LASSO-Driven Inference in Time and Space

Victor Chernozhukov
Wolfgang K. Härdle
Chen Huang
Weining Wang

The Institute for Fiscal Studies
Department of Economics,
UCL

cemmap working paper CWP20/19
Abstract

We consider the estimation and inference in a system of high-dimensional regression equations allowing for temporal and cross-sectional dependency in covariates and error processes, covering rather general forms of weak dependence. A sequence of regressions with many regressors using LASSO (Least Absolute Shrinkage and Selection Operator) is applied for variable selection purpose, and an overall penalty level is carefully chosen by a block multiplier bootstrap procedure to account for multiplicity of the equations and dependencies in the data. Correspondingly, oracle properties with a jointly selected tuning parameter are derived. We further provide high-quality de-biased simultaneous inference on the many target parameters of the system. We provide bootstrap consistency results of the test procedure, which are based on a general Bahadur representation for the $Z$-estimators with dependent data. Simulations demonstrate good performance of the proposed inference procedure. Finally, we apply the method to quantify spillover effects of textual sentiment indices in a financial market and to test the connectedness among sectors.

1 Introduction

Many applications in statistics, economics, finance, biology and psychology are concerned with a system of ultra high-dimensional objects that communicate within complex dependency channels. Given a complex system involving many factors, one builds a network model by taking a large set of regressions, i.e. regressing every factor in the system on a large subset of other factors. Examples include analysis of financial systemic risk by quantile predictive graphical...
models with LASSO (Härdle et al., 2016; Belloni et al., 2016), limit order book network modeling via the penalized vector autoregressive approach (Härdle et al., 2018), analysis of psychology data with temporal and cross-sectional dependencies (Epskamp et al., 2016). Another example is quantifying the spillover effects or externalities for a social network, especially when the social interactions (or the interconnectedness) is not obvious (Manresa, 2013). Besides, there are numerous applications concerning association network analysis in other fields of applied statistics, see Chapter 7 in Kolaczyk and Csárdi (2014) and Epskamp et al. (2018). In general, a step-by-step LASSO procedure is very helpful for the correlation network formation. In pursuing a highly structural approach, one certainly favors a simple set of regressions that allows multiple insights on the statistical structure of the data. Therefore, a sequence of regressions with LASSO is a natural path to take. Especially in cases of reduced forms of simultaneous equation models and structural vector autoregressive (VAR) models, one can attain valuable pre-information on the core structure by running a set of simple regressions with LASSO shrinkage.

A first important question arising in this framework is how to decide on a unified level of penalty. In this article we advocate an approach to selecting the overall level of the tuning parameter in a system of equations after performing a set of single step regressions with shrinkage. A feasible (block) bootstrap procedure is developed and the consistency of parameter estimation is studied. In addition, we provide a uniform near-oracle bound for the joint estimators. The proposed technique is applicable to ultra-high dimensional systems of regression equations with high-dimensional regressors.

A second crucial issue is to establish simultaneous inference on parameters, which is an important question regarding network topology inference. For example, in a large-scale linear factor pricing model, it is of great interest to check the significance of the intercepts of cross sectional regressions (connected with zero pricing errors), e.g. Pesaran and Yamagata (2017). Our approach is an alternative testing solution compared to the Wald test statistics proposed therein. To achieve the goal of simultaneous inference, we develop a uniform robust post-selection or post-regularization inference procedure for time series data. This method is generated from a uniform Bahadur representation of de-biased instrumental variable estimators. In particular, we need to establish maximal inequalities for empirical processes for a general Huber’s Z-estimation. Note that the commonly used technique for independent data, such as the symmetrization technique, is not directly applicable in the dependent data case, see Chapter 11.6 of Kosorok (2008) for a related overview.

Our contribution lies in three aspects. First, we select the penalty level by controlling the aggregated errors in a system of high-dimensional sparse regressions, and we establish the bounds on the estimated coefficients. Furthermore, we show the implication of the restricted eigenvalue (RE) condition at a population level. Secondly, an easily implemented algorithm for effective estimation and inference is proposed. In fact, the offered estimation scheme allows us to make local and global inference on any set of parameters of interest. Thirdly, we run numerical experiments to illustrate good performance of our joint penalty relative to the single equation estimation, and we show the finite sample improvement of our multiplier block bootstrap procedure on the parameter inference. Finally, an application of textual sentiment
spillover effects on the stock returns in a financial market is presented.

In the literature, the fundamental results on achieving near oracle rate for penalized $\ell_1$-norm estimators are developed by Bickel et al. (2009). There are many related articles on deriving near-oracle bounds using the $\ell_1$-norm penalization function for the i.i.d. case, such as Belloni et al. (2011), Belloni and Chernozhukov (2013). There are also many extensions to the LASSO estimation with dependent data. For example, Kock and Callot (2015) consider the high-dimensional near-oracle inequalities in large vector autoregressive models. However, the majority of the literature imposes a sub-Gaussian assumption on the error distribution; this is rather restrictive and excludes heavy tail distributions. For dependent data, Wu and Wu (2016) discuss the possibility of relaxing the sub-Gaussian assumption by generalizing Nagaev-type inequalities allowing for only moment assumptions. For the case of LASSO the analysis assumes the fixed design, which rules out the most important applications mentioned earlier in the introduction.

Theoretically, the LASSO tuning parameter selection requires characterizing the asymptotic distribution of the maximum of a high dimensional random vector. Chernozhukov et al. (2013a) develop a Gaussian approximation for the maximum of a sum of high-dimensional random vectors, which is in fact the basic tool for modern high-dimensional estimation. Here it is applied to the LASSO inference. Moreover, Chernozhukov et al. (2013b) deliver results for the case of $\beta$-mixing processes. Although it is quite common to assume a mixing condition which is at base a concept yielding asymptotic independence, it is not in general easy to verify the condition for a particular process, and some simple linear processes can be excluded from the strong mixing class. Andrews (1984). With an easily accessible dependency concept, Zhang and Wu (2017a) derive Gaussian approximation results for a wide class of stationary processes. Note that the dependence measure is linked to martingale decompositions and is therefore readily connected with a pool of results on tail probabilities, moment inequalities and central limit theorems of martingale theory. Our results are built on the above-mentioned theoretical works and we extend them substantially to fit into the estimation in a system of regression equations. In particular, our LASSO estimation is with random design for dependent data; therefore, we need to deal with the population implications of the Restricted Eigenvalue (RE) condition. Moreover, we show the interaction between the tail assumption and the dimensionality of the covariates in our theoretical results.

In the meantime, the issue of simultaneous inference is challenging and has motivated a series of research articles. For the case of i.i.d. data, Belloni et al. (2011, 2014), Zhang and Zhang (2014), Javanmard and Montanari (2014), van de Geer et al. (2014), Neykov et al. (2015), Chernozhukov et al. (2016), Zhu and Bradic (2018), among others, develop confidence intervals of low-dimensional variables in high-dimensional models with various forms of debiased/orthogonalization methods. Still in the case of i.i.d. data, Belloni et al. (2015b) establish a uniform post-selection inference for the target parameters defined via de-biased Huber’s Z-estimators when the dimension of the parameters of interest is potentially larger than the sample size, where they employ the multiplier bootstrap to the estimated residuals. Wild and residual bootstrap-assisted approaches are also studied in Dezeure et al. (2017), Zhang and Cheng (2017) for the case of mean regression. We pick up the line of the inference analysis of Belloni et al.
and employ it in a temporal and cross-sectional dependence framework, thus making it applicable to a rich class of high-dimensional time series. The core proof strategy is different, as it is well known that the technique for handling the suprema of empirical processes indexed by functional classes with dependent data is not the same as in i.i.d. cases. For instance, the key Bahadur representation in Belloni et al. (2015b) applies maximal inequalities derived in Chernozhukov et al. (2014) for i.i.d. random variables, while we derive the key concentration inequalities based on a martingale approximation method.

Our proposed estimation framework is complement to the literature on model selection for Gaussian Graphical model (GGM), see e.g. Yuan and Lin (2007), which has a wide spectrum of applications in statistics. A GGM can be connected with LASSO regression for estimating sparse correlation networks, and therefore is equivalent to our context with a partial correlation network, Meinshausen et al. (2006). In particular, we may find an equation-by-equation dependence can be developed. In addition, there is a big literature on social network analysis, which embeds the network information into a dynamic model in advance, see for example Zhu et al. (2017, 2019); Chen et al. (2019); Huang et al. (2016). Relatively, our approach is less structural as we treat the network structure to be unknown and uncover it using LASSO.

The following notations are adopted throughout this paper. For a vector $v = (v_1, \ldots, v_p)^\top$, let $|v|_\infty \overset{\text{def}}{=} \max_{1 \leq j \leq p} |v_j|$ and $|v|_s \overset{\text{def}}{=} (\sum_{j=1}^p |v_j|^s)^{1/s}$, $s \geq 1$. For a random variable $X$, let $\|X\|_q \overset{\text{def}}{=} (\mathbb{E} |X|^q)^{1/q}$, $q > 0$. For any function on a measurable space $g : \mathcal{W} \to \mathbb{R}$, $E_n(g) \overset{\text{def}}{=} n^{-1} \sum_{t=1}^n \{g(\omega_t)\}$ and $G_n(g) \overset{\text{def}}{=} n^{-1/2} \sum_{t=1}^n [g(\omega_t) - \mathbb{E}(g(\omega_t))]$. Given two sequences of positive numbers $x_n$ and $y_n$, write $x_n \lesssim y_n$ if there exists constant $C > 0$ such that $x_n/y_n \leq C$. For any finitely discrete measure $\mathcal{Q}$ on a measurable space, let $\mathcal{L}^q(\mathcal{Q})$ denote the space of all measurable functions $f : Z \to \mathbb{R}$ such that $\|f\|_{\mathcal{Q},q} \overset{\text{def}}{=} (\mathcal{Q}|f|^q)^{1/q} < \infty$, where $\mathcal{Q}f \overset{\text{def}}{=} \int f d\mathcal{Q}$. For a class of measurable functions $\mathcal{F}$, the $\epsilon$-covering number with respect to the $\mathcal{L}^q(\mathcal{Q})$-semimetric is denoted as $\mathcal{N}(\epsilon, \mathcal{F}, \| \cdot \|_{\mathcal{Q},q})$, and let $\text{ent}(\epsilon, \mathcal{F}) = \log \sup_{\mathcal{Q}} \mathcal{N}(\epsilon \|F\|_{\mathcal{Q},q}, \mathcal{F}, \| \cdot \|_{\mathcal{Q},q})$ with $F = \sup_{f \in \mathcal{F}} |f|$ (the envelope) denote the uniform entropy number. It should be noted that we suppress the notation of the outer expectation $\mathbb{E}^*$ to $\mathbb{E}$ and outer probability $\mathbb{P}^*$ to $\mathbb{P}$ when measurability issues are encountered. Details may be found in the Chapter 1 of Van Der Vaart and Wellner (1996).

The rest of the article is organized as follows. Section 2 shows the system model with a few examples. Section 3 introduces the sparsity method for effective prediction and provides an algorithm for the joint penalty level of LASSO via bootstrap. In Section 4 we propose approaches to implementing individual and simultaneous inference on the coefficients. Main theorems are listed in Section 5. In Section 6 and 7 we deliver the simulation studies and an empirical application on textual sentiment spillover effects. The technical proofs and other details are given in the supplementary materials. The codes to implement the algorithms are publicly accessible via the website www.quantlet.de.
2 The System Model

In this section, we present a general framework which covers many applications in statistics. Consider the system of regression equations (SRE):

\[ Y_{j,t} = X_{j,t}^\top \beta_j^0 + \varepsilon_{j,t}, \quad j = 1, \ldots, J, \quad t = 1, \ldots, n, \]

where \( X_{j,t} = (X_{jk,t})_{k=1}^{K_j} \). Without loss of generality, we assume the dimension of the covariates is identical among all equations thereafter, namely \( K_j = \dim(X_{j,t}) \equiv K \), for \( j = 1, \ldots, J \). We allow the dimension \( K \) of \( X_{j,t} \) and the number of equations, \( J \), to be large, potentially larger than \( n \), which creates an interplay with the tail assumptions on the error processes \( \varepsilon_{j,t} \). Both spatial and temporal dependency are allowed and we will obtain results on prediction and inference.

The SRE framework is a system of regression equations, which includes the following important special cases.

**Example 1 (Many Regression Models).** Suppose that we are interested in estimating the predictive models for the response variables \( U_{m,t} \):

\[ U_{m,t} = X_t^\top \gamma_m^0 + \varepsilon_{m,t}, \quad X_t \in \mathbb{R}^K, \quad \mathbb{E} \varepsilon_{m,t} X_t = 0, \quad m = 1, \ldots, M, \]

with auxiliary regressions to model predictive relations between covariates:

\[ X_{k,t} = X_{k,t}^\top \delta_k^0 + \nu_{k,t}, \quad \mathbb{E} \nu_{k,t} X_{k,t} = 0, \quad k = 1, \ldots, K, \]

where \( X_{k,t} = (X_{\ell,t})_{\ell \neq k} \in \mathbb{R}^{K-1} \), and \( \delta_k^0 \) is defined by the OLS estimator in population, namely

\[ \arg \min_{\delta_k} \frac{1}{n} \sum_{t=1}^n \mathbb{E} (X_{k,t} - X_{k,t}^\top \delta_k)^2. \]

This is a special SRE model with

\[ (Y_{j,t}, X_{j,t}, \varepsilon_{j,t}, \beta_j^0) = (U_{j,t}, X_t, \varepsilon_{j,t}, \gamma_j^0), \quad j = 1, \ldots, M, \]

\[ (Y_{j,t}, X_{j,t}, \varepsilon_{j,t}, \beta_j^0) = (X_{(j-M),t}, X_{(j-M),t}^\top, \nu_{(j-M),t}, \delta_{(j-M)}^0), \quad j = M + 1, \ldots, J = M + K. \]

It can be seen that we only put contemporaneous exogeneity conditions for \( X_t \). It is worth mentioning that this SRE case is closely related to the semiparametric estimation framework studied in Section 2.4 in [Belloni et al. (2015b)]. Here, the understanding of the predictive relations between covariates is important for constructing joint confidence intervals for the entire parameter vector \( \{ (\gamma_{mk}^0)_{k=1}^K \}_{m=1}^M \) in the main regression equations. Indeed, the construction relies on the semi-parametrically efficient point estimators obtained from the empirical analog of the following orthogonalized moment equation:

\[ \mathbb{E}[(U_{mk,t}^0 - X_{k,t}^\top \gamma_{mk}^0) \nu_{k,t}] = 0, \quad k = 1, \ldots, K, \quad m = 1, \ldots, M, \quad (2.1) \]

where \( U_{mk,t}^0 = U_{m,t} - X_{k,t}^\top \gamma_{mk}^0 \) is the response variable minus the part explained by the covariates other than \( k \). Note that the empirical analog would have all unknown nuisance parameters replaced by the estimators.

**Example 2 (Simultaneous Equation Systems (SES)).** Suppose there are many regression
equations in the following form:

\[ U_{m,t} = U_{m,t}^\top \gamma_{m,t}^0 + X_{t}^\top \gamma_{m,t}^0 + \varepsilon_{m,t}, \quad m = 1, \ldots, M. \]

Move all the endogenous variables to the left-hand side and rewrite the model in the vector form

\[ DU_t = \Gamma X_t + \varepsilon_t, \]

which is also called the structural form of the model. Suppose that \( D \) is invertible. Then the corresponding reduced form is given by

\[ U_t = BX_t + \nu_t, \quad \mathbb{E} \nu_{m,t} X_t = 0, \quad m = 1, \ldots, M, \tag{2.2} \]

with \( B = D^{-1} \Gamma \) and \( \nu_t = D^{-1} \varepsilon_t \). In this case the \( Y_{j,t} \)'s and \( X_{j,t} \)'s in SRE have no overlapping variables. A high-dimensional SES can be considered as a special case of SRE with

\[ (Y_{j,t}, X_{j,t}, \varepsilon_{j,t}, \beta_{j,t}^0) = (U_{j,t}, X_t, \nu_{j,t}, B_j^\top), \quad j = 1, \ldots, M. \]

**Example 3 (Large Vector Autoregression Models).** In the case where the covariates involve lagged variables of the response, SRE can be written as a large vector autoregression model. For example, the \( \text{VAR}(p) \) model,

\[ U_t = \sum_{\ell=1}^{p} B^\ell U_{t-\ell} + \varepsilon_t, \quad \mathbb{E} \varepsilon_{m,t} U_{t-\ell} = 0, \quad m = 1, \ldots, M, \tag{2.3} \]

where \( U_t = (U_{1,t}, U_{2,t}, \ldots, U_{M,t})^\top \), and \( \varepsilon_t \) is an \( M \)-dimensional white noise or innovation process; see e.g. Chapter 2.1 in Lütkepohl (2005). It is a special SRE case again with

\[ (Y_{j,t}, X_{j,t}, \varepsilon_{j,t}, \beta_{j,t}^0) = (U_{j,t}, (U_{t-1}^\top, \ldots, U_{t-p}^\top)^\top, \varepsilon_{j,t}, (B_j^1, \ldots, B_j^p)^\top), \quad j = 1, \ldots, M. \]

Such dynamics are of interest in biology to understand dynamic gene expression network association using micro array data, see for example Opgen-Rhein and Strimmer (2007); Ramirez et al. (2017); Dimitrakopoulou et al. (2011). It is understood that a crucial feature for many gene networks is their inherent sparsity. The issue of the number of variables involved is potentially larger than the sample size can be addressed by LASSO. Our methodology can help to analyze a gene interaction correlation network in a high dimensional regression scheme. In particular, suppose that each vertex represents a gene \( j \) collected at time point \( t \) with \( U_{j,t} \) as its gene expression and an edge connects two genes if they are correlated.

We refer to Section C.1 in the supplementary materials for more practical examples.

## 3 Effective Prediction Using Sparsity Method

In this section, we present our model setup and the LASSO estimation algorithm, including the joint penalty selection procedure.
3.1 Sparsity in SRE

The general SRE structure makes it possible to predict $Y_{j,t}$ using $X_{j,t}$ effectively. Note that the dimension of $X_{j,t}$ is large, potentially larger than $n$. Without loss of generality we assume exact sparsity of $\beta^0_j$ throughout the paper:

$$s_j = |\beta^0_j|_0 \leq s = o(n), \quad j = 1, \ldots, J.$$

(3.1)

Comment 3.1. It is now well understood that sparsity can be easily extended to approximate sparsity, in which the sorted absolute values of coefficients decrease fast to zero. To be more specific, when $\beta^0_{jk}$ is not sparse, we shall define an intermediary optimal value for our true coefficients, i.e. $\beta^*_jk$. Let $LC_p \overset{\text{def}}{=} \min_{|\beta_j|_0 \leq p} |\hat{\beta}_j - \beta^0_j|_0^2$, additionally with proper conditions on the design matrix, the optimal sparsity level is given by $s^*_j = \min_{0 \leq p \leq (K \wedge n)} LC_p^2 + \left( \max_{1 \leq k \leq K} \Psi_{jk}^2 \right) p/n$, where $\Psi_{jk}^2$ is the long run variance of $\frac{1}{\sqrt{n}} \sum_{t=1}^n \varepsilon_{j,t} X_{jk,t}$. Then the oracle $\beta^*_jk$ is defined to be $\arg \min_{|\beta_j|_0 \leq s^*_j} |\hat{\beta}_j - \beta^0_j|_0^2$. Thus an additional term involving $LC_p$ will appear in the bound in case of the true signal $\beta^0_{jk}$ is not sparse. With approximate sparsity we mean that the true signal is not sparse but nevertheless can be approximated by an exact sparsity set-up well, namely $|\beta^0_{jk}| \leq Ak^{-\gamma}$ (ranked in descending order), where $\gamma > 0.5$, and by taking $s^*_j \propto n^{1/(2\gamma)}$ the goal would be achieved.

For this situation one employs an $\ell_1$-penalized estimator of $\beta^0_j$ of the form:

$$\hat{\beta}_j = \arg \min_{\beta \in \mathbb{R}^K} \frac{1}{n} \sum_{t=1}^n \left( Y_{j,t} - X_{j,t}^T \hat{\beta} \right)^2 + \frac{\lambda}{n} \sum_{k=1}^K |\beta_{jk}| \Psi_{jk},$$

(3.2)

where $\lambda$ is the joint "optimal" penalty level and $\Psi_{jk}$'s are penalty loadings, which are defined below in (3.3).

A first aim is to obtain performance bounds with respect to the prediction norm:

$$|\hat{\beta}_j - \beta^0_j|_{j,pr} \overset{\text{def}}{=} \left[ \frac{1}{n} \sum_{t=1}^n \left( X_{j,t}^T (\hat{\beta}_j - \beta^0_j) \right)^2 \right]^{1/2},$$

where the outside $j$ indicates to use the covariates in the $j$th equation $X_{j,t}$ in computing the prediction norm, and the Euclidean norm:

$$|\hat{\beta}_j - \beta^0_j|_2 \overset{\text{def}}{=} \left\{ \sum_{k=1}^K (\hat{\beta}_{jk} - \beta^0_{jk})^2 \right\}^{1/2}.$$

To achieve good performance bounds, we first consider "ideal" choices of the penalty level and the penalty loadings. Let

$$S_{jk} = \frac{1}{\sqrt{n}} \sum_{t=1}^n \varepsilon_{j,t} X_{jk,t},$$

where for a moment we assume to be able to observe $\varepsilon_{j,t} = Y_{j,t} - X_{j,t}^T \beta^0_j$. In practice one obtains
an approximation by stepwise LASSO. Set
\[
\Psi_{jk} \overset{\text{def}}{=} \sqrt{\text{avar}(S_{jk})},
\]
\[
\lambda^0(1 - \alpha) \overset{\text{def}}{=} (1 - \alpha) - \text{quantile of } 2c\sqrt{n} \max_{1 \leq j \leq J, 1 \leq k \leq K} |S_{jk}/\Psi_{jk}|,
\]
where \( c > 1 \), e.g., \( c = 1.1 \), and \( 1 - \alpha \) is a confidence level, e.g. \( \alpha = 0.1 \), where the long run variance is denoted by \( \text{avar} \).

Theoretically, we can characterize the rate of \( \lambda^0(1 - \alpha) \) by the tail probability of \( S_{jk} \), see Theorem 5.1 also via Gaussian Approximation as in corollary 5.4. To calculate \( \lambda^0(1 - \alpha) \) from data, we can also use a Gaussian approximation based on:
\[
Q(1 - \alpha) \overset{\text{def}}{=} (1 - \alpha) - \text{quantile of } 2c\sqrt{n} \max_{1 \leq j \leq J, 1 \leq k \leq K} |Z_{jk}/\Psi_{jk}|,
\]
where \( \{Z_{jk}\} \) are multivariate Gaussian centered random variables with the same long run covariance structure as \( \{S_{jk}\} \). Alternatively, we can employ a multiplier bootstrap procedure to estimate IC empirically to achieve a better finite sample performance; see for example Cher- nozhukov et al. (2013a). In case of dependent observations over time, it is understood that data cannot be resampled directly as in the the i.i.d. case, as the dependency structure of the underlying processes will be lost. A usual solution to this problem is to consider a block bootstrap procedure, where the data are grouped into blocks, resampled and concatenated. In particular, we will adopt an estimate of IC by a multiplier block bootstrap procedure. The theoretical properties of LASSO and the tuning parameter choices are presented in Section 5.1-5.4.

### 3.2 Multiplier Bootstrap for the Joint Penalty Level

In this subsection, we introduce an algorithm to approximate the joint penalty level via a block multiplier bootstrap procedure, which is particularly nonoverlapping block bootstrap (NBB). Consider the system of equations with dependent data:
\[
Y_{j,t} = X_{j,t}^\top \beta_j^0 + \varepsilon_{j,t}, \quad \mathbb{E} \varepsilon_{j,t} X_{j,t} = 0, \quad j = 1, \ldots, J, \quad t = 1, \ldots, n,
\]

1. Run the initial \( \ell_1 \)-penalized regression equation by equation, i.e. for the \( j \)th equation,
\[
\tilde{\beta}_j = \arg \min_{\beta \in \mathbb{R}^K} \frac{1}{n} \sum_{t=1}^n (Y_{j,t} - X_{j,t}^\top \beta)^2 + \frac{\lambda_j}{n} \sum_{k=1}^K |\beta_{jk}|\Psi_{jk},
\]
where \( \lambda_j \) are the penalty levels and \( \Psi_{jk} \) are the penalty loadings. For instance, we can take the \( X \)-independence choice using Gaussian approximation (in the heteroscedasticity case): \( 2c' \sqrt{n} \Phi^{-1}\{1 - \alpha'/2K\} \) for \( \lambda_j \), where \( \Phi(\cdot) \) denotes the cdf of \( N(0,1) \), \( \alpha' = 0.1 \), \( c' = 0.5 \), and choose \( \sqrt{\text{ivar}(X_{jk,t} \tilde{\varepsilon}_{j,t})} \) for the penalty loadings, where \( \tilde{\varepsilon}_{j,t} \) are preliminary estimated errors and \( \text{ivar}(X_{jk,t} \tilde{\varepsilon}_{j,t}) \) is an estimate of the long-run variance.
\begin{equation}
\sum_{\ell=-\infty}^{\infty} \mathbb{E}(X_{jk,t} \hat{\varepsilon}_{j,t} X_{jk,(t-\ell)} \hat{\varepsilon}_{j,(t-\ell)}) \end{equation}
e.g. the Newey-West estimator is given by
\begin{equation}
\sum_{\ell=-p_n}^{p_n} k(\ell/p_n) \text{cov}(X_{jk,t} \hat{\varepsilon}_{j,t}, X_{jk,(t-\ell)} \hat{\varepsilon}_{j,(t-\ell)}),
\end{equation}
with \( k(z) = (1-|z|)1(|z| \leq 1) \). We note that the \(X\)-independent penalty (using Gaussian approximation) is more conservative, as the correlations among regressors can be adapted in the \(X\)-dependent case (using a multiplier bootstrap) with a less aggressive penalty level.

S2 Obtain the residuals for each equation by \( \hat{\varepsilon}_{j,t} = Y_{j,t} - X_{j,t}^\top \hat{\beta}_j \), and compute \( \Psi_{jk} = \sqrt{\text{var}(X_{jk,t} \hat{\varepsilon}_{j,t})} \).

S3 Divide \( \{ \hat{\varepsilon}_{j,t} \} \) into \( l_n \) blocks containing the same number of observations \( b_n \), \( n = b_n l_n \), where \( b_n, l_n \in \mathbb{Z} \). Then choose \( \lambda = 2c \sqrt{nq[B]} \), where \( q^{[B]}_{(1-\alpha)} \) is the \( (1-\alpha) \) quantile of \( \max_{1 \leq j \leq J, 1 \leq k \leq K} |Z_{jk}^{[B]} / \Psi_{jk}| \), and \( Z_{jk}^{[B]} \) are defined as
\begin{equation}
Z_{jk}^{[B]} = \frac{1}{\sqrt{n}} \sum_{i=1}^{l_n} e_{j,i} \sum_{l=(i-1)b_n+1}^{ib_n} \hat{\varepsilon}_{j,l} X_{jk,l}, \tag{3.7}
\end{equation}
e\( e_{j,i} \) are i.i.d. \( N(0,1) \) random variables independent of the data.

The bootstrap consistency regarding \( Z_{jk}^{[B]} \) is proved in Theorem 5.3

**Comment 3.2 (Block bootstrap procedures).** (i) Concerning the determination of \( b_n \), we shall report the prediction norm with several block sizes \( b_n \) and select the one with the best prediction performance in the simulation study. In addition, if it is the case that \( n \) cannot be divided by \( b_n \) with no remainder, one can simply take \( l_n = \lfloor n/b_n \rfloor \) and drop the remaining observations.

(ii) Other forms of multiplier bootstrap with any random multipliers centered around 0 can also be considered.

(iii) Alternative block bootstrap procedures can be adopted, such as the circular bootstrap and the stationary bootstrap among others; see for example Lahiri et al. (1999) for an overview.

4 Valid Inference on the Coefficients

With a reasonable fitting of LASSO on hand, we can proceed to investigate the issue of simultaneous inference. This section focuses on SRE of Example 2. We allow the covariates in each equation to be different.

The basic idea to facilitate inference is to formulate the estimation in a semi-parametric framework. With partialing out the effect of the nonparametric coefficient(s), we can achieve the desired estimation accuracy of the parametric component of interest. This trick is referred to as 'Neyman orthogonalization". Notably, the procedure is equivalent to the well known desparsification procedure in the mean square loss case, which is developed for the inference on the
estimated zero coefficients by LASSO. It thus serves the same purpose of generating a (robust) de-sparsified estimation for LASSO inference.

We list three algorithms to estimate $\beta_{jk}^0$. Algorithm 1 is easy to implement and algorithm 2 is tailored to the cases of heavy-tailed distribution of the error term, as Least Absolute Deviation (LAD) regression is well known to be robust against outliers. Algorithm 3 considers a double selection procedure aimed at remedying the bias due to omitted variables by one step selection, while also accounting for the cases of heteroscedastic errors.

Algorithm 1: LS-based algorithm

S1 Consider $Y_{j,t} = X_{jk,t} \beta_{jk}^0 + X_j^\top_{j(-k),t} \beta_{j(-k)}^0 + \varepsilon_{j,t}$, run (post) LS LASSO procedure (for each $j$), and keep the quantity $X_j^\top_{j(-k),t} \hat{\beta}_{j(-k)}^{[1]}$ for each $k$.

S2 Run LASSO (for each $j,k$) by regressing $X_{jk,t} = X_j^\top_{j(-k),t} \gamma_{j(-k)}^0 + v_{jk,t}$, and keep the residuals as $\hat{v}_{jk,t} = X_{jk,t} - X_j^\top_{j(-k),t} \hat{\gamma}_{j(-k)}$.

S3 Run LS IV regression of $Y_{j,t} - X_j^\top_{j(-k),t} \hat{\beta}_{j(-k)}^{[1]}$ on $X_{jk,t}$ using $\hat{v}_{jk,t}$ as an instrument variable, attaining the final estimator $\hat{\beta}_{jk}^{[2]}$.

Algorithm 2: LAD-based algorithm

S1 and S2 are the same as Algorithm 1.

S3′ Run LAD IV regression of $Y_{j,t} - X_j^\top_{j(-k),t} \hat{\beta}_{j(-k)}^{[1]}$ on $X_{jk,t}$ using $\hat{v}_{jk,t}$ as an instrument variable, attaining the final estimator $\hat{\beta}_{jk}^{[2]}$. We refer to Belloni et al. (2015b); Chernozhukov and Hansen (2008) for more details about how to achieve the estimator in this step.

The theoretical properties of the estimators $\hat{\beta}_{j(-k)}^{[1]}$ and $\hat{\gamma}_{j(-k)}$ in S1 and S2 are provided in Corollary 5.1 or 5.4 (see Corollary A.1 or A.4 in the supplementary correspondingly if the joint penalty over equations is employed), and Theorem A.4 for post LASSO, respectively. The uniform Bahadur representation and the Central Limit Theorem of the estimator $\hat{\beta}_{jk}^{[2]}$ in S3 or S3′ are established in Theorem 5.4 and 5.5.

Comment 4.1. Our algorithms follow patterns discussed in Belloni et al. (2015b,a) in the i.i.d. settings. The IV estimator obtained in S3 of Algorithm 1 reduced to the de-biased LASSO estimator (Zhang and Zhang, 2014; van de Geer et al., 2014) and is also first-order equivalent to the double LASSO method in Belloni et al. (2011, 2014). In particular, the estimator under LS IV regression (2-step least square regression) is given by

$$
\hat{\beta}_{jk}^{[2]} = (\hat{v}_{jk}^\top X_{jk})^{-1} \hat{v}_{jk}^\top (Y_j - X_j^\top_{j(-k),t} \hat{\beta}_{j(-k)}^{[1]})
= (\hat{v}_{jk}^\top X_{jk})^{-1} \hat{v}_{jk}^\top Y_j - \sum_{m \neq k} \hat{v}_{jk}^\top X_{jm} \hat{\beta}_{jm}^{[1]}.
$$

The second line in (4.1) is exactly the same as the de-biased or de-sparsified LASSO estimator given in Eq. (5) in Zhang and Zhang (2014) or Eq. (5) in van de Geer et al. (2014). As remarked in Belloni et al. (2015b,a), one can alternatively implement an algorithm via double selection as in Belloni et al. (2011, 2014). In particular, heteroscedastic LASSO is employed in S2″ and...
the IV regression is replaced by a either LASSO or LAD regression on the target variable and all covariates selected in the first two steps.

Algorithm 3: Double selection-based algorithm

S1′′ Run LS LASSO (for each j) of Yj,t on Xj,t:

$$\hat{\beta}_j^{[1]} = \arg \min_\beta \frac{1}{n} \sum_{t=1}^{n} (Y_{j,t} - X_{j,t}^\top \beta)^2 + \frac{\lambda}{n} |\hat{\Psi}_j|_1.$$ 

S2′′ Run Heteroscedastic LASSO (for each j, k) of Xjk,t on Xj(−k),t:

$$\hat{\gamma}_{j(−k)} = \arg \min_\gamma \frac{1}{n} \sum_{t=1}^{n} (X_{jk,t} - X_{j(−k),t}^\top \gamma)^2 + \frac{\lambda'}{n} |\hat{\Gamma}_j \gamma|_1,$$

where penalty loadings \(\hat{\Gamma}_j\) can be initialized as \(\sqrt{|\text{var}\{X_{jℓ,t}(X_{jk,t} - \frac{1}{n} \sum_{t=1}^{n} X_{jk,t}\})|}\) and then refined by \(\sqrt{|\text{var}(X_{jℓ,t}\hat{v}_{jk,t})|}\), for \(ℓ ≠ k\), and \(\hat{v}_{jk,t} = X_{jk,t} - X_{j(−k),t}\hat{\gamma}_{j(−k)}\) can be obtained by using the initial ones.

S3′′ Run LS regression of \(Y_{j,t}\) on \(X_{jk,t}\) and the covariates selected in S1′′ and S2′′:

$$\hat{\beta}_j^{[2]} = \arg \min_\beta \frac{1}{n} \sum_{t=1}^{n} (Y_{j,t} - X_{j,t}^\top \beta)^2 : \text{supp}(\beta_{−k}) \subseteq \text{supp}(\hat{\beta}_j^{[1]}_{(j(−k))}) \cup \text{supp}(\hat{\gamma}_{j(−k)}).$$

S3′′ Run LAD regression of \(Y_{j,t}\) on \(X_{jk,t}\) and the covariates selected in S1′′ and S2′′:

$$\hat{\gamma}_j^{[2]} = \arg \min_\gamma \frac{1}{n} \sum_{t=1}^{n} |Y_{j,t} - X_{j,t}^\top \beta| : \text{supp}(\beta_{−k}) \subseteq \text{supp}(\hat{\beta}_j^{[1]}_{(j(−k))}) \cup \text{supp}(\hat{\gamma}_{j(−k)}).$$

As shown in [Belloni et al. (2011)] and [Belloni et al. (2015a)], the double selection approach in S3′′ or S3′′′ creates an orthogonality condition with respect to the space spanned by the covariates selected by both steps, and thus generates an orthogonal relation to any space spanned by a linear projection of the covariates, e.g. \(\hat{v}_{jk,t}\). Therefore, the inference on the parameters may still be applied as in the framework of Algorithm 1 and 2. Therefore, one may still find the theoretical properties of estimators in S1′′, S2′′, S3′′ (S3′′′) in Section 5 according to the links mentioned above.

