Belloni, Alexandre; Chen, Mingli; Chernozhukov, Victor

**Working Paper**

# Quantile graphical models: Prediction and conditional independence with applications to systemic risk

cemmap working paper, No. CWP54/17

**Provided in Cooperation with:**
Institute for Fiscal Studies (IFS), London

*Suggested Citation:* Belloni, Alexandre; Chen, Mingli; Chernozhukov, Victor (2017) : Quantile graphical models: Prediction and conditional independence with applications to systemic risk, cemmap working paper, No. CWP54/17, Centre for Microdata Methods and Practice (cemmap), London,
https://doi.org/10.1920/wp.cem.2017.5417

This Version is available at:
https://hdl.handle.net/10419/189792

cemmap
centre for microdata methods and practice

# Quantile graphical models: prediction and conditional independence with applications to systemic risk

Alexandre Belloni
Mingli Chen
Victor Chernozhukov

The Institute for Fiscal Studies
Department of Economics, UCL

**cemmap** working paper CWP54/17

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL
An ESRC Research Centre

# QUANTILE GRAPHICAL MODELS: PREDICTION AND CONDITIONAL INDEPENDENCE WITH APPLICATIONS TO SYSTEMIC RISK

ALEXANDRE BELLONI*, MINGLI CHEN‡, AND VICTOR CHERNOZHUKOV†

Abstract. The understanding of co-movements, dependence, and influence between variables of interest is key in many applications. Broadly speaking such understanding can lead to better predictions and decision making in many settings. We propose Quantile Graphical Models (QGMs) to characterize prediction and conditional independence relationships within a set of random variables of interest. Although those models are of interest in a variety of applications, we draw our motivation and contribute to the financial risk management literature. Importantly, the proposed framework is intended to be applied to non-Gaussian settings, which are ubiquitous in many real applications, and to handle a large number of variables and conditioning events.

We propose two distinct QGMs. First, Condition Independence Quantile Graphical Models (CIQGMs) characterize conditional independence at each quantile index revealing the distributional dependence structure. Second, Prediction Quantile Graphical Models (PQGMs) characterize the best linear predictor under asymmetric loss functions. A key difference between those models is the (non-vanishing) misspecification between the best linear predictor and the conditional quantile functions.

We also propose estimators for those QGMs. Due to high-dimensionality, the two distinct QGMs require different estimators. The estimators are based on high-dimensional techniques including (a continuum of) $\ell_1$-penalized quantile regressions (and low biased equations), which allow us to handle the potential large number of variables. We build upon a recent literature to obtain new results for valid choice of the penalty parameters, rates of convergence, and confidence regions that are simultaneously valid.

We illustrate how to use QGMs to quantify tail interdependence (instead of mean dependence) between a large set of variables which is relevant in applications concerning with extreme events. We show that the associated tail risk network can be used for measuring systemic risk contributions. We also apply the framework to study international financial contagion and the impact of market downside movement on the dependence structure of assets' returns.

Key words: High-dimensional approximately sparse model, tail risk network, conditional independence, nonlinear correlation, penalized quantile regression, systemic risk, financial contagion, downside movement

---

*Date*: First version: November 2012, this version December 4, 2017.

We would like to thank Don Andrews, Debopam Bhattacharya, Peter Bickel, Marianne Bitler, Peter Bühlmann, Colin Cameron, Xu Cheng, Valentina Corradi, Alan Crawford, Francis Diebold, Iván Fernández-Val, Bulat Gafarov, Kenji Fukumizu, Dalia Ghanem, Bryan Graham, Jiaying Gu, Ran Gu, Jungbin Hwang, Cheng Hsiao, Hidehiko Ichimura, Michael Jansson, Oscar Jorda, Chihwa Kao, Kengo Kato, Shakeeb Khan, Roger Koenker, Chenlei Leng, Arthur Lewbel, Michael Leung, Song Liu, Whitney Newey, Dong Hwan Oh, David Pacini, Andrew Patton, Aureo de Paula, Victor de la Pena, Elisabeth Perlman, Stephen Portnoy, Demian Pouzo, James Powell, Geert Ridder, Stephen Ross, Shu Shen, Senay Sokullu, Sami Stouli, Aleksey Tetenov, Takuya Ura, Tiemen Woutersen, Zhijie Xiao, Wenkai Xu, and Yanos Zylberberg for comments and discussions. We would also like to thank the seminar and workshop participants from Boston College, 2015 Warwick Summer Workshop, 11th World Congress of the Econometric Society, 2017 UCL Workshop on the Theory of Big Data, 2017 California Econometrics Conference, University of Bristol, University of Connecticut, Humboldt University Berlin, the Institute of Statistical Mathematics, USC, UC Berkeley, UC Davis, Warwick Statistics Department and University of Tokyo.

*Duke University, e-mail:abn5@duke.edu.

‡University of Warwick, e-mail:m.chen.3@warwick.ac.uk.

†Massachusetts Institute of Technology, e-mail:vchern@mit.edu.

## 1. Introduction

The understanding of co-movements, dependence and influence between variables of interest is key in many applications. Such understanding can lead to better predictions and decision making in many settings. This is clearly of interest in financial management settings where quantifying risk is crucial. For example, the dependence between stock returns plays a key role in hedging strategies. Those hedging decisions are typically focused on the tail of the distribution of returns rather than the mean. Moreover, such strategies aiming to reduce risk are critical precisely during market downside movement. Therefore, it is also instructive to understand how the dependence (and policies) would change as the downside movement of the market becomes more extreme. Moreover, recent empirical evidence [7, 6, 66] points to non-Gaussianity of the distribution of stock returns, especially during market downturns. These issues require models that allow for non-Gaussian settings and can also accommodate various conditioning events (e.g., downside movements).

Motivated by those features, this work proposes Quantile Graphical Models (QGMs) to characterize (and visualize) the dependence structure of a set of random variables. The proposed framework allows us to understand prediction and conditional independence between those variables. Moreover, it also enables us to focus on specific parts of the distributions of those variables such as tail events. Such understanding plays an important role in applications like financial contagion and measuring systemic risk contributions where extreme events are the main interests for regulators. Our techniques are intended to be applicable to high-dimensional settings where the total number of variables (or additional conditioning variables) is large – possibly larger than the sample size.

QGMs provide an alternative route to learn conditional independence and prediction under asymmetric loss functions which is appealing in non-Gaussian settings. In the Gaussian setting, those notions essentially coincide; however, in non-Gaussian settings different estimation approaches are needed. Conditional independence hinges on the equivalence between conditional probabilities and conditional quantiles. Prediction under asymmetric loss function hinges on the solution of a quantile regression with non-vanishing misspecification ($M$-estimation problem). Although we build upon the quantile regression literature ([50, 18]), we derive new results for penalized quantile regression in high dimensional settings to handle misspecification, many controls and a continuum of additional conditioning events.

Conditional independence has a long history in statistical models with consequences towards parameter identification, causal inference, prediction sufficiency, and many others, see [32]. Conditional Independence Quantile Graphical Models (CIQGMs) aim to characterize conditional independence via the conditional quantile functions. In such models we consider a flexible specification that can approximate well the conditional quantile functions (up to a vanishing approximation error). In turn, this allows to detect which variables have a strong or near zero impact on others which can then be used to provide guidance on conditional independence.

Prediction Quantile Graphical Models (PQGMs) focus on prediction of a variable based on the other variables (a reduced form relation). An important motivation for proposing PQGMs is to allow for misspecification as the conditional quantile function is typically non-linear in non-Gaussian settings but a linear specification is widely used in practice. Misspecification in quantile regression models was first properly justified by [8] in which it was shown that we can recover a suitable "best approximation" for the conditional quantile function. We directly characterize the good prediction properties under asymmetric loss functions

which is appealing in empirical applications. Other papers investigated the impact of misspecification on the specification of the quantile function are [8], [49], [55] and [1]. Nonetheless, this work seems to be the first to accommodate non-vanishing misspecification for high-dimensional quantile regression.

Broadly speaking QGMs enhance our understanding of statistical dependence among the components of a $d$-dimensional random vector $X_V$, where the set $V$ contains the labels of the components. QGMs will provide a way to visualize such dependence via graphs where nodes represent components of $X_V$ and edges represent conditional relationships (as in Gaussian Graphical Models, see below). Given that for each specific quantile index $\tau$ we will obtain one such graph, we could have a graph process indexed by $\tau \in (0, 1)$. The structure represented by the $\tau$-quantile graph represents a local relation and can be valuable in applications where the tail interdependence (corresponding to high or low quantile index) might be of special interest. This is akin to the contrast between quantile regression and linear regression, where the latter provides information only on the conditional mean, while the former can provide a more complete description of the distribution of the outcome.

The graph process induced by QGMs has several important features. First, a $\tau$-quantile graph enables different values of edge strength in different directions. This is important because for undirected networks, the distinction between exposure and contribution is unclear. Second, QGMs are able to capture the tail interdependence through estimating at a high or low quantile index. Third, QGMs can capture the asymmetric dependence structure at different quantiles, which can be particularly useful in empirical applications (e.g., stock market dependence, exchange rate dependence). By considering all the quantiles at once we can characterize conditional independence structure for a set of variables which are not jointly Gaussian distributed, i.e. the case where the covariance matrix cannot characterize conditional independence completely.

A key feature of our work is it provides estimation procedures to learn QGMs from the data observed. The estimators are geared to cover high-dimensional settings where the size of the model is large relative to the sample size. Those estimators are based on $\ell_1$-penalized quantile regression and low biased equations. For CIQGMs, under mild regularity conditions, we provide rates of convergence and edge properties of the estimated graph that hold uniformly over a large class of data generating processes. We provide simultaneously valid confidence regions (post-selection) for the coefficients of the CIQGM that are uniformly valid, despite of possible model selection mistakes. Furthermore, based on proper thresholding, recovery of the QGMs pattern is possible when coefficients are well separated from zero which parallel the results for graph recovery in the Gaussian case based on the estimation of the precision matrix. [1] For PQGMs, we provide an estimator that achieves an adaptive rate of convergence which might differ under different conditioning events. Therefore we contribute to the recent literature on simultaneous valid confidence regions post-model selection that has been an active research area in econometrics [10, 17, 16, 41, 23, 28] and statistics [71, 80, 19, 18, 45, 64, 74]; in particular, the penalty choices and theoretical results are uniformly valid and adaptive to the relevant conditioning events.

QGMs can play important roles in applications where tail events are relevant. For example, with certain rescaling of the edge weights, we are able to capture the importance of each node or measure its systemic risk contribution. In parallel with [5], many approaches for measuring systemic risk fit naturally into QGMs. For

[1] Similar to graph recovery in the Gaussian case such exact recovery is subject to the lack of uniformity validity critiques of Leeb and Pötscher [56].

example, one can view $\Delta CoVaR_\tau^{b|a}$, a measure of the impact of institution $a$ on institution $b$, as the weight of an edge of a $\tau$-quantile graph. Then, the systemic risk contribution of institution $a$, equals to the sum of coefficients over $b \in V$, i.e. $\Delta CoVaR_\tau^{sys|a} = \sum_{b \in V} \Delta CoVaR_\tau^{b|a}$. Similarly, the summation over $a \in V$ measures exposures of individual institution to systemic shocks from the network. QGMs can also be used to study contagion and network spillover effects since it is useful for characterizing tail risk spillovers. We apply QGM to the study of international financial contagion in volatilities, specializing in learning the risk transmission channels, see [30] for an overview of international financial contagions. After learning the risk transmission channels, we can use our new network-cooperated $\Delta CoVaR$ to measure the contribution and exposure of each country to the whole market.

Our work is complementary to a large body of works that study a set of jointly Gaussian distributed random variables. Gaussian Graphical Model (GGM) provides a graphical model representation of those variables. It is well known that conditional independence structure is completely characterized by the support of the precision matrix (i.e. the inverse of the covariance matrix) of the random variables of interest, hence recovering the structure of an undirected Gaussian graph is equivalent to recovering the support of the precision matrix, [33, 53, 31, 38]. Methods work with the precision matrix (or covariance matrix) using hypothesis testing can be find in [38, 35, 36, 37]. Recently, due to the advance of regularization techniques, high-dimensional GGMs have been extensively studied in the statistical and machine learning literature: [62] propose neighborhood selection, i.e. using LASSO for each node in the graph and combine the results column-by-column; [79, 9, 43] penalize the log-likelihood function hence work with the precision matrix directly; other refined estimators including [78, 22, 60, 69, 59, 29]. [58] extended the result to a more general class of models called nonparanormal models or semiparametric Gaussian copula models, i.e., the variables have a multivariate Gaussian distribution after a set of unknown monotone transformations (see also [57, 76, 77]). However, those methods assume the (transformed) random vectors follows a joint Gaussian distribution. In addition, they characterize the conditional mean predictability by linear combinations of the other variables.

Our work also contributes to a growing literature that rely on quantile based models to characterize the data generating process. [81] considers a globally adaptive quantile regression model, establishes oracle properties and improved rates of convergence for the high-dimensional case. Screening procedures based on moment conditions motivated by the quantile models have been proposed and analyzed in [44] and [75] in the high-dimensional case. [46] considers tail dependence defined via conditional probabilities in a low dimensional setting. Several other quantile based models have been proposed, see e.g. [50]. Among the contributions of this work is to consider a high-dimensional setting and propose techniques that can be robust to small coefficients (i.e. allowing for model selection mistakes), non-vanishing misspecification in the conditional quantile function, and uniformly valid over additional conditioning events.

The rest of the paper is organized as follows. Section 2 lays out the foundation of QGMs. Section 3 contains estimators for QGMs while Section 4 contains the theoretical guarantees of the estimators. Section 5 provides empirical applications of QGMs to measure systemic risk contribution and to hedging conditional on the downside movements of the US stock market. Finally, the appendix contains proofs, simulations and implementation details of the estimators.

**Notation**. For an integer $k$, we let $[k] := \{1, \dots, k\}$ denote the set of integers from 1 to $k$. For a random variable $X$ we denote by $\mathcal{X}$ its support. We use the notation $a \vee b = \max\{a, b\}$ and $a \wedge b = \min\{a, b\}$. We use $\|v\|_p$ to denote the $p$-norm of a vector $v$. We denote the $\ell_0$-"norm" by $\|\cdot\|_0$ (i.e., the number of nonzero components). Given a vector $\delta \in \mathbb{R}^p$, and a set of indices $T \subset \{1, ..., d\}$, we denote by $\delta_T$ the vector in which $\delta_{Tj} = \delta_j$ if $j \in T$, $\delta_{Tj} = 0$ if $j \notin T$. We use $\mathbb{E}_n$ to abbreviate the notation $n^{-1}\sum_{i=1}^n$; for example, $\mathbb{E}_n[f] := \mathbb{E}_n[f(\omega_i)] := n^{-1}\sum_{i=1}^n f(\omega_i)$.

## 2. Quantile Graphical Models

In this section we describe quantile graphical models associated with a $d$-dimensional random vector $X_V$ where the set $V = [d] = \{1, \dots, d\}$ denotes the labels of the components. These models aim to provide a description of the dependence between the random variables in $X_V$. In particular, these models induce graphs that allow for visualizing dependence structures. Nonetheless, because of the non-Gaussianity, we consider two fundamentally distinct models (one geared towards conditional independence and one geared towards prediction).

### 2.1. Conditional Independence Quantile Graphical Models.

Conditional independence graphs have been used to provide visualization and insight on the dependence structure between random variables. Each node of the graph is associated with a component of $X_V$. We denote the conditional independence graph as $G^I = (V, E^I)$ where $G^I$ is an undirected graph with vertex set $V$ and edge set $E^I$ which is represented by an adjacency matrix ($E^I_{a,b} = 1$ if the edge $(a, b) \in G^I$, and $E^I_{a,b} = 0$ otherwise). An edge $(a, b)$ is not contained in the graph if and only if

$$X_a \perp X_b \;\mid\; X_{V\setminus\{a,b\}}, \tag{2.1}$$

namely $X_b$ and $X_a$ are independent conditional on all remaining variables $X_{V\setminus\{a,b\}} = \{X_k; k \in V\setminus\{a, b\}\}$.

**Comment 2.1** (Conditional Independence Under Gaussianity)**.** In the case that $X_V$ is jointly Gaussian distributed, $X_V \sim N(0, \Sigma)$ with $\Sigma$ as the covariance matrix of $X_V$, the conditional independence structure between two components is determined by the inverse of the covariance matrix, i.e. the precision matrix $\Theta = \Sigma^{-1}$. It follows that the nonzero elements in the precision matrix corresponds to the nonzero coefficients of the associated (high dimensional) mean regression. The family of Gaussian distributions with this property is known as a Gauss-Markov random field with respect to the graph $G$. This observation has motivated a large literature [53] and interesting extensions that allow for transformations of Gaussian variables [58, 57].■

In order to achieve a tractable concept for non-Gaussian settings, we use that (2.1) occurs if and only if

$$F_{X_a}(\cdot|X_{V\setminus\{a\}}) = F_{X_a}(\cdot|X_{V\setminus\{a,b\}}) \;\text{ for all }\; X_{V\setminus\{a\}} \in \mathcal{X}_{V\setminus\{a\}}. \tag{2.2}$$

In turn, by the equivalence between conditional probabilities and conditional quantiles to characterize a random variable, we have that (2.1) occurs if and only if

$$Q_{X_a}(\tau|X_{V\setminus\{a\}}) = Q_{X_a}(\tau|X_{V\setminus\{a,b\}}) \quad \text{for all } \tau \in (0, 1), \;\text{ and }\; X_{V\setminus\{a\}} \in \mathcal{X}_{V\setminus\{a\}}. \tag{2.3}$$

For a quantile index $\tau \in (0, 1)$, the $\tau$-quantile conditional independence graph is a directed graph $G^I(\tau) = (V, E^I(\tau))$ with vertex set $V$ and edge set $E^I(\tau)$. An edge $(a, b)$ is not contained in the edge set $E^I(\tau)$ if

and only if

$$Q_{X_a}(\tau|X_{V\setminus\{a\}}) = Q_{X_a}(\tau|X_{V\setminus\{a,b\}}) \text{ for all } X_{V\setminus\{a\}} \in \mathcal{X}_{V\setminus\{a\}}. \tag{2.4}$$

By the equivalence between (2.2) and (2.3), the union of $\tau$-quantile graphs over $\tau \in (0,1)$ represents the conditional independence structure of $X$, namely $E^I = \cup_{\tau\in(0,1)} E^I(\tau)$. We also consider a relaxation of (2.1). For a set of quantile indices $\mathcal{T} \subset (0,1)$, we say that

$$X_a \perp_{\mathcal{T}} X_b \mid X_{V\setminus\{a,b\}}, \tag{2.5}$$

$X_a$ and $X_b$ are $\mathcal{T}$-conditionally independent given $X_{V\setminus\{a,b\}}$, if (2.4) holds for all $\tau \in \mathcal{T}$. Thus, we have that (2.1) implies (2.5).We define the $\mathcal{T}$-quantile graph as $G^I(\mathcal{T}) = (V, E^I(\mathcal{T}))$ where

$$E^I(\mathcal{T}) = \cup_{\tau\in\mathcal{T}} E^I(\tau).$$

Although the conditional independence concept relates to all quantile indices, the quantile characterization described above also lends itself to quantile specific impacts which can be of independent interest.[2]

**Comment 2.2** (Simulation and Conditional Independence)**.** Although not pursued in this work, the tools developed here can also be used to develop simulation tools for high-dimensional random vectors. Specifically, we can simulate a random vector $X$ as follows

$$X_1 \sim Q_{X_1}(U_1), \quad X_2 \sim Q_{X_2}(U_2 \mid X_1), \quad X_3 \sim Q_{X_3}(U_3 \mid X_1, X_2), \quad \ldots, \quad X_d \sim Q_{X_d}(U_d \mid X_1, \ldots, X_{d-1})$$

where $U_j \sim \text{Uniform}(0,1)$, for all $j \in [d]$ and estimates of the conditional quantiles can be obtained based on a sample $(X_i \in \mathbb{R}^d)_{i=1}^n$ and the tools discussed here. It is clear that the order of the procedure can impact the estimation. In particular, if most variables are independent of (say) $X_d$, skipping them from the process are likely to increase the accuracy of the simulation procedure.

2.2. **Prediction Quantile Graphical Models.** Prediction Quantile Graphical Models (PQGMs) are motivated by prediction accuracy under an asymmetric loss function (instead of conditional independence as in Section 2.1). More precisely, for each $a \in V$, we are interested in predicting $X_a$ based on linear combinations of the remaining variables, $X_{V\setminus\{a\}}$, where accuracy is measured with respect to an asymmetric loss function. Formally, PQGMs measure accuracy as

$$\mathcal{L}_a(\tau \mid V\setminus\{a\}) = \min_{\beta} \mathrm{E}[\rho_\tau(X_a - X'_{-a}\beta)] \tag{2.6}$$

where $X_{-a} = (1, X'_{V\setminus\{a\}})'$, and the asymmetric loss function $\rho_\tau(t) = (\tau - 1\{t \leqslant 0\})t$ is the check function used in [51].

Importantly, PQGMs are concerned with the best linear predictor under the asymmetric loss function $\rho_\tau$ which is a specification that is widely used in practice. This is a fundamental distinction with respect to CIQGMs discussed in Section 2.1 where the specification of the conditional quantile was approximately a linear function of transformations of $X_{V\setminus\{a\}}$.[3] Indeed, we note that under suitable conditions the linear predictor that solves the minimization problem in (2.6) approximates the conditional quantile regression as shown in [15]. (In fact, the conditional quantile function would be linear if $X_V$ was jointly Gaussian distributed.) However, PQGMs do not assume that the conditional quantile function of $X_a$ is well approximated by a linear function and instead it focuses on the best linear predictor.

[2]For example, we might be interested in some extreme events which typically correspond to crises in financial systems.

[3]In Section 2.1 the vector $Z^a$ in equation (3.10) collects the functions of the vector $X_{V\setminus\{a\}}$.

We define that $X_b$ is predictively uninformative for $X_a$ given $X_{V\setminus\{a,b\}}$ if

$$\mathcal{L}_a(\tau \mid V\setminus\{a\}) = \mathcal{L}_a(\tau \mid V\setminus\{a,b\}) \quad \text{for all} \quad \tau \in (0,1),$$

i.e., considering a linear function of $X_b$ will not improve our performance of predicting $X_a$ with respect to the asymmetric loss function $\rho_\tau$.

Again we can visualize the predictive relationship through a graph process indexed by $\tau \in (0,1)$. That is, for each $\tau \in (0,1)$ we have a directed graph $G^P(\tau) = (V, E^P(\tau))$, where an edge $(a,b) \in G^P(\tau)$ only if $X_b$ is predictively informative for $X_a$ given $X_{V\setminus\{a,b\}}$ at the quantile $\tau$. Finally, it is also convenient to define the PQGM associated with a subset $\mathcal{T} \subset (0,1)$ as $G^P(\mathcal{T}) = (V, E^P(\mathcal{T}))$ where

$$E^P(\mathcal{T}) = \cup_{\tau\in\mathcal{T}} E^P(\tau).$$

2.3. **$\mathcal{W}$-Conditional Quantile Graphical Models.** In what follows, we discuss an extension of the QGMs discussed in Sections 2.1 and 2.2 to allows for conditioning on a (possible infinity) family of events $\varpi \in \mathcal{W}$.[4] Such extension is motivated by several applications in which the interdependence between the random variables in $X_V$ maybe substantially impacted by additional observable events (e.g. downside movements of the market). This general framework allows different forms of conditioning. The main implication of this extension is that QGMs are now graph processes indexed by $\tau \in \mathcal{T} \subset (0,1)$ and $\varpi \in \mathcal{W}$.

We define $X_a$ and $X_b$ are $(\mathcal{T}, \varpi)$-conditionally independent,

$$X_a \perp_{\mathcal{T}} X_b \;\mid\; X_{V\setminus\{a,b\}}, \varpi \tag{2.7}$$

if for all $\tau \in \mathcal{T}$ we have

$$Q_{X_a}(\tau | X_{V\setminus\{a\}}, \varpi) = Q_{X_a}(\tau | X_{V\setminus\{a,b\}}, \varpi). \tag{2.8}$$

The conditional independence edge set associated with $(\tau, \varpi)$ is defined analogously as before. We denote them by $E^I(\tau, \varpi)$ and $E^I(\mathcal{T}, \varpi) = \cup_{\tau\in\mathcal{T}} E^I(\tau, \varpi)$ for each $\varpi \in \mathcal{W}$.

The extension of PQGMs proceeds by defining the accuracy under the asymmetric loss function conditionally on $\varpi$. More precisely, we define

$$\mathcal{L}_a(\tau | V\setminus\{a\}, \varpi) = \min_\beta \mathrm{E}[\rho_\tau(X_a - X'_{-a}\beta) \mid \varpi]. \tag{2.9}$$

The prediction edge set associated with $(\tau, \varpi)$ is also defined analogously as before. We denote them by $E^P(\tau, \varpi)$ and $E^P(\mathcal{T}, \varpi) = \cup_{\tau\in\mathcal{T}} E^P(\tau, \varpi)$, for each $\varpi \in \mathcal{W}$.

**Example 1** (PQGMs for Stock Returns Under Market Downside Movements)**.** Hedging decisions rely on the dependence of various stocks returns. Moreover, hedging is even more relevant during market downside movements, which motivates us to understand interdependence conditional on those events. We can parameterize the downside movements by using a random variable $M$, which could be the market index, and conditional on the event $\Omega_\varpi = \{M \leqslant \varpi\}$. This allows us to define a $\varpi$-conditional-CIQGM as $G^I(\tau, \varpi) = (V, E^I(\tau, \varpi))$ and a $\varpi$-conditional-PQGM as $G^P(\tau, \varpi) = (V, E^P(\tau, \varpi))$, for each $\varpi \in \mathcal{W}$. ∎

[4]With a slight abuse of notation, we let $\varpi$ to denote the event and also the index of such event. For example, we write $\mathrm{P}(\varpi)$ as a shorthand for $\mathrm{P}(W \in \Omega_\varpi)$.

## 3. Estimators for High-Dimensional Quantile Graphical Models

In this section, we propose and discuss estimators for QGMs introduced in Section 2. Throughout it is assumed that we observe a $d$-dimensional i.i.d. random vector $X_V$, namely $\{X_{iV} : i = 1, \ldots, n\}$. Based on the data observed, unless additional assumptions are imposed we cannot estimate the quantities of interest for all $\tau \in (0, 1)$. Instead, in what follows we will consider a (compact) set of quantile index $\mathcal{T} \subset (0, 1)$. The estimators are intended to handle high dimensional models and a continuum of conditioning events in $\mathcal{W}$.

### 3.1. Estimators for CIQGMs.

We discuss the specification and propose an estimator for CIQGMs. Although in general it is potentially hard to correctly specify coherent models, the following are simple examples.

**Example 2** (Gaussian)**.** Consider the Gaussian case, $X_V \sim N(\mu, \Sigma)$. It follows that for each $a \in V$, the conditional distribution $X_a \mid X_{V\setminus\{a\}}$ satisfies

$$X_a \mid X_{V\setminus\{a\}} \sim N\left(\mu_a - \sum_{j\in V\setminus\{a\}} \frac{(\Sigma^{-1})_{aj}}{(\Sigma^{-1})_{aa}}(X_j - \mu_j), \frac{1}{(\Sigma^{-1})_{aa}}\right).$$

Therefore the conditional quantile function of $X_a$ is linear in $X_{V\setminus\{a\}}$ and is given by

$$Q_{X_a}(\tau|X_{V\setminus\{a\}}) = \frac{\Phi^{-1}(\tau)}{(\Sigma^{-1})_{aa}^{1/2}} + \mu_a - \sum_{j\in V\setminus\{a\}} \frac{(\Sigma^{-1})_{aj}}{(\Sigma^{-1})_{aa}}(X_j - \mu_j).$$

**Example 3** (Mixture of Gaussians)**.** Similar to the prior example, consider the case $X_V \mid \varpi \sim N(\mu_\varpi, \Sigma_\varpi)$ for each $\varpi \in \mathcal{W}$. It follows that for $a \in V$, the conditional distribution satisfies

$$X_a \mid X_{V\setminus\{a\}}, \varpi \sim N\left(\mu_{\varpi a} - \sum_{j\in V\setminus\{a\}} \frac{(\Sigma^{-1})_{\varpi aj}}{(\Sigma^{-1})_{\varpi aa}}(X_j - \mu_{\varpi j}), \frac{1}{(\Sigma^{-1})_{\varpi aa}}\right).$$

Again the conditional quantile function of $X_a$ is linear in $X_{V\setminus\{a\}}$ and is given by

$$Q_{X_a}(\tau|X_{V\setminus\{a\}}, \varpi) = \frac{\Phi^{-1}(\tau)}{(\Sigma^{-1})_{\varpi aa}^{1/2}} + \mu_{\varpi a} - \sum_{j\in V\setminus\{a\}} \frac{(\Sigma^{-1})_{\varpi aj}}{(\Sigma^{-1})_{\varpi aa}}(X_j - \mu_{\varpi j}).$$

**Example 4** (Monotone Transformations)**.** Consider the Gaussian case, for each $a \in V$, $X_a = h_a(Y_a)$ and $Y_V \sim N(\mu, \Sigma)$. It follows that for each $a \in V$, the conditional quantile function satisfies

$$Q_{X_a}(\tau|X_{V\setminus\{a\}}) = h_a\left(\frac{\Phi^{-1}(\tau)}{(\Sigma^{-1})_{aa}^{1/2}} + \mu_a - \sum_{j\in V\setminus\{a\}} \frac{(\Sigma^{-1})_{aj}}{(\Sigma^{-1})_{aa}}(h_j^{-1}(X_j) - \mu_j)\right).$$

In particular, if $(h_a : a \in V)$ are monotone polynomials, the expression above is a sum of monomials with fractional and integer exponents.

**Example 5** (Multiplicative Error Model)**.** Consider $d = 2$ so that $V = \{1, 2\}$. Assume that $X_2$ and $\varepsilon$ are independent positive random variables. Assume further that they relate to $X_1$ as

$$X_1 = \alpha + \varepsilon X_2.$$

In this case, we have that the conditional quantile functions are linear and given by

$$Q_{X_1}(\tau|X_2) = \alpha + F_\varepsilon^{-1}(\tau)X_2 \quad \text{and} \quad Q_{X_2}(\tau|X_1) = (X_1 - \alpha)/F_\varepsilon^{-1}(1 - \tau).$$

■

**Example 6** (Additive Error Model)**.** Consider $d = 2$ so that $V = \{1, 2\}$. Let $X_2 \sim U(0, 1)$ and $\varepsilon \sim U(0, 1)$ be independent random variables. Also define the random variable $X_1$ as

$$X_1 = \alpha + \beta X_2 + \varepsilon.$$

It follows that $Q_{X_1}(\tau|X_2) = \alpha + \beta X_2 + \tau$. However, if $\beta = 0$, we have $Q_{X_2}(\tau|X_1) = \tau$, and for $\beta > 0$, direct calculations yield that

$$Q_{X_2}(\tau|X_1) = \begin{cases} \frac{\tau}{\beta}(X_1 - \alpha), & \text{if} \quad X_1 \leqslant \alpha + \beta \\ \tau + (1 - \tau)(X_1 - \alpha - \beta), & \text{if} \quad X_1 \geqslant \alpha + \beta \end{cases}$$

where we note that $X_1 \in [\alpha, 1 + \alpha + \beta]$. ∎

Although a linear specification is correct for Examples 2 and 5, Example 6 illustrates that we need to consider a more general transformation of the covariates $X_V$ in the specification for each conditional quantile function. Nonetheless, specifications with additional non-linear terms can approximate non-drastic departures from normality.

We will consider a conditional quantile representation for each $a \in V$. It is based on transformations of the original covariates $X_{V\setminus\{a\}}$ that create a $p$-dimensional random vector $Z^a = Z^a(X_{V\setminus\{a\}})$ such that

$$Q_{X_a}(\tau|X_{V\setminus\{a\}}) = Z^a\beta_{a\tau} + r_{a\tau}, \quad \beta_{a\tau} \in \mathbb{R}^p, \quad \text{for all } \tau \in \mathcal{T}, \tag{3.10}$$

where $r_{a\tau}$ denotes a small approximation error. For $b \in V\setminus\{a\}$ we let $I_a(b) := \{j : Z^a_j \text{ depends on } X_b\}$. That is, $I_a(b)$ contains the components of $Z^a$ that are functions of $X_b$. Under correct specification, if $X_a$ and $X_b$ are conditionally independent, we have $\beta_{a\tau j} = 0$ for all $j \in I_a(b)$, $\tau \in (0, 1)$.

This allows us to connect the conditional independence quantile graph estimation problem with model selection with quantile regression. Indeed, the representation (3.10) has been used in several quantile regression models, see [50]. Under mild conditions this model allows us to identify the process $(\beta_{a\tau})_{\tau\in\mathcal{T}}$ as the solution of the following moment equation

$$\mathrm{E}[(\tau - 1\{X_a \leqslant Z^a\beta_{a\tau} + r_{a\tau}\})Z^a] = 0. \tag{3.11}$$

In order to allow for a flexible specification, so that the approximation errors are negligible, it is attractive to consider a high-dimensional $Z^a$ where its dimension $p$ is possibly larger than the sample size $n$. In turn, having a large number of technical controls creates an estimation challenge if the number of coefficients $p$ is not negligible with respect to the sample size $n$. In such a high dimensional setting, a widely applicable condition that makes estimation possible is approximate sparsity [40, 10, 17]. Formally we require

$$\max_{a\in V}\sup_{\tau\in\mathcal{T}}\|\beta_{a\tau}\|_0 \leqslant s, \quad \max_{a\in V}\sup_{\tau\in\mathcal{T}}\{\mathrm{E}[r^2_{a\tau}]\}^{1/2} \lesssim \sqrt{s/n}, \quad \text{and} \quad \max_{a\in V}\sup_{\tau\in\mathcal{T}}|\mathrm{E}[f_{a\tau}r_{a\tau}Z^a]| = o(n^{-1/2}), \tag{3.12}$$

where the sparsity parameter $s$ of the model is allowed to grow (at a slower rate) as $n$ grows, and $f_{a\tau} = f_{X_a|X_{V\setminus\{a\}}}(Q_{X_a}(\tau|X_{V\setminus\{a\}})|X_{V\setminus\{a\}})$ denotes the conditional density function evaluated at the corresponding conditional quantile value. This sparsity also has implications on the maximum degree of the associated quantile graph.

Algorithm 3.1 below contains our proposal to estimate $\beta_{a\tau}$, $a \in V$, $\tau \in \mathcal{T}$. It is based on three procedures in order to overcome high-dimensionality. In the first step, we apply a (post-)$\ell_1$-penalized quantile regression. The second step applies (post-)Lasso where the data is weighted by the conditional density function at the

conditional quantile.[5] Finally, the third step relies on constructing (orthogonal) score function that provides immunity to (unavoidable) model selection mistakes.

There are several parameters that need to be specified for Algorithm 3.1. The penalty parameter $\lambda_{V\mathcal{T}}$ is chosen to be larger than the $\ell_\infty$-norm of the (rescaled) score at the true quantile function. The work in [11] exploits the fact that this quantity is pivotal in their setting. Here, additional correlation structure would have an impact and the distribution is pivotal only for each $a \in V$. The penalty is based on the maximum of the quantiles of the following random variables (each with pivotal distribution), for $a \in V$

$$\Lambda_{a\mathcal{T}} = \sup_{\tau\in\mathcal{T}} \max_{j\in[p]} \frac{|\mathbb{E}_n[(1\{U \leqslant \tau\} - \tau)Z_j^a]|}{\sqrt{\tau(1-\tau)}\widehat{\sigma}_{aj}^Z} \tag{3.13}$$

where $\{U_i : i = 1, \ldots, n\}$ are i.i.d. uniform $(0, 1)$ random variables, and $\widehat{\sigma}_{aj}^Z = \{\mathbb{E}_n[(Z_j^a)^2]\}^{1/2}$ for $j \in [p]$. The penalty parameter $\lambda_{V\mathcal{T}}$ is defined as

$$\lambda_{V\mathcal{T}} := \max_{a\in V} \Lambda_{a\mathcal{T}}(1 - \xi/|V| \mid Z^a),$$

that is, the maximum of the $1-\xi/|V|$ conditional quantile of $\Lambda_{a\mathcal{T}}$ given in (3.13). Regarding the penalty term for the weighted Lasso in Step 2, we recommend a (theoretically valid) iterative choice. We refer to Appendix A for the implementation details of the algorithm. We denote $\|\beta\|_{1,\widehat{\sigma}^Z} := \sum_j \widehat{\sigma}_{aj}^Z|\beta_j|$ the standardized version of the $\ell_1$-norm.

**Algorithm 3.1.** (CIQGM Estimator.) *For each $a \in V$, $\tau \in \mathcal{T}$, and $j \in [p]$*

Step 1. *Compute $\widehat{\beta}_{a\tau}$ from $\|\cdot\|_{1,\widehat{\sigma}^Z}$-penalized $\tau$-quantile regression of $X_a$ on $Z^a$ with penalty $\lambda_{V\mathcal{T}}\sqrt{\tau(1-\tau)}$. Compute $\widetilde{\beta}_{a\tau}$ from $\tau$-quantile regression of $X_a$ on $\{Z_k^a : |\widehat{\beta}_{a\tau k}| \geqslant \lambda_{V\mathcal{T}}\sqrt{\tau(1-\tau)}/\widehat{\sigma}_{ak}^Z\}$.*

Step 2. *Compute $\widetilde{\gamma}_{a\tau}^j$ from the post-Lasso estimator of $f_{a\tau}Z_j^a$ on $f_{a\tau}Z_{-j}^a$.*

Step 3. *Construct the score function $\widehat{\psi}_i(\alpha) = (\tau - 1\{X_{ia} \leqslant Z_{ij}^a\alpha + Z_{i,-j}^a\widetilde{\beta}_{a\tau,-j}\})f_{ia\tau}(Z_{ij}^a - Z_{i,-j}^a\widetilde{\gamma}_{a\tau}^j)$ and for $L_{a\tau j}(\alpha) = |\mathbb{E}_n[\widehat{\psi}_i(\alpha)]|^2/\mathbb{E}_n[\widehat{\psi}_i^2(\alpha)]$, set $\check{\beta}_{a\tau j} \in \arg\min_{\alpha\in\mathcal{A}_{a\tau j}} L_{a\tau j}(\alpha)$.*

Algorithm 3.1 above has been studied in [18] where it is applied to a single triple $(a, \tau, j)$, and we have used the following parameter space for $\alpha$, $\mathcal{A}_{a\tau j} = \{\alpha \in \mathbb{R} : |\alpha - \widetilde{\beta}_{a\tau j}| \leqslant 10/\{\widehat{\sigma}_{aj}^Z \log n\}\}$. Under similar conditions, results that hold uniformly over $(a, \tau, j) \in V \times \mathcal{T} \times [p]$ are achievable (as shown in the next sections) building upon the tools developed in [11] and [25]. Algorithm 3.1 is tailored to achieve good rates of convergence in the $\ell_\infty$-norm. In particular, under standard regularity conditions, with probability approaching to 1 we have

$$\sup_{\tau\in\mathcal{T}} \|\beta_{a\tau} - \check{\beta}_{a\tau}\|_\infty \lesssim \sqrt{\frac{\log(p|V|n)}{n}}.$$

In order to create an estimate of $E^I(\tau) = \{(a, b) \in V \times V : \max_{j\in I_a(b)} |\beta_{a\tau j}| > 0\}$, we define

$$\widehat{E}^I(\tau) = \left\{(a, b) \in V \times V : \max_{j\in I_a(b)} \frac{|\check{\beta}_{a\tau j}|}{\mathrm{se}(\check{\beta}_{a\tau j})} > \overline{\mathrm{cv}}\right\}$$

where $\mathrm{se}(\check{\beta}_{a\tau j}) = \{\tau(1-\tau)\mathbb{E}_n[\widetilde{v}_{ia\tau j}^2]^{-1}\}^{1/2}$ with $\widetilde{v}_{ia\tau j} = \widehat{f}_{ia\tau}\{Z_{ij}^a - Z_{i,-j}^a\widetilde{\gamma}_{a\tau}^j\}$, is an estimate of the standard deviation of the estimator, and the critical value $\overline{\mathrm{cv}}$ is set to account for the uniformity over $a \in V$, $\tau \in \mathcal{T}$, and $j \in [p]$. We discuss in the following sections a data driven procedure based on multiplier bootstrap that is theoretically valid in this high dimensional setting.

[5] We note that an estimate for $f_{a\tau}$ is available from $\ell_1$-penalized quantile regression estimators for $\tau + h$ and $\tau - h$ where $h$ is a bandwidth parameter, see [50, 18] and Comment 3.2.

**Comment 3.1** (Stepdown Procedure for $\overline{\text{cv}}$)**.** Setting a critical value $\overline{\text{cv}}$ that accounts for the multiple hypotheses being tested plays an important role to estimate the graph $\widehat{E}^I(\tau)$. Further improvements can be obtained by considering the stepdown procedure of [68] for multiple hypothesis testing that was studied for the high-dimensional case in [24]. The procedure iteratively creates a suitable sequence of decreasing critical values. In each step only null hypotheses that were not rejected are considered to determine the critical value. Thus, as long as any hypothesis is rejected at a step, the critical value decreases and we continue to the next iteration. The procedure stops when no hypothesis in the current active set is rejected. ∎

**Comment 3.2** (Estimation of Conditional Density Function)**.** The algorithm above requires the conditional density function $f_{a\tau}$ which typically needs to be estimated in practice. It turns out that estimation of conditional quantiles yields a natural estimator for the conditional density function as

$$f_{a\tau} = \frac{1}{\partial Q_{X_a}(\tau|Z^a)/\partial\tau}.$$

Therefore, based on $\ell_1$-penalized quantile regression estimates at the $\tau + h_n$ and $\tau - h_n$ quantile, where $h = h_n \to 0$ denotes a bandwidth parameter, we have

$$\widehat{f}_{a\tau} = \frac{2h}{\widehat{Q}_{X_a}(\tau + h|Z^a) - \widehat{Q}_{X_a}(\tau - h|Z^a)} \tag{3.14}$$

as an estimator of $f_{a\tau}$. Under smoothness conditions, it has an bias of order $h^2$. See [18] and the references therein for additional comments and estimators. ∎

3.2. **Estimators for PQGMs.** In this section we propose an estimator for PQGMs in which case we are interested in the prediction of $X_a$, $a \in V$, using a linear combination of $X_{V\setminus\{a\}}$ under the asymmetric loss discussed in (2.6). We will add an intercept as one of the variables for the sake of notation so that $X_{-a} = (1, X'_{V\setminus\{a\}})'$. Given the loss function $\rho_\tau$, the target $d$-dimensional vector of parameters $\beta_{a\tau}$ is defined as (part of) the solution of the following optimization problem

$$\beta_{a\tau} \in \arg\min_{\beta} \ \mathrm{E}[\rho_\tau(X_a - X'_{-a}\beta)]. \tag{3.15}$$

As we are interested in the case that $d$ is large, the use of high-dimensional tools to achieve consistent estimators is needed. The estimation procedure we proposed is based on $\ell_1$-penalized quantile regression but additional issues need to be considered to cope with the (non-vanishing) difference between the best linear predictor and the conditional quantile function. Again we consider models that satisfy an approximately sparse condition. Formally, we require the existence of sparse coefficients $\{\bar{\beta}_{a\tau} : a \in V, \tau \in \mathcal{T}\}$ such that

$$\max_{a\in V} \sup_{\tau\in\mathcal{T}} \|\bar{\beta}_{a\tau}\|_0 \leqslant s \quad \text{and} \quad \max_{a\in V} \sup_{\tau\in\mathcal{T}} \{\mathrm{E}[\{X'_{-a}(\beta_{a\tau} - \bar{\beta}_{a\tau})\}^2]\}^{1/2} \lesssim \sqrt{s/n}, \tag{3.16}$$

where (again) the sparsity parameter $s$ of the model is allowed to grow as $n$ grows. The high-dimensionality prevents us from using (standard) quantile regression methods and regularization methods are needed to achieve good prediction properties.

A key issue is to set the penalty parameter properly so that it bounds from above

$$\max_{a\in V} \sup_{\tau\in\mathcal{T}} \max_{j\in[d]} |\mathbb{E}_n[(1\{X_a \leqslant X'_{-a}\beta_{a\tau}\} - \tau)X_{-a,j}]|. \tag{3.17}$$

However, it is important to note that we do not assume that the conditional quantile of $X_a$ is a linear function of $X_{-a}$. Under correct linear specification of the conditional quantile function, $\ell_1$-penalized quantile

regression estimator has been studied in [11]. The case that the conditional quantile function differs from a linear specification by vanishing approximation errors has been considered in [47] and [18]. The analysis proposed here aims to allow for non-vanishing misspecification of the quantile function relative to a linear specification while still guarantees good rates of convergence in the $\ell_2$-norm to the best linear specification. Thus the penalty parameter in the penalized quantile regression needs to account for such misspecification and is no longer pivotal as in [11].

In order to handle this issue we propose a two step estimation procedure. In the first step, the penalty parameter $\lambda_0$ is conservative and is set via bounds constructed based on symmetrization arguments, similar in spirit to [70, 12]. This leads to $\lambda_0 = 2(1 + 1/16)\sqrt{2\log(8|V|^2/\xi)/n}$. Although this is conservative, under mild conditions this would lead to estimates that can be leverage to fine tune the penalty choice. The second step uses the preliminary estimator to bootstrap (3.17) based on the tools in [24] as follows. Specifically, for estimates $\widehat{\varepsilon}_{ia\tau}$ of the "noise" $\varepsilon_{ia\tau} = 1\{X_{ia} \leqslant X'_{i,-a}\beta_{a\tau}\} - \tau$ for $i \in [n]$, for $a \in V$ define

$$\bar{\Lambda}_{a\mathcal{T}} := 1.1 \sup_{\tau\in\mathcal{T}} \max_{j\in[d]} \frac{|\mathbb{E}_n[g_i \widehat{\varepsilon}_{ia\tau} X_{i,-aj}]|}{\{\mathbb{E}_n[\widehat{\varepsilon}^2_{ia\tau} X^2_{i,-aj}]\}^{1/2}} \tag{3.18}$$

where $(g_i)_{i=1}^n$ is a sequence of i.i.d. standard Gaussian random variables. The new penalty parameter $\bar{\lambda}_{V\mathcal{T}}$ is defined as

$$\bar{\lambda}_{V\mathcal{T}} := \max_{a\in V} \bar{\Lambda}_{a\mathcal{T}}(1 - \xi|X_{-a}) \tag{3.19}$$

that is, the maximum of the $(1 - \xi)$ conditional quantile of $\bar{\Lambda}_{a\mathcal{T}}$. The penalty choice above adapts to the unknown correlation structure across components and quantile indices. The following algorithm states the procedure where we denote weighted $\ell_1$-norms by $\|\beta\|_{1,\widehat{\sigma}^X} := \sum_j \widehat{\sigma}^X_{aj}|\beta_j|$ with $\widehat{\sigma}^X_{aj} = \{\mathbb{E}_n[X^2_j]\}^{1/2}$ the standardized version of the $\ell_1$-norm and $\|\beta\|_{1,\widehat{\varepsilon}} := \sum_j \widehat{\sigma}^{\varepsilon X}_{a\tau j}|\beta_j|$ with $\widehat{\sigma}^{\varepsilon X}_{a\tau j} = \{\mathbb{E}_n[\widehat{\varepsilon}^2_{a\tau} X^2_{-a,j}]\}^{1/2}$ a norm based on the estimated residuals.

**Algorithm 3.2.** (PQGM Estimator.) *For each $a \in V$, and $\tau \in \mathcal{T}$*

Step 1. *Compute $\widehat{\beta}_{a\tau}$ from $\|\cdot\|_{1,\widehat{\sigma}^X}$-penalized $\tau$-quantile regression of $X_a$ on $X_{-a}$ with penalty $\lambda_0$. Compute $\widetilde{\beta}_{a\tau}$ from $\tau$-quantile regression of $X_a$ on $\{X_k : |\widehat{\beta}_{a\tau k}| \geqslant \lambda_0/\widehat{\sigma}^X_{ak}\}$.*

Step 2. *For $\widehat{\varepsilon}_{ia\tau} = 1\{X_{ia} \leqslant X'_{i,-a}\widetilde{\beta}_{a\tau}\} - \tau$ for $i \in [n]$, and $\xi = 1/n$, compute $\bar{\lambda}_{V\mathcal{T}}$ via (3.19).*

Step 3. *Recompute $\widehat{\beta}_{a\tau}$ from $\|\cdot\|_{1,\widehat{\varepsilon}}$-penalized $\tau$-quantile regression of $X_a$ on $X_{-a}$ with penalty $\bar{\lambda}_{V\mathcal{T}}$. Compute $\check{\beta}_{a\tau}$ from $\tau$-quantile regression of $X_a$ on $\{X_k : |\widehat{\beta}_{a\tau k}| \geqslant \bar{\lambda}_{V\mathcal{T}}/\widehat{\sigma}^{\varepsilon X}_{a\tau k}\}$.*

Under regularity conditions stated in Section 4, with probability approaching 1, we have

$$\max_{a\in V} \sup_{\tau\in\mathcal{T}} \|\beta_{a\tau} - \check{\beta}_{a\tau}\| \lesssim \sqrt{\frac{s\log(|V|n)}{n}}.$$

The estimate of the prediction quantile graph is given by the support of $(\check{\beta}_{a\tau})_{a\in V,\tau\in\mathcal{T}}$, namely

$$\widehat{E}^P(\tau) = \left\{(a,b) \in V \times V : \ |\widehat{\beta}_{a\tau b}| > \bar{\lambda}_{V\mathcal{T}}/\widehat{\sigma}^{\varepsilon X}_{a\tau b}\right\}.$$

That is, it is induced by covariates selected by the $\ell_1$-penalized estimator. Those thresholded estimators not only have the same rates of convergence as of the original penalized estimators but also possess additional sparsity guarantees.

3.3. **Estimators for $\mathcal{W}$-Conditional Quantile Graphical Models.** In order to handle the additional conditioning events $\Omega_\varpi$, $\varpi \in \mathcal{W}$, we propose to modify Algorithms 3.1 and 3.2 based on kernel smoothing. To that extent, we assume the observed data is of the form $\{(X_{iV}, W_i) : i = 1, \ldots, n\}$, where $W_i$ might be defined through additional variables. Furthermore, we assume for each conditioning event $\varpi \in \mathcal{W}$ we have access to a kernel function $K_\varpi$ that is applied to $W$, to represent the relevant observations associated with $\varpi$ (recall that we denote $\mathrm{P}(W \in \Omega_\varpi)$ as $\mathrm{P}(\varpi)$). We assume that $K_\varpi(W) = 1\{W \in \Omega_\varpi\}$.

**Example 7** (PQGMs for Stock Returns Under Market Downside Movements, continued)**.** In Example 1, we have $W$ as the market return and the conditioning event as $\Omega_\varpi = \{W \leqslant \varpi\}$ which is parameterized by $\varpi \in \mathcal{W}$, a closed interval in $\mathbb{R}$. We might be interest in a fixed $\varpi$ or on a family of values $\varpi \in (-\bar{\varpi}, 0]$. The latter induces $\mathcal{W} = \{\Omega_\varpi = \{W \leqslant \varpi\} : \varpi \in (-\bar{\varpi}, 0]\}$. The kernel function is simply $K_\varpi(t) = 1\{t \leqslant \varpi\}$.

This framework encompasses the previous framework by having $K_\varpi(W) = 1$ for all $W$. Moreover, it allows for a richer class of estimands which require estimators whose properties should hold uniformly over $\varpi \in \mathcal{W}$ as well. Next we propose estimators for this setting, i.e. we generalize the previous methods to account for the additional conditioning on $\varpi \in \mathcal{W}$. In what follows, with a slight abuse of notation we use $\varpi$ to denote not only the index but also the event $\Omega_\varpi$. For further notational convenience, we denote $u = (a, \tau, \varpi) \in \mathcal{U} := V \times \mathcal{T} \times \mathcal{W}$ so that the set $\mathcal{U}$ collects all the three relevant indices. With $\widehat{\sigma}^Z_{a\varpi j} = \{\mathbb{E}_n[K_\varpi(W)(Z^a_j)^2]\}^{1/2}$, we define the following weighted $\ell_1$-norm $\|\beta\|_{1,\varpi} = \sum_{j\in[p]} \widehat{\sigma}^Z_{a\varpi j}|\beta_j|$. This norm is $\varpi$ dependent and provides the proper adjustments as we condition on different events associated with different $\varpi$'s.

We first consider estimators of CIGMs conditional on the events in $\mathcal{W}$. In this setting, the model is correctly specified up to small approximation errors. The definition of the penalty parameter will be based on the random variable

$$\Lambda_{a\mathcal{T}\mathcal{W}} = \sup_{\tau\in\mathcal{T},\varpi\in\mathcal{W}} \max_{j\in[p]} \left| \frac{\mathbb{E}_n[K_\varpi(W)(1\{U \leqslant \tau\} - \tau)Z^a_j]}{\sqrt{\tau(1-\tau)}\widehat{\sigma}^Z_{a\varpi j}} \right|$$

where $U_i$ are independent uniform $(0, 1)$ random variables, and set the penalty

$$\lambda_{V\mathcal{T}\mathcal{W}} = \max_{a\in V} \Lambda_{a\mathcal{T}\mathcal{W}}(1 - \xi/\{|V|n^{1+2d_W}\}|Z^a, W),$$

that is, the maximum of the $(1 - \xi/\{|V|n^{1+2d_W}\})$ conditional quantile of $\Lambda_{a\mathcal{T}\mathcal{W}}$. Algorithm 3.3 provides the definition of the estimator. Here $\mathcal{A}_{uj} = \{\alpha \in \mathbb{R} : |\alpha - \widetilde{\beta}_{uj}| \leqslant 10/\{\widehat{\sigma}^Z_{a\varpi j} \log n\}\}$, and denote $\lambda_u := \lambda_{V\mathcal{T}\mathcal{W}}\sqrt{\tau(1-\tau)}$.

