Testing for sparse idiosyncratic components in factor-augmented regression models

Jad Beyhum
Department of Economics, KU Leuven, Belgium
and
Jonas Striaukas
Department of Finance, Copenhagen Business School, Denmark

August 25, 2024

Abstract

We propose a novel bootstrap test of a dense model, namely factor regression, against a sparse plus dense alternative augmented model with sparse idiosyncratic components. The asymptotic properties of the test are established under time series dependence and polynomial tails. We outline a data-driven rule to select the tuning parameter and prove its theoretical validity. In simulation experiments, our procedure exhibits high power against sparse alternatives and low power against dense deviations from the null. Moreover, we apply our test to various datasets in macroeconomics and finance and often reject the null. This suggests the presence of sparsity — on top of a dense component — in commonly studied economic applications. The R package 'FAS' implements our approach.

Keywords: sparse plus dense, high-dimensional inference, LASSO, factor models

1 Introduction

In this paper, we investigate a factor-augmented sparse regression model. Our analysis involves an observed sample of T real-valued outcomes $y_1, \ldots, y_T$, and high-dimensional regressors $x_1, \ldots, x_T \in \mathbb{R}^p$, which are interconnected as follows:

$$\begin{aligned} y_t &= f_t^\top \gamma^* + u_t^\top \beta^* + \varepsilon_t, \\ x_t &= B f_t + u_t, \quad t = 1 \ldots, T. \end{aligned} \tag{1}$$

Here, $\varepsilon_t \in \mathbb{R}$ represents a random error, u_t is a p-dimensional random vector of idiosyncratic shocks, f_t is a K-dimensional random vector of factors, and B is a $p \times K$ (nonrandom) matrix of loadings. The parameters of interest are $\gamma^* \in \mathbb{R}^K$ and $\beta^* \in \mathbb{R}^p$ and the right-hand side of (1) is unobserved. We consider the case where the number p of regressors is large with respect to the sample size T and a sparsity condition on the high-dimensional parameter vector β^* is imposed. In the asymptotic regimes we study, T goes to infinity, p is allowed to grow with T while K remains fixed. The model formulation in equation (1) effectively merges two popular approaches in handling high-dimensional datasets: factor regression (Stock & Watson (2002), Bai & Ng (2006)) and sparse high-dimensional regression (Tibshirani (1996), Bickel et al. (2009)). Such a model allows the outcome to be related to the regressors through both common and idiosyncratic shocks and may better explain the data than factor regression or sparse regression alone (see Fan et al. (2024, 2023), which introduce and study model (1)). As noted in Fan et al. (2023), this type of structure has many applications in forecasting, causal inference and to describe correlation networks. Note that, as in Stock & Watson (2002), Bai & Ng (2006), Fan et al. (2023), we could augment the model (1) with additional regressors w_t entering the first equation of (1) but not the second one. This case is discussed in the Appendix Section A.

We develop a test for the hypothesis:

$$H_0 : \beta^* = 0 \quad \text{against} \quad H_1 : \beta^* \neq 0 \text{ is sparse,} \tag{2}$$

where our theory outlines the set of sparse alternatives against which our test has power. Our specification test sheds light on the data-generating process by allowing us to determine if the underlying model is dense (as is the factor regression model) or sparse plus dense, as is the factor-augmented sparse regression model. This determination will then tell us if the relation between the regressors and the outcome is only driven by common shocks (factor regression) or if idiosyncratic shocks also play a role (factor-augmented sparse regression). The question of the adequacy of sparse or dense representations has recently garnered significant attention (see, e.g., Giannone et al. (2021), Kolesár et al. (2023)). However, existing studies mostly focus on the differences between sparse and dense models and have found that dense models are often more adequate. In contrast, we compare a dense model with a sparse plus dense alternative.

In this paper, we propose a new bootstrap test for (2). Our test's principle is to compare two estimators of $\sum_{t=1}^{T} u_t \varepsilon_t$. The first estimator is only consistent under the null, while the second estimator relies on the LASSO, and is, therefore, consistent under sparse alternatives. Our proposed test does not require estimating covariance matrices and is easy to implement. Following Lederer & Vogt (2021), we outline a data-driven rule to select the tuning parameter of the LASSO estimator and prove its theoretical validity. We establish the validity of the test within a theoretical framework that accommodates scenarios where the number of variables, denoted by p, can significantly exceed T, and the explanatory variables exhibit strong mixing and possess polynomial tails. We use simulations to evaluate the finite sample properties of our procedure. Our test controls size and exhibits good power against sparse alternatives even when p greatly exceeds T or the data are heavy-tailed and serially correlated. A potential limitation of our approach might be that our approach also rejects when β^* is nonzero but has a dense structure. To assess this issue, we conduct simulations with dense β^*. We find that our test exhibits very low power against such alternatives (in absolute terms and also relatively to sparse alternatives with the same signal-to-noise ratio). Hence, our test exhibits some robustness against dense alternatives. This result is intuitive: the LASSO estimator on which our test relies sets to zero very small coefficients pertaining to dense alternatives. Finally, we apply our test to several commonly studied datasets in macroeconomics and finance and often reject the null. This suggests that sparsity can help describe economic data once a dense component (here, modeled through the factors) is included in the model. This result complements the recent studies Giannone et al. (2021), Kolesár et al. (2023), which concluded that dense representations were often more appropriate for economic data. The R package `'FAS'` implements our approach.

Related literature. Our paper relates to the growing literature on factor-augmented sparse models, see Hansen & Liao (2019), Fan, Masini & Medeiros (2022), Fan et al. (2024, 2023), Vogt et al. (2022), Beyhum & Striaukas (2023), Barigozzi et al. (2024) among others. In particular, Fan et al. (2023) proposes a general framework for factor-augmented sparse models. As noted by a reviewer, our test can be used to check that the sparse idiosyncratic components are jointly significant after the third step of the methodology by Fan et al. (2023). Note that Fan et al. (2023)'s model includes lags of idiosyncratic terms and additional regressors. In the Appendix, we explain how to extend our test to the case with additional regressors. Section A of the Online Appendix outlines an adapted test with a lag. We do not formally prove that our test works in these cases, but we conjecture that the asymptotic properties extend to these more general models. Simulations reported in Section B of the Online Appendix corroborate this presumption. We also note that the factor-augmented sparse model studied in the present paper is a generalization of an earlier model studied in Fosten (2017*b*,*a*) which augments the factor regression of Stock & Watson (2002) with a low-dimensional set of idiosyncratic terms.

Fan et al. (2024) recently introduced the Factor-Adjusted deBiased Test (FabTest) for evaluating (2). However, the FabTest exhibits several limitations. The test relies on a desparsified LASSO estimator based on model (1). To achieve desparsification, Fan et al. (2024) utilized the nodewise LASSO method proposed by Zhang & Zhang (2014) and van de Geer et al. (2014) for estimating the precision matrix of the idiosyncratic shocks. However, this approach introduces p additional tuning parameters, in addition to the one used in the original LASSO regression. Although the tuning parameters are selected through cross-validation in practice, Fan et al. (2024) did not provide a theoretical justification for this selection procedure. Besides, inferential theory for LASSO-type regressions is not well understood when the tuning parameter is selected by cross-validation. Moreover, the test's performance may deteriorate due to errors associated with the nodewise LASSO estimates, and it incurs a heavy computational cost. Another limitation of the FabTest is its reliance on estimating the variance of ε_t, which can lead to imprecise results where variance estimation is challenging. Additionally, Fan et al. (2024) only established the validity of the FabTest for i.i.d. sub-Gaussian data (see Section 2 in Fan et al. (2024)). An alternative is to use the partial covariance test developed by Fan et al. (2023) and applied in Fan, Masini & Medeiros (2022). This test's principle is to estimate the covariance matrix of the idiosyncratic terms (including the idiosyncratic term of y_t, that is, $u_t^\top \beta^* + \varepsilon_t$) and then use a Gaussian bootstrap under the null to obtain the critical value of the test statistic. In contrast to our test, this partial covariance test does not make use of the LASSO estimator and requires estimating a high-dimensional covariance matrix, which is challenging.

Finally, we would like to note that this paper contributes to various other strands of literature. First, it connects to recent literature considering testing for high-dimensional parameters. There exists several approaches, see Fan et al. (2015), Zhu & Bradic (2018), Chernozhukov et al. (2019), Lederer & Vogt (2021), He et al. (2023) and references therein. Our strategy draws inspiration from Lederer & Vogt (2021), a recent paper that introduces a bootstrap procedure for selecting the penalty parameter of LASSO in a standard sparse linear regression. They employ this procedure to test the null hypothesis that a specific high-dimensional parameter equals zero. We adapt their approach to the case with unobserved factors, time series dependence and polynomial tails, which poses a challenge beyond the scope of the results in Lederer & Vogt (2021). Second, our work is related to the literature on inference on parameters of additional low-dimensional regressors in the factor regression model of Stock & Watson (2002), see Bai & Ng (2006), Gonçalves & Perron (2014, 2020). Third, our work connects with the literature on specification tests for models involving unobserved factors. Many papers test for the validity of the assumption that loadings are time-independent in the approximate factor model itself — the second equation in (1) — (Breitung & Eickmeier (2011), Chen et al. (2014), Han & Inoue (2015), Yamamoto & Tanaka (2015), Su & Wang (2017, 2020), Baltagi et al. (2021), Xu (2022), Fu et al. (2023)), while Corradi & Swanson (2014) tests for time-independence of all coefficients in the factor regression model of Stock & Watson (2002). Our approach complements this literature by proposing a specification test of the factor regression model under a different alternative, namely the factor-augmented sparse regression model.

Outline. The paper is organized as follows. In Section 2, we outline our testing procedure. Its asymptotic properties are studied in Section 3. Then, Section 4 contains Monte Carlo simulations. The empirical applications can be found in Section 5. Section 6 concludes. The Appendix contains an adapted testing procedure for a model with additional regressors and lags. The Online Appendix includes all proofs and additional discussions and simulation and empirical results.

Notation. For an integer $N \in \mathbb{N}$, let $[N] = \{1, \ldots, N\}$. The transpose of a $n_1 \times n_2$ matrix A is written $A^\top$. Its k^{th} singular value is $\sigma_k(A)$. Let us also define the Euclidean norm $\|A\|_2^2 = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} A_{ij}^2$ and the sup-norm $\|A\|_\infty = \max_{i\in[n_1], j\in[n_2]} |A_{ij}|$. The quantity $n_1 \vee n_2$ is the maximum of n_1 and n_2, $n_1 \wedge n_2$ is the minimum of n_1 and n_2. For $N \in \mathbb{N}$, I_N is the identity matrix of size $N \times N$. For a real-valued random variable Z and $g > 0$, we let $\|Z\| = \mathbb{E}[|Z|^g]^{\frac{1}{g}}$. For a d-dimensional random vector Z, we define $\|Z\|_g = \sup_{u\in\mathbb{R}^d:\ \|u\|_2 \leq 1} \|u^\top Z\|_g$.

2 The test

2.1 Testing procedure

In this subsection, we explain our testing procedure, which is then summarized in algorithmic form in subsection 2.2. To facilitate understanding, we rewrite the model in matrix form as follows:

$$Y = F\gamma^* + U\beta^* + \mathcal{E},$$
$$X = FB^\top + U,$$

where $Y = (y_1, \ldots, y_T)^\top$, $F = (f_1, \ldots, f_T)^\top$ is a $T \times K$ matrix, $U = (u_1, \ldots, u_T)^\top$ and $X = (x_1, \ldots, x_T)^\top$ are $T \times p$ matrices and $\mathcal{E} = (\varepsilon_1, \ldots, \varepsilon_T)^\top$.

It is important to note that, under the null hypothesis H_0, we have $U^\top(Y - F\gamma^*) = U^\top\mathcal{E}$. This observation suggests a testing procedure that involves computing an estimate $2T^{-1} \|U^\top(Y - F\gamma^*)\|_\infty$ and comparing it with the (estimated) quantiles of $2T^{-1} \|U^\top\mathcal{E}\|_\infty$.[1]

[1]We have a factor 2 in front of $T^{-1} \|U^\top(Y - F\gamma^*)\|_\infty$ and $T^{-1} \|U^\top\mathcal{E}\|_\infty$ because $2T^{-1} \|U^\top\mathcal{E}\|_\infty$ is the effective noise of the problem, a natural concept in the literature on the LASSO, see Lederer & Vogt (2021).

We can estimate $U^\top(Y - F\gamma^*)$ by principal components analysis. First, let $\widehat{K}$ be one of the many estimators of the number of factors K available in the literature (see for instance Bai & Ng (2002), Onatski (2010), Ahn & Horenstein (2013), Bai & Ng (2019), Fan, Guo & Zheng (2022)). As in Fan et al. (2024), we let the columns of $\widehat{F}/\sqrt{T}$ be the eigenvectors corresponding to the leading $\widehat{K}$ eigenvalues of $XX^\top$ and $\widehat{B} = X^\top \widehat{F}(\widehat{F}^\top \widehat{F})^{-1} = T^{-1}X^\top \widehat{F}$. Then, we project the data on the orthogonal of the vector space generated by the estimated factors. Let $\widehat{P} = T^{-1}\widehat{F}\left(\widehat{F}^\top \widehat{F}/T\right)^{-1}\widehat{F}^\top = T^{-1}\widehat{F}\widehat{F}^\top$ be the projector on the vector space generated by the columns of $\widehat{F}$. A natural estimate for U is $\widehat{U} = X - \widehat{F}\widehat{B}^\top = \left(I_T - \widehat{P}\right)X$. Similarly, we let $\widetilde{Y} = \left(I_T - \widehat{P}\right)Y$ be an estimate of $Y - F\gamma^*$. The final estimate of $2T^{-1}\left\|U^\top(Y - F\gamma^*)\right\|_\infty$ is our test statistic

$$2T^{-1}\left\|\widehat{U}^\top \widetilde{Y}\right\|_\infty. \tag{3}$$

Next, to estimate the quantiles of the distribution of $2T^{-1}\left\|U^\top \mathcal{E}\right\|_\infty$, we need an estimate of $\mathcal{E}$. We obtain it through the following LASSO estimator:

$$\widehat{\beta}_\lambda = \arg\min_{\beta \in \mathbb{R}^p} \frac{1}{T}\left\|\widetilde{Y} - \widehat{U}\beta\right\|_2^2 + \lambda\|\beta\|_1, \tag{4}$$

where $\lambda > 0$ is a penalty parameter, the choice of which will be fully data-driven in both theory and practice. For $t \in [T]$, we denote by $\widetilde{y}_t$ the t^{th} element of $\widetilde{Y}$ and $\widehat{u}_t$ as the $T \times 1$ vector corresponding to the t^{th} row of $\widehat{U}$. For a given $\lambda > 0$, let $\widehat{\varepsilon}_{\lambda,t} = \widetilde{y}_t - \widehat{u}_t^\top \widehat{\beta}_\lambda,\ t \in [T]$ be the estimate of ε_t. For a fixed $\alpha \in (0,1)$, we can then estimate q_α, the $(1-\alpha)$ quantile of the distribution of $2T^{-1}\|U^\top \mathcal{E}\|_\infty$, by the Gaussian multiplier bootstrap. Let $e = (e_1, \ldots, e_T)$ be a standard normal random vector independent of the data (X, Y) and define the criterion

$$\widehat{Q}(\lambda, e) = \left\|\frac{2}{T}\sum_{t=1}^T \widehat{u}_t \widehat{\varepsilon}_{\lambda,t} e_t\right\|_\infty.$$

The estimate $\widehat{q}_\alpha(\lambda)$ of q_α is then the $(1-\alpha)$-quantile of the distribution of $\widehat{Q}(\lambda, e)$ given X and Y. Formally, $\widehat{q}_\alpha(\lambda) = \inf\left\{q : \mathbb{P}_e(\widehat{Q}(\lambda, e) \leq q) \geq 1 - \alpha\right\}$, where $\mathbb{P}_e(\cdot) = \mathbb{P}(\cdot|X, Y)$.

The only remaining element is the procedure to select λ. We adapt the approach of Lederer & Vogt (2021) to our setting. Our choice of λ is

$$\widehat{\lambda}_\alpha = \inf\{\lambda > 0 \ : \ \widehat{q}_\alpha(\lambda') \leq \lambda' \text{ for all } \lambda' \geq \lambda\}. \tag{5}$$

We explain in Section 2.2 how to compute $\widehat{\lambda}_\alpha$ in practice. The infimum in (5) exists because for all $\lambda \geq \bar{\lambda} = 2T^{-1} \left\| \widehat{U}^\top \widetilde{Y} \right\|_\infty$, it holds that $\widehat{\beta}_\lambda = \widehat{\beta}_{\bar{\lambda}} = 0$. Moreover, since $\widehat{U}\widehat{\beta}_\lambda$ is a continuous function of λ, $\widehat{q}_\alpha(\lambda)$ is also continuous in λ and the infimum is attained at a point $\widehat{\lambda}_\alpha > 0$ such that $q_\alpha\left(\widehat{\lambda}_\alpha\right) = \widehat{\lambda}_\alpha$. Let us recall briefly the heuristics behind the choice of λ and refer the reader to Lederer & Vogt (2021) for more details. First, note that when λ is close to q_α, standard convergence bounds for the LASSO suggest that $\widehat{\beta}_\lambda$ is a precise estimate of β^*, so that $\widehat{\varepsilon}_{\lambda,t}$ is a good estimate of ε_t and, in turn, $\widehat{q}_\alpha(\lambda)$ is close to q_α. Second, when λ becomes (much) larger than q_α, the error $\widehat{\varepsilon}_{\lambda,t} - \varepsilon_t$ becomes large and dependent of $\widehat{u}_t$, which in turn increases $\widehat{q}_\alpha(\lambda)$ and leads it to be larger than q_α. We then let our estimator of q_α be $\widehat{\lambda}_\alpha = \widehat{q}_\alpha\left(\widehat{\lambda}_\alpha\right)$.

The test rejects H_0 at the level α when our test statistic is given in (3) is larger than the estimate $\widehat{\lambda}_\alpha$ of q_α. Therefore, our testing procedure is free of tuning parameters stemming from the LASSO regression in equation (4).

2.2 Computation

Algorithm 1 below explains how to conduct the test in practice. Let us discuss Step 4 of Algorithm 1 in detail. It approximates $\widehat{\lambda}_\alpha$ as defined in (5). It is advisable to set the grid size M and the number of bootstrap samples L to be as large as possible. As mentioned in Lederer & Vogt (2021), one can speed up Step 4b. by computing the LASSO with a warm start along the penalty parameter path (see, e.g., Friedman et al. (2007)), i.e., for each decreasing λ, the new coefficient estimate is computed by using the previous (i.e., the one that was computed with a larger value of λ) as a starting value. Furthermore, Step 4c. can be accelerated through parallelization techniques. In our implementation, we use both suggestions, which greatly speed up the computations. We also note that to compute the p-value of the test, it suffices to conduct it on a grid of values of α and let the p-value be equal to the largest value of α in this grid such that the test of level α rejects H_0.

1. Estimate $\widehat{K}$ by one of the available estimators of the number of factors.

2. Let the columns of $\widehat{F}/\sqrt{T}$ be the eigenvectors corresponding to the leading $\widehat{K}$ eigenvalues of $XX^\top$.

3. Compute $\widehat{U} = \left(I_T - \widehat{P}\right) X$ and $\widetilde{Y} = \left(I_T - \widehat{P}\right) Y$, where $\widehat{P} = T^{-1}\widehat{F}\widehat{F}^\top$.

4. Calculate an approximation $\widehat{\lambda}_{\alpha,emp}$ of $\widehat{\lambda}_\alpha$ as follows:

4a. Specify a grid $0 < \lambda_1 < \cdots < \lambda_M < \bar{\lambda}$, with $\bar{\lambda} = 2T^{-1}\left\|\widehat{U}^\top\widetilde{Y}\right\|_\infty$.

4b. For $m \in [M]$ compute $\left\{\widehat{Q}\left(\lambda_m, e^{(\ell)}\right) : \ \ell \in [L]\right\}$ for L draws of $e \sim \mathcal{N}(0, I_T)$ and the corresponding empirical $(1-\alpha)$-quantile $\widehat{q}_{\alpha,emp}(\lambda_m)$ from them.

4c. Let $\widehat{\lambda}_{\alpha,emp} = \widehat{q}_{\alpha,emp}(\lambda_{\widehat{m}})$, with $\widehat{m} = \min\{m \in [M] : \ \widehat{q}_{\alpha,emp}(\lambda_{m'}) \leq \lambda_{m'} \text{ for all } m' \geq m\}$.

5. Reject H_0 when $2T^{-1}\left\|\widehat{U}^\top\widetilde{Y}\right\|_\infty > \widehat{\lambda}_{\alpha,emp}$.

Algorithm 1: Conducting a test of level $\alpha \in (0,1)$.

3 Asymptotic theory

In this section, we provide the asymptotic properties of the test in a theoretical framework allowing for time series dependence in the factors and the idiosyncratic shocks and polynomial tails. We place ourselves in an asymptotic regime where T goes to infinity and p goes to infinity as a function of T. The number of factors K is fixed with T. It would be possible to let it grow, see, for instance, Beyhum & Gautier (2023). The distributions of the factors f_t and the error terms ε_t do not depend on T, while the distribution of the other variables are allowed to vary with T. All the constants we introduce are universal in the sense that they do not vary with the sample size. Our assumptions are similar to that of Fan et al. (2023) but significantly weaker than that of Fan et al. (2024), which imposes that the variables are i.i.d. sub-Gaussian.

We introduce further notation. The loading b_{jk} corresponds to the j^{th} element of the k^{th} column of B. Let also $b_j = (b_{j1}, \dots, b_{jK})^\top$. For $t \in [T]$, let

$$z_t = \left(u_t^\top, f_t^\top, \varepsilon_t, \frac{1}{\sqrt{p}}\sum_{\ell=1}^{p} u_{t\ell} b_\ell^\top\right)^\top.$$

Finally, define $\Sigma = \mathbb{E}[u_t u_t^\top]$.

We make the following assumptions.

Assumption 1 *The estimator* $\widehat{K}$ *is such that* $\mathbb{P}(\widehat{K} = K) \to 1$.

Assumption 2 *It holds that*

(*i*) *For all* $t \in [T]$, $\mathbb{E}[f_t] = 0$, $\mathbb{E}[f_t f_t^\top] = I_K$ *and* $B^\top B$ *is diagonal;*

(*ii*) *All the eigenvalues of the* $K \times K$ *matrix* $p^{-1}B^\top B$ *are bounded away from* 0 *and* ∞ *as* $p \to \infty$*;*

(*iii*) $\|\Sigma - BB^\top\|_2 = O(1)$*;*

(*iv*) $\|B\|_\infty = O(1)$.

Assumption 3 *The following holds:*

(*i*) *The process* $\{z_t\}_t$ *is weakly stationary. Moreover, it holds that*

$$\mathbb{E}[u_{tj}] = \mathbb{E}[u_{tj} f_{tk}] = 0,$$

for all $t \in [T], j \in [p], k \in [K]$.

(*ii*) *There exist* $\kappa_1, \kappa_2 > 0$ *such that, for all* $t \in [T]$, $\sigma_p(\mathbb{E}\left[\varepsilon_t^2 u_t u_t^\top\right]) > \kappa_1$, $\left\|\mathbb{E}\left[\varepsilon_t^2 u_t u_t^\top\right]\right\|_\infty < \kappa_2$, $\sigma_p(\Sigma) > \kappa_1$, $\max_{j\in[p]} \sum_{\ell=1}^{p} |\Sigma_{j\ell}| < \kappa_2$*;*

(*iii*) *There exist* $q \geq 8$, $C_1 > 0$ *and* $\zeta > 0$, *such that for* $s, t \in [T]$, *we have*

$$\left\| z_t \right\|_{q+\zeta} < C_1;$$
$$\left\| p^{-1/2}\left(u_s^\top u_t - \mathbb{E}\left[u_s^\top u_t\right]\right) \right\|_q < C_1;$$

(*iv*) $\{u_t \varepsilon_t\}_t$ *is uncorrelated across* t, *and, for all* $t \in [T], j \in [p], k \in [K]$,

$$\mathbb{E}[u_{tj}\varepsilon_t] = \mathbb{E}[f_{tk}\varepsilon_t] = 0.$$

Assumption 4 *Let* $\tilde{\alpha}$ *denote the strong mixing coefficients of* $\{z_t\}_t$. *There exist* $C_2, c > 0$ *such that* $c > \left[\left(\frac{h+\xi}{\xi}\right)\left(\frac{h}{2} - 1\right)\right] \vee \left(\frac{2}{1-\frac{2}{h}}\right)$ *and* $\kappa = \left(\frac{\frac{1}{2}+\frac{h}{4(h+1)}}{c+\frac{h}{2(h+1)}}\right) < \frac{1}{2}$, *where* $h = \frac{q}{2}$ *and* $\xi = \frac{\zeta}{2}$, *and, for all* $t \in \mathbb{Z}_+$, *we have* $\tilde{\alpha}(t) \leq C_2 t^{-c}$.

Assumption 1 means that the estimator $\widehat{K}$ of the number of factors K is consistent. Examples of $\widehat{K}$ and sufficient conditions for its consistency can be found in Bai & Ng (2002), Onatski (2010), Ahn & Horenstein (2013), Bai & Ng (2019), Fan, Guo & Zheng (2022). Assumption 2 is the same as Assumption 3 in Fan et al. (2023). Its conditions (i) and (ii) constitute a strong factor assumption (Bai (2003)). Assumption 3 restricts the moments and the tail behavior of the variables. Assumption 3 (i),(ii),(iv) contain conditions on the moments of the different variables similar to that of the literature (Bai & Ng (2006), Fan et al. (2013)). We assume that the variables in Assumption 3 (iii) have polynomial tails with common parameter $q + \zeta$. It would be possible to have different tail parameters for each variable but we avoid doing so to simplify our presentation. As in Fan et al. (2023) the number of finite moments is at least 8. In the similar context of inference on factor regression models, Assumption E.2. in Bai & Ng (2006) and Assumption 7 in Gonçalves & Perron (2014) impose conditions analogous to the restriction that $\{u_t \varepsilon_t\}_t$ is uncorrelated across t in Assumption 3 (iv). A sufficient condition for the latter restriction is that $\{\varepsilon_t\}_t$ is uncorrelated across t and independent of $\{u_t\}_t$. Note that this assumption could be avoided by using a block bootstrap method, but this would complicate the test. It may also not be justified because (most of) the serial correlation in the data may be picked up by the factors f_t and not by the error term ε_t. Assumption 4 means that the process $\{z_t\}_t$ has strong mixing coefficients decaying polynomially, which is a restriction on the time-series dependence of the variables. A similar assumption is made in Fan et al. (2023). Note that Assumption 4 restricts the full distribution of the process $\{z_t\}_t$, while Assumption 3 (iv) just imposes a condition on a particular serial correlation.

Let us introduce $\varphi^* = \gamma^* - B^\top \beta^*$. To interpret φ^*, note that the first equation of (1) can be rewritten $y_t = f_t^\top \varphi^* + x_t^\top \beta^* + \varepsilon_t$, which becomes a usual high-dimensional sparse regression model when $\varphi^* = 0$. The next assumption concerns the relative growth rates of $T, p, \|\varphi^*\|_2$ and $\|\beta^*\|_1$.

Assumption 5 *The following holds:*

(i) $p = O(T^r)$, *where* $r < \left(\frac{h}{4} - \frac{1}{2}\right) \wedge \left[(h+1)(c\kappa - \frac{1}{2})\right]$;

(ii) $\frac{\log(T \vee p)^5}{\sqrt{T}} \|\beta^*\|_1 = o(1)$;

(*iii*) $\log(T \vee p)^{5/2} \frac{\sqrt{T}}{T \wedge p}(\|\varphi^*\|_2 \vee 1) = o(1)$.

