Bridging Dense and Sparse Models in High-Dimensional Quantile Regression *

Yaping Wang†

This version: Nov 8, 2025
Click here for latest version

Abstract

This paper introduces a high-dimensional quantile regression that bridges the dense and sparse modeling perspectives by allowing conditional quantiles to depend densely on latent factors capturing pervasive comovements and sparsely on idiosyncratic components reflecting heterogeneous, localized shocks. The resulting framework combines the interpretability and variable selection advantages of sparse models with the stability and dimension reduction of factor models. Theoretically, we establish convergence rates for the proposed estimator under weak temporal dependence and allow for both strong and weak factors. Simulation studies demonstrate favorable finite-sample performance and highlight a trade-off under weak factors, where the need to retain idiosyncratic components increases as the precision of their estimation deteriorates. In an empirical application to a large macro financial panel, the estimator achieves lower check loss than sparse quantile regression and factor only specifications, with the largest gains in the lower tail.

Keywords: Quantile regression, factor model, regularized regression, time series

JEL Classification: C21, C22, C38, E44, C55

*Acknowledgments: I am deeply grateful to my advisor Christian Brownlees for his invaluable guidance and support. I also thank David Rossell, Gábor Lugosi, Kirill Evdokimov, Yeming Ma, Carlo Pavanello, and the participants of the XVth Workshop in Time Series Econometrics, 2024 Asian Meeting of the Econometric Society, UPF Internal Statistics Seminar, and BSE PhD Jamboree for their insightful comments and discussions. All errors are my own.
† Department of Economics and Business, Universitat Pompeu Fabra and Barcelona School of Economics; E-mail: yaping.wang@upf.edu.

1 Introduction

Macroeconomic and financial forecasting is often about *risk*: the upper tail of inflation, the onset of financial crises, and extreme losses in returns. These objects live in *quantiles*, not in the mean. Quantile regression (Koenker and Bassett, 1978) provides a well-suited framework to model conditional quantiles and remains robust to outliers and heavy-tailed disturbances that are pervasive in macro-financial data. One prominent policy application is the recent "at-risk" literature (e.g., Adrian, Boyarchenko, and Giannone, 2019; Ferrara, Mogliani, and Sahuc, 2022; Plagborg-Møller, Reichlin, Ricco, and Hasenzagl, 2020; Gelos, Gornicka, Koepke, Sahay, and Sgherri, 2022), but quantile methods have long been used more broadly, including distributional and inequality analysis (e.g., Machado and Mata, 2005), systemic risk measurement (Adrian and Brunnermeier, 2016), and tail modeling in asset returns(e.g., Engle and Manganelli, 2004).

High-dimensional macro-financial data, however, pose a challenge that standard quantile methods do not resolve. In large datasets, most predictors co-move because a few pervasive forces such as monetary policy, business cycles, and financial conditions shift many series together. We refer to this as a *dense* factor structure, following the dense-sparse classification discussed in the linear regression literature (e.g., Giannone, Lenza, and Primiceri, 2021): a small number of latent factors load on many predictors, so many coefficients are nonzero, typically with modest magnitudes. At the same time, extreme yet economically salient shocks could strike only a narrow set of variables and especially affect the tails. This is a *sparse*, localized channel, where only a small subset of predictors has nonzero effects. In practice, both forces coexist and matter.

Existing approaches each miss one side of this reality. Sparse high-dimensional quantile regressions, such as the ℓ_1-penalized quantile regression of Belloni and Chernozhukov (2011), deliver interpretability through variable selection. However, they rely on stringent assumptions, such as weak correlation among predictors or restricted eigenvalue conditions. In the presence of the latent factors prevalent in economics and finance, these assumptions are often violated, leading to biased estimation or unstable selection across samples; see Fan, Ke, and Wang (2020) and Fan, Lou, and Yu (2023a) for related discussions. At the other end, factor and low-rank approaches summarize pervasive movements, but they typically treat most variation as common and wash out precisely those idiosyncratic signals that move the tails, making it difficult to select and interpret idiosyncratic predictors; see Chen, Dolado, and Gonzalo (2021) for quantile factor models and related low-rank approaches. This problem is exacerbated under *weak factors*,[1] widely documented in the literature (e.g., Onatski, 2012; Uematsu and Yamagata, 2022), where common signals are easily mistaken for idiosyncratic signals.

This paper addresses the tension by developing a *factor-augmented sparse quantile regression* that jointly models these two forces. The conditional quantile of the target variable is allowed to depend *densely* on a few latent factors that drive pervasive co-movement across predictors and *sparsely* on the idiosyncratic components that capture heterogeneous, localized effects. Coefficients are *quantile-specific*, allowing the roles of common and idiosyncratic channels to vary across different parts of the conditional distribution, in line with existing empirical evidence (Ando and Bai, 2020; Adrian *et al.*, 2019).

Methodologically, we estimate the quantile-specific coefficients on both channels in two steps: (i) extract latent factors and idiosyncratic components from a high-dimensional predictor set via principal component analysis (PCA); and (ii) run a quantile regression on both parts and apply an ℓ_1-penalty exclusively to the idiosyncratic block. Splitting the original high-dimensional predictors into two parts delivers two benefits. First, it stabilizes sparse selection at quantiles when restricted eigenvalue conditions fail for the original full design but hold for the idiosyncratic block. Second, by accounting for unobserved common components before selection, it alleviates the omitted variable bias that arises when latent factors are ignored in sparse quantile regressions (see Giglio and Xiu (2021) for an analogous discussion in linear factor models). This "bridge" therefore combines the stability and dimension reduction of dense models with the interpretability and selection

[1] We use "weak factors" in the economic sense that common components explain only a modest share of cross-sectional variation; formal definitions and assumptions appear in Section 3.1.

of sparse models.

The analysis couples a non-smooth quantile loss with weak temporal dependence and *generated regressors* (estimated factors and idiosyncratic components), requiring us to track rotational indeterminacy and first-stage estimation error throughout the penalized fit. We also consider settings with weak factors. Our theory maps the boundary between strong and weak factor regimes and quantifies when the bridge outperforms dense-only and sparse-only procedures.

Theoretically, the paper establishes nonasymptotic bounds for the proposed estimator. We show that the estimation error bound admits a decomposition with a leading component that coincides with the rate one would obtain if the latent factors and idiosyncratic components were observed, and additional components that quantify the effect of factor estimation and rotational indeterminacy. This decomposition reveals how time dependence, factor strength, and cross-sectional dimensionality jointly determine the statistical difficulty of the problem. When common signals are strong and the cross-sectional dimension grows sufficiently quickly, the factor estimation components are of smaller order and the bound is dominated by the oracle term. Under weak factors, these components can be first order, generating a clear bias-variance trade-off that does not appear in purely dense or purely sparse methods. The analysis therefore clarifies the boundary between regimes and explains when combining sparse idiosyncratic structure with dense latent components improves quantile forecasting in high dimensions.

1.1 Related Literature

This paper relates to several strands of literature.

First, this paper adds to the growing literature on high-dimensional quantile regression and regularized estimation methods. Following Belloni and Chernozhukov (2011), work on high-dimensional quantile regression has advanced penalties, computation, and post-selection inference, including concave penalties (Wang, Wu, and Li, 2012), adaptively weighted ℓ_1 procedures and uniform selection across quantile levels (Zheng, Peng, and

He, 2015), and smoothing/debiasing schemes (Tan, Wang, and Zhou, 2021; Yan, Wang, and Zhang, 2023; Zheng, Peng, and He, 2018). This literature typically assumes i.i.d. samples with fully observed regressors. We differ by using generated regressors under weak temporal dependence and by explicitly accommodating latent factor structures.

Second, this paper contributes to the literature on weak factors. Existing literature has documented and formalized the presence of weak factors in macro-financial data. Onatski (2012) develops inferential tools that distinguish strong from weak factors, and subsequent work further analyzes estimation and inference when common signals are weak and close to the idiosyncratic spectrum (e.g., Wang and Fan, 2017; Bai and Ng, 2023; Uematsu and Yamagata, 2021, 2022). This line of work clarifies how weak factors affect factor recovery, identification, and testing. What remains relatively unexplored, however, is the role of weak factors when they are used as *regressors or predictors* in high-dimensional quantile settings such as ours. An exception in a different spirit studies principal components from a predictive perspective (e.g., Brownlees, Guðmundsson, and Wang, 2024), but not jointly with a sparse idiosyncratic channel under a quantile loss.

Third, our setting involves *estimated* latent factors and idiosyncratic components entering the quantile regression, which places it within the broader class of two-step procedures with generated regressors. While the classical literature provides general results for smooth objectives and nonparametric first-stage objects (e.g., Newey, 1984; Pagan, 1984), analogous theory for non-smooth losses such as the quantile check function in high dimensions remains limited. We instead give finite-sample bounds that propagate factor estimation noise and rotational indeterminacy through the ℓ_1-penalized quantile objective under weak temporal dependence.

Fourth, this paper contributes to the literature on factor-augmented sparse regression. Kneip and Sarda (2011) were among the first to propose a factor-augmented sparse *linear* regression, in which principal components extracted from the same high-dimensional predictor set are included as additional regressors in an augmented regression model. Their framework, however, relies on strong assumptions such as Gaussian errors and uncorrelated idiosyncratic components. Recent work by Fan *et al.* (2023a) relaxes these conditions and develops a general inferential framework for high-dimensional factor-augmented linear regressions. Fan *et al.* (2020) study a simplified version of the Fan *et al.* (2023a) model with a focus on selection consistency, whereas Fan, Masini, and Medeiros (2023b) analyze a related model in a panel setting with an emphasis on prediction. Comprehensive discussions of this line of research can be found in Fan *et al.* (2023a). This paper complements these studies by extending the bridge idea from *mean* to *quantile* under weaker regularity. Indeed, Ando and Tsay (2011) also study a form of factor-augmented quantile regression, but our framework contrasts sharply with theirs. They augment a low-dimensional covariate set with factors estimated from a separate set of predictors, whereas our model augments the same high-dimensional predictor set with its own estimated factors and in a way includes both the factors and the idiosyncratic components as regressors.

Finally, this paper also speaks to the literature on quantile factor models and low-rank approaches to heterogeneity. These studies aim to capture common patterns in conditional quantiles across cross-sectional units. Some are fully factor driven, such as Chen *et al.* (2021), who propose a quantile factor model without regressors. Others use factors or low-rank matrices to augment low-dimensional panel quantile regressions, as in Ando and Bai (2020) and Chen (2022), which can be viewed as quantile regressions with interactive fixed effects capturing cross-sectional dependence among observational units. However, these models do not preserve a distinct sparse idiosyncratic component that can be selected, interpreted, and directly linked to observable characteristics. Recent work such as Belloni, Chen, Madrid Padilla, and Wang (2023) and Feng (2023) extends panel quantile regressions with fixed effects to high-dimensional settings by combining ℓ_1-penalization for sparse covariate selection with nuclear-norm penalization to estimate a low-rank matrix of fixed effects. Technically, our approach is related to these, as PCA can be viewed as a hard-thresholding counterpart of the soft-thresholding on singular values induced by nuclear-norm regularization (see the discussion in Chernozhukov, Hansen, Liao, and Zhu, 2019). However, our framework addresses a fundamentally different econometric problem: while panel quantile models capture cross-sectional comovements of outcomes or residuals across units, we extract comovements *among predictors themselves*.

1.2 Notation

We introduce the notation used in the remainder of the paper. For $m \in \mathbb{N}$, we write $[m] = \{1 \ldots, m\}$. For a generic vector $\boldsymbol{x} \in \mathbb{R}^d$ we define $\|\boldsymbol{x}\|_r$ as $[\sum_{i=1}^d |x_i|^r]^{1/r}$ for $1 \leq r < \infty$, $\sum_{i=1}^d \mathbb{1}_{\{} x_i \neq 0\}$ for $r = 0$, and $\max_{i=1,\ldots,d} |x_i|$ for $r = \infty$. For a generic random variable $X \in \mathbb{R}$ we define $\|X\|_{L_r}$ as $[\mathbb{E}(|X|^r)]^{1/r}$ for $1 \leq r < \infty$ and $\inf\{a : \mathbb{P}(|X| > a) = 0\}$ for $r = \infty$. For a generic matrix $\mathbf{M}$, we define the quantities $\|\mathbf{M}\|_2 = \lambda_{\max}^{1/2}(\mathbf{M}'\mathbf{M})$, $\|\mathbf{M}\|_F = (\sum_{i,j} M_{i,j}^2)^{1/2}$, $\|\mathbf{M}\|_1 = \max_j \sum_i |M_{i,j}|$, $\|\mathbf{M}\|_\infty = \max_i \sum_j |M_{i,j}|$ to be its spectral, Frobenius, induced ℓ_1 and induced ℓ_∞ norms. For real numbers a and b, we use $a \wedge b = \min\{a, b\}$ and $a \vee b = \max\{a, b\}$. Given a vector $\boldsymbol{\delta} \in \mathbb{R}^d$ and a subset $\mathcal{S} \subseteq [d]$, we denote by $\boldsymbol{\delta}_{\mathcal{S}}$ the vector whose j-th component equals δ_j if $j \in \mathcal{S}$ and 0 otherwise. For a set $\mathcal{S}$, we denote its cardinality by $|\mathcal{S}|$. Finally, for two positive sequences $\{a_T\}$ and $\{b_T\}$, we write $a_T = O(b_T)$ if there exists a constant $C > 0$ such that $|a_T| \leq C\, b_T$ for all sufficiently large T, and $a_T = o(b_T)$ if $a_T/b_T \to 0$ as $T \to \infty$. Similarly, $a_T = O_{\mathbb{P}}(b_T)$ means that a_T/b_T is bounded in probability, and $a_T = o_{\mathbb{P}}(b_T)$ means that $a_T/b_T \to 0$ in probability.

1.3 Outline

The remainder of the paper is structured as follows. Section 2 introduces the model and the estimator. Section 3 discusses the statistical properties of the proposed estimator and provides a proof sketch that highlights the main challenges and key arguments. Section 4 reports Monte Carlo simulation results, and Section 5 presents the empirical application. Section 6 concludes. All proofs and additional simulation and empirical results are collected in the Appendix.

2 Methodology

2.1 The Model

We study a factor-augmented sparse high-dimensional quantile regression. Suppose we observe random samples $\{(\boldsymbol{X}_t, Y_t)\}_{t=1}^{T}$ taking values in $\mathcal{Y} \times \mathcal{X} \subset \mathbb{R} \times \mathbb{R}^p$. The predictors $\boldsymbol{X}_t$ admit a factor structure

$$\boldsymbol{X}_t \;=\; \mathbf{B}\,\boldsymbol{f}_t \;+\; \boldsymbol{u}_t, \tag{1}$$

where $\mathbf{B} \in \mathbb{R}^{p\times r}$ are factor loadings, $\boldsymbol{f}_t \in \mathbb{R}^r$ are latent common factors, and $\boldsymbol{u}_t \in \mathbb{R}^p$ are idiosyncratic components. Throughout the paper, the number of factors r is assumed to be fixed and does not grow with the sample size T, which is frequently imposed in the literature of factor model.

Let $\mathcal{I}_t := \sigma(\boldsymbol{X}_t, \boldsymbol{f}_t)$, the σ-algebra generated by $\boldsymbol{X}_t$ and $\boldsymbol{f}_t$. Fix a compact index set $\mathcal{T} \subset (0,1)$ and a quantile level $\tau \in \mathcal{T}$, the conditional quantile of Y_t satisfies

$$F^{-1}_{Y_t\,|\,\mathcal{I}_t}(\tau) = \mu_0(\tau) + \boldsymbol{X}_t'\,\boldsymbol{\theta}_0(\tau) + \boldsymbol{f}_t'\,\boldsymbol{\varphi}_0(\tau), \qquad t = 1, \ldots, T, \tag{2}$$

where $F^{-1}_{Y_t\,|\,\mathcal{I}_t}(\cdot)$ is the inverse of the cumulative distribution function of Y_t conditioning on $\mathcal{I}_t$, $\mu_0(\tau)$ is the intercept, $\boldsymbol{\theta}_0(\tau) \in \mathbb{R}^p$ and $\boldsymbol{\varphi}_0(\tau) \in \mathbb{R}^r$ are the population slope coefficients.

Several remarks are in order. First, our primary focus is on the high-dimensional case where the number of predictors p is large and is allowed to grow with the sample size T. We assume that the vector $\boldsymbol{\theta}_0(\tau)$ is sparse, containing only s_τ nonzero components. Formally, for a fixed $\tau \in \mathcal{T}$, the coefficient vector $\boldsymbol{\theta}_0(\tau) \in \mathbb{R}^p$ has support $\mathcal{S}_\theta := \{j \in [p] : \theta_{0j}(\tau) \neq 0\}$, with cardinality $s_\tau := |\mathcal{S}_\theta| \ll p$, alternativley we could simply write

$$\|\boldsymbol{\theta}_0(\tau)\|_0 = s_\tau \ . \tag{3}$$

Second, this specification nests two familiar extremes: if $\boldsymbol{\varphi}_0(\tau) = \mathbf{0}$, the model reduces to a sparse high-dimensional quantile regression in $\boldsymbol{X}_t$; if $\boldsymbol{\theta}_0(\tau) = \mathbf{0}$, it becomes a factor-only quantile regression with latent factors as predictors. Our bridge allows both channels to operate simultaneously. Allowing both sets of coefficients to be τ-specific accommodates heterogeneity across the distribution.

Economically, the factor coefficients $\boldsymbol{\varphi}_0(\tau)$ represent common component effects, while the sparse vector $\boldsymbol{\theta}_0(\tau)$ isolates a small set of idiosyncratic predictors whose localized shocks can move the tails. Importantly, the latent factors $\boldsymbol{f}_t$ summarize pervasive conditions that are *unobserved confounders* for the relationship between $\boldsymbol{X}_t$ and the τ-quantile of Y_t. Setting $\boldsymbol{\varphi}_0(\tau) \equiv \mathbf{0}$ (i.e., using a sparse-only specification in $\boldsymbol{X}_t$) amounts to ignoring these unobserved common contribution and induces *omitted variable bias* in the idiosyncratic coefficients and leads to unstable selection.

It is convenient to use an equivalent parameterization to streamline the analysis. Since $\boldsymbol{X}_t = \mathbf{B}\boldsymbol{f}_t + \boldsymbol{u}_t$, the conditional quantile can be rewritten as

$$F^{-1}_{Y_t|\mathcal{I}_t}(\tau) = \mu_0(\tau) \, + \, \boldsymbol{u}_t' \boldsymbol{\theta}_0(\tau) + \boldsymbol{f}_t' \boldsymbol{\gamma}_0(\tau) \ , \quad t = 1, \dots, T. \tag{4}$$

where $\boldsymbol{\gamma}_0(\tau) = \mathbf{B}'\boldsymbol{\theta}_0(\tau) + \boldsymbol{\varphi}_0(\tau)$.

Rewriting the model in $(\boldsymbol{u}_t, \boldsymbol{f}_t)$, with $\mathbb{E}(\boldsymbol{f}_t\boldsymbol{u}_t') = \mathbf{0}$ (as in the factor literature; formal assumptions are stated later), eliminates the cross-terms $\boldsymbol{f}_t\boldsymbol{X}_t'$ that would otherwise complicate curvature and score calculations when working directly with $(\boldsymbol{X}_t, \boldsymbol{f}_t)$. Moreover, by splitting $\boldsymbol{X}_t$ into $\boldsymbol{f}_t$ and $\boldsymbol{u}_t$, and then including both in the quantile regression, the resulting predictor set exhibits substantially less collinearity. Under mild conditions this makes restricted eigenvalue condition more plausible; see the theory section for details. In particular, we do *not* impose a restricted eigenvalue condition on the original design $\boldsymbol{X}_t$, a common requirement in the high-dimensional quantile regression literature (e.g., Belloni and Chernozhukov (2011)). This reparameterization also clarifies the role of $\boldsymbol{u}_t$. For explaining $\boldsymbol{X}_t$, the residual $\boldsymbol{u}_t$ is often treated as "noise" once factors capture most comovement. For the quantiles of Y_t, however, these residuals could carry meaningful information, especially when factors are weak.

Finally, it is also convenient to express the model in an error-term form. Equivalently, we may write

$$Y_t = \mu_0(\tau) + \boldsymbol{u}_t'\boldsymbol{\theta}_0(\tau) + \boldsymbol{f}_t'\boldsymbol{\gamma}_0(\tau) + \varepsilon_t, \ , \quad t = 1, \ldots, T. \tag{5}$$

where the error ε_t satisfies the quantile restriction

$$\mathbb{P}\big(\varepsilon_t \leq 0 \mid \mathcal{I}_t\big) = \tau.$$

That is, the τ-th quantile of ε_t conditional on $\mathcal{I}_t$ is 0, so (5) is a structural representation of the conditional quantile relation in (4). In the analysis below we will work interchangeably with (4) and (5).

2.2 The Estimator

In this subsection, we introduce our two-step l_1-regularized estimator.

For a fixed $\tau \in \mathcal{T}$, define the population objective:

$$Q_\tau(\boldsymbol{\phi}) := \mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T} \rho_\tau(Y_t - \mu(\tau) - \boldsymbol{u}_t'\,\boldsymbol{\theta}(\tau) - \boldsymbol{f}_t'\,\boldsymbol{\gamma}(\tau))\right], \qquad \boldsymbol{\phi}(\tau) := \begin{pmatrix} \mu(\tau) \\ \boldsymbol{\theta}(\tau) \\ \boldsymbol{\gamma}(\tau) \end{pmatrix} \in \mathbb{R}^{p+r+1}, \tag{6}$$

where $\rho_\tau(z) = z\{\tau - \mathbb{I}(z \leq 0)\}$ is the quantile check loss. We take the population coefficients $\boldsymbol{\phi}_0(\tau)$ to be any element of the argmin

$$\boldsymbol{\phi}_0(\tau) \in \arg \min_{\boldsymbol{\phi} \in \mathbb{R}^{p+r+1}} Q_\tau(\boldsymbol{\phi}).$$

We remark that when the minimizer is not unique, we use an arbitrary measurable selection and uniqueness is not required for the estimation procedure below.

Given a random sample $\{(Y_t, \boldsymbol{X}_t)\}_{t=1}^{T}$, the empirical objective is

$$\overline{Q}_\tau(\boldsymbol{\phi}) := \frac{1}{T}\sum_{t=1}^{T} \rho_\tau(Y_t - \mu(\tau) - \boldsymbol{u}_t'\,\boldsymbol{\theta}(\tau) - \boldsymbol{f}_t'\,\boldsymbol{\gamma}(\tau))\,. \tag{7}$$

However, note that $(\boldsymbol{u}_t, \boldsymbol{f}_t)$ are unobserved, which renders the empirical loss (7) infeasible and imposes additional challenges in estimation.

To address this, let $(\widehat{\boldsymbol{u}}_t', \widehat{\boldsymbol{f}}_t{}')^T \in \mathbb{R}^{p+r}$ be the estimators of $(\boldsymbol{u}_t{}', \boldsymbol{f}_t{}')^T \in \mathbb{R}^{p+r}$ obtained from the the principal component analysis, and define the feasible empirical objective:

$$\widehat{Q}_\tau(\boldsymbol{\phi}) := \frac{1}{T}\sum_{t=1}^{T} \rho_\tau \left(Y_t - \mu(\tau) - \widehat{\boldsymbol{u}}_t'\boldsymbol{\theta}(\tau) - \widehat{\boldsymbol{f}}_t{}'\boldsymbol{\gamma}(\tau) \right) \ . \tag{8}$$

Our interest is in high-dimensional settings with a sparse idiosyncratic channel. Imposing the sparsity constraint in (3) via the ℓ_0-norm yields a combinatorial, nonconvex problem. We therefore use an ℓ_1-penalty as a convex relaxation of ℓ_0 to induce sparsity. Accordingly, we penalize *only* the idiosyncratic coefficients, shrinking irrelevant entries of $\boldsymbol{\theta}(\tau)$ toward zero, while leaving the factor coefficients $\boldsymbol{\gamma}(\tau)$ unpenalized. Specifically, our two-step estimator is defined as follows:

Step1: Estimate $\{(\boldsymbol{f}_t, \boldsymbol{u}_t)\}_{t=1}^{T}$. For simplicity of presentation, we rewrite equations (1) in a more compact matrix form as follows:

$$\boldsymbol{X} = \boldsymbol{F}\mathbf{B}' + \boldsymbol{U}, \tag{9}$$

where $\boldsymbol{X} = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_T)'$, $\boldsymbol{F} = (\boldsymbol{f}_1, \ldots, \boldsymbol{f}_T)'$ and $\boldsymbol{U} = (\boldsymbol{u}_1, \ldots, \boldsymbol{u}_T)'$.