**Algorithm 3.3.** ($\mathcal{W}$-Conditional CIQGM Estimator.) *For $(a, \tau, \varpi) \in V \times \mathcal{T} \times \mathcal{W}$ and $j \in [p]$*

Step 1. *Compute $\widehat{\beta}_u$ from $\|\cdot\|_{1,\varpi}$-penalized $\tau$-quantile regression of $K_\varpi(W)(X_a; Z^a)$ with penalty $\lambda_u$. Compute $\widetilde{\beta}_u$ from $\tau$-quantile regression of $K_\varpi(W)(X_a; \{Z^a_k : |\widehat{\beta}_{uk}| \geqslant \lambda_u/\widehat{\sigma}^Z_{a\varpi j}\})$.*

Step 2. *Compute $\widetilde{\gamma}^j_u$ from the post-Lasso estimator of $K_\varpi(W)f_u Z^a_j$ on $K_\varpi(W)f_u Z^a_{-j}$.*

Step 3. *Construct the score function $\widehat{\psi}_i(\alpha) = K_\varpi(W_i)(\tau - 1\{X_{ia} \leqslant Z^a_{ij}\alpha + Z^a_{i,-j}\widetilde{\beta}_{u,-j}\})f_{iu}(Z^a_{ij} - Z^a_{i,-j}\widetilde{\gamma}^j_u)$ and for $L_{uj}(\alpha) = |\mathbb{E}_n[\widehat{\psi}_i(\alpha)]|^2/\mathbb{E}_n[\widehat{\psi}^2_i(\alpha)]$, set $\check{\beta}_{uj} \in \arg\min_{\alpha\in\mathcal{A}_{uj}} L_{uj}(\alpha)$ .*

Next we consider estimators of PQGMs conditional on the events in $\mathcal{W}$. Similar to the previous case, for $a \in V$ define

$$\bar{\Lambda}_{a\mathcal{T}\mathcal{W}} := 1.1 \sup_{\tau\in\mathcal{T},\varpi\in\mathcal{W}} \max_{j\in[d]} \frac{|\mathbb{E}_n[K_\varpi(W)g\widehat{\varepsilon}_{a\tau\varpi}X_{-a,j}]|}{\{\mathbb{E}_n[K_\varpi(W)\widehat{\varepsilon}^2_{a\tau\varpi}X^2_{-a,j}]\}^{1/2}} \tag{3.20}$$

where $(g_i)_{i=1}^n$ is a sequence of i.i.d. standard Gaussian random variables. The new penalty parameter $\bar{\lambda}_{V\mathcal{T}}$ is defined as

$$\bar{\lambda}_{V\mathcal{T}\mathcal{W}} := \max_{a\in V} \bar{\Lambda}_{a\mathcal{T}\mathcal{W}}(1-\xi|X_{-a}) \tag{3.21}$$

that is, the maximum of the $(1-\xi)$ conditional quantile of $\bar{\Lambda}_{a\mathcal{T}\mathcal{W}}$. It will also be useful to define another weighted $\ell_1$-norm, $\|\beta\|_{1,\varpi\widehat{\varepsilon}} := \sum_j \widehat{\sigma}^{\varepsilon X}_{a\tau\varpi j}|\beta_j|$ with $\widehat{\sigma}^{\varepsilon X}_{a\tau\varpi j} = \{\mathbb{E}_n[K_\varpi(W)\widehat{\varepsilon}^2_{a\tau\varpi}X^2_{-a,j}]\}^{1/2}$. We also denote $\widehat{\sigma}^X_{a\varpi j} = \{\mathbb{E}_n[K_\varpi(W)X^2_{-a,j}]\}^{1/2}$. The penalty choice and weighted $\ell_1$-norm adapt to the unknown correlation structure across components and quantile indices. The following algorithm states the procedure, with $\lambda_{0\mathcal{W}} = 2(1+1/16)\sqrt{2\log(8|V|^2\{ne/d_W\}^{2d_W}/\xi)/n}$.

**Algorithm 3.4.** ($\mathcal{W}$-Conditional PQGM Estimator.) *For $(a,\tau,\varpi)\in V\times\mathcal{T}\times\mathcal{W}$*

Step 1. *Compute $\widehat{\beta}_u$ from $\|\cdot\|_{1,\varpi}$-penalized $\tau$-quantile regression of $X_a$ on $X_{-a}$ with penalty $\lambda_{0\mathcal{W}}$.*
*Compute $\widetilde{\beta}_u$ from $\tau$-quantile regression of $K_\varpi(W)(X_a;\{X_{-a,k}:|\widehat{\beta}_{uk}|\geqslant\lambda_{0\mathcal{W}}/\widehat{\sigma}^X_{a\varpi k}\})$.*

Step 2. *For $\widehat{\varepsilon}_{iu} = 1\{X_{ia}\leqslant X'_{i,-a}\widetilde{\beta}_u\}-\tau$ for $i\in[n]$, and $\xi = 1/n$, compute $\bar{\lambda}_{V\mathcal{T}\mathcal{W}}$ via (3.21).*

Step 3. *Recompute $\widehat{\beta}_u$ from $\|\cdot\|_{1,\varpi\widehat{\varepsilon}}$-penalized $\tau$-quantile regression of $K_\varpi(W)(X_a;X_{-a})$ with penalty $\bar{\lambda}_{V\mathcal{T}\mathcal{W}}$.*
*Compute $\check{\beta}_u$ from $\tau$-quantile regression of $K_\varpi(W)(X_a;\{X_{-a,k}:|\widehat{\beta}_{uk}|\geqslant\bar{\lambda}_{V\mathcal{T}\mathcal{W}}/\widehat{\sigma}^{\varepsilon X}_{uk}\})$.*

**Comment 3.3** (Computation of Penalty Parameter over $\mathcal{W}$)**.** The penalty choices require one to maximize over $a\in V$, $\tau\in\mathcal{T}$ and $\varpi\in\mathcal{W}$. The set $V$ is discrete and does not pose a significant challenge. However both other sets are continuous and additional care is needed. In most applications we are concerned with the case that $\mathcal{W}$ is a low dimensional VC class of sets and it impacts the calculation only through indicator functions, which is precisely the case of $\mathcal{T}$. It follows that only a polynomial number (in $n$) of different values of $\tau$ and $\varpi$ would need to be considered. ∎

## 4. Main Theoretical Results

This section is devoted to theoretical guarantees associated with the proposed estimators. We will establish rates of convergence results for the proposed estimators as well as the (uniform) validity of confidence regions. These results build upon and contribute to an increasing literature on the estimation of many processes of interest with (high-dimensional) nuisance parameters.

Throughout, we will provide results for the estimators of the $\mathcal{W}$-conditional quantile graphical models as those can be generalized the other models by setting $K_\varpi(W)=1$. Although some of the tools are similar, CIQGMs and PQGMs require different estimators and are subject to different assumptions. Thus, substantial different analyses are required.

### 4.1. $\mathcal{W}$-Conditional CIQGM.

For $u=(a,\tau,\varpi)\in\mathcal{U}$, define the $\tau$-conditional quantile function of $X_a$ given $X_{V\setminus\{a\}}$ and $\varpi$ as

$$Q_{X_a}(\tau|X_{V\setminus\{a\}},\varpi) = Z^a\beta_u + r_u, \tag{4.22}$$

where $Z^a$ is a $p$-dimensional vector of (known) transformations of $X_{V\setminus\{a\}}$, and $r_u$ is an approximation error. The event $\varpi\in\mathcal{W}$ will be used for further conditioning through the function $K_\varpi(W)=1\{W\in\varpi\}$.

We let $f_{X_a|X_{V\setminus\{a\}},\varpi}(\cdot|X_{V\setminus\{a\}},\varpi)$ denote the conditional density function of $X_a$ given $X_{V\setminus\{a\}}$ and $\varpi\in\mathcal{W}$. We define $f_u := f_{X_a|X_{V\setminus\{a\}},\varpi}(Q_{X_a}(\tau|X_{V\setminus\{a\}},\varpi)|X_{V\setminus\{a\}},\varpi)$ as the value of the conditional density function

evaluated at the $\tau$-conditional quantile. In our analysis we will consider for $u \in \mathcal{U}$

$$\underline{f}_u = \inf_{\|\delta\|=1} \frac{\mathrm{E}[f_u\{Z^a\delta\}^2|\varpi]}{\mathrm{E}[\{Z^a\delta\}^2|\varpi]} \quad \text{and} \quad \underline{f}_{\mathcal{U}} = \min_{u\in\mathcal{U}} \underline{f}_u. \tag{4.23}$$

Moreover, for each $u \in \mathcal{U}$ and $j \in [p]$ we define

$$\gamma_u^j = \arg\min_{\gamma} \mathrm{E}[f_u^2 K_\varpi(W)(Z_j^a - Z_{-j}^a\gamma)^2]. \tag{4.24}$$

This provides a weighted projection to construct the residuals

$$v_{uj} = f_u(Z_j^a - Z_{-j}^a\gamma_u^j)$$

that satisfy $\mathrm{E}[f_u Z_{-j}^a v_{uj}|\varpi] = 0$ for each $(u, j) \in \mathcal{U} \times [p]$.

The estimands of interest are $\beta_u \in \mathbb{R}^p$, $u \in \mathcal{U}$, and can be written as the solution of (a continuum of) moment equations. Letting $\beta_{uj}$ denote the $j$th component of $\beta_u$ so that $\beta_{uj} \in \mathbb{R}$ solves

$$\mathrm{E}[\psi_{uj}(X, W, \beta, \eta_{uj})] = 0,$$

where the function $\psi_{uj}$ is given by

$$\psi_{uj}(X, W, \beta, \eta_{uj}) = K_\varpi(W)(\tau - 1\{X_a \leqslant Z_j^a\beta + Z_{-j}^a\eta_{uj}^{(1)} + \eta_{uj}^{(3)}\})f_u(Z_j^a - Z_{-j}^a\eta_{uj}^{(2)}),$$

and the true value of the nuisance parameter is given by $\eta_{uj} = (\eta_{uj}^{(1)}, \eta_{uj}^{(2)}, \eta_{uj}^{(3)})$ with $\eta_{uj}^{(1)} = \beta_{u,-j}$, $\eta_{uj}^{(2)} = \gamma_u^j$, and $\eta_{uj}^{(3)} = r_u$. In what follows $c, C$ denote some fixed constant, $\delta_n$ and $\Delta_n$ denote sequences go to zero with $\delta_n = n^{-\mu}$ for some sufficiently small $\mu$. Denote $\mu_{\mathcal{W}} = \inf_{\varpi\in\mathcal{W}} \mathrm{P}(\varpi)$.

**Condition CI.** *Let $u = (a, \tau, \varpi) \in \mathcal{U} := V \times \mathcal{T} \times \mathcal{W}$ and $(X_i, W_i)_{i=1}^n$ denote a sequence of independent and identically distributed random vectors generated accordingly to models (4.22) and (4.24):*

*(i) Suppose $\sup_{u\in\mathcal{U},j\in[p]}\{\|\beta_u\| + \|\gamma_u^j\|\} \leqslant C$ and $\mathcal{T}$ is a fixed compact set: (a) there exists $s = s_n$ such that $\sup_{u\in\mathcal{U},j\in[p]}\{\|\beta_u\|_0 + \|\bar\gamma_u^j\|_0\} \leqslant s$, $\sup_{u\in\mathcal{U},j\in[p]} \|\bar\gamma_u^j - \gamma_u^j\| + s^{-1/2}\|\bar\gamma_u^j - \gamma_u^j\|_1 \leqslant C\{n^{-1}s\log(|V|pn)\}^{1/2}$, where $\bar\gamma_u^j$ is approximately sparse; (b) the conditional distribution function of $X_a$ given $X_{V\setminus\{a\}}$ and $\varpi$ is absolutely continuous with continuously differentiable density $f_{X_a|X_{V\setminus\{a\}},\varpi}(t|X_{V\setminus\{a\}}, \varpi)$ bounded by $\bar f$ and its derivative bounded by $\bar f'$ uniformly over $u \in \mathcal{U}$; (c) $|f_u - f_{u'}| \leqslant L_f\|u - u'\|$, $\|\beta_u - \beta_{u'}\| \leqslant L_\beta\|u - u'\|^\kappa$ with $\kappa \in [1/2, 1]$, and $\mathrm{E}[|K_\varpi(W) - K_{\varpi'}(W)|] \leqslant L_K\|\varpi - \varpi'\|$; (d) the VC dimension $d_W$ of the set $\mathcal{W}$ is fixed, $\{Q_{X_a}(\tau|X_{V\setminus\{a\}}, \varpi) : (\tau, \varpi) \in \mathcal{T} \times \mathcal{W}\}$ is a VC-subgraph with VC-dimension $1 + Cd_W$ for every $a \in V$;*

*(ii) The following moment conditions hold uniformly over $u \in \mathcal{U}$ and $j \in [p]$: $\mathrm{E}[|f_u v_{uj} Z_k^a|^2|\varpi]^{1/2} \leqslant C\underline{f}_u$, $\min_{a\in V}\inf_{\|\delta\|=1}\mathrm{E}[\{(X_a, Z^a)\delta\}^2|\varpi] \geqslant c$, $\max_{a\in V}\sup_{\|\delta\|=1}\mathrm{E}[\{(X_a, Z^a)\delta\}^4|\varpi] \leqslant C$, $\mathrm{E}[f_u^2(Z^a\delta)^2|\varpi] \leqslant C\underline{f}_u^2\mathrm{E}[(Z^a\delta)^2|\varpi]$, $\max_{j,k}\frac{\mathrm{E}[|f_u v_{uj} Z_k^a|^3|\varpi]^{1/3}}{\mathrm{E}[|f_u v_{uj} Z_k^a|^2|\varpi]^{1/2}}\log^{1/2}(pn|V|) \leqslant \delta_n\{n\mathrm{P}(\varpi)\}^{1/6}$;*

*(iii) Furthermore, for some fixed $q \geqslant 4\vee(1+2d_W)$, $\sup_{u\in\mathcal{U},\|\delta\|=1}\mathrm{E}[|(X_a, Z^a)\delta|^2 r_u^2|\varpi] \leqslant C\mathrm{E}[r_u^2|\varpi] \leqslant Cs/n$, $\max_{u\in\mathcal{U},j\in[p]}|\mathrm{E}[f_u r_u v_{uj}|\varpi]| \leqslant \delta_n n^{-1/2}$, $\mathrm{E}[\max_{i\leqslant n}\sup_{u\in\mathcal{U}}|K_\varpi(W)r_{iu}|^q] \leqslant C$, and with probability $1 - \Delta_n$, uniformly over $u \in \mathcal{U}, j \in [p]$: $\mathbb{E}_n[r_u^2 v_{uj}^2|\varpi] + \mathbb{E}_n[r_u^2|\varpi] \lesssim n^{-1}s\log(p|V|n)$, $\mathbb{E}_n[K_\varpi(W)\{|r_u| + r_u^2\}(Z^a\delta)^2] \leqslant \delta_n\mathbb{E}_n[K_\varpi(W)f_u(Z^a\delta)^2]$;*

*(iv) For a fixed $q \geqslant 4 \vee (1 + d_W)$, $\mathrm{diam}(\mathcal{W}) \leqslant n^{1/2q}$, $\mathrm{E}[\max_{i\leqslant n}\|X_{iV}\|_\infty^q \vee \max_{a\in V}\|Z_i^a\|_\infty^q]^{1/q}/\mu_{\mathcal{W}} \leqslant M_n$, $\mathrm{E}[\max_{i\leqslant n}\sup_{u\in\mathcal{U},j\in[p]}|v_{iuj}|^q]^{1/q} \leqslant L_n$, $(L_f + L_K)^2 M_n^2\log^2(p|V|n) \leqslant \delta_n n\mu_{\mathcal{W}}^3\underline{f}_{\mathcal{U}}^6$, $M_n^4\log(p|V|n)\log n \leqslant \delta_n^2 n\mu_{\mathcal{W}}^2\underline{f}_{\mathcal{U}}^2$, $s^2\log^2(p|V|n) \leqslant \delta_n^2 n\underline{f}_{\mathcal{U}}^4\mu_{\mathcal{W}}^6$, $s^3\log^3(p|V|n) \leqslant \delta_n^4 n\underline{f}_{\mathcal{U}}^2\mu_{\mathcal{W}}^3$, $L_n^2 s\log^{3/2}(p|V|n) \leqslant \delta_n\underline{f}_{\mathcal{U}}(n\mu_{\mathcal{W}})^{1/2}$, $M_n s\sqrt{\log(p|V|n)} \leqslant \delta_n n^{1/2}\mu_{\mathcal{W}}\underline{f}_{\mathcal{U}}$.*

Condition CI assumes various conditional moment conditions to allow for the estimation to be conditional on $\varpi \in \mathcal{W}$. Those are analogous to the (unconditional) conditions in the high-dimensional literature in quantile regression models, [18]. In particular, condition CI(i) assumes smoothness of the density function, and of coefficients. Condition CI(ii) assumes conditions on the (conditional) population design matrices such as the ratio between eigenvalues. Condition CI(iii) pertains to the approximations errors and assumes mild moment conditions. Finally Condition CI(iv) provides sufficient conditions on the allowed growth of the model via $p$ and $|V|$ relative to the available sample size $n$. Based on Condition CI, we derive our main results regarding the proposed estimator. Moreover, we also establish new results for $\ell_1$-penalized quantile regression methods that hold uniformly over the indices $u \in \mathcal{U}$. The following theorems summarize these results. Note, Condition CI(iii) also assume $d_W$ is bounded by fixed $q$, and the proof can easily be extended to other cases.

**Theorem 1** (Uniform Rates of Convergence for $\mathcal{W}$-Conditional Penalized Quantile Regression)**.** *Under Condition CI, we have that with probability at least* $1 - o(1)$

$$\|\widehat{\beta}_u - \beta_u\| \lesssim \sqrt{\frac{s(1+d_W)\log(p|V|n)}{n\underline{f}_u \mathrm{P}(\varpi)}}, \quad \textit{uniformly over } u = (a,\tau,\varpi) \in \mathcal{U}$$

*Moreover, the thresholded estimator* $\widehat{\beta}^{\bar{\lambda}}$*, with* $\bar{\lambda} = \sqrt{(1+d_W)\log(p|V|n)/n}$ *and* $\widehat{\beta}^{\bar{\lambda}}_{uj} = \widehat{\beta}_{uj} 1\{|\widehat{\beta}_{uj}| > \bar{\lambda}\widehat{\sigma}^Z_{a\varpi j}\}$*, satisfies the same rate and* $\|\widehat{\beta}^{\bar{\lambda}}\|_0 \lesssim s$*.*

Theorem 1 builds upon ideas in [11] however the proof strategy is designed to derive rates that are adaptive to each $u \in \mathcal{U}$. Indeed the rates of convergence are $u$-dependent and they show a slower rate for rare events $\varpi \in \mathcal{W}$.

**Theorem 2** (Uniform Rates of Convergence for $\mathcal{W}$-Conditional Weighted Lasso)**.** *Under Condition CI, we have that with probability at least* $1 - o(1)$

$$\|\widehat{\gamma}^j_u - \gamma^j_u\| \lesssim \frac{1}{\underline{f}_u}\sqrt{\frac{s(1+d_W)\log(p|V|n)}{n\mathrm{P}(\varpi)}} \quad \textit{and} \quad \|\widehat{\gamma}^j_u\|_0 \lesssim s, \quad \textit{uniformly over } u = (a,\tau,\varpi) \in \mathcal{U},\ j \in [p].$$

The following result establishes a uniform Bahadur representation for the final estimators.

**Theorem 3** (Uniform Bahadur Representation for $\mathcal{W}$-Conditional CIQGM)**.** *Under Condition CI, the estimator* $(\check{\beta}_{uj})_{u\in\mathcal{U}, j\in[p]}$ *satisfies*

$$\sigma^{-1}_{uj}\sqrt{n}(\check{\beta}_{uj} - \beta_{uj}) = \mathbb{U}_n(u,j) + O_P(\delta_n) \textit{ in } \ell^\infty(\mathcal{U}\times[p]),$$

*where* $\sigma^2_{uj} = \tau(1-\tau)\mathrm{E}[K_\varpi(W)v^2_{uj}]^{-1}$ *and*

$$\mathbb{U}_n(u,j) := \frac{\{\tau(1-\tau)\mathrm{E}[K_\varpi(W)v^2_{uj}]\}^{-1/2}}{\sqrt{n}}\sum_{i=1}^n(\tau - 1\{U_i(a,\varpi) \leqslant \tau\})K_\varpi(W_i)v_{i,uj},$$

*where* $U_1(a,\varpi),\ldots,U_n(a,\varpi)$ *are i.i.d. uniform* $(0,1)$ *random variables, independent of* $v_{1,uj},\ldots,v_{n,uj}$*.*

Theorem 3 plays a key role. However, it is important to note that the marginal distribution of $\mathbb{U}_n(u,j)$ is pivotal. Nonetheless, there is a non-trivial correlation structure between $U(a,\varpi)$ and $U(\tilde{a},\tilde{\varpi})$. In order to construct confidence regions with non-conservative guarantees, we rely on a multiplier bootstrap method. We will approximate the process $\mathcal{N} = (\mathcal{N}_{uj})_{u\in\mathcal{U}, j\in[p]}$ by the Gaussian multiplier bootstrap

based on estimates $\widehat{\psi}_{uj} := \{\tau(1-\tau)\mathbb{E}_n[K_\varpi(W)\widehat{v}_{uj}^2]\}^{-1/2}(\tau - 1\{X_a \leqslant Z^a\widehat{\beta}_u\})K_\varpi(W_i)\widehat{v}_{uj}$ of $\bar{\psi}_{uj}(U,W) = \{\tau(1-\tau)\mathbb{E}_n[K_\varpi(W)v_{uj}^2]\}^{-1/2}(\tau - 1\{U(a,\varpi) \leqslant \tau\})K_\varpi(W)v_{uj}$, namely

$$\widehat{\mathcal{G}} = (\widehat{\mathcal{G}}_{uj})_{u\in\mathcal{U},j\in[p]} = \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^n g_i \widehat{\psi}_{uj}(X_i, W_i) \right\}_{u\in\mathcal{U},j\in[p]}$$

where $(g_i)_{i=1}^n$ are independent standard normal random variables which are independent from the data $(W_i)_{i=1}^n$. Based on Theorem 5.2 of [24], the following result shows that the multiplier bootstrap provides a valid approximation to the large sample probability law of $\sqrt{n}(\check{\beta}_{uj} - \beta_{uj})_{u\in\mathcal{U},j\in[p]}$ which is suitable for the construction of uniform confidence bands over the set of indices associated with $I_a(b)$ for all $a, b \in V$.

**Corollary 1** (Gaussian Multiplier Bootstrap for $\mathcal{W}$-Conditional CIQGM)**.** *Under Condition CI with* $\delta_n = o(\{(1+d_W)\log(p|V|n)\}^{-1/2})$*, and* $(1+d_W)\log(p|V|n) = o(\{(n/L_n^2)^{1/7} \wedge (n^{1-2/q}/L_n^2)^{1/3}\})$*, we have that*

$$\sup_{P\in\mathcal{P}_n} \sup_{t,t'\in\mathbb{R},u\in\mathcal{U},b\in V} \left| \mathrm{P}_P \left( \max_{j\in I_a(b)} \frac{|\check{\beta}_{uj} - \beta_{uj}|}{n^{-1/2}\sigma_{uj}} \in [t,t'] \right) - \mathrm{P}_P \left( \max_{j\in I_a(b)} |\widehat{\mathcal{G}}_{uj}| \in [t,t'] \mid (X_i, W_i)_{i=1}^n \right) \right| = o(1)$$

Corollary 1 allows the construction of simultaneous confidence regions for the coefficients that are uniformly valid over the set of data generating processes induced by Condition CI. Based on the coefficients whose intervals do not overlap zero, we can construct a conditional independence graph process $\widehat{E}^I(\tau,\varpi), \tau \in \mathcal{T}, \varpi \in \mathcal{W}$ that contains the true conditional independence quantile graph with a specified probability.

4.2. **$\mathcal{W}$-Conditional PQGM.** In this section, we derive theoretical guarantees for the $\mathcal{W}$-conditional predictive quantile estimators uniformly over $u = (a,\tau,\varpi) \in \mathcal{U}$. For each $u \in \mathcal{U}$ the estimand of interest is $\beta_u \in \mathbb{R}^p$ that corresponds to the best linear predictor under asymmetric loss function, namely

$$\beta_u \in \arg\min_\beta \mathrm{E}[\rho_\tau(X_a - X'_{-a}\beta) \mid \varpi] \tag{4.25}$$

where the event $\varpi \in \mathcal{W}$ is used for further conditioning. In the analysis below, the conditioning is implemented through the function $K_\varpi(W) = 1\{W \in \varpi\}$.

In the analysis of this case, the main issue is to handle the inherent misspecification of the linear form $X'_{-a}\beta_u$ with respect to the true conditional quantile. The first consequence is to handle the identification condition. Given $X_{-a}$ and $\varpi \in \mathcal{W}$, we let $f_u := f_{X_a|X_{-a},\varpi}(X'_{-a}\beta_u|X_{-a},\varpi)$ denote the value of the conditional density function evaluated at $X'_{-a}\beta_u$. In our analysis, we will consider

$$\underline{f}_u = \inf_{\|\delta\|=1} \frac{\mathrm{E}[f_u\{X'_{-a}\delta\}^2 \mid \varpi]}{\mathrm{E}[(X'_{-a}\delta)^2 \mid \varpi]} \quad \text{and} \quad \underline{f}_{\mathcal{U}} = \min_{u\in\mathcal{U}} \underline{f}_u. \tag{4.26}$$

We remark that $\underline{f}_u$ defined in (4.26) differs from (4.23) which is the standard conditional density at the true quantile value. It turns out that Knight's identity can be used by exploiting the first order condition associated with the optimization problem (4.25) which yields zero mean condition similar to the conditional quantile condition.

A second consequence of the misspecification is the lack of pivotality of the score. Such pivotal property was convenient in the previous section to define penalty parameters and to conduct inference. We will exploit bounds on the VC-dimension of the relevant classes of sets formally stated below.

**Condition P.** *Let $\mathcal{U} = V \times \mathcal{T} \times \mathcal{W}$ and $(X_i, W_i)_{i=1}^n$ denote a sequence of independent and identically distributed random vectors generated accordingly to models (4.25):*

*(i) Suppose that $\sup_{u\in\mathcal{U}} \|\beta_u\| \leqslant C$ and $\mathcal{T}$ is a fixed compact set: (a) there exists $s = s_n$ and $\bar\beta_u$ such that $\sup_{u\in\mathcal{U}} \|\bar\beta_u\|_0 \leqslant s$, $\sup_{u\in\mathcal{U}} \|\bar\beta_u - \beta_u\| + s^{-1/2}\|\bar\beta_u - \beta_u\|_1 \leqslant \sqrt{s/n}$; (b) the conditional distribution function of $X_a$ given $X_{-a}$ and $\varpi$ is absolutely continuous with continuously differentiable density $f_{X_a|X_{-a},\varpi}(t \mid X_{-a}, \varpi)$ such that its values are bounded by $\bar f$ and its derivative is bounded by $\bar f'$ uniformly over $u \in \mathcal{U}$; (c) $|f_u - f_{u'}| \leqslant L_f\|u - u'\|$, $\|\beta_u - \beta_{u'}\| \leqslant L_\beta\|u - u'\|^\kappa$ with $\kappa \in [1/2, 1]$, and $\mathrm{E}[|K_\varpi(W) - K_{\varpi'}(W)|] \leqslant L_K\|\varpi - \varpi'\|$; (d) the VC dimension $d_W$ of the set $\mathcal{W}$ is fixed, $\{1\{X_a \leqslant X'_{-a}\beta_u\} : (\tau, \varpi) \in \mathcal{T} \times \mathcal{W}\}$ is a VC-class with VC-dimension $1 + d_W$ for every $a \in V$;*

*(ii) The following moment conditions hold uniformly over $u \in \mathcal{U}$: $\min_{a\in V} \inf_{\|\delta\|=1} \mathrm{E}[\{X'_{-a}\delta\}^2|\varpi] \geqslant c$, $\max_{a\in V} \sup_{\|\delta\|=1} \mathrm{E}[\{X'_{-a}\delta\}^4|\varpi] \leqslant C$;*

*((iii) With probability $1 - \Delta_n$, uniformly over $u \in \mathcal{U}$ and $a \in V$: $\mathbb{E}_n[K_\varpi(W)\{|X'_{-a}(\bar\beta_u - \beta_u)| + |X'_{-a}(\bar\beta_u - \beta_u)|^2\}(Z^a\delta)^2] \leqslant \delta_n\mathbb{E}_n[K_\varpi(W)f_u(X'_{-a}\delta)^2]$;*

*(iv) For a fixed $q \geqslant 4 \vee (1 + d_W)$, we have that: $\mathrm{diam}(\mathcal{W}) \leqslant n^{1/2q}$, $\mathrm{E}[\max_{i\leqslant n} \|X_{iV}\|_\infty^q]^{1/q}/\mu_\mathcal{W} \leqslant M_n$, $M_n^2 \log^7(n|V|) \leqslant \delta_n n\underline{f}_\mathcal{U}^2\mu_\mathcal{W}^2$, $M_n^4 \log(n|V|)\log n \leqslant \delta_n n\mu_\mathcal{W}$, $(L_f + L_K)^2 M_n^2 \log^2(|V|n) \leqslant \delta_n n\mu_\mathcal{W}^3\underline{f}_\mathcal{U}^6$, $M_n^2 s\log^{3/2}(n|V|) \leqslant \delta_n\underline{f}_\mathcal{U}(n\mu_\mathcal{W})^{1/2}$, $M_n s\sqrt{\log(n|V|)} \leqslant \delta_n(n\mu_\mathcal{W})^{1/2}$, and $s^3\log^5(n|V|) \leqslant \delta_n n\underline{f}_\mathcal{U}^2\mu_\mathcal{W}^2$.*

Condition P is a high-level condition. It allows to cover conditioning events $\varpi \in \mathcal{W}$ whose probability can decrease to zero (although slower than $n^{-1/4}$).

Next we derive our main results regarding the proposed estimator for the best linear predictor. These results are also new $\ell_1$-penalized quantile regression methods as it holds under possible misspecification of the conditional quantile function and hold uniformly over the indices $u \in \mathcal{U}$. The following theorem summarizes the result.

**Theorem 4** (Uniform Rates of Convergence for $\mathcal{W}$-Conditional Penalized Quantile Regression under Misspecification)**.** *Under Condition P, we have that with probability at least $1 - o(1)$, uniformly over $u = (a, \tau, \varpi) \in \mathcal{U}$,*

$$\|\widehat\beta_u - \beta_u\| \lesssim \sqrt{\frac{s(1 + d_W)\log(|V|n)}{n\underline{f}_u\mathrm{P}(\varpi)}}.$$

The data-driven choice of penalty parameter helps diminish the regularization bias and also allow to obtain sparse estimators with provably rates of convergence (through thresholding). Moreover, the $u$ specific penalty parameter combined with the new analysis yields an adaptive rate of convergence to each $u \in \mathcal{U}$ unlike previous works.

**Comment 4.1** (Simultaneous Confidence Bands for Coefficients in PQGMs)**.** We note that in some applications we might be interested in constructing (simultaneous) confidence bands for the coefficients in PQGMs. In particular, this would include the cases practitioners are using a misspecified linear specification in a quantile regression model. Provided the conditional density function at $X'_{-a}\beta_u$ can be estimated, a version of Algorithm 3.3 using the penalty parameters in Algorithm 3.4 for the initial step can deliver such confidence regions via a multiplier bootstrap.

## 5. Empirical Applications of QGMs

5.1. **CoVaR, Network Spillover Effects, and Systemic Risk.** Traditional risk measures, such as Value of Risk (VaR), focus on the loss of an individual institution only. CoVaR proposed by [4] is to measure the VaR of the whole financial system or a particular financial institution by conditioning on another institution being in distress.

[4] define institution $b$'s CoVaR at level $\tau$ conditional on a particular outcome of institution $a$, as the value of $CoVaR_\tau^{b|a}$ that solves

$$\mathrm{P}(X_b \leqslant CoVaR_\tau^{b|a} | \mathbb{C}(X_a)) = \tau. \tag{5.27}$$

A special case is $\mathbb{C}(X_a) = \{X_a = VaR_\tau^a\}$ which, as interpreted by [4], means with probability $\tau$ institution $b$ is in trouble given that institution $a$ is in trouble. The estimation procedure is defined as: firstly, quantile regression $X_b$ on $X_a$ gives the value at risk of institution $b$ conditional on institution $a$,

$$VaR_\tau^b | X_a = \beta_0^b(\tau) + \beta_a^b(\tau) X_a, \tag{5.28}$$

then replacing variable $X_a$ by its $\tau$-th quantile, i.e. $VaR_\tau^a$, yields

$$CoVaR_\tau^{b|X_a = VaR_\tau^a} = \beta_0^b(\tau) + \beta_a^b(\tau) VaR_\tau^a, \tag{5.29}$$

and

$$\Delta CoVaR_\tau^{b|a} = \beta_a^b(\tau)(VaR_\tau^a - VaR_{50\%}^a). \tag{5.30}$$

Below, we show with QGM we can incorporate (tail) network spillover effects into risk measuring. (Note the identified risk connections between all financial institutions constitute a systemic risk network.) Define

$$\mathrm{P}(X_b \leqslant CoVaR_\tau^{b|a,V\setminus\{a,b\}} | \mathbb{C}(X_a, X_{V\setminus\{a,b\}})) = \tau, \tag{5.31}$$

we then have

$$CoVaR_\tau^{b|X_a = VaR_\tau^a, X_{V\setminus\{a,b\}} = VaR_\tau^{V\setminus\{a,b\}}} = \check{\beta}_0^b(\tau) + \check{\beta}_a^b(\tau) VaR_\tau^a + \check{\beta}_{V\setminus\{a,b\}}^b(\tau) VaR_\tau^{V\setminus\{a,b\}}, \tag{5.32}$$

and

$$\Delta CoVaR_\tau^{b|a,V\setminus\{a,b\}} = \check{\beta}_a^b(\tau)(VaR_\tau^a - VaR_{50\%}^a), \tag{5.33}$$

where $\check{\beta}^b(\tau) = \{\check{\beta}_0^b(\tau), \check{\beta}_{V\setminus\{b\}}^b(\tau)\}$ is estimated via Algorithm 3.1 or 3.2.

We stack $\Delta CoVaR_\tau^{b|a,V\setminus\{a,b\}}$ as the $(a,b)$-th element of an $d \times d$ matrix $E^\beta(\tau)$ which represents a weighted and directed network of institutions. Following [5], the systemic risk contribution of institution $a$ is called the network to-degree, defined as $\delta_a^{to} = \Delta CoVaR^{sys|a} = \sum_k \Delta CoVaR^{k|a,V\setminus\{a,k\}}$. To-degrees measure contributions of individual institutions to the overall risk of systemic network events. Similarly, from-degree of institution $a$ is defined as $\delta_a^{from} = \Delta CoVaR^{a|sys} = \sum_k \Delta CoVaR_\tau^{a|k,V\setminus\{a,k\}}$. From-degrees measure exposure of individual institutions to systemic shocks from the network. The total degree $\delta := \sum_a \Delta CoVaR^{sys|a}$, aggregates institution-specific systemic risk across institutions hence provides a measure of total systemic risk in the whole financial system. Finally, for institution $a$, we define its net contribution as $net\text{-}\Delta CoVaR^a = \delta_a^{to} - \delta_a^{from}$. For more about network analysis, see [52].

5.2. **International Financial Contagion.** In this section we apply PQGMs to the study of international financial contagion and then with the estimated network structure we can measure the systemic risk contributions of each country, as mentioned in Section 5.1. We focus on examining financial contagion through the volatility spillovers perspective. [39] reported that international stock markets are related through their volatilities instead of returns. [34] studied the return and volatility spillovers of 19 countries and found differences in return and volatility spillovers. For a survey of financial contagion see [30]. We also illustrate how PQGMs can highlight asymmetric dependence between the random variables.

We use daily equity index returns, September 2009 to September 2013 (1044 observations), from Morgan Stanley Capital International (MSCI). The returns are all translated into dollar-equivalents as of September 6th 2013. We use absolute returns as a proxy for volatility. We have a total of 45 countries in our sample, there are 21 developed markets (Australia, Austria, Belgium, Canada, Denmark, France, Germany, Hong Kong, Ireland, Italy, Japan, Netherlands, New Zealand, Norway, Portugal, Singapore, Spain, Sweden, Switzerland, the United Kingdom, the United States), 21 emerging markets (Brazil, Chile, Mexico, Greece, Israel, China, Colombia, Czech Republic, Egypt, Hungary, India, Indonesia, Korea, Malaysia, Peru, Philippines, Poland, Russia, Taiwan, Thailand, Turkey), and 3 frontier markets (Argentina, Morocco, Jordan).

Below in Figure 1 we provide a full-sample analysis of global volatility spillovers at different tails. We denote 20% quantile as Low Tail, 50% quantile as Median, 80% quantile as Up Tail. Both PQGMs and GGM are presented. Our purpose is to show the usefulness of PQGM in representing nonlinear tail interdependence allowing for heteroscedasticity and to show that PQGM can measure correlation asymmetry through looking at the tails of the distribution (not specific to any model).

There are significant differences in the network structure in terms of volatility spillovers when using PQGM and GGM. PQGM permits asymmetries in correlation dynamics, suited to investigate the presence of asymmetric responses. We find significant increase interdependence at the up tail between the volatility series, i.e. we find downside correlations (high volatility) are much larger than upside correlations (low volatility). This confirms findings in the finance literature that financial markets become more interdependent during high volatility periods.

We also find if two countries locate in the same geographic region, with many similarities in terms of market structure and history, they tend to be more closely connected (homophily effect as stated in network terminology); while two economies locate in separate geographic regions are less likely directly connected. In addition, we find among European Union member countries, Germany appears to play a major role in the transmission of shocks to others; while in Asia, Hong Kong, Thailand, and Singapore appear to play major roles; and among all the north and south American countries, Canada and US play major roles.

In addition, we present *net*-$\Delta CoVaR$ discussed in Section 5.1 with $\tau = 0.8$, i.e. the Up Tail, in Figure **??** which shows that: globally, total volatility spillovers from Germany and France to the others are much larger than total volatility spillovers from the others to them, and their *net*-$\Delta CoVaR$ are positive. Both Greece and Spain have negative *net*-$\Delta CoVaR$. The estimated network structure is important here as it demonstrates that shocks originated in some markets may be amplified during their transmission throughout the system, posing greater risks to the whole market than other shocks' origination.

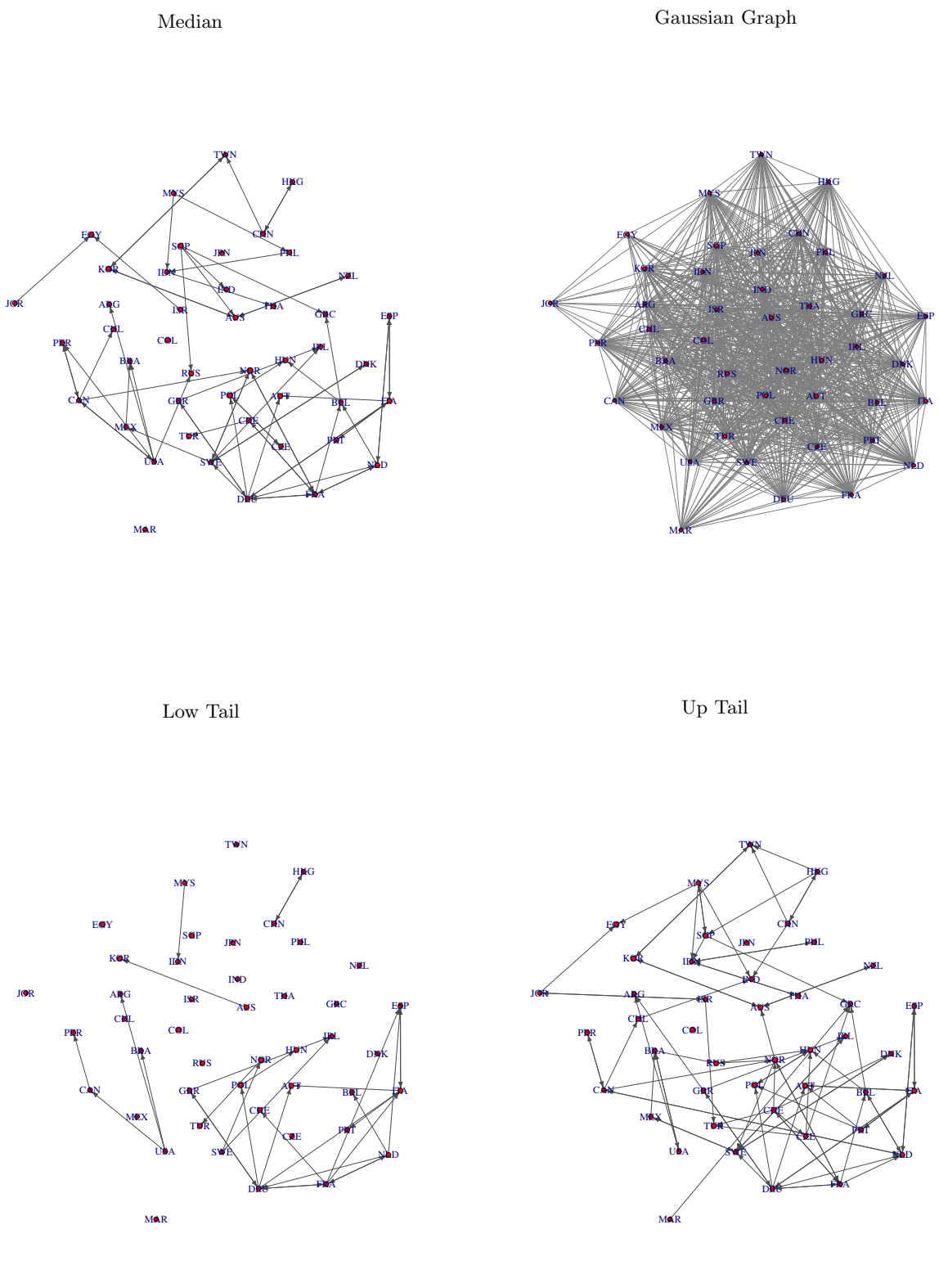



FIGURE 1. International Financial Contagion. Note: we show the volatility transmission channel is asymmetric (at different tails).

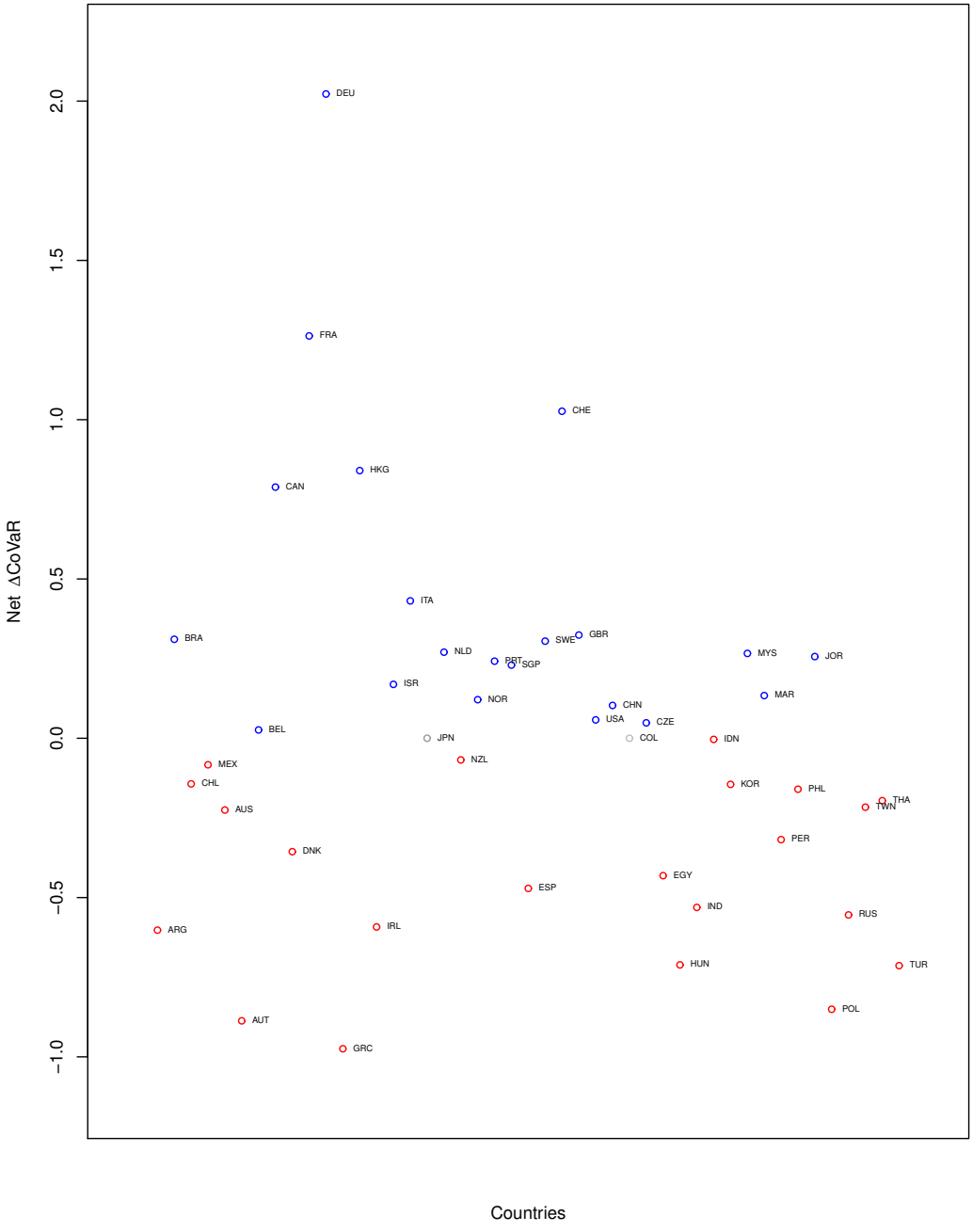



FIGURE 2. Volatility Spillovers Net Contribution of Each Country

5.3. **Stock Returns Under Market Downside Movements.** Stock returns are in general non-Gaussian. [6] find correlation asymmetries in the data and reject the null hypothesis of multivariate Gaussian distributions at daily, weekly, and monthly frequencies, conditional on market "downside" movements. See also [61, 67], among other studies, in the empirical finance literature for evidence of the non-Gaussian feature of financial markets. Hence, generally, in the financial market context, linear correlation measures only convey partial and often misleading information on the actual underlying conditional dependencies.

In fact, when we are particularly interested in the conditional dependencies of stocks returns conditioning on market "downside" movements, this can be modeled by the $\mathcal{W}$-Conditional CIQGM with $\mathcal{W} = \{\text{Market return} \leqslant \varpi\}$, $\varpi$ as the $\tau_m$-th quantile of the market index returns, and $\tau_m = \{0.15, 0.5, 0.75, 0.9, 1\}$. In the following example, we obtain daily stock returns from CRSP and use S&P 500 as the market index. The full sample consists of 2769 observations for 86 stocks from Jan 2, 2003 to December 31, 2013. The total number of stocks is 86 due to data availability at CRSP. In this case, we define market downside movement as when the market index returns are below a pre-specified level (e.g. $\tau_m$-th quantile), hence conditioning on a particular $\varpi$ simply corresponds to consider the subsample based on whether the corresponding date's market return is less equal to the $\tau_m$-th quantile of the market index returns. We reported the number of edges (there is no linkage between two stocks if they are conditional independent), under different market conditions in Table 1 below.

TABLE 1. Edges Produced by Different Graph Estimators

| | | obs | PQGM | CIQGM | CIQGM(0.1) | CIQGM(0.3) |
|---|---|---|---|---|---|---|
| $\tau_m$ | 0.15 | 416 | 82 | 7302 | 4086 | 388 |
| | 0.5 | 1385 | 196 | 7254 | 1308 | 74 |
| | 0.75 | 2077 | 238 | 7226 | 908 | 50 |
| | 0.9 | 2492 | 272 | 7202 | 740 | 46 |
| | 1 | 2769 | 304 | 7274 | 784 | 54 |
| $\varpi$ | $> 0$ | 1246 | 186 | 7262 | 1516 | 80 |
| | $< 0$ | 1522 | 180 | 7254 | 1308 | 64 |

Note: The results are computed by Algorithm 3.3 and 3.4. CIQGM(0.1) means additional thresholding at 0.1, i.e. we keep the edges that have correlation stronger than 0.1. Similarly, CIQGM(0.3) means additional thresholding at 0.3.

As shown in Table 1, there are significant differences between PQGM and CIQGMs, hence assuming Gaussianity for the distribution of stock returns could result in false correlation conclusions (e.g., estimation bias due to asymmetry in correlations conditional on market upside or downside moves). When using PQGM, the number of edges increases with $\tau_m$. The number of edges in CIQGM is significantly higher when no thresholding is applied. The results also show stronger correlation (and more connections) under market downside moments. All those empirical findings support evidence from the empirical finance literature.

## References


[1] Alberto Abadie, Guido Imbens, and Fanyin Zheng. Inference for misspecified models with fixed regressors. *Journal of the American Statistical Association*, 109(508):1601–1614, 2014.

[2] Daron Acemoglu, Asuman Ozdaglar, and Alireza Tahbaz-Salehi. Cascades in networks and aggregate volatility. Technical report, National Bureau of Economic Research, 2010.

[3] Daron Acemoglu, Asuman Ozdaglar, and Alireza Tahbaz-Salehi. Systemic risk and stability in financial networks. *The American Economic Review*, 105(2):564–608, 2015.

[4] Tobias Adrian and Markus Brunnermeier. Covar. *The American Economic Review*, 106(7):1705–1741, 2016.

[5] Torben Andersen, Tim Bollerslev, Peter Christoffersen, and Francis Diebold. Financial risk measurement for financial risk management. *Handbook of the Economics of Finance*, 2:1127–1220, 2013.

[6] Andrew Ang and Joseph Chen. Asymmetric correlations of equity portfolios. *Journal of Financial Economics*, 63(3):443–494, 2002.

[7] Andrew Ang, Joseph Chen, and Yuhang Xing. Downside risk. *Review of Financial Studies*, 19(4):1191–1239, 2006.

[8] Joshua Angrist, Victor Chernozhukov, and Iván Fernández-Val. Quantile regression under misspecification, with an application to the us wage structure. *Econometrica*, 74(2):539–563, 2006.

[9] Onureena Banerjee, Laurent El Ghaoui, and Alexandre d'Aspremont. Model selection through sparse maximum likelihood estimation for multivariate gaussian or binary data. *The Journal of Machine Learning Research*, 9:485–516, 2008.

[10] Alexandre Belloni, Daniel Chen, Victor Chernozhukov, and Christian Hansen. Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica*, 80(6):2369–2429, 2012.

[11] Alexandre Belloni and Victor Chernozhukov. $\ell_1$-penalized quantile regression for high dimensional sparse models. *Annals of Statistics*, 39(1):82–130, 2011.

[12] Alexandre Belloni and Victor Chernozhukov. Inference methods for high-dimensional sparse econometric models. *Advances in Economics and Econometrics, 10th World Congress of Econometric Society*, III:245–295, 2013.

[13] Alexandre Belloni and Victor Chernozhukov. Least squares after model selection in high-dimensional sparse models. *Bernoulli*, 19(2):521–547, 2013. ArXiv, 2009.

[14] Alexandre Belloni, Victor Chernozhukov, Denis Chetverikov, and Ying Wei. Uniformly valid post-regularization confidence regions for many functional parameters in z-estimation framework. *arXiv:1512.07619*, 2015.

[15] Alexandre Belloni, Victor Chernozhukov, and Iván Fernández-Val. Conditional quantile processes based on series or many regressors. *arXiv*, 2011.

[16] Alexandre Belloni, Victor Chernozhukov, Ivan Fernández-Val, and Christian Hansen. Program evaluation and causal inference with high-dimensional data. *Econometrica*, 85(1):233–298, 2017.

[17] Alexandre Belloni, Victor Chernozhukov, and Christian Hansen. Inference on treatment effects after selection among high-dimensional controls. *The Review of Economic Studies*, 81(2):608–650, 2014.

[18] Alexandre Belloni, Victor Chernozhukov, and Kengo Kato. Robust inference in high-dimensional approximately sparse quantile regression models. *arXiv preprint arXiv:1312.7186*, 2013.

[19] Alexandre Belloni, Victor Chernozhukov, and Kengo Kato. Uniform post-selection inference for least absolute deviation regression and other z-estimation problems. *Biometrika*, 102(1):77–94, 2015.

[20] Alexandre Belloni, Victor Chernozhukov, and Lie Wang. Square-root-lasso: Pivotal recovery of sparse signals via conic programming. *Biometrika*, 98(4):791–806, 2011. Arxiv, 2010.

[21] Peter Bickel, Yaacov Ritov, and Alexandre Tsybakov. Simultaneous analysis of Lasso and Dantzig selector. *Annals of Statistics*, 37(4):1705–1732, 2009.

[22] Tony Cai, Weidong Liu, and Xi Luo. A constrained $\ell_1$ minimization approach to sparse precision matrix estimation. *Journal of the American Statistical Association*, 106(494), 2011.

[23] Mehmet Caner and Anders Kock. Asymptotically honest confidence regions for high dimensional parameters by the desparsified conservative lasso. *Journal of Econometrics*, forthcoming.

[24] Victor Chernozhukov, Denis Chetverikov, and Kengo Kato. Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *The Annals of Statistics*, 41(6):2786–2819, 2013.

[25] Victor Chernozhukov, Denis Chetverikov, and Kengo Kato. Gaussian approximation of suprema of empirical processes. *The Annals of Statistics*, 42(4):1564–1597, 2014.