4.1 Confidence Interval for a Single Coefficient

We discuss an inference framework developed for a single coefficient obtained from the aforementioned algorithms.

Let \(ψ_{jk}(Z_{jt}, β_{jk}, h_{jk})\) denote the score function, where \(Z_{jt} = (Y_{j,t}, X_{jt}^\top, h_{jk}(X_{j(−k),t}) = (X_{j(−k),t} β_{j(−k)}, X_{j(−k),t} β_{j(−k)}, X_{j(−k),t} β_{j(−k)})^\top\). Consider the LAD-based case with \(ψ_{jk}(Z_{jt}, β_{jk}, h_{jk}) = \{1/2 - \mathbf{1}(Y_{j,t} ≤ X_{jk,t} β_{jk} + X_{j(−k),t} β_{j(−k)})\} v_{jk,t}\), define \(ω_{jk} \overset{\text{def}}{=} \mathbb{E}\{(\frac{1}{\sqrt{n}} \sum_{t=1}^{n} ψ_{jk,t}^0)^2\} = \sum_{\ell=−(n−1)}^{n−1}(1 − \frac{|\ell|}{n}) \text{cov}(ψ_{jk,t}^0, \psi_{jk,(t−\ell)}^0)\) with \(ψ_{jk,t}^0 \overset{\text{def}}{=} ψ_{jk}(Z_{jt}, β_{jk}^0, h_{jk})\), and \(ϕ_{jk} \overset{\text{def}}{=} \partial \mathbb{E}\{(ψ_{jk}(Z_{jt}, β_{jk}^0, h_{jk}^0))\} / \partial β_{jk}^0\).
Suppose we are interested in testing $H_0 : \beta_{jk}^0 = 0$. For this purpose we employ the uniform Bahadur representation (Theorem 5.4) to construct the confidence interval via a multiplier bootstrap procedure. In particular, the distribution of the asymptotically pivotal statistics:

$$T_{jk} = \frac{\sqrt{n}(\hat{\beta}_{jk}^2 - \beta_{jk}^0)}{\hat{\sigma}_{jk}},$$

is approximated via its block multiplier bootstrap counterpart:

$$T_{jk}^* = \frac{1}{\sqrt{n}} \sum_{i=1}^{l_n} e_{j,i} \sum_{l=(i-1)b_n+1}^{ib_n} \hat{\zeta}_{jk,l},$$

where $\hat{\zeta}_{jk,l}$ are pre-estimators of $\zeta_{jk,l} = -\phi_{jk}^{-1} \sigma_{jk}^{-1} \psi_{jk,l}$ such that $\max_{(j,k), (j',k')} |\sum_{l=1}^{l_n} \hat{\eta}_{j'k',i} n_{j,k,l} - \sum_{l=1}^{l_n} \eta_{j'k',i} n_{j,k,l}| = o_p(\{\log(JK)\}^{-2})$, with $\eta_{j,k,l} = \frac{1}{\sqrt{n}} \sum_{l=(i-1)b_n+1}^{ib_n} \zeta_{jk,l}$ and $\hat{\eta}_{j,k,l} \equiv \frac{1}{\sqrt{n}} \sum_{l=(i-1)b_n+1}^{ib_n} \hat{\zeta}_{jk,l}$, $e_{j,i}$ are independently drawn from $N(0,1)$, $l_n$ and $b_n$ are the numbers of blocks and block size, respectively.

Let $\hat{\sigma}_{jk}$ be any consistent estimator of $\sigma_{jk}$. Then the confidence interval is given by

$$\text{CI}_{jk}(\alpha) : [\hat{\beta}_{jk}^2 - \hat{\sigma}_{jk} n^{-1/2} q_{jk}^*(1 - \alpha), \hat{\beta}_{jk}^2 + \hat{\sigma}_{jk} n^{-1/2} q_{jk}^*(1 - \alpha)],$$

where $q_{jk}^*(1 - \alpha)$ is the $(1 - \alpha)$ quantile of the bootstrapped distribution of $|T_{jk}^*|$.

**Comment 4.2** (Asymptotic Normality of $\hat{\beta}_{jk}^2$). As shown in Corollary 5.5 we have the limit distribution of $\hat{\beta}_{jk}^2$:

$$\sigma_{jk}^{-1} n^{1/2}(\hat{\beta}_{jk}^2 - \beta_{jk}^0) \overset{D}{\to} N(0,1),$$

where $\sigma_{jk} = (\phi_{jk}^{-2} \omega_{jk})^{1/2}$. Therefore, the two-sided $100(1 - \alpha)$ confidence interval by asymptotic normality for $\beta_{jk}^0$ is given by

$$\text{CI}_{jk}(\alpha) : [\hat{\beta}_{jk}^2 - \hat{\sigma}_{jk} n^{-1/2} \Phi^{-1}(1 - \alpha/2), \hat{\beta}_{jk}^2 + \hat{\sigma}_{jk} n^{-1/2} \Phi^{-1}(1 - \alpha/2)].$$

**Comment 4.3** (Residual Multiplier Bootstrap). Alternative bootstrap procedures may be considered as well, e.g. the residual multiplier bootstrap procedure:

$$\hat{\varepsilon}_{j,t} = Y_{j,t} - X_{j,t}^\top \hat{\beta}_{j}^{[1]} ,$$

then divide $\{\hat{\varepsilon}_{j,t}\}$ into $l_n$ blocks of size $b_n$, where $b_n l_n = n$, and for each block $i = 1, \ldots, l_n$, $i b_n + 1, \ldots, i b_n$.

Define $Y_{j,t}^* = X_{j,t}^\top \hat{\beta}_{j}^{[1]} + \hat{\varepsilon}_{j,t}^*$ and compute the bootstrap counterpart as

$$T_{jk}^* = \frac{\sqrt{n}(\hat{\beta}_{jk}^2 - \hat{\beta}_{jk}^0)}{\hat{\sigma}_{jk}^*}.$$
where \( \hat{\beta}_{jk}^* \) and \( \tilde{\sigma}_{jk}^* \) are estimated using the bootstrap sample \( \{Y^*_{j,t}, X_{j,t}\} \).

4.2 Joint Confidence Region for Simultaneous Inference

We now continue to extend the single coefficient inference to simultaneous inference on a set of coefficients. As shown in the practical examples in Section [C.1], it is essential to conduct simultaneous inference on a group of parameters \( G \). In this case, the null hypothesis is: \( H_0 : \beta^0_{jk} = 0, \forall (j,k) \in G, \) and the alternative \( H_A : \beta^0_{jk} \neq 0, \) for some \( (j,k) \in G, \) where the group \( G \) is a set of coefficients with cardinality \( |G| \). Suppose for the \( j \)-th equation there are \( p_j \) target coefficients and the cardinality \( |G| = \sum_{j=1}^J p_j \). This can be understood as a multiple estimation problem compared to Section 4.1. Without loss of generality, we can rearrange the order of the variables and rewrite the regression equation for each \( j \) as (consider the LAD-based model here)

\[
Y_{j,t} = \sum_{l=1}^{p_j} X_{j,t}^l \beta^0_{jl} + \sum_{l=p_j+1}^{K} X_{j,t}^l \beta^0_{jl} + \varepsilon_{j,t}, \quad F_{\varepsilon_j}(0) = 1/2
\]

(4.7)

One follows the algorithms to obtain \( \hat{\beta}_{jl}^* \) \((1 \leq l \leq p_j)\) for each \( j \). Then the idea of simultaneous inference is very straightforward. We aggregate the statistics \( T_{jk} \) in (4.2) by taking the maximum and minimum over the set \( G \). Finally, the component-wise confidence interval is constructed with the quantiles of the bootstrap statistics over all bootstrap samples. Denote \( q_G^*(1 - \alpha) \) as the \((1 - \alpha)\) quantile of \( \max_{(j,k) \in G} |T^*_{jk}| \). A joint confidence region is then:

\[
\{ \beta \in \mathbb{R}^{|G|} : \max_{(j,k) \in G} T_{jk} \leq q_G^*(1 - \alpha) \text{ and } \min_{(j,k) \in G} T_{jk} \geq -q_G^*(1 - \alpha) \},
\]

(4.8)

and for each component \( (j,k) \in G \), the confidence interval \( \tilde{CI}_{jk}(\alpha) \) is given by \( [\hat{\beta}_{jk}^* - \tilde{\sigma}_{jk}^* n^{-1/2} q_G^*(1 - \alpha), \hat{\beta}_{jk}^* + \tilde{\sigma}_{jk}^* n^{-1/2} q_G^*(1 - \alpha)] \). We show in Corollary 5.7 the consistency of this bootstrap confidence band for simultaneous inference. Note that when there is only one parameter in \( G \) for inference, the joint confidence region (4.8) will reduce to the single parameter confidence interval (4.4) as a special case.

5 Main Theorems

In this section, we present the theoretical foundations for the procedures given earlier. In particular, we discuss the properties of the theoretical choices of penalty level and the validity of the other two empirical choices, as well as the theoretical support for the simultaneous inference.

Throughout the whole section, we define \( S_{jk} \equiv n^{-1/2} \sum_{t=1}^n \varepsilon_{j,t} X_{j,k,t}, \) \( S_j = (S_{jk})_{k=1}^K, \) and \( \Psi_{jk} \equiv \sqrt{\text{avar}(S_{jk})} \), which is the square root of the long-run variance of \( X_{j,k,t} \varepsilon_{j,t} \), namely \( \{ \sum_{t=-\infty}^\infty \text{E}(X_{j,k,t} X_{j,k,(t-\ell)} \varepsilon_{j,t} \varepsilon_{j,(t-\ell)}) \}^{1/2} \). Recall that for a single equation LASSO, we select the penalty in the following ways:

a) theoretically, for each regression, \( \lambda_j = \lambda_j^0(1 - \alpha) \) (IC), i.e. the \((1 - \alpha)\) quantile of \( 2c \sqrt{n} \max_{1 \leq k \leq K} |S_{jk}/\Psi_{jk}| \) (note that this penalty takes into account the correlation among regressors and is design adaptive):
b) an empirical choice given a Gaussian approximation result is $Q_j(1 - \alpha)$, which is defined to be the $(1 - \alpha)$ quantile of $2c \max_{1 \leq k \leq K} \sqrt{n} |Z_{jk}/\Psi_{jk}|$, where $Z_{jk}$'s are multivariate Gaussian centered random variables with the same long run covariance structure as $S_{jk}$. Alternatively, a canonical choice disregarding the correlation among regressors can be considered as $\tilde{Q}_j(1 - \alpha) \defeq 2c\sqrt{n} \Phi^{-1}(1 - \alpha/(2K))$. We shall note that $Q_j(1 - \alpha)$ is not feasible but can be estimated by simulations of Gaussian random variable $Z_{jk}$ with estimated long run variance covariance matrix. Typically $\tilde{Q}_j(1 - \alpha)$ is more conservative than $Q_j(1 - \alpha)$.

c) another empirical choice of the penalty level is $\Lambda_j(1 - \alpha)$ as the $(1 - \alpha)$ quantile of $2c \max_{1 \leq k \leq K} |Z_{jk}^{(\beta)}/\hat{\Psi}_{jk}|$'s are defined in (3.7), and obtainable via the multiplier block bootstrap technique.

5.1 Near Oracle Inequalities under IC

We first provide the near oracle inequalities for the single equation LASSO estimation $\hat{\beta}_j$ obtained from (5.6) under the ideal choices (IC). For this purpose, a few assumptions and definitions are required.

(A1) For $j = 1, \ldots, J, k = 1, \ldots, K$, let $X_{jk,t}$ and $\varepsilon_{j,t}$ be stationary processes admitting the following representation forms $X_{jk,t} = g_{jk}(F_t) = g_{jk}(\ldots, \xi_{t-1}, \xi_t)$ and $\varepsilon_{j,t} = h_j(F_t) = h_j(\ldots, \eta_{t-1}, \eta_t)$, where $\xi_t, \eta_t$ are i.i.d. random elements (innovations or shocks, allowing for overlap, see Comment 5.1) across $t$, $\xi_t = (\ldots, \xi_{t-1}, \xi_t, \eta_t)$, $g_{jk}(\cdot)$ and $h_j(\cdot)$ are measurable functions (filters). $E(X_{jk,t}\varepsilon_{j,t}) = 0$, for any $j, k \in 1, \ldots, J, 1, \ldots, K$.

**Definition 5.1.** Let $\xi_0$ be replaced by an i.i.d. copy of $\xi_0^*$, and $X_{jk,t}^* = g_{jk}(\ldots, \xi_0^*, \ldots, \xi_{t-1}, \xi_t)$. For $q \geq 1$, define the functional dependence measure $\delta_{q,j,k,t} \defeq \|X_{jk,t} - X_{jk,t}^*\|_q$, which measures the dependency of $\xi_0$ on $X_{jk,t}$. Also define $\Delta_{m,q,j,k} \defeq \sum_{t=m}^{\infty} \delta_{q,j,k,t}$, which measures the cumulative effect of $\xi_0$ on $X_{jk,t}$, $m \geq m_0$. Moreover, we introduce the dependence adjusted norm of $X_{jk,t}$ as $\|X_{jk,\cdot}\|_{q,\xi} \defeq \sup_{m \geq m_0} (m + 1)^{\xi} \Delta_{m,q,j,k}(\xi > 0)$. Similarly, let $\eta_0$ be replaced by an i.i.d. copy of $\eta_0^*$, and $\varepsilon_{j,t}^* = h_j(\ldots, \eta_0^*, \ldots, \eta_{t-1}, \eta_t)$, we define $\|\varepsilon_{j,\cdot}\|_{q,\xi} \defeq \sup_{m \geq m_0} (m + 1)^{\xi} \sum_{t=m}^{\infty} \|\varepsilon_{j,t} - \varepsilon_{j,t}^*\|_q$ and $\|X_{jk,\cdot}\|_{q,\xi} \defeq \sup_{m \geq m_0} (m + 1)^{\xi} \sum_{t=m}^{\infty} \|X_{jk,t}\varepsilon_{j,t} - X_{jk,t}\varepsilon_{j,t}^*\|_q$.

It should be noted that [A1] admits a wide class of processes. The largest value of $\xi$ which ensures a finite dependence adjusted norm characterizes the dependency structure of the process. The moment-based measure is directly connected with the impulse functions. A few examples for univariate time series $Z_t$ are listed in Appendix [C.2] in the supplementary materials.

(A2) Restricted eigenvalue (RE): given $\tilde{c} \geq 1$, for $\delta \in \mathbb{R}^K$, with probability $1 - o(1)$,

$$\kappa_j(\tilde{c}) \defeq \min_{|\delta_{T_j}| \leq \tilde{c}|\delta_{T_j}|, \delta_{T_j} \neq 0} \frac{\sqrt{\sigma_j} |\delta|_{j,pr}}{|\delta_{T_j}|} > 0,$$

where $T_j \defeq \{k : \beta_{jk}^0 \neq 0\}$ and $s_j = |T_j| = o(n)$, $\delta_{T_j,k} = \delta_k$ if $k \in T_j$, $\delta_{T_j,k} = 0$ if $k \notin T_j$.

(A3) $\|\varepsilon_{j,\cdot}\|_{q,\xi} < \infty$ and $\|X_{jk,\cdot}\|_{q,\xi} < \infty$ ($q \geq 8$).
Comment 5.1. We allow for overlap in the elements in $\xi_t$ and $\eta_t$, as long as the contemporaneous exogeneity condition $E(X_{jk,t\xi_{j,t} = 0}$ is satisfied. For example, consider the VAR(1) model: $Y_t = AY_{t-1} + \varepsilon_t$, with $Y_t, \varepsilon_t \in \mathbb{R}^J$, and suppose that $Y_t$ admits the representation $Y_t = \sum_{t=0}^{\infty} A^t \varepsilon_{t-1}$ with $\varepsilon_{t-1}$ as measurable functions of $\xi_{-\infty}, \ldots, \xi_{t-1}$. Thus $X_{jk,t} = g_{jk}(\varepsilon_{-\infty}, \ldots, \xi_{t-1}) = \sum_{t=0}^{\infty} [A^t]_{jk,t} \varepsilon_{t-1}$, where $[A^t]_{jk}$ is the $k$th row of the matrix $A^t$, $k = 1, \ldots, J$. In this case no serial correlation in the innovations $\varepsilon_t$'s would be sufficient for $E(X_{jk,t\xi_{j,t} = 0}$.

Comment 5.2. We show in Theorem B.1 (see the supplementary materials) that the RE (A2) and RSE (A5) conditions can be implied by assumptions on the corresponding population variance-covariance matrix. This illustrates the feasibility of the RE/RSE assumption.

Lemma 5.1 (Prediction Performance Bound of Single Equation LASSO). Suppose (A1) and (A2) (with $\bar{c} = \frac{q}{q} + 1$, $c > 1$), under the exact sparsity assumption (3.1) and given the event $\lambda_j > 2c\sqrt{n} \max_{1 \leq k \leq K} |S_{jk}/\Psi_{jk}|$ and another event which RE holds, then with probability $1 - o(1)$, $\beta_j$ obtained from (3.6) satisfy

$$|\tilde{\beta}_j - \beta_j|_{j,pr} \leq (1 + 1/c) \frac{\sqrt{\lambda_j}}{n\kappa_j(\bar{c})} \max_{1 \leq k \leq K} \Psi_{jk}. \tag{5.1}$$

In addition, if (A2) (with $2\bar{c}$) holds, then with probability $1 - o(1)$,

$$|\tilde{\beta}_j - \beta_j|_1 \leq \frac{(1 + 2\bar{c})\sqrt{\lambda_j}}{\kappa_j(2\bar{c})} |\tilde{\beta}_j - \beta_j|_{j,pr}. \tag{5.2}$$

Lemma 5.1 follows Theorem 1 of [Belloni and Chernozhukov, 2013]. As the proof is built on inequalities and for the case of dependent data [A1] they remain unchanged, we omit the detailed proof here. To further characterize the rate of IC, we provide a tail probability for $2c\sqrt{n} \max_{1 \leq k \leq K} |S_{jk}/\Psi_{jk}|$ under the moment assumption (A3). In particular, the rate depends on the dependence adjusted norm $\|X_{jk}, \varepsilon_j, \|_{q,\varsigma}$.

Theorem 5.1. Under (A1) and (A3), we have

$$P(2c\sqrt{n} \max_{1 \leq k \leq K} |S_{jk}/\Psi_{jk}| \geq r) \leq C_1 \varpi_n nr^{-q} \sum_{k=1}^{K} \frac{\|X_{jk}, \varepsilon_j, \|_{q,\varsigma}^q}{\Psi_{jk}^q} + C_2 \sum_{k=1}^{K} \exp \left( -\frac{C_3 r^2 \Psi_{jk}^2}{n \|X_{jk}, \varepsilon_j, \|_{2,\varsigma}^2} \right), \tag{5.3}$$

where for $\varsigma > 1/2 - 1/q$ (weak dependence case), $\varpi_n = 1$; for $\varsigma < 1/2 - 1/q$ (strong dependence case), $\varpi_n = n^{q/2-1-c}$. $C_1, C_2, C_3$ are constants depending on $q$ and $\varsigma$.

Comment 5.3. It can be seen in Theorem 5.1 that the rate of the dependence adjusted norm $\|X_{jk}, \varepsilon_j, \|_{q,\varsigma}$ plays an important role in the tail probability for $2c\sqrt{n} \max_{1 \leq k \leq K} |S_{jk}/\Psi_{jk}|$. Here we discuss the rate under some special cases.

1. **VAR(1):** Consider the VAR(1) model given by $Y_t = AY_{t-1} + \varepsilon_t$, with $Y_t, \varepsilon_t \in \mathbb{R}^J$, and $\varepsilon_t \sim i.i.d. \text{N}(0, \Sigma)$. In this case $X_{jk,t} = Y_{j,t-1}$ and $K = J$. Suppose there exists a stationary representation of the model as $Y_t = \sum_{t=0}^{\infty} A^t \varepsilon_{t-1}$. Then we have $\|X_{jk,t}, \varepsilon_{j,t} - X_{jk,t}^{*,e_{j,t}}\|_q = \|Y_{j,t-1} \varepsilon_{j,t} - Y_{j,t-1}^{*,e_{j,t}}\|_q = \|([A^t]_j (\varepsilon_{0} - \varepsilon_{0}^*) \varepsilon_{j,t} \|_q \leq 2\|([A^{t-1}]_j_\mu_q, where $\mu_q = \def$}
max_j \|\varepsilon_{j,t}\|_q \text{ and } [A^{-1}]_j \text{ is the jth row of the matrix } A^{-1}. \text{ Assume } \max_j |[A^t]|_1 \leq |c|^t \text{ with } |c| < 1 \text{ (a geometric decay rate). It follows that } ||X_{jk,t}\varepsilon_{j,t},\eta_{k,t}||_\infty = \frac{2\nu}{|c|^t-1} \text{ sup}_{m \geq 0} \left((m+1)^t \sum_{z=m}^\infty |c|^{-1} \leq (C/|c|) \vee (C(m^*+1)|c|m^{-1})\right), \text{ where } m^* = (-c/ \log |c| - 1) \vee 0 \text{ and } C > 0 \text{ depends on } \mu_q. \text{ Moreover, to justify the geometric decay rate, we consider the example of Network Autoregressive (NAR) model as in } \text{[Zhu et al. (2017)]} \text{ with } A = \rho W, \text{ where } W \text{ is a row-normalized adjacency matrix which is pre-specified to indicate the social network connectedness and } \rho \text{ is the network parameter suggesting the strength of the network effects. In that case, assuming a geometric decay rate } \max_j |[A^t]|_1 \leq |c|^t \text{ with } |c| < 1 \text{ again gives similar results.}

2. Spatial MA structure in } \varepsilon_t: \text{ Consider the model } Y_{j,t} = X_{j,t}^T \beta_j + \varepsilon_{j,t}, \text{ with } \varepsilon_t = \rho W \varepsilon_t + \eta_t, \text{ where } W \text{ is a spatial weight matrix, } \eta_t \text{ are i.i.d. and have finite qth moments } \mu^\eta_q \triangleq \max_j \|\eta_{j,t}\|_q. \text{ For simplicity, here we assume } X_{j,t} \text{ and } \varepsilon_{j,t} \text{ are independent. Suppose there exists a stationary representation of the error process given by } \varepsilon_t = \sum_{l=0}^\infty \rho^l W^l \eta_{t-l}. \text{ Then we have } ||X_{jk,t}\varepsilon_{j,t} - X_{jk,t}^* \varepsilon_{j,t}||_q \leq ||(X_{jk,t} - X_{jk,t}^*) \varepsilon_{j,t}||_q + ||X_{jk,t} \varepsilon_{j,t} - X_{jk,t}^* \varepsilon_{j,t}||_q \leq ||X_{jk,t} - X_{jk,t}^*||_q ||\varepsilon_{j,t}||_q + ||X_{jk,t}||_q ||\rho W||_1 (\varepsilon_{j,t} - \varepsilon_{j,t}^*)||_q \leq ((1 - \rho \varepsilon_{j,t} - \varepsilon_{j,t}^*))||_q \leq \max_j \|\rho W||_1 \leq |c|^t \text{ with } |c| < 1. \text{ It follows that } ||X_{jk,t} \varepsilon_{j,t}||_q \leq C_1 ||X_{j,k} \varepsilon_{j,k}||_\infty + C_2 \sup_{m \geq 0} (m+1)^t \sum_{z=m}^\infty |c|_t \leq C_1 ||X_{j,k} \varepsilon_{j,k}||_\infty + C_3 (m^*+1)|c|m^{-1}, \text{ where } m^* = (-c/ \log |c| - 1) \vee 0 \text{ and } C_1, C_2, C_3 > 0 \text{ depend on } \mu^\eta_q \text{ and } ||X_{j,k,t}||_q. $

3. General linear processes: \text{ To study more general spatial and temporal dependency, consider the model } Y_{j,t} = X_{j,t}^T \beta_j + \varepsilon_{j,t}, \text{ with } \varepsilon_t = \sum_{l=0}^\infty A^l \eta_{t-l}. \text{ Again } \eta_t \text{ are i.i.d. and have finite qth moments } \mu^\eta_q \triangleq \max_j \|\eta_{j,t}\|_q. \text{ If all the } A^t \text{ are diagonal matrices, there is just temporal dependence, and if } A^t = 0 \text{ for } t \geq 1 \text{ there exists only spatial dependence. Let } a^j_{k,t} \triangleq |A^t|_{jk} \text{ be the element on the jth row and kth column of } A^t. \text{ Assume } \sum_{l=0}^\infty \sum_k |a^j_{k,t}| < \infty, X_{j,t} \text{ and } \varepsilon_{j,t} \text{ to be independent. We have } ||X_{jk,t} \varepsilon_{j,t}||_q \leq C_1 \|\varepsilon_{j,t}||_\infty + C_2 \sup_{m \geq 0} (m+1)^t \sum_{l=m}^\infty \sum_k |a^j_{ik,t}|, \text{ where } C_1, C_2 > 0 \text{ depend on } \mu^\eta_q \text{ and } ||X_{j,k}||_q. \text{ Moreover, we have } ||\max_j (X_{jk,t} \varepsilon_{j,t})||_\infty \leq ||\max_j X_{j,k} \varepsilon_{j,k}||_\infty \leq |\max_j X_{jk,t} \varepsilon_{j,t}||_q \text{ and especially } ||\varepsilon_{j,t}||_q \leq ||\max_j \sum_k a^j_{ik,t} (\eta_{k,t} - \eta_{k_0})||_q \leq \max_k \sum_{j=1}^J \left( \sum_{k} \sum_{j=1}^J (a^j_{k,t})^2 (\mu^\eta_q)^2 \right)^{1/2} \leq \sum_k \max_j |a^j_{k,t}| \mu^\eta_q \vee \sqrt{q \log J} \left( \sum_k \max_j |a^j_{k,t}|^2 \right)^{1/2} \mu^\eta_q \text{, where the Rosenthal-Burkholder inequality is applied. Suppose that } \sum_{l=0}^\infty \sum_k |a^j_{k,t}| \leq J(m \vee 1)^{-c}, \text{ for some constant } c > 0. \text{ If } \xi < c, \text{ we have } ||\max_j \varepsilon_{j,t}||_q \leq C_3 \sup_{m \geq 1} (m+1)^t (m \vee 1)^{-c} J \sqrt{\log J} \leq C_3 \sup_{m \geq 1} (m+1)^t (m \vee 1)^{-c} J \sqrt{\log J}, \text{ where } C_3 > 0 \text{ depends on } \mu^\eta_q. \text{ To summarize, if the qth moments are bounded by constant, the dependence adjusted norm } ||X_{jk,t} \varepsilon_{j,t}||_\infty \text{ is also bounded in the first two examples where a geometric decay rate on the coefficients is assumed; while in the case of general linear processes, it would depend on the rate of } \sum_{l=0}^\infty \sum_k |a^j_{k,t}|. \text{ In particular, suppose } \sum_{l=0}^\infty \sum_k |a^j_{k,t}| \leq (m \vee 1)^{-c} \text{ for } c > 0. \text{ If } c > \xi, \text{ } ||X_{jk,t} \varepsilon_{j,t}||_q \text{ is bounded (assume } ||X_{j,k}||_\infty \text{ is bounded).}$

Under the choice (IC) \( \lambda_j^0 (1 - \alpha) \) is given by the (1 - \alpha) quantile of \( 2c \sqrt{n} \max_{1 \leq j \leq \lambda} |S_{jk}/\Psi_{jk}| \), combining the results of Lemma 5.1 and Theorem 5.1, we can get the bounds for \( \lambda_j^0 (1 - \alpha) \) and further obtain the oracle inequalities as in Corollary 5.1.
Corollary 5.1 (Bounds for $\lambda^0_j(1 - \alpha)$ and Oracle Inequalities under IC). Under (A1) (A3), given $\lambda^0_j(1 - \alpha)$ satisfying

$$
\lambda^0_j(1 - \alpha) \lesssim \max_{1 \leq k \leq K} \left\{ \|X_{jk, \varepsilon_{jk}}\|_{\varphi, \kappa} \sqrt{n \log(K/\alpha)} \vee \|X_{jk, \varepsilon_{jk}}\|_{q, \varsigma}(n \varpi_n K/\alpha)^{1/q} \right\},
$$

(5.4)

and the exact sparsity assumption (3.1), then $\tilde{\beta}_j$ obtained from (3.6) under IC satisfies

$$
|\tilde{\beta}_j - \beta^0_j|_{\varphi, r} \lesssim \max_{1 \leq k \leq K} \Psi_{jk}\left\{ \|X_{jk, \varepsilon_{jk}}\|_{\varphi, \kappa} \sqrt{n \log(K/\alpha)} \vee \|X_{jk, \varepsilon_{jk}}\|_{q, \varsigma} n^{1/q - 1}(n \varpi_n K/\alpha)^{1/q} \right\},
$$

(5.5)

with probability $1 - \alpha - o(1)$, where for $\varsigma > 1/2 - 1/q$ (weak dependence case), $\varpi_n = 1$; for $\varsigma < 1/2 - 1/q$ (strong dependence case), $\varpi_n = n^{q/2 - 1 - \varsigma q}$.

Comment 5.4. The Nagaev type of inequality in (5.3) has two terms, namely an exponential term and a polynomial term. It should be noted that if the polynomial term dominates, the above bound does not allow for ultra high dimension of $K$. Basically, we only allow for a polynomial rate $K = O(n^\epsilon)$, and the rate of $K$ interplays with the dependence adjusted norm $\|X_{jk, \varepsilon_{jk}}\|_{q, \kappa}$. In particular, to make sure that the estimators are consistent (i.e. the error bounds tend to zero for sufficiently large $n$), for example, we need $\hat{c} < q - 1 - \nu q/2 - dq$, if there exists $q$ to guarantee $\|X_{jk, \varepsilon_{jk}}\|_{q, \kappa} = O(n^d)$ and $0 < \nu < 1$ such that $s_j = O(n^\nu)$.

We now discuss the case of sub-Gaussian tail or sub-exponential tail, which is mostly assumed in the literature.

Comment 5.5. Suppose a stronger exponential moment condition is satisfied,

$$
\|X_{jk, \varepsilon_{jk}}\|_{\psi, \kappa} = \sup_{q \geq 2} q^{-\nu}\|X_{jk, \varepsilon_{jk}}\|_{q, \kappa} < \infty,
$$

(5.6)

where $\|X_{jk, \varepsilon_{jk}}\|_{\psi, \kappa}$ is interpreted as the dependence adjusted sub-exponential ($\nu = 2$) or sub-Gaussian ($\nu = 1$) norm. Consider the special case of VAR(1). As shown above, we have $\|X_{jk, \varepsilon_{jk}, t} - X_{jk, \varepsilon_{jk}, t}\|_q \leq 2\|A^{t-1}\|_1 \mu_2^2$. In particular, it is known that $\mu_q \lesssim q$ for sub-exponential variables and $\mu_q \lesssim \sqrt{q}$ for sub-Gaussian variables. Let $\nu = 2$ and $\nu = 1$ for the two cases respectively, $\|X_{jk, \varepsilon_{jk}}\|_{\psi, \kappa} \lesssim (m^* + 1)|c|^{m^* - 1}$. Then applying the exponential tail bounds as in Lemma B.3 in the supplementary material, we arrive at the following error bounds with probability $1 - \alpha - o(1)$,

$$
|\tilde{\beta}_j - \beta^0_j|_{\varphi, r} \lesssim \max_{1 \leq k \leq K} \left\{ \Psi_{jk}\frac{\log(K/\alpha)}{\sqrt{n}} \right\}^{1/\gamma}, \quad \gamma = 2/(2\nu + 1),
$$

(5.7)

as $\lambda^0_j(1 - \alpha) \lesssim \sqrt{n}(\log K)^{1/\gamma} \max_{1 \leq k \leq K} \|X_{jk, \varepsilon_{jk}}\|_{\psi_0, 0}$. The bound (5.7) works with ultra-high dimensional rate $\exp(n^{r\gamma})$ ($r < 1$) of $K$ as only the exponential term shows in the inequality. In particular, suppose $s_j = O(n^\nu)$, and $\|X_{jk, \varepsilon_{jk}}\|_{\psi_0, 0} = O(n^d)$, then $r + d + \nu/2 < 1/2$ is required to ensure the consistency.
5.2 Gaussian Approximation for Dependent Data

Now we look at the validity of the choice of \( Q_j(1-\alpha) \), which relies on a Gaussian approximation theorem. First we define the Kolmogorov distance between any two \( K \)-dim random vectors.

**Definition 5.2.** Let \( X = (X_1, \cdots, X_K)^T \in \mathbb{R}^K, \ Y = (Y_1, \cdots, Y_K)^T \in \mathbb{R}^K \). The Kolmogorov distance between \( X \) and \( Y \) is defined as

\[
\rho(X, Y) = \sup_{r \geq 0} | P(|X|_\infty \geq r) - P(|Y|_\infty \geq r) |.
\]

For each single equation \( j \), aggregate the dependence adjusted norm over \( k = 1, \ldots, K \):

\[
\|X_j,|_\infty\|_{\varphi,\gamma} \overset{\text{def}}{=} \sup_{m \geq 0} (m+1)^{\gamma} \sum_{t=m}^{\infty} \delta_{q,j,t}, \quad \delta_{q,j,t} \overset{\text{def}}{=} \|X_{j,t} - X_{j,t}^\ast\|_\varphi \quad \overset{(5.8)}{=} \quad \frac{\sum_{k=1}^{K} \|X_{jk} - \varphi\|_{\varphi,\gamma}}{q,\gamma},
\]

where \( q \geq 1 \) and \( \gamma > 0 \). Moreover, define the following quantities

\[
\Phi_{j,\varphi,\gamma} \overset{\text{def}}{=} 2 \max_{1 \leq k \leq K} \|X_{jk},|_\infty\|_{\varphi,\gamma}, \quad \Gamma_{j,\varphi,\gamma} \overset{\text{def}}{=} 2 \|\varepsilon_{j,|}_{\infty}\|_{\varphi,\gamma}, \quad \Theta_{j,\varphi,\gamma} \overset{\text{def}}{=} \Gamma_{j,\varphi,\gamma} \wedge \{2\|X_{j,|}_{\infty}\|_{\varphi,\gamma}(\log K)^{3/2}\} \quad \overset{(5.9)}{=} \quad \{2\|X_{j,|}_{\infty}\|_{\varphi,\gamma}(\log K)^{3/2}\}.
\]

It is worth noting that the norm \( \|X_{j,|}_{\infty}\|_{\varphi,\gamma} \) is a kind of aggregated dependence adjusted norm for a vector of processes in comparison to the dependence adjusted norm for a univariate process as in Definition 5.1.

