Internet Appendix for "Asset Pricing with Omitted Factors"

Stefano Giglio*
Yale School of Management
NBER and CEPR

Dacheng Xiu†
Booth School of Business
University of Chicago

This Version: August 19, 2020

Abstract

This appendix contains additional theoretical results, Monte Carlo simulations, additional empirical analysis, and supplementary mathematical proofs.

I Additional Theoretical Results

In what follows, we first provide a consistent estimator of the number of factors, p, and show the robustness of the risk premia estimator with respect to the number of factors used. We then develop the limiting distribution of the risk premia estimator and the zero-beta rate estimator in a more general setting that allows for pricing errors. Next, we develop the limiting distribution of the estimated factors. Finally we conclude this section by showing the consistency of the asymptotic variances involved.

I.1 Determining the Number of Factors

We propose to determine the number of factors using the following criterion:

$$\widehat{p} = \arg \min_{1 \leq j \leq p_{\max}} \left(n^{-1}T^{-1}\lambda_j(\bar{R}^\intercal \bar{R}) + j \times \phi(n,T) \right) - 1,$$

where $p_{\max}$ is some upper bound of p and $\phi(n,T)$ is some penalty function.

While our estimator makes use of a penalty function, in the same spirit as Bai and Ng (2002) do, our criterion takes on a simpler form. The objective function in Bai and Ng (2002) barring from penalty is equal to $\arg\min_k \frac{1}{nT}\sum_{j=k+1}^{n} \lambda_j(\bar{R}^\intercal \bar{R})$. It is rather challenging to analyze and control the growth rate of the sum of many eigenvalues required by this objective function, at least under only moment conditions we assume. Random matrix theory is likely unavoidable. In contrast, the plot of the objective function in our case against j is a penalized version of the scree plot. We show in the next theorem that our proposed estimator is consistent for the true number of factors p under appropriate conditions on the

*Address: 165 Whitney Avenue, New Haven, CT 06520, USA. E-mail address: `stefano.giglio@yale.edu`.
†Address: 5807 S Woodlawn Avenue, Chicago, IL 60637, USA. E-mail address: `dacheng.xiu@chicagobooth.edu`.

penalty function. On the one hand, the penalty function increases as j increases, so that it penalizes the choice of smaller eigenvalues. On the other hand, the penalty function is sufficiently small that it is dominated by the large eigenvalues. These two aspects together dictate the selected number of factors. Our choice of $p_{\max}$ is an economically reasonable upper bound for the number of factors, imposed only to improve the finite sample performance, which is not needed in the asymptotic analysis.

Theorem I.1. *Suppose Assumptions A.1, A.2, A.4 – A.7 hold, and suppose that $\phi(n,T) \to 0$ as n, $T \to \infty$, then it follows that $\mathrm{P}(\widehat{p} \geq p) \to 1$. If, in addition, $\phi(n,T)/(n^{-1/2} + T^{-1/2}) \to \infty$, then it follows that $\widehat{p} \xrightarrow{p} p$.*

Other estimators for the number of factors could be applied instead, including but not limited to those proposed by Onatski (2010) and Ahn and Horenstein (2013). However, to prove the consistency of these alternative estimators needs random matrix theory, which in turn requires stronger assumptions than ours.[1] Notably, the criterion of Ahn and Horenstein (2013) does not rely on any tuning parameter, which makes it appealing in certain scenarios. When applied empirically, their criterion often selects one factor, because the first eigenvalue is a bit stronger than the next few ones. However, it is unlikely that a single principal component summarizes all the risk factors in the financial markets. As we show in our simulations, selecting insufficient number of factors harms the inference on risk premia because of the omitted variable bias, whereas the risk premia estimates are robust to the inclusion of additional principal components. Fan et al. (2013) also find in their simulations that selecting more factors than necessary does not affect the performance of their factor-based covariance matrix estimates.

This robustness is a useful property, because in a finite sample, it is likely that $\widehat{p} \neq p$, although $\widehat{p}$ is a consistent estimator of p. The next theorem formally establishes the robustness of our estimates with respect to a few extra principal components. As long as our selected number of factors, denoted by $\breve{p}$, is greater than or equal to p, yet is not too large relative to n and T, then the risk premia estimator based on $\breve{p}$, denoted by $\breve{\gamma}_g$, remains consistent.

Theorem I.2. *Suppose Assumptions A.2, A.4 – A.11 hold. In addition, assume that z_t is i.i.d. and independent of u_t. If $\breve{p} \geq p$, $\breve{p} = o(n \wedge T)$, and $\lambda_{p+\breve{p}}(\bar{U}\bar{U}^\intercal) \geq K(n \vee T)$ for some $K > 0$ with probability approaching 1,*[2] *then $\breve{\gamma}_g$ is consistent with respect to $\eta\gamma$, and it holds that*

$$\breve{\gamma}_g - \widehat{\gamma}_g = o_p(1).$$

I.2 Allowing for Pricing Errors and Zero-beta Rate

Now we extend the main results to a more general setting, in which the zero-beta rate is unrestricted, and in which mispricing is allowed for in the model.

[1] We only need moment conditions to prove Theorem I.1. Nonetheless, our rate condition on the penalty function $\phi(n,T)$ is not sharp, which could be improved to $\phi(n,T)/(n^{-1} + T^{-1}) \to \infty$ using results of random matrix theory.

[2] The assumptions on the lower bound of the eigenvalue of $\bar{U}\bar{U}^\intercal$ can be replaced by more primitive assumptions on u_t. For instance, if u_t is $\overset{i.i.d.}{\sim} (0, \sigma_u^2)$ and $n/T \to c \in (0,\infty)$, then a direct result of the random matrix theory leads to such a bound, see, e.g., Theorem 5.11 of Bai and Silverstein (2009). Ahn and Horenstein (2013) show that a similar bound holds for somewhat more general u_t with time-series and cross-sectional dependence.

Assumption I.1. *Suppose the cross-section of asset returns r_t follows*

$$r_t = \alpha + \iota_n \gamma_0 + \beta\gamma + \beta v_t + u_t, \tag{I.1}$$

where the cross-sectional pricing error α is i.i.d., independent of β, u and v, with mean 0, standard deviation $\sigma^\alpha > 0$, and a finite fourth moment.

There is a large body of literature on testing the APT by exploring the deviation of α from 0, including Connor and Korajczyk (1988), Gibbons et al. (1989), MacKinlay and Richardson (1991), and more recently, Pesaran and Yamagata (2012) and Fan et al. (2015). This is, however, not the focus of this paper. Empirically, the pricing errors may exist for many reasons such as limits to arbitrage, transaction costs, market inefficiency, and so on, so that it is important to allow for a misspecified linear factor model. Gospodinov et al. (2014) and Kan et al. (2013) also consider this type of model misspecification in their two-pass cross-sectional regression setting.

Next, we assume

Assumption I.2. *There exists a $p \times 1$ vector β_0, such that $\left\| n^{-1}\beta^\intercal \iota_n - \beta_0 \right\|_{\text{MAX}} = o_p(1)$. Moreover, the matrix*

$$\begin{pmatrix} 1 & \beta_0^\intercal \\ \beta_0 & \Sigma^\beta \end{pmatrix} \quad \textit{is of full rank.}$$

This rank condition ensures that in the limit the factor loadings and ι_n are not perfectly correlated in the cross section, and in particular, that the zero-beta rate γ_0 is identifiable.

Finally, we need the following assumption, which imposes restrictions on the time series dependence of u_t. Assumption I.3 is similar to part of Assumption C in Bai (2003). Stationarity of u_t is not required. Eigenvalues of the residual covariance matrices $\mathrm{E}(u_t u_t^\intercal)$ are not necessarily bounded.

Assumption I.3. *Define, for any $i, i' \le n$, $t, t' \le T$,*

$$\mathrm{E}(u_{it}u_{i't}) = \sigma_{ii',t}, \quad \textit{and} \quad \mathrm{E}(u_{it}u_{i't'}) = \sigma_{ii',tt'}.$$

The following moment conditions hold, for all n and T, and $i, j \le p$, $l \le d$,

$$(i) \quad \max_{1\le t\le T} |\sigma_{ii',t}| \le |\sigma_{ii'}|, \quad \textit{for some } \sigma_{ii'}. \quad \textit{In addition, } n^{-1}\sum_{i=1}^{n}\sum_{i'=1}^{n}|\sigma_{ii'}| \le K.$$

$$(ii) \quad n^{-1}T^{-1}\sum_{i=1}^{n}\sum_{i'=1}^{n}\sum_{t=1}^{T}\sum_{t'=1}^{T}|\sigma_{ii',tt'}| \le K.$$

$$(iii) \quad \mathrm{E}\left(\sum_{t=1}^{T}\sum_{k=1}^{n} v_{jt}u_{kt}\right)^2 \le KnT.$$

In this scenario, we employ the alternative estimator (9), which also yields an estimate of the zero-beta rate. The next theorem establishes their limiting distributions.

Theorem I.3. *Under Assumptions A.2, A.4 – A.11, I.1 – I.3, and suppose $\widehat{p} \xrightarrow{p} p$, then as $n, T \to \infty$, we have*

$$n^{1/2}\left(\widehat{\gamma}_0 - \gamma_0\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \left(1 - \beta_0^\intercal (\Sigma^\beta)^{-1}\beta_0\right)^{-1} (\sigma^\alpha)^2\right),$$
$$\left(T^{-1}\Phi + n^{-1}\Upsilon\right)^{-1/2}\left(\widetilde{\gamma}_g - \eta\gamma\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \mathbb{I}_d\right),$$

where the asymptotic covariance matrices Φ is given by (11), and Υ is defined by

$$\Upsilon = (\sigma^\alpha)^2 \eta \left(\Sigma^\beta - \beta_0\beta_0^\intercal\right)^{-1} \eta^\intercal.$$

Unlike the CLT in Theorem 1, Theorem I.3 does not impose any restrictions on the relative rates of n and T. Note that this result assumes that the factor loading β is uncorrelated with the pricing error α, which means that the mispricing is not related to risk exposures. In fact, even if they were correlated, our estimator would instead converge to the "pseudo-true" parameter $\eta\left(\gamma + \text{plim}_{n\to\infty}(\beta^\intercal \mathbb{M}_{\iota_n}\beta)^{-1}\beta^\intercal\alpha\right)$, which is difficult to interpret, see, e.g., Kan et al. (2013).

To measure the goodness-of-fit in the cross-section of expected returns, we define the usual (population) cross-sectional R^2 for the latent factors in (I.1):

$$\mathrm{R}_v^2 = \frac{\gamma^\intercal(\Sigma^\beta - \beta_0\beta_0^\intercal)\gamma}{(\sigma^\alpha)^2 + \gamma^\intercal(\Sigma^\beta - \beta_0\beta_0^\intercal)\gamma},$$

which can be estimated in finite sample by

$$\widehat{\mathrm{R}}_v^2 = \frac{\bar{r}^\intercal \mathbb{M}_{\iota_n}\widehat{\beta}(\widehat{\beta}^\intercal \mathbb{M}_{\iota_n}\widehat{\beta})^{-1}\widehat{\beta}^\intercal \mathbb{M}_{\iota_n}\bar{r}}{\bar{r}^\intercal \mathbb{M}_{\iota_n}\bar{r}}.$$

We can consistently estimate the cross-sectional R^2 for the latent factors as well as the time-series R^2 for each observable factor g, introduced in Section 4.4 of the main text.

Theorem I.4. *Under Assumptions A.2, A.4 – A.11, I.1 – I.3, and suppose $\widehat{p} \xrightarrow{p} p$, then as $n, T \to \infty$, we have*

$$\widehat{\mathrm{R}}_v^2 \xrightarrow{p} \mathrm{R}_v^2 \quad \textit{and} \quad \widehat{\mathrm{R}}_g^2 \xrightarrow{p} \mathrm{R}_g^2.$$

I.3 Limiting Distribution of the Denoised Factors

As discussed above, our framework allows for measurement error in the observable factor proxies g. Theorem I.4 above indicates that we can separate the error from the factors using the extracted PCs. Moreover, we can conduct inference on $\widehat{g}_t$, provided additional assumptions:

Assumption I.4. *The following conditions hold:*

$$(i) \quad \sum_{t'=1}^{T} |\gamma_{n,tt'}| \leq K, \quad \textit{for all } t.$$

$$(ii) \quad \sum_{i'=1}^{n} |\sigma_{ii'}| \leq K, \quad \text{for all } i.$$

This assumption is identical to Assumption E in Bai (2003). It restricts the eigenvalues of $\mathrm{E}(u_t u_t^\intercal)$ and $\mathrm{E}(u_t^\intercal u_t)$ to be bounded as the dimension increases. We need this to bound the estimation error of factors uniformly over t, which in turn leads to the consistency of the asymptotic variance estimation to be discussed later.

For the same reason, we also need Assumption I.5, which Fan et al. (2011) and Fan et al. (2015) also adopt:

Assumption I.5. *For all $t', t \leq T$, we have*

$$\mathrm{E}\left(u_t^\intercal u_{t'} - \mathrm{E}u_t^\intercal u_{t'}\right)^4 \leq Kn^2, \quad \mathrm{E}\left\|\beta^\intercal u_t\right\|^4 \leq Kn^2.$$

The next assumption we need is identical to Assumption F3 in Bai (2003), which is used to describe the asymptotic distribution of the estimated factors at each point in time.

Assumption I.6. *For each t, as $n \to \infty$,*

$$n^{-1/2}\beta^\intercal u_t \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \Omega_t\right),$$

where, writing $\beta = (\beta_1 : \beta_2 : \ldots : \beta_n)^\intercal$,

$$\Omega_t = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \sum_{i'=1}^{n} \beta_i \beta_{i'}^\intercal \mathrm{E}(u_{it} u_{i't}). \tag{I.2}$$

Theorem I.5. *Under Assumptions A.2, and A.4 – A.11, I.1 – I.6, and suppose that $\widehat{p} \xrightarrow{p} p$, then as n, $T \to \infty$, we have*

$$\Psi_t^{-1/2}\left(\widehat{g}_t - \eta v_t\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \mathbb{I}_d),$$

where $\Psi_t = T^{-1}\Psi_{1t} + n^{-1}\Psi_{2t}$,

$$\begin{aligned}
\Psi_{1t} =& \Big\{ \left(v_t^\intercal \left(\Sigma^v\right)^{-1} \otimes \mathbb{I}_d\right) \Pi_{11} \left(\left(\Sigma^v\right)^{-1} v_t \otimes \mathbb{I}_d\right) - \left(v_t^\intercal \left(\Sigma^v\right)^{-1} \otimes \mathbb{I}_d\right) \Pi_{12}\eta^\intercal \\
&- \eta\Pi_{12}^\intercal \left(\left(\Sigma^v\right)^{-1} v_t \otimes \mathbb{I}_d\right) + \eta\Pi_{22}\eta^\intercal \Big\}, \quad \text{and} \\
\Psi_{2t} =& \eta \left(\Sigma^\beta\right)^{-1} \Omega_t \left(\Sigma^\beta\right)^{-1} \eta^\intercal.
\end{aligned}$$

In Bai (2003), the latent factors can be estimated at the $n^{-1/2}$-rate, provided that $n^{1/2}T^{-1} \to 0$. In our setting, the estimation error consists of the errors in estimating $\widehat{\eta}$ and $\widehat{v}_t$. Because $\widehat{\eta}$ is estimated up to a $T^{-1/2}$-rate error which dominates T^{-1} terms, the convergence rate of $\widehat{g}_t$ is $n^{-1/2} + T^{-1/2}$, which does not require any relative rate restrictions between n and T.

I.4 Consistency of the Asymptotic Covariance Estimators

In this section, we propose asymptotic variance estimators used in this paper, as well as establish their consistency. We only consider the more general setup in Theorem I.3. The case for Theorem 1 of the main text is simpler.

We construct the following estimators of the asymptotic variances, simply by using the sample analogues of their theoretical counterparts:

$$\widehat{\Phi} = \left(\widetilde{\gamma}^{\intercal}(\widehat{\Sigma}^{v})^{-1} \otimes \mathbb{I}_{d}\right) \widehat{\Pi}_{11} \left((\widehat{\Sigma}^{v})^{-1}\widetilde{\gamma} \otimes \mathbb{I}_{d}\right) + \left(\widetilde{\gamma}^{\intercal}(\widehat{\Sigma}^{v})^{-1} \otimes \mathbb{I}_{d}\right) \widehat{\Pi}_{12}\widehat{\eta}^{\intercal} + \widehat{\eta}\widehat{\Pi}_{21} \left((\widehat{\Sigma}^{v})^{-1}\widetilde{\gamma} \otimes \mathbb{I}_{d}\right) + \widehat{\eta}\widehat{\Pi}_{22}\widehat{\eta}^{\intercal},$$
$$\widehat{\Upsilon} = \widehat{\sigma^{\alpha}}^{2}\widehat{\eta}\left(\widehat{\Sigma}^{\beta} - \widehat{\beta}_{0}\widehat{\beta}_{0}^{\intercal}\right)^{-1}\widehat{\eta}^{\intercal},$$

where $\widehat{\Pi}_{11}, \widehat{\Pi}_{12}, \widehat{\Pi}_{22}, \widehat{Z}, \widehat{\Sigma}^{\beta}$, and $\widehat{\Sigma}^{v}$, are defined in Section 4.5, and

$$\widehat{\beta}_{0} = n^{-1}\widehat{\beta}^{\intercal}\iota_{n}, \quad \widehat{\sigma^{\alpha}}^{2} = n^{-1}\left\|\bar{r} - (\iota_{n} : \widehat{\beta})\widetilde{\Gamma}\right\|_{\mathrm{F}}^{2}, \quad \widetilde{\gamma} = \left(\widehat{\beta}^{\intercal}\mathbb{M}_{\iota_{n}}\widehat{\beta}\right)^{-1}\widehat{\beta}^{\intercal}\mathbb{M}_{\iota_{n}}\bar{r}, \quad \widetilde{\Gamma} = (\widehat{\gamma}_{0} : \widetilde{\gamma}^{\intercal})^{\intercal}.$$

Theorem I.6. *The sequence of $\{v_t, z_t\}_{t\geq 1}$ satisfies the exponential-type tail condition. Under Assumptions A.2, and A.4 – A.11, I.1 – I.5, and suppose that $\widehat{p} \xrightarrow{p} p$, then as n, $T \to \infty$, $n^{-3}T \to 0$, $q(T^{-1/4} + n^{-1/4}) \to 0$, $\widehat{\Phi} \xrightarrow{p} \Phi$ and $\widehat{\Upsilon} \xrightarrow{p} \Upsilon$.*

We say a sequence of centered multivariate random variables $\{y_t\}_{t\geq 1}$ satisfy the exponential-type tail condition, if there exist some constants a and b, such that $\mathrm{P}\left(|y_{it}| > y\right) \leq \exp\{-(y/b)^{a}\}$, for all i and t. This exponential-tail assumption is perhaps overly restrictive for financial returns, which feature heavy-tailness. The recent literature (e.g., Fan et al. (2017) and Fan et al. (2019)) replaces this assumption by moment conditions, and correspondingly, constructs estimators using Huber's loss. It is therefore possible to relax the exponential-tail condition using their techniques, though we do not explore this in this paper.

To estimate the asymptotic covariance matrices Ψ_{1t} and Ψ_{2t} in Theorem I.5, we can similarly replace v_t, Σ^{v}, Π_{11}, Π_{12}, Π_{22}, η, Σ^{β} by their sample analogues, $\widehat{v}_t$, $\widehat{\Sigma}^{v}$, $\widehat{\Pi}_{11}$, $\widehat{\Pi}_{12}$, $\widehat{\Pi}_{22}$, $\widehat{\eta}$, $\widehat{\Sigma}^{\beta}$, in the $\widehat{\Psi}_{1t}$ and $\widehat{\Psi}_{2t}$ constructions. With respect to Ω_t, we need an additional assumption:

Assumption I.7. *The innovation u_{it} is stationary and strongly mixing, and its covariance matrix Σ^{u} is sparse, i.e., there exists some $h \in [0, 1/2)$, with $\omega_T = (\log n)^{1/2}T^{-1/2} + n^{-1/2}$, such that*

$$s_n = \max_{1\leq i\leq n}\sum_{i'=1}^{n}|\Sigma^{u}_{ii'}|^{h}, \quad \textit{where} \quad s_n = o_p\left(\left(\omega_T^{1-h} + n^{-1} + T^{-1}\right)^{-1}\right).$$

Given this assumption, equation (I.2) and its estimator can be rewritten as

$$\Omega = \lim_{n\to\infty}\frac{1}{n}\beta^{\intercal}\Sigma^{u}\beta, \quad \text{and} \quad \widehat{\Omega}_t = \widehat{\Omega} = \frac{1}{n}\widehat{\beta}^{\intercal}\widehat{\Sigma}^{u}\widehat{\beta}, \tag{I.3}$$

where, for $1 \le i, i' \le n$,

$$\widehat{\Sigma}^u_{ii'} = \begin{cases} \widetilde{\Sigma}^u_{ii}, & i = i' \\ s_{ii'}(\Sigma^u_{ii'}), & i \neq i' \end{cases}, \quad \widetilde{\Sigma}^u = \frac{1}{T}\sum_{t=1}^{T} \widehat{u}_t \widehat{u}_t^{\intercal},$$

and $s_{ii'}(z) : \mathbb{R} \to \mathbb{R}$ is a general thresholding function with an entry dependent threshold $\tau_{ii'}$ such that (i) $s_{ii'}(z) = 0$ if $|z| < \tau_{ii'}$; (ii) $|s_{ii'}(z) - z| \le \tau_{ii'}$; and (iii) $|s_{ii'}(z) - z| \le a\tau^2_{ii'}$, if $|z| > b\tau_{ii'}$, with some $a > 0$ and $b > 1$. $\tau_{ii'}$ can be chosen as:

$$\tau_{ii'} = c(\widehat{\Sigma}_{ii}\widehat{\Sigma}_{i'i'})^{1/2}\omega_T, \quad \text{for some constant } c > 0.$$

Bai and Liao (2013) adopt a similar estimator of Σ^u for efficient estimation of factor models.

With their components constructed, our estimators for Ψ_{1t} and Ψ_{2t} are defined as:

$$\begin{aligned}
\widehat{\Psi}_{1t} =& T^{-1}\Big\{ \left(\widehat{v}_t^{\intercal}(\widehat{\Sigma}^v)^{-1} \otimes \mathbb{I}_d\right) \widehat{\Pi}_{11} \left((\widehat{\Sigma}^v)^{-1}\widehat{v}_t \otimes \mathbb{I}_d\right) - \left(\widehat{v}_t^{\intercal}(\widehat{\Sigma}^v)^{-1} \otimes \mathbb{I}_d\right) \widehat{\Pi}_{12}\widehat{\eta}^{\intercal} - \widehat{\eta}\widehat{\Pi}_{12}^{\intercal} \left((\widehat{\Sigma}^v)^{-1}\widehat{v}_t \otimes \mathbb{I}_d\right) \\
&+ \widehat{\eta}\widehat{\Pi}_{22}\widehat{\eta}^{\intercal}\Big\}, \\
\widehat{\Psi}_{2t} =& n^{-1}\widehat{\eta}\left(\widehat{\Sigma}^{\beta}\right)^{-1}\widehat{\Omega}_t\left(\widehat{\Sigma}^{\beta}\right)^{-1}\widehat{\eta}^{\intercal},
\end{aligned}$$

where $\widehat{\Omega}_t$ is given by (I.3). The next theorem establishes the desired consistency of $\widehat{\Psi}_{1t}$ and $\widehat{\Psi}_{2t}$:

Theorem I.7. *The sequence of $\{u_t, v_t, z_t\}_{t\ge1}$ satisfies the exponential-type tail condition. Under Assumptions A.2, A.3 – A.11, and I.1 – I.7, we have*

$$\widehat{\Psi}_{1t} - \Psi_{1t} \xrightarrow{p} 0, \quad \textit{and} \quad \widehat{\Psi}_{2t} - \Psi_{2t} \xrightarrow{p} 0.$$

II Simulations

In this section, we study the finite sample performance of our inference procedure using Monte Carlo simulations. We consider a five-factor data-generating process following (I.1), where the latent factors are calibrated to match the de-noised five Fama-French factors (RmRf, SMB, HML, RMW, CMA, see Fama and French (2015)) from our empirical study. Suppose that we do not observe all five factors, but instead some noisy version of the three Fama-French factors (RmRf, SMB, HML, see Fama and French (1993)), plus a potentially spurious macro factor calibrated to industrial production growth (IP) in our empirical study. Our simulations, therefore, include both the issue of omitted factors and that of a spurious factor. We calibrate the parameters γ_0, γ, η, Σ^v, Σ^z, Σ^u, $(\sigma^\alpha)^2$, β_0, and Σ^β to exactly match their counterparts in the data (in our estimation of the Fama-French five-factor model). We then generate the realizations of v_t, z_t, u_t, α, and β from a multivariate normal using the calibrated means and covariances.

We report in Tables II.1, II.2, and II.3 the bias and the root-mean-square error of the estimates using standard two-pass regressions and our three-pass approach. We choose different numbers of factors to estimate the model, $\breve{p} = 4$, 5, and 6, whereas the true value is 5. The five rows in each panel provide the results for the zero-beta rate, RmRf, SMB, HML, and IP, respectively. Throughout these tables, we find that the three-pass estimators with $\breve{p} = 5$ or 6 outperform the other estimators, in particular when n and T are large. By comparison, the two-pass estimates have substantial biases. Moreover, the biases for the market factor premium are substantial and negative even when n and T are large. The three-pass estimator with $\breve{p} = 4$ has an obvious bias, compared to the cases with $\breve{p} = 5$ and 6, because an omitted-factor problem still affects it (4 factors do not span the entire factor space).

We then plot in Figure II.1 the histograms of the standardized risk premia estimates using the estimated asymptotic standard errors for the two-pass estimator (right column) and for the three-pass method with $\breve{p} = 5$ (left column), respectively. The histograms on the right deviate substantially from the standard normal distribution, whereas those on the left match the normal distribution very well, which verifies our central limit results despite a small sample size $T = 240$ and a moderate dimension $n = 200$.[3] There exist some small higher order biases for γ_0, which would disappear with larger n and T in simulations not included here.