Condition (i) restricts the rate at which p can grow with respect to T. We allow it to grow at a polynomial rate, which is more restrictive than the standard restriction for the LASSO (exponential rate). We need to be more restrictive because of the presence of polynomial tails and time series dependence. Assumption 5 (ii) contains sparsity restrictions on the alternative hypotheses. When $\|\beta^*\|_\infty = O(1)$, condition (ii) corresponds, up to logarithmic factors, to the standard consistency condition for the LASSO with bounded regressors and errors with sub-Gaussian tails that is $\sqrt{\log(p)/T}(s_0 \vee 1) = o(1)$, where s_0 is the number of nonzero coefficients of β^*. We can impose a similar sparsity condition as in the standard LASSO literature with sub-Gaussian errors, despite dealing with polynomial tails and serial correlation, because we utilize the high-dimensional central limit theorem for polynomial-tailed time series from Fan et al. (2023). Under the rate condition specified in Assumption 5 (i), this theorem allows us to essentially revert to the scenario with sub-Gaussian errors. Condition (iii) is a slightly more restrictive version of the standard condition that $\sqrt{T}/(T \wedge p) = o(1)$ for inference in the factor regression model.[2] Indeed, since φ^* is of size K, it is reasonable to assume that $\|\varphi^*\|_2 = O(1)$. Under this condition, (iii) corresponds to $\sqrt{T}/(T \wedge p) = o(1)$ up to logarithmic factors. As noted by a referee, the role of (iii) is to ensure that the error coming from the estimation of the factors is asymptotically negligible (see for instance Bai & Ng (2006)). It may be possible to relax it using a more elaborate bootstrap scheme, see Gonçalves & Perron (2014). Additionally, it is worth noting that our proofs reveal that Assumption 5 is stronger than necessary, and the validity of the test could be established under more complex but weaker rate conditions. However, for the sake of clarity, we present Assumption 5 instead of a more intricate condition.

We have the following theorem.

Theorem 1 *Let Assumptions 1, 2, 3, 4 and 5 hold. For all $\alpha \in (0,1)$, we have*

(*i*) *If $\beta^* = 0$, then* $\mathbb{P}\left(T^{-1}\left\|\widehat{U}^\top \widetilde{Y}\right\|_\infty > \widehat{\lambda}_\alpha\right) \leq \alpha + o(1)$.

(*ii*) *If* $\sqrt{\frac{\log(T \vee p)}{T \wedge p}} = o_P\left(T^{-1}\left\|U^\top U \beta^*\right\|_\infty\right)$, *then* $\mathbb{P}\left(T^{-1}\left\|\widehat{U}^\top \widetilde{Y}\right\|_\infty > \widehat{\lambda}_\alpha\right) \to 1$.

[2] This condition is equivalently stated as $\sqrt{T}/p = o(1)$ in Bai & Ng (2006), Corradi & Swanson (2014) and many others.

The proof of Theorem 1 can be found in Online Appendix D. Statement (i) means that the empirical size of the test tends to the nominal size. Statement (ii) shows that the test has asymptotic power equal to 1 against sequences of alternatives such that $\sqrt{\frac{\log(T \vee p)}{T \wedge p}} = o_P\left(T^{-1}\left\|U^\top U \beta^*\right\|_\infty\right)$. As noted in Lederer & Vogt (2021), such a condition is inevitable because the presence of the error ε_t prevents us from distinguishing true $U\beta^*$ and ε_t when $U\beta^*$ is too small. We discuss further this condition in Section D.8 of the Online Appendix.

4 Monte Carlo simulations

In this section, we provide a Monte Carlo study shedding light on the finite sample performance of our proposed testing procedure. We use the following model as our data-generating process (DGP):

$$y_t = f_t^\top \gamma^* + u_t^\top \beta^* + \varepsilon_t,$$
$$x_t = B f_t + u_t, \quad t = 1 \dots, T.$$

We generate samples with $T = \{200, 400\}$ observations, $p = \{T, 5T\}$ variables and $K = 2$ factors. The loadings $B = \{b_{jk}\}_{j\in[p],k\in[K]}$ are such that $b_{jk} \sim \mathcal{U}[-1, 1]$. The factors are generated as $f_t = \rho_f f_{t-1} + \tilde{f}_t$ for $t = 2, \dots, T$, where $\tilde{f}_t$ are i.i.d. $\mathcal{N}\left(0, I_K\left(1 - \rho_f^2\right)\right)$. The idiosyncratic components $\{u_t\}$ are such that $u_t = \rho_u u_{t-1} + \tilde{u}_t$ for $t = 2, \dots, T$, where $\tilde{u}_t$ are i.i.d. $\mathcal{N}\left(0, \Sigma\left(1 - \rho_u^2\right)\right)$, with $\Sigma_{ij} = c_u^{|i-j|}, i, j \in [p]$., where c_u reflects the amount of cross-sectional dependence. We also let $\varepsilon_t = \rho_e \varepsilon_{t-1} + \tilde{\varepsilon}_t$ for $t = 2, \dots, T$, where $\tilde{\varepsilon}_t$ are i.i.d. $\mathcal{N}\left(0, \left(1 - \rho_e^2\right)\right)$.

The parameters ρ_f, ρ_u, and ρ_e control the level of time series dependence. The stationary distributions of f_t, u_t, ε_t are, respectively, $\mathcal{N}(0, I_K)$, $\mathcal{N}(0, \Sigma)$ and $\mathcal{N}(0, 1)$. We initialize f_0, u_0 and ε_0 as such. We consider three dependency designs, where we vary the cross-sectional dependence via parameter s and the time series dependence via parameters (ρ_f, ρ_u, ρ_e) as follows:

Design 1. $c_u = \rho_f = \rho_u = \rho_e = 0$, so that the data are i.i.d. across j and t.

Design 2. $c_u = 0.1$, $\rho_f = 0.6$, $\rho_u = 0.1$ and $\rho_e = 0$, which introduces cross-sectional dependence in the idiosyncratic shocks and time-series dependence in the factors and the idiosyncratic shocks.

Design 3. $c_u = 0.1$, $\rho_f = 0.6$ and $\rho_u = \rho_e = 0.1$, where, on top of the cross-sectional dependence, there is time series dependence in the factors, the idiosyncratic shocks and the error terms.

Our theory does not formally allow the third design, but we want to show that our test performs well even under weak serial correlation of $\{\varepsilon_t\}_t$.

Finally, we consider two cases for the target parameter β^*. For the first case we set $\beta^* = (1, 0, \ldots, 0)^\top \times m$, where $m \in \{0, 0.1, 0.2, 0.3, 0.4\}$ controls the signal strength. This choice of β^* corresponds to a *sparse* design. For the second case, we consider $\beta^* = (1/\sqrt{p}, \ldots, 1/\sqrt{p})^\top \times m$, which corresponds to *dense* design. In both cases, we set $\gamma^* = (0.5, 0.5)^\top$. Note that the choice of β^* ensures that for the dependence Design 1, the signal-to-noise ratio is the same for both sparse and dense cases.

We compute the rejection probabilities of our test at the significance levels $\alpha \in \{0.1, 0.05, 0.01\}$ over 2000 replications. To implement our test, we set $M = 100$ and choose an equidistant grid of values for λ, using $L = 1000$ bootstrap replications. The results are insensitive to the choice of L and M as long as they are sufficiently large, which is expected since their primary role is in the approximation of theoretical quantities. In our experience, $L = M = 100$ already yields very precise results. The number of factors K is estimated through the eigenvalue ratio estimator of Ahn & Horenstein (2013).

4.1 Main results

The results for $T = 200$ and sparse and dense alternatives are reported in Tables 1-2. In the Online Appendix, we present simulations under the same data-generating processes, but with the larger sample size $T = 400$ (Tables OA.1 and OA.2). First, we observe that the empirical size is close to the nominal levels for all designs, indicating that both cross-sectional and serial dependence have little effect on the empirical size of our testing procedure. Notably, the results show consistent power when the alternative hypothesis is sparse (Table 1) across all data-generating processes (DGPs), suggesting that the dependency design does not significantly impact the power of our test. However, the power tends to slightly decrease in scenarios with large p cases for both $T = 200$ and $T = 400$. Interestingly, in the case of dense alternatives (Table 2), we see a significant drop (relative to sparse alternatives) in power across all designs. This highlights that our testing procedure has low power when $\beta^* \neq 0$ is dense. An intuition for this result is as follows. Our test uses the LASSO estimator, which enforces sparsity. When β^* is dense, the LASSO estimator often sets $\widehat{\beta}_{\widehat{\lambda}_\alpha}$ to 0, leading to non-rejection of the null.

	p/T = 200/200			p/T = 1000/200		
m	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$
			Design 1: $s = \rho_f = \rho_u = \rho_e = 0$			
0.0	0.084	0.033	0.005	0.090	0.040	0.005
0.1	0.122	0.062	0.015	0.108	0.049	0.010
0.2	0.604	0.507	0.336	0.274	0.184	0.082
0.3	0.984	0.973	0.916	0.704	0.614	0.418
0.4	1.000	1.000	1.000	0.958	0.927	0.833
			Design 2: $s = 0.1$, $\rho_f = 0.6$, $\rho_u = 0.1$ and $\rho_e = 0$			
0.0	0.092	0.045	0.010	0.090	0.040	0.004
0.1	0.122	0.066	0.016	0.104	0.048	0.009
0.2	0.631	0.534	0.362	0.274	0.184	0.080
0.3	0.984	0.971	0.931	0.731	0.633	0.435
0.4	1.000	1.000	1.000	0.961	0.931	0.853
			Design 3: $s = 0.1$, $\rho_f = 0.6$ and $\rho_u = \rho_e = 0.1$			
0.0	0.102	0.048	0.011	0.099	0.044	0.004
0.1	0.130	0.070	0.017	0.108	0.054	0.009
0.2	0.620	0.528	0.356	0.273	0.186	0.084
0.3	0.980	0.970	0.925	0.723	0.631	0.429
0.4	1.000	1.000	1.000	0.962	0.933	0.849

Table 1: Rejection probabilities for the three dependence designs we consider and sparse β^*. The data are generated with Gaussian variables. The sample size is $T = 200$ while the number of regressors is $p \in \{200, 1000\}$.

	p/T = 200/200			p/T = 1000/200		
m	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$
			Design 1: $s = \rho_f = \rho_u = \rho_e = 0$			
0.0	0.084	0.033	0.005	0.090	0.040	0.005
0.1	0.097	0.039	0.008	0.092	0.035	0.009
0.2	0.114	0.060	0.007	0.100	0.045	0.009
0.3	0.140	0.074	0.011	0.124	0.055	0.010
0.4	0.176	0.096	0.017	0.152	0.068	0.015
			Design 2: $s = 0.1$, $\rho_f = 0.6$, $\rho_u = 0.1$ and $\rho_e = 0$			
0.0	0.084	0.040	0.006	0.081	0.032	0.007
0.1	0.085	0.041	0.009	0.091	0.034	0.005
0.2	0.108	0.051	0.012	0.114	0.051	0.008
0.3	0.155	0.072	0.015	0.153	0.072	0.011
0.4	0.210	0.109	0.023	0.199	0.100	0.018
			Design 3: $s = 0.1$, $\rho_f = 0.6$ and $\rho_u = \rho_e = 0.1$			
0.0	0.094	0.046	0.005	0.091	0.035	0.006
0.1	0.100	0.046	0.010	0.104	0.044	0.006
0.2	0.121	0.060	0.012	0.124	0.055	0.009
0.3	0.171	0.087	0.018	0.168	0.080	0.012
0.4	0.228	0.116	0.025	0.222	0.108	0.020

Table 2: Rejection probabilities for the three dependence designs we consider and dense β^*. The data are generated with Gaussian variables. The sample size is $T = 200$ while the number of regressors is $p \in \{200, 1000\}$.

4.2 Heavy-tailed data, number of factors and lagged idiosyncratic shocks

In this subsection, we consider the same data-generating process as in the main design Table 1, but we change elements of it to obtain results for heavy-tailed data and in cases when the number of factors is over- and under-estimated. For the heavy-tailed data scenarios, we generate factors, idiosyncratic shocks and regression errors from student-$t(5)$ distribution. This allows us to investigate the impact of the heavy tails on the proposed testing procedure. Note that our theoretical analysis imposes a condition for the existence of first 8 finite moments, so that generating student-$t(5)$ errors goes beyond what our assumptions allow. We generate heavy-tailed factors f_t, idiosyncratic components u_t and the error terms ϵ_t by generating $\tilde{f}_t$, $\tilde{u}_t$ and $\tilde{\epsilon}_t$ from a student-$t(5)$ rather than a Gaussian distribution. Besides changing the distribution of the data, we also compute the performance of our test when the number of factors is either over-estimated or under-estimated. In this case, our data generating process is the same as for Table 1, but rather than estimating the number of factors K using the eigenvalue ratio estimator, we set it to $K = 5$ (over-estimated case) and $K = 1$ (under-estimated case).

Results are reported in Tables OA.3-OA.4 and OA.5-OA.8 for heavy-tailed data and different number of factors, respectively. In the former case, we see a slight deterioration of our method compared to the Gaussian DGPs. This is not surprising since the residuals of the LASSO regression, as well as the factors, may be less accurately estimated in finite samples due to heavy tails, see, e.g., Babii et al. (2024). Next, we analyze the results of the case of the over-estimated number of factors. The performance is slightly affected compared to the case where we use the eigenvalue ratio estimator to determine the number of factors (see Table 1), but the differences are small. When comparing with the under-estimated case, we see that our testing procedure is over-sized. These results are in line with the literature on inference with factors models, see, e.g., Moon & Weidner (2015).

It is also interesting to investigate the performance of our testing procedure with lagged idiosyncratic shocks. In this case, we generate the data from the following model

$$\begin{aligned} y_t &= f_t^\top \gamma^* + u_t^\top \beta_1^* + u_{t-1}^\top \beta_2^* + \varepsilon_t, \\ x_t &= B f_t + u_t, \quad t \in [T], \end{aligned} \tag{6}$$

where all the elements are as in the main DGPs except that we add lagged idiosyncratic shock, set $\beta_1^* = \beta^* = (1, 0, \ldots, 0)^\top \times m$ and all elements of $\beta_2^* = 0$. The algorithm to compute the test for this model appears in the Online Appendix OA.1. We report results for sparse and Gaussian DGPs which appear in Online Appendix Tables OA.9-OA.10. Results show a slight deterioration of performance in terms of power, while the empirical size appears to be similar as in the main scenario Table 1 (see also Table OA.1 in Online Appendix for a large T comparison). The small decrease in power is due to an increase in the dimensionality of the regressors.

5 Empirical applications

In this section, we present three empirical examples using well-established macroeconomic and financial datasets.

First, we examine the FRED-MD dataset, which includes 121 monthly macroeconomic series covering various sectors of the US economy. For further details, see McCracken & Ng (2016). We also study a quarterly variant of FRED-MD with 202 variables, named FRED-QD, for which results are reported in the Online Appendix. Second, we analyze a financial dataset comprising 100 representative anomalies from the literature, which we use to explain aggregate market returns and industry portfolio returns. For more information, we refer to Dong et al. (2022). Lastly, we investigate the network structure in asset returns using a finance dataset. Following the approach of Fan et al. (2023), we consider a cross-section of monthly stock returns and a set of observable factors to study the relationships among financial firms.

We study these datasets for several reasons. First, FRED-MD and its quarterly variant FRED-QD are among the most widely studied high-dimensional macroeconomic datasets. Providing further insights into these datasets could be valuable for the extensive empirical macroeconomic literature. Second, the empirical asset pricing literature has proposed and examined many factors, known as anomalies, that explain asset prices. Our research offers new insights into applying long-short anomalies to explain the market and industry portfolio returns rather than individual asset returns. Third, applying our method to firm-level stock returns illustrates using a model with observed regressors, as detailed in Appendix Section A. Overall, these diverse applications, using macroeconomic and financial datasets, demonstrate the versatility of our approach across different econometric settings, varying in sample sizes relative to the number of regressors. Lastly, we note that in all empirical applications, we select the number of factors using the eigenvalue ratio estimator of Ahn & Horenstein (2013).

5.1 Macroeconomic application

In the macroeconomic application, we investigate the FRED-MD dataset over the sample period 1980 January to 2019 December containing $T = 480$ observations. This is a commonly used dataset in various macroeconomic studies. In our analysis, we regress each of the 121 variables available in the dataset on common and idiosyncratic shocks. Specifically, we estimate the following model:

$$\begin{aligned} y_{t+1} &= f_t^\top \gamma^* + u_t^\top \beta^* + \varepsilon_{t+1}, \\ x_t &= B f_t + u_t, \quad t = 1 \ldots, T, \end{aligned} \tag{7}$$

where y_{t+1} represents one of the variables in the dataset at time $t+1$, and x_t includes all the remaining regressors at time t. Note that we transform the original series using commonly applied transformations as suggested by McCracken & Ng (2016). To estimate f_t, we apply PCA using the eigenvalue ratio estimator to determine the number of common factors. We apply our procedure to test H_0. Results for each category of variable appear in Table 3.

First, results show that in most categories, we frequently reject the null hypothesis at the 10%, 5%, and even 1% significance levels. This provides evidence of sparsity in the idiosyncratic shocks of the macroeconomic data. The categories where we reject the null most often are *Output and Income*, *Consumption, Orders, and Inventories*, *Labor Market*, *Interest and Exchange Rates*, and *Prices*. For the remaining categories, the null hypothesis is not frequently rejected, suggesting that the sparse component is not important for those series. Interestingly, for the housing category, we never reject the null, while the other two categories for which we reject less frequently, i.e., *Money and Credit* and *Stock Market*, could be classified as financial rather than macroeconomic data. In addition to the results for the FRED-MD dataset, we also obtain similar rejection ratios for the FRED-QD dataset, which is a quarterly macroeconomic dataset similar to FRED-MD, but containing more variables and less observations. Results appear in Table OA.11. The results for the FRED-QD dataset also show a similar pattern, providing evidence of sparsity in the idiosyncratic shocks.

Category	10%	5%	1%
Output and Income (16)	0.500	0.438	0.125
Consumption, Orders, and Inventories (9)	1.000	0.778	0.222
Labor Market (31)	0.774	0.677	0.419
Housing (10)	0.000	0.000	0.000
Money and Credit (12)	0.250	0.167	0.083
Stock Market (3)	0.333	0.000	0.000
Interest and Exchange Rates (19)	0.789	0.737	0.474
Prices (20)	0.800	0.650	0.450

Table 3: Rejection ratios for each category of the FRED-MD dataset over the sample period 1980 January to 2019 December. In parentheses, we report the number of series per category. Data source: McCracken & Ng (2016).

5.2 Finance application I

In our second application, we test the sparse component of a financial dataset comprising a set of regressor variables shown to predict market excess returns (Dong et al. 2022). Specifically, we examine whether the idiosyncratic sparse component is important in explaining market excess returns as well as returns of 49 industry portfolios when regressing on a representative sample of 100 long-short anomaly portfolio returns. The data spans from January 1970 to December 2017, therefore, the effective sample size is $T = 575$. We estimate the same model as in (7).

Dong et al. (2022) have shown that these regressors lead to accurate forecasts of stock market excess return by employing various machine learning and forecast combination methods. For more details on the data and a full list of target variables, see Section C.2 of the Online Appendix. Results are reported in Table 4.

Our results indicate that for more than 50% of 49 industry returns, the idiosyncratic shocks are significant at the 10% level, meaning we reject the null hypothesis at the 10% level. For the aggregate market returns, the p-value is 0.042, providing evidence in favor of the sparse component. Additionally, we reject the null at the 1% significance level for

Market**	0.042	**Clths*****	0.001	FabPr	0.272	**Oil****	0.014	**Boxes****	0.015
Agric	0.184	Hlth	0.500	**Mach****	0.012	Util	0.252	**Trans****	0.024
Food**	0.012	MedEq	0.198	**ElcEq****	0.034	Telcm	0.625	Whlsl	0.117
Soda	0.129	Drugs	0.286	**Autos*****	0.007	**PerSv****	0.043	**Rtail***	0.099
Beer	0.150	**Chems****	0.014	**Aero*****	0.006	**BusSv***	0.051	**Meals****	0.043
Smoke	0.130	**Rubbr*****	0.002	**Ships****	0.034	Hardw	0.636	**Banks****	0.011
Toys	0.205	**Txtls****	0.023	Guns	0.173	Softw	0.206	**Insur****	0.024
Fun*	0.057	**BldMt*****	0.004	Gold	0.504	Chips	0.646	**RlEst*****	0.003
Books***	0.001	Cnstr	0.123	**Mines*****	0.005	LabEq	0.618	Fin	0.195
Hshld	0.189	Steel	0.278	Coal	0.268	**Paper*****	0.001	**Other****	0.010

Table 4: p-Values of market and industry excess returns regressed on 100 long-short anomaly characteristics. Bold entries with *, **, and *** indicate significance at 10%, 5% and 1% significance level, respectively.

nine industry portfolios: *Apparel* (Clths), *Automobiles and Trucks* (Autos), *Aircraft* (Aero), *Rubber and Plastic Products* (Rubbr), *Construction Materials* (BldMt), *Real Estate* (RlEst), *Printing and Publishing* (Books), *Non-Metallic and Industrial Metal Mining* (Mines), and *Business Supplies* (Paper). Furthermore, we analyze returns of 10 industry portfolios which are based on a broader classification compared to 49 industries. Results are presented in Table OA.13 in the Online Appendix, which confirm a similar pattern. Therefore, we find strong evidence supporting the presence of sparse idiosyncratic components in equity returns for a variety of industries and the aggregate market.

5.3 Finance application II

In the third empirical application, we examine the sparse idiosyncratic components of individual stock returns using the dataset from Jensen et al. (2023). Specifically, we use monthly stock return data for a sample of 721 firms with no missing data from January 1991 to December 2022, resulting in $T = 384$ and $p = 721$ (the regressors x_{it} are the returns of other firms). We test a model similar to Fan et al. (2023) where our test can be seen as a diagnostic check after the third step in the approach put forward by Fan et al. (2023). However, our analysis diverges from Fan et al. (2023) by focusing on a balanced panel of firms. We select firms for which we have a complete time series of returns and observed regressors. Furthermore, in our industry or sector analyses, we group firms into 49 industries — the same as in finance I application —, whereas Fan et al. (2023) use a different firm classification.

Specifically, denote $y_t^{(i)}$ the stock excess return of firm i at time t. We regress the firm's excess return on a set of observable factors — see Table OA.15 for the list of factors — as well as common and idiosyncratic components stemming from all other returns. The model is thus an example of the extension of the main model with additional covariates, see Online Appendix A. Denoting the returns of all firms except the i^{th} firm as $x_{it} = (y_{jt})_{j\in[n]/i}$, where n is the total number of firms, we report the rejection ratios β^* by testing $\beta^* = 0$ in the following model for each firm i:

$$\begin{aligned} y_t^{(i)} &= f_t^\top \gamma^* + w_t^\top \delta^* + u_{it}^\top \beta^* + \varepsilon_t, \\ x_{it} &= B f_{it} + u_{it}, \quad t = 1\ldots, T, \end{aligned} \tag{8}$$

where w_t denotes the observable factors.

Results are reported in Table 5. First, we see that the rejection ratios (the proportion of firms i for which we reject the null) are relatively low and only slightly above the nominal significance levels. This suggests that the sparse component is much less significant in individual stock returns compared to other applications we considered, i.e., we find weak evidence of the presence of sparse idiosyncratic shocks in firm-level returns. To gain further insight, we report results by grouping firms into 49 industries, as reported in Table OA.16 in the Online Appendix. We find that for most industries, the rejection ratios are small. However, a few industries exhibit a larger proportion of rejections, notably *Precious Metals* (Gold) and *Communications* (Telcm).

	10 %	5 %	1 %
Rejection rates	0.130	0.082	0.038

Table 5: Rejection rates for the firm-level financial returns dataset.

6 Conclusion

This paper proposes a new bootstrap test for the adequacy of the factor regression model against factor-augmented sparse alternatives. We establish the asymptotic validity of our test under time series dependence and polynomial tails. In a Monte Carlo study, we show that our procedure has excellent finite sample properties against sparse alternatives but low power against dense alternatives. We often reject the null when we apply our testing procedure to standard datasets in macroeconomics and finance. This suggests that sparsity is present - on top of a dense model - in several economic environments.

This message complements previous studies. Indeed, based on different approaches, Giannone et al. (2021), Kolesár et al. (2023) found evidence against the presence of sparsity in several economic datasets, comparing only sparse and only dense models. Our analysis instead suggests that sparsity may still have a role to play on top of a dense component. These findings also constitute arguments in favor of using sparse plus dense models, which have recently gathered a lot of interest in the econometrics literature.

A potential limitation of our analysis is that we may reject H_0 because β^* is nonzero but dense. In our Monte-Carlo simulations, we compare DGPs with sparse and dense β^* with the same signal-to-noise ratio and find that our test has high power against the sparse alternative but low power against dense deviations from the null. This finding should mitigate the previous concern that rejection might be due to dense $\beta^* \neq 0$. A complementary approach free of this concern is testing sparsity directly. Kolesár et al. (2023) takes this road by proposing a test for the null hypothesis of sparsity in a high-dimensional regression model (without factors). The limitations of their approach are that it is only valid when the number of variables is smaller than the sample size and that it does not reject the null when the regression coefficient is equal to 0. This is a problem because, in our context, we would not conclude in favor of the existence of sparsity when $\beta^* = 0$. For future research, it would be interesting to adapt the test of Kolesár et al. (2023) to the factor-augmented regression model and apply it to the datasets we consider in the present paper.

Appendix: extension to additional regressors

A A model with additional regressors

As in Stock & Watson (2002), Bai & Ng (2006), Lederer & Vogt (2021), in empirical applications we can augment the model with additional observed low-dimensional regressors $w_1, \dots, w_t \in \mathbb{R}^\ell$ (where ℓ is fixed with T). We therefore can consider the alternative model.

$$\begin{aligned} y_t &= f_t^\top \gamma^* + w_t^\top \delta^* + u_t^\top \beta^* + \varepsilon_t, \\ x_t &= B f_t + u_t, \quad t = 1 \dots, T, \end{aligned} \tag{9}$$

Here, again, $\varepsilon_t \in \mathbb{R}$ represents a random error, u_t is a p-dimensional random vector of idiosyncratic shocks, f_t is a K-dimensional random vector of factors, and B is a $p \times K$ random matrix of loadings. The parameters are $\gamma^* \in \mathbb{R}^K$, $\delta^* \in \mathbb{R}^\ell$, $\beta^* \in \mathbb{R}^p$. Note that here w_t plays the role of an observed factor (with loading equal to 0). This will be key to understanding the alternative testing procedure of Section B.

We focus on testing

$$H_0 : \beta^* = 0 \quad \text{against} \quad H_1 : \beta^* \neq 0. \tag{10}$$

To facilitate understanding, we again rewrite the model in matrix form as follows:

$$\begin{aligned} Y &= F^\top \gamma^* + W \delta^* + U^\top \beta^* + \mathcal{E}, \\ X &= BF + U, \end{aligned}$$

where $Y = (y_1, \dots, y_T)^\top$, $F = (f_1, \dots, f_T)^\top$ is a $T \times K$ matrix, $U = (u_1, \dots, u_T)^\top$, $W = (w_1, \dots, w_T)^\top$ and $X = (x_1, \dots, x_T)^\top$ are $T \times p$ matrices and $\mathcal{E} = (\varepsilon_1, \dots, \varepsilon_T)^\top$.

B Testing procedure of the extended model

Algorithm 2 present the test in this extended model. It is similar to Algorithm 1. The only difference is that $\widehat{P}$ is now the projector on the columns of the $T \times (\widehat{K} + \ell)$ matrix $(\widehat{F} \; W)$ in Step 3. Essentially, w_t is treated as an observed factor.

1. Estimate $\widehat{K}$ by one of the available estimators of the number of factors.

2. Let the columns of $\widehat{F}/\sqrt{T}$ be the eigenvectors corresponding to the leading $\widehat{K}$ eigenvalues of $XX^\top$.