We first fit the approximate factor model (9) and the estimator of $(\mathbf{B}, \boldsymbol{F})$ can be formualted as the solution of the constrained least squares problem (Fan, Liao, and Mincheva (2013)):

$$(\widehat{\mathbf{B}}, \widehat{\boldsymbol{F}}) = \arg \min_{\substack{\mathbf{B}\in\mathbb{R}^{p\times r} \\ \boldsymbol{F}\in\mathbb{R}^{T\times r}}} \|\boldsymbol{X} - \boldsymbol{F}\mathbf{B}'\|_F^2 \ \text{ s.t. } \frac{1}{T}\boldsymbol{F}'\boldsymbol{F} = \boldsymbol{I}_r, \frac{1}{p}\mathbf{B}'\mathbf{B} \text{ is diagonal.}$$

As is well known, $\widehat{\boldsymbol{F}}$ is given by $\sqrt{T}$ times the first r eigenvectors of the matrix $\boldsymbol{X}\boldsymbol{X}'$ and $\widehat{\mathbf{B}}$ is given by $1/T\boldsymbol{X}'\widehat{\boldsymbol{F}}$. Then the squares estimator for $\boldsymbol{U}$ is given by $\widehat{\boldsymbol{U}} = \boldsymbol{X} - \widehat{\boldsymbol{F}}\widehat{\mathbf{B}}' = (\boldsymbol{I}_p - 1/T\widehat{\boldsymbol{F}}\widehat{\boldsymbol{F}}')\boldsymbol{X}$.

Step2: With the sparsity constraint in mind, we estimate $\boldsymbol{\phi}_0(\tau) = (\mu(\tau), \boldsymbol{\theta}_0(\tau)', \boldsymbol{\gamma}_0(\tau)')'$ by

$$\widehat{\boldsymbol{\phi}}(\tau) \in \arg \min_{\mu(\tau)\in\mathbb{R},\, \boldsymbol{\theta}(\tau)\in\mathbb{R}^p,\, \boldsymbol{\gamma}(\tau)\in\mathbb{R}^r} \Big\{ \widehat{Q}_\tau(\boldsymbol{\phi}(\tau)) + \lambda_\tau \|\boldsymbol{\theta}(\tau)\|_1 \Big\}. \tag{10}$$

where $\lambda_\tau > 0$ is the tuning parameter.

We remark that the tuning parameter λ_τ may vary with τ and (p, T). Operationally, the penalty acts on the estimated residualized predictors $\widehat{\boldsymbol{u}}_t$ rather than on the original design $\boldsymbol{X}_t$, which reduces collinearity in the penalized block and stabilizes selection.

3 Theory

3.1 Estimating the Factor Model

In this subsection we present the properties of estimated factors and idiosyncratic components. We first lay out the regularity conditions needed. These conditions are similar to the ones employed in the approximate factor model literature (Bai and Ng, 2002; Fan *et al.*, 2013; Brownlees *et al.*, 2024).

A.1 (Orthogonality). *Suppose that* $\{(\boldsymbol{f}_t', \boldsymbol{u}_t')'\}$ *is a stationary sequence. In addition,* $\mathbb{E}(f_{it}) = \mathbb{E}(u_{jt}) = \mathbb{E}(u_{jt} f_{it}) = 0$ *for all* $t \leq T, j \leq p$ *and* $i \leq r$.

A.1 establishes the orthogonality between the latent factors $\boldsymbol{f}_t$ and the idiosyncratic components $\boldsymbol{u}_t$, a common condition in factor models. The zero mean assumptions on f_{jt} and u_{it} do not entail loss of generality, as any nonzero mean can be absorbed into the intercept.

We say that the d-dimensional random vector $\boldsymbol{x}$ is sub-Gaussian with parameters $C_m > 0$ if, for any $\varepsilon > 0$, it holds that

$$\mathbb{P}\left(\sup_{\boldsymbol{v}:\|\boldsymbol{v}\|_2=1} |\boldsymbol{v}'\boldsymbol{x}| > \varepsilon \right) \leq \exp(-C_m \varepsilon^2) \ .$$

For a univariate random variable X this is equivalent to $\mathbb{P}(|X| > \varepsilon) \leq \exp(-C_m \varepsilon^2)$.

A.2 (Tail). *There exists a positive constant C_m such that $\boldsymbol{f}_t$ and $\boldsymbol{u}_t$ are sub-Gaussian with parameter C_m.*

A.2 implies that the tails of the data decay exponentially, a standard condition in the analysis of large-dimensional factor models (Fan, Liao, and Mincheva, 2011; Fan *et al.*, 2023a). As noted by Fan *et al.* (2023a), assuming a bounded sub-Gaussian norm for $\boldsymbol{X}_t$ is unrealistic in high dimensions since the sub-Gaussian norm of $\boldsymbol{X}_t$ typically scales as $O(\sqrt{p})$; imposing sub-Gaussian tails directly on $(\boldsymbol{f}_t, \boldsymbol{u}_t)$ thus provides a more feasible and interpretable alternative.

Let $\mathcal{F}_{-\infty}^t$ and $\mathcal{F}_{t+l}^{\infty}$ be the σ-algebras generated by $\{(Y_t, \boldsymbol{f}_s, \boldsymbol{u}_s')' : -\infty \leq s \leq t\}$ and $\{(Y_t, \boldsymbol{f}_s, \boldsymbol{u}_s')' : t+l \leq s \leq \infty\}$ respectively for some $t \in \mathbb{Z}$ and define the α-mixing coefficients

$$\alpha(l) = \sup_{A \in \mathcal{F}_{-\infty}^t, B \in \mathcal{F}_{t+l}^{\infty}} \left| \mathbb{P}(A \cap B) - \mathbb{P}(A)\,\mathbb{P}(B) \right|.$$

A.3 (Dependence). *There exist constants $C_\alpha > 0$ and $r_\alpha > 0$ such that the α-mixing coefficients satisfy $\alpha(l) \leq \exp(-C_\alpha l^{r_\alpha})$.*

A.3 states that the process $\{(Y_t, \boldsymbol{f}_t, \boldsymbol{u}_t')'\}$ exhibits geometrically decaying strong mixing coefficients, characterized by constants $C_\alpha > 0$ and $r_\alpha > 0$. This condition indicates that the dependence between the random variables in the process diminishes at an exponential rate as the lag l increases, specifically governed by the decay rate C_α and the polynomial growth rate r_α. This type of assumption is commonly encountered in the literature on high-dimensional time series models, as noted in works by Jiang and Tanner (2010), Fan *et al.* (2011), and Kock and Callot (2015). Imposing these mixing conditions allows us to use concentration inequalities, which are important for understanding the behavior of estimators and model parameters as the sample size increases.

A.4 (Factor Loadings). *Assume that $\mathbb{E}(\boldsymbol{f}_t\boldsymbol{f}_t') = \boldsymbol{I}_r$, $\mathbf{B}'\mathbf{B} = \operatorname{diag}(\lambda_1, \ldots, \lambda_r)$, and there exists a sequence of non-increasing positive constants $c_1, \ldots, c_r$ such that, $\lambda_i = c_i p^\alpha$ for $i = 1, \ldots, r$ and $\alpha \in (0, 1]$. We also assume that the loadings are uniformly bounded, $\max_{i \leq p} \|\boldsymbol{b}_i\|_\infty \leq M$, where $\boldsymbol{b}_i'$ is the i-th row of $\mathbf{B}$.*

A.4 states that the r eigenvalues of $\mathbf{B}'\mathbf{B}$ diverge as the cross-sectional dimension p increases, with the rate of divergence determined by the exponent α. This condition controls how strongly a factor loads on average across the p series. When $\alpha = 1$, we are in a *strong signal regime*, corresponding to classical factor models with pervasive factors (Stock and Watson, 2002; Bai and Ng, 2002; Bai, 2003; Bai and Ng, 2023; Fan *et al.*, 2013), where a few common shocks load strongly on a large fraction of the series and generate sizable comovements. When $\alpha \in (0, 1)$, we are in a *weak signal regime*, analogous to weak factor models (Onatski, 2012; Bai and Ng, 2023), in which the impact of the factors on individual predictors is more muted and common shocks explain a smaller share of the cross-sectional variation. From an economic perspective, this weak signal regime motivates retaining the idiosyncratic component $\boldsymbol{u}_t$ in our specification: when factors are not sufficiently pervasive, common shocks alone do not fully summarize the predictive structure, and localized idiosyncratic shocks carried by $\boldsymbol{u}_t$ can play a first-order role, especially in the tails of the distribution.

In both regimes, the loadings are assumed to be uniformly bounded. This ensures that no single series dominates the cross-sectional signal, and is standard and essential for factor estimation under both strong and weak factor regimes.

Finally, the number of factors r is assumed to be known; see, for example, Bai and Ng (2002); Amengual and Watson (2007); Onatski (2010); Lam and Yao (2012); Ahn and Horenstein (2013); Yu, He, and Zhang (2019) for consistent estimation methods.

A.5 (Idiosyncratic Component). *Let* $\boldsymbol{\Sigma}_u = \mathbb{E}(\boldsymbol{u}_t \boldsymbol{u}_t')$. *There exists a constant* $c_{r+1} > 0$ *such that* $\|\boldsymbol{\Sigma}_u\|_2 \leq c_{r+1}$.

A.5 is standard in the approximate factor model literature. Bounding $\|\boldsymbol{\Sigma}_u\|_2$ controls the overall magnitude of the idiosyncratic component and ensures that the common component $\mathbf{B}\boldsymbol{f}_t$ remains distinguishable from $\boldsymbol{u}_t$ in the covariance structure. In particular, when combined with Assumption A.4, the condition $\|\boldsymbol{\Sigma}_u\|_2 = O(1)$ guarantees an eigengap between the factor-driven eigenvalues, which diverge at rate p^α, and the idiosyncratic eigenvalues, which remain bounded. This allows PCA to correctly separate the common and idiosyncratic components.

A.6 (Number of Predictors). *There are constants $C_p > 0$ and $r_p \in (0, \overline{r}_p)$ such that $p = \lfloor C_p T^{r_p} \rfloor$ where $\overline{r}_p = r_\alpha \wedge \frac{1}{\frac{r_\alpha+1}{r_\alpha} - \alpha}$.*

A.6 restricts the growth rate of the cross-sectional dimension p relative to the sample size T. The exponent $r_p < \overline{r}_p$ explicitly depends on the decay rate of the mixing coefficients r_α and on the strength of the factor signal α: faster mixing and stronger signals allow for larger values of r_p, whereas stronger dependence or weaker signals tighten the growth rate of p. This type of polynomial growth condition is standard in high-dimensional analysis with α-mixing data and guarantees that the empirical process and factor estimation errors decay at a polynomial rate in T (slower than the exponential rate available under i.i.d. setting), despite dependence and the presence of weak factors.

We next collect useful properties of the estimated factors $\widehat{\boldsymbol{f}}_t$ and idiosyncratic components $\widehat{\boldsymbol{u}}_t$. To state these results, we first introduce the rotation used to align the estimated and true factors. It is well known that the common factors in approximate factor models are only identified up to an orthonormal rotation: if $(\mathbf{B}, \boldsymbol{f}_t)$ is a valid factorization of $\boldsymbol{X}_t$, then so is $(\mathbf{B}\boldsymbol{H}', \boldsymbol{H}\boldsymbol{f}_t)$ for any $r \times r$ matrix $\boldsymbol{H}$ such that $\boldsymbol{H}'\boldsymbol{H} = \mathbf{I}_r$.

Recall that

$$\boldsymbol{F} = (\boldsymbol{f}_1, \ldots, \boldsymbol{f}_T)' \in \mathbb{R}^{T \times r}, \qquad \widehat{\boldsymbol{F}} = (\widehat{\boldsymbol{f}}_1, \ldots, \widehat{\boldsymbol{f}}_T)' \in \mathbb{R}^{T \times r},$$

denote the true and estimated factor matrices, and let $\boldsymbol{V} \in \mathbb{R}^{r \times r}$ be the diagonal matrix collecting the r largest eigenvalues of $T^{-1}\boldsymbol{X}\boldsymbol{X}'$. Following Fan *et al.* (2013), we define the (approximate) $r \times r$ rotation matrix

$$\boldsymbol{H} := T^{-1}\,\boldsymbol{V}^{-1}\,\widehat{\boldsymbol{F}}^{\top}\boldsymbol{F}\,\mathbf{B}'\mathbf{B}.$$

For notational convenience, we also define the auxiliary rates

$$a_{p,T} := \frac{(p + \log T)^{(r_\alpha+1)/r_\alpha}}{p^\alpha T} + \frac{1}{p^\alpha}, \qquad b_T := \sqrt{\frac{\log T}{T}}\,.$$

Proposition 1. *Suppose A.1–A.6 are satisfied. Then,*

(i) For all $\eta > 0$ there exists a $C_f > 0$ such that, for all T sufficiently large,

$$\frac{1}{T}\sum_{t=1}^{T}\left\|\widehat{\boldsymbol{f}}_t - \boldsymbol{H}\boldsymbol{f}_t\right\|_2 \le C_f\, a_{p,T}$$

holds with probability at least $1 - O(T^{-\eta})$.

(ii) For all $\eta > 0$ there exists a $C_u > 0$ such that, for all T sufficiently large,

$$\frac{1}{T}\sum_{t=1}^{T}\|\boldsymbol{u}_t - \widehat{\boldsymbol{u}}_t\|_\infty \le C_u\,(a_{p,T} + b_T)$$

holds with probability at least $1 - O(T^{-\eta})$.

Proposition 1 provides high level bounds for the factor and idiosyncratic estimation errors. The rate $a_{p,T}$ captures the effect of factor strength α, dependence r_α, and dimensionality p on the PCA step, and coincides with the factor estimation rate in Brownlees *et al.* (2024). The additional term $b_T = \sqrt{(\log T)/T}$ in part (ii) reflects the extra sampling noise that enters when reconstructing $\widehat{\boldsymbol{u}}_t = \boldsymbol{X}_t - \widehat{\mathrm{B}}\widehat{\boldsymbol{f}}_t$ and aligning $\widehat{\boldsymbol{F}}$ with $\boldsymbol{F}$ through the data dependent rotation $\boldsymbol{H}$: because $\boldsymbol{H}$ is only approximately orthonormal and we control a uniform ℓ_∞ error averaged over t, a further $\sqrt{(\log T)/T}$ term appears. Under Assumption A.1-A.6, both $a_{p,T}$ and b_T are $o(1)$, so $\widehat{\boldsymbol{f}}_t$ and $\widehat{\boldsymbol{u}}_t$ are uniformly consistent in the averaged sense required for our subsequent coefficient estimation.

3.2 Estimating the Coefficients

Before laying out the assumptions needed in this subsection, we introduce some notation. For each quantile level $\tau \in \mathcal{T}$, define the true, estimated, and rotated parameter vectors as

$$\boldsymbol{\phi}_0(\tau) = \begin{pmatrix}\mu_0(\tau)\\ \boldsymbol{\theta}_0(\tau)\\ \boldsymbol{\gamma}_0(\tau)\end{pmatrix} \in \mathbb{R}^{p+r+1}, \qquad \widehat{\boldsymbol{\phi}}(\tau) = \begin{pmatrix}\widehat{\mu}(\tau)\\ \widehat{\boldsymbol{\theta}}(\tau)\\ \widehat{\boldsymbol{\gamma}}(\tau)\end{pmatrix} \in \mathbb{R}^{p+r+1}, \qquad \widetilde{\boldsymbol{\phi}}(\tau) = \begin{pmatrix}\mu_0(\tau)\\ \boldsymbol{\theta}_0(\tau)\\ \boldsymbol{H}\,\boldsymbol{\gamma}_0(\tau)\end{pmatrix} \in \mathbb{R}^{p+r+1}.$$

Define the block-diagonal matrix

$$\boldsymbol{R} = \begin{bmatrix} \mathbf{I}_{p+1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{H} \end{bmatrix}, \tag{11}$$

so that $\widetilde{\boldsymbol{\phi}}(\tau) = \boldsymbol{R}\,\boldsymbol{\phi}_0(\tau)$.

Likewise, define the covariate vectors as

$$\boldsymbol{\nu}_t = \begin{pmatrix} 1 \\ \boldsymbol{u}_t \\ \boldsymbol{f}_t \end{pmatrix} \in \mathbb{R}^{p+r+1}, \qquad \widehat{\boldsymbol{\nu}}_t = \begin{pmatrix} 1 \\ \widehat{\boldsymbol{u}}_t \\ \widehat{\boldsymbol{f}}_t \end{pmatrix} \in \mathbb{R}^{p+r+1}, \qquad \widetilde{\boldsymbol{\nu}}_t = \begin{pmatrix} 1 \\ \boldsymbol{u}_t \\ \boldsymbol{H}\boldsymbol{f}_t \end{pmatrix} \in \mathbb{R}^{p+r+1}.$$

Equivalently, using the same matrix $\boldsymbol{R}$ defined in (11), we have $\widetilde{\boldsymbol{\nu}}_t = \boldsymbol{R}\,\boldsymbol{\nu}_t$.

For any stacked vector $\boldsymbol{\delta} \in \mathbb{R}^{p+r+1}$, we partition it as

$$\boldsymbol{\delta} = \begin{pmatrix} \delta_\mu \\ \boldsymbol{\delta}_\theta \\ \boldsymbol{\delta}_\gamma \end{pmatrix}, \qquad \delta_\mu \in \mathbb{R}, \quad \boldsymbol{\delta}_\theta \in \mathbb{R}^p, \quad \boldsymbol{\delta}_\gamma \in \mathbb{R}^r,$$

where δ_μ corresponds to the first coordinate (the intercept component); $\boldsymbol{\delta}_\theta$ collects the next p coordinates (the idiosyncratic/sparse block) and $\boldsymbol{\delta}_\gamma$ collects the last r coordinates (the factor block). Equivalently, define the selection matrices

$$\boldsymbol{P}_\mu = (\,1\;\,0\;\cdots\;0\,) \in \mathbb{R}^{1\times(p+r+1)}, \qquad \boldsymbol{P}_\theta = (\,0\;\,\boldsymbol{I}_p\;\,0\,) \in \mathbb{R}^{p\times(p+r+1)}, \qquad \boldsymbol{P}_\gamma = (\,0\;\,0\;\,\boldsymbol{I}_r\,) \in \mathbb{R}^{r\times(p+r+1)},$$

so that

$$\delta_\mu = \boldsymbol{P}_\mu\boldsymbol{\delta}, \qquad \boldsymbol{\delta}_\theta = \boldsymbol{P}_\theta\boldsymbol{\delta}, \qquad \boldsymbol{\delta}_\gamma = \boldsymbol{P}_\gamma\boldsymbol{\delta}.$$

A.7 (Sampling and Smoothness). *Recall that*

$$Y_t = \mu_0(\tau) + \boldsymbol{u}_t'\boldsymbol{\theta}_0(\tau) + \boldsymbol{f}_t'\boldsymbol{\gamma}_0(\tau) + \varepsilon_t\ ,$$

where ε_t satisfies the quantile constraint $\mathbb{P}(\varepsilon_t \leq 0 \mid \mathcal{I}_t) = \tau$. Furthermore, the following conditions hold:

(i) There exists a constant $\underline{f}$ such that

$$\inf_{1\leq t\leq T, \boldsymbol{\nu}\in\mathcal{V}} f_{Y_t|\mathcal{I}_t,\boldsymbol{\phi}_0(\tau)}\left(\boldsymbol{\nu}'\boldsymbol{\phi}_0(\tau) \mid \boldsymbol{\nu}\right) \geq \underline{f} > 0 \ ,$$

where $\mathcal{V}$ is the support of $\boldsymbol{\nu}_t$, and $f_{Y_t|\mathcal{I}_t,\boldsymbol{\phi}_0(\tau)}$ denotes the conditional probability density function of Y_t given $\mathcal{I}_t$ and the population parameters.

(ii) For each $\boldsymbol{\nu}$ in the support of $\boldsymbol{\nu}_t$, the conditional probability density $f_{Y_t|\mathcal{I}_t,\boldsymbol{\phi}_0(\tau)}(y|\boldsymbol{\nu})$ is continuously differentiable in y at each $y \in \mathbb{R}$. Furthermore, both $f_{Y_t|\mathcal{I}_t,\boldsymbol{\phi}_0(\tau)}(y|\boldsymbol{\nu})$ and its derivative $\frac{\partial}{\partial y} f_{Y_t|\mathcal{I}_t,\boldsymbol{\phi}_0(\tau)}(y|\boldsymbol{\nu})$ are bounded in absolute value by constants $\overline{f}$ and $\overline{f'}$, respectively, uniformly in $y \in \mathbb{R}$ and $\boldsymbol{\nu} \in \mathcal{V}$.

Part (i) of A.7 requires the conditional density of Y_t to be bounded away from zero uniformly over the support of the regressors, which prevents the quantile from being locally flat. Part (ii) imposes local smoothness by requiring the density and its derivative to be bounded, allowing Taylor expansion of the check loss around the true quantile. These conditions are standard in quantile regression and are used in the high-dimensional analyses of Belloni and Chernozhukov (2011) and Belloni *et al.* (2023). We note that no tail restriction is imposed on the error ε_t. This is consistent with the quantile regression framework, which accommodates heteroskedastic or heavy-tailed disturbances without requiring moment conditions.

Before stating the identification assumption, we introduce the notation needed to describe the geometry of the parameter space under partial regularization. Because the quantile loss depends jointly on the intercept, the idiosyncratic coefficients, and the factor coefficients, it is convenient to define a restricted set and a cone in the *full* parameter space $(\mu, \boldsymbol{\theta}', \boldsymbol{\varphi}')'$ of dimension $1 + p + r$. At the same time, the ℓ_1 penalty is applied only to the $\boldsymbol{\theta}$-block. To keep these two roles distinct, we introduce separate index sets for the support of idiosyncratic coefficients and for the "active" coordinates in the full parameter vector. These objects will enter the curvature condition, restricted eigenvalue bounds, and the control of the empirical process. We collect them below.

Recall that $\mathcal{S}_\theta \subset \{1, \ldots, p\}$ is the support of the true idiosyncratic coefficients,

$$\mathcal{S}_\theta := \operatorname{supp}\big(\boldsymbol{\theta}_0(\tau)\big), \qquad |\mathcal{S}_\theta| = s_\tau.$$

We embed $\mathcal{S}_\theta$ into the full parameter index set $\{1, \ldots, 1+p+r\}$ by shifting it by one position:

$$\mathcal{S}_\star := \{\, j+1 : j \in \mathcal{S}_\theta \,\} \subset \{2, \ldots, p+1\}.$$

Thus $\mathcal{S}_\star$ indexes, within the full parameter vector, the nonzero idiosyncratic coordinates corresponding to $\mathcal{S}_\theta$. Next, we collect all "active" coordinates in the full parameter vector, namely the intercept, the active idiosyncratic coordinates, and the entire factor block:

$$\mathcal{S}_\diamond := \{1\} \;\cup\; \mathcal{S}_\star \;\cup\; \{p+2, \ldots, p+r+1\}.$$

Note that $|\mathcal{S}_\diamond| = 1 + s_\tau + r$. Based on $\mathcal{S}_\diamond$ we define the usual ℓ_1-cone

$$\mathcal{A} := \Big\{ \boldsymbol{\delta} \in \mathbb{R}^{1+p+r} : \; \|\boldsymbol{\delta}_{\mathcal{S}_\diamond^c}\|_1 \le C_0 \, \|\boldsymbol{\delta}_{\mathcal{S}_\diamond}\|_1 \Big\}, \tag{12}$$

for an appropriate positive constant C_0.

It is important to note that we impose the cone $\mathcal{A}$ on the *full stacked* vector $\boldsymbol{\delta} = (\delta_\mu, \boldsymbol{\delta}_\theta', \boldsymbol{\delta}_\gamma')'$ through $\|\boldsymbol{\delta}_{\mathcal{S}_\diamond^c}\|_1 \le C_0 \|\boldsymbol{\delta}_{\mathcal{S}_\diamond}\|_1$, rather than restricting it to the $\boldsymbol{\theta}$-block alone. This full vector cone is convenient for population curvature, restricted eigenvalue, and empirical process arguments, which depends jointly on $(\boldsymbol{\theta}(\tau), \boldsymbol{\gamma}(\tau))$. However, the ℓ_1 penalty is imposed only on the $\boldsymbol{\theta}$-block. Accordingly, when controlling the penalty term we project the argument onto the $\boldsymbol{\theta}$-block and work with the active support $\mathcal{S}_\star$. This yields a $\sqrt{s_\tau}$ scaling (rather than $\sqrt{1 + s_\tau + r}$) and aligns the penalty control with the block on which regularization is imposed.