[26] Victor Chernozhukov, Denis Chetverikov, and Kengo Kato. Comparison and anti-concentration bounds for maxima of gaussian random vectors. *Probability Theory and Related Fields*, 162:47–70, 2015.

[27] Victor Chernozhukov, Denis Chetverikov, and Kengo Kato. Empirical and multiplier bootstraps for supreme of empirical processes of increasing complexity, and related gaussian couplings. *Stochastic Processes and their Applications*, 126(12):3632–3651, 2016.

[28] Victor Chernozhukov, Christian Hansen, and Martin Spindler. Post-selection and post-regularization inference in linear models with very many controls and instruments. *American Economic Review: Papers and Proceedings*, 105(5):486–490, 2015.

[29] Khai Chiong and Roger Moon. Estimation of graphical lasso using the l1, 2 norm. *The Econometrics Journal*, forthcoming.

[30] Stijn Claessens and Kristin Forbes. *International Financial Contagion*. Springer, 2001.

[31] David Cox and Nanny Wermuth. *Multivariate dependencies: Models, analysis and interpretation*, volume 67. CRC Press, 1996.

[32] Philip Dawid. Conditional independence in statistical theory. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 1–31, 1979.

[33] Arthur Dempster. Covariance selection. *Biometrics*, pages 157–175, 1972.

[34] Francis Diebold and Kamil Yilmaz. Measuring financial asset return and volatility spillovers, with application to global equity markets. *The Economic Journal*, 119(534):158–171, 2009.

[35] Mathias Drton and Michael Perlman. Model selection for gaussian concentration graphs. *Biometrika*, 91(3):591–602, 2004.

[36] Mathias Drton and Michael Perlman. Multiple testing and error control in gaussian graphical model selection. *Statistical Science*, 22(3):430–449, 2007.

[37] Mathias Drton and Michael Perlman. A sinful approach to gaussian graphical model selection. *Journal of Statistical Planning and Inference*, 138(4):1179–1200, 2008.

[38] David Edwards. *Introduction to graphical modelling*. Springer, 2000.

[39] Robert Engle and Raul Susmel. Common volatility in international equity markets. *Journal of Business & Economic Statistics*, 11(2):167–176, 1993.

[40] Jianqing Fan, Jinchi Lv, and Lei Qi. Sparse high dimensional models in economics. *Annual review of economics*, 3:291, 2011.

[41] Max Farrell. Robust inference on average treatment effects with possibly more covariates than observations. *Journal of Econometrics*, 174(2):1–23, 2015.

[42] Rina Foygel and Mathias Drton. Extended bayesian information criteria for gaussian graphical models. In *Advances in Neural Information Processing Systems*, pages 604–612, 2010.

[43] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441, 2008.

[44] Xuming He, Lan Wang, and Hyokyoung Grace Hong. Quantile-adaptive model-free variable screening for high-dimensional heterogeneous data. *The Annals of Statistics*, 41(1):342–369, 2013.

[45] Jana Jankova and Sara van de Geer. Confidence intervals for high-dimensional inverse covariance estimation. *Electronic Journal of Statistics*, 9(1):1205–1229, 2015.

[46] Harry Joe. *Multivariate models and multivariate dependence concepts*. CRC Press, 1997.

[47] Kengo Kato. Group lasso for high dimensional sparse quantile regression models. Preprint, ArXiv, 2011.

[48] Keith Knight. Limiting distributions for $L_1$ regression estimators under general conditions. *The Annals of Statistics*, 26:755–770, 1998.

[49] Keith Knight. Asymptotics of the regression quantile basic solution under misspecification. *Applications of Mathematics*, 53(3):223–234, 2008.

[50] Roger Koenker. *Quantile regression*. Cambridge University Press, New York, 2005.

[51] Roger Koenker and Gilbert Bassett. Regression quantiles. *Econometrica*, 46(1):33–50, 1978.

[52] Eric Kolaczyk. *Statistical analysis of network data*. Springer, 2009.

[53] Steffen Lauritzen. *Graphical models*. Oxford University Press, 1996.

[54] Michel Ledoux and Michel Talagrand. *Probability in Banach Spaces (Isoperimetry and processes)*. Ergebnisse der Mathematik undihrer Grenzgebiete, Springer-Verlag, 1991.

[55] Ying-Ying Lee. Efficiency bounds for semiparametric estimation of quantile regression under misspecification. *University of Wisconsin-Madison*, 2009.

[56] Hannes Leeb and Benedikt Pötscher. Can one estimate the unconditional distribution of post-model-selection estimators? *Econometric Theory*, 24(02):338–376, 2008.

[57] Han Liu, Fang Han, Ming Yuan, John Lafferty, and Larry Wasserman. High-dimensional semiparametric gaussian copula graphical models. *The Annals of Statistics*, 40(4):2293–2326, 2012.

[58] Han Liu, John Lafferty, and Larry Wasserman. The nonparanormal: Semiparametric estimation of high dimensional undirected graphs. *The Journal of Machine Learning Research*, 10:2295–2328, 2009.

[59] Han Liu and Lie Wang. Tiger: A tuning-insensitive approach for optimally estimating gaussian graphical models. *Electronic Journal of Statistics*, 11(1):241–294, 2017.

[60] Weidong Liu and Xi Luo. High-dimensional sparse precision matrix estimation via sparse column inverse operator. *arXiv preprint arXiv:1203.3896*, 2012.

[61] Francois Longin and Bruno Solnik. Extreme correlation of international equity markets. *The Journal of Finance*, 56(2):649–676, 2001.

[62] Nicolai Meinshausen and Peter Bühlmann. High-dimensional graphs and variable selection with the lasso. *The Annals of Statistics*, pages 1436–1462, 2006.

[63] Sahand Negahban, Pradeep Ravikumar, Martin Wainwright, and Bin Yu. A unified framework for high-dimensional analysis of m-estimators with decomposable regularizers. *Statistical Science*, 27(4):538–557, 2012.

[64] Yang Ning and Han Liu. A general theory of hypothesis tests and confidence regions for sparse high dimensional models. *The Annals of Statistics*, 45(1):158–195, 2017.

[65] Roberto Imbuzeiro Oliveira. The lower tail of random quadratic forms, with applications to ordinary least squares and restricted eigenvalue properties. *Probability Theory and Related Fields*, 166(3-4):1175–1194, 2016.

[66] Andrew Patton. On the out-of-sample importance of skewness and asymmetric dependence for asset allocation. *Journal of Financial Econometrics*, 2(1):130–168, 2004.

[67] Andrew Patton. Modelling asymmetric exchange rate dependence. *International economic review*, 47(2):527–556, 2006.

[68] Joseph Romano and Michael Wolf. Exact and approximate stepdown methods for multiple hypothesis testing. *Journal of the American Statistical Association*, 100(469):94–108, 2005.

[69] Tingni Sun and Cun-Hui Zhang. Sparse matrix inversion with scaled lasso. *The Journal of Machine Learning Research*, 14(1):3385–3418, 2013.

[70] Sara van de Geer. High-dimensional generalized linear models and the lasso. *Annals of Statistics*, 36(2):614–645, 2008.

[71] Sara Van de Geer, Peter Bühlmann, Yaacov Ritov, and Ruben Dezeure. On asymptotically optimal confidence regions and tests for high-dimensional models. *The Annals of Statistics*, 42(3):1166–1202, 2014.

[72] Aad van der Vaart and Jon Wellner. *Weak Convergence and Empirical Processes*. Springer Series in Statistics, 1996.

[73] Aad van der Vaart and Jon Wellner. Empirical process indexed by estimated functions. *IMS Lecture Notes-Monograph Series*, 55:234–252, 2007.

[74] Jialei Wang and Mladen Kolar. Inference for high-dimensional exponential family graphical models. In *Proc. of AISTATS*, volume 51, pages 751–760, 2016.

[75] Yuanshan Wu and Guosheng Yin. Conditional quantile screening in ultrahigh-dimensional heterogeneous data. *Biometrika*, 102(1):65–76, 2015.

[76] Lingzhou Xue and Hui Zou. Regularized rank-based estimation of high-dimensional nonparanormal graphical models. *The Annals of Statistics*, 40(5):2541–2571, 2012.

[77] Lingzhou Xue, Hui Zou, and Tianxi Cai. Nonconcave penalized composite conditional likelihood estimation of sparse ising models. *The Annals of Statistics*, 40(3):1403–1429, 2012.

[78] Ming Yuan. High dimensional inverse covariance matrix estimation via linear programming. *The Journal of Machine Learning Research*, 99:2261–2286, 2010.

[79] Ming Yuan and Yi Lin. Model selection and estimation in the gaussian graphical model. *Biometrika*, 94(1):19–35, 2007.

[80] Cun-Hui Zhang and Stephanie Zhang. Confidence intervals for low dimensional parameters in high dimensional linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(1):217–242, 2014.

[81] Qi Zheng, Limin Peng, and Xuming He. Globally adaptive quantile regression with ultra-high dimensional data. *Annals of statistics*, 43(5):2225, 2015.

## Appendix A. Implementation Details of Algorithms

This section provides details of the algorithms mentioned in Section 3. Note, for the weighted-Lasso estimator, the choice of penalty level $\lambda := 1.1n^{-1/2}2\Phi^{-1}(1-\xi/N_n)$ and penalty loading $\widehat{\Gamma}_\tau = \text{diag}[\widehat{\Gamma}_{\tau kk}, k \in [p]\backslash\{j\}]$ is a diagonal matrix defined by the following procedure: (1) Compute the post-Lasso estimator $\widetilde{\gamma}^j_{a\tau}$ based on $\lambda$ and initial values $\widehat{\Gamma}_{\tau kk} = \max_{i\leqslant n}\|f_{ia\tau}Z^a_i\|_\infty\{\mathbb{E}_n[|f_{a\tau}Z^a_k|^2]\}^{1/2}$. (2) Compute the residuals $\widehat{v}_{ia\tau j} = f_{ia\tau}(Z^a_{ij} - Z^a_{i,-j}\widetilde{\gamma}^j_{a\tau})$ and update the loadings

$$\widehat{\Gamma}_{\tau kk} = \sqrt{\mathbb{E}_n[f^2_{a\tau}|Z^a_k\widehat{v}_{a\tau j}|^2]},\ k \in [p]\backslash\{j\} \tag{A.34}$$

and use them to recompute the post-Lasso estimator $\widetilde{\gamma}^j_{a\tau}$. In the case of Algorithm 3.1 we can take $N_n = |V|p^3n^3$, in the case of Algorithm 3.3 we take $N_n = |V|p^2\{pn^3\}^{1+d_W}$. Denote $\widehat{\sigma}^Z_{aj} = \{\mathbb{E}_n[(Z^a_j)^2]\}^{1/2}$, $\widehat{\sigma}^X_{aj} = \{\mathbb{E}_n[X^2_{-a,j}]\}^{1/2}$, $\widehat{\sigma}^Z_{a\varpi j} = \{\mathbb{E}_n[K_\varpi(W)(Z^a_j)^2]\}^{1/2}$, and $\widehat{\sigma}^X_{a\varpi j} = \{\mathbb{E}_n[K_\varpi(W)X^2_{-a,j}]\}^{1/2}$.

**Detailed version of Algorithm 3.1 (CIQGM)**

For each $a \in V$, $\tau \in \mathcal{T}$, and $j \in [p]$, perform the following:

(1) Run Post-$\ell_1$-quantile regression of $X_a$ on $Z^a$; keep fitted value $Z^a_{-j}\widetilde{\beta}_{a\tau,-j}$,

$$\widehat{\beta}_{a\tau} \in \arg\min_\beta \mathbb{E}_n[\rho_\tau(X_a - Z^a\beta)] + \lambda_{V\mathcal{T}}\sqrt{\tau(1-\tau)}\textstyle\sum_{j=1}^p \widehat{\sigma}^Z_{aj}|\beta_j|$$
$$\widetilde{\beta}_{a\tau} \in \arg\min_\beta \mathbb{E}_n[\rho_\tau(X_a - Z^a\beta)]\ :\ \beta_j = 0 \ \text{ if } |\widehat{\beta}_{a\tau j}| \leqslant \lambda_{V\mathcal{T}}\sqrt{\tau(1-\tau)}/\widehat{\sigma}^Z_{aj}.$$

(2) Run Post-Lasso of $f_{a\tau}Z^a_j$ on $f_{a\tau}Z^a_{-j}$; keep the residual $\widetilde{v}_i := f_{ia\tau}\{Z^a_{ij} - Z^a_{i,-j}\widetilde{\gamma}^j_{a\tau}\}$,

$$\widehat{\gamma}^j_{a\tau} \in \arg\min_\gamma \mathbb{E}_n[f^2_{a\tau}(Z^a_j - Z^a_{-j}\gamma)^2] + \lambda\|\widehat{\Gamma}_\tau\gamma\|_1$$
$$\widetilde{\gamma}^j_{a\tau} \in \arg\min_\gamma \mathbb{E}_n[f^2_{a\tau}(Z^a_j - Z^a_{-j}\gamma)^2]\ :\ \text{support}(\gamma) \subseteq \text{support}(\widehat{\gamma}^j_{a\tau}).$$

(3) Run Instrumental Quantile Regression of $X_a - Z^a_{-j}\widetilde{\beta}_{a\tau,-j}$ on $Z^a_j$ using $\widetilde{v}$ as the instrument for $Z^a_j$,

$$\check{\beta}_{a\tau,j} \in \arg\min_{\alpha\in\mathcal{A}_{a\tau j}} \frac{\{\mathbb{E}_n[(1\{X_a \leqslant Z^a_j\alpha + Z^a_{-j}\widetilde{\beta}_{a\tau,-j}\} - \tau)\widetilde{v}]\}^2}{\mathbb{E}_n[(1\{X_a \leqslant Z^a_j\alpha + Z^a_{-j}\widetilde{\beta}_{a\tau,-j}\} - \tau)^2\widetilde{v}^2]},$$

with $\mathcal{A}_{a\tau j} = \{\alpha \in \mathbb{R} : |\alpha - \widetilde{\beta}_{a\tau j}| \leqslant 10/\{\widehat{\sigma}^Z_{aj}\log n\}\}$.

**Detailed version of Algorithm 3.2 (PQGM)**

For each $a \in V$, and $\tau \in \mathcal{T}$, perform the following:

(1) Run Post-$\ell_1$-quantile regression of $X_a$ on $X_{-a}$,

$$\widehat{\beta}_{a\tau} \in \arg\min_\beta \mathbb{E}_n[\rho_\tau(X_a - X'_{-a}\beta)] + \lambda_0\textstyle\sum_{j\in[d]}\widehat{\sigma}^X_{aj}|\beta_j|$$
$$\widetilde{\beta}_{a\tau} \in \arg\min_\beta \mathbb{E}_n[\rho_\tau(X_a - X'_{-a}\beta)]\ :\ \beta_j = 0 \ \text{ if } |\widehat{\beta}_{a\tau j}| \leqslant \lambda_0/\widehat{\sigma}^X_{aj}.$$

(2) Set $\widehat{\varepsilon}_{ia\tau} = 1\{X_{ia} \leqslant X'_{i,-a}\tilde{\beta}_{a\tau}\} - \tau$ for $i \in [n]$. Compute the penalty level $\bar{\lambda}_{V\mathcal{T}}$ via (3.19).

(3) Run Post-$\ell_1$-quantile regression of $X_a$ on $X_{-a}$,

$$\widehat{\beta}_{a\tau} \in \arg\min_\beta \mathbb{E}_n[\rho_\tau(X_a - X'_{-a}\beta)] + \bar{\lambda}_{V\mathcal{T}}\textstyle\sum_{j\in[d]}\{\mathbb{E}_n[\widehat{\varepsilon}^2_{a\tau}X^2_{-a,j}]\}^{1/2}|\beta_j|$$
$$\check{\beta}_{a\tau} \in \arg\min_\beta \mathbb{E}_n[\rho_\tau(X_a - X'_{-a}\beta)]\ :\ \beta_j = 0 \ \text{ if } |\widehat{\beta}_{a\tau j}| \leqslant \bar{\lambda}_{V\mathcal{T}}/\{\mathbb{E}_n[\widehat{\varepsilon}^2_{a\tau}X^2_{-a,j}]\}^{1/2}.$$

**Detailed version of Algorithm 3.3 ($\mathcal{W}$-Conditional CIQGM)**

For each $u = (a, \tau, \varpi) \in \mathcal{U} = V \times \mathcal{T} \times \mathcal{W}$, and $j \in [p]$, perform the following:

(1) Run Post-$\ell_1$-quantile regression of $X_a$ on $Z^a$; keep fitted value $Z^a_{-j}\widetilde{\beta}_{u,-j}$,

$$\widehat{\beta}_u \in \arg\min_\beta \mathbb{E}_n[K_\varpi(W)\rho_\tau(X_a - Z^a\beta)] + \lambda_u \textstyle\sum_{j=1}^p \widehat{\sigma}^Z_{a\varpi j}|\beta_j|$$
$$\widetilde{\beta}_u \in \arg\min_\beta \mathbb{E}_n[K_\varpi(W)\rho_\tau(X_a - Z^a\beta)] \ : \ \beta_j = 0 \ \text{ if } |\widehat{\beta}_{uj}| \leqslant \lambda_u/\widehat{\sigma}^Z_{a\varpi j}.$$

(2) Run Post-Lasso of $f_u Z^a_j$ on $f_u Z^a_{-j}$; keep the residual $\widetilde{v} := f_u(Z^a_j - Z^a_{-j}\widetilde{\gamma}^j_u)$,

$$\widehat{\gamma}^j_u \in \arg\min_\theta \mathbb{E}_n[K_\varpi(W) f^2_u (Z^a_j - Z^a_{-j}\gamma)^2] + \lambda\|\widehat{\Gamma}_u\gamma\|_1$$
$$\widetilde{\gamma}^j_u \in \arg\min_\gamma \mathbb{E}_n[K_\varpi(W) f^2_u (Z^a_j - Z^a_{-j}\gamma)^2] \ : \ \text{support}(\gamma) \subseteq \text{support}(\widehat{\gamma}^j_u).$$

(3) Run Instrumental Quantile Regression of $X_a - Z^a_{-j}\widetilde{\beta}_{u,-j}$ on $Z^a_j$ using $\widetilde{v}$ as the instrument,

$$\check{\beta}_{uj} \in \arg\min_{\alpha \in \mathcal{A}_{uj}} \frac{\{\mathbb{E}_n[K_\varpi(W)(1\{X_a \leqslant Z^a_j\alpha + Z^a_{-j}\widetilde{\beta}_{u,-j}\} - \tau)\widetilde{v}]\}^2}{\mathbb{E}_n[K_\varpi(W)(1\{X_a \leqslant Z^a_j\alpha + Z^a_{-j}\widetilde{\beta}_{u,-j}\} - \tau)^2\widetilde{v}^2]}$$

where $\mathcal{A}_{uj} := \{\alpha \in \mathbb{R} : |\alpha - \widetilde{\beta}_{uj}| \leqslant 10/\{\widehat{\sigma}^Z_{a\varpi j}\log n\}\}$.

**Detailed version of Algorithm 3.4 ($\mathcal{W}$-Conditional PQGM)**

For each $u = (a, \tau, \varpi) \in \mathcal{U} = V \times \mathcal{T} \times \mathcal{W}$ perform the following:

(1) Run Post-$\ell_1$-quantile regression of $X_a$ on $X_{-a}$,

$$\widehat{\beta}_u \in \arg\min_\beta \mathbb{E}_n[K_\varpi(W)\rho_\tau(X_a - X'_{-a}\beta)] + \lambda_{0\mathcal{W}} \textstyle\sum_{j\in[d]} \widehat{\sigma}^X_{a\varpi j}|\beta_j|$$
$$\widetilde{\beta}_u \in \arg\min_\beta \mathbb{E}_n[K_\varpi(W)\rho_\tau(X_a - X'_{-a}\beta)] : \ \beta_j = 0 \ \text{ if } |\widehat{\beta}_{uj}| \leqslant \lambda_{0\mathcal{W}}/\widehat{\sigma}^X_{a\varpi j}.$$

(2) Set $\widehat{\varepsilon}_{iu} = 1\{X_{ia} \leqslant X'_{i,-a}\widetilde{\beta}_u\} - \tau$ for $i \in [n]$, compute $\bar{\lambda}_{V\mathcal{T}\mathcal{W}}$ via (3.21).

(3) Run Post-$\ell_1$-quantile regression of $X_a$ on $X_{-a}$,

$$\widehat{\beta}_u \in \arg\min_\beta \mathbb{E}_n[K_\varpi(W)\rho_\tau(X_a - X'_{-a}\beta)] + \bar{\lambda}_{V\mathcal{T}\mathcal{W}} \textstyle\sum_{j\in[d]} \{\mathbb{E}_n[K_\varpi(W)\widehat{\varepsilon}^2_u X^2_{-a,j}]\}^{1/2}|\beta_j|.$$
$$\check{\beta}_u \in \arg\min_\beta \mathbb{E}_n[K_\varpi(W)\rho_\tau(X_a - X'_{-a}\beta)] \ : \ \beta_j = 0 \ \text{ if } |\widehat{\beta}_{uj}| \leqslant \bar{\lambda}_{V\mathcal{T}\mathcal{W}}/\{\mathbb{E}_n[K_\varpi(W)\widehat{\varepsilon}^2_u X^2_{-a,j}]\}^{1/2}.$$

## Appendix B. Simulations of Quantile Graphical Models

In this section, we perform numerical examples to illustrate the performance of the estimators proposed for QGMs. We will consider several different designs. In order to compare with other proposals we will consider both Gaussian and non-Gaussian examples.

B.1. **Isotropic Non-Gaussian Example.** In general, the equivalence between a zero in the inverse covariance matrix and a pair of conditional independent variables will break down for non-gaussian distributions. The nonparanormal graphical models extends Gaussian Graphical Models to Semiparametric Gaussian Copula models by transforming the variables with smooth functions. We illustrate the applicability of CIQGM in representing the conditional independence structure of a set of variables when the random variables are not even jointly nonparanormal.

Consider i.i.d. copies of an $d$-dimensional random vector $\tilde{X}_V = (\tilde{X}_1, \ldots, \tilde{X}_{d-1}, \tilde{X}_d)$ from the following multivariate normal distribution, $\tilde{X}_V \sim N(0, I_{d\times d})$, where $I_{d\times d}$ is the identity matrix. Further, we generate

$$X_d = -\sqrt{\tfrac{2}{3\pi-2}} + \sqrt{\tfrac{\pi}{3\pi-2}}\tilde{X}^2_{d-1}|\tilde{X}_d|. \tag{B.35}$$

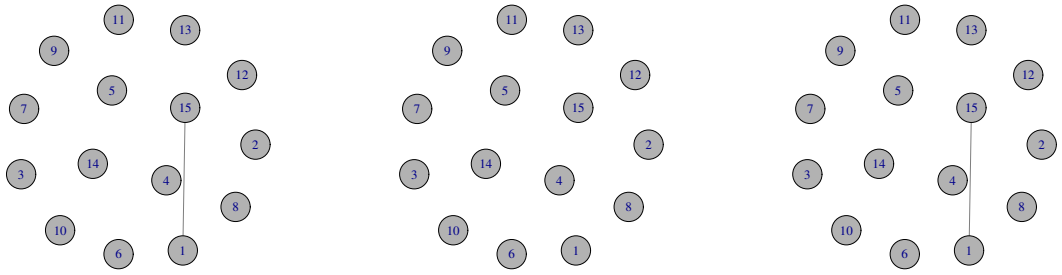
It follows that $\mathrm{E}[X_d] = \sqrt{\frac{\pi}{3\pi-2}}(\mathrm{E}[|\tilde{X}_d|] - \sqrt{2/\pi}) = 0$ and $\mathrm{Var}(X_d) = \frac{\pi}{3\pi-2}(\mathrm{E}[\tilde{X}_d^2 \cdot \tilde{X}_{d-1}^4] - \frac{2}{\pi}) = 1$. In addition, equation (B.35) is a location-scale-shift model in which the conditional median of the response is zero while quantile functions other than the median are nonzero. We define vector $X_V$ as

$$X_V = (X_d, \tilde{X}_1, ..., \tilde{X}_{d-1})'.$$

In this new set of variables, only $X_d$ and $\tilde{X}_{d-1}$ (i.e., node 1 and 15, when $d = 15$) are not conditionally independent. Nonetheless, the covariance matrix of $X_V$ is still $I_{d\times d}$.

Next we consider an example with $n = 300$ and $d = 15$. We show graphs, in Figure 3 and 4, estimated by both CIQGM(s) and GGMs in this non-Gaussian setting.

(A) CIQGM(0.2) (B) CIQGM(0.5) (C) CIQGM(0.8)



(D) CIQGM-Union (E) Gaussian (F) Nonparanormal

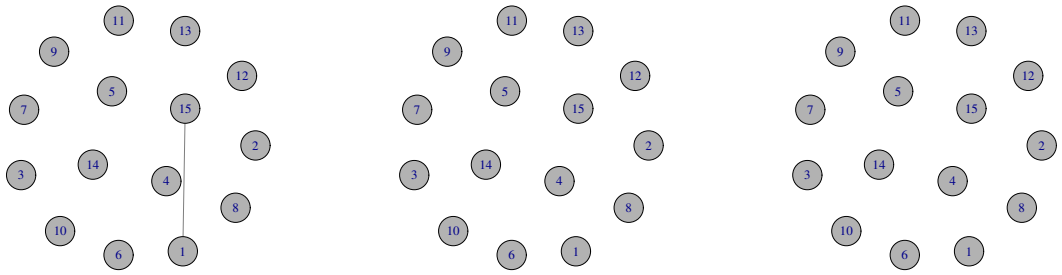


FIGURE 3. QGM and GGM

In Figure 3, Gaussian means the graph is estimated by using graphical lasso without any transformation of $X_V$, and the final graph is chosen by Extended Bayesian Information Criterion (ebic), see [42]. Nonparanormal means the graph is estimated using graphical lasso (likelihood based approach) with nonparanormal

transformation of $X_V$, see [58], and again the final graph is chosen by ebic. Both graphs are estimated using R-package **huge**.

In Figure 4, as a robustness check, we also compare results produced by CIQGM with those produced by neighborhood selection methods (pseudo-likelihood approach), e.g. TIGER of [59] in R-package **flare** the left graph is when choosing the turning parameter to be $\sqrt{\frac{\log d}{n}}$ while the right graph is when choosing the tuning parameter to be $2\sqrt{\frac{\log d}{n}}$. Throughout, we use Tiger2 represent TIGER with penalty level $2\sqrt{\frac{\log d}{n}}$. As expected, GGM cannot detect the correct dependence structure when the joint distribution is non-Gaussian while CIQGM can still represent the right conditional independence structure.

(A) Tiger1 (B) Tiger2

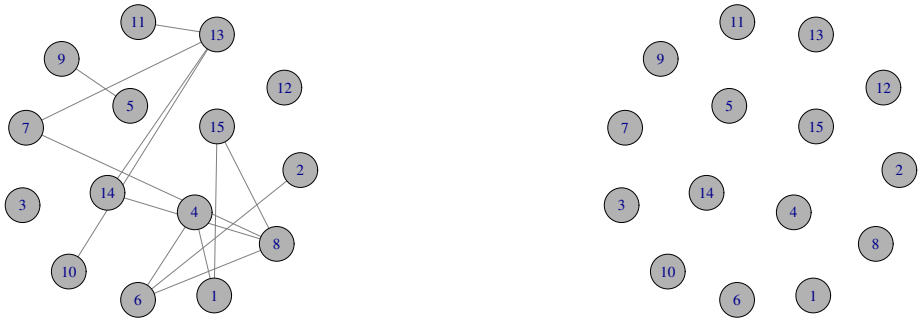



FIGURE 4. TIGER

## B.2. **Gaussian Examples.**

B.2.1. *Graph Recovery.* In this subsection, we start with comparing the numerical performance of QGM and other methods, e.g. TIGER of [59] and graphical lasso algorithm (Glasso) of [43], in graph recovery using simulated datasets with different pairs of $(n, d)$. We start with one simulation for illustration purpose (the results are summarized in Figure 5), and then we show the performance of QGM through estimated degree distribution with 100 simulations (the results are summarized in Figure 6).

We mainly consider the Hub graph, as mentioned in [59], which also corresponds to the star network mentioned in [2, 3]. In line with [59], we generate a $d$-dimensional sparse graph $G^I = (V, E^I)$ represents the conditional independence structure between the variables. In our simulations, we consider 12 settings to compare these methods: (A) $n = 200$, $d = 10$; (B) $n = 200$, $d = 20$; (C) $n = 200$, $d = 40$; (D) $n = 400$, $d = 10$; (E) $n = 400$, $d = 20$; (F) $n = 400$, $d = 40$; (G) $n = 200$, $d = 100$; (H) $n = 200$, $d = 200$; (I) $n = 200$, $d = 400$; (J) $n = 400$, $d = 100$; (K) $n = 400$, $d = 200$; (L) $n = 400$, $d = 400$. We adopt the following model for generating undirected graphs and precision matrices.

**Hub graph.** The $d$ nodes are evenly partitioned into $d/20$ (or $d/10$ when $d < 20$) disjoint groups with each group contains 20 (or 10) nodes. Within each group, one node is selected as the hub and we add edges between the hub and the other 19 (or 9) nodes in that group. For example, the resulting graph has 190 edges when $d = 200$ and 380 edges when $d = 400$. Once the graph is obtained, we generate an adjacency matrix $E^I$ by setting the nonzero off-diagonal elements to be 0.3 and the diagonal elements to be 0. We calculate its smallest eigenvalue $\Lambda_{\min}(E^I)$. The precision matrix is constructed as

$$\Theta = \mathbf{D}[E^I + (|\Lambda_{\min}(E^I)| + 0.2) \cdot I_{d\times d}]\mathbf{D} \tag{B.36}$$

where $\mathbf{D} \in \mathbb{R}^{d\times d}$ is a diagonal matrix with $\mathbf{D}_{jj} = 1$ for $j = 1, ..., d/2$ and $\mathbf{D}_{jj} = 1.5$ for $j = d/2 + 1, ..., d$. The covariance matrix $\Sigma := \Theta^{-1}$is then computed to generate the multivariate normal data: $X_1, ...., X_d \sim N(0, \Sigma)$. Below we provide simulation results using different estimators: PQGM[6], TIGER and Glasso. We start with one simualtion as an illustration:

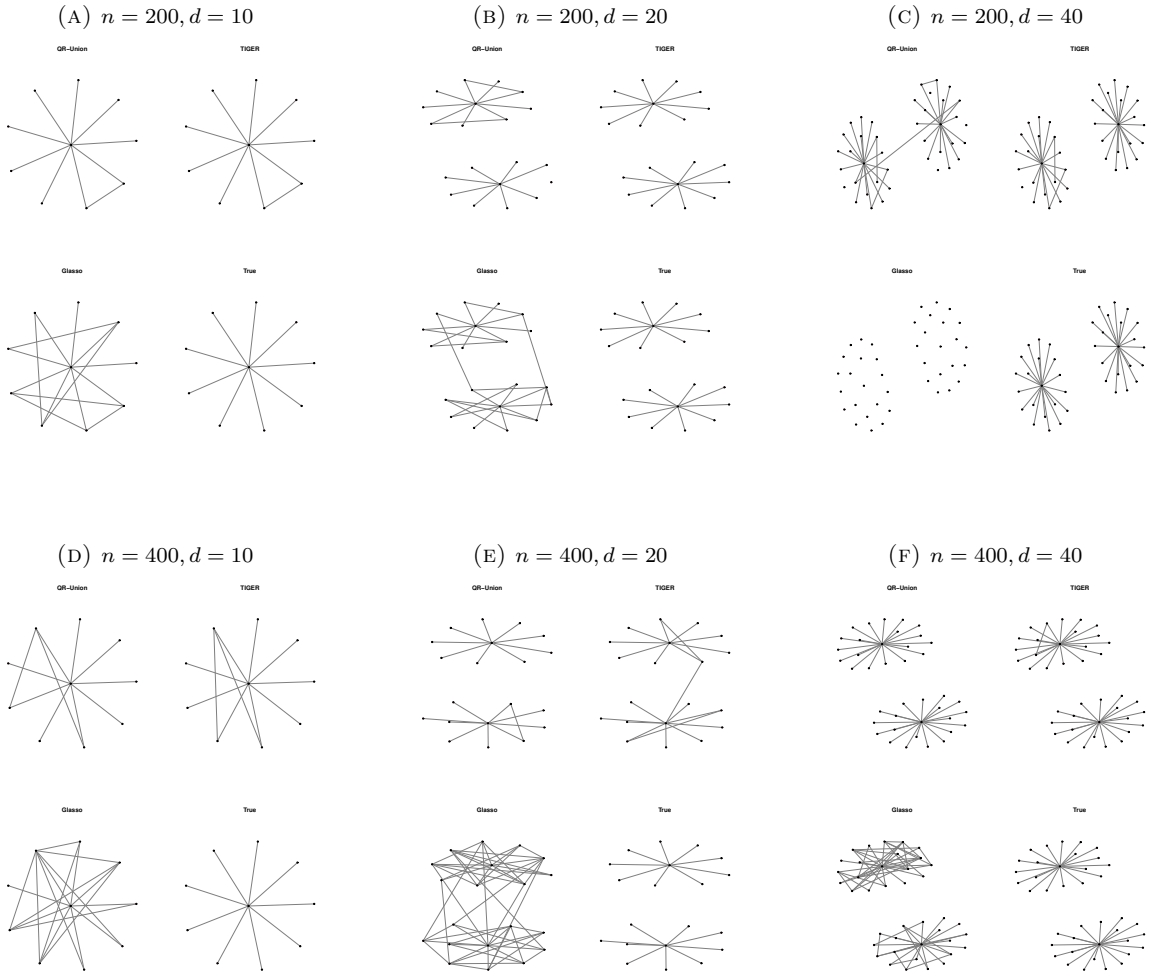



FIGURE 5. One simulation

[6]Given the graphs are generated from multivariate Gaussian distribution we can use PQGM to simplify the computation.

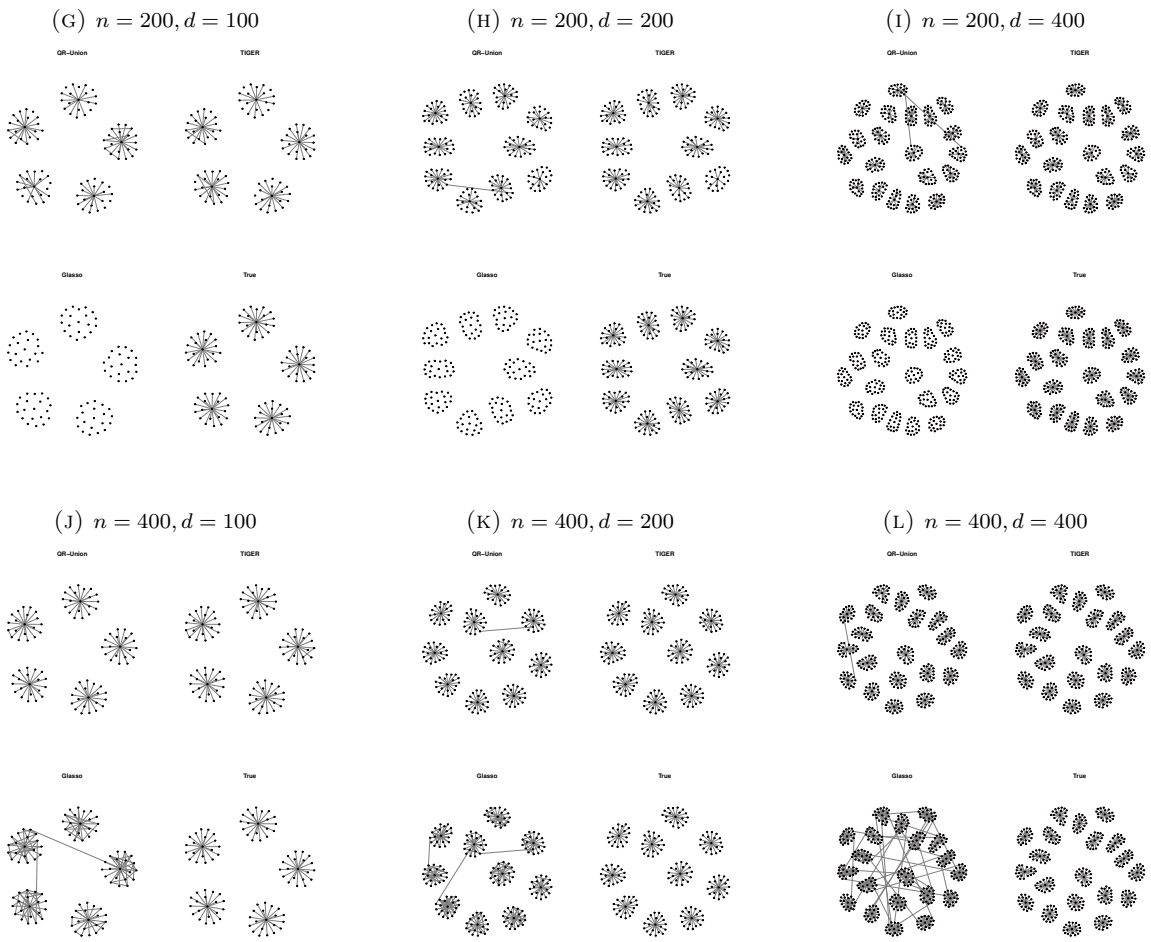



FIGURE 5. One simulation (Cont.)

Figure 5 shows that: for the low dimensional cases, $d = 10, 20, 40$, $n$ is large compared to $d$, CIQGM is comparable to TIGER and both are better than Glasso in terms of false positives; for the high dimensional cases, $d = \{100, 200, 300\}$, we can compare the performance of different graph estimators through looking at the denseness of the estimated graph (e.g., whether it is even or not), and again, both CIQGM and TIGER perform well in terms of graph recovery as compared to Glasso, and their performance are getting better when $n$ is increasing.

In what follows, Figure 6 shows the degree distribution of true graph, the estimated ones, and the standard deviations of the degree difference (between the true graph and the estimated ones). It is based on simulations of Hub graph with $n = 500$ and $d = 40$. Simulated 100 times.

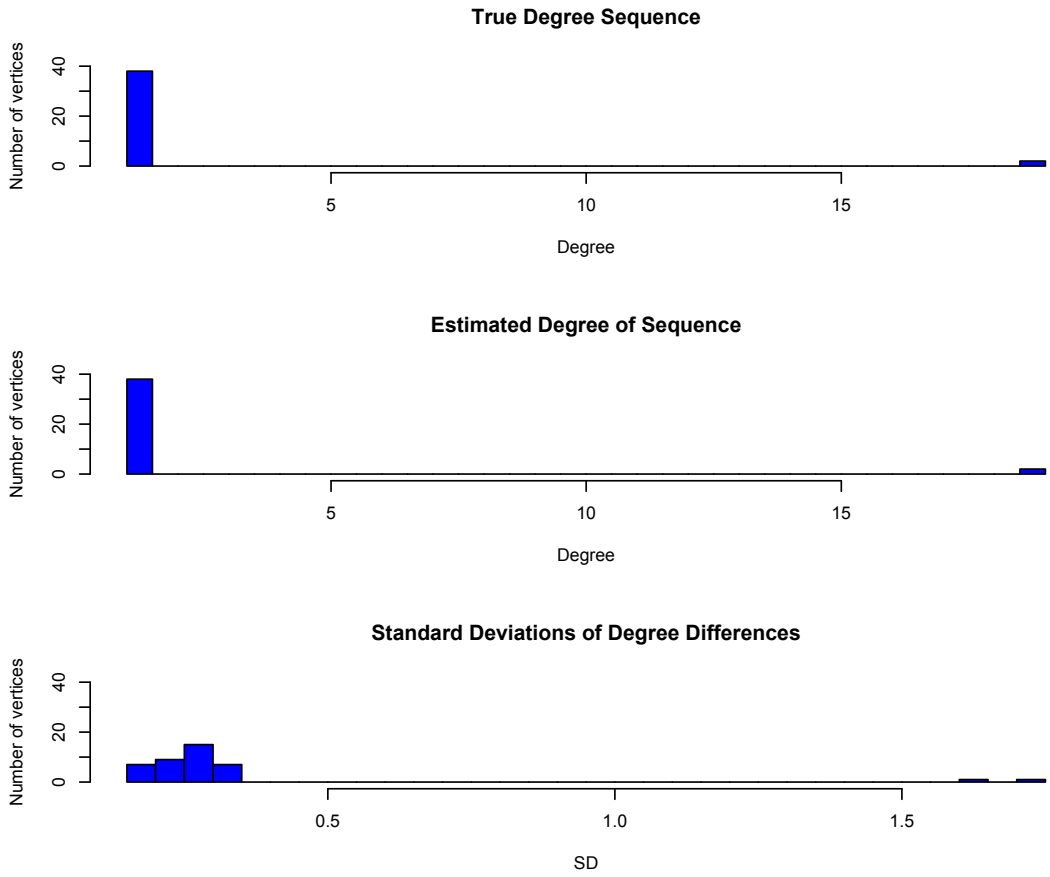



FIGURE 6. Upper panel shows the degree distribution of the true graphs. Middle panel shows the degree distribution of the estimated graphs. Bottom panel shows the standard deviations of the degree difference (between the true graph and the estimated ones). Hub graph with $n = 500$ and $d = 40$. Simulated 100 times.

B.2.2. *Inference.* In this subsection Table 2 shows the numerical performance of CIQGM, based on Algorithm 3.1, on estimating Erdős-Rényi random graphs. More precisely, we construct approximate 90% confidence intervals for $\beta_{ab}$ with $\tau = 0.5$, and we report the coverage probabilities. Note, in the jointly Gaussian distributed case, we have closed form solution of $\beta_{ab}$ as shown in Example 2.

**Erdős-Rényi random graph.** We add an edge between each pair of nodes with probability $\sqrt{\log d/n}/d$ independently. Once the graph is obtained, we construct the adjacency matrix $E^I$ and generate the precision matrix $\Theta$ using (B.36) but setting $\mathbf{D}_{jj} = 1$ for $j = 1, ..., d/2$ and $\mathbf{D}_{jj} = 1.5$ for $j = d/2 + 1, ..., d$. We then invert $\Theta$ to get the covariance matrices $\Sigma := \Theta^{-1}$ and generate the multivariate Gaussian data: $X_1, ...., X_d \sim N(0, \Sigma)$.

TABLE 2. Erdős-Rényi Random Graph

| | $(a,b)$ | $n=200$ | $n=500$ | $n=1000$ |
|---|---|---|---|---|
| d =20 | (1,20) | 84.0 | 87.5 | 91.5 |
| | (10,11) | 84.0 | 88.0 | 92.5 |
| | (19, 20) | 86.5 | 86.0 | 90.0 |
| | ACP | 86.3 | 89.4 | 89.8 |
| d=50 | (1,50) | 86.5 | 93.0 | 90.5 |
| | (25,26) | 88.0 | 87.0 | 91.0 |
| | (49, 50) | 87.5 | 90.5 | 87.5 |
| | ACP | 86.7 | 89.4 | 90.1 |
| d=100 | (1,100) | 82.5 | 81 | 89.0 |
| | (50,51) | 85.0 | 86 | 92.0 |
| | (99, 100) | 78.5 | 84 | 87.0 |
| | ACP | 86.5 | 86.8 | 90 |

$(a,b)$, coverage probability for $\beta_{ab}$; ACP, average coverage probability for $\beta_{ab}$, with $a \in V$, $b \in V \backslash \{a\}$. Simulated 200 times.

## APPENDIX C. PROOFS OF SECTION 4

*Proof of Theorem 1.* By Lemma 6, under Condition CI, for any $\theta$ such that $\|\theta\|_0 \leqslant Cs\ell_n$, $\ell_n \to \infty$ slowly, we have that

$$\|\sqrt{f_u} Z^a\theta\|_{n,\varpi} / \{\mathrm{E}[K_\varpi(W) f_u (Z^a\theta)^2]\}^{1/2} = 1 + o_P(1).$$

Moreover, $\mathrm{E}[K_\varpi(W) f_u (Z^a\theta)^2] \geqslant \underline{f}_u \mathrm{E}[K_\varpi(W)(Z^a\theta)^2]$, $\mathrm{E}[K_\varpi(W)(Z^a\theta)^2] = \mathrm{E}[(Z^a\theta)^2|\varpi]\mathrm{P}(\varpi)$, and $\mathrm{E}[(Z^a\theta)^2 \mid \varpi] \geqslant c\|\theta\|^2$ by Condition CI. Lemma 6 further implies that the ratio of the minimal and maximal eigenvalues of order $s\ell_n$ are bounded away from zero and from above uniformly over $\varpi \in \mathcal{W}$ and $a \in V$ with probability $1 - o(1)$. Therefore, since $c\{\mathrm{P}(\varpi)\}^{1/2}\|\delta\|_1 \leqslant \|\delta\|_{1,\varpi} \leqslant C\{\mathrm{P}(\varpi)\}^{1/2}\|\delta\|_1$, we have $\kappa_{u,2\mathbf{c}} \geqslant c$ uniformly over $u \in \mathcal{U}$ with the same probability for $n$ large enough, see for instance [21].

To establish rates of convergence of the estimator obtained in Step 1 we will apply Lemma 1. Consider the events $\Omega_1, \Omega_2$, and $\Omega_3$ as defined in (D.49), (D.50) and (D.51). By the choice of $\lambda_u$ we have $\mathrm{P}(\Omega_1) \geqslant 1 - o(1)$. By Condition CI with $\bar{R}_{u\xi} \leqslant Cs\log(p|V|n)/n$ and Lemma 2 we have $\mathrm{P}(\Omega_2) \geqslant 1 - o(1)$. Moreover, $\mathrm{P}(\Omega_3) \geqslant 1 - o(1)$ by Lemma 3 with $t_3 \leqslant Cn^{-1/2}\sqrt{(1+d_W)\log(p|V|nL_f)}$.

Using the same argument (with $Z^a$ replacing $X_{-a}$) as in (C.46), (C.47), and (C.48), for

$$\delta \in A_u := \Delta_{\varpi,2\mathbf{c}} \cup \{v : \|v\|_{1,\varpi} \leqslant 2\mathbf{c}\bar{R}_{u\xi}/\lambda_u, \|\sqrt{f_u}Z^a v\|_{n,\varpi} \geqslant C\sqrt{s(1+d_W)\log(p|V|n)/n}/\kappa_{u,2\mathbf{c}}\},$$

with the restricted set defined as $\Delta_{u,2\tilde{\mathbf{c}}} = \{\delta : \|\delta_{T_u^c}\|_1 \leqslant 2\tilde{\mathbf{c}}\|\delta_{T_u}\|_1\}$ for $u \in \mathcal{U}$. we have $\bar{q}_{A_u} \geqslant c(\underline{f}_{\mathcal{U}}^{3/2}/\bar{f}')\mu_{\mathcal{W}}^{1/2}/\{\sqrt{s}\max_{a\in V,i\leqslant n}\|Z_i^a\|_\infty\}$ where $\max_{a\in V,i\leqslant n}\|Z_i^a\|_\infty \lesssim_P M_n$. Thus the conditions on $\bar{q}_{A_u}$ are satisfied since Condition CI assumes $M_n^2 s^2\log(p|V|n) \leqslant n\mu_{\mathcal{W}}\underline{f}_{\mathcal{U}}^3$. The conditions on the approximation error are assumed in Condition CI.

Therefore, setting $\xi = 1/\log n$, by Lemma 1 we have uniformly over $u = (a,\tau,\varpi) \in \mathcal{U}$

$$\begin{aligned} \|\sqrt{f_u}Z^a(\widehat{\beta}_u - \beta_u)\|_{n,\varpi} &\lesssim \sqrt{(1+(t_3/\lambda_u)\bar{R}_{u\xi}} + (\lambda_u + t_3)\sqrt{s} \lesssim \sqrt{\tfrac{s(1+d_W)\log(p|V|n)}{n\tau(1-\tau)}} \\ \|\widehat{\beta}_u - \beta_u\|_{1,\varpi} &\lesssim s\sqrt{\tfrac{(1+d_W)\log(p|V|n)}{n}} \end{aligned} \tag{C.37}$$

here we used that $\lambda_u \leqslant C\sqrt{\frac{(1+d_W)\log(p|V|n)}{n}}$. Indeed by Lemma 15 with $\tilde{x}_{ij} = K_\varpi(W_i)Z_{ij}^a$ and $\widehat{\sigma}_j \geqslant c\mathrm{P}(\varpi)^{1/2}$, we can bound $\Lambda_{a\tau\varpi}(1-\xi/\{|V|n^{1+2d_W}\}|X_{-a},W)$ under $M_n^2\log(p|V|n/\{\tau(1-\tau)\}) = o(n\tau(1-\tau)\mu_{\mathcal{W}})$ for all $\tau\in\mathcal{T}$ and $\varpi\in\mathcal{W}$, and the bound on $\lambda_u$ follows from the union bound.

Let $\delta_u = \widehat{\beta}_u - \beta_u$. By triangle inequality it follows that

$$\{\mathrm{E}[K_\varpi(W)f_u(Z^a\delta_u)^2]\}^{1/2} \leqslant \|\sqrt{f_u}Z^a\delta_u\|_{n,\varpi} + \|\delta_u\|_1\{|(\mathbb{E}_n-\mathrm{E})[K_\varpi(W)f_u(Z^a\delta_u)^2]/\|\delta_u\|_1^2|\}^{1/2} \tag{C.38}$$

and the last term can be bounded by

$$\begin{aligned} \sup_{\|\delta\|_1\leqslant 1}|(\mathbb{E}_n-\mathrm{E})[K_\varpi(W)f_u(Z^a\delta)^2/\|\delta\|_1^2]| &\leqslant \max_{k,j}|(\mathbb{E}_n-\mathrm{E})[K_\varpi(W)f_u Z_k^a Z_j^a]| \\ &\lesssim \sqrt{\tfrac{(1+d_W)\log(p|V|n)}{n}} \end{aligned}$$

with probability $1-o(1)$ by Lemma 19 under our conditions.

Combining the relations above with (C.38), under $(1+d_W)s^2\log(p|V|n) = o(n)$ we have uniformly over $u\in\mathcal{U}$

$$\begin{aligned} \|\delta_u\| &\lesssim \{\mathrm{E}[(Z^a\delta_u)^2 \mid \varpi]\}^{1/2} \lesssim \{\mathrm{P}(\varpi)\}^{-1/2}\{\mathrm{E}[K_\varpi(W)(Z^a\delta_u)^2]\}^{1/2} \\ &\lesssim \{\mathrm{P}(\varpi)\underline{f}_u\}^{-1/2}\{\mathrm{E}[K_\varpi(W)f_u(Z^a\delta_u)^2]\}^{1/2} \\ &\leqslant \{\mathrm{P}(\varpi)\underline{f}_u\}^{-1/2}\|\sqrt{f_u}Z^a\delta_u\|_{n,\varpi} + \{\mathrm{P}(\varpi)\underline{f}_u\}^{-1/2}\sqrt[4]{\tfrac{(1+d_W)\log(p|V|n)}{n}}\|\delta_u\|_1 \\ &\leqslant C\sqrt{\tfrac{s(1+d_W)\log(p|V|n)}{n\underline{f}_u\mathrm{P}(\varpi)}}. \end{aligned}$$

given $\|\delta_u\|_1 \leqslant \|\delta_u\|_{1,\varpi}/\mathrm{P}(\varpi)^{1/2}$, (C.37), and $s^2(1+d_W)\log(p|V|n) = o(n\mu_{\mathcal{W}}^4\underline{f}_{\mathcal{U}}^2)$ assumed in Condition CI.

Finally, let $\widehat{\beta}_u^{\bar{\lambda}}$ obtained by thresholding the estimator $\widehat{\beta}_u$ with $\bar{\lambda} := \sqrt{(1+d_W)\log(p|V|n)/n}$ (note that each component is weighted by $\mathbb{E}_n[K_\varpi(W)(Z_j^a)^2]^{1/2}$). By Lemma 17, we have with probability $1-o(1)$

$$\begin{aligned} \|Z^a(\widehat{\beta}_u^{\bar{\lambda}} - \beta_u)\|_{n,\varpi} &\lesssim \sqrt{s(1+d_W)\log(p|V|n)/n} \\ \|\widehat{\beta}_u^{\bar{\lambda}} - \beta_u\|_{1,\varpi} &\lesssim s\sqrt{(1+d_W)\log(p|V|n)/n} \\ |\mathrm{support}(\widehat{\beta}_u^{\bar{\lambda}})| &\lesssim s \end{aligned}$$

by the choice of $\bar{\lambda}$ and the rates in (C.37)

■

*Proof of Theorem 2.* We verify Assumption C4 and Condition WL for the weighted Lasso model with index set $\mathcal{U} \times [p]$ where $Y_u = K_\varpi(W)Z_j^a$, $X_u = K_\varpi(W)Z_{-j}^a$, $\theta_u = \bar\gamma_u^j$, $a_u = (f_u, \bar r_{uj})$, $\bar r_{uj} = K_\varpi(W)Z_{-j}^a(\gamma_u^j - \bar\gamma_u^j)$, $S_{uj} = K_\varpi(W)f_u^2(Z_j^a - Z_{-j}^a\gamma_u^j)Z_{-j}^a = K_\varpi(W)f_u v_{uj}Z_{-j}^a$, and $w_u = K_\varpi(W)f_u^2$. We will take $N_n = |V|p^2\{pn^3\}^{1+d_W}$ in the definition of $\lambda$.