3. Compute $\widehat{U} = \left(I_T - \widehat{P}\right) X$ and $\widetilde{Y} = \left(I_T - \widehat{P}\right) Y$, where $\widehat{P}$ is the projector on the columns of the $T \times \left(\widehat{K} + \ell\right)$ matrix $\left(\widehat{F}\ W\right)$. Denote by $\widehat{u}_t$ the $T \times 1$ vector corresponding to the transpose of the t^{th} row of $\widehat{U}$.

4. Calculate an approximation $\widehat{\lambda}_{\alpha,emp}$ of $\widehat{\lambda}_\alpha$ as follows:

4a. Specify a grid $0 < \lambda_1 < \cdots < \lambda_M < \bar{\lambda}$, with $\bar{\lambda} = 2T^{-1}\left\|\widehat{U}^\top \widetilde{Y}\right\|_\infty$.

4b. For $\lambda > 0$, and $e \in \mathbb{R}^T$, let $\widehat{Q}(\lambda, e) = \left\|\frac{2}{T}\sum_{t=1}^T \widehat{u}_t \widehat{\varepsilon}_{\lambda,t} e_t,\right\|_\infty$, where $\widehat{\varepsilon}_{\lambda,t} = \widetilde{y}_t - \widehat{u}_t^\top \widehat{\beta}_\lambda,\ t \in [T]$, for $\widehat{\beta}_\lambda = \arg\min_{\beta \in \mathbb{R}^p} \frac{1}{T}\left\|\widetilde{Y} - \widehat{U}\beta\right\|_2^2 + \lambda\|\beta\|_1$. For $m \in [M]$, compute $\left\{\widehat{Q}\left(\lambda_m, e^{(\ell)}\right) :\ \ell \in [L]\right\}$ for L draws of $e \sim \mathcal{N}(0, I_T)$ and the corresponding empirical $(1-\alpha)$-quantile $\widehat{q}_{\alpha,emp}(\lambda_m)$ from them.

4c. Let $\widehat{\lambda}_{\alpha,emp} = \widehat{q}_{\alpha,emp}(\lambda_{\widehat{m}})$, with $\widehat{m} = \min\{m \in [M] :\ \widehat{q}_{\alpha,emp}(\lambda_{m'}) \leq \lambda_{m'} \text{ for all } m' \geq m\}$.

5. Reject H_0 when $2T^{-1}\left\|\widehat{U}^\top \widetilde{Y}\right\|_\infty > \widehat{\lambda}_{\alpha,emp}$.

Algorithm 2: Conducting a test of level $\alpha \in (0, 1)$ with additional regressors.

Supplementary material

Online Appendix: Additional empirical and simulation results, details on the data and the proof of Theorem 1 (.pdf file).

R package: the R package `'FAS'` that implements our test is available on CRAN: https://cran.r-project.org/web/packages/FAS/index.html.

Acknowledgments

The authors thank the Associate Editor and two anonymous referees as well as Andrii Babii, Otilia Boldea, Peter Boswijk, Geert Dhaene, Artūras Juodis, Frank Kleibergen, Lasse Heje Pedersen, seminar participants at Copenhagen Business School and BI Norwegian Business School, and conference attendants at the FinEML conference 2023 at Erasmus University (Rotterdam), the Leuven Statistics Days 2023, the Conference in Financial Econometrics (CFE) 2023 in Berlin, and the Netherlands Econometrics Study Group 2024 in Maastricht University. The authors also thank Michael Vogt for sharing his code for the procedure of Lederer & Vogt (2021). Jad Beyhum undertook part of this work while employed by CREST, ENSAI (Rennes). Jad Beyhum gratefully acknowledges financial support from the Research Fund KU Leuven through the grant STG/23/014. Jonas Striaukas gratefully acknowledges the financial support from the European Commission, MSCA-2022-PF Individual Fellowship, Project 101103508. Project Acronym: MACROML.

References

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

Babii, A., Ghysels, E. & Striaukas, J. (2024), 'High-dimensional granger causality tests with an application to vix and news', *Journal of Financial Econometrics* **22**(3), 605–635.

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

Bai, J. & Ng, S. (2002), 'Determining the number of factors in approximate factor models', *Econometrica* **70**(1), 191–221.

Bai, J. & Ng, S. (2006), 'Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions', *Econometrica* **74**(4), 1133–1150.

Bai, J. & Ng, S. (2019), 'Rank regularized estimation of approximate factor models', *Journal of Econometrics* **212**(1), 78–96.

Baltagi, B. H., Kao, C. & Wang, F. (2021), 'Estimating and testing high dimensional factor models with multiple structural changes', *Journal of Econometrics* **220**(2), 349–365.

Barigozzi, M., Cho, H. & Owens, D. (2024), 'Fnets: Factor-adjusted network estimation and forecasting for high-dimensional time series', *Journal of Business & Economic Statistics* **42**(3), 890–902.

Beyhum, J. & Gautier, E. (2023), 'Factor and factor loading augmented estimators for panel regression with possibly nonstrong factors', *Journal of Business & Economic Statistics* **41**(1), 270–281.

Beyhum, J. & Striaukas, J. (2023), 'Sparse plus dense MIDAS regressions and nowcasting during the COVID pandemic', *arXiv preprint arXiv:2306.13362* .

Bickel, P. J., Ritov, Y. & Tsybakov, A. B. (2009), 'Simultaneous analysis of lasso and dantzig selector', *Annals of Statistics* **37**(4), 1705 – 1732.

Breitung, J. & Eickmeier, S. (2011), 'Testing for structural breaks in dynamic factor models', *Journal of Econometrics* **163**(1), 71–84.

Chen, L., Dolado, J. J. & Gonzalo, J. (2014), 'Detecting big structural breaks in large factor models', *Journal of Econometrics* **180**(1), 30–48.

Chernozhukov, V., Chetverikov, D. & Kato, K. (2019), 'Inference on causal and structural parameters using many moment inequalities', *Review of Economic Studies* **86**(5), 1867–1900.

Corradi, V. & Swanson, N. R. (2014), 'Testing for structural stability of factor augmented forecasting models', *Journal of Econometrics* **182**(1), 100–118.

Dong, X., Li, Y., Rapach, D. E. & Zhou, G. (2022), 'Anomalies and the expected market return', *The Journal of Finance* **77**(1), 639–681.

Fan, J., Guo, J. & Zheng, S. (2022), 'Estimating number of factors by adjusted eigenvalues thresholding', *Journal of the American Statistical Association* **117**(538), 852–861.

Fan, J., Liao, Y. & Mincheva, M. (2013), 'Large covariance estimation by thresholding principal orthogonal complements', *Journal of the Royal Statistical Society. Series B, Statistical methodology* **75**(4).

Fan, J., Liao, Y. & Yao, J. (2015), 'Power enhancement in high-dimensional cross-sectional tests', *Econometrica* **83**(4), 1497–1541.

Fan, J., Lou, Z. & Yu, M. (2024), 'Are latent factor regression and sparse regression adequate?', *Journal of the American Statistical Association* **119**(546), 1076–1088.

Fan, J., Masini, R. & Medeiros, M. C. (2022), 'Do we exploit all information for counterfactual analysis? benefits of factor models and idiosyncratic correction', *Journal of the American Statistical Association* **117**(538), 574–590.

Fan, J., Masini, R. P. & Medeiros, M. C. (2023), 'Bridging factor and sparse models', *The Annals of Statistics* **51**(4), 1692–1717.

Fosten, J. (2017*a*), 'Confidence intervals in regressions with estimated factors and idiosyncratic components', *Economics letters* **157**, 71–74.

Fosten, J. (2017*b*), 'Model selection with estimated factors and idiosyncratic components', *Journal of Applied Econometrics* **32**(6), 1087–1106.

Friedman, J., Hastie, T., Höfling, H. & Tibshirani, R. (2007), 'Pathwise coordinate optimization', *The annals of applied statistics* **1**(2), 302–332.

Fu, Z., Hong, Y. & Wang, X. (2023), 'Testing for structural changes in large dimensional factor models via discrete fourier transform', *Journal of Econometrics* **233**(1), 302–331.

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

Gonçalves, S. & Perron, B. (2014), 'Bootstrapping factor-augmented regression models', *Journal of Econometrics* **182**(1), 156–173.

Gonçalves, S. & Perron, B. (2020), 'Bootstrapping factor models with cross sectional dependence', *Journal of Econometrics* **218**(2), 476–495.

Han, X. & Inoue, A. (2015), 'Tests for parameter instability in dynamic factor models', *Econometric Theory* **31**(5), 1117–1152.

Hansen, C. & Liao, Y. (2019), 'The factor-lasso and k-step bootstrap approach for inference in high-dimensional economic applications', *Econometric Theory* **35**(3), 465–509.

He, Y., Jaidee, S. & Gao, J. (2023), 'Most powerful test against a sequence of high dimensional local alternatives', *Journal of Econometrics* **234**(1), 151–177.

Jensen, T. I., Kelly, B. & Pedersen, L. H. (2023), 'Is there a replication crisis in finance?', *The Journal of Finance* **78**(5), 2465–2518.

Kolesár, M., Müller, U. K. & Roelsgaard, S. T. (2023), 'The fragility of sparsity', *arXiv preprint arXiv:2311.02299* .

Lederer, J. & Vogt, M. (2021), 'Estimating the Lasso's effective noise.', *Journal of Machine Learning Research* **22**, 276–1.

McCracken, M. W. & Ng, S. (2016), 'FRED-MD: a monthly database for macroeconomic research', *Journal of Business & Economic Statistics* **34**(4), 574–589.

Moon, H. R. & Weidner, M. (2015), 'Linear regression for panel with unknown number of factors as interactive fixed effects', *Econometrica* **83**(4), 1543–1579.

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

Stock, J. H. & Watson, M. W. (2002), 'Forecasting using principal components from a large number of predictors', *Journal of the American Statistical Association* **97**(460), 1167–1179.

Su, L. & Wang, X. (2017), 'On time-varying factor models: Estimation and testing', *Journal of Econometrics* **198**(1), 84–101.

Su, L. & Wang, X. (2020), 'Testing for structural changes in factor models via a nonparametric regression', *Econometric Theory* **36**(6), 1127–1158.

Tibshirani, R. (1996), 'Regression shrinkage and selection via the lasso', *Journal of the Royal Statistical Society Series B: Statistical Methodology* **58**(1), 267–288.

van de Geer, S., Bühlmann, P., Ritov, Y. & Dezeure, R. (2014), 'On asymptotically optimal confidence regions and tests for high-dimensional models', *Annals of Statistics* **42**(3), 1166 – 1202.

Vogt, M., Walsh, C. & Linton, O. (2022), 'CCE estimation of high-dimensional panel data models with interactive fixed effects', *arXiv preprint arXiv:2206.12152* .

Xu, W. (2022), 'Testing for time-varying factor loadings in high-dimensional factor models', *Econometric Reviews* **41**(8), 918–965.

Yamamoto, Y. & Tanaka, S. (2015), 'Testing for factor loading structural change under common breaks', *Journal of Econometrics* **189**(1), 187–206.

Zhang, C.-H. & Zhang, S. S. (2014), 'Confidence intervals for low dimensional parameters in high dimensional linear models', *Journal of the Royal Statistical Society Series B: Statistical Methodology* **76**(1), 217–242.

Zhu, Y. & Bradic, J. (2018), 'Linear hypothesis testing in dense high-dimensional linear models', *Journal of the American Statistical Association* **113**(524), 1583–1600.

Online appendix of "Testing for sparse idiosyncratic components in factor-augmented regression models"

Jad Beyhum
Department of Economics, KU Leuven, Belgium

Jonas Striaukas
Department of Finance, Copenhagen Business School, Denmark.

July 27, 2024

Contents

A Testing procedure of the model with lagged idiosyncratic terms

Algorithm OA.1 presents the test for the model with lagged idiosyncratic elements, see (6). This algorithm differs from the main procedure in Algorithm 1 by augmenting the lagged idiosyncratic terms into the sparse component. The procedure could be extended to accommodate observed factors, similar to Algorithm 2. The algorithm with lagged idiosyncratic terms is outlined as follows and can be easily adapted in cases where there is more than one lag.

---

1. Estimate $\widehat{K}$ by one of the available estimators of the number of factors.

2. Let the columns of $\widehat{F}/\sqrt{T}$ be the eigenvectors corresponding to the leading $\widehat{K}$ eigenvalues of $XX^\top$.

3. Compute $\widehat{U} = \left(I_T - \widehat{P}\right) X$ and $\widetilde{Y} = \left(I_T - \widehat{P}\right) Y$, where $\widehat{P} = T^{-1}\widehat{F}\widehat{F}^\top$.

3a. Denote by $\widehat{u}_t$ the $T \times 1$ vector corresponding to transpose of the t^{th} row of $\widehat{U}$. For all $t \in \{2, \ldots T\}$, let $\widetilde{u}_t = \left(\widehat{u}_t^\top, \widehat{u}_{t-1}^\top\right)$.

4. Calculate an approximation $\widehat{\lambda}_{\alpha,emp}$ of $\widehat{\lambda}_\alpha$ as follows:

4a. Specify a grid $0 < \lambda_1 < \cdots < \lambda_M < \bar{\lambda}$, with $\bar{\lambda} = 2(T-1)^{-1} \left\| \sum_{t=2}^T \widetilde{u}_t \widetilde{y}_t \right\|_\infty$.

4b. For $\lambda > 0$, and $e \in \mathbb{R}^{\widetilde{T}}$, let $\widehat{Q}\left(\lambda, e\right) = \left\| \frac{2}{T-1} \sum_{t=2}^T \widetilde{u}_t \widehat{\varepsilon}_{\lambda,t} e_t, \right\|_\infty$, where $\widehat{\varepsilon}_{\lambda,t} = \widetilde{y}_t - \widetilde{u}_t^\top \widehat{\beta}_\lambda,\ t \in \{2, \ldots, T\}$, for $\widehat{\beta}_\lambda = \arg\min_{\beta \in \mathbb{R}^p} \frac{1}{T-1} \sum_{t=2}^T \left(\widetilde{y}_t - \widetilde{u}_t^\top \widehat{\beta}_\lambda\right)^2 + \lambda \|\beta\|_1$. For $m \in [M]$, compute $\left\{\widehat{Q}\left(\lambda_m, e^{(\ell)}\right) :\ \ell \in [L]\right\}$ for L draws of $e \sim \mathcal{N}(0, I_{T-1})$ and the corresponding empirical $(1-\alpha)$-quantile $\widehat{q}_{\alpha,emp}(\lambda_m)$ from them.

4c. Let $\widehat{\lambda}_{\alpha,emp} = \widehat{q}_{\alpha,emp}(\lambda_{\widehat{m}})$, with $\widehat{m} = \min\{m \in [M] :\ \widehat{q}_{\alpha,emp}(\lambda_{m'}) \leq \lambda_{m'}$ for all $m' \geq m\}$.

5. Reject H_0 when $2(T-1)^{-1} \left\| \sum_{t=2}^T \widetilde{u}_t \widetilde{y}_t \right\|_\infty > \widehat{\lambda}_{\alpha,emp}$.

Algorithm OA.1: Conducting a test of level $\alpha \in (0, 1)$ with lagged idiosyncratic terms.

B Additional simulation results

	p/T = 400/400			p/T = 2000/400		
m	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$
			Design 1: $s = \rho_f = \rho_u = \rho_e = 0$			
0.0	0.068	0.024	0.002	0.072	0.026	0.005
0.1	0.085	0.038	0.005	0.075	0.034	0.005
0.2	0.460	0.367	0.210	0.170	0.101	0.034
0.3	0.949	0.926	0.847	0.563	0.456	0.272
0.4	1.000	1.000	0.999	0.898	0.856	0.734
			Design 2: $s = 0.1$, $\rho_f = 0.6$, $\rho_u = 0.1$ and $\rho_e = 0$			
0.0	0.078	0.028	0.006	0.070	0.027	0.005
0.1	0.098	0.046	0.011	0.074	0.030	0.004
0.2	0.473	0.378	0.230	0.174	0.101	0.034
0.3	0.961	0.942	0.873	0.551	0.448	0.272
0.4	1.000	0.999	0.997	0.900	0.850	0.720
			Design 3: $s = 0.1$, $\rho_f = 0.6$ and $\rho_u = \rho_e = 0.1$			
0.0	0.088	0.035	0.007	0.084	0.032	0.005
0.1	0.112	0.052	0.014	0.086	0.037	0.007
0.2	0.470	0.373	0.226	0.185	0.102	0.035
0.3	0.961	0.935	0.870	0.555	0.439	0.264
0.4	1.000	0.999	0.996	0.895	0.843	0.721

Table OA.1: Rejection probabilities for the three dependence designs we consider and sparse β^*. The data are generated with Gaussian variables. The sample size is $T = 400$ while the number of regressors is $p \in \{400, 2000\}$.

	p/T = 400/400			p/T = 2000/400		
m	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$
			Design 1: $s = \rho_f = \rho_u = \rho_e = 0$			
0.0	0.068	0.024	0.002	0.072	0.026	0.005
0.1	0.074	0.028	0.003	0.070	0.030	0.005
0.2	0.086	0.032	0.004	0.074	0.029	0.005
0.3	0.094	0.041	0.005	0.084	0.034	0.004
0.4	0.104	0.052	0.006	0.088	0.037	0.004
			Design 2: $s = 0.1$, $\rho_f = 0.6$, $\rho_u = 0.1$ and $\rho_e = 0$			
0.0	0.073	0.035	0.005	0.070	0.031	0.006
0.1	0.082	0.037	0.007	0.070	0.032	0.008
0.2	0.092	0.041	0.009	0.078	0.036	0.007
0.3	0.101	0.041	0.006	0.080	0.036	0.008
0.4	0.112	0.048	0.007	0.092	0.037	0.007
			Design 3: $s = 0.1$, $\rho_f = 0.6$ and $\rho_u = \rho_e = 0.1$			
0.0	0.090	0.043	0.006	0.078	0.035	0.007
0.1	0.094	0.044	0.009	0.083	0.038	0.008
0.2	0.110	0.046	0.008	0.087	0.041	0.009
0.3	0.116	0.052	0.007	0.094	0.046	0.009
0.4	0.131	0.056	0.007	0.103	0.044	0.010

Table OA.2: Rejection probabilities for the three dependence designs we consider and dense β^*. The data are generated with Gaussian variables. The sample size is $T = 400$ while the number of regressors is $p \in \{400, 2000\}$.

	p/T = 200/200			p/T = 1000/200		
m	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$
			Design 1: $s = \rho_f = \rho_u = \rho_e = 0$			
0.0	0.062	0.020	0.002	0.054	0.018	0.002
0.1	0.082	0.034	0.005	0.060	0.021	0.002
0.2	0.529	0.422	0.230	0.200	0.116	0.035
0.3	0.936	0.886	0.766	0.579	0.445	0.235
0.4	0.994	0.987	0.950	0.872	0.785	0.585
			Design 2: $s = 0.1$, $\rho_f = 0.6$, $\rho_u = 0.1$ and $\rho_e = 0$			
0.0	0.053	0.023	0.004	0.050	0.014	0.002
0.1	0.090	0.040	0.006	0.056	0.018	0.003
0.2	0.537	0.414	0.220	0.196	0.114	0.040
0.3	0.942	0.894	0.761	0.599	0.469	0.246
0.4	0.992	0.987	0.950	0.878	0.801	0.594
			Design 3: $s = 0.1$, $\rho_f = 0.6$ and $\rho_u = \rho_e = 0.1$			
0.0	0.065	0.025	0.003	0.060	0.018	0.002
0.1	0.096	0.048	0.006	0.070	0.023	0.004
0.2	0.530	0.410	0.218	0.210	0.114	0.040
0.3	0.936	0.892	0.755	0.593	0.467	0.242
0.4	0.993	0.986	0.950	0.876	0.799	0.590

Table OA.3: Rejection probabilities for the three dependence designs we consider, sparse β^* and with data generated with student-$t(5)$ variables. The sample size is $T = 200$ while the number of regressors is $p \in \{200, 1000\}$.

	p/T = 400/400			p/T = 2000/400		
m	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$
			Design 1: $s = \rho_f = \rho_u = \rho_e = 0$			
0.0	0.044	0.016	0.001	0.037	0.010	0.001
0.1	0.056	0.024	0.004	0.040	0.012	0.002
0.2	0.364	0.259	0.121	0.118	0.055	0.015
0.3	0.880	0.822	0.632	0.416	0.301	0.136
0.4	0.991	0.981	0.922	0.761	0.643	0.418
			Design 2: $s = 0.1$, $\rho_f = 0.6$, $\rho_u = 0.1$ and $\rho_e = 0$			
0.0	0.042	0.015	0.001	0.031	0.009	0.002
0.1	0.054	0.021	0.004	0.034	0.010	0.002
0.2	0.372	0.272	0.139	0.104	0.049	0.014
0.3	0.890	0.829	0.641	0.420	0.305	0.127
0.4	0.989	0.974	0.921	0.760	0.657	0.410
			Design 3: $s = 0.1$, $\rho_f = 0.6$ and $\rho_u = \rho_e = 0.1$			
0.0	0.050	0.013	0.001	0.039	0.011	0.002
0.1	0.056	0.021	0.004	0.040	0.012	0.001
0.2	0.374	0.268	0.142	0.114	0.054	0.014
0.3	0.883	0.820	0.633	0.414	0.302	0.128
0.4	0.989	0.974	0.919	0.753	0.656	0.404

Table OA.4: Rejection probabilities for the three dependence designs we consider, sparse β^* and with data generated with student-$t(5)$ variables. The sample size is $T = 400$ while the number of regressors is $p \in \{400, 2000\}$.

	p/T = 200/200			p/T = 1000/200		
m	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$
			Design 1: $s = \rho_f = \rho_u = \rho_e = 0$			
0.0	0.084	0.036	0.006	0.094	0.040	0.007
0.1	0.126	0.063	0.016	0.106	0.052	0.011
0.2	0.594	0.494	0.322	0.270	0.176	0.080
0.3	0.982	0.961	0.907	0.669	0.580	0.391
0.4	1.000	1.000	0.999	0.944	0.908	0.799
			Design 2: $s = 0.1$, $\rho_f = 0.6$, $\rho_u = 0.1$ and $\rho_e = 0$			
0.0	0.092	0.046	0.008	0.091	0.044	0.005
0.1	0.120	0.066	0.016	0.111	0.055	0.009
0.2	0.609	0.518	0.348	0.268	0.180	0.072
0.3	0.983	0.968	0.920	0.704	0.613	0.408
0.4	1.000	1.000	0.999	0.950	0.922	0.838
			Design 3: $s = 0.1$, $\rho_f = 0.6$ and $\rho_u = \rho_e = 0.1$			
0.0	0.091	0.041	0.006	0.091	0.039	0.001
0.1	0.115	0.063	0.014	0.109	0.052	0.005
0.2	0.607	0.517	0.347	0.267	0.178	0.067
0.3	0.980	0.964	0.918	0.700	0.611	0.405
0.4	0.997	0.999	0.997	0.948	0.922	0.835

Table OA.5: Rejection probabilities for the three dependence designs we consider, sparse β^* and using $K = 5$ as the number of factors while the true number of factors is $K = 2$. The data are generated with Gaussian variables. The sample size is $T = 200$ while the number of regressors is $p \in \{200, 1000\}$.

	p/T = 400/400			p/T = 2000/400		
m	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$
			Design 1: $s = \rho_f = \rho_u = \rho_e = 0$			
0.0	0.079	0.028	0.002	0.083	0.031	0.005
0.1	0.094	0.042	0.004	0.083	0.034	0.005
0.2	0.451	0.362	0.212	0.184	0.106	0.030
0.3	0.948	0.923	0.845	0.548	0.456	0.268
0.4	1.000	1.000	0.998	0.895	0.847	0.723
			Design 2: $s = 0.1$, $\rho_f = 0.6$, $\rho_u = 0.1$ and $\rho_e = 0$			
0.0	0.080	0.035	0.008	0.084	0.029	0.006
0.1	0.102	0.044	0.013	0.088	0.032	0.005
0.2	0.467	0.374	0.233	0.180	0.106	0.030
0.3	0.956	0.936	0.872	0.545	0.442	0.262
0.4	1.000	0.999	0.998	0.898	0.846	0.711
			Design 3: $s = 0.1$, $\rho_f = 0.6$ and $\rho_u = \rho_e = 0.1$			
0.0	0.077	0.034	0.007	0.084	0.029	0.004
0.1	0.102	0.040	0.011	0.088	0.032	0.003
0.2	0.465	0.371	0.231	0.176	0.103	0.029
0.3	0.956	0.933	0.870	0.540	0.438	0.259
0.4	0.995	0.996	0.996	0.896	0.842	0.711

Table OA.6: Rejection probabilities for the three dependence designs we consider, sparse β^* and using $K = 5$ as the number of factors while the true number of factors is $K = 2$. The data are generated with Gaussian variables. The sample size is $T = 400$ while the number of regressors is $p \in \{400, 2000\}$.

	p/T = 200/200			p/T = 1000/200		
m	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$
			Design 1: $s = \rho_f = \rho_u = \rho_e = 0$			
0.0	0.893	0.875	0.837	0.843	0.812	0.743
0.1	0.895	0.876	0.837	0.840	0.813	0.739
0.2	0.929	0.906	0.858	0.853	0.824	0.740
0.3	0.987	0.973	0.931	0.907	0.876	0.780
0.4	0.999	0.998	0.984	0.969	0.934	0.847
			Design 2: $s = 0.1$, $\rho_f = 0.6$, $\rho_u = 0.1$ and $\rho_e = 0$			
0.0	0.929	0.910	0.881	0.877	0.856	0.800
0.1	0.928	0.913	0.879	0.877	0.851	0.799
0.2	0.952	0.930	0.891	0.882	0.858	0.810
0.3	0.986	0.969	0.928	0.923	0.890	0.826
0.4	0.997	0.994	0.974	0.964	0.936	0.858
			Design 3: $s = 0.1$, $\rho_f = 0.6$ and $\rho_u = \rho_e = 0.1$			
0.0	0.925	0.912	0.879	0.884	0.858	0.804
0.1	0.929	0.915	0.885	0.879	0.850	0.804
0.2	0.951	0.925	0.892	0.888	0.855	0.805
0.3	0.984	0.971	0.930	0.918	0.885	0.821
0.4	0.997	0.994	0.976	0.964	0.934	0.856

Table OA.7: Rejection probabilities for the three dependence designs we consider, sparse β^* and using $K = 1$ as the number of factors while the true number of factors is $K = 2$. The data are generated with Gaussian variables. The sample size is $T = 200$ while the number of regressors is $p \in \{200, 1000\}$.

	p/T = 400/400			p/T = 2000/400		
m	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$
			Design 1: $s = \rho_f = \rho_u = \rho_e = 0$			
0.0	0.914	0.898	0.859	0.867	0.839	0.777
0.1	0.917	0.895	0.859	0.866	0.838	0.780
0.2	0.934	0.911	0.863	0.868	0.843	0.780
0.3	0.973	0.959	0.906	0.894	0.859	0.789
0.4	0.998	0.994	0.971	0.943	0.909	0.824
			Design 2: $s = 0.1$, $\rho_f = 0.6$, $\rho_u = 0.1$ and $\rho_e = 0$			
0.0	0.931	0.919	0.892	0.890	0.869	0.829
0.1	0.935	0.921	0.892	0.889	0.871	0.818
0.2	0.936	0.924	0.892	0.889	0.869	0.820
0.3	0.978	0.961	0.919	0.914	0.884	0.820
0.4	0.997	0.994	0.972	0.946	0.914	0.840
			Design 3: $s = 0.1$, $\rho_f = 0.6$ and $\rho_u = \rho_e = 0.1$			
0.0	0.930	0.919	0.892	0.890	0.866	0.829
0.1	0.933	0.922	0.891	0.889	0.865	0.819
0.2	0.940	0.923	0.892	0.891	0.869	0.815
0.3	0.977	0.961	0.920	0.914	0.884	0.821
0.4	0.998	0.993	0.972	0.945	0.914	0.838

Table OA.8: Rejection probabilities for the three dependence designs we consider, sparse β^* and using $K = 1$ as the number of factors while the true number of factors is $K = 2$. The data are generated with Gaussian variables. The sample size is $T = 400$ while the number of regressors is $p \in \{400, 2000\}$.

	p/T = 200/200			p/T = 1000/200		
m	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$
			Design 1: $s = \rho_f = \rho_u = \rho_e = 0$			
0.0	0.084	0.038	0.007	0.073	0.032	0.003
0.1	0.115	0.063	0.015	0.078	0.035	0.004
0.2	0.537	0.442	0.290	0.216	0.135	0.052
0.3	0.973	0.958	0.896	0.633	0.530	0.346
0.4	1.000	1.000	0.999	0.933	0.902	0.794
			Design 2: $s = 0.1$, $\rho_f = 0.6$, $\rho_u = 0.1$ and $\rho_e = 0$			
0.0	0.074	0.036	0.006	0.070	0.027	0.004
0.1	0.097	0.052	0.010	0.084	0.034	0.006
0.2	0.565	0.468	0.288	0.222	0.138	0.052
0.3	0.968	0.953	0.898	0.654	0.540	0.360
0.4	0.999	0.999	0.999	0.943	0.907	0.794
			Design 3: $s = 0.1$, $\rho_f = 0.6$ and $\rho_u = \rho_e = 0.1$			
0.0	0.089	0.042	0.010	0.084	0.030	0.004
0.1	0.112	0.059	0.014	0.093	0.038	0.008
0.2	0.571	0.460	0.282	0.226	0.135	0.055
0.3	0.969	0.951	0.895	0.647	0.535	0.355
0.4	0.999	0.999	0.999	0.935	0.897	0.792

Table OA.9: Rejection probabilities for the three dependence designs we consider, sparse β^* and lagged idiosyncratic terms. The data are generated with Gaussian variables. The sample size is $T = 200$ while the number of regressors is $p \in \{200, 1000\}$.

	p/T = 400/400			p/T = 2000/400		
m	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$
			Design 1: $s = \rho_f = \rho_u = \rho_e = 0$			
0.0	0.070	0.028	0.005	0.061	0.027	0.005
0.1	0.081	0.038	0.011	0.064	0.028	0.004
0.2	0.400	0.320	0.184	0.127	0.066	0.016
0.3	0.932	0.903	0.823	0.467	0.373	0.206
0.4	1.000	0.999	0.995	0.862	0.812	0.664
			Design 2: $s = 0.1$, $\rho_f = 0.6$, $\rho_u = 0.1$ and $\rho_e = 0$			
0.0	0.078	0.036	0.006	0.059	0.020	0.001
0.1	0.088	0.042	0.008	0.062	0.024	0.002
0.2	0.418	0.320	0.184	0.130	0.070	0.021
0.3	0.936	0.905	0.823	0.491	0.388	0.236
0.4	0.999	0.998	0.995	0.878	0.826	0.693
			Design 3: $s = 0.1$, $\rho_f = 0.6$ and $\rho_u = \rho_e = 0.1$			
0.0	0.091	0.043	0.006	0.068	0.028	0.002
0.1	0.106	0.048	0.007	0.067	0.029	0.003
0.2	0.420	0.322	0.173	0.137	0.077	0.024
0.3	0.931	0.904	0.812	0.490	0.387	0.230
0.4	1.000	0.998	0.995	0.875	0.823	0.682

Table OA.10: Rejection probabilities for the three dependence designs we consider, sparse β^*, and lagged idiosyncratic terms. The data are generated with Gaussian variables. The sample size is $T = 400$ while the number of regressors is $p \in \{400, 2000\}$.