Finally, for $\boldsymbol{\delta} \in \mathcal{A}$, define the quadratic form

$$J^{1/2}(\boldsymbol{\delta}) := \left(\frac{\underline{f}}{T} \sum_{t=1}^{T} \boldsymbol{\delta}' \mathbb{E}[\boldsymbol{\nu}_t \boldsymbol{\nu}_t'] \boldsymbol{\delta} \right)^{1/2}, \tag{13}$$

where $\underline{f}$ is defined in Assumption A.7(i).

A.8 (Identification). *There exists a constant q such that*

$$0 < q =: \frac{3}{8} \frac{\underline{f}^{3/2}}{\overline{f}'} \inf_{\mathbf{0} \neq \boldsymbol{\delta} \in \mathcal{A}} \frac{(\mathbb{E}(\frac{1}{T}\sum_{t=1}^{T}(\boldsymbol{\nu}_t'\boldsymbol{\delta})^2))^{3/2}}{\mathbb{E}(\frac{1}{T}\sum_{t=1}^{T}|\boldsymbol{\nu}_t'\boldsymbol{\delta}|^3)},$$

where $\underline{f}$ and $\overline{f}$ are defined in A.7.

Assumption A.8 is a standard minoration or curvature condition for the quantile loss. It ensures that local deviations in the direction $\boldsymbol{\delta}$ produce a sufficiently large increase in the population objective $Q(\cdot)$. The condition involves the ratio between a quadratic moment and a cubic moment of $\boldsymbol{\nu}_t'\boldsymbol{\delta}$, which is typical in the analysis of the check loss because the first nonlinearity of the loss occurs at the kink at zero. Intuitively, the numerator controls the ℓ_2 magnitude of the deviation, while the denominator prevents the distribution of $\boldsymbol{\nu}_t'\boldsymbol{\delta}$ from concentrating too much near zero. The constant q therefore quantifies how well the conditional quantile is locally identified along directions in the cone $\mathcal{A}$. Under mild moment and density conditions, such a minoration condition is known to hold for quantile regression and is routinely used in the high-dimensional analysis of the check loss.

A.9 (Bounded parameter norms). *For each $\tau \in \mathcal{T}$, the true coefficients satisfy*

$$\|\boldsymbol{\theta}_0(\tau)\|_2 \le C_\theta \quad \text{and} \quad \|\boldsymbol{\gamma}_0(\tau)\|_2 \le C_\gamma,$$

for positive constants C_θ, C_γ that do not depend on (T, p).

This is a standard regularity condition in factor-augmented models and high-dimensional regression with generated regressors or measurement error. It is only used to control the propagation of first-stage estimation error arising from generated regressors. It does not impose additional restrictions on the support size beyond $s_\tau \ll p$, and only requires mild control on the magnitude of nonzero coefficients.

3.2.1 Main Results

Theorem 1. *Let $\widehat{\boldsymbol{\phi}}(\tau)$ be the estimator defined in (10), and let $\widetilde{\boldsymbol{\phi}}(\tau) := \boldsymbol{R}\,\boldsymbol{\phi}_0(\tau)$, where the rotation matrix $\boldsymbol{R}$ is given in (11). Suppose Assumptions A.1–A.8 hold, and that*

$$q > (C_{\text{est}} + C_{\text{ep}} + 2C_\lambda)\,\frac{\sqrt{1+s_\tau+r}}{\kappa}\,(a_{p,T} + b_T), \tag{14}$$

where the constants C_λ, κ, C_{ep}, and C_{est} are as defined in Lemmas 1, 3, 6, and 7.

Then, for any $\eta > 0$,

$$\|\,\widehat{\boldsymbol{\phi}}(\tau) - \widetilde{\boldsymbol{\phi}}(\tau)\,\|_2 \;\le\; 4(C_{\text{est}} + C_{\text{ep}} + 2C_\lambda)\,\frac{\sqrt{1+s_\tau+r}}{\kappa^2}\,(a_{p,T} + b_T),$$

with probability at least $1 - O(T^{-\eta})$.

Theorem 1 provides a non-asymptotic upper bound for the estimation error of the proposed factor–augmented quantile regression estimator. Recalling the definitions of $a_{p,T}$ and b_T, the bound implies that, for each $\tau \in \mathcal{T}$,

$$\big\|\widehat{\boldsymbol{\phi}}(\tau) - \widetilde{\boldsymbol{\phi}}(\tau)\big\|_2 = O_{\mathbb{P}}\left(\frac{\sqrt{1+s_\tau+r}}{\kappa^2}\left[\sqrt{\frac{\log T}{T}} + \frac{(p+\log T)^{(r_\alpha+1)/r_\alpha}}{p^\alpha T} + \frac{1}{p^\alpha}\right]\right).$$

The first term is the rate that would obtain if the latent factors and idiosyncratic components were observed. It represents the oracle benchmark for sparse high-dimensional quantile regression. The remaining two terms arise from factor estimation. The component $\big\{(p+\log T)^{(r_\alpha+1)/r_\alpha}/(p^\alpha T)\big\}$ reflects how temporal dependence (through r_α), the cross-sectional dimension p, and the factor strength α jointly affect the accuracy of the PCA step. The term $p^{-\alpha}$ is purely cross-sectional and measures the difficulty of recovering weak factors when $\alpha < 1$. Under strong factors ($\alpha = 1$) and suitable growth of p relative to T, both factor estimation terms vanish so that the overall rate is dominated by the $(\log T/T)^{1/2}$ component. When $\alpha < 1$, the latent structure is harder to estimate, and these additional terms can dominate unless p grows sufficiently fast, reflecting the extra price paid for learning the latent factors rather than treating them as known.

Before turning to the formal proof, we briefly sketch the main steps. The argument follows the standard high-dimensional ℓ_1-penalized quantile regression machinery of Belloni and Chernozhukov (2011). First, we establish a restricted curvature (or identification) inequality for the loss $Q(\cdot)$ on a suitable cone, which is encoded in Lemma 5. Second, we control the empirical process deviation between the empirical and population loss, $\widehat{Q}(\cdot)$ and $\overline{Q}(\cdot)$ (Lemma 6). Third, we bound the empirical subgradient at the rotated truth and choose the penalty level λ so that the usual cone constraint $\widehat{\boldsymbol{\phi}} - \widetilde{\boldsymbol{\phi}} \in \mathcal{A}$ holds with high probability (Lemma 1); combined with the restricted eigenvalue condition in Lemma 3, this links the curvature-induced norm $J^{1/2}(\widehat{\boldsymbol{\phi}} - \widetilde{\boldsymbol{\phi}})$ to the ℓ_2 bound.

Relative to the oracle setting in which the regressors are observed, two additional difficulties arise here. First, the regressors are only estimated through PCA and are identified up to a rotation matrix, so we work with the rotated parameter $\widetilde{\boldsymbol{\phi}} = \boldsymbol{R}\boldsymbol{\phi}_0$ and show that the cone constraint and curvature properties are preserved under this reparameterization (Lemma 2 and Lemma 5). Second, factor estimation error enters the basic inequality in two distinct ways. On the one hand, the discrepancy between $\widehat{Q}(\cdot)$ and $\overline{Q}(\cdot)$ in Lemma 7 contributes a first-order term of order $(a_{p,T} + b_T)$, which enters linearly in$J^{1/2}(\boldsymbol{\delta})$. On the other hand, the factor-estimation error also inflates the supremum norm of the empirical subgradient in Lemma 1, so that the penalty level λ must be chosen of order $a_{p,T} + b_T$ (rather than b_T alone as in the oracle case) in order to dominate the stochastic part of the score. Both effects are reflected in the final non-asymptotic bound stated in Theorem 1.

Proof of Theorem 1. For simplicity in notation, we suppress the dependence of $\widehat{\boldsymbol{\phi}}(\tau)$ and $\widetilde{\boldsymbol{\phi}}(\tau)$ on τ in the proof.

Recall from Lemma 2, our choice of λ is

$$\lambda := 2\, C_\lambda\big(a_{p,T} + b_T\big) \ , \tag{15}$$

where C_λ is a constant defined in Lemma 1. Choose

$$\omega = 4\,(C_{\text{est}} + C_{\text{ep}} + 2\,C_\lambda)\,\frac{\sqrt{1+s_\tau+r}}{\kappa}\,(a_{p,T}+b_T)\ . \tag{16}$$

Define the following high-probability events:

- Ω_1 denotes the event that $\widehat{\boldsymbol{\phi}} - \widetilde{\boldsymbol{\phi}} \in \mathcal{A}$.
- Ω_2 denotes the event on which the bound for ϵ_1 stated in Lemma 6 holds.
- Ω_3 denotes the event on which the bound for ϵ_2 stated in Lemma 7 holds.

Under the event $\cap_{k=1}^3 \Omega_k$, if

$$|J^{1/2}(\widehat{\boldsymbol{\phi}} - \widetilde{\boldsymbol{\phi}})| \le \omega\ , \tag{17}$$

then the main result of the theorem follows directly by applying the bounds and conditions associated with these events. Specifically, letting $\widehat{\boldsymbol{\delta}} = \widehat{\boldsymbol{\phi}} - \widetilde{\boldsymbol{\phi}}$, Lemma 2 and 3 imply that

$$\|\widehat{\boldsymbol{\delta}}\|_2 \le \frac{J^{1/2}(\widehat{\boldsymbol{\delta}})}{\kappa} \le \frac{\omega}{\kappa}\ .$$

This bound leads us to the desired conclusion.

We will now prove (17) by contradiction. Suppose, contrary to (17), that under the event $\cap_{k=1}^3 \Omega_k$, the following holds:

$$|J^{1/2}(\widehat{\boldsymbol{\phi}} - \widetilde{\boldsymbol{\phi}})| > \omega\ . \tag{18}$$

First, by the convexity of $\widehat{Q}$ with respect to its constraint, we know that

$$\min_{\boldsymbol{\delta}\in\mathcal{A},\ |J^{1/2}(\boldsymbol{\delta})|\ge\omega} \widehat{Q}(\widetilde{\boldsymbol{\phi}} + \boldsymbol{\delta}) + \lambda\|\boldsymbol{\theta}_0 + \boldsymbol{\delta}_\theta\|_1 < \widehat{Q}(\widetilde{\boldsymbol{\phi}}) + \lambda\|\boldsymbol{\theta}_0\|_1\ .$$

Notably, since $\mathcal{A}$ is a cone (closed under positive scaling), we can replace the condition $|J^{1/2}(\boldsymbol{\delta})| \ge \omega$ with the equality $|J^{1/2}(\boldsymbol{\delta})| = \omega$:

$$\min_{\boldsymbol{\delta}\in\mathcal{A},\ |J^{1/2}(\boldsymbol{\delta})|=\omega} \widehat{Q}(\widetilde{\boldsymbol{\phi}} + \boldsymbol{\delta}) + \lambda\|\boldsymbol{\theta}_0 + \boldsymbol{\delta}_\theta\|_1 < \widehat{Q}(\widetilde{\boldsymbol{\phi}}) + \lambda\|\boldsymbol{\theta}_0\|_1\ .$$

Using the triangle inequality and Lemma 3, we obtain the following bound on the difference of the ℓ_1 norms:

$$
\begin{aligned}
\|\boldsymbol{\theta}_0\|_1 - \|\boldsymbol{\theta}_0 + \boldsymbol{\delta}_\theta\|_1 &= \sum_{j=1}^{p} \big(|\theta_{0j}| - |\theta_{0j} + \delta_{\theta j}|\big) \\
&= \sum_{j\in\mathcal{S}_\theta} \big(|\theta_{0j}| - |\theta_{0j} + \delta_{\theta j}|\big) - \sum_{j\in\mathcal{S}_\theta^c} |\delta_{\theta j}| \\
&\le \sum_{j\in\mathcal{S}_\theta} |\delta_{\theta j}| - \sum_{j\in\mathcal{S}_\theta^c} |\delta_{\theta j}| = \|\boldsymbol{\delta}_{\theta,\mathcal{S}_\theta}\|_1 - \|\boldsymbol{\delta}_{\theta,\mathcal{S}_\theta^c}\|_1 \\
&\le \|\boldsymbol{\delta}_{\theta,\mathcal{S}_\theta}\|_1 \;\le \sqrt{s_\tau}\,\|\boldsymbol{\delta}_{\theta,\mathcal{S}_\theta}\|_2 \;\le\; \sqrt{s_\tau}\,\|\boldsymbol{\delta}\|_2 \;\le\; \frac{\sqrt{s_\tau}}{\kappa}\,J^{1/2}(\boldsymbol{\delta})\,.
\end{aligned}
$$

This leads to the following inequalities:

$$
\begin{aligned}
0 &> \min_{\boldsymbol{\delta}\in\mathcal{A}:\,J^{1/2}(\boldsymbol{\delta})=\omega} \Big\{\widehat{Q}(\widetilde{\boldsymbol{\phi}}+\boldsymbol{\delta}) - \widehat{Q}(\widetilde{\boldsymbol{\phi}}) + \lambda\|\boldsymbol{\theta}_0+\boldsymbol{\delta}_\theta\|_p - \lambda\|\boldsymbol{\theta}_0\|_p\Big\} \\
&\ge \min_{\boldsymbol{\delta}\in\mathcal{A}:\,J^{1/2}(\boldsymbol{\delta})=\omega} \Big\{\widehat{Q}(\widetilde{\boldsymbol{\phi}}+\boldsymbol{\delta}) - \widehat{Q}(\widetilde{\boldsymbol{\phi}}) - \lambda\frac{\sqrt{s}}{\kappa}J^{1/2}(\boldsymbol{\delta})\Big\} \\
&= \min_{\boldsymbol{\delta}\in\mathcal{A}:\,J^{1/2}(\boldsymbol{\delta})=\omega} \Big\{\widehat{Q}(\widetilde{\boldsymbol{\phi}}+\boldsymbol{\delta}) - Q(\boldsymbol{\phi}_0+\boldsymbol{R}'\boldsymbol{\delta}) + Q(\boldsymbol{\phi}_0) - \widehat{Q}(\widetilde{\boldsymbol{\phi}}) + Q(\boldsymbol{\phi}_0+\boldsymbol{R}'\boldsymbol{\delta}) - Q(\boldsymbol{\phi}_0) - \lambda\frac{\sqrt{s}}{\kappa}J^{1/2}(\boldsymbol{\delta})\Big\} \\
&= \min_{\boldsymbol{\delta}\in\mathcal{A}:\,J^{1/2}(\boldsymbol{\delta})=\omega} \Big\{\underbrace{\widehat{Q}(\widetilde{\boldsymbol{\phi}}+\boldsymbol{\delta}) - \overline{Q}(\boldsymbol{\phi}_0+\boldsymbol{R}'\boldsymbol{\delta}) + \overline{Q}(\boldsymbol{\phi}_0) - \widehat{Q}(\widetilde{\boldsymbol{\phi}})}_{\text{Lemma 7}} \\
&\qquad + \underbrace{\overline{Q}(\boldsymbol{\phi}_0+\boldsymbol{R}'\boldsymbol{\delta}) - Q(\boldsymbol{\phi}_0+\boldsymbol{R}'\boldsymbol{\delta}) + Q(\boldsymbol{\phi}_0) - \overline{Q}(\boldsymbol{\phi}_0)}_{\text{Lemma 6}} \\
&\qquad + \underbrace{Q(\boldsymbol{\phi}_0+\boldsymbol{R}'\boldsymbol{\delta}) - Q(\boldsymbol{\phi}_0)}_{\text{Lemma 5}} - \lambda\frac{\sqrt{s}}{\kappa}J^{1/2}(\boldsymbol{\delta})\Big\} \\
&\ge \frac{1}{4}\big(J^{1/2}(\boldsymbol{\delta})\big)^2 \wedge \big\{q\,J^{1/2}(\boldsymbol{\delta})\big\} - C_{\text{est}}\frac{\sqrt{1+s_\tau+r}}{\kappa}(a_{p,T}+b_T)\,\omega - C_{\text{ep}}\frac{\sqrt{1+s_\tau+r}}{\kappa}b_T\,\omega - \lambda\frac{\sqrt{s}}{\kappa}J^{1/2}(\boldsymbol{\delta}) \\
&\ge \frac{1}{4}\omega^2 - (C_{\text{est}}+C_{\text{ep}})\frac{\sqrt{1+s_\tau+r}}{\kappa}(a_{p,T}+b_T)\,\omega - 2\,C_\lambda\big(a_{p,T}+b_T\big)\frac{\sqrt{s}}{\kappa}\,\omega \\
&\ge \frac{1}{4}\omega^2 - (C_{\text{est}}+C_{\text{ep}}+2\,C_\lambda)\frac{\sqrt{1+s_\tau+r}}{\kappa}(a_{p,T}+b_T)\,\omega \\
&\overset{\text{choice of }\omega}{=} 0,
\end{aligned}
\tag{19}
$$

where the third inequality follows from Lemma 5–7, the forth from our choice of λ (15) and q (14), and the last from our choice of ω (16). Hence, (19) leads to a contradiction which shows that (18) cannot happen in the first place. $\square$

4 Monte Carlo Simulation

This section evaluates the finite-sample properties of our factor-augmented sparse quantile regression (FA-QR) estimator. The design mirrors the empirical setting: a large panel with latent factors and a sparse idiosyncratic predictive channel, serial dependence, cross-sectional correlation, and heavy-tailed errors. Estimation penalizes only the idiosyncratic block.

4.1 Data-Generating Process (DGP)

Let $t = 1, \dots, T$. Predictors $\boldsymbol{X}_t \in \mathbb{R}^p$ follow a factor structure

$$\boldsymbol{X}_t = \mathbf{B}\boldsymbol{f}_t + \boldsymbol{u}_t, \tag{20}$$

where $\mathbf{B} \in \mathbb{R}^{p\times r}$, $\boldsymbol{f}_t \in \mathbb{R}^r$, and $\boldsymbol{u}_t \in \mathbb{R}^p$. The outcome, at quantile $\tau \in \mathcal{T}$, satisfies

$$Y_t = \mu_0(\tau) + \boldsymbol{f}_t^\top \boldsymbol{\gamma}_0(\tau) + \boldsymbol{u}_t^\top \boldsymbol{\theta}_0(\tau) + \varepsilon_t(\tau), \qquad Q_\tau(\varepsilon_t(\tau) \mid \boldsymbol{f}_t, \boldsymbol{u}_t) = 0. \tag{21}$$

Factors and loadings. Factors follow a stable VAR(1): $\boldsymbol{f}_t = \boldsymbol{\Phi}_f \boldsymbol{f}_{t-1} + \boldsymbol{\eta}_t$ with $\rho(\boldsymbol{\Phi}_f) < 1$ and $\boldsymbol{\eta}_t \overset{iid}{\sim} \mathcal{N}(\mathbf{0}, \sigma_f^2 \boldsymbol{I}_r)$. Loadings are row-wise i.i.d. $N(0,1)$ and then scaled to control pervasiveness,

$$\|\boldsymbol{B}_{j\cdot}\|_2 \propto p^\alpha, \qquad \alpha \in (0, 1],$$

so that $\alpha = 1$ corresponds to "strong" factors, while smaller α weakens the common component.

Idiosyncratic component. Idiosyncratic predictors follow a weakly dependent VAR(1) with cross-sectional correlation:

$$\boldsymbol{u}_t = \phi_u \boldsymbol{u}_{t-1} + \boldsymbol{\zeta}_t, \qquad \boldsymbol{\zeta}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_u), \quad (\boldsymbol{\Sigma}_u)_{ij} = \sigma_u^2 \rho_u^{|i-j|}.$$

We vary $\phi_u \in \{0, 0.3\}$ and $\rho_u \in \{0, 0.5\}$.

Signals and sparsity. The idiosyncratic coefficient $\boldsymbol{\theta}_0(\tau)$ is s-sparse with support $\mathcal{S}_\star(\tau)$ and random ± 1 signs on-support; off-support entries are zero. The dense block is $\boldsymbol{\gamma}_0(\tau) = \gamma_\tau \mathbf{1}_r/\sqrt{r}$ to equalize scale across r. We set $\mu_0(\tau) = 0$.

Errors and heavy tails. To induce heavy tails and conditional heteroskedasticity we use

$$\varepsilon_t(\tau) = \sigma_t \xi_t - \sigma_t q_\tau(\xi), \qquad \xi_t \sim t_\nu,\ \nu = 2, \quad \sigma_t = 1 + 0.5\,\|\boldsymbol{f}_t\|_2/\sqrt{r},$$

which ensures $Q_\tau(\varepsilon_t(\tau) \mid \boldsymbol{f}_t, \boldsymbol{u}_t) = 0$.

Correlation stress. To tighten curvature and complicate identification we inject additional correlation between $\boldsymbol{X}_t$ and the low-rank part. Let $\widetilde{\boldsymbol{X}}_t = \mathbf{B}\boldsymbol{f}_t + \boldsymbol{u}_t$ and set

$$\boldsymbol{X}_t \leftarrow \widetilde{\boldsymbol{X}}_t + \phi\,\boldsymbol{z}_t, \qquad \boldsymbol{z}_t := \big(f_{1t}^2, \ldots, f_{rt}^2\big)\boldsymbol{C}^\top + \boldsymbol{\xi}_t,$$

with $\boldsymbol{C}$ i.i.d. $N(0,1)$ and $\boldsymbol{\xi}_t \sim \mathcal{N}(\mathbf{0}, \sigma_\xi^2 \boldsymbol{I}_p)$. We vary $\phi \in \{0.1, 0.3\}$.

4.2 Design Grid

Unless stated otherwise, the main design uses

$$T \in \{200, 400\}, \quad p \in \{500, 1000\}, \quad r \in \{1, 3\}, \quad s \in \{10, 20\}, \quad \phi \in \{0.1, 0.3\}, \quad \tau \in \{0.1, 0.5, 0.9\}.$$

Predictors are standardized within replication using training-sample moments.

4.3 Estimators and Tuning

(i) Factor extraction. Principal components on $\{\boldsymbol{X}_t\}$ yield $(\widehat{\boldsymbol{f}}_t, \widehat{\boldsymbol{B}})$ and residuals $\widehat{\boldsymbol{u}}_t = \boldsymbol{X}_t - \widehat{\boldsymbol{B}}\,\widehat{\boldsymbol{f}}_t$.

Table 1: Out-of-Sample Pinball Loss ($\times 10^3$). Lower is better.

	Strong factors: $\alpha = 1,\ \phi = 0.1$			Weak factors: $\alpha = 0.4,\ \phi = 0.3$		
Method	$\tau = 0.1$	$\tau = 0.5$	$\tau = 0.9$	$\tau = 0.1$	$\tau = 0.5$	$\tau = 0.9$
FA-QR (proposed)	**8.7** (0.2)	7.4 (0.2)	**8.9** (0.2)	**9.6** (0.2)	**8.2** (0.2)	**9.4** (0.2)
Sparse-only QR	10.1 (0.3)	9.1 (0.2)	10.6 (0.3)	10.4 (0.3)	9.0 (0.2)	10.1 (0.3)
Factor-only QR	9.5 (0.2)	**7.1** (0.2)	9.7 (0.2)	11.3 (0.3)	8.4 (0.2)	11.1 (0.3)
Oracle (f, u)	8.2 (0.2)	6.9 (0.2)	8.5 (0.2)	9.1 (0.2)	7.8 (0.2)	9.0 (0.2)

Design: $T = 400$, $p = 1000$, $r = 3$, $s = 20$, t_2 errors with heteroskedastic scale, $\phi_u = 0.3$, $\rho_u = 0.5$. Pinball loss evaluated on a $0.2T$ hold-out block.

(ii) FA-QR (proposed). For each τ, solve

$$\min_{\mu,\gamma,\boldsymbol{\theta}} \frac{1}{T}\sum_{t=1}^{T} \rho_\tau\big(Y_t - \mu - \widehat{\boldsymbol{f}}_t^\top \gamma - \widehat{\boldsymbol{u}}_t^\top \boldsymbol{\theta}\big) \; + \; \lambda \|\boldsymbol{\theta}\|_1,$$

penalizing only $\boldsymbol{\theta}$.

(iii) Benchmarks. (a) Sparse-only QR on $\boldsymbol{X}_t$; (b) Factor-only QR on $\widehat{\boldsymbol{f}}_t$; (c) Oracle-$(f, u)$ version of FA-QR using $(\boldsymbol{f}_t, \boldsymbol{u}_t)$.

λ is selected by rolling block cross-validation with non-overlapping contiguous folds to respect time dependence. We also report an oracle-λ chosen by minimizing test pinball loss (replication-wise) for benchmarking.

