Uniform Inference for High-Frequency Data*

Qiyuan Li†

November 11, 2023

JOB MARKET PAPER

For the latest draft, click here.

Abstract

We address the uniform inference problem for high-frequency data that includes prices, volumes, and trading flows. Such data is modeled with a general state-space framework, where latent state process is the corresponding risk indicators, e.g., volatility, price jump, average order size, and arrival of events. The functional estimators are formed as the 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. We apply the proposed methodology to distinguish the informative part from the Federal Open Market Committee speeches, and to analyze the impact of social media activities on cryptocurrency markets.

Keywords: uniform inference, high-frequency data, strong approximation.
JEL Classification: C14, C22, C58.

*I am indebted to my advisor Jia Li for his insightful guidance and constant support on this paper. I would like to thank Chunrong Ai, Liyu Dou, Yifan Li, Morten Nielsen, Peter Phillips, Jun Yu, Yichong Zhang, and other participants at SMU econometrics workshops for valuable discussions and comments.

†School of Economics, Singapore Management University, Singapore; e-mail: qyli.2019@phdecons.smu.edu.sg.

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 the 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, Itô semimartingale plus noise, is clearly not suitable for other market indicators, such as volumes and trading flows.[1] 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 is approximately equal to a general transformation of state process and some random disturbance. Special cases include price, volume, and trading flow, with the corresponding states being volatility, average order size, and trading intensity. Without specific constraints 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 of Li and Xiu (2016). Our emphasis is on the *uniform inference*, which speaks to global properties of the entire state process. Specifically, functional estimators and associated inference procedures are developed for distributional features of transformed state process. The 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)). Based on this insight, our contribution in this paper is to establish a Gaussian coupling theory for functional estimators of both conditional mean process (Theorem 1) and conditional quantile process (Theorem 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 literature has developed involving 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

[1] In contrast to prices, volumes and trading flows are discrete-valued and may not exhibit long-memory properties. Therefore, they cannot fit in the conventional Itô semimartingale model.

volatility functionals (see, e.g., Andersen et al. (2003), Barndorff-Nielsen and Shephard (2004), Mykland and Zhang (2009)) have been extensively explored in the literature.[2,3] However, the uniform inference for the entire volatility process is an emerging problem, as recently explored by Jacod et al. (2021) and Bollerslev et al. (2021). In alignment with this strand of literature, the spot estimation for state process under the 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 the more general state-space setting.[4]

Meanwhile, the inference concerning quantiles is a relatively scarce topic in the realm of high-frequency literature. In a recent paper, Shephard (2022) introduced an estimator of integrated variance based on in-fill medians. The concept of using quantiles holds particular significance when the returns exhibit heavy tails, a common characteristic observed in cryptocurrency markets (see, e.g., Kolokolov (2022)). Our Gaussian strong approximation regarding conditional quantile process is derived, in part, by a novel uniform Bahadur representation for all in-fill quantiles (Lemma A.1). Such representation has been established for i.i.d. data (Bahadur (1966) and Ghosh (1971)) and for weakly dependent stationary data (Hesse (1990) and Wu (2005)), whereas the observations here are nonstationary and can 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 can be applied to tackle other econometric problems. As a byproduct, we provide an application to constructing confidence sets for ranks of spot values of the studied process, which is typically useful in determining arrivals of certain events. Specifically, we rely on Mogstad et al. (2023) to reframe the construction into a multiple hypotheses testing problem, where the valid critical value can be determined through our strong approximation. The paper is also related to prior studies in Gaussian coupling, such as Chernozhukov et al.

[2] These 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)).

[3] Another problem, which is orthogonal to the nonparametric setting, is estimating parameters in the specified volatility dynamics. Such models are suggested by Nelson (1990) and Heston (1993), where the associated estimation methods are proposed in Harvey et al. (1994), Andersen and Sørensen (1996), Durbin and Koopman (1997), and Knight and Yu (2002), among others.

[4] An 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.

(2013), Belloni et al. (2015), Li and Liao (2020), the distinguishing feature of our work lies in its emphasis on a nonstationary time series setting in contrast to the high-dimensional context.

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, Bollerslev et al. (2023) observed that press conferences might sometimes cause more pronounced market impact than the initial release of FOMC statements. We apply the uniform inference procedure to the analysis of trading intensity processes to determine the arrivals of additional information during the press conferences. The comparison of our results to the stand-alone textual analysis suggests the latter tends to smooth out information flow improperly. In view of the growing attention towards generative AI tools and large language models, where tasks are mainly performed with in-context learning, our method serves as a compliment that permits a deployment of supervised learning to achieve higher accuracy. Additionally, we provide an 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 becomes diverging, rendering the detection of abnormal returns no longer valid. Comparing to the outcomes of mean-based t-test in Ante (2023), results using quantile-based measurements of volatile levels indicate more significant 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 2. In Section 3, a Monte-Carlo experiment analysis is conducted. Two empirical studies are presented in Section 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 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 $C \geq 1$.

2 Theory

In Section 2.1, we introduce the state-space model employed in our research. In Section 2.2, three running examples are provided to illustrate adaptability of our framework for modeling different market indicators. Sections 2.3 and 2.4 present constructions and strong approximation results for the functional estimators of both conditional mean processes and conditional quantile processes, respectively. Section 2.5 provides an application in constructing confidence sets for ranks of spot values of the investigated process.

2.1 State-space Model for High-Frequency Data

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

$$Y_{i\Delta_n} = \mathscr{Y}(\zeta_{i\Delta_n}, \varepsilon_{n,i}) + R_{n,i}, \qquad \text{for } 1 \le i \le n, \tag{2.1}$$

where $(\zeta_t)_{t\in[0,T]}$ is a càdlàd 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 $\big(\Omega^{(0)}, \mathcal{F}^{(0)}, (\mathcal{F}_t^{(0)})_{t\ge 0}, \mathbb{P}^{(0)}\big)$. The function $\mathscr{Y}(\cdot,\cdot)$ 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.[5] 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 $(\varepsilon_{n,i})_{1\le i\le n}$ is a $\mathcal{F}^{(0)}$-conditionally independently and identically distributed (i.i.d.) sequence.[6] This is not a necessary condition, as the framework presented here can be extended to accommodate conditionally stationary and weakly dependent disturbances by employing 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 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 A.4 of 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 $(\varepsilon_{n,i})_{1\le i\le n}$

[5]The 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).

[6]There 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 $\mathscr{Y}(\cdot,\cdot)$ account for the dependence structure such that $\varepsilon_{n,i}$ is independent from $\zeta_{i\Delta_n}$.

with its marginal distribution denoted by $\mathbb{P}_\varepsilon$. 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 $\big(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P}\big)$, 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 $\mathscr{Y}(\cdot,\cdot)$ 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).

2.2 Motivating Examples

To facilitate a better understanding of broad implications of the general model (2.1), 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

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

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 (2.1) by setting

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

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

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

where μ is the drift process, σ is the stochastic volatility process, L is a stable process with Blumenthal–Getoor index $\beta \in (0, 2]$ and is assumed to be independent with σ.[7,8] The exten-

[7]Note that for a stable process, Blumenthal–Getoor 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 $\beta \in (0, 2)$.

[8]The 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).

sion 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 (2.1) by setting

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

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 2.3 and 2.4, the residual term can be proved to be uniformly negligible providing processes μ and σ satisfying some fairly weak regularity conditions. □

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 $\Delta_n \mu_{i\Delta_n}$.[9] Consequently, there exists a sequence

[9]The approximation has been explored from a different direction as well, see, e.g., Section 1.6 of Karr (1991)

of independent, uniformly distributed variables $(\varepsilon_{n,i})_{1\le i\le n}$ on $[0,1]$ which are also independent of process μ such that

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

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).[10] Note that increments over disjoint intervals can 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 (2.1) by setting

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

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).[11] 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. □

Aforementioned examples show the general state-space model (2.1) can be cast to model various market indicators such as high-frequency volumes, returns, and trading flows. In the following

where they discuss the optimal approximation of a Bernoulli point process by a Poisson process.

[10] In 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 $\|\mathcal{L}(N) - \mathcal{L}(N^n)\|_{\text{TV}} \le K\Delta_n \sup_{t\in[0,T]} |\mu_t^2|$, where $\|\cdot\|_{\text{TV}}$ denotes the total variation norm of measures. This result aligns with the asymptotic equivalence of statistical experiments in Le Cam's sense, see, e.g., Le Cam (1986) 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).

[11] Timestamps 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.

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.

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.[12] Formally, we define

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

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 $\mathscr{Y}(\cdot,\cdot)$ and the distribution $\mathbb{P}_\varepsilon$. 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 $\{1, \ldots, n\} = \cup_{j=1}^{m_n} \mathcal{I}_{n,j}$, where $\mathcal{I}_{n,j}$ denote the set of $k_{n,j}$ consecutive indices contained in the jth block. Specifically, we define $\iota(i,j) \equiv \min \mathcal{I}_{n,j} + i - 1$ as the ith index in the jth 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 block j such that $t \in \mathcal{T}_{n,j}$, define $\hat{g}_t$ as the local average of

[12] In 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.

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 $\mathscr{Y}(\cdot, \varepsilon)$, as $k_{n,j} \to \infty$ and $k_{n,j}\Delta_n \to 0$,

$$\sqrt{k_{n,j}}(\hat{g}_{n,t} - g_t) \xrightarrow{\mathcal{L}\text{-s}} \mathcal{MN}(0, \sigma_t^2), \tag{2.2}$$

where $\sigma_t^2 \equiv \int_{\mathcal{D}} \mathscr{Y}(\zeta_t, \varepsilon)^2 \mathbb{P}_\varepsilon(d\varepsilon) - \left(\int_{\mathcal{D}} \mathscr{Y}(\zeta_t, \varepsilon)\mathbb{P}_\varepsilon(d\varepsilon)\right)^2$ 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

$$\widehat{C}_{n,t}^{\pm}(\alpha) \equiv \hat{g}_{n,t} \pm z_{1-\alpha/2} \times k_{n,j}^{-1/2} \hat{\sigma}_{n,t}, \tag{2.3}$$

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.,

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

Above results can be easily extended to the case when $R_{n,i} \neq 0$ yet remains uniformly negligible, and furthermore, joint convergence of $\hat{g}_{n,\cdot}$ on a *finite* set of time points $\{t_1, \ldots, t_\ell\} \subset [0, T]$. By a classic Bonferroni approach, the hyperrectangle $C_{n,t_1}^{\pm}(\alpha/\ell) \times \cdots \times C_{n,t_\ell}^{\pm}(\alpha/\ell)$ serves as a valid confidence set for vector $(g_{t_1}, \ldots, g_{t_\ell})$. 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 jth 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 (2.2) 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.[13] To help fix ideas, we define the *sup-t statistic* as

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

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 $\hat{\sigma}_{n,j}^2 \equiv k_{n,j}^{-1} \sum_{i\in\mathcal{I}_{n,j}} Y_{i\Delta_n}^2 - \big(k_{n,j}^{-1} \sum_{i\in\mathcal{I}_{n,j}} Y_{i\Delta_n}\big)^2$. Theorem 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 1. *The observation process $(Y_{i\Delta_n})_{1\leq i\leq n}$ is given by (2.1). 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 $z \neq z'$, $Var\big(\mathscr{Y}(z,\varepsilon)\big)^{-1} + \|\mathscr{Y}(z,\varepsilon) - \mathscr{Y}(z',\varepsilon)\|_{L_2}/\|z - z'\| \leq K_m$;

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

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

Assumption 1 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,

[13] A 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)).

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 $\mathscr{Y}(z,\varepsilon)$ to be locally bounded away from zero, and the random mapping $z \mapsto \mathscr{Y}(z,\varepsilon)$ 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 Chakrabortty (2022) for a detailed discussion of sub-Weilbull tails. This condition holds for any $\eta \geq 1/2$ (resp. $\eta \geq 1$) if $\mathscr{Y}(z,\varepsilon)$ 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 1 (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 1(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 $\mathbb{P}_\varepsilon$ has a sub-Weibull tail, Assumption 1(iii) is met. This, combined with σ maintaining bounded away from zero, leads to the fulfillment of Assumption 1(ii). Recall that in this example residual terms $R_{n,i} = 0$ for all $1 \leq i \leq n$, Assumption 1(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 7, conditional mean process is coherently well-defined only when $\beta = 2$, i.e., L is a Brownian motion.[14] 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 2.4. Assumption 1(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 1(iii) holds for any $\eta \geq 1$, Assumption 1(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 $p \geq 1$,

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

confirming that Assumption 1(iv) holds for any $0 < r < 1$. □

Example 3 (Cox Trading Flows, continued). In the context of Cox trading flow model, recall that state process is $\Delta_n \mu_t$. Assumption 1(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,[15] 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). Note that the transformation takes binary values, Assumption 1(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 $r > 0$,

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

confirming that Assumption 1(iv) also holds for any $r > 0$. □

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

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

[15] This 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)).

Theorem 1. *Suppose that (i) Assumption 1 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 $(Z_1, Z_2, \ldots, Z_{m_n})^\top$ be a standard Gaussian random vector in $\mathbb{R}^{m_n}$. Then for some positive constant ϵ,*

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

Comment 1. Theorem 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 metric.[16] In that case, there exist sequences on a common probability space $\widehat{T}_n' \overset{\mathcal{L}}{=} \widehat{T}_n^*$ and $Z_n' \overset{\mathcal{L}}{=} \max_{1\leq j\leq m_n} |Z_j|$ 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.[17] Consequently, due to the particular usefulness in making inference, Theorem 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 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.[18] Then Theorem 1 implies $\big|\mathbb{P}\big(\widehat{T}_n^* \leq cv_n(\alpha)\big) - \mathbb{P}\big(\max_{1\leq j\leq m_n} |Z_j| \leq cv_n(\alpha)\big)\big| \leq K\Delta_n^{\epsilon}$. Consequently, let

$$\widehat{B}^{\pm}_{n,t}(\alpha) \equiv \hat{g}_{n,t} \pm cv_n(\alpha) \times k_{n,j}^{-1/2}\hat{\sigma}_{n,t}, \qquad \text{for all } t \in \mathcal{T}_{n,j}, \text{ and } 1 \leq j \leq m_n, \tag{2.4}$$

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.,

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

Observing that the uniform confidence band (2.4) is generally wider than pointwise confidence intervals (2.3), this difference magnifies as the number of blocks m_n becomes larger. To better

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

[17] The density of $\max_{1\leq j\leq m_n} |Z_j|$ is given by $f(x) \equiv 2m_n\big(2\Phi(x)-1\big)^{m_n-1}\phi(x)\mathbb{1}\{x \geq 0\}$, where $\phi(\cdot)$ and $\Phi(\cdot)$ denote the density and distribution functions of standard Gaussian distribution, respectively. Note that the Mills ratio $\big(1-\Phi(x)\big)/\phi(x) \to 1/x$, 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 $x \to \sqrt{2\log m_n} + \big(2\log 2 - \log\log m_n - \log(4\pi)\big)/\sqrt{8\log m_n}$ and $m_n \to \infty$.

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

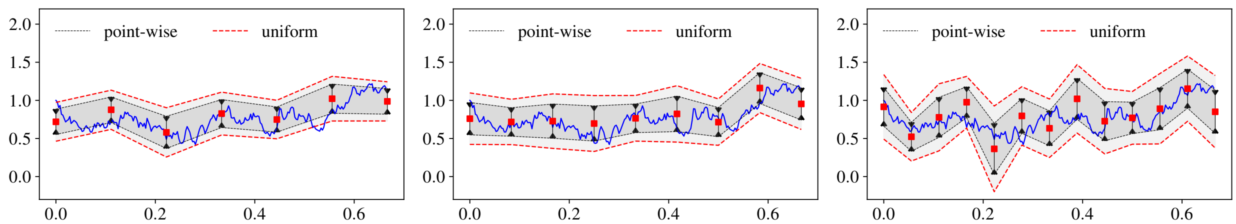

Figure 1: **Comparison of Confidence Bands under Different Numbers of Blocks**. In each panel, we mark spot estimates in red squares, 90% pointwise confidence interval in black vertical segments, 90% uniform confidence band in red dashed lines, and the true process in blue lines. The pointwise confidence band is constructed by connecting each confidence interval computed using (2.3), the uniform confidence band is computed using (2.4). 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.

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 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 proportional 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.

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 $\mathbb{P}_\varepsilon$.[19] 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 $\mathscr{Y}(\zeta_t, x)$, i.e.,

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

[19]Sample 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.

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 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 $Y^o_{1,j} \le \cdots \le Y^o_{k_{n,j},j}$. The spot estimator for conditional quantile, in this scheme, is defined as $\lceil k_{n,j}\chi\rceil$-order statistic.[20] Analogous to the previous section, we form a functional estimator as the collection of all blockwise estimates

$$\hat{q}_{n,j}(\chi) \equiv Y^o_{\lceil k_{n,j}\chi\rceil,j}, \quad \hat{q}_{n,t}(\chi) \equiv \hat{q}_{n,j}(\chi) \text{ for } t \in \mathcal{T}_{n,j} \text{ and } 1 \le j \le m_n.$$

Although observations from model (2.1) 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 (A.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 2. *The observation process $(Y_{i\Delta_n})_{1\le i\le n}$ is given by (2.1). There exists a sequence $(T_m)_{m\ge1}$ of stopping times increasing to infinity, a sequence of compact subsets $(\mathcal{K}_m)_{m\ge1}$ of $\mathcal{Z}$, and a sequence $(K_m)_{m\ge1}$ of positive constants such that:*

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

[20] Note 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 $\lceil k_{n,j}\chi\rceil$-order statistic to avoid unnecessary complexity.

(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* $F(\cdot, x) \equiv \mathbb{P}_\varepsilon\big(\mathscr{Y}(\cdot, \varepsilon) \leq x\big)$;

(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 1, 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.[21] Condition (iv) is the same high-level requirement as in Assumption 1, 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.[22] 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 abnormal 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 1(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 $t \in [0, T]$,

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

[21] Observe 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 absolutely continuity of the distribution. Analysis of sample quantiles for discretely distributed data deserves its own research.

[22] The 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).

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,[23] 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 2, see, e.g., Zolotarev (1986).[24] 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 $0 < r < 1$. □

Analogous to Theorem 1, we present Theorem 2 below, which states the strong approximation result for our functional quantile estimator using the Kolmogorov–Smirnov metric.

Theorem 2. *Suppose that (i) Assumption 2 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 $(Z_1, Z_2, \ldots, Z_{m_n})^\top \sim \mathcal{MN}(0, \mathrm{diag}\{\nu_1^2, \ldots, \nu_{m_n}^2\})$ be a mixed Gaussian random vector in $\mathbb{R}^{m_n}$ such that $\nu_j^2 \equiv \chi(1-\chi)/f_{\tau(1,j)}\big(q_{\tau(1,j)}(\chi)\big)^2$. Then for any $\chi \in (0,1)$, for some positive constant ϵ,*

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

Comment. In contrast to Theorem 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

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

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

[24] Note 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).

