Factors Common to Individual Stock and Sorted Portfolio Returns *

Elena Andreou† **Patrick Gagliardini**‡ **Eric Ghysels**§ **Mirco Rubin**¶

This Draft: February 7, 2022

Abstract

There exist two commonly used approaches to cross-sectional asset pricing, each with pros and cons. One consists of collecting stocks into portfolios and subsequently estimate risk exposures. The other consists of estimating cross-sectional risk premia using the entire universe of stocks. Applying a novel test, we identify the factor space common between individual stocks and sorted portfolios - neither affected by time-varying betas nor by the sorting characteristics. We find three factors - which can only partially be explained by Fama-French five factors with(out) momentum. Our three factors also feature superior out-of-sample pricing performance compared to standard pricing models.

Keywords: Testing common factors, portfolio sorting, factor zoo

JEL Codes: C22, C38, C53, C55, G10, G12 .

*We would like to thank Andrei Gonçalves, Cam Harvey, Abraham Lioui, Daniele Massacci, Filippos Papakonstantinou and Raman Uppal for insightful comments, as well as seminar participants at EDHEC Business School, Nice, Kenan-Flagler Finance and UNC Economics seminars, LUISS Business School and King's College London, and participants to the 2021 Annual SoFiE Conference at UCSD Rady School of Business, San Diego and the 2021 Annual Conference of the IAAE at Erasmus School of Economics, Rotterdam. The first author would like to acknowledge that this work was funded by the Republic of Cyprus through the Research and Innovation Foundation (Project: INTERNATIONAL/USA/0118/0043). The authors are extremely grateful to Andrew Chen from the Board of Governors of the Federal Reserve for providing us with his extensive factor zoo data set.

†University of Cyprus and CEPR, e-mail: elena.andreou@ucy.ac.cy.

‡Università della Svizzera italiana, Lugano and Swiss Finance Institute, e-mail: patrick.gagliardini@usi.ch.

§University of North Carolina - Chapel Hill and CEPR, e-mail: eghysels@unc.edu.

¶EDHEC Business School, Nice, e-mail: mirco.rubin@edhec.edu.

1 Introduction

The dominant research theme in empirical asset pricing is the low dimensional factor representation of a large set of asset returns. Ideally any high dimensional set of asset returns should contain the information necessary to recover the factors. In practice, the literature has taken two different approaches. Jensen, Black, and Scholes (1972) and Fama and MacBeth (1973), among many others, have advocated to collect stocks into portfolios and subsequently run cross-sectional regressions using portfolios as test assets. An alternative approach is to estimate cross-sectional risk premia using the entire universe of stocks as advocated by Litzenberger and Ramaswamy (1979), among others.

One might think that the choice between individual stocks versus sorted portfolios should only be a matter of practical implementation and ultimately should uncover the same low dimensional factor space. Unfortunately this is not the case. Extracting factors from individual stocks using static models is believed to overstate the "true" set because time-varying loadings may add spurious factors (see e.g. Breitung and Eickmeier (2011), among others). The main advantage of using portfolios is that their risk exposures are more stable across time. This being said, it is also known that portfolios might diversify away and therefore mask relevant risk- or return-related features of individual assets. Moreover, as Lewellen, Nagel, and Shanken (2010) point out, sorting on characteristics also results in a strong factor structure across test portfolios and indeed even factors that are weakly correlated with the sorting characteristics would explain the differences in average returns across test portfolios regardless of the economic theories underlying the factors.

It seems that we want to find the low dimensional factor representation able to price the cross-section of sorted portfolios *and* that of individual stocks. To put it differently, when we extract factors from stocks we want them to also correctly price sorted portfolios and vice versa. Using terminology from the factor model literature, we want to find the factor space that is *common* between panels of individual stocks and panels of sorted portfolio returns as it provides a path toward extracting the true set of factors neither affected by sorting characteristics nor by varying risk exposures and recalcitrant features of individual stocks. In the remainder of the paper we will call these the *common* factors.

The task of finding common factors is not trivial and requires theoretical insights so far not explored in the empirical asset pricing literature. The approach we use in this paper was first alluded to in the last section of the Roll and Ross (1980) paper on the empirical testing of the APT and remained largely unresolved since then. Numerous attempts have been made to address the problem, including recently by Pukthuanthong, Roll, and Subrahmanyam (2019). To achieve the task set forth we need to expand the theory underpinning a procedure recently proposed by Andreou, Gagliardini, Ghysels, and Rubin (2019) (henceforth AGGR). They study a situation where (latent) factors $h_{1,\tau}$ and $h_{2,\tau}$, are estimated from two separate panels of data, and one is interested in testing how many factors are common between them. AGGR show that the common factor space is identified by examining how many linear combinations of respectively $h_{1,\tau}$ and $h_{2,\tau}$ are perfectly correlated. Equivalently, they introduce a test for the number of canonical correlations between $h_{1,\tau}$ and $h_{2,\tau}$ equal to one and derive its asymptotic distribution.[1] Their analysis does not directly translate into a procedure suitable for asset pricing applications. One of the contributions of our paper is to provide the theory for such applications. It should parenthetically be noted that when we refer to latent factors, we do not necessarily mean principal component analysis (PCA). Indeed, our analysis also covers recently proposed procedures such as for example those advocated by Lettau and Pelger (2020a and 2020b).

The novel testing procedure identifies 3 factors at the intersection of individual stock returns and sorted portfolios. Surprisingly, we find that neither the Fama and French 3 (FF3) nor 5 (FF5) factor models, both with or without a momentum factor (hence up to 6 factors), span the factor space common between individual stocks and sorted portfolios. In fact out of the 6 factors considered, only the excess market returns factor seems to be the most related to the common factors, while all the other 5 factors are only partially spanned by the common factors, and a large part of their variability are specific to portfolio sorting. For convenience we will call the 3 common factors 3CF.

The search for factors has been on steroids with literally hundreds of potential additional candidate factors beyond FF3 suggested in the literature. The endeavor has been dubbed the factor zoo by Cochrane (2011) and terms such as p-hacking (meaning data-snooping or data-mining) have been used to describe the hunt for factors.[2] The literature started of with the pretty tame single factor model, i.e. the CAPM. It is perhaps more appropriate to say that we moved from a petting zoo to

[1] To sort out genuine risk factors Pukthuanthong et al. (2019) also rely on canonical correlations, but do not present a formal statistical procedure.

[2] See Harvey, Liu, and Zhu (2016), McLean and Pontiff (2016), Chorida, Goyal, and Saretto (2020) Hou, Xue, and Zhang (2020), Feng, Giglio, and Xiu (2020), Chen (2019) among others.

a jungle. For example, Harvey and Liu (2019) have documented over 400 factors published in top journals. In our empirical application we use a data set of over a thousand portfolios associated with 205 characteristics. It takes up to 10 PCs from this factor zoo to span the space of 3CF.

We perform a comprehensive in- and out-of-sample (OOS) analysis of the pricing performance of 3CF compared to a wide range of standard models as well as the factor zoo. Using multiple in-sample performance evaluation measures we find that the three common factors perform better than a large collection of observable and latent factor models in pricing individual stock and sorted portfolios assets. Turning to the out-of-sample analysis the results yield several interesting empirical findings. The three common factors yield again the highest total, pricing and predictive OOS R^2s with respect to the same benchmark models. For the individual stocks as well as sorted portfolios the OOS predictive R^2s gains using 3CF can be 80% and 50% vis-à-vis for example the Fama-French factors.

We regress factors from the zoo onto 3CF and document which ones yield the best fit. We find two of the three Fama and French factors (CAPM Beta and Size) along with portfolios based on market beta put forward by Frazzini and Pedersen (2014), and different measures of idiosyncratic risk and liquidity or uncertainty, such as Bid-ask Spread, Cash-flow to price variance, Volume to market equity, EPS Forecast Dispersion, Days with zero trades, Volume Variance, and Price delay R-square.

The rest of the paper is organized as follows. Section 2 introduces the spanning test and details the data on the various cross-sections of asset returns used to estimate common latent factor spaces. Section 3 covers the empirical implementation of the testing procedure, followed by Section 4 where we report the results of an extensive empirical study comparing the asset pricing performance of the common factors with widely used factors in the asset pricing literature. Section 5 revisits the topic of the factor zoo. Conclusions appear in Section 6.

2 Factor Space Spanning Test

We consider a situation where we have a panel of individual stocks as well as a panel of portfolios returns combining those individual stocks. In a first subsection we present a formal framework for the characterization of the factor model representation for both panels. A second subsection provides the details regarding the testing procedure to identify and estimate the factors common between the two panels. A final subsection described data sources.

2.1 Sorting and time-varying loadings

We assume that individual stock excess returns have the following factor structure:

$$r_{i,\tau} = b'_{i,\tau-1} f^c_\tau + e_{i,\tau} \qquad i = 1, \dots, N_1 \quad \tau = 1, \dots, t, \tag{2.1}$$

with $b_{i,\tau-1}$ is the vector of conditional betas and f^c_τ a set of k^c factors, where the superscript c will be clarified later. The conditional betas are driven by common ($Z_{\tau-1}$) as well as stock-specific ($\tilde{Z}_{i,\tau-1}$) variables (see e.g. Gagliardini, Ossola, and Scaillet (2016), namely: $b_{i,\tau-1} = b^0_i + B_i Z_{\tau-1} + C_i \tilde{Z}_{i,\tau-1}$, and therefore:

$$r_{i,\tau} = [b^0_i + B_i Z_{\tau-1} + C_i \tilde{Z}_{i,\tau-1}]' f^c_\tau + e_{i,\tau} \ . \tag{2.2}$$

Moreover, we can write the time-varying loadings in equation (2.2) as follows:

$$\left[B_i Z_{\tau-1} + C_i \tilde{Z}_{i,\tau-1}\right]' f^c_\tau = M^{c\prime}_i f^c_\tau + M^{s\prime}_i f^s_{1,\tau} + \delta_{i,\tau} \tag{2.3}$$

with $f^c_\tau \perp f^s_{1,\tau}$ and both factor vectors orthogonal to the errors $\delta_{i,\tau}$ (an illustration of the above equation appears in Appendix Section A.1).[3] In the above equation the (sort of scaled) factors $f^s_{1,\tau}$ emerge from the product of factors f^c_τ times the common component driving the loadings. The remainder term $\delta_{i,\tau}$ is assumed idiosyncratic, or (at most) weakly cross-sectionally correlated among the individual stocks. Then:

$$r_{i,\tau} = [b^0_i + M^c_i]' f^c_\tau + [M^s_i]' f^s_{1,\tau} + \tilde{e}_{i,\tau}, \tag{2.4}$$

which means that the constant loadings representation of individual stocks features the genuine factors f^c_τ augmented by $f^s_{1,\tau}$, which are spurious factors present because of the time-varying betas (with super/subscripts s and 1 referring to factors specific to the first panel).

Let us now examine the creation of spurious factors generated by characteristic-sorted portfolios which is the more popular approach in empirical asset pricing in part because the easy availability of

[3]It should parenthetically be noted that stock-specific variables $\tilde{Z}_{i,\tau-1}$ typically represent characteristics but also could represent interactions of characteristics. Moreover, in addition to single sorting, our analysis also covers portfolios built on double, triple etc. characteristic sorting. See in particular Appendix Section A.2 for further details.

test asset portfolios from websites such as the one maintained by Kenneth French. The usual argument in favor of test asset portfolios is that they dramatically reduces the noise stemming from idiosyncratic risk, see e.g. the discussion in the recent work of Harvey and Liu (2021). We start with sorting along a single characteristic and consider more complex sorting schemes later. Namely, suppose we create j = 1, ..., N_2 portfolios with weight for stock i in portfolio j : $\alpha_{j,i,\tau-1}/N_1$ = $g_j(\tilde{z}_{i,\tau-1})/N_1$, with N_1 the number of individual stocks (as specified in equation (2.1)) and $\tilde{z}_{i,\tau-1} \in \tilde{Z}_{i,\tau-1}$ is a single asset i characteristic. Then the portfolio excess returns are:

$$r_{j,\tau}^p = \left[\frac{1}{N_1}\sum_{i=1}^{N_1}\alpha_{j,i,\tau-1}b_{i,\tau-1}\right]' f_\tau^c + u_{j,\tau} \qquad j = 1, \ldots, N_2 \quad \tau = 1, \ldots, t, \tag{2.5}$$

where $u_{j,t} = \frac{1}{N_1}\sum_{i=1}^{N_1}\alpha_{j,i,\tau-1}e_{i,\tau}$. Note that without loss of generality the portfolios involve the entire cross-section of stocks, with the obvious implication that only a subset of weights might be non-zero. Moreover, suppose that the betas for portfolio j are such that:

$$\begin{aligned}\frac{1}{N_1}\sum_{i=1}^{N_1}\alpha_{j,i,\tau-1}b_{i,\tau-1} &= \frac{1}{N_1}\sum_{i=1}^{N_1}g_j(\tilde{z}_{i,\tau-1})\left[b_i^0 + B_i Z_{\tau-1} + C_i\tilde{Z}_{i,\tau-1}\right] \\ &\simeq \tilde{b}_j^0 + W_j Z_{\tau-1}^* \qquad \text{as } N_1 \to \infty, \quad j = 1, \ldots, N_2 \quad \tau = 1, \ldots, t, \end{aligned} \tag{2.6}$$

for some $\tilde{b}_j^0$, $Z_{\tau-1}^*$ and W_j. Therefore, we have portfolio returns:

$$r_{j,\tau}^p = \left[\tilde{b}_j^0 + W_j Z_{\tau-1}^*\right]' f_\tau^c + u_{j,\tau}, \qquad j = 1, \ldots, N_2 \quad \tau = 1, \ldots, t. \tag{2.7}$$

Let us define the linear projection (with super/subscripts s and 2 referring to factors specific to the second panel):

$$\left[W_j Z_{\tau-1}^*\right]' f_\tau^c = K_j^{c\prime} f_\tau^c + K_j^{p\prime} f_{2,\tau}^s + \eta_{j,\tau} \qquad j = 1, \ldots, N_2 \quad \tau = 1, \ldots, t,$$

with $f_\tau^c \perp f_{2,\tau}^s \perp \eta_{j,\tau}$ for all j and $\eta_{j,\tau}$ weakly cross-sectionally correlated. This yields the following constant loading factor representation of sorted portfolio returns:

$$r_{j,\tau}^p = \left[\tilde{b}_j^0 + K_j^c\right]' f_\tau^c + K_j^{p\prime} f_{2,\tau}^s + \tilde{u}_{j,\tau}, \tag{2.8}$$

which means that the constant loadings representation of sorted portfolios features the genuine factors f_τ^c augmented by spurious factors $f_{2,\tau}^s$ to account for characteristics-based portfolio sorts and time-varying betas.

Under some mild regularity conditions appearing in the Appendix Section A.3 pertaining to masking of factors in portfolio sorts and convergence of portfolio betas, we have two panels - one consisting of individual stock excess returns and the other of sorted portfolios with returns described by equations (2.4) and (2.8), respectively. It is worth describing here some of the regularity conditions which are detailed in Appendix Section A.3. Assumption A.1 allows for some individual stocks and some portfolios to mask some factors, but not a significant collection of stocks or portfolios is allowed to block out any specific factor. In addition, Assumption A.2 guarantees that the common factors between the two panels to be f_τ^c only, instead of f_τ^c and/or $Z'_{\tau-1}f_\tau^c$. In the Appendix we discuss various scenarios, showing that overall the regularity conditions are mild. They imply that:

$$\begin{aligned}
&\mathbb{E}\left[f_\tau^c\right] = \mu^c\ , \qquad \mathbb{E}\left[f_{1,\tau}^s\right] = \mu_1^s\ , \qquad \mathbb{E}\left[f_{2,\tau}^s\right] = \mu_2^s\ , \\
&\mathbb{E}\left[(f_\tau^c - \mu^c)(f_{1,\tau}^s - \mu_1^s)'\right] = 0 && \text{by construction,} \\
&\mathbb{E}\left[(f_{1,\tau}^s - \mu_1^s)(f_{2,\tau}^s - \mu_2^s)'\right] = \Phi\ , \\
&\mathbb{E}\left[(f_\tau^c - \mu^c)(f_{2,\tau}^s - \mu_2^s)'\right] = 0 && \text{by construction.}
\end{aligned} \tag{2.9}$$

The equations in (2.9) imply that factors appearing in panels of individual stock or portfolio excess returns due to time-varying betas and factors appearing in portfolios due to characteristic-based sorting and time-varying betas may be mutually dependent via the covariance matrix Φ. What is important, however, is the orthogonality of f_τ^c with respect to both $f_{1,\tau}^s$ and $f_{2,\tau}^s$, by the linear projection arguments.

Summarizing, we have shown that if individual stocks were used, a researcher would conclude that f_τ^c and $f_{1,\tau}^s$ are risk factors. Conversely, a researcher starting from sorted portfolio returns would conclude that the risk factors are instead f_τ^c and $f_{2,\tau}^s$. Whilst to the best of our knowledge, other methods cannot address this, our testing procedure allows us to identify f_τ^c, namely the factors which are the drivers of both individual stock and sorted portfolios excess returns.

2.2 Estimation and testing

Since the testing approach applies to different settings beyond panels of individual stock and sorted portfolio returns we use the generic notation $y_{j,\tau} = [y_{j,1\tau}, ..., y_{j,N_j\tau}]'$ collects N_j observations in panel (group) j = 1, 2. It will be convenient to use the terms of group factor models, and interchangeably refer to groups and panels 1 and 2. To formulate the various hypotheses of interest we borrow the notation from AGGR, for the (two) group factor model setting:

$$\begin{bmatrix} y_{1,\tau} \\ y_{2,\tau} \end{bmatrix} = \begin{bmatrix} \Lambda_1^c & \Lambda_1^s & 0 \\ \Lambda_2^c & 0 & \Lambda_2^s \end{bmatrix} \begin{bmatrix} f_\tau^c \\ f_{1,\tau}^s \\ f_{2,\tau}^s \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,\tau} \\ \varepsilon_{2,\tau} \end{bmatrix}, \qquad \tau = 1, \ldots, t \tag{2.10}$$

where $\Lambda_j^c = [\lambda_{j,1}^c, ..., \lambda_{j,N_j}^c]'$ and $\Lambda_j^s = [\lambda_{j,1}^s, ..., \lambda_{j,N_j}^s]'$ are the matrices of factor loadings and $\varepsilon_{j,\tau} = [\varepsilon_{j,1\tau}, ..., \varepsilon_{j,N_j\tau}]'$ the error terms, with $\tau = 1, \ldots, t$. The dimensions of the common factor f_τ^c and the group-specific factors $f_{1,\tau}^s$, $f_{2,\tau}^s$ are respectively k^c, k_1^s and k_2^s. In the remainder of this subsection we go through the four steps of the procedure, the technical details and formal definitions appear in Appendix A.5.

Step 1: We start with extracting factors from each panel separately. For the application in the paper, this means extracting factors from individual stock returns and doing the same for sorted portfolios. As is typically done (see e.g. Lehmann and Modest (1988) and recently Kim and Korajczyk (2021), among many others), we will study non-overlapping sample blocks of generic length w of the panels covering data from $t - w + 1$ to t.[4] We focus here on one sample and resort to either PCA or a version of principal component analysis with a penalty term accounting for the cross-sectional pricing error in expected returns recently suggested by Lettau and Pelger (2020a and 2020b), and summarized in Appendix A.4. In particular, their estimator searches for factors that can explain both the expected return and covariance structure.

Step 2: Let $k_j = k^c + k_j^s$, for $j = 1, 2$, be the dimensions of the factor spaces for the two panels,

[4]We use a block sampling scheme to avoid look ahead biases in full sample factor extraction as well as survivorship biases for individual firms (see section OA.1 in the Online Appendix for further discussion). Our theory is based on asymptotic expansions, but as Andreou et al. (2019) show via simulation, it is also suitable to describe finite sample behavior in settings corresponding to the empirical application of the paper.

and define $\underline{k} = \min(k_1, k_2)$. We collect the factors of each group in the k_j-dimensional vectors $h_{j,\tau}$ then $h_{1,\tau} = \mathcal{H}_1 \left[f_\tau^{c\prime}\ ,\ f_{1,\tau}^{s\prime}\right]'$ and $h_{2,\tau} = \mathcal{H}_2 \left[f_\tau^{c\prime}\ ,\ f_{2,\tau}^{s\prime}\right]'$, meaning that the factors we extract from each group are some linear transformation $\mathcal{H}_j$ of the underlying factors, with $\mathcal{H}_j$ being a $k_j \times k_j$ full rank matrix, for $j = 1, 2$. This means that some linear combinations of $h_{1,\tau}$ - namely those corresponding to f_τ^c - are *perfectly* correlated with linear combinations of $h_{2,\tau}$ and vice versa. Let us recall at this point the purpose of canonical correlation analysis. In general canonical correlation applies to a setting where we have two random vectors, in our application $h_{1,\tau}$ and $h_{2,\tau}$, and finds linear combinations of respectively $h_{1,\tau}$ and $h_{2,\tau}$ which have maximum correlation with each other. Therefore we are interested in finding how many of these linear combinations, also known as canonical variables, are perfectly correlated, i.e. have canonical correlation equal to one.

Step 3: Proposition A.1 in the Appendix tells us that the dimension k^c is the number of unitary canonical correlations between $h_{1,\tau}$ and $h_{2,\tau}$. The largest possible number of common factors is $\underline{k} = \min(k_1, k_2)$. We develop a test for k^c : $H_0(r)$: $k^c = r$ against $H_1(r)$: $k^c < r$, for any given $r = \underline{k}, \underline{k} - 1, \ldots, 1$. More precisely, we sort the canonical correlations from high to low and let $\hat{\rho}_\ell$ be the ℓ-th sorted sample canonical correlation between the factors $\hat{h}_{1,\tau}$ and $\hat{h}_{2,\tau}$ estimated on a sample of length w, and let:

$$\hat{\xi}(r) = \sum_{\ell=1}^{r} \hat{\rho}_\ell, \tag{2.11}$$

be the sum of the r largest sample canonical correlations. We reject the null for $r = k^c$ common factors $H_0 = H(k^c)$ when $\hat{\xi}(k^c) - k^c$ is negative and large - namely the sum of the largest k^c estimated canonical correlations is substantially less than k^c.[5] The test statistic is:

$$\tilde{\xi}(k^c) := N\sqrt{w} \left(\frac{1}{2} tr\{\hat{\Sigma}_U^2\}\right)^{-1/2} \left[\hat{\xi}(k^c) - k^c + \frac{1}{2N} tr\left\{\hat{\Sigma}_U\right\}\right], \tag{2.12}$$

with $N = \min\{N_1, N_2\}$ and the term $\hat{\Sigma}_U$ is defined in the technical appendix. In the generic case, under the null hypothesis $H_0(r)$: $k^c = r$ we have: $\tilde{\xi}(r) \xrightarrow{d} N(0, 1)$, and under the alternative hypothesis $H_1(r)$: $k^c < r$, $\tilde{\xi}(r) \xrightarrow{p} -\infty$ as N and T grow large.[6]

[5] When we reject the null $H(\underline{k})$ we look at the null hypothesis: $H(\underline{k} - 1) = \left\{\rho_1 = \ldots = \rho_{\underline{k}-1} = 1\right\}$, and so forth until we identify the dimension of the common factor space. Sequential testing issues are addressed in Andreou et al. (2019).

[6] See Theorem 2 of AGGR, and its extension Theorem A.2 in Appendix Section A.5.2. The asymptotic distribution and rate of convergence of the test statistic $\tilde{\xi}(k^c)$ in Theorem A.2 are unchanged when the true numbers of factors k_1 and k_2

Step 4: Once the dimension k^c is identified, we can recover the common factors f_τ^c via the canonical directions - i.e. the weights of the linear combinations yielding unitary canonical correlations - applied to the factors estimated from each of the separate panels.

It is worth highlighting a number of theoretical contributions of the paper. The theory in AGGR only covers panel data centered at zero. Appendix A.5 extends the estimators and theoretical results of Andreou et al. (2019) to the case where factors are allowed to have any finite mean, compatible with model (2.9). This general set-up is more relevant for asset pricing applications. Moreover, the current paper shows that the testing and estimation procedures for common factors across different panels based on canonical correlation and directions can be applied (a) to "classical" PCA estimators of the factors, and more importantly for asset pricing (b) to the more recent variations of PCA as proposed by Lettau and Pelger (2020a). Appendix A.4 shows the formal relationship between the different PC-type estimators for factors.

2.3 Data

To conclude a few words about the data. In Online Appendix, henceforth OA, Section OA.1 we provide a detailed description of the data. Broadly speaking we can summarize the data as follows. We consider three panels of monthly returns in our analysis, namely (i) individual US stock returns from CRSP, (ii) the panel of test asset portfolios from the April 2021 release of the database "Open Source Cross-Sectional Asset Pricing" created by Chen and Zimmermann (2021), CZ21 hereafter, and (iii) the panel of factors from the zoo considered by CZ21. For all three panels from Jan. 1966 to Dec. 2020 we split the 660 months into $B = 11$ non-overlapping blocks of 60 months, denoted as $b = 1, ..., B$. The first block is from Jan. 1966 to Dec. 1970 and the last block is from Jan. 2016 to Dec. 2020. Within each block, we consider only a balanced sample of individual stocks and test asset portfolios, that is we only include assets with returns available for all the 60 months. We work with 5-year non-overlapping samples, analogous to the empirical application of Lehmann and Modest (1988), to address the concern of survivorship bias if we were to use the full sample of individual stocks. Similar to the arguments in Kim and Korajczyk (2021), one can view the 5-year span as a compromise between a sample large enough for our test procedure to have desirable small sample

are unknown, and are estimated by some consistent empirical selection method.

properties and the concern of capturing new and disappearing stocks.

3 Testing Test Assets

In order to identify the factor space neither affected by portfolio sorting characteristics nor by varying risk exposures and other specific features of individual stocks, we implement the four-steps procedure described in Section 2.2 for each of non-overlapping sample blocks from end of 1970 until end of 2020 with 5-years increments. These are 5-years samples $y_{1,\tau}$ of balanced panels of individual stock returns and $y_{2,\tau}$ balanced panels of test asset portfolios.[7] Similar to Pukthuanthong et al. (2019) we decide to fix a priori the maximum number of pervasive factors in each panel k_1 and k_2 to 10.

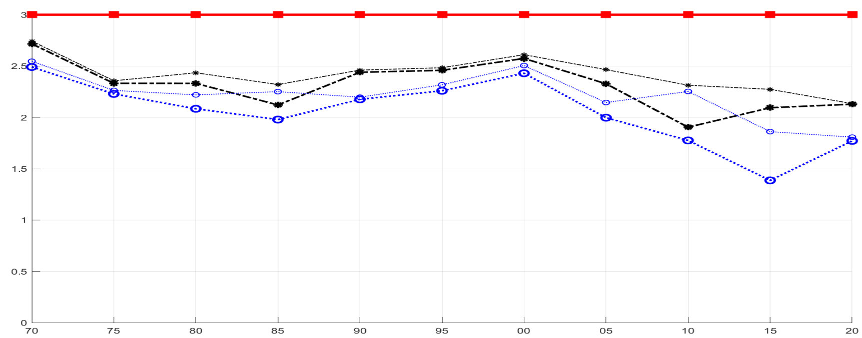
The pervasive factors in each balanced panel of assets are computed using Lettau and Pelger's Risk-premium PCs, or RP-PCs , fixing $\gamma_{LP} = -1$, and we simply refer to them as PCs. The number k_t^c of common factors between the first 10 PCs of individual stocks and the first 10 PCs of portfolio test assets is 3 for all 5-years blocks except for the blocks ending in 2000, 2005, 2015, 2020 where it is 4. While there is some variation we will proceed with the number of factors being equal to 3 across the entire sample. Henceforth we will refer to these three factors as the "common" factors and use the acronym 3CF. A first subsection is devoted to the testing of the common factor space and the second subsection covers the economic interpretation of the common factors.

Figure 1 displays the sum of the canonical correlations of the three factors common to the CRSP and CZ21 test assets with the 3 Fama and French factors (FF3): Market, SMB and HML factors, and the sum of the canonical correlations of the common factors with the 5 Fama and French factors (FF5): FF3 plus RMW-operating profitability, CMA-investment style. The figure also displays the sum of canonical correlations between common factors and FF3/FF5 factors augmented with the momentum factor. The red line across the plot marks the 3-factor benchmark common factor space.

The results in the figure convey a surprisingly simple and clear message. We observe that over the entire sample the canonical correlations between FF3 and common factors (blue circles, thick dotted line) are well below 3. This implies that over this sample period FF3 does not span the common factor space. What happens if we move from FF3 to FF5, i.e. we add RMW-operating profitability

[7] More precisely, we have 11 panels ending at t = Dec. 1970, Dec. 1975, ..., Dec. 2020 with observations $y_{1,\tau}$ and $y_{2,\tau}$ for $\tau = t - 59, t - 58, \ldots, t$.

Figure 1: Sum of canonical correlations of three factors common between Chen and Zimmermann (2021) test asset portfolios and CRSP stocks with Fama and French 3, 5 factors and momentum.



The figure displays the sum of the canonical correlations of the three factors common across CRSP and CZ21 test assets with the 3 Fama and French factors (FF3): Market, SMB and HML factors (blue circles, thick dotted line), and the sum of the canonical correlations of the three common factors with the 5 Fama and French factors (FF5) - adding RMW-operating profitability, CMA-investment style (black stars, thick dashed line). In addition, it also displays FF3/FF5 augmented with the momentum factor (FF3+Mom - blue circles, thin dotted line, FF5+Mom - black stars, thin dashed line). The red line across the plot marks the 3-factor benchmark common factor space. All quantities are computed on non-overlapping blocks of 5 years of monthly data, that is for each year y we report results computed on the block starting in year $y-4$ and ending in year y, for each $y = 1970, ..., 2020$.

and CMA-investment style? In the same figure, using the same approach, the black thick dashed line with stars shows that adding two FF factors falls again short of spanning the common factor space. In fact, in most years the improvements of the two additional factors appears to only be minor. The same analysis is repeated with as observable factors FF3 and FF5 plus momentum (FF3+Mom - blue circles, thin dotted line, FF5+Mom - black stars, thin dashed line). While the higher number of observable factors increases mechanically the value of the sum of non-zero canonical correlations, it remains the case that adding the momentum factor is not enough to span the common factors. This means that the popular models fall short of capturing the three common factors.[8] In addition, we examine formally whether the differences between the FF5+momentum factors versus the common factors are statistically significant. Andreou, Gagliardini, Ghysels, and Rubin (2021) derive a formal test between latent and observed factors similar in spirit to the test of AGGR explained in the previous section. Applying such a test, we reject the null that the FF5 and momentum factors span the k^c = 3 common factors - put differently all the lines in Figure 1 are significantly below 3. Finally, this also begs the

[8]In Online Appendix Section OA.3 we provide details about the composition of the three common factors in terms of their rescaled factor loadings.

question whether members of the zoo might help us out in recovering the common factor space, a topic addressed in Section 5.

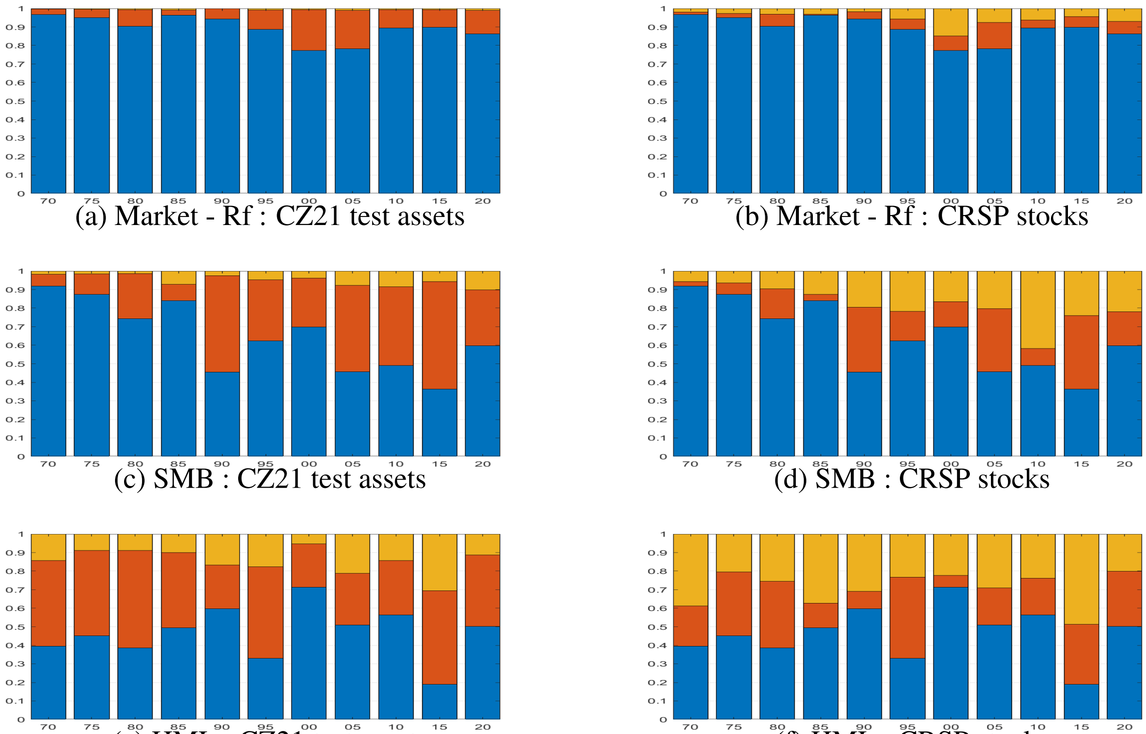
So far we established the existence of common factors between individual stocks and sorted portfolios, and shown that they are only partially spanned by FF5 and momentum. We now study how each of the FF5 and momentum observable factors taken one-by-one are related to (a) the common factors and (b) the factors which are specific respectively to the sorted portfolios or to the CRSP individual stocks. This analysis allows us to check (a) whether the common factors span at least one of the observable factors, and (b) to understand the nature of the panel-specific factors: portfolio-specific and individual stock-specific.

To achieve the task at hand, we regress each of the 6 observable factors on (i) the 3CF factors and (ii) all the panel-specific factors, and report the R^2s of these regression for FF3 in Figure 2.[9] For each of the three FF3 factors the figure displays the fraction of variance (R^2) explained by the common factors (blue bars which are the same in both panels), the CZ21 group-specific factors (orange bars, left panels), the CRSP group-specific factors (orange bars, right panels), and unexplained by common and group-specific factors (yellow bars).

Not surprisingly, on average (across the different rolling windows) 85% of the variability of the market factor is explained by the common factors, and specific factors of portfolio test assets tend to explain almost all the remaining part of its variability, with their R^2 ranging between 2% and 20%, depending on the time period. In the right panel of Figure 2 we observe that the factors specific to individual stocks are not able to capture the same amount of the variability unexplained by the common factors, as their R^2s are below 10%. The fact that CRSP stock-specific factors explain much less of the variability of observable factors compared to the CZ21 group-specific factors is an empirical regularity that we observe across all the FF5 and momentum factors. For instance, CZ21 group-specific factors explain between 10% and 60% of SMB while, CRSP stock-specific factors explain only 1% to 40%. Moreover, on average 50% of the variability of SMB is explained by the 3CF, and for the remaining 50% we note that between 4 and 20 % (resp. 20 to 45%) is not explained by portfolio sort specific factors (resp. individual stock specific factors). Analogous conclusions can be drawn for HML, as well as the other three factors reported in the Online Appendix.

[9]Figure OA.3 in the Online Appendix covers the other three considered so far - RMW, CMA and momentum.

Figure 2: Variability of FF3 factors explained by common and specific factors in Chen and Zimmermann (2021) test assets and CRSP individual stocks.



(a) Market - Rf : CZ21 test assets

(b) Market - Rf : CRSP stocks

(c) SMB : CZ21 test assets

(d) SMB : CRSP stocks

(e) HML : CZ21 test assets

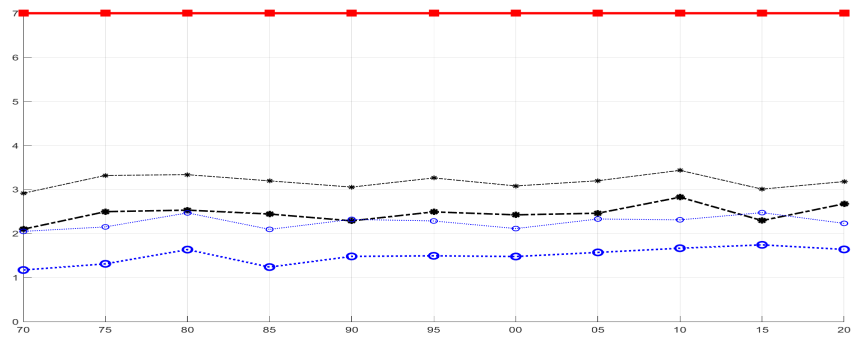
(f) HML : CRSP stocks

For each of the FF factors the figure displays the fraction of variance (R^2) explained by 3CF (blue bars which are the same in both panels), CZ21's group-specific factors (orange bars, left panels), CRSP group-specific factors (orange bars, right panels), and unexplained by common and group-specific factors (yellow bars). For each year y we report results based on the block starting in year *y-4* and ending in year y, for each y = 1970, ..., 2020.

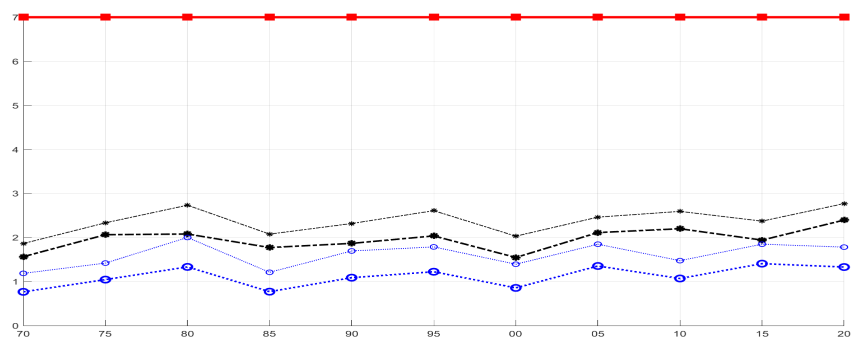
The finding implies that out of the 6 factors considered, only the market seems to be the one which is the most related to the common factors, while all the other FF and momentum factors identified in the literature based on sorting stocks on characteristics are only partially spanned by common factors, and a large part of their variability is due to a risk dimension which is specific to portfolio sorting.

We can examine the same question from a different angle, similar to what appears in Figure 1, by replacing the common factors by the group-specific factors. Recall that we started with 10 PCs in each panel and found three common factors. Therefore, we have 7 remaining group-specific factors in each panel. Figure 3 displays the sum of the canonical correlations of the 7 group-specific factors (Panel (a) the CZ21- and Panel (b) CRSP-specific) with FF3 (blue circles, thick dotted line), FF5 (black stars, thick dashed line), FF3 factors and momentum (blue circles, thin dotted line), and FF5 factors and momentum (black stars, thin dashed line). Recall that the group-specific factors reflect (1)

Figure 3: Sum of canonical correlations of three group-specific factors in Chen and Zimmermann (2021) test assets and CRSP stocks with Fama and French 3, 5 factors and momentum.



(a) CZ21 test assets



(b) CRSP individual stocks

Panel (a) displays the sum of the canonical correlations of the seven specific factors in CZ21 test asset portfolios with: (i) FF3 (blue circles, thick dotted line): (ii) FF5 (black stars, thick dashed line); (iii) the FF3 factors and momentum (blue circles, thin dotted line); (iv) the FF5 factors and momentum (black stars, thin dashed line). Panel (b) displays the sum of the canonical correlations of the seven specific factors in CRSP individual stocks with the same four sets of observable factors. For each year y we report results computed on the rolling window starting in year $y-4$ and ending in year y, for each $y = 1970, ..., 2020$.

spurious factors due to time variation in betas and (2) spurious factors due to characteristic-sorting. Panel (a) shows that FF factors with or without momentum relate to the CZ21-specific factors. This finding is perhaps not surprising, since the FF factors are constructed by sorting. The sum of canonical correlations is equal to one for FF3, two for FF3 plus momentum and FF5 and finally the sum equals three for FF5 plus momentum. If we combine the findings in Figures 1 and 3 it appears that one linear combination of FF3 (putting most of the weight on the market) is perfectly correlate with one common factor and another linear combination of FF3 perfectly correlates with a sorting-specific factor (Panels (c) through (f) in Figure 3 suggest this is a combination of SMB and HML). Similar arguments can be made for the others FF and momentum configurations.

The finding in Panel (b) may appear a bit as more surprising, showing that there is also some correlation – although not as strong as in Panel (a) – with factors specific to individual stocks pertaining to time-varying betas. It is worth recalling, however, that the beta dynamics are often instrumented via characteristics also used for sorting (see the many papers on the topic starting with Ferson and Harvey (1991) up to the recent applications of Instrumented Principal Component Analysis of Kelly, Pruitt, and Su (2019)), and the fact that in our model group-specific factors are also allowed to be mutually, although not perfectly, correlated (through the covariance matrix Φ in equation (2.9)).

4 Asset Pricing Performance of Common Factors

How do the common factors perform as predictors? How does their performance compare with the widely used factors in the asset pricing literature? These are questions we address in this section. In a first subsection we characterize the empirical models used in the forecasting evaluation, followed by subsections covering in-sample and out-of-sample empirical results.

4.1 Empirical models

We model the excess returns of CRSP individual stocks and test asset sorted portfolios as linear functions of different sets of K factors. In particular, we consider the following sets of factors:

(i) *FF + mom*: Under this header we have a set of models starting with the market factor only, defined as the value-weighted index of all CRSP stocks minus the risk-free ($K = 1$); FF3 and FF5 only ($K = 3, 5$); FF3 + Momentum, FF5 + Momentum, ($K = 4, 6$);

(ii) 3CF: $K = k^c = 3$ common factors;

(iii) 3CF + *CRSP-spec.*: $k^c = 3$ common factors and $k^s_1 = 1, 2, 3$ group-specific factors from the panel of CRSP stocks ($K = 4, 5, 6$);

(iv) 3CF + *CZ21-spec.*: $k^c = 3$ common factors and $k^s_2 = 1, 2, 3$ group-specific factors from the panel of CZ21 test assets portfolios ($K = 4, 5, 6$);

(v) *PCA on CRSP*: factors estimated as the first $K = 1, 3, 4, 5, 6$ PCs on CRSP individual stocks;

(vi) *PCA on CZ21*: factors estimated as the first $K = 1, 3, 4, 5, 6$ PCs on CZ21 test assets portfolios.[10]

[10]Note that are our CZ21 test-asset portfolios are a special type (i.e. long-only) of "managed portfolios" using the

All the models, factors and betas/loadings are estimated in each 5-years block b using the data available in that block only. Therefore, loadings/betas are constant for all the dates τ in each block b, but are allowed to change in the B different blocks. Let $y_{j,i,\tau}$ be the excess return in month τ belonging to block b of the i-th asset in group j, with $j = 1$ corresponding to individual stocks, and $j = 2$ to CZ21 test asset portfolios. Each model m for $y_{j,i,\tau}$ can be expressed as

$$y_{j,i,\tau} = \beta^{m\,\prime}_{j,i,b}\, f^m_{j,\tau} + \varepsilon^m_{j,i,\tau}\,, \qquad \text{with} \quad \tau \in b\,, \qquad j = 1,2 \tag{4.13}$$

where $f^m_{j,\tau}$ = $[f^{m,c\prime}_{\tau}, f^{j,m,s\prime}_{\tau}]'$, $\beta^m_{j,i,b} = [\lambda^{m,c\prime}_{j,i,b}\ ,\ \lambda^{m,s\prime}_{j,i,b}]'$, while $f^{m,c}_{\tau}$ and $\lambda^{m,c\prime}_{j,i,b}$ (resp. $f^{m,s}_{j,\tau}$ and $\lambda^{m,s\prime}_{j,i,b}$) are the common (resp. group-specific) factors and betas/loadings. For all models in (*i*), the factors are observable, while for all the remaining models (*ii*) - (*vi*) the factors are latent and need to be estimated either using the procedure for group-factor models described in Section 2.2 and detailed in Appendix A.5 (models (*ii*) - (*iv*)), or by performing PCA performed in only one panel of excess returns (models (*v*) - (*vi*)). In models (*i*), (*ii*), (*v*) and (*vi*) we have $\lambda^{m,s}_{j,i,b} = 0$ and $K = k^c$, as all the factors are assumed to be common across the two groups of assets.

4.2 Performance evaluation measures

We describe the in-sample and out-of-sample performance evaluation measures. The technical details appear in Online Appendix Section OA.4.

In-sample performance evaluation

We compute the following performance measures across the entire sample, that is across all B blocks:

- *Total* R^2 of Kelly et al. (2019) which represents the fraction of return variance for all the assets explained by both the dynamic behavior of the loadings and the contemporaneous factor realizations across different blocks, aggregated over all assets and all time periods.
- *Predictive* R^2 from Kelly et al. (2019) which represents the fraction of realized return variation

terminology of Kelly et al. (2019), and to the extent that PCs computed on our test-asset portfolios are similar to PCs computed on manged portfolios constructed in their way, the factors computed for our model (*vi*) are similar to the first step estimation of their (Instrumented) Principal Components.

explained by the model's description of conditional expected returns, and summarizes the model's ability to describe risk compensation only through exposure to systematic risk.

- *Pricing error* R^2 of Kelly, Palhares, and Pruitt (2020) which pertains to the fraction of the squared unconditional mean excess returns that is described by factors and betas.

In the Online Appendix we also report results using *Average* RMS_α, an alternative to the *Pricing error* R^2, which is computed as the average over different blocks of the RMS_α measure considered by Lettau and Pelger (2020a) and computed block by block.[11]

Out-of-sample performance evaluation

We implement the out-of-sample version of the $Total\ R^2$, $Pricing\ R^2$ and $Predictive\ R^2$ with betas and factor loadings computed using information from block $b-1$ to price date τ assets in block b. Analogously to Lettau and Pelger (2020b) we also compute the annualized Sharpe Ratio of the "Maximum Sharpe-ratio portfolio" that can be obtained by an optimal (in a mean-variance sense) linear combination of the factors, which are ultimately portfolios of individual stocks. The out-of-sample performance measures are defined as: (a) ***OOS Total*** R^2, (b) ***OOS Pricing*** R^2, (c) ***OOS Predictive*** R^2, and (d) ***Maximum Sharpe-ratio***, $Max.\ SR$.

4.3 Empirical results

The goal is to compare the role of the factor model specifications in explaining the variation of returns for individual CRSP stocks as well as CZ21 test assets, both in- and out-of-sample, during the period 1966-2020. Panels A - C in Table 1 present respectively the Total, Pricing and Predictive R^2 evaluation measures. The first two rows in each panel pertain to the Total/Pricing/Predictive R^2s of the benchmark models which consist of the FF and momentum factors, starting with the one-factor market (CAPM), then FF3, FF3 with momentum, FF5 and finally FF5 and momentum. The columns are therefore labeled 1, 3, 4, 5 and 6 corresponding to the number of factors K in each model. For comparison, the next two rows refer to the corresponding R^2s of the three common factors (Comm). These are followed by the models which consider both the three common factors as well as the panel-specific

[11]Table OA.1 in the Online Appendix covers Average RMS_α and an alternative version of Total R^2 computed from regressions with constant similar to Lettau and Pelger (2020b).

Table 1: In- and Out-of-sample performamce evaluation factor models

	In-Sample					Out-of-Sample				
N. of factors, K	1	3	4	5	6	1	3	4	5	6
Panel A: Total R^2										
r: CRSP, f: FF + mom	14.3	22.8	25.2	26.4	28.6	8.5	7.3	4.0	1.5	< 0
r: CZ21, f: FF + mom	73.9	89.8	91.5	90.7	92.3	70.8	83.6	84.7	83.6	84.8
r: CRSP, f: 3CF		27.0					13.7			
r: CZ21, f: 3CF		92.8					86.8			
r: CRSP, f: 3CF + CRSP spec.			30.1	32.9	35.4			14.0	14.3	14.4
r: CZ21, f: 3CF + CZ21 spec.			94.3	95.2	95.8			89.3	90.8	91.4
r: CRSP, f: PCA on CZ21	18.0	25.4	27.7	29.8	32.2	12.0	10.6	9.8	8.8	8.2
r: CZ21, f: PCA on CZ21	89.3	94.5	95.3	95.8	96.2	87.3	91.9	92.5	92.9	93.1
r: CRSP, f: PCA on CRSP	17.3	27.0	30.1	32.9	35.4	12.5	13.8	14.1	14.4	14.6
r: CZ21, f: PCA on CRSP	86.5	92.3	92.8	93.5	93.9	83.7	86.0	86.4	86.5	86.6
Panel B: Pricing R^2										
r: CRSP, f: FF + mom	33.7	50.9	54.2	52.7	55.5	4.0	9.3	11.1	11.5	11.6
r: CZ21, f: FF + mom	75.2	89.0	88.4	90.0	89.7	78.8	85.5	84.4	86.7	86.4
r: CRSP, f: 3CF		45.4					19.5			
r: CZ21, f: 3CF		91.8					91.4			
r: CRSP, f: 3CF + CRSP spec.			45.8	46.0	46.2			18.7	19.5	21.8
r: CZ21, f: 3CF + CZ21 spec.			93.3	93.6	95.9			92.3	92.5	94.0
r: CRSP, f: PCA on CZ21	51.6	59.1	61.3	62.7	64.3	15.7	11.5	11.1	11.3	11.5
r: CZ21, f: PCA on CZ21	89.7	94.9	96.1	97.3	97.9	90.9	93.5	94.6	95.4	95.8
r: CRSP, f: PCA on CRSP	50.0	55.4	56.3	57.9	58.8	15.2	18.6	19.4	21.9	18.8
r: CZ21, f: PCA on CRSP	91.0	92.3	92.4	92.6	92.7	91.3	91.8	92.1	92.0	92.0
Panel C: Predictive R^2										
r: CRSP, f: FF + mom	0.66	1.09	1.20	1.17	1.26	0.02	0.03	< 0	< 0	< 0
r: CZ21, f: FF + mom	2.60	4.00	4.03	4.07	4.12	0.05	0.96	1.05	0.97	1.05
r: CRSP, f: 3CF		0.99					0.24			
r: CZ21, f: 3CF		4.17					2.02			
r: CRSP, f: 3CF + CRSP spec.			1.02	1.02	1.03			0.21	0.20	0.19
r: CZ21, f: 3CF + CZ21 spec.			4.28	4.30	4.40			1.84	1.73	1.76
r: CRSP, f: PCA on CZ21	1.10	1.31	1.37	1.42	1.46	0.19	0.10	0.02	< 0	< 0
r: CZ21, f: PCA on CZ21	4.03	4.37	4.43	4.50	4.54	1.52	1.63	1.65	1.67	1.69
r: CRSP, f: PCA on CRSP	1.06	1.21	1.25	1.29	1.31	0.31	0.28	0.25	0.23	0.22
r: CZ21, f: PCA on CRSP	4.07	4.19	4.20	4.23	4.23	2.07	2.13	2.12	2.12	2.11

Panels A - C of the table report Total/Pricing and Predictive R^2s in percent for observable factor models (lines 1-2 in each panel), 3CF a latent factor model with 3 factors common between individual stocks and CZ21 portfolios (lines 3-4 in each panel), the same 3 common factors together with 1, 2, or 3 CRSP-specific factors (line 5 in each panel), again the same 3 common factors together with 1, 2, or 3 CZ21-specific factors (line 6 in each panel), a latent factor model where the factors are K PCs extracted from the CZ21 portfolios only (lines 7-8 in each panel), and a latent factor model where the factors are K PCs extracted from the CRSP individual stocks only (lines 9-10). Observable factor model specifications are CAPM, FF3, FF3 + Momentum, FF5, and FF5 + Momentum in the $K = 1$,3 ,4 ,5 ,6 columns, respectively. The models are estimated on the rolling window starting in year $y - 4$ and ending in year y, for each $y = 1970, ..., 2020$. $Total\ R^2$'s in-sample (left table) and out-of sample (right table) are computed either for the excess returns of individual stocks (r : CRSP) or CZ21 portfolios (r : CZ21) as described in Section 4.2.

factors (individual CRSP stocks and CZ21 test assets) in order to evaluate whether the latter have additional explanatory power for the variation of returns beyond the three common factors. Finally, the last four rows in each panel of Table 1 pertain to the R^2s for models with factors based on the PCs from each of the two panels as well as the PCs across the two panels (namely PCs from CRSP are used to price CZ21 assets and vice versa). Last but not least, Table 2 displays the maximum Sharpe ratio portfolios.

From the in-sample analysis reported in Table 1 we can draw the following two important observations. First, the three common factors typically yield better or comparable in-sample Total, Pricing and Predictive R^2s vis-à-vis the benchmark models (i.e. CAPM K =1, FF3 K = 3, FF3 plus momentum, K = 4, FF5 K = 5 and FF5 plus momentum K = 6). There are number of cases where the traditional models with $K > 3$ do better in-sample. Second, adding the corresponding group-specific factors from the two panels (of individual stock and test assets) leads only to marginal improvements compared to models with the three common factors.

Of greater interest are the out-of-sample results in Table 1. They yield the following key empirical findings:

- the three common factors yield the highest OOS Total, Pricing and Predictive R^2s compared to *any* FF (plus momentum) benchmark model (with up to K = 6). The relative gains of the OOS Total, Pricing and Predictive R^2s are the largest for individual stocks

- adding to 3CF the corresponding panel-specific factors (from individual stock and CZ21 assets) leads sometimes to only marginal improvements

- PCA on CZ21 yields better results than 3CF for OOS Total and Pricing (but not Predictive) R^2s for CZ21 returns. However, those factors poorly predict individual stocks out-of-sample according to the three types of R^2s considered

- PCA on CRSP is comparable to 3CF, both for CZ21 and CRSP returns, but under-perform when panel-specific factors are added to the common ones (with the exception of OOS Predictive R^2s)

Last but not least, Table 2 presents the out-of-sample annualized maximum Sharpe Ratio (SR) for the different factor model specifications. Interestingly, we find that the three common factors perform

Table 2: Out-of-sample factor portfolio Sharpe ratios

N. of factors, K	1	3	4	5	6
FF + mom	0.39	0.30	0.62	0.63	0.81
3CF		0.58			
3CF + CZ21 spec.			0.47	0.40	0.67
3CF + CRSP spec.			0.50	0.51	0.47
PCA on CZ21	0.49	0.54	0.81	0.98	0.84
PCA on CRSP	0.61	0.49	0.51	0.47	0.43

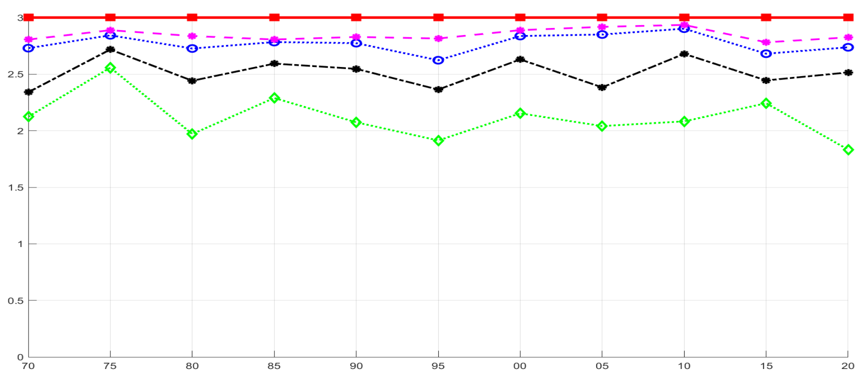
The table reports out-of-sample annualized Sharpe ratios for the mean-variance efficient portfolio of factors in each model - see caption Table 1 for details.

relatively better producing a SR of 0.58, vis-à-vis the CAPM and FF3 models which have SRs of 0.30 and 0.39, respectively. Nevertheless, the FF5 factor model plus momentum (K = 6) produces the highest SR of 0.81 among the traditional benchmark models. If we limit ourselves to K = 3, then the common factor model outperforms all other specifications in Table 2. Adding factors beyond K = 3 does increase SR, however, and the best model is PCA on CZ21. Moreover, PCA on CRSP, which according the performance measures reported in Table 1 is similar to common factors in pricing both panels of excess returns, features among the smallest maximum Sharpe Ratios. Overall, it is worth highlighting that the 3CF model does better than the strongest competitors appearing in Table 1.

5 Revisiting the Factor Zoo

The factor zoo is represented by all the factors collected by CZ21. As detailed in Online Appendix Section OA.1, when we refer to the factor zoo we use a data set of over a thousand portfolios associated with 205 characteristics. In this section we investigate the ability of PCs extracted form the factor zoo, which we call zoo PCs, to price and explain the variability of the panels of individual stocks and the CZ21 portfolios. In addition, we study the relationship between factors in the zoo and (a) the 3CF factors and (b) the CZ21-specific factors. Finally, we address the question whether the new factors entering in the zoo in a certain year provide additional information relative to the set of previously published factors.

Figure 4: Number of common factors between Chen and Zimmermann (2021)'s test assets and CRSP stocks, and sum of common factors canonical correlations with 3, 5, 10 and 15 PCs from the zoo.



The figure displays the sum of the canonical correlations of 3CF factors with the first 15 PCs (magenta circles, dashed line), 10 PCs (blue circles, dotted thin line), 5 PCs (black stars, dashed thin line), and finally the first 3 PCs form the factor zoo (green diamonds, dotted thin line). For each year y we report results computed on the block starting in year $y-4$, for each $y = 1970, 1975, ..., 2020$.

5.1 Common Factors and the Zoo

Figure 4 shows the sum of the canonical correlations of the three common factors with the first 3, 5, 10 and 15 zoo PCs. As before, all PCs and common factors are estimated from non-overlapping 5-year balanced panels of monthly data over the period 1966-2020. The figure allows us to understand whether the PCs from the zoo panel span the space of the common factors. The first observation emerging from Figure 4 is that 3 zoo PCs yield a sum of canonical correlations roughly equal to 2 as if there is constantly a missing factor. Going to 5 zoo PCs gets us to 2.5 and it takes up to 10 PCs from the factor zoo to approximately span the set of 3 common factors.

Table 3 documents which factors from the zoo are the most related (a) to 3CF, and (b) to 3CF augmented with the first three group-specific factors of the CZ21 portfolios. Panel A reports the twenty factors with the largest average R^2s - across all our 5 non-overlapping windows - when regressed on the 3CF factors. Among them, we find two of the three Fama-French factors (CAPM Beta and Size) along with portfolios based on market beta put forward by Frazzini and Pedersen (2014), and different measures of idiosyncratic risk and liquidity or uncertainty, such as Bid-ask Spread, Cash-flow to price variance, Volume to market equity, EPS Forecast Dispersion, Days with zero trades, Volume Variance, and Price delay R-square. Turning to Panel B, we report the factors in the zoo showing the highest increase in the R^2 - averaged across all non-overlapping 5 years windows - when the first 3 CZ21-

Table 3: Variability of the factors in the zoo explained by the 3 common factors, and 3 first three group-specific factors from Chen and Zimmermann (2021)

Panel A: 3 Common factors only		**Panel B: 3 CZ21-specif. added to 3 common factors**	
Factor	R^2	**Factor**	ΔR^2
CAPM beta (1973)	92.8	Cash Productivity (2009)	39.2
Frazzini-Pedersen Beta (2014)	87.8	Intangible return using BM (2006)	38.0
Bid-ask spread (1986)	83.6	Book to market using most recent ME (1985)	37.3
Cash-flow to price variance (1996)	80.7	Total assets to market (1992)	36.5
Volume to market equity (1996)	79.9	Market leverage (1988)	35.2
Price (1972)	79.9	Off season long-term reversal (2008)	34.7
Idiosyncratic risk (AHT) (2003)	78.0	Intangible return using CFtoP (2006)	34.6
Idiosyncratic risk (3 factor) (2006)	72.9	Sales-to-price (1996)	32.8
EPS Forecast Dispersion (2002)	71.8	Option to stock volume (2012)	32.7
Tail risk beta (2014)	71.8	Book to market using December ME (1992)	32.0
Days with zero trades (2006)	71.7	Efficient frontier index (2009)	31.3
Idiosyncratic risk (2006)	71.3	Momentum without the seasonal part (2008)	31.1
52 week high (2004)	70.9	Past trading volume (1998)	30.5
Days with zero trades (2006)	69.1	Initial Public Offerings (1991)	29.9
Size (1981)	68.7	Change in current operating liabilities (2005)	29.6
Days with zero trades (2006)	67.7	Change in equity to assets (2005)	29.5
Maximum return over month (2010)	67.1	Momentum (12 month) (1993)	29.0
Volume Variance (2001)	66.7	Intangible return using Sale2P (2006)	29.0
Analyst earnings per share (2006)	65.1	Employment growth (2014)	28.9
Price delay R-square (2005)	64.5	Intangible return using EP (2006)	28.8

We regress each of the factors in the zoo present in the 5-years rolling window ending in y on (a) the 3CF factors, and (b) 3CF + first 3 group-specific factors in CZ21 test assets. Factors names correspond to those in the Online Appendix of CZ21. For each factor, we compute the average across all years of the R^2 of these regressions. Panel A reports the 20 factors with the largest average R^2 for regression (a), i.e. the most related to the three common factors. Panel B reports the sorted 20 factors with the highest average increase in R^2 for regression (b) when added to the 3 common factors in regression (a), that is the factors which are most related to the first three CZ21-specific factors. Factor names correspond to those in the Online Appendix of CZ21. We consider years $y = 1970, 1975, ..., 2020$.

specific factors are added as regressors to 3CF. Interestingly, the majority of the factors which are the most related to the three CZ21 group-specific factors are associated with Book-to-Market (i.e. Total assets to market, and alternative ways to compute this ratio) or other valuation factors (i.e. Sales-to-Price), Momentum (i.e. Momentum 12 months, and without seasonal part), Long-Term reversal (i.e. Off-season long term reversal, and Intangibles using BM, or CFtoP, or EP) and investment (i.e. Change in Equity-to-assets, and Employment growth)

These findings complement the results of Figure 2 (and Figure OA.3) confirming the market and size are the factors most correlated with 3CF (for Size this is especially true for the first half of our sample), while the majority of the variability of book-to-market and momentum is mostly explained by CZ21-specific factors (i.e. are due to sorting).

Table 4 reports the performance evaluation measures for the zoo PCs (compared to Table 1 we added a column for K = 10 using the insights gained from Figure 4). For convenience of comparison

Table 4: Total, Pricing and Predictive R^2s - Common factors versus zoo PCs

	In-Sample						Out-of-Sample					
N. of factors, K	1	3	4	5	6	10	1	3	4	5	6	10
Panel A: Total R^2												
r: CRSP, f: 3CF		27.0						13.7				
r: CZ21, f: 3CF		92.8						86.8				
r: CRSP, f: PCA on Zoo	12.7	21.8	24.7	27.1	29.6	37.2	3.6	2.6	2.8	2.3	1.8	< 0
r: CZ21, f: PCA on Zoo	42.2	68.6	72.6	75.6	77.7	83.0	29.4	49.0	53.5	56.0	59.0	57.7
Panel B: Pricing R^2												
r: CRSP, f: 3CF		45.4						19.5				
r: CZ21, f: 3CF		91.8						91.4				
r: CRSP, f: PCA on Zoo	< 0	< 0	12.6	23.9	24.9	38.5	3.2	3.9	4.4	< 0	< 0	< 0
r: CZ21, f: PCA on Zoo	< 0	< 0	< 0	9.2	7.7	28.6	< 0	< 0	< 0	< 0	< 0	< 0
Panel c: Predictive R^2												
r: CRSP, f: 3CF		0.99						0.24				
r: CZ21, f: 3CF		4.17						2.02				
r: CRSP, f: PCA on Zoo	< 0	0.13	0.36	0.60	0.63	0.95	< 0	< 0	< 0	< 0	< 0	< 0
r: CZ21, f: PCA on Zoo	< 0	< 0	0.07	0.85	1.04	1.99	< 0	< 0	< 0	< 0	< 0	< 0

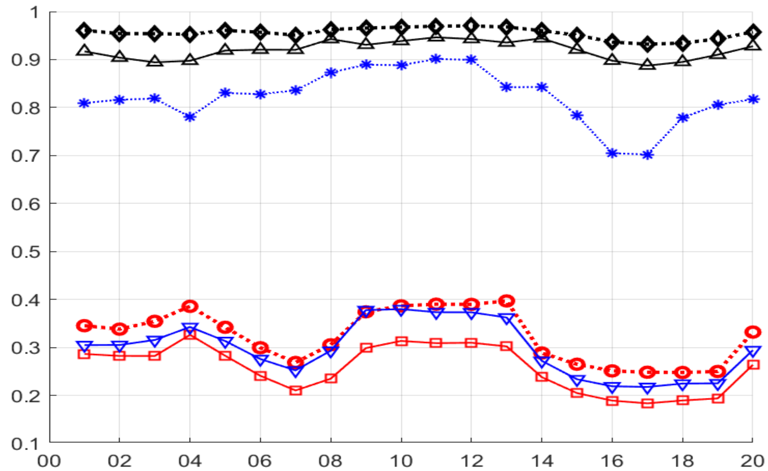
The table reports Total, Pricing and Predictive R^2s in percent for a latent factor model with only 3 common factors (lines 1-2 in each of the three panels) - a repeat of lines 3-4 in Panels A - C of Table 1, and a latent factor model where the factors are K PCs extracted from the factors in the zoo only (lines 3-4). The models are estimated on the rolling window starting in year $y-4$ and ending in year y, for each $y = 1970, ..., 2020$. R^2's in-sample (left table) and out-of sample (right table) are computed either for the excess returns of individual stocks (r : CRSP) or CZ21 portfolios (r : CZ21) as described in Section 4.2.

we repeat the results for 3CF from Panels A - C in Table 1. The zoo PCs have positive in- and out-of sample Total R^2s for individual stocks better than the observable factors, but worse than all the other latent factor models (comparing with results in Table 1). Moreover, they perform worse than the other set of factors in explaining CZ21 portfolio returns. The pricing performance of the zoo PCs is also the worst, as evident from their low or negative Pricing R^2 appearing in Table 4.

5.2 Old and New Factors

We can also address the question whether the new factors entering the zoo in a certain year provide additional information relative to the previously published factors. So far we used a "chronological time" sample which include all data available in each data set from Jan. 1966 to Dec. 2020. Here we consider a "publication time" sample which goes from Jan. 1996 to Dec. 2020, where the CZ21

Figure 5: $Total\ R^2$ from old factors



For each 5-years rolling window ending in year y we compute the percentage $TotalR^2$ generated by a linear factor model with 3CF only (model (i)), with 3CF + three CZ21-specific factors (model (ii)), and by linear factor model where the factors are the first six PCs from the old factor zoo (model(iii)). $TotalR^2$ is computed using as test assets either the individual stocks or the CZ21 portfolios available in year y for both model (i) and model (ii). When test assets are individual stocks we report the $TotalR^2$ for model (i) as red squares, for model (ii) as red circles and for model (iii) as blue downward triangles. When test assets are CZ21 test assets we report the $TotalR^2$ for model (i) as black upward triangles, for model (ii) as black diamonds and for model (iii) as blue stars. The models are estimated on the rolling window starting in year $y-4$ and ending in year y, for each $y = 2001, \ldots, 2020$. $Total\ R^2$'s are computed as described in Section 4.2, but taking into account only the 5-year window ending in year y.

test assets portfolios and factors enter with their publication date in the database. So far we used non-overlapping block samples. To do the analysis here we proceed on an annual basis instead and use a 5-years rolling sample scheme which allows us to examine the so called "new" factors being introduced every year versus the pre-existing factors, called the "old" factors.

For each 5-years rolling window ending in year y, we define the old CZ21 portfolios as the those corresponding to the factors in the zoo available in the 5-years rolling window ending in year $y-1$ to distinguish them from the new CZ21 portfolios, that is those corresponding to the new factors entering in the database in year y according to their publication date. In every 5-years window, we can compute common and panel-specific factors as we did in the prior sections. Then, we regress each of the new factors on the old three common factors and first three old CZ21-specific factors.

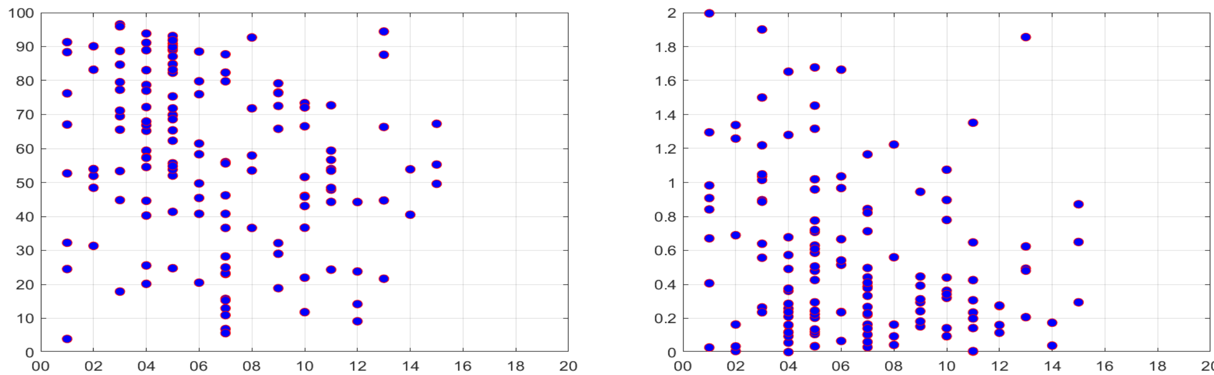
In Figure 5 we report the $TotalR^2$ generated by linear factor models involving the factors only form the old zoo. In particular, we consider a model with 3CF only (model (i)), another model with 3CF augmented with three CZ21-specific factors (model (ii)), and a final linear factor model where the factors are the first 6 PCs from the old factor zoo (model(iii)). As before, the models are estimated on 5-years rolling windows starting in year $y - 4$ and ending in year y, for each $y = 2001, ..., 2020$. We note that although there is some time variation, the $TotalR^2$ of models (i) and (ii) for both set of test assets remained relatively stable in our sample, indicating that the old factors tend to explain a relatively constant fraction of the time series variation of test assets excess returns. Interestingly though, all the $TotalR^2$s tend to be higher during rolling windows including one of the three stress periods in our sample, namely: the 2001 dot-com bubble, the 2008 financial crisis and the 2020 COVID pandemic.

As expected, the models explain a much larger fraction of the variability of CZ21 portfolios, as measured by $TotalR^2$, than of individual stocks. Finally, we note that for CZ21 portfolios the 6 PCs from the old zoo (model (iii)) always produce a lower $TotalR^2$ compared to models (i) and (ii). For individual stocks the 6 PCs form the old zoo always produce a lower $TotalR^2$ than the 3CF plus the 3 CZ21-specific factors, but comparable to a model with 3CF only. Similar conclusions can be drawn when looking at the pricing performance of the same sets of factors for the panels of individual stocks and sorted portfolios, as can be seen from Figure OA.8 in the OA, where we plot the Root Mean Squared α" (RMS_{α}), which is an alternative way to assess the pricing performance for each model (see also the Pricing R^2s defined in Appendix OA.4).

Next we turn to Figure 6 where for each 5-years rolling window ending in year y we regress each of the new factors (entering the database in year y) on old 3CF (i.e. year $y - 1$) plus first three old CZ21-specific factors. Figure 6 (a) displays the R^2's for each of these regressions, while Figure 6 (b) displays the absolute value of the intercepts ($|\alpha|$). The new factors with low R^2s and those with high $|\alpha|$ and/or significant t-stat provide additional information to the space already spanned by the 'old' factors. The detailed results are reported in Online Appendix Tables OA.2 and OA.3 respectively. We find that 4 factors have R^2s smaller or equal than 10%, and 7 R^2s $\leq 15\%$. Most are related to seasonality (2 of the 4 at 10% and 4 out of 7 at 15%).[12] Next, looking at the alphas we consider three cases: (a) $|\alpha| > 1.5\%$, (b) estimates with t-stats above 2, and (c) a combination of both (a) and (b). The exercise is somewhat similar to Kozak, Nagel, and Santosh (2018) in their Section II, where they check whether factors have significant alphas with respect to PCs computed from the factors themselves. There are 7

[12] According to the results appearing in Online Appendix Table OA.2 the factors with R^2s below 10% in chronological order are: Up Forecast (2002), Off season reversal years 16 to 20 (2008), Return seasonality years 6 to 10 (2008), and R&D ability (2013). In addition for the higher threshold of 15% to following are included: Return seasonality years 16 to 20 (2008), Put volatility minus call volatility (2011), and Dividend seasonality (2013).

Figure 6: Regressions of new factors entering the zoo on 3CF and 3 first three CZ21-specific factors



(a) R_i^2 (b) $|\alpha_i|$

For each 5-years rolling window ending in year y we regress each of the new factors (entering the database in year y) on old 3CF (i.e. year $y-1$) plus first three old CZ21-specific factors. Figure (a) displays the R^2's for each of these regressions, while Figure (b) displays the absolute value of the intercepts. We consider years $y = 2001, 2002, ..., 2020$.

factors with $|\alpha| > 1.5\%$. Several among them have insignificant estimates. Counting the alphas that have t-stats above 2 regardless of the magnitude, we count 11.[13] There are only a handful of factors with high alphas and significant t-stats: Consensus Recommendation (2002), Net Operating Assets (2004), Analyst earnings per share (2006), Net external financing (2006), Industry return of big firms (2007) and Frazzini-Pedersen Beta (2014).

Our analysis is different from the previous literature as we use the factors common between CZ21 and CRSP as the reference model. Section OA.5 reports extensive additional results showing that the increments in $TotalR^2$ and RMS_α generated by the addition of new factors with respect to the 6 old common and CZ21-specific factors is relatively small. So the conclusion one can draw is that new factors in the zoo seem only to improve marginally the ability of a model including the 6 old factors in pricing, and explaining the time series variability, of the two large sets of test assets we consider. Moreover, no factor seems to both generate an incremental contribution to the 6 old factors in pricing

[13]According to the results in Online Appendix Table OA.3 the factors with $|\alpha| > 1.5\%$ are in chronological order: Consensus Recommendation (2002), Firm Age - Momentum (2004), Net Operating Assets (2004), Institutional ownership among high short interest (2005), Analyst earnings per share (2006), Industry return of big firms (2007) and finally: Frazzini-Pedersen Beta (2014) and those with t-stats above 2 are: Consensus Recommendation (2002), Probability of Informed Trading (2002), Pastor-Stambaugh liquidity beta (2003), Net Operating Assets (2004), Mohanram G-score (2005), Analyst earnings per share (2006), Net equity financing (2006), Net external financing (2006), Industry return of big firms (2007), Efficient frontier index (2009), Intermediate Momentum (2012), and Frazzini-Pedersen Beta (2014).

of both sets of test assets, and in explaining their time series variation.[14] These findings are compatible with the idea that newly proposed factors might help to price and explain the time series variability of few portfolios built on the same characteristics of the factors themselves, but might not be relevant in explaining the returns of many other portfolios sorted on different characteristics, or all individual stocks.

It is worth comparing our findings with those in Table 2 of Feng et al. (2020). Although similar in spirit to the analysis considered here, their approach is based on a different approach. Namely, their procedure combines the double-selection LASSO method of Belloni, Chernozhukov, and Hansen (2014) with the Fama and MacBeth (1973) two-pass regressions to evaluate the pricing contribution of a factor in a high-dimensional setting and identifies a selection of factors which *contains* factors with incremental value. By contain we mean that their procedure potentially identifies a selection of factors that is larger than the set with genuine pricing contributions. The results in their Table 2 show that while most of the new factors are redundant relative to the existing factors, a few have statistically significant explanatory power beyond the hundreds of factors proposed in the past. In this respect our results agree with their findings, but the specifics are somewhat different. They have a total of twelve out of roughly one hundred factors (a smaller set than what we considered here) over the sample 2000 until 2015 which appear significant according to their testing procedure. There is one factor in common identified by our approach and theirs: betting against beta from Frazzini and Pedersen (2014). No other factors identified by the Feng et al. (2020) procedure have an R^2 lower than 50%.

6 Conclusions

The projection arguments put forward in Hansen and Jagannathan (1991) imply, as noted by Kozak et al. (2018), that there exists a factor representation of the stochastic discount factor (SDF). Moreover, there is practically no disagreement that the space of factors spanning the SDF is low-dimensional. In this paper we found 3 factors which were selected via a novel procedure addressing a longstanding

[14] A closer inspection of Table OA.5 shows that Frazzini-Pedersen Beta (2014), Change in net financial assets (2005), Equity Duration (2004), Operating Cash flows to price (2004), and 52 week high (2004) are among the top improves of both $TotalR^2$ and RSM_α for the panel of of CZ21 portfolios when added to the old factors, nevertheless none of them appears a a top 10 contributor for the same two measures computed for the panel of individual stocks. Actually, Frazzini-Pedersen Beta (2014) seems to be detrimental to the pricing of individual stock returns when added to the old factors.

debate in the empirical asset pricing literature.

We started from the idea that equity risk factors reside at the intersection of two panels: (a) individual stock returns and (b) sorted portfolio returns. We extract factors from both panels and find the common factor space between the two panels, yielding factors which price both individual stocks and sorted portfolios. We labeled these three common factors 3CF. We show that this provides a path toward extracting factors neither affected by sorting characteristics nor by varying risk exposures and recalcitrant features of individual stocks.

We also find that *at any point during our sample* neither FF3 nor FF5, both with or without a momentum factor, span the 3CF factor space. In fact we also find that out of the 6 factors considered, only the market seems to be the one which is the most related to the common factors, while all the other 5 factors, are only partially spanned by common factors, and a large part of their variability is specific to portfolio sorting.

Regarding the factor zoo we find that over the sample period 1996-2020 it takes 10 PCs from the factor zoo panel to span the set of common factors. Moreover, we also address the question whether the new factors entering in the zoo in a given year provide additional information relative to the previously published factors. We find that new factors being added to the zoo seem only to improve marginally the empirical performance compared to existing factors.

Last but not least, it should be noted that the testing procedure introduced in our paper can be applied in many other asset pricing settings. A few examples are: comparing panels of private equity and publicly traded companies, international asset pricing comparing stock in different countries, etc.

References

ANDERSON, C. W. AND L. GARCIA-FEIJOO (2006): "Empirical evidence on capital investment, growth options, and security returns," *Journal of Finance*, 61, 171–194.

ANDREOU, E., P. GAGLIARDINI, E. GHYSELS, AND M. RUBIN (2019): "Inference in Group Factor Models with an Application to Mixed-frequency Data," *Econometrica*, 87, 1267–1305.

——— (2020): "Mixed Frequency Macro-Finance Factor Models: Theory and Applications," *Journal of Financial Econometrics*, 18, 585–628.

——— (2021): "Spanning Latent and Observable Factors," Discussion Paper EDHEC, UCY, USI and UNC.

ASNESS, C. S., R. B. PORTER, AND R. L. STEVENS (2000): "Predicting stock returns using industry-relative firm characteristics," Available at SSRN 213872.

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

BAI, J. AND S. NG (2002): "Determining the Number of Factors in Approximate Factor Models," *Econometrica*, 70, 191–221.

BELLONI, A., V. CHERNOZHUKOV, AND C. HANSEN (2014): "Inference on treatment effects after selection among high-dimensional controls," *Review of Economic Studies*, 81, 608–650.

BILLINGSLEY, P. (1995): *Probability and Measure*, Wiley.

BRADSHAW, M. T., S. A. RICHARDSON, AND R. G. SLOAN (2006): "The relation between corporate financing activities, analysts' forecasts and stock returns," *Journal of Accounting and Economics*, 42, 53–85.

BREEDEN, D. T., M. R. GIBBONS, AND R. H. LITZENBERGER (1989): "Empirical test of the consumption-oriented CAPM," *Journal of Finance*, 2, 231–262.

BREITUNG, J. AND S. EICKMEIER (2011): "Testing for Structural Breaks in Dynamic Factor Models," *Journal of Econometrics*, 163, 71–84.

CHEN, A. Y. (2019): "The Limits of p-Hacking: A Thought Experiment," Available at SSRN `https://ssrn.com/abstract=3272572`.

CHEN, A. Y. AND T. ZIMMERMANN (2021): "Open Source Cross-Sectional Asset Pricing," *Working Paper*.

CHORDIA, T., A. SUBRAHMANYAM, AND V. R. ANSHUMAN (2001): "Trading activity and expected stock returns," *Journal of Financial Economics*, 59, 3–32.

CHORIDA, T., A. GOYAL, AND A. SARETTO (2020): "Anomalies and False Rejections," *Review of Financial Studies*, 35, 2134–2179.

COCHRANE, J. H. (2011): "Presidential address: Discount rates," *Journal of Finance*, 66, 1047–1108.

CONNOR, G. AND R. KORAJCZYK (1988): "Risk and return in an equilibrium APT: Application of a new test methodology," *Journal of Financial Economics*, 21, 255–289.

DAVIDSON, J. (1994): *Stochastic Limit Theory*, Oxford University Press.

FAIRFIELD, P. M., S. WHISENANT, AND T. L. YOHN (2003): "The differential persistence of accruals and cash flows for future operating income versus future profitability," *Review of Accounting Studies*, 8, 221–243.

FAMA, E. AND K. FRENCH (2006): "Profitability, Investment, and Average Returns," *Journal of Finacial Economics*, 82, 491–518.

FAMA, E. F. AND K. R. FRENCH (2015): "A five-factor asset pricing model," *Journal of Financial Economics*, 116, 1–22.

FAMA, E. F. AND J. D. MACBETH (1973): "Risk, return, and equilibrium: Empirical tests," *Journal of Political Economy*, 81, 607–636.

FENG, G., S. GIGLIO, AND D. XIU (2020): "Taming the factor zoo," *Journal of Finance*, 70, 1327–1370.

FERSON, W. E. AND C. R. HARVEY (1991): "The variation of economic risk premiums," *Journal of Political Economy*, 99, 385–415.

FRAZZINI, A. AND L. H. PEDERSEN (2014): "Betting against beta," *Journal of Financial Economics*, 111, 1–25.

GAGLIARDINI, P., E. OSSOLA, AND O. SCAILLET (2016): "Time-varying risk premium in large cross-sectional equity data sets," *Econometrica*, 84, 985–1046.

GIGLIO, S. AND D. XIU (2021): "Asset pricing with omitted factors," *Journal of Political Economy*, 129, 1947–1990.

GOYAL, A., C. PÉRIGNON, AND C. VILLA (2008): "How Common are Common Return Factors across the NYSE and Nasdaq?" *Journal of Financial Economics*, 90, 252–271.

GREEN, J., J. R. M. HAND, AND X. F. ZHANG (2017): "The characteristics that provide independent information about average us monthly stock returns," *Review of Financial Studies*, 30, 4389–4436.

HANSEN, L. P. AND R. JAGANNATHAN (1991): "Implications of security market data for models of dynamic economies," *Journal of Political Economy*, 99, 225–262.

HARVEY, C. R. AND Y. LIU (2019): "A census of the factor zoo," Available at SSRN `https://ssrn.com/abstract=3341728`.

——— (2021): "Lucky factors," *Journal of Financial Economics*, 141, 413–435.

HARVEY, C. R., Y. LIU, AND H. ZHU (2016): "... and the cross-section of expected returns," *Review of Financial Studies*, 29, 5–68.

HIRSHLEIFER, D., K. HOU, S. H. TEOH, AND Y. ZHANG (2004): "Do investors overvalue firms with bloated balance sheets?" *Journal of Accounting and Economics*, 38, 297–331.

HOU, K., C. XUE, AND L. ZHANG (2015): "Digesting anomalies: An investment approach," *Review of Financial Studies*, 28, 650–705.

——— (2020): "Replicating anomalies," *Review of Financial Studies*, 35, 2019–2133.

HUBERMAN, G., S. KANDEL, AND R. F. STAMBAUGH (1987): "Mimicking portfolios and exact arbitrage pricing," *Journal of Finance*, 1, 1–9.

JEGADEESH, N. AND J. LIVNAT (2006): "Revenue surprises and stock returns," *Journal of Accounting and Economics*, 41, 147–171.

JENSEN, M. C., F. BLACK, AND M. S. SCHOLES (1972): "The capital asset pricing model: Some empirical tests," in *Studies in the Theory of Capital Markets*, ed. by M. C. Jensen, Praeger Publishers Inc.

KAMA, I. (2009): "On the market reaction to revenue and earnings surprises," *Journal of Business Finance and Accounting*, 36, 31–50.

KELLY, B. T., D. PALHARES, AND S. PRUITT (2020): "Modeling Corporate Bond Returns," *Working Paper*.

KELLY, B. T., S. PRUITT, AND Y. SU (2019): "Characteristics are Covariances: A Unified Model of Risk and Return," *Journal of Financial Economics*, 134, 501–524.

KIM, S. AND R. KORAJCZYK (2021): "Large Sample Estimators of the Stochastic Discount Factor," *Working paper*.

KOZAK, S., S. NAGEL, AND S. SANTOSH (2018): "Interpreting factor models," *Journal of Finance*, 73, 1183–1223.

——— (2020): "Shrinking the cross-section," *Journal of Financial Economics*, 135, 271–292.

LEHMANN, B. AND D. MODEST (1988): "The Empirical Foundations of the Arbitrage Pricing Theory," *Journal of Financial Economics*, 21, 213–254.

LEHMANN, B. N. AND D. M. MODEST (2005): "Diversification and the Optimal Construction of Basis Portfolios," *Management Science*, 51, 581–598.