Some additional assumptions are required. Define \( L_{1,j} = \{\Phi_{j,\varphi,\gamma,\gamma}\Phi_{j,\varphi,\gamma,\gamma}\}^{1/\gamma}, \ W_{1,j} = \{\Phi_{j,\varphi,\gamma,\gamma}\Phi_{j,\varphi,\gamma,\gamma}\}^{1/\gamma}, \ W_{2,j} = \Phi_{j,\varphi,\gamma,\gamma}\}^{1/\gamma}, \ W_{3} = [n^{-\gamma}\{\log(Kn)\}]^{3/2}\Theta_{j,\varphi,\gamma}\}^{1/(1/2-\gamma-1/2)}\}, \ N_{1,j} = (n/\log K)^{2/\gamma}\Theta_{j,\varphi,\gamma}\}, \ N_{2,j} = (n/\log K)^{-2}\Phi_{j,\varphi,\gamma}\}, \ N_{1,j} = (n/\log K)^{2/\gamma}\Theta_{j,\varphi,\gamma}\}, \ N_{2,j} = (n/\log K)^{-2}\Phi_{j,\varphi,\gamma}\}^{1/(1/2-\gamma)}\}

(A4) i) (weak dependency case) Given \( \Theta_{j,2\varphi,\gamma} < \infty \) with \( q \geq 4 \) and \( \gamma > 1/2 - 1/q \), then \( \Theta_{j,2\varphi,\gamma,n1/q-1/2}\}^{3/2}\} \) and \( L_1 \max(W_{1,j}, W_{2,j}) = c(1) \min(N_{1,j}, N_{2,j}). \)

ii) (strong dependency case) Given \( \Theta_{j,2\varphi,\gamma,\{\log(K\}}^{1/2} = c(n^{\gamma}) \) and \( L_1 \max(W_{1,j}, W_{2,j}, W_{3,j}) = c(1) \min(N_{2,j}, N_{3,j}). \)

The assumptions impose mild restrictions on the dependency structure of covariates and error terms. They include a wide class of potential correlation and heterogeneity (including conditional heteroscedasticity), with possible allowance of the lagged dependent variables. Two examples of large VAR and ARCH for high-dimensional time series can be found in Appendix C.2 in the supplementary materials.

**Comment 5.6.** [Admissible Dimension Rates by the Conditions for Gaussian Approximation] As discussed in [Zhang and Wu 2017a], consider the case with \( \Theta_{j,2\varphi,\gamma} = \mathcal{O}(K^{1/\gamma}) \) and \( \Phi_{j,2\varphi,\gamma} = \mathcal{O}(1) \), where \( \gamma > 1/2 - 1/q \). Then \( \Theta_{j,2\varphi,\gamma,n1/q-1/2}\}^{3/2}\) becomes \( \mathcal{O}(\log(nK))^{3q/2} = c(n^{q/2-1}) \), which implies that \( L_1 \max(W_1, W_2) = c(1) \min(N_1, N_2) = c(n^{\gamma}) \). This means with \( \gamma \) the dimension \( K \) has to satisfy the condition \( K \mathcal{O}(K^{3q/2} = c(n^{q/2-1}). \)

**Theorem 5.2** (Gaussian Approximation Results for Dependent Data). Under \( \{A1\} \) and \( \{A4\} \), for each \( j = 1, \ldots, J \) assume that there exists a constant \( c_j > 0 \) such that \( \min_{1 \leq k \leq K} \text{var}(S_{jk}) \geq \)}
\( c_j, \) then we have
\[
\rho(D_j^{-1}S_j, D_j^{-1}Z_j) \to 0, \quad \text{as } n \to \infty, \tag{5.10}
\]
where \( Z_j \sim N(0, \Sigma_j), \Sigma_j \) is the \( K \times K \) long-run variance-covariance matrix of \( X_{j,t} \varepsilon_{j,t}, \) and \( D_j \) is a diagonal matrix with the square root of the diagonal elements of \( \Sigma_j, \) namely
\[
\left\{ \sum_{\ell = -\infty}^{\infty} \mathbb{E}(X_{j,k,t}X_{j,(t-\ell)} \varepsilon_{j,t} \varepsilon_{j,(t-\ell)}) \right\}^{1/2} = \sqrt{\var(C_j)}, \quad \text{for } k = 1, \ldots, K.
\]

**Corollary 5.4.** Let the conditions of Theorem 5.2 hold, and assume we use the estimate in (5.12). Moreover, the following corollary ensures that the Gaussian approximation results still hold if the long-run variance-covariance matrix of \( Z_j \) is a diagonal matrix with the square root of the diagonal elements of \( \Sigma_j, \) namely
\[
\left\{ \sum_{\ell = -\infty}^{\infty} \mathbb{E}(X_{j,k,t}X_{j,(t-\ell)} \varepsilon_{j,t} \varepsilon_{j,(t-\ell)}) \right\}^{1/2} = \sqrt{\var(C_j)}, \quad \text{for } k = 1, \ldots, K.
\]

**Comment 5.7.** The conclusion in Theorem 5.2 can be held with stronger tail assumptions, following Theorem 5.2 in Zhang and Wu (2017a).

Theorem 5.2 justifies the choice of \( \lambda_j \) and \( \hat{Q}_j(1 - \alpha), \) which leads to the following corollary:

**Corollary 5.2.** Under the conditions of Theorem 5.2, for each \( j \) we have
\[
\sup_{\alpha \in (0,1)} \mathbb{P}\{ \max_{1 \leq k \leq K} 2c\sqrt{n}|S_{j,k}|/\Psi_{j,k} \geq Q_j(1 - \alpha)\} - \alpha \to 0, \quad \text{as } n \to \infty. \tag{5.11}
\]

It is worth noting that in practice the variance involved in the Gaussian approximation in (5.2) is not known; we shall discuss how we estimate the variance and also the validity of the Gaussian approximation result with an estimated variance. Given the realization \( X_{j,1}\varepsilon_{j,1}, \ldots, X_{j,n}\varepsilon_{j,n}, \) we propose to estimate the \( K \times K \) long-run variance-covariance matrix \( \Sigma_j \) for \( j = 1, \ldots, J \) as follows, given \( \mathbb{E}(X_{j,t}\varepsilon_{j,t}) = 0, \) and consider:
\[
\hat{\Sigma}_j = \frac{1}{b_n} \sum_{i=1}^{b_n} \left( \sum_{l=(i-1)b_n+1}^{ib_n} X_{j,l}\varepsilon_{j,l} \right) \left( \sum_{l=(i-1)b_n+1}^{ib_n} X_{j,l}\varepsilon_{j,l} \right)' . \tag{5.12}
\]

Moreover, the following corollary ensures that the Gaussian approximation results still hold if we use the estimate in (5.12).

**Corollary 5.3.** Let the conditions of Theorem 5.2 hold, and assume \( \Phi_{j,2q,\varsigma} < \infty \) with \( q > 4, \)
\( b_n = O(n^\eta) \) for some \( 0 < \eta < 1. \) Let \( F_\varsigma = n, \) for \( \varsigma > 1 - 2/q; \) \( F_\varsigma = \eta_n^q/2-q/2, \) for \( 1/2 - 2/q < \varsigma < 1 - 2/q; \) \( F_\varsigma = \eta_n^q/2-q/2, \) for \( \varsigma < 1/2 - 2/q. \) Further assume
\[
n^{-1} \log^2 K \max \left\{ n^{1/2}b_n^{1/2} \Phi_{j,2q,\varsigma}^2, n^{1/2}b_n^{1/2} \sqrt{\log K} \Phi_{j,8,\varsigma}^2, F_{2q,\varsigma}^2 / F_{2q,\varsigma}^2, K^{2/4}, \Phi_{j,2q,\varsigma} v'(b_n)n/\sqrt{\log K} \right\} = O(1), \quad \text{with } v'(b_n) = (b_n + 1)^{1-\varsigma} + 2v_{n,2}/b_n, \quad v_{n,2} = \log b_n \ (\text{resp. } b_n^{-\varsigma+1} \text{ or } 1) \quad \text{for } \varsigma = 1 \ (\text{resp. } \varsigma < 1 \text{ or } \varsigma > 1). \quad \text{Then for each } j \text{ we have}
\]
\[
\rho(D_j^{-1}S_j, D_j^{-1}Z_j) \to 0, \quad \text{as } n \to \infty, \tag{5.13}
\]
where \( D_j = \{\text{diag}(\hat{\Sigma}_j)\}^{1/2}. \)

It should be noted that given the Gaussian approximation results in Theorem 5.2, we can have a refined bound for \( \lambda_j^0(1 - \alpha) \) and also the oracle inequalities under IC.

**Corollary 5.4 (Bounds for \( \lambda_j^0(1 - \alpha) \) and Oracle Inequalities under IC with Gaussian Approximation Results).** Under the conditions of Theorem 5.2 together with (A2), let \( 2(\log K)^{-1/2} + \rho(D_j^{-1}S_j, D_j^{-1}Z_j) = o(\alpha) \) and \( Z_0 \geq 2\hat{c}\sqrt{n}\log K, \) where \( \hat{c} \) is no less than the \( c \) in the definition
of \( \lambda^0_j(1-\alpha) \), then we have \( \lambda^0_j(1-\alpha) \leq Z_\alpha \),

\[
\lambda^0_j(1-\alpha) \leq Z_\alpha, \tag{5.14}
\]

and given the exact sparsity assumption \([3.1]\), then \( \tilde{\beta}_j \) obtained from \([3.6]\) under IC satisfies

\[
|\tilde{\beta}_j - \beta^0_j|_{j,pr} \leq \sqrt{\frac{\beta_j}{\kappa_j(\epsilon)}} \max_{1 \leq k \leq K} \Psi_{jk} \sqrt{\log K \frac{\sqrt{n}}{n}}, \tag{5.15}
\]

with probability \( 1 - \alpha - \alpha(1) \).

5.3 Multiplier Block Bootstrap Procedure

In this subsection, we discuss how \( \Lambda_j(1-\alpha) \) is attainable via block bootstrap. The data over \( t = 1, \ldots, n \) are divided into \( \ell_n \) blocks with the same number of observations \( b_n, n = b_n \ell_n \) (without loss of generality), where \( b_n, \ell_n \in \mathbb{Z} \).

Recall that \( \Lambda_j(1-\alpha) = 2c \sqrt{n} q_{j,(1-\alpha)}^{[B]} \), where \( q_{j,(1-\alpha)}^{[B]} \) is the \((1-\alpha)\) quantile of \( \max_{1 \leq k \leq K} |Z_{jk}^{[B]} / \Psi_{jk}| \), where \( Z_{jk}^{[B]} \) are defined as

\[
Z_{jk}^{[B]} = \frac{1}{\sqrt{n}} \sum_{i=1}^{\ell_n} e_{j,i} \sum_{l=(i-1)b_n+1}^{\ell_n} \varepsilon_{j,l} X_{j,k,l}, \tag{5.16}
\]

and \( e_{j,i} \) are i.i.d. \( N(0,1) \) random variables independent of \( X \) and \( \varepsilon \).

In fact, the above construction relies on knowing the true residuals \( \varepsilon_{j,i} \). In practice, one needs to pre-estimate them using a conservative choice of penalty levels and loadings. The issue of generated errors can be dealt with using a similar argument as in the proof of Corollary 5.3.

Theorem 5.3 (Validity of Multiplier Block Bootstrap Method). Under \([A1]\) and \([A3]\), and assume \( \Phi_{j,2q,s} < \infty \) with \( q > 4, b_n = O(n^\eta) \) for some \( 0 < \eta < 1 \) (the detailed rate is calculated in \([B.2]\) in the supplementary materials), then we have

\[
\sup_{\alpha \in (0,1)} \left| \mathbb{P} \left( \max_{1 \leq k \leq K} |S_{jk} / \Psi_{jk}| \geq q_{j,(1-\alpha)}^{[B]} \right) - \alpha \right| \to 0, \text{ as } n \to \infty. \tag{5.17}
\]

5.4 Joint Penalty over Equations

Recall that the theoretical choice \( \lambda^0_j(1-\alpha) \) is defined as the \((1-\alpha)\) quantile of \( \max_{1 \leq k \leq K, 1 \leq j \leq J} 2c \sqrt{n} |S_{jk} / \Psi_{jk}| \).

The empirical choices of the joint penalty level can be:

a) \( Q(1-\alpha) \): the \((1-\alpha)\) quantile of \( 2c \ max_{1 \leq k \leq K, 1 \leq j \leq J} \sqrt{n} |Z_{jk} / \Psi_{jk}| \). In practice, one can take an alternative choice such that \( \tilde{Q}(1-\alpha) \) \( \equiv 2c \sqrt{n} \Phi^{-1} \{ 1 - \alpha / (2KJ) \} \).

b) \( \Lambda(1-\alpha) \) \( \equiv 2c \sqrt{n} q_{j,(1-\alpha)}^{[B]} \), where \( q_{j,(1-\alpha)}^{[B]} \) is the \((1-\alpha)\) quantile of \( \max_{1 \leq k \leq K, 1 \leq j \leq J} |Z_{jk}^{[B]} / \Psi_{jk}| \).

Section \([A]\) in the supplementary material provides the main theorems for joint equation estimation. In particular, the dimension along \( k = 1, \ldots, K \) and \( j = 1, \ldots, J \) will be considered
together by vectorization, resulting in the dimension of $KJ$. Following the results for the single equation (where $j$ is fixed), we generalize the theorems above to multiple equations case by changing the dimension from $K$ to $KJ$, see Section A in the Appendix for more details.

5.5 Post-Model Selection Estimation

LASSO estimation is known to be biased especially for large coefficients. Therefore, a post-selection step helps to reduce the bias by running an OLS as a second step on the selected covariates in the first step. In particular, we consider the 2-step OLS post-LASSO estimator:

i) $\ell_1$-penalized regression (LASSO selection)

$$\tilde{\beta}_j = \arg \min_{\beta \in \mathbb{R}^K} \frac{1}{n} \sum_{t=1}^n (Y_{j,t} - X_{j,t}^\top \beta)^2 + \frac{\lambda}{n} \sum_{k=1}^K |\beta_{jk}| \Psi_{jk},$$  \hspace{1cm} (5.18)

where $\lambda$ is the joint penalty level.

ii) We run the post-selection regression (OLS estimation)

$$\hat{\beta}^{[P]}_j = \arg \min_{\beta \in \mathbb{R}^K} \left\{ \frac{1}{n} \sum_{t=1}^n (Y_{j,t} - X_{j,t}^\top \beta)^2 : \beta_k = 0, k \notin \hat{T}_j \right\},$$  \hspace{1cm} (5.19)

where $\hat{T}_j \overset{\text{def}}{=} \text{supp}(\tilde{\beta}_j) = \{k \in \{1, \ldots, K\} : \tilde{\beta}_{jk} \neq 0\}$.

To provide the prediction performance bounds for the OLS post-LASSO estimators, we need the following restricted sparse eigenvalue (RSE) condition:

(A5) Restricted sparse eigenvalue (RSE): given $p < n$, for $\delta \in \mathbb{R}^K$, with probability $1 - o(1)$,

$$\bar{\kappa}_j(p) = \min_{|\delta_{T_j^c}| = p} \frac{|\delta_{j,pr}|^2}{|\delta_{T_j^c}|^2} > 0, \hspace{1cm} \phi_j(p) = \max_{|\delta_{T_j}| = p, \delta \neq 0} \frac{|\delta_{j,pr}|^2}{|\delta_{T_j}|^2} > 0.$$

Here $p$ denotes the restriction on the length of the active set of $T_j^c$. When $T_j = \emptyset$, (A5) is reduced to the standard sparse eigenvalue condition. Moreover, let $\mu_j(p) = \frac{\sqrt{\phi_j(p)}}{\sqrt{\phi_j(p)}}$, and denote by $\hat{\rho}_j = |\hat{T}_j \setminus T_j|$ the number of components outside $T_j = \text{supp}(\beta_0^j) = \{k \in \{1, \ldots, K\} : \beta_{0jk} \neq 0\}$ selected by LASSO in the first step.

The performance bounds for the OLS post-LASSO estimator are shown in Theorem A.4 in the supplementary materials.

5.6 Simultaneous Inference

This subsection develops theory corresponding to Section 4. A key Bahadur representation which linearize the estimator for a proper application of the central limit theorem for inference is provided.
Recall that for each \( j = 1, \ldots, J \), the following model is considered

\[
Y_{j,t} = \sum_{k=1}^{p_j} X_{jk,t}^0 \beta_{jk}^0 + \sum_{k=p_j+1}^{K} X_{jk,t}^0 \beta_{jk}^0 + \varepsilon_{j,t}, \quad \mathbb{E}(\varepsilon_{j,t}X_{j,t}) = 0, \quad F_{\varepsilon_j}(0) = 1/2,
\]

\[
X_{jk,t} = X_{j(k),t}^0 + v_{jk,t}, \quad \mathbb{E}(v_{jk,t}X_{j(k),t}) = 0, \quad k = 1, \ldots, p_j,
\]

where we define \( \gamma_{j(k)}^0 = \arg \min_{\gamma_{j(k)}} \mathbb{E}(X_{j(k),t} - X_{j(k),t}^0 \gamma_{j(k)})^2 \), and let \( F_{\varepsilon_j} \) denote the distribution function of \( \varepsilon_{j,t} \). In this subsection, we show the validity of the joint confidence region for simultaneous inference on \( H \beta_{jk}^0 \). Moreover, we assume that the nuisance function \( \gamma_{j(k)}^0 \) admits a sparse estimator \( \hat{\gamma}_{jk} \) defined by

\[
\hat{\gamma}_{jk,m}(X_{j(k),t}) = X_{j(k),t}^\top \hat{\beta}_{jk,m}, \quad \| \hat{\beta}_{jk,m} \|_0 \leq s, \quad m = 1, \ldots, M,
\]

where the sparsity level \( s \) is small compared to \( n (s \ll n) \).

The true parameter \( \beta_{jk}^0 \) is identified as a unique solution to the moment condition

\[
\mathbb{E}[\psi_{jk}\{Z_{j,t}, \beta_{jk}, h_{jk}(X_{j(k),t})\}] = 0.
\]

However, the object \( \arg \min_{\beta_{jk} \in \hat{B}_{jk}} \mathbb{E}_n[|\psi_{jk}\{Z_{j,t}, \beta_{jk}^0, h_{jk}^0(X_{j(k),t})\}] \) does not necessarily exist due to the discontinuity of the function \( \psi_{jk} \). The estimator \( \hat{\beta}_{jk} \) is obtained as a Z-estimator by solving the sample analogue of (5.21)

\[
\mathbb{E}_n[\psi_{jk}\{Z_{j,t}, \hat{\beta}_{jk}, \hat{h}_{jk}(X_{j(k),t})\}] \leq \inf_{\beta_{jk} \in \hat{B}_{jk}} \mathbb{E}_n[|\psi_{jk}\{Z_{j,t}, \beta_{jk}, \hat{h}_{jk}(X_{j(k),t})\}] + c(n^{-1/2}g_n^{-1}),
\]

where \( g_n \) is defined in (C2) and \( \hat{B}_{jk} \) is defined in (C2).

We now lay out the following conditions needed in this section, which are assumed to hold uniformly over \( (j,k) \in G \).

**(C1)** Orthogonality condition:

\[
\mathbb{E} \left[ \partial_h \mathbb{E}(\psi_{jk}(Z_{j,t}, \beta_{jk}^0, h)X_{j(k),t}) \bigg| h = h_{jk}^0(X_{j(k),t}) \right] = 0,
\]

for any \( h \in \mathcal{H}_{jk} \cup \{h_{jk}^0\} \), where \( \mathcal{H}_{jk} \) is defined in (C5).
(C2) The true parameter $\beta_{jk}^0$ satisfies \[\{5.21\}. Let $B_{jk}$ be a fixed and closed interval and $B_{jk}$ be a possibly stochastic interval such that with probability $1 - o(1)$, \[|\beta_{jk}^0| \geq 1\] \(\rho_n \) is defined in \[\{5.5\}\], \(a_n \) \(\max(JK,n,e)\) and \(\psi_{jk,t}^0 = \psi_{jk}\{Z_{jt},\vartheta_{jk},h_{jk}^0(X_{j(k),t})\} \). \[r_x = n^{1/4} \text{ for } \varsigma > 1/2 - 1/q \text{ and } r_x = n^{1/2-\varsigma} \text{ for } \varsigma < 1/2 - 1/q.\]

(C3) Properties of the score function: the map \((\beta, h) \mapsto E\{\psi_{jk}(Z_{jt}, \beta, h)|X_{j(k),t}\}\) is twice continuously differentiable, and for every \(h \in \{\beta, h_1, \ldots, h_M\}\), \[
E[\sup_{(\beta, h) \in B_{jk}} |\partial_\beta E\{\psi_{jk}(Z_{jt}, \beta, h)|X_{j(k),t}\}|^2] \leq C_1; \text{ moreover, there exist measurable functions } \
\ell_1(\cdot), \ell_2(\cdot), \text{ constants } L_{1n}, L_{2n} \geq 1, \nu > 0 \text{ and a cube } T_{jk}(X_{j(k),t}) = \
\times_{m=1}^M T_{jk,m}(X_{j(k),t}) \text{ in } \mathbb{R}^M \text{ with center } h_{jk}^0(X_{j(k),t}) \text{ such that for every } \vartheta, \vartheta' \in \{\beta, h_1, \ldots, h_M\}, we have sup_{(\beta, h) \in B_{jk} \times T_{jk}(X_{j(k),t})} |\partial_\vartheta \partial_\vartheta' E\{\psi_{jk}(Z_{jt}, \beta, h)|X_{j(k),t}\}| \leq \ell_1(X_{j(k),t}), \\
E\{|\ell_1(X_{j(k),t})|^4\} \leq L_{1n}, \text{ and for every } \beta, \beta' \in B_{jk}, h, h' \in T_{jk}(X_{j(k),t}) \text{ we have } E\{|\psi_{jk}(Z_{jt}, \beta, h) - \\
\psi_{jk}(Z_{jt}, \beta', h')|^2|X_{j(k),t}\} \leq \ell_2(X_{j(k),t}) \text{ for all } X_{j(k),t} \in H_{jk}\}. \]

(C4) Identifiability: \(2|E\{\psi_{jk}(Z_{jt}, \beta, h_0^0(X_{j(k),t}))\}| \geq |\phi_{jk}(\beta - \beta_0)| \wedge c_1 \) for all \(\beta \in B_{jk}\), where \(\phi_{jk} = \partial_\beta E\{\psi_{jk}(Z_{jt}, \beta_{jk}, h_{jk}^0(X_{j(k),t}))\}\) and \(|\phi_{jk}| \geq c_1\). \[\text{Prop. of the nuisance function: with probability } 1 - o(1), \hat{h}_{jk} \in \mathcal{H}_{jk}\), where \(\mathcal{H}_{jk} = \
\times_{m=1}^M H_{jk,m} \text{ being the class of functions of the form } \hat{h}_{jk,m}(X_{j(k),t}) = \
X_{j(k),t}^{T} \theta_{jk,m}, |\theta_{jk,m}| \leq s, \hat{h}_{jk,m} \in T_{jk,m}. \text{ There exists sequence of constants } \rho_n \downarrow 0 \text{ such that } E\{\|\hat{h}_{jk,m}(X_{j(k),t}) - h_{0j}^0(X_{j(k),t})\|^2\} \leq \rho_n^2. \]

(C6) The class of functions \(F_{jk} = \{z \mapsto \psi_{jk}\{z, \beta, h(x_{j(k)})\} : \beta \in B_{jk}, \hat{h} \in \mathcal{H}_{jk} \cup \{h_{0j}^0\}\} \). \[\text{is a random vector taking values in a Borel subset of a Euclidean space which contains the vectors } x_{j(k)} \text{ as subvectors is pointwise measurable and has measurable envelope } \]
\[F_{jk} \sup\limits_{f \in F_{jk}} |f|, \text{ such that } F = \max \limits_{(j,k) \in G} F_{jk} \text{ satisfies } E\{F^q(z)\} < \infty \text{ for some } q \geq 4. \]

(C7) The second-order moments of scores are bounded away from zero: \(\omega_{jk} = E\{(1/n) \sum_{l=1}^n \psi_{jk,\ell}\}^2 \geq c_1.\)

(C8) Dimension growth rates: \(\rho_n = (L_{2n}s \log a_n)^{1/2} + n^{-1/2} r_x(s \log a_n)^{3/2} |F(z)|_q + s^{2} = o(g^{-1}). \) In particular, for the mean regression case \(\rho_n = \rho_n s \) and \(\rho_n = \rho_n^1/2 = \rho_n^1/2 = o(n^{-1}). \) For the median regression case, \(n^{-1/2} (s \log a_n)^{1/2} \max_{f \in F} |f(z)|_2 + n^{-1/2} r_x(s \log a_n)^{3/2} |F(z)|_q = O(p_n). \) \[\text{Prop. of the score function: the map } \]
\[(\beta, h) \mapsto E\{\psi_{jk}(Z_{jt}, \beta, h)|X_{j(k),t}\}\) is twice continuously differentiable, and for every \(h \in \{\beta, h_1, \ldots, h_M\}\), \[
E[\sup_{(\beta, h) \in B_{jk}} |\partial_\beta E\{\psi_{jk}(Z_{jt}, \beta, h)|X_{j(k),t}\}|^2] \leq C_1; \text{ moreover, there exist measurable functions } \
\ell_1(\cdot), \ell_2(\cdot), \text{ constants } L_{1n}, L_{2n} \geq 1, \nu > 0 \text{ and a cube } T_{jk}(X_{j(k),t}) = \
\times_{m=1}^M T_{jk,m}(X_{j(k),t}) \text{ in } \mathbb{R}^M \text{ with center } h_{jk}^0(X_{j(k),t}) \text{ such that for every } \vartheta, \vartheta' \in \{\beta, h_1, \ldots, h_M\}, we have sup_{(\beta, h) \in B_{jk} \times T_{jk}(X_{j(k),t})} |\partial_\vartheta \partial_\vartheta' E\{\psi_{jk}(Z_{jt}, \beta, h)|X_{j(k),t}\}| \leq \ell_1(X_{j(k),t}), \\
E\{|\ell_1(X_{j(k),t})|^4\} \leq L_{1n}, \text{ and for every } \beta, \beta' \in B_{jk}, h, h' \in T_{jk}(X_{j(k),t}) \text{ we have } E\{|\psi_{jk}(Z_{jt}, \beta, h) - \\
\psi_{jk}(Z_{jt}, \beta', h')|^2|X_{j(k),t}\} \leq \ell_2(X_{j(k),t}) \text{ for all } X_{j(k),t} \in H_{jk}\}. \]

(C9) Let \(B_{\Phi} = \max_{m \in (1,2)} \Phi_{m,2,\varsigma}, B_{\Omega} = \max_{m \in (1,2)} \Omega_{m,\varsigma}\) and \(B_{\Phi} = \max_{m \in (1,2)} \Phi_{m,2,\varsigma}, B_{\Omega} = \max_{m \in (1,2)} \Omega_{m,\varsigma}\) (see \[\{8.9\}, \{8.10\} \text{ and } \{8.15\} \text{ in the supplementary for the definitions of } \Phi_{m,\varsigma}, \Omega_{m,\varsigma}, \Phi_{2,\varsigma}, \Omega_{2,\varsigma}, \Phi_{3,\varsigma}, \Omega_{3,\varsigma}\). The following restrictions are assumed:

\[s \rho_n (s \log a_n)^{1/2} B_{\Phi} + n^{-1/2} r_x \rho_n s^2 (s \log a_n)^{3/2} B_{\Omega} = o(g^{-1}), \]

\[\rho_n (s \log a_n)^{1/2} \Phi_{2,\varsigma} + n^{-1/2} r_x \rho_n (s \log a_n)^{3/2} \Omega_{2,\varsigma} = o(g^{-1}), \]

\[\rho_n \]
The density of error that our proposed algorithms can produce the estimator of the form \( L \) model. Following the discussion in Comment 5.3, given a geometric decay rate, we have Comment 5.8.

moments of the envelopes are assumed to be finite in (C6). (C9) Consider the stronger exponential moment condition as in (5.6) and corresponding to (C5) assume that \( E \{|\tilde{h}_{j,k,m}(X_{j(-k),t}) - h_{j,k,m}(X_{j(-k),t})|^2 \} \lesssim (\rho_n^0)^2 \). Recall the definitions of \( \Phi_{m,\psi,0}^h, \Phi_{\psi,0}^\beta, \Phi_{m,\psi,0}^h, \Phi_{\psi,0}^\beta \) in (B.17) and (B.20) in the supplementary. The following restrictions are assumed:

\[
B_{\Phi,\rho_n}s^{1/2} = O(\max_{f \in F} \|f(z_t)\|_2), \quad B_{\Omega,\rho_n}s^{1/2} = O(\|F'(z_t)\|_q),
\]

\[
\Phi_{\rho_n}^{\beta} = O(\max_{f \in F} \|f(z_t)\|_2), \quad \Omega_{\rho_n}^{\beta} = O(\|F'(z_t)\|_q).
\]

(C10) The density of error \( f_{x_j}(\cdot) \) is continuously differentiable and both of \( f_{x_j}(\cdot) \) and \( f'_{x_j}(\cdot) \) are bounded from the above.

Conditions (C1), (C4), and (C7) assume mild restrictions on the Z-estimation problems. They include the LAD-based regression (used in Algorithm 2) with nonsmooth score function. Conditions (C2) and (C8) imply that \( \max_{(j,k) \in G} \|\psi_{j,k,1}\|_2 \lesssim s^{1/2} \max_{f \in F} \|f(z_t)\|_2 \) and \( \max_{(j,k) \in G} \|\psi_{j,k,1}\|_q \lesssim s^{3/2} \|F'(z_t)\|_q \). In (C5) we suppose that the nuisance parameters have estimators with good sparsity and convergence rate properties. As discussed in previous sections, given the ideal choice of the tuning parameter, the oracle inequalities provided in Corollary 5.1 ensures that our proposed algorithms can produce the estimator of the form \( \|\hat{\beta}_{j(-k)} - \beta_{j(-k)}\|_{J,pr} \lesssim \{s^{1/2}\log(n)\}/n \vee n^{1/q-1}(\omega_n\alpha_n)^{1/q} \max_{1 \leq k < K} \|X_{j,k}\|_{q,\xi} \). Moreover, suppose these quantities are bounded by constant and let \( d_n \triangleq \{1 \vee s\} \). We have \( B_{\hat{\Omega},\rho_n} \lesssim d_n^{3/(4q)}(1 \vee s^{1/2} \rho_n) \), \( \Omega_{\rho_n}^{\beta} \lesssim d_n^{3/(4q)}s^{1/2} \rho_n \) for mean regression case, and \( B_{\hat{\Omega},\rho_n} \lesssim d_n^{3/(4q)}(1 \vee s^{1/2} \rho_n) \), \( \Omega_{\rho_n}^{\beta} \lesssim d_n^{3/(4q)}s^{1/2} \rho_n \) for the median regression. Moreover, \( \|F(z_t)\|_q \), \( \|F'(z_t)\|_q \lesssim d_n^{1/4}(1 \vee \rho_n) \), \( \max_{(j,k) \in G} \|\psi_{j,k,1}\|_q \lesssim d_n^{1/4}(1 \vee \rho_n) \). The detailed derivation of these rates can be found in the Comment B.3 in the supplementary. Inserting

\[
\rho_n = (\log(n)/n)\vee n^{1/q-1}(\omega_n\alpha_n)^{1/q} \max_{1 \leq k < K} \|X_{j,k}\|_{q,\xi}.
\]

\[
\rho_n^0 = (\log(n)/n)\vee n^{1/q-1}(\omega_n\alpha_n)^{1/q} \max_{1 \leq k < K} \|X_{j,k}\|_{q,\xi}.
\]
them into (C8) and (C9) yields
\[
n^{-1/2} s^2 \left( \log a_n \right)^{3/2} + n^{-1} r \sigma^3 \left( \log a_n \right)^{5/2} d^{1/q}_n + n^{-1/2} r \sigma^3 \left( \log a_n \right)^{3/2} d^{3/(4q)}_n = o(1),
\]
and
\[
n^{-1/4} s^{3/4} \left( \log a_n \right)^{5/4} + n^{-1/2} r \sigma^3 \left( \log a_n \right)^{5/4} d^{3/(4q)}_n + n^{-1/2} r \sigma^3 \left( \log a_n \right)^{5/4} d^{3/(4q)}_n = o(1),
\]
for the smooth and non-smooth cases respectively. As a result, we only allow the dimension \((|G| \lor J)\) is of polynomial order with respect to \(n\) if \(q\) is not tending to infinity. In particular, under the case of \(\zeta > 1/2\) and \(q = \infty\), the required rate reduces to \(n^{-1/2} s^2 \left( \log a_n \right)^{3/2} + n^{-1} s^3 \left( \log a_n \right)^{5/2} + n^{-1/2} s^3 \left( \log a_n \right)^{3/2} = o(1)\) or \(n^{-1/4} s^{3/4} \left( \log a_n \right)^{5/4} + n^{-1/2} s^3 \left( \log a_n \right)^{5/4} d^{3/(4q)}_n = o(1)\), respectively. In the ideal case where we have weak dependency, the dimension growth rates are slightly slower than the i.i.d. case as in Belloni et al. (2015b) (i.e., \(s^2 \log a^3 = o(n)\) or \(s^3 \log a^5 = o(n)\) for the smooth or non-smooth case, respectively), as we apply a different way to bound the dependence adjusted norm in the concentration inequality.

More generally, suppose \(\max \{L_{2n}, B_{t}^h, B_{t}'^h, \Phi_{2c}, \Phi_{2c}' \}, \max_{f \in F} \|f(z_t)\|_2, \max_{(j,k) \in G} \|v^0_{jk}\|_{2,2} = O(n^k)\), and \(\max \{B_{t1}^h, B_{t1}'^h, \Omega_{q,<}^0, \Omega_{q,<}^0 \}, \|F(z_t)\|_q, \|F'(z_t)\|_q, \max_{(j,k) \in G} \|v^0_{jk}\|_{q,<} = O(n^k)\), with \(0 \leq k_1 \leq k_2\), and let \(s = O(n^n)\), log \(a_n = O(n^r)\). Then (C8) and (C9) imply that
\[
r < \max \left\{ \frac{1 - 4v - 2k_1}{3}, -\frac{2}{5q} + \frac{2 - 6v - 2k_2}{5}, -\frac{1}{2q} + \frac{1 - 3v - 2k_2}{4} \right\}, \text{ if } \zeta > 1/2 - 1/q,
\]
\[
r < \max \left\{ \frac{1 - 4v - 2k_1}{3}, \frac{2\zeta + 1 - 6v - 2k_2}{5}, \frac{2\zeta - 3v - 2k_2}{4} \right\}, \text{ if } \zeta < 1/2 - 1/q,
\]

and
\[
r < \max \left\{ \frac{1 - 4v - 2k_1}{3}, \frac{2}{7q} + \frac{2 - 5v - 2k_2}{7}, \frac{1 - 3v - 2k_2}{4} \right\}, \text{ if } \zeta > 1/2 - 1/q,
\]
\[
r < \max \left\{ \frac{1 - 4v - 2k_1}{3}, \frac{2\zeta + 1 - 5v - 2k_2}{7}, \frac{2\zeta - 3v - 2k_2}{4} \right\}, \text{ if } \zeta < 1/2 - 1/q,
\]
for the smooth and non-smooth cases.

**Theorem 5.4.** [Uniform Bahadur Representation] Under conditions (A1), (A4) and (C10) with probability \(1 - o(1)\), we have
\[
\max_{(j,k) \in G} \left| n^{1/2} \sigma^{-1} \beta_{jk}^0 + n^{-1/2} \sigma^{-1} \phi_{jk} \sum_{t=1}^{n} \psi_{jkt} \right| = o(g_n^{-1}), \text{ as } n \to \infty, \tag{5.23}
\]
where \(\sigma_{jk}^2 \coloneqq \sigma_{jk}^2 \omega_{jk}, \omega_{jk} \coloneqq E\left( \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \psi_{jkt}^0 \right)^2 \).