Next, we report in Table II.4 the estimated number of factors. We choose $\phi(n,T) = K(\log n + \log T)(n^{-1/2} + T^{-1/2})$, where $K = 0.5 \times \widehat{\lambda}$, $\widehat{\lambda}$ is the median of the first $p_{\max}$ eigenvalues of $n^{-1}T^{-1}\bar{R}^{\intercal}\bar{R}$. The median eigenvalue helps adjust the magnitude of the penalty function for better finite sample accuracy. Although the estimator is consistent, it cannot give the true number of factors without error, in particular when n or T is small, potentially due to the ad-hoc choice of tuning parameters.[4] In the empirical study, we apply this estimator of p and select slightly more factors to ensure the robustness of the estimates, as suggested by Theorem I.2.

Then we evaluate the size and power properties of the proposed test in Section 4.4. To check the size control, we create a purely noisy factor with $\eta = 0$ and variance calibrated to be the average variance of the four factors we consider. The top panel of Table II.5 reports the rejection probabilities of the test statistic under the null. In spite of slight over-rejection, the size control is acceptable given the moderate sizes of n and T. To evaluate the power, we report on the lower panel the average rejection probabilities when the null is false ($\eta \neq 0$). We test for factors with a variety of the signal-to-noise strength measured by R_g^2. These factors only load on the market factor, and share the same total variance calibrated to be the average variance as above, with R_g^2s being 2.5%, 5%, and 10%, respectively. As expected, we observe the rejection probability elevates to 100% as R_g^2 increases.

Finally, we compare the performance of these estimators with the mimicking portfolio estimators under more restrictive dynamics in which γ_0 is known and $\alpha = 0$. So we estimate the model using (8) and excess returns. We consider two sets of mimicking portfolios: one set (MP3) uses three portfolios as spanning assets to project factors, where portfolio weights are exactly proportional to the market, SMB, and HML beta. Using three base assets clearly leads to an omitted variable problem because these three assets cannot span the space of five factors. The second set of mimicking portfolios (MP) uses all assets as basis assets for projection. There is no omitted variable bias in this case as we prove

[3] Fan et al. (2017) show that the empirical eigenvectors can be estimated with very little finite sample bias. This might explain why the asymptotic approximation for risk premia is rather accurate despite the finite sample errors that accumulate from each step of our procedure.

[4] The eigenvalue ratio-based test by Ahn and Horenstein (2013) does not work well in our simulation setting because the first eigenvalue dominates the rest by a wide margin, so that their test often suggests 1 factor.

Table II.1: Simulation Results for $n = 50$

			Two-Pass Estimator		Three-Pass Estimators					
					$\breve{p} = 4$		$\breve{p} = 5$		$\breve{p} = 6$	
T	Param	True	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
120	γ_0	0.546	0.252	0.339	0.028	0.210	0.019	0.206	0.042	0.206
	RmRf	0.372	-0.230	0.514	0.020	0.425	0.019	0.425	-0.016	0.425
	SMB	0.229	-0.037	0.305	-0.024	0.275	-0.020	0.276	-0.012	0.276
	HML	0.209	0.010	0.349	-0.111	0.227	-0.088	0.221	-0.071	0.217
	IP	-0.003	-0.015	0.116	0.000	0.009	0.000	0.010	0.000	0.010
240	γ_0	0.546	0.265	0.324	0.085	0.191	0.060	0.177	0.048	0.173
	RmRf	0.372	-0.214	0.396	-0.033	0.319	-0.029	0.318	-0.020	0.317
	SMB	0.229	-0.129	0.250	-0.052	0.199	-0.037	0.196	-0.035	0.196
	HML	0.209	0.082	0.278	-0.074	0.168	-0.049	0.159	-0.043	0.158
	IP	-0.003	-0.024	0.136	0.001	0.006	0.001	0.006	0.001	0.007
480	γ_0	0.546	0.348	0.380	0.051	0.173	0.008	0.146	-0.001	0.145
	RmRf	0.372	-0.333	0.415	-0.047	0.252	-0.006	0.237	0.003	0.237
	SMB	0.229	-0.165	0.231	-0.067	0.158	-0.038	0.146	-0.036	0.146
	HML	0.209	0.211	0.296	-0.008	0.115	-0.022	0.115	-0.020	0.116
	IP	-0.003	-0.040	0.159	0.001	0.004	0.000	0.005	0.000	0.005

Note: In this table, we report the bias (Column "Bias") and the root-mean-square error (Column "RMSE") of the zero-beta rate and risk premia estimates using two-pass and three-pass estimators with $\breve{p} = 4$, 5, and 6, for $n = 50$, and $T = 120$, 240, and 480, respectively. The true data-generating process has five factors, and the parameters are calibrated based on the de-noised five Fama-French factors (RmRf, SMB, HML, RMW, and CMA). The true zero-beta rate is 0.546, and the true risk premia of four noisy yet observed factors (RmRf, SMB, HML, and IP) are provided in the "True" column. All numbers are in percentages.

in Proposition 1, but these estimators are not as efficient as the three-pass estimators. They become infeasible when $n > T$. Figure II.2 verifies these statements. Indeed, the deviation from normality is clearly visible for all estimators but ours. MP3 and two-pass estimates show visible biases whereas MP estimates display distortion due to the curse of dimensionality (n is of a similar scale to T). Tables II.6 - II.8 further illustrate that the RMSEs of the mimicking portfolio estimators are often larger than those of the three-pass estimators, due to large biases of MP3 and large variances of MP.

Overall, the three-pass estimator outperforms the two-pass and mimicking portfolio estimators by a large margin in almost all cases. The MP estimator using all assets ranks the second, despite being infeasible when n is greater than T. The biases in the two-pass and MP3 are substantial, yet they are unfortunately the most common choices in the empirical literature.

III Additional Empirical Results

In this section we provide more details on the construction of the test assets and present additional empirical results and robustness tests.

III.1 Additional Details on the Datasets

We assemble the set of test portfolios as follows. We start from a set of 202 standard equity portfolios from Kenneth French's website, that span the most well-known dimensions of risk: 25 portfolios sorted

Table II.2: Simulation Results for $n = 100$

			Two-Pass Estimator		Three-Pass Estimators					
					$\breve{p} = 4$		$\breve{p} = 5$		$\breve{p} = 6$	
T	Param	True	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
120	γ_0	0.546	0.434	0.482	0.102	0.191	0.083	0.167	0.076	0.163
	RmRf	0.372	-0.422	0.612	-0.088	0.421	-0.069	0.414	-0.063	0.414
	SMB	0.229	-0.087	0.305	-0.021	0.269	-0.018	0.269	-0.017	0.269
	HML	0.209	0.138	0.356	-0.023	0.202	-0.026	0.208	-0.023	0.209
	IP	-0.003	-0.011	0.100	0.000	0.010	0.000	0.010	0.000	0.011
240	γ_0	0.546	0.425	0.453	0.095	0.167	0.038	0.121	0.035	0.120
	RmRf	0.372	-0.431	0.538	-0.103	0.322	-0.041	0.298	-0.037	0.297
	SMB	0.229	-0.144	0.256	-0.043	0.197	-0.018	0.192	-0.018	0.192
	HML	0.209	0.312	0.399	0.058	0.165	0.009	0.153	0.006	0.153
	IP	-0.003	-0.020	0.107	0.000	0.006	0.000	0.007	0.000	0.007
480	γ_0	0.546	0.371	0.391	0.067	0.125	0.025	0.102	0.020	0.101
	RmRf	0.372	-0.374	0.439	-0.069	0.228	-0.022	0.216	-0.019	0.216
	SMB	0.229	-0.028	0.155	0.003	0.139	-0.001	0.140	0.002	0.140
	HML	0.209	0.033	0.203	-0.025	0.115	-0.019	0.114	-0.015	0.113
	IP	-0.003	-0.043	0.170	0.001	0.005	0.000	0.005	0.000	0.005

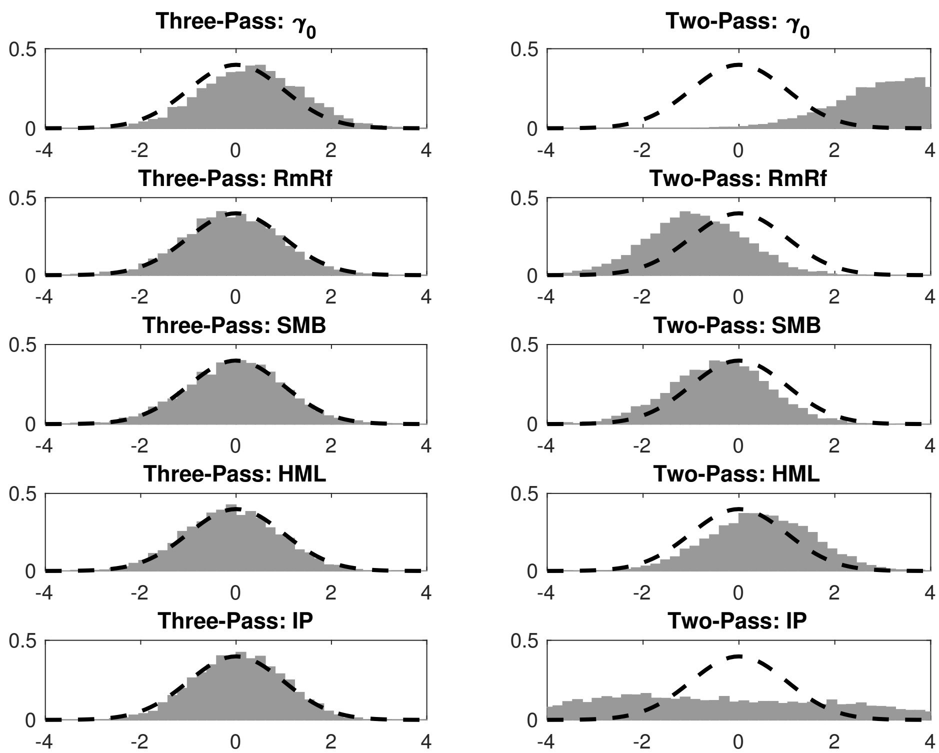
Note: In this table, we report the bias (Column "Bias") and the root-mean-square error (Column "RMSE") of the zero-beta rate and risk premia estimates using two-pass and three-pass estimators with $\breve{p} = 4$, 5, and 6, for $n = 100$, and $T = 120$, 240, and 480, respectively. The true data-generating process has five factors, and the parameters are calibrated based on the de-noised five Fama-French factors (RmRf, SMB, HML, RMW, and CMA). The true zero-beta rate is 0.546, and the true risk premia of four noisy yet observed factors (RmRf, SMB, HML, and IP) are provided in the "True" column. All numbers are in percentages.

Table II.3: Simulation Results for $n = 200$

			Two-Pass Estimator		Three-Pass Estimators					
					$\breve{p} = 4$		$\breve{p} = 5$		$\breve{p} = 6$	
T	Param	True	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
120	γ_0	0.546	0.423	0.460	0.069	0.132	0.059	0.113	0.058	0.112
	RmRf	0.372	-0.366	0.561	-0.047	0.400	-0.038	0.396	-0.038	0.396
	SMB	0.229	-0.137	0.323	-0.024	0.270	-0.020	0.271	-0.019	0.271
	HML	0.209	0.169	0.358	-0.011	0.208	-0.013	0.210	-0.012	0.210
	IP	-0.003	-0.012	0.092	0.000	0.010	0.000	0.011	0.000	0.011
240	γ_0	0.546	0.277	0.300	0.030	0.089	0.022	0.074	0.020	0.073
	RmRf	0.372	-0.270	0.408	-0.024	0.291	-0.023	0.290	-0.021	0.290
	SMB	0.229	-0.084	0.227	-0.011	0.196	-0.004	0.196	-0.004	0.196
	HML	0.209	0.105	0.266	-0.018	0.152	-0.013	0.155	-0.011	0.155
	IP	-0.003	-0.027	0.135	0.000	0.007	0.000	0.007	0.000	0.007
480	γ_0	0.546	0.256	0.273	0.051	0.105	0.013	0.067	0.012	0.067
	RmRf	0.372	-0.246	0.331	-0.054	0.226	-0.011	0.209	-0.010	0.209
	SMB	0.229	-0.089	0.175	-0.002	0.138	-0.007	0.138	-0.007	0.139
	HML	0.209	0.121	0.226	-0.010	0.112	-0.010	0.112	-0.009	0.112
	IP	-0.003	-0.046	0.168	0.001	0.005	0.000	0.005	0.000	0.005

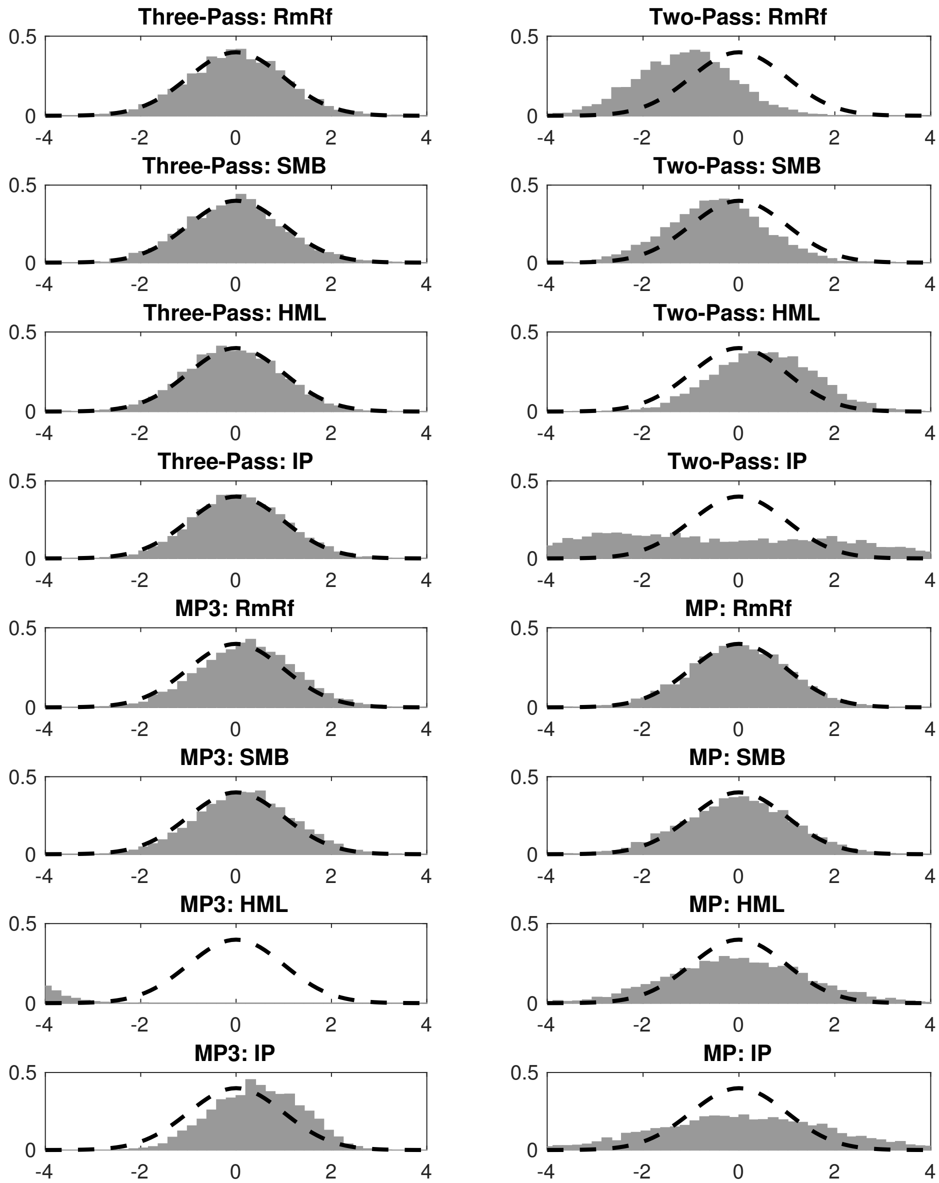
Note: In this table, we report the bias (Column "Bias") and the root-mean-square error (Column "RMSE") of the zero-beta rate and risk premia estimates using two-pass and three-pass estimators with $\breve{p} = 4$, 5, and 6, for $n = 200$, and $T = 120$, 240, and 480, respectively. The true data-generating process has five factors, and the parameters are calibrated based on the de-noised five Fama-French factors (RmRf, SMB, HML, RMW, and CMA). The true zero-beta rate is 0.546, and the true risk premia of four noisy yet observed factors (RmRf, SMB, HML, and IP) are provided in the "True" column. All numbers are in percentages.

Figure II.1: Histograms of the Standardized Estimates in Simulations



Note: The left panels provide the histograms of the standardized three-pass estimates using asymptotic standard errors, whereas the right panels provide those of the standardized two-pass risk premia estimates using the Fama-MacBeth approach for standard error estimation. We simulate the models with $n = 200$ and $T = 240$.

Figure II.2: Histograms of the Standardized Estimates in Simulations



Note: The top panel plots the histograms of the standardized estimates using the three-pass estimator (top left) and the two-pass estimator (top right), respectively, for four parameters. The bottom panels provide those of the standardized mimicking portfolio estimators, using three (bottom left) or all assets (bottom right), respectively. We simulate the models with $n = 200$ and $T = 240$.

Table II.4: Simulation Results for the Number of Factors

	$n = 50$		$n = 100$		$n = 200$	
T	Median	Stderr	Median	Stderr	Median	Stderr
120	3	0.87	4	0.69	5	0.18
240	3	0.58	3	0.92	5	0.18
480	3	0.03	5	0.42	5	0.41

Note: In this table, we report the median (Column "Median") and the standard error (Column "Stderr") of the estimates for the number of factors. The true number of factors in the data generating process is five.

Table II.5: Size and Power of the Test Statistic

	Size								
		$n = 50$			$n = 100$			$n = 200$	
α-level	1.0%	5.0%	10.0%	1.0%	5.0%	10.0%	1.0%	5.0%	10.0%
120	6.9%	15.8%	23.4%	7.8%	17.5%	25.2%	7.4%	17.1%	25.2%
240	3.7%	10.6%	17.3%	3.1%	9.8%	17.1%	3.8%	10.6%	17.2%
480	2.3%	8.3%	14.2%	2.3%	7.6%	13.8%	2.1%	8.1%	14.8%
					Power				
R_g^2	2.5%	5.0%	10.0%	2.5%	5.0%	10.0%	2.5%	5.0%	10.0%
120	41.1%	61.8%	87.3%	40.4%	62.6%	87.9%	40.2%	62.0%	88.6%
240	55.1%	83.6%	99.1%	53.6%	83.1%	99.0%	54.2%	83.3%	99.2%
480	79.5%	98.3%	100.0%	81.0%	98.5%	100.0%	80.8%	98.4%	100.0%

Note: In this table, we report on the upper panel the rejection probabilities for the level-α tests when $\mathbb{H}_0 : \eta = 0$ holds. The lower panel provides the rejection probabilities when the null hypothesis is false ($\mathrm{R}_g^2 = 2.5\%, 5\%, 10\%$).

by size and book-to-market ratio, 17 industry portfolios, 25 portfolios sorted by operating profitability and investment, 25 portfolios sorted by size and variance, 35 portfolios sorted by size and net issuance, 25 portfolios sorted by size and accruals, 25 portfolios sorted by size and beta, and 25 portfolio sorted by size and momentum.

We augment this set with a large set of additional anomaly portfolios sorted by various characteristics. Specifically, we obtain from WRDS a list of 103 characteristics, which we use to compute value-weighted quintile portfolios sorted by each characteristic (using NYSE breakpoints, restricting to share code 10 and 11, and exchange code 1, 2 and 3).[5] We remove portfolios for which we have missing returns during our sample period, yielding 413 additional test portfolios.

Finally, we add 10 maturity-sorted government bond portfolios, 10 corporate bond portfolios sorted on yield spread, 6 currency portfolios sorted on interest rate differentials, and 6 currency portfolios sorted on currency momentum obtained from Asaf Manela's website, for a total of 647 test portfolios.

The factors whose risk premia we estimate are listed in the main text. We report here the data sources. All tradable factors except BAB and QMJ are obtained from Kenneth French's website; BAB and QMJ from AQR's website; IP from the Federal Reserve Bank of St. Louis; the Macro PCs from Sydney Ludvigson's website; liquidity from Lubos Pastor's website; the intermediary factors from Bryan Kelly's website; the Novy-Marx factors from the various sources indicated in Novy-Marx (2014); the

[5] See Appendix B of the WRDS Factors Manuals for details of the sorting signals: https://wrds-www.wharton.upenn.edu/documents/1109/Backtest_Manual_v2.pdf. We use their raw signals to reconstruct test portfolios that are more conformable to the convention of the asset pricing literature.

Table II.6: Simulation Results for $n = 50$

T	Param	True	Two-Pass Estimator Bias	Two-Pass Estimator RMSE	Mimicking Portfolios using three assets Bias	Mimicking Portfolios using three assets RMSE	Mimicking Portfolios using all assets Bias	Mimicking Portfolios using all assets RMSE
	RmRf	0.372	-0.212	0.484	0.044	0.389	0.004	0.403
120	SMB	0.229	-0.133	0.326	0.039	0.262	-0.023	0.282
	HML	0.209	0.149	0.372	-0.344	0.359	0.003	0.263
	IP	-0.003	-0.009	0.088	0.003	0.008	0.000	0.028
	RmRf	0.372	-0.573	0.665	0.001	0.271	0.007	0.277
240	SMB	0.229	-0.046	0.218	0.176	0.248	-0.002	0.196
	HML	0.209	0.075	0.266	-0.380	0.388	-0.025	0.164
	IP	-0.003	-0.023	0.124	0.002	0.005	0.000	0.013
	RmRf	0.372	-0.288	0.373	0.014	0.201	0.010	0.204
480	SMB	0.229	-0.081	0.172	0.124	0.179	-0.013	0.138
	HML	0.209	0.057	0.216	-0.367	0.371	-0.032	0.112
	IP	-0.003	-0.053	0.160	0.002	0.004	0.000	0.007

Three-Pass Estimators

T	Param	True	$\check{p} = 4$ Bias	$\check{p} = 4$ RMSE	$\check{p} = 5$ Bias	$\check{p} = 5$ RMSE	$\check{p} = 6$ Bias	$\check{p} = 6$ RMSE
	RmRf	0.372	0.006	0.397	0.008	0.397	0.005	0.397
120	SMB	0.229	-0.072	0.273	-0.060	0.271	-0.046	0.269
	HML	0.209	0.027	0.208	0.015	0.209	0.002	0.210
	IP	-0.003	0.001	0.009	0.001	0.009	0.001	0.010
	RmRf	0.372	0.020	0.274	0.019	0.275	0.009	0.275
240	SMB	0.229	-0.016	0.190	-0.011	0.190	-0.002	0.191
	HML	0.209	-0.082	0.161	-0.074	0.158	-0.058	0.154
	IP	-0.003	0.000	0.006	0.000	0.007	0.000	0.007
	RmRf	0.372	0.024	0.204	0.024	0.204	0.013	0.203
480	SMB	0.229	-0.053	0.145	-0.045	0.142	-0.031	0.139
	HML	0.209	-0.043	0.110	-0.053	0.111	-0.036	0.105
	IP	-0.003	0.000	0.004	0.000	0.004	0.001	0.004

Note: In this table, we report the bias (Column "Bias") and the root-mean-square error (Column "RMSE") of the risk premia estimates using mimicking portfolios with three assets or all assets, two-pass and three-pass estimators with $\check{p} = 4$, 5, and 6, for $n = 50$, and $T = 120$, 240, and 480, respectively. The true data-generating process has five factors, and the parameters are calibrated based on the de-noised five Fama-French factors (RmRf, SMB, HML, RMW, and CMA). In this setting, there is no α. The true zero-beta rate is 0, and the true risk premia of four noisy yet observed factors (RmRf, SMB, HML, and IP) are provided in the "True" column. All numbers are in percentages.

Table II.7: Simulation Results for $n = 100$

T	Param	True	Two-Pass Estimator Bias	Two-Pass Estimator RMSE	Mimicking Portfolios using three assets Bias	Mimicking Portfolios using three assets RMSE	Mimicking Portfolios using all assets Bias	Mimicking Portfolios using all assets RMSE
120	RmRf	0.372	-0.228	0.478	0.047	0.395	0.015	0.417
	SMB	0.229	-0.114	0.315	0.070	0.272	-0.008	0.321
	HML	0.209	0.109	0.347	-0.417	0.433	0.004	0.439
	IP	-0.003	-0.010	0.098	0.002	0.008	0.000	0.073
240	RmRf	0.372	-0.223	0.377	0.043	0.283	-0.002	0.287
	SMB	0.229	-0.141	0.255	0.117	0.222	-0.010	0.202
	HML	0.209	0.220	0.328	-0.348	0.353	-0.001	0.181
	IP	-0.003	-0.024	0.114	0.002	0.006	0.000	0.020
480	RmRf	0.372	-0.301	0.373	0.030	0.202	0.001	0.205
	SMB	0.229	-0.095	0.177	0.083	0.156	-0.006	0.139
	HML	0.209	0.173	0.252	-0.389	0.393	-0.002	0.117
	IP	-0.003	-0.049	0.147	0.002	0.004	0.000	0.009

Three-Pass Estimators

T	Param	True	$\check{p} = 4$ Bias	$\check{p} = 4$ RMSE	$\check{p} = 5$ Bias	$\check{p} = 5$ RMSE	$\check{p} = 6$ Bias	$\check{p} = 6$ RMSE
120	RmRf	0.372	0.009	0.396	0.019	0.397	0.017	0.397
	SMB	0.229	-0.016	0.268	-0.018	0.269	-0.016	0.269
	HML	0.209	0.000	0.208	-0.013	0.211	-0.012	0.212
	IP	-0.003	0.000	0.010	0.000	0.011	0.000	0.011
240	RmRf	0.372	0.007	0.285	-0.005	0.285	-0.002	0.285
	SMB	0.229	-0.034	0.198	-0.018	0.196	-0.016	0.196
	HML	0.209	0.003	0.148	0.003	0.149	-0.003	0.149
	IP	-0.003	0.000	0.007	0.000	0.007	0.000	0.007
480	RmRf	0.372	0.008	0.204	0.004	0.204	0.001	0.204
	SMB	0.229	-0.030	0.140	-0.016	0.138	-0.012	0.138
	HML	0.209	-0.003	0.107	-0.004	0.108	-0.004	0.108
	IP	-0.003	0.000	0.004	0.000	0.005	0.000	0.005

Note: In this table, we report the bias (Column "Bias") and the root-mean-square error (Column "RMSE") of the risk premia estimates using mimicking portfolios with three assets or all assets, two-pass and three-pass estimators with $\check{p} = 4$, 5, and 6, for $n = 100$, and $T = 120$, 240, and 480, respectively. The true data-generating process has five factors, and the parameters are calibrated based on the de-noised five Fama-French factors (RmRf, SMB, HML, RMW, and CMA). In this setting, there is no α. The true zero-beta rate is 0, and the true risk premia of four noisy yet observed factors (RmRf, SMB, HML, and IP) are provided in the "True" column. All numbers are in percentages.

