

Is Industrial Production Still the Dominant Factor for the US Economy?*

E. Andreou,[†] P. Gagliardini,[‡] E. Ghysels,[§] M. Rubin[¶]

This version: September 14, 2016

First draft: October, 2014

Abstract

We propose a new class of large approximate factor models which enable us to study the full spectrum of quarterly Industrial Production (IP) sector data combined with annual non-IP sectors of the economy. We derive the large sample properties of the estimators and test statistics for the new class of unobservable factor models involving mixed frequency data and common as well as frequency-specific factors. Despite the growth of service sectors, we find that a single common factor explaining 90% of the variability in IP output growth index also explains 60% of total GDP output growth fluctuations. A single low frequency factor unrelated to manufacturing explains 14% of GDP growth. The picture with a structural factor model featuring technological innovations is quite different. Last but not least, our identification and inference methodology rely on novel results on group factor models that are of general interest beyond the mixed frequency framework and the application of the paper.

*We thank K. A. Aastveit, G. Barone-Adesi, M. Deistler, M. Del Negro, R. Engle, D. Giannone, T. Götz, K. Horl'ová, C. Hurlin, A. Levchenko, S. Ng, A. Onatski, C. Pérignon, G. Urga, M. Watson, B. Werker for their useful comments as well as participants at the 2013 (*EC*)² Conference on “The Econometric Analysis of Mixed Frequency Data” in Nicosia, the 2014 “8th ECB Workshop on Forecasting Techniques” in Frankfurt am Main, the 2014 ESEM in Toulouse, the 2015 “8th Financial Risk International Forum on Scenarios, Stress and Forecast in Finance” in Paris, the 2015 “2nd Workshop on High-Dimensional Time Series in Macroeconomics and Finance” in Vienna, the “Swiss Finance Institute Research Days 2015” in Gerzensee, the 2015 “EABCN/Norges Bank Conference: Econometric methods for business cycle analysis, forecasting and policy simulations” in Oslo, the 2015 “21st International Panel Data Conference” in Budapest, the 2015 “NBER-NSF Time Series Conference” in Vienna, the 2015 “Econometrics of high-dimensional risk networks” conference at University of Chicago, the 2016 “International Association for Applied Econometrics Annual Conference” at University of Milan-Bicocca, and seminar participants at the University of Alberta, the University of Amsterdam, the University of Cambridge, the University of Pennsylvania, and the Swiss National Bank. The first author would also like to thank the European Community Horizon 2020 ERC-POC 2014 grant 640924.

[†]University of Cyprus and CEPR (elena.andreou@ucy.ac.cy).

[‡]Università della Svizzera Italiana (USI, Lugano) and Swiss Finance Institute (patrick.gagliardini@usi.ch).

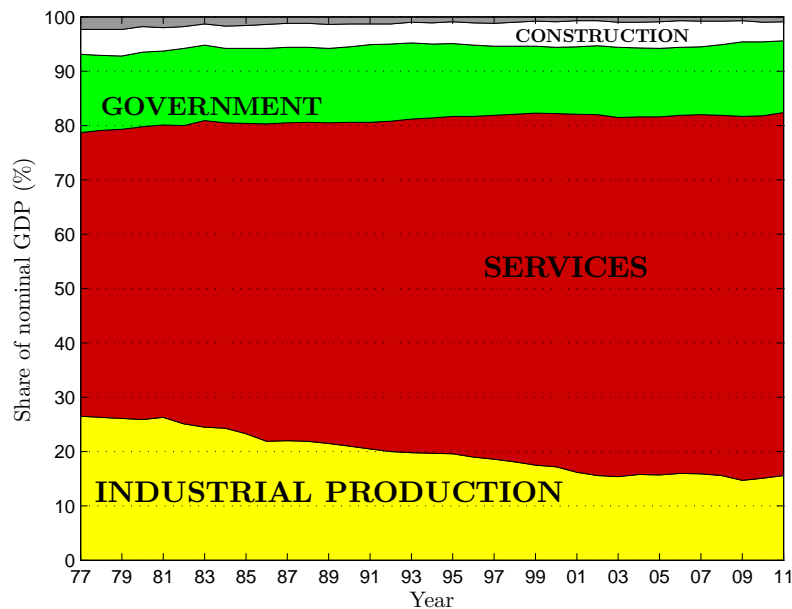
[§]University of North Carolina - Chapel Hill and CEPR (eghysels@unc.edu).

[¶]University of Bristol (mirco.rubin@bristol.ac.uk).

1 Introduction

In the public arena it is often claimed that manufacturing has been in decline in the US and most jobs have migrated overseas to lower wage countries. First, we would like to nuance this observation somewhat. It is true, as Figure 1 below clearly shows, that the share of the Industrial Production (IP) sector has been in decline since the late 70's, which is the beginning of our sample period. However, does size matter? The fact that the size shrank does not necessarily exclude the possibility that the IP sector still is a key factor, or even the dominant factor, of total US output. We study the validity of this question using novel econometric methods designed to deal with some of the challenging data issues one encounters when trying to address the problem. When studying the role of the IP sector we face

Figure 1: Sectoral decomposition of US nominal GDP.



The figure displays the evolution from 1977 to 2011 of the sectoral decomposition of US nominal GDP. We aggregate the shares of different sectors available from the website of the US Bureau of Economic Analysis, according to their North American Industry Classification System (NAICS) codes, in 5 different *macro* sectors: Industrial Production (yellow), Services (red), Government (green), Construction (white), Others (grey).

a conundrum. On the one hand, we have fairly extensive data on industrial production which consists of 117 sectors that make up aggregate IP, each sector roughly corresponding to a four-digit industry classification using NAICS. These data are published monthly, and therefore cover a rich time series

and cross-section. In our analysis, the data are sampled at quarterly frequency, for reasons explained later in the paper, and consists of over 16,000 time series data points counting all quarters from 1977 until 2011 (end of our data set) for each of the 42 sectors. On the other hand, contrary to IP, we do not have monthly or quarterly data about the cross-section of US output across non-IP sectors, but we do so on an annual basis. Indeed, the US Bureau of Economic Analysis provides Gross Domestic Product (GDP) and Gross Output by industry - not only IP sectors - annually. In our empirical analysis we use data on 42 non-IP sectors. If we were to study all sectors annually, we would be left with roughly 4000 data points for IP - a substantial loss of information.

Economists have proposed different models to explain how various sectors in the economy interact. Some rely on aggregate shocks which affect all sectors at once. Foerster, Sarte, and Watson (2011), who use an approximate factor model estimated with quarterly data, find that nearly all of IP variability is associated with a small number of common factors - even a single common factor suffices according to their findings. To what extent does the single common factor which drives the cross-sectional variation of IP sectors also affect the rest of the economy, in particular in light of the fact that the services sector grew in relative size? To put it differently, can we maintain a common factor view if we expand beyond IP sectors? Or should we think about sector-specific shocks affecting aggregate US output? If so, are these IP sector shocks, or rather services sector ones?

The technical problem we solve and the theoretical contributions we make go far beyond the specific application studied in the current paper. The technical problem can be described as follows. Suppose one has two sets of factors, say $h_{1,t}$, $h_{2,t}$, estimated from two separate panels, and we want to know how many factors are common between them. We introduce a test for the number of canonical correlations between $h_{1,t}$ and $h_{2,t}$ equal to one and do this in a large T and large N panel data context. What complicates the asymptotics is the fact that we deal with estimated factors, i.e. the first stage estimation error affecting the canonical correlation analysis. In fact the asymptotics are non-standard in terms of convergence rates and a non-trivial bias correction.

Testing the hypothesis how many canonical correlations are equal to one is of interest in several instances far beyond the context of the empirical challenge we address. For example what is the common space spanned by principal components extracted from stock returns (akin to rotations of Fama and French factors) and principal components extracted from macroeconomic series (like Stock and Watson factors)? Surprisingly, there are no formal tests available. If we consider the variance-

covariance matrix of the stacked factors, i.e. $V[(h'_{1,t}, h'_{2,t})']$, then our results also relate to testing the rank of a symmetric positive semi-definite matrix. Our theoretical results also contribute to a number of open questions in the literature on testing the rank of such matrices. In Section 4.3 we discuss how our contributions relate to the various existing literatures.

How does testing for common factor spaces relate to our empirical problem of interest? Using the terminology of the approximate factor model literature, we have a panel consisting of N_H IP sector growth series sampled across MT time periods, where $M = 4$ for quarterly data and $M = 12$ for monthly data, with T the number of years. Moreover, we also have a panel of N_L non-IP sectors - such as services and construction for example - which is only observed over T periods. Hence, generically speaking we have a high frequency panel data set of size $N_H \times MT$ and a corresponding low frequency panel data set of size $N_L \times T$. The issue we are interested in can be thought of as follows. We allow for the presence of three types of factors: (1) those which explain variations in both panels - say g^C , and therefore are economy-wide factors, (2) those exclusively pertaining to IP sector movements - say g^H , and finally (3) those exclusively affecting non-IP, denoted by g^L . Hence, we could have (1) common, (2) high frequency and (3) low frequency factors. We use superscripts C , H and L because the theory we develop is generic and pertains to common (C), high frequency (H) and low frequency (L) factors.

The purpose of this paper is to propose large scale approximate factor models in the spirit of Bai and Ng (2002), Stock and Watson (2002a), Bai (2003), Bai and Ng (2006a), and extend their analysis to mixed frequency data settings. A number of mixed frequency factor models have been proposed in the literature, although they almost exclusively rely on small cross-sections.¹ We approach the problem from a different angle. We start with a setup which identifies factors common to both high and low frequency data panels, the aforementioned g^C , and factors specific to the high and low frequency data. Our approach amounts to writing the model as a grouped factor model. The idea to apply grouped factor analysis to mixed frequency data is novel and has many advantages in terms of identification and estimation. In the proposed identification strategy, the groups correspond to panels observed at

¹See for example, Mariano and Murasawa (2003), Nunes (2005), Aruoba, Diebold, and Scotti (2009) Frale and Monteforte (2010), Marcellino and Schumacher (2010) and Banbura and Rünstler (2011), among others. Stock and Watson (2002b) in their Appendix A, propose a modification of the EM algorithm of Dempster, Laird, and Rubin (1977) to estimate high frequency factors from potentially large unbalanced panels, with mixed-frequency being a special case. Moreover, Jungbacker, Koopman, and van der Wel (2011) introduce a computationally efficient EM algorithm for the maximum likelihood estimation of a high-dimensional linear factor model with missing data.

different sampling frequencies. While there is a literature on how to estimate factors in a grouped model setting, there does not exist a general unifying asymptotic theory for large panel data.² As already noted, we propose estimators for the common and group specific factors, and an inference procedure for the number of common and group specific factors based on canonical correlation analysis of the principal components (PCs) estimators on each subgroup. The procedure is therefore general in scope and also of interest in many applications other than the one considered in the current paper. We study the large sample properties of our estimators and inference procedure as $T, N_H, N_L \rightarrow \infty$.

Our empirical application revisits the analysis of Foerster, Sarte, and Watson (2011) who use factor analytic methods to decompose industrial production into components arising from aggregate shocks and idiosyncratic sector-specific shocks. They focus exclusively on the IP sectors of the US economy. We find that a single common factor explains around 90% of the variability in the aggregate IP output growth index, and a factor specific to IP has very little additional explanatory power. This implies that the single common factor can be interpreted as an Industrial Production factor. Moreover, more than 60% of the variability of GDP output growth in service sectors, such as Transportation and Warehousing services, is also explained by the common factor. A single low frequency factor unrelated to manufacturing, explaining around 14 % of GDP growth fluctuations, drives the comovement of non-IP sectors such as Construction and Government. Note the great advantage of the mixed frequency setting - compared to the single frequency one - in the context of our IP and GDP sector application. The mixed frequency panel setting allows us to identify and estimate the *high frequency* values of factors common to IP and non-IP sectors. With IP (i.e. high frequency) data only we cannot assess what is common with the non-IP sectors. With low frequency data only, we cannot estimate the high frequency common factors from a large cross-section.

We re-examine whether the common factor reflects sectoral shocks that have propagated by way of input-output linkages between service sectors and manufacturing. A structural factor analysis indicates that both low and high frequency aggregate shocks continue to be the dominant source of variation in the US economy. The propagation mechanisms are very different, however, from those identified by Foerster, Sarte, and Watson (2011). Looking at technology shocks instead of output growth, it does not appear that a common factor explaining IP fluctuations is a dominant one for the entire

²In Section 2.4 we provide a short review of the literature on group factor models, including recent contributions related to the topic of the current paper.

economy. A factor specific to technological innovations in IP sectors is more important for the IP sector shocks and a low frequency factor which appears to explain variation in information industry as well as professional and business services innovations plays relatively speaking a more important role. Hence, when it comes to innovation shocks, IP is no longer the dominant factor.

The rest of the paper is organized as follows. In Section 2 we introduce the formal model and discuss identification. In Section 3 we study estimation and inference on the number of common factors. The large sample theory appears in Section 4. Section 5 presents briefly the results of a Monte Carlo study. Section 6 covers the empirical application. Section 7 concludes the paper. The Technical Appendix of the paper provides regularity conditions (Section A) and proofs of propositions and theorems (Section B). The Online Appendix (henceforth OA), Section C, contains additional theoretical results on identification and estimation, an extensive description of the dataset used in the empirical application, additional empirical results, and a procedure for the extraction of technology shocks in a mixed frequency setting. Finally, the OA, Section D, contains the details about the Monte Carlo simulation design and results.³

2 Model specification and identification

We consider a setting where both low and high frequency data are available. Let $t = 1, 2, \dots, T$ be the low frequency time units. Each period $(t-1, t]$ is divided into M subperiods with high frequency dates $t-1 + m/M$, with $m = 1, \dots, M$. Moreover, we assume a panel data structure with a cross-section of size N_H of high frequency data and N_L of low frequency data. It will be convenient to use a double time index to differentiate low and high frequency data. Specifically, we let $x_{m,t}^{Hi}$, for $i = 1, \dots, N_H$, be the high frequency data observation i during subperiod m of low frequency period t . Likewise, we let x_t^{Li} , with $i = 1, \dots, N_L$, be the observation of the i^{th} low-frequency series at t . These observations are gathered into the N_H -dimensional vectors $x_{m,t}^H, \forall m$, and the N_L -dimensional vector x_t^L , respectively.

We have a latent factor structure in mind to explain the panel data variation for both the low and high frequency observations. To that end, we assume that there are three types of factors, which we denote by respectively $g_{m,t}^C, g_{m,t}^H$ and $g_{m,t}^L$. The former represents factors which affect both high and low frequency data (throughout we use superscript C for common), whereas the other two types of

³The OA, Sections C and D are available at <https://sites.google.com/site/mircorubin/>.

factors affect exclusively high (superscript H) and low (marked by L) frequency data. We denote by k^C , k^H and k^L , the dimensions of these factors. The latent factor model with high frequency data sampling is:

$$\begin{aligned} x_{m,t}^H &= \Lambda_{HC} g_{m,t}^C + \Lambda_H g_{m,t}^H + e_{m,t}^H, \\ x_{m,t}^{L*} &= \Lambda_{LC} g_{m,t}^C + \Lambda_L g_{m,t}^L + e_{m,t}^L, \end{aligned} \quad (2.1)$$

where $m = 1, \dots, M$ and $t = 1, \dots, T$, and Λ_{HC} , Λ_H , Λ_{LC} and Λ_L are matrices of factor loadings. The vector $x_{m,t}^{L*}$ is unobserved for each high frequency subperiod and the measurements, denoted by x_t^L , depend on the observation scheme, which can be either flow sampling or stock sampling (or some general linear scheme). In the remainder of this section we study identification of the model for the case of flow sampling, corresponding to the empirical application covered later in the paper.⁴ For this purpose, we develop a group-factor model representation.

2.1 Group-factor model representation

In the case of flow sampling, the low frequency observations are the sum (or average) of all $x_{m,t}^{L*}$ across all m , that is: $x_t^L = \sum_{m=1}^M x_{m,t}^{L*}$. Then, model (2.1) implies:

$$\begin{aligned} x_{m,t}^H &= \Lambda_{HC} g_{m,t}^C + \Lambda_H g_{m,t}^H + e_{m,t}^H, \quad m = 1, \dots, M, \\ x_t^L &= \Lambda_{LC} \sum_{m=1}^M g_{m,t}^C + \Lambda_L \sum_{m=1}^M g_{m,t}^L + \sum_{m=1}^M e_{m,t}^L. \end{aligned} \quad (2.2)$$

Let us define the aggregated variables and innovations $x_t^H := \sum_{m=1}^M x_{m,t}^H$, $\bar{e}_t^U := \sum_{m=1}^M e_{m,t}^U$, $U = H, L$, and the aggregated factors:

$$\bar{g}_t^U := \sum_{m=1}^M g_{m,t}^U, \quad U = C, H, L.$$

⁴The identification with stock sampling is discussed in the OA, Section C.1. It is worth noting though that any linear sampling scheme leading to a representation of the model analogous to the group-factor model in equation (2.3) or (2.4) - as discussed shortly in Section C.1 - is compatible with the identification and estimation strategies of this paper.

Then we can stack the observations x_t^H and x_t^L and write:

$$\begin{bmatrix} x_t^H \\ x_t^L \end{bmatrix} = \begin{bmatrix} \Lambda_{HC} & \Lambda_H & 0 \\ \Lambda_{LC} & 0 & \Lambda_L \end{bmatrix} \begin{bmatrix} \bar{g}_t^C \\ \bar{g}_t^H \\ \bar{g}_t^L \end{bmatrix} + \begin{bmatrix} \bar{e}_t^H \\ \bar{e}_t^L \end{bmatrix}. \quad (2.3)$$

The last equation corresponds to a group factor model, with common factor \bar{g}_t^C and group-specific factors \bar{g}_t^H, \bar{g}_t^L .

To further generalize the setup, and draw directly upon the group-factor structure, we will consider the generic specification. To separate the specific from the generic case, we will change notation slightly. Namely, we keep the notation introduced so far with high and low frequency data, temporal aggregation, etc. for the mixed frequency setting further used in the empirical application, and use the following notation for the generic grouped factor model setting:

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} \Lambda_1^c & \Lambda_1^s & 0 \\ \Lambda_2^c & 0 & \Lambda_2^s \end{bmatrix} \begin{bmatrix} f_t^c \\ f_{1,t}^s \\ f_{2,t}^s \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix}, \quad (2.4)$$

where $y_{j,t} = [y_{j,1t}, \dots, y_{j,N_j t}]'$, $\Lambda_j^c = [\lambda_{j,1}^c, \dots, \lambda_{j,N_j}^c]'$, $\Lambda_j^s = [\lambda_{j,1}^s, \dots, \lambda_{j,N_j}^s]'$ and $\varepsilon_{j,t} = [\varepsilon_{j,1t}, \dots, \varepsilon_{j,N_j t}]'$, with $j = 1, 2$. The dimensions of the common factor f_t^c and the group-specific factors $f_{1,t}^s, f_{2,t}^s$ are k^c, k_1^s and k_2^s , respectively. In the case of no common factors, we set $k^c = 0$, while in the case of no group-specific factors we set $k_j^s = 0, j = 1, 2$. The group-specific factors $f_{1,t}^s$ and $f_{2,t}^s$ are orthogonal to the common factor f_t^c . Since the unobservable factors can be standardized, we assume:

$$E \begin{bmatrix} f_t^c \\ f_{1,t}^s \\ f_{2,t}^s \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \quad \text{and} \quad V \begin{bmatrix} f_t^c \\ f_{1,t}^s \\ f_{2,t}^s \end{bmatrix} = \begin{bmatrix} I_{k^c} & 0 & 0 \\ 0 & I_{k_1^s} & \Phi \\ 0 & \Phi' & I_{k_2^s} \end{bmatrix}, \quad (2.5)$$

where I_k denotes the identity matrix of order k . We allow for a non-zero covariance Φ between group-specific factors.⁵ In the mixed-frequency model (2.1), the latent common and group-specific factors

⁵In the main body of the text we only highlight some of the key assumptions underpinning our analysis. In the Appendix, Section A, we provide the detailed list of assumptions.

are normalized such that \bar{g}_t^U , $U = C, H, L$, satisfy the counterpart of (2.5).

2.2 Separation of common and group-specific factors

In standard linear latent factor models, the normalization induced by an identity factor variance-covariance matrix identifies the factor space up to an orthogonal rotation (and change of signs). Under a suitable identification condition, the rotational invariance of the group factor model (2.4) - (2.5) allows only for separate rotations among the components of $f_{1,t}^s$, among those of $f_{2,t}^s$, and among those of f_t^c . The rotational invariance of model (2.4) - (2.5) therefore maintains the interpretation of common factor and specific factors. More formally, the following proposition gives a sufficient condition for the identification of the group factor model.

PROPOSITION 1. *Assume that the matrices $\Lambda_1 = \begin{bmatrix} \Lambda_1^c & \Lambda_1^s \end{bmatrix}$ and $\Lambda_2 = \begin{bmatrix} \Lambda_2^c & \Lambda_2^s \end{bmatrix}$ are full column-rank, for N_1, N_2 large enough. Then, the factor model is identifiable: the data $[y'_{1,t}, y'_{2,t}]'$ satisfy a group factor model as (2.4) - (2.5) with stacked factor $(f_t^{c'}, f_{1,t}^{s'}, f_{2,t}^{s'})'$ replaced by $(\tilde{f}_t^{c'}, \tilde{f}_{1,t}^{s'}, \tilde{f}_{2,t}^{s'})'$ defined by the linear transformation*

$$\begin{bmatrix} f_t^c \\ f_{1,t}^s \\ f_{2,t}^s \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix} \begin{bmatrix} \tilde{f}_t^c \\ \tilde{f}_{1,t}^s \\ \tilde{f}_{2,t}^s \end{bmatrix} \quad (2.6)$$

if, and only if, the matrix $A = (A_{i,j})$ is a block-diagonal orthogonal matrix.

Proof: See Appendix B.1.

The full-rank condition in Proposition 1 is a standard condition for separate identification of the pervasive factor spaces in the two subgroups.⁶ Proposition 1 shows that this condition - together with the normalization restrictions in (2.5) - is also sufficient for identifiability of the common factor f_t^c , the group-specific factors $f_{j,t}^s$, and the factor loadings Λ_j^c, Λ_j^s , up to separate rotations.

By the same token in the mixed frequency setting of equation (2.3), the aggregated factors $\bar{g}_t^C, \bar{g}_t^H, \bar{g}_t^L$, and the factor loadings $\Lambda_{HC}, \Lambda_{LC}, \Lambda_H, \Lambda_L$, are identified. Once the factor loadings are identified from (2.3), the values of the common and high frequency factors for subperiods $m = 1, \dots, M$ are

⁶The identification condition in Proposition 1 is implied by Assumption A.2 in the Appendix, Section A, and implies that the matrix of loadings in the right hand side of equation (2.4) is full-rank.

identifiable by cross-sectional regression of the high frequency data on loadings Λ_{HC} and Λ_H in (2.1). More precisely, $g_{m,t}^C$ and $g_{m,t}^H$ are identified by regressing $x_{m,t}^{Hi}$ on $\lambda_{HC,i}$ and $\lambda_{H,i}$ across $i = 1, 2, \dots$, for any $m = 1, \dots, M$ and any t . Hence, with flow sampling, we can identify the common factor $g_{m,t}^C$ and the high frequency factor $g_{m,t}^H$ at all high frequency subperiods. On the other hand, only $\bar{g}_t^L = \sum_{m=1}^M g_{m,t}^L$, i.e. the within-period sum of the low frequency factor, is identifiable by the paired panel data set consisting of x_t^H combined with x_t^L . This is not surprising, since we have no HF observations available for the LF process.

2.3 Identification of the common and group-specific factor spaces from canonical correlations and directions

We now show that, under the conditions of Proposition 1, the factor space dimensions k^c , k_1^s , k_2^s are identifiable as well, and a constructive identification strategy for these dimensions and the corresponding factor spaces is made possible by canonical correlation analysis. In the interest of generality, let us again consider the generic setting of equation (2.4) and let $k_j = k^c + k_j^s$, for $j = 1, 2$, be the dimensions of the factor spaces for the two groups, and define $\underline{k} = \min(k_1, k_2)$. We collect the factors of each group in the k_j -dimensional vectors $h_{j,t}$:

$$h_{j,t} := \begin{bmatrix} f_t^c \\ f_{j,t}^s \end{bmatrix}, \quad j = 1, 2, \quad t = 1, \dots, T,$$

and define their variance and covariance matrices:

$$V_{j\ell} := E(h_{j,t} h_{\ell,t}'), \quad j, \ell = 1, 2.$$

Before stating the main identification result, let us first recall a few basic facts from canonical analysis (see e.g. Anderson (2003) and Magnus and Neudecker (2007)). Let ρ_ℓ , $\ell = 1, \dots, \underline{k}$ denote the canonical correlations between $h_{1,t}$ and $h_{2,t}$. The largest \underline{k} eigenvalues of matrices

$$R = V_{11}^{-1} V_{12} V_{22}^{-1} V_{21}, \quad \text{and} \quad R^* = V_{22}^{-1} V_{21} V_{11}^{-1} V_{12},$$

are the same, and are equal to the squared canonical correlations ρ_ℓ^2 , $\ell = 1, \dots, \underline{k}$ between $h_{1,t}$ and $h_{2,t}$. The associated eigenvectors $w_{1,\ell}$ (resp. $w_{2,\ell}$), with $\ell = 1, \dots, \underline{k}$, 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 allow to construct the canonical variables from vector $h_{1,t}$ (resp. $h_{2,t}$). The matrices $w_j = [w_{j,1}, \dots, w_{j,\underline{k}}]$, $j = 1, 2$, are such that $w'_j V_{jj} w_j = I_{\underline{k}}$, $j = 1, 2$. Moreover, when $\rho_\ell \neq 0$, then

$$w_{1,\ell} = \frac{1}{\rho_\ell} V_{11}^{-1} V_{12} w_{2,\ell}, \quad w_{2,\ell} = \frac{1}{\rho_\ell} V_{22}^{-1} V_{21} w_{1,\ell}. \quad (2.7)$$

PROPOSITION 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 smaller 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, with $j = 1, 2$. Then, we have $f_t^c = W'_j h_{j,t}$ (up to a rotation matrix), for $j = 1, 2$.*
- iii) *If $k^c = 0$, all canonical correlations between $h_{1,t}$ and $h_{2,t}$ are strictly smaller 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 $f_{j,t}^s = W_j^{s'} h_{j,t}$ (up to a rotation matrix) for $j = 1, 2$.*

Proof: See Appendix B.2.

Proposition 2 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 $f_{j,t}^s$, $j = 1, 2$, (up to a rotation) are identifiable from information that can be inferred by disjoint Principal Component Analysis (PCA) on the two subgroups. Indeed, disjoint PCA on the two subgroups allows us to identify the dimensions k_1 , k_2 , and vectors $h_{1,t}$ and $h_{2,t}$ up to linear transformations. The latter indeterminacy does not prevent identifiability of the common and group-specific factors from Proposition 2. More precisely, from the subpanel j we can identify the vector $h_{j,t}$ up to a non-singular matrix U_j , say, $j = 1, 2$. Under the transformation $h_{j,t} \rightarrow U_j h_{j,t}$, the matrices R and R^* are transformed such that $R \rightarrow (U'_1)^{-1} R U'_1$ and $R^* \rightarrow (U'_2)^{-1} R^* U'_2$. Therefore, the matrices of canonical directions W_1 and W_2 are transformed such as $W_j \rightarrow (U'_j)^{-1} W_j$, $j = 1, 2$. Therefore, the quantities

$W_j' h_{j,t}$, $j = 1, 2$, are invariant under such transformations.

One may wonder why we do not apply canonical correlation analysis directly to the (aggregated) high and low frequency data - avoiding the first step of computing PCs since the extra step considerably complicates the asymptotics and actually entails a novel contribution of the paper. What makes the first step of computing PCs necessary is the fact that canonical correlations applied to the raw data may not necessarily uncover pervasive factors.⁷ One may also wonder why we cannot stack all groups into one panel and apply standard PCA to estimate common factors as in Bekaert, Hodrick, and Zhang (2009) and Korajczyk and Sadka (2008), for instance. Unfortunately, this is also not a solution either, as discussed in Boivin and Ng (2006), Goyal, Pérignon, and Villa (2008), Wang (2012) and Breitung and Eickmeier (2016). In fact, in the case of a model with k^c common factors, a finite number of groups, and a positive number of group-specific factors, the estimate of the common factor obtained from the first k^c principal components of the pooled data is inconsistent due to the correlation in the residuals terms arising from the group specific factors.

2.4 Related literature

By related literature we mean specifically the literature on group factor models. In Subsection 4.3, which comes after we present the detailed theoretical results, we will discuss a broader literature related to our paper.

There exist a number of papers on group factor models, sometimes also named “multilevel factor models”, or “hierarchical factor models”. Many of these are rooted in the statistics literature and deal with large T and finite cross-sections, i.e. $\max_j (N_j) < \infty$.⁸ Recently, Bayesian methods for state space models have been applied by Moench and Ng (2011) and Moench, Ng, and Potter (2013) for relatively large scale hierarchical factor models. Moreover, Hallin and Liska (2011) extend the estimator based on dynamic PCs to their dynamic factor model with block structure, which is similar to the grouped factor models studied by the above literature.⁹ The main contribution of Goyal, Pérignon,

⁷A simple example would be to add an anomalous series to one panel and repeat the series to the other one. The canonical correlation analysis applied to the raw data will uncover the presence of the anomalous series in both panels, creating a “spurious” unit canonical correlation.

⁸See for example, Tucker (1958), Dauxois and Pousse (1975), Browne (1979), Krzanowski (1979), Browne (1980), Cho (1984), Flury (1984), Flury (1988), Chen and Robinson (1989), Schott (1988), Schott (1991), Dauxois, Romain, and Viguier (1993), Gregory and Head (1999), Schott (1999), Viguier-Pla (2004), and Kose, Otrok, and Whiteman (2008).

⁹Hallin and Liska (2011) deploy the information criteria of Hallin and Liska (2007) for inference on factor dimensions.

and Villa (2008) is the extension of the results of the classical statistical literature on group factor models to the case of approximate group factor models, even if they do not derive analytically any asymptotic results.

Our work is most closely related to Chen (2010, 2012), Wang (2012), Ando and Bai (2015) and Breitung and Eickmeier (2016), who handle the large dimensional T and N case, where $N = \min_j (N_j)$. To the best of our knowledge, the existing literature does not give a comprehensive asymptotic treatment yet of group factor models in a large dimensional setting. Hence, our contribution is to provide a general comprehensive analysis that deals with (1) asymptotic distributional theory for inference on the numbers of common and group-specific factors, and (2) consistency and asymptotic normality of the estimators for factor spaces and loadings, within a unified framework based on principal component and canonical correlation analysis.

More specifically, we use the results of canonical correlation analysis in Proposition 2 for conducting inference on the factor space dimensions by developing an asymptotic theory for estimating the number of canonical correlations equal to one. This asymptotic theory is nonstandard. Indeed, if the PCs in the two groups were observed, then the problem of testing for unit canonical correlations among them would have a degenerate feature, because it involves testing for deterministic relationships between random vectors. The estimation errors of the PCs drive the asymptotic distribution of the statistic, with a nonstandard convergence rate of $N\sqrt{T}$. Moreover, we also have to deal with a bias adjustment term, due to a problem akin to errors in the variables. The positive bias adjustment term re-centers the distribution to yield an asymptotic Gaussian density of the test statistic. The next two sections provide the details of these derivations.

It is worth explaining in more depth the relationship with the aforementioned existing literature. Our Proposition 1 corresponds to Proposition 1 in Wang (2012). However, Wang (2012) does not exploit canonical correlation analysis - as we do in Proposition 2 - for the identification of factor space dimensions, which are instead deduced from the number of pervasive factors in the subgroups, and in the overall panel, under a restrictive condition. For given factor space dimensions, Wang (2012), Ando and Bai (2015) and Breitung and Eickmeier (2016) consider an estimator defined by the minimization of a Least Square (LS) criterion with respect to factor values and factor loadings. The consistency of this procedure is formally proven by Ando and Bai (2015).¹⁰ To compute the minimizer of the non-

¹⁰Ando and Bai (2015) generalize the group-factor model by including both observable and unobservable factors. They

convex LS criterion, Wang (2012), Ando and Bai (2015) and Breitung and Eickmeier (2016) propose various iterative procedures. In particular, Wang (2012) exploits the first order conditions and describes an iterative PC estimator. The large sample properties of the iterated PC estimator are, however, not provided. In addition, we show in OA, Section C.3, that the iterative procedure is not operational as the resulting fixed point equations do not have a unique solution. We avoid such issues by proposing an estimator in closed form (up to the eigenvalue-eigenvector decompositions underlying principal component and canonical correlation analysis). In a framework with two groups and abstracting from classification issues, the estimation procedure of Chen (2012) relies on the fact that the number of common factors corresponds to the number of eigenvalues equal to 2 of the variance-covariance matrix of the stacked pervasive factors extracted in the two groups, and the common factors are spanned by the associated eigenvectors. Chen (2012) uses this result to estimate the common factors and derive the consistency of the estimators - assuming that the number of common factors is known. However, the asymptotic theory for the distribution of the eigenvalues equal to two is not developed in Chen (2012) and therefore cannot be used for inference on factor space dimensions.