**Comment 5.9.** The same conclusion as in Theorem 5.4 can be drawn with assuming stronger exponential moment conditions in (5.9) and using (C9') instead of (C6), (C8) and (C9). This is implied by Lemma B.8, B.9 and B.10 in the supplementary material.

We now discuss the rates implication under (C9'). Suppose all the dependence adjusted norms are bounded by constant with an appropriately chosen \(\nu\), the restrictions in (C9') would
imply \( n^{-1/2}(\log a_n)^{2/\gamma+1/2}s^{2/\gamma+1} = o(1) \) for the case of smooth score, and
\( n^{-1/4}(\log a_n)^{3/(2\gamma)s^{3/(2\gamma)+1/2} = o(1) \) for the non-smooth case, where \( \gamma = 2/(2\nu + 1) \). For example, when \( \nu = 1/2, \gamma = 1 \) the required rates would be \( s^6\log^5 a_n = o(n) \) and \( s^6\log^8 a_n = o(n) \) for the smooth and non-smooth cases respectively.

The results in Theorem 5.4 imply the asymptotic normality of the proposed estimator by Algorithm 1 and 2 by applying central limit theorems and Gaussian Approximation.

**Corollary 5.5.** Under conditions [A1, A4] and [C10], for any \((j, k) \in G\) the estimators obtained by Algorithm 1 and 2 satisfy
\[
\sigma_{jk}^{-1}n^{1/2}(\hat{\beta}_{jk}^{[2]} - \beta_{jk}^0) \overset{d}{\to} N(0, 1).
\]

**Theorem 5.5.** [Uniform-Dimensional Central Limit Theorem] Under the same conditions as in Theorem 5.4, assume that \( \|\psi_{jk}(\cdot)\|_{2,\varsigma} < \infty \), we have
\[
\sigma_{jk}^{-1}n^{1/2}(\hat{\beta}_{jk} - \beta_{jk}^0) \overset{d}{\to} N(0, 1),
\]
uniformly over \((j, k) \in G\).

Consider the vector \( \tilde{\zeta} \overset{\text{def}}{=} \text{vec}\{\zeta_{jk,t}(j, k) \in G\} \), \( \zeta_{jk,t} \overset{\text{def}}{=} -\sigma_{jk}^{-1}\phi_{jk}^{-1}\psi_{jk,t}^0 \), and define the aggregated dependence adjusted norm as follows:
\[
\|\tilde{\zeta}\|_{q,\varsigma} \overset{\text{def}}{=} \sup_{m \geq 0} (m + 1)^c \sum_{t=m}^{\infty} \|\zeta_t - \tilde{\zeta}_t\|_{q,\varsigma}, \quad (5.24)
\]
where \( q \geq 1 \), and \( c > 0 \). Moreover, define the following quantities
\[
\Phi_{q,\varsigma}^c \overset{\text{def}}{=} \max_{(j,k) \in G} \|\zeta_{jk,\cdot}\|_{q,\varsigma}, \quad \Gamma_{q,\varsigma}^c \overset{\text{def}}{=} \left( \sum_{(j,k) \in G} \|\zeta_{jk,\cdot}\|_{q,\varsigma}^q \right)^{1/q},
\]
\[
\Theta_{q,\varsigma}^c \overset{\text{def}}{=} \Gamma_{q,\varsigma}^c \wedge \{\|\zeta\|_{q,\varsigma}(\log |G|)^{3/2}\}. \quad (5.25)
\]

Define \( L^c_1 = \{\Phi_{q,\varsigma}^c, \Phi_{q,\varsigma}^c (\log |G|)^{3/2}\}^{1/\varsigma}, W^c_1 = \{(\Phi_{q,\varsigma}^c)^6 + (\Phi_{q,\varsigma}^c)^4\}(\log(|G|)^7, W^c_2 = (\Phi_{q,\varsigma}^c)^2(\log(|G|)^{1/2})^2, W^c_3 = [n^{-c}\{\log(|G|)^{1/2}\}^{-\varsigma - 1/2}, N^c_1 = (n/\log |G|)^{3/2}(\Theta_{q,\varsigma}^c)^q, N^c_2 = n(\log |G|)^{-2}(\Phi_{q,\varsigma}^c)^{-2}, N^c_3 = n^{1/2}(\log |G|)^{-1/2}(\Theta_{q,\varsigma}^c)^{1/2}, (A6)
\]
i (weak dependency case) Given \( \Theta_{q,\varsigma}^c < \infty \) with \( q \geq 2 \) and \( c > 1/2 - 1/q \), then \( \Theta_{q,\varsigma}^c n^{1/2-1/2}(\log(|G|)^{3/2} = o(1) \) and \( L^c_1 \max(W^c_1, W^c_2) = o(1) \min(N^c_1, N^c_2) \).

(ii) (strong dependency case) Given \( 0 < c < 1/2 - 1/q \), then \( \Theta_{q,\varsigma}^c (\log |G|)^{1/2} = o(n^c) \) and \( L^c_1 \max(W^c_1, W^c_2, W^c_3) = o(1) \min(N^c_2, N^c_3) \).

**Corollary 5.6** (Consistency of the Bootstrap Confidence Interval). Under [A6] and the same conditions as in Theorem 5.4, for each \((j, k) \in G\) assume that there exists a constant \( c > 0 \) such that \( \min_{(j,k) \in G} \text{var}(n^{-1/2} \sum_{t=1}^{n} \zeta_{jk,t}) \geq c \), with probability \( 1 - o(1) \), we have
\[
\sup_{a \in (0,1)} |P(\beta_{jk}^0 \in \tilde{C}I_{jk}(\alpha), \forall (j, k) \in G) - (1 - \alpha)| = o(1), \quad \text{as } n \to \infty, \quad (5.26)
\]
where $\tilde{C}_j^i(\alpha) \overset{\text{def}}{=} \left[ \hat{\beta}_j \pm \hat{\sigma}_{jk} n^{-1/2} q(1-\alpha) \right]$, and $q(1-\alpha)$ is the $(1-\alpha)$ quantile of the $\max_{(j,k) \in G} |Z_{jk}|$, where $Z_{jk}$'s are the standard normal random variables and $\hat{\sigma}_{jk}$ is a consistent estimator of $\sigma_{jk}$.

Following Theorem 5.4, a joint confidence region and the corresponding confidence interval for each component can be constructed via a block bootstrap method. In particular, the bootstrap statistic are defined by $\frac{1}{\sqrt{n}} \sum_{i=1}^{l_n} \epsilon_{j,i} \sum_{l=(i-1)b_n+1}^{ib_n} \hat{\gamma}_{jk,t}$, where $\epsilon_{j,i}$'s are independent and identically distributed draws of standard normal random variables and are independent with respect to the data sample $(Z_{j,t})_{t=1}^T$. Recall that $\hat{\gamma}_{jk,t}$ are pre-estimators with a certain range of accuracy.

Corollary 5.7 (Validity of Multiplier Bootstrap). Under the same conditions as in Theorem 5.4, assume $\Phi_{q,\xi} < \infty$ with $q > 4$, $b_n = O(n^\eta)$ for some $0 < \eta < 1$ (the detailed rate is specified in (B.27)), we have

$$\sup_{\alpha \in (0,1)} |P(\beta_j^0 \in \tilde{C}_j^i(\alpha), \forall (j,k) \in G) - (1-\alpha)| = o(1), \text{ as } n \to \infty,$$

where $\tilde{C}_j^i(\alpha) \overset{\text{def}}{=} \left[ \hat{\beta}_j \pm \hat{\sigma}_{jk} n^{-1/2} q^*(1-\alpha) \right]$, and $q^*(1-\alpha)$ is the $(1-\alpha)$ conditional quantile of $\max_{(j,k) \in G} \frac{1}{\sqrt{n}} |\sum_{i=1}^{l_n} \epsilon_{j,i} \sum_{l=(i-1)b_n+1}^{ib_n} \hat{\gamma}_{jk,t}|$.

6 Simulation Study

In this section, we illustrate the performance of our proposed methodology under different simulation scenarios. The first part concerns the performance of the jointly selected penalty level over equations, and the second part discusses the simultaneous inference.

6.1 Estimation with a Jointly Selected Penalty Level

Consider the system of regression equations:

$$Y_{j,t} = X_t^\top \beta_j^0 + \epsilon_{j,t}, \quad t = 1, \ldots, n, j = 1, \ldots, J,$$

where $X_t \in \mathbb{R}^K$. We generate $X_t$ independently from $N(0, \Sigma)$, where $\Sigma_{k_1,k_2} = \gamma |k_1-k_2|$, $\gamma = 0.5$, $\epsilon_{j,t} \overset{i.i.d.}{\sim} N(0,1)$. The coefficient vectors $\beta_j$ are assumed to be sparse. In particular, we divide the indices $\{1, \ldots, K\}$ evenly into blocks with fixed block size 5. $\beta_{j,k}^0 = 10$ if $k$ and $j$ belong to the same block and 0 otherwise.

We take $n = 100$, $\#$ of bootstrap replications = 5000. We set $J, K = 50, 100$ and 150. The prediction norm $|\hat{\beta}_j - \beta_j^0|_{1,pr}$ and the Euclidean norm $|\hat{\beta}_j - \beta_j^0|_2$ ratios are presented in Table 6.1. The ratios measure the relative difference between the results using the penalty level determined from the equation-by-equation case and from the joint equation case ($\lambda_j$ and $\lambda$ are selected by the multiplier bootstrap procedure). In particular, a ratio smaller than 1 indicates a better performance of using the jointly selected penalty level.
\[
\begin{array}{cccc}
J = K = 50 & J = K = 100 & J = K = 150 \\
\hline
\text{Prediction norm} & & & \\
\text{Mean} & 0.9634 & 0.9474 & 0.9347 \\
\text{Median} & 0.9695 & 0.9516 & 0.9371 \\
\text{Std.} & 0.0323 & 0.0272 & 0.0254 \\
\end{array}
\]
\[
\begin{array}{cccc}
\text{Euclidean norm} & & & \\
\text{Mean} & 0.9590 & 0.9429 & 0.9286 \\
\text{Median} & 0.9679 & 0.9468 & 0.9316 \\
\text{Std.} & 0.0367 & 0.0292 & 0.0286 \\
\end{array}
\]

Table 6.1: Prediction norm and Euclidean norm ratios (overall \(\lambda_j\) relative to equation-by-equation \(\lambda_j\)'s, average over equations). Results (mean, median and standard deviation) are computed over 1000 replications.

It is evident from Table 6.1 that the proposed estimation procedure delivers much better performance in terms of the two measures. In particular, the superiority tends to be more evident (more than 10%) with higher dimension of the covariates and more equations.

Still consider the system of regression equations as in (6.1), but here we generate the data with dependency by following the Appendix D in Zhang and Wu (2017b). In particular, assume the linear process such that \(X_t = \sum_{\ell=0}^{\infty} A_\ell \xi_{t-\ell} \), with \(A_\ell = (\ell + 1)^{-\rho-1} M_\ell \), where \(M_\ell \) are independently drawn from Ginibre matrices, i.e. all the entries of \(M_\ell \) are i.i.d. \(N(0,1)\), and in practice the sum is truncated to \(\sum_{\ell=0}^{1000} \). We set \(\rho\) to be 1.0 for the weaker dependence and 0.1 for the stronger dependence cases respectively. Let \(\xi_{k,t} = e_{k,t} (0.8 e_{k,t-1}^2 + 0.2)^{1/2} \) where \(e_{k,t}\) are i.i.d. distributed as \(t(d)/\sqrt{d/(d-2)}\) and \(t(d)\) is the Student’s \(t\) with degree of freedom \(d\) (take \(d = 8\) for example). \(\varepsilon_t\) are generated by following the same fashion independently.

We take \(n = 100\), \# of bootstrap replications = 5000, \(J, K = 50, 100\) and 150. Based on bias-variance tradeoff, several approaches were suggested to determine the optimal choice of \(b_n\) for univariate case. Concerning the high-dimensional case, we propose to take the one which gives the lowest prediction norm as the optimal choice. Below we report the average prediction norm \(J^{-1} \sum_{j=1}^{J} |\hat{\beta}_j - \beta_{0,j}|_{j,pr} \) with several block sizes \(b_n\) under different settings and the minimal ones are in bold.

\[
\begin{array}{cccc}
\rho = 0.1 \text{ (stronger dependency)} & & & \rho = 1.0 \text{ (weaker dependency)} \\
J = K = 50 & J = K = 100 & J = K = 150 & J = K = 50 & J = K = 100 & J = K = 150 \\
\hline
b_n = 2 & 2.0721 & 2.9122 & 3.5932 & 2.0165 & 2.6270 & 3.2286 \\
\ \ = 4 & 2.0627 & 2.8924 & 3.5617 & 2.0303 & 2.6183 & 3.2225 \\
\ \ = 6 & 2.0487 & 2.9007 & 3.5235 & 2.0834 & 2.6288 & 3.2198 \\
\ \ = 8 & 2.0388 & 2.8841 & 3.5073 & 2.2149 & 2.6502 & 3.3230 \\
\ \ = 10 & 2.0521 & 2.8836 & 3.5268 & 2.3576 & 2.7099 & 3.2975 \\
\ \ = 12 & 2.0581 & 2.9065 & 3.5687 & 2.5592 & 2.8310 & 3.3895 \\
\end{array}
\]

Table 6.2: The prediction norm (average over equations) using several choices of \(b_n\). Results are computed over 1000 simulations.

From Table 6.2 it is apparent that a larger block size is required for the stronger dependency
case. Moreover, the choice also depends on the dimensionality, which is more evident for relatively weaker dependent data. We note that when \( J = K = 50, \rho = 1.0 \) the ordinary multiplier bootstrap (with \( b_n = 1 \)) produces 2.1003 as the average prediction norm, therefore we suggest \( b_n = 2 \) for this case.

The prediction norm \( |\hat{\beta}_j - \beta_0^j|_{\text{pr}} \) and the Euclidean norm \( |\hat{\beta}_j - \beta_0^j|_2 \) ratios (using the optimal \( b_n \) suggested in Table 6.2 for each case correspondingly) are presented in Table 6.3. Again we report the results with the jointly estimated \( \lambda \) (selected by multiplier block bootstrap) relative to using the single equation \( \lambda_j \)'s (selected by the multiplier bootstrap).

<table>
<thead>
<tr>
<th>( \rho = 0.1 ) (stronger dependency)</th>
<th>( \rho = 1.0 ) (weaker dependency)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( J = K = 50 )</td>
<td>( J = K = 100 )</td>
</tr>
<tr>
<td><strong>Prediction norm</strong></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.9141</td>
</tr>
<tr>
<td>Median</td>
<td>0.9165</td>
</tr>
<tr>
<td>Std.</td>
<td>0.0436</td>
</tr>
<tr>
<td><strong>Euclidean norm</strong></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.9017</td>
</tr>
<tr>
<td>Median</td>
<td>0.9062</td>
</tr>
<tr>
<td>Std.</td>
<td>0.0515</td>
</tr>
</tbody>
</table>

Table 6.3: Prediction norm and Euclidean norm ratios (overall \( \lambda \) relative to equation-by-equation \( \lambda_j \)'s, average over equations). Results (mean, median and standard deviation) are computed over 1000 replications.

The results show that the coefficient estimation performance measured by both the prediction norm and the Euclidean norm is in favor of the joint penalty level with multiplier block bootstrap approach. The results are robust over different dimension cases with stronger or weaker dependency.

### 6.2 Simultaneous Inference

In this subsection we consider the following regression model for the purpose of simultaneous inference on the parameters within a system of equations

\[
Y_{jt} = d_{jt}\alpha_j^0 + X_{jt}^T\beta_j^0 + \varepsilon_{jt}, \quad d_{jt} = X_{jt}^T\theta_j^0 + v_{jt}, \quad t = 1, \ldots, n, \ j = 1, \ldots, J,
\]

(6.2)

where \( \alpha_j^0 = \alpha^0 \) for all \( j \). Also, \( \beta_j^0, \theta_j^0 \in \mathbb{R}^K \) are assumed to be sparse. In particular, we divide the indices \( 1, \ldots, K \) evenly into blocks with a fixed block size 5, \( \beta_{jk}^0 \) and \( \theta_{jk}^0 \) are independently drawn from Unif\([0,5]\) and Unif\([0,0.25]\) respectively, if \( k \) and \( j \) belong to the same block and 0 otherwise. The way to generate \( X_{jt}, \varepsilon_t \) and \( v_t \) is same as the dependent data setting above.

We consider the sample size \( n = 100 \). Our goal is to estimate and make inferences on the target variables \( d_{jt} \)'s based on the procedure proposed in Section \[\text{[ ]}\]. We evaluate and compare the empirical power and size performance of the confidence intervals constructed by the asymptotic distribution theory (4.6), block bootstrap (4.4) and the simultaneous confidence regions via block bootstrap (4.8). The bootstrap statistics are computed based on 5000 replications and we also take the optimal block size according to the numerical comparison conducted above.
Note that the case of $\alpha^0 = 0$ gives the size performance under the null hypothesis, while $\alpha^0$ uniformly lies in $[0, 2.5]$ and $[0, 5]$ illustrate the power results.

Table 6.4 shows the average rejection rate of $H^j_0: \alpha^0_j = 0$ over $j$ for individual (or multiple) inference and the rejection rate of $H_0: \alpha^0_1 = \cdots = \alpha^0_J = 0$ for simultaneous inference under different settings of $J, K$ and $\rho$. Multiple testing procedure via step-down method, see e.g. Romano and Wolf (2005); Chernozhukov et al. (2013a), is considered to control the false positives in evaluating the power performance. The rejection rates are computed over 1000 simulation samples.

<table>
<thead>
<tr>
<th>$\rho = 0.1$ (stronger dependency)</th>
<th>$\rho = 1.0$ (weaker dependency)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$J = K = 50$</td>
<td>$J = K = 100$</td>
</tr>
<tr>
<td>$\alpha^0 = 0$</td>
<td></td>
</tr>
<tr>
<td>Ind. Asym.</td>
<td>0.0166</td>
</tr>
<tr>
<td>Ind. Boot.</td>
<td>0.0303</td>
</tr>
<tr>
<td>Simult. Boot.</td>
<td>0.0260</td>
</tr>
<tr>
<td>$\alpha^0 \sim \text{Unif}[0, 2.5]$</td>
<td></td>
</tr>
<tr>
<td>Ind. Asym.</td>
<td>0.8714</td>
</tr>
<tr>
<td>Ind. Boot.</td>
<td>0.8746</td>
</tr>
<tr>
<td>Mult. Boot.</td>
<td>0.8413</td>
</tr>
<tr>
<td>$\alpha^0 \sim \text{Unif}[0, 5]$</td>
<td></td>
</tr>
<tr>
<td>Ind. Asym.</td>
<td>0.9376</td>
</tr>
<tr>
<td>Ind. Boot.</td>
<td>0.9390</td>
</tr>
<tr>
<td>Mult. Boot.</td>
<td>0.9282</td>
</tr>
</tbody>
</table>

Table 6.4: Average rejection rate of $H^j_0: \alpha^0_j = 0$ over $j$ for the individual (or multiple) inference and the rejection rate of $H_0: \alpha^0_1 = \cdots = \alpha^0_J = 0$ for simultaneous inference under several true $\alpha^0$ values (given the significance level $= 0.05$).

It is shown that for individual inference our proposed individual bootstrap approach provides a closer size control to the nominal $\alpha$ and more powerful empirical rejection probabilities compared to constructing the confidence intervals by asymptotic normality in most of the cases. Moreover, the simultaneous inference outperforms the individual inference in size accuracy and in terms of the power performance, the multiple testing is relatively conservative after controlling the false positives. Overall, we observe that the results using bootstrap approach are robust over different dimension settings under either stronger or weaker dependency cases.

7 Empirical Analysis: Textual Sentiment Spillover Effects

Financial markets are driven by information, and this is a well-known phenomenon among investors. More frequent news and availability of sentiment data allows study of the impact of firm-specific investor sentiment on market behavior such as stock returns, volatility and liquidity; see Baker and Wurgler (2006) Tetlock (2007) among others. Moreover, powerful statistical tools (e.g. LASSO-type estimators) are being used to model complex relationships among individuals. For example, Audrino and Tetereva (2017) analyze the influence of news on US and European companies by constructing a sparse predictive network via adaptive LASSO and
related testing procedures. In this section the developed technology is applied to study textual sentiment spillover effects across individual stocks. This is different from the 'equation-by-equation' analysis in [Audrino and Tetereva (2017)], since we build up a system of regression equations and implement the estimation and the inference of the network jointly.

7.1 Data Source

The empirical study in this paper is carried out based on the financial news articles published on the NASDAQ community platform from January 2, 2015 to December 29, 2015 (252 trading days). The data were gathered via a self-written web scraper to automate the downloading process. The dataset is available at the Research Data Centre (RDC), Humboldt-Universität zu Berlin. Moreover, unsupervised learning approaches are employed to extract sentiment variables from the articles. Two sentiment dictionaries: the BL option lexicon [Hu and Liu (2004)] and the LM financial sentiment dictionary [Loughran and McDonald (2011)] were used in Zhang et al. (2016). For each article $i$ (published on day $t$), the average proportion of positive/negative words using BL or LM lexica - $\text{Pos}_{j,i,t}^{\text{BL}}$, $\text{Neg}_{j,i,t}^{\text{BL}}$, $\text{Pos}_{j,i,t}^{\text{LM}}$, $\text{Neg}_{j,i,t}^{\text{LM}}$ - are considered as the text sentiment variables. Furthermore, the bullishness indicator for stock $j$ on day $t$ with the related articles $i = 1, \ldots, m$ (based on a particular lexicon) is constructed by following Antweiler and Frank (2004)

$$B_{j,t} = \log\left\{ \frac{1 + m^{-1} \sum_{i=1}^{m} 1(\text{Pos}_{j,i,t}^{\text{BL}} > \text{Neg}_{j,i,t}^{\text{BL}})}{1 + m^{-1} \sum_{i=1}^{m} 1(\text{Pos}_{j,i,t}^{\text{BL}} < \text{Neg}_{j,i,t}^{\text{BL}})} \right\}. \quad (7.1)$$

We refer to Zhang et al. (2016) for more details about the data gathering and processing procedure. 63 individual stocks which are S&P 500 component stocks from 9 Global Industrial Classification Standard (GICS) sectors are considered. They are traded at NSDAQ Stock Exchange or NYSE. The list of the stock symbols and the corresponding company names can be found in Table D.1 in Appendix D in the supplementary materials.

The daily log returns $R_{j,t}$ and log volatilities $\log(\sigma_{j,t}^2)$ for the stocks over the same time span are taken as response variables. More precisely, the Garman and Klass (1980) range-based measure to represent the volatility level is employed:

$$\sigma_{j,t}^2 = 0.511(u_{j,t} - d_{j,t})^2 - 0.019\{r_{j,t}(u_{j,t} + d_{j,t}) - 2u_{j,t}d_{j,t}\} - 0.383r_{j,t}^2, \quad (7.2)$$

where $u_{j,t} = \log(P_{j,t}^H) - \log(P_{j,t}^O)$, $d_{j,t} = \log(P_{j,t}^L) - \log(P_{j,t}^O)$, $r_{j,t} = \log(P_{j,t}^C) - \log(P_{j,t}^O)$, with $P_{j,t}^H$, $P_{j,t}^L$, $P_{j,t}^O$, and $P_{j,t}^C$ denote the highest, lowest, opening and closing prices, respectively. In addition, the S&P 500 index returns and Chicago Board Options Exchange volatility index (VIX) are included as the state variables. The financial time series data were originally obtained from Datastream, and GICS sector information was found at Compustat.
7.2 Model Setting and Results

We now construct a network model to detect the spillover effects from sentiment variables to financial variables by

\[
\begin{align*}
    r_{j,t} &= c_j + B_t^\top \beta_j + z_t^\top \gamma_j + r_{j,t-1} \delta_j + \varepsilon_{j,t}, \\
    \log \sigma^2_{j,t} &= c_j + B_t^\top \beta_j + z_t^\top \gamma_j + \log \sigma^2_{j,t-1} \delta_j + \varepsilon_{j,t},
\end{align*}
\]

where \( j = 1, \ldots, J \) indicate the stock symbols, \( B_t = (B_{1,t}, \ldots, B_{J,t})^\top \) and \( z_t \) includes the state variables.

It is of interest to make inferences on the parameters \( \beta_j \in \mathbb{R}^J, j = 1, \ldots, J \). Following the framework introduced in Section 4, an estimation procedure with three steps needs to be implemented.

S1 For each \( j \), run LASSO on (7.3) and keep the estimator \( \hat{\beta}_{j}^{[1]}(j), \hat{\gamma}_{j}^{[1]} \) and \( \hat{\delta}_{j}^{[1]} \) and \( \hat{c}_{j}^{[1]} \).

S2 For each \( j \), run LASSO on \( B_{j,t} = (B_{j,t}^\top, z_t^\top, r_{j,t-1})^\top \theta_j + v_{j,t} \) to model the dependence among sentiment variables. In particular, we propose to take the joint penalty level obtained via block multiplier bootstrap (discussed in Section 3.2) for this regression system. Keep the residuals as \( \hat{\varepsilon}_{j,t} = B_{j,t} - (B_{j,t}^\top, z_t^\top, r_{j,t-1})^\top \hat{\theta}_j \).

S3 For each \((j,k)\), run IV regression of \( r_{j,t} - \hat{c}_{j}^{[1]} - B_{j,t}^\top \hat{\beta}_{j}^{[1]}(j-1) - z_t^\top \hat{\gamma}_{j}^{[1]} - r_{j,t-1} \hat{\delta}_{j}^{[1]} \) on \( B_{k,t} \) using \( \hat{v}_{k,t} \) as an instrument variable. Then we obtain the final estimator \( \hat{\beta}_{jk}^{[2]} \).

If for stock \( j \), the sentiment variable of firm \( k \) is selected into the active set after the individual significance test i.e., the null hypothesis \( H_{0}^{jk} : \beta_{jk} = 0 \) is rejected under the block multiplier bootstrap procedure (as discussed in Section 6.1 we pre-determine \( b_n = 5 \) by choosing the one gives the lowest prediction norm in the LASSO estimation in S1 on a grid search), then we put a directional edge from \( k \) to \( j \). As a result, we achieve a \( 0-1 \) adjacency matrix describing the dependency network from sentiment variable to financial variable. Note that the diagonal elements in the matrix show the self-effect of stocks.

The graphical network for stock returns and volatility modelled by (7.3) based on BL and LM lexica (from 01/02/15 to 12/29/15) is depicted in Figures 7.1-7.2.
Figure 7.1: The dependency network among individual stocks from sentiment variables to return.
Figures 7.1-7.2 depict the dependency networks among individual stocks. Given that the time series of returns and volatility are scaled and centered before implementing the estimation procedure, we find even denser spillover effects in the volatility analysis. This indicates the stock volatility is more sensitive to sentiment than returns. Moreover, the relationships between sectors are also of interest. The simultaneous confidence region constructed via the bootstrap approach introduced in Section 4.2 may help us to detect whether the sentiment information...
from one sector has joint influence on the returns of the stocks in another sector. In particular, we look at the null hypothesis:

\[ H_{S_1,S_2}^0 : \beta_{jk} = 0, \forall j \in S_1, k \in S_2, \] where \( S_1 \) and \( S_2 \) represent two groups of stocks that belong to two sectors, respectively. The conclusion that the sentiment from sector \( S_2 \) has a joint effect on the returns or volatility of sector \( S_1 \) can be drawn if the null hypothesis is rejected with the simultaneous confidence region (LS) under the significance level \( = 0.05 \).

Figure 7.3: The dependency network among sectors from sentiment variables to financial variables.

Figure 7.3 describes the spillover effect network from sentiment to financial variables on the sector levels. In particular, the connections from energy to health care is found to be significant in the analysis of stock returns; while if volatility is focused on then the spillover effects from financials to health care, from information technology to energy, also from consumer discretionary to utilities are detected.

**Comment 7.1** (Link to GGM). Another popular way to conduct the network analysis in the literature is the GGM, which is corresponding to the estimation of a high dimensional precision matrix. And under the Gaussian assumption our SRE can be linked to a nodal wise GGM. In particular, one can estimate the coefficients in each equation of SRE by using a sparse Graphical model estimation, for example the LASSO type estimation as in Yuan and Lin (2007), and thus we build the link equation-by-equation.
Consider the VAR(1) as an example, the $j$th equation in the SRE is given by $Y_{j,t} = \Phi_j Y_{t-1} + \varepsilon_{j,t}$, where $Y_t$ is covariance stationary with $\text{Var}(Y_t) = \Gamma$ (p.d.). Correspondingly, we look at the vector $\tilde{Y}_{j,t} = (Y_{j,t}, Y_{1:t-1}, \ldots, Y_{J:t-1})^\top$ belonging to an undirected graph $(V_j, E_j)$ with vertex set $(1, \ldots, J + 1)$. Suppose $\tilde{Y}_{j,t} \sim \text{MVN}(0, \Sigma_j)$, $\Sigma_j = \begin{bmatrix} \Gamma_{jj} & \Phi_j \Gamma \\ (\Phi_j \Gamma)^\top & \Gamma \end{bmatrix}$. Define $C_j \overset{\text{def}}{=} \Phi_j \Gamma \Phi_j^\top$, then we have the precision matrix as $\Theta_j = \Sigma_j^{-1} = \begin{bmatrix} (\Gamma_{jj} - C_j)^{-1} & -((\Gamma_{jj} - C_j)^{-1}) \Phi_j \\ -\Phi_j (\Gamma_{jj} - C_j)^{-1} \Gamma^{-1} + \Phi_j^\top (\Gamma_{jj} - C_j)^{-1} \Phi_j \end{bmatrix}$.

It can be seen that $\Phi_{jk} = 0$ would imply that the $(1, k + 1)$th element of $\Theta_j$ is zero and vice versa. In addition, a LASSO type estimator proposed in [Yuan and Lin (2007)] can be obtained by solving

$$
\hat{\Theta}_j = \arg\max_{\Theta} \left\{ -\log \det(\Theta) + \text{trace}(S_j \Theta) + \lambda_j \sum_{k} |\Theta_{jk}| \right\},
$$

where $S_j \overset{\text{def}}{=} n^{-1} \sum_{t=1}^n \tilde{Y}_{j,t} \tilde{Y}_{j,t}^\top$.

In an unreported simulation study we compare the estimation performance between our proposed approach and the nodal wise GGM under the VAR(1) model. The results show that the nodal wise GGM which is approximated to SRE has worse prediction performance than our method, which can be obtained from the authors upon request.
Supplementary Material

A  Theorems for Joint Penalty over Equations

Recall that the theoretical choice $\lambda^0(1-\alpha)$ is defined as the $(1-\alpha)$ quantile of
\[ \max_{1 \leq k \leq K, 1 \leq j \leq J} 2c\sqrt{\tau}|S_{jk}/\Psi_{jk}|. \]
First, we provide the analogue results of Theorem 5.1 and Corollary 5.1.

**Theorem A.1.** Under (A1) and (A3), we have
\[
P(2c\sqrt{n} \max_{1 \leq k \leq K, 1 \leq j \leq J} |S_{jk}/\Psi_{jk}| \geq r) \leq C_1 \varpi_n n r^{-q} \sum_{j=1}^{J} \sum_{k=1}^{K} \frac{\|X_{jk}, \varepsilon_j, \|q_{jk}, \Psi_{jk}}{q_{jk}^n} + C_2 \sum_{j=1}^{J} \sum_{k=1}^{K} \exp \left( -\frac{C_3 r^2 q_{jk}^2}{n \|X_{jk}, \varepsilon_j, \|_2^2} \right),
\]
(A.1)

where for $\varsigma > 1/2 - 1/q$ (weak dependence case), $\varpi_n = 1$; for $\varsigma < 1/2 - 1/q$ (strong dependence case), $\varpi_n = n^{q/2-1-\varsigma}$. $C_1, C_2, C_3$ are constants depending on $q$ and $\varsigma$.

**Corollary A.1** (Bound for $\lambda^0(1-\alpha)$ and Oracle Inequalities under IC). Under (A1) and (A3), given $\lambda^0(1-\alpha)$ satisfies
\[
\lambda^0(1-\alpha) \leq \max_{1 \leq k \leq K, 1 \leq j \leq J} \left\{ \|X_{jk}, \varepsilon_j, \|_{2,\varsigma} \sqrt{n \log(KJ/\alpha)} \vee \|X_{jk}, \varepsilon_j, \|_{q,\varsigma}(n \varpi_n KJ/\alpha)^{1/q} \right\},
\]
(A.2)

additionally assume that the RE condition (A2) holds uniformly over equations $j = 1, \ldots, J$ with probability $1 - o(1)$, and under the exact sparsity assumption (3.1), then $\tilde{\beta}_j$ obtained from (3.2) under IC satisfy
\[
|\tilde{\beta}_j - \beta_j^{(j,pr)}| \leq C \varsigma \max_{1 \leq k \leq K} \Psi_{jk} \max_{1 \leq j \leq J} \left\{ \|X_{jk}, \varepsilon_j, \|_{2,\varsigma} \sqrt{n \log(KJ/\alpha)} \vee \|X_{jk}, \varepsilon_j, \|_{q,\varsigma} n^{1/q-1}(n \varpi_n KJ/\alpha)^{1/q} \right\},
\]
(A.3)

with probability $1 - \alpha - o(1)$, where for $\varsigma > 1/2 - 1/q$ (weak dependence case), $\varpi_n = 1$; for $\varsigma < 1/2 - 1/q$ (strong dependence case), $\varpi_n = n^{q/2-1-\varsigma}$, and the constant $C$ depends on the RE constants.

