Uniform Inference of State-space Model for High Frequency Data

Qiyuan Li*

September 13, 2023

JOB MARKET PAPER

For the latest draft, click here.

Abstract

In this study, we consider the inference problem of a continuous-time state-space model that may exhibit non-linearity and non-Gaussianity over a fixed time span. The model is aligned with numerous high-frequency financial econometric applications, including location-scale volume, Lévy-driven returns, and the Cox trading flow process. Our focus lies on the nonparametric uniform inference for the entire conditional mean process and the conditional quantile process of the transformed state under in-fill asymptotics. The estimators are constructed by collecting the blockwise estimates. Notably, while these estimators do not admit a functional central limit theorem, we establish Gaussian strong approximation which facilitates the feasible uniform inference. As a distinctive empirical application of the proposed methodology, we conduct a sentence-by-sentence analysis of trading flow process to discern the informative part from the Federal Open Market Committee (FOMC) press conference speeches. Our results suggest that stand-alone natural language processing methods may overestimate the information level, and often fail to pinpoint the actual informative sentences. This observation highlights the potential of our approach as a compliment to the volatility-based detection mechanisms and conventional natural language processing tools.

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

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

1 Introduction

Since its first introduction by Kalman (1960) in control engineering, the *state-space* model has gained widespread adoption across diverse fields, including environmental science, computer science, medicine, macroeconomics, and financial economics. This model offers a general framework for the analysis of stochastic dynamic systems, partially characterized by an underlying state process. A quintessential state-space model comprises two equations. The first, known as the observation equation, delineates the relationship between observed measurements and latent state variables. The second, known as the transition equation, captures the dynamics and evolution of these state variables. See Durbin and Koopman (2001) for a comprehensive introduction of the state-space methodology.

The inherent flexibility of the state-space model renders it an effective and powerful tool for practical time series analysis and forecasting. Notably, numerous time series models can either be articulated directly in a state-space framework or admit some observationally equivalent state-space representations. Illustrative examples include the classical ARMA processes, unit root processes, and cointegrating systems (Aoki and Havenner (1991), Durbin and Koopman (2001), and Bauer and Wagner (2012)), and structural time series models (Harvey (1989)). The majority of these state-space formulations alluded to are discrete-time and parametric in nature, Harvey (1989), West and Harrison (1997), Kim et al. (1999), and Harvey et al. (2004) provide general treatments and inference techniques for these state-space models.

Prompted by the increased availability of high-frequency data in the financial markets, the continuous-time state-space framework has emerged as a favored approach for modeling observable intraday price movements and latent stochastic volatility patterns, as explored by Nelson (1990) and Heston (1993), among others. Distinct from the long-span asymptotics usually employed in the discrete-time state-space frameworks, the profusion of high-frequency data suggests a shift towards in-fill asymptotics for analyzing the continuous-time state-space models. Eschewing restrictive assumptions on the dynamics of volatility and the ensuing parameter estimations, recent research predominantly concentrates on the direct estimation of the stochastic volatility process, hence is distinctly nonparametric, see Andersen and Bollerslev (2018) for an introductory discussion.

In their seminal work, Li and Xiu (2016) propose a continuous-time Markov state-space model. Within this framework, each observed variable is modeled as a noisy transformation of the contemporary state. Owing to the absence of constraints on the functional form of the transformation, this framework exhibits greater versatility than the classical linear state-space specification, which facilitates its applicability to various financial models, such as derivative pricing or the Poisson volume-volatility model as introduced by Andersen (1996). The paper then develop a generalized method of moments estimation based on the integrated moment condition involving the transformed state process. Subsequent to this, in a more recent paper, Bollerslev et al. (2018) develop the estimation and inference method for the instantaneous mean of transformed state at some given time point. However, one inherent constraint of their framework is the adoption of the Markovian assumption. As a complement, Bugni et al. (2023) proffers to model the observed variable as a noisy transformation of the contemporary state plus a residual term, which serves to absorb the biases stemming from local non-Markovian characteristics and nuisance variables/processes. Crucially, the inclusion of the additional residual term paves the way for modeling high-frequency returns when the (log) price follows an Itô semimartingale, in which the state process is the volatility. Large amount literature has been devoted to studying the inference of volatility using high-frequency data, see Foster and Nelson (1996), Comte and Renault (1998), and Kristensen (2010) for estimating spot volatility, Lee and Mykland (2008), Jacod and Rosenbaum (2013), and Li et al. (2017) for estimating integrated volatility functionals. However, given the absence stationarity assumption of volatility dynamics, the uniform inference for the entire volatility process is a relatively nascent research topic, which is broached and explored in Jacod et al. (2021) and Bollerslev et al. (2021).

In this paper, we follow the general state-space framework in Li and Xiu (2016) to move beyond the constraints of linearity and Gaussianity In addition, we adopt the strategy of Bugni et al. (2023) to accommodate potential non-Markovian aspects. Our emphasis diverges from the mere estimation of the integrated tate process or the state at specific time points; instead, we target the uniform inference for the entire state process. Specifically, we construct functional estimators and associated inference procedures for the conditional mean process and the conditional quantile process of the transformed state. The major challenge in uniform inference arises due to the asymptotic independence of estimation errors of the proposed functional estimators at any two distinct time points. Consequently, the functional estimators do not admit a functional central limit theorem. Recent literature like Chernozhukov et al. (2013), Jacod et al. (2021), and Bollerslev et al. (2021) shed lights on the uniform inference for the non-Donsker problems, highlighting the use of strong approximation. Based on this insight, we derive the Gaussian strong approximation of our proposed functional estimators for both conditional mean process (Theorem 1) and conditional quantile process (Theorem 2). For the conditional mean process, our result can be contextualized as an extension of Jacod et al. (2021) to a more abstract state-space framework.

For the conditional quantile process, which is of great theoretical importance when the model exhibits heavy tails, we establish a novel Gaussian strong approximation result facilitated by a local Bahadur representation. Notably, our results can be applied to the volatility of Lévy-driven price and stochastic intensity of Cox process. To the best of our knowledge, this is the first paper contributing to the uniform inference for these processes.

The rest of the paper is organized as the following. In subsection 2.1, we delineate the state-space model employed in our research. In subsection 2.2, we provide three running examples that used to illustrate adaptability of our framework to various models and to demonstrate the fulfillment of our assumptions. The construction and strong approximation results for the functional estimators of both the conditional mean process and the conditional quantile process are given in subsections 2.3 and 2.4, respectively. In Subsection 2.5, we offer an application to constructing confidence set for ranks of the spot values of investigated process to illustrate how our strong approximation findings can be utilized to tackle other econometric problems. Section 3 contains a Monte-Carlo experiment analysis. An empirical application is presented in Section 4, where we employ the proposed inference methodology to discern information flows during FOMC speeches. 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 $a_n/C \leq b_n \leq C a_n$ for some finite constant $C \geq 1$.

2 Theory

2.1 State-space Model for High Frequency Data

We observe a data sequence $(Y_{i\Delta_n})$ at some discrete times where $1 \leq i \leq n \equiv \lfloor T/\Delta_n \rfloor$, within a fixed time span $[0, T]$. It is assumed that the data is generated based on the subsequent non-linear non-Gaussian state-space model

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

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 denoted as $\left(\Omega^{(0)}, \mathcal{F}^{(0)}, (\mathcal{F}_t^{(0)})_{t\geq 0}, \mathbb{P}^{(0)}\right)$. The function $\mathscr{Y}(\cdot, \cdot)$ represents a deterministic noisy transform of the *current* state $\zeta_{i\Delta_n}$ and 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.[1] This residual term can be considered uniformly negligible in comparison with the dominant term, as per the requirement provided in the subsequent subsections.

We will make the assumption that the random disturbance $(\varepsilon_{n,i})_{1\leq i\leq n}$ is a $\mathcal{F}^{(0)}$-conditionally independently and identically distributed (i.i.d.) sequence.[2] Note that the framework presented here can be extended to accommodate conditionally stationary and weakly dependent disturbances by employing the methodologies developed in works such as Zhang and Cheng (2014), Chernozhukov et al. (2019), and Li and Liao (2020). However, it is worth mentioning that, in many empirical scenarios illustrated in the examples provided in subsection 2.2, the disturbance exhibits conditional independence. Hence, in order to avoid unnecessary technical complexities, our primary focus lies on the assumption of conditional independence for the disturbance. In order to formally delineate the framework, we introduce another probability space denoted as $(\Omega^{(1)}, \mathcal{F}^{(1)}, \mathbb{P}^{(1)})$ endowed with an i.i.d. sequence $(\varepsilon_{n,i})_{1\leq i\leq n}$ with its marginal distribution denoted by $\mathbb{P}_\varepsilon$. 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 the 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 the current state ζ_t through the function $\mathscr{Y}(\cdot,\cdot)$ is not overly restrictive owing to the inclusion of the additional residual term. Although, from an intuitive standpoint, Y_t could potentially depend on the historical states. However, given that the state process exhibits sufficient smoothness, the information encapsulated in the difference between past state and current state could be effectively captured within the residual term $R_{n,i}$. For example, when the observation Y_t depends on a local window of historical state $(\zeta_s)_{s\in[t-h,t]}$ through some noisy functional, this "approximate Markovian" property 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, i.e. $\sup_{s,r\in[t-h,t]} \|\zeta_s - \zeta_r\| = O_p(h)$; and (iii) the window size is shrinking, i.e. $h = o(1)$. On the other

[1] The residual terms are assumed to be zero in Li and Xiu (2016) and Bollerslev et al. (2018).

[2] There is no loss of generality to impose independence between disturbance and state process here. One can always selection an appropriate normalization of this representation so that $\varepsilon_{n,i}$ is independent from $\zeta_{i\Delta_n}$ and let $\mathscr{Y}(\cdot,\cdot)$ account for the dependence.

hand, the residual term can also absorb the dependence of observations on some nuisance process when the effect is negligible. Consequently, the incorporation of the residual term $R_{n,i}$ renders our framework to an essentially "approximately Markovian" setting, which is more comprehensive comparing with the simpler Markov state-space models employed in studies such as Li and Xiu (2016) and Bollerslev et al. (2018).