One final comment is in order. Our analysis characterizes mixed frequency panel data models as group factor models. This connection, which is key for identification and estimation, actually eliminates one topic of concern in group factor models - the so called classification issue - we do not deal with. Namely, in a mixed data sampling setting we know a priori to which group observations in the panel belong. Some papers in group factor models (see e.g. Chen (2010, 2012), Ando and Bai (2016), and the references therein) assume instead that the researcher has first to figure out the allocation of observations across the different groups.

3 Estimation and inference on the number of common factors

In Section 3.1 we provide estimators of the common and group-specific factors, based on canonical correlations and canonical directions, when the true number of group-specific and common factors are known. In Section 3.2 we propose a sequential testing procedure for determining the number of common factors when only the dimensions k_1 and k_2 are known. The test statistic is based on the

add a shrinkage penalty in the LS criterion and prove asymptotic normality for the estimator of the observable regressors coefficients.

canonical correlations between the estimated factors in each subgroup of observables. In Section 3.3 we explain why the asymptotic results concerning the test statistic and the factors estimators obtained under the assumption that the number of pervasive factors k_1 and k_2 in each group is known, remain unchanged when the number of pervasive factors is consistently estimated. Finally, in Section 3.4 we use these results to define estimators and test statistics for the mixed frequency factor model. Readers who are only interested in the empirical application can go directly to Section 6 which starts with a summary of the novel econometric procedure.

3.1 Estimation of common and group-specific factors when the number of common and group-specific factors is known

Let us assume that the true number of factors $k_j > 0$ in each subgroup $j = 1, 2$, is known, and also that the true number of common factors $k^c > 0$, is known. Proposition 2 suggests the following estimation procedure for the common factors. Let $h_{1,t}$ and $h_{2,t}$ be estimated (up to a rotation) by extracting the first k_j Principal Components (PCs) from each subpanel j , and denote by $\hat{h}_{j,t}$ these PC estimates of the factors, $j = 1, 2$. Let $\hat{H}_j = [\hat{h}_{j,1}, \dots, \hat{h}_{j,T}]'$ be the (T, k_j) matrix of estimated PCs extracted from panel $Y_j = [y_{j,1}, \dots, y_{j,T}]'$ associated with the largest k_j eigenvalues of matrix $\frac{1}{N_j T} Y_j Y_j'$, $j = 1, 2$. Let $\hat{V}_{j\ell}$ denote the empirical covariance matrix of the estimated vectors $\hat{h}_{j,t}$ and $\hat{h}_{\ell,t}$, with $j, \ell = 1, 2$:

$$\hat{V}_{j\ell} = \frac{\hat{H}_j' \hat{H}_\ell}{T} = \frac{1}{T} \sum_{t=1}^T \hat{h}_{j,t} \hat{h}_{\ell,t}', \quad j, \ell = 1, 2,$$

and let matrices \hat{R} and \hat{R}^* be defined as:

$$\hat{R} := \hat{V}_{11}^{-1} \hat{V}_{12} \hat{V}_{22}^{-1} \hat{V}_{21}, \quad \text{and} \quad \hat{R}^* := \hat{V}_{22}^{-1} \hat{V}_{21} \hat{V}_{11}^{-1} \hat{V}_{12}. \quad (3.1)$$

Matrices \hat{R} and \hat{R}^* have the same non-zero eigenvalues. From Anderson (2003) and Magnus and Neudecker (2007), we know that the largest k^c eigenvalues of \hat{R} (resp. \hat{R}^*), denoted by $\hat{\rho}_\ell^2$, $\ell = 1, \dots, k^c$, are the first k^c squared sample canonical correlation between $\hat{h}_{1,t}$ and $\hat{h}_{2,t}$. We also know that the associated k^c canonical directions, collected in the (k_1, k^c) (resp. (k_2, k^c)) matrix \hat{W}_1 (resp. \hat{W}_2), are the eigenvectors associated with the largest k^c eigenvalues of matrix \hat{R} (resp. \hat{R}^*), normalized to have

length 1 w.r.t. matrix \hat{V}_{11} (resp. \hat{V}_{22}). It also holds:

$$\hat{W}_1' \hat{V}_{11} \hat{W}_1 = I_{k^c}, \text{ and } \hat{W}_2' \hat{V}_{22} \hat{W}_2 = I_{k^c}. \quad (3.2)$$

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

From equation (3.2) we have: $\frac{1}{T} \sum_{t=1}^T \hat{f}_t^c \hat{f}_t^{c'} = I_{k^c}$, and similarly for \hat{f}_t^{c*} , i.e. the estimated common factor values match in sample the normalization condition of identity variance-covariance matrix in (2.5). Let matrix \hat{W}_1^s (resp. \hat{W}_2^s) be the (k_1, k_1^s) (resp. (k_2, k_2^s)) matrix collecting k_1^s (resp. k_2^s) eigenvectors associated with the k_1^s (resp. k_2^s) smallest eigenvalues of matrix \hat{R} (resp. \hat{R}^*), normalized to have length 1 w.r.t. matrix \hat{V}_{11} (resp. \hat{V}_{22}). It also holds:

$$\hat{W}_1^s' \hat{V}_{11} \hat{W}_1^s = I_{k_1^s}, \text{ and } \hat{W}_2^s' \hat{V}_{22} \hat{W}_2^s = I_{k_2^s}.$$

The estimators of the group-specific factors can be defined analogously to the estimators of the common factors: $\check{f}_{1,t}^s = \hat{W}_1^s' \hat{h}_{1,t}$ and $\check{f}_{2,t}^s = \hat{W}_2^s' \hat{h}_{2,t}$. By construction, \hat{f}_t^c and $\check{f}_{1,t}^s$ (resp. \hat{f}_t^{c*} and $\check{f}_{2,t}^s$) are orthogonal in sample.

An alternative 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.¹¹ Let $\hat{F}^c = [\hat{f}_1^{c'}, \dots, \hat{f}_T^{c'}]'$ be the (T, k^c) matrix of estimated common factors, and $\hat{\Lambda}_j^c = [\hat{\lambda}_{j,1}^c, \dots, \hat{\lambda}_{j,N_j}^c]'$ the (N_j, k^c) matrix collecting the loadings estimators:

$$\hat{\Lambda}_j^c = Y_j' \hat{F}^c (\hat{F}^{c'} \hat{F}^c)^{-1} = \frac{1}{T} Y_j' \hat{F}^c, \quad j = 1, 2. \quad (3.3)$$

Let $\xi_{j,i,t} = y_{j,i,t} - \hat{\lambda}_{j,i}^c \hat{f}_t^c$ be the residuals of the regression of $y_{j,t}$ on the estimated common factor \hat{f}_t^c , and define $\xi_{j,t} = [\xi_{j,1,t}, \dots, \xi_{j,N_j,t}]'$, for $j = 1, 2$. Let $\Xi_j = [\xi_{j,1}, \dots, \xi_{j,T}]'$ be the (T, N_j) matrix of the regression residuals, for $j = 1, 2$.

DEFINITION 2. *An estimator of the specific factor vector is $\hat{f}_{1,t}^s$ (resp. $\hat{f}_{2,t}^s$), defined as the first k_1^s (resp. k_2^s) Principal Components of subpanel Ξ_1 (resp. Ξ_2).*

We denote by $\hat{F}_j^s = [\hat{f}_{j,1}^s, \dots, \hat{f}_{j,T}^s]'$ the (T, k_j^s) matrix of estimated group-specific factors, corresponding

¹¹This alternative estimation method for the group-specific factors corresponds to the method proposed by Chen (2012).

to the PCs extracted from panel Ξ_j associated with the largest k_j^s eigenvalues of matrix $\frac{1}{N_j T} \Xi_j \Xi_j'$, normalized to have $\hat{F}_j^{s'} \hat{F}_j^s / T = I_{k_j^s}$ for $j = 1, 2$. Then, \hat{f}_t^c is orthogonal in sample both to $\hat{f}_{t,1}^s$ and to $\hat{f}_{t,2}^s$. The orthogonality of both group-specific factor estimates $\hat{f}_{j,t}^s$, $j = 1, 2$, with the common factor estimate explains why we focus of the estimation procedure in Definition 2 compared to $\check{f}_{j,t}^s$, $j = 1, 2$. Moreover, we define $\hat{\Lambda}_j^s = [\hat{\lambda}_{j,1}^s, \dots, \hat{\lambda}_{j,N_j}^s]'$ as the (N_j, k_j^s) matrix collecting the loadings estimators:

$$\hat{\Lambda}_j^s = Y_j' \hat{F}_j^s (\hat{F}_j^{s'} \hat{F}_j^s)^{-1} = \frac{1}{T} \Xi_j' \hat{F}_j^s, \quad j = 1, 2, \quad (3.4)$$

where the second equality follows from the in-sample orthogonality of \hat{F}^c and \hat{F}_j^s , and the normalization of \hat{F}_j^s for $j = 1, 2$.

3.2 Inference on the number of common factors based on canonical correlations

We first consider the case where the number of factors k_1 and k_2 in each subpanel is assumed to be known, and hence $\underline{k} = \min(k_1, k_2)$ is also known, and we consider the problem of inferring the dimension k^c of the common factor space. In the next section we relax this assumption. From Proposition 2, dimension k^c is the number of unit canonical correlations between $h_{1,t}$ and $h_{2,t}$. We consider the following set of hypotheses:

$$\begin{aligned} H(0) &= \{1 > \rho_1 \geq \dots \geq \rho_{\underline{k}}\}, \\ H(1) &= \{\rho_1 = 1 > \rho_2 \geq \dots \geq \rho_{\underline{k}}\} \\ &\dots \\ H(k^c) &= \{\rho_1 = \dots = \rho_{k^c} = 1 > \rho_{k^c+1} \geq \dots \geq \rho_{\underline{k}}\}, \\ &\dots \\ H(\underline{k}) &= \{\rho_1 = \dots = \rho_{\underline{k}} = 1\}, \end{aligned}$$

where $\rho_1, \dots, \rho_{\underline{k}}$ are the canonical correlations of $h_{1,t}$ and $h_{2,t}$. Hypothesis $H(0)$ corresponds to the case of no common factor in the two groups of observables Y_1 and Y_2 . 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. The

largest possible number of common factors is the minimum between k_1 and k_2 , i.e. \underline{k} , and corresponds to hypothesis $H(\underline{k})$. In order to select the number of common factors, let us consider the following sequence of tests:

$$H_0 = H(k^c) \quad \text{against} \quad H_1 = \bigcup_{0 \leq r < k^c} H(r),$$

for each $k^c = \underline{k}, \underline{k} - 1, \dots, 1$. We propose the following statistic to test H_0 against H_1 , for any given $k^c = \underline{k}, \underline{k} - 1, \dots, 1$:

$$\hat{\xi}(k^c) = \sum_{\ell=1}^{k^c} \hat{\rho}_\ell. \quad (3.5)$$

The statistic $\hat{\xi}(k^c)$ corresponds to the sum of the k^c largest sample canonical correlations. We reject the null $H_0 = H(k^c)$ when $\hat{\xi}(k^c) - k^c$ is negative and large. The critical value is deduced by the large sample distribution of the statistic provided in Section 4. The number of common factors is estimated by performing sequentially the tests starting from $k^c = \underline{k}$.

3.3 Estimation and inference when k_1 and k_2 are unknown

The tests defined in Section 3.2 require the knowledge of the true number of pervasive factors $k_j > 0$ in each subgroup, $j = 1, 2$. When the true number of pervasive factors is not known, but consistent estimators \hat{k}_1 and \hat{k}_2 , say, are available, the asymptotic distributions and rates of convergence for the test statistic $\hat{\xi}(k^c)$ based on \hat{k}_1 and \hat{k}_2 are the same as those of the test based on the true number of factors. Intuitively, this holds because the consistency of estimators \hat{k}_j , implies that $P(\hat{k}_j = k_j) \rightarrow 1$ for $j = 1, 2$, which means that the error due to the estimation of the number of pervasive factors is asymptotically negligible.¹²

The estimators based on the penalized information criteria of Bai and Ng (2002) applied on the two subgroups, are examples of consistent estimators for the numbers of pervasive factors. Therefore, in the next Section 4, the asymptotic distributions and rates of convergence of the test statistic and factors estimators are derived assuming that the true numbers of factors $k_j > 0$ in each subgroup, $j = 1, 2$, are known.

¹²This argument is formalized using similar arguments as, for instance, in footnote 5 of Bai (2003).

3.4 Estimation and inference in the mixed frequency factor model

The estimators and test statistics defined in Sections 3.1 - 3.3 for the group factor model (2.4) allow to define estimators for the loadings matrices Λ_{HC} , Λ_H , Λ_{LC} , Λ_L , the aggregated factor values \bar{g}_t^U , $U = C, H, L$ and the test statistic for the common factor space dimension k^C in equation (2.3). We denote these estimators $\hat{\Lambda}_{HC}$, $\hat{\Lambda}_H$, $\hat{\Lambda}_{LC}$, $\hat{\Lambda}_L$, \hat{g}_t^U , and the test statistic $\hat{\xi}(k^C)$. Then, the estimators of the common and high frequency factor values are:

$$\begin{pmatrix} \hat{g}_{m,t}^C \\ \hat{g}_{m,t}^H \end{pmatrix} = \left(\hat{\Lambda}'_1 \hat{\Lambda}_1 \right)^{-1} \hat{\Lambda}'_1 x_{m,t}^H, \quad m = 1, \dots, M, \quad t = 1, \dots, T, \quad (3.6)$$

where $\hat{\Lambda}_1 = [\hat{\Lambda}_{HC} : \hat{\Lambda}_H]$.

4 Large sample theory

In this section we derive the large sample distributions of the estimators of factor spaces and factor loadings, and of the test statistic for the dimension of the common factor space. We also define a consistent selection procedure for the number of common factors. We consider the joint asymptotics $N_1, N_2, T \rightarrow \infty$ under Assumptions A.1-A.8 provided in Appendix, Sections A.1 and A.2. From the asymptotic theory of PCA estimators in large panels, estimates $\hat{h}_{j,t}$, for $j = 1, 2$, satisfy asymptotic expansions as in the next Assumption 1 (see e.g. Bai and Ng (2002), Stock and Watson (2002a), Bai (2003), Bai and Ng (2006a) for sufficient conditions).

ASSUMPTION 1. *As $N_1, N_2, T \rightarrow \infty$, we have:*

$$\hat{h}_{j,t} \simeq \hat{\mathcal{H}}_j \left(h_{j,t} + \frac{1}{\sqrt{N_j}} u_{j,t} + \frac{1}{T} b_{j,t} \right), \quad j = 1, 2, \quad (4.1)$$

up to negligible terms, where $b_{j,t}$ is a deterministic bias term, the matrix $\hat{\mathcal{H}}_j$ converges to a non-singular matrix, and:

$$u_{j,t} := \left(\frac{1}{N_j} \sum_{i=1}^{N_j} \lambda_{j,i} \lambda'_{j,i} \right)^{-1} \frac{1}{\sqrt{N_j}} \sum_{i=1}^{N_j} \lambda_{j,i} \varepsilon_{j,i,t}.$$

Note that the terms $u_{j,t}$ depend also from the cross-sectional dimension N_j , but for notational convenience, we omit the index N_j in $u_{j,t}$. From Assumptions A.2 and A.5 d), the error terms $u_{j,t}$ are asymptotically Gaussian as $N_j \rightarrow \infty$:

$$u_{j,t} \xrightarrow{d} N(0, \Sigma_{u,j}), \quad (4.2)$$

where the asymptotic variance is:

$$\Sigma_{u,j} = \Sigma_{\Lambda,j}^{-1} \Omega_j \Sigma_{\Lambda,j}^{-1},$$

and

$$\begin{aligned} \Sigma_{\Lambda,j} &= \lim_{N_j \rightarrow \infty} \frac{1}{N_j} \sum_{i=1}^{N_j} \lambda_{j,i} \lambda'_{j,i}, \\ \Omega_j &= \lim_{N_j \rightarrow \infty} \frac{1}{N_j} \sum_{i=1}^{N_j} \sum_{\ell=1}^{N_j} \lambda_{j,i} \lambda'_{j,\ell} \text{Cov}(\varepsilon_{j,i,t}, \varepsilon_{j,\ell,t}), \quad j = 1, 2. \end{aligned}$$

Without loss of generality, let $N_2 \leq N_1$. We assume $\sqrt{N_1}/T = o(1)$ (as stated in Assumption A.6), which allows to neglect the bias terms $b_{j,t}/T$ in the asymptotic expansion (4.1). We also assume $T/N_2 = o(1)$, which further simplifies the asymptotic distributions derived in the next sections.

The high-level Assumption 1 clarifies the conditions on the asymptotic expansions of factor estimators from different groups which are necessary for our asymptotic theory to hold. We remark that such asymptotic expansions arise not only in the linear factor structures estimable by PCA of this work, but also in the general class of large scale non-linear latent factor models, as the ones considered in Gagliardini and Gourieroux (2014). Therefore, Assumption 1 extends the applicability of the asymptotic theory developed in the next subsections to a very general class of large scale, linear and non-linear group-factor models for data potentially observed at different frequencies.

4.1 Asymptotic results for the group factor model

In this section we collect the main results concerning the asymptotic distributions of estimators and test statistics for the group factor model. Define the matrices:

$$\begin{aligned}\Omega_{j,k}(h) &= \lim_{N_j, N_k \rightarrow \infty} \frac{1}{\sqrt{N_j N_k}} \sum_{i=1}^{N_j} \sum_{\ell=1}^{N_k} \lambda_{j,i} \lambda'_{k,\ell} \text{Cov}(\varepsilon_{j,i,t}, \varepsilon_{k,\ell,t-h}), \\ \Sigma_{u,jk}(h) &= \Sigma_{\Lambda,j}^{-1} \Omega_{jk}(h) \Sigma_{\Lambda,k}^{-1},\end{aligned}$$

for $j, k = 1, 2$, and $h = \dots, -1, 0, 1, \dots$. Matrix $\Sigma_{u,jk}(h)$ is the asymptotic covariance between $u_{j,t}$ and $u_{k,t-h}$. Moreover, we have $\Omega_j \equiv \Omega_{j,j}(0)$ and $\Sigma_{u,j} \equiv \Sigma_{u,jj}(0)$, and similarly we define $\Sigma_{u,12} \equiv \Sigma_{u,12}(0) = \Sigma'_{u,21}$. Let us denote $N = \min\{N_1, N_2\} = N_2$ the minimal cross-sectional dimension among the two groups, and $\mu_N^2 = N_2/N_1 \leq 1$. Let $\mu_N \rightarrow \mu$, with $\mu \in [0, 1]$. The boundary value $\mu = 0$ accounts for the possibility that N_1 grows faster than N_2 .

THEOREM 1. *Under Assumptions A.1 - A.6, and the null hypothesis $H_0 = H(k^c)$ of k^c common factors, we have:*

$$N\sqrt{T} \left[\hat{\xi}(k^c) - k^c + \frac{1}{2N} \text{tr} \left\{ \tilde{\Sigma}_{cc}^{-1} \Sigma_{U,N} \right\} \right] \xrightarrow{d} N \left(0, \frac{1}{4} \Omega_U \right),$$

where

$$\begin{aligned}\tilde{\Sigma}_{cc} &= \frac{1}{T} \sum_{t=1}^T f_t^c f_t^{c'}, \\ \Omega_U &= 2 \sum_{h=-\infty}^{\infty} \text{tr} \{ \Sigma_U(h) \Sigma_U(h)' \}, \\ \Sigma_U(h) &= \mu^2 \Sigma_{u,11}^{(cc)}(h) + \Sigma_{u,22}^{(cc)}(h) - \mu \Sigma_{u,12}^{(cc)}(h) - \mu \Sigma_{u,21}^{(cc)}(h), \\ \Sigma_{U,N} &= \mu_N^2 \Sigma_{u,1}^{(cc)} + \Sigma_{u,2}^{(cc)} - \mu_N \Sigma_{u,12}^{(cc)} - \mu_N \Sigma_{u,21}^{(cc)},\end{aligned}$$

and the upper index (c, c) denotes the upper-left (k^c, k^c) block of a matrix.

Proof: See Appendix B.3.

The asymptotic distribution of $\hat{\xi}(k^c) - k^c$ after appropriate recentering and rescaling is Gaussian. The convergence rate is $N\sqrt{T}$. The asymptotic expansion of $\hat{\xi}(k^c) - k^c$ involves a time series average of

squared estimation errors on group factors.¹³ Since these estimation errors are of order $1/\sqrt{N}$, the expected value of their square will be of order $1/N$, originating a recentering term of the second order analogous to an error-in-variable bias adjustment. Moreover, the averaging over time of the recentered squared estimation errors allows to apply a root- T central limit theorem for weakly dependent processes, originating a total estimation uncertainty for the test statistic of order $1/(N\sqrt{T})$.

PROPOSITION 3. *Under Assumptions A.1 - A.6 we have:*

$$\sqrt{N_1}(\hat{\mathcal{H}}_c \hat{f}_t^c - f_t^c) \xrightarrow{d} N\left(0, \Sigma_{u,1}^{(cc)}\right), \quad (4.3)$$

$$\sqrt{N_2}(\hat{\mathcal{H}}_c^* \hat{f}_t^{c*} - f_t^c) \xrightarrow{d} N\left(0, \Sigma_{u,2}^{(cc)}\right), \quad (4.4)$$

$$\sqrt{N_j} \left[\hat{\mathcal{H}}_{s,j} \hat{f}_{j,t}^s - (f_{j,t}^s - (F_j^s)' F^c) (F^c' F^c)^{-1} f_t^c \right] \xrightarrow{d} N\left(0, (\Sigma_{\Lambda,j}^{(ss)})^{-1} \Omega_j^{(ss)} (\Sigma_{\Lambda,j}^{(ss)})^{-1}\right), \quad (4.5)$$

for any j, t , where $\hat{\mathcal{H}}_c, \hat{\mathcal{H}}_c^*$ and $\hat{\mathcal{H}}_{s,j}$ are non-singular matrices, $F^c = [f_1^c, \dots, f_T^c]'$, $F_j^s = [f_{j,1}^s, \dots, f_{j,T}^s]'$ and the upper index (ss) in the asymptotic variance of $\hat{f}_{j,t}^s$ denotes the lower-right (k_j^s, k_j^s) block of a matrix.

Proof: See Appendix B.4.

From Proposition 3 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. $1/\sqrt{N_2}$). The variance of the asymptotic Gaussian distribution is the upper-left (c, c) block of matrix $\Sigma_{u,1}$ (resp. $\Sigma_{u,2}$), 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 $1/\sqrt{NT}$, 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 $1/\sqrt{N_j}$.

Let us now derive the asymptotic distribution of the factor loadings estimators.¹⁴ Define the matri-

¹³See Appendix B, Section B.3.4, for the asymptotic expansion.

¹⁴We assume that \hat{f}_t^c is used for the estimation of the factor loadings as in Definition 2. The distribution of the loadings estimators is analogous when using \hat{f}_t^{c*} as common factor estimator.

ces:

$$\begin{aligned}\Phi_{j,i}^c &= \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \sum_{r=1}^T E[f_t^c f_r^{c'}] \text{cov}(\varepsilon_{j,i,t}, \varepsilon_{j,i,r}), \\ \Phi_{j,i}^s &= \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \sum_{r=1}^T E[f_{j,t}^s f_{j,r}^{s'}] \text{cov}(\varepsilon_{j,i,t}, \varepsilon_{j,i,r}), \\ \Psi_j &= \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \sum_{r=1}^T E[f_{j,t}^s f_{j,r}^{s'} \otimes f_t^c f_r^{c'}].\end{aligned}$$

PROPOSITION 4. *Under Assumptions A.1 - A.6 we have:*

$$\sqrt{T} \left[\left(\hat{\mathcal{H}}_c' \right)^{-1} \hat{\lambda}_{j,i}^c - \lambda_{j,i}^c \right] \xrightarrow{d} N \left(0, \Phi_{j,i}^c + (\lambda_{j,i}^{s'} \otimes I_{k^c}) \Psi_j (\lambda_{j,i}^s \otimes I_{k^c}) \right), \quad (4.6)$$

$$\sqrt{T} \left[\left(\hat{\mathcal{H}}_{s,j}' \right)^{-1} \hat{\lambda}_{j,i}^s - \lambda_{j,i}^s \right] \xrightarrow{d} N \left(0, \Phi_{j,i}^s \right), \quad (4.7)$$

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 3.

Proof: See Appendix B.4.

The factor loadings are estimated at rate \sqrt{T} . Matrix $\Phi_{j,i}^c$ 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.

To get a feasible distributional result for the statistic $\hat{\xi}(k^c)$, we need consistent estimators for the unknown scalar $\text{tr} \left\{ \tilde{\Sigma}_{cc}^{-1} \Sigma_{U,N} \right\}$ and matrix Ω_U in Theorem 1. To simplify the analysis, we assume at this stage that the errors $\varepsilon_{j,i,t}$ are uncorrelated across subpanels j , individuals i and dates t (Assumption A.7).¹⁵ Then, we have:

$$\Sigma_{U,N} = \mu_N^2 \Sigma_{u,1}^{(cc)} + \Sigma_{u,2}^{(cc)}, \quad \Sigma_U(0) = \mu^2 \Sigma_{u,1}^{(cc)} + \Sigma_{u,2}^{(cc)}, \quad \Omega_U = 2 \text{tr} \left\{ \Sigma_U(0)^2 \right\}.$$

In Theorem 2 below, we replace $\tilde{\Sigma}_{cc}$ with its large sample limit I_{k^c} , and $\Sigma_{U,N}$ and $\Sigma_U(0)$ by consistent estimators. We show that the estimation error for $\text{tr}(\tilde{\Sigma}_{cc}^{-1} \Sigma_{U,N})$ in the bias adjustment is $o_p(1/\sqrt{T})$, and therefore, the asymptotic distribution of the statistic is unchanged.

¹⁵If the errors are weakly correlated across series and/or time, consistent estimation of $\Sigma_{U,N}$ and Ω_U requires thresholding of estimated cross-sectional covariances and/or HAC-type estimators.

THEOREM 2. Let $\hat{\Sigma}_U = (N_2/N_1)\hat{\Sigma}_{u,1}^{(cc)} + \hat{\Sigma}_{u,2}^{(cc)}$, with

$$\hat{\Sigma}_{u,j} = \left(\frac{\hat{\Lambda}'_j \hat{\Lambda}_j}{N_j} \right)^{-1} \left(\frac{1}{N_j} \hat{\Lambda}'_j \hat{\Gamma}_j \hat{\Lambda}_j \right) \left(\frac{\hat{\Lambda}'_j \hat{\Lambda}_j}{N_j} \right)^{-1}, \quad j = 1, 2, \quad (4.8)$$

where $\hat{\Lambda}_j = [\hat{\Lambda}_j^c; \hat{\Lambda}_j^s]$, estimates $\hat{\Lambda}_j^c$ and $\hat{\Lambda}_j^s$ are the loadings estimators defined in equations (3.3) and (3.4), $\hat{\Gamma}_j = \text{diag}(\hat{\gamma}_{j,ii}, i = 1, \dots, N_j)$ with

$$\hat{\gamma}_{j,ii} = \frac{1}{T} \sum_{t=1}^T \hat{\varepsilon}_{j,i,t}^2$$

and $\hat{\varepsilon}_{j,i,t} = y_{j,i,t} - \hat{\lambda}_{j,i}^{c'} f_t^c - \hat{\lambda}_{j,i}^{s'} f_{j,t}^s$, for $j = 1, 2$. Define the test statistic:

$$\tilde{\xi}(k^c) := N\sqrt{T} \left(\frac{1}{2} \text{tr} \{ \hat{\Sigma}_U^2 \} \right)^{-1/2} \left[\hat{\xi}(k^c) - k^c + \frac{1}{2N} \text{tr} \{ \hat{\Sigma}_U \} \right], \quad (4.9)$$

and let Assumptions A.1 - A.7 hold. Then:

- (i) Under the null hypothesis $H_0 = H(k^c)$ of k^c common factors, we have: $\tilde{\xi}(k^c) \xrightarrow{d} N(0, 1)$.
- (ii) Under the alternative hypothesis $H_1 = \bigcup_{0 \leq r < k^c} H(r)$, we have: $\tilde{\xi}(k^c) \xrightarrow{p} -\infty$.

Proof: See Appendix B.5.

The feasible asymptotic distribution in Theorem 2 (i) is the basis for a one-sided test of the null hypothesis of k^c common factors. If $\tilde{\xi}(k^c) < -1.64$, this null hypothesis is rejected at 5% level against the alternative hypothesis of less than k^c common factors. From Theorem 2 (ii), the test is consistent under the alternative.

One way to implement the model selection procedure to estimate the number of common factors k^c proposed in Section 3.2 consists in testing sequentially the null hypothesis $H_0 = H(r)$, against the alternative $H_1 = \bigcup_{0 \leq \ell < r} H(\ell)$, using the test statistic $\tilde{\xi}(r)$ defined as in Theorem 2 for any generic number r of common factors in the model. More specifically, a “naive” procedure can be defined as follows. The procedure is initiated with $r = \underline{k}$, proceeds backwards and is stopped at the largest integer $\hat{k}_{naive}^c = r$ such that the null $H(r)$ cannot be rejected, i.e. $\tilde{\xi}(k^c) > z_\alpha$, where z_α is the α -quantile of the standard Gaussian distribution. Otherwise, set $\hat{k}_{naive}^c = 0$ if the test rejects the null $H(r)$ for all $r = \underline{k}, \dots, 1$. This “naive” procedure is not a consistent estimator of the number of common factors.

Indeed, there exists asymptotically a non-zero probability α of underestimating k^c coming from the type I error of the test of $H(k_0^c)$ against $\bigcup_{0 \leq \ell < k_0^c} H(\ell)$, when the true number of common factors is $k_0^c > 0$.

Building on the results in Pötscher (1983), and on those for rank testing of Cragg and Donald (1997), and Robin and Smith (2000), a consistent estimator of the number of common factors k_0^c , for any integer $k_0^c \geq 0$, is obtained allowing the asymptotic size α to go to zero as $N, T \rightarrow \infty$. The following Proposition 5 defines a consistent inference procedure for the number of common factors.

PROPOSITION 5. *Let $\alpha_{N,T}$ be a sequence of real scalars defined in the interval $(0, 1)$ for any N, T , such that (i) $\alpha_{N,T} \rightarrow 0$ and (ii) $(N\sqrt{T})^{-1}z_{\alpha_{N,T}} \rightarrow 0$ for $N, T \rightarrow \infty$. Then, under Assumptions A.1 - A.7 the estimator of the number of common factors defined as:*

$$\hat{k}^c = \max \left\{ r : 1 \leq r \leq \underline{k}, \tilde{\xi}(r) \geq z_{\alpha_{N,T}} \right\},$$

and $\hat{k}^c = 0$, if $\tilde{\xi}(r) < z_{\alpha_{N,T}}$ for all $r = 1, \dots, \underline{k}$, is consistent, i.e. $P(\hat{k}^c = k_0^c) \rightarrow 1$ under $H(k_0^c)$, for any integer $k_0^c \in [0, \underline{k}]$.

Proof: See Appendix B.6.

Condition (i) ensures asymptotically zero probability of type I error when testing $H(k_0^c)$ against $\bigcup_{0 \leq \ell < k_0^c} H(\ell)$. Condition (ii) is a lower bound on the convergence rate to zero of the asymptotic size, and is used to keep asymptotically zero probability of type II error of each step of the procedure. The conditions in Proposition 5 are satisfied e.g. for

$$z_{N,T} \equiv z_{\alpha_{N,T}} = -c(N\sqrt{T})^\gamma, \quad (4.10)$$

with constants $c > 0$ and $0 < \gamma < 1$.