C Additional empirical results

C.1 Macro application — FRED-QD

Category	10%	5%	1%
NIPA (23)	0.478	0.435	0.261
Industrial Production (16)	0.625	0.500	0.125
Employment and Unemployment (48)	0.688	0.562	0.375
Housing (10)	0.400	0.200	0.200
Inventories, Orders, and Sales (7)	0.286	0.286	0.286
Prices (36)	0.417	0.306	0.194
Earnings and Productivity (13)	0.692	0.692	0.692
Interest Rates (15)	0.400	0.400	0.133
Money and Credit (14)	0.571	0.571	0.500
Household Balance Sheets (8)	0.250	0.125	0.125
Stock Markets (1)	0.000	0.000	0.000
Exchange Rates (6)	0.833	0.833	0.333
Other (2)	0.000	0.000	0.000
Non-Household Balance Sheets (2)	0.500	0.500	0.000

Table OA.11: Rejection ratios for each category of the FRED-QD dataset over the sample period 1959 Q3 to 2019 Q4. The sample size is $T = 241$ while the number of regressors is $p = 202$. In parentheses, we report the number of series per category. Data source: McCracken & Ng (2020).

C.2 Finance I: additional details on data and results

	Variable name	Full name	Source
1	Market	Market	Dong et al. (2022)
2	Agric	Agriculture	French (2024)
3	Food	Food Products	French (2024)

4	Soda	Candy & Soda	French (2024)
5	Beer	Beer & Liquor	French (2024)
6	Smoke	Tobacco Products	French (2024)
7	Toys	Recreation	French (2024)
8	Fun	Entertainment	French (2024)
9	Books	Printing and Publishing	French (2024)
10	Hshld	Consumer Goods	French (2024)
11	Clths	Apparel	French (2024)
12	Hlth	Healthcare	French (2024)
13	MedEq	Medical Equipment	French (2024)
14	Drugs	Pharmaceutical Products	French (2024)
15	Chems	Chemicals	French (2024)
16	Rubbr	Rubber and Plastic Products	French (2024)
17	Txtls	Textiles	French (2024)
18	BldMt	Construction Materials	French (2024)
19	Cnstr	Construction	French (2024)
20	Steel	Steel Works Etc	French (2024)
21	FabPr	Fabricated Products	French (2024)
22	Mach	Machinery	French (2024)
23	ElcEq	Electrical Equipment	French (2024)
24	Autos	Automobiles and Trucks	French (2024)
25	Aero	Aircraft	French (2024)
26	Ships	Shipbuilding, Railroad Equipment	French (2024)
27	Guns	Defense	French (2024)
28	Gold	Precious Metals	French (2024)
29	Mines	Non-Metallic and Industrial Metal Mining	French (2024)
30	Coal	Coal	French (2024)
31	Oil	Petroleum and Natural Gas	French (2024)
32	Util	Utilities	French (2024)
33	Telcm	Communication	French (2024)
34	PerSv	Personal Services	French (2024)

35	BusSv	Business Services	French (2024)
36	Hardw	Computers	French (2024)
37	Softw	Computer Software	French (2024)
38	Chips	Electronic Equipment	French (2024)
39	LabEq	Measuring and Control Equipment	French (2024)
40	Paper	Business Supplies	French (2024)
41	Boxes	Shipping Containers	French (2024)
42	Trans	Transportation	French (2024)
43	Whlsl	Wholesale	French (2024)
44	Rtail	Retail	French (2024)
45	Meals	Restaurants, Hotels, Motels	French (2024)
46	Banks	Banking	French (2024)
47	Insur	Insurance	French (2024)
48	RlEst	Real Estate	French (2024)
49	Fin	Trading	French (2024)
50	Other	Other	French (2024)

Table OA.12: A list of return portfolios for the results in reported in Table 4. We use average value weighted return.

NoDur**	0.013	**Durbl*****	0.004	**Manuf*****	0.001	**Enrgy****	0.016	HiTec	0.533
Telcm	0.625	**Shops****	0.049	Hlth	0.265	Utils	0.252	**Other*****	0.004

Table OA.13: P-values of industry excess returns (10 industry classification, see Table OA.14) regressed on 100 characteristics. Bold entries with $*$, $**$, and $***$ indicate significance at the 10%, 5% and 1% levels, respectively.

	Variable name	Full name	Source
1	NoDur	Consumer Nondurables	French (2024)
2	Durbl	Consumer Durables	French (2024)
3	Manuf	Manufacturing	French (2024)

4	Enrgy	Oil, Gas, and Coal	French (2024)
5	HiTec	Computers, Software, and Business/Electronic Equipment	French (2024)
6	Telcm	Telephone and Television Transmission	French (2024)
7	Shops	Wholesale, Retail, and Some Services	French (2024)
8	Hlth	Healthcare, Medical Equipment, and Drugs	French (2024)
9	Utils	Utilities	French (2024)
10	Other	Other	French (2024)

Table OA.14: A list of return portfolios for the results in reported in Table OA.13. We use average value weighted return.

C.3 Finance II: additional details on data

	Variable name	Factor	Source
1	MKT	Market	French (2024)
2	SMB	Small-minus-Big	French (2024)
3	HML	High-minus-Low	French (2024)
4	CMA	Conservative-minus-Aggressive	French (2024)
5	RMW	Robust-minus-Weak	French (2024)
6	ni_me	Earnings/price ratio	Jensen et al. (2023)
7	fcf_me	Cash-flow/price ratio	Jensen et al. (2023)
8	div12m_me	Dividend/price ratio	Jensen et al. (2023)
9	taccruals_at	Accruals	Jensen et al. (2023)
10	beta_60m	60 Month CAPM Beta	Jensen et al. (2023)
11	turnover_126d	Share turnover	Jensen et al. (2023)
12	rmax5_rvol_21d	Max Return to Volatility	Jensen et al. (2023)
13	ivol_ff3_21d	Idiosync. vol. from the Fama-French 3-factor model	Jensen et al. (2023)
14	ret_3_1	Momentum 1-3 Months	Jensen et al. (2023)
15	ret_60_12	Momentum 12-60 Months	Jensen et al. (2023)

Table OA.15: A list of observable factors used in Section 5.3.

	10 %	5 %	1 %	Number of firms
Agric	0.000	0.000	0.000	2
Food	0.000	0.000	0.000	19
Soda	0.000	0.000	0.000	4
Beer	0.333	0.167	0.000	6
Smoke	0.000	0.000	0.000	1
Toys	0.000	0.000	0.000	5
Fun	0.000	0.000	0.000	2
Books	0.375	0.125	0.125	8
Hshld	0.105	0.053	0.000	19
Clths	0.000	0.000	0.000	9
Hlth	0.100	0.100	0.100	10
MedEq	0.053	0.000	0.000	19
Drugs	0.053	0.053	0.000	19
Chems	0.133	0.067	0.000	15
Rubbr	0.000	0.000	0.000	3
Txtls	0.000	0.000	0.000	5
BldMt	0.000	0.000	0.000	15
Cnstr	0.250	0.167	0.083	12
Steel	0.222	0.111	0.111	9
FabPr	0.000	0.000	0.000	1
Mach	0.163	0.116	0.047	43
ElcEq	0.167	0.167	0.083	12
Autos	0.000	0.000	0.000	14
Aero	0.083	0.083	0.000	12
Ships	0.000	0.000	0.000	1
Guns	0.000	0.000	0.000	4
Gold	1.000	1.000	0.833	6
Mines	0.000	0.000	0.000	2
Oil	0.286	0.286	0.036	28
Util	0.154	0.077	0.019	52

Telcm	0.455	0.455	0.364	11
PerSv	0.000	0.000	0.000	5
BusSv	0.065	0.032	0.000	31
Hardw	0.000	0.000	0.000	12
Softw	0.100	0.050	0.000	20
Chips	0.206	0.059	0.029	34
LabEq	0.211	0.211	0.105	19
Paper	0.000	0.000	0.000	11
Boxes	0.000	0.000	0.000	3
Trans	0.050	0.000	0.000	20
Whlsl	0.179	0.143	0.071	28
Rtail	0.138	0.069	0.069	29
Meals	0.000	0.000	0.000	8
Banks	0.096	0.027	0.014	73
Insur	0.175	0.050	0.050	40
RlEst	0.000	0.000	0.000	5
Fin	0.077	0.077	0.000	13
Other	0.000	0.000	0.000	2

Table OA.16: Rejection ratios for each industry of the returns of firms in the dataset of the application in Section 5.3 over the sample period January 1991 to December 2022, thus $T = 384$ and $p = 727$. In the Column *Number of firms* , we report the number of firms per industry in our sample. Industry classification is based on 49 industry portfolios of French (2024), see Table OA.12 for a full list of industries. Data source: Jensen et al. (2023).

D On Theorem 1

This section contains material related to Theorem 1. In Section D.1, we define some useful mathematical objects. The proof of Theorem 1 is given in Section D.2 and makes use of results proved in later sections. Section D.3 contains some auxiliary lemmas on distribution functions of random variables used in the proof of Theorem 1. Then, in Section D.4, we state and prove some lemmas on the probability of some events. Section D.5, contains results on some sequences introduced in the proof of Theorem 1. Furthermore, Section D.6 introduces results on the factors, the loadings and their estimators. Finally, Section D.7 recalls pre-existing results on strong mixing sequences and high-dimensional Gaussian vectors. Finally, Section D.8 discusses the rate condition in statement (ii) of Theorem 1. Our proofs borrow ideas and results from Chernozhukov et al. (2013), Chernozhukov et al. (2015), Lederer & Vogt (2021), Fan et al. (2024) and Fan et al. (2023).

D.1 Preliminaries

We introduce some concepts which are latter useful in proving Theorem 1.

First, we define (infeasible) estimators using the true number of factors. These are all denoted using a "check" symbol. We let the columns of $\breve{F}/\sqrt{T}$ be the eigenvectors corresponding to the leading K eigenvalues of $XX^\top$ and $\breve{B} = X^\top \breve{F}(\breve{F}^\top \breve{F})^{-1} = T^{-1}X^\top \breve{F}$. Let also $\breve{P} = T^{-1}\breve{F}\left(\breve{F}^\top \breve{F}/T\right)^{-1}\breve{F}^\top = T^{-1}\breve{F}\breve{F}^\top$ be the projector on the vector space generated by the columns of $\breve{F}$. The estimator of U is $\breve{U} = X - \breve{F}\breve{B}^\top = \left(I_T - \breve{P}\right)X$. Similarly, we let $\breve{Y} = \left(I_T - \breve{P}\right)Y$ be the estimator of $Y - F\gamma^*$. The estimated loading $\breve{b}_{jk}$ corresponds to the j^{th} element of the k^{th} column of $\breve{B}$. Let also $\breve{b}_j = (\breve{b}_{j1}, \ldots, \breve{b}_{jK})^\top$. The second-step LASSO estimator using the true number of factors is then given by

$$\breve{\beta}_\lambda = \underset{\beta \in \mathbb{R}^p}{\arg\min} \frac{1}{T}\left\|\breve{Y} - \breve{U}\beta\right\|_2^2 + \lambda\|\beta\|_1,$$

For $t \in [T]$, we denote by $\breve{y}_t$ the t^{th} element of $\breve{Y}$ and $\breve{u}_t$ as the $T \times 1$ vector corresponding to the t^{th} row of $\breve{U}$. For a given $\lambda > 0$, let $\breve{\varepsilon}_{\lambda,t} = \breve{y}_t - \breve{u}_t^\top \breve{\beta}_\lambda,\ t \in [T]$. The equivalent of $\widehat{Q}(\lambda, e)$ is then

$$\breve{Q}(\lambda, e) = \left\|\frac{2}{T}\sum_{t=1}^{T} \breve{u}_t \breve{\varepsilon}_{\lambda,t} e_t,\right\|_\infty.$$

The estimator $\breve{q}_\alpha(\lambda)$ of q_α is the $(1-\alpha)$-quantile of the distribution of $\breve{Q}(\lambda, e)$ given X and Y. Formally, $\breve{q}_\alpha(\lambda) = \inf\left\{q : \mathbb{P}_e(\breve{Q}(\lambda, e) \leq q) \geq 1 - \alpha\right\}$ (recall that $\mathbb{P}_e(\cdot) = \mathbb{P}(\cdot|X, Y)$). The analog of $\widehat{\lambda}_\alpha$ is given by

$$\breve{\lambda}_\alpha = \inf\{\lambda > 0 \ :\ \breve{q}_\alpha(\lambda') \leq \lambda' \text{ for all } \lambda' \geq \lambda\}.$$

Remark that, on the event $\{\widehat{K} = K\}$, we have $\widehat{F} = \breve{F}$, $\widehat{B} = \breve{B}$, $\widehat{U} = \breve{U}$, $\widetilde{Y} = \breve{Y}$, $\widehat{\beta}_\lambda = \breve{\beta}_\lambda$ and $\widehat{\lambda}_\alpha = \breve{\lambda}_\alpha$.

As in Lederer & Vogt (2021), we re-scale some quantities by multiplying them with $\sqrt{T}/2$. This re-scaling is convenient to apply some probabilistic results. For instance, we let $\breve{\Pi}(\mu, e) = \left\| \breve{W}(\mu, e) \right\|_\infty$, where

$$\breve{W}(\mu, e) = \left(\breve{W}_1(\mu, e), \ldots, \breve{W}_p(\mu, e) \right)^\top, \text{ with } \breve{W}_j(\mu, e) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \breve{u}_{tj} \breve{\varepsilon}_{\frac{2}{\sqrt{T}}\mu, t} e_t.$$

Note that $\breve{\Pi}(\mu, e) = \frac{\sqrt{T}}{2} \breve{Q}(\lambda, e)$, for $\lambda = \frac{2}{\sqrt{T}}\mu$. Similarly, for $\alpha \in (0,1)$, we define

$$\breve{\pi}_\alpha(\mu) = \inf\{q : \mathbb{P}_e(\breve{\Pi}(\mu, e) \leq q) \geq 1 - \alpha\};$$
$$\breve{\mu}_\alpha = \inf\{\mu > 0 \ : \ \breve{\pi}_\alpha(\mu') \leq \mu' \text{ for all } \mu' \geq \mu\},$$

where $\breve{\mu}_\alpha = \frac{\sqrt{T}}{2} \breve{\lambda}_\alpha$.

Next, to be able to compare $\breve{\Pi}(e)$ with population analogs, we define several additional quantities. Let $\Pi(e) = \|W(e)\|_\infty$, where

$$W(e) = (W_1(e), \ldots, W_p(e))^\top, \text{ with } W_j(e) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} u_{tj} \varepsilon_t e_t$$

and let μ_α be the $(1-\alpha)-$quantile of $\Pi(e)$ conditionally on $(F, U, \mathcal{E})$. Formally, $\mu_\alpha = \inf\{q : \mathbb{P}^*_e(\Pi(e) \leq q) \geq 1 - \alpha\}$, where $\mathbb{P}^*_e(\cdot) = \mathbb{P}(\cdot | F, U, \mathcal{E})$.

Moreover, we define $\Pi^* = \|W^*\|_\infty$, where

$$W^* = \left(W^*_1, \ldots, W^*_p \right)^\top, \text{ with } W^*_j = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} u_{tj} \varepsilon_t,$$

where μ^*_α is the $(1-\alpha)$ quantile of Π^*. Finally, we also set $\Pi^G = \|G\|_\infty$ with G a Gaussian vector with same covariance structure as W^* and let μ^G_α be the $(1-\alpha)$-quantile of Π^G. Auxiliary lemmas concerning the distributions of $\Pi(e)$, Π^* and Π^G can be found in Section D.3.

We also introduce the following useful quantities

$$\Delta = \left\| \frac{1}{T} \sum_{t=1}^{T} u_t u_t^\top \varepsilon_t^2 - \mathbb{E}\left[u_t u_t^\top \varepsilon_t^2 \right] \right\|_\infty, \ R(\mu, e) = \frac{1}{\sqrt{T}} \left\| \breve{W}(\mu, e) - W(e) \right\|_\infty,$$

for $\mu > 0$ and the event $\mathcal{S}_\mu = \left\{ \frac{2}{T} \left\| \breve{U}^\top (\breve{Y} - \breve{U}\beta^*) \right\|_\infty \leq \frac{2}{\sqrt{T}}\mu \right\}$. The above terms and events are controlled in Section D.4.

The following sequences allow to bound some important terms in the proofs.

$$s_T^{(1)} = \sqrt{\log(T)}\frac{p^{4/q}}{\sqrt{T}};$$

$$s_T^{(2)} = \sqrt{\log(T)}\left(\frac{1}{p} + \frac{p^{\frac{2}{q}}}{T} + \frac{p^{\frac{2}{q}-\frac{1}{2}}}{\sqrt{T}}\right)(\|\varphi^*\|_2 \vee 1);$$

$$s_T^{(3)} = \sqrt{\log(T)}\left(1 + \frac{p^{\frac{4}{q}}}{T} + \frac{1}{p}\right);$$

$$s_T^{(4)} = \sqrt{\log(T)}\left(\left(p^{\frac{4}{q}}T^{\frac{2}{q}-1} + \frac{T^{\frac{2}{q}}}{p}\right) + \left(\frac{1}{T} + \frac{1}{p}\right)\|\varphi^*\|_2^2\right);$$

$$s_T^{(5)} = \sqrt{\log(T)}\frac{p^{\frac{2}{q}}}{\sqrt{T}};$$

$$s_T^{(6)} = \frac{2}{T^{1/4}}\sqrt{\log(Tp)2\|\beta^*\|_1 s_T^{(3)}};$$

$$s_T^{(7)} = \sqrt{\log(Tp)\frac{s_T^{(4)}}{T}};$$

$$s_T^{(8)} = K_1(\Delta)^{1/3}(1 \vee 2\log(2p) \vee \log(1/\Delta))^{1/3}\log(2p)^{1/3};$$

$$\begin{aligned} s_T^{(9)} = K_2\Big[&\left(T^{\kappa-1/2} + T^{1-\frac{c}{2}(1-\frac{2}{h})}\right)\log(T)\log(p) + T^{-1/4}\log(p)^{3/2}\log(T) + p^{\frac{2}{h}}T^{\frac{1}{h}-\frac{1}{2}}\log(p)^2\log(T) \\ &+ \left(p\log(p)^{\frac{3}{2}h-4}\log(T)\log(Tp)\right)^{\frac{1}{h-2}}T^{-\frac{1}{4}} + T^{\frac{1}{2}-c\kappa}\left(p^{\frac{1}{h+1}}\sqrt{1 \vee \log(p)}\right)\Big]; \end{aligned}$$

$$s_T^{(10)} = \frac{1}{T \vee p} + s_T^{(9)};$$

$$s_T^{(11)} = \kappa_2\left(\sqrt{2\log(2p)} + \sqrt{2\log(T \vee p)}\right);$$

$$s_T^{(12)} = s_T^{(6)}\left(1 + s_T^{(11)}\right) + \frac{\sqrt{T}}{2}s_T^{(2)} + s_T^{(6)}\left(1 + (1 + s_T^{(6)})s_T^{(11)} + \frac{\sqrt{T}}{2}s_T^{(2)}\right) + s_T^{(7)};$$

$$s_T^{(13)} = s_T^{(6)}\left(1 + s_T^{(11)}\right) + s_T^{(7)};$$

$$s_T^{(14)} = s_T^{(9)} + K_3 s_T^{(12)}\sqrt{1 \vee \log\left(2p/s_T^{(12)}\right)} + s_T^{(8)} + \frac{3}{T},$$

where K_1, K_2 and K_3 are constants introduced in Lemmas D.2, D.1 and D.3, respectively. The constant κ_2 is defined in Assumption 3 and h and q are introduced in Assumptions 3 (iii) and 4. In Lemma D.8, we show that these sequences all go to 0 under Assumption 5.

Finally, we introduce the following events

$$\mathcal{S}_T^{(1)} = \left\{\Delta \leq s_T^{(1)}\right\};$$

$$\mathcal{S}_T^{(2)} = \left\{\left\|\frac{\check{U}^\top(\check{Y} - \check{U}\beta^*)}{T} - \frac{U^\top\mathcal{E}}{T}\right\|_\infty \leq s_T^{(2)}\right\};$$

$$\mathcal{S}_T^{(3)} = \left\{ \max_{j\in[p]} \frac{1}{T} \sum_{t=1}^{T} \breve{u}_{tj}^2 \leq s_T^{(3)} \right\};$$
$$\mathcal{S}_T^{(4)} = \left\{ \max_{j\in[p]} \frac{1}{T} \sum_{t=1}^{T} \left(\breve{u}_{tj}\widetilde{\varepsilon}_t + \widetilde{f}_t^\top \varphi^* - u_{tj}\varepsilon_t \right)^2 \leq s_T^{(4)} \right\};$$
$$\mathcal{S}_T^{(5)} = \left\{ \left\| \frac{U^\top \mathcal{E}}{T} \right\|_\infty \leq s_T^{(5)} \right\},$$

where $\widetilde{\varepsilon}_t$ denotes the t^{th} element of $\left(I_T - \breve{P}\right)\mathcal{E}$ and $\widetilde{f}_t$ is the $K \times 1$ vector corresponding to the t^{th} row of $\left(I_T - \breve{P}\right) F$. We show that the probabilities of these events go to 1 with T in Lemma D.5.

D.2 Proof of Theorem 1

Proof of (i). We want to show that when $\beta^* = 0$, we have

$$\mathbb{P}\left(\left\| \frac{\widehat{U}^\top \widetilde{Y}}{T} \right\|_\infty > \widehat{\lambda}_\alpha \right) \leq \alpha + o(1). \tag{OA.1}$$

First, we move from quantities depending on $\widehat{K}$ to quantities only depending on K. Notice that

$$\begin{aligned}
&\mathbb{P}\left(\left\| \frac{\widehat{U}^\top \widetilde{Y}}{T} \right\|_\infty > \widehat{\lambda}_\alpha \right) \\
&= \mathbb{P}\left(\left\{ \left\| \frac{\widehat{U}^\top \widetilde{Y}}{T} \right\|_\infty > \widehat{\lambda}_\alpha \right\} \cap \left\{ \widehat{K} = K \right\} \right) + \mathbb{P}\left(\left\{ \left\| \frac{\widehat{U}^\top \widetilde{Y}}{T} \right\|_\infty > \widehat{\lambda}_\alpha \right\} \cap \left\{ \widehat{K} \neq K \right\} \right) \\
&\leq \mathbb{P}\left(\left\{ \left\| \frac{\breve{U}^\top \breve{Y}}{T} \right\|_\infty > \breve{\lambda}_\alpha \right\} \cap \left\{ \widehat{K} = K \right\} \right) + \mathbb{P}\left(\widehat{K} \neq K \right),
\end{aligned}$$

where, we used the fact that, on the event $\{\widehat{K} = K\}$, we have $\widehat{U} = \breve{U}$, $\widetilde{Y} = \breve{Y}$ and $\widehat{\lambda}_\alpha = \breve{\lambda}_\alpha$. Now, using the formula $\mathbb{P}(C \cap D) = \mathbb{P}(C) + \mathbb{P}(D) - \mathbb{P}(C \cup D)$ (where C and D are two probabilistic events), we obtain

$$\begin{aligned}
\mathbb{P}\left(\left\| \frac{\widehat{U}^\top \widetilde{Y}}{T} \right\|_\infty > \widehat{\lambda}_\alpha \right) &\leq \mathbb{P}\left(\left\| \frac{\breve{U}^\top \breve{Y}}{T} \right\|_\infty > \breve{\lambda}_\alpha \right) + \mathbb{P}(\widehat{K} = K) \\
&\quad - \mathbb{P}\left(\left\{ \left\| \frac{\breve{U}^\top \breve{Y}}{T} \right\|_\infty > \breve{\lambda}_\alpha \right\} \cup \left\{ \widehat{K} = K \right\} \right) + \mathbb{P}\left(\widehat{K} \neq K \right) \\
&\leq \mathbb{P}\left(\left\| \frac{\breve{U}^\top \breve{Y}}{T} \right\|_\infty > \breve{\lambda}_\alpha \right) + \mathbb{P}\left(\widehat{K} \neq K \right)
\end{aligned} \tag{OA.2}$$

$$\leq \mathbb{P}\left(\left\|\frac{\breve{U}^\top \breve{Y}}{T}\right\|_\infty > \breve{\lambda}_\alpha\right) + o(1), \tag{OA.3}$$

where, in equation (OA.2), we used $\mathbb{P}(C \cup D) \geq \mathbb{P}(C)$ and, in equation (OA.3), we used Assumption 1. This yields

$$\begin{aligned}
\mathbb{P}\left(\left\|\frac{\widehat{U}^\top \widetilde{Y}}{T}\right\|_\infty > \widehat{\lambda}_\alpha\right) &\leq \mathbb{P}\left(\left\|\frac{\breve{U}^\top \breve{Y}}{T}\right\|_\infty > \breve{\lambda}_\alpha\right) + o(1) & \text{(OA.4)}\\
&\leq \mathbb{P}\left(\left\|\frac{U^\top \mathcal{E}}{T}\right\|_\infty + \left\|\frac{\breve{U}^\top \breve{Y}}{T} - \frac{U^\top \mathcal{E}}{T}\right\|_\infty > \breve{\lambda}_\alpha\right) + o(1)\\
&\leq \mathbb{P}\left(\left\{\left\|\frac{U^\top \mathcal{E}}{T}\right\|_\infty > \breve{\lambda}_\alpha - s_T^{(2)}\right\} \cap \mathcal{S}_T^{(2)}\right) + \mathbb{P}\left((\mathcal{S}_T^{(2)})^c\right) + o(1)\\
&\leq \mathbb{P}\left(\left\|\frac{U^\top \mathcal{E}}{T}\right\|_\infty > \breve{\lambda}_\alpha - s_T^{(2)}\right) + o(1), & \text{(OA.5)}
\end{aligned}$$

where in the last line we used Lemma D.5 (ii).

