

Hansen, Christian; Liao, Yuan

Working Paper

The Factor-Lasso and K-Step bootstrap approach for inference in high-dimensional economic applications

Working Paper, No. 2016-10

Provided in Cooperation with:

Department of Economics, Rutgers University

Suggested Citation: Hansen, Christian; Liao, Yuan (2016) : The Factor-Lasso and K-Step bootstrap approach for inference in high-dimensional economic applications, Working Paper, No. 2016-10, Rutgers University, Department of Economics, New Brunswick, NJ

This Version is available at:

<https://hdl.handle.net/10419/162932>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

THE FACTOR-LASSO AND K-STEP BOOTSTRAP APPROACH FOR INFERENCE IN HIGH-DIMENSIONAL ECONOMIC APPLICATIONS

CHRISTIAN HANSEN AND YUAN LIAO

ABSTRACT. We consider inference about coefficients on a small number of variables of interest in a linear panel data model with additive unobserved individual and time specific effects and a large number of additional time-varying confounding variables. We allow the number of these additional confounding variables to be larger than the sample size, and suppose that, in addition to unrestricted time and individual specific effects, these confounding variables are generated by a small number of common factors and high-dimensional weakly-dependent disturbances. We allow that both the factors and the disturbances are related to the outcome variable and other variables of interest. To make informative inference feasible, we impose that the contribution of the part of the confounding variables not captured by time specific effects, individual specific effects, or the common factors can be captured by a relatively small number of terms whose identities are unknown. Within this framework, we provide a convenient computational algorithm based on factor extraction followed by lasso regression for inference about parameters of interest and show that the resulting procedure has good asymptotic properties. We also provide a simple k-step bootstrap procedure that may be used to construct inferential statements about parameters of interest and prove its asymptotic validity. The proposed bootstrap may be of substantive independent interest outside of the present context as the proposed bootstrap may readily be adapted to other contexts involving inference after lasso variable selection and the proof of its validity requires some new technical arguments. We also provide simulation evidence about performance of our procedure and illustrate its use in two empirical applications.

JEL classification: C33, C38.

Keywords: panel data, treatment effects, high-dimensional

Christian Hansen, Booth School of Business, University of Chicago, Chicago, IL 60637. Christian.Hansen@chicagobooth.edu. Yuan Liao, Department of Economics, Rutgers University, New Brunswick, NJ 08901. yuan.liao@rutgers.edu. The authors are grateful to Shakheeb Khan, Roger Moon, and seminar participants at the Australasian Meetings of the Econometric Society, University of Chile, National University of Singapore, Xiamen University, University of Toronto, and Stevens Institute of Technology for helpful comments. This material is based upon work supported by the National Science Foundation under Grant No. 1558636 and the University of Chicago Booth School of Business. First version: June 2016. This version: November 28, 2016.

1. INTRODUCTION

Data in which there are many observable variables available for each observation, i.e. “high-dimensional data,” are increasingly common and available for use in empirical applications. Having rich high-dimensional data offers many opportunities for empirical researchers but also poses statistical challenges in that regularization or dimension reduction will generally be needed for informative data analysis. The success of regularized estimation for either forecasting or inference using high-dimensional data relies on using a regularization device that is appropriate for the type of data at hand. Effective regularization imposes substantive restrictions in estimation, and the resulting estimates can perform very poorly, for example suffering from large biases and missing important explanatory power, when the restrictions provide poor approximations to the underlying data generating mechanism. It is thus important to employ regularized estimators that accommodate sensible beliefs about the structure of an underlying econometric model.

Two structures which are common in the econometrics literature are sparse structures and factor structures. To fix ideas, consider the linear regression model

$$y_i = x_i' \theta + \varepsilon_i \tag{1.1}$$

where $i \leq n$ indexes individual observations, y_i is the observed outcome of interest, x_i is a $p \times 1$ vector of observed predictor variables with $p \gg n$ allowed, and ε_i is a regression disturbance. A sparse structure essentially imposes that the number of non-zero elements in θ is small. Intuitively, the sparse structure relies on the belief that the majority of the explanatory power in the observed predictor variables concentrates within a small number of the available variables. Estimators that are appropriate for sparse models, such as the lasso or variable selection procedures, may perform very poorly when the true model is “dense” in the sense that there are many non-zero elements in β that are moderate in magnitude.

A commonly employed version of a linear factor model employs a different structure where

$$y_i = f_i' \xi + \varepsilon_i \tag{1.2}$$

$$x_i = \Lambda f_i + U_i. \tag{1.3}$$

Here f_i denotes a latent $K \times 1$ vector of factors with $K \ll n$ that are important in determining both the observed outcome of interest, y_i , and the observed $p \times 1$, with $p \gg n$, vector of observed predictor variables x_i . Within this structure, one may obtain estimates

of the latent factors and build a model for the outcome given the extracted factors; see, e.g. Bai (2003), Bai and Ng (2002), Stock and Watson (2002) and Fan et al. (2016). The basic factor model differs markedly from the sparse linear model (1.1). Importantly, data generated from model (1.2)-(1.3) would generally result in a dense coefficient vector θ in the regression of y_i onto x_i , and sparsity based estimation strategies would tend to perform poorly as a result. Of course, if the data generated by the sparse model (1.1), common factors will generally not capture the explanatory power, which loads on a small number of the raw regressors, and pure factor-based estimation will perform poorly.

In this paper, we propose a simple model that nests both the sparsity-based and factor-based structures. The model allows for the observed predictors to have a factor structure but then allows both the factors and the factor residuals, the U_i in equation (1.3), to load in the outcome equation. That is, we replace (1.2) with

$$y_i = f_i' \xi + U_i' \theta + \varepsilon_i \quad (1.4)$$

and impose that θ is sparse. This model allows for the fact that all of the relevant explanatory power in the predictors may not be captured entirely by the factors but imposes that any predictive power not captured by the factors concentrates on only a few elements of the high-dimensional covariate vector. (1.4) clearly reduces to (1.1) when there is no factor structure in x and reduces to (1.2) when $\theta = 0$. We note that this model shares much in common with factor augmented regression models, e.g. Bai and Ng (2006) and Bernanke et al. (2005), with the key points of departure being that we do not assume the identity of the additional variables to include in the model is known and that U is not observable. Hahn et al. (2013) consider a model that shares the essential structure of (1.4) and (1.3) from a Bayesian standpoint. They show that forecasts obtained from their Bayesian estimator of this model tend to outperform forecasts obtained based on either pure sparsity or pure factor based models.

The first key contribution of the present paper is offering a practical estimation and inference procedure that is appropriate for inference in a panel generalization of the model given by equations (1.4) and (1.3) and providing a formal treatment of the procedure's theoretical properties. Specifically, we proceed by first running a factor extraction step and taking residuals from regressing each observed variable on the estimated factors. Using these residuals, we then follow the lasso-based estimation and inference procedures of Belloni et al. (2015). We show that the resulting estimator of parameters of interest specified ex ante by the researcher is asymptotically normal with readily estimated asymptotic variance under sensible conditions. These conditions allow for errors in selection of the

elements of the covariate vector that load after controlling for the factors but maintain sufficiently strong conditions to allow oracle selection of the number of factors. The theoretical analysis is substantially complicated by the fact that factors and factor-residuals are not observed and must be extracted from the data. The estimation error in this extraction then enters the second step nonlinear and non-smooth lasso problem. Due to this complication, the theoretical results in this paper make use of arguments that, to our knowledge, are not implied by results existing in the current factor modeling literature or the current lasso literature. These results may be of some interest outside of the context of establishing the properties of our proposed inferential procedure.

By addressing estimation and inference in an interesting high-dimensional factor augmented regression model appropriate for panel data, our paper contributes to the rapidly growing literature dealing with obtaining valid inferential statements following regularized estimation. See, for example, Belloni et al. (2012, 2013a,b, 2014a,b, 2015), Berk et al. (2013), Chernozhukov et al. (2016), Dezeure et al. (2016), Fan and Li (2001), Fan and Lv (2011), Farrell (2015), Gautier and Tsybakov (2011), G'Sell et al. (2013), Fithian et al. (2014), Javanmard and Montanari (2014), Kozbur (2015), Lee and Taylor (2014), Lee et al. (2016), Lockhart et al. (2014), Loftus and Taylor (2014), Taylor et al. (2014), van de Geer et al. (2014), Wager and Athey (2015), and Zhang and Zhang (2014) for approaches to obtaining valid inferential statements in a variety of different high-dimensional settings.

As a second main contribution, we offer a new, computationally convenient bootstrap method for inference. Specifically, we consider a bootstrap where we apply our main procedure, including extraction of factors and lasso estimation steps, within each bootstrap replication. As computation of the lasso estimator within each bootstrap sample may be demanding, we explicitly consider a k -step bootstrap following Andrews (2002) where we start at the lasso solution from the full sample and then iterate a numeric solution algorithm for the lasso estimator for k -steps. We make use of solution algorithms for which the updates are available in closed form which leads to fast computation. We provide high-level conditions under which the procedure provides asymptotically valid inference for parameters of interest and provide specific examples with lower level conditions. The k -step bootstrap we propose complements other bootstrap procedures that have been proposed for lasso-based inference, for example, Belloni et al. (2014a), Chatterjee and Lahiri (2011), Chernozhukov et al. (2013), and Dezeure et al. (2016). In particular, the approach we take is something of a middle ground between Chernozhukov et al. (2013), which uses resampling of model scores to avoid recomputation of the lasso estimator, and Dezeure et al. (2016) which fully recomputes the lasso solution within each bootstrap replication. The former

approach is extremely computationally convenient and asymptotically valid but does not capture any finite sample uncertainty introduced in the lasso selection, while the latter may be computationally cumbersome due to fully recomputing the lasso solution within each iteration. We note that the bootstrap procedure could be easily applied outside of the specific model considered in this paper and that the technical analysis here is new and may be of interest outside of the present context.

The remainder of this paper is organized as follows. In Section 2, we describe the panel factor-lasso model and outline the basic algorithm we will employ for inference. We present formal results for the proposed procedure in Section 3, providing regularity conditions under which the estimator of parameters of interest is asymptotically normal and valid confidence statements may be obtained. Section 4 describes the k-step bootstrap approach in detail and provides a formal analysis establishing the validity of the resulting bootstrap inference. Section 5 discusses the factor extraction part of the problem in more detail and provides examples with accompanying low-level conditions that are sufficient for the high-level conditions stated in Section 3. We then provide simulation and empirical examples that motivate the model we consider and illustrate the use of the estimation procedure in Section 6. Key proofs are collected in an appendix with additional results provided in a supplementary appendix.

Throughout the paper, we use $\|\beta\|_1$ and $\|\beta\|_2$ to respectively denote the ℓ_1 - and ℓ_2 -norms of a vector β ; use $\|A\|$ and $\|A\|_F$ to respectively denote the spectral and Frobenius norms of a matrix A . In addition, denote by $|J|_0$ as the cardinality of a finite set J . Finally, for two positive sequences a_n, b_n , we write $a_n \asymp b_n$ if $a_n = O(b_n)$ and $b_n = O(a_n)$.

2. PANEL FACTOR-LASSO MODEL AND ALGORITHM

2.1. Panel Partial Factor Model. Consider the linear panel model defined by

$$y_{it} = \alpha d_{it} + \xi'_t f_i + U'_{it} \theta + g_i + \nu_t + \epsilon_{it} \tag{2.1}$$

$$d_{it} = \delta'_{dt} f_i + U'_{it} \gamma_d + \zeta_i + \mu_t + \eta_{it} \tag{2.2}$$

$$X_{it} = \Lambda_t f_i + w_i + \rho_t + U_{it} \tag{2.3}$$

where $i \leq n$ indexes cross-sectional observations, $t \leq T$ indexes time series observations, X_{it} are observed potentially confounding variables, and d_{it} is an a priori specified ‘‘treatment’’

variable of interest.¹ f_i is a $K \times 1$ vector of latent factors with time-varying $K \times 1$ factor loading vectors ξ_t , δ_{dt} and $p \times K$ dimensional factor-loading matrix Λ_t . We will take asymptotics where $\dim(X_{it}) = p \rightarrow \infty$, $n \rightarrow \infty$, and T is either fixed or growing slowly relative to n and p when stating our formal results, and we explicitly allow for scenarios where $p \gg nT$. K is assumed fixed throughout the paper. Our object of interest is the parameter α on the variable of interest d_{it} . Following Hahn et al. (2013), we refer to the model (2.1)-(2.3) as the ‘‘panel partial factor model’’ (PPFM).²

In each equation, we also allow for additive unobserved individual effects, (g_i, ζ_i, w'_i) , and time specific effects, (ν_t, μ_t, ρ'_t) , where g_i , ζ_i , ν_t , and μ_t are scalars and w_i and ρ_t are $p \times 1$ vectors. We do not impose structure over the individual or time specific effects and thus treat them as fixed effects. This treatment differentiates the common factors, f_i , from the additive heterogeneity (g_i, ζ_i, w'_i) and (ν_t, μ_t, ρ'_t) as we impose that the f_i are common to each observed series with common, time-varying loadings. Term U_{it} represents the part of the observed X_{it} that is orthogonal to the factors and unobserved time and individual specific heterogeneity. We allow U_{it} to be correlated to both the outcome and variable of interest after controlling for the factors and individual and time fixed effects. Because $p \gg nT$, we assume that θ and γ_d are approximately sparse vectors. We assume that observed right-hand side variables are strictly exogenous so that $E[\eta_{it}|X_{i1}, \dots, X_{iT}] = 0$ and $E[\epsilon_{it}|X_{i1}, \dots, X_{iT}, d_{i1}, \dots, d_{iT}] = 0$. We will assume that data are iid across i but allow for dependence across time periods, t . Finally, we note that while we treat the PPFM defined in (2.1)-(2.3) in the formal analysis, the results clearly apply to models without additive fixed effects or to a single cross-section.³

As noted in the Introduction, the PPFM generalizes the high-dimensional sparse fixed effects model examined in Belloni et al. (2015) and conventional large-dimensional factor models and factor augmented regression models; e.g. Bai and Ng (2006). The PPFM is also related to, but distinct from, interactive fixed effects models as in, for example, Bai (2009); Bai and Li (2014), Moon and Weidner (2015a,b), Pesaran (2006) and Su and Chen

¹Our results will immediately apply to the case where d_{it} is an $r \times 1$ vector with r fixed. The analysis could also be extended to handle unbalanced panels where observations are missing at random. We omit both cases for convenience.

²Hahn et al. (2013) consider a similar structure to (2.1)-(2.3) which excludes the individual and time effects and imposes that the ϵ_{it} are i.i.d. Gaussian innovations. They refer to this model as a partial factor model.

³We consider a cross-sectional instrumental variables version of the model in both a simulation and an empirical example.

(2013).⁴ A simple version of the interactive fixed effects model analogous to (2.1) is

$$y_{it} = \alpha d_{it} + z'_{it}\beta + \lambda_t f_i + \epsilon_{it}.$$

In this model, z_{it} represents a known, low-dimensional set of variables that must be controlled for in addition to the factors in f_i . There appear to be three key distinctions between the high-dimensional PPFM and interactive fixed effects approaches. First, we relax the assumption that one knows the exact identity of the variables that should appear in the model, z_{it} , by allowing for a high-dimensional set of observed potential confounds in X_{it} . Second, we allow for the fact that the relevant explanatory power in the predictors may not be captured entirely by the factors, but impose that any predictive power not captured by the factors concentrates on only a few elements of the high-dimensional vector U . Third, we directly extract estimates of the factors and U from X which can proceed even when T is small. Approaches to estimating the interactive fixed effects structure rely on having a large number of observations in both the time series and cross-sectional dimensions. We thus view the PPFM and interactive fixed effects approaches as complementary where one may prefer one or the other depending on the nature of the data at hand.

2.2. Estimation Algorithm. To estimate α , we begin by taking the within transformation of all observed variables to remove the fixed effects. To this end, let

$$\tilde{z}_{it} = z_{it} - \bar{z}_{.t} - \bar{z}_i + \bar{\bar{z}}$$

for any variable z_{it} where $\bar{z}_{.t} = \frac{1}{n} \sum_{i=1}^n z_{it}$, $\bar{z}_i = \frac{1}{T} \sum_{t=1}^T z_{it}$, and $\bar{\bar{z}} = \frac{1}{nT} \sum_{i=1, t=1}^{n, T} z_{it}$. We can then define a demeaned model as

$$\tilde{y}_{it} = \alpha \tilde{d}_{it} + \tilde{\xi}'_t \tilde{f}_i + \tilde{U}'_{it} \theta + \tilde{\epsilon}_{it} \tag{2.4}$$

$$\tilde{d}_{it} = \tilde{\delta}'_{dt} \tilde{f}_i + \tilde{U}'_{it} \gamma_d + \tilde{\eta}_{it}, \tag{2.5}$$

$$\tilde{X}_{it} = \tilde{\Lambda}_t \tilde{f}_i + \tilde{U}_{it}. \tag{2.6}$$

After removing the additive unobserved heterogeneity, we estimate the (demeaned) latent factors as well as the (demeaned) idiosyncratic components from the model $\tilde{X}_{it} = \tilde{\Lambda}'_t \tilde{f}_i + \tilde{U}_{it}$.⁵ Let $\hat{F} = (\hat{f}_1, \dots, \hat{f}_n)'$ be the $n \times K$ matrix of estimated factors. We shall discuss

⁴See also Bonhomme and Manresa (2015) for a distinct but related approach based on a grouped fixed effects model.

⁵We note that recovering the untransformed f_i and U_{it} would only be possible with large n and T due to the presence of the unrestricted fixed effects. Fortunately, recovering these quantities is unnecessary within the model with common coefficients θ , γ_d , and α as only \tilde{f}_i and \tilde{U}_{it} appear in the equations of interest. This simplification would not generally occur if we allowed heterogeneity in θ , γ_d , or α over time or across

some examples of \widehat{F} in Section 5. Given \widehat{F} , we estimate $\widehat{\Lambda}_t$ and \widehat{U}_{it} by least squares:

$$\widehat{\Lambda}_t = \sum_{i=1}^n \widetilde{X}_{it} \widehat{f}'_i (\widehat{F}' \widehat{F})^{-1}, \quad \widehat{U}_{it} = \widetilde{X}_{it} - \widehat{\Lambda}_t \widehat{f}_i, \quad i \leq n, t \leq T. \quad (2.7)$$

Substituting (2.5) to (2.4), we obtain

$$\begin{aligned} \widetilde{y}_{it} &= \alpha(\widetilde{\delta}'_{dt} \widetilde{f}_i + \widetilde{U}'_{it} \gamma_d + \widetilde{\eta}_{it}) + \widetilde{\xi}'_t \widetilde{f}_i + \widetilde{U}'_{it} \theta + \widetilde{\epsilon}_{it} \\ &:= \widetilde{\delta}'_{yt} \widetilde{f}_i + \widetilde{U}'_{it} \gamma_y + \widetilde{\epsilon}_{it}. \end{aligned}$$

Now let $\widetilde{Y}_t = (\widetilde{y}_{1t}, \dots, \widetilde{y}_{nt})'$ and $\widetilde{D}_t = (\widetilde{d}_{1t}, \dots, \widetilde{d}_{nt})'$ denote the vectors of outcome and treatment variable within each time period t . We next regress \widetilde{Y}_t and \widetilde{D}_t onto the extracted factors \widehat{F} time period by time period to obtain $\{\widehat{\delta}_{yt}\}_{t=1}^T$ and $\{\widehat{\delta}_{dt}\}_{t=1}^T$ for

$$\widehat{\delta}_{yt} = (\widehat{F}' \widehat{F})^{-1} \widehat{F}' \widetilde{Y}_t \text{ and } \widehat{\delta}_{dt} = (\widehat{F}' \widehat{F})^{-1} \widehat{F}' \widetilde{D}_t. \quad (2.8)$$

We then run the lasso with the residuals from each of these factor regressions as dependent variable and the estimated factor disturbances \widehat{U}_{it} as predictors. That is, we obtain

$$\widetilde{\gamma}_y = \arg \min_{\gamma \in \mathbb{R}^p} \frac{1}{nT} \sum_{t=1}^T \sum_{i=1}^n (\widetilde{y}_{it} - \widehat{\delta}'_{yt} \widehat{f}_i - \widehat{U}'_{it} \gamma)^2 + \kappa_n \|\widehat{\Psi}^y \gamma\|_1, \quad (2.9)$$

$$\widetilde{\gamma}_d = \arg \min_{\gamma \in \mathbb{R}^p} \frac{1}{nT} \sum_{t=1}^T \sum_{i=1}^n (\widetilde{d}_{it} - \widehat{\delta}'_{dt} \widehat{f}_i - \widehat{U}'_{it} \gamma)^2 + \kappa_n \|\widehat{\Psi}^d \gamma\|_1. \quad (2.10)$$

where the tuning parameter κ_n is chosen as, for some $c_0 > 1$ and $q_n \rightarrow 0$,

$$\kappa_n = \frac{2c_0}{\sqrt{nT}} \Phi^{-1}(1 - q_n/(2p)), \quad \log(q_n^{-1}) = O(\log p) \quad (2.11)$$

and $\widehat{\Psi}^y$ and $\widehat{\Psi}^d$ are diagonal penalty loading matrices. Given the fixed effects panel structure, we use the clustered penalty loadings of Belloni et al. (2015) which have diagonal elements defined as

$$[\widehat{\Psi}^y]_{j,j} = \sqrt{\frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T \sum_{t'=1}^T \widehat{U}_{it,j} \widehat{U}_{it',j} \widehat{\epsilon}_{it} \widehat{\epsilon}_{it'}} \quad (2.12)$$

$$[\widehat{\Psi}^d]_{j,j} = \sqrt{\frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T \sum_{t'=1}^T \widehat{U}_{it,j} \widehat{U}_{it',j} \widehat{\eta}_{it} \widehat{\eta}_{it'}} \quad (2.13)$$

individuals, and we would need to consider incidental parameters bias introduced by removing the additive fixed effects. We leave exploration of this issue to future research.

where \widehat{e}_{it} is an estimator of $\tilde{e}_{it} = \tilde{y}_{it} - \tilde{\delta}'_{yt}\tilde{f}_i - \tilde{U}'_{it}\gamma_y$ and $\widehat{\eta}_{it}$ is an estimator of $\tilde{\eta}_{it} = \tilde{d}_{it} - \tilde{\delta}'_{dt}\tilde{f}_i - \tilde{U}'_{it}\gamma_d$.⁶

For the final step, we adopt the post-double-selection procedure of Belloni et al. (2014b). Let $\widehat{J} = \{j \leq p : \tilde{\gamma}_{y,j} \neq 0\} \cup \{j \leq p : \tilde{\gamma}_{d,j} \neq 0\}$, and let $\widehat{U}_{it,\widehat{J}}$ be a subvector of \widehat{U}_{it} whose elements are $\{\widehat{U}_{it,j} : j \in \widehat{J}\}$. We then run the regression of $\tilde{y}_{it} - \widehat{\delta}'_{yt}\widehat{f}_i$ on $\widehat{U}_{it,\widehat{J}}$ and $\tilde{d}_{it} - \widehat{\delta}'_{dt}\widehat{f}_i$ on $\widehat{U}_{it,\widehat{J}}$ and obtain

$$\widehat{\gamma}_y = \left(\sum_{i=1}^n \sum_{t=1}^T \widehat{U}_{it,\widehat{J}} \widehat{U}'_{it,\widehat{J}} \right)^{-1} \sum_{i=1}^n \sum_{t=1}^T \widehat{U}_{it,\widehat{J}} (\tilde{y}_{it} - \widehat{\delta}'_{yt}\widehat{f}_i), \quad (2.14)$$

$$\widehat{\gamma}_d = \left(\sum_{i=1}^n \sum_{t=1}^T \widehat{U}_{it,\widehat{J}} \widehat{U}'_{it,\widehat{J}} \right)^{-1} \sum_{i=1}^n \sum_{t=1}^T \widehat{U}_{it,\widehat{J}} (\tilde{d}_{it} - \widehat{\delta}'_{dt}\widehat{f}_i). \quad (2.15)$$

The final estimator of α is then given by

$$\widehat{\alpha} = \left(\sum_{i=1}^n \sum_{t=1}^T \widehat{\eta}_{it}^2 \right)^{-1} \sum_{i=1}^n \sum_{t=1}^T \widehat{\eta}_{it} \widehat{e}_{it} \quad (2.16)$$

where $\widehat{e}_{it} = \tilde{y}_{it} - \widehat{\delta}'_{yt}\widehat{f}_i - \widehat{U}'_{it,\widehat{J}}\widehat{\gamma}_y$ and $\widehat{\eta}_{it} = \tilde{d}_{it} - \widehat{\delta}'_{dt}\widehat{f}_i - \widehat{U}'_{it,\widehat{J}}\widehat{\gamma}_d$ are the residuals from the regressions specified in (2.14) and (2.15).

The estimator $\widehat{\alpha}$ can be expressed more compactly in matrix form. Write

$$\begin{aligned} \tilde{Y} &= \begin{pmatrix} \tilde{Y}_1 \\ \vdots \\ \tilde{Y}_T \end{pmatrix}_{nT \times 1}, & \tilde{D} &= \begin{pmatrix} \tilde{D}_1 \\ \vdots \\ \tilde{D}_T \end{pmatrix}_{nT \times 1}, & \widehat{U}_{\widehat{J}} &= \begin{pmatrix} \widehat{U}_{1,\widehat{J}} \\ \vdots \\ \widehat{U}_{T,\widehat{J}} \end{pmatrix}_{nT \times |\widehat{J}|_0}, \\ \widehat{e} &= \begin{pmatrix} \widehat{e}_1 \\ \vdots \\ \widehat{e}_T \end{pmatrix}_{nT \times 1}, & \text{and } \widehat{\eta} &= \begin{pmatrix} \widehat{\eta}_1 \\ \vdots \\ \widehat{\eta}_T \end{pmatrix}_{nT \times 1}. \end{aligned}$$

In addition, for a matrix A , define $M_A = I - A(A'A)^-A'$, where $(A'A)^-$ represents a generalized inverse of $A'A$. Then it is straightforward to verify that

$$\widehat{e} = M_{\widehat{U}_{\widehat{J}}}(I_T \otimes M_{\widehat{F}})\tilde{Y}, \text{ and } \widehat{\eta} = M_{\widehat{U}_{\widehat{J}}}(I_T \otimes M_{\widehat{F}})\tilde{D}$$

⁶We obtain \widehat{e}_{it} and $\widehat{\eta}_{it}$ through an iterative algorithm similar to that of Belloni et al. (2014b), which starts from a preliminary estimate. In addition, we use $c_0 = 1.1$ and $q_n = .1/\log(n)$ in the simulation and empirical examples.

are the estimated residuals $(\tilde{e}'_1, \dots, \tilde{e}'_T)'$ and $(\tilde{\eta}'_1, \dots, \tilde{\eta}'_T)'$ defined above. Then

$$\hat{\alpha} = (\hat{\eta}'\hat{\eta})^{-1}\hat{\eta}'\hat{e}.$$

Note that the estimator $\hat{\alpha}$ is numerically equivalent to the coefficient on \tilde{d}_{it} in the regression of \tilde{y}_{it} on \tilde{d}_{it} , \hat{f}_i interacted with time dummy variables, and $\hat{U}_{it,\hat{J}}$. In Theorem 3.1 of the next section, we verify that inference for $\hat{\alpha}$ can proceed using the output from this OLS regression as long as clustered standard errors (e.g. Arellano (1987), Bertrand et al. (2004), and Hansen (2007)) are used.

The following algorithm summarizes the estimation strategy detailed above.

Algorithm (**Factor-Lasso Estimation of α** .)

- (1) Obtain $\{\hat{f}_i, \hat{U}_{it}\}_{i \leq n, t \leq T}$ by extracting factors from the model $\tilde{X}_{it} = \tilde{\Lambda}'_t \tilde{f}_i + \tilde{U}_{it}$.
- (2) For $\hat{\delta}_{yt}$ and $\hat{\delta}_{dt}$ defined in (2.8), run the cluster-lasso programs (2.9) and (2.10) to obtain $\tilde{\gamma}_y$ and $\tilde{\gamma}_d$.
- (3) Obtain the estimator $\hat{\alpha}$ and corresponding estimated standard error as the coefficient on $\tilde{d}_{it} - \hat{\delta}'_{dt} \hat{f}_i$ and associated clustered standard error from the regression of $\tilde{y}_{it} - \hat{\delta}'_{yt} \hat{f}_i - \hat{U}'_{it,\hat{J}} \tilde{\gamma}_y$ on $\tilde{d}_{it} - \hat{\delta}'_{dt} \hat{f}_i - \hat{U}'_{it,\hat{J}} \tilde{\gamma}_d$ where $\hat{U}_{it,\hat{J}}$ is the subvector of \hat{U}_{it} whose elements are $\{\hat{U}_{it,j} : j \in \hat{J}\}$.

3. ASSUMPTIONS AND ASYMPTOTIC THEORY

In this section, we present a set of sufficient conditions under which we establish asymptotic normality of $\hat{\alpha}$ and provide a consistent estimator of its asymptotic variance. Throughout we consider sequences of data generating processes (DGPs) where p increases as n and T increase and where model parameters are allowed to depend on n and T . We suppress this dependence for notational simplicity. We use the term “absolute constants” to mean given constants that do not depend on the DGP.

3.1. Regularity Conditions. Write $\epsilon_t = (\epsilon_{1t}, \dots, \epsilon_{nt})'$, $\eta_t = (\eta_{1t}, \dots, \eta_{nt})'$, and $U_t = (U'_{1t}, \dots, U'_{nt})'$. Similarly, let $\epsilon_i = (\epsilon_{i1}, \dots, \epsilon_{iT})'$ and $\eta_i = (\eta_{i1}, \dots, \eta_{iT})'$, $U_i = (U'_{i1}, \dots, U'_{iT})'$.

Our first two conditions collect various restrictions on dependence, tail behavior, and moments of the unobserved features of the model. We assume there are positive absolute constants C_1, C_2 and C_3 such that the following assumption holds.

Assumption 3.1 (DGP). (i) $\{f_i, \eta_i, \epsilon_i, U_i\}_{i \leq n}$ are independent and identically distributed across $i = 1, 2, \dots, n$ and satisfy

$$E(\eta_i | \epsilon_i, U_i, f_i) = 0, \quad E(\epsilon_i | \eta_i, U_i, f_i) = 0, \quad E(U_i | \eta_i, \epsilon_i, f_i) = 0.$$

In addition, given $\{f_i\}_{i \leq n}$, the sequence $\{U_i, \eta_i, \epsilon_i\}_{i \leq n, t \leq T}$ is also conditionally independent across i .

(ii) Given $\{f_i\}_{i \leq n}$, the sequence $\{U_t, \eta_t, \epsilon_t\}_{t \leq T}$ is stationary across t , and satisfies a strong-mixing condition. That is, there exists an absolute constant $r > 0$ such that for all $T \in \mathbb{R}^+$,

$$\sup_{A \in \mathcal{F}_{-\infty}^0, B \in \mathcal{F}_T^\infty} |P(A)P(B) - P(AB)| \leq \exp(-C_1 T^r),$$

where $\mathcal{F}_{-\infty}^0$ and \mathcal{F}_T^∞ denote the σ -algebras generated by $\{(U_t, \eta_t, \epsilon_t) : -\infty \leq t \leq 0\}$ and $\{(U_t, \eta_t, \epsilon_t) : T \leq t \leq \infty\}$ respectively.

(iii) Almost surely,

$$\max_{i \leq n, m \leq p, t \leq T} \sum_{k=1}^p \sum_{s=1}^T |E(U_{it,k} U_{is,m} | f_i, \epsilon_i, \eta_i)| < C_2.$$

(iv) For any $s > 0$, $i \leq n$, $j \leq p$ and $k \leq K$,

$$\begin{aligned} P(|U_{it,j}| > s) &\leq \exp(-C_3 s^2), & P(|f_{ik}| > s) &\leq \exp(-C_3 s^2), \\ P(|\eta_{it}| > s) &\leq \exp(-C_3 s^2), & P(|\epsilon_{it}| > s) &\leq \exp(-C_3 s^2). \end{aligned}$$

(v) Let θ_m and $\gamma_{d,m}$ be the m^{th} entries of θ and γ_d , and λ'_{tm} be the m^{th} row of Λ_t .

$$|\alpha| + \max_{t \leq T} (\|\xi_t\| + \|\delta_{dt}\|) + \max_{m \leq p} (|\theta_m| + |\gamma_{d,m}|) + \max_{m \leq p, t \leq T} \|\lambda_{tm}\| < C_2.$$

Assumption 3.1 collects reasonably standard regularity conditions that restrict the dependence across observations and tail behavior of random variables. These conditions impose that the unobserved variables in the model are cross-sectionally independent, are weakly dependent and stationary in the time series, and have sub-Gaussian tails. Assumption 3.1.(iii) further imposes weak conditional dependence in the factor residuals, U_{it} . In the simple case where U_{it} is independent of f_i , η_i , and ϵ_i for all t , this condition reduces to weak intertemporal correlation and no strong dependence among the columns of U_{it} . Importantly, it does not imply that all correlation among the observed X_{it} is captured by factors but allows for the presence of a rich covariance structure in the part of X_{it} that is not

linearly explained by the factors. The condition also allows for some dependence between “control” variables U_{it} and structural unobservables η_i and ϵ_i but restricts the magnitude of any such dependence so that it is asymptotically negligible. Finally, condition (v) requires that all the low dimensional parameters are well bounded.

Recall that $e_{it} = \alpha\eta_{it} + \epsilon_{it}$.

Assumption 3.2 (Moment bounds). *For $m \leq p, i \leq n, t \leq T$, define*

$$W_{im} = \frac{1}{\sqrt{T}} \sum_{t=1}^T (U_{it,m} - \bar{U}_{i,\cdot,m})(e_{it} - \bar{e}_{i,\cdot}).$$

There are absolute constants $c, C > 0$, such that

(i) $\max_{i \leq n, m \leq p} E|W_{im}|^3 \leq C$ and $c < \min_{i \leq n, m \leq p} EW_{im}^2 \leq \max_{i \leq n, m \leq p} EW_{im}^2 < C$, and

$$\text{Var} \left(\frac{1}{\sqrt{nT}} \sum_{i=1}^n \sum_{t=1}^T (\eta_{it} - \bar{\eta}_{i,\cdot})(\epsilon_{it} - \bar{\epsilon}_{i,\cdot}) \right) > c.$$

(ii) *almost surely in $F = (f_1, \dots, f_n)'$,*

$$\max_{m \leq p, t \leq T} \frac{1}{n} \sum_{i=1}^n E(U_{it,m}^8 | F) < C, \quad \max_{t \leq T} \frac{1}{n} \sum_{i=1}^n E(e_{it}^8 | F) < C.$$

Assumption 3.2 collects additional high-level moment bounds. The bounds on moments of normalized sums in Condition (i) could be established under a variety of sufficient lower level conditions. Condition (ii) places restrictions on the dependence between $\{U_{it}, e_{it}\}_{i=1, t=1}^{n, T}$ and $\{f_i\}_{i=1}^n$.