We first verify Condition WL. We have $\mathrm{E}[S_{ujk}^2] \leqslant \bar f^2\mathrm{E}[|v_{uj}Z_{-j,k}^a|^2] \leqslant \bar f^2\{\mathrm{E}[|v_{uj}|^4|Z_{-j,k}^a|^4]\}^{1/2} \leqslant C$ by the bounded fourth moment condition. We have that

$$\frac{\mathrm{E}[|S_{ujk}|^3]^{1/3}}{\mathrm{E}[|S_{ujk}|^2]^{1/2}} = \frac{\mathrm{E}[|S_{ujk}|^3 \mid \varpi]^{1/3}}{\mathrm{E}[|S_{ujk}|^2 \mid \varpi]^{1/2}}\{\mathrm{P}(\varpi)\}^{-1/6} = \frac{\mathrm{E}[|f_u v_{uj}Z_{-jk}^a|^3 \mid \varpi]^{1/3}}{\mathrm{E}[|f_u v_{uj}Z_{-jk}^a|^2 \mid \varpi]^{1/2}}\{\mathrm{P}(\varpi)\}^{-1/6} =: M_{uk}$$

By the choice of $N_n$ and $\Phi^{-1}(1-t) \leqslant C\sqrt{\log(1/t)}$, we have $M_{uk}\Phi^{-1}(1 - \xi/\{2pN_n\}) \leqslant M_{uk}C(1 + d_W)\log^{1/2}(pn|V|) \leqslant C\delta_n n^{1/6}$ where the last inequality holds by Condition CI so Condition WL(i) holds.

To verify Condition WL(ii) we will establish the validity of the choice of $N_n$. We will consider $u = (a, \tau, \varpi) \in \mathcal{U}$ and $u' = (a, \tau', \varpi') \in \mathcal{U}$. By Condition CI we have that

$$|f_u - f_{u'}| \leqslant L_f\|u - u'\| \quad \text{and} \quad \mathrm{E}[|K_\varpi(W) - K_{\varpi'}(W)|] \leqslant L_K\|\varpi - \varpi'\|. \tag{C.39}$$

Further, by Lemma 5 we have

$$\|\gamma_u^j - \gamma_{u'}^j\| \leqslant L_\gamma\{\|u - u'\| + \|\varpi - \varpi'\|^{1/2}\}. \tag{C.40}$$

By definition we have

$$S_{ujk} - S_{u'jk} = \{K_\varpi(W)f_u^2 - K_{\varpi'}(W)f_{u'}^2\}\{Z_j^a - Z_{-j}^a\gamma_u^j\}Z_k^a - K_{\varpi'}(W)f_{u'}^2\{Z_{-j}^a(\gamma_u^j - \gamma_{u'}^j)\}Z_k^a$$

and note that $f_u + f_{u'} \leqslant 2f_u + L\|u - u'\|$, $|Z_j - Z_{-j}^a\gamma_u^j| \cdot |Z_k^a| \leqslant |Z_j|^2 + 2|Z_k^a|^2 + |Z_{-j}^a\gamma_u^j|^2$. Moreover,

$$\begin{aligned}|K_\varpi(W)f_u^2 - K_{\varpi'}(W)f_{u'}^2| &\leqslant K_\varpi(W)K_{\varpi'}(W)|f_u^2 - f_{u'}^2| + (f_u + f_{u'})^2|K_\varpi(W) - K_{\varpi'}(W)| \\ &\leqslant 2\bar f K_\varpi(W)K_{\varpi'}(W)|f_u - f_{u'}| + 4\bar f^2|K_\varpi(W) - K_{\varpi'}(W)|\end{aligned}$$

Using these relations we have $|\mathbb{E}_n[S_{ujk} - S_{u'jk}]| \leqslant (I) + (II)$ where

$$\begin{aligned}(I) \quad &= \mathbb{E}_n[|K_\varpi(W)f_u^2 - K_{\varpi'}(W)f_{u'}^2| \cdot |\{Z_j^a - Z_{-j}^a\gamma_u^j\}Z_k^a|] \\ &\leqslant \max_{i\leqslant n}\|Z_i^a\|_\infty^2(1 + \|\gamma_u^j\|_1)\mathbb{E}_n[2\bar f K_\varpi(W)K_{\varpi'}(W)|f_u - f_{u'}| + 4\bar f^2|K_\varpi(W) - K_{\varpi'}(W)|] \\ &\leqslant (\bar f + \bar f^2)C\sqrt{s}\max_{i\leqslant n}\|Z_i^a\|_\infty^2\{L_f\|u - u'\| + \mathbb{E}_n[|K_\varpi(W) - K_{\varpi'}(W)|]\} \\ (II) \quad &= \mathbb{E}_n[K_{\varpi'}(W)f_{u'}^2|Z_{-j}^a(\gamma_u^j - \gamma_{u'}^j)Z_k^a|] \\ &\leqslant \bar f^2\mathbb{E}_n[\|Z^a\|_\infty^2]\|\gamma_u^j - \gamma_{u'}^j\|_1 \\ &\leqslant \bar f^2\mathbb{E}_n[\|Z^a\|_\infty^2]\sqrt{p}L_\gamma\{\|u - u'\| + \|\varpi - \varpi'\|^{1/2}\}\end{aligned}$$

Moreover, we have that $\max_{i\leqslant n}\|Z_i^a\|_\infty^2 \lesssim_P M_n^2$. For $d_{\mathcal{U}} = \|\cdot\|$, an uniform $\epsilon$-cover of $\mathcal{U}$ satisfies $(6\mathrm{diam}(\mathcal{U})/\epsilon)^{1+d_W} \geqslant N(\epsilon, \mathcal{U}, \|\cdot\|)$. We will set $1/\epsilon = (1 + \bar f^2)\{L_\gamma + L_f\}^2 pnM_n^2\log^2(p|V|n)/\{\mu_{\mathcal{W}}\underline{f}_{\mathcal{U}}^2\} \leqslant pn^3$ so that with probability $1 - o(1)$, for any pair $u, u' \in \mathcal{U}$, $\|u - u'\| \leqslant \epsilon$, we have

$$\begin{aligned}|\mathbb{E}_n[S_{ujk} - S_{u'jk}]| &\lesssim (\bar f + \bar f^2)\sqrt{s}\max_{i\leqslant n}\|Z_i^a\|_\infty^2\{L_f\epsilon + \mathbb{E}_n[|K_\varpi(W) - K_{\varpi'}(W)|]\} \\ &\quad + \bar f^2\mathbb{E}_n[\|Z_i^a\|_\infty^2]\sqrt{p}L_\gamma\{\epsilon + \epsilon^{1/2}\} \\ &\lesssim \delta_n n^{-1/2}\{\mu_{\mathcal{W}}\underline{f}_{\mathcal{U}}^2\}^{1/2} + \sqrt{s}M_n\log(n)\mathbb{E}_n[|K_\varpi(W) - K_{\varpi'}(W)|]\end{aligned}$$

by the choice of $\epsilon$. To control the last term, note that $\mathcal{W}$ is a VC-class of events with VC dimension $d_W$. Thus by Lemma 19, with probability $1-o(1)$

$$\begin{aligned}\mathbb{E}_n[|K_\varpi(W)-K_{\varpi'}(W)|] &\leqslant |(\mathbb{E}_n-\mathrm{E})[|K_\varpi(W)-K_{\varpi'}(W)|]| + \mathrm{E}[|K_\varpi(W)-K_{\varpi'}(W)|]\\ &\leqslant \sup_{\varpi,\varpi'\in\mathcal{W},\|\varpi-\varpi'\|\leqslant\epsilon} |(\mathbb{E}_n-\mathrm{E})[|K_\varpi(W)-K_{\varpi'}(W)|]| + L_K\epsilon\\ &\lesssim \sqrt{\tfrac{d_W\log(n/\epsilon)}{n}}\epsilon^{1/2} + \tfrac{d_W\log(n/\epsilon)}{n} + L_K\epsilon\end{aligned}$$

which yields uniformly over $u\in\mathcal{U}$ and $j\in[p]$

$$|\mathbb{E}_n[S_{ujk}-S_{u'jk}]| \lesssim \delta_n n^{-1/2}\mu_{\mathcal{W}}^{1/2}\underline{f}_{\mathcal{U}} \tag{C.41}$$

under $\sqrt{\epsilon d_W\log(n/\epsilon)}M_n\log n = o(\mu_{\mathcal{W}}^{1/2}\underline{f}_{\mathcal{U}})$ and $d_W\log(n/\epsilon)M_n\log n = o(n^{1/2}\mu_{\mathcal{W}}^{1/2}\underline{f}_{\mathcal{U}})$ assumed in Condition CI. In turn this implies

$$\sup_{|u-u'|\leqslant\epsilon}\max_{j,k\in[p],j\neq k}\frac{|\mathbb{E}_n[S_{ujk}-S_{u'jk}]|}{\mathrm{E}[S_{ujk}^2]^{1/2}} \leqslant \delta_n n^{-1/2}$$

since $\mathrm{E}[S_{ujk}^2]\geqslant c\mu_{\mathcal{W}}\underline{f}_{\mathcal{U}}^2$. Using the same choice of $\epsilon$, similar arguments also imply

$$\sup_{|u-u'|\leqslant\epsilon}\max_{j,k\in[p],j\neq k}\frac{|\mathrm{E}[S_{ujk}^2-S_{u'jk}^2]|}{\mathrm{E}[S_{ujk}^2]} \leqslant \delta_n \tag{C.42}$$

To establish the last requirement of Condition WL(ii), note that

$$\sup_{u\in\mathcal{U}}\max_{j,k\in[p],j\neq k}|(\mathbb{E}_n-\mathrm{E})[S_{ujk}^2]| \leqslant \sup_{u\in\mathcal{U}^\epsilon}\max_{j,k\in[p],j\neq k}|(\mathbb{E}_n-\mathrm{E})[S_{ujk}^2]| + \Delta_n \tag{C.43}$$

where $\Delta_n := \sup_{u,u'\in\mathcal{U},\|u-u'\|\leqslant\epsilon}\max_{j,k\in[p],j\neq k}|(\mathbb{E}_n-\mathrm{E})[S_{ujk}^2]-(\mathbb{E}_n-\mathrm{E})[S_{u'jk}^2]|$.

To bound the first term, we will apply Corollary 2 with $k=1$, $\widehat{\mathcal{U}} := \mathcal{U}^\epsilon\times[p]$ and the vector $\{(\bar{X})_{uj} = S_{uj}, (u,j)\in\widehat{\mathcal{U}}\}$. In this case note that

$$K^2 = \mathrm{E}[\max_{i\leqslant n}\sup_{u\in\mathcal{U}}\max_{j,k\in[p],j\neq k}S_{ujk}^2] \leqslant \mathrm{E}[\max_{i\leqslant n}\sup_{u\in\mathcal{U},j\in[p]}|v_{iuj}|^2\|f_{iu}Z_i^a\|_\infty^2] \leqslant \bar{f}^2M_n^2L_n^2.$$

Therefore, by Corollary 2 and Markov inequality, we have with probability $1-o(1)$ that

$$\sup_{u\in\mathcal{U}}\max_{j,k\in[p],j\neq k}|(\mathbb{E}_n-\mathrm{E})[S_{ujk}^2]| \leqslant Cn^{-1/2}M_nL_n\log^{1/2}(p|V|n) \leqslant C\delta_n\mu_{\mathcal{W}}\underline{f}_{\mathcal{U}} + \Delta_n$$

under $M_n^2L_n^2\log(p|V|n)\leqslant\delta_n n\mu_{\mathcal{W}}^2\underline{f}_{\mathcal{U}}^2$.

To control $\Delta_n$, note that

$$\begin{aligned}|(\mathbb{E}_n-\mathrm{E})[S_{ujk}^2]-(\mathbb{E}_n-\mathrm{E})[S_{u'jk}^2]| &\leqslant |\mathbb{E}_n[S_{ujk}^2-S_{u'jk}^2]| + |\mathrm{E}[S_{ujk}^2-S_{u'jk}^2]|\\ &\leqslant \mathbb{E}_n[|S_{ujk}-S_{u'jk}|]\sup_{u\in\mathcal{U}}\max_{i\leqslant n}|2S_{iu'jk}| + |\mathrm{E}[S_{ujk}^2-S_{u'jk}^2]|\\ &\lesssim \delta_n n^{-1/2}\mu_{\mathcal{W}}^{1/2}\underline{f}_{\mathcal{U}}\bar{f}\sup_{u\in\mathcal{U}}\max_{i\leqslant n}|v_{iuj}|\|Z_i^a\|_\infty + \delta_n\mu_{\mathcal{W}}\underline{f}_{\mathcal{U}}^2\\ &\lesssim \delta_n n^{-1/2}\mu_{\mathcal{W}}^{1/2}\underline{f}_{\mathcal{U}}M_nL_n\log n + \delta_n\mu_{\mathcal{W}}\underline{f}_{\mathcal{U}}^2\end{aligned}$$

with probability $1-o(1)$ where we used (C.41) and (C.42). Therefore $\Delta_n\lesssim\delta_n\mu_{\mathcal{W}}\underline{f}_{\mathcal{U}}^2$ with probability $1-o(1)$ as required.

To verify Assumption C4(a), note that $[\partial_\theta M_u(Y_u,X,\theta_u)-\partial_\theta M_u(Y_u,X,\theta_u,a_u)]'\delta = -f_u^2K_\varpi(W)\bar{r}_{uj}Z_{-j}^a\delta$, so that by Cauchy-Schwartz, we have

$$\mathbb{E}_n[\partial_\theta M_u(Y_u,X,\theta_u)-\partial_\theta M_u(Y_u,X,\theta_u,a_u)]'\delta \leqslant \|f_u\bar{r}_{uj}\|_{n,\varpi}\|f_uZ_{-j}^a\delta\|_{n,\varpi} \leqslant C_{un}\|f_uZ_{-j}^a\delta\|_{n,\varpi}$$

where we choose $C_{un}$ so that $\{C_{un} \geqslant \max_{j\in[p]} \|f_u \bar{r}_{uj}\|_{n,\varpi} : u \in \mathcal{U}\}$ with probability $1-o(1)$. To bound $C_{un}$, by Lemma 4, uniformly over $u \in \mathcal{U}, j \in [p]$ we have with probability $1-o(1)$

$$\|f_u \bar{r}_{uj}\|_{n,\varpi} = \|f_u Z^a_{-j}(\gamma^j_u - \bar{\gamma}^j_u)\|_{n,\varpi} \lesssim \underline{f}_u \{\mathrm{P}(\varpi)\}^{1/2} \{n^{-1} s \log(p|V|n)\}^{1/2}$$

so that setting $C_{un} = \underline{f}_u \{\mathrm{P}(\varpi)\}^{1/2} \{n^{-1} s \log(p|V|n)\}^{1/2}$ suffices.

Next we show that Assumption C4(b) holds. First, by (C.43) and the corresponding bounds, note the uniform convergence of the loadings

$$\sup_{u\in\mathcal{U}, j,k\in[p], j\neq k} (|\mathbb{E}_n[S^2_{ujk}] - \mathrm{E}[S^2_{ujk}]| + |(\mathbb{E}_n - \mathrm{E})[K_\varpi(W) f^2_u |Z^a_j Z^a_{-jk}|^2]|) \leqslant \delta_n \mu_{\mathcal{W}} \underline{f}^2_{\mathcal{U}}$$

so that $\mathbb{E}_n[S^2_{ujk}]/\mathrm{E}[S^2_{ujk}] = 1 + o_P(1)$. It follows that $\tilde{c}$ is bounded above by a constant for $n$ large enough. Indeed, uniformly over $u \in \mathcal{U}$, $j \in [p]$, since $c\underline{f}_u \leqslant \mathrm{E}[|f_u v_{uj} Z^a_k|^2 \mid \varpi]^{1/2} \leqslant C\underline{f}_u$, with probability $1-o(1)$ we have $c\underline{f}_u \mathrm{P}(\varpi)^{1/2} \leqslant \widehat{\Psi}_{u0jj} \leqslant C\underline{f}_u \mathrm{P}(\varpi)^{1/2}$ so that $c/C \leqslant \|\widehat{\Psi}_{u0}\|_\infty \|\widehat{\Psi}^{-1}_{u0}\|_\infty \leqslant C/c$.

Assumption C4(c) follows directly from the choice of $M_u(Y_u, X_u, \theta) = K_\varpi(W) f^2_u (Z^a_j - Z^a_{-j}\theta)^2$ with $\bar{q}_{A_u} = \infty$.

The result for the rate of convergence then follows from Lemma 22, namely

$$\|f_u X'_u(\widehat{\gamma}^j_u - \gamma^j_u)\|_{n,\varpi} \lesssim \frac{\|\widehat{\Psi}_{u0}\|_\infty}{\bar{\kappa}_{u,2\mathbf{c}}} \sqrt{\frac{s\log(p|V|n)}{n}} + C_{un} \lesssim \frac{\underline{f}_u \mathrm{P}(\varpi)^{1/2}}{\bar{\kappa}_{u,2\mathbf{c}}} \sqrt{\frac{s\log(p|V|n)}{n}} \tag{C.44}$$

By Lemma 6 we have that for sparse vectors, $\|\theta\|_0 \leqslant \ell_n s$ satisfies

$$\|f_u Z^a_{-j}\theta\|^2_{n,\varpi} / \mathrm{E}[K_\varpi(W) f^2_u (Z^a_{-j}\theta)^2] = 1 + o_P(1)$$

so that $\phi_{\max}(\ell_n s, uj) \leqslant C\underline{f}^2_u \mathrm{P}(\varpi)$ and $\widehat{s}_{uj} \leqslant \min_{m\in\mathcal{M}_u} \phi_{\max}(m, uj) L^2_u \leqslant Cs$ provided $L^2_u \lesssim s\{\underline{f}^2_u \mathrm{P}(\varpi)\}^{-1}$. Indeed, with probability $1-o(1)$, we have $\|\widehat{\Psi}^{-1}_{u0}\|_\infty \leqslant C\underline{f}^{-1}_u \mathrm{P}(\varpi)^{-1/2}$, so that $L_u \lesssim \underline{f}^{-1}_u \mathrm{P}(\varpi)^{-1/2} \frac{n}{\lambda}\{C_{un} + L_{un}\}$. Moreover, we can take $C_{un} \lesssim \underline{f}_u \{\mathrm{P}(\varpi) n^{-1} s \log(p|V|n)\}^{1/2}$, and $L_{un} \lesssim \{n^{-1} s \log(p|V|n)\}^{1/2}$ in Assumption C4 because

$$\begin{aligned}
&|\{\mathbb{E}_n[\partial_\gamma M_u(Y_u, X_u, \widehat{\gamma}^j_u) - \partial_\gamma M_u(Y_u, X_u, \gamma^j_u)]\}'\delta| \\
&= 2|\mathbb{E}_n[K_\varpi(W) f^2_u \{X'_u(\widehat{\gamma}^j_u - \gamma^j_u)\} X'_u \delta]| \\
&\leqslant 2\|f_u X'_u(\widehat{\gamma}^j_u - \gamma^j_u)\|_{n,\varpi} \|f_u X'_u \delta\|_{n,\varpi} =: L_{un} \|f_u X'_u \delta\|_{n,\varpi},
\end{aligned}$$

where the last inequality hold by (C.44) since $\bar{\kappa}_{u,2\mathbf{c}} \geqslant c\underline{f}_u \{\mathrm{P}(\varpi)\}^{1/2}$. The bound on the restricted eigenvalue $\bar{\kappa}_{u,2\mathbf{c}}$ holds[7] by arguments similar to (C.46) and using that $\|\delta\|_1 \leqslant C\sqrt{s}\|\delta\|$ for any $\delta \in \Delta_{u,2\mathbf{c}}$, and since for any $\|\delta\| = 1$, we have

$$\begin{aligned}
c\underline{f}_u \mathrm{P}(\varpi) &\leqslant \mathrm{E}[K_\varpi(W) f_u (Z^a \delta)^2] \\
&\leqslant \{\mathrm{E}[K_\varpi(W) f^2_u (Z^a\delta)^2]\}^{1/2} \{\mathrm{E}[K_\varpi(W)(Z^a\delta)^2]\}^{1/2} \\
&\leqslant \{\mathrm{E}[K_\varpi(W) f^2_u (Z^a\delta)^2]\}^{1/2} C \{\mathrm{P}(\varpi)\}^{1/2}
\end{aligned}$$

where the first inequality follows from the definition of $\underline{f}_u$, $\|\delta\| = 1$, and Condition CI, so that we have $\{\mathrm{E}[K_\varpi(W) f^2_u (Z^a\delta)^2]\}^{1/2} \geqslant c'\underline{f}_u \{\mathrm{P}(\varpi)\}^{1/2}$.

[7]Note that there are two restricted eigenvalues definitions, one used for the quantile regression ($\kappa_{u,2\mathbf{c}}$), and another used here for the weighted lasso ($\bar{\kappa}_{u,2\mathbf{c}}$). It is a consequence of the use of different norms.

Return to the rate of convergence we have by (C.44) and $\bar{\kappa}_{u,2\mathbf{c}} \geqslant c\underline{f}_u\{\mathrm{P}(\varpi)\}^{1/2}$ that

$$\|f_u X_u'(\widehat{\gamma}_u^j - \gamma_u^j)\|_{n,\varpi} \lesssim \frac{\underline{f}_u \mathrm{P}(\varpi)^{1/2}}{\bar{\kappa}_{u,2\mathbf{c}}} \sqrt{\frac{s\log(p|V|n)}{n}} \lesssim \sqrt{\frac{s\log(p|V|n)}{n}} \tag{C.45}$$

and the result follows by noting that $\|f_u X_u'(\widehat{\gamma}_u^j - \gamma_u^j)\|_{n,\varpi} \geqslant c\underline{f}_u \mathrm{P}(\varpi)^{1/2}\|\widehat{\gamma}_u^j - \gamma_u^j\|$ with probability $1 - o(1)$ by arguments similar to (C.46) under Condition CI.

The sparsity result follows from Lemma 21. The result for Post Lasso follows from Lemma 20 under the growth requirements in Condition CI.

■

*Proof of Theorem 3.* We will verify Assumptions C1 and C2, and the result follows from Theorem 5. The estimate of the nuisance parameter is constructed from the estimators in Steps 1 and 2 of the Algorithm.

For each $u = (a, \tau, \varpi) \in \mathcal{U}$ and $j \in [p]$, let $W_{uj} = (W, X_a, Z^a, v_{uj}, r_u)$, where $v_{uj} = f_u(Z_j^a - Z_{-j}^a\gamma_u^j)$ and let $\theta_{uj} \in \Theta_{uj} = \{\theta \in \mathbb{R} : |\theta - \beta_{uj}| \leqslant c/\log n\}$ (Assumption C1(i) holds). The score function is

$$\psi_{uj}(W_{uj}, \theta, \eta_{uj}) = K_\varpi(W)\{\tau - 1\{X_a \leqslant Z_j^a\theta + Z_{-j}^a\beta_{u,-j} + r_u\}\}f_u(Z_j^a - Z_{-j}^a\gamma_u^j)$$

where the nuisance parameter is $\eta_{uj} = (\beta_{u,-j}, \gamma_u^j, r_u)$ and the last component is a function $r_u = r_u(X)$. Recall that $K_\varpi(W) \in \{0,1\}$ and let $a_n = \max(n, p, |V|)$. Define the nuisance parameter set $\mathcal{H}_{uj} = \{\eta = (\eta^{(1)}, \eta^{(2)}, \eta^{(3)}) : \|\eta - \eta_{uj}\|_e \leqslant \tau_n\}$ where $\|\eta - \eta_{uj}\|_e = \|(\delta_\eta^{(1)}, \delta_\eta^{(2)}, \delta_\eta^{(3)})\|_e = \max\{\|\delta_\eta^{(1)}\|, \|\delta_\eta^{(2)}\|, \mathrm{E}[|\delta_\eta^{(3)}|^2]^{1/2}\}$, and

$$\tau_n := C\sup_{u\in\mathcal{U}} \frac{1}{1 \wedge \underline{f}_{\mathcal{U}}} \sqrt{\frac{s\log a_n}{n\mu_{\mathcal{W}}}}$$

The differentiability of the mapping $(\theta, \eta) \in \Theta_{uj} \times \mathcal{H}_{uj} \mapsto \mathrm{E}\psi_{uj}(W_{uj}, \theta, \eta)$ follows from the differentiability of the conditional probability distribution of $X_a$ given $X_{V\setminus\{a\}}$ and $\varpi$. Let $\eta = (\eta^{(1)}, \eta^{(2)}, \eta^{(3)})$, $\delta_\eta = (\delta_\eta^{(1)}, \delta_\eta^{(2)}, \delta_\eta^{(3)})$, and $\theta_{\bar{r}} = \theta + \bar{r}\delta_\theta$, $\eta_{\bar{r}} = \eta + \bar{r}\delta_\eta$.

To verify Assumption C1(v)(a) with $\alpha = 2$, for any $(\theta, \eta), (\bar{\theta}, \bar{\eta}) \in \Theta_{uj} \times \mathcal{H}_{uj}$ note that $f_{X_a|X_{-a},\varpi}$ is uniformly bounded from above by $\bar{f}$, therefore

$$\begin{aligned}
&\mathrm{E}[\{\psi_{uj}(W_{uj}, \theta, \eta) - \psi_{uj}(W_{uj}, \bar{\theta}, \bar{\eta})\}^2]^{1/2} \\
&\leqslant \bar{f}\mathrm{E}[|Z_{-j}^a(\eta^{(2)} - \bar{\eta}^{(2)})|^2]^{1/2} + \bar{f}^2\mathrm{E}[(Z_j^a - Z_{-j}^a\bar{\eta}^{(2)})^2\{|\eta^{(3)} - \bar{\eta}^{(3)}| + |Z_{-j}^a(\eta^{(1)} - \bar{\eta}^{(1)})| + |Z_j^a(\theta - \bar{\theta})|\}]^{1/2} \\
&\leqslant C\|\eta^{(2)} - \bar{\eta}^{(2)}\| + \bar{f}\mathrm{E}[(Z_j^a - Z_{-j}^a\bar{\eta}^{(2)})^4]^{1/4}\{\mathrm{E}[|\eta^{(3)} - \bar{\eta}^{(3)}|^2]^{1/4} + C\|\eta^{(1)} - \bar{\eta}^{(1)}\| + |\theta - \bar{\theta}|\}^{1/2} \\
&\leqslant C'|\theta - \bar{\theta}|^{1/2} \vee \|\eta - \bar{\eta}\|_e^{1/2}
\end{aligned}$$

for some constance $C' < \infty$ since by Condition CI we have $\mathrm{E}[|Z^a\bar{\xi}|^2]^{1/2} \leqslant C\|\bar{\xi}\|$ for all vectors $\bar{\xi}$, and the conditions $\sup_{u\in\mathcal{U}, j\in[p]} \|\gamma_u^j\| \leqslant C$, $\sup_{\theta\in\Theta_{uj}} |\theta| \leqslant C$, and $\sqrt{s\log(a_n)} \leqslant \delta_n\sqrt{n}$. This implies that $\|\eta^{(2)} - \bar{\eta}^{(2)}\| \leqslant \|\eta^{(2)} - \eta_{uj}^{(2)}\| + \|\eta_{uj}^{(2)} - \bar{\eta}^{(2)}\| \leqslant 1$ so that $\|\eta^{(2)} - \bar{\eta}^{(2)}\| \leqslant \|\eta^{(2)} - \bar{\eta}^{(2)}\|^{1/2}$.

To verify Assumption C1(v)(b), let $t_{\bar{r}} = Z_j^a\theta_{\bar{r}} + Z_{-j}^a\eta_{\bar{r}}^{(1)} + \eta_{\bar{r}}^{(3)}$. We have

$$\begin{aligned}
&\partial_r \mathrm{E}(\psi_{uj}(W_{uj}, \theta + r\delta_\theta, \eta + r\delta_\eta))|_{r=\bar{r}} = \\
&-\mathrm{E}[K_\varpi(W)f_{X_a|X_{-a},\varpi}(t_{\bar{r}})(Z_j^a - Z_{-j}^a\eta_{\bar{r}}^{(2)})\{Z_j^a\delta_\theta + Z_{-j}^a\delta_\eta^{(1)} + \delta_\eta^{(3)}\}] \\
&-\mathrm{E}[K_\varpi(W)\{\tau - F_{X_a|X_{-a},\varpi}(t_{\bar{r}})\}Z_{-j}^a\delta_\eta^{(2)}]
\end{aligned}$$

Applying Cauchy-Schwartz we have that

$$\begin{aligned}&\left|\partial_r \mathrm{E}(\psi_{uj}(W_{uj}, \theta + r\delta_\theta, \eta + r\delta_\eta))|_{r=\bar{r}}\right| \\ &\leqslant \bar{f}\mathrm{E}[(Z_j^a - Z_{-j}^a \eta_{\bar{r}}^{(2)})^2]^{1/2}\{\mathrm{E}[(Z_j^a)^2]^{1/2}|\delta_\theta| + \mathrm{E}[(Z_{-j}^a \delta_\eta^{(1)})^2]^{1/2} + \mathrm{E}[|\delta_\eta^{(3)}|^2]^{1/2}\} + \bar{f}\mathrm{E}[(Z_{-j}^a \delta_\eta^{(2)})^2]^{1/2} \\ &\leqslant \bar{B}_{1n}(|\delta_\theta| \vee \|\eta - \eta_{uj}\|_e)\end{aligned}$$

where $\bar{B}_{1n} \leqslant C$ by the same arguments of bounded (second) moments of linear combinations.

Assumption C1(v)(c) follows similarly as

$$\begin{aligned}&\partial_r^2 \mathrm{E}(\psi_{uj}(W_{uj}, \theta + r\delta_\theta, \eta + r\delta_\eta))\big|_{r=\bar{r}} = \\ &-\mathrm{E}[K_\varpi(W) f'_{X_a|X_{-a},\varpi}(t_{\bar{r}})(Z_j^a - Z_{-j}^a \eta_{\bar{r}}^{(2)})\{Z_j^a \delta_\theta + Z_{-j}^a \delta_\eta^{(1)} + \delta_\eta^{(3)}\}^2] \\ &+2\mathrm{E}[K_\varpi(W) f_{X_a|X_{-a},\varpi}(t_{\bar{r}})(Z_{-j}^a \delta_\eta^{(2)})\{Z_j^a \delta_\theta + Z_{-j}^a \delta_\eta^{(1)} + \delta_\eta^{(3)}\}]\end{aligned}$$

and under $|f'_{X_a|X_{-a},\varpi}| \leqslant \bar{f}'$, from Cauchy-Schwartz inequality we have

$$\begin{aligned}&\left|\partial_r^2 \mathrm{E}(\psi_{uj}(W_{uj}, \theta + r\delta_\theta, \eta + r\delta_\eta))\big|_{r=\bar{r}}\right| \\ &\leqslant |\bar{f}'_n \mathrm{E}[(Z_j^a - Z_{-j}^a \eta_{\bar{r}}^{(2)})^2]^{1/2}\{\mathrm{E}[(Z_j^a)^4]|\delta_\theta|^2 + \mathrm{E}[(Z_{-j}^a \delta_\eta^{(1)})^4]^{1/2}\} + C\mathrm{E}[\{\delta_\eta^{(3)}\}^2] \\ &+2\bar{f}\mathrm{E}[(Z_{-j}^a \delta_\eta^{(2)})^2]^{1/2}\{\mathrm{E}[(Z_j^a)^2]^{1/2}|\delta_\theta| + \mathrm{E}[(Z_{-j}^a \delta_\eta^{(1)})^2]^{1/2} + \mathrm{E}[\{\delta_\eta^{(3)}\}^2]^{1/2}\} \\ &\leqslant \bar{B}_{2n}(\delta_\theta^2 \vee \|\eta - \eta_{uj}\|_e^2)\end{aligned}$$

where $\bar{B}_{2n} \leqslant C(1 + \bar{f}'_n)$ by the same arguments of bounded (fourth) moments as before and using that $|\mathrm{E}[(Z_j^a - Z_{-j}^a \eta_{\bar{r}}^{(2)})(\delta_\eta^{(3)})^2]| \leqslant \{\mathrm{E}[(Z_j^a - Z_{-j}^a \eta_{\bar{r}}^{(2)})^2 (\delta_\eta^{(3)})^2]\}^{1/2} \mathrm{E}[(\delta_\eta^{(3)})^2]^{1/2} \leqslant C\mathrm{E}[(\delta_\eta^{(3)})^2]$.

To verify the near orthogonality condition, note that for all $u \in \mathcal{U}$ and $j \in [p]$, since by definition $f_u = f_{X_a|X_{-a},\varpi}(Z^a \beta_u + r_u)$ we have

$$|\mathrm{D}_{u,j,0}[\tilde{\eta}_{uj} - \eta_{uj}]| = |-\mathrm{E}[K_\varpi(W) f_u \{Z_{-j}^a (\tilde{\eta}^{(2)} - \eta_{uj}^{(2)}) + r_u\} v_{uj}]| \leqslant \delta_n n^{-1/2}$$

by the relations $\mathrm{E}[K_\varpi(W)(\tau - F_{X_a|X_{-a},\varpi}(Z^a \beta_u + r_u)) Z_{-j}^a] = 0$ and $\mathrm{E}[K_\varpi(W) f_u Z_{-j}^a v_{uj}] = 0$ implied by the model, and $|\mathrm{E}[K_\varpi(W) f_u r_u v_{uj}]| \leqslant \delta_n n^{-1/2}$ by Condition CI. Thus, condition (H.79) holds.

Furthermore, since $\Theta_{uj} \subset \theta_{uj} \pm C/\log n$, for $J_{uj} = \partial_\theta \mathrm{E}[\psi_{uj}(W_{uj}, \theta_{uj}, \eta_{uj})] = \mathrm{E}[K_\varpi(W) f_u Z_j^a v_{uj}] = \mathrm{E}[K_\varpi(W) v_{uj}^2] = \mathrm{E}[v_{uj}^2 | \varpi]\mathrm{P}(\varpi)$ as $\mathrm{E}[K_\varpi(W) f_u Z_{-j}^a v_{uj}] = 0$, we have that for all $\theta \in \Theta_{uj}$

$$\mathrm{E}[\psi_{uj}(W_{uj}, \theta, \eta_{uj})] = J_{uj}(\theta - \theta_{uj}) + \frac{1}{2}\partial_\theta^2 \mathrm{E}[\psi_{uj}(W_{uj}, \bar{\theta}, \eta_{uj})](\theta - \theta_{uj})^2$$

where $|\partial_\theta^2 \mathrm{E}[\psi_{uj}(W_{uj}, \bar{\theta}, \eta_{uj})]| \leqslant \bar{f}'\mathrm{E}[|Z_j^a|^2 |v_{uj}| \mid \varpi]\mathrm{P}(\varpi) \leqslant \bar{f}'\mathrm{E}[|Z_j^a|^4 | \varpi]^{1/2}\mathrm{E}[|v_{uj}|^2 | \varpi]^{1/2}\mathrm{P}(\varpi) \leqslant C\bar{f}'\mathrm{P}(\varpi)$ so that for all $\theta \in \Theta_{uj}$

$$|\mathrm{E}[\psi_{uj}(W_{uj}, \theta, \eta_{uj})]| \geqslant \{|\mathrm{E}[v_{uj}^2 \mid \varpi]| - (C^2 \bar{f}')/\log n\}\mathrm{P}(\varpi)|\theta - \theta_{uj}|$$

and we can take $j_n \geqslant c \inf_{\varpi \in \mathcal{W}} \mathrm{P}(\varpi) = c\mu_{\mathcal{W}}$.

Next we verify Assumption C2 with $\mathcal{H}_{ujn} = \{\eta = (\beta, \gamma, 0) : \|\beta\|_0 \leqslant Cs, \|\gamma\|_0 \leqslant Cs, \|\beta - \beta_{u,-j}\| \leqslant C\tau_n, \|\gamma - \gamma_u^j\| \leqslant C\tau_n, \|\gamma - \gamma_u^j\|_1 \leqslant C\sqrt{s}\tau_n\}$. We will show that $\widehat{\eta}_{uj} = (\widetilde{\beta}_{u,-j}, \widetilde{\gamma}_u^j, 0) \in \mathcal{H}_{ujn}$ with probability $1 - o(1)$, uniformly over $u \in \mathcal{U}$ and $j \in [p]$.

Under Condition CI and the choice of penalty parameters, by Theorems 1 and 2, with probability $1 - o(1)$, uniformly over $u \in \mathcal{U}$ we have

$$\|\widetilde{\beta}_u - \beta_u\| \leqslant C\tau_n, \quad \max_{j \in [p]} \sup_{u \in \mathcal{U}} \|\widetilde{\gamma}_u^j - \gamma_u^j\| \leqslant C\tau_n, \quad \text{and} \quad \max_{j \in [p]} \sup_{u \in \mathcal{U}} \|\widetilde{\gamma}_u^j\|_0 \leqslant \bar{C}s,$$

further by thresholding we can achieve $\sup_{u\in\mathcal{U}}\|\widetilde{\beta}_u\|_0 \leqslant \bar{C}s$ using Lemma 17.

Next we establish the entropy bounds. For $\eta \in \mathcal{H}_{ujn}$ we have that

$$\psi_{uj}(W_{uj},\theta,\eta) = K_\varpi(W)(\tau - 1\{X_a \leqslant Z_j^a\theta + Z_{-j}^a\beta_{-j}\})f_u\{Z_j^a - Z_{-j}^a\gamma\}$$

It follows that $\mathcal{F}_1 \subset \mathcal{W}\mathcal{G}_1\mathcal{G}_2\mathcal{G}_3 \cup \bar{\mathcal{F}}_0$ where $\bar{\mathcal{F}}_0 = \{\psi_{uj}(W_{uj},\theta,\eta_{uj}) : u\in\mathcal{U}, j\in[p], \theta\in\Theta_{uj}\}$, $\mathcal{G}_1 = \{\tau - 1\{X_a \leqslant Z^a\beta\} : \|\beta\|_0 \leqslant Cs, \tau\in\mathcal{T}, a\in V\}$, $\mathcal{G}_2 = \{Z^a \to Z^a(1,-\gamma), \|\gamma\|_0 \leqslant Cs, \|\gamma\| \leqslant C, a\in V\}$, $\mathcal{G}_3 = \{f_u : u\in\mathcal{U}\}$. Under Condition CI, $\mathcal{W}$ is a VC class of sets with VC index $d_W$ (fixed). It follows that $\mathcal{G}_1$ and $\mathcal{G}_2$ are $p$ choose $O(s)$ VC-subgraph classes with VC indices at most $O(s)$. Therefore, $\mathrm{ent}(\mathcal{G}_1)\vee\mathrm{ent}(\mathcal{G}_2)\vee\mathrm{ent}(\mathcal{W}) \leqslant Cs\log(a_n/\epsilon) + Cd_W\log(e/\epsilon)$ by Theorem 2.6.7 in [72] and by standard arguments. Also, since $f_u$ is Lipschitz in $u$ by Condition CI, we have $\mathrm{ent}(\mathcal{G}_3) \leqslant (1+d_W)\log(a_nL_f/\epsilon)$. Moreover, an envelope $F_G$ for $\mathcal{F}_1$ satisfies

$$\begin{aligned}\mathrm{E}[F_G^q] &= \mathrm{E}[\sup_{u\in\mathcal{U},j\in[p],\|\gamma-\gamma_u^j\|_1\leqslant C\sqrt{s}\tau_n}|v_{uj} - f_uZ_{-j}^a(\gamma-\gamma_u^j)|^q]\\ &\leqslant 2^{q-1}\mathrm{E}[\sup_{u\in\mathcal{U},j\in[p]}|v_{uj}|^q] + 2^{q-1}\bar{f}\mathrm{E}[\max_{a\in V}\|Z^a\|_\infty^q]\{C\sqrt{s}\tau_n\}^q\\ &\leqslant 2^{q-1}L_n^q + 2^{q-1}\bar{f}\{M_nC\sqrt{s}\tau_n\}^q \leqslant 2^qL_n^q\end{aligned}$$

since $M_nC\sqrt{s}\tau_n \leqslant \delta_nL_n/\bar{f}$ and $\delta_n \leqslant 1$ for $n$ large.

Next we bound the entropy in $\bar{\mathcal{F}}_0$. Note that for any $\psi_{uj}(W_{uj},\theta,\eta_{uj}) \in \bar{\mathcal{F}}_0$, there is some $\delta\in[-C,C]$ such that

$$\psi_{uj}(W_{uj},\theta,\eta_{uj}) = K_\varpi(W)\{\tau - 1\{X_a \leqslant Z_j^a\delta + Q_{X_a}(\tau \mid X_{-a},\varpi)\}\}v_{uj}$$

and therefore $\bar{\mathcal{F}}_0 \subset \mathcal{W}\{\mathcal{T} - \phi(\mathcal{V})\}\mathcal{L}$ where $\phi(t) = 1\{t\leqslant 0\}$, $\mathcal{V} = \cup_{a\in V,j\in[p]}\mathcal{V}_{aj}$ with

$$\mathcal{V}_{aj} := \{X_a - Z_j^a\delta - Q_{X_a}(\tau \mid X_{-a},\varpi) : \tau\in\mathcal{T}, \varpi\in\mathcal{W}, |\delta| \leqslant C\},$$

and $\mathcal{L} = \cup_{a\in V,j\in[p]}(\mathcal{L}_{aj} + \{v_{\bar{u}j}\})$ where $\mathcal{L}_{aj} = \{(X,W)\mapsto v_{uj} - v_{\bar{u}j} = f_uZ_{-j}^a(\gamma_u^j - \gamma_{\bar{u}}^j) : u\in\mathcal{U}\}$. Note that each $\mathcal{V}_{aj}$ is a VC subgraph class of functions with index $1+Cd_W$ as $\{Q_{X_a}(\tau|X_{-a},\varpi) : (\tau,\varpi)\in\mathcal{W}\times\mathcal{T}\}$ is a VC-subgraph with VC-dimension $Cd_W$ for every $a\in V$. Since $\phi$ is monotone, $\phi(\mathcal{V})$ is also the union of VC-dimension of order $1+Cd_W$.

Letting $F_1 = 1$ be an envelope for $\mathcal{W}$ and $\mathcal{T}-\phi(\mathcal{V})$. By Lemma 5, it follows that $\|\gamma_u^j - \gamma_{u'}^j\| \leqslant L_\gamma\{\|u-u'\| + \|\varpi-\varpi'\|^{1/2}\}$ for some $L_\gamma$ satisfying $\log(L_\gamma) \leqslant C\log(p|V|n)$ under Condition CI. Therefore, $|v_{uj} - v_{\bar{u}j}| = |Z_{-j}^a(\gamma_u^j - \gamma_{\bar{u}}^j)| \leqslant \|Z^a\|_\infty\sqrt{p}\|\gamma_u^j - \gamma_{\bar{u}}^j\|$. For a choice of envelope $F_a = M_n^{-1}\|Z^a\|_\infty + 2\sup_{u\in\mathcal{U}}|v_{uj}|$ which satisfies $\|F_a\|_{P,q} \lesssim L_n$, we have

$$\begin{aligned}\log N(\epsilon\|F_a\|_{Q,2}, \mathcal{L}_{aj}, \|\cdot\|_{Q,2}) &\leqslant \log N(\tfrac{\epsilon}{M_n}\|\,\|Z^a\|_\infty\|_{Q,2}, \mathcal{L}_{aj}, \|\cdot\|_{Q,2})\\ &\leqslant \log N(\epsilon/\{M_n\sqrt{p}L_\gamma\}, \mathcal{U}, |\cdot|) \leqslant Cd_u\log(M_npL_\gamma/\epsilon)\end{aligned}$$

Since $\mathcal{L} = \cup_{a\in V,j\in[p]}(\mathcal{L}_{aj} + \{v_{\bar{u}j}\})$, taking $F_L = \max_{a\in V}F_a$, we have that

$$\begin{aligned}\log N(\epsilon\|F_LF_1\|_{Q,2}, \bar{\mathcal{F}}_0, \|\cdot\|_{Q,2}) &\leqslant \log N(\tfrac{\epsilon}{4}\|F_1\|_{Q,2}, \mathcal{W}, \|\cdot\|_{Q,2}) + \log N(\tfrac{\epsilon}{4}\|F_1\|_{Q,2}, \mathcal{T}-\phi(\mathcal{V}), \|\cdot\|_{Q,2})\\ &\quad + \log\textstyle\sum_{a\in V,j\in[p]} N(\tfrac{\epsilon}{2}\|F_a\|_{Q,2}, \mathcal{L}_{aj}, \|\cdot\|_{Q,2})\\ &\leqslant \log(p|V|) + 1 + C'\{d_W + d_u\}\log(4eM_n|V|pL_\gamma/\epsilon)\end{aligned}$$

where the last line follows from the previous bounds.

Next we verify the growth conditions in Assumption C2 with the proposed $\mathcal{F}_1$ and $K_n \lesssim CL_n$. We take $s_{n(\mathcal{U},p)} = (1+d_W)s$ and $a_n = \max\{n,p,|V|\}$. Recall that $\bar{B}_{1n} \leqslant C$, $\bar{B}_{2n} \leqslant C$, $j_n \geqslant c\mu_{\mathcal{W}}$. Thus, we have $\sqrt{n}(\tau_n/j_n)^2 \lesssim \sqrt{n}\frac{s\log(p|V|n)}{n(1\wedge\underline{f}_{\mathcal{U}}^2)\mu_{\mathcal{W}}^3} \leqslant \delta_n$ under $s^2\log^2(p|V|n) \leqslant n(1\wedge\underline{f}_{\mathcal{U}}^4)\mu_{\mathcal{W}}^6$. Moreover,

$(\tau_n/j_n)^{\alpha/2}\sqrt{s_{n(\mathcal{U},p)}\log(a_n)} \lesssim \sqrt[4]{\frac{(1+d_W)^3 s^3\log^3(p|V|n)}{n(1\wedge \underline{f}_{\mathcal{U}}^2)\mu_{\mathcal{W}}^3}} \lesssim \delta_n$ under $d_W$ fixed and $s^3\log^3(p|V|n) \leqslant \delta_n^4 n(1\wedge \underline{f}_{\mathcal{U}}^2)\mu_{\mathcal{W}}^3$ and $s_{n(\mathcal{U},p)}n^{-\frac{1}{2}}K_n\log(a_n)\log n \lesssim (1+d_W)sn^{\frac{1}{q}-\frac{1}{2}}M_n\log(p|V|n)\log n \leqslant \delta_n$ under our conditions. Finally, the conditions of Corollary 4 hold with $\rho_n = (1+d_W)$ since the score is the product of VC-subgraph classes of function with VC index bounded by $C(1+d_W)$. ∎

*Proof of Theorem 4.* We will invoke Lemma 7 with $\bar\beta_u$ as the estimand and $r_{iu} = X'_{i,-a}(\beta_u - \bar\beta_u)$, therefore $\mathrm{E}[K_\varpi(W)(\tau - 1\{X_a \leqslant X'_{-a}\bar\beta_u + r_u\})X_{-a}] = 0$. To invoke the lemma we verify that the events $\Omega_1$, $\Omega_2$, $\Omega_3$ and $\Omega_4$ hold with probability $1-o(1)$

$$\begin{aligned}
\Omega_1 &:= \{\lambda_u \geqslant c|S_{uj}|/\widehat\sigma^X_{a\varpi j}, \text{ for all } u\in\mathcal{U}, j\in V\},\\
\Omega_2 &:= \{\widehat R_u(\bar\beta_u) \leqslant \bar R_{u\xi} : u\in\mathcal{U}\}\\
\Omega_3 &:= \left\{\sup_{u\in\mathcal{U},1/\sqrt{n}\leqslant\|\delta\|_{1,\varpi}\leqslant\sqrt{n}} |\mathbb{E}_n[g_u(\delta,X,W) - \mathrm{E}[g_u(\delta,X,W)\mid X_{-a},W]]|/\|\delta\|_{1,\varpi} \leqslant t_3\right\}\\
\Omega_4 &:= \{K_u\widehat\sigma^X_{a\varpi j} \geqslant |\mathbb{E}_n[h_{uj}(X_{-a},W)]|, \text{for all } u\in\mathcal{U}, j\in V\backslash\{a\}\}
\end{aligned}$$

where $g_u(\delta,X,W) = K_\varpi(W)\{\rho_\tau(X_a - X'_{-a}(\bar\beta_u+\delta)) - \rho_\tau(X_a - X'_{-a}\bar\beta_u)\}$, $h_{uj}(X_{-a},W) = \mathrm{E}[K_\varpi(W)\{\tau - F_{X_a|X_{-a},W}(X'_{-a}\bar\beta_u + r_u)\}X_j \mid X_{-a},W]$.

By Lemma 8 with $\xi = 1/n$, by setting $\lambda_u = \lambda_0 = c2(1+1/16)\sqrt{2\log(8|V|^2\{ne/d_W\}^{2d_W}n)/n}$, we have $\mathrm{P}(\Omega_1) = 1-o(1)$. By Lemma 9, setting $\bar R_{u\xi} = Cs(1+d_W)\log(|V|n)/n$ we have $\mathrm{P}(\Omega_2) = 1-o(1)$ for some $\xi = o(1)$. By Lemma 10 we have $\mathrm{P}(\Omega_3) = 1-o(1)$ by setting $t_3 := C\sqrt{(1+d_W)\log(|V|nM_n/\xi)}$. Finally, by Lemma 11 with $K_u = C\sqrt{\frac{(1+d_W)\log(|V|n)}{n}}$ we have $\mathrm{P}(\Omega_4) = 1-o(1)$.

Moreover, we have that $\|\bar\beta_u\|_{1,\varpi} \leqslant \sqrt{s}\|\bar\beta_u\|_{2,\varpi} \leqslant C\sqrt{s} = o(\sqrt{n})$ and $\frac{1}{\lambda_u(1-1/c)}\bar R_{u\xi} = o(\sqrt{n})$ for all $u\in\mathcal{U}$. Finally, we verify condition (F.64) holds for all

$$\delta \in A_u := \Delta_{u,2\mathbf{c}} \cup \{v : \|v\|_{1,\varpi} \leqslant 2\mathbf{c}\bar R_{u\xi}/\lambda_u, \|\sqrt{f_u}X'_{-a}v\|_{n,\varpi} \geqslant C\sqrt{s(1+d_W)\log(n|V|)/n}/\kappa_{u,2\mathbf{c}}\},$$

$$\bar q_{A_u}/4 \geqslant (\sqrt{\bar f}+1)\|r_u\|_{n,\varpi} + [\lambda_u + t_3 + K_u]\tfrac{3\mathbf{c}\sqrt{s}}{\kappa_{u,2\mathbf{c}}} \text{ and } \bar q_{A_u} \geqslant \{2\mathbf{c}\left(1+\tfrac{t_3+K_u}{\lambda_u}\right)\bar R_{u\gamma}\}^{1/2}.$$

Consider the matrices $\mathbb{E}_n[K_\varpi(W)f_uX_{-a}X'_{-a}]$ and $\mathrm{E}[K_\varpi(W)f_uX_{-a}X'_{-a}]$. By Lemma 6, with probability $1-o(1)$, it follows that we can take $\eta = \eta_n = CM_n\sqrt{s(1+d_W)\log(|V|n)}\log(1+s)\{\log n\}/\sqrt{n}$ and $D_{kk} = 2\eta$ in Lemma 16. (Note that we increase $\delta_n$ by a factor of $\sqrt{\log n}$.) Therefore, with at least the same probability we have (taking $s\geqslant 2$)

$$\delta'\mathbb{E}_n[K_\varpi(W)f_uX_{-a}X'_{-a}]\delta \geqslant \delta'\mathrm{E}[K_\varpi(W)f_uX_{-a}X'_{-a}]\delta - 4\eta\|\delta\|_1^2/s \tag{C.46}$$

and by definition of $\underline{f}_u$ we have

$$\mathbb{E}_n[K_\varpi(W)f_u|X'_{-a}\delta|^2] \geqslant \underline{f}_u\mathrm{E}[K_\varpi(W)|X'_{-a}\delta|^2] - 4\eta\|\delta\|_1^2/s \geqslant c\underline{f}_u\mathrm{P}(\varpi)\|\delta\|^2 - 4\eta\|\delta\|_1^2/s.$$

For $\delta\in\Delta_{u,2\mathbf{c}}$ we have $\|\delta\|_1 \leqslant C\|\delta\|_{1,\varpi}/\{\mathrm{P}(\varpi)\}^{1/2} \leqslant C'\|\delta_{T_u}\|_{1,\varpi}/\{\mathrm{P}(\varpi)\}^{1/2} \leqslant C'\sqrt{s}\|\delta_{T_u}\|_2$. Note that we can assume $\|\delta\| \geqslant c\sqrt{s(1+d_W)\log(n|V|)/n}$ otherwise we are done. So that for $\delta\in A_u\backslash\Delta_{u,2\mathbf{c}}$ we have that $\|\delta\|_1/\|\delta\|_2 \leqslant Cs\sqrt{\log(|V|n)/n}/\sqrt{s(1+d_W)\log(n|V|)/n} \leqslant C'\sqrt{s}$.