Table II.8: Simulation Results for $n = 200$

			Two-Pass Estimator		Mimicking Portfolios using three assets		Mimicking Portfolios using all assets	
T	Param	True	Bias	RMSE	Bias	RMSE	Bias	RMSE
	RmRf	0.372	-0.357	0.554	0.022	0.396	NA	NA
120	SMB	0.229	-0.074	0.302	0.088	0.280	NA	NA
	HML	0.209	0.083	0.327	-0.386	0.400	NA	NA
	IP	-0.003	-0.009	0.087	0.003	0.008	NA	NA
	RmRf	0.372	-0.344	0.459	0.059	0.285	0.000	0.293
240	SMB	0.229	-0.095	0.229	0.033	0.191	-0.003	0.227
	HML	0.209	0.122	0.266	-0.382	0.388	-0.008	0.297
	IP	-0.003	-0.026	0.130	0.002	0.006	0.002	0.051
	RmRf	0.372	-0.276	0.351	0.030	0.201	0.000	0.202
480	SMB	0.229	-0.097	0.180	0.088	0.158	-0.005	0.142
	HML	0.209	0.068	0.201	-0.419	0.422	-0.003	0.127
	IP	-0.003	-0.060	0.188	0.002	0.004	0.000	0.014
			Three-Pass Estimators					
			$\breve{p} = 4$		$\breve{p} = 5$		$\breve{p} = 6$	
T	Param	True	Bias	RMSE	Bias	RMSE	Bias	RMSE
	RmRf	0.372	-0.017	0.401	-0.005	0.402	-0.005	0.402
120	SMB	0.229	-0.005	0.273	-0.008	0.274	-0.008	0.274
	HML	0.209	0.003	0.212	-0.012	0.214	-0.012	0.214
	IP	-0.003	0.001	0.010	0.000	0.011	0.000	0.011
	RmRf	0.372	0.001	0.284	0.001	0.284	0.001	0.284
240	SMB	0.229	-0.010	0.193	-0.008	0.194	-0.008	0.194
	HML	0.209	-0.009	0.149	-0.010	0.150	-0.010	0.151
	IP	-0.003	0.000	0.007	0.000	0.007	0.000	0.007
	RmRf	0.372	0.024	0.203	0.002	0.201	0.001	0.201
480	SMB	0.229	-0.031	0.141	-0.008	0.138	-0.008	0.139
	HML	0.209	-0.032	0.110	-0.009	0.107	-0.009	0.107
	IP	-0.003	0.000	0.005	0.000	0.005	0.000	0.005

Note: In this table, we report the bias (Column "Bias") and the root-mean-square error (Column "RMSE") of the risk premia estimates using mimicking portfolios with three assets or all assets, two-pass and three-pass estimators with $\breve{p} = 4$, 5, and 6, for $n = 200$, and $T = 120$, 240, and 480, respectively. The true data-generating process has five factors, and the parameters are calibrated based on the de-noised five Fama-French factors (RmRf, SMB, HML, RMW, and CMA). In this setting, there is no α. The true zero-beta rate is 0, and the true risk premia of four noisy yet observed factors (RmRf, SMB, HML, and IP) are provided in the "True" column. All numbers are in percentages. NA means the estimators are "infeasible."

consumption factors from Toby Moskowitz's website.

III.2 Robustness to the Number of Latent Factors

Theorem I.2 shows that our results are theoretically robust to using "a few" too many latent factors in our analysis, compared to the true number p. Given the potential concern that our baseline choice of p, $\check{p} = 7$, might omit some latent factor with small eigenvalues, we explore here how the results change as we increase the number of factors. We choose $\check{p} = 10$ and $\check{p} = 13$ based on the scree plot (Figure III.3).

The results are reported in the first three columns of Table III.9, and appear mostly robust to the change in number of factors. While the significance changes in some cases (with more factors, estimates of risk premia tend to become less precise), the signs and magnitudes of the estimated risk premia remain similar as $\check{p}$ varies.

Of course, we should not expect (and Theorem I.2 does not guarantee) that the results remain the same as $\check{p}$ increases arbitrarily: in the limit, as $\check{p}$ approaches n, the estimator becomes the mimicking-portfolio estimator with all assets (which in this case is infeasible). It is however reassuring to see that almost doubling the number of factors included gives similar results.

III.3 Estimating the Zero-beta Rate

Column 4 of Table III.9 shows the results produced by our more general estimator (9), that allows the zero-beta rate to be different from the T-bill rate.[6]

The estimated zero-beta rate (which does not depend on the choice of g_t by construction) is 49bp, close to the average T-bill rate of 46bp per month in our sample. Given that the unconstrained estimates of the zero-beta rate are close to the average T-bill rate, it should not be surprising that the results for the risk premia are similar to the ones obtained when the zero-beta rate is constrained to be equal to the T-bill rate.

III.4 Robustness to the Presence of Weak Latent Factors

As we discuss in the main text, our procedure works even if the *observable* factor g_t is weak (in fact, we propose a test for whether g_t is weak); however, PCA will not necessarily recover the entire factor space if the underlying *latent* factors are weak. In this section we summarize our main theoretical and empirical arguments for using PCA in practice, and propose an additional robustness test to mitigate the concern that the presence of weak factors may distort our results.

In theory, weak latent factors — unobservable factors for which the dispersion of risk exposures is small in the cross-section — can affect our estimator because they have low eigenvalues, and PCA might fail to separate them from noise. However, for weak factors to bias our estimates of risk premia for observable factors, they also need to have themselves high risk premia, which allows them to explain a significant portion of the cross-section of average returns. But large risk premia for factors with low eigenvalues imply high Sharpe ratios. A first theoretical argument in favor of focusing on the PCs with

[6]The inference based on Theorem I.3 in this case is also robust to the presence of pricing errors (alphas) that satisfy our assumptions.

largest eigenvalues are *good-deal bounds*, which impose a theoretical upper bound on the potential bias from weak factors (Kozak et al. (2018) make precisely this argument to support using PCA in this context).

A second, empirical, argument is that we can easily add additional PCs with lower and lower eigenvalues, and verify that the risk premia estimates are stable (as shown in Table III.9).

A third way to verify that weak latent factors are not driving our empirical results is the comparison of the risk premia estimated for tradable factors using our three-pass procedure with those obtained as time-series average of the portfolios' excess returns. As discussed in the text, the two should be the same if the factor model is correctly specified. Biases due to the presence of weak latent factors should produce significant differences between the estimates using cross-sectional methods and the time-series averages.

Finally, we propose here an additional robustness test with respect to the possibility of weak factors, based on changing the objective function when extracting the statistical factors from the panel of returns. Recall that the first step towards PCA is to calculate eigenvalues of the covariance matrix of returns, which equal the variances of the corresponding PCs, and that the constructed factors are eigenvectors associated with the largest few eigenvalues.

Since weak factors are factors with low eigenvalues, which however explain the cross-section of returns, we can modify the objective function to account for the contribution to the cross-sectional variation. That is, rather than finding factors that best explain the time-series comovement of stock returns, we find factors that strike a balance between explaining the time-series comovement of stock returns and the cross-sectional variation of expected returns. This alternative objective function was first proposed by Connor and Korajczyk (1986), and has been recently extended by Lettau and Pelger (2018). It is a convenient reference point because it puts equal weight on the two components of the objective function — the time-series and the cross-sectional variation.

As shown in Bai and Ng (2002), our PCA formula given in (7) is the solution to the following optimization problem:

$$\min_{\beta,\bar{V}} n^{-1}T^{-1} \left\|\bar{R} - \beta\bar{V}\right\|_{\mathrm{F}}^2, \quad \text{subject to} \quad T^{-1}\bar{V}\bar{V}^{\intercal} = \mathbb{I}_{\widehat{p}},$$

where $\|\cdot\|_{\mathrm{F}}$ is the Frobenius norm of a matrix. By our rotation invariance result, it would give the same risk premia estimates if we were to use an alternative normalization $n^{-1}\beta^{\intercal}\beta = \mathbb{I}_n$. Connor and Korajczyk (1986) suggest another optimization problem (henceforth CK):

$$\min_{\beta,\bar{V},\gamma} n^{-1}T^{-1} \left\|\bar{R} - \beta\bar{V}\right\|_{\mathrm{F}}^2 + wn^{-1}\left\|\bar{r} - \beta\gamma\right\|_{\mathrm{F}}^2, \quad \text{subject to} \quad n^{-1}\beta^{\intercal}\beta = \mathbb{I}_n,$$

where they choose $w = 1$. The solution turns out to be

$$\tilde{\beta} = n^{1/2}(\tilde{\varsigma}_1 : \tilde{\varsigma}_2 : \ldots : \tilde{\varsigma}_{\widehat{p}}), \quad \text{and} \quad \tilde{V} = (\tilde{\beta}^{\intercal}\tilde{\beta})^{-1}\tilde{\beta}^{\intercal}\bar{R},$$

where $\tilde{\varsigma}_1, \tilde{\varsigma}_2, \ldots, \tilde{\varsigma}_{\widehat{p}}$ are the eigenvectors associated with the largest eigenvalues of the matrix $n^{-1}T^{-1}\bar{R}\bar{R}^{\intercal} + wn^{-1}\bar{r}\bar{r}^{\intercal}$. Note that starting from CK's formulation, setting $w = 0$ (thus focusing entirely on time-series comovement) is equivalent to PCA.

The CK approach can be used instead of the standard PCA in step (i) of our three-step procedure. Since the second term of the objective function is the cross-sectional R^2, it may help select latent factors that have large risk premia but are weak. We can then continue steps (ii) and (iii) as in Section 3 using the estimated latent factors together with g_t to estimate risk premia. Note that the CK approach does not allow for an unrestricted zero-beta rate or pricing errors.

In Table III.9, column CK, we report the results using the CK approach (we have not derived the standard error of our estimator when CK is used in the first step; so we only report the point estimates). The table shows that there is almost no difference in the risk premia estimates relative to the baseline, which suggests that weak factors are either not present in this dataset we consider, or if they are, they have small enough risk premia that ignoring them has little consequence for our estimates.

Taken together, these considerations lead us to conclude that for the purposes of estimating risk premia, using PCA to recover the factor space represents a simple yet robust solution.

III.5 Ridge Regression

Section 4.3 shows that instead of using PCA to reduce the dimensionality of the returns' space, we could instead use ridge regression and still obtain a consistent estimator of risk premia. Table III.9 reports the risk premia estimates in the second-to-last column. While we do not derive the asymptotic distribution of the estimator (so we do not report standard errors), the point estimates we obtain are in general quite similar to our baseline results that instead use PCA. Note that we fix the tuning parameter μ in the ridge equation (14) to be $\lambda_{2\times\check{p}}(\bar{R}\bar{R}^\intercal)$, so that the ridge results effectively serve as a benchmark of a large factor model.

III.6 Robustness to the Choice of Test Portfolios

Our main empirical results are obtained using a large set of 647 portfolios, that spans equities, bonds, and currencies. It is natural, however, to wonder to what extent the results are affected by the particular selection of test assets.

We explore robustness with respect to the choice of test portfolios in two ways. First, we perform the estimation using only equity portfolios. The results are in the last column of Table III.9. The results are similar to those of the baseline.

Second, we perform a bootstrap-type analysis that excludes systematically random subsets of assets. In particular, from the 647 test portfolios we use in our empirical exercise, we randomly select (without replacement) half of them, and we re-estimate the risk premia of all observable factors in this subsample. We repeat this exercise 10,000 times, thus obtaining a distribution of risk premia estimates across subsamples of 323 portfolios each, randomly selected.

Appendix Figure III.5 shows the results for several factors. Note that all panels of the figure report the same range of risk premia (x axis, between -20bp and 100bp), so that the histograms are easily comparable across panels. The results are heterogeneous across factors. In the top left panel, we see that the risk premium for the market return is precisely estimated, from essentially any of the random subsets of the assets. The top right panel shows that the risk premia estimates for SMB and HML vary more across subsets of assets. The next panel shows that momentum's risk premium varies even more across subsets, but it is still estimated to be between 25bp and 70bp in almost all subsamples.

The last three panels show interesting results for non-tradable factors. Confirming the results of Table 1, IP is a useless factor, with a risk premium of effectively zero in all subsamples. On the contrary, liquidity and intermediary capital factors all appear positively priced across subsamples.

Overall, our subsample results show that the conclusions of our empirical analysis are very robust to the selection of the test assets, at least within the universe of assets we consider (equity, bonds, currencies). One caveat worth keeping in mind when interpreting these results is that this analysis randomly selects (without replacement) half of the assets within the original set of 647 portfolios – but the original universe was itself not randomly selected in the first place, since it is based on characteristics proposed in the existing literature.

III.7 Robustness to the Choice of Estimation Time Period

A potential concern when working with PCs is the stability of the estimated loadings and factors over time. The extent to which our risk premia estimates are consistent across time periods is an empirical question that we explore in this section.

Similarly to the robustness with respect to the test assets, we perform our robustness check with respect to the sample period by resampling half of the time periods randomly without replacement, and looking at the variability of the risk premia estimates. Simple resampling in the time series is possible in our context because of the low serial correlation of returns and factor innovations over time.

Figure III.6 shows the results. Interestingly, the estimates are more variable across time subsamples compared to the case in which test assets were resampled (previous section). That said, the estimated risk premia remain positive across almost all subsamples, for all factors (except IP, which is clearly a weak factor, and whose risk premium is precisely estimated to be zero); our results are therefore quite stable across subsamples.

Note that the variability of estimates in this analysis is necessarily higher than the variability captured by the standard errors of our full-sample estimates: we are resampling periods with *half* as many observations than in our full-sample test.

Despite the increased variability, the estimates are quite stable across subsamples. This may appear surprising, because our estimator is based on PCA, which is known to give different factor estimates (rotations) in different subsamples. However, it is useful to note that our risk premia estimator is not *only* based on PCA. Instead, a key step is the projection of the factor of interest g_t onto the extracted PCs. So any rotation that makes the extracted factors differ across subsamples will be entirely offset by a corresponding rotation of the loading of g_t onto those factors – resulting in stable risk premia estimates for the observable factors.

III.8 Separating the Measurement Error from the Factors

The R^2 of the projection of each factor g_t onto the latent factors (R^2_g in Table 1) reveals the amount of measurement error in the factor. Figure III.7 further explores how our method allows us to "clean" the factors from measurement error. It shows the time series of cumulated innovations in the original and cleaned (i.e., fitted) factors, for a few of them. The figures provide a graphical representation of the extent to which the PCs of returns capture the variation in each factor. While for several factors the original and cleaned factors track each other closely (e.g., for the market, SMB and HML plotted in the figure), for others the cleaned factor displays much lower variation than the original factor: the difference is the measurement error that our procedure has eliminated (a nice example is IP, plotted in the figure, which is identified as a weak factor).

III.9 Individual Assets vs. Portfolios

In this paper, we recommend using characteristic-sorted portfolios instead of individual stocks. The main advantage of using portfolios is that their risk exposures are more stable over time, as discussed at length in the asset pricing literature. This is particularly important in our setting, because we assume the betas of the test assets are constant.

To see this intuition more formally, call $\tilde{r}_t$ the vector of time-t returns for m individual stocks, and c_t a $m \times n$ matrix of characteristics (or their functions) observed at time t for the m stocks. The typical procedure to construct characteristic-sorted portfolios in asset pricing categorizes stocks at each time $t-1$ into groups based on one or more observed characteristics, and then obtains the portfolio return at time t using equal or market-value weights for stocks in each group.

The sorting procedure can be represented mathematically by constructing the matrix c_{t-1} stacking side-by-side the n dummy variables corresponding to each characteristic-sorted group. For example, to construct 10 size-based portfolios, c_{t-1} would be an $m \times 10$ matrix containing 10 dummy variables, each indicating the size group to which each stock belongs at time $t-1$. The n characteristic-sorted portfolio returns from $t-1$ to t are simply the coefficients of a cross-sectional regression of $\tilde{r}_t$ onto c_{t-1}, since c_{t-1} contains only dummies.

More generally, given any matrix c_{t-1} (that could include dummies or continuous variables), the characteristics-weighted portfolio returns at time t are:

$$r_t = (c_{t-1}^{\intercal} c_{t-1})^{-1} c_{t-1}^{\intercal} \tilde{r}_t, \tag{III.4}$$

where the term $(c_{t-1}^{\intercal} c_{t-1})^{-1} c_{t-1}$ therefore represents the time-$(t-1)$ portfolio weights.

Using this expression that links r_t and $\tilde{r}_t$, it is immediate to find that if individual factor exposures are linear functions of c_{t-1} (e.g., Rosenberg (1974)), then the sorted portfolios have constant factor exposures. Specifically, extending our setup (1) to include time-varying factor exposures for individual asset returns, we have:

$$\tilde{r}_t = \beta_{t-1}\gamma_{t-1} + \beta_{t-1}\tilde{v}_t + \tilde{u}_t,$$

where $\tilde{r}_t$ and $\tilde{u}_t$ are $m \times 1$ vectors, β_{t-1} is an $m \times n$ matrix of time-varying exposures, following

$$\beta_{t-1} = c_{t-1}\beta + \varepsilon_{t-1}, \tag{III.5}$$

for some $n \times p$ matrix β, $m \times n$ matrix of observable characteristics c_{t-1}, and some $n \times p$ matrix of residuals ε_{t-1}. Prior to applying our three-pass estimation procedure, we construct characteristics-sorted portfolios:

$$r_t = (c_{t-1}^{\intercal} c_{t-1})^{-1} c_{t-1}^{\intercal} \tilde{r}_t = \beta\gamma + \beta v_t + u_t,$$

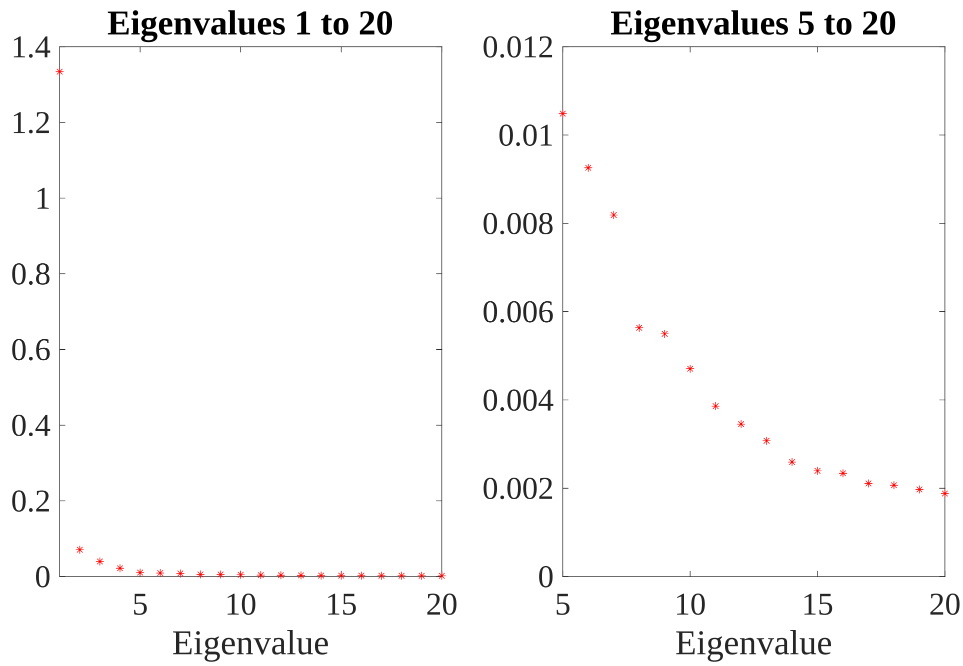
where

$$\gamma = \mathrm{E}(\gamma_{t-1}), \quad v_t = \tilde{v}_t + \gamma_{t-1} - \mathrm{E}(\gamma_{t-1}), \quad u_t = (c_{t-1}^{\intercal} c_{t-1})^{-1} c_{t-1}^{\intercal} \left(\tilde{u}_t + \varepsilon_{t-1}(\gamma_{t-1} + \tilde{v}_t)\right).$$

Therefore, our methodology to estimate risk premia can be applied even if individual stock risk exposures are time-varying, as long as characteristic-sorted portfolios that have constant factor exposures are used as test assets, provided that u_t and v_t satisfy assumptions in this paper. Also, we can interpret the estimated risk premia as estimates of their time-series average.

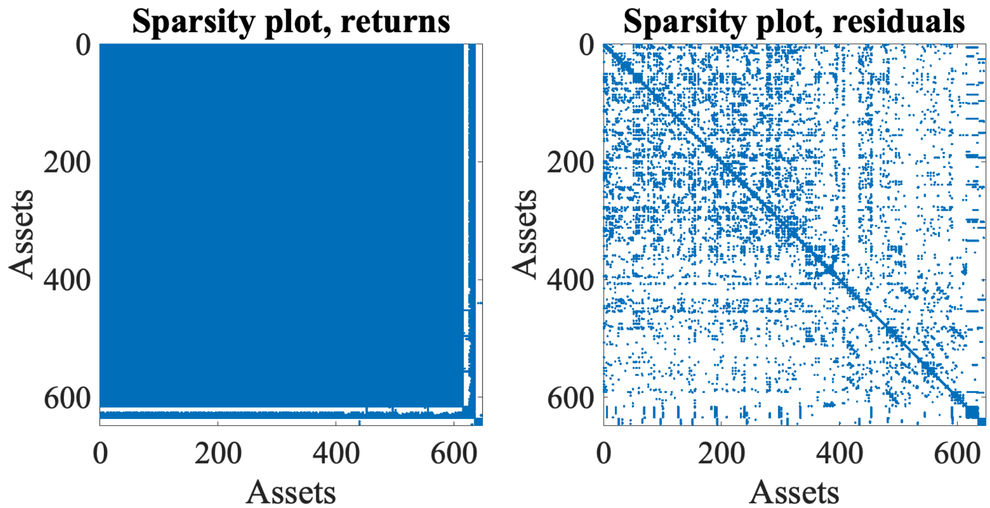
In this paper, we take the portfolio-formation step as given, and use characteristic-sorted portfolios that have been proposed in the literature. In contrast, Kelly et al. (2019) construct such portfolios using characteristics and individual stocks for a model specification test. Their results show that PCs based on such portfolios explain more cross-sectional variations than those based on individual stocks, which is consistent with the formal result shown above that characteristic-sorted portfolios will have constant betas if the characteristics are chosen appropriately.

Figure III.3: First Fifteen Eigenvalues of the Covariance Matrix of 647 Test Portfolios



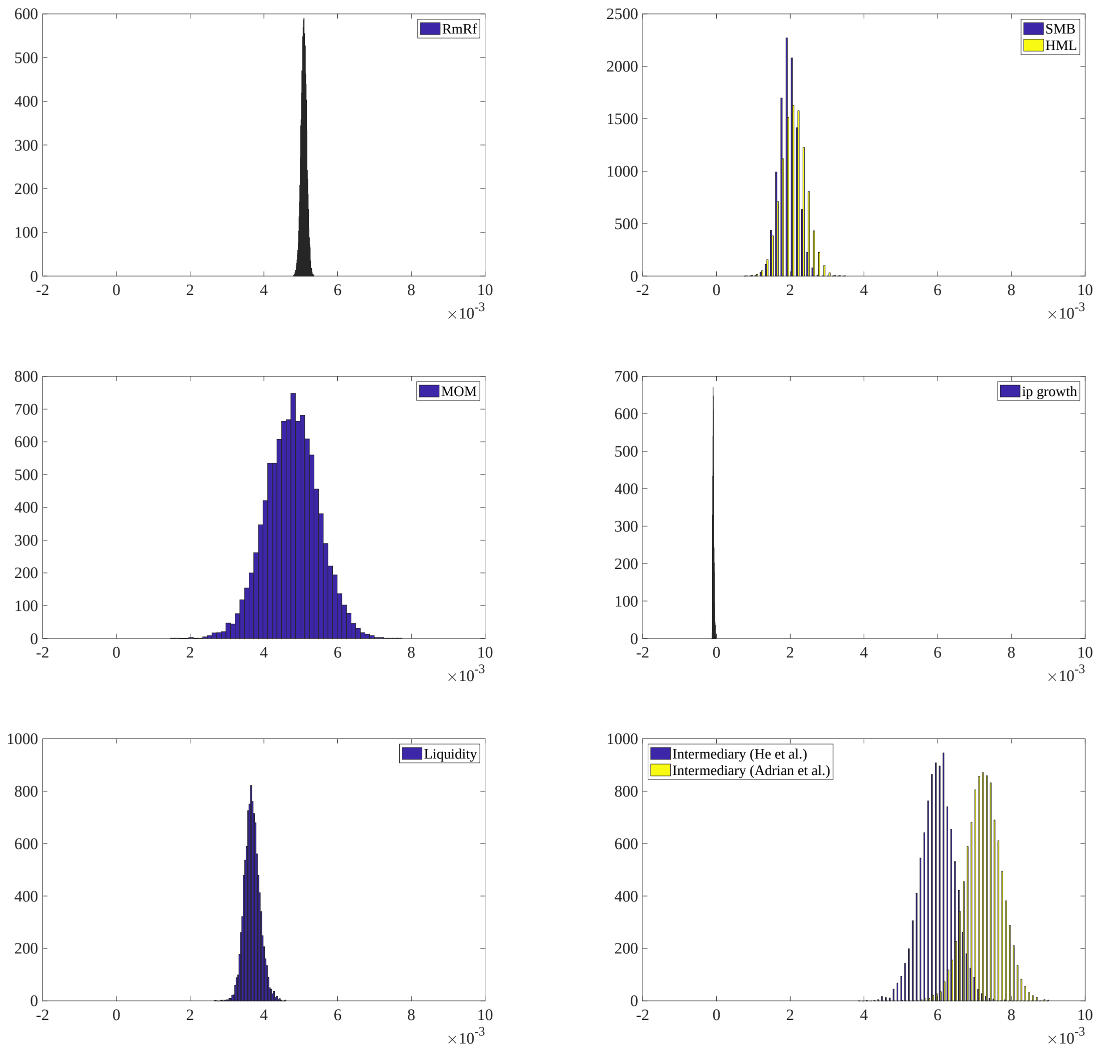
Note: The left panel reports the first 20 eigenvalues of the covariance matrix of our 647 test portfolios. The right panel zooms in to the eigenvalues 5 through 20.

