0)} \right] \leq K(\Delta_n^{\rho/2} \sqrt{L_n} + \Delta_n^{\rho/2} L_n^{1+2\eta}).$$

Therefore, a Fuk–Nagaev type inequality (see Theorem 4 in Einmahl and Li (2008))

implies that for every $x > 0$,

$$\begin{aligned} & \mathbb{P} \left(\max_{1 \leq j \leq m_n} \left| \frac{1}{\bar{k}_n} \sum_{i=1}^{\bar{k}_n} \tilde{U}_{i,j}^2 - 1 \right| > K \Delta_n^{\rho/2} L_n^{1+2\eta} + x \middle| \mathcal{F}^{(0)} \right) \\ & \leq \exp\{-K' x^2 \Delta_n^{-\rho}\} + K' x^{-2} \Delta_n^{\rho} L_n^{4\eta}. \end{aligned}$$

Taking $x \asymp \Delta_n^{\rho(1-\varpi)/2} L_n^{2\eta}$ where $0 < \varpi < 1$, the right-hand side is bounded by $\exp\{-K \Delta_n^{-\rho\varpi} L_n^{4\eta}\} + K \Delta_n^{\rho\varpi} \leq K' \Delta_n^{\rho\varpi}$. Consequently, we have

$$\mathbb{P} \left(\max_{1 \leq j \leq m_n} \left| \frac{1}{\bar{k}_n} \sum_{i=1}^{\bar{k}_n} \tilde{U}_{i,j}^2 - 1 \right| > K \Delta_n^{\rho(1-\varpi)/2} L_n^{1+2\eta} \middle| \mathcal{F}^{(0)} \right) \leq K' \Delta_n^{\rho\varpi}. \quad (\text{C.11})$$

Similarly, noting that $\bar{k}_n^{-1} \sum_{i=1}^{\bar{k}_n} \mathbb{E}[\tilde{U}_{i,j} | \mathcal{F}^{(0)}] = 0$ and by (C.10) together with a maximal inequality, we have $\mathbb{E}[\max_{1 \leq i \leq \bar{k}_n} \max_{1 \leq j \leq m_n} \tilde{U}_{i,j}^2 | \mathcal{F}^{(0)}] \leq K \Delta_n^{-\rho/2} L_n^{2\eta}$.

Applying Lemma 8 in Chernozhukov et al. (2015) again, we can obtain

$$\mathbb{E} \left[\max_{1 \leq j \leq m_n} \left| \frac{1}{\bar{k}_n} \sum_{i=1}^{\bar{k}_n} \tilde{U}_{i,j} \right| \middle| \mathcal{F}^{(0)} \right] \leq K(\Delta_n^{\rho/2} \sqrt{L_n} + \Delta_n^{3\rho/4} L_n^{1+\eta}). \quad (\text{C.12})$$

Then the Fuk–Nagaev type inequality implies that for every $x > 0$,

$$\begin{aligned} & \mathbb{P}\left(\max_{1 \leq j \leq m_n} \left| \frac{1}{\bar{k}_n} \sum_{i=1}^{\bar{k}_n} \tilde{U}_{i,j} \right| > K(\Delta_n^{\rho/2} L_n^{1+\eta}) + x \mid \mathcal{F}^{(0)}\right) \\ & \leq \exp\{-K'x^2 \Delta_n^{-\rho}\} + K'x^{-4} \Delta_n^{3\rho} L_n^{4\eta}. \end{aligned}$$

Taking $x \asymp \Delta_n^{\rho/4} L_n^\eta$, the right-hand side is bounded by $\exp\{-K \Delta_n^{-\rho/2} L_n^{2\eta}\} + K \Delta_n^{2\rho} \leq K' \Delta_n^{2\rho}$. Consequently, we have

$$\mathbb{P}\left(\max_{1 \leq j \leq m_n} \left| \frac{1}{\bar{k}_n} \sum_{i=1}^{\bar{k}_n} \tilde{U}_{i,j} \right| > K \Delta_n^{\rho/4} L_n^{1+\eta} \mid \mathcal{F}^{(0)}\right) \leq K' \Delta_n^{2\rho}. \quad (\text{C.13})$$

Combining (C.9), (C.11) and (C.13), noting that $\rho(1-\varpi)/2 < \rho/2$, by the law of iterated expectation, for all $\varpi \geq \epsilon/\rho$, we obtain

$$\mathbb{P}\left(\max_{1 \leq j \leq m_n} \left| \frac{\tilde{\sigma}_{n,j}}{\sigma_{n,j}} - 1 \right| > K \Delta_n^{\rho(1-\varpi)/2} L_n^{1+2\eta}\right) \leq K' \Delta_n^\epsilon.$$

Also note that $|a-1| \leq x/(x+1)$ implies $|a^{-1}-1| \leq x$, combining the above inequality with (C.8), we conclude that for $\varpi > (\epsilon/\rho) \vee (1-2\gamma/\rho)$,

$$\mathbb{P}\left(\max_{1 \leq j \leq m_n} \left| \frac{\sigma_{n,j}}{\hat{\sigma}_{n,j}} - 1 \right| > K \Delta_n^{\rho(1-\varpi)/2} L_n^{1+2\eta}\right) \leq K' \Delta_n^\epsilon. \quad (\text{C.14})$$

Moreover, recall (C.12) and the definition of $\tilde{U}_{i,j}$, by the law of iterated expectation and the Markov inequality, for all $\varpi < 1-4\epsilon/\rho$, we can show

$$\mathbb{P}\left(\max_{1 \leq j \leq m_n} \left| \frac{1}{\sqrt{k_{n,j}}} \sum_{i=1}^{k_{n,j}} \tilde{Y}_{i,j} \right| > K \Delta_n^{\rho(\varpi-1)/4} \sqrt{L_n}\right) \leq K' \Delta_n^\epsilon, \quad (\text{C.15})$$

Combining (C.4), (C.7), (C.14), and (C.15), by the Markov inequality, the desired inequality (C.2) follows by taking

$$\delta_n \asymp \Delta_n^{\rho(1-\varpi)/2} L_n^{1+2\eta} \times \Delta_n^{\rho(\varpi-1)/4} \sqrt{L_n} = \Delta_n^{\rho(1-\varpi)/4} L_n^{3/2+2\eta},$$

where $(\epsilon/\rho) \vee (1-2\gamma/\rho) < \varpi < 1-4\epsilon/\rho$, such ϖ exists since $\epsilon/\rho < 1/6$ and $2\epsilon < \gamma$. Note that the choice of sequence δ_n satisfies $\delta_n \rightarrow 0$ and $\delta_n \sqrt{L_n} \leq K \Delta_n^\epsilon$.

This completes the proof of Step 1.

STEP 2. For each $1 \leq i \leq \bar{k}_n$ and $1 \leq j \leq 2m_n$, we define $\tilde{U}_{i,j}^\dagger$ as

$$\tilde{U}_{i,j}^\dagger \equiv \tilde{U}_{i,j} \mathbb{1}_{\{1 \leq j \leq m_n\}} - \tilde{U}_{i,j-m_n} \mathbb{1}_{\{m_n+1 \leq j \leq 2m_n\}}.$$

Observing that by the definition of \tilde{T}_n^* and $\tilde{U}_{i,j}^\dagger$, we can rewrite

$$\tilde{T}_n^* \equiv \max_{1 \leq j \leq m_n} \left| \frac{1}{\sqrt{k_{n,j}}} \sum_{i=1}^{k_{n,j}} \frac{\tilde{Y}_{i,j}}{\sigma_{n,j}} \right| = \max_{1 \leq j \leq m_n} \left| \frac{1}{\sqrt{\bar{k}_n}} \sum_{i=1}^{\bar{k}_n} \tilde{U}_{i,j} \right| = \max_{1 \leq j \leq 2m_n} \frac{1}{\sqrt{\bar{k}_n}} \sum_{i=1}^{\bar{k}_n} \tilde{U}_{i,j}^\dagger.$$

Recall that $(\tilde{U}_{i,j})_{1 \leq i \leq \bar{k}_n, 1 \leq j \leq m_n}$ are $\mathcal{F}^{(0)}$ -conditionally independent, centered random variables. Let $(\tilde{Z}_{i,j})_{1 \leq i \leq \bar{k}_n, 1 \leq j \leq m_n}$ be a sequence of $\mathcal{F}^{(0)}$ -conditionally independent, centered Gaussian random variables with conditional variance given by $\mathbb{E}[\tilde{Z}_{i,j}^2 | \mathcal{F}^{(0)}] = \mathbb{E}[\tilde{U}_{i,j}^2 | \mathcal{F}^{(0)}] = v_{i,j}$. Further, for each $1 \leq i \leq \bar{k}_n$ and $1 \leq j \leq 2m_n$, let

$$\tilde{Z}_{i,j}^\dagger \equiv \tilde{Z}_{i,j} \mathbb{1}\{1 \leq j \leq m_n\} - \tilde{Z}_{i,j-m_n} \mathbb{1}\{m_n + 1 \leq j \leq 2m_n\},$$

which implies $\mathbb{E}[\tilde{Z}_{i,j}^\dagger \tilde{Z}_{i',j'}^\dagger | \mathcal{F}^{(0)}] = \mathbb{E}[\tilde{U}_{i,j}^\dagger \tilde{U}_{i',j'}^\dagger | \mathcal{F}^{(0)}]$ for all $1 \leq i, i' \leq \bar{k}_n$ and $1 \leq j \leq 2m_n$. The proof of this part relies on a conditional version of Gaussian approximations for maxima of sums, see [Chernozhukov et al. \(2013\)](#).

Generally, the bound in the conditional approximation may depend on ζ , hence some specific random variable $K^{(0)}$ involved in $\mathcal{F}^{(0)}$. In our case, since by Assumption 5(i), ζ takes value in a compact set, the bound obtained in the approximation can be universal. This universality property ensures that, after applying the law of iterated expectation, the bound obtained from the Gaussian approximation remains the same.

Note that Assumption 5(ii) implies, for $p \in \{3, 4\}$, and $1 \leq j \leq 2m_n$,

$$\frac{1}{\bar{k}_n} \sum_{i=1}^{\bar{k}_n} \mathbb{E}[|\tilde{U}_{i,j}^\dagger|^p | \mathcal{F}^{(0)}] \leq K_p \frac{1}{k_{n,j}} \sum_{i=1}^{k_{n,j}} G_p(\zeta_{\tau(i,j)}) / \sigma_{n,j}^p \leq K_p.$$

Combining with Assumption 5(iii) and (C.10), by Proposition 2.1 in [Chernozhukov et al. \(2017\)](#), we obtain for all $\epsilon < \rho/6$ that

$$\begin{aligned} & \sup_{x \in \mathbb{R}} \left| \mathbb{P}(\tilde{T}_n^* \leq x | \mathcal{F}^{(0)}) - \mathbb{P}\left(\max_{1 \leq j \leq 2m_n} \frac{1}{\sqrt{\bar{k}_n}} \sum_{i=1}^{\bar{k}_n} \tilde{Z}_{i,j}^\dagger \leq x \mid \mathcal{F}^{(0)} \right) \right| \\ & \leq K(\Delta_n^{\rho/6} L_n^{7/6+\eta/3} + \Delta_n^{\rho/6} L_n^{1+2\eta/3}) \leq K \Delta_n^\epsilon. \end{aligned}$$

For $1 \leq j \leq m_n$, define $Z_j \equiv \bar{k}_n^{-1/2} \sum_{i=1}^{\bar{k}_n} \tilde{Z}_{i,j}$. Recalling the definition of $\tilde{Z}_{i,j}$ and $\sigma_{n,j}$, we conclude

$$(Z_1, Z_2, \dots, Z_{m_n})^\top | \mathcal{F}^{(0)} \sim \mathcal{N}(0, I_{m_n}).$$

Since the right hand side is a pivot, $(Z_j)_{1 \leq j \leq m_n}$ remains standard Gaussian uncon-

ditionally, hence satisfies the requirement in the assertion. Note that by construction we have

$$\max_{1 \leq j \leq 2m_n} \frac{1}{\sqrt{\bar{k}_n}} \sum_{i=1}^{\bar{k}_n} \tilde{Z}_{i,j}^\dagger = \max_{1 \leq j \leq m_n} \left| \frac{1}{\sqrt{\bar{k}_n}} \sum_{i=1}^{\bar{k}_n} \tilde{Z}_{i,j} \right| = \max_{1 \leq j \leq m_n} |Z_j|.$$

Equation (C.3) then follows by applying the law of iterated expectation. This completes the proof of our second step.

STEP 3. We are now ready to prove the assertion of Theorem 4.1. Combining the results in (C.2) and (C.3), we observe that

$$\begin{aligned} & \sup_{x \in \mathbb{R}} \left(\mathbb{P}(\widehat{T}_n^* \leq x) - \mathbb{P}\left(\max_{1 \leq j \leq m_n} |Z_j| \leq x\right) \right) \\ & \leq \mathbb{P}(|\widehat{T}_n^* - \widetilde{T}_n^*| > \delta_n) + \sup_{x \in \mathbb{R}} \left(\mathbb{P}(\widetilde{T}_n^* \leq x + \delta_n) - \mathbb{P}\left(\max_{1 \leq j \leq m_n} |Z_j| \leq x + \delta_n\right) \right) \\ & \quad + \sup_{x \in \mathbb{R}} \mathbb{P}\left(x < \max_{1 \leq j \leq m_n} |Z_j| \leq x + \delta_n\right) \\ & \leq K \Delta_n^\epsilon, \end{aligned}$$

where the last term is bounded by $K \Delta_n^\epsilon$ using the anti-concentration inequality (see Corollary 2.1 in Chernozhukov et al. (2015)), together with the fact that

$$\mathbb{E}\left[\max_{1 \leq j \leq m_n} |Z_j|\right] \leq K \sqrt{L_n},$$

and $\delta_n \sqrt{L_n} \leq K \Delta_n^\epsilon$ by construction of δ_n . Similarly, we can show

$$\sup_{x \in \mathbb{R}} \left(\mathbb{P}\left(\max_{1 \leq j \leq m_n} |Z_j| \leq x\right) - \mathbb{P}(\widehat{T}_n^* \leq x) \right) \leq K \Delta_n^\epsilon.$$

This completes the proof of required statement.

Q.E.D.

C.1.2 Proof of Theorem 4.2

For notation simplicity, we suppress the dependence on χ and write $\hat{q}_{n,j}(\chi)$ as $\hat{q}_{n,j}$ and $q_t(\chi)$ as q_t . Further denote $q_{n,j} \equiv q_{\tau(1,j)}$ and $f_{n,j}(x) \equiv f_{\tau(1,j)}(x)$. By a standard localization procedure, we can strengthen Assumption 4 by assuming $T_1 = \infty$, $\mathcal{K}_m = \mathcal{K}$, and $K_m = K$ for some fixed compact set \mathcal{K} and positive constant $K > 0$. That is, it suffices to prove the results under Assumption 6.

Assumption 6. *There exist a positive constant K , and a compact subset $\mathcal{K} \subset \mathcal{Z}$ such that: (i) ζ takes value in \mathcal{K} ; for all $s, t \in \mathcal{T}_{n,j}$ where $1 \leq j \leq m_n$, and for each*

$p > 0$, $\mathbb{E}[\|\zeta_t - \zeta_s\|^p] \leq K_p |t - s|^{p/2}$ for some constant K_p ; (ii) for each $x \in \mathbb{R}$, for all $z, z' \in \mathcal{K}$, $|F(z, x) - F(z', x)| \vee |\partial_x F(z, x) - \partial_x F(z', x)| \leq K \|z - z'\|$; (iii) for each $t \in [0, T]$ and x in some neighborhood of q_t , $f_t(x) + f_t^{-1}(x) + |\partial_x f_t(x)| < K$; (iv) $\max_{1 \leq i \leq n} |R_{n,i}| = o_p(\Delta_n^r)$ for some $r > 0$.

The proof of Theorem 4.2 is based on a uniform Bahadur type representation of infill sample quantiles, where the approximation error can be controlled uniformly, as shown in the following lemma.