Similarly we have

$$\mathbb{E}_n[K_\varpi(W)|X'_{-a}\delta|^2] \leqslant \mathrm{E}[K_\varpi(W)|X'_{-a}\delta|^2] + 4\eta\|\delta\|_1^2/s \leqslant C\mathrm{P}(\varpi)\|\delta\|^2 - 4\eta\|\delta\|_1^2/s. \tag{C.47}$$

Under the condition that $\eta = o(\underline{f}_{\mathcal{U}}\mu_{\mathcal{W}})$, which holds by Condition P, for $n$ sufficiently large we have with probability $1-o(1)$ that

$$\begin{aligned}\bar{q}_{A_u} &\geqslant \frac{c}{\bar{f}'}\inf_{\delta\in A_u}\frac{\mathbb{E}_n[K_\varpi(W)f_u|X'_{-a}\delta|^2]^{3/2}}{\mathbb{E}_n[K_\varpi(W)|X'_{-a}\delta|^3]} \geqslant \frac{c}{\bar{f}'}\inf_{\delta\in A_u}\frac{\mathbb{E}_n[K_\varpi(W)f_u|X'_{-a}\delta|^2]^{3/2}}{\mathbb{E}_n[K_\varpi(W)|X'_{-a}\delta|^2]\max_{i\leqslant n}\|X_i\|_\infty\|\delta\|_1}\\ &\geqslant \frac{c}{\bar{f}'}\inf_{\delta\in A_u}\frac{\{c'\underline{f}_u\mathrm{P}(\varpi)\|\delta\|^2\}^{3/2}}{C'\mathrm{P}(\varpi)\|\delta\|^2\max_{i\leqslant n}\|X_i\|_\infty\|\delta\|_1} \geqslant \frac{c}{\bar{f}'}\inf_{\delta\in A_u}\frac{c'\underline{f}_u^{3/2}\mathrm{P}(\varpi)^{1/2}\|\delta\|}{C'\max_{i\leqslant n}\|X_i\|_\infty\|\delta\|_1}\\ &\geqslant C''\frac{\underline{f}_{\mathcal{U}}^{3/2}}{\bar{f}'}\frac{\mu_{\mathcal{W}}^{1/2}}{\sqrt{s}\max_{i\leqslant n}\|X_i\|_\infty}\end{aligned} \tag{C.48}$$

where $\max_{i\leqslant n}\|X_i\|_\infty \leqslant \ell_n M_n$ with probability $1-o(1)$ for any $\ell_n\to\infty$. Therefore, under the condition $M_n s\sqrt{\log(p|V|n)} = o(\sqrt{n\mu_{\mathcal{W}}})$ assumed in Condition P, the conditions on $\bar{q}_{A_u}$ are satisfied.

By Lemma 7, we have uniformly over all $u=(a,\tau,\varpi)\in\mathcal{U}:=V\times\mathcal{T}\times\mathcal{W}$

$$\|\sqrt{f_u}X'_{-a}(\widehat{\beta}_u-\beta_u)\|_{n,\varpi} \leqslant C\sqrt{\frac{(1+d_W)\log(n|V|)}{n}}\frac{\sqrt{s}}{\kappa_{u,2\mathbf{c}}} \quad\text{and}\quad \|\widehat{\beta}_u-\beta_u\|_{1,\varpi} \leqslant C\sqrt{\frac{(1+d_W)\log(n|V|)}{n}}\frac{s}{\kappa_{u,2\mathbf{c}}}$$

where $\kappa_{u,2\mathbf{c}}$ is bounded away from zero with probability $1-o(1)$ for $n$ sufficiently large. Consider the thresholded estimators $\widehat{\beta}_u^{\bar{\lambda}}$ for $\bar{\lambda}=\{(1+d_W)\log(n|V|)/n\}^{1/2}$. By Lemma 17 we have $\|\widehat{\beta}_u^{\bar{\lambda}}\|_0\leqslant Cs$ and the same rates of convergence as $\widehat{\beta}_u$. Therefore, by refitting over the support of $\widehat{\beta}_u^{\bar{\lambda}}$ we have by Lemma 14, the estimator $\widetilde{\beta}_u$ has the same rate of convergence where we used that $\widehat{Q}_u \leqslant \lambda_u\|\widehat{\beta}_u^{\bar{\lambda}}-\beta_u\|_{1,\varpi} \lesssim Cs(1+d_W)\log(|V|n)/n$ (the other conditions of Lemma 14 hold as for the conditions in Lemma 7).

Next we will invoke Lemma 7 for the new penalty choice and penalty loadings. (We note that minor modifications cover the new penalty loadings.)

$$\begin{aligned}\Omega_1 &:= \{\lambda_u \geqslant c|S_{uj}|/\{\mathbb{E}_n[K_w(W)\varepsilon_u^2X_{-a,j}^2]\}^{1/2},\ \text{for all}\ \ u\in\mathcal{U}, j\in V\},\\ \Omega_2 &:= \{\widehat{R}_u(\bar{\beta}_u)\leqslant \bar{R}_{u\xi} : u\in\mathcal{U}\}\\ \Omega_3 &:= \Big\{\sup_{u\in\mathcal{U},1/\sqrt{n}\leqslant\|\delta\|_{1,\varpi}\leqslant\sqrt{n}}|\mathbb{E}_n[g_u(\delta,X,W)-\mathrm{E}[g_u(\delta,X,W)|X_{-a},W]]|/\{\theta_u\|\delta\|_{1,\varpi}\}\leqslant t_3\Big\}\\ \Omega_4 &:= \{K_u\theta_u\widehat{\sigma}^X_{a\varpi j}\geqslant|\mathbb{E}_n[h_{uj}(X_{-a},W)]|, \text{for all } u\in\mathcal{U}, j\in V\backslash\{a\}\}\\ \Omega_5 &:= \{\theta_u\geqslant\max_{j\in V}\widehat{\sigma}^X_{a\varpi j}/\{\mathbb{E}_n[K_\varpi(W)\varepsilon_u^2X_j^2]\}^{1/2}\}\end{aligned}$$

where event $\Omega_5$ simply makes the relevant norms equivalent, $\|\cdot\|_{1,u}\leqslant\|\cdot\|_{1,\varpi}\leqslant\theta_u\|\cdot\|_{1,u}$. Note that we can always take $\theta_u\leqslant 1/\{\tau(1-\tau)\}\leqslant C$ since $\mathcal{T}$ is a fixed compact set.

Next we show that the bootstrap approximation of the score provides a valid choice of penalty parameter. Let $\widehat{\varepsilon}_u := 1\{X_a\leqslant X'_{-a}\widetilde{\beta}_u\}-\tau$. For notational convenience for $u\in\mathcal{U}$, $j\in V\backslash\{a\}$ define

$$\bar{\psi}_{iuj} = \frac{K_\varpi(W_i)\varepsilon_{iu}X_{ij}}{\mathrm{E}[K_\varpi(W)\varepsilon_u^2X_j^2]^{1/2}},\quad \psi_{iuj} = \frac{K_\varpi(W_i)\varepsilon_{iu}X_{ij}}{\mathbb{E}_n[K_\varpi(W)\varepsilon_u^2X_j^2]^{1/2}},\quad \widehat{\psi}_{iuj} = \frac{K_\varpi(W_i)\widehat{\varepsilon}_{iu}X_{ij}}{\mathbb{E}_n[K_\varpi(W)\widehat{\varepsilon}_u^2X_j^2]^{1/2}}.$$

We will consider the following processes:

$$\bar{S}_{uj}=\tfrac{1}{\sqrt{n}}\textstyle\sum_{i=1}^n\bar{\psi}_{iuj},\quad S_{uj}=\tfrac{1}{\sqrt{n}}\sum_{i=1}^n\psi_{iuj},\quad \overline{\mathcal{G}}_{uj}=\tfrac{1}{\sqrt{n}}\sum_{i=1}^n g_i\bar{\psi}_{iuj},\quad \widehat{\mathcal{G}}_{uj}=\frac{1}{\sqrt{n}}\sum_{i=1}^n g_i\widehat{\psi}_{iuj},$$

and $\mathcal{N}$ is a tight zero-mean Gaussian process with covariance operator given by $\mathrm{E}[\bar{\psi}_{uj}\bar{\psi}_{u'j'}]$. Their supremum are denoted by $\bar{Z}_S := \sup_{u\in\mathcal{U},j\in V\backslash\{a\}}|\bar{S}_{uj}|$, $Z_S := \sup_{u\in\mathcal{U},j\in V\backslash\{a\}}|S_{uj}|$, $\bar{Z}_G^* := \sup_{u\in\mathcal{U},j\in V\backslash\{a\}}|\overline{\mathcal{G}}_{uj}|$, $\widehat{Z}_G^* := \sup_{u\in\mathcal{U},j\in V\backslash\{a\}}|\widehat{\mathcal{G}}_{uj}|$, and $Z_N := \sup_{u\in\mathcal{U},j\in V\backslash\{a\}}|\mathcal{N}_{uj}|$.

The penalty choice should majorate $Z_S$ and we simulate via $\widehat{Z}_G^*$. We have that

$$\begin{aligned}|\mathrm{P}(Z_S \leqslant t) - \mathrm{P}(\widehat{Z}_G^* \leqslant t)| \quad &\leqslant |\mathrm{P}(Z_S \leqslant t) - \mathrm{P}(\bar{Z}_S \leqslant t)| + |\mathrm{P}(\bar{Z}_S \leqslant t) - \mathrm{P}(Z_N \leqslant t)| \\ &+|\mathrm{P}(Z_N \leqslant t) - \mathrm{P}(\bar{Z}_G^* \leqslant t)| + |\mathrm{P}(\bar{Z}_G^* \leqslant t) - \mathrm{P}(\widehat{Z}_G^* \leqslant t)|\end{aligned}$$

We proceed to bound each term. We have that

$$\begin{aligned}|Z_S - \bar{Z}_S| \quad &\leqslant \bar{Z}_S \sup_{u\in\mathcal{U}, j\in V\setminus\{a\}} \left| \frac{\mathrm{E}[K_\varpi(W)\varepsilon_u^2 X_j^2]^{1/2}}{\mathbb{E}_n[K_\varpi(W)\varepsilon_u^2 X_j^2]^{1/2}} - 1 \right| \\ &\leqslant \bar{Z}_S \sup_{u\in\mathcal{U}, j\in V\setminus\{a\}} \left| \frac{(\mathbb{E}_n - \mathrm{E})[K_\varpi(W)\varepsilon_u^2 X_j^2]}{\mathbb{E}_n[K_\varpi(W)\varepsilon_u^2 X_j^2]^{1/2}\{\mathbb{E}_n[K_\varpi(W)\varepsilon_u^2 X_j^2]^{1/2} + \mathrm{E}[K_\varpi(W)\varepsilon_u^2 X_j^2]^{1/2}\}} \right|\end{aligned}$$

Therefore, since $\{1\{X_a \leqslant X'_{-a}\beta_u\} : u \in \mathcal{U}\}$ is a VC-subgraph of VC dimension $1 + d_W$, and $\mathcal{W}$ is a VC class of sets of dimension $d_W$, we apply Lemma 19 with envelope $F = \|X\|_\infty^2$ and $\sigma^2 \leqslant \max_{j\in V} \mathrm{E}[X_j^4] \leqslant C$ to obtain with probability $1 - o(1)$

$$\sup_{u\in\mathcal{U}, j\in V\setminus\{a\}} |(\mathbb{E}_n - \mathrm{E})[K_\varpi(W)\varepsilon_u^2 X_j^2]| \lesssim \delta'_{1n} := \sqrt{\frac{(1+d_W)\log(|V|n)}{n}} + \frac{M_n^2(1+d_W)\log(|V|n)}{n}$$

where $\delta_{1n} = o(\mu_{\mathcal{W}}^2)$ under Condition P. Note that this implies that the denominator above is bounded away from zero by $c\mu_{\mathcal{W}}$. Therefore,

$$|Z_S - \bar{Z}_S| \lesssim_P \delta_{1n} := \bar{Z}_S \delta'_{1n}/\mu_{\mathcal{W}}.$$

where $\bar{Z}_S \lesssim_P \{(1+d_W)\log(n|V|)\}^{1/2}$. By Theorem 2.1 in [27], since $\mathrm{E}[\bar{\psi}_{uj}^4] \leqslant C$, there is a version of $Z_N$ such that

$$|\bar{Z}_S - Z_N| \lesssim_P \delta_{2n} := \left( \frac{M_n(1+d_W)\log(n|V|)}{n^{1/2}} + \frac{M_n^{1/3}((1+d_W)\log(n|V|))^{2/3}}{n^{1/6}} \right)$$

and by Theorem 2.2 in [27], there is also a version of

$$|Z_N - \bar{Z}_G^*| \lesssim_P \left( \frac{M_n(1+d_W)\log(n|V|)}{n^{1/2}} + \frac{M_n^{1/2}((1+d_W)\log(n|V|))^{3/4}}{n^{1/4}} \right)$$

Finally, we have that

$$|\bar{Z}_G^* - \widehat{Z}_G^*| \leqslant \sup_{u\in\mathcal{U}, j} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n g_i(\widehat{\psi}_{iuj} - \bar{\psi}_{iuj}) \right|$$

where conditional on $(X_i, W_i), i = 1, \ldots, n$, $\frac{1}{\sqrt{n}}\sum_{i=1}^n g_i(\widehat{\psi}_{iuj} - \bar{\psi}_{iuj})$ is a zero-mean Gaussian with variance $\mathbb{E}_n[(\widehat{\psi}_{iuj} - \bar{\psi}_{iuj})^2] \leqslant \bar{\delta}_n^2$. Next we bound $\bar{\delta}_n$. We have

$$\bar{\delta}_n \quad \leqslant \mathbb{E}_n[(\widehat{\psi}_{uj} - \psi_{uj})^2]^{1/2} + \mathbb{E}_n[(\psi_{uj} - \bar{\psi}_{uj})^2]^{1/2} \leqslant \mathbb{E}_n[(\widehat{\psi}_{uj} - \psi_{uj})^2]^{1/2} + \delta_{1n}/\mu_{\mathcal{W}},$$

and

$$\begin{aligned}\mathbb{E}_n[(\widehat{\psi}_{uj} - \psi_{uj})^2]^{1/2} \quad &\leqslant \tfrac{\mathbb{E}_n[(K_\varpi(W)X_{ij}|\widehat{\varepsilon}_u - \varepsilon_u|)^2]^{1/2}}{\mathbb{E}_n[K_\varpi(W)X_{ij}^2\widehat{\varepsilon}_u^2]^{1/2}} \\ &+ \tfrac{\mathbb{E}_n[K_\varpi(W)X_{ij}^2\varepsilon_u^2]^{1/2}}{c\mathrm{P}(\varpi)} |\mathbb{E}_n[K_\varpi(W)X_{ij}^2\widehat{\varepsilon}_u^2]^{1/2} - \mathbb{E}_n[K_\varpi(W)X_{ij}^2\varepsilon_u^2]^{1/2}| \\ &\leqslant \mathbb{E}_n[K_\varpi(W)X_{ij}^2|\widehat{\varepsilon}_u - \varepsilon_u|^2]^{1/2} \left\{ \tfrac{1}{\mathbb{E}_n[K_\varpi(W)X_{ij}^2\widehat{\varepsilon}_u^2]^{1/2}} + \tfrac{\mathbb{E}_n[K_\varpi(W)X_{ij}^2\varepsilon_u^2]^{1/2}}{c\mathrm{P}(\varpi)} \right\},\end{aligned}$$

note that the term in the curly brackets is bounded by $C/\mathrm{P}(\varpi)^{1/2}$ with probability $1 - o(1)$. To bound the other term note that $|\widehat{\varepsilon}_u - \varepsilon_u|^2 = |1\{X_a \leqslant X'_{-a}\widetilde{\beta}_u\} - 1\{X_a \leqslant X'_{-a}\beta_u\}|$. Note that $\widehat{\varepsilon}_u = 1\{X_a \leqslant X'_{-a}\widetilde{\beta}_u\} - \tau$

where $\|\widetilde{\beta}_u\|_0 \leqslant Cs$. Therefore, we have $\{1\{X_a \leqslant X'_{-a}\widetilde{\beta}_u\} : u \in \mathcal{U}\} \subset \{1\{X_a \leqslant X'_{-a}\beta\} : \|\beta\|_0 \leqslant Cs\}$ which is the union of $\binom{|V|}{Cs}$ VC subgraph classes of functions with VC dimension $C's$. Moreover, we have

$$\begin{aligned}
\mathrm{E}[K_\varpi(W)X_{ij}^2|\widehat{\varepsilon}_u - \varepsilon_u|^2] &= \mathrm{E}[K_\varpi(W)X_{ij}^2|1\{X_a \leqslant X'_{-a}\widetilde{\beta}_u\} - 1\{X_a \leqslant X'_{-a}\beta_u\}|] \\
&\leqslant \bar{f}\mathrm{E}[K_\varpi(W)X_{ij}^2|X'_{-a}(\widetilde{\beta}_u - \beta_u)|] \\
&\leqslant \bar{f}\mathrm{E}[K_\varpi(W)X_{ij}^4]^{1/2}\mathrm{E}[K_\varpi(W)|X'_{-a}(\widetilde{\beta}_u - \beta_u)|^2]^{1/2} \\
&\leqslant C(\bar{f}/\underline{f}_{\mathcal{U}}^{1/2})\mathrm{P}(\varpi)^{1/2}\sqrt{s(1+d_W)\log(n|V|)/n}
\end{aligned}$$

Therefore, by Lemma 19, with probability $1 - o(1)$ we have

$$\left|(\mathbb{E}_n - \mathrm{E})[K_\varpi(W)X_{ij}^2|\widehat{\varepsilon}_u - \varepsilon_u|^2]\right| \lesssim \sqrt{\frac{s(1+d_W)\log(n|V|)}{n}}C(\bar{f}/\underline{f}_{\mathcal{U}}^{1/2})\sqrt{s(1+d_W)\log(n|V|)/n}$$

Under $\sqrt{s(1+d_W)\log(n|V|)/n} = o(\underline{f}_{\mathcal{U}}\mu_{\mathcal{W}})$ we have that with probability $1 - o(1)$ that

$$\bar{\delta}_n \leqslant C\{s(1+d_W)\log(n|V|)/n\}^{1/4}.$$

Therefore, using again the sparsity of $\widetilde{\beta}_u$ in the definition of $\widehat{\psi}_{iuj}$

$$\begin{aligned}
\sup_{u\in\mathcal{U}, j\in V\setminus\{a\}}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n g_i(\widehat{\psi}_{iuj} - \bar{\psi}_{iuj})\right| &\lesssim_P \bar{\delta}_n\sqrt{s(1+d_W)\log(|V|n)} \\
&\lesssim_P \delta_{3n} := \{s\log(|V|n)/n\}^{1/4}\sqrt{s(1+d_W)\log(|V|n)}
\end{aligned}$$

The rest of the proof follows similarly to Corollary 2.2 in [14] since under Condition P (and the bounds above) we have that $r_n := \delta_{1n} + \delta_{2n} + \delta_{3n} = o(\{\mathrm{E}[Z_N]\}^{-1}))$ where $\mathrm{E}[Z_N] \lesssim \{(1+d_W)\log(|V|n)\}^{1/2}$. Then we have $\sup_t |\mathrm{P}(Z_S \leqslant t) - \mathrm{P}(\widehat{Z}_G^* \leqslant t)| = o_P(1)$ which in turn implies that

$$\begin{aligned}
\mathrm{P}(\Omega_1) &= \mathrm{P}(Z_S \leqslant \widehat{c}_G^*(\xi)) \\
&\geqslant \mathrm{P}(\widehat{Z}_G^* \leqslant \widehat{c}_G^*(\xi)) - |\mathrm{P}(Z_S \leqslant \widehat{c}_G^*(\xi)) - \mathrm{P}(\widehat{Z}_G^* \leqslant \widehat{c}_G^*(\xi))| \\
&\geqslant 1 - \xi + o_P(1)
\end{aligned}$$

Note that the occurrence of the events $\Omega_2$, $\Omega_3$ and $\Omega_4$ follows by similar arguments. The result follows by Lemma 7, thresholding and applying Lemma 17 and Lemma 14 similarly to before.

■

## Appendix D. Technical Lemmas for Conditional Independence Quantile Graphical Model

Let $u = (a, \tau, \varpi) \in \mathcal{U} := V \times \mathcal{T} \times \mathcal{W}$, and $T_u = \mathrm{support}(\beta_u)$ where $|T_u| \leqslant s$ for all $u \in \mathcal{U}$.

Define the pseudo-norms

$$\|v\|_{n,\varpi}^2 := \frac{1}{n}\sum_{i=1}^n K_\varpi(W_i)(v_i)^2, \quad \|\delta\|_{2,\varpi} := \left\{\sum_{j=1}^p \{\widehat{\sigma}_{a\varpi j}^Z\}^2|\delta_j|^2\right\}^{1/2}, \quad \text{and} \quad \|\delta\|_{1,\varpi} := \sum_{j=1}^p \widehat{\sigma}_{a\varpi j}^Z|\delta_j|,$$

where $\widehat{\sigma}_{a\varpi j}^Z = \{\mathbb{E}_n[\{K_\varpi(W)Z_j^a\}^2]\}^{1/2}$. These pseudo-norms induce the following restricted eigenvalue as

$$\kappa_{u,\mathbf{c}} = \min_{\|\delta_{T_u^c}\|_{1,\varpi} \leqslant \mathbf{c}\|\delta_{T_u}\|_{1,\varpi}} \frac{\|\sqrt{f_u}Z^a\delta\|_{n,\varpi}}{\|\delta\|_{1,\varpi}/\sqrt{s}}.$$

The restricted eigenvalue $\kappa_{u,\mathbf{c}}$ is an counterpart of the restricted eigenvalue proposed in [21] for our setting. We note that $\kappa_{u,\mathbf{c}}$ typically will vary with the events $\varpi \in \mathcal{W}$.

We will consider three key events in our analysis. Let

$$\Omega_1 := \{\lambda_u \geqslant c|S_{uj}|/\widehat{\sigma}^Z_{a\varpi j}, \text{ for all } u \in \mathcal{U}, j \in [p]\} \tag{D.49}$$

which occurs with probability at least $1-\xi$ by the choice of $\lambda_u$. For CIQGMs, we have $S_{uj} := \mathbb{E}_n[K_\varpi(W)(\tau - 1\{X_a \leqslant Z^a\beta_u + r_u\})Z^a_j]$, and $\lambda_u = \lambda_{V\mathcal{T}\mathcal{W}}\sqrt{\tau(1-\tau)}$. (In the case of PQGMs, we have $S_{uj} := \mathbb{E}_n[K_\varpi(W)(\tau - 1\{X_a \leqslant X'_{-a}\beta_u\})X_{-a}]$, $\widehat{\sigma}^X_{a\varpi j} = \{\mathbb{E}_n[K_\varpi(W)X^2_{-a,j}]\}^{1/2}$ and $\lambda_u = \lambda_0$.)

To define the next event, for each $u \in \mathcal{U}$, consider the function defined as

$$\widehat{R}_u(\beta_u) = \mathbb{E}_n[K_\varpi(W)\{\rho_u(X_a - Z^a\beta) - \rho_u(X_a - Z^a\beta_u - r_u) - (\tau - 1\{X_a \leqslant Z^a\beta_u + r_u\})(Z^a\beta - Z^a\beta_u - r_u)\}]$$

in the case of CIQGMs. (In the case of PQGMs, we replace $Z^a$ with $X_{-a}$.) By convexity we have $\widehat{R}_u(\beta_u) \geqslant 0$. The event

$$\Omega_2 := \{\widehat{R}_u(\beta_u) \leqslant \bar{R}_{u\xi} : u \in \mathcal{U}\} \tag{D.50}$$

where $\bar{R}_{u\xi}$ are chosen so that $\Omega_2$ occurs with probability at least $1-\xi$. Note that by Lemma 2, we have $\mathbb{E}_n[\mathrm{E}[\widehat{R}_u(\beta_u)|X_{-a}, W]] \leqslant \bar{f}\|r_u\|^2_{n,\varpi}/2$ and with probability at least $1-\xi$, $\widehat{R}_u(\beta_u) \leqslant \bar{R}_{u\xi} := 4\max\{\bar{f}\|r_u\|^2_{n,\varpi},\ \|r_u\|_{n,\varpi}C\sqrt{\log(n^{1+d_W}p/\xi)/n}\} \leqslant C's\log(n^{1+d_W}p/\xi)/n$.

Define $g_u(\delta, X, W) = K_\varpi(W)\{\rho_\tau(X_a - Z^a(\beta_u + \delta)) - \rho_\tau(X_a - Z^a\beta_u)\}$ so that event $\Omega_3$ is defined as

$$\Omega_3 := \left\{\sup_{u\in\mathcal{U}, 1/\sqrt{n}\leqslant\|\delta\|_{1,\varpi}\leqslant\sqrt{n}} \frac{|\mathbb{E}_n[g_u(\delta, X, W) - \mathrm{E}[g_u(\delta, X, W)|X_{-a}, W]]|}{\|\delta\|_{1,\varpi}} \leqslant t_3\right\} \tag{D.51}$$

where $t_3$ is given in Lemma 3 so that $\Omega_3$ holds with probability at least $1-\xi$.

**Lemma 1.** *Suppose that $\Omega_1$, $\Omega_2$ and $\Omega_3$ holds. Further assume $2\frac{1+1/c}{1-1/c}\|\beta_u\|_{1,\varpi} + \frac{1}{\lambda_u(1-1/c)}\bar{R}_{u\xi} \leqslant \sqrt{n}$ for all $u \in \mathcal{U}$, and (F.64) holds for all $\delta \in A_u := \Delta_{u,2\mathbf{c}} \cup \{v : \|v\|_{1,\varpi} \leqslant 2\mathbf{c}\bar{R}_{u\xi}/\lambda_u\}$, $\bar{q}_{A_u}/4 \geqslant (\sqrt{\bar{f}}+1)\|r_u\|_{n,\varpi} + [\lambda_u + t_3]\frac{3\mathbf{c}\sqrt{s}}{\kappa_{u,2\mathbf{c}}}$ and $\bar{q}_{A_u} \geqslant \{2\mathbf{c}\left(1 + \frac{t_3}{\lambda_u}\right)\bar{R}_{u\xi}\}^{1/2}$. Then uniformly over all $u \in \mathcal{U}$ we have*

$$\begin{aligned}
\|\sqrt{f_u}Z^a(\widehat{\beta}_u - \beta_u)\|_{n,\varpi} &\leqslant \sqrt{8\mathbf{c}\left(1 + \tfrac{t_3}{\lambda_u}\right)\bar{R}_{u\xi}} + (\bar{f}^{1/2}+1)\|r_u\|_{n,\varpi} + \tfrac{3\mathbf{c}\lambda_u\sqrt{s}}{\kappa_{u,2\mathbf{c}}} + t_3\tfrac{(1+\mathbf{c})\sqrt{s}}{\kappa_{u,2\mathbf{c}}} \\
\|\widehat{\beta}_u - \beta_u\|_{1,\varpi} &\leqslant (1+2\mathbf{c})\sqrt{s}\|\sqrt{f_u}Z^a\delta_u\|_{n,\varpi}/\kappa_{u,2\mathbf{c}} + \tfrac{2\mathbf{c}}{\lambda_u}\bar{R}_{u\xi}
\end{aligned}$$

*Proof of Lemma 1.* Let $u = (a, \tau, \varpi) \in \mathcal{U}$ and $\delta_u = \widehat{\beta}_u - \beta_u$. By convexity and definition of $\widehat{\beta}_u$ we have

$$\begin{aligned}
&\widehat{R}_u(\widehat{\beta}_u) - \widehat{R}_u(\beta_u) + S_u'\delta_u \\
&= \mathbb{E}_n[K_\varpi(W)\rho_u(X_a - Z^a\widehat{\beta}_u)] - \mathbb{E}_n[K_\varpi(W)\rho_u(X_a - Z^a\beta_u)] \\
&\leqslant \lambda_u\|\beta_u\|_{1,\varpi} - \lambda_u\|\widehat{\beta}_u\|_{1,\varpi}
\end{aligned} \tag{D.52}$$

where $S_u$ is defined as in (D.49) so that under $\Omega_1$ we have $\lambda_u \geqslant c|S_{uj}|/\widehat{\sigma}^Z_{a\varpi j}$.

Under $\Omega_1 \cap \Omega_2$, and since $\widehat{R}_u(\widehat{\beta}_u) \geqslant 0$, we have

$$\begin{aligned}
-\widehat{R}_u(\beta_u) - \tfrac{\lambda_u}{c}\|\delta_u\|_{1,\varpi} &\leqslant \widehat{R}_u(\beta_u + \delta_u) - \widehat{R}_u(\beta_u) + \mathbb{E}_n[K_\varpi(W)(\tau - 1\{X_a \leqslant Z^a\beta_u + r_u\})Z^a\delta_u] \\
&= \mathbb{E}_n[K_\varpi(W)\rho_u(X_a - Z^a(\delta_u + \beta_u))] - \mathbb{E}_n[K_\varpi(W)\rho_u(X_a - Z^a\beta_u)] \\
&\leqslant \lambda_u\|\beta_u\|_{1,\varpi} - \lambda_u\|\delta_u + \beta_u\|_{1,\varpi}
\end{aligned} \tag{D.53}$$

so that for $\mathbf{c} = (c+1)/(c-1)$

$$\|\delta_{T_u^c}\|_{1,\varpi} \leqslant \mathbf{c}\|\delta_{T_u}\|_{1,\varpi} + \frac{c}{\lambda_u(c-1)}\widehat{R}_u(\beta_u).$$

To establish that $\delta_u \in A_u := \Delta_{u,2\mathbf{c}} \cup \{v : \|v\|_{1,\varpi} \leqslant 2\mathbf{c}\bar{R}_{u\xi}/\lambda_u\}$ we consider two cases. If $\|\delta_{T_u^c}\|_{1,\varpi} \geqslant 2\mathbf{c}\|\delta_{T_u}\|_{1,\varpi}$ we have

$$\frac{1}{2}\|\delta_{T_u^c}\|_{1,\varpi} \leqslant \frac{c}{\lambda_u(c-1)}\widehat{R}_u(\beta_u)$$

and consequentially

$$\|\delta_u\|_{1,\varpi} \leqslant \{1+1/(2c)\}\|\delta_{T_u^c}\|_{1,\varpi} \leqslant \frac{2\mathbf{c}}{\lambda_u}\widehat{R}_u(\beta_u).$$

Otherwise, we have $\|\delta_{T_u^c}\|_{1,\varpi} \leqslant 2\mathbf{c}\|\delta_{T_u}\|_{1,\varpi}$ which implies

$$\|\delta_u\|_{1,\varpi} \leqslant (1+2\mathbf{c})\|\delta_{T_u}\|_{1,\varpi} \leqslant (1+2\mathbf{c})\sqrt{s}\|\sqrt{f_u}Z^a\delta_u\|_{n,\varpi}/\kappa_{u,2\mathbf{c}}$$

by definition of $\kappa_{u,2\mathbf{c}}$. Thus we have $\delta_u \in A_u$ under $\Omega_1 \cap \Omega_2$.

Furthermore, (D.53) also implies that

$$\begin{aligned}\|\delta_u+\beta_u\|_{1,\varpi} &\leqslant \|\beta_u\|_{1,\varpi} + \tfrac{1}{c}\|\delta_u\|_{1,\varpi} + \widehat{R}_u(\beta_u)/\lambda_u \\ &\leqslant (1+1/c)\|\beta_u\|_{1,\varpi} + (1/c)\|\delta_u+\beta_u\|_{1,\varpi} + \widehat{R}_u(\beta_u)/\lambda_u.\end{aligned}$$

which in turn establishes

$$\|\delta_u\|_{1,\varpi} \leqslant 2\frac{1+1/c}{1-1/c}\|\beta_u\|_{1,\varpi} + \frac{1}{\lambda_u(1-1/c)}\widehat{R}_u(\beta_u) \leqslant 2\frac{1+1/c}{1-1/c}\|\beta_u\|_{1,\varpi} + \frac{1}{\lambda_u(1-1/c)}\bar{R}_{u\xi}$$

where the last inequality holds under $\Omega_2$. Thus, $\|\delta_u\|_{1,\varpi} \leqslant \sqrt{n}$ under our condition. In turn, $\delta_u$ is considered in the supremum that defines $\Omega_3$.

Under $\Omega_1 \cap \Omega_2 \cap \Omega_3$ we have

$$\begin{aligned}&\mathbb{E}_n[\mathrm{E}[K_\varpi(W)\{\rho_u(X_a - Z^a(\beta_u+\delta_u)) - \rho_u(X_a - Z^a\beta_u)\} \mid X_{-a}, W] \\ &\leqslant \mathbb{E}_n[K_\varpi(W)\{\rho_u(X_a - Z^a(\beta_u+\delta_u)) - \rho_u(X_a - Z^a\beta_u)\}] + t_3\|\delta_u\|_{1,\varpi} \\ &\leqslant \lambda_u\|\delta_u\|_{1,\varpi} + t_3\|\delta_u\|_{1,\varpi} \\ &\leqslant 2\mathbf{c}\left(1+\tfrac{1}{\lambda_u}t_3\right)\bar{R}_{u\xi} + \|\sqrt{f_u}Z^a\delta_u\|_{n,\varpi}\left[\lambda_u + t_3\right]\tfrac{3\mathbf{c}\sqrt{s}}{\kappa_{u,2\mathbf{c}}}\end{aligned} \tag{D.54}$$

here we used the bound $\|\delta_u\|_{1,\varpi} \leqslant (1+2\mathbf{c})\sqrt{s}\|\sqrt{f_u}Z^a\delta_u\|_{n,\varpi}/\kappa_{u,2\mathbf{c}} + \frac{2\mathbf{c}}{\lambda_u}\bar{R}_{u\xi}$ under $\Omega_1 \cap \Omega_2$.

Using Lemma 12, since (F.64) holds, we have for each $u \in \mathcal{U}$

$$\begin{aligned}&\mathbb{E}_n[\mathrm{E}[K_\varpi(W)\{\rho_u(X_a - Z^a(\beta_u+\delta_u)) - \rho_u(X_a - Z^a\beta_u)\} \mid X_{-a}, W] \\ \geqslant\ &-(\sqrt{\bar{f}}+1)\|r_u\|_{n,\varpi}\|\sqrt{f_u}Z^a\delta_u\|_{n,\varpi} - \sup_{u\in\mathcal{U}, j\in[p]}|\mathbb{E}_n[\mathrm{E}[S_{uj}|X_{-a},W]/\widehat{\sigma}^Z_{a\varpi j}]|\ \|\delta_u\|_{1,\varpi} \\ &+ \tfrac{\|\sqrt{f_u}Z^a\delta_u\|^2_{n,\varpi}}{4} \wedge \{\bar{q}_{A_u}\|\sqrt{f_u}Z^a\delta_u\|_{n,\varpi}\}\end{aligned}$$

here we have $\mathrm{E}[S_{iuj}|X_{i,-a}, W_i] = 0$ since $\tau = \mathrm{P}(X_a \leqslant Z^a\beta_u + r_u|X_{-a}, W)$ by the definition of conditional quantile.

Note that for positive numbers $(t^2/4)\wedge qt \leqslant A + Bt$ implies $t^2/4 \leqslant A + Bt$ provided $q/2 > B$ and $2q^2 > A$. (Indeed, otherwise $(t^2/4) \geqslant qt$ so that $t \geqslant 4q$ which in turn implies that $2q^2 + qt/2 \leqslant (t^2/4)\wedge qt \leqslant A + Bt$.) Since $\bar{q}_{A_u}/4 \geqslant (\sqrt{\bar{f}}+1)\|r_u\|_{n,\varpi} + \left[\{\lambda_u + t_3\}\frac{3\mathbf{c}\sqrt{s}}{\kappa_{u,2\mathbf{c}}}\right]$ and $\bar{q}_{A_u} \geqslant \{2\mathbf{c}\left(1+\frac{t_3}{\lambda_u}\right)\bar{R}_{u\xi}\}^{1/2}$, the minimum on the right hand side is achieved by the quadratic part for all $u \in \mathcal{U}$. Therefore we have uniformly over $u \in \mathcal{U}$

$$\frac{\|\sqrt{f_u}Z^a\delta_u\|^2_{n,\varpi}}{4} \leqslant 2\mathbf{c}\left(1+\frac{t_3}{\lambda_u}\right)\bar{R}_{u\xi} + \|\sqrt{f_u}Z^a\delta_u\|_{n,\varpi}\left[(\sqrt{\bar{f}}+1)\|r_u\|_{n,\varpi} + \{\lambda_u+t_3\}\frac{3\mathbf{c}\sqrt{s}}{\kappa_{u,2\mathbf{c}}}\right]$$

which implies that

$$\|\sqrt{f_u}Z^a\delta_u\|_{n,\varpi} \leqslant \sqrt{8\mathbf{c}\left(1+\tfrac{t_3}{\lambda_u}\right)\bar{R}_{u\xi}} + \left[(\sqrt{\bar{f}}+1)\|r_u\|_{n,\varpi} + \{\lambda_u+t_3\}\tfrac{3\mathbf{c}\sqrt{s}}{\kappa_{u,2\mathbf{c}}}\right].$$

■

**Lemma 2** (CIQGM, Event $\Omega_2$)**.** *Under Condition CI we have* $\mathbb{E}_n[\mathrm{E}[\widehat{R}_u(\beta_u)|X_{-a},\varpi]] \leqslant \bar{f}\|r_u\|^2_{n,\varpi}/2$, $\widehat{R}_u(\beta_u) \geqslant 0$ *and*

$$\mathrm{P}\left(\sup_{u\in\mathcal{U}} \widehat{R}_u(\beta_u) \leqslant C\{1+\bar{f}\}\{n^{-1}s(1+d_W)\log(p|V|n)\}\right) = 1-o(1).$$

*Proof of Lemma 2.* We have that $\widehat{R}_u(\beta_u) \geqslant 0$ by convexity of $\rho_\tau$. Let $\varepsilon_{iu} = X_{ia} - Z_i^a\beta_u - r_{iu}$ where $\|\beta_u\|_0 \leqslant s$ and $r_{iu} = Q_{X_a}(\tau|X_{-a},\varpi) - Z^a\beta_u$.

By Knight's identity (F.65), $\widehat{R}_u(\beta_u) = -\mathbb{E}_n[K_\varpi(W)r_u\int_0^1 1\{\varepsilon_u \leqslant -tr_u\} - 1\{\varepsilon_u \leqslant 0\}\,dt] \geqslant 0$.

$$\begin{aligned}\mathbb{E}_n[\mathrm{E}[\widehat{R}_u(\beta_u)|X_{-a},\varpi] &= \mathbb{E}_n[K_\varpi(W)r_u\textstyle\int_0^1 F_{X_a|X_{-a},\varpi}(Z^a\beta_u + (1-t)r_u) - F_{X_a|X_{-a},\varpi}(Z^a\beta_u + r_u)\,dt]\\ &\leqslant \mathbb{E}_n[K_\varpi(W)r_u\textstyle\int_0^1 \bar{f}tr_u dt] \leqslant \bar{f}\|r_u\|^2_{n,\varpi}/2 \leqslant C\bar{f}s/n.\end{aligned}$$

Since Condition CI assumes $\mathrm{E}[\|r_u\|^2_{n,\varpi}] \leqslant \mathrm{P}(\varpi)s/n$, by Markov's inequality we have $\mathrm{P}(\widehat{R}_u(\beta_u) \leqslant C\bar{f}s/n) \geqslant 1/2$. Define $z_{iu} := -\int_0^1 1\{\varepsilon_{iu} \leqslant -tr_{iu}\} - 1\{\varepsilon_{iu}\leqslant 0\}\,dt$, so that $\widehat{R}_u(\beta_u) = \mathbb{E}_n[K_\varpi(W)r_uz_u]$ where $|z_{iu}| \leqslant 1$. By Lemma 2.3.7 in [73] (note that the Lemma does not require zero mean stochastic processes), for $t \geqslant 2C\bar{f}s/n$ we have

$$\frac{1}{2}\mathrm{P}\left(\sup_{u\in\mathcal{U}}|\mathbb{E}_n[K_\varpi(W)r_uz_u]| \geqslant t\right) \leqslant 2\mathrm{P}\left(\sup_{u\in\mathcal{U}}|\mathbb{E}_n[\varepsilon K_\varpi(W)r_uz_u]| > t/4\right)$$

where $\varepsilon_i, i=1,\ldots,n$ are Rademacher random variables independent of the data.

Next consider the class of functions $\mathcal{F} = \{-K_\varpi(W)r_u(1\{\varepsilon_{iu}\leqslant -B_ir_{iu}\} - 1\{\varepsilon_{iu}\leqslant 0\}) : u\in\mathcal{U}\}$ where $B_i \sim \mathrm{Uniform}(0,1)$ independent of $(X_i,W_i)_{i=1}^n$. It follows that $K_\varpi(W)r_uz_u = \mathrm{E}[-K_\varpi(W)r_u(1\{\varepsilon_{iu}\leqslant -B_ir_{iu}\} - 1\{\varepsilon_{iu}\leqslant 0\})|X_i,W_i]$ where the expectation is taken over $B_i$ only. Thus we will bound the entropy of $\overline{\mathcal{F}} = \{\mathrm{E}[f|X,W] : f\in\mathcal{F}\}$ via Lemma 25. Note that $\mathcal{R} := \{r_u = Q_{X_a}(\tau|X_{-a},\varpi) - Z^a\beta_u : u\in\mathcal{U}\}$ where $\mathcal{G} := \{Z^a\beta_u : u\in\mathcal{U}\}$ is contained in the union of at most $|V|\binom{p}{s}$ VC-classes of dimension $Cs$ and $\mathcal{H} := \{Q_{X_a}(\tau|X_{-a},\varpi) : u\in\mathcal{U}\}$ is the union of $|V|$ VC-class of functions of dimension $(1+d_W)$ by Condition CI. Finally note that $\mathcal{E} := \{\varepsilon_{iu} : u\in\mathcal{U}\} \subset \{X_{ia} : a\in V\} - \mathcal{G} - \mathcal{R}$.

Therefore, we have

$$\begin{aligned}\sup_Q \log N(\epsilon\|\bar{F}\|_{Q,2}, \overline{\mathcal{F}}, \|\cdot\|_{Q,2}) &\leqslant \sup_Q \log N((\epsilon/4)^2\|F\|_{Q,2}, \mathcal{F}, \|\cdot\|_{Q,2})\\ &\leqslant \sup_Q \log N(\tfrac{1}{8}(\epsilon^2/16), \mathcal{W}, \|\cdot\|_{Q,2})\\ &\quad + \sup_Q \log N(\tfrac{1}{8}(\epsilon^2/16)\|F\|_{Q,2}, \mathcal{R}, \|\cdot\|_{Q,2})\\ &\quad + \sup_Q \log N(\tfrac{1}{8}(\epsilon^2/16), 1\{\mathcal{E}+\{B\}\mathcal{R}\leqslant 0\} - 1\{\mathcal{E}\leqslant 0\}, \|\cdot\|_{Q,2})\end{aligned}$$

We will apply Lemma 19 with envelope $\bar{F} = \sup_{u\in\mathcal{U}}|K_\varpi(W)r_u|$, so that $\mathrm{E}[\max_{i\leqslant n}\bar{F}_i^2] \leqslant C$, and $\sup_{u\in\mathcal{U}}\mathrm{E}[K_\varpi(W)r_u^2] \leqslant Cs/n =: \sigma^2$ by Condition CI. Thus, we have that with probability $1-o(1)$

$$\sup_{u\in\mathcal{U}}|\mathbb{E}_n[\varepsilon K_\varpi(W)r_uz_u]| \lesssim \sqrt{\frac{s(1+d_W)\log(p|V|n)}{n}}\sqrt{\frac{s}{n}} + \frac{s(1+d_W)\log(p|V|n)}{n} \lesssim \frac{s(1+d_W)\log(p|V|n)}{n}$$

under $M_n\sqrt{s^2/n} \leqslant C$. ∎

**Lemma 3** (CIQGM, Event $\Omega_3$)**.** *For* $u=(a,\tau,\varpi)\in\mathcal{U} := V\times\mathcal{T}\times\mathcal{W}$*, define the function* $g_u(\delta,X,W) = K_\varpi(W)\{\rho_\tau(X_a - Z^a(\beta_u+\delta)) - \rho_\tau(X_a - Z^a\beta_u)\}$*, and the event*

$$\Omega_3 := \left\{\sup_{u\in\mathcal{U},1/\sqrt{n}\leqslant\|\delta\|_{1,\varpi}\leqslant\sqrt{n}} \frac{|\mathbb{E}_n[g_u(\delta,X,W) - \mathrm{E}[g_u(\delta,X,W)\mid X_{-a},W]]|}{\|\delta\|_{1,\varpi}} < t_3\right\}.$$

*Then, under Condition CI we have* $P(\Omega_3) \geqslant 1-\xi$ *for any* $t_3$ *satisfying*

$$t_3\sqrt{n} \geqslant 12 + 16\sqrt{2\log(64|V|p^2n^{3+2d_W}\log(n)L_\beta^{1+d_W/\kappa}M_n/\xi)}$$

*Proof.* We have that $\Omega_3^c := \{\max_{a\in V} A_a \geqslant t_3\sqrt{n}\}$ for

$$A_a := \sup_{(\tau,\varpi)\in\mathcal{T}\times\mathcal{W}, \underline{N}\leqslant\|\delta\|_{1,\varpi}\leqslant\bar{N}} \sqrt{n}\left|\frac{\mathbb{E}_n[g_u(\delta,X,W) - \mathrm{E}[g_u(\delta,X,W)\mid X_{-a},W]]}{\|\delta\|_{1,\varpi}}\right|.$$

Therefore, for $\underline{N} = 1/\sqrt{n}$ and $\bar{N} = \sqrt{n}$ we have by Lemma 13 with $\rho=\kappa$, $L_\eta = L_\beta$, $\tilde{x} = Z^a$

$$\begin{aligned}
\mathrm{P}(\Omega_3^c) &= \mathrm{P}(\max_{a\in V} A_a \geqslant t_3\sqrt{n}) \\
&\leqslant |V|\max_{a\in V}\mathrm{P}(A_a \geqslant t_3\sqrt{n}) \\
&= |V|\max_{a\in V}\mathrm{E}_{X_{-a},W}\{\mathrm{P}(A_a \geqslant t_3\sqrt{n}\mid X_{-a},W)\} \\
&\leqslant |V|\max_{a\in V}\mathrm{E}_{X_{-a},W}\left\{8p|\widehat{\mathcal{N}}|\cdot|\widehat{\mathcal{W}}|\cdot|\widehat{\mathcal{T}}|\exp(-(t_3\sqrt{n}/4-3)^2/32)\right\} \\
&\leqslant \exp(-(t_3\sqrt{n}/4-3)^2/32)|V|64pn^{1+d_W}\log(n)L_\beta\mathrm{E}_{X_{-a}}\left\{\frac{\max_{i\leqslant n}\|Z_i^a\|_\infty^{1+d_W/\kappa}}{\underline{N}^{1+d_W}}\right\} \\
&\leqslant \xi
\end{aligned}$$

by the choice of $t_3$ and noting that $M_n^{(1+d_W/\kappa)/q} \geqslant \mathrm{E}_{X_{-a}}[\max_{i\leqslant n}\|Z_i^a\|_\infty^{1+d_W/\kappa}]$, $1+d_W/\kappa \leqslant q$ and $M_n \geqslant 1$. ∎

**Lemma 4** (CIQGM, Uniform Control of Approximation Error in Auxiliary Equation)**.** *Under Condition CI, with probability* $1-o(1)$ *uniformly over* $u\in\mathcal{U}$ *and* $j\in[p]$ *we have*

$$\mathbb{E}_n[K_\varpi(W)f_u^2\{Z_{-j}^a(\gamma_u^j - \bar{\gamma}_u^j)\}^2] \lesssim \underline{f}_u^2\mathrm{P}(\varpi)\{n^{-1}s\log(p|V|n)\}^{1/2}.$$

*Proof.* Define the class of functions $\mathcal{G} = \cup_{a\in V, j\in[p]}\mathcal{G}_{aj}$ with $\mathcal{G}_{aj} := \{Z_{-j}^a(\gamma_u^j - \bar{\gamma}_u^j) : \tau\in\mathcal{T}, \varpi\in\mathcal{W}\}$. Under Condition CI we have $\sup_{u\in\mathcal{U}}\|\bar{\gamma}_u^j\|_0 \leqslant Cs$, $\sup_{u\in\mathcal{U},j\in[p]}\|\bar{\gamma}_u^j - \gamma_u^j\| \vee \frac{\|\bar{\gamma}_u^j-\gamma_u^j\|_1}{\sqrt{s}} \leqslant \{n^{-1}s\log(p|V|n)\}^{1/2}$. Without loss of generality we can set $\bar{\gamma}_{uk}^j = \gamma_{uk}^j$ for $k\in\mathrm{support}(\bar{\gamma}_u^j)$. Letting $\mathcal{G}_{aj,T} := \{Z_{-j}^a(\gamma_u^j - \gamma_{uT}^j) : \tau\in\mathcal{T}, \varpi\in\mathcal{W}\}$ for $T\subset\{1,\ldots,p\}$, it follows that $\mathcal{G}\subset\cup_{a\in V,j\in[p]}\cup_{|T|\leqslant Cs}\mathcal{G}_{aj,T}$.