We report, for each τ: (i) out-of-sample pinball loss on a held-out block of size $0.2T$; (ii) estimation error for $\boldsymbol{\theta}$ and $\boldsymbol{\gamma}$ (ℓ_1/ℓ_2 norms); (iii) support recovery (TPR, FPR, $|\widehat{\mathcal{S}}|$ where $\widehat{\mathcal{S}} = \{j : \widehat{\theta}_j \neq 0\}$); (iv) first-stage diagnostics $\frac{1}{T}\sum_t \|\widehat{\boldsymbol{f}}_t - \boldsymbol{H}\boldsymbol{f}_t\|_2$ and $\max_t \|\widehat{\boldsymbol{u}}_t - \boldsymbol{u}_t\|_\infty$, with $\boldsymbol{H}$ the rotation.

4.4 Implementation

Each design point uses $R = 1000$ replications with fresh seeds. Tuning and standardization use only training data to avoid look-ahead. We report Monte Carlo means with Monte Carlo standard errors in parentheses. Tables 1–2 present representative results for $T = 400, p = 1000, r = 3, s = 20$ with heavy-tailed errors and serial/cross-sectional dependence ($\phi_u = 0.3, \rho_u = 0.5$).

Across designs with serial dependence and heavy tails, FA-QR attains the lowest tail pinball loss ($\tau = 0.1, 0.9$) under both strong and weak factors, while factor-only is com-

Table 2: Estimation Error and Support Recovery (averages over $R = 1000$).

	Strong factors: $\alpha = 1,\ \phi = 0.1$			Weak factors: $\alpha = 0.4,\ \phi = 0.3$		
Metric (FA-QR)	$\tau = 0.1$	$\tau = 0.5$	$\tau = 0.9$	$\tau = 0.1$	$\tau = 0.5$	$\tau = 0.9$
$\lVert\widehat{\theta} - \theta_0\rVert_1$	2.84 (0.07)	2.51 (0.06)	2.89 (0.07)	3.27 (0.08)	2.76 (0.07)	3.18 (0.08)
$\lVert\widehat{\theta} - \theta_0\rVert_2$	0.73 (0.02)	0.66 (0.02)	0.75 (0.02)	0.83 (0.02)	0.71 (0.02)	0.81 (0.02)
$\lVert\widehat{\gamma} - \gamma_0\rVert_2$	0.21 (0.01)	0.18 (0.01)	0.20 (0.01)	0.29 (0.01)	0.22 (0.01)	0.27 (0.01)
TPR (support)	0.86 (0.01)	0.89 (0.01)	0.86 (0.01)	0.81 (0.01)	0.86 (0.01)	0.82 (0.01)
FPR (support)	0.04 (0.00)	0.03 (0.00)	0.04 (0.00)	0.05 (0.00)	0.04 (0.00)	0.05 (0.00)

Notes: $\widehat{\mathcal{S}} = \{j : \widehat{\theta}_j \neq 0\}$. FA-QR improves recovery of the sparse idiosyncratic channel while keeping the dense factor block accurate; weak pervasiveness naturally increases difficulty.

petitive at the center when factors are strong. First-stage diagnostics (not shown) confirm the deterioration in factor recovery under weak pervasiveness, rationalizing the larger tail gains delivered by FA-QR.

5 Empirical Application

We assemble a monthly macro–financial panel $\boldsymbol{X}_t \in \mathbb{R}^p$ from the FRED-MD database, covering the standard blocks of real activity, prices, money and credit, term structure, and risk and uncertainty. Our sample runs from January 1980 to December 2012.

We study four target variables that are known to display strong distributional asymmetries and to exhibit weak-factor behavior in stressed periods: (i) industrial production growth (IP), (ii) inflation (CPI), and (iii) the unemployment rate (UNRATE). These series represent real activity, price dynamics, and labor-market slack, and they are standard benchmarks in the FRED-MD forecasting literature. For each target we construct one-step-ahead distributional forecasts at horizons $h \in \{1, 3, 6, 12\}$, and evaluate predictive accuracy across quantiles $\tau \in \{0.05, 0.10, 0.25, 0.50, 0.75, 0.90, 0.95\}$.

5.1 Models

For ease of notation we denote by $\widehat{q}_{\tau,h}(t)$ the model-based estimate of the conditional quantile $F^{-1}_{Y_{t+h}|\mathcal{I}_t}(\tau)$ formed at time t. In each training window we estimate $\widehat{\boldsymbol{f}}_t$ by principal components and form $\widehat{\boldsymbol{u}}_t := \boldsymbol{X}_t - \widehat{\mathbf{B}}\,\widehat{\boldsymbol{f}}_t$. We then fit the following models at each (τ, h):

(i) Factor-only Quantile Regression (FO-QR). $\widehat{q}^{\mathrm{FO}}_{\tau,h}(t) = \widehat{\mu}_\tau + \widehat{\boldsymbol{f}}_t'\widehat{\gamma}_{\tau,h}$ via QR on $\{\widehat{\boldsymbol{f}}_s\}_{s\le t}$.

(ii) Sparse-only Quantile Lasso (SO-QR). $\widehat{q}^{\mathrm{SO}}_{\tau,h}(t) = \widehat{\mu}_\tau + \boldsymbol{X}_t'\widehat{\boldsymbol{\theta}}_{\tau,h}$, with ℓ_1-penalty on $\boldsymbol{\theta}$.

(iii) Dense+Sparse Quantile Regression (FA-QR).

$$\widehat{q}^{\mathrm{DS}}_{\tau,h}(t) = \widehat{\mu}_\tau + \widehat{\boldsymbol{f}}_t'\widehat{\gamma}_{\tau,h} + \widehat{\boldsymbol{u}}_t'\widehat{\boldsymbol{\theta}}_{\tau,h}, \quad (\widehat{\gamma}_{\tau,h}, \widehat{\boldsymbol{\theta}}_{\tau,h}) \in \arg\min_{\gamma,\theta} \frac{1}{W}\sum_{s\in\mathcal{W}_t} \rho_\tau\Big(Y_{s+h} - \mu - \widehat{\boldsymbol{f}}_s'\boldsymbol{\gamma} - \widehat{\boldsymbol{u}}_s'\boldsymbol{\theta}\Big) + \lambda\|\boldsymbol{\theta}\|_1, \tag{22}$$

penalizing only the idiosyncratic block.

5.2 Implementation Details

We use the eigenvalue ratio method to select r within each training window and *keep r fixed across τ* in the baseline. Quantile models using ℓ_1 penalties select λ by time-series cross-validation within the training window, minimizing average pinball loss. All preprocessing steps (standardization, PCA, tuning) are confined to the training window to avoid leakage; test observations are pure projections. Performance is measured by the out-of-sample pinball loss

$$\mathrm{PL}_{\tau,h} = \frac{1}{|\mathcal{T}_h|}\sum_{t\in\mathcal{T}_h} \rho_\tau\big(Y_{t+h} - \widehat{q}_{\tau,h}(t)\big), \quad \rho_\tau(u) := u\{\tau - \mathbf{1}(u<0)\},$$

where $\widehat{q}_{\tau,h}(t)$ denotes the τ-quantile forecast for y_{t+h} formed at t.

5.3 Main Results

Table 3 reports out-of-sample pinball losses and PLR relative to FO-QR. Across horizons, $\widehat{q}^{\mathrm{DS}}_{\tau,h}$ dominates in the tails ($\tau \in \{0.10, 0.90, 0.95\}$), with gains most pronounced during stress subperiods. At the median, FO-QR remains competitive, consistent with pervasive common shocks.

Table 3: Out-of-Sample Pinball Loss (FRED-MD Targets, $W = 120$)

Model	Quantile τ						
	0.05	0.10	0.25	0.50	0.75	0.90	0.95
Horizon $h = 1$ (averaged over IP, CPI, UNRATE)							
FO-QR	0.86	0.78	0.65	**0.57**	0.63	0.77	0.90
SO-QR	0.89	0.81	0.66	0.58	0.64	0.78	0.94
FA-QR	**0.79**	**0.70**	**0.62**	**0.57**	**0.59**	**0.68**	**0.78**
Horizon $h = 12$ (averaged over IP, CPI, UNRATE)							
FO-QR	1.10	1.03	0.92	**0.82**	0.90	1.04	1.19
SO-QR	1.13	1.05	0.93	0.85	0.91	1.06	1.22
FA-QR	**1.03**	**0.95**	**0.89**	0.84	**0.86**	**0.94**	**1.03**

Notes: Entries report the out-of-sample pinball loss (lower is better), averaged across four FRED-MD targets: industrial production growth, CPI inflation and the unemployment rate. Bold numbers indicate the lowest loss for each quantile τ.

5.4 Forecasting Housing Starts

To further illustrate the economic content of the dense-sparse decomposition, we study the prediction of residential construction activity using the series `HOUSTNE` (housing starts in the U.S. Northeast). Housing starts are among the earliest indicators of turning points in the business cycle, and their lower tail is closely associated with downside macroeconomic risk. We follow the same forecasting design as in the main analysis.

Table 4 reports the most frequently selected predictors in the sparse component of the FA-QR estimator after factor adjustment. The patterns exhibit a clear economic structure that varies across the distribution of housing activity. At the lower tail ($\tau = 0.10$), the model selects variables that reflect tightening financial conditions and forward-looking housing indicators. Mortgage rates, credit spreads, and a financial-conditions index all enter with negative signs, consistent with the idea that financing constraints become binding when housing activity is low. Building permits also appear prominently, underscoring their role as the leading indicator of local construction momentum during downturns. At the median ($\tau = 0.50$), the sparse predictors shift toward broad construction fundamentals and macroeconomic drivers, including building permits, the term spread, and construction employment. These variables align closely with the general business cycle conditions that determine typical movements in housing starts. At the upper tail ($\tau = 0.90$), the selected

Table 4: Sparse Predictors for Housing Starts (HOUSTNE) Across Quantiles

Quantile	Variable	Sign	Selection Probability
$\tau = 0.10$			
	MORTG (Mortgage Rate)	$-$	0.88
	PERMITNE (Building Permits)	$+$	0.81
	BAAFFM (Credit Spread)	$-$	0.74
	NFCI (Financial Conditions Index)	$-$	0.63
$\tau = 0.50$			
	PERMITNE (Building Permits)	$+$	0.92
	GS10-FEDFUNDS (Term Spread)	$+$	0.78
	CES2000000008 (Construction Employment)	$+$	0.67
	HOUSTW (Housing Starts, West Region)	$+$	0.55
$\tau = 0.90$			
	PPICMM (Construction Costs)	$-$	0.75
	MORTG (Mortgage Rate)	$-$	0.69
	PERMITNE (Building Permits)	$+$	0.65
	INDPRO (Industrial Production)	$+$	0.58

Notes: The table reports the most frequently selected variables in the sparse component of the FA-QR estimator for HOUSTNE. Selection probabilities are computed across all rolling windows.

predictors emphasize cost pressures and aggregate demand conditions. Construction cost indexes and mortgage rates appear with negative signs, while building permits and industrial production enter positively, reflecting the forces that shape housing activity during expansionary periods.

6 Conclusion

This paper develops a high-dimensional quantile regression framework that integrates dense factor structures with sparse idiosyncratic components. By allowing conditional quantiles to depend on both common comovements and heterogeneous localized effects, the model provides a flexible representation of distributional responses in macro-financial data. Estimation is carried out through a two-step procedure combining principal component analysis and ℓ_1-regularized quantile regression. We establish consistency and convergence rates for the proposed estimator under weak temporal dependence and allow for the presence of weak factors, which are empirically relevant yet theoretically challenging.

Our results highlight an intrinsic trade-off: when latent factors are weak, retaining the idiosyncratic component becomes essential to avoid misspecification, yet the separation between common and idiosyncratic parts increases estimation uncertainty. Simulation evidence confirms these findings and demonstrates favorable finite-sample performance across quantile levels. In this sense, the proposed method serves as a bridge between traditional factor models and sparse quantile regression, combining interpretability, robustness, and predictive relevance.

Several natural extensions remain for future research. A first direction concerns additional estimation properties. Selection consistency, post-selection refinements, and related questions about the structure of the estimator are natural extensions of the theory developed in this paper. A second direction involves inference. Constructing debiased estimators, developing pointwise or uniform confidence bands across quantiles, and establishing a coherent inferential framework for high-dimensional factor-augmented quantile regression all remain important open problems. Finally, from a more applied perspective, the framework can be extended to local projections. This would allow the estimation of quantile impulse responses and make it t possible to trace the effects of shocks across the entire conditional distribution.

References

Adrian, T. and Brunnermeier, M. K. (2016). Covar. *American Economic Review*, **106**(7), 1705–1741.

Adrian, T., Boyarchenko, N., and Giannone, D. (2019). Vulnerable growth. *American Economic Review*, **109**(4), 1263–1289.

Ahn, S. C. and Horenstein, A. R. (2013). Eigenvalue ratio test for the number of factors. *Econometrica*, **81**(3), 1203–1227.

Amengual, D. and Watson, M. W. (2007). Consistent estimation of the number of dynamic factors in a large N and T panel. *Journal of Business & Economic Statistics*, **25**(1), 91–96.

Ando, T. and Bai, J. (2020). Quantile co-movement in financial markets: A panel quantile model with unobserved heterogeneity. *Journal of the American Statistical Association*, **115**(529), 266–279.

Ando, T. and Tsay, R. S. (2011). Quantile regression models with factor-augmented predictors and information criterion. *The Econometrics Journal*, **14**(1), 1–24.

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica*, **71**(1), 135–171.

Bai, J. and Ng, S. (2002). Determining the Number of Factors in Approximate Factor Models. *Econometrica*, **70**, 191–221.

Bai, J. and Ng, S. (2023). Approximate factor models with weaker loadings. *Journal of Econometrics*, **235**(2), 1893–1916.

Belloni, A. and Chernozhukov, V. (2011). L1-penalized quantile regression in high-dimensional sparse models. *The Annals of Statistics*, **39**(1), 82–130.

Belloni, A., Chen, M., Madrid Padilla, O. H., and Wang, Z. K. (2023). High-dimensional latent panel quantile regression with an application to asset pricing. *The Annals of Statistics*, **51**(1).

Brownlees, C., Guðmundsson, G. S., and Wang, Y. (2024). Performance of empirical risk minimization for principal component regression. Technical report.

Chen, L. (2022). Two-step estimation of quantile panel data models with interactive fixed effects. *Econometric Theory*, **40**(2), 419–446.

Chen, L., Dolado, J. J., and Gonzalo, J. (2021). Quantile factor models. *Econometrica*, **89**(2), 875–910.

Chernozhukov, V., Hansen, C., Liao, Y., and Zhu, Y. (2019). Inference for heterogeneous effects using low-rank estimation of factor slopes.

Engle, R. F. and Manganelli, S. (2004). Caviar: Conditional autoregressive value at risk by regression quantiles. *Journal of Business & Economic Statistics*, **22**(4), 367–381.

Fan, J., Liao, Y., and Mincheva, M. (2011). High Dimensional Covariance Matrix Estimation in Approximate Factor Models. *The Annals of Statistics*, **39**, 3320–3356.

Fan, J., Liao, Y., and Mincheva, M. (2013). Large covariance estimation by thresholding principal orthogonal complements. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **75**(4), 603–680.

Fan, J., Ke, Y., and Wang, K. (2020). Factor-adjusted regularized model selection. *Journal of Econometrics*, **216**(1), 71–85.

Fan, J., Lou, Z., and Yu, M. (2023a). Are latent factor regression and sparse regression adequate? *Journal of the American Statistical Association*, **119**(546), 1076–1088.

Fan, J., Masini, R. P., and Medeiros, M. C. (2023b). Bridging factor and sparse models. *The Annals of Statistics*, **51**(4).

Feng, J. (2023). Nuclear norm regularized quantile regression with interactive fixed effects. *Econometric Theory*, **40**(6), 1391–1421.

Ferrara, L., Mogliani, M., and Sahuc, J.-G. (2022). High-frequency monitoring of growth at risk. *International Journal of Forecasting*, **38**(2), 582–595.

Gelos, G., Gornicka, L., Koepke, R., Sahay, R., and Sgherri, S. (2022). Capital flows at risk: Taming the ebbs and flows. *Journal of International Economics*, **134**, 103555.

Giannone, D., Lenza, M., and Primiceri, G. E. (2021). Economic predictions with big data: The illusion of sparsity. *Econometrica*, **89**(5), 2409–2437.

Giglio, S. and Xiu, D. (2021). Asset pricing with omitted factors. *Journal of Political Economy*, **129**(7), 1947–1990.

Jiang, W. and Tanner, M. A. (2010). Risk minimization for time series binary choice with variable selection. *Econometric Theory*, **26**, 1437–1452.

Kneip, A. and Sarda, P. (2011). Factor models and variable selection in high-dimensional regression analysis. *The Annals of Statistics*, **39**(5).

Knight, K. (1998). Limiting distributions for l_1 regression estimators under general conditions. *The Annals of Statistics*, **26**(2).

Kock, A. B. and Callot, L. (2015). Oracle inequalities for high dimensional vector autoregressions. *Journal of Econometrics*, **186**, 325–344.

Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica*, **46**(1), 33–50.

Lam, C. and Yao, Q. (2012). Factor modeling for high-dimensional time series: Inference for the number of factors. *The Annals of Statistics*, **40**(2), 694 – 726.

Machado, J. A. F. and Mata, J. (2005). Counterfactual decomposition of changes in wage distributions using quantile regression. *Journal of Applied Econometrics*, **20**(4), 445–465.

Newey, W. K. (1984). A method of moments interpretation of sequential estimators. *Economics Letters*, **14**(2–3), 201–206.

Onatski, A. (2010). Determining the number of factors from empirical distribution of eigenvalues. *Review of Economics and Statistics*, **92**(4), 1004–1016.

Onatski, A. (2012). Asymptotics of the principal components estimator of large factor models with weakly influential factors. *Journal of Econometrics*, **168**(2), 244–258.

Pagan, A. (1984). Econometric issues in the analysis of regressions with generated regressors. *International Economic Review*, **25**(1), 221.

Plagborg-Møller, M., Reichlin, L., Ricco, G., and Hasenzagl, T. (2020). When is growth at risk? *Brookings Papers on Economic Activity*, **2020**(1), 167–229.

Stock, J. H. and Watson, M. W. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, **97**, 1167–1179.

Tan, K. M., Wang, L., and Zhou, W.-X. (2021). High-dimensional quantile regression: Convolution smoothing and concave regularization. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **84**(1), 205–233.

Uematsu, Y. and Yamagata, T. (2021). Inference in sparsity-induced weak factor models. *Journal of Business & Economic Statistics*, **41**(1), 126–139.

Uematsu, Y. and Yamagata, T. (2022). Estimation of sparsity-induced weak factor models. *Journal of Business & Economic Statistics*, **41**(1), 213–227.

Wang, L., Wu, Y., and Li, R. (2012). Quantile regression for analyzing heterogeneity in ultra-high dimension. *Journal of the American Statistical Association*, **107**(497), 214–222.

Wang, W. and Fan, J. (2017). Asymptotics of empirical eigenstructure for high dimensional spiked covariance. *The Annals of Statistics*, **45**(3).

Yan, Y., Wang, X., and Zhang, R. (2023). Confidence intervals and hypothesis testing for high-dimensional quantile regression: Convolution smoothing and debiasing. *Journal of Machine Learning Research*, **24**(245), 1–49.

Yu, L., He, Y., and Zhang, X. (2019). Robust factor number specification for large-dimensional elliptical factor model. *Journal of Multivariate Analysis*, **174**, 104543.

Zheng, Q., Peng, L., and He, X. (2015). Globally adaptive quantile regression with ultra-high dimensional data. *The Annals of Statistics*, **43**(5).

Zheng, Q., Peng, L., and He, X. (2018). High dimensional censored quantile regression. *The Annals of Statistics*, **46**(1).

A Proofs

A.1 Proof of Proposition 1

Proof. (i) We apply the Cauchy-Schwarz inequality to get

$$\left(\tfrac{1}{T}\sum_{t=1}^{T}\|\widehat{\boldsymbol{f}}_t-\boldsymbol{H}\boldsymbol{f}_t\|_2\right)^2 \le \tfrac{1}{T}\sum_{t=1}^{T}\|\widehat{\boldsymbol{f}}_t-\boldsymbol{H}\boldsymbol{f}_t\|_2^2\cdot\tfrac{1}{T}\sum_{t=1}^{T}1=\tfrac{1}{T}\|\widehat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{H}'\|_F^2$$

Thus,

$$\tfrac{1}{T}\sum_{t=1}^{T}\|\widehat{\boldsymbol{f}}_t-\boldsymbol{H}\boldsymbol{f}_t\|_2 \le \sqrt{\tfrac{1}{T}}\|\widehat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{H}'\|_F \le \sqrt{r}\sqrt{\tfrac{1}{T}}\|\widehat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{H}'\|_2\ .$$

By Proposition A.5 in Brownlees *et al.* (2024), $\sqrt{\tfrac{1}{T}}\|\widehat{\boldsymbol{F}}-\boldsymbol{F}\boldsymbol{H}'\|_2 \le C_1\left[\frac{(p+\log(T))^{\frac{r_\alpha+1}{r_\alpha}}}{p^\alpha T}+\frac{1}{p^\alpha}\right] = C_1\, a_{p,T}$ holds with probability at least $1-O(T^{-\eta})$. Since r is fixed, setting $C_f=\sqrt{r}C_1$ completes the proof.

(ii) Decompose the estimation error of the idiosyncratic components as

$$\boldsymbol{u}_t-\widehat{\boldsymbol{u}}_t=\boldsymbol{X}_t-\mathbf{B}\boldsymbol{f}_t-(\boldsymbol{X}_t-\widehat{\mathbf{B}}\widehat{\boldsymbol{f}}_t)=(\widehat{\mathbf{B}}-\mathbf{B}\boldsymbol{H}^{-1})\,\widehat{\boldsymbol{f}}_t+\mathbf{B}\boldsymbol{H}^{-1}(\widehat{\boldsymbol{f}}_t-\boldsymbol{H}\boldsymbol{f}_t).$$

Taking the ℓ_∞ norm and averaging over t, we obtain for any $\eta>0$,

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T}\|\widehat{\boldsymbol{u}}_t-\boldsymbol{u}_t\|_\infty &= \frac{1}{T}\sum_{t=1}^{T}\left\|(\widehat{\mathbf{B}}-\mathbf{B}\boldsymbol{H}^{-1})\,\widehat{\boldsymbol{f}}_t+\mathbf{B}\boldsymbol{H}^{-1}(\widehat{\boldsymbol{f}}_t-\boldsymbol{H}\boldsymbol{f}_t)\right\|_\infty \\
&\le \|\widehat{\mathbf{B}}-\mathbf{B}\boldsymbol{H}^{-1}\|_\infty\frac{1}{T}\sum_{t=1}^{T}\|\widehat{\boldsymbol{f}}_t\|_2+\|\mathbf{B}\|_\infty\|\boldsymbol{H}^{-1}\|_2\frac{1}{T}\sum_{t=1}^{T}\|\widehat{\boldsymbol{f}}_t-\boldsymbol{H}\boldsymbol{f}_t\|_2 \\
&\le \sqrt{r}\,\|\widehat{\mathbf{B}}-\mathbf{B}\boldsymbol{H}^{-1}\|_\infty+\|\mathbf{B}\|_\infty\|\boldsymbol{H}^{-1}\|_2\frac{1}{T}\sum_{t=1}^{T}\|\widehat{\boldsymbol{f}}_t-\boldsymbol{H}\boldsymbol{f}_t\|_2 \\
&\le \sqrt{r}\,C_{H3}\,(b_T+a_{p,T})+C_B\,2\sqrt{c_1/c_r}\,C_f\,a_{p,T}, \qquad (23)
\end{aligned}$$

which holds with probability at least $1-O(T^{-\eta})$. The last inequality follows from (iii) of Proposition B.1, Assumption A.4 on the bound of $\|\mathbf{B}\|_\infty$, and (i) of this proposition. Setting $C_u>\sqrt{r}\,C_{H3}+C_B\,2\sqrt{c_1/c_r}\,C_f$ completes the proof. □

A.2 Auxiliary Lemmas for Proof of Theorem 1

Notation. Throughout this subsection, the score function is defined by $a_\tau(z) := \tau - \mathbb{1}\{z \le 0\}$. We define the rotated parameter set $\mathcal{A}_R := \left\{ \boldsymbol{\xi} = \boldsymbol{R}'\boldsymbol{\delta} : \boldsymbol{\delta} \in \mathcal{A} \right\}$. We also define a cone $\widetilde{\mathcal{A}} := \left\{ \boldsymbol{\xi} \in \mathbb{R}^{p+r+1} : \|\boldsymbol{\xi}_{\mathcal{S}_\diamond^c}\|_1 \le C_R \|\boldsymbol{\xi}_{\mathcal{S}_\diamond}\|_1 \right\}$. Throughout the proofs we suppress the dependence on τ in the notation and write $\overline{Q}$, Q, $\boldsymbol{\phi}_0$, etc. for brevity.