Figure III.4: Sparsity plot of portfolio returns



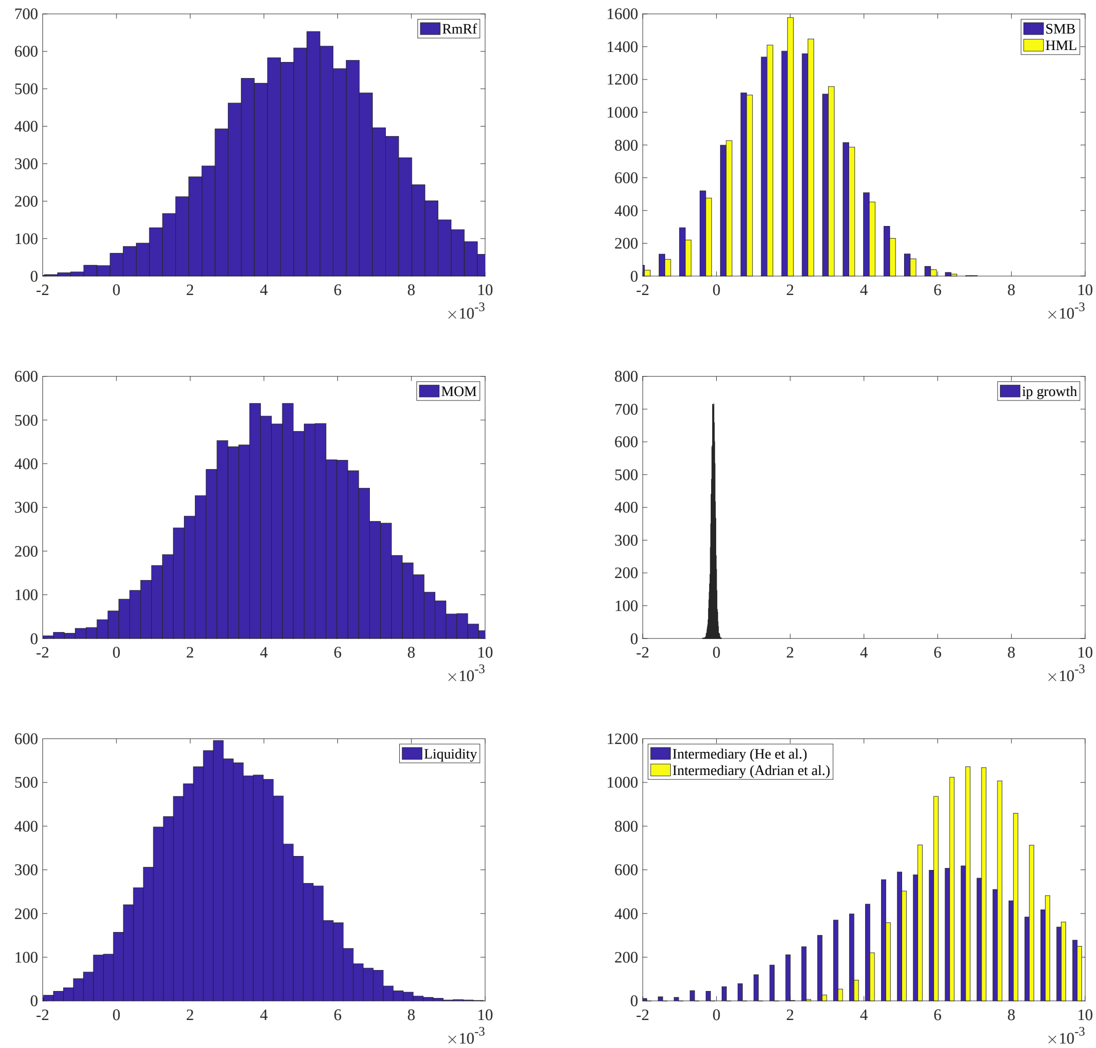
Note: The left panel reports a sparsity plot based on the correlation matrix of excess returns of the 647 portfolios. This matrix plots a dot if the corresponding correlation is above 0.25, and no dot if it is below 0.25. The right-hand side reports the same sparsity plot, but for the residuals of our 7-factor model. Assets are our 647 portfolios: assets 1-413 are the characteristic-sorted portfolios from WRDS, 414-438 are the portfolios sorted by size and book-to-market, 439-455 are industry portfolios, 456-480 are portfolios sorted by operating profitability and investment, 481-505 are portfolios sorted by size and variance, 506-540 are portfolios sorted by size and net issuance, 541-565 are portfolios sorted by size and beta, 566-590 are portfolios sorted by size and accruals, 591-615 are portfolios sorted by size and momentum, 616-635 are bond portfolios and 636-647 are currency portfolios.

Figure III.5: Robustness to the Set of Test Portfolios: Resampling Exercise



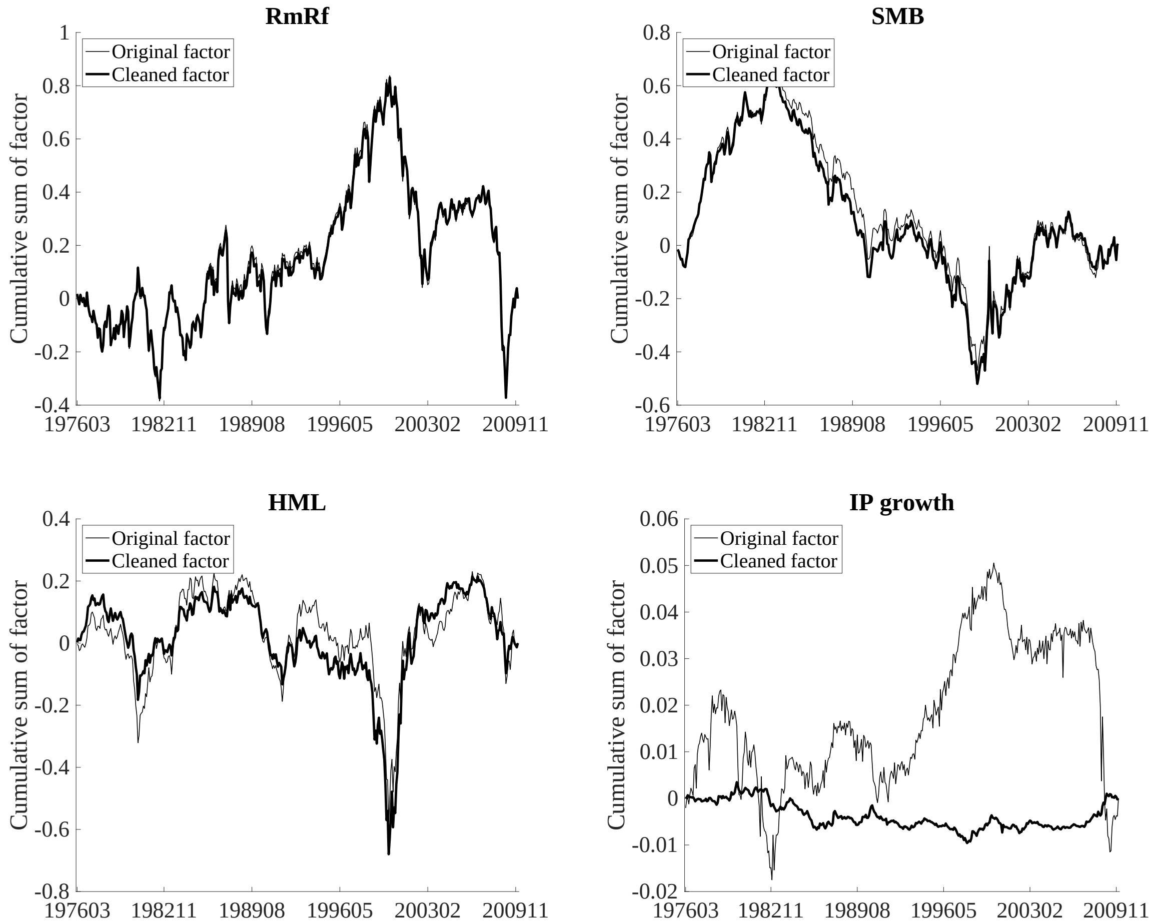
Note: This figure reports the histograms of risk premia estimated using the three-pass estimator across subsamples of the set of 647 test portfolios. We generate 10,000 subsamples by randomly drawing (without replacement) half of the portfolios from the baseline set of 647 portfolios. In each sample we estimate the risk premium of each factor using the three-pass estimator, setting $\check{p} = 7$. The histogram reports the frequency of the risk premia estimates across samples. All figures report the same range for the risk premia, between -20bp and 100bp per month.

Figure III.6: Robustness to the Time Period: Resampling Exercise



Note: This figure reports the histograms of risk premia estimated using the three-pass estimator across subsamples of the time period. We generate 10,000 subsamples by randomly drawing (without replacement) half of the available time periods (using all 647 portfolios). In each sample we estimate the risk premium of each factor using the three-pass estimator, setting $\check{p} = 7$. The histogram reports the frequency of the risk premia estimates across samples. All figures report the same range for the risk premia, between -20bp and 100bp per month.

Figure III.7: Cumulative Factor Time Series with and without Measurement Error



Note: This figure reports the time series of cumulative factor innovations for RmRf, SMB, HML, and IP (thin line) together with the time series obtained from removing measurement error from the factor (thick line).

Table III.9: Additional Empirical Results

Factors	baseline $\check{p} = 7$ γ	stderr	$\check{p} = 10$ γ	stderr	$\check{p} = 13$ γ	stderr	free zero-beta γ	stderr	CK γ	Ridge γ	only equities γ	stderr
Market	0.51^{**}	(0.23)	0.51^{**}	(0.23)	0.50^{**}	(0.23)	0.48^{**}	(0.23)	0.51	0.51	0.51^{**}	(0.23)
SMB	0.20	(0.16)	0.20	(0.16)	0.17	(0.16)	0.20	(0.16)	0.21	0.21	0.19	(0.16)
HML	0.20	(0.15)	0.26^{*}	(0.15)	0.35^{**}	(0.16)	0.20	(0.16)	0.22	0.27	0.20	(0.15)
MOM	0.49^{**}	(0.23)	0.53^{**}	(0.24)	0.61^{**}	(0.24)	0.48^{**}	(0.23)	0.54	0.50	0.46^{**}	(0.23)
RMW	0.22^{*}	(0.11)	0.08	(0.12)	0.12	(0.12)	0.22^{*}	(0.12)	0.24	0.22	0.20^{*}	(0.11)
CMA	0.14	(0.10)	0.28^{***}	(0.10)	0.30^{***}	(0.10)	0.14	(0.10)	0.16	0.22	0.14	(0.09)
BAB	0.57^{***}	(0.15)	0.48^{***}	(0.14)	0.46^{***}	(0.16)	0.56^{***}	(0.15)	0.61	0.51	0.56^{***}	(0.14)
QMJ	0.06	(0.13)	0.02	(0.14)	0.03	(0.14)	0.06	(0.13)	0.08	0.10	0.03	(0.13)
Liquidity	0.37^{**}	(0.16)	0.32^{**}	(0.16)	0.23	(0.16)	0.35^{**}	(0.16)	0.37	0.20	0.37^{***}	(0.13)
Interm. (He)	0.60^{**}	(0.31)	0.60^{*}	(0.31)	0.69^{**}	(0.32)	0.56^{*}	(0.31)	0.60	0.65	0.54^{*}	(0.30)
Interm. (Adrian)	0.72^{***}	(0.16)	0.87^{***}	(0.19)	1.01^{***}	(0.19)	0.70^{***}	(0.17)	0.77	0.86	0.69^{***}	(0.16)
NY temp.	−0.69	(13.90)	−16.02	(19.62)	−23.62	(17.66)	0.30	(13.97)	-1.26	10.00	−0.44	(9.11)
Global temp.	0.05	(0.21)	0.06	(0.33)	0.13	(0.36)	0.03	(0.21)	0.08	-0.14	0.06	(0.16)
El Niño	0.41	(0.82)	0.77	(1.10)	1.66	(1.08)	0.38	(0.82)	0.43	0.56	0.41	(0.56)
Sunspots	4.01	(35.63)	−3.49	(53.13)	2.04	(54.53)	7.10	(35.35)	4.31	21.86	1.77	(23.50)
IP growth	-0.01^{*}	(0.00)	−0.01	(0.01)	−0.01	(0.01)	-0.01^{*}	(0.00)	-0.01	-0.01	-0.01^{**}	(0.00)
Macro PC 1	3.26^{**}	(1.58)	2.61	(2.16)	2.56	(2.19)	3.13^{**}	(1.58)	3.54	3.31	3.02^{**}	(1.32)
Macro PC 2	−0.88	(1.27)	0.28	(1.71)	−0.24	(1.57)	−0.77	(1.28)	-0.65	-0.82	−1.33	(1.06)
Macro PC 3	−1.25	(1.51)	−1.87	(1.81)	−1.29	(1.90)	−1.33	(1.56)	-1.30	-1.88	−1.58	(1.36)
Cons. growth	0.00	(0.01)	0.01	(0.01)	0.00	(0.01)	0.00	(0.01)	0.01	-0.00	0.00	(0.01)
Stockholder cons.	0.17^{**}	(0.08)	0.08	(0.10)	0.10	(0.11)	0.16^{**}	(0.08)	0.18	0.04	0.16^{***}	(0.06)

Note: For each factor, the table reports the risk premia estimates using different versions of the three-pass estimator. The first column reports the baseline case for convenience (7 factors). Columns 2 and 3 vary the number of factors used (to 10 and 13, respectively), restricting the zero-beta rate to be equal to the T-bill rate (as in table 1). Columns 4 estimates the zero-beta rate instead of setting it to the t-bill rate. Column 5 reports the point estimates using the Connor and Korajczyk (1986) methodology to extract factors. Column 6 uses ridge regression instead of PCA when building mimicking portfolios. Column 7 uses only the equity portfolios instead of the full set of 647 portfolios.

IV Technical Lemmas and Their Proofs

To prove the main theorems of the paper, we need the following lemmas:

Lemma 1. *Under Assumptions A.1, A.2, A.4, A.5, A.6, and A.7, and suppose that $\widehat{p} = p$, we have*

$$\left\| \widehat{V} - H\bar{V} \right\|_{\mathrm{F}} = O_p(n^{-1/2}T^{1/2} + 1).$$

Proof. We make use of the following decomposition:

$$\widehat{V} - H\bar{V} = n^{-1}T^{-1}\left(\widehat{\Lambda}^{-1}\widehat{V}\bar{R}^{\intercal}\bar{R} - \widehat{\Lambda}^{-1}\widehat{V}\bar{V}^{\intercal}\beta^{\intercal}\beta\bar{V}\right) = n^{-1}T^{-1}\widehat{\Lambda}^{-1}\widehat{V}\left(\bar{U}^{\intercal}\beta\bar{V} + \bar{V}^{\intercal}\beta^{\intercal}\bar{U} + \bar{U}^{\intercal}\bar{U}\right). \quad \text{(IV.6)}$$

Note that by (V.36), we have

$$\left\|\bar{V}\right\|_{\mathrm{F}} \leq K\left\|\bar{V}\right\| = O_p(T^{1/2}). \quad \text{(IV.7)}$$

Also, we have

$$\left\|\widehat{V}\right\| \leq \left\|\widehat{V}\right\|_{\mathrm{F}} \leq \left\|\widehat{V}\widehat{V}^{\intercal}\right\|_{\mathrm{F}}^{1/2} = T^{1/2}. \quad \text{(IV.8)}$$

Using (V.39), we have

$$\left\|\widehat{\Lambda}\right\|_{\mathrm{MAX}} = O_p(1), \quad \text{and} \quad \left\|\widehat{\Lambda}^{-1}\right\|_{\mathrm{MAX}} = O_p(1). \quad \text{(IV.9)}$$

Combining these estimates with (V.34), we obtain

$$n^{-1}T^{-1}\left\|\widehat{\Lambda}^{-1}\widehat{V}\bar{U}^{\intercal}\beta\bar{V}\right\|_{\mathrm{F}} \leq Kn^{-1}T^{-1}\left\|\widehat{\Lambda}^{-1}\right\|_{\mathrm{MAX}}\left\|\widehat{V}\right\|\left\|\bar{V}\bar{U}^{\intercal}\beta\right\| = O_p(n^{-1/2}T^{1/2}).$$

The same bound holds for another term:

$$n^{-1}T^{-1}\left\|\widehat{\Lambda}^{-1}\widehat{V}\bar{V}^{\intercal}\beta^{\intercal}\bar{U}\right\|_{\mathrm{F}} = O_p(n^{-1/2}T^{1/2}).$$

Using (V.31), (IV.8), and (IV.9), we obtain

$$n^{-1}T^{-1}\left\|\widehat{\Lambda}^{-1}\widehat{V}\bar{U}^{\intercal}\bar{U}\right\|_{\mathrm{F}} \leq Kn^{-1}T^{-1}\left\|\widehat{\Lambda}^{-1}\widehat{V}\bar{U}^{\intercal}\bar{U}\right\| \leq K\left(n^{-1/2}T^{1/2} + 1\right),$$

which concludes the proof.

□

Lemma 2. *Under Assumptions A.1, A.2, A.4, A.5, A.6, and A.7, and suppose that $\widehat{p} = p$, if follows that H is invertible with probability approaching 1. Moreover, $\|H\| = O_p(1)$, $\left\|H^{-1}\right\| = O_p(1)$, and $\left\|H^{\intercal}H - (\Sigma^v)^{-1}\right\| = O_p(n^{-1/2} + T^{-1/2})$.*

Proof. Note that

$$\left\|\bar{V}\right\| = O_p(T^{1/2}), \quad \left\|\widehat{V}\right\| \le \left\|\widehat{V}\right\|_{\mathrm{F}} = T^{1/2}, \quad n^{-1}\left\|\beta^\intercal\beta\right\| \le \left\|n^{-1}\beta^\intercal\beta - \Sigma^\beta\right\| + \left\|\Sigma^\beta\right\| \le O_p(1),$$

it follows from (B.2) that

$$\|H\| \le n^{-1}T^{-1}\left\|\widehat{V}\right\|\left\|\bar{V}\right\|\left\|\beta^\intercal\beta\right\|\left\|\widehat{\Lambda}^{-1}\right\| = O_p(1).$$

Moreover, by triangle inequalities, Assumption A.5, and Lemma 1, we have

$$\begin{aligned}\|H\Sigma^v H^\intercal - \mathbb{I}_p\| \le& \left\|H\Sigma^v H^\intercal - T^{-1}H\bar{V}\bar{V}^\intercal H^\intercal\right\| + \left\|T^{-1}H\bar{V}\bar{V}^\intercal H^\intercal - \mathbb{I}_p\right\| \\ \le& \|H\|^2\left(\left\|\Sigma^v - T^{-1}VV^\intercal\right\| + \|\bar{v}\bar{v}^\intercal\|\right) + T^{-1}\left\|\widehat{V} - H\bar{V}\right\|_{\mathrm{F}}\left\|H\bar{V}\right\| + T^{-1}\left\|\widehat{V}\right\|\left\|\widehat{V} - H\bar{V}\right\|_{\mathrm{F}} \\ =& O_p(n^{-1/2} + T^{-1/2}). \end{aligned} \tag{IV.10}$$

By Weyl's inequality, we have $\lambda_{\min}(H\Sigma^v H^\intercal) > 1/2$ with probability approaching 1. This implies that $H(\Sigma^v)^{1/2}$ is invertible with probability approaching 1, so is H. Moreover, since $\lambda_{\min}(H\Sigma^v H^\intercal) \le \lambda_{\max}(\Sigma^v)\|H\|^2$, there exists $\epsilon > 0$ such that $\|H\| \ge \epsilon$ with probability approaching 1, and hence $\left\|H^{-1}\right\| = O_p(1)$.

Multiplying H^{-1} and $H^{-\intercal}$ from each side of (IV.10) respectively, we obtain

$$\left\|\Sigma^v - H^{-1}H^{-\intercal}\right\| = O_p(n^{-1/2} + T^{-1/2}), \tag{IV.11}$$

so that multiplying $(\Sigma^v)^{-1}$ from the right hand side gives

$$\left\|\mathbb{I}_p - H^{-1}H^{-\intercal}(\Sigma^v)^{-1}\right\| = O_p(n^{-1/2} + T^{-1/2}).$$

Finally, multiplying $H^\intercal H$ from the left gives the desired result. □

Lemma 3. *Under Assumptions A.1, A.2, A.4, A.5, A.6, A.7, A.9, and $\widehat{p} = p$, we have*

$$\begin{aligned} &(a) \quad T^{-1}\left\|\left(H\bar{V} - \widehat{V}\right)\bar{V}^\intercal\right\|_{\mathrm{MAX}} = O_p(n^{-1} + T^{-1}). \\ &(b) \quad \left\|\beta - \widehat{\beta}H\right\|_{\mathrm{F}} = O_p(1 + n^{1/2}T^{-1/2}). \end{aligned}$$

Proof. (a) By (IV.6) and (IV.9), we have

$$\left\|\left(H\bar{V} - \widehat{V}\right)\bar{V}^\intercal\right\|_{\mathrm{MAX}} \le Kn^{-1}T^{-1}\left\|\widehat{V}\bar{U}^\intercal\beta\bar{V}\bar{V}^\intercal + \widehat{V}\bar{V}^\intercal\beta^\intercal\bar{U}\bar{V}^\intercal + \widehat{V}\bar{U}^\intercal\bar{U}\bar{V}^\intercal\right\|_{\mathrm{MAX}}.$$

(i) To bound the first term, we note that

$$\left\|\widehat{V}\bar{U}^\intercal\beta\right\|_{\mathrm{MAX}} \le K\left\|\widehat{V} - H\bar{V}\right\|_{\mathrm{F}}\left\|\bar{U}^\intercal\beta\right\|_{\mathrm{F}} + K\|H\|_{\mathrm{MAX}}\left\|\bar{V}\bar{U}^\intercal\beta\right\|_{\mathrm{MAX}}.$$

Since $\|\bar{v}\|_{\text{MAX}} = O_p(T^{-1/2})$, it follows from (V.33) that

$$\|\bar{v}\bar{u}^\intercal \beta\|_{\text{MAX}} \leq K \|\bar{v}\|_{\text{MAX}} \|\bar{u}^\intercal \beta\|_{\text{MAX}} = O_p(n^{1/2}T^{-1}). \tag{IV.12}$$

Combining with Assumption A.9(ii), we have

$$\left\|\bar{V}\bar{U}^\intercal \beta\right\|_{\text{MAX}} \leq \|VU^\intercal \beta\|_{\text{MAX}} + T\|\bar{v}\bar{u}^\intercal \beta\|_{\text{MAX}} = O_p(n^{1/2}T^{1/2}). \tag{IV.13}$$

By Lemmas 1 and 2, (V.34), and (IV.13), we obtain

$$\left\|\widehat{V}\bar{U}^\intercal \beta\right\|_{\text{MAX}} = O_p(T + n^{1/2}T^{1/2}). \tag{IV.14}$$

Therefore, by (V.35), we have

$$\left\|\widehat{V}\bar{U}^\intercal \beta\bar{V}\bar{V}^\intercal\right\|_{\text{MAX}} \leq K\left\|\widehat{V}\bar{U}^\intercal \beta\right\|_{\text{MAX}} \|\bar{V}\bar{V}^\intercal\|_{\text{MAX}} = O_p(T^2 + n^{1/2}T^{3/2}).$$

(ii) To bound the second term, by Weyl's inequalities and Assumption A.6,

$$\left|\lambda_{\min}\left(\frac{1}{n}\beta^\intercal \beta\right) - \lambda_{\min}\left(\Sigma^\beta\right)\right| \leq \left\|\frac{1}{n}\beta^\intercal\beta - \Sigma^\beta\right\| = o_p(1).$$

Therefore, there exists some $0 < \varepsilon < \lambda_{\min}(\Sigma^\beta)$, such that

$$\lambda_{\min}(\beta^\intercal\beta) \geq n\left(\lambda_{\min}(\Sigma^\beta) - \varepsilon\right),$$

which establishes that

$$\left\|(n^{-1}\beta^\intercal\beta)^{-1}\right\| = n\lambda_{\min}^{-1}(\beta^\intercal\beta) = O_p(1), \tag{IV.15}$$

so that

$$\left\|\widehat{V}\bar{V}^\intercal\right\|_{\text{MAX}} = nT\left\|\widehat{\Lambda}H(\beta^\intercal\beta)^{-1}\right\|_{\text{MAX}} \leq KnT\left\|\widehat{\Lambda}\right\|_{\text{MAX}} \|H\|_{\text{MAX}} \left\|(\beta^\intercal\beta)^{-1}\right\| = O_p(T). \tag{IV.16}$$