2.2 Motivating Examples

To facilitate a better understanding of the broader implications of the general model (2.1), it is beneficial to outline the discussion using some empirically relevant running examples. In this subsection, we provide three motivating examples, showing how commonly used financial econometric models align with our state-space model 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 the process σ captures potential heteroskedasticity in time. This additive structure directly fits in the 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 the dynamics of the 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 $\sigma_{i\Delta_n}\varepsilon_{n,i}$ may be the pricing error. Liu and Tang (2013) employ the additive state-space model to devise an expectation-maximization algorithm tailored for the estimation of integrated volatility matrices, particularly when asset prices are observed with microstructure noise. In their model, $Y_{i\Delta_n}$ is the observed price, the state $\mu_{i\Delta_n}$ is the associated latent price and is assumed to have a VAR dynamics, $\sigma_{i\Delta_n}\varepsilon_{n,i}$ is the microstructure noise component where $\sigma_{i\Delta_n}$ captures the time-varying heterogeneity in the noise magnitude. Bugni et al. (2023) also use the additive state-space model to describe the trading volume process, where $\mu_{i\Delta_n}$ is the local mean of volume, and $\varepsilon_{n,i}$ captures the time-varying heterogeneity in order size. A particularly fitting application of the additive state-space model emerges when the observation is, in itself, an estimation of the spot state process. This specification aligns closely with the fixed-k estimation framework introduced by Bollerslev et al. (2021). Specifically, let $\log(\hat{\sigma}_{n,i})$ be the logarithm fixed-k estimator of the spot variance at time

$i\Delta_n$, and $\log(\sigma_{n,i})$ be the logarithm of true spot variance, Bollerslev et al. (2021) show 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. □

Example 2 (Lévy-driven Asset Returns). The proposed state-space model can also be applied to characterize a wide range of price dynamics studied in the high-frequency financial econometrics literature. Consider the log price which has the drift component and a jump-diffusion component driven by a Lévy martingale L, i.e., the 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 the Blumenthal–Getoor index $\beta \in (0,2]$ and is assumed to be independent with σ.[3][4] 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, we observe that the scaled Lévy increments $\Delta_n^{-1/\beta}(L_{(i+1)\Delta_n} - L_{i\Delta_n})$ are i.i.d. across $1 \le i \le n$ and have a non-degenerate distribution. Therefore, upon expanding the above display and collecting dominant terms, the model can be rewritten in the form of model (2.1) by setting

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

[3] Note that for a stable process, the Blumenthal–Getoor index and the stability index agree. A general stable process has the 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 and σ/K generate the same price process. Therefore, to avoid the non-identification issue between σ and the "scale" of L, we make additional restriction that $c = 1$ if $\beta = 2$ and $a = 1/\pi$ if $\beta \in (0,2)$.

[4] The independence assumption between L and σ rules out the interaction between price and volatility, 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 increment term. That is, the independence requirement can be dropped here, for the case when L is a Brownian motion, see Jacod et al. (2021).

Distinct with the preceding example, 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 the current volatility $\sigma_{i\Delta_n}$ and remains dependent on the ancillary process μ, it may still conform to an "approximate Markovian" characterization involving the current volatility. As discussed in the subsequent sections 2.3 and 2.4, the residual term can be rendered uniformly negligible providing the processes μ and σ satisfy some fairly weak assumptions. □

Example 3 (Cox Trading Flows). Consider the number of trades during the time span $[0, t]$ as denoted by N_t. It is cogent to model the trading flows as a Cox process — or referred to as the doubly stochastic Poisson process — which is originally introduced by Cox (1955) for modeling the fibrous threads, i.e. conditional on the process μ_t, N_t behaves as an inhomogeneous Poisson process with an intensity function μ_t. Let $Y_{i\Delta_n} = N_{(i+1)\Delta_n} - N_{i\Delta_n}$ denote the 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) $\mathbb{P}(Y_{i\Delta_n} \geq 2|\mu_t) = o(\Delta_n)$ and (ii) $\mathbb{P}(Y_{i\Delta_n} = 1|\mu_t) = \Delta_n\mu_{i\Delta_n} + o(\Delta_n)$. This naturally suggest a compelling approximation of $Y_{i\Delta_n}$ by a mixed Bernoulli random variable with parameter $\Delta_n\mu_{i\Delta_n}$.[5] Consequently, there exists a sequence of independent, uniformly distributed variables $(\varepsilon_{n,i})_{1\leq i\leq n}$ on $[0, 1]$ which are also independent of process μ such that

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

where the residual term takes value in $\{-1\} \cup \mathbb{N}$ and satisfies $\mathbb{P}(R_{n,i} \neq 0|\mu_t) = o(\Delta_n)$ according to property (i) and (ii).[6] Note that the increments over disjoint intervals can be in general dependent in a Cox process through the $\mu_{i\Delta_n}$ part, as contrasted with the postulated independence in the conventional Poisson process. Above display shows the increment of trading flow process can be expressed in the form of our state-space model (2.1) by setting

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

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

[6] 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 $\|\mathcal{L}(N) - \mathcal{L}(N^n)\|_{\mathrm{TV}} \to 0$, where $\|\cdot\|_{\mathrm{TV}}$ denotes the total variation norm of measures. This result aligns with the asymptotic equivalence of statistical experiments in the 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).

We stress the importance of analyzing trading flow process for the following reason. In the TAQ (Trade and Quote) database, each trade is recorded with a precision of nanoseconds (10^{-9} seconds).[7] Our mixed Bernoulli approximation indeed matches remarkably well with empirical data: we record whether a trade has transpired within the preceding nanosecond window. Comparing with the volume which may also oscillate due to unobserved trader-specific heterogeneity and the recorded prices which is often contaminated by the microstructure noise, the trading intensity is more closely related to the information flow. That being said, as a compliment to the price movement, which contains consensual decisions and viewpoints of market participants, the trading frequency also reflects the speed at which market participants are reacting to and incorporating new information into their idiosyncratic trading strategies. A surge in trading frequency usually indicates a higher level of information flow and potentially reflects real-time changes in market sentiment, news announcements, or other factors that influence trading activity. □

The aforementioned examples show the general state-space model (2.1) can be cast to model various market indicators such as high-frequency volumes, prices, and trading flows. In the following subsections, we will construct functional estimators and inference procedure for the conditional mean process and conditional quantile process of transformed state, and provide further practical implementation details of these examples.

2.3 Uniform Inference on Conditional Mean Process

Our primary interest lies not in estimating the latent state process ζ per se, but rather in the specific transformations of the state process. In this subsection, we focus on estimating the instantaneous conditional mean process g. 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 the conditional mean process may not always be well-defined, especially when the disturbance exhibits heavy tails. As a supplementary measure, we discuss the estimation and inference of conditional quantile process in the next subsection, which always exists. The precise implications of these processes, along with the identification procedure of the state process ζ from these processes, intrinsically depend on the specific form of transformation $\mathscr{Y}(\cdot, \cdot)$ and the distribution $\mathbb{P}_\varepsilon$. These aspects should be analyzed on a meticulous case-by-case basis.

[7] 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.

In preparation for a deep dive into the estimation procedure estimation procedure, we first introduce some additional notations concerning the block sampling scheme which is particularly useful in the inference for high-frequency data. This scheme divides our observational window into distinct, manageable blocks, facilitating the construction of our estimators, 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, denote $\iota(i,j) \equiv \min \mathcal{I}_{n,j} + i - 1$ 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 the 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, then it naturally suggests forming an estimator by taking the local average within the block which contains time t, while keeping the 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 $t \in \mathcal{T}_{n,j}$, it is straightforward to define $\hat{g}_t$ as the local average of observations $Y_{i\Delta_n}$ over this block $\hat{g}_{n,t} \equiv k_{n,j}^{-1} \sum_{i \in \mathcal{I}_{n,j}} Y_{i\Delta_n}$. Theorem 1 in Bollerslev et al. (2018) shows that when $R_{n,i} = 0$, under fairly weak conditions on the local smoothness of ζ and bounded second conditional moments of $\mathscr{Y}(\cdot, \varepsilon)$, 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 $\sigma_t^2 \equiv \int_{\mathcal{D}} \mathscr{Y}(\zeta_t, \varepsilon)^2 \mathbb{P}_\varepsilon(d\varepsilon) - \left(\int_{\mathcal{D}} \mathscr{Y}(\zeta_t, \varepsilon)\mathbb{P}_\varepsilon(d\varepsilon)\right)^2$, $\xrightarrow{\mathcal{L}\text{-s}}$ denotes stable convergence in law, and $\mathcal{MN}$ denotes the 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 that the 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 asymptotic variance σ_t^2, we can derive the feasible central limit theorem

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

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

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

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

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

Above results can be easily extend to the case when $R_{n,i} \neq 0$ yet remains uniformly negligible, and furthermore, the joint convergence of g on a *finite* set of time points $\{t_1, \ldots, t_\ell\} \subset [0, T]$. By a classical Bonferroni approach, the hyper-rectangle $C^{\pm}_{n,t_1}(\alpha/\ell) \times \cdots \times C^{\pm}_{n,t_\ell}(\alpha/\ell)$ serves as a valid confidence set for the vector $(g_{t_1}, \ldots, g_{t_\ell})$. However, the difficulty arises in extending this to the estimation of 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 the spot estimator for the jth block similar as before

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

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

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

Note that the blocks are non-overlapping, the estimation errors within different blocks are asymptotically independent. Consequently, the pointwise central limit theorem (2.2) shows that the process of spot estimators have a path structure similar to Gaussian white noise, hence is non-Donsker in nature. That being said, such non-Donsker problems that commonly arise from uniform inference in nonparametric settings, can be addressed by using strong approximation of the nonparametric 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 time series data. To help fix ideas, we define the *sup-t statistic* as

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

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 $\hat{\sigma}^2_{n,j} \equiv k^{-1}_{n,j} \sum_{i \in \mathcal{T}_{n,j}} Y^2_{i\Delta_n} - \big(k^{-1}_{n,j} \sum_{i \in \mathcal{I}_{n,j}} Y_{i\Delta_n}\big)^2$. Theorem 1, presented subsequently, shows the sup-t statistic can be strongly approximated by the 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 the random disturbance, and the residual term, which allow for essentially unrestricted nonstationary state process and heavy tailed disturbance. Specifically, condition (i) requires the state process to be locally taken value in compact set. Condition (i) also imposes smoothness of the state process *within* each block, namely, it requires the 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. It holds if the state process is a continuous Itô semimartingale or long-memory process within each block, and it also allows the 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 the aforementioned examples. Condition (iii) requires the transformed disturbance to have a sub-Weibull tail with parameter η, which is a generalization of the 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 analysis 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 deep neural networks, see Hayou et al. (2019). We highlight that condition (iii) also ensures the existence of conditional mean process. Condition (iv) is a high-level condition which requires the 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 the sup-t statistic, we provide some additional implementation details by revisiting the three examples outlined in the preceding subsection. The discussion of implementation details primarily aims to shed light on the interplay between the conditional mean process and the state process, and the validation of Assumption 1 (especially condition iv), within those specific models.