Lemma C.1 (Uniform Bahadur Representation). *Suppose Assumption 6 holds.*

For $1 \leq j \leq m_n$, denote

$$\tilde{q}_{n,j} \equiv q_{n,j} + \frac{\sqrt{\chi(1-\chi)}}{f_{n,j}(q_{n,j})} \frac{1}{k_{n,j}} \sum_{i=1}^{k_{n,j}} \frac{F(\zeta_{\tau(i,j)}, q_{n,j}) - \mathbb{1}\{\mathcal{Y}(\zeta_{\tau(i,j)}, \varepsilon_{n,t(i,j)}) \leq q_{n,j}\}}{\sqrt{F(\zeta_{\tau(i,j)}, q_{n,j})(1-F(\zeta_{\tau(i,j)}, q_{n,j}))}}.$$

Then we have for each $\chi \in (0, 1)$, and for some positive ϵ and γ ,

$$\mathbb{P}\left(\max_{1 \leq j \leq m_n} \sqrt{k_{n,j}} |\hat{q}_{n,j} - \tilde{q}_{n,j}| > K \Delta_n^\epsilon\right) \leq K' \Delta_n^\epsilon.$$

PROOF OF LEMMA C.1. We prove the validity of the assertion for all positive ϵ and γ such that

$$\epsilon + \gamma < \frac{\rho}{4} \wedge \left(\frac{1}{2} - \rho\right) \wedge \left(r - \frac{\rho}{2}\right).$$

Let $\tilde{Y}_{i,j} \equiv \mathcal{Y}(\zeta_{\tau(i,j)}, \varepsilon_{n,t(i,j)})$, within each block j reindex the sequence $(\tilde{Y}_{i,j})_{1 \leq i \leq k_{n,j}}$ in the non-decreasing order and denote as $\tilde{Y}_{1,j}^o \leq \dots \leq \tilde{Y}_{k_{n,j},j}^o$. Note that in each block, there are at least $\lceil k_{n,j} \chi \rceil$ of $\tilde{Y}_{i,j}$ no larger than $Y_{\lceil k_{n,j} \chi \rceil, j}^o + \max_{i \in \mathcal{I}_{n,j}} |R_{n,i}|$, which implies $\tilde{Y}_{\lceil k_{n,j} \chi \rceil, j}^o \leq Y_{\lceil k_{n,j} \chi \rceil, j}^o + \max_{i \in \mathcal{I}_{n,j}} |R_{n,i}|$. Similarly, there are at least $k_{n,j} - \lceil k_{n,j} \chi \rceil$ of $\tilde{Y}_{i,j}$ no smaller than $Y_{\lceil k_{n,j} \chi \rceil, j}^o - \max_{i \in \mathcal{I}_{n,j}} |R_{n,i}|$, which implies $\tilde{Y}_{\lceil k_{n,j} \chi \rceil, j}^o \geq Y_{\lceil k_{n,j} \chi \rceil, j}^o - \max_{i \in \mathcal{I}_{n,j}} |R_{n,i}|$. Therefore, Assumption 6(iv) implies that

$$\max_{1 \leq j \leq m_n} \sqrt{k_{n,j}} |\hat{g}_{n,j} - \tilde{Y}_{\lceil k_{n,j} \chi \rceil, j}^o| \leq K \Delta_n^{-\rho/2} \max_{1 \leq i \leq n} |R_{n,i}| = o_p(\Delta_n^{r-\rho/2}) = o_p(\Delta_n^{\epsilon+\gamma}). \quad (\text{C.16})$$

For each $1 \leq j \leq m_n$, let $\tilde{F}_{n,j}(x) \equiv k_{n,j}^{-1} \sum_{i=1}^{k_{n,j}} \mathbb{1}\{\tilde{Y}_{i,j} \leq x\}$ be the empirical distribution function of $(\tilde{Y}_{i,j})_{1 \leq i \leq k_{n,j}}$. The rest of the proof is divided into three steps. In Step 1, we show that the averaged distribution function $k_{n,j}^{-1} \sum_{i=1}^{k_{n,j}} F(\zeta_{\tau(i,j)}, \cdot)$ can be well approximated by the empirical distribution function $\tilde{F}_{n,j}(\cdot)$ in some small neighborhood of true quantile $q_{n,j}$, uniformly over $1 \leq j \leq m_n$. In Step 2, we show

that with large probability, the sample quantile $\tilde{Y}_{[k_{n,j}\chi],j}^o$ falls in the neighborhood described in Step 1 for all $1 \leq j \leq m_n$. Step 3 derives the asserted statement.

STEP 1. For $1 \leq j \leq m_n$, denote

$$S_{n,j}(x) \equiv \tilde{F}_{n,j}(x) - \tilde{F}_{n,j}(q_{n,j}) - \frac{1}{k_{n,j}} \sum_{i=1}^{k_{n,j}} (F(\zeta_{\tau(i,j)}, x) - \chi). \quad (\text{C.17})$$

For any set $A \subseteq \mathbb{R}$, denote $\bar{S}_{n,j}(A) \equiv \sup_{x \in A} |S_{n,j}(x)|$. Let $\varkappa_{1,n} \asymp \Delta_n^{\rho/2} L_n$ be a positive real sequence, and let $\varkappa_{2,n} \asymp \Delta_n^{-\rho/4}$ be a positive integer sequence, denote interval $\bar{I}_{n,j} \equiv (q_{n,j} - \varkappa_{1,n}, q_{n,j} + \varkappa_{1,n})$. For any integer ℓ , let $\psi_{n,j}(\ell) \equiv q_{n,j} + \varkappa_{1,n} \varkappa_{2,n}^{-1} \ell$, denote interval $I_{n,j}(\ell) \equiv [\psi_{n,j}(\ell), \psi_{n,j}(\ell + 1)]$, then we have $\bar{I}_{n,j} \subseteq \bigcup_{\ell=-\varkappa_{2,n}}^{\varkappa_{2,n}-1} I_{n,j}(\ell)$. Note that both $\tilde{F}_{n,j}(\cdot)$ and $F(z, \cdot)$ are nondecreasing functions, we have for $x \in I_{n,j}(\ell)$,

$$\begin{aligned} S_{n,j}(x) &\leq \tilde{F}_{n,j}(\psi_{n,j}(\ell + 1)) - \tilde{F}_{n,j}(q_{n,j}) - \frac{1}{k_{n,j}} \sum_{i=1}^{k_{n,j}} (F(\zeta_{\tau(i,j)}, \psi_{n,j}(\ell)) - \chi) \\ &\leq S_{n,j}(\psi_{n,j}(\ell + 1)) + \vartheta_{n,j}(\ell), \end{aligned}$$

where

$$\vartheta_{n,j}(\ell) \equiv \frac{1}{k_{n,j}} \sum_{i=1}^{k_{n,j}} F(\zeta_{\tau(i,j)}, \psi_{n,j}(\ell + 1)) - \frac{1}{k_{n,j}} \sum_{i=1}^{k_{n,j}} F(\zeta_{\tau(i,j)}, \psi_{n,j}(\ell)).$$

Similarly, we also have $S_{n,j}(x) \geq S_{n,j}(\psi_{n,j}(\ell)) - \vartheta_{n,j}(\ell)$. Denote

$$\bar{\vartheta}_{n,j} \equiv \max_{-\varkappa_{2,n} \leq \ell \leq \varkappa_{2,n}-1} \vartheta_{n,j}(\ell).$$

Then it follows the definition of $\bar{I}_{n,j}$ that

$$\bar{S}_{n,j}(\bar{I}_{n,j}) \leq \bar{S}_{n,j} \left(\bigcup_{\ell=-\varkappa_{2,n}}^{\varkappa_{2,n}-1} I_{n,j}(\ell) \right) \leq \max_{-\varkappa_{2,n} \leq \ell \leq \varkappa_{2,n}} |S_{n,j}(\psi_{n,j}(\ell))| + \bar{\vartheta}_{n,j}. \quad (\text{C.18})$$

For the second term, note that $|\psi_{n,j}(\ell) - q_{n,j}| \leq \varkappa_{1,n} \rightarrow 0$ for $|\ell| \leq \varkappa_{2,n}$. Then by Assumption 6(iii) and the mean value theorem, recall that $\gamma < \rho/4 - \epsilon$, we have for n sufficiently large,

$$\begin{aligned} \max_{1 \leq j \leq m_n} \sqrt{k_{n,j}} \bar{\vartheta}_{n,j} &\leq K \max_{1 \leq j \leq m_n} \max_{-\varkappa_{2,n} \leq \ell \leq \varkappa_{2,n}-1} \sqrt{k_{n,j}} |\psi_{n,j}(\ell + 1) - \psi_{n,j}(\ell)| \\ &= K \varkappa_{1,n} \varkappa_{2,n}^{-1} \max_{1 \leq j \leq m_n} \sqrt{k_{n,j}} \\ &\leq K \Delta_n^{\rho/4} L_n \leq K \Delta_n^{\epsilon+\gamma}. \end{aligned} \quad (\text{C.19})$$

For the first term in the right-hand side of (C.18), first consider a fixed $1 \leq j \leq m_n$.

For each $-\varkappa_{2,n} \leq \ell \leq \varkappa_{2,n}$, let $(\xi_{i,j}(\ell))_{1 \leq i \leq k_{n,j}}$ be a sequence of $\mathcal{F}^{(0)}$ -conditionally independent, Bernoulli random variables with parameter

$$(|F(\zeta_{\tau(i,j)}, \psi_{n,j}(\ell)) - F(\zeta_{\tau(i,j)}, q_{n,j})|)_{1 \leq i \leq k_{n,j}},$$

respectively. Let $\Xi_{n,j}(\ell) \equiv \sum_{i=1}^{k_{n,j}} \xi_{i,j}(\ell)$ denote their convolution. Note that by construction and (C.17),

$$k_{n,j} |S_{n,j}(\psi_{n,j}(\ell))| \stackrel{\mathcal{L}|\mathcal{F}^{(0)}}{=} \left| \Xi_{n,j}(\ell) - \sum_{i=1}^{k_{n,j}} (F(\zeta_{\tau(i,j)}, \psi_{n,j}(\ell)) - \chi) \right|.$$

In view of above equation, by the triangle inequality, we have for all $x \in \mathbb{R}$,

$$\begin{aligned} & \left\{ \max_{1 \leq j \leq m_n} \max_{-\varkappa_{2,n} \leq \ell \leq \varkappa_{2,n}} \sqrt{k_{n,j}} S_{n,j}(\psi_{n,j}(\ell)) \geq x \right\} \\ & \subseteq \left\{ \max_{1 \leq j \leq m_n} \max_{-\varkappa_{2,n} \leq \ell \leq \varkappa_{2,n}} \frac{1}{\sqrt{k_{n,j}}} \left| \Xi_{n,j}(\ell) - \sum_{i=1}^{k_{n,j}} (F(\zeta_{\tau(i,j)}, \psi_{n,j}(\ell)) - F(\zeta_{\tau(i,j)}, q_{n,j})) \right| \geq \frac{x}{2} \right\} \\ & \quad \cup \left\{ \max_{1 \leq j \leq m_n} \frac{1}{\sqrt{k_{n,j}}} \sum_{i=1}^{k_{n,j}} |F(\zeta_{\tau(i,j)}, q_{n,j}) - \chi| \geq \frac{x}{2} \right\} \\ & = \left\{ \max_{1 \leq j \leq m_n} \max_{-\varkappa_{2,n} \leq \ell \leq \varkappa_{2,n}} \mathfrak{B}_{n,j}^{(I)}(\ell) \geq \frac{x}{2} \right\} \cup \left\{ \max_{1 \leq j \leq m_n} \mathfrak{B}_{n,j}^{(II)} \geq \frac{x}{2} \right\}, \quad (\text{C.20}) \end{aligned}$$

where for $1 \leq j \leq m_n$ and $-\varkappa_{2,n} \leq \ell \leq \varkappa_{2,n}$,

$$\begin{aligned} \mathfrak{B}_{n,j}^{(I)}(\ell) & \equiv \frac{1}{\sqrt{k_{n,j}}} \left| \Xi_{n,j}(\ell) - \sum_{i=1}^{k_{n,j}} (F(\zeta_{\tau(i,j)}, \psi_{n,j}(\ell)) - F(\zeta_{\tau(i,j)}, q_{n,j})) \right|, \\ \mathfrak{B}_{n,j}^{(II)} & \equiv \frac{1}{\sqrt{k_{n,j}}} \sum_{i=1}^{k_{n,j}} |F(\zeta_{\tau(i,j)}, q_{n,j}) - \chi|. \end{aligned}$$

For the second term, note that Assumption 6(iii) implies for each $t \in [0, T]$, $f_t(x)$ is Lipschitz in some neighborhood of q_t , and $F(\zeta_t, \cdot)$ has no mass at q_t , hence $F(\zeta_t, q_t) = \chi$ by the definition of q_t . Therefore, we deduce

$$\begin{aligned} \mathbb{P}(|q_t - q_s| > x) & \leq \mathbb{P}(q_t - q_s > x) + \mathbb{P}(q_s - q_t > x) \\ & \leq \mathbb{P}(F(\zeta_t, q_s + x) < \chi) + \mathbb{P}(F(\zeta_s, q_t + x) < \chi) \\ & \leq \mathbb{P}(F(\zeta_s, q_s + x) - K \|\zeta_s - \zeta_t\| < \chi) \\ & \quad + \mathbb{P}(F(\zeta_t, q_t + x) - K \|\zeta_s - \zeta_t\| < \chi) \\ & \leq 2\mathbb{P}(\|\zeta_s - \zeta_t\| > Kx), \quad (\text{C.21}) \end{aligned}$$

where the second line is by the fact that $F(z, x)$ is increasing in x , the third line is by Assumption 6(ii). Also note that by Fubini's theorem $\mathbb{E}[X^p] = \int_0^\infty px^{p-1}\mathbb{P}(X > x)dx$ for nonnegative random variable X . Therefore, it follows Assumption 6(i) and (C.21) that the instantaneous conditional quantile process q is also $1/2$ -Hölder continuous under the L_p -norm. Then by a maximal inequality, we have

$$\left\| \max_{1 \leq j \leq m_n} \mathfrak{B}_{n,j}^{(II)} \right\|_{L_p} \leq K_p m_n^{1/p} \max_{1 \leq j \leq m_n} \sqrt{k_{n,j}} (k_{n,j} \Delta_n)^{1/2} \leq K_p \Delta_n^{(\rho-1)/p+1/2-\rho}.$$

Taking $p > (1-\rho)/(1/2-\rho-\epsilon-\gamma)$, the right-hand side becomes $o(\Delta_n^{\epsilon+\gamma})$. Therefore, by the Markov inequality and the law of iterated expectation, we conclude that

$$\mathbb{P}\left(\max_{1 \leq j \leq m_n} \mathfrak{B}_{n,j}^{(II)} \geq K \Delta_n^\epsilon\right) \leq K' \Delta_n^\epsilon. \quad (\text{C.22})$$

For the first term inside the max operator in the right-hand side of (C.20), by the Bernstein inequality (see, e.g., bound (2.13) under Theorem 3 of Hoeffding (1963)), we have for all $x \in \mathbb{R}^+$,

$$\begin{aligned} & \mathbb{P}(\sqrt{k_{n,j}} \mathfrak{B}_{n,j}^{(I)}(\ell) \geq x \mid \mathcal{F}^{(0)}) \\ & \leq 2 \exp\left\{-\frac{x^2/2}{\sum_{i=1}^{k_{n,j}} |F(\zeta_{\tau(i,j)}, \psi_{n,j}(\ell)) - F(\zeta_{\tau(i,j)}, q_{n,j})| + x}\right\}. \end{aligned} \quad (\text{C.23})$$

According Assumption 6(iv), we can choose and fix a positive constant M_1 such that $\partial_x F(\zeta_t, q_t) < M_1$ for all $t \in [0, T]$. Then by the definition of $\psi_{n,j}(\ell)$, we have

$$\sum_{i=1}^{k_{n,j}} |F(\zeta_{\tau(i,j)}, \psi_{n,j}(\ell)) - F(\zeta_{\tau(i,j)}, q_{n,j})| \leq M_1 k_{n,j} \varkappa_{1,n}. \quad (\text{C.24})$$

Note that the right-hand side bound of above equation is deterministic and does not depend on ℓ . Therefore, combining (C.23) and (C.24), we can conclude that

$$\begin{aligned} & \mathbb{P}\left(\max_{-\varkappa_{2,n} \leq \ell \leq \varkappa_{2,n}} \mathfrak{B}_{n,j}^{(I)}(\ell) \geq M_2 \Delta_n^{\rho/4} L_n \mid \mathcal{F}^{(0)}\right) \\ & \leq \sum_{\ell=-\varkappa_{2,n}}^{\varkappa_{2,n}} \mathbb{P}(\mathfrak{B}_{n,j}^{(I)}(\ell) \geq M_2 \Delta_n^{\rho/4} L_n \mid \mathcal{F}^{(0)}) \\ & \leq 4\varkappa_{2,n} \exp\left\{-\frac{M_2^2 k_{n,j} \Delta_n^{\rho/2} L_n^2 / 2}{M_1 k_{n,j} \varkappa_{1,n} + M_2 \sqrt{k_{n,j}} \Delta_n^{\rho/4} L_n}\right\}. \end{aligned}$$

Let $\mathcal{O}_n(M_1, M_2)$ denote the right-hand side bound of the above display. Note that by the definition of $\varkappa_{1,n}$, as $\Delta_n \rightarrow 0$ (or equivalently, as $n \rightarrow \infty$), we have

$$\frac{\log(\mathcal{O}_n(M_1, M_2))}{\log n} \rightarrow \frac{\rho}{4} - \frac{M_2^2}{2M_1}.$$