LETTAU, M. AND M. PELGER (2020a): "Estimating latent asset-pricing factors," *Journal of Econometrics*, 218, 1–31.

——— (2020b): "Factors That Fit the Time Series and Cross-Section of Stock Returns," *Review of Financial Studies*, 33, 2274–2325.

LEWELLEN, J., S. NAGEL, AND J. SHANKEN (2010): "A skeptical appraisal of asset pricing tests," *Journal of Financial Economics*, 96, 175–194.

LITZENBERGER, R. H. AND K. RAMASWAMY (1979): "The effect of personal taxes and dividends on capital asset prices: Theory and empirical evidence," *Journal of Financial Economics*, 7, 163–195.

MAGNUS, J. R. AND H. NEUDECKER (2007): *Matrix Differential Calculus with Applications in Statistics and Econometrics*, John Wiley and Sons: Chichester/New York.

MCLEAN, R. D. AND J. PONTIFF (2016): "Does academic research destroy stock return predictability?" *Journal of Finance*, 71, 5–32.

PUKTHUANTHONG, K., R. ROLL, AND A. SUBRAHMANYAM (2019): "A protocol for factor identification," *Review of Financial Studies*, 32, 1573–1607.

ROLL, R. AND S. A. ROSS (1980): "An Empirical Investigation of the Arbitrage Pricing Theory," *Journal of Finance*, 35, 1073–1103.

ROSS, S. A. (1976): "The arbitrage theory of capital asset pricing," *Journal of Economic Theory*, 13, 341–360.

SCHOTT, J. R. (2005): *Matrix Analysis for Statistics*, Wiley, New York, 2 ed.

STOCK, J. H. AND M. W. WATSON (2002): "Forecasting Using Principal Components from a Large Number of Predictors," *Journal of the American Statistical Association*, 97, 1167 – 1179.

ZAFFARONI, P. (2019): "Factor Models for Conditional Asset Pricing," *Working Paper*.

Appendix

Section A.1 provides an illustrative example for equation (2.3) whereas Section A.2 extends our analysis to multiple sorting and interaction terms. A discussion of regularity conditions appears in Section A.3. Section A.4 summarizes variations of Principal Components Analysis (PCA) discussed by Lettau and Pelger (2020a and 2020b) and Zaffaroni (2019), among others, allowing to estimate non-zero mean factors from a panel of excess returns. Section A.5 extend Theorems 1 and 2 in AGGR to the case in which the latent factors are estimated by the variations of PCA described in Section A.4. As the factors are not zero-mean, the estimators of the canonical correlations among the factors from the two groups and the test statistics in Theorems 1 and 2 of AGGR need to be adjusted accordingly. Proofs of propositions and theorems, together with the assumptions, are provided in the Online Appendix. We will denote the sample mean a generic sequence z_t, $t = 1, ..., T$ as $\bar{z} = \frac{1}{T}\sum_{t=1}^{T} z_t$, the T-dimensional vector of ones as $\mathbb{1}_T = [1, ..., 1]'$, and the identity matrix of order T as I_T.

A.1 Illustrative example for equation (2.3)

It might be worth providing a simple illustrative example of the result appearing in equation (2.3). Assume there are two mutually independent factors with Gaussian marginal distributions $f_j \sim$ N(m_j,1) where $m_j > 0$, j = 1,2 are the expected values (risk premia) of the factors. Both factors also feature order one autocovariance parameterized by $\rho_i \neq 0$ and volatility prediction $\mathbb{E}(f_{i,\tau-1}f_{i,\tau}^2)$ = $\rho_i^v \neq 0$ for i = 1, 2. Assume that f_1 is the "true/common" factor in f^c, and the time-varying betas $b_{i,\tau-1}$ are driven by $Z_{\tau-1} = f_{1,\tau-1}f_{2,\tau-1}$ so that f_2 only affects the loadings. According to equation (2.3) we need to regress $Z_{\tau-1}f_{1,\tau}$ onto $f_{1,\tau}$ and show that also $f_{2,\tau}$ appears in that projection. We do so sequentially, first projecting on f_1 and then regress the residual onto f_2 since both regressors are mutually independent (cfr. Frisch-Waugh-Lovell theorem). Denote by β_1 the slope coefficient of the projection onto f_1, i.e. β_1 = cov($Z_{\tau-1}f_{1,\tau}, f_{1,\tau}$)/ var($f_{1,\tau}$) = cov($Z_{\tau-1}f_{1,\tau}, f_{1,\tau}$) = cov($f_{1,\tau-1}f_{2,\tau-1}f_{1,\tau}, f_{1,\tau}$) = $(\rho_1^v - m_1\rho_1 - m_1^3)m_2$. The residuals projected onto $f_{2,\tau}$ yield a slope β_2 = cov($Z_{\tau-1}f_{1,\tau} - (\rho_1^v - m_1\rho_1 - m_1^3)m_2 f_{1,\tau}, f_{2,\tau}$)/ var($f_{2,\tau}$) = cov($f_{1,\tau-1}f_{2,\tau-1}f_{1,\tau}, f_{2,\tau}$) = $(\rho_1 + m_1^2)\rho_2$. Hence, the time-invariant factor representation of model (2.1) features the "true" factor $f_{1,\tau}$ with loading proportional to the slope β_1 = $(\rho_1^v - m_1\rho_1 - m_1^3)m_2 \neq 0$ and a factor driving the time-varying betas which has loading proportional to the slope β_2 = $(\rho_1 + m_1^2)\rho_2 \neq 0$: this is an illustration of what happens in equation (2.3). Furthermore, our assumption implies that the residuals after projection on $f_{1,\tau}$ and $f_{2,\tau}$, i.e. $\delta_{i,\tau}$ in equation (2.3), feature weak cross-sectional dependence.

A.2 Multiple sorting and interaction terms

In this section we show that our analysis also covers portfolios built on double, triple etc. characteristic sorting, as long as Assumptions A.1 and A.2 are satisfied for the entire set of portfolios considered. To see this, note that we referred to $w_{i,\tau-1}$ as the vector collecting the N_w characteristics of each asset i, that is $w_{i,\tau-1} = [w_{i,\tau-1,1}, ..., w_{i,\tau-1,N_w}]'$. Portfolio weights $g_j(w_{i,\tau-1})$ are defined as generic functions allowing to construct for example the j-th portfolio by sorting stocks on one characteristic, say $g_j(w_{i,\tau-1}) = c \cdot \mathbf{1}\{q_{\alpha-1}[w_{\cdot,\tau-1,h}] < w_{i,\tau-1,h} \leq q_\alpha[w_{\cdot,\tau-1,h}]\}$ where $q_\alpha[w_{\cdot,\tau-1,h}]$ is the α-decile of the cross-sectional distribution at date $\tau-1$ of the characteristic $w_{i,\tau-1,h}$, and $\mathbf{1}\{\cdot\}$ denotes the indicator function. Moreover, our generic function $g_j(w_{i,\tau-1})$ allows also for portfolios built by double-sorting stocks on two characteristics of stock i, say size ($w_{i,\tau-1,1}$) and book-to-market ($w_{i,\tau-1,2}$): $g_j(w_{i,\tau-1}) = c \cdot \mathbf{1}\{q_{\alpha-1}[w_{\cdot,\tau-1,1}] < w_{i,\tau-1,1} \leq q_\alpha[w_{\cdot,\tau-1,1}]\ ,\ q_{\beta-1}[w_{\cdot,\tau-1,2}] < w_{i,\tau-1,2} \leq q_\beta[w_{\cdot,\tau-1,2}]\}$, where $q_\beta[w_{\cdot,\tau-1,2}]$ is the β-decile of the cross-sectional distribution at date $\tau-1$ of the characteristic $w_{i,\tau-1,2}$.

We also noted in the main body of the paper that our analysis can handle interactions of characteristics. To see this, consider another generic set N_z individual characteristics $z_{i,\tau,1}$, ..., z_{i,τ,N_z}, not necessarily overlapping with $w_{i,\tau}$. Then, each element $\tilde{Z}_{i,\tau-1,h}$ of the vector of specific variables $\tilde{Z}_{i,\tau-1}$ appearing in betas of individual stocks could simply be some (transformations of) a specific characteristic, say $\tilde{Z}_{i,\tau-1,h} = z_{i,\tau-1,1}$, but can also be a function of multiple characteristic, say $\tilde{Z}_{i,\tau-1,h} = z_{i,\tau-1,1} \cdot z_{i,\tau-1,2}$.

A.3 Regularity conditions

In this section we digress on two technical regularity conditions which result in the constant loading factor representation for the two panels - one of individual stock excess returns and the other of sorted portfolios with returns described by equations (2.4) and (2.8).

We start with an assumption which tells us that some masking of factors may occur, but only for an asymptotically vanishing fraction of the individual stocks and portfolios. Put differently, some stocks/portfolios may mask some factors, but not a significant collection of portfolios is allowed to block out any specific factor.

ASSUMPTION A.1 (No Masking). *Consider matrix $\Lambda_2^c \equiv \left[\tilde{b}_j^0 + K_j^c\right]'$ appearing in equation (2.8). For any element $k \in \{1, \ldots, k^c\}$, the fraction $N_{2,k}$ out of N_2 portfolios with non-zero $\Lambda_{2,k}^c$ for factor k is such that $N_{2,k}/N_2$ is bounded away from zero. Similarly, for $\Lambda_1^c \equiv [b_i^0 + M_i^c]'$ the fraction $N_{1,k}$ out of N_1 individual stocks with non-zero $\Lambda_{1,k}^c$ for factor k is such that $N_{1,k}/N_1$ is bounded away from zero.*

In addition, we also need the following assumption for the common factors between the two panels of respectively individual stocks and sorted portfolios to be f_τ^c only, instead of f_τ^c and/or $Z'_{\tau-1} f_\tau^c$.

ASSUMPTION A.2 (Convergence of portfolio betas)**.** *Consider the term $W_j Z^*_{\tau-1}$ appearing in the vector of portfolio j's betas $\left[\tilde{b}_j^0 + W_j Z^*_{\tau-1}\right]$ in equation (2.7). The fraction N_2^* out of N_2 portfolios for which $W_j Z^*_{\tau-1} = A_j Z_{\tau-1}$, is such that N_2^*/N_2 converges to zero as $N_1, N_2, N_2^* \to \infty$, for $A_j \neq 0$.*

Assumption A.2 holds if $\frac{1}{N_1}\sum_{i=1}^{N_1} g_j(w_{i,\tau-1}) C_i \tilde{Z}_{i,\tau-1} \not\to \breve{W} Z_{\tau-1}$, and $\frac{1}{N_1}\sum_{i=1}^{N_1} g_j(w_{i,\tau-1}) B_i$ converges to a time dependent limit $N_1 \to \infty$, say $\breve{W}_j \breve{Z}_{\tau-1}$, for a sufficiently large number of portfolios. This implies $\frac{1}{N_1}\sum_{i=1}^{N_1} g_j(w_{i,\tau-1}) B_i Z_{\tau-1} \to \breve{W}_j \breve{Z}_{\tau-1} Z_{\tau-1}$, which is not linear in $Z_{\tau-1}$. To illustrate the condition that $\frac{1}{N_1}\sum_i g(w_{i,\tau}) B_i$ converges to a time-dependent limit, consider a situation where the portfolio weights for portfolio j load on stocks with characteristic $w_{i,\tau}$ in a given interval $\mathcal{I}_j$ (think of size, book-to-market etc.). Then, the large sample limit of $\frac{1}{N_1}\sum_i g(w_{i,\tau}) B_i$ for portfolio j is: $\mathbb{E}[B_i | w_{i,\tau} \in \mathcal{I}_j]$. The time dependent limit in Assumption A.2 holds for example when the group of stocks having characteristics in a certain range, varies over time, and there is sufficient cross-sectional heterogeneity in the coefficients in B_i. As an illustrative example, we can consider size sorted portfolios. Assumption A.2 implies that the mix of firms in a particular decile size portfolio varies through time in terms of characteristics other than size. Finally, of less interest but worth noting is the fact that a sufficient condition for Assumption A.2 to hold is that $W_j = 0$ for a fraction of N_2^{**} out of N_2 portfolios such that $1 - N_2^{**}/N_2$ converges to zero, i.e. the betas with respect the common factors f_τ^c are time invariant for the majority of the portfolios: $\frac{1}{N_1}\sum_{i=1}^{N_1} \alpha_{j,i,\tau-1} b_{i,\tau-1} \to \tilde{b}_j^0$.

A.4 Factor estimation: PCA and its recent extensions

To simplify the exposition we will use a generic notation here for the discussion of various estimators which can be applied to different panel data settings. Let y_t be N-dimensional vector of returns, and assume that the data generating process of y_t is a linear factor model as the APT of Ross (1976), that is:

$$y_t = \Lambda h_t + \varepsilon_t, \qquad t = 1, ..., T, \tag{A.1}$$

where h_t is the $(k,1)$ vector of (unobservable) factors with expected value $\mu_h \equiv E[h_t]$, possibly different from zero, $\Lambda = [\lambda_1 , ..., \lambda_N]'$ is the (N,k) full column-rank matrix of unknown loadings, and the idiosyncratic innovations $E[\varepsilon_t] = 0$. These assumptions imply $E[y_t] = \Lambda\mu_h$, possibly different from zero. Model (A.1) can be written as:

$$Y = H\Lambda' + \varepsilon, \tag{A.2}$$

where $Y = [y_1, ..., y_T]'$ is the (T, N)-dimensional matrix of observed excess returns, and $H = [h_1, ..., h_T]'$ is the (T, k)-dimensional matrix of factor values.

Lettau and Pelger (2020a, 2020b) and Zaffaroni (2019) suggest that estimating model (A.2) by performing PCA on the demeaned returns $\tilde{y}_t = y_t - \bar{y}$, as typically done in the finance and macroeconomics literature, is restrictive as the mean of the factors and the returns should contain information on the factor structure. Let $\tilde{h}_t = h_t - \bar{h}$ be the demeaned factors, and $\bar{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{i,t}$ be the time series mean of the returns of the i-th asset, with $i = 1, ..., N$. Lettau and Pelger (2020a) address the estimation of the non-demeaned factors in model (A.1) with their RP-PCA procedure, which consists in solving the following minimization problem:

$$\min_{\substack{\lambda_1, ..., \lambda_N, \\ h_1, ..., h_T,}} \quad \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(\tilde{y}_{i,t} - \tilde{h}_t'\lambda_i)^2 \; + \; (1+\gamma_{RP})\frac{1}{N}\sum_{i=1}^{N}(\bar{y}_i - \bar{h}'\lambda_i)^2 \; . \tag{A.3}$$

The first double summation in (A.3) corresponds to the average unexplained (time-series) variation of the data, the second summation correspond to the (cross-sectional) average of the squared "pricing errors" across all N assets, and $\gamma_{RP} \in [-1, +\infty)$ is a constant which can be interpreted as a tuning parameter: as it increases more weight is given to the pricing errors in the factor estimation. They show that the solution to (A.3) can be obtained by performing the following two steps:

(i) Estimate the loading matrix Λ_j is as $\sqrt{N_j}$ times the (N, k) matrix of the eigenvectors associated to the largest k eigenvalues of matrix

$$M_{RP}(\gamma_{RP}) \;\; := \;\; \frac{1}{T}\sum_{t=1}^{T} y_t y_t' + \gamma_{RP}\left(\frac{1}{T}\sum_{t=1}^{T} y_t\right)\left(\frac{1}{T}\sum_{t=1}^{T} y_t\right)' \; . \tag{A.4}$$

The estimated loadings, that we denote as $\hat{\Lambda}_{RP}$, are such that $\hat{\Lambda}_{RP}'\hat{\Lambda}_{RP}/N = I_k$. [15]

(ii) Estimate the latent factors in model (A.1) at each date t by a cross-sectional regression of the returns y_t on the estimated loadings $\hat{\Lambda}_{RP}$:

$$\hat{h}_{t,RP} \;\; := \;\; \left(\hat{\Lambda}_{RP}'\hat{\Lambda}_{RP}\right)^{-1}\hat{\Lambda}_{RP}'\, y_t \; . \tag{A.5}$$

We denote as $\hat{H}_{RP} = [\hat{h}_{1,RP}, ..., \hat{h}_{T,RP}]'$ the (T, k) matrix of estimated factors.

[15] Lettau and Pelger (2020a), in their online appendix, show that $\hat{\Lambda}_{j,RP}$ can be obtained as the conventional PCA estimator of the loadings applied to the "projected" model: $\check{Y}_j = \check{H}_j\Lambda_j + \check{\varepsilon}_j$ where $\check{Y} := W(\gamma_{RP})Y$, $\check{H} := W(\gamma_{RP})H$, $\check{\varepsilon} := W(\gamma_{RP})\varepsilon$, and $W(\gamma_{RP}) = I_T + \left(\sqrt{\gamma_{RP}+1} - 1\right)\frac{\mathbb{1}_T\mathbb{1}_T'}{T}$. That is, the loading matrix $\Lambda_{j,RP}$ can be estimated as $\sqrt{N}$ times the (N, k) matrix of the eigenvectors associated to the largest k eigenvalues of $M_{RP}(\gamma_{RP}) = \frac{1}{T}\check{Y}'\check{Y}$.

As linear latent factor models are identified up to an invertible transformation, an equivalent estimator $\hat{H}^{*}_{RP}$ of the factors is obtained by rescaling $\hat{H}_{RP}$ such that the (uncentered) second moment of the estimated factors is $\hat{H}^{*\prime}_{RP}\hat{H}^{*}_{RP}/T = I_k$, that is $\hat{H}^{*}_{RP} := \hat{H}_{RP}\left(\hat{H}^{\prime}_{RP}\hat{H}_{RP}/T\right)^{-1/2}$. Following Lettau and Pelger (2020a), we refer to $\hat{H}_{RP}$ and $\hat{H}^{*}_{RP}$ as "RP-PCA estimators". Importantly, the factors estimated by RP-PCA have a mean $\hat{\bar{h}} = \frac{1}{T}\sum_{t=1}^{T}\hat{h}_t$ which is not necessarily equal to zero. In fact, equation (A.5) shows that $\hat{h}_{t,RP}$ is a linear combination of the original returns which are not-demeaned.[16]

Special case of the RP-PCA: $\gamma_{RP} = 0$

When $\gamma_{RP} = 0$, the matrix $M_{RP}(\gamma_{RP})$ characterizing the RP-PCA estimator (A.5) coincides with the conventional PCA estimator but with loadings estimated from the uncentered second moment matrix of the returns $M_{RP}(\gamma_{RP} = 0) = \frac{1}{T}\sum_{t=1}^{T} y_t y_t^{\prime}$. The RP-PCA estimators of the factors and the loadings with $\gamma_{RP} = 0$ coincides with those proposed by Zaffaroni (2019).

Special cases of the RP-PCA: $\gamma_{RP} = -1$

When $\gamma_{RP} = -1$ the RP-PCA estimator of the loadings, denoted by $\hat{\Lambda}_{PCA}$, is computed as $\sqrt{N}$ times the eigenvectors of the sample variance-covariance matrix of the returns $\hat{V}(y_t) = M_{RP}(\gamma_{RP} = -1) = \frac{1}{T}\sum_{t=1}^{T}\tilde{y}_t\tilde{y}_t^{\prime}$. We denote the RP-PCA factor estimator in this special case as $\hat{h}_{t,PCA}$:

$$\hat{h}_{t,PCA} := \left(\hat{\Lambda}^{\prime}_{PCA}\hat{\Lambda}_{PCA}\right)^{-1}\hat{\Lambda}^{\prime}_{PCA}\, y_t, \tag{A.6}$$

and name $\hat{\Lambda}_{PCA}$ and $\hat{h}_{t,PCA}$ as the "conventional PCA" estimators of the loadings and factors, respectively, as they are used by most of the financial literature. For instance, the factor estimators used in Connor and Korajczyk (1988), Lehmann and Modest (2005), Kozak et al. (2018), Kozak, Nagel, and Santosh (2020), Giglio and Xiu (2021), and Pukthuanthong et al. (2019), among others all coincide with $\hat{h}_{t,PCA}$. Another frequently used estimator, denoted by $\hat{\hat{h}}_{t,PCA}$, is obtained by a cross-sectional regression of the demeaned returns $\tilde{y}_t$ on $\hat{\Lambda}_{PCA}$:

$$\hat{\hat{h}}_{t,PCA} := \left(\hat{\Lambda}^{\prime}_{PCA}\hat{\Lambda}_{PCA}\right)^{-1}\hat{\Lambda}^{\prime}_{PCA}\,\tilde{y}_t\,. \tag{A.7}$$

[16] Zaffaroni (2019), in his Section 3, notices the estimated factors $\hat{h}_{t,RP}$ are portfolio (excess-) returns, and correspond to "the feasible PCA-estimators" of the infeasible "mimicking portfolios" (of the true the latent factors) proposed by Huberman, Kandel, and Stambaugh (1987) and Breeden, Gibbons, and Litzenberger (1989). See Lehmann and Modest (2005) for a discussion of factor-mimicking portfolio estimators.

Differently from $\hat{h}_{t,PCA}$ and $\hat{h}_{t,RP}$, factors $\hat{\hat{h}}_{t,PCA}$ have zero-mean as they are linear combinations of the demeaned data $\tilde{y}_t$.[17]

Let $\tilde{Y} = [\tilde{y}_1, ..., \tilde{y}_T]'$ be (T, N) matrix collecting the demeaned returns. AGGR consider the estimator $\hat{\hat{H}}^*_{PCA} = [\hat{\hat{h}}^*_{1,PCA}, ..., \hat{\hat{h}}^*_{T,PCA}]'$ of the k factors which is defined as $\sqrt{T}$ times the eigenvectors associated to the k largest eigenvalues of the matrix $\frac{1}{NT}\tilde{Y}\tilde{Y}'$. By construction the estimated factors are zero mean, and their (sample) variance-covariance matrix is $\hat{\hat{H}}^{*\prime}_{PCA}\hat{\hat{H}}^*_{PCA}/T = I_k$. Using the arguments in Bai and Ng (2002), it can be shown that $\hat{\hat{H}}^{*\prime}_{PCA}$ is equal to the PCA estimator in (A.7) rescaled to have unit variance: $\hat{\hat{H}}^*_{PCA} = \hat{\hat{H}}_{PCA}\left(\hat{\hat{H}}'_{PCA}\hat{\hat{H}}_{PCA}/T\right)^{-1/2}$.

Importantly, $\hat{\hat{h}}_{t,PCA}$ and $\hat{\hat{h}}^*_{t,PCA}$ are consistent estimators of the latent factors only when these are assumed to have zero expected value, as in Assumption A.2 of AGGR. In the next Section A.5 we show that relaxing this assumption does not change the main results of their paper, but requires modifications to their canonical correlations estimator as well as other statistics.

A.5 Identification, estimation and test for common and group-specific factors with generic mean

Consider the group-factor model appearing in equation (2.10). As in AGGR we assume, without loss of generality, that the group-specific factors $f^s_{1,t}$ and $f^s_{2,t}$ are orthogonal to the common factor f^c_t. Since the unobservable factors can be standardized, we have:

$$E\begin{bmatrix} f^c_t \\ f^s_{1,t} \\ f^s_{2,t} \end{bmatrix} = \begin{bmatrix} \mu^c \\ \mu^s_1 \\ \mu^s_2 \end{bmatrix}, \quad \text{and} \quad \Sigma_F := V\begin{bmatrix} f^c_t \\ f^s_{1,t} \\ f^s_{2,t} \end{bmatrix} = \begin{bmatrix} I_{k^c} & 0 & 0 \\ 0 & I_{k^s_1} & \Phi \\ 0 & \Phi' & I_{k^s_2} \end{bmatrix}, \tag{A.8}$$

where the expected values of the factors are finite, and matrix Σ_F is positive-definite. We allow for a non-zero covariance Φ between group-specific factors, but differently from AGGR, we allow the factors to have expected value different from zero. We refer to (A.8) as Assumption B.2 in the list of regularity conditions in Appendix B.1. Model (2.10) together with Assumption B.2 is identified by the same arguments used by AGGR.

Let $h_{j,t} = [f^{c\prime}_t, f^{s\prime}_{j,t}]'$, with $j = 1, 2$, and $V_{j\ell} = Cov(h_{j,t}, h_{\ell,t})$, with $j, \ell = 1, 2$. The $\underline{k} = \min(k_1, k_2)$ largest eigenvalues of the matrices $R = V^{-1}_{11}V_{12}V^{-1}_{22}V_{21}$ and $R^* = V^{-1}_{22}V_{21}V^{-1}_{11}V_{12}$ are the same, and are equal to the squared canonical correlations ρ^2_ℓ, $\ell = 1, ..., \underline{k}$, between $h_{1,t}$ and $h_{2,t}$. The

[17] As discussed Section 2 of Zaffaroni (2019), $\hat{\hat{h}}_{t,PCA}$ is the estimator of the (demeaned) latent factors $\tilde{h}_t := h_t - \bar{h}$ of model (A.1) for the the demeaned data $\tilde{y}_t$. This can be easily seen by noting that the model for the demeaned data can be written as: $\tilde{y}_t = \Lambda_j(h_t - \bar{h}) + (\varepsilon_t - \bar{\varepsilon})$.

associated eigenvectors $w_{1,\ell}$ (resp. $w_{2,\ell}$), with $\ell = 1, ..., \underline{k}$, of matrix R (resp. R^*) standardized such that $w_{1,\ell}' V_{11} w_{1,\ell} = 1$ (resp. $w_{2,\ell}' V_{22} w_{2,\ell} = 1$) are the canonical directions which yield the canonical variables $w_{1,\ell}' h_{1,t}$ (resp. $w_{2,\ell}' h_{2,t}$). The next Proposition A.1 deals with determining k^c, the number of common factors, using canonical correlations between the vectors $h_{1,t}$ and $h_{2,t}$, which are unobserved and estimated by PCA or its variations described in Section A.4. It corresponds to Proposition 1 in AGGR where the zero mean assumption of the factors is replaced with our new Assumption B.2.

PROPOSITION A.1. *Under Assumption B.2, the following hold:*

(i) If $k^c > 0$, the largest k^c canonical correlations between $h_{1,t}$ and $h_{2,t}$ are equal to 1, and the remaining $\underline{k} - k^c$ canonical correlations are strictly less than 1.

(ii) Let W_j be the (k_j, k^c) matrix whose columns are the canonical directions for $h_{j,t}$ associated with the k^c canonical correlations equal to 1, for $j = 1, 2$. Then, $f_t^c = W_j' h_{j,t}$ (up to an orthogonal matrix), for $j = 1, 2$.

(iii) If $k^c = 0$, all canonical correlations between $h_{1,t}$ and $h_{2,t}$ are strictly less than 1.

(iv) Let W_1^s (resp. W_2^s) be the (k_1, k_1^s) (resp. (k_2, k_2^s)) matrix whose columns are the eigenvectors of matrix R (resp. R^*) associated with the smallest k_1^s (resp. k_2^s) eigenvalues.

Then $f_{j,t}^s = W_j^{s\prime} h_{j,t}$ (up to an orthogonal matrix) for $j = 1, 2$.

Proposition A.1 shows that the number of common factors k^c, the common factor space spanned by f_t^c, and the spaces spanned by group-specific factors, can be identified from the canonical correlations and canonical variables of $h_{1,t}$ and $h_{2,t}$. Therefore, the factor space dimensions k^c, k_j^s, and factors f_t^c and $f_{j,t}^s$, $j = 1, 2$, are identifiable (up to a rotation) from information that can be inferred by disjoint PCA, or its variations described in Section A.4, on the two subgroups.

A.5.1 Estimators

When the true number of factors $k_j > 0$ in each subgroup $j = 1, 2$ and $k^c > 0$ are known, Proposition A.1 suggests the following estimation procedure. Let $h_{1,t}$ and $h_{2,t}$ be estimated by extracting the first k_j PCs (or its variations) from each sub-panel j, and denote by $\hat{h}_{j,t}$ these PC estimates of the factors, $j = 1, 2$. Let $\hat{V}_{j\ell}$ denote the empirical covariance matrix between $\hat{h}_{j,t}$ and $\hat{h}_{\ell,t}$, i.e.:

$$\hat{V}_{j\ell} = \frac{1}{T} \sum_{t=1}^{T} \hat{h}_{j,t} \hat{h}_{\ell,t}' - \left(\frac{1}{T} \sum_{t=1}^{T} \hat{h}_{j,t} \right) \left(\frac{1}{T} \sum_{t=1}^{T} \hat{h}_{\ell,t}' \right), \qquad j, \ell = 1, 2, \tag{A.9}$$

and let:

$$\hat{R} \ := \ \hat{V}_{11}^{-1}\hat{V}_{12}\hat{V}_{22}^{-1}\hat{V}_{21}, \ \text{ and } \ \hat{R}^* := \hat{V}_{22}^{-1}\hat{V}_{21}\hat{V}_{11}^{-1}\hat{V}_{12}. \tag{A.10}$$

be the estimators of matrices matrices R and R^*, respectively. Differently from AGGR, the estimators of the variance-covariance matrices $\hat{V}_{j\ell}$ take into account that the estimated factors might have non-zero mean, compatible with the RP-PCA estimators described above. Matrices $\hat{R}$ and $\hat{R}^*$ have the same non-zero eigenvalues. The k^c largest eigenvalues of $\hat{R}$ (resp. $\hat{R}^*$), denoted by $\hat{\rho}_\ell^2$, $\ell = 1, ..., k^c$, are the first k^c squared sample canonical correlation between $\hat{h}_{1,t}$ and $\hat{h}_{2,t}$. The associated k^c canonical directions, collected in the (k_1, k^c) matrix $\hat{W}_1$ (resp. (k_2, k^c) matrix $\hat{W}_2$), are the eigenvectors associated with the k^c largest eigenvalues of matrix $\hat{R}$ (resp. $\hat{R}^*$), normalized to have length 1 with respect to $\hat{V}_{11}$ (resp. $\hat{V}_{22}$). It also holds that:

$$\hat{W}_1'\hat{V}_{11}\hat{W}_1 = I_{k^c}, \ \text{ and } \ \hat{W}_2'\hat{V}_{22}\hat{W}_2 = I_{k^c}. \tag{A.11}$$

DEFINITION 1. *Two estimators of the common factors vector are* $\hat{f}_t^c = \hat{W}_1'\hat{h}_{1,t}$ *and* $\hat{f}_t^{c*} = \hat{W}_2'\hat{h}_{2,t}$.

Definition 1 and equation (A.11) imply that the estimated common factors have identity sample variance-covariance matrix:

$$\hat{V}(\hat{f}_t^c) := \frac{1}{T}\sum_{t=1}^{T}\hat{f}_t^c\hat{f}_t^{c\prime} - \left(\frac{1}{T}\sum_{t=1}^{T}\hat{f}_t^c\right)\left(\frac{1}{T}\sum_{t=1}^{T}\hat{f}_t^{c\prime}\right) = I_{k^c} \ , \tag{A.12}$$

and analogously $\hat{V}(\hat{f}_t^{c*}) = \frac{1}{T}\sum_{t=1}^{T}\hat{f}_t^{c*}\hat{f}_t^{c*\prime} - \left(\frac{1}{T}\sum_{t=1}^{T}\hat{f}_t^{c*}\right)\left(\frac{1}{T}\sum_{t=1}^{T}\hat{f}_t^{c*\prime}\right) = I_{k^c}$, i.e. the estimated common factor values match in-sample the normalization condition of identity variance-covariance matrix in (A.8). An estimator for the group-specific factors $f_{1,t}^s$ (resp. $f_{2,t}^s$) is obtained by computing the first k_1^s (resp. k_2^s) principal components of the variance-covariance matrix of the residuals of the regression of $y_{1,t}$ (resp. $y_{2,t}$) on the estimated common factors.

Let $\breve{F}^c = [\breve{f}_1^{c\prime}, ..., \breve{f}_T^{c\prime}]'$ be the (T, k^c) matrix of estimated demeaned common factors, and $\hat{\Lambda}_j^c = [\hat{\lambda}_{j,1}^c, ..., \hat{\lambda}_{j,N_j}^c]'$ the (N_j, k^c) matrix collecting the estimated loadings, and let $\breve{Y}_j$ be the (T, N_j) matrix of (time-series) demeaned observations for group j:

$$\hat{\Lambda}_j^c = \breve{Y}_j'\breve{F}^c(\breve{F}^{c\prime}\breve{F}^c)^{-1} = \frac{1}{T}\breve{Y}_j'\breve{F}^c, \qquad j = 1, 2. \tag{A.13}$$

As shown in the OA, this estimator is unbiased for Λ^c when group-specific factors $\hat{F}_t^s$ have expected value different from zero. In this case, indeed, regressing the non-demeaned original data Y_j on the non-demeaned estimated factors $\hat{F}^c$ would produce a biased estimator for Λ^c as the residuals of this regression model would be a linear function of $\hat{F}_t^s$ which, in general, are not zero-mean, as allowed by the first of the two conditions in (A.8).

Define the residual $(N_j, 1)$ vector $\xi_{j,t} = y_{j,t} - \hat{\Lambda}_j^c \hat{f}_t^c$ and the (T, N_j) matrix of the regression residuals $\Xi_j = [\xi_{j,1}, ..., \xi_{j,T}]'$ for $j = 1, 2$.

DEFINITION 2. *Estimators of the specific factors $\hat{f}_{1,t}^s$ (resp. $\hat{f}_{2,t}^s$) are defined as the first k_1^s (resp. k_2^s) Risk Premium Principal Components of sub-panel Ξ_1 (resp. Ξ_2).*

Note that $\breve{f}_t^c := \hat{f}_t^c - \bar{\hat{f}}^c$, the demeaned estimated common factor, is orthogonal in-sample both to $\breve{f}_{t,1}^s := \hat{f}_{t,1}^s - \bar{\hat{f}}_1^s$ and to $\breve{f}_{t,2}^s := \hat{f}_{t,2}^s - \bar{\hat{f}}_2^s$, that is the demeaned group-specific factors, matching the population orthogonality assumption in (A.8). Let us define $\breve{\Xi}_j$ as the (T, N_j) matrix of (time-series) demeaned estimated residuals for group j. Then, the (N_j, k_j^s) matrix of the loadings estimators $\hat{\Lambda}_j^s = [\hat{\lambda}_{j,1}^s, ..., \hat{\lambda}_{j,N_j}^s]'$ is:

$$\hat{\Lambda}_j^s = \breve{\Xi}_j' \breve{F}_j^s (\breve{F}_j^{s\prime} \breve{F}_j^s)^{-1} = \frac{1}{T} \breve{\Xi}_j' \breve{F}_j^s, \qquad j = 1, 2\ . \tag{A.14}$$

In a follow-up paper, Andreou, Gagliardini, Ghysels, and Rubin (2020) explore the idea that linear combinations of the two estimators $\hat{f}_t^c$ and $\hat{f}_t^{c*}$ also yield valid estimators when they are estimated by PCA on demeaned data as in AGGR. Following their arguments, we can consider the generic estimator

$$\hat{f}_t^{c\star} = S(\omega)(\hat{f}_t^c + \omega \hat{f}_t^{c*}), \tag{A.15}$$

obtained as the linear combination of $\hat{f}_t^c$ and $\hat{f}_t^{c*}$ estimated by RP-PCs as in our Definition 1 and where the scalar parameter ω is the weight. The transformation by matrix $S(\omega) = [(1 + \omega^2) I_{k^c} + 2\omega \hat{D}]^{-1/2}$, with $\hat{D} = \text{diag}(\hat{\rho}_1, ..., \hat{\rho}_{k^c})$, ensures that the new estimator has identity sample covariance matrix, that is:

$$\hat{V}(\hat{f}_t^{c\star}) := \frac{1}{T} \sum_{t=1}^{T} \hat{f}_t^{c\star} \hat{f}_t^{c\star\prime} - \left(\frac{1}{T} \sum_{t=1}^{T} \hat{f}_t^{c\star} \right) \left(\frac{1}{T} \sum_{t=1}^{T} \hat{f}_t^{c\star\prime} \right) = I_{k^c}\ . \tag{A.16}$$

The parametric family (A.15) encompasses the estimators $\hat{f}_t^c$ and $\hat{f}_t^{c*}$ in our Definition 1, which correspond to choices $\omega = 0$ and $\omega = +\infty$, respectively. By using arguments analogous to those in Andreou et al. (2020), it can be shown that choosing $\omega = 1$, the common factor estimator $\hat{f}_t^{c\star}$ is an equally-weighted linear combination of the two basis estimators $\hat{f}_t^c$ and $\hat{f}_t^{c*}$. This new estimator extends the one proposed by Goyal, Pérignon, and Villa (2008), which was originally derived for zero mean data and factors, by allowing factors with possibly non-zero mean.[18]

[18] Andreou et al. (2020) also show that an alternative choice for ω is provided by the optimal weight which minimizes the Asymptotic Mean Square Error (AMSE) of the factor estimator. In a simplified setting $k^c = 1$, under the same set of assumptions of Theorem A.2 and the additional assumption $N_1/T^2 = o(1)$, they show that the average AMSE is

For a given choice of the weight ω, let $\hat{F}^{c\star} = [\hat{f}_1^{c\star\prime}, ..., \hat{f}_T^{c\star\prime}]'$ be the (T, k^c) matrix of estimated common factors, the estimators of the common factor loadings, group-specific factors and loadings are the same as in equation (A.13), Definition 2, and equation (A.14), respectively, where the estimator $\hat{F}^c$ is replaced by $\hat{F}^{c\star}$. In unreported Monte Carlo experiments we find that the estimator $\hat{f}_t^{c\star}$ with weight $\omega = 1$, has better small sample properties than the estimators $\hat{f}_t^c$ and $\hat{f}_t^{c*}$. For this reason in the empirical application we use $\hat{f}_t^{c\star}$ as the estimator of the common factors with weight $\omega = 1$.

A.5.2 Inference on the number of common factors via canonical correlations

In order to infer the dimension k^c of the common factor space, we consider the case where the number of pervasive factors k_1 and k_2 in each sub-panel is known, hence $\underline{k} = \min(k_1, k_2)$ is also known. As explained in AGGR, all the results remain unchanged when the numbers of pervasive factors k_1 and k_2 are estimated consistently. From Proposition A.1, dimension k^c is the number of unit canonical correlations between $h_{1,t}$ and $h_{2,t}$.

We consider the hypotheses:

$$H(0) = \{1 > \rho_1 \geq \ldots \geq \rho_{\underline{k}}\}, \quad H(1) = \{\rho_1 = 1 > \rho_2 \geq \ldots \geq \rho_{\underline{k}}\}, \ldots,$$
$$H(k^c) = \{\rho_1 = \ldots = \rho_{k^c} = 1 > \rho_{k^c+1} \geq \ldots \geq \rho_{\underline{k}}\}, \ldots,$$
$$\text{and finally } H(\underline{k}) = \{\rho_1 = ... = \rho_{\underline{k}} = 1\},$$

where $\rho_1, ..., \rho_{\underline{k}}$ are the ordered canonical correlations of $h_{1,t}$ and $h_{2,t}$. Generically, $H(k^c)$ corresponds to the case of k^c common factors and $k_1 - k^c$ and $k_2 - k^c$ group-specific factors in each group, and $H(0)$ corresponds to the absence of common factors. In order to select the number of common factors, let us consider the following sequence of tests: $H_0 = H(k^c)$ against $H_1 = \bigcup_{0 \leq r < k^c} H(r)$, for each $k^c = \underline{k}, \underline{k} - 1, ..., 1$. To test H_0 against H_1, for any given $k^c = \underline{k}, \underline{k} - 1, ..., 1$ we consider the test statics $\hat{\xi}(k^c)$ defined in equation (2.11). The null hypothesis $H_0 = H(k^c)$ is rejected when $\hat{\xi}(k^c) - k^c$ is negative and large. The critical value is obtained from the large sample distribution of the statistic when N_1, N_2, $T \to \infty$, provided below. The number of common factors is estimated by sequentially applying the tests starting from $k^c = \underline{k}$, the maximum number of common factors.

Let us denote $N = \min\{N_1, N_2\}$ and $\mu_N = \sqrt{N_2/N_1}$. Without loss of generality, we set $N = N_2$, which implies $\mu_N \leq 1$. We assume that:

$$\sqrt{T}/N = o(1), \quad N/T^2 = o(1) \text{ and } \mu_N \to \mu, \text{ with } \mu \in [0, 1]. \tag{A.17}$$

Note that the assumption $N/T^2 = o(1)$ is made by Lettau and Pelger (2020a), and is more restrictive than the assumption $N/T^{5/2} = o(1)$ made by AGGR in their equation (4.1).

minimized for $\omega = [N_2 \Sigma_{u,11}^{(cc)}]/[N_1 \Sigma_{u,22}^{(cc)}]$. This result also holds when the factors have possibly non-zero mean as in the set-up of this paper.

The large sample distribution of the test statistic for the number of common factors is derived following the same arguments as in AGGR. In Proposition B.2 of the Online Appendix we show that, for $t = 1, \ldots, T$ and $j = 1, 2$, the estimator $\hat{h}_{j,t}$ is asymptotically equivalent, up to negligible terms, to $\hat{\mathcal{H}}_j(h_{j,t} + u_{j,t}/\sqrt{N_j} + \breve{b}_{j,t}/T)$ where

$$
\begin{aligned}
u_{j,t} &= \left(\frac{1}{N_j}\sum_{i=1}^{N_j}\lambda_{j,i}\lambda_{j,i}'\right)^{-1} \frac{1}{\sqrt{N_j}}\sum_{i=1}^{N_j}\lambda_{j,i}\varepsilon_{j,i,t}, \\
\breve{b}_{j,t} &= \left(\frac{1}{N_j}\sum_{i=1}^{N_j}\lambda_{j,i}\lambda_{j,i}'\right)^{-1} \left(\frac{1}{T}\sum_{t=1}^{T}\breve{h}_{j,t}\breve{h}_{j,t}'\right)^{-1} \eta_{j,t}^2\breve{h}_{j,t}
\end{aligned}
$$

and $\eta_{j,t}^2 = \underset{N_j\to\infty}{\text{plim}} \frac{1}{N_j}\sum_{i=1}^{N_j} E[\varepsilon_{j,i,t}^2|\mathcal{F}_t]$ is the limit average error variance conditional on the sigma field $\mathcal{F}_t = \sigma(F_s, s \leq t)$ generated by current and past factor values $F_t = (f_t^{c\prime}, f_{1,t}^{s\prime}, f_{2,t}^{s\prime})'$, and $\lambda_{j,i}$ = $(\lambda_{j,i}^{c\prime}, \lambda_{j,i}^{s\prime})'$. The zero-mean term $u_{j,t}$ drives the randomness in group factor estimates conditional on factor path. Vector $b_{j,t}$ is measurable with respect to the factor path and induces a bias term at order T^{-1} in principal components estimates. Vectors $u_{j,t}$ and $b_{j,t}$ depend on sample sizes but, for convenience, we omit the indices N_j, T. and $\hat{\mathcal{H}}_j$ is a nonsingular stochastic factor rotation matrix. This expansion extends the results in Lettau and Pelger (2020a).

Let $\tilde{\Sigma}_{u,jk,t}(h) = Cov(u_{j,t}, u_{k,t-h}|\mathcal{F}_t)$ be the conditional covariance between $u_{j,t}$ and $u_{k,t-h}$, i.e.

$$
\tilde{\Sigma}_{u,jk,t}(h) = \left(\frac{1}{N_j}\sum_{i=1}^{N_j}\lambda_{j,i}\lambda_{j,i}'\right)^{-1} \frac{1}{\sqrt{N_jN_k}}\sum_{i=1}^{N_j}\sum_{\ell=1}^{N_k}\lambda_{j,i}\lambda_{k,\ell}'Cov(\varepsilon_{j,i,t}, \varepsilon_{k,\ell,t-h}|\mathcal{F}_t)\left(\frac{1}{N_k}\sum_{i=1}^{N_k}\lambda_{k,i}\lambda_{k,i}'\right)^{-1},
$$

and $\tilde{\Sigma}_{u,jk,t}(-h)$ = $\tilde{\Sigma}_{u,kj,t}(h)'$, for j, $k = 1, 2$ and h= 0, 1, ... We define $\tilde{\Sigma}_{u,jj,t} \equiv \tilde{\Sigma}_{u,jj,t}(0)$, and set $\Sigma_{u,jk,t}(h)$ = $\underset{N_j,N_k\to\infty}{\text{plim}} \tilde{\Sigma}_{u,jk,t}(h)$ and $\Sigma_{\lambda,j}$ = $\lim_{N_j\to\infty}\frac{1}{N_j}\sum_{i=1}^{N_j}\lambda_{j,i}\lambda_{j,i}'$, and let

$$
\breve{\breve{b}}_{j,t} := \Sigma_{\lambda,j}^{-1}\eta_{j,t}^2\breve{h}_{j,t} \tag{A.18}
$$

be the large sample counterpart of $b_{j,t}$, from Assumptions B.2 - B.4. The following Theorem A.1 provides the asymptotic distribution of the infeasible test statistic $\hat{\xi}(k^c)$.

THEOREM A.1. *Under Assumptions B.1 - B.7, and the null hypothesis $H_0 = H(k^c)$ of k^c common factors, we have:*

$$
N\sqrt{T} \cdot \Omega_{U,1}^{-1/2} \cdot \left[\hat{\xi}(k^c) - k^c + \frac{1}{2N}tr\left\{\tilde{\Sigma}_{cc}^{-1}\tilde{\Sigma}_U\right\} + \frac{1}{2T}tr\left\{\tilde{\Sigma}_{cc}^{-1}\tilde{\Sigma}_B\right\}\right] \xrightarrow{d} N(0,1), \tag{A.19}
$$

where $\tilde{\Sigma}_{cc} = \frac{1}{T}\sum_{t=1}^{T} \breve{f}_t^c \breve{f}_t^{c\prime}$. *Moreover* $\tilde{\Sigma}_B = \frac{1}{T}\sum_{t=1}^{T} \widetilde{\Delta b}_t^{(c)} \widetilde{\Delta b}_t^{(c)\prime}$, *and*

$$\begin{aligned}
\widetilde{\Delta b}_t^{(c)} &= \breve{b}_{1,t} - \breve{b}_{2,t} - \left(\frac{1}{T}\sum_{t=1}^{T}(\breve{b}_{1,t} - \breve{b}_{2,t})\breve{F}_s'\right)\left(\frac{1}{T}\sum_{t=1}^{T}\breve{F}_t\breve{F}_t'\right)^{-1}\breve{F}_t \\
\tilde{\Sigma}_U &= \frac{1}{T}\sum_{t=1}^{T}\left(\mu_N^2\tilde{\Sigma}_{u,11,t}^{(cc)} + \tilde{\Sigma}_{u,22,t}^{(cc)} - \mu_N\tilde{\Sigma}_{u,12,t}^{(cc)} - \mu_N\tilde{\Sigma}_{u,21,t}^{(cc)}\right), \\
\Omega_{U,1} &= \frac{1}{2}\sum_{h=-\infty}^{\infty} E\left[tr\left\{\Sigma_{U,t}(h)\Sigma_{U,t}(h)'\right\}\right] \\
\Sigma_{U,t}(h) &= \mu^2\Sigma_{u,11,t}^{(cc)}(h) + \Sigma_{u,22,t}^{(cc)}(h) - \mu\Sigma_{u,12,t}^{(cc)}(h) - \mu\Sigma_{u,21,t}^{(cc)}(h), \qquad h = ..., -1, 0, 1, ...,
\end{aligned}$$

where the upper index (c) *denotes the upper* $(k^c, 1)$ *block of a vector, the upper index* (c, c) *denotes the upper-left* (k^c, k^c) *block of a matrix, and* $\breve{F}_t = [\breve{f}_t^{c\prime}, \breve{f}_{1,t}^{s\prime}, \breve{f}_{2,t}^{s\prime}]' = [(f_t^c - \bar{f}^c)', (f_{1,t}^s - \bar{f}_1^s)', (f_{2,t}^s - \bar{f}_2^s)']'$, *with* $\bar{f}^c = \sum_{t=1}^{T} f_t^c/T$, *and* $\bar{f}_j^s = \sum_{t=1}^{T} f_{j,t}^s/T$, *for* $j = 1, 2$.

Theorem A.1 corresponds to Theorem 1 in AGGR, where estimator of the canonical correlations of the estimated factors (used to compute $\hat{\xi}(k^c)$), and the sample covariance matrix of the true factors $\tilde{\Sigma}_{cc}$ have been modified to take into account that under our new Assumption B.2 the factors are allowed to have a non-zero mean.

To get a feasible distributional result for the statistic $\hat{\xi}(k^c)$, we need consistent estimators for the unknown scalar tr$\left\{\tilde{\Sigma}_{cc}^{-1}\tilde{\Sigma}_U\right\}$ and tr$\left\{\tilde{\Sigma}_{cc}^{-1}\tilde{\Sigma}_B\right\}$, and matrix $\Omega_{U,1}$ in Theorem A.1. To simplify the analysis, we make the simplifying assumptions that the errors $\varepsilon_{j,i,t}$ are (i) uncorrelated across sub-panels j and individuals i, at all leads and lags, and (ii) a conditionally homoscedastic martingale difference sequence for each individual i, conditional on the factor path, that is,

$$\begin{aligned}
Cov(\varepsilon_{j,i,t}, \varepsilon_{k,\ell,t-h}|\mathcal{F}_t) &= 0, \qquad \text{if either } j \neq k, \text{ or } i \neq \ell, \\
E[\varepsilon_{j,i,t}|\{\varepsilon_{j,i,t-h}\}_{h\geq 1}, \mathcal{F}_t] &= 0, \qquad E[\varepsilon_{j,i,t}^2|\{\varepsilon_{j,i,t-h}\}_{h\geq 1}, \mathcal{F}_t] = \gamma_{j,ii} \text{ (say)},
\end{aligned} \tag{A.20}$$

for all j, i, t, h, see Assumption B.9 in the Online Appendix for more details.[19] Then, we have

$$\tilde{\Sigma}_U = \mu_N^2\tilde{\Sigma}_{u,11}^{(cc)} + \tilde{\Sigma}_{u,22}^{(cc)}, \quad \Sigma_U(0) \equiv \Sigma_U = \mu^2\Sigma_{u,11}^{(cc)} + \Sigma_{u,22}^{(cc)}, \quad \Omega_{U,1} = \frac{1}{2}tr\left\{\Sigma_U^2\right\}. \tag{A.21}$$

Matrices $\tilde{\Sigma}_{u,jj}$ and $\Sigma_{u,jj} \equiv \Sigma_{u,jj}(0)$ do not depend on time. In Theorem A.2 we show that, by replacing $\tilde{\Sigma}_{cc}$ with its large sample limit I_{k^c}, and matrix $\tilde{\Sigma}_U$ by a consistent estimator $\hat{\Sigma}_U$ (defined in the Thoerem), the asymptotic distribution of the feasible statistic is unchanged with respect to that of Theorem A.1.

[19] Our simplifying Assumption B.9 is the same as Assumption A.9 in AGGR.

THEOREM A.2. *Let* $\hat{\Sigma}_U = (N_2/N_1)\hat{\Sigma}^{(cc)}_{u,11} + \hat{\Sigma}^{(cc)}_{u,22}$, *with* $\hat{\Sigma}_{u,jj} = \left(\frac{1}{N_j}\hat{\Lambda}_j'\hat{\Lambda}_j\right)^{-1}\left(\frac{1}{N_j}\hat{\Lambda}_j'\hat{\Gamma}_j^*\hat{\Lambda}_j\right)\left(\frac{1}{N_j}\hat{\Lambda}_j'\hat{\Lambda}_j\right)^{-1}$ *where* $\hat{\Lambda}_j = [\hat{\Lambda}_j^c \vdots \hat{\Lambda}_j^s]$, $\hat{\Lambda}_j^c$ *and* $\hat{\Lambda}_j^s$ *are the loadings estimators defined in equations (A.13) and (A.14),* $\hat{\Gamma}_j^* = diag(\hat{\gamma}^*_{j,ii},\ i = 1, ..., N_j)$ *with*

$$\hat{\gamma}^*_{j,ii} := \frac{1}{T}\sum_{t=1}^{T}(\hat{\varepsilon}_{j,i,t} - \bar{\hat{\varepsilon}}_{j,i})^2 \tag{A.22}$$

where $\bar{\hat{\varepsilon}}_{j,i} = \frac{1}{T}\sum_{t=1}^{T}\hat{\varepsilon}_{j,i,t}$, *and* $\hat{\varepsilon}_{j,i,t} = y_{j,i,t} - \hat{\lambda}^{c\,\prime}_{j,i}\hat{f}^c_t - \hat{\lambda}^{s\,\prime}_{j,i}\hat{f}^s_{j,t}$, *for* $j = 1, 2$. *Define the test statistic:*

$$\tilde{\xi}(k^c) := N\sqrt{T}\left(\frac{1}{2}tr\{\hat{\Sigma}_U^2\}\right)^{-1/2}\left[\hat{\xi}(k^c) - k^c + \frac{1}{2N}tr\left\{\hat{\Sigma}_U\right\}\right], \tag{A.23}$$

and let Assumptions B.1 - B.9 hold. Then:
(i) Under the null hypothesis $H_0 = H(k^c)$ *we have:* $\tilde{\xi}(k^c) \xrightarrow{d} N(0,1)$.
(ii) Under the alternative hypothesis $H_1 = \bigcup_{0\le r<k^c} H(r)$, *we have:* $\tilde{\xi}(k^c) \xrightarrow{p} -\infty$.

Theorem A.2 corresponds to Theorem 2 in AGGR, where their original estimator residuals'variance $\hat{\gamma}_{j,ii} = \frac{1}{T}\sum_{t=1}^{T}\hat{\varepsilon}^2_{j,i,t}$ has been substituted by $\hat{\gamma}^*_{j,ii}$ in equation (A.22). As the dependent variables in regressions the (A.13) and (A.14) are not demeaned and both regressions do not include the constant terms, the residuals of these regressions might not be zero mean by construction, and therefore $\hat{\gamma}^*_{j,ii}$ is an appropriate estimator. We also note that if the errors are weakly correlated across series and/or time, consistent estimation of $\tilde{\Sigma}_U$ requires thresholding of estimated cross-sectional covariance and/or HAC-type estimators. Finally, Theorem A.2 remains the same when the estimator $\hat{f}^{c\star}_t$ (or $\hat{f}^{c*}_t$) is used instead of $\hat{f}^c_t$.

ONLINE APPENDIX

"Factors Common to Individual Stock and Sorted Portfolio Returns".

Elena Andreou[1] Patrick Gagliardini[2] Eric Ghysels[3] M. Rubin[4]

Date: February 7, 2022

[1]University of Cyprus and CEPR, e-mail: elena.andreou@ucy.ac.cy
[2]Università della Svizzera italiana, Lugano and Swiss Finance Institute, e-mail: patrick.gagliardini@usi.ch
[3]University of North Carolina - Chapel Hill and CEPR, e-mail: eghysels@unc.edu
[4]EDHEC Business School, Nice, France, e-mail: mirco.rubin@edhec.edu

OA.1 Data Description

We consider three panels of monthly returns in our analysis, namely (i) individual US stock returns from CRSP, (ii) the panel of test asset portfolios from the April 2021 release of the database "Open Source Cross-Sectional Asset Pricing" created by Chen and Zimmermann (2021), CZ21 hearafter, and (iii) the panel of factors from the zoo considered by CZ21.[5] For all three panels, we consider two samples: (i) the chronological time sample which include all data available in each dataset from Jan. 1966 to Dec. 2020, and (ii) the publication time sample which goes from Jan. 1996 to Dec. 2020, where the CZ21 test assets portfolios and factors enter with their publication date in the database. We split the 660 (resp. 300) months in the chronological time (resp. publication time) sample into $B = 11$ (resp. 5) non-overlapping blocks of 60 months, denoted as $b = 1, ..., B$. The first block in the chronological time (resp. publication time) sample is from Jan. 1966 to Dec. 1970 (resp. Jan. 1996 to Dec. 2000) and the last block is from Jan. 2016 to Dec. 2020. Within each block, we consider only a balanced sample of individual stocks and test asset portfolios, that is we only include assets with returns available for all the 60 months. We work with 5-year non-overlapping samples to address the concern of survivorship bias if we were to use the full sample of individual stocks. Similar to the arguments in Kim and Korajczyk (2021), one can view the 5-year span as a compromise between a sample large enough for our test procedure to have desirable small sample properties and the concern of capturing new and disappearing stocks. Figures OA.1 and OA.2 report the number of individual CRSP stocks, test assets portfolios and factors available in each of the B blocks in the chronological time and publication time samples, respectively. Both samples are described in more detail below.

The first chronological time sample panel of test assets consists of individual stocks available from the Center for Research in Security Prices (CRSP) traded on the New York Stock Exchange (NYSE), the American Stock Exchange (AMEX) and the NASDAQ for the period from January 1966 through December 2020. We focus on common stocks (CRSP share codes 10 and 11) and delete all stocks having less than 60 consecutive monthly returns. We end up having an unbalanced panel for the returns of 14948 different stocks. The average cross-sectional size, computed in each month, is about 4270 stocks. In the first (resp. last) block, that is the block 1966-1970 (resp. 2016-2020), we have 1539 (resp. 2668) stocks. The publication time sample is constructed analogously but goes from January 1996 to December 2020. Applying the same filters as above, we end up having an unbalanced panel for the returns of 8131 different stocks. The average cross-sectional size, computed in each month, is about 4170 stocks. In the first (resp. last) block, that is the block 1996-2000 (resp. 2016-2020), we have 3779 (resp. 2668) stocks.

Turning to the test assets portfolios and factors from CZ21, we consider the unbalanced panel of

[5] Data for the "Open Source Cross-Sectional Asset Pricing" project are available on Andrew Chen's website: `https://sites.google.com/site/chenandrewy/open-source-ap`

1215 portfolios formed starting from the 205 firm-level characteristic, or predictors, having predictive ability for firm-level returns according to the four asset pricing meta-studies by McLean and Pontiff (2016), Green, Hand, and Zhang (2017), Hou et al. (2020), Harvey et al. (2016). The returns of the 205 factors in our zoo panel are those of long-short portfolios of the upper and bottom quantile portfolios constructed by sorting stocks according to each characteristic.[6] Following CZ21, we consider test asset portfolios and factors associated only with characteristics classified either as "clearly" or "likely" returns predictors in their study.[7]

In the chronological time sample we include all the quantile portfolios and factors available in the baseline version of the database of CZ21, leading to an unbalanced panel of 1214 portfolios associated with 205 characteristics.[8] The average cross-sectional size, computed in each month, is about 1113 portfolios. In the first (resp. last) block, that is the block 1966-2000 (resp. 2016-2020), we have 855 (resp. 1001) test asset portfolios and 141 (resp. 171) factors.

In the publication time sample we include all the quantile portfolios available in the baseline version of the database of CZ21, after excluding all (binary) portfolios associated to binary characteristics. This leads to 1159 portfolios associated to 177 characteristics.[9] In each 5-years block going from January of year $y - 4$ to December of year y, a factor and the relative test assets portfolios from CZ21 are included for all the dates corresponding to the rolling window only if the paper introducing the factor was published in year $y + 1$, or before.[10] These choices allow us to have in the first rolling window (resp. the last), that is the window 1996-2000 (resp. 2015-2020), 59 (resp. 171) factors, and 276 (resp. 959) test asset portfolios.

Finally, for both samples we also download from Kenneth French website the 5 Fama and French factors: Market, SMB, HML, Operating Profitability (RMW), and Investment Style (CMA), together with the momentum factor (and based on prior 2-12 months returns), and the 1 month risk free rate which is used to compute excess returns for the panels of test assets.

[6]CZ21 construct factors following the methodology of the papers where they have been introduced, therefore most factors are constructed from long-short portfolios of equal-weighted quintiles. Value-weighting or other quantiles are used in the factor construction only for the few papers that emphasize these constructions.

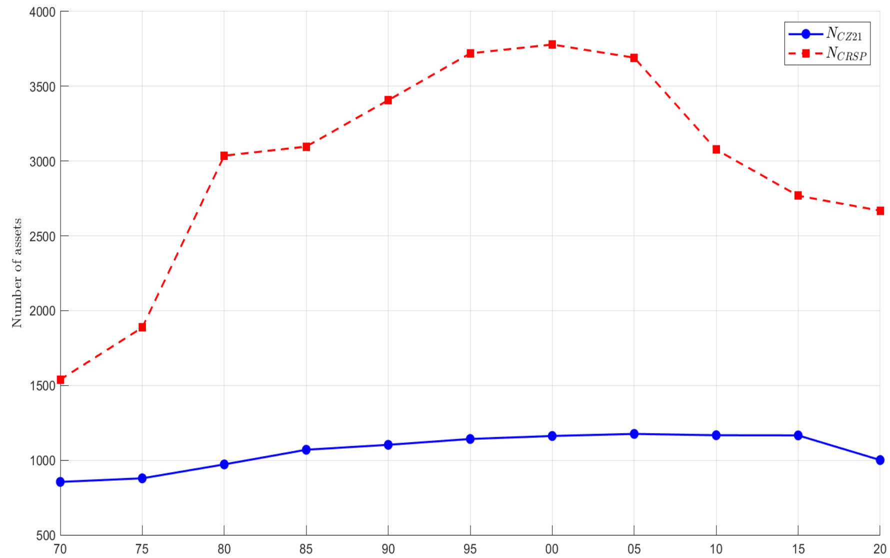
[7]CZ21 define as "clear predictor" a characteristic which is expected to achieve statistically significant mean raw returns in long-short portfolios (e.g. t-stat $>$ 2.5 in a long-short portfolio, monotonic portfolio sort with 80 bps spread, t-stat $>$ 4 in a regression, t-stat $>$ 3 in 6-month event study). On the other hand, a "likely predictor" is a characteristic expected to achieve borderline evidence for the significance of mean raw returns in long-short portfolios (e.g. t-stat $=$ 2.0 in long-short with factor adjustments, t-stat between 2 and 3 in a regression, large t-stat in 3-day event study).

[8]For 28 characteristics only 2 quantile portfolios are available, for 7 characteristics 3 quantile portfolios are available, for 5 characteristics 4 quantile portfolios are available, for 105 characteristics 5 quintile portfolios are available, for 1 characteristic 6 quantile portfolios are available, for 1 characteristic 7 quantile portfolios are available, and finally for 58 characteristics all 10 decile portfolios are available.

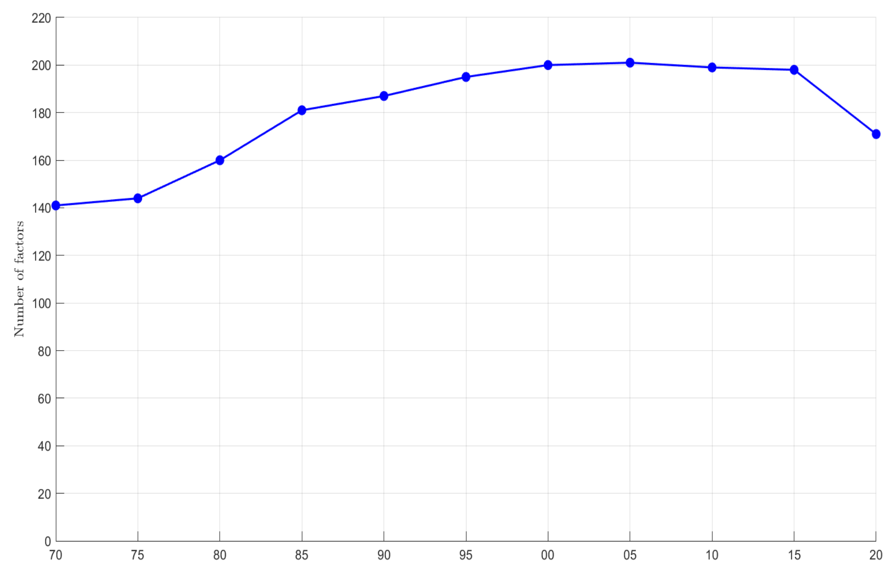
[9]More precisely, for 7 characteristics only 3 quantile portfolios are available, for 5 characteristics 4 quantile portfolios are available, for 105 characteristics 5 quintile portfolios are available, for 1 characteristic 6 quantile portfolios are available, for 1 characteristic 7 quantile portfolios are available, and finally for 58 characteristics all 10 decile portfolios are available.

[10]Publication dates are also available in CZ21.

Figure OA.1: Number test assets and factors in the Zoo, full sample: 1966-2020



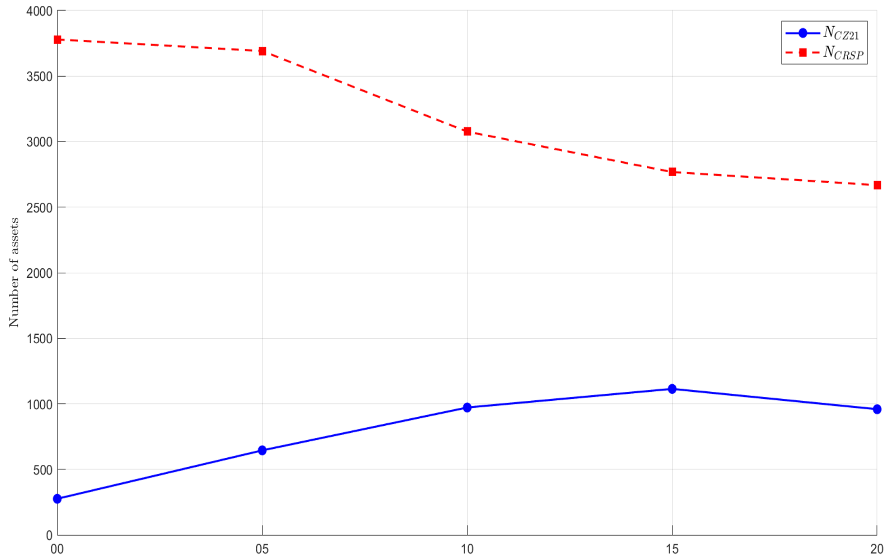
(a) Crsp and CZ21 test assets



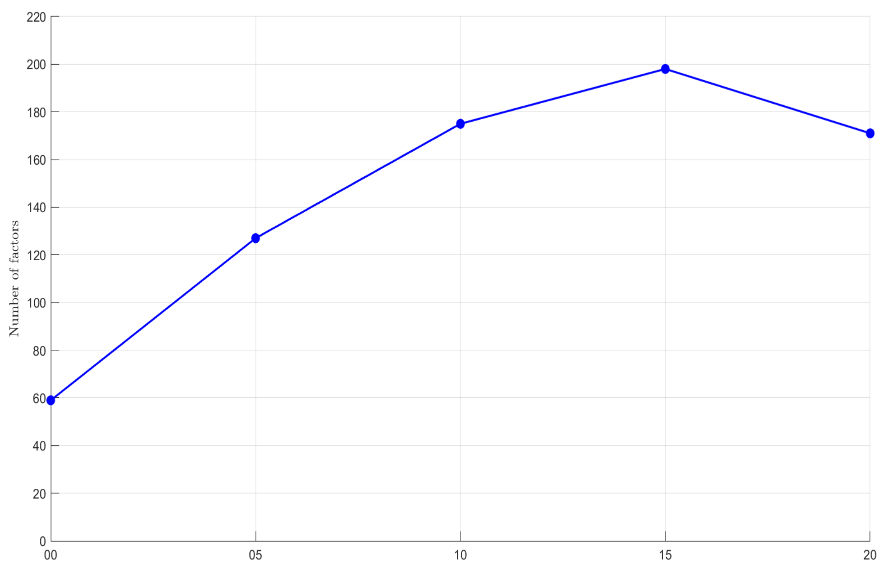
(b) CZ21 Factors in the zoo

Panel (a) displays the number of assets in each balanced panel of CRSP individual stocks (red squares) and CZ21 test assets (blue dots). These two panels of assets are constructed in every year y based on the 5-years non-overlapping window starting in year $y-4$ and ending in year y, for each $y = 1970, ..., 2020$. Panel (b) displays the number of factors in the our zoo, that is the number of facotrs in the CZ21 dataset (blue dots). In every 5-year window we include all assets and factors with non-missing returns for all the 60 months as available in the CZ21 dataset.

Figure OA.2: Number test assets and factors in the zoo, publication time dataset 1996 - 2020



(a) Crsp and CZ21 test assets

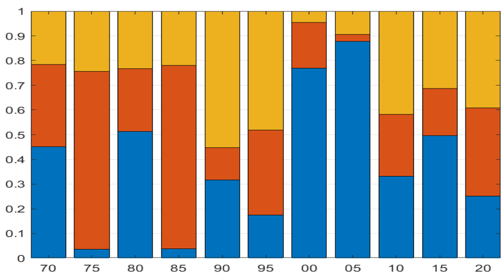


(b) CZ21 Factors in the zoo

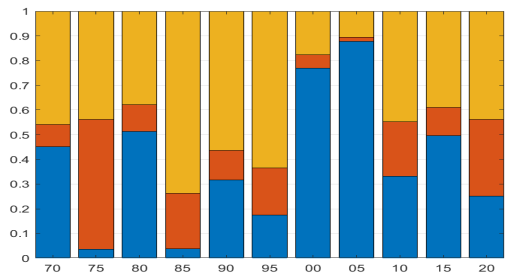
Panel (a) displays the number of assets in each balanced panel of CRSP individual stocks (red squares) and CZ21 test assets (blue dots). These two panels of assets are constructed in every year y based on the 5-years non-overlapping window starting in year $y-4$ and ending in year y, for each $y = 2000, ..., 2020$. Panel (b) displays the number of factors in the our zoo, that is the number of facotrs in the CZ21 dataset (blue dots). In every 5-year window we only include assets and factors with non-missing returns for all the 60 months. Let y_{pub} be the publication year of a certain factor. CZ21 test assets and the corresponding factor are included in our dataset, for all years $y \geq y_{pub}$, that is we include them in our sample only if their full history is available at least from $y_{pub} - 4$.

OA.2 Supplementary empirical results

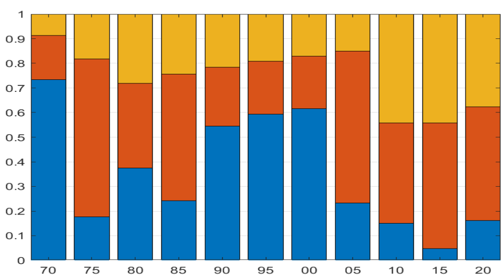
Figure OA.3: Variability of the FF factors RMW, CMA and Momentum explained by common and specific factors in Chen and Zimmermann (2021) test assets and CRSP individual stocks.



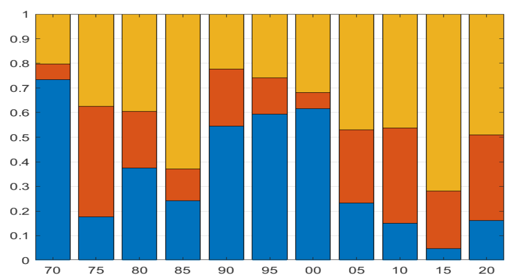
(a) RWM (operating profitability): CZ21 test assets



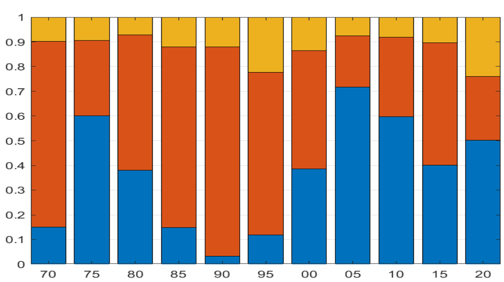
(b) RWM (operating profitability): CRSP stocks



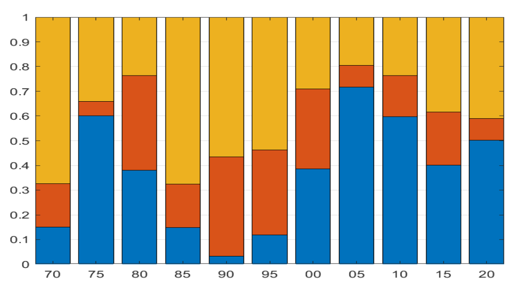
(c) CMA (investment style): CZ21 test assets



(d) CMA (investment style): CRSP stocks



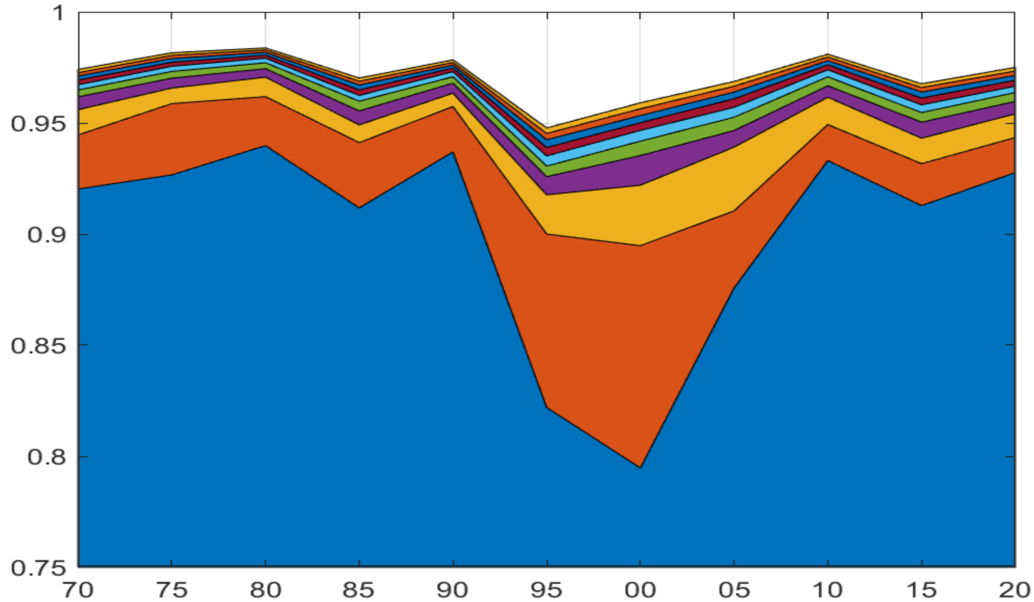
(e) Momentum : CZ21 test assets



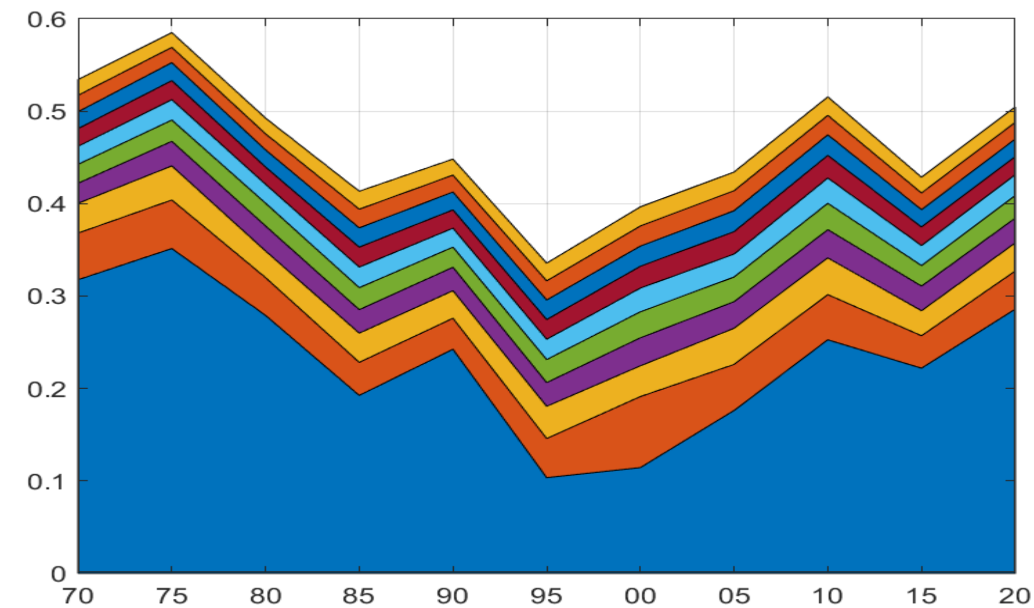
(f) Momentum : CRSP stocks

For each the Fama and French factors and Momentum the figure displays the fraction of variance (R^2) explained by the three common factors between CRSP and CZ21 test assets (blue bars which are the same in both panels), CZ21's group-specific factors (orange bars, left panels), CRSP group-specific factors (orange bars, right panels), and unexplained by common and group-specific factors (yellow bars). For each year y we report results based on the block starting in year *y-4* and ending in year y, for each *y* = 1970, ..., 2020.

Figure OA.4: Average R^2 of first 10 pervasive factors for Chen and Zimmermann (2021) test assets and CRSP individual stocks, full sample: 1966-2020



(a) CZ21 test assets



(b) CRSP individual stocks

Panel (a) [resp. Panel (b)] displays the average fraction of variance (R^2) of the individuals on the balanced panel of CZ21 test assets [resp. CRSP individual stocks] explained by the first 10 RP-PCs extracted from the same panel. The bottom blue area in each panel represents the average R^2 of the first RP-PC, the second (from the bottom) orange area represents the average R^2 of the second RP-PC, and so on. Lettau and Pelger's RP-PCs are computed (fixing $\gamma_{LP} = -1$) on balanced panel of assets. In every year y each panel of assets is constructed for the rolling window starting in year *y-4* and ending in year y, for each *y* = 1970, ..., 2019. In every 5-years rolling window we only include assets with non-missing returns for all the 60 months.

OA.3 Composition of Common Factors

Figures OA.5 - OA.7 show the composition of the three common factors in terms of their rescaled factor-loadings for the test asset portfolios for four different 5 years windows: 1966-1970 (first window in our sample), 1996-2000 (includes dot-com bubble and its burst), 2005-2010 (includes financial crisis), and 2015-2020 (last window, includes Covid). Weights are grouped according to the 34 categories defined by CZ21 listed in alphabetical order (see supplementary files to their paper). The factor weights of the lowest quantile portfolios (e.g. first decile or quintile) are shown in red while those of highest quantile portfolios (e.g. 10^{th} decile or 5^{th}quintile) are in blue. Each bar shows the total weight of a category with the contribution of each quantile-portfolio in the categories. Our group-factor model (2.10) can be written for each date τ in the 5-year window b as:

$$y_\tau = \Lambda_b \, f_\tau + \varepsilon_\tau \, , \qquad \text{with} \qquad \tau \in b \, , \tag{OA.1}$$

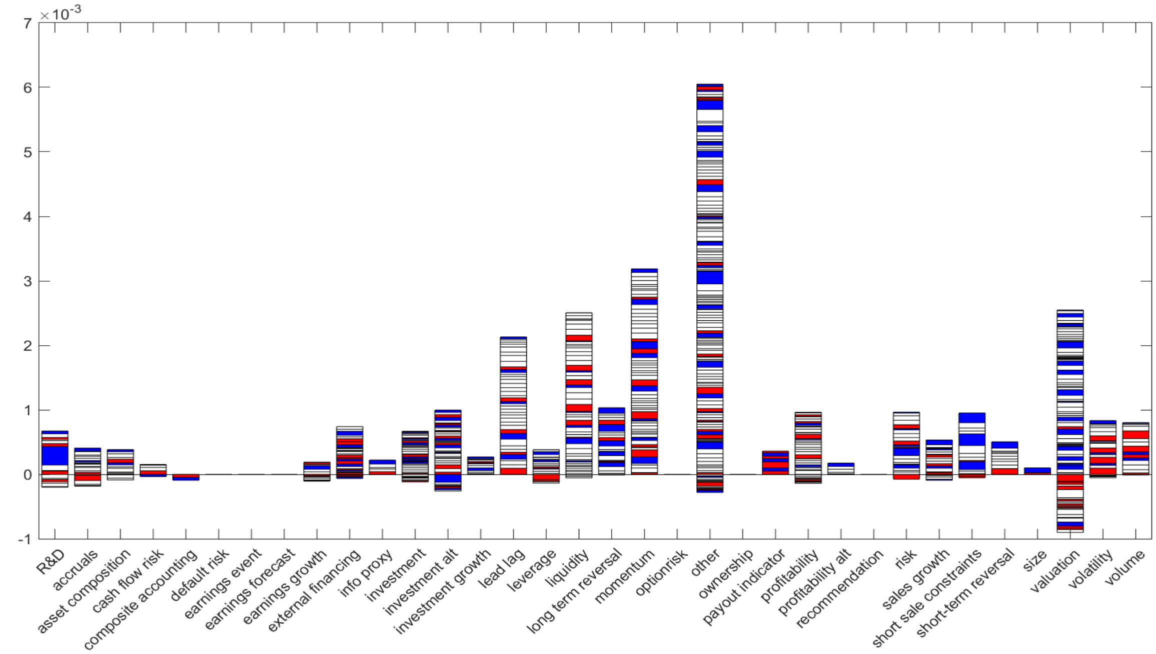
where $y_\tau = [y'_{1,\tau}, \; y'_{2,\tau}]'$, $f_\tau = [f^{c\,\prime}_\tau, \; f^{s\,\prime}_{1,\tau}, \; f^{s\,\prime}_{2,\tau}]'$, $\varepsilon_\tau = [\varepsilon'_{1,\tau}, \; \varepsilon'_{2,\tau}]'$, and

$$\Lambda_b = \begin{bmatrix} \Lambda^c_{b,1} & \Lambda^s_{b,1} & 0 \\ \Lambda^c_{b,2} & 0 & \Lambda^s_{b,2} \end{bmatrix} . \tag{OA.2}$$

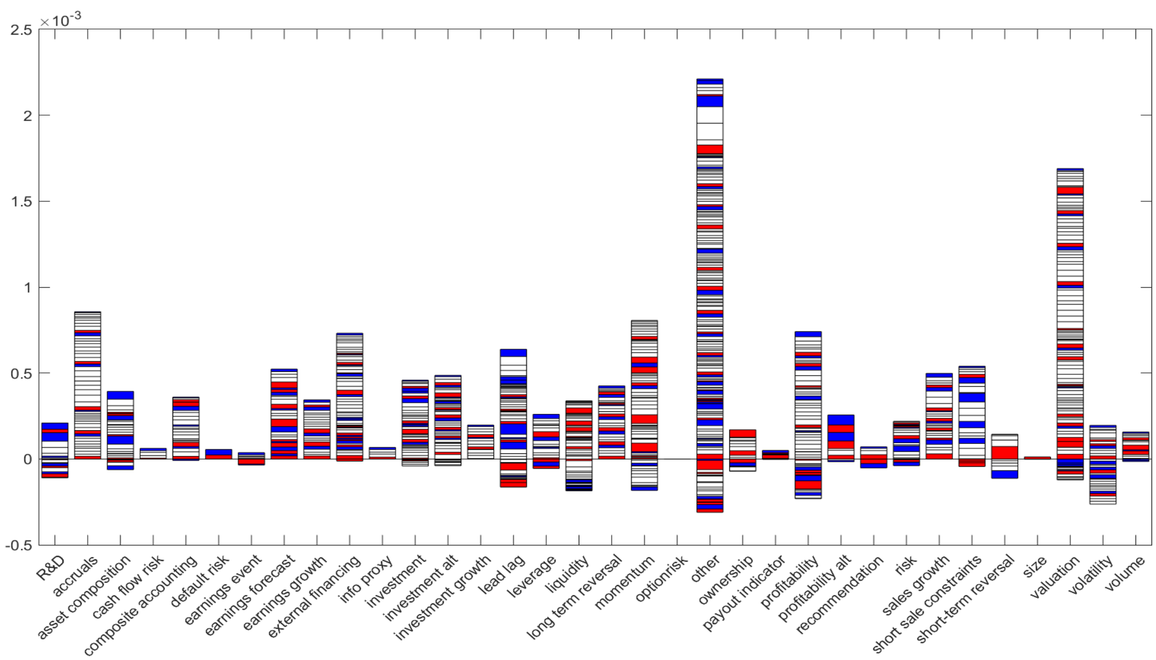
Let $\hat{\Lambda}_b$ be the estimate of Λ_b obtained by the estimation procedure for group-factor model of Section A.5 applied for dates $\tau \in b$. Instead of representing the upper $(N_1 \times k^c)$ block of $\hat{\Lambda}_b$, that is $\hat{\Lambda}^c_{b,1}$, we represent the upper $(N_1 \times k^c)$ block of $\hat{\Lambda}_b(\hat{\Lambda}'_b\hat{\Lambda}_b)^{-1}$, which are the weights of the N_1 test asset portfolios in the k^c portfolios mimicking the common factors $\hat{f}^c_\tau$ when combined with N_2 individual stocks (with weights equal to the lower $(N_2 \times k^c)$ block of $\hat{\Lambda}_b$), that is $(\hat{\Lambda}'_b\hat{\Lambda}_b)^{-1}\hat{\Lambda}'_b \; y_\tau$. This choice allows us to understand the composition of the common factors in terms of all the test asset portfolios. We note that the sign of each of the common factors and the corresponding loadings are not identified, due to the sign indeterminacy of principal components applied group by group and also of the canonical correlation analysis applied on the PCs. At this stage we have not imposed any sign restrictions to represent the loadings.

Figure OA.5 shows that cross the the vast majority of all the loadings of test asset portfolios on the first common factor have the same sign, therefore this factor can be mostly interpreted as a "level" factor. The (absolute value of) correlations of the first common factor with the CRSP-VW index return are 0.88, 0.69, 0.93 and 0.93 in the 4 windows considered, therefore the first common factor does proxy relatively well, but not perfectly, for the aggregate market return.