Using (IV.13) we have

$$\left\|\widehat{V}\bar{V}^\intercal\beta^\intercal\bar{U}\bar{V}^\intercal\right\|_{\text{MAX}} \leq K\left\|\widehat{V}\bar{V}^\intercal\right\|_{\text{MAX}} \|\beta^\intercal\bar{U}\bar{V}^\intercal\|_{\text{MAX}} = O_p(n^{1/2}T^{3/2}).$$

(iii) Finally, we have

$$\left\|\widehat{V}\bar{U}^\intercal\bar{U}\bar{V}^\intercal\right\|_{\text{MAX}} \leq \left\|(\widehat{V} - H\bar{V})\bar{U}^\intercal\bar{U}\bar{V}^\intercal\right\|_{\text{MAX}} + \|H\bar{V}\bar{U}^\intercal\bar{U}\bar{V}^\intercal\|_{\text{MAX}}.$$

On the one hand, note that

$$\left\|(\widehat{V} - H\bar{V})\bar{U}^\intercal\bar{U}\bar{V}^\intercal\right\|_{\text{MAX}} \leq K\left\|\widehat{V} - H\bar{V}\right\|_{\text{F}} \|\bar{U}^\intercal\bar{U}\| \|\bar{V}^\intercal\|$$

Combining with (V.36), (V.31), and Lemma 1, we have

$$\left\|(\widehat{V} - H\bar{V})\bar{U}^\intercal\bar{U}\bar{V}^\intercal\right\|_{\text{MAX}} = O_p(T^2 + nT). \tag{IV.17}$$

On the other hand, we have

$$\left\|\bar{U}\bar{V}^\intercal\right\|_{\text{F}} \leq \|UV^\intercal\|_{\text{F}} + T\left\|\bar{u}\bar{v}^\intercal\right\|_{\text{F}}.$$

By Assumption A.9(i):

$$\text{E}\left\|UV^\intercal\right\|_{\text{F}}^2 \leq \sum_{i=1}^{p}\sum_{j=1}^{n}\text{E}\left(\sum_{t=1}^{T} v_{it}u_{jt}\right)^2 \leq KnT.$$

Also, by Assumption A.5 and Equation (V.27),

$$\|\bar{v}\|_{\text{F}} \leq p^{1/2}\|\bar{v}\|_{\text{MAX}} = O_p(T^{-1/2}), \quad \|\bar{u}\|_{\text{F}} = O_p(n^{1/2}T^{-1/2}),$$

so that $T\left\|\bar{u}\bar{v}^\intercal\right\|_{\text{F}} = O_p(n^{1/2})$, hence we obtain that

$$\left\|\bar{U}\bar{V}^\intercal\right\|_{\text{F}} = O_p(n^{1/2}T^{1/2}), \tag{IV.18}$$

and that

$$\left\|H\bar{V}\bar{U}^\intercal\bar{U}\bar{V}^\intercal\right\|_{\text{MAX}} \leq K\|H\|_{\text{MAX}}\left\|\bar{U}\bar{V}^\intercal\right\|_{\text{F}}^2 = O_p(nT).$$

Therefore, we obtain

$$\left\|\widehat{V}\bar{U}^\intercal\bar{U}\bar{V}^\intercal\right\|_{\text{MAX}} = O_p\left(T^2 + nT\right).$$

Combining (i), (ii), and (iii), we have

$$T^{-1}\left\|\left(H\bar{V} - \widehat{V}\right)\bar{V}^\intercal\right\|_{\text{MAX}} = O_p(n^{-1} + T^{-1}).$$

To prove (b), using the following decomposition,

$$\beta - \widehat{\beta}H = -T^{-1}\left(\beta H^{-1}(H\bar{V} - \widehat{V})\widehat{V}^\intercal H + \bar{U}\left(\widehat{V}^\intercal - \bar{V}^\intercal H^\intercal\right)H + \bar{U}\bar{V}^\intercal H^\intercal H\right),$$

it follows from Lemmas 1 and 2, (V.27), (V.28), (IV.8), (IV.18), and $\|\beta\|_{\text{F}} \leq p^{1/2}\|\beta\| = O_p(n^{1/2})$, that

$$\begin{aligned} n^{-1}\left\|\beta - \widehat{\beta}H\right\|_{\text{F}} \leq& Kn^{-1}T^{-1}\|\beta\|_{\text{F}}\left\|H^{-1}\right\|_{\text{MAX}}\left\|H\bar{V} - \widehat{V}\right\|_{\text{F}}\left\|\widehat{V}^\intercal\right\|_{\text{F}}\|H\|_{\text{MAX}} \\ &+ Kn^{-1}T^{-1}\left\|\bar{U}\right\|_{\text{F}}\left\|\widehat{V}^\intercal - \bar{V}^\intercal H^\intercal\right\|_{\text{F}}\|H\|_{\text{MAX}} + Kn^{-1}T^{-1}\left\|\bar{U}\bar{V}^\intercal\right\|_{\text{F}}\|H\|_{\text{MAX}}^2 \\ =& O_p(n^{-1} + n^{-1/2}T^{-1/2}). \end{aligned}$$

□

Lemma 4. *Under Assumptions A.1, A.2, A.4, A.5, A.6, A.7, A.9, and $\widehat{p} = p$, we have*

$$
\begin{aligned}
&(a) \quad n^{-1} \left\| H^{-\intercal}\beta^\intercal \bar{u} \right\|_{\mathrm{MAX}} = O_p\left(n^{-1/2}T^{-1/2}\right).\\
&(b) \quad n^{-1} \left\| H^{-\intercal}\beta^\intercal(\beta - \widehat{\beta}H) \right\|_{\mathrm{MAX}} = O_p\left(n^{-1} + T^{-1}\right).\\
&(c) \quad n^{-1} \left\| \left(\widehat{\beta} - \beta H^{-1}\right)^\intercal \bar{u} \right\|_{\mathrm{MAX}} = O_p(n^{-1} + T^{-1}).\\
&(d) \quad n^{-1} \left\| H^{-\intercal}\beta^\intercal(\beta - \widehat{\beta}H)\bar{v} \right\|_{\mathrm{MAX}} = O_p(n^{-1}T^{-1/2} + T^{-3/2}).\\
&(e) \quad n^{-1} \left\| (\widehat{\beta}^\intercal - H^{-\intercal}\beta^\intercal)(\beta - \widehat{\beta}H) \right\|_{\mathrm{MAX}} = O_p(n^{-1} + T^{-1}).
\end{aligned}
$$

Proof. For (a), by (V.33), we have

$$
n^{-1} \left\| H^{-\intercal}\beta^\intercal \bar{u} \right\|_{\mathrm{MAX}} \leq K n^{-1} \left\| H^{-1} \right\|_{\mathrm{MAX}} \left\| \beta^\intercal \bar{u} \right\|_{\mathrm{F}} = O_p\left(n^{-1/2}T^{-1/2}\right).
$$

As to (b), we have

$$
\begin{aligned}
&n^{-1}H^{-\intercal}\beta^\intercal(\beta - \widehat{\beta}H)\\
=&-n^{-1}T^{-1}\left(H^{-\intercal}\beta^\intercal\beta H^{-1}(H\bar{V} - \widehat{V})\widehat{V}^\intercal H + H^{-\intercal}\beta^\intercal\bar{U}\left(\widehat{V}^\intercal - \bar{V}^\intercal H^\intercal\right)H + H^{-\intercal}\beta^\intercal\bar{U}\bar{V}^\intercal H^\intercal H\right).
\end{aligned}
$$

(i) We need the following result, which can be shown by Lemmas 1 and 3:

$$
\begin{aligned}
&T^{-1}\left\|(H\bar{V} - \widehat{V})\widehat{V}^\intercal\right\|_{\mathrm{MAX}}\\
\leq & T^{-1}\left\|(H\bar{V} - \widehat{V})\left(\widehat{V}^\intercal - \bar{V}^\intercal H^\intercal\right)\right\|_{\mathrm{MAX}} + T^{-1}\left\|(H\bar{V} - \widehat{V})\bar{V}^\intercal H^\intercal\right\|_{\mathrm{MAX}}\\
\leq & KT^{-1}\left\|\widehat{V} - H\bar{V}\right\|_{\mathrm{F}}^2 + KT^{-1}\left\|\left(H\bar{V} - \widehat{V}\right)\bar{V}^\intercal\right\|_{\mathrm{MAX}}\|H\|_{\mathrm{MAX}}\\
= & O_p(n^{-1} + T^{-1}). \qquad \text{(IV.19)}
\end{aligned}
$$

Combined with that $\|\beta^\intercal\beta\| = O_p(n)$, we obtain

$$
n^{-1}T^{-1}\left\|H^{-\intercal}\beta^\intercal\beta H^{-1}(H\bar{V} - \widehat{V})\widehat{V}^\intercal H\right\| \leq K n^{-1}\left\|\beta^\intercal\beta\right\|T^{-1}\left\|(H\bar{V} - \widehat{V})\widehat{V}^\intercal\right\|_{\mathrm{MAX}} = O_p(n^{-1} + T^{-1}).
$$

(ii) It immediately follows from (V.34) and Lemma 2 that

$$
\left\|H^{-\intercal}\beta^\intercal\bar{U}\right\|_{\mathrm{F}} = O_p(n^{1/2}T^{1/2}).
$$

Using this and Lemma 1, we have

$$
\begin{aligned}
n^{-1}T^{-1}\left\|H^{-\intercal}\beta^\intercal\bar{U}\left(\widehat{V}^\intercal - \bar{V}^\intercal H^\intercal\right)H\right\|_{\mathrm{MAX}} \leq & n^{-1}T^{-1}\left\|H^{-1}\beta^\intercal\bar{U}\right\|_{\mathrm{F}}\left\|\widehat{V}^\intercal - \bar{V}^\intercal H^\intercal\right\|_{\mathrm{F}}\|H\|_{\mathrm{F}}\\
= & O_p(n^{-1} + n^{-1/2}T^{-1/2}).
\end{aligned}
$$

(iii) By (IV.13) and Lemma 2, we have

$$n^{-1}T^{-1}\left\|H^{-\intercal}\beta^\intercal\bar{U}\bar{V}^\intercal H^\intercal H\right\|_{\text{MAX}} \leq Kn^{-1}T^{-1}\left\|H^{-1}\right\|_{\text{MAX}}\|H\|_{\text{MAX}}^2\left\|\bar{V}\bar{U}^\intercal\beta\right\|_{\text{MAX}} = O_p(n^{-1/2}T^{-1/2}).$$

(i), (ii), and (iii) yield (b).

To show (c), note that

$$\begin{aligned}
n^{-1}\bar{u}^\intercal(\widehat{\beta}-\beta H^{-1}) =& -n^{-1}\bar{u}^\intercal(\beta-\widehat{\beta}H)H^{-1}\\
=& n^{-1}T^{-1}\left(\bar{u}^\intercal\beta H^{-1}(H\bar{V}-\widehat{V})\widehat{V}^\intercal+\bar{u}^\intercal\bar{U}\left(\widehat{V}^\intercal-\bar{V}^\intercal H^\intercal\right)+\bar{u}^\intercal\bar{U}\bar{V}^\intercal H^\intercal\right).
\end{aligned}$$

(iv) By (V.33), we have

$$\|\bar{u}^\intercal\beta\|_{\text{F}} = O_p(n^{1/2}T^{-1/2}).$$

Combined with (IV.19), we have

$$\begin{aligned}
n^{-1}T^{-1}\left\|\bar{u}^\intercal\beta H^{-1}(H\bar{V}-\widehat{V})\widehat{V}^\intercal\right\|_{\text{MAX}} \leq& n^{-1}T^{-1}\|\bar{u}^\intercal\beta\|_{\text{F}}\left\|H^{-1}\right\|_{\text{MAX}}\left\|(H\bar{V}-\widehat{V})\widehat{V}^\intercal\right\|_{\text{MAX}}\\
=& O_p\left(n^{-3/2}T^{-1/2}+n^{-1/2}T^{-3/2}\right).
\end{aligned}$$

(v) Next, note that by (V.31)

$$\begin{aligned}
\left\|\bar{u}^\intercal\bar{U}\right\|_{\text{F}} =& T^{-1}\left\|\iota_T^\intercal U^\intercal U\right\|_{\text{F}}+\|\iota_T\bar{u}^\intercal\bar{u}\|_{\text{F}} \leq KT^{-1}\|\iota_T\|\left\|U^\intercal U\right\|+\|\iota_T\|_{\text{F}}\|\bar{u}^\intercal\bar{u}\|_{\text{MAX}}\\
=& O_p(n^{1/2}T^{1/2}+n).
\end{aligned} \tag{IV.20}$$

With this, we obtain

$$n^{-1}T^{-1}\left\|\bar{u}^\intercal\bar{U}\left(\widehat{V}^\intercal-\bar{V}^\intercal H^\intercal\right)\right\|_{\text{MAX}} \leq n^{-1}T^{-1}\left\|\bar{u}^\intercal\bar{U}\right\|_{\text{F}}\left\|\widehat{V}^\intercal-\bar{V}^\intercal H^\intercal\right\|_{\text{F}} = O_p\left(n^{-1}+T^{-1}\right).$$

(vi) By (IV.18), we have

$$n^{-1}T^{-1}\left\|\bar{u}^\intercal\bar{U}\bar{V}^\intercal\right\|_{\text{MAX}} \leq n^{-1}T^{-1}\|\bar{u}\|_{\text{F}}\left\|\bar{U}\bar{V}^\intercal\right\|_{\text{F}} = O_p(T^{-1}).$$

(iv), (v), and (vi) yield (c).

As to (d), by Assumption A.5, Lemmas 2 and 4(b), we have

$$n^{-1}\left\|H^{-\intercal}\beta^\intercal(\beta-\widehat{\beta}H)\bar{v}\right\|_{\text{MAX}} \leq Kn^{-1}\left\|H^{-\intercal}\beta^\intercal(\beta-\widehat{\beta}H)\right\|_{\text{MAX}}\|\bar{v}\|_{\text{MAX}} = O_p(n^{-1}T^{-1/2}+T^{-3/2}).$$

For (e), we use Lemmas 2 and 3(b):

$$\left\|n^{-1}(\widehat{\beta}^\intercal-H^{-\intercal}\beta^\intercal)(\beta-\widehat{\beta}H)\right\|_{\text{MAX}} \leq n^{-1}\left\|\beta-\widehat{\beta}H\right\|_{\text{F}}^2\left\|H^{-1}\right\| = O_p(n^{-1}+T^{-1}),$$

which concludes the proof. □

Lemma 5. *Under Assumptions A.1, A.2, A.4, A.5, A.6, A.7, A.8, A.9, A.10, and* $\widehat{p} = p$, *we have*

$$(a) \quad \left\| \eta H^{-1} \left(H\bar{V} - \widehat{V} \right) \widehat{V}^{\intercal} (\widehat{V}\widehat{V}^{\intercal})^{-1} \right\|_{\text{MAX}} = O_p(n^{-1} + T^{-1}).$$
$$(b) \quad T^{-1} \left\| \left(H\bar{V} - \widehat{V} \right) \bar{Z}^{\intercal} \right\|_{\text{MAX}} = O_p(n^{-1} + T^{-1}).$$
$$(c) \quad \left\| \bar{Z}\bar{V}^{\intercal} H^{\intercal} (\widehat{V}\widehat{V}^{\intercal})^{-1} \right\|_{\text{MAX}} = O_p(T^{-1/2}).$$

Proof. To show (a), by (IV.19) and the fact that $T^{-1}\widehat{V}\widehat{V}^{\intercal} = \mathbb{I}_p$, we have

$$\begin{aligned}\left\| \eta H^{-1} \left(H\bar{V} - \widehat{V} \right) \widehat{V}^{\intercal} (\widehat{V}\widehat{V}^{\intercal})^{-1} \right\|_{\text{MAX}} \leq & \|\eta\|_{\text{MAX}} \|H\|_{\text{MAX}} \left\| T^{-1} \left(H\bar{V} - \widehat{V} \right) \widehat{V}^{\intercal} \right\|_{\text{MAX}} \\ = & O_p(n^{-1} + T^{-1}).\end{aligned}$$

For (b) we use the following decomposition

$$\left\| \left(H\bar{V} - \widehat{V} \right) \bar{Z}^{\intercal} \right\|_{\text{MAX}} \leq K n^{-1} T^{-1} \left\| \widehat{V}\bar{U}^{\intercal}\beta\bar{V}\bar{Z}^{\intercal} + \widehat{V}\bar{V}^{\intercal}\beta^{\intercal}\bar{U}\bar{Z}^{\intercal} + \widehat{V}\bar{U}^{\intercal}\bar{U}\bar{Z}^{\intercal} \right\|_{\text{MAX}}.$$

(i) By Assumption A.8 and (IV.14), we have

$$\left\| \widehat{V}\bar{U}^{\intercal}\beta\bar{V}\bar{Z}^{\intercal} \right\|_{\text{MAX}} \leq K \left\| \widehat{V}\bar{U}^{\intercal}\beta \right\|_{\text{MAX}} (\|VZ^{\intercal}\|_{\text{MAX}} + T\|\bar{v}\|_{\text{MAX}} \|\bar{z}\|_{\text{MAX}}) = O_p(T^{3/2} + n^{1/2}T).$$

(ii) Next, since $\|\bar{z}\|_{\text{MAX}} = O_p(T^{-1/2})$, it follows from (V.33) that

$$\|\beta^{\intercal}\bar{u}\bar{z}^{\intercal}\|_{\text{MAX}} \leq \|\beta^{\intercal}\bar{u}\|_{\text{MAX}} \|\bar{z}^{\intercal}\|_{\text{MAX}} = O_p(n^{1/2}T^{-1}).$$

By Assumption A.10(ii), we have

$$\mathrm{E}\,\|\beta^{\intercal}UZ^{\intercal}\|_{\text{MAX}}^2 \leq K \sum_{l=1}^{d}\sum_{j=1}^{p} \mathrm{E}\left(\sum_{s=1}^{T}\sum_{k=1}^{n} z_{ls}u_{ks}\beta_{kj} \right)^2 = O_p(nT),$$

and hence

$$\|\beta^{\intercal}\bar{U}\bar{Z}^{\intercal}\|_{\text{MAX}} \leq \|\beta^{\intercal}UZ^{\intercal}\|_{\text{MAX}} + T\|\beta^{\intercal}\bar{u}\bar{z}^{\intercal}\|_{\text{MAX}} = O_p(n^{1/2}T^{1/2}).$$

Therefore, we obtain

$$\left\| \widehat{V}\bar{V}^{\intercal}\beta^{\intercal}\bar{U}\bar{Z}^{\intercal} \right\|_{\text{MAX}} \leq K \left\| \widehat{V}\bar{V}^{\intercal} \right\|_{\text{MAX}} \|\beta^{\intercal}\bar{U}\bar{Z}^{\intercal}\|_{\text{MAX}} = O_p(n^{1/2}T^{3/2}).$$

(iii) Finally, we note that

$$\left\| \widehat{V}\bar{U}^{\intercal}\bar{U}\bar{Z}^{\intercal} \right\|_{\text{MAX}} \leq \left\| (\widehat{V} - H\bar{V})\bar{U}^{\intercal}\bar{U}\bar{Z}^{\intercal} \right\|_{\text{MAX}} + \|H\bar{V}\bar{U}^{\intercal}\bar{U}\bar{Z}^{\intercal}\|_{\text{MAX}}.$$

On the one hand, using the same argument as in the proof of (IV.17) and $\left\|\bar{Z}\right\| = O_p(T^{1/2})$, we obtain

$$\left\|(\widehat{V} - H\bar{V})\bar{U}^\intercal\bar{U}\bar{Z}^\intercal\right\|_{\text{MAX}} = O_p(T^2 + nT). \tag{IV.21}$$

On the other hand, we have

$$\left\|\bar{U}\bar{Z}^\intercal\right\|_{\text{F}} \leq \left\|UZ^\intercal\right\|_{\text{F}} + T\left\|\bar{u}\bar{z}^\intercal\right\|_{\text{F}}.$$

By Assumption A.10(i):

$$\text{E}\left\|UZ^\intercal\right\|_{\text{F}}^2 \leq K\sum_{i=1}^{d}\sum_{j=1}^{n}\text{E}\left(\sum_{t=1}^{T} z_{it}u_{jt}\right)^2 = O_p(nT).$$

Also, by Assumption A.8 and Equation (V.27),

$$\left\|\bar{z}\right\|_{\text{F}} \leq p^{1/2}\left\|\bar{z}\right\|_{\text{MAX}} = O_p(T^{-1/2}), \quad \left\|\bar{u}\right\|_{\text{F}} = O_p(n^{1/2}T^{-1/2}),$$

so that $T\left\|\bar{u}\bar{z}^\intercal\right\|_{\text{F}} = O_p(n^{1/2})$, hence we obtain

$$\left\|\bar{U}\bar{Z}^\intercal\right\|_{\text{F}} = O_p(n^{1/2}T^{1/2}). \tag{IV.22}$$

Combined with (IV.18), we have

$$\left\|H\bar{V}\bar{U}^\intercal\bar{U}\bar{Z}^\intercal\right\|_{\text{MAX}} \leq K\left\|H\right\|_{\text{MAX}}\left\|\bar{U}\bar{V}^\intercal\right\|_{\text{F}}\left\|\bar{U}\bar{Z}^\intercal\right\|_{\text{F}} = O_p(nT).$$

Therefore, we obtain

$$\left\|\widehat{V}\bar{U}^\intercal\bar{U}\bar{Z}^\intercal\right\|_{\text{MAX}} = O_p\left(T^2 + nT\right).$$

Combining (i), (ii), and (iii), we have

$$T^{-1}\left\|\left(H\bar{V} - \widehat{V}\right)\bar{Z}^\intercal\right\|_{\text{MAX}} = O_p(n^{-1} + T^{-1}).$$

For (c), by Lemma 2, Assumptions A.5 and A.8,

$$\left\|\bar{Z}\bar{V}^\intercal H^\intercal(\widehat{V}\widehat{V}^\intercal)^{-1}\right\|_{\text{MAX}} = \left\|T^{-1}\bar{Z}\bar{V}^\intercal H^\intercal\right\|_{\text{MAX}} \leq \left\|T^{-1}ZV^\intercal H^\intercal\right\|_{\text{MAX}} + \left\|\bar{z}\bar{v}^\intercal H^\intercal\right\|_{\text{MAX}} = O_p(T^{-1/2}).$$

□

Lemma 6. *Under Assumptions A.2, A.4, A.5, A.6, A.7, A.8, A.9, I.1, I.2, I.3, and* $\widehat{p} = p$, *we have*

$$(a) \quad n^{-1}\left\|\iota_n^\intercal\bar{u}\right\|_{\text{MAX}} = O_p\left(n^{-1/2}T^{-1/2}\right).$$

$$(b) \quad n^{-1}\left\|\iota_n^\intercal(\beta - \widehat{\beta}H)\right\|_{\text{MAX}} = O_p\left(n^{-1} + T^{-1}\right).$$

$$(c) \quad \left\| n^{-1}\iota_n^\intercal(\beta - \widehat{\beta}H)\bar{v} \right\|_{\text{MAX}} = O_p(n^{-1}T^{-1/2} + T^{-3/2}).$$

Proof. To show (a), we note that by Assumption I.3(ii),

$$\mathrm{E}\,\|\iota_n^\intercal U \iota_T\|^2 \leq \sum_{s=1}^{T}\sum_{s'=1}^{T}\sum_{i=1}^{n}\sum_{i'=1}^{n} |\sigma_{ii',ss'}| \leq KnT, \tag{IV.23}$$

so that

$$n^{-1}\,\|\iota_n^\intercal \bar{u}\|_{\text{MAX}} = n^{-1}T^{-1}\,\|\iota_n^\intercal U \iota_T\|_{\text{MAX}} = O_p\left(n^{-1/2}T^{-1/2}\right).$$

To show (b), we start from the following decomposition:

$$n^{-1}\iota_n^\intercal(\beta - \widehat{\beta}H) = -\,n^{-1}T^{-1}\left(\iota_n^\intercal \beta H^{-1}(H\bar{V} - \widehat{V})\widehat{V}^\intercal H + \iota_n^\intercal \bar{U}\left(\widehat{V}^\intercal - \bar{V}^\intercal H^\intercal\right)H + \iota_n^\intercal \bar{U}\bar{V}^\intercal H^\intercal H\right).$$

(i) By Assumption I.2, $n^{-1}\,\|\iota_n^\intercal \beta\|_\infty = O_p(1)$. By Lemmas 1, 2, and 3, we have

$$\begin{aligned}
& n^{-1}T^{-1}\left\|\iota_n^\intercal \beta H^{-1}(H\bar{V} - \widehat{V})\widehat{V}^\intercal H\right\|_{\text{MAX}} \\
\leq & n^{-1}T^{-1}\left\|\iota_n^\intercal \beta H^{-1}(H\bar{V} - \widehat{V})\left(\widehat{V}^\intercal - \bar{V}^\intercal H^\intercal\right)H\right\|_{\text{MAX}} + n^{-1}T^{-1}\left\|\iota_n^\intercal \beta H^{-1}(H\bar{V} - \widehat{V})\bar{V}^\intercal H^\intercal H\right\|_{\text{MAX}} \\
\leq & Kn^{-1}T^{-1}\,\|\iota_n^\intercal \beta\|_\infty \left\|H^{-1}\right\|_{\text{MAX}} \|H\|_{\text{MAX}} \left\|\widehat{V} - H\bar{V}\right\|_{\text{F}}^2 \\
& + Kn^{-1}T^{-1}\,\|\iota_n^\intercal \beta\|_\infty \left\|H^{-1}\right\|_{\text{MAX}} \left\|\left(H\bar{V} - \widehat{V}\right)\bar{V}^\intercal\right\|_{\text{MAX}} \|H\|_{\text{MAX}}^2 \\
= & O_p(n^{-1} + T^{-1}).
\end{aligned}$$

(ii) We note that

$$\left\|\iota_n^\intercal \bar{U}\right\|_{\text{F}} \leq \|\iota_n^\intercal U\|_{\text{F}} + \left\|\iota_n^\intercal \bar{u}\iota_T^\intercal\right\|_{\text{F}},$$

and that by Assumption I.3(i),

$$\mathrm{E}\,\|\iota_n^\intercal U\|_{\text{F}}^2 = \mathrm{E}\sum_{t=1}^{T}\left(\sum_{i=1}^{n} u_{it}\right)^2 \leq K\sum_{t=1}^{T}\sum_{i=1}^{n}\sum_{i'=1}^{n} |\sigma_{ii',t}| \leq KnT,$$

and that by Assumption I.3(ii) again,

$$\mathrm{E}\left\|\iota_n^\intercal \bar{u}\iota_T^\intercal\right\|_{\text{F}}^2 \leq KT^{-1}\sum_{i=1}^{n}\sum_{i'=1}^{n}\sum_{t=1}^{T}\sum_{t'=1}^{T} |\sigma_{ii',tt'}| \leq Kn,$$

so that by Lemma 1 again, we obtain

$$\begin{aligned}
n^{-1}T^{-1}\left\|\iota_n^\intercal \bar{U}\left(\widehat{V}^\intercal - \bar{V}^\intercal H^\intercal\right)H\right\|_{\text{MAX}} &\leq n^{-1}T^{-1}\left\|\iota_n^\intercal \bar{U}\right\|_{\text{F}}\left\|\widehat{V}^\intercal - \bar{V}^\intercal H^\intercal\right\|_{\text{F}}\|H\|_{\text{F}} \\
&= O_p(n^{-1} + n^{-1/2}T^{-1/2}).
\end{aligned}$$

(iii) By Assumption I.3(iii), we have

$$\|\iota_n^\intercal U V^\intercal\|_{\text{MAX}} = O_p\left(n^{1/2}T^{1/2}\right). \tag{IV.24}$$

Moreover, it follows from (IV.23) that

$$\|\iota_n^\intercal U \iota_T \bar{v}^\intercal\|_{\text{MAX}} \leq \|\iota_n^\intercal U \iota_T\| \, \|\bar{v}\|_{\text{MAX}} = O_p(n^{1/2}), \tag{IV.25}$$

and hence that

$$\left\|\iota_n^\intercal \bar{U}\bar{V}^\intercal\right\|_{\text{MAX}} = \left\|\iota_n^\intercal U\bar{V}^\intercal\right\|_{\text{MAX}} \leq \|\iota_n^\intercal U V^\intercal\|_{\text{MAX}} + \|\iota_n^\intercal U \iota_T \bar{v}^\intercal\|_{\text{MAX}} = O_p(n^{1/2}T^{1/2}).$$

Therefore, we have

$$n^{-1}T^{-1}\left\|\iota_n^\intercal \bar{U}\bar{V}^\intercal H^\intercal H\right\|_{\text{MAX}} = O_p(n^{-1/2}T^{-1/2}).$$

Combining (i), (ii), and (iii) leads to (b).