Taking $M_2 > \sqrt{2M_1(1 + \rho/4)}$, the above limit is less than -1 . By the property of Harmonic p -series, this implies $\sum_{n=1}^{\infty} \mathcal{O}(M_1, M_2) < \infty$. Then by the Borel–Cantelli lemma, we conclude that

$$\mathbb{P}\left(\limsup_{n \rightarrow \infty} \max_{-\varkappa_{2,n} \leq \ell \leq \varkappa_{2,n}} \mathfrak{B}_{n,j}^{(I)}(\ell) \geq M_2 \Delta_n^{\rho/4} L_n \mid \mathcal{F}^{(0)}\right) = 0.$$

Note that $\gamma < \rho/4 - \epsilon$, then by the law of iterated expectation, we have for n sufficiently large

$$\begin{aligned} & \mathbb{P}\left(\max_{1 \leq j \leq m_n} \max_{-\varkappa_{2,n} \leq \ell \leq \varkappa_{2,n}} \mathfrak{B}_{n,j}^{(I)}(\ell) \geq K \Delta_n^\gamma\right) \\ & \leq \sum_{j=1}^{m_n} \mathbb{P}\left(\max_{-\varkappa_{2,n} \leq \ell \leq \varkappa_{2,n}} \mathfrak{B}_{n,j}^{(I)}(\ell) \geq M_2 \Delta_n^{\rho/4} L_n\right) = 0. \end{aligned} \quad (\text{C.25})$$

Combining (C.18)-(C.22), and (C.25), we conclude that

$$\mathbb{P}\left(\max_{1 \leq j \leq m_n} \sqrt{k_{n,j}} \bar{S}_{n,j}(\bar{I}_{n,j}) \geq K \Delta_n^\gamma\right) \leq K' \Delta_n^\epsilon. \quad (\text{C.26})$$

STEP 2. Recall the definition of $\tilde{Y}_{[k_{n,j}\chi],j}^o$ and $\tilde{F}_{n,j}(\cdot)$, for each $1 \leq j \leq m_n$, we have $\tilde{Y}_{[k_{n,j}\chi],j}^o \leq p_{n,j} - \varkappa_{1,n}$ if and only if $k_{n,j} \tilde{F}_{n,j}(q_{n,j} - \varkappa_{1,n}) \geq [k_{n,j}\chi]$. Therefore,

$$\begin{aligned} & \{\exists 1 \leq j \leq m_n \text{ such that } \tilde{Y}_{[k_{n,j}\chi],j}^o \leq p_{n,j} - \varkappa_{1,n}\} \\ & = \left\{ \max_{1 \leq j \leq m_n} (k_{n,j} \tilde{F}_{n,j}(q_{n,j} - \varkappa_{1,n}) - [k_{n,j}\chi]) \geq 0 \right\}. \end{aligned} \quad (\text{C.27})$$

Let $(\xi'_{i,j})$ be a sequence of $\mathcal{F}^{(0)}$ -conditionally independent, Bernoulli random variables with parameter

$$(F(\zeta_{\tau(i,j)}, q_{n,j} - \varkappa_{1,n}))_{1 \leq i \leq k_{n,j}},$$

respectively. Let $\Xi'_{n,j} \equiv \sum_{i=1}^{k_{n,j}} \xi'_{i,j}$ denote their convolution. By the construction, we have

$$k_{n,j} \tilde{F}_{n,j}(q_{n,j} - \varkappa_{1,n}) \stackrel{\mathcal{L} \mid \mathcal{F}^{(0)}}{=} \Xi'_{n,j}. \quad (\text{C.28})$$

Note that Assumption 6(i)-(iii) imply that

$$\max_{1 \leq i \leq k_{n,j}} |F(\zeta_{\tau(i,j)}, q_{n,j} - \varkappa_{1,n}) - \chi| \leq M \left(\max_{1 \leq i \leq k_{n,j}} \|\zeta_{\tau(i,j)} - \zeta_{\tau(1,j)}\| + \varkappa_{1,n} \right).$$

Observe that in the right-hand side of above display, by Assumption 6(i), we have

$$\begin{aligned} \left\| \max_{1 \leq j \leq m_n} \max_{1 \leq i \leq k_{n,j}} (\zeta_{\tau(i,j)} - \zeta_{\tau(1,j)}) \right\|_{L_p} & \leq K_p m_n^{1/p} \max_{1 \leq j \leq m_n} (k_{n,j} \Delta_n)^{1/2} \\ & \leq K_p \Delta_n^{(\rho-1)/p + (1-\rho)/2}. \end{aligned} \quad (\text{C.29})$$

Taking $p > (1 - \rho)/(1/2 - \rho - \epsilon)$, the right-hand side becomes $o(\varkappa_{1,n} \Delta_n^\epsilon)$. Let $E_{n,1}$ be the event such that

$$E_{n,1} \equiv \left\{ \max_{1 \leq j \leq m_n} \max_{1 \leq i \leq k_{n,j}} \|\zeta_{\tau(i,j)} - \zeta_{\tau(1,j)}\| < \varkappa_{1,n} \right\}.$$

Therefore, from (C.29), by the Markov inequality and the law of iterated expectation, we conclude that $\mathbb{P}(E_{n,1}^c) \leq K \Delta_n^\epsilon$. In view of (C.28), and noting that

$$\max_{1 \leq j \leq m_n} (\lceil k_{n,j} \chi \rceil - k_{n,j} \chi) < 1,$$

we can rewrite

$$\begin{aligned} & \left\{ \max_{1 \leq j \leq m_n} (k_{n,j} \widetilde{F}_{n,j}(q_{n,j} - \varkappa_{1,n}) - \lceil k_{n,j} \chi \rceil) \geq 0 \right\} \cap E_{n,1} \\ & \subseteq \left\{ \max_{1 \leq j \leq m_n} \left(\mathbb{E}'_{n,j} - \sum_{i=1}^{k_{n,j}} F(\zeta_{\tau(i,j)}, q_{n,j} - \varkappa_{1,n}) \right) \right. \\ & \quad \left. \geq 1 - (M + K \Delta_n^{-\rho}) \varkappa_{1,n} \right\} \cap E_{n,1} \\ & \subseteq \left\{ \max_{1 \leq j \leq m_n} \left(\mathbb{E}'_{n,j} - \sum_{i=1}^{k_{n,j}} F(\zeta_{\tau(i,j)}, q_{n,j} - \varkappa_{1,n}) \right) \geq -K \Delta_n^{-\rho} \varkappa_{1,n} \right\} \cap E_{n,1}. \end{aligned}$$

For the term inside the max operator of above display, it follows the Bernstein inequality that

$$\begin{aligned} & \mathbb{P} \left(\left\{ \mathbb{E}'_{n,j} - \sum_{i=1}^{k_{n,j}} F(\zeta_{\tau(i,j)}, q_{n,j} - \varkappa_{1,n}) \geq -K \Delta_n^{-\rho} \varkappa_{1,n} \right\} \cap E_{n,1} \mid \mathcal{F}^{(0)} \right) \\ & \leq \exp \left\{ - \frac{(-K \Delta_n^{-\rho} \varkappa_{1,n})^2}{2 \left(\sum_{i=1}^{k_{n,j}} F(\zeta_{\tau(i,j)}, q_{n,j} - \varkappa_{1,n}) - K \Delta_n^{-\rho} \varkappa_{1,n} \right)} \right\} \\ & \leq \exp \left\{ - \frac{K \Delta_n^{-2\rho} \varkappa_{1,n}^2}{2 k_{n,j} \chi} \right\}, \end{aligned}$$

where the last line is by the fact that

$$\left| \sum_{i=1}^{k_{n,j}} F(\zeta_{\tau(i,j)}, q_{n,j} - \varkappa_{1,n}) - k_{n,j} \chi \right| \leq K \Delta_n^{-\rho} \varkappa_{1,n},$$

on $E_{n,1}$. Note that the expression inside the exponential operator has an order of $\Delta_n^{-2\rho} \varkappa_{1,n}^2 / \Delta_n^{-\rho} \asymp L_n^2$, observing that $\int_0^\infty \exp\{-\log(x)^2\} dx < \infty$, which implies the right-hand side is summable. Then by the Borel–Cantelli lemma, we conclude that on the event $E_{n,1}$,

$$\mathbb{P} \left(\left\{ \limsup_{n \rightarrow \infty} k_{n,j} \widetilde{F}_{n,j}(q_{n,j} - \varkappa_{1,n}) \geq \lceil k_{n,j} \chi \rceil \right\} \cap E_{n,1} \mid \mathcal{F}^{(0)} \right) = 0.$$

Then by the law of iterated expectation, we have for n sufficiently large

$$\begin{aligned} & \mathbb{P}\left(\left\{\max_{1 \leq j \leq m_n} (\tilde{F}_{n,j}(q_{n,j} - \varkappa_{1,n}) - \lceil k_{n,j} \chi \rceil) \geq 0\right\} \cap E_{n,1}\right) \\ & \leq \sum_{j=1}^{m_n} \mathbb{P}(\{\tilde{F}_{n,j}(q_{n,j} - \varkappa_{1,n}) - \lceil k_{n,j} \chi \rceil \geq 0\} \cap E_{n,1}) = 0. \end{aligned} \quad (\text{C.30})$$

Combining (C.27) and (C.30) yields for n sufficiently large,

$$\begin{aligned} & \mathbb{P}(\{\exists 1 \leq j \leq m_n \text{ such that } \tilde{Y}_{\lceil k_{n,j} \chi \rceil, j}^o \leq p_{n,j} - \varkappa_{1,n}\}) \\ & \leq \mathbb{P}\left(\left\{\max_{1 \leq j \leq m_n} (\tilde{F}_{n,j}(q_{n,j} - \varkappa_{1,n}) - \lceil k_{n,j} \chi \rceil) \geq 0\right\} \cap E_{n,1}\right) + \mathbb{P}(E_{n,1}^c) \\ & \leq K \Delta_n^\epsilon. \end{aligned} \quad (\text{C.31})$$

Following a similar argument as driving (C.31), we can also show

$$\mathbb{P}(\{\exists 1 \leq j \leq m_n \text{ such that } \tilde{Y}_{\lceil k_{n,j} \chi \rceil, j}^o \geq p_{n,j} + \varkappa_{1,n}\}) \leq K \Delta_n^\epsilon. \quad (\text{C.32})$$

Combining (C.31) and (C.32), recall the definition of $\bar{I}_{n,j}$, we conclude that

$$\mathbb{P}(\tilde{Y}_{\lceil k_{n,j} \chi \rceil, j}^o \in \bar{I}_{n,j} \text{ for all } 1 \leq j \leq m_n) \geq 1 - K \Delta_n^\epsilon. \quad (\text{C.33})$$

Now, let $E_{n,2}$ be the event such that

$$\begin{aligned} E_{n,2} & \equiv \left\{ \max_{1 \leq j \leq m_n} \sqrt{k_{n,j}} \bar{S}_{n,j}(\bar{I}_{n,j}) \leq K \Delta_n^\gamma \right\} \\ & \quad \cap \{\tilde{Y}_{\lceil k_{n,j} \chi \rceil, j}^o \in \bar{I}_{n,j} \text{ for all } 1 \leq j \leq m_n\}. \end{aligned}$$

Then (C.26) and (C.33) imply $\mathbb{P}(E_{n,2}^c) \leq K' \Delta_n^\epsilon$. Recall that Assumption 6(iii) implies $\partial_x f_{n,j}(x)$ is uniformly bounded over $x \in \bigcup_{j=1}^{m_n} \bar{I}_{n,j}$ for n sufficiently large.

On the event $E_{n,2}$, by the second order Taylor expansion, we have

$$\begin{aligned} & \max_{1 \leq j \leq m_n} \max_{1 \leq i \leq k_{n,j}} \sqrt{k_{n,j}} \left| F(\zeta_{\tau(i,j)}, \tilde{Y}_{\lceil k_{n,j} \chi \rceil, j}^o) - F(\zeta_{\tau(i,j)}, q_{n,j}) \right. \\ & \quad \left. - (\tilde{Y}_{\lceil k_{n,j} \chi \rceil, j}^o - q_{n,j}) f_{\tau(i,j)}(q_{n,j}) \right| \\ & \leq K \Delta_n^{-\rho/2} \varkappa_{1,n}^2 \leq K \Delta_n^\gamma. \end{aligned} \quad (\text{C.34})$$

It follows Assumption 6(ii) and (C.29) that

$$\begin{aligned} & \max_{1 \leq j \leq m_n} \max_{1 \leq i \leq k_{n,j}} \sqrt{k_{n,j}} (|F(\zeta_{\tau(i,j)}, q_{n,j}) - \chi| + |f_{\tau(i,j)}(q_{n,j}) - f_{n,j}(q_{n,j})|) \\ & = o_p(\Delta_n^{\epsilon+\gamma}). \end{aligned} \quad (\text{C.35})$$

Combining (C.34) and (C.35) yields

$$\mathbb{P}\left(\left\{\max_{1 \leq j \leq m_n} \max_{1 \leq i \leq k_{n,j}} \sqrt{k_{n,j}} \left| F(\zeta_{\tau(i,j)}, \tilde{Y}_{\lceil k_{n,j} \chi \rceil, j}^o) - F(\zeta_{\tau(i,j)}, q_{n,j}) \right| \right.\right.$$

$$\begin{aligned}
& -(\tilde{Y}_{[k_{n,j}\chi],j}^o - q_{n,j})f_{n,j}(q_{n,j}) \Big| \geq K\Delta_n^\gamma \Big\} \cap E_{n,2}) \\
& \leq K'\Delta_n^\epsilon. \tag{C.36}
\end{aligned}$$

On the event $E_{n,2}$, by the definition of $\bar{S}_{n,j}$, we have

$$\begin{aligned}
& \max_{1 \leq j \leq m_n} \sqrt{k_{n,j}} \left| \frac{[k_{n,j}\chi]}{k_{n,j}} - \tilde{F}_{n,j}(q_{n,j}) - \frac{1}{k_{n,j}} \sum_{i=1}^{k_{n,j}} (F(\zeta_{\tau(i,j)}, \tilde{Y}_{[k_{n,j}\chi],j}^o) \right. \\
& \quad \left. - F(\zeta_{\tau(i,j)}, q_{n,j})) \right| \leq \max_{1 \leq j \leq m_n} \sqrt{k_{n,j}} \bar{S}_{n,j}(\bar{I}_{n,j}) \\
& \leq K\Delta_n^\gamma. \tag{C.37}
\end{aligned}$$

By simple algebra we have $\sqrt{k_{n,j}}|[k_{n,j}\chi]/k_{n,j} - \chi| \leq k_{n,j}^{-1/2} \leq K\Delta_n^{\rho/2} \leq K\Delta_n^\gamma$.