The other empirical choices of the joint penalty level can be:

a) $Q(1-\alpha)$: the $(1-\alpha)$ quantile of $2c\sqrt{n} \max_{1 \leq k \leq K, 1 \leq j \leq J} |Z_{jk}/\Psi_{jk}|$. In practice, one can take an alternative choice such that $\tilde{Q}(1-\alpha) \equiv 2c\sqrt{n} \Phi^{-1} \{1 - \alpha/(2KJ)\}.$

b) $\Lambda(1-\alpha) \equiv 2c\sqrt{\varpi_n} \{B \}^{1-\alpha}$, where $\varpi_n \equiv \Phi^{-1} \{1 - \alpha/(2KJ)\}$ is the $(1-\alpha)$ quantile of $\max_{1 \leq k \leq K, 1 \leq j \leq J} |Z_{jk}/\Psi_{jk}|.$

For a) again we need the Gaussian approximation results for the vectorized process $\vec{\Sigma} \equiv \text{vec}\{(S_{jk})_{k=1}^K \}_{j=1}^J = \frac{1}{\sqrt{n}} \sum_{t=1}^{\tau_n} \hat{\Lambda}_t$, where $\hat{\Lambda}_t \equiv \text{vec}\{(X_{jk}, t \varepsilon_k)_{k=1}^K \}_{j=1}^J$ similar to Theorem 5.2 and Corollary 5.2 to justify the choice of $\lambda$ as $Q(1-\alpha)$.
Let \( X_t \overset{\text{def}}{=} \text{vec}[[ (X_{jk,t})_{k=1}^K]_{j=1}^J \]}. We first aggregate the dependence adjusted norm over \( j = 1, \ldots, J \) and \( k = 1, \ldots, K \):

\[
\|\mathcal{X}_t\|_{q,\varsigma} \overset{\text{def}}{=} \sup_{m \geq 0} (m + 1)^{\varsigma} \sum_{t=m}^{\infty} \delta_{q,t}, \quad \delta_{q,t} \overset{\text{def}}{=} \|\mathcal{X}_t - \mathcal{X}_t^{\varsigma}\|_{q},
\]

(A.4)

where \( q \geq 1 \), and \( \varsigma > 0 \). Moreover, define the following functions:

\[
\Phi_{q,\varsigma} \overset{\text{def}}{=} 2 \max_{1 \leq k \leq K, 1 \leq j \leq J} \|X_{jk,\cdot}\|_{q,\varsigma} \|\varepsilon_{j,\cdot}\|_{q,\varsigma}, \quad \Gamma_{q,\varsigma} \overset{\text{def}}{=} 2 \left( \sum_{j=1}^{J} \|\varepsilon_{j,\cdot}\|_{q,\varsigma}^{2/q} \right)^{2/q} \left( \sum_{k=1}^{K} \sum_{j=1}^{J} \|X_{jk,\cdot}\|_{q,\varsigma}^{q/2} \right)^{2/q},
\]

\[
\Theta_{q,\varsigma} \overset{\text{def}}{=} \Gamma_{q,\varsigma} \wedge \left\{ \|\mathcal{X}_t\|_{q,\varsigma} \|\varepsilon_{j,\cdot}\|_{q,\varsigma} (\log K J)^{3/2} \right\}.
\]

(A.5)

Let \( L_1 = [\Phi_{1,\varsigma} \Phi_{4,0} (\log(KJ))^{1/2}]^{1/q}, \quad W_1 = (\Phi_{0,0}^2 + \Phi_{4,0}^2) (\log(KJ_1))^{1/2}, \quad W_2 = (\Phi_{4,\varsigma}^2 (\log(KJ_1))^{1/2}, \quad W_3 = (n^{-\varsigma} (\log(KJ_1))^{3/2} \Theta_{2q,\varsigma}^{1/2 - 1/2 - 1/q}, \quad N_1 = \{ n / \log(KJ) \}^{q/2} \Theta_{2q,\varsigma}^{3/2}, \quad N_2 = n (\log(KJ))^{-2} \Phi_{4,\varsigma}^{2}, \quad N_3 = n^{-1} (\log(KJ))^{-1/2} \Theta_{2q,\varsigma}^{3/2}.

(A.4')

i) (weak dependency case) Given \( \Theta_{2q,\varsigma} < \infty \) with \( q \geq 4 \) and \( \varsigma > 1/2 - 1/q \), then

\[
\Theta_{2q,\varsigma} n^{1/4 - q/2 - \varsigma/2} (\log(KJ_1))^{1/2} \to 0 \quad \text{and} \quad L_1 \max(W_1, W_2) = o(1) \min(N_1, N_2).
\]

ii) (strong dependency case) Given \( 0 < \varsigma < 1/2 - 1/q \), then \( \Theta_{2q,\varsigma} (\log(KJ))^{1/2} = o(n^{-\varsigma}) \)

and \( L_1 \max(W_1, W_2, W_3) = o(1) \min(N_2, N_3) \).

Theorem A.2. Under (A1), (A3) and (A4'), for each \( k = 1, \ldots, K, j = 1, \ldots, J \) assume that there exists a constant \( c > 0 \) such that \( \min_{1 \leq k \leq K, 1 \leq j \leq J} \text{var}(S_{jk}) \geq c \), then we have

\[
\rho(D^{-1} \tilde{S}, D^{-1} \tilde{Z}) \to 0, \quad \text{as} \quad n \to \infty,
\]

(A.6)

where \( \tilde{Z} \sim N(0, \Sigma_{\tilde{X}}), \Sigma_{\tilde{X}} \) is the \( JK \times JK \) long-run variance-covariance matrix of \( \tilde{X}_t \), and \( D \) is a diagonal matrix with the square root of the diagonal elements of \( \Sigma_{\tilde{X}} \), namely

\[
\left\{ \sum_{t=-\infty}^{\infty} \mathbb{E}(X_{jk,t} X_{jk,(t-\ell)} \varepsilon_{j,t} \varepsilon_{j,(t-\ell)}) \right\}^{1/2} = \text{var}(S_{jk}), \quad \text{for} \quad k = 1, \ldots, K, j = 1, \ldots, J.
\]

Corollary A.2. Under the conditions of Theorem A.2, we have

\[
\sup_{\alpha \in (0,1)} \left\{ \mathbb{P} \left[ \max_{1 \leq k \leq K, 1 \leq j \leq J} 2c^{\sqrt{n}} |S_{jk}/\Psi_{jk}| \geq Q(1 - \alpha) \right] - \alpha \right\} \to 0, \quad \text{as} \quad n \to \infty.
\]

(A.7)

Corollary A.3. Under the conditions of Theorem A.2 and assume \( \Phi_{2q,\varsigma} < \infty \) with \( q > 4 \), \( b_n = \mathcal{O}(n^\varsigma) \) for some \( 0 < \varsigma < 1 \). Let \( F_\varsigma = n, \quad \text{for} \quad \varsigma > 1 - 2/2 - \varsigma/2, \quad F_\varsigma = n^{2/2 - \varsigma/2}, \quad \text{for} \quad 1/2 - 2/2 < \varsigma < 1 - 2/2 - \varsigma/2, \quad F_\varsigma = n^{2/2 - \varsigma/2}, \quad \text{for} \quad \varsigma < 1 < 1/2 - 2/2. \quad \text{Given} \quad n^{-1} (\log(KJ))^{2} \max\{ n^{1/2b_n^{1/2} \Phi_{2q,\varsigma}^{1/2}, n^{1/2b_n^{1/2} \Phi_{2q,\varsigma}^{1/2}} \} \mathbb{E}(F_\varsigma^{2/2} / \sqrt{\log(KJ)} \Phi_{2q,\varsigma}^{1/2} F_\varsigma^{2/2} / \sqrt{\log(KJ)} \Phi_{2q,\varsigma}^{1/2} \} (\log(KJ)) = o(1), \quad \text{where} \quad v'(b_n) = (b_n + 1)^{-\varsigma} + 2v_{n,2}/b_n, \quad v_{n,2} = \log b_n \quad (\text{resp.} \quad b_n^{1/2} \quad \text{or} \quad 1), \quad \text{for} \quad \varsigma = 1 \quad (\text{resp.} \quad \varsigma < 1 \quad \text{or} \quad \varsigma > 1), \quad \text{then we have}

\[
\rho(D^{-1} \tilde{S}, D^{-1} \tilde{Z}) \to 0, \quad \text{as} \quad n \to \infty,
\]

(A.8)

where \( \tilde{D} = \{ \text{diag}(\Sigma_{\tilde{X}^\varsigma}) \}^{1/2}, \quad \Sigma_{\tilde{X}} = \frac{1}{b_n} \sum_{t=1}^{b_n} \left( \sum_{t=(t-1)b_n+1}^{tb_n} \tilde{X}_t \right) \left( \sum_{t=(t-1)b_n+1}^{tb_n} \tilde{X}_t^\varsigma \right)^T. \)
Similar to Corollary 5.4, we can provide a refined bound for $\lambda^0(1 - \alpha)$ and also the oracle inequalities under IC as follows.

**Corollary A.4 (Bounds for $\lambda^0(1 - \alpha)$ and Oracle Inequalities under IC with Gaussian Approximation Results).** Under the conditions of Theorem A.2, suppose $2\log(KJ)^{-1/2} + \rho(D^{-1}\tilde{S}, D^{-1}\tilde{Z}) = o(\alpha)$ and let $Z_\alpha \geq 2\tilde{c}\sqrt{\log(KJ)}$, where $\tilde{c}$ is no less than the $c$ in the definition of $\lambda^0(1 - \alpha)$, then we have $\lambda^0(1 - \alpha)$ satisfying

$$\lambda^0(1 - \alpha) \leq Z_\alpha,$$

(A.9)

additionally assume that the RE condition (A2) holds uniformly over equations $j = 1, \ldots, J$ with probability $1 - o(1)$, and given the exact sparsity assumption (3.1), then $\tilde{\beta}_j$ obtained from (3.2) under IC satisfies

$$|\tilde{\beta}_j - \beta^0|_{\delta, pr} \leq C\sqrt{s} \max_{1 \leq k \leq K} \Psi_{jk} \frac{\log(KJ)}{\sqrt{n}},$$

(A.10)

with probability $1 - \alpha - o(1)$, where the constant $C$ depends on the RE constants.

Next, we need to show the validity of b) Let $\tilde{Z}^{[B]} = \text{vec}\{(Z_{jk})_{k=1}^J\}_{j=1}^J$ and $\tilde{\Psi} = \text{vec}\{((\Psi_{jk})_{k=1}^J)_{j=1}^J\}$. Similarly to Theorem 5.3 we have the following results:

**Theorem A.3.** Under (A1), (A3), and assume $\Phi_{2q, \varsigma} < \infty$ with $q > 4$, $b_n = \mathcal{O}(n^{\eta})$ for some $0 < \eta < 1$ (the detailed rate is calculated in (B.3)), then

$$\tilde{p}_n = \sup_{r \in \mathbb{R}} \left| P(\tilde{Z}^{[B]} / \tilde{\Psi}_{\infty} \leq r | \mathcal{X}_t, \varepsilon) - P(\tilde{Z} / \tilde{\Psi}_{\infty} \leq r) \right| \to 0, \text{ as } n \to \infty,$$

(A.11)

and

$$\sup_{\alpha \in (0,1)} \left| P(|\tilde{Z} / \tilde{\Psi}_{\infty} | \geq q^{[B]}_{(1-\alpha)}) - \alpha \right| \to 0, \text{ as } n \to \infty.$$

(A.12)

Lastly, we show the performance bounds for the OLS post-LASSO estimator in the following theorem.

For each $p \leq n$, $T_j \subset \{1, \ldots, K\}$, $|T_j| \leq p$, we define the class of functions $G_{T_j} = \{\varepsilon_j, X_{j,l}^\delta | \varepsilon_j, X_{j,l}^\delta, \text{supp}(\delta) \subset T_j, |\delta|_2 = 1\}$. The covering number of the function class is given by $\sup_{\mathcal{C}} N(\epsilon, G_{T_j}, \|\cdot\|_{Q, 1})$. Also define $\mathcal{F}_{j,p} = \{G_{T_j} : T_j \subset \{1, \ldots, K\}, |T_j| \leq p\}$. For any $f \in \mathcal{F}_{j,p}$, there exists a set $F_{j,p}$ such that $\min_{f' \in F_{j,p}} \|f - f'\|_{Q, 1} \leq \epsilon$, and the cardinality of the set is denoted by $|F_{j,p}|$. Consider the vector $\vartheta_{l,t}$ of length $|F_{j,p}|$, such that for $l = 1, \ldots, |F_{j,p}|$, there is $\vartheta_{l,t} = (f - E f) / \psi_{l,t}$ with $\psi_{l,t} = \{\text{var}(G_{n}(f))\}^{1/2}$, corresponding to each $f \in F_{j,p}$. The aggregated dependence adjusted norm is given by

$$\|\vartheta_{l,t}\|_{q, \varsigma} \overset{def}{=} \sup_{m \geq 0} (m + 1)^{\varsigma} \sum_{t=m}^{\infty} \|\vartheta_t - \vartheta_t^*\|_{\infty},$$

(A.13)

where $q \geq 1$, and $\varsigma > 0$. Moreover, define the following quantities (for simplicity we drop the
To evoke the Gaussian approximation on $G_n(f)/\psi_f$, we need to impose the following assumptions additionally. Define $L_1^q = \{\Phi_{2e_\zeta}(\log |F_{j,p}|)\}^{1/4}$, $W_2^q = \{\Phi_{2e_\zeta}(\log |F_{j,p}|)\}^2$, $W_4^q = [\log(|F_{j,p}|)]^{2}\Phi_{2e_\zeta}(\log |F_{j,p}|)^{1/2-1/4}$, $N_1^q = \log |F_{j,p}|^{q/2}(\Theta_{q,ς})^q$, $N_2^q = n\log |F_{j,p}|^{-2}(\Theta_{q,ς})^q$, $N_3^q = n\log |F_{j,p}|^{-2}(\Theta_{q,ς})^{1/2-q}$.

\(A7\) (weak dependency case) Given $\Theta_{q,ς} < \infty$ with $q \geq 2$ and $ς > 1/2 - 1/4$, then $\Theta_{q,ς} n^{1/4-1/2}\log(|F_{j,p}|)^{1/2} \rightarrow 0$ and $L_1 \max(W_1^q, W_2^q, W_4^q) = o(1) \min(N_1^q, N_2^q)$.

Comment A.1. For a random vector $z_t \in R^K$, suppose there exist constants $C, D > 0$, such that $\max_k \mathbb{E}\{\exp(|z_{kt}|/|D|)\} \leq C$. Then by Jensen’s inequality it follows that $\|z_t\|_\infty \leq D(\log K + \log C)^{1/4}$. In particular, for the case of sub-Gaussian random variables, there exists constant $D > 0$ such that $\mathbb{E}\{\exp(|z_{kt}|/|D|)\} - 1 \leq 1$, which implies $\|z_t\|_\infty \leq D\sqrt{\log K}$.

Similar to the discussion in Remark 5.6, consider the case with $\Theta_{q,ς} = O(\log |F_{j,p}|)^{1/4}$ and $\Phi_{2e_\zeta} = O(1)$, where $ς > 1/2 - 1/4$. Then $\Theta_{q,ς} n^{1/4-1/2}\log(|F_{j,p}|)^{1/2} \rightarrow 0$ becomes $\log |F_{j,p}|^{q/2}(\Theta_{q,ς})^{q-1} = o(n^{q/2-1})$, which implies that $L_1 \max(W_1^q, W_2^q, W_4^q) = o(1) \min(N_1^q, N_2^q)$.

As shown in the proof of Theorem A.4, $|F_{j,p}| \lesssim K^p(6\mu_j(p)|\sigma|^{1+\epsilon})^{1+\gamma}$ with $\epsilon = \sqrt{p\log K + (p + s)\log(6\mu_j(p)|\sigma|)/4}$, which means with $A7$, the dimension $K$ has to satisfy the condition $\{p\log K + (s + p)\log(\sqrt{K})\}^{1+\gamma/2} = o(n^{q/2-1})$, where we consider the case such that $|F_{j,p}|$ is larger than $n$.

Theorem A.4 (Prediction Performance Bounds for OLS Post-LASSO). Given (A1), (A3) and (A7), suppose (A2) with $c = \frac{c_1}{c_2} > 1$, and (A5) with $\tilde{j} = |\tilde{T}_j \setminus T_j|$, then under the exact sparsity assumption (3.1), for any $\tau > 0$, there is a constant $C_\tau$ independent of $n$, for all $j = 1, \ldots, J$ we have

$$\begin{align*}
\beta_j^{[p]} - \beta_j^{(0)}_{j,p} &\leq C_\tau \max_{1 \leq k \leq K} \Psi_j \sqrt{\frac{p \log K + (p + s)|\log(6\mu_j(p)|\sigma|) + \log n}{n}} \\
 &+ 1(T_j \not\subset \tilde{T}_j) C \sqrt{s} \max_{1 \leq k \leq K} \Psi_j \max_{1 \leq j \leq J} \left\{ \|X_{jk}, e_j\|_2, \frac{\sqrt{\log(K/J)}/(\alpha \log n)}{\sqrt{n}} \right\} \|X_{jk}, e_j\|_{q,ς} \sqrt{n^{1/4-1/4}}(\varpi_{n\log K/J}\alpha)^{1/4},
\end{align*}$$

(A15)

with probability $1 - \alpha - \tau - o(1)$, where for $ς > 1/2 - 1/4$ (weak dependence case), $\varpi_n = 1$; for $ς < 1/2 - 1/4$ (strong dependence case), $\varpi_n = n^{q/2-1-\varsigma}$. $\sigma = \max_j \{\text{var}(n^{-1/2}\sum_{t=1}^n e_{jt})\}^{1/2}$ and the constant $C$ depends on the RE constants.

In particular, suppose the Gaussian approximation results hold for $\lambda^{0}(1 - \alpha)$, the bound for it can be replaced according to Corollary A.4.
B Detailed Proofs

B.1 Proofs of Single Equation Estimation

Proof of Theorem 5.1. For each $j = 1, \ldots, J$, $k = 1, \ldots, K$, applying Theorem 2 of Wu (2016) gives

$$P(\sqrt{n}|S_{jk}| \geq x) \leq \frac{C_1 w_n n\|X_{jk}, \varepsilon_j, \varepsilon_j^*\|_q^q}{x^q} + C_2 \exp \left( \frac{-C_3 x^2}{n \|X_{jk}, \varepsilon_j, \varepsilon_j^*\|_2^2} \right),$$

where for $\varsigma > 1/2 - 1/q$, $w_n = 1$; for $\varsigma < 1/2 - 1/q$, $w_n = n^{q/2-1-\varsigma q}$. $C_1, C_2, C_3$ are three constants depending on $q$ and $\varsigma$. It follows that the conclusion holds if we set $x = (2c)^{-1} \Psi_j k r$. \qed

Proof of Theorem 5.2. According to the Minkowski’s inequality and Hölder’s inequality, we have

$$\sum_{t=m}^{\infty} \|X_{jk,t} \varepsilon_j, t - X_{jk,t}^* \varepsilon_j^*\|_q \leq \sum_{t=m}^{\infty} \left\{ \|X_{jk,t} (\varepsilon_j, t - \varepsilon_j^*)\|_q + \|X_{jk,t} - X_{jk,t}^* \varepsilon_j^*\|_q \right\} \leq \sum_{t=m}^{\infty} \left\{ \|X_{jk,t}\|_q \|\varepsilon_j - \varepsilon_j^*\|_2 + \|X_{jk,t} - X_{jk,t}^*\|_2 \|\varepsilon_j^*\|_2 \right\}.$$

Thus, it is easy to see that

$$\|X_{jk}, \varepsilon_j, \varepsilon_j^*\|_q \leq \|X_{jk,\cdot}\|_q \|\varepsilon_j, \varepsilon_j^*\|_q \leq 2 \|X_{jk,\cdot}\|_2 \|\varepsilon_j, \varepsilon_j^*\|_2 \leq 2 \|X_{jk,\cdot}\|_2 \|\varepsilon_j, \varepsilon_j^*\|_2.$$

Consequently, we have the following relationships:

$$\max_{1 \leq k \leq K} \|X_{jk, \varepsilon_j}, \varepsilon_j^*\|_q \leq 2 \max_{1 \leq k \leq K} \|X_{jk, \cdot}\|_2 \|\varepsilon_j, \varepsilon_j^*\|_2,$$

$$\left( \sum_{k=1}^{K} \|X_{jk, \varepsilon_j, \varepsilon_j^*}\|_q^{q/q_q} \right)^{1/q_q} \leq 2 \|\varepsilon_j, \varepsilon_j^*\|_2 \left( \sum_{k=1}^{K} \|X_{jk, \cdot}\|_2 \right)^{1/q_q},$$

$$\|X_{jk, \varepsilon_j, \varepsilon_j^*}\|_q \leq 2 \|X_{jk, \cdot}\|_2 \|\varepsilon_j, \varepsilon_j^*\|_2.$$

Therefore, the conditions in Theorem 3.2 of Zhang and Wu (2017a) can be verified for the $K$-dimensional stationary process $X_{j,t} \varepsilon_j, t$. Finally, applying that theorem yields the Gaussian approximation results. \qed

Proof of Corollary 5.2. It follows directly from the Gaussian approximation results in Theorem 5.2. \qed

Proof of Corollary 5.3. The proof follows that of Corollary 5.4 in Zhang and Wu (2017a). For $w > 0$, we have

$$\rho(\hat{D}^{-1}_{j} S_j, D^{-1}_{j} Z_j) = \sup_{r \geq 0} |P(|\hat{D}^{-1}_{j} S_j|_\infty \geq r) - P(|D^{-1}_{j} Z_j|_\infty \geq r)| \leq \rho(D^{-1}_{j} S_j, D^{-1}_{j} Z_j) + \sup_{r \geq 0} P(|D^{-1}_{j} Z_j|_\infty - r | \leq w) + P(|(D^{-1}_{j} - \hat{D}^{-1}_{j}) S_j|_\infty \geq w) \leq \rho(D^{-1}_{j} S_j, D^{-1}_{j} Z_j) + w \sqrt{\log K} + P(|(D^{-1}_{j} - \hat{D}^{-1}_{j}) S_j|_\infty \geq w),$$

where $\hat{D}^{-1}_{j}$ is the estimated $D^{-1}_{j}$. Hence, we have

$$\rho(\hat{D}^{-1}_{j} S_j, D^{-1}_{j} Z_j) \leq \rho(D^{-1}_{j} S_j, D^{-1}_{j} Z_j) + w \sqrt{\log K} + P(|(D^{-1}_{j} - \hat{D}^{-1}_{j}) S_j|_\infty \geq w).$$

Finally, the conditions in Theorem 3.2 of Zhang and Wu (2017a) can be verified for the $K$-dimensional stationary process $X_{j,t} \varepsilon_j, t$. Finally, applying that theorem yields the Gaussian approximation results. \qed
where the last line uses the arguments of Theorem 3 in Chernozhukov et al. (2015). Let $V_{n,j} \overset{\text{def}}{=} \max_{1 \leq k \leq K} |\Psi_{jk}/\hat{\Psi}_{jk} - 1|$ and $L_{n,j} \overset{\text{def}}{=} \max_{1 \leq k \leq K} |\hat{\Psi}_{jk}^2 - \hat{\Psi}_{jk}|$. Then $|(D_j^{-1} - \hat{D}_j^{-1})S_j|_\infty \leq V_{n,j}|D_j^{-1}S_j|_\infty$.

As $\min_{1 \leq k \leq K} \hat{\Psi}_{jk}^2 \geq c_j$, let $w = xy$, $0 < x < c_j/2$, $y > 0$, then

$$P(|(D_j^{-1} - \hat{D}_j^{-1})S_j|_\infty \geq w) \leq P(V_{n,j} \geq 2x/c_j) + P(|D_j^{-1}S_j|_\infty \geq c_j y/2)$$

$$\leq P(L_{n,j} \geq x) + \rho(D_j^{-1}S_j, D_j^{-1}Z_j) + P(|D_j^{-1}Z_j|_\infty \geq c_j y/2).$$

It follows that

$$\rho(D_j^{-1}S_j, D_j^{-1}Z_j) \leq \rho(D_j^{-1}S_j, D_j^{-1}Z_j) + x y \sqrt{\log K} + P(L_{n,j} \geq x) + P(|D_j^{-1}Z_j|_\infty \geq c_j y/2).$$

In particular, $L_{n,j} \leq L_{n,j,1} + L_{n,j,2}$, with $L_{n,j,1} = \max_{1 \leq k \leq K} |\Psi_{jk}^2 - E\hat{\Psi}_{jk}^2|$ and $L_{n,j,2} = \max_{1 \leq k \leq K} |E\hat{\Psi}_{jk}^2 - \hat{\Psi}_{jk}|$.

As for $L_{n,j,1}$, applying Theorem 5.1 of Zhang and Wu (2017a), for $u \geq n^{-1/2}b_{n,2\varsigma}^{1/2}$, we have

$$P(nL_{n,j,1} \geq u) \leq \frac{F_2^{2,\varsigma}(u)}{n^2b_{n,2\varsigma}^{1/2}} + K \exp\left(-\frac{Cn\mu^2}{\eta b_{n,\varsigma}^{1/2}}\right),$$

where the constants $C_j$ depend on $\eta$, $q$, and $c$. Then we have $P(L_{n,j,1} > x) \to 0$, as $n \to \infty$, if we set $x = \frac{\sqrt{\log K}}{n} \max\{n^{-1/2}b_{n,2\varsigma}^{1/2}, n^{-1/2}b_{n,2\varsigma}^{1/2} \sqrt{\log K} \Phi_{2,\varsigma}, F_{\varsigma}^{2/2,\varsigma} \Phi_{2,\varsigma}^{2/2,\varsigma}, \Phi_{2,\varsigma}v'(b_n)n/\sqrt{\log K}\}$.

For $L_{n,j,2}$, define $v'(b_n) = (b_n + 1)^{-\varsigma} + 2v_{n,2}/b_n$, $v_{n,2} = \log b_n$ (resp. $b_n^{-1+\varsigma}$ or $1$ for $\varsigma = 1$ (resp. $\varsigma < 1$ or $\varsigma > 1$). It can be shown that $L_{n,j,2} \leq \Phi_{2,0}\Phi_{2,0}v'(b_n)$. Note that $v'(b_n)$ is a special case of $v(b_n)$ in the proof of Theorem 5.3 given $n \to \infty$, and the conclusion follows similarly.

It follows that $P(L_{n,j} > x) \to 0$, as $n \to \infty$, if we set

$$x = \frac{\sqrt{\log K}}{n} \max\{n^{-1/2}b_{n,2\varsigma}^{1/2}, n^{-1/2}b_{n,2\varsigma}^{1/2} \sqrt{\log K} \Phi_{2,\varsigma}, F_{\varsigma}^{2/2,\varsigma} \Phi_{2,\varsigma}^{2/2,\varsigma}, \Phi_{2,\varsigma}v'(b_n)n/\sqrt{\log K}\}.$$

Moreover, given Theorem 5.2 and choosing $y = C\sqrt{\log K}$ (the constant $C > 0$ is sufficiently large) yields the conclusion.

**Proof of Corollary 5.4** Let $\tilde{\rho}_n \overset{\text{def}}{=} \rho(D_j^{-1}S_j, D_j^{-1}Z_j)$ and by its definition, we have

$$P(2c\sqrt{n} \max_{1 \leq k \leq K} |S_{jk}/\Psi_{jk}| \leq Z_\alpha) \geq P(2c\sqrt{n} \max_{1 \leq k \leq K} |Z_{jk}/\Psi_{jk}| \leq Z_\alpha) - \tilde{\rho}_n \geq 1 - \sum_{k=1}^{K} P(|Z_{jk}/\Psi_{jk}| \geq Z_\alpha/(2c\sqrt{n})) - \tilde{\rho}_n \geq 1 - \sum_{k=1}^{K} 2[Z_\alpha/(2c\sqrt{n})]^{-1} \exp[-Z_\alpha^2/(2(2c\sqrt{n})^2)] - \tilde{\rho}_n \geq 1 - 2(\log K)^{-1/2} - \tilde{\rho}_n,$$

where we have applied the union bound, the tail probability of Gaussian random variable and the fact that $Z_\alpha = 2c\sqrt{n}\log K \geq 2c\sqrt{n}\log K$ ($\tilde{c}$ is no less than the $c$ in the definition of $\lambda_0^j(1 - \alpha)$).

It follows that $\lambda_0^j(1 - \alpha) \leq Z_\alpha$ as $1 - \alpha = P(2c\sqrt{n} \max_{1 \leq k \leq K} |S_{jk}/\Psi_{jk}| \leq \lambda_0^j(1 - \alpha)) \leq$
\[ P(\sqrt{n} \max_{1 \leq k \leq K} |S_{jk}/\Psi_{jk}| \leq Z_\alpha), \] given \( 2(\log K)^{-1/2} + \tilde{\rho}_n = o(\alpha) \) (note that Theorem 5.2 ensures that \( \tilde{\rho}_n \to 0 \) with a polynomial rate as \( n \to \infty \)).

**Proof of Theorem 5.3.** Let \( S_{jk,i} = \frac{1}{n} \sum_{t=-(i-1)b_n+1}^{b_n} X_{jk,t} \varepsilon_{j,t} \), we first need to prove that

\[
\rho_{n,j} \equiv \sup_{r \in \mathbb{R}} \left| P \left\{ \max_{1 \leq k \leq K} (Z_{jk}^2 / \Psi_{jk}) \leq r \right| X_{j,.}, \varepsilon_{j,} \right] - P \left\{ \max_{1 \leq k \leq K} (\tilde{Z}_{jk} / \Psi_{jk}) \leq r \right\} \right| 
\]

\[
= \sup_{r \in \mathbb{R}} \left| P \left\{ \max_{1 \leq k \leq K} \left( \sum_{i=1}^{n} \varepsilon_{j,i} S_{jk,i} / \Psi_{jk} \right) \leq r \right| X_{j,.}, \varepsilon_{j,} \right] - P \left\{ \max_{1 \leq k \leq K} (\tilde{Z}_{jk} / \Psi_{jk}) \leq r \right\} \right| \to 0, \text{ as } n \to \infty.
\]

Given the sample variance covariance matrix \((K \times K) \Sigma_{j,n} = \sum_{t=-n}^{n} (1 - |\ell|/n) \Gamma_j(\ell)\), where \( \Gamma_j(\ell) = \mathbb{E}(X_{j,t} \varepsilon_{j,t} X_{j,\ell}^T \varepsilon_{j,\ell}) \), let \( \tilde{Z}_{jk} = Z_{jk}(K)_{k=1}^K \sim \text{N}(0, \Sigma_{j,n}) \). In addition, define \( \Sigma_{j,b_n} = \sum_{t=-b_n}^{b_n} (1 - |\ell|/b_n) \Gamma_j(\ell) \) and \( \tilde{\Sigma}_j = \sum_{t=1}^{b_n} S_{j,t} S_{j,t}^T \), where \( S_{j,t} = (S_{j,k,i})_{k=1}^K \). Let \( \Psi_j = \text{diag}(\Psi_{jk}) \),

\[
\delta_j = \delta_{j,1} + \delta_{j,2}, \text{ with } \delta_{j,1} = |\Psi_j^{-1} \tilde{\Sigma}_j - \Psi_j^{-1} \Sigma_{j,b_n} \Psi_j^{-1}|_{\max} \text{ and } \delta_{j,2} = |\Psi_j^{-1} \Sigma_{j,b_n} \Psi_j^{-1} - \Psi_j^{-1} \Sigma_{j,n} \Psi_j^{-1}|_{\max},
\]

where \(| \cdot |_{\max}\) is the maximum norm of a matrix. According to Theorem 2 of [Chernozhukov et al. (2015)], \( \rho_{n,j} \) is bounded by \( \pi(\delta_{j,1}) \vee \pi(\delta_{j,2}) \), with \( \pi(\delta_j) = C_3^{1/3} \{ 1 + a \delta_j \}^{1/3}(\log K)^{1/3} \), where \( a = \mathbb{E}(\max_{1 \leq k \leq K} Z_{jk}/\Psi_{jk}) \leq \sqrt{2} \log K \).

For the first part,

\[
\delta_{j,1} = \max_{1 \leq k_1, k_2 \leq K} \left| \sum_{t=1}^{b_n} S_{j,k_1,i} S_{j,k_2,i} - \frac{l_n}{\Psi_{jk_1} \Psi_{jk_2}} \mathbb{E}(S_{j,k_1,i} S_{j,k_2,i}) \right| \leq \max_{1 \leq k_1, k_2 \leq K} \left| \sum_{t=1}^{b_n} S_{j,k_1,i} S_{j,k_2,i} - \frac{l_n}{\Psi_{jk_1} \Psi_{jk_2}} \mathbb{E}(S_{j,k_1,i} S_{j,k_2,i}) \right| \cdot \min_{1 \leq k_1, k_2 \leq K} \Psi_{jk_1} \Psi_{jk_2}.
\]

We need to analyze the tail probability of \( \delta_{j,1} \). Applying Theorem 5.1 of [Zhang and Wu (2017a)], for \( x \geq n^{1/2} b_n^{1/2} \Psi_{j,2q_0} \), we have

\[
P \left( n \delta_{j,1} \geq \frac{x}{\min_{1 \leq k_1, k_2 \leq K} \Psi_{jk_1} \Psi_{jk_2}} \right) \leq \frac{K F \Gamma_j^{3/2} q_0}{x^{1/2}} + K^2 \exp \left( - \frac{C_j x^2}{n b_n^{3} \Psi_{j,8,8}} \right),
\]

for all large \( n \), where \( F = n, \) for \( \zeta > 1 - 2/q; \) \( F = l_n b_n^{q/2-\zeta} \), for \( 1/2 - 2/q < \zeta < 1 - 2/q; \)

\( F = l_n^{q/2-\zeta} b_n^{q/4-\zeta} \), for \( \zeta < 1/2 - 2/q \). The constants \( C_j \) depend on \( \eta, q, \) and \( \varsigma \). This ensures that when \( x = \max \{ n^{1/2} b_n^{1/2} \Psi_{j,2q_0}, n^{1/2} b_n^{1/2} (\log K)^{1/2} \Phi_{j,8,8} c_n, K^{2/3} l_n b_n^{2/3} \Gamma_j^{3/2} \}, c_n^{-1} = o(1), \) the tail probability tends to 0, as \( n \to \infty \).