4.2 Asymptotic results for the mixed frequency factor model

In this section we give the asymptotic distribution for the estimators of factor values in the mixed frequency factor model. The asymptotics is for $N_H, N_L, T \rightarrow \infty$, such that $N_L \leq N_H$, $\sqrt{N_H}/T =$

$o(1), T/N_L = o(1)$. Define the matrices:

$$\Omega_{\Lambda,m}^* = \lim_{N_H \rightarrow \infty} \frac{1}{N_H} \sum_{i=1}^{N_H} \sum_{\ell=1}^{N_H} \lambda_{1,i} \lambda'_{1,\ell} Cov(e_{m,t}^{i,H}, e_{m,t}^{\ell,H}), \quad m = 1, \dots, M,$$

where $\lambda'_{1,i}$ is the i -th row of the $(N_H, k^C + k^H)$ matrix $\Lambda_1 = [\Lambda_{HC} : \Lambda_H]$.

PROPOSITION 6. *Under Assumptions A.1 - A.8 we have:*

$$\begin{aligned} \sqrt{N_H}(\hat{\mathcal{H}}_c \hat{g}_{m,t}^C - g_{m,t}^C) &\xrightarrow{d} N(0, [\Sigma_{\Lambda,1}^{-1} \Omega_{\Lambda,m}^* \Sigma_{\Lambda,1}^{-1}]^{(CC)}), \\ \sqrt{N_H} \left[\hat{\mathcal{H}}_{1,s} \hat{g}_{m,t}^H - (g_{m,t}^H - (\bar{g}^{H'} \bar{g}^C)(\bar{g}^{C'} \bar{g}^C)^{-1}) g_{m,t}^C \right] &\xrightarrow{d} N(0, [\Sigma_{\Lambda,1}^{-1} \Omega_{\Lambda,m}^* \Sigma_{\Lambda,1}^{-1}]^{(HH)}), \end{aligned}$$

for any m, t , where $\hat{\mathcal{H}}_c$ and $\hat{\mathcal{H}}_{1,s}$ are the same non-singular matrices of Proposition 3, $\bar{g}^C = [\bar{g}_1^C, \dots, \bar{g}_T^C]'$, $\bar{g}^H = [\bar{g}_1^H, \dots, \bar{g}_T^H]'$, $\Sigma_{\Lambda,1} = \lim_{N_H \rightarrow \infty} \frac{1}{N_H} \sum_{i=1}^{N_H} \lambda_{1,i} \lambda'_{1,i}$, and indices (CC) and (HH) denote the upper-left (k^C, k^C) block and the lower-right (k^H, k^H) block of a matrix, respectively.

Proof: See Appendix B.7.

From Proposition 6, a linear transformation of vector $\hat{g}_{m,t}^C$, resp. $\hat{g}_{m,t}^H$, estimates the common factor $g_{m,t}^C$, resp. the residual of the sample projection of the high-frequency factor on the common factor. The estimation rate is $\sqrt{N_H}$. There is no asymptotic effect from the error-in-variable problem induced by using estimated factor loadings in the cross-sectional regression when $T/N_H = o(1)$. The asymptotic distribution of the estimator \hat{g}_t^L of the aggregated low-frequency factor is deduced from Proposition 3.

4.3 Theoretical extensions beyond group factor models

The theoretical contributions, in particular Theorems 1 and 2, are of interest beyond (mixed frequency) group factor models. It is the purpose of this subsection to discuss several cases where our theory readily applies.

Inference regarding the rank of an unknown, real-valued matrix is an important and well-studied problem.¹⁶ For indefinite matrix estimators there is a well-developed framework, see Donald, Fortuna, and Pipiras (2007). The case of semi-definite matrix estimators still poses many challenges, however,

¹⁶See for instance Gill and Lewbel (1992), Cragg and Donald (1996), Robin and Smith (2000), Kleibergen and Paap (2006), and the review article by Camba-Mendez and Kapetanios (2009).

as discussed by Bai and Ng (2007) and more recently in Donald, Fortuna, and Pipiras (2014) who argue that the tests suggested in the literature are not suitable. In fact, when the rank of a generic (positive) semi-definite matrix, say M , needs to be estimated using a semi-definite estimator, say \hat{M} , the asymptotic variance-covariance matrix of this estimator - denoted as W_0 - is necessarily singular, as shown in Proposition 2.1 of Donald, Fortuna, and Pipiras (2007). Therefore standard rank tests cannot be applied as they assume matrix W_0 to be full rank.

Intimately related to testing the rank of a semi-definite matrix is testing the dimension of a common factor space. This problem appears in many applications. For example what is the common space between on the one hand the principal components extracted from stock returns (akin to rotations of Fama and French factors), call them $h_{1,t}$, and on the other hand $h_{2,t}$ which are the principal components extracted from macroeconomic series (like Stock and Watson factors)? Surprisingly, there are no formal tests available.

The connection is as follows. If we consider the variance-covariance matrix of the stacked factors, i.e. $V[(h'_{1,t}, h'_{2,t})']$, then testing for common factors relates to testing the rank of a symmetric semi-definite matrix and this is where our theory makes a novel contribution. The key ingredient is the link between the eigenvalues and eigenvectors of the variance-covariance matrix of stacked PCs extracted from the groups, and the distributional theory developed in Section 4. We provide in the OA, Section C.2, a throughout discussion on the link between the unit canonical correlations of $h_{1,t}$, $h_{2,t}$ and the eigenvalues equal to two, or equivalently the eigenvalues equal to zero, of the variance-covariance matrix $V[(h'_{1,t}, h'_{2,t})']$ of the stacked PCs, when these are normalized such that $V(h_{j,t}) = I_{k_j}$, for $j = 1, 2$.¹⁷ Therefore, using our theory developed in Theorems 1 and 2, it is straightforward to derive the asymptotic distribution for the number of eigenvalues equal to two of the matrix $V[(h'_{1,t}, h'_{2,t})']$. Analogous arguments, could be used to derive the asymptotic properties of the criteria for the selection of the number of common and group-specific factors in Goyal, Pérignon, and Villa (2008), as well.

Our results in Section 4.1 provide the guidance for the construction of the asymptotic distribution of the (sum of the) eigenvalues of a semi-definite matrix, and develop a sequential testing procedure for determining the rank of the matrix itself. This test, for example, would enable us to determine the

¹⁷Specifically, in the OA, Section C.2, we provide Corollary 2, which is analogous to Proposition 3.1 in Chen (2012). Our new proof, alternative to the one of Chen (2012), provides the explicit link between the number of canonical correlations equal to one, and the number of eigenvalues equal to two of matrix $V[(h'_{1,t}, h'_{2,t})']$. This number is also equal to the number of zero eigenvalues of that matrix.

number of latent dynamic factors in large panels of data, without having to estimate them, a problem tackled by Bai and Ng (2007). In their paper, first a number - say r - of static factors should be estimated by PCA from a large panel. Different from their methodology, and also different from the solution proposed by Amengual and Watson (2007), we can directly test the rank - say $q \leq r$ - of the residual covariance (or correlation matrix) of a VAR model estimated on the factors themselves.

Other potential applications for our test for the number of canonical correlations equal to one, consist of testing if some estimated factors are spanned by a small number of observed factors (observed at the same, or different frequencies), a problem also considered by Bai and Ng (2006b). Such a test is a special case of ours, in which one group of factors (the observable ones) are estimated with infinite precision. In our notation, this means that for either $j=1$, or $j=2$, we have that $\hat{h}_{j,t} \approx \hat{\mathcal{H}}_j h_{j,t}$ holds as an equality with $\hat{\mathcal{H}}_j = I_{k_j}$, and not as an (asymptotic) approximation as in (4.1), and the variance-covariance matrix of the observational errors is $\Sigma_{u,j} = 0$. It is important to note that the asymptotic distributions of Section 3.2 in Bai and Ng (2006b), for the empirical canonical correlations among a group of factors estimated from a large panel, and a group of exact ones, are valid for all values of the true canonical correlations different from zero and one. The same holds for the asymptotic distribution of the estimated canonical correlations among two groups of factors observed without measurement error, considered in the classical statistical literature - see Anderson (2003) - and used by Bai and Ng (2006b) to derive their results, showing that the problem of testing for unit canonical correlations is a degenerate one, in this special, but important case. On the other hand our results allow one to exploit the observational error of the factors to derive an asymptotically normal distribution for the estimator of a true canonical correlation equal to one. Therefore, unlike Bai and Ng (2006b), we are able to test for the number of canonical correlations equal to one. Moreover, Bai and Ng (2006b) note that the squared canonical correlation between a single series measured without error, and a group of estimated factors must be equal to the R^2 of the regression of the observed factor on the estimated ones. Therefore our test could also be used to derive the asymptotic distribution of R^2 , when this is equal to one under the null hypothesis.

In related recent work Pelger (2015) also deals with the question whether a set of statistical factors coincides with a set of economic candidate factors. Hence, one simply tests whether the factor spaces are equal. To circumvent the use of canonical correlations he proposes to look at the sum of squared canonical correlations. If two sets of factors coincide then the sum of squared canonical correlations

should be equal to the dimension of the factor set. As the sum of squared canonical correlations is simply the trace of a specific product of (realized) covariance matrices (see p. 26 of Pelger (2015)), it is straightforward to derive the asymptotic distribution under very general conditions. Based on the distribution result he develops a new test to determine if a set of estimated statistical factors can be written as a linear combination of observed economic variables. As in the previous examples, our theoretical results can address the problem posed by Pelger (2015) more directly and straightforwardly. Indeed, from an empirical point of view, our testing methodology can also be used to test for the presence of common factors among large panels of high-frequency financial variables, and low frequency observable macro economic variables, or large panels of (same frequency) financial and/or macroeconomic data for different regions or countries, extending the works of Cho (1984), and Kose, Otrok, and Whiteman (2008), among many others.

5 Monte Carlo simulations

In order to assess the small sample properties of our estimator for the number of common factors we conduct an extensive Monte Carlo simulation study, which can also evaluate the advantages of our testing procedure compared to the one proposed in Chen (2012). The Online Appendix, Section D, includes the detailed description of the simulation design and the tables of results. The Data Generating Process (DGP) considered in each design corresponds to the high frequency model (2.1) with flow sampling for the LF variables, and possibly autocorrelated common and specific factors, and possibly cross-sectionally correlated idiosyncratic innovations. Following our empirical application, we fix the number of high frequency subperiods $M = 4$. We consider different numbers of common and specific factors across the DGPs, given by $k^C = 0, 1, 2$, and $k^H = k^L = 1, 2, 5, 6$. The model for the factors is defined by stacking the latent factor vectors $g_{m,t}^C$, $g_{m,t}^H$, and $g_{m,t}^L$ into the new vector $g_{m,t} = [g_{m,t}^C, g_{m,t}^H, g_{m,t}^L]'$, and is characterized by the following dynamic:

$$g_{m,t} = a_F g_{m-1,t} + \eta_{m,t},$$

where a_F is the common scalar AR(1) coefficient for the $k^C + k^H + k^L$ factors. The innovations $\eta_{m,t}$ are such that:

$$\eta_{m,t} \sim i.i.N(0, \Sigma_\eta), \quad \Sigma_\eta = (1 - a_F^2) \begin{bmatrix} I_{k^C} & 0 & 0 \\ 0 & I_{k^H} & \Phi_{sim} \\ 0 & \Phi'_{sim} & I_{k^L} \end{bmatrix},$$

where $\Phi_{sim} = \phi_{sim} I_{k^H}$. The scalar ϕ_{sim} generates correlation between pairs of HF and LF factors. We consider different values of a_F and ϕ_{sim} in the simulation designs. The factor loadings are simulated in a setup such that the distribution of $R^{2'}$ s of the regressions of data on factors in the Monte Carlo design mimics the one in the empirical applications.¹⁸ We ran 2000 Monte Carlo simulations for each DGP, and consider cross-sectional and time series dimensions N_H, N_L, T as small as the ones in our empirical applications, and progressively increase their sizes. In order to estimate the number of common factors we consider *i)* our consistent sequential testing procedure defined in Proposition 5, and *ii)* the selection procedure based on the penalized information criterion of Theorem 3.7 in Chen (2012). The critical value for our selection procedure is as in equation (4.10), with $\gamma = 0.1$, and $c = 0.95$ such that $z_{NT} = -1.64 \sim z_{0.05}$ for $N_1 = N_2 = 40$, and $T = 35$, which are analogous to the cross-sectional and time series dimensions in our empirical application. The properties of the estimators are evaluated comparing *a)* the percentage of estimates \hat{k}^C that are less than, equal to, and greater than k^C , and *b)* the average estimated number of common, high-frequency specific, and low-frequency specific factors. In order to have comparable results for the different estimators, we assume that the true numbers of pervasive factors $k^C + k^H$ and $k^C + k^L$ in the two panels are known, and only k^C needs to be estimated, for both selection procedures.

In the special case of a small number, say 1 or 2, of uncorrelated specific factors, the penalized information criterion proposed in Chen (2012) yields the correct number of factors in almost all Monte Carlo simulations for any sample size, confirming the results in Chen (2012). For a small number of specific factors and small sample size, our selection procedure is less accurate, with frequencies of detection of the correct number of factors ranging from 80% to 100%, depending on the design. Yet, the frequency of correct estimation of k^C for our selection procedure increases monotonically with the sample size, approaching 100%. This last result holds for all simulation designs considered,

¹⁸See OA, Section D.1.3, for a thorough description of the simulation design setup for the loadings matrices.

confirming the ability of the asymptotic distribution in approximating the finite sample distribution of the test statistic for relatively small sample sizes.

Importantly, as the correlation ϕ_{sim} among the specific factors increases, ranging from 0 to 0.7 and 0.9 (and in some designs to 0.95 and 0.99), the frequencies of correct detection of the true number of factors by the procedure of Chen's (2012) decreases very quickly toward zero, for all sample sizes, including very large ones. This deterioration in the performance is much less dramatic for our selection procedure based on sequential tests, which clearly dominates the one based on the penalized criteria. As expected, we also observe a monotonic decrease in the precision of all the estimators when the number of specific factors gets relatively large, say 5 or 6. Also in these cases, the deterioration of the precision of our sequential procedure is much less dramatic than Chen's (2012) procedure, for all sample sizes, suggesting that our sequential testing estimator is clearly preferable in these more general cases. Interestingly, the sequential testing procedure exhibits also a more rapid improvement in performance as the sample size increases, compared to the penalized information criteria.

The above results are qualitatively similar *i)* when different values of the factor autocorrelation α_F are considered, namely 0 and 0.6, *ii)* for different levels of the cross-sectional correlation of the idiosyncratic errors of the observables, and *iii)* for different magnitudes of the pervasiveness of the factors, both common and specific, as measured by the theoretical R^2 s of the simulated observables on the factors themselves. We refer to the OA, Section D for full details.

6 Empirical application

Before turning to the empirical application with results on factor analysis related to US sectoral output growth, it is worth summarizing our methodology for the benefit of the readers who skipped the previous sections. This is done in a first subsection.

6.1 Practical implementation of the procedure

Let us first assume that k^C , k^H , k^L , i.e. the number of respectively common, high and low frequency factors in equation (2.1), are known and are all strictly larger than zero. The identification strategy presented in Section 2 directly implies a simple estimation procedure for the factor values and the factor loadings, which consists of the following three steps:

1. *PCA performed on the HF and LF panels separately*

Define the (T, N_H) matrix of temporally aggregated (in our application flow-sampled) HF observables as $X^H = [x_1^H, \dots, x_T^H]'$, and the (T, N_L) matrix of LF observables as $X^L = [x_1^L, \dots, x_T^L]'$. The estimated pervasive factors of the HF data, which are collected in $(T, k^C + k^H)$ matrix $\hat{h}_H = [\hat{h}_{H,1}, \dots, \hat{h}_{H,T}]'$, are obtained performing PCA on the HF data:

$$\left(\frac{1}{TN_H} X^H X^{H'} \right) \hat{h}_H = \hat{h}_H \hat{V}_H,$$

where \hat{V}_H is the diagonal matrix of the eigenvalues of $(TN_H)^{-1} X^H X^{H'}$. Analogously, the estimated pervasive factors of the LF data, which are collected in the $(T, k^C + k^L)$ matrix $\hat{h}_L = [\hat{h}_{L,1}, \dots, \hat{h}_{L,T}]'$, are obtained performing PCA on the LF data:

$$\left(\frac{1}{TN_L} X^L X^{L'} \right) \hat{h}_L = \hat{h}_L \hat{V}_L,$$

where \hat{V}_L is the diagonal matrix of the eigenvalues of $(TN_L)^{-1} X^L X^{L'}$.

2. *Canonical correlation analysis performed on estimated principal components*

Let \hat{W}_H^C be the $(k^C + k^H, k^C)$ matrix whose columns are the canonical directions for $\hat{h}_{H,t}$ associated with the k^C largest canonical correlations between \hat{h}_H and \hat{h}_L . Then, the estimator of the (in our application flow sampled) common factor is $\hat{g}_t^C = \hat{W}_H^C{}' \hat{h}_{H,t}$, for $t = 1, \dots, T$, and the estimated loadings matrices $\hat{\Lambda}_{HC}$ and $\hat{\Lambda}_{LC}$ are obtained from the least squares regressions of x_t^H and x_t^L on estimated factor \hat{g}_t^C . Collect the residuals of these regressions:

$$\begin{aligned} \hat{\xi}_t^H &= x_t^H - \hat{\Lambda}_{HC} \hat{g}_t^C, \\ \hat{\xi}_t^L &= x_t^L - \hat{\Lambda}_{LC} \hat{g}_t^C, \end{aligned}$$

in the following (T, N_U) , with $U = H, L$, matrices:

$$\hat{\Xi}^U = \left[\hat{\xi}_1^U, \dots, \hat{\xi}_T^U \right]', \quad U = H, L.$$

Then, the estimators of the HF and LF factors, collected in the (T, k^U) , $U = H, L$, matrices:

$$\hat{G}^U = \left[\hat{g}_1^U, \dots, \hat{g}_T^U \right]', \quad U = H, L,$$

are obtained extracting the first k^H and k^L PCs from the matrices of residuals:

$$\left(\frac{1}{TN_U} \hat{\Xi}^U \hat{\Xi}^{U'} \right) \hat{G}^U = \hat{G}^U \hat{V}_S^U, \quad U = H, L,$$

where \hat{V}_S^U , with $U = H, L$, are the diagonal matrices of the associated eigenvalues. Next, the estimated loadings matrices $\hat{\Lambda}_H$ and $\hat{\Lambda}_C$ are obtained from the least squares regression of $\hat{\xi}_t^H$ and $\hat{\xi}_t^L$ on respectively the estimated factors \hat{g}_t^H and \hat{g}_t^L .

3. *Reconstruction of the common and high frequency-specific factors*

The estimates of the common and HF factors for each HF subperiod, denoted by $\hat{g}_{m,t}^C$ and $\hat{g}_{m,t}^H$, for any $m = 1, \dots, M$ and $t = 1, \dots, T$, are obtained by cross-sectional regression of $x_{m,t}$ on the estimated loadings $[\hat{\Lambda}_{HC} : \hat{\Lambda}_H]$ obtained from the second step.

Since the factors dimensions are unknown, the aforementioned procedure is implemented with estimated factors dimensions \hat{k}^C , \hat{k}^H , and \hat{k}^L . Inference on the number of common, low and high-frequency specific factors proceeds as follows:

1. Estimate $k_{XH} = k^C + k^H$ and $k_{XL} = k^C + k^L$, i.e. the numbers of pervasive factors in panels X^H and X^L , by some consistent estimators, as the IC_{p1} and IC_{p2} criteria of Bai and Ng (2002).
2. Let $\underline{k} := \min(\hat{k}_{XH}, \hat{k}_{XL})$. Test sequentially:

$$H_0 = H(r) : k^C = r \quad \text{against} \quad H_1 : k^C < r,$$

for any given $r = \underline{k}, \underline{k} - 1, \dots, 1$. We use the statistic $\tilde{\xi}(r)$ defined in equation (4.9), which is based on $\hat{\xi}(r) = \sum_{\ell=1}^r \hat{\rho}_\ell$, where the $\hat{\rho}_\ell$, for $\ell = 1, \dots, r$, are the r largest canonical correlations between $\hat{h}_{H,t}$ and $\hat{h}_{L,t}$. Here, $\hat{h}_{H,t}$ and $\hat{h}_{L,t}$ are the first \hat{k}_{XH} and \hat{k}_{XL} PCs extracted from the X^H and X^L panels, respectively, and the canonical correlations are the squared roots of the eigenvalues of matrix \hat{R} defined in equation (3.1). We reject $H_0 = H(r)$ if $\hat{\xi}(r) < z_{NT}$, where critical value z_{NT} is set as in equation (4.10), with $\gamma = 0.1$ and constant $c = 0.95$ as in the Monte Carlo study. Estimate \hat{k}^C is the largest dimension r such that H_0 is not rejected, or $\hat{k}^C = 0$ if H_0 is rejected for all r .

3. The dimensions of frequency specific factors are obtained by difference: $\hat{k}^H = \hat{k}_{XH} - \hat{k}^C$, and $\hat{k}^L = \hat{k}_{XL} - \hat{k}^C$.

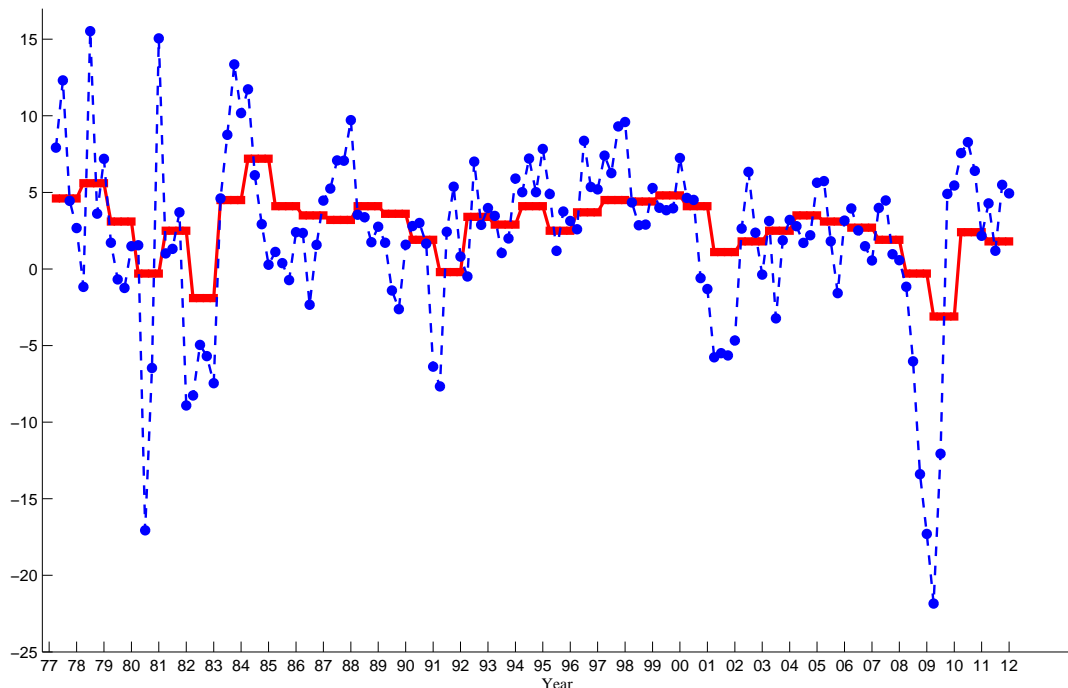
6.2 Data description

The data consists of a combination of Industrial Production (IP) and non-IP indices for the different sectors. For industrial production we use the same data on 117 IP sectoral growth rates indices considered by Foerster, Sarte, and Watson (2011), sampled at quarterly frequency from 1977.Q1 to 2011.Q4.¹⁹ These indices correspond to the finest level of disaggregation for the sectoral components of the IP aggregate index which can be matched with the available sectors in the *Input-Output* and *Capital Use* tables used in the structural analysis in Section 6.4. The data for all the remaining non-IP sectors consist of the annual growth rates of real GDP for the following 42 sectors: 35 services, Construction, Farms, Forestry-fishing and related activities, General government (federal), Government enterprises (federal), General government (state and local) and Government enterprises (state

¹⁹The IP data are available also at monthly frequency. Following Foerster, Sarte, and Watson (2011), we focus only on quarterly IP data, as they share the main feature of the monthly ones, but are less noisy.

and local). These LF data are also available from 1977 until 2011 and are published by the Bureau of Economic Analysis (BEA).²⁰ Moreover, as IP is a Gross Output (GO) measure, in the structural analysis it is convenient also to consider the yearly growth rates of real GO for the non-IP sectors. These data are available from 1988 until 2011, and are also published by the BEA. Following the sectoral productivity literature, in the structural analysis we focus exclusively on the private sectors, and therefore exclude four Government Gross Output indices, reducing the sample size to 38 non-IP sectors indices. All growth rates refer to seasonally adjusted real output indices, and are expressed in percentage points.²¹ Figure 2 displays the series of quarterly growth rates of the aggregate Industrial

Figure 2: Growth rates of the Industrial Production (IP) and Gross Domestic Product (GDP) indices



The dotted (blue) line corresponds to the quarterly growth rates of the aggregate IP index for sample period 1977.Q1-2011.Q4, while solid (red) line represents the annual growth rates of GDP for the entire US economy for the sample period 1977-2011.

Production and annual growth rates of Gross Domestic Product over the sample period from 1977 to 2011. The objective of our empirical application is to use our mixed frequency factor model to capture

²⁰GDP data are available at quarterly frequency for the aggregate index, but not for sectoral ones. As in the remaining part of the paper we study comovements among different sectors, we consider the panel of yearly GDP sectoral data.

²¹An exhaustive description of the dataset is provided in the OA, Section C.4.

the major sources of comovement among the sectoral constituents of these two indices, which are the most reliable measures of US economic activity. ²²

6.3 Factors common to all US sectors

We assume that our dataset follows the factor structure for flow sampling as in equation (2.2), with $x_{m,t}^H$ and x_t^L corresponding to quarterly IP and annual non-IP data, respectively. Let X^H be the (T, N_H) panel of the yearly observations of the IP indices growth rates (computed as the sum of the quarterly growth rates $x_{m,t}^H$, $m = 1, \dots, 4$, for year t), and let X^L be the (T, N_L) panel of the yearly growth rates of the non-IP indices as defined in Section 6.1. Let also $X_{HF} = [x_{1,1}^H, x_{2,1}^H, \dots, x_{m,t}^H, \dots, x_{4,T}^H]'$ be the $(4T, N_H)$ panel of IP indices quarterly growth rates.

We start by selecting the number of factors in each subpanel, which are of dimensions $k_{X^H} = k^C + k^H$ for X^H and X_{HF} and $k_{X^L} = k^C + k^L$ for X^L , respectively. We use the IC_{p2} information criteria of Bai and Ng (2002), and report the results in Table 1. Results for the IC_{p1} information criteria are reported in the OA, Section C.5. Table 1 corroborates the evidence in Foerster, Sarte,

Table 1: Estimated number of pervasive factors in HF and LF panels

	X_{HF}	X^H	X^L
IP data: 1977.Q1-2011.Q4. Non-IP data: Gross Domestic Product, 1977-2011			
IC_{p2}	1	2	1
IP data: 1988.Q1-2011.Q4. Non-IP data: Gross Output, 1988-2011			
IC_{p2}	1	1	2

The number of latent pervasive factors selected by the IC_{p2} information criteria is reported for different subpanels. Subpanels X_{HF} and X^H correspond to IP data sampled at quarterly and yearly frequency, respectively. Panels X^L correspond to non-IP data. We use $k_{max} = 15$ as maximum number of factors when computing IC_{p2} . In the first line the quarterly IP data are for sample period 1977.Q1-2011.Q4, the annual non-IP data are GDP growth rates for the sample period 1977-2011, while in the second line the quarterly IP data are for sample period 1988.Q1-2011.Q4, the annual non-IP data are Gross Output growth rates for the sample period 1988-2011.

and Watson (2011) suggesting that there is either one or perhaps two pervasive factors in the IP data.

²²In the OA, Section C.5.1, we replicate the analysis in Section II.B of Foerster, Sarte, and Watson (2011), in order to rule out the possibilities that *a*) sectoral weights in GDP and IP aggregate indexes are the major determinants in explaining the variability of the indexes themselves, and *b*) that their aggregate variability is driven mainly by sector-specific variability. As in Foerster, Sarte, and Watson (2011), our analysis confirms that covariance among different sectors is the main source of variation in the growth rate of the entire US economy, and justifies the use of our mixed frequency factor model to study the comovement among sectors.

Likewise, for the non-IP data, we also find evidence in favor of either one or two pervasive factors. We adopt a conservative approach, and select a specification with two pervasive factors in both panels, i.e. $k_{XH} = k_{XL} = 2$.

Let us consider the dataset where the HF data are quarterly IP indices, and the LF data are annual GDP non-IP indices. In order to select the number of common and frequency-specific factors, we follow the procedure detailed in Section 6.1. In Table 2 we report the estimated canonical correlations

Table 2: Canonical correlations and test statistics for common factors

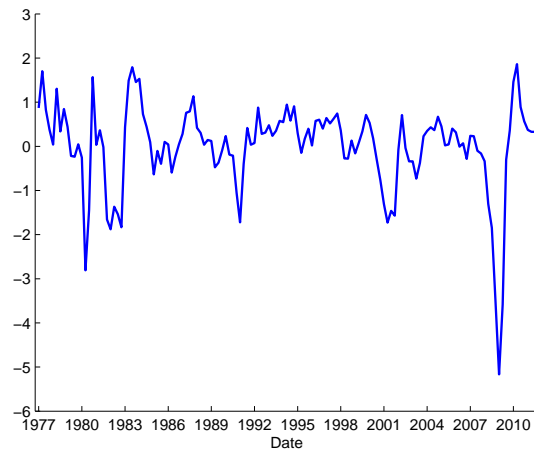
$\hat{\rho}_1$	$\hat{\rho}_2$	$\tilde{\xi}(2)$	$\tilde{\xi}(1)$
0.84	0.06	-3.56	-1.56

We report the two largest canonical correlations among the PCs computed from each subpanel of IP and non-IP data, and the values of statistic $\xi(r)$, i.e. the feasible standardized value of the test statistic $\hat{\xi}(r)$, for the null hypotheses of $r = 2$ and $r = 1$ common factors, respectively. The quarterly IP data are for sample period 1977.Q1-2011.Q4, the annual non-IP data are GDP growth rates for the sample period 1977-2011.

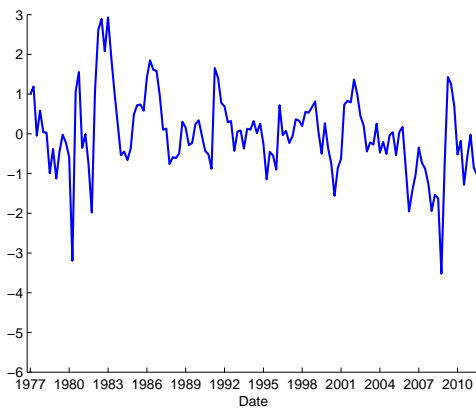
of the first two PC's estimated in each subpanel X^H and X^L , which are used to compute the value of the test statistic $\hat{\xi}(r)$, for the null hypotheses of $r = 2$ and $r = 1$ common factors. We note that the first canonical correlation is close to one, which is consistent with the presence of one common factor in each of the two mixed frequency datasets considered. The test rejects the null hypothesis $r = 2$, i.e. the presence of two common factors, for significance levels as small as 0.1%, while we cannot reject the null of one common factor with a 5% significance level. Our selection procedure recalled in Section 6.1 produces the estimate $\hat{k}^C = 1$. In light of the results in Tables 1 and 2 we select a model with $k^C = k^H = k^L = 1$.