Combing with (C.35)-(C.37), by the triangle inequality, we conclude that

$$\begin{aligned}
& \mathbb{P} \left(\max_{1 \leq j \leq m_n} \sqrt{k_{n,j}} \left| \tilde{Y}_{[k_{n,j}\chi],j}^o - q_{n,j} - \frac{1}{f_{n,j}(q_{n,j})} \left(\frac{1}{k_{n,j}} \sum_{i=1}^{k_{n,j}} F(\zeta_{\tau(i,j)}, q_{n,j}) \right. \right. \right. \\
& \quad \left. \left. \left. - \tilde{F}_{n,j}(q_{n,j}) \right) \right| \geq K\Delta_n^\gamma \right) \\
& \leq \mathbb{P} \left(\left\{ \max_{1 \leq j \leq m_n} \sqrt{k_{n,j}} \left| \tilde{Y}_{[k_{n,j}\chi],j}^o - q_{n,j} - \frac{1}{f_{n,j}(q_{n,j})} \right. \right. \right. \\
& \quad \left. \left. \left. \times \left(\frac{1}{k_{n,j}} \sum_{i=1}^{k_{n,j}} F(\zeta_{\tau(i,j)}, q_{n,j}) - \tilde{F}_{n,j}(q_{n,j}) \right) \right| \geq K\Delta_n^\gamma \right\} \cap E_{n,2} \right) \\
& + \mathbb{P}(E_{n,2}^c) \\
& \leq K\Delta_n^\epsilon. \tag{C.38}
\end{aligned}$$

STEP 3. Combining (C.16) and (C.38), by the triangle inequality and the Markov inequality, we obtain

$$\begin{aligned}
& \mathbb{P} \left(\max_{1 \leq j \leq m_n} \sqrt{k_{n,j}} \left| \hat{q}_{n,j} - q_{n,j} - \frac{1}{f_{n,j}(q_{n,j})} \frac{1}{k_{n,j}} \sum_{i=1}^{k_{n,j}} (F(\zeta_{\tau(i,j)}, q_{n,j}) \right. \right. \\
& \quad \left. \left. - \mathbb{1}_{\{\tilde{Y}_{i,j} \leq q_{n,j}\}}) \right| \geq K\Delta_n^\gamma \right) \\
& \leq \mathbb{P} \left(\max_{1 \leq j \leq m_n} \sqrt{k_{n,j}} \left| \tilde{Y}_{[k_{n,j}\chi],j}^o - q_{n,j} - \frac{1}{f_{n,j}(q_{n,j})} \right. \right. \\
& \quad \left. \left. \times \left(\frac{1}{k_{n,j}} \sum_{i=1}^{k_{n,j}} F(\zeta_{\tau(i,j)}, q_{n,j}) - \tilde{F}_{n,j}(q_{n,j}) \right) \right| \geq K\Delta_n^\gamma \right) \\
& + \mathbb{P} \left(\max_{1 \leq j \leq m_n} \sqrt{k_{n,j}} |\hat{g}_{n,j} - \tilde{Y}_{[k_{n,j}\chi],j}^o| \geq K\Delta_n^\gamma \right) \\
& \leq K\Delta_n^\epsilon. \tag{C.39}
\end{aligned}$$

Recall $|\sqrt{a} - 1| \leq |a - 1|$ for positive a , note that by (C.29) and Assumption 6(ii), we have

$$\begin{aligned}
& \max_{1 \leq j \leq m_n} \max_{1 \leq i \leq k_{n,j}} \left| \sqrt{\frac{\chi(1-\chi)}{F(\zeta_{\tau(i,j)}, q_{n,j})(1-F(\zeta_{\tau(i,j)}, q_{n,j}))}} - 1 \right| \\
& \leq \max_{1 \leq j \leq m_n} \max_{1 \leq i \leq k_{n,j}} \left| \frac{F(\zeta_{\tau(1,j)}, q_{n,j})(1-F(\zeta_{\tau(1,j)}, q_{n,j}))}{F(\zeta_{\tau(i,j)}, q_{n,j})(1-F(\zeta_{\tau(i,j)}, q_{n,j}))} - 1 \right| \\
& \leq K \max_{1 \leq j \leq m_n} \max_{1 \leq i \leq k_{n,j}} \|\zeta_{\tau(i,j)} - \zeta_{\tau(1,j)}\| = o_p(\Delta_n^{\epsilon+\gamma}). \tag{C.40}
\end{aligned}$$

Combining (C.39) and (C.40) completes the proof of Lemma C.1. *Q.E.D.*

PROOF OF THEOREM 4.2. We are now ready to prove strong approximation result for the functional quantile estimator $(\hat{q}_{n,t})_{t \in [0, T]}$. With a slightly stronger restriction on ϵ than in the proof of Lemma C.1, we prove the validity of the assertion for all positive ϵ satisfying

$$\epsilon < \frac{\rho}{6} \wedge \left(\frac{1}{2} - \rho\right) \wedge \left(r - \frac{\rho}{2}\right).$$

Correspondingly, let γ be a positive constant satisfying

$$\gamma < \left(\frac{\rho}{4} - \epsilon\right) \wedge \left(\frac{1}{2} - \rho - \epsilon\right) \wedge \left(r - \frac{\rho}{2} - \epsilon\right).$$

By the triangle inequality, we have

$$\begin{aligned}
& \max_{1 \leq j \leq m_n} \sup_{t \in \mathcal{T}_{n,j}} \sqrt{k_{n,j}} |\hat{q}_{n,t} - q_{n,t}| \\
& \leq \max_{1 \leq j \leq m_n} \sup_{t \in \mathcal{T}_{n,j}} \sqrt{k_{n,j}} |q_{n,j} - q_t| + \max_{1 \leq j \leq m_n} \sqrt{k_{n,j}} |\hat{q}_{n,j} - \tilde{q}_{n,j}| \\
& \quad + \max_{1 \leq j \leq m_n} |\tilde{q}_{n,j} - q_{n,j}|. \tag{C.41}
\end{aligned}$$

For the first term, by (C.21) and (C.29), we have

$$\max_{1 \leq j \leq m_n} \sup_{t \in \mathcal{T}_{n,j}} \sqrt{k_{n,j}} |q_{n,j} - q_t| = o_p(\Delta_n^{\epsilon+\gamma}). \tag{C.42}$$

Let $\bar{k}_n \equiv \max_{1 \leq j \leq m_n} k_{n,j}$, then $\bar{k}_n \asymp \Delta_n^{-\rho}$ and $1/K \leq \bar{k}_n/k_{n,j} \leq K$ uniformly for all $1 \leq j \leq m_n$. For each $1 \leq i \leq \bar{k}_n$ and $1 \leq j \leq m_n$, define $\tilde{\mathcal{U}}_{i,j}$ and $v_{i,j}$ as follows:

$$\begin{aligned}
\tilde{\mathcal{U}}_{i,j} & \equiv \sqrt{\frac{\bar{k}_n}{k_{n,j}}} \frac{\sqrt{\chi(1-\chi)}}{f_{n,j}(q_{n,j})} \\
& \quad \times \frac{F(\zeta_{\tau(i,j)}, q_{n,j}) - \mathbb{1}\{\mathcal{Y}(\zeta_{\tau(i,j)}, \varepsilon_{n,u(i,j)}) \leq q_{n,j}\}}{\sqrt{F(\zeta_{\tau(i,j)}, q_{n,j})(1-F(\zeta_{\tau(i,j)}, q_{n,j}))}} \mathbb{1}\{1 \leq i \leq k_{n,j}\}, \\
\tilde{v}_{i,j}^2 & \equiv \frac{\bar{k}_n}{k_{n,j}} \frac{\chi(1-\chi)}{f_{n,j}(q_{n,j})^2} \mathbb{1}\{1 \leq i \leq k_{n,j}\}.
\end{aligned}$$

By construction the variables $\tilde{\mathcal{U}}_{i,j}$ are $\mathcal{F}^{(0)}$ -conditionally independent across different values of $1 \leq i \leq \bar{k}_n$ and $1 \leq j \leq m_n$ with mean zero and conditional variance $\tilde{v}_{i,j}^2$. Note that

$$\sqrt{k_{n,j}}(\tilde{q}_{n,j} - q_{n,j}) = \frac{1}{\sqrt{\bar{k}_n}} \sum_{i=1}^{\bar{k}_n} \tilde{\mathcal{U}}_{i,j}, \quad \text{for } 1 \leq j \leq m_n.$$

Therefore, for each $1 \leq i \leq \bar{k}_n$ and $1 \leq j \leq 2m_n$, define $\tilde{\mathcal{U}}_{i,j}^\dagger$ as

$$\tilde{\mathcal{U}}_{i,j}^\dagger \equiv \tilde{\mathcal{U}}_{i,j} \mathbb{1}\{1 \leq j \leq m_n\} - \tilde{\mathcal{U}}_{i,j-m_n} \mathbb{1}\{m_n + 1 \leq j \leq 2m_n\}.$$

We can thus rewrite

$$\max_{1 \leq j \leq m_n} \sqrt{k_{n,j}} |\hat{q}_{n,j} - q_{n,j}| = \max_{1 \leq j \leq 2m_n} \frac{1}{\sqrt{\bar{k}_n}} \sum_{i=1}^{\bar{k}_n} \tilde{\mathcal{U}}_{i,j}^\dagger.$$

Let $(\tilde{\mathcal{Z}}_{i,j})_{1 \leq i \leq \bar{k}_n, 1 \leq j \leq m_n}$ be a sequence of centered mixed Gaussian variables with $\mathcal{F}^{(0)}$ -conditional variance $\mathbb{E}[\tilde{\mathcal{Z}}_{i,j}^2 | \mathcal{F}^{(0)}] = \mathbb{E}[\tilde{\mathcal{U}}_{i,j}^2 | \mathcal{F}^{(0)}] = \tilde{v}_{i,j}^2$. Further, for each $1 \leq i \leq \bar{k}_n$ and $1 \leq j \leq 2m_n$, let

$$\tilde{\mathcal{Z}}_{i,j}^\dagger \equiv \tilde{\mathcal{Z}}_{i,j} \mathbb{1}\{1 \leq j \leq m_n\} - \tilde{\mathcal{Z}}_{i,j-m_n} \mathbb{1}\{m_n + 1 \leq j \leq 2m_n\},$$

which implies $\mathbb{E}[\tilde{\mathcal{Z}}_{i,j} \tilde{\mathcal{Z}}_{i',j'} | \mathcal{F}^{(0)}] = \mathbb{E}[\tilde{\mathcal{Z}}_{i,j} \tilde{\mathcal{Z}}_{i',j'} | \mathcal{F}^{(0)}]$ for all $1 \leq i, i' \leq \bar{k}_n$ and $1 \leq j \leq 2m_n$. Recall that the variables $\tilde{\mathcal{U}}_{i,j}$ are bounded, by Proposition 2.1 in Chernozhukov et al. (2017), we obtain for all $\epsilon < \rho/6$ that

$$\begin{aligned} & \sup_{x \in \mathbb{R}} \mathbb{P} \left(\max_{1 \leq j \leq m_n} \sqrt{k_{n,j}} |\tilde{q}_{n,j} - q_{n,j}| \leq x \mid \mathcal{F}^{(0)} \right) \\ & - \mathbb{P} \left(\max_{1 \leq j \leq 2m_n} \frac{1}{\sqrt{\bar{k}_n}} \sum_{i=1}^{\bar{k}_n} \tilde{\mathcal{Z}}_{i,j}^\dagger \leq x \mid \mathcal{F}^{(0)} \right) \leq K \Delta_n^\epsilon. \end{aligned} \quad (\text{C.43})$$

For $1 \leq j \leq m_n$, define $Z_j \equiv \bar{k}_n^{-1/2} \sum_{i=1}^{\bar{k}_n} \tilde{\mathcal{Z}}_{i,j}$. Recalling the definition of $\tilde{\mathcal{Z}}_{i,j}$ and $\tilde{v}_{i,j}$, we have $\mathbb{E}[Z_j^2 | \mathcal{F}^{(0)}] = \chi(1 - \chi) / f_{n,j}(q_{n,j})^2 \equiv v_j^2$ for $1 \leq j \leq m_n$, hence

$$(Z_1, \dots, Z_{m_n})^\top \sim \mathcal{MN}(0, \text{diag}\{v_1^2, \dots, v_{m_n}^2\}).$$

Also note that by construction we have

$$\max_{1 \leq j \leq 2m_n} \frac{1}{\sqrt{\bar{k}_n}} \sum_{i=1}^{\bar{k}_n} \tilde{\mathcal{Z}}_{i,j}^\dagger = \max_{1 \leq j \leq m_n} \left| \frac{1}{\sqrt{\bar{k}_n}} \sum_{i=1}^{\bar{k}_n} \tilde{\mathcal{Z}}_{i,j} \right| = \max_{1 \leq j \leq m_n} |Z_j|. \quad (\text{C.44})$$

Therefore, it follows (C.41) and the triangle inequality that

$$\sup_{x \in \mathbb{R}} \left(\mathbb{P} \left(\max_{1 \leq j \leq m_n} \sup_{t \in \mathcal{T}_{n,j}} \sqrt{k_{n,j}} |\hat{q}_{n,t} - q_t| \leq x \right) - \mathbb{P} \left(\max_{1 \leq j \leq m_n} |Z_j| \leq x \right) \right)$$

$$\begin{aligned}
&\leq \mathbb{P}\left(\max_{1 \leq j \leq m_n} \sup_{t \in \mathcal{T}_{n,j}} \sqrt{k_{n,j}} |q_{n,j} - q_t| > K\Delta_n^\gamma\right) \\
&\quad + \mathbb{P}\left(\max_{1 \leq j \leq m_n} \sqrt{k_{n,j}} |\hat{q}_{n,j} - \tilde{q}_{n,j}| > K\Delta_n^\gamma\right) \\
&\quad + \sup_{x \in \mathbb{R}} \left(\mathbb{P}\left(\max_{1 \leq j \leq m_n} \sqrt{k_{n,j}} |\tilde{q}_{n,j} - q_{n,j}| \leq x + 2K\Delta_n^\gamma\right) \right. \\
&\quad \quad \left. - \mathbb{P}\left(\max_{1 \leq j \leq m_n} |Z_j| \leq x + 2K\Delta_n^\gamma\right) \right) \\
&\quad + \sup_{x \in \mathbb{R}} \mathbb{P}\left(x < \max_{1 \leq j \leq m_n} |Z_j| \leq x + 2K\Delta_n^\gamma\right) \\
&\leq K\Delta_n^\epsilon,
\end{aligned}$$

where the first term is bounded by $K\Delta_n^\epsilon$ using (C.42) and the Markov inequality, the second term uses Lemma C.1, the third term is bounded by $K\Delta_n^\epsilon$ using (C.43), (C.44) and the law of iterated expectation, the last term is bounded by $K\Delta_n^\epsilon$ using the anti-concentration inequality (see Corollary 2.1 in Chernozhukov et al. (2015)), together with the fact that

$$\mathbb{E}\left[\max_{1 \leq j \leq m_n} |Z_j|\right] \leq K\sqrt{L_n}.$$

Similarly, we can show

$$\sup_{x \in \mathbb{R}} \left(\mathbb{P}\left(\max_{1 \leq j \leq m_n} |Z_j| \leq x\right) - \mathbb{P}\left(\max_{1 \leq j \leq m_n} \sup_{t \in \mathcal{T}_{n,j}} \sqrt{k_{n,j}} |\hat{q}_{n,t} - q_t| \leq x\right) \right) \leq K\Delta_n^\epsilon.$$

This completes the proof of required statement.

Q.E.D.

C.1.3 Proof of Theorem 4.3

As mentioned in the main text, we prove a stronger result that the statement in Theorem 4.3 holds for all $\mathcal{S}_n \subset \mathcal{S}_n^{\text{all}}$ with $|\mathcal{S}_n| \geq 3$. Let $\mathcal{G}_n \equiv \mathcal{F}^{(0)} \vee \sigma(Y_{i\Delta_n} : 1 \leq i \leq n)$ denote the smallest σ -algebra contains $\mathcal{F}^{(0)} \cup \sigma(Y_{i\Delta_n} : 1 \leq i \leq n)$. Also, we strengthen Assumption 3 to Assumption 5 by a using of Localization procedure. We prove assertions of the theorem for positive ϵ satisfying

$$\epsilon < \frac{\rho}{7} \wedge \left(\frac{1}{6} - \frac{\rho}{3}\right) \wedge \left(\frac{r}{3} - \frac{\rho}{6}\right).$$

To facilitate our analysis, we adopt the notations from the proof of Theorem 4.1, and introduce some additional notations. For $1 \leq i \leq k_n$ and $(j, j') \in \mathcal{S}_n$, denote

$$V_{n,i}(j, j') \equiv Y_{\tau(i,j)} - Y_{\tau(i,j')},$$

$$\begin{aligned}
\tilde{V}_{n,i}(j, j') &\equiv \mathcal{Y}(\zeta_{\tau(i,j)}, \varepsilon_{n,t(i,j)}) - \mathcal{Y}(\zeta_{\tau(i,j')}, \varepsilon_{n,t(i,j')}), \\
\mu_{n,i}(j, j') &\equiv g_{\tau(i,j)} - g_{\tau(i,j')}, \\
\bar{\mu}_n(j, j') &\equiv g_{n,j} - g_{n,j'}, \\
\varsigma_n(j, j')^2 &\equiv \sigma_{n,j}^2 + \sigma_{n,j'}^2.
\end{aligned}$$

Using above notations, we further define

$$\begin{aligned}
\bar{D}_n &\equiv \max_{(j,j') \in \mathcal{S}_n} \frac{1}{\sqrt{k_n}} \sum_{i=1}^{k_n} \frac{V_{n,i}(j, j') - \mu_{n,i}(j, j')}{\hat{\varsigma}_n(j, j')}, \\
\tilde{D}_n &\equiv \max_{(j,j') \in \mathcal{S}_n} \frac{1}{\sqrt{k_n}} \sum_{i=1}^{k_n} \frac{\tilde{V}_{n,i}(j, j') - \bar{\mu}_n(j, j')}{\varsigma_n(j, j')}, \\
\hat{D}_n^B &\equiv \max_{(j,j') \in \mathcal{S}_n} \frac{1}{\sqrt{k_n}} \sum_{i=1}^{k_n} \frac{e_i(V_{n,i}(j, j') - (\hat{g}_{n,j} - \hat{g}_{n,j'}))}{\hat{\varsigma}_n(j, j')} \\
&= \max_{(j,j') \in \mathcal{S}_n} \frac{\sqrt{k_n}(\hat{g}_{n,j}^B - \hat{g}_{n,j'}^B)}{\hat{\varsigma}_n(j, j')}, \\
\tilde{D}_n^B &\equiv \max_{(j,j') \in \mathcal{S}_n} \frac{1}{\sqrt{k_n}} \sum_{i=1}^{k_n} \frac{e_i(\tilde{V}_{n,i}(j, j') - (\hat{g}_{n,j} - \hat{g}_{n,j'}))}{\varsigma_n(j, j')}.
\end{aligned}$$