Figure OA.5: Loadings of first common factor in different 5-year windows



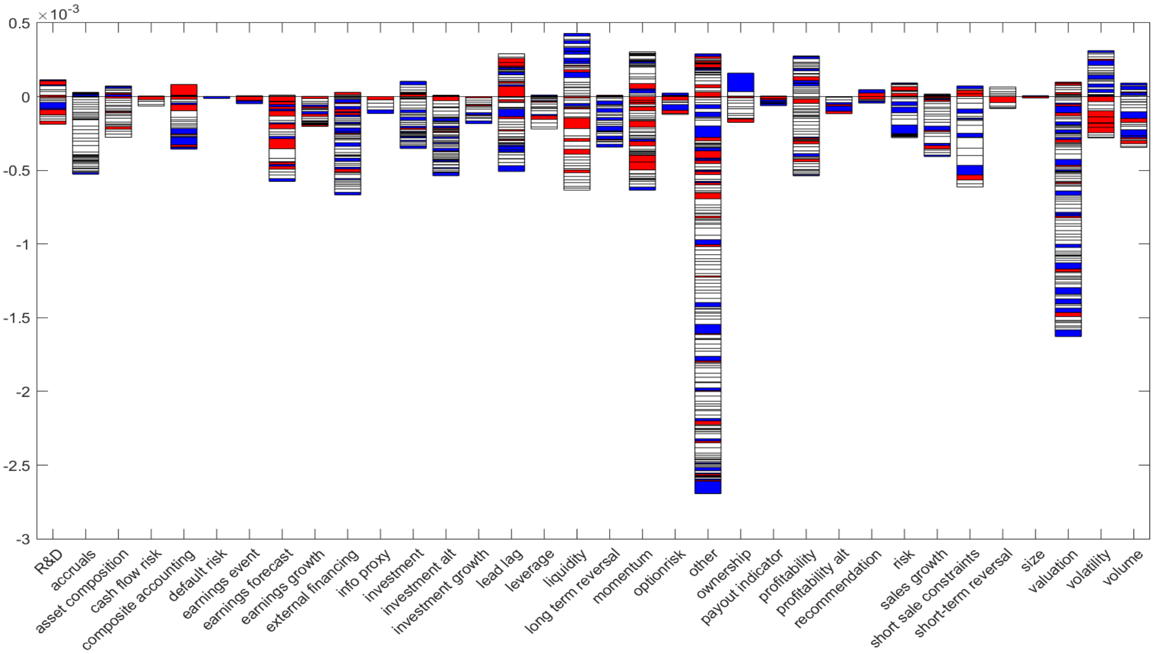
(a) 1966-1970



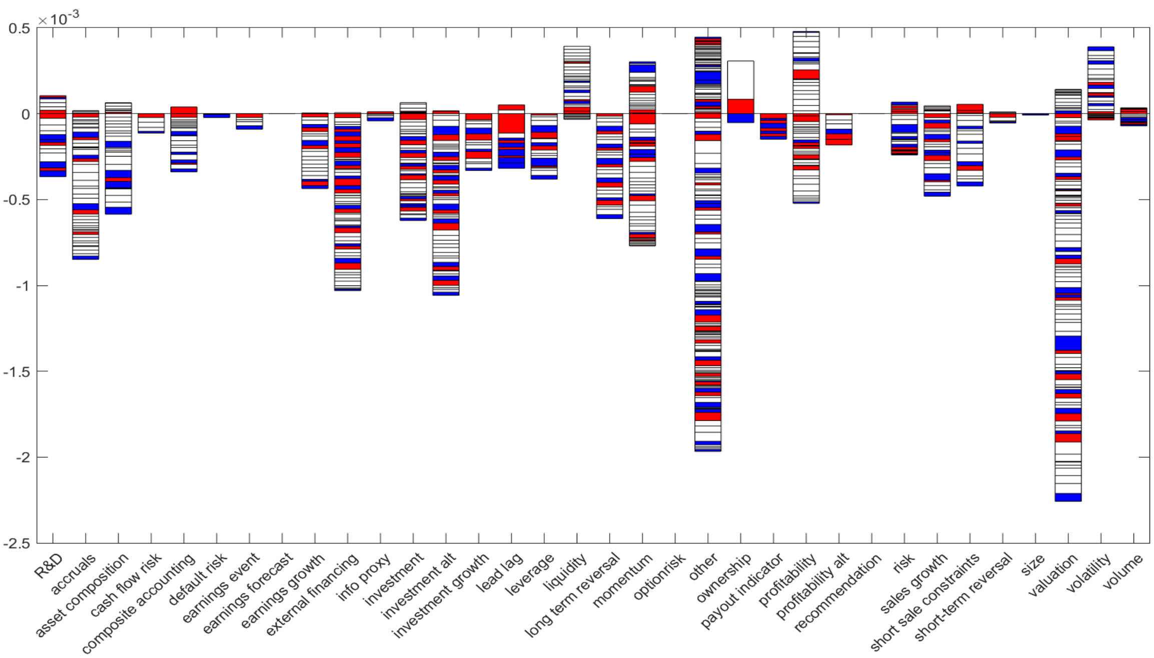
(b) 1996-2000

Continues into next page.

Figure OA.5 (cont'd)



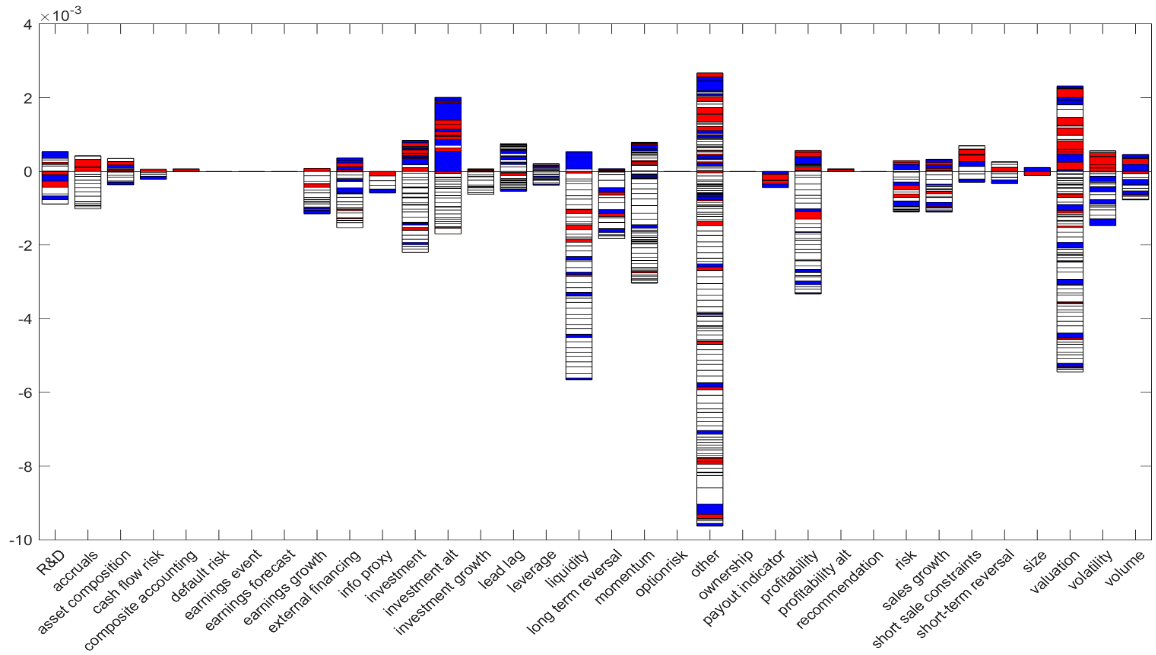
(c) 2005-2010



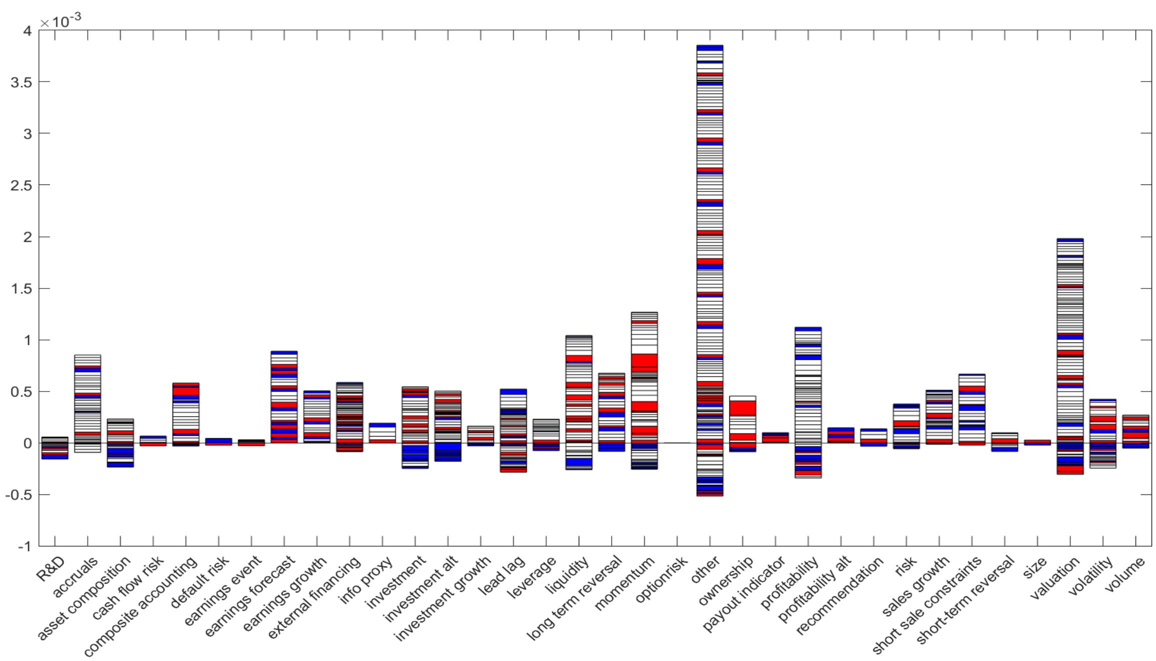
(d) 2016-2020

The figure displays the factor loadings of the first, out of the three, common factor for our group-factor model estimated on the balanced panels of individual stocks and CZ21-portfolios in four different 5-years windows: 1966-1970 in Panel (a), 1996-2000 in Panel (b), 2005-2010 in Panel (c), and 2016-2020 in Panel (d). Weights are grouped according to the 34 categories defined by CZ21 listed in alphabetical order (see supplementary files to their paper and our online appendix). Factor weights of the lowest quantile portfolios (e.g. first decile or quintile) are shown in red while those of highest quantile portfolios (e.g. 10^{th} decile or 5^{th}quintile) are in blue. Each bar shows the total weight of a category with the contribution of each quantile-portfolio in the categories. The loadings of the common factors are computed as described in Section OA.3. No sign restriction is imposed on the sign of the loadings and factors.

Figure OA.6: Loadings of second common factor in different 5-year windows



(a) 1966-1970

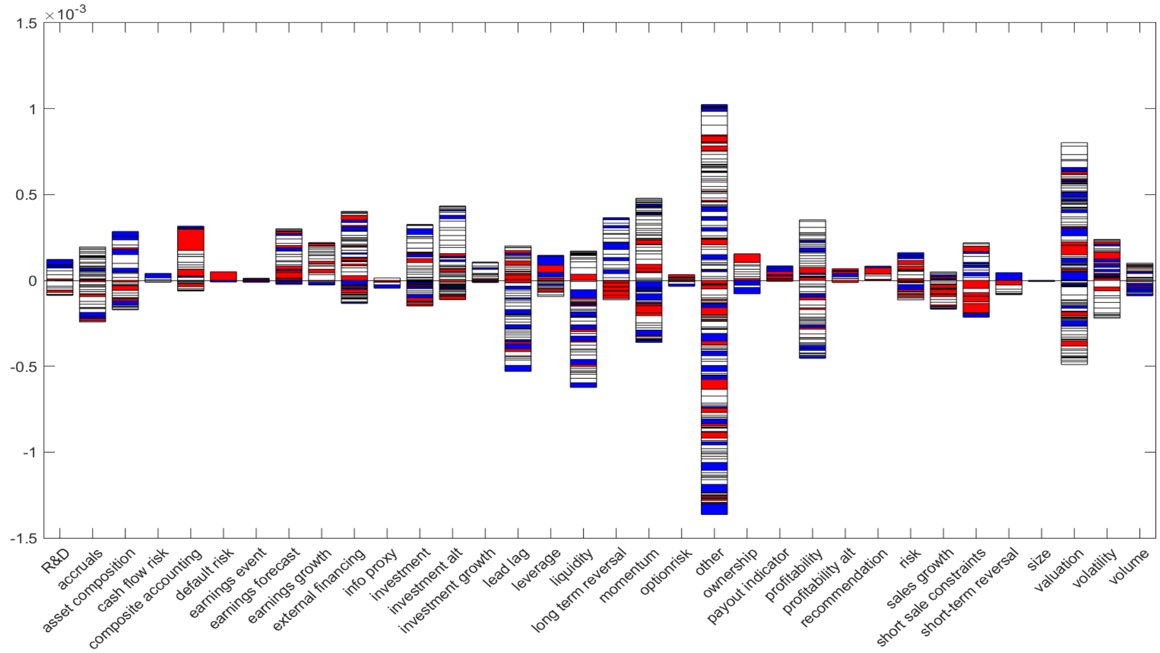


(b) 1996-2000

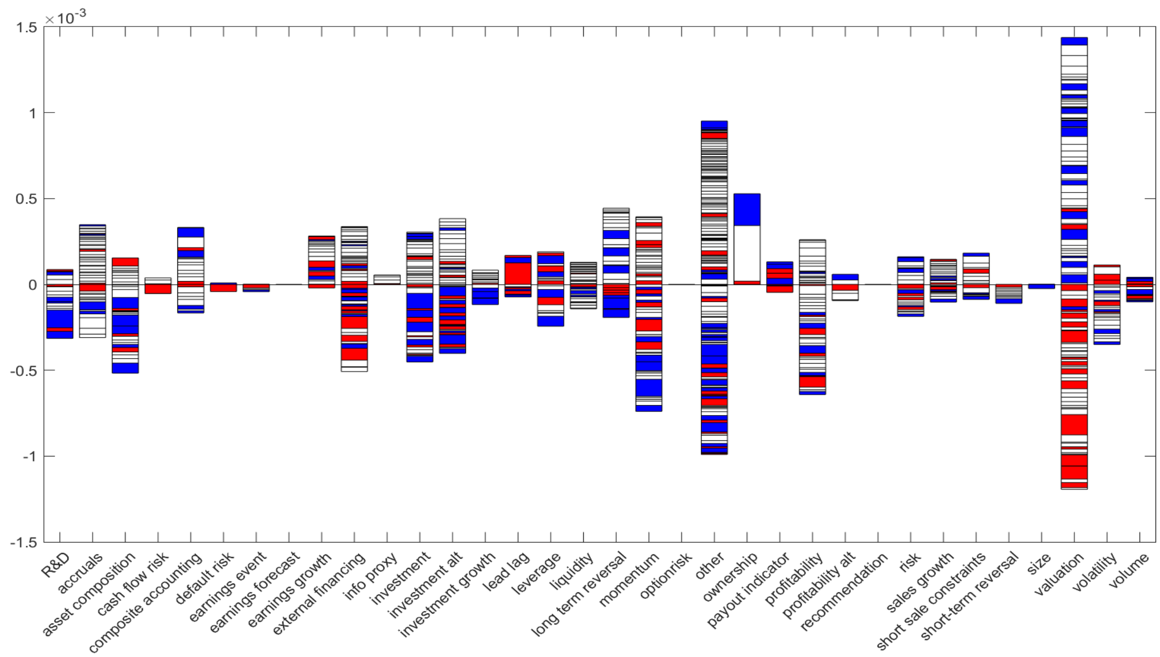
Continues into next page.

Figures OA.6 and OA.7 show the composition of the second and third common factors,

Figure OA.6 (cont'd)



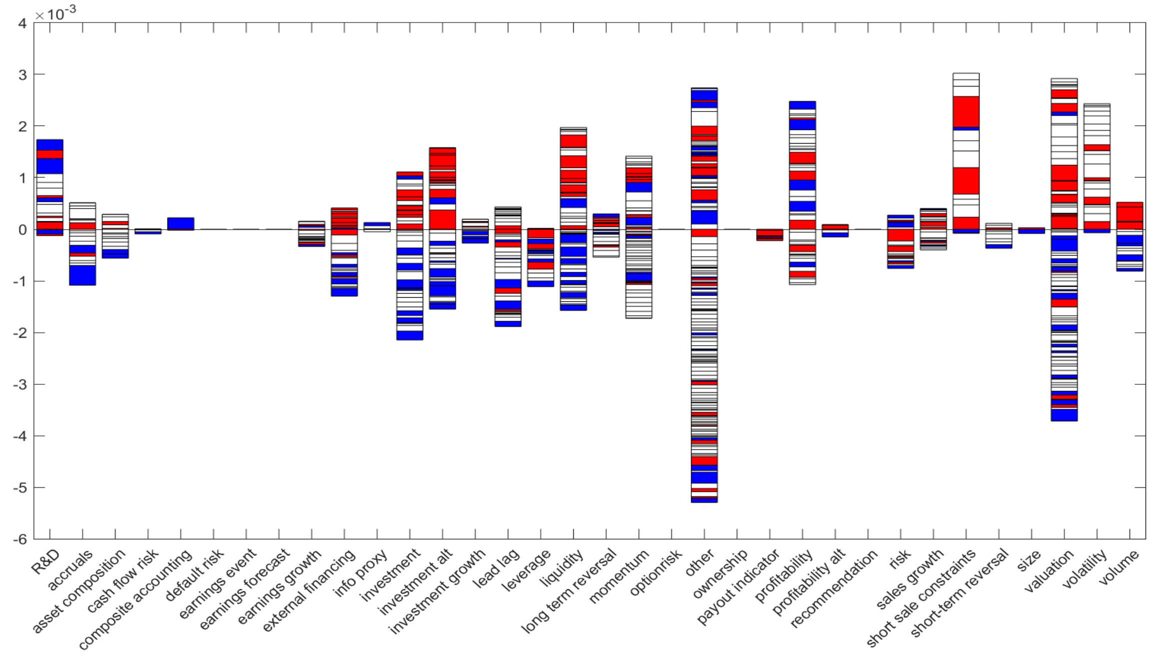
(c) 2005-2010



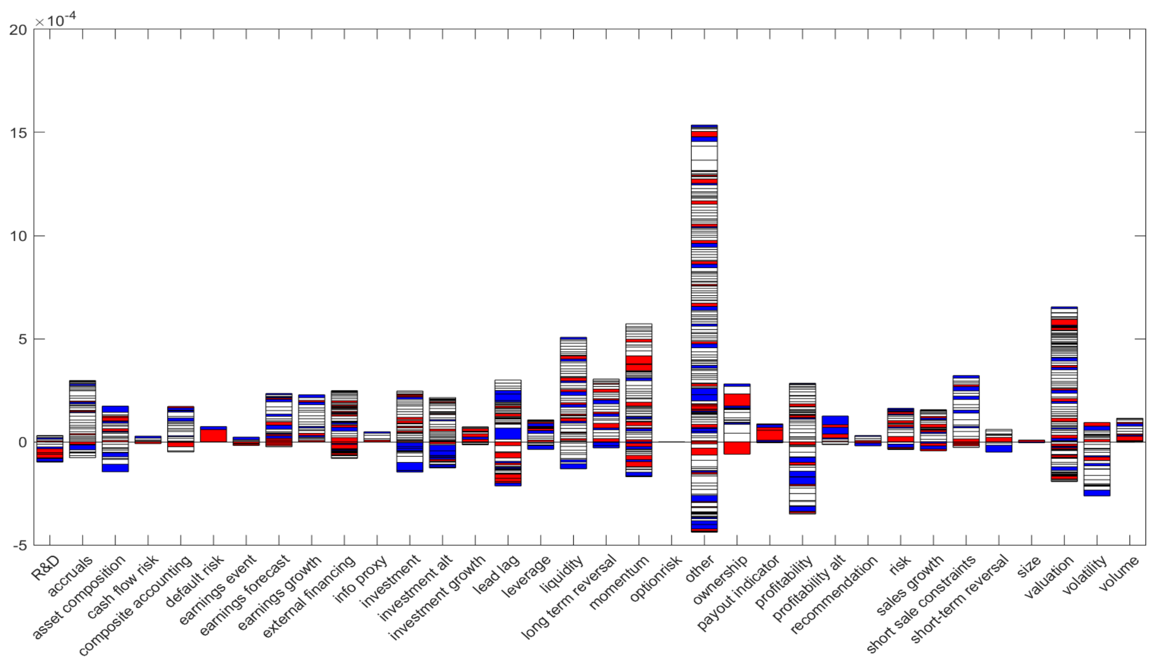
(d) 2016-2020

The figure displays the factor loadings of the second, out of the three, common factor for our group-factor model estimated on the balanced panels of individual stocks and CZ21-portfolios in four different 5-years windows: 1966-1970 in Panel (a), 1996-2000 in Panel (b), 2005-2010 in Panel (c), and 2016-2020 in Panel (d). Weights are grouped according to the 34 categories defined by CZ21 listed in alphabetical order (see supplementary files to their paper and our online appendix). Factor weights of the lowest quantile portfolios (e.g. first decile or quintile) are shown in red while those of highest quantile portfolios (e.g. 10^{th} decile or 5^{th}quintile) are in blue. Each bar shows the total weight of a category with the contribution of each quantile-portfolio in the categories. The loadings of the common factors are computed as described in Section **??**. No sign restriction is imposed on the sign of the loadings and factors.

Figure OA.7: Loadings of third common factor in different 5-year windows



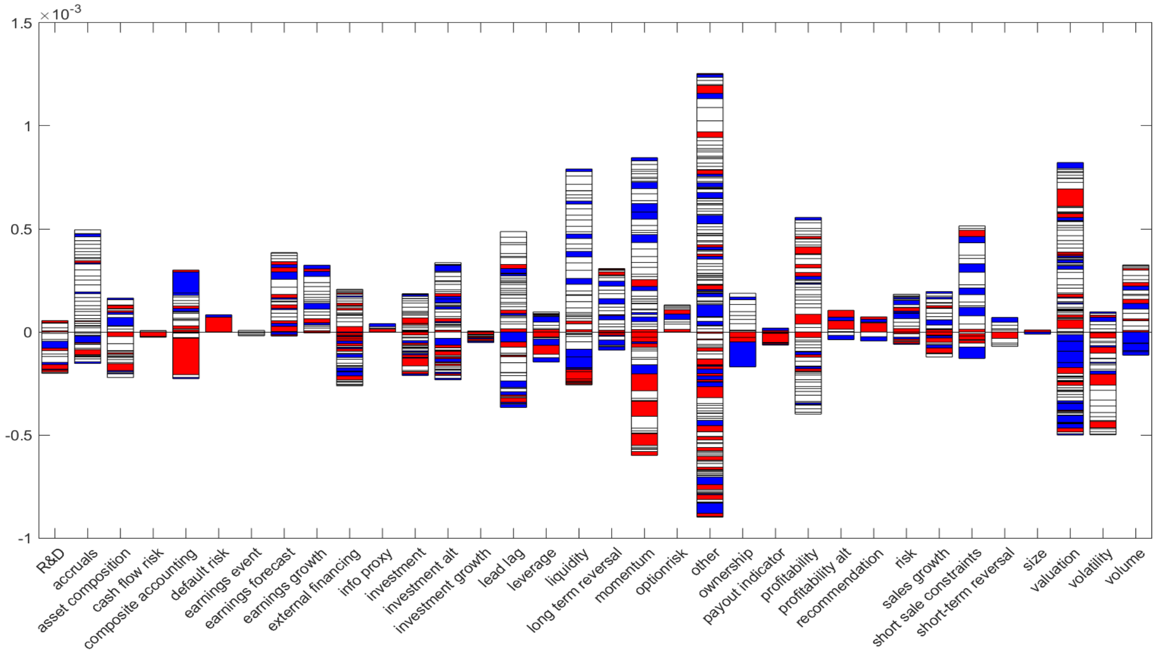
(a) 1966-1970



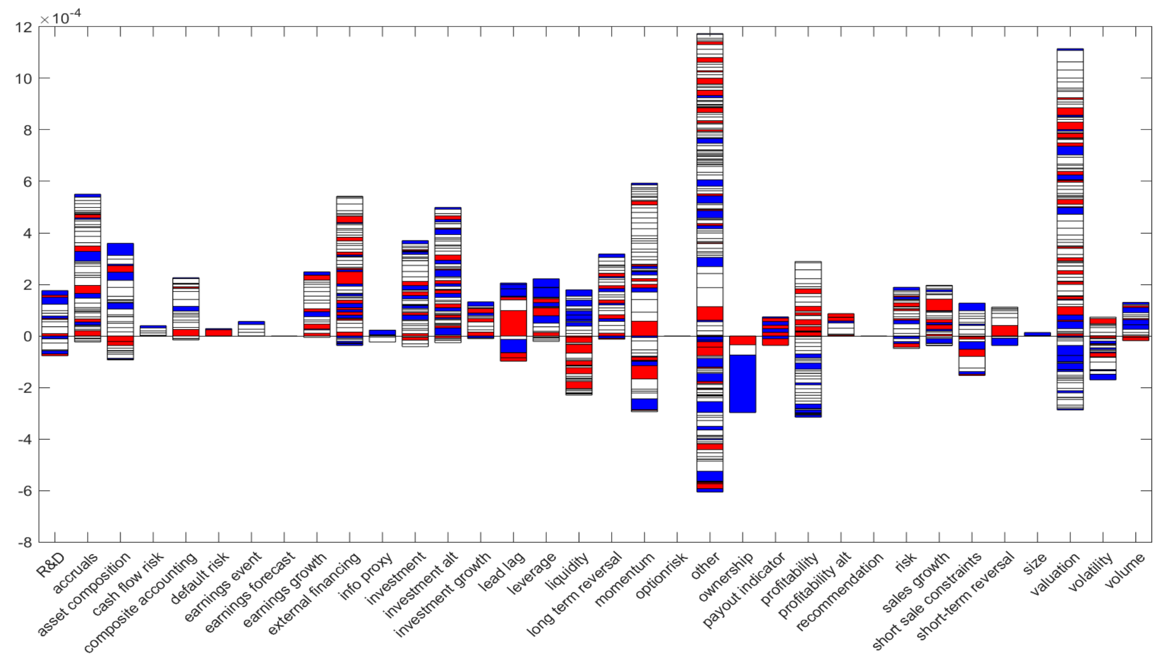
(b) 1996-2000

Continues into next page.

Figure OA.7 (cont'd)



(c) 2005-2010



(d) 2016-2020

The figure displays the factor loadings of the third, out of the three, common factor for our group-factor model estimated on the balanced panels of individual stocks and CZ21-portfolios in four different 5-years windows: 1966-1970 in Panel (a), 1996-2000 in Panel (b), 2005-2010 in Panel (c), and 2016-2020 in Panel (d). Weights are grouped according to the 34 categories defined by CZ21 listed in alphabetical order (see supplementary files to their paper and our online appendix). Factor weights of the lowest quantile portfolios (e.g. first decile or quintile) are shown in red while those of highest quantile portfolios (e.g. 10^{th} decile or 5^{th}quintile) are in blue. Each bar shows the total weight of a category with the contribution of each quantile-portfolio in the categories. The loadings of the common factors are computed as described in Section **??**. No sign restriction is imposed on the sign of the loadings and factors.

respectively. Differently from the mostly the uniformity of sign of the weights of first factor, the second factors is constituted by long and short positions of individual test asset portfolios. In some windows, we can identify groups of categories in which the lowest and largest quantile portfolio loadings (mostly) have opposite signs. For instance, in 1966-1970 and 2016-2020 the second factor seem to have mostly opposite exposures to the extreme quantiles of “valuation”, “volatility” and ”leverage” (last window only). The second common factor in 1996-2000 (resp. the third one in 2005-2010) have sizable exposure of the same signs to the bottom (resp. bottom and top) quantiles of “investment” and “investment alternative” portfolios. Finally, the third factor in 2016-2020 has clear opposite exposures to the extreme quantiles of “profitability” portfolios.

Some caution should nevertheless be placed on this kind of analysis because model (2.10) implies that the k^c common factors are identified in our model up to a rotation, implying that their loadings and their interpretation of each one of the factors can change, depending on which linear combination of them is chosen. It is also possible that applying RP-PC as in Lettau and Pelger (2020b) can generate common factors of a different nature with respect to those we found simply applying PCA in the first step of our estimation.

OA.4 Performance evaluation measures

We describe the various performance evaluation measures both in-sample and out-of-sample starting with the former.

In-sample performance evaluation

Let $N_{j,b}$ be the total number of assets for which the full sample of returns is available in group j and block b, with $j = 1, 2$ and $b = 1, ..., B$. For each model m and for each group of assets j, we compute the following six performance measures across the entire sample, that is across all B blocks:

1. *Total* R^2 of Kelly et al. (2019), which for our model with betas changing across blocks can be expressed as:

$$Tot.\ R_j^2(m) = 1 - \frac{\sum_{b=1}^{B} \sum_{i=1}^{N_{j,b}} \sum_{\tau \in b} \left(y_{j,i,\tau} - \hat{\beta}_{j,i,b}^{m\,\prime} f_{\tau}^{m} \right)^2}{\sum_{b=1}^{B} \sum_{i=1}^{N_{j,b}} \sum_{\tau \in b} y_{j,i,\tau}^2} .$$

It represents the fraction of return variance for all the assets present in group j explained by both the dynamic behavior of the loadings across different blocks, as well as by the contemporaneous factor realizations, aggregated over all assets and all time periods, that is across all B blocks. The

$Total\ R^2$ summarizes how well the systematic factor in a given model specification describes the realized riskiness in the panel of individual stocks. In the case of observable factors, i.e. models in (*i*), the coefficients $\beta^{m\,\prime}_{j,i,b}$ are estimated by an OLS regression without intercept of excess returns on factors, compatible with model (4.13). By construction, the $\beta^{m\,\prime}_{j,i,b}$ and factors for all other models (ii) - (vi) are also estimated by PCA, or variation of it, compatible with a linear model without intercept. [11]

2. *Predictive* R^2 from Kelly et al. (2019), which for our model with betas changing across blocks can be expressed as:

$$Pred.\ R^2_j(m) = 1 - \frac{\sum_{b=1}^{B}\sum_{i=1}^{N_{j,b}}\sum_{\tau\in b}\left(y_{j,i,\tau} - \hat{\beta}^{m\,\prime}_{j,i,b}\bar{f}^m_\tau\right)^2}{\sum_{b=1}^{B}\sum_{i=1}^{N_{j,b}}\sum_{\tau\in b} y^2_{j,i,\tau}}\ .$$

where $\bar{f}^m_\tau = \frac{1}{T_b}\sum_{\tau\in b} f^m_\tau$ is the sample average of the factors'realizations within all the T_b dates in block b only, that is the same block in which the $\beta^{m\,\prime}_{j,i,\tau}$ are estimated compatible with a model without intercept. *Predictive* R^2 represents the fraction of realized return variation explained by the model's description of conditional expected returns, and summarizes the model's ability to describe risk compensation only through exposure to systematic risk. Our measure of the *Predictive* R^2 is slightly different from the in-sample *Predictive* R^2 of Kelly et al. (2019) as ours allows for factor risk premia which vary across different blocks, while theirs imposes constant risk premia across dates.[12]

3. *Pricing error* R^2 of Kelly et al. (2020), which is defined as:

$$Pr.Err.\ R^2_j(m) = 1 - \frac{\sum_{b=1}^{B}\sum_{i=1}^{N_{j,b}}\left(\frac{1}{T_b}\sum_{\tau\in b} y_{j,i,\tau} - \hat{\beta}^{m\,\prime}_{j,i,b} f^m_\tau\right)^2}{\sum_{b=1}^{B}\sum_{i=1}^{N_{j,b}}\left(\frac{1}{T_b}\sum_{\tau\in b} y_{j,i,\tau}\right)^2}\ ,$$

that is the fraction of the squared unconditional mean excess returns that is described by factors and betas. In contrast to the previous two R^2 measures, this focuses on whether the model's

[11] We also consider the $Tot.\ R^2$ *with constant*, an alternative way to compute *Total* R^2, defined as:

$$Tot.\ R^2_{const,j}(m) = 1 - \frac{\sum_{b=1}^{B}\sum_{i=1}^{N_{j,b}}\sum_{\tau\in b}\left(y_{j,i,\tau} - \hat{\alpha}^m_{j,i,b} - \hat{\beta}^{m\,\prime}_{j,i,b} f^m_\tau\right)^2}{\sum_{b=1}^{B}\sum_{i=1}^{N_{j,b}}\sum_{\tau\in b} y^2_{j,i,\tau}}\ ,$$

and where $\hat{\alpha}^m_{j,i,b}$ and $\hat{\beta}^{m\,\prime}_{j,i,b}$ are estimated by regressions including the intercept of the excess returns on the factors. This measure is related to the *idiosyncratic variation* measure considered by Lettau and Pelger (2020a), which they define as the average variance of the residuals after regressing the returns of test assets on the factors and including the intercept in the regression.

[12] Due to the rotational indeterminacy of factors which are re-estimated across different blocks, we cannot impose a constant factor average across different blocks.

fitted values do a good job of explaining assets' average returns. This metric is close in flavor to formal statistical tests (like the GRS test) of whether or not a cross section of test assets' pricing errors are zero.

4. *Average* RMS_α, an alternative to the *Pricing error* R^2, which is computed as the average over different blocks of the RMS_α measure considered by Lettau and Pelger (2020a) and computed block by block. For each block b, group j and model m, RMS_α is computed as $RMS_{\alpha,j,b}(m) = \sqrt{(1/N_{j,b}) \cdot \sum_{i=1}^{N_{j,b}} (\hat{\alpha}_{j,i,b}^m)^2}$, and $\hat{\alpha}_{j,i,b}^m$ is the estimated intercept of the same regressions described in the construction of $Tot.\ R^2$ *with constant*. Then, *Average* $RMS_{j,\alpha}(m)$ $= (1/B) \cdot \sum_{b=1}^{B} RMS_{\alpha,j,b}(m)$. It assesses the model ability to characterize average excess returns of individual assets. Also this measure is close in flavor to formal statistical tests (like the GRS test) of whether or not a cross section of test assets' pricing errors are zero.

Out-of-sample performance evaluation

We implement the out-of-sample version of the $Total\ R^2$, $Pricing\ R^2$ and $Predicitve\ R^2$ where betas and factor loadings, needed to reconstruct the latent factors out-of-sample for date τ in block b are computed using information form the previous block $b-1$. Analogously to Lettau and Pelger (2020b) we also compute the annualized Sharpe Ratio of the "Maximum Sharpe-ratio portfolio" that can be obtained by an optimal (in a mean-variance sense) linear combination of the factors, which are ultimately portfolios of individual stocks. Our out-of-sample performance measures are defined as:

1. ***OOS Total*** R^2, which for our model with betas changing across blocks can be expressed as:

$$OOS\ Tot.\ R_j^2(m) = 1 - \frac{\sum_{b=2}^{B} \sum_{i=1}^{N_{j,b-1}} \sum_{\tau \in b} \left(y_{j,i,\tau} - \hat{\beta}_{j,i,b}^{m\ \prime} f_{\tau|b-1}^m \right)^2}{\sum_{b=2}^{B} \sum_{i=1}^{N_{j,b-1}} \sum_{\tau \in b} y_{j,i,\tau}^2} \ .$$

The beta coefficients $\hat{\beta}_{j,i,b}^m$ are estimated using information available in block $b-1$ only, while returns $y_{j,i,\tau}$ are observed at dates τ in block b. For models with observable factors, $f_{\tau|b-1}^m$ is simply the observed value of the factor at date τ in block b, as all our observable factors are returns of portfolios of individual stocks observed at date τ with weights computed at date $\tau-1$. Instead, when a model includes latent factors, we compute their values at date τ by running cross-sectional regressions of the returns $y_{j,i,\tau}$ for all assets available both in the previous block $b-1$, and in the current one b, on the factor loadings estimated in the previous block $b-1$ only. More specifically, model *v* (resp *vi*) implies that in block b the DGP for the return of individual stocks (resp. test assets) is:

$$y_{j,\tau} = \Lambda_{j,b}^m\ f_\tau^m + \varepsilon_{j,\tau}^m\ , \qquad \text{with} \qquad \tau \in b\ . \tag{OA.3}$$

Let $\hat{\Lambda}^m_{j,b}$ be the PC estimator of matrix $\Lambda_{j,b}$ obtained using the returns of all assets in group j for dates $\tau \in b$. Then, compatible with model (OA.3), factors $f^m_{\tau|b-1}$ are computed as $f^m_{\tau|b-1} = (\hat{\Lambda}^{m\,\prime}_{j,b-1}\hat{\Lambda}^m_{j,b-1})^{-1}\hat{\Lambda}^{m\,\prime}_{j,b-1}\, y_{j,\tau}$, for all dates $\tau \in b$. Analogously, we can write the DGP (2.10), corresponding to models (*iii*) and (*iv*), as in equations (OA.1) and (OA.2). Let $\hat{\Lambda}_b$ be the estimate of Λ_b in equation (OA.1) obtained by the estimation procedure for group-factor model of Section A.5, then $f^m_{\tau|b-1}$ is an appropriate subset of $(\hat{\Lambda}'_{j,b-1}\hat{\Lambda}_{j,b-1})^{-1}\hat{\Lambda}'_{j,b-1}\, y_{j,\tau}$, for dates $\tau \in b$.

2. ***OOS Pricing*** R^2, which can be expressed as:

$$OOS\ Pr.Err.\ R^2_j(m) = 1 - \frac{\sum_{b=2}^{B}\sum_{i=2}^{N_{j,b-1}}\left(\frac{1}{T_b}\sum_{\tau\in b} y_{j,i,\tau} - \hat{\beta}^{m\,\prime}_{j,i,b-1} f^m_{\tau|b-1}\right)^2}{\sum_{b=1}^{B}\sum_{i=1}^{N_{j,b}}\left(\frac{1}{T_b}\sum_{\tau\in b} y_{j,i,\tau}\right)^2},$$

where all quantities are computed as described for the ***OOS Total*** R^2.

3. ***OOS Predictive*** R^2, which can be expressed as:

$$OOS\ Pred.\ R^2_j(m) = 1 - \frac{\sum_{b=2}^{B}\sum_{i=1}^{N_{j,b-1}}\sum_{\tau\in b}\left(y_{j,i,\tau} - \hat{\beta}^{m\,\prime}_{j,i,b}\bar{f}^m_{\tau|b-1,\tau-1}\right)^2}{\sum_{b=2}^{B}\sum_{i=1}^{N_{j,b-1}}\sum_{\tau\in b} y^2_{j,i,\tau}}.$$

where $\bar{f}^m_{\tau|b-1,\tau-1}$ is the sample average of the factor realizations computed over the 60 months ending at date $\tau-1$, and where the factor is reconstructed (if necessary) for each date as described for the computation of the ***OOS Total*** R^2, that is regressing returns in each month τ on loadings estimated in the previous block $b-1$.

4. ***Maximum Sharpe-ratio***, $Max.\ SR$, that is the realized Sharpe Ratio of a portfolio of "factors" (returns) of each model $f^m_{\tau|b-1}$ combined at each date τ in block b with weights $w^m_{f,b} = (\hat{\Sigma}^m_{f,b-1})^{-1}\hat{\mu}^m_{f,b-1}$, where $\hat{\mu}^m_{f,b-1}$ and $\hat{\Sigma}^m_{f,b-1}$ are the sample mean and covariance, respectively, of all factors in model m computed using their observations in block $b-1$. Therefore, both factors and their weights in the Maximum Sharpe Ratio portfolio in block b are computed using the factor loadings estimated in block $b-1$.

Table OA.1: Average RMS_α, and Total R^2 computed from regressions with intercept.

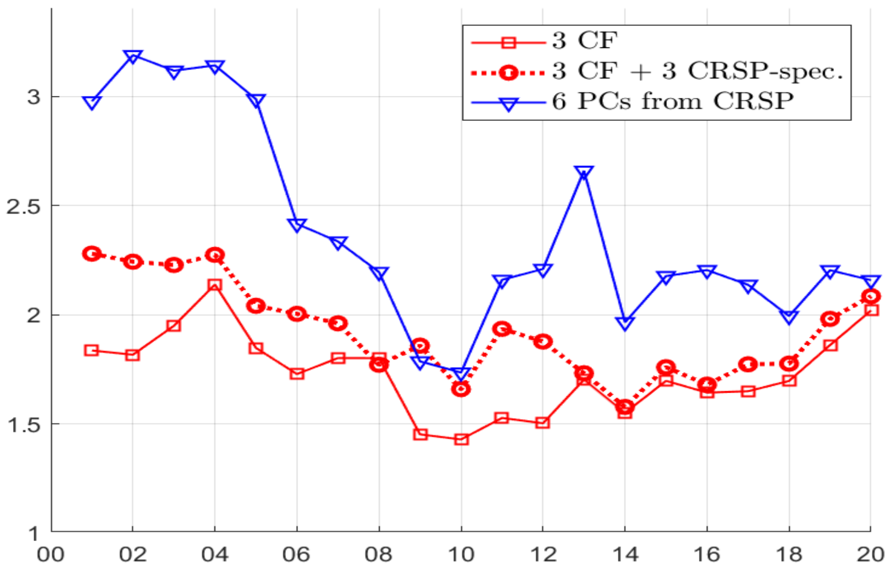
	In-Sample									
	Average RMS_α					**Total R^2, with constant**				
N. of factors, K	1	3	4	5	6	1	3	4	5	6
r: CRSP, f: FF + mom	1.99	1.90	1.91	2.07	2.08	15.86	24.12	26.40	27.79	29.91
r: CZ21, f: FF + mom	0.91	0.60	0.62	0.65	0.66	76.03	90.59	92.35	91.53	93.09
r: CRSP, f: 3CF		1.69					28.18			
r: CZ21, f: 3CF		0.44					93.35			
r: CRSP, f: 3CF + CRSP spec.			1.67	1.67	1.66			31.32	34.02	36.58
r: CZ21, f: 3CF + CZ21 spec.			0.38	0.38	0.33			94.72	95.55	96.10
r: CRSP, f: PCA on CZ21	1.73	1.84	1.86	1.94	1.97	19.22	26.52	28.76	30.92	33.29
r: CZ21, f: PCA on CZ21	0.49	0.37	0.35	0.32	0.31	89.92	94.87	95.54	96.02	96.36
r: CRSP, f: PCA on CRSP	1.73	1.67	1.65	1.63	1.63	18.54	28.09	31.15	33.88	36.41
r: CZ21, f: PCA on CRSP	0.50	0.47	0.47	0.46	0.46	87.13	92.86	93.37	93.99	94.39

The table reports RMS_α (left panel) and the $Total\ R^2$ (right panel) in percent for observable factor models (lines 1-2), a latent factor model with only 3 common factors (lines 3-4), a latent factor model with only 3 common factors between individual stocks and CZ21 portfolios (lines 3-4), a latent factor model with 3 common factors between individual stocks and CZ21 portfolios together with 1, 2, or 3 CRSP-specific factors (line 5), a latent factor model with 3 common factors between individual stocks and CZ21 portfolios together with 1, 2, or 3 CZ21-specific factors (line 6), a latent factor model where the factors are K PCs extracted from the CZ21 portfolios only (lines 7-8), a latent factor model where the factors are K PCs extracted from the CRSP individual stocks only (lines 9-10). Observable factor model specifications are CAPM, FF3, FF3 + Momentum, FF5, and FF5 + Momentum in the $K = 1$,3 ,4 ,5 ,6 columns, respectively. The models are estimated on the rolling window starting in year $y-4$ and ending in year y, for each $y = 1970$, ..., 2020. Both RMS_α and the $Total\ R^2$ are computed in-sample either for the excess returns of individual stocks (r : CRSP) or CZ21 portfolios (r : CZ21) as described in Section 4.2, therefore taking into account all the estimation windows. All the linear models used to compute the $Total\ R^2$ include an intercept, differently from the models used to compute the RMS_α, and from those used to produce the results in Table 1.

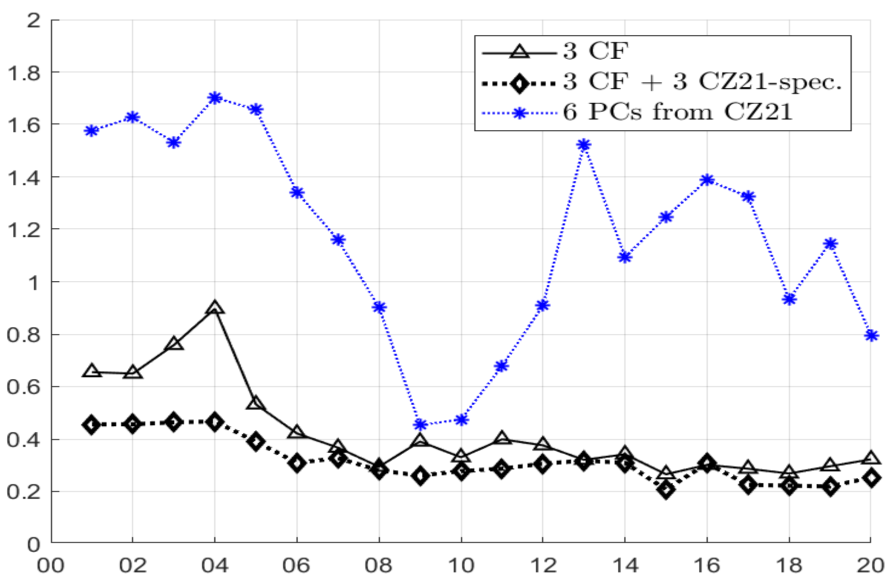
OA.5 Old and New Factors - Supplementary Results

Feng et al. (2020) find the following factors as having incremental contributions to the pricing of the cross-section (with our R^2s and $|\alpha|$ with respect to 3 common and 3 CZ21-specific factors appearing in parenthesis): growth in long term net operating assets from Fairfield, Whisenant, and Yohn (2003, $R^2 = 53\%$, $|\alpha| = 0.00\%$), net operating assets from Hirshleifer, Hou, Teoh, and Zhang (2004, $R^2 = 73\%$, $|\alpha| = 1.20\%$), three-year investment growth from Anderson and Garcia-Feijoo (2006, $R^2 = 65\%$, $|\alpha| = 0.79\%$), net external finance from Bradshaw, Richardson, and Sloan (2006, $R^2 = 92\%$, $|\alpha| = 1.23\%$), revenue surprise from Kama (2009, $R^2 = 56\%$, $|\alpha| = 0.89\%$, for the characteristic "Revenue Surprise" which enters the CZ21 database with the paper of Jegadeesh and Livnat (2006)), betting against beta from Frazzini and Pedersen (2014, $R^2 = 95\%$, $|\alpha| = 1.64\%$), robust minus weak from Fama and French (2015, $R^2 = 92\%$, $|\alpha| = 0.90\%$), for the characteristic "operating profits / book equity" entering the CZ21 database with the paper Fama and French (2006).

Figure OA.8: RMS_α from old factors



(a) Test assets: CRSP individual stocks



(b) Test assets: CZ21 portfolios

For each 5-years rolling window ending in year y we compute the percentage RMS_α generated by a linear factor model with 3 common factors between individual stocks and CZ21 portfolios only (model (i)), with 3 common factors between individual stocks and CZ21 portfolios together with three CZ21-specific factors (model (ii)), and by linear factor model where the factors are the first six PCs from the old factor zoo (model(iii)). Panel (a) displays results considering as test assets individual stocks: we report the RMS_α for model (i) as red squares, for model (ii) as red circles and for model (iii) as blue downward triangles. Panel (b) displays results considering as test assets CZ21 portfolios: we report the RMS_α for model (i) as black upward triangles, for model (ii) as black diamonds and for model (iii) as blue stars. The models are estimated on the rolling window starting in year $y - 4$ and ending in year y, for each $y = 2001, ..., 2020$. RMS_α are computed as described in Section 4.2, but taking into account only the 5-year window ending in year y, and are reported in Panels (a) and (b) respectively.

HXZ Investment and HXZ profitability from Hou, Xue, and Zhang (2015), were present in older versions of the CZ21 database as "Change in Return on Assets" and "Change in Return on Equity" entering 2018 with the paper Hou et al. (2020)). In the 2021 version of the database they are attributed to Balakrishnan, Bartov and Faurel (2010), but are classified as "Indirect Signals" (see Section 2.4 and their Table 4), i.e. are modifications of other characteristics showing"only suggestive evidence of predictive power (e.g. correlated with earnings/price, modified version of a different characteristic, in-sample evidence only)", and therefore are not available. The characteristics "industry-adjusted change in employees", "industry-adjusted size" both from Asness, Porter, and Stevens (2000) are not present in the CZ21 dataset, while "volatility of liquidity (dollar trading volume)" from Chordia, Subrahmanyam, and Anshuman (2001) although in the CZ21 dataset, by construction are not present in the rolling windows we consider in the current version of our analysis.

These results indicate that some of the new factors could provide additional information with respect to the old ones in explaining US stock returns. To assess this issue, in the spirit of the cross-sectional pricing exercises of Feng et al. (2020), we investigate which factors contribute to substantial improvement of either $TotalR^2$ or RMS_α for the panels of individual stocks and/or test asset portfolios when added to the three common factors between individual stocks and the old CZ21 portfolios, and first three old CZ21-specific factors, that is we assess the ability of the new factors to explain the variability and the mean of the returns of both groups of test assets when added to those six factors. The exercise is somewhat similar the the two pass estimator in Feng et al. (2020). Their procedure combines the double-selection LASSO method of Belloni et al. (2014) with the Fama and MacBeth (1973) two-pass regressions to evaluate the contribution of a factor to explaining asset prices specifically in a high-dimensional setting. The results in their Table 2 show that while most of the new factors are redundant relative to the existing factors, a few have statistically significant explanatory power beyond the hundreds of factors proposed in the past. Our procedure differs from the two above papers in that we approach dimensionality reduction differently. We rely on a relatively small number of factors, that is the three common factors and the first three CZ21-specific factors, as we have shown they perform better than PCs form the factor zoo in explain test assets returns.

The double-selection estimator of Feng et al. (2020) is a Fama-MacBeth double machine learning regularized regression approach. The double-selection procedure of Feng et al. (2020) and our regressions result in analogous findings: the majority of factors in the zoo are redundant and only few of them contain genuine new pricing information, as detailed below. In Table 2 of Feng et al. (2020) a total of twelve out of roughly one hundred factors over the sample 2000 until 2015 appear significant and robust according to their testing procedure.

From the results in our Table OA.5, we first note that the increase of $TotalR^2$ generated by the addition of new factor to the 6 old common and CZ21-specific factors is relatively small compared

Table OA.2: Variability of factors in the zoo explained by 3 common factors between CRSP and the old CZ21 test assets, and 3 first three group-specific factors old CZ21 test assets: smallest values.

New factor	R^2
Consensus Recommendation (2002)	24.5
Down forecast EPS (2002)	32.2
Up Forecast (2002)	3.9
Pastor-Stambaugh liquidity beta (2003)	31.3
Change in recommendation (2004)	17.9
Active shareholders (2005)	20.1
Inst own among high short interest (2005)	25.6
Systematic volatility (2006)	24.7
Earnings surprise of big firms (2007)	20.5
Change in Asset Turnover (2008)	23.1
Change in Net Working Capital (2008)	15.7
Customer momentum (2008)	15.3
Off season reversal years 6 to 10 (2008)	28.2
Off season reversal years 11 to 15 (2008)	23.0
Off season reversal years 16 to 20 (2008)	6.8
Return seasonality years 6 to 10 (2008)	5.6
Return seasonality years 11 to 15 (2008)	23.3
Return seasonality years 16 to 20 (2008)	13.0
Return seasonality last year (2008)	25.0
Return seasonality years 2 to 5 (2008)	10.9
Customers momentum (2010)	18.9
Suppliers momentum (2010)	32.1
Real estate holdings (2010)	29.0
Percent Operating Accruals (2011)	22.0
Put volatility minus call volatility (2011)	11.8
Inventory Growth (2012)	24.3
Dividend seasonality (2013)	14.2
Organizational capital (2013)	23.8
R&D ability (2013)	9.1
Growth in advertising expenses (2014)	21.7

For each 5-years rolling window ending in year y we regress each of the new factors (entering the database in year y) on the three common factors between individual stocks and the old CZ21 portfolios, and first three old CZ21-specific factors, that is those computed using CZ21-portfolios available only in year $y-1$. The Table displays the name of Factors with a value $R^2 < 35\%$ in these regressions in chronological order of publication, together with the value of the R^2.The publication date in parenthesis next to each factor. We consider years $y = 2001, 2002, ..., 2020$.

Table OA.3: Absolute value of the intercept of the factors in the zoo when regressed on the 3 common factors between CRSP and the old CZ21 test assets, and 3 first three group-specific factors old CZ21 test assets: largest 15 absolute values and their t-stat.

New factor	$\lvert\alpha_i\rvert$	(tstat)
Consensus Recommendation (2002)	2.0	(2.6)
Probability of Informed Trading (2002)	1.3	(2.6)
Pastor-Stambaugh liquidity beta (2003)	1.3	(2.6)
Idiosyncratic risk (AHT) (2003)	1.3	(1.8)
Firm Age - Momentum (2004)	1.9	(1.8)
Net Operating Assets (2004)	1.5	(3.7)
Inst own among high short interest (2005)	1.7	(1.0)
Mohanram G-score (2005)	1.3	(2.6)
Analyst earnings per share (2006)	1.7	(4.2)
Net equity financing (2006)	1.3	(4.1)
Net external financing (2006)	1.5	(3.8)
Industry return of big firms (2007)	1.7	(3.0)
Efficient frontier index (2009)	1.2	(5.1)
Intermediate Momentum (2012)	1.4	(2.0)
Frazzini-Pedersen Beta (2014)	1.9	(-5.7)

For each 5-years rolling window ending in year y we regress each of the new factors (entering the database in year y) on the three common factors between individual stocks and the old CZ21 portfolios, and first three old CZ21-specific factors, that is those computed using CZ21-portfolios available only in year $y-1$. The Table displays the name of factors with with the largest absolute values of the intercept in these regressions, in chronological order of publication, together with absolute value of the intercept (α_i). The publication date in parenthesis next to each factor. The *t-statistics* for the test of significance of α_i is computed using OLD standard errors are reported in parenthesis. We consider years $y = 2001, 2002, ..., 2020$.

Table OA.4: Regressions of significant factors in table 2 of Feng et al. (2020) on 3 common factors between CRSP and the old CZ21 test assets only, and on the 3 common factors together with 3 first three group-specific factors old CZ21 test assets.

New factor	R^2 on 3 CF	R^2 on 3 CF and 3 CZ21-factors	$\lvert\alpha_i\rvert$ on 3 CF and 3 CZ21-factors	(tstat)
Growth in long term operating assets (2003)	50.28	53.95	0.03	(0.1)
Net Operating Assets (2004)	33.92	77.29	1.50	(3.7)
Change in capex (three years) (2006)	30.14	71.82	0.61	(-2.6)
Net external financing (2006)	88.23	91.84	1.45	(3.8)
Revenue Surprise (2006)	51.31	54.84	1.02	(4.3)
Frazzini-Pedersen Beta (2014)	94.12	94.43	1.86	(-5.7)
operating profits / book equity (2006)	92.43	93.12	0.96	(3.6)
Asset growth (2008)	64.34	87.72	0.16	(-0.7)

We regress each of the significant factors in table 2 of Feng et al. (2020) in the 5-years rolling window ending in year y, with y being the date in which the factor enters in our database, on the three common factors between individual stocks and the old CZ21 portfolios only (regression (a)), and on the three common factors together with the first three old CZ21-specific factors (regression (b)), that is those computed using CZ21-portfolios available only in year $y-1$. The Table displays the name of factors, their publication date, the R^2 of regressions (a) and (b), and the absolute value of the intercept (α_i) in regression (b). The *t-statistics* for the test of significance of α_i is computed using OLD standard errors are reported in parenthesis.

the to $Total\ R^2$ achievable with the 6 old factors only: the increase in $Total\ R^2$ ranges from 1.1% to 2.51% (resp. 0.06% to 0.38%) for individual stocks (resp. CZ21 portfolios) compared to a $Total\ R^2$ achievable with the 6 old factors, which ranges from 26% to 38% (resp. 94% to 97%). Analogous considerations can be made when looking at the RMS_α, which, in the best 10 cases, exhibit a decrease ranging form 0.01 to 0.06 (resp. 0.04 to 0.02) for individual stocks (resp. CZ21 portfolios) compared to a RMS_α achievable with the 6 old factors, which ranges from 1.88 to 2.58 (resp. 0.27 to 0.45). Second, none of the factors which appear as the best 10 ones in decreasing the RMS_α of CZ21 portfolios, appear also as the best 10 ones in decreasing the RMS_α of individual stocks, and only 2 factors, namely Net external financing (2006) and Book-to-market and accruals (2004) appear to be among the top 10 contributors to $Total\ R^2$ both for CZ21 portfolios and individual stocks. Moreover, two factors, namely Analyst earnings per share (2006), Frazzini-Pedersen Beta (2014), although they seem to improve the pricing of the CZ21 portfolios, if anything they seem to be detrimental for the pricing of individual stocks, as they increase the RMS_α.

Out of the 18 factors with large absolute value of α when regresses on the 6 old factors, we find that Analyst earnings per share (2006), Frazzini-Pedersen Beta (2014), and Net equity financing (2006) are among those generating the largest decrease in RMS_α for CZ21 portfolios. Interestingly, the last two are also among the significant factors identified with the methodology of Feng et al. (2020). Out of these three factors, only Net equity financing (2006) appears as one of the top 10 contributors to the increase of $Total\ R^2$ for CZ21 portfolios. Additionally, Revenue surprise (2006) is the only other factor found significant by Feng et al. (2020) which also appears in the top 10 contributors to the decrease in RMS_α of the CZ21 portfolios.

Out of the 34 factors with small value of R^2 when regresses on the 6 old factors, we find that none are among those generating the largest decrease in RMS_α for CZ21 portfolios, while Inst own among high short interest (2005), Inventory Growth (2012) are among those generating the largest decrease in $Total\ R^2$ for CZ21 portfolios. Moreover, Systematic volatility (2006), Suppliers Momentum (2010), Organizational Capital(2013) are among those generating the largest decrease in RMS_α for individual stocks, while Percent Total Accruals (2011) is among those generating the largest decrease in $Total\ R^2$ for individual stocks.

Table OA.5: Changes in $Total\ R^2$ and RMS_α due to new factors.

Panel A: $Total\ R^2$, test assets: CRSP stocks

New factor	Old fac.	ΔR^2
Top 10		
Growth in book equity (2010)	37.41	2.58
Investment to revenue (2004)	35.45	2.15
Employment growth (2014)	39.68	2.11
Volatility smirk near the money (2010)	37.41	1.96
Inventory Growth (2002)	34.53	1.91
Growth in advertising expenses (2014)	39.68	1.90
Taxable income to income (2004)	35.45	1.87
Maximum return over month (2010)	37.41	1.84
Cash to assets (2012)	39.01	1.83
Net external financing (2006)	34.20	1.82
Bottom 10		
Inst own among high short interest (2005)	38.56	1.21
Price delay coeff (2005)	38.56	1.20
Leverage component of BM (2007)	29.94	1.19
Inst Own and Market to Book (2005)	38.56	1.17
Dividend seasonality (2013)	38.97	1.17
change in net operating assets (2004)	35.45	1.17
Breadth of ownership (2002)	34.53	1.17
gross profits / total assets (2013)	38.97	1.16
Brand capital investment (2014)	39.68	1.12
Return seasonality last year (2008)	26.86	1.11

Panel B: RMS_α, test assets: CRSP stocks

New factor	Old fac.	ΔRMS_α
Top 10		
Organizational capital (2013)	1.88	-0.05
Conglomerate return (2012)	1.93	-0.03
Suppliers momentum (2010)	1.86	-0.02
Volatility smirk near the money (2010)	1.86	-0.02
Cash Productivity (2009)	1.77	-0.01
Earnings consistency (2009)	1.77	-0.01
Change in long-term investment (2005)	2.27	-0.01
Systematic volatility (2006)	2.04	-0.01
Inst Own and Turnover (2005)	2.27	-0.01
Growth in book equity (2010)	1.86	-0.00
Bottom 10		
Change in financial liabilities (2005)	2.27	0.32
Net equity financing (2006)	2.04	0.34
Composite equity issuance (2006)	2.04	0.37
Put volatility minus call volatility (2011)	1.66	0.41
Efficient frontier index (2009)	1.77	0.45
Down forecast EPS (2002)	2.28	0.47
Net Operating Assets (2004)	2.23	0.48
Up Forecast (2002)	2.28	0.55
Change in recommendation (2004)	2.23	0.62
Frazzini-Pedersen Beta (2014)	1.73	0.64

Panel C: $Total\ R^2$, test assets: CZ21 portfolios

New factor	Old fac.	ΔR^2
Top 10		
Total accruals (2005)	95.25	0.43
Operating Cash flows to price (2004)	95.39	0.37
Net external financing (2006)	96.07	0.37
Change in equity to assets (2005)	95.25	0.36
Change in net financial assets (2005)	95.25	0.34
Net equity financing (2006)	96.07	0.33
Equity Duration (2004)	95.39	0.33
Firm Age - Momentum (2004)	95.39	0.32
52 week high (2004)	95.39	0.30
Growth in book equity (2010)	96.56	0.29
Bottom 10		
Real dirty surplus (2011)	96.74	0.08
Dividend seasonality (2013)	97.01	0.08
Return skewness (2015)	96.01	0.08
Leverage component of BM (2007)	95.69	0.08
Idiosyncratic skewness (3F model) (2015)	96.01	0.08
R&D ability (2013)	97.01	0.08
Enterprise component of BM (2007)	95.69	0.07
Change in Asset Turnover (2008)	95.09	0.07
Organizational capital (2013)	97.01	0.06
gross profits / total assets (2013)	97.01	0.06

Panel D: RMS_α, test assets: CZ21 portfolios

New factor	Old fac.	ΔRMS_α
Top 10		
Taxable income to income (2004)	0.46	-0.05
Analyst earnings per share (2006)	0.39	-0.05
Net debt financing (2006)	0.39	-0.05
Frazzini-Pedersen Beta (2014)	0.32	-0.05
Change in net financial assets (2005)	0.47	-0.04
Equity Duration (2004)	0.46	-0.04
Operating Cash flows to price (2004)	0.46	-0.04
52 week high (2004)	0.46	-0.03
Revenue Surprise (2006)	0.39	-0.03
Tail risk beta (2014)	0.32	-0.03
Bottom 10		
Net Operating Assets (2004)	0.46	0.02
Price delay coeff (2005)	0.47	0.02
Change in capex (three years) (2006)	0.39	0.02
Percent Operating Accruals (2011)	0.28	0.02
Up Forecast (2002)	0.46	0.02
Pastor-Stambaugh liquidity beta (2003)	0.46	0.03
change in net operating assets (2004)	0.46	0.03
Book-to-market and accruals (2004)	0.46	0.04
Down forecast EPS (2002)	0.46	0.04
Change in recommendation (2004)	0.46	0.07

For each 5-years rolling window ending in year y we compute the percentage $TotalR^2$ and RMS_α generated by a latent factor model with 3 common factors between individual stocks and CZ21 portfolios together with three CZ21-specific factors (model (i)). When a new factor form the zoo enters in the dataset we add this factor only to the six factors of model (i) and recompute the $TotalR^2$ with this new set of seven factors (model (ii)). $TotalR^2$ and RMS_α are computed using as test assets either the individual stocks or the CZ21 portfolios available in year y for both model (i) and model (ii). For each set of test assets we report the top 10 and bottom 10 increases in $TotalR^2$ ($\Delta\ R^2$), and the top 10 and bottom 10 decreases in RMS_α ($\Delta\ RMS_\alpha$) when a new factor is added in model (ii) to the 6 factors in model (i). We also report the $TotalR^2$ and RMS_α obtained with the old factors only (Old fac.), that is with the factors in model (i). The models are estimated on the rolling window starting in year $y-4$ and ending in year y, for each $y = 2001, ..., 2020$. $Total\ R^2$'s and RMS_α are computed as described in Section 4.2, but taking into account only the 5-year window ending in year y.

Table OA.6: Changes in $Total\ R^2$ and RMS_α due to new factors. Old factors: 3 common factors only.

Panel A: $Total\ R^2$, test assets: CRSP stocks

New factor	Old fac.	ΔR^2
Top 10		
Taxable income to income (2004)	28.22	3.17
52 week high (2004)	28.22	3.10
Net Operating Assets (2004)	28.22	2.75
Long-vs-short EPS forecasts (2011)	31.35	2.74
Unexpected R&D increase (2004)	28.22	2.70
Deferred Revenue (2012)	30.89	2.62
Percent Total Accruals (2011)	31.35	2.60
Growth in book equity (2010)	29.86	2.58
Growth in advertising expenses (2014)	30.21	2.56
Cash Productivity (2009)	23.48	2.48
Bottom 10		
Idiosyncratic risk (AHT) (2003)	28.23	1.24
Change in recommendation (2004)	28.22	1.24
Up Forecast (2002)	28.64	1.23
Real estate holdings (2010)	29.86	1.22
Return seasonality last year (2008)	21.03	1.21
Put volatility minus call volatility (2011)	31.35	1.21
Inst own among high short interest (2005)	32.62	1.21
Pastor-Stambaugh liquidity beta (2003)	28.23	1.15
Dividend seasonality (2013)	30.98	1.02
Organizational capital (2013)	30.98	0.99

Panel B: RMS_α, test assets: CRSP stocks

New factor	Old fac.	ΔRMS_α
Top 10		
Cash Productivity (2009)	1.80	-0.08
Inst Own and Market to Book (2005)	2.14	-0.05
Earnings consistency (2009)	1.80	-0.05
Change in Taxes (2011)	1.43	-0.03
R&D capital-to-assets (2011)	1.43	-0.02
Momentum based on FF3 residuals (2011)	1.43	-0.02
Growth in book equity (2010)	1.45	-0.02
Long-vs-short EPS forecasts (2011)	1.43	-0.01
Return seasonality years 16 to 20 (2008)	1.80	-0.01
Change in Forecast and Accrual (2004)	1.95	-0.01
Bottom 10		
Analyst earnings per share (2006)	1.85	0.46
Net external financing (2006)	1.85	0.46
Put volatility minus call volatility (2011)	1.43	0.49
Net equity financing (2006)	1.85	0.51
change in net operating assets (2004)	1.95	0.58
Inventory Growth (2002)	1.84	0.59
Up Forecast (2002)	1.84	0.74
Frazzini-Pedersen Beta (2014)	1.70	0.77
Change in recommendation (2004)	1.95	0.79
Net Operating Assets (2004)	1.95	0.83

Panel C: $Total\ R^2$, test assets: CZ21 portfolios

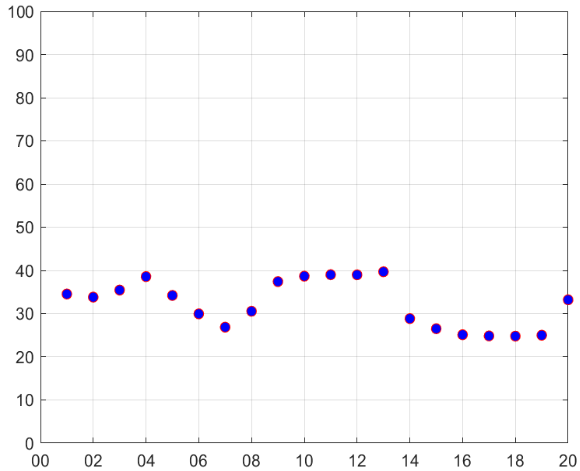
New factor	Old fac.	ΔR^2
Top 10		
52 week high (2004)	89.42	3.13
Price delay R-square (2005)	89.73	2.75
Price delay SE adjusted (2005)	89.73	2.69
Firm Age - Momentum (2004)	89.42	2.58
Inst Own and Idio Vol (2005)	89.73	2.43
Net Operating Assets (2004)	89.42	2.32
Inst Own and Turnover (2005)	89.73	2.18
Amihud's illiquidity (2002)	91.70	2.18
Inst Own and Market to Book (2005)	89.73	2.12
Intangible return using Sale2P (2006)	91.87	2.08
Bottom 10		
Return seasonality years 16 to 20 (2008)	92.02	0.16
Idiosyncratic skewness (3F model) (2015)	94.41	0.15
Organizational capital (2013)	94.27	0.15
R&D ability (2013)	94.27	0.12
Percent Operating Accruals (2011)	93.85	0.12
Consensus Recommendation (2002)	91.70	0.11
Change in Net Working Capital (2008)	92.02	0.11
Return seasonality years 6 to 10 (2008)	92.02	0.10
Sin Stock (selection criteria) (2009)	94.26	0.10
Real estate holdings (2010)	93.07	0.09

Panel D: RMS_α, test assets: CZ21 portfolios

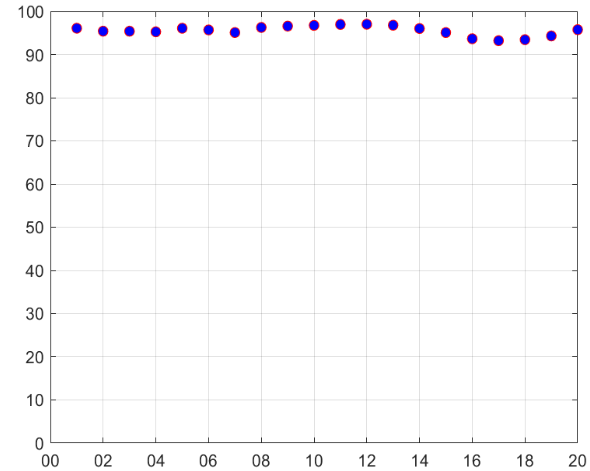
New factor	Old fac.	ΔRMS_α
Top 10		
Price delay SE adjusted (2005)	0.90	-0.38
Price delay R-square (2005)	0.90	-0.35
Inst Own and Idio Vol (2005)	0.90	-0.34
Inst Own and Turnover (2005)	0.90	-0.31
Change in current operating liabilities (2005)	0.90	-0.26
Inst Own and Market to Book (2005)	0.90	-0.26
change in net operating assets (2004)	0.76	-0.25
Equity Duration (2004)	0.76	-0.25
Change in equity to assets (2005)	0.90	-0.20
Firm Age - Momentum (2004)	0.76	-0.20
Bottom 10		
Mohanram G-score (2005)	0.90	0.05
Up Forecast (2002)	0.65	0.06
Operating Cash flows to price (2004)	0.76	0.06
Industry concentration (assets) (2006)	0.53	0.06
Efficient frontier index (2009)	0.29	0.06
Taxable income to income (2004)	0.76	0.07
Industry concentration (equity) (2006)	0.53	0.07
Return on assets (qtrly) (2010)	0.39	0.07
Book-to-market and accruals (2004)	0.76	0.07
Composite equity issuance (2006)	0.53	0.12

Details see Table OA.5

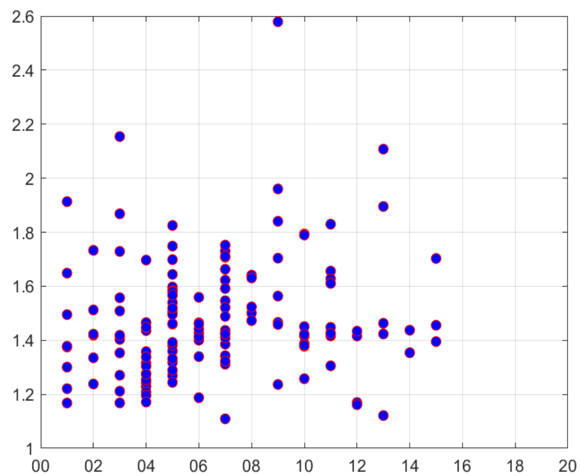
Figure OA.9: $Total\ R^2$ generated by old factors and its change due new factors. Old factors: 3 common factors and first 3 CZ21-specific factors.



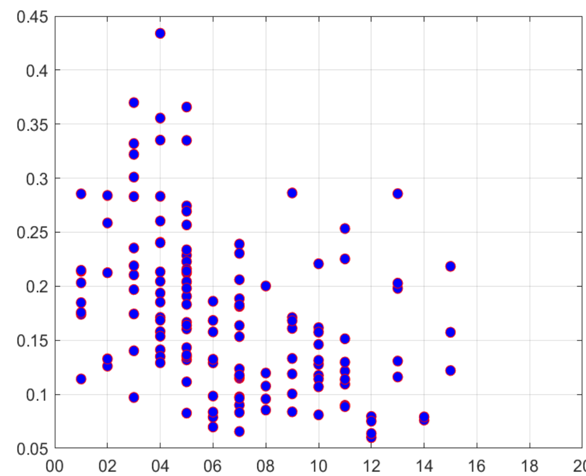
(a) Total R^2. Test assets: CRSP stocks



(b) Total R^2. Test assets: CZ21 portfolios



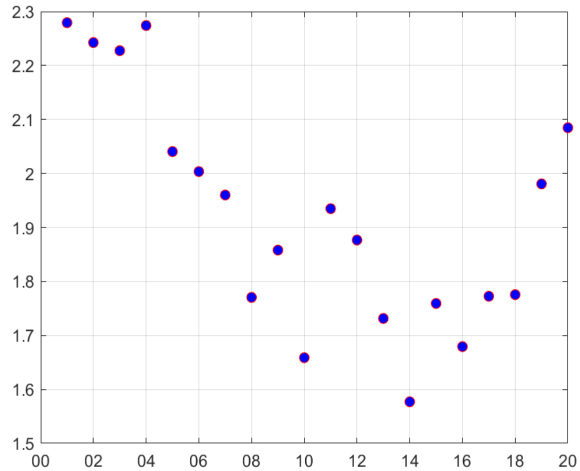
(c) Increase in Total R^2. Test assets: CRSP stocks



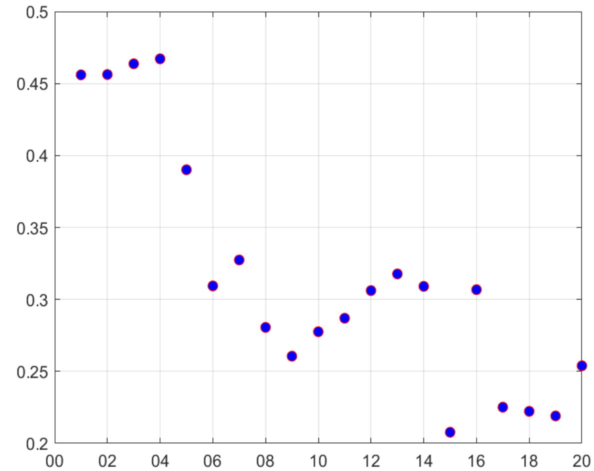
(d) Increase in Total R^2. Test assets: CZ21 portfolios

For each 5-years rolling window ending in year y we compute the percentage $TotalR^2$ generated by a latent factor model with 3 common factors between individual stocks and CZ21 portfolios (model (i)). When a new factor form the zoo enters in the dataset we add this factor only to the six factors of model (a) and recompute the $TotalR^2$ with this new set of seven factors (model (ii)). $TotalR^2$ is computed using as test assets either the individual stocks or the CZ21 portfolios available in year y for both model (i) and model (ii). In Panels (a) and (b) we report the $TotalR^2$ obtained with the old factors only (Old fac.), that is with the factors in model (i). We also report all the increases in $TotalR^2$ when a new factor is added in model (ii) to the 6 factors in model (i). The models are estimated on the rolling window starting in year $y-4$ and ending in year y, for each $y = 2001, ..., 2020$. $Total\ R^2$'s are computed as described in Section 4.2, but taking into account only the 5-year window ending in year y.

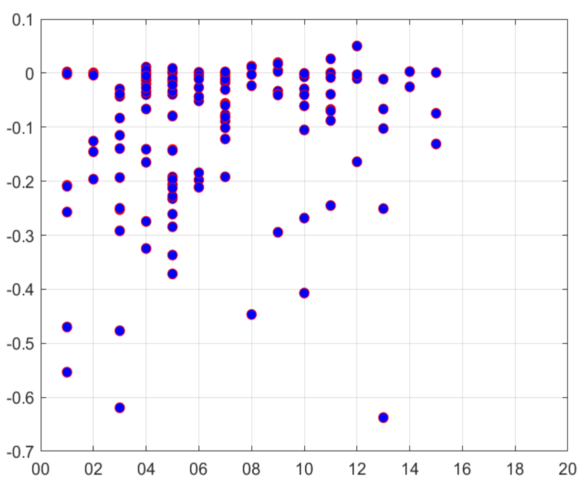
Figure OA.10: RMS_α generated by old factors and its change due new factors. Old factors: 3 common factors and first 3 CZ21-specific factors.



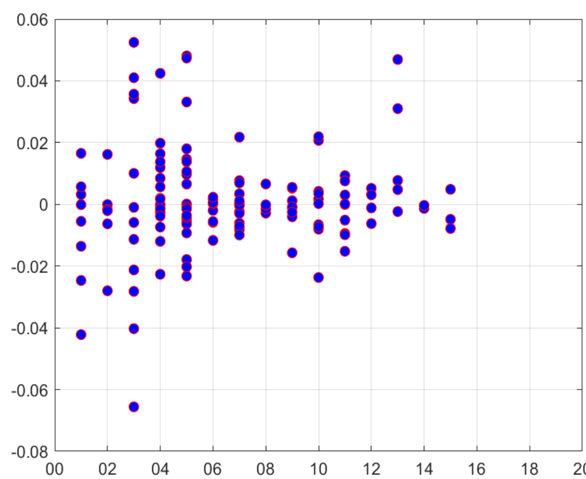
(a) RMS_α. Test assets: CRSP stocks



(b) RMS_α. Test assets: CZ21 portfolios



(c) Increase in RMS_α. Test assets: CRSP stocks



(d) Increase in RMS_α. Test assets: CZ21 portfolios

For each 5-years rolling window ending in year y we compute the RMS_α generated by a latent factor model with 3 common factors between individual stocks and CZ21 portfolios (model (i)). When a new factor form the zoo enters in the dataset we add this factor only to the six factors of model (a) and recompute the RMS_α with this new set of seven factors (model (ii)). RMS_α is computed using as test assets either the individual stocks or the CZ21 portfolios available in year y for both model (i) and model (ii). In Panels (a) and (b) we report the RMS_α obtained with the old factors only (Old fac.), that is with the factors in model (i). We also report all the increases in RMS_α when a new factor is added in model (ii) to the 6 factors in model (i). The models are estimated on the rolling window starting in year $y-4$ and ending in year y, for each $y = 2001, ..., 2020$. RMS_α's are computed as described in Section 4.2, but taking into account only the 5-year window ending in year y.

B Assumptions and proofs

Section B.1 includes all the Assumptions required to prove Proportion A.1 and Theorems A.1 and A.2 in Appendix A. Section B.2 provides the proof of Proposition A.1, while Sections B.3 and B.4 provide the proofs of Theorems A.1 and A.2, respectively. Section B.5 contains additional technical results required in the previous sections. Finally, Sections B.5 and B.7 provides the uniform asymptotic expansions and distributions, respectively, of factors and loadings in the group factor model when factors are estimated by RP-PCA: these results are useful in themselves, but also instrumental to some of the proofs of the previous results.

In this appendix, we denote by $a_t = [A]_t$ the column vector corresponding to the t-th row a_t' of matrix $A = [a_1, ..., a_t, ..., a_T]'$.

B.1 Assumptions for Proportion A.1 and all Theorems (NEED TO BE CHANGED/ADAPTED TO THE NEW PROOFS!!!!)

We make the following assumptions:

Assumption B.1. *We have $N_1, N_2, T \to \infty$ such that the conditions in (A.17) hold, that is:* $\sqrt{T}/N = o(1)$, $N/T^2 = o(1)$, *and* $\mu_N = \sqrt{N_2/N_1} \to \mu$, *with* $\mu \in [0, 1]$.

Assumption B.2. *The unobservable factor process $F_t = [\ f_t^{c\prime},\ f_{1,t}^{s\prime},\ f_{2,t}^{s\prime}\ ]'$ has vector of means, and covariance matrix as defined in (A.8), that is:*

$$E[F_t] = \begin{bmatrix} \mu^c \\ \mu_1^s \\ \mu_2^s \end{bmatrix}, \qquad \text{and} \qquad \Sigma_F := V(F_t) = \begin{bmatrix} I_{k^c} & 0 & 0 \\ 0 & I_{k_1^s} & \Phi \\ 0 & \Phi' & I_{k_2^s} \end{bmatrix},$$

with all the elements of vector $E[F_t]$ being finite, and where Σ_F is positive-definite.

Assumption B.3. *The loadings matrix $\Lambda_j = [\ \Lambda_j^c \ \vdots \ \Lambda_j^s\ ] = [\ \lambda_{j,1},\ \ldots,\ \lambda_{j,N_j}\ ]'$ is such that $\lim_{N_j\to\infty} \frac{1}{N_j}\Lambda_j'\Lambda_j = \Sigma_{\lambda,j}$, where $\Sigma_{\lambda,j}$ is a positive-definite (k_j, k_j) matrix with distinct eigenvalues and $k_j = k^c + k_j^s$, for $j = 1, 2$.*

Assumption B.4. *The error terms $\varepsilon_{j,i,t}$ and the factors $h_{j,t} = [f_t^{c\prime}, f_{j,t}^{s\prime}]'$ are such that for $j = 1, 2$ and all $i, t \geq 1$: a) $E[\varepsilon_{j,i,t}|\mathcal{F}_t] = 0$ and $E[\varepsilon_{j,i,t}^2|\mathcal{F}_t] \leq M$, a.s., where $\mathcal{F}_t = \sigma(F_s, s \leq t)$, b) $E[\varepsilon_{j,i,t}^8] \leq M$ and $E[\|h_{j,t}\|^{2r\vee 8}] \leq M$, for a constant $M < \infty$, where $r > 2$ is defined in Assumption B.5 b).*

Assumption B.5. *Define the variables $\xi_{j,t} = \frac{1}{\sqrt{N_j}} \sum_{i=1}^{N_j} \lambda_{j,i}\varepsilon_{j,i,t}$ and $\kappa_{j,t} = \frac{1}{\sqrt{N_j}} \sum_{i=1}^{N_j}(\varepsilon_{j,i,t}^2 - \eta_{j,t}^2)$, indexed by N_1, N_2, where $\eta_{j,t}^2 = \operatorname*{plim}_{N_j\to\infty} \frac{1}{N_j} \sum_{i=1}^{N_j} E[\varepsilon_{j,i,t}^2|\mathcal{F}_t]$, for $j = 1, 2$. a) For any $t \geq 1$ and $h \geq 0$*

have:

$$[\xi'_{1,t}, \xi'_{2,t}, \xi'_{1,t-h}, \xi'_{2,t-h}]' \xrightarrow{d} N(0, \Omega_t(h)), \quad (\mathcal{F}_t\text{-stably}),$$

as $N_1, N_2 \to \infty$, where the asymptotic variance matrix is:

$$\Omega_t(h) = \begin{bmatrix} \Omega_{11,t}(0) & \Omega_{12,t}(0) & \Omega_{11,t}(h) & \Omega_{12,t}(h) \\ & \Omega_{22,t}(0) & \Omega_{21,t}(h) & \Omega_{22,t}(h) \\ & & \Omega_{11,t-h}(0) & \Omega_{12,t-h}(0) \\ & & & \Omega_{22,t-h}(0) \end{bmatrix},$$

for $\Omega_{jk,t}(h) = \underset{N_1,N_2\to\infty}{\text{plim}} \frac{1}{\sqrt{N_j N_k}} \sum_{i=1}^{N_j} \sum_{\ell=1}^{N_k} \lambda_{j,i}\lambda'_{k,\ell} cov(\varepsilon_{j,i,t}, \varepsilon_{k,\ell,t-h}|\mathcal{F}_t)$, for any j, k, h. Moreover, for all $N_1, N_2 \geq 1$ and $j = 1, 2$, we have: b) $E(\|\xi_{j,t}\|^{2r}|\mathcal{F}_t) \leq M$, a.s., and c) $E\left[|\kappa_{j,t}|^4\right] \leq M$, for constants $M < \infty$ and $r > 2$. Finally, let $\breve{\aleph}_{j,i} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \breve{\varepsilon}_{j,i,t}\breve{h}_t$, then d) $E[\|\breve{\aleph}_{j,i}\|^2] \leq M$.

Assumption B.6. *a) The triangular array processes $V_t \equiv V_{N_1,N_2,t} = [h'_{j,t}, \xi'_{j,t}, j = 1, 2]'$ and $V_t^* \equiv V_{N_1,N_2,t}^* = [\kappa_{j,t}, \eta_{j,t}^2, j = 1, 2]'$ are strong mixing of size $-\frac{r}{r-2}$, uniformly in $N_1, N_2 \geq 1$.* [13] *Moreover*
b) $\|E(\xi_{j,t}\xi'_{k,t}|\mathcal{F}_t) - E(\xi_{j,t}\xi'_{k,t}|F_t, ..., F_{t-m})\|_2 = O(m^{-\psi})$, as $m \to \infty$, uniformly in $N_1, N_2 \geq 1$, and
c) $\|\eta_{j,t}^2 - E(\eta_{j,t}^2|\mathcal{V}_{t-m}^{t+m})\|_8 = O(m^{-\psi})$, as $m \to \infty$, uniformly in $N_1, N_2 \geq 1$,
for $j, k = 1, 2$, where $\mathcal{V}_{t-m}^{t+m} = \sigma(V_s, t - m \leq s \leq t + m)$ and $\psi > 1$.