Once the factor space dimensions are determined, the factor values are obtained using the estimation procedure of Section 6.1. In Figure 3 we plot the estimated factor paths from the panels of 42 GDP sectors and 117 IP indices on the entire sample going from 1977 to 2011. All factors are standardized to have zero mean and unit variance, and their sign is chosen so that the majority of the associated loadings are positive. A visual inspection of the plots in Figure 3 reveals that the common factor in Panel (a) resembles the IP index of Figure 2, with a large decline corresponding to the Great Recession following the financial crisis of 2007-2008 and the positive spike associated to the recent economic recovery. On the other hand, the LF-specific factor features a less dramatic fall during the Great Recession, and actually features a positive spike in 2008, followed by large negative values in

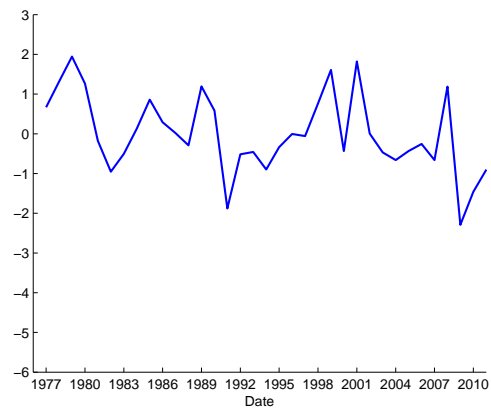
Figure 3: Sample paths of the estimated common and specific factors



(a) Common factor



(b) HF specific factor



(c) LF specific factor

Panel (a) displays the time series plot of the estimated common factor. Panel (b) displays that of the HF-specific factor and finally Panel (c) that of the LF-specific factor. The factors are estimated from the panels of 42 annual non-IP GDP sectoral series and 117 quarterly IP indices using a mixed frequency factor model with $k^C = k^H = k^L = 1$. The sample period is 1977.Q1-2011.Q4.

the following years. This constitutes preliminary evidence suggesting that some non-IP sectors could feature different responses to the financial crisis of 2007-2008. The economic interpretation of factors is easier when they are used as explanatory variables in standard regression analysis. We start with a disaggregated analysis, and look at the relative importance of the common and frequency specific factors in explaining the variability across all sectoral growth rates. For each sector in the panel, we regress the index growth rates (i) on the common factor only, (ii) on the specific factor only, for non-IP and IP series respectively, and (iii) on both common and specific factors. In Table 3 we report the

Table 3: Adjusted \bar{R}^2 of regressions on common factors from indices growth rates

Panel A						Panel B					
Factors	\bar{R}^2 : Quantiles					Factors	\bar{R}^2 : Quantiles				
	10%	25%	50%	75%	90%		10%	25%	50%	75%	90%
<i>Observables: Gross Domestic Product, 1977-2011</i>						<i>Observables: Gross Output, 1988-2011</i>					
common	-2.2	-0.5	11.5	28.9	42.9	common	-2.0	6.6	28.2	45.6	64.5
common, LF-spec.	0.1	9.2	25.4	34.5	60.3	common, LF-spec.	2.8	15.2	45.0	63.7	70.8
LF-spec.	-2.8	-2.3	5.7	15.7	22.4	LF-spec.	-4.5	-3.8	3.2	13.4	40.7
<i>Observables: IP, 1977.Q1-2011.Q4</i>						<i>Observables: IP, 1988.Q1-2011.Q4</i>					
common	0.3	4.8	20.3	36.0	60.0	common	0.1	3.5	10.5	29.8	48.2
common, HF-spec.	1.1	6.8	28.7	45.3	63.4	common, HF-spec.	0.8	7.9	28.2	43.2	65.4
HF-spec.	-0.7	-0.1	3.0	11.2	23.5	HF-spec.	-0.8	2.0	10.0	21.9	33.9

Panel A. The regressions in the first three lines involve the growth rates of the 42 non-IP sectors as dependent variables, while those in the last three lines involve the growth rates of the 117 IP indices as dependent variables. The explanatory variables are factors estimated from the same indices using a mixed frequency factor model with $k^C = k^H = k^L = 1$. The sample period for the estimation of both the factor model and the regressions is 1977-2011. *Panel B.* The regressions in the first three lines involve the Gross Output growth rates of the 38 non-IP sectors as dependent variables, while those in the last three lines involve the growth of the 117 IP indices as dependent variables. The explanatory variables are factors estimated from the same indices using a mixed frequency factor model with $k^C = k^H = k^L = 1$. The sample period for the estimation of both the factor model and the regressions is 1988-2011.

quantiles of the empirical distribution of the adjusted \bar{R}^2 (denoted $\bar{\bar{R}}^2$) of these regressions. In each panel, the first and fourth rows report the quantiles of $\bar{\bar{R}}^2$ of the regressions involving as explanatory variable the common factor only for the IP and non-IP series respectively, while the second and fifth rows report the quantiles of $\bar{\bar{R}}^2$ when the explanatory variables are the common and frequency-specific factors. Finally, the quantiles of $\bar{\bar{R}}^2$ in the third and sixth rows refer to regressions where the explanatory variable is the frequency-specific factor only.²³

From the first three lines of Panel A we observe that adding the LF specific factor to the common factor regressions for the non-IP indices yields an increment of the median $\bar{\bar{R}}^2$ around 14%, going from 11.5% to 25.4%, and the 90% quantile increases by more than 17%. On the other hand, the HF-specific factor, when added to the common factor, contributes less to the increments in $\bar{\bar{R}}^2$ for the IP sectors. In Panel B we note that for both the IP and non-IP Gross Output sectoral indices, the frequency-specific factor increases the median $\bar{\bar{R}}^2$ by more than 15%, when added to the common factor. Overall, Table 3 confirms that the common and frequency-specific factors explain a significant part of the variability

²³The regressions in the second and third rows are restricted MIDAS regressions. The regressions in fourth, fifth and sixth rows impose the estimated coefficients of the common and HF-specific factors to be the same for each quarter, as they are estimated as HF regressions. The empirical distribution of the $\bar{\bar{R}}^2$ corresponding to the first and second lines of Table 3, Panel A, are represented in the histograms available in OA, Figures C.9 (a) and (b). The empirical distribution of the $\bar{\bar{R}}^2$ corresponding to the fourth and fifth lines of Table 3, Panel A, are represented in the histograms available in OA, Figures C.10 (a) and (b).

of output growth for the majority of the sectors of the US economy. Moreover, the common factor is pervasive for most of the IP and non-IP sectors alike.

In order to provide the economic interpretation to the estimated factors, we list in Table 4 the top and bottom ten GDP non-IP sectors in terms of \bar{R}^2 when regressed on the common factor only, and both the common and LF-specific factors. We also report the top and bottom ten GDP non-IP sectors with the highest and lowest absolute increments in \bar{R}^2 when the LF-specific factor is added to the common one.²⁴ From Panel A we first note that the common factor explains most of the variability of service sectors with direct economic links to industrial production sectors like Transportation and Warehousing: for instance, “Truck Transportation”, “Other Transportation & Support Activities”, and “Warehousing & Storage” have an \bar{R}^2 of 63%, 43% and 41%, respectively, when regressed on the common factor only. This is another clear indication that the common factor could be interpreted as IP factor, as already noticed on Figure 3. On the other hand, the common factor is completely unrelated to Agriculture, forestry, fishing & hunting, as well as to most of the Financial and Information services sectors.

Turning to Panel C of Table 4, we note that the LF-specific factor explains more than 20% of the variability of output for very heterogeneous services sectors like “Miscellaneous professional, scientific, & technical services, Administrative & support services”, “Legal services”, “Real Estate”, some important financial services like “Federal Reserve banks, Credit intermediation, & Related activities”, “Rental & Leasing Services” but also “Government (state & local)”. Interpreting these results, we can conclude that the LF-specific factor is completely unrelated to service sectors which depend almost exclusively on IP output, and is a common factor driving the comovement of non-IP sectors such as some other service sectors, Construction and government sectors.

In Table 4 we highlight further differences in the dynamics of output growth between the two sub-sectors of the financial services industry which are particularly revealing: “Securities” and “Credit intermediation”, extensively studied by Greenwood and Scharfstein (2013). We find that the subsectors “Funds, trusts, & other financial vehicles” and “Securities, commodity contracts, & investments” are unrelated to both the common and LF-specific factors, indicating that their output growth is uncorrelated with the common component of real output growth across the other sectors of the US economy.

²⁴The entire list of non-IP sectors ranked by the three criteria used in Table 4, are available in Tables C.20-C.22 in the OA, Section C.5.

Table 4: Regression of yearly sectoral GDP growth on the common and LF-specific factors: adjusted R^2

Panel A. Regressor: common factor		Panel B. Regressors: common and LF spec. factors		Panel C. Increment in adjusted R^2	
Sector	\bar{R}^2	Sector	\bar{R}^2	Sector	$\Delta \bar{R}^2$
<i>Ten sectors with largest \bar{R}^2</i>					
Truck transportation	63.10	Misc. prof., scient., & tech. serv.	66.67	Misc. prof., scient., & tech. serv.	49.69
Accommodation	62.43	Administrative & support services	62.63	Government enterprises (state & local)	34.69
Construction	44.05	Truck transportation	62.51	Rental & leasing serv. & lessors of int. assets	29.52
Other transp. & support activ.	43.31	Accommodation	61.48	General government (state & local)	24.90
Administrative & support services	42.69	Construction	59.75	Legal services	24.32
Other services, except government	42.53	Warehousing & storage	52.53	Motion picture & sound recording ind.	22.77
Warehousing & storage	40.95	Government enterprises (state & local)	45.78	Fed. Reserve banks, credit interm., & rel. activ.	20.31
Air transportation	31.58	Other services, except government	41.75	Administrative & support services	19.95
Retail trade	30.70	Other transportation & support activities	41.71	Social assistance	19.91
Amusem., gambling, & recr. ind.	29.17	Government enterprises (federal)	37.78	Real estate	18.14
<i>Ten sectors with smallest \bar{R}^2</i>					
Funds, trusts, & other finan. vehicles	-1.23	Ambulatory health care services	7.76	Accommodation	-0.96
Motion picture & sound record. ind.	-1.68	Management of companies & enterprises	7.52	Rail transportation	-1.16
Pipeline transportation	-1.74	Funds, trusts, & other fin. vehicles	6.15	Other transportation & support activities	-1.59
Information & data processing services	-1.84	Information & data processing services	1.96	Air transportation	-1.77
Transit & ground passenger transp.	-2.05	Educational services	1.35	Retail trade	-2.15
General government (state & local)	-2.12	Insurance carriers & related activities	0.36	Amusements, gambling, & recreation ind.	-2.15
Forestry, fishing & related activities	-2.33	Water transportation	-0.64	Educational services	-2.62
Water transportation	-2.94	Farms	-1.87	Farms	-2.80
Securities, commodity contracts, & investm.	-2.99	Forestry, fishing, & related activities	-5.31	Forestry, fishing, & related activities	-2.98
Insurance carriers & related activities	-3.03	Securities, commodity contracts, & investm.	-5.99	Securities, commodity contracts, & investm.	-3.00

In the table we report the adjusted R^2 , denoted \bar{R}^2 , for restricted MIDAS regressions of the growth rates of 42 GDP non-IP sectoral indices on the estimated factors. The factors are estimated from the panel of 42 GDP sectors and 117 IP indices using a mixed frequency factor model with $k^C = k^H = k^L = 1$. The sample period for the estimation of both factor model and regressions is 1977-2011. Regressions in *Panel A* involve a LF explained variable and the estimated common factor. Regressions in *Panel B* involve a LF explained variable and both the common and LF-specific estimated factors. The regressions in both cases are restricted MIDAS regressions. In *Panel C* we report the difference in \bar{R}^2 (denoted as $\Delta \bar{R}^2$) between the regressions in *Panel B* and regressions in *Panel A*.

In contrast, the “Credit intermediation” industry comoves with the other IP and non-IP sectors (see also Tables C.20 and C.21 in the OA).

Up to this point, we have looked at the explanatory power of the factors for sectoral output indices. For both the non-IP GDP and Gross Output, these indices correspond to the finest level of disaggregation of output growth by sector. In Table 5 we report the results of regressions with aggregated indices instead. In particular, we regress the output of each aggregate index either on the estimated common factor or the common and frequency specific factors, and focus on the adjusted \bar{R}^2 s of these regressions. It is important to note that we also include the GDP Manufacturing aggregate index which is *not* used in the estimation of the factors. This will help us with the interpretation of our estimated common and frequency-specific factors. Panel A of Table 5 shows that the common factor explains around 90% of the variability in the aggregate IP index, confirming that the common factor can be interpreted as an Industrial Production factor. This is further corroborated in Panel B where we find an \bar{R}^2 around 82% for the regression of the GDP Manufacturing Index on the common factor only. As most of the sectors included in the Industrial Production index are Manufacturing sectors, this result is not surprising, but is still worth noting because, as noted earlier, the GDP data on Manufacturing have not been used in the factor estimation, in order not to double-count these sectors in our mixed frequency sectoral panel.²⁵ As expected from the results in Table 4, more than 60% of the variability of GDP of Transportation and Warehousing services index is explained by the common factor only, and the LF-specific factor has no explanatory power. On the other hand, the HF-specific factor seems not to be important in explaining the aggregate IP index, as the \bar{R}^2 increases only by 1% when it is added as a regressor to the common factor.²⁶ This suggests that the HF-specific factor is pervasive only for a subgroup of IP sectors which have relatively low weights in the index, meaning that their aggregate output is a negligible part of the output of the entire IP sector and, consequently, also the entire US economy.²⁷

Looking at the aggregate GDP index, we first note that even if the weight of Industrial Production sectors in the aggregate nominal GDP index has always been below 30%, as evident from Figure 1,

²⁵A detailed discussion of the difference in the sectoral components of the IP index and the GDP Manufacturing index is provided in OA, Section C.4.

²⁶See also Table C.23 in OA, Section C.5, for the \bar{R}^2 of the regression of all GDP indices on the HF factor only, and all the 3 factors together.

²⁷These results corroborate the findings of Foerster, Sarte, and Watson (2011), who claim that the main results of their paper are qualitatively the same when considering either one or two common factors extracted from the same 117 IP indices of our study.

Table 5: Adj. R^2 of aggregate IP and selected GDP indices growth rates on estimated factors

Panel A <i>Quarterly observations, 1977.Q1-2011.Q4</i>				
Sector	(1) $\bar{R}^2(C)$	(2) $\bar{R}^2(H)$	(3) $\bar{R}^2(C + H)$	(3) - (1)
Industrial Production	89.06	5.02	90.26	1.20
Panel B <i>Yearly observations, 1977-2011</i>				
Sector	(1) $\bar{R}^2(C)$	(2) $\bar{R}^2(L)$	(3) $\bar{R}^2(C + L)$	(3) - (1)
GDP	60.54	8.59	74.21	13.67
GDP - Manufacturing	81.88	-3.03	81.53	-0.35
GDP - Agriculture, forestry, fishing, and hunting	1.43	-2.52	-1.26	-2.69
GDP - Construction	44.05	11.22	59.75	15.70
GDP - Wholesale trade	20.35	7.90	30.83	10.48
GDP - Retail trade	30.70	-2.86	28.56	-2.15
GDP - Transportation and warehousing	62.14	-2.95	60.97	-1.17
GDP - Information	12.14	22.28	37.57	25.43
GDP - Finance, insurance, real estate, rental, and leasing	-1.42	21.22	21.11	22.53
GDP - Professional and business serv.	30.02	30.21	65.61	35.59
GDP - Educational serv., health care, and social assist.	-1.38	18.38	18.18	19.56
GDP - Arts, entert., recreat., accomm., and food serv.	53.51	-2.23	53.70	0.18
GDP - Government	-2.12	22.37	20.47	22.59

In the table we report the adjusted R^2 , denoted \bar{R}^2 , of the regression of growth rates of the aggregate IP index and selected aggregated sectoral GDP non-IP output indices on the common factor (column $\bar{R}^2(C)$), the specific HF and LF factors (columns $\bar{R}^2(H)$ and $\bar{R}^2(L)$) only, and the common and frequency-specific factors together (column (3)). The last column displays the difference between the values in the third and first columns. The factors are estimated from the panel of 42 GDP non-IP sectors and 117 IP indices using a mixed frequency factor model with $k^C = k^H = k^L = 1$. The sample period for the estimation of both factor model and regressions is 1977-2011.

still 60% of its total variability can be explained exclusively by the common factor which - as shown in Panel A - is primarily an IP factor. This implies that there must be substantial comovement between IP and some important service sectors. Moreover, it appears from the first line in Panel B that a relevant part of the variability of the aggregate GDP index not due to the common factor is explained by the LF-specific factor (the \bar{R}^2 increases by about 14% to 74%).²⁸ This indicates that significant comovements are present among the most important sectors of the US economy which are not related to manufacturing. Indeed, Panel B in Table 5 indicates that some services sectors such as Professional & Business Services and Information and Construction load significantly both on the common and

²⁸See the results in Table C.23 in the OA.

the LF-specific factor, while some other sectors like Finance and Government load exclusively on the LF-specific factor.²⁹

Our sample covers what is known as the Great Moderation, which refers to a reduction in the volatility of business cycle fluctuations starting in the mid-1980s. We performed therefore an analysis on subsamples. Results are provided in the OA, Section C.5.4, and here we discuss briefly the main findings. There is a deterioration of the overall fit of approximate factor models during the Great Moderation sample starting in 1984 and ending in 2007 – a finding also reported by Foerster, Sarte, and Watson (2011) – and the common factor plays a lesser role during that period. Interestingly, when the financial crisis is added to the Great Moderation (sample 1984-2011), we find patterns closer to the full sample results presented above. The other findings, i.e. the exposure of the various subindices, appear to be similar in subsamples and in the full sample.

6.4 Structural model and productivity shocks

The macroeconomics literature, with the works of Long and Plosser (1983), Horvath (1998) and Carvalho (2010), among many others, has recognized that input-output linkages in both intermediate materials and capital goods lead to propagation of sector-specific shocks in a way that generates comovements across sectors. An important contribution of the work of Foerster, Sarte, and Watson (2011) is to describe the conditions under which an approximate linear factor structure for sectoral output growth arises from standard neoclassical multisector models including those linkages. In particular, these authors develop a generalized version of the multisector growth model of Horvath (1998), which allows them to filter out the effects of these linkages, and reconstruct the time series of productivity shocks using sector data on output growth when input-output tables for intermediate materials and capital goods are available. We can characterize this as statistical versus structural factor analysis.

The main objective of this section is to verify the presence of a common factor in the *innovations* of productivity for *all* the sectors (not just IP) of the US economy by means of our mixed frequency factor model. If a common factor is present also in the productivity shocks, then the factor structure uncovered by the reduced form analysis of output growth in Section 6.3 is not only due to interlinkages in materials and capital use among different sectors.

²⁹The results change when we look at the Finance sector disaggregated in “Fed. Reserve banks, credit interm., & rel. activ.,” “Securities, commodity contracts, & investm.,” “Insurance carriers & related activities”, as evident in Table 4.

We rely on the same multi-industry real business cycle model described in Section IV of Foerster, Sarte, and Watson (2011) to extract productivity shocks from the time series of the growth rates of the same 117 IP indices considered in the previous section, and the growth rates of 38 non-IP private sectors Gross Output, therefore excluding the 4 Government indices considered previously.³⁰ One challenge due to the mixed frequency nature of our output growth dataset consists in the extraction of mixed frequency technological shocks. In the OA, Section C.6 we explain how to adapt the algorithm proposed by Foerster, Sarte, and Watson (2011) to estimate technological shocks for our mixed frequency output series. Specifically, the multi-sector business cycle model that we use to filter out the technological shocks correspond to the “Benchmark” model considered by Foerster, Sarte, and Watson (2011) in their Section IV, while the data on input-output and capital use matrices necessary to estimate the model are built from the BEA’s 1997 “use table” and “capital flow table”, respectively.³¹ Using the extracted productivity shocks for the IP and non-IP sectors, denoted $\hat{\zeta}_{m,t}^H$ and $\hat{\zeta}_t^L$, respectively, we estimate a mixed frequency factor model with these productivity shock series. The sample period for the estimation of both the factor model and the regressions is 1989-2011, because the productivity shocks can not be computed for the first year of the sample (see Foerster, Sarte, and Watson (2011), especially equation (B38) on page 10 of their Appendix B). For a direct comparison between the statistical factor model covered in the previous subsection and the structural factor analysis, we need to first re-estimate our model with one common, one HF-specific and one LF-specific factors on the panels of growth rates of annual Gross Output non-IP indices (as opposed to the GDP growth indices in Table 5) and the same 117 quarterly sectoral IP indices. The results of the corresponding in-sample regressions are reported in Table 6.

For the moment we focus exclusively on the shaded areas of Table 6, as the non-shaded areas pertain to the productivity shocks which will be covered later. We expect some difference with the results displayed in Table 5 for at least two reasons. First, the dataset in which the non-IP data are Gross Output indices, refers to shorter time period going from 1988, instead of 1977, to the end of 2011, as Gross Output indices are not available before 1988. Second, as the panel in Table 6 does not include the four governmental sectors, we expect that the common and frequency-specific factors may have different dynamics when compared to those extracted from the panel with GDP non-IP sectors.

³⁰The exclusion of the public sector from the analysis is a standard choice in the sectoral productivity literature.

³¹The last year for which sectoral capital use tables have been constructed by the BEA is 1997.

Table 6: Adj. R^2 of aggregate IP and selected Gross Output indices growth rates on estimated factors from raw data (shaded) and estimated factors from productivity innovations (not shaded)

Panel A <i>Quarterly observations, 1988.Q1-2011.Q4</i>				
Sector	(1) $\bar{R}^2(C)$	(2) $\bar{R}^2(H)$	(3) $\bar{R}^2(C + H)$	(3) - (1)
Industrial Production	63.71	38.32	89.48	25.78
	31.21	50.15	77.25	46.05
Panel B <i>Yearly observations, 1988-2011</i>				
Sector	(1) $\bar{R}^2(C)$	(2) $\bar{R}^2(L)$	(3) $\bar{R}^2(C + L)$	(3) - (1)
GO (all sectors)	68.54	12.20	89.66	21.12
	42.17	13.97	57.60	15.43
GO - Manufacturing	86.08	-3.05	88.94	2.86
	62.29	-0.20	64.42	2.13
GO - Agriculture, forestry, fishing, and hunting	-3.21	3.35	-0.25	2.96
	0.96	-4.23	-3.35	-4.31
GO - Wholesale trade	80.82	-3.85	79.97	-0.85
	74.73	-3.08	74.74	0.01
GO - Construction	25.30	34.16	67.15	41.84
	6.64	20.55	27.78	21.14
GO - Retail trade	64.72	-4.50	63.15	-1.57
	47.02	-4.35	45.04	-1.98
GO - Transportation and warehousing	83.82	-4.51	83.22	-0.60
	70.42	-2.69	70.58	0.15
GO - Information	33.70	38.59	81.54	47.84
	17.78	42.45	61.76	43.98
GO - Finance, insurance, real estate, rental, and leasing	3.37	50.30	59.29	55.92
	-4.09	17.55	13.96	18.05
GO - Professional and business services	45.13	21.97	75.48	30.36
	25.17	44.89	71.81	46.64
GO - Educational serv., health care, and social assist.	-4.19	-1.58	-6.17	-1.98
	-4.73	-4.48	-9.66	-4.93
GO - Arts, entert., recreat., accomm., and food serv.	71.06	-3.74	71.90	0.84
	55.64	-2.29	55.49	-0.16

In the table we display the adjusted R^2 , denoted \bar{R}^2 , of the regressions of growth rates of the aggregate IP index and selected aggregated sectoral Gross Output non-IP output indices on the common factor (column $\bar{R}^2(C)$), the specific HF and LF factors (columns $\bar{R}^2(H)$ and $\bar{R}^2(L)$) only, and the common and frequency-specific factor together (column (3)). The last column displays the difference between the values in the third and first columns. The factors are estimated from the panel of 38 Gross Output non-IP sectors and 117 IP indices using a mixed frequency factor model with $k^C = k^H = k^L = 1$. The sample period for the estimation of both factor model and regressions is 1988-2011 for the regressions in the shaded areas, while it is 1989-2011 for the regressions in the non-shaded areas. The regression results in the shaded areas are obtained using as regressors the factors extracted from raw data, while the regression results in the non-shaded areas are obtained using as regressors the factors extracted from productivity innovations.

Compared to the previous Section 6.3, we obtain qualitatively similar results, as shown in Table 6. There appear to be only two notable differences with the results reported in Table 5. We see an increased importance of the HF-specific factor in explaining the variability of the IP aggregate index (see Panel A in Table 6), at the expense of a lower explanatory power for the common factor. Moreover, there is also an increased importance of both the common and LF-specific factors in explaining the total variability of total aggregate output (measured as total Gross Output, in the first line of Panel B in Table 6). Still the common factor explains roughly 65 % of the variation in the panel of IP data.

Table 7: Adjusted R^2 of regressions on common factors from productivity innovations

Panel A						Panel B					
Factors	Adjusted R^2 : Quantile					Factors	Adjusted R^2 : Quantile				
	10%	25%	50%	75%	90%		10%	25%	50%	75%	90%
<i>Observables: Gross Output productivity innovations, 1989-2011</i>						<i>Observables: Gross Output, 1988-2011</i>					
common	-3.3	-0.3	11.0	33.6	46.1	common	-2.4	3.7	21.2	31.5	55.8
common, LF-spec.	-2.6	4.8	26.3	45.0	60.7	common, LF-spec.	-0.9	7.8	28.2	56.9	68.0
LF-spec.	-4.2	-3.6	-0.1	17.7	33.1	LF-spec.	-4.6	-3.3	1.3	20.6	43.8
<i>Observables: IP productivity innovations, 1989.Q1-2011.Q4</i>						<i>Observables: IP, 1988.Q1-2011.Q4</i>					
common	-1.0	-0.4	1.5	12.1	22.4	common	-0.8	0.2	4.5	17.7	34.7
common, HF-spec.	-0.6	3.1	13.1	28.4	40.1	common, HF-spec.	1.2	5.9	25.7	40.8	63.8
HF-spec.	-0.7	0.6	6.2	18.7	28.2	HF-spec.	-0.3	2.2	14.7	29.2	37.8

Panel A: The regressions in the first three lines involve the productivity innovations of the 38 non-IP sectors as dependent variables, while the regressions in the last three lines involve the productivity innovations of the 117 IP indices as dependent variables. Productivity innovations are computed using the panel of Gross Output growth rates for the LF observables adapting the procedure of Foerster, Sarte, and Watson (2011). The explanatory variables are factors estimated from a mixed frequency factor model with $k^C = k^H = k^L = 1$, applied on the panels of productivity innovations. The sample period for the estimation of both the factor model and the regressions is 1989.Q1-2011.Q4. Panel B: The regressions in the first three lines involve the Gross Output growth rates of the 38 non-IP sectors as dependent variables, while the regressions in the last three lines involve the growth of the 117 IP indices as dependent variables. The explanatory variables are the same factors used in the regressions of Panel A extracted from productivity innovations. The sample period for the estimation of both the factor model and the regressions is 1989.Q1-2011.Q4.

What do we learn from the structural analysis with common and frequency-specific factors of productivity shocks? First, it is quite interesting to find that again there is one common factor in productivity shocks. Indeed, the selection of the number of common factors is performed as in the previous section, and our testing methodology suggests the presence of one common factor. Therefore we estimate a model for the productivity innovations with $k^C = k^H = k^L = 1$.³² As in the previous section, we start with a disaggregated analysis and look at the relative importance of the new common and frequency specific factors in explaining the variability of the constituents of the panel of produc-

³²The values of the penalized selection criteria of Bai and Ng (2002) performed on different subpanels and the test for the number of common factors are available in Tables C.24 and C.25 in the OA, Section C.5.

tivity innovations, and the panels of all output growth rates used for the extraction of the productivity innovations themselves. For each sector, we regress both the productivity innovations and the index growth rates on the common factor only, on the specific factor only, and on both common and specific factors. In Table 7 we report the quantiles of the empirical distribution of \bar{R}^2 of these regressions.³³

Panel A of Table 7 confirms that both the common and the frequency-specific factors are pervasive for the panels of productivity innovations. From the first two rows we note that the common factor alone explains at least 11% of the variability of half of the non-IP series considered, and this fraction increases to more than 26% when the LF-specific factor is an additional regressor to the common one. On the other hand, from the last three rows of Panel B we note that for the IP sectors the high frequency specific factor seems to explain the majority of the variability of the productivity indices, while the explanatory power of the common factor seems to have some relevance only for 50% of the IP sectors. Panel B reports the \bar{R}^2 of the regressions of the GO indices growth rates on the factors estimated on the panels of productivity shocks themselves. Therefore, they give an indication of the fraction of variability of the indices explained by the common components of the output growth which is not due to input-output linkages between sectors, as captured by the structural “Benchmark” of Foerster, Sarte, and Watson (2011). Panel B of Table 7 can be compared with Panel B of Table 3. As expected, as part of the comovement among different sectors is due to input-output and capital use linkages, all the \bar{R}^2 in Panel B of Table 7 are strictly lower than those in Table 3, if we exclude the negative ones and those very close to zero. For instance, the median \bar{R}^2 of regressions including the common factor only for the non-IP sectors decreases from 28% to 21%, and the median \bar{R}^2 of regressions including the common and LF-specific factors decreases from 45% to 28%. A similar pattern is observed for the higher quantiles, and for the IP indices. However, overall Panel B in Table 7 gives a first indication of the presence of commonality in the comovement on the majority of the sectors of the US economy even when the output growth rates are purged of the input-output linkages in both intermediate materials and capital goods.

We conclude the analysis by repeating the same exercise as for the shaded areas of Table 6, and regress the Industrial Production and aggregate (mostly non-IP) Gross Output indices growth on the factors extracted from productivity innovations and look at the adjusted R^2 s. The results are provided

³³The regressions in the second and third rows are restricted MIDAS regressions. The regressions in fourth, fifth and sixth rows impose the estimated coefficients of the common and HF-specific factors to be the same at each quarter, as they are estimated as HF regressions.

in the non-shaded rows of Table 6. From Panel A we observe that the common factor extracted from productivity innovations explains around 31% of the variability of the aggregate IP index, i.e. around half of the variability explained by the common factor extracted directly from the output series. Moreover, when the high frequency-specific productivity factor is added as explanatory variable, the \bar{R}^2 increases to 77%, which is also significantly smaller than the 89% \bar{R}^2 obtained using as regressors the factors extracted from the output series.³⁴ Hence, the case of a common pervasive factor in innovation shocks across the entire economy mainly related to IP sector technology shocks is less compelling. From Panel B of Table 6 we observe that 42% of the variability of the aggregate Gross Output of the US economy can be explained by the common factor of productivity shocks, and when the factor specific to non-IP sector is added, the \bar{R}^2 grows to 57%.

From this analysis we learn something interesting which Foerster, Sarte, and Watson (2011) were not able to address since they exclusively examined IP sectors. Overall, there is a difference in the explanatory power of factors in structural versus non-structural factor models - as they found. However, it seems that looking at technology shocks instead of output, it does not appear that a common factor explaining IP fluctuations is a dominant factor for the entire economy. A factor specific to technological innovations in IP sectors is more important for the IP sector shocks and a low frequency factor which appears to explain variation in information industry as well as professional and business services innovations plays, relatively speaking, a more important role.

7 Conclusions

Panels with data sampled at different frequencies are the rule rather than the exception in economic applications. We develop a novel approximate factor modeling approach which allows us to estimate factors which are common across all data regardless of their sample frequency, versus factors which are specific to subpanels stratified by sampling frequency. To develop the generic theoretical framework, we cast our analysis into a group factor structure and develop a unified asymptotic theory for the identification of common and group- or frequency-specific factors, for the determination of the number of common and specific factor values, for the estimation of loadings and the factor values via principal component analysis and canonical correlation analysis in a setting with large dimensional data sets,

³⁴This result is in line with the findings of Foerster, Sarte, and Watson (2011) in their Section IV C.

using asymptotic expansions both in the cross-sections and the time series.

There are a plethora of applications to which our theoretical analysis applies. We selected a specific example based on the work of Foerster, Sarte, and Watson (2011) who analyzed the dynamics of comovements across 117 industrial production sectors using both statistical and structural factor models. We revisit their analysis and incorporate the rest - and most dominant part - of the US economy, namely the non-IP sectors whose growth rate we only observe annually. We find evidence for a common factor for the entire US economy, but this common factor appears not to be an IP factor after accounting for input-output linkages.

Despite the generality of our analysis, we can think of many possible extensions, such as models with loadings which change across subperiods, i.e. periodic loadings, or loadings which vary stochastically or feature structural breaks. Moreover, we could consider the problem of specification and estimation of a joint dynamic model for the common and frequency-specific factors extracted with our methodology (see Ghysels (2016) and the references therein for structural Vector Autoregressive (VAR) models with mixed frequency sampling). Further, in the interest of conciseness we have focused our analysis on models with two sampling frequencies, leading to group factor models with two groups. Results could be extended to cover the cases with more than two groups / more than two sampling frequencies. For conducting inference on the factor spaces and their dimensions, the key ingredients are the link with the eigenvalues and eigenvectors of the variance-covariance matrix of stacked PCs extracted from the groups, and the distributional theory developed in Section 4. All these extensions are left for future research.