First, we compute the approximation bounds of these variables and their conditional quantiles. Our analysis relies on the following decomposition of $|\bar{D}_n - \tilde{D}_n|$,

$$\begin{aligned}
&|\bar{D}_n - \tilde{D}_n| \\
&\leq \max_{(j,j') \in \mathcal{S}_n} \left| \frac{\varsigma_n(j, j')}{\hat{\varsigma}_n(j, j')} - 1 \right| \times \left(\max_{(j,j') \in \mathcal{S}_n} \left| \frac{1}{\sqrt{k_n}} \sum_{i=1}^{k_n} \frac{\tilde{V}_{n,i}(j, j') - \mu_{n,i}(j, j')}{\varsigma_n(j, j')} \right| \right) \\
&\quad + \max_{(j,j') \in \mathcal{S}_n} |\mathfrak{E}_n(j, j')|,
\end{aligned}$$

where for $(j, j') \in \mathcal{S}_n$, $\mathfrak{E}_n(j, j') \equiv \mathfrak{E}_n^{(I)}(j, j') + \mathfrak{E}_n^{(II)}(j, j')$ with

$$\begin{aligned}
\mathfrak{E}_n^{(I)}(j, j') &\equiv \frac{1}{\sqrt{k_n}} \sum_{i=1}^{k_n} \frac{R_{n,t(i,j)} - R_{n,t(i,j')}}{\varsigma_n(j, j')}, \\
\mathfrak{E}_n^{(II)}(j, j') &\equiv \frac{1}{\sqrt{k_n}} \sum_{i=1}^{k_n} \frac{\mu_{n,i}(j, j') - \bar{\mu}_n(j, j')}{\varsigma_n(j, j')}.
\end{aligned}$$

By the triangle inequality and (C.6), for $p > (1 - \rho)/(1/2 - \rho - \epsilon - \gamma)$, we have

$$\left\| \max_{(j,j') \in \mathcal{S}_n} \frac{1}{\sqrt{k_n}} \sum_{i=1}^{k_n} (\mu_{n,i}(j, j') - \bar{\mu}_n(j, j')) \right\|_{L_p} \leq K_p m_n^{1/p} \Delta_n^{1/2-\rho} = o(\Delta_n^{\epsilon+\gamma}). \tag{C.45}$$

Then combing (C.5) and (C.45), it follows the triangle inequality again that

$$\max_{(j,j') \in \mathcal{S}_n} |\mathfrak{C}_n(j, j')| \leq \max_{(j,j') \in \mathcal{S}_n} |\mathfrak{C}_n^{(I)}(j, j')| + \max_{(j,j') \in \mathcal{S}_n} |\mathfrak{C}_n^{(II)}(j, j')| = o_p(\Delta_n^{\epsilon+\gamma}). \quad (\text{C.46})$$

Note that for positive a, b, c, d , we have $a/b \leq c/d$ implies $a/b \leq (a+c)/(b+d) \leq c/d$. Combing with (C.14), we obtain that for $\epsilon/\rho \leq \varpi < 1 - 2\gamma/\rho$,

$$\mathbb{P}\left(\max_{(j,j') \in \mathcal{S}_n} \left| \frac{\varsigma_n(j, j')}{\hat{\varsigma}_n(j, j')} - 1 \right| > K \Delta_n^{\rho(1-\varpi)/2} L_n^{2\eta} \log(|\mathcal{S}_n|)\right) \leq K' \Delta_n^\epsilon. \quad (\text{C.47})$$

Combing (C.46) and (C.47), following the similar procedure as deriving (C.2), we can show that

$$\mathbb{P}(|\bar{D}_n - \tilde{D}_n| > K \varrho_n) \leq K' \Delta_n^\epsilon, \quad (\text{C.48})$$

for some sequence $\varrho_n \asymp \Delta_n^{\rho(1-\varpi)/4} L_n^{2\eta} \log(|\mathcal{S}_n|)^{3/2}$ where $(\epsilon/\rho) \vee (1 - 2\gamma/\rho) < \varpi < 1 - 4\epsilon/\rho$. Note that $|\mathcal{S}_n| \leq m_n(m_n - 1)$ by construction. On the other hand, we have the following decomposition of $|\hat{D}_n^B - \tilde{D}_n^B|$ as

$$\begin{aligned} & |\hat{D}_n^B - \tilde{D}_n^B| \\ & \leq \max_{(j,j') \in \mathcal{S}_n} \left| \frac{\varsigma_n(j, j')}{\hat{\varsigma}_n(j, j')} - 1 \right| \\ & \quad \times \left(\max_{(j,j') \in \mathcal{S}_n} \frac{1}{\sqrt{k_n}} \sum_{i=1}^{k_n} \frac{e_i (\tilde{V}_{n,i}(j, j') - (\hat{g}_{n,j} - \hat{g}_{n,j'}))}{\varsigma_n(j, j')} \right) \\ & \quad + \max_{(j,j') \in \mathcal{S}_n} |\mathfrak{D}_n(j, j')|, \end{aligned} \quad (\text{C.49})$$

where for $(j, j') \in \mathcal{S}_n$, $\mathfrak{D}_n(j, j') \equiv k_n^{-1/2} \sum_{i=1}^{k_n} e_i (R_{n,t(i,j)} - R_{n,t(i,j')}) / \varsigma_n(j, j')$. Recall that $(e_i)_{1 \leq i \leq k_n}$ follows i.i.d. standard Gaussian distribution, hence

$$\max_{1 \leq i \leq k_n} |e_i|^2 = O_p(L_n),$$

by the maximal inequality. Applying the Cauchy–Schwartz inequality and combining with (C.46), we have

$$\max_{(j,j') \in \mathcal{S}_n} |\mathfrak{D}_n(j, j')| \leq \sqrt{\max_{1 \leq i \leq k_n} |e_i|^2 \times \max_{(j,j') \in \mathcal{S}_n} |\mathfrak{C}_n^{(I)}(j, j')|^2} = o_p(\Delta_n^{\epsilon+\gamma} \sqrt{L_n}). \quad (\text{C.50})$$

Let $E_{n,3}$ be the event such that

$$\begin{aligned} E_{n,3} & \equiv \left\{ \max_{(j,j') \in \mathcal{S}_n} \left| \frac{\varsigma_n(j, j')}{\hat{\varsigma}_n(j, j')} - 1 \right| \leq \Delta_n^{\rho(1-\varpi)/2} L_n^{2\eta} \log(|\mathcal{S}_n|) \right\} \\ & \quad \cap \left\{ \max_{(j,j') \in \mathcal{S}_n} |\mathfrak{D}_n(j, j')| \leq \Delta_n^{\gamma/2} \right\}, \end{aligned}$$

by (C.47), (C.50) and the Markov inequality, we have shown $\mathbb{P}(E_{n,3}) > 1 - K' \Delta_n^\epsilon$.

Note that conditional on \mathcal{G}_n , the normalized t -statistics $(k_n^{-1/2} \sum_{i=1}^{k_n} e_i (\tilde{V}_{n,i}(j, j') - (\hat{g}_{n,j} - \hat{g}_{n,j'})) / \varsigma_n(j, j'))_{(j, j') \in \mathcal{S}_n}$ follow a Gaussian distribution with bounded variance, which implies $\mathbb{E}[\tilde{D}_n^B | \mathcal{G}_n] \leq K \sqrt{\log(|\mathcal{S}_n|)}$. Therefore, it follows the Markov inequality and (C.49) that

$$\begin{aligned} & \mathbb{P}(\{|\hat{D}_n^B - \tilde{D}_n^B| > \varrho_n\} \cap E_{n,3} | \mathcal{G}_n) \\ & \leq \varrho_n^{-1} \left(\max_{(j, j') \in \mathcal{S}_n} \left| \frac{\varsigma_n(j, j')}{\hat{\varsigma}_n(j, j')} - 1 \right| \times \left(\mathbb{E}[\tilde{D}_n^B | \mathcal{G}_n] + 2 \max_{(j, j') \in \mathcal{S}_n} |\mathfrak{D}_n(j, j')| \right) \right) \\ & \leq \frac{\Delta_n^{\rho(1-\varpi)/2} L_n^{2\eta} \log(|\mathcal{S}_n|) (K \sqrt{\log(|\mathcal{S}_n|)} + \Delta_n^{\gamma/2})}{K' \Delta_n^{\rho(1-\varpi)/4} L_n^{2\eta} \log(|\mathcal{S}_n|)^{3/2}} \leq K \Delta_n^{\rho(1-\varpi)/4}. \end{aligned}$$

With K denoting the same constant as in the above display, by the law of iterated expectation, we can conclude that

$$\mathbb{P}(\mathbb{P}(|\hat{D}_n^B - \tilde{D}_n^B| > \varrho_n | \mathcal{G}_n) > K \Delta_n^{\rho(1-\varpi)/4}) \leq \mathbb{P}(E_{n,3}^c) \leq K' \Delta_n^\epsilon. \quad (\text{C.51})$$

Let $\tilde{X}_n(j, j')$ be centered mixed Gaussian variables indexed by (j, j') with $\mathcal{F}^{(0)}$ -conditional covariance matrix such that for all $(j, j'), (\ell, \ell') \in \mathcal{S}_n$,

$$\begin{aligned} & \mathbb{E}[\tilde{X}_n(j, j') \tilde{X}_n(\ell, \ell') | \mathcal{F}^{(0)}] \\ & = \mathbb{E} \left[\left(\frac{1}{\sqrt{k_n}} \sum_{i=1}^{k_n} \frac{\tilde{V}_{n,i}(j, j') - \mu_{n,i}(j, j')}{\varsigma_n(j, j')} \right) \right. \\ & \quad \left. \times \left(\frac{1}{\sqrt{k_n}} \sum_{i=1}^{k_n} \frac{\tilde{V}_{n,i}(\ell, \ell') - \mu_{n,i}(\ell, \ell')}{\varsigma_n(\ell, \ell')} \right) \middle| \mathcal{F}^{(0)} \right]. \end{aligned}$$

Then by Proposition 2.1 in Chernozhukov et al. (2017), we have for all $\epsilon < \rho/6$,

$$\begin{aligned} & \sup_{x \in \mathbb{R}} \left| \mathbb{P} \left(\max_{(j, j') \in \mathcal{S}_n} \frac{1}{\sqrt{k_n}} \sum_{i=1}^{k_n} \frac{\tilde{V}_{n,i}(j, j') - \mu_{n,i}(j, j')}{\varsigma_n(j, j')} \leq x \middle| \mathcal{F}^{(0)} \right) \right. \\ & \quad \left. - \mathbb{P} \left(\max_{(j, j') \in \mathcal{S}_n} \tilde{X}_n(j, j') \leq x \middle| \mathcal{F}^{(0)} \right) \right| \\ & \leq K (\Delta_n^{\rho/6} L_n^{\eta/3} (L_n + \log(|\mathcal{S}_n|)))^{7/6} + \Delta_n^{\rho/6} L_n^{2\eta/3} (L_n + \log(|\mathcal{S}_n|)) \\ & \leq K \Delta_n^\epsilon. \end{aligned} \quad (\text{C.52})$$

By Corollary 4.2 in Chernozhukov et al. (2017), for all $\epsilon < \rho/7$, with probability at least $1 - K \Delta_n^\epsilon$,

$$\begin{aligned} & \sup_{x \in \mathbb{R}} \left| \mathbb{P}(\tilde{D}_n^B \leq x | \mathcal{G}_n) - \mathbb{P} \left(\max_{(j, j') \in \mathcal{S}_n} \tilde{X}_n(j, j') \leq x \middle| \mathcal{F}^{(0)} \right) \right| \\ & \leq K' (\Delta_n^{\rho/6} L_n^{(1+\eta)/3} (L_n + \log(|\mathcal{S}_n|)))^{5/6} + \Delta_n^{(\rho-\epsilon)/6} L_n^{2\eta/3} (L_n + \log(|\mathcal{S}_n|)) \\ & \leq K' \Delta_n^\epsilon. \end{aligned} \quad (\text{C.53})$$

Let $\tilde{c}v_n(\cdot, \mathcal{S}_n)$ denote the $\mathcal{F}^{(0)}$ -conditional $1 - (\cdot)$ quantile of $\max_{(j,j') \in \mathcal{S}_n} \tilde{X}_n(j, j')$, i.e.,

$$\tilde{c}v_n(\cdot, \mathcal{S}_n) \equiv \inf \left\{ C \in \mathbb{R} : \mathbb{P} \left(\max_{(j,j') \in \mathcal{S}_n} \tilde{X}_n(j, j') \leq C \mid \mathcal{F}^{(0)} \right) \geq 1 - (\cdot) \right\}.$$

Note that $\mathbb{E}[\max_{(j,j') \in \mathcal{S}_n} \tilde{X}_n(j, j') \mid \mathcal{F}^{(0)}] \leq K \sqrt{\log(|\mathcal{S}_n|)}$. Also note that Assumption 5(i) implies the bounds obtained in the previous equation and in the approximation (C.52), (C.53) are universal. Consequently, we can fix a positive universal constant M satisfying the previous equation. Therefore, for $\alpha \in (0, 1 - M\varrho_n \sqrt{\log(|\mathcal{S}_n|)})$, by the anti-concentration inequality, we have

$$\mathbb{P} \left(\max_{(j,j') \in \mathcal{S}_n} \tilde{X}_n(j, j') \leq \tilde{c}v_n(\alpha + M\varrho_n \sqrt{\log(|\mathcal{S}_n|)}, \mathcal{S}_n) + \varrho_n \mid \mathcal{F}^{(0)} \right) \leq 1 - \alpha. \quad (\text{C.54})$$

Let $E_{n,4}$ be the event such that

$$\begin{aligned} E_{n,4} \equiv & \{ \mathbb{P}(|\widehat{D}_n^B - \widetilde{D}_n^B| > \varrho_n \mid \mathcal{G}_n) \leq M\Delta_n^{\rho(1-\varpi)/4} \} \\ & \cap \left\{ \sup_{x \in \mathbb{R}} \left| \mathbb{P}(\widetilde{D}_n^B \leq x \mid \mathcal{G}_n) - \mathbb{P} \left(\max_{(j,j') \in \mathcal{S}_n} \tilde{X}_n(j, j') \leq x \mid \mathcal{F}^{(0)} \right) \right| \leq M\Delta_n^\epsilon \right\}, \end{aligned}$$

by (C.51) and (C.53), we have shown $\mathbb{P}(E_{n,4}) \geq 1 - K'\Delta_n^\epsilon$. Therefore, we have

$$\begin{aligned} & \mathbb{P}(\{\widehat{D}_n^B \leq \tilde{c}v_n(\alpha + M(\Delta_n^\epsilon + \varrho_n \sqrt{\log(|\mathcal{S}_n|)}), \mathcal{S}_n)\} \cap E_{n,4} \mid \mathcal{G}_n) \\ & \leq \mathbb{P}(\{\widetilde{D}_n^B \leq \tilde{c}v_n(\alpha + M(\Delta_n^\epsilon + \varrho_n \sqrt{\log(|\mathcal{S}_n|)}), \mathcal{S}_n) + \varrho_n\} \cap E_{n,4} \mid \mathcal{G}_n) \\ & \quad + M\Delta_n^{\rho(1-\varpi)/4} \\ & \leq \mathbb{P} \left(\max_{(j,j') \in \mathcal{S}_n} \tilde{X}_n(j, j') \leq \tilde{c}v_n(\alpha + M(\Delta_n^\epsilon + \varrho_n \sqrt{\log(|\mathcal{S}_n|)}), \mathcal{S}_n) + \varrho_n \mid \mathcal{F}^{(0)} \right) \\ & \quad + M\Delta_n^\epsilon \\ & \leq 1 - \alpha - M\Delta_n^\epsilon + M\Delta_n^\epsilon = 1 - \alpha, \end{aligned}$$

where the third line uses the fact that $\rho(1 - \varpi)/4 > \epsilon$, and the fourth line is by (C.54). By the law of iterated expectation and the definition of $cv_n^B(\alpha, \mathcal{S}_n)$, we can conclude that

$$\mathbb{P}(cv_n^B(\alpha, \mathcal{S}_n) < \tilde{c}v_n(\alpha + M(\Delta_n^\epsilon + \varrho_n \sqrt{\log(|\mathcal{S}_n|)}), \mathcal{S}_n)) \leq \mathbb{P}(E_{n,4}^c) \leq K'\Delta_n^\epsilon. \quad (\text{C.55})$$