Let us define

$$\mathcal{T}_1 = \mathcal{S}_{\mu^*_{\alpha+s_T^{(14)}}} \cap \mathcal{S}_T^{(1)} \cap \mathcal{S}_T^{(2)} \cap \mathcal{S}_T^{(3)} \cap \mathcal{S}_T^{(4)}.$$

Note that, by Lemmas D.5 and D.7, and the fact that $s_T^{(14)} \to 0$ by Lemma D.8 (iv), (v) and (vi), the event $\mathcal{T}_1$ has probability going to $1-\alpha$.

Hence, by (OA.5), to show (OA.1), it suffices to prove that, on $\mathcal{T}_1$, we have

$$\breve{\lambda}_\alpha \geq \frac{2}{\sqrt{T}}\mu^*_{\alpha+s_T^{(14)}} + s_T^{(2)}. \tag{OA.6}$$

Indeed, in this case, we would have

$$\begin{aligned}
\mathbb{P}\left(\left\|\frac{\breve{U}^\top \breve{Y}}{T}\right\|_\infty > \breve{\lambda}_\alpha\right) &\leq \mathbb{P}\left(\left\|\frac{U^\top \mathcal{E}}{T}\right\|_\infty > \breve{\lambda}_\alpha - s_T^{(2)}\right) + o(1)\\
&\leq \mathbb{P}\left(\left\{\left\|\frac{U^\top \mathcal{E}}{T}\right\|_\infty > \breve{\lambda}_\alpha - s_T^{(2)}\right\} \cap \mathcal{T}_1\right) + \mathbb{P}(\mathcal{T}_1^c) + o(1)\\
&\leq \mathbb{P}\left(\left\{\left\|\frac{U^\top \mathcal{E}}{T}\right\|_\infty > \frac{2}{\sqrt{T}}\mu^*_{\alpha+s_T^{(14)}}\right\} \cap \mathcal{T}_1\right) + \mathbb{P}(\mathcal{T}_1^c) + o(1)\\
&= 0 + \mathbb{P}(\mathcal{T}_1^c) + o(1) = \alpha + o(1),
\end{aligned}$$

where, on the last line, we used $\mathcal{S}_{\mu^*_{\alpha+s_T^{(14)}}} \subset \mathcal{T}_1$.

Let us therefore prove that, on $\mathcal{T}_1$, (OA.6) holds. To do so, we show that, on $\mathcal{T}_1$,

$$\mathbb{P}_e\left(\breve{\Pi}(\mu, e) > \mu\right) > \alpha \tag{OA.7}$$

for $\mu = (1+s_T^{(6)})\mu^*_{\alpha+s_T^{(14)}} + \frac{\sqrt{T}}{2}s_T^{(2)} > \mu^*_{\alpha+s_T^{(14)}} + \frac{\sqrt{T}}{2}s_T^{(2)}$, which implies that (OA.6) is true by definition of $\check{\lambda}_\alpha$ and the fact that $\check{\mu}_\alpha = \frac{\sqrt{T}}{2}\check{\lambda}_\alpha$. We have

$$\begin{aligned}
\mathbb{P}_e\left(\check{\Pi}(\mu,e) > \mu\right) &\geq \mathbb{P}_e\left(\Pi(e) - R(\mu,e) > \mu\right) \\
&\geq \mathbb{P}_e\left(\Pi(e) - R(\mu,e) > \mu, R(\mu,e) \leq s_T^{(6)}\sqrt{\mu} + s_T^{(7)}\right) \\
&\geq \mathbb{P}_e\left(\Pi(e) > \mu + s_T^{(6)}\sqrt{\mu} + s_T^{(7)}\right) - \mathbb{P}_e\left(R(\mu,e) > s_T^{(6)}\sqrt{\mu} + s_T^{(7)}\right) \\
&\geq \mathbb{P}_e\left(\Pi(e) > \mu + s_T^{(6)}\sqrt{\mu} + s_T^{(7)}\right) - \frac{2}{T},
\end{aligned}$$

where, on the last line, we used Lemma D.6 and the facts that $\mu \geq \mu^*_{\alpha+s_T^{(14)}}$ and we work on $\mathcal{S}_T^{(3)} \cap \mathcal{S}_T^{(4)} \cap \mathcal{S}_{\mu^*_{\alpha+s_T^{(14)}}} \subset \mathcal{T}_1$ to obtain that $\mathbb{P}_e\left(R(\mu,e) > s_T^{(6)}\sqrt{\mu} + s_T^{(7)}\right) \leq \frac{2}{T}$. By Lemma D.2, we obtain

$$\mathbb{P}_e(\Pi(e) > \mu + s_T^{(6)}\sqrt{\mu} + s_T^{(7)}) \geq \mathbb{P}(\Pi^G \geq \mu + s_T^{(6)}\sqrt{\mu} + s_T^{(7)}) - s_T^{(8)} - \frac{2}{T}. \tag{OA.8}$$

Since $\sqrt{\mu} \leq (1+\mu)$, for T large enough, it holds that

$$\begin{aligned}
&\mathbb{P}\left((\Pi^G > \mu + s_T^{(6)}\sqrt{\mu} + s_T^{(7)}\right) \\
&\geq \mathbb{P}\left(\Pi^G > \mu + s_T^{(6)}(1+\mu) + s_T^{(7)}\right) \\
&\geq \mathbb{P}\left(\Pi^G > (1+s_T^{(6)})\mu^*_{\alpha+s_T^{(14)}} + \frac{\sqrt{T}}{2}s_T^{(2)} + s_T^{(6)}\left(1 + (1+s_T^{(6)})\mu^*_{\alpha+s_T^{(14)}} + \frac{\sqrt{T}}{2}s_T^{(2)}\right) + s_T^{(7)}\right) \\
&\geq \mathbb{P}\left(\Pi^G > \mu^*_{\alpha+s_T^{(14)}} + s_T^{(12)}\right) && \text{(OA.9)} \\
&\geq \mathbb{P}\left(\Pi^G > \mu^*_{\alpha+s_T^{(14)}}) - \mathbb{P}(|\Pi^G - \mu^*_{\alpha+s_T^{(14)}}| \leq s_T^{(12)}\right) \\
&\geq \mathbb{P}\left(\Pi^G > \mu^*_{\alpha+s_T^{(14)}}\right) - K_3 s_T^{(12)}\sqrt{1 \vee \log\left(2p/s_T^{(12)}\right)} && \text{(OA.10)} \\
&\geq \mathbb{P}\left(\Pi^* > \mu^*_{\alpha+s_T^{(14)}}\right) - s_T^{(9)} - K_3 s_T^{(12)}\sqrt{1 \vee \log\left(2p/s_T^{(12)}\right)} && \text{(OA.11)} \\
&= \alpha + s_T^{(14)} - s_T^{(9)} - K_3 s_T^{(12)}\sqrt{1 \vee \log\left(2p/s_T^{(12)}\right)},
\end{aligned}$$

where, in (OA.9), we used Lemma D.4 and the fact that $s_T^{(14)} \to 0$ by Lemma D.8 (iv), (v) and (vi), to obtain that $\mu^*_{\alpha+s_T^{(14)}} \leq s_T^{(11)}$ for T large enough, in (OA.10), we leveraged Lemma D.3 and (OA.11) follows from Lemma D.1. This and (OA.8), therefore yield

$$\mathbb{P}_e\left(\Pi(e) > \mu\right) \geq \alpha + s_T^{(14)} - s_T^{(9)} - K_3 s_T^{(12)}\sqrt{1 \vee \log\left(2p/s_T^{(12)}\right)} - s_T^{(8)} - \frac{2}{T} = \alpha + \frac{1}{T} > \alpha,$$

by definition of $s_T^{(14)}$. This shows (OA.7) and therefore concludes the proof of (i).

Proof of (ii). We want to show that if $\sqrt{\frac{\log(T\vee p)}{T\wedge p}} = o_P\left(\left\|\frac{U^\top U\beta^*}{T}\right\|_\infty\right)$, we have

$$\mathbb{P}\left(\left\|\frac{\widehat{U}^\top \widehat{Y}}{T}\right\|_\infty > \widehat{\lambda}_\alpha\right) \to 1. \tag{OA.12}$$

As in the proof of (i), we first move from quantities depending on $\widehat{K}$ to quantities only depending on K. Notice that

$$\begin{aligned}
&\mathbb{P}\left(\left\|\frac{\widehat{U}^\top \widetilde{Y}}{T}\right\|_\infty > \widehat{\lambda}_\alpha\right)\\
&\geq \mathbb{P}\left(\left\{\left\|\frac{\widehat{U}^\top \widehat{Y}}{T}\right\|_\infty > \widehat{\lambda}_\alpha\right\} \cap \left\{\widehat{K} = K\right\}\right)\\
&= \mathbb{P}\left(\left\{\left\|\frac{\check{U}^\top \check{Y}}{T}\right\|_\infty > \check{\lambda}_\alpha\right\} \cap \left\{\widehat{K} = K\right\}\right)\\
&= \mathbb{P}\left(\left\|\frac{\check{U}^\top \check{Y}}{T}\right\|_\infty > \check{\lambda}_\alpha\right) - \mathbb{P}\left(\left\{\left\|\frac{\check{U}^\top \check{Y}}{T}\right\|_\infty > \check{\lambda}_\alpha\right\} \cap \left\{\widehat{K} \neq K\right\}\right) &\text{(OA.13)}\\
&\geq \mathbb{P}\left(\left\|\frac{\check{U}^\top \check{Y}}{T}\right\|_\infty > \check{\lambda}_\alpha\right) - \mathbb{P}\left(\widehat{K} \neq K\right)\\
&\geq \mathbb{P}\left(\left\|\frac{\check{U}^\top \check{Y}}{T}\right\|_\infty > \check{\lambda}_\alpha\right) + o(1), &\text{(OA.14)}
\end{aligned}$$

where, in (OA.13), we used the formula $\mathbb{P}(C) = \mathbb{P}(C \cap D) + \mathbb{P}(C \cap D^c)$, and (OA.14) follows from Assumption 1. Hence, it holds that

$$\begin{aligned}
&\mathbb{P}\left(\left\|\frac{\widehat{U}^\top \widetilde{Y}}{T}\right\|_\infty > \widehat{\lambda}_\alpha\right)\\
&\geq \mathbb{P}\left(\left\|\frac{\check{U}^\top \check{Y}}{T}\right\|_\infty > \check{\lambda}_\alpha\right)\\
&\geq \mathbb{P}\left(\left\|\frac{U^\top U\beta^*}{T}\right\|_\infty - \left\|\frac{U^\top \mathcal{E}}{T}\right\|_\infty - \left\|\frac{\check{U}^\top(\check{Y} - \check{U}\beta^*)}{T} - \frac{U^\top \mathcal{E}}{T}\right\|_\infty > \check{\lambda}_\alpha\right)\\
&\geq \mathbb{P}\left(\left\{\left\|\frac{U^\top U\beta^*}{T}\right\|_\infty > \check{\lambda}_\alpha + s_T^{(2)} + s_T^{(5)}\right\} \cap \mathcal{S}_T^{(2)} \cap \mathcal{S}_T^{(5)}\right)\\
&\geq \mathbb{P}\left(\left\|\frac{U^\top U\beta^*}{T}\right\|_\infty > \check{\lambda}_\alpha + s_T^{(2)} + s_T^{(5)}\right) - \mathbb{P}\left(\left(\mathcal{S}_T^{(2)} \cap \mathcal{S}_T^{(5)}\right)^c\right)\\
&= \mathbb{P}\left(\left\|\frac{U^\top U\beta^*}{T}\right\|_\infty > \check{\lambda}_\alpha + s_T^{(2)} + s_T^{(5)}\right) + o(1), &\text{(OA.15)}
\end{aligned}$$

where, in the last line, we used Lemma D.5.

Let us define

$$\mathcal{T}_2 = \mathcal{S}_{\mu^*_{2s_T^{(10)}}} \cap \mathcal{S}_T^{(1)} \cap \mathcal{S}_T^{(2)} \cap \mathcal{S}_T^{(3)} \cap \mathcal{S}_T^{(4)} \cap \mathcal{S}_T^{(5)}.$$

Note that, by Lemmas D.5 and D.7, and the fact that $s_T^{(10)} = \frac{1}{T \vee p} + s_T^{(9)} \to 0$ by Lemma D.8 (v), the event $\mathcal{T}_2$ has probability going to 1.

Hence, by (OA.15), to show (OA.12), it suffices to prove that, on $\mathcal{T}_2$, we have

$$\check{\lambda}_\alpha \leq \frac{2}{\sqrt{T}} \mu^*_{2s_T^{(10)}}, \tag{OA.16}$$

for T large enough. Indeed, in this case, we would have

$$\begin{aligned}
\mathbb{P}\left(\left\|\frac{\check{U}^\top \check{Y}}{T}\right\|_\infty \geq \check{\lambda}_\alpha\right) &\geq \mathbb{P}\left(\left\|\frac{U^\top U \beta^*}{T}\right\|_\infty > \check{\lambda}_\alpha + s_T^{(2)} + s_T^{(5)}\right) + o(1), \\
&\geq \mathbb{P}\left(\left\{\left\|\frac{U^\top U \beta^*}{T}\right\|_\infty > \frac{2}{\sqrt{T}} \mu^*_{2s_T^{(10)}} + s_T^{(2)} + s_T^{(5)}\right\} \cap \mathcal{T}_2\right) + o(1) \\
&\geq \mathbb{P}\left(\left\{\left\|\frac{U^\top U \beta^*}{T}\right\|_\infty > \frac{2}{\sqrt{T}} s_T^{(11)} + s_T^{(2)} + s_T^{(5)}\right\} \cap \mathcal{T}_2\right) + o(1) \\
&\geq \mathbb{P}\left(\left\|\frac{U^\top U \beta^*}{T}\right\|_\infty > \frac{2}{\sqrt{T}} s_T^{(11)} + s_T^{(2)} + s_T^{(5)}\right) - \mathbb{P}(\mathcal{T}_2^c) + o(1) \to 1,
\end{aligned}$$

where, in the third line, we used $\mu^*_{2s_T^{(10)}} \leq s_T^{(11)}$ by Lemma D.4 and, in the last line, we leveraged the facts $\frac{2}{\sqrt{T}} s_T^{(11)} + s_T^{(2)} + s_T^{(5)} = O_P\left(\sqrt{\frac{\log(T \vee p)}{T \wedge p}}\right)$ by Lemma D.8 (ii) and that $\sqrt{\frac{\log(T \vee p)}{T \wedge p}} = o_P\left(\left\|\frac{U^\top U \beta^*}{T}\right\|_\infty\right)$ to obtain that $\mathbb{P}\left(\left\|\frac{U^\top U \beta^*}{T}\right\|_\infty > \frac{2}{\sqrt{T}} s_T^{(11)} + s_T^{(2)} + s_T^{(5)}\right) \to 1$.

Let us therefore prove that, on $\mathcal{T}_2$, (OA.16) holds for T large enough. To do so, we show that, on $\mathcal{T}_2$, for T large enough,

$$\mathbb{P}_e\left(\check{\Pi}(\mu^*_{2s_T^{(10)}}, e) > \mu^*_{2s_T^{(10)}}\right) \leq \alpha, \tag{OA.17}$$

which implies (OA.16) by definition of $\check{\lambda}_\alpha$. On $\mathcal{T}_2$, we have

$$\begin{aligned}
\mathbb{P}_e\left(\check{\Pi}(\mu^*_{2s_T^{(10)}}, e) > \mu^*_{2s_T^{(10)}}\right) &\leq \mathbb{P}_e\left(\Pi(e) + R(\mu^*_{2s_T^{(10)}}, e) > \mu\right) \\
&\leq \mathbb{P}_e\left(\Pi(e) > \mu^*_{2s_T^{(10)}} - R(\mu^*_{2s_T^{(10)}}, e), R(\mu^*_{2s_T^{(10)}}, e) \leq s_T^{(6)} \sqrt{\mu^*_{2s_T^{(10)}}} + s_T^{(7)}\right) \\
&\quad + \mathbb{P}_e\left(R(\mu^*_{2s_T^{(10)}}, e) > s_T^{(6)} \sqrt{\mu^*_{2s_T^{(10)}}} + s_T^{(7)}\right) \\
&\leq \mathbb{P}_e\left(\Pi(e) > \mu^*_{2s_T^{(10)}} - s_T^{(6)} \sqrt{\mu^*_{2s_T^{(10)}}} - s_T^{(7)}\right) + \frac{2}{T},
\end{aligned}$$

where, in the last line, we used Lemma D.6. By Lemma D.2, we obtain

$$\mathbb{P}_e\left(\Pi(e) > \mu^*_{2s_T^{(10)}}\right) \leq \mathbb{P}\left(\Pi^G \geq \mu^*_{2s_T^{(10)}} - s_T^{(6)}\sqrt{\mu^*_{2s_T^{(10)}}} - s_T^{(7)}\right) + s_T^{(8)} + \frac{2}{T}. \tag{OA.18}$$

Since $\sqrt{\mu^*_{2s_T^{(10)}}} \leq 1 + \mu^*_{2s_T^{(10)}}$, it holds that

$$\begin{aligned}
&\mathbb{P}\left(\Pi^G > \mu^*_{2s_T^{(10)}} - s_T^{(6)}\sqrt{\mu^*_{2s_T^{(10)}}} - s_T^{(7)}\right) \\
&\leq \mathbb{P}\left(\Pi^G > \mu^*_{2s_T^{(10)}} - s_T^{(6)}(1 + \mu^*_{2s_T^{(10)}}) - s_T^{(7)}\right) \\
&\leq \mathbb{P}\left(\Pi^G > \mu^*_{2s_T^{(10)}} - s_T^{(6)}(1 + s_T^{(11)}) - s_T^{(7)}\right) && \text{(OA.19)} \\
&= \mathbb{P}\left(\Pi^G > \mu^*_{2s_T^{(10)}} - s_T^{(13)}\right) \\
&\leq \mathbb{P}\left(\Pi^G > \mu^*_{2s_T^{(10)}}\right) + \mathbb{P}\left(|\Pi^G - \mu^*_{2s_T^{(10)}}| \leq s_T^{(13)}\right) \\
&\leq \mathbb{P}\left(\Pi^G > \mu^*_{2s_T^{(10)}}\right) + K_3 s_T^{(13)}\sqrt{1 \vee \log\left(2p/s_T^{(13)}\right)} && \text{(OA.20)} \\
&\leq \mathbb{P}\left(\Pi^* > \mu^*_{2s_T^{(10)}}\right) + s_T^{(9)} + K_3 s_T^{(13)}\sqrt{1 \vee \log\left(2p/s_T^{(13)}\right)} && \text{(OA.21)} \\
&= 2s_T^{(10)} + s_T^{(9)} + K_3 s_T^{(13)}\sqrt{1 \vee \log\left(2p/s_T^{(13)}\right)},
\end{aligned}$$

where, in (OA.19), we used Lemma D.4 to obtain that $\mu^*_{2s_T^{(10)}} \leq s_T^{(11)}$, in (OA.20), we leveraged Lemma D.3 and (OA.21) follows from Lemma D.1. This and (OA.18), therefore yield

$$\mathbb{P}_e\left(\Pi(e) > \mu^*_{2s_T^{(10)}}\right) \leq 2s_T^{(10)} + s_T^{(9)} + K_3 s_T^{(13)}\sqrt{1 \vee \log\left(2p/s_T^{(13)}\right)} + s_T^{(8)} + \frac{2}{T} \leq \alpha,$$

for T large enough by Lemma D.8 (iv), (v), (vi). This shows that (OA.17) holds and therefore concludes the proof of (ii).

D.3 Auxiliary lemmas on distributions

Lemma D.1 *Under the assumptions of Theorem 1, it holds that*

$$\sup_{z\in\mathbb{R}}\left|\mathbb{P}(\Pi^* \leq z) - \mathbb{P}\left(\Pi^G \leq z\right)\right| < s_T^{(9)}.$$

Proof. The result is a direct consequence of Lemma D.18 applied to $Z_t = u_t\varepsilon_t$ (and the constant K_1 used in the definition of $s_T^{(9)}$ is introduced in Lemma D.18). Condition (i) of

Lemma D.18 is satisfied by Lemma D.15 and Assumption 3 (iii). Assumption 4 implies that conditions (ii) and (iii) hold. Concerning condition (iv), note that, by Assumption 3 (iv), we have

$$\mathbb{E}\left[\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}u_t\varepsilon_t\right)\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}u_t\varepsilon_t\right)^{\top}\right]=\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{E}\left[u_t\varepsilon_t u_s^{\top}\varepsilon_s\right]=\mathbb{E}\left[u_t u_t^{\top}\varepsilon_t^2\right].$$

By Assumption 3 (ii), this implies that

$$\sigma_p\left(\mathbb{E}\left[\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}u_t\varepsilon_t\right)\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}u_t\varepsilon_t\right)^{\top}\right]\right)=\sigma_p\left(E\left[u_t u_t^{\top}\varepsilon_t^2\right]\right)\geq\kappa_1>0,$$

and therefore that condition (iv) holds. □

Lemma D.2 *Let the assumptions of Theorem 1 hold. On the event* $\mathcal{S}_T^{(1)}$,

$$\sup_{z\in\mathbb{R}}\left|\mathbb{P}_e(\Pi(e)\leq z)-\mathbb{P}\left(\Pi^G\leq z\right)\right|\leq s_T^{(8)}.$$

Proof. Conditionally on $U,\mathcal{E}$, $W(e)$ is a centered Gaussian vector with covariance matrix $T^{-1}\sum_{t=1}^{T}u_t u_t^{\top}\varepsilon_t^2$. Moreover, G is a centered Gaussian vector with covariance matrix

$$\mathbb{E}\left[\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}u_t\varepsilon_t\right)\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}u_t\varepsilon_t\right)^{\top}\right]=\mathbb{E}\left[\varepsilon_t^2 u_t u_t^{\top}\right],$$

see the proof of Lemma D.1 for a justification of this equality. Remark that, by Assumption 3 (ii),

$$\kappa_2>\mathbb{E}\left[u_{tj}u_{tj}^{\top}\varepsilon_t^2\right]>\kappa_1,$$

for all $j\in[p]$. We can therefore apply Lemma D.20 to get

$$\sup_{z\in\mathbb{R}}\left|\mathbb{P}_e(\Pi(e)\leq z)-\mathbb{P}\left(\Pi^G\leq z\right)\right|\leq\pi(\Delta),$$

where $\pi(\Delta)=K_1\Delta^{1/3}(1\vee\log(2p)\vee\log(1/\Delta))^{1/3}\log(2p)^{1/3}$.

□

Lemma D.3 *Under the assumptions of Theorem 1, there exists a constant* $K_2>0$ *such that, for all* $z_1,z_2>0$, *we have*

$$\mathbb{P}\left(\left|\Pi^G-z_1\right|\leq z_2\right)\leq K_2 z_2\sqrt{1\vee\log(2p/z_2)}.$$

Proof. This is a direct consequence of Lemma D.19 of which the conditions are satisfied by Assumption 3 (see the proofs of Lemmas D.1 and D.2 for more details). □

Lemma D.4 *For every $\alpha > s_T^{(10)}$, we have*

$$\mu_\alpha^* \leq s_T^{(11)}.$$

Proof. Notice that, by Assumption 3 (iv),

$$\mathbb{E}\left[(W_j^*)^2\right] = \mathbb{E}\left[\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} u_{tj}\varepsilon_t\right)^2\right] = \mathbb{E}\left[u_{tj}^2\varepsilon_t^2\right],$$

which, by Assumption 3 (ii) and (iv), is bounded uniformly in j and t by $\kappa_2 > 0$. Using Lemma 7 in Chernozhukov et al. (2015) and remark A.8 in Lederer & Vogt (2021), we have, for every $r > 0$,

$$\mathbb{P}\left(\|G\|_\infty \geq \mathbb{E}\left[\|G\|_\infty\right] + r\right) \leq \exp\left(-\frac{r^2}{2\kappa_2}\right).$$

Taking $r = \kappa_2\sqrt{2\log(T \vee p)}$, we get

$$\mathbb{P}\left(\|G/\kappa_2\|_\infty \geq \mathbb{E}\left[\|G/\kappa_2\|_\infty\right] + \sqrt{2\log(T \vee p)}\right) \leq \frac{1}{T \vee p}.$$

By the Gaussian maximal inequality (see e.g., Exercise 2.17 in Boucheron et al. (2013)), it holds that $\mathbb{E}\left[\|G/\kappa_2\|_\infty\right] \leq \sqrt{2\log(2p)}$, which yields

$$\mathbb{P}\left(\|G\|_\infty \geq \kappa_2\left(\sqrt{2\log(2p)} + \sqrt{2\log(T \vee p)}\right)\right) \leq \frac{1}{T \vee p},$$

so that $\mu_\alpha^G \leq \kappa_2\left(\sqrt{2\log(2p)} + \sqrt{2\log(T \vee p)}\right)$ for $\alpha > 1/(T \vee p)$ by definition of μ_α^G. Now, for $\alpha > s_T^{(10)} = (T \vee p)^{-1} + s_T^{(9)}$, by Lemma D.1, we have

$$\mathbb{P}\left(\Pi^* \geq \mu^G_{\alpha - s_T^{(9)}}\right) \leq \mathbb{P}\left(\Pi^G \geq \mu^G_{\alpha - s_T^{(9)}}\right) + s_T^{(9)} \leq \alpha - s_T^{(9)} + s_T^{(9)} = \alpha.$$

Hence, we obtain $\mu_\alpha^* \leq \mu^G_{\alpha - s_T^{(9)}} \leq \kappa_2\left(\sqrt{2\log(2p)} + \sqrt{2\log(T \vee p)}\right)$. □

D.4 Auxiliary lemmas on probabilistic events

Lemma D.5 *Under the assumptions of Theorem 1, it holds that*

(i) $\mathbb{P}\left(\mathcal{S}_T^{(1)}\right) \to 1$;

(ii) $\mathbb{P}\left(\mathcal{S}_T^{(2)}\right) \to 1$;

(iii) $\mathbb{P}\left(\mathcal{S}_T^{(3)}\right) \to 1$;

(iv) $\mathbb{P}\left(\mathcal{S}_T^{(4)}\right) \to 1$;

(v) $\mathbb{P}\left(\mathcal{S}_T^{(5)}\right) \to 1$.

Proof.