By Lemma 5, we have $\|\gamma_u^j - \gamma_{u'}^j\| \leqslant L_\gamma(\|u-u'\| + \|u-u'\|^{1/2})$ for each $a\in V$, $j\in[p]$. (Note that although $\bar{\gamma}_u^j$ might not be Lipschitz in $u$, however, for each $T$, $\gamma_{uT}^j$ satisfies the same Lipschitz relation as $\gamma_u^j$, in fact $\|\bar{\gamma}_{uT}^j - \gamma_{u'T}^j\| \leqslant \|\gamma_u^j - \gamma_{u'}^j\|$ by construction.) Therefore, for each $T$ we have

$$\begin{aligned}
&\|\{Z_{-j}^a(\bar{\gamma}_{uT}^j - \gamma_u^j)\}^2 - \{Z_{-j}^a(\bar{\gamma}_{u'T}^j - \gamma_{u'}^j)\}^2\|_{Q,2} \\
&\leqslant \|Z_{-j}^a(\bar{\gamma}_{uT}^j - \bar{\gamma}_{u'T}^j + \gamma_{u'}^j - \gamma_u^j)Z_{-j}^a(\bar{\gamma}_{uT}^j - \gamma_u^j + \bar{\gamma}_{u'T}^j - \gamma_{u'}^j)\|_{Q,2} \\
&\leqslant \|\|Z_{-j}^a\|_\infty^2\|_{Q,2}\|\bar{\gamma}_{uT}^j - \bar{\gamma}_{u'T}^j + \gamma_{u'}^j - \gamma_u^j\|_1\|\bar{\gamma}_{uT}^j - \gamma_u^j + \bar{\gamma}_{u'T}^j - \gamma_{u'}^j\|_1 \\
&\leqslant 4\|\|Z_{-j}^a\|_\infty^2\|_{Q,2}\sup_{u\in\mathcal{U}}\|\bar{\gamma}_{uT}^j - \gamma_u^j\|_1\sqrt{2p}\|\gamma_u^j - \gamma_{u'}^j\| \\
&\leqslant \|\|Z_{-j}^a\|_\infty^2\|_{Q,2}L'_\gamma(\|u-u'\| + \|u-u'\|^{1/2}).
\end{aligned}$$

where $L'_\gamma = 4\{n^{-1}s^2\log(p|V|n)\}^{1/2}\sqrt{2p}L_\gamma$. Thus, for the envelope $G = \max_{a\in V}\|Z^a\|_\infty^2\sup_{u\in\mathcal{U}}\|\bar{\gamma}_u^j - \gamma_u^j\|_1^2$ that

$$\log N(\epsilon\|G\|_{Q,2}, \mathcal{G}, \|\cdot\|_{Q,2}) \leqslant Cs\log(|V|p) + \log N\left(\epsilon\tfrac{\sup_{u\in\mathcal{U}}\|\bar{\gamma}_u^j-\gamma_u^j\|_1^2}{L'_\gamma}, \mathcal{U}, d_{\mathcal{U}}\right) \leqslant Cs(1+d_W)^2\log(L'_\gamma n/\epsilon).$$

Next define the functions $\mathcal{W}_0 = \{K_\varpi(W)f_u^2 : u\in\mathcal{U}\}$, $\mathcal{W}_1 = \{\mathrm{P}(\varpi)^{-1} : \varpi\in\mathcal{W}\}$ and $\mathcal{W}_2 = \{K_\varpi(W) : \varpi\in\mathcal{W}\}$. We have that $\mathcal{W}_2$ is VC class with VC index $Cd_W$ and $\mathcal{W}_1$ is bounded by $\mu_{\mathcal{W}}^{-1}$ and covering number bounded by $(Cd_W/\{\mu_{\mathcal{W}}\epsilon\})^{1+d_W}$. Finally, since $|K_\varpi(W)f_u^2 - K_{\varpi'}(W)f_{u'}^2| \leqslant K_\varpi(W)K_{\varpi'}(W)|f_u^2 - f_{u'}^2| +$

$\bar{f}^2|K_\varpi(W)-K_{\varpi'}(W)| \leqslant 2\bar{f}L_f\|u-u'\| + \bar{f}^2|K_\varpi(W)-K_{\varpi'}(W)|$, we have $N(\epsilon,\mathcal{U},|\cdot|) \leqslant (C(1+d_W)/\epsilon)^{1+d_W}$. Therefore, using standard bounds we have

$$\log N(\epsilon\|\mu_{\mathcal{W}}^{-1}G\bar{f}\|_{Q,2}, \mathcal{W}_0\mathcal{W}_1\mathcal{W}_2\mathcal{G}, \|\cdot\|_{Q,2}) \lesssim s(1+d_W)^2\log(L'_\gamma L_f n/\epsilon)$$

By Lemma 19 we have that with probability $1-o(1)$ that

$$\begin{aligned}
&\sup_{u\in\mathcal{U},j\in[p]} |(\mathbb{E}_n-\mathrm{E})[f_u^2\{Z_{-j}^a(\gamma_u^j-\bar{\gamma}_u^j)\}^2/\mathrm{P}(\varpi)]| \\
&\lesssim \sqrt{\frac{s(1+d_W)^2\log(p|V|n)\sup_{u\in\mathcal{U}}\mathrm{E}[K_\varpi(W)f_u^4\{Z_{-j}^a(\gamma_u^j-\bar{\gamma}_u^j)\}^4]/\mathrm{P}(\varpi)^2}{n}} + \frac{s(1+d_W)^2M_n^2\mu_{\mathcal{W}}^{-1}\sup_{u\in\mathcal{U}}\|\bar{\gamma}_u^j-\gamma_u^j\|_1^2\log(p|V|n)}{n} \\
&\lesssim \sqrt{\frac{s(1+d_W)^2\log(p|V|n)}{\mu_{\mathcal{W}}n}\frac{s\log(p|V|n)}{n}} + \frac{(1+d_W)^2M_n^2s^2\log(p|V|n)}{n\mu_{\mathcal{W}}}\frac{s\log(p|V|n)}{n} \\
&\lesssim \frac{s\log(p|V|n)}{n}\mu_{\mathcal{W}}\underline{f}_{\mathcal{U}}\left\{\sqrt{\frac{s(1+d_W)^2\log(p|V|n)}{\mu_{\mathcal{W}}^3\underline{f}_{\mathcal{U}}^2n}} + \frac{(1+d_W)^2M_n^2s^2\log(p|V|n)}{n\mu_{\mathcal{W}}^2\underline{f}_{\mathcal{U}}}\right\} \\
&\lesssim \frac{s\log(p|V|n)}{n}\mu_{\mathcal{W}}\underline{f}_{\mathcal{U}}\{\delta_n^{1/2}+\delta_n^2\}
\end{aligned}$$

here we used that $\mathrm{E}[f_u^4\{Z^a\delta\}^4|\varpi] \leqslant \bar{f}^4\mathrm{E}[\{Z^a\delta\}^4|\varpi] \leqslant C\|\delta\|^4$, $\|\bar{\gamma}_u^j-\gamma_u^j\| + s^{-1/2}\|\bar{\gamma}_u^j-\gamma_u^j\|_1 \leqslant \{n^{-1}s\log(p|V|n)\}^{1/2}$, $s(1+d_W)^2\log(p|V|n) \leqslant \delta_n n\underline{f}_{\mathcal{U}}^2\mu_{\mathcal{W}}^3$ and $(1+d_W)M_n s\log^{1/2}(p|V|n) \leqslant \delta_n n^{1/2}\mu_{\mathcal{W}}\underline{f}_{\mathcal{U}}$ by Condition CI. Furthermore, by Condition CI, the result follows from $\mathrm{E}[f_u^2\{Z_{-j}^a(\bar{\gamma}_u^j-\gamma_u^j)\}^2|\varpi] \leqslant C\underline{f}_u^2\|\bar{\gamma}_u^j-\gamma_u^j\|^2 \leqslant C\underline{f}_u^2n^{-1}s\log(p|V|n)$. ∎

**Lemma 5.** *Under Condition CI, for $u=(a,\tau,\varpi)\in\mathcal{U}$ and $u'=(a,\tau',\varpi')\in\mathcal{U}$ we have that*

$$\|\gamma_u^j-\gamma_{u'}^j\| \leqslant \frac{C'}{\underline{f}_{u'}^2\mathrm{P}(\varpi')}\{\bar{f}^2\mathrm{E}[\{K_{\varpi'}(W)-K_\varpi(W)\}^2]^{1/2} + \mathrm{E}[K_\varpi(W)K_{\varpi'}(W)\{f_{u'}^2-f_u^2\}^2]^{1/2}\}.$$

*In particular, we have $\|\gamma_u^j-\gamma_{u'}^j\| \leqslant L_\gamma\{\|\varpi-\varpi'\|^{1/2}+\|u-u'\|\}$ for $L_\gamma = C\{L_f+L_K\}/\{\underline{f}_{\mathcal{U}}^2\mu_{\mathcal{W}}\}$ under $\mathrm{E}[|K_\varpi(W)-K_{\varpi'}(W)|] \leqslant L_K\|\varpi-\varpi'\|$, $K_\varpi(W)K_{\varpi'}(W)|f_{u'}-f_u| \leqslant L_f\|u'-u\|$, and $f_u \leqslant \bar{f} \leqslant C$.*

*Proof.* Let $u=(a,\tau,\varpi)$ and $u'=(a,\tau',\varpi')$. By Condition CI we have

$$\|\gamma_u^j-\gamma_{u'}^j\|^2 \leqslant C\mathrm{E}[\{Z_{-j}^a(\gamma_u^j-\gamma_{u'}^j)\}^2|\varpi'] \leqslant \{C/\mathrm{P}(\varpi')\}\mathrm{E}[K_{\varpi'}(W)\{Z_{-j}^a(\gamma_u^j-\gamma_{u'}^j)\}^2]$$

To bound the last term of the right hand side above, by the definition of $\underline{f}_{u'}$ and Cauchy-Schwarz's inequality we have

$$\begin{aligned}
\underline{f}_{u'}\mathrm{E}[K_{\varpi'}(W)\{Z_{-j}^a(\gamma_u^j-\gamma_{u'}^j)\}^2] &\leqslant \mathrm{E}[K_{\varpi'}(W)f_{u'}\{Z_{-j}^a(\gamma_u^j-\gamma_{u'}^j)\}^2] \\
&\leqslant \{\mathrm{E}[K_{\varpi'}(W)f_{u'}^2\{Z_{-j}^a(\gamma_u^j-\gamma_{u'}^j)\}^2]\ \mathrm{E}[K_{\varpi'}(W)\{Z_{-j}^a(\gamma_u^j-\gamma_{u'}^j)\}^2]\}^{1/2}
\end{aligned}$$

so that $\mathrm{E}[K_{\varpi'}(W)\{Z_{-j}^a(\gamma_u^j-\gamma_{u'}^j)\}^2]^{1/2} \leqslant \{\mathrm{E}[K_{\varpi'}(W)f_{u'}^2\{Z_{-j}^a(\gamma_u^j-\gamma_{u'}^j)\}^2]\}^{1/2}/\underline{f}_{u'}$. Therefore

$$\|\gamma_u^j-\gamma_{u'}^j\|^2 \leqslant \{1/\underline{f}_{u'}\}^2\{C/\mathrm{P}(\varpi)\}\mathrm{E}[K_{\varpi'}(W)f_{u'}^2\{Z_{-j}^a(\gamma_u^j-\gamma_{u'}^j)\}^2]. \tag{D.55}$$

We proceed to bound the last term. The optimality of $\gamma_u^j$ and $\gamma_{u'}^j$ yields

$$\mathrm{E}[K_\varpi(W)f_u^2Z_{-j}^a(Z_j^a-Z_{-j}^a\gamma_u^j)] = 0 \quad\text{and}\quad \mathrm{E}[K_{\varpi'}(W)f_{u'}^2Z_{-j}^a(Z_j^a-Z_{-j}^a\gamma_{u'}^j)] = 0$$

Therefore, we have

$$\begin{aligned}
\mathrm{E}[K_{\varpi'}(W)f_{u'}^2\{Z_{-j}^a(\gamma_u^j-\gamma_{u'}^j)\}Z_{-j}^a] &= -\mathrm{E}[K_{\varpi'}(W)f_{u'}^2\{Z_j^a-Z_{-j}^a\gamma_u^j\}Z_{-j}^a] \\
&= -\mathrm{E}[\{K_{\varpi'}(W)f_{u'}^2-K_\varpi(W)f_u^2\}\{Z_j^a-Z_{-j}^a\gamma_u^j\}Z_{-j}^a]
\end{aligned} \tag{D.56}$$

Multiplying by $(\gamma_u^j - \gamma_{u'}^j)$ both sides of (D.56), we have

$$\begin{aligned}
&\mathrm{E}[K_{\varpi'}(W) f_{u'}^2 \{Z_{-j}^a (\gamma_u^j - \gamma_{u'}^j)\}^2] \\
&\leqslant \mathrm{E}[\{K_{\varpi'}(W) f_{u'}^2 - K_{\varpi}(W) f_u^2\}^2]^{1/2} \{\mathrm{E}[\{Z_j^a - Z_{-j}^a \gamma_u^j\}^2 \{Z_{-j}^a (\gamma_u^j - \gamma_{u'}^j)\}^2]\}^{1/2} \\
&\leqslant \mathrm{E}[\{K_{\varpi'}(W) f_{u'}^2 - K_{\varpi}(W) f_u^2\}^2]^{1/2} C \|\gamma_u^j - \gamma_{u'}^j\|
\end{aligned}$$

by the fourth moment assumption in Condition CI. By Condition CI, $f_u, f_{u'} \leqslant \bar{f}$, and it follows that

$$|K_{\varpi}(W) f_u^2 - K_{\varpi'}(W) f_{u'}^2| \leqslant K_{\varpi}(W) K_{\varpi'}(W) |f_u^2 - f_{u'}^2| + \bar{f}^2 |K_{\varpi}(W) - K_{\varpi'}(W)| \tag{D.57}$$

From (D.55) we obtain

$$\|\gamma_u^j - \gamma_{u'}^j\| \leqslant \frac{C'}{\underline{f}_{u'}^2 \mathrm{P}(\varpi')} \{\bar{f}^2 \mathrm{E}[\{K_{\varpi'}(W) - K_{\varpi}(W)\}^2]^{1/2} + \mathrm{E}[K_{\varpi}(W) K_{\varpi'}(W) \{f_{u'}^2 - f_u^2\}^2]^{1/2}\}.$$

■

**Lemma 6.** *Let $\mathcal{U} = V \times \mathcal{T} \times \mathcal{W}$. Under Condition CI, for $m = 1, 2$, we have*

$$\mathrm{E}\left[\sup_{u \in \mathcal{U}, \|\theta\|_0 \leqslant k, \|\theta\| = 1} |(\mathbb{E}_n - \mathrm{E})[K_{\varpi}(W) f_u^m (Z^a \theta)^2]|\right] \lesssim C \delta_n \sup_{u \in \mathcal{U}, \|\theta\|_0 \leqslant k, \|\theta\| = 1} \{\mathrm{E}[K_{\varpi}(W) f_u^m (Z^a \theta)^2]\}^{1/2}$$

*where $\delta_n = M_n \sqrt{k(1 + d_W) C \log(p|V|n)} \log(1 + k) \sqrt{\log n / n}$. Moreover, under Condition CI, $\delta_n = o(\mu_{\mathcal{W}})$.*

*Proof.* By symmetrization we have

$$\mathrm{E}\left[\sup_{u \in \mathcal{U}, \|\theta\|_0 \leqslant k, \|\theta\| = 1} |(\mathbb{E}_n - \mathrm{E})[K_{\varpi}(W) f_u^m (Z^a \theta)^2]|\right] \leqslant 2\mathrm{E}\left[\sup_{u \in \mathcal{U}, \|\theta\|_0 \leqslant k, \|\theta\| = 1} |\mathbb{E}_n[\varepsilon K_{\varpi}(W) f_u^m (Z^a \theta)^2]|\right]$$

where $\varepsilon_i$ are i.i.d. Rademacher random variables. We have that $|K_{\varpi}(W) f_u^m - K_{\varpi'}(W) f_{u'}^m| \leqslant K_{\varpi}(W) K_{\varpi'}(W) |f_u - f_{u'}| (1 + 2\bar{f}) + \bar{f}^m |K_{\varpi}(W) - K_{\varpi'}(W)|$ for $m = 1, 2$ where $u$ and $u'$ have the same $a \in V$. However, conditional on $\{(W_i, X_i), i = 1, \ldots, n\}$, $\{K_{\varpi}(W_i) : i = 1, \ldots, n, \varpi \in \mathcal{W}\}$ induces at most $n^{d_W}$ different sequences by Corollary 2.6.3 in [72]. This induces (at most) $n^{d_W}$ partitions of $\mathcal{W}$ such that $K_{\varpi}(W) = K_{\varpi'}(W)$ for any $\varpi, \varpi'$ in the same partition given the conditioning. Thus, for such suitable $\varpi'$ we have $|K_{\varpi}(W) f_u^m - K_{\varpi'}(W) f_{u'}^m| \leqslant K_{\varpi}(W) |f_u - f_{u'}| (1 + 2\bar{f})$ for $m = 1, 2$. (Thus it suffices to create a net for each partition.) We can take a cover $\widehat{\mathcal{U}}$ of $V \times \mathcal{T} \times \mathcal{W}$ such that $\|u - u'\| \leqslant \{L_f (1 + 2\bar{f}) nk \max_{i \leqslant n} \|Z_i^a\|_\infty^2\}^{-1}$ so that $|f_u - f_{u'}| (Z^a \theta)^2 \leqslant |f_u - f_{u'}| \|Z^a\|_\infty^2 \|\theta\|_1^2 \leqslant |f_u - f_{u'}| \|Z^a\|_\infty^2 k \|\theta\|^2$ which implies

$$\left|\sup_{u \in \mathcal{U}, \|\theta\|_0 \leqslant k, \|\theta\| = 1} |\mathbb{E}_n[\varepsilon K_{\varpi}(W) f_u^m (Z^a \theta)^2]| - \sup_{u \in \widehat{\mathcal{U}}, \|\theta\|_0 \leqslant k, \|\theta\| = 1} |\mathbb{E}_n[\varepsilon K_{\varpi}(W) f_u^m (Z^a \theta)^2]|\right| \leqslant n^{-1}$$

Consequentially

$$\mathrm{E}\left[\sup_{u \in \mathcal{U}, \|\theta\|_0 \leqslant k, \|\theta\| = 1} |\mathbb{E}_n[\varepsilon K_{\varpi}(W) f_u^m (Z^a \theta)^2]|\right] \leqslant \mathrm{E}\left[\sup_{u \in \widehat{\mathcal{U}}, \|\theta\|_0 \leqslant k, \|\theta\| = 1} |\mathbb{E}_n[\varepsilon K_{\varpi}(W) f_u^m (Z^a \theta)^2]|\right] + \frac{1}{n}$$

where $|\widehat{\mathcal{U}}| \leqslant |V| n^{d_W} \{L_f (1 + 2\bar{f}) nk \max_{i \leqslant n} \|Z_i^a\|_\infty^2\}^{(1 + d_W)}$.

By Lemma 18 with $K = K(W, X) = (1 + \bar{f}^2) \sup_{a \in V} \max_{i \leqslant n} \|Z_i^a\|_\infty$ and

$$\begin{aligned}
\delta_n(W, X) &:= \bar{C} K(W, X) \sqrt{k} \left(\sqrt{\log |\widehat{\mathcal{U}}|} + \sqrt{1 + \log p} + \log k \sqrt{\log(p \vee n)} \sqrt{\log n}\right) / \sqrt{n} \\
&\lesssim K(W, X) \sqrt{k(1 + d_W) C \log(p|V|nK(W, X))} \log(1 + k) \sqrt{\log n} / \sqrt{n}
\end{aligned}$$

so that conditional on $(W, X)$ we have

$$\mathrm{E}\left[\sup_{u\in\widehat{\mathcal{U}},\|\theta\|_0\leqslant k,\|\theta\|=1}|\mathbb{E}_n[\varepsilon K_\varpi(W)f_u^m(Z^a\theta)^2]|\right]\lesssim\delta_n(W,X)\sup_{u\in\widehat{\mathcal{U}},\|\theta\|_0\leqslant k,\|\theta\|=1}\sqrt{\mathbb{E}_n[K_\varpi(W)f_u^m(Z^a\theta)^2]}$$

Therefore,

$$\begin{aligned}
&\mathrm{E}\left[\sup_{u\in\mathcal{U},\|\theta\|_0\leqslant k,\|\theta\|=1}|\mathbb{E}_n[\varepsilon K_\varpi(W)f_u^m(Z^a\theta)^2]|\right]\\
&\leqslant\mathrm{E}_{W,X}\left[\delta_n(W,X)\sup_{u\in\widehat{\mathcal{U}},\|\theta\|_0\leqslant k,\|\theta\|=1}\sqrt{\mathbb{E}_n[K_\varpi(W)f_u^m(Z^a\theta)^2]}\right]+\frac{1}{n}\\
&\leqslant\mathrm{E}_{W,X}[\delta_n^2(W,X)]+\mathrm{E}_{W,X}[\delta_n^2(W,X)]^{1/2}\sup_{u\in\widehat{\mathcal{U}},\|\theta\|_0\leqslant k,\|\theta\|=1}\mathrm{E}[K_\varpi(W)f_u^m(Z^a\theta)^2]^{1/2}+\frac{1}{n}
\end{aligned}$$

Note that for a random variable $A\geqslant 1$, we have that $\mathrm{E}[A^2\sqrt{\log(CA)}]\leqslant\mathrm{E}[A^2]\sqrt{\log(C)}+\mathrm{E}[A^2\sqrt{\log(A)}]\leqslant\mathrm{E}[A^2]\sqrt{\log(C)}+\mathrm{E}[A^{2+1/4}]$. Therefore, under Condition CI, since $q\geqslant 2+1/4$ in the definition of $M_n$, we have

$$\mathrm{E}_{W,X}[\delta_n^2(W,X)]^{1/2}\lesssim M_n\sqrt{k(1+d_W)C\log(p|V|n)}\log(1+k)\sqrt{\log n}/\sqrt{n}.$$

The results follows by setting $\delta_n=\mathrm{E}_{W,X}[\delta_n^2(W,X)]^{1/2}$. ∎

## Appendix E. Results for Prediction Quantile Graphical Models

In the analysis of PQGM we also use the following event for some sequence $(K_u)_{u\in\mathcal{U}}$

$$\Omega_4=\{K_u\widehat{\sigma}_{a\varpi j}^X\geqslant\mathbb{E}_n[\mathrm{E}[K_\varpi(W)(\tau-1\{X_a\leqslant X_{-a}'\beta_u+r_u\})X_{-a}|X_j,W]],\ u\in\mathcal{U},j\in V\backslash\{a\}\}.\tag{E.58}$$

**Lemma 7** (Rate for PQGM)**.** *Suppose that $\Omega_1$, $\Omega_2$, $\Omega_3$ and $\Omega_4$ hold. Further assume $2\frac{1+1/c}{1-1/c}\|\beta_u\|_{1,\varpi}+\frac{1}{\lambda_u(1-1/c)}\bar{R}_{u\xi}\leqslant\sqrt{n}$ for all $u\in\mathcal{U}$, and (F.64) holds for all $\delta\in A_u:=\Delta_{\varpi,2\mathbf{c}}\cup\{v:\|v\|_{1,\varpi}\leqslant 2\mathbf{c}\bar{R}_{u\xi}/\lambda_u\}$, $\bar{q}_{A_u}/4\geqslant(\sqrt{\bar{f}}+1)\|r_u\|_{n,\varpi}+[\lambda_u+t_3+K_u]\frac{3\mathbf{c}\sqrt{s}}{\kappa_{u,2\mathbf{c}}}$ and $\bar{q}_{A_u}\geqslant\{2\mathbf{c}\left(1+\frac{t_3+K_u}{\lambda_u}\right)\bar{R}_{u\xi}\}^{1/2}$. Then uniformly over all $u=(a,\tau,\varpi)\in\mathcal{U}:=V\times\mathcal{T}\times\mathcal{W}$, the $\|\cdot\|_{1,\varpi}$-penalized estimator $\widehat{\beta}_u$ satisfies*

$$\begin{aligned}
\|\sqrt{f_u}X_{-a}'(\widehat{\beta}_u-\beta_u)\|_{n,\varpi}&\leqslant\sqrt{8\mathbf{c}\left(1+\tfrac{t_3}{\lambda_u}\right)\bar{R}_{u\xi}}+(\bar{f}^{1/2}+1)\|r_u\|_{n,\varpi}+[\lambda_u+t_3+K_u]\tfrac{3\mathbf{c}\sqrt{s}}{\kappa_{u,2\mathbf{c}}}\\
\|\widehat{\beta}_u-\beta_u\|_{1,\varpi}&\leqslant(1+2\mathbf{c})\sqrt{s}\|\sqrt{f_u}X_{-a}'\delta_u\|_{n,\varpi}/\kappa_{u,2\mathbf{c}}+\tfrac{2\mathbf{c}}{\lambda_u}\bar{R}_{u\xi}
\end{aligned}$$

*Proof of Lemma 7.* The proof proceeds similarly to the proof of Lemma 1 by defining

$$\begin{aligned}
\widehat{R}_u(\beta)&=\mathbb{E}_n[K_\varpi(W)\{\rho_u(X_a-X_{-a}'\beta)-\rho_u(X_a-X_{-a}'\beta_u-r_u)\}]\\
&\quad-\mathbb{E}_n[K_\varpi(W)\{(\tau-1\{X_a\leqslant X_{-a}'\beta_u+r_u\})(X_{-a}'\beta-X_{-a}'\beta_u-r_u)\}].
\end{aligned}$$

The same argument yields $\delta_u=\widehat{\beta}_u-\beta_u\in A_u:=\Delta_{\varpi,2\mathbf{c}}\cup\{v:\|v\|_{1,\varpi}\leqslant 2\mathbf{c}\bar{R}_{u\xi}/\lambda_u\}$ under $\Omega_1\cap\Omega_2$. (Similarly we also have $\|\delta_u\|_{1,\varpi}\leqslant\sqrt{n}$.) Furthermore, under $\Omega_1\cap\Omega_2\cap\Omega_3$ we have that (D.54) also holds which implies

$$\mathbb{E}_n[\mathrm{E}[K_\varpi(W)\{\rho_u(X_a-X_{-a}'(\beta_u+\delta_u))-\rho_u(X_a-X_{-a}'\beta_u)\}|X_{-a},W]\leqslant(\lambda_u+t_3)\|\delta_u\|_{1,\varpi}$$

Since the conditions of Lemma 12 hold we have

$$\begin{aligned}
&\mathbb{E}_n[\mathrm{E}[K_\varpi(W)\{\rho_\tau(X_a-X_{-a}'(\beta_u+\delta_u))-\rho_\tau(X_a-X_{-a}'\beta_u)\}|X_{-a},W]\\
&\qquad\geqslant-(\sqrt{\bar{f}}+1)\|r_u\|_{n,\varpi}\|\sqrt{f_u}X_{-a}'\delta\|_{n,\varpi}-K_u\|\delta_u\|_{1,\varpi}\\
&\qquad\quad+\tfrac{\|\sqrt{f_u}X_{-a}'\delta_u\|_{n,\varpi}^2}{4}\wedge\bar{q}_{A_u}\|\sqrt{f_u}X_{-a}'\delta_u\|_{n,\varpi}
\end{aligned}$$

where $K_u$ is given in $\Omega_4$ which accounts for the misspecification the conditional quantile condition. Therefore, we have

$$\begin{aligned}\tfrac{\|\sqrt{f_u}X'_{-a}\delta_u\|_{n,\varpi}^2}{4} \wedge \bar{q}_{A_u}\|\sqrt{f_u}X'_{-a}\delta_u\|_{n,\varpi} &\leqslant (\bar{f}^{1/2}+1)\|r_u\|_{n,\varpi}\|\sqrt{f_u}X'_{-a}\delta_u\|_{n,\varpi} + (\lambda_u + t_3 + K_u)\|\delta_u\|_{1,\varpi} \\ &\leqslant \{(\bar{f}^{1/2}+1)\|r_u\|_{n,\varpi} + \tfrac{3\mathbf{c}\sqrt{s}}{\kappa_{u,2\mathbf{c}}}(\lambda_u + t_3 + K_u)\}\|\sqrt{f_u}X'_{-a}\delta_u\|_{n,\varpi} \\ &\quad +(\lambda_u + t_3 + K_u)\tfrac{2\mathbf{c}}{\lambda_u}\bar{R}_{u\xi}\end{aligned}$$

The result then follows with the same argument under the current assumptions that account for $K_u$. ∎

**Lemma 8** (PQGM, Event $\Omega_1$)**.** *Under Condition P, we have*

$$\mathrm{P}\left(\sup_{u\in\mathcal{U},j\in[d]} \frac{|\mathbb{E}_n[K_\varpi(W)(\tau - 1\{X_a \leqslant X'_{-a}\beta_u + r_u\})X_{-a,j}]|}{\widehat{\sigma}^X_{a\varpi j}} > t\right) \leqslant 8|V|(\frac{ne}{d_W})^{2d_W}\exp\left(-\left\{\frac{t/(1+\bar{\delta}_n)}{2(1+1/16)}\right\}^2\right)$$

*where $t \geqslant 4\sup_{u\in\mathcal{U}}\{\mathrm{E}[K_\varpi(W)(\tau - 1\{X_a \leqslant X'_{-a}\beta_u + r_u\})^2X^2_{-a,j}]\}^{1/2}$ and $\bar{\delta}_n = o(1)$. In particular, the RHS is less than $\xi$ if $t \geqslant 2(1+\bar{\delta}_n)(1+1/16)\sqrt{\log(8|V|\{ne/d_W\}^{2d_W}/\xi)}$.*

*Proof.* Set $\sigma^X_{a\varpi j} := \mathrm{E}[K_\varpi(W)X^2_{-a,j}]^{1/2}$. We have that for any $\bar{\delta}_n \to 0$

$$\begin{aligned}&\mathrm{P}(\lambda_0 \leqslant \sup_{u\in\mathcal{U},j\in[d]} |\mathbb{E}_n[K_\varpi(W)(\tau - 1\{X_a \leqslant X'_{-a}\beta_u\})X_{-a,j}]|/\widehat{\sigma}^X_{a\varpi j}) \\ &\leqslant \mathrm{P}(\lambda_0 \leqslant (1+\bar{\delta}_n)\sup_{u\in\mathcal{U},j\in[d]} |\mathbb{E}_n[K_\varpi(W)(\tau - 1\{X_a \leqslant X'_{-a}\beta_u\})X_{-a,j}]|/\sigma^X_{a\varpi j}) \\ &\qquad +\mathrm{P}(\sup_{u\in\mathcal{U},j\in[d]} \sigma^X_{a\varpi j}/\widehat{\sigma}^X_{a\varpi j} \geqslant (1+\bar{\delta}_n))\end{aligned} \tag{E.59}$$

To bound the last term in (E.59), note that under Condition P, $c\mu_{\mathcal{W}} \leqslant (\sigma^X_{a\varpi j})^2 \leqslant C$ and $\mathcal{W}$ is a VC class of set with VC dimension $d_W$. Therefore, by Lemma 19 we have that with probability $1 - o(1)$

$$\sup_{u\in\mathcal{U},j\in[d]} (\mathbb{E}_n - \mathrm{E})[K_\varpi(W)X^2_{-a,j}] \lesssim \sqrt{\frac{(1+d_W)\log(|V|M_n/\sigma_1)}{n}} + \frac{(1+d_W)M_n^2\log(|V|M_n/\sigma_1)}{n} \tag{E.60}$$

for $\sigma_1^2 = \max_{u\in\mathcal{U},j\in[d]}\mathrm{E}[K_\varpi(W)X^2_{-a,j}] \leqslant \max_{j\in V}\mathrm{E}[X^2_{-a,j}] \leqslant C$ and envelope $F = \|X\|^2_\infty$ so that $\|F\|_{P,2} \leqslant \|\max_{i\leqslant n}F_i\|_{P,2} \leqslant M_n^2$. Thus for $\bar{\delta}_n \to 0$, provided $(1+d_W)M_n^2\log(|V|n) = o(n^{1/2})$ and $(1+d_W)\log(|V|n) = o(n\bar{\delta}_n^2\mu^2_{\mathcal{W}})$, so that the RHS of (E.60) is $o(\bar{\delta}_n\mu_{\mathcal{W}})$, we have

$$\mathrm{P}\left(\frac{1}{1+\bar{\delta}_n} \leqslant \frac{\sigma^X_{a\varpi j}}{\widehat{\sigma}^X_{a\varpi j}} \leqslant (1+\bar{\delta}_n),\ \text{for all } u\in\mathcal{U}, j\in[d]\right) = 1 - o(1). \tag{E.61}$$

Now we bound the first term of the RHS of (E.59). and let $\sigma_2^2 = \sup_{u\in\mathcal{U},j\in[d]}\mathrm{Var}(K_\varpi(W)(\tau - 1\{X_a \leqslant X'_{-a}\beta_u\})X_{-a,j}/\sigma^X_{a\varpi j}) \leqslant 1$. By symmetrization (adapting Lemma 2.3.7 in [73] to replace the "arbitrary" factor 2 with $1+\bar{\delta}_n$), for $\delta := 1/(2(1+n\sigma_2^2/t^2)) < 1/2$ we have

$$\begin{aligned}(*) &:= \mathrm{P}(\sup_{u\in\mathcal{U},j\in[d]}|\textstyle\sum_{i=1}^n K_\varpi(W_i)(\tau - 1\{X_{ia} \leqslant X'_{i,-a}\beta_u\})X_{i,-aj}/\sigma^X_{a\varpi j}| \geqslant t) \\ &\leqslant 2\mathrm{P}(\sup_{u\in\mathcal{U},j\in[d]}|\textstyle\sum_{i=1}^n \varepsilon_i K_\varpi(W_i)(\tau - 1\{X_{ia} \leqslant X'_{i,-a}\beta_u\})X_{i,-aj}/\sigma^X_{a\varpi j}| \geqslant t\delta) \\ &\leqslant 2\mathrm{P}\left(\sup_{u\in\mathcal{U},j\in[d]} \frac{|\sum_{i=1}^n \varepsilon_i K_\varpi(W_i)(\tau - 1\{X_{ia}\leqslant X'_{i,-a}\beta_u\})X_{i,-aj}|}{\widehat{\sigma}^X_{a\varpi j}} \geqslant t\delta/(1+\bar{\delta}_n)\right) + o(1)\end{aligned}$$

where $\varepsilon_i, i = 1,\ldots,n$, are Rademacher random variables independent of the data, and the last inequality follows from (E.61).

Therefore, by the union bound and symmetry, and iterated expectations we have

$$(*) \leqslant 4|V| \max_{j\in[d]} \mathrm{E}_{W,X}\left[\mathrm{P}_\varepsilon\left(\sup_{u\in\mathcal{U}} \frac{\left|\sum_{i=1}^n \varepsilon_i K_\varpi(W_i)(\tau - 1\{X_{ia} \leqslant X_{i,-a}'\beta_u\})X_{i,-aj}\right|}{\widehat{\sigma}^X_{a\varpi j}} \geqslant t\delta/(1+\bar{\delta}_n) \mid W, X\right)\right]$$

Next we use that $\{\varpi \in \mathcal{W}\}$ is a VC class of sets with VC dimension bounded by $d_W$ and $\{1\{X_a \leqslant X_{-a}'\beta_u\} : (\tau,\varpi) \in \mathcal{T}\times\mathcal{W}\}$ is a VC class of sets with VC dimension bounded by $1+d_W$. By Corollary 2.6.3 in [72], we have that conditionally on $(W_i, X_i)_{i=1}^n$, the set of (binary) sequences $\{(K_\varpi(W_i))_{i=1,\ldots,n} : \varpi \in \mathcal{W}\}$ has at most $\sum_{j=0}^{d_W-1}\binom{n}{j}$ different values. Similarly, $\{(1\{X_{ia} \leqslant X_{i,-a}'\beta_u\})_{i=1,\ldots,n} : u \in \mathcal{U}\}$ assumes at most $\sum_{j=0}^{d_W}\binom{n}{j}$ different values. Assuming that $n \geqslant d_W$, we have $\sum_{j=0}^k \binom{n}{j} \leqslant \{ne/k\}^k$ and

$$\begin{aligned}
&\mathrm{P}_\varepsilon\left(\sup_{u\in\mathcal{U}} \frac{\left|\sum_{i=1}^n \varepsilon_i K_\varpi(W_i)(\tau-1\{X_{ia}\leqslant X_{i,-a}'\beta_u\})X_{i,-aj}\right|}{\widehat{\sigma}^X_{a\varpi j}} \geqslant t\delta/(1+\bar{\delta}_n) \mid W,X\right)\\
&\leqslant \{\tfrac{ne}{d_W-1}\}^{d_W-1}\{\tfrac{ne}{d_W}\}^{d_W} \sup_{u\in\mathcal{U}} \mathrm{P}_\varepsilon\left(\sup_{\tilde\tau\in\mathcal{T}} \frac{\left|\sum_{i=1}^n \varepsilon_i K_\varpi(W_i)(\tilde\tau-1\{X_{ia}\leqslant X_{i,-a}'\beta_u\})X_{i,-aj}\right|}{\widehat{\sigma}^X_{a\varpi j}} \geqslant t\delta/(1+\bar{\delta}_n) \mid W,X\right)\\
&\leqslant \{ne/d_W\}^{2d_W} \sup_{u\in\mathcal{U},\tilde\tau\in[\underline{\tau},\bar\tau]} \mathrm{P}_\varepsilon\left(\frac{\left|\sum_{i=1}^n \varepsilon_i K_\varpi(W_i)(\tilde\tau-1\{X_{ia}\leqslant X_{i,-a}'\beta_u\})X_{i,-aj}\right|}{\widehat{\sigma}^X_{a\varpi j}} \geqslant t\delta/(1+\bar{\delta}_n) \mid W,X\right)\\
&\leqslant 2\{ne/d_W\}^{2d_W}\exp(-\{t\delta/[1+\bar{\delta}_n]\}^2)
\end{aligned}$$

here we used that the expression is linear in $\tau$ and so it is maximized at the extremes. Combining the bounds in the last two displayed equations we have

$$(*) \leqslant 8|V|\{ne/d_W\}^{2d_W}\exp(-\{t\delta/[1+\bar{\delta}_n]\}^2).$$

Therefore, setting $\lambda_0 = ct/n$ where $t \geqslant 4\sqrt{n}\sigma_2$ and $t \geqslant 2(1+\bar{\delta}_n)(1+1/16)\sqrt{2\log(8p|V|\{ne/d_W\}^{2d_W}/\xi)}$. (Note that $t \geqslant 4\sqrt{n}\sigma_2$ implies that $\delta \geqslant 1/\{2(1+1/16)\}$.) ∎

**Lemma 9** (PQGM, Event $\Omega_2$)**.** *Under Condition P we have*

$$\mathrm{P}\left(\sup_{u\in\mathcal{U}} \widehat{R}_u(\bar\beta_u) \leqslant C\{1+\bar f\}\{n^{-1}s(1+d_W)\log(|V|n)\}\right) = 1 - o(1).$$

*Proof of Lemma 9.* We have that $\widehat{R}_u(\bar\beta_u) \geqslant 0$ by convexity of $\rho_\tau$. Let $\varepsilon_{iu} = X_{ia} - X_{i,-a}\bar\beta_u - r_{iu}$ where $\|\bar\beta_u\|_0 \leqslant s$ and $r_{iu} = X_{-a}'(\beta_u - \bar\beta_u)$. By Knight's identity (F.65), $\widehat{R}_u(\bar\beta_u) = -\mathbb{E}_n[K_\varpi(W)r_u\int_0^1 1\{\varepsilon_u \leqslant -tr_u\} - 1\{\varepsilon_u \leqslant 0\}\,dt] \geqslant 0$.

$$\begin{aligned}
\mathbb{E}_n[\mathrm{E}[\widehat{R}_u(\bar\beta_u)] &= \mathbb{E}_n[K_\varpi(W)r_u\textstyle\int_0^1 F_{X_a|X_{-a},\varpi}(X_{-a}'\bar\beta_u + (1-t)r_u) - F_{X_a|X_{-a},\varpi}(X_{-a}'\bar\beta_u + r_u)\,dt]\\
&\leqslant \mathbb{E}_n[K_\varpi(W)r_u\textstyle\int_0^1 \bar f t r_u dt] \leqslant \bar f[\|r_u\|_{n,\varpi}^2]/2 \leqslant C\bar f s/n.
\end{aligned}$$

Thus, by Markov's inequality we have $\inf_{u\in\mathcal{U}} \mathrm{P}(\widehat{R}_u(\bar\beta_u) \leqslant C\bar f s/n) \geqslant 1/2$.

Define $z_{iu} := -\int_0^1 1\{\varepsilon_{iu} \leqslant -tr_{iu}\} - 1\{\varepsilon_{iu}\leqslant 0\}\,dt$, so that $\widehat{R}_u(\bar\beta_u) = \mathbb{E}_n[K_\varpi(W)r_u z_u]$ with $|z_{iu}| \leqslant 1$. By Lemma 2.3.7 in [73] (note that the Lemma does not require zero mean stochastic processes), for $t \geqslant 2C\bar f s/n$ we have

$$\frac{1}{2}\mathrm{P}\left(\sup_{u\in\mathcal{U}} |\mathbb{E}_n[K_\varpi(W)r_u z_u]| \geqslant t\right) \leqslant 2\mathrm{P}\left(\sup_{u\in\mathcal{U}} |\mathbb{E}_n[\varepsilon K_\varpi(W)r_u z_u]| > t/4\right)$$

where $\varepsilon_i, i = 1,\ldots,n$, are Rademacher random variables independent of the data.

Consider the class of functions $\mathcal{F} = \{-K_\varpi(W)r_u(1\{\varepsilon_{iu} \leqslant -B_i r_{iu}\} - 1\{\varepsilon_{iu}\leqslant 0\}) : u \in \mathcal{U}\}$ where $B_i \sim \mathrm{Uniform}(0,1)$ independent of $(X_i, W_i)_{i=1}^n$. It follows that $K_\varpi(W)r_u z_u = \mathrm{E}[-K_\varpi(W)r_u(1\{\varepsilon_{iu} \leqslant -B_i r_{iu}\} - 1\{\varepsilon_{iu}\leqslant 0\})|X_i, W_i]$ where the expectation is taken over $B_i$ only. Thus we will bound the

entropy of $\overline{\mathcal{F}} = \{\mathrm{E}[f|X,W] : f \in \mathcal{F}\}$ via Lemma 25. Note that $\mathcal{R} := \{r_u = X'_{-a}\beta_u - X'_{-a}\bar{\beta}_u : u \in \mathcal{U}\}$ where $\mathcal{G} := \{X'_{-a}\bar{\beta}_u : u \in \mathcal{U}\}$ is contained in the union of at most $|V|\binom{p}{s}$ VC-classes of dimension $Cs$ and $\mathcal{H} := \{X'_{-a}\beta_u : u \in \mathcal{U}\}\}$ is a VC-class of functions of dimension $(1+d_W)$ by Condition P. Finally note that $\mathcal{E} := \{\varepsilon_{iu} : u \in \mathcal{U}\} \subset \{X_{ia} : a \in V\} - \mathcal{G} - \mathcal{R}$.

Therefore, we have

$$\begin{aligned}\sup_Q \log N(\epsilon\|\bar{F}\|_{Q,2}, \overline{\mathcal{F}}, \|\cdot\|_{Q,2}) &\leqslant \sup_Q \log N((\epsilon/4)^2\|F\|_{Q,2}, \mathcal{F}, \|\cdot\|_{Q,2}) \\ &\leqslant \sup_Q \log N(\tfrac{1}{8}(\epsilon^2/16), \mathcal{W}, \|\cdot\|_{Q,2}) \\ &+ \sup_Q \log N(\tfrac{1}{8}(\epsilon^2/16)\|F\|_{Q,2}, \mathcal{R}, \|\cdot\|_{Q,2}) \\ &+ \sup_Q \log N(\tfrac{1}{8}(\epsilon^2/16), 1\{\mathcal{E} + \{B\}\mathcal{R} \leqslant 0\} - 1\{\mathcal{E} \leqslant 0\}, \|\cdot\|_{Q,2})\end{aligned}$$

By Lemma 19 with envelope $\bar{F} = \|X\|_\infty \sup_{u\in\mathcal{U}} \|\beta_u - \bar{\beta}_u\|_1$, and $(\sigma^r_{\max})^2 = \sup_{u\in\mathcal{U}} \mathrm{E}[K_\varpi(W) r_u^2] \lesssim s/n$ by Condition P, we have that with probability $1 - o(1)$

$$\sup_{u\in\mathcal{U}} |\mathbb{E}_n[\varepsilon K_\varpi(W) r_u z_u]| \lesssim \sqrt{\frac{s(1+d_W)\log(|V|n)}{n}}\sqrt{\frac{s}{n}} + \frac{M_n\sqrt{s^2/n}\log(|V|n)}{n} \lesssim \frac{s(1+d_W)\log(|V|n)}{n}$$

under $M_n\sqrt{s^2/n} \leqslant C$. ∎

**Lemma 10** (PQGM, Event $\Omega_3$)**.** *Under Condition P, for $u = (a, \tau, \varpi) \in V \times \mathcal{T} \times \mathcal{W}$, define $g_u(\delta, X, W) = K_\varpi(W)\{\rho_\tau(X_a - X'_{-a}(\beta_u + \delta)) - \rho_\tau(X_a - X'_{-a}\beta_u)\}$, and*

$$\Omega_3 := \left\{ \sup_{u\in\mathcal{U}, 1/\sqrt{n} \leqslant \|\delta\|_{1,\varpi} \leqslant \sqrt{n}} \frac{|\mathbb{E}_n[g_u(\delta, X, W) - \mathrm{E}[g_u(\delta, X, W)|X_{-a}, W]]|}{\|\delta\|_{1,\varpi}} < t_3 \right\}.$$

*Then, under Condition CI we have $\mathrm{P}(\Omega_3) \geqslant 1 - \xi$ for any*

$$t_3\sqrt{n} \geqslant 12 + 16\sqrt{2\log\left(64|V|^2 n^{1+d_W}\log(n)(L_\beta M_n\sqrt{n})^{1+d_W/\kappa}/\xi\right)}$$

*Proof.* We have that $\Omega_3^c := \{\max_{a\in V} A_a \geqslant t_3\sqrt{n}\}$ for

$$A_a := \sup_{(\tau,\varpi)\in\mathcal{T}\times\mathcal{W}, \underline{N} \leqslant \|\delta\|_{1,\varpi} \leqslant \bar{N}} \sqrt{n}\left|\frac{\mathbb{E}_n[g_u(\delta, X, W) - \mathrm{E}[g_u(\delta, X, W)|X_{-a}, W]]}{\|\delta\|_{1,\varpi}}\right|.$$

We will apply Lemma 13 with $\rho = \kappa$, $L_\eta = L_\beta$, $\tilde{x} = X_{-a}$ (so we take $p = |V|$), $\underline{N} = 1/\sqrt{n}$ and $\bar{N} = \sqrt{n}$. Therefore, we have by Lemma 13 and the union bound

$$\begin{aligned}\mathrm{P}(\Omega_3^c) &= \mathrm{P}(\max_{a\in V} A_a \geqslant t_3\sqrt{n}) \\ &\leqslant |V| \max_{a\in V} \mathrm{P}(A_a \geqslant t_3\sqrt{n}) \\ &= |V| \max_{a\in V} \mathrm{E}_{X_{-a},W}\{\mathrm{P}(A_a \geqslant t_3\sqrt{n} \mid X_{-a}, W)\} \\ &\leqslant |V| \max_{a\in V} \mathrm{E}_{X_{-a},W}\left\{8|V|\ |\widehat{\mathcal{N}}| \cdot |\widehat{\mathcal{W}}| \cdot |\widehat{\mathcal{T}}| \exp(-(t_3\sqrt{n}/4 - 3)^2/32)\right\} \\ &\leqslant C|V|^2 n^{1+d_W}\log(n) L_\beta^{1+d_W/\kappa} \mathrm{E}_{X_{-a}}\left\{\frac{\max_{i\leqslant n}\|X_{i,-a}\|_\infty^{1+d_W/\kappa}}{\underline{N}^{1+d_W/\kappa}}\right\}\exp(-(t_3\sqrt{n}/4 - 3)^2/32) \\ &\leqslant \xi\end{aligned}$$

where the last step follows by the choice of $t_3$. ∎

**Lemma 11** (PQGM, Event $\Omega_4$)**.** *Under Condition P, and setting $K_u = C\sqrt{\frac{(1+d_W)\log(|V|n)}{n}}$, we have that $\mathrm{P}(\Omega_4) = 1 - o(1)$.*

*Proof.* First note that by Lemma 19 we have that with probability $1 - o(1)$

$$\sup_{\varpi \in \mathcal{W}, j \in V} (\mathbb{E}_n - \mathrm{E})[K_\varpi(W) X_{-a,j}^2] \lesssim \sqrt{\frac{(1+d_W)\log(|V|M_n/\sigma)}{n}} + \frac{(1+d_W)M_n^2 \log(|V|M_n/\sigma)}{n}$$

and $\sigma_{a\varpi j}^X \geqslant c\mathrm{P}(\varpi)$. Under $(1+d_W)\log(|V|M_n/\sigma) \leqslant \delta_n^2 \mu_{\mathcal{W}}^2$ and $(1+d_W)M_n^2 \log(|V|M_n/\sigma) \leqslant \delta_n n \mu_{\mathcal{W}}$, we have that $|(\mathbb{E}_n - \mathrm{E})[K_\varpi(W)X_{-a,j}^2]| = o(\sigma_{a\varpi j}^X)$ for all $u \in \mathcal{U}$. Therefore, we have

$$\begin{aligned}
&\mathrm{P}(\sup_{u\in\mathcal{U}} |\mathbb{E}_n[h_{uj}(X_{-a}, W)]|/\{K_u \widehat{\sigma}_{a\varpi j}^X\} > 1) \\
&\leqslant \mathrm{P}(\sup_{u\in\mathcal{U}} |\mathbb{E}_n[h_{uj}(X_{-a}, W)]|/\{K_u \sigma_{a\varpi j}^X\} > 1 + O(\delta_n)) + o(1)
\end{aligned}$$

Applying Lemma 19 to $\mathcal{F} = \{h_{uj}(X_{-a}, W)/\sigma_{a\varpi j}^X : u \in \mathcal{U}\}$. For convenience define $\bar{\mathcal{H}}_j = \{h_{uj}(X_{-a}, W) : u \in \mathcal{U}\}$ and $\bar{\mathcal{K}}_j := \{\mathrm{E}[K_\varpi(W)X_j^2] : \varpi \in \mathcal{W}\}$. Note that $\bar{\mathcal{K}}_j$ has covering numbers bounded by the covering number of $\mathcal{K}_j := \{K_\varpi(W)X_j^2 : \varpi \in \mathcal{W}\}$ hence $\sup_Q \log N(\epsilon \|\bar{K}_j\|_{Q,2}, \bar{\mathcal{K}}_j, \|\cdot\|_{Q,2}) \leqslant \log \sup_{\tilde{Q}} N((\epsilon/4)^2 \|F\|_{\tilde{Q},2}, \mathcal{K}_j, \|\cdot\|_{\tilde{Q},2})$ by Lemma 25. Similarly, Lemma 25 also allows us to bound covering numbers of $\bar{\mathcal{H}}_j$ via covering numbers of $\mathcal{H}_j = \{K_\varpi(W)(\tau - 1\{X_a \leqslant X'_{-a}\beta_u\})X_j : u \in \mathcal{U}\}$.

$$\begin{aligned}
&\sup_Q \log N(\epsilon \|F\|_{Q,2}, \mathcal{F}, \|\cdot\|_{Q,2}) \leqslant p \max_{j\in[p]} \sup_Q \log N(\epsilon \|F_j\|_{Q,2}, \mathcal{F}_j, \|\cdot\|_{Q,2}) \\
&\leqslant p \max_{j\in[p]} \sup_Q \{\log N((1/2)\epsilon \|\bar{H}_j\|_{Q,2}, \bar{\mathcal{H}}_j, \|\cdot\|_{Q,2}) + \sup_Q \log N((1/2)\epsilon c \mu_{\mathcal{W}}^{1/2}, 1/\bar{\mathcal{K}}_j^{1/2}, \|\cdot\|_{Q,2})\} \\
&\leqslant p \max_{j\in[p]} \sup_Q \{\log N((1/2)\epsilon \|\bar{H}_j\|_{Q,2}, \bar{\mathcal{H}}_j, \|\cdot\|_{Q,2}) + \sup_Q \log N((1/2)\epsilon c \mu_{\mathcal{W}}^{3/2}, \bar{\mathcal{K}}_j^{1/2}, \|\cdot\|_{Q,2})\} \\
&\leqslant p \max_{j\in[p]} \sup_Q \{\log N((1/2)\epsilon \|\bar{H}_j\|_{Q,2}, \bar{\mathcal{H}}_j, \|\cdot\|_{Q,2}) + \sup_Q \log N(C(1/2)\epsilon c \mu_{\mathcal{W}}^{3/2}, \bar{\mathcal{K}}_j, \|\cdot\|_{Q,2})\} \\
&\leqslant p \max_{j\in[p]} \sup_Q \{\log N((1/2)\epsilon \|\bar{H}_j\|_{Q,2}, \bar{\mathcal{H}}_j, \|\cdot\|_{Q,2}) + \sup_Q \log N(1/(2C)\epsilon c \mu_{\mathcal{W}}^{3/2}, \bar{\mathcal{K}}_j, \|\cdot\|_{Q,2})\} \\
&\leqslant p \max_{j\in[p]} \sup_{\tilde{Q}} \log N((1/4)\epsilon^2 \|H_j\|_{Q,2}, \mathcal{H}_j, \|\cdot\|_{Q,2}) \\
&+ p \max_{j\in[p]} \sup_{\tilde{Q}} \log N((1/(4C^2))\epsilon^2 c^2 \mu_{\mathcal{W}}^3, \mathcal{K}_j, \|\cdot\|_{\tilde{Q},2})
\end{aligned}$$

where $F_j = c\|X\|_\infty/\mu_{\mathcal{W}}^{1/2}$, $H_j = \|X\|_\infty$. Since $\mathcal{K}_j$ is the product of a VC subgraph of dimension $d_W$ with a single function, and $\mathcal{H}_j$ is the product of two VC subgraph of dimension $1 + d_W$ and a single function, by Lemma 19 with $\sigma^2 = 1$, we have with probability $1 - o(1)$

$$\sup_{u\in\mathcal{U}} \left| \frac{(\mathbb{E}_n - \mathrm{E})[h_{uj}(X_{-a}, W)]}{\mathrm{E}[K_\varpi(W)X_j^2]^{1/2}} \right| \leqslant C\sqrt{\frac{(1+d_W)\log(|V|n)}{n}} + C\frac{M_n(1+d_W)\log(|V|n)}{n\mu_{\mathcal{W}}^{1/2}}.$$

Thus, under $M_n(1+d_W)\log(|V|n) \leqslant n^{1/2}\mu_{\mathcal{W}}^{1/2}$ we have that we can take $K_u = C\sqrt{\frac{(1+d_W)\log(|V|n)}{n}}$. ∎

## Appendix F. Technical Results for High-Dimensional Quantile Regression

In this section we provide technical results for high-dimensional quantile regression. It is based on a sample $(\tilde{y}_i, \tilde{x}_i, W_i)_{i=1}^n$, independent across $i$, $\rho_\tau(t) = (\tau - 1\{t \leqslant 0\})t$, $\tau \in \mathcal{T} \subset (0,1)$ a compact interval, and a family of indicator functions $K_w(W) = 1$ if $W \in \Omega_\varpi$, $K_w(W) = 0$ otherwise, here $\Omega_\varpi \in \mathcal{W}$. For convenience we index the sets $\Omega_\varpi$ by $\varpi \in B_W \subset \mathbb{R}^{d_W}$ where we normalize the diameter of $B_W$ to be less or equal than 1/6. Let $f_{\tilde{y}|\tilde{x},r_u,\varpi}(\cdot)$ denote the conditional density function, $f_{\tilde{y}|\tilde{x},r_u,\varpi}(\cdot) \leqslant \bar{f}$, $|f'_{\tilde{y}|\tilde{x},r_u,\varpi}(\cdot)| \leqslant \bar{f}'$ and $f_u := f_{\tilde{y}|\tilde{x},r_u,\varpi}(\tilde{x}'\eta_u)$. Moreover, we assume that

$$\|\eta_u - \eta_{\tilde{u}}\|_1 \leqslant L_\eta\{|\tau - \tilde{\tau}| + \|\varpi - \tilde{\varpi}\|^\rho\}. \tag{F.62}$$

Although the results can be applied more generally, these results will be used for $(\eta_u, r_u), u = (\tilde{y}, \tau, \varpi) \in \mathcal{U} := \{\tilde{y}\} \times \mathcal{T} \times \mathcal{W}$ satisfying

$$\mathrm{E}[K_\varpi(W)(\tau - 1\{\tilde{y} \leqslant \tilde{x}'\eta_u + r_u\})\tilde{x}] = 0.$$

Note that this generality is flexible enough to allow us to cover the case that the $\tau$-conditional quantile function $Q_{\tilde{y}}(\tau|\tilde{x},\varpi) = \tilde{x}'\tilde{\eta}_u + \tilde{r}_u$ by setting $\eta_u = \tilde{\eta}_u$ and $r_u = \tilde{r}_u$ in which case $\mathrm{E}[(\tau - 1\{\tilde{y} \leqslant \tilde{x}'\eta_u + r_u\})|\tilde{x},\varpi] = 0$. It also covers the case that

$$\tilde{\eta}_u \in \arg\min_{\beta} \mathrm{E}[K_\varpi(W)\rho_\tau(\tilde{y} - \tilde{x}'\beta)]$$

so that $\mathrm{E}[K_\varpi(W)(\tau - 1\{\tilde{y} \leqslant \tilde{x}'\tilde{\eta}_u\})\tilde{x}] = 0$ holds by the first order condition by setting $\eta_u = \tilde{\eta}_u$ and $r_u = 0$. Moreover, it also covers the case that we work with a sparse approximation $\bar{\eta}_u$ of $\tilde{\eta}_u$ by setting $\eta_u = \bar{\eta}_u$ and $r_u = \tilde{x}'(\tilde{\eta}_u - \bar{\eta}_u)$.