By the anti-concentration inequality, for $\alpha \in (M\varrho_n \sqrt{\log(|\mathcal{S}_n|)}, 1)$, we have

$$\mathbb{P} \left(\max_{(j,j') \in \mathcal{S}_n} \tilde{X}_n(j, j') \leq \tilde{c}v_n(\alpha - M\varrho_n \sqrt{\log(|\mathcal{S}_n|)}, \mathcal{S}_n) - \varrho_n \mid \mathcal{F}^{(0)} \right) \geq 1 - \alpha.$$

Similarly, we can show

$$\mathbb{P}(cv_n^B(\alpha, \mathcal{S}_n) > \tilde{c}v_n(\alpha - M(\Delta_n^\epsilon + \varrho_n \sqrt{\log(|\mathcal{S}_n|)}), \mathcal{S}_n)) \leq \mathbb{P}(E_{n,4}^c) \leq K' \Delta_n^\epsilon. \quad (\text{C.56})$$

We are now ready to prove the asserted statements in the theorem, starting from assertion (i). Assume that $\max_{(j,j') \in \mathcal{S}_n} (g_{n,j} - g_{n,j'}) \leq 0$, this implies $\bar{\mu}_n(j, j') \leq 0$ for all $(j, j') \in \mathcal{S}_n$. Combing with (C.45) yields

$$\mathbb{P}\left(\max_{(j,j') \in \mathcal{S}_n} \frac{1}{\sqrt{k_n}} \sum_{i=1}^{k_n} \mu_{n,i}(j, j') > K\varrho_n\right) \leq K' \Delta_n^\epsilon.$$

Therefore, by (C.47) and the Markov inequality, this gives $\mathbb{P}(\widehat{D}_n - \bar{D}_n > \varrho_n/2) \leq K\Delta_n^\epsilon$. Hence

$$\begin{aligned} & \mathbb{P}(\widehat{D}_n > cv_n^B(\alpha, \mathcal{S}_n)) \\ & \leq \mathbb{P}(\bar{D}_n > cv_n^B(\alpha, \mathcal{S}_n) - \varrho_n/2) + \mathbb{P}(\widehat{D}_n - \bar{D}_n > \varrho_n/2) \\ & \leq \mathbb{P}(\widetilde{D}_n > cv_n^B(\alpha, \mathcal{S}_n) - \varrho_n/2 - \varrho_n/2) + K\Delta_n^\epsilon \\ & \leq \mathbb{P}(\widetilde{D}_n > \tilde{c}v_n(\alpha + M(\Delta_n^\epsilon + \varrho_n \sqrt{\log(|\mathcal{S}_n|)}), \mathcal{S}_n) - \varrho_n) + K\Delta_n^\epsilon, \end{aligned} \quad (\text{C.57})$$

where the second line is by (C.48), and the last line is by (C.55). For the first term, we have

$$\begin{aligned} & \mathbb{P}(\widetilde{D}_n > \tilde{c}v_n(\alpha + M(\Delta_n^\epsilon + \varrho_n \sqrt{\log(|\mathcal{S}_n|)}), \mathcal{S}_n) - \varrho_n \mid \mathcal{F}^{(0)}) \\ & \leq \mathbb{P}\left(\max_{(j,j') \in \mathcal{S}_n} \widetilde{X}_n(j, j') > \tilde{c}v_n(\alpha + M(\Delta_n^\epsilon + \varrho_n \sqrt{\log(|\mathcal{S}_n|)}), \mathcal{S}_n) - \varrho_n \mid \mathcal{F}^{(0)}\right) \\ & \quad + K\Delta_n^\epsilon \\ & \leq \mathbb{P}\left(\max_{(j,j') \in \mathcal{S}_n} \widetilde{X}_n(j, j') > \tilde{c}v_n(\alpha + 2M(\Delta_n^\epsilon + \varrho_n \sqrt{\log(|\mathcal{S}_n|)}), \mathcal{S}_n) \mid \mathcal{F}^{(0)}\right) \\ & \quad + K\Delta_n^\epsilon \\ & \leq \alpha + 2M(\Delta_n^\epsilon + \varrho_n \sqrt{\log(|\mathcal{S}_n|)}) + K\Delta_n^\epsilon \\ & \leq \alpha + K\Delta_n^\epsilon, \end{aligned} \quad (\text{C.58})$$

where the second line is by (C.45), (C.52), and the law of iterated expectation, the third line is by (C.54), the last line is by the definition of $\tilde{c}v_n(\cdot, \mathcal{S}_n)$ and the fact that $\mathcal{S}_n \subset \{1, \dots, m_n\} \times \{1, \dots, m_n\}$ hence

$$\varrho_n \sqrt{\log(|\mathcal{S}_n|)} \leq K\varrho_n \sqrt{L_n} \leq K' \Delta_n^\epsilon.$$

Combing (C.57), (C.58), and applying the law of iterated expectation again, we can conclude that

$$\mathbb{P}(\widehat{D}_n > cv_n^B(\alpha, \mathcal{S}_n)) \leq \alpha + K\Delta_n^\epsilon, \quad \text{if } \max_{(j,j') \in \mathcal{S}_n} (g_{n,j} - g_{n,j'}) \leq 0, \quad (\text{C.59})$$

which is the first part of assertion (i). For the second part, assume $\bar{\mu}_n(j, j') = 0$, then (C.45) yields

$$\mathbb{P}\left(\max_{(j,j') \in \mathcal{S}_n} \left| \frac{1}{\sqrt{k_n}} \sum_{i=1}^{k_n} \mu_{n,i}(j, j') \right| > K\varrho_n\right) \leq K'\Delta_n^\epsilon.$$

Therefore, by (C.47) and the Markov inequality, this gives $\mathbb{P}(\bar{D}_n - \widehat{D}_n > \varrho_n/2) \leq K\Delta_n^\epsilon$. Hence

$$\begin{aligned} & \mathbb{P}(\widehat{D}_n > cv_n^B(\alpha, \mathcal{S}_n)) \\ & \geq \mathbb{P}(\bar{D}_n > cv_n^B(\alpha, \mathcal{S}_n) + \varrho_n/2) - \mathbb{P}(\bar{D}_n - \widehat{D}_n > \varrho_n/2) \\ & \geq \mathbb{P}(\widetilde{D}_n > \widetilde{c}v_n(\alpha - M(\Delta_n^\epsilon + \varrho_n \sqrt{\log(|\mathcal{S}_n|)}), \mathcal{S}_n) + \varrho_n) - K\Delta_n^\epsilon, \end{aligned} \quad (\text{C.60})$$

where the second line is by (C.48) and (C.56). For the first term, we have

$$\begin{aligned} & \mathbb{P}(\widetilde{D}_n > \widetilde{c}v_n(\alpha - M(\Delta_n^\epsilon + \varrho_n \sqrt{\log(|\mathcal{S}_n|)}), \mathcal{S}_n) + \varrho_n \mid \mathcal{F}^{(0)}) \\ & \geq \mathbb{P}\left(\max_{(j,j') \in \mathcal{S}_n} \widetilde{X}_n(j, j') > \widetilde{c}v_n(\alpha - 2M(\Delta_n^\epsilon + \varrho_n \sqrt{\log(|\mathcal{S}_n|)}), \mathcal{S}_n) \mid \mathcal{F}^{(0)}\right) \\ & \quad - K\Delta_n^\epsilon \\ & \geq \alpha - 2M(\Delta_n^\epsilon + \varrho_n \sqrt{\log(|\mathcal{S}_n|)}) - K\Delta_n^\epsilon \\ & \geq \alpha - K\Delta_n^\epsilon, \end{aligned} \quad (\text{C.61})$$

where the second line is by (C.45), (C.52), and (C.54). Combing (C.59)-(C.61), and the law of iterated expectation completes the proof of assertion (i).

For assertion (ii), assume that $\max_{(j,j') \in \mathcal{S}_n} \bar{\mu}_n(j, j') \geq \Upsilon$ for some positive Υ . Combining with (C.45) and (C.47) gives $\mathbb{P}(\bar{D}_n - \widehat{D}_n + \Delta_n^{-\rho/2}\Upsilon > \varrho_n/2) \leq K\Delta_n^\epsilon$. Therefore, we have

$$\begin{aligned} & \mathbb{P}(\widehat{D}_n > cv_n^B(\alpha, \mathcal{S}_n)) \\ & \geq \mathbb{P}(\bar{D}_n + \Delta_n^{-\rho/2}\Upsilon > cv_n^B(\alpha, \mathcal{S}_n) + \varrho_n/2) - \mathbb{P}(\bar{D}_n - \widehat{D}_n > \varrho_n/2) \\ & \geq \mathbb{P}\left(\max_{(j,j') \in \mathcal{S}_n} \widetilde{X}_n(j, j') + \Delta_n^{-\rho/2}\Upsilon > \widetilde{c}v_n(\alpha - 2M(\Delta_n^\epsilon + \varrho_n \sqrt{\log(|\mathcal{S}_n|)}), \mathcal{S}_n)\right) \\ & \quad - K\Delta_n^\epsilon \end{aligned}$$

$$\begin{aligned}
&\geq \mathbb{P}\left(\max_{(j,j') \in \mathcal{S}_n} \tilde{X}_n(j, j') + \Delta_n^{-\rho/2} \Upsilon > K(\sqrt{\log(|\mathcal{S}_n|)} + \sqrt{L_n})\right) - K\Delta_n^\epsilon \\
&\geq 1 - K\Delta_n^{\rho/2} - K'\Delta_n^\epsilon \geq 1 - K\Delta_n^\epsilon,
\end{aligned}$$

where the second line is by (C.45), (C.48), (C.52), (C.54), and (C.56). The third line is by Borell's concentration inequality (see, e.g., Proposition A.2.1 in van der Vaart and Wellner (1996)), which gives $\mathbb{P}(|\max_{(j,j') \in \mathcal{S}_n} \tilde{X}_n(j, j') - M\sqrt{\log(|\mathcal{S}_n|)}| \geq \lambda) \leq K \exp\{-\lambda^2/2K'\}$, setting the right hand side equaling to

$$\alpha - 2M(\Delta_n^\epsilon + \varrho_n \sqrt{\log(|\mathcal{S}_n|)}),$$

yields

$$\begin{aligned}
&\tilde{c}v_n(\alpha - 2M(\Delta_n^\epsilon + \varrho_n \sqrt{\log(|\mathcal{S}_n|)}), \mathcal{S}_n) \\
&\leq M\sqrt{\log(|\mathcal{S}_n|)} + K\sqrt{2\log\left(\frac{1}{\alpha - 2M(\Delta_n^\epsilon + \varrho_n \sqrt{\log(|\mathcal{S}_n|)})}\right)} \\
&\leq K(\sqrt{\log(|\mathcal{S}_n|)} + \sqrt{L_n}).
\end{aligned}$$

This completes the proof of required statement.

Q.E.D.

C.1.4 Proof of Corollary 4.1

The corollary is a direct consequence of Theorem 3.3 in Mogstad et al. (2023) and Theorem 4.3.

Q.E.D.

C.2 Extension to Dependent Disturbance

The strong approximation results derived in this paper can be extended to the case without assuming disturbances to be conditionally independent. In particular, we outline the main steps in constructing a similar approximation result of $\max_{1 \leq j \leq m_n} \sup_{t \in \mathcal{T}_{n,j}} \sqrt{k_{n,j}} |\hat{g}_{n,t} - g_t|$ for stationary β -mixing disturbance. For any sub σ -fields \mathcal{A}, \mathcal{B} of \mathcal{F} , denote

$$\beta(\mathcal{A}, \mathcal{B}) \equiv \frac{1}{2} \sup \left\{ \sum_{i=1}^I \sum_{j=1}^J |\mathbb{P}(A_i \cap B_j) - \mathbb{P}(A_i)\mathbb{P}(B_j)| \right\},$$

where the supremum is taken over all pairs of finite partitions $\{A_1, \dots, A_I\}$ and $\{B_1, \dots, B_J\}$ of Ω such that $A_i \in \mathcal{A}$ for each i and $B_j \in \mathcal{B}$ for each j . Define the k th β -mixing coefficient of $(\varepsilon_{n,i})_{1 \leq i \leq n}$ as $\beta(k) \equiv \max_{1 \leq \ell \leq n-k} \beta(\mathcal{H}_1^\ell, \mathcal{H}_{\ell+k}^n)$ where $\mathcal{H}_{i,j} \equiv \sigma(\varepsilon_{n,i}, \dots, \varepsilon_{n,j})$ for $1 \leq i \leq j \leq n$. Moreover, for each $1 \leq q \leq n$, and $\mathcal{E} \subset \mathcal{Z}$ define

$$\begin{aligned}\bar{\sigma}^2(\mathcal{E}, q) &\equiv \sup_{z \in \mathcal{E}} \frac{1}{q} \sum_{i=1}^q \sum_{j=1}^q \text{Cov}(\mathcal{Y}(z, \varepsilon_{n,i}), \mathcal{Y}(z, \varepsilon_{n,j})), \\ \underline{\sigma}^2(\mathcal{E}, q) &\equiv \inf_{z \in \mathcal{E}} \frac{1}{q} \sum_{i=1}^q \sum_{j=1}^q \text{Cov}(\mathcal{Y}(z, \varepsilon_{n,i}), \mathcal{Y}(z, \varepsilon_{n,j})).\end{aligned}$$

We follow the notations used in the proof of Theorem 4.1. Note that the derivation of (C.5) and (C.6) does not depend on conditional independence of $(\varepsilon_{n,i})_{1 \leq i \leq n}$, hence we have

$$\mathbb{P}\left(\left|\max_{1 \leq j \leq m_n} \sup_{t \in \mathcal{T}_{n,j}} \sqrt{k_{n,j}} |\hat{g}_{n,t} - g_t| - \max_{1 \leq j \leq m_n} \left| \frac{1}{\sqrt{k_{n,j}}} \sum_{i=1}^{k_{n,j}} \tilde{Y}_{i,j} \right| \right| > K \Delta_n^\gamma\right) \leq K' \Delta_n^\epsilon.$$

For $1 \leq i \leq \bar{k}_n$ and $1 \leq j \leq 2m_n$, denote

$$\begin{aligned}\tilde{Y}_{i,j}^\dagger &\equiv \sqrt{\frac{\bar{k}_n}{k_{n,j}}} (\tilde{Y}_{i,j} \mathbb{1}\{1 \leq i \leq k_{n,j}, 1 \leq j \leq m_n\} \\ &\quad - \tilde{Y}_{i,j-m_n} \mathbb{1}\{1 \leq i \leq k_{n,j}, m_n + 1 \leq j \leq 2m_n\}).\end{aligned}$$

Then we can rewrite

$$\max_{1 \leq j \leq m_n} \left| \frac{1}{\sqrt{k_{n,j}}} \sum_{i=1}^{k_{n,j}} \tilde{Y}_{i,j} \right| = \max_{1 \leq j \leq 2m_n} \frac{1}{\sqrt{\bar{k}_n}} \sum_{i=1}^{\bar{k}_n} \tilde{Y}_{i,j}^\dagger.$$

The key step is to reduce the summation on the right hand side of above display into an independent sum. To establish this, we need the following assumption which specifies the rate of convergence of β -mixing coefficient and boundedness of long-run variance.