To show (c), by Assumption A.5 and Lemma 6(b), we have

$$n^{-1}\left\|\iota_n^\intercal(\beta - \widehat{\beta}H)\bar{v}\right\|_{\text{MAX}} \leq Kn^{-1}\left\|\iota_n^\intercal(\beta - \widehat{\beta}H)\right\|_{\text{MAX}} \|\bar{v}\|_{\text{MAX}} = O_p(n^{-1}T^{-1/2} + T^{-3/2}),$$

which concludes the proof. □

Lemma 7. *Under Assumptions A.2, A.4, A.5, A.6, A.7, A.8, A.9, I.1, and $\widehat{p} = p$, we have*

$$n^{-1}\left\|(\widehat{\beta} - \beta H^{-1})^\intercal \alpha\right\|_{\text{MAX}} = O_p(n^{-1} + n^{-1/2}T^{-1/2}).$$

Proof. Note that

$$\begin{aligned} n^{-1}\alpha^\intercal(\widehat{\beta} - \beta H^{-1}) =& - n^{-1}\alpha^\intercal(\beta - \widehat{\beta}H)H^{-1} \\ =& n^{-1}T^{-1}\left(\alpha^\intercal\beta H^{-1}(H\bar{V} - \widehat{V})\widehat{V}^\intercal + \alpha^\intercal\bar{U}\left(\widehat{V}^\intercal - \bar{V}^\intercal H^\intercal\right) + \alpha^\intercal\bar{U}\bar{V}^\intercal H^\intercal\right). \end{aligned}$$

Since α is i.i.d., and α and β are independent, we have

$$\mathrm{E}\,\|\alpha^\intercal\beta\|_{\mathrm{F}}^2 = \mathrm{E}\sum_{k=1}^{p}\left(\sum_{j=1}^{n}\alpha_j\beta_{jk}\right)^2 \leq K\mathrm{E}\,\|\beta\|_{\mathrm{F}}^2 \leq Kn,$$

and by (IV.19), we obtain

$$\begin{aligned} n^{-1}T^{-1}\left\|\alpha^\intercal\beta H^{-1}(H\bar{V} - \widehat{V})\widehat{V}^\intercal\right\|_{\text{MAX}} \leq& Kn^{-1}T^{-1}\|\alpha^\intercal\beta\|_{\mathrm{F}}\left\|H^{-1}\right\|_{\text{MAX}}\left\|(H\bar{V} - \widehat{V})\widehat{V}^\intercal\right\|_{\text{MAX}} \\ =& O_p(n^{-3/2} + n^{-1/2}T^{-1}). \end{aligned}$$

Moreover, by Assumption I.1, we have

$$\begin{aligned}\mathrm{E}\left\|\alpha^{\intercal}U\right\|_{\mathrm{F}}^{2} =&\mathrm{E}\sum_{t=1}^{T}\left(\sum_{k=1}^{n}\alpha_{k}u_{kt}\right)^{2}\leq\sum_{t=1}^{T}\sum_{k=1}^{n}\mathrm{E}\alpha_{k}^{2}\mathrm{E}u_{kt}^{2}\leq KnT,\\ \mathrm{E}\left\|\alpha\right\|_{\mathrm{F}}^{2} \leq&\mathrm{E}\sum_{k=1}^{n}\alpha_{k}^{2}\leq Kn.\end{aligned}$$

Therefore, we obtain

$$\left\|\alpha^{\intercal}\bar{u}\iota_{T}^{\intercal}\right\|_{\mathrm{F}}\leq\left\|\alpha\right\|_{\mathrm{F}}\left\|\bar{u}\right\|_{\mathrm{F}}\left\|\iota_{T}\right\|_{\mathrm{F}}=O_{p}(n).$$

These imply that

$$\begin{aligned}n^{-1}T^{-1}\left\|\alpha^{\intercal}\bar{U}\left(\widehat{V}^{\intercal}-\bar{V}^{\intercal}H^{\intercal}\right)\right\|_{\mathrm{MAX}} \leq& n^{-1}T^{-1}\left(\left\|\alpha^{\intercal}U\right\|_{\mathrm{F}}+\left\|\alpha^{\intercal}\bar{u}\iota_{T}^{\intercal}\right\|_{\mathrm{F}}\right)\left\|\widehat{V}^{\intercal}-\bar{V}^{\intercal}H^{\intercal}\right\|_{\mathrm{F}}\\ =&O_{p}(n^{-1}+n^{-1/2}T^{-1/2}).\end{aligned}$$

Finally, by Assumption A.9(i), we have

$$\mathrm{E}\left\|\alpha^{\intercal}UV^{\intercal}\right\|_{\mathrm{F}}^{2}=\sum_{j=1}^{n}\mathrm{E}(\alpha_{j}^{2})\sum_{i=1}^{p}\mathrm{E}\left(\sum_{t=1}^{T}u_{jt}v_{it}\right)^{2}\leq KnT.$$

Using the fact that

$$T\left\|\alpha^{\intercal}\bar{u}\bar{v}^{\intercal}\right\|_{\mathrm{F}}\leq T\left\|\alpha\right\|_{\mathrm{F}}\left\|\bar{u}\right\|_{\mathrm{F}}\left\|\bar{v}\right\|_{\mathrm{F}}=O_{p}(n),$$

we obtain

$$n^{-1}T^{-1}\left\|\alpha^{\intercal}\bar{U}\bar{V}^{\intercal}H^{\intercal}\right\|_{\mathrm{MAX}}\leq n^{-1}T^{-1}\left(\left\|\alpha^{\intercal}UV^{\intercal}\right\|_{\mathrm{F}}+T\left\|\alpha^{\intercal}\bar{u}\bar{v}^{\intercal}\right\|_{\mathrm{F}}\right)\left\|H\right\|_{\mathrm{MAX}}=O_{p}(n^{-1/2}T^{-1/2}).$$

□

Lemma 8. *Suppose that v_t satisfies the exponential-type tail condition. Under Assumptions A.2, A.4, A.5, A.6, A.7, A.8, A.9, I.1, I.4, I.5, and $\widehat{p}=p$, we have*

$$\left\|\widehat{V}-H\bar{V}\right\|_{\mathrm{MAX}}=O_{p}\left(n^{-1/2}T^{1/4}+T^{-1/2}\right).$$

Proof. By (IV.6)

$$\widehat{V}-H\bar{V}=n^{-1}T^{-1}\widehat{\Lambda}^{-1}\widehat{V}\left(\bar{U}^{\intercal}\beta\bar{V}+\bar{V}^{\intercal}\beta^{\intercal}\bar{U}+\bar{U}^{\intercal}\bar{U}\right).$$

We bound each term on the right-hand side. First, note that by the exponential-tail condition,

$\left\|\bar{V}\right\|_{\text{MAX}} = O_p((\log T)^{1/a})$. Along with (IV.14), we obtain

$$\left\|\widehat{V}\bar{U}^\intercal\beta\bar{V}\right\|_{\text{MAX}} \le K\left\|\widehat{V}\bar{U}^\intercal\beta\right\|_{\text{MAX}}\left\|\bar{V}\right\|_{\text{MAX}} = O_p\left(T(\log T)^{1/a} + n^{1/2}T^{1/2}(\log T)^{1/a}\right).$$

Next, by Assumption I.5 and Bonferroni and Markov inequalities,

$$\mathrm{P}(\max_{t\le T}\|\beta^\intercal u_t\| > x) \le T\max_{t\le T}\mathrm{P}(\|\beta^\intercal u_t\| > x) \le Tx^{-4}\max_{t\le T}\mathrm{E}\,\|\beta^\intercal u_t\|^4 \le KTx^{-4}n^2,$$

which implies that

$$\max_{t\le T}\|\beta^\intercal u_t\| = O_p(T^{1/4}n^{1/2}).$$

Then by (V.33) and (IV.16), we have $\left\|\beta^\intercal\bar{u}\iota_T^\intercal\right\|_{\text{MAX}} = \|\beta^\intercal\bar{u}\|_{\text{MAX}} = O_p(n^{1/2}T^{-1/2})$, and hence

$$\left\|\widehat{V}\bar{V}^\intercal\beta^\intercal\bar{U}\right\|_{\text{MAX}} \le K\left\|\widehat{V}\bar{V}^\intercal\right\|_{\text{MAX}}\left(\max_{t\le T}\|\beta^\intercal u_t\| + \left\|\beta^\intercal\bar{u}\iota_T^\intercal\right\|_{\text{MAX}}\right) = O_p\left(n^{1/2}T^{5/4}\right).$$

Finally, by Cauchy-Schwartz inequality,

$$\mathrm{E}\,\|U^\intercal u_t - \mathrm{E}(U^\intercal u_t)\|_{\text{F}}^4 = \mathrm{E}\left(\sum_{s=1}^T (u_s^\intercal u_t - \mathrm{E}(u_s^\intercal u_t))^2\right)^2 \le T\sum_{s=1}^T \mathrm{E}(u_s^\intercal u_t - \mathrm{E}(u_s^\intercal u_t))^4 \le Kn^2T^2,$$

therefore, by Bonferroni and Markov inequalities again, we have

$$\mathrm{P}\left(\max_{t\le T}\|U^\intercal u_t - \mathrm{E}(U^\intercal u_t)\|_{\text{F}} > x\right) \le T\max_{t\le T}\mathrm{P}\left(\|U^\intercal u_t - \mathrm{E}(U^\intercal u_t)\|_{\text{F}} > x\right) \le KT^3x^{-4}n^2,$$

which implies that

$$\max_{t\le T}\|U^\intercal u_t - \mathrm{E}(U^\intercal u_t)\|_{\text{F}} = O_p(n^{1/2}T^{3/4}).$$

Also, by Assumption I.4 and the fact that $|\rho_{n,tt'}| \le 1$, we obtain

$$\max_{t\le T}\|\mathrm{E}(U^\intercal u_t)\|_{\text{F}} = \max_{t\le T}\left(\sum_{t'=1}^T (\mathrm{E}(u_t^\intercal u_{t'}))^2\right)^{1/2} = n\max_{t\le T}\left(\sum_{s=1}^T \gamma_{n,tt'}^2\right)^{1/2} \le Kn.$$

Since by (V.27), (IV.20), and (V.49), we have

$$\begin{aligned}\max_{t\le T}\left\|\bar{U}^\intercal\bar{u}_t\right\|_{\text{F}} \le& \left\|\bar{U}^\intercal\bar{u}\right\|_{\text{F}} + \max_{t\le T}\|\iota_T\|_{\text{F}}\left\|\bar{u}^\intercal\right\|_{\text{F}}\|u_t\|_{\text{F}} + \max_{t\le T}\|U^\intercal u_t - \mathrm{E}(U^\intercal u_t)\|_{\text{F}} + \max_{t\le T}\|\mathrm{E}(U^\intercal u_t)\|_{\text{F}}\\ =& O_p(n^{1/2}T^{3/4} + n),\end{aligned}$$

it follows that

$$\left\|\widehat{V}\bar{U}^{\intercal}\bar{U}\right\|_{\text{MAX}} \leq K \max_{t\leq T}\left\|\widehat{V}\bar{U}^{\intercal}\bar{u}_t\right\| \leq K\left\|\widehat{V}\right\|_{\text{F}} \max_{t\leq T}\left\|\bar{U}^{\intercal}\bar{u}_t\right\|_{\text{F}} = O_p(n^{1/2}T^{5/4}+nT^{1/2}).$$

This concludes the proof.

□

V Proofs of Theorems in I

Proof of Theorem I.1. We take two steps to prove it.
Step 1: Since

$$\bar{R}^{\intercal}\bar{R} - \bar{V}^{\intercal}\beta^{\intercal}\beta\bar{V} = \bar{U}^{\intercal}\beta\bar{V} + \bar{V}^{\intercal}\beta^{\intercal}\bar{U} + \bar{U}^{\intercal}\bar{U},$$

then by Weyl's inequality, we have, for $1 \leq j \leq p$,

$$\left|\lambda_j(\bar{R}^{\intercal}\bar{R}) - \lambda_j(\bar{V}^{\intercal}\beta^{\intercal}\beta\bar{V})\right| \leq \left\|\bar{U}^{\intercal}\bar{U}\right\| + \left\|\bar{U}^{\intercal}\beta\bar{V}\right\| + \left\|\bar{V}^{\intercal}\beta^{\intercal}\bar{U}\right\|.$$

We analyze the terms on the right-hand side one by one.
(i) To begin with, write $\Gamma^u = (\gamma_{n,tt'})$. Note that

$$\left\|\bar{U}^{\intercal}\bar{U} - n\Gamma^u\right\| \leq \left\|U^{\intercal}U - n\Gamma^u\right\|_{\text{F}} + 2\left\|\iota_T\bar{u}^{\intercal}U\right\|_{\text{F}} + \left\|\iota_T\bar{u}^{\intercal}\bar{u}\iota_T^{\intercal}\right\|_{\text{F}}.$$

By Assumption A.4(ii),

$$\text{E}\left\|U^{\intercal}U - n\Gamma^u\right\|_{\text{F}}^2 = \sum_{s=1}^{T}\sum_{t=1}^{T}\text{E}\left(\sum_{j=1}^{n}(u_{js}u_{jt} - \text{E}(u_{js}u_{jt}))\right)^2 \leq KnT^2, \tag{V.26}$$

and by Assumption A.4(i),

$$\text{E}\left\|\bar{u}\right\|_{\text{F}}^2 = T^{-2}\text{E}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{t'=1}^{T}u_{it}u_{it'} \leq nT^{-2}\sum_{t=1}^{T}\sum_{t'=1}^{T}\left|\gamma_{n,tt'}\right| \leq KnT^{-1}, \tag{V.27}$$

$$\text{E}\left\|U\right\|_{\text{F}}^2 = \sum_{i=1}^{n}\sum_{t=1}^{T}\text{E}u_{it}^2 \leq n\sum_{t=1}^{T}\gamma_{n,tt} \leq KnT, \tag{V.28}$$

it follows that

$$\left\|\iota_T\bar{u}^{\intercal}U\right\|_{\text{F}} \leq \left\|\iota_T\right\|_{\text{F}}\left\|\bar{u}^{\intercal}\right\|_{\text{F}}\left\|U\right\|_{\text{F}} = O_p(nT^{1/2}), \quad \left\|\iota_T\bar{u}^{\intercal}\bar{u}\iota_T^{\intercal}\right\|_{\text{F}} \leq \left\|\iota_T\right\|_{\text{F}}^2\left\|\bar{u}^{\intercal}\right\|_{\text{F}}^2 = O_p(n),$$

and hence that

$$\left\|\bar{U}^{\intercal}\bar{U} - n\Gamma^u\right\| = O_p(n^{1/2}T) + O_p(nT^{1/2}). \tag{V.29}$$

Next, writing $\rho_{n,st} = \gamma_{n,st}/\sqrt{\gamma_{n,ss}\gamma_{n,tt}}$, by Assumption A.4(i) and the fact that $|\rho_{n,st}| \leq 1$,

$$\begin{aligned}\|\Gamma^u\|_{\mathrm{F}}^2 =& \sum_{s=1}^{T}\sum_{t=1}^{T}\gamma_{n,st}^2 = \sum_{s=1}^{T}\sum_{t=1}^{T}\gamma_{n,ss}\gamma_{n,tt}\rho_{n,st}^2 \\ \leq& K\sum_{s=1}^{T}\sum_{t=1}^{T}|\gamma_{n,ss}\gamma_{n,tt}|^{1/2}|\rho_{n,st}| \leq K\sum_{s=1}^{T}\sum_{t=1}^{T}|\gamma_{n,st}| \leq KT,\end{aligned} \tag{V.30}$$

so we have $n\|\Gamma^u\| = O_p(nT^{1/2})$. Therefore, we obtain

$$\left\|\bar{U}^\intercal\bar{U}\right\| \leq \left\|\bar{U}^\intercal\bar{U} - n\Gamma^u\right\| + n\|\Gamma^u\| = O_p(nT^{1/2}) + O_p(n^{1/2}T). \tag{V.31}$$

(ii) By Assumption A.7, we have

$$\mathrm{E}\|U^\intercal\beta\|_{\mathrm{F}}^2 = \mathrm{E}\sum_{j=1}^{p}\sum_{t=1}^{T}\left(\sum_{i=1}^{n}\beta_{ij}u_{it}\right)^2 \leq KnT, \tag{V.32}$$

$$\mathrm{E}\|\bar{u}^\intercal\beta\|_{\mathrm{F}}^2 = \mathrm{E}\sum_{k=1}^{p}\left(\sum_{i=1}^{n}\bar{u}_i\beta_{ik}\right)^2 \leq KnT^{-1}, \tag{V.33}$$

it follows that

$$\left\|\bar{U}^\intercal\beta\right\|_{\mathrm{F}} \leq \|U^\intercal\beta\|_{\mathrm{F}} + \|\iota_T\|_{\mathrm{F}}\|\bar{u}^\intercal\beta\|_{\mathrm{F}} = O_p(n^{1/2}T^{1/2}). \tag{V.34}$$

Also, by Assumption A.5,

$$T^{-1}\left\|\bar{V}\bar{V}^\intercal\right\|_{\mathrm{MAX}} \leq \left\|T^{-1}VV^\intercal - \Sigma^v\right\|_{\mathrm{MAX}} + \|\Sigma^v\|_{\mathrm{MAX}} + \|\bar{v}\bar{v}^\intercal\|_{\mathrm{MAX}} \leq K, \tag{V.35}$$

we have

$$\left\|\bar{V}\right\| \leq \left\|\bar{V}\bar{V}^\intercal\right\|^{1/2} \leq K\left\|\bar{V}\bar{V}^\intercal\right\|_{\mathrm{MAX}}^{1/2} = O_p(T^{1/2}). \tag{V.36}$$

Therefore, we have

$$\left\|\bar{V}^\intercal\beta^\intercal\bar{U}\right\| = \left\|\bar{U}^\intercal\beta\bar{V}\right\| \leq \left\|\bar{U}^\intercal\beta\right\|_{\mathrm{F}}\left\|\bar{V}\right\| = O_p(n^{1/2}T).$$

Combining (i) and (ii), we have for $1 \leq j \leq p$,

$$n^{-1}T^{-1}\left|\lambda_j(\bar{R}^\intercal\bar{R}) - \lambda_j(\bar{V}^\intercal\beta^\intercal\beta\bar{V})\right| = O_p(n^{-1/2} + T^{-1/2}) = o_p(1). \tag{V.37}$$

(iii) Moreover, by Assumption A.6, (V.36), and Weyl's inequality again,

$$\left|n^{-1}T^{-1}\lambda_j(\bar{V}^\intercal\beta^\intercal\beta\bar{V}) - T^{-1}\lambda_j(\bar{V}^\intercal\Sigma^\beta\bar{V})\right| \leq \left\|n^{-1}\beta^\intercal\beta - \Sigma^\beta\right\|T^{-1}\left\|\bar{V}^\intercal\right\|\left\|\bar{V}\right\| = o_p(1),$$

and combined with Assumption A.5, and the fact that $\|\bar{v}\| \leq K \|\bar{v}\|_{\mathrm{MAX}} = O_p(T^{-1/2})$,

$$\left| T^{-1}\lambda_j(\bar{V}^\intercal \Sigma^\beta \bar{V}) - \lambda_j\left(\left(\Sigma^\beta\right)^{1/2} \Sigma^v \left(\Sigma^\beta\right)^{1/2}\right)\right| \\ \leq \left\|T^{-1}\bar{V}\bar{V}^\intercal - \Sigma^v\right\| \left\|\Sigma^\beta\right\| \leq \left(\left\|T^{-1}VV^\intercal - \Sigma^v\right\| + \|\bar{v}\bar{v}^\intercal\|\right) \left\|\Sigma^\beta\right\| = o_p(1),$$

where we also use the fact that the non-zero eigenvalues of $\bar{V}^\intercal \Sigma^\beta \bar{V}$ are identical to the non-zero eigenvalues of $\left(\Sigma^\beta\right)^{1/2} \bar{V}\bar{V}^\intercal \left(\Sigma^\beta\right)^{1/2}$. Therefore, for $1 \leq j \leq p$,

$$\left| n^{-1}T^{-1}\lambda_j(\bar{R}^\intercal \bar{R}) - \lambda_j\left(\left(\Sigma^\beta\right)^{1/2} \Sigma^v \left(\Sigma^\beta\right)^{1/2}\right)\right| = o_p(1). \tag{V.38}$$

Step 2: By Assumptions A.5 and A.6, there exists $0 < K_1, K_2 < \infty$, such that

$$K_1 < \lambda_{\min}(\Sigma^v)\lambda_{\min}(\Sigma^\beta) \leq \lambda_{\min}(\Sigma^v\Sigma^\beta) \leq \lambda_{\max}(\Sigma^v\Sigma^\beta) \leq \lambda_{\max}(\Sigma^v)\lambda_{\max}(\Sigma^\beta) < K_2.$$

Therefore the eigenvalues of $(\Sigma^\beta)^{1/2}\Sigma^v(\Sigma^\beta)^{1/2}$ are bounded away from 0 and ∞, we have by (V.38), for $1 \leq j \leq p$,

$$K_1 < n^{-1}T^{-1}\lambda_j(\bar{R}^\intercal \bar{R}) < K_2. \tag{V.39}$$

On the other hand, we can write

$$\bar{R}\bar{R}^\intercal = \tilde{\beta}\bar{V}\bar{V}^\intercal\tilde{\beta}^\intercal + \bar{U}\left(\mathbb{I}_T - \bar{V}^\intercal(\bar{V}\bar{V}^\intercal)^{-1}\bar{V}\right)\bar{U}^\intercal, \tag{V.40}$$

where $\tilde{\beta} = \beta + U\bar{V}^\intercal(\bar{V}\bar{V}^\intercal)^{-1}$. By (4.3.2a) of Theorem 4.3.1 and (4.3.14) of Corollary 4.3.12 in Horn and Johnson (2013), for $p+1 \leq j \leq n$, we have

$$\lambda_j(\bar{R}\bar{R}^\intercal) \leq \lambda_{j-p}\left(\bar{U}(\mathbb{I}_T - \bar{V}^\intercal(\bar{V}\bar{V}^\intercal)^{-1}\bar{V})\bar{U}^\intercal\right) + \lambda_{p+1}(\tilde{\beta}\bar{V}\bar{V}^\intercal\tilde{\beta}) \leq \lambda_{j-p}(\bar{U}\bar{U}^\intercal) \leq \lambda_1(\bar{U}\bar{U}^\intercal).$$

Moreover, by (V.31), we have

$$\lambda_1(\bar{U}\bar{U}^\intercal) = \left\|\bar{U}^\intercal\bar{U}\right\| = O_p(nT^{1/2}) + O_p(n^{1/2}T),$$

hence for $p+1 \leq j \leq n$, there exists some $K > 0$, such that

$$n^{-1}T^{-1}\lambda_j(\bar{R}^\intercal \bar{R}) \leq K(n^{-1/2} + T^{-1/2}). \tag{V.41}$$

Now we define, for $1 \leq j \leq n$,

$$f(j) = n^{-1}T^{-1}\lambda_j(\bar{R}^\intercal \bar{R}) + j \times \phi(n,T).$$