Before stating the next assumption, we decompose the high dimensional coefficients as

$$\gamma_y = \underbrace{\gamma_y^0}_{\text{exactly sparse}} + \underbrace{R_y}_{\text{remainder}} \quad \text{and} \quad \gamma_d = \underbrace{\gamma_d^0}_{\text{exactly sparse}} + \underbrace{R_d}_{\text{remainder}}$$

where γ_y^0 and γ_d^0 are sparse vectors that approximate the potentially dense true coefficient vectors γ_y and γ_d and R_y and R_d represent approximation errors. Let $J = \{j \leq p : \gamma_{y,j}^0 \neq 0\} \cup \{j \leq p : \gamma_{d,j}^0 \neq 0\}$ be the union of the support of the exactly sparse components.

Assumption 3.3 (Rate Conditions). (i) $\|R_d\|_1 + \|R_y\|_1 = o(\sqrt{\frac{\log p}{nT}})$.

(ii) $|J|_0^2 \log^3(p) = O(n)$.

(iii) $|J|_0^2 T = o(n)$. *In addition, the number of factors, K , stays constant.*

Assumption 3.3 collects restrictions on the quality of the approximation provided by γ_y^0 and γ_d^0 and rates of growth of model complexity as measured by J and p and sample sizes in the cross-sectional and time series dimension. Condition (iii) imposes the somewhat nonstandard requirement that T be much smaller than n . The need for this condition arises from the fact that we need to obtain high-quality estimates of the idiosyncratic term in the factor equation, U_{it} , which depends on accurately estimating both the unknown factors and the loadings. Estimating the loading matrix Λ_t well for any given t requires a relatively large n , and we thus require T to be smaller than n as the number of unknown loading matrices $\{\Lambda_t\}_{t \leq T}$ is $O(T)$.

Our next assumption restricts the covariance matrix of the within-transformed factor residuals \tilde{U}_{it} .

Assumption 3.4. For any $\delta \in \mathbb{R}^p / \{0\}$, write

$$\mathcal{R}(\delta) = \frac{\delta' \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T \tilde{U}_{it} \tilde{U}_{it}' \delta}{\delta' \delta}.$$

Define restricted and sparse eigenvalue constants:

$$\begin{aligned} \underline{\phi}(m) &= \inf_{\delta \in \mathbb{R}^p: \|\delta_{J^c}\|_1 \leq m \|\delta_J\|_1} \mathcal{R}(\delta), \\ \phi_{\min}(m) &= \inf_{\delta \in \mathbb{R}^p: 1 \leq \|\delta\|_0 \leq m} \mathcal{R}(\delta), \\ \phi_{\max}(m) &= \sup_{\delta \in \mathbb{R}^p: 1 \leq \|\delta\|_0 \leq m} \mathcal{R}(\delta). \end{aligned}$$

(i) (restricted eigenvalue) For any $m > 0$ there is an absolute constant $\underline{\phi} > 0$ so that with probability approaching one,

$$\underline{\phi}(m) > \underline{\phi}.$$

(ii) (sparse eigenvalue) There is a sequence of absolute constants $l_T \rightarrow \infty$ and $c_1, c_2 > 0$ so that with probability approaching one,

$$c_1 < \phi_{\min}(l_T | J|_0) \leq \phi_{\max}(l_T | J|_0) < c_2.$$

Maintaining Assumptions 3.1-3.3, a simple sufficient condition for Assumption 3.4 is that all the eigenvalues of $\frac{1}{nT} \sum_i \sum_t E(U_{it} - \bar{U}_{i\cdot})(U_{it} - \bar{U}_{i\cdot})'$ are well bounded. This is a typical condition in high-dimensional approximate factor models (e.g., Bai (2003); Stock and Watson (2002)). It ensures that the idiosyncratic components are weakly dependent and therefore the decomposition $\tilde{X}_{it} = \tilde{\Lambda}'_t \tilde{f}_i + \tilde{U}_{it}$ is asymptotically identified (as $p \rightarrow \infty$).

Finally, we present high-level conditions on the accuracy of \widehat{F} in Assumption 3.5. The high-level conditions potentially allow for many estimators of the factors, and we verify that these conditions hold under more primitive assumptions for the case of estimating the factors using PCA in Appendix C.

Assumption 3.5 (Quality of Factor Estimation in Original Data). *Suppose there is an invertible $\dim(f_i) \times \dim(f_i)$ matrix H with $\|H\| + \|H^{-1}\| = O_P(1)$, and non-negative sequences $\Delta_F, \Delta_{eg}, \Delta_{ud}, \Delta_{fum}, \Delta_{fe}, \Delta_{\max}$, so that for $\tilde{z}_{it} \in \{\tilde{\epsilon}_{it}, \tilde{\eta}_{it}\}$, $\tilde{w}_{tm} \in \{\tilde{\Lambda}'_t \gamma_d, \tilde{\Lambda}'_t \gamma_y, \tilde{\delta}_{dt}, \tilde{\delta}_{yt}, \tilde{\lambda}_{tm}\}$, $\tilde{h}_{tk} \in \{\tilde{\delta}_{dt}, \tilde{\delta}_{yt}, \tilde{\lambda}_{tk}\}$, and $\gamma \in \{\gamma_d, \gamma_y\}$,*

$$\begin{aligned} \max_{i \leq n} \|\widehat{f}_i - H' \tilde{f}_i\|_2 &= O_P(\Delta_{\max}), \quad \frac{1}{n} \sum_{i=1}^n \|\widehat{f}_i - H' \tilde{f}_i\|_2^2 = O_P(\Delta_F^2) \\ \frac{1}{T} \sum_{t=1}^T \left\| \frac{1}{n} \sum_{i=1}^n (\widehat{f}_i - H' \tilde{f}_i) \tilde{z}_{it} \right\|_2^2 &= O_P(\Delta_{fe}^2), \\ \max_{m \leq p} \left\| \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T (\widehat{f}_i - H' \tilde{f}_i) \tilde{z}_{it} \tilde{w}'_{tm} \right\|_F &= O_P(\Delta_{eg}), \\ \max_{m, k \leq p} \left\| \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T (\widehat{f}_i - H' \tilde{f}_i) \tilde{U}_{it,m} \tilde{h}'_{tk} \right\|_F &= O_P(\Delta_{ud}), \\ \max_{m \leq p, t \leq T} \left\| \frac{1}{n} \sum_{j=1}^n (\widehat{f}_j - H' \tilde{f}_j) \tilde{U}_{jt,m} \right\|_2 &= O_P(\Delta_{fum}). \end{aligned}$$

These sequences satisfy the following restrictions:

$$\begin{aligned} \sqrt{nT} |J|_0^2 \Delta_F^2 &= o(1), \quad \Delta_{eg} = o\left(\frac{1}{\sqrt{nT}}\right), \quad \Delta_{ud} = o\left(\sqrt{\frac{\log p}{nT}}\right), \quad |J|_0^2 \sqrt{\log p} \Delta_{ud} = o(1), \\ \Delta_{fum}^2 &= o\left(\frac{\log p}{T |J|^2 \log(pT)}\right), \quad \Delta_{fe}^2 = o\left(\frac{\log p}{T \log(pT)}\right), \quad \Delta_{\max}^2 = O(\log(n)), \quad \text{and} \\ \Delta_{\max}^2 |J|_0^2 T (\lambda_n^2 |J|_0 + \Delta_F^2 |J|_0^2 + \frac{|J|_0}{n}) &= o(1). \end{aligned}$$

One of the major technical tasks of this paper is to show that the effects of estimating the latent factor and idiosyncratic terms are stochastically dominated by the plug-in tuning parameter κ_n in (2.11). Since $\kappa_n \asymp \sqrt{\frac{\log p}{nT}}$, this is a strong requirement, and gives rise to Assumption 3.5 (and Assumption 4.4 below for the bootstrap sample). Technically, existing results in the literature on estimating factors models are not directly applicable to verify these conditions. In Appendix C, we show that

$$\frac{1}{n} \sum_{i=1}^n \|\widehat{f}_i - H' \tilde{f}_i\|_2^2 = O_P\left(\frac{1}{pT} + \frac{1}{n^2} + \frac{1}{nT^2}\right)$$

when \widehat{f}_t is estimated via PCA. While this result is essentially standard and allows conditions involving Δ_F to be directly verified, however, it does not imply the uniform convergence condition $\max_{t \leq T} \|\widehat{f}_t - H' \tilde{f}_t\|_2$. Nor is this result sufficient to verify the other stated conditions because other terms, e.g. $\Delta_{eg}, \Delta_{fum}, \Delta_{fe}$, involve “weighted averages” of $\{\widehat{f}_i - H' \tilde{f}_i\}$ whose rates of convergence can be derived and shown to be faster than that of $\Delta_F = \frac{1}{pT} + \frac{1}{n^2} + \frac{1}{nT^2}$. For instance, if we use a simple Cauchy-Schwarz inequality to bound Δ_{ud} , we would have

$$\max_{m, k \leq p} \left\| \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T (\widehat{f}_i - H' \tilde{f}_i) \tilde{U}_{it, m} \tilde{h}'_{tk} \right\|_F^2 \leq \frac{1}{n} \sum_{i=1}^n \|\widehat{f}_i - H' \tilde{f}_i\|_2^2 \max_{m, k \leq p} \frac{1}{n} \sum_{i=1}^n \left\| \frac{1}{T} \sum_{t=1}^T \tilde{U}_{it, m} \tilde{h}_{tk} \right\|_2^2.$$

It can be shown that $\max_{m, k \leq p} \frac{1}{n} \sum_{i=1}^n \left\| \frac{1}{T} \sum_{t=1}^T \tilde{U}_{it, m} \tilde{h}_{tk} \right\|_2^2 = O_P\left(\frac{\log p}{T}\right)$, so this crude bound gives us $\Delta_{ud} = \Delta_F \sqrt{\frac{\log p}{T}}$. Unfortunately, this bound is not sharp enough to verify the condition $\Delta_{ud} = o\left(\sqrt{\frac{\log p}{nT}}\right)$ unless $n = o(pT)$. In the special case that T is fixed, requiring $n = o(p)$ is a restrictive condition. Rather than relying on these crude bounds, we achieve sharper bounds by directly deriving the rate of convergence for each required term in Appendix C which relies on some novel technical work. These conditions only require $n = o(p^2T)$ which provides much more freedom on the ratio n/p .

3.2. Main results. The asymptotic variance of $\widehat{\alpha}$ will depend on the quantities

$$\sigma_{\eta\epsilon} = \text{Var} \left(\frac{1}{\sqrt{nT}} \sum_{i=1}^n \sum_{t=1}^T (\eta_{it} - \bar{\eta}_i)(\epsilon_{it} - \bar{\epsilon}_i) \right) \quad \text{and} \quad \sigma_\eta^2 = \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T \text{Var}(\eta_{it} - \bar{\eta}_i)$$

for which

$$\widehat{\sigma}_{\eta\epsilon} = \frac{1}{nT} \sum_{i=1}^n \left(\sum_{t=1}^T \widehat{\eta}_{it} \widehat{\epsilon}_{it} \right)^2 \quad \text{and} \quad \widehat{\sigma}_\eta^2 = \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T \widehat{\eta}_{it}^2$$

are natural estimators. Note that $\widehat{\sigma}_{\eta\epsilon}$ is just the usual clustered covariance estimator with clustering at the individual level.

Theorem 3.1. *Suppose $n, p \rightarrow \infty$, and T is either fixed or growing. Under Assumptions 3.1-3.5,*

$$\sqrt{nT} \sigma_{\eta\epsilon}^{-1/2} \sigma_\eta^2 (\widehat{\alpha} - \alpha) \rightarrow^d \mathcal{N}(0, 1),$$

In addition,

$$\sqrt{nT} \widehat{\sigma}_{\eta\epsilon}^{-1/2} \widehat{\sigma}_\eta^2 (\widehat{\alpha} - \alpha) \rightarrow^d \mathcal{N}(0, 1).$$

Corollary 3.1. *Let \mathcal{P} be a collection of all DGP's such that the assumptions of Theorem 3.1 hold uniformly over all the DGP's in \mathcal{P} . Let $\zeta_\tau = \Phi^{-1}(1 - \tau/2)$. Then as $n, p \rightarrow \infty$,*

and T is either fixed or growing with n , uniformly over $P \in \mathcal{P}$,

$$\lim_{n,p \rightarrow \infty} P \left(\alpha \in \left[\hat{\alpha} \pm \frac{\zeta_\tau}{\sqrt{nT}} \hat{\sigma}_{\eta^\epsilon}^{1/2} \hat{\sigma}_\eta^{-2} \right] \right) = 1 - \tau.$$

The main implication of Theorem 3.1 and Corollary 3.1 is that $\hat{\alpha}$ converges at a \sqrt{nT} rate and that inference may proceed using standard asymptotic confidence intervals and hypothesis tests. Importantly, the inferential results hold uniformly across a large class of approximately sparse models which include cases where perfect selection over which elements of \tilde{U}_{it} enter the model is impossible even in the limit. It is also important to highlight that the conditions on estimation of the factors do rule out the presence of weak factors, and the inferential results do not hold uniformly over sequences of models in which perfect selection of the number of factors and fast convergence of the factors and factor loadings do not hold. The difficulty with handling weak factors arises due to the entry of the estimation errors of the factors in the cluster-lasso problems (2.9) and (2.10) and the non-smooth and highly nonlinear nature of this problem. Extending the results to accommodate the presence of weak factors and imperfect selection of the number of factors would be an interesting direction for further research.

4. k -STEP BOOTSTRAP

In this section, we present a computationally tractable bootstrap procedure that can be used in lieu of the plug-in asymptotic inference formally presented in Theorem 3.1 and Corollary 3.1. While well-developed in low-dimensional settings, there are relatively few formal treatments of bootstrap procedures in high-dimensional settings, though see Chatterjee and Lahiri (2011), Chernozhukov et al. (2013), Belloni et al. (2014a), and Dezeure et al. (2016) for important existing treatments. In the following, we consider a bootstrap procedure which only approximately solves the cluster-lasso problem within each bootstrap replication and thus may remain computationally convenient while also intuitively capturing the sampling variation introduced in the lasso selection.

4.1. The k -Step Bootstrap. Let $D^* = \{\tilde{y}_{it}^*, \tilde{d}_{it}^*, \tilde{X}_{it}^*\}_{i \leq n, t \leq T}$ denote a sample of bootstrap data, and let $\hat{\alpha}^*$ be the estimator obtained by applying the factor-lasso estimator with data D^* . Let B denote the number of bootstrap repetitions.

A potential computational problem with bootstrap procedures for lasso estimation is that one needs to solve B lasso problems where B will typically be fairly large. To circumvent

this problem, we adopt the approach of Andrews (2002) by using the fact that the complete lasso estimator based on the original data, denoted by $\tilde{\gamma}_{lasso}$, should be close to the complete lasso estimator based on bootstrapped data D^* , denoted by $\tilde{\gamma}_{lasso}^*$. Hence, within each bootstrap replication, we can use $\tilde{\gamma}_{lasso}$ as the initial value for solving the lasso problem and iteratively update the estimator for k steps. Denote the resulting k -step bootstrap lasso estimator by $\tilde{\gamma}^*$. We simply use $\tilde{\gamma}^*$ in place of $\tilde{\gamma}_{lasso}^*$ wherever the solution to a lasso problem shows up in the factor-lasso problem. The main result of this section is showing that the k -step bootstrap procedure is first-order valid for statistical inference about α as long as the minimization error after k steps is less than the statistical error (i.e. $o_{P^*}((nT)^{-1/2})$).

The substantive difference between the present context and Andrews (2002) is that Andrews (2002) makes use of Newton-Raphson updates for the k -steps while face a regularized optimization problem at each iteration. Tractability relies on the fact that there are a variety of procedures for updating within the lasso problem that are available in closed form. Using these analytic updates greatly reduces the overall computational task and makes a k -step bootstrap procedure attractive within the lasso context.

Specifically, consider the following lasso problems on the bootstrap data. Let

$$\begin{aligned}\tilde{\gamma}_{y,lasso}^* &= \arg \min_{\gamma \in \mathbb{R}^p} \mathcal{L}_y^*(\gamma) + \kappa_n \|\widehat{\Psi}^y \gamma\|_1, \\ \tilde{\gamma}_{d,lasso}^* &= \arg \min_{\gamma \in \mathbb{R}^p} \mathcal{L}_d^*(\gamma) + \kappa_n \|\widehat{\Psi}^d \gamma\|_1,\end{aligned}\tag{4.1}$$

where

$$\begin{aligned}\mathcal{L}_y^*(\gamma) &= \frac{1}{nT} \sum_{t=1}^T \sum_{i=1}^n (\tilde{y}_{it}^* - \widehat{\delta}_{yt}^{*'} \widehat{f}_i^* - \widehat{U}_{it}^{*'} \gamma)^2, \\ \mathcal{L}_d^*(\gamma) &= \frac{1}{nT} \sum_{t=1}^T \sum_{i=1}^n (\tilde{d}_{it}^* - \widehat{\delta}_{dt}^{*'} \widehat{f}_i^* - \widehat{U}_{it}^{*'} \gamma)^2.\end{aligned}$$

The definitions of $\{\tilde{y}_{it}^*, \tilde{d}_{it}^*, \widehat{\delta}_{yt}^*, \widehat{\delta}_{dt}^*, \widehat{f}_i^*, \widehat{U}_{it}^*\}_{i \leq n, t \leq T}$ will be formally given below. Let $\tilde{\gamma}_y$ and $\tilde{\gamma}_d$ be the lasso solutions obtained from the original data. Also, note that we fix the value of κ_n and of the penalty loadings $\widehat{\Psi}^y$ and $\widehat{\Psi}^d$ to the same values as used to obtain the solutions $\tilde{\gamma}_y$ and $\tilde{\gamma}_d$ in the original data.

Within each bootstrap replication, we then approximately solve the lasso problems (4.1) by applying the following procedure. The maximum number of steps k to be taken should be determined on a case-by-case basis according to the available computational capacity.⁷

Algorithm (k -Step Lasso Iteration.)

Set k to be a pre-determined number of iterations.

(A1) Set $l = 0$ and initialize at $\gamma_{y,0} = \tilde{\gamma}_y$, $\gamma_{d,0} = \tilde{\gamma}_d$.

(A2) Determine one-step iteration mappings $\mathcal{S}_y, \mathcal{S}_d : \mathbb{R}^p \rightarrow \mathbb{R}^p$. Let

$$\gamma_{y,l+1} = \mathcal{S}_y(\gamma_{y,l}), \quad \gamma_{d,l+1} = \mathcal{S}_d(\gamma_{d,l}) \quad (4.2)$$

Set $l = l + 1$.

(A3) Repeat (A2) until $l = k$. Let the k -step lasso estimators be

$$\tilde{\gamma}_y^* = \gamma_{y,k}, \quad \tilde{\gamma}_d^* = \gamma_{d,k}.$$

There are a variety of iteration mappings that can be used in Step (A2) of the k -step lasso problem. A commonly used and simple mapping is the “coordinate descent method,” also known as the “shooting method,” studied by Fu (1998).⁸ For solving problem (4.1), write the solution after the l^{th} iteration as $\gamma_{y,l} = (\gamma_{y,l,1}, \dots, \gamma_{y,l,p})'$. The coordinate descent method updates $\gamma_{y,l+1}$ by iteratively cycling through all coordinates. Specifically, we solve the following one-dimensional optimization problem for $m = 1, \dots, p$,

$$\begin{aligned} \gamma_{y,l+1,m} = \arg \min_{g \in \mathbb{R}} \frac{1}{nT} \sum_{i,t} (\tilde{y}_{it}^* - \hat{\delta}_{yt}^{*'} \hat{f}_i^* - \hat{U}_{it,m^-}' \gamma_{y,l+1,m^-} - \hat{U}_{it,m^+}' \gamma_{y,l,m^+} - \hat{U}_{it,m}^* g)^2 \\ + \kappa_n |\hat{\Psi}_{m,g}^y|. \end{aligned} \quad (4.3)$$

Here $m^- = \{j : j < m\}$; and $\gamma_{y,l+1,m^-}$ and \hat{U}_{it,m^-}^* are \mathbb{R}^{m-1} dimensional vectors whose components are respectively those of $\{\gamma_{y,l+1,j} : j < m\}$ and $\{\hat{U}_{it,j}^* : j < m\}$. Similarly, $m^+ = \{j : j > m\}$; and γ_{y,l,m^+} and \hat{U}_{it,m^+}^* are \mathbb{R}^{p-m} dimensional vectors whose components are respectively those of $\{\gamma_{y,l,j} : j > m\}$ and $\{\hat{U}_{it,j}^* : j > m\}$. When $m = 1$, m^- is empty;

⁷In applications where obtaining the full lasso solution is not too burdensome, one may simply iterate to convergence.

⁸Another commonly used iterative scheme that could readily be applied in the present setting is the “composite gradient method” (e.g. Nesterov (2007) and Agarwal et al. (2012)). We choose to focus on the coordinate descent method as our concrete example as it does not rely on additional tuning parameters and performed well numerically in preliminary simulation experiments. In addition, coordinate descent requires weaker regularity conditions than the composite gradient method for our theoretical analysis.

and when $m = p$, m^+ is empty. In these cases, the corresponding subvectors, $\gamma_{y,l+1,m^-}$ and \widehat{U}_{it,m^-}^* or γ_{y,l,m^+} , and \widehat{U}_{it,m^+}^* , are defined as zero. Note that when $\gamma_{y,l+1,m}$ is being updated the previous $m-1$ elements have already been updated, while the remaining $p-m$ elements are yet to be updated. Thus, $\gamma_{y,l+1,m^-}$ is a subvector of $\gamma_{y,l+1}$, but γ_{y,l,m^+} is a subvector of $\gamma_{y,l}$. Denote by $\gamma_{y,l+1}^{(m)} := (\gamma_{y,l+1,m^-}, \gamma_{y,l+1,m}, \gamma_{y,l,m^+})'$ the vector that results immediately after the m^{th} coordinate has been updated during the $(l+1)^{\text{th}}$ iteration. When $m = p$, all the components have been updated; and we obtain $\gamma_{y,l+1} := \gamma_{y,l+1}^{(p)}$.

Importantly, (4.3) is a one-dimensional ℓ_1 -penalized quadratic problem which has an analytical solution given by the soft thresholding operation:

$$\begin{aligned} \gamma_{y,l+1,m} = & \left[\text{sgn} \left(\frac{1}{nT} \sum_{i=1}^T \sum_{t=1}^T Z_{it,l,m}^* \widehat{U}_{it,m}^* \right) \right] \\ & \times \left(\left| \frac{1}{nT} \sum_{i=1}^T \sum_{t=1}^T Z_{it,l,m}^* \widehat{U}_{it,m}^* - \frac{1}{2} \kappa_n \widehat{\Psi}_m^y \right|_+ \left(\frac{1}{nT} \sum_{i=1}^T \sum_{t=1}^T \widehat{U}_{it,m}^{*2} \right)^{-1} \right), \end{aligned} \quad (4.4)$$

where $Z_{it,l,m}^* := \widehat{y}_{it}^* - \widehat{\delta}_{yt}^{*'} \widehat{f}_i^* - \widehat{U}_{it,m^-}^{*'} - \gamma_{y,l+1,m^-} - \widehat{U}_{it,m^+}^{*'} - \gamma_{y,l,m^+}$, $(x)_+ = \max\{x, 0\}$, and $\text{sgn}(x)$ takes the sign of x . Therefore, the mappings in (4.2) are given by

$$\mathcal{S}_y(\gamma_{y,l}) = (\gamma_{y,l+1,1}, \dots, \gamma_{y,l+1,p})', \quad \text{where each } \gamma_{y,l+1,m} \text{ is given in (4.4).}$$

$\mathcal{S}_d(\gamma_{d,l})$ is obviously defined similarly.

With the k-step lasso program defined, we now state the complete algorithm for the proposed k-step bootstrap procedure. We make use of a wild residual bootstrap to generate the data at each bootstrap replication.

Algorithm (*k*-Step Wild Bootstrap.)

Let $\{\widehat{f}_i, \widehat{U}_{it}, \widehat{\Lambda}_t\}_{i \leq n, t \leq T}$ denote the estimates of the features of the factor model using the original data. Let $\widehat{\alpha}, \widehat{\delta}_{dt}, \widehat{\delta}_{yt}, \widehat{\gamma}_d, \widehat{\gamma}_y$ be the estimated coefficients from the original data, defined in (2.7) through (2.16). Also, let

$$\begin{aligned} \widehat{\xi}_t &= \widehat{\delta}_{yt} - \widehat{\alpha} \widehat{\delta}_{dt}, \quad t = 1, \dots, T, \text{ and} \\ \widehat{\theta} &= \widehat{\gamma}_y - \widehat{\alpha} \widehat{\gamma}_d. \end{aligned}$$

- (1) For each $i = 1, \dots, n$, let w_i^x ($x = U, Y, D$) be mutually independent random variables, where $\{w_i^x\}_{i \leq n}$ are i.i.d. with mean zero and variance one. Let

$$\widetilde{U}_{it}^* = w_i^U \widehat{U}_{it}, \quad \widetilde{\eta}_{it}^* = w_i^D \widehat{\eta}_{it}, \quad \widetilde{\epsilon}_{it}^* = w_i^Y \widehat{\epsilon}_{it}, \quad t = 1, \dots, T.$$

Define $\{\tilde{y}_{it}^*, \tilde{d}_{it}^*, \tilde{X}_{it}^*\}_{t \leq T}$ as

$$\begin{aligned}\tilde{y}_{it}^* &= \hat{\alpha} \tilde{d}_{it}^* + \hat{\xi}_t' \hat{f}_i + \tilde{U}_{it}^{*'} \hat{\theta} + \tilde{\varepsilon}_{it}^* \\ \tilde{d}_{it}^* &= \hat{\delta}_{dt}' \hat{f}_i + \tilde{U}_{it}^{*'} \hat{\gamma}_d + \tilde{\eta}_{it}^*, \\ \tilde{X}_{it}^* &= \hat{\Lambda}_t \hat{f}_i + \tilde{U}_{it}^*.\end{aligned}$$

(2) Apply the Factor-Lasso Algorithm to the bootstrap data $\{\tilde{y}_{it}^*, \tilde{d}_{it}^*, \tilde{X}_{it}^*\}_{i \leq n, t \leq T}$ to obtain an estimated alpha $\hat{\alpha}^*$ replacing the lasso estimation in Step (2) of the Factor-Lasso Algorithm with steps (A1)-(A3) from the k -Step Lasso Iteration defined above.

(3) Repeat the above steps (1)-(2) B times to obtain $\{\hat{\alpha}_b^*\}_{b \leq B}$.

Let q_τ^* be the τ^{th} upper quantile of $\{\sqrt{nT}|\hat{\alpha}_b^* - \hat{\alpha}|\}_{b \leq B}$, so that

$$P^*(\sqrt{nT}|\hat{\alpha}_b^* - \hat{\alpha}| \leq q_\tau^*) = 1 - \tau.$$

Construct the bootstrap confidence interval:

$$\left[\hat{\alpha} \pm \frac{q_\tau^*}{\sqrt{nT}} \right].$$

4.2. Validity of k -Step Bootstrap Confidence Interval. In the following, we present conditions under which we verify that the bootstrap confidence intervals are asymptotically valid:

$$P \left(\alpha \in \left[\hat{\alpha} \pm \frac{q_\tau^*}{\sqrt{nT}} \right] \right) \rightarrow 1 - \tau.$$

The first assumption imposes high-level conditions that will admit the use of general updating rules in (4.2) of the k -Step Lasso Iteration. The assumption provides high-level conditions on the computational properties and sparsity of the solution resulting after taking k iterations in the solution of the lasso problem. Recall that $\tilde{\gamma}_y^* = \gamma_{y,k}$ and $\tilde{\gamma}_d^* = \gamma_{d,k}$.

Assumption 4.1. *The following conditions hold for $x \in \{y, d\}$:*

(i) *Minimization Error: There is a deterministic sequence a_n such that $a_n \sqrt{nT} = o(1)$, and a $K_0 > 0$, such that when $k > K_0$,*

$$\mathcal{L}_x^*(\tilde{\gamma}_x^*) + \kappa_n \|\hat{\Psi}^x \tilde{\gamma}_x^*\|_1 \leq \mathcal{L}_x^*(\tilde{\gamma}_{x,lasso}^*) + \kappa_n \|\hat{\Psi}^x \tilde{\gamma}_{x,lasso}^*\|_1 + O_{P^*}(a_n).$$

(ii) *Sparsity*: $|\widehat{J}^*| = O_{P^*}(|J|_0)$, where $\widehat{J}^* = \{j \leq p : \tilde{\gamma}_{dj}^* \neq 0\} \cup \{j \leq p : \tilde{\gamma}_{yj}^* \neq 0\}$.

Condition (i) requires that the minimization error should be negligible compared to the statistical error after k iteration steps. Condition (ii) guarantees the sparsity of the iterated solutions. As a concrete example, we verify both conditions for the coordinate descent method. We note that, to the best of our knowledge, showing the $|J|_0$ -sparsity of the k -step iterated coordinate descent estimator has not been done previously when p is potentially much larger than n and may be of some independent interest.

Proposition 4.1. *The coordinate descent iteration as given in (4.4) satisfies Assumption 4.1.*

We next impose a fairly standard notion of regularity on the high-dimensional component \tilde{U}_{it} which shows up in the infeasible lasso problem with known factors.

Assumption 4.2 (Restricted Strong Convexity). *There is a constant $c > 0$, and a sequence $\tau_n = o(|J|_0^{-1})$ so that for all $\delta \in \mathbb{R}^p$,*

$$\delta' \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T \tilde{U}_{it} \tilde{U}'_{it} \delta \geq \frac{c}{2} \|\delta\|_2^2 - O_P(\tau_n) \|\delta\|_1^2.$$

This assumption has been discussed by many authors, and various sufficient conditions have been provided (e.g., Raskutti et al. (2010) and Loh and Wainwright (2015)). The following lemma provides a simple sufficient condition for both this assumption and the restricted/sparse eigenvalue assumption.

Lemma 4.1. *Suppose Assumption 3.1 holds. Let $\lambda_1 \leq \dots \leq \lambda_p$ be the eigenvalues of $\frac{1}{nT} \sum_i \sum_t E [(U_{it} - \bar{U}_{i,\cdot})(U_{it} - \bar{U}_{i,\cdot})']$. Suppose for some $0 < c < C$,*

$$c < \lambda_1 \leq \lambda_p < C.$$

Then Assumptions 3.4 and 4.2 are satisfied.

As we described earlier, even if $p/n \rightarrow \infty$, requiring that the eigenvalues of the population covariance matrix are well-bounded is not a stringent condition. Note that this condition is imposed only on the factor-residuals, U_{it} , and that similar conditions on the population covariance matrix of factor residuals are typically imposed in the formal analysis of large approximate factor models.

The following conditions are imposed on the bootstrap weights.

Assumption 4.3. For $x = U, Y, D$, $Ew_i^x = 0$ and $\text{Var}(w_i^x) = 1$. In addition, there exist $L, r > 0$, such that for any $s > 0$, $i \leq n$,

$$P(|w_i^x| > s) \leq \exp(-Ls^r).$$

The sub-exponential condition for the bootstrap weights enables us to bound many stochastic processes uniformly in $m \leq p$ and $t \leq T$. In our numerical studies, we follow Mammen (1993) and use $w_i^x = \zeta_{1,i}^x/\sqrt{2} + ((\zeta_{2,i}^x)^2 - 1)/2$ where $\zeta_{1,i}^x$ and $\zeta_{2,i}^x$ are independent standard normals and $x \in \{U, Y, D\}$.

Finally, we impose further regularity on the quality of estimation of the factors in the bootstrap data.

Assumption 4.4 (Quality of Factor Estimation in Bootstrap Data). *Suppose there is an invertible $\dim(f_i) \times \dim(f_i)$ matrix H^* with $\|H^*\| + \|H^{*-1}\| = O_{P^*}(1)$, and non-negative sequences Δ_F^* , Δ_{eg}^* , Δ_{ud}^* , Δ_{fe}^* , so that for $\tilde{z}_{it}^* \in \{\tilde{\eta}_{it}^*, \tilde{\epsilon}_{it}^*\}$, $\hat{g}_{tm} \in \{\hat{\Lambda}'_t \hat{\gamma}_d, \hat{\Lambda}'_t \hat{\gamma}_y, \hat{\delta}_{dt}, \hat{\delta}_{yt}, \hat{\lambda}_{tm}\}$, and $\hat{h}_{tm} \in \{\hat{\delta}_{dt}, \hat{\delta}_{yt}, \hat{\lambda}_{tm}\}$,*

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \|\hat{f}_i^* - H^{*'} \hat{f}_i\|_2^2 &= O_P(\Delta_F^{*2}) \\ \max_{m \leq p} \left\| \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T \tilde{z}_{it}^* \hat{g}_{tm} (\hat{f}_i^* - H^{*'} \hat{f}_i)' \right\|_F &= O_{P^*}(\Delta_{eg}^*) \\ \max_{m, k \leq p} \left\| \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T (\hat{f}_i^* - H^{*'} \hat{f}_i) \tilde{U}_{it,m}^* \hat{h}'_{tk} \right\|_F &= O_P(\Delta_{ud}^*). \end{aligned}$$

These sequences satisfy the following restrictions:

$$\begin{aligned} \sqrt{nT} |J|_0^2 \Delta_F^{*2} = o(1), \quad \Delta_{eg}^* = o\left(\frac{1}{\sqrt{nT}}\right), \quad \Delta_{ud}^* = o\left(\sqrt{\frac{\log p}{nT}}\right), \quad |J|_0^2 \sqrt{\log p} \Delta_{ud}^* = o(1), \\ \Delta_F^{*2} = o\left(\frac{\log p}{T \log(pT)}\right), \quad \text{and } \Delta_{\max}^2 |J|_0^2 T \Delta_F^{*2} = o(1). \end{aligned}$$

As with Assumption 3.5, we show that

$$\frac{1}{n} \sum_{i=1}^n \|\hat{f}_i^* - H' \tilde{f}_i\|_2^2 = O_{P^*}\left(\frac{1}{pT} + \frac{1}{n^2} + \frac{1}{nT^2}\right)$$

when \hat{f}_i^* is estimated using PCA in Appendix C which allows direct verification of conditions involving Δ_F^* . We handle the remaining terms by directly deriving the rates of convergences for each required term in Appendix C.

Under these additional conditions, we are able to verify that the confidence interval resulting from application the k -step bootstrap procedure has asymptotically correct coverage.

Theorem 4.1. *Suppose $n, p \rightarrow \infty$, and T is either fixed or growing. Under Assumptions 3.1-3.5 and 4.1-4.4,*

$$\sqrt{nT}\sigma_{\eta\epsilon}^{-1/2}\sigma_{\eta}^2(\hat{\alpha}^* - \hat{\alpha}) \rightarrow^{d^*} \mathcal{N}(0, 1).$$

In addition,

$$P(\sqrt{nT}|\hat{\alpha} - \alpha| \leq q_{\tau}^*) \rightarrow 1 - \tau.$$

5. ESTIMATING FACTORS USING PRINCIPAL COMPONENTS ANALYSIS

In this section, we discuss estimation of factors and factor residuals using principal components (PC).⁹ We also provide low-level conditions under which the high-level conditions used in establishing Theorem 3.1 are satisfied for PC.