Assumption B.7. *For $j = 1, 2$:*
a) $\frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{t-1} E[\eta_{j,ts}^4] \leq M$, $E\left[\left(\frac{1}{\sqrt{N_j}} \sum_{i=1}^{N_j} (\varepsilon_{j,i,t}\varepsilon_{j,i,s} - \eta_{j,ts}^2)\right)^2\right] \leq M$, for any $s < t$ and a constant M, where $\eta_{j,ts}^2 = \underset{N_j\to\infty}{\text{plim}} \frac{1}{N_j} \sum_{i=1}^{N_j} E[\varepsilon_{j,i,t}\varepsilon_{j,i,s}|\mathcal{F}_t]$; b) $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} (1 + \eta_{j,t}^2) h_{j,t}\alpha'_{j,t} = O_p(1)$, $\frac{1}{T} \sum_{t=1}^{T} \xi_{j,t}\alpha'_{j,t} = o_p(1)$, $E[\|\alpha_{j,t}\|^2] = O(1)$, where $\alpha_{j,t} = \frac{1}{\sqrt{N_j T}} \sum_{i=1}^{N_j} \sum_{s=1,s\neq t}^{T} \varepsilon_{j,i,t}\varepsilon_{j,i,s}h_{j,s}$;
c) $E[\|\beta_{j,t}\|^2] = O(1)$ and $E[\|\bar{\beta}_{j,t}\|^2] = O(1)$, where $\beta_{j,t} = \frac{1}{\sqrt{N_j T}} \sum_{i=1}^{N_j} \sum_{s=1,s\neq t}^{T} \varepsilon_{j,i,t}(\varepsilon_{j,i,s}\zeta_{j,s} - E[\varepsilon_{j,i,s}\zeta_{j,s}])$ and $\bar{\beta}_{j,t} = \frac{1}{T} \sum_{i=1}^{N_j} \sum_{s=1,s\neq t}^{T} \varepsilon_{j,i,t}E[\varepsilon_{j,i,s}\zeta_{j,s}]$, where $\zeta_{j,t} = (\eta_{j,t}^2 h'_{j,t}, \kappa_{j,t}h'_{j,t}, \xi'_{j,t}, \alpha'_{j,t})'$.

Assumption B.8. *For $j = 1, 2$:*
a) $P[\|h_{j,t}\| \geq \delta] \leq c_1 \exp(-c_2\delta^b)$, for large δ; b) $\sum_{\ell=1:\ell\neq i}^{N_j} E[\varepsilon_{j,\ell,t}\varepsilon_{j,i,t}] \leq M$, for all $i \geq 1$; c) $P[\|\frac{1}{T} \sum_{t=1}^{T} z_{j,t}\| \geq \delta] \leq c_1 T \exp(-c_2\delta^2 T^\eta) + c_3 T\delta^{-1} \exp(-c_4 T^{\bar{\eta}})$, for all $i \geq 1$ and $\delta > 0$, where either $z_{i,t} = h_{j,t}\varepsilon_{j,i,t}$, or $z_{i,t} = \varepsilon_{j,i,t}^2 - E[\varepsilon_{j,i,t}^2]$, or $z_{i,t} = \frac{1}{\sqrt{N_j}} \sum_{\ell=1:\ell\neq i}^{N_j} \varepsilon_{j,\ell,t}\varepsilon_{j,i,t} -$

[13] That is, $\alpha(h) = O(h^{-\phi})$ for some $\phi > \frac{r}{r-2}$, where $\alpha(h) = \sup_{N_1,N_2\geq 1} \sup_{t\geq 1} \sup_{A\in\mathcal{V}_{-\infty}^t, B\in\mathcal{V}_{t+h}^{\infty}} |P(A \cap B) - P(A)P(B)|$, where $\mathcal{V}_{t-m}^{t+m} = \sigma(V_s, t - m \leq s \leq t + m)$, and similarly for V_t^*.

$E[\frac{1}{\sqrt{N_j}}\sum_{\ell=1:\ell\neq i}^{N_j}\varepsilon_{j,\ell,t}\varepsilon_{j,i,t}]$; *d)* $\|\lambda_{j,i}\| \leq M$, *for all* $i \geq 1$; *where* $b, c_1, c_2, c_3, c_4, \eta, \bar{\eta}, M > 0$ *are constants, and* $\eta \geq 1/2$.

Assumption B.9. *The error terms are such that:* *a)* $Cov(\varepsilon_{j,i,t}, \varepsilon_{k,\ell,t-h}|\mathcal{F}_t) = 0$, *if either* $j \neq k$, *or* $i \neq \ell$, *b)* $E[\varepsilon_{j,i,t}|\{\varepsilon_{j,i,t-h}\}_{h\geq 1}, \mathcal{F}_t] = 0$, *c)* $E[\varepsilon_{j,i,t}^2|\{\varepsilon_{j,i,t-h}\}_{h\geq 1}, \mathcal{F}_t] = \gamma_{j,ii}$, *say, where* $\gamma_{j,ii} > 0$, *for all* j, i, t, h.

B.2 Proof of Proposition A.1

From the covariance matrix Σ_F of the factor vector $\left(f_t^c, f_{1,t}^s, f_{2,t}^s\right)'$ in equation (A.8), and the definition of matrices R and R^* given in Section A.5, it follows that:

$$R = \begin{pmatrix} I_{k^c} & 0 \\ 0 & \Phi\Phi' \end{pmatrix}, \qquad R^* = \begin{pmatrix} I_{k^c} & 0 \\ 0 & \Phi'\Phi \end{pmatrix}. \tag{B.1}$$

Noting that also in our set-up matrix Σ_F is assumed to be positive definite (see Assumption B.2), then the proof of Proposition A.1 is omitted as it is analogous to the proof of Proposition 1 in AGGR (see Section C.1 in their OA).

B.3 Proof of Theorem A.1

The proof of Theorem A.1 is structured analogously to the proof of Theorem 1 in AGGR. The main novelty in the current paper consists in the derivation of the asymptotic expansion for the estimates of the pervasive factors extracted by RP-PCA in each group (Subsection B.3.1), which are different in some higher order terms from that derived in AGGR, who instead derived the asymptotic expansion of the classical PCA estimators. Our asymptotic expansion provides an higher order terms compared to the one derived by Lettau and Pelger (2020a) for the RP-PCA estimators. Then, all the steps in the proof of Theorem 1 in AGGR are re-done taking into account of the new asymptotic expansion of the factors' estimators, and the fact that the formulas for the canonical correlations of the estimated factors are different from those used in AGGR, as they now need to include the factor mean.

The proof starts with the asymptotic expansion of the factor estimates $\hat{h}_{j,t}$ (see Proposition B.2 in Section B.3.1). This result allows to derive the asymptotic expansion for the sample canonical correlation matrix $\hat{R}$ (Section B.3.2), and the asymptotic expansions of the eigenvalues (and eigenvectors) of matrix $\hat{R}$ by perturbation methods (Sections B.3.3 and B.3.4). This yields the asymptotic expansions of the canonical correlations and of the test statistic $\hat{\xi}(k^c)$ (Section B.3.5), and the asymptotic Gaussian distribution of the test statistic (Section B.3.6).

B.3.1 Asymptotic expansion of the factor estimates $\hat{h}_{j,t}$

PROPOSITION B.2. *Under Assumptions B.1-B.4, B.5 b), c), B.6 a), and B.7, we have:*

$$\hat{h}_{j,t} = \hat{\mathcal{H}}_j(h_{j,t} + \psi_{j,t}), \qquad \psi_{j,t} := \frac{1}{\sqrt{N_j}} u_{j,t} + \frac{1}{T} b_{j,t} + \frac{1}{\sqrt{N_j T}} d_{j,t} + \vartheta_{j,t}, \tag{B.2}$$

for groups $j = 1, 2$, and dates $t = 1, \ldots, T$, where:

$$\begin{aligned}
u_{j,t} &= \left(\frac{1}{N_j} \sum_{i=1}^{N_j} \lambda_{j,i} \lambda'_{j,i} \right)^{-1} \frac{1}{\sqrt{N_j}} \sum_{i=1}^{N_j} \lambda_{j,i} \varepsilon_{j,i,t}, \\
b_{j,t} &= \left(\frac{1}{N_j} \sum_{i=1}^{N_j} \lambda_{j,i} \lambda'_{j,i} \right)^{-1} \left[\frac{1}{T} \sum_{r=1}^{T} \breve{h}^*_{j,r} \breve{h}^{*\prime}_{j,r} \right]^{-1} \left(\frac{1}{N_j} \sum_{i=1}^{N_j} \lambda_{j,i} \lambda'_{j,i} \right)^{-1} \\
&\quad \times \left[\eta^2_{j,t} \breve{h}_{j,t} + \frac{1}{N_j T} \sum_{i=1}^{N_j} \sum_{s=1}^{T} \sum_{q=1}^{T} \breve{\varepsilon}^*_{is} \breve{\varepsilon}^*_{iq} \breve{h}^*_s \breve{h}^{*\prime}_q \left(\frac{1}{N_j} \sum_{i=1}^{N_j} \lambda_{j,i} \lambda'_{j,i} \right) h_t \right], \\
d_{j,t} &= \left(\frac{1}{N_j} \sum_{i=1}^{N_j} \lambda_{j,i} \lambda'_{j,i} \right)^{-1} \left[\frac{1}{T} \sum_{r=1}^{T} \breve{h}^*_{j,r} \breve{h}^{*\prime}_{j,r} \right]^{-1} \left(\frac{1}{N_j} \sum_{i=1}^{N_j} \lambda_{j,i} \lambda'_{j,i} \right)^{-1} \\
&\quad \times \left\{ \frac{1}{\sqrt{N_j T}} \sum_{i=1}^{N_j} \sum_{s=1}^{T} \varepsilon_{j,i,s} \left[\breve{h}^*_{j,s} \lambda'_{j,i} + \lambda_{j,i} h'_{j,s} \left(\frac{1}{N_j} \sum_{i=1}^{N_j} \lambda_{j,i} \lambda'_{j,i} \right) \right] \right\} h_{j,t},
\end{aligned}$$

with $\breve{h}^*_{j,t} := h_{j,t} + \tilde{\gamma}_{RP} \bar{h}_j$, $\breve{\varepsilon}^*_{j,i,t} := \varepsilon_{j,i,t} + \tilde{\gamma}_{RP} \bar{\varepsilon}_{j,i}$ $\lambda_{j,i} = (\lambda^{c\prime}_{j,i}, \lambda^{s\prime}_{j,i})'$. The terms $\vartheta_{j,t}$ are such that

$$\frac{1}{T} \sum_{t=1}^{T} \left(\frac{1}{\sqrt{N_j}} u_{j,t} + \frac{1}{T} b_{j,t} + \frac{1}{\sqrt{N_j T}} d_{j,t} + \vartheta_{j,t} \right) \vartheta'_{k,t} = o_p \left(\frac{1}{N\sqrt{T}} \right)$$

and $\frac{1}{T} \sum_{t=1}^{T} h_{j,t} \vartheta'_{k,t} = O_p \left(\frac{1}{N} \right)$, *as $N_1, N_2, T \to \infty$. Finally, matrix $\hat{\mathcal{H}}_j$ converges in probability to a nonstochastic positive definite (k_j, k_j) matrix, for $j, k = 1, 2$.*

Proposition B.2 extends Proposition 3 in AGGR, which was derived for factors estimated by PCA, to the more general case in which factors are estimated by RP-PCA, as in Lettau and Pelger (2020a,b). For each group of data $j = 1, 2$, Proposition B.2 provides a more accurate asymptotic expansion of RP-PC factor estimator compared to the results in Lettau and Pelger (2020a): this refined result is needed to control higher-order terms in the asymptotic expansion of the test statistic in our Theorem A.1. Notably, the term $\frac{1}{\sqrt{N_j}} u_{j,t}$ in the expression for $\psi_{j,t}$ appears in Proposition 3 of AGGR, and is also the only term appearing in the expansion of the RP-PC estimator in the OA of Lettau and Pelger (2020a). Notably, the term of stochastic order $1/\sqrt{N}$ is $u_t/\sqrt{N}$, where $u_t = (\Lambda' \Lambda / N)^{-1} \xi_t$ is zero mean as $E[\xi_t] = 0$ from Assumption B.4 a). This term is the usual first order term appearing in the asymptotic expansion of classical PC factor estimator (see Bai (2003)) and it is also the first order term appearing in the expansion of RP-PC estimator (see Lettau and Pelger (2020a)). All the other terms in the expression for $\psi_{j,t}$ are new compared to Lettau and Pelger (2020a), and are also different from those appearing in AGGR.

B.3.2 Asymptotic expansion of matrix $\hat{R}$

As canonical correlations and canonical directions are invariant to one-to-one transformations of the vectors $\hat{h}_{1,t}$ and $\hat{h}_{2,t}$, in the asymptotic analysis of the test statistic $\hat{\xi}(k^c)$, we can set $\hat{\mathcal{H}}_j = I_{k_j}$, $j = 1, 2$, in expansion (B.2) without loss of generality.

Let $\bar{\hat{h}}_j = \frac{1}{T}\sum_{t=1}^{T} \hat{h}_{j,t}$, then equation (B.2) implies $\bar{\hat{h}}_j = \hat{\mathcal{H}}_j(\bar{h}_{j,t} + \bar{\psi}_{j,t})$, with:

$$\bar{\psi}_{j,t} := \frac{1}{\sqrt{N_j}}\bar{u}_{j,t} + \frac{1}{T}\bar{b}_{j,t} + \frac{1}{\sqrt{N_j T}}\bar{d}_{j,t} + \bar{\vartheta}_{j,t}. \tag{B.3}$$

By defining

$$\breve{u}_{j,t} := u_{j,t} - \bar{u}_j \ , \tag{B.4}$$

$$\breve{b}_{j,t} := b_{j,t} - \bar{b}_j \ , \tag{B.5}$$

$$\breve{d}_{j,t} := d_{j,t} - \bar{d}_j \ , \tag{B.6}$$

$$\breve{\vartheta}_{j,t} := \vartheta_{j,t} - \bar{\vartheta}_j \ , \tag{B.7}$$

and

$$\begin{aligned} \breve{\psi}_{j,t} &:= \psi_{j,t} - \bar{\psi}_j \\ &= \frac{1}{\sqrt{N_j}}\breve{u}_{j,t} + \frac{1}{T}\breve{b}_{j,t} + \frac{1}{\sqrt{N_j T}}\breve{d}_{j,t} + \breve{\vartheta}_{j,t} \ , \end{aligned} \tag{B.8}$$

we get

$$\begin{aligned} \hat{h}_{j,t} - \bar{\hat{h}}_j &= (h_{j,t} + \psi_{j,t}) - (\bar{\psi}_j + \bar{h}_j) = (h_{j,t} - \bar{h}_j) + (\psi_{j,t} - \bar{\psi}_j) \\ &= \hat{\mathcal{H}}_j(\breve{h}_{j,t} + \breve{\psi}_{j,t}) \ , \end{aligned} \tag{B.9}$$

for all dates $t = 1, ..., T$. Therefore, we get:

$$\hat{V}_{j,k} = \frac{1}{T}\sum_{t=1}^{T} \hat{h}_{j,t}\hat{h}'_{k,t} - \bar{\hat{h}}_j \bar{\hat{h}}'_j = \frac{1}{T}\sum_{t=1}^{T} (\breve{h}_{j,t} + \breve{\psi}_{j,t})(\breve{h}_{j,t} + \breve{\psi}_{j,t})' = \tilde{V}_{j,k} + \hat{X}_{j,k}, \tag{B.10}$$

where:

$$\tilde{V}_{j,k} = \frac{1}{T}\sum_{t=1}^{T} \breve{h}_{j,t}\breve{h}'_{k,t}\,, \quad \hat{X}_{j,k} = \frac{1}{T}\sum_{t=1}^{T} (\breve{h}_{j,t}\breve{\psi}'_{k,t} + \breve{\psi}_{j,t}\breve{h}'_{k,t}) + \frac{1}{T}\sum_{t=1}^{T} \breve{\psi}_{j,t}\breve{\psi}'_{k,t}, \tag{B.11}$$

for $j, k = 1, 2$. From the definition of matrix $\hat{R}$ in (A.10), and by using (B.10) and $\hat{V}_{jj}^{-1} = \left(I_{k_j} + \tilde{V}_{jj}^{-1}\hat{X}_{jj}\right)^{-1}\tilde{V}_{jj}^{-1}$, we get:

$$\hat{R} = \left(I_{k_1} + \tilde{V}_{11}^{-1}\hat{X}_{11}\right)^{-1}\tilde{V}_{11}^{-1}\left(\tilde{V}_{12} + \hat{X}_{12}\right)\left(I_{k_2} + \tilde{V}_{22}^{-1}\hat{X}_{22}\right)^{-1}\tilde{V}_{22}^{-1}\left(\tilde{V}_{21} + \hat{X}_{21}\right). \tag{B.12}$$

By using the definitions of $\psi_{j,t}$ in Proposition B.2 and of $\bar{\psi}_{j,t}$ in equation (B.3), the next Lemma provides an upper bound for terms $\hat{X}_{j,k}$, $j, k = 1, 2$.

LEMMA B.1. *Under Assumptions B.1-B.4, B.5 b)-c), B.6 a) and B.7 we have* $\hat{X}_{j,k} = O_p(\delta_{N,T})$*, for* $j, k = 1, 2$*, where* $\delta_{N,T} := (\min\{N, T\})^{-1}$.

We now expand matrix $\hat{R}$ at second order in the $\hat{X}_{j,k}$. We do this because the first-order contribution of the $\hat{X}_{j,k}$ to the statistic of interest involves leading terms of stochastic order $O_p\left(\frac{1}{N\sqrt{T}}\right)$: see . The second-order remainder term is $O_p(\delta_{N,T}^2)$, and $\delta_{N,T}^2$ is not negligible with respect to $\frac{1}{N\sqrt{T}}$ when T is too small compared to N (that is the case when $m = T$ and $\frac{1}{N\sqrt{T}} \geq \frac{1}{T^2}$ when $T\sqrt{T} \ll N \ll T^2$). In order to get validity of our results for more general conditions on the relative growth rate of N and T such as in Assumption B.1, we consider a second-order expansion. By using $(I - X)^{-1} = I + X + X^2 + O_p(\delta_{N,T}^3)$ for $X = O_p(\delta_{N,T})$, from equation (B.12) we get the next Lemma.

LEMMA B.2. *Under Assumptions B.1-B.4, B.5 b)-c), B.6 a) and B.7, the second-order asymptotic expansion of matrix* $\hat{R}$ *is:*

$$\hat{R} = \tilde{R} + \hat{\Psi} + O_p(\delta_{N,T}^3), \tag{B.13}$$

where $\tilde{R} = \tilde{V}_{11}^{-1}\tilde{V}_{12}\tilde{V}_{22}^{-1}\tilde{V}_{21}$ *and* $\hat{\Psi} = \tilde{V}_{11}^{-1}\hat{\Psi}^*$*, with* $\hat{\Psi}^* = \hat{\Psi}^{*\,(I)} + \hat{\Psi}^{*\,(II)}$*,*

$$\hat{\Psi}^{*\,(I)} = -\hat{X}_{11}\tilde{R} + \hat{X}_{12}\tilde{B} - \tilde{B}'\hat{X}_{22}\tilde{B} + \tilde{B}'\hat{X}_{21}, \tag{B.14}$$

$$\hat{\Psi}^{*\,(II)} = -\hat{X}_{11}\tilde{V}_{11}^{-1}\hat{\Psi}^{*\,(I)} + \left(\hat{X}_{22}\tilde{B} - \hat{X}_{21}\right)'\tilde{V}_{22}^{-1}\left(\hat{X}_{22}\tilde{B} - \hat{X}_{21}\right), \tag{B.15}$$

and $\tilde{B} = \tilde{V}_{22}^{-1}\tilde{V}_{21}$.

This Lemma is analogous to Lemma B.2 in AGGR. In equation (B.13) matrix $\hat{R}$ is decomposed into the sum of the sample canonical correlation matrix $\tilde{R}$ computed with the true factor values, an estimation error term $\hat{\Psi}$ consisting of first-order and second-order components $\hat{\Psi}^{*\,(I)}$ and $\hat{\Psi}^{*\,(II)}$, respectively, and a third-order remainder term $O_p(\delta^3_{N,T})$.

B.3.3 Matrix $\tilde{R}$ and its eigenvalues and eigenvectors

We now characterize matrix $\tilde{R}$ and its eigenvalues, that are $\tilde{\rho}^2_1, ..., \tilde{\rho}^2_{k_1}$, i.e. the squared sample canonical correlations of $h_{1,t}$ and $h_{2,t}$, under the null hypothesis of $k^c > 0$ common factors among the two groups of observables. Since the vectors $h_{1,t}$ and $h_{2,t}$ have a common component of dimension k^c, we know that $\tilde{\rho}_1 = ... = \tilde{\rho}_{k^c} = 1$ a.s.. Using the notation:

$$
\begin{aligned}
\tilde{\Sigma}_{cc} &= \frac{1}{T}\sum_{t=1}^{T}(f^c_t - \bar{f}^c)(f^c_t - \bar{f}^c)' = \frac{1}{T}\sum_{t=1}^{T}\breve{f}^c_t\breve{f}^{c\prime}_t, \\
\tilde{\Sigma}_{c,j} &= \frac{1}{T}\sum_{t=1}^{T}(f^c_t - \bar{f}^c)(f^s_{j,t} - \bar{f}^s_j)' = \frac{1}{T}\sum_{t=1}^{T}\breve{f}^c_t\breve{f}^{s\prime}_{j,t}, \qquad \tilde{\Sigma}_{j,c} = \tilde{\Sigma}'_{c,j}\ , \\
\tilde{\Sigma}_{j,k} &= \frac{1}{T}\sum_{t=1}^{T}(f^s_{j,t} - \bar{f}^s_j)(f^s_{k,t} - \bar{f}^s_k) = \frac{1}{T}\sum_{t=1}^{T}\breve{f}^s_{j,t}\breve{f}^{s\prime}_{k,t}, \qquad j,k = 1,2,
\end{aligned}
$$

with $\breve{f}^c_t := f^c_t - \bar{f}^c$ and $\breve{f}^s_{j,t} := f^s_{j,t} - \bar{f}^s_j$, we can write matrices $\tilde{V}_{j,k}$, with $j,k = 1,2$, in (B.11) in block form as:

$$
\tilde{V}_{jj} = \begin{pmatrix} \tilde{\Sigma}_{cc} & \tilde{\Sigma}_{c,j} \\ \tilde{\Sigma}_{j,c} & \tilde{\Sigma}_{jj} \end{pmatrix}, \quad j = 1,2, \qquad \tilde{V}_{12} = \begin{pmatrix} \tilde{\Sigma}_{cc} & \tilde{\Sigma}_{c,2} \\ \tilde{\Sigma}_{1,c} & \tilde{\Sigma}_{12} \end{pmatrix} = \tilde{V}'_{21}.
$$

The last two equations and the definition of $\tilde{R}$ allow to obtain the next Lemma, which is analogous as the Lemma B.3 in AGGR, with the only fundamental difference being the definition of matrices $\tilde{\Sigma}_{cc}$, $\tilde{\Sigma}_{c,j}$, and $\tilde{\Sigma}_{j,k}$.

LEMMA B.3. *The matrix $\tilde{B}$ defined in Lemma B.2 is such that:*

$$
B = \tilde{V}^{-1}_{22}\tilde{V}_{21} = \begin{bmatrix} I_{k^c} & \tilde{\Sigma}^{-1}_{cc|2}\tilde{\Sigma}_{c1|2} \\ 0 & \tilde{\Sigma}^{-1}_{22|c}\tilde{\Sigma}_{21|c} \end{bmatrix} = \begin{bmatrix} I_{k^c} & \tilde{\Sigma}^{-1}_{cc}\left(\tilde{\Sigma}_{c1} - \tilde{\Sigma}_{c2}\tilde{\Sigma}^{-1}_{22|c}\tilde{\Sigma}_{21|c}\right) \\ 0 & \tilde{\Sigma}^{-1}_{22|c}\tilde{\Sigma}_{21|c} \end{bmatrix}. \quad \text{(B.16)}
$$

The matrix $\tilde{R} = \tilde{V}^{-1}_{11}\tilde{V}_{12}\tilde{V}^{-1}_{22}\tilde{V}_{21}$ is such that:

$$
\tilde{R} = \begin{bmatrix} I_{k^c} & \tilde{\Sigma}^{-1}_{cc}\tilde{\Sigma}_{c1}(I_{k_1-k^c} - \tilde{R}_{ss}) \\ 0 & \tilde{R}_{ss} \end{bmatrix},
$$

where $\tilde{R}_{ss} = \tilde{\Sigma}_{11|c}^{-1}\tilde{\Sigma}_{12|c}\tilde{\Sigma}_{22|c}^{-1}\tilde{\Sigma}_{21|c}$ *and* $\tilde{\Sigma}_{jk|c} := \tilde{\Sigma}_{jk} - \tilde{\Sigma}_{jc}\tilde{\Sigma}_{cc}^{-1}\tilde{\Sigma}_{ck}$ *for* $j, k = 1, 2$.

Matrix $\tilde{R}_{ss}$ is the sample canonical correlation matrix for the residuals of the sample orthogonal projections of $\breve{f}_{1,t}^s$ and $\breve{f}_{2,t}^s$ onto $\breve{f}_t^c$, with the latter three factors being the demeaned versions of $f_{1,t}^s$, $f_{2,t}^s$ and f_t^c. From Lemma B.3, the k^c largest eigenvalues of matrix $\tilde{R}$ are $\tilde{\rho}_1^2 = ... = \tilde{\rho}_{k^c}^2 = 1$, while the remaining $k_1 - k^c$ eigenvalues are the eigenvalues of matrix $\tilde{R}_{ss}$ and are such that $1 > \tilde{\rho}_{k^c+1}^2 \geq ... \geq \tilde{\rho}_{k_1}^2 > 0$, a.s.. Let us define:

$$\underset{(k_1 \times k^c)}{E_c} = \begin{bmatrix} I_{k^c} \\ 0 \end{bmatrix}, \qquad \underset{(k_1 \times (k_1 - k^c))}{E_s} = \begin{bmatrix} 0 \\ I_{k_1 - k^c} \end{bmatrix}. \tag{B.17}$$

Then, the eigenvectors associated with the first k^c unit eigenvalues of $\tilde{R}$ are spanned by the columns of matrix E_c. The columns of matrices E_c and E_s span the space $\mathbb{R}^{k_1}$.

B.3.4 Eigenvalues and eigenvectors of matrix $\hat{R}$ obtained by perturbation methods

The estimators of the first k^c canonical correlations are such that $\hat{\rho}_\ell^2$, with $\ell = 1, ..., k^c$ are the k^c largest eigenvalues of matrix $\hat{R}$. We now derive their asymptotic expansion under the null hypothesis $H(k^c)$ using perturbations arguments applied to equation (B.13). Let $\hat{W}_1^*$ be a (k_1, k^c) matrix whose columns are eigenvectors of matrix $\hat{R}$ associated with the eigenvalues $\hat{\rho}_\ell^2$, with $\ell = 1, ..., k^c$. We have:

$$\hat{R}\hat{W}_1^* = \hat{W}_1^*\hat{\Lambda}, \tag{B.18}$$

where $\hat{\Lambda} = diag(\hat{\rho}_\ell^2, \ell = 1, ..., k^c)$ is the (k^c, k^c) diagonal matrix containing the k^c largest eigenvalues of $\hat{R}$. We know from the previous subsection that the eigenspace associated with the largest eigenvalue of $\tilde{R}$ (equal to 1) has dimension k^c and is spanned by the columns of matrix E_c. Since the columns of E_c and E_s span $\mathbb{R}^{k_1}$, we can write the following expansions:

$$\hat{W}_1^* = E_c\hat{\mathcal{U}} + E_s\hat{\alpha}, \qquad \hat{\Lambda} = I_{k^c} + \hat{M}, \tag{B.19}$$

where E_c and E_s are defined in equation (B.17), the stochastic (k^c, k^c) matrix $\hat{\mathcal{U}}$ is nonsingular with probability approaching (w.p.a.) 1, stochastic matrix $\hat{M}$ is diagonal, and $\hat{\alpha}$ is a $(k_1 - k^c, k^c)$ stochastic matrix. By the continuity of the matrix eigenvalue and eigenfunction mappings, and Lemma B.1, we have that $\hat{\alpha}$ and $\hat{M}$ converge in probability to null matrices as $N_1, N_2, T \to \infty$ at rate $O_p(\delta_{N,T})$. By substituting the expansions (B.13) and (B.19) into the eigenvalue-eigenvector equation (B.18), using the characterization of matrix $\tilde{R}$ obtained in Lemma B.3, and keeping terms up to order $O_p(\delta_{N,T}^3)$, we get expressions for matrices $\hat{\alpha}$ and $\hat{M}$. These yield the asymptotic expansions of the eigenvalues and eigenvectors of matrix $\hat{R}$ provided in the next Lemma.

LEMMA B.4. *Under Assumptions B.1-B.4, B.5 b)-c), B.6 a) and B.7, we have:*

$$\begin{aligned}
\hat{\Lambda} &= I_{k^c} + \hat{\mathcal{U}}^{-1}\tilde{\Sigma}_{cc}^{-1}\left\{\hat{\Psi}_{cc}^* + \hat{\Psi}_{cs}^*(I_{k_1-k^c} - \tilde{R}_{ss})^{-1}\hat{\Psi}_{sc} - \tilde{\Sigma}_{c,1}(I_{k_1-k^c} - \tilde{R}_{ss})^{-1}\hat{\Psi}_{sc}\tilde{\Sigma}_{cc}^{-1}\hat{\Psi}_{cc}^*\right\}\hat{\mathcal{U}} \\
&\quad + O_p(\delta_{N,T}^3), \qquad \text{(B.20)} \\
\hat{W}_1^* &= \left(E_c + E_s(I_{k_1-k^c} - \tilde{R}_{ss})^{-1}\left[\hat{\Psi}_{sc} + \hat{\Psi}_{ss}(I_{k_1-k^c} - \tilde{R}_{ss})^{-1}\hat{\Psi}_{sc}\right.\right. \\
&\quad \left.\left. -(I_{k_1-k^c} - \tilde{R}_{ss})^{-1}\hat{\Psi}_{sc}\left(\tilde{\Sigma}_{cc}^{-1}\hat{\Psi}_{cc}^*\right)\right]\right)\hat{\mathcal{U}} + O_p(\delta_{N,T}^3), \qquad \text{(B.21)}
\end{aligned}$$

where $\hat{\Psi}_{cc}$, $\hat{\Psi}_{cs} = \hat{\Psi}_{sc}'$, $\hat{\Psi}_{ss}$ denote the upper-left (k^c, k^c) block, the upper-right (k^c, k_1^s) block and the lower-right (k_1^s, k_1^s) block of matrix $\hat{\Psi}$, and similarly for the blocks of $\hat{\Psi}^*$.

In equations (B.20) and (B.21), in the terms that are of second-order with respect to $\hat{\Psi}$, we can replace $\hat{\Psi}$ by $\hat{\Psi}^{(I)}$ without changing the order $O_p(\delta_{N,T}^3)$ of the remainder term. Note that the approximation in (B.20) holds for the terms in the main diagonal, as matrix $\hat{\Lambda}$ has been defined to be diagonal.

B.3.5 Asymptotic expansion of $\sum_{\ell=1}^{k^c} \hat{\rho}_\ell$

Let us now derive an asymptotic expansion for the sum of the k^c largest canonical correlations $\sum_{\ell=1}^{k^c} \hat{\rho}_\ell$. By using the expansion of the matrix square root function in a neighbourhood of the identity, i.e. $(I + X)^{1/2} = I + \frac{1}{2}X - \frac{1}{8}X^2 + O_p(\delta_{N,T}^3)$ for $X = O_p(\delta_{N,T})$, from equation (B.20) we have:

$$\begin{aligned}
\hat{\Lambda}^{1/2} &= I_{k^c} + \frac{1}{2}\hat{\mathcal{U}}^{-1}\tilde{\Sigma}_{cc}^{-1}\left\{\hat{\Psi}_{cc}^* - \frac{1}{4}\hat{\Psi}_{cc}^*\tilde{\Sigma}_{cc}^{-1}\hat{\Psi}_{cc}^* + \hat{\Psi}_{cs}^*(I_{k_1-k^c} - \tilde{R}_{ss})^{-1}\hat{\Psi}_{sc}\right. \\
&\quad \left. -\tilde{\Sigma}_{c,1}(I_{k_1-k^c} - \tilde{R}_{ss})^{-1}\hat{\Psi}_{sc}\tilde{\Sigma}_{cc}^{-1}\hat{\Psi}_{cc}^*\right\}\hat{\mathcal{U}} + O_p(\delta_{N,T}^3).
\end{aligned}$$

Using $\sum_{\ell=1}^{k^c} \hat{\rho}_\ell = tr\left\{\hat{\Lambda}^{1/2}\right\}$, this implies:

$$\begin{aligned}
\sum_{\ell=1}^{k^c} \hat{\rho}_\ell &= k^c + \frac{1}{2}tr\left\{\tilde{\Sigma}_{cc}^{-1}\left[\hat{\Psi}_{cc}^* - \frac{1}{4}\hat{\Psi}_{cc}^{*(I)}\tilde{\Sigma}_{cc}^{-1}\hat{\Psi}_{cc}^{*(I)} + \hat{\Psi}_{cs}^{*(I)}(I_{k_1-k^c} - \tilde{R}_{ss})^{-1}\hat{\Psi}_{sc}^{(I)}\right.\right. \\
&\quad \left.\left. -\tilde{\Sigma}_{c,1}(I_{k_1-k^c} - \tilde{R}_{ss})^{-1}\hat{\Psi}_{sc}^{(I)}\tilde{\Sigma}_{cc}^{-1}\hat{\Psi}_{cc}^{*(I)}\right]\right\} + O_p(\delta_{N,T}^3). \qquad \text{(B.22)}
\end{aligned}$$

The next Lemma provides the asymptotic expansions of the terms within the the trace operator in the r.h.s. of (B.22) by plugging the expressions of $\hat{\Psi}^*$ and its components from Lemma B.2.

LEMMA B.5. *Under Assumptions B.1-B.4, B.5 b)-c), B.6 a) and B.7 we have:*

$$\begin{aligned}\sum_{\ell=1}^{k^c}\hat{\rho}_\ell &= k^c - \frac{1}{2N} tr\left\{\tilde{\Sigma}_{cc}^{-1}\frac{1}{T}\sum_{t=1}^{T} E[(\mu_N \breve{u}_{1t}^{(c)} - \breve{u}_{2t}^{(c)})(\mu_N \breve{u}_{1t}^{(c)} - \breve{u}_{2t}^{(c)})'|\mathcal{F}_t]\right\} - \frac{1}{2T^2} tr\left\{\tilde{\Sigma}_{cc}^{-1}\frac{1}{T}\sum_{t=1}^{T}\widetilde{\Delta b}_t^{(c)}\widetilde{\Delta b}_t^{(c)\prime}\right\} \\ &\quad -\frac{1}{2N\sqrt{T}} tr\left\{\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[(\mu_N \breve{u}_{1t}^{(c)} - \breve{u}_{2t}^{(c)})(\mu_N \breve{u}_{1t}^{(c)} - \breve{u}_{2t}^{(c)})' - E[(\mu_N \breve{u}_{1t}^{(c)} - \breve{u}_{2t}^{(c)})(\mu_N \breve{u}_{1t}^{(c)} - \breve{u}_{2t}^{(c)})'|\mathcal{F}_t]\right]\right\} \\ &\quad + O_p\left(\delta_{N,T}^3\right) + o_p\left(\epsilon_{N,T}\right), \end{aligned} \qquad \text{(B.23)}$$

where $\epsilon_{N,T} := \frac{1}{N\sqrt{T}}$*. The terms in the curly brackets are* $O_p(1)$*.*

We have $\delta_{N,T}^3 = o(\epsilon_{N,T})$ from the definitions of $\delta_{N,T}$ in Lemma B.1 and of $\epsilon_{N,T}$ in Lemma B.5, and the condition $\sqrt{T} \ll N \ll T^2$ in Assumption B.1. Therefore, the leading stochastic terms in the difference $\sum_{\ell=1}^{k^c}\hat{\rho}_\ell - k^c$ are of order $O_p\left(\frac{1}{N}\right)$, $O_p\left(\frac{1}{T^2}\right)$ and $O_p\left(\frac{1}{N\sqrt{T}}\right)$. If the assumption $N \ll T^2$ was violated, an additional term of order the term of order $\frac{1}{T\sqrt{NT}}$ would appear on the r.h.s. of equation (B.23), analogously to what happens in the r.h.s. of the analogous equation in Lemma B.5 of AGGR.[14] This additional term is negligible w.r.t. the dominating term of order $\frac{1}{N\sqrt{T}}$, and therefore absorbed in the term $o_p\left(\epsilon_{N,T}\right)$ in our equation (B.23) under Assumption B.1.

From the definition of matrices $\tilde{\Sigma}_U$ and $\tilde{\Sigma}_B$ in Theorem A.1, we have $\frac{1}{T}\sum_{t=1}^{T} E[(\mu_N \breve{u}_{1t}^{(c)} - \breve{u}_{2t}^{(c)})(\mu_N \breve{u}_{1t}^{(c)} - \breve{u}_{2t}^{(c)})'|\mathcal{F}_t] = \tilde{\Sigma}_U$ and $\frac{1}{T}\sum_{t=1}^{T}\widetilde{\Delta b}_t^{(c)}\widetilde{\Delta b}_t^{(c)\prime} = \tilde{\Sigma}_B$. Moreover, let us define the process

$$U_t := \mu_N \breve{u}_{1t}^{(c)} - \breve{u}_{2t}^{(c)}. \qquad \text{(B.24)}$$

Process U_t depends on N_1, N_2, but we do not make this dependence explicit for expository purpose. By using these definitions, from Lemma B.5 we get:

$$\sum_{\ell=1}^{k^c}\hat{\rho}_\ell - k^c + \frac{1}{2N} tr\left\{\tilde{\Sigma}_{cc}^{-1}\tilde{\Sigma}_U\right\} + \frac{1}{2T^2} tr\left\{\tilde{\Sigma}_{cc}^{-1}\tilde{\Sigma}_B\right\} = -\frac{1}{2N\sqrt{T}}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[U_t'U_t - E(U_t'U_t|\mathcal{F}_t)\right]\right) + o_p\left(\epsilon_{N,T}\right). \qquad \text{(B.25)}$$

Under our set of assumptions the term $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[U_t'U_t - E(U_t'U_t|\mathcal{F}_t)\right]$ is $O_p(1)$, as in the next subsection we show that it is asymptotically Gaussian distributed. The remainder term $o_p\left(\epsilon_{N,T}\right)$ in the r.h.s. of (B.25) is negligible with respect to the first term in the r.h.s. The result in equation (B.25) is analogous to the one in equation (B.15) in AGGR, with the notable differences being the new definitions of the term in U_t provided in our equation (B.24), and of matrices $\tilde{\Sigma}_U$ and $\tilde{\Sigma}_B$. Moreover, as mentioned above, under the assumption $N \ll T^2$, the term of order $\frac{1}{T\sqrt{NT}}$ appearing in Lemma B.5 of AGGR is negligible w.r.t. the dominating term of order $\frac{1}{N\sqrt{T}}$.

[14] For example, assumption $N \ll T^2$ is violated in the case $T^2 \ll N \ll T^{5/2}$ allowed by AGGR, but not by our Assumption B.1

B.3.6 Asymptotic distribution of the test statistic under the null hypothesis $H(k^c)$

From the asymptotic expansion (B.25) we obtain the asymptotic distribution of $\hat{\xi}(k^c) = \sum_{\ell=1}^{k^c} \hat{\rho}_\ell$ under the null hypothesis $H(k^c)$ of k^c common factors. First, we apply a CLT for weakly dependent triangular array data to prove the asymptotic normality of $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \mathcal{Z}_{N,t}$ as $N, T \to \infty$, where $\mathcal{Z}_{N,t} := U_t' U_t - E(U_t' U_t | \mathcal{F}_t)$ depends on N_1, N_2 via process U_t defined in (B.24).

i) CLT for Near-Epoch Dependent (NED) processes

Let process $V_{N_1,N_2,t} \equiv V_t$ be as defined in Assumption B.6, and let $\mathcal{V}_{t-m}^{t+m} = \sigma(V_s, t-m \leq s \leq t+m)$ for any positive integer m, with $\mathcal{V}_t \equiv \mathcal{V}_{-\infty}^{t}$.

LEMMA B.6. *Under Assumptions B.3, B.4 a), b), B.5 b) and B.6 a)-c) we have:*

(i) $\mathcal{Z}_{N,t}$ *is measurable w.r.t.* $\mathcal{V}_t$*, and* $E[\mathcal{Z}_{N,t}] = 0$ *for all* $t \geq 1$ *and* $N_1, N_2 \geq 1$*,*

(ii) $\sup_{t \geq 1, N_1, N_2 \geq 1} E\left[\|\mathcal{Z}_{N,t}\|^r\right] < \infty$*, for a constant* $r > 2$*,*

(iii) Process $(\mathcal{Z}_{N,t})$ *is* L^2 *Near Epoch Dependent (*L^2*-NED) of size* -1 *on process* (V_t)*, and* (V_t) *is strong*

mixing of size $-r/(r-2)$*, uniformly in* $N_1, N_2 \geq 1$*,* [15]

(iv) Matrix $\Omega_U := \lim_{T,N \to \infty} V\left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \mathcal{Z}_{N,t}\right)$ *is positive definite and such that*

$$\Omega_U = \sum_{h=-\infty}^{\infty} \Gamma(h), \qquad \Gamma(h) := \lim_{N \to \infty} Cov\left(\mathcal{Z}_{N,t}, \mathcal{Z}_{N,t-h}\right). \tag{B.26}$$

Then, by an application of the univariate CLT in Corollary 24.7 in Davidson (1994) and the Cramér-Wold device, we have that:

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \mathcal{Z}_{N,t} \xrightarrow{d} N\left(0, \Omega_U\right), \tag{B.27}$$

as $T, N \to \infty$. Let us now compute the limit autocovariance matrix $\Gamma(h)$ explicitly. By the Law of Iterated Expectation and $E[\mathcal{Z}_{N,t}|\mathcal{F}_t] = 0$, we have:

$$\Gamma(h) = \lim_{N \to \infty} E\left[Cov\left(\mathcal{Z}_{N,t}, \mathcal{Z}_{N,t-h} | \mathcal{F}_t\right)\right]. \tag{B.28}$$

Moreover, from Assumptions B.3 and B.5 *a)*, vector $(U_t', U_{t-h}')'$ is asymptotically Gaussian for any h,

[15] That is, $\left\|\mathcal{Z}_{N,t} - E[\mathcal{Z}_{N,t}|\mathcal{V}_{t-m}^{t+m}]\right\|_2 \leq \xi(m)$, uniformly in $t \geq 1$ and $N_1, N_2 \geq 1$, where $\xi(m) = O(m^{-\psi})$ for some $\psi > 1$.

t as $N \to \infty$:

$$\begin{pmatrix} U_t \\ U_{t-h} \end{pmatrix} \xrightarrow{d} \begin{pmatrix} U_t^\infty \\ U_{t-h}^\infty \end{pmatrix} \sim N\left(0, \begin{bmatrix} \Sigma_{U,t}(0) & \Sigma_{U,t}(h) \\ \Sigma_{U,t}(h)' & \Sigma_{U,t}(0) \end{bmatrix}\right), \quad (\mathcal{F}_t\text{-stably}). \tag{B.29}$$

We use the Lebesgue Lemma to interchange the limes for $N \to \infty$ and the outer expectation in the r.h.s. of (B.28), and the fact that convergence in distribution plus uniform integrability imply convergence of the expectation for a sequence of random variables (see Theorem 25.12 in Billingsley (1995)) to show the next lemma.

LEMMA B.7. *Under Assumptions B.3 and B.5 b), we have:*

$$\Gamma(h) = E\left[Cov(U_t^{\infty\,\prime} U_t^\infty, U_{t-h}^{\infty\prime} U_{t-h}^\infty | \mathcal{F}_t)\right].$$

Lemma B.7 allows to deploy the joint asymptotic Gaussian distribution of $(U_t^{\infty\,\prime}, U_{t-h}^{\infty\prime})'$ to compute the limit autocovariance $\Gamma(h)$. To compute matrix $\Gamma(h)$, we use Theorem 12 p. 284 in Magnus and Neudecker (2007) and Theorem 10.21 in Schott (2005). We get $Cov(U_t^{\infty\,\prime} U_t^\infty, U_{t-h}^{\infty\prime} U_{t-h}^\infty | \mathcal{F}_t) = 2tr\left\{\Sigma_{U,t}(h)\Sigma_{U,t}(h)'\right\}$. Therefore from (B.26) and Lemma B.7 we get:

$$\Omega_U = \sum_{h=-\infty}^{\infty} 2tr\left\{E\left[\Sigma_{U,t}(h)\Sigma_{U,t}(h)'\right]\right\} = 4\Omega_{U,1}. \tag{B.30}$$

ii) Asymptotic Gaussian distribution of the test statistic

Let us define the constant $D_{N,T} = \frac{1}{2N\sqrt{T}}$. From equations (B.25) and (B.30), and by using: $(D_{N,T}^2 \Omega_U)^{1/2} = \frac{1}{N\sqrt{T}} \Omega_{U,1}^{1/2}$, and $N\sqrt{T}\Omega_{U,1}^{-1/2} = O\left(N\sqrt{T}\right) = O(\epsilon_{N,T}^{-1})$, under the hypothesis of k^c common factors in each group the statistics $\hat{\xi}(k^c) = \sum_{\ell=1}^{k^c} \hat{\rho}_\ell$ is such that:

$$\begin{aligned} & N\sqrt{T}\Omega_{U,1}^{-1/2}\left[\hat{\xi}(k^c) - k^c + \frac{1}{2N} tr\left\{\tilde{\Sigma}_{cc}^{-1}\tilde{\Sigma}_U\right\} + \frac{1}{2T^2} tr\left\{\tilde{\Sigma}_{cc}^{-1}\tilde{\Sigma}_B\right\}\right] \\ = \; & -(D_{N,T}^2 \Omega_U)^{-1/2} D_{N,T} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \mathcal{Z}_{N,T} + o_p(1). \end{aligned}$$

From equation (B.27), the r.h.s. converges in distribution to a standard normal distribution, which yields Theorem A.1.

B.4 Proof of Theorem A.2

To establish the asymptotic distribution of the feasible statistic in Theorem A.2 we need to control the effect of replacing the re-centering and scaling terms by means of their estimates. The latter involve factors and loadings estimates. Hence, in Section B.6 we derive uniform asymptotic expansions of factors and loadings estimators. These results are instrumental for the proof of Theorem A.2, as well as for the proofs of other results in this paper. In Subsection B.4.1 and B.4.2 we show the statements in Part i) and in Part ii) of Theorem A.2, respectively.

B.4.1 Proof of Part (i)

Let us first consider the asymptotic distribution of $\tilde{\xi}(k^c)$ under the null hypothesis of k^c common factors. Under the assumptions of Theorem A.2, the infeasible asymptotic distribution in Theorem A.1 becomes:

$$N\sqrt{T}\Omega_{U,1}^{-1/2}\left[\hat{\xi}(k^c) - k^c + \frac{1}{2N}tr\left\{\tilde{\Sigma}_{cc}^{-1}\tilde{\Sigma}_U\right\}\right] \xrightarrow{d} N(0,1), \tag{B.31}$$

where $\Omega_{U,1} = \frac{1}{2}tr\left\{\Sigma_U(0)^2\right\}$ and we use (A.21) and $\tilde{\Sigma}_B = 0$. Theorem A.2 i) follows, if we prove:

$$tr\left\{\hat{\Sigma}_U\right\} = tr\left\{\tilde{\Sigma}_{cc}^{-1}\tilde{\Sigma}_U\right\} + o_p\left(\frac{1}{\sqrt{T}}\right), \tag{B.32}$$

$$tr\left\{\hat{\Sigma}_U^2\right\} = tr\left\{\Sigma_U(0)^2\right\} + o_p(1). \tag{B.33}$$

Indeed, the statistic $\tilde{\xi}(k^c)$ can be rewritten as:

$$\begin{aligned}\tilde{\xi}(k^c) = & \left[\frac{1}{2}tr\left\{\hat{\Sigma}_U^2\right\}\Big/\Omega_{U,1}\right]^{-1/2}\left\{N\sqrt{T}\Omega_{U,1}^{-1/2}\left[\hat{\xi}(k^c) - k^c + \frac{1}{2N}tr\left\{\tilde{\Sigma}_{cc}^{-1}\tilde{\Sigma}_U\right\}\right]\right. \\ & \left. + O_p\left(\sqrt{T}\left[tr\left\{\hat{\Sigma}_U\right\} - tr\left\{\tilde{\Sigma}_{cc}^{-1}\tilde{\Sigma}_U\right\}\right]\right)\right\},\end{aligned}$$

where the ratio $\frac{1}{2}tr\left\{\hat{\Sigma}_U^2\right\}\Big/\Omega_{U,1}$ converges in probability to 1 from (B.33), the term within the curly brackets in the first line in the r.h.s. converges in distribution to a standard normal distribution from (B.31), and the term on the second line on the r.h.s. is $o_p(1)$ from (B.32).

Le us now prove equations (B.32) and (B.33) by deriving the asymptotic expansions of $\hat{\Sigma}_U$ and $\tilde{\Sigma}_{cc}^{-1}$. To derive the asymptotic expansion of $\hat{\Sigma}_U$, we use its definition $\hat{\Sigma}_U = \mu_N^2\hat{\Sigma}_{u,11}^{(cc)} + \hat{\Sigma}_{u,22}^{(cc)}$, where the matrices $\hat{\Sigma}_{u,jj} = \left(\frac{1}{N_j}\hat{\Lambda}_j'\hat{\Lambda}_j\right)^{-1}\left(\frac{1}{N_j}\hat{\Lambda}_j'\hat{\Gamma}_j^*\hat{\Lambda}_j\right)\left(\frac{1}{N_j}\hat{\Lambda}_j'\hat{\Lambda}_j\right)^{-1}$, $j = 1, 2$, involve the estimated loadings and residuals. We plug in the uniform asymptotic expansions from Proposition B.6 in Section B.6 to show the next result.

LEMMA B.8. *Under Assumptions B.1 - B.9, i) The asymptotic expansion of estimator $\hat{\Lambda}_j'\hat{\Lambda}_j/N_j$ is:*

$$\frac{\hat{\Lambda}_j'\hat{\Lambda}_j}{N_j} = \hat{\mathcal{U}}_j' \left[\tilde{\Sigma}_{\Lambda,j} + \frac{1}{\sqrt{T}}\left(L_{\Lambda,j} + L_{\Lambda,j}'\right)\right]\hat{\mathcal{U}}_j + o_p\left(\frac{1}{\sqrt{T}}\right), \tag{B.34}$$

for $j = 1, 2$, where $\tilde{\Sigma}_{\Lambda,j} = \frac{1}{N_j}\Lambda_j'\Lambda_j$ with $\Lambda_j = [\Lambda_j^c \vdots \Lambda_j^s]$, and $L_{\Lambda,j} = \tilde{\Sigma}_{\Lambda,j}Q_j$ and:

$$\hat{\mathcal{U}}_j = \begin{bmatrix} \hat{\mathcal{H}}_c & 0 \\ 0 & \hat{\mathcal{H}}_{s,j} \end{bmatrix}, \qquad Q_j = \begin{bmatrix} 0 & 0 \\ \sqrt{T}\tilde{\Sigma}_{j,c}\tilde{\Sigma}_{cc}^{-1} & 0 \end{bmatrix}, \tag{B.35}$$

and $\hat{\mathcal{H}}_c$, $\hat{\mathcal{H}}_{s,j}$ are non-singular matrices w.p.a. 1. ii) The asymptotic expansion of $\hat{\Lambda}_j'\hat{\Gamma}_j^*\hat{\Lambda}_j/N_j$ is:

$$\frac{1}{N_j}\hat{\Lambda}_j'\hat{\Gamma}_j^*\hat{\Lambda}_j = \hat{\mathcal{U}}_j' \left[\tilde{\Omega}_{jj} + \frac{1}{\sqrt{T}}\left(L_{\Omega,j} + L_{\Omega,j}'\right)\right]\hat{\mathcal{U}}_j + o_p\left(\frac{1}{\sqrt{T}}\right), \tag{B.36}$$

for $j = 1, 2$, where $\tilde{\Omega}_{jj} = \frac{1}{N_j}\Lambda_j'\Gamma_j^*\Lambda_j$, with $\Gamma_j^* = diag(\gamma_{j,ii}^*, i = 1, ..., N_j)$, and $L_{\Omega,j} = \tilde{\Omega}_{jj}Q_j$.

Equation (B.34) allows to compute the asymptotic approximation of $\left(\frac{1}{N_j}\hat{\Lambda}_j'\hat{\Lambda}_j\right)^{-1}$ by matrix inversion:

$$\left(\frac{1}{N_j}\hat{\Lambda}_j'\hat{\Lambda}_j\right)^{-1} = \hat{\mathcal{U}}_j^{-1}\left[\tilde{\Sigma}_{\Lambda,j}^{-1} - \frac{1}{\sqrt{T}}\tilde{\Sigma}_{\Lambda,j}^{-1}\left(L_{\Lambda,j} + L_{\Lambda,j}'\right)\tilde{\Sigma}_{\Lambda,j}^{-1}\right]\left(\hat{\mathcal{U}}_j'\right)^{-1} + o_p\left(\frac{1}{\sqrt{T}}\right). \tag{B.37}$$

Substituting equations (B.37) and (B.36) into the expression of $\hat{\Sigma}_{u,jj}$ and rearranging terms, we get:

$$\begin{aligned} \hat{\Sigma}_{u,jj} &= \hat{\mathcal{U}}_j^{-1}\tilde{\Sigma}_{\Lambda,j}^{-1}\left[\tilde{\Omega}_{jj} + \frac{1}{\sqrt{T}}\left(L_{\Omega,j} + L_{\Omega,j}'\right) - \frac{1}{\sqrt{T}}\tilde{\Omega}_{jj}\tilde{\Sigma}_{\Lambda,j}^{-1}\left(L_{\Lambda,j} + L_{\Lambda,j}'\right)\right. \\ &\left. - \frac{1}{\sqrt{T}}\left(L_{\Lambda,j} + L_{\Lambda,j}'\right)\tilde{\Sigma}_{\Lambda,j}^{-1}\tilde{\Omega}_{jj}\right]\tilde{\Sigma}_{\Lambda,j}^{-1}\left(\hat{\mathcal{U}}_j'\right)^{-1} + o_p\left(\frac{1}{\sqrt{T}}\right). \end{aligned}$$

Therefore, from the definitions of matrices $L_{\Omega,j}$ and $L_{\Lambda,j}$ in Lemma B.8, we have:

$$\hat{\Sigma}_{u,jj} = \hat{\mathcal{U}}_j^{-1}\left(\tilde{\Sigma}_{u,jj} + \frac{1}{\sqrt{T}}(L_{U,j} + L_{U,j}')\right)\left(\hat{\mathcal{U}}_j'\right)^{-1} + o_p\left(\frac{1}{\sqrt{T}}\right), \tag{B.38}$$

where $\tilde{\Sigma}_{u,jj} = \tilde{\Sigma}_{\Lambda,j}^{-1}\tilde{\Omega}_{jj}\tilde{\Sigma}_{\Lambda,j}^{-1}$ and $L_{U,j} = -Q_j\tilde{\Sigma}_{u,jj}$, for $j = 1, 2$. In particular, the upper-left (k^c, k^c) block of $L_{U,j}$ vanishes, i.e. $(L_{U,j})^{(cc)} = 0$ for $j = 1, 2$.

From equation (B.38) we get the asymptotic expansion for $\hat{\Sigma}_U = \mu_N^2 \hat{\Sigma}_{u,11}^{(cc)} + \hat{\Sigma}_{u,22}^{(cc)}$:

$$
\begin{aligned}
\hat{\Sigma}_U &= \hat{\mathcal{H}}_c^{-1} \left(\left[\mu_N^2 \tilde{\Sigma}_{u,11} + \tilde{\Sigma}_{u,22} \right]^{(cc)} + \frac{1}{\sqrt{T}} \left[\mu_N^2 (L_{U,1} + L_{U,1}') + L_{U,2} + L_{U,2}' \right]^{(cc)} \right) \left(\hat{\mathcal{H}}_c' \right)^{-1} + o_p\left(\frac{1}{\sqrt{T}} \right) \\
&= \hat{\mathcal{H}}_c^{-1} \tilde{\Sigma}_U \left(\hat{\mathcal{H}}_c' \right)^{-1} + o_p\left(\frac{1}{\sqrt{T}} \right). \qquad \text{(B.39)}
\end{aligned}
$$

Moreover, Proposition B.6 ii) implies $\tilde{\Sigma}_{cc}^{-1} = \left(\hat{\mathcal{H}}_c^{-1} \right)' \hat{\mathcal{H}}_c^{-1} + o_p\left(\frac{1}{\sqrt{T}} \right)$. This equation, together with the asymptotic expansion (B.39) and the commutative property of the trace operator, imply equation (B.32). Similarly, the asymptotic expansion (B.39) and the convergence $\tilde{\Sigma}_U \to \Sigma_U(0)$ imply equation (B.33).

B.4.2 Proof of Part (ii)

In order to prove Theorem A.2 *(ii)*, we consider the behavior of statistic $\tilde{\xi}(k^c)$ under the alternative hypothesis H_1 of less than k^c common factors. Specifically, let $r < k^c$ be the true number of common factors in the DGP. The statistic is given by: $\tilde{\xi}(k^c) = N\sqrt{T} \left(\frac{1}{2} tr\{\hat{\Sigma}_U^2\} \right)^{-1/2} \left[\sum_{\ell=1}^{k^c} \hat{\rho}_\ell - k^c + \frac{1}{2N} tr\left\{ \hat{\Sigma}_U \right\} \right]$. We rely on the following Lemma. For its proof we assume that $\hat{f}_t^c$ is used to estimate the common factor in panel $j = 1$, while estimator $\hat{f}_t^{c*}$ is used in panel $j = 2$.

LEMMA B.9. *Under the alternative hypothesis $H(r)$, with $r < k^c$, we have $\|\hat{\Sigma}_U\| \leq C$, w.p.a. 1, for a constant $C > 0$.*

From Lemma B.9 and using $\sum_{\ell=1}^{k^c} \hat{\rho}_\ell = \sum_{\ell=1}^{k^c} \rho_\ell + o_p(1)$, where the $o_p(1)$ term follows from the continuity of the eigenvalues mapping, we get:

$$
\tilde{\xi}(k^c) = N\sqrt{T} \left(\frac{1}{2} tr\{\hat{\Sigma}_U^2\} \right)^{-1/2} \left[\sum_{\ell=1}^{k^c} \rho_\ell - k^c + o_p(1) \right].
$$

Under $H(r)$, we have $r < k^c$ canonical correlations that are equal to 1, while the other ones are strictly smaller than 1. Therefore, $\sum_{\ell=1}^{k^c} \rho_\ell - k^c < 0$. Then, from Lemma B.9 we get $\tilde{\xi}(k^c) \leq -N\sqrt{T} c_1$, w.p.a. 1, for a constant $c_1 > 0$. The conclusion follows.

B.5 Additional proofs

This section contains the proofs of Proposition B.2 and Lemmas B.1 - B.9, stated in Sections B.3 and B.4. The section also includes the proof of the additional technical results needed to prove them.

B.5.1 Proof of Proposition B.2

The group factor model (2.10) implies that the ususal factor model for the N_j individuals in group j is:

$$y_{j,t} = \Lambda_j h_{j,t} + \varepsilon_{j,t}$$

where $y_{j,t} = [y_{j,1,t}, ..., y_{j,N_j,t}]'$, $\Lambda = [\lambda_{j,1}, ..., \lambda_{j,N_j}]' = [\Lambda_1^c,\ \Lambda_1^s]$, $h_{j,t} = [f_t^{c\prime},\ f_{j,t}^{s\prime}]'$ and $\varepsilon_{j,t} = [\varepsilon_{j,1,t}, ..., \varepsilon_{j,N_j,t}]'$. Therefore the model for the individual i in panel j is:

$$y_{j,i,t} = \lambda_{j,i}'\ h_{j,t} + \varepsilon_{j,i,t},$$

$i = 1, ..., N_j$, and $t = 1, ..., T$. Only for the remaining part of the proof of Proposition B.2 we omit the group index j since it is immaterial for the proof's arguments. We write the factor model for the generic individual i in group j simply as:

$$y_{i,t} = \lambda_i' h_t + \varepsilon_{i,t}, \quad i = 1, ..., N, \quad t = 1, ..., T, \tag{B.40}$$

where h_t is the $(k, 1)$ vector of unobservable factors for group j. In matrix notation, the model can be written as in equation (A.2), that is:

$$Y = H\Lambda' + \varepsilon,$$

where Y is the (T, N) matrix of observations and H is the (T, k) matrix of factor values. Analogously to Lettau and Pelger (2020a), define:

$$W := I_T + \left(\sqrt{\gamma_{RP} + 1} - 1\right) \frac{\mathbb{1}_T \mathbb{1}_T'}{T} = I_T + \tilde{\gamma}_{RP} \frac{\mathbb{1}_T \mathbb{1}_T'}{T} \tag{B.41}$$

with both $\gamma_{RP} \in [-1, +\infty)$, and $\tilde{\gamma}_{RP} := (\sqrt{\gamma_{RP} + 1} - 1) \in [-1, +\infty)$. These definitions imply:

$$W^2 = I_T + \gamma_{RP} \frac{\mathbb{1}_T \mathbb{1}_T'}{T}\ . \tag{B.42}$$

We introduce a set of high-level assumptions (Assumptions B.10-B.13 below) and show in Section B.5.1.7 that they are implied by Assumptions B.2-B.4, B.5 *b)-c)*, B.6 *a)*, B.7.

Assumption B.10. *The factors are such that their sample covariance matrix is* $H'W^2H/T =$

$\Sigma_h(W) + o_p(1)$ as $T \to \infty$, where the matrix

$$\Sigma_h(W) := E\left[\ (h_t + \tilde{\gamma}_{RP} E[h_t])\,(h_t + \tilde{\gamma}_{RP} E[h_t])'\ \right] \tag{B.43}$$

is positive definite. The loadings are such that $\Lambda'\Lambda/N = \Sigma_\lambda + o(1)$ as $N \to \infty$.

In the spacial case of $\tilde{\gamma}_{RP} = -1$ we have $\Sigma_h(W) = V(h_t) = I_k$, from Assumption B.2. Note that Assumption B.10 is different form Assumption C.1 in AGG, which instead imposed that $H'H/T = I_k + o_p(1)$. The (N, k) matrix of RP-PCA loading estimates $\hat{\Lambda} = [\hat{\lambda}_1, ..., \hat{\lambda}_N]'$ satisfies the following eigenvector-eigenvalue equation:

$$\frac{1}{NT}(Y'W^2Y)\,\hat{\Lambda} \;=\; \hat{\Lambda}\,\hat{V}, \tag{B.44}$$

where $\hat{V}$ is the (k, k) diagonal matrix of the k largest eigenvalues of matrix $Y'W^2Y/(NT)$, and the columns of matrix $\hat{\Lambda}$ are the associated normalized eigenvectors such that $\hat{\Lambda}'\hat{\Lambda}/N = I_k$.[16] Equivalently, $\hat{\Lambda}$ is defined as the eigenvectors matrix $Y'W^2Y/(NT)$ multiplied by $\sqrt{N}$. It is easy to see that $\hat{\Lambda}$ in (B.44) is an appropriate estimator for Λ. In fact, the latter is the (true) vector of loadings both in the original model (A.2), and also on the new "projected model" obtained by pre-multiplying the matrix of observations Y by W:

$$WY = WH\Lambda' + W\varepsilon. \tag{B.45}$$

This argument is used by Lettau and Pelger (2020a) to prove the asymptotic results for their RP-PC estimator extending those derived for classical PC in Bai (2003).[17] By defining $\breve{Y} := WY$, $\breve{H} := WH$ and $\breve{\varepsilon} := W\varepsilon$, the "projected model" (B.45) can be written as:

$$\breve{Y} = \breve{H}\Lambda' + \breve{\varepsilon}\,. \tag{B.46}$$

The (T, k) matrix of factor estimators $\hat{H} = [\hat{h}_1, ..., \hat{h}_T]'$ coincides with the estimator obtained as in Theorem 1 of the OA of Lettau and Pelger (2020a), and is obtained by cross-sectional regressions of the observed data on the estimated loadings at each date:

$$\hat{H} \;=\; Y\hat{\Lambda}(\hat{\Lambda}'\hat{\Lambda})^{-1} \;=\; \frac{1}{N}Y\hat{\Lambda}, \tag{B.47}$$

where the last equality follows from the normalization $\hat{\Lambda}'\hat{\Lambda}/N = I_k$. Equation (B.47) allows

[16] See Theorem 1 of the OA of Lettau and Pelger (2020a).

[17] Note that when $\gamma_{RP} = -1$, matrix $\frac{1}{NT}(Y'W^2Y)$ is a re-scaled version of the covariance matrix $\frac{1}{T}(Y'W^2Y)$ of the non-demeaned original data, and therefore the loadings estimators are a re-scaled version of the eigenvectors of the latter covariance matrix.

to establish an asymptotic expansion of the factor estimate with explicit characterization of the remainder term. This can be obtained by manipulating equations (B.44) and (B.47) using the next two Assumptions B.12, and by defining:

$$\breve{h}_t := h_t + \left(\sqrt{\gamma_{RP}+1}-1\right)\bar{h} = h_t + \tilde{\gamma}_{RP}\bar{h},$$

$$\breve{\varepsilon}_{i,t} := \varepsilon_{i,t} + \left(\sqrt{\gamma_{RP}+1}-1\right)\bar{\varepsilon}_{i,\cdot} = \varepsilon_{i,t} + \tilde{\gamma}_{RP}\bar{\varepsilon}_{i,\cdot}$$

where $\bar{h} = \frac{1}{T}\sum_{t=1}^{T} h_t$, and $\bar{\varepsilon}_{i,\cdot} = \frac{1}{T}\sum_{t=1}^{T}\varepsilon_{i,t}$,

$$\xi_t := \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_i\varepsilon_{i,t}, \qquad \bar{\xi} := \frac{1}{T}\sum_{t=1}^{T}\xi_t = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_i\bar{\varepsilon}_{i,\cdot},$$

and

$$\breve{\xi}_t := \xi_t + \tilde{\gamma}_{RP}\bar{\xi} = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_i\left[\varepsilon_{i,t} + \tilde{\gamma}_{RP}\bar{\varepsilon}_{i,\cdot}\right] = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_i\breve{\varepsilon}_{i,t} \ . \tag{B.48}$$

We also define $\breve{H} = [\breve{h}_1, ..., \breve{h}_T']'$. Then the following two Assumptions allow to prove Proposition B.3.

Assumption B.11. *We have (i)* $E[\breve{\varepsilon}_{it}] = 0$*; (ii)* $E[\breve{\varepsilon}_{it}^8] \leq M$ *and* $E[\|\breve{h}_{it}^{2r\vee 8}] \leq M$*; and (iii)* $E[\breve{\xi}_t^8] \leq M$ *and* $E[\|\breve{\xi}_{it}\|^{2r}] \leq M,$ *for a constant* $M < \infty$*, where* $r > 2$ *is defined in Assumption B.5 b).*
(iv) Let $\breve{\eta}_{ts} = \frac{1}{N}\sum_{i=1}^{N} E[\breve{\varepsilon}_{i,t}\breve{\varepsilon}_{i,s}]$*, with* $s < t$*. Then* $\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{t-1} E[\breve{\eta}_{ts}^4] \leq M$ *and* $E\left[\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\breve{\varepsilon}_{i,t}\breve{\varepsilon}_{i,s} - \breve{\eta}_{ts}^2\right)^2\right] \leq M.$

Assumption B.12. *We have (i)* $\frac{1}{\sqrt{NT}}H'W^2\varepsilon\Lambda = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\breve{h}_t\breve{\xi}_t' = O_p(1)$ *and* $E[\|\breve{\xi}_t\|^2] = O(1)$*,*

(ii) $\|\frac{1}{NT}\varepsilon W^2\varepsilon'\Lambda\| = O_p\left(\frac{1}{\sqrt{m}}\right)$*,*

(iii) $\|\frac{1}{NT}\varepsilon W^2\varepsilon'\| = O_p\left(\frac{1}{\sqrt{m}}\right)$*, where* $m := \min\{N, T\}$*,*

(iv) $\|\frac{1}{NT}\varepsilon\varepsilon' W^2 H\| = O_p\left(\frac{1}{\sqrt{m}}\right)$*,*

(v) Let $\breve{\aleph}_i := \frac{1}{\sqrt{T}}H'W^2\varepsilon = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\breve{h}_t\breve{\varepsilon}_{i,t}$*, then* $E[\|\breve{\aleph}_i\|^2] = O(1)$*.*

Note that by setting $\gamma_{RP} = 0$, Assumption B.12 implies: $\frac{1}{\sqrt{NT}}H'\varepsilon\Lambda = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}h_t\xi_t' = O_p(1)$, $E[\|\xi_t\|^2] = O(1)$, $\|\frac{1}{NT}\varepsilon\varepsilon' H\| = O_p\left(\frac{1}{\sqrt{m}}\right)$, $\|\frac{1}{NT}\varepsilon\varepsilon'\| = O_p\left(\frac{1}{\sqrt{m}}\right)$ and $E[\|\aleph_i\|^2] = O(1)$.

PROPOSITION B.3. *Under Assumptions B.10-B.12 we have:*

$$(\tilde{\mathscr{H}}')^{-1}\hat{h}_t - h_t \;=\; \frac{1}{\sqrt{N}}u_t + \frac{1}{T}\breve{b}_t + \frac{1}{\sqrt{NT}}\breve{d}_t + \breve{\vartheta}_t, \qquad t = 1, ..., T, \tag{B.49}$$

where matrix $\tilde{\mathscr{H}} = (\Lambda'\Lambda/N)(H'W^2H/T)(\Lambda'\hat{\Lambda}/N)\hat{V}^{-1}$ *is invertible w.p.a. 1, and:*

$$\begin{aligned}
u_t &= (\Lambda'\Lambda/N)^{-1}\xi_t \\
\breve{b}_t &= \left(\frac{\Lambda'\Lambda}{N}\right)^{-1}\breve{S}'\left[\eta_t^2\breve{h}_t \;+\; \left(\frac{H'W^2H}{T}\right)^{-1}\breve{\Pi}_4\left(\frac{\Lambda'\Lambda}{N}\right)h_t\right], \\
\breve{d}_t &= \left(\frac{\Lambda'\Lambda}{N}\right)^{-1}\breve{S}'\left[\breve{\Pi}_1 + \breve{\Pi}_1'\left(\frac{\Lambda'\Lambda}{N}\right)\right]h_t, \\
\breve{\vartheta}_t &= \left(\frac{\Lambda'\Lambda}{N}\right)^{-1}\breve{S}'\left[\frac{1}{\sqrt{NT}}\breve{\alpha}_t \;+\; \frac{1}{N}\breve{\Pi}_2 h_t\right] \;+\; \breve{r}_t \;+\; \breve{\mathcal{R}}_t,
\end{aligned}$$

with $\eta_t^2 = \underset{N\to\infty}{plim}\frac{1}{N}\sum_{i=1}^{N}E[\varepsilon_{i,t}^2|\mathcal{F}_t]$ *and* $\mathcal{F}_t$ *is the sigma-field generated by the* h_s *for* $s \leq t$*,*

$$r_t \;=\; r_{1,t} \;+\; r_{7,t} \;+\; r_{8,t} \;+\; \hat{\mathcal{R}}_t\;, \tag{B.50}$$

vector $r_{1,t}$ is defined as:

$$
\begin{aligned}
r_{1,t} &= \frac{1}{T}\left(\frac{\Lambda'\Lambda}{N}\right)^{-1}\hat{B}'\breve{S}'\left[\eta_t^2\breve{h}_t \;+\; \breve{\Pi}_4\left(\frac{\Lambda'\Lambda}{N}\right)h_t\right] \;+\; \frac{1}{N}\left(\frac{\Lambda'\Lambda}{N}\right)^{-1}\hat{B}'\breve{S}'\breve{\Pi}_2 h_t \\
&+ \frac{1}{\sqrt{NT}}\left(\frac{\Lambda'\Lambda}{N}\right)^{-1}\hat{B}'\breve{S}'\left[\breve{\Pi}_1+\breve{\Pi}_1'\left(\frac{\Lambda'\Lambda}{N}\right)\right]h_t \;+\; \frac{1}{\sqrt{NT}}\left(\frac{\Lambda'\Lambda}{N}\right)^{-1}\hat{B}'\breve{S}'\breve{\alpha}_t \\
&+ \frac{1}{\sqrt{NT}}M^{*\prime}(H'W^2H/T)^{-1}\breve{\Pi}_3 h_t \;+\; \frac{1}{\sqrt{NT}}M^{**\prime}\breve{\Pi}_3'\left(\frac{\Lambda'\Lambda}{N}\right)h_t \\
&+ \frac{1}{T\sqrt{N}}M^{*\prime}\kappa_t\breve{h}_t + \tilde{\gamma}_{RP}M^{*\prime}\left[\frac{1}{T^2}\eta_t^2\breve{h}_t + \frac{1}{T\sqrt{NT}}\breve{\alpha}_t^* + \frac{1}{T^3\sqrt{N}}\kappa_t\breve{h}_t\right] \\
&+ M^{*\prime}\left\{\frac{1}{\sqrt{NT}}\eta_t^2\breve{\xi}_t + \frac{1}{NT}\kappa_t\breve{\xi}_t + \frac{1}{N\sqrt{N}}\breve{\varphi}_t + \frac{1}{N\sqrt{T}}\breve{\gamma}_t + \tilde{\gamma}_{RP}\left[\frac{1}{T^2\sqrt{N}}\eta_t^2\breve{\xi}_t + \frac{1}{NT^2}\kappa_t\breve{\xi}_t + \frac{1}{NT\sqrt{T}}\breve{\alpha}_t^*\right]\right\} \\
&+ \frac{1}{N\sqrt{T}}\;M^{*\prime}\breve{\Pi}_1'\;\xi_t \;+\; M^{*\prime}\left\{\frac{1}{T^2}\breve{\eta}_t^4\breve{h}_t + \frac{1}{\sqrt{N}T^2}\breve{\kappa}_t\breve{\eta}_t^2\breve{h}_t + \frac{1}{\sqrt{NT}T}\breve{\breve{\alpha}}_t\right. \\
&+ \frac{1}{T^2\sqrt{N}}\breve{\kappa}_t\breve{\eta}_t^2\breve{h}_t + \frac{1}{NT^2}\breve{\kappa}_t^2\breve{h}_t + \frac{1}{NT\sqrt{N}}\breve{\breve{\varphi}}_t + \frac{1}{NT\sqrt{T}}\breve{\breve{\gamma}}_t \\
&+ \left.\frac{1}{T\sqrt{NT}}\breve{\eta}_t^2\breve{\breve{\alpha}}_t + \frac{1}{NT\sqrt{T}}\breve{\kappa}_t\breve{\breve{\alpha}}_t + \frac{1}{N\sqrt{NT}}\breve{\delta}_t + \frac{1}{NT}\breve{\chi}_t - \tilde{\gamma}_{RP}\left[\frac{1}{T^2\sqrt{N}}\breve{\beta}_{1,t} + \frac{1}{NT^2\sqrt{T}}\breve{\beta}_{2,t} + \frac{1}{NT^2}\breve{\beta}_{3,t}\right]\right\} \\
&+ \frac{1}{T\sqrt{N}}M^*\breve{\Pi}_4\xi_t + M^{**\prime}\left\{\frac{1}{NT}\left(\frac{1}{T}\sum_{t=1}^T\breve{\eta}_t^2\breve{\xi}_t\breve{\xi}_t'\right) + \frac{1}{NT\sqrt{N}}\left(\frac{1}{T}\sum_{t=1}^T\breve{\kappa}_t\breve{\xi}_t\breve{\xi}_t'\right)\right. \\
&+ \left.\frac{1}{N^2}\left(\frac{1}{T}\sum_{t=1}^T\breve{\xi}_t\breve{\breve{\varphi}}_t'\right)' + \frac{1}{N\sqrt{TN}}\left(\frac{1}{T}\sum_{t=1}^T\breve{\xi}_t\breve{\breve{\gamma}}_t\right)'\right\}h_t \\
&+ \frac{1}{N\sqrt{NT}}M^{**\prime}\breve{\Pi}_1'\left(\frac{1}{T}\sum_{t=1}^T\breve{\xi}_t\breve{\xi}_t'\right)h_t \;+\; \frac{1}{NT}M^{**\prime}\breve{\Pi}_1'\breve{\Pi}_1'\left(\frac{\Lambda'\Lambda}{N}\right)h_t \\
&+ M^{**\prime}\left[\frac{1}{NT\sqrt{T}}\breve{\beta}_{4,t} + \frac{1}{NT\sqrt{NT}}\breve{\beta}_{5,t} + \frac{1}{N^2\sqrt{T}}\breve{\beta}_{6,t} + \frac{1}{NT\sqrt{N}}\breve{\beta}_{7,t}\right] \\
&+ \frac{1}{N\sqrt{NT}}M^{**\prime}\breve{\Pi}_1'\left\{\frac{1}{\sqrt{N}T}\eta_t^2\breve{\xi}_t + \frac{1}{NT}\kappa_t\breve{\xi}_t + \frac{1}{N\sqrt{N}}\breve{\varphi}_t + \frac{1}{N\sqrt{T}}\breve{\gamma}_t\right. \\
&+ \left.\tilde{\gamma}_{RP}\left[\frac{1}{T^2\sqrt{N}}\eta_t^2\breve{\xi}_t + \frac{1}{NT^2}\kappa_t\breve{\xi}_t + \frac{1}{NT\sqrt{T}}\breve{\alpha}_t^*\right]\right\} + \frac{1}{N\sqrt{T}}M^{**\prime}\breve{\Pi}_3'\breve{\xi}_t \;+\; \frac{1}{NT\sqrt{N}}M^{**\prime}\breve{\Pi}_1'\breve{\Pi}_1'\breve{\xi}_t
\end{aligned}
\tag{B.51}
$$

vector $r_{7,t}$ is defined as:

$$
\begin{aligned}
r_{7,t} \quad = \quad & = M^{**\prime}(H'W^2H/T)^{-1}\Bigg\{ \left(\frac{H'\varepsilon'W^2\varepsilon\varepsilon W}{NT\sqrt{T}}\right)\left(\frac{W\varepsilon\varepsilon'W^2\Lambda}{N^2T\sqrt{T}}\right)h_t + \frac{1}{TN}\breve{\Pi}_4\breve{\Pi}_2 h_t \\
& + \left(\frac{H'\varepsilon'W^2\varepsilon\varepsilon W}{NT\sqrt{T}}\right)\left(\frac{W\varepsilon\varepsilon'W^2H}{NT\sqrt{T}}\right)\left(\frac{\Lambda'\Lambda}{N}\right)h_t + \frac{1}{TN}\breve{\Pi}_4\breve{\Pi}_1'(\Lambda'\Lambda/N)h_t \\
& + \frac{1}{NT}\sum_{s=1}^{T}\sum_{i=1}^{N}\varepsilon_{i,t}\breve{\varepsilon}_{i,s}\left\{\frac{1}{T^2}\breve{\eta}_s^4\breve{h}_s + \frac{1}{\sqrt{N}T^2}\breve{\kappa}_s\breve{\eta}_s^2\breve{h}_s + \frac{1}{\sqrt{NT}T}\breve{\breve{\alpha}}_s\right. \\
& + \frac{1}{T^2\sqrt{N}}\breve{\kappa}_s\breve{\eta}_s^2\breve{h}_s + \frac{1}{NT^2}\breve{\kappa}_s^2\breve{h}_s + \frac{1}{NT\sqrt{N}}\breve{\breve{\varphi}}_s + \frac{1}{NT\sqrt{T}}\breve{\breve{\gamma}}_s \\
& \left. + \frac{1}{T\sqrt{NT}}\breve{\eta}_s^2\breve{\breve{\alpha}}_s + \frac{1}{NT\sqrt{T}}\breve{\kappa}_s\breve{\breve{\alpha}}_s + \frac{1}{N\sqrt{NT}}\breve{\breve{\delta}}_s + \frac{1}{NT}\breve{\breve{\chi}}_s - \tilde{\gamma}_{RP}\left[\frac{1}{T^2\sqrt{N}}\breve{\beta}_{1,s} + \frac{1}{NT^2\sqrt{T}}\breve{\beta}_{2,s} + \frac{1}{NT^2}\breve{\beta}_{3,s}\right]\right\} \\
& + \frac{1}{T}\breve{\Pi}_4\left\{\frac{1}{\sqrt{NT}}\eta_t^2\breve{\xi}_t + \frac{1}{NT}\kappa_t\breve{\xi}_t + \frac{1}{N\sqrt{N}}\breve{\varphi}_t + \frac{1}{N\sqrt{T}}\breve{\gamma}_t + \tilde{\gamma}_{RP}\left[\frac{1}{T^2\sqrt{N}}\eta_t^2\breve{\xi}_t + \frac{1}{NT^2}\kappa_t\breve{\xi}_t + \frac{1}{NT\sqrt{T}}\breve{\alpha}_t^*\right]\right\} \\
& + \frac{1}{\sqrt{N}}\left(\frac{H'W^2\varepsilon\varepsilon'W}{NT\sqrt{T}}\right)\left(\frac{W\varepsilon\varepsilon'W^2H}{NT\sqrt{T}}\right)\xi_t + \frac{1}{NT\sqrt{T}}\breve{\Pi}_4\breve{\Pi}_1'\xi_t\Bigg\}, \qquad \text{(B.52)}
\end{aligned}
$$

vector $r_{8,t}$ is defined as:

$$
\begin{aligned}
& = M^{***\prime}\breve{\Pi}_1'\left\{\frac{1}{N\sqrt{NT}}\left(\frac{\Lambda'\varepsilon'W^2\varepsilon}{NT}\right)\left(\frac{\varepsilon'W^2\varepsilon\Lambda}{NT}\right) + \frac{1}{N^2T}\breve{\Pi}_1'\breve{\Pi}_2 + \frac{1}{NT}\breve{\Pi}_3'\left(\frac{\Lambda'\Lambda}{N}\right) + \frac{1}{NT\sqrt{NT}}(\breve{\Pi}_1')^2\left(\frac{\Lambda'\Lambda}{N}\right)\right\}h_t \\
& + M^{***\prime}\breve{\Pi}_1'\left\{\frac{1}{\sqrt{NT}}\left[\frac{1}{NT\sqrt{T}}\breve{\beta}_{4,t} + \frac{1}{NT\sqrt{NT}}\breve{\beta}_{5,t} + \frac{1}{N^2\sqrt{T}}\breve{\beta}_{6,t} + \frac{1}{NT\sqrt{N}}\breve{\beta}_{7,t}\right]\right. \\
& + \frac{1}{NT}\breve{\Pi}_1\left\{\frac{1}{\sqrt{NT}}\eta_t^2\breve{\xi}_t + \frac{1}{NT}\kappa_t\breve{\xi}_t + \frac{1}{N\sqrt{N}}\breve{\varphi}_t + \frac{1}{N\sqrt{T}}\breve{\gamma}_t + \tilde{\gamma}_{RP}\left[\frac{1}{T^2\sqrt{N}}\eta_t^2\breve{\xi}_t + \frac{1}{NT^2}\kappa_t\breve{\xi}_t + \frac{1}{NT\sqrt{T}}\breve{\alpha}_t^*\right]\right\} \\
& \left. + \frac{1}{NT\sqrt{N}}\breve{\Pi}_3'\xi_t + \frac{1}{N^2T\sqrt{T}}(\breve{\Pi}_1')^2\xi_t\right\} \\
& + M^{***\prime}\left(\frac{\Lambda'\varepsilon'W^2\varepsilon}{NT}\right)\left\{\frac{1}{N}\left(\frac{\Lambda'\varepsilon'W^2\varepsilon}{NT}\right)\left(\frac{\varepsilon'W^2\varepsilon}{NT}\right) + \frac{1}{N\sqrt{NT}}\breve{\Pi}_2\left(\frac{H'W^2\varepsilon}{\sqrt{NT}}\right)\right. \\
& \left. + \frac{1}{\sqrt{NT}}\left(\frac{\Lambda'\Lambda}{N}\right)\breve{\Pi}_3 + \frac{1}{NT}\left(\frac{\Lambda'\Lambda}{N}\right)\breve{\Pi}_1\left(\frac{\varepsilon'W^2H}{\sqrt{NT}}\right)\right\}h_t \\
& + M^{***\prime}\left(\frac{\Lambda'\varepsilon'W^2\varepsilon}{NT}\right)\Bigg\{ \\
& \frac{1}{\sqrt{NT}}\left(\frac{\varepsilon'W^2H}{\sqrt{NT}}\right)\left\{\frac{1}{\sqrt{NT}}\eta_t^2\breve{\xi}_t + \frac{1}{NT}\kappa_t\breve{\xi}_t + \frac{1}{N\sqrt{N}}\breve{\varphi}_t + \frac{1}{N\sqrt{T}}\breve{\gamma}_t + \tilde{\gamma}_{RP}\left[\frac{1}{T^2\sqrt{N}}\eta_t^2\breve{\xi}_t + \frac{1}{NT^2}\kappa_t\breve{\xi}_t + \frac{1}{NT\sqrt{T}}\breve{\alpha}_t^*\right]\right\} \\
& + \frac{1}{N\sqrt{T}}\left(\frac{\varepsilon'W^2\varepsilon\varepsilon'W^2H}{NT\sqrt{NT}}\right)\xi_t + \frac{1}{NT\sqrt{N}}\left(\frac{\varepsilon'W^2H}{\sqrt{NT}}\right)\breve{\Pi}_1'\xi_t\Bigg\} \\
& + M^{***\prime}\frac{1}{NT}\sum_{s=1}^{T}\sum_{i=1}^{N}\breve{\varepsilon}_{it}\breve{\varepsilon}_{is}\left[\frac{1}{NT\sqrt{T}}\breve{\beta}_{4,s} + \frac{1}{NT\sqrt{NT}}\breve{\beta}_{5,s} + \frac{1}{N^2\sqrt{T}}\breve{\beta}_{6,s} + \frac{1}{NT\sqrt{N}}\breve{\beta}_{7,s}\right]
\end{aligned}
$$

(B.53)

$\hat{\mathcal{R}}_t'$ is the t-th row of matrix $\hat{\mathcal{R}}$ defined by:

$$\hat{\mathcal{R}} = \frac{1}{N^4T^3}(H\Lambda' + \varepsilon)(\varepsilon'W^2\varepsilon + \Lambda H'W^2\varepsilon)^3(\hat{\Lambda}\hat{\mathscr{H}}^{-1} - \Lambda)\left[\breve{S}(I_k + \hat{B})\right]^3\left(\frac{\Lambda'\Lambda}{N}\right)^{-1}. \tag{B.54}$$

The term ξ_t is defined in Assumption B.5, and

$$\begin{aligned}
\breve{S} &= (\Lambda'\Lambda/N)^{-1}(H'W^2H/T)^{-1} \\
\hat{B} &= (H'W^2H/T)(\Lambda'\Lambda/N)\left[(I_k + \hat{A})^{-1} - I_k\right](\Lambda'\Lambda/N)^{-1}(H'W^2H/T)^{-1}, \\
\hat{A} &= (\Lambda'\Lambda/N)^{-1}\Lambda'(\hat{\Lambda}\hat{\mathscr{H}}^{-1} - \Lambda)/N \\
M^* &= \breve{S}(I_k + \hat{B})(\Lambda'\Lambda/N)^{-1}, \qquad M^{**} = \left[\breve{S}(I_k + \hat{B})\right]^2(\Lambda'\Lambda/N)^{-1} \\
M^{***} &= \left[\breve{S}(I_k + \hat{B})\right]^3(\Lambda'\Lambda/N)^{-1} \\
\breve{\aleph}_i &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\breve{\varepsilon}_{i,t}\breve{h}_t \\
\breve{\Pi}_1 &= \frac{1}{\sqrt{NT}}H'W^2\varepsilon\Lambda = \frac{1}{\sqrt{NT}}\breve{H}'\breve{\varepsilon}\Lambda = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\breve{h}_t\breve{\xi}_t', \qquad \text{(B.55)}
\end{aligned}$$

$$\begin{aligned}
\breve{\Pi}_2 &= \frac{1}{NT}\Lambda'\varepsilon'W^2\varepsilon\Lambda = \frac{1}{NT}\Lambda'\breve{\varepsilon}'\breve{\varepsilon}\Lambda = \frac{1}{T}\sum_{t=1}^{T}\breve{\xi}_t\breve{\xi}_t', \\
\breve{\Pi}_3 &= \frac{1}{NT\sqrt{NT}}\breve{H}'\breve{\varepsilon}\breve{\varepsilon}'\breve{\varepsilon}\Lambda = \frac{1}{\sqrt{N}}(\frac{1}{T}\sum_{t=1}^{T}\breve{\alpha}_t\breve{\xi}_t') + \frac{1}{T}(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\breve{\eta}_t^2\breve{h}_t\breve{\xi}_t') + \frac{1}{\sqrt{NT}}(\frac{1}{T}\sum_{t=1}^{T}\breve{h}_t\breve{\xi}_t'\breve{\kappa}_t), \\
\breve{\Pi}_4 &= \frac{1}{NT}H'W^2\varepsilon\varepsilon'W^2H = \frac{1}{NT}\breve{H}'\breve{\varepsilon}\breve{\varepsilon}'\breve{H} = \frac{1}{N}\sum_{i=1}^{N}\breve{\aleph}_i\breve{\aleph}_i', \qquad \text{(B.56)}
\end{aligned}$$

Moreover:

$$
\begin{aligned}
\eta_t^2 &= \operatorname*{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}E[\varepsilon_{i,t}^2|\mathcal{F}_t], \qquad \breve{\eta}_t^2 = \operatorname*{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}E[\breve{\varepsilon}_{i,t}^2|\mathcal{F}_t],\\
\kappa_t &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\varepsilon_{i,t}^2-\eta_t^2)\ , \qquad \breve{\kappa}_t = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\breve{\varepsilon}_{i,t}^2-\breve{\eta}_t^2)\ ,\\
\breve{\alpha}_t &= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{s=1,s\neq t}^{T}\varepsilon_{i,t}\breve{\varepsilon}_{i,s}\breve{h}_s\ , \qquad \breve{\breve{\alpha}}_t = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{s=1,s\neq t}^{T}\breve{\varepsilon}_{i,t}\breve{\varepsilon}_{i,s}\breve{h}_s\ ,\\
\breve{\alpha}_t^* &= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{s=1,s\neq t}^{T}\varepsilon_{i,t}\varepsilon_{i,s}\breve{h}_t\\
\breve{\varphi}_t &= \frac{1}{T}\sum_{i=1}^{N}\sum_{s=1,s\neq t}^{T}\varepsilon_{i,t}E[\breve{\varepsilon}_{i,s}\breve{\xi}_s], \qquad \breve{\gamma}_t = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{s=1,s\neq t}^{T}\varepsilon_{i,t}(\breve{\varepsilon}_{i,s}\breve{\xi}_s - E[\breve{\varepsilon}_{i,s}\breve{\xi}_s]),\\
\breve{\breve{\varphi}}_t &= \frac{1}{T}\sum_{i=1}^{N}\sum_{s=1,s\neq t}^{T}\breve{\varepsilon}_{i,t}E[\breve{\varepsilon}_{i,s}\breve{\xi}_s], \qquad \breve{\breve{\gamma}}_t = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{s=1,s\neq t}^{T}\breve{\varepsilon}_{i,t}(\breve{\varepsilon}_{i,s}\breve{\xi}_s - E[\breve{\varepsilon}_{i,s}\breve{\xi}_s]),\\
\breve{\breve{\alpha}}_t &= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{s=1,s\neq t}^{T}\varepsilon_{i,t}\breve{\varepsilon}_{i,s}\breve{\eta}_s^2\breve{h}_s, \qquad \breve{\breve{\varphi}}_t = \frac{1}{T}\sum_{i=1}^{N}\sum_{s=1,s\neq t}^{T}\varepsilon_{i,t}E[\breve{\varepsilon}_{i,s}\breve{\kappa}_s\breve{h}_s],\\
\breve{\breve{\gamma}}_t &= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{s=1,s\neq t}^{T}\varepsilon_{i,t}(\breve{\varepsilon}_{i,s}\breve{\kappa}_s\breve{h}_s - E[\breve{\varepsilon}_{i,s}\breve{\kappa}_s\breve{h}_s]),\\
\breve{\delta}_t &= \frac{1}{T}\sum_{i=1}^{N}\sum_{s=1,s\neq t}^{T}\varepsilon_{i,t}E[\breve{\varepsilon}_{i,s}\breve{\breve{\alpha}}_s], \qquad \breve{\chi}_t = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{s=1,s\neq t}^{T}\varepsilon_{i,t}(\breve{\varepsilon}_{i,s}\breve{\breve{\alpha}}_s - E[\breve{\varepsilon}_{i,s}\breve{\breve{\alpha}}_s]),
\end{aligned}
\tag{B.57}
$$

and

$$
\begin{aligned}
\breve{\beta}_{1,t} &:= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{s=1}^{T}\breve{h}_t\breve{\eta}_t^2\varepsilon_{i,t}\varepsilon_{i,s}, \qquad \breve{\beta}_{2,t} := \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{s=1}^{T}\breve{h}_t\breve{\kappa}_t^2\varepsilon_{i,t}\varepsilon_{i,s},\\
\breve{\beta}_{3,t} &:= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{s=1}^{T}\breve{\breve{\alpha}}_t\varepsilon_{i,t}\varepsilon_{i,s},\\
\breve{\beta}_{4,t} &:= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{s=1}^{T}\varepsilon_{it}\breve{\varepsilon}_{is}\breve{\eta}_s^2\breve{\xi}_s, \qquad \breve{\beta}_{5,t} := \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{s=1}^{T}\varepsilon_{it}\breve{\varepsilon}_{is}\breve{\kappa}_s\breve{\xi}_s,\\
\breve{\beta}_{6,t} &:= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{s=1}^{T}\varepsilon_{it}\breve{\varepsilon}_{is}\breve{\breve{\varphi}}_s, \qquad \breve{\beta}_{7,t} := \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{s=1}^{T}\varepsilon_{it}\breve{\varepsilon}_{is}\breve{\breve{\gamma}}_s.
\end{aligned}
\tag{B.58}
$$

Moreover, if the eigenvalues of matrix $\Sigma_\lambda^{-1/2}\Sigma_h(W)\Sigma_\lambda^{-1/2}$ for a specific choice of W (that is for a specific choice of $\tilde{\gamma}_{RP}$) in Assumption B.10 are distinct, then, for a suitable ordering and choice of the signs of the factor estimates, we have $\Sigma_\lambda^{-1/2}\tilde{\mathscr{H}} \xrightarrow{p} \mathscr{H}^*$, where the columns of the orthogonal matrix $\mathscr{H}^*$ are the normalized eigenvectors of $\Sigma_\lambda^{-1/2}\Sigma_h(W)\Sigma_\lambda^{-1/2}$.