(V.39) and (V.41) together imply that for $1 \le j \le p$,

$$f(j) - f(p+1) = n^{-1}T^{-1}\left(\lambda_j(\bar{R}^\intercal \bar{R}) - \lambda_{p+1}(\bar{R}^\intercal \bar{R})\right) + (j-p-1)\phi(n,T)$$
$$> \lambda_j\left(\left(\Sigma^\beta\right)^{1/2}\Sigma^v\left(\Sigma^\beta\right)^{1/2}\right) + o_p(1) > K,$$

for some $K > 0$. This establishes the first statement. Moreover, for $p+1 < j \le n$, we have

$$\mathrm{P}(f(j) < f(p+1)) = \mathrm{P}\left((j-p-1)\phi(n,T) < n^{-1}T^{-1}\left(\lambda_{p+1}(\bar{R}^\intercal \bar{R}) - \lambda_j(\bar{R}^\intercal \bar{R})\right)\right) \to 0.$$

Therefore, $p+1 = \arg\min_{1\le j\le p_{\max}} f(j)$ holds with probability approaching 1, and hence $\widehat{p} \xrightarrow{p} p$. □

Proof of Theorem I.2. We denote the estimators of $\bar{V}$ and β based on $\breve{p}$ as $\breve{V}$ and $\breve{\beta}$ respectively. Consider the singular value decomposition of $n^{-1/2}T^{-1/2}\bar{R}$ by scaling (B.10), we have

$$n^{-1/2}T^{-1/2}\varsigma_{p+1:\breve{p}}^\intercal \bar{R} = \Lambda_{p+1:\breve{p}}^{1/2}\xi_{p+1:\breve{p}}^\intercal \quad \text{and} \quad n^{-1/2}T^{-1/2}\bar{R}\xi_{p+1:\breve{p}} = \varsigma_{p+1:\breve{p}}\Lambda_{p+1:\breve{p}}^{1/2}, \tag{V.42}$$

where $\Lambda_{p+1:\breve{p}}$ is a $(\breve{p}-p)\times(\breve{p}-p)$ diagonal matrix with the ith entry on the diagonal being $n^{-1}T^{-1}\lambda_i(\bar{R}^\intercal \bar{R})$, $\xi_{p+1:\breve{p}} = (\xi_{p+1} : \xi_{p+2} : \cdots : \xi_{\breve{p}})$ is $T \times (\breve{p}-p)$, and $\varsigma_{p+1:\breve{p}} = (\varsigma_{p+1} : \varsigma_{p+2} : \ldots : \varsigma_{\breve{p}})$ is $n \times (\breve{p}-p)$. It is also easy to observe that

$$\breve{V}^\intercal = \left(\widehat{V}^\intercal : T^{1/2}\xi_{p+1:\breve{p}}\right), \quad \breve{\beta} = \left(\widehat{\beta} : n^{1/2}\varsigma_{p+1:\breve{p}}\Lambda_{p+1:\breve{p}}^{1/2}\right), \quad \widehat{V}\xi_{p+1:\breve{p}} = 0, \quad \text{and} \quad \varsigma_{p+1:\breve{p}}^\intercal \widehat{\beta} = 0.$$

By direct calculation, we have

$$\breve{\gamma}_g - \widehat{\gamma}_g = T^{-1/2}n^{-1/2}\bar{G}\xi_{p+1:\breve{p}}\Lambda_{p+1:\breve{p}}^{-1/2}\varsigma_{p+1:\breve{p}}^\intercal \bar{r}.$$

First, by Lemma 3(b) and (V.27), we have

$$\left\|\varsigma_{p+1:\breve{p}}^\intercal(\beta - \widehat{\beta}H)\right\| = O_p(1 + n^{1/2}T^{-1/2}), \quad \left\|\varsigma_{p+1:\breve{p}}^\intercal \bar{u}\right\| = O_p(n^{1/2}T^{-1/2}),$$

which, in turn, leads to

$$\left\|\varsigma_{p+1:\breve{p}}^\intercal \bar{r}\right\| \le \left\|\varsigma_{p+1:\breve{p}}^\intercal(\beta - \widehat{\beta}H)(\gamma + \bar{v})\right\| + \left\|\varsigma_{p+1:\breve{p}}^\intercal \bar{u}\right\| = O_p(1 + n^{1/2}T^{-1/2}).$$

Second, by Lemmas 1 and 2,

$$\left\|\eta\bar{V}\xi_{p+1:\breve{p}}\right\| = \left\|\eta(\bar{V} - H^{-1}\widehat{V})\xi_{p+1:\breve{p}}\right\| = O_p(1 + n^{-1/2}T^{1/2}).$$

Because z_t is i.i.d., and that it is independent of $\xi_{p+1:\breve{p}}$, we also have $\left\|\bar{Z}\xi_{p+1:\breve{p}}\right\|_{\mathrm{F}} = O_p(\breve{p}^{1/2})$. This establishes that

$$\left\|\bar{G}\xi_{p+1:\breve{p}}\right\| = O_p(1 + \breve{p}^{1/2} + n^{-1/2}T^{1/2}).$$

Third, it follows from (V.40) that

$$\bar{R}\bar{R}^\intercal + \bar{U}\bar{V}^\intercal(\bar{V}\bar{V}^\intercal)^{-1}\bar{V}\bar{U}^\intercal = \bar{U}\bar{U}^\intercal + \tilde{\beta}\bar{V}\bar{V}^\intercal\tilde{\beta}^\intercal,$$

where $\tilde{\beta} = \beta + U\bar{V}^\intercal(\bar{V}\bar{V}^\intercal)^{-1}$. By (4.3.2a) and (4.3.2b) of Theorem 4.3.1 in Horn and Johnson (2013), for $p+1 \leq j \leq \check{p}$,

$$\lambda_{j+p}(\bar{U}\bar{U}^\intercal) + \lambda_{n-1}(\tilde{\beta}\bar{V}\bar{V}^\intercal\tilde{\beta}^\intercal) \leq \lambda_{j+p}(\bar{R}\bar{R}^\intercal + \bar{U}\bar{V}^\intercal(\bar{V}\bar{V}^\intercal)^{-1}\bar{V}\bar{U}^\intercal) \leq \lambda_j(\bar{R}\bar{R}^\intercal) + \lambda_{p+1}(\bar{U}\bar{V}^\intercal(\bar{V}\bar{V}^\intercal)^{-1}\bar{V}\bar{U}^\intercal).$$

Since $\operatorname{rank}(\tilde{\beta}\bar{V}\bar{V}^\intercal\tilde{\beta}^\intercal) \leq p$ and $\operatorname{rank}(\bar{U}\bar{V}^\intercal(\bar{V}\bar{V}^\intercal)^{-1}\bar{V}\bar{U}^\intercal) \leq p$, we obtain,

$$\lambda_{\check{p}+p}(\bar{U}\bar{U}^\intercal) \leq \lambda_{j+p}(\bar{U}\bar{U}^\intercal) \leq \lambda_j(\bar{R}\bar{R}^\intercal) \leq \lambda_{j-p}(\bar{U}\bar{U}^\intercal) \leq \lambda_1(\bar{U}\bar{U}^\intercal).$$

This implies that with probability approaching 1,

$$\lambda_j(\bar{R}\bar{R}^\intercal) \geq K(n \vee T), \quad p+1 \leq j \leq \check{p},$$

so that $\left\|\Lambda_{p+1:\check{p}}^{-1/2}\right\| = O_p(n^{1/2} \wedge T^{1/2})$.

Combining results of the above three steps, we therefore obtain that

$$\left\|T^{-1/2}n^{-1/2}\bar{G}\xi_{p+1:\check{p}}\Lambda_{p+1:\check{p}}^{-1/2}\varsigma_{p+1:\check{p}}^\intercal\bar{r}\right\| = o_p(1),$$

which concludes the proof. □

Proof of Theorem I.3. We summarize the parameters of interest in $\Gamma = (\gamma_0 : (\eta\gamma)^\intercal)^\intercal$, and denote

$$\widetilde{\Gamma} := (\widetilde{\gamma}_0, \widetilde{\gamma}^\intercal)^\intercal = \left((\iota_n : \widehat{\beta})^\intercal(\iota_n : \widehat{\beta})\right)^{-1}(\iota_n : \widehat{\beta})^\intercal\bar{r}, \quad \widehat{\Gamma} := \begin{pmatrix} \widehat{\gamma}_0 \\ \widehat{\gamma}_g \end{pmatrix} := \begin{pmatrix} 1 & 0 \\ 0 & \widehat{\eta} \end{pmatrix}\widetilde{\Gamma} = \begin{pmatrix} \widetilde{\gamma}_0 \\ \widehat{\eta}\widetilde{\gamma} \end{pmatrix}.$$

Because $\widehat{\beta}$ and $\widehat{\eta}$ only rely on $\bar{R}$ and $\bar{G}$, which do not depend on $\gamma_0\iota_n$ and α, we can recycle the estimates derived in Lemmas 1 – 5, despite that the DGP is given by Assumption I.1 instead of Assumption A.1.

We use the following decomposition:

$$\begin{aligned}
&\widetilde{\Gamma} - \begin{pmatrix} \gamma_0 \\ H\gamma \end{pmatrix} \\
&= \left((\iota_n : \widehat{\beta})^\intercal(\iota_n : \widehat{\beta})\right)^{-1}(\iota_n : \widehat{\beta})^\intercal\left(\left(\beta - \widehat{\beta}H\right)\gamma + \beta\bar{v} + \alpha + \bar{u}\right) \\
&= \begin{pmatrix} 0 \\ H\bar{v} \end{pmatrix} + \left\{\frac{1}{n}\begin{pmatrix} \iota_n^\intercal\iota_n & \iota_n^\intercal\widehat{\beta} \\ \widehat{\beta}^\intercal\iota_n & \widehat{\beta}^\intercal\widehat{\beta} \end{pmatrix}\right\}^{-1}\left\{\frac{1}{n}\begin{pmatrix} \iota_n^\intercal\alpha \\ H^{-\intercal}\beta^\intercal\alpha \end{pmatrix} + \frac{1}{n}\begin{pmatrix} \iota_n^\intercal\bar{u} + \iota_n^\intercal(\beta - \widehat{\beta}H)\gamma \\ H^{-\intercal}\beta^\intercal\bar{u} + H^{-\intercal}\beta^\intercal(\beta - \widehat{\beta}H)\gamma \end{pmatrix}\right. \\
&\quad \left. + \frac{1}{n}\begin{pmatrix} \iota_n^\intercal(\beta - \widehat{\beta}H)\bar{v} \\ (\widehat{\beta} - \beta H^{-1})^\intercal(\alpha + \bar{u}) + H^{-\intercal}\beta^\intercal(\beta - \widehat{\beta}H)\bar{v} + (\widehat{\beta}^\intercal - H^{-\intercal}\beta^\intercal)(\beta - \widehat{\beta}H)(\gamma + \bar{v}) \end{pmatrix}\right\}. \qquad \text{(V.43)}
\end{aligned}$$

By Lemma 6(b), we have

$$n^{-1}\left\|\iota_n^\intercal(\widehat{\beta}-\beta H^{-1})\right\|_{\text{MAX}} = O_p(n^{-1}+T^{-1}).$$

Therefore, we have

$$\frac{1}{n}\begin{pmatrix} \iota_n^\intercal\iota_n & \iota_n^\intercal\widehat{\beta} \\ \widehat{\beta}^\intercal\iota_n & \widehat{\beta}^\intercal\widehat{\beta}\end{pmatrix} = \frac{1}{n}\begin{pmatrix} \iota_n^\intercal\iota_n & \iota_n^\intercal\beta H^{-1} \\ H^{-\intercal}\beta^\intercal\iota_n & H^{-\intercal}\beta^\intercal\beta H^{-1}\end{pmatrix} + O_p(n^{-1}+T^{-1}). \tag{V.44}$$

Using this, and by Lemmas 2, 4, 5, 6, and 7, we have

$$\widehat{\Gamma} - \begin{pmatrix} \gamma_0 \\ \eta\gamma \end{pmatrix} = \begin{pmatrix} 0 \\ T^{-1}\bar{Z}\bar{V}^\intercal(\Sigma^v)^{-1}\gamma + \eta\bar{v} \end{pmatrix}$$
$$+\begin{pmatrix} 1 & 0 \\ 0 & \eta \end{pmatrix}\left\{\frac{1}{n}\begin{pmatrix} \iota_n^\intercal\iota_n & \iota_n^\intercal\beta \\ \beta^\intercal\iota_n & \beta^\intercal\beta\end{pmatrix} + o_p(1)\right\}^{-1} \times \left\{\frac{1}{n}\begin{pmatrix} \iota_n^\intercal\alpha \\ \beta^\intercal\alpha\end{pmatrix} + o_p(n^{-1/2}+T^{-1/2})\right\}.$$

Moreover, by Cramér-Wold theorem and Lyapunov's central limit theorem, we can obtain

$$n^{-1/2}\begin{pmatrix} \iota_n^\intercal\alpha \\ \beta^\intercal\alpha\end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \beta_0^\intercal \\ \beta_0 & \Sigma^\beta \end{pmatrix}(\sigma^\alpha)^2\right), \tag{V.45}$$

where we use $\|n^{-1}\beta^\intercal\iota_n - \beta_0\|_{\text{MAX}} = o(1)$ and $\|n^{-1}\beta^\intercal\beta - \Sigma^\beta\|_{\text{MAX}} = o(1)$. Also, Assumptions A.6 and I.2 ensure that $(1-\beta_0^\intercal(\Sigma^\beta)^{-1}\beta_0)$ and $(\Sigma^\beta - \beta_0\beta_0^\intercal)$ are invertible. Therefore, by the Delta method, we have

$$n^{1/2}\left(\widehat{\gamma}_0 - \gamma_0\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \left(1-\beta_0^\intercal(\Sigma^\beta)^{-1}\beta_0\right)^{-1}(\sigma^\alpha)^2\right),$$

Similarly, we have

$$n^{1/2}\begin{pmatrix} 0 & \eta \end{pmatrix}\left\{\frac{1}{n}\begin{pmatrix} \iota_n^\intercal\iota_n & \iota_n^\intercal\beta \\ \beta^\intercal\iota_n & \beta^\intercal\beta\end{pmatrix} + o_p(1)\right\}^{-1} \times \frac{1}{n}\begin{pmatrix} \iota_n^\intercal\alpha \\ \beta^\intercal\alpha\end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \Upsilon\right),$$

where

$$\Upsilon = (\sigma^\alpha)^2\eta\left(\Sigma^\beta - \beta_0\beta_0^\intercal\right)^{-1}\eta^\intercal.$$

By the same asymptotic independence argument as in the proof of Theorem 3 in Bai (2003), we establish the desired result:

$$\left(T^{-1}\Phi + n^{-1}\Upsilon\right)^{-1/2}\left(\widetilde{\gamma}_g - \eta\gamma\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \mathbb{I}_d).$$

□

Proof of Theorem I.4. By Assumptions A.5, A.6, and I.2, Lemma 4, (V.33), and (V.45), we have

$$
\begin{aligned}
&n^{-1}\iota_n^\intercal \bar{r} = \gamma_0 + \beta_0^\intercal \gamma + O_p(n^{-1/2} + T^{-1/2}),\\
&n^{-1}\bar{r}^\intercal \bar{r} = \gamma^\intercal \Sigma^\beta \gamma + \gamma_0^2 + (\sigma^\alpha)^2 + \gamma^\intercal \beta_0 \gamma_0 + \beta_0^\intercal \gamma\gamma_0 + O_p(n^{-1/2} + T^{-1/2}),
\end{aligned}
$$

it then follows that

$$
n^{-1}\bar{r}^\intercal \mathbb{M}_{\iota_n} \bar{r} = n^{-1}\bar{r}^\intercal \bar{r} - (n^{-1}\iota_n^\intercal \bar{r})^2 = \gamma^\intercal(\Sigma^\beta - \beta_0\beta_0^\intercal)\gamma + (\sigma^\alpha)^2 + o_p(1).
$$

On the other hand, by Assumption A.5, Lemma 3, (V.27), we have

$$
\begin{aligned}
n^{-1}\left\|H^\intercal \widehat{\beta}^\intercal \mathbb{M}_{\iota_n}\bar{r} - \beta^\intercal \mathbb{M}_{\iota_n}\bar{r}\right\|_{\text{MAX}} =& \left\|(H^\intercal \widehat{\beta}^\intercal - \beta^\intercal)\mathbb{M}_{\iota_n}(\alpha + \beta\gamma + \beta\bar{v} + \bar{u})\right\|_{\text{MAX}}\\
\leq& n^{-1}\left\|H^\intercal \widehat{\beta}^\intercal - \beta^\intercal\right\|_{\text{F}} \|\alpha + \beta\gamma + \beta\bar{v} + \bar{u}\|_{\text{F}} = O_p(n^{-1/2} + T^{-1/2}).
\end{aligned}
$$

Similarly, we have

$$
\begin{aligned}
&n^{-1}\beta^\intercal \mathbb{M}_{\iota_n}\bar{r} = \left(\Sigma^\beta - \beta_0\beta_0^\intercal\right)\gamma + o_p(1),\\
&n^{-1}\beta^\intercal \mathbb{M}_{\iota_n}\beta = \Sigma^\beta - \beta_0\beta_0^\intercal + o_p(1),
\end{aligned}
$$

therefore, we obtain

$$
(n^{-1}\beta^\intercal \mathbb{M}_{\iota_n}\bar{r})^\intercal \left(n^{-1}\beta^\intercal \mathbb{M}_{\iota_n}\beta\right)^{-1} \left(n^{-1}\beta^\intercal \mathbb{M}_{\iota_n}\bar{r}\right) = \gamma^\intercal \left(\Sigma^\beta - \beta_0\beta_0^\intercal\right)\gamma + o_p(1),
$$

which establishes $\widehat{\mathrm{R}}_v^2 \xrightarrow{p} \mathrm{R}_v^2$.

By Lemma 2, (IV.11), (B.6) and the fact that $\|\eta\|_{\text{MAX}} \leq K$, we have

$$
\begin{aligned}
&\left\|T^{-1}\widehat{\eta}\widehat{V}\widehat{V}^\intercal \eta^\intercal - \eta\Sigma^v\eta^\intercal\right\|_{\text{MAX}}\\
\leq& \left\|(\widehat{\eta} - \eta H^{-1})(\widehat{\eta} - \eta H^{-1})^\intercal\right\|_{\text{MAX}} + \left\|(\widehat{\eta} - \eta H^{-1})H^{-\intercal}\eta^\intercal\right\|_{\text{MAX}} + \left\|\eta H^{-1}(\widehat{\eta} - \eta H^{-1})^\intercal\right\|_{\text{MAX}}\\
&+ \left\|\eta(H^{-1}H^{-\intercal} - \Sigma^v)\eta^\intercal\right\|_{\text{MAX}}\\
=& O_p(n^{-1/2} + T^{-1/2}).
\end{aligned}
$$

Also, by Assumptions A.5, A.8, and A.11, we have

$$
T^{-1}\bar{G}\bar{G}^\intercal = T^{-1}(\eta\bar{V} + \bar{Z})(\eta\bar{V} + \bar{Z})^\intercal \xrightarrow{p} \eta\Sigma^v\eta^\intercal + \Sigma^z,
$$

hence it follows that $\widehat{\mathrm{R}}_g^2 \xrightarrow{p} \mathrm{R}_g^2$. □

Proof of Theorem I.5. For any $1 \leq t \leq T$, we have

$$
\widehat{g}_t - \eta v_t = (\widehat{\eta} - \eta H^{-1})(\widehat{v}_t - H\bar{v}_t) + (\widehat{\eta} - \eta H^{-1})H\bar{v}_t + \eta H^{-1}(\widehat{v}_t - H\bar{v}_t) - \eta\bar{v} \tag{V.46}
$$

By (IV.6), we have

$$\widehat{v}_t - H\bar{v}_t = n^{-1}T^{-1}\widehat{\Lambda}^{-1}(\widehat{V} - H\bar{V})\left(\bar{U}^\intercal \beta \bar{v}_t + \bar{U}^\intercal \bar{u}_t\right) + n^{-1}T^{-1}\widehat{\Lambda}^{-1}\left(H\bar{V}\bar{U}^\intercal \beta \bar{v}_t + H\bar{V}\bar{U}^\intercal \bar{u}_t\right) + n^{-1}T^{-1}\widehat{\Lambda}^{-1}\widehat{V}\bar{V}^\intercal \beta^\intercal \bar{u}_t. \tag{V.47}$$

By Assumption I.6, we have $\|\beta^\intercal u_t\| = O_p(n^{1/2})$, so that using (V.33),

$$\|\beta^\intercal \bar{u}_t\|_{\mathrm{F}} \leq \|\beta^\intercal u_t\|_{\mathrm{F}} + \|\beta^\intercal \bar{u}\|_{\mathrm{F}} = O_p(n^{1/2}). \tag{V.48}$$

By Assumption A.4(i), Assumptions I.4 and I.5, using the fact that $|\rho_{n,st}| \leq 1$, we have

$$\begin{aligned}
\mathrm{E}\left\|U^\intercal u_t\right\|_{\mathrm{F}}^2 =& \mathrm{E}\sum_{s=1}^{T}\left(n\gamma_{n,st} + \sum_{k=1}^{n}\left(u_{ks}u_{kt} - \mathrm{E}(u_{ks}u_{kt})\right)\right)^2 \\
\leq& Kn^2\sum_{s=1}^{T}\gamma_{n,st}^2 + KnT \leq n^2\sum_{s=1}^{T}|\gamma_{n,st}| + KnT = Kn^2 + KnT, \\
\mathrm{E}\left\|u_t\right\|_{\mathrm{F}}^2 \leq& \sum_{k=1}^{n}\mathrm{E}u_{kt}^2 \leq \sum_{k=1}^{n}|\sigma_{kk'}| \leq K.
\end{aligned} \tag{V.49}$$

Then from (V.27) and (IV.20), it follows that

$$\left\|\bar{U}^\intercal \bar{u}_t\right\|_{\mathrm{F}} \leq \left\|\bar{U}^\intercal \bar{u}\right\|_{\mathrm{F}} + \left\|U^\intercal u_t\right\|_{\mathrm{F}} + \left\|\iota_T\right\|_{\mathrm{F}}\left\|\bar{u}^\intercal\right\|_{\mathrm{F}}\left\|u_t\right\|_{\mathrm{F}} = O_p(n + n^{1/2}T^{1/2}).$$

The above estimates, along with (V.34), Lemma 1, and $\|\bar{v}_t\| = O_p(1)$, lead to

$$\begin{aligned}
&\left\|n^{-1}T^{-1}\widehat{\Lambda}^{-1}(\widehat{V} - H\bar{V})\left(\bar{U}^\intercal \beta \bar{v}_t + \bar{U}^\intercal \bar{u}_t\right)\right\|_{\mathrm{MAX}} \\
\leq& n^{-1}T^{-1}\left\|\widehat{\Lambda}^{-1}\right\|_{\mathrm{MAX}}\left\|\widehat{V} - H\bar{V}\right\|_{\mathrm{F}}\left(\left\|\bar{U}^\intercal \beta\right\|_{\mathrm{F}}\left\|\bar{v}_t\right\| + \left\|\bar{U}^\intercal \bar{u}_t\right\|_{\mathrm{F}}\right) = O_p(n^{-1} + T^{-1}).
\end{aligned}$$

Moreover, it follows from (V.27), (IV.13), and (IV.18) that

$$\begin{aligned}
&\left\|n^{-1}T^{-1}\widehat{\Lambda}^{-1}\left(H\bar{V}\bar{U}^\intercal \beta \bar{v}_t + H\bar{V}\bar{U}^\intercal \bar{u}_t\right)\right\|_{\mathrm{MAX}} \\
\leq& Kn^{-1}T^{-1}\left\|\widehat{\Lambda}^{-1}\right\|_{\mathrm{MAX}}\|H\|\left(\left\|\bar{V}\bar{U}^\intercal \beta\right\|_{\mathrm{MAX}}\left\|\bar{v}_t\right\| + \left\|\bar{V}\bar{U}^\intercal\right\|_{\mathrm{F}}\left(\|u_t\|_{\mathrm{F}} + \|u\|_{\mathrm{F}}\right)\right) \\
=& O_p(n^{-1/2}T^{-1/2} + T^{-1}).
\end{aligned}$$

We thereby focus on the remaining term, which by Lemma 1, (V.36) and (V.48), satisfies

$$n^{-1}T^{-1}\left\|\widehat{\Lambda}^{-1}\widehat{V}\bar{V}^\intercal \beta^\intercal \bar{u}_t\right\|_{\mathrm{MAX}} \leq Kn^{-1}T^{-1}\left\|\widehat{\Lambda}^{-1}\right\|_{\mathrm{MAX}}\left\|\widehat{V}\right\|_{\mathrm{F}}\left\|\bar{V}^\intercal\right\|_{\mathrm{F}}\left\|\beta^\intercal \bar{u}_t\right\|_{\mathrm{MAX}} = O_p(n^{-1/2}).$$

Therefore, we have

$$\left\|\widehat{v}_t - H\bar{v}_t\right\|_{\mathrm{MAX}} = O_p(n^{-1/2} + T^{-1}). \tag{V.50}$$

Then by (V.46), (V.47), and (B.5), we have

$$\left\|\widehat{g}_t - \eta v_t - \left(T^{-1}\bar{Z}\bar{V}^\intercal H^\intercal H v_t + n^{-1}T^{-1}\eta H^{-1}\widehat{\Lambda}^{-1}\widehat{V}\bar{V}^\intercal\beta^\intercal u_t - \eta\bar{v}\right)\right\|_{\text{MAX}} = o_p(n^{-1/2}+T^{-1/2}).$$

Next, we note that by Assumption A.11 and Lemma 2,

$$T^{1/2}\begin{pmatrix} T^{-1}\text{vec}\left(\bar{Z}\bar{V}^\intercal H^\intercal H v_t\right) \\ \eta\bar{v}\end{pmatrix} = T^{1/2}\begin{pmatrix} (v_t^\intercal H^\intercal H\otimes \mathbb{I}_d)\text{vec}(\bar{Z}\bar{V}^\intercal) \\ \eta\bar{v}\end{pmatrix}$$
$$\xrightarrow{\mathcal{L}}\mathcal{N}\left(0,\begin{pmatrix}\left(v_t^\intercal\left(\Sigma^v\right)^{-1}\otimes\mathbb{I}_d\right)\Pi_{11}\left(\left(\Sigma^v\right)^{-1}v_t\otimes\mathbb{I}_d\right) & \left(v_t^\intercal\left(\Sigma^v\right)^{-1}\otimes\mathbb{I}_d\right)\Pi_{12}\eta^\intercal \\ \cdot & \eta\Pi_{22}\eta^\intercal\end{pmatrix}\right).$$

By (B.2) and Assumptions A.6 and I.6, we have

$$n^{-1/2}T^{-1}\eta H^{-1}\widehat{\Lambda}^{-1}\widehat{V}\bar{V}^\intercal\beta^\intercal u_t = n^{1/2}\eta(\beta^\intercal\beta)^{-1}\beta^\intercal u_t \xrightarrow{\mathcal{L}}\mathcal{N}\left(0,\eta\left(\Sigma^\beta\right)^{-1}\Omega_t\left(\Sigma^\beta\right)^{-1}\eta^\intercal\right).$$

The desired result follows from the same asymptotic independence argument as in Bai (2003). □

Proof of Theorem I.6. To prove the consistency of $\widehat{\Phi}$, without loss of generality, we focus on the case of Π_{12}, and show that

$$\left(\widetilde{\gamma}^\intercal\otimes\mathbb{I}_d\right)\widehat{\Pi}_{12}\widehat{\eta}^\intercal\xrightarrow{p}\left(\gamma^\intercal\left(\Sigma^v\right)^{-1}\otimes\mathbb{I}_d\right)\Pi_{12}\eta^\intercal. \tag{V.51}$$

The proof for the other two terms in $\widehat{\Phi}$ is similar and hence is omitted.