5.1. Principal Components Estimator. Let

$$\tilde{X} = \begin{pmatrix} \tilde{X}_{i1} & \cdots & \tilde{X}_{n1} \\ \vdots & & \vdots \\ \tilde{X}_{iT} & \cdots & \tilde{X}_{nT} \end{pmatrix}_{pT \times n}, \quad \tilde{\Lambda} = \begin{pmatrix} \tilde{\Lambda}_1 \\ \vdots \\ \tilde{\Lambda}_T \end{pmatrix}_{pT \times K}, \quad \tilde{F} = \begin{pmatrix} \tilde{f}'_1 \\ \vdots \\ \tilde{f}'_n \end{pmatrix}_{n \times K},$$

and define \tilde{U} similarly. The matrix form of the factor model is then

$$\tilde{X} = \tilde{\Lambda}\tilde{F}' + \tilde{U},$$

where the individual and time effects have already been removed.

One of the most commonly used factor estimators is based on the PC of the $n \times n$ matrix $\tilde{X}'\tilde{X}$. Let \hat{F} denote the $n \times K$ matrix of the estimated factors. The columns of \hat{F}/\sqrt{n} are the eigenvectors of the first K eigenvalues of $\tilde{X}'\tilde{X}/(npT)$. Let V be the K by K diagonal matrix consisting of the first K eigenvalues. Then the PC estimator estimates \tilde{F} up to a $K \times K$ rotation matrix (e.g., Stock and Watson (2002) and Bai (2003)) H defined by

$$H = \frac{1}{npT}\tilde{\Lambda}'\tilde{\Lambda}\tilde{F}'\hat{F}V^{-1}.$$

⁹We choose to focus on the PC estimator as a concrete example because it is relatively simple and is free of tuning parameters. One could consider other options which would also satisfy our assumed high-level conditions. For example, the weighted PC estimator (e.g. Choi (2012); Bai and Liao (2013)), can be more efficient than the standard PC estimator but requires additional tuning parameters for practical application.

The factor estimator in the bootstrap sampling space is defined similarly with \widehat{F}^* denoting the $n \times K$ matrix of the estimated factors whose columns are \sqrt{n} times the first K eigenvectors of $\widetilde{X}^{*'}\widetilde{X}^*/(npT)$. Finally, it is important to note that we do not need to estimate the factors but only need to estimate the space spanned by the factors for Theorem 3.1 to hold.

5.2. Regularity Conditions. We now present additional regularity conditions which are sufficient to verify that the PC estimator satisfies the conditions given in Assumption 3.5. These conditions are standard for high-dimensional approximate factor models.

Assumption 5.1 (Pervasiveness). *There are $c, C > 0$ so that*

$$c < \frac{1}{T} \sum_{t=1}^T \frac{1}{p} \widetilde{\Lambda}'_t \widetilde{\Lambda}_t < C.$$

Assumption 5.1 effectively implies that the factors do not load on a small number of series but rather are related to a large number of the available X -variables. The use of this assumption in high-dimensional factor models provides part of the motivation for the factor-lasso approach where at least some forms of association between factors that are not pervasive but instead load on only a few elements in X and an outcome can be captured through the presence of the factor residuals in the equations of interest.

Assumption 5.2 (Second Order Weak Dependence). *There is $C > 0$,*

$$\begin{aligned} \max_{mti} \sum_{s=1}^T \sum_{v=1}^p \text{Cov}(U_{it,m}^2, U_{is,v}^2) &< C, \\ \max_{imstv} \sum_{h=1}^T \sum_{l=1}^p |\text{Cov}(U_{it,v}U_{is,m}, U_{ih,l}U_{is,m})| &< C, \\ \max_i \frac{1}{T^2 p} \sum_{m,l \leq p} \sum_{t,s,h,v \leq T} \text{Cov}(U_{it,m}U_{is,m}, U_{ih,l}U_{iv,l}) &< C, \\ \max_{im} \frac{1}{T^2 p} \sum_{k,l \leq p} \sum_{t,s,h,v \leq T} |\text{Cov}(U_{it,k}U_{is,m}, U_{ih,l}U_{iv,m})| &< C. \end{aligned}$$

The left hand side of the third condition equals $\max_i \text{Var}(\frac{1}{\sqrt{p}} \sum_{m=1}^p (\sqrt{T}\bar{U}_{i,m})^2)$. In addition, if we ignore the absolute value, then the left hand side of the fourth condition equals $\max_{im} \text{Var}(\frac{1}{\sqrt{p}} (\sqrt{T}\bar{U}_{i,k})(\sqrt{T}\bar{U}_{i,m}))$. Hence the third condition means that the variance of the standardized squared average should be bounded, and the fourth condition is slightly

stronger than requiring $\max_{im} \text{Var}(\frac{1}{\sqrt{p}}(\sqrt{T}\bar{U}_{i,k})(\sqrt{T}\bar{U}_{i,m})) < C$. In the special case when $\{U_t\}$ is serially independent across t , all of the four conditions in Assumption 5.2 can be directly verified under various notions of weak cross-sectional dependence.

The following proposition show that the high-level Assumptions 3.5 and 4.4 are satisfied by the PC estimator.

Proposition 5.1. *Further assume $|J|_0^4 = o(nT^3)$, $|J|_0^4 n = o(p^2T)$ and $|J|_0^2 \log n = o(p)$. Then Assumptions 3.5 and 4.4 about \hat{F} and \hat{F}^* are satisfied.*

The conditions $|J|_0^4 n = o(p^2T)$ and $|J|_0^2 \log n = o(p)$ require lower bounds on the growth of p . These conditions differ from those used in the literature on inference in purely sparse high-dimensional, e.g. Belloni et al. (2014b), in that lower bounds on p are not required in the purely sparse setting. These lower bounds arise since accurately estimating the unknown factors using PCA requires a large number of observed series. Indeed, the ‘‘average rate of convergence’’ is

$$\frac{1}{n} \sum_{i=1}^n \|\hat{f}_i - H' \tilde{f}_i\|_2^2 = O_P\left(\frac{1}{n^2} + \frac{1}{nT^2} + \frac{1}{pT}\right),$$

where the product pT is the dimension of \tilde{X}_i . In the special case $|J|_0 = O(1)$, these conditions require

$$T \ll n \ll p^2T, \log^3 p = O(n), \text{ and } \log n = o(p).$$

The results developed in this paper will thus be inappropriate in settings where p is quite small relative to n . Of course, in the setting with p small relative to n , a simple approach is to just use all of the available variables without dimension reduction.

Finally, though we have been assuming the number of factors, K , is known a priori, our procedure admits data-dependent methods (e.g., Bai and Ng (2002) or Ahn and Horenstein (2013)) for selecting K . Under mild conditions such as those employed in Bai and Ng (2002) or Ahn and Horenstein (2013), \hat{K} , the estimator for K , is consistent. All the preceding results hold can then be shown to hold following first-step estimation of K by conducting the theoretical analysis conditional upon the event that $\hat{K} = K$ and then arguing that the results asymptotically hold unconditionally as $P(\hat{K} = K) \rightarrow 1$.

6. NUMERICAL STUDIES AND EXAMPLES

We now present simulation and empirical results in support of the formal analysis presented in the previous sections. The first simulation example is based directly on the PPFM

given in (2.1)-(2.3). The second simulation example is based on a purely cross-sectional model that allows for instrumental variables estimation of the parameter on an endogenous variable in the presence of a low-dimensional set of instrumental variables (IVs) and a large number of potential control variables.¹⁰ Following the simulation experiments, we then present results from two empirical applications. In the first, we apply the developed procedure to estimate the effects of gun prevalence on crime following Cook and Ludwig (2006) using the data from Belloni et al. (2015). In the second example, we apply the instrumental variables strategy of Acemoglu et al. (2001) to try to estimate the effect of institutions on growth.

6.1. Simulation Examples.

6.1.1. *Panel Partial Factor Model Simulations.* In our first set of simulations, we report results for estimation and inference on α with data generated according to

$$\begin{aligned} y_{it} &= \alpha d_{it} + (c_\xi \xi_t)' f_i + U_{it}'(c_\theta \theta) + g_i + \nu_t + \epsilon_{it} \\ d_{it} &= (c_\delta \delta_{dt})' f_i + U_{it}'(c_\gamma \gamma_d) + \zeta_i + \mu_t + \eta_{it} \\ X_{it} &= (c_\Lambda \Lambda_t) f_i + w_i + \rho_t + U_{it} \end{aligned}$$

with $n = 100$, $T = 10$, $K = 3$, and $p = 100$. We take $\epsilon_{it} \sim N(0, 1)$, $\eta_{it} \sim N(0, 1)$, and $U_{it} \sim N(0_p, \Sigma_U)$ where 0_p is a $p \times 1$ vector of zeros, Σ_U has (r, s) element given by $[\Sigma_U]_{[r,s]} = .7^{|r-s|}$, and ϵ_{it} , η_{it} , and U_{it} are i.i.d. over i and t and jointly independent of each other. We generate unobserved individual-specific and time-specific heterogeneity by taking n i.i.d. draws, one for each individual, $(g_i, \zeta_i, w_i) \sim N(0_{p+2}, I_{p+2})$ where I_{p+2} is a $(p+2) \times (p+2)$ identity matrix and taking T i.i.d. draws, one for each time period, $(\nu_t, \mu_t, \rho_t) \sim N(0_{p+2}, I_{p+2})$. The latent factors, f_i , are generated as i.i.d. draws from $N(0_K, I_K)$. The factor loading vectors ξ_t and δ_{dt} and factor loading matrix Λ_t are drawn independently over time with each entry generated as an independent draw from a standard normal random variable. The individual-specific, time-specific heterogeneity terms and factor loadings are drawn once, and the same values are used in each simulation replication.

We set the j^{th} entry of θ and γ_d as $\theta_j = \gamma_{d,j} = \frac{1}{j^2}$. c_Λ , c_δ , c_γ , c_ξ , and c_θ are scalars that are set to alter the relative strength of f_i and U_{it} in each equation. We choose c_Λ so that the average R^2 from the p regressions of $X_{it,j}$ on f_i is 0.5. We choose (c_δ, c_γ) so that the R^2 of the infeasible regression of $d_{it} - \zeta_i - \mu_t$ on $(c_\delta \delta_{dt})' f_i + U_{it}'(c_\gamma \gamma_d)$ is 0.7 and the factors

¹⁰The formal development in the IV case with a small number of instruments is a notationally burdensome but straightforward extension of the results developed in this paper.

account for 0%, 25%, 50%, 75%, or 100% of the explanatory power in this regression. We similarly choose (c_ξ, c_θ) so that the R^2 of the infeasible regression of $y_{it} - \alpha d_{it} - g_i - \nu_t$ on $(c_\xi \xi_t)' f_i + U'_{it}(c_\theta \theta)$ is 0.7 and the factors account for 0%, 25%, 50%, 75%, or 100% of the explanatory power in this regression. Finally, we set $\alpha = 1$.

We compare the performance of the procedure developed in this paper to several benchmarks. Because we consider a design with $p < nT$, ordinary least squares of y_{it} on d_{it} , X_{it} and a full set of individual and time dummy variables is feasible (OLS). We also consider estimating α based on the assumption that confounding is entirely captured by latent factors. To implement this procedure, we extract factors, \hat{f}_i , from \tilde{X}_{it} by PCA as discussed in Section 5. We then regress y_{it} on d_{it} , \hat{f}_i interacted with a complete set of time dummy variables, and a full set of individual and time dummy variables to obtain the estimator for α (Factor). For our third procedure, we directly apply the fixed effects double-selection procedure of Belloni et al. (2015) which is appropriate for a sparse high-dimensional model with fixed effects (Double Selection). We then consider two *ad hoc* variants of the double-selection approach. In the first, we extract the first 20 principal components and interact these with a full set of time dummies. We then apply the fixed effects double-selection procedure of Belloni et al. (2015) to the data (Y, D, X^*) where X^* denotes the original X variables augmented to include the interactions of principal components with time dummies (Double Selection F). The second *ad hoc* procedure extracts factors from \tilde{X}_{it} by PCA. We then obtain estimates \hat{U}_{it} as in (2.7) and apply the fixed effects double-selection procedure of Belloni et al. (2015) to the data (Y, D, \hat{U}^*) where \hat{U}^* denotes the matrix formed by combining \hat{U} with the interactions of principal components with time dummies (Double Selection U). Finally, we directly apply the factor-lasso approach outlined in this paper (Factor Lasso). We use the Ahn and Horenstein (2013) procedure to select the number of factors to use in obtaining the Factor, Double Selection U, and Factor Lasso results.

Figure 1 gives simulation RMSEs for the estimator of α resulting from applying each procedure. The RMSEs are truncated at 0.1 for readability of the figure. The most striking feature of Figure 1 is that only the proposed factor lasso procedure delivers uniformly good performance regardless of the relative strength of the factors and factor residuals in this simulation design. Each of the other procedures exhibits behavior that depends strongly on the exact strength of the factors in the different equations. In terms of RMSE, the factor-lasso procedure uniformly dominates regular OLS, Double Selection ignoring the factor structure, and the *ad hoc* procedure Double Selection F within the design considered. The factor-lasso estimator of α is outperformed by the pure factor model in the case where all of the explanatory power in the outcome equation is contained in the factors,

which corresponds to the case where the pure factor model is correctly specified and there is no additional confounding based on the factor residuals, and the Double Selection U procedure when the factors have no explanatory power in the treatment (D) equation but all explanatory power in the Y equation. It is also important to note that the performance loss is small in these few cases where the factor lasso is outperformed. A final interesting point to note is that the conventional lasso-based double selection procedure is outperformed by the factor lasso even when the factors do not load in either the treatment or outcome equation. It seems likely that the loss in this case is due to the presence of the factors in the observed explanatory variables which leads to strong correlation among these variables. This strong correlation among the X 's is well-known to pose challenges for lasso-type estimators.

We report size of 5% level tests based on standard asymptotic approximations for each of the six procedures considered in Figure 2 where the sizes are truncated at 0.3 for readability of the figure. In each panel, we report the rejection frequency of the standard t-test of the null hypothesis that $\alpha = 1$ with standard errors clustered at the individual level. The most striking feature of the figure is again the uniformly good performance of tests based on the proposed factor lasso procedure. Tests based on the factor-lasso procedure effectively control size, with size ranging between 3.3% and 5.3% across the design parameters considered in the simulation. This behavior is in sharp contrast to the other procedures considered which may have large size distortions depending upon exactly how large the relative contribution of the factors is in the D and Y equations. Importantly, this good behavior does not come at the cost of using an inferior estimator as evidenced by the RMSE results.

We conclude this discussion by looking at the performance of the k-step bootstrap. In Figure 3, we report size of 5% level tests using the factor-lasso estimator and the asymptotic approximation provided in Theorem 3.1, the k-step bootstrap, and a score bootstrap based on Belloni et al. (2014a). The k-step bootstrap and asymptotic approximation have similar performance that keeps size close to the promised level. Interestingly, the score-based bootstrap that does not reestimate the factors or the lasso parts of the model exhibits mild size distortions across all of the design settings in this example.

6.1.2. Instrumental Variables Model Simulations. We supplement the simulation results from the PPFM with additional simulations in a cross-sectional version of the model generalized to allow for an endogenous variable. Specifically, we generate data from the model

$$y_i = \alpha d_i + (c_\xi \xi)' f_i + U_i'(c_\theta \theta) + \nu + \epsilon_i$$

$$\begin{aligned} d_i &= \pi z_i + (c_{\delta_d} \delta_d)' f_i + U_i'(c_{\gamma_d} \gamma_d) + \mu + \eta_i \\ z_i &= (c_{\delta_z} \delta_z)' f_i + U_i'(c_{\gamma_z} \gamma_z) + \zeta + v_i \\ X_i &= (c_{\Lambda} \Lambda) f_i + \rho + U_i \end{aligned}$$

with $n = 100$, $K = 2$, and $p = 100$. Within this model, d_i is an endogenous variable with coefficient of interest α and z_i is an instrumental variable. We generate $\epsilon_i \sim N(0, 1)$ and $\eta_i \sim N(0, 1)$ with $E[\epsilon_i \eta_i] = .8$ i.i.d. across i and independent of all other random variables. We generate i.i.d. draws for U_i as before, and $v_i \sim N(0, 1)$ independently from U_i . We also generate $(\nu, \mu, \rho, \xi, \delta_d, \Lambda, c_{\Lambda}, c_{\gamma_z}, c_{\delta_d}, c_{\xi}, c_{\theta})$ as before. We set $\theta = \gamma_d = \gamma_z$ to be vectors with j^{th} entry given by $\theta_j = \gamma_{d,j} = \gamma_{z,j} = \frac{1}{j^2}$. To control the strength of the instrument, we choose $(c_{\delta_z}, c_{\gamma_z})$ so that the R^2 of the infeasible regression of $z_i - \zeta$ on $(c_{\delta_z} \delta_z)' f_i + U_i'(c_{\gamma_z} \gamma_z)$ is 0.7 and the factors account for 50% of the explanatory power in this regression. We set π so that the fraction of variation accounted for by z_i in the regression of d_i on z_i, f_i and U_i is 25%. Finally, we set $\alpha = 1$.

We again estimate α using six different IV procedures similar to those implemented in the previous simulation with one exception. As the number of features is equal to the sample size in these simulations, we consider an infeasible ‘‘oracle’’ estimator that estimates α from IV regression of $y_i - (c_{\xi} \xi)' f_i - U_i'(c_{\theta} \theta) - \nu$ on $d_i - (c_{\delta_d} \delta_d)' f_i - U_i'(c_{\gamma_d} \gamma_d) - \mu$ using $z_i - (c_{\delta_z} \delta_z)' f_i - U_i'(c_{\gamma_z} \gamma_z) - \zeta$ as instrument (Oracle). This estimator provides a type of best-case benchmark and allows us to ascertain that instruments are strong enough that the usual asymptotic approximation provides a reasonable approximation in the idealized scenario where one is able to perfectly remove the effect of confounding from all variables.

Figure 4 gives simulation RMSEs for the estimator of α resulting from applying each procedure. The RMSEs are truncated at 0.1 for readability of the figure.¹¹ Again, we see that the factor lasso procedure delivers good performance regardless of the relative strength of the factors and factor residuals in this simulation design. Each of the other procedures exhibits behavior that depends strongly on the exact strength of the factors in the different equations. It might be noted that the dominance of the factor-lasso estimator, in terms of RMSE, over the ‘‘Oracle’’ procedure is due to the definition of the oracle that we use which fully removes the variation in each variable due to factors and factor residuals even in situations in which some of these variables produce no confounding. For example, one need not remove the variation in the instruments due to the factors in cases where the

¹¹Theoretically, the MSE of the IV estimator does not exist in this context. We report root mean truncated squared error with a truncation point of 1.

factors have zero loadings in the outcome equation, but this variation is always removed due to the way we have defined the oracle model.

We report size of 5% level tests based on standard asymptotic approximations for each of the six procedures considered in Figure 5 with size truncated at 0.3 for readability of the figure. In each panel, we report the rejection frequency of the standard t-test of the null hypothesis that $\alpha = 1$ using heteroscedasticity robust standard errors. Here, we see that the only procedure that uniformly controls size is the infeasible oracle. Among the feasible procedures, the proposed factor lasso approach performs relatively well in keeping size distortions small across the majority of combinations of relative strengths of the factors. In this case, we do see that the factor-lasso procedure suffers from reasonably large size distortions when the factors account for all of the confounding in the outcome equation and a moderate amount of confounding in the treatment equation. We also see that the pure factor model controls size well in this case, but performs very poorly once all variation in the outcome equation is not due to the factors.

We again conclude by looking at the performance of the k-step bootstrap in Figure 6. We see that there is a modest, but clearly visible, improvement from using the k-step bootstrap relative to the asymptotic approximation. The score based bootstrap, on the other hand, lines up reasonably well with the asymptotic approximation.

6.1.3. Summary of Simulation Results. Overall, the results from the two simulation experiments are supportive of the asymptotic theory. We see that the factor-lasso approach delivers estimators with good properties relative to other feasible procedures that leverage either a pure factor structure or a pure sparse structure in partial factor model settings. We see that both point estimation properties, measured in terms of RMSE, and inferential quality, as measured by size of tests, are competitive or much better than the other procedures considered in our simulation design. The results also suggest that the proposed k-step bootstrap procedure works relatively well and may offer some gains relative to the asymptotic Gaussian approximation.

6.2. Empirical Examples.

6.2.1. Estimating the Effects of Gun Prevalence on Crime. In this example, we follow Belloni et al. (2015) who build upon the work of Cook and Ludwig (2006) and attempt to estimate the effect of gun prevalence on crime in a setting with a high-dimensional set of

potential controls. As in Belloni et al. (2015), we focus exclusively on trying to measure the effect of gun prevalence on homicide rates. An important difficulty with estimating the effect of gun prevalence in the United States is that exact gun-ownership numbers are difficult to obtain. Due to this difficulty, Cook and Ludwig (2006) use the fraction of suicides committed with a firearm (abbreviated FSS) within a county to proxy for county-level gun ownership rates. Cook and Ludwig (2006) provide a series of arguments and evidence from secondary data sources supporting the claim that FSS provides a useful proxy for gun ownership. For the analysis in this paper, we simply take it as given that estimating a causal effect of FSS on crime measures is worthwhile and abstract from any further measurement or data issues surrounding the use of this proxy.

Both Cook and Ludwig (2006) and Belloni et al. (2015) estimate linear fixed effects models of the form

$$\log Y_{it} = \alpha \log \text{FSS}_{it-1} + X'_{it}\beta + g_i + \nu_t + \epsilon_{it} \quad (6.1)$$

where g_i and ν_t are treated as parameters to be estimated, X_{it} are control variables, and Y_{it} is one of three dependent variables: the overall homicide rate within county i in year t , the firearm homicide rate within county i in year t , or the non-firearm homicide rate within county i in year t . Cook and Ludwig (2006) use the four variables percent African American, percent of households with female head, nonviolent crime rates, and percent of the population that lived in the same house five years earlier as their set of controls X_{it} . Belloni et al. (2015) maintain the assumption of approximate sparsity and employ their variable selection approach using a much larger set of potential controls generated by taking variables compiled by the US Census Bureau as X_{it} . Their variables include county-level measures of demographics, the age distribution, the income distribution, crime rates, federal spending, home ownership rates, house prices, educational attainment, voting patterns, employment statistics, and migration rates along with interactions of the initial (1980) values of all control variables with a linear, quadratic, and cubic term in time.

Rather than adopt the approximately sparse model in (6.1), we employ the PPFM, (2.1)-(2.3), and factor-lasso approach to estimate α using 909 variables in X_{it} constructed as in Belloni et al. (2015).¹² The PPFM model seems very appropriate for this data as it directly incorporates a mechanism to accommodate the concern that there are features of counties that are not directly observed, the f_i , but are related to the evolution of the outcome and treatment variable of interest, which is captured by the time-varying factor

¹²The exact identities of the variables are available upon request. The data is from the U.S. Census Bureau USA Counties Database, <http://www.census.gov/support/USACdataDownloads.html>.

loadings. Obviously, exclusion of these factors would then lead to omitted variables bias in any estimator of α that fails to capture them. Concern about the existence of such factors is common in empirical applications involving aggregate panel data.

The key assumption that we leverage to allow us to simply accommodate these latent factors is that the same correlated unobserved factors that lead to confounding are related to the evolution of other observed county-level aggregates and that we have access to a large number of these auxiliary aggregates. While this key assumption is strong, the PPFM also naturally provides some robustness to the presence of shocks (U_{it}) that are related to movements of the observed X_{it} series as well as movements in the variable of interest and outcome. Such shocks may be motivated, for example, by the factor structure being misspecified, by the presence of variables that are not strongly related to factors but are confounded with the treatment and outcome, and simply by the presence of local shocks not captured by the factors that are related to the observed series.

We present estimation results in Table 1 with results for each dependent variable presented across the columns and rows corresponding to different estimation approaches. As a baseline, we report numbers taken directly from the first row of Table 3 in Cook and Ludwig (2006) in the first row of Table 1 (“Cook and Ludwig (2006) Baseline”). We report results obtained from our data the remaining rows.¹³ For these results, we first report the point estimate and estimate of the asymptotic standard error obtained by clustering by county. Immediately below these results, we report the 95% confidence interval obtained from applying the k-step bootstrap procedure in brackets. The rows labeled “Post Double Selection” apply the procedure of Belloni et al. (2015). The rows labeled “Factor” are based on a pure factor model; the rows labeled “Factor-Lasso” use the proposed factor-lasso procedure. All factors are estimated using PCA and the number of factors is selected using Ahn and Horenstein (2013).

We see that the estimates and inferential statements produced for the firearm homicide rate (“Gun”) and the non-firearm homicide rate (“non-Gun”) are broadly consistent with each other. In all cases, there is a fairly large positive point estimate for the effect on the firearm homicide rate with corresponding 95% confidence intervals that exclude zero, suggesting positive association between the used measure of gun prevalence and gun homicides. For the non-firearm homicide rate, all point estimates are negative and modest and confidence intervals include both positive and negative values. The broad results for the

¹³All results are based on weighted regression where we weight by the within-county average population over 1980-1999.

overall homicide rate (“Overall”) are slightly more mixed. The baseline results for Cook and Ludwig (2006) and results from a pure factor model suggest a strongly significant, positive effect of gun prevalence on the overall homicide rate. Assuming sparsity and applying Belloni et al. (2015) yields a positive estimate of the effect which is statistically insignificant at the 5% level. Finally, the factor-lasso estimator is similar in magnitude to the sparsity-based estimator but borderline significant at the 5% level using the bootstrap confidence interval.

A more interesting comparison can be made by looking more closely and considering the variable and factor selection results. The “Post Double Selection” procedure ends up selecting three variables for estimating the effect on overall homicide rates, three variables for gun homicide rates, and two variables for non-gun homicide rates. The pure factor model uses one factor. The factor-lasso approach then uses one factor in all cases but selects eight additional variables for estimating the effect on the overall homicide rate, eight additional variables for the gun homicide rate, and five additional variables for the non-gun homicide rate. These results suggest that the “Post Double Selection” and “Factor” results may be based on models that fail to adequately capture the effect of potential confounds. We also see that the “Factor” estimates are substantially shifted away from the “Factor Lasso” estimates relative to standard errors and that the factor-lasso estimates are the most precise in the sense of having the shortest confidence intervals. Both findings are consistent with the asymptotic theory and with the simulation results.

6.2.2. *Estimating the Effects of Institutions on Output.* We revisit the example considered in Acemoglu et al. (2001). Acemoglu et al. (2001) are interested in the parameter α in a structural model of the form

$$\log(\text{GDP per capita}_i) = \alpha(\text{Protection from Expropriation}_i) + x_i'\beta + \varepsilon_i$$

based on aggregate country level data where “Protection from Expropriation” is a measure of the strength of individual property rights that is used as a proxy for the strength of institutions and x_i is a set of variables that are meant to control for geography. Acemoglu et al. (2001) adopt an IV strategy where they instrument for institution quality using early European settler mortality to estimate α as institutions are clearly potentially endogenous. They point out that their instrument would be invalid if there were other factors that are highly persistent and related to the development of institutions within a country and to the country’s GDP. A leading candidate for such a factor that they discuss is geography. To address this possibility, Acemoglu et al. (2001) control for the distance from the equator

in their baseline specifications and consider different sets of geographic controls such as continent dummies within their robustness checks.¹⁴

There are, of course, many other ways to measure geography besides distance to the equator or continent where a country is found. Rather than *ex ante* choose a small number of variables to proxy for geography, we put a large number of variables that potentially capture geography in x_i and then use the data to reduce dimension. Specifically, we consider dummies for Africa, Asia, North America, and South America as well as longitude, renewable water, land boundary, land area, amount of coastline, territorial seas, amount of arable land, average temperature, average high temperature, average low temperature, average precipitation, elevation of highest point, elevation of lowest point, fraction of area that is low-lying, latitude, and spherical distance from London.

We adapt the analysis of Acemoglu et al. (2001) to the present setting by considering estimation of a partial factor instrumental variables model

$$\begin{aligned}\log(\text{GDP per capita}_i) &= \alpha(\text{Protection from Expropriation}_i) + f_i'\xi + U_i'\theta + \varepsilon_i \\ \text{Protection from Expropriation}_i &= \pi\text{Early Settler Mortality}_i + f_i'\delta_d + U_i'\gamma_d + \eta_i \\ \text{Early Settler Mortality}_i &= f_i'\delta_z + U_i'\gamma_z + v_i \\ x_i &= \Lambda f_i + U_i\end{aligned}$$

using our 20 geography measures as x_i and the 64 countries from the original Acemoglu et al. (2001) data. The factor-lasso approach seems quite sensible in this setting. Each of the observed geography measures could reasonably be taken as a noisy proxy for a country’s geography. This relationship is likely to be complicated and uneven with the chief features leading to association between the geography proxies plausibly being only weakly related to the notions of geography that are important predictors of mortality and institutions. The factor-lasso approach, by allowing a small number of elements of U_i to enter the equation of interest in addition to any common geography factors, readily accommodates this latter possibility in a parsimonious, data-dependent way.

We report estimation results for the first stage coefficient on the instrument in Table 2. We report results from the factor-lasso approach in the row “Factor-Lasso”. For comparison, we also report results from a few natural alternative models. The row labeled “Latitude” uses the single variable distance from the equator to control for geography as in the baseline results from Acemoglu et al. (2001). We report results from applying OLS

¹⁴E.g. Acemoglu et al. (2001) Table 4.

using all 20 available geographic controls without dimension reduction in the row labeled “All Controls.” We apply the double selection approach of Belloni et al. (2014b) which would be appropriate if the relationship between geographic controls and the variables of interest were well-approximated by a sparse linear model in “Double Selection.” Finally, “Factor” reduces dimension through positing a conventional factor model. All factors are estimated using PCA with number of factors selected by applying the procedure from Ahn and Horenstein (2013).

The first-stage results using only the latitude control suggest there is a fairly strong relationship between the instrument and endogenous variable if latitude is a sufficient control for geography. The first stage F-statistic using just latitude is 10.9 which many would take to indicate that the instrument is sufficiently strong to identify the effect of interest.¹⁵ The results change in a potentially substantive way after allowing for the possibility that geography is not adequately captured by latitude. For each of the remaining approaches considered, the first-stage F-statistic drops substantially below 10, with all methods besides applying the pure factor model returning first-stage coefficients that are statistically insignificant at the 5% level.

One might dismiss the lack of significance after including all controls without dimension reduction as it seems likely that a model with 20 covariates in addition to the variables of interest and only 64 observations is overfit. The next strongest result is from the pure factor model which makes use of a single extracted component and produces a first-stage F-statistic of 7.5. As evidenced in the simulation example, inference results based on a pure factor model may be highly misleading when elements of U_i also have explanatory power. It is then interesting that the double-selection approach and the factor-lasso approach deliver almost identical results indicating a weak association between the endogenous variable and instrument after controlling parsimoniously for geography. The double-selection procedure selects four variables¹⁶, and the factor-lasso approach uses one factor and two additional variables.¹⁷ One might take this to mean that the four variables selected in the double-selection procedure approximately capture the same information as the single factor and two variables used in the factor-lasso results. In either case, the results suggest that, at best, identification of the structural effect of institutions as measured by “Protection from

¹⁵A benchmark that is commonly used in the applied literature to assess whether there is sufficient variation in the instrument to identify the effect of interest is to compare the first stage F-statistic to 10, with smaller values indicating weak identification.

¹⁶These variables are the Africa dummy, average temperature, average high temperature, and amount of arable land.

¹⁷The two selected variables in addition to the factor are the Africa and Asia dummies.

Expropriation” using settler mortality as instrument is weak after geography is controlled for in a parsimonious, data-dependent way. Given this apparent weak identification, we do not report second stage estimates of the structural effect.¹⁸

6.2.3. Summary of Empirical Examples. We believe the two empirical examples illustrate the potential applicability of partial factor models and the associated factor-lasso approach in applied economics. The model provides a natural generalization to standard factor models and sparse high-dimensional models and seems appropriate for many economic applications, especially those that make use of aggregate panel or cross-sectional data. The results in the first example based on Cook and Ludwig (2006) roughly line up with the original results, though they demonstrate the potential for efficiency gains from adopting the methods developed in this paper. In the second example, we draw substantively different conclusions about the strength of identification than one would draw following the approach in Acemoglu et al. (2001) due to the ability to control more flexibly for the leading candidate for confounding. Overall, the results suggest that application of the proposed methods may usefully complement the sensitivity analyses performed in empirical economics and also have the potential to strengthen the plausibility of any conclusions drawn.