Assumption 7. *There exists a positive constant K such that (i) $\beta(n) \leq Kn^{-\nu}$ for some positive ν ; (ii) $1/K \leq \underline{\sigma}^2(\mathcal{K}, q) \leq \bar{\sigma}^2(\mathcal{K}, q) \leq K$ for all $1 \leq q \leq n$.*

The construction is based on the method of ‘‘Bernstein sums,’’ which is widely used for analyzing dependent processes, see, e.g., Bernstein (1927) and Davidson (1992). Namely, let $q_{1,n} \asymp \Delta_n^{-2\kappa}$ and $q_{2,n} \asymp \Delta_n^{-\kappa}$ where $\rho/(2+\nu) < \kappa < \rho/2$ and $q_{1,n} + q_{2,n} < \bar{k}_n/2$. Denote $\bar{\ell}_n \equiv \lfloor \bar{k}_n/(q_{1,n} + q_{2,n}) \rfloor \asymp \Delta_n^{2\kappa-\rho}$. For $1 \leq j \leq 2m_n$

and $1 \leq \ell \leq \bar{\ell}_n$, define

$$\tilde{S}_{\ell,j} \equiv \sum_{i=(\ell-1)(q_{1,n}+q_{2,n})+1}^{(\ell-1)(q_{1,n}+q_{2,n})+q_{1,n}} \tilde{Y}_{i,j}^\dagger, \quad \text{and} \quad \mathring{S}_{\ell,j} \equiv \sum_{i=(\ell-1)(q_{1,n}+q_{2,n})+q_{1,n}+1}^{\ell(q_{1,n}+q_{2,n})} \tilde{Y}_{i,j}^\dagger.$$

Then we have the following decomposition

$$\frac{1}{\sqrt{\bar{k}_n}} \sum_{i=1}^{\bar{k}_n} \tilde{Y}_{i,j}^\dagger = \frac{1}{\sqrt{\bar{k}_n}} \sum_{\ell=1}^{\bar{\ell}_n} \tilde{S}_{\ell,j} + \frac{1}{\sqrt{\bar{k}_n}} \sum_{\ell=1}^{\bar{\ell}_n} \mathring{S}_{\ell,j} + \frac{1}{\sqrt{\bar{k}_n}} \sum_{i=\ell(q_{1,n}+q_{2,n})}^{\bar{k}_n} \tilde{Y}_{i,j}^\dagger.$$

Therefore, by the triangle inequality,

$$\begin{aligned} & \left| \max_{1 \leq j \leq 2m_n} \frac{1}{\sqrt{\bar{k}_n}} \sum_{i=1}^{\bar{k}_n} \tilde{Y}_{i,j} - \max_{1 \leq j \leq 2m_n} \frac{1}{\sqrt{\bar{k}_n}} \sum_{\ell=1}^{\bar{\ell}_n} \tilde{S}_{\ell,j} \right| \\ & \leq \max_{1 \leq j \leq 2m_n} \left| \frac{1}{\sqrt{\bar{k}_n}} \sum_{\ell=1}^{\bar{\ell}_n} \mathring{S}_{\ell,j} \right| + \max_{1 \leq j \leq 2m_n} \left| \frac{1}{\sqrt{\bar{k}_n}} \sum_{i=\ell(q_{1,n}+q_{2,n})+1}^{\bar{k}_n} \tilde{Y}_{i,j}^\dagger \right|. \end{aligned} \quad (\text{C.62})$$

Moreover, let $(\tilde{S}'_{\ell,j})_{1 \leq \ell \leq \bar{\ell}_n}$ and $(\mathring{S}'_{\ell,j})_{1 \leq \ell \leq \bar{\ell}_n}$ be two $\mathcal{F}^{(0)}$ -conditionally *independent* sequences such that $\tilde{S}'_{\ell,j} \stackrel{\mathcal{L}}{=} \tilde{S}_{\ell,j}$ and $\mathring{S}'_{\ell,j} \stackrel{\mathcal{L}}{=} \mathring{S}_{\ell,j}$ for $1 \leq \ell \leq \bar{\ell}_n$. Since the projection mapping is continuous, hence the Borel σ -algebra of \mathbb{R}^{2m_n} is equivalent to the σ -algebra generated by the Cartesian product of Borel sets of \mathbb{R} . Therefore, by Assumption 7(i), it follows Corollary 2.7 of Yu (1994) that

$$\begin{aligned} & \sup_{x \in \mathbb{R}} \left| \mathbb{P} \left(\max_{1 \leq j \leq 2m_n} \sum_{\ell=1}^{\bar{\ell}_n} \tilde{S}_{\ell,j} \leq x \mid \mathcal{F}^{(0)} \right) - \mathbb{P} \left(\max_{1 \leq j \leq 2m_n} \sum_{\ell=1}^{\bar{\ell}_n} \tilde{S}'_{\ell,j} \leq x \mid \mathcal{F}^{(0)} \right) \right| \\ & \leq K \bar{\ell}_n q_{2,n}^{-\nu} \leq K \Delta_n^{(2+\nu)\kappa-\rho}, \end{aligned} \quad (\text{C.63})$$

$$\begin{aligned} & \sup_{x \in \mathbb{R}} \left| \mathbb{P} \left(\max_{1 \leq j \leq 2m_n} \sum_{\ell=1}^{\bar{\ell}_n} \mathring{S}_{\ell,j} \leq x \mid \mathcal{F}^{(0)} \right) - \mathbb{P} \left(\max_{1 \leq j \leq 2m_n} \sum_{\ell=1}^{\bar{\ell}_n} \mathring{S}'_{\ell,j} \leq x \mid \mathcal{F}^{(0)} \right) \right| \\ & \leq K \bar{\ell}_n q_{1,n}^{-\nu} \leq K \Delta_n^{(2+2\nu)\kappa-\rho}. \end{aligned} \quad (\text{C.64})$$

Taking positive constants ϵ and γ such that $\epsilon + \gamma < ((2 + \nu)\kappa - \rho) \wedge (\kappa/2) \wedge (\rho/2 - \kappa)$. Combing (C.62)-(C.64) and by the law of iterated expectation, we have

$$\begin{aligned} & \mathbb{P} \left(\max_{1 \leq j \leq 2m_n} \frac{1}{\sqrt{\bar{k}_n}} \sum_{i=1}^{\bar{k}_n} \tilde{Y}_{i,j}^\dagger < x \right) \\ & \leq \mathbb{P} \left(\max_{1 \leq j \leq 2m_n} \frac{1}{\sqrt{\bar{k}_n}} \sum_{\ell=1}^{\bar{\ell}_n} \tilde{S}'_{\ell,j} \leq x + K \Delta_n^\gamma \right) + \mathbb{P} \left(\max_{1 \leq j \leq 2m_n} \mathfrak{E}_{n,j}^{(I)} \leq \frac{K}{2} \Delta_n^\gamma \right) \\ & \quad + \mathbb{P} \left(\max_{1 \leq j \leq 2m_n} \mathfrak{E}_{n,j}^{(II)} > \frac{K}{2} \Delta_n^\gamma \right) + K' \Delta_n^\epsilon, \end{aligned} \quad (\text{C.65})$$

where for $1 \leq j \leq 2m_n$,

$$\mathfrak{E}_{n,j}^{(I)} \equiv \left| \frac{1}{\sqrt{\bar{k}_n}} \sum_{\ell=1}^{\bar{\ell}_n} \tilde{S}'_{\ell,j} \right|, \quad \text{and} \quad \mathfrak{E}_{n,j}^{(II)} \equiv \left| \frac{1}{\sqrt{\bar{k}_n}} \sum_{i=\ell(q_{1,n}+q_{2,n})+1}^{\bar{k}_n} \tilde{Y}_{i,j}^\dagger \right|.$$

For the second term, by Assumption 5(iii) and Assumption 7(ii), it follows Lemma 8 in Chernozhukov et al. (2015) that

$$\begin{aligned} \mathbb{E} \left[\max_{1 \leq j \leq 2m_n} \mathfrak{E}_{n,j}^{(I)} \mid \mathcal{F}^{(0)} \right] &\leq K(q_{1,n}^{-1/2} q_{2,n}^{1/2} \sqrt{L_n} + \bar{k}_n^{-1/2} q_{2,n} L_n^{3/2}) \\ &\leq K \Delta_n^{(\kappa/2) \wedge (\rho/2 - \kappa)} L_n^{3/2}. \end{aligned}$$

Then by the Markov inequality and the law of iterated expectation, we obtain

$$\mathbb{P} \left(\max_{1 \leq j \leq 2m_n} \mathfrak{E}_{n,j}^{(I)} \leq K \Delta_n^\gamma \right) \leq K' \Delta_n^\epsilon. \quad (\text{C.66})$$

For the third term in the right hand side of (C.65), note that Assumption 7(i) implies α -mixing and hence, combining with Assumption 5(iii) and 7(ii) yields condition (1.3) in Rio (1995). Therefore, it follow the law of iterated logarithm for stationary mixing sequence (see Theorem 2 in Rio (1995)) that for each $1 \leq j \leq 2m_n$,

$$\mathbb{P}(\mathfrak{E}_{n,j}^{(II)} > K \Delta_n^\gamma \mid \mathcal{F}^{(0)}) \leq \mathbb{P}(\mathfrak{E}_{n,j}^{(II)} > K \bar{k}_n^{-1/2} q_{1,n}^{1/2} \sqrt{L_n} \mid \mathcal{F}^{(0)}) = 0.$$

Then by the law of iterated expectation, we have

$$\mathbb{P} \left(\max_{1 \leq j \leq 2m_n} \mathfrak{E}_{n,j}^{(II)} > K \Delta_n^\gamma \right) \leq \sum_{j=1}^{2m_n} \mathbb{P}(\mathfrak{E}_{n,j}^{(II)} > K \Delta_n^\gamma) = 0. \quad (\text{C.67})$$

Combining (C.65)-(C.67) yields

$$\mathbb{P} \left(\max_{1 \leq j \leq 2m_n} \frac{1}{\sqrt{\bar{k}_n}} \sum_{i=1}^{\bar{k}_n} \tilde{Y}_{i,j}^\dagger < x \right) \leq \mathbb{P} \left(\max_{1 \leq j \leq 2m_n} \frac{1}{\sqrt{\bar{k}_n}} \sum_{\ell=1}^{\bar{\ell}_n} \tilde{S}'_{\ell,j} \leq x + K \Delta_n^\gamma \right) + K' \Delta_n^\epsilon.$$

Following a similar argument, we can also show that

$$\mathbb{P} \left(\max_{1 \leq j \leq 2m_n} \frac{1}{\sqrt{\bar{k}_n}} \sum_{i=1}^{\bar{k}_n} \tilde{Y}_{i,j}^\dagger < x \right) \geq \mathbb{P} \left(\max_{1 \leq j \leq 2m_n} \frac{1}{\sqrt{\bar{k}_n}} \sum_{\ell=1}^{\bar{\ell}_n} \tilde{S}'_{\ell,j} \leq x - K \Delta_n^\gamma \right) - K' \Delta_n^\epsilon.$$

Recall for each $1 \leq j \leq 2m_n$, the summand $(\tilde{S}'_{\ell,j})_{1 \leq \ell \leq \bar{\ell}_n}$ is $\mathcal{F}^{(0)}$ -conditionally independent, then a similar strong approximation result can be established following the same proof as in Theorem 4.1.

Bibliography

- Admati, A. R. and Pfleiderer, P. (1988). A theory of intraday patterns: Volume and price variability. *The Review of Financial Studies*, 1(1):3–40.
- Ai, H. and Bansal, R. (2018). Risk preferences and the macroeconomic announcement premium. *Econometrica*, 86(4):1383–1430.
- Aït-Sahalia, Y. and Jacod, J. (2014). *High-frequency financial econometrics*. Princeton University Press.
- Aït-Sahalia, Y. and Xiu, D. (2019). A hausman test for the presence of market microstructure noise in high frequency data. *Journal of Econometrics*, 211(1):176–205.
- Alizadeh, S., Brandt, M. W., and Diebold, F. X. (2002). Range-based estimation of stochastic volatility models. *Journal of Finance*, 57(3):1047–1091.
- Ament, S. and O’Neil, M. (2018). Accurate and efficient numerical calculation of stable densities via optimized quadrature and asymptotics. *Statistics and Computing*, 28:171–185.
- Andersen, T., Thyrgaard, M., and Todorov, V. (2019). Time-varying periodicity in intraday volatility. *Journal of the American Statistical Association*, 114(528):1695–1707.
- Andersen, T. G. (1996). Return volatility and trading volume: An information flow interpretation of stochastic volatility. *The Journal of Finance*, 51(1):169–204.
- Andersen, T. G. and Bollerslev, T. (2018). *Volatility*. Edward Elgar Publishing.

- Andersen, T. G., Bollerslev, T., and Diebold, F. X. (2007). Roughing it up: Disentangling continuous and jump components in measuring, modeling and forecasting asset return volatility. *Review of Economics and Statistics*, 89(4):701–720.
- Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2000). Great realizations. *Risk*, 13:105–108.
- Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica*, 71(2):579–625.
- Andersen, T. G. and Sørensen, B. E. (1996). Gmm estimation of a stochastic volatility model: A monte carlo study. *Journal of Business & Economic Statistics*, 14(3):328–352.
- Ante, L. (2023). How elon musk’s twitter activity moves cryptocurrency markets. *Technological Forecasting and Social Change*, 186:122112.
- Asmussen, S., Glynn, P., and Pitman, J. (1995). Discretization error in simulation of one-dimensional reflecting brownian motion. *The Annals of Applied Probability*, 5(4):875–896.
- Bahadur, R. R. (1966). A note on quantiles in large samples. *The Annals of Mathematical Statistics*, 37(3):577–580.
- Bai, Y., Santos, A., and Shaikh, A. (2019). A practical method for testing many moment inequalities. *University of Chicago, Becker Friedman Institute for Economics Working Paper*, (2019-116).
- Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (2008). Designing realized kernels to measure the ex post variation of equity prices in the presence of noise. *Econometrica*, 76(6):1481–1536.
- Barndorff-Nielsen, O. E. and Shephard, N. (2002). Econometric analysis of realised volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society: Series B*, 64(2):253–280.

- Barndorff-Nielsen, O. E. and Shephard, N. (2004). Econometric analysis of realized covariation: High frequency based covariance, regression, and correlation in financial economics. *Econometrica*, 72(3):885–925.
- Belloni, A., Chernozhukov, V., Chetverikov, D., and Kato, K. (2015). Some new asymptotic theory for least squares series: Pointwise and uniform results. *Journal of Econometrics*, 186(2):345–366.
- Berkson, J. (1950). Are there two regressions? *Journal of the American Statistical Association*, 45(250):164–180.
- Bernanke, B. S. and Kuttner, K. N. (2005). What explains the stock market’s reaction to federal reserve policy? *The Journal of Finance*, 60(3):1221–1257.
- Bernstein, S. (1927). Sur l’extension du théorème limite du calcul des probabilités aux sommes de quantités dépendantes. *Mathematische Annalen*, 97:1–59.
- Bertsimas, D., King, A., and Mazumder, R. (2016). Best subset selection via a modern optimization lens. *The Annals of Statistics*, 44(2):813.
- Bollerslev, T., Li, J., and Li, Q. (2024). Optimal nonparametric range-based volatility estimation. *Journal of Econometrics*, 238(1):105548.
- Bollerslev, T., Li, J., and Liao, Z. (2021). Fixed- k inference for volatility. *Quantitative Economics*, 12(4):1053–1084.
- Bollerslev, T., Li, J., and Xue, Y. (2018). Volume, volatility, and public news announcements. *Review of Economic Studies*, 85(4):2005–2041.
- Bollerslev, T., Patton, A. J., and Quaedvlieg, R. (2016). Exploiting the errors: A simple approach for improved volatility forecasting. *Journal of Econometrics*, 192(1):1–18.
- Borodin, A. N. and Salminen, P. (2002). *Handbook of Brownian Motion-Facts and Formulae*. Birkhäuser, Basel.
- Brillinger, D. R. (1975). Statistical inference for stationary point processes. *Stochastic Processes and Related Topics*, 1:55–99.

- Brown, L. (1968). Inadmissibility of the usual estimators of scale parameters in problems with unknown location and scale parameters. *Annals of Mathematical Statistics*, 39(1):29–48.
- Bugni, F. A., Li, J., and Li, Q. (2023). Permutation-based tests for discontinuities in event studies. *Quantitative Economics*, 14(1):37–70.
- Cattaneo, M. D., Masini, R. P., and Underwood, W. G. (2022). Yurinskii’s coupling for martingales. *arXiv preprint arXiv:2210.00362*.
- Chaudhuri, P. (1991). Nonparametric estimates of regression quantiles and their local bahadur representation. *The Annals of Statistics*, 19(2):760–777.
- Cheng, L., Roulstone, D. T., and Van Buskirk, A. (2021). Are investors influenced by the order of information in earnings press releases? *The Accounting Review*, 96(2):413–433.
- Chernozhukov, V., Chetverikov, D., and Kato, K. (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *The Annals of Statistics*, pages 2786–2819.
- Chernozhukov, V., Chetverikov, D., and Kato, K. (2015). Comparison and anti-concentration bounds for maxima of gaussian random vectors. *Probability Theory and Related Fields*, 162:47–70.
- Chernozhukov, V., Chetverikov, D., and Kato, K. (2017). Central limit theorems and bootstrap in high dimensions. *The Annals of Probability*, 45:2309–2352.
- Chernozhukov, V., Chetverikov, D., and Kato, K. (2019). Inference on causal and structural parameters using many moment inequalities. *The Review of Economic Studies*, 86(5):1867–1900.
- Christensen, K., Hounyo, U., and Podolskij, M. (2018). Is the diurnal pattern sufficient to explain intraday variation in volatility? a nonparametric assessment. *Journal of Econometrics*, 205(2):336–362.
- Christensen, K. and Kolokolov, A. (2023). An unbounded intensity model for point processes. *Available at SSRN 4513848*.