References

- AMENGUAL, D., AND M. W. WATSON (2007): “Consistent Estimation of the Number of Dynamic Factors in a Large N and T Panel,” *Journal of Business and Economic Statistics*, 25(1), 91–96.
- ANDERSON, T. W. (2003): *An Introduction to Multivariate Statistical Analysis*. Wiley-Interscience.
- ANDO, T., AND J. BAI (2015): “Multifactor Asset Pricing with a Large Number of Observable Risk Factors and Unobservable Common and Group-specific Factors,” *Journal of Financial Econometrics*, 13(3), 556–604.
- (2016): “Panel Data Models with Grouped Factor Structure Under Unknown Group Membership,” *Journal of Applied Econometrics*, 31, 163–191.
- ARUOBA, S. B., F. X. DIEBOLD, AND C. SCOTTI (2009): “Real-time Measurement of Business Conditions,” *Journal of Business and Economic Statistics*, 27(4), 417–427.
- 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(1), 191–221.
- (2006a): “Confidence Intervals for Diffusion Index Forecasts and Inference for Factor-Augmented Regressions,” *Econometrica*, 74(4), 1133–1150.
- (2006b): “Evaluating Latent and Observed Factors in Macroeconomics and Finance,” *Journal of Econometrics*, 131(4), 507–537.
- (2007): “Determining the Number of Primitive Shocks in Factor Models,” *Journal of Business and Economic Statistics*, 25(1), 52–60.
- BANBURA, M., AND G. RÜNSTLER (2011): “A Look into the Factor Model Black Box: Publication Lags and the Role of Hard and Soft Data in Forecasting GDP,” *International Journal of Forecasting*, 27, 333–346.

- BEKAERT, G., R. J. HODRICK, AND X. ZHANG (2009): “International Stock Return Comovements,” *Journal of Finance*, 64, 2591–2626.
- BOIVIN, J., AND S. NG (2006): “Are More Data Always Better for Factor Analysis?,” *Journal of Econometrics*, 132(1), 169–194.
- BREITUNG, J., AND S. EICKMEIER (2016): “Analyzing International Business and Financial Cycles Using Multi-level Factor Models: A Comparison of Alternative Approaches,” *Advances in Econometrics*, 15, 177–214.
- BROWNE, M. W. (1979): “The Maximum-Likelihood Solution in Inter-battery Factor Analysis,” *British Journal of Mathematical and Statistical Psychology*, 32, 75–86.
- (1980): “Factor Analysis of Multiple Batteries by Maximum Likelihood,” *British Journal of Mathematical and Statistical Psychology*, 33(2), 184–199.
- CAMBA-MENDEZ, G., AND G. KAPETANIOS (2009): “Statistical Tests and Estimators of the Rank of a Matrix and Their Applications in Econometric Modelling,” *Econometric Reviews*, 28, 581–611.
- CARVALHO, V. M. (2010): “Aggregate Fluctuations and the Network Structure of Intersectoral Trade,” *Unpublished Manuscript*.
- CHAMBERLAIN, G., AND M. ROTHSCHILD (1983): “Arbitrage, Factor Structure and Mean-Variance Analysis on Large Asset Markets,” *Econometrica*, 51, 1281–1304.
- CHEN, K. H., AND J. ROBINSON (1989): “Comparison of Factor Spaces of two Related Populations,” *Journal of Multivariate Analysis*, 28, 190–203.
- CHEN, P. (2010): “A Grouped Factor Model,” *Working Paper*.
- (2012): “Common Factors and Specific Factors,” *Working Paper*.
- CHO, D. C. (1984): “On Testing the Arbitrage Pricing Theory: Inter-Battery Factor Analysis,” *Journal of Finance*, 39(5), 1458–1502.
- CRAGG, J. G., AND S. G. DONALD (1996): “On the Asymptotic Properties of LDU-Based Tests of the Rank of a Matrix,” *Journal of the American Statistical Association*, 91, 1301–1309.

- (1997): “Inferring the Rank of a Matrix,” *Journal of Econometrics*, 76, 223–250.
- DAUXOIS, J., AND A. POUSSE (1975): “Une Extension de l’Analyse Canonique. Quelques Applications,” *Ann. Inst. Henri Poincaré*, XI, 355–379.
- DAUXOIS, J., Y. ROMAIN, AND S. VIGUIER (1993): “Comparison of Two Factor Subspaces,” *Journal of Multivariate Analysis*, 44(1), 160–178.
- DEMPSTER, A., N. LAIRD, AND D. RUBIN (1977): “Maximum Likelihood Estimation from Incomplete Data,” *Journal of the Royal Statistical Society*, 14, 1–38.
- DONALD, S. G., N. FORTUNA, AND V. PIPIRAS (2007): “On Rank Estimation in Symmetric Matrices: The Case of Indefinite Matrix Estimators,” *Econometric Theory*, 23(6), 1103–1123.
- (2014): “On Rank Estimation in Semidefinite Matrices,” *Working Paper*.
- FLURY, B. N. (1984): “Common Principal Components in k Groups,” *Journal of the American Statistical Association*, 79, 892–898.
- FLURY, B. N. (1988): *Common Principal Components and Related Multivariate Models*. Wiley, New York.
- FOERSTER, A. T., P.-D. G. SARTE, AND M. W. WATSON (2011): “Sectoral versus Aggregate Shocks: A Structural Factor Analysis of Industrial Production,” *Journal of Political Economy*, 119(1), 1–38.
- FRALE, C., AND L. MONTEFORTE (2010): “FaMIDAS: A Mixed Frequency Factor Model with MIDAS Structure,” *Government of the Italian Republic, Ministry of Economy and Finance, Department of the Treasury Working Paper*, 3.
- GAGLIARDINI, P., AND C. GOURIEROUX (2014): “Efficiency in Large Dynamic Panel Models with Common Factors,” *Econometric Theory*, 30, 961–1020.
- GHYSELS, E. (2016): “Macroeconomics and the Reality of Mixed Frequency Data,” *Journal of Econometrics*, 193(2), 294–314.

- GILL, L., AND A. LEWBEL (1992): “Testing the Rank and Definiteness of Estimated Matrices with Applications to Factor, State-space and ARMA Models,” *Journal of the American Statistical Association*, 87, 766–776.
- 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.
- GREENWOOD, R., AND D. SCHARFSTEIN (2013): “The Growth of Finance,” *Journal of Economic Perspectives*, 27(2), 3–28.
- GREGORY, A. W., AND A. C. HEAD (1999): “Common and Country-specific Fluctuations in Productivity, Investment, and the Current Account,” *Journal of Monetary Economics*, 44(3), 423 – 451.
- HALLIN, M., AND R. LISKA (2007): “Determining the Number of Factors in the General Dynamic Factor Model,” *Journal of the American Statistical Association*, 102(478), 603–617.
- (2011): “Dynamic Factors in the Presence of Blocks,” *Journal of Econometrics*, 163, 29–41.
- HORVATH, M. (1998): “Cyclicalities and Sectoral Linkages: Aggregate Fluctuations from Independent Sectoral Shocks,” *Review of Economic Dynamics*, 1, 781–808.
- JUNGBACKER, B., S. KOOPMAN, AND M. VAN DER WEL (2011): “Maximum Likelihood Estimation for Dynamic Factor Models with Missing Data,” *Journal of Economic Dynamics and Control*, 35(8), 1358 – 1368.
- KLEIBERGEN, F., AND R. PAAP (2006): “Generalized Reduced Rank Tests Using the Singular Value Decomposition,” *Journal of Econometrics*, 133, 97–126.
- KORAJCZYK, R. A., AND R. SADKA (2008): “Pricing the Commonality across Alternative Measures of Liquidity,” *Journal of Financial Economics*, 87(1), 45 – 72.
- KOSE, A. M., C. OTROK, AND C. H. WHITEMAN (2008): “Understanding the Evolution of World Business Cycles,” *Journal of International Economics*, 75, 110–130.
- KRZANOWSKI, W. (1979): “Between-groups Comparison of Principal Components,” *Journal of the American Statistical Association*, 74, 703–707.

- LONG, J. B. J., AND C. I. PLOSSER (1983): "Real Business Cycles," *Journal of Political Economy*, 91, 39–69.
- MAGNUS, J. R., AND H. NEUDECKER (2007): *Matrix Differential Calculus with Applications in Statistics and Econometrics*. John Wiley and Sons: Chichester/New York.
- MARCELLINO, M., AND C. SCHUMACHER (2010): "Factor MIDAS for Nowcasting and Forecasting with Ragged-Edge Data: A Model Comparison for German GDP," *Oxford Bulletin of Economics and Statistics*, 72(4), 518–550.
- MARIANO, R. S., AND Y. MURASAWA (2003): "A New Coincident Index of Business Cycles Based on Monthly and Quarterly Series," *Journal of Applied Econometrics*, 18(4), 427–443.
- MOENCH, E., AND S. NG (2011): "A Hierarchical Factor Analysis of US Housing Market Dynamics," *The Econometrics Journal*, 14, C1–C24.
- MOENCH, E., S. NG, AND S. POTTER (2013): "Dynamic Hierarchical Factor Models," *Review of Economics and Statistics*, 95, 1811–1817.
- NUNES, L. C. (2005): "Nowcasting Quarterly GDP Growth in a Monthly Coincident Indicator Model," *Journal of Forecasting*, 24(8), 575–592.
- PELGER, M. (2015): "Large-Dimensional Factor Modeling Based on High-Frequency Observations," Discussion Paper, Stanford University.
- PÖTSCHER, B. M. (1983): "Order Estimation in ARMA-Models by Lagrangian Multiplier Tests," *The Annals of Statistics*, 11(3), 872–885.
- ROBIN, J. M., AND R. SMITH (2000): "Tests of Rank," *Econometric Theory*, 16, 151–175.
- SCHOTT, J. R. (1988): "Common Principal Components Subspaces in Two Groups," *Biometrika*, 75, 229–236.
- (1991): "Some Tests for Common Principal Component Subspaces in Several Groups," *Biometrika*, 78, 771–777.
- SCHOTT, J. R. (1999): "Partial Common Principal Component Subspaces," *Biometrika*, 86, 899–908.

——— (2005): *Matrix Analysis for Statistics*. Wiley, New York, 2 edn.

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

——— (2002b): “Macroeconomic Forecasting Using Diffusion Indexes,” *Journal of Business and Economic Statistics*, 20, 147–162.

TUCKER, L. R. (1958): “An Inter-Battery Method of Factor Analysis,” *Psychometrika*, 23(2), 111–136.

VIGUIER-PLA, S. (2004): “Factor-based Comparison of k Populations,” *Statistics*, 38(1), 1–15.

WANG, P. (2012): “Large Dimensional Factor Models with a Multi-Level Factor Structure: Identification, Estimation, and Inference,” *Working Paper*.

Technical Appendices

A Assumptions

A.1 Group factor model

Let $\|A\| = \sqrt{\text{tr}(A'A)}$ denote the Frobenius norm of matrix A . Let $k^F = k^c + k_1^s + k_2^s$, and define the k^F -dimensional vector of factors: $F_t = [f_t^{c'}, f_{1,t}^{s'}, f_{2,t}^{s'}]'$, and the (T, k^F) matrix $F = [F_1', \dots, F_T']'$. We make the following assumptions:

Assumption A.1. *The unobservable factor process is such that $F'F/T = \Sigma_F + O_p(1/\sqrt{T})$ as $T \rightarrow \infty$, where Σ_F is a positive definite $(k^F \times k^F)$ matrix defined as:*

$$\Sigma_F = \begin{bmatrix} I_{k^c} & 0 & 0 \\ 0 & I_{k_1^s} & \Phi \\ 0 & \Phi' & I_{k_2^s} \end{bmatrix}.$$

Assumption A.2. *The loadings matrices $\Lambda_j = [\Lambda_j^c; \Lambda_j^s] = [\lambda_{j,1}, \dots, \lambda_{j,N_j}]'$, for $j = 1, 2$, are full column-rank, for N_1, N_2 large enough, and such that:*

$$\frac{\Lambda_j' \Lambda_j}{N_j} = \Sigma_{\Lambda,j} + O\left(\frac{1}{\sqrt{N_j}}\right), \quad j = 1, 2,$$

as $N_j \rightarrow \infty$, where $\Sigma_{\Lambda,j} := \lim_{N_j \rightarrow \infty} \left(\frac{\Lambda_j' \Lambda_j}{N_j}\right)$ is a p.d. (k_j, k_j) matrix, for $j = 1, 2$.

Assumption A.3. *The error terms $(\varepsilon_{1,it} \ \varepsilon_{2,it})'$ are weakly dependent across i and t , and such that $E[\varepsilon_{j,it}] = 0$.*

Assumption A.4. *There exists a constant C_ε such that $E[\varepsilon_{j,it}^4] \leq C_\varepsilon$ for all j, i and t .*

Assumption A.5. *a) The variables F_t and $\varepsilon_{j,is}$ are independent, for all i, j, t and s .*

b) The processes $\{\varepsilon_{j,it}\}$ are stationary, for all j, i .

c) The process $\{F_t\}$ is stationary and weakly dependent over time.

d) For each j and t , as $N_j \rightarrow \infty$, it holds:

$$\frac{1}{\sqrt{N_j}} \sum_{i=1}^{N_j} \lambda_{j,i} \varepsilon_{j,i,t} \xrightarrow{d} N(0, \Omega_j),$$

where $\Omega_j = \lim_{N_j \rightarrow \infty} \frac{1}{N_j} \sum_{i=1}^{N_j} \sum_{\ell=1}^{N_j} \lambda_{j,i} \lambda_{j,\ell}' E[\varepsilon_{j,i,t} \varepsilon_{j,\ell,t}]$.

Assumption A.6. *The asymptotic analysis is for $N_1, N_2, T \rightarrow \infty$ such that $N_2 \leq N_1$, $T/N_2 = o(1)$, $\sqrt{N_1}/T = o(1)$.*

Assumption A.7. *The error terms $\varepsilon_{j,i,t}$ are uncorrelated across j, i and t , and $\varepsilon_{j,i,t} \sim (0, \gamma_{j,ii})$.*

Assumption A.1 requires that the empirical second-order moment matrix of the zero-mean factor process converges at rate $1/\sqrt{T}$ to the population variance-covariance matrix Σ_F . Such converge rate applies, for instance, if the factor process satisfies the conditions for a Central Limit Theorem (CLT) for stationary and weakly dependent data. Positive definiteness of matrix Σ_F is necessary for our model to have exactly $k^c + k_1^s + k_2^s$ pervasive factors. The zero restrictions on the matrix Σ_F , corresponding to the orthogonality of the common and group-specific factors, as well as the identity diagonal blocks, are identification conditions.

Assumption A.2 states that the empirical cross-sectional second-order moment matrix of the loadings converges to its population value at rate $1/\sqrt{N_j}$, for each group $j = 1, 2$. Positive definiteness of matrix $\Sigma_{\Lambda, j}$, for $j = 1, 2$, is also necessary for the existence of exactly $k^c + k_1^s + k_2^s$ pervasive factors. Note that we consider non-random loadings to simplify the assumptions and proofs. If the loadings were random, such a rate of (stochastic) convergence could be obtained with a DGP for the loadings which satisfies the conditions of the CLT for weakly dependent data. Assumptions A.1 and A.2 are similar to conditions used in the large scale factor model literature (see Assumptions A and B in Bai and Ng (2002), Bai (2003), and Bai and Ng (2006a), among others).

Assumption A.3 constrains the amount of admissible time series and cross-sectional dependence of the error terms across different individuals, in the spirit of the assumption - introduced by Chamberlain and Rothschild (1983) - of weak cross-sectional dependence characterizing “approximate factor models”. No distributional assumption is made on the idiosyncratic terms, instead their fourth moments are uniformly bounded in Assumption A.4. Assumptions A.5 a) and b) are standard in factor analysis, but they could be substituted, at the expense of more elaborated proofs, by weak dependence assumptions for factors and idiosyncratic errors analogous to Assumptions D, F.2 and F.4 in Bai (2003). Assumption A.5 c) rules out explosive behavior of the factors. Assumption A.5 d) states that a CLT holds for the error terms scaled by the factor loadings in each group, and is satisfied e.g. by a number of mixing processes. It corresponds to Assumption F.3 in Bai (2003).

Assumption A.6 establishes that in our asymptotic theory both the time-series and the cross-sectional dimensions of each group tend simultaneously to infinity, with constraints on their relative growth rate. The cross-sectional dimensions for both groups grow faster than the time-series dimension, but slower than time-series dimension squared. The latter condition eliminates some bias terms in asymptotic expansions. Assumption A.7 simplifies the derivation of the feasible asymptotic distribution of the statistic used to test the dimension of the common factor space. This condition is stronger than Assumptions A.3 and A.5 b). Moreover, under Assumption A.7, the matrix Ω_j in Assumption A.5 d) simplifies to $\Omega_j = \lim_{N_j \rightarrow \infty} (1/N_j) \sum_{i=1}^{N_j} \lambda_{j,i} \lambda'_{j,i} \gamma_{j,ii}$. We note that, Assumption A.7 simplifies substantially the proof of Theorem 2, but is not needed in the proofs of Theorem 1 and Propositions 3 - 4. Together with Assumption A.5, Assumption A.7 implies a “strict factor model” for each group.

A.2 Mixed frequency factor model

Let $\lambda'_{1,i}$ be the i -th row of the $(N_H, k^C + k^H)$ matrix $\Lambda_1 = [\Lambda_{HC} \ ; \ \Lambda_H]$. We make the following assumption:

Assumption A.8. *The variables $\lambda_{1,i}$ and $e_{m,t}^{i,H}$ are such that:*

$$\frac{1}{\sqrt{N_H}} \sum_{i=1}^{N_H} \lambda_{1,i} e_{m,t}^{i,H} \xrightarrow{d} N(0, \Omega_{\Lambda, m}^*),$$

as $N_H \rightarrow \infty$, where

$$\Omega_{\Lambda, m}^* = \lim_{N_H \rightarrow \infty} \frac{1}{N_H} \sum_{i=1}^{N_H} \sum_{\ell=1}^{N_H} \lambda_{1,i} \lambda'_{1,\ell} Cov(e_{m,t}^{i,H}, e_{m,t}^{\ell,H}), \quad m = 1, \dots, M.$$

Assumption A.8 is analogous to Assumption A.5 d) expressed for the high frequency DGP of the idiosyncratic innovation terms $e_{m,t}^{i,H}$.

B Proofs of Propositions

B.1 Proof of Proposition 1

By replacing equation (2.6) into model (2.4), we get

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} \Lambda_1^c A_{11} + \Lambda_1^s A_{21} & \Lambda_1^c A_{12} + \Lambda_1^s A_{22} & \Lambda_1^c A_{13} + \Lambda_1^s A_{23} \\ \Lambda_2^c A_{11} + \Lambda_2^s A_{31} & \Lambda_2^c A_{12} + \Lambda_2^s A_{32} & \Lambda_2^c A_{13} + \Lambda_2^s A_{33} \end{bmatrix} \begin{bmatrix} \tilde{f}_t^c \\ \tilde{f}_{1,t}^s \\ \tilde{f}_{2,t}^s \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix}.$$

This factor model satisfies the restrictions in the loading matrix appearing in equation (2.4) if, and only if, $\Lambda_1^c A_{13} + \Lambda_1^s A_{23} = 0$, and $\Lambda_2^c A_{12} + \Lambda_2^s A_{32} = 0$, which can be written as linear homogeneous systems of equations for the elements of matrices $[A'_{13} \ A'_{23}]'$ and $[A'_{12} \ A'_{32}]'$:

$$\begin{bmatrix} \Lambda_1^c \\ \Lambda_1^s \end{bmatrix} \begin{bmatrix} A_{13} \\ A_{23} \end{bmatrix} = 0, \quad \text{and} \quad \begin{bmatrix} \Lambda_2^c \\ \Lambda_2^s \end{bmatrix} \begin{bmatrix} A_{12} \\ A_{32} \end{bmatrix} = 0.$$

Since $\begin{bmatrix} \Lambda_1^c \\ \Lambda_1^s \end{bmatrix}$ and $\begin{bmatrix} \Lambda_2^c \\ \Lambda_2^s \end{bmatrix}$ are full column rank, it follows that $A_{13} = 0$, $A_{23} = 0$, $A_{12} = 0$, and $A_{32} = 0$. Therefore, the transformation of the factors that is compatible with the restrictions on the loading matrix in equation (2.4) is:

$$\begin{bmatrix} f_t^c \\ f_{1,t}^s \\ f_{2,t}^s \end{bmatrix} = \begin{bmatrix} A_{11} & 0 & 0 \\ A_{21} & A_{22} & 0 \\ A_{31} & 0 & A_{33} \end{bmatrix} \begin{bmatrix} \tilde{f}_t^c \\ \tilde{f}_{1,t}^s \\ \tilde{f}_{2,t}^s \end{bmatrix}.$$

We can invert this transformation and write:

$$\tilde{f}_t^c = A_{11}^{-1} f_t^c, \quad \tilde{f}_{1,t}^s = A_{22}^{-1} f_{1,t}^s - A_{22}^{-1} A_{21} A_{11}^{-1} f_t^c, \quad \tilde{f}_{2,t}^s = A_{33}^{-1} f_{2,t}^s - A_{33}^{-1} A_{31} A_{11}^{-1} f_t^c.$$

The transformed factors satisfy the normalization restrictions in (2.5) if, and only if,

$$\text{Cov}(\tilde{f}_{1,t}^s, \tilde{f}_t^c) = -A_{22}^{-1} A_{21} A_{11}^{-1} (A_{11}^{-1})' = 0, \quad (\text{B.1})$$

$$\text{Cov}(\tilde{f}_{2,t}^s, \tilde{f}_t^c) = -A_{33}^{-1} A_{31} A_{11}^{-1} (A_{11}^{-1})' = 0, \quad (\text{B.2})$$

$$V(\tilde{f}_t^c) = A_{11}^{-1} (A_{11}^{-1})' = I_{k^c}, \quad (\text{B.3})$$

$$V(\tilde{f}_{1,t}^s) = A_{22}^{-1} (A_{22}^{-1})' + A_{22}^{-1} A_{21} A_{11}^{-1} (A_{11}^{-1})' A_{21}' (A_{22}^{-1})' = I_{k_1^s}, \quad (\text{B.4})$$

$$V(\tilde{f}_{2,t}^s) = A_{33}^{-1} (A_{33}^{-1})' + A_{33}^{-1} A_{31} A_{11}^{-1} (A_{11}^{-1})' A_{31}' (A_{33}^{-1})' = I_{k_2^s}. \quad (\text{B.5})$$

Since the matrices A_{11} , A_{22} and A_{33} are nonsingular, equations (B.1) and (B.2) imply $A_{21} = 0$, and $A_{31} = 0$. Then, from equations (B.3) - (B.5), we get that matrices A_{11} , A_{22} and A_{33} are orthogonal.

Q.E.D.

B.2 Proof of Proposition 2

From equation (2.5) we have

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

Matrix R is block diagonal, and the upper-left block I_{k^c} has eigenvalue 1 with multiplicity k^c . The associated eigenspace is $\{(\xi', 0)'\}$, $\xi \in \mathbb{R}^{k^c}$. The lower-right block $\Phi\Phi'$ is a positive semi-definite matrix, and its largest eigenvalue is $\tilde{\rho}^2$, where $\tilde{\rho}^2 = \sup \{\xi_1' \Phi\Phi' \xi_1 : \xi_1 \in \mathbb{R}^{k_1^s}, \|\xi_1\| = 1\} < 1$ is the first squared canonical correlation of vectors $f_{1,t}^s$ and $f_{2,t}^s$. Therefore, we deduce that the largest eigenvalue of matrix R is equal to 1, with multiplicity k^c , and the associated eigenspace,

denoted by \mathcal{E}_c , is spanned by vectors $(\xi', 0)'$, with $\xi \in \mathbb{R}^{k^c}$. Let S_1 be an orthogonal (k^c, k^c) matrix, then the columns of the (k_1, k^c) matrix

$$W_1 = \begin{pmatrix} S_1 \\ 0_{k_1^s \times k^c} \end{pmatrix}$$

are an orthonormal basis of the eigenspace \mathcal{E}_c . We have:

$$W_1' h_{1,t} = S_1' f_t^c. \quad (\text{B.6})$$

Analogous arguments allow to show that the largest eigenvalue of matrix R^* is equal to 1, with multiplicity k^c and that the associated eigenspace, denoted by \mathcal{E}_c^* , is spanned by vectors $(\xi^{*'}, 0)'$, with $\xi^* \in \mathbb{R}^{k^c}$. We have $\mathcal{E}_c^* = \mathcal{E}_c$. Let S_2 be an orthogonal (k^c, k^c) matrix. Then, the columns of the (k_2, k^c) matrix

$$W_2 = \begin{pmatrix} S_2 \\ 0_{k_2^s \times k^c} \end{pmatrix}$$

are an orthonormal basis of the eigenspace \mathcal{E}_c^* . We have:

$$W_2' h_{2,t} = S_2' f_t^c, \quad (\text{B.7})$$

which yields parts *i*) and *ii*).

When there is no common factor, the matrix R becomes $R = \Phi\Phi'$, and matrix R^* becomes $R^* = \Phi'\Phi$. By the above arguments, the largest eigenvalue of matrix R , which is equal to the largest eigenvalue of matrix R^* , is not larger than $\tilde{\rho}^2$, where $\tilde{\rho}^2 < 1$ is the first squared canonical correlation between the two group-specific factors. This yields part *iii*).

Finally, we prove part *iv*). We showed that the lower-right block $\Phi\Phi'$ of matrix R is a positive semi-definite matrix and all its $k_1^s = k_1 - k^c$ eigenvalues are strictly smaller than one. These are also eigenvalues of matrix R . Let us denote the space spanned by the associated k_1^s eigenvectors of matrix R by $\mathcal{E}_{s,1}$. This space is spanned by vectors $(0', \tilde{\xi}')'$ with $\tilde{\xi} \in \mathbb{R}^{k_1^s}$. We note that, by construction, the vectors $(0', \tilde{\xi}')'$ are linearly independent of the vectors $(\xi', 0)'$ spanning the eigenspace \mathcal{E}_c . Let Q_1 be an orthogonal (k_1^s, k_1^s) matrix, then the columns of matrix

$$W_1^s = \begin{pmatrix} 0_{k^c \times k_1^s} \\ Q_1 \end{pmatrix}$$

are an orthonormal basis of the eigenspace $\mathcal{E}_{s,1}$. We have: $W_1^{s'} h_{1,t} = Q_1' f_{1,t}^s$.

Analogously, we have that the lower-right block $\Phi'\Phi$ of matrix R^* is a positive semi-definite matrix and all its $k_2^s = k_2 - k^c$ eigenvalues are strictly smaller than one. These are also eigenvalues of matrix R^* . Let us denote the space spanned by the associated k_2^s eigenvectors of matrix R^* by $\mathcal{E}_{s,2}$. This space is spanned by vectors $(0', \tilde{\xi}^{*'})'$ with $\tilde{\xi}^* \in \mathbb{R}^{k_2^s}$. We note that, by construction, the vectors $(0', \tilde{\xi}^{*'})'$ are linearly independent of the vectors $(\xi^{*'}, 0)'$ spanning the eigenspace \mathcal{E}_c^* . Let Q_2 be an orthogonal (k_2^s, k_2^s) matrix, then the columns of matrix

$$W_2^s = \begin{pmatrix} 0_{k^c \times k_2^s} \\ Q_2 \end{pmatrix}$$

are an orthonormal basis of the eigenspace $\mathcal{E}_{s,2}$. We have $W_2^{s'} h_{2,t} = Q_2' f_{2,t}^s$. *Q.E.D.*

B.3 Proof of Theorem 1

The Proof of Theorem 1 is structured as follows. We start by deriving an asymptotic expansion of matrix \hat{R} (Subsection B.3.1). This result is used to obtain the asymptotic expansions of the eigenvalues and eigenvectors of matrix \hat{R} by perturbation methods (Subsections B.3.2 and B.3.3), of the canonical correlations, and of the test statistic $\hat{\xi}(k^c)$ (Subsection B.3.4). Finally, the asymptotic distribution of the test statistic follows by applying a suitable CLT (Subsection B.3.5).