It follows that \( \pi(\delta_{j,1}) \to 0 \) as \( n \to \infty \), given \( x = o(n(\log K)^{-2}) \), which implies the following conditions on \( b_n \):

\[ b_n = c \{ n(\log K)^{-q} \Psi_{j,2q_0}^{-4} \wedge n(\log K)^{-q} \Phi_{j,8,8}^{-4} c_n^{-2} \}, \quad F = c \{ n^{q/2} (\log K)^{-q} K^{-1} \Gamma_j^{3/2} q_0^{-2} \} \]

For the second part, by defining \( \psi_j \equiv \min_{1 \leq k_1, k_2 \leq K} \Psi_{jk_1} \Psi_{jk_2} \), we have

\[
\delta_{j,2} \leq \left| \psi_j^{-1} \sum_{b_n < |\ell| \leq n} (1 - |\ell|/n) \Gamma_j(\ell) + \sum_{|\ell| = -b_n}^{b_n} |\ell|(-1/n + 1/b_n) \Gamma_j(\ell) \right|_{\max}.
\]
Recall that
\[
|\Gamma_{j,k_1,k_2}(\ell)| = \sum_{h=0}^{\infty} \mathbb{E}\{(\mathcal{P}_h(X_{j,k_0}e_{j,0})\mathcal{P}_h(X_{j,k_2}e_{j,\ell}))\} \\
\leq \sum_{h=0}^{\infty} \|X_{j,k_1,h}e_{j,h} - X_{j,k_1,h}e_{j,h}^*\|_2^2 \|X_{j,k_2,h+\ell}e_{j,h+\ell} - X_{j,k_2,h+\ell}e_{j,h+\ell}^*\|_2^2,
\]
where the operator is given by \(\mathcal{P}_h(\cdot) \overset{\text{def}}{=} \mathbb{E}(\cdot|\mathcal{F}_h) - \mathbb{E}(\cdot|\mathcal{F}_{h-1})\). It follows that
\[
\left| \sum_{b_n < |\ell| \leq n} (1 - |\ell|/n)\Gamma_{j,k_1,k_2}(\ell) + \sum_{\ell = -b_n}^{b_n} |\ell|(-1/n + 1/b_n)\Gamma_{j,k_1,k_2}(\ell) \right| \\
\leq \Delta_{0,2,j,k_1}\Delta_{b_n+2,j,k_2} + \frac{2}{n}\Delta_{0,2,j,k_1} \sum_{\ell = b_n+1}^{n} \Delta_{\ell,2,j,k_2} + \frac{2n - b_n}{nb_n}\Delta_{0,2,j,k_1} \sum_{\ell = 1}^{b_n} \Delta_{\ell,2,j,k_2},
\]
where \(\Delta_{m,2,j,k} = \sum_{\ell = m}^{\infty} \|X_{j,k,\ell}e_{j,\ell} - X_{j,k,\ell}e_{j,\ell}^*\|_2^2\). Given the fact that \(\Delta_{0,2,j,k} \leq \Phi_{j,4,0,\epsilon}, \Delta_{\ell,2,j,k} \leq \Phi_{j,4,0,\epsilon}^{-\epsilon}\), \(\mathbf{B.1}\) is bounded by \(\Phi_{j,4,0,\epsilon} \mathbb{P}_{j,4,\epsilon}(\{b_n + 1\}^{-\epsilon} + 2n^{-1} \sum_{\ell = b_n+1}^{\infty} \ell^{-\epsilon} + 2\frac{n - b_n}{nb_n}\sum_{\ell = 1}^{b_n} \ell^{-\epsilon}) = \Phi_{j,4,0,\epsilon} \mathbb{P}_{j,4,\epsilon} v(b_n)\) for any \(k_1, k_2\), where \(v(b_n)\) is a function with respect to \(b_n\). Note that \(v(b_n) \lesssim (b_n + 1)^{-\epsilon} + 2n^{-1}/(n - b_n)\), where \(v_{n,1} = \mathbb{E}(n/(b_n + 1))\) and \(v_{n,2} = \mathbb{E}(n/(b_n))\) for \(\epsilon < 1\) and \(\epsilon > 1\). Therefore, the bound of \(\delta_{j,2}\) would decrease as \(b_n\) increases. In particular, we need to impose an addition assumption such that \(\Phi_{j,4,0,\epsilon} \mathbb{P}_{j,4,\epsilon} v(b_n) = o\{(\log K)^{-2}\}\) to guarantee \(\pi(\delta_{j,2}) \rightarrow 0\). The results for the two parts above ensure that \(\rho_{n,j} \rightarrow 0\) as \(n \rightarrow \infty\), given \(x = o\{(n\log K)^{-2}\}\) and \(\Phi_{j,4,0,\epsilon} \mathbb{P}_{j,4,\epsilon} v(b_n) = o\{(\log K)^{-2}\}\), which imply the following conditions on \(b_n\):
\[
\begin{align*}
\alpha & = o\{(n\log K)^{-4}\Phi_{j,2,q,\epsilon}^{-4} \wedge (n\log K)^{-5}\Phi_{j,8,\epsilon}^{-1}\}, \\
\mathbb{F}_{n} & = o\{n^{q/2}(\log K)^{-q} K^{-1}\mathbb{I}_{j,2,q,\epsilon} c_n^{-q/2}\}, \text{ with } c_n^{-1} = o(1). \\
\Phi_{j,4,0,\epsilon} \mathbb{P}_{j,4,\epsilon} \{b_{n+1} - \log(n/b_n)/(n + (n - b_n) \log b_n/(nb_n))\} & (\log K)^2 = o(1), \text{ if } \epsilon = 1; \\
\Phi_{j,4,0,\epsilon} \mathbb{P}_{j,4,\epsilon} \{b_{n+1} - n^{-\epsilon} + (n - b_n)b_n^{-\epsilon+1}/(nb_n)\} & (\log K)^2 = o(1), \text{ if } \epsilon < 1; \\
\Phi_{j,4,0,\epsilon} \mathbb{P}_{j,4,\epsilon} \{b_{n+1} - n^{-\epsilon}b_n^{-\epsilon+1} + (n - b_n)/(nb_n)\} & (\log K)^2 = o(1), \text{ if } \epsilon > 1.
\end{align*}
\]
At last, combining the Gaussian approximation results for \(S_{jk}/\Psi_{jk}\) and applying Theorem 3.1 in [Chernozhukov et al. (2013a)], we have
\[
\sup_{\alpha \in (0,1)} \left| \mathbb{P} \left( \max_{1 \leq k \leq K} |S_{jk}/\Psi_{jk}| \geq q_{j,(1 - \alpha)} \right) - \alpha \right| \lesssim \rho_{n,j} + \pi'(z) + \mathbb{P}(\delta_j \geq z),
\]
where \(\pi'(z) = z^{1/3}(1 \vee \log(K^{1/2}z))^{2/3}\). We need to pick \(z\) such that \(\pi'(z) + \mathbb{P}(\delta_j \geq z) \to 0\) as \(n \to \infty\) and it can be obtained by taking \(z = r_n^{1/2}/(\log K)\), with
\[
\begin{align*}
r_n & = n^{-1} \max\{n^{1/2}b_{n}^{1/2}\Phi_{j,2,q,\epsilon}^{2}, n^{1/2}b_{n}^{1/2}(\log K)^{1/2}\Phi_{j,8,\epsilon}^{2} c_{n}^{4}, K^{2/4}\mathbb{F}_{n}^{2} \mathbb{I}_{j,2,q,\epsilon} c_{n}, n\Phi_{j,2,0,\epsilon} \mathbb{P}_{j,2,\epsilon} v(b_n)\}, \\
c_n^{-1} & \overset{\text{def}}{=} o(1).
\end{align*}
\]
\hfill \Box

\textbf{Comment B.1} (Admissible rate of \(b_n\)). Consider the special case with \(\Phi_{j,2,q,\epsilon} = O(1)\) and
\[ \Gamma_{j2q,\xi} = \mathcal{O}(1), \text{ for } q > 4. \] Let \( K = \mathcal{O}(n^r), \log K = \mathcal{O}(r \log n) = \mathcal{O}(n^r), c_n = n^s \text{ with } s > 0 \text{ and assume } 1/2 - 2/q < \xi < 1 - 2/q. \) Then (B.2) implies an admissible rate of \( b_n = \mathcal{O}(n^\eta) \) such that \( 2\bar{\xi}/\xi < \eta < \max\{1 - 5\bar{\xi}/2 \nu, (q/2 - q\bar{\xi} - q\bar{s}/2 - r - 1)/(q/2 - \xi q/2 - 1)\}. \)

**Comment B.2** (Validity of multiplier block bootstrap under stronger tail assumptions). Note that in case with stronger exponential moment conditions on the underlying processes, we shall change the tail probabilities to bound \( \delta_{j1}. \)

Let \( \Phi_{j,\psi,\xi} = \max_{1 \leq k \leq p \geq 2} q^{-\nu} |X_{jk}, \xi_j|, q, \xi < \infty, \) then according to Theorem 5.2 of Zhang and Wu (2017a), for all \( x > 0, \) we have

\[
P(n\delta_{j1} \geq \min_{1 \leq k_1, k_2 \leq K} \frac{x}{\Psi_{j,k_1}^{(1)} \Psi_{j,k_2}^{(1)}}) \lesssim K^2 \exp \left( - \frac{x^{\gamma}}{4c\gamma n^2 \Phi_{j,\psi,\xi}} \right),
\]

where \( \gamma = 1/(2\nu + 1). \) This implies that when \( x = (\log K^{\nu} b_n \Phi_{j,\psi,\xi}^{(1)})^{1/\gamma} c_n, \) with \( c_n^{-1} = o(1), \) the tail probability tends to 0, as \( n \to \infty. \) It follows that \( \pi(\delta_{j1}) \to 0 \) as \( n \to \infty, \) given \( x = (\log K^{\nu} b_n \Phi_{j,\psi,\xi}^{(1)})^{1/\gamma} = o\{n(\log K)^{-2}\}. \) As a result, (B.2) will be replaced by

\[
b_n = o\{n^{\gamma-1/2}(\log K)^{-2}\gamma - 1/2 \Phi_{j,\psi,\xi}^{(1)} c_n^{-1}\}, \text{ with } c_n^{-1} = o(1).
\]

\[
\Phi_{j,4.0} \Phi_{j,4.3} \{b_n^{-1} + \log(n/b_n)/n + (n - b_n) \log b_n/(nb_n)\} (\log K)^2 = o(1), \text{ if } \xi = 1;
\]

\[
\Phi_{j,4.0} \Phi_{j,4.3} \{b_n^{-1} + n^{-\xi} + (n - b_n)b_n^{-1}/(nb_n)\} (\log K)^2 = o(1), \text{ if } \xi < 1;
\]

\[
\Phi_{j,4.0} \Phi_{j,4.3} \{b_n^{-1} + n^{-\xi} + b_n^{-1}/(nb_n)\} (\log K)^2 = o(1), \text{ if } \xi > 1.
\]

**B.2 Proofs of Joint Equation Estimation**

**Proof of Theorem A.3**. Analogue to the proof of Theorem 5.3, the conclusions are implied by

\[
P \left( n\delta_{1} \geq \left( \min_{1 \leq k_1, k_2 \leq K, 1 \leq j_1, j_2 \leq J} \Psi_{j_1,k_1}^{2} \Psi_{j_2,k_2}^{2} \right)^{-1} x \right) \lesssim \frac{JKF_{1,2}^{1/2}}{\gamma \eta} + (JK)^{1/2} \exp \left( - \frac{C x^2}{nb_n \Phi_{1,2}^{2}} \right),
\]

for \( x \geq n^{1/2} b_n^{1/2} \Phi_{2q,\xi}^{2} \) and all large \( n, \) where

\[
\delta_{1} \overset{\text{def}}{=} \max_{1 \leq k_1, k_2 \leq K, 1 \leq j_1, j_2 \leq J} \left| \sum_{i=1}^{n} S_{j_1,k_1,i} S_{j_2,k_2,i} - \frac{E(S_{j_1,k_1,i} S_{j_2,k_2,i})}{\Psi_{j_1,k_1} \Psi_{j_2,k_2}} \right|.
\]

In particular, when \( x = \max\{n^{1/2} b_n^{1/2} \Phi_{2q,\xi}^{2}, n^{1/2} b_n^{1/2} \{\log(JK)\}^{1/2} \Phi_{1,2}^{2} c_n, (JK)^{2/4} F_{2q,\xi}^{2} \}, \) \( c_n^{-1} = o(1), \) the tail probability tends to 0, as \( n \to \infty. \)

By similar proof to that of Theorem 5.3, it follows that \( \hat{\rho}_n \to 0 \) as \( n \to \infty, \) given \( x = o\{n(\log(JK)\}^{-2}\} \) and \( \Phi_{4.0} \Phi_{4.3} \nu(b_n) = o\{(\log JK)^{-2}\}, \) which imply the following conditions on
Recall that \( F_\varsigma = n \), for \( \varsigma > 1 - 2/q \); \( F_\varsigma = l_n^q/n^{q/2} - q/2 \), for \( 1/2 - 2/q < \varsigma < 1 - 2/q \); \( F_\varsigma = l_n^{q/2 - q/2}/n^{q/2} - q/2 \), for \( \varsigma < 1/2 - 2/q \).

The rest of the proof is similar to that of Theorem 5.3 and thus is omitted.

**Proof of Theorem A.4.** For any \( \delta, \delta \in \mathbb{R}^K \) in \( G_{\tilde{T}_j} \), we have

\[
\mathbb{E}_n \left\{ \varepsilon_{j,t} \left( \frac{X_{j,t}^\top \delta}{|\delta|_{j,pr}} - \frac{X_{j,t}^\top \delta}{|\delta|_{j,pr}} \right) \right\} = \mathbb{E}_n \left\{ \varepsilon_{j,t} \left( \frac{X_{j,t}^\top (\delta - \tilde{\delta})}{|\delta|_{j,pr}} + \frac{X_{j,t}^\top \tilde{\delta}}{|\delta|_{j,pr}} - \frac{X_{j,t}^\top \delta}{|\delta|_{j,pr}} \right) \right\} \\
\leq \mathbb{E}_n \left\{ \varepsilon_{j,t} \right\} \left| \mathbb{E}_n \left\{ \frac{X_{j,t}^\top (\delta - \tilde{\delta})}{|\delta|_{j,pr}} \right\} \right| + \mathbb{E}_n \left\{ \varepsilon_{j,t} \right\} \left| \mathbb{E}_n \left\{ \frac{X_{j,t}^\top \tilde{\delta}}{|\delta|_{j,pr}} - \frac{X_{j,t}^\top \delta}{|\delta|_{j,pr}} \right\} \right| \\
\leq (\mathbb{E}_n \varepsilon_{j,t}^2)^{1/2} \left( \mathbb{E}_n \left| \frac{X_{j,t}^\top (\delta - \tilde{\delta})}{|\delta|_{j,pr}} \right|^2 \right)^{1/2} + (\mathbb{E}_n \varepsilon_{j,t}^2)^{1/2} \left( \frac{|\delta|_{j,pr} - |\delta|_{j,pr}}{\delta_{j,pr}} \right) \leq 2\sigma \mu_j(p)|\delta - \tilde{\delta}|_2.
\]

Then by following the proof of Lemma 5 (Step 2) in Belloni and Chernozhukov (2013), we have

\[
\sup_{Q} \mathcal{N}(\varepsilon, G_{\tilde{T}_j} \cdot \| \cdot \|_{Q,1}) \lesssim (6\mu_j(p)\sigma/\varepsilon)^{s+p}.
\]

And it follows that \( |F_{j,p}| \lesssim \left( \frac{K}{p} \right) (6\mu_j(p)\sigma/\varepsilon)^{s+p}. \)

Moreover, it is not hard to see that \( \sup_{f \in \mathcal{F}_{j,p}} |G_n(f)| \leq 2\sqrt{n}\varepsilon + \sup_{f \in \mathcal{F}_{j,p}} |G_n(f)|. \) Let \( \psi = \max_{f \in \mathcal{F}_{j,p}} \psi_f \) (assume \( \psi \) is bounded by constant) and applying the Gaussian approximation results on the vector \( G_n(f)/\psi_f \) (given (A6)), we have

\[
P \left\{ \sup_{f \in \mathcal{F}_{j,p}} |G_n(f)| \geq \kappa_n/2 \right\} \leq P \left\{ \sup_{f \in \mathcal{F}_{j,p}} |G_n(f)/\psi_f| \geq \kappa_n/(2\psi) \right\} \\
\leq 2|F_{j,p}| \{1 - \Phi(\kappa_n/(2\psi))\} + d_n \\
\leq 2K^p (6\mu_j(p)\sigma/\varepsilon)^{s+p} \exp\{-(\kappa_n^2/(8\psi^2))\} \{\kappa_n/(2\psi)\}^{-1} + d_n,
\]

as \( \left( \frac{K}{p} \right) \leq K^p. \) Therefore, for \( \kappa_n = \psi \sqrt{p \log K + (p + s) \log(6\mu_j(p)\sigma) + \log n}/2 \) and \( \varepsilon = \sqrt{p \log K + (p + s) \log(6\mu_j(p)\sigma)(4\sqrt{n})^{-1}} \), it follows that \( \sup_{f \in \mathcal{F}_{j,p}} |G_n(f)| \lesssim \kappa_n \) (note that \( d_n \to 0 \) with a polynomial rate as \( n \to \infty \)).

The rest of the proof is a direct application of Theorem 5 of Belloni and Chernozhukov (2013) by inserting the bound for \( \lambda^0(1 - \alpha) \) \( \{A.2\} \) provided in Corollary A.1 and thus is omitted.

**B.3 Plausibility of RE and RSE Conditions**

Define the \( s \)-sparse sphere as \( F_{\delta} = \{ \delta : |\delta|_0 \leq s, |\delta|_2 = 1 \} \). According to Rudelson and Zhou (2012), the \( \epsilon \)-covering number of \( F_{\delta} \) w.r.t. the Euclidean metric is \( l = \exp(s \log(3eK/m\epsilon)) \), with
$m \geq 1$. This is the cardinality of the $\epsilon$-cover set $\Pi_\delta$ of $F_\delta$. Moreover, for any point $\delta \in F_\delta$, let $\pi_\delta$ denote the closest point to $\delta$ within $\Pi_\delta$. Let $\tilde{X}_j(\delta) \overset{\text{def}}{=} (\tilde{X}_j(X_j^\top \pi(\delta)))^2 - n^{-1} \pi(\delta)^\top \mathbb{E}\{X_j X_j^\top\}\pi(\delta)$, where $\tilde{X}_j \overset{\text{def}}{=} n^{-1/2}X_j$ and $X_j(n \times K)$ is a matrix of $X_j$. Note that $\tilde{X}_j(\delta)$ is a vector of the cardinality of $\Pi_\delta$.

**Theorem B.1** (Plausibility of RE and RSE). For any $j = 1, \ldots, J$, suppose the vectors $X_{j,t}$ of length $K$ satisfy

$$0 < \kappa \leq \min_{|\delta| \leq s, |\delta|^2 = 1} \delta^T \mathbb{E}(X_{j,t} X_{j,t}^\top) \delta \leq \max_{|\delta| \leq s, |\delta|^2 = 1} \delta^T \mathbb{E}(X_{j,t} X_{j,t}^\top) \delta < \psi < \infty,$$

where $\psi$ and $\kappa$ are positive constants. Given $\Phi_{2,\kappa} \overset{\text{def}}{=} \max_{\pi(\delta) \in \Pi_\delta} \| \tilde{X}_j(\delta) \|^2_{2,\kappa} < \infty$, and for $q > 2$,

$$\max_{\pi(\delta) \in \Pi_\delta} \| \tilde{X}_j(\delta) \| q,\kappa < \infty,$$

$$n^{-1/2} (\log l)^{1/2} \Phi_{2,\kappa} + n^{-1} r_c (\log l)^{3/2} \max_{\pi(\delta) \in \Pi_\delta} \| \tilde{X}_j(\delta) \| q,\kappa = o(1),$$

where $r_c = n^{1/4}$ for $\zeta > 1/2 - 1/q$ and $r_c = n^{1/2 - \zeta}$ for $\zeta < 1/2 - 1/q$, then the RE and RSE conditions hold with probability $1 - o(1)$, with $p + s_j \leq s$.

**Proof of Theorem B.1**

Firstly, we need to check the implication of the population matrix. We know that $\delta^T X_j^T X_j \delta / n = |\tilde{X}_j|_2^2$. Then we have the following inequalities for any point $\delta \in F_\delta$,

$$-|\tilde{X}_j(\delta - \pi(\delta))|_2 + |\tilde{X}_j\pi(\delta)|_2 \leq |\tilde{X}_j|_2 \leq |\tilde{X}_j(\delta - \pi(\delta))|_2 + |\tilde{X}_j\pi(\delta)|_2. \quad (B.4)$$

We first check the right hand side of (B.4). Define $\| \tilde{X}_j \|_{2,F_\delta} \overset{\text{def}}{=} \sup_{\delta \in F_\delta} |\tilde{X}_j|_2$. As indicated in the proof of Theorem 16 in [[Rudelson and Zhou 2012]]^1, we have $|\tilde{X}_j(\delta - \pi(\delta))|_2 \leq \epsilon \| \tilde{X}_j \|_{2,F_\delta}$. To bound $\max_{\pi(\delta) \in \Pi_\delta} |\tilde{X}_j\pi(\delta)|_2$, we invoke the tail probability inequality in Lemma B.2, which gives

$$\mathbb{P} \left( \max_{\pi(\delta) \in \Pi_\delta} \left| \sum_{t=1}^n \tilde{X}_j(\delta) \right| \geq x \right) = \mathbb{P} \left[ \max_{\pi(\delta) \in \Pi_\delta} \left| \tilde{X}_j\pi(\delta) \right|_2 - \pi(\delta)^\top \mathbb{E}\{X_j X_j^\top\}\pi(\delta) \right] \geq x \right) \rightarrow 0, \text{ as } n \rightarrow \infty,$$

if $x \geq \sqrt{n \log l} \Phi_{2,\kappa} + r_c (\log l)^{3/2} \max_{\pi(\delta) \in \Pi_\delta} |\tilde{X}_j(\delta)|_q,\kappa.$

Therefore, given $\kappa, \psi > 0$, $\kappa - x_n \leq |\tilde{X}_j\pi(\delta)|_2 \leq x_n + \psi$ holds with probability $1 - o(1)$ for all $\pi(\delta) \in \Pi_\delta$, where $x_n \overset{\text{def}}{=} \sqrt{n \log l} \Phi_{2,\kappa} + r_c (\log l)^{3/2} \max_{\pi(\delta) \in \Pi_\delta} |\tilde{X}_j(\delta)|_q,\kappa$. In particular, the assumption

$$n^{-1/2} (\log l)^{1/2} \Phi_{2,\kappa} + n^{-1} r_c (\log l)^{3/2} \max_{\pi(\delta) \in \Pi_\delta} |\tilde{X}_j(\delta)|_q,\kappa = o(1)$$

ensures that $x_n = o(1)$.

Hence, the right inequality in (B.4) leads to $|\tilde{X}_j|_2 \leq \epsilon \| \tilde{X}_j \|_{2,F_\delta} + \sqrt{x_n} + \sqrt{\psi}$. Taking the supremum over all $\delta \in F_\delta$ on both sides shows that $\sup_{\delta \in F_\delta} |\tilde{X}_j|_2 \leq (\sqrt{x_n} + \sqrt{\psi})/(1 - \epsilon)$ with probability $1 - o(1)$. Moreover, by the left hand side of (B.4), we have $|\tilde{X}_j|_2 \geq \sqrt{n - x_n - \epsilon(\sqrt{x_n} + \sqrt{\psi})/(1 - \epsilon)}$, with probability $1 - o(1).
Collecting the results together, we have shown that for all \( \delta \in F_\delta \),
\[
\sqrt{n} - x_n - \frac{\epsilon(\sqrt{x_n + \sqrt{\psi}})}{(1 - \epsilon)} \leq |\bar{X}_j\delta|_2 \leq \frac{\sqrt{x_n + \sqrt{\psi}}}{(1 - \epsilon)},
\]
with probability \( 1 - o(1) \).

Let \( c^*(s) = \max_{\delta \in F_\delta} |\bar{X}_j\delta|_2, c_*(s) = \min_{\delta \in F_\delta} |\bar{X}_j\delta|_2 \), with properly chosen \( \epsilon \), \( c^*(s) \), \( c_*(s) \) are bounded from above and below, and the desired results follow by the fact \( \kappa_j^2(p) \geq c_*(s_j + p) \), \( \phi_j(p) \leq c^*(s_j + p) \), with \( s_j + p \leq s \).

\[ \square \]

### B.4 Proofs of Simultaneous Inference

#### B.4.1 Some Useful Lemmas

**Lemma B.1** (Burkholder (1988)). Let \( q > 1, q' = \min(q, 2) \). Let \( M_n = \sum_{t=1}^{n} \xi_t \) where \( \xi_t \in \mathcal{L}^q \) (i.e., \( \|\xi_t\|_q < \infty \)) are martingale differences. Then
\[
\|M_n\|_q^q \leq K_q^2 \sum_{t=1}^{n} \|\xi_t\|_q^q \quad \text{where} \quad K_q = \max((q - 1)^{-1}, \sqrt{q - 1}).
\]

**Lemma B.2** (Theorem 6.2 of Zhang and Wu (2017a) Tail probabilities for high dimensional partial sums). For a zero-mean \( p \)-dimensional random variable \( X_t \in \mathbb{R}^p \), let \( S_n = \sum_{t=1}^{n} X_t \) and assume that \( \|X_t\|_\infty \|q, < \infty \) where \( q > 2 \) and \( \varsigma \geq 0 \), and \( \Phi_{2,\varsigma} = \max_{1 \leq j \leq p} \|X_j\|_{2,\varsigma} < \infty \).

i) If \( \varsigma > 1/2 - 1/q \), then for \( x \geq \sqrt{n \log p} \Phi_{2,\varsigma} + n^{1/q} (\log p)^{3/2} \|X_t\|_{\infty} \|q, \),
\[
P(|S_n|_{\infty} \geq x) \leq \frac{C_{q,\varsigma} n (\log p)^{q/2} \|X_t\|_{\infty}^{q/2} }{x^{q/2}} + C_{q,\varsigma} \exp \left( \frac{-C_{q,\varsigma} x^2}{n \Phi_{2,\varsigma}^2} \right).
\]

ii) If \( 0 < \varsigma < 1/2 - 1/q \), then for \( x \geq \sqrt{n \log p} \Phi_{2,\varsigma} + n^{1/2-\varsigma} (\log p)^{3/2} \|X_t\|_{\infty} \|q, \),
\[
P(|S_n|_{\infty} \geq x) \leq \frac{C_{q,\varsigma} n^{q/2-\varsigma} (\log p)^{q/2} \|X_t\|_{\infty}^{q/2} }{x^{q/2}} + C_{q,\varsigma} \exp \left( \frac{-C_{q,\varsigma} x^2}{n \Phi_{2,\varsigma}^2} \right).
\]

**Lemma B.3** (Tail probabilities for high dimensional partial sums with strong tail assumptions). For a zero-mean \( p \)-dimensional random variable \( X_t \in \mathbb{R}^p \), let \( S_n = \sum_{t=1}^{n} X_t \) and assume that \( \Phi_{\psi,\varsigma} = \max_{1 \leq j \leq p} \sup_{q \geq 2} q^{-\nu} \|X_j\|_{q,\varsigma} < \infty \) for some \( \nu \geq 0 \), and let \( \gamma = 2/(1+2\nu) \). Then for all \( x > 0 \), we have
\[
P(|S_n|_{\infty} \geq x) \leq p \exp \left\{ -C_{\gamma,x^\gamma}/(\sqrt{n} \Phi_{\psi,0})^\gamma \right\}.
\]

**Lemma B.4** (Theorem 1 of El Machkouri et al. (2013)). Denote \( Y_t = f(F_t) \), where \( f \) is some measurable function. Let \( S_n = \sum_{t=1}^{n} Y_t \), and \( \delta_{\varsigma,t} = \|Y_t - Y^*_t\|_\varsigma \). If \( \mathbb{E}(Y_t) = 0 \), \( \sum_{t=0}^{\infty} \delta_{\varsigma,t} < \infty \), some \( \varsigma \geq 2 \), and \( \sigma_n^2 \overset{\text{def}}{=} \mathbb{E}(S_n^2) \rightarrow \infty \), then
\[
\sigma_n^{-1} S_n \overset{\mathcal{D}}{\rightarrow} \mathcal{N}(0,1).
\]

**Lemma B.5.** Under the same conditions as in Theorem 5.2, let \( \hat{\beta}_{jk} \) be any estimator such that
\[ |\hat{\beta}_{jk} - \beta_{jk}^0| \leq C\rho_n \text{ with probability } 1 - o(1). \] Then we have
\[ n^{-1} \max_{(j,k) \in G} \Delta_n \lesssim o(n^{-1/2}g_n^{-1}), \quad (B.6) \]
holds with probability \(1 - o(1)\), where \(\Delta_n \overset{\text{def}}{=} n^{1/2}G_n \{ \psi_{jk}(Z_{j,t}, \hat{\beta}_{jk}, \hat{h}_{jk}) - \psi_{jk}(Z_{j,t}, \beta_{jk}^0, h_{jk}^0) \} \).

**Proof of Lemma B.5**

As indicated in the proof of Theorem 2 in [Belloni et al. (2015b)], the entropy \( \text{ent}(\epsilon, \tilde{F}) \leq cs \log(an/\epsilon) \) for the function class \( \tilde{F} = \{ z \mapsto \psi_{jk}\{ z, \beta, \tilde{h}(x_{j(-k)}) \} - \psi_{jk}\{ z, \beta_{jk}^0, h_{jk}^0(x_{j(-k)}) \} : (j, k) \in G, \beta \in B_{jk}, |\beta - \beta_{jk}^0| \leq C\rho_n, \tilde{h} \in \mathcal{H}_{jk} \} \), which has \( 2F \) as the envelope (the definition of \( F \) is given in (C6)). Therefore, for any \( f \in \tilde{F} \), there exists a set \( F_n \) such that \( \min_{f \in F_n} \| f - f' \|_{Q,2} \leq \tilde{\epsilon} \), where \( \tilde{\epsilon} \overset{\text{def}}{=} \epsilon|2F|_{Q,2} \) and the cardinality of the set \( |F_n| = (an/\epsilon)^c_\nu \). Then we have
\[ \sup_{f \in \tilde{F}} \left| \frac{1}{n} \sum_{t=1}^{n} [ f - \pi(f) - E(f - \pi(f)) ] \right| \leq 2\tilde{\epsilon}n, \]
where \( \pi(f) \overset{\text{def}}{=} \arg \min_{f' \in F_n} \| f - f' \|_{Q,2} \). Hence, with probability \(1 - o(1)\),
\[ \max_{(j,k) \in G} \Delta_n \leq n^{1/2} \sup_{f \in \tilde{F}} |G_n(f)| \]
\[ = n \sup_{f \in \tilde{F}} \left| [E_n(f) - E_n\{ \pi(f) \} - E(f) + E\{ \pi(f) \} ] + [E_n\{ \pi(f) \} - E(\pi(f)) ] \right| \]
\[ \leq 2\tilde{n} + n \max_{f \in F_n} |E_n(f) - E(f)| \]
\[ \leq 2\tilde{n} + n \max_{f \in F_n} |E_n(f) - E_n E(f|\mathcal{F}_{t-1}, X_{j(-k),t})| + n \max_{f \in F_n} |E_n E(f|\mathcal{F}_{t-1}, X_{j(-k),t}) - E(f)| \]
\[ = 2\tilde{n} + K_n + N_n \quad (B.7) \]

Next, we look for the bounds for \( K_n \) and \( N_n \), respectively. Note the summation of \( K_n \) form martingale differences. Consider the function set \( F_n \), for each \( f \in F_n \), let \( \varphi_{l,t} \overset{\text{def}}{=} f(z_l) \) and \( \tilde{\varphi}_{l,t} \overset{\text{def}}{=} \varphi_{l,t} - E(\varphi_{l,t}|\mathcal{F}_{t-1}, X_{j(-k),t}) \). Note that \( \varphi_{l,t} \) and \( \tilde{\varphi}_{l,t} \) are vectors of length \( |F_n| = (an/\epsilon)^c_\nu \). For \( l = 1, \ldots, |F_n| \), the dependence adjusted norm of \( \tilde{\varphi}_{l,t} \) obeys that \( \| \tilde{\varphi}_{l,t} \|_{2,\kappa} \leq 2\| \tilde{\varphi}_{l,t} \|_2 \lesssim 8\| \varphi_{l,t} \|_2 \). Moreover, we have \( \| \varphi_{l,t} \|_2 \lesssim L_{2n}^{2}\rho_{n,v}^{2} \). In particular, for the mean regression case \( \rho_{n,v} = \rho_n s \), while \( \rho_{n,v} = \rho_n^{1/2} \) for the median regression case (by (C5)).

Apply the tail inequality as in Lemma B.2 to the vector \( \tilde{\varphi}_{l,t} \). As \( \max_{1 \leq i \leq |F_n|} \| \tilde{\varphi}_{l,t} \|_{2,\kappa} \lesssim \sqrt{L_{2n}\rho_{n,v}} \) and \( \max_{1 \leq i \leq |F_n|} \tilde{\varphi}_{l,t} \|_{q,\kappa} \lesssim \| 4F(z_l) \|_q \) (by (C6)), then we can see that with probability greater than \( 1 - O(|F_n|^{-1} + (\log |F_n|)^{-q}) \),
\[ K_n \lesssim \sqrt{ns \log(an/\epsilon)} \max_{1 \leq i \leq |F_n|} \| \tilde{\varphi}_{l,t} \|_{2,\kappa} + r_\kappa \{ s \log(an/\epsilon) \}^{3/2} \max_{1 \leq i \leq |F_n|} \tilde{\varphi}_{l,t} \|_{q,\kappa} \]
\[ \lesssim \sqrt{nL_{2n}s \log(an/\epsilon)\rho_n s} + r_\kappa \{ s \log(an/\epsilon) \}^{3/2} \| 4F(z_l) \|_q, \]
Hence, we have
\[ K_n \lesssim \rho K_n, \quad (B.8) \]
where \( \rho_{k_n} \overset{\text{def}}{=} r_{k_1} + r_{k_2} \) with \( r_{k_1} \overset{\text{def}}{=} \sqrt{n L_{2n} \log(a_n/\epsilon)\rho_{n,v}} \), \( r_{k_2} \overset{\text{def}}{=} \{ s \log(a_n/\epsilon) \}^{3/2} ||F(z_1)||_q \) and \( r_\epsilon \overset{\text{def}}{=} n^{1/q} \) for \( \epsilon > 1/2 - 1/q \) and \( r_\epsilon \overset{\text{def}}{=} n^{1/2-\epsilon} \) for \( \epsilon < 1/2 - 1/q \).

Then we handle the term \( N_n \). Again consider the function set \( F_n \), for each \( f \in F_n \), let \( \tilde{\varphi}_{l,t} \overset{\text{def}}{=} E(\varphi_{l,t}|F_{l-1}, X_{j(k),l}) - E(\tilde{\varphi}_{l,t}) \), where \( \varphi_{l,t} = f(z_1) \). Then

\[
N_n \leq \max_{1 \leq \ell \leq |F_n|} \left| \sum_{l=1}^{n} \tilde{\varphi}_{l,t} \right|.
\]

Moreover, for \( l = 1, \ldots, |F_n| \), there is a function \( g \) corresponding to each \( f \in F_n \) such that \( \tilde{\varphi}_{l,t} = g(z_l, \beta, \tilde{h}) \), where \( \beta \in B_{jk}, |\beta - \beta_{jk}^0| \leq C \rho_n, \tilde{h} \in H_{jk}, (j, k) \in G \). By the mean value theorem and the continuity of the function \( g \), we have

\[
g(Z_{j,t}, \beta, \tilde{h}) = \partial_\beta g(Z_{j,t}, \beta, \tilde{h})(\beta - \beta_{jk}^0)
\]

\[
+ \sum_{m=1}^{2} \partial_{hm} g(Z_{j,t}, \beta, \tilde{h}) \{ \tilde{h}_m(X_{j(k),l}) - h_{jk,m}^0(X_{j(k),l}) \},
\]

where \( (\beta, \tilde{h}(\cdot)) \) is the corresponding point which joins the line segment between \( (\beta, \tilde{h}(\cdot)) \) and \( (\beta_{jk}^0, h_{jk,m}^0(\cdot)) \). Then

\[
\max_{1 \leq \ell \leq |F_n|} \sum_{l=1}^{n} \tilde{\varphi}_{l,t} = \max_{\beta \in F_n} \sum_{l=1}^{n} \partial_\beta g(Z_{j,t}, \beta, \tilde{h})(\beta - \beta_{jk}^0)
\]

\[
+ \sum_{\tilde{h}_m \in F_n^h} \sum_{m=1}^{2} \partial_{hm} g(Z_{j,t}, \beta, \tilde{h}) \{ \tilde{h}_m(X_{j(k),l}) - h_{jk,m}^0(X_{j(k),l}) \},
\]

where \( F_n^\beta \) and \( F_n^h \) collect all the points of \( \beta \) and \( \tilde{h} \) according to \( F_n \), respectively.