APPENDIX A. PROOF OF THEOREM 3.1 AND COROLLARY 3.1

Define $(KT) \times 1$ matrices $\tilde{\Xi} = (\tilde{\xi}_1', \dots, \tilde{\xi}_T')'$ and $\tilde{\Delta}_d = (\tilde{\delta}_{d1}', \dots, \tilde{\delta}_{dT}')'$. Note that

$$\begin{aligned}\tilde{Y} &= \tilde{D}\alpha + (I_T \otimes \tilde{F})\tilde{\Xi} + \tilde{U}\theta + \tilde{\epsilon} \\ \tilde{D} &= (I_T \otimes \tilde{F})\tilde{\Delta}_d + \tilde{U}\gamma_d + \tilde{\eta}.\end{aligned}$$

Note that $\hat{\eta} = M_{\hat{U}_{\hat{J}}}(I_T \otimes M_{\hat{F}})\tilde{D}$. Hence,

$$\begin{aligned}\hat{\alpha} &= (\hat{\eta}'\hat{\eta})^{-1}\hat{\eta}'M_{\hat{U}_{\hat{J}}}(I_T \otimes M_{\hat{F}})\tilde{Y} \\ &= \alpha + (\hat{\eta}'\hat{\eta})^{-1}\hat{\eta}'M_{\hat{U}_{\hat{J}}}(I_T \otimes M_{\hat{F}})[(I_T \otimes \tilde{F})\tilde{\Xi} + \tilde{U}\theta + \tilde{\epsilon}] \\ &= \alpha + (\hat{\eta}'\hat{\eta})^{-1}(\hat{\eta} - \tilde{\eta})'M_{\hat{U}_{\hat{J}}}(I_T \otimes M_{\hat{F}})\tilde{\epsilon} + (\hat{\eta}'\hat{\eta})^{-1}\hat{\eta}'M_{\hat{U}_{\hat{J}}}(I_T \otimes M_{\hat{F}})\tilde{\epsilon} \\ &\quad + (\hat{\eta}'\hat{\eta})^{-1}\hat{\eta}'M_{\hat{U}_{\hat{J}}}(I_T \otimes M_{\hat{F}}\tilde{F})\tilde{\Xi} + (\hat{\eta}'\hat{\eta})^{-1}\hat{\eta}'M_{\hat{U}_{\hat{J}}}(I_T \otimes M_{\hat{F}})\tilde{U}\theta.\end{aligned}$$

Note that $\tilde{\eta}'M_{\hat{U}_{\hat{J}}}(I_T \otimes M_{\hat{F}})\tilde{\epsilon} = \tilde{\eta}'\tilde{\epsilon} - \tilde{\eta}'(I_T \otimes P_{\hat{F}})\tilde{\epsilon} - \tilde{\eta}'P_{\hat{U}_{\hat{J}}}\tilde{\epsilon} + \tilde{\eta}'P_{\hat{U}_{\hat{J}}}(I_T \otimes P_{\hat{F}})\tilde{\epsilon}$. Hence,

$$\sqrt{nT} \left(\frac{1}{nT} \hat{\eta}'\hat{\eta} \right) (\hat{\alpha} - \alpha) = \frac{1}{\sqrt{nT}} \tilde{\eta}'\tilde{\epsilon} + \sum_{i=1}^6 A_i \quad (\text{A.1})$$

¹⁸We note that it would be straightforward to adapt the weak-identification robust procedure of Chernozhukov and Hansen (2008) to the present setting. We do not pursue this extension for brevity.

where

$$\begin{aligned}
 A_1 &= \frac{1}{\sqrt{nT}}(\hat{\eta} - \tilde{\eta})' M_{\hat{U}_j} (I_T \otimes M_{\hat{F}}) \tilde{\epsilon}, & A_2 &= \frac{1}{\sqrt{nT}} \tilde{\eta}' M_{\hat{U}_j} (I_T \otimes M_{\hat{F}} \tilde{F}) \tilde{\Xi} \\
 A_3 &= -\frac{1}{\sqrt{nT}} \tilde{\eta}' (I_T \otimes P_{\hat{F}}) \tilde{\epsilon}, & A_4 &= \frac{1}{\sqrt{nT}} \tilde{\eta}' M_{\hat{U}_j} (I_T \otimes M_{\hat{F}}) \tilde{U} \theta \\
 A_5 &= -\frac{1}{\sqrt{nT}} \tilde{\eta}' P_{\hat{U}_j} \tilde{\epsilon}, & A_6 &= \frac{1}{\sqrt{nT}} \tilde{\eta}' P_{\hat{U}_j} (I_T \otimes P_{\hat{F}}) \tilde{\epsilon} = 0.
 \end{aligned}$$

We shall prove that $A_i = o_P(1)$ for $i = 1, \dots, 6$ and $\frac{1}{nT} \tilde{\eta}' \hat{\eta} - \frac{1}{nT} \tilde{\eta}' \tilde{\eta} = o_P(1)$. So

$$\begin{aligned}
 \hat{\eta} &= M_{\hat{U}_j} (I_T \otimes M_{\hat{F}}) \tilde{D} = M_{\hat{U}_j} (I_T \otimes M_{\hat{F}}) ((I_T \otimes \tilde{F}) \tilde{\Delta}_d + \tilde{U} \gamma_d + \tilde{\eta}) \\
 &= M_{\hat{U}_j} (I_T \otimes M_{\hat{F}} \tilde{F}) \tilde{\Delta}_d + M_{\hat{U}_j} (I_T \otimes M_{\hat{F}}) \tilde{U} \gamma_d + M_{\hat{U}_j} (I_T \otimes M_{\hat{F}}) \tilde{\eta}.
 \end{aligned}$$

Using the fact that $M_{\hat{F}} \hat{F} = 0$, it can be proven that

$$\begin{aligned}
 \frac{1}{\sqrt{nT}}(\hat{\eta} - \tilde{\eta}) &= \frac{1}{\sqrt{nT}} M_{\hat{U}_j} (I_T \otimes M_{\hat{F}} (\tilde{F}H - \hat{F})H^{-1}) \tilde{\Delta}_d \\
 &\quad + \frac{1}{\sqrt{nT}} M_{\hat{U}_j} (I_T \otimes M_{\hat{F}}) \tilde{U} \gamma_d \\
 &\quad - \frac{1}{\sqrt{nT}} P_{\hat{U}_j} \tilde{\eta} - \frac{1}{\sqrt{nT}} M_{\hat{U}_j} (I_T \otimes P_{\hat{F}}) \tilde{\eta}.
 \end{aligned} \tag{A.2}$$

In the subsequent subsections, we provide bounds for A_i for $i = 1, \dots, 6$ and for $\frac{1}{\sqrt{nT}} \|\hat{\eta} - \tilde{\eta}\|_2$.

A.1. Bounding $\hat{\eta} - \tilde{\eta}$. Write

$$\psi_n := \kappa_n |J|_0^{1/2} + \|R_y\|_1 + \Delta_F |J|_0 + \sqrt{\frac{|J|_0}{n}}. \tag{A.3}$$

Proposition A.1. $\frac{1}{\sqrt{nT}} \|\hat{\eta} - \tilde{\eta}\|_2 = O_P(\psi_n)$.

Proof. Note that $\|\tilde{\Delta}_d\|_2 = O(\sqrt{T})$. Hence by Lemma G.1,

$$\begin{aligned}
 \left\| \frac{1}{\sqrt{nT}} M_{\hat{U}_j} (I_T \otimes M_{\hat{F}} (\tilde{F}H - \hat{F})H^{-1}) \tilde{\Delta}_d \right\|_2 &\leq O_P(1) \frac{1}{\sqrt{nT}} \|\tilde{F}H - \hat{F}\|_F \|\tilde{\Delta}_d\|_2 = O_P(\Delta_F) \\
 \left\| \frac{1}{\sqrt{nT}} M_{\hat{U}_j} \tilde{U} \gamma_d \right\|_2 &= O_P \left(\kappa_n |J|_0^{1/2} + \|R_y\|_1 + \Delta_F |J|_0 + \sqrt{\frac{|J|_0}{n}} \right), \\
 \left\| \frac{1}{\sqrt{nT}} M_{\hat{U}_j} (I_T \otimes P_{\hat{F}}) \tilde{U} \gamma_d \right\|_2 &\leq \left\| \frac{1}{\sqrt{nT}} (I_T \otimes P_{\hat{F}}) \tilde{U} \gamma_d \right\|_2 = O_P \left(\sqrt{\frac{|J|_0}{n}} + \Delta_F |J|_0 \right), \\
 \left\| \frac{1}{\sqrt{nT}} M_{\hat{U}_j} (I_T \otimes P_{\hat{F}}) \tilde{\eta} \right\|_2 &\leq \left\| \frac{1}{\sqrt{nT}} (I_T \otimes P_{\hat{F}}) \tilde{\eta} \right\|_2 = O_P \left(\frac{1}{\sqrt{n}} + \Delta_F \right), \\
 \left\| \frac{1}{\sqrt{nT}} P_{\hat{U}_j} \tilde{\eta} \right\|_2 &= O_P \left(\sqrt{|J|_0 \frac{\log p}{nT}} \right).
 \end{aligned}$$

Hence, equation (A.2) implies $\frac{1}{\sqrt{nT}}\|\hat{\eta} - \tilde{\eta}\|_2 = O_P(\psi_n)$. ■

A.2. Showing $A_1, A_3, A_5, A_6 = o_P(1)$. By equation (A.2), Lemma F.2, $P_{\hat{U}_j}(I_T \otimes P_{\hat{F}}) = 0$, and noting that $P_{\hat{U}_j}M_{\hat{U}_j} = 0$, we can show that

$$\begin{aligned} A_1 &= \frac{1}{\sqrt{nT}}(\hat{\eta} - \tilde{\eta})'M_{\hat{U}_j}(I_T \otimes M_{\hat{F}})\tilde{\epsilon} \\ &= \tilde{\epsilon}'\frac{1}{\sqrt{nT}}M_{\hat{U}_j}(I_T \otimes M_{\hat{F}})\tilde{U}\gamma_d + \tilde{\epsilon}'\frac{1}{\sqrt{nT}}M_{\hat{U}_j}(I_T \otimes M_{\hat{F}}(\tilde{F}H - \hat{F})H^{-1})\tilde{\Delta}_d. \end{aligned}$$

It then follows from Lemma G.3 (i)(v) that $A_1 = o_P(1)$.

We can also immediately apply Lemma G.3 (iii) to establish that $A_3 = o_P(1)$.

Also, it follows from Lemma G.1 (iv) that

$$|A_5| = \left| \frac{1}{\sqrt{nT}}\tilde{\eta}'P_{\hat{U}_j}\tilde{\epsilon} \right| \leq \sqrt{nT} \left\| \frac{1}{\sqrt{nT}}P_{\hat{U}_j}\tilde{\epsilon} \right\|_2 \left\| \frac{1}{\sqrt{nT}}P_{\hat{U}_j}\tilde{\eta} \right\|_2 = O_P\left(\frac{|J|_0 \log p}{\sqrt{nT}}\right) = o_P(1)$$

since $|J|_0^2 \log^2 p = o(nT)$.

Finally, it follows immediately from Lemma F.2 that $A_6 = 0$. ■

A.3. Showing $A_2 = o_P(1)$. By (A.2),

$$\begin{aligned} A_2 &= \frac{1}{\sqrt{nT}}\tilde{\eta}'M_{\hat{U}_j}(I_T \otimes M_{\hat{F}}(\tilde{F}H - \hat{F})H^{-1})\tilde{\Xi} \\ &= \frac{1}{\sqrt{nT}}\tilde{\eta}'M_{\hat{U}_j}(I_T \otimes M_{\hat{F}}(\tilde{F}H - \hat{F})H^{-1})\tilde{\Xi} \end{aligned} \tag{A.4}$$

$$+ \tilde{\Xi}'(I_T \otimes H'^{-1}(\tilde{F}H - \hat{F})'M_{\hat{F}})\frac{1}{\sqrt{nT}}M_{\hat{U}_j}(I_T \otimes M_{\hat{F}}(\tilde{F}H - \hat{F})H^{-1})\tilde{\Delta}_d \tag{A.5}$$

$$- \tilde{\Xi}'(I_T \otimes H'^{-1}(\tilde{F}H - \hat{F})'M_{\hat{F}})\frac{1}{\sqrt{nT}}M_{\hat{U}_j}(I_T \otimes P_{\hat{F}})\tilde{\eta} \tag{A.6}$$

$$+ \tilde{\Xi}'(I_T \otimes H'^{-1}(\tilde{F}H - \hat{F})'M_{\hat{F}})\frac{1}{\sqrt{nT}}M_{\hat{U}_j}(I_T \otimes M_{\hat{F}})\tilde{U}\gamma_d. \tag{A.7}$$

It follows from Lemma G.3(i) that (A.4) is $o_P(1)$. By the Cauchy-Schwarz inequality and under the assumption that $\sqrt{nT}\Delta_{\hat{F}}^2 = o(1)$, (A.5) is bounded by

$$\begin{aligned} &|\tilde{\Xi}'(I_T \otimes H'^{-1}(\tilde{F}H - \hat{F})'M_{\hat{F}})\frac{1}{\sqrt{nT}}M_{\hat{U}_j}(I_T \otimes M_{\hat{F}}(\tilde{F}H - \hat{F})H^{-1})\tilde{\Delta}_d| \\ &\leq \frac{1}{\sqrt{nT}} \max_{G=\tilde{\Xi}, \tilde{\Delta}_d} \|G'(I_T \otimes H'^{-1}(\tilde{F}H - \hat{F})'M_{\hat{F}})M_{\hat{U}_j}\|_2^2 \\ &\leq \frac{1}{\sqrt{nT}} \max_{G=\tilde{\Xi}, \tilde{\Delta}_d} \|G'(I_T \otimes H'^{-1}(\tilde{F}H - \hat{F})'M_{\hat{F}})\|_2^2 \\ &\leq \frac{1}{\sqrt{nT}} \max_{g_t=\tilde{\xi}_t, \tilde{\delta}_{at}} \sum_t \|g_t'H'^{-1}(\tilde{F}H - \hat{F})'M_{\hat{F}}\|_2^2 \end{aligned}$$

$$\leq O_P\left(\frac{\sqrt{T}}{\sqrt{n}}\right)\|\tilde{F}H - \hat{F}\|_F^2 = O_P(\sqrt{nT}\Delta_F^2) = o_P(1).$$

Term (A.6) equals

$$\begin{aligned} & -\frac{1}{\sqrt{nT}}\tilde{\Xi}'(I_T \otimes H'^{-1}(\tilde{F}H - \hat{F})'M_{\hat{F}})M_{\hat{U}_j}(I_T \otimes P_{\hat{F}})\tilde{\eta} \\ & = -\frac{1}{\sqrt{nT}}\tilde{\Xi}'(I_T \otimes H'^{-1}(\tilde{F}H - \hat{F})'M_{\hat{F}})(I_T \otimes P_{\hat{F}})\tilde{\eta} = 0 \end{aligned}$$

where the first equality is due to $P_{\hat{U}_j}(I_T \otimes P_{\hat{F}}) = 0$ and the second equality is due to $M_{\hat{F}}P_{\hat{F}} = 0$ and the fact that the kronecker product satisfies $(A \otimes B)(C \otimes D) = AC \otimes BD$.

Finally, using $M_{\hat{F}}P_{\hat{F}} = 0$ and $P_{\hat{U}_j}(I_T \otimes P_{\hat{F}}) = 0$, (A.7) equals

$$\begin{aligned} & \tilde{\Xi}'(I_T \otimes H'^{-1}(\tilde{F}H - \hat{F})'M_{\hat{F}})\frac{1}{\sqrt{nT}}M_{\hat{U}_j}(I_T \otimes M_{\hat{F}})\tilde{U}\gamma_d \\ & = \tilde{\Xi}'(I_T \otimes H'^{-1}(\tilde{F}H - \hat{F})'M_{\hat{F}})\frac{1}{\sqrt{nT}}M_{\hat{U}_j}\tilde{U}\gamma_d \\ & \quad - \tilde{\Xi}'(I_T \otimes H'^{-1}(\tilde{F}H - \hat{F})'M_{\hat{F}})\frac{1}{\sqrt{nT}}M_{\hat{U}_j}(I_T \otimes P_{\hat{F}})\tilde{U}\gamma_d \\ & = \tilde{\Xi}'(I_T \otimes H'^{-1}(\tilde{F}H - \hat{F})'M_{\hat{F}})\frac{1}{\sqrt{nT}}M_{\hat{U}_j}\tilde{U}\gamma_d \\ & = o_P(1) \end{aligned} \tag{A.8}$$

where the last equality follows from Lemma G.3 (vi).

Hence, $A_2 = o_P(1)$. ■

A.4. Showing $A_4 = o_P(1)$.

$$\begin{aligned} A_4 & = \frac{1}{\sqrt{nT}}\tilde{\eta}'M_{\hat{U}_j}(I_T \otimes M_{\hat{F}})\tilde{U}\theta \\ & = \frac{1}{\sqrt{nT}}\tilde{\eta}'M_{\hat{U}_j}(I_T \otimes M_{\hat{F}})\tilde{U}\theta \end{aligned} \tag{A.9}$$

$$+ \frac{1}{\sqrt{nT}}\theta'\tilde{U}'(I_T \otimes M_{\hat{F}})M_{\hat{U}_j}(I_T \otimes M_{\hat{F}})\tilde{U}\gamma_d \tag{A.10}$$

$$+ \frac{1}{\sqrt{nT}}\theta'\tilde{U}'(I_T \otimes M_{\hat{F}})M_{\hat{U}_j}(I_T \otimes M_{\hat{F}}(\tilde{F}H - \hat{F})H^{-1})\tilde{\Delta}_d. \tag{A.11}$$

It follows from Lemma G.3 (iii) that term (A.9) is $o_P(1)$.

By Lemma G.1 (i) and (ii), we can bound term (A.10) by

$$\frac{1}{\sqrt{nT}}\theta'\tilde{U}'(I_T \otimes M_{\hat{F}})M_{\hat{U}_j}(I_T \otimes M_{\hat{F}})\tilde{U}\gamma_d$$

$$\begin{aligned}
&\leq \sqrt{nT} \max_{g=\theta, \gamma_d} \left\| \frac{1}{\sqrt{nT}} M_{\hat{U}_{\hat{J}}} (I_T \otimes M_{\hat{F}}) \tilde{U}g \right\|_2^2 \\
&\leq 2\sqrt{nT} \max_{g=\theta, \gamma_d} \left\| \frac{1}{\sqrt{nT}} M_{\hat{U}_{\hat{J}}} \tilde{U}g \right\|_2^2 + 2\sqrt{nT} \max_{g=\theta, \gamma_d} \left\| \frac{1}{\sqrt{nT}} (I_T \otimes P_{\hat{F}}) \tilde{U}g \right\|_2^2 \\
&\leq \sqrt{nT} O_P \left(\|R_y\|_1^2 + \kappa_n^2 |J|_0 + |J|_0^2 \Delta_F^2 + \frac{|J|_0}{n} \right) = o_P(1),
\end{aligned}$$

under the assumption $\left(\kappa_n^2 |J|_0 + \|R_y\|_1^2 + \Delta_F^2 |J|_0^2 + \frac{|J|_0}{n} \right) \sqrt{nT} = o(1)$.

The same argument as that employed in the bound given by equation (A.8) yields that term (A.11) is $o_P(1)$. ■

A.5. Proof of Theorem 3.1. (i) Write $\iota_{it} := (\eta_{it} - \bar{\eta}_i)^2$.

Step 1: Show $|\frac{1}{nT} \tilde{\eta}' \hat{\eta} - \frac{1}{nT} \sum_{it} E \iota_{it}| = o_P(1)$.

It follows from Proposition A.1 that $|\frac{1}{nT} \tilde{\eta}' \hat{\eta} - \frac{1}{nT} \tilde{\eta}' \tilde{\eta}| = o_P(1)$. Also,

$$\frac{1}{nT} \tilde{\eta}' \tilde{\eta} = \frac{1}{nT} \sum_{i,t} \tilde{\eta}_{it}^2 = \frac{1}{nT} \sum_{i,t} (\eta_{it} - \bar{\eta}_i)^2 - \frac{1}{T} \sum_t \bar{\eta}_t^2 + \bar{\eta}^2.$$

We have that $E \left[\frac{1}{T} \sum_t \bar{\eta}_t^2 \right] = \frac{1}{T} \sum_t \frac{1}{n^2} \sum_i E [\eta_{it}^2] = O(1/n)$ and that $\bar{\eta}^2 = o_P(1)$. Hence,

$$\frac{1}{nT} \tilde{\eta}' \tilde{\eta} = \frac{1}{nT} \sum_{it} (\eta_{it} - \bar{\eta}_i)^2 + o_P(1) = \frac{1}{nT} \sum_{it} \iota_{it} + o_P(1).$$

Note that

$$\text{Var} \left(\frac{1}{nT} \sum_{it} \iota_{it} \right) = \frac{1}{n^2 T^2} \sum_i \text{Var} \left(\sum_t \iota_{it} \right) = O(1/n).$$

Hence, $|\frac{1}{nT} \sum_{it} \iota_{it} - \frac{1}{nT} \sum_{it} E \iota_{it}| = o_P(1)$. We then have

$$\left| \frac{1}{nT} \tilde{\eta}' \tilde{\eta} - \frac{1}{nT} \sum_{it} E \iota_{it} \right| = o_P(1), \quad \left| \frac{1}{nT} \tilde{\eta}' \hat{\eta} - \frac{1}{nT} \sum_{it} E \iota_{it} \right| = o_P(1), \quad (\text{A.12})$$

and $\frac{1}{nT} \tilde{\eta}' \hat{\eta}$ is bounded away from zero.

Let $z_{n,i} = \frac{1}{\sqrt{T}} \sum_t (\eta_{it} - \bar{\eta}_i) (\epsilon_{it} - \bar{\epsilon}_i)$, $b_n = \left[\text{Var} \left(\frac{1}{\sqrt{n}} \sum_i z_{n,i} \right) \right]^{-1/2}$, and $x_{n,i} = b_n z_{n,i}$. In addition, let $s_n^2 = \sum_i \text{Var}(x_{n,i}) = \sum_i \text{Var}(z_{n,i}) b_n^2 = n$.

Step 2: Show $\frac{b_n}{\sqrt{nT}} \tilde{\eta}' \tilde{\epsilon} = \frac{1}{s_n} \sum_i x_{n,i} + o_P(1)$.

Note that

$$\begin{aligned}
 E \left(\frac{1}{T} \sum_t \bar{\eta}_{\cdot t} \bar{\epsilon}_{\cdot t} \right)^2 &= \frac{1}{T} \sum_t \frac{1}{n} \sum_i \frac{1}{n} \sum_j \frac{1}{T} \sum_{s=1}^T \frac{1}{n} \sum_{l=1}^n \frac{1}{n} \sum_{v=1}^n E \eta_{ls} \epsilon_{vs} \eta_{it} \epsilon_{jt} \\
 &= \sum_t \frac{1}{n^4} \sum_i \frac{1}{T^2} \sum_{s=1}^T E \eta_{is} \epsilon_{is} \eta_{it} \epsilon_{it} \\
 &\quad + \frac{1}{T} \sum_t \frac{1}{n^4} \sum_i \sum_{j \neq i} \frac{1}{T} \sum_{s=1}^T E \eta_{js} \epsilon_{jt} E \epsilon_{is} \eta_{it} \\
 &\quad + \frac{1}{T} \sum_t \frac{1}{n^4} \sum_i \sum_{j \neq i} \frac{1}{T} \sum_{s=1}^T E \eta_{is} \eta_{it} E \epsilon_{js} \epsilon_{jt} \\
 &\quad + \frac{1}{T} \sum_t \frac{1}{n^4} \sum_i \frac{1}{T} \sum_{s=1}^T \sum_{l \neq i} E \eta_{ls} \epsilon_{ls} \eta_{it} \epsilon_{it} \\
 &\quad + \frac{1}{T} \sum_t \frac{1}{n^4} \sum_i \sum_{j \neq i} \frac{1}{T} \sum_{s=1}^T E \eta_{js} \epsilon_{js} \eta_{it} \epsilon_{jt} \\
 &= O \left(\frac{1}{n^2} \right).
 \end{aligned} \tag{A.13}$$

We have $E[z_{n,i}] = 0$. We also have, under our assumptions, $\text{Var}(z_{n,i})$ and b_n bounded away from both zero and infinity uniformly in i . Then

$$\begin{aligned}
 \frac{b_n}{\sqrt{nT}} \tilde{\eta}' \tilde{\epsilon} &= \frac{b_n}{\sqrt{nT}} \sum_{it} (\eta_{it} - \bar{\eta}_{i \cdot}) (\epsilon_{it} - \bar{\epsilon}_{i \cdot}) - \frac{b_n \sqrt{nT}}{T} \sum_t \bar{\eta}_{\cdot t} \bar{\epsilon}_{\cdot t} + b_n \sqrt{nT} \bar{\eta} \bar{\epsilon} \\
 &= \frac{1}{s_n} \sum_i x_{n,i} + o_P(1).
 \end{aligned}$$

Step 3: Apply the CLT.

We now verify the Lindeberg condition for the triangular array $\{x_{n,i}\}$. For any $\varepsilon > 0$,

$$E \left(\frac{1}{n} \sum_i x_{n,i}^2 1\{|x_{n,i}| > \varepsilon \sqrt{n}\} \right) \leq E \left[\frac{1}{n} \sum_i x_{n,i}^2 \right] = 1.$$

Hence by the dominated convergence theorem,

$$s_n^{-2} \sum_i E [x_{n,i}^2 1\{|x_{n,i}| > \varepsilon s_n\}] = E \left(\frac{1}{n} \sum_i x_{n,i}^2 1\{|x_{n,i}| > \varepsilon \sqrt{n}\} \right) \rightarrow 0.$$

This implies, by the Lindeberg central limit theorem,

$$\frac{b_n}{\sqrt{nT}} \tilde{\eta}' \tilde{\epsilon} \rightarrow^d \mathcal{N}(0, 1).$$

In the previous subsections, we have proven $A_i = o_P(1)$ for $i = 1, \dots, 6$. Hence, it follows from (A.1) that $\sqrt{nT}(\frac{1}{nT}\hat{\eta}'\hat{\eta})(\hat{\alpha} - \alpha) = \frac{1}{\sqrt{nT}}\tilde{\eta}'\tilde{\epsilon} + o_P(1)$. In addition,

$$\begin{aligned} & b_n \left| \left(\frac{1}{nT}\hat{\eta}'\hat{\eta} - \frac{1}{nT} \sum_{i,t} E[l_{it}] \right) \sqrt{nT}(\hat{\alpha} - \alpha) \right| \\ & \leq b_n \left| \frac{1}{nT}\hat{\eta}'\hat{\eta} - \frac{1}{nT} \sum_{i,t} E[l_{it}] \right| \left(\frac{1}{nT}\hat{\eta}'\hat{\eta} \right)^{-1} \left| \frac{1}{\sqrt{nT}}\tilde{\eta}'\tilde{\epsilon} + o_P(1) \right| \\ & = o_P(1). \end{aligned}$$

Therefore,

$$\begin{aligned} \sigma_{\eta\epsilon}^{-1/2} \sigma_\eta^2 \sqrt{nT}(\hat{\alpha} - \alpha) &= b_n \frac{1}{nT} \sum_{i,t} E[l_{it}] \sqrt{nT}(\hat{\alpha} - \alpha) \\ &= b_n \frac{1}{nT} \hat{\eta}'\hat{\eta} \sqrt{nT}(\hat{\alpha} - \alpha) + o_P(1) \\ &= \frac{b_n}{\sqrt{nT}} \tilde{\eta}'\tilde{\epsilon} + o_P(1) \xrightarrow{d} \mathcal{N}(0, 1). \end{aligned}$$

(ii) To verify normality with the estimated asymptotic variance, we need to prove consistency of $\hat{\sigma}_{\eta\epsilon}$ and $\hat{\sigma}_\eta^2$. We have previously shown $|\frac{1}{nT}\hat{\eta}'\hat{\eta} - \frac{1}{nT} \sum_{i,t} E[l_{it}]| = o_P(1)$ which establishes consistency of $\sigma_\eta^2 = \frac{1}{nT} \sum_{i,t} E[l_{it}]$. Hence, it remains to prove $\hat{\sigma}_{\eta\epsilon} - \sigma_{\eta\epsilon} = o_P(1)$. Recall that

$$\sigma_{\eta\epsilon} = \text{Var} \left(\frac{1}{\sqrt{nT}} \sum_{i=1}^n \sum_{t=1}^T (\eta_{it} - \bar{\eta}_i) (\epsilon_{it} - \bar{\epsilon}_i) \right).$$

Step 1: Bound $\Delta_1 := \hat{\sigma}_{\eta\epsilon} - \frac{1}{nT} \sum_{i=1}^n (\sum_{t=1}^T \tilde{\eta}_{it} \tilde{\epsilon}_{it})^2$. We have

$$\begin{aligned} \Delta_1 &= \frac{1}{nT} \sum_{i=1}^n \left[\left(\sum_{t=1}^T \hat{\eta}_{it} \hat{\epsilon}_{it} \right)^2 - \left(\sum_{t=1}^T \tilde{\eta}_{it} \tilde{\epsilon}_{it} \right)^2 \right] \\ &= \frac{1}{nT} \sum_{i=1}^n \left[\sum_{t=1}^T (\hat{\eta}_{it} \hat{\epsilon}_{it} + \tilde{\eta}_{it} \tilde{\epsilon}_{it}) \right] \left[\sum_{t=1}^T (\hat{\eta}_{it} \hat{\epsilon}_{it} - \tilde{\eta}_{it} \tilde{\epsilon}_{it}) \right] \\ &= \frac{1}{nT} \sum_{i=1}^n \left[\sum_{t=1}^T (\hat{\eta}_{it} \hat{\epsilon}_{it} - \tilde{\eta}_{it} \tilde{\epsilon}_{it}) \right]^2 \tag{A.14} \end{aligned}$$

$$+ \frac{2}{n} \sum_{i=1}^n \left(\frac{1}{T} \sum_{t=1}^T \tilde{\eta}_{it} \tilde{\epsilon}_{it} \right) \sum_{s=1}^T (\hat{\eta}_{is} \hat{\epsilon}_{is} - \tilde{\eta}_{is} \tilde{\epsilon}_{is}). \tag{A.15}$$

By Lemma G.4, term (A.14) is $o_P(1)$.

For term (A.15), we have

$$E \left[\frac{1}{n} \sum_i \left| \frac{1}{T} \sum_{t=1}^T \tilde{\eta}_{it} \tilde{\epsilon}_{it} \right|^2 \right] = E \left[\frac{1}{n} \sum_i \left| \frac{1}{T} \sum_{t=1}^T (\eta_{it} - \bar{\eta}_t) (\epsilon_{it} - \bar{\epsilon}_t) - (\bar{\eta}_i - \bar{\eta}) (\bar{\epsilon}_i - \bar{\epsilon}) \right|^2 \right]$$

$$= O\left(\frac{1}{T}\right).$$

since the process $\{(\eta_t, \epsilon_t)\}_{t=-\infty}^{+\infty}$ satisfies the strong mixing condition with exponential tails. Thus by Cauchy-Schwarz,

$$\begin{aligned} & \left[\frac{1}{n} \sum_{i=1}^n \left(\frac{1}{T} \sum_{t=1}^T \tilde{\eta}_{it} \tilde{\epsilon}_{it} \right) \sum_{s=1}^T (\hat{\eta}_{is} \hat{\epsilon}_{is} - \tilde{\eta}_{is} \tilde{\epsilon}_{is}) \right]^2 \\ & \leq \left[\frac{1}{n} \sum_{i=1}^n \left(\frac{1}{T} \sum_{t=1}^T \tilde{\eta}_{it} \tilde{\epsilon}_{it} \right)^2 \right] \left[\frac{1}{n} \sum_{i=1}^n \left(\sum_{s=1}^T (\hat{\eta}_{is} \hat{\epsilon}_{is} - \tilde{\eta}_{is} \tilde{\epsilon}_{is}) \right)^2 \right] \\ & \leq O_P\left(\frac{1}{T}\right) \frac{1}{n} \sum_{i=1}^n \left(\sum_{s=1}^T (\hat{\eta}_{is} \hat{\epsilon}_{is} - \tilde{\eta}_{is} \tilde{\epsilon}_{is}) \right)^2 \\ & = o_P(1). \end{aligned}$$

This result then implies $\Delta_1 = o_P(1)$.

$$\text{Step 2: Bound } \Delta_2 := \frac{1}{nT} \sum_{i=1}^n \left(\sum_{t=1}^T \tilde{\eta}_{it} \tilde{\epsilon}_{it} \right)^2 - \frac{1}{nT} \sum_{i=1}^n \left(\sum_{t=1}^T (\eta_{it} - \bar{\eta}_i) (\epsilon_{it} - \bar{\epsilon}_i) \right)^2.$$

Note that

$$\sum_{t=1}^T \tilde{\eta}_{it} \tilde{\epsilon}_{it} = \sum_{t=1}^T (\eta_{it} - \bar{\eta}_i) (\epsilon_{it} - \bar{\epsilon}_i) - \underbrace{\sum_{t=1}^T \bar{\eta}_t (\epsilon_{it} - \bar{\epsilon}_t) - \sum_{t=1}^T \eta_{it} \bar{\epsilon}_t + T \bar{\eta} (\bar{\epsilon}_i - \bar{\epsilon}) + T \bar{\eta}_i \bar{\epsilon}}_{B_i}.$$

and that $\frac{1}{T} \sum_{t=1}^T (\eta_{it} - \bar{\eta}_i) (\epsilon_{it} - \bar{\epsilon}_i) = \frac{1}{T} \sum_{t=1}^T \eta_{it} \epsilon_{it} - \bar{\eta}_i \bar{\epsilon}_i$.

Hence,

$$\begin{aligned} \frac{1}{nT} \sum_{i=1}^n \left(\sum_{t=1}^T \tilde{\eta}_{it} \tilde{\epsilon}_{it} \right)^2 &= \frac{1}{nT} \sum_{i=1}^n \left(\sum_{t=1}^T (\eta_{it} - \bar{\eta}_i) (\epsilon_{it} - \bar{\epsilon}_i) + B_i \right)^2 \\ &= \frac{1}{nT} \sum_{i=1}^n \left(\sum_{t=1}^T (\eta_{it} - \bar{\eta}_i) (\epsilon_{it} - \bar{\epsilon}_i) \right)^2 + \frac{1}{nT} \sum_{i=1}^n B_i^2 \\ &\quad + \frac{2}{nT} \sum_{i=1}^n B_i \sum_{t=1}^T (\eta_{it} - \bar{\eta}_i) (\epsilon_{it} - \bar{\epsilon}_i). \end{aligned}$$

Note that

$$\begin{aligned} & \left[\frac{1}{nT} \sum_{i=1}^n B_i \sum_{t=1}^T (\eta_{it} - \bar{\eta}_i) (\epsilon_{it} - \bar{\epsilon}_i) \right]^2 \\ & \leq \frac{1}{nT} \sum_i B_i^2 \frac{1}{n} \sum_i \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T (\eta_{it} - \bar{\eta}_i) (\epsilon_{it} - \bar{\epsilon}_i) \right)^2 \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{nT} \sum_i B_i^2 O_P \left(\frac{1}{n} \sum_i \text{Var} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T (\eta_{it} - \bar{\eta}_i) (\epsilon_{it} - \bar{\epsilon}_i) \right) \right) \\
&= \frac{1}{nT} \sum_i B_i^2 O_P(\sigma_{\eta\epsilon}).
\end{aligned}$$

Therefore, $|\Delta_2| \leq \frac{1}{nT} \sum_i B_i^2 + (\frac{1}{nT} \sum_i B_i^2)^{1/2} O_P(1)$. It suffices to prove $\frac{1}{nT} \sum_i B_i^2 = o_P(1)$. In fact, $\frac{1}{nT} \sum_i B_i^2 \leq C \sum_{l=1}^4 \bar{A}_l$ for a constant $C > 0$ and

$$\begin{aligned}
\bar{A}_1 &= \frac{1}{nT} \sum_{i=1}^n \left(\sum_{t=1}^T \eta_{it} \bar{\epsilon}_t \right)^2, & \bar{A}_2 &= \frac{1}{nT} \sum_{i=1}^n \left(\sum_{t=1}^T \tilde{\eta}_t (\epsilon_{it} - \bar{\epsilon}_t) \right)^2, \\
\bar{A}_3 &= \frac{T}{n} \bar{\epsilon}^2 \sum_{i=1}^n \tilde{\eta}_i^2, & \bar{A}_4 &= \frac{T}{n} \bar{\eta}^2 \sum_{i=1}^n (\bar{\epsilon}_i - \bar{\epsilon})^2.
\end{aligned}$$

where each $\bar{A}_l = O_P(E\bar{A}_l)$. We then have

$$\begin{aligned}
E\bar{A}_1 &= \frac{1}{n^3 T} \sum_{j=1}^n \sum_{i=1}^n \sum_{m=1}^n \sum_{s=1}^T \sum_{t=1}^T E \eta_{it} \epsilon_{jt} \eta_{is} \epsilon_{ms} = \frac{1}{n^3 T} \sum_{i=1}^n \sum_{s=1}^T \sum_{t=1}^T E \eta_{it} \epsilon_{it} \eta_{is} \epsilon_{is} \\
&\quad + \frac{1}{n^3 T} \sum_{i=1}^n \sum_{j \neq i} \sum_{s=1}^T \sum_{t=1}^T E \eta_{it} \eta_{is} E \epsilon_{js} \epsilon_{jt} = O\left(\frac{T}{n}\right) = o(1).
\end{aligned}$$

Similarly, $E\bar{A}_2 = o(1)$. In addition, $\bar{\epsilon}^2 = O_P(n^{-1})$ and $\bar{\eta}^2 = O_P(n^{-1})$, so \bar{A}_3 and \bar{A}_4 are each $O_P(T/n) = o_P(1)$. Combining verifies that

$$\Delta_2 := \frac{1}{nT} \sum_{i=1}^n \left(\sum_{t=1}^T \tilde{\eta}_t \tilde{\epsilon}_{it} \right)^2 - \frac{1}{nT} \sum_{i=1}^n \left(\sum_{t=1}^T (\eta_{it} - \bar{\eta}_i) (\epsilon_{it} - \bar{\epsilon}_i) \right)^2 = o_P(1). \quad (\text{A.16})$$

Step 3: Bound $\Delta_3 := \frac{1}{nT} \sum_{i=1}^n (\sum_{t=1}^T (\eta_{it} - \bar{\eta}_i) (\epsilon_{it} - \bar{\epsilon}_i))^2 - \sigma_{\eta\epsilon}$.