**Lemma 12** (Identification Lemma)**.** *For $u = (a,\tau,\varpi) \in \mathcal{U}$, and a subset $A_u \subset \mathbb{R}^p$ let*

$$\bar{q}_{A_u} = 1/(2\bar{f}') \cdot \inf_{\delta \in A_u} \mathbb{E}_n\left[K_\varpi(W) f_u |\tilde{x}'\delta|^2\right]^{3/2} / \mathbb{E}_n\left[K_\varpi(W)|\tilde{x}'\delta|^3\right] \tag{F.63}$$

*and assume that for all $\delta \in A_u$*

$$\mathbb{E}_n\left[K_\varpi(W)|r_u| \cdot |\tilde{x}'\delta|^2\right] + \mathbb{E}_n\left[K_\varpi(W) r_u^2 \cdot |\tilde{x}'\delta|^2\right] \leqslant 1/(4\bar{f}')\mathbb{E}_n[K_\varpi(W) f_u |\tilde{x}'\delta|^2]. \tag{F.64}$$

*Then we have*

$$\begin{aligned}
&\mathbb{E}_n[\mathrm{E}[K_\varpi(W)\rho_\tau(\tilde{y} - \tilde{x}'(\eta_u + \delta)) \mid \tilde{x}, r_u, W]] - \mathbb{E}_n[\mathrm{E}[K_\varpi(W)\rho_\tau(\tilde{y} - \tilde{x}'\eta_u) \mid \tilde{x}, r_u, W]] \\
&\geqslant \tfrac{\|\sqrt{f_u}\tilde{x}'\delta\|_{n,\varpi}^2}{4} \wedge \left\{\bar{q}_{A_u}\|\sqrt{f_u}\tilde{x}'\delta\|_{n,\varpi}\right\} - K_{n2}\|\sqrt{f_u}\tilde{x}'\delta\|_{n,\varpi} - K_{n1}\|\delta\|_{1,\varpi}.
\end{aligned}$$

*where $K_{n2} := (\bar{f}^{1/2} + 1)\|r_u\|_{n,\varpi}$ and $K_{n1} := \sup_{u\in\mathcal{U}, j\in[p]} \frac{|\mathbb{E}_n[\mathrm{E}[K_\varpi(W)(\tau - 1\{\tilde{y} \leqslant \tilde{x}'\eta_u + r_u\})\tilde{x}_j|\tilde{x},W]]|}{\{\mathbb{E}_n[K_\varpi(W)\tilde{x}_j^2]\}^{1/2}}$.*

*Proof of Lemma 12.* Let $T_u = \mathrm{support}(\eta_u)$, and $Q_u(\eta) := \mathbb{E}_n\mathrm{E}[K_\varpi(W)\rho_\tau(\tilde{y} - \tilde{x}'\eta) \mid \tilde{x}, r_u, W]$. The proof proceeds in steps.

Step 1. (Minoration) Define the maximal radius over which the criterion function can be minorated by a quadratic function

$$r_{A_u} = \sup_r \left\{ \begin{array}{ll} r \ : & Q_u(\eta_u + \delta) - Q_u(\eta_u) + K_{n2}\|\sqrt{f_u}\tilde{x}'\delta\|_{n,\varpi} + K_{n1}\|\delta\|_{1,\varpi} \geqslant \frac{1}{4}\|\sqrt{f_u}\tilde{x}'\delta\|_{n,\varpi}^2, \\ & \forall \delta \in A_u,\ \|\sqrt{f_u}\tilde{x}'\delta\|_{n,\varpi} \leqslant r \end{array} \right\}.$$

Step 2 below shows that $r_{A_u} \geqslant \bar{q}_{A_u}$. By construction of $r_{A_u}$ and the convexity of $Q_u(\cdot)$, $\|\cdot\|_{1,\varpi}$ and $\|\cdot\|_{n,\varpi}$,

$$\begin{aligned}
&Q_u(\eta_u + \delta) - Q_u(\eta_u) + K_{n2}\|\sqrt{f_u}\tilde{x}'\delta\|_{n,\varpi} + K_{n1}\|\delta\|_{1,\varpi} \geqslant \\
&\geqslant \tfrac{\|\sqrt{f_u}\tilde{x}'\delta\|_{n,\varpi}^2}{4} \wedge \left\{ \begin{array}{ll} \frac{\|\sqrt{f_u}\tilde{x}'\delta\|_{n,\varpi}}{r_A} \cdot & \inf_{\tilde{\delta}} \quad Q_u(\eta_u + \tilde{\delta}) - Q_u(\eta_u) + K_{n2}\|\sqrt{f_u}\tilde{x}'\tilde{\delta}\|_{n,\varpi} + K_{n1}\|\tilde{\delta}\|_{1,\varpi} \\ & \tilde{\delta} \in A_u, \|\sqrt{f_u}\tilde{x}'\tilde{\delta}\|_u \geqslant r_{A_u} \end{array} \right\} \\
&\geqslant \tfrac{\|\sqrt{f_u}\tilde{x}'\delta\|_{n,\varpi}^2}{4} \wedge \left\{ \tfrac{\|\sqrt{f_u}\tilde{x}'\delta\|_{n,\varpi}}{r_{A_u}} \tfrac{r_{A_u}^2}{4} \right\} \\
&\geqslant \tfrac{\|\sqrt{f_u}\tilde{x}'\delta\|_{n,\varpi}^2}{4} \wedge \left\{ \bar{q}_{A_u}\|\sqrt{f_u}\tilde{x}'\delta\|_{n,\varpi} \right\}.
\end{aligned}$$

Step 2. ($r_{A_u} \geqslant \bar{q}_{A_u}$) Let $F_{\tilde{y}|\tilde{x},r_u,\varpi}$ denote the conditional distribution of $\tilde{y}$ given $\tilde{x}, r_u, \varpi$. From [48], for any two scalars $w$ and $v$ the Knight's identity is

$$\rho_\tau(w - v) - \rho_\tau(w) = -v(\tau - 1\{w \leqslant 0\}) + \int_0^v (1\{w \leqslant z\} - 1\{w \leqslant 0\})dz. \tag{F.65}$$

Using (F.65) with $w = \tilde{y}_i - \tilde{x}_i'\eta_u$ and $v = \tilde{x}_i'\delta$ and taking expectations with respect to $\tilde{y}$, we have

$$\begin{aligned}
Q_u(\eta_u + \delta) - Q_u(\eta_u) = \quad & -\mathbb{E}_n[\mathrm{E}\left[K_\varpi(W)(\tau - 1\{\tilde{y} \leqslant \tilde{x}'\eta_u\})\tilde{x}_i'\delta \mid \tilde{x}, r_u, W\right]] \\
& + \mathbb{E}_n\left[\textstyle\int_0^{K_\varpi(W)\tilde{x}'\delta} F_{\tilde{y}|\tilde{x},r_u,\varpi}(\tilde{x}'\eta_u + t) - F_{\tilde{y}|\tilde{x},r_u,\varpi}(\tilde{x}'\eta_u)dt\right].
\end{aligned}$$

Using the law of iterated expectations and mean value expansion, the relation

$$
\begin{aligned}
&|\mathbb{E}_n[\mathrm{E}\left[K_\varpi(W)(\tau - 1\{\tilde{y} \leqslant \tilde{x}'\eta_u\})\tilde{x}'\delta \mid \tilde{x}, r_u, W\right]]| \\
&= |\mathbb{E}_n\left[K_\varpi(W)\{F_{\tilde{y}|\tilde{x},r_u,\varpi}(\tilde{x}'\eta_u + r_u) - F_{\tilde{y}|\tilde{x},r_u,\varpi}(\tilde{x}'\eta_u)\}\tilde{x}'\delta\right] \\
&+\mathbb{E}_n\left[K_\varpi(W)\{\tau - F_{\tilde{y}|\tilde{x},r_u,\varpi}(\tilde{x}'\eta_u + r_u)\}\tilde{x}'\delta \mid \tilde{x}, r_u, W\right]| \\
&\leqslant \mathbb{E}_n[K_\varpi(W)f_u|r_u|\ |\tilde{x}'\delta|] + \bar{f}'\mathbb{E}_n[K_\varpi(W)|r_u|^2|\tilde{x}'\delta|] + K_{n1}\|\delta\|_{1,\varpi} \\
&\leqslant \|\sqrt{f_u}r_u\|_{n,\varpi}\|\sqrt{f_u}\tilde{x}'\delta\|_{n,\varpi} + \bar{f}'\|r_u\|_{n,\varpi}\|r_u\tilde{x}'\delta\|_{n,\varpi} + K_{n1}\|\delta\|_{1,\varpi} \\
&\leqslant (\bar{f}^{1/2} + 1)\|r_u\|_{n,\varpi}\|\sqrt{f_u}\tilde{x}'\delta\|_{n,\varpi} + K_{n1}\|\delta\|_{1,\varpi}
\end{aligned}
$$

where we used our assumption on the approximation error and we have $K_{n2} = (\bar{f}^{1/2}+1)\|r_u\|_{n,\varpi}$. With that and similar arguments we obtain for $\tilde{t}_{\tilde{x}_i,t} \in [0,t]$

$$
\begin{aligned}
&Q_u(\eta_u + \delta) - Q_u(\eta_u) + K_{n2}\|\sqrt{f_u}\tilde{x}'\delta\|_{n,\varpi} + K_{n1}\|\delta\|_{1,\varpi} \geqslant \\
&Q_u(\eta_u + \delta) - Q_u(\eta_u) + \mathbb{E}_n[\mathrm{E}\left[K_\varpi(W)(\tau - 1\{\tilde{y} \leqslant \tilde{x}'\eta_u\})\tilde{x}'\delta \mid \tilde{x}, r_u, W\right]] = \\
&= \mathbb{E}_n\left[\int_0^{K_\varpi(W)\tilde{x}'\delta} F_{\tilde{y}|\tilde{x},r_u,\varpi}(\tilde{x}'\eta_u + t) - F_{\tilde{y}|\tilde{x},r_u,\varpi}(\tilde{x}'\eta_u)dt\right] \\
&= \mathbb{E}_n\left[\int_0^{K_\varpi(W)\tilde{x}'\delta} tf_{\tilde{y}|\tilde{x},r_u,\varpi}(\tilde{x}'\eta_u) + \tfrac{t^2}{2}f'_{\tilde{y}|\tilde{x},r_u,\varpi}(\tilde{x}'\eta_u + \tilde{t}_{\tilde{x},t})dt\right] \\
&\geqslant \tfrac{1}{2}\|\sqrt{f_u}\tilde{x}'\delta\|^2_{n,\varpi} - \tfrac{1}{6}\bar{f}'\mathbb{E}_n[K_\varpi(W)|\tilde{x}'\delta|^3] - \mathbb{E}_n\left[\int_0^{K_\varpi(W)\tilde{x}_i'\delta} t[f_{\tilde{y}|\tilde{x},r_u,\varpi}(\tilde{x}'\eta_u) - f_{\tilde{y}|\tilde{x},r_u,\varpi}(\tilde{x}'\eta_u + r_u)]dt\right] \\
&\geqslant \tfrac{1}{4}\|\sqrt{f_u}\tilde{x}'\delta\|^2_{n,\varpi} + \tfrac{1}{4}\|\sqrt{f_u}\tilde{x}'\delta\|^2_{n,\varpi} - \tfrac{1}{6}\bar{f}'\mathbb{E}_n[K_\varpi(W)|\tilde{x}'\delta|^3] - (\bar{f}'/2)\mathbb{E}_n\left[K_\varpi(W)|\tilde{r}_u| \cdot |\tilde{x}'\delta|^2\right].
\end{aligned}
\tag{F.66}
$$

Moreover, by assumption we have

$$
\mathbb{E}_n\left[K_\varpi(W)|r_u| \cdot |\tilde{x}'\delta|^2\right] \leqslant \tfrac{1}{4\bar{f}'}\mathbb{E}_n[K_\varpi(W)f_u|\tilde{x}'\delta|^2] \tag{F.67}
$$

Note that for any $\delta$ such that $\|\sqrt{f_u}\tilde{x}'\delta\|_{n,\varpi} \leqslant \bar{q}_{A_u}$ we have

$$
\|\sqrt{f_u}\tilde{x}'\delta\|_{n,\varpi} \leqslant \bar{q}_{A_u} \leqslant 1/(2\bar{f}') \cdot \mathbb{E}_n\left[K_\varpi(W)f_u|\tilde{x}'\delta|^2\right]^{3/2} / \mathbb{E}_n\left[K_\varpi(W)|\tilde{x}'\delta|^3\right].
$$

It follows that $(1/6)\bar{f}'\mathbb{E}_n[K_\varpi(W)|\tilde{x}'\delta|^3] \leqslant (1/8)\mathbb{E}_n[K_\varpi(W)f_u|\tilde{x}'\delta|^2]$. Combining this with (F.67) we have

$$
\frac{1}{4}\mathbb{E}_n[K_\varpi(W)f_u|\tilde{x}'\delta|^2] - \frac{\bar{f}'}{6}\mathbb{E}_n[K_\varpi(W)|\tilde{x}'\delta|^3] - \frac{\bar{f}'}{2}\mathbb{E}_n\left[K_\varpi(W)|r_u| \cdot |\tilde{x}'\delta|^2\right] \geqslant 0. \tag{F.68}
$$

Combining (F.66) and (F.68) we have $r_{A_u} \geqslant \bar{q}_{A_u}$. ∎

**Lemma 13.** *Let $\mathcal{W}$ be a VC-class of sets with VC-index $d_W$. Conditional on $\{(W_i, \tilde{x}_i), i = 1, \ldots, n\}$ we have*

$$
P_{\tilde{y}}\left(\sup_{\substack{\tau \in \mathcal{T}, \varpi \in \mathcal{W}, \\ \underline{N} \leqslant \|\delta\|_{1,\varpi} \leqslant \bar{N}}} \left|\mathbb{G}_n\left(K_\varpi(W)\frac{\rho_\tau(\tilde{y} - \tilde{x}'(\eta_u + \delta)) - \rho_\tau(\tilde{y} - \tilde{x}'\eta_u)}{\|\delta\|_{1,\varpi}}\right)\right| \geqslant M \mid (W_i, \tilde{x}_i)_{i=1}^n\right) \leqslant S_n \exp(-(M/4 - 3)^2/32)
$$

*where $S_n \leqslant 8p|\widehat{\mathcal{N}}| \cdot |\widehat{\mathcal{W}}| \cdot |\widehat{\mathcal{T}}|$, with*

$$
|\widehat{\mathcal{N}}| \leqslant 1 + \left\lfloor 3\sqrt{n}\log(\bar{N}/\underline{N})\right\rfloor, \quad |\widehat{\mathcal{T}}| \leqslant 2\sqrt{n}\frac{\max_{i\leqslant n}\|\tilde{x}_i\|_\infty}{\underline{N}}L_\eta, \quad |\widehat{\mathcal{W}}| \leqslant n^{d_W} + \left\{2\sqrt{n}\frac{\max_{i\leqslant n}\|\tilde{x}_i\|_\infty}{\underline{N}}L_\eta\right\}^{d_W/\rho}.
$$

*Proof of Lemma 13.* Let $g_{i\tau\varpi}(b) = K_\varpi(W_i)\{\rho_\tau(\tilde{y}_i - \tilde{x}_i'\eta_{\tau\varpi} + b) - \rho_\tau(\tilde{y}_i - \tilde{x}_i'\eta_{\tau\varpi})\} \leqslant K_\varpi(W_i)|b|$ since $K_\varpi(W_i) \in \{0,1\}$. Note that $|g_{i\tau\varpi}(b) - g_{i\tau\varpi}(a)| \leqslant K_\varpi(W_i)|b - a|$. To easy the notation we omit the conditioning on $(\tilde{x}_i, W_i)$ from the probabilities.

For any $\delta \in \mathbb{R}^p$, since $\rho_\tau$ is 1-Lipschitz, we have

$$\mathrm{Var}\left(\mathbb{G}_n\left(\tfrac{g_{\tau\varpi}(\tilde{x}'\delta)}{\|\delta\|_{1,\varpi}}\right)\right) \leqslant \tfrac{\mathbb{E}_n[\{g_{\tau\varpi}(\tilde{x}'\delta)\}^2]}{\|\tilde{x}'\delta\|_{n,\varpi}^2} \leqslant \tfrac{\mathbb{E}_n[|K_\varpi(W)\tilde{x}'\delta|^2]}{\|\tilde{x}'\delta\|_{n,\varpi}^2} = 1$$

since by definition $\|\delta\|_{1,\varpi} = \sum_j \|\delta_j\|_{1,\varpi} = \sum_j \|\tilde{x}_j'\delta_j\|_{n,\varpi} \geqslant \|\tilde{x}'\delta\|_{n,\varpi}$.

Since we are conditioning on $(W_i, \tilde{x}_i)_{i=1}^n$ the process is independent across $i$. Then, by Lemma 2.3.7 in [72] (Symmetrization for Probabilities) we have for any $M > 1$

$$\begin{aligned} &\mathrm{P}\left(\sup_{\tau\in\mathcal{T},\varpi\in\mathcal{W},\underline{N}\leqslant\|\delta\|_{1,\varpi}\leqslant\bar{N}}\left|\mathbb{G}_n\left(\frac{g_{\tau\varpi}(\tilde{x}'\delta)}{\|\delta\|_{1,\varpi}}\right)\right| \geqslant M\right) \\ \leqslant\ &\frac{2}{1-M^{-2}}P\left(\sup_{\tau\in\mathcal{T},\varpi\in\mathcal{W},\underline{N}\leqslant\|\delta\|_{1,\varpi}\leqslant\bar{N}}\left|\mathbb{G}_n^o\left(\frac{g_{\tau\varpi}(\tilde{x}'\delta)}{\|\delta\|_{1,\varpi}}\right)\right| \geqslant M/4\right) \end{aligned}$$

where $\mathbb{G}_n^o$ is the symmetrized process.

Consider $\mathcal{F}_{t,\tau,\varpi} = \{\delta : \|\delta\|_{1,\varpi} = t\}$. We will consider the families of $\mathcal{F}_{t,\tau,\varpi}$ for $t \in [\underline{N}, \bar{N}]$, $\tau \in \mathcal{T}$ and $\varpi \in \mathcal{W}$.

We will construct a finite net $\widehat{\mathcal{T}} \times \widehat{\mathcal{W}} \times \widehat{\mathcal{N}}$ of $\mathcal{T} \times \mathcal{W} \times [\underline{N}, \bar{N}]$ such that

$$\sup_{\tau\in\mathcal{T},\varpi\in\mathcal{W},t\in[\underline{N},\bar{N}],\delta\in\mathcal{F}_{t,\tau,\varpi}}\left|\mathbb{G}_n^o\left(\frac{g_{\tau\varpi}(\tilde{x}'\delta)}{\|\delta\|_{1,\varpi}}\right)\right| \leqslant 3 + \sup_{\tau\in\widehat{\mathcal{T}},\varpi\in\widehat{\mathcal{W}},t\in\widehat{\mathcal{N}}}\sup_{\delta\in\mathcal{F}_{t,\tau,\varpi}}\left|\mathbb{G}_n^o\left(\frac{g_{\tau\varpi}(\tilde{x}'\delta)}{t}\right)\right| =: 3 + \mathcal{A}^o.$$

By triangle inequality we have

$$\begin{aligned} \left|\mathbb{G}_n^o\left(\tfrac{g_{\tau\varpi}(\tilde{x}'\delta)}{t} - \tfrac{g_{\tilde{\tau}\tilde{\varpi}}(\tilde{x}'\tilde{\delta})}{\tilde{t}}\right)\right| &\leqslant \left|\mathbb{G}_n^o\left(\tfrac{g_{\tau\varpi}(\tilde{x}'\delta)}{t} - \tfrac{g_{\tilde{\tau}\varpi}(\tilde{x}'\delta)}{t}\right)\right| + \left|\mathbb{G}_n^o\left(\tfrac{g_{\tilde{\tau}\varpi}(\tilde{x}'\delta)}{t} - \tfrac{g_{\tilde{\tau}\tilde{\varpi}}(\tilde{x}'\delta)}{t}\right)\right| \\ &+ \left|\mathbb{G}_n^o\left(\tfrac{g_{\tilde{\tau}\tilde{\varpi}}(\tilde{x}'\delta)}{t} - \tfrac{g_{\tilde{\tau}\tilde{\varpi}}(\tilde{x}'\tilde{\delta})}{\tilde{t}}\right)\right| \end{aligned} \tag{F.69}$$

The first term in (F.69) is such that

$$\begin{aligned} \left|\mathbb{G}_n^o\left(\tfrac{g_{\tau\varpi}(\tilde{x}'\delta)}{t} - \tfrac{g_{\tilde{\tau}\varpi}(\tilde{x}'\delta)}{t}\right)\right| &\leqslant \tfrac{2\sqrt{n}}{t}\mathbb{E}_n[K_\varpi(W)|\tilde{x}'(\eta_{\tau\varpi} - \eta_{\tilde{\tau}\varpi})|] \\ &\leqslant \tfrac{2\sqrt{n}}{\underline{N}}\max_{i\leqslant n}\|\tilde{x}_i\|_\infty\mathbb{E}_n[K_\varpi(W)]\|\eta_{\tau\varpi} - \eta_{\tilde{\tau}\varpi}\|_1 \\ &\leqslant \tfrac{2\sqrt{n}}{\underline{N}}\max_{i\leqslant n}\|\tilde{x}_i\|_\infty\mathbb{E}_n[K_\varpi(W)]L_\eta|\tau - \tau'|. \end{aligned} \tag{F.70}$$

Define a net $\widehat{\mathcal{T}} = \{\tau_1, \ldots, \tau_T\}$ such that

$$|\tau_{k+1} - \tau_k| \leqslant \left\{2\sqrt{n}\frac{\max_{i\leqslant n}\|\tilde{x}_i\|_\infty}{\underline{N}}L_\eta\right\}^{-1}.$$

To bound the second term in (F.69), note that $\mathcal{W}$ is a VC-class. Therefore, by Corollary 2.6.3 in [72] we have that conditional on $(W_i)_{i=1}^n$, there are at most $n^{d_W}$ different sets $\varpi \in \mathcal{W}$ that induce a different sequence $\{K_\varpi(W_1), \ldots, K_\varpi(W_n)\}$. Thus we can choose a (data-dependent) cover $\widehat{\mathcal{W}}$ with at most $n^{d_W}$ values of $\varpi$. Further, similarly to (F.71) we have $\|\eta_{\tilde{\tau}\varpi} - \eta_{\tilde{\tau}\tilde{\varpi}}\|_1 \leqslant L_\eta\|\varpi - \tilde{\varpi}\|^\rho$ and

$$\begin{aligned} \left|\mathbb{G}_n^o\left(\tfrac{g_{\tilde{\tau}\varpi}(\tilde{x}'\delta)}{t} - \tfrac{g_{\tilde{\tau}\tilde{\varpi}}(\tilde{x}'\delta)}{t}\right)\right| &\leqslant \tfrac{2\sqrt{n}}{t}\mathbb{E}_n[K_\varpi(W)|\tilde{x}'(\eta_{\tilde{\tau}\varpi} - \eta_{\tilde{\tau}\tilde{\varpi}})|] + \tfrac{2\sqrt{n}}{t}\mathbb{E}_n[|K_\varpi(W) - K_{\tilde{\varpi}}(W)||\tilde{x}'\delta|] \\ &\leqslant \tfrac{2\sqrt{n}}{\underline{N}}\max_{i\leqslant n}\|\tilde{x}_i\|_\infty\mathbb{E}_n[K_\varpi(W)]\|\eta_{\tilde{\tau}\varpi} - \eta_{\tilde{\tau}\tilde{\varpi}}\|_1 \\ &\leqslant \tfrac{2\sqrt{n}}{\underline{N}}\max_{i\leqslant n}\|\tilde{x}_i\|_\infty\mathbb{E}_n[K_\varpi(W)]L_\eta\|\varpi - \tilde{\varpi}\|^\rho. \end{aligned} \tag{F.71}$$

We define a net $\widehat{\mathcal{W}}$ such that $|\widehat{\mathcal{W}}| \leqslant n^{d_W} + \left\{2\sqrt{n}\frac{\max_{i\leqslant n}\|\tilde{x}_i\|_\infty}{\underline{N}}L_\eta\right\}^{d_W/\rho}$

To bound the third term in (F.69), note that for any $\delta \in \mathcal{F}_{t,\tau,\varpi}$, $t \leqslant \tilde{t}$, by considering $\tilde{\delta} := \delta(\tilde{t}/t) \in \mathcal{F}_{\tilde{t},\tau,\varpi}$ we have

$$\begin{aligned}\left|\mathbb{G}_n^o\left(\tfrac{g_{\tau\varpi}(\tilde{x}'\delta)}{t} - \tfrac{g_{\tau\varpi}(\tilde{x}'\delta(\tilde{t}/t))}{\tilde{t}}\right)\right| &\leqslant \left|\mathbb{G}_n^o\left(\tfrac{g_{\tau\varpi}(\tilde{x}'\delta)}{t} - \tfrac{g_{\tau\varpi}(\tilde{x}'\delta(\tilde{t}/t))}{t}\right)\right| + \left|\mathbb{G}_n^o\left(\tfrac{g_{\tau\varpi}(\tilde{x}'\delta(\tilde{t}/t))}{t} - \tfrac{g_{\tau\varpi}(\tilde{x}'\delta(\tilde{t}/t))}{\tilde{t}}\right)\right| \\ &= \tfrac{1}{t}\left|\mathbb{G}_n^o\left(g_{\tau\varpi}(\tilde{x}'\delta) - g_{\tau\varpi}(\tilde{x}'\delta[\tilde{t}/t])\right)\right| + \left|\mathbb{G}_n^o\left(g_{\tau\varpi}(\tilde{x}'\delta(\tilde{t}/t))\right)\right| \cdot \left|\tfrac{1}{t} - \tfrac{1}{\tilde{t}}\right| \\ &\leqslant \sqrt{n}\mathbb{E}_n\left(\tfrac{|K_\varpi(W)\tilde{x}'\delta|}{t}\right)\tfrac{|t-\tilde{t}|}{t} + \sqrt{n}\mathbb{E}_n\left(|K_\varpi(W)\tilde{x}'\delta|\right)\tfrac{\tilde{t}}{t}\left|\tfrac{1}{t} - \tfrac{1}{\tilde{t}}\right| \\ &= 2\sqrt{n}\mathbb{E}_n\left(\tfrac{|K_\varpi(W)\tilde{x}'\delta|}{t}\right)\left|\tfrac{t-\tilde{t}}{t}\right| \leqslant 2\sqrt{n}\left|\tfrac{t-\tilde{t}}{t}\right|.\end{aligned}$$

We let $\widehat{\mathcal{N}}$ be a $\varepsilon$-net $\{\underline{N} =: t_1, t_2, \ldots, t_K := \bar{N}\}$ of $[\underline{N}, \bar{N}]$ such that $|t_k - t_{k+1}|/t_k \leqslant 1/(2\sqrt{n})$. Note that we can achieve that with $|\widehat{\mathcal{N}}| \leqslant 1 + \lfloor 3\sqrt{n}\log(\bar{N}/\underline{N})\rfloor$.

By Markov bound, we have

$$\begin{aligned}\mathrm{P}(\mathcal{A}^o \geqslant K) &\leqslant \min_{\psi\geqslant 0}\exp(-\psi K)\mathrm{E}[\exp(\psi\mathcal{A}^o)] \\ &\leqslant 8p|\widehat{\mathcal{T}}|\cdot|\widehat{\mathcal{W}}|\cdot|\widehat{\mathcal{N}}|\min_{\psi\geqslant 0}\exp(-\psi K)\exp\left(8\psi^2\right) \\ &\leqslant 8p|\widehat{\mathcal{T}}|\cdot|\widehat{\mathcal{W}}|\cdot|\widehat{\mathcal{N}}|\exp(-K^2/32)\end{aligned}$$

here we set $\psi = K/16$ and bound $\mathrm{E}[\exp(\psi\mathcal{A}^o)]$ as follows

$$\begin{aligned}\mathrm{E}\left[\exp\left(\psi\mathcal{A}^o\right)\right] &\leqslant_{(1)} 2|\widehat{\mathcal{T}}|\cdot|\widehat{\mathcal{W}}|\cdot|\widehat{\mathcal{N}}|\sup_{(\tau,\varpi,t)\in\widehat{\mathcal{T}}\times\widehat{\mathcal{W}}\times\widehat{\mathcal{N}}}\mathrm{E}\left[\exp\left(\psi\sup_{\|\delta\|_{1,\varpi}=t}\mathbb{G}_n^o\left(\frac{g_{\tau\varpi}(\tilde{x}'\delta)}{t}\right)\right)\right] \\ &\leqslant_{(2)} 2|\widehat{\mathcal{T}}|\cdot|\widehat{\mathcal{W}}|\cdot|\widehat{\mathcal{N}}|\sup_{(\tau,\varpi,t)\in\widehat{\mathcal{T}}\times\widehat{\mathcal{W}}\times\widehat{\mathcal{N}}}\mathrm{E}\left[\exp\left(2\psi\sup_{\|\delta\|_{1,\varpi}=t}\mathbb{G}_n^o\left(\frac{K_\varpi(W)\tilde{x}'\delta}{t}\right)\right)\right] \\ &\leqslant_{(3)} 2|\widehat{\mathcal{T}}|\cdot|\widehat{\mathcal{W}}|\cdot|\widehat{\mathcal{N}}|\sup_{(\tau,\varpi,t)\in\widehat{\mathcal{T}}\times\widehat{\mathcal{W}}\times\widehat{\mathcal{N}}}\mathrm{E}\left[\exp\left(2\psi\left[\sup_{\|\delta\|_{1,\varpi}=t}\frac{\|\delta\|_{1,\varpi}}{t}\max_{j\leqslant p}\frac{|\mathbb{G}_n^o(K_\varpi(W)\tilde{x}_j)|}{\{\mathbb{E}_n[K_\varpi(W)\tilde{x}_j^2]\}^{1/2}}\right]\right)\right] \\ &=_{(4)} 2|\widehat{\mathcal{T}}|\cdot|\widehat{\mathcal{W}}|\cdot|\widehat{\mathcal{N}}|\sup_{(\tau,\varpi,t)\in\widehat{\mathcal{T}}\times\widehat{\mathcal{W}}\times\widehat{\mathcal{N}}}\mathrm{E}\left[\exp\left(2\psi\left[\max_{j\leqslant p}\frac{|\mathbb{G}_n^o(K_\varpi(W)\tilde{x}_j)|}{\{\mathbb{E}_n[K_\varpi(W)\tilde{x}_j^2]\}^{1/2}}\right]\right)\right] \\ &\leqslant_{(5)} 4p|\widehat{\mathcal{T}}|\cdot|\widehat{\mathcal{W}}|\cdot|\widehat{\mathcal{N}}|\max_{j\leqslant p}\sup_{\varpi\in\widehat{\mathcal{W}}}\mathrm{E}\left[\exp\left(4\psi\frac{\mathbb{G}_n^o(K_\varpi(W)\tilde{x}_j)}{\{\mathbb{E}_n[K_\varpi(W)\tilde{x}_j^2]\}^{1/2}}\right)\right] \\ &\leqslant_{(6)} 8p|\widehat{\mathcal{T}}|\cdot|\widehat{\mathcal{W}}|\cdot|\widehat{\mathcal{N}}|\exp\left(8\psi^2\right)\end{aligned}$$

here (1) follows by $\exp(\max_{i\in I}|z_i|) \leqslant 2|I|\max_{i\in I}\exp(z_i)$, (2) by contraction principle (apply Theorem 4.12 [54] with $t_i = K_\varpi(W_i)\tilde{x}_i'\delta$, and $\phi_i(t_i) = \rho_\tau(K_\varpi(W_i)\tilde{y}_i - K_\varpi(W_i)\tilde{x}_i'\eta_\tau + t_i) - \rho_\tau(K_\varpi(W_i)\tilde{y}_i - K_\varpi(W_i)\tilde{x}_i'\eta_\tau)$ so that $|\phi_i(s) - \phi_i(t)| \leqslant |s-t|$ and $\phi_i(0) = 0$, (3) follows by

$$|\mathbb{G}_n^o(K_\varpi(W)\tilde{x}'\delta)| \leqslant \|\delta\|_{1,\varpi}\max_{j\leqslant p}|\mathbb{G}_n^o(K_\varpi(W)\tilde{x}_j)/\{\mathbb{E}_n[K_\varpi(W)\tilde{x}_j^2]\}^{1/2}|,$$

(4) by the definition of suprema, (5) we again use $\exp(\max_{i\in I}|z_i|) \leqslant 2|I|\max_{i\in I}\exp(z_i)$, and (6) $\exp(z) + \exp(-z) \leqslant 2\exp(z^2/2)$.

■

**Lemma 14** (Estimation Error of Refitted Quantile Regression)**.** *Consider an arbitrary vector $\widehat{\eta}_u$ and suppose $\|\eta_u\|_0 \leqslant s$. Let $\|r_{iu} \leqslant \bar{r}_u\|_{n,\varpi}$, $|\mathrm{support}(\widehat{\eta}_u)| \leqslant \widehat{s}_u$ and $\mathbb{E}_n[K_\varpi(W)\{\rho_\tau(\tilde{y}_i - \tilde{x}_i'\widehat{\eta}_u) - \rho_\tau(\tilde{y}_i - \tilde{x}_i'\eta_u)\}] \leqslant \widehat{Q}_u$ for all $u \in \mathcal{U}$ hold. Furthermore, suppose that*

$$\sup_{u=(\tau,\varpi)\in\mathcal{U}}\left|\mathbb{E}_n\left(K_\varpi(W)\frac{\rho_\tau(\tilde{y} - \tilde{x}'\widetilde{\eta}_u) - \rho_\tau(\tilde{y} - \tilde{x}'\eta_u)}{\|\widetilde{\eta}_u - \eta_u\|_{1,\varpi}} - \mathrm{E}\left[K_\varpi(W)\frac{\rho_\tau(\tilde{y} - \tilde{x}'\widetilde{\eta}_u) - \rho_\tau(\tilde{y} - \tilde{x}'\eta_u)}{\|\widetilde{\eta}_u - \eta_u\|_{1,\varpi}} \mid W, \tilde{x}\right]\right)\right| \leqslant \frac{t_3}{\sqrt{n}}.$$

*Under these events, we have for $n$ large enough,*

$$\|\sqrt{f_u}\tilde{x}_i'(\widetilde{\eta}_u - \eta_u)\|_{n,\varpi} \lesssim \widetilde{N}_u := \sqrt{\frac{(\widehat{s}_u + s)}{\phi_{\min}(u, \widehat{s}_u + s)}}(K_{n1} + t_3/\sqrt{n}) + K_{n2} + \bar{f}\bar{r}_u + \widehat{Q}_u^{1/2}$$

*where* $\phi_{\min}(u,k) = \inf_{\|\delta\|_0=k} \|\sqrt{f_u}\tilde{x}'\delta\|_{n,\varpi}^2/\|\delta\|^2$, *provided that*

$$\sup_{u\in\mathcal{U},\|\bar{\delta}\|_0\leqslant\widehat{s}_u+s} \frac{\bar{f}'\mathbb{E}_n[K_\varpi(W)(|r_u|+|r_u|^2)|\tilde{x}'\bar{\delta}|^2]}{\mathbb{E}_n[K_\varpi(W)f_u|\tilde{x}'\bar{\delta}|^2]} + \widetilde{N}_u/\bar{q}_{A_u} \to 0. \tag{F.72}$$

*where* $A_u = \{\delta\in\mathbb{R}^p : \|\delta\|_0 \leqslant \widehat{s}_u + s\}$.

*Proof of Lemma 14.* Let $\widehat{\delta}_u = \widehat{\eta}_u - \eta_u$ which satisfies $\|\widehat{\delta}_u\|_0 \leqslant \widehat{s}_u + s$. By optimality of $\widetilde{\eta}_u$ in the refitted quantile regression we have

$$\begin{array}{c}\mathbb{E}_n[K_\varpi(W)\rho_\tau(\tilde{y}_i - \tilde{x}_i'\widetilde{\eta}_u)] - \mathbb{E}_n[K_\varpi(W)\rho_\tau(\tilde{y}_i - \tilde{x}_i'\eta_u)] \\ \leqslant \mathbb{E}_n[K_\varpi(W)\rho_\tau(\tilde{y}_i - \tilde{x}_i'\widehat{\eta}_u)] - \mathbb{E}_n[K_\varpi(W)\rho_\tau(\tilde{y}_i - \tilde{x}_i'\eta_u)] \leqslant \widehat{Q}_u\end{array} \tag{F.73}$$

where the second inequality holds by assumption.

Moreover, by assumption, uniformly over $u\in\mathcal{U}$, we have conditional on $(W_i,\tilde{x}_i,r_{iu})_{i=1}^n$ that

$$\left|\mathbb{G}_n\left(K_\varpi(W)\frac{\rho_\tau(\tilde{y}-\tilde{x}'(\eta_u+\widetilde{\delta}_u)) - \rho_\tau(\tilde{y}-\tilde{x}'\eta_u)}{\|\widetilde{\delta}_u\|_{1,\varpi}}\right)\right| \leqslant t_3. \tag{F.74}$$

Thus combining relations (F.73) and (F.74), we have

$$\mathbb{E}_n[\mathrm{E}[K_\varpi(W)\{\rho_u(\tilde{y}-\tilde{x}'(\eta_u+\widetilde{\delta}_u)) - \rho_u(\tilde{y}-\tilde{x}'\eta_u)\}|\tilde{x},\tilde{r},\varpi]] \leqslant \|\widetilde{\delta}_u\|_{1,\varpi}t_3/\sqrt{n} + \widehat{Q}_u.$$

Invoking the sparse identifiability relation of Lemma 12, since the required condition on the approximation errors $r_u$'s holds by assumption (F.72), for $n$ large enough

$$\frac{\|\sqrt{f_u}\tilde{x}'\widetilde{\delta}_u\|_{n,\varpi}^2}{4} \wedge \left\{\bar{q}_{A_u}\|\sqrt{f_u}\tilde{x}'\widetilde{\delta}_u\|_{n,\varpi}\right\} \leqslant K_{n2}\|\sqrt{f_u}\tilde{x}'\widetilde{\delta}_u\|_{n,\varpi} + \|\widetilde{\delta}_u\|_{1,\varpi}(K_{n1}+t_3/\sqrt{n}) + \widehat{Q}_u,$$

where $\bar{q}_{A_u}$ is defined with $A_u := \{\delta : \|\delta\|_0 \leqslant \widehat{s}_u + s\}$. Moreover, by the sparsity of $\tilde{\delta}_u$ we have $\|\widetilde{\delta}_u\|_{1,\varpi} \leqslant \sqrt{(\widehat{s}_u+s)/\phi_{\min}(u,\widehat{s}_u+s)}\|\sqrt{f_u}\tilde{x}'\widetilde{\delta}_u\|_{n,\varpi}$ so that we have for $t = \|\sqrt{f_u}\tilde{x}'\widetilde{\delta}_u\|_{n,\varpi}$,

$$\frac{t^2}{4}\wedge\{\bar{q}_{A_u}t\} \quad\leqslant\quad t(K_{n2} + \sqrt{(\widehat{s}_u+s)/\phi_{\min}(u,\widehat{s}_u+s)}\{K_{n1}+t_3/\sqrt{n}\}) + \widehat{Q}_u.$$

Note that for positive numbers $(t^2/4)\wedge(\bar{q}_{A_u}t) \leqslant A + Bt$ implies $t^2/4 \leqslant A + Bt$ provided $\bar{q}_{A_u}/2 > B$ and $2\bar{q}_{A_u}^2 > A$. (Indeed, otherwise $(t^2/4) \geqslant qt$ so that $t \geqslant 4q$, which in turn implies that $2\bar{q}_{A_u}^2 + \bar{q}_{A_u}t/2 \leqslant (t^2/4)\wedge\bar{q}_{A_u}t \leqslant A + Bt$.) Note that $\bar{q}_{A_u}/2 > B$ and $2\bar{q}_{A_u}^2 > A$ is implied by condition (F.72) when we set $A = \widehat{Q}_u$ and $B = (K_{n2} + \sqrt{(\widehat{s}_u+s)/\phi_{\min}(u,\widehat{s}_u+s_u)}\{K_{1n}+t_3/\sqrt{n}\})$. Thus the minimum is achieved in the quadratic part. Therefore, for $n$ sufficiently large, we have

$$\|\sqrt{f_u}\tilde{x}'\widetilde{\delta}_u\|_{n,\varpi} \leqslant \widehat{Q}_u^{1/2} + K_{n2} + (K_{n1}+t_3/\sqrt{n})\sqrt{(\widehat{s}_u+s)/\phi_{\min}(u,\widehat{s}_u+s_u)}.$$

■

Under the condition $\max_{i\leqslant n}\|\tilde{x}_i\|_\infty^2\log(n\vee p) = o(n\min_{\tau\in\mathcal{T}}\tau(1-\tau))$, the next result provides new bounds for the data driven penalty choice parameter when the quantile indices in $\mathcal{T}$ can approach the extremes.

**Lemma 15** (Pivotal Penalty Parameter Bound)**.** *Let* $\underline{\tau} = \min_{\tau\in\mathcal{T}}\tau(1-\tau)$ *and* $K_n = \max_{i\leqslant n, j\in[p]}|\tilde{x}_{ij}/\widehat{\sigma}_j|$, $\widehat{\sigma}_j = \mathbb{E}_n[\tilde{x}_j^2]^{1/2}$. *Under* $K_n^2\log(p/\underline{\tau}) = o(n\underline{\tau})$, *for* $n$ *large enough we have that for some constant* $\bar{C}$

$$\Lambda(1-\xi|\tilde{x}_1,\ldots,\tilde{x}_n) \leqslant \bar{C}\sqrt{\frac{\log(16p/(\underline{\tau}\xi))}{n}}$$

*where* $\Lambda(1-\xi|\tilde{x}_1,\ldots,\tilde{x}_n)$ *is the* $1-\xi$ *quantile of* $\max_{j\in[p]}\sup_{\tau\in\mathcal{T}}\left|\frac{\sum_{i=1}^n \tilde{x}_{ij}(\tau-1\{U_i\leqslant\tau\})}{\widehat{\sigma}_j\sqrt{\tau(1-\tau)}}\right|$ *conditional on* $\tilde{x}_1,\ldots,\tilde{x}_n$*, and* $U_i$ *are independent uniform*$(0,1)$ *random variables.*

*Proof.* Conditional on $\tilde{x}_1,\ldots,\tilde{x}_n$, letting $\widehat{\sigma}_j^2=\mathbb{E}_n[x_j^2]$, we have that

$$n\Lambda=\max_{j\in[p]}\sup_{\tau\in\mathcal{T}}\left|\frac{\sum_{i=1}^n \tilde{x}_j(\tau-1\{U\leqslant\tau\})}{\widehat{\sigma}_j\sqrt{\tau(1-\tau)}}\right|.$$

Step 1. (Entropy Calculation) Let $\mathcal{F}=\{\tilde{x}_{ij}(\tau-1\{U_i\leqslant\tau\})/\widehat{\sigma}_j:\tau\in\mathcal{T},j\in[p]\}$, $h_\tau=\sqrt{\tau(1-\tau)}$, and $\mathcal{G}=\{f_\tau/h_\tau:\tau\in\mathcal{T}\}$. We have that

$$\begin{aligned}d(f_\tau/h_\tau,f_{\bar\tau}/h_{\bar\tau}) &\leqslant d(f_\tau,f_{\bar\tau})/h_\tau+d(f_{\bar\tau}/h_\tau,f_{\bar\tau}/h_{\bar\tau})\\ &\leqslant d(f_\tau,f_{\bar\tau})/h_\tau+d(0,f_{\bar\tau}/h_{\bar\tau})|h_\tau-h_{\bar\tau}|/h_\tau\end{aligned}$$

Therefore, since $\|F\|_Q\leqslant\|G\|_Q$ by $h_\tau\leqslant 1$, and $d(0,f_{\bar\tau}/h_{\bar\tau})\leqslant 1/h_{\bar\tau}$ we have

$$N(\epsilon\|G\|_Q,\mathcal{G},Q)\leqslant N(\epsilon\|F\|_Q/\{2\min_{\tau\in\mathcal{T}}h_\tau\},\mathcal{F},Q)N(\epsilon/\{2\min_{\tau\in\mathcal{T}}h_\tau^2\},\mathcal{T},|\cdot|).$$

Thus we have for some constants $K$ and $v$ that

$$N(\epsilon\|G\|_Q,\mathcal{G},Q)\leqslant p(K/\{\epsilon\min_{\tau\in\mathcal{T}}h_\tau^2\})^v.$$

Step 2.(Symmetrization) Since we have $\mathrm{E}[g^2]=1$ for all $g\in\mathcal{G}$, by Lemma 2.3.7 in [72] we have

$$\mathrm{P}(\Lambda\geqslant t\sqrt{n})\leqslant 4\mathrm{P}(\max_{j\leqslant p}\sup_{\tau\in\mathcal{T}}|\mathbb{G}_n^o(g)|\geqslant t/4)$$

here $\mathbb{G}_n^o:\mathcal{G}\to\mathbb{R}$ is the symmetrized process generated by Rademacher variables. Conditional on $(x_1,u_1),\ldots,(x_n,u_n)$, we have that $\{\mathbb{G}_n^o(g):g\in\mathcal{G}\}$ is sub-Gaussian with respect to the $L_2(\mathbb{P}_n)$-norm by the Hoeffding inequality. Thus, by Lemma 16 in [11], for $\delta_n^2=\sup_{g\in\mathcal{G}}\mathbb{E}_n[g^2]$ and $\bar\delta_n=\delta_n/\|G\|_{\mathbb{P}_n}$, we have

$$\mathrm{P}(\sup_{g\in\mathcal{G}}|\mathbb{G}_n^o(g)|>CK\delta_n\sqrt{\log(pK/\underline{\tau})}\mid\{\tilde{x}_i,U_i\}_{i=1}^n)\leqslant\int_0^{\bar\delta_n/2}\epsilon^{-1}\{p(K/\{\epsilon\min_{\tau\in\mathcal{T}}h_\tau^2\})^v\}^{-C^2+1}d\epsilon$$

for some universal constant $K$.

In order to control $\delta_n$, note that $\delta_n^2=\sup_{g\in\mathcal{G}}\frac{1}{\sqrt{n}}\mathbb{G}_n(g^2)+\mathrm{E}[g^2]$. In turn, since $\sup_{g\in\mathcal{G}}\mathbb{E}_n[g^4]\leqslant\delta_n^2\max_{i\leqslant n}G_i^2$, we have

$$\mathrm{P}(\sup_{g\in\mathcal{G}}|\mathbb{G}_n^o(g^2)|>C\bar{K}\delta_n\max_{i\leqslant n}G_i\sqrt{\log(pK/\underline{\tau})}\mid\{\tilde{x}_i,U_i\}_{i=1}^n)\leqslant\int_0^{\bar\delta_n/2}\epsilon^{-1}\{p(K/\{\epsilon\underline{\tau}\})^v\}^{-C^2+1}d\epsilon.$$

Thus with probability $1-\int_0^{1/2}\epsilon^{-1}\{p(K/\epsilon\underline{\tau})^v\}^{-C^2+1}d\epsilon$, since $\mathrm{E}[g^2]=1$ and $\max_{i\leqslant n}G_i\leqslant K_n/\sqrt{\underline{\tau}}$, we have

$$\delta_n\leqslant 1+\frac{C'K_n\sqrt{\log(pK/\underline{\tau})}}{\sqrt{n}\sqrt{\underline{\tau}}}.$$

Therefore, under $K_n\sqrt{\log(pK/\underline{\tau})}=o(\sqrt{n}\sqrt{\underline{\tau}})$, conditionally on $\{\tilde{x}_i\}_{i=1}^n$ and $n$ sufficiently large, with probability $1-2\int_0^{1/2}\epsilon^{-1}\{p(K/\{\epsilon\underline{\tau}\})^v\}^{-C^2+1}d\epsilon$ we have that

$$\sup_{g\in\mathcal{G}}|\mathbb{G}_n^o(g)|\leqslant 2CK\sqrt{\log(pK/\underline{\tau})}$$

The stated bound follows since for $C > 2$

$$2\int_0^{1/2} \epsilon^{-1}\{p(K/\{\epsilon\underline{\tau}\})^v\}^{-C^2+1}d\epsilon \leqslant \{p/\underline{\tau}\}^{-C^2+1}2\int_0^{1/2}\epsilon^{-2+C^2}d\epsilon \leqslant \{p/\underline{\tau}\}^{-C^2+1}.$$

■

## Appendix G. Inequalities

**Lemma 16** (Transfer principle, [65])**.** *Let $\widehat{\Sigma}$ and $\Sigma$ be $p \times p$ matrices with non-negative diagonal entries, and assume that for some $\eta \in (0,1)$ and $s \leqslant p$ we have*

$$\forall v \in \mathbb{R}^p, \|v\|_0 \leqslant s, v'\widehat{\Sigma}v \geqslant (1-\eta)v'\Sigma v$$

*Let $D$ be a diagonal matrix such that $D_{kk} \geqslant \widehat{\Sigma}_{kk} - (1-\eta)\Sigma_{kk}$. Then for all $\delta \in \mathbb{R}^p$ we have*

$$\delta'\widehat{\Sigma}\delta \geqslant (1-\eta)\delta'\Sigma\delta - \|D^{1/2}\delta\|_1^2/(s-1).$$

**Lemma 17.** *Consider $\widehat{\beta}_u$ and $\beta_u$ with $\|\beta_u\|_0 \leqslant s$. Denote by $\widehat{\beta}_u^\lambda$ the vector with $\widehat{\beta}_{uj}^\lambda = \widehat{\beta}_{uj}1\{\widehat{\sigma}_{a\varpi j}^Z|\widehat{\beta}_{uj}| \geqslant \lambda\}$ where $\widehat{\sigma}_{a\varpi j}^Z = \{\mathbb{E}_n[K_\varpi(W)(Z_j^a)^2]\}^{1/2}$. We have that*

$$\begin{aligned}
\|\widehat{\beta}_u^\lambda - \beta_u\|_{1,\varpi} &\leqslant \|\widehat{\beta}_u - \beta_u\|_{1,\varpi} + \lambda s \\
|\text{support}(\widehat{\beta}_u^\lambda)| &\leqslant s + \|\widehat{\beta}_u - \beta_u\|_{1,\varpi}/\lambda \\
\|Z^a(\widehat{\beta}_u^\lambda - \beta_u)\|_{n,\varpi} &\leqslant \|Z^a(\widehat{\beta}_u - \beta_u)\|_{n,\varpi} + \sqrt{\tilde{\phi}_{\max}(s,\varpi)}\{2\sqrt{s}\lambda + \|\widehat{\beta}_u - \beta_u\|_{1,\varpi}/\sqrt{s}\}
\end{aligned}$$

*here $\tilde{\phi}_{\max}(m,\varpi) = \sup_{1\leqslant\|\theta\|_0\leqslant m}\|\tilde{Z}^a\theta\|_{n,\varpi}/\|\theta\|$ and $\tilde{Z}_{ij}^a = Z_{ij}^a/\{\mathbb{E}_n[K_\varpi(W)(Z_j^a)^2]\}^{1/2}$.*

*Proof.* Let $T_u = \text{support}(\beta_u)$. The first relation follows from the triangle inequality

$$\begin{aligned}
\|\widehat{\beta}_u^\lambda - \beta_u\|_{1,\varpi} &= \|(\widehat{\beta}_u^\lambda - \beta_u)_{T_u}\|_{1,\varpi} + \|(\widehat{\beta}_u^\lambda)_{T_u^c}\|_{1,\varpi} \\
&\leqslant \|(\widehat{\beta}_u^\lambda - \widehat{\beta}_u)_{T_u}\|_{1,\varpi} + \|(\widehat{\beta}_u - \beta_u)_{T_u}\|_{1,\varpi} + \|(\widehat{\beta}_u^\lambda)_{T_u^c}\|_{1,\varpi} \\
&\leqslant \lambda s + \|(\widehat{\beta}_u - \beta_u)_{T_u}\|_{1,\varpi} + \|(\widehat{\beta}_u)_{T_u^c}\|_{1,\varpi} \\
&= \lambda s + \|\widehat{\beta}_u - \beta_u\|_{1,\varpi}
\end{aligned}$$

To show the second result note that $\|\widehat{\beta}_u - \beta_u\|_{1,\varpi} \geqslant \{|\text{support}(\widehat{\beta}_u^\lambda)| - s\}\lambda$. Therefore,

$$|\text{support}(\widehat{\beta}_u^\lambda)| \leqslant s + \|\widehat{\beta}_u - \beta_u\|_{1,\varpi}/\lambda$$

which yields the result.

To show the third bound, we start using the triangle inequality

$$\|Z^a(\widehat{\beta}_u^\lambda - \beta_u)\|_{n,\varpi} \leqslant \|Z^a(\widehat{\beta}_u^\lambda - \widehat{\beta}_u)\|_{n,\varpi} + \|Z^a(\widehat{\beta}_u - \beta_u)\|_{n,\varpi}.$$

Without loss of generality, assume that the components are ordered so that $|(\widehat{\beta}_u^\lambda - \widehat{\beta}_u)_j|\widehat{\sigma}_{uj}$ is decreasing. Let $T_1$ be the set of $s$ indices corresponding to the largest values of $|(\widehat{\beta}_u^\lambda - \widehat{\beta}_u)_j|\widehat{\sigma}_{uj}$. Similarly define $T_k$ as the set of $s$ indices corresponding to the largest values of $|(\widehat{\beta}_u^\lambda - \widehat{\beta}_u)_j|\widehat{\sigma}_{uj}$ outside $\cup_{m=1}^{k-1}T_m$. Therefore,

$\widehat{\beta}_u^\lambda - \widehat{\beta}_u = \sum_{k=1}^{\lceil p/s \rceil} (\widehat{\beta}_u^\lambda - \widehat{\beta}_u)_{T_k}$. Moreover, given the monotonicity of the components, $\|(\widehat{\beta}_u^\lambda - \widehat{\beta}_u)_{T_k}\|_{2,\varpi} \leqslant \|(\widehat{\beta}_u^\lambda - \widehat{\beta}_u)_{T_{k-1}}\|_1/\sqrt{s}$. Then, we have

$$\begin{aligned}
\|Z^a(\widehat{\beta}_u^\lambda - \widehat{\beta}_u)\|_{n,\varpi} &= \|Z^a \textstyle\sum_{k=1}^{\lceil p/s \rceil} (\widehat{\beta}_u^\lambda - \widehat{\beta}_u)_{T_k}\|_{n,\varpi} \\
&\leqslant \|Z^a(\widehat{\beta}_u^\lambda - \widehat{\beta}_u)_{T_1}\|_{n,\varpi} + \textstyle\sum_{k \geqslant 2} \|Z^a(\widehat{\beta}_u^\lambda - \widehat{\beta}_u)_{T_k}\|_{n,\varpi} \\
&\leqslant \sqrt{\tilde{\phi}_{\max}(s,\varpi)}\|(\widehat{\beta}_u^\lambda - \widehat{\beta}_u)_{T_1}\|_{2,\varpi} + \sqrt{\tilde{\phi}_{\max}(s,\varpi)} \textstyle\sum_{k \geqslant 2} \|(\widehat{\beta}_u^\lambda - \widehat{\beta}_u)_{T_k}\|_{2,\varpi} \\
&\leqslant \sqrt{\tilde{\phi}_{\max}(s,\varpi)}\lambda\sqrt{s} + \sqrt{\tilde{\phi}_{\max}(s,\varpi)} \textstyle\sum_{k \geqslant 1} \|(\widehat{\beta}_u^\lambda - \widehat{\beta}_u)_{T_k}\|_{1,\varpi}/\sqrt{s} \\
&= \sqrt{\tilde{\phi}_{\max}(s,\varpi)}\lambda\sqrt{s} + \sqrt{\tilde{\phi}_{\max}(s,\varpi)}\|\widehat{\beta}_u^\lambda - \widehat{\beta}_u\|_{1,\varpi}/\sqrt{s} \\
&\leqslant \sqrt{\tilde{\phi}_{\max}(s,\varpi)}\{2\lambda\sqrt{s} + \|\widehat{\beta}_u - \beta_u\|_{1,\varpi}/\sqrt{s}\}
\end{aligned}$$

here the last inequality follows from the first result and the triangle inequality.