Recall that in our linear model setting, \( h_{jk,m}^0(X_{j(k),l}) = (X_{j(k),l}^\top, \gamma_{jk,m}^0(X_{j(k),l})^\top) = (X_{j(k),l}^\top, X_{j(k),l}^\top, \gamma_{jk,m}^0(X_{j(k),l})^\top) \), and \( \tilde{h}(X_{j(k),l}) = (X_{j(k),l}^\top, \tilde{\gamma}_{jk,m}^0(X_{j(k),l})^\top) \), where \( \gamma_{jk,m}^0 \) and \( \tilde{\gamma}_{jk,m}^0 \) (\( m = 1, 2 \)) are vectors of length \( K - 1 \). Let \( T_{j,k}^0 \overset{\text{def}}{=} \{ 1 \leq \ell \leq K - 1 : \theta_{jk,1,\ell}^0 \neq 0, \theta_{jk,2,\ell}^0 \neq 0 \} \), \( \tilde{T}_{j,k}^0 \overset{\text{def}}{=} \{ 1 \leq \ell \leq K - 1 : \tilde{\theta}_{jk,1,\ell}^0 \neq 0, \tilde{\theta}_{jk,2,\ell}^0 \neq 0 \} \), and \( X_{jk,m}^i \overset{\text{def}}{=} \text{vec}\{ (X_{j(k),l})_{\ell \in T_{j,k}^0} \cup \tilde{T}_{j,k}^0 \} \). Now we apply Lemma B.2 on \( \sum_{m=1}^{n} \partial_{hm} g(Z_{j,t}, \beta, \tilde{h}) \{ \tilde{h}_m(X_{j(k),l}) - h_{jk,m}^0(X_{j(k),l}) \} \) and \( \sum_{l=1}^{n} \partial_\beta g(Z_{j,t}, \beta, \tilde{h})(\beta - \beta_{jk}^0) \).

To this end, we define the following quantities:

\( \Phi_{m,2,\kappa}^h \overset{\text{def}}{=} \max_{h \in F_n^h} \left\| \tilde{\phi}_{jk} \partial_{hm} g(Z_{j,t}, \beta, \tilde{h}) \right\|_{q, \kappa} \),

\( \Omega_{m,2,\kappa}^h \overset{\text{def}}{=} \max_{h \in F_n^h} \left\| \tilde{\phi}_{jk} \partial_{hm} g(Z_{j,t}, \beta, \tilde{h}) \right\|_{q, \kappa} \).

Let \( \chi_{j,m}^i \overset{\text{def}}{=} \partial_{hm} g(Z_{j,t}, \beta, \tilde{h}) \{ \tilde{h}_m(X_{j(k),l}) - h_{jk,m}^0(X_{j(k),l}) \} \) and define the projector operator \( P_j(\chi_{j,m}^i) \overset{\text{def}}{=} E(\chi_{j,m}^i|F) - E(\chi_{j,m}^i|F_{j-1}) \). According to Theorem 1(i) of \cite{WEI20051}, it is not hard to see that \( \left\| \chi_{j,m}^i \right\|_{q, \kappa} \lesssim \sup_{m=0}^{d+1} \sum_{i=1}^{\infty} \left\| P_j(\chi_{j,m}^i) \right\|_q \), for \( m = 1, 2 \). Moreover, as \( |\tilde{\theta}_{jk,m} - \theta_{jk,m}^0| \lesssim \sqrt{s} \rho_n \), we have

\[
\left\| P_j(\chi_{j,m}^i) \right\|_q \leq \left( E[P_j(\partial_{hm} g(Z_{j,t}, \beta, \tilde{h}) \tilde{\phi}_{jk} | \kappa) | \tilde{\theta}_{jk,m} - \theta_{jk,m}^0(\varnothing) \right)^{1/q} \]

\[
\lesssim \sqrt{s} \rho_n \left( E[P_j(\partial_{hm} g(Z_{j,t}, \beta, \tilde{h}) \tilde{\phi}_{jk} | \kappa) \right)^{1/q}.
\]
It follows that \( \|\chi^m\|_{q,\kappa} \lesssim \sqrt{s}\rho_n \|\hat{X}^k|_{\infty}\|\partial_{h_m}g(Z_{j,:},\beta,\tilde{h})\|_{q,\kappa} \). Then applying the tail probability bounds in Lemma B.2 yields with probability approaching 1,

\[
\max_{h \in F_n^m} \left\{ \frac{1}{n} \sum_{t=1}^{n} \partial_{h_m}g(Z_{j,t},\beta,\tilde{h})\{h_m(X_{j-(k),t}) - h_{jk,m}(X_{j-(k),t})\} \right\} \lesssim r_{N1,m} + r_{\kappa}r_{N2,m},
\]

where \( r_{N1,m} = \sqrt{n}s\rho_n \{\log(an)/\epsilon\}^{1/2} \Phi_{m,2,\kappa}^h, \) \( r_{N2,m} = s^2\rho_n \{\log(an)/\epsilon\}^{3/2} \Omega_{m,q,\kappa}^h, \) and the rates of \( \Phi_{m,2,\kappa}^h \) and \( \Omega_{m,q,\kappa}^h \) are restricted in (C9).

Similarly, by defining

\[
\Phi_{2,\kappa}^\beta \overset{\text{def}}{=} \max_{\beta \in F_n^m} \|\partial_\beta g(Z_{j,:},\beta,\tilde{h})\|_{q,\kappa}, \quad \Omega_{q,\kappa}^\beta \overset{\text{def}}{=} \max_{\beta \in F_n^m} \|\partial_\beta g(Z_{j,:},\beta,\tilde{h})\|_{q,\kappa},
\]

we have

\[
\max_{\beta \in F_n^m} \left\{ \frac{1}{n} \sum_{t=1}^{n} \partial_\beta g(Z_{j,t},\beta,\tilde{h})\{\beta - \beta_{jk}\} \right\} \lesssim r_{N1,0} + r_{\kappa}r_{N2,0},
\]

where \( r_{N1,0} = \rho_n \sqrt{n}s\log(an)/\epsilon \Phi_{2,\kappa}^\beta, \) \( r_{N2,0} = \rho_n \{\log(an)/\epsilon\}^{3/2} \Omega_{q,\kappa}^\beta. \) And (C9) constrains the rates of \( \Phi_{2,\kappa}^\beta \) and \( \Omega_{q,\kappa}^\beta. \)

As a result, with probability \( 1 - o(1), \)

\[
N_n \lesssim \rho_{N_n},
\]

by letting \( \max_{m \in \{0,1,2\}} \{r_{N1,m} + r_{\kappa}r_{N2,m}\} = O(\rho_{N_n}). \)

As \( P(K_n + N_n \geq x) \leq P(K_n \geq x/2) + P(N_n \geq x/2) \) and collecting the results from (B.7), (B.8), and (B.11), we have shown that \( \Delta_n \) satisfies

\[
n^{-1} \max_{(j,k) \in G} \Delta_n \lesssim \rho_{\Delta_n},
\]

where \( \rho_{\Delta_n} = n^{-1}(\rho_n + \rho_{N_n}) = o(n^{-1/2}g_n^{-1}) \) (given \( \hat{\epsilon} \) is sufficiently small, and using (C8) and (C9)).

\( \Box \)

**Comment B.3.** [The rates of \( \Omega_{m,q,\kappa}^h \) and \( \Omega_{q,\kappa}^\beta \) by the definition under some special cases. For example, consider the VAR(1) model as in Comment 5.3 given by \( Y_t = AY_{t-1} + \varepsilon_t, \) where \( Y_t, \varepsilon_t \in \mathbb{R}^l, \) and \( \varepsilon_t \sim \text{i.i.d. N}(0, \Sigma). \) At first, as shown in the proof of Theorem 5.2, we have

\[
\Omega_{m,q,\kappa}^h = \max_{h \in F_n^m} \|\hat{X}_{j,k}^h\partial_{h_m}g(Z_{j,:},\beta,\tilde{h})\|_{q,\kappa} \lesssim \max_{(j,k) \in G} \|\hat{X}_{j,k}^h\|_{q,\kappa} \|\partial_{h_m}g(Z_{j,:},\beta,\tilde{h})\|_{2,\kappa}.
\]

For the first term, it is not hard to see that

\[
\max_{(j,k) \in G} \{\|\hat{X}_{j,k}^h\|_{q,\kappa} \} \lesssim |A|^{1/(2q)} \|\varepsilon_0\|_{q,\kappa} \lesssim J^{1/(2q)},
\]

where the last inequality is by the union bound, assuming \( |A|_{\infty} < 1, \) and the \( q \)th moments of
\(\varepsilon_{j,0} (\forall j)\) are bounded by a constant \(\mu_q\). As for the second term, let \(d_n \overset{\text{def}}{=} |G| \lor J\). In the mean regression case, for \(f \in \tilde{F}\), \(E(f(z_t)|\mathcal{F}_{t-1}) = \{X_{jk,t}(\beta^{0}_{jk} - \beta) + h^{0}_{jk} - \tilde{h}_1\}(v_{jk,t} + h^{0}_{jk} - \tilde{h}_2)\), it can be seen that

\[
\left\| \max_{h \in F_n^h} \{\partial h, g(Z_j,t, \beta, \tilde{h}) - \partial h, g(Z_j^*, t, \beta, \tilde{h})\} \right\|_{2q} \\
\leq \left\| \max_{(j,k) \in G} |v_{jk,t} - v_{jk,t}^*| \right\|_{2q} + \left\| \max_{(j,k) \in G} \{|X_{j(-k),t}^\top - (X_{j(-k),t})^*| \max_{\gamma_{j(-k)}} |\gamma_{j(-k)}^0 - \gamma_{j(-k)}| \right\|_{2q} \\
\lesssim d_n^{1/(2q)} (1 \lor s^{1/2} \rho_n),
\]

while in the median regression case, for \(f \in \tilde{F}\), \(E(f(z_t)|\mathcal{F}_{t-1}) = \left[\frac{1}{2} - F_{\varepsilon_j,t}|\mathcal{F}_{t-1}\right] \{X_{jk,t}(\beta^{0}_{jk} - \beta) + h^{0}_{jk} - \tilde{h}_1\}(v_{jk,t} + h^{0}_{jk} - \tilde{h}_2)\),

\[
\left\| \max_{h \in F_n^h} \{\partial h, g(Z_j,t, \beta, \tilde{h}) - \partial h, g(Z_j^*, t, \beta, \tilde{h})\} \right\|_{2q} \\
\leq \left\| \max_{(j,k) \in G} |v_{jk,t} - v_{jk,t}^*| \right\|_{4q} + \left\| \max_{(j,k) \in G} \{|X_{j(-k),t}^\top - (X_{j(-k),t})^*| \max_{\gamma_{j(-k)}} |\gamma_{j(-k)}^0 - \gamma_{j(-k)}| \right\|_{4q} \\
\lesssim d_n^{1/(4q)} (1 \lor s^{1/2} \rho_n),
\]

where we use the assumption such that the 4th moment of the conditional density is bounded. Moreover, we have

\[
\left\| \max_{h \in F_n^h} \{\partial h, g(Z_j,t, \beta, \tilde{h}) - \partial h, g(Z_j^*, t, \beta, \tilde{h})\} \right\|_{2q} \\
\leq \left\| \max_{(j,k) \in G} |(X_{j(-k),t} - X_{j(-k),t}^*)(\beta_{jk}^0 - \beta)| \right\|_{2q} \\
+ \left\| \max_{(j,k) \in G} \{|X_{j(-k),t}^\top - (X_{j(-k),t})^*| \max_{\beta_{j(-k)}} |\beta_{j(-k)}^0 - \beta_{j(-k)}| \right\|_{2q} \\
\lesssim d_n^{1/(2q)} (1 \lor s^{1/2} \rho_n),
\]

or \(\max_{h \in F_n^h} \{\partial h, g(Z_j,t, \beta, \tilde{h}) - \partial h, g(Z_j^*, t, \beta, \tilde{h})\} \right\|_{2q} = O(1)\) for the two cases. Therefore, we are able to conclude that \(\Omega_{m,q,\xi}^h \lesssim d_n^{1/q} (1 \lor s^{1/2} \rho_n)\) or \(\Omega_{m,q,\xi}^h \lesssim d_n^{3/(4q)} (1 \lor s^{1/2} \rho_n)\), respectively.

Similarly, it can be shown that \(\Omega_{q,\xi}^h \lesssim d_n^{1/q} s^{1/2} \rho_n\) or \(\Omega_{q,\xi}^h \lesssim d_n^{3/(2q)} s^{1/2} \rho_n\) for the two cases, since

\[
\left\| \max_{\beta \in F_n^\beta} |\partial \beta, g(Z_j, \tilde{\beta}, \tilde{h})| \right\|_q \\
\lesssim \left\| \max_{(j,k) \in G} |X_{j(-k),t}^\top - (X_{j(-k),t})^*| \right\|_{2q} \left\| \max_{(j,k) \in G} \{|X_{j(-k),t}^\top - (X_{j(-k),t})^*| \max_{\gamma_{j(-k)}} |\gamma_{j(-k)}^0 - \gamma_{j(-k)}| \right\|_{2q} \\
\lesssim d_n^{1/q} s^{1/2} \rho_n,
\]

52
or
\[
\|\max_{\beta \in F_n^*} |\partial_\beta g(Z_j, \ldots, \tilde{h}, \tilde{h})|\|_q \\
\lesssim \|\max_{(j,k) \in G} |X_{j(-k),t}^T - (X_{j(-k),t}^T)^*|_{4q}\| \max_{(j,k) \in G} \|\{X_{j(-k),t}^T - (X_{j(-k),t}^T)^*\}\{\bar{\gamma}_j^{0} - \bar{\gamma}_j^{(-k)}\}|_{4q} \\
\lesssim d_n^{1/(2q)} s^{1/2} \rho_n.
\]

In addition, a similar derivation can show that \(\|F(z_t)\|_q \lesssim d_n^{1/q} (1 \lor \rho_n)\) and \(\|\max_{(j,k) \in G} |\psi_{jk}^{0}|\|_{q,\varsigma} \lesssim d_n^{1/q} (1 \lor \rho_n)\).

**Lemma B.6.** Under the same conditions as in Theorem 5.4, we have with probability \(1 - o(1)\),
\[
\max_{(j,k) \in G} |E_n \psi_{jk}\{Z_{j,t}, \beta_{jk}^{0}, h_{jk}^{0}(X_{j(-k),t})\}| \lesssim r_n. \tag{B.12}
\]

**Proof of Lemma B.6.** Consider the class of function \(F_G = \{z \mapsto \psi_{jk}\{z, \beta_{jk}^{0}, h_{jk}^{0}(x_{j(-k)})\} : (j, k) \in G\}\), the cardinality of the set is \(|G|\). Therefore, the corresponding covering number is given by \(\text{sup}_{Q, N} (\epsilon\|\bar{F}_G\|_{Q,2}, F_G, \| \cdot \|_{Q,2}) = |G|/\epsilon\) with \(\bar{F}_G = \text{sup}_{f \in F_G} |f|\). Let \(\psi_{jk}^{0} \overset{\text{def}}{=} \psi_{jk}\{Z_{j,t}, \beta_{jk}^{0}, h_{jk}^{0}(X_{j(-k),t})\}\) and applying the tail probability bounds in Lemma B.2, we have with probability \(1 - o(1)\),
\[
\max_{(j,k) \in G} |E_n \psi_{jk}^{0}| \lesssim n^{-1}(r_1 + r_2) \lesssim r_n. \tag{B.13}
\]
where \(r_1 = (n \log a_n)^{1/2} \max_{(j,k) \in G} \|\psi_{jk}^{0}\|_{2,\varsigma}\), \(r_2 = (\log a_n)^{3/2} \max_{(j,k) \in G} |\psi_{jk}^{0}|\|_{q,\varsigma}\), \(r_\varsigma = n^{1/q} \) for \(\varsigma > 1/2 - 1/q\) and \(r_\varsigma = n^{1/2 - 1/q}\) for \(\varsigma < 1/2 - 1/q\).

**Lemma B.7.** Under the same conditions as in Theorem 5.4, consider the class of functions \(F' = \{z \mapsto \psi_{jk}\{z, \beta, \tilde{h}(x_{j(-k)})\} : (j, k) \in G, \beta \in B_{jk}, \tilde{h} \in H_{jk} \cup \{h_{jk}^{0}\}\}\), we have with probability \(1 - o(1)\),
\[
n^{-1/2} \text{sup}_{f \in F'} |G_n(f)| \lesssim \rho_n. \tag{B.14}
\]

**Proof of Lemma B.7.** The covering number of the function class \(F'\) is given by \(\text{sup}_{Q, N} (\epsilon\|\bar{F}'\|_{Q,2}, F', \| \cdot \|_{Q,2}) = (a_n/\epsilon)^{\varsigma_a}\) with \(\bar{F}' = \text{sup}_{f \in F'} |f|\). Also, for any \(f \in F'\), there exists a set \(F'_n\) such that \(\min_{f \in F'_n} \|f - f'\|_{Q,2} \leq \epsilon\|\bar{F}'\|_{Q,2}\) and the cardinality of the set \(|F'_n| = (a_n/\epsilon)^{\varsigma_a}\).

One can apply the technique we used in the proof of Lemma B.5 to achieve the concentration inequality. Similarly, consider the function set \(F'_n\), for each \(f \in F'_n\), let \(\varphi_{t,l} \overset{\text{def}}{=} f(z_{t})\) and \(\varphi_{t,l} \overset{\text{def}}{=} \varphi_{t,l} - E(\varphi_{t,l}|F_{t-1}, X_{j(-k),t})\), \(l = 1, \ldots, |F'_n|\). We have
\[
n|\max_{f \in F'_n} |E_n f - E_n E(f|F_{t-1}, X_{j(-k),t})| \lesssim 4 \sqrt{\text{ns} \log(a_n/\epsilon)} \max_{f \in F'} \|f(z_t)\|_2 + r_\varsigma \{\text{log}(a_n/\epsilon)^{3/2}\|\bar{F}'(z_t)\|_q\}. \]

For each \(f \in F'_n\), there exists a function \(g\) such that \(g(z_t, \beta, \tilde{h}) = E(f(z)\{F_{t-1}, X_{j(-k),t}\}) - E(f(z_t))\), where \(\beta \in B_{jk}, \tilde{h} \in H_{jk} \cup \{h_{jk}^{0}\}, (j, k) \in G\). As by the mean value theorem and the
continuity of the function \( g \), we have
\[
g(Z_{j,t}, \beta, \tilde{h}) = \partial_{\beta} g(Z_{j,t}, \beta, \tilde{h})(\beta - \beta_{jk}^0) + 2 \sum_{m=1}^{2} \partial_m g(Z_{j,t}, \beta, \tilde{h}) \{ \tilde{h}_m(X_{j(-k),t}) - h_{jk,m}^0(X_{j(-k),t}) \},
\]
where \((\beta, \tilde{h}(\cdot))\) is the corresponding point which joins the line segment between \((\beta, \tilde{h}(\cdot))\) and 
\((\beta_{jk}^0, h_{jk,m}^0(\cdot))\). Let \( F_n^\beta \) and \( F_n^\tilde{h} \) collect all the points of \( \beta \) and \( \tilde{h} \) according to \( F_n^\beta \), and define the following quantities \((m = 1, 2)\)

\[
\Phi^h_{m,2,\kappa} \defeq \max_{h \in F_n^h} \| \bar{X}_{jk} \partial_m g(Z_{j,-}, \beta, \tilde{h}) \|_{2,\kappa}, \quad \Omega^{\tilde{h}}_{m,\kappa,\varsigma} \defeq \max_{h \in F_n^\tilde{h}} \| \bar{X}_{jk} \partial_m g(Z_{j,-}, \beta, \tilde{h}) \|_{\kappa,\varsigma},
\]

\[
\Phi^{\beta}_{2,\kappa} \defeq \max_{\beta \in F_n^{\beta}} \| \partial_{\beta} g(Z_{j,-}, \beta, \tilde{h}) \|_{2,\kappa}, \quad \Omega^{\beta}_{\kappa,\varsigma} \defeq \max_{\beta \in F_n^{\beta}} \| \partial_{\beta} g(Z_{j,-}, \beta, \tilde{h}) \|_{\kappa,\varsigma}. \quad (B.15)
\]

Then we have with probability approaching 1,
\[
\max_{h \in F_n^\tilde{h}} \left| \sum_{t=1}^{n} \partial_m g(Z_{j,t}, \beta, \tilde{h}) \{ \tilde{h}_m(X_{j(-k),t}) - h_{jk,m}^0(X_{j(-k),t}) \} \right| \lesssim r_{N1,m} + r_{N2,m}, \quad m = 1, 2,
\]
\[
\max_{\beta \in F_n^{\beta}} \left| \sum_{t=1}^{n} \partial_{\beta} g(Z_{j,t}, \beta, \tilde{h})(\beta - \beta_{jk}^0) \right| \lesssim r_{N1,0} + r_{N2,0},
\]
where \( r_{N1,m} = \sqrt{n}s \rho_n \{ \log(a_n/\epsilon) \}^{1/2} \Phi^h_{m,2,\kappa} \), \( r_{N2,m} = s^2 \rho_n \{ \log(a_n/\epsilon) \}^{3/2} \Omega^{\tilde{h}}_{m,\kappa,\varsigma} \), and \( r_{N1,0} = \rho_n \sqrt{n}s \log(a_n/\epsilon) \Phi^{\beta}_{2,\kappa} \), \( r_{N2,0} = \rho_n \{ s \log(a_n/\epsilon) \}^{3/2} \Omega^{\beta}_{\kappa,\varsigma} \). Also \((C9)\) constrains the rates of \( \Phi^h_{m,2,\kappa}, \quad \Omega^{\tilde{h}}_{m,\kappa,\varsigma}, \Phi^{\beta}_{2,\kappa}, \) and \( \Omega^{\beta}_{\kappa,\varsigma} \).

The rest of the proof is similar as for Lemma \textbf{B.3} and thus is omitted. \( \square \)

\textbf{Lemma B.8.} Under the same conditions as in Lemma \textbf{B.5} with \((C9')\) instead of \((C6)\) \((C8)\) and \((C9)\)
\[
n^{-1} \max_{(j,k) \in G} \Delta_n \lesssim \sigma(n^{-1/2}g_n^{-1}), \quad (B.16)
\]
holds with probability \( 1 - o(1) \).

\textbf{Proof of Lemma \textbf{B.8}} We now study the tail probability under stronger tail assumptions. In particular, we need to carry out an analogue proof of Lemma \textbf{B.5} under \((C9')\).

Specifically, by Lemma \textbf{B.3} we have \( K_n \lesssim n^{1/2}(s \log a_n)^{1/\gamma} \rho^c_{u,v} \) (in particular, for the mean regression case \( \rho^c_{u,v} = \rho^c_s s \) and \( \rho^c_{u,v} = \sqrt{\rho^c_n} \)), and
\[
N_n \lesssim n^{1/2}(s \log(a_n/\epsilon)^{1/\gamma} \rho^c_n \{ s^{1/2} \max_{m \in \{1,2\}} \Phi^h_{m,\psi_v,0} \} \lor \Phi^{\beta}_{\psi_v,0} \},
\]
\[
\Phi^h_{m,\psi_v,0} \defeq \max_{h \in F_n^h} \| \bar{X}_{jk} \partial_m g(Z_{j,-}, \beta, \tilde{h}) \|_{\psi_v,0}, \quad \Phi^{\beta}_{\psi_v,0} \defeq \max_{\beta \in F_n^{\beta}} \| \partial_{\beta} g(Z_{j,-}, \beta, \tilde{h}) \|_{\psi_v,0}. \quad (B.17)
\]
The rest of the proof is similar as for Lemma \textbf{B.5} and thus is omitted. \( \square \)
Lemma B.9. Under the same conditions as in Lemma B.6 with (C9)' instead of (C6), (C8) and (C9), and assume that \( \max_{(j,k) \in G} \| \psi_{jk} \| \psi_{\nu,0} < \infty \), we have with probability \( 1 - o(1) \),

\[
\max_{(j,k) \in G} \left| E_n \psi_{jk} \{ Z_{j,t}, \beta_{jk}, h_{jk}^0 (X_{j(-k),t}) \} \right| \lesssim n^{-1/2} (\log a_n)^{1/\gamma} \max_{(j,k) \in G} \| \psi_{jk} \| \psi_{\nu,0} \lesssim r_n. \tag{B.18}
\]

Proof of Lemma B.9. The proof is similar to the proof of Lemma B.6 by replacing the tail probability bounds therein by Lemma B.3.

Lemma B.10. Under the same conditions as in Lemma B.7 with (C9)' instead of (C6), (C8) and (C9), and assume that \( \max_{f \in \mathcal{F}} |f(z)| \psi_{\nu,0} < \infty \), we have with probability \( 1 - o(1) \),

\[
n^{-1/2} \sup_{f \in \mathcal{F}_1} |G_n(f)| \lesssim \rho_n. \tag{B.19}
\]

Proof of Lemma B.10. The proof is similar to the proof of Lemma B.7 by replacing the tail probability bounds therein by Lemma B.3. In particular, it can be shown that

\[
n^{-1/2} \sup_{f \in \mathcal{F}_1} |G_n(f)| \lesssim n^{-1/2} \left( s \log (a_n / \epsilon) \right)^{1/\gamma} \left[ \max_{f \in \mathcal{F}_1} \| f(z) \| \psi_{\nu,0} \vee \rho_n \left\{ \left( s^{1/2} \max_{m \in \{1,2\}} \Phi_{m,\psi_{\nu,0}} \right) \wedge \Phi_{\psi_{\nu,0}} \right\} \right],
\]

\[
\Phi_{m,\psi_{\nu,0}} \overset{\text{def}}{=} \max_{h \in \mathcal{H}_n} \| \tilde{X}_{j,k} \partial_m g(Z_{j,t}, \beta, \tilde{h}) \| \psi_{\nu,0}, \Phi_{\psi_{\nu,0}} \overset{\text{def}}{=} \max_{\beta \in B_n} \| \partial g(Z_{j,t}, \beta, \tilde{h}) \| \psi_{\nu,0}. \tag{B.20}
\]

The final conclusion can be achieved by (C9)'.

B.4.2 Proofs of Section 5.6

Proof of Theorem 5.4. The sketch of the proof follows the proof of Theorem 2 in Belloni et al. (2015b).

Step 1: Let \( \hat{\beta}_{jk} \) be any estimator such that \( \max_{(j,k) \in G} |\hat{\beta}_{jk} - \beta_{jk}^0| \leq C \rho_n \) with probability \( 1 - o(1) \). By rewriting (using the fact that \( E_n [\psi_{jk} \{ Z_{j,t}, \beta_{jk}, h_{jk}^0 (X_{j(-k),t}) \}] = 0 \), we have

\[
E_n [\psi_{jk} \{ Z_{j,t}, \hat{\beta}_{jk}, \tilde{h}_{jk} (X_{j(-k),t}) \}] = E_n [\psi_{jk} \{ Z_{j,t}, \beta_{jk}, h_{jk}^0 (X_{j(-k),t}) \}]
+ E_n [\psi_{jk} \{ Z_{j,t}, \beta, \tilde{h} (X_{j(-k),t}) \}] |_{\beta = \hat{\beta}_{jk}, \tilde{h} = \tilde{h}_{jk}} + n^{-1} \Delta_n \tag{B.21}
\]

where \( \Delta_n \overset{\text{def}}{=} n^{1/2} G_n [\psi_{jk} \{ Z_{j,t}, \hat{\beta}_{jk}, \tilde{h}_{jk} (X_{j(-k),t}) \} - \psi_{jk} \{ Z_{j,t}, \beta_{jk}, h_{jk}^0 (X_{j(-k),t}) \}]. \)

We first observe that with probability \( 1 - o(1) \), \( \max_{(j,k) \in G} \Delta_n \leq \sqrt{n} \sup_{f \in \mathcal{F}_1} |G_n(f)| \), where \( \mathcal{F}_1 \) is the class of functions defined by \( \mathcal{F}_1 = \{ z \mapsto \psi_{jk} \{ z, \beta, \tilde{h} (x_{j(-k)}) \} - \psi_{jk} \{ z, \beta_{jk}, h_{jk}^0 (x_{j(-k)}) \} : (j,k) \in G, \beta \in B_n, |\beta - \beta_{jk}| \leq C \rho_n, \tilde{h} \in \mathcal{H}_n \} \). The key to our proof is to achieve a concentration inequality for \( \Delta_n \), such that \( n^{-1} \max_{(j,k) \in G} \Delta_n \lesssim o(n^{-1/2} g_n^{-1}) \) holds with probability \( 1 - o(1) \). This is done in Lemma B.3.

Then we expand the second term in (B.21) by Taylor expansion. Pick any \( \beta \in B_{jk} \) such that \( |\beta - \beta_{jk}| \leq C \rho_n \) and \( \tilde{h} \in \mathcal{H}_{jk} \). For any \( (j,k) \in G \), let \( (\beta, \tilde{h}, (X_{j(-k),t})^\top) \) lie on the line segment between \( (\beta, \tilde{h} (X_{j(-k),t}))^\top \) and \( (\beta_{jk}, h_{jk}^0 (X_{j(-k),t}))^\top \). Therefore, we can write
\[
E[\psi_{jk}(Z_{j,t}, \beta, \tilde{h}(X_{j(-k),t}))] \text{ as follows}
\]
\[
E[\psi_{jk}(Z_{j,t}, \beta_{jk}, h_{jk}^0(X_{j(-k),t}))] + E(\partial_\beta E[\psi_{jk}(Z_{j,t}, \beta_{jk}, h_{jk}^0(X_{j(-k),t}))|X_{j(-k),t}])(\beta - \beta_{jk}^0)
\]
\[+ \sum_{m=1}^{M} E(\partial_{h_m} E[\psi_{jk}(Z_{j,t}, \beta_{jk}, h_{jk}^0(X_{j(-k),t}))|X_{j(-k),t}]\{h_m(X_{j(-k),t}) - h_{jk,m}(X_{j(-k),t})\})
\]
\[+ \frac{1}{2} E(\partial_{\tilde{h}} E[\psi_{jk}(Z_{j,t}, \beta, \tilde{h}(X_{j(-k),t}))|X_{j(-k),t}])(\beta - \beta_{jk}^0)^2
\]
\[+ \frac{1}{2} \sum_{m,m'=1}^{M} E(\partial_{h_m} \partial_{h_{m'}} E[\psi_{jk}(Z_{j,t}, \beta, \tilde{h}(X_{j(-k),t}))|X_{j(-k),t}]\{h_m(X_{j(-k),t}) - h_{jk,m}(X_{j(-k),t})\})\{h_{m'}(X_{j(-k),t}) - h_{jk,m'}(X_{j(-k),t})\}((\beta - \beta_{jk}^0).
\]

(B.22)

It can be seen from the orthogonality condition \((5.22)\) that the third term in \((B.22)\) is zero. By \((C3)\) we have \(E(\partial_\beta E[\psi_{jk}(Z_{j,t}, \beta_{jk}, h_{jk}^0(X_{j(-k),t}))|X_{j(-k),t}] = E(\partial_\beta E[\psi_{jk}(Z_{j,t}, \beta_{jk}, h_{jk}^0(X_{j(-k),t}))]) = \phi_{jk}\). Moreover, each of the last three terms in \((B.22)\) is \(O(L_1n\beta_{jk}^2) = o(n^{-1/2}g_n^{-1})\) (by \((C3), (C5)\) and \((C8)\)). Therefore, we have shown that the second term in \((B.21)\) equals \(\phi_{jk}(\beta_{jk} - \beta_{jk}^0) + o(n^{-1/2}g_n^{-1})\), uniformly over \((j, k) \in G\). Then, combining the results in Lemma \(B.5\) gives

\[
E_n[\psi_{jk}(Z_{j,t}, \beta_{jk}, \tilde{h}_{jk}(X_{j(-k),t}))] = E_n[\psi_{jk}(Z_{j,t}, \beta_{jk}, h_{jk}^0(X_{j(-k),t}))] + \phi_{jk}(\beta_{jk} - \beta_{jk}^0) + o(n^{-1/2}g_n^{-1}).
\]

(B.23)

Step 2: Next, we need to prove that \(\inf_{\beta \in \hat{B}_{jk}} |E_n[\psi_{jk}(Z_{j,t}, \beta, \tilde{h}_{jk}(X_{j(-k),t}))]| = o(n^{-1/2}g_n^{-1})\) holds with probability \(1 - o(1)\). For any \((j, k) \in G\), we focus on any point \(\beta_{jk}^* = \beta_{jk}^0 - \phi_{jk}^{-1}E_n[\psi_{jk}(Z_{j,t}, \beta_{jk}, h_{jk}^0(X_{j(-k),t}))], \) thus

\[
\max_{(j,k)\in G} |\beta_{jk}^* - \beta_{jk}^0| \leq C \max_{(j,k)\in G} |E_n[\psi_{jk}(Z_{j,t}, \beta_{jk}, h_{jk}^0(X_{j(-k),t}))]|.
\]

By Lemma \(B.6\) we have \(|\beta_{jk}^* - \beta_{jk}^0| \lesssim r_n\) uniformly over \((j, k) \in G\). By \((C2)\) \([\beta_{jk}^0 \pm c_1r_n] \subset \hat{B}_{jk}\) with probability \(1 - o(1)\), thus \(\beta_{jk}^0\) is contained in \(\hat{B}_{jk}\) with probability \(1 - o(1)\). Using the continuity argument as in \((B.23)\) with \(\beta_{jk} = \beta_{jk}^0\) and combining the fact that \(\phi_{jk}(\beta_{jk}^* - \beta_{jk}^0) = -E_n[\psi_{jk}(Z_{j,t}, \beta_{jk}, h_{jk}^0(X_{j(-k),t}))]\), we have,

\[
E_n[\psi_{jk}(Z_{j,t}, \beta_{jk}^*, \tilde{h}_{jk}(X_{j(-k),t}))] = E_n[\psi_{jk}(Z_{j,t}, \beta_{jk}, h_{jk}^0(X_{j(-k),t}))] + \phi_{jk}(\beta_{jk}^* - \beta_{jk}^0) + o(n^{-1/2}g_n^{-1})
\]

\[= o(n^{-1/2}g_n^{-1}).
\]

Therefore,

\[
\max_{(j,k)\in G} \inf_{\beta \in \hat{B}_{jk}} |E_n[\psi_{jk}(Z_{j,t}, \beta, \tilde{h}_{jk}(X_{j(-k),t}))]| \leq |E_n[\psi_{jk}(Z_{j,t}, \beta_{jk}, \tilde{h}_{jk}(X_{j(-k),t}))]| = o(n^{-1/2}g_n^{-1}),
\]

(B.24)
holds with probability $1 - o(1)$ uniformly over $(j, k) \in G$.

Step 3: Lastly, it is left to prove that with probability $1 - o(1)$, $\max_{(j, k) \in G} |\hat{\beta}_{jk} - \beta_{0,jk}| \leq C \rho_n$, which will lead to the desired Bahadur representation. Consider the class of functions $\mathcal{F}' = \{ z \mapsto \psi_{jk}(z, \beta, \bar{h}(x_{j(-k)})) : (j, k) \in G, \beta \in B_{jk}, \bar{h} \in H_{jk} \cup \{ \eta_{jk} \} \}$. From (B.24) and by the definition of $\hat{\beta}_{jk}$ we have

$$|E_n[\psi_{jk}\{Z_{j,t}, \hat{\beta}_{jk}, \hat{h}_{jk}(X_{j(-k),t})\}]| \geq |E[\psi_{jk}\{Z_{j,t}, \beta, \bar{h}(X_{j(-k),t})\}]|_{\beta = \hat{\beta}_{jk}, \bar{h} = \hat{h}_{jk}} - n^{-1/2} \sup_{f \in \mathcal{F}'} |G_n(f)|,$$

holds with probability $1 - o(1)$ uniformly over $(j, k) \in G$.