Example 1 (Location-Scale Model, continued). In the simple location-scale model with additive structure, suppose that the disturbance is centered, then by definition the conditional mean process inherently translates into the 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-Weilbull 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 the residual term $R_{n,i} = 0$ for all $1 \leq i \leq n$, Assumption 1 (iv) trivially holds for any $r > 0$. □

Example 2 (Lévy-driven Asset Returns, continued). Recall the characteristic triple of stable Lévy process described in the footnote 3, the conditional mean process is coherently well-defined only when $\beta = 2$, i.e. L is a Brownian motion.[8] Therefore, the subsequent discussion in this subsection is confined to the case where $\beta = 2$, the case where $\beta \in (0, 2)$ will be addressed in the succeeding subsection. 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 state $(z\varepsilon)^2$ follows a scaled $\chi^2(1)$ distribution. As a result, the conditional mean process translates into the 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 the 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 context the Cox trading flow model, recall that the 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 even more critically, to be both bounded from above and away from zero,[9] which alludes to the "high traffic" assumption, as introduced in Kingman (1961). As a complement elaboration, Christensen and Kolokolov (2023) provide an alternative justification for the "high traffic" assumption by modeling the trading flow as the sum of n independent copies of Cox processes with conditional intensity $\Delta_n \mu_t$. The "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

[8]Generally, by the property of β, the instantaneous conditional mean diverges at a rate of $\Delta_n^{2/\beta - 1}$.

[9]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)).

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 that $\mathbb{P}(R_{n,i} \neq 0|\mu_t) = o(\Delta_n)$, by the law of iterated expectation we have for any $r > 0$,

$$\mathbb{P}\Big(\max_{1\leq i\leq n} |R_{n,i}| > \Delta_n^r\Big) \leq \sum_{i=1}^{n} \mathbb{P}(R_{n,i} \neq 0) = no(\Delta_n) = o(1),$$

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

We are now ready to formally state the result for the strong approximation of the sup-t statistic.

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 the 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.[10] In that case, there exist sequences on the common probability space $\widehat{T}_n' \stackrel{d}{=} \widehat{T}_n^*$ and $Z_n' \stackrel{d}{=} \max_{1\leq j\leq m_n} |Z_j|$ such that $\widehat{T}_n' = Z_n' + o_p(\Delta_n^{\epsilon})$. However, it is not straightforward that convergence under the Kantorovich–Monge–Rubinstein metric implies convergence in laws under the Kolmogorov–Smirnov distance, since the density of Z_n' is unbounded.[11] Therefore, the theorem is stated employing the Kolmogorov–Smirnov distance because of the particular usefulness in making inference.

Comment 2. We emphasize that the 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

[10] 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 $\inf\{\mathbb{E}[\|X-Y\|] : \mathcal{L}(X) = \mathbb{P}_1, \mathcal{L}(Y) = \mathbb{P}_2\}$.

[11] 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. 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 $m_n \to \infty$.

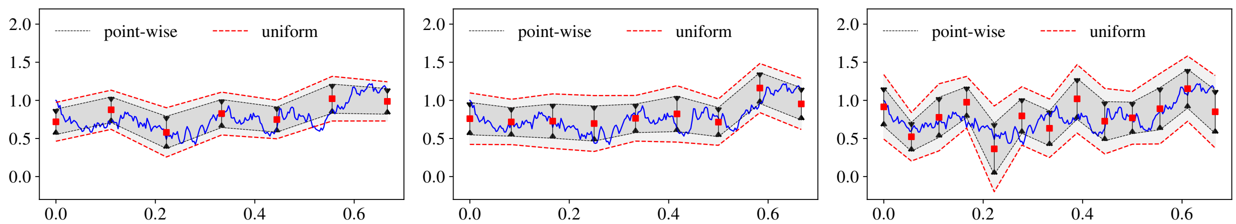

Figure 1: **Comparison of Confidence Bands under Different Number of Blocks**. In each panel, we mark the spot estimators in red squares, the 90% pointwise confidence interval in block vertical segments, the 90% uniform confidence band in red dashed lines, and the true process in blue lines. The pointwise confidence band is constructed by connecting each confidence interval computed using (2.3), the uniform confidence band is computed using (2.4). The three panels from left to right show the 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.

$(1-\alpha)$ quantile of $\max_{1\le j\le m_n}|Z_j|$, which can be computed numerically for any m_n. Theorem 1 implies $\left|\mathbb{P}\big(\widehat{T}_n^* \le cv_n(\alpha)\big) - \mathbb{P}(\max_{1\le j\le m_n}|Z_j| \le cv_n(\alpha)\right| \le K\Delta_n^{\epsilon}$. Consequently, let

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

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.,

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

Observe that the uniform confidence band (2.4) is generally wider than the pointwise confidence interval (2.3), the difference magnifies as the number of blocks m_n becomes larger. To better illustrate the intuition behind this difference, we present a simple comparative visualization for the uniform confidence band and the pointwise confidence interval under different numbers of blocks in Figure 1. Given that the 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, the probability of committing type I error across distinct blocks accumulates. Such accumulating errors are not accommodated for in the pointwise confidence interval.

2.4 Uniform Inference on Conditional Quantile Process

As we mentioned in the previous subsection, if the disturbance exhibits exceedingly heavy tails, the instantaneous conditional mean process is not well-defined. In contrast to the conditional mean process, the conditional quantile process remains well-defined, regardless of the nature of $\mathbb{P}_\varepsilon$. This subsection pivots to explore an alternative method of analyzing these heavy-tailed models, centering on the instantaneous conditional quantile of the transformed state as a supplemental measure.[12] To be precise, for some pre-determined level $\chi \in (0,1)$, we define

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

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 the 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 the estimator through the sample counterpart, and the theory of extreme estimators applies. Alternatively, although essentially equivalent in most cases, many statisticians opt to define the sample quantile estimator 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 establish the almost sure bound of representing the difference between the population quantile and corresponding order statistics as a sample average of some i.i.d. auxiliary variables. Ghosh (1971) give the simple proof for a weaker but sufficiently useful bound. The result is extended to nonparametric quantile regression by Chaudhuri (1991) and to the case of weakly dependent stationary data by Hesse (1990) and Wu (2005).

We adopt the idea from classic statistic methodology to define each spot estimator as the "in-fill order statistic" of observations inside the block, instead of through the convention of a minimization problem. Namely, within each block, we re-index the sequence $(Y_{i\Delta_n})_{i\in\mathcal{I}_{n,j}}$ in the non-decreasing order and denoted as $Y^o_{1,j} \leq \cdots \leq Y^o_{k_{n,j},j}$. The spot estimator for conditional quantile, in this scheme, is defined as the $\lceil k_{n,j}\chi\rceil$-order statistic with each block.[13] Analogous to the previous subsection, we construct the functional estimator as the collection of all the spot estimators

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

[12] 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.

[13] The results presented in this subsection 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 the $\lceil k_{n,j}\chi\rceil$-order statistic to avoid unnecessary complexity.

Although the observations from model (2.1) are neither independent nor identically distributed, in the appendix we show that a local Bahadur representation holds uniformly for all blockwise sample quantiles given some regularity conditions, implying that each spot estimator obeys a central limit theorem. More importantly, the local Bahadur representation forms the bedrock for deriving the strong approximation result for the functional conditional quantile process estimator. We first introduce the regularity conditions.

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

(i) $\zeta_{t\wedge T_m}$ 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 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 the Assumption 1, i.e., it requires the 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 the conditional density function at true state ζ_t evaluated at a neighborhood of quantile is positive and not too concentrate around the quantile, which holds if $f_t(\cdot)$ is continuous and has no point mass.[14] Condition (iv) is the same high-level requirement as in Assumption 1, which state the residual terms 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.[15]

[14] Observe that this requirement excludes the case where the random disturbance is discretely distributed. This is not surprising sine the classical Bahadur representation for i.i.d. data requires the absolutely continuous distribution. The analysis of sample quantiles of discretely distributed data deserves its own research.

[15] 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).