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 $\hat{\nu}_{n,j}^2 \equiv \chi(1-\chi)Q(L,\chi)^2\bar{f}\big(Q(L,\chi)\big)^2/\hat{q}_{n,j}(\chi)^2$ for all $1 \le j \le m_n$. Given that $\bar{f}(\cdot)$ is Lipschitz in the neighborhood of $Q(L,\chi)$ by Assumption A.2(iii), Theorem 2 then implies that $\max_{1\le j\le m_n}|\hat{\nu}_{n,j}^2 - \nu_j^2| = O_p\big(\Delta_n^{\rho/2}\log(\Delta_n^{-1})^{1/2}\big)$. Consequently, let $cv_n(\alpha)$ be defined identically as in (2.4), denote

$$\widehat{B}_{n,t}^{\prime\pm}(\alpha) \equiv \big(\hat{q}_{n,t}(\chi) \pm cv_n(\alpha) \times k_{n,j}^{-1/2}\hat{\nu}_{n,j}\big)/Q(L,\chi), \qquad \text{for all } t \in \mathcal{T}_{n,j}, \text{ and } 1 \le j \le m_n, \tag{2.5}$$

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

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

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 $\mathbb{P}(\hat{g}_{t_1} > \hat{g}_{t_2} | g_{t_1} < g_{t_2}) = 1 - \Phi\big(\sqrt{k_n}(g_{t_2} - g_{t_1})/2\big)$, 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 $\{t_1, \ldots, t_{m_n}\}$ 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 $(g_t)_{t\in[0,T]}$. Analogues results can be formulated for conditional quantile process via uniform Bahadur representation and Theorem 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

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

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

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

Let $\mathcal{S}_n^{\mathrm{all}} \equiv \{(j,j') : 1 \leq j, j' \leq m_n \text{ and } j \neq j'\}$ 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

$$H_{j,j'} : g_{n,j} \leq g_{n,j'} \quad \text{against} \quad K_{j,j'} : g_{n,j} > g_{n,j'}, \quad \text{where } (j,j') \in \mathcal{S}_n^{\mathrm{all}}. \tag{2.6}$$

According to which null hypotheses hold true, we can partition all paired indices into two subsets $\mathcal{S}_n^{\mathrm{all},-} \equiv \{(j,j') \in \mathcal{S}_n^{\mathrm{all}} : g_{n,j} \leq g_{n,j'}\}$, $\mathcal{S}_n^{\mathrm{all},+} \equiv \{(j,j') \in \mathcal{S}_n^{\mathrm{all}} : g_{n,j} \geq g_{n,j'}\}$. We also denote the set of rejected hypotheses as $\mathrm{Rej}_n^-(j) \equiv \{(j,j') \in \mathcal{S}_n^{\mathrm{all}} : H_{j',j} \text{ is rejected}\}$ and $\mathrm{Rej}_n^+(j) \equiv \{(j,j') \in \mathcal{S}_n^{\mathrm{all}} : H_{j,j'} \text{ is rejected}\}$. Moreover, define $\mathrm{Rej}_n^\pm \equiv \bigcup_{j=1}^{m_n} \mathrm{Rej}_n^\pm(j)$. Then the familywise error rate for testing family (2.6) can be formally expressed as

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

Our goal is to find a valid test such that $\limsup_{\Delta_n\to 0} \mathbb{P}(\mathrm{FWER}_n) \leq \alpha$. 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 $\mathcal{S}_n^{\mathrm{all},\pm}$ are both random in nature. Secondly, we allow the number of evaluated time points to diverge as $\Delta_n \to 0$ 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 $(j,j') \in \mathcal{S}_n^{\mathrm{all}}$, we construct tests statistic concerning the difference $\hat{g}_{n,j} - \hat{g}_{n,j'}$. Denote the corresponding variance estimator as $\hat{\varsigma}_n(j,j')^2 \equiv \hat{\sigma}_{n,j}^2 + \hat{\sigma}_{n,j'}^2$. Then we reject $H_{j,j'}$ whenever the associated t-statistic

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

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 $\widehat{D}_n \equiv \max_{(j,j')\in\mathcal{S}_n^{\mathrm{all}}} \hat{d}_n(j,j')$.[25] A direct application of Theorem 1 indicates a similar strong approximation result holding for $\widehat{D}_n$. 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 $\mathcal{S}_n^{\mathrm{all}}$ 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

$$cv_n^B(\alpha, \mathcal{S}_n^{\mathrm{all}}) \equiv \inf\left\{x \in \mathbb{R} : \mathbb{P}\left(\max_{(j,j')\in\mathcal{S}_n^{\mathrm{all}}} \frac{\sqrt{k_n}(\hat{g}_{n,j}^B - \hat{g}_{n,j'}^B)}{\hat{\varsigma}_n(j,j')} \leq x \,\middle|\, (Y_{i\Delta_n})_{1\leq i\leq n}\right) \geq 1-\alpha\right\}, \tag{2.7}$$

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

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

[25] Existing literature offers alternative test statistic formulations. For example Bai et al. (2019) suggest using $\widehat{D}_n' \equiv \max_{(j,j')\in\mathcal{S}_n^{\mathrm{all}}} \hat{d}_n(j,j') \vee 0$, 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 $\widehat{D}_n$ instead of $\widehat{D}_n'$ here.

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

(ii) $\mathbb{P}\big(\widehat{D}_n > cv_n^B(\alpha, \mathcal{S}_n^{\text{all}})\big) \geq 1 - K\Delta_n^\epsilon$ *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 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 $\max_{(j,j')\in\mathcal{S}_n^{\text{all}}}(g_{n,j} - g_{n,j'}) = o\big(\Delta_n^{\rho/2}\log(\Delta_n^{-1})^{1/2}\big)$.

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 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.

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 (2.7).

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.

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

Corollary 1. *Under the same setting as Theorem 3. For* $1 \leq j \leq m_n$*, let*

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

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

3 Monte Carlo Simulations

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 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:

$$\begin{aligned} &\text{DGP 1}: \qquad Y_{i\Delta_n} = \mu_{i\Delta_n} + \varepsilon_{n,i}, \text{ where } \varepsilon_{n,i} \sim_{i.i.d.} \mathcal{N}(0,1), \\ &\text{DGP 2}: \qquad Y_{i\Delta_n} = \mu_{i\Delta_n} + \sigma_{i\Delta_n}\varepsilon_{n,i}, \text{ where } \varepsilon_{n,i} \sim_{i.i.d.} t(3). \end{aligned}$$

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 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 Example 2, we simulate price processes with Blumenthal–Getoor 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:

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

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 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:

$$\text{DGP 6}: \ Y_{i\Delta_n} = N_{(i+1)\Delta_n} - N_{i\Delta_n}, \ \text{where } (N_t)_{t\in[0,T]} \text{ is a Cox process with intensity } (\mu_t)_{t\in[0,T]}.$$

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

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

where B and B' are two independent Brownian motions. Following empirical results calibrated in the literature, we choose two parameter configurations summarized in Table 1. Setting (a) is more conservative comparing with setting (b), 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).[26] 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,

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

Table 1: Parameter Specification for the Simulation Study

Setting	$\bar{\mu}_t$	ρ	ς	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.

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. 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 (2.4) and (2.5) for conditional mean processes and conditional median processes, respectively.

3.2 The Results

Table 2 shows the coverage rate of confidence bands (2.4) and (2.5) 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

Table 2: Coverage Rate of Uniform Confidence Band

	$\Delta_n = 1/390$			$\Delta_n = 1/23400$		
DGP	$k_n = 20$	$k_n = 30$	$k_n = 40$	$k_n = 300$	$k_n = 600$	$k_n = 1200$
$1(a)$	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 (2.4) for DGP $1(a)$-$4(b)$, DGP $7(a)$, and $7(b)$, according to (2.5) 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.

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 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 Empirical Illustration

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 (2018). 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. (2023) 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.[27] 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 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.

[27] For 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.

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) and Loughran and McDonald (2020). 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,[28] an expansive language model pioneered by OpenAI, to extract features that could be essential signals indicating a high level of forward guidance.[29] 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 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.

[28]ChatGPT 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.

[29]Recently, 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.

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. 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 $k_n\Delta_n = 1\,\text{sec}$ corresponds to one-second block. In Figure 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 2 shows, there are large

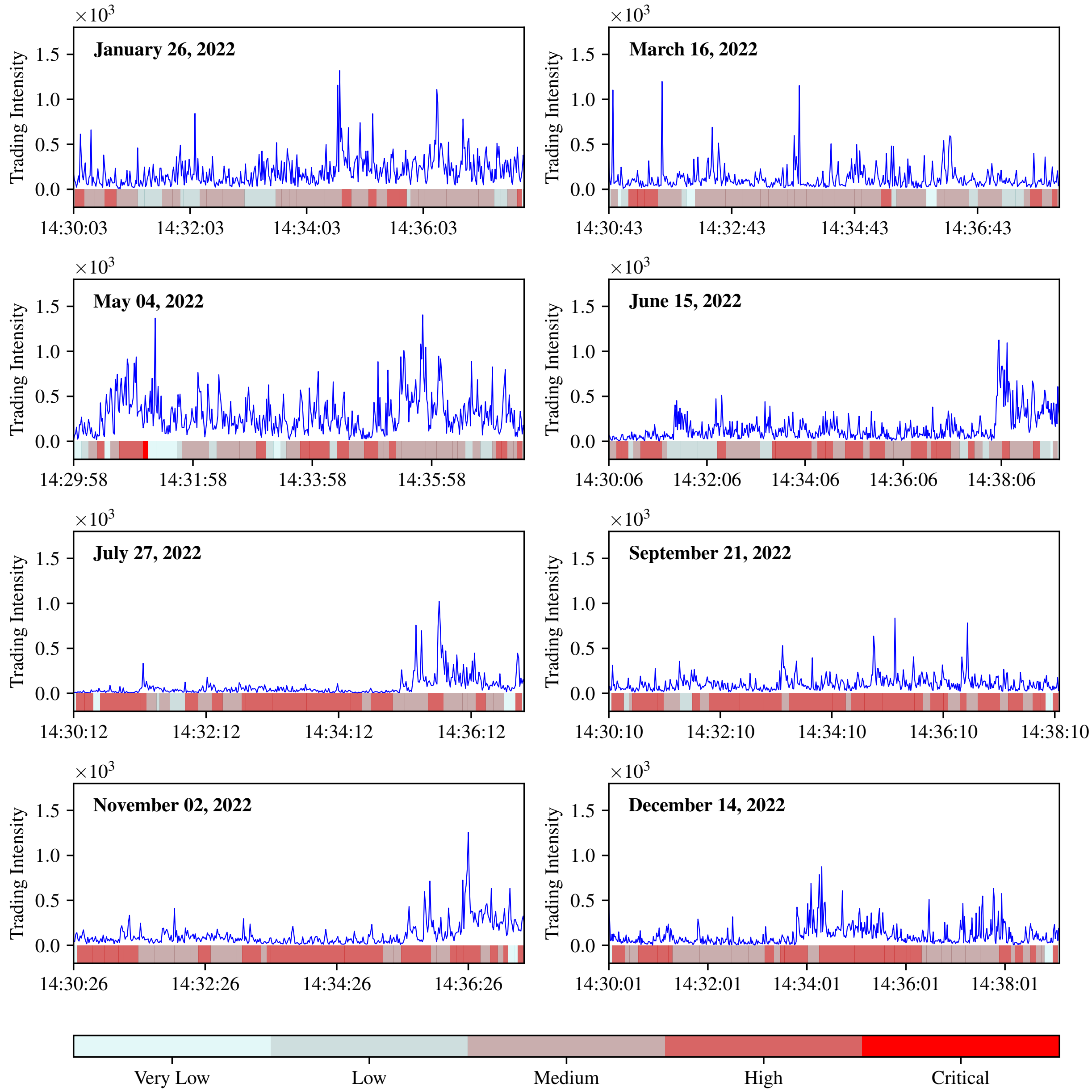

Figure 2: **Trading Intensities and Relative Information Levels during FOMC Press Conference Speeches**. The figure plots one-second trading intensities during eight FOMC press conference speeches in 2022. The horizontal axis is colored according to the relative information level embedded in potential 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$\alpha = \left(s/5, 1 - s/5, 1 - s/5, (s/5)^{1.25}\right)$ where s denotes the information level in the scale of 1 to 5, with 1 being "very low," 5 being "critical."

amount of informative sentences following by barely no intensity variation, indicating the market has 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 $\{0, 1, \ldots, 19\}$ seconds. Figure 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. 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,[30]

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

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 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 2.5, and construct a 90% confidence set for ranks of all second-level intensities. Based on this results, we can partition each speech

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

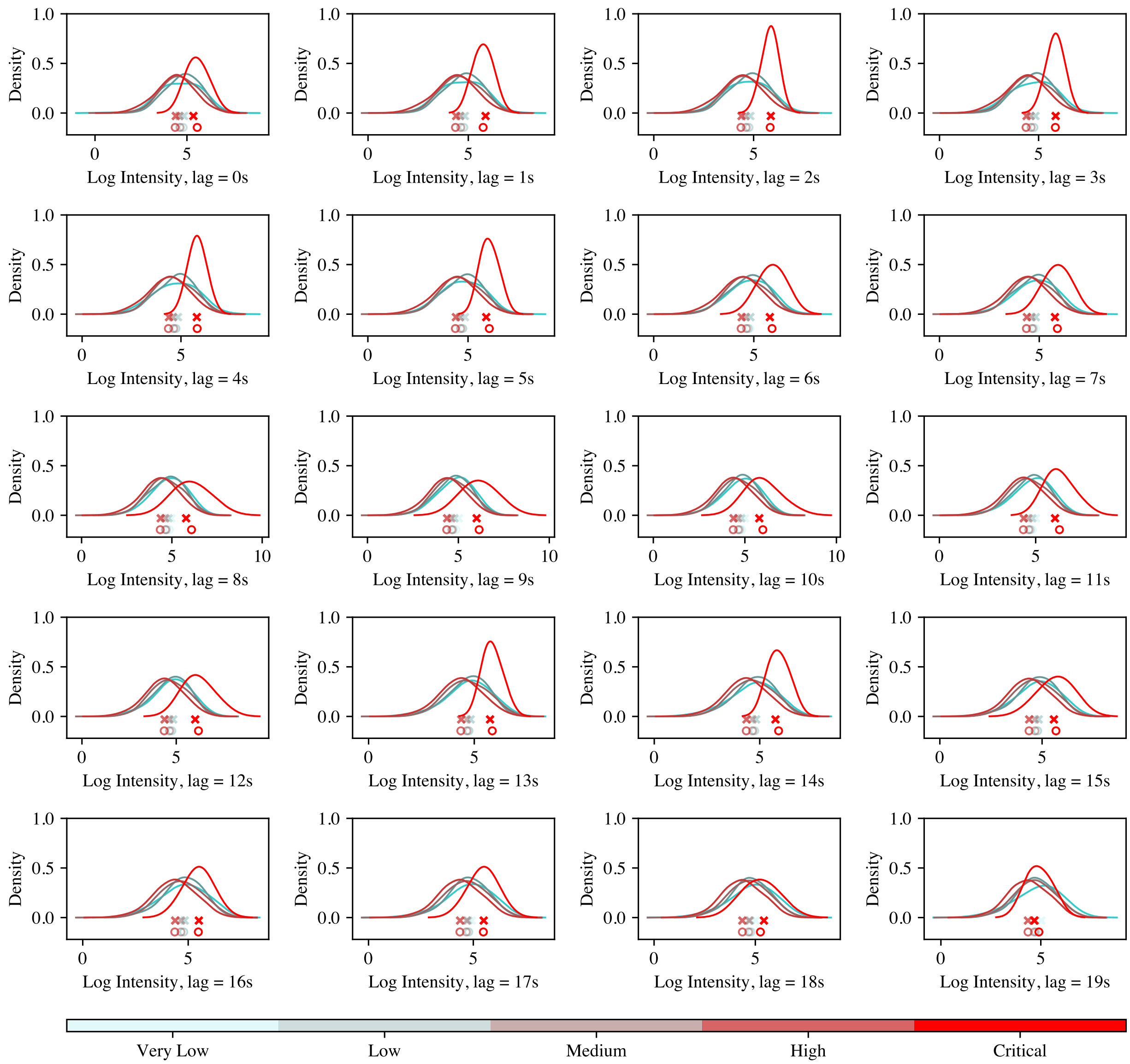

Figure 3: **Distribution of Intensity with Different Information Levels**. The figure plots the kernel density estimation of trading intensities with different relative information level embedded in the potential forward guidance contained in each sentence, which is determined using the algorithm described 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 2, the median and the mean of each group are marked in $\times$ and $\circ$ sign, respectively.

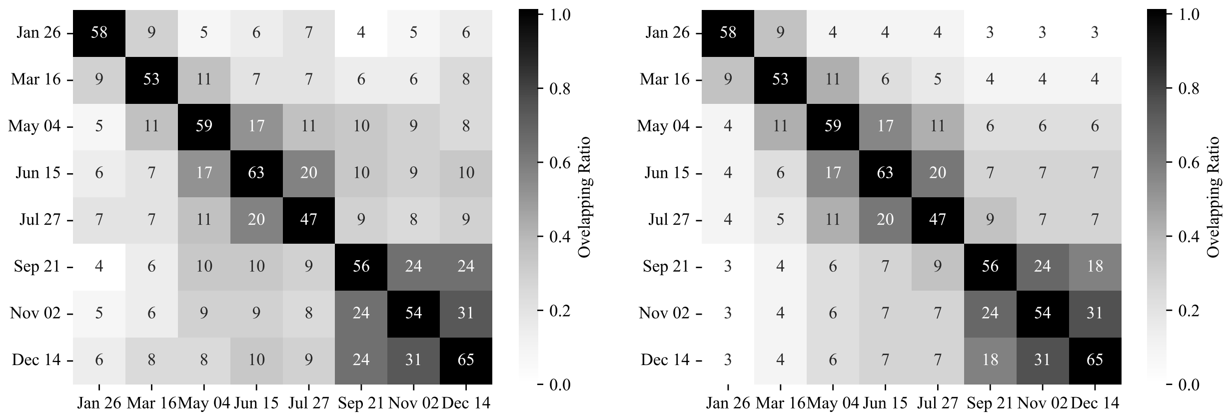

Figure 4: **Similarity of FOMC Press Conference Speeches**. The figure plots the overlapping ratio between different speeches. The overlapping ratio is defined as $\log S_{i,j}^{\text{type}}$, where type $\in \{\text{pw}, \text{cm}\}$. In the left panel, $S_{i,j}^{\text{pw}}$ is the pairwise Jaccard 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 $\{t_i, \ldots, t_j\}$ divided by the total number of sentences. Exact numbers of pairwise and cumulative overlapping sentences are displayed in each square.

into groups $G \in \{1, \ldots, \overline{G}\}$ via the following algorithm: First, we permute indices such that $\hat{g}_{\pi(1)} \leq \hat{g}_{\pi(2)} \leq \cdots \leq \hat{g}_{\pi(m_n)}$. Starting from $\pi(1)$, which initiates the first group $G = 1$, if $\widehat{\text{Rank}}_n\big(\pi(j+1)\big) \cap \widehat{\text{Rank}}_n\big(\pi(1)\big) \neq \varnothing$, then $\pi(j+1)$ belongs to the same group as $\pi(1)$; otherwise, $\pi(j+1)$ initiates a new group $G \leftarrow G + 1$. Repeats until the last second $j = m_n$. In Figure 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.[31]

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.*

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 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 5 indicates an overall absence of significant trading spikes during the December conference. This observation aligns with the result shown in Figure 4 that approximately half of the December speech mirrors

[31] See, 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.

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 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 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.

4.3 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,[32] 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

[32]Namely, 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).

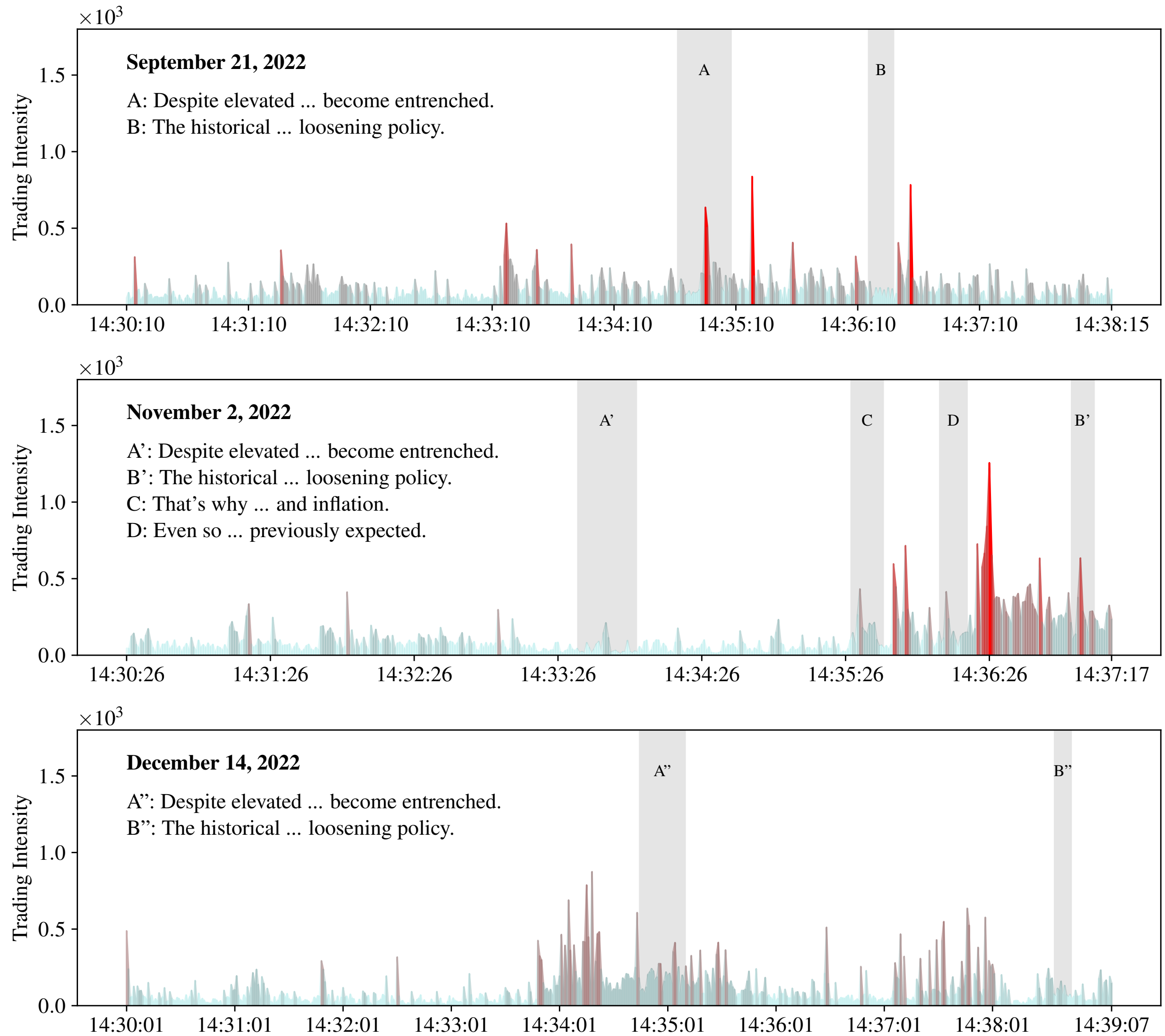

Figure 5: **Trading Intensity during FOMC Press Conference Speeches**. The figure shows the heatbar of estimated trading intensities during FOMC press conference speeches on September 21, November 2, and December 14 in 2022, arranged from the top panel to the bottom panel, respectively. On each of these dates, a 90% confidence set of joint ranks is constructed using Algorithm 1 proposed in section 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$\alpha = \left(G/\overline{G}, 1 - G/\overline{G}, 1 - G/\overline{G}, (G/\overline{G})^{1.25}\right)$ so that the color of each group remains the same as in Figure 2. The duration of target sentences are shaded light gray in each panel, where primes in the label indicate repetitions.