Result (i) follows directly from Lemma D.9 (v); (ii) is a consequence of Lemma D.14; (iii) comes from Lemmas D.10 (vi) and Lemma D.9 (i) and the triangle inequality, (iv) follows from Lemma D.13 and (v) is a direct consequence of Lemma D.9 (iii). □

Lemma D.6 *Let the assumptions of Theorem 1 hold. On the event* $\mathcal{S}_T^{(3)} \cap \mathcal{S}_T^{(4)} \cap \mathcal{S}_\mu$, *we have, for all* $\mu' \geq \mu$,

$$\mathbb{P}_e\left(R(\mu', e) \geq s_T^{(6)}\sqrt{\mu'} + s_T^{(7)}\right) \leq \frac{2}{T}.$$

Proof. Take $\mu' \geq \mu$. Remember that $\breve{Y} = \left(I_T - \breve{P}\right)(X\beta + F\varphi^* + \mathcal{E})$. This yields that

$$\breve{\varepsilon}_{\frac{2}{\sqrt{T}}\mu',t} = \widetilde{y}_t - \breve{u}_t^\top \breve{\beta}_{\frac{2}{\sqrt{T}}\mu'} = \breve{u}_t\left(\beta^* - \breve{\beta}_{\frac{2}{\sqrt{T}}\mu'}\right) + \widetilde{f}_t^\top\varphi^* + \widetilde{\varepsilon}_t,$$

where we recall that $\widetilde{\varepsilon}_t$ is the t^{th} element of $\left(I_T - \breve{P}\right)\mathcal{E}$ and $\widetilde{f}_t$ is the $K \times 1$ vector corresponding to the t^{th} row of $\left(I_T - \breve{P}\right)F$. This yields

$$\begin{aligned}
&R(\mu', e)\\
&= \frac{1}{\sqrt{T}}\left\|\breve{W}(\mu', e) - W(e)\right\|_\infty\\
&= \frac{1}{T}\max_{j\in[p]}\left|\sum_{t=1}^T \breve{u}_{tj}\breve{\varepsilon}_{\frac{2}{\sqrt{T}}\mu',t}e_t - \sum_{t=1}^T u_{tj}\varepsilon_t e_t\right|\\
&\leq \frac{1}{T}\max_{j\in[p]}\left|\sum_{t=1}^T \breve{u}_{tj}\breve{u}_t^\top\left(\beta^* - \breve{\beta}_{\frac{2}{\sqrt{T}}\mu'}\right)e_t\right| + \frac{1}{T}\max_{j\in[p]}\left|\sum_{t=1}^T\left(\breve{u}_{tj}\widetilde{\varepsilon}_t + \widetilde{f}_t^\top\varphi^* - u_{tj}\varepsilon_t\right)e_t\right|. \quad \text{(OA.22)}
\end{aligned}$$

Now, we bound the two terms in (OA.22). We start with $\max_{j\in[p]}\left|\sum_{t=1}^{T}\breve{u}_{tj}\breve{u}_t^\top\left(\beta^*-\breve{\beta}_{\frac{2}{\sqrt{T}}\mu'}\right)e_t\right|$. Remark that given $(F,U,\mathcal{E})$, we have

$$\frac{1}{T}\sum_{t=1}^{T}\breve{u}_{tj}\breve{u}_t^\top\left(\breve{\beta}_\lambda-\beta^*\right)e_t\sim\mathcal{N}\left(0,\frac{1}{T^2}\sum_{t=1}^{T}\left(\breve{u}_{tj}\breve{u}_t^\top\left(\breve{\beta}_\lambda-\beta^*\right)\right)^2\right)$$

By the Gaussian tail bound (equation (2.10) in Vershynin (2018)), for $z>0$, we obtain, for all $j\in[p]$ and $z>0$,

$$\mathbb{P}_e^*\left(\left|\frac{1}{T}\sum_{t=1}^{T}\breve{u}_{tj}\breve{u}_t^\top\left(\breve{\beta}_\lambda-\beta^*\right)e_t\right|>z\right)\leq 2\exp\left(-\frac{z^2}{\frac{1}{T^2}\sum_{t=1}^{T}\left(\breve{u}_{tj}\breve{u}_t^\top\left(\breve{\beta}_\lambda-\beta^*\right)\right)^2}\right).\tag{OA.23}$$

Next, let $\lambda=\frac{2}{\sqrt{T}}\mu$ and $\lambda'=\frac{2}{\sqrt{T}}\mu'$. By definition of $\breve{\beta}_{\lambda'}$, it holds that

$$\frac{1}{T}\left\|\breve{Y}-\breve{U}\breve{\beta}_{\lambda'}\right\|_2^2+\lambda'\left\|\breve{\beta}_\lambda\right\|_1\leq\frac{1}{T}\left\|\breve{Y}-\breve{U}\beta^*\right\|_2^2+\lambda\|\beta^*\|_1.$$

This yields

$$\begin{aligned}
&\frac{1}{T}\left\|\breve{U}(\beta^*-\breve{\beta}_{\lambda'})\right\|_2^2\\
&\leq\frac{2}{T}\left(\breve{Y}-\breve{U}\beta^*\right)^\top\breve{U}\left(\breve{\beta}_{\lambda'}-\beta^*\right)+\lambda'\left(\|\beta^*\|_1-\left\|\breve{\beta}_{\lambda'}\right\|_1\right)\\
&\leq\frac{2}{T}\left\|\breve{U}^\top\left(\breve{Y}-\breve{U}\beta^*\right)\right\|_\infty\left\|\breve{\beta}_{\lambda'}-\beta^*\right\|_1+\lambda'\left(\|\beta^*\|_1-\left\|\breve{\beta}_{\lambda'}\right\|_1\right)\\
&\leq\lambda'\|\breve{\beta}_{\lambda'}-\beta^*\|_1+\lambda'\left(\|\beta^*\|_1-\left\|\breve{\beta}_{\lambda'}\right\|_1\right)\\
&\leq 2\lambda'\|\beta^*\|_1.
\end{aligned}\tag{OA.24}$$

where we used Hölder's inequality and the fact that we work on $\mathcal{S}_\mu$. Moreover, we have

$$\begin{aligned}
\frac{1}{T^2}\sum_{t=1}^{T}(\breve{u}_{tj}\breve{u}_t^\top(\breve{\beta}_\lambda-\beta^*))^2&\leq\frac{1}{T}\sum_{t=1}^{T}\breve{u}_{tj}^2\frac{1}{T}\left\|\breve{U}\left(\beta^*-\breve{\beta}_\lambda\right)\right\|_2^2\\
&\leq s_T^{(3)}2\lambda'\|\beta^*\|_1,
\end{aligned}\tag{OA.25}$$

by (OA.24) and because we work on $\mathcal{S}_T^{(3)}$. Recall that $s_T^{(6)}=2\sqrt{\log(Tp)\|\beta^*\|_1 s_T^{(3)}T^{-1/2}}$. Using (OA.23), (OA.25) and the union bound, we get

$$\begin{aligned}
&\mathbb{P}_e^*\left(\frac{1}{T}\max_{j\in[p]}\left|\sum_{t=1}^{T}\breve{u}_{tj}\breve{u}_t^\top\left(\beta^*-\breve{\beta}_{\frac{2}{\sqrt{T}}\mu'}\right)e_t\right|_\infty>s_T^{(6)}\sqrt{\mu'}\right)\\
&\leq p\max_{j\in[p]}\mathbb{P}_e^*\left(\left|\frac{1}{T}\sum_{t=1}^{T}\breve{u}_{tj}\breve{u}_t^\top\left(\breve{\beta}_\lambda-\beta^*\right)e_t\right|>s_T^{(6)}\sqrt{\mu'}\right)
\end{aligned}$$

$$\leq \exp\left(-\frac{(s_T^{(6)})^2\mu'}{2\lambda'\|\beta^*\|_1 s_T^{(3)}} + \log(p)\right) = T^{-1}. \tag{OA.26}$$

Let us now bound the term $\max_{j\in[p]}\left|\sum_{t=1}^T\left(\breve{u}_{tj}\widetilde{\varepsilon}_t + \widetilde{f}_t^\top\varphi^* - u_{tj}\varepsilon_t\right)e_t\right|$. Conditional on $(F, U, \mathcal{E})$, we have

$$\frac{1}{T}\sum_{t=1}^T\left(\breve{u}_{tj}\widetilde{\varepsilon}_t + \widetilde{f}_t^\top\varphi^* - u_{tj}\varepsilon_t\right)e_t \sim \mathcal{N}\left(0, \frac{1}{T^2}\sum_{t=1}^T\left(\breve{u}_{tj}\widetilde{\varepsilon}_t + \widetilde{f}_t^\top\varphi^* - u_{tj}\varepsilon_t\right)^2\right).$$

Since we work on $\mathcal{S}_T^{(4)}$, by the Gaussian tail bound, this yields, for all $j \in [p]$ and $z > 0$,

$$\mathbb{P}_e^*\left(\left|\frac{1}{T}\sum_{t=1}^T\left(\breve{u}_{tj}\widetilde{\varepsilon}_t + \widetilde{f}_t^\top\varphi^* - u_{tj}\varepsilon_t\right)e_t\right| > z\right) \leq \exp\left(-\frac{Tz^2}{s_T^{(4)}}\right).$$

Recall that $s_T^{(7)} = \sqrt{\log(Tp)T^{-1}s_T^{(4)}}$. Using the union bound, we get

$$\begin{aligned}
&\mathbb{P}_e^*\left(\max_{j\in[p]}\left|\frac{1}{T}\sum_{t=1}^T\left(\breve{u}_{tj}\widetilde{\varepsilon}_t + \widetilde{f}_t^\top\varphi^* - u_{tj}\varepsilon_t\right)e_t\right| > s_T^{(7)}\right)\\
&\leq p\max_{j\in[p]}\mathbb{P}_e\left(\left|\frac{1}{T}\sum_{t=1}^T\left(\breve{u}_{tj}\widetilde{\varepsilon}_t + \widetilde{f}_t^\top\varphi^* - u_{tj}\varepsilon_t\right)e_t\right| > s_T^{(7)}\right)\\
&\leq p\exp\left(-\frac{\left(s_T^{(7)}\right)^2}{T^{-1}s_T^{(4)}}\right) = T^{-1}.
\end{aligned} \tag{OA.27}$$

Using the pigeonhole principle, (OA.22), (OA.26) and (OA.27), we get $\mathbb{P}_e^*\left(R(\mu', e) \geq s_T^{(6)}\sqrt{\mu'} + s_T^{(7)}\right) \leq 2T^{-1}$, which yields $\mathbb{P}_e\left(R(\mu', e) \geq s_T^{(6)}\sqrt{\mu'} + s_T^{(7)}\right) \leq 2T^{-1}$, integrating over the distribution of $(F, U, \mathcal{E})$.

□

Lemma D.7 *Under the assumptions of Theorem 1, we have*

$$\sup_{\alpha'\in(0,1)}\left|\mathbb{P}\left(\mathcal{S}_{\mu^*_{\alpha'}}\right) - (1-\alpha')\right| = o(1).$$

Proof. Let us first bound $\mathbb{P}\left(\mathcal{S}_{\mu^*_{\alpha'}}\right)$ from above. For $\alpha' \in (0,1)$, we have

$$\mathbb{P}\left(\mathcal{S}_{\mu^*_{\alpha'}}\right) = \mathbb{P}\left(2\left\|\frac{\breve{U}^\top(\breve{Y} - \breve{U}\beta^*)}{T}\right\|_\infty \leq \frac{2}{\sqrt{T}}\mu^*_{\alpha'}\right)$$

$$\leq \mathbb{P}\left(\left\|\frac{U^\top \mathcal{E}}{T}\right\|_\infty - \left\|\frac{\check{U}^\top(\check{Y}-\check{U}\beta^*)}{T} - \frac{U^\top \mathcal{E}}{T}\right\|_\infty \leq \frac{1}{\sqrt{T}}\mu^*_{\alpha'}\right)$$
$$\leq \mathbb{P}\left(\left\{\left\|\frac{U^\top \mathcal{E}}{T}\right\|_\infty - \left\|\frac{\check{U}^\top(\check{Y}-\check{U}\beta^*)}{T} - \frac{U^\top \mathcal{E}}{T}\right\|_\infty \leq \frac{1}{\sqrt{T}}\mu^*_{\alpha'}\right\} \cap \mathcal{S}_T^{(2)}\right) + \mathbb{P}\left(\left(\mathcal{S}_T^{(2)}\right)^c\right)$$
$$\leq \mathbb{P}\left(\left\|\frac{U^\top \mathcal{E}}{T}\right\|_\infty \leq \frac{1}{\sqrt{T}}\mu^*_{\alpha'} + s_T^{(2)}\right) + \mathbb{P}\left(\left(\mathcal{S}_T^{(2)}\right)^c\right). \tag{OA.28}$$

Now, we have

$$\begin{aligned}
\mathbb{P}\left(\left\|\frac{U^\top \mathcal{E}}{T}\right\|_\infty \leq \frac{1}{\sqrt{T}}\mu^*_{\alpha'} + s_T^{(2)}\right) &= \mathbb{P}\left(\Pi^* \leq \mu^*_{\alpha'} + \sqrt{T}s_T^{(2)}\right) \\
&\leq \mathbb{P}\left(\Pi^G \leq \mu^*_{\alpha'} + \sqrt{T}s_T^{(2)}\right) + s_T^{(9)} \\
&\leq \mathbb{P}\left(\Pi^G \leq \mu^*_{\alpha'}\right) + \mathbb{P}\left(\left|\Pi^G - \mu^*_{\alpha'}\right| \leq \sqrt{T}s_T^{(2)}\right) + s_T^{(9)} \\
&\leq \mathbb{P}\left(\Pi^* \leq \mu^*_{\alpha'}\right) + \mathbb{P}\left(\left|\Pi^G - \mu^*_{\alpha'}\right| \leq \sqrt{T}s_T^{(2)}\right) + 2s_T^{(9)} \\
&\leq 1-\alpha' + \mathbb{P}\left(\left|\Pi^G - \mu^*_{\alpha'}\right| \leq \sqrt{T}s_T^{(2)}\right) + 2s_T^{(9)}, \qquad \text{(OA.29)}
\end{aligned}$$

where we used Lemma D.1 in the second and fourth lines. By Lemma D.3, we have

$$\mathbb{P}\left(\left|\Pi^G - \mu^*_{\alpha'}\right| \leq s_T^{(2)}\right) \leq K_4\sqrt{T}s_T^{(2)}\sqrt{1 \vee \log\left(\frac{2p}{\sqrt{T}s_T^{(2)}}\right)}.$$

Combining this, (OA.28) and (OA.29), we get

$$\mathbb{P}\left(\mathcal{S}_{\mu^*_{\alpha'}}\right) \leq 1-\alpha' + \mathbb{P}\left(\left(\mathcal{S}_T^{(2)}\right)^c\right) + K_4\sqrt{T}s_T^{(2)}\sqrt{1 \vee \log\left(\frac{2p}{\sqrt{T}s_T^{(2)}}\right)} + 2s_T^{(9)}. \tag{OA.30}$$

By a similar reasoning, we can show that

$$\mathbb{P}\left(\mathcal{S}_{\mu^*_{\alpha'}}\right) \geq 1-\alpha' - \mathbb{P}\left(\left(\mathcal{S}_T^{(2)}\right)^c\right) - K_4\sqrt{T}s_T^{(2)}\sqrt{1 \vee \log\left(\frac{2p}{\sqrt{T}s_T^{(2)}}\right)} - 2s_T^{(9)}. \tag{OA.31}$$

Since $\sqrt{T}s_T^{(2)}\sqrt{1 \vee \log\left(\frac{2p}{\sqrt{T}s_T^{(2)}}\right)} \to 0$, $s_T^{(8)} \to 0$, $s_T^{(9)} \to 0$ by Lemma D.8 (iii), (iv), (v) and $\mathbb{P}\left(\left(\mathcal{S}_T^{(2)}\right)^c\right) \to 0$ by Lemma D.5 (ii), (OA.30) and (OA.31) yield the result □

D.5 Auxiliary lemma on sequences

Lemma D.8 *Under Assumption 5, we have* (*i*) $s_T^{(12)}\sqrt{1 \vee \log\left(2p/s_T^{(12)}\right)} = o(1)$;

(*ii*) $2T^{-1/2}s_T^{(11)} + s_T^{(2)} + s_T^{(5)} = O\left(\sqrt{\frac{\log(T\vee p)}{(T\wedge p)}}p^{\frac{2}{q}}\right)(\|\varphi^*\|_2 \vee 1)$;

(*iii*) $\sqrt{T}s_T^{(2)}\sqrt{1 \vee \log\left(\frac{2p}{\sqrt{T}s_T^{(2)}}\right)} = o(1)$;

(*iv*) $s_T^{(8)} = o_P(1)$;

(*v*) $s_T^{(9)} = o(1)$;

(*vi*) $s_T^{(13)}\sqrt{1 \vee \log\left(2p/s_T^{(13)}\right)} = o(1)$.

Proof.

Proof of (i). By Assumption 5 (i), we have $s_T^{(3)} = O\left(\sqrt{\log(T)}\right)$, so that

$$s_T^{(6)} = O\left(\left(\frac{\log(T\vee p)^2}{\sqrt{T}}\|\beta^*\|_1\right)^{1/2}\right). \tag{OA.32}$$

Since $s_T^{(11)} = O\left(\sqrt{\log(T\vee p)}\right)$, this yields

$$s_T^{(6)}s_T^{(11)} = O\left(\left(\frac{\log(T\vee p)^3}{\sqrt{T}}\|\beta^*\|_1\right)^{1/2}\right). \tag{OA.33}$$

We also have

$$\left(s_T^{(6)}\right)^2 s_T^{(11)} = O\left(\left(\frac{\log(T\vee p)^3}{\sqrt{T}}\|\beta^*\|_1\right)^{1/2}\right), \tag{OA.34}$$

because $s_T^{(6)} = o(1)$ by (OA.32) and Assumption 5 (ii). Next, it holds that

$$s_T^{(2)} = O\left(\frac{\sqrt{\log(T)}p^{\frac{2}{q}}}{T\wedge p}(\|\varphi^*\|_2 \vee 1)\right), \tag{OA.35}$$

so that

$$s_T^{(6)}s_T^{(2)} = o\left(s_T^{(2)}\right), \tag{OA.36}$$

since $s_T^{(6)} = o(1)$ by (OA.32) and Assumption 5 (ii). Moreover, it holds that

$$s_T^{(4)} = O\left(\sqrt{\log(T)}\frac{p^{\frac{4}{q}}T^{\frac{2}{q}}}{T\wedge p}(\|\varphi^*\|_2^2 \vee 1)\right)$$

and, therefore,

$$s_T^{(7)} = O\left(\sqrt{\frac{\log(T \vee p)^2 p^{\frac{4}{q}} T^{\frac{2}{q}}}{T(T \wedge p)}} \left(\|\varphi^*\|_2 \vee 1\right)\right). \tag{OA.37}$$

Recall that

$$s_T^{(12)} = 2s_T^{(6)} + 2s_T^{(6)} s_T^{(11)} + \left(s_T^{(6)}\right)^2 s_T^{(11)} + \frac{\sqrt{T}}{2} s_T^{(2)} + \frac{\sqrt{T}}{2} s_T^{(6)} s_T^{(2)} + s_T^{(7)}.$$

By (OA.32), (OA.33), (OA.34), (OA.35), (OA.36), (OA.37), we obtain

$$\begin{aligned} s_T^{(12)} &= O\left(\left(\frac{\log(T \vee p)^3}{\sqrt{T}} \|\beta^*\|_1\right)^{1/2} + \left(\frac{\log(T \vee p)^{3/2}\sqrt{T}}{(T \wedge p)} + \sqrt{\frac{\log(T \vee p)^2 p^{\frac{4}{q}} T^{\frac{2}{q}}}{T(T \wedge p)}}\right) \left(\|\varphi^*\|_2 \vee 1\right)\right) \\ &= O\left(\left(\frac{\log(T \vee p)^3}{\sqrt{T}} \|\beta^*\|_1\right)^{1/2} + \frac{\log(T \vee p)^{3/2}\sqrt{T}}{(T \wedge p)} \left(1 + \sqrt{\frac{p^{\frac{4}{q}} T^{\frac{2}{q}}}{T}}\right) \left(\|\varphi^*\|_2 \vee 1\right)\right). \end{aligned} \tag{OA.38}$$

Additionally, we have $(T(T \wedge p))^{-1/2} = o_P\left(s_T^{(7)}\right) = O\left(s_T^{(12)}\right)$ so that $\log\left(2p/s_T^{(12)}\right) = O\left(\log(2p) + \log\left(\sqrt{T(T \wedge p)}\right)\right) = O(\log(T \vee p))$. This and (OA.38) imply

$$\begin{aligned} & s_T^{(12)} \sqrt{1 \vee \log\left(2p/s_T^{(12)}\right)} \\ & = O\left(\left(\frac{\log(T \vee p)^5}{\sqrt{T}} \|\beta^*\|_1\right)^{1/2} + \frac{\log(T \vee p)^{5/2}\sqrt{T}}{(T \wedge p)} \left(1 + \sqrt{\frac{p^{\frac{4}{q}} T^{\frac{2}{q}}}{T}}\right) \left(\|\varphi^*\|_2 \vee 1\right)\right) = o(1), \end{aligned}$$

by Assumption 5 and the fact that $q \geq 8$.

Proof of (ii). The result follows directly from Assumption 5 and (OA.35).

Proof of (iii). We have $(T \wedge p)^{-1} = o_P\left(\sqrt{T} s_T^{(2)}\right)$, hence $\log\left(\frac{2p}{\sqrt{T} s_T^{(2)}}\right) = O(\log(T \vee p))$, so that

$$\begin{aligned} & \sqrt{T} s_T^{(2)} \sqrt{1 \vee \log\left(\frac{2p}{\sqrt{T} s_T^{(2)}}\right)} \\ & = O\left(\log(T \vee p)^{\frac{3}{2}} \sqrt{T} \left(\frac{1}{p} + \frac{p^{\frac{2}{q}}}{T} + \frac{p^{\frac{2}{q} - \frac{1}{2}}}{\sqrt{T}}\right) \left(\|\varphi^*\|_2 \vee 1\right)\right) = o(1) \end{aligned}$$

by the definition of $s_T^{(2)}$ and Assumption 5.

Proof of (iv). We have $\Delta = o_P(1)$ by the facts that $\mathbb{P}(\mathcal{S}_T^{(1)}) \to 1$ and that $s_T^{(1)} = o(1)$ by Assumption 5. This yields the result using that $\lim_{x\to 0^+} x\log(x) \to 0$.

Proof of (v). Recall that

$$
\begin{aligned}
s_T^{(9)} &= K_1\Big[\left(T^{\kappa-1/2} + T^{1-\frac{c}{2}(1-\frac{2}{h})}\right)\log(T)\log(p) + T^{-1/4}\log(p)^{3/2}\log(T) \\
&\quad + p^{\frac{2}{h}} T^{\frac{1}{h}-\frac{1}{2}}\log(p)^2\log(T) \\
&\quad + \left(p\log(p)^{\frac{3}{2}h-4}\log(T)\log(Tp)\right)^{\frac{1}{h-2}} T^{-\frac{1}{4}} + T^{\frac{1}{2}-c\kappa}\left(p^{\frac{1}{h+1}}\sqrt{1\vee\log(p)}\right)\Big] \\
&= O_P\Big(\left(T^{\kappa-1/2} + T^{1-\frac{c}{2}(1-\frac{2}{h})}\right)\log(T)\log(T^r) + T^{-1/4}\log(T^r)^{3/2}\log(T) \\
&\quad + T^{\frac{2r}{h}+\frac{1}{h}-\frac{1}{2}}\log(T^r)^2\log(T) \\
&\quad + \left(\log(T^r)^{\frac{3}{2}h-4}\log(T)\log(T^{r+1})\right)^{\frac{1}{h-2}} T^{\frac{r}{h-2}-\frac{1}{4}} + T^{\frac{r}{h+1}+\frac{1}{2}-c\kappa}\sqrt{1\vee\log(T^r)}\Big).
\end{aligned}
$$

By Assumption 4, we have $\kappa - 1/2 < 0$ and $1 - \frac{c}{2}(1-\frac{2}{h}) < 0$, so that

$$
\left(T^{\kappa-1/2} + T^{1-\frac{c}{2}(1-\frac{2}{h})}\right)\log(T)\log(T^r) \to 0.
$$

As r is finite, $T^{-1/4}\log(T^r)^{3/2}\log(T) \to 0$. Moreover, since $r < \frac{h}{4} - \frac{1}{2}$, we have

$$
T^{\frac{2r}{h}+\frac{1}{h}-\frac{1}{2}}\log(T^r)^2\log(T) \to 0
$$

and

$$
\left(\log(T^r)^{\frac{3}{2}h-4}\log(T)\log(T^{r+1})\right)^{\frac{1}{h-2}} T^{\frac{r}{h-2}-\frac{1}{4}} \to 0
$$

Additionally, it holds that $T^{\frac{r}{h+1}+\frac{1}{2}-c\kappa}\sqrt{1\vee\log(T^r)} \to 0$ because $r < (h+1)(c\kappa - \frac{1}{2})$. All of this yields (v).

Proof of (vi). The proof is similar to that of (i) and therefore omitted. □

D.6 Auxiliary lemmas on factors and loadings

In this Section, we prove useful results on the factors, the factor loadings and their estimators. Let $H = T^{-1}V^{-1}\breve{F}^\top F B^\top B$, where V is the $K\times K$ matrix corresponding the K largest eigenvalues of $T^{-1}XX^\top$. Recall that the estimated loadings are $\breve{B} = X^\top\breve{F}\left(\breve{F}^\top\breve{F}\right)^{-1} =$

$T^{-1}X^\top \breve{F}$ and $\breve{b}_j$ and b_j are the $K \times 1$ vectors corresponding to the j^{th} rows of $\breve{B}$ and B, respectively.

Lemma D.9 *Under the assumptions of Theorem 1, the following holds:*

(i) $\max_{j\in[p]} \left|\sum_{t=1}^T u_{tj}^2\right| = O_P(T)$;

(ii) $\max_{j\in[p],k\in[K]} \left|\sum_{t=1}^T u_{tj} f_{tk}\right| = O_P\left(p^{\frac{2}{q}}\sqrt{T}\right)$;

(iii) $\left\|U^\top \mathcal{E}\right\|_\infty = O_P\left(p^{\frac{2}{q}}\sqrt{T}\right)$;

(iv) $\max_{j\in[p],k\in[K]} \left|\sum_{t=1}^T u_{tj}\left(\sum_{\ell=1}^p u_{t\ell} b_{\ell k}\right)\right| = O_P\left(\sqrt{T}p^{\frac{2}{q}+\frac{1}{2}} + T\right)$;

(v) $\left\|\frac{1}{T}\sum_{t=1}^T u_t u_t^\top \varepsilon_t^2 - \mathbb{E}\left[u_t u_t^\top \varepsilon_t^2\right]\right\|_\infty = O_P\left(\frac{p^{4/q}}{\sqrt{T}}\right)$;

(vi) $\|\mathcal{E}\|_2 = O_P\left(\sqrt{T}\right)$;

(vii) $\|F\|_2 = O_P\left(\sqrt{T}\right)$;

(viii) $\left\|\frac{1}{T}F^\top F - I_K\right\|_2 = O_P\left(\frac{1}{\sqrt{T}}\right)$;

(ix) $\|U\|_2 = O_P\left(\sqrt{Tp}\right)$;

(x) $\left\|F^\top \mathcal{E}\right\|_2 = O_P\left(\sqrt{T}\right)$;

(xi) $\|UB\|_2 = O_P\left(\sqrt{Tp}\right)$;

(xii) $\left\|F^\top U\right\|_2 = O_P\left(\sqrt{Tp}\right)$;

(xiii) $\left\|\mathcal{E}^\top U\right\|_2 = O_P\left(\sqrt{Tp}\right)$;

(xiv) $\left\|F^\top UB\right\|_2^2 = O_P\left(Tp\right)$;

(xv) $\left\|\mathcal{E}^\top UB\right\|_2^2 = O_P\left(Tp\right)$.