Specifically, in the special case when $\beta = 2$ and choosing $\chi = 1/2$, Shephard (2022) consider estimating the integrated volatility over the interval $[0, T]$ through the normalized sum of "in-fill median" in each block. The asymptotic properties of his estimator are derived via the monotonicity of the first order condition of the minimization problem. Although the integrated volatility estimator using median is asymptotically less efficient than realized variance in the Brownian motion case, it remains robust to the extreme returns which often arise when the price contains jump. As a complement to Shephard (2022), in this example, our focus is on the uniform inference for entire volatility process even in the case when $\beta < 2$, a setting wherein the conditional mean process becomes not well-defined and Assumption 1 (iii) no longer holds. Consequently, the return-based estimation procedure becomes invalid. Nevertheless, recall the state-space formation of Lévy-driven returns, it is evident that for all $t \in [0, T]$,

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

where $Q(L, \chi)$ denote the χ-quantile of $\varepsilon_{n,i} = \Delta_n^{-2/\beta}(L_{(i+1)\Delta_n} - L_{i\Delta_n})^2$, hence is free of nuisance. The proportional structure between $q(\chi)$ and σ suggests that conditional quantile process can serve as a feasible proxy for the volatility process. Note that formally defining the volatility process via moment conditions is challenging in this case whereas the interquantile range effectively capture the volatile level of the price. Besides the theoretical importance, empirical data has unearthed evidence suggesting that the stable index in cryptocurrency prices (e.g. BTC) is strictly smaller than 2, i.e., the price is driven by a pure jump process. Although for the case $\beta \neq 1$ the closed-form density of $\varepsilon_{n,i}$ is almost never known, we do have the explicit closed-form of the characteristic function. This facilitates the numerical computation $Q(L, \chi)$ and the validation of condition (ii) and (iii) in Assumption 2, see e.g., Zolotarev (1986).[16] Moreover, a similar argument as in the previous example yields that condition (iv) remains valid for all $0 < r < 1$. $\square$

Analogous to Theorem 1, we present Theorem 2 below, which states the strong approximation result for the 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*

[16] 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. Many numerical computation procedures and associated error bounds are discussed in Ament and O'Neil (2018).

$\chi \in (0,1)$, *for some positive constant* ϵ,

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

COMMENT. In contrast to Theorem 1, the coupling variable Z_j here is not pivotal and the variance matrix remains unknown, which is not surprising in quantile-related inference. This challenge 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 asymptotically valid critical value, as justified by Zuo (2015) who derive a Bahadur representation for the empirical bootstrap quantiles. We stress that under certain special scenarios, the distribution can indeed be pivotalized. For example, in the multiplicative transformation scenario (see Example 2) where the conditional quantile estimation is proved extremely useful, we have for all $t\in[0,T]$ that

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

where $\bar{f}(\cdot)$ denotes the density of $\Delta_n^{-2/\beta}(L_{(i+1)\Delta_n}-L_{i\Delta_n})^2$ which is free of nuisance hence can be computed numerically. Let $\hat{\nu}_{n,j}^2\equiv\chi(1-\chi)Q(L,\chi)^2\bar{f}\big(Q(L,\chi)\big)^2/\hat{q}_{n,j}^2$ 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 $\max_{1\le j\le m_n}|\hat{\nu}_{n,j}^2-\nu_j^2|=O_p(\Delta_n^{\rho/2}\sqrt{\log m_n})$. Consequently, let $cv_n(\alpha)$ be defined identically as in (2.4), denote

$$\widehat{B}'^{\pm}_{n,t}(\alpha)=\big(\hat{q}_{n,t}\pm cv_n(\alpha)\times k_{n,j}^{-1/2}\hat{\nu}_{n,j}\big)/Q(L,\chi),\qquad\text{for all } t\in\mathcal{T}_{n,j},\ \text{and } 1\le j\le m_n,\tag{2.5}$$

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

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

2.5 Application: Inference for Ranks

Given a path of certain stochastic process, the rankings of the values at a set of time points are often of great interest. Notably, the interest stems when the process indicates varying level of some signals, where quantifying these signals is challenging hence we are interested instead in their relative magnitudes. Such rankings illuminate which segments of the process possess comparatively higher signal strength in relation to others. the vigor of trading activities can shed light on the real-time information level that affects the market. Usually, the realized path is unobservable. Thus, the 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 the rankings of the true process. To illustrate this inherent uncertainty, consider a simple example where $\sqrt{k_n}(\hat{\mu}_{t_i} - \mu_{t_i}) \sim \mathcal{N}(0,1)$ for $i \in \{1,2\}$, then we have $\mathbb{P}(\hat{\mu}_{t_1} > \hat{\mu}_{t_2} | \mu_{t_1} < \mu_{t_2}) = 1 - \Phi\big(\sqrt{k_n}(\mu_{t_2} - \mu_{t_1})/2\big)$, 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) provide a comprehensive framework for inferring ranks via the introduction of confidence sets designated for the ranks. This methodology is congruent with our setting. Given a designated set of inspected time points, observing 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 the conditional mean process $(g_t)_{t\in[0,T]}$. Analogues results can be formulated for the conditional quantile process via the local Bahadur representation. 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 the rank of $(g_{n,j})_{1\leq j\leq m_n}$ and the entire rank vector as

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

The joint $(1-\alpha)$ confidence set for the 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 pairwise indices. Based on the insight of Theorem 3.4 in Mogstad et al. (2023), the confidence level of a joint confidence set for all the ranks is bounded below by one minus the *familywise error rate*, denoted as FWER_n, for testing the family of multiple hypotheses

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

Now, denote $\mathcal{S}_n^{\mathrm{all},-} \equiv \{(j,j') \in \mathcal{S}_n^{\mathrm{all}} : g_{n,j} \leq g_{n,j'}\}$, $\mathcal{S}_n^{\mathrm{all},+} \equiv \{(j,j') \in \mathcal{S}_n^{\mathrm{all}} : g_{n,j} \geq g_{n,j'}\}$, and the set of rejected hypotheses as $\mathrm{Rej}_n^-(j) \equiv \{1 \leq j' \leq m_n : H_{j,j'} \text{ is rejected}\}$, $\mathrm{Rej}_n^+(j) \equiv \{1 \leq j' \leq$

$m_n : H_{j',j}$ 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

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

The goal becomes finding 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 remainder of this subsection. Before presenting the procedure, we highlight that the setting in this subsection differs from that in Mogstad et al. (2023) in two aspects. First, while Mogstad et al. (2023) focus on the rankings across diverse populations, on the contrary, we consider ranks that defined for a single realized path of the investigated process at different time points. Consequently, the observations are no longer identically distributed, and the ranks Rank_n hence the sets $\mathcal{S}_n^{\mathrm{all},\pm}$ are random in nature. Second, the number of evaluated time points m_n increasing with n, contrasting with the case in Mogstad et al. (2023) where the total number of populations remains fixed.

For sake of notational simplicity, we assume in this subsection that $k_{n,j} = k_n$ for $1 \leq j \leq m_n$, i.e. we partition the 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 the test 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 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 $\widehat{D}_n \equiv \max_{(j,j') \in \mathcal{S}_n^{\mathrm{all}}} \hat{d}_n(j,j')$.[17] A direct application of Theorem 1 indicates a similar strong approximation result holds for $\widehat{D}_n$. Nonetheless, additional difficulty arises since the distribution of coupling variable is typically unknown. This stems from the fact the covariance matrix becomes non-identity and contains unknown parameters given that $\mathcal{S}_n^{\mathrm{all}}$ contains pairs with coinciding indices. In light of this, we propose the employment of a Gaussian multiplier bootstrap technique to determine the requisite

[17] 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 two, we use $\widehat{D}_n$ instead of $\widehat{D}_n'$ here.

confidence value. Namely, we generate i.i.d. standard Gaussian variables $(e_i)_{1\le i\le k_n}$ independent of $(Y_{i\Delta_n})_{1\le i\le n}$. 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 $(\hat{g}_{n,j}^{B})_{1\le j\le m_n}$. Then we compute the conditional $(1-\alpha)$ quantile of the maximum of studentized bootstrap statistics via

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

The following theorem provides the 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,*

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

(ii) $\mathbb{P}\big(\widehat{D}_n > cv_n^B(\alpha,\mathcal{S}_n^{\text{all}})\big) \ge 1-K\Delta_n^{\epsilon}$ if $\max_{(j,j')\in\mathcal{S}_n^{\text{all}}}(g_{n,j}-g_{n,j'}) \ge \underline{\beta}$ for some $\underline{\beta}>0$.

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 $\mathbb{P}(\text{FWER}_n) \le \alpha + K\Delta_n^{\epsilon}$. Furthermore, the theorem also shows the proposed test is consistent against any (non-local) alternatives. Lemma 5.1 in Chernozhukov et al. (2019) indicates, under the 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 $\max_{(j,j')\in\mathcal{S}_n^{\text{all}}}(g_{n,j}-g_{n,j'}) = o\big(\log(\Delta_n^{-1})^{1/2}\Delta_n^{\rho/2}\big)$.

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 too conservative in application. In the appendix, we prove that Theorem 3 remains valid even when $\mathcal{S}_n^{\text{all}}$ in the formulation of $\widehat{D}_n$ and $cv_n^B(\alpha,\mathcal{S}_n^{\text{all}})$ are replaced with any arbitrary set $\mathcal{S}_n\subseteq\mathcal{S}_n^{\text{all}}$ with $|\mathcal{S}_n|\ge 3$. This stronger result facilitates the incorporation of a stepdown improvement akin to those presented in Romano and Wolf (2005). The ultimate algorithm contains the following steps.

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

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 any $H_{j,j'}$ according to $\hat{\phi}_n^{(i)}(j, j') = \mathbb{1}\{\hat{d}_n(j, j') > cv_n^{(i)}\}$. For $1 \le j \le m_n$, form $\text{Rej}_n^{(i),-}(j)$ and $\text{Rej}_n^{(i),+}(j)$ as 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.

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

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

where $\text{Rej}_n^{\pm}(j)$ is computed according to Algorithm 1. Then $\widehat{\text{Rank}}_n \equiv \prod_{j=1}^{m_n} \widehat{\text{Rank}}_n(j)$ constitutes a joint $(1 - \alpha)$ confidence set for the ranks at all time points.

3 Monte Carlo simulations

3.1 Data Generating Processes

We conduct a Monte Carlo experiment to evaluate the performance of the proposed inference procedures. Our simulation is anchored in the setting of the motivating examples mentioned in subsection 2.2. In each example, the parameters used in the data generating processes (DGP) and sampling schemes are selected to closely resemble the empirical data encountered in the application.