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,[33] and the detection of abnormal returns, as well as associated t-tests, would be invalid.

As discussed in Section 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 interquantile 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 2, and the critical value is computed using Bootstrap.

Table 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.

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

Table 3: Event Study Results for BTC/USD Price

			1 Hour ($m_n = 24$)			2 Hours ($m_n = 12$)		
No.	Time & Date	Tweet	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 (*Twitter bio change*)	1.94*	1.24	2.98**	1.76	9.87***	4.68***
4	08:18 Feb 10, 2021	This is true power haha (*picture about Bitcoin*)	-0.51	0.80	1.14	0.36	0.92	1.58
5	00:42 Feb 21, 2021	Cryptocurrency explained (*link to a video*)	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 (*picture about suspending Bitcoin*)	-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 (*picture for Bitcoin*)	-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] (*picture of a couple's conversation*)	-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. interquantile 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.

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." In particular, the model accommodates Lévy-driven returns and Cox trading flow processes. We allow state processes to have undefined dynamics, and propose uniform inference procedure for the entire conditional mean processes and the entire conditional quantile processes of transformed states.

To construct functional estimators for the investigated processes, we collect 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 that allows for valid uniform inference, and can be used to tackle other economics problems like constructing confidence set for ranks of spot values of certain 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 to compare trading intensity under one-second level and can be used to precisely pinpoint the informative part of a speech. The proposed inference procedure serves as a compliment to existing methodologies, like volatility-based detection mechanisms and conventional textual analysis tools. We also apply the proposed inference procedure to analyze the impact of Elon Musk's tweets on cryptocurrency markets, the case in which mean-based tests may fail due to heavy tails in returns. The results using quantile-based measurements of volatile levels indicate significant price impact over an extended time window following the posting of tweets.

Appendix: Proofs

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 $L_n \equiv \log(\Delta_n^{-1})$.

A.1 **Proofs for Section 2.3**

Proof of Theorem 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 1 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 A.1.

Assumption A.1. *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 $z \neq z'$, $Var\big(\mathscr{Y}(z,\varepsilon)\big)^{-1} + \|\mathscr{Y}(z,\varepsilon) - \mathscr{Y}(z',\varepsilon)\|_{L_2}/\|z - z'\| \leq K$; (iii) for all $x > 0$ and $z \in \mathcal{K}$, $\mathbb{P}_\varepsilon\big(|\mathscr{Y}(z,\varepsilon)| \geq x\big) \leq K \exp\{-(x/K)^{1/\eta}\}$ 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 $G_p(\cdot) \equiv \int_{\mathcal{D}} \mathscr{Y}(\cdot,\varepsilon)^p \mathbb{P}_\varepsilon(d\varepsilon)$, we have for all $z \in \mathcal{K}$, $Var\big(\mathscr{Y}(z,\varepsilon)\big) = G_2(z) - G_1^2(z)$ is bounded away from zero. Note that by Theorem 2.1 in Vladimirova et al. (2020), Assumption A.1(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)),

$$\sup_{z \in \mathcal{K}} \Big\| \max_{1 \leq j \leq m_n} \mathscr{Y}(z, \varepsilon_j) \Big\|_{L_p} \leq K_p (\log m_n)^\eta \leq K_p L_n^\eta. \tag{A.1}$$

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

$$\epsilon < \frac{\rho}{6} \wedge \Big(\frac{1}{6} - \frac{\rho}{3}\Big) \wedge \Big(\frac{r}{3} - \frac{\rho}{6}\Big).$$

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

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

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

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

Note that by the above construction, the variables $\widetilde{Y}_{i,j}$ 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

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

The proof is divided into three parts. In Step 1, we establish that $\widehat{T}_n^*$ can be strong approximated by $\widetilde{T}_n^*$ in the following sense:

$$\mathbb{P}(|\widehat{T}_n^* - \widetilde{T}_n^*| > \delta_n) \leq K \Delta_n^{\epsilon}, \tag{A.2}$$

for some real sequence satisfying $\delta_n \to 0$ and $\delta_n \sqrt{L_n} \leq K \Delta_n^{\epsilon}$. In Step 2, we construct $(Z_j)_{1 \leq j \leq m_n}$ and prove the validity of the following inequality for $\widetilde{T}_n^*$:

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P}(\widetilde{T}_n^* \leq x) - \mathbb{P}\Big(\max_{1 \leq j \leq m_n} |Z_j| \leq x \Big) \right| \leq K \Delta_n^{\epsilon}. \tag{A.3}$$

Step 3 concludes the proof by establishing the asserted statement.

Step 1. Note that we can rewrite $\widehat{T}_n^* = \max_{1 \leq j \leq m_n} \sup_{t \in \mathcal{T}_{n,j}} |\sqrt{k_{n,j}} (\hat{g}_{n,j} - g_t)/\hat{\sigma}_{n,j}|$. By simple algebra we can verify that $|(a-b)/c - a/d| \leq |d/c - 1| \times |(a-b)/d| + |b/d|$. Recall equation (2.1), the proof of this step thus relies on the following decomposition

$$|\widehat{T}_n^* - \widetilde{T}_n^*| \leq \max_{1 \leq j \leq m_n} \left| \frac{\sigma_{n,j}}{\hat{\sigma}_{n,j}} - 1 \right| \times \max_{1 \leq j \leq m_n} \left| \frac{1}{\sqrt{k_{n,j}}} \sum_{i=1}^{k_{n,j}} \frac{\widetilde{Y}_{i,j}}{\sigma_{n,j}} \right| + \max_{1 \leq j \leq m_n} |\mathfrak{A}_{n,j}|, \tag{A.4}$$

where for $1 \leq j \leq m_n$, $\mathfrak{A}_{n,j} \equiv \mathfrak{A}_{n,j}^{(I)} + \mathfrak{A}_{n,j}^{(II)}$ with

$$\begin{aligned}
\mathfrak{A}_{n,j}^{(I)} &\equiv \frac{1}{\sqrt{k_{n,j}}} \sum_{i=1}^{k_{n,j}} \frac{R_{n,\iota(i,j)}}{\sigma_{n,j}}, \\
\mathfrak{A}_{n,j}^{(II)} &\equiv \sup_{t \in \mathcal{T}_{n,j}} \frac{1}{\sqrt{k_{n,j}}} \sum_{i=1}^{k_{n,j}} \frac{g_{\tau(i,j)} - g_t}{\sigma_{n,j}}.
\end{aligned}$$

Note that by Assumption A.1(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 A.1(iv), together with $k_{n,j} \asymp \Delta_n^{-\rho}$, implies that

$$\max_{1\leq j\leq m_n} |\mathfrak{A}_{n,j}^{(I)}| \leq K\Delta_n^{-\rho/2} \max_{1\leq i\leq n} |R_{n,i}| = o_p(\Delta_n^{r-\rho/2}) = o_p(\Delta_n^{\epsilon+\gamma}). \tag{A.5}$$

Note that Assumption A.1(ii) implies function $G_1(\cdot)$ is Lipschitz since by the triangle inequality and Hölder inequality $|G_1(z) - G_1(z')| \leq \|\mathscr{Y}(z,\varepsilon) - \mathscr{Y}(z',\varepsilon)\|_{L_2}$. Also note that $m_n \asymp \Delta_n^{\rho-1}$ by $k_{n,j} \asymp \Delta_n^{-\rho}$, applying a maximal inequality, we have

$$\Big\|\max_{1\leq j\leq m_n} \mathfrak{A}_{n,j}^{(II)}\Big\|_{L_p} \leq K_p m_n^{1/p} \max_{1\leq j\leq m_n} \sqrt{k_{n,j}}(k_{n,j}\Delta_n)^{1/2} \leq K_p \Delta_n^{(\rho-1)/p+1/2-\rho}. \tag{A.6}$$

Taking $p > (1-\rho)/(1/2-\rho-\epsilon-\gamma)$, the right-hand side becomes $o(\Delta_n^{\epsilon+\gamma})$. Then combining (A.5) and (A.6), it follows the triangle inequality and the Hölder inequality that

$$\max_{1\leq j\leq m_n} |\mathfrak{A}_{n,j}| \leq \max_{1\leq j\leq m_n} |\mathfrak{A}_{n,j}^{(I)}| + \max_{1\leq j\leq m_n} |\mathfrak{A}_{n,j}^{(II)}| = o_p(\Delta_n^{\epsilon+\gamma}). \tag{A.7}$$

For $1 \leq j \leq m_n$ and $1 \leq i \leq k_{n,j}$, denote

$$\tilde{\sigma}_{n,j}^2 \equiv \frac{1}{k_{n,j}} \sum_{i=1}^{k_{n,j}} \widetilde{Y}_{i,j}^2 - \left(\frac{1}{k_{n,j}} \sum_{i=1}^{k_{n,j}} \widetilde{Y}_{i,j}\right)^2.$$

Equation (A.5) and (A.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

$$\max_{1\leq j\leq m_n}\left|\frac{\hat{\sigma}_{n,j}}{\sigma_{n,j}}-1\right| \leq \max_{1\leq j\leq m_n}\left|\frac{\tilde{\sigma}_{n,j}}{\sigma_{n,j}}-1\right|+K\max_{1\leq j\leq m_n}|\hat{\sigma}_{n,j}-\tilde{\sigma}_{n,j}| \leq \max_{1\leq j\leq m_n}\left|\frac{\tilde{\sigma}_{n,j}}{\sigma_{n,j}}-1\right|+o_p(\Delta_n^{\epsilon+\gamma}). \tag{A.8}$$

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 $\widetilde{U}_{i,j}$ and $v_{i,j}$ as follows:

$$\begin{aligned}
\widetilde{U}_{i,j} &\equiv \sqrt{\frac{\bar{k}_n}{k_{n,j}}}\frac{\widetilde{Y}_{i,j}}{\sigma_{n,j}}\mathbb{1}\{1 \leq i \leq k_{n,j}\},\\
v_{i,j} &\equiv \frac{\bar{k}_n(G_2(\zeta_{\tau(i,j)}) - g_{\tau(i,j)}^2)}{k_{n,j}\sigma_{n,j}^2}\mathbb{1}\{1 \leq i \leq k_{n,j}\}.
\end{aligned}$$

By construction the variables $\widetilde{U}_{i,j}$ 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

$$\frac{\tilde{\sigma}_{n,j}^2}{\sigma_{n,j}^2} - 1 = \left(\frac{1}{k_{n,j}} \sum_{i=1}^{k_{n,j}} \frac{\widetilde{Y}_{i,j}^2}{\sigma_{n,j}^2} - 1\right) - \left(\frac{1}{k_{n,j}} \sum_{i=1}^{k_{n,j}} \frac{\widetilde{Y}_{i,j}}{\sigma_{n,j}}\right)^2 = \left(\frac{1}{\bar{k}_n} \sum_{i=1}^{\bar{k}_n} \widetilde{U}_{i,j}^2 - 1\right) - \left(\frac{1}{\bar{k}_n} \sum_{i=1}^{\bar{k}_n} \widetilde{U}_{i,j}\right)^2.$$

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

$$\begin{aligned}\mathbb{P}\left(\max_{1\leq j\leq m_n}\left|\frac{\tilde{\sigma}_{n,j}}{\sigma_{n,j}}-1\right|>x\middle|\mathcal{F}^{(0)}\right) &\leq \mathbb{P}\left(\max_{1\leq j\leq m_n}\left|\frac{1}{\bar{k}_n}\sum_{i=1}^{\bar{k}_n}\widetilde{U}_{i,j}^2-1\right|>\frac{x}{2}\middle|\mathcal{F}^{(0)}\right)\\ &\quad+\mathbb{P}\left(\max_{1\leq j\leq m_n}\left|\frac{1}{\bar{k}_n}\sum_{i=1}^{\bar{k}_n}\widetilde{U}_{i,j}\right|>\sqrt{\frac{x}{2}}\middle|\mathcal{F}^{(0)}\right). \qquad (A.9)\end{aligned}$$

For the first term, noting that by Assumption A.1(ii), we have

$$\max_{1\leq j\leq m_n}\sum_{i=1}^{k_{n,j}}\mathbb{E}[\widetilde{U}_{i,j}^4|\mathcal{F}^{(0)}] \leq K\max_{1\leq j\leq m_n}\sum_{i=1}^{k_{n,j}}G_4(\zeta_{\tau(i,j)}) \leq K\Delta_n^{-\rho}.$$

By (A.1), we can further deduce for each $1\leq i\leq \bar{k}_n$,

$$\mathbb{E}\Big[\max_{1\leq j\leq m_n}\widetilde{U}_{i,j}^4\Big|\mathcal{F}^{(0)}\Big] \leq K\sup_{z\in\mathcal{K}}\mathbb{E}\Big[\max_{1\leq j\leq m_n}\mathscr{Y}(z,\varepsilon_{n,\iota(i,j)})^4\Big] \leq KL_n^{4\eta}. \qquad (A.10)$$

Then by a maximal inequality, we obtain

$$\mathbb{E}\Big[\max_{1\leq i\leq \bar{k}_n}\max_{1\leq j\leq m_n}\widetilde{U}_{i,j}^4\Big|\mathcal{F}^{(0)}\Big] \leq K\Delta_n^{-\rho}L_n^{4\eta}.$$

Observing that by the definition of $v_{i,j}$ and $\sigma_{n,j}$, we can verify $\bar{k}_n^{-1}\sum_{i=1}^{\bar{k}_n}\mathbb{E}[\widetilde{U}_{i,j}^2|\mathcal{F}^{(0)}] = \bar{k}_n^{-1}\sum_{i=1}^{\bar{k}_n}v_{i,j} = \sigma_{n,j}/\sigma_{n,j} = 1$. Then by Lemma 8 in Chernozhukov et al. (2015), we obtain

$$\mathbb{E}\left[\max_{1\leq j\leq m_n}\left|\frac{1}{\bar{k}_n}\sum_{i=1}^{\bar{k}_n}\widetilde{U}_{i,j}^2-1\right|\middle|\mathcal{F}^{(0)}\right] \leq K(\Delta_n^{\rho/2}\sqrt{L_n}+\Delta_n^{\rho/2}L_n^{1+2\eta}).$$

Therefore, a Fuk–Nagaev type inequality (see Theorem 4 in Einmahl and Li (2008)) implies that for every $x>0$,

$$\begin{aligned}&\mathbb{P}\left(\max_{1\leq j\leq m_n}\left|\frac{1}{\bar{k}_n}\sum_{i=1}^{\bar{k}_n}\widetilde{U}_{i,j}^2-1\right|>K\Delta_n^{\rho/2}L_n^{1+2\eta}+x\middle|\mathcal{F}^{(0)}\right)\\ &\leq \exp\{-K'x^2\Delta_n^{-\rho}\}+K'x^{-2}\Delta_n^{\rho}L_n^{4\eta}.\end{aligned}$$

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

$$\mathbb{P}\left(\max_{1\leq j\leq m_n}\left|\frac{1}{\bar{k}_n}\sum_{i=1}^{\bar{k}_n}\widetilde{U}_{i,j}^2-1\right|>K\Delta_n^{\rho(1-\varpi)/2}L_n^{1+2\eta}\middle|\mathcal{F}^{(0)}\right) \leq K'\Delta_n^{\rho\varpi}. \qquad (A.11)$$

Similarly, noting that $\bar{k}_n^{-1}\sum_{i=1}^{\bar{k}_n}\mathbb{E}[\widetilde{U}_{i,j}|\mathcal{F}^{(0)}]=0$ and by (A.10) together with a maximal inequality, we have $\mathbb{E}[\max_{1\le i\le k_n}\max_{1\le j\le m_n}\widetilde{U}_{i,j}^2|\mathcal{F}^{(0)}]\le K\Delta_n^{-\rho/2}L_n^{2\eta}$. Applying Lemma 8 in Chernozhukov et al. (2015) again, we can obtain

$$\mathbb{E}\left[\max_{1\le j\le m_n}\left|\frac{1}{\bar{k}_n}\sum_{i=1}^{\bar{k}_n}\widetilde{U}_{i,j}\right|\middle|\mathcal{F}^{(0)}\right]\le K(\Delta_n^{\rho/2}\sqrt{L_n}+\Delta_n^{3\rho/4}L_n^{1+\eta}). \tag{A.12}$$

Then the Fuk–Nagaev type inequality implies that for every $x>0$,

$$\begin{aligned}\mathbb{P}\left(\max_{1\le j\le m_n}\left|\frac{1}{\bar{k}_n}\sum_{i=1}^{\bar{k}_n}\widetilde{U}_{i,j}\right|>K(\Delta_n^{\rho/2}L_n^{1+\eta})+x\middle|\mathcal{F}^{(0)}\right)&\\ \le\exp\{-K'x^2\Delta_n^{-\rho}\}+K'x^{-4}\Delta_n^{3\rho}L_n^{4\eta}.&\end{aligned}$$

Taking $x\asymp\Delta_n^{\rho/4}L_n^{\eta}$, the right-hand side is bounded by $\exp\{-K\Delta_n^{-\rho/2}L_n^{2\eta}\}+K\Delta_n^{2\rho}\le K'\Delta_n^{2\rho}$. Consequently, we have

$$\mathbb{P}\left(\max_{1\le j\le m_n}\left|\frac{1}{\bar{k}_n}\sum_{i=1}^{\bar{k}_n}\widetilde{U}_{i,j}\right|>K\Delta_n^{\rho/4}L_n^{1+\eta}\middle|\mathcal{F}^{(0)}\right)\le K'\Delta_n^{2\rho}. \tag{A.13}$$

Combining (A.9), (A.11) and (A.13), noting that $\rho(1-\varpi)/2<\rho/2$, by the law of iterated expectation, for all $\varpi\ge\epsilon/\rho$, we obtain

$$\mathbb{P}\left(\max_{1\le j\le m_n}\left|\frac{\tilde{\sigma}_{n,j}}{\sigma_{n,j}}-1\right|>K\Delta_n^{\rho(1-\varpi)/2}L_n^{1+2\eta}\right)\le K'\Delta_n^{\epsilon}.$$

Also note that $|a-1|\le x/(x+1)$ implies $|a^{-1}-1|\le x$, combining the above inequality with (A.8), we conclude that for $\varpi>(\epsilon/\rho)\vee(1-2\gamma/\rho)$,

$$\mathbb{P}\left(\max_{1\le j\le m_n}\left|\frac{\sigma_{n,j}}{\hat{\sigma}_{n,j}}-1\right|>K\Delta_n^{\rho(1-\varpi)/2}L_n^{1+2\eta}\right)\le K'\Delta_n^{\epsilon}. \tag{A.14}$$

Moreover, recall (A.12) and the definition of $\widetilde{U}_{i,j}$, by the law of iterated expectation and the Markov inequality, for all $\varpi<1-4\epsilon/\rho$, we can show

$$\mathbb{P}\left(\max_{1\le j\le m_n}\left|\frac{1}{\sqrt{k_{n,j}}}\sum_{i=1}^{k_{n,j}}\frac{\widetilde{Y}_{i,j}}{\sigma_{n,j}}\right|>K\Delta_n^{\rho(\varpi-1)/4}\sqrt{L_n}\right)\le K'\Delta_n^{\epsilon}, \tag{A.15}$$

Combining (A.4), (A.7), (A.14), and (A.15), by the Markov inequality, the desired inequality (A.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 $\delta_n\to 0$ and $\delta_n\sqrt{L_n}\le K\Delta_n^{\epsilon}$. 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 $\widetilde{U}^{\dagger}_{i,j}$ as

$$\widetilde{U}^{\dagger}_{i,j} \equiv \widetilde{U}_{i,j} \mathbb{1}\{1 \leq j \leq m_n\} - \widetilde{U}_{i,j-m_n} \mathbb{1}\{m_n + 1 \leq j \leq 2m_n\}.$$