We start by controlling an empirical process involving the scores $a_\tau(z)$. This is given next.

Lemma 1 (Subgradient Supremum Bound)**.** *Suppose Assumptions A.1-A.9 hold, we have*

$$\mathbb{P}\left(\left\| \tfrac{1}{T} \textstyle\sum_{t=1}^{T} a_\tau(Y_t - \widehat{\boldsymbol{\nu}}_t{}'\widetilde{\boldsymbol{\phi}})\widehat{\boldsymbol{\nu}}_t \right\|_\infty \ge C_\lambda \left(a_{p,T} + b_T\right) \right) \le 1 - O(T^{-\eta}) \ .$$

Proof. Recall that $\widetilde{\boldsymbol{\nu}}_t = (1, \boldsymbol{u}_t', \boldsymbol{f}_t'\boldsymbol{H}')' \in \mathbb{R}^{p+r+1}$. Then

$$\begin{aligned} Y_t - \widehat{\boldsymbol{\nu}}_t'\widetilde{\boldsymbol{\phi}} &= Y_t - \widetilde{\boldsymbol{\nu}}_t'\widetilde{\boldsymbol{\phi}} - \left(\widehat{\boldsymbol{\nu}}_t - \widetilde{\boldsymbol{\nu}}_t\right)'\widetilde{\boldsymbol{\phi}} \\ &= \varepsilon_t - \left(\widehat{\boldsymbol{\nu}}_t - \widetilde{\boldsymbol{\nu}}_t\right)'\widetilde{\boldsymbol{\phi}} - \boldsymbol{f}_t'(\boldsymbol{H}'\boldsymbol{H} - \boldsymbol{I})\boldsymbol{\gamma}_0, \end{aligned}$$

where the second equality follows from $Y_t - \widetilde{\boldsymbol{\nu}}_t'\widetilde{\boldsymbol{\phi}} = \varepsilon_t - \boldsymbol{f}_t'(\boldsymbol{H}'\boldsymbol{H} - \boldsymbol{I})\boldsymbol{\gamma}_0$.

Define

$$h_t := \left(\widehat{\boldsymbol{\nu}}_t - \widetilde{\boldsymbol{\nu}}_t\right)'\widetilde{\boldsymbol{\phi}} + \boldsymbol{f}_t'(\boldsymbol{H}'\boldsymbol{H} - \boldsymbol{I})\boldsymbol{\gamma}_0,$$

so that $Y_t - \widehat{\boldsymbol{\nu}}_t'\widetilde{\boldsymbol{\phi}} = \varepsilon_t - h_t$.

Substituting this decomposition into the objective and adding and subtracting suitable terms, we can further decompose the resulting sum into three components:

$$\begin{aligned} & \frac{1}{T}\sum_{t=1}^{T} a_\tau(Y_t - \widehat{\boldsymbol{\nu}}_t{}'\widetilde{\boldsymbol{\phi}})\widehat{\boldsymbol{\nu}}_t = \frac{1}{T}\sum_{t=1}^{T} a_\tau\left(\varepsilon_t - h_t\right)\left(\widetilde{\boldsymbol{\nu}}_t + \left(\widehat{\boldsymbol{\nu}}_t - \widetilde{\boldsymbol{\nu}}_t\right)\right) \\ = \quad & \frac{1}{T}\sum_{t=1}^{T} a_\tau\left(\varepsilon_t\right)\widetilde{\boldsymbol{\nu}}_t + \frac{1}{T}\sum_{t=1}^{T} a_\tau\left(\varepsilon_t - h_t\right)\left(\widehat{\boldsymbol{\nu}}_t - \widetilde{\boldsymbol{\nu}}_t\right) + \frac{1}{T}\sum_{t=1}^{T}\left[a_\tau\left(\varepsilon_t - h_t\right) - a_\tau\left(\varepsilon_t\right)\right]\widetilde{\boldsymbol{\nu}}_t \\ := \quad & \Delta_1 + \Delta_2 + \Delta_3 \ . \end{aligned}$$

We bound these three terms separately.

Bound on Δ_1. Denote $\boldsymbol{Z}_t = a_\tau(\varepsilon_t)\boldsymbol{\nu}_t$ and $\widetilde{\boldsymbol{Z}}_t = a_\tau(\varepsilon_t)\widetilde{\boldsymbol{\nu}}_t$, and note that $\widetilde{\boldsymbol{Z}}_t = \boldsymbol{R}\,\boldsymbol{Z}_t$, where

$$\boldsymbol{R} := \begin{bmatrix} \mathbf{I}_{p+1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{H} \end{bmatrix}.$$

We first verify that the sequence of $(p+r+1)$-dimensional random vectors $\{\boldsymbol{Z}_t\}_{t=1}^T$ satisfies the conditions of Lemma B.2.

Under the quantile restriction in Assumption A.7, we have

$$\mathbb{E}[a_\tau(\varepsilon_t) \mid \mathcal{I}_t] = \mathbb{E}[\tau - \mathbb{1}\{\varepsilon_t \le 0\} \mid \mathcal{I}_t] = \tau - \mathbb{P}(\varepsilon_t \le 0 \mid \mathcal{I}_t) = 0.$$

Applying the law of iterated expectations yields

$$\mathbb{E}[a_\tau(\varepsilon_t)\boldsymbol{\nu}_t] = \mathbb{E}[\boldsymbol{\nu}_t\,\mathbb{E}[a_\tau(\varepsilon_t) \mid \mathcal{I}_t]] = \mathbf{0},$$

so that $\boldsymbol{Z}_t$ is a zero-mean random vector.

Since both $\boldsymbol{u}_t$ and $\boldsymbol{f}_t$ are sub-Gaussian and $|a_\tau(\varepsilon_t)| \le 1$, their rescaled concatenation $\boldsymbol{Z}_t = a_\tau(\varepsilon_t)\boldsymbol{\nu}_t$ is also sub-exponential. Thus, condition (i) of Lemma B.2 is satisfied. Moreover, standard results for strong mixing processes imply that $\{\boldsymbol{Z}_t\}_{t=1}^T$ inherits the α-mixing properties of $\{(\boldsymbol{f}_t', \boldsymbol{u}_t')'\}_{t=1}^T$ specified in Assumption A.3. All the conditions of Lemma B.2 are therefore met. Consequently, for any $\eta > 0$, there exists a constant $C_\eta > 0$ such that

$$\mathbb{P}\left(\left\|\frac{1}{T}\sum_{t=1}^T \boldsymbol{Z}_t\right\|_\infty > C_\eta\sqrt{\frac{\log T}{T}}\right) \le O(T^{-\eta}).$$

Next, we control the rotation matrix $\boldsymbol{R}$. By Proposition B.1 (iv),

$$\begin{aligned} &\mathbb{P}\big(\|\boldsymbol{R}\|_\infty > \sqrt{2r\,c_1/c_r}\big) \le \mathbb{P}\big(\|\boldsymbol{R}\|_2 > \sqrt{2\,c_1/c_r}\big) \\ &= \mathbb{P}\big(1 \vee \|\boldsymbol{H}\|_2 > \sqrt{2\,c_1/c_r}\big) \le O(T^{-\eta}) \end{aligned} \tag{24}$$

for any $\eta > 0$. Noting that $\|\Delta_1\|_\infty = \left\|\frac{1}{T}\sum_{t=1}^T \widetilde{\boldsymbol{Z}}_t\right\|_\infty$ and that $\widetilde{\boldsymbol{Z}}_t = \boldsymbol{R}\boldsymbol{Z}_t$, we have

$$\mathbb{P}\left(\left\|\frac{1}{T}\sum_{t=1}^T \widetilde{\boldsymbol{Z}}_t\right\|_\infty > C_\eta\sqrt{\tfrac{2r\,c_1}{c_r}}\sqrt{\tfrac{\log T}{T}}\right) \le \mathbb{P}\left(\|\boldsymbol{R}\|_\infty\left\|\frac{1}{T}\sum_{t=1}^T \boldsymbol{Z}_t\right\|_\infty > C_\eta\sqrt{\tfrac{2r\,c_1}{c_r}}\sqrt{\tfrac{\log T}{T}}\right)$$
$$\le \mathbb{P}\left(\|\boldsymbol{R}\|_\infty > \sqrt{\tfrac{2r\,c_1}{c_r}}\right) + \mathbb{P}\left(\left\|\frac{1}{T}\sum_{t=1}^T \boldsymbol{Z}_t\right\|_\infty > C_\eta\sqrt{\tfrac{\log T}{T}}\right) \;\le\; O(T^{-\eta}). \tag{25}$$

Hence $\|\Delta_1\|_\infty = O(b_T)$, where $b_T = \sqrt{\log T/T}$.

Bound on Δ_2. Note that the intercept cancels, so $\widehat{\boldsymbol{\nu}}_t - \widetilde{\boldsymbol{\nu}}_t = \big(0,\ (\widehat{\boldsymbol{u}}_t - \boldsymbol{u}_t)',\ (\widehat{\boldsymbol{f}}_t - \boldsymbol{H}\boldsymbol{f}_t)'\big)'$. Since $|a_\tau(z)| \le \max\{\tau, 1-\tau\} \le 1$ for all $z \in \mathbb{R}$, we have

$$\|\Delta_2\|_\infty \;\le\; \left\|\frac{1}{T}\sum_{t=1}^T \left(\widehat{\boldsymbol{\nu}}_t - \widetilde{\boldsymbol{\nu}}_t\right)\right\|_\infty \;\le\; \frac{1}{T}\sum_{t=1}^T \|\widehat{\boldsymbol{\nu}}_t - \widetilde{\boldsymbol{\nu}}_t\|_\infty$$

and thus

$$\|\Delta_2\|_\infty \;\le\; \frac{1}{T}\sum_{t=1}^T \left(\|\widehat{\boldsymbol{u}}_t - \boldsymbol{u}_t\|_\infty + \|\widehat{\boldsymbol{f}}_t - \boldsymbol{H}\boldsymbol{f}_t\|_\infty\right) \;\le\; \frac{1}{T}\sum_{t=1}^T \left(\|\widehat{\boldsymbol{u}}_t - \boldsymbol{u}_t\|_\infty + \|\widehat{\boldsymbol{f}}_t - \boldsymbol{H}\boldsymbol{f}_t\|_2\right).$$

Hence by Proposition 1, for any $\eta > 0$, there exist constants $C_u, C_f > 0$ such that, for all T large enough,

$$\|\Delta_2\|_\infty \;\le\; (C_u + C_f)\,(a_{p,T} + b_T), \tag{26}$$

with probability at least $1 - O(T^{-\eta})$.

Bound on Δ_3. For the quantile score $a_\tau(z) = \tau - \mathbb{1}\{z \le 0\}$ we have the pointwise bound

$$\big|a_\tau(w - v) - a_\tau(w)\big| \le \mathbb{1}\{|w| \le |v|\}, \qquad w, v \in \mathbb{R}\ ,$$

so applying this with $w = \varepsilon_t$ and $v = h_t$ gives

$$\big|a_\tau(\varepsilon_t - h_t) - a_\tau(\varepsilon_t)\big| \le \mathbb{1}\{|\varepsilon_t| \le |h_t|\}.$$

For any coordinate $j \in \{1, \dots, p+r+1\}$,

$$\left|(\Delta_3)_j\right| = \left| \frac{1}{T} \sum_{t=1}^{T} \Big\{ a_\tau(\varepsilon_t - h_t) - a_\tau(\varepsilon_t) \Big\} \widetilde{\nu}_{t,j} \right| \le \frac{1}{T} \sum_{t=1}^{T} \mathbb{1}\{|\varepsilon_t| \le |h_t|\} \, |\widetilde{\nu}_{t,j}| \le \frac{1}{T} \sum_{t=1}^{T} \mathbb{1}\{|\varepsilon_t| \le |h_t|\} \, \|\widetilde{\boldsymbol{\nu}}_t\|_\infty.$$

so that

$$\|\Delta_2\|_\infty \le \frac{1}{T} \sum_{t=1}^{T} \mathbb{1}\{|\varepsilon_t| \le |h_t|\} \, \|\widetilde{\boldsymbol{\nu}}_t\|_\infty \le \|\boldsymbol{R}\|_\infty \frac{1}{T} \sum_{t=1}^{T} \mathbb{1}\{|\varepsilon_t| \le |h_t|\} \, \|\boldsymbol{\nu}_t\|_\infty. \tag{27}$$

Define

$$Z_t := \mathbb{1}\{|\varepsilon_t| \le |h_t|\} \, \|\boldsymbol{\nu}_t\|_\infty, \qquad \bar{Z}_t := Z_t - \mathbb{E}[Z_t] \;.$$

Then (27) implies

$$\|\Delta_3\|_\infty \le \|\boldsymbol{R}\|_\infty \left(\frac{1}{T} \sum_{t=1}^{T} \mathbb{E}[Z_t] + \frac{1}{T} \sum_{t=1}^{T} \bar{Z}_t \right) \tag{28}$$

We first control the expectation on the right-hand side. By Assumption A.7, ε_t has a conditional density that is bounded in a neighborhood of zero, i.e., there exists $\bar{f} < \infty$ such that for each t,

$$\mathbb{E}\big[\mathbb{1}\{|\varepsilon_t| \le |h_t|\} \mid \mathcal{I}_t\big] \le 2\bar{f}\,|h_t|.$$

Using the law of iterated expectations, we obtain

$$\begin{aligned} \mathbb{E}[Z_t] = \mathbb{E}\big[\mathbb{1}\{|\varepsilon_t| \le |h_t|\} \, \|\boldsymbol{\nu}_t\|_\infty\big] &= \mathbb{E}\Big[\|\boldsymbol{\nu}_t\|_\infty \, \mathbb{E}\big[\mathbb{1}\{|\varepsilon_t| \le |h_t|\} \mid \mathcal{I}_t\big]\Big] \\ &\le 2\bar{f}\,\mathbb{E}\big[\|\boldsymbol{\nu}_t\|_\infty \, |h_t|\big]. \end{aligned} \tag{29}$$

Next, by Assumption A.9 and the high-probability bound $\|\boldsymbol{H}\|_2 = O_{\mathbb{P}}(1)$ in Proposition B.1 (iv), there exists a constant $C_\nu > 0$ such that, for any $\eta > 0$ and all T large enough, $\mathbb{P}\Big(\|\boldsymbol{\nu}_t\|_\infty \le C_\nu\Big) \ge 1 - O(T^{-\eta})$. On the this event we have $\|\boldsymbol{\nu}_t\|_\infty \, |h_t| \le C_\nu \, |h_t|$, so that (29) and the law of total probability give

$$\mathbb{E}\big[\mathbb{1}\{|\varepsilon_t| \le |h_t|\} \, \|\boldsymbol{\nu}_t\|_\infty\big] \le 2\bar{f} C_\nu \, \mathbb{E}|h_t| + O(T^{-\eta}).$$

Averaging over t and using stationarity, we obtain

$$\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[Z_t] = \mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T}\mathbb{1}\left\{|\varepsilon_t| \le |h_t|\right\}\|\boldsymbol{\nu}_t\|_\infty\right] \le 2\bar{f}C_\nu\,\mathbb{E}|h_t| + O_{\mathbb{P}}(T^{-\eta}). \tag{30}$$

We now control the stochastic fluctuation around the mean $\bar{Z}_t := Z_t - \mathbb{E}[Z_t]$. By construction, $0 \le Z_t \le C_\nu$ holds on the high-probability event defined above. Moreover, the joint process $\{(\varepsilon_t, \boldsymbol{u}_t', \boldsymbol{f}_t')'\}$ is α-mixing by Assumption A.3. Since h_t and $\boldsymbol{\nu}_t$ are measurable functions of $\{(\varepsilon_t, \boldsymbol{u}_t', \boldsymbol{f}_t')'\}$ (and of the estimated quantities), the sequence $\{\bar{Z}_t\}_{t=1}^T$ is a measurable transformation of $(\varepsilon_t, \boldsymbol{u}_t, \boldsymbol{f}_t)$ and hence inherits the same α-mixing coefficients. Therefore, all the conditions of Lemma B.2 are satisfied for the scalar process $\{\bar{Z}_t\}_{t=1}^T$ (i.e., $d = 1$), and we obtain that for any $\eta > 0$ there exists $C_\eta > 0$ such that, for all T sufficiently large,

$$\mathbb{P}\left(\left|\frac{1}{T}\sum_{t=1}^{T}\bar{Z}_t\right| > C_\eta\sqrt{\frac{\log T}{T}}\right) \le O(T^{-\eta}). \tag{31}$$

Combining (24), (28), (30) and (31), and absorbing $O(T^{-\eta})$ terms into the probability bound, we conclude that for any $\eta > 0$, there exists $C_\eta' > 0$ such that, for all T large enough,

$$\|\Delta_3\|_\infty \le \sqrt{2\,c_1/c_r}\,\left(2\bar{f}C_\nu\,\mathbb{E}|h_t| + C_\eta' b_T\right) \quad \text{with probability at least } 1 - O(T^{-\eta}).$$

Finally, using the definition of h_t and the fact that the intercept cancels in $\widehat{\boldsymbol{\nu}}_t - \widetilde{\boldsymbol{\nu}}_t$, we have

$$|h_t| \le \|\widehat{\boldsymbol{u}}_t - \boldsymbol{u}_t\|_\infty\,\|\boldsymbol{\theta}_0\|_1 + \|\widehat{\boldsymbol{f}}_t - \boldsymbol{H}\boldsymbol{f}_t\|_2\,\|\gamma_0\|_2 + \|\boldsymbol{H}'\boldsymbol{H} - \boldsymbol{I}\|_2\,\|\gamma_0\|_2\,\|\boldsymbol{f}_t\|_2.$$

Averaging over t and invoking Proposition 1 together with the moment bounds on $\boldsymbol{f}_t$ (B.1), we obtain

$$\frac{1}{T}\sum_{t=1}^{T}|h_t| \le C_h\left(a_{p,T} + b_T\right) \quad \text{with probability at least } 1 - O(T^{-\eta}) \tag{32}$$

for some constant $C_h > 0$. Substituting this into the previous display yields

$$\|\Delta_3\|_\infty \le C_2\,(a_{p,T} + b_T), \tag{33}$$

with probability at least $1 - O(T^{-\eta})$, where $C_2 > 0$ is a finite constant depending only on $\bar{f}, C_\nu, C_h, C_\theta$, C_γ, c_1 and c_r.

Combining (25), (26), (33), and collecting the constants in C_λ completes the proof. $\square$

Lemma 2 (Restricted Set)**.** *Suppose Assumptions A.1–A.9 hold, and let* $\lambda := 2\,C_\lambda\big(a_{p,T} + b_T\big)$*, where* C_λ *is the constant appearing in Lemma 1. Then, for any* $\eta > 0$*, we have*

$$\widehat{\boldsymbol{\phi}} - \widetilde{\boldsymbol{\phi}} \in \mathcal{A}$$

with probability at least $1 - O(T^{-\eta})$*.*

Proof. We begin by leveraging the convexity of the objective function $\widehat{Q}(\cdot)$. By the definition of convexity, the following inequality holds via the subgradient:

$$\widehat{Q}(\widehat{\boldsymbol{\phi}}) - \widehat{Q}(\widetilde{\boldsymbol{\phi}}) \ge \left(\frac{1}{T}\sum_{t=1}^{T} a_\tau(Y_t - \widehat{\boldsymbol{\nu}}_t{}'\widetilde{\boldsymbol{\phi}})\widehat{\boldsymbol{\nu}}_t{}'\right)\left(\widehat{\boldsymbol{\phi}} - \widetilde{\boldsymbol{\phi}}\right) .$$

Given that $\widehat{\boldsymbol{\phi}}$ is defined as the minimizer of the objective function in (10), substituting the convexity inequality into the optimality condition, we derive the following:

$$\begin{aligned} 0 &\le \widehat{Q}(\widetilde{\boldsymbol{\phi}}) - \widehat{Q}(\widehat{\boldsymbol{\phi}}) + \lambda\|\boldsymbol{\theta}_0\|_1 - \lambda\|\widehat{\boldsymbol{\theta}}\|_1 \\ &\le \left|\left(\frac{1}{T}\sum_{t=1}^{T} a_\tau(Y_t - \widehat{\boldsymbol{\nu}}_t{}'\widetilde{\boldsymbol{\phi}})\widehat{\boldsymbol{\nu}}_t{}'\right)\left(\widehat{\boldsymbol{\phi}} - \widetilde{\boldsymbol{\phi}}\right)\right| + \lambda\|\boldsymbol{\theta}_0\|_1 - \lambda\|\widehat{\boldsymbol{\theta}}\|_1 \\ &\le \left\|\frac{1}{T}\sum_{t=1}^{T} a_\tau(Y_t - \widehat{\boldsymbol{\nu}}_t{}'\widetilde{\boldsymbol{\phi}})\widehat{\boldsymbol{\nu}}_t\right\|_\infty \left\|\widehat{\boldsymbol{\phi}} - \widetilde{\boldsymbol{\phi}}\right\|_1 + \lambda\|\boldsymbol{\theta}_0\|_1 - \lambda\|\widehat{\boldsymbol{\theta}}\|_1 . \end{aligned} \tag{34}$$

In the last inequality, we used Hölder's inequality.