- Christensen, K. and Podolskij, M. (2007). Realized range-based estimation of integrated variance. *Journal of Econometrics*, 141(2):323–349.
- Cieslak, A., Morse, A., and Vissing-Jorgensen, A. (2019). Stock returns over the fomic cycle. *The Journal of Finance*, 74(5):2201–2248.
- Citi, L., Ba, D., Brown, E. N., and Barbieri, R. (2014). Likelihood methods for point processes with refractoriness. *Neural computation*, 26(2):237–263.
- Cochrane, J. H. and Piazzesi, M. (2002). The fed and interest rates - a high-frequency identification. *American Economic Review*, 92(2):90–95.
- Coeurjolly, J.-F. (2008). Hurst exponent estimation of locally self-similar gaussian processes using sample quantiles. *Annals of Statistics*, 36(3):1404–1434.
- Comte, F. and Renault, E. (1998). Long memory in continuous-time stochastic volatility models. *Mathematical Finance*, 8(4):291–323.
- Cox, D. R. (1955). Some statistical methods connected with series of events. *Journal of the Royal Statistical Society: Series B (Methodological)*, 17(2):129–157.
- Cox, J. C., Ingersoll Jr, J. E., and Ross, S. A. (1985). An intertemporal general equilibrium model of asset prices. *Econometrica*, pages 363–384.
- Da, R. and Xiu, D. (2021). When moving-average models meet high-frequency data: Uniform inference on volatility. *Econometrica*, 89(6):2787–2825.
- Davidson, J. (1992). A central limit theorem for globally nonstationary near-epoch dependent functions of mixing processes. *Econometric theory*, 8(3):313–329.
- Diebold, F. X. and Strasser, G. (2013). On the correlation structure of microstructure noise: A financial economic approach. *Review of Economic Studies*, 80(4):1304–1337.
- Du, S. and Zhu, H. (2017). What is the optimal trading frequency in financial markets? *The Review of Economic Studies*, 84(4):1606–1651.
- Durbin, J. and Koopman, S. J. (1997). Monte carlo maximum likelihood estimation for non-gaussian state space models. *Biometrika*, 84(3):669–684.

- Einmahl, U. and Li, D. (2008). Characterization of lil behavior in banach space. *Transactions of the American Mathematical Society*, 360(12):6677–6693.
- Engle, R. F., Giglio, S., Kelly, B., Lee, H., and Stroebel, J. (2020). Hedging climate change news. *The Review of Financial Studies*, 33(3):1184–1216.
- Feller, W. (1951). The asymptotic distribution of the range of sums of independent random variables. *The annals of mathematical statistics*, pages 427–432.
- Foster, D. P. and Nelson, D. B. (1996). Continuous record asymptotics for rolling sample variance estimators. *Econometrica*, 64(1):139–174.
- Gallant, A. R., Te Hsu, C., and Tauchen, G. (1999). Using daily range data to calibrate volatility diffusions and extract the forward integrated variance. *Review of Economics and Statistics*, 81(4):617–631.
- Garman, M. B. and Klass, M. J. (1980). On the estimation of security price volatilities from historical data. *Journal of Business*, 53(1):67–78.
- Gatheral, J., Jaisson, T., and Rosenbaum, M. (2018). Volatility is rough. *Quantitative Finance*, 18(6):933–949.
- Genon-Catalot, V., Laredo, C., and Nussbaum, M. (2002). Asymptotic equivalence of estimating a poisson intensity and a positive diffusion drift. *The Annals of Statistics*, 30(3):731–753.
- Gentzkow, M., Kelly, B., and Taddy, M. (2019). Text as data. *Journal of Economic Literature*, 57(3):535–574.
- Ghosh, J. K. (1971). A new proof of the bahadur representation of quantiles and an application. *The Annals of Mathematical Statistics*, pages 1957–1961.
- Grams, I. and Nussbaum, M. (1998). Asymptotic equivalence for nonparametric generalized linear models. *Probability Theory and Related Fields*, 111:167–214.
- Hájek, J. (1972). Local asymptotic minimax and admissibility in estimation. In *Proceedings of the 6th Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 175–194.

- Hansen, A. L. and Kazinnik, S. (2023). Can chatgpt decipher fedspeak? *Available at SSRN*.
- Hansen, B. E. (2022). A modern gauss-markov theorem. *Econometrica*, forthcoming.
- Hansen, P. R. and Lunde, A. (2006). Realized variance and market microstructure noise. *Journal of Business and Economic Statistics*, 24(2):127–161.
- Harris, L. (1987). Transaction data tests of the mixture of distributions hypothesis. *Journal of financial and Quantitative Analysis*, 22(2):127–141.
- Harvey, A., Ruiz, E., and Shephard, N. (1994). Multivariate stochastic variance models. *The Review of Economic Studies*, 61(2):247–264.
- Hayashi, T., Jacod, J., and Yoshida, N. (2011). Irregular sampling and central limit theorems for power variations: The continuous case. In *Annales de l’IHP Probabilités et statistiques*, volume 47, pages 1197–1218.
- Hayou, S., Doucet, A., and Rousseau, J. (2019). On the impact of the activation function on deep neural networks training. In *International conference on machine learning*, pages 2672–2680. PMLR.
- Helmers, R. and Zitikis, R. (1999). On estimation of poisson intensity functions. *Annals of the Institute of Statistical Mathematics*, 51:265–280.
- Hesse, C. (1990). A bahadur-type representation for empirical quantiles of a large class of stationary, possibly infinite-variance, linear processes. *The Annals of Statistics*, pages 1188–1202.
- Heston, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The Review of Financial Studies*, 6(2):327–343.
- Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30.
- Hong, H. and Wang, J. (2000). Trading and returns under periodic market closures. *The Journal of Finance*, 55(1):297–354.

- Hyslop, R. and Imbens, G. W. (2001). Bias from classical and other forms of measurement error. *Journal of Business and Economic Statistics*, 19(4):475–481.
- Ibragimov, R. and Müller, U. K. (2010). t-statistic based correlation and heterogeneity robust inference. *Journal of Business & Economic Statistics*, 28(4):453–468.
- Ito, R. (2013). Modeling dynamic diurnal patterns in high frequency financial data. *Cambridge Working Papers in Economics*.
- Jaccard, P. (1912). The distribution of the flora in the alpine zone. 1. *New Phytologist*, 11(2):37–50.
- Jacod, J., Li, J., and Liao, Z. (2021). Volatility coupling. *The Annals of Statistics*, 49(4):1982–1998.
- Jacod, J., Li, Y., Mykland, P. A., Podolskij, M., and Vetter, M. (2009). Microstructure noise in the continuous case: the pre-averaging approach. *Stochastic processes and their applications*, 119(7):2249–2276.
- Jacod, J., Li, Y., and Zheng, X. (2017). Statistical properties of microstructure noise. *Econometrica*, 85(4):1133–1174.
- Jacod, J. and Protter, P. (2012). *Discretization of Processes*, volume 67. Springer Science & Business Media.
- Jacod, J. and Rosenbaum, M. (2013). Quarticity and other functionals of volatility: Efficient estimation. *Annals of Statistics*, 41(3):1462–1484.
- Jeganathan, P. (1982). On the convergence of moments of statistical estimators. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 213–232.
- Jeganathan, P. (1983). Some asymptotic properties of risk functions when the limit of the experiment is mixed normal. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 66–87.
- Kalnina, I. and Linton, O. (2008). Estimating quadratic variation consistently in the presence of endogenous and diurnal measurement error. *Journal of Econometrics*, 147(1):47–59.

- Kandel, E. and Pearson, N. D. (1995). Differential interpretation of public signals and trade in speculative markets. *Journal of Political Economy*, 103(4):831–872.
- Karr, A. (1991). *Point Processes and Their Statistical Inference*, volume 7. CRC Press.
- Ke, Z. T., Kelly, B. T., and Xiu, D. (2019). Predicting returns with text data. Technical report, National Bureau of Economic Research.
- Kingman, J. F. (1961). The single server queue in heavy traffic. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 57, pages 902–904. Cambridge University Press.
- Knight, J. L. and Yu, J. (2002). Empirical characteristic function in time series estimation. *Econometric Theory*, 18(3):691–721.
- Koenker, R. (2005). *Quantile regression*, volume 38. Cambridge university press.
- Koenker, R. and Bassett Jr, G. (1978). Regression quantiles. *Econometrica: journal of the Econometric Society*, pages 33–50.
- Kolokolov, A. (2022). Estimating jump activity using multipower variation. *Journal of Business & Economic Statistics*, 40(1):128–140.
- Kristensen, D. (2010). Nonparametric filtering of the realized spot volatility: A kernel-based approach. *Econometric Theory*, 26(1):60–93.
- Kuchibhotla, A. K. and Chakraborty, A. (2022). Moving beyond sub-gaussianity in high-dimensional statistics: Applications in covariance estimation and linear regression. *Information and Inference: A Journal of the IMA*, 11(4):1389–1456.
- Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica*, 53(6):1315–1335.
- Le Cam, L. (1960). Locally asymptotically normal families of distributions. *University of California Publications in Statistics*, 3:37–98.
- Le Cam, L. (1986a). *Asymptotic methods in statistical decision theory*. Springer Series in Statistics.

- Le Cam, L. (1986b). *Asymptotic Theory of Statistical Inference*. John Wiley & Sons.
- Le Cam, L. and Yang, G. L. (2000). *Asymptotics in statistics: some basic concepts*. Springer Science & Business Media.
- Lee, S. (2012). Jumps and information flow in financial markets. *Review of Financial Studies*, 25(2):439–479.
- Lee, S. and Mykland, P. (2008). Jumps in financial markets: A new nonparametric test and jump dynamics. *Review of Financial Studies*, 21(6):2535–2563.
- Lehmann, E., Romano, J. P., and Popper, J. (2005). On optimality of stepdown and stepup multiple test procedures. *The Annals of Statistics*, 33(3):1084–1108.
- Lehmann, E. L. and Casella, G. (1998). *Theory of Point Estimation*. Springer Texts in Statistics.
- Li, J. and Liao, Z. (2020). Uniform nonparametric inference for time series. *Journal of Econometrics*, 219(1):38–51.
- Li, J. and Liu, Y. (2021). Efficient estimation of integrated volatility functionals under general volatility dynamics. *Econometric Theory*, 37(4):664–707.
- Li, J., Todorov, V., and Tauchen, G. (2017). Adaptive estimation of continuous-time regression models using high-frequency data. *Journal of Econometrics*, 200(1):36–47.
- Li, J., Wang, D., and Zhang, Q. (2022). Reading the candlesticks: An ok estimator for volatility. *Review of Economics and Statistics*, forthcoming.
- Li, J. and Xiu, D. (2016). Generalized method of integrated moments for high-frequency data. *Econometrica*, 84(4):1613–1633.
- Li, Z. M. and Linton, O. (2022). A ReMeDI for microstructure noise. *Econometrica*, 90(1):367–389.
- Liu, C. and Tang, C. Y. (2013). A state space model approach to integrated covariance matrix estimation with high frequency data. *Statistics and Its Interface*, 6(4):463.

- Loughran, T. and McDonald, B. (2020). Textual analysis in finance. *Annual Review of Financial Economics*, 12:357–375.
- Lucca, D. O. and Moench, E. (2015). The pre-fomc announcement drift. *Journal of Finance*, 70(1):329–371.
- Miranda-Agrippino, S. and Ricco, G. (2021). The transmission of monetary policy shocks. *American Economic Journal: Macroeconomics*, 13(3):74–107.
- Mogstad, M., Romano, J. P., Shaikh, A. M., and Wilhelm, D. (2023). Inference for ranks with applications to mobility across neighbourhoods and academic achievement across countries. *Review of Economic Studies*, pages 00, 01–23.
- Mykland, P. A. and Zhang, L. (2009). Inference for continuous semimartingales observed at high frequency. *Econometrica*, 77(5):1403–1445.
- Nakamura, E. and Steinsson, J. (2018a). High-frequency identification of monetary non-neutrality: the information effect. *The Quarterly Journal of Economics*, 133(3):1283–1330.
- Nakamura, E. and Steinsson, J. (2018b). Identification in macroeconomics. *Journal of Economic Perspectives*, 32(3):59–86.
- Nelson, D. B. (1990). Arch models as diffusion approximations. *Journal of Econometrics*, 45(1-2):7–38.
- Nison, S. (2001). *Japanese Candlestick Charting Techniques*. Prentice Hall Press, 2nd edn.
- Olver, F. W., Lozier, D. W., Boisvert, R. F., and Clark, C. W. (2010). *NIST Handbook of Mathematical Functions*. Cambridge University Press.
- Parkinson, M. (1980). The extreme value method for estimating the variance of the rate of return. *Journal of Business*, 53:61–65.
- Pötscher, B. M. and Preinerstorfer, D. (2022). A modern gauss-markov theorem? really? working paper, University of Vienna, Department of Statistics.

- Renault, E., Sarisoy, C., and Werker, B. J. (2017). Efficient estimation of integrated volatility and related processes. *Econometric Theory*, 33(2):439–478.
- Resnick, S. I. (2008). *Extreme values, regular variation, and point processes*, volume 4. Springer Science & Business Media.
- Rigobon, R. (2003). Identification through heteroskedasticity. *Review of Economics and Statistics*, 85(4):777–792.
- Rigobon, R. and Sack, B. (2004). The impact of monetary policy on asset prices. *Journal of Monetary Economics*, 51(8):1553–1575.
- Rigobon, R. and Sack, B. (2008). Noisy macroeconomic announcements, monetary policy, and asset prices. In *Asset prices and monetary policy*, pages 335–370. University of Chicago Press.
- Rio, E. (1995). The functional law of the iterated logarithm for stationary strongly mixing sequences. *The Annals of Probability*, 23(3):1188–1203.
- Romano, J. P. and Wolf, M. (2005). Stepwise multiple testing as formalized data snooping. *Econometrica*, 73(4):1237–1282.
- Romer, C. D. and Romer, D. H. (2004). A new measure of monetary shocks: Derivation and implications. *American economic review*, 94(4):1055–1084.
- Ross, S. M. (1995). *Stochastic Processes*. John Wiley & Sons.
- Ruzankin, P. (2004). On the rate of poisson process approximation to a bernoulli process. *Journal of Applied Probability*, 41(1):271–276.
- Savor, P. and Wilson, M. (2014). Asset pricing: A tale of two days. *Journal of Financial Economics*, 113(2):171–201.
- Schennach, S. M. (2020). Mismeasured and unobserved variables. In Durlauf, S. N., Hansen, L. P., Heckman, J. J., and Matzkin, R. L., editors, *Handbook of Econometrics, Volume 7A*, chapter 6, pages 487–565. Elsevier.
- Shaffer, J. P. (1991). The gauss-markov theorem and random regressors. *The American Statistician*, 45(4):269–273.

- Shen, D., Urquhart, A., and Wang, P. (2019). Does twitter predict bitcoin? *Economics letters*, 174:118–122.
- Shephard, N. (2022). Some properties of the sample median of an in-fill sequence of events with an application to high frequency financial econometrics. *Working Paper*.
- Shepp, L. A. (1979). The joint density of the maximum and its location for a wiener process with drift. *Journal of Applied probability*, 16(2):423–427.
- Szulga, A. (1983). On minimal metrics in the space of random variables. *Theory of Probability & Its Applications*, 27(2):424–430.
- Tandon, C., Revankar, S., and Parihar, S. S. (2021). How can we predict the impact of the social media messages on the value of cryptocurrency? insights from big data analytics. *International Journal of Information Management Data Insights*, 1(2):100035.
- Tauchen, G. E. and Pitts, M. (1983). The price variability-volume relationship on speculative markets. *Econometrica*, pages 485–505.
- van der Vaart, A. and Wellner, J. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer Science & Business Media.
- Van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press.
- Vladimirova, M., Girard, S., Nguyen, H., and Arbel, J. (2020). Sub-weibull distributions: Generalizing sub-gaussian and sub-exponential properties to heavier tailed distributions. *Stat*, 9(1):e318.
- Wang, X., Phillips, P. C. B., and Yu, J. (2011). Bias in estimating multivariate and univariate diffusions. *Journal of Econometrics*, 161(2):228–245.
- Welch, W. J. (1982). Algorithmic complexity: three np-hard problems in computational statistics. *Journal of Statistical Computation and Simulation*, 15(1):17–25.
- Wu, W. B. (2005). On the bahadur representation of sample quantiles for dependent sequences. *The Annals of Statistics*, 33(4):1934–1963.

- Yu, B. (1994). Rates of convergence for empirical processes of stationary mixing sequences. *The Annals of Probability*, pages 94–116.
- Zhang, L., Mykland, P. A., and Aït-Sahalia, Y. (2005). A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association*, 100(472):1394–1411.
- Zhang, X. and Cheng, G. (2014). Bootstrapping high dimensional time series. *arXiv preprint arXiv:1406.1037*.
- Zolotarev, V. M. (1986). *One-dimensional stable distributions*, volume 65. American Mathematical Soc.
- Zuo, Y. (2015). Bahadur representations for bootstrap quantiles. *Metrika: International Journal for Theoretical and Applied Statistics*, 78(5):597–610.