B.3.1 Asymptotic expansion of \hat{R}

The canonical correlations and the canonical directions are invariant to one-to-one transformations of the vectors $\hat{h}_{1,t}$ and $\hat{h}_{2,t}$ (see, among others, Anderson (2003)). Therefore, without loss of generality, for 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 approximation (4.1). Moreover, under Assumption A.6 the bias term is negligible, and we get:

$$\hat{h}_{j,t} \simeq h_{j,t} + \frac{1}{\sqrt{N_j}} u_{j,t}, \quad j = 1, 2. \quad (\text{B.8})$$

By using approximation (B.8), and $N_2 = N$, $N_1 = N/\mu_N^2$, we have:

$$\hat{V}_{12} = \frac{1}{T} \sum_{t=1}^T \hat{h}_{1,t} \hat{h}'_{2,t} \simeq \frac{1}{T} \sum_{t=1}^T \left(h_{1,t} + \frac{1}{\sqrt{N}} \mu_N u_{1,t} \right) \left(h_{2,t} + \frac{1}{\sqrt{N}} u_{2,t} \right)' = \tilde{V}_{12} + \hat{X}_{12},$$

where:

$$\tilde{V}_{12} = \frac{1}{T} \sum_{t=1}^T h_{1,t} h'_{2,t}, \quad \hat{X}_{12} = \frac{1}{T\sqrt{N}} \sum_{t=1}^T (h_{1,t} u'_{2,t} + \mu_N u_{1,t} h'_{2,t}) + \frac{\mu_N}{TN} \sum_{t=1}^T u_{1,t} u'_{2,t}. \quad (\text{B.9})$$

Similarly:

$$\hat{V}_{jj} = \frac{1}{T} \sum_{t=1}^T \hat{h}_{j,t} \hat{h}'_{j,t} \simeq \frac{1}{T} \sum_{t=1}^T \left(h_{j,t} + \frac{1}{\sqrt{N_j}} u_{j,t} \right) \left(h_{j,t} + \frac{1}{\sqrt{N_j}} u_{j,t} \right)' = \tilde{V}_{jj} \left(Id + \tilde{V}_{jj}^{-1} \hat{X}_{jj} \right), \quad j = 1, 2,$$

where:

$$\tilde{V}_{jj} = \frac{1}{T} \sum_{t=1}^T h_{j,t} h'_{j,t}, \quad j = 1, 2,$$

$$\hat{X}_{11} = \frac{\mu_N}{T\sqrt{N}} \sum_{t=1}^T (h_{1,t} u'_{1,t} + u_{1,t} h'_{1,t}) + \frac{\mu_N^2}{TN} \sum_{t=1}^T u_{1,t} u'_{1,t}, \quad (\text{B.10})$$

$$\hat{X}_{22} = \frac{1}{T\sqrt{N}} \sum_{t=1}^T (h_{2,t} u'_{2,t} + u_{2,t} h'_{2,t}) + \frac{1}{TN} \sum_{t=1}^T u_{2,t} u'_{2,t}. \quad (\text{B.11})$$

Therefore, from (3.1) we get:

$$\hat{R} \simeq \left(Id + \tilde{V}_{11}^{-1} \hat{X}_{11} \right)^{-1} \tilde{V}_{11}^{-1} \left(\tilde{V}_{12} + \hat{X}_{12} \right) \left(Id + \tilde{V}_{22}^{-1} \hat{X}_{22} \right)^{-1} \tilde{V}_{22}^{-1} \left(\tilde{V}_{21} + \hat{X}_{21} \right).$$

Using $\frac{1}{\sqrt{T}} \sum_{t=1}^T h_{j,t} u'_{k,t} = O_p(1)$ we have $\hat{X}_{j,k} = O_p\left(\frac{1}{\sqrt{NT}}\right)$. Let us expand matrix \hat{R} at first order in the $\hat{X}_{j,k}$. By using $(Id + X)^{-1} \simeq Id - X$ for $X \simeq 0$, we have:

$$\begin{aligned} \hat{R} &\simeq \left(Id - \tilde{V}_{11}^{-1} \hat{X}_{11} \right) \tilde{V}_{11}^{-1} \left(\tilde{V}_{12} + \hat{X}_{12} \right) \left(Id - \tilde{V}_{22}^{-1} \hat{X}_{22} \right) \tilde{V}_{22}^{-1} \left(\tilde{V}_{21} + \hat{X}_{21} \right) \\ &\simeq \tilde{V}_{11}^{-1} \tilde{V}_{12} \tilde{V}_{22}^{-1} \tilde{V}_{21} \\ &\quad - \tilde{V}_{11}^{-1} \hat{X}_{11} \tilde{V}_{11}^{-1} \tilde{V}_{12} \tilde{V}_{22}^{-1} \tilde{V}_{21} + \tilde{V}_{11}^{-1} \hat{X}_{12} \tilde{V}_{22}^{-1} \tilde{V}_{21} - \tilde{V}_{11}^{-1} \tilde{V}_{12} \tilde{V}_{22}^{-1} \hat{X}_{22} \tilde{V}_{22}^{-1} \tilde{V}_{21} + \tilde{V}_{11}^{-1} \tilde{V}_{12} \tilde{V}_{22}^{-1} \hat{X}_{21}. \end{aligned}$$

Defining the following quantities:

$$\begin{aligned} \tilde{A} &= \tilde{V}_{11}^{-1} \tilde{V}_{12}, \quad \tilde{B} = \tilde{V}_{22}^{-1} \tilde{V}_{21}, \quad \tilde{R} = \tilde{V}_{11}^{-1} \tilde{V}_{12} \tilde{V}_{22}^{-1} \tilde{V}_{21} = \tilde{A} \tilde{B}, \\ \hat{\Psi}^* &= -\hat{X}_{11} \tilde{R} + \hat{X}_{12} \tilde{B} - \tilde{B}' \hat{X}_{22} \tilde{B} + \tilde{B}' \hat{X}_{21}, \end{aligned} \quad (\text{B.12})$$

$$\hat{\Psi} = \tilde{V}_{11}^{-1} \hat{\Psi}^*, \quad (\text{B.13})$$

we get the asymptotic expansion of matrix \hat{R} :

$$\hat{R} = \tilde{R} + \hat{\Psi} + O_p\left(\frac{1}{NT}\right). \quad (\text{B.14})$$

B.3.2 Matrix \tilde{R} and its eigenvalues and eigenvectors

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

$$\begin{aligned} \tilde{\Sigma}_{cc} &= \frac{1}{T} \sum_{t=1}^T f_t^c f_t^{c'}, & \tilde{\Sigma}_{cj} &= \frac{1}{T} \sum_{t=1}^T f_t^c f_{j,t}^{s'}, & \tilde{\Sigma}_{jc} &= \tilde{\Sigma}_{cj}', & j &= 1, 2, \\ \tilde{\Sigma}_{jj} &= \frac{1}{T} \sum_{t=1}^T f_{j,t}^s f_{j,t}^{s'}, & j &= 1, 2, & \tilde{\Sigma}_{12} &= \frac{1}{T} \sum_{t=1}^T f_{1,t}^s f_{2,t}^{s'}, \end{aligned}$$

we can write matrices \tilde{V}_{jj} , with $j = 1, 2$, and \tilde{V}_{12} as:

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

By matrix algebra we get:

$$\tilde{V}_{11}^{-1} = \begin{bmatrix} \tilde{\Sigma}_{*,1}^{-1} & -\tilde{\Sigma}_{*,1}^{-1} \tilde{\Sigma}_{c1} \tilde{\Sigma}_{11}^{-1} \\ -\tilde{\Sigma}_{11}^{-1} \tilde{\Sigma}_{1c} \tilde{\Sigma}_{*,1}^{-1} & \tilde{\Sigma}_{11}^{-1} + \tilde{\Sigma}_{11}^{-1} \tilde{\Sigma}_{1c} \tilde{\Sigma}_{*,1}^{-1} \tilde{\Sigma}_{c1} \tilde{\Sigma}_{11}^{-1} \end{bmatrix}, \quad (\text{B.15})$$

where $\tilde{\Sigma}_{*,1} = \tilde{\Sigma}_{cc} - \tilde{\Sigma}_{c1} \tilde{\Sigma}_{11}^{-1} \tilde{\Sigma}_{1c}$. From Assumption A.1, we have:

$$\begin{aligned} \tilde{\Sigma}_{c1} &= O_p(1/\sqrt{T}), & \tilde{\Sigma}_{cc} &= I_{k^c} + O_p(1/\sqrt{T}), \\ \tilde{\Sigma}_{jj} &= I_{k_j^s} + O_p(1/\sqrt{T}), & j &= 1, 2, & \tilde{\Sigma}_{12} &= \Phi + O_p(1/\sqrt{T}), \end{aligned}$$

which imply:

$$\tilde{\Sigma}_{*,1} = \tilde{\Sigma}_{cc} + O_p(1/T), \quad (\text{B.16})$$

$$\tilde{\Sigma}_{*,1}^{-1} = \tilde{\Sigma}_{cc}^{-1} + O_p(1/T), \quad (\text{B.17})$$

$$-\tilde{\Sigma}_{*,1}^{-1} \tilde{\Sigma}_{c1} \tilde{\Sigma}_{11}^{-1} = -\tilde{\Sigma}_{cc}^{-1} \tilde{\Sigma}_{c1} \tilde{\Sigma}_{11}^{-1} + O_p(1/T) = -\tilde{\Sigma}_{c1} + O_p(1/T), \quad (\text{B.18})$$

$$\tilde{\Sigma}_{11}^{-1} \tilde{\Sigma}_{1c} \tilde{\Sigma}_{*,1}^{-1} \tilde{\Sigma}_{c1} \tilde{\Sigma}_{11}^{-1} = O_p(1/T). \quad (\text{B.19})$$

Substituting results (B.16) - (B.19) into equation (B.15) we get:

$$\tilde{V}_{11}^{-1} = \begin{bmatrix} \tilde{\Sigma}_{cc}^{-1} & -\tilde{\Sigma}_{c1} \\ -\tilde{\Sigma}_{1c} & \tilde{\Sigma}_{11}^{-1} \end{bmatrix} + O_p(1/T). \quad (\text{B.20})$$

Equation (B.15) allows to compute \tilde{A} :

$$\tilde{A} = \tilde{V}_{11}^{-1} \tilde{V}_{12} = \begin{bmatrix} \tilde{\Sigma}_{*,1}^{-1} & -\tilde{\Sigma}_{*,1}^{-1} \tilde{\Sigma}_{c1} \tilde{\Sigma}_{11}^{-1} \\ -\tilde{\Sigma}_{11}^{-1} \tilde{\Sigma}_{1c} \tilde{\Sigma}_{*,1}^{-1} & \tilde{\Sigma}_{11}^{-1} + \tilde{\Sigma}_{11}^{-1} \tilde{\Sigma}_{1c} \tilde{\Sigma}_{*,1}^{-1} \tilde{\Sigma}_{c1} \tilde{\Sigma}_{11}^{-1} \end{bmatrix} \begin{bmatrix} \tilde{\Sigma}_{cc} & \tilde{\Sigma}_{c2} \\ \tilde{\Sigma}_{1c} & \tilde{\Sigma}_{12} \end{bmatrix} = \begin{bmatrix} I_{k^c} & \tilde{A}_{cs} \\ 0 & \tilde{A}_{ss} \end{bmatrix}, \quad (\text{B.21})$$

where:

$$\begin{aligned}\tilde{A}_{cs} &= \tilde{\Sigma}_{*1}^{-1}\tilde{\Sigma}_{c2} - \tilde{\Sigma}_{*1}^{-1}\tilde{\Sigma}_{c1}\tilde{\Sigma}_{11}^{-1}\tilde{\Sigma}_{12} = O_p\left(\frac{1}{\sqrt{T}}\right), \\ \tilde{A}_{ss} &= -\tilde{\Sigma}_{11}^{-1}\tilde{\Sigma}_{1c}\tilde{\Sigma}_{*1}^{-1}\tilde{\Sigma}_{c2} + \tilde{\Sigma}_{11}^{-1}\tilde{\Sigma}_{12} + \tilde{\Sigma}_{11}^{-1}\tilde{\Sigma}_{1c}\tilde{\Sigma}_{*1}^{-1}\tilde{\Sigma}_{c1}\tilde{\Sigma}_{11}^{-1}\tilde{\Sigma}_{12} \\ &= \tilde{\Sigma}_{11}^{-1}\tilde{\Sigma}_{12} + O_p\left(\frac{1}{T}\right) = \Phi + O_p\left(\frac{1}{\sqrt{T}}\right).\end{aligned}$$

Let us compute:

$$\tilde{V}_{22}^{-1} = \begin{bmatrix} \tilde{\Sigma}_{*2}^{-1} & -\tilde{\Sigma}_{*2}^{-1}\tilde{\Sigma}_{c2}\tilde{\Sigma}_{22}^{-1} \\ -\tilde{\Sigma}_{22}^{-1}\tilde{\Sigma}_{2c}\tilde{\Sigma}_{*2}^{-1} & \tilde{\Sigma}_{22}^{-1} + \tilde{\Sigma}_{22}^{-1}\tilde{\Sigma}_{2c}\tilde{\Sigma}_{*2}^{-1}\tilde{\Sigma}_{c2}\tilde{\Sigma}_{22}^{-1} \end{bmatrix}, \quad (\text{B.22})$$

where $\tilde{\Sigma}_{*2} = \tilde{\Sigma}_{cc} - \tilde{\Sigma}_{c2}\tilde{\Sigma}_{22}^{-1}\tilde{\Sigma}_{2c}$. Equation (B.22) allows to compute \tilde{B} :

$$\tilde{B} = \tilde{V}_{22}^{-1}\tilde{V}_{21} = \begin{bmatrix} \tilde{\Sigma}_{*2}^{-1} & -\tilde{\Sigma}_{*2}^{-1}\tilde{\Sigma}_{c2}\tilde{\Sigma}_{22}^{-1} \\ -\tilde{\Sigma}_{22}^{-1}\tilde{\Sigma}_{2c}\tilde{\Sigma}_{*2}^{-1} & \tilde{\Sigma}_{22}^{-1} + \tilde{\Sigma}_{22}^{-1}\tilde{\Sigma}_{2c}\tilde{\Sigma}_{*2}^{-1}\tilde{\Sigma}_{c2}\tilde{\Sigma}_{22}^{-1} \end{bmatrix} \begin{bmatrix} \tilde{\Sigma}_{cc} & \tilde{\Sigma}_{c1} \\ \tilde{\Sigma}_{2c} & \tilde{\Sigma}_{21} \end{bmatrix} = \begin{bmatrix} I_{k^c} & \tilde{B}_{cs} \\ 0 & \tilde{B}_{ss} \end{bmatrix}, \quad (\text{B.23})$$

where:

$$\begin{aligned}\tilde{B}_{cs} &= \tilde{\Sigma}_{*2}^{-1}\tilde{\Sigma}_{c1} - \tilde{\Sigma}_{*2}^{-1}\tilde{\Sigma}_{c2}\tilde{\Sigma}_{22}^{-1}\tilde{\Sigma}_{21} = O_p\left(\frac{1}{\sqrt{T}}\right), \\ \tilde{B}_{ss} &= -\tilde{\Sigma}_{22}^{-1}\tilde{\Sigma}_{2c}\tilde{\Sigma}_{*2}^{-1}\tilde{\Sigma}_{c1} + \tilde{\Sigma}_{22}^{-1}\tilde{\Sigma}_{21} + \tilde{\Sigma}_{22}^{-1}\tilde{\Sigma}_{2c}\tilde{\Sigma}_{*2}^{-1}\tilde{\Sigma}_{c2}\tilde{\Sigma}_{22}^{-1}\tilde{\Sigma}_{21} \\ &= \tilde{\Sigma}_{22}^{-1}\tilde{\Sigma}_{21} + O_p\left(\frac{1}{T}\right) = \Phi' + O_p\left(\frac{1}{\sqrt{T}}\right).\end{aligned}$$

Finally, using results (B.21) and (B.23) we can compute:

$$\tilde{R} = \tilde{A}\tilde{B} = \begin{bmatrix} I_{k^c} & \tilde{A}_{cs} \\ 0 & \tilde{A}_{ss} \end{bmatrix} \begin{bmatrix} I_{k^c} & \tilde{B}_{cs} \\ 0 & \tilde{B}_{ss} \end{bmatrix} = \begin{bmatrix} I_{k^c} & \tilde{R}_{cs} \\ 0 & \tilde{R}_{ss} \end{bmatrix},$$

where

$$\begin{aligned}\tilde{R}_{cs} &= \tilde{B}_{cs} + \tilde{A}_{cs}\tilde{B}_{ss} = O_p(1/\sqrt{T}), \\ \tilde{R}_{ss} &= \tilde{A}_{ss}\tilde{B}_{ss} = \tilde{\Sigma}_{11}^{-1}\tilde{\Sigma}_{12}\tilde{\Sigma}_{22}^{-1}\tilde{\Sigma}_{21} + O_p(1/T) = \Phi\Phi' + O_p(1/\sqrt{T}).\end{aligned}$$

The eigenvalues of matrix \tilde{R} are $\hat{\rho}_1^2 = \dots = \hat{\rho}_{k^c}^2 = 1 > \hat{\rho}_{k^c+1}^2 \geq \dots \geq \hat{\rho}_{k_1}^2$. The eigenvectors associated with the first k^c eigenvalues are spanned by the columns of matrix:

$$E_c = \begin{bmatrix} I_{k^c} \\ 0 \end{bmatrix}. \quad (\text{B.24})$$

Define:

$$E_s = \begin{bmatrix} 0 \\ I_{k_1-k^c} \end{bmatrix}. \quad (\text{B.25})$$

We note $I_{k_1} = \begin{bmatrix} E_c & \vdots & E_s \end{bmatrix}$, so that the columns of matrices E_c and E_s span the space \mathbb{R}^{k_1} .

B.3.3 Perturbation of the eigenvalues and eigenvectors of matrix \hat{R}

The estimators of the first k^c canonical correlations are such that $\hat{\rho}_\ell^2$, with $\ell = 1, \dots, k^c$ are the k^c largest eigenvalues of matrix \hat{R} . Using (B.14), we now derive their asymptotic expansion using perturbations arguments. Under the null hypothesis $H(k^c)$, 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 $\ell = 1, \dots, k^c$. We have:

$$\hat{R}\hat{W}_1^* = \hat{W}_1^*\hat{\Lambda}, \quad (\text{B.26})$$

where $\hat{\Lambda} = \text{diag}(\hat{\rho}_\ell^2, \ell = 1, \dots, k^c)$ 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 \hat{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:

$$\hat{W}_1^* = E_c \hat{U} + E_s \alpha, \quad (\text{B.27})$$

$$\hat{\Lambda} = I_{k^c} + \hat{M}, \quad (\text{B.28})$$

where E_c and E_s are defined in equations (B.24) and (B.25), \hat{U} is a (k^c, k^c) nonsingular matrix, $\hat{M} = \text{diag}(\hat{\mu}_1, \dots, \hat{\mu}_{k^c})$, and α is a $(k_1 - k^c, k^c)$ matrix, with $\alpha, \hat{\mu}_1, \dots, \hat{\mu}_{k^c}$ converging to zero as $N_1, N_2, T \rightarrow \infty$. Substituting the expansions in equations (B.14) and (B.26) we get:

$$(\tilde{R} + \hat{\Psi})(E_c \hat{U} + E_s \alpha) \simeq (E_c \hat{U} + E_s \alpha)(I_{k^c} + \hat{M}).$$

By using $\tilde{R}E_c = E_c$, and keeping only the terms at first order, we get:

$$\tilde{R}E_s \alpha + \hat{\Psi}E_c \hat{U} \simeq E_s \alpha + E_c \hat{U} \hat{M}. \quad (\text{B.29})$$

Pre-multiplying equation (B.29) by E'_c , we get:

$$E'_c \tilde{R}E_s \alpha + E'_c \hat{\Psi}E_c \hat{U} \simeq \hat{U} \hat{M} \Leftrightarrow \hat{M} \simeq \hat{U}^{-1} \left(\tilde{R}_{cs} \alpha + \hat{\Psi}_{cc} \hat{U} \right), \quad (\text{B.30})$$

where we use the fact that \hat{U} is non-singular, $\hat{\Psi}_{cc} = E'_c \hat{\Psi} E_c$ and $\tilde{R}_{cs} = E'_c \tilde{R} E_s$. Pre-multiplying equation (B.29) by E'_s , we get:

$$E'_s \tilde{R}E_s \alpha + E'_s \hat{\Psi}E_c \hat{U} \simeq \alpha \Leftrightarrow \alpha \simeq \tilde{R}_{ss} \alpha + \hat{\Psi}_{sc} \hat{U},$$

where $\hat{\Psi}_{sc} = E'_s \hat{\Psi} E_c$. This implies:

$$\alpha \simeq (I_{k_1 - k^c} - \tilde{R}_{ss})^{-1} \hat{\Psi}_{sc} \hat{U}. \quad (\text{B.31})$$

Substituting the first order approximation of α from equation (B.31) into equation (B.27) we get:

$$\hat{W}_1^* \simeq \left(E_c + E_s (I_{k_1 - k^c} - \tilde{R}_{ss})^{-1} \hat{\Psi}_{sc} \right) \hat{U}. \quad (\text{B.32})$$

The normalized eigenvectors corresponding to the canonical directions are:

$$\hat{W}_1 = \hat{W}_1^* \cdot \text{diag}(\hat{W}_1^* \hat{V}_{11} \hat{W}_1^*)^{-1/2}. \quad (\text{B.33})$$

Substituting the first order approximation of α from equation (B.31) into (B.30), we get the first order approximation of matrix \hat{M} :

$$\hat{M} \simeq \hat{U}^{-1} \left(\hat{\Psi}_{cc} + \tilde{R}_{cs} (I_{k_1 - k^c} - \tilde{R}_{ss})^{-1} \hat{\Psi}_{sc} \right) \hat{U}. \quad (\text{B.34})$$

Substituting the first order approximation of \hat{M} from equation (B.34) into (B.28), matrix $\hat{\Lambda}$ can be approximated as:

$$\hat{\Lambda} \simeq I_{k^c} + \hat{U}^{-1} \left(\hat{\Psi}_{cc} + \tilde{R}_{cs} (I_{k_1 - k^c} - \tilde{R}_{ss})^{-1} \hat{\Psi}_{sc} \right) \hat{U}.$$

Note that this first order approximation holds for the terms in the main diagonal, as matrix $\hat{\Lambda}$ has been defined to be diagonal, and the out-of-diagonal terms are of higher order. Up to higher order terms we have:

$$\hat{\Lambda}^{1/2} \simeq I_{k^c} + \frac{1}{2}\hat{\mathcal{U}}^{-1} \left[\hat{\Psi}_{cc} + \tilde{R}_{cs}(I_{k_1-k^c} - \tilde{R}_{ss})^{-1}\hat{\Psi}_{sc} \right] \hat{\mathcal{U}},$$

which implies:

$$\begin{aligned} \sum_{\ell=1}^{k^c} \hat{\rho}_\ell &= \text{tr}(\hat{\Lambda}^{1/2}) = k^c + \frac{1}{2}\text{tr} \left[\hat{\mathcal{U}}^{-1} \left(\hat{\Psi}_{cc} + \tilde{R}_{cs}(I_{k_1-k^c} - \tilde{R}_{ss})^{-1}\hat{\Psi}_{sc} \right) \hat{\mathcal{U}} \right] + O_p \left(\frac{1}{NT} \right), \\ &= k^c + \frac{1}{2}\text{tr} \left[\hat{\Psi}_{cc} + \tilde{R}_{cs}(I_{k_1-k^c} - \tilde{R}_{ss})^{-1}\hat{\Psi}_{sc} \right] + O_p \left(\frac{1}{NT} \right), \end{aligned} \quad (\text{B.35})$$

by the commutative property of the trace.

B.3.4 Asymptotic expansion of $\sum_{\ell=1}^{k^c} \hat{\rho}_\ell$

Let us now derive an asymptotic expansion for the sum of the canonical correlations $\sum_{\ell=1}^{k^c} \hat{\rho}_\ell$. From equation (B.35), and using (B.13), we get:

$$\begin{aligned} \sum_{\ell=1}^{k^c} \hat{\rho}_\ell &= k^c + \frac{1}{2}\text{tr} \left\{ \left[I_{k^c} \ : \ \tilde{R}_{cs}(I_{(k_1-k^c)} - \tilde{R}_{ss})^{-1} \right] \hat{\Psi} E_c \right\} + O_p \left(\frac{1}{NT} \right) \\ &= k^c + \frac{1}{2}\text{tr} \left\{ \left[I_{k^c} \ : \ \tilde{R}_{cs}(I_{(k_1-k^c)} - \tilde{R}_{ss})^{-1} \right] \tilde{V}_{11}^{-1} \begin{bmatrix} \hat{\Psi}_{cc}^* \\ \hat{\Psi}_{sc}^* \end{bmatrix} \right\} + O_p \left(\frac{1}{NT} \right), \end{aligned} \quad (\text{B.36})$$

where $\hat{\Psi}_{cc}^*$ and $\hat{\Psi}_{sc}^*$ are blocks of matrix $\hat{\Psi}^*$. In order to further simplify this asymptotic expansion, let us consider the following matrix:

$$\left[I_{k^c} \ : \ \tilde{R}_{cs}(I_{(k_1-k^c)} - \tilde{R}_{ss})^{-1} \right] \tilde{V}_{11}^{-1}.$$

Using equation (B.20), we have:

$$\begin{aligned} \left[I_{k^c} \ : \ \tilde{R}_{cs}(I_{k_1-k^c} - \tilde{R}_{ss})^{-1} \right] \tilde{V}_{11}^{-1} &= \left[I_{k^c} \ : \ \tilde{R}_{cs}(I_{k_1-k^c} - \tilde{R}_{ss})^{-1} \right] \begin{bmatrix} \tilde{\Sigma}_{cc}^{-1} & -\tilde{\Sigma}_{c1} \\ -\tilde{\Sigma}_{1c} & \tilde{\Sigma}_{11}^{-1} \end{bmatrix} + O_p \left(\frac{1}{T} \right) \\ &= \left[\tilde{\Sigma}_{cc}^{-1} - \tilde{R}_{cs}(I_{k_1-k^c} - \tilde{R}_{ss})^{-1}\tilde{\Sigma}_{1c} \ : \ -\tilde{\Sigma}_{c1} + \tilde{R}_{cs}(I_{k_1-k^c} - \tilde{R}_{ss})^{-1}\tilde{\Sigma}_{11}^{-1} \right] + O_p \left(\frac{1}{T} \right) \\ &= \left[\tilde{\Sigma}_{cc}^{-1} \ : \ -\tilde{\Sigma}_{c1} + \tilde{R}_{cs}(I_{k_1-k^c} - \tilde{R}_{ss})^{-1} \right] + O_p \left(\frac{1}{T} \right), \end{aligned} \quad (\text{B.37})$$

where the last equality follows from the bounds $\tilde{R}_{cs} = O_p(1/\sqrt{T})$, $\tilde{\Sigma}_{1c} = O_p(1/\sqrt{T})$ and $\tilde{\Sigma}_{11} = I_{k_1} + O_p(1/\sqrt{T})$. Note that equation (B.37) can be further simplified by considering the asymptotic expansion of term \tilde{R}_{cs} . Let us consider the different terms in the equations of \tilde{R}_{cs} and \tilde{R}_{ss} :

$$\tilde{R}_{cs} = \tilde{B}_{cs} + \tilde{A}_{cs}\tilde{B}_{ss}, \quad \tilde{R}_{ss} = \tilde{A}_{ss}\tilde{B}_{ss}, \quad (\text{B.38})$$

where:

$$\tilde{A}_{cs} = \tilde{\Sigma}_{*1}^{-1}\tilde{\Sigma}_{c2} - \tilde{\Sigma}_{*1}^{-1}\tilde{\Sigma}_{c1}\tilde{\Sigma}_{11}^{-1}\tilde{\Sigma}_{12}, \quad \tilde{A}_{ss} = \tilde{\Sigma}_{11}^{-1}\tilde{\Sigma}_{12} + O_p \left(\frac{1}{T} \right), \quad (\text{B.39})$$

$$\tilde{B}_{cs} = \tilde{\Sigma}_{*2}^{-1}\tilde{\Sigma}_{c1} - \tilde{\Sigma}_{*2}^{-1}\tilde{\Sigma}_{c2}\tilde{\Sigma}_{22}^{-1}\tilde{\Sigma}_{21}, \quad \tilde{B}_{ss} = \tilde{\Sigma}_{22}^{-1}\tilde{\Sigma}_{21} + O_p \left(\frac{1}{T} \right). \quad (\text{B.40})$$

Substituting equations (B.39) - (B.40) into the equations in (B.38) we get:

$$\begin{aligned}\tilde{R}_{cs} &= \tilde{\Sigma}_{*2}^{-1}\tilde{\Sigma}_{c1} - \tilde{\Sigma}_{*2}^{-1}\tilde{\Sigma}_{c2}\tilde{\Sigma}_{22}^{-1}\tilde{\Sigma}_{21} + \left[\tilde{\Sigma}_{*1}^{-1}\tilde{\Sigma}_{c2} - \tilde{\Sigma}_{*1}^{-1}\tilde{\Sigma}_{c1}\tilde{\Sigma}_{11}^{-1}\tilde{\Sigma}_{12}\right] \left[\tilde{\Sigma}_{22}^{-1}\tilde{\Sigma}_{21} + O_p\left(\frac{1}{T}\right)\right] \\ &= \tilde{\Sigma}_{c1} \left[I_{k_1^s} - \tilde{\Sigma}_{11}^{-1}\tilde{\Sigma}_{12}\tilde{\Sigma}_{22}^{-1}\tilde{\Sigma}_{21}\right] + O_p\left(\frac{1}{T}\right),\end{aligned}$$

and

$$\tilde{R}_{ss} = \left[\tilde{\Sigma}_{11}^{-1}\tilde{\Sigma}_{12} + O_p\left(\frac{1}{T}\right)\right] \left[\tilde{\Sigma}_{22}^{-1}\tilde{\Sigma}_{21} + O_p\left(\frac{1}{T}\right)\right] = \tilde{\Sigma}_{11}^{-1}\tilde{\Sigma}_{12}\tilde{\Sigma}_{22}^{-1}\tilde{\Sigma}_{21} + O_p\left(\frac{1}{T}\right).$$

Therefore we have $\tilde{R}_{cs} = \tilde{\Sigma}_{c1}(I_{k_1-k^c} - \tilde{R}_{ss}) + O_p\left(\frac{1}{T}\right)$, which implies:

$$-\tilde{\Sigma}_{c1} + \tilde{R}_{cs}(I_{k_1-k^c} - \tilde{R}_{ss})^{-1} = O_p\left(\frac{1}{T}\right). \quad (\text{B.41})$$

Equations (B.37) and (B.41), bound $\hat{\Psi}_{sc}^* = O_p\left(\frac{1}{\sqrt{NT}}\right)$, together with the assumption $\sqrt{N}/T = o(1)$, imply:

$$\left[I_{k^c} \quad \tilde{R}_{cs}(I_{(k_1-k^c)} - \tilde{R}_{ss})^{-1}\right] \tilde{V}_{11}^{-1} \begin{bmatrix} \hat{\Psi}_{cc}^* \\ \hat{\Psi}_{sc}^* \end{bmatrix} = \tilde{\Sigma}_{cc}^{-1}\hat{\Psi}_{cc}^* + o_p\left(\frac{1}{N\sqrt{T}}\right).$$

Thus, from (B.36) we get the asymptotic expansion:

$$\sum_{\ell=1}^{k^c} \hat{\rho}_\ell = k^c + \frac{1}{2} \text{tr} \left\{ \tilde{\Sigma}_{cc}^{-1} \hat{\Psi}_{cc}^* \right\} + o_p\left(\frac{1}{N\sqrt{T}}\right). \quad (\text{B.42})$$

Now, let us derive the asymptotic expansion of $\hat{\Psi}_{cc}^*$. From equation (B.12), this term is given by:

$$\hat{\Psi}_{cc}^* = \left[-\hat{X}_{11}\tilde{R} + \hat{X}_{12}\tilde{B} - \tilde{B}'\hat{X}_{22}\tilde{B} + \tilde{B}'\hat{X}_{21}\right]_{(11)},$$

with $M_{(ij)}$ denoting the block in position (i, j) of matrix M . As matrices \tilde{R} and \tilde{B} have the same structure $[E_c \quad *]$, we have:

$$\hat{\Psi}_{cc}^* = \left[-\hat{X}_{11} + \hat{X}_{12} - \hat{X}_{22} + \hat{X}_{21}\right]_{(11)}. \quad (\text{B.43})$$

Let us compute the asymptotic expansions of the terms \hat{X}_{11} , \hat{X}_{12} , \hat{X}_{22} and \hat{X}_{21} . Vectors $u_{j,t}$, with $j = 1, 2$, can be partitioned into the k^c -dimensional vector $u_{jt}^{(c)}$ and the k_j^s -dimensional vector $u_{jt}^{(s)}$:

$$u_{jt} = \begin{bmatrix} u_{jt}^{(c)} \\ u_{jt}^{(s)} \end{bmatrix}, \quad j = 1, 2,$$

and from Assumption A.5 we can express $\Sigma_{u,j}$, $j = 1, 2$, as:³⁵

$$\Sigma_{u,j} = E[u_{jt}u_{jt}'] = E \begin{bmatrix} u_{jt}^{(c)} u_{jt}^{(c)'} & u_{jt}^{(c)} u_{jt}^{(s)'} \\ u_{jt}^{(s)} u_{jt}^{(c)'} & u_{jt}^{(s)} u_{jt}^{(s)'} \end{bmatrix} = \begin{bmatrix} \Sigma_{u,j}^{(cc)} & \Sigma_{u,j}^{(cs)} \\ \Sigma_{u,j}^{(sc)} & \Sigma_{u,j}^{(ss)} \end{bmatrix}, \quad j = 1, 2. \quad (\text{B.44})$$

³⁵Matrix $\Sigma_{u,j}$ is the asymptotic variance of $u_{j,t}$ as $N_j \rightarrow \infty$. We omit the limit for expository purpose.