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

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

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 the conditional mean process analyzed in subsection 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 limited. In contrast, DGP 2 introduces both heteroskedasticity in time and non-Gaussian disturbance. Regarding the Lévy driven returns discussed in Example 2, we simulate the 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 the Brownian motion W. Specifically, we focus on the following three data generating processes:

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

In forming the processes, we adopt a truncation technique analogous to the one employed in Bugni et al. (2023) for the stable distribution such that the normalized increment takes value in $[-30, 30]$ to avoid unrealistic price paths. The state process of interest is the variance process $(\sigma_t^2)_{t\in[0,T]}$, which is estimated through the the conditional mean process for DGP 3, or through the conditional median process (i.e. $\chi = 1/2$) analyzed in subsection 2.4 for DGP 4 and 5. Additionally, we focus on DGP 6 which serves as a representative illustration of the 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 is estimated through the conditional mean process.

Recall that we have two auxiliary processes μ and σ which serve as the 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 the 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,

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.

the mean process $\bar{\mu}$ and α are time variant and exhibit systematic moves in time, which the literature identifies as diurnal features. 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).[18] Moreover, the state processes under setting (b) are more volatile than those under the previous configuration, attributable to the smaller mean reverting parameters and larger variance magnitude. In summary, we have six types of DGPs in conjunction with two sets of parameter configurations. These 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 the 1-minute and 1-second data respectively. We stress that the 1-second sampling frequency is not practically feasible for DGP 3-5 to hold in reality, wherein the observed price in such high-frequency is contaminated by the so called microstructure noise, see e.g., the discussion in Zhang et al. (2005). Empirical evidence such as a signature plot of the realized volatility in relation to the sampling frequency shows the noise component overshadows when the 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 the returns. Conversely, given our application of DGP 6 in empirical illustrations wherein trading flow data is recorded at an ultra-high frequency and where approximations falter with coarser sampling frequency, for DGP 6 we exclusively consider 1-second data. The selection of the tuning parameter $k_{n,j}$ is described as follows. We partition the observations into equal-sized

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

blocks, i.e. $k_{n,j} = k_n$ for all $1 \leq j \leq m_n$. For the 1-minute data, we adopt $k_n \in \{10, 15, 30\}$, representing blocks of $\{10, 15, 30\}$ minutes respectively, the corresponding number of blocks is $m_n \in \{39, 26, 13\}$. For the 1-second data, we adopt $k_n \in \{120, 300, 600\}$, representing blocks of $\{2, 5, 10\}$ minutes respectively, the corresponding number of blocks is $m_n \in \{195, 78, 39\}$. 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, under different DGPs.

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. the data is observed every one minute, not surprisingly, the proposed confidence bands perform bad when the number of observation in each block is small, say $k_n = 10$, especially for DGP $2(a)$ and $2(b)$. This is particularly due to the poor approximation of Gaussian distribution for spot estimators in small sample. As k_n becomes larger, the coverage rate elevates remarkably. For instance, when $k_n = 30$, the coverage rates are above 80% for all DGPs with the exception of $2(a)$ and $2(b)$. Intriguingly, the 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 beneficial when the driving process of price data markedly deviates from Brownian motion. For a higher sampling frequency, $\Delta_n = 1/23400$, where the data is observed every one second, the coverage rates are above 85% for most DGPs when $k_n = 600$. There is a considerable increment in the time-variation effect of state processes within each block as block size expands. Notably, the coverage rates for DGPs equipped with parameter setting (b) are generally lower than the same DGP equipped with parameter setting (a) when k_n becomes larger. Drawing a parallel between the results for DGP $1(a)$ under column 4 and 7, both scenarios have a block length of 10 minutes and same number of blocks. Given the Gaussian nature of disturbance terms, each spot estimator maintains its Gaussianity in finite samples, the only difference lies in sampling frequency. There is a substantial improvement in convergence rate from $\Delta_n = 1/390$ to $\Delta_n = 1/23400$, indicating a finer sampling frequency can effectively take into account the time variation of state process.

In summary, the simulation results show that the proposed confidence bands aptly covers the true processes across all data generating processes aligned with an appropriate sampling frequency. Although under certain DGPs it appears to perform bad when the number of the observations in each block is insufficient, it can effectively addressed by adapting a larger block size and a finer

Table 2: Coverage Rate of Uniform Confidence Band

	$\Delta_n = 1/390$			$\Delta_n = 1/23400$		
DGP	$k_n = 10$	$k_n = 15$	$k_n = 30$	$k_n = 120$	$k_n = 300$	$k_n = 600$
$1(a)$	0.4739	0.6725	0.8257	0.8570	0.8907	0.8933
$1(b)$	0.4686	0.6787	0.8254	0.8528	0.8824	0.8834
$2(a)$	0.0830	0.2753	0.6326	0.5630	0.8115	0.8654
$2(b)$	0.0784	0.2894	0.6306	0.5509	0.7996	0.8580
$3(a)$	0.4733	0.6856	0.8311	–	–	–
$3(b)$	0.4756	0.6819	0.8339	–	–	–
$4(a)$	0.6972	0.7735	0.8147	–	–	–
$4(b)$	0.6981	0.7825	0.8044	–	–	–
$5(a)$	0.8941	0.9029	0.8916	–	–	–
$5(b)$	0.8914	0.9091	0.8912	–	–	–
$6(a)$	–	–	–	0.7610	0.8585	0.8868
$6(b)$	–	–	–	0.7643	0.8628	0.8782

Note: The table reports the coverage rates of a 90% 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 the 1-minute data, column 5-7 correspond to the 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.

sampling scheme. These observations stress that the proposed inference method remains robust in contexts analogous to the market settings. Moreover, in order to achieve better performance of the proposed inference procedure, one should employ the highest justifiable sampling frequency and opt the block size carefully in a suitable range to mitigate the time variation effect 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 led by the chair of the Federal Reserve, currently Jerome Powell, plays a pivotal role in disseminating the 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 the economy, its policy decisions, and the rationale behind those decisions. The statement provides insights into the 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., the analysis in Cochrane and Piazzesi (2002), Rigobon and Sack (2004), Bernanke and Kuttner (2005), and Nakamura and Steinsson (2018). On the other hand, Savor and Wilson (2014), Lucca and Moench (2015), and Bollerslev et al. (2021) also find evidence of pre-announcement effects of the initial release. With more accurate estimation of volatility, Bollerslev et al. (2022) finds that the announcement of new policy decision may not cause the most substantial shock during the FOMC days, especially when the decision has been well anticipated by the market. In that case, the 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 the FOMC statement release, the Federal Reserve 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 the chair to elaborate on the FOMC's decision-making process, provide additional context, and address questions from media. It allows for a more in-depth discussion of the committee's views on the economy and financial policy. During the press conference, the chair inevitably reveals some (possibly subtle) forward guidance, i.e., the information about the expected path of monetary policy in the future. This guidance 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 the central bank's future actions.

Pinpointing the exact sentence in the press conference that provides additional information and contains forward guidance, however, is a challenging task. Each sentence in the press conference is typically spoken within a few seconds, making it challenging to precisely associate specific sentences with changes in volatility. The rapid succession of sentences and the limited time span of each sentence can make it difficult to isolate their individual impact on market volatility. Analyzing volatility changes at the sentence level requires examining high-frequency data, such as tick-by-tick price movements. Ultra-high-frequency price data is often subject to microstructure noise, which can distort the identification of precise volatility patterns associated with specific sentences. Noise in the price data makes it harder to attribute volatility changes solely to the content of individual sentences. To mitigate the impact of noise on volatility analysis, existing procedures such as Jacod et al. (2009) often employ wider time windows when studying high-frequency data. However, using wider windows can make it more challenging to detect the specific volatility patterns related to the individual sentences.

Utilizing textual analysis on the conference scripts is another approach to studying the FOMC press conferences. With the developing natural language processing (NLP) methods, these textual analysis algorithms have found prevalent application in economics and finance, as discussed in the review articles Gentzkow et al. (2019) and Loughran and McDonald (2020). Nonetheless, in the formal announcing scenario like the FOMC meetings, the conventional NLP method based on experiences might exhibit considerable inaccuracies. To better understand the possible limitation of pure textual analysis, we deploy a 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, an expansive language model pioneered by OpenAI, to extract several features that could be essential signals of a high level of forward guidance.[19] Here is a brief overview of factors that the algorithm takes into account:

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

[19]ChatGPT 3.5 was trained with data up to September 2021, hence has no knowledge beyond that cutoff. This ensures that the identified features are intrinsically rooted in NLP proficiency without any “sneak peak” at contemporaneous market activities.

Level of Detail: Sentences that provide specific details about future policy actions are given higher scores. This could include the announcement of specific interest rate changes, plans for balance sheet reduction, or discussions about future meetings where policy changes might be discussed.

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 algorithm also considers the overall context of the sentence and how it fits within the broader speech or statement. It looks for patterns and consistency in the language used to convey future policy intentions.

Quantitative and Qualitative Aspects: The algorithm considers both quantitative aspects (e.g., specific percentages or values) and qualitative aspects (e.g., intentions, expectations) when assessing forward guidance.

Comparative Analysis: The algorithm compares the sentence with other sentences in the text to ensure a relative ranking of forward guidance strength. This takes into account the range of guidance provided throughout the text.