In equation (B.49), which corresponds to the expansion in (B.2), the difference $(\tilde{\mathscr{H}}')^{-1}\hat{h}_t - h_t$ is written as a sum of a zero-mean term at stochastic order $1/\sqrt{N}$, terms at orders $1/T$, $1/\sqrt{NT}$ and $1/N$, plus remainder terms r_t and $\hat{\mathcal{R}}_t$. Notably, the term of stochastic order $1/\sqrt{N}$ is $u_t/\sqrt{N}$, where $u_t = (\Lambda'\Lambda/N)^{-1}\xi_t$ is zero mean as $E[\xi_t] = 0$ from Assumption B.4 a). This term is the usual first order term appearing in the asymptotic expansion of classical PC factor estimator (see Bai (2003)) and it is also the first order term appearing in the expansion of RP-PC estimator (see Lettau and Pelger (2020a)). The remainder terms r_t and $\hat{\mathcal{R}}_t$ are either scaled by factors that converge to zero faster than $\max\{\frac{1}{T}, \frac{1}{\sqrt{NT}}, \frac{1}{N}\} = O(\frac{1}{m})$, where $m = \min\{N, T\}$.

We now control for the magnitude of the remainder terms r_t and $\hat{\mathcal{R}}_t$ in ϑ_t to show the bounds in Proposition B.2. The next Proposition B.4 provides an upper bound for $T^{-1/2}\|\hat{H}\tilde{\mathscr{H}}^{-1} - H\| = \left(\frac{1}{T}\sum_{t=1}^T \|(\tilde{\mathscr{H}}^{-1})'\hat{h}_t - h_t\|^2\right)^{1/2}$, namely the root MSE of the factor estimates. It is analogous to Lemma A.1 in Bai (2003), but is now derived for RP-PCs instead of only PCs, and yields a sharper upper bound. This result is used to derive a bound on the remainder term $\hat{\mathcal{R}}$, which is also provided in Proposition B.4. Let us define the matrix:

$$\hat{\mathscr{H}} := (\Lambda'\Lambda/N)^{-1}\tilde{\mathscr{H}} = (H'W^2H/T)(\Lambda'\hat{\Lambda}/N)\hat{V}^{-1}. \tag{B.59}$$

PROPOSITION B.4. *Under Assumptions B.10-B.12, we have:*
i) $N^{-1/2}\|\hat{\Lambda}\hat{\mathscr{H}}^{-1} - \Lambda\| = O_p(1/\sqrt{T})$ *and* $N^{-1/2}\|\hat{\Lambda}\tilde{\mathscr{H}}^{-1} - \Lambda\| = O_p(1/\sqrt{T})$. *Moreover,*
ii) $T^{-1/2}\|\hat{H}\tilde{\mathscr{H}}^{-1} - H\| = O_p(1/\sqrt{N})$, *and*
iii)

$$T^{-1/2}\|\hat{\mathcal{R}}\| = O_p\left(\frac{1}{m\sqrt{Tm}}\right). \tag{B.60}$$

From Proposition B.4 and Assumption B.10, we have term $\hat{A}$ defined in (B.55) is such that $\|\hat{A}\| = O_p(\frac{1}{\sqrt{T}})$. By the series representation of the inverse matrix function in a neighborhood of the identity, we deduce that $\|(I_k + \hat{A})^{-1} - I_k\| = O_p(\frac{1}{\sqrt{T}})$. Thus, from Proposition B.4 and Assumption B.10 we get that term $\hat{B}$ appearing in the remainder term r_t in the expansion of Proposition B.3 is such that:

$$\hat{B} = O_p\left(\frac{1}{\sqrt{T}}\right). \tag{B.61}$$

To control for the remainder term r_t we use the next assumption.

Assumption B.13. *We have: (i)* $E[\varepsilon_{i,t}^2|\mathcal{F}_t] \leq M$ *for all* $i \geq 1$ *and* $t \geq 1$*, and a constant* $M > 0$*, (ii)* $\frac{1}{\sqrt{T}}\sum_{t=1}^T h_t\alpha_t' = O_p(1)$*, (iii)* $\frac{1}{\sqrt{T}}\sum_{t=1}^T \eta_t^2 h_t\alpha_t' = O_p(1)$*, (iv)* $\frac{1}{T}\sum_{t=1}^T \xi_t\alpha_t' = o_p(1)$*, and (v)* $E[\|a_t\|^2] = O(1)$*, where* a_t *is any of the following processes:* $\kappa_t h_t$*,* α_t*,* $\kappa_t\xi_t$*,* φ_t*,* γ_t*,* $\bar{\alpha}_t$*,* $\kappa_t^2 h_t$*,* $\bar{\varphi}_t$*,* $\bar{\gamma}_t$*,* $\kappa_t\alpha_t$*,* δ_t*,* χ_t*.*

PROPOSITION B.5. *Under Assumptions B.1 and B.10-B.13, we have:*

$$\frac{1}{T}\sum_{t=1}^T \|\breve{r}_t\|^2 = O_p\left(\frac{1}{N}\right). \tag{B.62}$$

Moreover, $\breve{\vartheta}_t$ *satisfies* $\frac{1}{T}\sum_{t=1}^T \breve{\vartheta}_t h_t' = O_p\left(\frac{1}{N} + \frac{1}{T\sqrt{T}}\right)$ *and* $\frac{1}{T}\sum_{t=1}^T \left(\frac{1}{\sqrt{N}}u_t + \frac{1}{T}\breve{b}_t + \frac{1}{\sqrt{NT}}\breve{d}_t + \breve{\vartheta}_t\right)\breve{\vartheta}_t'$ $= o_p\left(\frac{1}{N\sqrt{T}}\right)$.

Propositions B.3 and B.5 yield Proposition B.2 (with $\mathcal{H} = \mathscr{H}'$ in each group). ■

In the remaining part of this subsection B.5.1 we provide the proofs of Propositions B.3-B.5 and show that Assumptions B.10-B.13 are implied by the Assumptions in Appendix B.1.

B.5.1.1 Proof of Proposition B.3

From equation (B.45) we have $Y'W^2Y = \Lambda H'W^2H\Lambda' + \Lambda H'W^2\varepsilon + \varepsilon'W^2H\Lambda' + \varepsilon'W^2\varepsilon$. By plugging this equation into (B.44), and rearranging the terms, we get:

$$\hat{\Lambda}\hat{V} - \Lambda\left(H'W^2H/T\right)\left(\Lambda'\hat{\Lambda}/N\right) = \frac{1}{NT}(\varepsilon'W^2\varepsilon\hat{\Lambda} + \Lambda H'W^2\varepsilon\hat{\Lambda} + \varepsilon'W^2H\Lambda'\hat{\Lambda}). \tag{B.63}$$

The large sample behaviours of the matrices $\Lambda'\hat{\Lambda}/N$ and $\hat{V}$ are given in the next Lemmas B.10 and B.11, respectively.

LEMMA B.10. *Under Assumptions B.10-B.12, the matrix* $\Lambda'\hat{\Lambda}/N$ *is invertible w.p.a.* 1*, and the inverse is such that* $\|(\Lambda'\hat{\Lambda}/N)^{-1}\| = O_p(1)$*.*

LEMMA B.11. *Under Assumptions B.10-B.12, we have* $\hat{V} \xrightarrow{p} V$*, where* V *is the* (k,k) *diagonal matrix with diagonal elements corresponding to the eigenvalues of matrix* $\Sigma_\lambda\Sigma_h(W)$*. These eigenvalues are the same as those of matrices* $\Sigma_h(W)\Sigma_\lambda$ *and* $\Sigma_\lambda^{1/2}\Sigma_h(W)\Sigma_\lambda^{1/2}$*.*

From Lemma B.11 and Assumption B.10, the matrix $\hat{V}$ is invertible w.p.a. 1. From Assumption B.10 and Lemmas B.10 and B.11, matrix $\hat{\mathscr{H}}$ is invertible w.p.a. 1, and its inverse is:

$$\hat{\mathscr{H}}^{-1} = \hat{V}(\Lambda'\hat{\Lambda}/N)^{-1}(H'W^2H/T)^{-1}. \tag{B.64}$$

From Assumption B.10 and Lemmas B.10 and B.11, matrix $\hat{\mathscr{H}}$ is invertible w.p.a. 1. By post-multiplication of equation (B.63) times the matrix $(\Lambda'\hat{\Lambda}/N)^{-1}(H'W^2H/T)^{-1}$, and using the definition of matrix $\hat{\mathscr{H}}$ given in (B.59), we get:

$$\hat{\Lambda}\hat{\mathscr{H}}^{-1} - \Lambda = \frac{1}{NT}(\varepsilon' W^2\varepsilon + \Lambda H'W^2\varepsilon)\hat{\Lambda}(\Lambda'\hat{\Lambda}/N)^{-1}(H'W^2H/T)^{-1} + \frac{1}{T}\varepsilon' W^2 H(H'W^2H/T)^{-1} \tag{B.65}$$

By using $\hat{\Lambda} = [\hat{\Lambda}\hat{\mathscr{H}}^{-1} - \Lambda]\hat{\mathscr{H}} + \Lambda\hat{\mathscr{H}}$, the last equation can be rewritten as:

$$\begin{aligned}\hat{\Lambda}\hat{\mathscr{H}}^{-1} - \Lambda \;&=\; \frac{1}{NT}(\varepsilon' W^2\varepsilon + \Lambda H'W^2\varepsilon)\Lambda\hat{\mathscr{H}}(\Lambda'\hat{\Lambda}/N)^{-1}(H'W^2H/T)^{-1} + \frac{1}{T}\varepsilon' W^2 H(H'W^2H/T)^{-1} \\ &\quad + \frac{1}{NT}(\varepsilon' W^2\varepsilon + \Lambda H'W^2\varepsilon)[\hat{\Lambda}\hat{\mathscr{H}}^{-1} - \Lambda]\hat{\mathscr{H}}(\Lambda'\hat{\Lambda}/N)^{-1}(H'W^2H/T)^{-1},\end{aligned} \tag{B.66}$$

and the following also holds:

$$\begin{aligned}(\Lambda'\hat{\Lambda}/N)^{-1} \quad &= \left[\frac{\Lambda'\Lambda}{N}\left(I_k + (\Lambda'\Lambda/N)^{-1}\Lambda'(\hat{\Lambda}\hat{\mathscr{H}}^{-1} - \Lambda)/N\right)\hat{\mathscr{H}}\right]^{-1} \\ &= \hat{\mathscr{H}}^{-1}(I_k + \hat{A})^{-1}(\Lambda'\Lambda/N)^{-1}\end{aligned} \tag{B.67}$$

where $\hat{A} = (\Lambda'\Lambda/N)^{-1}\Lambda'(\hat{\Lambda}\hat{\mathscr{H}}^{-1} - \Lambda)/N$. By substituting (B.67) in the first term in the r.h.s. of (B.66), and rearranging terms, we get:

$$\begin{aligned}&\hat{\Lambda}\hat{\mathscr{H}}^{-1} - \Lambda \\ &= \frac{1}{T}\varepsilon' W^2 H(H'W^2H/T)^{-1} + \frac{1}{NT}(\varepsilon' W^2\varepsilon + \Lambda H'W^2\varepsilon)\Lambda(I_k + \hat{A})^{-1}(\Lambda'\Lambda/N)^{-1}(H'W^2H/T)^{-1} \\ &+ \frac{1}{NT}(\varepsilon' W^2\varepsilon + \Lambda H'W^2\varepsilon)\left(\hat{\Lambda}\hat{\mathscr{H}}^{-1} - \Lambda\right)(I_k + \hat{A})^{-1}(\Lambda'\Lambda/N)^{-1}(H'W^2H/T)^{-1} \\ &= \frac{1}{T}\varepsilon' W^2 H(H'W^2H/T)^{-1} + \frac{1}{NT}(\varepsilon' W^2\varepsilon\Lambda + \Lambda H'W^2\varepsilon\Lambda)(\Lambda'\Lambda/N)^{-1}(H'W^2H/T)^{-1}(I_k + \hat{B}) \\ &+ \frac{1}{NT}(\varepsilon' W^2\varepsilon + \Lambda H'W^2\varepsilon)\left(\hat{\Lambda}\hat{\mathscr{H}}^{-1} - \Lambda\right)(\Lambda'\Lambda/N)^{-1}(H'W^2H/T)^{-1}(I_k + \hat{B})\end{aligned} \tag{B.68}$$

where

$$\hat{B} = (H'W^2H/T)(\Lambda'\Lambda/N)\left[(I_k + \hat{A})^{-1} - I_k\right](\Lambda'\Lambda/N)^{-1}(H'W^2H/T)^{-1}, \tag{B.69}$$

or equivalently

$$(\Lambda'\Lambda/N)^{-1}(H'W^2H/T)^{-1}(I_k + \hat{B}) = (I_k + \hat{A})^{-1}(\Lambda'\Lambda/N)^{-1}(H'W^2H/T)^{-1}.$$

Equation (B.68) is a recursive equation for $\hat{\Lambda}\hat{\mathscr{H}}^{-1} - \Lambda$, since this quantity appears also in the third term in the r.h.s. By iterating this equation B.68, we get:

$$
\begin{aligned}
&\hat{\Lambda}\hat{\mathscr{H}}^{-1} - \Lambda \\
&= \frac{1}{T}\varepsilon' W^2 H(H'W^2H/T)^{-1} + \frac{1}{NT}(\varepsilon' W^2 \varepsilon \Lambda + \Lambda H' W^2 \varepsilon \Lambda)(\Lambda'\Lambda/N)^{-1}(H'W^2H/T)^{-1}(I_k + \hat{B}) \\
&+ \frac{1}{NT}(\varepsilon' W^2 \varepsilon + \Lambda H' W^2 \varepsilon)\left\{\frac{1}{T}\varepsilon' W^2 H(H'W^2H/T)^{-1}\right\}(\Lambda'\Lambda/N)^{-1}(H'W^2H/T)^{-1}(I_k + \hat{B}) \\
&+ \frac{1}{NT}(\varepsilon' W^2 \varepsilon + \Lambda H' W^2 \varepsilon)\left\{\frac{1}{NT}(\varepsilon' W^2 \varepsilon \Lambda + \Lambda H' W^2 \varepsilon \Lambda)\right\}\left[(\Lambda'\Lambda/N)^{-1}(H'W^2H/T)^{-1}(I_k + \hat{B})\right]^2 \\
&+ \frac{1}{N^2T^2}(\varepsilon' W^2 \varepsilon + \Lambda H' W^2 \varepsilon)^2\left(\hat{\Lambda}\hat{\mathscr{H}}^{-1} - \Lambda\right)\left[(\Lambda'\Lambda/N)^{-1}(H'W^2H/T)^{-1}(I_k + \hat{B})\right]^2, \qquad \text{(B.70)}
\end{aligned}
$$

and by substituting again (B.68) into (B.70), we get:

$$
\begin{aligned}
&\hat{\Lambda}\hat{\mathscr{H}}^{-1} - \Lambda \\
&= \frac{1}{T}\varepsilon' W^2 H(H'W^2H/T)^{-1} + \frac{1}{NT}(\varepsilon' W^2 \varepsilon \Lambda + \Lambda H' W^2 \varepsilon \Lambda)(\Lambda'\Lambda/N)^{-1}(H'W^2H/T)^{-1}(I_k + \hat{B}) \\
&+ \frac{1}{NT}(\varepsilon' W^2 \varepsilon + \Lambda H' W^2 \varepsilon)\left\{\frac{1}{T}\varepsilon' W^2 H(H'W^2H/T)^{-1}\right\}(\Lambda'\Lambda/N)^{-1}(H'W^2H/T)^{-1}(I_k + \hat{B}) \\
&+ \frac{1}{NT}(\varepsilon' W^2 \varepsilon + \Lambda H' W^2 \varepsilon)\left\{\frac{1}{NT}(\varepsilon' W^2 \varepsilon \Lambda + \Lambda H' W^2 \varepsilon \Lambda)\right\}\left[(\Lambda'\Lambda/N)^{-1}(H'W^2H/T)^{-1}(I_k + \hat{B})\right]^2 \\
&+ \frac{1}{N^2T^2}(\varepsilon' W^2 \varepsilon + \Lambda H' W^2 \varepsilon)^2\left\{\frac{1}{T}\varepsilon' W^2 H(H'W^2H/T)^{-1}\right\}\left[(\Lambda'\Lambda/N)^{-1}(H'W^2H/T)^{-1}(I_k + \hat{B})\right]^2 \\
&+ \frac{1}{N^2T^2}(\varepsilon' W^2 \varepsilon + \Lambda H' W^2 \varepsilon)^2\left\{\frac{1}{NT}(\varepsilon' W^2 \varepsilon \Lambda + \Lambda H' W^2 \varepsilon \Lambda)\right\}\left[(\Lambda'\Lambda/N)^{-1}(H'W^2H/T)^{-1}(I_k + \hat{B})\right]^3 \\
&+ \frac{1}{N^3T^3}(\varepsilon' W^2 \varepsilon + \Lambda H' W^2 \varepsilon)^3\left(\hat{\Lambda}\hat{\mathscr{H}}^{-1} - \Lambda\right)\left[(\Lambda'\Lambda/N)^{-1}(H'W^2H/T)^{-1}(I_k + \hat{B})\right]^3, \qquad \text{(B.71)}
\end{aligned}
$$

We can now obtain the expansion for the RP-PC factor estimator $\hat{H}$. By using the definition of Y from (A.2) and the equality $\hat{\Lambda} = [\hat{\Lambda}\hat{\mathscr{H}}^{-1} - \Lambda]\hat{\mathscr{H}} + \Lambda\hat{\mathscr{H}}$, equation (B.47) can be re-written as:

$$
\hat{H} \;=\; \frac{1}{N}Y\hat{\Lambda} = H\left(\frac{\Lambda'\Lambda}{N}\right)\hat{\mathscr{H}} + \left(\frac{\varepsilon\Lambda}{N}\right)\hat{\mathscr{H}} + \frac{1}{N}(H\Lambda' + \varepsilon)[\hat{\Lambda}\hat{\mathscr{H}}^{-1} - \Lambda]\hat{\mathscr{H}}. \qquad \text{(B.72)}
$$

As matrix $\hat{\mathscr{H}}$ defined in equation (B.59) is invertible w.p.a. 1, then also $\tilde{\mathscr{H}}$ in invertible w.p.a. 1 by Assumption B.3, with its inverse being:

$$
\tilde{\mathscr{H}}^{-1} \;=\; \hat{\mathscr{H}}^{-1}(\Lambda'\Lambda/N)^{-1} \;=\; \hat{V}(\Lambda'\hat{\Lambda}/N)^{-1}(H'W^2H/T)^{-1}(\Lambda'\Lambda/N)^{-1}. \qquad \text{(B.73)}
$$

By post-multiplication of equation (B.72) times the matrix $\tilde{\mathscr{H}}^{-1}$ we get:

$$\hat{H}\tilde{\mathscr{H}}^{-1} - H = \frac{1}{N}\varepsilon\Lambda(\Lambda'\Lambda/N)^{-1} + \frac{1}{N}(H\Lambda' + \varepsilon)[\hat{\Lambda}\hat{\mathscr{H}}^{-1} - \Lambda](\Lambda'\Lambda/N)^{-1}. \tag{B.74}$$

By plugging in equation (B.71) into (B.74), using the definition of $\breve{S}$ and re-arranging terms we get:

$$\begin{aligned}
&\hat{H}\tilde{\mathscr{H}}^{-1} - H = \frac{1}{N}\varepsilon\Lambda(\Lambda'\Lambda/N)^{-1} \\
&+\left(\frac{1}{NT}H\Lambda'\varepsilon'W^2H\right)\breve{S}(I_k+\hat{B})(\Lambda'\Lambda/N)^{-1} + \left(\frac{1}{NT}\varepsilon\varepsilon'W^2H\right)\breve{S}(I_k+\hat{B})(\Lambda'\Lambda/N)^{-1} \\
&+\frac{1}{N}(H\Lambda'+\varepsilon)\left[\frac{1}{NT}(\varepsilon'W^2\varepsilon\Lambda + \Lambda H'W^2\varepsilon\Lambda)\right]\breve{S}(I_k+\hat{B})(\Lambda'\Lambda/N)^{-1} \\
&+\frac{1}{N}(H\Lambda'+\varepsilon)\cdot\frac{1}{NT}(\varepsilon'W^2\varepsilon + \Lambda H'W^2\varepsilon)\left\{\frac{1}{T}\varepsilon'W^2H(H'W^2H/T)^{-1}\right\}\breve{S}(I_k+\hat{B})(\Lambda'\Lambda/N)^{-1} \\
&+\frac{1}{N}(H\Lambda'+\varepsilon)\cdot\frac{1}{NT}(\varepsilon'W^2\varepsilon + \Lambda H'W^2\varepsilon)\left\{\frac{1}{NT}(\varepsilon'W^2\varepsilon\Lambda + \Lambda H'W^2\varepsilon\Lambda)\right\}\left[\breve{S}(I_k+\hat{B})\right]^2(\Lambda'\Lambda/N)^{-1} \\
&+\frac{1}{N}(H\Lambda'+\varepsilon)\frac{1}{N^2T^2}(\varepsilon'W^2\varepsilon + \Lambda H'W^2\varepsilon)^2\left\{\frac{1}{T}\varepsilon'W^2H(H'W^2H/T)^{-1}\right\}\left[\breve{S}(I_k+\hat{B})\right]^2(\Lambda'\Lambda/N)^{-1} \\
&+\frac{1}{N}(H\Lambda'+\varepsilon)\frac{1}{N^2T^2}(\varepsilon'W^2\varepsilon + \Lambda H'W^2\varepsilon)^2\left\{\frac{1}{NT}(\varepsilon'W^2\varepsilon\Lambda + \Lambda H'W^2\varepsilon\Lambda)\right\}\left[\breve{S}(I_k+\hat{B})\right]^3(\Lambda'\Lambda/N)^{-1} \\
&+\frac{1}{N}(H\Lambda'+\varepsilon)\frac{1}{N^3T^3}(\varepsilon'W^2\varepsilon + \Lambda H'W^2\varepsilon)^3\left(\hat{\Lambda}\hat{\mathscr{H}}^{-1} - \Lambda\right)\left[(\Lambda'\Lambda/N)^{-1}(H'W^2H/T)^{-1}(I_k+\hat{B})\right]^3(\Lambda'\Lambda/N)^{-1},
\end{aligned} \tag{B.75}$$

By using the definitions of M^*, M^{**}, M^{***} and $\hat{\mathcal{R}}$ provided in Proposition B.3 the last equation can be expressed as:

$$\begin{aligned}
\hat{H}\tilde{\mathscr{H}}^{-1} - H = {} & \frac{1}{N}\varepsilon\Lambda(\Lambda'\Lambda/N)^{-1} + \left(\frac{1}{NT}H\Lambda'\varepsilon'W^2H\right)M^* + \left(\frac{1}{NT}\varepsilon\varepsilon'W^2H\right)M^* \\
&+\frac{1}{N}(H\Lambda'+\varepsilon)\cdot\left[\frac{1}{NT}(\varepsilon'W^2\varepsilon\Lambda + \Lambda H'W^2\varepsilon\Lambda)\right]M^* \\
&+\frac{1}{N}(H\Lambda'+\varepsilon)\frac{1}{NT}(\varepsilon'W^2\varepsilon + \Lambda H'W^2\varepsilon)\left\{\frac{1}{T}\varepsilon'W^2H(H'W^2H/T)^{-1}\right\}M^* \\
&+\frac{1}{N}(H\Lambda'+\varepsilon)\cdot\frac{1}{NT}(\varepsilon'W^2\varepsilon + \Lambda H'W^2\varepsilon)\left\{\frac{1}{NT}(\varepsilon'W^2\varepsilon\Lambda + \Lambda H'W^2\varepsilon\Lambda)\right\}M^{**} \\
&+\frac{1}{N}(H\Lambda'+\varepsilon)\frac{1}{N^2T^2}(\varepsilon'W^2\varepsilon + \Lambda H'W^2\varepsilon)^2\left\{\frac{1}{T}\varepsilon'W^2H(H'W^2H/T)^{-1}\right\}M^{**} \\
&+\frac{1}{N}(H\Lambda'+\varepsilon)\frac{1}{N^2T^2}(\varepsilon'W^2\varepsilon + \Lambda H'W^2\varepsilon)^2\left\{\frac{1}{NT}(\varepsilon'W^2\varepsilon\Lambda + \Lambda H'W^2\varepsilon\Lambda)\right\}M^{***} \\
&+\hat{\mathcal{R}},
\end{aligned} \tag{B.76}$$

We now study each of the terms in the r.h.s. of the last equation in order to write expansion of

$\hat{H}\tilde{\mathscr{H}}^{-1} - H$ for each date t. The first term in the r.h.s. of equation (B.76) is such that:

$$\left[\frac{1}{N}\varepsilon\Lambda(\Lambda'\Lambda/N)^{-1}\right]_t \;=\; \frac{1}{\sqrt{N}}(\Lambda'\Lambda/N)^{-1}\xi_t \;=\; \frac{1}{\sqrt{N}}u_t. \tag{B.77}$$

The second term in the r.h.s. of equation (B.76) can be written as:

$$\frac{1}{NT}(H\Lambda'\varepsilon'W^2H)M^* \;=\; \frac{1}{\sqrt{NT}}H\left(\frac{1}{\sqrt{NT}}\Lambda'\varepsilon'W^2H\right)M^* \;=\; \frac{1}{\sqrt{NT}}H\breve{\Pi}_1'M^*$$

where $\breve{\Pi}_1$ is defined in Proposition B.3. Therefore:

$$\left[\frac{1}{NT}(H\Lambda'\varepsilon'W^2H)M^*\right]_t \;=\; M^{*\prime}\frac{1}{\sqrt{NT}}\breve{\Pi}_1 h_t \tag{B.78}$$

The third term in the r.h.s. of equation (B.76) can be written as $\left(\frac{1}{NT}\varepsilon\varepsilon'W^2H\right)M^*$. Noting that:

$$\begin{aligned}
\frac{1}{NT}[\varepsilon\varepsilon'W^2H]_t \;&=\; \frac{1}{NT}\sum_{i=1}^{N}\sum_{s=1}^{T}\varepsilon_{i,t}\breve{\varepsilon}_{i,s}\breve{h}_s = \frac{1}{NT}\sum_{i=1}^{N}\varepsilon_{i,t}\breve{\varepsilon}_{i,t}\breve{h}_t + \frac{1}{NT}\sum_{i=1}^{N}\sum_{s=1,s\neq t}^{T}\varepsilon_{i,t}\breve{\varepsilon}_{i,s}\breve{h}_s \\
&=\; \frac{1}{NT}\sum_{i=1}^{N}\varepsilon_{i,t}^2\breve{h}_t + \tilde{\gamma}_{RP}\frac{1}{NT}\sum_{i=1}^{N}\varepsilon_{i,t}\bar{\varepsilon}_{i,\cdot}\breve{h}_t + \frac{1}{NT}\sum_{i=1}^{N}\sum_{s=1,s\neq t}^{T}\varepsilon_{i,t}\breve{\varepsilon}_{i,s}\breve{h}_s \\
&=\; \frac{1}{NT}\sum_{i=1}^{N}\varepsilon_{i,t}^2\breve{h}_t + \tilde{\gamma}_{RP}\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{s=1}^{T}\varepsilon_{i,t}\varepsilon_{i,s}\breve{h}_t + \frac{1}{NT}\sum_{i=1}^{N}\sum_{s=1,s\neq t}^{T}\varepsilon_{i,t}\breve{\varepsilon}_{i,s}\breve{h}_s \\
&=\; \frac{1}{NT}\sum_{i=1}^{N}\varepsilon_{i,t}^2\breve{h}_t + \tilde{\gamma}_{RP}\frac{1}{NT^2}\sum_{i=1}^{N}\varepsilon_{i,t}^2\breve{h}_t + \tilde{\gamma}_{RP}\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{s=1,s\neq t}^{T}\varepsilon_{i,t}\varepsilon_{i,s}\breve{h}_t + \frac{1}{NT}\sum_{i=1}^{N}\sum_{s=1,s\neq t}^{T}\varepsilon_{i,t}\breve{\varepsilon}_{i,s}\breve{h}_s \\
&=\; \left(1+\frac{1}{T}\tilde{\gamma}_{RP}\right)\frac{1}{NT}\sum_{i=1}^{N}\varepsilon_{i,t}^2\breve{h}_t + \tilde{\gamma}_{RP}\frac{1}{T\sqrt{NT}}\left(\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{s=1,s\neq t}^{T}\varepsilon_{i,t}\breve{h}_t\varepsilon_{i,s}\right) + \frac{1}{NT}\sum_{i=1}^{N}\sum_{s=1,s\neq t}^{T}\varepsilon_{i,t}\breve{\varepsilon}_{i,s}\breve{h}_s \\
&=\; \left(1+\frac{1}{T}\tilde{\gamma}_{RP}\right)\left[\frac{1}{T}\left(\frac{1}{N}\sum_{i=1}^{N}E[\varepsilon_{i,t}^2|\mathcal{F}_t]\breve{h}_t\right) + \frac{1}{T\sqrt{N}}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\{\varepsilon_{i,t}^2 - E[\varepsilon_{i,t}^2|\mathcal{F}_t]\}\breve{h}_t\right)\right] \\
&\quad +\tilde{\gamma}_{RP}\frac{1}{T\sqrt{NT}}\left(\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{s=1,s\neq t}^{T}\varepsilon_{i,t}\breve{h}_t\varepsilon_{i,s}\right) \;+\; \frac{1}{\sqrt{NT}}\left(\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{s=1,s\neq t}^{T}\varepsilon_{i,t}\breve{\varepsilon}_{i,s}\breve{h}_s\right) \\
&=\; \left(1+\frac{1}{T}\tilde{\gamma}_{RP}\right)\left[\frac{1}{T}\eta_t^2\breve{h}_t + \frac{1}{T\sqrt{N}}\kappa_t\breve{h}_t\right] + \frac{1}{\sqrt{NT}}\breve{\alpha}_t + \tilde{\gamma}_{RP}\frac{1}{T\sqrt{NT}}\breve{\alpha}_t^* \\
&=\; \frac{1}{T}\eta_t^2\breve{h}_t + \frac{1}{T\sqrt{N}}\kappa_t\breve{h}_t + \frac{1}{\sqrt{NT}}\breve{\alpha}_t + \tilde{\gamma}_{RP}\left[\frac{1}{T^2}\eta_t^2\breve{h}_t + \frac{1}{T\sqrt{NT}}\breve{\alpha}_t^* + \frac{1}{T^3\sqrt{N}}\kappa_t\breve{h}_t\right],
\end{aligned} \tag{B.79}$$

where η_t^2, κ_t, $\breve{\alpha}_t$ and $\breve{\alpha}_t^*$ are defined in Proposition B.3. Therefore:

$$\left[\frac{1}{NT}(\varepsilon\varepsilon' W^2 H)M^*\right]_t = \frac{1}{T}M^{*\prime}\eta_t^2\breve{h}_t + \frac{1}{T\sqrt{N}}M^{*\prime}\kappa_t\breve{h}_t + \frac{1}{\sqrt{NT}}M^{*\prime}\breve{\alpha}_t + \tilde{\gamma}_{RP}M^{*\prime}\left[\frac{1}{T^2}\eta_t^2\breve{h}_t + \frac{1}{T\sqrt{NT}}\breve{\alpha}_t^* + \frac{1}{T^3\sqrt{N}}\kappa_t\breve{h}_t\right] \quad \text{(B.80)}$$

The fourth term in the r.h.s. of equation (B.76) can be written as:

$$\frac{1}{N^2T}(H\Lambda' + \varepsilon)\left[\varepsilon' W^2\varepsilon + \Lambda H' W^2\varepsilon\right]\Lambda M^*$$
$$= \frac{1}{N^2T}(H\Lambda'\varepsilon' W^2\varepsilon\Lambda)M^* + \frac{1}{N^2T}(H\Lambda'\Lambda H' W^2\varepsilon\Lambda)M^* + \frac{1}{N^2T}(\varepsilon\varepsilon' W^2\varepsilon\Lambda)M^* + \frac{1}{N^2T}(\varepsilon\Lambda H' W^2\varepsilon\Lambda)M^* \quad \text{(B.81)}$$

The date t - element of the first term in the r.h.s. of equation (B.81) can be computed from:

$$\left[\frac{1}{N^2T}(H\Lambda'\varepsilon' W^2\varepsilon\Lambda)M^*\right]_t = M^{*\prime}\frac{1}{N}\left(\frac{1}{NT}\Lambda'\varepsilon' W^2\varepsilon\Lambda\right)h_t = M^{*\prime}\frac{1}{N}\breve{\Pi}_2 h_t\,, \quad \text{(B.82)}$$

where $\breve{\Pi}_2$ is defined in Proposition (B.3). The date t - element of the second term in the r.h.s. of equation (B.81) can be computed as:

$$\left[\frac{1}{N^2T}(H\Lambda'\Lambda H' W^2\varepsilon\Lambda)M^*\right]_t = \frac{1}{\sqrt{NT}}M^{*\prime}\left(\frac{1}{\sqrt{NT}}\Lambda'\varepsilon' W^2 H\right)\left(\frac{\Lambda'\Lambda}{N}\right)h_t = \frac{1}{\sqrt{NT}}M^{*\prime}\breve{\Pi}_1'\left(\frac{\Lambda'\Lambda}{N}\right)h_t. \quad \text{(B.83)}$$

The date t - element of the third term in the r.h.s. of equation (B.81) can be computed from:

$$\begin{aligned}
\left[\frac{1}{N^2T}\varepsilon\varepsilon' W^2\varepsilon\Lambda\right]_t &= \frac{1}{N^2T}\sum_{i=1}^{N}\sum_{s=1}^{T}\sum_{\ell=1}^{N}\varepsilon_{i,t}\breve{\varepsilon}_{i,s}\breve{\varepsilon}_{\ell,s}\lambda_\ell = \frac{1}{N\sqrt{NT}}\sum_{i=1}^{N}\sum_{s=1}^{T}\varepsilon_{i,t}\breve{\varepsilon}_{i,s}\breve{\xi}_s \\
&= \frac{1}{N\sqrt{NT}}\sum_{i=1}^{N}\varepsilon_{i,t}\breve{\varepsilon}_{i,t}\breve{\xi}_t + \left[\frac{1}{N\sqrt{NT}}\sum_{i=1}^{N}\sum_{s=1,s\neq t}^{T}\varepsilon_{i,t}\breve{\varepsilon}_{i,s}\breve{\xi}_s\right] \\
&= \frac{1}{N\sqrt{NT}}\sum_{i=1}^{N}\varepsilon_{i,t}^2\breve{\xi}_t + \tilde{\gamma}_{RP}\frac{1}{N\sqrt{NT}}\sum_{i=1}^{N}\varepsilon_{i,t}\bar{\varepsilon}_{i,\cdot}\breve{\xi}_t \\
&+ \left[\frac{1}{N\sqrt{N}}\left(\frac{1}{T}\sum_{i=1}^{N}\sum_{s=1,s\neq t}^{T}\varepsilon_{i,t}E[\breve{\varepsilon}_{i,s}\breve{\xi}_s]\right) + \frac{1}{N\sqrt{T}}\left\{\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{s=1,s\neq t}^{T}\varepsilon_{i,t}(\breve{\varepsilon}_{i,s}\breve{\xi}_s - E[\breve{\varepsilon}_{i,s}\breve{\xi}_s])\right\}\right] \\
&= \frac{1}{N\sqrt{NT}}\sum_{i=1}^{N}\varepsilon_{i,t}^2\breve{\xi}_t + \tilde{\gamma}_{RP}\frac{1}{N\sqrt{NT^2}}\sum_{i=1}^{N}\sum_{s=1}^{T}\varepsilon_{i,t}\varepsilon_{i,s}\breve{\xi}_t + \left[\frac{1}{N\sqrt{N}}\breve{\varphi}_t + \frac{1}{N\sqrt{T}}\breve{\gamma}_t\right] \\
&= \left(1+\frac{1}{T}\tilde{\gamma}_{RP}\right)\frac{1}{N\sqrt{NT}}\sum_{i=1}^{N}\varepsilon_{i,t}^2\breve{\xi}_t + \tilde{\gamma}_{RP}\frac{1}{N\sqrt{NT^2}}\sum_{i=1}^{N}\sum_{s=1,s\neq t}^{T}\varepsilon_{i,t}\varepsilon_{i,s}\breve{\xi}_t + \left[\frac{1}{N\sqrt{N}}\breve{\varphi}_t + \frac{1}{N\sqrt{T}}\breve{\gamma}_t\right] \\
&= \left(1+\frac{1}{T}\tilde{\gamma}_{RP}\right)\frac{1}{\sqrt{NT}}\left\{\frac{1}{N}\sum_{i=1}^{N}E[\varepsilon_{i,t}^2]\breve{\xi}_t\right\} + \left(1+\frac{1}{T}\tilde{\gamma}_{RP}\right)\frac{1}{NT}\cdot\left\{\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\varepsilon_{i,t}^2\breve{\xi}_t - E[\varepsilon_{i,t}^2]\breve{\xi}_t)\right\} \\
&+ \tilde{\gamma}_{RP}\frac{1}{NT\sqrt{T}}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{s=1,s\neq t}^{T}\varepsilon_{i,t}\breve{\xi}_t\varepsilon_{i,s} + \left[\frac{1}{N\sqrt{N}}\breve{\varphi}_t + \frac{1}{N\sqrt{T}}\breve{\gamma}_t\right] \\
&= \left(1+\frac{1}{T}\tilde{\gamma}_{RP}\right)\left[\frac{1}{\sqrt{NT}}\eta_t^2\breve{\xi}_t + \frac{1}{NT}\kappa_t\breve{\xi}_t\right] + \frac{1}{N\sqrt{N}}\breve{\varphi}_t + \frac{1}{N\sqrt{T}}\breve{\gamma}_t + \tilde{\gamma}_{RP}\frac{1}{NT\sqrt{T}}\breve{\alpha}_t^* \\
&= \frac{1}{\sqrt{NT}}\eta_t^2\breve{\xi}_t + \frac{1}{NT}\kappa_t\breve{\xi}_t + \frac{1}{N\sqrt{N}}\breve{\varphi}_t + \frac{1}{N\sqrt{T}}\breve{\gamma}_t + \tilde{\gamma}_{RP}\left[\frac{1}{T^2\sqrt{N}}\eta_t^2\breve{\xi}_t + \frac{1}{NT^2}\kappa_t\breve{\xi}_t + \frac{1}{NT\sqrt{T}}\breve{\alpha}_t^*\right]
\end{aligned} \tag{B.84}$$

For the last equation to hold we need a condition such that:

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{s=1,s\neq t}^{T}\varepsilon_{i,t}\varepsilon_{i,s}\breve{\xi}_t = O_p(1)\ .$$

It is enough to add $\breve{\alpha}_t^*$ in the list of terms in Assumption B.13 v).

which implies:

$$\begin{aligned}
&\left[\frac{1}{N^2T}\varepsilon\varepsilon' W^2\varepsilon\Lambda M^*\right]_t \\
&= M^{*\prime}\left\{\frac{1}{\sqrt{NT}}\eta_t^2\breve{\xi}_t + \frac{1}{NT}\kappa_t\breve{\xi}_t + \frac{1}{N\sqrt{N}}\breve{\varphi}_t + \frac{1}{N\sqrt{T}}\breve{\gamma}_t + \tilde{\gamma}_{RP}\left[\frac{1}{T^2\sqrt{N}}\eta_t^2\breve{\xi}_t + \frac{1}{NT^2}\kappa_t\breve{\xi}_t + \frac{1}{NT\sqrt{T}}\breve{\alpha}_t^*\right]\right\}
\end{aligned} \tag{B.85}$$

The date t - element of the fourth term in the r.h.s. of equation (B.81) can be computed by noting that:

$$\frac{1}{N^2T}(\varepsilon\Lambda H'W^2\varepsilon\Lambda)M^* = \frac{1}{N\sqrt{T}}\left(\frac{\varepsilon\Lambda}{\sqrt{N}}\right)\left(\frac{1}{\sqrt{NT}}H'W^2\varepsilon\Lambda\right)M^* = \frac{1}{N\sqrt{T}}\left(\frac{\varepsilon\Lambda}{\sqrt{N}}\right)\breve{\Pi}_1 M^*$$

which implies:

$$\left[\frac{1}{N^2T}(\varepsilon\Lambda H'W^2\varepsilon\Lambda)M^*\right]_t = \frac{1}{N\sqrt{T}}\,M^{*\prime}\breve{\Pi}_1'\,\xi_t. \tag{B.86}$$

The fifth term in the r.h.s. of equation (B.76) can be written as:

$$\begin{aligned}
&\frac{1}{N^2T^2}(H\Lambda'+\varepsilon)\left[\varepsilon'W^2\varepsilon\varepsilon'W^2H+\Lambda H'W^2\varepsilon\varepsilon'W^2H\right](H'W^2H/T)^{-1}M^*\\
&=\left[\frac{1}{N^2T^2}H\Lambda'\varepsilon'W^2\varepsilon\varepsilon'W^2H+\frac{1}{N^2T^2}H\Lambda'\Lambda H'W^2\varepsilon\varepsilon'W^2H+\frac{1}{N^2T^2}\varepsilon\varepsilon'W^2\varepsilon\varepsilon'W^2H\right.\\
&\left.+\frac{1}{N^2T^2}\varepsilon\Lambda H'W^2\varepsilon\varepsilon'W^2H\right](H'W^2H/T)^{-1}M^*\,.
\end{aligned} \tag{B.87}$$

The date t - element of the first term in the r.h.s. of equation (B.87) can be computed by first noting that:

$$\frac{1}{N^2T^2}H\Lambda'\varepsilon'W^2\varepsilon\varepsilon'W^2H(H'W^2H/T)^{-1}M^* = \frac{1}{\sqrt{NT}}H\left(\frac{1}{NT\sqrt{NT}}\Lambda'\varepsilon'W^2\varepsilon\varepsilon'W^2H\right)(H'W^2H/T)^{-1}M^*,$$

and that the term in the brackets in the r.h.s. of the last equation is:

$$\begin{aligned}
\frac{1}{NT\sqrt{NT}}\Lambda'\varepsilon'W^2\varepsilon\varepsilon'W^2H &= \frac{1}{NT\sqrt{NT}}\sum_{i=1}^N\sum_{t=1}^T\sum_{\ell=1}^N\sum_{s=1}^T\lambda_i\breve{\varepsilon}_{i,t}\breve{\varepsilon}_{\ell,t}\breve{\varepsilon}_{\ell,s}\breve{h}_s'\\
&= \frac{1}{NT\sqrt{T}}\sum_{t=1}^T\sum_{\ell=1}^N\sum_{s=1}^T\breve{\xi}_t\breve{\varepsilon}_{\ell,t}\breve{\varepsilon}_{\ell,s}\breve{h}_s'\\
&= \frac{1}{NT\sqrt{T}}\sum_{t=1}^T\sum_{\ell=1}^N\sum_{s=1,s\neq t}^T\breve{\xi}_t\breve{\varepsilon}_{\ell,t}\breve{\varepsilon}_{\ell,s}\breve{h}_s'+\frac{1}{NT\sqrt{T}}\sum_{t=1}^T\sum_{\ell=1}^N\breve{\xi}_t\breve{\varepsilon}_{\ell,t}^2\breve{h}_t'\\
&= \frac{1}{\sqrt{N}}\left(\frac{1}{T}\sum_{t=1}^T\breve{\xi}_t\breve{\alpha}_t'\right)+\frac{1}{T}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^T\breve{\xi}_t\breve{\eta}_t^2\breve{h}_t'\right)+\frac{1}{\sqrt{NT}}\left(\frac{1}{T}\sum_{t=1}^T\breve{\xi}_t\breve{\kappa}_t\breve{h}_t'\right)=\breve{\Pi}_3',
\end{aligned}$$

which implies:

$$\left[\frac{1}{N^2T^2}H(\Lambda'\varepsilon'W^2\varepsilon\varepsilon'W^2H)(H'W^2H/T)^{-1}M^*\right]_t = \frac{1}{\sqrt{NT}}M^{*\prime}(H'W^2H/T)^{-1}\breve{\Pi}_3 h_t\,. \tag{B.88}$$

The date t - element of the second term in the r.h.s. of equation (B.87) can be computed by first noting that:

$$\begin{aligned}
\frac{1}{N^2T^2}H\Lambda'\Lambda H'W^2\varepsilon\varepsilon'W^2H &= \frac{1}{T}H\left(\frac{\Lambda'\Lambda}{N}\right)\frac{1}{NT}H'W^2\varepsilon\varepsilon'W^2H = \frac{1}{T}H\left(\frac{\Lambda'\Lambda}{N}\right)\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\breve{h}_{i,t}\breve{\varepsilon}_{i,t}\breve{\varepsilon}_{i,s}\breve{h}'_{i,s} \\
&= \frac{1}{T}H\left(\frac{\Lambda'\Lambda}{N}\right)\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\breve{h}_{i,t}\breve{\varepsilon}_{i,t}\right)\left(\frac{1}{\sqrt{T}}\sum_{s=1}^{T}\breve{\varepsilon}_{i,s}\breve{h}'_{i,s}\right) \\
&= \frac{1}{T}H\left(\frac{\Lambda'\Lambda}{N}\right)\frac{1}{N}\sum_{i=1}^{N}\breve{\aleph}_i\breve{\aleph}'_i \;=\; \frac{1}{T}H\left(\frac{\Lambda'\Lambda}{N}\right)\breve{\Pi}_4
\end{aligned}$$

which implies:

$$\left[\frac{1}{N^2T^2}H\Lambda'\Lambda H'W^2\varepsilon\varepsilon'W^2H(H'W^2H/T)^{-1}M^*\right]_t = \frac{1}{T}M^{*\prime}(H'W^2H/T)^{-1}\breve{\Pi}_4\left(\frac{\Lambda'\Lambda}{N}\right)\breve{h}_t \quad \text{(B.89)}$$

To compute the date t - element of the third term in the r.h.s. of equation (B.87), that is $\frac{1}{N^2T^2}\varepsilon\varepsilon'W^2\varepsilon\varepsilon'W^2HM^*$, we need an expression for the term $\frac{1}{NT}[W\varepsilon\varepsilon'W^2H]_t$. Using the definitions of terms $\breve{\eta}_t$, $\breve{\kappa}_t$, and $\breve{\alpha}_t$ in Proposition B.3, we get:

$$\begin{aligned}
\frac{1}{NT}[W\varepsilon\varepsilon'W^2H]_t &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{s=1}^{T}\breve{\varepsilon}_{i,t}\breve{\varepsilon}_{i,s}\breve{h}_s = \frac{1}{NT}\sum_{i=1}^{N}\breve{\varepsilon}_{i,t}\breve{\varepsilon}_{i,t}\breve{h}_t + \frac{1}{NT}\sum_{i=1}^{N}\sum_{s=1,s\neq t}^{T}\breve{\varepsilon}_{i,t}\breve{\varepsilon}_{i,s}\breve{h}_s \\
&= \frac{1}{T}\left(\frac{1}{N}\sum_{i=1}^{N}E[\breve{\varepsilon}^2_{i,t}|\mathcal{F}_t]\breve{h}_t\right) + \frac{1}{T\sqrt{N}}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\{\breve{\varepsilon}^2_{i,t} - E[\breve{\varepsilon}^2_{i,t}|\mathcal{F}_t]\}\breve{h}_t\right) + \frac{1}{NT}\sum_{i=1}^{N}\sum_{s=1,s\neq t}^{T}\breve{\varepsilon}_{i,t}\breve{\varepsilon}_{i,s}\breve{h}_s \\
&= \frac{1}{T}\breve{\eta}^2_t\breve{h}_t + \frac{1}{T\sqrt{N}}\breve{\kappa}_t\breve{h}_t + \frac{1}{\sqrt{NT}}\breve{\alpha}_t. \qquad \text{(B.90)}
\end{aligned}$$

The date t - element of the third term in the r.h.s. of equation (B.87) is:

$$
\begin{aligned}
&\frac{1}{N^2T^2}[\varepsilon\varepsilon' W^2 \varepsilon\varepsilon' W^2 H M^*]_t \\
&= M^{*\prime}\frac{1}{N^2T^2}\sum_{s=1}^{T}\sum_{i=1}^{N}\varepsilon_{i,t}\breve{\varepsilon}_{i,s}[W\varepsilon\varepsilon' W^2 H]_s = M^{*\prime}\frac{1}{NT}\sum_{s=1}^{T}\sum_{i=1}^{N}\varepsilon_{i,t}\breve{\varepsilon}_{i,s}\left(\frac{1}{T}\breve{\eta}_s^2\breve{h}_s + \frac{1}{T\sqrt{N}}\breve{\kappa}_s\breve{h}_s + \frac{1}{\sqrt{NT}}\breve{\breve{\alpha}}_s\right) \\
&= M^{*\prime}\Bigg\{\frac{1}{NT^2}\sum_{i=1}^{N}\varepsilon_{i,t}\breve{\varepsilon}_{i,t}\breve{\eta}_t^2\breve{h}_t + \frac{1}{NT^2}\sum_{s=1,s\neq t}^{T}\sum_{i=1}^{N}\varepsilon_{i,t}\breve{\varepsilon}_{i,s}\breve{\eta}_s^2\breve{h}_s \\
&+\frac{1}{NT^2\sqrt{N}}\sum_{i=1}^{N}\varepsilon_{i,t}\breve{\varepsilon}_{i,t}\breve{\kappa}_t\breve{h}_t + \frac{1}{NT^2\sqrt{N}}\sum_{s=1,s\neq t}^{T}\sum_{i=1}^{N}\varepsilon_{i,t}\breve{\varepsilon}_{i,s}\breve{\kappa}_s\breve{h}_s \\
&+\frac{1}{NT\sqrt{NT}}\sum_{i=1}^{N}\varepsilon_{i,t}\breve{\varepsilon}_{i,t}\breve{\breve{\alpha}}_t + \frac{1}{NT\sqrt{NT}}\sum_{s=1,s\neq t}^{T}\sum_{i=1}^{N}\varepsilon_{i,t}\breve{\varepsilon}_{i,s}\breve{\breve{\alpha}}_s\Bigg\} \\
&= M^{*\prime}\Bigg\{\frac{1}{NT^2}\sum_{i=1}^{N}\breve{\varepsilon}_{i,t}\breve{\varepsilon}_{i,t}\breve{\eta}_t^2\breve{h}_t + \frac{1}{NT^2}\sum_{s=1,s\neq t}^{T}\sum_{i=1}^{N}\varepsilon_{i,t}\breve{\varepsilon}_{i,s}\breve{\eta}_s^2\breve{h}_s - \tilde{\gamma}_{RP}\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{s=1}^{T}\varepsilon_{i,s}\varepsilon_{i,t}\breve{\eta}_t^2\breve{h}_t \\
&+\frac{1}{NT^2\sqrt{N}}\sum_{i=1}^{N}\varepsilon_{i,t}\breve{\varepsilon}_{i,t}\breve{\kappa}_t\breve{h}_t + \frac{1}{NT^2\sqrt{N}}\sum_{s=1,s\neq t}^{T}\sum_{i=1}^{N}\varepsilon_{i,t}\breve{\varepsilon}_{i,s}\breve{\kappa}_s\breve{h}_s - \tilde{\gamma}_{RP}\frac{1}{NT^3\sqrt{N}}\sum_{i=1}^{N}\sum_{s=1}^{T}\varepsilon_{i,s}\varepsilon_{i,t}\breve{\kappa}_t\breve{h}_t \\
&+\frac{1}{NT\sqrt{NT}}\sum_{i=1}^{N}\varepsilon_{i,t}\breve{\varepsilon}_{i,t}\breve{\breve{\alpha}}_t + \frac{1}{NT\sqrt{NT}}\sum_{s=1,s\neq t}^{T}\sum_{i=1}^{N}\varepsilon_{i,t}\breve{\varepsilon}_{i,s}\breve{\breve{\alpha}}_s - \tilde{\gamma}_{RP}\frac{1}{NT^2\sqrt{NT}}\sum_{i=1}^{N}\sum_{s=1}^{T}\varepsilon_{i,s}\varepsilon_{i,t}\breve{\breve{\alpha}}_t\Bigg\},
\end{aligned}
$$

and therefore we have:

$$
\begin{aligned}
&\frac{1}{N^2T^2}[\varepsilon\varepsilon' W^2\varepsilon\varepsilon' W^2 H(H'W^2H/T)^{-1}M^*]_t = M^{*\prime}(H'W^2H/T)^{-1}\Bigg\{\frac{1}{T^2}\breve{\eta}_t^4\breve{h}_t + \frac{1}{\sqrt{N}T^2}\breve{\kappa}_t\breve{\eta}_t^2\breve{h}_t + \frac{1}{\sqrt{NT}T}\breve{\breve{\alpha}}_t \\
&+\frac{1}{T^2\sqrt{N}}\breve{\kappa}_t\breve{\eta}_t^2\breve{h}_t + \frac{1}{NT^2}\breve{\kappa}_t^2\breve{h}_t + \frac{1}{NT\sqrt{N}}\breve{\varphi}_t + \frac{1}{NT\sqrt{T}}\breve{\gamma}_t \\
&+\frac{1}{T\sqrt{NT}}\breve{\eta}_t^2\breve{\breve{\alpha}}_t + \frac{1}{NT\sqrt{T}}\breve{\kappa}_t\breve{\breve{\alpha}}_t + \frac{1}{N\sqrt{NT}}\breve{\delta}_t + \frac{1}{NT}\breve{\chi}_t - \tilde{\gamma}_{RP}\left[\frac{1}{T^2\sqrt{N}}\breve{\beta}_{1,t} + \frac{1}{NT^2\sqrt{T}}\breve{\beta}_{2,t} + \frac{1}{NT^2}\breve{\beta}_{3,t}\right]\Bigg\},
\end{aligned}
\tag{B.91}
$$

where the three terms

$$
\breve{\beta}_{1,t} := \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{s=1}^{T}\breve{h}_t\breve{\eta}_t^2\varepsilon_{i,t}\varepsilon_{i,s}, \qquad \breve{\beta}_{2,t} := \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{s=1}^{T}\breve{h}_t\breve{\kappa}_t^2\varepsilon_{i,t}\varepsilon_{i,s}, \qquad \breve{\beta}_{3,t} := \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{s=1}^{T}\breve{\breve{\alpha}}_t\varepsilon_{i,t}\varepsilon_{i,s}
$$

are all $O_p(1)$ by the new Assumption B.7 d).

The date t-element of the fourth term in the r.h.s. of equation (B.87), that is

$\frac{1}{N^2T^2}\varepsilon\Lambda H'W^2\varepsilon M^*\varepsilon'W^2H$, can be easily computed by noting that:

$$\begin{aligned}\frac{1}{N^2T^2}\varepsilon\Lambda H'W^2\varepsilon\varepsilon'W^2H &= \frac{1}{T\sqrt{N}}\left(\frac{\varepsilon\Lambda}{\sqrt{N}}\right)\left[\frac{1}{N}\left(\frac{1}{\sqrt{T}}H'W^2\varepsilon\right)\left(\frac{1}{\sqrt{T}}\varepsilon'W^2H\right)\right] \\ &= \frac{1}{T\sqrt{N}}\left(\frac{\varepsilon\Lambda}{\sqrt{N}}\right)\left[\frac{1}{N}\sum_{i=1}^{N}\breve{\aleph}_i\breve{\aleph}_i'\right] = \frac{1}{T\sqrt{N}}\left(\frac{\varepsilon\Lambda}{\sqrt{N}}\right)\breve{\Pi}_4 ,\end{aligned}$$

which implies:

$$\left[\frac{1}{N^2T^2}\varepsilon\Lambda H'W^2\varepsilon\varepsilon'W^2H(H'W^2H/T)^{-1}M^*\right]_t = \frac{1}{T\sqrt{N}}M^*(H'W^2H/T)^{-1}\breve{\Pi}_4\xi_t \ . \quad \text{(B.92)}$$

The sixth term in the r.h.s. of equation (B.76) can be written as:

$$\begin{aligned}&\frac{1}{N^3T^2}(H\Lambda'+\varepsilon)\cdot\left[\varepsilon'W^2\varepsilon+\Lambda H'W^2\varepsilon\right]\cdot\left[\varepsilon'W^2\varepsilon\Lambda+\Lambda H'W^2\varepsilon\Lambda\right]M^{**} \\ &= \frac{1}{N^3T^2}H\Lambda'\varepsilon'W^2\varepsilon\varepsilon'W^2\varepsilon\Lambda M^{**}+\frac{1}{N^3T^2}H\Lambda'\varepsilon'W^2\varepsilon\Lambda H'W^2\varepsilon\Lambda M^{**} \\ &+\frac{1}{N^3T^2}H\Lambda'\Lambda H'W^2\varepsilon\varepsilon'W^2\varepsilon\Lambda M^{**}+\frac{1}{N^3T^2}H\Lambda'\Lambda H'W^2\varepsilon\Lambda H'W^2\varepsilon\Lambda M^{**} \\ &+\frac{1}{N^3T^2}\varepsilon\varepsilon'W^2\varepsilon\varepsilon'W^2\varepsilon\Lambda M^{**}+\frac{1}{N^3T^2}\varepsilon\varepsilon'W^2\varepsilon\Lambda H'W^2\varepsilon\Lambda M^{**} \\ &+\frac{1}{N^3T^2}\varepsilon\Lambda H'W^2\varepsilon\varepsilon'W^2\varepsilon\Lambda M^{**}+\frac{1}{N^3T^2}\varepsilon\Lambda H'W^2\varepsilon\Lambda H'W^2\varepsilon\Lambda M^{**} \quad \text{(B.93)}\end{aligned}$$

The date t - element of the first term in the r.h.s. of equation (B.93) can be computed by first noting that:

$$\begin{aligned}&\left[\frac{1}{N^2T}W\varepsilon\varepsilon'W^2\varepsilon\Lambda\right]_t = \frac{1}{N^2T}\sum_{i=1}^{N}\sum_{s=1}^{T}\sum_{\ell=1}^{N}\breve{\varepsilon}_{i,t}\breve{\varepsilon}_{i,s}\breve{\varepsilon}_{\ell,s}\lambda_\ell = \frac{1}{N\sqrt{NT}}\sum_{i=1}^{N}\sum_{s=1}^{T}\breve{\varepsilon}_{i,t}\breve{\varepsilon}_{i,s}\breve{\xi}_s \\ &= \frac{1}{N\sqrt{NT}}\sum_{i=1}^{N}\breve{\varepsilon}_{i,t}\breve{\varepsilon}_{i,t}\breve{\xi}_t+\left[\frac{1}{N\sqrt{NT}}\sum_{i=1}^{N}\sum_{s=1,s\neq t}^{T}\breve{\varepsilon}_{i,t}\breve{\varepsilon}_{i,s}\breve{\xi}_s\right] \\ &= \frac{1}{N\sqrt{NT}}\sum_{i=1}^{N}\breve{\varepsilon}_{i,t}^2\breve{\xi}_t+\left[\frac{1}{N\sqrt{N}}\left(\frac{1}{T}\sum_{i=1}^{N}\sum_{s=1,s\neq t}^{T}\breve{\varepsilon}_{i,t}E[\breve{\varepsilon}_{i,s}\breve{\xi}_s]\right)+\frac{1}{N\sqrt{T}}\left\{\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{s=1,s\neq t}^{T}\breve{\varepsilon}_{i,t}(\breve{\varepsilon}_{i,s}\breve{\xi}_s-E[\breve{\varepsilon}_{i,s}\breve{\xi}_s])\right\}\right] \\ &= \frac{1}{N\sqrt{NT}}\sum_{i=1}^{N}\breve{\varepsilon}_{i,t}^2\breve{\xi}_t+\left[\frac{1}{N\sqrt{N}}\breve{\varphi}_t+\frac{1}{N\sqrt{T}}\breve{\gamma}_t\right] \\ &= \frac{1}{\sqrt{NT}}\left\{\frac{1}{N}\sum_{i=1}^{N}E[\breve{\varepsilon}_{i,t}^2]\breve{\xi}_t\right\}+\frac{1}{NT}\cdot\left\{\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\breve{\varepsilon}_{i,t}^2\breve{\xi}_t-E[\breve{\varepsilon}_{i,t}^2]\breve{\xi}_t)\right\} + \left[\frac{1}{N\sqrt{N}}\breve{\varphi}_t+\frac{1}{N\sqrt{T}}\breve{\gamma}_t\right] \\ &= \frac{1}{\sqrt{NT}}\breve{\eta}_t^2\breve{\xi}_t+\frac{1}{NT}\breve{\kappa}_t\breve{\xi}_t+\frac{1}{N\sqrt{N}}\breve{\varphi}_t+\frac{1}{N\sqrt{T}}\breve{\gamma}_t, \quad \text{(B.94)}\end{aligned}$$

Then, date t - element of the first term in the r.h.s. of equation (B.93) is:

$$\begin{aligned}
&\frac{1}{N^3T^2}H\Lambda'\varepsilon'W^2\varepsilon\varepsilon'W^2\varepsilon\Lambda \\
&= H\cdot\frac{1}{\sqrt{N}}\left[\frac{1}{T}\left(\frac{\Lambda'\varepsilon'W}{\sqrt{N}}\right)\cdot\left(\frac{1}{N^2T}W\varepsilon\varepsilon'W^2\varepsilon\Lambda\right)\right] = H\cdot\frac{1}{\sqrt{N}}\frac{1}{T}\sum_{t=1}^{T}\breve{\xi}_t\left[\frac{1}{N^2T}W\varepsilon\varepsilon'W^2\varepsilon\Lambda\right]_t \\
&= H\cdot\frac{1}{\sqrt{N}}\left[\frac{1}{T}\sum_{t=1}^{T}\breve{\xi}_t\left(\frac{1}{\sqrt{NT}}\breve{\eta}_t^2\breve{\xi}_t' + \frac{1}{NT}\breve{\kappa}_t\breve{\xi}_t' + \frac{1}{N\sqrt{N}}\breve{\breve{\varphi}}_t' + \frac{1}{N\sqrt{T}}\breve{\breve{\gamma}}_t\right)\right] \\
&= H\cdot\frac{1}{\sqrt{N}}\left[\frac{1}{\sqrt{NT}}\left(\frac{1}{T}\sum_{t=1}^{T}\breve{\eta}_t^2\breve{\xi}_t\breve{\xi}_t'\right) + \frac{1}{NT}\left(\frac{1}{T}\sum_{t=1}^{T}\breve{\kappa}_t\breve{\xi}_t\breve{\xi}_t'\right) + \frac{1}{N\sqrt{N}}\left(\frac{1}{T}\sum_{t=1}^{T}\breve{\xi}_t\breve{\breve{\varphi}}_t'\right) + \frac{1}{N\sqrt{T}}\left(\frac{1}{T}\sum_{t=1}^{T}\breve{\xi}_t\breve{\breve{\gamma}}_t\right)\right]
\end{aligned} \tag{B.95}$$

which implies:

$$\begin{aligned}
&\left[\frac{1}{N^3T^2}H\Lambda'\varepsilon'W^2\varepsilon\varepsilon'W^2\varepsilon\Lambda M^{**}\right]_t \\
&= \frac{1}{\sqrt{N}}M^{**\prime}\left[\frac{1}{\sqrt{NT}}\left(\frac{1}{T}\sum_{t=1}^{T}\breve{\eta}_t^2\breve{\xi}_t\breve{\xi}_t'\right) + \frac{1}{NT}\left(\frac{1}{T}\sum_{t=1}^{T}\breve{\kappa}_t\breve{\xi}_t\breve{\xi}_t'\right) + \frac{1}{N\sqrt{N}}\left(\frac{1}{T}\sum_{t=1}^{T}\breve{\xi}_t\breve{\breve{\varphi}}_t'\right)' + \frac{1}{N\sqrt{T}}\left(\frac{1}{T}\sum_{t=1}^{T}\breve{\xi}_t\breve{\breve{\gamma}}_t\right)'\right]h_t
\end{aligned} \tag{B.96}$$

The date t - element of the second term in the r.h.s. of equation (B.93) can be computed by noting that:

$$\begin{aligned}
&\frac{1}{N^3T^2}H\Lambda'\varepsilon'W^2\varepsilon\Lambda H'W^2\varepsilon\Lambda = H\cdot\frac{1}{N\sqrt{NT}}\left[\frac{1}{T}\left(\frac{\Lambda'\varepsilon'W}{\sqrt{N}}\right)\cdot\left(\frac{W\varepsilon\Lambda}{\sqrt{N}}\right)\cdot\left(\frac{1}{\sqrt{NT}}H'W^2\varepsilon\Lambda\right)\right] \\
&= H\cdot\frac{1}{N\sqrt{NT}}\left(\frac{1}{T}\sum_{t=1}^{T}\breve{\xi}_t\breve{\xi}_t'\right)\breve{\Pi}_1
\end{aligned}$$

which implies:

$$\left[\frac{1}{N^3T^2}H\Lambda'\varepsilon'W^2\varepsilon\Lambda H'W^2\varepsilon\Lambda M^{**}\right]_t = \frac{1}{N\sqrt{NT}}M^{**\prime}\breve{\Pi}_1'\left(\frac{1}{T}\sum_{t=1}^{T}\breve{\xi}_t\breve{\xi}_t'\right)h_t \tag{B.97}$$

The date t - element of the third term in the r.h.s. of equation (B.93) can be computed by first noting that:

$$\begin{aligned}
&\left[\frac{1}{N^3T^2}H\Lambda'\Lambda H'W^2\varepsilon\varepsilon'W^2\varepsilon\Lambda M^{**}\right]_t = \left[H\frac{1}{\sqrt{NT}}\left(\frac{\Lambda'\Lambda}{N}\right)\left(\frac{1}{NT\sqrt{NT}}H'W^2\varepsilon\varepsilon'W^2\varepsilon\Lambda\right)M^{**}\right]_t \\
&= \left[H\frac{1}{\sqrt{NT}}\left(\frac{\Lambda'\Lambda}{N}\right)\breve{\Pi}_3M^{**}\right]_t = \frac{1}{\sqrt{NT}}M^{**\prime}\breve{\Pi}_3'\left(\frac{\Lambda'\Lambda}{N}\right)h_t.
\end{aligned} \tag{B.98}$$

The date t - element of the fourth term in the r.h.s. of equation (B.93) is:

$$\left[\frac{1}{N^3T^2}H\Lambda'\Lambda H'W^2\varepsilon\Lambda H'W^2\varepsilon\Lambda M^{**}\right]_t = \left[H\frac{1}{NT}\left(\frac{\Lambda'\Lambda}{N}\right)\cdot\left(\frac{H'W^2\varepsilon\Lambda}{\sqrt{NT}}\right)\left(\frac{H'W^2\varepsilon\Lambda}{\sqrt{NT}}\right)M^{**}\right]_t$$
$$= \left[H\frac{1}{NT}\left(\frac{\Lambda'\Lambda}{N}\right)\breve{\Pi}_1\breve{\Pi}_1 M^{**}\right]_t = \frac{1}{NT}M^{**\prime}\breve{\Pi}_1'\breve{\Pi}_1'\left(\frac{\Lambda'\Lambda}{N}\right)h_t \tag{B.99}$$

The date t - element of the fifth term in the r.h.s. of equation (B.93) is:

$$\begin{aligned}\left[\frac{1}{N^3T^2}\varepsilon\varepsilon'W^2\varepsilon\varepsilon'W^2\varepsilon\Lambda M^{**}\right]_t &= M^{**\prime}\frac{1}{NT}\sum_{s=1}^{T}\sum_{i=1}^{N}\varepsilon_{it}\breve{\varepsilon}_{is}\left[\frac{1}{N^2T}W\varepsilon\varepsilon'W^2\varepsilon\Lambda\right]_s \\ &= M^{**\prime}\frac{1}{NT}\sum_{s=1}^{T}\sum_{i=1}^{N}\varepsilon_{it}\breve{\varepsilon}_{is}\left(\frac{1}{\sqrt{NT}}\breve{\eta}_s^2\breve{\xi}_s + \frac{1}{NT}\breve{\kappa}_s\breve{\xi}_s + \frac{1}{N\sqrt{N}}\breve{\breve{\varphi}}_s + \frac{1}{N\sqrt{T}}\breve{\breve{\gamma}}_s\right) \\ &= M^{**\prime}\left[\frac{1}{NT\sqrt{T}}\breve{\beta}_{4,t} + \frac{1}{NT\sqrt{NT}}\breve{\beta}_{5,t} + \frac{1}{N^2\sqrt{T}}\breve{\beta}_{6,t} + \frac{1}{NT\sqrt{N}}\breve{\beta}_{7,t}\right], \end{aligned} \tag{B.100}$$

where:

$$\begin{aligned}\breve{\beta}_{4,t} &:= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{s=1}^{T}\varepsilon_{it}\breve{\varepsilon}_{is}\breve{\eta}_s^2\breve{\xi}_s, \qquad \breve{\beta}_{5,t} := \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{s=1}^{T}\varepsilon_{it}\breve{\varepsilon}_{is}\breve{\kappa}_s\breve{\xi}_s, \\ \breve{\beta}_{6,t} &:= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{s=1}^{T}\varepsilon_{it}\breve{\varepsilon}_{is}\breve{\breve{\varphi}}_s, \qquad \breve{\beta}_{7,t} := \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{s=1}^{T}\varepsilon_{it}\breve{\varepsilon}_{is}\breve{\breve{\gamma}}_s.\end{aligned}$$

The date t - element of the sixth term in the r.h.s. of equation (B.93) can be computed by first noting that:

$$\frac{1}{N^3T^2}\varepsilon\varepsilon'W^2\varepsilon\Lambda H'W^2\varepsilon\Lambda = \frac{1}{\sqrt{NT}}\cdot\left(\frac{1}{N^2T}\varepsilon\varepsilon'W^2\varepsilon\Lambda\right)\cdot\left(\frac{1}{\sqrt{NT}}H'W^2\varepsilon\Lambda\right) = \frac{1}{\sqrt{NT}}\cdot\left(\frac{1}{N^2T}\varepsilon\varepsilon'W^2\varepsilon\Lambda\right)\breve{\Pi}_1$$

which implies:

$$\left[\frac{1}{N^3T^2}\varepsilon\varepsilon'W^2\varepsilon\Lambda H'W^2\varepsilon\Lambda M^{**}\right]_t$$
$$= \frac{1}{N\sqrt{NT}}M^{**\prime}\breve{\Pi}_1'\left\{\frac{1}{\sqrt{NT}}\eta_t^2\breve{\xi}_t + \frac{1}{NT}\kappa_t\breve{\xi}_t + \frac{1}{N\sqrt{N}}\breve{\varphi}_t + \frac{1}{N\sqrt{T}}\breve{\gamma}_t + \tilde{\gamma}_{RP}\left[\frac{1}{T^2\sqrt{N}}\eta_t^2\breve{\xi}_t + \frac{1}{NT^2}\kappa_t\breve{\xi}_t + \frac{1}{NT\sqrt{T}}\breve{\alpha}_t^*\right]\right\} \tag{B.101}$$

The date t - element of the seventh term in the r.h.s. of equation (B.93) can be computed by first noting that:

$$\frac{1}{N^3T^2}\varepsilon\Lambda H'W^2\varepsilon\varepsilon'W^2\varepsilon\Lambda = \frac{1}{N\sqrt{T}}\left(\frac{\varepsilon\Lambda}{\sqrt{N}}\right)\cdot\left(\frac{1}{NT\sqrt{NT}}H'W^2\varepsilon\varepsilon'W^2\varepsilon\Lambda\right) = \frac{1}{N\sqrt{T}}\left(\frac{\varepsilon\Lambda}{\sqrt{N}}\right)\cdot\breve{\Pi}_3$$

which implies:

$$\left[\frac{1}{N^3T^2}\varepsilon\Lambda H'W^2\varepsilon\varepsilon'W^2\varepsilon\Lambda M^{**}\right]_t \quad = \quad \frac{1}{N\sqrt{T}}M^{**\prime}\breve{\Pi}_3'\breve{\xi}_t \tag{B.102}$$

The date t - element of the eighth, and final, term in the r.h.s. of equation (B.93) can be computed by first noting that:

$$\frac{1}{N^3T^2}\varepsilon\Lambda H'W^2\varepsilon\Lambda H'W^2\varepsilon\Lambda \quad = \quad \frac{1}{NT\sqrt{N}}\left(\frac{\varepsilon\Lambda}{\sqrt{N}}\right)\left(\frac{1}{\sqrt{NT}}H'W^2\varepsilon\Lambda\right)\cdot\left(\frac{1}{\sqrt{NT}}H'W^2\varepsilon\Lambda\right) = \frac{1}{NT\sqrt{N}}\left(\frac{\varepsilon\Lambda}{\sqrt{N}}\right)\cdot\breve{\Pi}_1\cdot\breve{\Pi}_1$$

which implies:

$$\left[\frac{1}{N^3T^2}\varepsilon\Lambda H'W^2\varepsilon\Lambda H'W^2\varepsilon\Lambda M^{**}\right]_t \quad = \quad \frac{1}{NT\sqrt{N}}M^{**\prime}\breve{\Pi}_1'\breve{\Pi}_1'\breve{\xi}_t\ . \tag{B.103}$$

The seventh term in the r.h.s. of equation (B.76) can be written as:

$$\frac{1}{N}(H\Lambda'+\varepsilon)\frac{1}{N^2T^2}(\varepsilon'W^2\varepsilon+\Lambda H'W^2\varepsilon)^2\frac{1}{T}\varepsilon'W^2H\cdot(H'W^2H/T)^{-1}M^{**}$$
$$=H\frac{1}{N^3T^3}\left[\Lambda'\varepsilon'W^2\varepsilon\varepsilon'W^2\varepsilon+\Lambda'\varepsilon'W^2\varepsilon\Lambda H'W^2\varepsilon+\Lambda'\Lambda H'W^2\varepsilon\varepsilon'W^2\varepsilon+\Lambda'\Lambda H'W^2\varepsilon\Lambda H'W^2\varepsilon\right]$$
$$\times\varepsilon'W^2H\cdot(H'W^2H/T)^{-1}M^{**}$$
$$+\left[\varepsilon\varepsilon'W^2\varepsilon\varepsilon'W^2\varepsilon+\varepsilon\varepsilon'W^2\varepsilon\Lambda H'W^2\varepsilon+\varepsilon\Lambda H'W^2\varepsilon\varepsilon'W^2\varepsilon+\varepsilon\Lambda H'W^2\varepsilon\Lambda H'W^2\varepsilon\right]$$
$$\times\varepsilon'W^2H\cdot(H'W^2H/T)^{-1}M^{**} \tag{B.104}$$

The date t - element of the first term in the r.h.s. of equation (B.104) can be computed by first noting that:

$$\frac{1}{N^3T^3}(\Lambda'\varepsilon'W^2\varepsilon\varepsilon'W^2\varepsilon)\varepsilon'W^2H$$
$$=\frac{1}{T}\left(\frac{\Lambda'\varepsilon'W^2\varepsilon\varepsilon'W}{N^2T}\right)\left(\frac{W\varepsilon\varepsilon'W^2H}{NT}\right)=\frac{1}{T}\sum_{t=1}^{T}\left[\frac{W\varepsilon\varepsilon'W^2\varepsilon\Lambda'}{N^2T}\right]_t\left(\left[\frac{W\varepsilon\varepsilon'W^2H}{NT}\right]_t\right)'$$
$$=\frac{1}{T}\sum_{t=1}^{T}\left[\frac{1}{\sqrt{NT}}\breve{\eta}_t^2\breve{\xi}_t+\frac{1}{NT}\breve{\kappa}_t\breve{\xi}_t+\frac{1}{N\sqrt{N}}\breve{\varphi}_t+\frac{1}{N\sqrt{T}}\breve{\gamma}_t\right]\left[\frac{1}{T}\breve{\eta}_t^2\breve{h}_t'+\frac{1}{T\sqrt{N}}\breve{\kappa}_t\breve{h}_t'+\frac{1}{\sqrt{NT}}\breve{\alpha}_t'\right] \tag{B.105}$$

which implies:

$$\left[H\frac{1}{N^3T^3}(\Lambda'\varepsilon'W^2\varepsilon\varepsilon'W^2\varepsilon)\varepsilon'W^2H\cdot(H'W^2H/T)^{-1}M^{**}\right]_t=M^{**\prime}(H'W^2H/T)^{-1}$$
$$\times\left\{\frac{1}{T}\sum_{s=1}^{T}\left[\frac{1}{T}\breve{\eta}_s^2\breve{h}_s+\frac{1}{T\sqrt{N}}\breve{\kappa}_s\breve{h}_s+\frac{1}{\sqrt{NT}}\breve{\alpha}_s\right]\left[\frac{1}{\sqrt{NT}}\breve{\eta}_t^2\breve{\xi}_s'+\frac{1}{NT}\breve{\kappa}_t\breve{\xi}_s'+\frac{1}{N\sqrt{N}}\breve{\varphi}_s'+\frac{1}{N\sqrt{T}}\breve{\gamma}_s'\right]\right\}h_t$$
$$=M^{**\prime}(H'W^2H/T)^{-1}\left(\frac{H'\varepsilon'W^2\varepsilon\varepsilon W}{NT\sqrt{T}}\right)\left(\frac{W\varepsilon\varepsilon'W^2\Lambda}{N^2T\sqrt{T}}\right)h_t \tag{B.106}$$