Note that $\sigma_{\eta\epsilon} = E \left[\frac{1}{nT} \sum_{i=1}^n \left(\sum_{t=1}^T (\eta_{it} - \bar{\eta}_i) (\epsilon_{it} - \bar{\epsilon}_i) \right)^2 \right]$, and let

$$z_{n,i} = \frac{1}{\sqrt{T}} \sum_t (\eta_{it} - \bar{\eta}_i) (\epsilon_{it} - \bar{\epsilon}_i).$$

Then $\Delta_3 = \frac{1}{n} \sum_{i=1}^n (z_{n,i}^2 - E z_{n,i}^2)$. Because $\frac{1}{n} \sum_i \text{Var}(z_{n,i}^2) = O(1)$, we have

$$E\Delta_3^2 = \text{Var}(\Delta_3) = \text{Var} \left(\frac{1}{n} \sum_i z_{n,i}^2 \right) = \frac{1}{n^2} \sum_i \text{Var}(z_{n,i}^2) = o(1), \quad (\text{A.17})$$

which implies

$$\Delta_3 := \frac{1}{nT} \sum_{i=1}^n \left(\sum_{t=1}^T (\eta_{it} - \bar{\eta}_i) (\epsilon_{it} - \bar{\epsilon}_i) \right)^2 - \sigma_{\eta\epsilon} = o_P(1). \quad (\text{A.18})$$

Combining the above three steps, we reach $|\hat{\sigma}_{\eta\epsilon} - \sigma_{\eta\epsilon}| = o_P(1)$. ■

Proof of Corollary 3.1 Given Theorem 3.1, the corollary follows from the same argument as that of Corollary 1 (i) of Belloni et al. (2014b). We thus refer to Belloni et al. (2014b) for details. ■

APPENDIX B. CONVERGENCE OF THE k -STEP BOOTSTRAP LASSO

In this section, we obtain the statistical convergence rate (in O_{P^*}) of the k -step bootstrap lasso estimators $\tilde{\gamma}_d^*$ $\tilde{\gamma}_y^*$. We focus on $\tilde{\gamma}_y^*$, as the proof of $\tilde{\gamma}_d^*$ is similar. Recall that

$$\begin{aligned} \mathcal{L}_y^*(\gamma) &= \frac{1}{nT} \sum_{t=1}^T \sum_{i=1}^n (\tilde{y}_{it}^* - \hat{\delta}_{yt}^{*'} \hat{f}_i^* - \hat{U}_{it}^{*'} \gamma)^2, \\ \tilde{\gamma}_{y,lasso}^* &= \arg \min_{\gamma \in \mathbb{R}^p} \mathcal{L}_y^*(\gamma) + \kappa_n \|\hat{\Psi}^y \gamma\|_1. \end{aligned} \tag{B.1}$$

and that

- $\hat{\gamma}_y$ = the post-lasso estimator based on the original data
- $\tilde{\gamma}_y^*$ = the k -step lasso estimator based on the bootstrap data
- $\tilde{\gamma}_{y,lasso}^*$ = the lasso estimator based on the bootstrap data if a complete lasso program is carried out.

In particular, $\hat{\gamma}_y$ is used as the coefficient when generating the bootstrap data.

We divide the proof into three subsections. Section B.1 proves the statistical convergence of $\|\tilde{\gamma}_{y,lasso}^* - \hat{\gamma}_y\|_1$ in the bootstrap sampling space. Section B.2 quantifies the computational error $\|\tilde{\gamma}_y^* - \tilde{\gamma}_{y,lasso}^*\|_1$ and shows that the computational error of the k -step lasso is negligible using the assumed high-level conditions on the iterative scheme $\mathcal{S}_y(\cdot)$. Section B.3 verifies the high-level conditions for the coordinate descent, or “shooting”, method (Fu (1998); Kadkhodaie et al. (2014)).

We employ the usual definition of $o_{P^*}(1)$ and $O_{P^*}(1)$. We say that a sequence X_n^* in the bootstrap sampling space is $o_{P^*}(1)$ if for any $\varepsilon, \delta > 0$,

$$P\{P^*(|X_n^*| > \varepsilon) > \delta\} \rightarrow 0,$$

and that $X_n^* = O_{P^*}(1)$ if for any $\delta > 0$, there is $M > 0$, such that

$$P\{P^*(|X_n^*| > M) > \delta\} \rightarrow 0.$$

B.1. The convergence of lasso on bootstrap data. The main result in this subsection is the following proposition.

Proposition B.1. $\|\tilde{\gamma}_{y,lasso}^* - \hat{\gamma}_y\|_1 = O_{P^*}(\kappa_n |J|_0)$.

Proof. Recall that \widehat{F} , $\widehat{\delta}_{yt}$, and $\widehat{\gamma}_y$ respectively denote the estimated factors, the estimator of δ_{yt} , and the post-lasso estimator of γ_y using the original data. We also have that \widetilde{U}_t^* denotes the wild bootstrapped idiosyncratic term in the factor equation and that the following relations hold:

$$\widetilde{Y}_t^* = \widehat{F}\widehat{\delta}_{yt} + \widetilde{U}_t^*\widehat{\gamma}_y + \widetilde{e}_t^*, \quad \widetilde{e}_t^* = \widetilde{\epsilon}_t^* + \widetilde{\eta}_t^*\widehat{\alpha}. \quad (\text{B.2})$$

In addition, recall that $\widehat{\delta}_{yt}^*$ and \widehat{U}_t^* denote the estimates obtained from the bootstrap data. Define

$$M_t^* = \widehat{F}\widehat{\delta}_{yt} - \widehat{F}^*\widehat{\delta}_{yt}^* + (\widetilde{U}_t^* - \widehat{U}_t^*)\widehat{\gamma}_y, \quad \Delta_\gamma^* = \widehat{\gamma}_y - \widetilde{\gamma}_{y,lasso}^*. \quad (\text{B.3})$$

By definition, $\mathcal{L}_y^*(\widetilde{\gamma}_{y,lasso}^*) + \kappa_n \|\widehat{\Psi}^y \widetilde{\gamma}_{y,lasso}^*\|_1 \leq \mathcal{L}_y^*(\widehat{\gamma}_y) + \kappa_n \|\widehat{\Psi}^y \widehat{\gamma}_y\|_1$, which implies

$$\frac{1}{nT} \sum_{t=1}^T \left(\|\widehat{U}_t^* \Delta_\gamma^*\|_2^2 + 2(\widetilde{e}_t^{*'} + M_t^{*'})\widehat{U}_t^* \Delta_\gamma^* \right) + \kappa_n \|\widehat{\Psi}^y \widetilde{\gamma}_{y,lasso}^*\|_1 \leq \kappa_n \|\widehat{\Psi}^y \widehat{\gamma}_y\|_1. \quad (\text{B.4})$$

By Lemma H.4 and $\kappa_n = \frac{2c_0}{\sqrt{nT}}\Phi^{-1}(1 - q_n/(2p))$ for some $c_0 > 1$,

$$\begin{aligned} & \left| \frac{1}{nT} \sum_{t=1}^T 2(\widetilde{e}_t^{*'} + M_t^{*'})\widehat{U}_t^* \Delta_\gamma^* \right| \\ & \leq \left\| \frac{1}{nT} \sum_{t=1}^T 2\widetilde{e}_t^{*'} \widetilde{U}_t^* \widehat{\Psi}^{y-1} \right\|_\infty \|\widehat{\Psi}^y \Delta_\gamma^*\|_1 + \left(\left\| \frac{1}{nT} \sum_{t=1}^T 2\widetilde{e}_t^{*'} (\widehat{U}_t^* - \widetilde{U}_t^*) \right\|_\infty \right. \\ & \quad \left. + \left\| \frac{1}{nT} \sum_{t=1}^T 2M_t^{*'} \widehat{U}_t^* \Delta_\gamma^* \right\|_\infty \right) \|\widehat{\Psi}^y \Delta_\gamma^*\|_1 \max_m \widehat{\Psi}_m^y \\ & \leq \left[\frac{2}{\sqrt{nT}} \Phi^{-1}(1 - \frac{q_n}{2p})(1 + o_{P^*}(1)) + o_{P^*} \left(\sqrt{\frac{\log p}{nT}} \right) \right] \|\widehat{\Psi}^y \Delta_\gamma^*\|_1 \\ & \leq \frac{(c_0 + 1)}{\sqrt{nT}} \Phi^{-1} \left(1 - \frac{q_n}{2p} \right) \|\widehat{\Psi}^y \Delta_\gamma^*\|_1 \\ & = \frac{c_0 + 1}{2c_0} \kappa_n \|\widehat{\Psi}^y \Delta_\gamma^*\|_1. \end{aligned}$$

with P^* approaching one. Equation (B.4) then implies, for the support set \widehat{J} of $\widehat{\gamma}_y$,

$$\frac{1}{nT} \sum_{t=1}^T \|\widehat{U}_t^* \Delta_\gamma^*\|_2^2 + \frac{c_0 - 1}{2c_0} \kappa_n \|(\widehat{\Psi}^y \Delta_\gamma^*)_{\widehat{J}^c}\|_1 \leq \kappa_n \|(\widehat{\Psi}^y \Delta_\gamma^*)_{\widehat{J}}\|_1 \frac{3c_0 + 1}{2c_0}. \quad (\text{B.5})$$

Hence, $\|(\Delta_\gamma^*)_{\widehat{J}^c}\|_1 \leq c \|(\Delta_\gamma^*)_{\widehat{J}}\|_1$ for some $c > 0$. This also implies for some generic $C > 0$, $\|\Delta_\gamma^*\|_1^2 \leq C \|\Delta_{\gamma, \widehat{J}}^*\|_1^2 \leq C \|\Delta_\gamma^*\|_2^2 O_P(|\widehat{J}|_0)$ as $|\widehat{J}|_0 = O_P(|J|_0)$.

We can now apply Lemma H.4 to obtain

$$\frac{1}{nT} \sum_{t=1}^T \|\widehat{U}_t^* \Delta_\gamma^*\|_2^2 \geq \frac{1}{nT} \sum_{t=1}^T \|\widetilde{U}_t \Delta_\gamma^*\|_2^2 - \|\Delta_\gamma^*\|_2^2 o_{P^*}(1).$$

In addition, the sparse-eigenvalue condition implies the restricted eigenvalue condition: For any $m > 0$, there is $\underline{\phi} > 0$, so that, for $|\widehat{J}| = O_P(|J|_0)$, with probability arbitrarily close to one,

$$\inf_{\delta \in \mathbb{R}^p: \|\delta_{\widehat{J}}\|_1 \leq m \|\delta_{\widehat{J}^c}\|_1} \frac{\delta' \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T \tilde{U}_{it} \tilde{U}'_{it} \delta}{\delta' \delta} \geq \underline{\phi}.$$

Hence $\frac{1}{nT} \sum_{t=1}^T \|\widehat{U}_t^* \Delta_\gamma^*\|_2^2 \geq \underline{\phi} \|\Delta_\gamma^*\|_2^2 / 2$.

$$\frac{1}{nT} \sum_{t=1}^T \|\widehat{U}_t^* \Delta_\gamma^*\|_2^2 = O_{P^*}(\kappa_n^2 |J|_0), \quad \|\Delta_\gamma^*\|_1 = O_{P^*}(\kappa_n |J|_0). \quad (\text{B.6})$$

■

B.2. The computational error of the k-step lasso. The main result in this subsection is the following proposition.

Proposition B.2. (i) $\|\tilde{\gamma}_y^* - \tilde{\gamma}_{y,lasso}^*\|_1 \leq c \|\widehat{\gamma}_y - \tilde{\gamma}_{y,lasso}^*\|_1 + O_{P^*} \left(\frac{a_n}{\kappa_n} + \sqrt{a_n |J|_0} \right)$ for some $c > 0$.
 (ii) $\|\widehat{\gamma}_y - \tilde{\gamma}_y^*\|_1 = O_{P^*} \left(\kappa_n |J|_0 + \frac{a_n}{\kappa_n} + \sqrt{a_n |J|_0} \right)$,
 (iii) $\frac{1}{nT} \sum_{t=1}^T \|\widehat{U}_t^* (\widehat{\gamma}_y - \tilde{\gamma}_y^*)\|_2^2 = O_{P^*}(\kappa_n^2 |J|_0 + a_n)$.

Proof. (i) We apply Lemma B.1 below. Note that condition (B.7) in this lemma is satisfied under Assumption 4.1 with $b_n = O_{P^*}(a_n)$. Hence applying Lemma B.1 with $\gamma = \tilde{\gamma}_y^*$ immediately implies the result.

(ii) The conclusion follows immediately from part (i) and Proposition B.1.

(iii) By equation (B.8) given below in the proof of Lemma B.1 with $b_n = O_{P^*}(a_n)$,

$$\frac{2}{nT} \sum_{it} \|\widehat{U}_{it}^* (\tilde{\gamma}_y^* - \tilde{\gamma}_{y,lasso}^*)\|_2^2 = O_{P^*}(a_n).$$

Hence by (B.6), $\frac{1}{nT} \sum_{t=1}^T \|\widehat{U}_t^* (\widehat{\gamma}_y - \tilde{\gamma}_y^*)\|_2^2 \leq O_{P^*}(\kappa_n^2 |J|_0 + a_n)$. ■

Lemma B.1. For each γ , suppose for some b_n (either stochastic or deterministic),

$$\mathcal{L}^*(\gamma) + \kappa_n \|\widehat{\Psi} \gamma\|_1 \leq \mathcal{L}^*(\tilde{\gamma}_{lasso}^*) + \kappa_n \|\widehat{\Psi} \tilde{\gamma}_{lasso}^*\|_1 + b_n, \quad (\text{B.7})$$

then

$$\begin{aligned} \|\gamma - \tilde{\gamma}_{lasso}^*\|_1 &\leq C \|\widehat{\gamma}_y - \tilde{\gamma}_{y,lasso}^*\|_{\widehat{J}} + \frac{b_n}{\kappa_n} + O_{P^*}(\sqrt{b_n |J|_0}), \\ \|\gamma - \tilde{\gamma}_{lasso}^*\|_2 &\leq b_n^{1/2} + o_{P^*} \left(|J|_0^{-1/2} \right) \left(C \|\widehat{\gamma}_y - \tilde{\gamma}_{y,lasso}^*\|_{\widehat{J}} + \frac{b_n}{\kappa_n} \right). \end{aligned}$$

Proof. We prove for $\mathcal{L}^* = \mathcal{L}_y^*$. The case with $\mathcal{L}^* = \mathcal{L}_d^*$ follows by the same argument.

Step 1: Show $\|\Delta\|_2^2 \leq O_{P^*}(b_n) + \|\Delta\|_1^2 O_{P^*}(|J|_0^{-1})$. Here $\Delta = \gamma - \tilde{\gamma}_{y,lasso}^*$.

Since $\mathcal{L}_y^*(\gamma)$ is quadratic, for any γ_1, γ_2 ,

$$\mathcal{L}_y^*(\gamma_1) - \mathcal{L}_y^*(\gamma_2) = (\gamma_1 - \gamma_2)' \nabla \mathcal{L}_y^*(\gamma_2) + (\gamma_1 - \gamma_2)' \nabla^2 \mathcal{L}_y^*(\gamma_2) (\gamma_1 - \gamma_2),$$

where

$$\begin{aligned} \nabla \mathcal{L}_y^*(\gamma_2) &= -\frac{2}{nT} \sum_{t=1}^T \sum_{i=1}^n \widehat{U}_{it}^* (\tilde{y}_{it}^* - \widehat{\delta}_{yt}^{*'} \widehat{f}_i^* - \widehat{U}_{it}^{*'} \gamma_2), \\ \nabla^2 \mathcal{L}_y^*(\gamma_2) &= \frac{2}{nT} \sum_{t=1}^T \sum_{i=1}^n \widehat{U}_{it}^* \widehat{U}_{it}^{*'} . \end{aligned}$$

Now let $\gamma_1 = \gamma$, and $\gamma_2 = \tilde{\gamma}_{y,lasso}^*$. Condition (B.7) then implies

$$\begin{aligned} \Delta' \frac{2}{nT} \sum_{t=1}^T \sum_{i=1}^n \widehat{U}_{it}^* \widehat{U}_{it}^{*'} \Delta &\leq b_n + \kappa_n \|\widehat{\Psi}^y \tilde{\gamma}_{y,lasso}^*\|_1 - \kappa_n \|\widehat{\Psi}^y \gamma\|_1 - \Delta' \nabla \mathcal{L}_y^*(\tilde{\gamma}_{y,lasso}^*) \\ &\leq b_n \end{aligned} \tag{B.8}$$

where, to establish the last inequality, we used $\kappa_n \|\widehat{\Psi}^y \tilde{\gamma}_{y,lasso}^*\|_1 - \kappa_n \|\widehat{\Psi}^y \gamma\|_1 - \Delta' \nabla \mathcal{L}_y^*(\tilde{\gamma}_{y,lasso}^*) \leq 0$ which follows due to the first order condition of (B.1) and the convexity of $\|\cdot\|_1$. (See the proof of Lemma 11 of Agarwal et al. (2012).)

We now establish a lower bound for the left hand side of (B.8).

$$\begin{aligned} \Delta' \frac{2}{nT} \sum_{t=1}^T \sum_{i=1}^n \widehat{U}_{it}^* \widehat{U}_{it}^{*'} \Delta &= \frac{2}{nT} \sum_{t=1}^T \|\widehat{U}_t^* \Delta\|_2^2 \\ &\geq^{(a)} \frac{2}{nT} \sum_{t=1}^T \|\tilde{U}_t \Delta\|_2^2 - \|\Delta\|_1^2 O_{P^*}(|J|_0^{-1}) \\ &\geq^{(b)} c \|\Delta\|_2^2 - \|\Delta\|_1^2 O_{P^*}(|J|_0^{-1}) \end{aligned}$$

where (a) follows from (H.16) and (b) follows from Assumption 4.2. Substituting this lower bound in for the left-hand-side of (B.8) then yields

$$\|\Delta\|_2^2 \leq b_n + \|\Delta\|_1^2 O_{P^*}(|J|_0^{-1}). \tag{B.9}$$

Step 2: Show $\|\Delta\|_1 \leq O_{P^*}(b_n/\kappa_n) + \|(\widehat{\gamma}_y - \tilde{\gamma}_{y,lasso}^*)_{\widehat{J}}\|_1 + \|\Delta\|_2 O_P(\sqrt{|J|_0})$.

We re-visit the proof of Proposition B.1. Note that (B.5) implies, for some $c > 0$,

$$\kappa_n \|(\widehat{\gamma}_y - \tilde{\gamma}_{y,lasso}^*)_{\widehat{J}^c}\|_1 \leq \kappa_n \|(\widehat{\gamma}_y - \tilde{\gamma}_{y,lasso}^*)_{\widehat{J}}\|_1 c.$$

The same argument also applies using γ in place of $\tilde{\gamma}_{y,lasso}^*$ due to Condition (B.7), yielding

$$\kappa_n \|(\widehat{\gamma}_y - \gamma)_{\widehat{J}^c}\|_1 \leq \kappa_n \|(\widehat{\gamma}_y - \gamma)_{\widehat{J}}\|_1 c + b_n.$$

Adding these two inequalities and using the triangle inequality, we have

$$\|(\Delta)_{\widehat{J}^c}\|_1 \leq \|(\widehat{\gamma}_y - \tilde{\gamma}_{y,lasso}^*)_{\widehat{J}^c}\|_1 + \|(\widehat{\gamma}_y - \gamma)_{\widehat{J}^c}\|_1$$

$$\begin{aligned}
 &\leq \|(\widehat{\gamma}_y - \widetilde{\gamma}_{y,lasso}^*)_{\widehat{J}}\|_1 c + \|(\widehat{\gamma}_y - \gamma)_{\widehat{J}}\|_1 c + \frac{b_n}{\kappa_n} \\
 &\leq 2c \|(\widehat{\gamma}_y - \widetilde{\gamma}_{y,lasso}^*)_{\widehat{J}}\|_1 + \|(\Delta)_{\widehat{J}}\|_1 c + \frac{b_n}{\kappa_n} \\
 &\leq 2c \|(\widehat{\gamma}_y - \widetilde{\gamma}_{y,lasso}^*)_{\widehat{J}}\|_1 + \|(\Delta)_{\widehat{J}}\|_2 c \sqrt{|\widehat{J}|_0} + \frac{b_n}{\kappa_n}.
 \end{aligned}$$

We then obtain

$$\begin{aligned}
 \|\Delta\|_1 &\leq \|(\Delta)_{\widehat{J}^c}\|_1 + \|(\Delta)_{\widehat{J}}\|_1 \\
 &\leq 2c \|(\widehat{\gamma}_y - \widetilde{\gamma}_{y,lasso}^*)_{\widehat{J}}\|_1 + \|\Delta\|_2 O_{P^*}(\sqrt{|J|_0}) + \frac{b_n}{\kappa_n}.
 \end{aligned} \tag{B.10}$$

Step 3: Complete the proof. Substituting (B.10) in for the right-hand-side of (B.9) gives

$$\|\Delta\|_2^2 \leq b_n + o_{P^*}(|J|_0^{-1}) \left(C \|(\widehat{\gamma}_y - \widetilde{\gamma}_{y,lasso}^*)_{\widehat{J}}\|_1 + \frac{b_n}{\kappa_n} \right)^2 + \|\Delta\|_2^2 o_{P^*}(1),$$

yielding $\|\Delta\|_2^2 \leq b_n + o_{P^*}(|J|_0^{-1}) \left(C \|(\widehat{\gamma}_y - \widetilde{\gamma}_{y,lasso}^*)_{\widehat{J}}\|_1 + \frac{b_n}{\kappa_n} \right)^2$, and thus

$$\|\Delta\|_2 \leq b_n^{1/2} + o_{P^*}(|J|_0^{-1/2}) \left(C \|(\widehat{\gamma}_y - \widetilde{\gamma}_{y,lasso}^*)_{\widehat{J}}\|_1 + \frac{b_n}{\kappa_n} \right).$$

Substituting this bound back in for $\|\Delta\|_2$ in (B.10) then yields

$$\|\Delta\|_1 \leq C \|(\widehat{\gamma}_y - \widetilde{\gamma}_{y,lasso}^*)_{\widehat{J}}\|_1 + \frac{b_n}{\kappa_n} + O_{P^*}(\sqrt{b_n |J|_0}). \tag{B.11}$$

■

B.3. Verifying Assumption 4.1. We now prove Proposition 4.1, which states that the shooting method of Fu (1998) satisfies Assumption 4.1.

We make use of the following lemma in proving Lemma B.3.

Lemma B.2. *Recall that $\widehat{\gamma}$ denotes the post-lasso estimator using the original data and γ_{lasso}^* denotes the lasso estimator using the bootstrap data. We have that*

$$0 \leq \mathcal{L}^*(\widehat{\gamma}) + \kappa_n \|\widehat{\Psi}\widehat{\gamma}\|_1 - (\mathcal{L}^*(\widehat{\gamma}_{lasso}^*) + \kappa_n \|\widehat{\Psi}\widehat{\gamma}_{lasso}^*\|_1) = O_{P^*}(\kappa_n^2 |J|_0). \tag{B.12}$$

Proof. The first inequality follows from the definition of $\widehat{\gamma}_{lasso}^*$.

We now show the equality. Note that for each γ ,

$$\mathcal{L}^*(\gamma) = \frac{1}{nT} \sum_{t=1}^T \left(\|\widehat{U}_t^*(\widehat{\gamma} - \gamma)\|_2^2 + \|M_t^* + \widetilde{e}_t^*\|_2^2 + 2(\widetilde{e}_t^{*'} + M_t^{*'}) \widehat{U}_t^*(\widehat{\gamma} - \gamma) \right) + \kappa_n \|\widehat{\Psi}^y \gamma\|_1$$

where M_t^* and \tilde{e}_t^* are defined in the proof of Proposition B.1. Hence by Proposition B.1 and Lemma H.4,

$$\begin{aligned}
& \mathcal{L}^*(\hat{\gamma}) + \kappa_n \|\widehat{\Psi}\hat{\gamma}\|_1 - (\mathcal{L}^*(\hat{\gamma}_{lasso}^*) + \kappa_n \|\widehat{\Psi}\hat{\gamma}_{lasso}^*\|_1) \\
&= \kappa_n \|\widehat{\Psi}^y \hat{\gamma}\|_1 - \kappa_n \|\widehat{\Psi}\hat{\gamma}_{lasso}^*\|_1 \\
&\quad - \frac{1}{nT} \sum_{t=1}^T \left(\|\widehat{U}_t^*(\hat{\gamma} - \hat{\gamma}_{lasso}^*)\|_2^2 + 2(\tilde{e}_t^{*'} + M_t^{*'}) \widehat{U}_t^*(\hat{\gamma} - \hat{\gamma}_{lasso}^*) \right) \\
&\leq \kappa_n \|\widehat{\Psi}^y(\hat{\gamma} - \hat{\gamma}_{lasso}^*)\|_1 + \left\| \frac{2}{nT} \sum_{t=1}^T (\tilde{e}_t^{*'} + M_t^{*'}) \widehat{U}_t^* \right\|_\infty \|\hat{\gamma} - \hat{\gamma}_{lasso}^*\|_1 \\
&\leq O_{P^*}(\kappa_n) \|\hat{\gamma} - \hat{\gamma}_{lasso}^*\|_1 = O_{P^*}(\kappa_n^2 |J|_0).
\end{aligned}$$

■

Lemma B.3. For the shooting method, (i) $\mathcal{L}^*(\gamma_l) + \kappa_n \|\widehat{\Psi}\gamma_l\|_1 \leq \mathcal{L}^*(\gamma_{l-1}) + \kappa_n \|\widehat{\Psi}\gamma_{l-1}\|_1$.

(ii) $\mathcal{L}^*(\tilde{\gamma}^*) + \kappa_n \|\widehat{\Psi}\tilde{\gamma}^*\|_1 \leq \mathcal{L}^*(\hat{\gamma}_{lasso}^*) + \kappa_n \|\widehat{\Psi}\hat{\gamma}_{lasso}^*\|_1 + O_{P^*}(\kappa_n^2 |J|_0)$.

(iii) $|\widehat{J}^*|_0 = O_{P^*}(|J|_0)$.

Proof. Write $\gamma_l = (\gamma_{l,1}, \dots, \gamma_{l,p})'$, where γ_l can be either $\gamma_{d,l}$ or $\gamma_{y,l}$, to denote the solution after the l^{th} iteration. Note that $\gamma_k = \tilde{\gamma}^*$ is the k -step lasso estimator.

(i) For the shooting method, each $\gamma_{l,m}$ for $m \leq p$ is defined as

$$\gamma_{l,m} = \arg \min_g \frac{1}{nT} \sum_{i,t} (\tilde{y}_{it}^* - \widehat{\delta}_{yt}^{*'} \widehat{f}_i^* - \widehat{U}_{it,m}^{*'} \gamma_{l,m} - \widehat{U}_{it,m}^{*'} \gamma_{l-1,m} - \widehat{U}_{it,m}^{*'} g)^2 + \kappa_n |\widehat{\Psi}_m g|.$$

As is discussed in Section 4.1, after the m^{th} element is updated in the l^{th} iteration, the vector becomes $\gamma_l^{(m)} := (\gamma_{l,m^-}, \gamma_{l,m}, \gamma_{l-1,m^+})'$, where $m^- = \{j : j < m\}$, and $m^+ = \{j : j > m\}$; γ_{l,m^-} represents the subvector of γ_l whose indices are in m^- and γ_{l-1,m^+} represents the subvector of γ_{l-1} whose indices are in m^+ . With this notation, after the $(m-1)^{\text{th}}$ element is updated in the l^{th} iteration, the current solution vector is $\gamma_l^{(m-1)} = (\gamma_{l,(m-1)^-}, \gamma_{l,m-1}, \gamma_{l-1,(m-1)^+})'$. This vector can be rearranged as

$$(\gamma_{l,(m-1)^-}, \gamma_{l,m-1}, \gamma_{l-1,(m-1)^+})' = (\gamma_{l,m^-}, \gamma_{l-1,m}, \gamma_{l-1,m^+})'.$$

It can be seen that the loss function is non-increasing after the m^{th} element is updated:

$$\begin{aligned}
& \mathcal{L}^*(\gamma_l^{(m)}) + \kappa_n \|\widehat{\Psi}\gamma_l^{(m)}\|_1 \\
&= \mathcal{L}^*((\gamma_{l,m^-}, \gamma_{l,m}, \gamma_{l-1,m^+})) \\
&\quad + \kappa_n \|\widehat{\Psi}_{m^-} \gamma_{l,m^-}\|_1 + \kappa_n |\widehat{\Psi}_m \gamma_{l,m}| + \kappa_n \|\widehat{\Psi}_{m^+} \gamma_{l-1,m^+}\|_1 \\
&\leq \mathcal{L}^*((\gamma_{l,m^-}, \gamma_{l-1,m}, \gamma_{l-1,m^+})) \\
&\quad + \kappa_n \|\widehat{\Psi}_{m^-} \gamma_{l,m^-}\|_1 + \kappa_n |\widehat{\Psi}_m \gamma_{l-1,m}| + \kappa_n \|\widehat{\Psi}_{m^+} \gamma_{l-1,m^+}\|_1
\end{aligned}$$

$$\begin{aligned}
 &= \mathcal{L}^*((\gamma_{l,(m-1)^-}, \gamma_{l,m-1}, \gamma_{l-1,(m-1)^+})) \\
 &\quad + \kappa_n \|\widehat{\Psi}_{(m-1)^-} \gamma_{l,(m-1)^-}\|_1 + \kappa_n |\widehat{\Psi}_{m-1} \gamma_{lm-1}| + \kappa_n \|\widehat{\Psi}_{(m-1)^+} \gamma_{l-1,(m-1)^+}\|_1 \\
 &= \mathcal{L}^*(\gamma_l^{(m-1)}) + \kappa_n \|\widehat{\Psi} \gamma_l^{(m-1)}\|_1.
 \end{aligned}$$

Note that $\gamma_l^{(p)} = \gamma_l$. Hence by (B.12) in Lemma B.2,

$$\begin{aligned}
 \mathcal{L}^*(\gamma_l) + \kappa_n \|\widehat{\Psi} \gamma_l\|_1 &\leq \mathcal{L}^*(\gamma_l^{(1)}) + \kappa_n \|\widehat{\Psi} \gamma_l^{(1)}\|_1 \leq \mathcal{L}^*(\gamma_{l-1}^{(p)}) + \kappa_n \|\widehat{\Psi} \gamma_{l-1}^{(p)}\|_1 \\
 &= \mathcal{L}^*(\gamma_{l-1}) + \kappa_n \|\widehat{\Psi} \gamma_{l-1}\|_1.
 \end{aligned}$$

(ii) From (i), $\mathcal{L}^*(\gamma_k) + \kappa_n \|\widehat{\Psi} \gamma_k\|_1 \leq \mathcal{L}^*(\gamma_0) + \kappa_n \|\widehat{\Psi} \gamma_0\|_1 = \mathcal{L}^*(\widehat{\gamma}) + \kappa_n \|\widehat{\Psi} \widehat{\gamma}\|_1$. In addition, by (B.12) in Lemma B.2,

$$\mathcal{L}^*(\widehat{\gamma}) + \kappa_n \|\widehat{\Psi} \widehat{\gamma}\|_1 - (\mathcal{L}^*(\widehat{\gamma}_{lasso}^*) + \kappa_n \|\widehat{\Psi} \widehat{\gamma}_{lasso}^*\|_1) = O_{P^*}(\kappa_n^2 |J|_0)$$

for $\widehat{\gamma}$ and $\widehat{\gamma}_{lasso}^*$ respectively denoting the completed lasso estimator (as opposed to the k-step lasso solution) using the original data and the bootstrap data. Note that $\kappa_n^2 |J|_0 \sqrt{nT} = o(1)$ and $\gamma_k = \tilde{\gamma}^*$, so

$$\begin{aligned}
 \mathcal{L}^*(\tilde{\gamma}^*) + \kappa_n \|\widehat{\Psi} \tilde{\gamma}^*\|_1 &\leq \mathcal{L}^*(\widehat{\gamma}) + \kappa_n \|\widehat{\Psi} \widehat{\gamma}\|_1 \\
 &\leq \mathcal{L}^*(\widehat{\gamma}_{lasso}^*) + \kappa_n \|\widehat{\Psi} \widehat{\gamma}_{lasso}^*\|_1 + o_{P^*}((nT)^{-1/2})
 \end{aligned}$$

which verifies Assumption 4.1(i).