Observing that by the definition of $\widetilde{T}^*_n$ and $\widetilde{U}^{\dagger}_{i,j}$, we can rewrite

$$\widetilde{T}^*_n \equiv \max_{1 \leq j \leq m_n} \left| \frac{1}{\sqrt{k_{n,j}}} \sum_{i=1}^{k_{n,j}} \frac{\widetilde{Y}_{i,j}}{\sigma_{n,j}} \right| = \max_{1 \leq j \leq m_n} \left| \frac{1}{\sqrt{\bar{k}_n}} \sum_{i=1}^{\bar{k}_n} \widetilde{U}_{i,j} \right| = \max_{1 \leq j \leq 2m_n} \frac{1}{\sqrt{\bar{k}_n}} \sum_{i=1}^{\bar{k}_n} \widetilde{U}^{\dagger}_{i,j}.$$

Recall that $(\widetilde{U}_{i,j})_{1 \leq i \leq \bar{k}_n, 1 \leq j \leq m_n}$ are $\mathcal{F}^{(0)}$-conditionally independent, centered random variables. Let $(\widetilde{Z}_{i,j})_{1 \leq i \leq \bar{k}_n, 1 \leq j \leq m_n}$ be a sequence of $\mathcal{F}^{(0)}$-conditionally independent, centered Gaussian random variables with conditional variance $\mathbb{E}[\widetilde{Z}^2_{i,j} | \mathcal{F}^{(0)}] = \mathbb{E}[\widetilde{U}^2_{i,j} | \mathcal{F}^{(0)}] = v_{i,j}$. Further, for each $1 \leq i \leq \bar{k}_n$ and $1 \leq j \leq 2m_n$, let

$$\widetilde{Z}^{\dagger}_{i,j} \equiv \widetilde{Z}_{i,j} \mathbb{1}\{1 \leq j \leq m_n\} - \widetilde{Z}_{i,j-m_n} \mathbb{1}\{m_n + 1 \leq j \leq 2m_n\},$$

which implies $\mathbb{E}[\widetilde{Z}^{\dagger}_{i,j} \widetilde{Z}^{\dagger}_{i',j} | \mathcal{F}^{(0)}] = \mathbb{E}[\widetilde{U}^{\dagger}_{i,j} \widetilde{U}^{\dagger}_{i',j} | \mathcal{F}^{(0)}]$ 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 A.1(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 A.1(ii) implies, for $p \in \{3, 4\}$, and $1 \leq j \leq 2m_n$,

$$\frac{1}{\bar{k}_n} \sum_{i=1}^{\bar{k}_n} \mathbb{E}[|\widetilde{U}^{\dagger}_{i,j}|^p | \mathcal{F}^{(0)}] \leq K_p \frac{1}{k_{n,j}} \sum_{i=1}^{k_{n,j}} G_p(\zeta_{\tau(i,j)}) / \sigma^p_{n,j} \leq K_p.$$

Combining with Assumption A.1(iii) and (A.10), by Proposition 2.1 in Chernozhukov et al. (2017), we obtain for all $\epsilon < \rho/6$ that

$$\begin{aligned} &\sup_{x \in \mathbb{R}} \left| \mathbb{P}(\widetilde{T}^*_n \leq x | \mathcal{F}^{(0)}) - \mathbb{P}\left(\max_{1 \leq j \leq 2m_n} \frac{1}{\sqrt{\bar{k}_n}} \sum_{i=1}^{\bar{k}_n} \widetilde{Z}^{\dagger}_{i,j} \leq x \middle| \mathcal{F}^{(0)} \right) \right| \\ &\quad \leq K(\Delta_n^{\rho/6} L_n^{7/6 + \eta/3} + \Delta_n^{\rho/6} L_n^{1 + 2\eta/3}) \leq K \Delta_n^{\epsilon}. \end{aligned}$$

For $1 \leq j \leq m_n$, define $Z_j \equiv \bar{k}_n^{-1/2} \sum_{i=1}^{\bar{k}_n} \widetilde{Z}_{i,j}$. Recalling the definition of $\widetilde{Z}_{i,j}$ and $\sigma_{n,j}$, we conclude

$$(Z_1, Z_2, \ldots, Z_{m_n})^{\top} | \mathcal{F}^{(0)} \sim \mathcal{N}(0, I_{m_n}).$$

Since the right hand side is a pivot, $(Z_j)_{1\le j\le m_n}$ remains standard Gaussian unconditionally, hence satisfies the requirement in the assertion. Note that by construction we have

$$\max_{1\le j\le 2m_n} \frac{1}{\sqrt{\bar{k}_n}} \sum_{i=1}^{\bar{k}_n} \widetilde{Z}^{\dagger}_{i,j} = \max_{1\le j\le m_n} \left| \frac{1}{\sqrt{\bar{k}_n}} \sum_{i=1}^{\bar{k}_n} \widetilde{Z}_{i,j} \right| = \max_{1\le j\le m_n} |Z_j|.$$

Equation (A.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 1. Combining the results in (A.2) and (A.3), we observe that

$$\begin{aligned}
&\sup_{x\in\mathbb{R}}\bigg(\mathbb{P}(\widehat{T}^*_n \le x) - \mathbb{P}\Big(\max_{1\le j\le m_n}|Z_j| \le x\Big)\bigg)\\
&\le \mathbb{P}(|\widehat{T}^*_n - \widetilde{T}^*_n| > \delta_n) + \sup_{x\in\mathbb{R}}\bigg(\mathbb{P}(\widetilde{T}^*_n \le x+\delta_n) - \mathbb{P}\Big(\max_{1\le j\le m_n}|Z_j| \le x+\delta_n\Big)\bigg)\\
&\quad + \sup_{x\in\mathbb{R}}\mathbb{P}\Big(x < \max_{1\le j\le m_n}|Z_j| \le x+\delta_n\Big)\\
&\le K\Delta_n^{\epsilon},
\end{aligned}$$

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

$$\mathbb{E}\Big[\max_{1\le j\le m_n}|Z_j|\Big] \le K\sqrt{L_n},$$

and $\delta_n\sqrt{L_n} \le K\Delta_n^{\epsilon}$ by construction of δ_n. Similarly, we can show

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

This competes the proof of required statement. *Q.E.D.*

A.2 Proofs for Section 2.4

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 2 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 A.2.

Assumption A.2. *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 $1 \le j \le m_n$, and for each $p > 0$, $\mathbb{E}[\|\zeta_t - \zeta_s\|^p] \le K_p|t-s|^{p/2}$ for some constant K_p; (ii) for each $x \in \mathbb{R}$, for all $z, z' \in \mathcal{K}$, $|F(z,x) - F(z',x)| \vee |\partial_x F(z,x) - \partial_x F(z',x)| \le K\|z - z'\|$; (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) $\max_{1\le i\le n}|R_{n,i}| = o_p(\Delta_n^r)$ for some $r > 0$.*

The proof of Theorem 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 A.1 (Uniform Bahadur Representation)**.** *Suppose Assumption A.2 holds. For $1 \leq j \leq m_n$, denote*

$$\tilde{q}_{n,j} \equiv q_{n,j} + \frac{\sqrt{\chi(1-\chi)}}{f_{n,j}(q_{n,j})} \frac{1}{k_{n,j}} \sum_{i=1}^{k_{n,j}} \frac{F(\zeta_{\tau(i,j)}, q_{n,j}) - \mathbb{1}\{\mathscr{Y}(\zeta_{\tau(i,j)}, \varepsilon_{n,\iota(i,j)}) \leq q_{n,j}\}}{\sqrt{F(\zeta_{\tau(i,j)}, q_{n,j})\big(1 - F(\zeta_{\tau(i,j)}, q_{n,j})\big)}}.$$

Then we have for each $\chi \in (0,1)$, and for some positive ϵ and γ,

$$\mathbb{P}\Big(\max_{1\leq j\leq m_n} \sqrt{k_{n,j}}|\hat{q}_{n,j} - \tilde{q}_{n,j}| > K\Delta_n^{\gamma}\Big) \leq K'\Delta_n^{\epsilon}.$$

Proof of Lemma A.1. We prove the validity of the assertion for all positive ϵ and γ such that

$$\epsilon + \gamma < \frac{\rho}{4} \wedge \Big(\frac{1}{2} - \rho\Big) \wedge \Big(r - \frac{\rho}{2}\Big).$$

Let $\widetilde{Y}_{i,j} \equiv \mathscr{Y}(\zeta_{\tau(i,j)}, \varepsilon_{n,\iota(i,j)})$, within each block j reindex the sequence $(\widetilde{Y}_{i,j})_{1\leq i\leq k_{n,j}}$ in the non-decreasing order and denote as $\widetilde{Y}^o_{1,j} \leq \cdots \leq \widetilde{Y}^o_{k_{n,j},j}$. Note that in each block, there are at least $\lceil k_{n,j}\chi\rceil$ of $\widetilde{Y}_{i,j}$ no larger than $Y^o_{\lceil k_{n,j}\chi\rceil,j} + \max_{i\in\mathcal{I}_{n,j}} |R_{n,i}|$, which implies $\widetilde{Y}^o_{\lceil k_{n,j}\chi\rceil,j} \leq Y^o_{\lceil k_{n,j}\chi\rceil,j} + \max_{i\in\mathcal{I}_{n,j}} |R_{n,i}|$. Similarly, there are at least $k_{n,j} - \lceil k_{n,j}\chi\rceil$ of $\widetilde{Y}_{i,j}$ no smaller than $Y^o_{\lceil k_{n,j}\chi\rceil,j} - \max_{i\in\mathcal{I}_{n,j}} |R_{n,i}|$, which implies $\widetilde{Y}^o_{\lceil k_{n,j}\chi\rceil,j} \geq Y^o_{\lceil k_{n,j}\chi\rceil,j} - \max_{i\in\mathcal{I}_{n,j}} |R_i|$. Therefore, assumption A.2(iv) implies that

$$\max_{1\leq j\leq m_n} \sqrt{k_{n,j}}|\hat{g}_{n,j} - \widetilde{Y}^o_{\lceil k_{n,j}\chi\rceil,j}| \leq K\Delta_n^{-\rho/2} \max_{1\leq i\leq n} |R_{n,i}| = o_p(\Delta_n^{r-\rho/2}) = o_p(\Delta_n^{\epsilon+\gamma}). \tag{A.16}$$

For each $1 \leq j \leq m_n$, let $\widetilde{F}_{n,j}(x) \equiv k_{n,j}^{-1}\sum_{i=1}^{k_{n,j}} \mathbb{1}\{\widetilde{Y}_{i,j} \leq x\}$ be the empirical distribution function of $(\widetilde{Y}_{i,j})_{1\leq i\leq k_{n,j}}$. 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 $\widetilde{F}_{n,j}(\cdot)$ 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 $\widetilde{Y}^o_{\lceil k_{n,j}\chi\rceil,j}$ 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

$$S_{n,j}(x) \equiv \widetilde{F}_{n,j}(x) - \widetilde{F}_{n,j}(q_{n,j}) - \frac{1}{k_{n,j}} \sum_{i=1}^{k_{n,j}} \big(F(\zeta_{\tau(i,j)}, x) - \chi\big). \tag{A.17}$$

For any set $A \subseteq \mathbb{R}$, denote $\overline{S}_{n,j}(A) \equiv \sup_{x\in A}|S_{n,j}(x)|$. 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 $\overline{I}_{n,j} \equiv (q_{n,j} - \varkappa_{1,n}, q_{n,j} + \varkappa_{1,n})$. 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 $\overline{I}_{n,j} \subseteq \bigcup_{\ell=-\varkappa_{2,n}}^{\varkappa_{2,n}-1} I_{n,j}(\ell)$. Note that both $\widetilde{F}_{n,j}(\cdot)$ and $F(z,\cdot)$ are nondecreasing functions, we have for $x \in I_{n,j}(\ell)$,

$$
\begin{aligned}
S_{n,j}(x) &\leq \widetilde{F}_{n,j}\big(\psi_{n,j}(\ell+1)\big) - \widetilde{F}_{n,j}(q_{n,j}) - \frac{1}{k_{n,j}}\sum_{i=1}^{k_{n,j}}\big(F(\zeta_{\tau(i,j)}, \psi_{n,j}(\ell)) - \chi\big) \\
&\leq S_{n,j}\big(\psi_{n,j}(\ell+1)\big) + \vartheta_{n,j}(\ell),
\end{aligned}
$$

where

$$
\vartheta_{n,j}(\ell) \equiv \frac{1}{k_{n,j}}\sum_{i=1}^{k_{n,j}} F\big(\zeta_{\tau(i,j)}, \psi_{n,j}(\ell+1)\big) - \frac{1}{k_{n,j}}\sum_{i=1}^{k_{n,j}} F\big(\zeta_{\tau(i,j)}, \psi_{n,j}(\ell)\big).
$$

Similarly, we also have $S_{n,j}(x) \geq S_{n,j}\big(\psi_{n,j}(\ell)\big) - \vartheta_{n,j}(\ell)$. 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 $\overline{I}_{n,j}$ that

$$
\overline{S}_{n,j}(\overline{I}_{n,j}) \leq \overline{S}_{n,j}\left(\bigcup_{\ell=-\psi_{2,n}}^{\psi_{2,n}-1} I_{n,j}(\ell)\right) \leq \max_{-\varkappa_{2,n}\leq\ell\leq\varkappa_{2,n}} \big|S_{n,j}\big(\psi_{n,j}(\ell)\big)\big| + \bar{\vartheta}_{n,j}. \tag{A.18}
$$

For the second term, note that $|\psi_{n,j}(\ell) - q_{n,j}| \leq \varkappa_{1,n} \to 0$ for $|\ell| \leq \varkappa_{2,n}$. Then by Assumption A.2(iii) and the mean value theorem, recall that $\gamma < \rho/4 - \epsilon$, we have for n sufficiently large,

$$
\begin{aligned}
\max_{1\leq j\leq m_n}\sqrt{k_{n,j}}\bar{\vartheta}_{n,j} &\leq K\max_{1\leq j\leq m_n}\max_{-\varkappa_{2,n}\leq\ell\leq\varkappa_{2,n}-1}\sqrt{k_{n,j}}|\psi_{n,j}(\ell+1) - \psi_{n,j}(\ell)| \\
&= K\varkappa_{1,n}\varkappa_{2,n}^{-1}\max_{1\leq j\leq m_n}\sqrt{k_{n,j}} \\
&\leq K\Delta_n^{\rho/4}L_n \leq K\Delta_n^{\epsilon+\gamma}.
\end{aligned} \tag{A.19}
$$

For the first term in the right-hand side of (A.18), first consider a fixed $1 \leq j \leq m_n$. For each $-\varkappa_{2,n} \leq \ell \leq \varkappa_{2,n}$, let $\big(\xi_{i,j}(\ell)\big)_{1\leq i\leq k_{n,j}}$ be a sequence of $\mathcal{F}^{(0)}$-conditionally independent, Bernoulli random variables with parameter $\big(|F(\zeta_{\tau(i,j)}, \psi_{n,j}(\ell)) - F(\zeta_{\tau(i,j)}, q_{n,j})|\big)_{1\leq i\leq k_{n,j}}$ 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 (A.17),

$$
k_{n,j}\big|S_{n,j}\big(\psi_{n,j}(\ell)\big)\big| \overset{\mathcal{L}|\mathcal{F}^{(0)}}{=} \left|\Xi_{n,j}(\ell) - \sum_{i=1}^{k_{n,j}}\big(F(\zeta_{\tau(i,j)}, \psi_{n,j}(\ell)) - \chi\big)\right|.
$$

In view of above equation, by the triangle inequality, we have for all $x \in \mathbb{R}$,

$$
\begin{aligned}
&\Big\{\max_{1\le j\le m_n}\max_{-\varkappa_{2,n}\le \ell\le \varkappa_{2,n}} \sqrt{k_{n,j}}S_{n,j}\big(\psi_{n,j}(\ell)\big)\ge x\Big\}\\
&\subseteq \Big\{\max_{1\le j\le m_n}\max_{-\varkappa_{2,n}\le \ell\le \varkappa_{2,n}} \frac{1}{\sqrt{k_{n,j}}}\Big|\Xi_{n,j}(\ell)-\sum_{i=1}^{k_{n,j}}\big(F(\zeta_{\tau(i,j)},\psi_{n,j}(\ell))-F(\zeta_{\tau(i,j)},q_{n,j})\big)\Big|\ge \frac{x}{2}\Big\}\\
&\quad\cup\Big\{\max_{1\le j\le m_n}\frac{1}{\sqrt{k_{n,j}}}\sum_{i=1}^{k_{n,j}}|F(\zeta_{\tau(i,j)},q_{n,j})-\chi|\ge\frac{x}{2}\Big\}\\
&=\Big\{\max_{1\le j\le m_n}\max_{-\varkappa_{2,n}\le \ell\le \varkappa_{2,n}}\mathfrak{B}^{(I)}_{n,j}(\ell)\ge\frac{x}{2}\Big\}\cup\Big\{\max_{1\le j\le m_n}\mathfrak{B}^{(II)}_{n,j}\ge\frac{x}{2}\Big\}, \qquad (A.20)
\end{aligned}
$$

where for $1 \le j \le m_n$ and $-\varkappa_{2,n} \le \ell \le \varkappa_{2,n}$,

$$
\begin{aligned}
\mathfrak{B}^{(I)}_{n,j}(\ell) &\equiv \frac{1}{\sqrt{k_{n,j}}}\Big|\Xi_{n,j}(\ell)-\sum_{i=1}^{k_{n,j}}\big(F(\zeta_{\tau(i,j)},\psi_{n,j}(\ell))-F(\zeta_{\tau(i,j)},q_{n,j})\big)\Big|,\\
\mathfrak{B}^{(II)}_{n,j} &\equiv \frac{1}{\sqrt{k_{n,j}}}\sum_{i=1}^{k_{n,j}}|F(\zeta_{\tau(i,j)},q_{n,j})-\chi|.
\end{aligned}
$$

For the second term, note that Assumption A.2(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

$$
\begin{aligned}
\mathbb{P}(|q_t-q_s|>x) &\le \mathbb{P}(q_t-q_s>x)+\mathbb{P}(q_s-q_t>x)\\
&\le \mathbb{P}\big(F(\zeta_t,q_s+x)<\chi\big)+\mathbb{P}\big(F(\zeta_s,q_t+x)<\chi\big)\\
&\le \mathbb{P}\big(F(\zeta_s,q_s+x)-K\|\zeta_s-\zeta_t\|<\chi\big)+\mathbb{P}\big(F(\zeta_t,q_t+x)-K\|\zeta_s-\zeta_t\|<\chi\big)\\
&\le 2\mathbb{P}\big(\|\zeta_s-\zeta_t\|>Kx\big), \qquad (A.21)
\end{aligned}
$$

where the second line is by the fact that $F(z,x)$ is increasing in x, the third line is by Assumption A.2(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 A.2(i) and (A.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

$$
\Big\|\max_{1\le j\le m_n}\mathfrak{B}^{(II)}_{n,j}\Big\|_{L_p}\le K_p m_n^{1/p}\max_{1\le j\le m_n}\sqrt{k_{n,j}}(k_{n,j}\Delta_n)^{1/2}\le K_p\Delta_n^{(\rho-1)/p+1/2-\rho}.
$$

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

$$
\mathbb{P}\Big(\max_{1\le j\le m_n}\mathfrak{B}^{(II)}_{n,j}\ge K\Delta_n^{\gamma}\Big)\le K'\Delta_n^{\epsilon}. \qquad (A.22)
$$

For the first term inside the max operator in the right-hand side of (A.20), by the Bernstein inequality (see, e.g., bound (2.13) under Theorem 3 of Hoeffding (1963)), we have for all $x \in \mathbb{R}^+$,

$$\mathbb{P}\big(\sqrt{k_{n,j}}\mathfrak{B}_{n,j}^{(I)}(\ell) \geq x \big| \mathcal{F}^{(0)}\big) \leq 2\exp\Big\{-\frac{x^2/2}{\sum_{i=1}^{k_{n,j}}\big|F\big(\zeta_{\tau(i,j)}, \psi_{n,j}(\ell)\big) - F(\zeta_{\tau(i,j)}, q_{n,j})\big| + x}\Big\}. \quad \text{(A.23)}$$

According Assumption A.2(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

$$\sum_{i=1}^{k_{n,j}}\big|F\big(\zeta_{\tau(i,j)}, \psi_{n,j}(\ell)\big) - F(\zeta_{\tau(i,j)}, q_{n,j})\big| \leq M_1 k_{n,j}\varkappa_{1,n}. \quad \text{(A.24)}$$

Note that the right-hand side bound of above equation is deterministic and does note depend on ℓ. Therefore, combining (A.23) and (A.24), we can conclude that

$$\begin{aligned}\mathbb{P}\Big(\max_{-\varkappa_{2,n}\leq \ell \leq \varkappa_{2,n}} \mathfrak{B}_{n,j}^{(I)}(\ell) \geq M_2\Delta_n^{\rho/4}L_n\Big|\mathcal{F}^{(0)}\Big) &\leq \sum_{\ell=-\varkappa_{2,n}}^{\varkappa_{2,n}} \mathbb{P}\big(\mathfrak{B}_{n,j}^{(I)}(\ell) \geq M_2\Delta_n^{\rho/4}L_n\big|\mathcal{F}^{(0)}\big) \\ &\leq 4\varkappa_{2,n}\exp\Big\{-\frac{M_2^2 k_{n,j}\Delta_n^{\rho/2}L_n^2/2}{M_1 k_{n,j}\varkappa_{1,n} + M_2\sqrt{k_{n,j}}\Delta_n^{\rho/4}L_n}\Big\}.\end{aligned}$$