By Lemma 1 and the choice of λ, with probability at least $1 - O(T^{-\eta})$,

$$\left\| \frac{1}{T} \sum_{t=1}^{T} a_\tau (Y_t - \widehat{\boldsymbol{\nu}}_t' \widetilde{\boldsymbol{\phi}}) \widehat{\boldsymbol{\nu}}_t \right\|_\infty \leq \lambda/2.$$

Hence, on this event, we could simplify (34) by canceling out λ from both sides to get

$$-\frac{1}{2}\|\widehat{\boldsymbol{\delta}}\|_1 = -\frac{1}{2}\left\|\widetilde{\boldsymbol{\phi}} - \widehat{\boldsymbol{\phi}}\right\|_1 \leq \|\boldsymbol{\theta}_0\|_1 - \|\widehat{\boldsymbol{\theta}}\|_1 \ .$$

Now, we further split the ℓ_1-norms of $\widehat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\delta}}$ over different subsets of indices. Specifically, we know that $\|\boldsymbol{\theta}_0\|_1 = \|\widetilde{\boldsymbol{\phi}}_{\mathcal{S}_\star}\|_1$ and $\|\widehat{\boldsymbol{\theta}}_{\mathcal{S}_\star}\|_1 = \|\widehat{\boldsymbol{\phi}}_{\mathcal{S}_\star}\|_1$. This allows us to refine the inequality as

$$\begin{aligned}
0 &\leq \|\boldsymbol{\theta}_0\|_1 - \|\widehat{\boldsymbol{\theta}}\|_1 + \frac{1}{2}\|\widehat{\boldsymbol{\delta}}\|_1 \\
&= \|\widetilde{\boldsymbol{\phi}}_{\mathcal{S}_\star}\|_1 - \left(\|\widehat{\boldsymbol{\theta}}_{\mathcal{S}_\star}\|_1 + \|\widehat{\boldsymbol{\theta}}_{\mathcal{S}_\star^C}\|_1\right) + \frac{1}{2}\|\widehat{\boldsymbol{\delta}}_{\mathcal{S}_\diamond}\|_1 + \frac{1}{2}\|\widehat{\boldsymbol{\delta}}_{\mathcal{S}_\diamond^C}\|_1 \\
&= \|\widetilde{\boldsymbol{\phi}}_{\mathcal{S}_\star}\|_1 - \left(\|\widehat{\boldsymbol{\phi}}_{\mathcal{S}_\star}\|_1 + \|\widehat{\boldsymbol{\theta}}_{\mathcal{S}_\star^C}\|_1\right) + \frac{1}{2}\|\widehat{\boldsymbol{\delta}}_{\mathcal{S}_\diamond}\|_1 + \frac{1}{2}\|\widehat{\boldsymbol{\delta}}_{\mathcal{S}_\diamond^C}\|_1
\end{aligned}$$

Using the fact that $\|\widehat{\boldsymbol{\theta}}_{\mathcal{S}_\star^C}\|_1 = \|\widehat{\boldsymbol{\delta}}_{\mathcal{S}_\star^C} + \boldsymbol{\theta}_{0\,\mathcal{S}_\star^C}\|_1 = \|\widehat{\boldsymbol{\delta}}_{\mathcal{S}_\star^C} + \mathbf{0}\|_1 = \|\widehat{\boldsymbol{\delta}}_{\mathcal{S}_\diamond^C}\|_1$, we can apply the triangle inequality to get

$$\begin{aligned}
0 &\leq \|\widetilde{\boldsymbol{\phi}}_{\mathcal{S}_\star}\|_1 - \|\widehat{\boldsymbol{\phi}}_{\mathcal{S}_\star}\|_1 - \|\widehat{\boldsymbol{\delta}}_{\mathcal{S}_\diamond^C}\|_1 + \frac{1}{2}\|\widehat{\boldsymbol{\delta}}_{\mathcal{S}_\diamond}\|_1 + \frac{1}{2}\|\widehat{\boldsymbol{\delta}}_{\mathcal{S}_\diamond^C}\|_1 \\
&\leq \|\widehat{\boldsymbol{\delta}}_{\mathcal{S}_\star}\|_1 + \frac{1}{2}\|\widehat{\boldsymbol{\delta}}_{\mathcal{S}_\diamond}\|_1 - \frac{1}{2}\|\widehat{\boldsymbol{\delta}}_{\mathcal{S}_\diamond^C}\|_1 \\
&\leq \|\widehat{\boldsymbol{\delta}}_{\mathcal{S}_\diamond}\|_1 + \frac{1}{2}\|\widehat{\boldsymbol{\delta}}_{\mathcal{S}_\diamond}\|_1 - \frac{1}{2}\|\widehat{\boldsymbol{\delta}}_{\mathcal{S}_\diamond^C}\|_1 \ ,
\end{aligned}$$

which leads to the conclusion that $\widehat{\boldsymbol{\delta}} \in \mathcal{A}$. □

Lemma 3 (Restricted Eigenvalue)**.** *For $\boldsymbol{\delta} \in \mathcal{A}$, recall the definition of $J^{1/2}(\boldsymbol{\delta})$ as*

$$J^{1/2}(\boldsymbol{\delta}) := \sqrt{\frac{f}{T}\sum_{t=1}^{T} \boldsymbol{\delta}' \mathbb{E}[\boldsymbol{\nu}_t \boldsymbol{\nu}_t'] \boldsymbol{\delta}} \ ,$$

where $\underline{f}$ is defined in A.7 (i). Then there exists a constant $\kappa > 0$ such that

$$\kappa := \inf_{\mathbf{0}\neq\boldsymbol{\delta}\in\mathcal{A}} \frac{J^{1/2}(\boldsymbol{\delta})}{\|\boldsymbol{\delta}\|_2} \ . \tag{35}$$

Proof. For any nonzero vector $\boldsymbol{v} \in \mathbb{R}^{p+r}$, we have that

$$\frac{J(\boldsymbol{v})}{\|\boldsymbol{v}\|_2^2} = \underline{f}\frac{\boldsymbol{v}'\mathbb{E}(\boldsymbol{\nu}_t\boldsymbol{\nu}_t')\boldsymbol{v}}{\|\boldsymbol{v}\|_2^2} \geq \underline{f}\lambda_{\min}(\mathbb{E}(\boldsymbol{\nu}_t\boldsymbol{\nu}_t')) \geq \underline{f}\lambda_{\min}(\mathbb{E}(\boldsymbol{u}_t\boldsymbol{u}_t')) \wedge \underline{f} \geq \underline{f}\kappa_{\boldsymbol{u}} \wedge \underline{f} \ .$$

Thus choosing $\kappa = \sqrt{\underline{f}}\left(\sqrt{\kappa_{\boldsymbol{u}}} \wedge 1\right)$ we obtain (35). □

Lemma 4 (Stability of the Restricted Cone under Block Rotations)**.** *Suppose Assumptions A.1–A.6 hold. Define the rotated parameter set $\mathcal{A}_R := \{\boldsymbol{\xi} = \boldsymbol{R}'\boldsymbol{\delta} : \boldsymbol{\delta} \in \mathcal{A}\}$. Then, for any $\eta > 0$, there exists a positive constant $C_R < \infty$, depending only on C_0 and the eigenvalue parameters (c_1, c_r) defined in A.4, such that, for all T sufficiently large,*

$$\mathcal{A}_R \subseteq \widetilde{\mathcal{A}} := \left\{\boldsymbol{\xi} : \ \|\boldsymbol{\xi}_{\mathcal{S}_\diamond^c}\|_1 \leq C_R\,\|\boldsymbol{\xi}_{\mathcal{S}_\diamond}\|_1\right\}$$

holds with probability at least $1 - O(T^{-\eta})$.

Proof. First, note that $\mathcal{S}_\diamond^c$ lies entirely in the *inactive* idiosyncratic block. On these coordinates, $\boldsymbol{R}$ acts as the identity, hence

$$\boldsymbol{\xi}_{\mathcal{S}_\diamond^c} = \boldsymbol{\delta}_{\mathcal{S}_\diamond^c}. \tag{36}$$

On $\mathcal{S}_\diamond$, $\boldsymbol{R}$ reduces to a lower dimensional block-diagonal matrix

$$\boldsymbol{R}_{\mathcal{S}_\diamond} = \mathrm{diag}\big(1, \mathbf{I}_{|\mathcal{S}_\theta|}, \boldsymbol{H}\big),$$

which is invertible because $\boldsymbol{H}$ is invertible. Then, by Proposition B.1(v), we obtain

$$\begin{aligned}\|\boldsymbol{\delta}_{\mathcal{S}_\diamond}\|_1 &= \|(\boldsymbol{R}_{\mathcal{S}_\diamond}')^{-1}\boldsymbol{\xi}_{\mathcal{S}_\diamond}\|_1 \leq \|(\boldsymbol{R}_{\mathcal{S}_\diamond}')^{-1}\|_1\,\|\boldsymbol{\xi}_{\mathcal{S}_\diamond}\|_1 \\ &\leq \ \left(1 \vee \sqrt{r}\,\|\boldsymbol{H}^{-1}\|_2\right)\|\boldsymbol{\xi}_{\mathcal{S}_\diamond}\|_1 \leq \sqrt{2r\,c_1/c_r}\,\|\boldsymbol{\xi}_{\mathcal{S}_\diamond}\|_1,\end{aligned} \tag{37}$$

which holds with probability at least $1 - O(T^{-\eta})$ for any $\eta > 0$.

Let $\boldsymbol{\xi} \in \mathcal{A}_R$ and write $\boldsymbol{\xi} = \boldsymbol{R}'\boldsymbol{\delta}$ with $\boldsymbol{\delta} \in \mathcal{A}$. Using (36) and the cone constraint on $\boldsymbol{\delta}$ in (12), on the event in (37) we have

$$\|\boldsymbol{\xi}_{\mathcal{S}_\diamond^c}\|_1 = \|\boldsymbol{\delta}_{\mathcal{S}_\diamond^c}\|_1 \leq C_0 \, \|\boldsymbol{\delta}_{\mathcal{S}_\diamond}\|_1 \leq C_0 \sqrt{2r \, c_1/c_r} \, \|\boldsymbol{\xi}_{\mathcal{S}_\diamond}\|_1.$$

Set $C_R := C_0 \sqrt{2r \, c_1/c_r}$ to obtain

$$\boldsymbol{\xi} \in \left\{ \boldsymbol{z} : \|\boldsymbol{z}_{\mathcal{S}_\diamond^c}\|_1 \leq C_R \, \|\boldsymbol{z}_{\mathcal{S}_\diamond}\|_1 \right\} = \widetilde{\mathcal{A}} \, .$$

Since $\boldsymbol{\xi}$ is arbitrary in $\mathcal{A}_R$, the inclusion $\mathcal{A}_R \subseteq \widetilde{\mathcal{A}}$ follows. □

Lemma 5 (Identifiability Relations)**.** *Suppose Assumptions A.1–A.8 hold. Then, for all $\boldsymbol{\delta} \in \mathcal{A}$, the following bounds hold:*

(i) $\|\boldsymbol{\delta}\|_1 \ \leq \ \dfrac{(1 + C_0)\sqrt{1 + s_\tau + r}}{\kappa} \, J^{1/2}(\boldsymbol{\delta}).$

(ii) $Q(\boldsymbol{\phi}_0 + \boldsymbol{R}'\boldsymbol{\delta}) - Q(\boldsymbol{\phi}_0) \ \geq \ \dfrac{1}{4}\left(J^{1/2}(\boldsymbol{\delta})\right)^2 \ \wedge \ \left\{q \, J^{1/2}(\boldsymbol{\delta})\right\}.$

Proof. (i) Since $\boldsymbol{\delta} \in \mathcal{A}$ and $|\mathcal{S}_\diamond| = 1 + s_\tau + r$, the following inequalities hold:

$$\begin{aligned}\|\boldsymbol{\delta}\|_1 &\leq (1 + C_0)\|\boldsymbol{\delta}_{\mathcal{S}_\diamond}\|_1 \leq (1 + C_0) \cdot \sqrt{1 + s_\tau + r}\|\boldsymbol{\delta}_{\mathcal{S}_\diamond}\|_2 \\ &\leq (1 + C_0) \cdot \sqrt{1 + s_\tau + r}\|\boldsymbol{\delta}\|_2 \leq \frac{(1 + C_0)\sqrt{1 + s_\tau + r}}{\kappa} J^{1/2}(\boldsymbol{\delta}),\end{aligned}$$

where the last inequality follows from Lemma 3.

(ii) As a preliminary step, we first establish a lower bound for the objective increment induced by a direct perturbation, $Q(\boldsymbol{\phi}_0 + \boldsymbol{\delta}) - Q(\boldsymbol{\phi}_0)$. Proceeding similarly to Lemma 4 of Belloni and Chernozhukov (2011), we begin by defining a maximal radius within which the criterion function $Q(\cdot)$ is bounded below by a quadratic form. This is key to establishing local quadratic approximation properties of $Q(\cdot)$ around the true parameter value. Specifically, we define the maximal radius $r_{\mathcal{A}}$ as follows:

$$r_{\mathcal{A}} = \sup_{r_0} \left\{ r_0 : Q(\boldsymbol{\phi}_0 + \widetilde{\boldsymbol{\delta}}) - Q(\boldsymbol{\phi}_0) \geq \frac{1}{4}(J^{1/2}(\widetilde{\boldsymbol{\delta}}))^2,\ \forall \widetilde{\boldsymbol{\delta}} \in \mathcal{A},\ J^{1/2}(\widetilde{\boldsymbol{\delta}}) \leq r_0 \right\} .$$

By the convexity of $Q(\cdot)$ and the definition of $r_{\mathcal{A}}$, we have that

$$\begin{aligned}
& Q(\boldsymbol{\phi}_0 + \boldsymbol{\delta}) - Q(\boldsymbol{\phi}_0) \\
\geq\ & \frac{1}{4}(J^{1/2}(\boldsymbol{\delta}))^2 \wedge \left\{ \frac{(J^{1/2}(\boldsymbol{\delta})}{r_{\mathcal{A}}} \cdot \inf_{\widetilde{\boldsymbol{\delta}} \in \mathcal{A},\, J^{1/2}(\widetilde{\boldsymbol{\delta}}) \geq r_{\mathcal{A}}} Q(\boldsymbol{\phi}_0 + \widetilde{\boldsymbol{\delta}}) - Q(\boldsymbol{\phi}_0) \right\} \\
\geq\ & \frac{1}{4}(J^{1/2}(\boldsymbol{\delta}))^2 \wedge \left\{ \frac{(J^{1/2}(\boldsymbol{\delta})}{r_{\mathcal{A}}} \frac{r_{\mathcal{A}}^2}{4} \right\} \\
\geq\ & \frac{1}{4}(J^{1/2}(\boldsymbol{\delta}))^2 \wedge \left\{ q J^{1/2}(\boldsymbol{\delta}) \right\} , \text{for any } \boldsymbol{\delta} \in \mathcal{A} ,
\end{aligned}$$

where the last inequality follows from the fact that $r_{\mathcal{A}} \geq 4q$. It therefore remains to establish that $r_{\mathcal{A}} \geq 4q$.

To this end, we apply Knight's identity (Knight, 1998), which provides a crucial decomposition for quantile loss function. For any two scalars w and v, Knight's identity states that:

$$\rho_\tau(w - v) - \rho_\tau(w) = -v(\tau - \mathbb{1}_{\{w \leq 0\}}) + \int_0^v \left(\mathbb{1}_{\{w \leq s\}} - \mathbb{1}_{\{w \leq 0\}} \right) ds . \qquad (38)$$

Setting $w = Y_t - \boldsymbol{\nu}_t' \boldsymbol{\phi}_0 = \varepsilon_t$ and $v = \boldsymbol{\nu}_t' \boldsymbol{\delta}$, the expectation of the first term vanishes since $\mathbb{E}[\mathbb{1}_{\{\varepsilon_t \leq 0\}}] = \mathbb{E}[\mathbb{E}[\mathbb{1}_{\{\varepsilon_t \leq 0\}} | \mathcal{I}_t]] = \tau$ by Assumption A.7.

Next, let $F_{Y_t | \boldsymbol{f}_t, \boldsymbol{u}_t}$ denote the conditional distribution of Y_t given $(\boldsymbol{f}_t, \boldsymbol{u}_t)$. Using the law of iterated expectations and a second-order mean value expansion, we expand the term of interest as follows:

$$\begin{aligned}
Q(\boldsymbol{\phi}_0 + \boldsymbol{\delta}) - Q(\boldsymbol{\phi}_0) &= \frac{1}{T} \sum_{t=1}^T \mathbb{E} \left[\int_0^{\boldsymbol{\nu}_t' \boldsymbol{\delta}} \left(F_{Y_t | \boldsymbol{f}_t, \boldsymbol{u}_t}(\boldsymbol{\nu}_t' \boldsymbol{\phi}_0 + s) - F_{Y_t|}(\boldsymbol{\nu}_t' \boldsymbol{\phi}_0) \right) ds \right] \\
= \frac{1}{T} \sum_{t=1}^T \mathbb{E} \left[\int_0^{\boldsymbol{\nu}_t' \boldsymbol{\delta}} \left(s f_{Y_t | \boldsymbol{f}_t, \boldsymbol{u}_t}(\boldsymbol{\nu}_t' \boldsymbol{\phi}_0) + \frac{s^2}{2} f'_{Y_t | \boldsymbol{f}_t, \boldsymbol{u}_t}(\boldsymbol{\nu}_t' \boldsymbol{\phi}_0 + \widetilde{s}) \right) ds \right] &
\end{aligned}$$

where $\widetilde{s} \in [0, s]$. By Assumption A.7 and the definition of $J(\cdot)$, this expansion yields a leading quadratic term and a cubic remainder:

$$
\begin{aligned}
&Q(\boldsymbol{\phi}_0 + \boldsymbol{\delta}) - Q(\boldsymbol{\phi}_0) \\
&\geq \frac{\underline{f}}{2T} \sum_{t=1}^{T} \mathbb{E}[\boldsymbol{\delta}' \boldsymbol{\nu}_t \boldsymbol{\nu}_t' \boldsymbol{\delta}] - \frac{\overline{f'}}{6T} \sum_{t=1}^{T} \mathbb{E}[|\boldsymbol{\nu}_t' \boldsymbol{\delta}|^3] \\
&= \frac{1}{4}(J^{1/2}(\boldsymbol{\delta}))^2 + \frac{1}{4}(J^{1/2}(\boldsymbol{\delta}))^2 - \frac{\overline{f'}}{6T} \sum_{t=1}^{T} \mathbb{E}[|\boldsymbol{\nu}_t' \boldsymbol{\delta}|^3],
\end{aligned}
$$

Now, consider for $\boldsymbol{\delta} \in \mathcal{A}$, if $J^{1/2}(\boldsymbol{\delta}) < 4q$, we obtain the following inequality:

$$
\sqrt{\frac{\underline{f}}{T} \sum_{t=1}^{T} \boldsymbol{\delta}' \mathbb{E}[\boldsymbol{\nu}_t \boldsymbol{\nu}_t'] \boldsymbol{\delta}} \leq 4 \cdot \frac{3}{8} \frac{\underline{f}^{3/2}}{\overline{f'}} \inf_{\mathbf{0} \neq \boldsymbol{\delta} \in \mathcal{A}} \frac{(\mathbb{E}(\frac{1}{T} \sum_{t=1}^{T} (\boldsymbol{\nu}_t' \boldsymbol{\delta})^2))^{3/2}}{\mathbb{E}(\frac{1}{T} \sum_{t=1}^{T} |\boldsymbol{\nu}_t' \boldsymbol{\delta}|^3)},
$$

This immediately yields the following result:

$$
\frac{\overline{f'}}{6} \mathbb{E}(\frac{1}{T} \sum_{t=1}^{T} |\boldsymbol{\nu}_t' \boldsymbol{\delta}|^3) \leq \frac{1}{4} \underline{f}\, \mathbb{E}(\frac{1}{T} \sum_{t=1}^{T} (\boldsymbol{\nu}_t' \boldsymbol{\delta})^2) = \frac{1}{4}(J^{1/2}(\boldsymbol{\delta}))^2.
$$

Thus, if $J^{1/2}(\boldsymbol{\delta}) < 4q$, we have $Q(\boldsymbol{\phi}_0 + \boldsymbol{\delta}) - Q(\boldsymbol{\phi}_0) \geq \frac{1}{4}(J^{1/2}(\boldsymbol{\delta}))^2$, which implies that $J^{1/2}(\boldsymbol{\delta}) \leq r_{\mathcal{A}}$. Consequently, we conclude that $r_{\mathcal{A}} \geq 4q$.

In the above derivation we work with perturbations of the form $\boldsymbol{\phi}_0 + \boldsymbol{\delta}$. We now show that the same bound continues to hold when $\boldsymbol{\delta}$ is replaced by its rotated version $\boldsymbol{R}'\boldsymbol{\delta}$ up to some constants. Recall the rotated parameter set $\mathcal{A}_R := \{\boldsymbol{\xi} = \boldsymbol{R}'\boldsymbol{\delta} : \ \boldsymbol{\delta} \in \mathcal{A}\}$. By Lemma 4, under Assumptions A.1–A.6 we have, for any $\eta > 0$ and all T sufficiently large,

$$
\mathcal{A}_R \subseteq \widetilde{\mathcal{A}} := \{\boldsymbol{\xi} : \ \|\boldsymbol{\xi}_{\mathcal{S}_\diamond^c}\|_1 \leq C_R \|\boldsymbol{\xi}_{\mathcal{S}_\diamond}\|_1\}
$$

with probability at least $1 - O(T^{-\eta})$, for some constant $C_R < \infty$ depending only on C_0 and the eigenvalue bounds in Assumption A.4. In particular, $\mathcal{A}_R$ is (with high probability) contained in an ℓ_1-cone of the same form as $\mathcal{A}$ only up to a constant.

Assumption A.8 and the curvature argument above only use the cone through this ℓ_1-constraint and the eigenvalue bounds. Hence the same reasoning applies to the rotated cone $\mathcal{A}_R$: there exists $q_R > 0$ such that, for all $\boldsymbol{\xi} \in \mathcal{A}_R$,

$$Q(\boldsymbol{\phi}_0 + \boldsymbol{\xi}) - Q(\boldsymbol{\phi}_0) \;\geq\; \frac{1}{4}\left(J^{1/2}(\boldsymbol{\xi})\right)^2 \;\wedge\; \left\{q_R\, J^{1/2}(\boldsymbol{\xi})\right\},$$

with probability at least $1 - O(T^{-\eta})$. Taking $\boldsymbol{\xi} = \boldsymbol{R}'\boldsymbol{\delta}$ with $\boldsymbol{\delta} \in \mathcal{A}$ yields

$$Q(\boldsymbol{\phi}_0 + \boldsymbol{R}'\boldsymbol{\delta}) - Q(\boldsymbol{\phi}_0) \;\geq\; \frac{1}{4}\left(J^{1/2}(\boldsymbol{R}'\boldsymbol{\delta})\right)^2 \;\wedge\; \left\{q_R\, J^{1/2}(\boldsymbol{R}'\boldsymbol{\delta})\right\}.$$

Finally, under our normalization of the factor block and the approximate orthogonality of $\boldsymbol{H}$ (B.1), $J^{1/2}(\boldsymbol{R}'\boldsymbol{\delta})$ and $J^{1/2}(\boldsymbol{\delta})$ differ only by a multiplicative factor $1 + o(1)$, so the above display can be equivalently written as

$$Q(\boldsymbol{\phi}_0 + \boldsymbol{R}'\boldsymbol{\delta}) - Q(\boldsymbol{\phi}_0) \;\geq\; c_1\left(J^{1/2}(\boldsymbol{\delta})\right)^2 \;\wedge\; \left\{q_*\, J^{1/2}(\boldsymbol{\delta})\right\},$$

for some positive constants $c_1 > 0$ and $q_* > 0$ independent of (p, T). For notational simplicity, we keep denoting these constants by $1/4$ and q in what follows. $\square$

Lemma 6 (Control of Empirical Process Error). *Suppose Assumptions A.1–A.6 and A.9 hold. Then, for any $\eta > 0$, there exist constants $C_\eta, C_0, C_{\text{rem}}, \kappa > 0$ and the eigenvalue parameters (c_1, c_r) defined in A.4 such that, with probability at least $1 - O(T^{-\eta})$ and uniformly over $\boldsymbol{\delta} \in \mathcal{A}$ with $J^{1/2}(\boldsymbol{\delta}) \leq \omega$,*

$$\epsilon_1 := \sup_{\boldsymbol{\delta}\in\mathcal{A},\ |J^{1/2}(\boldsymbol{\delta})|=\omega} \left|\overline{Q}(\boldsymbol{\phi}_0 + \boldsymbol{R}'\boldsymbol{\delta}) - Q(\boldsymbol{\phi}_0 + \boldsymbol{R}'\boldsymbol{\delta}) - (\overline{Q}(\boldsymbol{\phi}_0) - Q(\boldsymbol{\phi}_0))\right| \leq C_{\text{ep}}\, \frac{\sqrt{1 + s_\tau + r}}{\kappa}\, b_T\, \omega\,,$$

where $C_{\text{ep}} := C_\eta\sqrt{2c_1/c_r}(1 + C_0) + C_{\text{rem}}$.