We also define:

$$\Sigma_{u,12} := E[u_{1t}u'_{2t}] := E \begin{bmatrix} u_{1t}^{(c)}u_{2t}^{(c)'} & u_{1t}^{(c)}u_{2t}^{(s)'} \\ u_{1t}^{(s)}u_{2t}^{(c)'} & u_{1t}^{(s)}u_{2t}^{(s)'} \end{bmatrix} = \begin{bmatrix} \Sigma_{u,12}^{(cc)} & \Sigma_{u,12}^{(cs)} \\ \Sigma_{u,12}^{(sc)} & \Sigma_{u,12}^{(ss)} \end{bmatrix},$$

and $\Sigma_{u,21} = \Sigma'_{u,12}$. From equation (B.10) we have:

$$\begin{aligned} \hat{X}_{11} &= \frac{\mu_N}{T\sqrt{N}} \sum_{t=1}^T (h_{1,t}u'_{1,t} + u_{1,t}h'_{1,t}) + \frac{\mu_N^2}{TN} \sum_{t=1}^T u_{1,t}u'_{1,t} \\ &= \frac{\mu_N}{\sqrt{TN}} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T \begin{bmatrix} f_t^c u_{1t}^{(c)'} + u_{1t}^{(c)} f_t^{c'} & f_t^c u_{1t}^{(s)'} + u_{1t}^{(c)} f_t^{s'} \\ f_{1t}^s u_{1t}^{(c)'} + u_{1t}^{(s)} f_t^{c'} & f_{1t}^s u_{1t}^{(s)'} + u_{1t}^{(s)} f_t^{s'} \end{bmatrix} \right) + \frac{\mu_N^2}{TN} \sum_{t=1}^T \begin{bmatrix} u_{1t}^{(c)}u_{1t}^{(c)'} & u_{1t}^{(c)}u_{1t}^{(s)'} \\ u_{1t}^{(s)}u_{1t}^{(c)'} & u_{1t}^{(s)}u_{1t}^{(s)'} \end{bmatrix}, \end{aligned}$$

and using the definition of matrix $\Sigma_{u,1}$ in equation (B.44) we get:

$$\begin{aligned} \hat{X}_{11} &= \frac{\mu_N}{\sqrt{TN}} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T \begin{bmatrix} f_t^c u_{1t}^{(c)'} + u_{1t}^{(c)} f_t^{c'} & f_t^c u_{1t}^{(s)'} + u_{1t}^{(c)} f_t^{s'} \\ f_{1t}^s u_{1t}^{(c)'} + u_{1t}^{(s)} f_t^{c'} & f_{1t}^s u_{1t}^{(s)'} + u_{1t}^{(s)} f_t^{s'} \end{bmatrix} \right) \\ &\quad + \frac{\mu_N^2}{N} \begin{bmatrix} \Sigma_{u,1}^{(cc)} & \Sigma_{u,1}^{(cs)} \\ \Sigma_{u,1}^{(sc)} & \Sigma_{u,1}^{(ss)} \end{bmatrix} + \frac{\mu_N^2}{N\sqrt{T}} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T \begin{bmatrix} u_{1t}^{(c)}u_{1t}^{(c)'} - \Sigma_{u,1}^{(cc)} & u_{1t}^{(c)}u_{1t}^{(s)'} - \Sigma_{u,1}^{(cs)} \\ u_{1t}^{(s)}u_{1t}^{(c)'} - \Sigma_{u,1}^{(sc)} & u_{1t}^{(s)}u_{1t}^{(s)'} - \Sigma_{u,1}^{(ss)} \end{bmatrix} \right). \end{aligned}$$

Analogously, from (B.11) we have:

$$\begin{aligned} \hat{X}_{22} &= \frac{1}{\sqrt{TN}} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T \begin{bmatrix} f_t^c u_{2t}^{(c)'} + u_{2t}^{(c)} f_t^{c'} & f_t^c u_{2t}^{(s)'} + u_{2t}^{(c)} f_t^{s'} \\ f_{2t}^s u_{2t}^{(c)'} + u_{2t}^{(s)} f_t^{c'} & f_{2t}^s u_{2t}^{(s)'} + u_{2t}^{(s)} f_t^{s'} \end{bmatrix} \right) \\ &\quad + \frac{1}{N} \begin{bmatrix} \Sigma_{u,2}^{(cc)} & \Sigma_{u,2}^{(cs)} \\ \Sigma_{u,2}^{(sc)} & \Sigma_{u,2}^{(ss)} \end{bmatrix} + \frac{1}{N\sqrt{T}} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T \begin{bmatrix} u_{2t}^{(c)}u_{2t}^{(c)'} - \Sigma_{u,2}^{(cc)} & u_{2t}^{(c)}u_{2t}^{(s)'} - \Sigma_{u,2}^{(cs)} \\ u_{2t}^{(s)}u_{2t}^{(c)'} - \Sigma_{u,2}^{(sc)} & u_{2t}^{(s)}u_{2t}^{(s)'} - \Sigma_{u,2}^{(ss)} \end{bmatrix} \right). \end{aligned}$$

From equation (B.9), the term \hat{X}_{12} results to be:

$$\begin{aligned} \hat{X}_{12} &= \frac{1}{T\sqrt{N}} \sum_{t=1}^T (h_{1,t}u'_{2,t} + \mu_N u_{1,t}h'_{2,t}) + \frac{\mu_N}{TN} \sum_{t=1}^T u_{1,t}u'_{2,t} \\ &= \frac{1}{\sqrt{TN}} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T \begin{bmatrix} f_t^c u_{2t}^{(c)'} + \mu_N u_{1t}^{(c)} f_t^{c'} & f_t^c u_{2t}^{(s)'} + \mu_N u_{1t}^{(c)} f_t^{s'} \\ f_{1t}^s u_{2t}^{(c)'} + \mu_N u_{1t}^{(s)} f_t^{c'} & f_{1t}^s u_{2t}^{(s)'} + \mu_N u_{1t}^{(s)} f_t^{s'} \end{bmatrix} \right) + \frac{\mu_N}{N} \begin{bmatrix} \Sigma_{u,12}^{(cc)} & \Sigma_{u,12}^{(cs)} \\ \Sigma_{u,12}^{(sc)} & \Sigma_{u,12}^{(ss)} \end{bmatrix} \\ &\quad + \frac{\mu_N}{N\sqrt{T}} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T \begin{bmatrix} u_{1t}^{(c)}u_{2t}^{(c)'} - \Sigma_{u,12}^{(cc)} & u_{1t}^{(c)}u_{2t}^{(s)'} - \Sigma_{u,12}^{(cs)} \\ u_{1t}^{(s)}u_{2t}^{(c)'} - \Sigma_{u,12}^{(sc)} & u_{1t}^{(s)}u_{2t}^{(s)'} - \Sigma_{u,12}^{(ss)} \end{bmatrix} \right). \end{aligned}$$

Finally we have:

$$\begin{aligned} \hat{X}_{21} &= \hat{X}'_{12} = \frac{1}{\sqrt{TN}} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T \begin{bmatrix} u_{2t}^{(c)} f_t^{c'} + \mu_N f_t^c u_{1t}^{(c)'} & u_{2t}^{(c)} f_{1t}^{s'} + \mu_N f_t^c u_{1t}^{(s)'} \\ u_{2t}^{(s)} f_t^{c'} + \mu_N f_{2t}^s u_{1t}^{(c)'} & u_{2t}^{(s)} f_{1t}^{s'} + \mu_N f_{2t}^s u_{1t}^{(s)'} \end{bmatrix} \right) \\ &\quad + \frac{\mu_N}{N} \begin{bmatrix} \Sigma_{u,21}^{(cc)} & \Sigma_{u,21}^{(cs)} \\ \Sigma_{u,21}^{(sc)} & \Sigma_{u,21}^{(ss)} \end{bmatrix} + \frac{\mu_N}{N\sqrt{T}} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T \begin{bmatrix} u_{2t}^{(c)}u_{1t}^{(c)'} - \Sigma_{u,21}^{(cc)} & u_{2t}^{(c)}u_{1t}^{(s)'} - \Sigma_{u,21}^{(cs)} \\ u_{2t}^{(s)}u_{1t}^{(c)'} - \Sigma_{u,21}^{(sc)} & u_{2t}^{(s)}u_{1t}^{(s)'} - \Sigma_{u,21}^{(ss)} \end{bmatrix} \right). \end{aligned}$$

We can now compute directly term $\hat{\Psi}_{cc}^*$. From equation (B.43), we get:

$$\begin{aligned}
& \hat{\Psi}_{cc}^* \\
&= \frac{1}{\sqrt{TN}} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T \left[-\mu_N f_t^c u_{1t}^{(c)'} - \mu_N u_{1t}^{(c)} f_t^{c'} + f_t^c u_{2t}^{(c)'} + \mu_N u_{1t}^{(c)} f_t^{c'} - f_t^c u_{2t}^{(c)'} - u_{2t}^{(c)} f_t^{c'} + u_{2t}^{(c)} f_t^{c'} + \mu_N f_t^c u_{1t}^{(c)'} \right] \right) \\
&+ \frac{1}{N} [-\mu_N^2 \Sigma_{u,1}^{(cc)} - \Sigma_{u,2}^{(cc)} + \mu_N \Sigma_{u,12}^{(cc)} + \mu_N \Sigma_{u,21}^{(cc)}] \\
&+ \frac{1}{N\sqrt{T}} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T \left[-\mu_N^2 [u_{1t}^{(c)} u_{1t}^{(c)'} - \Sigma_{u,1}^{(cc)}] + \mu_N [u_{1t}^{(c)} u_{2t}^{(c)'} - \Sigma_{u,12}^{(cc)}] - [u_{2t}^{(c)} u_{2t}^{(c)'} - \Sigma_{u,2}^{(cc)}] + \mu_N [u_{2t}^{(c)} u_{1t}^{(c)'} - \Sigma_{u,21}^{(cc)}] \right] \right) \\
&= -\frac{1}{N} E[(\mu_N u_{1t}^{(c)} - u_{2t}^{(c)})(\mu_N u_{1t}^{(c)} - u_{2t}^{(c)})'] \\
&- \frac{1}{N\sqrt{T}} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T \left[(\mu_N u_{1t}^{(c)} - u_{2t}^{(c)})(\mu_N u_{1t}^{(c)} - u_{2t}^{(c)})' - E[(\mu_N u_{1t}^{(c)} - u_{2t}^{(c)})(\mu_N u_{1t}^{(c)} - u_{2t}^{(c)})'] \right] \right),
\end{aligned}$$

where the term at order $O_p(1/\sqrt{NT})$ vanishes. Using the limit $\mu_N \rightarrow \mu$, we get the asymptotic expansion of $\hat{\Psi}_{cc}^*$:

$$\begin{aligned}
\hat{\Psi}_{cc}^* &= -\frac{1}{N} E[(\mu_N u_{1t}^{(c)} - u_{2t}^{(c)})(\mu_N u_{1t}^{(c)} - u_{2t}^{(c)})'] \\
&- \frac{1}{N\sqrt{T}} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T \left[(\mu u_{1t}^{(c)} - u_{2t}^{(c)})(\mu u_{1t}^{(c)} - u_{2t}^{(c)})' - E[(\mu u_{1t}^{(c)} - u_{2t}^{(c)})(\mu u_{1t}^{(c)} - u_{2t}^{(c)})'] \right] \right) + o_p\left(\frac{1}{N\sqrt{T}}\right).
\end{aligned} \tag{B.45}$$

By substituting (B.45) into (B.42), and using $\tilde{\Sigma}_{cc} = I_{k^c} + o_p(1)$, we get:

$$\begin{aligned}
\sum_{\ell=1}^{k^c} \hat{\rho}_\ell &= k^c - \frac{1}{2N} tr \left\{ \tilde{\Sigma}_{cc}^{-1} E[(\mu_N u_{1t}^{(c)} - u_{2t}^{(c)})(\mu_N u_{1t}^{(c)} - u_{2t}^{(c)})'] \right\} \\
&- \frac{1}{2N\sqrt{T}} tr \left\{ \frac{1}{\sqrt{T}} \sum_{t=1}^T \left[(\mu u_{1t}^{(c)} - u_{2t}^{(c)})(\mu u_{1t}^{(c)} - u_{2t}^{(c)})' - E[(\mu u_{1t}^{(c)} - u_{2t}^{(c)})(\mu u_{1t}^{(c)} - u_{2t}^{(c)})'] \right] \right\} \\
&+ o_p\left(\frac{1}{N\sqrt{T}}\right).
\end{aligned} \tag{B.46}$$

From the definition of matrix $\Sigma_{U,N}$ we have $E[(\mu_N u_{1t}^{(c)} - u_{2t}^{(c)})(\mu_N u_{1t}^{(c)} - u_{2t}^{(c)})'] = \Sigma_{U,N}$. Moreover, let us define the process

$$U_t := \mu u_{1t}^{(c)} - u_{2t}^{(c)}.$$

Using these definitions together with the commutativity and linearity properties of the trace operator, from equation (B.46) we get the asymptotic expansion:

$$\sum_{\ell=1}^{k^c} \hat{\rho}_\ell = k^c - \frac{1}{2N} tr \left\{ \tilde{\Sigma}_{cc}^{-1} \Sigma_{U,N} \right\} - \frac{1}{2N\sqrt{T}} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T [U_t' U_t - E(U_t' U_t)] \right) + o_p\left(\frac{1}{N\sqrt{T}}\right). \tag{B.47}$$

B.3.5 Asymptotic distribution of the test statistic under the null hypothesis of k^c common factors

From the asymptotic expansion (B.47) we obtain the asymptotic distribution of $\hat{\xi}(k^c) = \sum_{\ell=1}^{k^c} \hat{\rho}_\ell$ under the null hypothesis $H(k^c)$. By a CLT for weakly dependent data we have:

$$\frac{1}{\sqrt{T}} \sum_{t=1}^T [U_t' U_t - E(U_t' U_t)] \xrightarrow{d} N(0, \Omega_U), \quad (\text{B.48})$$

where the long-run variance-covariance matrix is given by:

$$\Omega_U = \lim_{T \rightarrow \infty} V \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T U_t' U_t \right) = \sum_{h=-\infty}^{\infty} \text{Cov}(U_t' U_t, U_{t-h}' U_{t-h}). \quad (\text{B.49})$$

From equations (B.47) and (B.48), the asymptotic distribution of $\hat{\xi}(k^c) = \sum_{\ell=1}^{k^c} \hat{\rho}_\ell$, under the hypothesis of k^c common factors in each group, is:

$$N\sqrt{T} \left[\hat{\xi}(k^c) - k^c + \frac{1}{2N} \text{tr} \left\{ \tilde{\Sigma}_{cc}^{-1} \Sigma_{U,N} \right\} \right] \xrightarrow{d} N \left(0, \frac{1}{4} \Omega_U \right).$$

To conclude the proof of Theorem 1, let us derive the expression of matrix Ω_U defined in equation (B.49). For this purpose, note that vector $(U_t', U_{t-h}')'$ is asymptotically Gaussian for any h :

$$\begin{pmatrix} U_t \\ U_{t-h} \end{pmatrix} \xrightarrow{d} N \left(0, \begin{bmatrix} \Sigma_U(0) & \Sigma_U(h) \\ \Sigma_U(h)' & \Sigma_U(0) \end{bmatrix} \right).$$

We use the following lemma (see Theorem 12 p. 284 in Magnus and Neudecker (2007) and Theorem 10.21 in Schott (2005) for the proof).

LEMMA B.1. *Let the $(n, 1)$ random vector x and the $(m, 1)$ random vector y be such that*

$$\begin{pmatrix} x \\ y \end{pmatrix} \sim N \left(0, \begin{bmatrix} \Omega_{xx} & \Omega_{xy} \\ \Omega_{xy}' & \Omega_{yy} \end{bmatrix} \right),$$

and let A and B be symmetric (n, n) and (m, m) matrices, respectively. Then $\text{Cov}(x'Ax, y'By) = 2\text{tr} \{ A\Omega_{xy}B\Omega_{xy}' \}$.

From Lemma B.1 we get (asymptotically) $\text{Cov}(U_t' U_t, U_{t-h}' U_{t-h}) = 2\text{tr} \{ \Sigma_U(h)\Sigma_U(h)' \}$ and:

$$\Omega_U = 2 \sum_{h=-\infty}^{\infty} \text{tr} \{ \Sigma_U(h)\Sigma_U(h)' \}.$$

The conclusion follows.

Q.E.D.

B.4 Proofs of Propositions 3 and 4

We prove Propositions 3 and 4 by establishing the asymptotic distribution of the estimators of the common factor values (Section B.4.1), the common factor loadings (Section B.4.2), the specific factor values (Section B.4.3), and the specific factor loadings (Section B.4.4).

B.4.1 Asymptotic distribution of \hat{f}_t^c and \hat{f}_t^{c*}

Equation (B.32) and $\hat{\Psi}_{sc} = O_p\left(\frac{1}{\sqrt{NT}}\right)$ imply $\hat{W}_1^* = E_c \hat{U} + O_p\left(\frac{1}{\sqrt{NT}}\right)$. Recall from equation (B.33) that the normalized eigenvectors corresponding to the canonical directions are: $\hat{W}_1 = \hat{W}_1^* \hat{D}$, where $\hat{D} = \text{diag}(\hat{W}_1^{*'} \hat{V}_{11} \hat{W}_1^*)^{-1/2}$. Then, we get:

$$\begin{aligned}\hat{f}_t^c &= \hat{W}_1' \hat{h}_{1,t} = \hat{D} \hat{U}' E_c' \left(h_{1,t} + \frac{1}{\sqrt{N_1}} u_{1,t} \right) + O_p\left(\frac{1}{\sqrt{NT}}\right) \\ &= \hat{D} \hat{U}' \left(f_t^c + \frac{1}{\sqrt{N_1}} u_{1,t}^{(c)} \right) + O_p\left(\frac{1}{\sqrt{NT}}\right).\end{aligned}$$

Therefore the estimated factor can be written as:

$$\hat{f}_t^c = \hat{\mathcal{H}}_c^{-1} \left(f_t^c + \frac{1}{\sqrt{N_1}} u_{1,t}^{(c)} \right) + O_p\left(\frac{1}{\sqrt{NT}}\right), \quad (\text{B.50})$$

where $\hat{\mathcal{H}}_c^{-1} = \hat{D} \hat{U}'$. Given that $\frac{1}{T} \sum_{t=1}^T \hat{f}_t^c \hat{f}_t^{c'} = I_{k^c}$, and $\frac{1}{T} \sum_{t=1}^T f_t^c f_t^{c'} = I_{k^c} + o_p(1)$, we have

$$\hat{\mathcal{H}}_c \hat{\mathcal{H}}_c' = I_{k^c} + o_p(1). \quad (\text{B.51})$$

Equations (B.50) and (4.2) imply:

$$\sqrt{N_1} \left(\hat{\mathcal{H}}_c \hat{f}_t^c - f_t^c \right) = u_{1,t}^{(c)} + o_p(1) \xrightarrow{d} N\left(0, \Sigma_{u,1}^{(cc)}\right),$$

which proves the result in (4.3). The derivation of the asymptotic distribution of $\sqrt{N_2} \left(\hat{\mathcal{H}}_c^* \hat{f}_t^{c*} - f_t^c \right)$ in (4.4) from the canonical direction \hat{W}_2 is analogous, and is omitted.

B.4.2 Asymptotic distribution of $\hat{\lambda}_{j,i}^c$

Let us derive the asymptotic expansion of the loading estimator $\hat{\lambda}_{j,i}^c = \hat{F}^{c'} y_{j,i} / T$, where $y_{j,i}$ is the i -th column of matrix Y_j . From equation (B.50) we can express $\hat{F}^c = [\hat{f}_1^c, \dots, \hat{f}_T^c]'$ as:

$$\begin{aligned}\hat{F}^c &= \left(F^c + \frac{1}{\sqrt{N_1}} U_1^{(c)} \right) \left(\hat{\mathcal{H}}_c^{-1} \right)' + O_p\left(\frac{1}{\sqrt{NT}}\right) \\ &= F^c \left(\hat{\mathcal{H}}_c^{-1} \right)' + \frac{1}{\sqrt{N_1}} U_1^{(c)} \left(\hat{\mathcal{H}}_c^{-1} \right)' + O_p\left(\frac{1}{\sqrt{NT}}\right),\end{aligned} \quad (\text{B.52})$$

where $U_1^{(c)} = [u_{1,1}^{(c)}, \dots, u_{1,T}^{(c)}]'$. Equation (B.52) implies:

$$\hat{F}^c \hat{\mathcal{H}}_c' - F^c = \frac{1}{\sqrt{N_1}} U_1^{(c)} + O_p\left(\frac{1}{\sqrt{NT}}\right). \quad (\text{B.53})$$

Then, denoting with $\xi_{j,i}$ the i -th column of matrix Ξ_j , we get:

$$\begin{aligned}\hat{\lambda}_{j,i}^c &= \frac{1}{T} \hat{F}^{c'} y_{j,i} = \frac{1}{T} \hat{F}^{c'} \left(F^c \lambda_{j,i}^c + F_j^s \lambda_{j,i}^s + \varepsilon_{j,i} \right) \\ &= \frac{1}{T} \hat{F}^{c'} \left[\left(F^c - \hat{F}^c \hat{\mathcal{H}}_c' + \hat{F}^c \hat{\mathcal{H}}_c' \right) \lambda_{j,i}^c + F_j^s \lambda_{j,i}^s + \varepsilon_{j,i} \right] \\ &= \hat{\mathcal{H}}_c' \lambda_{j,i}^c + \frac{1}{T} \hat{F}^{c'} \varepsilon_{j,i} + \frac{1}{T} \hat{F}^{c'} \left(F^c - \hat{F}^c \hat{\mathcal{H}}_c' \right) \lambda_{j,i}^c + \frac{1}{T} \hat{F}^{c'} F_j^s \lambda_{j,i}^s, \quad j = 1, 2.\end{aligned} \quad (\text{B.54})$$

Equations (B.52) and (B.53) allow to compute:

$$\frac{1}{T}\hat{F}^{c'}\left(F^c - \hat{F}^c\hat{\mathcal{H}}_c'\right) \simeq -\frac{1}{T\sqrt{N_1}}\hat{\mathcal{H}}_c^{-1}F^{c'}U_1^{(c)} - \frac{1}{N_1T}\hat{\mathcal{H}}_c^{-1}U_1^{(c)'}U_1^{(c)} = O_p\left(\frac{1}{\sqrt{NT}}\right), \quad (\text{B.55})$$

and:

$$\frac{1}{T}\hat{F}^{c'}\varepsilon_{j,i} = \hat{\mathcal{H}}_c^{-1}\left(\frac{1}{T}F^{c'}\varepsilon_{j,i} + \frac{1}{T\sqrt{N_1}}U_1^{(c)'}\varepsilon_{j,i}\right) = \hat{\mathcal{H}}_c^{-1}\frac{1}{T}F^{c'}\varepsilon_{j,i} + O_p\left(\frac{1}{\sqrt{NT}}\right). \quad (\text{B.56})$$

We also have:

$$\frac{1}{T}\hat{F}^{c'}F_j^s = \hat{\mathcal{H}}_c^{-1}\left(\frac{1}{T}F^{c'}F_j^s + \frac{1}{T\sqrt{N_1}}U_1^{(c)'}F_j^s\right) = \hat{\mathcal{H}}_c^{-1}\frac{1}{T}F^{c'}F_j^s + O_p\left(\frac{1}{\sqrt{NT}}\right). \quad (\text{B.57})$$

Substituting approximations (B.55) - (B.57) into equation (B.54) we get:

$$\hat{\lambda}_{j,i}^c \simeq \hat{\mathcal{H}}_c'\lambda_{j,i}^c + \hat{\mathcal{H}}_c^{-1}\frac{1}{T}F^{c'}\varepsilon_{j,i} + \hat{\mathcal{H}}_c^{-1}\frac{1}{T}F^{c'}F_j^s\lambda_{j,i}^s + O_p\left(\frac{1}{\sqrt{NT}}\right), \quad j = 1, 2.$$

The last equation and (B.51) imply:

$$\sqrt{T}\left[\left(\hat{\mathcal{H}}_c'\right)^{-1}\hat{\lambda}_{j,i}^c - \lambda_{j,i}^c\right] = \varphi_{j,i} + K_j\lambda_{j,i}^s + o_p(1), \quad (\text{B.58})$$

where:

$$\varphi_{j,i} = \frac{1}{\sqrt{T}}F^{c'}\varepsilon_{j,i}, \quad K_j = \frac{1}{\sqrt{T}}F^{c'}F_j^s.$$

The r.h.s. of equation (B.58) can be rewritten to get:

$$\sqrt{T}\left[\left(\hat{\mathcal{H}}_c'\right)^{-1}\hat{\lambda}_{j,i}^c - \lambda_{j,i}^c\right] = \frac{1}{\sqrt{T}}\sum_{t=1}^T f_t^c(\varepsilon_{j,i,t} + f_{j,t}^{s'}\lambda_{j,i}^s) + o_p(1) \equiv w_{j,i}^c + o_p(1). \quad (\text{B.59})$$

Term $w_{j,i}^c$ can be written as:

$$w_{j,i}^c = \frac{1}{\sqrt{T}}\sum_{t=1}^T [f_t^c\varepsilon_{j,i,t} + (\lambda_{j,i}^{s'} \otimes I_{k^c})(f_{j,t}^{s'} \otimes f_t^c)].$$

Then, since the errors and the factors are independent (Assumption A.5 a)), a CLT for weakly dependent data yields equation (4.6).

B.4.3 Asymptotic distribution of $\hat{f}_{j,t}^s$

Let us now derive the asymptotic expansion of term $\hat{f}_{j,t}^s$. We start by computing the asymptotic expansion of the regression residuals $y_{j,i,t} - \hat{f}_t^{c'}\hat{\lambda}_{j,i}^c$:

$$\begin{aligned} y_{j,i,t} - \hat{f}_t^{c'}\hat{\lambda}_{j,i}^c &= f_{j,t}^{s'}\lambda_{j,i}^s + \varepsilon_{j,i,t} - \left(\hat{f}_t^{c'}\hat{\lambda}_{j,i}^c - f_t^{c'}\lambda_{j,i}^c\right) \\ &\simeq f_{j,t}^{s'}\lambda_{j,i}^s + \varepsilon_{j,i,t} - \left[\left(f_t^c + \frac{1}{\sqrt{N_1}}u_{1,t}^{(c)}\right)'\left(\lambda_{j,i}^c + \frac{1}{\sqrt{T}}\varphi_{j,i} + \frac{1}{\sqrt{T}}K_j\lambda_{j,i}^s\right) - f_t^{c'}\lambda_{j,i}^c\right] \\ &\simeq g_{j,t}'\lambda_{j,i}^s + e_{j,i,t}, \end{aligned} \quad (\text{B.60})$$

where we use equation (B.50) and (B.58), and we define:

$$g_{j,t} := f_{j,t}^s - \frac{1}{\sqrt{T}} K_j' f_t^c \simeq f_{j,t}^s - (F_j^{s'} F^c)(F^c' F^c)^{-1} f_t^c, \quad e_{j,i,t} := \varepsilon_{j,i,t} - \frac{1}{\sqrt{T}} f_t^{c'} \varphi_{j,i}.$$

Then, the residuals $y_{j,i,t} - \hat{f}_t^c \hat{\lambda}_{j,i}^s$, with $i = 1, \dots, N_j$ and $t = 1, \dots, T$, satisfy an approximate factor structure with factors $g_{j,t}$ and errors $e_{j,i,t}$. From asymptotic theory of the PC estimators in large panels, we know that:

$$\sqrt{N_j} \left[\hat{\mathcal{H}}_{s,j} \hat{f}_{j,t}^s - g_{j,t} \right] = v_{j,t}^{*s} + o_p(1), \quad j = 1, 2, \quad (\text{B.61})$$

where $\hat{\mathcal{H}}_{s,j}$, $j = 1, 2$, is a non-singular matrix and:

$$\begin{aligned} v_{j,t}^{*s} &= \left(\frac{\Lambda_j^{s'} \Lambda_j^s}{N_j} \right)^{-1} \frac{1}{\sqrt{N_j}} \Lambda_j^{s'} e_{j,t} \\ &= \left(\frac{\Lambda_j^{s'} \Lambda_j^s}{N_j} \right)^{-1} \frac{1}{\sqrt{N_j}} \sum_{i=1}^{N_j} \lambda_{j,i}^s \varepsilon_{j,i,t} - \left(\frac{\Lambda_j^{s'} \Lambda_j^s}{N_j} \right)^{-1} \frac{1}{\sqrt{NT}} \sum_{i=1}^{N_j} \lambda_{j,i}^s f_t^{c'} \left(\frac{1}{\sqrt{T}} \sum_{r=1}^T f_r^c \varepsilon_{j,i,r} \right) \\ &= \left(\frac{\Lambda_j^{s'} \Lambda_j^s}{N_j} \right)^{-1} \frac{1}{\sqrt{N_j}} \sum_{i=1}^{N_j} \lambda_{j,i}^s \varepsilon_{j,i,t} + o_p(1). \end{aligned}$$

Therefore we have

$$\sqrt{N_j} \left[\hat{\mathcal{H}}_{s,j} \hat{f}_{j,t}^s - (f_{j,t}^s - (F_j^{s'} F^c)(F^c' F^c)^{-1} f_t^c) \right] = v_{j,t}^s + o_p(1), \quad j = 1, 2,$$

where $v_{j,t}^s = \left(\frac{\Lambda_j^{s'} \Lambda_j^s}{N_j} \right)^{-1} \frac{1}{\sqrt{N_j}} \sum_{i=1}^{N_j} \lambda_{j,i}^s \varepsilon_{j,i,t}$, which yields equation (4.5).

B.4.4 Asymptotic distribution of $\hat{\lambda}_{j,i}^s$

From asymptotic theory of the PC estimators in large panels, we know that the following result must hold for the loadings estimator of factor model (B.60):

$$\sqrt{T} \left[\left(\hat{\mathcal{H}}_{s,j}' \right)^{-1} \hat{\lambda}_{j,i}^s - \lambda_{j,i}^s \right] = w_{j,i}^{*s} + o_p(1), \quad j = 1, 2,$$

where $\hat{\mathcal{H}}_{s,j}$, $j = 1, 2$ are the same non-singular matrices in equation (B.61), and

$$\begin{aligned} w_{j,i}^{*s} &= \frac{1}{\sqrt{T}} \sum_{t=1}^T \left(f_{j,t}^s + \frac{1}{\sqrt{T}} K_j' f_t^c \right) e_{j,i,t} = \frac{1}{\sqrt{T}} \sum_{t=1}^T \left(f_{j,t}^s + \frac{1}{\sqrt{T}} K_j' f_t^c \right) \left(\varepsilon_{j,i,t} - \frac{1}{\sqrt{T}} f_t^{c'} \varphi_{j,i} \right), \\ &= \frac{1}{\sqrt{T}} \sum_{t=1}^T f_{j,t}^s \varepsilon_{j,i,t} - \frac{1}{T} \sum_{t=1}^T f_{j,t}^s f_t^{c'} \varphi_{j,i} + K_j' \frac{1}{T} \sum_{t=1}^T f_t^c \varepsilon_{j,i,t} - K_j' \frac{1}{T\sqrt{T}} \sum_{t=1}^T f_t^c f_t^{c'} \varphi_{j,i} \\ &= \frac{1}{\sqrt{T}} \sum_{t=1}^T f_{j,t}^s \varepsilon_{j,i,t} + o_p(1), \end{aligned}$$

since $\frac{1}{T} \sum_{t=1}^T f_{j,t}^s f_t^{c'} = o_p(1)$. Therefore, we get:

$$\sqrt{T} \left[\left(\hat{\mathcal{H}}_{s,j}' \right)^{-1} \hat{\lambda}_{j,i}^s - \lambda_{j,i}^s \right] = \frac{1}{\sqrt{T}} \sum_{t=1}^T f_{j,t}^s \varepsilon_{j,i,t} + o_p(1) \equiv w_{j,i}^s + o_p(1), \quad (\text{B.62})$$

which yields equation (4.7).

Q.E.D.

B.5 Proof of Theorem 2

B.5.1 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. Theorem 2 i) follows from Theorem 1 since we have:

$$\text{tr} \left\{ \hat{\Sigma}_U \right\} = \text{tr} \left\{ \tilde{\Sigma}_{cc}^{-1} \Sigma_{U,N} \right\} + o_p(1/\sqrt{T}), \quad (\text{B.63})$$

$$\text{tr} \left\{ \hat{\Sigma}_U^2 \right\} = \text{tr} \left\{ \Sigma_U(0)^2 \right\} + o_p(1), \quad (\text{B.64})$$

and $\Omega_U = 2\text{tr} \left\{ \Sigma_U(0)^2 \right\}$ under Assumption A.7. Equations (B.63) and (B.64) are proved next, by deriving the asymptotic expansions of $\tilde{\Sigma}_{cc}^{-1}$ and $\hat{\Sigma}_U$.

a) Asymptotic expansion of $\tilde{\Sigma}_{cc}^{-1}$

Substituting the expression of \hat{f}_t^c from equation (B.50) into the equality $\frac{1}{T} \sum_{t=1}^T \hat{f}_t^c \hat{f}_t^{c'} = I_{k^c}$, we get:

$$\begin{aligned} I_{k^c} &= \frac{1}{T} \sum_{t=1}^T \hat{\mathcal{H}}_c^{-1} \left(f_t^c + \frac{1}{\sqrt{N_1}} u_{1,t}^{(c)} \right) \left(f_t^c + \frac{1}{\sqrt{N_1}} u_{1,t}^{(c)} \right)' \left(\hat{\mathcal{H}}_c^{-1} \right)' + O_p \left(\frac{1}{\sqrt{NT}} \right) \\ &= \hat{\mathcal{H}}_c^{-1} \tilde{\Sigma}_{cc} \left(\hat{\mathcal{H}}_c^{-1} \right)' + O_p \left(\frac{1}{\sqrt{NT}} \right). \end{aligned}$$

This implies:

$$\tilde{\Sigma}_{cc}^{-1} = \left(\hat{\mathcal{H}}_c^{-1} \right)' \hat{\mathcal{H}}_c^{-1} + O_p \left(\frac{1}{\sqrt{NT}} \right). \quad (\text{B.65})$$

To derive the asymptotic expansion of $\hat{\Sigma}_U$, we use its definition $\hat{\Sigma}_U = \mu_N^2 \hat{\Sigma}_{u,1}^{(cc)} + \hat{\Sigma}_{u,2}^{(cc)}$, and expand each component of $\hat{\Sigma}_{u,j}^{(cc)}$, $j = 1, 2$, in (4.8).

b) Asymptotic expansion of $\frac{\hat{\Lambda}_j' \hat{\Lambda}_j}{N_j}$.