Proof. In this proof, we will often apply Lemmas D.15, D.16 and D.17 to some specific processes. Following the arguments of the proof of Lemma D.1, it can be checked that the conditions of these Lemmas hold for these processes under the assumptions of Theorem 1.

Proof of (i). We apply Lemmas D.15 and D.17 to $Z_t = \left(u_{tj}^2 - \mathbb{E}\left[u_{tj}^2\right]\right)_{j=1}^p$ and get

$$\max_{j\in[p]}\left|\frac{1}{T}\sum_{t=1}^T u_{tj}^2 - \mathbb{E}\left[u_{tj}^2\right]\right| = O_P\left(\frac{p^{\frac{2}{q}}}{\sqrt{T}}\right). \tag{OA.39}$$

By the triangle inequality, we obtain

$$\begin{aligned}\max_{j\in[p]}\left|\sum_{t=1}^T u_{tj}^2\right| &\leq T\max_{j\in[p]}\left|\frac{1}{T}\sum_{t=1}^T u_{tj}^2 - \mathbb{E}[u_{tj}^2]\right| + T\max_{j\in[p]}\mathbb{E}\left[u_{tj}^2\right]\\ &= O_P\left(T + \sqrt{T}p^{\frac{2}{q}}\right) = O_P(T),\end{aligned}$$

where we used $\max_{j\in[p]}\mathbb{E}[u_{tj}^2] \leq \|\Sigma\|_\infty = O(1)$ by Assumption 3 (ii) and the fact that $p^{\frac{2}{q}}/\sqrt{T} \to 0$ by Assumption 5 (i).

Proof of (ii), (iii). We apply Lemmas D.15 and D.17 to

$$\begin{aligned}Z_t &= ((u_{tj}f_{tk})_{j=1}^p)_{k=1}^K;\\ Z_t &= (u_{tj}\varepsilon_t)_{j=1}^p,\end{aligned}$$

and obtain (ii), (iii).

Proof of (iv). We apply Lemmas D.15 and D.17 to

$$Z_t = \left(\left(u_{tj}\left(p^{-1/2}\sum_{\ell=1}^p u_{t\ell}b_{\ell k}\right) - \mathbb{E}\left[u_{tj}\left(p^{-1/2}\sum_{\ell=1}^p u_{t\ell}b_{\ell k}\right)\right]\right)_{j=1}^p\right)_{k=1}^K,$$

and obtain

$$\max_{j\in[p],k\in[K]}\left|\sum_{t=1}^T\left(u_{tj}\left(p^{-1/2}\sum_{\ell=1}^p u_{t\ell}b_{\ell k}\right) - \mathbb{E}\left[u_{tj}\left(p^{-1/2}\sum_{\ell=1}^p u_{t\ell}b_{\ell k}\right)\right]\right)\right| = O_P\left(p^{\frac{2}{q}}\sqrt{T}\right). \tag{OA.40}$$

Next, by Assumption 3 (v), we have

$$\max_{j\in[p],k\in[K]}\left|\sum_{t=1}^T\mathbb{E}\left[u_{tj}\left(\sum_{\ell=1}^p u_{t\ell}b_{\ell k}\right)\right]\right| = O(T). \tag{OA.41}$$

By the triangle inequality and equations (OA.40) and (OA.41), we obtain

$$\max_{j\in[p],k\in[K]}\left|\sum_{t=1}^T u_{tj}\left(\sum_{\ell=1}^p u_{t\ell}b_{\ell k}\right)\right|$$

$$\leq \sqrt{p} \max_{j\in[p],k\in[K]} \left| \sum_{t=1}^{T} \left(u_{tj} \left(p^{-1/2} \sum_{\ell=1}^{p} u_{t\ell} b_{\ell k} \right) - \mathbb{E}\left[u_{tj} \left(p^{-1/2} \sum_{\ell=1}^{p} u_{t\ell} b_{\ell k} \right) \right] \right) \right|$$
$$+ \max_{j\in[p],k\in[K]} \left| \sum_{t=1}^{T} \mathbb{E}\left[u_{tj} \left(\sum_{\ell=1}^{p} u_{t\ell} b_{\ell k} \right) \right] \right| = O_P\left(\sqrt{T} p^{\frac{2}{q}+\frac{1}{2}} + T \right).$$

Proof of (v). The result directly follows from the application of Lemmas D.15 and D.17 to $Z_t = u_t u_t^\top \varepsilon_t^2 - \mathbb{E}\left[u_t u_t^\top \varepsilon_t^2\right]$.

Proof of (vi). The result follows from applying Lemmas D.15 and D.17 to $Z_t = \varepsilon_t^2 - \mathbb{E}\left[\varepsilon_t^2\right]$ and using the triangle inequality.

Proof of (vii). To obtain this statement, we apply Lemmas D.15 and D.17 to $Z_t = f_{tk}^2 - \mathbb{E}[f_{tk}^2]$, sum over k and use the triangle inequality, noticing that $\mathbb{E}[f_{tk}^2] = 1$ by Assumption 2 (i).

Proof of (viii). Statement (viii) follows from the application of Lemmas D.15 and D.17 to $Z_t = f_{tk} f_{t\ell} - \mathbb{E}[f_{tk} f_{t\ell}]$, summing over k, ℓ and using the fact that $\mathbb{E}[f_t f_t^\top] = I_K$ by Assumption 2 (ii) from the main text.

Proof of (ix). We use the fact that $\mathbb{E}\left[\|U\|_2^2\right] = \mathbb{E}\left[\sum_{t=1}^{T} \sum_{j=1}^{p} u_{tj}^2\right] = O(Tp)$ by Assumption 3 (ii) and Markov's inequality.

Proof of (x). We apply Lemmas D.15 and D.17 to $Z_t = \varepsilon_t f_{tk}$ and obtain $\sum_{t=1}^{T} \varepsilon_t f_{tk} = O_P\left(\sqrt{T}\right)$. This yields (x), by $\left\|F^\top \mathcal{E}\right\|_2 = \sqrt{\sum_{k=1}^{K} \left(\sum_{t=1}^{T} \varepsilon_t f_{tk}\right)^2} = O_P\left(\sqrt{T}\right)$.

Proof of (xi), (xii) *and* (xiii). We apply Lemmas D.15 and D.17 to

$$Z_t = \left(p^{-1/2} \sum_{\ell=1}^{p} u_{t\ell} b_{\ell k} \right)^2 - \mathbb{E}\left[\left(p^{-1/2} \sum_{\ell=1}^{p} u_{t\ell} b_{\ell k} \right)^2 \right]$$

and obtain

$$\max_{k\in[K]} \left| \sum_{t=1}^{T} \left(\left(p^{-1/2} \sum_{\ell=1}^{p} u_{t\ell} b_{\ell k} \right)^2 - \mathbb{E}\left[\left(p^{-1/2} \sum_{\ell=1}^{p} u_{t\ell} b_{\ell k} \right)^2 \right] \right) \right| = O_P\left(\sqrt{T}\right). \quad \text{(OA.42)}$$

Note that, by Assumption 3 (iii),

$$\max_{k\in[K]} \mathbb{E}\left[\left(p^{-1/2}\sum_{\ell=1}^{p} u_{t\ell}b_{\ell k}\right)^2\right] = O(1). \tag{OA.43}$$

Then, we obtain the result using the triangle inequality and equations (OA.67) and (OA.69):

$$\begin{aligned}
\|UB\|_2^2 &= p\sum_{t=1}^{T}\sum_{k=1}^{K}\left(p^{-1/2}\sum_{\ell=1}^{p} u_{t\ell}b_{\ell k}\right)^2 \\
&\leq Kp\max_{k\in[K]}\left|\sum_{t=1}^{T}\left(\left(p^{-1/2}\sum_{\ell=1}^{p} u_{t\ell}b_{\ell k}\right)^2 - \mathbb{E}\left[\left(p^{-1/2}\sum_{\ell=1}^{p} u_{t\ell}b_{\ell k}\right)^2\right]\right)\right| \\
&\quad + KTp\max_{k\in[K]}\mathbb{E}\left[\left(p^{-1/2}\sum_{\ell=1}^{p} u_{t\ell}b_{\ell k}\right)^2\right] = O_P\left(Tp\right).
\end{aligned}$$

The proofs of (xii) and (xiii) are similar and therefore omitted.

Proof of (xiv), (xv). We apply Lemmas D.15 and D.17 to

$$Z_t = f_t\left(p^{-1/2}\sum_{\ell=1}^{p} u_{t\ell}b_{\ell k}\right);$$

$$Z_t = \varepsilon_t\left(p^{-1/2}\sum_{\ell=1}^{p} u_{t\ell}b_{\ell k}\right)$$

and obtain the result by summing over k. □

Lemma D.10 *Under the assumptions of Theorem 1, the following holds:*

(i) $\left\|\breve{F} - FH^\top\right\|_2^2 = O_P\left(\frac{T}{p} + 1\right)$;

(ii) $\left\|H^\top H - I_K\right\|_2^2 = O_P\left(\frac{1}{T} + \frac{1}{p}\right)$;

(iii) $\max_{j\in[p]}\left\|\breve{b}_j - Hb_j\right\|_2 = O_P\left(\frac{1}{\sqrt{p}} + \frac{p^{\frac{2}{q}}}{\sqrt{T}}\right)$;

(iv) $\|V^{-1}\|_2 = O_P\left(\frac{1}{p}\right)$;

(v) $\left\|\breve{U} - U\right\|_\infty = o_P\left(\frac{p^{\frac{2}{q}}}{T^{1/2-1/q}} + \frac{T^{1/q}}{\sqrt{p}}\right)$;

(vi) $\max_{j\in[p]}\sum_{t=1}^{T}|\breve{u}_{tj} - u_{tj}|^2 = O_P\left(p^{4/q} + \frac{T}{p}\right)$;

Proof. The results (i) to (v) follow from Lemmas S.9 and S.11 and Theorem 2 in Fan et al. (2023). Assumption 2 in Fan et al. (2023) is satisfied by our Assumption 3. Assumption 3 in Fan et al. (2023) corresponds to our Assumption 2.

To show (vi), note that, for all $t \in [T]$ and $j \in [p]$, it holds that

$$
\begin{aligned}
\breve{u}_{tj} - u_{tj} &= x_{tj}^\top - \breve{b}_j^\top \breve{f}_t - u_{tj} \\
&= b_j^\top f_t - \breve{b}_j^\top \breve{f}_t \\
&= b_j^\top \left(I_K - H^\top H\right) f_t - \left(\breve{b}_j - H b_j\right)^\top \breve{f}_t - b_j^\top H^\top \left(\breve{f}_t - H f_t\right).
\end{aligned}
$$

This yields

$$
\begin{aligned}
\max_{j\in[p]} \sum_{t=1}^{T} |\breve{u}_{tj} - u_{tj}|^2 \leq K\|B\|_\infty^2 \left\|H^\top H - I_K\right\|_2^2 \|F\|_2^2 \\
+ \max_{j\in[p]} \left\|\breve{b}_j - H b_j\right\|_2^2 T \\
+ K\|B\|_\infty^2 \|H\|_2^2 \left\|\breve{F} - F H^\top\right\|_2^2 = O_P\left(p^{4/q} + \frac{T}{p}\right),
\end{aligned}
$$

where we used (i), (ii), (iii), Lemma D.9 (vii) and Assumption 2 (iv). □

Lemma D.11 *Under the assumptions of Theorem 1, it holds that*

$$
\breve{F} - F H^\top = \frac{1}{T} F B^\top U^\top \breve{F} V^{-1} + \frac{1}{T} U B F^\top \breve{F} V^{-1} + \frac{1}{T} U U^\top \breve{F} V^{-1}.
$$

Proof. Recall that $H = \frac{1}{T} V^{-1} \breve{F}^\top F B^\top B$ and $\breve{F} V = \frac{1}{T} X X^\top \breve{F}$. As a result, we have

$$
\begin{aligned}
\breve{F} V &= T^{-1} X X^\top \breve{F} \\
&= \frac{1}{T}(F B^\top + U)(F B^\top + U)^\top \breve{F} \\
&= \frac{1}{T} F B^\top B F^\top \breve{F} + \frac{1}{T} F B^\top U^\top \breve{F} + \frac{1}{T} U B F^\top \breve{F} + T^{-1} U U^\top \breve{F}.
\end{aligned}
$$

Multiplying both sides by V^{-1}, we get the result. □

Lemma D.12 *Under the assumptions of Theorem 1, we have*

$$
\left\|\left(\breve{F} - F H^\top\right)^\top \mathcal{E}\right\|_2 = O_P\left(\sqrt{\frac{T}{p}} + 1\right).
$$

Proof. By Lemma D.11, we have

$$\left\|(\breve{F} - FH^\top)^\top \mathcal{E}\right\|_2 \le J_1 + J_2 + J_3, \tag{OA.44}$$

where

$$\begin{aligned} J_1 &= \frac{1}{T}\left\|\mathcal{E}^\top F B^\top U^\top \breve{F} V^{-1}\right\|_2; \\ J_2 &= \frac{1}{T}\left\|\mathcal{E}^\top U B F^\top \breve{F} V^{-1}\right\|_2; \\ J_3 &= \frac{1}{T}\left\|\mathcal{E}^\top U U^\top \breve{F} V^{-1}\right\|_2. \end{aligned}$$

We have

$$\begin{aligned} J_1 &\le \frac{1}{T}\left\|\mathcal{E}^\top F\right\|_2 \left(\|UB\|_2 \left\|\breve{F} - FH^\top\right\|_2 + \|H\|_2 \left\|B^\top U^\top F\right\|_2\right) \|V^{-1}\|_2 \\ &= O_P\left(\frac{1}{T}\sqrt{T}\left(\sqrt{Tp}\sqrt{\frac{T}{p}+1} + \sqrt{Tp}\right)\frac{1}{p}\right) = O_P\left(\frac{\sqrt{T}}{p} + \frac{1}{\sqrt{p}}\right), \end{aligned} \tag{OA.45}$$

by Lemmas D.10 (i), (ii), (iv) and D.9 (x), (xi), (xiv). Moreover, it holds that

$$\begin{aligned} J_2 &\le \frac{1}{T}\|\mathcal{E}^\top U B\|_2 \|F\|_2 \left\|\breve{F}\right\|_2 \|V^{-1}\|_2 \\ &= O_P\left(\frac{1}{T}\sqrt{T}\sqrt{Tp}\sqrt{T}\frac{1}{p}\right) = O_P\left(\sqrt{\frac{T}{p}}\right), \end{aligned} \tag{OA.46}$$

by Lemmas D.10 (iv) and D.9 (vii), (xv) and the fact that $\left\|\breve{F}\right\|_2 = \sqrt{KT}$. We also have

$$\begin{aligned} J_3 &\le \frac{1}{T}\left\|\mathcal{E}^\top U\right\|_2 \left(\|U\|_2 \left\|\breve{F} - FH^\top\right\|_2 + \left\|U^\top F\right\|_2\right) \|V^{-1}\|_2 \\ &= O_P\left(\frac{1}{T}\sqrt{Tp}\left(\sqrt{Tp}\sqrt{\frac{T}{p}+1} + \sqrt{Tp}\right)\frac{1}{p}\right) \\ &= O_P\left(\sqrt{\frac{T}{p}} + \frac{1}{\sqrt{p}} + 1\right), \end{aligned} \tag{OA.47}$$

where we used Lemmas D.10 (i), (iv) and D.9 (ix), (xii), (xiii). We obtain the result by (OA.44), (OA.45), (OA.46) and (OA.47). □

Lemma D.13 *Under the assumptions of Theorem 1, we have*

$$\max_{j\in[p]} \sum_{t=1}^{T} \left(\breve{u}_{tj}\widetilde{\varepsilon}_t + \widetilde{f}_t^\top \varphi^* - u_{tj}\varepsilon_t\right)^2 = O_P\left(\left(p^{\frac{4}{q}}T^{\frac{2}{q}} + \frac{T^{\frac{2}{q}+1}}{p}\right) + \left(1 + \frac{T}{p}\right)\|\varphi^*\|_2^2\right).$$

Proof. First, notice that, by the triangle inequality,

$$\begin{aligned}&\sqrt{\sum_{t=1}^{T}\left(\breve{u}_{tj}\widetilde{\varepsilon}_t+\widetilde{f}_t^\top\varphi^*-u_{tj}\varepsilon_t\right)^2}\\&=\sqrt{\sum_{t=1}^{T}\left(\breve{u}_{tj}\left(\widetilde{\varepsilon}_t-\varepsilon_t\right)+\widetilde{f}_t^\top\varphi^*+\left(\breve{u}_{tj}-u_{tj}\right)\varepsilon_t\right)^2}\\&\leq\sqrt{\sum_{t=1}^{T}\left(\breve{u}_{tj}\left(\widetilde{\varepsilon}_t-\varepsilon_t\right)\right)^2}+\sqrt{\sum_{t=1}^{T}\left(\widetilde{f}_t^\top\varphi^*\right)^2}+\sqrt{\sum_{t=1}^{T}\left(\left(\breve{u}_{tj}-u_{tj}\right)\varepsilon_t\right)^2}.\end{aligned}\tag{OA.48}$$

We first bound the term $\sum_{t=1}^{T}\left(\breve{u}_{tj}\left(\widetilde{\varepsilon}_t-\varepsilon_t\right)\right)^2$. Remark that

$$\sum_{t=1}^{T}\left(\breve{u}_{tj}\left(\widetilde{\varepsilon}_t-\varepsilon_t\right)\right)^2\leq\left\|\breve{U}\right\|_\infty^2\left\|\left(I_T-\breve{P}\right)\mathcal{E}-\mathcal{E}\right\|_2^2=\left\|\breve{U}\right\|_\infty^2\left\|\breve{P}\mathcal{E}\right\|_2^2.\tag{OA.49}$$

Now, using Lemma D.16 and the tail bound in Assumption 3 (iii), we obtain $\|U\|_\infty=O_P\left((Tp)^{1/q}\right)$. Combining this with Lemma D.10 (v) and $\left\|\breve{U}\right\|_\infty\leq\left\|\breve{U}-U\right\|_\infty+\|U\|_\infty$, we get

$$\left\|\breve{U}\right\|_\infty^2=O_P\left((Tp)^{2/q}\right).\tag{OA.50}$$

Next, recall that $\breve{P}=T^{-1}\breve{F}\breve{F}^\top\mathcal{E}$ and $\left\|\breve{F}\right\|_2=\sqrt{KT}$. This yields

$$\begin{aligned}\left\|\breve{P}\mathcal{E}\right\|_2&\leq\frac{1}{T}\left\|\breve{F}\right\|_2\left\|\left(\breve{F}-FH^\top\right)^\top\mathcal{E}\right\|_2+\frac{1}{T}\left\|\breve{F}\right\|_2\|H\|_2\left\|F^\top\mathcal{E}\right\|_2\\&=\frac{1}{\sqrt{T}}O_P\left(\sqrt{\frac{T}{p}}+1\right)=O_P\left(\frac{1}{\sqrt{T}}+\frac{1}{\sqrt{p}}\right),\end{aligned}\tag{OA.51}$$

by Lemmas D.10 (ii), D.9 (x) and D.12. Thanks to (OA.49), (OA.50) and (OA.51), we obtain

$$\sum_{t=1}^{T}\left(\breve{u}_{tj}\left(\widetilde{\varepsilon}_t-\varepsilon_t\right)\right)^2=O_P\left((Tp)^{\frac{2}{q}}\left(\frac{1}{T}+\frac{1}{p}\right)\right).\tag{OA.52}$$

Let us now bound the term $\sum_{t=1}^{T}\left(\widetilde{f}_t^\top\varphi^*\right)^2$. We have

$$\sum_{t=1}^{T}\left(\widetilde{f}_t^\top\varphi^*\right)^2=\left\|\left(I_T-\breve{P}\right)F\varphi^*\right\|_2^2\leq\left\|\left(I_T-\breve{P}\right)F\right\|_2^2\|\varphi^*\|_2^2.\tag{OA.53}$$

Next, notice that

$$\left\|\left(I_T-\breve{P}\right)F\right\|_2=\left\|\left(I_T-\frac{1}{T}\breve{F}\breve{F}^\top\right)F\right\|_2$$

$$\leq \left\| \frac{1}{T} \left(\breve{F} - FH^\top \right) \left(FH^\top \right)^\top F \right\|_2 + \left\| \frac{1}{T} FH^\top \left(\breve{F} - FH^\top \right)^\top F \right\|_2$$
$$+ \left\| \left(I_T - \frac{1}{T} FH^\top \left(FH^\top \right)^\top \right) F \right\|_2 . \tag{OA.54}$$

Then, notice that

$$\left\| \frac{1}{T} \left(\breve{F} - FH^\top \right) \left(FH^\top \right)^\top F \right\|_2 + \left\| \frac{1}{T} FH^\top \left(\breve{F} - FH^\top \right)^\top F \right\|_2$$
$$\leq \frac{2}{T} \left\| \breve{F} - FH^\top \right\|_2 \|F\|_2^2 \|H\|_2 = O_P \left(\sqrt{\frac{T}{p}} + 1 \right), \tag{OA.55}$$

by Lemmas D.10 (i), (ii) and D.9 (vii). Moreover, we have

$$\left\| \left(I_T - \frac{1}{T} FH^\top \left(FH^\top \right)^\top \right) F \right\|_2$$
$$\leq \left\| \left(I_T - \frac{1}{T} FF^\top \right) F \right\|_2 + \left\| \frac{1}{T} F \left(H^\top H - I_K \right) F^\top F \right\|_2$$
$$\leq \|F\|_2 \left\| I_K - \frac{1}{T} F^\top F \right\|_2 + \frac{1}{T} \|F\|_2 \left\| H^\top H - I_K \right\|_2 \|F\|_2^2 = O_P \left(1 + \sqrt{\frac{T}{p}} \right), \tag{OA.56}$$

by Lemmas D.9 (vii), (viii) and D.10 (ii). Combining (OA.53), (OA.54), (OA.55) and (OA.56), we get

$$\sum_{t=1}^{T} \left(\widetilde{f}_t^\top \varphi^* \right)^2 = O_P \left(1 + \frac{T}{p} \right) \|\varphi^*\|_2^2 . \tag{OA.57}$$

Finally, we bound $\sum_{t=1}^{T} \left(\left(\breve{u}_{tj} - u_{tj} \right) \varepsilon_t \right)^2$. Notice that

$$\max_{j \in [p]} \sum_{t=1}^{T} \left((\breve{u}_{tj} - u_{tj}) \varepsilon_t \right)^2 \leq \|\mathcal{E}\|_\infty^2 \max_{j \in [p]} \sum_{t=1}^{T} (\breve{u}_{tj} - u_{tj})^2 \tag{OA.58}$$

Next, using Lemma D.16 and the tail bound in Assumption 3 (iii), we have $\|\mathcal{E}\|_\infty^2 = O_P \left(T^{\frac{2}{q}} \right)$. This, Lemma D.10 (vi) and equation (OA.58) yield that

$$\max_{j \in [p]} \sum_{t=1}^{T} \left(\left(\breve{u}_{tj} - u_{tj} \right) \varepsilon_t \right)^2 = O_P \left(T^{\frac{2}{q}} p^{\frac{4}{q}} + \frac{T^{1+\frac{2}{q}}}{p} \right). \tag{OA.59}$$

Combining (OA.48), (OA.52), (OA.57) and (OA.59), we obtain the result. □

Lemma D.14 *Under the assumptions of Theorem 1, we have*

$$\left\| \breve{U}^\top \left(\breve{Y} - \breve{U} \beta^* \right) - U^\top \mathcal{E} \right\|_\infty = (\|\varphi^*\|_2 \vee 1) O_P \left(\frac{T}{p} + p^{\frac{2}{q}} + \sqrt{T} p^{\frac{2}{q} - \frac{1}{2}} \right).$$

Proof. In all this proof, we work on the event $\mathcal{E}_\sigma = \{\sigma_p\left(H^\top H\right) \geq 1/2\}$ which has probability going to 1 by Lemma D.10 (ii). Note that, on $\mathcal{E}_\sigma$, we have

$$\left\|\left(H^\top\right)^{-1}\right\|_2 \leq \sqrt{K}\left\|\left(H^\top\right)^{-1}\right\|_{op} \leq \sqrt{K}\sigma_p\left(H^\top H\right)^{-1/2} \leq \sqrt{2K}. \tag{OA.60}$$

Recall that $\check{Y} = \left(I_T - \check{P}\right)(X\beta^* + F\varphi^* + \mathcal{E})$. This yields

$$\begin{aligned}\left\|\check{U}^\top\left(\check{Y} - \check{U}\beta^*\right) - U^\top\mathcal{E}\right\|_\infty &\leq \left\|\check{U}^\top(F\varphi^* + \mathcal{E}) - U^\top\mathcal{E}\right\|_\infty \\ &\leq \left\|\check{U}^\top F\varphi^*\right\|_\infty + \left\|\left(\check{U} - U\right)^\top\mathcal{E}\right\|_\infty.\end{aligned} \tag{OA.61}$$

Let us first bound $\left\|\check{U}^\top F\varphi^*\right\|_\infty$. Since $\check{U}^\top\check{F} = 0$ and $H^\top$ is invertible on the event $\mathcal{E}_\sigma$, it holds that

$$\left\|\check{U}^\top F\varphi^*\right\|_\infty \leq \left\|\left(\check{U} - U\right)^\top\left(FH^\top - \check{F}\right)\left(H^\top\right)^{-1}\varphi^*\right\|_\infty + \left\|U^\top\left(FH^\top - \check{F}\right)\left(H^\top\right)^{-1}\varphi^*\right\|_\infty. \tag{OA.62}$$

We now bound the first term on the right-hand side of (OA.62). By the inequality of Cauchy-Schwartz, we have

$$\begin{aligned}&\left\|\left(\check{U} - U\right)^\top\left(FH^\top - \check{F}\right)\left(H^\top\right)^{-1}\varphi\right\|_\infty \\ &= \max_{j\in[p]}\left|\left(\left(\check{U} - U\right)^\top\left(FH^\top - \check{F}\right)\left(H^\top\right)^{-1}\varphi^*\right)_j\right| \\ &\leq \left(\max_{j\in[p]}\sum_{t=1}^T\left|\check{u}_{tj} - u_{tj}\right|^2\right)^{1/2}\left\|\check{F} - FH^\top\right\|_2\left\|\left(H^\top\right)^{-1}\right\|_2\|\varphi^*\|_2 \\ &= \|\varphi^*\|_2 O_P\left(\sqrt{p^{\frac{4}{q}} + \frac{T}{p}}\sqrt{\frac{T}{p} + 1}\right) = \|\varphi^*\|_2 O_P\left(\frac{T}{p} + p^{\frac{2}{q}} + p^{\frac{2}{q}-\frac{1}{2}}\sqrt{T}\right),\end{aligned} \tag{OA.63}$$

where we used Lemma D.10 (i), (ii), (vi) and equation (OA.60). Next, we control the second term on the right-hand side of (OA.62). By Lemma D.11, it holds that

$$\left\|U^\top\left(FH^\top - \check{F}\right)\left(H^\top\right)^{-1}\varphi^*\right\|_\infty \leq J_1 + J_2 + J_3, \tag{OA.64}$$

where

$$\begin{aligned}J_1 &= \frac{1}{T}\left\|U^\top FB^\top U^\top\check{F}V^{-1}\left(H^\top\right)^{-1}\varphi^*\right\|_\infty; \\ J_2 &= \frac{1}{T}\left\|U^\top UBF^\top\check{F}V^{-1}\left(H^\top\right)^{-1}\varphi^*\right\|_\infty;\end{aligned}$$

$$J_3 = \frac{1}{T} \left\| U^\top U U^\top \breve{F} \left(H^\top\right)^{-1} \varphi^* \right\|_\infty .$$