The date t - element of the second term in the r.h.s. of equation (B.104) can be computed by first noting that:

$$\frac{1}{N^3T^3}(\Lambda'\varepsilon'W^2\varepsilon\Lambda H'W^2\varepsilon)\varepsilon'W^2H=\frac{1}{TN}\left(\frac{\Lambda'\varepsilon'W^2\varepsilon\Lambda}{NT}\right)\left(\frac{H'W^2\varepsilon\varepsilon'W^2H}{NT}\right)=\frac{1}{TN}\breve{\Pi}_2\breve{\Pi}_4 \tag{B.107}$$

which implies:

$$\left[H\frac{1}{N^3T^3}(\Lambda'\varepsilon'W^2\varepsilon\Lambda H'W^2\varepsilon)\varepsilon'W^2H\cdot(H'W^2H/T)^{-1}M^{**}\right]_t = M^{**\prime}(H'W^2H/T)^{-1}\frac{1}{TN}\breve{\Pi}_4\breve{\Pi}_2h_t \tag{B.108}$$

The date t - element of the third term in the r.h.s. of equation (B.104) can be computed by first noting that:

$$\begin{aligned}&\frac{1}{N^3T^3}(\Lambda'\Lambda H'W^2\varepsilon\varepsilon'W^2\varepsilon)\varepsilon'W^2H\\&=\frac{1}{T}(\Lambda'\Lambda/N)\left(\frac{H'W^2\varepsilon\varepsilon'W}{NT}\right)\left(\frac{W\varepsilon\varepsilon'W^2H}{NT}\right) = (\Lambda'\Lambda/N)\frac{1}{T}\sum_{t=1}^{T}\left[\frac{W\varepsilon\varepsilon'W^2H}{NT}\right]_t\left(\left[\frac{W\varepsilon\varepsilon'W^2H}{NT}\right]_t\right)'\\&=(\Lambda'\Lambda/N)\frac{1}{T}\sum_{t=1}^{T}\left[\frac{1}{T}\breve{\eta}_t^2\breve{h}_t+\frac{1}{T\sqrt{N}}\breve{\kappa}_t\breve{h}_t+\frac{1}{\sqrt{NT}}\breve{\breve{\alpha}}_t\right]\cdot\left[\frac{1}{T}\breve{\eta}_t^2\breve{h}_t'+\frac{1}{T\sqrt{N}}\breve{\kappa}_t\breve{h}_t'+\frac{1}{\sqrt{NT}}\breve{\breve{\alpha}}_t'\right]\end{aligned} \tag{B.109}$$

which implies:

$$\begin{aligned}&\left[H\frac{1}{N^3T^3}(\Lambda'\Lambda H'W^2\varepsilon\varepsilon'W^2\varepsilon)\varepsilon'W^2H\cdot(H'W^2H/T)^{-1}M^{**}\right]_t = M^{**\prime}(H'W^2H/T)^{-1}\\&\times\left\{\frac{1}{T}\sum_{t=1}^{T}\left[\frac{1}{T}\breve{\eta}_t^2\breve{h}_t+\frac{1}{T\sqrt{N}}\breve{\kappa}_t\breve{h}_t+\frac{1}{\sqrt{NT}}\breve{\breve{\alpha}}_t\right]\left[\frac{1}{T}\breve{\eta}_t^2\breve{h}_t'+\frac{1}{T\sqrt{N}}\breve{\kappa}_t\breve{h}_t'+\frac{1}{\sqrt{NT}}\breve{\breve{\alpha}}_t'\right](\Lambda'\Lambda/N)\right\}h_t\\&=M^{**\prime}(H'W^2H/T)^{-1}\left(\frac{H'\varepsilon'W^2\varepsilon\varepsilon W}{NT\sqrt{T}}\right)\left(\frac{W\varepsilon\varepsilon'W^2H}{NT\sqrt{T}}\right)\left(\frac{\Lambda'\Lambda}{N}\right)h_t\end{aligned} \tag{B.110}$$

The date t - element of the fourth term in the r.h.s. of equation (B.104) can be computed by first noting that:

$$\begin{aligned}\frac{1}{N^3T^3}(\Lambda'\Lambda H'W^2\varepsilon\Lambda H'W^2\varepsilon)\varepsilon'W^2H &= \frac{1}{T\sqrt{NT}}(\Lambda'\Lambda/N)\left(\frac{H'W^2\varepsilon\Lambda}{\sqrt{NT}}\right)\left(\frac{H'W\varepsilon\varepsilon'W^2H}{NT}\right)\\&=\frac{1}{T\sqrt{NT}}(\Lambda'\Lambda/N)\breve{\Pi}_1\breve{\Pi}_4\end{aligned} \tag{B.111}$$

which implies:

$$\begin{aligned}&\left[H\frac{1}{N^3T^3}(\Lambda'\Lambda H'W^2\varepsilon\Lambda H'W^2\varepsilon)\varepsilon'W^2H\cdot(H'W^2H/T)^{-1}M^{**}\right]_t\\&=M^{**\prime}(H'W^2H/T)^{-1}\frac{1}{TN}\breve{\Pi}_4\breve{\Pi}_1'(\Lambda'\Lambda/N)h_t\end{aligned} \tag{B.112}$$

The date t - element of the fifth term in the r.h.s. of equation (B.104) is

$$\begin{aligned}
&\left[\frac{1}{N^3T^3}(\varepsilon\varepsilon' W^2\varepsilon\varepsilon' W^2\varepsilon)\varepsilon' W^2 H\cdot(H'W^2H/T)^{-1}M^{**}\right]_t\\
&= M^{**\prime}(H'W^2H/T)^{-1}\frac{1}{NT}\sum_{s=1}^{T}\sum_{i=1}^{N}\varepsilon_{i,t}\breve{\varepsilon}_{i,s}\left[\frac{W\varepsilon\varepsilon' W^2\varepsilon\varepsilon' W^2 H}{N^2T^2}\right]_s\\
&= M^{**\prime}(H'W^2H/T)^{-1}\frac{1}{NT}\sum_{s=1}^{T}\sum_{i=1}^{N}\varepsilon_{i,t}\breve{\varepsilon}_{i,s}\left\{\frac{1}{T^2}\breve{\eta}_s^4\breve{h}_s+\frac{1}{\sqrt{N}T^2}\breve{\kappa}_s\breve{\eta}_s^2\breve{h}_s+\frac{1}{\sqrt{NT}T}\breve{\breve{\alpha}}_s\right.\\
&+\frac{1}{T^2\sqrt{N}}\breve{\kappa}_s\breve{\eta}_s^2\breve{h}_s+\frac{1}{NT^2}\breve{\kappa}_s^2\breve{h}_s+\frac{1}{NT\sqrt{N}}\breve{\breve{\varphi}}_s+\frac{1}{NT\sqrt{T}}\breve{\breve{\gamma}}_s\\
&\left.+\frac{1}{T\sqrt{NT}}\breve{\eta}_s^2\breve{\breve{\alpha}}_s+\frac{1}{NT\sqrt{T}}\breve{\kappa}_s\breve{\breve{\alpha}}_s+\frac{1}{N\sqrt{NT}}\breve{\delta}_s+\frac{1}{NT}\breve{\chi}_s-\tilde{\gamma}_{RP}\left[\frac{1}{T^2\sqrt{N}}\breve{\beta}_{1,s}+\frac{1}{NT^2\sqrt{T}}\breve{\beta}_{2,s}+\frac{1}{NT^2}\breve{\beta}_{3,s}\right]\right\}
\end{aligned} \tag{B.113}$$

The date t - element of the sixth term in the r.h.s. of equation (B.104) is

$$\begin{aligned}
&\left[\frac{1}{N^3T^3}(\varepsilon\varepsilon' W^2\varepsilon\Lambda H' W^2\varepsilon)\varepsilon' W^2 H\cdot(H'W^2H/T)^{-1}M^{**}\right]_t\\
&= M^{**\prime}(H'W^2H/T)^{-1}\frac{1}{T}\left(\frac{H'W^2\varepsilon\varepsilon' W^2 H}{NT}\right)\left[\frac{\varepsilon\varepsilon' W^2\varepsilon\Lambda}{N^2T}\right]_t\\
&= M^{**\prime}(H'W^2H/T)^{-1}\frac{1}{T}\breve{\Pi}_4\left\{\frac{1}{\sqrt{NT}}\eta_t^2\breve{\xi}_t+\frac{1}{NT}\kappa_t\breve{\xi}_t+\frac{1}{N\sqrt{N}}\breve{\varphi}_t+\frac{1}{N\sqrt{T}}\breve{\gamma}_t+\right.\\
&\left.+\tilde{\gamma}_{RP}\left[\frac{1}{T^2\sqrt{N}}\eta_t^2\breve{\xi}_t+\frac{1}{NT^2}\kappa_t\breve{\xi}_t+\frac{1}{NT\sqrt{T}}\breve{\alpha}_t^*\right]\right\}
\end{aligned} \tag{B.114}$$

The date t - element of the seventh term in the r.h.s. of equation (B.104) is

$$\begin{aligned}
&\left[\frac{1}{N^3T^3}(\varepsilon\Lambda H'W^2\varepsilon\varepsilon'W^2\varepsilon)\varepsilon'W^2H\cdot(H'W^2H/T)^{-1}M^{**}\right]_t\\
&=M^{**\prime}(H'W^2H/T)^{-1}\frac{1}{\sqrt{N}}\left(\frac{H'W^2\varepsilon\varepsilon'W}{NT\sqrt{T}}\right)\left(\frac{W\varepsilon\varepsilon'W^2H}{NT\sqrt{T}}\right)\left[\frac{\varepsilon\Lambda}{\sqrt{N}}\right]_t\\
&=M^{**\prime}(H'W^2H/T)^{-1}\frac{1}{\sqrt{N}}\left\{\frac{1}{T}\sum_{s=1}^{T}\left[\frac{W\varepsilon\varepsilon'W^2H}{NT}\right]_s\left(\left[\frac{W\varepsilon\varepsilon'W^2H}{NT}\right]_s\right)'\right\}\xi_t\\
&=M^{**\prime}(H'W^2H/T)^{-1}\frac{1}{T\sqrt{N}}\\
&\times\left\{\frac{1}{T}\sum_{s=1}^{T}\left[\frac{1}{T}\breve{\eta}_s^2\breve{h}_s+\frac{1}{T\sqrt{N}}\breve{\kappa}_s\breve{h}_s+\frac{1}{\sqrt{NT}}\breve{\breve{\alpha}}_s\right]\cdot\left[\frac{1}{T}\breve{\eta}_s^2\breve{h}_s'+\frac{1}{T\sqrt{N}}\breve{\kappa}_s\breve{h}_s'+\frac{1}{\sqrt{NT}}\breve{\breve{\alpha}}_s'\right]\right\}\xi_t\\
&=M^{**\prime}(H'W^2H/T)^{-1}\frac{1}{\sqrt{N}}\left(\frac{H'W^2\varepsilon\varepsilon'W}{NT\sqrt{T}}\right)\left(\frac{W\varepsilon\varepsilon'W^2H}{NT\sqrt{T}}\right)\xi_t \qquad \text{(B.115)}
\end{aligned}$$

The date t - element of the eighth term in the r.h.s. of equation (B.104) is

$$\begin{aligned}
&\left[\frac{1}{N^3T^3}(\varepsilon\Lambda H'W^2\varepsilon\Lambda H'W^2\varepsilon)\varepsilon'W^2H\cdot(H'W^2H/T)^{-1}M^{**}\right]_t\\
&=M^{**\prime}(H'W^2H/T)^{-1}\frac{1}{NT\sqrt{T}}\left(\frac{H'W^2\varepsilon\varepsilon'W^2H}{NT}\right)\left(\frac{\Lambda\varepsilon'W^2H}{\sqrt{NT}}\right)\left[\frac{\varepsilon\Lambda}{\sqrt{N}}\right]_t\\
&=M^{**\prime}(H'W^2H/T)^{-1}\frac{1}{NT\sqrt{T}}\breve{\Pi}_4\breve{\Pi}_1'\xi_t \qquad \text{(B.116)}
\end{aligned}$$

Therefore, the date t - element (B.104), which we denote ad $r_{7,t}$ is:

$$\begin{aligned}
r_{7,t} \quad = \quad &= M^{**\prime}(H'W^2H/T)^{-1}\Bigg\{\left(\frac{H'\varepsilon'W^2\varepsilon\varepsilon W}{NT\sqrt{T}}\right)\left(\frac{W\varepsilon\varepsilon'W^2\Lambda}{N^2T\sqrt{T}}\right)h_t+\frac{1}{TN}\breve{\Pi}_4\breve{\Pi}_2h_t\\
&+\left(\frac{H'\varepsilon'W^2\varepsilon\varepsilon W}{NT\sqrt{T}}\right)\left(\frac{W\varepsilon\varepsilon'W^2H}{NT\sqrt{T}}\right)\left(\frac{\Lambda'\Lambda}{N}\right)h_t+\frac{1}{TN}\breve{\Pi}_4\breve{\Pi}_1'(\Lambda'\Lambda/N)h_t\\
&+\frac{1}{NT}\sum_{s=1}^{T}\sum_{i=1}^{N}\varepsilon_{i,t}\breve{\varepsilon}_{i,s}\Bigg\{\frac{1}{T^2}\breve{\eta}_s^4\breve{h}_s+\frac{1}{\sqrt{N}T^2}\breve{\kappa}_s\breve{\eta}_s^2\breve{h}_s+\frac{1}{\sqrt{NT}T}\breve{\breve{\alpha}}_s\\
&+\frac{1}{T^2\sqrt{N}}\breve{\kappa}_s\breve{\eta}_s^2\breve{h}_s+\frac{1}{NT^2}\breve{\kappa}_s^2\breve{h}_s+\frac{1}{NT\sqrt{N}}\breve{\breve{\varphi}}_s+\frac{1}{NT\sqrt{T}}\breve{\breve{\gamma}}_s\\
&+\frac{1}{T\sqrt{NT}}\breve{\eta}_s^2\breve{\breve{\alpha}}_s+\frac{1}{NT\sqrt{T}}\breve{\kappa}_s\breve{\breve{\alpha}}_s+\frac{1}{N\sqrt{NT}}\breve{\delta}_s+\frac{1}{NT}\breve{\chi}_s-\tilde{\gamma}_{RP}\left[\frac{1}{T^2\sqrt{N}}\breve{\beta}_{1,s}+\frac{1}{NT^2\sqrt{T}}\breve{\beta}_{2,s}+\frac{1}{NT^2}\breve{\beta}_{3,s}\right]\Bigg\}\\
&+\frac{1}{T}\breve{\Pi}_4\left\{\frac{1}{\sqrt{NT}}\eta_t^2\breve{\xi}_t+\frac{1}{NT}\kappa_t\breve{\xi}_t+\frac{1}{N\sqrt{N}}\breve{\varphi}_t+\frac{1}{N\sqrt{T}}\breve{\gamma}_t+\tilde{\gamma}_{RP}\left[\frac{1}{T^2\sqrt{N}}\eta_t^2\breve{\xi}_t+\frac{1}{NT^2}\kappa_t\breve{\xi}_t+\frac{1}{NT\sqrt{T}}\breve{\alpha}_t^*\right]\right\}\\
&+\frac{1}{\sqrt{N}}\left(\frac{H'W^2\varepsilon\varepsilon'W}{NT\sqrt{T}}\right)\left(\frac{W\varepsilon\varepsilon'W^2H}{NT\sqrt{T}}\right)\xi_t+\frac{1}{NT\sqrt{T}}\breve{\Pi}_4\breve{\Pi}_1'\xi_t\Bigg\}. \qquad \text{(B.117)}
\end{aligned}$$

The eighth term in the r.h.s. of equation (B.76) can be written as:

$$\frac{1}{N^4T^3}(H\Lambda'+\varepsilon)(\varepsilon'W^2\varepsilon+\Lambda H'W^2\varepsilon)^2(\varepsilon'W^2\varepsilon\Lambda+\Lambda H'W^2\varepsilon\Lambda)M^{***}$$

$$= H\Bigg\{\frac{1}{N\sqrt{NT}}\left(\frac{\Lambda'\varepsilon'W^2\varepsilon}{NT}\right)\left(\frac{\varepsilon'W^2\varepsilon\Lambda}{NT}\right)+\frac{1}{N^2T}\left(\frac{\Lambda'\varepsilon'W^2\varepsilon\Lambda}{NT}\right)\left(\frac{H'W^2\varepsilon\Lambda}{\sqrt{NT}}\right)$$

$$+\frac{1}{NT}\left(\frac{\Lambda'\Lambda}{N}\right)\left(\frac{H'W^2\varepsilon\varepsilon'W^2\varepsilon\Lambda}{NT\sqrt{NT}}\right)+\frac{1}{NT\sqrt{NT}}\left(\frac{\Lambda'\Lambda}{N}\right)\left(\frac{H'W^2\varepsilon\Lambda}{\sqrt{NT}}\right)\left(\frac{H'W^2\varepsilon\Lambda}{\sqrt{NT}}\right)\Bigg\}\left(\frac{H'W^2\varepsilon\Lambda}{\sqrt{NT}}\right)M^{***}$$

$$+\Bigg\{\frac{1}{\sqrt{NT}}\left(\frac{\varepsilon\varepsilon'W^2\varepsilon\varepsilon'W^2\varepsilon\Lambda}{N^3T^2}\right)+\frac{1}{NT}\left(\frac{\varepsilon\varepsilon'W^2\varepsilon\Lambda}{N^2T}\right)\left(\frac{H'W^2\varepsilon\Lambda}{\sqrt{NT}}\right)$$

$$+\frac{1}{NT\sqrt{N}}\left(\frac{\varepsilon\Lambda}{\sqrt{N}}\right)\left(\frac{H'W^2\varepsilon\varepsilon'W^2\varepsilon\Lambda}{NT\sqrt{NT}}\right)+\frac{1}{N^2T\sqrt{T}}\left(\frac{\varepsilon\Lambda}{\sqrt{N}}\right)\left(\frac{H'W^2\varepsilon\Lambda}{\sqrt{NT}}\right)\left(\frac{H'W^2\varepsilon\Lambda}{\sqrt{NT}}\right)\Bigg\}\left(\frac{H'W^2\varepsilon\Lambda}{\sqrt{NT}}\right)M^{***}$$

$$+H\Bigg\{\frac{1}{N}\left(\frac{\Lambda'\varepsilon'W^2\varepsilon}{NT}\right)\left(\frac{\varepsilon'W^2\varepsilon}{NT}\right)+\frac{1}{N\sqrt{NT}}\left(\frac{\Lambda'\varepsilon'W^2\varepsilon\Lambda}{NT}\right)\left(\frac{H'W^2\varepsilon}{\sqrt{NT}}\right)$$

$$+\frac{1}{\sqrt{NT}}\left(\frac{\Lambda'\Lambda}{N}\right)\left(\frac{H'W^2\varepsilon\varepsilon'W^2\varepsilon}{NT\sqrt{NT}}\right)+\frac{1}{NT}\left(\frac{\Lambda'\Lambda}{N}\right)\left(\frac{H'W^2\varepsilon\Lambda}{\sqrt{NT}}\right)\left(\frac{H'W^2\varepsilon}{\sqrt{NT}}\right)\Bigg\}\left(\frac{\varepsilon'W^2\varepsilon\Lambda}{NT}\right)M^{***}$$

$$+\Bigg\{\frac{1}{\sqrt{NT}}\left(\frac{\varepsilon\varepsilon'W^2\varepsilon\Lambda}{N^2T}\right)\left(\frac{H'W^2\varepsilon}{\sqrt{NT}}\right)$$

$$+\frac{1}{N\sqrt{T}}\left(\frac{\varepsilon\Lambda}{\sqrt{N}}\right)\left(\frac{H'W^2\varepsilon}{NT}\right)\left(\frac{\varepsilon'W^2\varepsilon}{\sqrt{NT}}\right)+\frac{1}{NT\sqrt{N}}\left(\frac{\varepsilon\Lambda}{\sqrt{N}}\right)\left(\frac{H'W^2\varepsilon\Lambda}{\sqrt{NT}}\right)\left(\frac{H'W^2\varepsilon}{\sqrt{NT}}\right)\Bigg\}\left(\frac{\varepsilon'W^2\varepsilon\Lambda}{NT}\right)M^{***}$$

$$+\left(\frac{\varepsilon'W^2\varepsilon}{NT}\right)\left(\frac{\varepsilon\varepsilon'W^2\varepsilon\varepsilon'W^2\varepsilon\Lambda}{N^3T^2}\right)M^{***}, \tag{B.118}$$

which can be written as

$$r_{8,t} = H\Bigg\{\frac{1}{N\sqrt{NT}}\left(\frac{\Lambda'\varepsilon'W^2\varepsilon}{NT}\right)\left(\frac{\varepsilon'W^2\varepsilon\Lambda}{NT}\right)+\frac{1}{N^2T}\breve{\Pi}_2\breve{\Pi}_1+\frac{1}{NT}\left(\frac{\Lambda'\Lambda}{N}\right)\breve{\Pi}_3+\frac{1}{NT\sqrt{NT}}\left(\frac{\Lambda'\Lambda}{N}\right)\breve{\Pi}_1^2\Bigg\}\breve{\Pi}_1M^{***}$$

$$+\Bigg\{\frac{1}{\sqrt{NT}}\left(\frac{\varepsilon\varepsilon'W^2\varepsilon\varepsilon'W^2\varepsilon\Lambda}{N^3T^2}\right)+\frac{1}{NT}\left(\frac{\varepsilon\varepsilon'W^2\varepsilon\Lambda}{N^2T}\right)\breve{\Pi}_1+\frac{1}{NT\sqrt{N}}\left(\frac{\varepsilon\Lambda}{\sqrt{N}}\right)\breve{\Pi}_3+\frac{1}{N^2T\sqrt{T}}\left(\frac{\varepsilon\Lambda}{\sqrt{N}}\right)\breve{\Pi}_1^2\Bigg\}\breve{\Pi}_1M^{***}$$

$$+H\Bigg\{\frac{1}{N}\left(\frac{\Lambda'\varepsilon'W^2\varepsilon}{NT}\right)\left(\frac{\varepsilon'W^2\varepsilon}{NT}\right)+\frac{1}{N\sqrt{NT}}\breve{\Pi}_2\left(\frac{H'W^2\varepsilon}{\sqrt{NT}}\right)$$

$$+\frac{1}{\sqrt{NT}}\left(\frac{\Lambda'\Lambda}{N}\right)\breve{\Pi}_3+\frac{1}{NT}\left(\frac{\Lambda'\Lambda}{N}\right)\breve{\Pi}_1\left(\frac{H'W^2\varepsilon}{\sqrt{NT}}\right)\Bigg\}\left(\frac{\varepsilon'W^2\varepsilon\Lambda}{NT}\right)M^{***}$$

$$+\Bigg\{\frac{1}{\sqrt{NT}}\left(\frac{\varepsilon\varepsilon'W^2\varepsilon\Lambda}{N^2T}\right)\left(\frac{H'W^2\varepsilon}{\sqrt{NT}}\right)$$

$$+\frac{1}{N\sqrt{T}}\left(\frac{\varepsilon\Lambda}{\sqrt{N}}\right)\left(\frac{H'W^2\varepsilon}{NT}\right)\left(\frac{\varepsilon'W^2\varepsilon}{\sqrt{NT}}\right)+\frac{1}{NT\sqrt{N}}\left(\frac{\varepsilon\Lambda}{\sqrt{N}}\right)\breve{\Pi}_1\left(\frac{H'W^2\varepsilon}{\sqrt{NT}}\right)\Bigg\}\left(\frac{\varepsilon'W^2\varepsilon\Lambda}{NT}\right)M^{***}$$

$$+\left(\frac{\varepsilon'W^2\varepsilon}{NT}\right)\left(\frac{\varepsilon\varepsilon'W^2\varepsilon\varepsilon'W^2\varepsilon\Lambda}{N^3T^2}\right)M^{***}. \tag{B.119}$$

Therefore, the date t - element (B.119), which we denote ad $r_{8,t}$ is:

$$\begin{aligned}
r_{8,t} &= M^{***\prime}\breve{\Pi}_1'\left\{\frac{1}{N\sqrt{NT}}\left(\frac{\Lambda'\varepsilon'W^2\varepsilon}{NT}\right)\left(\frac{\varepsilon'W^2\varepsilon\Lambda}{NT}\right)+\frac{1}{N^2T}\breve{\Pi}_1'\breve{\Pi}_2+\frac{1}{NT}\breve{\Pi}_3'\left(\frac{\Lambda'\Lambda}{N}\right)+\frac{1}{NT\sqrt{NT}}(\breve{\Pi}_1')^2\left(\frac{\Lambda'\Lambda}{N}\right)\right\}h_t \\
&\quad +M^{***\prime}\breve{\Pi}_1'\left\{\frac{1}{\sqrt{NT}}\left[\frac{\varepsilon\varepsilon'W^2\varepsilon\varepsilon'W^2\varepsilon\Lambda}{N^3T^2}\right]_t+\frac{1}{NT}\breve{\Pi}_1\left[\frac{\varepsilon\varepsilon'W^2\varepsilon\Lambda}{N^2T}\right]_t+\frac{1}{NT\sqrt{N}}\breve{\Pi}_3'\xi_t+\frac{1}{N^2T\sqrt{T}}(\breve{\Pi}_1')^2\xi_t\right\} \\
&\quad +M^{***\prime}\left(\frac{\Lambda'\varepsilon'W^2\varepsilon}{NT}\right)\left\{\frac{1}{N}\left(\frac{\Lambda'\varepsilon'W^2\varepsilon}{NT}\right)\left(\frac{\varepsilon'W^2\varepsilon}{NT}\right)+\frac{1}{N\sqrt{NT}}\breve{\Pi}_2\left(\frac{H'W^2\varepsilon}{\sqrt{NT}}\right)\right. \\
&\quad \left.+\frac{1}{\sqrt{NT}}\left(\frac{\Lambda'\Lambda}{N}\right)\breve{\Pi}_3+\frac{1}{NT}\left(\frac{\Lambda'\Lambda}{N}\right)\breve{\Pi}_1\left(\frac{\varepsilon'W^2H}{\sqrt{NT}}\right)\right\}h_t \\
&\quad +M^{***\prime}\left(\frac{\Lambda'\varepsilon'W^2\varepsilon}{NT}\right)\left\{\frac{1}{\sqrt{NT}}\left(\frac{\varepsilon'W^2H}{\sqrt{NT}}\right)\left[\frac{\varepsilon\varepsilon'W^2\varepsilon\Lambda}{N^2T}\right]_t\right. \\
&\quad \left.+\frac{1}{N\sqrt{T}}\left(\frac{\varepsilon'W^2\varepsilon}{\sqrt{NT}}\right)\left(\frac{H'W^2\varepsilon}{NT}\right)'\xi_t+\frac{1}{NT\sqrt{N}}\left(\frac{\varepsilon'W^2H}{\sqrt{NT}}\right)\breve{\Pi}_1'\xi_t\right\} \\
&\quad +M^{***\prime}\frac{1}{NT}\sum_{s=1}^{T}\sum_{i=1}^{N}\breve{\varepsilon}_{it}\breve{\varepsilon}_{is}\left[\frac{\varepsilon\varepsilon'W^2\varepsilon\varepsilon'W^2\varepsilon\Lambda}{N^3T^2}\right]_s
\end{aligned} \tag{B.120}$$

which is equal to:

$$
\begin{aligned}
r_{8,t} &= M^{***\prime}\breve{\Pi}_1'\Bigg\{\frac{1}{N\sqrt{NT}}\left(\frac{\Lambda'\varepsilon'W^2\varepsilon}{NT}\right)\left(\frac{\varepsilon'W^2\varepsilon\Lambda}{NT}\right)+\frac{1}{N^2T}\breve{\Pi}_1'\breve{\Pi}_2+\frac{1}{NT}\breve{\Pi}_3'\left(\frac{\Lambda'\Lambda}{N}\right)+\frac{1}{NT\sqrt{NT}}(\breve{\Pi}_1')^2\left(\frac{\Lambda'\Lambda}{N}\right)\Bigg\}h_t \\
&+M^{***\prime}\breve{\Pi}_1'\Bigg\{\frac{1}{\sqrt{NT}}\left[\frac{1}{NT\sqrt{T}}\breve{\beta}_{4,t}+\frac{1}{NT\sqrt{NT}}\breve{\beta}_{5,t}+\frac{1}{N^2\sqrt{T}}\breve{\beta}_{6,t}+\frac{1}{NT\sqrt{N}}\breve{\beta}_{7,t}\right] \\
&+\frac{1}{NT}\breve{\Pi}_1\left\{\frac{1}{\sqrt{NT}}\eta_t^2\breve{\xi}_t+\frac{1}{NT}\kappa_t\breve{\xi}_t+\frac{1}{N\sqrt{N}}\breve{\varphi}_t+\frac{1}{N\sqrt{T}}\breve{\gamma}_t+\tilde{\gamma}_{RP}\left[\frac{1}{T^2\sqrt{N}}\eta_t^2\breve{\xi}_t+\frac{1}{NT^2}\kappa_t\breve{\xi}_t+\frac{1}{NT\sqrt{T}}\breve{\alpha}_t^*\right]\right\} \\
&+\frac{1}{NT\sqrt{N}}\breve{\Pi}_3'\xi_t+\frac{1}{N^2T\sqrt{T}}(\breve{\Pi}_1')^2\xi_t\Bigg\} \\
&+M^{***\prime}\left(\frac{\Lambda'\varepsilon'W^2\varepsilon}{NT}\right)\Bigg\{\frac{1}{N}\left(\frac{\Lambda'\varepsilon'W^2\varepsilon}{NT}\right)\left(\frac{\varepsilon'W^2\varepsilon}{NT}\right)+\frac{1}{N\sqrt{NT}}\breve{\Pi}_2\left(\frac{H'W^2\varepsilon}{\sqrt{NT}}\right) \\
&+\frac{1}{\sqrt{NT}}\left(\frac{\Lambda'\Lambda}{N}\right)\breve{\Pi}_3+\frac{1}{NT}\left(\frac{\Lambda'\Lambda}{N}\right)\breve{\Pi}_1\left(\frac{\varepsilon'W^2H}{\sqrt{NT}}\right)\Bigg\}h_t \\
&+M^{***\prime}\left(\frac{\Lambda'\varepsilon'W^2\varepsilon}{NT}\right)\Bigg\{ \\
&\frac{1}{\sqrt{NT}}\left(\frac{\varepsilon'W^2H}{\sqrt{NT}}\right)\left\{\frac{1}{\sqrt{NT}}\eta_t^2\breve{\xi}_t+\frac{1}{NT}\kappa_t\breve{\xi}_t+\frac{1}{N\sqrt{N}}\breve{\varphi}_t+\frac{1}{N\sqrt{T}}\breve{\gamma}_t+\tilde{\gamma}_{RP}\left[\frac{1}{T^2\sqrt{N}}\eta_t^2\breve{\xi}_t+\frac{1}{NT^2}\kappa_t\breve{\xi}_t+\frac{1}{NT\sqrt{T}}\breve{\alpha}_t^*\right]\right\} \\
&+\frac{1}{N\sqrt{T}}\left(\frac{\varepsilon'W^2\varepsilon}{\sqrt{NT}}\right)\left(\frac{H'W^2\varepsilon}{NT}\right)'\xi_t+\frac{1}{NT\sqrt{N}}\left(\frac{\varepsilon'W^2H}{\sqrt{NT}}\right)\breve{\Pi}_1'\xi_t\Bigg\} \\
&+M^{***\prime}\frac{1}{NT}\sum_{s=1}^{T}\sum_{i=1}^{N}\breve{\varepsilon}_{it}\breve{\varepsilon}_{is}\left[\frac{1}{NT\sqrt{T}}\breve{\beta}_{4,s}+\frac{1}{NT\sqrt{NT}}\breve{\beta}_{5,s}+\frac{1}{N^2\sqrt{T}}\breve{\beta}_{6,s}+\frac{1}{NT\sqrt{N}}\breve{\beta}_{7,s}\right]
\end{aligned}
\tag{B.121}
$$

Therefore, the expansion in (B.76) for date t reads:

$$
\begin{aligned}
(\tilde{\mathscr{H}}')^{-1}\hat{h}_t - h_t \quad = & \quad \frac{1}{\sqrt{N}}(\Lambda'\Lambda/N)^{-1}\xi_t \; + \; M^{*\prime}\frac{1}{\sqrt{NT}}\breve{\Pi}_1 h_t \\
+ & \quad \frac{1}{T}M^{*\prime}\eta_t^2\breve{h}_t + \frac{1}{T\sqrt{N}}M^{*\prime}\kappa_t\breve{h}_t + \frac{1}{\sqrt{NT}}M^{*\prime}\breve{\alpha}_t + \tilde{\gamma}_{RP}M^{*\prime}\left[\frac{1}{T^2}\eta_t^2\breve{h}_t + \frac{1}{T\sqrt{NT}}\breve{\alpha}_t^* + \frac{1}{T^3\sqrt{N}}\kappa_t\breve{h}_t\right] \\
+ & \quad M^{*\prime}\frac{1}{N}\breve{\Pi}_2 h_t \; + \; \frac{1}{\sqrt{NT}}M^{*\prime}\breve{\Pi}_1'\left(\frac{\Lambda'\Lambda}{N}\right)h_t \\
+ & \quad M^{*\prime}\left\{\frac{1}{\sqrt{NT}}\eta_t^2\breve{\xi}_t + \frac{1}{NT}\kappa_t\breve{\xi}_t + \frac{1}{N\sqrt{N}}\breve{\varphi}_t + \frac{1}{N\sqrt{T}}\breve{\gamma}_t + \tilde{\gamma}_{RP}\left[\frac{1}{T^2\sqrt{N}}\eta_t^2\breve{\xi}_t + \frac{1}{NT^2}\kappa_t\breve{\xi}_t + \frac{1}{NT\sqrt{T}}\breve{\alpha}_t^*\right]\right\} \\
+ & \quad \frac{1}{N\sqrt{T}}\; M^{*\prime}\breve{\Pi}_1'\;\xi_t \\
+ & \quad \frac{1}{\sqrt{NT}}M^{*\prime}(H'W^2H/T)^{-1}\breve{\Pi}_3 h_t \; + \; M^{*\prime}(H'W^2H/T)^{-1}\frac{1}{T}\breve{\Pi}_4\left(\frac{\Lambda'\Lambda}{N}\right)h_t \\
+ & \quad M^{*\prime}\left\{\frac{1}{T^2}\breve{\eta}_t^4\breve{h}_t + \frac{1}{\sqrt{N}T^2}\breve{\kappa}_t\breve{\eta}_t^2\breve{h}_t + \frac{1}{\sqrt{NT}T}\breve{\breve{\alpha}}_t\right. \\
+ & \quad \frac{1}{T^2\sqrt{N}}\breve{\kappa}_t\breve{\eta}_t^2\breve{h}_t + \frac{1}{NT^2}\breve{\kappa}_t^2\breve{h}_t + \frac{1}{NT\sqrt{N}}\breve{\breve{\varphi}}_t + \frac{1}{NT\sqrt{T}}\breve{\breve{\gamma}}_t \\
+ & \quad \left. + \frac{1}{T\sqrt{NT}}\breve{\eta}_t^2\breve{\breve{\alpha}}_t + \frac{1}{NT\sqrt{T}}\breve{\kappa}_t\breve{\breve{\alpha}}_t + \frac{1}{N\sqrt{NT}}\breve{\delta}_t + \frac{1}{NT}\breve{\chi}_t - \tilde{\gamma}_{RP}\left[\frac{1}{T^2\sqrt{N}}\breve{\beta}_{1,t} + \frac{1}{NT^2\sqrt{T}}\breve{\beta}_{2,t} + \frac{1}{NT^2}\breve{\beta}_{3,t}\right]\right\} \\
+ & \quad \frac{1}{T\sqrt{N}}M^{*}(H'W^2H/T)^{-1}\breve{\Pi}_4\xi_t \; + \; M^{**\prime}\left\{\frac{1}{NT}\left(\frac{1}{T}\sum_{t=1}^{T}\breve{\eta}_t^2\breve{\xi}_t\breve{\xi}_t'\right) + \frac{1}{NT\sqrt{N}}\left(\frac{1}{T}\sum_{t=1}^{T}\breve{\kappa}_t\breve{\xi}_t\breve{\xi}_t'\right)\right. \\
+ & \quad \left.\frac{1}{N^2}\left(\frac{1}{T}\sum_{t=1}^{T}\breve{\xi}_t\breve{\varphi}_t'\right)' + \frac{1}{N\sqrt{TN}}\left(\frac{1}{T}\sum_{t=1}^{T}\breve{\xi}_t\breve{\gamma}_t\right)'\right\}h_t \\
+ & \quad \frac{1}{N\sqrt{NT}}M^{**\prime}\breve{\Pi}_1'\left(\frac{1}{T}\sum_{t=1}^{T}\breve{\xi}_t\breve{\xi}_t'\right)h_t \; + \; \frac{1}{\sqrt{NT}}M^{**\prime}\breve{\Pi}_3'\left(\frac{\Lambda'\Lambda}{N}\right)h_t \; + \; \frac{1}{NT}M^{**\prime}\breve{\Pi}_1'\breve{\Pi}_1'\left(\frac{\Lambda'\Lambda}{N}\right)h_t \\
+ & \quad M^{**\prime}\left[\frac{1}{NT\sqrt{T}}\breve{\beta}_{4,t} + \frac{1}{NT\sqrt{NT}}\breve{\beta}_{5,t} + \frac{1}{N^2\sqrt{T}}\breve{\beta}_{6,t} + \frac{1}{NT\sqrt{N}}\breve{\beta}_{7,t}\right] \\
+ & \quad \frac{1}{N\sqrt{NT}}M^{**\prime}\breve{\Pi}_1'\left\{\frac{1}{\sqrt{NT}}\eta_t^2\breve{\xi}_t + \frac{1}{NT}\kappa_t\breve{\xi}_t + \frac{1}{N\sqrt{N}}\breve{\varphi}_t + \frac{1}{N\sqrt{T}}\breve{\gamma}_t\right. \\
+ & \quad \left.\tilde{\gamma}_{RP}\left[\frac{1}{T^2\sqrt{N}}\eta_t^2\breve{\xi}_t + \frac{1}{NT^2}\kappa_t\breve{\xi}_t + \frac{1}{NT\sqrt{T}}\breve{\alpha}_t^*\right]\right\} \\
+ & \quad \frac{1}{N\sqrt{T}}M^{**\prime}\breve{\Pi}_3'\breve{\xi}_t \; + \; \frac{1}{NT\sqrt{N}}M^{**\prime}\breve{\Pi}_1'\breve{\Pi}_1'\breve{\xi}_t \; + \; r_{7,t} \; + \; r_{8,t} \; + \; \breve{\mathcal{R}}_t\,. \qquad \text{(B.122)}
\end{aligned}
$$

By rearranging and decomposing the terms of order $1/\sqrt{N}$, $1/\sqrt{NT}$, $1/T$, and $1/N$ in the last equations, the expansion in (B.49) follows.

Let us now prove the convergence of matrices $\hat{\mathscr{H}}$ and $\tilde{\mathscr{H}}$. From Proposition B.4 and Assumption B.10, we have $o_p(1) = \Lambda'(\hat{\Lambda} - \Lambda\hat{\mathscr{H}})/N = (\Lambda'\hat{\Lambda}/N) - \Sigma_\lambda\hat{\mathscr{H}} + o_p(1)$, which implies:

$$(\Lambda'\hat{\Lambda}/N) = \Sigma_\lambda\hat{\mathscr{H}} + o_p(1). \qquad \text{(B.123)}$$

By combining equations (B.59) and (B.123), and using Lemma B.11 and Assumption B.10, we get:

$$\Sigma_h(W^2)\Sigma_\lambda\hat{\mathscr{H}} = \hat{\mathscr{H}}V + o_p(1). \tag{B.124}$$

The definition of matrix $\tilde{\mathscr{H}}$ and Assumption B.10 imply:

$$\tilde{\mathscr{H}} \;=\; \Sigma_\lambda\hat{\mathscr{H}} + o_p(1)\;, \tag{B.125}$$

which combined with equation (B.124) implies:

$$\Sigma_\lambda\Sigma_h(W^2)\tilde{\mathscr{H}} = \tilde{\mathscr{H}}V + o_p(1), \tag{B.126}$$

and

$$\Sigma_\lambda^{1/2}\Sigma_h(W^2)\Sigma_\lambda^{1/2}(\Sigma_\lambda^{-1/2}\tilde{\mathscr{H}}) = (\Sigma_\lambda^{-1/2}\tilde{\mathscr{H}})V + o_p(1). \tag{B.127}$$

The last equation shows that $\Sigma_\lambda^{-1/2}\tilde{\mathscr{H}}$ are the eigenvectors of $\Sigma_\lambda^{1/2}\Sigma_h(W^2)\Sigma_\lambda^{1/2}$.

Moreover, equation (B.125) implies:

$$\Sigma_\lambda^{-1/2}\tilde{\mathscr{H}} \;=\; \Sigma_\lambda^{1/2}\hat{\mathscr{H}} + o_p(1)\;. \tag{B.128}$$

By substituting the quality $\hat{\Lambda} = (\hat{\Lambda} - \Lambda\hat{\mathscr{H}}) + \Lambda\hat{\mathscr{H}}$ into the RP-PCA loadings constraint $\hat{\Lambda}'\hat{\Lambda}/N = I_k$, Assumption B.10, Proposition B.4 and equation (B.123), we get:

$$\hat{\mathscr{H}}'\Sigma_\lambda\hat{\mathscr{H}} = \hat{\mathscr{H}}'\Sigma_\lambda^{1/2}\Sigma_\lambda^{1/2}\hat{\mathscr{H}} = I_k + o_p(1). \tag{B.129}$$

The last equation, combined with (B.128), implies also:

$$\tilde{\mathscr{H}}'\Sigma_\lambda^{-1/2}\Sigma_\lambda^{-1/2}\tilde{\mathscr{H}} = I_k + o_p(1). \tag{B.130}$$

Recall that V is the diagonal matrix with diagonal elements corresponding to the eigenvalues of the symmetric matrix $\Sigma_\lambda^{1/2}\Sigma_h(W^2)\Sigma_\lambda^{1/2}$. Then, if these eigenvalues are distinct, equations (B.127) and (B.130) imply that the columns of matrix $\Sigma_\lambda^{-1/2}\tilde{\mathscr{H}}$ converge in probability to the orthonormal eigenvectors of matrix $\Sigma_\lambda^{1/2}\Sigma_h(W)\Sigma_\lambda^{1/2}$. Proposition B.3 follows. ■

In the remaining part of this subsection we provides the proofs of all the Lemmas and checks of the Assumptions.

B.5.1.2 Proof of Proposition B.4

By computing the norms of both sides of equation (B.66), using the triangular inequality and the Cauchy-Schwarz inequality, Lemmas B.10 and B.11, and Assumption B.10, we get:

$$\begin{aligned}\|\hat{\Lambda}\hat{\mathscr{H}}^{-1} - \Lambda\| \;&=\; O_p\left(\|\frac{1}{NT}\varepsilon'W^2\varepsilon\Lambda\| + \|\frac{1}{NT}\Lambda H'W^2\varepsilon\Lambda\| + \|\frac{1}{T}\varepsilon'W^2H\|\right)\\ &\quad+O_p\left[\left(\|\frac{1}{NT}\varepsilon'W^2\varepsilon\| + \|\frac{1}{NT}\Lambda H'W^2\varepsilon\|\right)\|\hat{\Lambda}\hat{\mathscr{H}}^{-1} - \Lambda\|\right],\end{aligned} \tag{B.131}$$

To control the term in the r.h.s. we use the first three results in next lemma.

LEMMA B.12. *Under Assumptions B.10 and B.12, we have: (i)* $\|\frac{1}{T}\varepsilon' W^2 H\| = O_p\left(\sqrt{\frac{N}{T}}\right)$*,*
(ii) $\|\frac{1}{NT}\Lambda H' W^2 \varepsilon\| = O_p\left(\frac{1}{\sqrt{T}}\right)$ *and (iii)* $\|\frac{1}{NT}\Lambda H' W^2 \varepsilon \Lambda\| = O_p\left(\frac{1}{\sqrt{T}}\right)$*. Moreover, we also have:*
(iv) $\|\frac{1}{N}\breve{\varepsilon}\Lambda\| = O_p\left(\sqrt{\frac{T}{N}}\right)$*, which implies* $\|\frac{1}{N}\varepsilon\Lambda\| = O_p\left(\sqrt{\frac{T}{N}}\right)$*.*
(v) $\|\frac{1}{NT}H\Lambda'\varepsilon'\| = O_p\left(\frac{1}{\sqrt{N}}\right)$ *and (vi)* $\|\frac{1}{NT}H\Lambda'\varepsilon' H\| = O_p\left(\frac{1}{\sqrt{N}}\right)$*.*
(check if and where the last 2 results are used).

By multiplying both sides of equation (B.131) times $N^{-1/2}$, and using Assumption B.12 ii)-iii) and Lemma B.12, we get:

$$\begin{aligned} N^{-1/2}\|\hat{\Lambda}\hat{\mathscr{H}}^{-1} - \Lambda\| &= O_p\left(\frac{1}{\sqrt{Nm}} + \frac{1}{\sqrt{NT}} + \frac{1}{\sqrt{N}}\frac{\sqrt{N}}{\sqrt{T}}\right) + O_p\left(\frac{1}{\sqrt{m}} + \frac{1}{\sqrt{T}}\right) N^{-1/2}\|\hat{\Lambda}\tilde{\mathscr{H}}^{-1} - \Lambda\| \\ &= O_p\left(\frac{1}{\sqrt{T}} + \frac{1}{\sqrt{NT}} + \frac{1}{\sqrt{Nm}}\right) + o_p(N^{-1/2}\|\hat{\Lambda}\hat{\mathscr{H}}^{-1} - \Lambda\|), \end{aligned}$$

where $m = \min\{N, T\}$. The last equation simplifies as:

$$N^{-1/2}\|\hat{\Lambda}\hat{\mathscr{H}}^{-1} - \Lambda\| = O_p\left(\frac{1}{\sqrt{T}}\right), \tag{B.132}$$

using that $\frac{1}{\sqrt{T}} > \frac{1}{\sqrt{Nm}}$, which is implied by Assumption B.1. This last result shows part i) of Proposition B.4.

By plugging in equation (B.66) into equation (B.74) we get:

$$\begin{aligned} &\hat{H}\tilde{\mathscr{H}}^{-1} - H \\ &= \frac{1}{N}\varepsilon\Lambda(\Lambda'\Lambda/N)^{-1} \\ &+ \frac{1}{N}(H\Lambda' + \varepsilon)\left\{\frac{1}{NT}(\varepsilon' W^2 \varepsilon\Lambda + \Lambda H' W^2 \varepsilon\Lambda)\hat{\mathscr{H}}(\Lambda'\hat{\Lambda}/N)^{-1}(H' W^2 H/T)^{-1}\right\}(\Lambda'\Lambda/N)^{-1} \\ &+ \frac{1}{N}(H\Lambda' + \varepsilon)\left\{\frac{1}{T}\varepsilon' W^2 H(H' W^2 H/T)^{-1}\right\}(\Lambda'\Lambda/N)^{-1} \\ &+ \frac{1}{N}(H\Lambda' + \varepsilon)\left\{\frac{1}{NT}(\varepsilon' W^2 \varepsilon + \Lambda H' W^2 \varepsilon)[\hat{\Lambda}\hat{\mathscr{H}}^{-1} - \Lambda]\hat{\mathscr{H}}(\Lambda'\hat{\Lambda}/N)^{-1}(H' W^2 H/T)^{-1}\right\}(\Lambda'\Lambda/N)^{-1} \end{aligned} \tag{B.133}$$

By computing the norms of both sides of the last equation, pre-multiplying by $1/\sqrt{T}$, using the triangular inequality and the Cauchy-Schwarz inequality, Lemmas B.10 and B.11, and Assumption B.10, we get:

$$
\begin{aligned}
&\frac{1}{\sqrt{T}}\|\hat{H}\tilde{\mathscr{H}}^{-1}-H\| \\
&= \frac{1}{\sqrt{T}}\left\|\frac{1}{N}\varepsilon\Lambda\right\|\cdot O_p(1)+\left\|\frac{1}{N^2T}H\Lambda'\varepsilon'W^2\varepsilon\Lambda\right\|\cdot O_p(1)+\left\|\frac{1}{N^2T}\varepsilon\varepsilon'W^2\varepsilon\Lambda\right\|\cdot O_p(1) \\
&+\frac{1}{\sqrt{T}}\left\|\frac{1}{N^2T}H\Lambda'\Lambda H'W^2\varepsilon\Lambda\right\|\cdot O_p(1)+\frac{1}{\sqrt{T}}\left\|\frac{1}{N^2T}\varepsilon\Lambda H'W^2\varepsilon\Lambda\right\|\cdot O_p(1) \\
&+\frac{1}{\sqrt{T}}\left\|\frac{1}{NT}H\Lambda'\varepsilon'W^2H\right\|\cdot O_p(1)+\frac{1}{\sqrt{T}}\left\|\frac{1}{NT}\varepsilon\varepsilon'W^2H\right\|\cdot O_p(1) \\
&+\sqrt{\frac{N}{T}}\left\|\frac{1}{N^2T}H\Lambda'\varepsilon'W^2\varepsilon\right\|\frac{1}{\sqrt{N}}\|\hat{\Lambda}\hat{\mathscr{H}}^{-1}-\Lambda\|O_p(1)+\sqrt{\frac{N}{T}}\left\|\frac{1}{N^2T}\varepsilon\varepsilon'W^2\varepsilon\right\|\frac{1}{\sqrt{N}}\|\hat{\Lambda}\hat{\mathscr{H}}^{-1}-\Lambda\|O_p(1) \\
&+\sqrt{\frac{N}{T}}\left\|\frac{1}{N^2T}H\Lambda'\Lambda H'W^2\varepsilon\Lambda\right\|\frac{1}{\sqrt{N}}\|\hat{\Lambda}\hat{\mathscr{H}}^{-1}-\Lambda\|O_p(1)+\sqrt{\frac{N}{T}}\left\|\frac{1}{N^2T}\varepsilon\Lambda'W^2\varepsilon\Lambda\right\|\frac{1}{\sqrt{N}}\|\hat{\Lambda}\hat{\mathscr{H}}^{-1}-\Lambda\|O_p(1) \\
&= \frac{1}{\sqrt{T}}\left\|\frac{1}{N}\varepsilon\Lambda\right\|\cdot O_p(1)+\frac{1}{T}\left\|\frac{1}{\sqrt{T}}H\right\|\left\|\frac{1}{N}\breve{\varepsilon}\Lambda\right\|^2\cdot O_p(1)+\frac{1}{\sqrt{N}}\left\|\frac{1}{\sqrt{NT}}\varepsilon\right\|\left\|\frac{1}{NT}\varepsilon'W^2\varepsilon\Lambda\right\|\cdot O_p(1) \\
&+\frac{1}{\sqrt{N}}\left\|\frac{1}{\sqrt{NT}}H\Lambda'\right\|\left\|\frac{1}{NT}\Lambda H'W^2\varepsilon\Lambda\right\|\cdot O_p(1)+\frac{1}{\sqrt{N}}\left\|\frac{1}{\sqrt{NT}}\varepsilon\right\|\left\|\frac{1}{NT}\Lambda H'W^2\varepsilon\Lambda\right\|\cdot O_p(1) \\
&+\frac{1}{\sqrt{NT}}\left\|\frac{1}{\sqrt{T}}H\right\|\left\|\frac{1}{\sqrt{NT}}\Lambda'\varepsilon'W^2H\right\|\cdot O_p(1)+\frac{1}{\sqrt{T}}\left\|\frac{1}{NT}\varepsilon\varepsilon'W^2H\right\|\cdot O_p(1) \\
&+\frac{\sqrt{N}}{N}\left\|\frac{1}{\sqrt{T}}H\right\|\left\|\frac{1}{NT}\Lambda'\varepsilon'W^2\varepsilon\right\|O_P\left(\frac{1}{\sqrt{T}}\right)+\sqrt{\frac{N}{N}}\left\|\frac{1}{\sqrt{NT}}\varepsilon\right\|\left\|\frac{1}{NT}\varepsilon'W^2\varepsilon\right\|O_P\left(\frac{1}{\sqrt{T}}\right) \\
&+\sqrt{\frac{N}{N}}\left\|\frac{1}{\sqrt{NT}}H\Lambda'\right\|\left\|\frac{1}{NT}{}'\Lambda H'W^2\varepsilon\Lambda\right\|O_P\left(\frac{1}{\sqrt{T}}\right) \\
&+\sqrt{\frac{N}{T}}\left\|\frac{1}{\sqrt{NT}}\varepsilon\right\|\left\|\frac{1}{NT}\Lambda H'W^2\varepsilon\Lambda\right\|O_P\left(\frac{1}{\sqrt{T}}\right)\ . \qquad \text{(B.134)}
\end{aligned}
$$

Therefore, by using Assumptions B.10, B.12 ii), iii), iv) and Lemma B.12 ii), iv), we get:

$$
\begin{aligned}
\frac{1}{\sqrt{T}}\|\hat{H}\tilde{\mathscr{H}}^{-1}-H\| &= O_P\left(\frac{1}{\sqrt{N}}\right)+O_P\left(\frac{1}{N}\right)+O_P\left(\frac{1}{\sqrt{Nm}}\right)+O_P\left(\frac{1}{\sqrt{NT}}\right)+O_P\left(\frac{1}{\sqrt{Tm}}\right) \\
&\quad +O_P\left(\frac{1}{T}\right)+O_P\left(\frac{1}{\sqrt{NTm}}\right)+O_P\left(\frac{1}{T\sqrt{N}}\right) \\
&= O_P\left(\frac{1}{\sqrt{N}}\right)+O_P\left(\frac{1}{T}\right)+O_P\left(\frac{1}{\sqrt{Tm}}\right)=O_P\left(\frac{1}{\sqrt{N}}\right),
\end{aligned}
$$

where the last two equalities follow from Assumption B.1. Therefore we get:

$$
T^{-1/2}\|\hat{H}\tilde{\mathscr{H}}^{-1}-H\| = O_p\left(\frac{1}{\sqrt{N}}\right). \qquad \text{(B.135)}
$$

The reminder term $\hat{\mathcal{R}}$ is controlled by first noting that, from its definition in equation (B.54), we have

$$\frac{1}{\sqrt{T}}\hat{\mathcal{R}} \;=\; \frac{1}{\sqrt{NT}}(H\Lambda' + \varepsilon)\left[\frac{1}{NT}(\varepsilon' W^2 \varepsilon + \Lambda H' W^2 \varepsilon)\right]^3 \frac{1}{\sqrt{N}}(\hat{\Lambda}\mathscr{H}^{-1} - \Lambda)\left[\breve{S}(I_k + \hat{B})\right]^3 \left(\frac{\Lambda'\Lambda}{N}\right)^{-1}.$$

and that from equations and (B.135), Assumptions B.10 and B.12 ii), and Lemmas B.10, B.11 and B.12 ii), and the Cauchy-Schwarz inequality we have:

$$\begin{aligned} T^{-1/2}\|\hat{\mathcal{R}}\| \;&=\; O_p\left[\left(\left\|\frac{H\Lambda}{\sqrt{NT}}\right\| + \left\|\frac{\varepsilon}{\sqrt{NT}}\right\|\right)\left(\left\|\frac{\varepsilon' W^2 \varepsilon}{NT}\right\|^3 + \left\|\frac{\Lambda H' W^2 \varepsilon}{NT}\right\|^3\right)\left(\frac{1}{\sqrt{N}}\|\hat{\Lambda}\tilde{\mathscr{H}}^{-1} - \Lambda\|\right)\right] \\ &=\; O_p\left[\left(\frac{1}{m\sqrt{m}} + \frac{1}{T\sqrt{T}}\right)\frac{1}{\sqrt{T}}\right] \;=\; O_p\left(\frac{1}{m\sqrt{Tm}}\right). \end{aligned}$$

This last result concludes the proofs of Proposition B.4. ■

B.5.1.3 Proof of Proposition B.5

Let us first establish the MSE bound for remainder term r_t, which can be computed by first noting that, from its definition we have:

$$\frac{1}{T}\sum_{t=1}^{T}\|r_t\|^2 \le \frac{2}{T}\sum_{t=1}^{T}\left(\|r_{1,t}\|^2 + \|r_{7,t}\|^2 + \|r_{8,t}\|^2 + \|\hat{\mathcal{R}}_t\|^2\right). \tag{B.136}$$

Moreover, from the definition of $r_{1,t}$ in (B.50) and Assumption B.13 we have

$$\begin{aligned} \left(\frac{1}{T}\sum_{t=1}^{T}\|r_{1,t}\|^2\right)^{1/2} \;&=\; O_p\left[\|\hat{B}\|\left(\frac{1}{T} + \frac{1}{N} + \frac{1}{\sqrt{NT}}\right)\right] + O_p\left(\frac{1}{N\sqrt{T}} + \frac{1}{T\sqrt{N}} + \frac{1}{T^2} + \frac{1}{N\sqrt{N}}\right) \\ &=\; O_p\left[\frac{1}{\sqrt{T}}\left(\frac{1}{T} + \frac{1}{N} + \frac{1}{\sqrt{NT}}\right)\right] + O_p\left(\frac{1}{N\sqrt{T}} + \frac{1}{T\sqrt{N}} + \frac{1}{T^2} + \frac{1}{N\sqrt{N}}\right) \\ &=\; O_p\left(\frac{1}{N\sqrt{T}} + \frac{1}{T\sqrt{N}} + \frac{1}{T\sqrt{T}} + \frac{1}{N\sqrt{N}}\right). \end{aligned} \tag{B.137}$$

From Assumption ... also have:

$$\left(\frac{1}{T}\sum_{t=1}^{T}\|r_{7,t}\|^2\right)^{1/2} \;=\; ...., \tag{B.138}$$

$$\left(\frac{1}{T}\sum_{t=1}^{T}\|r_{8,t}\|^2\right)^{1/2} \;=\; ...., \tag{B.139}$$

and

$$\left(\frac{1}{T}\sum_{t=1}^{T}\|\hat{\mathcal{R}}_t\|^2\right)^{1/2} = O_p\left(\frac{1}{m\sqrt{Tm}}\right), \tag{B.140}$$

where the last equality follows directly from bound (B.60). Combining the last four results, and by using Assumption B.1 we get:

$$\left(\frac{1}{T}\sum_{t=1}^{T}\|r_{1,t}\|^2\right)^{1/2} = O_p\left(\frac{1}{N\sqrt{T}}+\frac{1}{T\sqrt{N}}+\frac{1}{T\sqrt{T}}+\frac{1}{N\sqrt{N}}\right). \tag{B.141}$$

Let us now show that $\frac{1}{T}\sum_{t=1}^{T}\vartheta_t h_t' = O_p(\frac{1}{N}+\frac{1}{T\sqrt{T}})$. We use $\breve{\vartheta}_t = \tilde{\vartheta}_t + \hat{\mathcal{R}}_t$ where

$$\tilde{\vartheta}_t = \left(\frac{\Lambda'\Lambda}{N}\right)^{-1}\breve{S}'\left(\frac{1}{\sqrt{NT}}\alpha_t + \frac{1}{N}\breve{\Pi}_2 h_t\right) + r_t.$$

From the Cauchy-Schwarz inequality and the bound in (B.60), we have:

$$\frac{1}{T}\sum_{t=1}^{T}\breve{\vartheta}_t h_t' = \frac{1}{T}\sum_{t=1}^{T}\tilde{\vartheta}_t h_t' + \frac{1}{T}\sum_{t=1}^{T}\hat{\mathcal{R}}_t h_t' = \frac{1}{T}\sum_{t=1}^{T}\tilde{\vartheta}_t h_t' + O_p\left(\frac{1}{m\sqrt{Tm}}\right).$$

Moreover, by using Assumption B.13, bound (B.141) and $\frac{1}{T\sqrt{N}} = o(\frac{1}{N})$ (which follows from Assumption B.1), we have:

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T}\tilde{\vartheta}_t h_t' &= O_p\left[\frac{1}{\sqrt{NT}}\breve{S}'\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\alpha_t h_t'\right)\right] + O_p\left[\frac{1}{N}\breve{S}'\breve{\Pi}_2\left(\frac{1}{T}\sum_{t=1}^{T}h_t h_t'\right)\right] + O_p\left(\frac{1}{N\sqrt{T}}+\frac{1}{T\sqrt{N}}+\frac{1}{T\sqrt{T}}+\frac{1}{N\sqrt{N}}\right)\\ &= O_p\left(\frac{1}{N}+\frac{1}{T\sqrt{T}}\right).\end{aligned}$$

Then, $\frac{1}{T}\sum_{t=1}^{T}\breve{\vartheta}_t h_t' = O_p\left(\frac{1}{N}+\frac{1}{T\sqrt{T}}\right)$ follows.

Let us finally show that $\frac{1}{T}\sum_{t=1}^{T}(\frac{1}{\sqrt{N}}u_t + \frac{1}{T}\breve{b}_t + \frac{1}{\sqrt{NT}}\breve{d}_t + \breve{\vartheta}_t)\breve{\vartheta}_t' = o_p(\frac{1}{N\sqrt{T}})$. We have:

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T}(\frac{1}{\sqrt{N}}u_t + \frac{1}{T}\breve{b}_t + \frac{1}{\sqrt{NT}}\breve{d}_t + \breve{\vartheta}_t)\breve{\vartheta}_t' &= \frac{1}{T}\sum_{t=1}^{T}(\frac{1}{\sqrt{N}}u_t + \frac{1}{T}\breve{b}_t + \frac{1}{\sqrt{NT}}\breve{d}_t + \tilde{\vartheta}_t)\tilde{\vartheta}_t'\\ &+\frac{1}{T}\sum_{t=1}^{T}\hat{\mathcal{R}}_t\tilde{\vartheta}_t' + \frac{1}{T}\sum_{t=1}^{T}[(\hat{\mathscr{H}}')^{-1}\hat{h}_t - h_t]\hat{\mathcal{R}}_t'.\end{aligned}$$

Moreover, by using bound (B.141) and Assumption B.13:

$$(\frac{1}{T}\sum_{t=1}^{T}\|\tilde{\vartheta}_t\|^2)^{1/2} = O_p\left(\frac{1}{N}+\frac{1}{T\sqrt{T}}+\frac{1}{\sqrt{NT}}\right) = O_p\left(\frac{1}{N}+\frac{1}{\sqrt{NT}}\right), \tag{B.142}$$

where the last equation follows from Assumption B.1. Thus, from (B.60) and $\sqrt{T} \ll N \ll T^2$ (Assumption B.1), we get:

$$\begin{aligned} \frac{1}{T}\sum_{t=1}^{T}\hat{\mathcal{R}}_t\tilde{\vartheta}_t' &= O_p\left[T^{-1/2}\|\hat{\mathcal{R}}\|(\frac{1}{T}\sum_{t=1}^{T}\|\tilde{\vartheta}_t\|^2)^{1/2}\right] \\ &= O_p\left[\frac{1}{m\sqrt{Tm}}\left(\frac{1}{N}+\frac{1}{\sqrt{NT}}\right)\right] = o_p(\frac{1}{N\sqrt{T}}). \end{aligned}$$

Further, from Proposition B.4 ii) bound (B.60)

$$\frac{1}{T}\sum_{t=1}^{T}[(\tilde{\mathscr{H}}')^{-1}\hat{h}_t - h_t]\hat{\mathcal{R}}_t' = O_p\left(T^{-1/2}\|\hat{H}\mathscr{H}^{-1} - H\|T^{-1/2}\|\hat{\mathcal{R}}\|\right) = O_p\left[\frac{1}{\sqrt{N}}\frac{1}{m\sqrt{Tm}}\right] = o_p(\frac{1}{N\sqrt{T}}),$$

since $\sqrt{T} \ll N \ll T^2$. Finally, from Assumption B.13 and the bound in (B.142), we have:

$$\begin{aligned} &\frac{1}{T}\sum_{t=1}^{T}(\frac{1}{\sqrt{N}}u_t + \frac{1}{T}\breve{b}_t + \frac{1}{\sqrt{NT}}\breve{d}_t + \tilde{\vartheta}_t)\tilde{\vartheta}_t' \\ =\ & \frac{1}{T}\sum_{t=1}^{T}\frac{1}{\sqrt{N}}u_t\tilde{\vartheta}_t' + \frac{1}{T}\sum_{t=1}^{T}\frac{1}{T}\breve{b}_t\tilde{\vartheta}_t' + O_p\left[\left(\frac{1}{\sqrt{NT}}+\frac{1}{N}\right)^2\right] \\ =\ & \frac{1}{T\sqrt{N}}\sum_{t=1}^{T}u_t\tilde{\vartheta}_t' + \frac{1}{T^2}\sum_{t=1}^{T}\breve{b}_t\tilde{\vartheta}_t' + o_p\left(\frac{1}{N\sqrt{T}}\right), \end{aligned}$$

and:

$$\begin{aligned} \frac{1}{T\sqrt{N}}\sum_{t=1}^{T}u_t\tilde{\vartheta}_t' &= \frac{1}{N\sqrt{T}}(\frac{1}{T}\sum_{t=1}^{T}u_t\breve{\alpha}_t)S\left(\frac{\Lambda'\Lambda}{N}\right)^{-1} + \frac{1}{\sqrt{NT}}(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}u_t h_t') \\ &\quad\times\left\{\frac{1}{T}O_p\left(\hat{B}\right) + \frac{1}{N}O_p\left(\hat{B}\right) + \frac{1}{\sqrt{NT}}O_p\left(\hat{B}\right) + \frac{1}{\sqrt{NT}}O_p\left(\breve{\Pi}_3\right)\right\} \\ &\quad+\frac{1}{\sqrt{NT}}(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}u_t\breve{h}_t')\frac{1}{T}O_p\left(\hat{B}\right) \\ &\quad+O_p\left[\frac{1}{\sqrt{NT}}+\frac{1}{N\sqrt{N}}+\frac{1}{N\sqrt{T}}+\frac{1}{T^2}\right] + o_p(\frac{1}{N\sqrt{T}}) \\ &= \frac{1}{N\sqrt{T}}(\Lambda'\Lambda/N)^{-1}(\frac{1}{T}\sum_{t=1}^{T}\xi_t\breve{\alpha}_t')S\left(\frac{\Lambda'\Lambda}{N}\right)^{-1} + o_p(\frac{1}{N\sqrt{T}}) = o_p(\frac{1}{N\sqrt{T}}), \end{aligned}$$

from Assumption B.13 (iv), and:

$$\begin{aligned} \frac{1}{T^2}\sum_{t=1}^{T}b_t\tilde{\vartheta}_t' &= \frac{1}{\sqrt{N}T^2}(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}b_t\breve{\alpha}_t')S\left(\frac{\Lambda'\Lambda}{N}\right)^{-1} + ... \\ &\quad+... + o_p(\frac{1}{N\sqrt{T}}) \\ &= \frac{1}{\sqrt{N}T^2}\left(\frac{\Lambda'\Lambda}{N}\right)^{-1}S'(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\eta_t^2\breve{h}_t\breve{\alpha}_t')S\left(\frac{\Lambda'\Lambda}{N}\right)^{-1} + o_p(\frac{1}{N\sqrt{T}}) = o_p(\frac{1}{N\sqrt{T}}), \end{aligned}$$

since $\sqrt{T} \ll N \ll T^2$. Hence, $\frac{1}{T}\sum_{t=1}^{T}(\frac{1}{\sqrt{N}}u_t + \frac{1}{T}\breve{b}_t + \frac{1}{\sqrt{NT}}\breve{d}_t + \breve{\vartheta}_t)\breve{\vartheta}_t' = o_p(\frac{1}{N\sqrt{T}})$ follows. ■

B.5.1.4 Proof of Lemma B.10

The proof follows closely the proof of Proposition 1 (ii) in Bai (2009), applied to the projected model (B.46), that is: $\breve{Y} = \breve{H}\Lambda' + \breve{\varepsilon}$. Let us denote by Λ^0 the matrix of true factor loadings, in order to distinguish it from a matrix Λ of generic factor values. The estimator $\hat{\Lambda}$ is obtained from minimization of the LS criterion:

$$\min_{\breve{H},\Lambda:\Lambda'\Lambda/N=I_k} tr[(\breve{Y} - \breve{H}\Lambda')'(\breve{Y} - \breve{H}\Lambda')]. \tag{B.143}$$

The criterium in (B.143), after concentration w.r.t. H, that is by using $\hat{H} = \breve{Y}\Lambda(\Lambda'\Lambda)^{-1}$, becomes $tr(\breve{Y}M_\Lambda\breve{Y}')$, where $M_\Lambda = I_N - P_H$ and $P_\Lambda = \Lambda(\Lambda'\Lambda)^{-1}\Lambda'$. Let us divide the criterium by NT, and subtract its value at Λ^0, to get:

$$S_{NT}(\Lambda) = \frac{1}{NT}tr(\breve{Y}M_\Lambda\breve{Y}') - \frac{1}{NT}tr(\breve{\varepsilon}M_{\Lambda^0}\breve{\varepsilon}').$$

The matrix of factor estimates $\hat{\Lambda}$ is the minimizer of function $S_{NT}(\Lambda)$ w.r.t. Λ such that $\Lambda'\Lambda/N = I_k$. By using $\breve{Y} = \breve{H}\Lambda^{0\prime} + \breve{\varepsilon}$, we get:

$$S_{NT}(\Lambda) = \frac{1}{NT}tr(\breve{H}\Lambda^{0\prime}M_\Lambda\Lambda^0\breve{H}') + 2\frac{1}{NT}tr(\breve{H}\Lambda^{0\prime}M_\Lambda\breve{\varepsilon}') + \frac{1}{NT}tr(\breve{\varepsilon}(P_\Lambda - P_{\Lambda^0})\breve{\varepsilon}'). \tag{B.144}$$

Now, let us show that the second and third terms in the r.h.s. are $o_p(1)$ uniformly w.r.t. the (N,k) matrix Λ such that $\Lambda'\Lambda/T = I_k$. We follow here different arguments compared to the ones in the proof of Lemma A.1 in Bai (2009), since we deploy slightly different assumptions. We have:

$$\begin{aligned}
\frac{1}{NT}tr(\breve{H}\Lambda^{0\prime}M_\Lambda\breve{\varepsilon}') &= \frac{1}{NT}tr(\Lambda^{0\prime}\breve{\varepsilon}'\breve{H}) - tr\left[\frac{1}{N}\Lambda^{0\prime}\Lambda(\frac{1}{N}\Lambda'\Lambda)^{-1}\frac{1}{NT}\Lambda'\breve{\varepsilon}'\breve{H}\right] \\
&= O_p(\|\frac{1}{NT}\Lambda^{0\prime}\breve{\varepsilon}'\breve{H}\|) + O_p(\|\frac{1}{NT}\Lambda'\breve{\varepsilon}'\breve{H}\|) = O_p(\|\frac{1}{\sqrt{N}T}\breve{\varepsilon}'\breve{H}\|) = O_p(\frac{1}{\sqrt{T}}),
\end{aligned}$$

and:

$$\begin{aligned}
\frac{1}{NT}tr(\varepsilon(P_\Lambda - P_{\Lambda^0})\varepsilon') &= \frac{1}{N}tr\left[\frac{1}{N}\breve{\varepsilon}\Lambda(\frac{1}{N}\Lambda'\Lambda)^{-1}\frac{1}{N}\Lambda'\breve{\varepsilon}'\right] - tr\left[\frac{1}{N}\breve{\varepsilon}\Lambda^0(\frac{1}{N}\Lambda^{0\prime}\Lambda^0)^{-1}\frac{1}{N}\Lambda^{0\prime}\breve{\varepsilon}'\right] \\
&= \frac{1}{N}tr\left[\Lambda'(\frac{1}{NT}\breve{\varepsilon}'\breve{\varepsilon})\Lambda\right] - \frac{1}{N}tr\left[(\frac{1}{N}\Lambda^{0\prime}\Lambda^0)^{-1}\Lambda^{0\prime}(\frac{1}{NT}\breve{\varepsilon}'\breve{\varepsilon})\Lambda^0\right] \\
&= O_p(\|\frac{1}{NT}\breve{\varepsilon}'\breve{\varepsilon}\|) = o_p(1),
\end{aligned}$$

uniformly w.r.t. the (N,k) matrix Λ such that $\Lambda'\Lambda/T = I_k$, using Assumptions B.10 and B.12 i) and iii), Lemma B.12 i), and the invariance of the trace under cyclical permutations.

Thus, from (B.144) we get $S_{NT}(\Lambda) = \tilde{S}_{NT}(\Lambda) + o_p(1)$, where:

$$\tilde{S}_{NT}(\Lambda) = \frac{1}{NT}tr(\breve{H}\Lambda^{0\prime}M_\Lambda\Lambda^0\breve{H}') = tr[(\Lambda^{0\prime}M_\Lambda\Lambda^0/N)(\breve{H}'\breve{H}/T)], \tag{B.145}$$

and the $o_p(1)$ term is uniform w.r.t. Λ such that $\Lambda'\Lambda/N = I_k$. We have:

$$\begin{aligned}\tilde{S}_{NT}(\hat{\Lambda}) &\geq 0, \\ 0 = S_{NT}(\Lambda^0) &\geq S_{NT}(\hat{\Lambda}) = \tilde{S}_{NT}(\hat{\Lambda}) + o_p(1),\end{aligned}$$

which imply $\tilde{S}_{NT}(\hat{\Lambda}) = o_p(1)$. Then, from equation (B.145), Assumption B.10 and $\hat{\Lambda}'\hat{\Lambda}/N = I_k$, it follows:

$$\Lambda^{0\prime}\Lambda^0/N - (\Lambda^{0\prime}\hat{\Lambda}/N)(\hat{\Lambda}'\Lambda^0/N) = o_p(1).$$

Thus, from Assumption B.10, we have $(\Lambda^{0\prime}\hat{\Lambda}/N)(\hat{\Lambda}'\Lambda^0/N) = \Sigma_\lambda + o_p(1)$. Lemma B.10 follows. ■

B.5.1.5 Proof of Lemma B.11

Let us multiply both sides of equation (B.63) by $N^{-1}\Lambda'$ to get:

$$(\Lambda'\hat{\Lambda}/N)\hat{V} - (\Lambda'\Lambda/N)\left(H'W^2H/T\right)\left(\Lambda'\hat{\Lambda}/N\right) = \frac{1}{N^2T}\Lambda'(\varepsilon'W^2\varepsilon\hat{\Lambda} + \Lambda H'W^2\varepsilon\hat{\Lambda} + \varepsilon'W^2H\Lambda'\hat{\Lambda}).$$

By applying the Cauchy-Schwarz inequality, Assumption B.12 ii), Lemmas B.10 and B.12 *(i)*, and $N^{-1/2}\|\hat{\Lambda}\| = \sqrt{k}$, we get:

$$(\Lambda'\hat{\Lambda}/N)\hat{V} - (\Lambda'\Lambda/N)\left(H'W^2H/T\right)\left(\Lambda'\hat{\Lambda}/N\right) = o_p(1).$$

Then, from Lemma B.10 and Assumption B.10, we get:

$$\begin{aligned}\hat{V} &= (\Lambda'\hat{\Lambda}/N)^{-1}(\Lambda'\Lambda/N)\left(H'W^2H/T\right)\left(\Lambda'\hat{\Lambda}/N\right) + o_p(1) \\ &= (\Lambda'\hat{\Lambda}/N)^{-1}\,\Sigma_\lambda\Sigma_h(W)\,(\Lambda'\hat{\Lambda}/N) + o_p(1)\end{aligned}$$

We deduce that the eigenvalues of matrix $\hat{V}$ converge in probability to the eigenvalues of matrix $\Sigma_\lambda\Sigma_h(W)$. Since matrix $\hat{V}$ is diagonal, the conclusion follows. ■

B.5.1.6 Proof of Lemma B.12

(i) Using $\frac{1}{\sqrt{T}}[\breve{\varepsilon}'H]_i = \frac{1}{\sqrt{T}}\sum_{t=1}^T \breve{\varepsilon}_{i,t}\breve{h}_t = \breve{\aleph}_i$ and Assumption B.12 v), we have:

$$\begin{aligned}\|\frac{1}{T}\varepsilon W^2H\| &= \|\frac{1}{T}\breve{\varepsilon}\breve{H}\| = \frac{1}{\sqrt{T}}\left[tr\left(\sum_{i=1}^N \breve{\aleph}_i\breve{\aleph}_i'\right)\right]^{1/2} \\ &= \sqrt{\frac{N}{T}}\left[tr\left(\frac{1}{N}\sum_{t=1}^T \breve{\aleph}_i\breve{\aleph}_i'\right)\right]^{1/2} = O_p\left(\sqrt{\frac{N}{T}}\right). \qquad \text{(B.146)}\end{aligned}$$

(ii) By using (B.146) and $N^{-1/2}\|\Lambda\| = O_p(1)$ we have:

$$\|\frac{1}{NT}\Lambda\breve{H}'\breve{\varepsilon}\| \leq \frac{1}{\sqrt{N}}\frac{1}{\sqrt{N}}\|\Lambda\|\|\frac{1}{T}\breve{H}'\breve{\varepsilon}\| = O_p\left(\frac{1}{\sqrt{T}}\right).$$

(iii) By using Assumption B.12 i), we have

$$\|\frac{1}{NT}\Lambda\breve{H}'\breve{\varepsilon}\Lambda\| \leq \frac{1}{\sqrt{T}}\frac{1}{\sqrt{N}}\|\Lambda\|\|\frac{1}{\sqrt{NT}}\breve{H}'\breve{\varepsilon}\Lambda\| = O_p(\frac{1}{\sqrt{T}}),$$

(iv) Using $\frac{1}{\sqrt{N}}[\breve{\varepsilon}\Lambda]_t = \frac{1}{\sqrt{N}}\sum_{i=1}^N \lambda_i\breve{\varepsilon}_{i,t} = \breve{\xi}_t$ and Assumption B.12 i), we have:

$$\|\frac{1}{N}\breve{\varepsilon}\Lambda\| = \frac{1}{\sqrt{N}}\left[tr\left(\sum_{t=1}^T \breve{\xi}_t\breve{\xi}_t'\right)\right]^{1/2} = \sqrt{\frac{T}{N}}\left[tr\left(\frac{1}{T}\sum_{t=1}^T \breve{\xi}_t\breve{\xi}_t'\right)\right]^{1/2} = O_p\left(\sqrt{\frac{T}{N}}\right). \quad \text{(B.147)}$$

In case v) and vii) are needed, proof is the same as in AGGR (v) By using (B.147) and $T^{-1/2}\|H\| = O_p(1)$, we have:

$$\|\frac{1}{NT}H\Lambda'\varepsilon'\| \leq \frac{1}{T}\|H\|\|\frac{1}{N}\varepsilon\Lambda\| = O_p\left(\frac{1}{\sqrt{N}}\right).$$

(vi) We have:

$$\|\frac{1}{NT}H\Lambda'\varepsilon'H\| \leq \frac{1}{\sqrt{N}}T^{-1/2}\|H\|\|\frac{1}{\sqrt{NT}}\Lambda'\varepsilon'H\| = O_p(\frac{1}{\sqrt{N}}),$$

by using $\frac{1}{\sqrt{NT}}\Lambda'\varepsilon'H = \frac{1}{\sqrt{T}}\sum_{t=1}^T \xi_t h_t' = O_p(1)$ from Assumption B.12 i). ■

B.5.1.7 Check of the conditions in Assumptions B.11-B.10

a) Check of Assumption B.10

Assumption B.10 is standard in the factor literature, see e.g. Bai and Ng (2002), Stock and Watson (2002), Bai (2003). It is implied by Assumptions B.2 and B.3. ■

b) Check of Assumption B.11

[TO BE WRITTEN] ■

c) Check of Assumption B.12

Using the definitions of $\breve{h}_t$ and $\breve{\xi}_t$, we get:

$$\begin{aligned}
\frac{1}{\sqrt{T}}\sum_{t=1}^T \breve{h}_t\breve{\xi}_t' &= \frac{1}{\sqrt{T}}\sum_{t=1}^T (h_t + \tilde{\gamma}_{RP}\bar{h})(\xi_t + \tilde{\gamma}_{RP}\bar{\xi})' \\
&= \frac{1}{\sqrt{T}}\sum_{t=1}^T h_t\xi_t' + \tilde{\gamma}_{RP}\left(\frac{1}{T}\sum_{t=1}^T \xi_t\right)\frac{1}{\sqrt{T}}\sum_{t=1}^T h_t + \tilde{\gamma}_{RP}\left(\frac{1}{T}\sum_{t=1}^T h_t\right)\frac{1}{\sqrt{T}}\sum_{t=1}^T \xi_t' \\
&\quad + \tilde{\gamma}_{RP}^2\frac{1}{\sqrt{T}}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^T h_t\right)\left(\frac{1}{\sqrt{T}}\sum_{t=1}^T \xi_t\right)' = O_p(1) + O_P(1/\sqrt{T}).
\end{aligned}$$

where the last equality follows from Assumptions B.4 b), B.5 b) and B.6 a) which imply $\frac{1}{\sqrt{T}}\sum_{t=1}^{T} h_t = O_p(1)$ and $\frac{1}{\sqrt{T}}\sum_{t=1}^{T} \xi_t = O_p(1)$. Moreover, Assumption B.6 a) also implies $\frac{1}{\sqrt{T}}\sum_{t=1}^{T} h_t\xi_t' = O_p(1)$.

Equality $E[\|\breve{\xi}_t\|^2] = O(1)$ follows directly from Assumption B.11 (iii). And this completes the proof of part (i).