The algorithm then computes a weighted average of the scores of the aforementioned aspects. Note that this algorithm is designed to identify potential forward guidance based on linguistic patterns and context, and the scores are indicative rather than definitive. The assessment also accounts for variations in language and communication styles, so the scores may represent a nuanced interpretation of forward guidance strength in the given contest. Based on this algorithm, we can classify the sentences of 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 on the press conference speeches on the eight 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 the speeches has been designated to carry medium

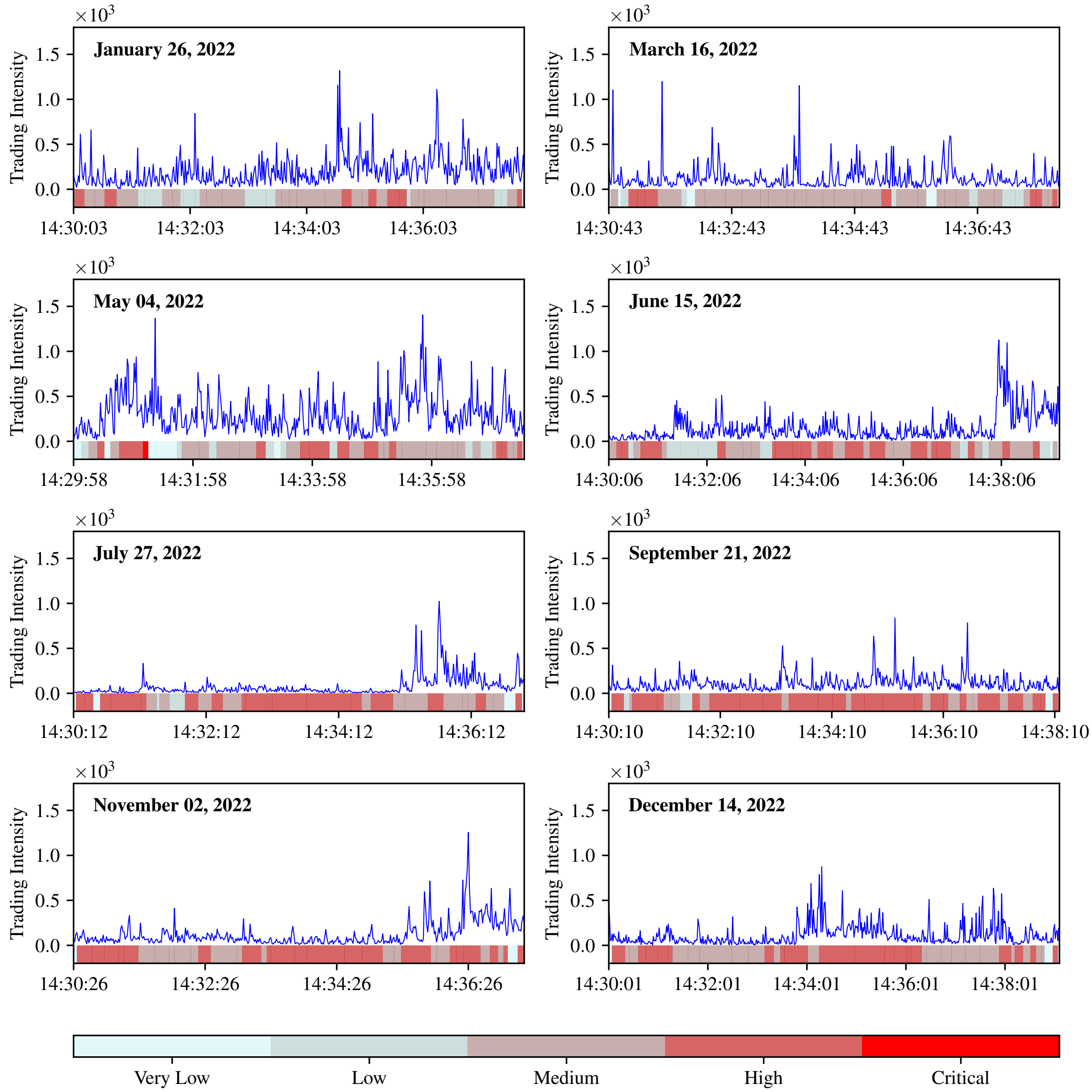

Figure 2: **Trading Intensities and Relative Information Levels during FOMC Press Conference Speeches**. The figure plots the one-second trading intensity during the eight FOMC press conference speeches. The horizontal axis is colored according to the relative information level embedded in the potential forward guidance contained in each sentence, which is computed using the algorithm described in this section. The color follows the rule $\text{RGB}\alpha = \big(s/5, 1 - s/5, 1 - s/5, (s/5)^{1.25}\big)$ where s denotes the information level in the scale of 1 to 5, with 1 being "very low," 5 being "critical."

and high level of information. To gain a direct insight of the accuracy of this procedure, we mark the relative information level in the time line, along with the estimated trading intensity to conduct a visual comparison. For the intensity, we use the nanosecond tick-by-tick data of S&P 500 ETF (ticker: SPY), downloaded from Trade and Quote (TAQ) database. We estimate the second-level trading intensity during each FOMC press conference speech, i.e. $\Delta_n = 1/(2.34 \times 10^{13})$, $k_n = 10^9$ so that $k_n\Delta_n = 1\,\mathrm{sec}$ corresponds to one-second block. In Figure 2, we plot the estimated trading intensities during the press conference speeches, and colored the horizontal line in the gradient spectrum such that the sentences with lowest information level (i.e., labeled "very low") tends to be transparent green, where the sentences with highest information level (i.e., labeled "critical") tends to be red.

Next, we delve deeper into the textual analysis outcomes, exploring the trading intensities across the categorized groups. Considering the potential reactive latency between the information arrivals and the correspondent trading actions, we shift the observation window 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 the medians and means with each group. We further conduct Welch's t-tests to test if sentences identified with a higher information level truly exhibit an elevated average trading intensity. The results indicate that, even under the best case (a 14 seconds lag), where the group labeled "critical" has significantly higher intensity than other four groups, we still cannot conclusively negate the possibility of no significant distinctions among the other four groups.

The main inherent challenge of the pure textual analysis approach stems from the carefully crafted nature of the speech scripts and the potential overlap 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, since the scripts are designed to convey information while maintaining stability and avoiding unnecessary shocks. Moreover, the speech part of press conferences can 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 the the similarity by computing Jaccard similarity

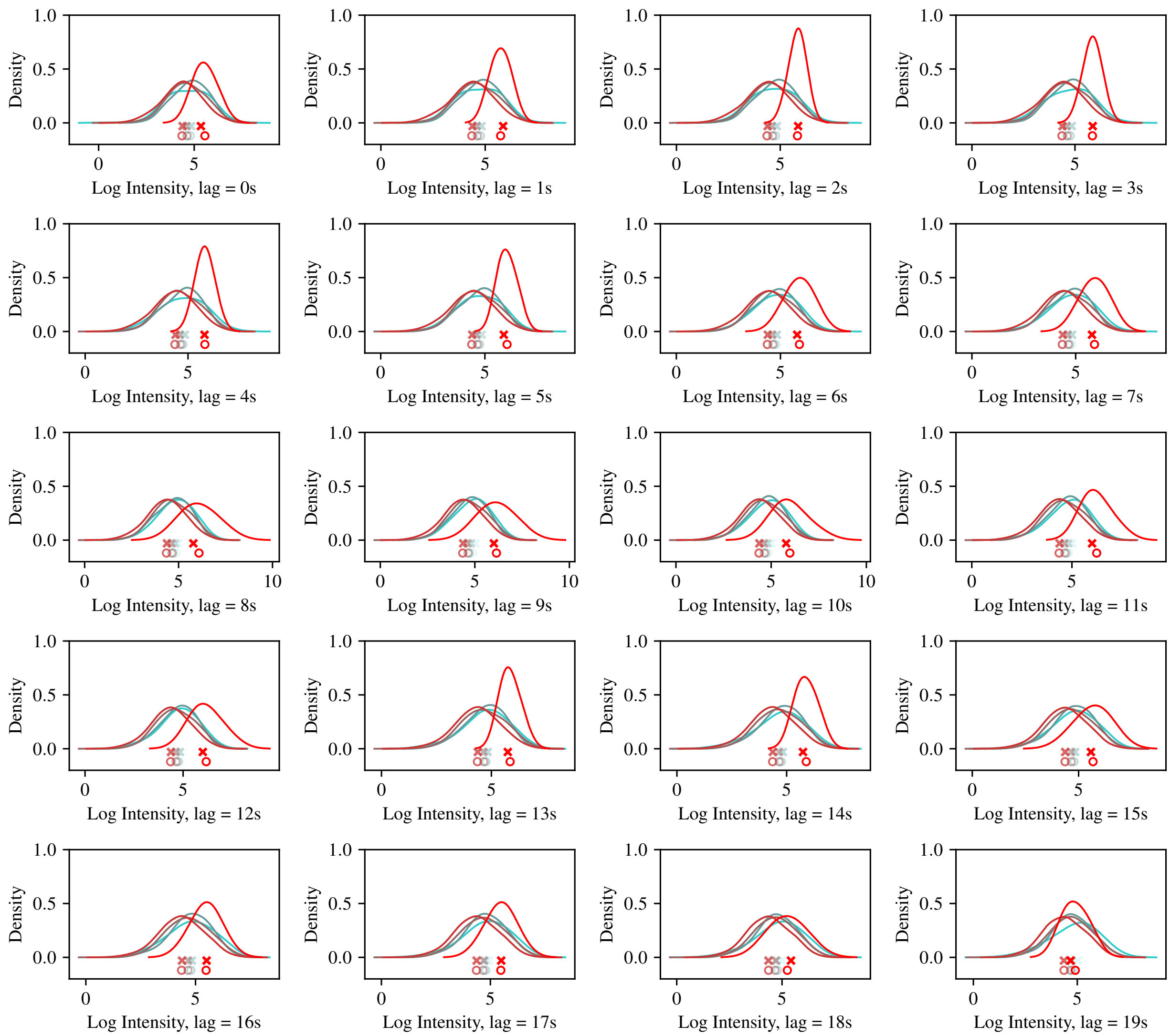

Figure 3: **Distribution of Intensity with Different Information Levels**. The figure plots the kernel density estimation of the logarithm trading intensities with different relative information level embedded in the potential forward guidance contained in each sentence, which is computed using the algorithm described in this section. In each panel, we shift the window by several seconds to cover the effect of reaction time between the information arrivals and tradings. The color of each line is the same as in Figure 2, the median and the mean of each group are marked in $\times$ and $\circ$ sign respectively.