Note that by (IV.19), Lemma 2, Lemma 3(a), and Assumption A.5, we have

$$\begin{aligned}&\left\|T^{-1}H^{-1}\widehat{V}\widehat{V}^\intercal H^{-\intercal}-\Sigma^v\right\|_{\text{MAX}}\\ =&\left\|T^{-1}H^{-1}(\widehat{V}-H\bar{V})\widehat{V}^\intercal H^{-\intercal}+T^{-1}\bar{V}(\widehat{V}^\intercal-\bar{V}^\intercal H^\intercal)H^{-\intercal}+T^{-1}VV^\intercal-\Sigma^v-\bar{v}\bar{v}^\intercal\right\|_{\text{MAX}}\\ =&O_p(n^{-1}+T^{-1/2}).\end{aligned}$$

By (B.6), Lemma 2, and the proof of Theorem 1, we have

$$\|\widehat{\eta}H-\eta\|_{\text{MAX}}=O_p(n^{-1}+T^{-1/2}),\quad \left\|H^{-1}\widetilde{\gamma}-\gamma\right\|_{\text{MAX}}=O_p(n^{-1/2}+T^{-1/2}). \tag{V.52}$$

Therefore, to prove (V.51), we only need to show that

$$\widetilde{\Pi}_{12}:=(H^{-1}\otimes\mathbb{I}_d)\widehat{\Pi}_{12}H^{-\intercal}\xrightarrow{p}\Pi_{12}, \tag{V.53}$$

with which, and by the continuous mapping theorem, we have

$$\begin{aligned}\left(\widetilde{\gamma}^\intercal\left(\widehat{\Sigma}^v\right)^{-1}\otimes\mathbb{I}_d\right)\widehat{\Pi}_{12}\widehat{\eta}^\intercal&=\left((H^{-1}\widetilde{\gamma})^\intercal\left(H^{-1}\widehat{\Sigma}^vH^{-\intercal}\right)^{-1}\otimes\mathbb{I}_d\right)(H^{-1}\otimes\mathbb{I}_d)\widehat{\Pi}_{12}H^{-\intercal}(\widehat{\eta}H)^\intercal\\ &\xrightarrow{p}\left(\gamma^\intercal\left(\Sigma^v\right)^{-1}\otimes\mathbb{I}_d\right)\Pi_{12}\eta^\intercal.\end{aligned}$$

Writing $\widetilde{V} = H^{-1}\widehat{V}$, we have

$$\widetilde{\Pi}_{12,(i-1)d+j,i'} = \mathrm{vec}(e_j e_i^\intercal)^\intercal (H^{-1} \otimes \mathbb{I}_d)\widehat{\Pi}_{12} H^{-\intercal} e_{i'} = \mathrm{vec}(e_j e_i^\intercal H^{-1})^\intercal \widehat{\Pi}_{12} H^{-\intercal} e_{i'} = T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T} \widehat{z}_{jt}\widetilde{v}_{it} Q_{ts}\widetilde{v}_{i's},$$

where $Q_{st} = \left(1 - \frac{|s-t|}{q+1}\right) 1_{|s-t|\le q}$.

In fact, to show (V.53), by Lemma 2 we only need to prove for any fixed $1 \le i, i' \le p$, and $1 \le j, j' \le d$,

$$\widetilde{\Pi}_{12,(i-1)d+j,i'} - T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T} z_{jt} v_{it} Q_{ts} v_{i's} \xrightarrow{p} 0, \tag{V.54}$$

since by the identical proof of Theorem 2 in Newey and West (1987), we have

$$T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T} z_{jt} v_{it} Q_{ts} v_{i's} - \Pi_{12,(i-1)d+j,i'} \xrightarrow{p} 0.$$

Note that

$$\begin{aligned}
&\text{the left-hand side of (V.54)}\\
=&T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T}\Big\{(\widehat{z}_{jt} - z_{jt})(\widetilde{v}_{it} - v_{it})Q_{ts}(\widetilde{v}_{i's} - v_{i's}) + (\widehat{z}_{jt} - z_{jt})(\widetilde{v}_{it} - v_{it})Q_{ts}v_{i's}\\
&+ (\widehat{z}_{jt} - z_{jt})v_{it}Q_{ts}\widetilde{v}_{i's} + z_{jt}(\widetilde{v}_{it} - v_{it})Q_{ts}\widetilde{v}_{i's} + z_{jt}\widetilde{v}_{it}Q_{ts}(\widetilde{v}_{i's} - v_{i's})\Big\}.
\end{aligned}$$

We analyze these terms one by one. Since we have

$$\widehat{Z} - \bar{Z} = \eta\bar{V} - \widehat{\eta}\widehat{V} = (\eta H^{-1} - \widehat{\eta})H\bar{V} - (\widehat{\eta} - \eta H^{-1})(\widehat{V} - H\bar{V}) - \eta H^{-1}(\widehat{V} - H\bar{V}), \tag{V.55}$$

it follows from (B.6), (IV.7), and Lemmas 1 and 2 that

$$\begin{aligned}
&T^{-1}\left\|\widehat{Z} - \bar{Z}\right\|_{\mathrm{F}}\\
\le&KT^{-1}\left(\left\|\eta H^{-1} - \widehat{\eta}\right\|_{\mathrm{MAX}}\|H\|\left\|\bar{V}\right\|_{\mathrm{F}} + \left\|\widehat{\eta} - \eta H^{-1}\right\|_{\mathrm{F}}\left\|\widehat{V} - H\bar{V}\right\|_{\mathrm{F}} + \left\|\eta H^{-1}\right\|\left\|\widehat{V} - H\bar{V}\right\|_{\mathrm{F}}\right)\\
=&O_p(n^{-1/2}T^{-1/2} + T^{-1}).
\end{aligned}$$

Moreover, by Lemma 8, Assumption I.5, (V.55), and (V.52), we have

$$\begin{aligned}
&\left\|\widehat{Z} - \bar{Z}\right\|_{\mathrm{MAX}}\\
\le&\left\|\eta H^{-1} - \widehat{\eta}\right\|_{\mathrm{MAX}}\|H\|\left\|\bar{V}\right\|_{\mathrm{MAX}} + \left\|\widehat{\eta} - \eta H^{-1}\right\|_{\mathrm{MAX}}\left\|\widehat{V} - H\bar{V}\right\|_{\mathrm{MAX}} + \left\|\eta H^{-1}\right\|\left\|\widehat{V} - H\bar{V}\right\|_{\mathrm{MAX}}\\
=&O_p((\log T)^{1/a}T^{-1/2} + n^{-1/2}T^{1/4}).
\end{aligned}$$

By Cauchy-Schwartz inequality, Lemmas 1, 8, and using the fact that $|Q_{ts}| \le 1_{|t-s|\le q}$ and $\left\|\bar{v}\iota_T^\intercal\right\|_{\mathrm{F}} = \|\bar{v}\|_{\mathrm{F}} \left\|\iota_T^\intercal\right\|_{\mathrm{F}} \le KT^{1/2}\|\bar{v}\|_{\mathrm{MAX}} = O_p(1)$, we have

$$\begin{aligned}
&\left|T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T}(\widehat{z}_{jt}-z_{jt})(\widetilde{v}_{it}-v_{it})Q_{ts}(\widetilde{v}_{i's}-v_{i's})\right| \\
\le& KqT^{-1}\left(\left\|\widetilde{V}-\bar{V}\right\|_{\mathrm{MAX}}+\|\bar{v}\iota_T^\intercal\|_{\mathrm{MAX}}\right)\left(\left\|\widetilde{V}-\bar{V}\right\|_{\mathrm{F}}+\|\bar{v}\iota_T^\intercal\|_{\mathrm{F}}\right)\left(\left\|\widehat{Z}-\bar{Z}\right\|_{\mathrm{F}}+\|\bar{z}\iota_T^\intercal\|_{\mathrm{F}}\right) \\
=& O_p\left(q(T^{-1}+n^{-1})(T^{1/4}n^{-1/2}+T^{-1})\right).
\end{aligned}$$

Similarly, because of $\left\|\widetilde{V}\right\|_{\mathrm{F}} = O_p(T^{1/2})$ implied by (IV.8), $\|Z\|_{\mathrm{MAX}} = O_p((\log T)^{1/a})$ by Assumption I.5 and Lemma 2, and by Assumptions A.5 and A.8, we have

$$\begin{aligned}
&\left|T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T}(\widehat{z}_{jt}-z_{jt})(\widetilde{v}_{it}-v_{it})Q_{ts}v_{i's}\right| \\
\le& KqT^{-1}\|V\|_{\mathrm{MAX}}\left(\left\|\widetilde{V}-\bar{V}\right\|_{\mathrm{F}}+\|\bar{v}\iota_T^\intercal\|_{\mathrm{F}}\right)\left(\left\|\widehat{Z}-\bar{Z}\right\|_{\mathrm{F}}+\|\bar{z}\iota_T^\intercal\|_{\mathrm{F}}\right) = O_p\left(q(\log T)^{1/a}(n^{-1}+T^{-1})\right), \\
&\left|T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T}(\widehat{z}_{jt}-z_{jt})v_{it}Q_{ts}\widetilde{v}_{i's}\right| \\
\le& KqT^{-1}\|V\|_{\mathrm{MAX}}\left\|\widetilde{V}\right\|_{\mathrm{F}}\left(\left\|\widehat{Z}-\bar{Z}\right\|_{\mathrm{F}}+\|\bar{z}\iota_T^\intercal\|_{\mathrm{F}}\right) = O_p\left(q(\log T)^{1/a}\left(n^{-1/2}+T^{-1/2}\right)\right), \\
&\left|T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T}z_{jt}(\widetilde{v}_{it}-v_{it})Q_{ts}\widetilde{v}_{i's}\right| \\
\le& KqT^{-1}\|Z\|_{\mathrm{MAX}}\left\|\widetilde{V}\right\|_{\mathrm{F}}\left(\left\|H^{-1}\widehat{V}-\bar{V}\right\|_{\mathrm{F}}+\|\bar{v}\iota_T^\intercal\|_{\mathrm{F}}\right) = O_p\left(q(\log T)^{1/a}\left(n^{-1/2}+T^{-1/2}\right)\right), \\
&\left|T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T}z_{jt}\widetilde{v}_{it}Q_{ts}(\widetilde{v}_{i's}-v_{i's})\right| \\
\le& KqT^{-1}\|Z\|_{\mathrm{MAX}}\left\|\widetilde{V}\right\|_{\mathrm{F}}\left(\left\|H^{-1}\widehat{V}-\bar{V}\right\|_{\mathrm{F}}+\|\bar{v}\iota_T^\intercal\|_{\mathrm{F}}\right) = O_p\left(q(\log T)^{1/a}\left(n^{-1/2}+T^{-1/2}\right)\right).
\end{aligned}$$

All the above terms converge to 0, as $T, n \to \infty$, with $qT^{-1/4} + qn^{-1/4} \to 0$ and $n^{-3}T \to 0$, which establishes (V.54).

Finally, to show the consistency of $\widehat{\Upsilon}$, we first note

$$\left\|H^\intercal\left(\widehat{\Sigma}^\beta-\widehat{\beta}_0\widehat{\beta}_0^\intercal\right)H-\left(\Sigma^\beta-\beta_0\beta_0^\intercal\right)\right\|_{\mathrm{MAX}} \le \left\|H^\intercal\widehat{\Sigma}^\beta H-\Sigma^\beta\right\|_{\mathrm{MAX}}+\left\|H^\intercal\widehat{\beta}_0\widehat{\beta}_0^\intercal H-\beta_0\beta_0^\intercal\right\|_{\mathrm{MAX}}.$$

By Lemmas 2, 4(b), (e), and Assumption A.6,

$$\begin{aligned}
&\left\|H^\intercal\widehat{\Sigma}^\beta H-\Sigma^\beta\right\|_{\mathrm{MAX}} \\
\le&\left\|n^{-1}H^\intercal\widehat{\beta}^\intercal\widehat{\beta}H-n^{-1}\beta^\intercal\beta\right\|_{\mathrm{MAX}}+\left\|n^{-1}\beta^\intercal\beta-\Sigma^\beta\right\|_{\mathrm{MAX}} \\
\le&\left\|n^{-1}\left(H^\intercal\widehat{\beta}^\intercal-\beta^\intercal\right)(\widehat{\beta}H-\beta)+n^{-1}\left(H^\intercal\widehat{\beta}^\intercal-\beta^\intercal\right)\beta-n^{-1}\beta^\intercal(\beta-\widehat{\beta}H)\right\|_{\mathrm{MAX}}+o_p(1) \\
=&o_p(1).
\end{aligned} \tag{V.56}$$

$$
\begin{aligned}
&\left\|H^\intercal\widehat{\beta}_0\widehat{\beta}_0^\intercal H-\beta_0\beta_0^\intercal\right\|_{\text{MAX}}\\
\leq&\left\|\left(H^\intercal\widehat{\beta}_0-\beta_0\right)\left(\widehat{\beta}_0^\intercal H-\beta_0^\intercal\right)+\beta_0\left(\widehat{\beta}_0^\intercal H-\beta_0^\intercal\right)+\left(H^\intercal\widehat{\beta}_0-\beta_0\right)\beta_0^\intercal\right\|_{\text{MAX}}\\
=&o_p(1),
\end{aligned}
$$

where we also use Lemma 6(b):

$$
\left\|H^\intercal\widehat{\beta}_0-\beta_0\right\|_{\text{MAX}}=n^{-1}\left\|\left(H^\intercal\widehat{\beta}^\intercal-\beta^\intercal\right)\iota_n\right\|_{\text{MAX}}=o_p(1).
$$

Next, by Lemma 3(b) and (V.52), we have

$$
\begin{aligned}
\widehat{\sigma^\alpha}^2-(\sigma^\alpha)^2=&n^{-1}\left\|\bar{r}-\iota_n\widetilde{\gamma}_0-\widehat{\beta}\widetilde{\gamma}\right\|_{\text{F}}^2-(\sigma^\alpha)^2\\
=&n^{-1}\left\|\iota_n(\gamma_0-\widetilde{\gamma}_0)+\beta\gamma-\widehat{\beta}\widetilde{\gamma}+\beta\bar{v}+\bar{u}\right\|_{\text{F}}^2+n^{-1}\left\|\alpha\right\|_{\text{F}}^2-(\sigma^\alpha)^2\\
\leq&n^{-1}\left\|\iota_n\right\|_{\text{F}}^2\left\|\gamma_0-\widetilde{\gamma}_0\right\|_{\text{F}}^2+n^{-1}\left\|\beta\right\|_{\text{F}}^2\left\|\bar{v}\right\|_{\text{F}}^2+n^{-1}\left\|\bar{u}\right\|_{\text{F}}^2+n^{-1}\left\|(\widehat{\beta}H-\beta)\gamma\right\|_{\text{F}}^2\\
&+n^{-1}\left\|(\widehat{\beta}H-\beta)(H^{-1}\widetilde{\gamma}-\gamma)\right\|_{\text{F}}^2+n^{-1}\left\|\beta(H^{-1}\widetilde{\gamma}-\gamma)\right\|_{\text{F}}^2+o_p(1).
\end{aligned}
$$

Therefore, by (B.6) and the continuous mapping theorem,

$$
\widehat{\sigma^\alpha}^2\widehat{\eta}HH^{-1}\left(\widehat{\Sigma}^\beta-\widehat{\beta}_0\widehat{\beta}_0^\intercal\right)^{-1}H^{-\intercal}H^\intercal\widehat{\eta}^\intercal\xrightarrow{p}\Upsilon,
$$

which concludes the proof. □

Proof of Theorem I.7. For $\widehat{\Psi}_{1t}$, we can follow exactly the same proof as that of Theorem I.6, since, similar to (V.52) for $\widetilde{\gamma}$, we have the same estimate for $\widehat{v}_t$ by (V.50).

As to $\widehat{\Psi}_{2t}$, similarly, we only need to show

$$
\left\|H^\intercal\widehat{\Omega}H-\Omega\right\|_{\text{MAX}}=o_p(1).
$$

Then by the continuous mapping theorem, along with (V.52) and (V.56), we have

$$
\widehat{\Psi}_{2t}=\widehat{\eta}H\left(H^\intercal\widehat{\Sigma}^\beta H\right)^{-1}H^\intercal\widehat{\Omega}_tH\left(H^\intercal\widehat{\Sigma}^\beta H\right)^{-1}H^\intercal\widehat{\eta}^\intercal\xrightarrow{p}\Psi_{2t}.
$$

Note that by Fan et al. (2013), we have

$$
\left\|\widehat{\Sigma}^u-\Sigma^u\right\|=O_p(s_n\omega_T^{1-h}). \tag{V.57}
$$

Then by (V.57) and Lemmas 3(b), 4(b), and using the fact that $\|\beta\|_{\text{F}}=O_p(n^{1/2})$ and $\|\Sigma^u\|\leq\|\Sigma^u\|_1=O_p(s_n)$, writing $\widetilde{\beta}=\widehat{\beta}H$, we have

$$
\frac{1}{n}\left\|(\widetilde{\beta}-\beta)^\intercal(\widehat{\Sigma}^u-\Sigma^u)(\widetilde{\beta}-\beta)\right\|_{\text{MAX}}\leq\frac{1}{n}\left\|\widetilde{\beta}-\beta\right\|_{\text{F}}^2\left\|\widehat{\Sigma}^u-\Sigma^u\right\|=O_p\left(s_n\omega_T^{1-h}(n^{-1}+T^{-1})\right),
$$

$$\begin{aligned}
\frac{1}{n}\left\|(\widetilde{\beta}-\beta)^{\intercal}\Sigma^u(\widetilde{\beta}-\beta)\right\|_{\text{MAX}} \leq& \frac{1}{n}\left\|\widetilde{\beta}-\beta\right\|_{\text{F}}^2\|\Sigma^u\| = O_p\left(s_n(n^{-1}+T^{-1})\right),\\
\frac{1}{n}\left\|\beta^{\intercal}(\widehat{\Sigma}^u-\Sigma^u)\beta\right\|_{\text{MAX}} \leq& \frac{1}{n}\|\beta\|_{\text{F}}^2\left\|\widehat{\Sigma}^u-\Sigma^u\right\| = O_p\left(s_n\omega_T^{1-h}\right),\\
\frac{1}{n}\left\|\beta^{\intercal}(\widehat{\Sigma}^u-\Sigma^u)(\widetilde{\beta}-\beta)\right\|_{\text{MAX}} \leq& \frac{1}{n}\left\|(\widehat{\Sigma}^u-\Sigma^u)(\widetilde{\beta}-\beta)\beta^{\intercal}\right\| \leq \frac{1}{n}\left\|\widehat{\Sigma}^u-\Sigma^u\right\|\left\|\beta^{\intercal}(\widetilde{\beta}-\beta)\right\|\\
\leq& \frac{K}{n}\left\|\beta^{\intercal}(\widetilde{\beta}-\beta)\right\|_{\text{MAX}}\left\|\widehat{\Sigma}^u-\Sigma^u\right\| = O_p\left(s_n\omega_T^{1-h}(n^{-1}+T^{-1})\right),\\
\frac{1}{n}\left\|\beta^{\intercal}\Sigma^u(\widetilde{\beta}-\beta)\right\|_{\text{MAX}} \leq& \frac{K}{n}\left\|\beta^{\intercal}(\widetilde{\beta}-\beta)\right\|_{\text{MAX}}\|\Sigma^u\| = O_p\left(s_n(n^{-1}+T^{-1})\right).
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\left\|H^{\intercal}\widehat{\Omega}H-\Omega\right\|_{\text{MAX}} = \frac{1}{n}\left\|H^{\intercal}\widehat{\beta}^{\intercal}\widehat{\Sigma}^u\widehat{\beta}H-\beta^{\intercal}\Sigma^u\beta\right\|_{\text{MAX}}\\
\leq&\frac{1}{n}\left\|(\widetilde{\beta}-\beta)^{\intercal}(\widehat{\Sigma}^u-\Sigma^u)(\widetilde{\beta}-\beta)+(\widetilde{\beta}-\beta)^{\intercal}\Sigma^u(\widetilde{\beta}-\beta)\right\|_{\text{MAX}}+\frac{1}{n}\left\|\beta^{\intercal}(\widehat{\Sigma}^u-\Sigma^u)\beta\right\|_{\text{MAX}}\\
&+\frac{1}{n}\left\|\beta^{\intercal}(\widehat{\Sigma}^u-\Sigma^u)(\widetilde{\beta}-\beta)+(\widetilde{\beta}-\beta)^{\intercal}(\widehat{\Sigma}^u-\Sigma^u)\beta\right\|_{\text{MAX}}+\frac{1}{n}\left\|\beta^{\intercal}\Sigma^u(\widetilde{\beta}-\beta)+(\widetilde{\beta}-\beta)^{\intercal}\Sigma^u\beta\right\|_{\text{MAX}}\\
=&O_p\left(s_n\left(\omega_T^{1-h}+n^{-1}+T^{-1}\right)\right)=o_p(1),
\end{aligned}$$

which concludes the proof. □

References

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

Bai, J. (2003). Inferential Theory for Factor Models of Large Dimensions. *Econometrica 71*(1), 135–171.

Bai, J. and Y. Liao (2013). Statistical inferences using large estimated covariances for panel data and factor models. Technical report, Columbia University.

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

Bai, Z. and J. W. Silverstein (2009). *Spectral Analysis of Large Dimensional Random Matrices*. Springer.

Connor, G. and R. A. Korajczyk (1986). Performance measurement with the arbitrage pricing theory: A new framework for analysis. *Journal of Financial Economics 15*(3), 373–394.

Connor, G. and R. A. Korajczyk (1988). Risk and return in an equilibrium APT: Application of a new test methodology. *Journal of Financial Economics 21*(2), 255–289.

Fama, E. F. and K. R. French (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics 33*(1), 3–56.

Fama, E. F. and K. R. French (2015). A five-factor asset pricing model. *Journal of Financial Economics 116*(1), 1–22.

Fan, J., Q. Li, and Y. Wang (2017). Estimation of high dimensional mean regression inthe absence of symmetry and light tail assumptions. *Journal of the Royal Statistical Society, B 79*(1), 247–265.

Fan, J., Y. Liao, and M. Mincheva (2011). High-dimensional covariance matrix estimation in approximate factor models. *Annals of Statistics 39*(6), 3320–3356.

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

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

Fan, J., D. Wang, K. Wang, and Z. Zhu (2017). Distributed estimation of principal eigenspaces. *arXiv preprint arXiv:1702.06488*.

Fan, J., W. Wang, and Y. Zhong (2019). Robust covariance estimation for approximate factor models. *Journal of Econometrics 208*(1), 5–22.

Gibbons, M., S. A. Ross, and J. Shanken (1989). A test of the efficiency of a given portfolio. *Econometrica 57*(5), 1121–1152.

Gospodinov, N., R. Kan, and C. Robotti (2014). Misspecification-Robust Inference in Linear Asset-Pricing Models with Irrelevant Risk Factors. *The Review of Financial Studies 27*(7), 2139–2170.

Horn, R. A. and C. R. Johnson (2013). *Matrix Analysis* (Second ed.). Cambridge University Press.

Kan, R., C. Robotti, and J. Shanken (2013). Pricing model performance and the two-pass cross-sectional regression methodology. *The Journal of Finance 68*(6), 2617–2649.

Kelly, B., S. Pruitt, and Y. Su (2019). Characteristics are covariances: A unified model of risk and return. *Journal of Financial Economics, forthcoming*.

Kozak, S., S. Nagel, and S. Santosh (2018). Interpreting factor models. *Journal of Finance 73*(3), 1183–1223.

Lettau, M. and M. Pelger (2018). Factors that fit the time series and cross-section of stock returns. Technical report, National Bureau of Economic Research.

MacKinlay, A. C. and M. P. Richardson (1991). Using Generalized Method of Moments to Test Mean-Variance Efficiency. *The Journal of Finance 46*(2), 511–527.

Newey, W. K. and K. D. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*, 703–708.

Novy-Marx, R. (2014). Predicting anomaly performance with politics, the weather, global warming, sunspots, and the stars. *Journal of Financial Economics 112*(2), 137–146.

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

Pesaran, M. and T. Yamagata (2012). Testing CAPM with a large number of assets. Technical report, University of South California.

Rosenberg, B. (1974). Extra-market components of covariance in security returns. *Journal of Financial and Quantitative Analysis 9*(2), 263–274.