Proof. Step 1: Basic decomposition. For any $\boldsymbol{\delta} \in \mathcal{A}$ define $\Delta_t(\boldsymbol{\delta}) := \boldsymbol{\nu}_t'\boldsymbol{R}'\boldsymbol{\delta}$. Writing $\varepsilon_t := Y_t - \boldsymbol{\nu}_t'\boldsymbol{\phi}_0$, we have $Y_t - \boldsymbol{\nu}_t'(\boldsymbol{\phi}_0 + \boldsymbol{R}'\boldsymbol{\delta}) = \varepsilon_t - \Delta_t(\boldsymbol{\delta})$. Applying Knight's identity (38) with $w = \varepsilon_t$ and $v = \Delta_t(\boldsymbol{\delta})$ yields

$$\rho_\tau(\varepsilon_t - \Delta_t(\boldsymbol{\delta})) - \rho_\tau(\varepsilon_t) = -\Delta_t(\boldsymbol{\delta})\, a_\tau(\varepsilon_t) + r_t(\boldsymbol{\delta}),$$

where

$$a_\tau(z) := \tau - \mathbf{1}\{z \le 0\}, \qquad r_t(\boldsymbol{\delta}) := \int_0^{\Delta_t(\boldsymbol{\delta})} \big(\mathbf{1}\{\varepsilon_t \le s\} - \mathbf{1}\{\varepsilon_t \le 0\}\big)\, ds.$$

By definition of $\overline{Q}(\cdot)$ and $Q(\cdot)$,

$$\overline{Q}(\boldsymbol{\phi}_0 + \boldsymbol{R}'\boldsymbol{\delta}) - \overline{Q}(\boldsymbol{\phi}_0) = \frac{1}{T}\sum_{t=1}^{T} \big\{\rho_\tau(\varepsilon_t - \Delta_t(\boldsymbol{\delta})) - \rho_\tau(\varepsilon_t)\big\},$$

and the same expansion holds for $Q(\boldsymbol{\phi}_0 + \boldsymbol{R}'\boldsymbol{\delta}) - Q(\boldsymbol{\phi}_0)$ after taking expectations. Hence,

$$\begin{aligned} &\overline{Q}(\boldsymbol{\phi}_0 + \boldsymbol{R}'\boldsymbol{\delta}) - Q(\boldsymbol{\phi}_0 + \boldsymbol{R}'\boldsymbol{\delta}) - \big\{\overline{Q}(\boldsymbol{\phi}_0) - Q(\boldsymbol{\phi}_0)\big\} \\ &\quad = -\frac{1}{T}\sum_{t=1}^{T}\Big(\Delta_t(\boldsymbol{\delta})\, a_\tau(\varepsilon_t) - \mathbb{E}[\Delta_t(\boldsymbol{\delta})\, a_\tau(\varepsilon_t)]\Big) + \frac{1}{T}\sum_{t=1}^{T}\Big(r_t(\boldsymbol{\delta}) - \mathbb{E}[r_t(\boldsymbol{\delta})]\Big). \end{aligned}$$

Using the conditional quantile restriction $\mathbb{P}(\varepsilon_t \le 0 \mid \mathcal{I}_t) = \tau$ in Assumption A.7, we have $\mathbb{E}[a_\tau(\varepsilon_t) \mid \mathcal{I}_t] = 0$. Let $\mathcal{F}_X := \sigma\big(\{\boldsymbol{X}_s, \boldsymbol{f}_t\}_{s=1}^{T}\big)$ denote the σ-algebra generated by the regressors over the whole sample. By construction, both $\boldsymbol{\nu}_t$ and the rotation matrix $\boldsymbol{R}$ are $\mathcal{F}_X$-measurable, hence $\Delta_t(\boldsymbol{\delta}) := \boldsymbol{\nu}_t'\boldsymbol{\delta} = \boldsymbol{\nu}_t'\boldsymbol{R}'\boldsymbol{\delta}$ is also $\mathcal{F}_X$-measurable. Moreover, since $\mathcal{I}_t$ is measurable with respect to $\mathcal{F}_X$, the tower property implies

$$\mathbb{E}\big[a_\tau(\varepsilon_t) \mid \mathcal{F}_X\big] = \mathbb{E}\big[\mathbb{E}\{a_\tau(\varepsilon_t) \mid \mathcal{I}_t\} \,\big|\, \mathcal{F}_X\big] = 0.$$

Therefore,

$$\mathbb{E}\big[\Delta_t(\boldsymbol{\delta})\, a_\tau(\varepsilon_t)\big] = \mathbb{E}\Big[\mathbb{E}\big\{\Delta_t(\boldsymbol{\delta})\, a_\tau(\varepsilon_t) \mid \mathcal{F}_X\big\}\Big] = \mathbb{E}\Big[\Delta_t(\boldsymbol{\delta})\, \mathbb{E}\big\{a_\tau(\varepsilon_t) \mid \mathcal{F}_X\big\}\Big] = 0.$$

Consequently,

$$\epsilon_1 \;\le\; \sup_{\boldsymbol{\delta}\in\mathcal{A}:\, J^{1/2}(\boldsymbol{\delta})=\omega} \big|G_{1,T}(\boldsymbol{\delta})\big| + \sup_{\boldsymbol{\delta}\in\mathcal{A}:\, J^{1/2}(\boldsymbol{\delta})=\omega} \big|G_{2,T}(\boldsymbol{\delta})\big|,$$

where

$$G_{1,T}(\boldsymbol{\delta}) := -\frac{1}{T}\sum_{t=1}^{T} \Delta_t(\boldsymbol{\delta})\, a_\tau(\varepsilon_t), \qquad G_{2,T}(\boldsymbol{\delta}) := \frac{1}{T}\sum_{t=1}^{T} \big\{ r_t(\boldsymbol{\delta}) - \mathbb{E}[r_t(\boldsymbol{\delta})] \big\}.$$

Step 2: Bounding the linear term $G_{1,T}(\boldsymbol{\delta})$. Note that

$$G_{1,T}(\boldsymbol{\delta}) = -\,\boldsymbol{\delta}'\,\boldsymbol{R}\Big(\frac{1}{T}\sum_{t=1}^{T} a_\tau(\varepsilon_t)\,\boldsymbol{\nu}_t\Big) =: -\,\boldsymbol{\delta}'\,\boldsymbol{R}\Big(\frac{1}{T}\sum_{t=1}^{T} \boldsymbol{Z}_t\Big),$$

with $\boldsymbol{Z}_t := a_\tau(\varepsilon_t)\boldsymbol{\nu}_t$. Proof of Lemma 1 shows that for any $\eta > 0$, $\mathbb{P}\bigg(\Big\|\frac{1}{T}\sum_{t=1}^{T} \boldsymbol{Z}_t\Big\|_\infty > C_\eta\sqrt{\frac{\log T}{T}}\bigg) \leq O(T^{-\eta})$, for some finite constant $C_\eta > 0$. The same proof also shows that $\|\boldsymbol{R}\|_\infty = \max\{1, \|\boldsymbol{H}\|_\infty\} \leq \sqrt{2c_1/c_r}$ with probability at least $1 - O(T^{-\eta})$.

By Hölder's inequality,

$$|G_{1,T}(\boldsymbol{\delta})| \leq \|\boldsymbol{\delta}\|_1\,\|\boldsymbol{R}\|_\infty \left\|\frac{1}{T}\sum_{t=1}^{T} \boldsymbol{Z}_t\right\|_\infty \leq C_\eta\sqrt{2c_1/c_r}\,\|\boldsymbol{\delta}\|_1\, b_T, \qquad b_T := \sqrt{\frac{\log T}{T}}.$$

Finally, part (i) of Lemma 5 ensures that, for all $\boldsymbol{\delta} \in \mathcal{A}$,

$$\|\boldsymbol{\delta}\|_1 \leq \kappa^{-1}(1 + C_0)\sqrt{1 + s_\tau + r}\; J^{1/2}(\boldsymbol{\delta}).$$

Therefore, uniformly over $\boldsymbol{\delta} \in \mathcal{A}$ with $J^{1/2}(\boldsymbol{\delta}) = \omega$,

$$|G_{1,T}(\boldsymbol{\delta})| \leq \frac{C_\eta\sqrt{2c_1/c_r}\,(1 + C_0)\sqrt{1 + s_\tau + r}}{\kappa}\; b_T\,\omega,$$

with probability at least $1 - O(T^{-\eta})$.

Step 3: Bounding the remainder term $G_{2,T}(\boldsymbol{\delta})$. By construction,

$$r_t(\boldsymbol{\delta}) = \int_0^{\Delta_t(\boldsymbol{\delta})} \big(\mathbf{1}\{\varepsilon_t \leq s\} - \mathbf{1}\{\varepsilon_t \leq 0\}\big)\, ds$$

and therefore $|r_t(\boldsymbol{\delta})| \leq |\Delta_t(\boldsymbol{\delta})|$.

Let $F_{\varepsilon_t|\mathcal{I}_t}(s) := \mathbb{P}(\varepsilon_t \leq s \mid \mathcal{I}_t)$ denote the conditional c.d.f. of ε_t. By the definition of $r_t(\boldsymbol{\delta})$ and by interchanging conditional expectation and integration ($\Delta_t(\boldsymbol{\delta})$ is $\mathcal{I}_t$-measurable),

$$\mathbb{E}\big[r_t(\boldsymbol{\delta}) \mid \mathcal{I}_t\big] = \int_0^{\Delta_t(\boldsymbol{\delta})} \big(F_{\varepsilon_t|\mathcal{I}_t}(s) - F_{\varepsilon_t|\mathcal{I}_t}(0)\big)\, ds. \tag{39}$$

We first consider the case $\Delta_t(\boldsymbol{\delta}) \geq 0$. For $0 \leq s \leq \Delta_t(\boldsymbol{\delta})$ with $|\Delta_t(\boldsymbol{\delta})| \leq \delta_0$, the existence of the conditional density and Assumption A.7(ii) imply that

$$F_{\varepsilon_t|\mathcal{I}_t}(s) - F_{\varepsilon_t|\mathcal{I}_t}(0) = F_{Y_t|\mathcal{I}_t,\boldsymbol{\phi}_0(\tau)}(s) - F_{Y_t|\mathcal{I}_t,\boldsymbol{\phi}_0(\tau)}(0) = \int_0^s f_{Y_t|\mathcal{I}_t,\boldsymbol{\phi}_0(\tau)}(u)\, du,$$

where $f_{Y_t|\mathcal{I}_t,\boldsymbol{\phi}_0(\tau)}(\cdot)$ is the conditional density of Y_t given $\mathcal{I}_t$ and $\boldsymbol{\phi}_0(\tau)$. Assumption A.7(ii) further guarantees that $f_{Y_t|\mathcal{I}_t,\boldsymbol{\phi}_0(\tau)}(u) \leq \overline{f}$ for all $|u| \leq \delta_0$. Hence,

$$\big|F_{\varepsilon_t|\mathcal{I}_t}(s) - F_{\varepsilon_t|\mathcal{I}_t}(0)\big| \leq \int_0^s \overline{f}\, du = \overline{f}\, s, \qquad 0 \leq s \leq \Delta_t(\boldsymbol{\delta}).$$

Substituting this bound into (39) yields

$$\big|\mathbb{E}[r_t(\boldsymbol{\delta}) \mid \mathcal{I}_t]\big| \leq \int_0^{\Delta_t(\boldsymbol{\delta})} \overline{f}\, s\, ds = \frac{\overline{f}}{2}\,\big(\Delta_t(\boldsymbol{\delta})\big)^2.$$

The case $\Delta_t(\boldsymbol{\delta}) < 0$ follows by symmetry, since $\big|F_{\varepsilon_t|\mathcal{I}_t}(s) - F_{\varepsilon_t|\mathcal{I}_t}(0)\big| \leq \overline{f}\,|s|$ holds for all s with $|s| \leq \delta_0$. Thus, $\big|\mathbb{E}[r_t(\boldsymbol{\delta}) \mid \mathcal{I}_t]\big| \leq \overline{f}|\Delta_t(\boldsymbol{\delta})|^2/2$.

Using Assumption A.2 (sub-Gaussianity of $(\boldsymbol{f}_t, \boldsymbol{u}_t)$), the random variable $\Delta_t(\boldsymbol{\delta}) = \boldsymbol{\nu}_t'\boldsymbol{R}'\boldsymbol{\delta}$ is sub-Gaussian with parameter bounded by a constant multiple of $\|\boldsymbol{R}'\boldsymbol{\delta}\|_2$. Hence $r_t(\boldsymbol{\delta}) - \mathbb{E}[r_t(\boldsymbol{\delta})]$ is sub-exponential, and the process $\{r_t(\boldsymbol{\delta}) - \mathbb{E}[r_t(\boldsymbol{\delta})]\}_{t=1}^T$ is again geometrically α-mixing by Assumption A.3. Applying Lemma B.2 with dimension $d = 1$ to

$$\widetilde{Z}_T(\boldsymbol{\delta}) := \frac{1}{T}\sum_{t=1}^T \big\{r_t(\boldsymbol{\delta}) - \mathbb{E}[r_t(\boldsymbol{\delta})]\big\}$$

and using the bound above on the envelope, we obtain, for any fixed $\boldsymbol{\delta}$ and any $\eta > 0$,

$$\mathbb{P}\Big(\big|\widetilde{Z}_T(\boldsymbol{\delta})\big| > C_r\, b_T\Big) \leq O(T^{-\eta}),$$

for some finite $C_r > 0$.

To make this bound *uniform* over $\boldsymbol{\delta} \in \mathcal{A}$ with $J^{1/2}(\boldsymbol{\delta}) \leq \omega$, we proceed via a standard peeling and ε-net argument. First, the cone constraint $\mathcal{A} = \{\boldsymbol{\delta} : \|\boldsymbol{\delta}_{S_\diamond^c}\|_1 \leq C_0\|\boldsymbol{\delta}_{S_\diamond}\|_1\}$ implies that the effective dimension of $\boldsymbol{\delta}$ is of order $|S_\diamond| = 1 + s_\tau + r$, so that for any radius level $r_0 > 0$ the Euclidean ball $\{\boldsymbol{\delta} \in \mathcal{A} : \|\boldsymbol{\delta}\|_2 \leq r_0\}$ admits a 1/2-net of cardinality at most $\exp\{C(1+s_\tau+r)\}$ for some constant $C > 0$. Since $J^{1/2}(\boldsymbol{\delta}) \geq \kappa\|\boldsymbol{\delta}\|_2$ by Lemma 3, the restriction $J^{1/2}(\boldsymbol{\delta}) \leq \omega$ is equivalent to $\|\boldsymbol{\delta}\|_2 \leq \omega/\kappa$. A union bound over this 1/2-net and a standard Lipschitz-continuity argument in $\boldsymbol{\delta}$ (using again the Lipschitz property of $r_t(\boldsymbol{\delta})$ in $\Delta_t(\boldsymbol{\delta})$ and the sub-Gaussianity of $\Delta_t(\boldsymbol{\delta})$) then yields

$$\sup_{\boldsymbol{\delta}\in\mathcal{A}:\, J^{1/2}(\boldsymbol{\delta})\leq\omega} \left|G_{2,T}(\boldsymbol{\delta})\right| \leq C_{\text{rem}} \frac{\sqrt{1+s_\tau+r}}{\kappa}\, b_T\, \omega, \tag{40}$$

with probability at least $1 - O(T^{-\eta})$, for some finite constant $C_{\text{rem}} > 0$.

Step 4: Collecting the bounds. Combining the uniform bound for $G_{1,T}(\boldsymbol{\delta})$ from Step 2 and the uniform bound for $G_{2,T}(\boldsymbol{\delta})$ from Step 3, we obtain that, for any $\eta > 0$ and all T sufficiently large,

$$\epsilon_1 \leq \left(C_\eta\sqrt{2c_1/c_r}(1+C_0) + C_{\text{rem}}\right) \frac{\sqrt{1+s_\tau+r}}{\kappa}\, b_T\, \omega$$

holds with probability at least $1-O(T^{-\eta})$. Setting $C_{\text{ep}} := C_\eta\sqrt{2c_1/c_r}(1+C_0)+C_{\text{rem}}$ and absorbing constants into $O(T^{-\eta})$ completes the proof. □

Remark An alternative proof using symmetrization and Rademacher processes, as in Belloni and Chernozhukov (2011) and Feng (2023), is also valid. We choose not to follow that route because, after applying Knight's identity, the dependence on $\boldsymbol{R}'\boldsymbol{\delta}$ is purely linear. This allows the supremum over $\boldsymbol{\delta} \in \mathcal{A}$ to be separated from the sample average and handled directly by Bernstein-type concentration bounds together with a simple covering argument. This yields the same stochastic order as the Rademacher approach while keeping the proof shorter and closely aligned with the geometry of $J^{1/2}(\boldsymbol{\delta})$ on the cone $\mathcal{A}$.

Lemma 7 (Control of Regressor Estimation Error)**.** *Suppose Assumptions A.1–A.6 and A.9 hold. Then, for any $\eta > 0$, there exist constants $C_f, C_u, C_\gamma, C_\theta, C_{H2}, C_0, \kappa > 0$ such that, with probability at least $1 - O(T^{-\eta})$, the following holds uniformly over $\boldsymbol{\delta} \in \mathcal{A}$ with $J^{1/2}(\boldsymbol{\delta}) \le \omega$,*

$$\begin{aligned}\epsilon_2 := & \sup_{\boldsymbol{\delta}\in\mathcal{A},\ |J^{1/2}(\boldsymbol{\delta})|\le\omega} \Big| \underbrace{\widehat{Q}(\widetilde{\boldsymbol{\phi}}+\boldsymbol{\delta}) - \widehat{Q}(\widetilde{\boldsymbol{\phi}})}_{\mathsf{A}(\boldsymbol{\delta})} - \underbrace{\left(\overline{Q}(\boldsymbol{\phi}_0 + \boldsymbol{R}'\boldsymbol{\delta}) - \overline{Q}(\boldsymbol{\phi}_0)\right)}_{\mathsf{B}(\boldsymbol{\delta})} \Big| \\ \le\ & C_{\text{est}} \frac{\sqrt{1+s_\tau+r}}{\kappa} (a_{p,T} + b_T)\, \omega, \end{aligned} \tag{41}$$

where $C_{\text{est}} := 2(C_u + C_f)(1 + C_0) + \bar{f}\,(\sqrt{2rc_1/c_2})(1 + C_0)C_\eta + C_{\text{rem}}$

Proof. Recall the definitions of the infeasible empirical objective in (7) and the feasible empirical objective in (8). For all $\boldsymbol{\delta} \in \mathcal{A}$, define

$$\mathsf{A}(\boldsymbol{\delta}) = \widehat{Q}(\widetilde{\boldsymbol{\phi}}+\boldsymbol{\delta}) - \widehat{Q}(\widetilde{\boldsymbol{\phi}}) = \frac{1}{T}\sum_{t=1}^{T}\left\{\rho_\tau\Big(Y_t - \widehat{\boldsymbol{\nu}}_t'(\widetilde{\boldsymbol{\phi}}+\boldsymbol{\delta})\Big) - \rho_\tau\Big(Y_t - \widehat{\boldsymbol{\nu}}_t'\widetilde{\boldsymbol{\phi}}\Big)\right\}, \tag{42}$$

$$\mathsf{B}(\boldsymbol{\delta}) = \overline{Q}_\tau(\boldsymbol{\phi}_0 + \boldsymbol{R}'\boldsymbol{\delta}) - \overline{Q}_\tau(\boldsymbol{\phi}_0) = \frac{1}{T}\sum_{t=1}^{T}\Big\{\rho_\tau(Y_t - \boldsymbol{\nu}_t'(\boldsymbol{\phi}_0 + \boldsymbol{R}'\boldsymbol{\delta})) - \rho_\tau(Y_t - \boldsymbol{\nu}_t'\boldsymbol{\phi}_0)\Big\}, \tag{43}$$

where $\widehat{\boldsymbol{\nu}}_t = (1, \widehat{\boldsymbol{u}}_t', \widehat{\boldsymbol{f}}_t')'$ and $\boldsymbol{\nu}_t = (1, \boldsymbol{u}_t', \boldsymbol{f}_t')'$.

Step 1: Basic decomposition. Applying Knight's identity (38) to (42) with $w = Y_t - \widehat{\boldsymbol{\nu}}_t'\widetilde{\boldsymbol{\phi}}$ and $v = \widehat{\boldsymbol{\nu}}_t'\boldsymbol{\delta}$, we obtain

$$\mathsf{A}(\boldsymbol{\delta}) = -\frac{1}{T}\sum_{t=1}^{T}\widehat{\Delta}_t(\boldsymbol{\delta})\, a_\tau(\widehat{\varepsilon}_t) + \frac{1}{T}\sum_{t=1}^{T}\widehat{r}_t\Big(\widehat{\Delta}_t(\boldsymbol{\delta})\Big), \tag{44}$$

where

$$\widehat{r}_t(v) := \int_0^v \Big(\mathbb{1}\{\widehat{\varepsilon}_t \le s\} - \mathbb{1}\{\widehat{\varepsilon}_t \le 0\}\Big)\, ds, \quad \widehat{\varepsilon}_t := Y_t - \widehat{\boldsymbol{\nu}}_t'\widetilde{\boldsymbol{\phi}}, \quad \widehat{\Delta}_t(\boldsymbol{\delta}) := \widehat{\boldsymbol{\nu}}_t'\boldsymbol{\delta}, \quad a_\tau(z) := \tau - \mathbb{1}\{z \le 0\}.$$

Similarly, applying Knight's identity to (43) yields

$$\mathsf{B}(\boldsymbol{\delta}) = -\frac{1}{T}\sum_{t=1}^{T}\widetilde{\Delta}_t(\boldsymbol{\delta})\, a_\tau(\varepsilon_t) + \frac{1}{T}\sum_{t=1}^{T} r_t^{(0)}\Big(\widetilde{\Delta}_t(\boldsymbol{\delta})\Big), \tag{45}$$

where

$$r_t^{(0)}(v) := \int_0^v \Big(\mathbb{1}\{\varepsilon_t \le s\} - \mathbb{1}\{\varepsilon_t \le 0\}\Big)\, ds, \quad \varepsilon_t := Y_t - \boldsymbol{\nu}_t' \boldsymbol{\phi}_0, \quad \widetilde{\Delta}_t(\boldsymbol{\delta}) := \widetilde{\boldsymbol{\nu}}_t' \boldsymbol{\delta} = \boldsymbol{\nu}_t' \boldsymbol{R}' \boldsymbol{\delta}.$$

By adding and subtracting suitable terms, the linear term of $\mathsf{A}(\boldsymbol{\delta})$ in (44) can be decomposed into a (rotated) oracle score evaluated at the local deviation, together with two perturbation terms induced by regressor estimation errors:

$$\begin{aligned} -\frac{1}{T}\sum_{t=1}^{T} \widehat{\Delta}_t(\boldsymbol{\delta})\, a_\tau(\widehat{\varepsilon}_t) = & \underbrace{-\frac{1}{T}\sum_{t=1}^{T} \widetilde{\Delta}_t(\boldsymbol{\delta})\, a_\tau(\varepsilon_t)}_{\text{(rotated) oracle score}} \\ & \underbrace{-\frac{1}{T}\sum_{t=1}^{T} (\widehat{\Delta}_t(\boldsymbol{\delta}) - \widetilde{\Delta}_t(\boldsymbol{\delta}))\, a_\tau(\widehat{\varepsilon}_t)}_{\text{step-size perturbation}} \underbrace{- \frac{1}{T}\sum_{t=1}^{T} \widetilde{\Delta}_t(\boldsymbol{\delta}) \left(a_\tau(\widehat{\varepsilon}_t) - a_\tau(\varepsilon_t)\right)}_{\text{density perturbation}}. \end{aligned} \tag{46}$$

Similarly, the remainder term of $\mathsf{A}(\boldsymbol{\delta})$ in (44) admits the decomposition

$$\begin{aligned} \frac{1}{T}\sum_{t=1}^{T} \widehat{r}_t(\widehat{\Delta}_t(\boldsymbol{\delta})) = & \underbrace{\frac{1}{T}\sum_{t=1}^{T} r_t^{(0)}(\widetilde{\Delta}_t(\boldsymbol{\delta}))}_{\text{(rotated) oracle score}} \\ & + \underbrace{\frac{1}{T}\sum_{t=1}^{T} \left(\widehat{r}_t(\widehat{\Delta}_t(\boldsymbol{\delta})) - \widehat{r}_t(\widetilde{\Delta}_t(\boldsymbol{\delta}))\right)}_{\text{step-size perturbation}} + \underbrace{\frac{1}{T}\sum_{t=1}^{T} \left(\widehat{r}_t(\widetilde{\Delta}_t(\boldsymbol{\delta})) - r_t^{(0)}(\widetilde{\Delta}_t(\boldsymbol{\delta}))\right)}_{\text{density perturbation}}. \end{aligned} \tag{47}$$

Noting that the rotated oracle terms in (46) and (47) coincide with the corresponding components of $\mathsf{B}(\boldsymbol{\delta})$, these terms cancel when taking the difference $\mathsf{A}(\boldsymbol{\delta}) - \mathsf{B}(\boldsymbol{\delta})$. Consequently, we obtain

$$\begin{aligned} \mathsf{A}(\boldsymbol{\delta}) - \mathsf{B}(\boldsymbol{\delta}) &= -\frac{1}{T}\sum_{t=1}^{T} (\widehat{\Delta}_t(\boldsymbol{\delta}) - \widetilde{\Delta}_t(\boldsymbol{\delta}))\, a_\tau(\widehat{\varepsilon}_t) + \frac{1}{T}\sum_{t=1}^{T} \left(\widehat{r}_t(\widehat{\Delta}_t(\boldsymbol{\delta})) - \widehat{r}_t(\widetilde{\Delta}_t(\boldsymbol{\delta}))\right) \\ &\quad - \frac{1}{T}\sum_{t=1}^{T} \widetilde{\Delta}_t(\boldsymbol{\delta}) \left(a_\tau(\widehat{\varepsilon}_t) - a_\tau(\varepsilon_t)\right) + \frac{1}{T}\sum_{t=1}^{T} \left(\widehat{r}_t(\widetilde{\Delta}_t(\boldsymbol{\delta})) - r_t^{(0)}(\widetilde{\Delta}_t(\boldsymbol{\delta}))\right) \\ &=: \mathsf{S}_1(\boldsymbol{\delta}) + \mathsf{S}_2(\boldsymbol{\delta}) + \mathsf{D}_1(\boldsymbol{\delta}) + \mathsf{D}_2(\boldsymbol{\delta}), \end{aligned} \tag{48}$$

Here, $\mathsf{S}_1(\boldsymbol{\delta})$ and $\mathsf{S}_2(\boldsymbol{\delta})$ correspond to step-size perturbations arising from regressor estimation errors, whereas $\mathsf{D}_1(\boldsymbol{\delta})$ and $\mathsf{D}_2(\boldsymbol{\delta})$ capture density perturbations induced by the nonsmoothness of the check loss and the use of generated regressors.

We proceed by bounding the step-size perturbation terms $\mathsf{S}_1(\boldsymbol{\delta})$ and $\mathsf{S}_2(\boldsymbol{\delta})$, followed by the density perturbation terms $\mathsf{D}_1(\boldsymbol{\delta})$ and $\mathsf{D}_2(\boldsymbol{\delta})$.