Let $\mathscr{O}_n(M_1, M_2)$ denote the right-hand side bound of the above display. Note that by the definition of $\varkappa_{1,n}$, as $\Delta_n \to 0$ (or equivalently, as $n \to \infty$), we have

$$\frac{\log\big(\mathscr{O}_n(M_1, M_2)\big)}{\log n} \to \frac{\rho}{4} - \frac{M_2^2}{2M_1}.$$

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 $\sum_{n=1}^{\infty}\mathscr{O}(M_1, M_2) < \infty$. Then by the Borel–Cantelli lemma, we conclude that

$$\mathbb{P}\Big(\limsup_{n\to\infty}\max_{-\varkappa_{2,n}\leq \ell \leq \varkappa_{2,n}} \mathfrak{B}_{n,j}^{(I)}(\ell) \geq M_2\Delta_n^{\rho/4}L_n\Big|\mathcal{F}^{(0)}\Big) = 0.$$

Note that $\gamma < \rho/4 - \epsilon$, then by the law of iterated expectation, we have for n sufficiently large

$$\begin{aligned}&\mathbb{P}\Big(\max_{1\leq j\leq m_n}\max_{-\varkappa_{2,n}\leq \ell \leq \varkappa_{2,n}} \mathfrak{B}_{n,j}^{(I)}(\ell) \geq K\Delta_n^{\gamma}\Big) \\ &\quad\leq \sum_{j=1}^{m_n}\mathbb{P}\Big(\max_{-\varkappa_{2,n}\leq \ell \leq \varkappa_{2,n}} \mathfrak{B}_{n,j}^{(I)}(\ell) \geq M_2\Delta_n^{\rho/4}L_n\Big) = 0. \end{aligned} \quad \text{(A.25)}$$

Combining (A.18)-(A.22), and (A.25), we conclude that

$$\mathbb{P}\Big(\max_{1\leq j\leq m_n}\sqrt{k_{n,j}}\,\overline{S}_{n,j}(\overline{I}_{n,j}) \geq K\Delta_n^{\gamma}\Big) \leq K'\Delta_n^{\epsilon}. \quad \text{(A.26)}$$

Step 2. Recall the definition of $\widetilde{Y}^o_{\lceil k_{n,j}\chi\rceil,j}$ and $\widetilde{F}_{n,j}(\cdot)$, for each $1\leq j\leq m_n$, we have $\widetilde{Y}^o_{\lceil k_{n,j}\chi\rceil,j}\leq p_{n,j}-\varkappa_{1,n}$ if and only if $k_{n,j}\widetilde{F}_{n,j}(q_{n,j}-\varkappa_{1,n})\geq\lceil k_{n,j}\chi\rceil$. Therefore,

$$\begin{aligned}&\{\exists 1\leq j\leq m_n \text{ such that } \widetilde{Y}^o_{\lceil k_{n,j}\chi\rceil,j}\leq p_{n,j}-\varkappa_{1,n}\}\\&\quad=\Big\{\max_{1\leq j\leq m_n}\big(k_{n,j}\widetilde{F}_{n,j}(q_{n,j}-\varkappa_{1,n})-\lceil k_{n,j}\chi\rceil\big)\geq 0\Big\}.\end{aligned}\tag{A.27}$$

Let $(\xi'_{i,j})$ be a sequence of $\mathcal{F}^{(0)}$-conditionally independent, Bernoulli random variables with parameter $\big(F(\zeta_{\tau(i,j)},q_{n,j}-\varkappa_{1,n})\big)_{1\leq i\leq k_{n,j}}$ respectively. Let $\Xi'_{n,j}\equiv\sum_{i=1}^{k_{n,j}}\xi'_{i,j}$ denote their convolution. By the construction, we have

$$k_{n,j}\widetilde{F}_{n,j}(q_{n,j}-\varkappa_{1,n})\overset{\mathcal{L}|\mathcal{F}^{(0)}}{=}\Xi'_{n,j}.\tag{A.28}$$

Note that Assumption A.2(i)-(iii) imply that

$$\max_{1\leq i\leq k_{n,j}}|F(\zeta_{\tau(i,j)},q_{n,j}-\varkappa_{1,n})-\chi|\leq M\Big(\max_{1\leq i\leq k_{n,j}}\|\zeta_{\tau(i,j)}-\zeta_{\tau(1,j)}\|+\varkappa_{1,n}\Big).$$

Observe that in the right-hand side of above display, by Assumption A.2(i), we have

$$\Big\|\max_{1\leq j\leq m_n}\max_{1\leq i\leq k_{n,j}}(\zeta_{\tau(i,j)}-\zeta_{\tau(1,j)})\Big\|_{L_p}\leq K_p m_n^{1/p}\max_{1\leq j\leq m_n}(k_{n,j}\Delta_n)^{1/2}\leq K_p\Delta_n^{(\rho-1)/p+(1-\rho)/2}.\tag{A.29}$$

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

$$E_{n,1}\equiv\Big\{\max_{1\leq j\leq m_n}\max_{1\leq i\leq k_{n,j}}\|\zeta_{\tau(i,j)}-\zeta_{\tau(1,j)}\|<\varkappa_{1,n}\Big\}.$$

Therefore, from (A.29), by the Markov inequality and the law of iterated expectation, we conclude that $\mathbb{P}(E_{n,1}^{\complement})\leq K\Delta_n^{\epsilon}$. In view of (A.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

$$\begin{aligned}&\Big\{\max_{1\leq j\leq m_n}\big(k_{n,j}\widetilde{F}_{n,j}(q_{n,j}-\varkappa_{1,n})-\lceil k_{n,j}\chi\rceil\big)\geq 0\Big\}\cap E_{n,1}\\&\quad\subseteq\Big\{\max_{1\leq j\leq m_n}\Big(\Xi'_{n,j}-\sum_{i=1}^{k_{n,j}}F(\zeta_{\tau(i,j)},q_{n,j}-\varkappa_{1,n})\Big)\geq 1-(M+K\Delta_n^{-\rho})\varkappa_{1,n}\Big\}\cap E_{n,1}\\&\quad\subseteq\Big\{\max_{1\leq j\leq m_n}\Big(\Xi'_{n,j}-\sum_{i=1}^{k_{n,j}}F(\zeta_{\tau(i,j)},q_{n,j}-\varkappa_{1,n})\Big)\geq -K\Delta_n^{-\rho}\varkappa_{1,n}\Big\}\cap E_{n,1}.\end{aligned}$$

For the term inside the max operator of above display, it follows the Bernstein inequality that

$$\begin{aligned}&\mathbb{P}\bigg(\Big\{\Xi'_{n,j}-\sum_{i=1}^{k_{n,j}}F(\zeta_{\tau(i,j)},q_{n,j}-\varkappa_{1,n})\geq -K\Delta_n^{-\rho}\varkappa_{1,n}\Big\}\cap E_{n,1}\bigg|\mathcal{F}^{(0)}\bigg)\\&\quad\leq\exp\bigg\{-\frac{(-K\Delta_n^{-\rho}\varkappa_{1,n})^2}{2\Big(\sum_{i=1}^{k_{n,j}}F(\zeta_{\tau(i,j)},q_{n,j}-\varkappa_{1,n})-K\Delta_n^{-\rho}\varkappa_{1,n}\Big)}\bigg\}.\\&\quad\leq\exp\bigg\{-\frac{K\Delta_n^{-2\rho}\varkappa_{1,n}^2}{2k_{n,j}\chi}\bigg\},\end{aligned}$$

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 $E_{n,1}$,

$$\mathbb{P}\Big(\Big\{\limsup_{n\to\infty} k_{n,j}\widetilde{F}_{n,j}(q_{n,j} - \varkappa_{1,n}) \geq \lceil k_{n,j}\chi\rceil\Big\} \cap E_{n,1}\Big|\mathcal{F}^{(0)}\Big) = 0.$$

Then by the law of iterated expectation, we have for n sufficiently large

$$\begin{aligned}&\mathbb{P}\Big(\Big\{\max_{1\leq j\leq m_n}\big(\widetilde{F}_{n,j}(q_{n,j} - \varkappa_{1,n}) - \lceil k_{n,j}\chi\rceil\big) \geq 0\Big\} \cap E_{n,1}\Big)\\ &\quad \leq \sum_{j=1}^{m_n}\mathbb{P}\big(\{(\widetilde{F}_{n,j}(q_{n,j} - \varkappa_{1,n}) - \lceil k_{n,j}\chi\rceil) \geq 0\} \cap E_{n,1}\big) = 0. \end{aligned} \tag{A.30}$$

Combining (A.27) and (A.30) yields for n sufficiently large,

$$\begin{aligned}&\mathbb{P}(\{\exists 1 \leq j \leq m_n \text{ such that } \widetilde{Y}^o_{\lceil k_{n,j}\chi\rceil,j} \leq p_{n,j} - \varkappa_{1,n}\})\\ &\quad\leq \mathbb{P}\Big(\Big\{\max_{1\leq j\leq m_n}\big(\widetilde{F}_{n,j}(q_{n,j} - \varkappa_{1,n}) - \lceil k_{n,j}\chi\rceil\big) \geq 0\Big\} \cap E_{n,1}\Big) + \mathbb{P}(E_{n,1}^{\complement})\\ &\quad\leq K\Delta_n^{\epsilon}. \end{aligned} \tag{A.31}$$

Following a similar argument as driving (A.31), we can also show

$$\mathbb{P}(\exists 1 \leq j \leq m_n \text{ such that } \widetilde{Y}^o_{\lceil k_{n,j}\chi\rceil,j} \geq p_{n,j} + \varkappa_{1,n}) \leq K\Delta_n^{\epsilon}. \tag{A.32}$$

Combining (A.31) and (A.32), recall the definition of $\overline{I}_{n,j}$, we conclude that

$$\mathbb{P}(\widetilde{Y}^o_{\lceil k_{n,j}\chi\rceil,j} \in \overline{I}_{n,j} \text{ for all } 1 \leq j \leq m_n) \geq 1 - K\Delta_n^{\epsilon}. \tag{A.33}$$

Now, let $E_{n,2}$ be the event such that

$$E_{n,2} \equiv \Big\{\max_{1\leq j\leq m_n}\sqrt{k_{n,j}}\,\overline{S}_{n,j}(\overline{I}_{n,j}) \leq K\Delta_n^{\gamma}\Big\} \cap \{\widetilde{Y}^o_{\lceil k_{n,j}\chi\rceil,j} \in \overline{I}_{n,j} \text{ for all } 1 \leq j \leq m_n\}.$$

Then (A.26) and (A.33) imply $\mathbb{P}(E_{n,2}^{\complement}) \leq K'\Delta_n^{\epsilon}$. Recall that Assumption A.2(iii) implies $\partial_x f_{n,j}(x)$ is uniformly bounded over $x \in \bigcup_{j=1}^{m_n}\overline{I}_{n,j}$ for n sufficiently large. On the event $E_{n,2}$, by the second order Taylor expansion, we have

$$\begin{aligned}\max_{1\leq j\leq m_n}\max_{1\leq i\leq k_{n,j}}\sqrt{k_{n,j}}\Big|&F(\zeta_{\tau(i,j)}, \widetilde{Y}^o_{\lceil k_{n,j}\chi\rceil,j}) - F(\zeta_{\tau(i,j)}, q_{n,j})\\ &- (\widetilde{Y}^o_{\lceil k_{n,j}\chi\rceil,j} - q_{n,j})f_{\tau(i,j)}(q_{n,j})\Big| \leq K\Delta_n^{-\rho/2}\varkappa_{1,n}^2 \leq K\Delta_n^{\gamma}. \end{aligned} \tag{A.34}$$

It follows Assumption A.2(ii) and (A.29) that

$$\max_{1\leq j\leq m_n}\max_{1\leq i\leq k_{n,j}}\sqrt{k_{n,j}}\big(|F(\zeta_{\tau(i,j)}, q_{n,j}) - \chi| + |f_{\tau(i,j)}(q_{n,j}) - f_{n,j}(q_{n,j})|\big) = o_p(\Delta_n^{\epsilon+\gamma}). \tag{A.35}$$

Combining (A.34) and (A.35) yields

$$\begin{aligned}\mathbb{P}\Big(\Big\{\max_{1\le j\le m_n}\max_{1\le i\le k_{n,j}}\sqrt{k_{n,j}}\Big|F(\zeta_{\tau(i,j)},\widetilde{Y}^o_{\lceil k_{n,j}\chi\rceil,j})-F(\zeta_{\tau(i,j)},q_{n,j})\\ -(\widetilde{Y}^o_{\lceil k_{n,j}\chi\rceil,j}-q_{n,j})f_{n,j}(q_{n,j})\Big|\ge K\Delta_n^\gamma\Big\}\cap E_{n,2}\Big)\le K'\Delta_n^\epsilon.\end{aligned} \tag{A.36}$$

On the event $E_{n,2}$, by the definition of $\overline{S}_{n,j}$, we have

$$\begin{aligned}\max_{1\le j\le m_n}\sqrt{k_{n,j}}\bigg|\frac{\lceil k_{n,j}\chi\rceil}{k_{n,j}}-\widetilde{F}_{n,j}(q_{n,j})-\frac{1}{k_{n,j}}\sum_{i=1}^{k_{n,j}}\big(F(\zeta_{\tau(i,j)},\widetilde{Y}^o_{\lceil k_{n,j}\chi\rceil,j})\\ -F(\zeta_{\tau(i,j)},q_{n,j})\big)\bigg|\le\max_{1\le j\le m_n}\sqrt{k_{n,j}}\overline{S}_{n,j}(\overline{I}_{n,j})\le K\Delta_n^\gamma.\end{aligned} \tag{A.37}$$

By simple algebra we have $\sqrt{k_{n,j}}|\lceil k_{n,j}\chi\rceil/k_{n,j}-\chi|\le k_{n,j}^{-1/2}\le K\Delta_n^{\rho/2}\le K\Delta_n^\gamma$. Combing with (A.35)-(A.37), by the triangle inequality, we conclude that

$$\begin{aligned}&\mathbb{P}\bigg(\max_{1\le j\le m_n}\sqrt{k_{n,j}}\bigg|\widetilde{Y}^o_{\lceil k_{n,j}\chi\rceil,j}-q_{n,j}-\frac{1}{f_{n,j}(q_{n,j})}\bigg(\frac{1}{k_{n,j}}\sum_{i=1}^{k_{n,j}}F(\zeta_{\tau(i,j)},q_{n,j})-\widetilde{F}_{n,j}(q_{n,j})\bigg)\bigg|\ge K\Delta_n^\gamma\bigg)\\ &\le\mathbb{P}\bigg(\bigg\{\max_{1\le j\le m_n}\sqrt{k_{n,j}}\bigg|\widetilde{Y}^o_{\lceil k_{n,j}\chi\rceil,j}-q_{n,j}-\frac{1}{f_{n,j}(q_{n,j})}\\ &\qquad\times\bigg(\frac{1}{k_{n,j}}\sum_{i=1}^{k_{n,j}}F(\zeta_{\tau(i,j)},q_{n,j})-\widetilde{F}_{n,j}(q_{n,j})\bigg)\bigg|\ge K\Delta_n^\gamma\bigg\}\cap E_{n,2}\bigg)+\mathbb{P}(E_{n,2}^{\complement})\\ &\le K\Delta_n^\epsilon.\end{aligned} \tag{A.38}$$

Step 3. Combining (A.16) and (A.38), by the triangle inequality and the Markov inequality, we obtain

$$\begin{aligned}&\mathbb{P}\bigg(\max_{1\le j\le m_n}\sqrt{k_{n,j}}\bigg|\hat{q}_{n,j}-q_{n,j}-\frac{1}{f_{n,j}(q_{n,j})}\frac{1}{k_{n,j}}\sum_{i=1}^{k_{n,j}}\big(F(\zeta_{\tau(i,j)},q_{n,j})-\mathbb{1}\{\widetilde{Y}_{i,j}\le q_{n,j}\}\big)\bigg|\ge K\Delta_n^\gamma\bigg)\\ &\le\mathbb{P}\bigg(\max_{1\le j\le m_n}\sqrt{k_{n,j}}\bigg|\widetilde{Y}^o_{\lceil k_{n,j}\chi\rceil,j}-q_{n,j}-\frac{1}{f_{n,j}(q_{n,j})}\\ &\qquad\times\bigg(\frac{1}{k_{n,j}}\sum_{i=1}^{k_{n,j}}F(\zeta_{\tau(i,j)},q_{n,j})-\widetilde{F}_{n,j}(q_{n,j})\bigg)\bigg|\ge K\Delta_n^\gamma\bigg)\\ &\quad+\mathbb{P}\Big(\max_{1\le j\le m_n}\sqrt{k_{n,j}}|\hat{g}_{n,j}-\widetilde{Y}^o_{\lceil k_{n,j}\chi\rceil,j}|\ge K\Delta_n^\gamma\Big)\\ &\le K\Delta_n^\epsilon.\end{aligned} \tag{A.39}$$

Recall $|\sqrt{a}-1| \leq |a-1|$ for positive a, note that by (A.29) and Assumption A.2(ii), we have

$$\begin{aligned}
&\max_{1\leq j\leq m_n}\max_{1\leq i\leq k_{n,j}}\left|\sqrt{\frac{\chi(1-\chi)}{F(\zeta_{\tau(i,j)},q_{n,j})\big(1-F(\zeta_{\tau(i,j)},q_{n,j})\big)}}-1\right| \\
&\leq \max_{1\leq j\leq m_n}\max_{1\leq i\leq k_{n,j}}\left|\frac{F(\zeta_{\tau(1,j)},q_{n,j})\big(1-F(\zeta_{\tau(1,j)},q_{n,j})\big)}{F(\zeta_{\tau(i,j)},q_{n,j})\big(1-F(\zeta_{\tau(i,j)},q_{n,j})\big)}-1\right| \\
&\leq K\max_{1\leq j\leq m_n}\max_{1\leq i\leq k_{n,j}}\|\zeta_{\tau(i,j)}-\zeta_{\tau(1,j)}\| = o_p(\Delta_n^{\epsilon+\gamma}). \qquad \text{(A.40)}
\end{aligned}$$

Combining (A.39) and (A.40) completes the proof of Lemma A.1. *Q.E.D.*

Proof of Theorem 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 sightly stronger restriction on ϵ than in the proof of Lemma A.1, we prove the validity of the assertion for all positive ϵ satisfying

$$\epsilon < \frac{\rho}{6}\wedge\Big(\frac{1}{2}-\rho\Big)\wedge\Big(r-\frac{\rho}{2}\Big).$$

Correspondingly, let γ be a positive constant satisfying

$$\gamma < \Big(\frac{\rho}{4}-\epsilon\Big)\wedge\Big(\frac{1}{2}-\rho-\epsilon\Big)\wedge\Big(r-\frac{\rho}{2}-\epsilon\Big).$$

By the triangle inequality, we have

$$\begin{aligned}
\max_{1\leq j\leq m_n}\sup_{t\in\mathcal{T}_{n,j}}\sqrt{k_{n,j}}|\hat{q}_{n,t}-q_{n,t}| \leq\; &\max_{1\leq j\leq m_n}\sup_{t\in\mathcal{T}_{n,j}}\sqrt{k_{n,j}}|q_{n,j}-q_t| + \max_{1\leq j\leq m_n}\sqrt{k_{n,j}}|\hat{q}_{n,j}-\tilde{q}_{n,j}| \\
&+\max_{1\leq j\leq m_n}|\tilde{q}_{n,j}-q_{n,j}|. \qquad \text{(A.41)}
\end{aligned}$$

For the first term, by (A.21) and (A.29), we have

$$\max_{1\leq j\leq m_n}\sup_{t\in\mathcal{T}_{n,j}}\sqrt{k_{n,j}}|q_{n,j}-q_t| = o_p(\Delta_n^{\epsilon+\gamma}). \qquad \text{(A.42)}$$

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 $\widetilde{U}_{i,j}$ and $\nu_{i,j}$ as follows:

$$\begin{aligned}
\widetilde{U}_{i,j} &\equiv \sqrt{\frac{\bar{k}_n}{k_{n,j}}}\frac{\sqrt{\chi(1-\chi)}}{f_{n,j}(q_{n,j})}\frac{F(\zeta_{\tau(i,j)},q_{n,j})-\mathbb{1}\{\mathscr{Y}(\zeta_{\tau(i,j)},\varepsilon_{n,\iota(i,j)})\leq q_{n,j}\}}{\sqrt{F(\zeta_{\tau(i,j)},q_{n,j})\big(1-F(\zeta_{\tau(i,j)},q_{n,j})\big)}}\mathbb{1}\{1\leq i\leq k_{n,j}\}, \\
\tilde{\nu}_{i,j}^2 &\equiv \frac{\bar{k}_n}{k_{n,j}}\frac{\chi(1-\chi)}{f_{n,j}(q_{n,j})^2}\mathbb{1}\{1\leq i\leq k_{n,j}\}.
\end{aligned}$$