Lemma B.7 ensures that $n^{-1/2} \sup_{f \in \mathcal{F}'} |G_n(f)| = \mathcal{O}(\rho_n)$. Furthermore, applying the expansion in (B.22) with $\beta_{0,jk} = \beta$ implies that

$$|E[\psi_{jk}\{Z_{j,t}, \beta, \bar{h}(X_{j(-k),t})\}] - E[\psi_{jk}\{Z_{j,t}, \beta, \bar{h}_{jk}(X_{j(-k),t})\}]| \leq C(\rho_n + L_{12}n\rho_n^2) = \mathcal{O}(\rho_n).$$

By (C3) along with the fact that $E[\{\bar{h}_m(X_{j(-k),t}) - \bar{h}_{jk,m}(X_{j(-k),t})\}^2] \leq C\rho_n^2$ for all $m = 1, \ldots, M$ and any $\bar{h} = (\bar{h}_m)_{m=1}^M \in H_{jk}$, we have with probability $1 - o(1),$

$$|E[\psi_{jk}\{Z_{j,t}, \beta, \bar{h}(X_{j(-k),t})\}]|_{\beta = \hat{\beta}_{jk}, \bar{h} = \hat{h}_{jk}} \geq |E[\psi_{jk}\{Z_{j,t}, \beta, \bar{h}_{jk}(X_{j(-k),t})\}]|_{\beta = \hat{\beta}_{jk}} - \mathcal{O}(\rho_n),$$

uniformly over $(j, k) \in G$. From (B.24) we can see that the left-hand side of (B.25) is $o(n^{-1/2}g_n^{-1})$. Moreover, due to the identification condition (C4) the first term on the right-hand side of (B.25) is bounded from below by $\frac{1}{2} \{ |\phi_{jk}(\hat{\beta}_{jk} - \beta_{0,jk})| \cap c_1 \}$ and this results in $|\hat{\beta}_{jk} - \beta_{0,jk}| \leq o(n^{-1/2}g_n^{-1}) + \mathcal{O}(\rho_n)$, with probability $1 - o(1)$.

In summary, we have shown that, with probability $1 - o(1)$,

$$E_n[\psi_{jk}\{Z_{j,t}, \hat{\beta}_{jk}, \bar{h}_{jk}(X_{j(-k),t})\}] = E_n[\psi_{jk}\{Z_{j,t}, \beta_{0,jk}, \bar{h}_{jk}(X_{j(-k),t})\}] + \phi_{jk}(\hat{\beta}_{jk} - \beta_{0,jk}) + o(n^{-2}g_n^{-1}),$$

uniformly over $(j, k) \in G$. And with probability $1 - o(1)$, the left-hand side is $o(n^{-1/2}g_n^{-1})$ uniformly over $(j, k) \in G$. Lastly, the uniform Bahadur representation can be obtained by solving (B.26) with respect to $|\hat{\beta}_{jk} - \beta_{0,jk}|$.

\[ \square \]

**Proof of Corollary 5.5**. The proof is an application of Theorem 5.4 with verification of conditions (C1)-(C9).

Here we focus on the estimator by Algorithm 2 as the proof of Algorithm 1 is basically the same. In particular, with the LAD regression case, we have $|G| = 1$, $a_n = \max(JK,n)$, $g_n = 1$, $M = 2$, $\bar{h}_{jk}(X_{j(-k),t}) = (X_{j(-k),t}^\top \beta_{0,jk}^0, X_{j(-k),t}^\top \gamma_{j(-k)}^0)^\top$, $\psi_{jk}\{Z_{j,t}, \beta_{jk}, \bar{h}_{jk}(X_{j(-k),t})\} = \{1/2 - 1(Y_{j,t} \leq X_{jk,t} \beta_{jk} + X_{j(-k),t}^\top \beta_{0,jk} + X_{j(-k),t}^\top \gamma_{j(-k)}^0)\}(X_{jk,t} - X_{j(-k),t}^\top \gamma_{j(-k)}^0)$. Verification of (C1) Our model setting assumes $F_{\epsilon_j}(0) = 1/2$ and $E(v_{jk,t}X_{j(-k),t}) = 0$.
hence we have

\[
\begin{align*}
E(\partial_{\beta_k} E[\psi_{jk} \{ Z_{j,t}, \beta_k^0, h_k(X_{j(-k),t}) \}] | X_{j(-k),t}) & = -\beta_k^T (-k) E(f_{j,t}(0)v_{j,t}X_{j(-k),t}) = 0 \\
E(\partial_{\gamma_k} E[\psi_{jk} \{ Z_{j,t}, \beta_k^0, h_k(X_{j(-k),t}) \}] | X_{j(-k),t}) & = -\gamma_k^T (-k) E[\{1/2 - F_{j,t}(0)\} X_{j(-k),t}] = 0
\end{align*}
\]

Verification of (C2) The true parameter \( \beta_k^0 \) satisfies (5.21) given \( F_{j,t}(0) = 1/2 \). Moreover, based on the fact that \( |\beta_k^1| - \beta_k^0 |_{pr} \lesssim \sqrt{s(\log a_n)/n} \) (according to Corollary 5.4), and by Remark 2 in Belloni et al. (2015a), with probability \( 1 - o(1) \), \( |\beta_k^2| - \beta_k^0 |_{pr} \lesssim o(1/\log n) \), so that for some sufficiently small \( c > 0 \), \( [\beta_k^0 + c/\log n] \subset \hat{B}_{jk} \subset B_{jk} \), with probability \( 1 - o(1) \). Then the condition holds.

Verification of (C3) The map

\[
(\beta, h) \mapsto E[\psi_{jk}(Z_{j,t}, \beta, h) | X_{j(-k),t})
\]

is twice continuously differentiable as \( f'_{j,t} \) is continuous. For every \( \vartheta \in \{ \beta, h_1, h_2 \} \),

\[
\partial_{\beta_k} E[\psi_{jk}(Z_{j,t}, \beta, h) | X_{j(-k),t}) = - E[\bar{f}_{j,t} \{ X_{j,t}, \beta - \beta_k^0 \} | X_{j(-k),t}, h_1] (X_{j,t} - h_2) | X_{j(-k),t})
\]

(w.r.t. \( \beta \)) or \(- E[\bar{f}_{j,t} \{ X_{j,t} \} | X_{j(-k),t}, h_1] (X_{j,t} - h_2) | X_{j(-k),t}) \) (w.r.t. \( \beta_j \)) or

\[
- E[1/2 - F_{j,t} \{ X_{j,t} \} | X_{j(-k),t}) \) (w.r.t. \( h_2 \)). Hence, for every \( \beta \in B_{jk} \),

\[
|\partial_{\beta} E[\psi_{jk}(Z_{j,t}, \beta, h) | X_{j(-k),t}) | X_{j(-k),t}) | \lesssim C_1 E[|X_{j,t}v_{j,t} | | X_{j(-k),t}) \] \( \lor \) \( C_1 E[|v_{j,t} | | X_{j(-k),t}) \lor 1.
\]

Therefore, the expectation of the square of the right-hand side is bounded. Moreover, let \( T_{jk}(X_{j(-k),t}) = \{ \tau \in \mathbb{R}^2 : \tau_2 - X_{j(-k),t}^T \beta_k^0 (-k) \leq c_3 \} \), where \( c_3 > 0 \) is a constant. Then for every \( \vartheta, \vartheta' \in \{ \beta, h_1, h_2 \} \), \( \beta \in B_{jk} \), \( h \in T_{jk}(X_{j(-k),t}) \), we have

\[
|\partial_{\beta} \partial_{\vartheta'} E[\psi_{jk}(Z_{j,t}, \beta, h) | X_{j(-k),t}) | \]

\[
\lesssim C_1 [1 \lor E([X_{j,t}^2 | X_{j,t} - h_2]) | | X_{j(-k),t}) \lor E([X_{j,t} | X_{j,t} - h_2]) | | X_{j(-k),t}) \lor E([X_{j,t} - h_2] | X_{j(-k),t})].
\]

In particular,

\[
E([X_{j,t}^2 | X_{j,t} - h_2]) | | X_{j(-k),t}) \leq E([X_{j(-k),t}^T \gamma_{j(-k)} + v_{j,k,t})^2 (c_3 + |v_{j,k,t}|) | | X_{j(-k),t})
\]

\[
\leq 2E([((X_{j(-k),t}^T \gamma_{j(-k)})^2 + v_{j,k,t})^2 (c_3 + |v_{j,k,t}|) | | X_{j(-k),t}) \]

\[
\leq C|X_{j(-k),t}^T \gamma_{j(-k)}|^2.
\]

And by similar computation we can show that \( |\partial_{\beta} \partial_{\beta'} E[\psi_{jk}(Z_{j,t}, \beta, h) | X_{j(-k),t}) | \leq \ell_1(X_{j(-k),t}) = C'|X_{j(-k),t}^T \gamma_{j(-k)}|^2 \), where the constants \( C, C' \) depend on \( c_3 \) and \( C_1 \). Lastly, for every \( \beta, \beta' \in \)
where constant $C''$ depends on $c_3$ and $C_1$. Consequently, we have verified the last condition in (C3) by taking $\ell_2(X_{j(-k),t}) = \sqrt{2}(C''|X_{j(-k),t}^{\top}\gamma_{j(-k)}^0| + 2c_3)$ and $v = 1$. And given the finite moments conditions on $X_t$, we have $E\{\ell^2_1(X_{j(-k),t})\} \leq L_1n$, $E\{\ell^2_2(X_{j(-k),t})\} \leq L_2n$.

Verification of (C4) For any $\beta \in B_{jk}$, there exists $\beta'$ between $\beta_{jk}^0$ and $\beta$ such that

\[
E[\psi_{jk}\{Z_{j,t},\beta,\phi_{jk}(X_{j(-k),t})\}] = \partial_{\beta} E[\psi_{jk}\{Z_{j,t},\beta_{jk}^0,\phi_{jk}(X_{j(-k),t})\}](\beta - \beta_{jk}^0) + \frac{1}{2} \partial_{\beta}^2 E[\psi_{jk}\{Z_{j,t},\beta',\phi_{jk}(X_{j(-k),t})\}](\beta - \beta_{jk}^0)^2.
\]

Let $\phi_{jk} = \partial_{\beta} E[\psi_{jk}\{Z_{j,t},\beta,\phi_{jk}(X_{j(-k),t})\}] \geq c_2^2$. Since $\partial_{\beta}^2 E[\psi_{jk}\{Z_{j,t},\beta',\phi_{jk}(X_{j(-k),t})\}] \leq C_1E|X_{jk,t}^2v_{jk,t}| \leq C_2$, we have

\[
2E[\psi_{jk}\{Z_{j,t},\beta,\phi_{jk}(X_{j(-k),t})\}] \geq 2\phi_{jk} |\beta - \beta_{jk}^0| - C_2(\beta - \beta_{jk}^0)^2 \geq \phi_{jk}|\beta - \beta_{jk}^0|,
\]

whenever $|\beta - \beta_{jk}^0| \leq c_2^2/C_2$.

Verification of (C5) According to Corollary 5.4 with probability $1 - o(1)$ we have

\[
\|\tilde{\beta}_{j(-k)}^{[1]} - \beta_{jk}^0\|_{j,pr} \lesssim \sqrt{s(\log a_n)/n}, \quad \|\hat{\gamma}_{j(-k)} - \gamma_{j(-k)}^0\|_{j,pr} \lesssim \sqrt{s(\log a_n)/n},
\]

which means the algorithms can provide an estimator of the nuisance function with good sparsity and rate properties given IC $\lambda$. Thus, by Lemma 7 in Belloni et al. (2015a), we have (C5) holds.

Verification of (C6) We refer to the proof of Theorem 1 in Belloni et al. (2015a).

Verification of (C7) Recall that $\psi_{jk}^0 = \{1/2 - 1(\varepsilon_{j,t} \leq 0)\}v_{jk,t}$. Hence, $E(\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \psi_{jk,t}^0)^2 = \sum_{t=-(n-1)}^{n-1}(1 - |\ell|/n)E(\psi_{jk,t}^0\psi_{jk,t-\ell}) \geq \frac{1}{4} \sum_{t=-(n-1)}^{n-1}(1 - |\ell|/n)E(v_{jk,t}v_{jk,t-\ell}) \geq c_1/4$.

Verification of (C8) and (C9) See Comment 5.8 where we discuss the admissible dimension rates either under the special case of VAR(1) with geometric decay rate (which gives bounded dependence adjusted norm) or more generally with finite dependence adjusted norm in polynomial rates.

Verification of (C9') See Comment 5.9 and the discussion can be generalized to the case of finite dependence adjusted norm in polynomial rates easily.

\[\text{Lemma B.11. Let } \psi_{jk}^0 \overset{\text{def}}{=} \psi_{jk}\{Z_{j,t},\beta_{jk}^0,\phi_{jk}(X_{j(-k),t})\}, T_{jk}^n \overset{\text{def}}{=} \sigma_{jk}^{-1}\phi_{jk}^{-1} \sum_{t=1}^{n} \psi_{jk,t}, \text{ and assume that } \|\psi_{jk}^0\|_{2,\varsigma} < \infty. \text{ Then}
\]

\[
\|T_{jk}^n\|_2 = O(\sqrt{n}\|\psi_{jk}^0\|_{2,\varsigma}), \text{ and } n^{-1/2}T_{jk}^n \overset{\mathcal{D}}{\to} N(0,1).
\]
Proof of Lemma B.11. Define the projector operator $P_l(X_t) \overset{\text{def}}{=} E(X_t|\mathcal{F}_l) - E(X_t|\mathcal{F}_{l-1})$. Note that the projection operator is directly linked to the dependence adjusted norm for $X_{jk,t} = g_{jk}(\mathcal{F}_t) = g_{jk}(\xi_t)$, and $\|P_0(X_{jk,t})\|^2 \leq \|g_{jk}(\mathcal{F}_t) - g_{jk}(\mathcal{F}_1)\|^2 \leq 2\|P_0(X_{jk,t})\|^2$ (by Theorem 1(i) in [Wu, 2005]).

Let $J_{l,n}^{jk} \overset{\text{def}}{=} \sigma_{jk}^{-1}(\phi_{jk}^{-1})\sum_{t=1}^{\infty} P_{l-1}(\psi_{jk,l,t})$, and it is not hard to see that $T_n^{jk} = \sum_{t=0}^{\infty} J_{l,n}^{jk}$. As $\sigma_{jk}^{-1}(\phi_{jk}^{-1})P_{l-1}(\psi_{jk,l,t})$ form the martingale differences over $t$, according to Lemma B.11 we can apply the Burkholder Inequality and get $\|J_{l,n}^{jk}\|^2 \leq (\sigma_{jk}^{-1}(\phi_{jk}^{-1}))^{-2} \sum_{t=1}^{\infty} \|P_{l-1}(\psi_{jk,l,t})\|^2 \leq n(\sigma_{jk}^{-1}(\phi_{jk}^{-1}))^{-2}$, where $\sigma_{jk}^{-1}(\phi_{jk}^{-1}) = \|\psi_{jk,l,t}\|_{l,n}$. Thus, $\|T_n^{jk}\|^2 \leq \sqrt{n} \sum_{t=0}^{\infty} \|\psi_{jk,l,t}\|_{l,n}^2 = O(\sqrt{n}\|\psi_{jk,l,t}\|_{l,n}^2)$. Then the conclusion that $n^{-1/2} \tilde{T}_n^{jk} \overset{\text{d}}{\to} N(0,1)$ follows from Lemma B.4 in light of the fact that $E\psi_{jk,l,t} = 0$ and $\|\psi_{jk,l,t}\|_{l,n} < \infty$.

Proof of Theorem 5.5. The proof follows directly from Lemma B.11.

Proof of Corollary 5.6. We apply the high-dimensional central limit theorem (Theorem 3.2 in Zhang and Wu [2017a]) to the vector $\tilde{\zeta} \overset{\text{def}}{=} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \tilde{\zeta}_t$ and $\tilde{Z} \overset{\text{def}}{=} \text{vec}[[Z_{jk,k_{l-1}}]_{k=1}^{J}]$ is the corresponding standard Gaussian random vector, with the same correlation structure. Then we have $p(D^{-1}\tilde{\zeta}, D^{-1}\tilde{Z}) \to 0$, as $n \to \infty$, where $D$ is a diagonal matrix with the square root of the diagonal elements of the long-run variance-covariance matrix of $\tilde{\zeta}_t$, namely $\{\sum_{t=\infty}^{\infty} E(\tilde{\zeta}_{jk,t}\tilde{\zeta}_{jk,t-\ell})\}^{1/2}$, for $k = 1, \ldots, K$, $j = 1, \ldots, J$. The rest of the proof is similar to Corollary 5.2 and thus is omitted.

Proof of Corollary 5.7. The proof is similar to that of Theorem 5.3 and Theorem A.3 therefore, we omit the detailed proof here. In particular, the following conditions on $b_n$ are required:

$$b_n = o(n(\log |G|)^{-4}(\Phi_{q,t})^{-4} \wedge n(\log |G|)^{-5}(\Phi_{q,t})^{-4}c_n^{-2}),$$

$$F_{q,t} = o(n^{q/2}(\log |G|)^{-q}G^{-1}(\Phi_{q,t}^{-1})^{-q}c_n^{-q/2}),$$

with $c_n^{-1} = o(1)$.

$$\Phi_{2,0}\Phi_{2,0}^{-1}\{b_n^{-1} + \log(n/b_n)/n + (n - b_n)\log b_n/(nb_n)\}(\log |G|)^2 = o(1),$$

$$\Phi_{2,0}\Phi_{2,0}^{-1}\{b_n^{-1} + n^{-\zeta} + (n - b_n)b_n^{-\zeta+1}/(nb_n)\}(\log |G|)^2 = o(1),$$

$$\Phi_{2,0}\Phi_{2,0}^{-1}\{b_n^{-1} - n+1 - n^{-1}b_n^{-\zeta+1} + (n - b_n)/(nb_n)\}(\log |G|)^2 = o(1),$$

for $q > 1 - 2/q$: $F_{\zeta} = l_n b_n^{q/2-cq/2}/2$, for $1/2 < q < 1 - 2/q$: $F_{\zeta} = l_n^{q/2-cq/2} b_n^{2-cq/2}$.

C Supplementary Examples

C.1 Practical Examples of SRE

Example 4 (Identification Test for Large Structural Vector Autoregression Models). Denote $U_t = (U_{1,t}, U_{2,t}, \ldots, U_{M,t})^\top$. A large structural VAR can be represented in the following form (without loss of generality, consider only lag one):

$$AU_t = BU_{t-1} + \varepsilon_t,$$
where \( A \) (invertible) and \( B \) are \( M \times M \) matrices. The structural shocks \( \varepsilon_t \) satisfy \( \mathbb{E}(\varepsilon_t) = 0 \) and \( \text{Var}(\varepsilon_t) = I_M \). The corresponding reduced form is given by

\[
U_t = DU_{t-1} + \nu_t,
\]

with \( D = A^{-1}B \) and \( \nu_t = A^{-1}\varepsilon_t \), where \( \nu_t \) is denoted as the reduced form VAR shocks. Suppose \( \nu_t \) spans the space of \( \varepsilon_t \). The crucial question is the identification of \( A \). Typically, the covariance matrix of the reduced form shock \( \nu_t \) is estimated with \( M(M+1)/2 \) restrictions, which are smaller than the \( M^2 \) restrictions needed to pin down \( \varepsilon_t \). Adopting the identification approach proposed by Stock and Watson (2012), we may use external instruments that are correlated with the shock of interest and are uncorrelated with other shocks. Without loss of generality, suppose the structural shock of interest is \( \varepsilon_{j,t} \). Then we can define \( z_{j,t} \) as an external instrument for the \( j \)th structural shock satisfying

\[
E(\varepsilon_{j,t}z_{j,t}) \neq 0,
\]

\[
E(\varepsilon_{j',t}z_{j,t}) = 0, \quad \text{for } j' \neq j.
\]

Thus, we propose to regress \( z_{j,t} \) on \( \nu_t \):

\[
z_{j,t} = \nu_t^\top \delta_j + e_{j,t}.
\]

In practice, \( \nu_t \) are replaced by the residuals obtained from a large VAR reduced form regression as in Example 3. The estimator of \( \delta_j \) is denoted as \( \hat{\delta}_j \). It can be obtained by LASSO estimation, which gives us a sparse estimator of the \( j \)th row of the matrix \( A^{-1} \) up to a scaling factor. Repeating this step for any \( j \), one may formulate estimators for each row and perform simultaneous inference/hypothesis testing on the structural matrix \( A^{-1} \).

In summary, this is also a special case of SRE with

\[
(Y_{j,t}, X_{j,t}, \varepsilon_{j,t}, \beta^0_j) = (U_{j,t}, U_{-j,t-1}, \nu_t, D_j^\top), \quad j = 1, \ldots, M,
\]

\[
(Y_{j,t}, X_{j,t}, \varepsilon_{j,t}, \beta^0_j) = (z_{(j-M),t}, \nu_t, e_{(j-M),t}, \delta_{(j-M)}), \quad j = M + 1, \ldots, 2M.
\]

**Example 5 (Cross-sectional Asset Pricing).** Denote \( Y_{j,t} \) as the excess return for asset \( j \) and period \( t \). Asset pricing models explain the cross sectional variation in expected returns across assets; see e.g. Cochrane (2009). In particular, the variation of expected cross sectional returns is explained by the exposure to \( K - 1 \) factors \( X_{jk,t}, k = 1, \ldots, K - 1 \). One commonly used way to estimate an asset pricing model is to run a system of regression equations:

\[
Y_{j,t} = \beta_{j0} + \sum_{k=1}^{K-1} \beta_{jk}X_{jk,t} + \varepsilon_{j,t}, \quad (C.2)
\]

where \( X_{jk,t} \)'s are the factor returns (assumed to be excess returns of zero-cost portfolios). The selection of factors is a critical issue and the SRE framework addresses this issue, in particular when the number of factors \( K \) is large. See Feng et al. (2017) for a detailed model-selection exercise on picking asset pricing factors. The factor premiums are \( \mathbb{E}(X_{jk,t}) \) and the
pricing errors are $\beta_{j0}$. Usually, asset pricing imposes the restriction that all $\beta_{j0}$’s are zero. Our simultaneous inference framework naturally serves the purpose of simultaneously testing the zero pricing errors in a cross-sectional regression setup. Namely, we are interested in testing $H_0 : \beta_{j0} = 0, \forall j = 1, \ldots, J$ versus $H_A : \exists j$ such that $\beta_{j0} \neq 0$. Our test procedure in Section 4.2 can be directly applied to achieve this goal.

Example 6 (Network Formation and Spillover Effects). There is an emerging literature in economics concerning quantifying spillover effects and network formation. One leading example is as in Manresa (2013), which attempts to quantify social returns to research and development (R&D). Here, $U_{j,t}$ is taken to be the log output for firm $j$ and time $t$. This output is loading on $D_{j,t}$ (capital stock for firm $j$ and period $t$), and the aggregated spillovers from the capital stock of other firms $\sum_{i \neq j} w_{ij} D_{i,t}$. The regression equation also controls for other covariates $X_{j,t}$ (e.g., log labor, log capital etc.):

$$U_{j,t} = \beta_j D_{j,t} + \sum_{i \neq j} \omega_{ij} D_{i,t} + \gamma_j^T X_{j,t} + \varepsilon_{j,t}, \quad \text{(C.3)}$$

where $\omega_{ij}$ is referred to as the spillover effects of the R&D development of firm $i$ on firm $j$. This again is contained in the SRE with

$$(Y_{j,t}, X_{j,t}, \varepsilon_{j,t}, \beta_{j0}^0) = (U_{j,t}, (D_{j,t}, D_{j,t}^T, X_{j,t}^T)^T, \varepsilon_{j,t}, (\beta_j, \omega_{(\cdot j)}, \gamma_j^T)^T), \quad j = 1, \ldots, J.$$  

Our simultaneous inference procedure (Section 4.2) can be applied to check the significance of the spillover effects for any set of parameters of interest. As an analogy, the presented framework displays a general class of network models, where $U_{j,t}$ is taken to be the nodal response, and $D_{i,t}$ are the nodal covariates. Global or local inference on the network parameters $\omega_{ij}$ is the subject of research. Section 7 is devoted to inference on the spillover effects of a textual sentiment index.

Comment C.1. Suppose there is unobserved heterogeneity in $U_{j,t}$, e.g. $U_{j,t} = \alpha_j + \sum_{i \neq j} w_{ij} D_{i,t} + \varepsilon_{j,t}$, where $w_{ij}$ characterizes the spillover of individual $i$ on $j$, and $\alpha_j$ is the individual fixed effect. For this situation consider the demeaned version to eliminate the individual specific effects and work with the new model: $\tilde{U}_{j,t} = \sum_{i \neq j} w_{ij} \tilde{D}_{i,t} + \tilde{\varepsilon}_{j,t}$, where $\tilde{U}_{j,t} = U_{j,t} - \frac{1}{n} \sum_{j=1}^n U_{j,t}$, $\tilde{D}_{i,t} = D_{i,t} - \frac{1}{n} \sum_{i=1}^n D_{i,t}$, $\tilde{\varepsilon}_{j,t} = \varepsilon_{j,t} - \frac{1}{n} \sum_{i=1}^n \varepsilon_{j,t}$, under the condition that $U_{j,t}$ has no feedback effects on $D_{i,t}$ (for example, $D_{i,t}$ should not be the lagged variable of $U_{j,t}$).

C.2 Examples of the Dependence Measure

1. AR(1): $Y_t$ follows $Y_t = aY_{t-1} + \varepsilon_t$, with $|a| < 1$, $\varepsilon_t \sim \text{i.i.d.}(0, \sigma^2)$. Therefore, the MA representation is given by $Y_t = \sum_{l=0}^\infty a^l \varepsilon_{t-l}$ and $Y_t^* = \sum_{l=0}^\infty a^l \varepsilon_{t-l} + a^l \varepsilon_0$. $\|Y_t - Y_t^*\|_q = |a|^q \|\varepsilon_0 - \varepsilon_0^*\|_q$, $\Delta_{m,q} \lesssim |a|^m$, $\|Y_t\|_{q,\infty} \lesssim \sup_{m \geq 0} (m+1)^{\frac{1}{q}} |a|^m < \infty$.  

2. ARCH(1): An ARCH (Autoregressive conditionally heteroscedastic) model is given by $Z_t = \sigma_t \varepsilon_t$, $\sigma_t^2 = \alpha_0 + \alpha_2 Z_{t-1}^2$, with $\alpha_2 > 0$, $\varepsilon_t$ are i.i.d. shocks and $\text{Var}(Z_t) = \sigma^2 < \infty$. Thus, it is not hard to see that $Z_t^2 = w \sum_{l=0}^\infty \alpha_2 \varepsilon_{t-l}^2$. Rewriting the model as $Z_t = R(Z_{t-1}, \varepsilon_t) = \sqrt{(w + \alpha^2 Z_{t-1}^2)} \varepsilon_t$. According to Wu and Shao (2004), we have
the Lipschitz constant involved in the Lyapunov type condition ensuring the forward iteration contraction $\sup_{x \neq x'} |R(x, \epsilon) - R(x', \epsilon)| \leq |\alpha \epsilon|$. Let $\mu \overset{\text{def}}{=} E|\alpha \epsilon| < 1$ and assume $|\alpha \epsilon| + |R(t_0, \epsilon_0)|$ has finite qth moment. Then the process $Z_t$ has stationary solutions. Moreover, $\|Z_t - Z_t^\ast\|_q \leq |\mu|^q |\epsilon_0 - \epsilon_0^\ast|_q$, and thus $\Delta_{m,q} \overset{\preceq}{=} |\mu|^m$. Given $|\mu| < 1$, then we have $\|Z\|_{q,\kappa} \overset{\preceq}{=} \sup_{m \geq 0} (m + 1)^{q/2} |\mu|^m < \infty$.

3. TAR (Threshold autoregressive model): $Y_t = \theta_1 Y_{t-1} 1\{Y_{t-1} < \tau\} + \theta_2 Y_{t-1} 1\{Y_{t-1} \geq \tau\} + \epsilon_t$, where $\theta_1$ and $\theta_2$ are two parameters and $\epsilon_t$ are i.i.d. shocks. If $\theta \overset{\text{def}}{=} \max(\|\theta_1\|, \|\theta_2\|) < 1$ and $\epsilon_t$ has a finite $\alpha$-th order moment, then the TAR model admits a stationary solution with $\|Y\|_{q,\kappa} \overset{\preceq}{=} \sup_{m \geq 0} (m + 1)^{q/2}|\theta|^m < \infty$.

4. VAR (Vector autoregressive model): Without loss of generality we focus on VAR(1) given by $Y_t = A Y_{t-1} + \epsilon_t$, where $Y_t, \epsilon_t \in \mathbb{R}^J$, and $\epsilon_t \sim$ i.i.d. $N(0, \Sigma)$. If the spectral radius of $A^\top A$, $\rho(A^\top A) < 1$, then $\lim_{m \to \infty} \|A\|^m \to 0$, where $\| \cdot \|$ denotes the spectral norm of a matrix. Rewrite the model as $Y_t = \sum_{t=0}^\infty A^t \epsilon_{t-1}$. The existence of a stationary solution can be checked by Kolmogorov’s three series theorem. For each equation $j$, $Y_{j,t} - Y_{j,t}^\ast = [A]_j j (\epsilon_0 - \epsilon_0^\ast)$, where $[A]_j$ is the jth row of the matrix $A^t$. $(E[(Y_{j,t} - Y_{j,t}^\ast)|\theta]) = \sum_{i=0}^\infty (\epsilon_0 - \epsilon_0^\ast) [A]_j (\theta - \theta^\ast)^{1/2}$, where $\theta^\ast = \text{vech}(\sum_{i=0}^\infty (\epsilon_0 - \epsilon_0^\ast) [A]_j (\theta - \theta^\ast)^{1/2})$. The vector notation of the iterations follows $h_t = \text{vech}(\sum_{i=0}^\infty (\epsilon_0 - \epsilon_0^\ast) [A]_j (\theta - \theta^\ast)^{1/2})$. The partial derivative matrix is $\Delta_t = \Delta(h_t, \epsilon_t) = \partial h_{t+1} / \partial \epsilon_{t+1} = \tilde{A} \tilde{D}^\top_{\theta}(H_t^{1/2} \epsilon_{t+1}^\top H_t^{-1/2} \otimes I_J) \tilde{D}_{\theta}$, and $\Sigma = \text{E} \Delta_t = \tilde{A}$. Therefore, the spectral radius of $A A^\top < 1$ ensures a stationary solution to the process $h_t$.

According to [Hahner and Preminger (2009)], the iteration formulae are given by $h_t = \varpi(h_{t-1}, \epsilon_{t-1}) + \sum_{k=1}^{m} \Pi^k \Delta(h_{t-k}, \epsilon_{t-k}) \varpi(h_{t-k}, \epsilon_{t-k}) h_{t-m}$, where $\varpi(h, \epsilon) = w + G(h^\ast, \epsilon) - \Delta(h, \epsilon) h^\ast$, $h^\ast$ is the contraction state, and $h_{t-k}$’s lie on the line segment between $h^\ast$ and $h_{t-k}$. For ease of derivation, we assume a strong assumption such that $E \sup_{m, q} \|\Delta(h_m, \epsilon_m)\|^q < \infty$ for all $m \geq 1$ and $q \geq 2$, where $\| \cdot \|$ denotes...
the spectral norm of a matrix. Let $h^m = \{(h_1^T, \ldots, h_m^T)^T : |h_t|^2 = 1, t = 1, \ldots, m\}$, it follows $E_{h^m} \|\prod_{k=1}^m \Delta(h_{m-k+1}, \varepsilon_{m-k+1})\|^q \leq \prod_{k=1}^m E_{h_{m-k+1}} \|\Delta(h_{m-k+1}, \varepsilon_{m-k+1})\|^q \leq s^m$. Hence, $\max_{1 \leq j \leq J(J+1)/2} \|h_j\|_{q, \varsigma} \leq C$, $\|h_\cdot\|_{\infty, q, \varsigma} \lesssim \|h_t\|_{q} \lesssim \{J(J+1)/2\}^{1/q}$, and $(\sum_{j=1}^{J(J+1)/2} \|h_j\|_{q, \varsigma})^{1/q} \lesssim \{J(J+1)/2\}^{1/q}$. 
## D Additional Details for Empirical Analysis

<table>
<thead>
<tr>
<th>Consumer Discretionary (11)</th>
<th>Financials (8)</th>
<th>Industrials (10)</th>
<th>Information Technology (11)</th>
<th>Materials (3)</th>
<th>Utilities (2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMZN Amazon.com, Inc.</td>
<td>AIG American International Group, Inc.</td>
<td>BAC Bank of America Corporation</td>
<td>GD General Dynamics Corporation</td>
<td>AAPL Apple Inc.</td>
<td>AA Alcoa Corporation</td>
</tr>
<tr>
<td>BBY Best Buy Co. Inc.</td>
<td>AMT American Tower Corporation (REIT)</td>
<td>GS Genmark Limited</td>
<td>GE General Electric Company</td>
<td>ACN Accenture plc</td>
<td>DD EI du Pont de Nemours &amp; Co</td>
</tr>
<tr>
<td>CBS CBS Corporation</td>
<td>AXP American Express Company</td>
<td>JPM J P Morgan Chase &amp; Co</td>
<td>HON Honeywell International Inc.</td>
<td>ADP Automatic Data Processing, Inc.</td>
<td>DOW Dow Chemical</td>
</tr>
<tr>
<td>CMCSA Comcast Corporation</td>
<td>BAC Bank of America Corporation</td>
<td>Health Care (8)</td>
<td>IBM International Business Machines Corporation</td>
<td>BIIB Biogen Inc.</td>
<td>DUK Duke Energy Corp.</td>
</tr>
<tr>
<td>CMG Chipotle Mexican Grill, Inc.</td>
<td>C Cigtigroup Inc.</td>
<td>JNJ Johnson &amp; Johnson</td>
<td>EMC EMC Corporation</td>
<td>EBAY eBay Inc.</td>
<td>EXC Exelon Corporation</td>
</tr>
<tr>
<td>DIS Walt Disney Company (The)</td>
<td>E*TRADE Financial Corporation</td>
<td>LLY Eli Lilly and Company</td>
<td>FSLR First Solar, Inc.</td>
<td>EA Electronic Arts Inc.</td>
<td></td>
</tr>
<tr>
<td>GM General Motors Company</td>
<td>JPM J P Morgan Chase &amp; Co</td>
<td>Health Care (8)</td>
<td>IBM International Business Machines Corporation</td>
<td>JNJ Johnson &amp; Johnson</td>
<td></td>
</tr>
<tr>
<td>GPS Gap, Inc. (The)</td>
<td>JPM J P Morgan Chase &amp; Co</td>
<td></td>
<td>INTC Intel Corporation</td>
<td>LLY Eli Lilly and Company</td>
<td></td>
</tr>
<tr>
<td>HD Home Depot, Inc. (The)</td>
<td>AET Aetna Inc.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LEN Lennar Corporation</td>
<td>AMGN Amgen Inc.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>BIIB Biogen Inc.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>BMY Bristol-Myers Squibb Company</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CVS Health Corporation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CELG Celgene Corporation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>KO Coca-Cola Company (The)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>LLY Eli Lilly and Company</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>KR Kroger Company (The)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
References


69