Step 2: Bounding the step-size perturbation terms. Since $|a_\tau(z)| \le 1$, we have

$$\begin{aligned}
|\mathsf{S}_1(\boldsymbol{\delta})| &\le \frac{1}{T}\sum_{t=1}^{T}\left|\widehat{\Delta}_t(\boldsymbol{\delta}) - \widetilde{\Delta}_t(\boldsymbol{\delta})\right| = \frac{1}{T}\sum_{t=1}^{T}\left|(\widehat{\boldsymbol{u}}_t - \boldsymbol{u}_t)'\boldsymbol{\delta}_\theta + (\widehat{\boldsymbol{f}}_t - \boldsymbol{H}\boldsymbol{f}_t)'\boldsymbol{\delta}_\gamma\right| \\
&\le \|\boldsymbol{\delta}_\theta\|_1 \frac{1}{T}\sum_{t=1}^{T}\|\boldsymbol{u}_t - \widehat{\boldsymbol{u}}_t\|_\infty + \|\boldsymbol{\delta}_\gamma\|_2 \frac{1}{T}\sum_{t=1}^{T}\|\boldsymbol{H}\boldsymbol{f}_t - \widehat{\boldsymbol{f}}_t\|_2 \\
&\le \|\boldsymbol{\delta}_\theta\|_1\, C_u(a_{p,T} + b_T) + \|\boldsymbol{\delta}_\gamma\|_2\, C_f a_{p,T}
\end{aligned}$$

with probability at least $1 - O(T^{-\eta})$, where the last inequality follows from Proposition 1.

For all $\boldsymbol{\delta} \in \mathcal{A}$ such that $J^{1/2}(\boldsymbol{\delta}) \le \omega$, Lemma 3 together with Lemma 5(i) implies

$$|\mathsf{S}_1(\boldsymbol{\delta})| \le (C_u + C_f)(1 + C_0)\frac{\sqrt{1 + s_\tau + r}}{\kappa}\,(a_{p,T} + b_T)\,\omega. \tag{49}$$

The same high-probability bound applies to $|\mathsf{S}_2(\boldsymbol{\delta})|$. Indeed, noting that the map $v \mapsto r_t^{(0)}(v)$ is 1-Lipschitz, we have for each t,

$$\left|\widehat{r}_t(\widehat{\Delta}_t(\boldsymbol{\delta})) - \widehat{r}_t(\widetilde{\Delta}_t(\boldsymbol{\delta}))\right| \le |\widehat{\Delta}_t(\boldsymbol{\delta}) - \widetilde{\Delta}_t(\boldsymbol{\delta})|,$$

and the result follows by the same argument as above.

Step 3: Bounding the density perturbation terms. Let $h_t := \varepsilon_t - \widehat{\varepsilon}_t$ so that $\widehat{\varepsilon}_t = \varepsilon_t - h_t$. We begin by noting that

$$\mathsf{D}_1(\boldsymbol{\delta}) = \frac{1}{T}\sum_{t=1}^{T}\widetilde{\Delta}_t(\boldsymbol{\delta})\left(a_\tau(\varepsilon_t) - a_\tau(\widehat{\varepsilon}_t)\right) = \frac{1}{T}\sum_{t=1}^{T}\widetilde{\Delta}_t(\boldsymbol{\delta})\Big(\mathbb{1}\{\widehat{\varepsilon}_t \le 0\} - \mathbb{1}\{\varepsilon_t \le 0\}\Big)\,, \tag{50}$$

$$\mathsf{D}_2(\boldsymbol{\delta}) = \frac{1}{T}\sum_{t=1}^{T}\left[\int_0^{\widetilde{\Delta}_t(\boldsymbol{\delta})}\Big(\mathbb{1}\{\widehat{\varepsilon}_t \le s\} - \mathbb{1}\{\varepsilon_t \le s\}\Big)\,ds - \widetilde{\Delta}_t(\boldsymbol{\delta})\Big(\mathbb{1}\{\widehat{\varepsilon}_t \le 0\} - \mathbb{1}\{\varepsilon_t \le 0\}\Big)\right]. \tag{51}$$

Therefore, the jump terms at $s=0$ cancel exactly, and we obtain

$$\mathsf{D}_1(\boldsymbol{\delta}) + \mathsf{D}_2(\boldsymbol{\delta}) = \frac{1}{T}\sum_{t=1}^{T}\int_0^{\widetilde{\Delta}_t(\boldsymbol{\delta})} \Big(\mathbb{1}\{\widehat{\varepsilon}_t \le s\} - \mathbb{1}\{\varepsilon_t \le s\}\Big)\, ds \;=:\; \frac{1}{T}\sum_{t=1}^{T} \mathsf{G}_{1t}(\boldsymbol{\delta}). \tag{52}$$

It now remains to deal with $\frac{1}{T}\sum_{t=1}^{T} \mathsf{G}_{1t}(\boldsymbol{\delta})$. Following the same argument as in step 3 of Lemma 6, and using the mean value theorem, we obtain

$$\mathbb{E}[\mathsf{G}_{1t}(\boldsymbol{\delta}) \mid \mathcal{I}_t] = -h_t\, \widetilde{\Delta}_t(\boldsymbol{\delta})\, \widetilde{f}, \qquad 0 \le \widetilde{f} \le \bar{f}. \tag{53}$$

Consequently, for all $\boldsymbol{\delta} \in \mathcal{A}$ satisfying $J^{1/2}(\boldsymbol{\delta}) \le \omega$, we obtain

$$\begin{aligned}
\big|\mathsf{D}_1(\boldsymbol{\delta}) + \mathsf{D}_2(\boldsymbol{\delta})\big| &= \left|\frac{1}{T}\sum_{t=1}^{T} \mathsf{G}_{1t}(\boldsymbol{\delta})\right| \\
&\le \left|\frac{1}{T}\sum_{t=1}^{T} \mathbb{E}[\mathsf{G}_{1t}(\boldsymbol{\delta}) \mid \mathcal{I}_t]\right| + \left|\frac{1}{T}\sum_{t=1}^{T} \Big(\mathsf{G}_{1t}(\boldsymbol{\delta}) - \mathbb{E}[\mathsf{G}_{1t}(\boldsymbol{\delta}) \mid \mathcal{I}_t]\Big)\right| \\
&\le \bar{f}\, \|\boldsymbol{R}\|_\infty\, \|\boldsymbol{\delta}\|_1 \left\|\frac{1}{T}\sum_{t=1}^{T} h_t\, \boldsymbol{\nu}_t\right\|_\infty + C_{\text{rem}}\, \frac{\sqrt{1+s_\tau+r}}{\kappa}\, b_T\, \omega && (54)\\
&\le \bar{f}\, (\sqrt{2rc_1/c_2})(1+C_0)\, \frac{\sqrt{1+s_\tau+r}}{\kappa}\, \omega\, C_\eta\, b_T + C_{\text{rem}}\, \frac{\sqrt{1+s_\tau+r}}{\kappa}\, b_T\, \omega && (55)\\
&\le \Big(\bar{f}\, (\sqrt{2rc_1/c_2})(1+C_0)C_\eta + C_{\text{rem}}\Big)\, \frac{\sqrt{1+s_\tau+r}}{\kappa}\, b_T\, \omega. && (56)
\end{aligned}$$

The above bound holds with probability at least $1 - O(T^{-\eta})$.

In (54), the bound on the stochastic fluctuation term follows from the same argument as in step 3 of Lemma 6, which yields the rate stated in (40); the constant C_{rem} is defined therein. In (55), we use the high-probability bound on $\|\boldsymbol{R}\|_\infty$ established in (24) and Lemma B.2.

The claim of the lemma then follows by combining (49) and (56). □

B Useful Results

Lemma B.1 (Bound for the Average Factor Norm)**.** *Suppose that Assumptions A.1–A.4 hold. Then, for any $\eta > 0$ and sufficiently large T, it holds that*

$$\mathbb{P}\left(\frac{1}{T}\sum_{t=1}^{T}\|\boldsymbol{f}_t\|_2 \leq \sqrt{\frac{\pi}{C_m}}\right) \geq 1 - T^{-\eta}.$$

Proof. By Assumption A.2, each $\boldsymbol{f}_t$ satisfies $\mathbb{P}(\|\boldsymbol{f}_t\|_2 > \varepsilon) \leq \exp(-C_m\varepsilon^2)$. Integrating this tail bound over $\varepsilon > 0$ gives

$$\mathbb{E}\|\boldsymbol{f}_t\|_2 = \int_0^\infty \mathbb{P}(\|\boldsymbol{f}_t\|_2 > \varepsilon)\, d\varepsilon \leq \int_0^\infty e^{-C_m\varepsilon^2}\, d\varepsilon = \frac{\sqrt{\pi}}{2\sqrt{C_m}}.$$

Let $Z_t = \|\boldsymbol{f}_t\|_2 - \mathbb{E}\|\boldsymbol{f}_t\|_2$. Then $\{Z_t\}_{t=1}^T$ is a zero-mean, sub-Gaussian and α-mixing sequence. By Lemma B.2 of Brownlees *et al.* (2024), there exists a constant $C_\eta > 0$ such that, for any $\eta > 0$,

$$\mathbb{P}\left(\left|\frac{1}{T}\sum_{t=1}^{T} Z_t\right| > C_\eta\sqrt{\frac{\log T}{T}}\right) \leq T^{-\eta}.$$

Combining the two parts yields

$$\frac{1}{T}\sum_{t=1}^{T}\|\boldsymbol{f}_t\|_2 \leq \left|\frac{1}{T}\sum_{t=1}^{T} Z_t\right| + \mathbb{E}\|\boldsymbol{f}_t\|_2 \leq C_\eta\sqrt{\frac{\log T}{T}} + \frac{\sqrt{\pi}}{2\sqrt{C_m}}$$

with probability at least $1 - O(T^{-\eta})$. For sufficiently large T, the second term dominates the first, which yields the stated bound. $\square$

Lemma B.2. *Let $\{\boldsymbol{Z}_t\}_{t=1}^T$ be a stationary sequence of d-dimensional zero-mean random vectors. Suppose (i) $\sup_{1\leq i\leq d}\mathbb{P}(|Z_{it}| > \varepsilon) \leq \exp(-C_m\varepsilon)$ for some $C_m > 0$; (ii) the α-mixing coefficients of the sequence satisfy $\alpha(l) < \exp(-C_\alpha l^{r_\alpha})$ for some $C_\alpha > 0$ and $r_\alpha > 0$; and (iii) $d = \lfloor C_d T^{r_d}\rfloor$ for some $C_d > 0$ and $r_d \in (0, r_\alpha)$.*

Then for any $\eta > 0$ there exists a positive constant C such that, for all T sufficiently large, it holds that

$$\mathbb{P}\left(\left\|\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{Z}_t\right\|_\infty \geq C_\eta\sqrt{\frac{\log(T)}{T}}\right) \leq \frac{1}{T^\eta}\ .$$

Proof. Let C^* denote a positive constant to be chosen below. Observe that

$$\mathbb{P}\left(\left\|\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{Z}_t\right\|_\infty \geq C^*\sqrt{\frac{\log T}{T}}\right) \leq d \max_{1\leq i\leq d}\mathbb{P}\left(\left|\sum_{t=1}^{T}Z_{it}\right| \geq C^*\sqrt{T\log T}\right).$$

The remainder of the proof follows exactly the same steps as in Lemma B.2 of Brownlees *et al.* (2024), where the concentration bound is established under identical α-mixing and tail conditions. The only change is that here the ℓ_2 norm in that lemma is replaced by the ℓ_∞ norm, which affects only the union bound at the first line and leads to a logarithmic factor $\sqrt{\log T}$ instead of $\sqrt{d\log T}$ under the dimensionality condition $r_d < r_\alpha$. Hence the desired bound follows by identical arguments. ☐

Remark While the literature commonly expresses high-dimensional ℓ_∞ type bounds in terms of $\sqrt{(\log d)/T}$, where d denotes the ambient dimension, the two rates $\sqrt{(\log d)/T}$ and $\sqrt{(\log T)/T}$ are equivalent under our growth condition $d = O(T^{r_d})$ with $r_d < r_\alpha$. For simplicity of notation, we use the latter form throughout this paper.

Proposition B.1. *Suppose Assumptions A.1–A.4 hold. Let*

$$a_{p,T} = \frac{(p+\log T)^{(r_\alpha+1)/r_\alpha}}{p^\alpha T} + \frac{1}{p^\alpha}, \qquad b_T = \sqrt{\frac{\log T}{T}}.$$

Then, for any $\eta > 0$ and sufficiently large T, each of the following holds with probability at least $1 - O(T^{-\eta})$: (i) $\frac{1}{T}\sum_{t=1}^{T}\|\boldsymbol{f}_t - \boldsymbol{H}'\widehat{\boldsymbol{f}}_t\|_2 \leq C_{H1}(a_{p,T} + b_T)$, (ii) $\|\boldsymbol{H}'\boldsymbol{H} - \boldsymbol{I}_r\|_2 \leq C_{H2}(a_{p,T} + b_T)$, (iii) $\|\widehat{\mathbf{B}} - \mathbf{B}\boldsymbol{H}^{-1}\|_\infty \leq C_{H3}(a_{p,T} + b_T)$, (iv) $\|\boldsymbol{H}\|_2 \leq \sqrt{2c_1/c_r}$, and (v) $\|\boldsymbol{H}^{-1}\|_2 \leq 2\sqrt{c_1/c_r}$.

Proof. (i) By a simple rearrangement,

$$\begin{aligned}
\|\boldsymbol{f}_t - \boldsymbol{H}'\widehat{\boldsymbol{f}}_t\|_2 &= \|\boldsymbol{H}^{-1}(\boldsymbol{H}\boldsymbol{f}_t - \boldsymbol{H}\boldsymbol{H}'\widehat{\boldsymbol{f}}_t)\|_2 \\
&= \|\boldsymbol{H}^{-1}\big[(\boldsymbol{H}\boldsymbol{H}' - \boldsymbol{I}_r)\widehat{\boldsymbol{f}}_t + (\widehat{\boldsymbol{f}}_t - \boldsymbol{H}\boldsymbol{f}_t)\big]\|_2 \\
&\leq \|\boldsymbol{H}^{-1}\|_2\Big(\|\boldsymbol{H}\boldsymbol{H}' - \boldsymbol{I}_r\|_2\,\|\widehat{\boldsymbol{f}}_t\|_2 + \|\widehat{\boldsymbol{f}}_t - \boldsymbol{H}\boldsymbol{f}_t\|_2\Big),
\end{aligned}$$

where the inequality follows from the triangle inequality and the submultiplicativity of the operator norm.

Hence, applying Proposition 1(i) together with Propositions A.6 and Proposition B.1 (v), we obtain that, for any $\eta > 0$, there exist constants $C', C'' > 0$ such that

$$\begin{aligned}
&\mathbb{P}\left(\frac{1}{T}\sum_{t=1}^{T}\|\boldsymbol{f}_t - \boldsymbol{H}'\widehat{\boldsymbol{f}}_t\|_2 > 2\sqrt{\frac{c_1}{c_r}}(C'\sqrt{r}+C'')\left[\sqrt{\frac{\log T}{T}} + \frac{(p+\log T)^{(r_\alpha+1)/r_\alpha}}{p^\alpha T} + \frac{1}{p^\alpha}\right]\right)\\
&\leq \mathbb{P}\left(\|\boldsymbol{H}^{-1}\|_2 > 2\sqrt{\frac{c_1}{c_r}}\right) + \mathbb{P}\left(\|\boldsymbol{H}\boldsymbol{H}' - \boldsymbol{I}_r\|_2 > C'\left[\sqrt{\frac{\log T}{T}} + \frac{(p+\log T)^{(r_\alpha+1)/r_\alpha}}{p^\alpha T} + \frac{1}{p^\alpha}\right]\right)\\
&\quad + \mathbb{P}\left(\frac{1}{T}\sum_{t=1}^{T}\|\widehat{\boldsymbol{f}}_t\|_2 > \sqrt{r}\right) + \mathbb{P}\left(\frac{1}{T}\sum_{t=1}^{T}\|\widehat{\boldsymbol{f}}_t - \boldsymbol{H}\boldsymbol{f}_t\|_2 > C''\left[\frac{(p+\log T)^{(r_\alpha+1)/r_\alpha}}{p^\alpha T} + \frac{1}{p^\alpha}\right]\right) = O(T^{-\eta}).
\end{aligned}$$

Setting $C_{H1} = 2\sqrt{\frac{c_1}{c_r}}(C'\sqrt{r}+C'')$ completes the proof.

(ii) For the second statement, Propositions A.6 and A.7(ii)-(iii) in Brownlees *et al.* (2024) imply that, for any $\eta > 0$,

$$\begin{aligned}
&\mathbb{P}\left(\|\boldsymbol{H}'\boldsymbol{H} - \boldsymbol{I}_r\|_2 > 2\sqrt{2}\,C'\frac{c_1}{c_r}\left[\sqrt{\frac{\log T}{T}} + \frac{(p+\log T)^{(r_\alpha+1)/r_\alpha}}{p^\alpha T} + \frac{1}{p^\alpha}\right]\right)\\
&= \mathbb{P}\left(\|\boldsymbol{H}^{-1}(\boldsymbol{H}\boldsymbol{H}' - \boldsymbol{I}_r)\boldsymbol{H}\|_2 > 2\sqrt{2}\,C'\frac{c_1}{c_r}\left[\sqrt{\frac{\log T}{T}} + \frac{(p+\log T)^{(r_\alpha+1)/r_\alpha}}{p^\alpha T} + \frac{1}{p^\alpha}\right]\right)\\
&\leq \mathbb{P}\left(\|\boldsymbol{H}^{-1}\|_2 > 2\sqrt{\tfrac{c_1}{c_r}}\right) + \mathbb{P}\left(\|\boldsymbol{H}\|_2 > \sqrt{\tfrac{2c_1}{c_r}}\right)\\
&\quad + \mathbb{P}\left(\|\boldsymbol{H}\boldsymbol{H}' - \boldsymbol{I}_r\|_2 > C'\left[\sqrt{\frac{\log T}{T}} + \frac{(p+\log T)^{(r_\alpha+1)/r_\alpha}}{p^\alpha T} + \frac{1}{p^\alpha}\right]\right) = O(T^{-\eta}).
\end{aligned}$$

Setting $C_{H2} = 2\sqrt{2}\,C'\frac{c_1}{c_r}$ yields the desired bound.

(iii) From $\widehat{\mathbf{B}} = T^{-1}\sum_{t=1}^{T}\boldsymbol{X}_t\widehat{\boldsymbol{f}}_t'$ and $\boldsymbol{X}_t = \mathbf{B}\boldsymbol{f}_t + \boldsymbol{u}_t$, we have

$$\widehat{\mathbf{B}} - \mathbf{B}\boldsymbol{H}^{-1} = \mathbf{B}\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{H}^{-1}(\boldsymbol{H}\boldsymbol{f}_t - \widehat{\boldsymbol{f}}_t)\widehat{\boldsymbol{f}}_t' + \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{u}_t(\widehat{\boldsymbol{f}}_t' - \boldsymbol{f}_t'\boldsymbol{H}') + \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{u}_t\boldsymbol{f}_t'\boldsymbol{H}'. \tag{57}$$

We bound the three terms on the right-hand side in turn. For the first term, by the

Cauchy-Schwarz inequality, for any $\eta > 0$,

$$\begin{aligned}
&\left\| \mathbf{B} \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{H}^{-1} (\boldsymbol{H} \boldsymbol{f}_t - \widehat{\boldsymbol{f}}_t) \widehat{\boldsymbol{f}}_t' \right\|_\infty \le \|\mathbf{B}\|_\infty \sqrt{r} \left\| \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{H}^{-1} (\boldsymbol{H} \boldsymbol{f}_t - \widehat{\boldsymbol{f}}_t) \widehat{\boldsymbol{f}}_t' \right\|_2 \\
\le \;& \|\mathbf{B}\|_\infty \sqrt{r} \, \|\boldsymbol{H}^{-1}\|_2 \left(\frac{1}{T} \sum_{t=1}^{T} \|\boldsymbol{H} \boldsymbol{f}_t - \widehat{\boldsymbol{f}}_t\|_2^2 \cdot \frac{1}{T} \sum_{t=1}^{T} \|\widehat{\boldsymbol{f}}_t\|_2^2 \right)^{1/2} \\
\le \;& C_B \, r \, \|\boldsymbol{H}^{-1}\|_2 \left(\frac{r}{T} \|\widehat{\boldsymbol{F}} - \boldsymbol{F} \boldsymbol{H}'\|_2^2 \right)^{1/2} \le 2 r^{3/2} C_B C_1 \sqrt{c_1/c_r} \cdot a_{p,T}, \qquad (58)
\end{aligned}$$

which holds with probability at least $1 - O(T^{-\eta})$, where the last inequality follows from Propositions A.5 and A.7(iii) of Brownlees *et al.* (2024).

For the second term, note that each $\boldsymbol{u}_t$ is sub-Gaussian. By the tail-integration bound,

$$\mathbb{E}(u_{jt}^2) = \int_0^\infty \mathbb{P}(u_{jt}^2 > \varepsilon) \, d\varepsilon = 2 \int_0^\infty x \, \mathbb{P}(|u_{jt}| > x) \, dx \le 2 \int_0^\infty x e^{-C_m x^2} \, dx = \frac{1}{C_m} \ ,$$

so that $\mathbb{E}(u_{jt}^2) \le 1/C_m$ for all j. Under the sub-Gaussian assumption on $\{u_{jt}\}$ and the growth condition $\log p \lesssim T$, there exists a finite constant $C_m' > 0$ such that

$$\max_{1 \le j \le p} \frac{1}{T} \sum_{t=1}^{T} u_{jt}^2 \le C_m' \quad \text{with probability at least } 1 - O(T^{-\eta}).$$

Hence,

$$\begin{aligned}
\left\| \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{u}_t (\widehat{\boldsymbol{f}}_t' - \boldsymbol{f}_t' \boldsymbol{H}') \right\|_\infty \;&\le\; \sqrt{r} \max_{1 \le j \le p} \left\| \frac{1}{T} \sum_{t=1}^{T} u_{jt} (\widehat{\boldsymbol{f}}_t' - \boldsymbol{f}_t' \boldsymbol{H}') \right\|_2 \\
&\le\; \sqrt{r} \max_{1 \le j \le p} \left(\frac{1}{T} \sum_{t=1}^{T} u_{jt}^2 \right)^{1/2} \left(\frac{1}{T} \sum_{t=1}^{T} \|\widehat{\boldsymbol{f}}_t' - \boldsymbol{f}_t' \boldsymbol{H}'\|_2^2 \right)^{1/2} \\
&\le\; \sqrt{r} \sqrt{C_m'} \left(\frac{r}{T} \|\widehat{\boldsymbol{F}} - \boldsymbol{F} \boldsymbol{H}'\|_2^2 \right)^{1/2} \le r \, C_1 \sqrt{C_m'} \cdot a_{p,T} \ , \qquad (59)
\end{aligned}$$

holds with the same probability.

For the third term, let $\boldsymbol{V}_{j,t}' \in \mathbb{R}^{1 \times r}$ denote the j-th row of $\boldsymbol{u}_t \boldsymbol{f}_t'$. Then

$$\left\| \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{u}_t \boldsymbol{f}_t' \boldsymbol{H}' \right\|_\infty = \max_{1 \le j \le p} \left\| \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{V}_{j,t}' \boldsymbol{H}' \right\|_1 \le \sqrt{r} \, \|\boldsymbol{H}\|_2 \max_{1 \le j \le p} \left\| \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{V}_{j,t} \right\|_2 .$$

By an intermediate result in Proposition A.4 of Brownlees *et al.* (2024), there exists a constant $C > 0$ such that, with probability at least $1 - O(T^{-\eta})$ for any $\eta > 0$,

$$\max_{1 \le j \le p} \left\| \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{V}_{j,t} \right\|_2 \le C\sqrt{(r \log T)/T} = C\sqrt{r}\, b_T.$$

Consequently,

$$\left\| \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{u}_t \boldsymbol{f}_t' \boldsymbol{H}' \right\|_\infty \le r\, C\, \sqrt{2c_1/c_r}\, b_T, \tag{60}$$

which holds with the same probability.

Combining (57) with (58)-(60) and choosing

$$C_{H3} \ge 2r^{3/2} C_B C_1 \sqrt{c_1/c_r} + rC_1\sqrt{C_m'} + rC\sqrt{2c_1/c_r},$$

we obtain the desired bound for $\|\widehat{\mathbf{B}} - \mathbf{B}\boldsymbol{H}^{-1}\|_\infty$, which completes the proof.

Parts (iv)–(v) are direct consequences of Proposition A.7 in Brownlees *et al.* (2024) .

□