To bound $\frac{1}{NT}\varepsilon W^2\varepsilon'\Lambda$, we first need to bound the j-th column of matrix $\frac{1}{NT}\varepsilon'\varepsilon\Lambda$, that is $\frac{1}{NT}[\varepsilon'\varepsilon\Lambda]_j = \frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\varepsilon_{j,t}\varepsilon_{i,t}\lambda_i$:

$$
\begin{aligned}
\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\varepsilon_{j,t}\varepsilon_{i,t}\lambda_i &= \frac{1}{NT}\sum_{t=1}^{T}\varepsilon_{j,t}^2\lambda_i + \frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1,i\neq j}^{N}\varepsilon_{j,t}\varepsilon_{i,t}\lambda_i \\
&= \frac{1}{NT}\sum_{t=1}^{T}E[\varepsilon_{j,t}^2]\lambda_i + \frac{1}{NT}\sum_{t=1}^{T}(\varepsilon_{j,t}^2 - E[\varepsilon_{j,t}^2])\lambda_i + \frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1,i\neq j}^{N}\varepsilon_{j,t}\varepsilon_{i,t}\lambda_i \\
&= \frac{1}{N}\eta_j^*\lambda_j + \frac{1}{N\sqrt{T}}\kappa_j^*\lambda_j + \frac{1}{\sqrt{NT}}\alpha_j^*
\end{aligned}
$$

where

$$
\begin{aligned}
\eta_j^* &= \frac{1}{T}\sum_{t=1}^{T}E[\varepsilon_{j,t}^2] \\
\kappa_j^* &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\varepsilon_{j,t}^2 - E[\varepsilon_{j,t}^2]) \\
\alpha_j^* &= \frac{1}{\sqrt{NT}}\sum_{t=1}^{T}\sum_{i=1,i\neq j}^{N}\varepsilon_{j,t}\varepsilon_{i,t}\lambda_i
\end{aligned}
$$

Analogously, by using the definitions of $\breve{\varepsilon}_{i,t}$, we get:

$$
\begin{aligned}
&\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\breve{\varepsilon}_{j,t}\breve{\varepsilon}_{i,t}\lambda_i\\
&=\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\varepsilon_{j,t}\varepsilon_{i,t}\lambda_i+\tilde{\gamma}_{RP}\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{s=1}^{T}\varepsilon_{j,s}\right)\varepsilon_{i,t}\lambda_i\\
&+\tilde{\gamma}_{RP}\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\varepsilon_{j,t}\left(\frac{1}{T}\sum_{r=1}^{T}\varepsilon_{i,r}\right)\lambda_i+\tilde{\gamma}_{RP}^2\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{s=1}^{T}\varepsilon_{j,s}\right)\left(\frac{1}{T}\sum_{r=1}^{T}\varepsilon_{i,r}\right)\varepsilon_{i,t}\lambda_i\\
&=\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\varepsilon_{j,t}\varepsilon_{i,t}\lambda_i\ +\tilde{\gamma}_{RP}\frac{1}{NT^2}\sum_{t=1}^{T}\sum_{i=1}^{N}\varepsilon_{j,t}\varepsilon_{i,t}\lambda_i+\tilde{\gamma}_{RP}\frac{1}{NT^2}\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{r=1,r\neq t}^{T}\varepsilon_{j,t}\varepsilon_{i,r}\lambda_i\\
&+\tilde{\gamma}_{RP}\frac{1}{NT^2}\sum_{t=1}^{T}\sum_{i=1}^{N}\varepsilon_{j,t}\varepsilon_{i,t}\lambda_i+\tilde{\gamma}_{RP}\frac{1}{NT^2}\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{s=1,s\neq t}^{T}\varepsilon_{j,s}\varepsilon_{i,t}\lambda_i\\
&+\tilde{\gamma}_{RP}^2\frac{1}{NT^2}\sum_{t=1}^{T}\sum_{i=1}^{N}\varepsilon_{j,t}\varepsilon_{i,t}\lambda_i+\tilde{\gamma}_{RP}^2\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{s=1}^{T}\sum_{r=1,r,s\neq t}^{T}\varepsilon_{j,s}\varepsilon_{i,r}\lambda_i\\
&=\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\varepsilon_{j,t}\varepsilon_{i,t}\lambda_i+o_p\left(\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\varepsilon_{j,t}\varepsilon_{i,t}\lambda_i\right)\\
&+\tilde{\gamma}_{RP}\frac{1}{NT^2}\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{r=1,r\neq t}^{T}\varepsilon_{j,t}\varepsilon_{i,r}\lambda_i+\tilde{\gamma}_{RP}\frac{1}{NT^2}\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{s=1,s\neq t}^{T}\varepsilon_{j,s}\varepsilon_{i,t}\lambda_i+\tilde{\gamma}_{RP}^2\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{s=1}^{T}\sum_{r=1,r,s\neq t}^{T}\varepsilon_{j,s}\varepsilon_{i,r}\lambda_i\\
&=\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\varepsilon_{j,t}\varepsilon_{i,t}\lambda_i+o_p\left(\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\varepsilon_{j,t}\varepsilon_{i,t}\lambda_i\right)\\
&+\tilde{\gamma}_{RP}\frac{1}{T\sqrt{NT}}\sum_{t=1}^{T}\left\{\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{r=1,r\neq t}^{T}\varepsilon_{j,t}\varepsilon_{i,r}\lambda_i\right\}+\tilde{\gamma}_{RP}\frac{1}{T\sqrt{NT}}\sum_{t=1}^{T}\left\{\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{s=1,s\neq t}^{T}\varepsilon_{j,s}\varepsilon_{i,t}\lambda_i\right\}\\
&+\tilde{\gamma}_{RP}^2\frac{1}{TN}\sum_{i=1}^{N}\left\{\frac{1}{T}\sum_{s=1}^{T}\sum_{r=1,r,s\neq t}^{T}\varepsilon_{j,s}\varepsilon_{i,r}\lambda_i\right\}
\end{aligned}
$$

which implies:

$$
\begin{aligned}
\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\breve{\varepsilon}_{j,t}\breve{\varepsilon}_{i,t}\lambda_i &= \frac{1}{N}\eta_j^*\lambda_j+\frac{1}{N\sqrt{T}}\kappa_j^*\lambda_j+\frac{1}{\sqrt{NT}}\alpha_j^*+o_p\left(\frac{1}{N}\eta_j^*\lambda_j+\frac{1}{N\sqrt{T}}\kappa_j^*\lambda_j+\frac{1}{\sqrt{NT}}\alpha_j^*\right)\\
&+\tilde{\gamma}_{RP}\frac{1}{\sqrt{NT}}\frac{1}{T}\sum_{t=1}^{T}(\delta_{j,t}^*+\delta_{j,t}^{**})+\tilde{\gamma}_{RP}^2\frac{1}{TN}\sum_{i=1}^{N}\delta_{i,j}^{***}
\end{aligned}
$$

Therefore,

$$\begin{aligned}
\|\frac{1}{NT}\breve{\varepsilon}'\breve{\varepsilon}\Lambda\|^2 &= \sum_{j=1}^{N}(\frac{1}{NT}[\breve{\varepsilon}'\breve{\varepsilon}\Lambda]_j)'(\frac{1}{NT}[\breve{\varepsilon}'\breve{\varepsilon}\Lambda]_j) = \sum_{j=1}^{N}\|\frac{1}{NT}[\breve{\varepsilon}'\breve{\varepsilon}\Lambda]_j\|^2 = \sum_{j=1}^{N}\|\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\breve{\varepsilon}_{j,t}\breve{\varepsilon}_{i,t}\lambda_i\|^2 \\
&\leq 2\sum_{j=1}^{N}\left\{\frac{1}{N^2}\|\eta_j^*\lambda_j\|^2 + \frac{1}{N^2T}\|\kappa_j^*\lambda_j\|^2 + \frac{1}{NT}\|\alpha_j^*\|^2 + o_p\left(\frac{1}{N^2}\|\eta_j^*\lambda_j\|^2 + \frac{1}{N^2T}\|\kappa_j^*\lambda_j\|^2 + \frac{1}{NT}\|\alpha_j^*\|^2\right)\right. \\
&\left. +\tilde{\gamma}_{RP}\frac{1}{NT}\frac{1}{T^2}\|\sum_{t=1}^{T}\delta_{j,t}^*\|^2 + \tilde{\gamma}_{RP}\frac{1}{NT}\frac{1}{T^2}\|\sum_{t=1}^{T}\delta_{j,t}^{**}\|^2 + \tilde{\gamma}_{RP}^2\frac{1}{T^2N^2}\|\sum_{i=1}^{N}\delta_{i,j}^{***}\|^2\right\} \\
&\leq 2\frac{1}{N}\sum_{j=1}^{N}\left\{\frac{1}{N}\|\eta_j^*\lambda_j\|^2 + \frac{1}{NT}\|\kappa_j^*\lambda_j\|^2 + \frac{1}{T}\|\alpha_j^*\|^2 + o_p\left(\frac{1}{N}\|\eta_j^*\lambda_j\|^2 + \frac{1}{NT}\|\kappa_j^*\lambda_j\|^2 + \frac{1}{T}\|\alpha_j^*\|^2\right)\right\} \\
&+4\tilde{\gamma}_{RP}\left\{\frac{1}{T^2}\frac{1}{NT}\sum_{j=1}^{N}\sum_{t=1}^{T}\|\delta_{j,t}^*\|^2 + \frac{1}{T^2}\frac{1}{NT}\sum_{j=1}^{N}\sum_{t=1}^{T}\|\delta_{j,t}^{**}\|^2 + \tilde{\gamma}_{RP}\frac{1}{T^2}\frac{1}{N^2}\sum_{j=1}^{N}\sum_{i=1}^{N}\|\delta_{i,j}^{***}\|^2\right\} \\
&= O_p\left(\frac{1}{T} + \frac{1}{N}\right),
\end{aligned}$$

Therefore part (ii) follows from Assumptions B.4 a), ..., Assumption B.13 (which is checked below), Assumption B.13 v), since $\eta_t^2 \leq M$ (Assumption B.4 a)).

Let us now show the validity of Assumption B.12 (iii). We have:

$$\begin{aligned}
\|\frac{1}{NT}\breve{\varepsilon}\breve{\varepsilon}'\|^2 &= \frac{1}{N^2T^2}Tr[\breve{\varepsilon}\breve{\varepsilon}'\breve{\varepsilon}\breve{\varepsilon}'] = \frac{1}{N^2T^2}\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{s=1}^{T}\breve{\varepsilon}_{i,t}\breve{\varepsilon}_{i,s}\breve{\varepsilon}_{j,s}\breve{\varepsilon}_{j,t} \\
&= \frac{1}{N^2T^2}\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\breve{\varepsilon}_{i,t}^2\breve{\varepsilon}_{j,t}^2 + \frac{1}{N^2T^2}\sum_{t=1}^{T}\sum_{s=1,s\neq t}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\breve{\varepsilon}_{i,t}\breve{\varepsilon}_{i,s}\breve{\varepsilon}_{j,t}\breve{\varepsilon}_{j,s}.
\end{aligned}$$

The first term in the r.h.s. is $O_p(T^{-1})$ from Assumption B.4 b). Let us now consider the second term in the r.h.s.. We have:

$$\begin{aligned}
&\frac{1}{N^2T^2}\sum_{t=1}^{T}\sum_{s=1,s\neq t}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\breve{\varepsilon}_{i,t}\breve{\varepsilon}_{i,s}\breve{\varepsilon}_{j,t}\breve{\varepsilon}_{j,s} \\
=\ &\frac{2}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{t-1}\left(\frac{1}{N}\sum_{i=1}^{N}\breve{\varepsilon}_{i,t}\breve{\varepsilon}_{i,s}\right)^2 \\
=\ &\frac{2}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{t-1}\breve{\eta}_{ts}^4 + \frac{4}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{t-1}\left(\frac{1}{N}\sum_{i=1}^{N}(\breve{\varepsilon}_{i,t}\breve{\varepsilon}_{i,s} - \breve{\eta}_{ts}^2)\right)\breve{\eta}_{ts}^2 + \frac{2}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{t-1}\left(\frac{1}{N}\sum_{i=1}^{N}(\breve{\varepsilon}_{i,t}\breve{\varepsilon}_{i,s} - \breve{\eta}_{ts}^2)\right)^2,
\end{aligned}$$

where $\breve{\eta}_{ts}^2 := \operatorname*{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\breve{\varepsilon}_{i,t}\breve{\varepsilon}_{i,s}$. By taking expectations, and using the Cauchy-Schwarz inequality and Assumption B.11 (iv) (in AGGR it was B.7 a)), we get that $\frac{1}{N^2T^2}\sum_{t=1}^{T}\sum_{s=1,s\neq t}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\breve{\varepsilon}_{i,t}\breve{\varepsilon}_{i,s}\breve{\varepsilon}_{j,t}\breve{\varepsilon}_{j,s} =$

$O_p(\frac{1}{T}+\frac{1}{N})$. Assumption B.12 iii) follows.

Assumption B.12 (iv) is implied by Assumptions **??** a) and Assumption B.13 (which is checked below). Indeed, from (B.79) we have:

$$\begin{aligned}\|\frac{1}{NT}\varepsilon\varepsilon' W^2 H\|^2 &= \|\frac{1}{NT}\varepsilon\breve{\varepsilon}'\breve{H}\|^2 = \sum_{t=1}^{T}(\frac{1}{NT}[\varepsilon\breve{\varepsilon}'\breve{H}]_t)'(\frac{1}{NT}[\varepsilon\breve{\varepsilon}'\breve{H}]_t) = \sum_{t=1}^{T}\|\frac{1}{NT}[\varepsilon\breve{\varepsilon}'\breve{H}]_t\|^2 \\ &\leq 2\sum_{t=1}^{T}\left(\frac{1}{T^2}\|\eta_t^2\breve{h}_t\|^2+\frac{1}{T^2N}\|\kappa_t\breve{h}_t\|^2+\frac{1}{NT}\|\breve{\alpha}_t\|^2\right. \\ &\quad \left.+\tilde{\gamma}_{RP}\left[\frac{1}{T^4}\|\eta_t^2\breve{h}_t\|^2+\frac{1}{T^3N}\|\breve{\alpha}_t^*\|^2+\frac{1}{T^6N}\|\kappa_t\breve{h}_t\|^2\right]\right) = O_p(\frac{1}{T}+\frac{1}{N}),\end{aligned}$$

under Assumption B.13 v), ..., since $\eta_t^2 \leq M$ (Assumption **??** a)) and ... .

Finally, Assumption B.12 (iv) follows directly from Assumption B.5 d) .

■

d) Check of the conditions in Assumption B.13

Assumption B.13 i) corresponds to Assumption B.4 a). Assumptions B.13 (ii)-(iv) are implied by Assumption B.7 b). Assumption B.13 (v) is implied by Assumptions B.5 b), c), and B.7 c). ■

This concludes the proofs of all the technical results needed to prove Proposition B.2. ■

B.5.2 Proof of Lemma B.1

(**TO BE ADAPTED**) We prove the bound for $\hat{X}_{1,2}$; the bounds for the other terms are obtained similarly. We substitute the definition $\breve{\psi}_{j,t} = \frac{1}{\sqrt{N_j}}\breve{u}_{j,t} + \frac{1}{T}\breve{b}_{j,t} + \frac{1}{\sqrt{N_j T}}\breve{d}_{j,t} + \breve{\vartheta}_{j,t}$ into (B.11) and use $N_2 = N$, $N_1 = N/\mu_N^2$. We get:

$$\begin{aligned}
\hat{X}_{12} \quad = \quad & \frac{1}{T\sqrt{N}}\sum_{t=1}^{T}(\breve{h}_{1,t}\breve{u}_{2,t}' + \mu_N \breve{u}_{1,t}\breve{h}_{2,t}') + \frac{\mu_N}{TN}\sum_{t=1}^{T}\breve{u}_{1,t}\breve{u}_{2,t}' \\
& + \frac{1}{T^2}\sum_{t=1}^{T}\left(\breve{h}_{1,t}\breve{b}_{2,t}' + \breve{b}_{1,t}\breve{h}_{2,t}'\right) + \frac{1}{T^2\sqrt{N}}\sum_{t=1}^{T}\left(\breve{b}_{1,t}\breve{u}_{2,t}' + \mu_N\breve{u}_{1,t}\breve{b}_{2,t}'\right) + \frac{1}{T^3}\sum_{t=1}^{T}\breve{b}_{1,t}\breve{b}_{2,t}' \\
& + \frac{1}{T\sqrt{NT}}\sum_{t=1}^{T}\left(\breve{h}_{1,t}\breve{d}_{2,t}' + \mu_N\breve{d}_{1,t}\breve{h}_{2,t}'\right) + \frac{\mu_N}{TN\sqrt{T}}\sum_{t=1}^{T}\left(\breve{u}_{1,t}\breve{d}_{2,t}' + \breve{d}_{1,t}\breve{u}_{2,t}'\right) \\
& + \frac{1}{T^2\sqrt{NT}}\sum_{t=1}^{T}\left(\breve{b}_{1,t}\breve{d}_{2,t}' + \mu_N\breve{d}_{1,t}\breve{b}_{2,t}'\right) + \frac{\mu_N}{NT^2}\sum_{t=1}^{T}\breve{d}_{1,t}\breve{d}_{2,t}' + \frac{1}{T}\sum_{t=1}^{T}(\breve{h}_{1,t}\breve{\vartheta}_{2,t}' + \breve{\vartheta}_{1,t}\breve{h}_{2,t}') \\
& + \frac{1}{T}\sum_{t=1}^{T}\left[\left(\frac{\mu_N}{\sqrt{N}}\breve{u}_{1,t} + \frac{1}{T}\breve{b}_{1,t} + \frac{\mu_N}{\sqrt{NT}}\breve{d}_{1,t} + \breve{\vartheta}_{1,t}\right)\breve{\vartheta}_{2,t}' + \vartheta_{1,t}\left(\frac{1}{\sqrt{N}}\breve{u}_{2,t} + \frac{1}{T}\breve{b}_{2,t} + \frac{1}{\sqrt{NT}}\breve{d}_{2,t}\right)'\right].
\end{aligned} \tag{B.148}$$

To bound the terms in the r.h.s. of (B.148), we use that under Assumptions B.2-B.4, B.5 *b)-c)* and B.6 *a)* we have:

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}h_{j,t}u_{k,t}' = O_p(1), \qquad \frac{1}{T}\sum_{t=1}^{T}u_{j,t}u_{k,t}' = O_p(1), \tag{B.149}$$

$$\frac{1}{T}\sum_{t=1}^{T}h_{j,t}b_{k,t}' = O_p(1), \tag{B.150}$$

$$\frac{1}{T}\sum_{t=1}^{T}b_{j,t}u_{k,t}' = O_p\left(\frac{1}{\sqrt{T}}\right), \tag{B.151}$$

$$\frac{1}{T}\sum_{t=1}^{T}b_{j,t}b_{k,t}' = O_p(1), \tag{B.152}$$

$$\frac{1}{T}\sum_{t=1}^{T}h_{j,t}d_{k,t}' = O_p(1), \tag{B.153}$$

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}u_{j,t}d_{k,t}' = O_p(1), \tag{B.154}$$

$$\frac{1}{T}\sum_{t=1}^{T}b_{j,t}d_{k,t}' = O_p(1), \tag{B.155}$$

$$\frac{1}{T}\sum_{t=1}^{T}d_{j,t}d_{k,t}' = O_p(1), \tag{B.156}$$

for $j, k = 1, 2$. These bounds are shown below by using the definitions of $u_{j,t}$, $b_{j,t}$, $d_{j,t}$ in Proposition B.2. Therefore, the first nine summation terms in the r.h.s. of (B.148) are of order $O_p(\frac{1}{\sqrt{NT}})$, $O_p(\frac{1}{N})$, $O_p(\frac{1}{T})$, $O_p(\frac{1}{T\sqrt{NT}})$, $O_p(\frac{1}{T^2})$, $O_p(\frac{1}{\sqrt{NT}})$, $O_p(\frac{1}{NT})$, $O_p(\frac{1}{T\sqrt{NT}})$ and $O_p(\frac{1}{NT})$, respectively. From Proposition B.2, the last two summation terms in the r.h.s. of (B.148) are of order $O_p(\frac{1}{N}+\frac{1}{T^2})$ and $o_p(\frac{1}{N\sqrt{T}})$, respectively. Therefore, we get $\hat{X}_{1,2} = O_p(\delta_{N,T})$, where $\delta_{N,T} = \max\{\frac{1}{N}, \frac{1}{T}\} = (\min\{N, T\})^{-1}$.

Proof of (B.150). We have:

$$\frac{1}{T}\sum_{t=1}^{T} h_{j,t}b'_{k,t} = \left(\frac{1}{T}\sum_{t=1}^{T} h_{j,t}h'_{k,t}\eta^2_{k,t}\right)\left(\frac{1}{T}\sum_{t=1}^{T} h_{k,t}h'_{k,t}\right)^{-1}\left(\frac{1}{N_k}\sum_{i=1}^{N_k}\lambda_{k,i}\lambda'_{k,i}\right)^{-1}.$$

The first and second terms in the r.h.s. are $O_p(1)$ by Assumptions B.2, B.4 *b)* and B.6 *a)* and an application of a LLN for mixing processes. The third term in the r.h.s. is $O_p(1)$ by Assumption B.3. Then, (B.150) follows.

Proof of (B.151). We have:

$$b_{j,t} = \left(\frac{1}{N_j}\sum_{i=1}^{N_j}\lambda_{j,i}\lambda'_{j,i}\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T} h_{j,t}h'_{j,t}\right)^{-1} h_{j,t}\eta^2_{j,t}, \tag{B.157}$$

and:

$$u_{j,t} = \left(\frac{1}{N_j}\sum_{i=1}^{N_j}\lambda_{j,i}\lambda'_{j,i}\right)^{-1}\xi_{j,t},$$

where $\eta^2_{j,t}$ and $\xi_{j,t}$ are defined as in Assumption B.5. Then, we have:

$$\frac{1}{T}\sum_{t=1}^{T} b_{j,t}u'_{k,t} = \frac{1}{\sqrt{T}}\left(\frac{1}{N_j}\sum_{i=1}^{N_j}\lambda_{j,i}\lambda'_{j,i}\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T} h_{j,t}h'_{j,t}\right)^{-1}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\eta^2_{j,t}h_{j,t}\xi'_{k,t}\right]\left(\frac{1}{N_k}\sum_{i=1}^{N_k}\lambda_{k,i}\lambda'_{k,i}\right)^{-1}.$$

Now, $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\eta^2_{j,t}h_{j,t}\xi'_{k,t} = O_p(1)$ follows from the bound $\|\eta^2_{j,t}h_{j,t}\xi'_{k,t}\|_r \leq M$ with $r > 2$ (implied by Assumptions B.4 *a)-b)* and B.5 *b)* and Cauchy-Schwarz inequality), the mixing property with size $r/(r-2)$ in B.6 *a)*, and an application of Corollary 14.3 in Davidson (1994). Then, (B.151) follows.

The proofs of the other bounds are established by similar arguments and are omitted. ■

B.5.3 Proofs of Lemmas B.2, B.3 and B.4

The proofs of Lemma B.2, B.3 and B.4 are analogous to the proofs of Lemmas B.2, B.3 and B.4 in AGGR (see their Online Appendices C.5, C.6 and C.7), respectively, and therefore are omitted. ■

B.5.4 Proof of Lemma B.5

The proof is based on the asymptotic expansions of the terms within the trace operator in the r.h.s. of equation (B.22). We distinguish the terms that are first-order, resp. second-order, with respect to the $\hat{X}_{j,k}$.

i) Asymptotic expansion of first-order term $\hat{\Psi}^{*(I)}_{cc}$

From equation (B.14), we have $\hat{\Psi}^{*(I)}_{cc} = \left[-\hat{X}_{11}\tilde{R} + \hat{X}_{12}\tilde{B} - \tilde{B}'\hat{X}_{22}\tilde{B} + \tilde{B}'\hat{X}_{21} \right]^{(cc)}$. As matrices $\tilde{R}$ and $\tilde{B}$ have the same structure $[\ E_c \ \vdots \ * \ ]$ (see Lemma B.3), we have:

$$\hat{\Psi}^{*(I)}_{cc} = -\hat{X}^{(cc)}_{11} + \hat{X}^{(cc)}_{12} - \hat{X}^{(cc)}_{22} + \hat{X}^{(cc)}_{21}. \tag{B.158}$$

From the expressions of the matrices $\hat{X}_{j,k}$ in (B.11), and using the fact that upper k^c-dimensional subvector of both $\breve{h}_{1,t}$ and $\breve{h}_{2,t}$ is $\breve{f}^c_t$, the upper-left (k^c, k^c) blocks of the first and second matrices in the r.h.s. vanish. Therefore, from (B.158) we get:

$$\hat{\Psi}^{*(I)}_{cc} = -\frac{1}{T}\sum_{t=1}^{T}(\breve{\psi}^{(c)}_{1,t} - \breve{\psi}^{(c)}_{2,t})(\breve{\psi}^{(c)}_{1,t} - \breve{\psi}^{(c)}_{2,t})', \tag{B.159}$$

where $\breve{\psi}^{(c)}_{j,t}$ denotes the upper $(k^c, 1)$ block of vector $\breve{\psi}_{j,t}$. To compute the matrix in the r.h.s., we plug the expressions $\breve{\psi}_{j,t} = \frac{1}{\sqrt{N_j}}\breve{u}_{j,t} + \frac{1}{T}\breve{b}_{j,t} + \frac{1}{\sqrt{N_j T}}\breve{d}_{j,t} + \breve{\vartheta}_{j,t}$ for $j = 1, 2$ from (B.8), and use Proposition B.2 (**CHECK notation in PROPOSITION B.2**) and Assumptions B.1-B.4, B.5 *b)-c)* and B.6 *a)* to bound negligible terms up to $o_p(\epsilon_{N,T})$, where $\epsilon_{N,T} = (N\sqrt{T})^{-1}$.

LEMMA B.13. *Under Assumptions B.1-B.4, B.5 b)-c), B.6 a) and B.7 we have:*

$$\begin{aligned}
\hat{\Psi}^{*(I)}_{cc} \;=\;& -\frac{1}{N}\left(\frac{1}{T}\sum_{t=1}^{T}E[(\mu_N\breve{u}^{(c)}_{1t} - \breve{u}^{(c)}_{2t})(\mu_N\breve{u}^{(c)}_{1t} - \breve{u}^{(c)}_{2t})'|\mathcal{F}_t]\right) \\
& -\frac{1}{N\sqrt{T}}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[(\mu_N\breve{u}^{(c)}_{1t} - \breve{u}^{(c)}_{2t})(\mu_N\breve{u}^{(c)}_{1t} - \breve{u}^{(c)}_{2t})' - E[(\mu_N\breve{u}^{(c)}_{1t} - \breve{u}^{(c)}_{2t})(\mu_N\breve{u}^{(c)}_{1t} - \breve{u}^{(c)}_{2t})'|\mathcal{F}_t]\right]\right) \\
& -\frac{1}{T^2}\left(\frac{1}{T}\sum_{t=1}^{T}(\breve{b}^{(c)}_{1,t} - \breve{b}^{(c)}_{2,t})(\breve{b}^{(c)}_{1,t} - \breve{b}^{(c)}_{2,t})'\right) + o_p(\epsilon_{N,T}),
\end{aligned}$$

where the terms in the parentheses are $O_p(1)$.

Lemma B.13 shows that the leading stochastic terms in $\hat{\Psi}^{*(I)}_{cc}$ are of order $O_p\left(\frac{1}{N}\right)$, $O_p\left(\frac{1}{N\sqrt{T}}\right)$ and $O_p\left(\frac{1}{T^2}\right)$.

ii) Asymptotic expansion of the second-order terms in the r.h.s. of (B.22)

The asymptotic expansion of the second-order term $\hat{\Psi}^{*(II)}_{cc} - \frac{1}{4}\hat{\Psi}^{*(I)}_{cc}\tilde{\Sigma}^{-1}_{cc}\hat{\Psi}^{*(I)}_{cc} + \hat{\Psi}^{*(I)}_{cs}(I_{k_1-k^c} - \tilde{R}_{ss})^{-1}\hat{\Psi}^{(I)}_{sc} - \tilde{\Sigma}_{c,1}(I_{k_1-k^c} - \tilde{R}_{ss})^{-1}\hat{\Psi}^{(I)}_{sc}\tilde{\Sigma}^{-1}_{cc}\hat{\Psi}^{*(I)}_{cc}$ is provided in the next lemma.

LEMMA B.14. *Under Assumptions B.1-B.4, B.5 b)-c), B.6 a) and B.7 we have:*

$$
\begin{aligned}
& \hat{\Psi}_{cc}^{*(II)} - \frac{1}{4}\hat{\Psi}_{cc}^{*(I)}\tilde{\Sigma}_{cc}^{-1}\hat{\Psi}_{cc}^{*(I)} + \hat{\Psi}_{cs}^{*(I)}(I_{k_1-k^c} - \tilde{R}_{ss})^{-1}\hat{\Psi}_{sc}^{(I)} - \tilde{\Sigma}_{c,1}(I_{k_1-k^c} - \tilde{R}_{ss})^{-1}\hat{\Psi}_{sc}^{(I)}\tilde{\Sigma}_{cc}^{-1}\hat{\Psi}_{cc}^{*(I)} \\
= \; & \frac{1}{T^2}\left\{\left[\frac{1}{T}\sum_{t=1}^{T}\left(\breve{b}_{1,t}^{(c)} - \breve{b}_{2,t}^{(c)}\right)\breve{F}_t'\right]\tilde{\Sigma}_F^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}\breve{F}_t\left(\breve{b}_{1,t}^{(c)} - \breve{b}_{2,t}^{(c)}\right)'\right]\right\} + o_p(\epsilon_{N,T}),
\end{aligned}
$$

where $\tilde{\Sigma}_F = \frac{1}{T}\sum_{t=1}^{T}\breve{F}_t\breve{F}_t'$*, and the terms in the curly brackets are* $O_p(1)$.

From Lemmas B.13 and B.14, the asymptotic expansion of the term within the square brackets in the r.h.s of (B.22) is:

$$
\begin{aligned}
& \hat{\Psi}_{cc}^{*} - \frac{1}{4}\hat{\Psi}_{cc}^{*(I)}\tilde{\Sigma}_{cc}^{-1}\hat{\Psi}_{cc}^{*(I)} + \hat{\Psi}_{cs}^{*(I)}(I_{k_1-k^c} - \tilde{R}_{ss})^{-1}\hat{\Psi}_{sc}^{(I)} - \tilde{\Sigma}_{c,1}(I_{k_1-k^c} - \tilde{R}_{ss})^{-1}\hat{\Psi}_{sc}^{(I)}\tilde{\Sigma}_{cc}^{-1}\hat{\Psi}_{cc}^{*(I)} \\
= \; & -\frac{1}{N}\left(\frac{1}{T}\sum_{t=1}^{T}E[(\mu_N u_{1t}^{(c)} - u_{2t}^{(c)})(\mu_N u_{1t}^{(c)} - u_{2t}^{(c)})'|\mathcal{F}_t]\right) - \frac{1}{T^2}\left\{\frac{1}{T}\sum_{t=1}^{T}\widetilde{\Delta b}_t^{(c)}\widetilde{\Delta b}_t^{(c)\prime}\right\} \\
& -\frac{1}{N\sqrt{T}}\left\{\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[(\mu_N u_{1t}^{(c)} - u_{2t}^{(c)})(\mu_N u_{1t}^{(c)} - u_{2t}^{(c)})' - E[(\mu_N u_{1t}^{(c)} - u_{2t}^{(c)})(\mu_N u_{1t}^{(c)} - u_{2t}^{(c)})'|\mathcal{F}_t]\right]\right\} \\
& + o_p\left(\epsilon_{N,T}\right), \qquad \text{(B.160)}
\end{aligned}
$$

where $\widetilde{\Delta b}_t$ are the sample residuals defined in Theorem A.1.

Moreover, from Assumptions B.2, B.4 *b)* and B.6 *a)*, and Corollary 14.3 in Davidson (1994), we have:

$$
\tilde{V}_{jj} = I_{k_j} + O_p(T^{-1/2}), \quad j = 1, 2, \quad \tilde{V}_{12} = \begin{bmatrix} I_{k^c} & 0 \\ 0 & \Phi \end{bmatrix} + O_p(T^{-1/2}). \qquad \text{(B.161)}
$$

By plugging (B.160) into (B.22), and $\tilde{\Sigma}_{cc} = I_{k^c} + O_p(T^{-1/2})$ from (B.161), the conclusion follows. ■

B.5.4.1 Proof of Lemma B.13

We substitute the expressions $\breve{\psi}_{j,t} = \frac{1}{\sqrt{N_j}}\breve{u}_{j,t} + \frac{1}{T}\breve{b}_{j,t} + \frac{1}{\sqrt{N_j T}}\breve{d}_{j,t} + \breve{\vartheta}_{j,t}$ for $j = 1, 2$ into the r.h.s. of (B.159). We use $N_2 = N$ and $N_1 = N/\mu_N^2$, and partition vectors $\breve{u}_{j,t}$ and $\breve{b}_{j,t}$ in block-form as:

$$
\breve{u}_{j,t} = \begin{bmatrix} \breve{u}_{jt}^{(c)} \\ \breve{u}_{jt}^{(s)} \end{bmatrix}, \quad \breve{b}_{j,t} = \begin{bmatrix} \breve{b}_{jt}^{(c)} \\ \breve{b}_{jt}^{(s)} \end{bmatrix}, \;\; j = 1, 2.
$$

Moreover, we use that from Proposition B.2 the contribution of the remainder terms $\breve{\vartheta}_{j,t}$ in the r.h.s. of (B.159) is of order $o_p\left(\epsilon_{N,T}\right)$, and that under Assumptions B.2-B.4, B.5 *b)-c)* and B.6 *a)* we have $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\breve{u}_{j,t}\breve{d}_{k,t}' =$

$O_p(1)$ and $\frac{1}{T}\sum_{t=1}^{T} \breve{d}_{j,t}\breve{d}_{k,t}' = O_p(1)$ (see (B.154) and (B.156)). Therefore, we get:

$$
\begin{aligned}
\hat{\Psi}_{cc}^{*(I)} &= -\frac{1}{TN}\sum_{t=1}^{T}(\mu_N \breve{u}_{1,t}^{(c)} - \breve{u}_{2,t}^{(c)})(\mu_N \breve{u}_{1,t}^{(c)} - \breve{u}_{2,t}^{(c)})' \\
&\quad -\frac{1}{T^2\sqrt{N}}\sum_{t=1}^{T}\left[(\breve{b}_{1,t}^{(c)} - \breve{b}_{2,t}^{(c)})(\mu_N \breve{u}_{1,t}^{(c)} - \breve{u}_{2,t}^{(c)})' + (\mu_N \breve{u}_{1,t}^{(c)} - \breve{u}_{2,t}^{(c)})(\breve{b}_{1,t}^{(c)} - \breve{b}_{2,t}^{(c)})'\right] \\
&\quad -\frac{1}{T^3}\sum_{t=1}^{T}(\breve{b}_{1,t}^{(c)} - \breve{b}_{2,t}^{(c)})(\breve{b}_{1,t}^{(c)} - \breve{b}_{2,t}^{(c)})' \\
&\quad -\frac{1}{T^2\sqrt{NT}}\sum_{t=1}^{T}\left[(\breve{b}_{1,t}^{(c)} - \breve{b}_{2,t}^{(c)})(\mu_N \breve{d}_{1,t}^{(c)} - \breve{d}_{2,t}^{(c)})' + (\mu_N \breve{d}_{1,t}^{(c)} - \breve{d}_{2,t}^{(c)})(\breve{b}_{1,t}^{(c)} - \breve{b}_{2,t}^{(c)})'\right] + o_p(\epsilon_{N,T}).
\end{aligned}
$$

By recentering the first term in the r.h.s., and highlighting the convergence rates, we have:

$$
\begin{aligned}
\hat{\Psi}_{cc}^{*(I)} &= -\frac{1}{N}\left(\frac{1}{T}\sum_{t=1}^{T} E[(\mu_N \breve{u}_{1t}^{(c)} - \breve{u}_{2t}^{(c)})(\mu_N \breve{u}_{1t}^{(c)} - \breve{u}_{2t}^{(c)})'|\mathcal{F}_t]\right) \\
&\quad -\frac{1}{N\sqrt{T}}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[(\mu_N \breve{u}_{1t}^{(c)} - \breve{u}_{2t}^{(c)})(\mu_N \breve{u}_{1t}^{(c)} - \breve{u}_{2t}^{(c)})' - E[(\mu_N \breve{u}_{1t}^{(c)} - \breve{u}_{2t}^{(c)})(\mu_N \breve{u}_{1t}^{(c)} - \breve{u}_{2t}^{(c)})'|\mathcal{F}_t]\right]\right) \\
&\quad -\frac{1}{T\sqrt{NT}}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[(\breve{b}_{1,t}^{(c)} - \breve{b}_{2,t}^{(c)})(\mu_N \breve{u}_{1,t}^{(c)} - \breve{u}_{2,t}^{(c)})' + (\mu_N \breve{u}_{1,t}^{(c)} - \breve{u}_{2,t}^{(c)})(\breve{b}_{1,t}^{(c)} - \breve{b}_{2,t}^{(c)})'\right]\right) \\
&\quad -\frac{1}{T^2}\left(\frac{1}{T}\sum_{t=1}^{T}(\breve{b}_{1,t}^{(c)} - \breve{b}_{2,t}^{(c)})(\breve{b}_{1,t}^{(c)} - \breve{b}_{2,t}^{(c)})'\right) \\
&\quad -\frac{1}{T\sqrt{NT}}\left(\frac{1}{T}\sum_{t=1}^{T}\left[(\breve{b}_{1,t}^{(c)} - \breve{b}_{2,t}^{(c)})(\mu_N \breve{d}_{1,t}^{(c)} - \breve{d}_{2,t}^{(c)})' + (\mu_N \breve{d}_{1,t}^{(c)} - \breve{d}_{2,t}^{(c)})(\breve{b}_{1,t}^{(c)} - \breve{b}_{2,t}^{(c)})'\right]\right) + o_p(\epsilon_{N,T}).
\end{aligned}
\tag{B.162}
$$

Finally, by using ... and $N \ll T^2$ we get the expansion in Lemma B.13. ■

B.5.4.2 Proof of Lemma B.14

i) Asymptotic expansion of $\hat{\Psi}_{cc}^{*(II)}$

Let us start with $\hat{\Psi}_{cc}^{*(II)}$. From the definitions of the matrices $\hat{X}_{j,k}$ in equation (B.11), bounding the higher-order terms as in the proof of Lemma B.1, and using that $\frac{1}{T\sqrt{NT}} \leq \frac{1}{\sqrt{NT}} \leq \frac{1}{2}\left(\frac{1}{N} + \frac{1}{T^2}\right)$, we have:

$$
\hat{X}_{j,k} = \frac{1}{T}\tilde{\Xi}_{j,k} + \frac{1}{\sqrt{NT}}\hat{S}_{j,k} + O_p\left(\frac{1}{N} + \frac{1}{T^2}\right), \tag{B.163}
$$

where:

$$\tilde{\Xi}_{j,k} = \frac{1}{T}\sum_{t=1}^{T}(\breve{h}_{j,t}\breve{b}'_{k,t} + \breve{b}_{j,t}\breve{h}'_{k,t}), \tag{B.164}$$

$$\hat{S}_{j,k} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\mu_{N,k}\breve{h}_{j,t}\breve{u}'_{k,t} + \mu_{N,j}\breve{u}_{j,t}\breve{h}'_{k,t}) + \frac{1}{T}\sum_{t=1}^{T}(\mu_{N,k}\breve{h}_{j,t}\breve{d}'_{k,t} + \mu_{N,j}\breve{d}_{j,t}\breve{h}'_{k,t}), \tag{B.165}$$

with $\mu_{N,1} = \mu_N$ and $\mu_{N,2} = 1$. Terms $\tilde{\Xi}_{j,k}$ and $\hat{S}_{j,k}$ are $O_p(1)$ under Assumptions B.2-B.4, B.5 *b)-c)* and B.6 *a)*. Then, from the definition of $\hat{\Psi}^{*(II)}$ in (B.15), the bounds $\left(\frac{1}{T} + \frac{1}{\sqrt{NT}}\right)\left(\frac{1}{N} + \frac{1}{T^2}\right) = o(\epsilon_{N,T})$ and $\left(\frac{1}{N} + \frac{1}{T^2}\right)^2 = o(\epsilon_{N,T})$ which hold if $T^{1/2} \ll N \ll T^{5/2}$, we get:

$$\begin{aligned}
\hat{\Psi}^{*\,(II)} &= \frac{1}{T^2}\left\{-\tilde{\Xi}_{11}\tilde{V}_{11}^{-1}\left[-\tilde{\Xi}_{11}\tilde{R} + \tilde{\Xi}_{12}\tilde{B} - \tilde{B}'\tilde{\Xi}_{22}\tilde{B} + \tilde{B}'\tilde{\Xi}_{21}\right] + \left(\tilde{\Xi}_{22}\tilde{B} - \tilde{\Xi}_{21}\right)'\tilde{V}_{22}^{-1}\left(\tilde{\Xi}_{22}\tilde{B} - \tilde{\Xi}_{21}\right)\right\} \\
&\quad + \frac{1}{T\sqrt{NT}}\left\{-\tilde{\Xi}_{11}\tilde{V}_{11}^{-1}\left[-\hat{S}_{11}\tilde{R} + \hat{S}_{12}\tilde{B} - \tilde{B}'\hat{S}_{22}\tilde{B} + \tilde{B}'\hat{S}_{21}\right]\right. \\
&\qquad -\hat{S}_{11}\tilde{V}_{11}^{-1}\left[-\tilde{\Xi}_{11}\tilde{R} + \tilde{\Xi}_{12}\tilde{B} - \tilde{B}'\tilde{\Xi}_{22}\tilde{B} + \tilde{B}'\tilde{\Xi}_{21}\right] \\
&\qquad \left. + \left(\hat{S}_{22}\tilde{B} - \hat{S}_{21}\right)'\tilde{V}_{22}^{-1}\left(\tilde{\Xi}_{22}\tilde{B} - \tilde{\Xi}_{21}\right) + \left(\tilde{\Xi}_{22}\tilde{B} - \tilde{\Xi}_{21}\right)'\tilde{V}_{22}^{-1}\left(\hat{S}_{22}\tilde{B} - \hat{S}_{21}\right)\right\} \\
&\qquad + o_p(\epsilon_{N,T}).
\end{aligned}$$

Neglecting terms at order $o_p(\epsilon_{N,T})$ when we further assume $N \ll T^2$ we get:

$$\begin{aligned}
\hat{\Psi}^{*\,(II)} &= \frac{1}{T^2}\left\{-\tilde{\Xi}_{11}\tilde{V}_{11}^{-1}\left[-\tilde{\Xi}_{11}\tilde{R} + \tilde{\Xi}_{12}\tilde{B} - \tilde{B}'\tilde{\Xi}_{22}\tilde{B} + \tilde{B}'\tilde{\Xi}_{21}\right] + \left(\tilde{\Xi}_{22}\tilde{B} - \tilde{\Xi}_{21}\right)'\tilde{V}_{22}^{-1}\left(\tilde{\Xi}_{22}\tilde{B} - \tilde{\Xi}_{21}\right)\right\} \\
&\quad + o_p(\epsilon_{N,T}),
\end{aligned}$$

Let us now compute the (cc) block of this expansion. We get:

$$\begin{aligned}
\hat{\Psi}^{*\,(II)}_{cc} &= \frac{1}{T^2}\left\{-\tilde{\Xi}_{11}\tilde{V}_{11}^{-1}\left[-\tilde{\Xi}_{11}\tilde{R} + \tilde{\Xi}_{12}\tilde{B} - \tilde{B}'\tilde{\Xi}_{22}\tilde{B} + \tilde{B}'\tilde{\Xi}_{21}\right] + \left(\tilde{\Xi}_{22}\tilde{B} - \tilde{\Xi}_{21}\right)'\tilde{V}_{22}^{-1}\left(\tilde{\Xi}_{22}\tilde{B} - \tilde{\Xi}_{21}\right)\right\}_{(cc)} \\
&\qquad + o_p(\epsilon_{N,T}).
\end{aligned} \tag{B.166}$$

ii) Asymptotic expansion of $\hat{\Psi}_{cs}^{*(I)}(I_{k_1-k^c}-\tilde{R}_{ss})^{-1}\hat{\Psi}_{sc}^{(I)}$

Let us now consider the term $\hat{\Psi}_{cs}^{*(I)}(I_{k_1-k^c}-\tilde{R}_{ss})^{-1}\hat{\Psi}_{sc}^{(I)}$. By the formula of the partitioned inverse for $\tilde{V}_{11}^{-1}$, and Lemmas B.1 and B.13, we have:

$$\begin{aligned}
&\hat{\Psi}_{cs}^{*(I)}(I_{k_1-k^c}-\tilde{R}_{ss})^{-1}\hat{\Psi}_{sc}^{(I)}\\
=\;&\hat{\Psi}_{cs}^{*(I)}(I_{k_1-k^c}-\tilde{R}_{ss})^{-1}\left[(\tilde{V}_{11}^{-1})_{ss}\hat{\Psi}_{sc}^{*(I)}+O_p\left(T^{-1/2}\hat{\Psi}_{cc}^{*(I)}\right)\right]\\
=\;&\hat{\Psi}_{cs}^{*(I)}(I_{k_1-k^c}-\tilde{R}_{ss})^{-1}(\tilde{V}_{11}^{-1})_{ss}\hat{\Psi}_{sc}^{*(I)}+O_p\left(\delta_{N,T}\frac{1}{\sqrt{T}}\left(\frac{1}{N}+\frac{1}{T^2}+\frac{1}{T\sqrt{NT}}+\epsilon_{N,T}\right)\right)\\
=\;&\hat{\Psi}_{cs}^{*(I)}(I_{k_1-k^c}-\tilde{R}_{ss})^{-1}(\tilde{V}_{11}^{-1})_{ss}\hat{\Psi}_{sc}^{*(I)}+o_p(\epsilon_{N,T}),
\end{aligned}\tag{B.167}$$

if $N \ll T^{5/2}$. Let us consider $\hat{\Psi}_{cs}^{*(I)}(I_{k_1-k^c}-\tilde{R}_{ss})^{-1}(\tilde{V}_{11}^{-1})_{ss}\hat{\Psi}_{sc}^{*(I)}$. By using $\hat{\Psi}^{*(I)}=-\hat{X}_{11}\tilde{R}+\hat{X}_{12}\tilde{B}-\tilde{B}'\hat{X}_{22}\tilde{B}+\tilde{B}'\hat{X}_{21}$, the expansion for $\hat{X}_{j,k}$ in (B.163), $\tilde{R}_{ss}=\Phi\Phi'+o_p(1)$, and the condition $T^{1/2}\ll N\ll T^{4/2}$ to control negligible terms, we get:

$$\begin{aligned}
&\hat{\Psi}_{cs}^{*(I)}(I_{k_1-k^c}-\tilde{R}_{ss})^{-1}(\tilde{V}_{11}^{-1})_{ss}\hat{\Psi}_{sc}^{*(I)}\\
=\;&\frac{1}{T^2}\left\{\left[-\tilde{\Xi}_{11}\tilde{R}+\tilde{\Xi}_{12}\tilde{B}-\tilde{B}'\tilde{\Xi}_{22}\tilde{B}+\tilde{B}'\tilde{\Xi}_{21}\right]_{cs}(I_{k_1-k^c}-\tilde{R}_{ss})^{-1}(\tilde{V}_{11}^{-1})_{ss}\left[-\tilde{\Xi}_{11}\tilde{R}+\tilde{\Xi}_{12}\tilde{B}-\tilde{B}'\tilde{\Xi}_{22}\tilde{B}+\tilde{B}'\tilde{\Xi}_{21}\right]_{sc}\right\}\\
&+o_p(\epsilon_{N,T}).
\end{aligned}$$

iii) Asymptotic expansion of $\hat{\Psi}_{cc}^{*\,(II)}+\hat{\Psi}_{cs}^{*(I)}(I_{k_1-k^c}-\tilde{R}_{ss})^{-1}\hat{\Psi}_{sc}^{(I)}$

By putting the expansions (B.166) and (B.167) together, we get the asymptotic expansion:

$$\begin{aligned}
&\hat{\Psi}_{cc}^{*\,(II)}+\hat{\Psi}_{cs}^{*(I)}(I_{k_1-k^c}-\tilde{R}_{ss})^{-1}\hat{\Psi}_{sc}^{(I)}\\
=\;&\frac{1}{T^2}\Bigg\{\left(-\tilde{\Xi}_{11}\tilde{V}_{11}^{-1}\left[-\tilde{\Xi}_{11}\tilde{R}+\tilde{\Xi}_{12}\tilde{B}-\tilde{B}'\tilde{\Xi}_{22}\tilde{B}+\tilde{B}'\tilde{\Xi}_{21}\right]+\left(\tilde{\Xi}_{22}\tilde{B}-\tilde{\Xi}_{21}\right)'\tilde{V}_{22}^{-1}\left(\tilde{\Xi}_{22}\tilde{B}-\tilde{\Xi}_{21}\right)\right)_{cc}\\
&\quad+\left[-\tilde{\Xi}_{11}\tilde{R}+\tilde{\Xi}_{12}\tilde{B}-\tilde{B}'\tilde{\Xi}_{22}\tilde{B}+\tilde{B}'\tilde{\Xi}_{21}\right]_{cs}(I_{k_1-k^c}-\tilde{R}_{ss})^{-1}(\tilde{V}_{11}^{-1})_{ss}\times\\
&\quad\left[-\tilde{\Xi}_{11}\tilde{R}+\tilde{\Xi}_{12}\tilde{B}-\tilde{B}'\tilde{\Xi}_{22}\tilde{B}+\tilde{B}'\tilde{\Xi}_{21}\right]_{sc}\Bigg\}+o_p(\epsilon_{N,T}).
\end{aligned}\tag{B.168}$$

Let us now rework the term at order T^{-2}. For this purpose we use the equations:

$$\begin{aligned}
\left[-\tilde{\Xi}_{11}\tilde{R}+\tilde{\Xi}_{12}\tilde{B}-\tilde{B}'\tilde{\Xi}_{22}\tilde{B}+\tilde{B}'\tilde{\Xi}_{21}\right]_{cc}&=-\tilde{\Xi}_{11,cc}+\tilde{\Xi}_{12,cc}-\tilde{\Xi}_{22,cc}+\tilde{\Xi}_{21,cc}=0,\\
\left[-\tilde{\Xi}_{11}\tilde{R}+\tilde{\Xi}_{12}\tilde{B}-\tilde{B}'\tilde{\Xi}_{22}\tilde{B}+\tilde{B}'\tilde{\Xi}_{21}\right]_{sc}&=-\tilde{\Xi}_{11,sc}+\tilde{\Xi}_{12,sc}-\tilde{B}_{cs}'\tilde{\Xi}_{22,cc}\\
&\quad-\tilde{B}_{ss}'\tilde{\Xi}_{22,sc}+\tilde{B}_{cs}'\tilde{\Xi}_{21,cc}+\tilde{B}_{ss}'\tilde{\Xi}_{21,sc},\\
\left[-\tilde{\Xi}_{11}\tilde{R}+\tilde{\Xi}_{12}\tilde{B}-\tilde{B}'\tilde{\Xi}_{22}\tilde{B}+\tilde{B}'\tilde{\Xi}_{21}\right]_{cs}&=-\tilde{\Xi}_{11,cc}\tilde{R}_{cs}-\tilde{\Xi}_{11,cs}\tilde{R}_{ss}+\tilde{\Xi}_{12,cc}\tilde{B}_{cs}+\tilde{\Xi}_{12,cs}\tilde{B}_{ss}\\
&\quad-\tilde{\Xi}_{22,cc}\tilde{B}_{cs}-\tilde{\Xi}_{22,cs}\tilde{B}_{ss}+\tilde{\Xi}_{21,cs}.
\end{aligned}$$

Then, a block product computation yields:

$$
\begin{aligned}
& \left(-\tilde{\Xi}_{11}\tilde{V}_{11}^{-1}\left[-\tilde{\Xi}_{11}\tilde{R}+\tilde{\Xi}_{12}\tilde{B}-\tilde{B}'\tilde{\Xi}_{22}\tilde{B}+\tilde{B}'\tilde{\Xi}_{21}\right]\right)_{cc} \\
& \qquad +\left[-\tilde{\Xi}_{11}\tilde{R}+\tilde{\Xi}_{12}\tilde{B}-\tilde{B}'\tilde{\Xi}_{22}\tilde{B}+\tilde{B}'\tilde{\Xi}_{21}\right]_{cs}(I_{k_1-k^c}-\tilde{R}_{ss})^{-1}(\tilde{V}_{11}^{-1})_{ss}\times \\
& \qquad \left[-\tilde{\Xi}_{11}\tilde{R}+\tilde{\Xi}_{12}\tilde{B}-\tilde{B}'\tilde{\Xi}_{22}\tilde{B}+\tilde{B}'\tilde{\Xi}_{21}\right]_{sc} \\
= \;& -\left[\tilde{\Xi}_{11,cc}(\tilde{V}_{11}^{-1})_{cs}+\tilde{\Xi}_{11,cs}(\tilde{V}_{11}^{-1})_{ss}\right]\times \\
& \qquad \left[-\tilde{\Xi}_{11,sc}+\tilde{\Xi}_{12,sc}-\tilde{B}'_{cs}\tilde{\Xi}_{22,cc}-\tilde{B}'_{ss}\tilde{\Xi}_{22,sc}+\tilde{B}'_{cs}\tilde{\Xi}_{21,cc}+\tilde{B}'_{ss}\tilde{\Xi}_{21,sc}\right] \\
& +\left[-\tilde{\Xi}_{11,cc}\tilde{R}_{cs}-\tilde{\Xi}_{11,cs}\tilde{R}_{ss}+\tilde{\Xi}_{12,cc}\tilde{B}_{cs}+\tilde{\Xi}_{12,cs}\tilde{B}_{ss}-\tilde{\Xi}_{22,cc}\tilde{B}_{cs}-\tilde{\Xi}_{22,cs}\tilde{B}_{ss}+\tilde{\Xi}_{21,cs}\right] \\
& \times(I_{k_1-k^c}-\tilde{R}_{ss})^{-1}(\tilde{V}_{11}^{-1})_{ss}\left[-\tilde{\Xi}_{11,sc}+\tilde{\Xi}_{12,sc}-\tilde{B}'_{cs}\tilde{\Xi}_{22,cc}-\tilde{B}'_{ss}\tilde{\Xi}_{22,sc}+\tilde{B}'_{cs}\tilde{\Xi}_{21,cc}+\tilde{B}'_{ss}\tilde{\Xi}_{21,sc}\right] \\
= \;& \left[-\tilde{\Xi}_{11,cc}\left((\tilde{V}_{11}^{-1})_{cs}(\tilde{V}_{11}^{-1})_{ss}^{-1}(I_{k_1-k^c}-\tilde{R}_{ss})+\tilde{R}_{cs}\right)\right. \\
& \qquad \left.-\tilde{\Xi}_{11,cs}+\tilde{\Xi}_{12,cc}\tilde{B}_{cs}+\tilde{\Xi}_{12,cs}\tilde{B}_{ss}-\tilde{\Xi}_{22,cc}\tilde{B}_{cs}-\tilde{\Xi}_{22,cs}\tilde{B}_{ss}+\tilde{\Xi}_{21,cs}\right] \\
& \times(I_{k_1-k^c}-\tilde{R}_{ss})^{-1}(\tilde{V}_{11}^{-1})_{ss}\left[-\tilde{\Xi}_{11,sc}+\tilde{\Xi}_{12,sc}-\tilde{B}'_{cs}\tilde{\Xi}_{22,cc}-\tilde{B}'_{ss}\tilde{\Xi}_{22,sc}+\tilde{B}'_{cs}\tilde{\Xi}_{21,cc}+\tilde{B}'_{ss}\tilde{\Xi}_{21,sc}\right].
\end{aligned}
$$

Let us show that the term $(\tilde{V}_{11}^{-1})_{cs}(\tilde{V}_{11}^{-1})_{ss}^{-1}(I_{k_1-k^c}-\tilde{R}_{ss})+\tilde{R}_{cs}$ vanishes. Indeed, from equation (**??**) we have:

$$
\begin{aligned}
(\tilde{V}_{11}^{-1})_{cs}(\tilde{V}_{11}^{-1})_{ss}^{-1}(I_{k_1-k^c}-\tilde{R}_{ss})+\tilde{R}_{cs} &= \left[(\tilde{V}_{11}^{-1})_{cs}(\tilde{V}_{11}^{-1})_{ss}^{-1}+\tilde{R}_{cs}(I_{k_1-k^c}-\tilde{R}_{ss})^{-1}\right](I_{k_1-k^c}-\tilde{R}_{ss}) \\
&= \left[(\tilde{V}_{11}^{-1})_{cs}(\tilde{V}_{11}^{-1})_{ss}^{-1}+\tilde{\Sigma}_{cc}^{-1}\tilde{\Sigma}_{c,1}\right](I_{k_1-k^c}-\tilde{R}_{ss}) \\
&= \tilde{\Sigma}_{cc}^{-1}\left[\tilde{\Sigma}_{cc}(\tilde{V}_{11}^{-1})_{cs}+\tilde{\Sigma}_{c,1}(\tilde{V}_{11}^{-1})_{ss}\right](\tilde{V}_{11}^{-1})_{ss}^{-1}(I_{k_1-k^c}-\tilde{R}_{ss}) \\
&= \tilde{\Sigma}_{cc}^{-1}\left[(\tilde{V}_{11})_{cc}(\tilde{V}_{11}^{-1})_{cs}+(\tilde{V}_{11})_{cs}(\tilde{V}_{11}^{-1})_{ss}\right](\tilde{V}_{11}^{-1})_{ss}^{-1}(I_{k_1-k^c}-\tilde{R}_{ss}) \\
&= 0.
\end{aligned}
$$

Therefore, we get:

$$
\begin{aligned}
& \left(-\tilde{\Xi}_{11}\tilde{V}_{11}^{-1}\left[-\tilde{\Xi}_{11}\tilde{R}+\tilde{\Xi}_{12}\tilde{B}-\tilde{B}'\tilde{\Xi}_{22}\tilde{B}+\tilde{B}'\tilde{\Xi}_{21}\right]\right)_{cc} \\
& \qquad +\left[-\tilde{\Xi}_{11}\tilde{R}+\tilde{\Xi}_{12}\tilde{B}-\tilde{B}'\tilde{\Xi}_{22}\tilde{B}+\tilde{B}'\tilde{\Xi}_{21}\right]_{cs}(I_{k_1-k^c}-\tilde{R}_{ss})^{-1}(\tilde{V}_{11}^{-1})_{ss}\times \\
& \qquad \left[-\tilde{\Xi}_{11}\tilde{R}+\tilde{\Xi}_{12}\tilde{B}-\tilde{B}'\tilde{\Xi}_{22}\tilde{B}+\tilde{B}'\tilde{\Xi}_{21}\right]_{sc} \\
= \;& \left[-\tilde{\Xi}_{11,cs}+\tilde{\Xi}_{12,cc}\tilde{B}_{cs}+\tilde{\Xi}_{12,cs}\tilde{B}_{ss}-\tilde{\Xi}_{22,cc}\tilde{B}_{cs}-\tilde{\Xi}_{22,cs}\tilde{B}_{ss}+\tilde{\Xi}_{21,cs}\right] \\
& \times(I_{k_1-k^c}-\tilde{R}_{ss})^{-1}(\tilde{V}_{11}^{-1})_{ss}\left[-\tilde{\Xi}_{11,sc}+\tilde{\Xi}_{12,sc}-\tilde{B}'_{cs}\tilde{\Xi}_{22,cc}-\tilde{B}'_{ss}\tilde{\Xi}_{22,sc}+\tilde{B}'_{cs}\tilde{\Xi}_{21,cc}+\tilde{B}'_{ss}\tilde{\Xi}_{21,sc}\right] \\
= \;& \left[-\tilde{\Xi}_{11,sc}+\tilde{\Xi}_{12,sc}-\tilde{B}'_{cs}\tilde{\Xi}_{22,cc}-\tilde{B}'_{ss}\tilde{\Xi}_{22,sc}+\tilde{B}'_{cs}\tilde{\Xi}_{21,cc}+\tilde{B}'_{ss}\tilde{\Xi}_{21,sc}\right]' \\
& \times(I_{k_1-k^c}-\tilde{R}_{ss})^{-1}(\tilde{V}_{11}^{-1})_{ss}\left[-\tilde{\Xi}_{11,sc}+\tilde{\Xi}_{12,sc}-\tilde{B}'_{cs}\tilde{\Xi}_{22,cc}-\tilde{B}'_{ss}\tilde{\Xi}_{22,sc}+\tilde{B}'_{cs}\tilde{\Xi}_{21,cc}+\tilde{B}'_{ss}\tilde{\Xi}_{21,sc}\right].
\end{aligned}
$$

Let us consider the term $-\tilde{\Xi}_{11,sc}$ + $\tilde{\Xi}_{12,sc}$ - $\tilde{B}'_{cs}\tilde{\Xi}_{22,cc}$ - $\tilde{B}'_{ss}\tilde{\Xi}_{22,sc}$ + $\tilde{B}'_{cs}\tilde{\Xi}_{21,cc}$ + $\tilde{B}'_{ss}\tilde{\Xi}_{21,sc}$ = $-\left[(\tilde{\Xi}_{11,sc}-\tilde{\Xi}_{12,sc})-\tilde{B}'_{cs}(\tilde{\Xi}_{21,cc}-\tilde{\Xi}_{22,cc})-\tilde{B}'_{ss}(\tilde{\Xi}_{21,sc}-\tilde{\Xi}_{22,sc})\right]$.
Using $\tilde{\Xi}_{11,sc}-\tilde{\Xi}_{12,sc}=\frac{1}{T}\sum_t \breve{f}^s_{1,t}\left(\breve{b}^{(c)}_{1,t}-\breve{b}^{(c)}_{2,t}\right)'$,
$\tilde{\Xi}_{21,cc}-\tilde{\Xi}_{22,cc}=\frac{1}{T}\sum_t \breve{f}^c_t\left(\breve{b}^{(c)}_{1,t}-\breve{b}^{(c)}_{2,t}\right)'$ and
$\tilde{\Xi}_{21,sc}-\tilde{\Xi}_{22,sc}=\frac{1}{T}\sum_t \breve{f}^s_{2,t}\left(\breve{b}^{(c)}_{1,t}-\breve{b}^{(c)}_{2,t}\right)'$, we can write it as:

$$\begin{aligned}&-\tilde{\Xi}_{11,sc}+\tilde{\Xi}_{12,sc}-\tilde{B}'_{cs}\tilde{\Xi}_{22,cc}-\tilde{B}'_{ss}\tilde{\Xi}_{22,sc}+\tilde{B}'_{cs}\tilde{\Xi}_{21,cc}+\tilde{B}'_{ss}\tilde{\Xi}_{21,sc}\\&\qquad=-\frac{1}{T}\sum_{t=1}^{T}\left[\breve{f}^s_{1,t}-\tilde{B}'_{cs}\breve{f}^c_t-\tilde{B}'_{ss}\breve{f}^s_{2,t}\right]\left(\breve{b}^{(c)}_{1,t}-\breve{b}^{(c)}_{2,t}\right)'.\end{aligned}$$

Noting that

$$\tilde{B}'=\tilde{V}_{12}\tilde{V}_{22}^{-1}=\begin{bmatrix} I & 0\\ \tilde{B}'_{cs} & \tilde{B}'_{ss}\end{bmatrix},$$

we deduce that:

$$\tilde{f}_{1\perp 2c,t}=\breve{f}^s_{1,t}-\tilde{B}'_{cs}\breve{f}^c_t-\tilde{B}'_{ss}\breve{f}^s_{2,t},\qquad t=1,...,T,$$

are the residuals in the sample orthogonal projection of $\breve{f}^s_{1,t}$ on $\breve{f}^s_{2,t}$ and $\breve{f}^c_t$. Let us now show that $(I_{k_1-k^c}-\tilde{R}_{ss})^{-1}(\tilde{V}_{11}^{-1})_{ss}$ is the inverse of the sample variance of that residuals. Indeed, the sample variance is:

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T}\tilde{f}_{1\perp 2c,t}\tilde{f}'_{1\perp 2c,t} &= \frac{1}{T}\sum_{t=1}^{T}\left[\breve{f}^s_{1,t}-\tilde{B}'_{cs}\breve{f}^c_t-\tilde{B}'_{ss}\breve{f}_{2,t}\right]\breve{f}^{s\prime}_{1,t}\\ &= \tilde{\Sigma}_{11}-\tilde{B}'_{cs}\tilde{\Sigma}_{c,1}-\tilde{B}'_{ss}\tilde{\Sigma}_{2,1}=\left(\tilde{V}_{11}-\tilde{B}'\tilde{V}_{21}\right)_{ss}\\ &= \left(\tilde{V}_{11}(I_{k_1}-\tilde{R})\right)_{ss}\\ &= -\tilde{\Sigma}_{1c}\tilde{R}_{cs}+\tilde{\Sigma}_{11}(I_{k_1-k^c}-R_{ss})\\ &= \left[-\tilde{\Sigma}_{1c}\tilde{R}_{cs}(I_{k_1-k^c}-R_{ss})^{-1}+\tilde{\Sigma}_{11}\right](I_{k_1-k^c}-R_{ss})\\ &= \left(-\tilde{\Sigma}_{1c}\tilde{\Sigma}_{cc}^{-1}\tilde{\Sigma}_{c1}+\tilde{\Sigma}_{11}\right)(I_{k_1-k^c}-R_{ss})=[(\tilde{V}_{11}^{-1})_{ss}]^{-1}(I_{k_1-k^c}-R_{ss}),\end{aligned}$$

from Equation (C.67) in the OA of AGGR . By gathering these results, we get:

$$\begin{aligned}&\left(-\tilde{\Xi}_{11}\tilde{V}_{11}^{-1}\left[-\tilde{\Xi}_{11}\tilde{R}+\tilde{\Xi}_{12}\tilde{B}-\tilde{B}'\tilde{\Xi}_{22}\tilde{B}+\tilde{B}'\tilde{\Xi}_{21}\right]\right)_{cc}\\&\quad+\left[-\tilde{\Xi}_{11}\tilde{R}+\tilde{\Xi}_{12}\tilde{B}-\tilde{B}'\tilde{\Xi}_{22}\tilde{B}+\tilde{B}'\tilde{\Xi}_{21}\right]_{cs}(I_{k_1-k^c}-\tilde{R}_{ss})^{-1}(\tilde{V}_{11}^{-1})_{ss}\times\\&\qquad\left[-\tilde{\Xi}_{11}\tilde{R}+\tilde{\Xi}_{12}\tilde{B}-\tilde{B}'\tilde{\Xi}_{22}\tilde{B}+\tilde{B}'\tilde{\Xi}_{21}\right]_{sc}\\ &= \left[\frac{1}{T}\sum_{t=1}^{T}\left(\breve{b}^{(c)}_{1,t}-\breve{b}^{(c)}_{2,t}\right)\tilde{f}'_{1\perp 2c,t}\right]\left(\frac{1}{T}\sum_{t=1}^{T}\tilde{f}_{1\perp 2c,t}\tilde{f}'_{1\perp 2c,t}\right)^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}\tilde{f}_{1\perp 2c,t}\left(\breve{b}^{(c)}_{1,t}-\breve{b}^{(c)}_{2,t}\right)'\right].\end{aligned}$$

Let us now consider the term $\left[\left(\tilde{\Xi}_{22}\tilde{B}-\tilde{\Xi}_{21}\right)'\tilde{V}_{22}^{-1}\left(\tilde{\Xi}_{22}\tilde{B}-\tilde{\Xi}_{21}\right)\right]_{cc}$ also showing at order T^{-2} in the r.h.s. of the asymptotic expansion (B.168). Direct computation yields:

$$
\begin{aligned}
&\left[\left(\tilde{\Xi}_{22}\tilde{B}-\tilde{\Xi}_{21}\right)'\tilde{V}_{22}^{-1}\left(\tilde{\Xi}_{22}\tilde{B}-\tilde{\Xi}_{21}\right)\right]_{cc} \\
&\qquad = \left[(\tilde{\Xi}_{22,cc}-\tilde{\Xi}_{21,cc})' \vdots (\tilde{\Xi}_{22,sc}-\tilde{\Xi}_{21,sc})'\right]\tilde{V}_{22}^{-1}\begin{bmatrix}\tilde{\Xi}_{22,cc}-\tilde{\Xi}_{21,cc}\\ \tilde{\Xi}_{22,sc}-\tilde{\Xi}_{21,sc}\end{bmatrix} \\
&\qquad = \left[\frac{1}{T}\sum_{t=1}^{T}\left(\breve{b}_{1,t}^{(c)}-\breve{b}_{2,t}^{(c)}\right)\breve{h}_{2,t}'\right]\left(\frac{1}{T}\sum_{t=1}^{T}\breve{h}_{2,t}\breve{h}_{2,t}'\right)^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}\breve{h}_{2,t}\left(\breve{b}_{1,t}^{(c)}-\breve{b}_{2,t}^{(c)}\right)'\right].
\end{aligned}
$$

Hence, the term at order T^{-2} in the r.h.s. of (B.168) becomes:

$$
\begin{aligned}
&\left(-\tilde{\Xi}_{11}\tilde{V}_{11}^{-1}\left[-\tilde{\Xi}_{11}\tilde{R}+\tilde{\Xi}_{12}\tilde{B}-\tilde{B}'\tilde{\Xi}_{22}\tilde{B}+\tilde{B}'\tilde{\Xi}_{21}\right]+\left(\tilde{\Xi}_{22}\tilde{B}-\tilde{\Xi}_{21}\right)'\tilde{V}_{22}^{-1}\left(\tilde{\Xi}_{22}\tilde{B}-\tilde{\Xi}_{21}\right)\right)_{cc} \\
&\quad+\left[-\tilde{\Xi}_{11}\tilde{R}+\tilde{\Xi}_{12}\tilde{B}-\tilde{B}'\tilde{\Xi}_{22}\tilde{B}+\tilde{B}'\tilde{\Xi}_{21}\right]_{cs}(I_{k_1-k^c}-\tilde{R}_{ss})^{-1}(\tilde{V}_{11}^{-1})_{ss}\times \\
&\quad\left[-\tilde{\Xi}_{11}\tilde{R}+\tilde{\Xi}_{12}\tilde{B}-\tilde{B}'\tilde{\Xi}_{22}\tilde{B}+\tilde{B}'\tilde{\Xi}_{21}\right]_{sc} \\
=\;&\left[\frac{1}{T}\sum_{t=1}^{T}\left(\breve{b}_{1,t}^{(c)}-\breve{b}_{2,t}^{(c)}\right)\breve{h b}_{2,t}'\right]\left(\frac{1}{T}\sum_{t=1}^{T}\breve{h}_{2,t}\breve{h}_{2,t}'\right)^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}\breve{h}_{2,t}\left(\breve{b}_{1,t}^{(c)}-\breve{b}_{2,t}^{(c)}\right)'\right] \\
&+\left[\frac{1}{T}\sum_{t=1}^{T}\left(\breve{b}_{1,t}^{(c)}-\breve{b}_{2,t}^{(c)}\right)\tilde{f}_{1\perp 2c,t}'\right]\left(\frac{1}{T}\sum_{t=1}^{T}\tilde{f}_{1\perp 2c,t}\tilde{f}_{1\perp 2c,t}'\right)^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}\tilde{f}_{1\perp 2,c}\left(\breve{b}_{1,t}^{(c)}-\breve{b}_{2,t}^{(c)}\right)'\right] \\
=\;&\left[\frac{1}{T}\sum_{t=1}^{T}\left(\breve{b}_{1,t}^{(c)}-\breve{b}_{2,t}^{(c)}\right)\breve{F}_t'\right]\tilde{\Sigma}_F^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}\breve{F}_t\left(\breve{b}_{1,t}^{(c)}-\breve{b}_{2,t}^{(c)}\right)'\right],
\end{aligned}
\tag{B.169}
$$

where:

$$
\tilde{\Sigma}_F=\frac{1}{T}\sum_{t=1}^{T}\breve{F}_t\breve{F}_t',
$$

because $\tilde{f}_{1\perp 2c,t}$ is orthogonal in-sample to $\breve{h}_{2,t}$, and $(\tilde{f}_{1\perp 2,c}', h_{2,t}')'$ is a linear transformation of $(\breve{f}_t^{c\prime}, \breve{f}_{1,t}^{s\prime}, \breve{f}_{2,t}^{s\prime})'$.

By substituting (B.169) into (B.168), we get:

$$
\begin{aligned}
&\hat{\Psi}_{cc}^{*\,(II)}+\hat{\Psi}_{cs}^{*(I)}(I_{k_1-k^c}-\tilde{R}_{ss})^{-1}\hat{\Psi}_{sc}^{(I)} \\
=\;&\frac{1}{T^2}\left\{\left[\frac{1}{T}\sum_{t=1}^{T}\left(b_{1,t}^{(c)}-b_{2,t}^{(c)}\right)F_t'\right]\tilde{\Sigma}_F^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}F_t\left(b_{1,t}^{(c)}-b_{2,t}^{(c)}\right)'\right]\right\}+o_p(\epsilon_{N,T}).
\end{aligned}
\tag{B.170}
$$

iv) Conclusion

We finally consider the other second-order terms in the r.h.s. of (B.22).

By $\hat{\Psi}_{cc}^{*(I)} = O_p\left(\frac{1}{N} + \frac{1}{T^2} + \epsilon_{N,T}\right)$ from Lemma B.13, we have:

$$\hat{\Psi}_{cc}^{*(I)} \tilde{\Sigma}_{cc}^{-1} \hat{\Psi}_{cc}^{*(I)} = o_p(\epsilon_{N,T}), \tag{B.171}$$

if $T^{1/2} \ll N \ll T^2$. Moreover, by using $\tilde{\Sigma}_{c,1}(I_{k_1-k^c} - \tilde{R}_{ss})^{-1} = O_p(T^{-1/2})$ from (B.161), $\hat{\Psi}_{sc}^{(I)} = O_p(\delta_{N,T})$, and $\hat{\Psi}_{cc}^{*(I)} = O_p\left(\frac{1}{N} + \frac{1}{T^2} + \epsilon_{N,T}\right)$, we have:

$$\begin{aligned} \tilde{\Sigma}_{c,1}(I_{k_1-k^c} - \tilde{R}_{ss})^{-1} \hat{\Psi}_{sc}^{(I)} \tilde{\Sigma}_{cc}^{-1} \hat{\Psi}_{cc}^{*(I)} &= O_p\left[\frac{1}{\sqrt{T}} \delta_{N,T} \left(\frac{1}{N} + \frac{1}{T^2} + \epsilon_{N,T}\right)\right] \\ &= o_p(\epsilon_{N,T}), \end{aligned} \tag{B.172}$$

if $T^{1/2} \ll N \ll T^2$. From (B.170), (B.171) and (B.172), the conclusion follows. ■

B.5.5 Proof of Lemma B.6

We show the conditions in parts (i)-(iv) of Lemma B.6. Part (i) follows by the Law of Iterated Expectation and $E(U_t|\mathcal{F}_t) = 0$, which is implied by Assumption B.4 *a)*. Part (ii) is implied by Assumptions B.3, B.4 *b)* and B.5 *b)*. The NED property in part (iii) holds true because conditional expectations given $\mathcal{F}_t$ can be well approximated by elements in the sigma-field $\mathcal{V}_{t-m}^{t+m}$ generated by the mixing process (V_t), for large m, by Assumptions B.3, B.4 *b)*, B.5 *b)* and B.6 *a)-c)*, as we show in the next lemma.

LEMMA B.15. *Assumptions B.3, B.4 b), B.5 b) and B.6 a)-c) imply part (iii) in Lemma B.6.*

To check part (iv) in Lemma B.6 we use:

$$\begin{aligned} \lim_{T,N\to\infty} V\left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \mathcal{Z}_{N,t}\right) &= \lim_{T,N\to\infty} \frac{1}{T} \sum_{h=-T+1}^{T-1} (T - |h|) Cov\left(\mathcal{Z}_{N,t}, \mathcal{Z}_{N,t-h}\right) \\ &= \lim_{N\to\infty} \sum_{h=-\infty}^{\infty} Cov\left(\mathcal{Z}_{N,t}, \mathcal{Z}_{N,t-h}\right), \end{aligned}$$

where the first equality follows from stationarity of the data. The series converges because the zero-mean process $\mathcal{Z}_{N,t}$ is a L^2-mixingale with size -1, [18] by Theorem 17.5 in Davidson (1994) and Conditions (ii)-(iii), which implies $\|Cov\left(\mathcal{Z}_{N,t}, \mathcal{Z}_{N,t-h}\right)\| = \left\|E\left[E(\mathcal{Z}_{N,t}|\mathcal{V}_{t-h})\mathcal{Z}'_{N,t-h}\right]\right\| \leq \|E(\mathcal{Z}_{N,t}|\mathcal{V}_{t-h})\|_2 \|\mathcal{Z}_{N,t-h}\|_2 = O\left(h^{-\psi}\right)$, uniformly in $N_1, N_2 \geq 1$, for some $\psi > 1$. The latter uniform bound also allows for an application of the Lebesgue Lemma to get:

$$\Omega_U = \lim_{T,N\to\infty} V\left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \mathcal{Z}_{N,t}\right) = \sum_{h=-\infty}^{\infty} \Gamma(h),$$

where $\Gamma(h) = \lim_{N\to\infty} Cov\left(\mathcal{Z}_{N,t}, \mathcal{Z}_{N,t-h}\right)$, which yields equation (B.26). The computations in Subsection B.3.6, and in particular Lemma B.7, show that the limit in $\Gamma(h)$ is well-defined.

[18] That is, $\|E[\mathcal{Z}_{N,t}|\mathcal{V}_{t-m}]\|_2 \leq \zeta(m)$, uniformly in $t \geq 1$ and $N_1, N_2 \geq 1$, where $\zeta(m) = O(m^{-\psi})$ for some $\psi > 1$.

B.5.5.1 Proof of Lemma B.15

Assumption B.6 *a)* gives the strong mixing condition for process V_t. Since $U_t = \mu_N \left(\tilde{\Sigma}_{\Lambda,1}^{-1}\breve{\xi}_{1,t}\right)^{(c)} - \left(\tilde{\Sigma}_{\Lambda,2}^{-1}\breve{\xi}_{2,t}\right)^{(c)}$, where $\tilde{\Sigma}_{\Lambda,j} = \Lambda_j'\Lambda_j/N_j$,

$$\breve{\xi}_{j,t} := \xi_{j,t} - \bar{\xi}_{j,t} \tag{B.173}$$

for $j = 1, 2$, and process U_t is function of the components of process V_t. Therefore, to prove the NED property for process $\mathcal{Z}_{N,t}$, we simply have to show that processes $X_{N,t} = E(U_t'U_t|\mathcal{F}_t)$ is L^2-NED on (V_t). We have:

$$\begin{aligned}\|X_{N,t} - E(X_{N,t}|\mathcal{V}_{t-m}^{t+m})\|_2 &\leq \|X_{N,t} - E(X_{N,t}|F_t, ..., F_{t-m})\|_2 \\ &= \|E(U_t'U_t|\mathcal{F}_t) - E(U_t'U_t|F_t, ..., F_{t-m})\|_2 = O(m^{-\psi}),\end{aligned}$$

for $\psi > 1$, by the Law of Iterated Expectation and Assumption B.6 *b)*. The conclusion follows. ■

B.5.6 Proof of Lemma B.7

The proof of Lemma B.7 is analogous to the proof of Lemma B.7 in AGGR (see their Online Appendix C.10), with their assumption A.5 *b)* replaced by our similar A ssumption B.5 *b)*, and therefore is omitted.

B.5.7 Proof of Lemma B.8

The proof of Lemma B.8 deploys the following uniform asymptotic expansions of factors and loadings estimates:

$$\hat{f}_t^c = \hat{\mathcal{H}}_c^{-1}\left[f_t^c + \frac{1}{\sqrt{N_1}}u_{1,t}^{(c)}\right] + o_p\left(T^{-1/2}\right), \tag{B.174}$$

$$\hat{f}_{j,t}^s = \hat{\mathcal{H}}_{s,j}^{-1}\left[\tilde{f}_{j,t}^s + \frac{1}{\sqrt{N_j}}u_{j,t}^{(s)}\right] + o_p(T^{-1/2}), \qquad j = 1, 2, \tag{B.175}$$

$$\hat{\lambda}_{j,i}^c = \hat{\mathcal{H}}_c'\left[\lambda_{j,i}^c + \tilde{\Sigma}_{cc}^{-1}\tilde{\Sigma}_{c,j}\lambda_{j,i}^s + \frac{1}{\sqrt{T}}\breve{w}_{j,i}^c\right] + o_p\left(T^{-1/2}\right), \qquad j = 1, 2, \tag{B.176}$$

$$\hat{\lambda}_{j,i}^s = \hat{\mathcal{H}}_{s,j}'\left[\lambda_{j,i}^s + \frac{1}{\sqrt{T}}\breve{w}_{j,i}^s\right] + o_p\left(T^{-1/2}\right), \qquad j = 1, 2, \tag{B.177}$$

where the $o_p(T^{-1/2})$ terms are uniform w.r.t. $1 \leq t \leq T$ and $1 \leq i \leq N_j$, vector $u_{j,t}$ is defined in Proposition B.2, $\tilde{f}_{j,t}^s = f_{j,t}^s - \tilde{\Sigma}_{j,c}\tilde{\Sigma}_{cc}^{-1}f_t^c$, $\breve{w}_{j,i}^c = \tilde{\Sigma}_{cc}^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^T \breve{f}_t^c\breve{\varepsilon}_{j,i,t}$ and $\breve{w}_{j,i}^s = (\breve{\tilde{F}}_j^{s\prime}\breve{\tilde{F}}_j^s/T)^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^T \breve{\tilde{f}}_{j,t}^s\breve{\varepsilon}_{j,i,t}$, and matrices $\hat{\mathcal{H}}_c$ and $\hat{\mathcal{H}}_{s,j}$ are such that $\hat{\mathcal{H}}_c'\hat{\mathcal{H}}_c = I_{k^c} + o_p(1)$ and $\hat{\mathcal{H}}_{s,j}'\hat{\mathcal{H}}_{s,j} = I_{k_j^s} + o_p(1)$.

These asymptotic expansions hold under Assumptions B.1-B.4, B.5 *b)-c)*, B.6 *a)*, B.7, B.8, and are derived in Proposition B.6 in Section B.6.

B.5.7.1 Proof of Lemma B.8 Part (i)

To derive the asymptotic expansion of matrix $\hat{\Lambda}_j'\hat{\Lambda}_j/N_j$, we work with the matrix versions of the asymptotic expansions in equations (B.176) and (B.177). Stacking the loadings $\hat{\lambda}_{j,i}^c$ in matrix $\hat{\Lambda}_j^c = [\hat{\lambda}_{j,1}^c, ..., \hat{\lambda}_{j,N_j}^c]'$ we get:

$$\hat{\Lambda}_j^c = \left[\Lambda_j^c + \frac{1}{\sqrt{T}}(G_j^c + \Lambda_j^s \sqrt{T}\tilde{\Sigma}_{j,c}\tilde{\Sigma}_{cc}^{-1})\right]\hat{\mathcal{H}}_c + o_p\left(T^{-1/2}\right),$$

where

$$G_j^c = \frac{1}{\sqrt{T}}\breve{\varepsilon}_j'\breve{F}^c, \tag{B.178}$$

and $o_p(T^{-1/2})$ denotes a matrix whose rows are $(k^c, 1)$ vectors uniformly of order $o_p(T^{-1/2})$. Similarly, stacking the loadings $\hat{\lambda}_{j,i}^s$ in matrix $\hat{\Lambda}_j^s = [\hat{\lambda}_{j,1}^s, ..., \hat{\lambda}_{j,N_j}^s]'$ we get:

$$\hat{\Lambda}_j^s = \left[\Lambda_j^s + \frac{1}{\sqrt{T}}G_j^s\right]\hat{\mathcal{H}}_{s,j} + o_p\left(T^{-1/2}\right),$$

where

$$G_j^s = \frac{1}{\sqrt{T}}\breve{\varepsilon}_j'\breve{F}_j^s. \tag{B.179}$$

By gathering these expansions into matrix $\hat{\Lambda}_j = [\hat{\Lambda}_j^c \ \vdots \ \hat{\Lambda}_j^s]$, we get:

$$\hat{\Lambda}_j = \left(\Lambda_j + \frac{1}{\sqrt{T}}G_j + \frac{1}{\sqrt{T}}\Lambda_j Q_j\right)\hat{\mathcal{U}}_j + o_p\left(T^{-1/2}\right), \qquad j = 1, 2, \tag{B.180}$$

where

$$G_j = \left[\begin{array}{ccc} G_j^c & \vdots & G_j^s \end{array}\right] = \frac{1}{\sqrt{T}}\breve{\varepsilon}_j'\breve{H}_j, \qquad \breve{H}_j = [\breve{F}^c \vdots \breve{F}_j^s], \tag{B.181}$$

$$\hat{\mathcal{U}}_j = \begin{bmatrix} \hat{\mathcal{H}}_c & 0 \\ 0 & \hat{\mathcal{H}}_{s,j} \end{bmatrix}, \tag{B.182}$$

$$Q_j = \begin{bmatrix} 0 & 0 \\ \sqrt{T}\tilde{\Sigma}_{j,c}\tilde{\Sigma}_{cc}^{-1} & 0 \end{bmatrix}. \tag{B.183}$$

To compute $\frac{\hat{\Lambda}_j'\hat{\Lambda}_j}{N_j}$, we consider the matrix product:

$$\begin{aligned}
& \frac{1}{N_j}\left[\Lambda_j + \frac{1}{\sqrt{T}}G_j + \frac{1}{\sqrt{T}}\Lambda_j Q_j\right]'\left[\Lambda_j + \frac{1}{\sqrt{T}}G_j + \frac{1}{\sqrt{T}}\Lambda_j Q_j\right] \\
= & \frac{1}{N_j}\Lambda_j'\Lambda_j + \frac{1}{N_j\sqrt{T}}\left(\Lambda_j' G_j + G_j'\Lambda_j\right) + \frac{1}{N_j T}G_j' G_j + \frac{1}{\sqrt{T}}\left[\left(\frac{1}{N_j}\Lambda_j'\Lambda_j\right)Q_j + Q_j'\left(\frac{1}{N_j}\Lambda_j'\Lambda_j\right)\right] \\
& + \frac{1}{N_j T}\left(Q_j'\Lambda_j' G_j + G_j'\Lambda_j Q_j\right) + \frac{1}{T}Q_j'\left(\frac{1}{N_j}\Lambda_j'\Lambda_j\right)Q_j. \qquad \text{(B.184)}
\end{aligned}$$

Let us now bound the different terms. We have:

$$\frac{1}{\sqrt{N_j}}\Lambda_j' G_j = \frac{1}{\sqrt{N_j T}}\Lambda_j'\breve{\varepsilon}_j'\breve{H}_j = \frac{1}{\sqrt{N_j T}}\sum_{i=1}^{N_j}\sum_{t=1}^{T}\lambda_{j,i}\breve{h}_{j,t}'\breve{\varepsilon}_{j,it} = O_p(1),$$

and:

$$\frac{1}{N_j}G_j' G_j = \frac{1}{N_j}\sum_{i=1}^{N_j}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\breve{h}_{j,t}\breve{\varepsilon}_{j,i,t}\right)\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\breve{h}_{j,t}\breve{\varepsilon}_{j,i,t}\right)' = O_p(1),$$

by arguments similar to the proof of Lemma B.1. Thus, by using these bounds and $\Lambda_j'\Lambda_j/N_j = O(1)$ and $Q_j = O_p(1)$, from equation (B.184) we get:

$$\begin{aligned}
\frac{1}{N_j}\left[\Lambda_j + \frac{1}{\sqrt{T}}G_j + \frac{1}{\sqrt{T}}\Lambda_j Q_j\right]'\left[\Lambda_j + \frac{1}{\sqrt{T}}G_j + \frac{1}{\sqrt{T}}\Lambda_j Q_j\right] \; = \; & \frac{1}{N_j}\Lambda_j'\Lambda_j + \frac{1}{\sqrt{T}}\left(L_{\Lambda,j} + L_{\Lambda,j}'\right) \\
& + O_p\left(\frac{1}{\sqrt{NT}} + \frac{1}{T}\right),
\end{aligned}$$

where

$$L_{\Lambda,j} \;=\; \left(\frac{\Lambda_j'\Lambda_j}{N_j}\right)Q_j. \qquad \text{(B.185)}$$

Therefore we have:

$$\frac{\hat{\Lambda}_j'\hat{\Lambda}_j}{N_j} \;=\; \hat{\mathcal{U}}_j'\left[\frac{\Lambda_j'\Lambda_j}{N_j} + \frac{1}{\sqrt{T}}\left(L_{\Lambda,j} + L_{\Lambda,j}'\right)\right]\hat{\mathcal{U}}_j + o_p\left(\frac{1}{\sqrt{T}}\right).$$

B.5.7.2 Proof of Lemma B.8 Part (ii)

a) Asymptotic expansion of $\hat{\Gamma}_j$

We start by deriving the uniform asymptotic expansion for the residuals. The asymptotic expansions in (B.174)-

(B.177) allow to compute the asymptotic expansion of $\hat{\varepsilon}_{j,i,t}$:

$$\begin{aligned}
\hat{\varepsilon}_{j,i,t} &= y_{j,i,t} - \hat{\lambda}_{j,i}^{c\,\prime}\hat{f}_t^c - \hat{\lambda}_{j,i}^{s\,\prime}\hat{f}_{j,t}^s = \varepsilon_{j,i,t} - \left[\hat{\lambda}_{j,i}^{c\,\prime}\hat{f}_t^c - \lambda_{j,i}^{c\,\prime}f_t^c\right] - \left[\hat{\lambda}_{j,i}^{s\,\prime}\hat{f}_{j,t}^s - \lambda_{j,i}^{s\,\prime}f_{j,t}^s\right] \\
&= \varepsilon_{j,i,t} - \left[\left(\lambda_{j,i}^c + \tilde{\Sigma}_{cc}^{-1}\tilde{\Sigma}_{c,j}\lambda_{j,i}^s + \frac{1}{\sqrt{T}}\breve{w}_{j,i}^c + o_p(T^{-1/2})\right)'\left(f_t^c + \frac{1}{\sqrt{N_1}}u_{1,t}^{(c)} + o_p(T^{-1/2})\right) - \lambda_{j,i}^{c\,\prime}f_t^c\right] \\
&\quad - \left[\left(\lambda_{j,i}^s + \frac{1}{\sqrt{T}}\breve{w}_{j,i}^s + o_p(T^{-1/2})\right)'\left(f_{j,t}^s - \tilde{\Sigma}_{j,c}\tilde{\Sigma}_{cc}^{-1}f_t^c + \frac{1}{\sqrt{N_j}}u_{j,t}^{(s)} + o_p(T^{-1/2})\right) - \lambda_{j,i}^{s\,\prime}f_{j,t}^s\right] \\
&= \varepsilon_{j,i,t} - \left(\frac{1}{\sqrt{N_1}}\lambda_{j,i}^{c\,\prime}u_{1,t}^{(c)} + \frac{1}{\sqrt{T}}\breve{w}_{j,i}^{c\prime}f_t^c\right) - \left(\frac{1}{\sqrt{N_j}}\lambda_{j,i}^{s\,\prime}u_{j,t}^{(s)} + \frac{1}{\sqrt{T}}\breve{w}_{j,i}^{s\prime}f_{j,t}^s\right) + o_p\left(T^{-1/2}\right). \quad \text{(B.186)}
\end{aligned}$$

Here the $o_p(T^{-1/2})$ term is uniform w.r.t. $1 \leq i \leq N_j$, $1 \leq t \leq T$ by the bounds in the next Lemma B.16 and Assumption B.8 *d)*.