■

**Lemma 18** (Supremum of Sparse Vectors on Symmetrized Random Matrices)**.** *Let $\widehat{\mathcal{U}}$ denote a finite set and $(X_{iu})_{u \in \widehat{\mathcal{U}}}$, $i = 1, \ldots, n$, be fixed vectors such that $X_{iu} \in \mathbb{R}^p$ and $\max_{1 \leqslant i \leqslant n} \max_{u \in \widehat{\mathcal{U}}} \|X_{iu}\|_\infty \leqslant K$. Furthermore define*

$$\delta_n := \bar{C}K\sqrt{k}\left(\sqrt{\log|\widehat{\mathcal{U}}|} + \sqrt{1+\log p} + \log k\sqrt{\log(p \vee n)}\sqrt{\log n}\right)/\sqrt{n},$$

*where $\bar{C}$ is a universal constant. Then,*

$$\mathrm{E}\left[\sup_{\|\theta\|_0 \leqslant k, \|\theta\|=1} \max_{u \in \widehat{\mathcal{U}}} \left|\mathbb{E}_n[\varepsilon(\theta' X_u)^2]\right|\right] \leqslant \delta_n \sup_{\|\theta\|_0 \leqslant k, \|\theta\|=1, u \in \widehat{\mathcal{U}}} \sqrt{\mathbb{E}_n[(\theta' X_u)^2]}.$$

*Proof.* See [14] for the proof. ■

**Corollary 2** (Supremum of Sparse Vectors on Many Random Matrices)**.** *Let $\widehat{\mathcal{U}}$ denote a finite set and $(X_{iu})_{u \in \widehat{\mathcal{U}}}$, $i = 1, \ldots, n$, be independent (across $i$) random vectors such that $X_{iu} \in \mathbb{R}^p$ and*

$$\sqrt{\mathrm{E}[\max_{1 \leqslant i \leqslant n} \max_{u \in \widehat{\mathcal{U}}} \|X_{iu}\|_\infty^2]} \leqslant K.$$

*Furthermore define*

$$\delta_n := \bar{C}K\sqrt{k}\left(\sqrt{\log|\widehat{\mathcal{U}}|} + \sqrt{1+\log p} + \log k\sqrt{\log(p \vee n)}\sqrt{\log n}\right)/\sqrt{n},$$

*here $\bar{C}$ is a universal constant. Then,*

$$\mathrm{E}\left[\sup_{\|\theta\|_0 \leqslant k, \|\theta\|=1} \max_{u \in \widehat{\mathcal{U}}} \left|\mathbb{E}_n\left[(\theta' X_u)^2 - \mathrm{E}[(\theta' X_u)^2]\right]\right|\right] \leqslant \delta_n^2 + \delta_n \sup_{\|\theta\|_0 \leqslant k, \|\theta\|=1, u \in \widehat{\mathcal{U}}} \sqrt{\mathbb{E}_n[\mathrm{E}[(\theta' X_u)^2]]}.$$

We will also use the following result of [25].

**Lemma 19** (Maximal Inequality)**.** *Work with the setup above. Suppose that $F \geqslant \sup_{f \in \mathcal{F}} |f|$ is a measurable envelope for $\mathcal{F}$ with $\|F\|_{P,q} < \infty$ for some $q \geqslant 2$. Let $M = \max_{i \leqslant n} F(W_i)$ and $\sigma^2 > 0$ be any positive constant such that $\sup_{f \in \mathcal{F}} \|f\|_{P,2}^2 \leqslant \sigma^2 \leqslant \|F\|_{P,2}^2$. Suppose that there exist constants $a \geqslant e$ and $v \geqslant 1$ such that*

$$\log \sup_Q N(\epsilon\|F\|_{Q,2}, \mathcal{F}, \|\cdot\|_{Q,2}) \leqslant v \log(a/\epsilon), \ 0 < \epsilon \leqslant 1.$$

*Then*

$$\mathrm{E}_P[\sup_{f\in\mathcal{F}}|\mathbb{G}_n(f)|] \leqslant K\left(\sqrt{v\sigma^2\log\left(\frac{a\|F\|_{P,2}}{\sigma}\right)}+\frac{v\|M\|_{P,2}}{\sqrt{n}}\log\left(\frac{a\|F\|_{P,2}}{\sigma}\right)\right),$$

*here $K$ is an absolute constant. Moreover, for every $t\geqslant 1$, with probability $>1-t^{-q/2}$,*

$$\sup_{f\in\mathcal{F}}|\mathbb{G}_n(f)| \quleq (1+\alpha)\mathrm{E}_P[\sup_{f\in\mathcal{F}}|\mathbb{G}_n(f)|] + K(q)\Big[(\sigma+n^{-1/2}\|M\|_{P,q})\sqrt{t}+\alpha^{-1}n^{-1/2}\|M\|_{P,2}t\Big],$$

*$\forall\alpha>0$ where $K(q)>0$ is a constant depends only on $q$. In particular, setting $a\geqslant n$ and $t=\log n$, with probability $>1-c(\log n)^{-1}$,*

$$\sup_{f\in\mathcal{F}}|\mathbb{G}_n(f)| \leqslant K(q,c)\left(\sigma\sqrt{v\log\left(\frac{a\|F\|_{P,2}}{\sigma}\right)}+\frac{v\|M\|_{P,q}}{\sqrt{n}}\log\left(\frac{a\|F\|_{P,2}}{\sigma}\right)\right), \tag{G.75}$$

*here $\|M\|_{P,q}\leqslant n^{1/q}\|F\|_{P,q}$ and $K(q,c)>0$ is a constant depending only on $q$ and $c$.*

## Appendix H. Confidence Regions for Function-Valued Parameters Based on Moment Conditions

For completeness, in this section we collect an adaptation of the results of [14] that are invoked in our proofs. The main difference is the weakening of the identification condition (which is allowed to decrease to zero, see the parameter $j_n$ in Condition C1 below). We are interested in function-valued target parameters indexed by $u\in\mathcal{U}\subset\mathbb{R}^{d_u}$. The true value of the target parameter is denoted by

$$\theta^0=(\theta_{uj})_{u\in\mathcal{U},j\in[\tilde{p}]},\quad \text{where}\quad \theta_{uj}\in\Theta_{uj}\text{ for each }u\in\mathcal{U}\quad\text{and}\quad j\in[\tilde{p}].$$

For each $u\in\mathcal{U}$ and $j\in[\tilde{p}]$, the parameter $\theta_{uj}$ is characterized as the solution to the following moment condition:

$$\mathrm{E}[\psi_{uj}(W_{uj},\theta_{uj},\eta_{uj})]=0, \tag{H.76}$$

where $W_{uj}$ is a random vector that takes values in a Borel set $\mathcal{W}_{uj}\subset\mathbb{R}^{d_w}$, $\eta^0=(\eta_{uj})_{u\in\mathcal{U},j\in[\tilde{p}]}$ is a nuisance parameter where $\eta_{uj}\in T_{uj}$ a convex set, and the moment function

$$\psi_{uj}:\mathcal{W}_{uj}\times\Theta_{uj}\times T_{ujn}\mapsto\mathbb{R},\quad (w,\theta,t)\mapsto\psi_{uj}(w,\theta,t) \tag{H.77}$$

is a Borel measurable map.

We assume that the (continuum) nuisance parameter $\eta^0$ can be modelled and estimated by $\widehat{\eta}=(\widehat{\eta}_{uj})_{u\in\mathcal{U},j\in[\tilde{p}]}$. We will discuss examples where the corresponding $\eta^0$ can be estimated using modern regularization and post-selection methods such as Lasso and Post-Lasso (although other procedures can be applied). The estimator $\check{\theta}_{uj}$ of $\theta_{uj}$ is constructed as any approximate $\epsilon_n$-solution in $\Theta_{uj}$ to a sample analog of the moment condition (H.76), i.e.,

$$\max_{j\in[\tilde{p}]}\sup_{u\in\mathcal{U}}\left\{|\mathbb{E}_n[\psi_{uj}(W_{uj},\check{\theta}_{uj},\widehat{\eta}_{uj})]|-\inf_{\theta_j\in\Theta_{uj}}|\mathbb{E}_n[\psi_{uj}(W_{uj},\theta,\widehat{\eta}_{uj})]|\right\}\leqslant\epsilon_n=o_P(n^{-1/2}\delta_n). \tag{H.78}$$

As discussed before, we rely on an orthogonality condition for regular estimation of $\theta_{uj}$, which we will state next.

**Definition 1** (**Near Orthogonality Condition**). For each $u \in \mathcal{U}$ and $j \in [\tilde{p}]$, we say that $\psi_{uj}$ obeys a general form of orthogonality with respect to $\mathcal{H}_{uj}$ uniformly in $u \in \mathcal{U}$, if the following conditions hold: the Gâteaux derivative map

$$\mathrm{D}_{u,j,\bar{r}}[\tilde{\eta}_{uj} - \eta_{uj}] := \partial_r \mathrm{E}\left( \psi_{uj}\Big\{ W_{uj}, \theta_{uj}, \eta_{uj} + r\Big[\tilde{\eta}_{uj} - \eta_{uj}\Big]\Big\}\right)\Big|_{r=\bar{r}}$$

exists for all $r \in [0,1)$, $\tilde{\eta} \in \mathcal{H}_{uj}$, $j \in \tilde{p}$, and $u \in \mathcal{U}$ and vanishes at $r = 0$, namely,

$$|\mathrm{D}_{u,j,0}[\tilde{\eta}_{uj} - \eta_{uj}]| \leqslant \delta_n n^{-1/2} \quad \text{for all } \tilde{\eta}_{uj} \in \mathcal{H}_{uj}. \tag{H.79}$$

In what follows, we shall denote by $c_0$, $c$, and $C$ some positive constants.

**Assumption C1** (Moment Condition). *Consider a random element $W$, taking values in a measure space $(\mathcal{W}, \mathcal{A}_{\mathcal{W}})$, with law determined by a probability measure $P \in \mathcal{P}_n$. The observed data $((W_{iu})_{u\in\mathcal{U}})_{i=1}^n$ consist of $n$ i.i.d. copies of a random element $(W_u)_{u\in\mathcal{U}}$ which is generated as a suitably measurable transformation with respect to $W$ and $u$. Uniformly for all $n \geqslant n_0$ and $P \in \mathcal{P}_n$, the following conditions hold: (i) The true parameter value $\theta_{uj}$ obeys (H.76) and is interior relative to $\Theta_{uj}$, namely there is a ball of radius $Cn^{-1/2}u_n \log n$ centered at $\theta_{uj}$ contained in $\Theta_{uj}$ for all $u \in \mathcal{U}$, $j \in [\tilde{p}]$ with $u_n := \mathrm{E}[\sup_{u\in\mathcal{U}, j\in[\tilde{p}]} |\sqrt{n}\mathbb{E}_n[\psi_{uj}(W_{uj}, \theta_{uj}, \eta_{uj})]|]$; (ii) For each $u \in \mathcal{U}$ and $j \in [\tilde{p}]$, the map $(\theta, \eta) \in \Theta_{uj} \times \mathcal{H}_{uj} \mapsto \mathrm{E}[\psi_{uj}(W_{uj}, \theta, \eta)]|$ is twice continuously differentiable; (iii) For all $u \in \mathcal{U}$ and $j \in [\tilde{p}]$, the moment function $\psi_{uj}$ obeys the orthogonality condition given in Definition 1 for the set $\mathcal{H}_{uj} = \mathcal{H}_{ujn}$ specified in Assumption C2; (iv) The following identifiability condition holds: $|\mathrm{E}[\psi_{uj}(W_{uj}, \theta, \eta_{uj})]| \geqslant \frac{1}{2}|J_{uj}(\theta - \theta_{uj})| \wedge c_0$ for all $\theta \in \Theta_{uj}$, with $J_{uj} := \partial_\theta \mathrm{E}[\psi_{uj}(W_{uj}, \theta, \eta_{uj})]|_{\theta=\theta_{uj}}$ satisfies $0 < j_n < |J_{uj}| < C < \infty$ for all $u \in \mathcal{U}$ and $j \in [\tilde{p}]$; (v) The following smoothness conditions holds*

(a) $\sup_{u\in\mathcal{U}, j\in[\tilde{p}], (\theta,\bar{\theta})\in\Theta_{uj}^2, (\eta,\bar{\eta})\in\mathcal{H}_{ujn}^2} \frac{\mathrm{E}[\{\psi_{uj}(W_{uj},\theta,\eta) - \psi_{uj}(W_{uj},\bar{\theta},\bar{\eta})\}^2]}{\{|\theta-\bar{\theta}|\vee\|\eta-\bar{\eta}\|_e\}^\alpha} \leqslant C,$

(b) $\sup_{u\in\mathcal{U}, (\theta,\eta)\in\Theta_{uj}\times\mathcal{H}_{ujn}, r\in[0,1)} |\partial_r \mathrm{E}\left[\psi_{uj}(W_{uj}, \theta, \eta_{uj} + r\{\eta - \eta_{uj}\})\right]|/\|\eta - \eta_{uj}\|_e \leqslant \bar{B}_{1n},$

(c) $\sup_{u\in\mathcal{U}, j\in[\tilde{p}], (\theta,\eta)\in\Theta_{uj}\times\mathcal{H}_{ujn}, r\in[0,1)} \frac{|\partial_r^2 \mathrm{E}[\psi_{uj}(W_{uj},\theta_{uj}+r\{\theta-\theta_{uj}\},\eta_{uj}+r\{\eta-\eta_{uj}\})]|}{\{|\theta-\theta_{uj}|^2\vee\|\eta-\eta_{uj}\|_e^2\}} \leqslant \bar{B}_{2n}.$

Next we state assumptions on the nuisance functions. In what follows, let $\Delta_n \searrow 0$, $\delta_n \searrow 0$, and $\tau_n \searrow 0$ be sequences of constants approaching zero from above at a speed at most polynomial in $n$ (for example, $\delta_n \geqslant 1/n^c$ for some $c > 0$).

**Assumption C2** (Estimation of Nuisance Functions). *The following conditions hold for each $n \geqslant n_0$ and all $P \in \mathcal{P}_n$. The estimated functions $\widehat{\eta}_{uj} \in \mathcal{H}_{ujn}$ with probability at least $1 - \Delta_n$, $\mathcal{H}_{ujn}$ is the set of measurable maps $\tilde{\eta}_{uj}$ such that*

$$\sup_{u\in\mathcal{U}} \max_{j\in[\tilde{p}]} \|\tilde{\eta}_{uj} - \eta_{uj}\|_e \leqslant \tau_n,$$

*here the e-norm is the same as in Assumption C1, and whose complexity does not grow too quickly in the sense that $\mathcal{F}_1 = \{\psi_{uj}(W_{uj}, \theta, \eta) : u \in \mathcal{U}, j \in [\tilde{p}], \theta \in \Theta_{uj}, \eta \in \mathcal{H}_{ujn} \cup \{\eta_{uj}\}\}$ is suitably measurable and its uniform covering entropy obeys:*

$$\sup_Q \log N(\epsilon\|F_1\|_{Q,2}, \mathcal{F}_1, \|\cdot\|_{Q,2}) \leqslant s_{n(\mathcal{U},\tilde{p})}(\log(a_n/\epsilon)) \vee 0,$$

*where $F_1(W)$ is an envelope for $\mathcal{F}_1$ which is measurable with respect to $W$ and satisfies $F_1(W) \geqslant \sup_{u\in\mathcal{U}, j\in[\tilde{p}], \theta\in\Theta_{uj}, \eta\in\mathcal{H}_{ujn}} |\psi_{uj}(W_{uj}, \theta, \eta)|$ and $\|F_1\|_{P,q} \leqslant K_n$ for $q \geqslant 2$. The complexity characteristics*

*$a_n \geqslant \max(n, K_n, \mathrm{e})$ and $s_{n(\mathcal{U},\tilde{p})} \geqslant 1$ obey the growth conditions:*

$$n^{-1/2}\sqrt{s_{n(\mathcal{U},\tilde{p})}\log(a_n)} + n^{-1}s_{n(\mathcal{U},\tilde{p})}n^{\frac{1}{q}}K_n\log(a_n) \leqslant \tau_n$$
$$\{(1 \vee \bar{B}_{1n})(\tau_n/j_n)\}^{\alpha/2}\sqrt{s_{n(\mathcal{U},\tilde{p})}\log(a_n)} + s_{n(\mathcal{U},\tilde{p})}n^{\frac{1}{q}-\frac{1}{2}}K_n\log(a_n)\log n \leqslant \delta_n,$$
$$\text{and} \quad \sqrt{n}\bar{B}_{2n}(1 \vee \bar{B}_{1n})(\tau_n/j_n)^2 \leqslant \delta_n$$

*here $\bar{B}_{1n}$, $\bar{B}_{2n}$, $j_n$, $q$ and $\alpha$ are defined in Assumption C1.*

**Theorem 5** (Uniform Bahadur representation for a Continuum of Target Parameters)**.** *Under Assumptions C1 and C2, for an estimator $(\check{\theta}_{uj})_{u\in\mathcal{U},j\in[\tilde{p}]}$ that obeys equation (H.78),*

$$\sqrt{n}\sigma_{uj}^{-1}(\check{\theta}_{uj} - \theta_{uj}) = \mathbb{G}_n\bar{\psi}_{uj} + O_P(\delta_n) \text{ in } \ell^\infty(\mathcal{U}\times[\tilde{p}]), \text{ uniformly in } P \in \mathcal{P}_n,$$

*here $\bar{\psi}_{uj}(W) := -\sigma_{uj}^{-1}J_{uj}^{-1}\psi_{uj}(W_{uj}, \theta_{uj}, \eta_{uj})$ and $\sigma_{uj}^2 = \mathrm{E}[J_{uj}^{-2}\psi_{uj}^2(W_{uj}, \theta_{uj}, \eta_{uj})]$.*

The uniform Bahadur representation derived in Theorem 5 is useful in the construction of simultaneous confidence bands for $(\theta_{uj})_{u\in\mathcal{U},j\in[\tilde{p}]}$. This is achieved by new high-dimensional central limit theorems that have recently been developed in [24] and [25]. We will make use of the following regularity condition. In what follows $\bar{\delta}_n$ and $\Delta_n$ are fixed sequences going to zero, and we denote $\widehat{\psi}_{uj}(W) := -\widehat{\sigma}_{uj}^{-1}\widehat{J}_{uj}^{-1}\psi_{uj}(W_{uj}, \check{\theta}_{uj}, \widehat{\eta}_{uj})$ be the estimators of $\bar{\psi}_{uj}(W)$, with $\widehat{J}_{uj}$ and $\widehat{\sigma}_{uj}$ being suitable estimators of $J_{uj}$ and $\sigma_{uj}$. In what follows, $\|\cdot\|_{\mathbb{P}_n,2}$ denotes the empirical $L_2(\mathbb{P}_n)$-norm with $\mathbb{P}_n$ as the empirical measure of the data.

**Assumption C3** (Score Regularity)**.** *The following conditions hold for each $n \geqslant n_0$ and all $P \in \mathcal{P}_n$. (i) The class of function induced by the score $\mathcal{F}_0 = \{\bar{\psi}_{uj}(W) : u \in \mathcal{U}, j \in [\tilde{p}]\}$ is suitably measurable and its uniform covering entropy obeys:*

$$\sup_Q \log N(\epsilon\|F_0\|_{Q,2}, \mathcal{F}_0, \|\cdot\|_{Q,2}) \leqslant \varrho_n(\log(A_n/\epsilon)) \vee 0,$$

*here $F_0(W)$ is an envelope for $\mathcal{F}_0$ which is measurable with respect to $W$ and satisfies $F_0(W) \geqslant \sup_{u\in\mathcal{U},j\in[\tilde{p}]}|\bar{\psi}_{uj}(W)|$ and $\|F_0\|_{P,q} \leqslant L_n$ for $q \geqslant 4$. Furthermore, $c \leqslant \sup_{u\in\mathcal{U},j\in[\tilde{p}]}\mathrm{E}[|\bar{\psi}_{uj}(W)|^k] \leqslant CL_n^{k-2}$ for $k = 2, 3, 4$. (ii) The set $\widehat{\mathcal{F}}_0 = \{\bar{\psi}_{uj}(W) - \widehat{\psi}_{uj}(W) : u \in \mathcal{U}, j \in [\tilde{p}]\}$ satisfies the conditions $\log N(\epsilon, \widehat{\mathcal{F}}_0, \|\cdot\|_{\mathbb{P}_n,2}) \leqslant \bar{\varrho}_n(\log(\bar{A}_n/\epsilon)) \vee 0$, and $\sup_{u\in\mathcal{U},j\in[\tilde{p}]}\mathbb{E}_n[\{\bar{\psi}_{uj}(W) - \widehat{\psi}_{uj}(W)\}^2] \leqslant \bar{\delta}_n\{\rho_n\bar{\rho}_n\log(A_n \vee n)\log(\bar{A}_n \vee n)\}^{-1}$ with probability $1 - \Delta_n$.*

Assumption C3 imposes condition on the class of functions induced by $\bar{\psi}_{uj}$ and on its estimators $\widehat{\psi}_{uj}$. Typically the bound $L_n$ on the moment of the envelope is smaller than $K_n$, and in many settings $\bar{\rho}_n = \rho_n \lesssim d_{\mathcal{U}}$ the dimension of $\mathcal{U}$.

Next let $\mathcal{N}$ denote a mean zero Gaussian process indexed by $\mathcal{U}\times[\tilde{p}]$ with covariance operator given by $\mathrm{E}[\bar{\psi}_{uj}(W)\bar{\psi}_{u'j'}(W)]$ for $j, j' \in [\tilde{p}]$ and $u, u' \in \mathcal{U}$. Because of the high-dimensionality, indeed $\tilde{p}$ can be larger than the sample size $n$, the central limit theorem will be uniformly valid over "rectangles". This class of sets are rich enough to construct many confidence regions of interest in applications accounting for multiple testing. Let $\mathcal{R}$ denote the set of rectangles $R = \{z \in \mathbb{R}^{\tilde{p}} : \max_{j\in A} z_j \leqslant t, \max_{j\in B}(-z_j) \leqslant t\}$ for all $A, B \subset [\tilde{p}]$ and $t \in \mathbb{R}$. The following result is a consequence of Theorem 5 above and Corollary 2.2 of [26].

**Corollary 3.** *Under Assumptions C1, C2 and Assumption C3(i), with $\delta_n = o(\{\rho_n\log(A_n \vee n)\}^{-1/2})$, and $\rho_n\log(A_n \vee n) = o(\{(n/L_n^2)^{1/7} \wedge (n^{1-2/q}/L_n^2)^{1/3}\})$, we have that*

$$\sup_{P\in\mathcal{P}_n}\sup_{R\in\mathcal{R}}\left|\mathrm{P}_P\left(\{\sup_{u\in\mathcal{U}} n^{1/2}\sigma_{uj}^{-1}(\check{\theta}_{uj} - \theta_{uj})\}_{j=1}^{\tilde{p}} \in R\right) - \mathrm{P}_P(\mathcal{N} \in R)\right| = o(1).$$

In order to derive a method to build confidence regions we approximate the process $\mathcal{N}$ by the Gaussian multiplier bootstrap based on estimates $\widehat{\psi}_{uj}$ of $\bar{\psi}_{uj}$, namely

$$\widehat{\mathcal{G}} = (\widehat{\mathcal{G}}_{uj})_{u\in\mathcal{U}, j\in[\tilde{p}]} = \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g_i \widehat{\psi}_{uj}(W_i) \right\}_{u\in\mathcal{U}, j\in[\tilde{p}]}$$

here $(g_i)_{i=1}^n$ are independent standard normal random variables which are independent from the data $(W_i)_{i=1}^n$. Based on Theorem 5.2 of [24], the following result shows that the multiplier bootstrap provides a valid approximation to the large sample probability law of $\sqrt{n}(\check{\theta}_{uj} - \theta_{uj})_{u\in\mathcal{U}, j\in[\tilde{p}]}$ over rectangles.

**Corollary 4** (**Uniform Validity of Gaussian Multiplier Bootstrap**)**.** *Under Assumptions C1, C2 and Assumption C3, with* $\delta_n = o(\{(1 + d_W)\rho_n \log(A_n \vee n)\}^{-1/2})$ *and* $\rho_n \log(A_n \vee n) = o(\{(n/L_n^2)^{1/7} \wedge (n^{1-2/q}/L_n^2)^{1/3}\})$*, we have that*

$$\sup_{P\in\mathcal{P}_n} \sup_{R\in\mathcal{R}} \left| \mathrm{P}_P \left( \{\sup_{u\in\mathcal{U}} n^{1/2} \sigma_{uj}^{-1} (\check{\theta}_{uj} - \theta_{uj})\}_{j=1}^{\tilde{p}} \in R \right) - \mathrm{P}_P(\widehat{\mathcal{G}} \in R \mid (W_i)_{i=1}^n) \right| = o(1)$$

## Appendix I. Continuum of $\ell_1$-Penalized M-Estimators

For the reader's convenience, this section collects results on the estimation of a continuum of estimation of high-dimensional models via $\ell_1$-penalized estimators. We refer to [14] for the proofs.

Consider a data generating process with a response variable $(Y_u)_{u\in\mathcal{U}}$ and observable covariates $(X_u)_{u\in\mathcal{U}}$ satisfies for each $u \in \mathcal{U}$,

$$\theta_u \in \arg\min_{\theta\in\mathbb{R}^p} \mathrm{E}[M_u(Y_u, X_u, \theta, a_u)], \tag{I.80}$$

here $\theta_u$ is a $p$-dimensional vector, $a_u$ is a nuisance function that capture the misspecification of the model, $M_u$ is a pre-specified function, and the $p_u$-dimensional ($p_u \leqslant p$) covariate $X_u$ could have been constructed based on transformations of other variables. This implies that

$$\partial_\theta \mathrm{E}[M_u(Y_u, X_u, \theta_u, a_u)] = 0 \quad \text{for all } u \in \mathcal{U}.$$

The solution $\theta_u$ is assumed to be sparse in the sense that for some process $(\theta_u)_{u\in\mathcal{U}}$ satisfies

$$\|\theta_u\|_0 \leqslant s \text{ for all } u \in \mathcal{U}.$$

Because of the nuisance function, such sparsity assumption is very mild and formulation (I.80) encompasses several cases of interest including approximate sparse models. We focus on the estimation of $(\theta_u)_{u\in\mathcal{U}}$ and we assume that an estimate $\widehat{a}_u$ of the nuisance function $a_u$ is available and the criterion $M_u(Y_u, X_u, \theta_u) := M_u(Y_u, X_u, \theta_u, \widehat{a}_u)$ is used as a proxy for $M_u(Y_u, X, \theta_u, a_u)$.

In the case of linear regression we have $M_u(y, x, \theta) = \frac{1}{2}(y - x'\theta)^2$. In the logistic regression case, we have $M_u(y, x, \theta) = -\{1(y = 1) \log G(x'\theta) + 1(y = 0) \log(1 - G(x'\theta))\}$ with $G$ is the logistic link function $G(t) = \exp(t)/\{1 + \exp(t)\}$. Additional examples include quantile regression models for $u \in (0, 1)$.

**Example 8** (Quantile Regression Model)**.** Consider a data generating process $Y = F_{Y|X}^{-1}(U) = X'\theta_U + r_U(X)$, with $U \sim \mathrm{Unif}(0, 1)$, and $X$ is a $p$-dimensional vector of covariates. The criterion $M_u(y, x, \theta) = (u - 1\{y \leqslant x'\theta\})(y - x'\theta)$ with the (trivial) estimate $\widehat{a}_u = 0$ for the nuisance parameter $a_u = r_u$.

**Example 9** (Lasso with Estimated Weights)**.** We consider a linear model defined as $f_u Y = f_u X'\theta_u + \bar{r}_u + \zeta_u$, $\mathrm{E}[f_u X \zeta_u] = 0$, here $X$ are $\bar{p}$-dimensional covariates, $\theta_u$ is a $s$-sparse vector, and $\bar{r}_u$ is an approximation error satisfies $\sup_{u\in\mathcal{U}} \mathbb{E}_n[\bar{r}_u^2] \lesssim_P s \log \bar{p}/n$. In this setting, $(Y, X)$ are observed and only an estimator $\widehat{f}_u$ of $f_u$ is available. This corresponds to nuisance parameter $a_u = (f_u, \bar{r}_u)$ and $\widehat{a}_u = (\widehat{f}_u, 0)$ so that $\mathbb{E}_n[M_u(Y, X, \theta, a_u)] = \mathbb{E}_n[f_u^2(Y - X'\theta - \bar{r}_u)^2]$ and $\mathbb{E}_n[M_u(Y, X, \theta)] = \mathbb{E}_n[\widehat{f}_u^2(Y - X'\theta)^2]$.

We assume that $n$ i.i.d. observations from dgps with (I.80) holds, $\{(Y_{iu}, X_{iu})_{u\in\mathcal{U}}\}_{i=1}^n$, are observed to estimate $(\theta_u)_{u\in\mathcal{U}}$. For each $u \in \mathcal{U}$, a penalty level $\lambda$, and a diagonal matrix of penalty loadings $\widehat{\Psi}_u$, we define the $\ell_1$-penalized $M_u$-estimator (Weighed-Lasso) as

$$\widehat{\theta}_u \in \arg\min_\theta \mathbb{E}_n[M_u(Y_u, X_u, \theta)] + \frac{\lambda}{n}\|\widehat{\Psi}_u\theta\|_1. \tag{I.81}$$

Furthermore, for each $u \in \mathcal{U}$, the post-penalized estimator (Post-Lasso) based on a set of covariates $\widetilde{T}_u$ is then defined as

$$\widetilde{\theta}_u \in \arg\min_\theta \mathbb{E}_n[M_u(Y_u, X_u, \theta)] \quad : \quad \mathrm{support}(\theta) \subseteq \widetilde{T}_u. \tag{I.82}$$

Potentially, the set $\widetilde{T}_u$ contains $\mathrm{support}(\widehat{\theta}_u)$ and possibly additional variables deemed as important (although in that case the total number of additional variables should also obey the same growth conditions that $s$ obeys). We will set $\widetilde{T}_u = \mathrm{support}(\widehat{\theta}_u)$ unless otherwise noted.

In order to handle the functional response data, the penalty level $\lambda$ and penalty loading $\widehat{\Psi}_u = \mathrm{diag}(\{\widehat{l}_{uk}, k = 1, \ldots, p\})$ need to be set to control selection errors uniformly over $u \in \mathcal{U}$. The choice of loading matrix is problem specific and we suggest to mimic the following "ideal" choice $\widehat{\Psi}_{u0} = \mathrm{diag}(\{l_{uk}, k = 1, \ldots, p\})$ with

$$l_{uk} = \{\mathbb{E}_n\left[\{\partial_{\theta_k} M_u(Y_u, X_u, \theta_u, a_u)\}^2\right]\}^{1/2} \tag{I.83}$$

which is motivated by the use of self-normalized moderate deviation theory. In that case, it is suitable to set $\lambda$ so that with high probability

$$\frac{\lambda}{n} \geqslant c \sup_{u\in\mathcal{U}} \left\|\widehat{\Psi}_{u0}^{-1}\mathbb{E}_n\left[\partial_\theta M_u(Y_u, X_u, \theta_u, a_u)\right]\right\|_\infty, \tag{I.84}$$

here $c > 1$ is a fixed constant. Indeed, in the case that $\mathcal{U}$ is a singleton the choice above is similar to [21], [13], and [20]. This approach was first employed for a continuum of indices $\mathcal{U}$ in the context of $\ell_1$-penalized quantile regression processes by [11].

To implement (I.84), we propose setting the penalty level as

$$\lambda = c\sqrt{n}\Phi^{-1}(1 - \xi/\{2pN_n\}), \tag{I.85}$$

here $N_n$ is a measure of the class of functions indexed by $\mathcal{U}$, $1 - \xi$ (with $\xi = o(1)$) is a confidence level associated with the probability of event (I.84), and $c > 1$ is a slack constant. In many settings we can take $N_n = n^{d_\mathcal{U}}$. If the set $\mathcal{U}$ is a singleton, $N_n = 1$ suffices which corresponds to what is used in [17].

I.1. **Generic Finite Sample Bounds.** In this subsection we derive finite sample bounds based on Assumption C4 below. This assumption provides sufficient conditions that are implied by a variety of settings including generalized linear models.

**Assumption C4** (M-Estimation Conditions)**.** *Let $\{(Y_{iu}, X_{iu}, u \in \mathcal{U}), i = 1, \ldots, n\}$ be $n$ i.i.d. observations of the model (I.80) and let $T_u = \text{support}(\theta_u)$, here $\|T_u\|_0 \leqslant s$, $u \in \mathcal{U}$. With probability $1 - \Delta_n$ we have that for all $u \in \mathcal{U}$ there are weights $w_u = w_u(Y_u, X_u)$ and $C_{un}$ such that:*

(a) $|\mathbb{E}_n[\partial_\theta M_u(Y_u, X_u, \theta_u) - \partial_\theta M_u(Y_u, X_u, \theta_u, a_u)]'\delta| \leqslant C_{un}\|\sqrt{w_u}X_u'\delta\|_{\mathbb{P}_n,2}$;
(b) $\ell\widehat{\Psi}_{u0} \leqslant \widehat{\Psi}_u \leqslant L\widehat{\Psi}_{u0}$ *for* $\ell > 1/c$, *and let* $\tilde{c} = \frac{Lc+1}{\ell c-1} \sup_{u\in\mathcal{U}} \|\widehat{\Psi}_{u0}\|_\infty \|\widehat{\Psi}_{u0}^{-1}\|_\infty$;
(c) *for all* $\delta \in A_u$ *there is* $\bar{q}_{A_u} > 0$ *such that*

$$\mathbb{E}_n[M_u(Y_u, X_u, \theta_u + \delta)] - \mathbb{E}_n[M_u(Y_u, X_u, \theta_u)] - \mathbb{E}_n[\partial_\theta M_u(Y_u, X_u, \theta_u)]'\delta + 2C_{un}\|\sqrt{w_u}X_u'\delta\|_{\mathbb{P}_n,2} \geqslant \left\{\|\sqrt{w_u}X_u'\delta\|^2_{\mathbb{P}_n,2}\right\} \wedge \left\{\bar{q}_{A_u}\|\sqrt{w_u}X_u'\delta\|_{\mathbb{P}_n,2}\right\}.$$

In many applications we take the weights to be $w_u = w_u(X_u) = 1$ but we allow for more general weights. Assumption C4(a) bounds the impact of estimating the nuisance functions uniformly over $u \in \mathcal{U}$. In the setting with $s$-sparse estimands, we typically have $C_{un} \lesssim \{n^{-1}s\log(pn)\}^{1/2}$. The loadings $\widehat{\Psi}_u$ are assumed larger (but not too much larger) than the ideal choice $\widehat{\Psi}_{u0}$ defined in (I.83). This is formalized in Assumption C4(b). Assumption C4(c) is an identification condition that will be imposed for specific choices of $A_u$ and $q_{A_u}$. It relates to conditions in the literature derived for the case of a singleton $\mathcal{U}$ and no nuisance functions, see the restricted strong convexity[8] used in [63] and the non-linear impact coefficients used in [11] and [18].

The following results establish rates of convergence for the $\ell_1$-penalized solution with estimated nuisance functions (I.81), sparsity bounds and rates of convergence for the post-selection refitted estimator (I.82). They are based on restricted eigenvalue type conditions and sparse eigenvalue conditions. With the restricted eigenvalue is defined as $\bar{\kappa}_{u,2\tilde{\mathbf{c}}} = \inf_{\delta\in\Delta_{u,2\tilde{\mathbf{c}}}} \|\sqrt{w_u}X_u'\delta\|_{\mathbb{P}_n,2}/\|\delta_{T_u}\|$ In the results for sparsity and post-selection refitted models, the minimum and maximum sparse eigenvalues,

$$\phi_{\min}(m, u) = \min_{1\leqslant\|\delta\|_0\leqslant m} \frac{\|\sqrt{w_u}X_u'\delta\|^2_{\mathbb{P}_n,2}}{\|\delta\|^2} \quad \text{and} \quad \phi_{\max}(m, u) = \max_{1\leqslant\|\delta\|_0\leqslant m} \frac{\|\sqrt{w_u}X_u'\delta\|^2_{\mathbb{P}_n,2}}{\|\delta\|^2},$$

are also relevant quantities to characterize the behavior of the estimators.

**Lemma 20.** *Suppose that Assumption C4 holds with $\delta \in A_u = \{\delta : \|\delta_{T_u^c}\|_1 \leqslant 2\tilde{\mathbf{c}}\|\delta_{T_u}\|_1\} \cup \{\delta : \|\delta\|_1 \leqslant \frac{6c\|\widehat{\Psi}_{u0}^{-1}\|_\infty}{\ell c-1}\frac{n}{\lambda}C_{un}\|\sqrt{w_u}X_u'\delta\|_{\mathbb{P}_n,2}\}$ and $\bar{q}_{A_u} > 3\left\{(L + \frac{1}{c})\|\widehat{\Psi}_{u0}\|_\infty\frac{\lambda\sqrt{s}}{n\bar{\kappa}_{u,2\tilde{\mathbf{c}}}} + 9\tilde{\mathbf{c}}C_{un}\right\}$. Suppose that $\lambda$ satisfies condition (I.84) with probability $1 - \Delta_n$. Then, with probability $1 - 2\Delta_n$ we have uniformly over $u \in \mathcal{U}$*

$$\begin{aligned}\|\sqrt{w_u}X_u'(\widehat{\theta}_u - \theta_u)\|_{\mathbb{P}_n,2} &\leqslant 3\left\{(L + \tfrac{1}{c})\|\widehat{\Psi}_{u0}\|_\infty\tfrac{\lambda\sqrt{s}}{n\bar{\kappa}_{u,2\tilde{\mathbf{c}}}} + 9\tilde{\mathbf{c}}C_{un}\right\}\\ \|\widehat{\theta}_u - \theta_u\|_1 &\leqslant 3\left\{\tfrac{(1+2\tilde{\mathbf{c}})\sqrt{s}}{\bar{\kappa}_{u,2\tilde{\mathbf{c}}}} + \tfrac{6c\|\widehat{\Psi}_{u0}^{-1}\|_\infty}{\ell c-1}\tfrac{n}{\lambda}C_{un}\right\}\left\{(L + \tfrac{1}{c})\|\widehat{\Psi}_{u0}\|_\infty\tfrac{\lambda\sqrt{s}}{n\bar{\kappa}_{u,2\tilde{\mathbf{c}}}} + 9\tilde{\mathbf{c}}C_{un}\right\}\end{aligned}$$

**Lemma 21** (M-Estimation Sparsity)**.** *In addition to conditions of Lemma 20, assume that with probability $1 - \Delta_n$ for all $u \in \mathcal{U}$ and $\delta \in \mathbb{R}^p$ we have*

$$|\{\mathbb{E}_n[\partial_\theta M_u(Y_u, X_u, \widehat{\theta}_u) - \partial_\theta M_u(Y_u, X_u, \theta_u)]\}'\delta| \leqslant L_{un}\|\sqrt{w_u}X_u'\delta\|_{\mathbb{P}_n,2}.$$

*Let $\mathcal{M}_u = \{m \in \mathbb{N} : m \geqslant 2\phi_{\max}(m, u)L_u^2\}$ with $L_u = \frac{c\|\widehat{\Psi}_{u0}^{-1}\|_\infty}{c\ell-1}\frac{n}{\lambda}\{C_{un} + L_{un}\}$, then with probability $1 - 3\Delta_n$ we have that*

$$\widehat{s}_u \leqslant \min_{m\in\mathcal{M}_u} \phi_{\max}(m, u)L_u^2 \quad \text{for all} \quad u \in \mathcal{U}.$$

[8]Assumption C4 (a) and (c) could have been stated with $\{C_{un}/\sqrt{s}\}\|\delta\|_1$ instead of $C_{un}\|\sqrt{w_u}X_u'\delta\|_{\mathbb{P}_n,2}$.

**Lemma 22.** *Let $\widetilde{T}_u, u \in \mathcal{U}$, be the support used for post penalized estimator (I.82) and $\tilde{s}_u = \|\widetilde{T}_u\|_0$ its cardinality. In addition to conditions of Lemma 20, suppose that Assumption C4(c) holds also for $A_u = \{\delta : \|\delta\|_0 \leqslant \tilde{s}_u + s\}$ with probability $1 - \Delta_n$, $\bar{q}_{A_u} > 2\left\{\frac{\sqrt{\tilde{s}_u + s_u}\|\mathbb{E}_n[S_u]\|_\infty}{\sqrt{\phi_{\min}(\tilde{s}_u + s_u, u)}} + 3C_{un}\right\}$ and $\bar{q}_{A_u} > 2\{\mathbb{E}_n[M_u(Y_u, X_u, \tilde{\theta}_u)] - \mathbb{E}_n[M_u(Y_u, X_u, \theta_u)]\}_+^{1/2}$. Then, we have uniformly over $u \in \mathcal{U}$*

$$\|\sqrt{w_u}X_u'(\tilde{\theta}_u - \theta_u)\|_{\mathbb{P}_n,2} \leqslant \{\mathbb{E}_n[M_u(Y_u, X_u, \tilde{\theta}_u)] - \mathbb{E}_n[M_u(Y_u, X_u, \theta_u)]\}_+^{1/2} + \frac{\sqrt{\tilde{s}_u + s_u}\|\mathbb{E}_n[S_u]\|_\infty}{\sqrt{\phi_{\min}(\tilde{s}_u + s_u, u)}} + 3C_{un}.$$

In Lemma 22, if $\widetilde{T}_u = \text{support}(\widehat{\theta}_u)$, we have that

$$\mathbb{E}_n[M_u(Y_u, X_u, \tilde{\theta}_u)] - \mathbb{E}_n[M_u(Y_u, X_u, \theta_u)] \leqslant \mathbb{E}_n[M_u(Y_u, X_u, \widehat{\theta}_u)] - \mathbb{E}_n[M_u(Y_u, X_u, \theta_u)] \leqslant \lambda C'\|\widehat{\theta}_u - \theta_u\|_1$$

and $\sup_{u\in\mathcal{U}} \|\mathbb{E}_n[S_u]\|_\infty \leqslant C'\lambda$ with high probability, $C' \leqslant L \sup_{u\in\mathcal{U}} \|\widehat{\Psi}_{u0}\|_\infty$.

These results generalize important results of the $\ell_1$-penalized estimators to the case of functional response data and estimated of nuisance functions. A key assumption in Lemmas 20-22 is that the choice of $\lambda$ satisfies (I.84). We next provide a set of simple generic conditions that will imply the validity of the proposed choice. These generic conditions can be verified in many applications of interest.

**Condition WL.** *For each $u \in \mathcal{U}$, let $S_u = \partial_\theta M_u(Y_u, X_u, \theta_u, a_u)$, suppose that:*
*(i)* $\sup_{u\in\mathcal{U}} \max_{k\leqslant p} \{\mathrm{E}[|S_{uk}|^3]\}^{1/3}/\{\mathrm{E}[|S_{uk}|^2]\}^{1/2}\Phi^{-1}(1 - \xi/\{2pN_n\}) \leqslant \delta_n n^{1/6}$, *for all $u \in \mathcal{U}$, $k \in [p]$;*
*(ii) $N_n \geqslant N(\epsilon, \mathcal{U}, d_\mathcal{U})$, here $\epsilon$ is such that with probability $1 - \Delta_n$:*
$\sup_{d_\mathcal{U}(u,u')\leqslant\epsilon} \max_{k\leqslant p} \frac{\|\mathbb{E}_n[S_u - S_{u'}]\|_\infty}{\mathrm{E}[|S_{uk}|^2]^{1/2}} \leqslant \delta_n n^{-\frac{1}{2}}$, *and* $\sup_{d_\mathcal{U}(u,u')\leqslant\epsilon} \max_{k\leqslant p} \frac{|\mathrm{E}[S_{uk}^2 - S_{u'k}^2]| + |(\mathbb{E}_n - \mathrm{E})[S_{uk}^2]|}{\mathrm{E}[|S_{uk}|^2]} \leqslant \delta_n$.

The following technical lemma justifies the choice of penalty level $\lambda$. It is based on self-normalized moderate deviation theory.

**Lemma 23** (Choice of $\lambda$)**.** *Suppose Condition WL holds, let $c' > c > 1$ be constants, $\xi \in [1/n, 1/\log n]$, and $\lambda = c'\sqrt{n}\Phi^{-1}(1 - \xi/\{2pN_n\})$. Then for $n \geqslant n_0$ large enough depends only on Condition WL,*

$$\mathrm{P}\left(\lambda/n \geqslant c \sup_{u\in\mathcal{U}} \|\widehat{\Psi}_{u0}^{-1}\mathbb{E}_n[\partial_\theta M_u(Y_u, X_u, \theta_u, a_u)]\|_\infty\right) \geqslant 1 - \xi - o(\xi) - \Delta_n.$$

We note that Condition WL(ii) contains high level conditions. See [16] for examples that satisfy these conditions. The following corollary summarizes these results for many applications of interest in well behaved designs.

**Corollary 5** (Rates under Simple Conditions)**.** *Suppose that with probability $1 - o(1)$ we have that $C_{un} \vee L_{un} \leqslant C\{n^{-1}s\log(pn)\}^{1/2}$, $(Lc+1)/(\ell c-1) \leqslant C$, $w_u = 1$, and Condition WL holds with $\log N_n \leqslant C\log(pn)$. Further suppose that with probability $1 - o(1)$ the sparse minimal and maximal eigenvalues are well behaved, $c \leqslant \phi_{\min}(s\ell_n, u) \leqslant \phi_{\max}(s\ell_n, u) \leqslant C$ for some $\ell_n \to \infty$ uniformly over $u \in \mathcal{U}$. Then with probability $1 - o(1)$ we have*

$$\sup_{u\in\mathcal{U}} \|X_u'(\widehat{\theta}_u - \theta_u)\|_{\mathbb{P}_n,2} \lesssim \sqrt{\frac{s\log(pn)}{n}}, \quad \sup_{u\in\mathcal{U}} \|\widehat{\theta}_u - \theta_u\|_1 \lesssim \sqrt{\frac{s^2\log(pn)}{n}}, \quad \text{and} \quad \sup_{u\in\mathcal{U}} \|\widehat{\theta}_u\|_0 \lesssim s.$$

*Moreover, if $\widetilde{T}_u = \text{support}(\widehat{\theta}_u)$, we have that*

$$\sup_{u\in\mathcal{U}} \|X_u'(\tilde{\theta}_u - \theta_u)\|_{\mathbb{P}_n,2} \lesssim \sqrt{\frac{s\log(pn)}{n}}$$

## Appendix J. Bounds on Covering entropy

Let $(W_i)_{i=1}^n$ be a sequence of independent copies of a random element $W$ taking values in a measurable space $(\mathcal{W}, \mathcal{A}_{\mathcal{W}})$ according to a probability law $P$. Let $\mathcal{F}$ be a set of suitably measurable functions $f\colon \mathcal{W} \to \mathbb{R}$, equipped with a measurable envelope $F\colon \mathcal{W} \to \mathbb{R}$. The proofs for the following lemmas can be found in [16].

**Lemma 24** (Algebra for Covering Entropies)**.** *Work with the setup above.*
*(1) Let $\mathcal{F}$ be a VC subgraph class with a finite VC index $k$ or any other class whose entropy is bounded above by that of such a VC subgraph class, then the uniform entropy numbers of $\mathcal{F}$ obey*

$$\sup_Q \log N(\epsilon \|F\|_{Q,2}, \mathcal{F}, \|\cdot\|_{Q,2}) \lesssim \{1 + k \log(1/\epsilon)\} \vee 0$$

*(2) For any measurable classes of functions $\mathcal{F}$ and $\mathcal{F}'$ mapping $\mathcal{W}$ to $\mathbb{R}$,*

$$\log N(\epsilon \|F + F'\|_{Q,2}, \mathcal{F} + \mathcal{F}', \|\cdot\|_{Q,2}) \leqslant \log N\left(\tfrac{\epsilon}{2}\|F\|_{Q,2}, \mathcal{F}, \|\cdot\|_{Q,2}\right) + \log N\left(\tfrac{\epsilon}{2}\|F'\|_{Q,2}, \mathcal{F}', \|\cdot\|_{Q,2}\right),$$
$$\log N(\epsilon \|F \cdot F'\|_{Q,2}, \mathcal{F} \cdot \mathcal{F}', \|\cdot\|_{Q,2}) \leqslant \log N\left(\tfrac{\epsilon}{2}\|F\|_{Q,2}, \mathcal{F}, \|\cdot\|_{Q,2}\right) + \log N\left(\tfrac{\epsilon}{2}\|F'\|_{Q,2}, \mathcal{F}', \|\cdot\|_{Q,2}\right),$$
$$N(\epsilon \|F \vee F'\|_{Q,2}, \mathcal{F} \cup \mathcal{F}', \|\cdot\|_{Q,2}) \leqslant N\left(\epsilon\|F\|_{Q,2}, \mathcal{F}, \|\cdot\|_{Q,2}\right) + N\left(\epsilon\|F'\|_{Q,2}, \mathcal{F}', \|\cdot\|_{Q,2}\right).$$

*(3) For any measurable class of functions $\mathcal{F}$ and a fixed function $f$ mapping $\mathcal{W}$ to $\mathbb{R}$,*

$$\log \sup_Q N(\epsilon \||f| \cdot F\|_{Q,2}, f \cdot \mathcal{F}, \|\cdot\|_{Q,2}) \leqslant \log \sup_Q N\left(\epsilon/2 \|F\|_{Q,2}, \mathcal{F}, \|\cdot\|_{Q,2}\right)$$

*(4) Given measurable classes $\mathcal{F}_j$ and envelopes $F_j$, $j = 1, \ldots, k$, mapping $\mathcal{W}$ to $\mathbb{R}$, a function $\phi\colon \mathbb{R}^k \to \mathbb{R}$ such that for $f_j, g_j \in \mathcal{F}_j$, $|\phi(f_1, \ldots, f_k) - \phi(g_1, \ldots, g_k)| \leqslant \sum_{j=1}^k L_j(x)|f_j(x) - g_j(x)|$, $L_j(x) \geqslant 0$, and fixed functions $\bar{f}_j \in \mathcal{F}_j$, the class of functions $\mathcal{L} = \{\phi(f_1, \ldots, f_k) - \phi(\bar{f}_1, \ldots, \bar{f}_k)\colon f_j \in \mathcal{F}_j, j = 1, \ldots, k\}$ satisfies*

$$\log \sup_Q N\left(\epsilon \Big\| \sum_{j=1}^k L_j F_j \Big\|_{Q,2}, \mathcal{L}, \|\cdot\|_{Q,2}\right) \leqslant \sum_{j=1}^k \log \sup_Q N\left(\tfrac{\epsilon}{k}\|F_j\|_{Q,2}, \mathcal{F}_j, \|\cdot\|_{Q,2}\right).$$

*Proof.* See Lemma L.1 in [16]. ∎

**Lemma 25** (Covering Entropy for Classes obtained as Conditional Expectations)**.** *Let $\mathcal{F}$ denote a class of measurable functions $f\colon \mathcal{W} \times \mathcal{Y} \to \mathbb{R}$ with a measurable envelope $F$. For a given $f \in \mathcal{F}$, let $\bar{f}\colon \mathcal{W} \to \mathbb{R}$ be the function $\bar{f}(w) := \int f(w, y) d\mu_w(y)$ here $\mu_w$ is a regular conditional probability distribution over $y \in \mathcal{Y}$ conditional on $w \in \mathcal{W}$. Set $\bar{\mathcal{F}} = \{\bar{f}\colon f \in \mathcal{F}\}$ and let $\bar{F}(w) := \int F(w, y) d\mu_w(y)$ be an envelope for $\bar{\mathcal{F}}$. Then, for $r, s \geqslant 1$,*

$$\log \sup_Q N(\epsilon \|\bar{F}\|_{Q,r}, \bar{\mathcal{F}}, \|\cdot\|_{Q,r}) \leqslant \log \sup_{\widetilde{Q}} N((\epsilon/4)^r \|F\|_{\widetilde{Q},s}, \mathcal{F}, \|\cdot\|_{\widetilde{Q},s}),$$

*here $Q$ belongs to the set of finitely-discrete probability measures over $\mathcal{W}$ such that $0 < \|\bar{F}\|_{Q,r} < \infty$, and $\widetilde{Q}$ belongs to the set of finitely-discrete probability measures over $\mathcal{W} \times \mathcal{Y}$ such that $0 < \|F\|_{\widetilde{Q},s} < \infty$. In particular, for every $\epsilon > 0$ and any $k \geqslant 1$,*

$$\log \sup_Q N(\epsilon, \bar{\mathcal{F}}, \|\cdot\|_{Q,k}) \leqslant \log \sup_{\widetilde{Q}} N(\epsilon/2, \mathcal{F}, \|\cdot\|_{\widetilde{Q},k}).$$

*Proof.* See Lemma L.2 in [16]. ∎