By construction the variables $\widetilde{U}_{i,j}$ 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 $\tilde{\nu}_{i,j}^2$. Note that

$$\sqrt{k_{n,j}}(\tilde{q}_{n,j} - q_{n,j}) = \frac{1}{\sqrt{\bar{k}_n}} \sum_{i=1}^{\bar{k}_n} \widetilde{U}_{i,j}, \quad \text{for } 1 \leq j \leq m_n.$$

Therefore, for each $1 \leq i \leq \bar{k}_n$ and $1 \leq j \leq 2m_n$, define $\widetilde{U}_{i,j}^{\dagger}$ as

$$\widetilde{U}_{i,j}^{\dagger} \equiv \widetilde{U}_{i,j} \mathbb{1}\{1 \leq j \leq m_n\} - \widetilde{U}_{i,j-m_n} \mathbb{1}\{m_n + 1 \leq j \leq 2m_n\}.$$

We can thus rewrite

$$\max_{1 \leq j \leq m_n} \sqrt{k_{n,j}} |\hat{q}_{n,j} - q_{n,j}| = \max_{1 \leq j \leq 2m_n} \frac{1}{\sqrt{\bar{k}_n}} \sum_{i=1}^{\bar{k}_n} \widetilde{U}_{i,j}^{\dagger}.$$

Let $(\widetilde{Z}_{i,j})_{1 \leq i \leq \bar{k}_n, 1 \leq j \leq m_n}$ be a sequence of centered mixed Gaussian variables with $\mathcal{F}^{(0)}$-conditional variance $\mathbb{E}[\widetilde{Z}_{i,j}^2 | \mathcal{F}^{(0)}] = \mathbb{E}[\widetilde{U}_{i,j}^2 | \mathcal{F}^{(0)}] = \tilde{\nu}_{i,j}^2$. Further, for each $1 \leq i \leq \bar{k}_n$ and $1 \leq j \leq 2m_n$, let

$$\widetilde{Z}_{i,j}^{\dagger} \equiv \widetilde{Z}_{i,j} \mathbb{1}\{1 \leq j \leq m_n\} - \widetilde{Z}_{i,j-m_n} \mathbb{1}\{m_n + 1 \leq i \leq 2m_n\},$$

which implies $\mathbb{E}[\widetilde{Z}_{i,j} \widetilde{Z}_{i',j} | \mathcal{F}^{(0)}] = \mathbb{E}[\widetilde{Z}_{i,j} \widetilde{Z}_{i',j} | \mathcal{F}^{(0)}]$ for all $1 \leq i, i' \leq \bar{k}_n$ and $1 \leq j \leq 2m_n$. Recall that the variables $\widetilde{U}_{i,j}$ are bounded, by Proposition 2.1 in Chernozhukov et al. (2017), we obtain for all $\epsilon < \rho/6$ that

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P}\left(\max_{1 \leq j \leq m_n} \sqrt{k_{n,j}} |\tilde{q}_{n,j} - q_{n,j}| \leq x \middle| \mathcal{F}^{(0)} \right) - \mathbb{P}\left(\max_{1 \leq j \leq 2m_n} \frac{1}{\sqrt{\bar{k}_n}} \sum_{i=1}^{\bar{k}_n} \widetilde{Z}_{i,j}^{\dagger} \leq x \middle| \mathcal{F}^{(0)} \right) \right| \leq K \Delta_n^{\epsilon}. \quad \text{(A.43)}$$

For $1 \leq j \leq m_n$, define $Z_j \equiv \bar{k}_n^{-1/2} \sum_{i=1}^{\bar{k}_n} \widetilde{Z}_{i,j}$. Recalling the definition of $\widetilde{Z}_{i,j}$ and $\tilde{\nu}_{i,j}$, we have $\mathbb{E}[Z_j^2 | \mathcal{F}^{(0)}] = \chi(1-\chi)/f_{n,j}(q_{n,j})^2 \equiv \nu_j^2$ for $1 \leq j \leq m_n$, hence

$$(Z_1, \ldots, Z_{m_n})^{\top} \sim \mathcal{MN}(0, \mathrm{diag}\{\nu_1^2, \ldots, \nu_{m_n}^2\}).$$

Also note that by construction we have

$$\max_{1 \leq j \leq 2m_n} \frac{1}{\sqrt{\bar{k}_n}} \sum_{i=1}^{\bar{k}_n} \widetilde{Z}_{i,j}^{\dagger} = \max_{1 \leq j \leq m_n} \left| \frac{1}{\sqrt{\bar{k}_n}} \sum_{i=1}^{\bar{k}_n} \widetilde{Z}_{i,j} \right| = \max_{1 \leq j \leq m_n} |Z_j|. \quad \text{(A.44)}$$

Therefore, it follows (A.41) and the triangle inequality that

$$
\begin{aligned}
&\sup_{x\in\mathbb{R}}\bigg(\mathbb{P}\Big(\max_{1\le j\le m_n}\sup_{t\in\mathcal{T}_{n,j}}\sqrt{k_{n,j}}|\hat q_{n,t}-q_t|\le x\Big)-\mathbb{P}\Big(\max_{1\le j\le m_n}|Z_j|\le x\Big)\bigg)\\
&\quad\le \mathbb{P}\Big(\max_{1\le j\le m_n}\sup_{t\in\mathcal{T}_{n,j}}\sqrt{k_{n,j}}|q_{n,j}-q_t|>K\Delta_n^{\gamma}\Big)+\mathbb{P}\Big(\max_{1\le j\le m_n}\sqrt{k_{n,j}}|\hat q_{n,j}-\tilde q_{n,j}|>K\Delta_n^{\gamma}\Big)\\
&\qquad+\sup_{x\in\mathbb{R}}\bigg(\mathbb{P}\Big(\max_{1\le j\le m_n}\sqrt{k_{n,j}}|\tilde q_{n,j}-q_{n,j}|\le x+2K\Delta_n^{\gamma}\Big)-\mathbb{P}\Big(\max_{1\le j\le m_n}|Z_j|\le x+2K\Delta_n^{\gamma}\Big)\bigg)\\
&\qquad+\sup_{x\in\mathbb{R}}\mathbb{P}\Big(x<\max_{1\le j\le m_n}|Z_j|\le x+2K\Delta_n^{\gamma}\Big)\\
&\quad\le K\Delta_n^{\epsilon},
\end{aligned}
$$

where the first term is bounded by $K\Delta_n^{\epsilon}$ using (A.42) and the Markov inequality, the second term uses Lemma A.1, the third term is bounded by $K\Delta_n^{\epsilon}$ using (A.43), (A.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

$$
\mathbb{E}\Big[\max_{1\le j\le m_n}|Z_j|\Big]\le K\sqrt{L_n}.
$$

Similarly, we can show

$$
\sup_{x\in\mathbb{R}}\bigg(\mathbb{P}\Big(\max_{1\le j\le m_n}|Z_j|\le x\Big)-\mathbb{P}\Big(\max_{1\le j\le m_n}\sup_{t\in\mathcal{T}_{n,j}}\sqrt{k_{n,j}}|\hat q_{n,t}-q_t|\le x\Big)\bigg)\le K\Delta_n^{\epsilon}.
$$

This competes the proof of required statement. *Q.E.D.*

A.3 **Proofs for Section 2.5**

Proof of Theorem 3. As mentioned in the main text, we prove a stronger result that the statement in Theorem 3 holds for all $\mathcal{S}_n\subset\mathcal{S}_n^{\mathrm{all}}$ with $|\mathcal{S}_n|\ge 3$. Let $\mathcal{G}_n\equiv\mathcal{F}^{(0)}\vee\sigma(Y_{i\Delta_n}:1\le i\le n)$ denote the smallest σ-algebra contains $\mathcal{F}^{(0)}\cup\sigma(Y_{i\Delta_n}:1\le i\le n)$. Also, we strengthen Assumption 1 to Assumption A.1 by a using of Localization procedure. We prove assertions of the theorem for positive ϵ satisfying

$$
\epsilon<\frac{\rho}{7}\wedge\Big(\frac{1}{6}-\frac{\rho}{3}\Big)\wedge\Big(\frac{r}{3}-\frac{\rho}{6}\Big).
$$

To facilitate our analysis, we adopt the notations from the proof of Theorem 1, and introduce some additional notations. For $1 \le i \le k_n$ and $(j, j') \in \mathcal{S}_n$, denote

$$\begin{aligned}
V_{n,i}(j,j') &\equiv Y_{\tau(i,j)} - Y_{\tau(i,j')}, \\
\widetilde{V}_{n,i}(j,j') &\equiv \mathscr{Y}(\zeta_{\tau(i,j)}, \varepsilon_{n,\iota(i,j)}) - \mathscr{Y}(\zeta_{\tau(i,j')}, \varepsilon_{n,\iota(i,j')}), \\
\mu_{n,i}(j,j') &\equiv g_{\tau(i,j)} - g_{\tau(i,j')}, \\
\bar{\mu}_n(j,j') &\equiv g_{n,j} - g_{n,j'}, \\
\varsigma_n(j,j')^2 &\equiv \sigma_{n,j}^2 + \sigma_{n,j'}^2.
\end{aligned}$$

Using above notations, we further define

$$\begin{aligned}
\overline{D}_n &\equiv \max_{(j,j')\in\mathcal{S}_n} \frac{1}{\sqrt{k_n}} \sum_{i=1}^{k_n} \frac{V_{n,i}(j,j') - \mu_{n,i}(j,j')}{\hat{\varsigma}_n(j,j')}, \\
\widetilde{D}_n &\equiv \max_{(j,j')\in\mathcal{S}_n} \frac{1}{\sqrt{k_n}} \sum_{i=1}^{k_n} \frac{\widetilde{V}_{n,i}(j,j') - \bar{\mu}_n(j,j')}{\varsigma_n(j,j')}, \\
\widehat{D}_n^B &\equiv \max_{(j,j')\in\mathcal{S}_n} \frac{1}{\sqrt{k_n}} \sum_{i=1}^{k_n} \frac{e_i\big(V_{n,i}(j,j') - (\hat{g}_{n,j} - \hat{g}_{n,j'})\big)}{\hat{\varsigma}_n(j,j')} = \max_{(j,j')\in\mathcal{S}_n} \frac{\sqrt{k_n}(\hat{g}_{n,j}^B - \hat{g}_{n,j'}^B)}{\hat{\varsigma}_n(j,j')}, \\
\widetilde{D}_n^B &\equiv \max_{(j,j')\in\mathcal{S}_n} \frac{1}{\sqrt{k_n}} \sum_{i=1}^{k_n} \frac{e_i\big(\widetilde{V}_{n,i}(j,j') - (\hat{g}_{n,j} - \hat{g}_{n,j'})\big)}{\varsigma_n(j,j')}.
\end{aligned}$$

First, we compute the approximation bounds of these variables and their conditional quantiles. Our analysis relies on the following decomposition of $|\overline{D}_n - \widetilde{D}_n|$,

$$|\overline{D}_n - \widetilde{D}_n| \le \max_{(j,j')\in\mathcal{S}_n} \left| \frac{\varsigma_n(j,j')}{\hat{\varsigma}_n(j,j')} - 1 \right| \times \left(\max_{(j,j')\in\mathcal{S}_n} \left| \frac{1}{\sqrt{k_n}} \sum_{i=1}^{k_n} \frac{\widetilde{V}_{n,i}(j,j') - \mu_{n,i}(j,j')}{\varsigma_n(j,j')} \right| \right) + \max_{(j,j'\in\mathcal{S}_n)} |\mathfrak{C}_n(j,j')|,$$

where for $(j, j') \in \mathcal{S}_n$, $\mathfrak{C}_n(j,j') \equiv \mathfrak{C}_n^{(I)}(j,j') + \mathfrak{C}_n^{(II)}(j,j')$ with

$$\begin{aligned}
\mathfrak{C}_n^{(I)}(j,j') &\equiv \frac{1}{\sqrt{k_n}} \sum_{i=1}^{k_n} \frac{R_{n,\iota(i,j)} - R_{n,\iota(i,j')}}{\varsigma_n(j,j')}, \\
\mathfrak{C}_n^{(II)}(j,j') &\equiv \frac{1}{\sqrt{k_n}} \sum_{i=1}^{k_n} \frac{\mu_{n,i}(j,j') - \bar{\mu}_n(j,j')}{\varsigma_n(j,j')}.
\end{aligned}$$

By the triangle inequality and (A.6), for $p > (1-\rho)/(1/2-\rho-\epsilon-\gamma)$, we have

$$\left\| \max_{(j,j')\in\mathcal{S}_n} \frac{1}{\sqrt{k_n}} \sum_{i=1}^{k_n} \big(\mu_{n,i}(j,j') - \bar{\mu}_n(j,j')\big) \right\|_{L_p} \le K_p m_n^{1/p} \Delta_n^{1/2-\rho} = o(\Delta_n^{\epsilon+\gamma}). \tag{A.45}$$

Then combing (A.5) and (A.45), it follows the triangle inequality again that

$$\max_{(j,j')\in\mathcal{S}_n} |\mathfrak{C}_n(j,j')| \le \max_{(j,j')\in\mathcal{S}_n} |\mathfrak{C}_n^{(I)}(j,j')| + \max_{(j,j')\in\mathcal{S}_n} |\mathfrak{C}_n^{(II)}(j,j')| = o_p(\Delta_n^{\epsilon+\gamma}). \tag{A.46}$$

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 (A.14), we obtain that for $\epsilon/\rho \leq \varpi < 1 - 2\gamma/\rho$,

$$\mathbb{P}\left(\max_{(j,j')\in\mathcal{S}_n} \left|\frac{\varsigma_n(j,j')}{\hat{\varsigma}_n(j,j')} - 1\right| > K\Delta_n^{\rho(1-\varpi)/2} L_n^{2\eta} \log(|\mathcal{S}_n|)\right) \leq K'\Delta_n^{\epsilon}. \tag{A.47}$$

Combining (A.46) and (A.47), following the similar procedure as deriving (A.2), we can show that

$$\mathbb{P}(|\overline{D}_n - \widetilde{D}_n| > K\varrho_n) \leq K'\Delta_n^{\epsilon}, \tag{A.48}$$

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 $|\widehat{D}_n^B - \widetilde{D}_n^B|$ as

$$|\widehat{D}_n^B - \widetilde{D}_n^B| \leq \max_{(j,j')\in\mathcal{S}_n} \left|\frac{\varsigma_n(j,j')}{\hat{\varsigma}_n(j,j')} - 1\right| \times \left(\max_{(j,j')\in\mathcal{S}_n} \frac{1}{\sqrt{k_n}} \sum_{i=1}^{k_n} \frac{e_i\big(\widetilde{V}_{n,i}(j,j') - (\hat{g}_{n,j} - \hat{g}_{n,j'})\big)}{\varsigma_n(j,j')}\right) + \max_{(j,j')\in\mathcal{S}_n} |\mathfrak{D}_n(j,j')|, \tag{A.49}$$

where for $(j,j') \in \mathcal{S}_n$, $\mathfrak{D}_n(j,j') \equiv k_n^{-1/2} \sum_{i=1}^{k_n} e_i(R_{n,\iota(i,j)} - R_{n,\tau(i,j')})/\varsigma_n(j,j')$. 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 Cauchy–Schwartz inequality and combining with (A.46), we have

$$\max_{(j,j')\in\mathcal{S}_n} |\mathfrak{D}_n(j,j')| \leq \sqrt{\max_{1\leq i\leq k_n} |e_i|^2 \times \max_{(j,j')\in\mathcal{S}_n} |\mathfrak{C}_n^{(I)}(j,j')|^2} = o_p(\Delta_n^{\epsilon+\gamma}\sqrt{L_n}). \tag{A.50}$$

Let $E_{n,3}$ be the event such that

$$E_{n,3} \equiv \left\{\max_{(j,j')\in\mathcal{S}_n} \left|\frac{\varsigma_n(j,j')}{\hat{\varsigma}_n(j,j')} - 1\right| \leq \Delta_n^{\rho(1-\varpi)/2} L_n^{2\eta} \log(|\mathcal{S}_n|)\right\} \cap \left\{\max_{(j,j')\in\mathcal{S}_n} |\mathfrak{D}_n(j,j')| \leq \Delta_n^{\gamma/2}\right\},$$

by (A.47), (A.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 $\big(k_n^{-1/2} \sum_{i=1}^{k_n} e_i(\widetilde{V}_{n,i}(j,j') - (\hat{g}_{n,j} - \hat{g}_{n,j'}))/\varsigma_n(j,j')\big)_{(j,j')\in\mathcal{S}_n}$ follow a Gaussian distribution with bounded variance, which implies $\mathbb{E}[\widetilde{D}_n^B|\mathcal{G}_n] \leq K\sqrt{\log(|\mathcal{S}_n|)}$. Therefore, it follows the Markov inequality and (A.49) that

$$\begin{aligned}
&\mathbb{P}(\{|\widehat{D}_n^B - \widetilde{D}_n^B| > \varrho_n\} \cap E_{n,3}|\mathcal{G}_n) \\
&\quad \leq \varrho_n^{-1}\left(\max_{(j,j')\in\mathcal{S}_n} \left|\frac{\varsigma_n(j,j')}{\hat{\varsigma}_n(j,j')} - 1\right| \times \left(\mathbb{E}[\widetilde{D}_n^B|\mathcal{G}_n] + 2\max_{(j,j')\in\mathcal{S}_n} |\mathfrak{D}_n(j,j')|\right)\right) \\
&\quad \leq \frac{\Delta_n^{\rho(1-\varpi)/2} L_n^{2\eta} \log(|\mathcal{S}_n|)\big(K\sqrt{\log(|\mathcal{S}_n|)} + \Delta_n^{\gamma/2}\big)}{K'\Delta_n^{\rho(1-\varpi)/4} L_n^{2\eta} \log(|\mathcal{S}_n|)^{3/2}} \leq K\Delta_n^{\rho(1-\varpi)/4}.
\end{aligned}$$

With K denoting the same constant as in the above display, by the law of iterated expectation, we can conclude that

$$\mathbb{P}\big(\mathbb{P}(|\widehat{D}_n^B - \widetilde{D}_n^B| > \varrho_n|\mathcal{G}_n) > K\Delta_n^{\rho(1-\varpi)/4}\big) \leq \mathbb{P}(E_{n,3}^{\complement}) \leq K'\Delta_n^{\epsilon}. \tag{A.51}$$

Let $\widetilde{X}_n(j,j')$ be centered mixed Gaussian variables indexed by (j,j') with $\mathcal{F}^{(0)}$-conditional covariance matrix such that for all $(j,j'),(\ell,\ell')\in\mathcal{S}_n$,

$$\begin{aligned}&\mathbb{E}[\widetilde{X}_n(j,j')\widetilde{X}_n(\ell,\ell')|\mathcal{F}^{(0)}]\\&\quad=\mathbb{E}\Big[\Big(\frac{1}{\sqrt{k_n}}\sum_{i=1}^{k_n}\frac{\widetilde{V}_{n,i}(j,j')-\mu_{n,i}(j,j')}{\varsigma_n(j,j')}\Big)\Big(\frac{1}{\sqrt{k_n}}\sum_{i=1}^{k_n}\frac{\widetilde{V}_{n,i}(\ell,\ell')-\mu_{n,i}(\ell,\ell')}{\varsigma_n(\ell,\ell')}\Big)\Big|\mathcal{F}^{(0)}\Big].\end{aligned}$$

Then by Proposition 2.1 in Chernozhukov et al. (2017), we have for all $\epsilon<\rho/6$,

$$\begin{aligned}&\sup_{x\in\mathbb{R}}\Big|\mathbb{P}\Big(\max_{(j,j')\in\mathcal{S}_n}\frac{1}{\sqrt{k_n}}\sum_{i=1}^{k_n}\frac{\widetilde{V}_{n,i}(j,j')-\mu_{n,i}(j,j')}{\varsigma_n(j,j')}\le x\Big|\mathcal{F}^{(0)}\Big)-\mathbb{P}\Big(\max_{(j,j')\in\mathcal{S}_n}\widetilde{X}_n(j,j')\le x\Big|\mathcal{F}^{(0)}\Big)\Big|\\&\quad\le K\big(\Delta_n^{\rho/6}L_n^{\eta/3}(L_n+\log(|\mathcal{S}_n|))^{7/6}+\Delta_n^{\rho/6}L_n^{2\eta/3}(L_n+\log(|\mathcal{S}_n|))\big)\le K\Delta_n^{\epsilon}.\end{aligned}\tag{A.52}$$