(iii) We now focus on the k -step lasso estimator $\gamma_k = \tilde{\gamma}^*$ and let $\gamma_{k,m}$ denote its m^{th} element. By the KKT condition, if $\gamma_{k,m} \neq 0$, then

$$\begin{aligned}
 -\kappa_n \widehat{\Psi}_m \text{sgn}(\gamma_{k,m}) &= \frac{2}{nT} \sum_{it} \widehat{U}_{it,m}^* (\tilde{y}_{it}^* - \widehat{\delta}_{yt}^{*'} \widehat{f}_i^* - \widehat{U}_{it,m^-}^{*'} \gamma_{k,m^-} - \widehat{U}_{it,m^+}^{*'} \gamma_{k-1,m^+} - \widehat{U}_{it,m}^* \gamma_{k,m}) \\
 &= \frac{2}{nT} \sum_{it} \widehat{U}_{it,m}^* (\tilde{y}_{it}^* - \widehat{\delta}_{yt}^{*'} \widehat{f}_i^* - \widehat{U}_{it}^{*'} \gamma_k^{(m)}) \\
 &= \frac{2}{nT} \sum_{it} \widehat{U}_{it,m}^* (M_{it}^* + \tilde{e}_{it}^* + \widehat{U}_{it}^{*'} (\widehat{\gamma} - \gamma_k^{(m)}))
 \end{aligned} \tag{B.13}$$

where $\gamma_k^{(m)} := (\gamma_{k,m^-}, \gamma_{k,m}, \gamma_{k-1,m^+})'$, and M_{it}^* , \tilde{e}_{it}^* are respectively defined in (B.2)(B.3). Let $\widehat{U}_{it,\widehat{J}^*}^*$ denote the subvector of \widehat{U}_{it}^* , consisting of $\{\widehat{U}_{it,m}^* : \gamma_{k,m} \neq 0, m \leq p\} = \{\widehat{U}_{it,m}^* : m \in \widehat{J}^*\}$. Then the vector form of (B.13) is

$$-\kappa_n \widehat{\Psi}(\widehat{J}^*) \text{sgn}(\gamma_{k,m} : m \in \widehat{J}^*) = \frac{2}{nT} \sum_{it} \widehat{U}_{it,\widehat{J}^*}^* (M_{it}^* + \tilde{e}_{it}^*) + A,$$

where (without loss of generality, we assume $\{m \leq p : \gamma_{k,m} \neq 0\} = \{1, \dots, |\hat{J}^*|_0\}$)

$$\begin{aligned} A &= \begin{pmatrix} \frac{2}{nT} \sum_{it} \hat{U}_{it,1}^* \hat{U}_{it}^{*'} (\hat{\gamma} - \gamma_k^{(1)}) \\ \vdots \\ \frac{2}{nT} \sum_{it} \hat{U}_{it,|\hat{J}^*|_0}^* \hat{U}_{it}^{*'} (\hat{\gamma} - \gamma_k^{(|\hat{J}^*|_0)}) \end{pmatrix} = \begin{pmatrix} \frac{2}{nT} \hat{U}_1^{*'} \hat{U}^* (\hat{\gamma} - \gamma_k^{(1)}) \\ \vdots \\ \frac{2}{nT} \hat{U}_{|\hat{J}^*|_0}^{*'} \hat{U}^* (\hat{\gamma} - \gamma_k^{(|\hat{J}^*|_0)}) \end{pmatrix} \\ &= \begin{pmatrix} (\hat{\gamma} - \gamma_k^{(1)})' \frac{2}{nT} \hat{U}^{*'} & 0 \cdots & 0 \\ 0 & \ddots & \\ 0 & \cdots & (\hat{\gamma} - \gamma_k^{(|\hat{J}^*|_0)})' \frac{2}{nT} \hat{U}^{*'} \end{pmatrix} \begin{pmatrix} \hat{U}_1^* \\ \vdots \\ \hat{U}_{|\hat{J}^*|_0}^* \end{pmatrix} := \frac{2}{nT} B \hat{U}_{\hat{J}^*}. \end{aligned}$$

Therefore

$$\kappa_n \|\hat{\Psi}(\hat{J}^*)\|_2 \leq \max_j \left| \frac{2}{nT} \sum_{it} \hat{U}_{it,j}^* (M_{it}^* + \tilde{e}_{it}^*) \right| \sqrt{|\hat{J}^*|_0} + \left\| \frac{2}{\sqrt{nT}} B \right\| \left\| \frac{1}{\sqrt{nT}} \hat{U}_{\hat{J}^*} \right\|. \quad (\text{B.14})$$

Note that here the norm in both $\left\| \frac{2}{\sqrt{nT}} B \right\|$ and $\left\| \frac{1}{\sqrt{nT}} \hat{U}_{\hat{J}^*} \right\|$ is the operator norm and we have used the inequality $\|B \hat{U}_{\hat{J}^*}\|_2 \leq \|B\| \|\hat{U}_{\hat{J}^*}\|$.

Now by (H.14),

$$\begin{aligned} \frac{1}{nT} \|\hat{U}_{\hat{J}^*}^*\|^2 &\leq \frac{2}{nT} \|\tilde{U}_{\hat{J}^*}^*\|^2 + \frac{2}{nT} \|\hat{U}_{\hat{J}^*}^* - \tilde{U}_{\hat{J}^*}^*\|^2 \\ &\leq 2\phi_{\max}(|\hat{J}^*|_0) + \frac{2}{nT} \sum_{t=1}^T \sum_{i=1}^n \sum_{m \in \hat{J}^*} (\hat{U}_{it,m}^* - \tilde{U}_{it,m}^*)^2 \\ &\leq 2\phi_{\max}(|\hat{J}^*|_0) + O_{P^*}(\Delta_F^{*2} + \frac{\log(pT)}{n}) |\hat{J}^*|_0, \text{ and} \\ \frac{1}{nT} \|B\|^2 &= \max_{m \in \hat{J}^*} \frac{4}{nT} \|\hat{U}^* (\hat{\gamma} - \gamma_k^{(m)})\|^2 \\ &\leq \frac{8}{nT} \|\hat{U}^* (\hat{\gamma} - \tilde{\gamma}_{lasso}^*)\|_2^2 + \max_{m \in \hat{J}^*} \frac{8}{nT} \|\hat{U}^* (\tilde{\gamma}_{lasso}^* - \gamma_k^{(m)})\|_2^2 \\ &\leq O_{P^*}(\kappa_n^2 |J|_0). \end{aligned} \quad (\text{B.15})$$

where $\frac{1}{nT} \|\hat{U}^* (\hat{\gamma} - \tilde{\gamma}_{lasso}^*)\|_2^2 = O_{P^*}(\kappa_n^2 |J|_0)$ follows from (B.6). To show

$$\max_{m \in \hat{J}^*} \frac{8}{nT} \|\hat{U}^* (\tilde{\gamma}_{lasso}^* - \gamma_k^{(m)})\|_2^2 = O_{P^*}(\kappa_n^2 |J|_0),$$

we note that part (i) and (B.12) demonstrate

$$\begin{aligned} \mathcal{L}^*(\gamma_k^{(m)}) + \kappa_n \|\hat{\Psi} \gamma_k^{(m)}\|_1 &\leq \mathcal{L}^*(\gamma_0) + \kappa_n \|\hat{\Psi} \gamma_0\|_1 \\ &= \mathcal{L}^*(\hat{\gamma}) + \kappa_n \|\hat{\Psi} \hat{\gamma}\|_1 \leq \mathcal{L}^*(\tilde{\gamma}_{lasso}^*) + \kappa_n \|\hat{\Psi} \tilde{\gamma}_{lasso}^*\|_1 + O_{P^*}(\kappa_n^2 |J|_0). \end{aligned}$$

Thus, the same argument as used in equation (B.8) leads to

$$\max_{m \in \hat{J}^*} \frac{8}{nT} \|\hat{U}^* (\tilde{\gamma}_{lasso}^* - \gamma_k^{(m)})\|_2^2 \leq O_{P^*}(\kappa_n^2 |J|_0).$$

By Lemma H.4, $\max_j |\frac{2}{nT} \sum_{it} \widehat{U}_{it,j}^* (M_{it}^* + \tilde{e}_{it}^*)| = o_{P^*}(\kappa_n)$ and $\kappa_n \|\widehat{\Psi}(\widehat{J}^*)\|_2 \geq c\kappa_n$. Hence, (B.14) and (B.15) imply

$$\kappa_n^2 |\widehat{J}^*|_0 \leq (C\phi_{\max}(|\widehat{J}^*|_0) + O_{P^*}(\Delta_F^{*2} + \frac{\log(pT)}{n})) |\widehat{J}^*|_0 \kappa_n^2 |J|_0.$$

We thus obtain exactly the same inequality as given (D.3). The rest of the proof then follows from the same argument as used to show Proposition D.1 (ii). We conclude $|\widehat{J}^*|_0 = O_{P^*}(|J|_0)$ which verifies Assumption 4.1(ii). ■

The following lemma is useful to uniformly bound terms in the bootstrap sampling space.

Lemma B.4. *Suppose the following conditions hold:*

(i) $\{Z_{ijm}\}$ is a sequence of random variables in the original sampling space, satisfying

$$\max_{m \leq p, i \leq n} \frac{1}{n} \sum_{j=1}^n Z_{ijm}^2 = O_P(a_n^2)$$

for some deterministic sequence $a_n > 0$.

(ii) $\{X_i^*, Y_i^*\}_{i \leq n}$ is an i.i.d. sequence in the bootstrap sampling space such that $\{X_i^*\}$ is independent of $\{Y_i^*\}$, $EX_i^* = EY_i^* = 0$, and $\text{Var}^*(X_i) < C$ and $\text{Var}^*(Y_i) < C$ for a constant $C > 0$ where C is non-random in both the original and bootstrap sampling space.

(iii) Both X_i^* and Y_i^* are sub-exponential random variables satisfying Assumption 3.1 (iv).

Then for any $\varepsilon_1, \varepsilon_2 > 0$, there is a $C_{\varepsilon_1, \varepsilon_2} > 0$ such that

$$P \left(P^* \left(\max_{m \leq p} \left| \frac{1}{n^2} \sum_{i,j \leq n} X_i^* Y_j^* Z_{ijm} \right| > \frac{2a_n C \sqrt{C_{\varepsilon_1, \varepsilon_2} \log p \log(pn)}}{n} \right) > \varepsilon_1 \right) < \varepsilon_2.$$

Thus $\max_{m \leq p} |\frac{1}{n^2} \sum_{i,j \leq n} X_i^* Y_j^* Z_{ijm}| = O_{P^*} \left(\frac{a_n \sqrt{\log p \log(np)}}{n} \right)$.

Proof. By condition (i), for any $\delta > 0$, there is $C_\delta > 0$ such that with probability at least $1 - \delta$ the event $A_\delta := \{\max_{m \leq p, i \leq n} \frac{1}{n} \sum_{j=1}^n Z_{ijm}^2 < a_n^2 C_\delta\}$ holds.

Let $V^* = \max_{mi} |\frac{1}{n} \sum_j Y_j^* Z_{ijm}|$. Define $W_{im}^* = X_i^* \frac{1}{n} \sum_j Y_j^* Z_{ijm}$ and $Y^* = \{Y_i^*\}_{i \leq n}$. Since $\{X_i^*\}$ and $\{Y_i^*\}$ are independent, then on the event A_δ ,

$$\begin{aligned} \max_{m,i} \frac{1}{n} \sum_j \text{Var}^*(Y_j^* Z_{ijm}) &= \max_{m,i} \frac{1}{n} \sum_j Z_{ijm}^2 \text{Var}(Y_j^*) \\ &< a_n^2 C C_\delta \\ \max_m \frac{1}{n} \sum_i \text{Var}^*(W_{im}^* | Y^*) &= \max_m \frac{1}{n} \sum_i \left(\frac{1}{n} \sum_j Y_j^* Z_{ijm} \right)^2 \text{Var}^*(X_i^*) \\ &\leq CV^{*2}. \end{aligned} \tag{B.16}$$

In the bootstrap sampling space (BSS), $\{Y_j^* Z_{ijm}\}_{j \leq n}$ is independent across j and $E^* Y_j^* Z_{ijm} = 0$. By the Bernstein inequality, for $y = (2a_n^2 C C_\delta \log(pn)/n)^{1/2}$,

$$\begin{aligned} P^*(V^* > y) 1\{A_\delta\} &\leq pn \max_{m,i} P^* \left(\left| \frac{1}{n} \sum_j Y_j^* Z_{ijm} \right| > y \right) 1\{A_\delta\} \\ &\leq \exp \left(\log(pn) - \frac{ny^2}{\max_{m,i} \frac{1}{n} \sum_j \text{Var}^*(Y_j^* Z_{ijm})} \right) 1\{A_\delta\} \\ &\leq \exp \left(\log(pn) - \frac{ny^2}{a_n^2 C C_\delta} \right) = (pn)^{-1}. \end{aligned} \quad (\text{B.17})$$

In the BSS, $\{W_{im}^*\}_{i \leq n}$ is independent across i conditional on Y^* . By (B.16) and the Bernstein inequality, for $x = y \sqrt{\frac{2C \log p}{n}} = \frac{2a_n C \sqrt{C_\delta \log p \log(pn)}}{n}$,

$$\begin{aligned} P^* \left(\max_{m \leq p} \left| \frac{1}{n} \sum_i W_{im}^* \right| > x | Y^* \right) 1\{V^* < y\} \\ &\leq p \max_m P^* \left(\left| \frac{1}{n} \sum_i W_{im}^* \right| > x | Y^* \right) 1\{V^* < y\} \\ &\leq \exp \left(\log p - \frac{nx^2}{\max_m \frac{1}{n} \sum_i \text{Var}^*(W_{im}^* | Y^*)} \right) 1\{V^* < y\} \\ &\leq \exp \left(\log p - \frac{nx^2}{C V^{*2}} \right) 1\{V^* < y\} \\ &\leq \exp \left(\log p - \frac{nx^2}{C y^2} \right) = p^{-1}. \end{aligned} \quad (\text{B.18})$$

Let E_{Y^*} denote the expectation operator with respect to the marginal distribution of Y^* in the bootstrap sampling space; i.e., E_{Y^*} is the conditional distribution of Y^* given the original data. By the law of iterated expectations, $E_{Y^*} [P^*(\cdot | Y^*)] = P^*(\cdot)$. We then have

$$\begin{aligned} &P^* \left(\max_{m \leq p} \left| \frac{1}{n^2} \sum_{i,j \leq n} X_i^* Y_j^* Z_{ijm} \right| > x \right) \\ &\leq P^* \left(\max_{m \leq p} \left| \frac{1}{n^2} \sum_{i,j \leq n} X_i^* Y_j^* Z_{ijm} \right| > x \right) 1\{A_\delta\} + 1\{A_\delta^c\} \\ &= P^* \left(\max_{m \leq p} \left| \frac{1}{n} \sum_i W_{im}^* \right| > x \right) 1\{A_\delta\} + 1\{A_\delta^c\} \\ &= E_{Y^*} P^* \left(\max_{m \leq p} \left| \frac{1}{n} \sum_i W_{im}^* \right| > x | Y^* \right) 1\{A_\delta\} + 1\{A_\delta^c\} \end{aligned}$$

$$\begin{aligned}
 &\leq^{(a)} E_{Y^*} P^* \left(\max_{m \leq p} \left| \frac{1}{n} \sum_i W_{im}^* \right| > x \mid Y^* \right) 1\{V^* < y\} 1\{A_\delta\} + E_{Y^*} 1\{V^* \geq y\} 1\{A_\delta\} + 1\{A_\delta^c\} \\
 &\leq^{(b)} p^{-1} + P^*(V^* \geq y) 1\{A_\delta\} + 1\{A_\delta^c\} \\
 &\leq^{(c)} p^{-1} + (pn)^{-1} + 1\{A_\delta^c\},
 \end{aligned}$$

where we used $P^*(\cdot \mid Y^*) \leq P^*(\cdot \mid Y^*) 1\{V^* < y\} + 1\{V^* \geq y\}$ in (a), (B.18) in (b), and (B.17) in (c). Because $P(A_\delta^c) \leq \delta$, taking the expectation with respect to the distribution of the original data on both sides yields

$$EP^* \left(\max_{m \leq p} \left| \frac{1}{n^2} \sum_{i,j \leq n} X_i^* Y_j^* Z_{ijm} \right| > x \right) \leq p^{-1} + (pn)^{-1} + \delta.$$

For any $\varepsilon_1, \varepsilon_2 > 0$, let $\delta = \varepsilon_1 \varepsilon_2 / 2$, and call C_δ in x to be $C_{\varepsilon_1, \varepsilon_2}$. By the Markov Inequality, we then have

$$\begin{aligned}
 P \left(P^* \left(\max_{m \leq p} \left| \frac{1}{n^2} \sum_{i,j \leq n} X_i^* Y_j^* Z_{ijm} \right| > x \right) > \varepsilon_1 \right) &\leq \frac{1}{\varepsilon_1} (p^{-1} + (pn)^{-1} + \delta) \\
 &\leq \frac{\varepsilon_1 \varepsilon_2 / 2 + \delta}{\varepsilon_1} = \varepsilon_2.
 \end{aligned}$$

■

APPENDIX C. VERIFYING CONDITIONS FOR ESTIMATING THE FACTORS

This section verifies Assumptions 3.5 and 4.4 when factors are estimated using PCA.

C.1. Proof of Proposition 5.1 (for \widehat{F} using the original data). (i) By Assumption 5.1, it can be shown that $\|H\| = O_P(1) = \|V^{-1}\|$. In addition, we have the following identity:

$$\widehat{f}_i - H' \tilde{f}_i = V^{-1} \sum_{l=1}^5 A_{il},$$

where

$$\begin{aligned}
A_{i1} &= \frac{1}{pTn} \sum_{j=1}^n \widehat{f}_j \sum_{t=1}^T (\tilde{U}'_{it} \tilde{U}_{jt} - U'_{it} U_{jt}), \\
A_{i2} &= \frac{1}{pTn} \sum_{j=1}^n \widehat{f}_j \sum_{t=1}^T (U'_{it} U_{jt} - EU'_{it} U_{jt}), \\
A_{i3} &= \frac{1}{pTn} \sum_{j=1}^n \widehat{f}_j \sum_{t=1}^T EU'_{it} U_{jt}, \\
A_{i4} &= \frac{1}{pTn} \sum_{j=1}^n \widehat{f}_j \sum_{t=1}^T \tilde{f}'_j \tilde{\Lambda}'_t \tilde{U}_{it}, \\
A_{i5} &= \frac{1}{pTn} \sum_{j=1}^n \widehat{f}_j \sum_{t=1}^T \tilde{f}'_i \tilde{\Lambda}'_t \tilde{U}_{jt}.
\end{aligned} \tag{C.1}$$

Each term can be written in the form $A_{il} = \frac{1}{pTn} \sum_{j=1}^n \widehat{f}_j \sum_{t=1}^T B_{ijt,l}$. By Cauchy-Schwarz,

$$\begin{aligned}
\frac{1}{n} \sum_{i=1}^n \|\widehat{f}_i - H' \tilde{f}_i\|_2^2 &= O_P(1) \sum_{l=1}^5 \frac{1}{n} \sum_{i=1}^n \|A_{il}\|_2^2 \\
&\leq O_P(1) \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \left(\frac{1}{pT} \sum_{t=1}^T B_{ijt,l} \right)^2.
\end{aligned} \tag{C.2}$$

We bound the terms in (C.2) in Lemmas J.1 and J.2 below. Then, applying the bounds in Lemmas J.1 and J.2 and using $T = o(n)$, we have

$$\frac{1}{n} \sum_{i=1}^n \|\widehat{f}_i - H' \tilde{f}_i\|_2^2 = O_P(\Delta_F^2), \quad \Delta_F^2 = \frac{1}{n^2} + \frac{1}{nT^2} + \frac{1}{pT}.$$

It is then straightforward to verify that $|J|_0^2 \Delta_F^2 = o\left(\sqrt{\frac{1}{nT}}\right)$ holds when $|J|_0^4 = o(nT^3)$, $|J|_0^4 n = o(p^2 T)$, $|J|_0^2 \sqrt{\log p} \log(pT) = o(n)$, and $|J|_0^2 T = o(n)$. For example, to show $|J|_0^2 \sqrt{\log p} \log(T) = o(n)$, note that $|J|_0 = o(\sqrt{n}/(\log p))$ and $|J|_0 = o(\sqrt{n}/T)$ implies

$$|J|_0^2 \sqrt{\log p} \log(T) = o(n \log T \sqrt{\log p} / \sqrt{T \log p}) = o(n).$$

(ii) We now verify that we can produce sequences Δ_{eg} so that $\Delta_{eg} = o\left(\frac{1}{\sqrt{nT}}\right)$. First, note that we can set $g_{tm} \in \{\gamma'_d \tilde{\Lambda}_t, \tilde{\lambda}_{tm}, \tilde{\delta}_t\}$ in applying Lemma C.2, each of which yields $\omega_n = O(|J|_0^2)$ for ω_n defined in Lemma C.2. It then follows from Lemma C.2 that we can take $\Delta_{eg} = \left(\frac{1}{\sqrt{npT}} + \frac{1}{n}\right) |J|_0$ so that $\Delta_{eg} = o\left(\sqrt{\frac{1}{nT}}\right)$, given $T|J|_0^2 = o(n)$ and $|J|_0^2 = o(p)$. Note that $|J|_0^2 = o(p)$ is implied by the assumption that $|J|_0^4 n = o(p^2 T)$.

(iii) By Lemma J.5, we take $\Delta_{fum} = \frac{1}{n} + \frac{1}{T\sqrt{n}} + \frac{1}{T\sqrt{p}} + \sqrt{\frac{\log(pT)}{npT}}$, and $\Delta_{fe}^2 = \frac{1}{n^2} + \frac{1}{Tpn}$. We then have

$$\max_{s \leq T, m \leq p} \left| \frac{1}{n} \sum_{i=1}^n (H' \tilde{f}_i - \hat{f}_i) \tilde{U}_{is,m} \right| = O_P(\Delta_{fum})$$

and

$$\frac{1}{T} \sum_{t=1}^T \left\| \frac{1}{n} \sum_{i=1}^n \tilde{e}_{it} (\hat{f}_i - H' \tilde{f}_i) \right\|_2^2 = O_P(\Delta_{fe}^2).$$

Then, it is straightforward to check $\Delta_{fum}^2 = o\left(\frac{\log p}{T|J|_0^2 \log(pT)}\right)$ and $\Delta_{fe}^2 = o\left(\frac{\log p}{T \log(pT)}\right)$.

(iv) By Lemma C.3, we can define $\Delta_{ud} = \frac{1}{\sqrt{n}} \sqrt{\frac{\log(pT)}{nT}} + \sqrt{\frac{\log p}{nT}} \left(\frac{1}{\sqrt{pT}} + \frac{1}{T\sqrt{n}}\right) + \frac{1}{pT}$. Given $|J|_0^4 n = o(p^2 T)$ and $|J|_0^4 = o(nT^3)$, it is straightforward to verify that $\Delta_{ud} = o\left(\sqrt{\frac{\log p}{nT}}\right)$ and $|J|_0^2 \sqrt{\log p} \Delta_{ud} = o(1)$. This result follows by verifying $\sqrt{\frac{\log p}{nT}} \frac{1}{\sqrt{pT}} |J|_0^2 \sqrt{\log p} = o(1)$ which can be shown by noting that $|J|_0^4 n = o(p^2 T)$ implies $|J_0|^2 = o(p\sqrt{T}/n)$ and that $|J|_0^2 \log(p) = o(n)$. Thus, because $\log^2 p = o(n)$,

$$\left(\sqrt{\frac{\log p}{nT}} \frac{|J|_0^2 \sqrt{\log p}}{\sqrt{pT}} \right)^2 = \frac{\log^2 p |J|_0^4}{npT^2} = o\left(\frac{\log(p)pn\sqrt{T}}{npT^2\sqrt{n}}\right) = o\left(\frac{\log(p)\sqrt{T}}{T^2\sqrt{n}}\right) = o(1).$$

(v) First, by Lemma C.1, we can take $\Delta_{\max} = \frac{1}{\sqrt{n}} + \sqrt{\frac{\log n}{Tp}}$. Also $\Delta_F^2 = \frac{1}{n^2} + \frac{1}{nT^2} + \frac{1}{pT}$. This implies $\Delta_F^2 |J|_0^2 + \frac{|J|_0}{n} = O\left(\frac{|J|_0}{n} + \frac{|J|_0^2}{nT^2} + \frac{|J|_0^2}{pT}\right)$. In addition, $\kappa_n^2 |J|_0 \sqrt{nT} = o(1)$ and $\|R_y\|_1^2 = o\left(\frac{\log p}{nT}\right)$ imply $\lambda_n^2 |J|_0 + \|R_y\|_1^2 = O\left(\frac{1}{\sqrt{nT}}\right)$. Hence with the conditions $|J|_0^4 = o(nT^3)$, $|J|_0^4 n = o(p^2 T)$, we have

$$\lambda_n^2 |J|_0 + \|R_y\|_1^2 + \Delta_F^2 |J|_0^2 + \frac{|J|_0}{n} = O\left(\frac{1}{\sqrt{nT}}\right).$$

Thus, in order to verify $\Delta_{\max}^2 |J|_0^2 T (\lambda_n^2 |J|_0 + \|R_y\|_1^2 + \Delta_F^2 |J|_0^2 + \frac{|J|_0}{n}) = o(1)$, it suffices to verify

$$\left(\frac{1}{n} + \frac{\log n}{Tp}\right) |J|_0^2 T \frac{1}{\sqrt{nT}} = o(1),$$

which holds given the conditions $|J|_0^2 T = o(n)$ and $|J|_0^2 = o(p)$. Note that $|J|_0^2 = o(p)$ is implied by $|J|_0^4 n = o(p^2 T)$. ■

Lemma C.1. $\max_i \|\hat{f}_i - H' \tilde{f}_i\|_2 = O_P\left(\frac{1}{\sqrt{n}} + \sqrt{\frac{\log n}{Tp}}\right) := O_P(\Delta_{\max}^2)$.

Proof. We first bound $\max_i \frac{1}{pT} \sum_{tm} U_{it,m}^2$ and $\max_i \left\| \frac{1}{pT} \sum_{t=1}^T \tilde{\Lambda}'_t U_{it} \right\|_2$. Since

$$\max_i \text{Var} \left(\frac{1}{p} \sum_m U_{it,m}^2 \right) = O(p^{-1}),$$

we have

$$\begin{aligned} \max_i \frac{1}{pT} \sum_{tm} U_{it,m}^2 &\leq \max_i \left| \frac{1}{T} \sum_t \left(\frac{1}{p} \sum_m U_{it,m}^2 - E \frac{1}{p} \sum_m U_{it,m}^2 \right) \right| + \max_{itm} EU_{it,m}^2 \\ &= O_P \left(\sqrt{\frac{\log n}{pT}} + 1 \right) \end{aligned}$$

by the Bernstein inequality for strong mixing sequences.

For the k^{th} row $\tilde{\lambda}'_{tk}$ of $\tilde{\Lambda}'_t$,

$$\begin{aligned} \max_{itk} \text{Var} \left(\frac{1}{p} \tilde{\lambda}'_{tk} U_{it} \right) &= \frac{1}{p^2} \tilde{\lambda}_{tk,m} \tilde{\lambda}_{tk,l} \sum_{m,l} EU_{it,m} U_{it,l} \\ &\leq O \left(\frac{1}{p} \right) \max_{itm} \sum_l |EU_{it,m} U_{it,l}| \\ &= O \left(\frac{1}{p} \right). \end{aligned}$$

Employing the Bernstein inequality for weakly dependent data (e.g., Merlevède et al. (2011)), we then have

$$\begin{aligned} \max_i \left\| \frac{1}{pT} \sum_{t=1}^T \tilde{\Lambda}'_t U_{it} \right\|_2 &\leq O_P \left(\sqrt{\frac{\log n}{T}} \right) \sqrt{\max_{i \leq n, k \leq K, t \leq T} \text{Var} \left(\frac{1}{p} \tilde{\lambda}'_{tk} U_{it} \right)} \\ &= O_P \left(\sqrt{\frac{\log n}{Tp}} \right). \end{aligned} \tag{C.3}$$

Now, $\max_i \|\hat{f}_i - H' \tilde{f}_i\|_2 \leq \sum_{l=1}^3 G_l$ where each G_l is defined and bounded below. Specifically,

$$\begin{aligned} G_1 &= \max_i \left\| \frac{1}{pTn} \sum_{t=1}^T \tilde{U}'_t \sum_{j=1}^n \tilde{U}_{jt} \hat{f}_j \right\|_2 \\ &\leq \max_i \left\| \frac{1}{pTn} \sum_{t=1}^T U'_{it} \sum_{j=1}^n \tilde{U}_{jt} \hat{f}_j \right\|_2 + \max_i \left\| \frac{1}{pTn} \sum_{t=1}^T \bar{U}'_t \sum_{j=1}^n \tilde{U}_{jt} \hat{f}_j \right\|_2 \\ &\leq \left(2 \max_i \frac{1}{pT} \sum_{tm} U_{it,m}^2 + 2 \frac{1}{pT} \sum_{tm} \bar{U}_{t,m}^2 \right)^{1/2} \left(\frac{1}{pT} \sum_{tm} \left\| \frac{1}{n} \sum_j \tilde{U}_{jt,m} \hat{f}_j \right\|_2^2 \right)^{1/2} \\ &= O_P \left(\left(\sqrt{\frac{\log n}{pT}} + 1 \right) \left(\frac{1}{T\sqrt{p}} + \frac{1}{\sqrt{n}} \right) \right) \end{aligned}$$

by equation (C.6) given below in Lemma C.2.

$$G_2 = \max_i \left\| \frac{1}{pTn} \sum_{j=1}^n \hat{f}_j \sum_{t=1}^T \tilde{f}'_j \tilde{\Lambda}'_t \tilde{U}_{it} \right\|_2$$

$$\begin{aligned}
 &\leq O_P(1) \max_i \left\| \frac{1}{pT} \sum_{t=1}^T \tilde{\Lambda}'_t U_{it} \right\|_2 + O_P(1) \left\| \frac{1}{pT} \sum_{t=1}^T \tilde{\Lambda}'_t \tilde{U}_{.t} \right\|_2 \\
 &= O_P \left(\sqrt{\frac{\log n}{Tp}} \right)
 \end{aligned}$$

by (C.3). Finally,

$$\begin{aligned}
 G_3 &= \max_i \left\| \frac{1}{pTn} \sum_{t=1}^T \tilde{f}'_t \tilde{\Lambda}'_t \sum_{j=1}^n \tilde{U}_{jt} \hat{f}'_j \right\|_2 \\
 &\leq \max_i \|\tilde{f}_i\|_2 \left\| \frac{1}{pTn} \sum_{t=1}^T \tilde{\Lambda}'_t \sum_{j=1}^n \tilde{U}_{jt} \hat{f}'_j \right\|_F \\
 &\leq O_P(\sqrt{\log n}) \left\| \frac{1}{pTn} \sum_{t=1}^T \tilde{\Lambda}'_t \sum_{j=1}^n \tilde{U}_{jt} \tilde{f}'_j \right\|_F \\
 &\quad + O_P(\sqrt{\log n}) \left\| \frac{1}{pTn} \sum_{t=1}^T \tilde{\Lambda}'_t \sum_{j=1}^n \tilde{U}_{jt} (\hat{f}_j - H' \tilde{f}_j)' \right\|_F \\
 &= O_P \left(\sqrt{\frac{\log n}{pTn}} \right) + O_P(\sqrt{\log n} \Delta_F) \left(\frac{1}{n} \sum_j \left\| \frac{1}{pT} \sum_t \tilde{\Lambda}'_t \tilde{U}_{jt} \right\|_2^2 \right)^{1/2} \\
 &= O_P \left(\sqrt{\frac{\log n}{pTn}} + \Delta_F \sqrt{\frac{\log n}{pT}} \right).
 \end{aligned}$$

Hence, $\max_i \|\hat{f}_i - H' \tilde{f}_i\|_2 = O_P \left(\frac{1}{\sqrt{n}} + \sqrt{\frac{\log n}{Tp}} \right)$. ■

Lemma C.2. *Let $\{z_{it}\}$ be a random sequence with $E(z_{it}|f_t, U_t) = 0$ and $\text{Var}(z_{it}) > 0$. In addition, let $\{g_{tm}\}$ be a deterministic sequence of vectors with a fixed dimension, $m \leq p$. Then for $\tilde{z}_{it} = z_{it} - \bar{z}_i - \bar{z}_{.t} + \bar{z}$, and $\omega_n = \max_{m \leq p} \frac{1}{T} \sum_{t=1}^T \|g_{tm}\|_2^2$,*

$$\max_{m \leq p} \left\| \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T g_{tm} \tilde{z}_{it} (\hat{f}_i - H' \tilde{f}_i)' \right\|_F = O_P \left(\frac{1}{\sqrt{npT}} + \frac{1}{n} \right) \omega_n^{1/2}.$$

Proof. It follows from equation (C.1) that

$$\max_{m \leq p} \left\| \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T g_{tm} \tilde{z}_{it} (\hat{f}_i - H' \tilde{f}_i)' \right\|_F \leq \sum_{l=1}^3 \bar{C}_l O_P(1),$$

where each term \bar{C}_l is defined and bounded in below.

$$\bar{C}_1 = \max_{m \leq p} \left\| \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T \frac{1}{pTn} \sum_{j=1}^n \hat{f}_j \sum_{s=1}^T \tilde{U}'_{is} \tilde{U}_{js} \tilde{z}_{it} g'_{tm} \right\|_F$$

$$\begin{aligned}
&\leq \max_{m \leq p} \left\| \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T \frac{1}{pTn} \sum_{j=1}^n (\hat{f}_j - H' \tilde{f}_j) \sum_{s=1}^T \tilde{U}'_{is} \tilde{U}_{js} \tilde{z}_{it} g'_{tm} \right\|_F \\
&\quad + O_P(1) \max_{m \leq p} \left\| \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T \frac{1}{pTn} \sum_{j=1}^n \tilde{f}_j \sum_{s=1}^T \tilde{U}'_{is} \tilde{U}_{js} \tilde{z}_{it} g'_{tm} \right\|_F \\
&\leq (x_n \omega_n)^{1/2} \left(\left(\frac{1}{npT} \|\tilde{U}\|_F^2 \right)^{1/2} O_P(\Delta_F) + O_P(1) \left(\frac{1}{pT} \sum_s \left\| \frac{1}{n} \sum_{j=1}^n \tilde{f}_j \tilde{U}'_{js} \right\|_F^2 \right)^{1/2} \right) \\
&= (x_n \omega_n)^{1/2} O_P\left(\Delta_F + \frac{1}{\sqrt{n}}\right) \\
&= \omega_n^{1/2} O_P\left(\frac{1}{\sqrt{npT}} + \frac{1}{n}\right).
\end{aligned}$$

where $x_n = \frac{1}{pT^2} \sum_{t=1}^T \sum_{s=1}^T \left\| \frac{1}{n} \sum_{i=1}^n \tilde{z}_{it} \tilde{U}_{is} \right\|_2^2 = O_P\left(\frac{1}{n}\right)$ and $\omega_n = \max_{m \leq p} \frac{1}{T} \sum_{t=1}^T \|g_{tm}\|_2^2$. Next,

$$\begin{aligned}
\bar{C}_2 &= \max_{m \leq p} \left\| \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T \frac{1}{pTn} \sum_{j=1}^n \hat{f}_j \sum_{s=1}^T \tilde{f}'_j \tilde{\Lambda}'_s \tilde{U}_{is} \tilde{z}_{it} g'_{tm} \right\|_F \\
&\leq O_P(\omega_n^{1/2}) \left(\frac{1}{T} \sum_{t=1}^T \left\| \frac{1}{pT} \frac{1}{n} \sum_{i=1}^n \sum_{s=1}^T \tilde{\Lambda}'_s \tilde{U}_{is} \tilde{z}_{it} \right\|_2^2 \right)^{1/2} \\
&= O_P\left(\omega_n^{1/2} \frac{1}{\sqrt{npT}}\right)
\end{aligned}$$

by Lemma J.3). Finally,

$$\begin{aligned}
\bar{C}_3 &= \max_{m \leq p} \left\| \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T \frac{1}{pTn} \sum_{j=1}^n \hat{f}_j \sum_{s=1}^T \tilde{f}'_i \tilde{\Lambda}'_s \tilde{U}_{js} \tilde{z}_{it} g'_{tm} \right\|_F \\
&\leq \max_{m \leq p} \left\| \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T \frac{1}{pTn} \sum_{j=1}^n (\hat{f}_j - H' \tilde{f}_j) \sum_{s=1}^T \tilde{f}'_i \tilde{\Lambda}'_s \tilde{U}_{js} \tilde{z}_{it} g'_{tm} \right\|_F \\
&\quad + O_P(1) \max_{m \leq p} \left\| \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T \frac{1}{pTn} \sum_{j=1}^n \tilde{f}_j \sum_{s=1}^T \tilde{f}'_i \tilde{\Lambda}'_s \tilde{U}_{js} \tilde{z}_{it} g'_{tm} \right\|_F \\
&\leq (\omega_n c_n)^{1/2} \left(O_P(\Delta_F) \left(\frac{1}{n} \sum_j \left\| \frac{1}{pT} \sum_{s=1}^T \tilde{\Lambda}'_s \tilde{U}_{js} \right\|_2^2 \right)^{1/2} + O_P(1) \max_{m \leq p} \left\| \frac{1}{pTn} \sum_{j=1}^n \sum_{s=1}^T \tilde{\Lambda}'_s \tilde{U}_{js} \tilde{f}'_j \right\|_F \right) \\
&= \omega_n^{1/2} O_P\left(\frac{1}{\sqrt{npT}} + \frac{1}{n}\right)
\end{aligned}$$

where $c_n = \frac{1}{T} \sum_{t=1}^T \left\| \frac{1}{n} \sum_{i=1}^n \tilde{f}_i \tilde{z}_{it} \right\|_2^2 = O_P\left(\frac{1}{n}\right)$. Therefore,

$$\max_{m \leq p} \left\| \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T g_{tm} \tilde{z}_{it} (\hat{f}_i - H' \tilde{f}_i)' \right\|_F = O_P\left(\frac{1}{\sqrt{npT}} + \frac{1}{n}\right) \omega_n^{1/2}.$$

■

Lemma C.3. *Let $\{g_{tk} : k \leq p\}$ be a deterministic sequence of vectors of fixed dimension with $\max_{tk} \|g_{tk}\|_2 = O(1)$, and let $\tilde{g}_{tk} = g_{tk} - \frac{1}{T} \sum_t g_{tk}$. Then*

$$\begin{aligned} \max_{m,k \leq p} \left\| \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T (\hat{f}_i - H' \tilde{f}_i) \tilde{U}_{it,m} \tilde{g}'_{tk} \right\|_F &= O_P \left(\frac{1}{\sqrt{n}} \sqrt{\frac{\log(pT)}{nT}} + \sqrt{\frac{\log p}{nT}} \left(\frac{1}{\sqrt{pT}} + \frac{1}{T\sqrt{n}} \right) + \frac{1}{pT} \right) \\ &:= O_P(\Delta_{ud}). \end{aligned}$$

Proof. First, note that $\sum_t \tilde{g}_{tk} = 0$. Hence,

$$\begin{aligned} \max_{m,k \leq p} \left\| \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T (\hat{f}_i - H' \tilde{f}_i) \tilde{U}_{it,m} \tilde{g}'_{tk} \right\|_F \\ \leq \max_{m,k \leq p} \left\| \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T (\hat{f}_i - H' \tilde{f}_i) U_{it,m} \tilde{g}'_{tk} \right\|_F \end{aligned} \quad (\text{C.4})$$

$$+ \left\| \frac{1}{n} \sum_{i=1}^n (\hat{f}_i - H' \tilde{f}_i) \right\|_2 \max_{m,k \leq p} \left\| \frac{1}{T} \sum_{t=1}^T \tilde{U}_{\cdot,t,m} \tilde{g}'_{tk} \right\|_F. \quad (\text{C.5})$$

Term (C.5) is $O_P \left(\Delta_F \sqrt{\frac{\log p}{nT}} \right)$. Term (C.4) is bounded by $\sum_{l=1}^7 C_l$, where each C_l is defined and bounded blow.