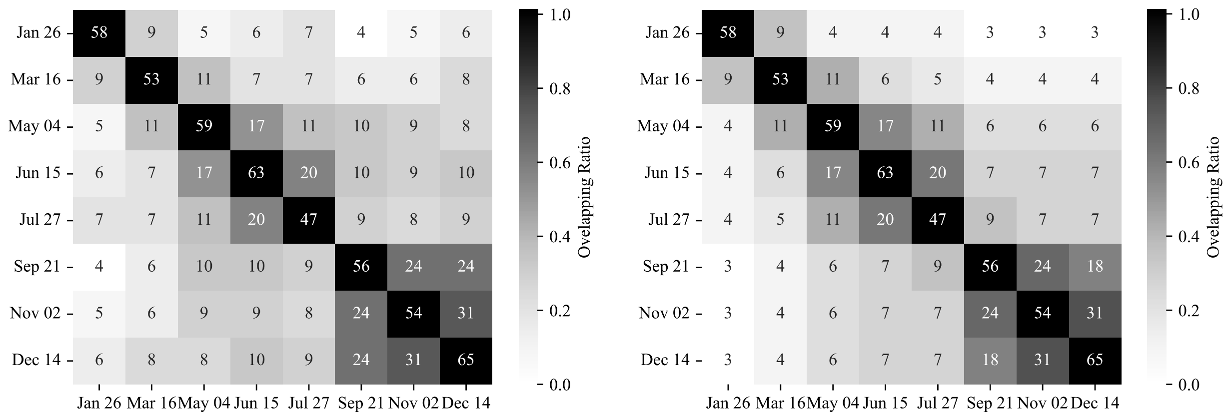

Figure 4: **Similarity of FOMC Press Conference Speeches**. The figure plots the overlapping ratio between different speeches. The overlapping ratio is defined as the log $S_{i,j}^{\text{type}}$, where type $\in \{\text{pw}, \text{cm}\}$. In the left panel, $J_{i,j}^{\text{pw}}$ is the pairwise Jaccard 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 sentences between speeches within $\{t_i, \ldots, t_j\}$ divided by the total number of sentences. The exact numbers of pairwise and cumulative over-lapping sentences are shown in each square. All the ratios are normalized to $[0, 1]$.

coefficient between these sets, as proposed by Jaccard (1912),[20]

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

The repetition of certain phrases or topics will diminish their impact on market expectations over time. Textual analysis techniques that focus solely on keyword detection might identify familiar terms without considering the 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 group the 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 the ranks of all second-level intensities. Based on this results, we can partition each speech into groups $G \in \{1, \ldots, \overline{G}\}$ using the following algorithm: First, we permute the index 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 $\widehat{R}_n(\pi(j+1)) \cap \widehat{R}_n(\pi(1)) \neq \varnothing$, then $\pi(j+1)$ belongs to the same group as $\pi(1)$; otherwise, $\pi(j+1)$

[20] 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 the speeches under consideration do not exhibit significant difference.

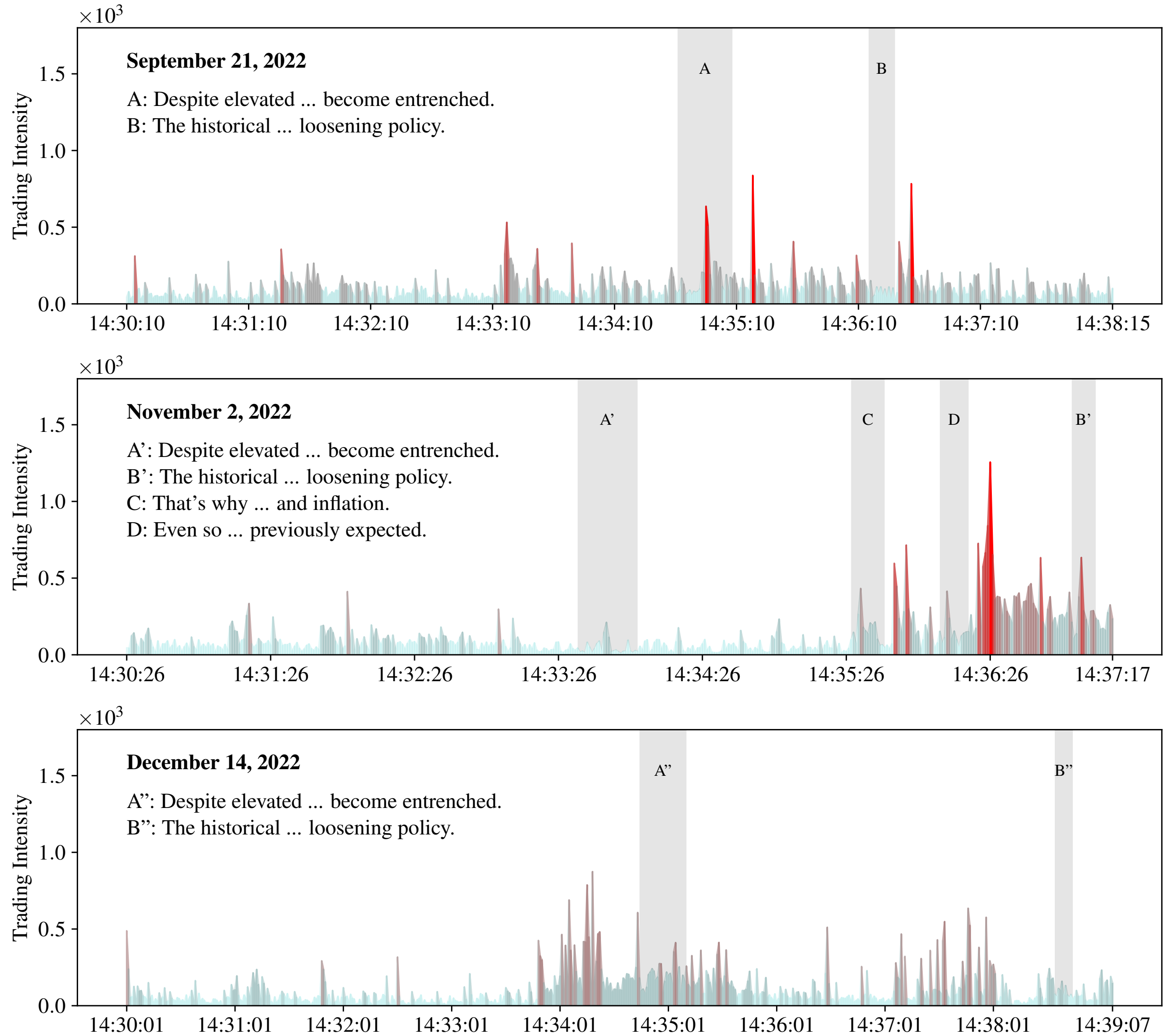

Figure 5: **Trading Intensity during FOMC Press Conference Speeches**. The figure shows the heatbar of estimated trading intensities during the FOMC press conference speeches on September 21, November 2, and December 14 in 2022 from the top panel to the bottom panel, respectively. On each day, a 90% confidence set of joint ranks is constructed using the multiple testing procedure proposed in section 2.5. Further, each speech was partitioned into groups and colored using the strategy described in the main text. The heatbar is colored according to the 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 is the same as in Figure 2. The duration time of target sentences are shaded in light gray in each panel.

initiates a new group $G + 1$. Repeats until the last second $j = m_n$. In Figure 5, we present a heatbar of speech according to the trading intensity and color it in the same way such that the group with lowest intensity tends to be transparent light green, the group 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 the speeches actually impart minimal information, as evinced by low trading intensities. Most of them are recurring sentences across consecutive speeches, which theoretically, should not disseminate any novel information after their inaugural mention. Meanwhile, we detected more sentences that were markedly informative, eliciting high trading intensities. Most of them do not contain the keywords or phrases, suggesting that even subtle tonal shifts on certain topics may result in a substantial market effect.

In conclusion, the conventional textual analysis procedure tends to overstate the information level of individual sentences, and in the meantime fail to accurately identify the most informative parts. The proposed intensity-based analysis offers a compliment to textual analysis, one can refine the analysis procedure by deploying a supervised learning, i.e., utilizing the intensity-level-labeled text to train algorithms in order to classify the text more accurately.

4.2 Case Study

Next, we conduct a case study to better illustrate the preceding findings, opting for specific sentences from the speeches that stand out as high level of information about forward guidance and followed with considerable intensity spikes. The first sentence is the shift in tone about longer-term inflation expectations which presents a double twist, first mentioned during 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 the 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 notable 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, the aforementioned sentences [A] and [B] recur in both the November and the December conference. On the contrary, these repetitions do not elicit similar spikes in intensity. In fact, the bottom panel of Figure 5 indicates an 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 the content from the preceding conference. Above observation coincides with our intuition that new information only when it is introduced for the first time. After the reaction, the market quickly accept it and subsequent repetitions of the same sentences are lack of novelty.

During the September conference, inquiries emerged concerning the 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, the Fed incorporate specific remarks about such lags in both the official statement and the subsequent 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 several seconds after sentence [C]. After mentioning the 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 the 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.

5 Concluding Remarks

We introduce a valid methodology for conducting inference on a non-linear, non-Gaussian continuous-time state-space model over a fixed time span. Through the integration of a residual term, we allow the model to be "approximately Markovian," which accommodates the Lévy driven returns and Cox trading flow process. We allow the state process to have undefined dynamics, and propose a uniform inference procedure for the entire conditional mean process and the entire conditional quantile process of the transformed state.

To construct functional estimators for the investigated processes, we collect the blockwise estimates, where each block size shrinks to zero. The challenge of achieving uniform inference for these functional estimators arises from its non-Donsker nature. To address this, we establish Gaussian strong approximation that facilitates valid uniform inference. The proposed framework can be used to tackle other economics problems like constructing confidence set for ranks of spot values of certain process.

We apply the proposed inference procedure to analyze the trading flow process and detects the 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. This serves as a compliment to existing methodologies, like the volatility-based detection mechanisms and conventional textual analysis tools.

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 distinguish between them, we use M to denote some positive constant defined in the content 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_t 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_t globally takes values in the compact set $\mathcal{K}$ and is 1/2-Hölder continuous under the L_p norm. 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 notation. For $1 \leq j \leq m_n$ and $1 \leq i \leq k_{n,j}$, denote

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

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 $G_2(\zeta_{\tau(i,j)}) - g^2_{\tau(i,j)}$. 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:

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

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

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

Step 3 concludes the proof by establishing the asserted statement.