By Corollary 4.2 in Chernozhukov et al. (2017), for all $\epsilon<\rho/7$, with probability at least $1-K\Delta_n^{\epsilon}$,

$$\begin{aligned}&\sup_{x\in\mathbb{R}}\Big|\mathbb{P}(\widetilde{D}_n^B\le x|\mathcal{G}_n)-\mathbb{P}\Big(\max_{(j,j')\in\mathcal{S}_n}\widetilde{X}_n(j,j')\le x\Big|\mathcal{F}^{(0)}\Big)\Big|\\&\quad\le K'\big(\Delta_n^{\rho/6}L_n^{(1+\eta)/3}(L_n+\log(|\mathcal{S}_n|))^{5/6}+\Delta_n^{(\rho-\epsilon)/6}L_n^{2\eta/3}(L_n+\log(|\mathcal{S}_n|))\big)\le K'\Delta_n^{\epsilon}.\end{aligned}\tag{A.53}$$

Let $\widetilde{cv}_n(\cdot,\mathcal{S}_n)$ denote the $\mathcal{F}^{(0)}$-conditional $1-(\cdot)$ quantile of $\max_{(j,j')\in\mathcal{S}_n}\widetilde{X}_n(j,j')$, i.e.,

$$\widetilde{cv}_n(\cdot,\mathcal{S}_n)\equiv\inf\Big\{C\in\mathbb{R}:\mathbb{P}\Big(\max_{(j,j')\in\mathcal{S}_n}\widetilde{X}_n(j,j')\le C\Big|\mathcal{F}^{(0)}\Big)\ge 1-(\cdot)\Big\}.$$

Note that $\mathbb{E}[\max_{(j,j')\in\mathcal{S}_n}\widetilde{X}_n(j,j')|\mathcal{F}^{(0)}]\le K\sqrt{\log(|\mathcal{S}_n|)}$. Also note that Assumption A.1(i) implies the bounds obtained in the previous equation and in the approximation (A.52), (A.53) are universal. Therefore, we can fix a positive universal constant M satisfying the previous equation. Therefore, for $\alpha\in\big(0,1-M\varrho_n\sqrt{\log(|\mathcal{S}_n|)}\big)$, by the anti-concentration inequality, we have

$$\mathbb{P}\Big(\max_{(j,j')\in\mathcal{S}_n}\widetilde{X}_n(j,j')\le\widetilde{cv}_n\big(\alpha+M\varrho_n\sqrt{\log(|\mathcal{S}_n|)},\mathcal{S}_n\big)+\varrho_n\Big|\mathcal{F}^{(0)}\Big)\le 1-\alpha.\tag{A.54}$$

Let $E_{n,4}$ be the event such that

$$\begin{aligned}E_{n,4}\ \equiv\ &\{\mathbb{P}(|\widehat{D}_n^B-\widetilde{D}_n^B|>\varrho_n|\mathcal{G}_n)\le M\Delta_n^{\rho(1-\varpi)/4}\}\\&\cap\Big\{\sup_{x\in\mathbb{R}}\Big|\mathbb{P}(\widetilde{D}_n^B\le x|\mathcal{G}_n)-\mathbb{P}\Big(\max_{(j,j')\in\mathcal{S}_n}\widetilde{X}_n(j,j')\le x\Big|\mathcal{F}^{(0)}\Big)\Big|\le M\Delta_n^{\epsilon}\Big\},\end{aligned}$$

by (A.51) and (A.53), we have shown $\mathbb{P}(E_{n,4})\ge 1-K'\Delta_n^{\epsilon}$. Therefore, we have

$$\begin{aligned}&\mathbb{P}\big(\{\widehat{D}_n^B\le\widetilde{cv}_n(\alpha+M(\Delta_n^{\epsilon}+\varrho_n\sqrt{\log(|\mathcal{S}_n|)}),\mathcal{S}_n)\}\cap E_{n,4}\big|\mathcal{G}_n\big)\\&\quad\le\mathbb{P}\big(\{\widetilde{D}_n^B\le\widetilde{cv}_n(\alpha+M(\Delta_n^{\epsilon}+\varrho_n\sqrt{\log(|\mathcal{S}_n|)}),\mathcal{S}_n)+\varrho_n\}\cap E_{n,4}\big|\mathcal{G}_n\big)+M\Delta_n^{\rho(1-\varpi)/4}\\&\quad\le\mathbb{P}\Big(\max_{(j,j')\in\mathcal{S}_n}\widetilde{X}_n(j,j')\le\widetilde{cv}_n\big(\alpha+M(\Delta_n^{\epsilon}+\varrho_n\sqrt{\log(|\mathcal{S}_n|)}),\mathcal{S}_n\big)+\varrho_n\Big|\mathcal{F}^{(0)}\Big)+M\Delta_n^{\epsilon}\\&\quad\le 1-\alpha-M\Delta_n^{\epsilon}+M\Delta_n^{\epsilon}=1-\alpha,\end{aligned}$$

where the third line uses the fact that $\rho(1-\varpi)/4 > \epsilon$, and the fourth line is by (A.54). By the law of iterated expectation and the definition of $cv_n^B(\alpha, \mathcal{S}_n)$, we can conclude that

$$\mathbb{P}\big(cv_n^B(\alpha, \mathcal{S}_n) < \widetilde{cv}_n(\alpha + M(\Delta_n^\epsilon + \varrho\sqrt{\log(|\mathcal{S}_n|)}), \mathcal{S}_n)\big) \le \mathbb{P}(E_{n,4}^{\complement}) \le K'\Delta_n^\epsilon. \tag{A.55}$$

By the anti-concentration inequality, for $\alpha \in \big(M\varrho_n\sqrt{\log(|\mathcal{S}_n|)}, 1\big)$, we have $\mathbb{P}\big(\max_{(j,j')\in\mathcal{S}_n} \widetilde{X}_n(j,j') \le \widetilde{cv}_n(\alpha - M\varrho_n\sqrt{\log(|\mathcal{S}_n|)}, \mathcal{S}_n) - \varrho_n \big| \mathcal{F}^{(0)}\big) \ge 1-\alpha$. Similarly, we can show

$$\mathbb{P}\big(cv_n^B(\alpha, \mathcal{S}_n) > \widetilde{cv}_n(\alpha - M(\Delta_n^\epsilon + \varrho_n\sqrt{\log(|\mathcal{S}_n|)}), \mathcal{S}_n)\big) \le \mathbb{P}(E_{n,4}^{\complement}) \le K'\Delta_n^\epsilon. \tag{A.56}$$

We are now ready to prove the asserted statements in the theorem, starting from assertion (i). Assume that $\max_{(j,j')\in\mathcal{S}_n}(g_{n,j} - g_{n,j'}) \le 0$, this implies $\bar{\mu}_n(j,j') \le 0$ for all $(j,j') \in \mathcal{S}_n$. Combing with (A.45) yields

$$\mathbb{P}\bigg(\max_{(j,j')\in\mathcal{S}_n} \frac{1}{\sqrt{k_n}} \sum_{i=1}^{k_n} \mu_{n,i}(j,j') > K\varrho_n\bigg) \le K'\Delta_n^\epsilon.$$

Therefore, by (A.47) and the Markov inequality, this gives $\mathbb{P}(\widehat{D}_n - \overline{D}_n > \varrho_n/2) \le K\Delta_n^\epsilon$. Hence

$$\begin{aligned}
\mathbb{P}\big(\widehat{D}_n > cv_n^B(\alpha, \mathcal{S}_n)\big) &\le \mathbb{P}\big(\overline{D}_n > cv_n^B(\alpha, \mathcal{S}_n) - \varrho_n/2\big) + \mathbb{P}(\widehat{D}_n - \overline{D}_n > \varrho_n/2) \\
&\le \mathbb{P}\big(\widetilde{D}_n > cv_n^B(\alpha, \mathcal{S}_n) - \varrho_n/2 - \varrho_n/2\big) + K\Delta_n^\epsilon \\
&\le \mathbb{P}\big(\widetilde{D}_n > \widetilde{cv}_n(\alpha + M(\Delta_n^\epsilon + \varrho_n\sqrt{\log(|\mathcal{S}_n|)}), \mathcal{S}_n) - \varrho_n\big) + K\Delta_n^\epsilon,
\end{aligned} \tag{A.57}$$

where the second line is by (A.48), and the last line is by (A.55). For the first term, we have

$$\begin{aligned}
&\mathbb{P}\big(\widetilde{D}_n > \widetilde{cv}_n(\alpha + M(\Delta_n^\epsilon + \varrho_n\sqrt{\log(|\mathcal{S}_n|)}), \mathcal{S}_n) - \varrho_n \big| \mathcal{F}^{(0)}\big) \\
&\le \mathbb{P}\Big(\max_{(j,j')\in\mathcal{S}_n} \widetilde{X}_n(j,j') > \widetilde{cv}_n(\alpha + M(\Delta_n^\epsilon + \varrho_n\sqrt{\log(|\mathcal{S}_n|)}), \mathcal{S}_n) - \varrho_n \Big| \mathcal{F}^{(0)}\Big) + K\Delta_n^\epsilon \\
&\le \mathbb{P}\Big(\max_{(j,j')\in\mathcal{S}_n} \widetilde{X}_n(j,j') > \widetilde{cv}_n(\alpha + 2M(\Delta_n^\epsilon + \varrho_n\sqrt{\log(|\mathcal{S}_n|)}), \mathcal{S}_n) \Big| \mathcal{F}^{(0)}\Big) + K\Delta_n^\epsilon \\
&\le \alpha + 2M(\Delta_n^\epsilon + \varrho_n\sqrt{\log(|\mathcal{S}_n|)}) + K\Delta_n^\epsilon \le \alpha + K\Delta_n^\epsilon,
\end{aligned} \tag{A.58}$$

where the second line is by (A.45), (A.52), and the law of iterated expectation, the third line is by (A.54), the last line is by the definition of $\widetilde{cv}_n(\cdot, \mathcal{S}_n)$ and the fact that $\mathcal{S}_n \subset \{1, \ldots, m_n\} \times \{1, \ldots, m_n\}$ hence

$$\varrho_n\sqrt{\log(|\mathcal{S}_n|)} \le K\varrho_n\sqrt{L_n} \le K'\Delta_n^\epsilon.$$

Combing (A.57), (A.58), and applying the law of iterated expectation again, we can conclude that

$$\mathbb{P}\big(\widehat{D}_n > cv_n^B(\alpha, \mathcal{S}_n)\big) \le \alpha + K\Delta_n^\epsilon, \quad \text{ if } \max_{(j,j')\in\mathcal{S}_n}(g_{n,j} - g_{n,j'}) \le 0, \tag{A.59}$$

which is the first part of assertion (i). For the second part, assume $\bar{\mu}_n(j,j')=0$, then (A.45) yields

$$\mathbb{P}\Big(\max_{(j,j')\in\mathcal{S}_n}\Big|\frac{1}{\sqrt{k_n}}\sum_{i=1}^{k_n}\mu_{n,i}(j,j')\Big|>K\varrho_n\Big)\leq K'\Delta_n^{\epsilon}.$$

Therefore, by (A.47) and the Markov inequality, this gives $\mathbb{P}(\overline{D}_n-\widehat{D}_n>\varrho_n/2)\leq K\Delta_n^{\epsilon}$. Hence

$$\begin{aligned}\mathbb{P}\big(\widehat{D}_n>cv_n^B(\alpha,\mathcal{S}_n)\big) &\geq \mathbb{P}\big(\overline{D}_n>cv_n^B(\alpha,\mathcal{S}_n)+\varrho_n/2\big)-\mathbb{P}(\overline{D}_n-\widehat{D}_n>\varrho_n/2) \\ &\geq \mathbb{P}\big(\widetilde{D}_n>\widetilde{cv}_n(\alpha-M(\Delta_n^{\epsilon}+\varrho_n\sqrt{\log(|\mathcal{S}_n|)}),\mathcal{S}_n)+\varrho_n\big)-K\Delta_n^{\epsilon},\end{aligned}\tag{A.60}$$

where the second line is by (A.48) and (A.56). For the first term, we have

$$\begin{aligned}&\mathbb{P}\big(\widetilde{D}_n>\widetilde{cv}_n(\alpha-M(\Delta_n^{\epsilon}+\varrho_n\sqrt{\log(|\mathcal{S}_n|)}),\mathcal{S}_n)+\varrho_n\Big|\mathcal{F}^{(0)}\big)\\ &\geq \mathbb{P}\Big(\max_{(j,j')\in\mathcal{S}_n}\widetilde{X}_n(j,j')>\widetilde{cv}_n(\alpha-2M(\Delta_n^{\epsilon}+\varrho_n\sqrt{\log(|\mathcal{S}_n|)}),\mathcal{S}_n)\Big|\mathcal{F}^{(0)}\Big)-K\Delta_n^{\epsilon}\\ &\geq \alpha-2M(\Delta_n^{\epsilon}+\varrho_n\sqrt{\log(|\mathcal{S}_n|)})-K\Delta_n^{\epsilon}\leq\alpha-K\Delta_n^{\epsilon},\end{aligned}\tag{A.61}$$

where the second line is by (A.45), (A.52), and (A.54). Combing (A.59)-(A.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 (A.45) and (A.47) gives $\mathbb{P}(\overline{D}_n-\widehat{D}_n+\Delta_n^{-\rho/2}\Upsilon>\varrho_n/2)\leq K\Delta_n^{\epsilon}$. Therefore, we have

$$\begin{aligned}&\mathbb{P}\big(\widehat{D}_n>cv_n^B(\alpha,\mathcal{S}_n)\big)\\ &\geq \mathbb{P}\big(\overline{D}_n+\Delta_n^{-\rho/2}\Upsilon>cv_n^B(\alpha,\mathcal{S}_n)+\varrho_n/2\big)-\mathbb{P}(\overline{D}_n-\widehat{D}_n>\varrho_n/2)\\ &\geq \mathbb{P}\Big(\max_{(j,j')\in\mathcal{S}_n}\widetilde{X}_n(j,j')+\Delta_n^{-\rho/2}\Upsilon>\widetilde{cv}_n\big(\alpha-2M(\Delta_n^{\epsilon}+\varrho_n\sqrt{\log(|\mathcal{S}_n|)}),\mathcal{S}_n\big)\Big)-K\Delta_n^{\epsilon}\\ &\geq \mathbb{P}\Big(\max_{(j,j')\in\mathcal{S}_n}\widetilde{X}_n(j,j')+\Delta_n^{-\rho/2}\Upsilon>K\big(\sqrt{\log(|\mathcal{S}_n|)}+\sqrt{L_n}\big)\Big)-K\Delta_n^{\epsilon}\\ &\geq 1-K\Delta_n^{\rho/2}-K'\Delta_n^{\epsilon}\geq 1-K\Delta_n^{\epsilon},\end{aligned}$$

where the second line is by (A.45), (A.48), (A.52), (A.54), and (A.56). The third line is by the Borell's concentration inequality (see, e.g., Proposition A.2.1 in van der Vaart and Wellner (1996)), which gives $\mathbb{P}\big(|\max_{(j,j')\in\mathcal{S}_n}\widetilde{X}_n(j,j')-M\sqrt{\log(|\mathcal{S}_n|)}|\geq\lambda\big)\leq K\exp\{-\lambda^2/2K'\}$, setting the right hand side equaling to $\alpha-2M\big(\Delta_n^{\epsilon}+\varrho_n\sqrt{\log(|\mathcal{S}_n|)}\big)$ yields

$$\begin{aligned}&\widetilde{cv}_n\big(\alpha-2M(\Delta_n^{\epsilon}+\varrho_n\sqrt{\log(|\mathcal{S}_n|)}),\mathcal{S}_n\big)\\ &\leq M\sqrt{\log(|\mathcal{S}_n|)}+K\sqrt{2\log\Big(\frac{1}{\alpha-2M\big(\Delta_n^{\epsilon}+\varrho_n\sqrt{\log(|\mathcal{S}_n|)}\big)}\Big)}\\ &\leq K\big(\sqrt{\log(|\mathcal{S}_n|)}+\sqrt{L_n}\big).\end{aligned}$$

This competes the proof of required statement. *Q.E.D.*

Proof of Corollary 1. The corollary is a direct consequence of Theorem 3.3 in Mogstad et al. (2023) and Theorem 3. *Q.E.D.*

A.4 **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 $\max_{1\le j\le m_n}\sup_{t\in\mathcal{T}_{n,j}}\sqrt{k_{n,j}}|\hat{g}_{n,t}-g_t|$ for stationary β-mixing disturbance. For any sub σ-fields $\mathcal{A},\mathcal{B}$ of $\mathcal{F}$, denote

$$\beta(\mathcal{A},\mathcal{B})\equiv\frac{1}{2}\sup\Big\{\sum_{i=1}^{I}\sum_{j=1}^{J}|\mathbb{P}(A_i\cap B_j)-\mathbb{P}(A_i)\mathbb{P}(B_j)|\Big\},$$

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 kth β-mixing coefficient of $(\varepsilon_{n,i})_{1\le i\le n}$ as $\beta(k)\equiv\max_{1\le\ell\le n-k}\beta(\mathcal{H}_1^{\ell},\mathcal{H}_{\ell+k}^{n})$ where $\mathcal{H}_{i,j}\equiv\sigma(\varepsilon_{n,i},\dots,\varepsilon_{n,j})$ for $1\le i\le j\le n$. Moreover, for each $1\le q\le n$, define

$$\begin{aligned}\bar{\sigma}^2(q) &\equiv \sup_{\zeta\in\mathcal{Z}}\frac{1}{q}\sum_{i=1}^{q}\sum_{j=1}^{q}Cov\big(\mathscr{Y}(z,\varepsilon_{n,i}),\mathscr{Y}(z,\varepsilon_{n,j})\big),\\ \underline{\sigma}^2(q) &\equiv \inf_{\zeta\in\mathcal{Z}}\frac{1}{q}\sum_{i=1}^{q}\sum_{j=1}^{q}Cov\big(\mathscr{Y}(z,\varepsilon_{n,i}),\mathscr{Y}(z,\varepsilon_{n,j})\big).\end{aligned}$$

We follow the notations used in the proof of Theorem 1. Note that the derivation of (A.5) and (A.6) does not depend on conditional independence of $(\varepsilon_{n,i})_{1\le i\le n}$, hence we have

$$\mathbb{P}\left(\left|\max_{1\le j\le m_n}\sup_{t\in\mathcal{T}_{n,j}}\sqrt{k_{n,j}}|\hat{g}_{n,t}-g_t|-\max_{1\le j\le m_n}\left|\frac{1}{\sqrt{k_{n,j}}}\sum_{i=1}^{k_{n,j}}\widetilde{Y}_{i,j}\right|\right|>K\Delta_n^{\gamma}\right)\le K'\Delta_n^{\epsilon}.$$

For $1\le i\le\bar{k}_n$ and $1\le j\le 2m_n$, denote

$$\widetilde{Y}_{i,j}^{\dagger}\equiv\sqrt{\frac{\bar{k}_n}{k_{n,j}}}(\widetilde{Y}_{i,j}\mathbb{1}\{1\le i\le k_{n,j},1\le j\le m_n\}-\widetilde{Y}_{i,j-m_n}\mathbb{1}\{1\le i\le k_{n,j},m_n+1\le j\le 2m_n\}).$$

Then we can rewrite

$$\max_{1\le j\le m_n}\left|\frac{1}{\sqrt{k_{n,j}}}\sum_{i=1}^{k_{n,j}}\widetilde{Y}_{i,j}\right|=\max_{1\le j\le 2m_n}\frac{1}{\sqrt{\bar{k}_n}}\sum_{i=1}^{\bar{k}_n}\widetilde{Y}_{i,j}^{\dagger}.$$

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 A.3. *There exists a positive constant K such that (i) $\beta(n) \leq Kn^{-\upsilon}$ for some positive υ; (ii) $1/K \leq \underline{\sigma}^2(q) \leq \bar{\sigma}^2(q) \leq K$ 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 $\rho/(2+\upsilon) < \kappa < \rho/2$ 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

$$\widetilde{S}_{\ell,j} \equiv \sum_{i=(\ell-1)(q_{1,n}+q_{2,n})+1}^{(\ell-1)(q_{1,n}+q_{2,n})+q_{1,n}} \widetilde{Y}_{i,j}^{\dagger}, \quad \text{and} \quad \mathring{S}_{\ell,j} \equiv \sum_{i=(\ell-1)(q_{1,n}+q_{2,n})+q_{1,n}+1}^{\ell(q_{1,n}+q_{2,n})} \widetilde{Y}_{i,j}^{\dagger}.$$

Then we have the following decomposition

$$\frac{1}{\sqrt{\bar{k}_n}} \sum_{i=1}^{\bar{k}_n} \widetilde{Y}_{i,j}^{\dagger} = \frac{1}{\sqrt{\bar{k}_n}} \sum_{\ell=1}^{\bar{\ell}_n} \widetilde{S}_{\ell,j} + \frac{1}{\sqrt{\bar{k}_n}} \sum_{\ell=1}^{\bar{\ell}_n} \mathring{S}_{\ell,j} + \frac{1}{\sqrt{\bar{k}_n}} \sum_{i=\ell(q_{1,n}+q_{2,n})}^{\bar{k}_n} \widetilde{Y}_{i,j}^{\dagger}.$$