LEMMA B.16. *Let $X = O_{p,\ell}(a_{N,T})$ mean $X = O_p[a_{N,T}(\log T)^{\bar{b}}]$ for some $\bar{b} > 0$. Under Assumption B.8 we have the following uniform bounds:*

$$\begin{aligned}
\sup_{1\leq t\leq T}\|h_{j,t}\| &= O_{p,\ell}(1), && \text{(B.187)} \\
\sup_{1\leq t\leq T}\|u_{j,t}\| &= O_{p,\ell}(1), && \text{(B.188)} \\
\sup_{1\leq i\leq N_j}\|\frac{1}{T}\sum_{t=1}^T h_{j,t}\varepsilon_{j,i,t}\| &= O_{p,\ell}(T^{-\eta/2}), && \text{(B.189)}
\end{aligned}$$

where $\eta \geq 1/2$.

If we adopt $\hat{f}_t^c$ to compute residuals in panel $j = 1$, and $\hat{f}_t^{c*}$ for $j = 2$, we have:

$$\hat{\varepsilon}_{j,i,t} = \varepsilon_{j,i,t} - \frac{1}{\sqrt{T}}\left(\breve{w}_{j,i}^{c\prime}f_t^c + \breve{w}_{j,i}^{s\prime}f_{j,t}^s\right) - \frac{1}{\sqrt{N_j}}\left(\lambda_{j,i}^{c\,\prime}u_{j,t}^{(c)} + \lambda_{j,i}^{s\,\prime}u_{j,t}^{(s)}\right) + o_p\left(T^{-1/2}\right). \quad \text{(B.190)}$$

Equation (B.190) allows us to compute:

$$\begin{aligned}
\hat{\gamma}_{j,ii}^* &= \frac{1}{T}\sum_{t=1}^T \breve{\hat{\varepsilon}}_{j,i,t}^2 = \frac{1}{T}\sum_{t=1}^T\left[\breve{\varepsilon}_{j,i,t} - \frac{1}{\sqrt{T}}\left(\breve{w}_{j,i}^{c\prime}\breve{f}_t^c + \breve{w}_{j,i}^{s\prime}\breve{f}_{j,t}^s\right) - \frac{1}{\sqrt{N_j}}\left(\lambda_{j,i}^{c\,\prime}\breve{u}_{j,t}^{(c)} + \lambda_{j,i}^{s\,\prime}\breve{u}_{j,t}^{(s)}\right)\right]^2 + o_p\left(T^{-1/2}\right) \\
&= \frac{1}{T}\sum_{t=1}^T \varepsilon_{j,i,t}^2 - \frac{2}{T\sqrt{T}}\sum_{t=1}^T \varepsilon_{j,i,t}\left(\breve{w}_{j,i}^{c\,\prime}\breve{f}_t^c + \breve{w}_{j,i}^{s\,\prime}\breve{f}_{j,t}^s\right) - \frac{2}{T\sqrt{N_j}}\sum_{t=1}^T \varepsilon_{j,i,t}\left(\lambda_{j,i}^{c\,\prime}\breve{u}_{j,t}^{(c)} + \lambda_{j,i}^{s\,\prime}\breve{u}_{j,t}^{(s)}\right) \\
&\quad + \frac{1}{T^2}\sum_{t=1}^T\left(\breve{w}_{j,i}^{c\,\prime}\breve{f}_t^c + \breve{w}_{j,i}^{s\,\prime}\breve{f}_{j,t}^s\right)^2 + \frac{1}{TN_j}\sum_{t=1}^T\left(\lambda_{j,i}^{c\,\prime}u_{j,t}^{(c)} + \lambda_{j,i}^{s\,\prime}\breve{u}_{j,t}^{(s)}\right)^2 \\
&\quad + \frac{2}{T\sqrt{TN_j}}\sum_{t=1}^T\left(\breve{w}_{j,i}^{c\,\prime}\breve{f}_t^c + \breve{w}_{j,i}^{s\,\prime}\breve{f}_{j,t}^s\right)\left(\lambda_{j,i}^{c\,\prime}\breve{u}_{j,t}^{(c)} + \lambda_{j,i}^{s\,\prime}\breve{u}_{j,t}^{(s)}\right) + o_p\left(T^{-1/2}\right).
\end{aligned}$$

By solving out the parentheses, using $\breve{w}_{j,i}^c = \frac{1}{\sqrt{T}}\sum_{t=1}^T \breve{\varepsilon}_{j,i,t}\breve{f}_t^c = O_p(1)$, $\breve{w}_{j,i}^s =$ $(\breve{\breve{F}}_j^s{}'\breve{\breve{F}}_j^s/T)^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^T \breve{\breve{f}}_{j,t}^s\breve{\varepsilon}_{j,i,t} = O_p(1)$, $\frac{1}{\sqrt{T}}\sum_{t=1}^T \breve{\varepsilon}_{j,i,t}\breve{u}_{j,t}^{(c)} = O_p(1)$ and $\frac{1}{\sqrt{T}}\sum_{t=1}^T \breve{\varepsilon}_{j,i,t}\breve{u}_{j,t}^{(s)} = O_p(1)$, uniformly in $1 \leq i \leq N_j$, we get:

$$\hat{\gamma}_{j,ii}^* \;=\; \frac{1}{T}\sum_{t=1}^T \breve{\varepsilon}_{j,i,t}^2 + O_p\left(\frac{1}{N}\right) + o_p\left(T^{-1/2}\right),$$

uniformly in $1 \leq i \leq N_j$. Using that $1/N = o(1/\sqrt{T})$ when $\sqrt{T} \ll N$, we get:

$$\hat{\gamma}_{j,ii} \;=\; \frac{1}{T}\sum_{t=1}^T \breve{\varepsilon}_{j,i,t}^2 + o_p\left(T^{-1/2}\right) = \gamma_{j,ii} + \frac{1}{\sqrt{T}}w_{j,i}^{\varepsilon} + o_p\left(T^{-1/2}\right),$$

uniformly in $1 \leq i \leq N_j$, where

$$w_{j,i}^{\varepsilon} \;:=\; \frac{1}{\sqrt{T}}\sum_{t=1}^T (\breve{\varepsilon}_{j,i,t}^2 - \gamma_{j,ii}).$$

Therefore, we have:

$$\hat{\Gamma}_j \;=\; \Gamma_j + \frac{1}{\sqrt{T}}W_j^{\varepsilon} + o_p\left(T^{-1/2}\right), \tag{B.191}$$

where $\Gamma_j = diag(\gamma_{j,ii},\ i = 1,...,N_j)$ and $W_j^{\varepsilon} = diag(w_{j,i}^{\varepsilon},\ i = 1,...,N)$, for $j = 1,2$.

b) Asymptotic expansion of $\frac{1}{N_j}\hat{\Lambda}_j'\hat{\Gamma}_j\hat{\Lambda}_j$

From (B.180) and (B.191) we have:

$$\frac{1}{N_j}\hat{\Lambda}_j'\Gamma_j\hat{\Lambda}_j = \hat{\mathcal{U}}_j'\hat{\Omega}_{jj}^*\hat{\mathcal{U}}_j + o_p\left(T^{-1/2}\right), \tag{B.192}$$

where we define:

$$\begin{aligned}
\hat{\Omega}_{jj}^* \;&:=\; \frac{1}{N_j}\left(\Lambda_j + \frac{1}{\sqrt{T}}G_j + \frac{1}{\sqrt{T}}\Lambda_j Q_j\right)'\left(\Gamma_j + \frac{1}{\sqrt{T}}W_j^{\varepsilon}\right)\left(\Lambda_j + \frac{1}{\sqrt{T}}G_j + \frac{1}{\sqrt{T}}\Lambda_j Q_j\right) \\
&=\; \tilde{\Omega}_{jj} + \hat{\Omega}_{jj,I}^* + \hat{\Omega}_{jj,II}^* + \hat{\Omega}_{jj,II}^*{}' + \hat{\Omega}_{jj,III}^* + \hat{\Omega}_{jj,III}^*{}' + \hat{\Omega}_{jj,IV}^* + \hat{\Omega}_{jj,V}^* \\
&\quad + \frac{1}{\sqrt{T}}(\tilde{\Omega}_{jj}Q_j + Q_j'\tilde{\Omega}_{jj}) + \frac{1}{\sqrt{T}}(\hat{\Omega}_{jj,I}^*Q_j + Q_j'\hat{\Omega}_{jj,I}^*) + \frac{1}{\sqrt{T}}(\hat{\Omega}_{jj,II}^*Q_j + Q_j'\hat{\Omega}_{jj,II}^*{}') \\
&\quad + \frac{1}{\sqrt{T}}(\hat{\Omega}_{jj,III}^*Q_j + Q_j'\hat{\Omega}_{jj,III}^*{}') + \frac{1}{T}Q_j'\tilde{\Omega}_{jj}Q_j + \frac{1}{T}Q_j'\hat{\Omega}_{jj,I}^*Q_j,
\end{aligned}$$

and:

$$
\begin{aligned}
\tilde{\Omega}_{jj} &:= \frac{1}{N_j}\Lambda_j'\Gamma_j\Lambda_j, \\
\hat{\Omega}^*_{jj,I} &:= \frac{1}{N_j\sqrt{T}}\Lambda_j'W_j^{\varepsilon}\Lambda_j = O_p\left(\frac{1}{\sqrt{NT}}\right), \\
\hat{\Omega}^*_{jj,II} &:= \frac{1}{N_j\sqrt{T}}G_j'\Gamma_j\Lambda_j = O_p\left(\frac{1}{\sqrt{NT}}\right), \\
\hat{\Omega}^*_{jj,III} &:= \frac{1}{N_jT}G_j'W_j^{\varepsilon}\Lambda_j = O_p\left(\frac{1}{T}\right), \\
\hat{\Omega}^*_{jj,IV} &:= \frac{1}{N_jT}G_j'\Gamma_jG_j = O_p\left(\frac{1}{T}\right), \\
\hat{\Omega}^*_{jj,V} &:= \frac{1}{N_jT\sqrt{T}}G_j'W_j^{\varepsilon}G_j = O_p\left(\frac{1}{T\sqrt{T}}\right).
\end{aligned}
$$

Collecting the previous results, we get:

$$
\hat{\Omega}^*_{jj} = \tilde{\Omega}_{jj} + \frac{1}{\sqrt{T}}\left(L_{\Omega,j} + L_{\Omega,j}'\right) + O_p\left(\frac{1}{\sqrt{NT}} + \frac{1}{T}\right), \tag{B.193}
$$

where:

$$
L_{\Omega,j} = \tilde{\Omega}_{jj}Q_j. \tag{B.194}
$$

By substituting into equation (B.192) we get:

$$
\frac{1}{N_j}\hat{\Lambda}_j'\hat{\Gamma}_j\hat{\Lambda}_j = \hat{\mathcal{U}}_j'\left[\tilde{\Omega}_{jj} + \frac{1}{\sqrt{T}}\left(L_{\Omega,j} + L_{\Omega,j}'\right)\right]\hat{\mathcal{U}}_j + o_p\left(T^{-1/2}\right), \qquad j = 1, 2.
$$

■

B.5.7.3 Proof of Lemma B.16

We prove the uniform bounds in (B.187) and (B.189). The proof of bound (B.188) follows by similar arguments.

Proof of (B.187). Let $\delta = c(\log T)^{\bar{b}}$, for constants $c > 0$ and $\bar{b} = 1/b$, where $b > 0$ is defined in Assumption B.8 a). Then:

$$
\begin{aligned}
P[\sup_{1\leq t\leq T}\|h_{j,t}\| \geq \delta] &\leq \sum_{t=1}^{T} P[\|h_{j,t}\| \geq \delta] \leq c_1T\exp(-c_2\delta^b) = c_1T\exp[-c_2c^b(\log T)] \\
&= c_1T^{1-c_2c^b} = o(1),
\end{aligned}
$$

if $c > (1/c_2)^{1/b}$. Thus, $\sup_{1\leq t\leq T}\|h_{j,t}\| = O_p[(\log T)^{\bar{b}}]$.

Proof of (B.189). Let $\delta = c(\log T)^{1/2}T^{-\eta/2}$, for constants $c > 0$ and η, where $\eta \geq 1/2$ is defined in

Assumption B.8 c). Then:

$$\begin{aligned}
P[\sup_{1\leq i\leq N_j}\|\frac{1}{T}\sum_{t=1}^{T}h_{j,t}\varepsilon_{j,i,t}\|\geq\delta] &\leq \sum_{i=1}^{N_j}P[\|\frac{1}{T}\sum_{t=1}^{T}h_{j,t}\varepsilon_{j,i,t}\|\geq\delta]\leq N_j\sup_{1\leq i\leq N_j}P[\|\frac{1}{T}\sum_{t=1}^{T}h_{j,t}\varepsilon_{j,i,t}\|\geq\delta] \\
&\leq c_1N_jT\exp(-c_2\delta^2T^{\eta})+c_3TN_j\delta^{-1}\exp(-c_4T^{\bar{\eta}}) \\
&= c_1N_jT\exp(-c_2c^2(\log T))+c_3TN_j\delta^{-1}\exp(-c_4T^{\bar{\eta}}) \\
&= O(T^{7/2-c_2c^2})+o(1)=o(1),
\end{aligned}$$

if $c>(\frac{7}{2c_2})^{1/2}$. Thus, $\sup_{1\leq i\leq N_j}\|\frac{1}{T}\sum_{t=1}^{T}h_{j,t}\varepsilon_{j,i,t}\|=O_p[(\log T)^{1/2}T^{-\eta/2}]=O_{p,\ell}(T^{-\eta/2})$.

B.5.8 Proof of Lemma B.9

We assume that estimator $\hat{f}_t^c$ is used to get factor loadings on panel $j=1$, and estimator $\hat{f}_t^{c\,*}$ is used to get factor loadings on panel $j=2$. Recall $\hat{\Sigma}_U=(N_2/N_1)\hat{\Sigma}_{u,11}^{(cc)}+\hat{\Sigma}_{u,22}^{(cc)}$. Let r be the true number of common factors, and let k^c denote the number of common factors used in the estimation procedure. We consider the case with $r<k^c\leq\underline{k}\equiv\min\{k_1,k_2\}$.

Let us first consider panel $j=1$. The common factor estimator is $\hat{f}_t^c=\hat{W}_1'\hat{h}_{1,t}$ where $\hat{W}_1$ is the $k_1\times k^c$ matrix whose columns are eigenvectors of $\hat{R}$ associated with the k^c largest eigenvalues, normalized to have $\hat{W}_1'\hat{W}_1=I_{k^c}$. Without loss of generality, let $\hat{\mathcal{H}}_j=I_{k_j}$ in Proposition B.2. Then, we have $\hat{R}=R+o_p(1)$, where $R=\begin{pmatrix} I_r & 0\\ 0 & \Phi\Phi'\end{pmatrix}$. The large-sample limit of $\hat{W}_1$ is the matrix of normalized eigenvectors associated to the k^c largest eigenvalues of matrix R. These eigenvalues are 1, with multiplicity r, and ρ_{r+1}^2, ..., $\rho_{k^c}^2$, that are the k^c-r largest eigenvalues of matrix $\Phi\Phi'$ (assumed distinct, to simplify the proof). Let α denote the $(k_1-r)\times(k^c-r)$ matrix whose columns are the corresponding normalized eigenvectors of $\Phi\Phi'$. Then, we have $\hat{W}_1=W_1+o_p(1)$ where

$$W_1=\begin{bmatrix}\mathcal{U} & 0\\ 0 & \alpha\end{bmatrix},$$

$r\times r$ matrix $\mathcal{U}$ is possibly stochastic and such that $\mathcal{U}'\mathcal{U}=I_r$, and $\alpha'\alpha=I_{k^c-r}$. For later use, we denote by β the $(k_1-r)\times(k_1-k^c)$ matrix whose columns are an orthonormal basis of the orthogonal complement to the columns space of α. Then, $[\alpha\vdots\beta]$ is an orthogonal matrix, $\beta'\beta=I_{k_1-k^c}$, $\alpha'\beta=0$, and:

$$\alpha\alpha'+\beta\beta'=I_{k_1-r}. \tag{B.195}$$

From Proposition B.2 with $\hat{\mathcal{H}}_j=I_{k_j}$ we have $\hat{h}_{j,t}\simeq h_{j,t}$, where symbol $\simeq$ means equality up to terms that are asymptotically negligible for determining large-sample limits. Then:

$$\hat{f}_t^c\simeq W_1'h_{1,t}=\begin{bmatrix}\mathcal{U}'f_t^c\\ \alpha'f_{1,t}^s\end{bmatrix}.$$

Let us consider the estimation of the factor loadings on the panel with $j = 1$. From (B.195) the model for this panel can be written as:

$$\begin{aligned} y_{1,i,t} &= f_t^{c\,\prime}\lambda_{1,i}^c + f_{1,t}^{s\,\prime}\lambda_{1,i}^s + \varepsilon_{1,i,t} = [\mathcal{U}f_t^{c\,\prime}][\mathcal{U}\lambda_{1,i}^c] + [\alpha' f_{1,t}^s]'[\alpha'\lambda_{1,i}^s] + [\beta' f_{1,t}^s]'[\beta'\lambda_{1,i}^s] + \varepsilon_{1,i,t} \\ &= \underline{f}_t^{c\,\prime}\underline{\lambda}_{1,i}^c + \underline{f}_{1,t}^{s\,\prime}\underline{\lambda}_{1,i}^s + \varepsilon_{1,i,t}, \end{aligned}$$

where $\underline{f}_t^c = \begin{bmatrix} \mathcal{U}' f_t^c \\ \alpha' f_{1,t}^s \end{bmatrix}$, $\underline{\lambda}_{1,i}^c = \begin{bmatrix} \mathcal{U}'\lambda_{1,i}^c \\ \alpha'\lambda_{1,i}^s \end{bmatrix}$, $\underline{f}_{1,t}^s = \beta' f_{1,t}^s$ and $\underline{\lambda}_{1,i}^s = \beta'\lambda_{1,i}^s$. Note that the transformed factors $\underline{f}_t^c$ and $\underline{f}_{1,t}^s$ are orthogonal, and have dimensions k^c and $k_1 - k^c$ respectively. Since $\hat{f}_t^c$ converges to $\underline{f}_t^c$, by regressing $\breve{y}_{1,i,t}$ onto $\check{\hat{f}}_t^c$ we estimate $\underline{\lambda}_{1,i}^c$. Then, the residuals satisfy the model:

$$\xi_{1,i,t} \simeq \underline{f}_{1,t}^{s\,\prime}\underline{\lambda}_{1,i}^s + \varepsilon_{1,i,t}.$$

The group-specific factor is estimated by extracting the first $k_1 - k^c$ RP-PCs (RP-Principal components) from the residuals, which yields asymptotically $\hat{f}_{1,t}^s \simeq \mathcal{V}\underline{f}_t^s$, where $\mathcal{V}$ is p.d. matrix. So for the estimated factor loadings we have:

$$\hat{\lambda}_{1,i}^c \simeq \underline{\lambda}_{1,i}^c = \begin{bmatrix} \mathcal{U}'\lambda_{1,i}^c \\ \alpha'\lambda_{1,i}^s \end{bmatrix}, \qquad \hat{\lambda}_{1,i}^s \simeq \mathcal{V}\underline{\lambda}_{1,i}^s = \mathcal{V}\beta'\lambda_{1,i}^s.$$

Thus, $\hat{\lambda}_{1,i}$ is asymptotically an orthogonal transformation of $\lambda_{1,i}$, i.e. $\hat{\lambda}_{1,i} \simeq \mathcal{R}_1\lambda_{1,i}$, say. Using $\hat{\varepsilon}_{1,i,t} \simeq \varepsilon_{1,i,t}$, we get $\hat{\Sigma}_{u,11} \simeq \mathcal{R}_1\Sigma_{u,11}\mathcal{R}_1'$, which implies $\hat{\Sigma}_{u,11} = O_p(1)$.

Let us now consider the estimation of factor loadings in panel $j = 2$. By paralleling the above arguments, we have $\hat{\Sigma}_{u,22} = O_p(1)$. Thus, $\|\hat{\Sigma}_U\| = O_p(1)$. The conclusion follows. ■

B.6 Uniform asymptotic expansions of factor values and factor loadings in the group factor model

In order to prove Lemma B.8, we need to the asymptotic expansions of factor values and factor loadings in our group factor model, which is provided in the following Proposition B.6. This proposition provides the uniform asymptotic expansions for the estimators of the factor values and factor loadings in Definitions 1 and 2 and equations (A.13) and (A.14), up to terms $o_p(\bar{N}^{-1/2})$, where $\bar{N} := \max\{N_1, T\}$.

PROPOSITION B.6. *i) Under Assumption B.1 with $\mu > 0$, and Assumptions B.2-B.4, B.5 b)-c), B.6 a), B.7, B.8 (TO BE CHECKED !!!!) the asymptotic expansions of the factors estimators are given by:*

$$\hat{f}_t^c = \hat{\mathcal{H}}_c^{-1}\left[f_t^c + \frac{1}{\sqrt{N_1}}u_{1,t}^{(c)}\right] + o_p\left(\bar{N}^{-1/2}\right), \tag{B.196}$$

and:

$$\hat{f}_{j,t}^s = \hat{\mathcal{H}}_{s,j}^{-1}\left[\tilde{f}_{j,t}^s + \frac{1}{\sqrt{N_j}}u_{j,t}^{(s)}\right] + o_p(\bar{N}^{-1/2}), \qquad j = 1, 2, \tag{B.197}$$

where $\tilde{f}^s_{j,t} = f^s_{j,t} - \tilde{\Sigma}_{j,c}\tilde{\Sigma}^{-1}_{cc} f^c_t$ *and the* o_p *terms are uniform w.r.t.* $1 \leq t \leq T$. *The asymptotic expansions of the loadings estimators are:*

$$\hat{\lambda}^c_{j,i} = \hat{\mathcal{H}}'_c \left[\lambda^c_{j,i} + \tilde{\Sigma}^{-1}_{cc}\tilde{\Sigma}_{c,j}\lambda^s_{j,i} + \frac{1}{\sqrt{T}}\breve{w}^c_{j,i} \right] + o_p\left(\bar{N}^{-1/2}\right), \qquad j = 1,\ 2, \tag{B.198}$$

and:

$$\hat{\lambda}^s_{j,i} = \hat{\mathcal{H}}'_{s,j} \left[\lambda^s_{j,i} + \frac{1}{\sqrt{T}}\breve{w}^s_{j,i} \right] + o_p\left(\bar{N}^{-1/2}\right), \qquad j = 1, 2, \tag{B.199}$$

where the o_p *terms are uniform w.r.t.* $1 \leq i \leq N_j$. *Matrices* $\hat{\mathcal{H}}_c$ *and* $\hat{\mathcal{H}}_{s,j}$ *are such that:*

$$\hat{\mathcal{H}}_c\hat{\mathcal{H}}'_c = \tilde{\Sigma}_{cc} + o_p(\bar{N}^{-1/2}), \qquad \hat{\mathcal{H}}_{s,j}\hat{\mathcal{H}}'_{s,j} = (\frac{1}{T}\tilde{F}^{s}_j{}'\tilde{F}^s_j)^{-1} + o_p(\bar{N}^{-1/2}), \quad j = 1, 2, \tag{B.200}$$

where $\tilde{F}^s_j = [\tilde{f}^s_{j,1}, ..., \tilde{f}^s_{j,T}]'$. *Vector* $u_{j,t}$ *is defined in Proposition B.2, and* $\breve{w}^c_{j,i} = \tilde{\Sigma}^{-1}_{cc}\frac{1}{\sqrt{T}}\sum_{t=1}^T \breve{\tilde{f}}^c_t \breve{\varepsilon}_{j,i,t}$ *and* $\breve{w}^s_{j,i} = \frac{1}{\sqrt{T}}\sum_{t=1}^T \breve{\tilde{f}}^s_{j,t}\breve{\varepsilon}_{j,i,t}$.

This proposition is analogous to Proposition D.4 in AGGR (see their Online Appendix D.4). In the asymptotic expansion of $\hat{f}^c_t$, the stochastic term at order $N_1^{-1/2}$ comes from the estimation of the principal components in the first subgroup. Interestingly, no bias term of order $1/T$ appears in the expansions of $\hat{f}^c_t$ and $\hat{f}^s_{1,t}$, as these bias terms can be absorbed into the terms $o_p(\bar{N}^{-1/2})$ under Assumption B.1, whihc implies $\sqrt{T} \ll N \ll T^2$. Instead, bias terms of order $1/T$ were present in AGGR, who used the assumptions $\sqrt{T} \ll N \ll T^{5/2}$. Similarly, bias terms of order $1/T$ in the expansions of the loadings estimators appearing in Propositon 4 of AGGR are also absorbed in the terms $o_p(\bar{N}^{-1/2})$ in our Proposition B.6.

In the asymptotic expansion of $\hat{\lambda}^c_{j,i}$, the term $\tilde{\Sigma}^{-1}_{cc}\tilde{\Sigma}_{c,j}\lambda^s_{j,i}$ is induced by the fact that the common and frequency-specific factors are not orthogonal in-sample. The expansion of $\hat{\lambda}^c_{j,i}$ does not contain explicitly a bias component at order N_j^{-1}, since $N_j^{-1} = o_p(\bar{N}^{-1/2})$ under Assumption B.1.

The uniform asymptotic expansions at order $o_p(T^{-1/2})$ in Proposition B.6 ii) suffice for the proof of Theorem A.2.

B.6.1 Proof of Proposition B.6

We start by providing some uniform bounds in Subsection B.6.1 a), that are instrumental for the rest of the proof of Proposition B.6. Then, in Subsections B.6.1 b)-e) we establish the uniform asymptotic expansions of factors and loadings up to order $o_p(\bar{N}^{-1/2})$, where $\bar{N} = \max\{N_1, T\}$ (proof of part i)). Finally, in Subsection B.6.1 f) we show how to get the uniform asymptotic expansions up to order $o_p(T^{-1/2})$ under a less restrictive asymptotic scheme (proof of part ii)).

a) Uniform bounds (TO BE CHECKED !!!!)

Let $X = O_{p,\ell}(a_{N,T})$ mean $X = O_p[a_{N,T}(\log T)^{\bar{b}}]$ for some $\bar{b} > 0$. Under Assumption B.8 we have the following uniform bounds, which complement those in Lemma B.16:

$$\sup_{1\leq t\leq T} \|b_{j,t}\| = O_{p,\ell}(1), \tag{B.201}$$

$$\sup_{1\leq t\leq T} \|d_{j,t}\| = O_{p,\ell}(1), \tag{B.202}$$

$$\sup_{1\leq t\leq T} \|\hat{h}_{j,t}\| = O_{p,\ell}(1), \tag{B.203}$$

$$\sup_{1\leq t\leq T} \|\beta^c_{j,t}\| = O_{p,\ell}(1), \tag{B.204}$$

$$\sup_{1\leq i\leq N_j} \|\frac{1}{T}\sum_{t=1}^{T} \beta^c_{j,t}\varepsilon_{j,i,t}\| = O_{p,\ell}(T^{-\eta/2}), \tag{B.205}$$

$$\sup_{1\leq i\leq N_j} \|\frac{1}{T}\sum_{t=1}^{T} \varepsilon^2_{j,i,t}\| = O_p(1), \tag{B.206}$$

$$\sup_{1\leq i\leq N_j} \frac{1}{N_j T}\sum_{\ell=1,\ell\neq i}^{N_j}\sum_{t=1}^{T} \lambda_{j,\ell}\varepsilon_{j,\ell,t}\varepsilon_{j,i,t} = O_{p,\ell}(\frac{1}{\sqrt{NT^{\eta}}}) + O(\frac{1}{N}), \tag{B.207}$$

where $\eta \geq 1/2$. We prove below the uniform bound in (B.207). The proofs of the other ones follow by similar arguments.

Proof of (B.207). We have:

$$\begin{aligned}\frac{1}{N_j T}\sum_{\ell=1,\ell\neq i}^{N_j}\sum_{t=1}^{T} \lambda_{j,\ell}\varepsilon_{j,\ell,t}\varepsilon_{j,i,t} &= \frac{1}{\sqrt{N_j}}\left[\frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{\sqrt{N_j}}\sum_{\ell=1,\ell\neq i}^{N_j} \varepsilon_{j,\ell,t}\varepsilon_{j,i,t} - E[\frac{1}{\sqrt{N_j}}\sum_{\ell=1,\ell\neq i}^{N_j} \lambda_{j,\ell}\varepsilon_{j,\ell,t}\varepsilon_{j,i,t}]\right)\right] \\ &+ \frac{1}{N_j}\sum_{\ell=1,\ell\neq i}^{N_j} \lambda_{j,\ell} E[\varepsilon_{j,\ell,t}\varepsilon_{j,i,t}].\end{aligned}$$

From Assumption B.8 c) we have $\frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{\sqrt{N_j}}\sum_{\ell=1,\ell\neq i}^{N_j} \varepsilon_{j,\ell,t}\varepsilon_{j,i,t} - E[\frac{1}{\sqrt{N_j}}\sum_{\ell=1,\ell\neq i}^{N_j} \lambda_{j,\ell}\varepsilon_{j,\ell,t}\varepsilon_{j,i,t}]\right) = O_{p,\ell}(T^{-\eta/2})$, uniformly in $1 \leq i \leq N_j$, similarly as in the proof of (B.189). From Assumptions B.8 b) and d) we have $\sum_{\ell=1,\ell\neq i}^{N_j} \lambda_{j,\ell} E[\varepsilon_{j,\ell,t}\varepsilon_{j,i,t}] = O(1)$, uniformly in $1 \leq i \leq N_j$. Then, (B.207) follows.

b) Asymptotic expansion of $\hat{f}_t^c$

Let us start by establishing the asymptotic expansion of $\hat{f}_t^c$ up to order $o_p(\bar{N}^{-1/2})$. Equation (B.21) and $\hat{\Psi} = O_p(\delta_{N,T})$ imply $\hat{W}_1^* = [E_c + E_s(I_{k_1-k^c} - \tilde{R}_{ss})^{-1}\hat{\Psi}_{sc}^{(I)}]\hat{\mathcal{U}} + O_p(\delta_{N,T}^2)$. The normalized eigenvectors corresponding to the canonical directions are: $\hat{W}_1 = \hat{W}_1^* \hat{D}$, where $\hat{D} = diag(\hat{W}_1^{*\prime}\hat{V}_{11}\hat{W}_1^*)^{-1/2}$. Then, from Definition 1 and equation (B.2), we get:

$$
\begin{aligned}
\hat{f}_t^c &= \hat{W}_1'\hat{h}_{1,t} = \hat{D}\hat{\mathcal{U}}'\left[E_c'\hat{h}_{1,t} + \hat{\Psi}_{sc}^{(I)\prime}(I_{k_1-k^c} - \tilde{R}_{ss})^{-1}E_s'\hat{h}_{1,t}\right] + O_{p,l}\left(\delta_{N,T}^2\right) \\
&= \hat{D}\hat{\mathcal{U}}'\Bigg[f_t^c + \frac{1}{\sqrt{N_1}}u_{1,t}^{(c)} + \frac{1}{T}\breve{b}_{1,t}^{(c)} + \frac{1}{\sqrt{N_1 T}}\breve{d}_{1,t}^{(c)} + \breve{\vartheta}_{1,t}^{(c)} \\
&\quad + \hat{\Psi}_{sc}^{(I)\prime}(I_{k_1-k^c} - \tilde{R}_{ss})^{-1}\left(f_{1,t}^s + \frac{1}{\sqrt{N_1}}u_{1,t}^{(s)} + \frac{1}{T}\breve{b}_{1,t}^{(s)} + \frac{1}{\sqrt{N_1 T}}\breve{d}_{1,t}^{(s)} + \breve{\vartheta}_{1,t}^{(s)}\right)\Bigg] + O_{p,l}\left(\delta_{N,T}^2\right),
\end{aligned} \tag{B.208}
$$

uniformly in $1 \le t \le T$, where we use the expansion of the factor estimates in Proposition B.2, and (B.203). Under Assumption B.1 with $\mu > 0$, $N = N_2$ and N_1 grow at the same rate such that $T^{1/2} \ll N \ll T^2$. Therefore, $(\log T)^{\bar{b}}\delta_{N,T}^2 = o(\bar{N}^{-1/2})$, for any $\bar{b} > 0$, $\frac{1}{\sqrt{N_1}}\delta_{N,T} = o(\bar{N}^{-1/2})$ and $\frac{1}{T}\delta_{N,T} = o(\bar{N}^{-1/2})$ under Assumption B.1 with $\mu > 0$. By using uniform bounds in Lemma B.16 (**this Lemma needs to be written and proved ... but should hold!**) and (B.201)-(B.202), and keeping only terms up to $o_p(\bar{N}^{-1/2})$, we get:

$$
\hat{f}_t^c = \hat{\mathcal{H}}_c^{-1}\left[f_t^{(c)} + \frac{1}{\sqrt{N_1}}u_{1,t}^{(c)} + \frac{1}{T}\breve{\breve{b}}_{1,t}^{(c)} + \hat{\Psi}_{sc}^{(I)\prime}(I_{k_1-k^c} - \tilde{R}_{ss})^{-1}f_{1,t}^s\right] + o_p\left(\bar{N}^{-1/2}\right), \tag{B.209}
$$

uniformly in $1 \le t \le T$, where $\hat{\mathcal{H}}_c^{-1} = \hat{D}\hat{\mathcal{U}}'$ and $\breve{\breve{b}}_{1,t}$ is defined in equation (A.18).

To further develop this asymptotic expansion, we need the asymptotic behavior of $\hat{\Psi}_{sc}^{(I)}$. From equation $\hat{\Psi} = \tilde{V}_{11}^{-1}\hat{\Psi}^*$ (see Lemma B.2) we have $\hat{\Psi}_{sc}^{(I)} = (\tilde{V}_{11}^{-1})_{sc}\hat{\Psi}_{cc}^{*(I)} + (\tilde{V}_{11}^{-1})_{ss}\hat{\Psi}_{sc}^{*(I)}$. From Lemma B.13, we have $\hat{\Psi}_{cc}^{*(I)} = O_p\left(\frac{1}{N} + \frac{1}{T^2} + \frac{1}{T\sqrt{NT}}\right) = o_p(\bar{N}^{-1/2})$ under Assumption B.1 with $\mu > 0$. Moreover, from (B.14) and Lemma B.3 we get:

$$
\hat{\Psi}_{sc}^{*(I)} = -(\hat{X}_{11,sc} - \hat{X}_{12,sc}) + \tilde{B}_{cs}'(\hat{X}_{21,cc} - \hat{X}_{22,cc}) + \tilde{B}_{ss}'(\hat{X}_{21,sc} - \hat{X}_{22,sc}).
$$

From Lemmas B.1 and B.3, and equation (B.161), the second term in the r.h.s. is $O_p(T^{-1/2}\delta_{N,T}) = o_p(\bar{N}^{-1/2})$ under Assumption B.1 with $\mu > 0$. Now, we substitute in the definitions of terms $\hat{X}_{j,k}$ from (B.11), and use that $\frac{1}{T}\sum_{t=1}^T \breve{\psi}_{j,t}\breve{\psi}_{k,t}' = o_p(\bar{N}^{-1/2})$. We get:

$$
\hat{\Psi}_{sc}^{*(I)} = -\frac{1}{T}\sum_{t=1}^T(\breve{f}_{1,t}^s - \tilde{B}_{ss}'\breve{f}_{2,t}^s)[\breve{\psi}_{1,t}^{(c)} - \breve{\psi}_{2,t}^{(c)}]' + o_p(\bar{N}^{-1/2}).
$$

By using the definition of $\breve{\psi}_{j,t}$, $\tilde{B}_{ss} = \Phi' + O_p(T^{-1/2})$, and keeping terms up to $o_p(\bar{N}^{-1/2})$, we get:

$$\begin{aligned}\hat{\Psi}^{*(I)}_{sc} &= -\frac{1}{T}\left(\frac{1}{T}\sum_{t=1}^{T}(\breve{f}^s_{1,t} - \Phi\breve{f}^s_{2,t})[\breve{\bar{b}}^{(c)}_{1,t} - \breve{\bar{b}}^{(c)}_{2,t}]'\right) + o_p(\bar{N}^{-1/2}) \\ &= -\frac{1}{T}E[(\breve{f}^s_{1,t} - \Phi\breve{f}^s_{2,t})(\breve{\bar{b}}^{(c)}_{1,t} - \breve{\bar{b}}^{(c)}_{2,t})'] + o_p(\bar{N}^{-1/2}).\end{aligned}$$

Thus, by using $(\tilde{V}_{11}^{-1})_{ss} = I_{k_1-k^c} + O_p(T^{-1/2})$ and $N \ll T^3$, we get:

$$\hat{\Psi}^{(I)}_{sc} = -\frac{1}{T}E[(\breve{f}^s_{1,t} - \Phi\breve{f}^s_{2,t})(\breve{\bar{b}}^{(c)}_{1,t} - \breve{\bar{b}}^{(c)}_{2,t})'] + o_p(\bar{N}^{-1/2}). \tag{B.210}$$

Thus, from (B.209) and (B.210), and by using $(I_{k_1-k^c} - \tilde{R}_{ss})^{-1} = (I_{k_1-k^c} - \Phi\Phi')^{-1} + O_p(T^{-1/2})$ and $N \ll T^3$, we get:

$$\hat{f}^c_t = \hat{\mathcal{H}}_c^{-1}\left[f^{(c)}_t + \frac{1}{\sqrt{N_1}}u^{(c)}_{1,t} + \frac{1}{T}\breve{\beta}^c_{1,t}\right] + o_p\left(\bar{N}^{-1/2}\right), \tag{B.211}$$

uniformly in $1 \leq t \leq T$, where:

$$\breve{\beta}^c_{1,t} = \breve{\bar{b}}^{(c)}_{1,t} - E[(\breve{\bar{b}}^{(c)}_{1,t} - \breve{\bar{b}}^{(c)}_{2,t})(\breve{f}^s_{1,t} - \Phi\breve{f}^s_{2,t})'](I_{k_1-k^c} - \Phi\Phi')^{-1}f^s_{1,t},$$

which yields (B.196). By noting that $1/T = o_p(\bar{N}^{-1/2})$ under Assumption B.1, we get:

$$\hat{f}^c_t = \hat{\mathcal{H}}_c^{-1}\left[f^{(c)}_t + \frac{1}{\sqrt{N_1}}u^{(c)}_{1,t}\right] + o_p\left(\bar{N}^{-1/2}\right), \tag{B.212}$$

The asymptotic expansion for estimator $\hat{f}^{c\,*}_t$ is obtained by interchanging the roles of panels $j = 1$ and $j = 2$. Hence,

$$\hat{f}^{c\,*}_t = \hat{\mathcal{H}}_{*c}^{-1}\left[f^{(c)}_t + \frac{1}{\sqrt{N_2}}u^{(c)}_{2,t}\right] + o_p\left(\bar{N}^{-1/2}\right),$$

uniformly in $1 \leq t \leq T$.

Finally, let us show the asymptotic expansion for $\hat{\mathcal{H}}_c\hat{\mathcal{H}}_c'$. We first need to compute:

$$\begin{aligned}\breve{\hat{f}}^c_t &= \hat{f}^c_t - \bar{f}^c \\ &= \hat{\mathcal{H}}_c^{-1}\left[\breve{f}^{(c)}_t + \frac{1}{\sqrt{N_1}}\breve{u}^{(c)}_{1,t}\right] + o_p\left(\bar{N}^{-1/2}\right),\end{aligned} \tag{B.213}$$

Substituting the expression of $\hat{f}^c_t$ from equation (B.213) into the equality $\frac{1}{T}\sum_{t=1}^{T}\breve{\hat{f}}^c_t\breve{\hat{f}}^{c\prime}_t = I_{k^c}$ from equation

(A.12), we get:

$$\begin{aligned} I_{k^c} &= \hat{\mathcal{H}}_c^{-1}\frac{1}{T}\sum_{t=1}^{T}\left(\breve{f}_t^c+\frac{1}{\sqrt{N_1}}\breve{u}_{1,t}^{(c)}\right)\left(\breve{f}_t^c+\frac{1}{\sqrt{N_1}}\breve{u}_{1,t}^{(c)}\right)'\left(\hat{\mathcal{H}}_c^{-1}\right)'+o_p\left(\bar{N}^{-1/2}\right) \\ &= \hat{\mathcal{H}}_c^{-1}\tilde{\Sigma}_{cc}\left(\hat{\mathcal{H}}_c^{-1}\right)'+o_p\left(\bar{N}^{-1/2}\right), \end{aligned} \tag{B.214}$$

using arguments similar to the proof of Lemma B.1 and Assumption B.1 with $\mu > 0$. Thus, we get $\hat{\mathcal{H}}_c\hat{\mathcal{H}}_c' = \tilde{\Sigma}_{cc} + o_p\left(\bar{N}^{-1/2}\right)$, which yields the first equation in (B.200). By using (B.161) it follows:

$$\hat{\mathcal{H}}_c\hat{\mathcal{H}}_c' = I_{k^c} + O_p(T^{-1/2}). \tag{B.215}$$

c) Asymptotic expansion of $\hat{\lambda}_{j,i}^c$

Let us now derive the asymptotic expansion of the loading estimator

$$\hat{\lambda}_{j,i}^c = \left(\hat{\breve{F}}^{c\,\prime}\hat{\breve{F}}^c\right)^{-1}\hat{\breve{F}}^{c\,\prime}\breve{y}_{j,i} = \hat{\breve{F}}^{c\,\prime}\breve{y}_{j,i}/T \tag{B.216}$$

up to order $o_p\left(\bar{N}^{-1/2}\right)$, where $y_{j,i}$ is the i-th column of matrix $\breve{Y}_j$ and $\hat{\breve{F}}^c = [\hat{\breve{f}}_1^c, ..., \hat{\breve{f}}_T^c]'$. From equation (B.213) we have $\hat{\breve{F}}^c = \left(\breve{F}^c + \frac{1}{\sqrt{N_1}}\breve{U}_1^c\right)\left(\hat{\mathcal{H}}_c^{-1}\right)' + o_p\left(\bar{N}^{-1/2}\right)$, where $\breve{U}_1^c = [\breve{u}_{1,1}^{(c)}, ..., \breve{u}_{1,T}^{(c)}]'$, which implies:

$$\hat{\breve{F}}^c\hat{\mathcal{H}}_c' - \breve{F}^c = \frac{1}{\sqrt{N_1}}\breve{U}_1^c + o_p\left(\bar{N}^{-1/2}\right). \tag{B.217}$$

Here $o_p\left(\bar{N}^{-1/2}\right)$ denotes a matrix whose rows are uniformly of stochastic order $o_p\left(\bar{N}^{-1/2}\right)$. Then:

$$\begin{aligned} \hat{\lambda}_{j,i}^c &= (\hat{\breve{F}}^{c\,\prime}\hat{\breve{F}}^c)^{-1}\hat{\breve{F}}^{c\,\prime}\breve{y}_{j,i} = \frac{1}{T}\hat{\breve{F}}^{c\,\prime}\breve{y}_{j,i} \\ &= \hat{\breve{F}}^{c\,\prime}\left(\breve{F}^c\lambda_{j,i}^c + \breve{F}_j^s\lambda_{j,i}^s + \varepsilon_{j,i}\right) \\ &= \frac{1}{T}\hat{\breve{F}}^{c\,\prime}\left(\left[\hat{\breve{F}}^c\hat{\mathcal{H}}_c' - \left(\hat{\breve{F}}^c\hat{\mathcal{H}}_c' - \breve{F}^c\right)\right]\lambda_{j,i}^c + \breve{F}_j^s\lambda_{j,i}^s + \breve{\varepsilon}_{j,i}\right) \\ &= \hat{\mathcal{H}}_c'\lambda_{j,i}^c - \frac{1}{T}\hat{\breve{F}}^{c\,\prime}\left(\hat{\breve{F}}^c\hat{\mathcal{H}}_c' - \breve{F}^c\right)\lambda_{j,i}^c + \frac{1}{T}\hat{F}^{c\,\prime}\breve{F}_j^s\lambda_{j,i}^s + \frac{1}{T}\hat{\breve{F}}^{c\,\prime}\breve{\varepsilon}_{j,i}, \end{aligned}$$

for $j = 1, 2$. By writing $\hat{\breve{F}}^c = \left[\breve{F}^c + (\hat{\breve{F}}^c\hat{\mathcal{H}}_c' - \breve{F}^c)\right](\hat{\mathcal{H}}_c')^{-1}$, and rearranging terms, we get:

$$\begin{aligned} \hat{\lambda}_{j,i}^c = \hat{\mathcal{H}}_c'\Bigg\{&\lambda_{j,i}^c + (\hat{\mathcal{H}}_c')^{-1}(\hat{\mathcal{H}}_c)^{-1}\frac{1}{T}\breve{F}^{c\,\prime}\varepsilon_{j,i} + (\hat{\mathcal{H}}_c')^{-1}(\hat{\mathcal{H}}_c)^{-1}\frac{1}{T}\breve{F}^{c\,\prime}\breve{F}_j^s\lambda_{j,i}^s \\ &+(\hat{\mathcal{H}}_c')^{-1}(\hat{\mathcal{H}}_c)^{-1}(\hat{\breve{F}}^c\hat{\mathcal{H}}_c' - \breve{F}^c)'\varepsilon_{j,i} + (\hat{\mathcal{H}}_c')^{-1}(\hat{\mathcal{H}}_c)^{-1}\frac{1}{T}(\hat{\breve{F}}^c\hat{\mathcal{H}}_c' - \breve{F}^c)'\breve{F}_j^s\lambda_{j,i}^s \\ &-(\hat{\mathcal{H}}_c')^{-1}(\hat{\mathcal{H}}_c)^{-1}\frac{1}{T}\left[\breve{F}^c + (\hat{\breve{F}}^c\hat{\mathcal{H}}_c' - \breve{F}^c)\right]'\left(\hat{\breve{F}}^c\hat{\mathcal{H}}_c' - \breve{F}^c\right)\lambda_{j,i}^c\Bigg\}. \end{aligned} \tag{B.218}$$

We use equation (B.217) to bound the different terms. We have:

$$\begin{aligned}
\frac{1}{T}(\hat{\breve{F}}^c\hat{\mathcal{H}}_c' - \breve{F}^c)'\breve{\varepsilon}_{1,i} &= \frac{1}{\sqrt{N_1}T}\breve{U}_1^{c\,\prime}\breve{\varepsilon}_{1,i} + o_p(\bar{N}^{-1/2}) \\
&= (\Lambda_1'\Lambda_1/N_1)^{-1}\frac{1}{N_1T}\sum_{\ell=1}^{N_1}\sum_{t=1}^{T}\lambda_{1,\ell}\breve{\varepsilon}_{1,\ell,t}\breve{\varepsilon}_{1,i,t} + o_p(\bar{N}^{-1/2}) \\
&= (\Lambda_1'\Lambda_1/N_1)^{-1}\frac{1}{N_1T}\sum_{t=1}^{T}\lambda_{1,i}\breve{\varepsilon}_{1,i,t}^2 + (\Lambda_1'\Lambda_1/N_1)^{-1}\frac{1}{N_1T}\sum_{\ell=1,\ell\neq i}^{N_1}\sum_{t=1}^{T}\lambda_{1,\ell}\breve{\varepsilon}_{1,\ell,t}\breve{\varepsilon}_{1,i,t} \\
&\quad + o_p(\bar{N}^{-1/2}) = O_p(N_1^{-1}) + O_{p,l}[(N_1T^{\eta})^{-1/2}] + o_p(\bar{N}^{-1/2}),
\end{aligned}$$

uniformly in $1 \leq i \leq N_1$, using bounds (B.205)-(B.206) and Assumption B.8 *d)*. A similar bound holds for $j = 2$. Since N_1 grows at the same rate as N and $T^{1/2} \ll N$, we have $N_1^{-1} = o(\bar{N}^{-1/2})$. Moreover, from $\eta \geq 1/2$ and $T^{1/2} \ll N$, we have $O_{p,l}[(N_1T^{\eta})^{-1/2}] = o_p(\bar{N}^{-1/2})$. Hence, $\frac{1}{T}(\hat{\breve{F}}^c\hat{\mathcal{H}}_c' - \breve{F}^c)'\breve{\varepsilon}_{j,i} = o_p(\bar{N}^{-1/2})$, uniformly in $1 \leq i \leq N_1$. Moreover:

$$\begin{aligned}
\frac{1}{T}(\hat{\breve{F}}^c\hat{\mathcal{H}}_c' - \breve{F}^c)'\breve{F}_j^s &= \frac{1}{T\sqrt{N_1}}\breve{U}_1^{c\,\prime}\breve{F}_j^s + o_p(\bar{N}^{-1/2}) \\
&= O_p((N_1T)^{-1/2}) + o_p(\bar{N}^{-1/2}) = o_p(\bar{N}^{-1/2}),
\end{aligned}$$

and:

$$\begin{aligned}
&\frac{1}{T}\left[\breve{F}^c + (\hat{\breve{F}}^c\hat{\mathcal{H}}_c' - \breve{F}^c)\right]'\left(\hat{\breve{F}}^c\hat{\mathcal{H}}_c' - \breve{F}^c\right) \\
=\ & \frac{1}{T\sqrt{N_1}}\breve{F}^{c\,\prime}\breve{U}_1^c \\
=\ & O_p((N_1T)^{-1/2} + N_1^{-1}) + o_p(\bar{N}^{-1/2}) = o_p(\bar{N}^{-1/2}).
\end{aligned}$$

Further, from (B.215) we have $(\hat{\mathcal{H}}_c)^{-1}(\hat{\mathcal{H}}_c')^{-1} = (\hat{\mathcal{H}}_c'\hat{\mathcal{H}}_c)^{-1} = \tilde{\Sigma}_{cc}^{-1} + o_p(\bar{N}^{-1/2}) = I_{k^c} + o_p(T^{-1/2})$. Then, from (B.218) and Assumption B.8 *d)* we get:

$$\hat{\lambda}_{j,i}^c = \hat{\mathcal{H}}_c'\left[\lambda_{j,i}^c + \tilde{\Sigma}_{cc}^{-1}\frac{1}{T}\breve{F}^{c\,\prime}\breve{\varepsilon}_{j,i}\right] + o_p(\bar{N}^{-1/2}),$$

uniformly in $1 \leq i \leq N_j$. The last equation can be rewritten as

$$\hat{\lambda}_{j,i}^c = \hat{\mathcal{H}}_c'\left[\lambda_{j,i}^c + \tilde{\Sigma}_{cc}^{-1}\tilde{\Sigma}_{c,j}\lambda_{j,i}^s + \frac{1}{\sqrt{T}}\breve{w}_{j,i}^c\right] + o_p\left(\bar{N}^{-1/2}\right), \qquad j = 1,\ 2, \tag{B.219}$$

where:

$$
\begin{aligned}
\breve{w}^c_{j,i} &:= \tilde{\Sigma}_{cc}^{-1}\frac{1}{\sqrt{T}}\breve{F}^{c\,\prime}\breve{\varepsilon}_{j,i} = \tilde{\Sigma}_{cc}^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\breve{f}^c_t\breve{\varepsilon}_{j,i,t}, \\
\tilde{\Sigma}_{cc} &= \frac{1}{T}\breve{F}^{c\,\prime}\breve{F}^c = \frac{1}{T}\sum_{t=1}^{T}\breve{f}^c_t\breve{f}^{c\,\prime}_t, \qquad \tilde{\Sigma}_{c,j} = \frac{1}{T}\breve{F}^{c\,\prime}\breve{F}^s_j = \frac{1}{T}\sum_{t=1}^{T}\breve{f}^c_t\breve{f}^{s\prime}_{j,t}.
\end{aligned}
$$

d) Asymptotic expansion of $\hat{f}^s_{j,t}$

Let us now derive the asymptotic expansion of term $\hat{f}^s_{j,t}$. We start by computing the asymptotic expansion of the regression residuals $\xi_{j,i,t} := y_{j,i,t} - \hat{f}^{c\,\prime}_t\hat{\lambda}^c_{j,i}$, where we replace $\hat{f}^c_t$ with $\hat{f}^{c\,*}_t$ for $j = 2$. By substituting the asymptotic expansions in equations (B.212) and (B.219), have:

$$
\begin{aligned}
\xi_{j,i,t} &= f^{s\prime}_{j,t}\lambda^s_{j,i} + \varepsilon_{j,i,t} - \left(\hat{f}^{c\,\prime}_t\hat{\lambda}^c_{j,i} - f^{c\,\prime}_t\lambda^c_{j,i}\right) \\
&= f^{s\prime}_{j,t}\lambda^s_{j,i} + \varepsilon_{j,i,t} \\
&\quad - \left[\left(f^c_t + \frac{1}{\sqrt{N_j}}u^{(c)}_{j,t}\right)^{\prime}\left(\lambda^c_{j,i} + \tilde{\Sigma}_{cc}^{-1}\tilde{\Sigma}_{c,j}\lambda^s_{j,i} + \frac{1}{\sqrt{T}}\breve{w}^c_{j,i}\right) - f^{c\,\prime}_t\lambda^c_{j,i}\right] \\
&\quad + o_p(\bar{N}^{-1/2}) \\
&= \tilde{f}^{s\prime}_{j,t}\lambda^s_{j,i} + e_{j,i,t} + o_p(\bar{N}^{-1/2}),
\end{aligned} \tag{B.220}
$$

where we define:

$$
\tilde{f}^s_{j,t} := f^s_{j,t} - \tilde{\Sigma}_{j,c}\tilde{\Sigma}_{cc}^{-1}f^c_t, \tag{B.221}
$$

$$
e_{j,i,t} := \varepsilon_{j,i,t} - \frac{1}{\sqrt{T}}f^{c\prime}_t\breve{w}^c_{j,i} - \frac{1}{\sqrt{N_j}}u^{(c)\,\prime}_{j,t}\lambda^c_{j,i}. \tag{B.222}
$$

The term $o_p(\bar{N}^{-1/2})$ is uniform in $i = 1, ..., N_j$ and $t = 1, ..., T$ by bounds (B.187)-(B.188) and (B.203)-(B.204), and Assumption B.8 *d)*. Then, the residuals $\xi_{j,i,t}$, with $i = 1, ..., N_j$ and $t = 1, ..., T$, satisfy an approximate factor structure with factors $\tilde{f}^s_{j,t}$, loadings $\lambda^s_{j,i}$ and errors $e_{j,i,t}$, up to $o_p(\bar{N}^{-1/2})$. Differently from the proof of Proposition D.4 d) in AGGR, our error terms $e_{j,i,t}$ do not contain a factor structure at order T^{-1}.

The RP-PC estimator to the panel of residuals $\xi_{j,i,t}$ has an asymptotic expansion analogous to the one of Proposition B.2:

$$
\hat{f}^s_{j,t} = \hat{\mathcal{H}}_{s,j}^{-1}\left[\tilde{f}^s_{j,t} + \frac{1}{\sqrt{N_j}}v^{*\,s}_{j,t} + \frac{1}{T}b^{*\,s}_{j,t} + \frac{1}{\sqrt{N_jT}}d^{*\,s}_{j,t} + \vartheta^{*\,s}_{j,t}\right], \qquad j = 1, 2, \tag{B.223}
$$

where $\hat{\mathcal{H}}_{s,j}$, $j = 1, 2$, is a non-singular matrix w.p.a. 1, and: (**TO BE CHECKED!!!!!!**)

$$\begin{aligned}
v^{*\,s}_{j,t} &= \left(\frac{1}{N_j}\Lambda_j^{s\prime}\Lambda_j^s\right)^{-1}\frac{1}{\sqrt{N_j}}\Lambda_j^{s\prime}e_{j,t} \\
b^{*\,s}_{j,t} &= \left(\frac{1}{N_j}\Lambda_j^{s\,\prime}\Lambda_j^s\right)^{-1}\left(\frac{1}{T}\breve{\tilde{F}}_j^{s\,\prime}\breve{\tilde{F}}_j^s\right)^{-1}(\eta^*_{j,t})^2\breve{\tilde{f}}^s_{j,t}, \\
d^{*\,s}_{j,t} &= \left(\frac{1}{N_j}\Lambda_j^{s\,\prime}\Lambda_j^s\right)^{-1}\left(\frac{1}{T}\breve{\tilde{F}}_j^{s\,\prime}\breve{\tilde{F}}_j^s\right)^{-1}\left(\frac{1}{N_jT}\sum_{i=1}^{N_j}\sum_{r=1}^{T}e_{j,i,r}\tilde{f}^s_{j,r}\lambda^{s\,\prime}_{j,i}+...\right)\tilde{f}^s_{j,t}+...,
\end{aligned}$$

where $(\eta^*_{j,t})^2 = \underset{N_j\to\infty}{plim}\frac{1}{N_j}\sum_{i=1}^{N_j}E[e^2_{j,i,t}|\mathcal{F}_t]$ and $\tilde{F}^s_j$ denotes the matrix with rows $\tilde{f}^{s\,\prime}_{j,t}$. We have

$$\frac{1}{N_j}\Lambda_j^{s\,\prime}e_{j,t} = \frac{1}{N_j}\sum_{i=1}^{N_j}\lambda^s_{j,i}\varepsilon_{j,i,t} - \frac{1}{\sqrt{T}}\left(\frac{1}{N_j}\sum_{i=1}^{N_j}\lambda^s_{j,i}\breve{w}^{c\,\prime}_{j,i}\right)f^c_t - \frac{1}{\sqrt{N_j}}\left(\frac{\Lambda_j^{s\prime}\Lambda_j^c}{N_j}\right)u^{(c)}_{j,t}.$$

We have $\frac{1}{N_j}\sum_{i=1}^{N_j}\lambda^s_{j,i}\breve{w}^{c\,\prime}_{j,i} = O_p(N_j^{-1/2})$, $\frac{1}{N_j}\Lambda_j^{s\,\prime}\Lambda_j^c = \Sigma^{(sc)}_{\lambda,j} + O(N_j^{-1/2})$. Thus:

$$\frac{1}{N_j}\Lambda_j^{s\,\prime}e_{j,t} = \frac{1}{N_j}\sum_{i=1}^{N_j}\lambda^s_{j,i}\varepsilon_{j,i,t} - \frac{1}{\sqrt{N_j}}\left(\frac{\Lambda_j^{s\prime}\Lambda_j^c}{N_j}\right)u^{(c)}_{j,t} + o_p(\bar{N}^{-1/2}),$$

uniformly w.r.t. $t = 1, ..., T$, and:

$$\frac{1}{\sqrt{N_j}}v^{*\,s}_{j,t} = \frac{1}{\sqrt{N_j}}v^s_{j,t} + o_p(\bar{N}^{-1/2}),$$

where $v^s_{j,t} = \left(\frac{\Lambda_j^{s\prime}\Lambda_j^s}{N_j}\right)^{-1}\frac{1}{\sqrt{N_j}}\sum_{i=1}^{N_j}\lambda^s_{j,i}\varepsilon_{j,i,t} - \left(\frac{\Lambda_j^{s\prime}\Lambda_j^s}{N_j}\right)^{-1}\left(\frac{\Lambda_j^{s\prime}\Lambda_j^c}{N_j}\right)u^{(c)}_{j,t}$. Moreover:

$$b^{*\,s}_{j,t} = [\Sigma^{(ss)}_{\lambda,j}]^{-1}\eta^2_{j,t}\breve{f}^s_{j,t} + O_p(T^{-1/2} + N^{-1/2}).$$

Therefore, we have:

$$\hat{f}^s_{j,t} = \hat{\mathcal{H}}^{-1}_{s,j}\left[f^s_{j,t} - \tilde{\Sigma}_{jc}\tilde{\Sigma}^{-1}_{cc}f^c_t + \frac{1}{\sqrt{N_j}}v^s_{j,t}\right] + o_p(N_j^{-1/2}), \qquad j = 1, 2, \tag{B.224}$$

uniformly w.r.t. $t = 1, ..., T$.

Let us now show that $v^s_{j,t} = u^{(s)}_{j,t}$, the lower k^s_j-dimensional component of $u_{j,t}$. For this purpose, let us denote by $\tilde{\Sigma}_{ab}$ and $(\tilde{\Sigma}^{-1})_{ab}$, with $a, b = c, s$ the blocks of matrix $\tilde{\Sigma} \equiv \tilde{\Sigma}_{\lambda,j}$ and of its inverse $\tilde{\Sigma}^{-1}$. Then, we have:

$$v^s_{j,t} = \tilde{\Sigma}^{-1}_{ss}\frac{1}{\sqrt{N_j}}\sum_{i=1}^{N_j}\lambda^s_{j,i}\varepsilon_{j,i,t} - \tilde{\Sigma}^{-1}_{ss}\tilde{\Sigma}_{sc}u^{(c)}_{j,t},$$

and:

$$u_{j,t}^{(c)} = (\tilde{\Sigma}^{-1})_{cc} \frac{1}{\sqrt{N_j}} \sum_{i=1}^{N_j} \lambda_{j,i}^c \varepsilon_{j,i,t} + (\tilde{\Sigma}^{-1})_{cs} \frac{1}{\sqrt{N_j}} \sum_{i=1}^{N_j} \lambda_{j,i}^s \varepsilon_{j,i,t}.$$

Therefore, we get:

$$v_{j,t}^s = \tilde{\Sigma}_{ss}^{-1} [I_{k_j} - \tilde{\Sigma}_{sc} (\tilde{\Sigma}^{-1})_{cs}] \frac{1}{\sqrt{N_j}} \sum_{i=1}^{N_j} \lambda_{j,i}^s \varepsilon_{j,i,t} - \tilde{\Sigma}_{ss}^{-1} \tilde{\Sigma}_{sc} (\tilde{\Sigma}^{-1})_{cc} \frac{1}{\sqrt{N_j}} \sum_{i=1}^{N_j} \lambda_{j,i}^c \varepsilon_{j,i,t}.$$

From the property of the matrix inverse, $I_{k_j} - \tilde{\Sigma}_{sc}(\tilde{\Sigma}^{-1})_{cs} = \tilde{\Sigma}_{ss}(\tilde{\Sigma}^{-1})_{ss}$ and $\tilde{\Sigma}_{sc}(\tilde{\Sigma}^{-1})_{cc} = -\tilde{\Sigma}_{ss}(\tilde{\Sigma}^{-1})_{sc}$. Therefore, we get:

$$v_{j,t}^s = (\tilde{\Sigma}^{-1})_{ss} \frac{1}{\sqrt{N_j}} \sum_{i=1}^{N_j} \lambda_{j,i}^s \varepsilon_{j,i,t} + (\tilde{\Sigma}^{-1})_{sc} \frac{1}{\sqrt{N_j}} \sum_{i=1}^{N_j} \lambda_{j,i}^c \varepsilon_{j,i,t} = \left(\tilde{\Sigma}^{-1} \frac{1}{\sqrt{N_j}} \sum_{i=1}^{N_j} \lambda_{j,i} \varepsilon_{j,i,t} \right)^{(s)} = u_{j,t}^{(s)}.$$

Plugging the latter equation in (B.224) yields (B.197).

e) Asymptotic expansion of $\hat{\lambda}_{j,i}^s$

Let us now derive the asymptotic expansion of factor loadings estimator $\hat{\lambda}_{j,i}^s$ up to order $o_p\left(\bar{N}^{-1/2}\right)$. The analysis parallels the one in Subsection B.6.1 c). We have

$$\hat{\lambda}_{j,i}^s = \left(\breve{\hat{F}}^{s\,\prime} \breve{\hat{F}}^s \right)^{-1} \breve{\hat{F}}_j^{s\,\prime} \breve{\xi}_{j,i} = \frac{\breve{\hat{F}}_j^{s\,\prime} \breve{\xi}_{j,i}}{T},$$

where $\breve{\xi}_{j,i}$ is the i-th column of matrix $\breve{\Xi}_j$ and $\breve{\hat{F}}_j^s = [\breve{\hat{f}}_{j,1}^s, ..., \breve{\hat{f}}_{j,T}^s]'$. From equation (B.224) we have $\breve{\hat{F}}_j^s = \left(\breve{\tilde{F}}_j^s + \frac{1}{\sqrt{N_j}} \breve{U}_j^s \right) \left(\hat{\mathcal{H}}_{s,j}^{-1} \right)' + o_p\left(\bar{N}^{-1/2}\right)$, where $\breve{U}_j^s = [\breve{u}_{j,1}^{(s)}, ..., \breve{u}_{j,T}^{(s)}]'$, which implies:

$$\breve{\hat{F}}_j^s \hat{\mathcal{H}}_{j,s}' - \breve{\tilde{F}}_j^s = \frac{1}{\sqrt{N_j}} \breve{U}_j^s + o_p\left(\bar{N}^{-1/2}\right). \tag{B.225}$$

Then:

$$\begin{aligned}
\hat{\lambda}_{j,i}^s &= \frac{1}{T} \breve{\hat{F}}_j^{s\,\prime} \breve{\xi}_{j,i} = \frac{1}{T} \breve{\hat{F}}_j^{s\,\prime} \left(\breve{\tilde{F}}_j^s \lambda_{j,i}^s + \breve{e}_{j,i} \right) + o_p(\bar{N}^{-1/2}) \\
&= \frac{1}{T} \breve{\hat{F}}_j^{s\,\prime} \left(\left[\breve{\hat{F}}_j^s \hat{\mathcal{H}}_{j,s}' - \left(\breve{\hat{F}}_j^s \hat{\mathcal{H}}_{j,s}' - \breve{\tilde{F}}_j^s \right) \right] \lambda_{j,i}^s + \breve{e}_{j,i} \right) + o_p(\bar{N}^{-1/2}) \\
&= \hat{\mathcal{H}}_{j,s}' \lambda_{j,i}^s - \frac{1}{T} \breve{\hat{F}}_j^{s\,\prime} \left(\breve{\hat{F}}_j^s \hat{\mathcal{H}}_{j,s}' - \breve{\tilde{F}}_j^s \right) \lambda_{j,i}^s + \frac{1}{T} \breve{\hat{F}}_j^{s\,\prime} \breve{e}_{j,i} + o_p(\bar{N}^{-1/2}), \quad j = 1, 2,
\end{aligned}$$

uniformly in $i = 1, ..., N_j$. By writing $\breve{F}_j^s = \left[\breve{\tilde{F}}_j^s + (\breve{\hat{F}}_j^s \hat{\mathcal{H}}_{j,s}' - \breve{\tilde{F}}_j^s)\right] (\hat{\mathcal{H}}_{j,s}')^{-1}$, and rearranging terms, we get:

$$\begin{aligned}
\hat{\lambda}_{j,i}^s &= \hat{\mathcal{H}}_{s,j}' \Big\{ \lambda_{j,i}^s + (\hat{\mathcal{H}}_{j,s}')^{-1} (\hat{\mathcal{H}}_{j,s})^{-1} \frac{1}{T} \breve{\tilde{F}}_j^{s\,\prime} \breve{e}_{j,i} \\
&\quad + (\hat{\mathcal{H}}_{j,s}')^{-1} (\hat{\mathcal{H}}_{j,s})^{-1} \frac{1}{T} (\breve{\hat{F}}_j^s \hat{\mathcal{H}}_{j,s}' - \breve{\tilde{F}}_j^s)' \breve{e}_{j,i} \\
&\quad - (\hat{\mathcal{H}}_{j,s}')^{-1} (\hat{\mathcal{H}}_{j,s})^{-1} \frac{1}{T} \left[\breve{\tilde{F}}_j^s + (\breve{\hat{F}}_j^s \hat{\mathcal{H}}_{j,s}' - \breve{\tilde{F}}_j^s)\right]' (\breve{\hat{F}}_j^s \hat{\mathcal{H}}_{j,s}' - \breve{\tilde{F}}_j^s) \lambda_{j,i}^s \Big\} + o_p(\bar{N}^{-1/2}).
\end{aligned} \tag{B.226}$$

By using equations $\breve{e}_{j,i} = \breve{\varepsilon}_{j,i} - \frac{1}{\sqrt{T}} \breve{F}^c \breve{w}_{j,i}^c - \frac{1}{\sqrt{N_j}} \breve{U}_j^c \lambda_{j,i}^c$ and $\breve{\tilde{F}}_j^{s\,\prime} \breve{F}^c = 0$, equation (B.225), and paralleling the computations in Subsection B.6.1 c), we get:

$$\begin{aligned}
\frac{1}{T} \breve{\tilde{F}}_j^{s\,\prime} \breve{e}_{j,i} &= \frac{1}{T} \breve{\tilde{F}}_j^{s\,\prime} \breve{\varepsilon}_{j,i} + o_p(\bar{N}^{-1/2}), \\
\frac{1}{T} (\breve{\hat{F}}_j^s \hat{\mathcal{H}}_{j,s}' - \breve{\tilde{F}}_j^s)' \breve{e}_{j,i} &= o_p(\bar{N}^{-1/2}), \\
\frac{1}{T} \left[\breve{\tilde{F}}_j^s + (\breve{\hat{F}}_j^s \hat{\mathcal{H}}_{j,s}' - \breve{\tilde{F}}_j^s)\right]' (\breve{\hat{F}}_j^s \hat{\mathcal{H}}_{j,s}' - \breve{\tilde{F}}_j^s) &= o_p(\bar{N}^{-1/2}), \\
(\hat{\mathcal{H}}_{j,s}')^{-1} (\hat{\mathcal{H}}_{j,s})^{-1} &= (\breve{\tilde{F}}_j^{s\,\prime} \breve{\tilde{F}}_j^s / T)^{-1} + o_p(\bar{N}^{-1/2}),
\end{aligned}$$

uniformly in $i = 1, ..., N_j$. Thus, from (B.226) we get:

$$\hat{\lambda}_{j,i}^s = \hat{\mathcal{H}}_{s,j}' \left\{ \lambda_{j,i}^s + (\breve{\tilde{F}}_j^{s\,\prime} \breve{\tilde{F}}_j^s / T)^{-1} \frac{1}{T} \breve{\tilde{F}}_j^{s\,\prime} \breve{\varepsilon}_{j,i} \right\} + o_p(\bar{N}^{-1/2}),$$

uniformly in $i = 1, ..., N_j$. This equation can be written as:

$$\hat{\lambda}_{j,i}^s = \hat{\mathcal{H}}_{s,j}' \left[\lambda_{j,i}^s + \frac{1}{\sqrt{T}} \breve{w}_{j,i}^s \right] + o_p(\bar{N}^{-1/2}),$$

where:

$$\breve{w}_{j,i}^s = (\breve{\tilde{F}}_j^{s\,\prime} \breve{\tilde{F}}_j^s / T)^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^T \breve{\tilde{f}}_{j,t}^s \breve{\varepsilon}_{j,i,t}.$$

f) Asymptotic expansions up to order $o_p(T^{-1/2})$

Let us start by establishing the uniform asymptotic expansion of estimator $\hat{f}_t^c$ at order $o_p(T^{-1/2})$. From (B.208), using $(\log T)^{\bar{b}} \delta_{N,T} = o(T^{-1/2})$, for any $\bar{b} > 0$, and the uniform bounds (B.187)-(B.203), we get:

$$\hat{f}_t^c = \hat{\mathcal{H}}_c^{-1} \left(f_t^c + \frac{1}{\sqrt{N_1}} u_{1,t}^{(c)} \right) + o_p(T^{-1/2}),$$

uniformly in $t = 1, ..., T$, which yields the uniform bound for $\hat{f}_t^c$. The uniform bounds for the other estimators follow by paralleling the arguments in Subsection B.6.1 c)-e). ■

B.7 Asymptotic distribution of factors and loadings in generic group factor model

The next proposition provides the asymptotic distribution of the common and group-specific factors estimators introduced in Definitions 1 and 2 in the main body of the paper. To simplify the proof, we assume that N_1 and N_2, with $N_2 \leq N_1$, grow at the same rate, i.e., $N_2/N_1 \to \mu$ with $\mu > 0$. This condition could be relaxed at the expense of a more involved restriction on N_1, N_2, T.

PROPOSITION B.7. *Under Assumption B.1 with $\mu > 0$, and Assumptions B.2 - B.8 we have:*

$$\sqrt{N_1} \begin{bmatrix} \hat{\mathcal{H}}_c \hat{f}_t^c - f_t^c \\ \hat{\mathcal{H}}_{s,1} \hat{f}_{1,t}^s - \left(f_{1,t}^s - (F_1^s{}' F^c)(F^c{}' F^c)^{-1} f_t^c \right) \end{bmatrix} \xrightarrow{d} N\left(0, \Sigma_{u,11,t}\right), \quad (\mathcal{F}_t\text{-stably}), \tag{B.227}$$

and:

$$\sqrt{N_2} \begin{bmatrix} \hat{\mathcal{H}}_c^* \hat{f}_t^{c\,*} - f_t^c \\ \hat{\mathcal{H}}_{s,2} \hat{f}_{2,t}^s - \left(f_{2,t}^s - (F_2^s{}' F^c)(F^c{}' F^c)^{-1} f_t^c \right) \end{bmatrix} \xrightarrow{d} N\left(0, \Sigma_{u,22,t}\right), \quad (\mathcal{F}_t\text{-stably}), \tag{B.228}$$

for any t, where matrices $\hat{\mathcal{H}}_c$, $\hat{\mathcal{H}}_c^*$ and $\hat{\mathcal{H}}_{s,j}$ are such that $\hat{\mathcal{H}}_c \hat{\mathcal{H}}_c' = (\frac{1}{T} F^c{}' F^c)^{-1} + o_p(N_1^{-1/2})$, $\hat{\mathcal{H}}_c^* \hat{\mathcal{H}}_c^*{}' = (\frac{1}{T} F^c{}' F^c)^{-1} + o_p(N_2^{-1/2})$ and $\hat{\mathcal{H}}_{s,j} \hat{\mathcal{H}}_{s,j}' = (\frac{1}{T} \tilde{F}_j^s{}' \tilde{F}_j^s)^{-1} + o_p(N_j^{-1/2})$, we define $F^c = [f_1^c, ..., f_T^c]'$, $F_j^s = [f_{j,1}^s, ..., f_{j,T}^s]'$ and $\tilde{F}_j^s = F_j^s - F^c(F^{c\prime} F^c)^{-1}(F^{c\prime} F_j^s)$ for $j = 1, 2$. and $\tilde{\beta}_{j,t}^c = \beta_{j,t}^c - E[\beta_{j,t}^c f_t^c{}'] f_t^c$ is the residual of the orthogonal projection of $\beta_{j,t}^c$ onto f_t^c.

From Proposition B.7 a linear transformation of vector $\hat{f}_t^c$ (resp. $\hat{f}_t^{c*}$) estimates the common factor f_t^c at rate $1/\sqrt{N_1}$ (resp. $1/\sqrt{N_2}$) with no bias of order $1/T$. Compared to the analogous asymptotic expansion derived in Proposition 5 of AGGR, the bias terms of order $1/T$ are negligible under our Assumption B.1. The variance of the asymptotic Gaussian distribution is the upper-left (c, c) block of matrix $\Sigma_{u,11,t}$ (resp. $\Sigma_{u,22,t}$), i.e. the asymptotic variance of the estimation error $u_{1,t}$ (resp. $u_{2,t}$) for the PC vector in group 1 (resp. group 2). The estimation error for recovering the common factors from the group PC's is of order $o_p(N_1^{-1/2})$, and therefore asymptotically negligible. The estimator $\hat{f}_{j,t}^s$ approximates the residual of the sample projection of the group-j specific factor on the common factor, up to a linear transformation, at rate $1/\sqrt{N_j}$ and with an asymptotic bias of order $1/T$.

Let us now derive the asymptotic distribution of the factor loadings estimators in equations (A.13) and (A.14). For this purpose, we introduce the next assumption.

Assumption B.14. *We have for any $j = 1, 2$ and $i \geq 1$:*

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \begin{bmatrix} f_t^c \varepsilon_{j,i,t} \\ f_{j,t}^s \varepsilon_{j,i,t} \\ f_{j,t}^s \otimes f_t^c \end{bmatrix} \xrightarrow{d} N\left(0, \begin{bmatrix} \Phi_{j,i}^{cc} & \Phi_{j,i}^{cs} & 0 \\ \Phi_{j,i}^{sc} & \Phi_{j,i}^{ss} & 0 \\ 0 & 0 & \Psi_j \end{bmatrix}\right),$$

as $T \to \infty$, where:

$$\Phi_{j,i}^{cc} = \sum_{h=-\infty}^{\infty} E[f_t^c f_{t-h}^{c\prime} \varepsilon_{j,i,t} \varepsilon_{j,i,t-h}], \qquad \Phi_{j,i}^{cs} = \sum_{h=-\infty}^{\infty} E[f_t^c f_{j,t-h}^{s\prime} \varepsilon_{j,i,t} \varepsilon_{j,i,t-h}] = (\Phi_{j,i}^{sc})',$$

$$\Phi_{j,i}^{ss} = \sum_{h=-\infty}^{\infty} E[f_{j,t}^s f_{j,t-h}^{s\prime} \varepsilon_{j,i,t} \varepsilon_{j,i,t-h}], \qquad \Psi_j = \sum_{h=-\infty}^{\infty} E\left[f_{j,t}^s f_{j,t-h}^{s\prime} \otimes f_t^c f_{t-h}^{c\prime}\right].$$

Assumption B.14 states that time series averages of the error terms scaled by the factors, as well as time series averages of the cross-products of common and specific factors, are asymptotically Gaussian. It is used to show the asymptotic normality of the loadings estimators in Proposition B.8, and is implied by e.g. a mixing condition on the individual error series jointly with the factor process. The part of Assumption B.14 concerning scaled error terms corresponds to Assumption F.4 in Bai (2003).

PROPOSITION B.8. *Under Assumption B.1 with $\mu > 0$, Assumptions B.2 - B.8 and B.14 we have:*

$$\sqrt{T} \begin{bmatrix} \left(\hat{\mathcal{H}}_c'\right)^{-1} \hat{\lambda}_{j,i}^c - \lambda_{j,i}^c \\ \left(\hat{\mathcal{H}}_{s,j}'\right)^{-1} \hat{\lambda}_{j,i}^s - \lambda_{j,i}^s \end{bmatrix} \xrightarrow{d} N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \left(\Phi_{j,i}^{cc} + (\lambda_{j,i}^{s\prime} \otimes I_{k^c})\Psi_j(\lambda_{j,i}^s \otimes I_{k^c})\right) & \Psi_{j,i}^{cs} \\ \Psi_{j,i}^{sc} & \Psi_{j,i}^{ss} \end{bmatrix}\right), \quad \text{(B.229)}$$

for any j, i, where $\hat{\mathcal{H}}_c$ and $\hat{\mathcal{H}}_{s,j}$, $j = 1, 2$, are the same non-singular matrices of Proposition B.7.

The factor loadings are estimated at rate $\sqrt{T}$. Matrix $\Phi_{j,i}^{cc}$ is the asymptotic variance for cross-sectional OLS regression of data in group j on the true values of the common factor. The additional component in the asymptotic variance of estimator $\hat{\lambda}_{j,i}^c$ is due to the fact that the true values of common and group-specific factors are not orthogonal in-sample. This fact is not taken into account by the estimator of factor loadings. Finally, there are no bias terms at order N_1^{-1}, N_2^{-1} in the large sample distributions of factor loadings, since in our asymptotics $\sqrt{T}/N = o(1)$ and hence such bias terms are negligible.

B.7.1 Proof of Proposition B.7

We use the asymptotic expansions in Proposition B.6 i). Specifically, equations (B.196) and (B.197) for $j = 1$ imply:

$$\sqrt{N_1} \begin{bmatrix} \hat{\mathcal{H}}_c \hat{f}_t^c - f_t^c \\ \hat{\mathcal{H}}_{s,1} \hat{f}_{1,t}^s - (f_{1,t}^s - (F_1^{s\prime} F^c)(F^{c\prime} F^c)^{-1} f_t^c) \end{bmatrix} = u_{1,t} + o_p(1).$$

From Assumptions B.3 and B.5 *a)*, we have $u_{1,t} \xrightarrow{d} N\left(0, \Sigma_{u,11,t}\right)$, $\mathcal{F}_t$-stably. Then, the asymptotic distribution in (B.227) follows. The asymptotic distribution in (B.228) can be establish along similar lines.

B.7.2 Proof of Proposition B.8

We prove Proposition B.8 by the asymptotic expansions in Proposition B.6 i), by keeping only terms up to $o_p(T^{-1/2})$. Specifically, equation (B.198) implies:

$$\begin{aligned}
\sqrt{T}\left[\left(\hat{\mathcal{H}}_c'\right)^{-1}\hat{\lambda}_{j,i}^c - \lambda_{j,i}^c\right] &= w_{j,i}^c + (F^{c\prime}F_j^s/\sqrt{T})\lambda_{j,i}^s + o_p(1) \\
&= \frac{1}{\sqrt{T}}\sum_{t=1}^{T} f_t^c(\varepsilon_{j,i,t} + f_{j,t}^{s\,\prime}\lambda_{j,i}^s) + o_p(1) \\
&= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[f_t^c\varepsilon_{j,i,t} + (\lambda_{j,i}^{s\prime}\otimes I_{k^c})(f_{j,t}^s\otimes f_t^c)\right] + o_p(1).
\end{aligned}$$

Moreover, equation (B.199) imply:

$$\sqrt{T}\left[\left(\hat{\mathcal{H}}_{s,j}'\right)^{-1}\hat{\lambda}_{j,i}^s - \lambda_{j,i}^s\right] = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} f_{j,t}^s\varepsilon_{j,i,t} + o_p(1).$$

Thus, we get:

$$\sqrt{T}\begin{bmatrix}\left(\hat{\mathcal{H}}_c'\right)^{-1}\hat{\lambda}_{j,i}^c - \lambda_{j,i}^c \\ \left(\hat{\mathcal{H}}_{s,j}'\right)^{-1}\hat{\lambda}_{j,i}^s - \lambda_{j,i}^s\end{bmatrix} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\begin{bmatrix} f_t^c\varepsilon_{j,i,t} + (\lambda_{j,i}^{s\prime}\otimes I_{k^c})(f_{j,t}^s\otimes f_t^c) \\ f_{j,t}^s\varepsilon_{j,i,t}\end{bmatrix} + o_p(1).$$

Then, Assumption B.14 yields (B.229). ■