First, note that applying Lemma J.5 gives

$$\begin{aligned} \frac{1}{\sqrt{p}} \left(\frac{1}{T} \sum_t \left\| \frac{1}{n} \sum_j \hat{f}_j \tilde{U}'_{jt} \right\|_F^2 \right)^{1/2} &\leq \frac{2}{\sqrt{p}} \left(\frac{1}{T} \sum_t \left\| \frac{1}{n} \sum_j (\hat{f}_j - H' \tilde{f}_j) \tilde{U}'_{jt} \right\|_F^2 \right)^{1/2} \\ &\quad + \frac{2}{\sqrt{p}} \left(\frac{1}{T} \sum_t \left\| \frac{1}{n} \sum_j H' \tilde{f}_j \tilde{U}'_{jt} \right\|_F^2 \right)^{1/2} \\ &= O_P \left(\frac{1}{n} + \frac{1}{T\sqrt{n}} + \frac{1}{T\sqrt{p}} + \sqrt{\frac{\log(pT)}{npT}} + \frac{1}{\sqrt{n}} \right) \\ &= O_P \left(\frac{1}{T\sqrt{p}} + \frac{1}{\sqrt{n}} \right) \text{ and} \\ \max_{tl} \left\| \frac{1}{n} \sum_{j=1}^n \hat{f}_j \tilde{U}_{jt,l} \right\|_2 &= O_P \left(\sqrt{\frac{\log(pT)}{n}} + \frac{1}{T\sqrt{p}} \right). \end{aligned} \quad (\text{C.6})$$

We then have, up to an $\|V^{-1}\| = O_P(1)$ term,

$$C_1 = \max_{m,k \leq p} \left\| \frac{1}{nT} \sum_{i=1}^n \sum_{s=1}^T \frac{1}{pTn} \sum_{j=1}^n \hat{f}_j \sum_{t=1}^T \tilde{U}'_{jt} (\tilde{U}_{it} - U_{it}) U_{is,m} \tilde{g}'_{sk} \right\|_F$$

$$\begin{aligned}
&= \max_{m,k \leq p} \left\| \frac{1}{nT} \sum_{i=1}^n \sum_{s=1}^T \frac{1}{pTn} \sum_{j=1}^n \hat{f}_j \sum_{t=1}^T \tilde{U}'_{jt} \bar{U}_{\cdot t} U_{is,m} \tilde{g}'_{sk} \right\|_F \quad (\text{because } \frac{1}{T} \sum_t \tilde{U}_{jt} = 0) \\
&\leq \left(\frac{1}{T} \sum_t \left\| \frac{1}{n} \sum_j \hat{f}_j \tilde{U}'_{jt} \right\|_F^2 \right)^{1/2} \left(\frac{1}{T} \sum_t \|\bar{U}_{\cdot t}\|_2^2 \right)^{1/2} \max_{mk} \left\| \frac{1}{nTp} \sum_{i=1}^n \sum_{s=1}^T U_{is,m} \tilde{g}'_{sk} \right\|_2 \\
&= O_P \left(\frac{1}{T\sqrt{p}} + \frac{1}{\sqrt{n}} \right) \frac{1}{\sqrt{n}} \sqrt{\frac{\log p}{nT}} \\
&= O_P \left(\frac{1}{T\sqrt{pn}} + \frac{1}{n} \right) \sqrt{\frac{\log p}{nT}}.
\end{aligned}$$

$$\begin{aligned}
C_2 &= \max_{m,k \leq p} \left\| \frac{1}{nT} \sum_{i=1}^n \sum_{s=1}^T \frac{1}{pTn} \sum_{j=1}^n \hat{f}_j \sum_{t=1}^T \tilde{U}'_{jt} (U_{it} U_{is,m} - EU_{it} U_{is,m}) \tilde{g}'_{sk} \right\|_F \\
&\leq \left(\frac{1}{T} \sum_t \left\| \frac{1}{n} \sum_j \hat{f}_j \tilde{U}'_{jt} \right\|_F^2 \right)^{1/2} \max_{mkt} \left\| \frac{1}{nTp} \sum_{i=1}^n \sum_{s=1}^T (U_{it} U_{is,m} - EU_{it} U_{is,m}) \tilde{g}'_{sk} \right\|_F \quad (\text{C.7}) \\
&= O_P \left(\frac{1}{T\sqrt{p}} + \frac{1}{\sqrt{n}} \right) \left(\sqrt{\frac{\log(pT)}{nT}} \right)
\end{aligned}$$

using Lemma J.4.

$$\begin{aligned}
C_3 &= \max_{m,k \leq p} \left\| \frac{1}{nT} \sum_{i=1}^n \sum_{s=1}^T \frac{1}{pTn} \sum_{j=1}^n \hat{f}_j \sum_{t=1}^T \tilde{U}'_{jt} (EU_{it} U_{is,m}) \tilde{g}'_{sk} \right\|_F \\
&\leq \max_{itm} \sum_{s=1}^T \sum_{l=1}^p |(EU_{it,l} U_{is,m})| \max_{sk} \|\tilde{g}_{sk}\|_2 \frac{1}{pT} \max_{tl} \left\| \frac{1}{n} \sum_{j=1}^n \hat{f}_j \tilde{U}_{jt,l} \right\|_2 \\
&= O_P \left(\frac{1}{pT} \right) O_P \left(\sqrt{\frac{\log(pT)}{n}} + \frac{1}{T\sqrt{p}} \right).
\end{aligned}$$

$$\begin{aligned}
C_4 &= \max_{m,k \leq p} \left\| \frac{1}{nT} \sum_{i=1}^n \sum_{s=1}^T \frac{1}{pTn} \sum_{j=1}^n \hat{f}_j \sum_{t=1}^T \tilde{f}'_j \tilde{\Lambda}'_t (\tilde{U}_{it} - U_{it}) U_{is,m} \tilde{g}'_{sk} \right\|_F \\
&= \max_{m,k \leq p} \left\| \frac{1}{nT} \sum_{i=1}^n \sum_{s=1}^T \frac{1}{pTn} \sum_{j=1}^n \hat{f}_j \sum_{t=1}^T \tilde{f}'_j \tilde{\Lambda}'_t \bar{U}_{\cdot t} U_{is,m} \tilde{g}'_{sk} \right\|_F \quad (\text{because } \sum_t \tilde{\Lambda}_t = 0) \\
&\leq O_P(1) \left\| \frac{1}{pT} \sum_{t=1}^T \tilde{\Lambda}'_t \bar{U}_{\cdot t} \right\|_2 \max_{m,k \leq p} \left\| \frac{1}{nT} \sum_{i=1}^n \sum_{s=1}^T U_{is,m} \tilde{g}'_{sk} \right\|_2
\end{aligned}$$

$$\begin{aligned}
 &= O_P\left(\frac{1}{\sqrt{npT}}\right) \sqrt{\frac{\log p}{nT}}. \\
 C_5 &= \max_{m,k \leq p} \left\| \frac{1}{nT} \sum_{i=1}^n \sum_{s=1}^T \frac{1}{pTn} \sum_{j=1}^n \widehat{f}_j \sum_{t=1}^T \widetilde{f}'_j \widetilde{\Lambda}'_t (U_{it} U_{is,m} - EU_{it} U_{is,m}) \widetilde{g}'_{sk} \right\|_F \\
 &\leq O_P(1) \max_{m,k \leq p} \left\| \frac{1}{nT} \sum_{i=1}^n \sum_{s=1}^T \frac{1}{pT} \sum_{t=1}^T \widetilde{\Lambda}'_t (U_{it} U_{is,m} - EU_{it} U_{is,m}) \widetilde{g}'_{sk} \right\|_F \\
 &= O_P\left(\sqrt{\frac{\log p}{npT^2}}\right)
 \end{aligned} \tag{C.8}$$

using the same proof as that of Lemma J.4 (ii).

$$\begin{aligned}
 C_6 &= \max_{m,k \leq p} \left\| \frac{1}{nT} \sum_{i=1}^n \sum_{s=1}^T \frac{1}{pTn} \sum_{j=1}^n \widehat{f}_j \sum_{t=1}^T \widetilde{f}'_j \widetilde{\Lambda}'_t (EU_{it} U_{is,m}) \widetilde{g}'_{sk} \right\|_F \\
 &\leq O_P(1) \max_{m,k \leq p} \left\| \frac{1}{nT} \sum_{i=1}^n \sum_{s=1}^T \frac{1}{pT} \sum_{t=1}^T \widetilde{\Lambda}'_t (EU_{it} U_{is,m}) \widetilde{g}'_{sk} \right\|_F \\
 &= O_P\left(\frac{1}{pT}\right).
 \end{aligned}$$

$$\begin{aligned}
 C_7 &= \max_{m,k \leq p} \left\| \frac{1}{nT} \sum_{i=1}^n \sum_{s=1}^T \frac{1}{pTn} \sum_{j=1}^n \widehat{f}_j \sum_{t=1}^T \widetilde{f}'_i \widetilde{\Lambda}'_t \widetilde{U}_{jt} U_{is,m} \widetilde{g}'_{sk} \right\|_F \\
 &\leq \left\| \frac{1}{pTn} \sum_{j=1}^n \sum_{t=1}^T \widehat{f}_j \widetilde{U}'_{jt} \widetilde{\Lambda}_t \right\|_2 \max_{m,k \leq p} \left\| \frac{1}{nT} \sum_{i=1}^n \sum_{s=1}^T \widetilde{f}_i U_{is,m} \widetilde{g}'_{sk} \right\|_F \\
 &\leq O_P\left(\sqrt{\frac{\log p}{nT}}\right) \left(\left\| \frac{1}{pTn} \sum_{j=1}^n \sum_{t=1}^T \widetilde{f}_j \widetilde{U}'_{jt} \widetilde{\Lambda}_t \right\|_2 + \left\| \frac{1}{n} \sum_{j=1}^n (\widehat{f}_j - H' \widetilde{f}_j) \widetilde{U}_{jt,m} \right\|_2 \right) \\
 &= O_P\left(\sqrt{\frac{\log p}{nT}}\right) \left(\frac{1}{n} + \frac{1}{T\sqrt{n}} + \frac{1}{T\sqrt{p}} + \sqrt{\frac{\log(pT)}{npT}} \right).
 \end{aligned}$$

Combining the above, we reach

$$\begin{aligned}
 &\max_{m,k \leq p} \left\| \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T (\widehat{f}_i - H' \widetilde{f}_i) \widetilde{U}_{it,m} \widetilde{g}'_{tk} \right\|_F \\
 &= O_P\left(\frac{1}{\sqrt{n}} \sqrt{\frac{\log(pT)}{nT}} + \sqrt{\frac{\log p}{nT}} \left(\frac{1}{\sqrt{pT}} + \frac{1}{T\sqrt{n}} + \Delta_F \right) + \frac{1}{pT}\right)
 \end{aligned}$$

$$= O_P \left(\frac{1}{\sqrt{n}} \sqrt{\frac{\log(pT)}{nT}} + \sqrt{\frac{\log p}{nT}} \left(\frac{1}{\sqrt{pT}} + \frac{1}{T\sqrt{n}} \right) + \frac{1}{pT} \right).$$

■

C.2. Proof of Proposition 5.1 (for \widehat{F}^* using the bootstrap data). (i) Similar to (C.1), it can be proven that there is $\|H^*\| = O_{P^*}(1)$ and $\|V^{*-1}\| = O_{P^*}(1)$ such that

$$\widehat{f}_i^* - H^{*'} \widehat{f}_i = V^{*-1} \sum_{l=1}^4 A_{il}^*,$$

where

$$\begin{aligned} A_{i1}^* &= \frac{1}{pTn} \sum_{j=1}^n \widehat{f}_j^* \sum_{t=1}^T (\tilde{U}_{it}^{*'} \tilde{U}_{jt}^* - E^* \tilde{U}_{it}^{*'} \tilde{U}_{jt}^*), & A_{i2}^* &= \frac{1}{pTn} \sum_{j=1}^n \widehat{f}_j^* \sum_{t=1}^T E^* \tilde{U}_{it}^{*'} \tilde{U}_{jt}^*, \\ A_{i3}^* &= \frac{1}{pTn} \sum_{j=1}^n \widehat{f}_j^* \sum_{t=1}^T \widehat{f}_j^* \widehat{\Lambda}_t' \tilde{U}_{it}^*, & A_{i4}^* &= \frac{1}{pTn} \sum_{j=1}^n \widehat{f}_j^* \sum_{t=1}^T \widehat{f}_i^* \widehat{\Lambda}_t' \tilde{U}_{jt}^*. \end{aligned} \tag{C.9}$$

We first treat $A_{i2}^* - A_{i4}^*$. Because \tilde{U}_{it}^* and \tilde{U}_{jt}^* are independent if $i \neq j$, we have

$$\begin{aligned} \frac{1}{n} \sum_i \|A_{i2}^*\|_2^2 &= \frac{1}{n} \sum_i \left\| \frac{1}{pTn} \sum_{j=1}^n \widehat{f}_j^* \sum_{t=1}^T E^* \tilde{U}_{it}^{*'} \tilde{U}_{jt}^* \right\|_2^2 \\ &= \frac{1}{n} \sum_i \left\| \frac{1}{pTn} \widehat{f}_i^* \sum_{t=1}^T E^* \tilde{U}_{it}^{*'} \tilde{U}_{it}^* \right\|_2^2 \\ &= O_{P^*} \left(\frac{1}{n^2} \right). \end{aligned}$$

By Lemma J.7, $\frac{1}{n} \sum_i \|A_{i3}^*\|_2^2 + \frac{1}{n} \sum_i \|A_{i4}^*\|_2^2 = O_{P^*}(\Delta_F^2)$. Hence

$$\sum_{l=2}^4 \frac{1}{n} \sum_i \|A_{il}^*\|_2^2 = O_{P^*}(\Delta_F^2). \tag{C.10}$$

Now we bound $\frac{1}{n} \sum_{i=1}^n \|A_{i1}^*\|_2^2$. A preliminary rate is provided by Lemma J.6 where we have that

$$\frac{1}{n} \sum_{i=1}^n \|A_{i1}^*\|_2^2 = O_{P^*} \left(\Delta_F^2 + \frac{1}{n} \right).$$

However, this rate is not sharp due to the $O_{P^*}(n^{-1})$ term and can be improved. Specifically, the proof of Lemma J.6 (iii) uses a Cauchy-Schwarz inequality and is not sharp for terms involving $E[U_{it}' U_{jt}]$. To see this intuitively, consider a simple example where we bound

$$\frac{1}{n} \sum_{i=1}^n \left\| \frac{1}{nTp} \sum_{j=1}^n \sum_{t=1}^T f_j E U_{it}' U_{jt} \right\|_2^2.$$

Since U_{it} and U_{jt} are independent when $i \neq j$, this term may be simplified to

$$\frac{1}{n} \sum_{i=1}^n \left\| \frac{1}{nTp} \sum_{t=1}^T f_i EU'_{it} U_{it} \right\|_2^2 = O_P \left(\frac{1}{n^2} \right).$$

In contrast, using the Cauchy-Schwarz inequality gives

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \left\| \frac{1}{nTp} \sum_{j=1}^n \sum_{t=1}^T f_j EU'_{it} U_{jt} \right\|_2^2 &\leq \frac{1}{n} \sum_j \|\widehat{f}_j\|_2^2 \frac{1}{n^2} \sum_{j=1}^n \sum_{i=1}^n \left(\frac{1}{Tp} \sum_{t=1}^T EU'_{it} U_{jt} \right)^2 \\ &= O_P \left(\frac{1}{n} \right). \end{aligned}$$

Lemma J.6 (iii) does provide a useful preliminary rate to build upon. Applying Lemma J.6 (iii) and (C.10), we obtain a preliminary rate

$$\frac{1}{n} \sum_{i=1}^n \|\widehat{f}_i^* - H^{*'} \widehat{f}_i\|_2^2 = O_{P^*} \left(\Delta_F^2 + \frac{1}{n} \right).$$

Our goal is to remove the term $\frac{1}{n}$ through improving the bound for $\frac{1}{n} \sum_{i=1}^n \|A_{i1}^*\|_2^2$. By the triangle inequality,

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \|A_{i1}^*\|_2^2 &\leq \frac{2}{n} \sum_{i=1}^n \left\| \frac{1}{pTn} \sum_{j=1}^n H^{*'} \widehat{f}_j \sum_{t=1}^T (\widetilde{U}_{it}^* \widetilde{U}_{jt}^* - E^* \widetilde{U}_{it}^* \widetilde{U}_{jt}^*) \right\|_2^2 \\ &\quad + \frac{2}{n} \sum_{i=1}^n \left\| \frac{1}{n} \sum_{j=1}^n (\widehat{f}_j^* - H^{*'} \widehat{f}_j) \frac{1}{pT} \sum_{t=1}^T (\widetilde{U}_{it}^* \widetilde{U}_{jt}^* - E^* \widetilde{U}_{it}^* \widetilde{U}_{jt}^*) \right\|_2^2 \\ &\leq^{(a)} O_{P^*}(\Delta_F^2) + O_{P^*}(\Delta_F^2) \frac{1}{n^2} \sum_{ij} \left(\frac{1}{pT} \sum_{t=1}^T (\widetilde{U}_{it}^* \widetilde{U}_{jt}^* - E^* \widetilde{U}_{it}^* \widetilde{U}_{jt}^*) \right)^2 \\ &= O_{P^*}(\Delta_F^2) \end{aligned}$$

where in (a) we used Lemma J.8 and the last equality follows from (J.5). Hence combining with (C.10), we have $\frac{1}{n} \sum_{i=1}^n \|\widehat{f}_i^* - H^{*'} \widehat{f}_i\|_2^2 = O_{P^*}(\Delta_F^2)$. Thus, we have $\Delta_F^* = \Delta_F$.

(ii) We now verify the conditions in Assumption 4.4: $\sqrt{nT}|J|_0^2 \Delta_F^{*2} = o(1)$, $\Delta_{eg}^* = o(\frac{1}{\sqrt{nT}})$, $\Delta_{ud}^* = o(\sqrt{\frac{\log p}{nT}})$, $|J|_0^2 \sqrt{\log p} \Delta_{ud}^* = o(1)$, $\Delta_F^{*2} = o(\frac{\log p}{T \log(pT)})$, and $\Delta_{\max}^2 |J|_0^2 T \Delta_F^{*2} = o(1)$.

Δ_F^* : We have previously proven that $\Delta_F^2 = \frac{1}{n^2} + \frac{1}{nT^2} + \frac{1}{pT}$. In addition, Lemma C.1 gives $\Delta_{\max} = \frac{1}{\sqrt{n}} + \sqrt{\frac{\log n}{Tp}}$. Hence it is straightforward to verify the required conditions involving Δ_F^* , given the assumption that $|J|_0^2 \log n = o(p)$.

Δ_{ud}^* : Δ_{ud}^* is defined in Lemma C.5 which gives $\Delta_{ud}^* = b_n + \Delta_{ud}$ for

$$b_n = \frac{1}{T} \sqrt{\frac{\log p \log(np)}{np}} + \sqrt{\frac{\log n}{npT}} + \frac{\log p}{n\sqrt{T}} + \frac{1}{pT} + \sqrt{\frac{\log p}{npT}} + \frac{\sqrt{\log p}}{n} + \frac{\sqrt{|J|_0}}{nT} + \sqrt{\frac{|J|_0}{npT}}.$$

In the proof of Proposition 5.1, we verified $\Delta_{ud} = o\left(\sqrt{\frac{\log p}{nT}}\right)$ and $|J|_0^2 \sqrt{\log p} \Delta_{ud} = o(1)$. It is also straightforward to verify that $b_n = o\left(\sqrt{\frac{\log p}{nT}}\right)$ and that $|J|_0^2 \sqrt{\log p} b_n = o(1)$ given that $\log n = o(p)$, $|J|_0^4 n = o(p^2 T)$, $|J|_0^2 \log^3 p = o(n)$, and $|J|_0^4 = o(nT^3)$. In particular, we need to verify $|J|_0^5 \log p \left(\frac{1}{n^2 T^2} + \frac{1}{npT}\right) = o(1)$. To verify this condition, we use $|J|_0^4 n = o(p^2 T)$ and $|J|_0^4 = o(nT^3)$ to show

$$\begin{aligned} |J|_0^5 \log p \left(\frac{1}{n^2 T^2} + \frac{1}{npT}\right) &= |J|_0^3 \log p \left(\frac{|J|_0^2}{n^2 T^2} + \frac{|J|_0^2 n^{1/2}}{n^{3/2} pT}\right) \\ &= o(1) |J|_0^3 \log p \left(\frac{n^{1/2} T^{3/2}}{n^2 T^2} + \frac{pT^{1/2}}{n^{3/2} pT}\right) \\ &= o(1) \frac{|J|_0^3 \log p}{n^{3/2} T^{1/2}} \\ &= o(1) \left(\frac{|J|_0 \log p}{n^{1/2}}\right)^3 = o(1). \end{aligned}$$

Δ_{eg}^* : Note that for $\hat{g}_{tm} \in \{\hat{\Lambda}'_t \hat{\gamma}_d, \hat{\Lambda}'_t \hat{\gamma}_y, \hat{\delta}_{dt}, \hat{\delta}_{yt}, \hat{\lambda}_{tm}\}$, we have

$$\omega_n^* = \max_{m \leq p} \frac{1}{T} \sum_{t=1}^T \|\hat{g}_{tm}\|_2^2 = O_P(|J|_0^2).$$

Hence by Lemma C.4, $\Delta_{eg}^{*2} = \left(\frac{1}{n^2} + \frac{\log n}{npT} + \frac{\log n}{n^2 T^2} + \frac{\log^{1/2} n}{n^2 T^{1/2}}\right) |J|_0^2$. Given $|J|_0^2 \log n = o(p)$ and $|J|_0^2 T = o(n)$, it is then straightforward to verify $\Delta_{eg}^{*2} = o\left(\frac{1}{nT}\right)$ which follows by verifying $\frac{|J|_0^2 \log n}{nT} = o(1)$. To see $\frac{|J|_0^2 \log n}{nT} = o(1)$, note that we have, by $|J|_0^{4/3} = o(n^{1/3} T)$,

$$\frac{|J|_0^2 \log n}{nT} = \frac{|J|_0^{2/3} |J|_0^{4/3} \log n}{nT} = o(1) \frac{|J|_0^{2/3} n^{1/3} T \log n}{nT} = o(1) \frac{|J|_0^{2/3} \log n}{n^{2/3}} = o(1).$$

■

Lemma C.4. *In the bootstrap sampling space, let $\tilde{z}_{it}^* = \hat{z}_{it} w_i^Z$ where $\{w_i^Z\}_{i=1}^n$ are i.i.d. with mean zero and bounded variance and independent of $\{w_i^U\}$ and $\hat{z}_{it} = \hat{\eta}_{it}$ or $\hat{z}_{it} = \hat{\epsilon}_{it}$. In addition, let $\{\hat{g}_{tm}\}$ be a deterministic sequence (in the bootstrap sampling space) of vectors with a fixed dimension, $m \leq p$. Then for $\omega_n^* = \max_{m \leq p} \frac{1}{T} \sum_{t=1}^T \|\hat{g}_{tm}\|_2^2$,*

$$\max_{m \leq p} \left\| \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T \hat{g}_{tm} \tilde{z}_{it}^* (\hat{f}_i^* - H^{*'} \hat{f}_i)' \right\|_F^2 = O_{P^*} \left(\frac{1}{n^2} + \frac{\log n}{npT} + \frac{\log n}{n^2 T^2} + \frac{\log^{1/2} n}{n^2 T^{1/2}} \right) \omega_n^*$$

where the term $O_{P^*} \left(\frac{1}{n^2} + \frac{\log n}{npT} + \frac{\log n}{n^2 T^2} + \frac{\log^{1/2} n}{n^2 T^{1/2}} \right) \omega_n^*$ defines Δ_{eg}^{*2} .

Proof. It follows from (C.9) that

$$\max_{m \leq p} \left\| \frac{1}{nT} \sum_{i=1}^n \sum_{s=1}^T \tilde{z}_{is}^* (\hat{f}_i^* - H^{*'} \hat{f}_i) \hat{g}'_{sm} \right\|_F \leq \sum_{l=1}^3 \bar{C}_l O_P(1),$$

where each term \bar{C}_l is defined and bounded below.

First, we have

$$\begin{aligned} \bar{C}_1 &= \max_{m \leq p} \left\| \frac{1}{nT} \sum_{i=1}^n \sum_{s=1}^T \tilde{z}_{is}^* \frac{1}{pTn} \sum_{j=1}^n \hat{f}_j^* \sum_{t=1}^T \tilde{U}_{it}^{*'} \tilde{U}_{jt}^* \hat{g}'_{sm} \right\|_F \\ &\leq \left(\frac{1}{pT} \sum_t \left\| \frac{1}{n} \sum_{j=1}^n \hat{f}_j^* \tilde{U}_{jt}^{*'} \right\|_F^2 \right)^{1/2} \left(\frac{1}{T^2 p} \sum_{s,t} \left\| \frac{1}{n} \sum_{i=1}^n \tilde{z}_{is}^* \tilde{U}_{it}^* \right\|_2^2 \right)^{1/2} \omega_n^{*1/2} \\ &= O_{P^*} \left(\Delta_F + \frac{1}{\sqrt{n}} \right) \frac{1}{\sqrt{n}} \omega_n^{*1/2} \\ &= \omega_n^{*1/2} O_{P^*} \left(\frac{1}{\sqrt{npT}} + \frac{1}{n} \right), \end{aligned}$$

where we used

$$\begin{aligned} \frac{1}{pT} \sum_t \left\| \frac{1}{n} \sum_{j=1}^n \hat{f}_j^* \tilde{U}_{jt}^{*'} \right\|_F^2 &\leq O_{P^*}(\Delta_F^2) + O_{P^*}(1) \frac{1}{pT} \sum_t \left\| \frac{1}{n} \sum_{j=1}^n \hat{f}_j^* \tilde{U}_{jt}^{*'} \right\|_F^2 \\ &= O_{P^*} \left(\Delta_F^2 + \frac{1}{n} \right), \\ \frac{1}{T^2 p} \sum_{s,t} \left\| \frac{1}{n} \sum_{i=1}^n \tilde{z}_{is}^* \tilde{U}_{it}^* \right\|_2^2 &= \frac{1}{T^2 p} \sum_{s,t} \left\| \frac{1}{n} \sum_{i=1}^n \hat{z}_{is} w_i^Z w_i^U \hat{U}_{it} \right\|_2^2 \\ &= O_{P^*} \left(\frac{1}{n} \right), \end{aligned} \tag{C.11}$$

and $\Delta_F = \Delta_F^*$.

For the second term, we have by Lemma J.8 that

$$\begin{aligned} \bar{C}_2 &= \max_{m \leq p} \left\| \frac{1}{nT} \sum_{i=1}^n \sum_{s=1}^T \tilde{z}_{is}^* \frac{1}{pTn} \sum_{j=1}^n \hat{f}_j^* \sum_{t=1}^T \hat{f}_j^* \hat{\Lambda}'_t \tilde{U}_{it}^* \hat{g}'_{sm} \right\|_F \\ &\leq O_{P^*}(\omega_n^{*1/2}) \left(\frac{1}{T} \sum_{s=1}^T \left\| \frac{1}{pTn} \sum_{i=1}^n \sum_{t=1}^T \hat{\Lambda}'_t \tilde{U}_{it}^* \tilde{z}_{is}^* \right\|_2^2 \right)^{1/2} \\ &= O_{P^*} \left(\sqrt{\frac{\log n}{npT}} + \frac{\sqrt{\log n}}{nT} + \frac{1}{n} + \frac{1}{n} \left(\frac{\log n}{T} \right)^{1/4} \right) \omega_n^{*1/2}. \end{aligned}$$

Finally,

$$\begin{aligned}
\bar{C}_3 &= \max_{m \leq p} \left\| \frac{1}{nT} \sum_{i=1}^n \sum_{s=1}^T \tilde{z}_{is}^* \frac{1}{pTn} \sum_{j=1}^n \hat{f}_j^* \sum_{t=1}^T \hat{f}_i' \hat{\Lambda}_t \tilde{U}_{jt}^* \hat{g}'_{sm} \right\|_F \\
&\leq \left(\frac{1}{T} \sum_{s=1}^T \left(\frac{1}{n} \sum_{i=1}^n \tilde{z}_{is}^* \hat{f}_i' \right)^2 \right)^{1/2} \omega_n^{*1/2} \left\| \frac{1}{nTp} \sum_{j=1}^n \sum_{t=1}^T \hat{f}_j^* \hat{\Lambda}_t \tilde{U}_{jt}^* \right\|_2 \\
&= O_{P^*} \left(\omega_n^{*1/2} \frac{1}{\sqrt{n}} \right) \left(\Delta_F^2 + \sqrt{\frac{\Delta_F^2 \log n}{n}} \right)
\end{aligned}$$

where we used Lemma H.1 to obtain $\frac{1}{T} \sum_t \left\| \frac{1}{n} \sum_{i=1}^n \tilde{z}_{it}^* \hat{f}_i' \right\|_2^2 = O_{P^*} \left(\frac{1}{n} \right)$.

Combining the above, we have

$$\max_{m \leq p} \left\| \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T \hat{g}_{tm} \tilde{z}_{it}^* (\hat{f}_i^* - H^{*'} \hat{f}_i)' \right\|_F^2 = O_{P^*} \left(\frac{1}{n^2} + \frac{\log n}{npT} + \frac{\log n}{n^2 T^2} + \frac{\log^{1/2} n}{n^2 T^{1/2}} \right) \omega_n^*.$$

■

Lemma C.5. For $\hat{h}_{tk} \in \{\hat{\delta}_{yt}, \hat{\delta}_{dt}, \hat{\lambda}_{tk}\}$, we have

$$\begin{aligned}
&\max_{m, k \leq p} \left\| \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T (\hat{f}_i^* - H^{*'} \hat{f}_i) \tilde{U}_{it, m}^* \hat{h}'_{tk} \right\|_F \\
&\leq O_{P^*} \left(\frac{1}{T} \sqrt{\frac{\log p \log(np)}{np}} + \sqrt{\frac{\log n}{npT}} + \frac{\log p}{n\sqrt{T}} + \frac{1}{pT} + \sqrt{\frac{\log p}{npT}} + \frac{\sqrt{\log p}}{n} + \frac{\sqrt{|J|_0}}{nT} + \sqrt{\frac{|J|_0}{npT}} \right) \\
&\quad + O_{P^*}(\Delta_{ud})
\end{aligned}$$

for Δ_{ud} defined as in Lemma C.3. The term on the right-hand-side of the inequality defines Δ_{ud}^* .