Therefore, by the triangle inequality,

$$\begin{aligned} &\left| \max_{1\leq j\leq 2m_n} \frac{1}{\sqrt{\bar{k}_n}} \sum_{i=1}^{\bar{k}_n} \widetilde{Y}_{i,j} - \max_{1\leq j\leq 2m_n} \frac{1}{\sqrt{\bar{k}_n}} \sum_{\ell=1}^{\bar{\ell}_n} \widetilde{S}_{\ell,j} \right| \\ &\leq \max_{1\leq j\leq 2m_n} \left| \frac{1}{\sqrt{\bar{k}_n}} \sum_{\ell=1}^{\bar{\ell}_n} \mathring{S}_{\ell,j} \right| + \max_{1\leq j\leq 2m_n} \left| \frac{1}{\sqrt{\bar{k}_n}} \sum_{i=\ell(q_{1,n}+q_{2,n})+1}^{\bar{k}_n} \widetilde{Y}_{i,j}^{\dagger} \right|. \end{aligned} \tag{A.62}$$

Moreover, let $(\widetilde{S}'_{\ell,j})_{1\leq \ell\leq \bar{\ell}_n}$ and $(\mathring{S}'_{\ell,j})_{1\leq \ell\leq \bar{\ell}_n}$ be two $F^{(0)}$-conditionally *independent* sequences such that $\widetilde{S}'_{\ell,j} \stackrel{\mathcal{L}}{=} \widetilde{S}_{\ell,j}$ and $\mathring{S}'_{\ell,j} \stackrel{\mathcal{L}}{=} \mathring{S}_{\ell,j}$. 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 A.3(i), it follows Corollary 2.7 of Yu (1994) that

$$\sup_{x\in\mathbb{R}} \left| \mathbb{P}\left(\max_{1\leq j\leq 2m_n} \sum_{\ell=1}^{\bar{\ell}_n} \widetilde{S}_{\ell,j} \leq x \middle| \mathcal{F}^{(0)} \right) - \mathbb{P}\left(\max_{1\leq j\leq 2m_n} \sum_{\ell=1}^{\bar{\ell}_n} \widetilde{S}'_{\ell,j} \leq x \middle| \mathcal{F}^{(0)} \right) \right| \leq K\bar{\ell}_n q_{2,n}^{-\upsilon} \leq K\Delta_n^{(2+\upsilon)\kappa-\rho}, \tag{A.63}$$

$$\sup_{x\in\mathbb{R}} \left| \mathbb{P}\left(\max_{1\leq j\leq 2m_n} \sum_{\ell=1}^{\bar{\ell}_n} \mathring{S}_{\ell,j} \leq x \middle| \mathcal{F}^{(0)} \right) - \mathbb{P}\left(\max_{1\leq j\leq 2m_n} \sum_{\ell=1}^{\bar{\ell}_n} \mathring{S}'_{\ell,j} \leq x \middle| \mathcal{F}^{(0)} \right) \right| \leq K\bar{\ell}_n q_{1,n}^{-\upsilon} \leq K\Delta_n^{(2+2\upsilon)\kappa-\rho}. \tag{A.64}$$

Taking positive constants ϵ and γ such that $\epsilon + \gamma < \big((2+\upsilon)\kappa - \rho\big) \wedge (\kappa/2) \wedge (\rho/2 - \kappa)$. Combing (A.62)-(A.64) and by the law of iterated expectation, we have

$$\begin{aligned}
&\mathbb{P}\Big(\max_{1\le j\le 2m_n} \frac{1}{\sqrt{\bar{k}_n}} \sum_{i=1}^{\bar{k}_n} \widetilde{Y}_{i,j}^{\dagger} < x\Big) \\
&\le \mathbb{P}\Big(\max_{1\le j\le 2m_n} \frac{1}{\sqrt{\bar{k}_n}} \sum_{\ell=1}^{\bar{\ell}_n} \widetilde{S}'_{\ell,j} \le x + K\Delta_n^{\gamma}\Big) + \mathbb{P}\Big(\max_{1\le j\le 2m_n} \mathfrak{E}_{n,j}^{(I)} \le \frac{K}{2}\Delta_n^{\gamma}\Big) \\
&\quad + \mathbb{P}\Big(\max_{1\le j\le 2m_n} \mathfrak{E}_{n,j}^{(II)} > \frac{K}{2}\Delta_n^{\gamma}\Big) + K'\Delta_n^{\epsilon},
\end{aligned} \tag{A.65}$$

where for $1 \le j \le 2m_n$, $\mathfrak{E}_{n,j}^{(I)} \equiv \big|\bar{k}_n^{-1/2} \sum_{\ell=1}^{\bar{\ell}_n} \mathring{S}'_{\ell,j}\big|$ and $\mathfrak{E}_{n,j}^{(II)} \equiv \big|\bar{k}_n^{-1/2} \sum_{i=\ell(q_{1,n}+q_{2,n})+1}^{\bar{k}_n} \widetilde{Y}_{i,j}^{\dagger}\big|$. For the second term, by Assumption A.1(iii) and Assumption A.3(ii), it follows Lemma 8 in Chernozhukov et al. (2015) that

$$\mathbb{E}\Big[\max_{1\le j\le 2m_n} \mathfrak{E}_{n,j}^{(I)} \Big| \mathcal{F}^{(0)}\Big] \le K\big(q_{1,n}^{-1/2} q_{2,n}^{1/2} \sqrt{L_n} + \bar{k}_n^{-1/2} q_{2,n} L_n^{3/2}\big) \le K\Delta_n^{(\kappa/2)\wedge(\rho/2-\kappa)} L_n^{3/2}.$$

Then by the Markov inequality and the law of iterated expectation, we obtain

$$\mathbb{P}\Big(\max_{1\le j\le 2m_n} \mathfrak{E}_{n,j}^{(I)} \le K\Delta_n^{\gamma}\Big) \le K'\Delta_n^{\epsilon}. \tag{A.66}$$

For the third term in the right hand side of (A.65), note that Assumption A.3(i) implies α-mixing and hence, combining with Assumption A.1(iii) and A.3(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 $1 \le j \le 2m_n$,

$$\mathbb{P}(\mathfrak{E}_{n,j}^{(II)} > K\Delta_n^{\gamma} | \mathcal{F}^{(0)}) \le \mathbb{P}(\mathfrak{E}_{n,j}^{(II)} > K\bar{k}_n^{-1/2} q_{1,n}^{1/2} \sqrt{L_n} | \mathcal{F}^{(0)}) = 0.$$

Then by the law of iterated expectation, we have

$$\mathbb{P}\Big(\max_{1\le j\le 2m_n} \mathfrak{E}_{n,j}^{(II)} > K\Delta_n^{\gamma}\Big) \le \sum_{j=1}^{2m_n} \mathbb{P}(\mathfrak{E}_{n,j}^{(II)} > K\Delta_n^{\gamma}) = 0. \tag{A.67}$$

Combining (A.65)-(A.67) yields

$$\mathbb{P}\Big(\max_{1\le j\le 2m_n} \frac{1}{\sqrt{\bar{k}_n}} \sum_{i=1}^{\bar{k}_n} \widetilde{Y}_{i,j}^{\dagger} < x\Big) \le \mathbb{P}\Big(\max_{1\le j\le 2m_n} \frac{1}{\sqrt{\bar{k}_n}} \sum_{\ell=1}^{\bar{\ell}_n} \widetilde{S}'_{\ell,j} \le x + K\Delta_n^{\gamma}\Big) + K'\Delta_n^{\epsilon}.$$

Following a similar argument, we can also show that

$$\mathbb{P}\Big(\max_{1\le j\le 2m_n} \frac{1}{\sqrt{\bar{k}_n}} \sum_{i=1}^{\bar{k}_n} \widetilde{Y}_{i,j}^{\dagger} < x\Big) \ge \mathbb{P}\Big(\max_{1\le j\le 2m_n} \frac{1}{\sqrt{\bar{k}_n}} \sum_{\ell=1}^{\bar{\ell}_n} \widetilde{S}'_{\ell,j} \le x - K\Delta_n^{\gamma}\Big) - K'\Delta_n^{\epsilon}.$$

Recall for each $1 \le j \le 2m_n$, the summand $(\widetilde{S}'_{\ell,j})_{1\le\ell\le\bar{\ell}_n}$ is $\mathcal{F}^{(0)}$-conditionally independent, then a similar strong approximation result can be established following the same proof as in Theorem 1.

References

Admati, A. R. and P. Pfleiderer (1988): "A theory of intraday patterns: Volume and price variability," *The Review of Financial Studies*, 1, 3–40.

Aït-Sahalia, Y. and J. Jacod (2014): *High-frequency financial econometrics*, Princeton University Press.

Ament, S. and M. O'Neil (2018): "Accurate and efficient numerical calculation of stable densities via optimized quadrature and asymptotics," *Statistics and Computing*, 28, 171–185.

Andersen, T., M. Thyrsgaard, and V. Todorov (2019): "Time-Varying Periodicity in Intraday Volatility," *Journal of the American Statistical Association*, 114, 1695–1707.

Andersen, T. G. (1996): "Return volatility and trading volume: An information flow interpretation of stochastic volatility," *The Journal of Finance*, 51, 169–204.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2003): "Modeling and forecasting realized volatility," *Econometrica*, 71, 579–625.

Andersen, T. G. and B. E. Sørensen (1996): "GMM estimation of a stochastic volatility model: A Monte Carlo study," *Journal of Business & Economic Statistics*, 14, 328–352.

Ante, L. (2023): "How Elon Musk's twitter activity moves cryptocurrency markets," *Technological Forecasting and Social Change*, 186, 122112.

Bahadur, R. R. (1966): "A note on quantiles in large samples," *The Annals of Mathematical Statistics*, 37, 577–580.

Bai, Y., A. Santos, and A. Shaikh (2019): "A practical method for testing many moment inequalities," *University of Chicago, Becker Friedman Institute for Economics Working Paper*.

Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard (2008): "Designing realized kernels to measure the ex post variation of equity prices in the presence of noise," *Econometrica*, 76, 1481–1536.

Barndorff-Nielsen, O. E. and N. Shephard (2004): "Econometric analysis of realized covariation: High frequency based covariance, regression, and correlation in financial economics," *Econometrica*, 72, 885–925.

Belloni, A., V. Chernozhukov, D. Chetverikov, and K. Kato (2015): "Some new asymptotic theory for least squares series: Pointwise and uniform results," *Journal of Econometrics*, 186, 345–366.

Bernanke, B. S. and K. N. Kuttner (2005): "What Explains the Stock Market's Reaction to Federal Reserve Policy?" *The Journal of Finance*, 60, 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.

Bollerslev, T., J. Li, and Q. Li (2023): "Optimal Nonparametric Range-Based Volatility Estimation," *Journal of Econometrics*, forthcoming.

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

Bollerslev, T., J. Li, and Y. Xue (2018): "Volume, Volatility, and Public News Announcements," *Review of Economic Studies*, 85, 2005–2041.

Brillinger, D. R. (1975): "Statistical inference for stationary point processes," *Stochastic Processes and Related Topics*, 1, 55–99.

Bugni, F. A., J. Li, and Q. Li (2023): "Permutation-based tests for discontinuities in event studies," *Quantitative Economics*, 14, 37–70.

Cattaneo, M. D., R. P. Masini, and W. G. Underwood (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, 760–777.

Chernozhukov, V., D. Chetverikov, and K. Kato (2013): "Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors," *The Annals of Statistics*, 2786–2819.

——— (2015): "Comparison and anti-concentration bounds for maxima of Gaussian random vectors," *Probability Theory and Related Fields*, 162, 47–70.

——— (2017): "Central limit theorems and bootstrap in high dimensions," *The Annals of Probability*, 45, 2309–2352.

——— (2019): "Inference on causal and structural parameters using many moment inequalities," *The Review of Economic Studies*, 86, 1867–1900.

Christensen, K., U. Hounyo, and M. Podolskij (2018): "Is the diurnal pattern sufficient to explain intraday variation in volatility? A nonparametric assessment," *Journal of Econometrics*, 205, 336–362.

Christensen, K. and A. Kolokolov (2023): "An unbounded intensity model for point processes," *Available at SSRN 4513848*.

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

Coeurjolly, J.-F. (2008): "Hurst exponent estimation of locally self-similar Gaussian processes using sample quantiles," *Annals of Statistics*, 36, 1404–1434.

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

COX, D. R. (1955): "Some statistical methods connected with series of events," *Journal of the Royal Statistical Society: Series B (Methodological)*, 17, 129–157.

COX, J. C., J. E. INGERSOLL JR, AND S. A. ROSS (1985): "An intertemporal general equilibrium model of asset prices," *Econometrica*, 363–384.

DAVIDSON, J. (1992): "A central limit theorem for globally nonstationary near-epoch dependent functions of mixing processes," *Econometric theory*, 8, 313–329.

DU, S. AND H. ZHU (2017): "What is the optimal trading frequency in financial markets?" *The Review of Economic Studies*, 84, 1606–1651.

DURBIN, J. AND S. J. KOOPMAN (1997): "Monte Carlo maximum likelihood estimation for non-Gaussian state space models," *Biometrika*, 84, 669–684.

EINMAHL, U. AND D. LI (2008): "Characterization of LIL behavior in Banach space," *Transactions of the American Mathematical Society*, 360, 6677–6693.

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

GARMAN, M. B. AND M. J. KLASS (1980): "On the Estimation of Security Price Volatilities from Historical Data," *Journal of Business*, 53, 67–78.

GENON-CATALOT, V., C. LAREDO, AND M. NUSSBAUM (2002): "Asymptotic equivalence of estimating a Poisson intensity and a positive diffusion drift," *The Annals of Statistics*, 30, 731–753.

GENTZKOW, M., B. KELLY, AND M. TADDY (2019): "Text as data," *Journal of Economic Literature*, 57, 535–574.

GHOSH, J. K. (1971): "A new proof of the Bahadur representation of quantiles and an application," *The Annals of Mathematical Statistics*, 1957–1961.

GRAMA, I. AND M. NUSSBAUM (1998): "Asymptotic equivalence for nonparametric generalized linear models," *Probability Theory and Related Fields*, 111, 167–214.

HANSEN, A. L. AND S. KAZINNIK (2023): "Can ChatGPT Decipher Fedspeak?" *Available at SSRN*.

HARVEY, A., E. RUIZ, AND N. SHEPHARD (1994): "Multivariate stochastic variance models," *The Review of Economic Studies*, 61, 247–264.

HAYASHI, T., J. JACOD, AND N. YOSHIDA (2011): "Irregular sampling and central limit theorems for power variations: The continuous case," in *Annales de l'IHP Probabilités et statistiques*, vol. 47, 1197–1218.

HAYOU, S., A. DOUCET, AND J. ROUSSEAU (2019): "On the impact of the activation function on deep neural networks training," in *International conference on machine learning*, PMLR, 2672–2680.

HELMERS, R. AND R. ZITIKIS (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*, 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, 327–343.

HOEFFDING, W. (1963): "Probability Inequalities for Sums of Bounded Random Variables," *Journal of the American Statistical Association*, 58, 13–30.

HONG, H. AND J. WANG (2000): "Trading and returns under periodic market closures," *The Journal of Finance*, 55, 297–354.

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, 37–50.

JACOD, J., J. LI, AND Z. LIAO (2021): "Volatility Coupling," *The Annals of Statistics*, 49, 1982–1998.

JACOD, J., Y. LI, P. A. MYKLAND, M. PODOLSKIJ, AND M. VETTER (2009): "Microstructure noise in the continuous case: the pre-averaging approach," *Stochastic processes and their applications*, 119, 2249–2276.

JACOD, J., Y. LI, AND X. ZHENG (2017): "Statistical properties of microstructure noise," *Econometrica*, 85, 1133–1174.

JACOD, J. AND P. PROTTER (2012): *Discretization of Processes*, vol. 67, Springer Science & Business Media.

JACOD, J. AND M. ROSENBAUM (2013): "Quarticity and Other Functionals of Volatility: Efficient Estimation," *Annals of Statistics*, 41, 1462–1484.

KARR, A. (1991): *Point Processes and Their Statistical Inference*, vol. 7, CRC Press.

KINGMAN, J. F. (1961): "The single server queue in heavy traffic," in *Mathematical Proceedings of the Cambridge Philosophical Society*, Cambridge University Press, vol. 57, 902–904.

KNIGHT, J. L. AND J. YU (2002): "Empirical characteristic function in time series estimation," *Econometric Theory*, 18, 691–721.

KOENKER, R. (2005): *Quantile regression*, vol. 38, Cambridge university press.

KOENKER, R. AND G. BASSETT JR (1978): "Regression quantiles," *Econometrica: journal of the Econometric Society*, 33–50.

KOLOKOLOV, A. (2022): "Estimating jump activity using multipower variation," *Journal of Business & Economic Statistics*, 40, 128–140.

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

Kuchibhotla, A. K. and A. Chakrabortty (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, 1389–1456.

Le Cam, L. (1986): *Asymptotic methods in statistical decision theory*, Springer Series in Statistics.

Le Cam, L. and G. L. Yang (2000): *Asymptotics in statistics: some basic concepts*, Springer Science & Business Media.

Lehmann, E., J. P. Romano, and J. Popper (2005): "On Optimality of Stepdown and Stepup Multiple Test Procedures," *The Annals of Statistics*, 33, 1084–1108.

Li, J. and Z. Liao (2020): "Uniform nonparametric inference for time series," *Journal of Econometrics*, 219, 38–51.

Li, J., V. Todorov, and G. Tauchen (2017): "Adaptive estimation of continuous-time regression models using high-frequency data," *Journal of Econometrics*, 200, 36–47.

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

Li, Z. M. and O. Linton (2022): "A ReMeDI for Microstructure Noise," *Econometrica*, 90, 367–389.

Liu, C. and C. Y. Tang (2013): "A state space model approach to integrated covariance matrix estimation with high frequency data," *Statistics and Its Interface*, 6, 463.

Loughran, T. and B. McDonald (2020): "Textual analysis in finance," *Annual Review of Financial Economics*, 12, 357–375.

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

Mogstad, M., J. P. Romano, A. M. Shaikh, and D. Wilhelm (2023): "Inference for Ranks with Applications to Mobility across Neighbourhoods and Academic Achievement across Countries," *Review of Economic Studies*, 00, 01–23.

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

Nakamura, E. and J. Steinsson (2018): "Identification in Macroeconomics," *Journal of Economic Perspectives*, 32, 59–86.

Nelson, D. B. (1990): "ARCH models as diffusion approximations," *Journal of Econometrics*, 45, 7–38.

Parkinson, M. (1980): "The Extreme Value Method for Estimating the Variance of the Rate of Return," *Journal of Business*, 53, 61–65.

Renault, E., C. Sarisoy, and B. J. Werker (2017): "Efficient Estimation of Integrated Volatility and Related Processes," *Econometric Theory*, 33, 439–478.

Resnick, S. I. (2008): *Extreme values, regular variation, and point processes*, vol. 4, Springer Science & Business Media.

Rigobon, R. and B. Sack (2004): "The Impact of Monetary Policy on Asset Prices," *Journal of Monetary Economics*, 51, 1553–1575.

Rio, E. (1995): "The functional law of the iterated logarithm for stationary strongly mixing sequences," *The Annals of Probability*, 23, 1188–1203.

Romano, J. P. and M. Wolf (2005): "Stepwise multiple testing as formalized data snooping," *Econometrica*, 73, 1237–1282.

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, 271–276.

Savor, P. and M. Wilson (2014): "Asset Pricing: A Tale of Two Days," *Journal of Financial Economics*, 113, 171–201.

Shen, D., A. Urquhart, and P. Wang (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*.

Szulga, A. (1983): "On minimal metrics in the space of random variables," *Theory of Probability & Its Applications*, 27, 424–430.

Tandon, C., S. Revankar, and S. S. Parihar (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, 100035.

van der Vaart, A. and J. Wellner (1996): *Weak Convergence and Empirical Processes: With Applications to Statistics*, Springer Science & Business Media.

Vladimirova, M., S. Girard, H. Nguyen, and J. Arbel (2020): "Sub-Weibull distributions: Generalizing sub-Gaussian and sub-Exponential properties to heavier tailed distributions," *Stat*, 9, e318.

Wu, W. B. (2005): "On the Bahadur representation of sample quantiles for dependent sequences," *The Annals of Statistics*, 33, 1934–1963.

Yu, B. (1994): "Rates of convergence for empirical processes of stationary mixing sequences," *The Annals of Probability*, 94–116.

Zhang, L., P. A. Mykland, and Y. Aït-Sahalia (2005): "A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High-Frequency Data," *Journal of the American Statistical Association*, 100, 1394–1411.

Zhang, X. and G. Cheng (2014): "Bootstrapping high dimensional time series," *arXiv preprint arXiv:1406.1037*.

Zolotarev, V. M. (1986): *One-dimensional stable distributions*, vol. 65, American Mathematical Soc.

Zuo, Y. (2015): "Bahadur representations for bootstrap quantiles," *Metrika: International Journal for Theoretical and Applied Statistics*, 78, 597–610.