Remark that

$$\begin{aligned}
\left\| B^\top U^\top \breve{F} \right\|_2 &\le \left\| B^\top U^\top \right\|_2 \left\| \breve{F} - F H^\top \right\|_2 + \|H\|_2 \left\| B^\top U^\top F \right\|_2 \\
&= O_P\left(\sqrt{Tp} \sqrt{\frac{T}{p} + 1} + \sqrt{Tp} \right) = O_P(T + \sqrt{Tp}), \qquad \text{(OA.65)}
\end{aligned}$$

by Lemmas D.9 (xi), (xiv) and D.10 (i), (ii). By the inequality of Cauchy-Schwartz, this yields

$$\begin{aligned}
J_1 &= \frac{1}{T} \max_{j\in[p]} \left| \left(U^\top F B^\top U^\top \breve{F} V^{-1} \left(H^\top\right)^{-1} \varphi^* \right)_j \right| \\
&= \frac{1}{T} \max_{j\in[p]} \left| \sum_{k=1}^K \left(\sum_{t=1}^T u_{tj} f_{tk} \right) \left(B^\top U^\top \breve{F} V^{-1} \left(H^\top\right)^{-1} \varphi^* \right)_k \right| \\
&\le \frac{1}{T} \left(\max_{j\in[p]} \left| \sum_{k=1}^K \left(\sum_{t=1}^T u_{tj} f_{tk} \right)^2 \right| \right)^{1/2} \left\| B^\top U^\top \breve{F} \right\|_2 \left\| V^{-1} \right\|_2 \left\| (H^{-1})^\top \right\|_2 \|\varphi^*\|_2 \\
&\le \frac{1}{T} \sqrt{K} \max_{j\in[p]} \left| \sum_{t=1}^T u_{tj} f_{tk} \right| \left\| B^\top U^\top \breve{F} \right\|_2 \left\| V^{-1} \right\|_2 \left\| (H^{-1})^\top \right\|_2 \|\varphi^*\|_2 \\
&= O_P\left(\frac{1}{Tp} \left(T + \sqrt{Tp} \right) \sqrt{T} p^{\frac{2}{q}} \right) \|\varphi^*\|_2 = O_P\left(\sqrt{T} p^{\frac{2}{q}-1} + p^{\frac{2}{q}-\frac{1}{2}} \right) \|\varphi^*\|_2, \qquad \text{(OA.66)}
\end{aligned}$$

where we used Lemmas D.9 (ii) and D.10 (ii), (iv) and equations (OA.65) and (OA.60). Then, notice that, by Lemma D.10 (i) and (ii), we have

$$\left\| F^\top \breve{F} \right\|_2 \le \|F\|_2 \left\| \breve{F} - F H^\top \right\|_2 + \|F\|_2^2 \|H\|_2 = O_P(T). \qquad \text{(OA.67)}$$

This allows to bound J_2. Indeed, by the inequality of Cauchy-Schwartz, it holds that

$$\begin{aligned}
J_2 &= \frac{1}{T} \max_{j\in[p]} \left| \left(U^\top U B F^\top \breve{F} V^{-1} \left(H^\top\right)^{-1} \varphi^* \right)_j \right| \\
&= \frac{1}{T} \max_{j\in[p]} \left| \sum_{k=1}^K \sum_{t=1}^T u_{tj} \left(\sum_{\ell=1}^p u_{t\ell} b_{\ell k} \right) \left(F^\top \breve{F} V^{-1} \left(H^\top\right)^{-1} \varphi^* \right)_k \right| \\
&\le \frac{1}{T} \max_{j\in[p]} \sqrt{ \sum_{k=1}^K \left(\sum_{t=1}^T u_{tj} \left(\sum_{\ell=1}^p u_{t\ell} b_{\ell k} \right) \right)^2 } \left\| F^\top \breve{F} V^{-1} \left(H^\top\right)^{-1} \varphi^* \right\|_2 \\
&\le \frac{1}{T} \sqrt{K} \max_{j\in[p],k\in[K]} \left| \sum_{t=1}^T u_{tj} \left(\sum_{\ell=1}^p u_{t\ell} b_{\ell k} \right) \right| \left\| F^\top \breve{F} \right\|_2 \left\| V^{-1} \right\|_2 \left\| (H^{-1})^\top \right\|_2 \|\varphi^*\|_2
\end{aligned}$$

$$= O_P\left(\frac{1}{Tp}T\left(T+p^{\frac{2}{q}+\frac{1}{2}}\sqrt{T}\right)\right)\|\varphi^*\|_2 = O_P\left(\frac{T}{p}+\sqrt{T}p^{\frac{2}{q}-\frac{1}{2}}\right)\|\varphi^*\|_2, \tag{OA.68}$$

by Lemmas D.9 (iv), (vii) and D.10 (ii), (iv) and equations (OA.60) and (OA.67). Finally note that

$$\begin{aligned}\left\|U^\top \breve{F}\right\|_2 &\leq \left\|U^\top F\right\|_2\left\|H^\top\right\|_2 + \|U\|_2\left\|\breve{F}-FH^\top\right\|_2\\ &= O_P\left(\sqrt{T}+\sqrt{T}\sqrt{\frac{T}{p}+1}\right) = O_P\left(\sqrt{T}+\frac{T}{\sqrt{p}}\right),\end{aligned} \tag{OA.69}$$

by Lemmas D.10 (i), (ii) and D.9 (ix), (xii). Thanks to this, we can bound J_3. Indeed, by the inequality of Cauchy-Schwartz, we have

$$\begin{aligned}J_3 &= \frac{1}{T}\max_{j\in[p]}\left|\left(U^\top U U^\top \breve{F}\left(H^\top\right)^{-1}\varphi^*\right)_j\right|\\ &= \frac{1}{T}\max_{j\in[p]}\left|\sum_{\ell=1}^{p}\left(\sum_{t=1}^{T}u_{tj}u_{t\ell}\right)\left(U^\top\breve{F}\left(H^\top\right)^{-1}\varphi^*\right)_\ell\right|\\ &\leq \frac{1}{T}\max_{j\in[p]}\sqrt{\sum_{\ell=1}^{p}\left(\sum_{t=1}^{T}u_{tj}u_{t\ell}\right)^2}\left\|U^\top\breve{F}\right\|_2\left\|V^{-1}\right\|_2\left\|(H^{-1})^\top\right\|_2\|\varphi^*\|_2 \qquad \text{(OA.70)}\\ &\leq \frac{1}{T}\sqrt{p}\max_{j\in[p]}\left|\sum_{t=1}^{T}u_{tj}^2\right|\left\|U^\top\breve{F}\right\|_2\left\|V^{-1}\right\|_2\left\|(H^{-1})^\top\right\|_2\|\varphi^*\|_2\\ &= O_P\left(\frac{1}{Tp}T\sqrt{p}\left(\sqrt{T}+\frac{T}{\sqrt{p}}\right)\right)\|\varphi^*\|_2 = O_P\left(\sqrt{\frac{T}{p}}+\frac{T}{p}\right)\|\varphi^*\|_2, \qquad \text{(OA.71)}\end{aligned}$$

where we used Lemmas D.9 (i) and D.10 (iv) and equations (OA.60) and (OA.69). Then, (OA.62), (OA.63), (OA.64), (OA.66), (OA.68), (OA.71) imply that

$$\left\|\breve{U}^\top F\varphi^*\right\|_\infty = O_P\left(\frac{T}{p}+p^{\frac{2}{q}}+\sqrt{T}p^{\frac{2}{q}-\frac{1}{2}}\right)\|\varphi^*\|_2. \tag{OA.72}$$

Let us now bound the second term on the right-hand side of (OA.61), that is $\left\|\left(\breve{U}-U\right)^\top\mathcal{E}\right\|_\infty$. Note that

$$\begin{aligned}\breve{U}^\top - U^\top &= X^\top - \breve{B}\breve{F}^\top - U^\top\\ &= BF^\top - \breve{B}\breve{F}^\top\\ &= B\left(I_K - H^\top H\right)F^\top - \left(\breve{B}-BH^\top\right)\breve{F}^\top - BH^\top\left(\breve{F}-FH\right)^\top.\end{aligned}$$

This yields

$$\left\|\left(\breve{U}-U\right)^\top\mathcal{E}\right\|_\infty \leq K_1+K_2+K_3, \tag{OA.73}$$

where

$$
\begin{aligned}
K_1 &= \left\|B\left(I_K - H^\top H\right)F^\top \mathcal{E}\right\|_\infty ;\\
K_2 &= \left\|\left(\breve{B} - BH^\top\right)\breve{F}^\top \mathcal{E}\right\|_\infty ;\\
K_3 &= \left\|BH^\top\left(\breve{F} - FH\right)^\top \mathcal{E}\right\|_\infty .
\end{aligned}
$$

By the inequality of Cauchy-Schwartz, Lemmas D.9 (x) and D.10 (ii) and Assumption 2 (iv), it holds that

$$
\begin{aligned}
K_1 &= \max_{j\in[p]}\left|\sum_{k=1}^{K} b_{jk}\left(\left(I_K - H^\top H\right)F^\top \mathcal{E}\right)_k\right|\\
&\leq \sqrt{K}\|B\|_\infty \left\|I_K - H^\top H\right\|_2 \left\|F^\top \mathcal{E}\right\|_2\\
&= O_P\left(\sqrt{\frac{1}{T}+\frac{1}{p}}\sqrt{T}\right) = O_P\left(1+\sqrt{\frac{T}{p}}\right). \qquad \text{(OA.74)}
\end{aligned}
$$

Next, we have

$$
\begin{aligned}
K_2 &= \max_{j\in[p]}\left|\sum_{k=1}^{K}\left(\breve{b}_j - Hb_j\right)_k\left(\breve{F}^\top \mathcal{E}\right)_k\right|\\
&\leq \max_{j\in[p]}\left\|\breve{b}_j - Hb_j\right\|_2\left\|\breve{F}^\top \mathcal{E}\right\|_2\\
&\leq \max_{j\in[p]}\left\|\breve{b}_j - Hb_j\right\|_2\left(\left\|\left(\breve{F} - FH^\top\right)^\top \mathcal{E}\right\|_2 + \|H\|_2\left\|F^\top \mathcal{E}\right\|_2\right)\\
&= O_P\left(\left(\frac{1}{\sqrt{p}}+\frac{p^{\frac{2}{q}}}{\sqrt{T}}\right)\left(\sqrt{T}+\sqrt{\frac{T}{p}}\right)\right)\\
&= O_P\left(\sqrt{\frac{T}{p}}+p^{\frac{2}{q}}\right). \qquad \text{(OA.75)}
\end{aligned}
$$

where we used the inequality of Cauchy-Schwarz, Lemmas D.9 (x), D.10 (ii), (iii), and D.12. Finally, by the inequality of Cauchy-Schwartz, Lemmas D.9 (ii) and D.12 and Assumption 2 (iv), it holds that

$$
\begin{aligned}
K_3 &= \max_{j\in[p]}\left|\sum_{k=1}^{K} b_{jk}\left(H^\top\left(\breve{F} - FH\right)^\top \mathcal{E}\right)_k\right|\\
&\leq \sqrt{K}\,\|B\|_\infty\left\|\left(\breve{F} - FH^\top\right)^\top \mathcal{E}\right\|_2\|H\|_2\\
&= O_P\left(\sqrt{\frac{T}{p}}+1\right). \qquad \text{(OA.76)}
\end{aligned}
$$

Combining (OA.73), (OA.74), (OA.75) and (OA.76) yields

$$\frac{1}{T}\left\|\left(\breve{U}-U\right)^{\top}\mathcal{E}\right\|_{\infty}=O_P\left(\sqrt{\frac{T}{p}}+p^{\frac{2}{q}}\right). \tag{OA.77}$$

We obtain the result of the lemma by (OA.61), (OA.72) and (OA.77). □

D.7 Pre-existing results on strong mixing sequences and high-dimensional Gaussian vectors

In this section, we state some useful lemmas and reformulate some results of Fan et al. (2023) and Lederer & Vogt (2021) that we use to prove Theorem 1.

D.7.1 Results on variables with polynomial tails

The following result is a direct consequence of the inequality of Cauchy-Schwarz. This lemma allows to show that products of variables in $u_{tj}, f_{tk}, \varepsilon_t$, $p^{-1/2}\sum_{j=1}^{p} b_j u_{tj}$ have polynomial tails.

Lemma D.15 *Let Z_1 and Z_2 be random variables such that $\|Z_1\|_q \leq C$ and $\|Z_2\|_q \leq C$, for some constants $C, q > 0$. Then, we have $\|Z_1 Z_2\|_{q/2} \leq C^2$.*

The next lemma serves to bound the sup norm of some variables.

Lemma D.16 *Let Z be a mean-zero p-dimensional random vector. Assume that there exist constants $C, h > 0$ such that we have $\|Z\|_h \leq C$. Then, it holds that $\|Z\|_\infty = O_P\left(p^{1/h}\right)$.*

Proof. For all $j \in [p]$ and $z > 0$, we have $\mathbb{P}(|Z_j| \leq z) = \mathbb{P}(|Z_j|^h \leq z^h) \leq \mathbb{E}[|Z_j|^h]/z^h$ by Markov's inequality. By the union bound, we obtain $\mathbb{P}(|Z|_\infty \geq z) \leq p \max_{j\in[p]} \mathbb{P}(|Z_j|_\infty \geq z) \leq p\left(\mathbb{E}[|Z_j|^h]/z^h\right)$. Taking $z \propto p^{1/h}$ we obtain the result. □

D.7.2 Results on strong mixing sequences

The next Lemma is a direct consequence of Lemma S.4 and Remark 4 in Fan et al. (2023).

Lemma D.17 *Let $S_T = \sum_{t=1}^{T} Z_t$, where $\{Z_t\}_t$ is a sequence of mean-zero p-dimensional random vectors such that*

(i) There exist constants $C_1, \xi > 0$ and $h \geq 2$ such that, for all $t \in [T]$, we have

$$\left\|\left|Z_t\right|\right\|_{h+\xi} \leq C_1;$$

(ii) Theres exist constants $C_2, c > 0$ such that the strong mixing coefficients of the sequence $\{Z_t\}_t$ satisfy $\tilde{\alpha}(t) \leq C_2 t^{-c}$ for all $t \in \mathbb{Z}_+$;

(iii) $c > \frac{(h-1)(h+\xi)}{\xi}$.

Then, it holds that $\|S_T\|_\infty = O_P\left(p^{1/h}\sqrt{T}\right)$.

The last result of this subsection is a direct consequence of the high-dimensional central limit theorem for strong mixing sequences due to Theorem S.7 (a) in Fan et al. (2023).

Lemma D.18 *Let $S_T = n^{-1/2}\sum_{t=1}^{T} Z_t$, where $\{Z_t\}_t$ is a sequence of mean-zero p-dimensional random vectors, such that*

(i) There exist $C_1, \xi > 0$ and $h \geq 4$ such that, for all $t \in [T]$, $j \in [p]$ and $z > 0$, we have $\left\|\left|Z_t\right|\right\|_{h+\xi} \leq C_1$;

(ii) There exist constants $C_2, c > 0$ such that the strong mixing coefficients of the sequence $\{Z_t\}_t$ satisfy $\tilde{\alpha}(t) \leq C_2 t^{-c}$ for all $t \geq 2$;

(iii) $c > \left[\left(\frac{h+\xi}{\xi}\right)\left(\frac{h}{2}-1\right)\right] \vee \left(\frac{2}{1-\frac{2}{h}}\right)$ and $\kappa = \left(\frac{\frac{1}{2}+\frac{h}{4(h+1)}}{c+\frac{h}{2(h+1)}}\right) < \frac{1}{2}$

(iv) There exists $\sigma_* > 0$ such that $\sigma_p(\Sigma) \geq \sigma_*^2$, where $\Sigma = \mathbb{E}\left[S_T S_T^\top\right]$;

Let also $G \sim \mathcal{N}(0, \Sigma)$. Then, there exists a constant $K_2 > 0$ such that, for all $z \geq 0$, we have

$$\begin{aligned}
&\sup_{z\in\mathbb{R}_+} \left|\mathbb{P}\left(\|S_T\|_\infty \leq z\right) - \mathbb{P}(|G|_\infty \leq z)\right| \\
&\leq K_2\Big[\left(T^{\kappa-1/2} + T^{1-\frac{c}{2}(1-\frac{2}{h})}\right)\log(T)\log(p) + T^{-1/4}\log(p)^{3/2}\log(T) + p^{\frac{2}{h}} T^{\frac{1}{h}-\frac{1}{2}}\log(p)^2\log(T) \\
&\quad + \left(p\log(p)^{\frac{3}{2}h-4}\log(T)\log(Tp)\right)^{\frac{1}{h-2}} T^{-\frac{1}{4}} + T^{\frac{1}{2}-c\kappa}\left(p^{\frac{1}{h+1}}\sqrt{1\vee\log(p)}\right)\Big].
\end{aligned}$$

D.7.3 Results on high-dimensional Gaussian vectors

The following two lemmas are direct consequences of Lemmas A.4 and A.5 and Remark A.8 in Lederer & Vogt (2021). (Note that the lemmas in Lederer & Vogt (2021) themselves follow from results in Chernozhukov et al. (2013) and Chernozhukov et al. (2015).)

Lemma D.19 *Let $G := (G_1, \ldots, G_p)^\top$ be a mean zero p-dimensional Gaussian vector. Suppose that there exist constants c_3, C_3 such that $c_3 \leq \mathbb{E}[G_j^2] \leq C_3$ for all $j \in [p]$, then, for every $z, \delta > 0$, we have*

$$\mathbb{P}\left(\left|\|G\|_\infty - z\right| \leq \delta\right) \leq C\delta\sqrt{1 \vee \log(2p/\delta)},$$

where $C > 0$ depends only on c_3, C_3.

Lemma D.20 *Let $G := (G_1, \ldots, G_p)^\top$ and $G' := (G'_1, \ldots, G'_p)^\top$ be two mean zero p-dimensional Gaussian vectors with respective covariance matrices Σ^G and $\Sigma^{G'}$. Define $\delta = \left\|\Sigma^G - \Sigma^{G'}\right\|_\infty$. Suppose that there exist constants c_3, C_3 such that $c_3 \leq \mathbb{E}[G_j^2] \leq C_3$ for all $j \in [p]$. Then, there exists a constant $C > 0$ depending only on c_3, C_3 such that*

$$\sup_{z \in \mathbb{R}} \left|\mathbb{P}\left(\|G\|_\infty \leq z\right) - \mathbb{P}\left(\|G'\|_\infty \leq z\right)\right| \leq C\delta^{1/3}(1 \vee 2\log(2p) \vee \log(1/\delta))^{1/3}(\log(2p))^{1/3}.$$

D.8 On the rate condition in statement (ii) of Theorem 1.

D.8.1 Discussion

The power analysis in Theorem 1 contains the rate conditon

$$\sqrt{\frac{\log(T \vee p)}{T \wedge p}} = o_P\left(\frac{1}{T}\left\|U^\top U \beta^*\right\|_\infty\right), \tag{OA.78}$$

which we analyze in this subsection. First, we state the following Lemma, which contains sufficient conditions in terms of nonrandom quantities for (OA.78) to hold (concerning (i)) and not hold (concerning (ii)).

Lemma D.21 *Let the assumptions of Theorem 1 hold. We have*

(i) If $\sqrt{\frac{\log(T\vee p)}{T\wedge p}} + \frac{p^{\frac{4}{q}}}{\sqrt{T}}\|\beta^*\|_1 = o\left(\|\Sigma\beta^*\|_\infty\right)$, then (OA.78) holds.

(ii) If $\|\Sigma\beta^*\|_\infty + \frac{p^{\frac{4}{q}}}{\sqrt{T}}\|\beta^*\|_1 = o\left(\sqrt{\frac{\log(T\vee p)}{T\wedge p}}\right)$, then (OA.78) does not hold.

Lemma D.21 is proved in Section D.8.2. This allows us to give examples of sequences of β^* and Σ such that (OA.78) holds or does not hold under our assumptions.

Example where (OA.78) holds. Let $\beta^* = (b, 0, \ldots, 0)$ and $\Sigma_{11} \geq \underline{\sigma}$, where $b \neq 0$ and $\underline{\sigma} > 0$ are constants that do not depend on T. Assume also that $p = o(T^{\frac{q}{8}})$. Then, we have $\|\Sigma\beta^*\|_\infty \geq \Sigma_{11}|b| \geq \underline{\sigma}|b|$ and, therefore, by the fact that $p = o(T^{\frac{q}{8}})$, we obtain

$$\sqrt{\frac{\log(T \vee p)}{T \wedge p}} + \frac{p^{\frac{4}{q}}}{\sqrt{T}}\|\beta^*\|_1 = \sqrt{\frac{\log(T \vee p)}{T \wedge p}} + \frac{p^{\frac{4}{q}}}{\sqrt{T}}|b| = o(\underline{\sigma}|b|) = o\left(\|\Sigma\beta^*\|_\infty\right).$$

By Lemma D.21 (i), this shows that (OA.78) holds.

Example where (OA.78) does not hold. Let $\beta^* = \left(\left(\frac{p^{\frac{4}{q}}}{\sqrt{T}}\right)^{-1}\frac{1}{T \wedge p}, 0, \ldots, 0\right)$ and $\max_{j\in[p]}|\Sigma_{1j}| = O(1)$. Assume also that $p = o(T^{\frac{q}{8}})$. Then, we have

$$\|\Sigma\beta^*\|_\infty \leq \left(\max_{j\in[p]}|\Sigma_{1j}|\right)\left(\frac{p^{\frac{4}{q}}}{\sqrt{T}}\right)^{-1}\frac{1}{T \wedge p}$$

and, therefore, we obtain

$$\|\Sigma\beta^*\|_\infty + \frac{p^{\frac{4}{q}}}{\sqrt{T}}\|\beta^*\|_1 \leq \left(\max_{j\in[p]}|\Sigma_{1j}|\right)\left(\frac{p^{\frac{4}{q}}}{\sqrt{T}}\right)^{-1}\frac{1}{T \wedge p} + \frac{1}{T \wedge p} = o\left(\sqrt{\frac{\log(T \vee p)}{T \wedge p}}\right).$$

By Lemma D.21 (ii), this shows that (OA.78) does not hold.

D.8.2 Proof of Lemma D.21

First, we apply Lemmas D.15 and D.17 to $Z_t = \left(\left(u_{tj}u_{tk} - \mathbb{E}\left[u_{tj}u_{tk}\right]\right)_{j=1}^p\right)_{k=1}^p$ (it can be shown that the conditions of these lemmas hold following the arguments of the proof of Lemma D.1). These Lemmas yield

$$\left\|\frac{U^\top U}{T} - \Sigma\right\|_\infty \leq \max_{j,k\in[p]}\left|\frac{1}{T}\sum_{t=1}^T u_{tj}u_{tk} - \mathbb{E}\left[u_{tj}u_{tk}\right]\right| = O_P\left(\frac{p^{\frac{4}{q}}}{\sqrt{T}}\right). \tag{OA.79}$$

Next, we show (i). By the triangle inequality, we have

$$\|\Sigma\beta^*\|_\infty \leq \left\|\left(\frac{U^\top U}{T} - \Sigma\right)\beta^*\right\|_\infty + \left\|\frac{U^\top U}{T}\beta^*\right\|_\infty.$$

By Hölder's inequality applied to each component of the vector $\left(\frac{U^\top U}{T} - \Sigma\right)\beta^*$, it holds that

$$\left\|\left(\frac{U^\top U}{T} - \Sigma\right)\beta^*\right\|_\infty \leq \left\|\frac{U^\top U}{T} - \Sigma\right\|_\infty \|\beta^*\|_1.$$

This leads to

$$\left\|\frac{U^\top U}{T}\beta^*\right\|_\infty \geq \|\Sigma\beta^*\|_\infty - \left\|\frac{U^\top U}{T} - \Sigma\right\|_\infty \|\beta^*\|_1.$$

Hence, by (OA.79) and since, under (i), it holds that $\sqrt{\frac{\log(T\vee p)}{T\wedge p}} + \frac{p^{\frac{4}{q}}}{\sqrt{T}}\|\beta^*\|_1 = o\left(\|\Sigma\beta^*\|_\infty\right)$, we have

$$\begin{aligned}\sqrt{\frac{\log(T\vee p)}{T\wedge p}} &= o\left(\|\Sigma\beta^*\|_\infty - \frac{p^{\frac{4}{q}}}{\sqrt{T}}\|\beta^*\|_1\right)\\ &= o_P\left(\|\Sigma\beta^*\|_\infty - \left\|\frac{U^\top U}{T} - \Sigma\right\|_\infty \|\beta^*\|_1\right)\\ &= o_P\left(T^{-1}\left\|U^\top U\beta^*\right\|_\infty\right),\end{aligned}$$

which proves (i).

Finally, we show (ii). By a reasoning similar to that of the proof of (i), we have

$$\begin{aligned}\left\|\frac{U^\top U}{T}\beta^*\right\|_\infty &\leq \left\|\left(\frac{U^\top U}{T} - \Sigma\right)\beta^*\right\|_\infty + \|\Sigma\beta^*\|_\infty\\ &\leq \left\|\frac{U^\top U}{T} - \Sigma\right\|_\infty \|\beta^*\|_1 + \|\Sigma\beta^*\|_\infty.\end{aligned}$$

Hence, by (OA.79) and since, under (ii), it holds that $\|\Sigma\beta^*\|_\infty + \frac{p^{\frac{4}{q}}}{\sqrt{T}}\|\beta^*\|_1 = o\left(\sqrt{\frac{\log(T\vee p)}{T\wedge p}}\right)$, we have

$$\begin{aligned}\left\|\frac{U^\top U}{T}\beta^*\right\|_\infty &\leq \left\|\frac{U^\top U}{T} - \Sigma\right\|_\infty \|\beta^*\|_1 + \|\Sigma\beta^*\|_\infty\\ &= O_P\left(\|\Sigma\beta^*\|_\infty + \frac{p^{\frac{4}{q}}}{\sqrt{T}}\|\beta^*\|_1\right)\\ &= o_P\left(\sqrt{\frac{\log(T\vee p)}{T\wedge p}}\right),\end{aligned}$$

which proves (ii) (i.e., that (OA.78) cannot hold).

References

Boucheron, S., Lugosi, G. & Massart, P. (2013), *Concentration inequalities: A nonasymptotic theory of independence*, Oxford university press.

Chernozhukov, V., Chetverikov, D. & Kato, K. (2013), 'Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors', *Annals of Statistics* **41**(6), 2786 – 2819.

Chernozhukov, V., Chetverikov, D. & Kato, K. (2015), 'Comparison and anti-concentration bounds for maxima of gaussian random vectors', *Probability Theory and Related Fields* **162**, 47–70.

Dong, X., Li, Y., Rapach, D. E. & Zhou, G. (2022), 'Anomalies and the expected market return', *The Journal of Finance* **77**(1), 639–681.

Fan, J., Lou, Z. & Yu, M. (2024), 'Are latent factor regression and sparse regression adequate?', *Journal of the American Statistical Association* **119**(546), 1076–1088.

Fan, J., Masini, R. P. & Medeiros, M. C. (2023), 'Bridging factor and sparse models', *The Annals of Statistics* **51**(4), 1692–1717.

French, K. R. (2024), 'Kenneth R. French-data library', *Tuck-MBA program web server. http://mba. tuck. dartmouth. edu/pages/faculty/ken. french/data_library. html* .

Jensen, T. I., Kelly, B. & Pedersen, L. H. (2023), 'Is there a replication crisis in finance?', *The Journal of Finance* **78**(5), 2465–2518.

Lederer, J. & Vogt, M. (2021), 'Estimating the Lasso's effective noise.', *Journal of Machine Learning Research* **22**, 276–1.

McCracken, M. & Ng, S. (2020), FRED-QD: A quarterly database for macroeconomic research, Technical report, National Bureau of Economic Research.

Vershynin, R. (2018), *High-dimensional probability: An introduction with applications in data science*, Vol. 47, Cambridge university press.