Step 1. Recall the model (2.1), 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 $|(a-b)/c - a/d| \leq |d/c - 1| \times |(a-b)/d| + |b/d|$. 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 $1 \leq j \leq m_n$, $\mathfrak{A}_{n,j} \equiv \mathfrak{A}^{(I)}_{n,j} + \mathfrak{A}^{(II)}_{n,j}$ with

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

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 $k_{n,j} \asymp \Delta_n^{-\rho}$ imply 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 $m_n \asymp \Delta_n^{\rho-1}$, 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

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

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 $1\leq i\leq \bar{k}_n$,

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

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

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

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 $x>0$,

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

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 $1 \le i \le \bar{k}_n$ and $1 \le j \le 2m_n$, we define $\widetilde{U}^{\dagger}_{i,j}$ as

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

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

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

Recall that $(\widetilde{U}_{i,j})_{1\le i\le \bar{k}_n, 1\le j\le m_n}$ are $\mathcal{F}^{(0)}$-conditionally independent, centered random variables. Let $(\widetilde{Z}_{i,j})_{1\le i\le \bar{k}_n, 1\le j\le m_n}$ 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 $1 \le i \le \bar{k}_n$ and $1 \le j \le 2m_n$, let

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

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 $1 \le i, i' \le \bar{k}_n$ and $1 \le j \le 2m_n$. 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 $1 \le j \le 2m_n$,

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

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

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

For $1 \le j \le m_n$, 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, $(Z_j)_{1\leq j\leq m_n}$ remains standard Gaussian unconditionally, hence satisfies the requirement in the assertion. Note that by construction we have

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

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

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

where the last term is bounded by $K\Delta_n^{\epsilon}$ using the anti-concentration inequality (see Corollary 2.1 in Chernozhukov et al. (2015)) together with the fact that

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

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

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

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_t takes value in $\mathcal{K}$; for all $s, t \in \mathcal{T}_{n,j}$ where $1 \leq j \leq m_n$, and for each $p > 0$, $\mathbb{E}[\|\zeta_t - \zeta_s\|^p] \leq K_p|t-s|^{p/2}$ for some constant K_p; (ii) for each $x \in \mathbb{R}$, for all $z, z' \in \mathcal{K}$, $|F(z,x) - F(z',x)| \vee |\partial_x F(z,x) - \partial_x F(z',x)| \leq K\|z - z'\|$; (iii) for each $t \in [0,T]$ and x in some neighborhood of q_t, $f_t(x) + f_t^{-1}(x) + |\partial_x f_t(x)| < K$; (iv) $\max_{1\leq i\leq n} |R_{n,i}| = o_p(\Delta_n^r)$ for some $r > 0$.*

The proof of Theorem 2 is based on a local Bahadur type representation of sample quantiles, where the approximation error can be controlled uniformly, as shown in the following lemma.

Lemma A.1 (Local Bahadur Representation). *Suppose Assumption A.2 holds. For $1 \le j \le m_n$, denote*

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

Then we have for some positive ϵ and γ,

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

Proof of Lemma A.1. We prove the validity of the assertion for all positive ϵ satisfying

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

Correspondingly, let γ be positive constant satisfying

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

Denote $\widetilde{Y}_{i,j} \equiv \mathscr{Y}(\zeta_{\tau(i,j)}, \varepsilon_{n,\iota(i,j)})$, within each block j re-index the sequence $(\widetilde{Y}_{i,j})_{1\le i\le k_{n,j}}$ in the non-decreasing order and denoted as $\widetilde{Y}^o_{1,j} \le \cdots \le \widetilde{Y}^o_{k_{n,j},j}$. 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 $\widetilde{Y}^o_{\lceil k_{n,j}\chi\rceil,j} \le Y^o_{\lceil k_{n,j}\chi\rceil,j} + \max_{i\in\mathcal{I}_{n,j}}|R_{n,i}|$. 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 $\widetilde{Y}^o_{\lceil k_{n,j}\chi\rceil,j} \ge Y^o_{\lceil k_{n,j}\chi\rceil,j} - \max_{i\in\mathcal{I}_{n,j}}|R_i|$. Therefore, assumption A.2(iv) implies that

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

For each $1 \le j \le m_n$, let $\widetilde{F}_{n,j}(x) \equiv k_{n,j}^{-1}\sum_{i=1}^{k_{n,j}} \mathbb{1}\{\widetilde{Y}_{i,j} \le x\}$ be the empirical distribution function of $(\widetilde{Y}_{i,j})_{1\le i\le k_{n,j}}$. 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 $1 \le j \le m_n$. 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 for all $1 \le j \le m_n$. Step 3 derives the asserted statement.

Step 1. For $1 \le j \le m_n$, denote

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

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 $x \in I_{n,j}(\ell)$,

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

where

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

Similarly, we also have $S_{n,j}(x) \ge S_{n,j}\bigl(\psi_{n,j}(\ell)\bigr) - \vartheta_{n,j}(\ell)$. Denote $\bar{\vartheta}_{n,j} \equiv \max_{-\varkappa_{2,n} \le \ell \le \varkappa_{2,n}-1} \vartheta_{n,j}(\ell)$. Then it follows the definition of $\overline{I}_{n,j}$ that

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

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

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

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

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

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

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

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

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

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

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

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

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

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,

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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 $(\hat{q}_{n,t})_{t\in[0,T]}$. Similar as 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 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

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

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

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

Let $\bar{k}_n \equiv \max_{1\le j\le m_n} k_{n,j}$, then $\bar{k}_n \asymp \Delta_n^{-\rho}$ and $1/K \le \bar{k}_n/k_{n,j} \le K$ uniformly for all $1\le j\le m_n$. For each $1\le i\le \bar{k}_n$ and $1\le j\le m_n$, define $\widetilde{U}$ and $\nu_{i,j}$ as follows:

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

By construction the variables $\widetilde{U}_{i,j}$ are $\mathcal{F}^{(0)}$-conditionally independent across different values of $1 \le i \le \bar{k}_n$ and $1 \le j \le m_n$ with mean zero and conditional variance $\tilde{\nu}_{i,j}^2$. Note that

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

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

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

We can thus rewrite

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

Let $(\widetilde{Z}_{i,j})_{1 \le i \le \bar{k}_n, 1 \le j \le m_n}$ 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 $1 \le i \le \bar{k}_n$ and $1 \le j \le 2m_n$, let

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

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 $1 \le i, i' \le \bar{k}_n$ and $1 \le j \le 2m_n$. 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

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

For $1 \le j \le m_n$, 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 $1 \le j \le m_n$, hence

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

Also note that by construction we have

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

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

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

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

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

Similarly, we can show

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

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 the theorem holds for all $\mathcal{S}_n\subset\mathcal{S}_n^{\mathrm{all}}$ with $|\mathcal{S}_n|\geq 3$. Let $\mathcal{G}_n\equiv\mathcal{F}^{(0)}\vee\sigma(Y_{i\Delta_n}:1\leq i\leq n)$ denote the smallest σ-algebra contains $\mathcal{F}^{(0)}\cup\sigma(Y_{i\Delta_n}:1\leq i\leq n)$. Also we strengthen Assumption 1 to Assumption A.1 by a using of Localization procedure. We prove the assertions of 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 notation. 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 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 $|\overline{D}_n - \widetilde{D}_n|$,

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

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

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

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

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

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 $\epsilon/\rho \leq \varpi < 1 - 2\gamma/\rho$,

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

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

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

where for $(j,j') \in \mathcal{S}_n$, $\mathfrak{C}_n(j,j') \equiv k_n^{-1/2}\sum_{i=1}^{k_n} e_i(R_{n,\iota(i,j)} - R_{n,\tau(i,j')})/\varsigma_n(j,j')$. 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

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

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

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

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

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

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 $(j,j'),(\ell,\ell')\in\mathcal{S}_n$,

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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 $\max_{(j,j')\in\mathcal{S}_n}\bar{\mu}_n(j,j')\ge\underline{\beta}>0$. Combining with (A.45) and (A.47) gives $\mathbb{P}(\overline{D}_n-\widehat{D}_n+\Delta_n^{-\rho/2}\underline{\beta}>\varrho_n/2)\le K\Delta_n^{\epsilon}$. Therefore, we have

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

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

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

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

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

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.

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

Aoki, M. and A. Havenner (1991): "State space modeling of multiple time series," *Econometric Reviews*, 10, 1–59.

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*.

Bauer, D. and M. Wagner (2012): "A state space canonical form for unit root processes," *Econometric Theory*, 28, 1313–1349.

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.

Bollerslev, T., J. Li, and Q. Li (2022): "Optimal Nonparametric Range-Based Volatility Estimation," *Available at SSRN 4129450*.

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.

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.

Durbin, J. and S. J. Koopman (2001): *Time series analysis by state space methods*, Oxford University Press.

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.

Harvey, A. C. (1989): "Forecasting, structural time series models and the Kalman filter," .

Harvey, A. C., S. J. Koopman, and N. Shephard (2004): *State space and unobserved component models: Theory and applications*, Cambridge University Press Cambridge.

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.*

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.

Kalman, R. (1960): "A new approach to linear filtering and prediction problems," *Transactions of the ASME–Journal of Basic Engineering*, 82, 35–44.

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

Kim, C.-J., C. R. Nelson, et al. (1999): "State-space models with regime switching: classical and Gibbs-sampling approaches with applications," *MIT Press Books*, 1.

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.

Koenker, R. and G. Bassett Jr (1978): "Regression quantiles," *Econometrica: journal of the Econometric Society*, 33–50.

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.

Lee, S. S. and P. A. Mykland (2008): "Jumps in financial markets: A new nonparametric test and jump dynamics," *The Review of Financial Studies*, 21, 2535–2563.

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.

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.

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.

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.

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.

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.

West, M. and J. Harrison (1997): *Bayesian forecasting and dynamic models*, Springer Science & Business Media.

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

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.