Proof. We have $\max_{m, k \leq p} \left\| \frac{1}{nT} \sum_{i=1}^n \sum_{s=1}^T (\hat{f}_i^* - H^{*'} \hat{f}_i) \tilde{U}_{is, m}^* \hat{h}'_{sk} \right\|_F \leq \sum_{l=1}^5 D_l$, where each D_l is defined and bounded below.

$$\begin{aligned}
D_1 &= \max_{m, k \leq p} \left\| \frac{1}{T} \sum_{t=1}^T \frac{1}{n} \sum_{j=1}^n \hat{f}_j^* \tilde{U}_{jt}^{*'} \frac{1}{nTp} \sum_{i=1}^n \sum_{s=1}^T (\tilde{U}_{it}^* \tilde{U}_{is, m}^* - E^* \tilde{U}_{it}^* \tilde{U}_{is, m}^*) \hat{h}'_{sk} \right\|_F \\
&\leq \left(\frac{1}{T} \sum_{t=1}^T \left\| \frac{1}{n} \sum_{j=1}^n \hat{f}_j^* \tilde{U}_{jt}^{*'} \right\|_2^2 \right)^{1/2} \\
&\quad \times \max_{m, k \leq p} \left(\frac{1}{T} \sum_{t=1}^T \left\| \frac{1}{nTp} \sum_{i=1}^n \sum_{s=1}^T (\tilde{U}_{it}^* \tilde{U}_{is, m}^* - E^* \tilde{U}_{it}^* \tilde{U}_{is, m}^*) \hat{h}'_{sk} \right\|_F^2 \right)^{1/2} \\
&= O_{P^*} \left(\sqrt{\frac{\log(pT)}{n}} \left(\frac{\log n \log(pT)}{n} + \frac{\log(np)}{T} \right) \right) \left(\Delta_F + \frac{1}{\sqrt{n}} \right)
\end{aligned}$$

where we used Lemma J.8 and equation (C.11) from the proof of Lemma C.4 which gives

$$\frac{1}{pT} \sum_t \left\| \frac{1}{n} \sum_{j=1}^n \widehat{f}_j^* \widetilde{U}_{jt}^* \right\|_F^2 \leq O_{P^*} \left(\Delta_F^2 + \frac{1}{n} \right).$$

Next,

$$\begin{aligned} D_2 &= \max_{m,k \leq p} \left\| \sum_{l=1}^p \frac{1}{T} \sum_{t=1}^T \frac{1}{pn} \sum_{j=1}^n \widehat{f}_j^* \widetilde{U}_{jt,l}^* \frac{1}{nT} \sum_{i=1}^n \sum_{s=1}^T E^*(\widetilde{U}_{it,l}^* \widetilde{U}_{is,m}^*) \widehat{h}'_{sk} \right\|_F \\ &\leq \max_{tl} \left\| \frac{1}{n} \sum_{j=1}^n \widehat{f}_j^* \widetilde{U}_{jt,l}^* \right\|_2 \max_{m,k \leq p} \frac{1}{pT} \sum_{l=1}^p \sum_{t=1}^T \left\| \frac{1}{nT} \sum_{i=1}^n \sum_{s=1}^T E^*(\widetilde{U}_{it,l}^* \widetilde{U}_{is,m}^*) \widehat{h}'_{sk} \right\|_F \\ &= O_{P^*} \left(\sqrt{\frac{\log(pT)}{n}} + \Delta_F \right) \left(\Delta_F + \frac{\sqrt{\log(pT) \log p}}{n} + \sqrt{\frac{\log(pT)}{nT}} \right) \end{aligned}$$

where we used Lemma J.8 and $\max_{mt} \left\| \frac{1}{n} \widehat{F}^{*'} \widetilde{U}_{t,m}^* \right\|_2 = O_{P^*} \left(\sqrt{\frac{\log(pT)}{n}} + \Delta_F \right)$ due to Lemma H.1.

We then have

$$\begin{aligned} D_3 &= \max_{m,k \leq p} \left\| \frac{1}{nT} \sum_{i=1}^n \sum_{s=1}^T \frac{1}{pTn} \sum_{j=1}^n \widehat{f}_j^* \sum_{t=1}^T \widehat{f}_j^* \widehat{\Lambda}'_t (\widetilde{U}_{it}^* \widetilde{U}_{is,m}^* - E^* \widetilde{U}_{it}^* \widetilde{U}_{is,m}^*) \widehat{h}'_{sk} \right\|_F \\ &\leq O_{P^*}(1) \max_{m,k \leq p} \left\| \frac{1}{nT} \sum_{i=1}^n \sum_{s=1}^T \frac{1}{pT} \sum_{t=1}^T \widehat{\Lambda}'_t (\widetilde{U}_{it}^* \widetilde{U}_{is,m}^* - E^* \widetilde{U}_{it}^* \widetilde{U}_{is,m}^*) \widehat{h}'_{sk} \right\|_F \\ &= O_{P^*} \left(\sqrt{\frac{\log p}{n}} \left(\frac{\log n \log(pT)}{n} + \frac{\log(np)}{T} \right) \Delta_F^2 \right) \end{aligned}$$

by Lemma J.8. We also have

$$\begin{aligned} D_4 &= \max_{m,k \leq p} \left\| \frac{1}{nT} \sum_{i=1}^n \sum_{s=1}^T \frac{1}{pTn} \sum_{j=1}^n \widehat{f}_j^* \sum_{t=1}^T \widehat{f}_j^* \widehat{\Lambda}'_t E^* \widetilde{U}_{it}^* \widetilde{U}_{is,m}^* \widehat{h}'_{sk} \right\|_F \\ &\leq O_{P^*}(1) \max_{m,k \leq p} \left\| \frac{1}{nT} \sum_{i=1}^n \sum_{s=1}^T \frac{1}{pT} \sum_{t=1}^T \widehat{\Lambda}'_t E^* \widetilde{U}_{it}^* \widetilde{U}_{is,m}^* \widehat{h}'_{sk} \right\|_F \\ &= O_{P^*} \left(\frac{1}{pT} + \sqrt{\frac{\log p}{npT}} + \frac{\sqrt{\log p}}{n} + \frac{\sqrt{|J|_0}}{nT} + \sqrt{\frac{|J|_0}{npT}} \right) + O_{P^*}(\Delta_{ud}) \end{aligned}$$

where the inequality follows from Lemma J.9 (iii). Finally,

$$\begin{aligned} D_5 &= \max_{m,k \leq p} \left\| \frac{1}{pTn} \sum_{j=1}^n \widehat{f}_j^* \sum_{t=1}^T \widehat{\Lambda}'_t \widetilde{U}_{jt}^* \frac{1}{nT} \sum_{s=1}^T \sum_{i=1}^n \widehat{f}_i^* \widetilde{U}_{is,m}^* \widehat{h}'_{sk} \right\|_F \\ &\leq \left\| \frac{1}{pTn} \sum_{j=1}^n \widehat{f}_j^* \sum_{t=1}^T \widehat{\Lambda}'_t \widetilde{U}_{jt}^* \right\|_2 \max_{m,k \leq p} \left\| \frac{1}{nT} \sum_{s=1}^T \sum_{i=1}^n \widehat{f}_i^* \widetilde{U}_{is,m}^* \widehat{h}'_{sk} \right\|_F \end{aligned}$$

$$\begin{aligned}
&=^{(a)} O_{P^*} \left(\Delta_F^2 + \sqrt{\frac{\Delta_F^2 \log n}{n}} \right) \max_{m,k \leq p} \left\| \frac{1}{nT} \sum_{s=1}^T \sum_{i=1}^n \widehat{f}_i \widetilde{U}_{is,m}^* \widehat{h}'_{sk} \right\|_F \\
&= O_{P^*} \left(\Delta_F^2 + \sqrt{\frac{\Delta_F^2 \log n}{n}} \right)
\end{aligned}$$

where equality (a) results by applying Lemma J.8 (iii). Note that the upper bound achieved in the last equality is not sharp but is sufficient to verify Assumptions about Δ_{ud}^* .

Combining the above terms, we reach

$$\begin{aligned}
&\max_{m,k \leq p} \left\| \frac{1}{nT} \sum_{i=1}^n \sum_{s=1}^T (\widehat{f}_i^* - H^{*'} \widehat{f}_i) \widetilde{U}_{is,m}^* \widehat{h}'_{sk} \right\|_F \\
&= O_{P^*} \left(\sqrt{\frac{\log(pT)}{n}} \left(\frac{\log n \log(pT)}{n} + \frac{\log(np)}{T} \right) \left(\Delta_F + \frac{1}{\sqrt{n}} \right) \right. \\
&\quad + O_{P^*} \left(\sqrt{\frac{\log(pT)}{n}} + \Delta_F \right) \left(\Delta_F + \frac{\sqrt{\log(pT) \log p}}{n} + \sqrt{\frac{\log(pT)}{nT}} \right) \\
&\quad + O_{P^*} (\Delta_{ud} + \Delta_F \sqrt{\frac{\log n}{n}}) + O_{P^*} \left(\Delta_F \sqrt{\frac{\log p}{n}} \left(\frac{\log n \log(pT)}{n} + \frac{\log(np)}{T} \right) \right) \\
&\quad + O_{P^*} \left(\frac{1}{pT} + \sqrt{\frac{\log p}{npT}} + \frac{\sqrt{\log p}}{n} + \frac{\sqrt{|J|_0}}{nT} + \sqrt{\frac{|J|_0}{npT}} \right) \\
&= O_{P^*} \left(\frac{1}{T} \sqrt{\frac{\log p \log(np)}{np}} + \sqrt{\frac{\log n}{npT}} + \frac{\log p}{n\sqrt{T}} + \frac{1}{pT} + \sqrt{\frac{\log p}{npT}} + \frac{\sqrt{\log p}}{n} + \frac{\sqrt{|J|_0}}{nT} + \sqrt{\frac{|J|_0}{npT}} \right) \\
&\quad + O_{P^*} (\Delta_{ud})
\end{aligned}$$

where Δ_{ud} is defined as in Lemma C.3. ■

REFERENCES

- ACEMOGLU, D., JOHNSON, S. and ROBINSON, J. A. (2001). The colonial origins of comparative development: An empirical investigation. *American Economic Review* **91** 1369–1401.
- AGARWAL, A., NEGAHBAN, S., WAINWRIGHT, M. J. ET AL. (2012). Fast global convergence of gradient methods for high-dimensional statistical recovery. *The Annals of Statistics* **40** 2452–2482.
- AHN, S. and HORENSTEIN, A. (2013). Eigenvalue ratio test for the number of factors. *Econometrica* **81** 1203–1227.
- ANDREWS, D. W. (2002). Higher-order improvements of a computationally attractive k-step bootstrap for extremum estimators. *Econometrica* **70** 119–162.
- ARELLANO, M. (1987). Computing robust standard errors for within-groups estimators. *Oxford Bulletin of Economics and Statistics* **49** 431–434.
- BAI, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica* **71** 135–171.

- BAI, J. (2009). Panel data models with interactive fixed effects. *Econometrica* **77** 1229–1279.
- BAI, J. and LI, K. (2014). Theory and methods of panel data models with interactive effects. *The Annals of Statistics* **42** 142–170.
- BAI, J. and LIAO, Y. (2013). Statistical inferences using large estimated covariances for panel data and factor models. Tech. rep., University of Maryland.
- BAI, J. and NG, S. (2002). Determining the number of factors in approximate factor models. *Econometrica* **70** 191–221.
- BAI, J. and NG, S. (2006). Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions. *Econometrica* **74** 133–1150.
- BELLONI, A., CHEN, D., CHERNOZHUKOV, V. and HANSEN, C. (2012). Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica* **80** 2369–2429.
- BELLONI, A., CHERNOZHUKOV, V., FERNÁNDEZ-VAL, I. and HANSEN, C. (2014a). Program evaluation with high-dimensional data. *Econometrica* Forthcoming, arXiv:1311.2645.
- BELLONI, A., CHERNOZHUKOV, V. and HANSEN, C. (2014b). Inference on treatment effects after selection among high-dimensional controls. *The Review of Economic Studies* **81** 608–650.
- BELLONI, A., CHERNOZHUKOV, V., HANSEN, C. and KOZBUR, D. (2015). Inference in high dimensional panel models with an application to gun control. *Journal of Business and Economic Statistics* Forthcoming, arXiv:1411.6507.
- BELLONI, A., CHERNOZHUKOV, V. and KATO, K. (2013a). Uniform post selection inference for lad regression models and other z-estimation problems. *arXiv preprint arXiv:1304.0282* ArXiv, 2013; Oberwolfach, 2012, Luminy, 2012.
- BELLONI, A., CHERNOZHUKOV, V. and WEI, Y. (2013b). Honest confidence regions for logistic regression with a large number of controls. *arXiv preprint arXiv:1304.3969* ArXiv, 2013.
- BERK, R., BROWN, L., BUJA, A., ZHANG, K. and ZHAO, L. (2013). Valid post-selection inference. *Annals of Statistics* **41** 802–837.
- BERNANKE, B., BOIVIN, J. and ELIASZ, P. (2005). Factor augmented vector autoregressions (fvars) and the analysis of monetary policy. *Quarterly Journal of Economics* **120** 387–422.
- BERTRAND, M., DUFLO, E. and MULLAINATHAN, S. (2004). How much should we trust differences-in-differences estimates? *Quarterly Journal of Economics* **119** 249–275.
- BONHOMME, S. and MANRESA, E. (2015). Grouped patterns of heterogeneity in panel data. *Econometrica* **83** 1147–1184.
- CHATTERJEE, A. and LAHIRI, S. N. (2011). Bootstrapping lasso estimators. *Journal of the American Statistical Association* **106** 608–625.
- CHERNOZHUKOV, V., CHETVERIKOV, D., DEMIRER, M., DUFLO, E., HANSEN, C. and NEWEY, W. (2016). Double machine learning for treatment and causal parameters. *arXiv:1608.00060* .
- CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *Annals of Statistics* **41** 2786–2819.
- CHERNOZHUKOV, V. and HANSEN, C. (2008). The reduced form: A simple approach to inference with weak instruments. *Economics Letters* **100** 68–71.
- CHOI, I. (2012). Efficient estimation of factor models. *Econometric Theory* **28** 274–308.
- COOK, P. J. and LUDWIG, J. (2006). The social costs of gun ownership. *Journal of Public Economics* **90** 379–391.

- DEZEURE, R., BÜHLMANN, P. and ZHANG, C.-H. (2016). High-dimensional simultaneous inference with the bootstrap. *arXiv preprint arXiv:1606.03940* .
- FAN, J. and LI, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association* **96** 1348–1360.
- FAN, J. and LV, J. (2011). Non-concave penalized likelihood with np-dimensionality. *IEEE Transactions on Information Theory* **57** 5467–5484.
- FAN, J., XUE, L. and YAO, J. (2016). Sufficient forecasting using factor models. *Journal of Econometrics, forthcoming* .
- FARRELL, M. (2015). Robust inference on average treatment effects with possibly more covariates than observations. *Journal of Econometrics* **174** 1–23.
- FITHIAN, W., SUN, D. and TAYLOR, J. (2014). Optimal inference after model selection. *arXiv preprint arXiv:1410.2597v1* ArXiv, 2014.
- FU, W. J. (1998). Penalized regressions: the bridge versus the lasso. *Journal of computational and graphical statistics* **7** 397–416.
- GAUTIER, E. and TSYBAKOV, A. B. (2011). High-dimensional instrumental variables regression and confidence sets. *ArXiv working report* .
- G'SELL, M. G., TAYLOR, J. and TIBSHIRANI, R. (2013). Adaptive testing for the graphical lasso. *arXiv preprint arXiv:1307.4765* .
- HAHN, P. R., MUKEHERJEE, S. and CARVALHO, C. (2013). Partial factor modeling: Predictor dependent shrinkage for linear regression. *JASA* **108** 999–1008.
- HANSEN, C. B. (2007). Asymptotic properties of a robust variance matrix estimator for panel data when t is large. *Journal of Econometrics* **141** 597–620.
- JAVANMARD, A. and MONTANARI, A. (2014). Confidence intervals and hypothesis testing for high-dimensional regression. *arXiv:1306.3171v2* .
- KADKHODAIE, M., SANJABI, M. and LUO, Z.-Q. (2014). On the linear convergence of the approximate proximal splitting method for non-smooth convex optimization. *Journal of the Operations Research Society of China* **2** 123–141.
- KOZBUR, D. (2015). Testing-based forward model selection. *arXiv:1512.02666* .
- LEE, J. D., SUN, D. L., SUN, Y. and TAYLOR, J. E. (2016). Exact post-selection inference with the lasso. *Annals of Statistics* **44** 907–927.
- LEE, J. D. and TAYLOR, J. E. (2014). Exact post model selection inference for marginal screening. *Advances in Neural Information Processing Systems* .
- LOCKHART, R., TAYLOR, J., TIBSHIRANI, R. J. and TIBSHIRANI, R. (2014). A significance test for the lasso (with discussion). *Annals of Statistics* **42** 413–468.
- LOFTUS, J. R. and TAYLOR, J. E. (2014). A significance test for forward stepwise model selection. *arXiv preprint arXiv:1405.3920* .
- LOH, P.-L. and WAINWRIGHT, M. J. (2015). Regularized m-estimators with nonconvexity: Statistical and algorithmic theory for local optima. *Journal of Machine Learning Research* **16** 559–616.
- MAMMEN, E. (1993). Bootstrap and wild bootstrap for high dimensional linear models. *The Annals of Statistics* 255–285.
- MERLEVÈDE, F., PELIGRAD, M. and RIO, E. (2011). A bernstein type inequality and moderate deviations for weakly dependent sequences. *Probability Theory and Related Fields* **151** 435–474.

- MOON, R. and WEIDNER, M. (2015a). Dynamic linear panel regression models with interactive fixed effects. *Econometric Theory* Forthcoming.
- MOON, R. and WEIDNER, M. (2015b). Linear regression for panel with unknown number of factors as interactive fixed effects. *Econometrica* **83** 1543–1579.
- NESTEROV, Y. (2007). Gradient methods for minimizing composite objective function. Tech. rep., UCL.
- PESARAN, H. (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica* **74** 967–1012.
- RASKUTTI, G., WAINWRIGHT, M. J. and YU, B. (2010). Restricted eigenvalue properties for correlated gaussian designs. *Journal of Machine Learning Research* **99** 2241–2259.
- STOCK, J. and WATSON, M. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association* **97** 1167–1179.
- SU, L. and CHEN, Q. (2013). Testing homogeneity in panel data models with interactive fixed effects. *Econometric Theory* **29** 1079–1135.
- TAYLOR, J., LOCKHART, R., TIBSHIRANI, R. J. and TIBSHIRANI, R. (2014). Post-selection adaptive inference for least angle regression and the lasso. *arXiv preprint arXiv:1401.3889* .
- VAN DE GEER, S., BÜHLMANN, P., RITOV, Y. and DEZEURE, R. (2014). On asymptotically optimal confidence regions and tests for high-dimensional models. *Annals of Statistics* **42** 1166–1202.
- WAGER, S. and ATHEY, S. (2015). Estimation and inference of heterogeneous treatment effects using random forests. *arXiv:1510.04342* .
- ZHANG, C.-H. and ZHANG, S. S. (2014). Confidence intervals for low dimensional parameters in high dimensional linear models. *Journal of the Royal Statistical Society: Series B* **76** 217–242.

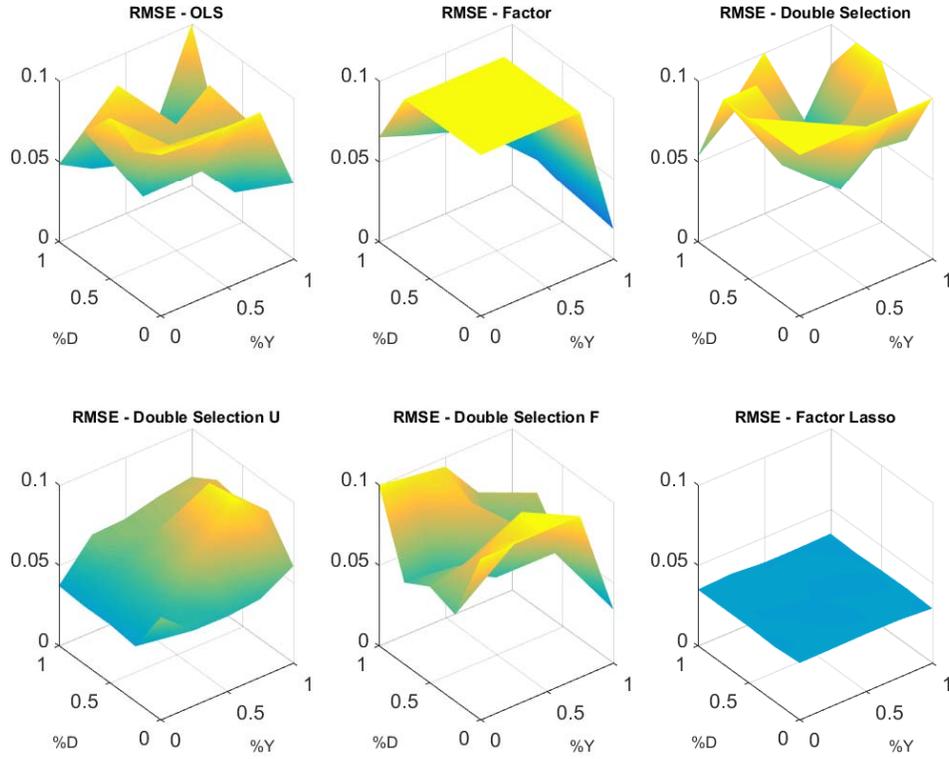


FIGURE 1. This figure shows the simulation RMSE of each of the estimators described in the text for estimating the coefficient of interest in a panel partial factor model. RMSE (truncated at 0.1) is shown in the vertical axis. The horizontal axes give the fraction of the explanatory power in an infeasible regression of Y on factors and factor residuals, “%Y,” and the fraction of the explanatory power in an infeasible regression of D on factors and factor residuals, “%D,” where the infeasible regressions are described in the text.

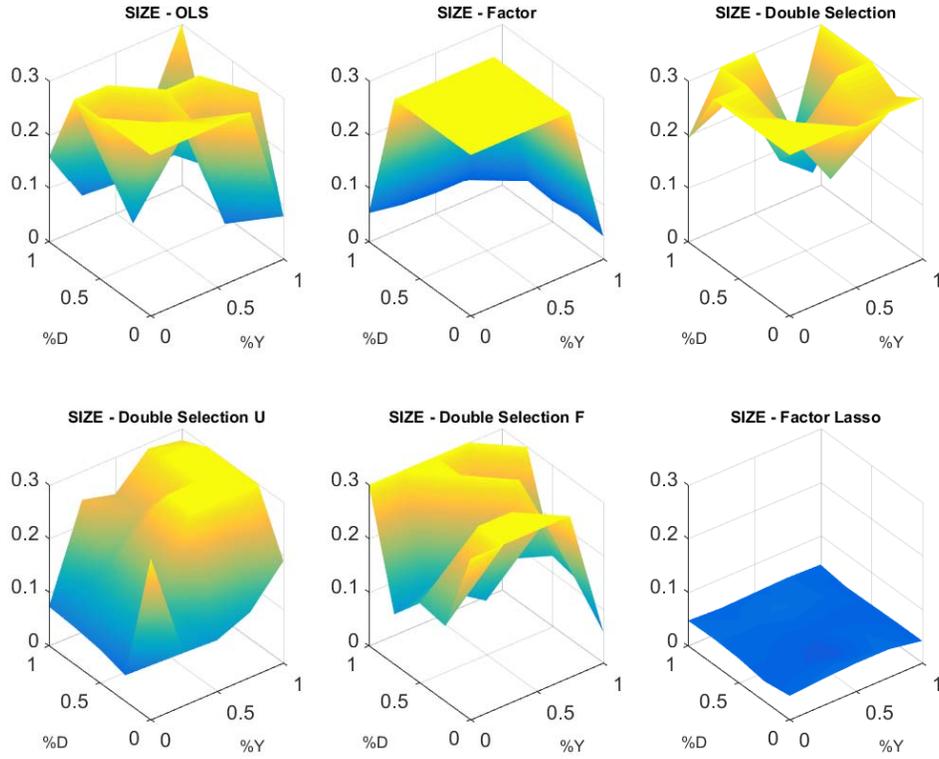


FIGURE 2. This figure shows the simulation size of 5% level tests based on each of the estimators described in the text for the PPFM. Size (truncated at 0.3) is shown in the vertical axis. The horizontal axes give the fraction of the explanatory power in an infeasible regression of Y on factors and factor residuals, “%Y,” and the fraction of the explanatory power in an infeasible regression of D on factors and factor residuals, “%D,” where the infeasible regressions are described in the text.

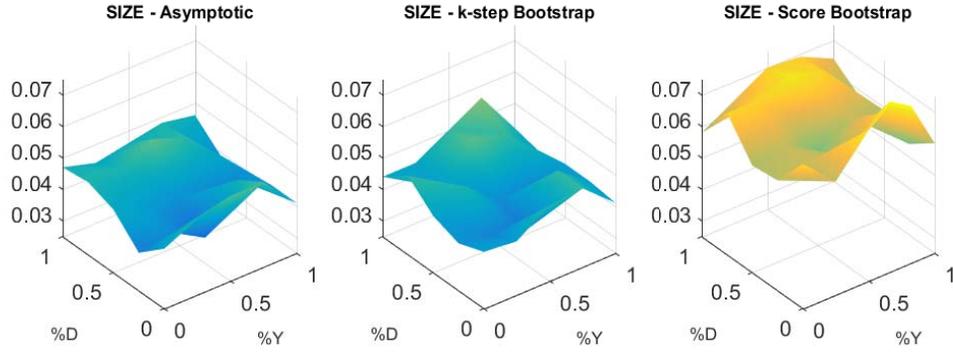


FIGURE 3. This figure shows the simulation size of 5% level tests based on the factor-lasso estimator in the PPFM and the asymptotic Gaussian approximation, the k -step bootstrap, and a score based bootstrap. Size is shown in the vertical axis. The horizontal axes give the fraction of the explanatory power in an infeasible regression of Y on factors and factor residuals, “%Y,” and the fraction of the explanatory power in an infeasible regression of D on factors and factor residuals, “%D,” where the infeasible regressions are described in the text.

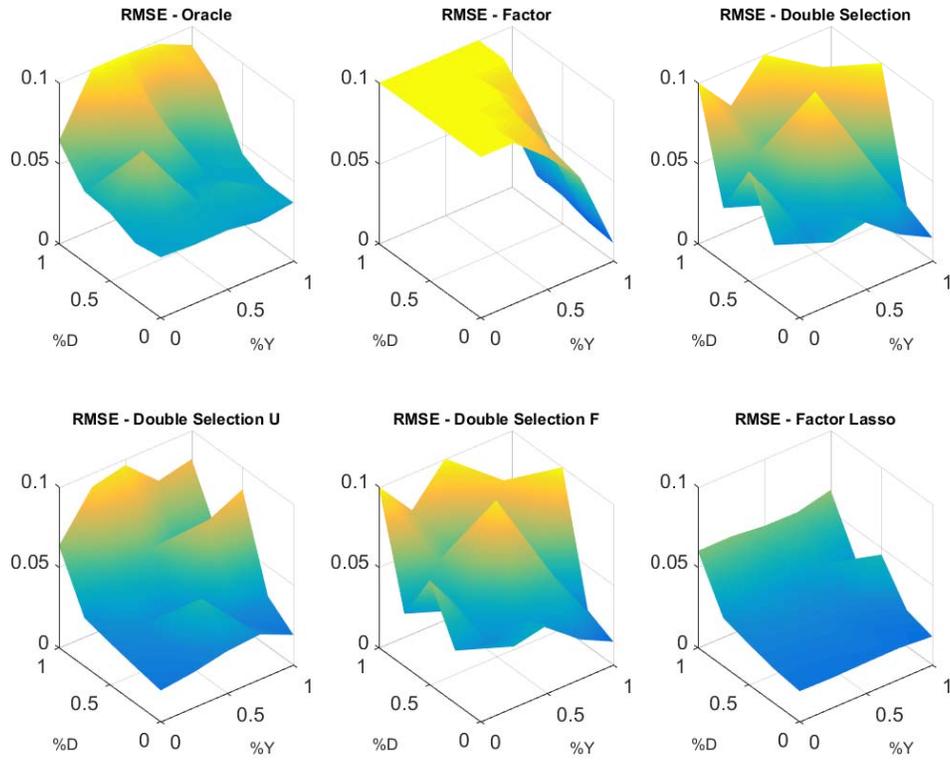


FIGURE 4. This figure shows the simulation RMSE of each of the estimators described in the text for estimating the coefficient of interest in an IV partial factor model. RMSE (truncated at 0.1) is shown in the vertical axis. The horizontal axes give the fraction of the explanatory power in an infeasible regression of Y on factors and factor residuals, “%Y,” and the fraction of the explanatory power in an infeasible regression of D on factors and factor residuals, “%D,” where the infeasible regressions are described in the text.

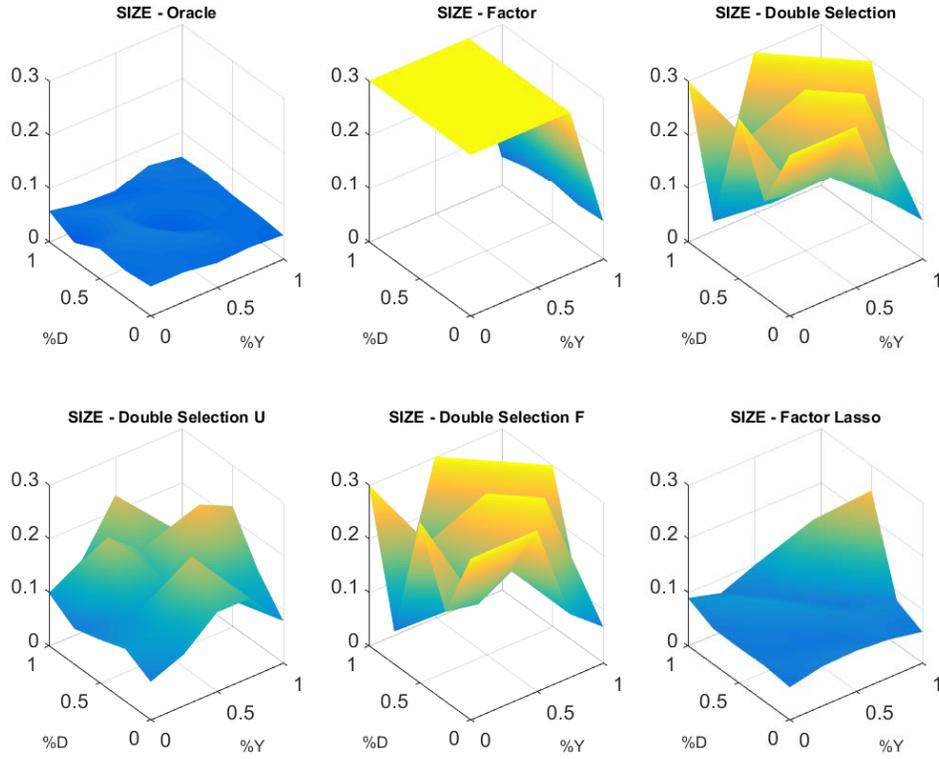


FIGURE 5. This figure shows the simulation size of 5% level tests based on each of the estimators described in the text for the IV partial factor model. Size (truncated at 0.3) is shown in the vertical axis. The horizontal axes give the fraction of the explanatory power in an infeasible regression of Y on factors and factor residuals, "%Y," and the fraction of the explanatory power in an infeasible regression of D on factors and factor residuals, "%D," where the infeasible regressions are described in the text.

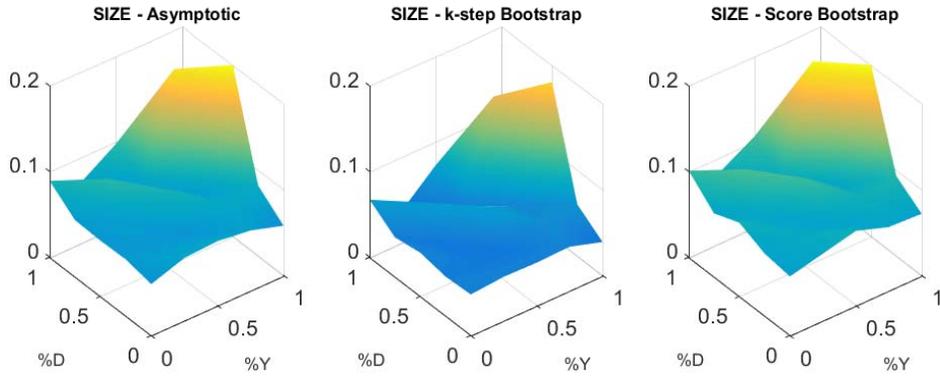


FIGURE 6. This figure shows the simulation size of 5% level tests based on the factor-lasso estimator in the IV partial factor model and the asymptotic Gaussian approximation, the k-step bootstrap, and a score based bootstrap. Size is shown in the vertical axis. The horizontal axes give the fraction of the explanatory power in an infeasible regression of Y on factors and factor residuals, “%Y,” and the fraction of the explanatory power in an infeasible regression of D on factors and factor residuals, “%D,” where the infeasible regressions are described in the text.

TABLE 1. Estimates of the Effect of Gun Prevalence on Homicide Rates

	Overall	Gun	non-Gun
Cook and Ludwig (2006) Baseline	0.086 (0.038)	0.173 (0.049)	-0.033 (0.040)
Post Double Selection	0.062 (0.042)	0.138 (0.059)	-0.055 (0.042)
	[-0.019,0.143]	[0.036,0.240]	[-0.139,0.029]
Factor	0.104 (0.043)	0.210 (0.064)	-0.022 (0.040)
	[0.019,0.189]	[0.097,0.323]	[-0.099,0.055]
Factor-Lasso	0.069 (0.036)	0.167 (0.046)	-0.048 (0.040)
	[0.000,0.138]	[0.078,0.256]	[-0.128,0.032]

This table presents estimates of the effect of gun ownership on homicide rates for a panel of 195 US Counties over the years 1980-1999. The columns “Overall”, “Gun”, and “non-Gun” respectively report the estimated effect of gun prevalence on the log of the overall homicide rate, the log of the gun homicide rate, and the log of the non-gun homicide rate. Each row corresponds to a different specification as described in the text. In each specification, the outcome corresponding to the column label is regressed on lagged log(FSS) (a proxy for gun ownership) and additional covariates as described in the text. Each specification includes a full set of year and county fixed effects. Standard errors clustered by county are provided in parentheses. k-step bootstrap 95% confidence intervals are given in brackets.

TABLE 2. Estimates of the First-Stage Relationship between Settler Mortality and Protection from Expropriation

	$\hat{\pi}$	Estimated s.e.	Bootstrap C.I.
Latitude	-0.549	(0.166)	[-0.851,-0.246]
All Controls	-0.218	(0.168)	[-0.778,0.341]
Double Selection	-0.364	(0.178)	[-0.885,0.158]
Factor	-0.475	(0.173)	[-0.880,-0.070]
Factor-Lasso	-0.353	(0.183)	[-0.708,0.002]

This table presents estimates of the coefficient on the instrument (Settler Mortality) in the first-stage regression of the endogenous variable from the Acemoglu et al. (2001) example (Protection from Expropriation) on the instrument and geographic controls using different methods. The row labeled “Latitude” uses the single variable distance from the equator to control for geography. “All Controls” uses all 20 geographic controls without dimension reduction. “Double Selection” uses the approach of Belloni et al. (2014b) to select important controls from among the 20 potential geography measures. “Factor” reduces dimension through positing a conventional factor model. “Factor-Lasso” makes use of the approach developed in this paper. Point estimates from each method are provided in the column “ $\hat{\pi}$ ” and the associated estimated asymptotic standard errors are given in “Estimated s.e.”. The k-step bootstrap 95% confidence interval is reported in “Bootstrap C.I.”.

BOOTH SCHOOL OF BUSINESS, UNIVERSITY OF CHICAGO, CHICAGO, IL 60637

E-mail address: Christian.Hansen@chicagobooth.edu

DEPARTMENT OF ECONOMICS, RUTGERS UNIVERSITY, NEW BRUNSWICK, NJ 08901

E-mail address: yuan.liao@rutgers.edu