To derive the asymptotic expansion of matrix $\hat{\Lambda}_j' \hat{\Lambda}_j / N_j$, it is useful to write the matrix versions of the quantities defined in equations (B.59) and (B.62). Stacking the loadings $\hat{\lambda}_{j,i}^c$ in matrix $\hat{\Lambda}_j^c = [\hat{\lambda}_{j,1}^c, \dots, \hat{\lambda}_{j,N}^c]'$ we get:

$$\hat{\Lambda}_j^c = \left[\Lambda_j^c + \frac{1}{\sqrt{T}} G_j^c \right] \hat{\mathcal{H}}_c + o_p \left(\frac{1}{\sqrt{T}} \right),$$

where

$$G_j^c = \frac{1}{\sqrt{T}} \varepsilon_j' F^c + \Lambda_j^s \left(\frac{1}{\sqrt{T}} F_j^{s'} F^c \right) \quad (\text{B.66})$$

$$= \frac{1}{\sqrt{T}} \varepsilon_j' F^c + \Lambda_j^s \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T f_{j,t}^s f_t^{c'} \right). \quad (\text{B.67})$$

Similarly, stacking the loadings $\hat{\lambda}_{j,i}^s$ in matrix $\hat{\Lambda}_j^s = [\hat{\lambda}_{j,1}^s, \dots, \hat{\lambda}_{j,N}^s]'$ we get:

$$\hat{\Lambda}_j^s = \left[\Lambda_j^c + \frac{1}{\sqrt{T}} G_j^s \right] \hat{\mathcal{H}}_{j,s} + o_p \left(\frac{1}{\sqrt{T}} \right),$$

where

$$G_j^s = \frac{1}{\sqrt{T}} \varepsilon_j' F_j^s. \quad (\text{B.68})$$

By gathering these expansions, we get:

$$\hat{\Lambda}_j \simeq \left(\Lambda_j + \frac{1}{\sqrt{T}} G_j \right) \hat{U}_j, \quad j = 1, 2, \quad (\text{B.69})$$

where

$$G_j = \begin{bmatrix} G_j^c & \vdots & G_j^s \end{bmatrix}, \quad (\text{B.70})$$

$$\hat{U}_j = \begin{bmatrix} \hat{\mathcal{H}}_c & 0 \\ 0 & \hat{\mathcal{H}}_{s,j} \end{bmatrix}. \quad (\text{B.71})$$

We start by computing the asymptotic expansion of $\frac{\hat{\Lambda}_j' \hat{\Lambda}_j}{N_j}$. From Assumptions A.1, A.2 and A.5 we get:

$$\frac{1}{N_j} \left[\Lambda_j + \frac{1}{\sqrt{T}} G_j \right]' \left[\Lambda_j + \frac{1}{\sqrt{T}} G_j \right] \simeq \frac{1}{N_j} \Lambda_j' \Lambda_j + \frac{1}{N \sqrt{T}} (\Lambda_j' G_j + G_j' \Lambda_j) + \frac{1}{NT} G_j' G_j. \quad (\text{B.72})$$

Let us compute the asymptotic expansion of $\frac{1}{N \sqrt{T}} \Lambda_j' G_j$:

$$\frac{1}{N_j \sqrt{T}} \Lambda_j' G_j = \frac{1}{N_j \sqrt{T}} \begin{bmatrix} \Lambda_j^c G_j^c & \Lambda_j^c G_j^s \\ \Lambda_j^s G_j^c & \Lambda_j^s G_j^s \end{bmatrix}.$$

Using equation (B.66) we get:

$$\begin{aligned} \frac{1}{N_j \sqrt{T}} \Lambda_j^c G_j^c &= \frac{1}{N_j \sqrt{T}} \Lambda_j^c \left[\frac{1}{\sqrt{T}} \varepsilon_j' F^c + \Lambda_j^s \left(\frac{1}{\sqrt{T}} F_j^{s'} F^c \right) \right] = \frac{1}{N_j T} \Lambda_j^c \varepsilon_j' F^c + \frac{1}{N_j T} \Lambda_j^c \Lambda_j^s (F_j^{s'} F^c) \\ &= \left(\frac{\Lambda_j^c \Lambda_j^s}{N_j} \right) \frac{1}{T} \sum_{t=1}^T f_{j,t}^s f_t^c + O_p \left(\frac{1}{\sqrt{N_j T}} \right), \end{aligned}$$

Using analogous arguments and equation (B.68), we get:

$$\frac{1}{N_j \sqrt{T}} \Lambda_j^s G_j^c = \left(\frac{\Lambda_j^s \Lambda_j^s}{N_j} \right) \frac{1}{T} \sum_{t=1}^T f_{j,t}^s f_t^c + O_p \left(\frac{1}{\sqrt{N_j T}} \right),$$

$$\frac{1}{N_j \sqrt{T}} \Lambda_j^c G_j^s = \frac{1}{N_j \sqrt{T}} \Lambda_j^c \varepsilon_j' F^s = O_p \left(\frac{1}{\sqrt{N_j T}} \right),$$

$$\frac{1}{N_j \sqrt{T}} \Lambda_j^s G_j^s = \frac{1}{N_j \sqrt{T}} \Lambda_j^s \varepsilon_j' F^s = O_p \left(\frac{1}{\sqrt{N_j T}} \right).$$

The last four equations imply:

$$\begin{aligned} \frac{1}{N_j \sqrt{T}} \Lambda'_j G_j &= \begin{bmatrix} \left(\frac{\Lambda'_j \Lambda_j^s}{N_j} \right) \frac{1}{T} \sum_{t=1}^T f_{j,t}^s f_{j,t}^{c'} & 0 \\ \left(\frac{\Lambda_j^{s'} \Lambda_j^s}{N_j} \right) \frac{1}{T} \sum_{t=1}^T f_{j,t}^s f_{j,t}^{c'} & 0 \end{bmatrix} + O_p \left(\frac{1}{\sqrt{N_j T}} \right) \\ &= \begin{bmatrix} \left(\frac{\Lambda'_j \Lambda_j^s}{N_j} \right) \frac{1}{T} \sum_{t=1}^T f_{j,t}^s f_{j,t}^{c'} & \vdots & 0_{(k_j \times k_j^s)} \end{bmatrix} + O_p \left(\frac{1}{\sqrt{N_j T}} \right). \end{aligned} \quad (\text{B.73})$$

Using analogous arguments, we have:

$$\frac{1}{N_j T} G_j^{c'} G_j^c = \frac{1}{N_j T} \left[\frac{1}{\sqrt{T}} \varepsilon_j^{c'} F^c + \Lambda_j^s \left(\frac{1}{\sqrt{T}} F_j^{s'} F^c \right) \right]' \left[\frac{1}{\sqrt{T}} \varepsilon_j^{c'} F^c + \Lambda_j^s \left(\frac{1}{\sqrt{T}} F_j^{s'} F^c \right) \right] = o_p \left(\frac{1}{\sqrt{T}} \right)$$

and

$$\frac{1}{N_j T} G_j^{c'} G_j = o_p \left(\frac{1}{\sqrt{T}} \right). \quad (\text{B.74})$$

Substituting (B.73) and (B.74) into equation (B.72) we get:

$$\frac{1}{N_j} \left[\Lambda_j + \frac{1}{\sqrt{T}} G_j \right]' \left[\Lambda_j + \frac{1}{\sqrt{T}} G_j \right] \simeq \Sigma_{\Lambda,j} + \frac{1}{\sqrt{T}} (L_{1,j} + L'_{1,j}) + O_p \left(\frac{1}{\sqrt{N}} \right)$$

where

$$L_{1,j} = \begin{bmatrix} \left(\frac{\Lambda'_j \Lambda_j^s}{N_j} \right) \left(\frac{1}{\sqrt{T}} F_j^{s'} F^c \right) & \vdots & 0_{(k_j \times k_j^s)} \end{bmatrix}. \quad (\text{B.75})$$

Therefore we have:

$$\frac{\hat{\Lambda}'_j \hat{\Lambda}_j}{N_j} = \hat{U}'_j \left[\Sigma_{\Lambda,j} + \frac{1}{\sqrt{T}} (L_{1,j} + L'_{1,j}) \right] \hat{U}_j + o_p \left(\frac{1}{\sqrt{T}} \right). \quad (\text{B.76})$$

c) Asymptotic expansion of $\hat{\Gamma}_j$

The approximations in Propositions 3 and 4 allow to compute the asymptotic expansion of $\hat{\varepsilon}_{j,i,t}$:

$$\begin{aligned} \hat{\varepsilon}_{j,i,t} &= y_{j,i,t} - \hat{\lambda}_{j,i}^{c'} \hat{f}_t^c - \hat{\lambda}_{j,i}^{s'} \hat{f}_{j,t}^s = \varepsilon_{j,i,t} - \left[\hat{\lambda}_{j,i}^{c'} \hat{f}_t^c - \lambda_{j,i}^{c'} f_t^c \right] - \left[\hat{\lambda}_{j,i}^{s'} \hat{f}_{j,t}^s - \lambda_{j,i}^{s'} f_{j,t}^s \right] \\ &\simeq \varepsilon_{j,i,t} - \left[\left(\lambda_{j,i}^c + \frac{1}{\sqrt{T}} w_{j,i}^c \right)' \left(f_t^c + \frac{1}{\sqrt{N_1}} u_{1,t}^{(c)} \right) - \lambda_{j,i}^{c'} f_t^c \right] \\ &\quad - \left[\left(\lambda_{j,i}^s + \frac{1}{\sqrt{T}} w_{j,i}^s \right)' \left(f_{j,t}^s - \frac{1}{\sqrt{T}} K_j' f_t^c + \frac{1}{\sqrt{N_j}} v_{j,t}^s \right) - \lambda_{j,i}^{s'} f_{j,t}^s \right] \\ &\simeq \varepsilon_{j,i,t} - \left(\frac{1}{\sqrt{N_1}} \lambda_{j,i}^{c'} u_{1,t}^{(c)} + \frac{1}{\sqrt{T}} w_{j,i}^{c'} f_t^c \right) - \left(\frac{1}{\sqrt{N_j}} \lambda_{j,i}^{s'} v_{j,t}^s + \frac{1}{\sqrt{T}} w_{j,i}^{s'} f_{j,t}^s \right) \\ &\quad + \lambda_{j,i}^{s'} \frac{1}{\sqrt{T}} K_j' f_t^c. \end{aligned} \quad (\text{B.77})$$

Since $T/N_j = o(1)$, we keep only the terms of order $1/\sqrt{T}$ in equation (B.77), and we get:

$$\hat{\varepsilon}_{j,i,t} = \varepsilon_{j,i,t} - \frac{1}{\sqrt{T}} (w_{j,i}^{c'} f_t^c + w_{j,i}^{s'} f_{j,t}^s) + \lambda_{j,i}^{s'} \frac{1}{\sqrt{T}} K_j' f_t^c + o_p \left(\frac{1}{\sqrt{T}} \right). \quad (\text{B.78})$$

From the definition of $w_{j,i}^c$ in Proposition 4 we get:

$$w_{j,i}^{c'} f_t^c = \frac{1}{\sqrt{T}} \left(\sum_{r=1}^T \varepsilon_{j,ir} f_r^{c'} \right) f_t^c + \lambda_{j,i}^{s'} K_j' f_t^c,$$

which implies:

$$\hat{\varepsilon}_{j,i,t} = \varepsilon_{j,i,t} - \frac{1}{\sqrt{T}} (\tilde{w}_{j,i}^{c'} f_t^c + w_{j,i}^{s'} f_{j,t}^s) + o_p \left(\frac{1}{\sqrt{T}} \right),$$

where:

$$\tilde{w}_{j,i}^c = \frac{1}{\sqrt{T}} \sum_{r=1}^T f_r^c \varepsilon_{j,ir}.$$

Equation (B.78) allows us to compute:

$$\begin{aligned} \hat{\gamma}_{j,ii} &= \frac{1}{T} \sum_{t=1}^T \hat{\varepsilon}_{j,i,t}^2 \simeq \frac{1}{T} \sum_{t=1}^T \left[\varepsilon_{j,i,t} - \frac{1}{\sqrt{T}} (\tilde{w}_{j,i}^{c'} f_t^c + w_{j,i}^{s'} f_{j,t}^s) \right]^2 \\ &= \frac{1}{T} \sum_{t=1}^T \varepsilon_{j,i,t}^2 - \frac{2}{T\sqrt{T}} \sum_{t=1}^T \varepsilon_{j,i,t} (\tilde{w}_{j,i}^{c'} f_t^c + w_{j,i}^{s'} f_{j,t}^s) + \frac{1}{T^2} \sum_{t=1}^T (\tilde{w}_{j,i}^{c'} f_t^c + w_{j,i}^{s'} f_{j,t}^s)^2. \end{aligned}$$

Using $\frac{1}{\sqrt{T}} \sum_{t=1}^T \varepsilon_{j,i,t} f_t^c = O_p(1)$ and $\frac{1}{\sqrt{T}} \sum_{t=1}^T \varepsilon_{j,i,t} f_{j,t}^s = O_p(1)$ we get:

$$\hat{\gamma}_{j,ii} = \frac{1}{T} \sum_{t=1}^T \varepsilon_{j,i,t}^2 + O_p \left(\frac{1}{T} \right),$$

which implies:

$$\hat{\gamma}_{j,ii} = \frac{1}{T} \sum_{t=1}^T \varepsilon_{j,i,t}^2 + o_p \left(\frac{1}{\sqrt{T}} \right) = \gamma_{j,ii} + \frac{1}{\sqrt{T}} w_{j,i} + o_p \left(\frac{1}{\sqrt{T}} \right),$$

where

$$w_{j,i} = \frac{1}{\sqrt{T}} \sum_{t=1}^T (\varepsilon_{j,i,t}^2 - \gamma_{j,ii}) = O_p(1),$$

from Assumptions A.4 and A.7. Therefore, we have:

$$\hat{\Gamma}_j \simeq \Gamma_j + \frac{1}{\sqrt{T}} W_j.$$

where $\Gamma_j = \text{diag}(\gamma_{j,ii}, i = 1, \dots, N)$ and $W_j = \text{diag}(w_{j,i}, i = 1, \dots, N)$, for $j = 1, 2$.

d) Asymptotic expansion of $\frac{1}{N_j} \hat{\Lambda}'_j \hat{\Gamma}_j \hat{\Lambda}_j$ Let us define

$$\begin{aligned}\hat{\Omega}_j^* &:= \frac{1}{N_j} \left(\Lambda_j + \frac{1}{\sqrt{T}} G_j \right)' \hat{\Gamma}_j \left(\Lambda_j + \frac{1}{\sqrt{T}} G_j \right) \\ &= \frac{1}{N_j} \left(\Lambda_j + \frac{1}{\sqrt{T}} G_j \right)' \left(\Gamma_j + \frac{1}{\sqrt{T}} W_j \right) \left(\Lambda_j + \frac{1}{\sqrt{T}} G_j \right) \\ &= \frac{1}{N_j} \Lambda_j' \Gamma_j \Lambda_j + \hat{\Omega}_{j,I}^* + \hat{\Omega}_{j,II}^* + \hat{\Omega}_{j,III}^* + \hat{\Omega}_{j,II}'^* + \hat{\Omega}_{j,III}'^* + \hat{\Omega}_{j,IV}^* + \hat{\Omega}_{j,V}^*,\end{aligned}$$

where

$$\begin{aligned}\hat{\Omega}_{j,I}^* &= \frac{1}{N_j \sqrt{T}} \Lambda_j' W_j \Lambda_j = O_p \left(\frac{1}{\sqrt{NT}} \right), & \hat{\Omega}_{j,III}^* &= \frac{1}{N_j T} \Lambda_j' W_j G_j = O_p \left(\frac{1}{T} \right), \\ \hat{\Omega}_{j,IV}^* &= \frac{1}{N_j T} G_j' \Gamma_j G_j = O_p \left(\frac{1}{T} \right), & \hat{\Omega}_{j,V}^* &= \frac{1}{N_j T \sqrt{T}} G_j' W_j G_j = O_p \left(\frac{1}{T \sqrt{T}} \right).\end{aligned}$$

Moreover, similarly as for (B.73) we have:

$$\begin{aligned}\hat{\Omega}_{j,II}^* &= \frac{1}{N_j \sqrt{T}} \Lambda_j' \Gamma_j G_j = \left[\frac{1}{N_j} \Lambda_j' \Gamma_j \Lambda_j^s \left(\frac{1}{T} \sum_{t=1}^T f_{j,t}^s f_t^{c'} \right) \vdots 0_{(k_j \times k_j^s)} \right] + o_p \left(\frac{1}{\sqrt{T}} \right), \\ &= \frac{1}{\sqrt{T}} \left[\left(\frac{1}{N_j} \Lambda_j' \Gamma_j \Lambda_j^s \right) \left(\frac{1}{\sqrt{T}} F_j^s' F_c \right) \vdots 0_{(k_j \times k_j^s)} \right] + o_p \left(\frac{1}{\sqrt{T}} \right) = \frac{1}{\sqrt{T}} L_{2,j} + o_p \left(\frac{1}{\sqrt{T}} \right),\end{aligned}$$

where

$$L_{2,j} = \left[\left(\frac{1}{N} \Lambda_j' \Gamma_j \Lambda_j^s \right) \left(\frac{1}{\sqrt{T}} F_j^s' F_c \right) \vdots 0_{(k_j \times k_j^s)} \right].$$

Collecting the previous results, using $T/N = o_p(1)$, and defining $\Omega_j^* = \lim_{N \rightarrow \infty} \frac{1}{N} \Lambda_j' \Gamma_j \Lambda_j$ we get:

$$\begin{aligned}\hat{\Omega}_j^* &= \frac{1}{N} \Lambda_j' \Gamma_j \Lambda_j + \frac{1}{\sqrt{T}} (L_{2,j} + L'_{2,j}) + o_p \left(\frac{1}{\sqrt{T}} \right) \\ &= \Omega_j^* + \frac{1}{\sqrt{T}} (L_{2,j} + L'_{2,j}) + o_p \left(\frac{1}{\sqrt{T}} \right).\end{aligned}\tag{B.79}$$

Substituting equation (B.69) into $\frac{1}{N_j} \hat{\Lambda}'_j \hat{\Gamma}_j \hat{\Lambda}_j$, and using equation (B.79) we get:

$$\frac{1}{N_j} \hat{\Lambda}'_j \hat{\Gamma}_j \hat{\Lambda}_j = \hat{U}'_j \hat{\Omega}_j^* \hat{U}_j = \hat{U}'_j \left[\Omega_j^* + \frac{1}{\sqrt{T}} (L_{2,j} + L'_{2,j}) \right] \hat{U}_j + o_p \left(\frac{1}{\sqrt{T}} \right), \quad j = 1, 2.\tag{B.80}$$

e) Asymptotic expansion of $\hat{\Sigma}_U$

The estimator of $\Sigma_{u,j}$ is given in equation (4.8). Equation (B.76) allows to compute the asymptotic approximation of $\left(\frac{\hat{\Lambda}'_j \hat{\Lambda}_j}{N_j} \right)^{-1}$:

$$\left(\frac{\hat{\Lambda}'_j \hat{\Lambda}_j}{N_j} \right)^{-1} \simeq \hat{U}_j^{-1} \left[\Sigma_{\Lambda,j}^{-1} - \frac{1}{\sqrt{T}} \Sigma_{\Lambda,j}^{-1} (L_{1,j} + L'_{1,j}) \Sigma_{\Lambda,j}^{-1} \right] (\hat{U}'_j)^{-1}.\tag{B.81}$$

Substituting equations (B.81) and (B.80) into equation (4.8), we get:

$$\begin{aligned}
\hat{\Sigma}_{u,j} &\simeq \hat{U}_j^{-1} \left[\Sigma_{\Lambda,j}^{-1} - \frac{1}{\sqrt{T}} \Sigma_{\Lambda,j}^{-1} (L_{1,j} + L'_{1,j}) \Sigma_{\Lambda,j}^{-1} \right] \left[\Omega_j^* + \frac{1}{\sqrt{T}} (L_{2,j} + L'_{2,j}) \right] \\
&\quad \times \left[\Sigma_{\Lambda,j}^{-1} - \frac{1}{\sqrt{T}} \Sigma_{\Lambda,j}^{-1} (L_{1,j} + L'_{1,j}) \Sigma_{\Lambda,j}^{-1} \right] (\hat{U}'_j)^{-1} \\
&\simeq \hat{U}_j^{-1} \Sigma_{\Lambda,j}^{-1} \left[I - \frac{1}{\sqrt{T}} (L_{1,j} + L'_{1,j}) \Sigma_{\Lambda,j}^{-1} \right] \left[\Omega_j^* + \frac{1}{\sqrt{T}} (L_{2,j} + L'_{2,j}) \right] \left[I - \frac{1}{\sqrt{T}} \Sigma_{\Lambda,j}^{-1} (L_{1,j} + L'_{1,j}) \right] \Sigma_{\Lambda,j}^{-1} (\hat{U}'_j)^{-1} \\
&\simeq \hat{U}_j^{-1} \Sigma_{\Lambda,j}^{-1} \left[\Omega_j^* + \frac{1}{\sqrt{T}} (L_{2,j} + L'_{2,j}) - \frac{1}{\sqrt{T}} \Omega_j^* \Sigma_{\Lambda,j}^{-1} (L_{1,j} + L'_{1,j}) - \frac{1}{\sqrt{T}} (L_{1,j} + L'_{1,j}) \Sigma_{\Lambda,j}^{-1} \Omega_j^* \right] \Sigma_{\Lambda,j}^{-1} (\hat{U}'_j)^{-1},
\end{aligned}$$

which implies:

$$\hat{\Sigma}_{u,j} = \hat{U}_j^{-1} \Sigma_{u,j} (\hat{U}'_j)^{-1} + \frac{1}{\sqrt{T}} \hat{U}_j^{-1} L_{3,j} (\hat{U}'_j)^{-1} + o_p \left(\frac{1}{\sqrt{T}} \right),$$

where

$$L_{3,j} = \Sigma_{\Lambda,j}^{-1} \left[(L_{2,j} + L'_{2,j}) - \Omega_j^* \Sigma_{\Lambda,j}^{-1} (L_{1,j} + L'_{1,j}) - (L_{1,j} + L'_{1,j}) \Sigma_{\Lambda,j}^{-1} \Omega_j^* \right] \Sigma_{\Lambda,j}^{-1}. \quad (\text{B.82})$$

From equation (B.71) we have:

$$\begin{aligned}
\hat{\Sigma}_U &= \mu_N^2 \hat{\Sigma}_{u,1}^{(cc)} + \hat{\Sigma}_{u,2}^{(cc)} \\
&= \hat{\mathcal{H}}_c^{-1} [\mu_N^2 \Sigma_{u,1} + \Sigma_{u,2}]^{(cc)} (\hat{\mathcal{H}}'_c)^{-1} + \frac{1}{\sqrt{T}} \hat{\mathcal{H}}_c^{-1} (\mu_N^2 L_{3,1} + L_{3,2})^{(cc)} (\hat{\mathcal{H}}'_c)^{-1} + o_p \left(\frac{1}{\sqrt{T}} \right) \\
&= \hat{\mathcal{H}}_c^{-1} \Sigma_{U,N} (\hat{\mathcal{H}}'_c)^{-1} + \frac{1}{\sqrt{T}} \hat{\mathcal{H}}_c^{-1} (\mu_N^2 L_{3,1} + L_{3,2})^{(cc)} (\hat{\mathcal{H}}'_c)^{-1} + o_p \left(\frac{1}{\sqrt{T}} \right). \quad (\text{B.83})
\end{aligned}$$

This expansion, the convergence $\Sigma_{U,N} \rightarrow \Sigma_U(0)$ and the commutative property of the trace, imply equation (B.64).

f) Asymptotic expansion of $tr \left\{ \tilde{\Sigma}_{cc}^{-1} \hat{\Sigma}_U \right\}$

Results (B.65) and (B.83), and the commutative property of the trace, imply:

$$tr \left\{ \hat{\Sigma}_U \right\} = tr \left\{ \tilde{\Sigma}_{cc}^{-1} \Sigma_{U,N} \right\} + \frac{1}{\sqrt{T}} tr \left\{ \tilde{\Sigma}_{cc}^{-1} (\mu_N^2 L_{3,1} + L_{3,2})^{(cc)} \right\} + o_p \left(\frac{1}{\sqrt{T}} \right).$$

Noting that $L_{3,j} = O_p(1)$, for $j = 1, 2$, and recalling that $\tilde{\Sigma}_{cc} = I_{k^c} + O_p(1/\sqrt{T})$ and $\mu_N = \mu + o(1)$, the last equation can be further simplified to

$$tr \left\{ \hat{\Sigma}_U \right\} = tr \left\{ \tilde{\Sigma}_{cc}^{-1} \Sigma_{U,N} \right\} + \frac{1}{\sqrt{T}} tr \left\{ (\mu^2 L_{3,1} + L_{3,2})^{(cc)} \right\} + o_p \left(\frac{1}{\sqrt{T}} \right). \quad (\text{B.84})$$

Let us compute $L_{3,j}$ explicitly. From equation (B.75) we get:

$$\begin{aligned}
L_{1,j} &= \begin{bmatrix} \left(\frac{\Lambda_j^{c'} \Lambda_j^s}{N} \right) \left(\frac{1}{\sqrt{T}} F_j^{s'} F^c \right) & 0_{(k^c \times k_j^s)} \\ \left(\frac{\Lambda_j^{s'} \Lambda_j^s}{N} \right) \left(\frac{1}{\sqrt{T}} F_j^{s'} F^c \right) & 0_{(k_j^s \times k_j^s)} \end{bmatrix} \\
&= \begin{bmatrix} \Sigma_{\Lambda,j,cs} \left(\frac{1}{\sqrt{T}} F_j^{s'} F^c \right) & 0 \\ \Sigma_{\Lambda,j,ss} \left(\frac{1}{\sqrt{T}} F_j^{s'} F^c \right) & 0 \end{bmatrix} + O_p \left(\frac{1}{\sqrt{N}} \right) = \Sigma_{\Lambda,j} \begin{bmatrix} 0_{(k^c \times k^c)} & 0_{(k^c \times k_j^s)} \\ K_j' & 0_{(k_j^s \times k_j^s)} \end{bmatrix} + O_p \left(\frac{1}{\sqrt{N}} \right). \quad (\text{B.85})
\end{aligned}$$

Equation (B.85) implies:

$$\Omega_j^* \Sigma_{\Lambda,j}^{-1} L_{1,j} = L_{2,j} + O_p\left(\frac{1}{\sqrt{N}}\right). \quad (\text{B.86})$$

Substituting results (B.85) and (B.86) into equation (B.82) we get:

$$L_{3,j} = -\Sigma_{\Lambda,j}^{-1} \left[\Omega_j^* \Sigma_{\Lambda,j}^{-1} L'_{1,j} + L_{1,j} \Sigma_{\Lambda,j}^{-1} \Omega_j^* \right] \Sigma_{\Lambda,j}^{-1} = -\Sigma_{u,j} L'_{1,j} \Sigma_{\Lambda,j}^{-1} - \Sigma_{\Lambda,j}^{-1} L_{1,j} \Sigma_{u,j} + O_p\left(\frac{1}{\sqrt{N}}\right).$$

Moreover, noting that:

$$\Sigma_{\Lambda,j}^{-1} L_{1,j} = \begin{bmatrix} 0_{(k^c \times k^c)} & 0_{(k^c \times k_j^s)} \\ \left(\frac{1}{\sqrt{T}} F_j^s\right)' F^c & 0_{(k_j^s \times k_j^s)} \end{bmatrix} + O_p\left(\frac{1}{\sqrt{N}}\right),$$

we get:

$$\Sigma_{\Lambda,j}^{-1} L_{1,j} \Sigma_{u,j} = \begin{bmatrix} 0_{(k^c \times k^c)} & 0_{(k^c \times k_j^s)} \\ * & * \end{bmatrix} + O_p\left(\frac{1}{\sqrt{N}}\right). \quad (\text{B.87})$$

Equation (B.87) implies:

$$(L_{3,j})^{(cc)} = O_p\left(\frac{1}{\sqrt{N}}\right). \quad (\text{B.88})$$

Finally, substituting result (B.88) into equation (B.84), equation (B.63) follows, which concludes the proof of part (i).

B.5.2 Proof of part (ii)

In order to prove Theorem 2 (ii), we consider the behaviour 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:

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

Since the eigenvalues are continuous functions of the matrix, and \hat{R} converges to R in probability, we have:

$$\sum_{\ell=1}^{k^c} \hat{\rho}_\ell = \sum_{\ell=1}^{k^c} \rho_\ell + o_p(1). \quad (\text{B.89})$$

Moreover, we rely on the following Lemma:

LEMMA B.2. *Under the alternative hypothesis $H(r)$, with $r < k^c$, we have $\|\hat{\Sigma}_U\| \leq C$, with probability approaching (w.p.a.) 1, for a constant $C > 0$.*

Proof: See Online Appendix, Section C.4.

From (B.89) and Lemma B.2, we have:

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

Under $H(r)$, we have $r < k^c$ canonical correlations that are equal to 1, while the other ones are strictly smaller than 1.

Thus, $\sum_{\ell=1}^{k^c} \rho_\ell - k^c < 0$. Then, from Lemma B.2 (ii) we get:

$$\tilde{\xi}(k^c) \leq -N\sqrt{T}c_1, \quad w.p.a. 1, \quad (\text{B.90})$$

for a constant $c_1 > 0$. The conclusion follows.

Q.E.D.

B.6 Proof of Proposition 5

Let us define the events $\Omega_{r,\alpha_{N,T}} \equiv \{\tilde{\xi}(r) < z_{\alpha_{N,T}}\}$, for $r = 1, \dots, \underline{k}$, and their complementary events $\Omega_{r,\alpha_{N,T}}^c = \{\tilde{\xi}(r) \geq z_{\alpha_{N,T}}\}$. For any integer k^* we have:

$$\begin{aligned} \{\hat{k}^c = k^*\} &= \Omega_{k^*,\alpha_{N,T}}^c, \quad \text{if } k^* = \underline{k}, \\ &= \left(\bigcap_{r=k^*+1}^{\underline{k}} \Omega_{r,\alpha_{N,T}} \right) \cap \Omega_{k^*,\alpha_{N,T}}^c, \quad \text{if } 0 < k^* < \underline{k}, \\ &= \bigcap_{r=k^*+1}^{\underline{k}} \Omega_{r,\alpha_{N,T}}, \quad \text{if } k^* = 0. \end{aligned}$$

We prove Proposition 5 by distinguishing three cases according to the true number of common factors: $k_0^c = \underline{k}$, $0 < k_0^c \leq \underline{k}$, $k_0^c = 0$. Moreover, we use the convergence results:

$$P(\Omega_{r,\alpha_{N,T}}) \rightarrow 1, \quad r > k_0^c, \quad (\text{B.91})$$

$$P(\Omega_{r,\alpha_{N,T}}) \rightarrow 0, \quad r = k_0^c, \quad (\text{B.92})$$

which are proved at the end of the section.

B.6.1 Case $k_0^c = \underline{k}$

We have $P(\hat{k}^c = k_0^c) = P(\Omega_{\underline{k},\alpha_{N,T}}^c) = 1 - P(\Omega_{\underline{k},\alpha_{N,T}}) \rightarrow 1$, from equation (B.92).

B.6.2 Case $0 < k_0^c < \underline{k}$

The event $\{\hat{k}^c = k_0^c\}$ can be written as:

$$\{\hat{k}^c = k_0^c\} = \left(\bigcap_{r=k_0^c+1}^{\underline{k}} \Omega_{r,\alpha_{N,T}} \right) \cap \Omega_{k_0^c,\alpha_{N,T}}^c.$$

The events $\Omega_{r,\alpha_{N,T}}$, for $r = k_0^c + 1, \dots, \underline{k}$, have all probability tending to 1 from equation (B.91), and so do events $\bigcap_{r=k_0^c+1}^{\underline{k}} \Omega_{r,\alpha_{N,T}}$ and $\left(\bigcap_{r=k_0^c+1}^{\underline{k}} \Omega_{r,\alpha_{N,T}} \right) \cup \Omega_{k_0^c,\alpha_{N,T}}^c$. Moreover, $P(\Omega_{k_0^c,\alpha_{N,T}}^c) = 1 - P(\Omega_{k_0^c,\alpha_{N,T}}) \rightarrow 1$ from equation (B.92). Thus, we get:

$$\begin{aligned} P(\hat{k}^c = k_0^c) &= P \left[\left(\bigcap_{r=k_0^c+1}^{\underline{k}} \Omega_{r,\alpha_{N,T}} \right) \cap \Omega_{k_0^c,\alpha_{N,T}}^c \right] \\ &= P \left(\bigcap_{r=k_0^c+1}^{\underline{k}} \Omega_{r,\alpha_{N,T}} \right) + P(\Omega_{k_0^c,\alpha_{N,T}}^c) - P \left[\left(\bigcap_{r=k_0^c+1}^{\underline{k}} \Omega_{r,\alpha_{N,T}} \right) \cup \Omega_{k_0^c,\alpha_{N,T}}^c \right] \rightarrow 1. \end{aligned}$$

B.6.3 Case $k_0^c = 0$

We have:

$$P(\hat{k}^c = k_0^c) = P\left(\bigcap_{r=1}^{\underline{k}} \Omega_{r,\alpha_{N,T}}\right) \rightarrow 1,$$

because the events $\Omega_{r,\alpha_{N,T}}$, for $r = 1, \dots, \underline{k}$, have all probability tending to 1, from equation (B.91).

B.6.4 Proof of result (B.91)

We have:

$$P(\Omega_{r,\alpha_{N,T}}) = P\left(\frac{\tilde{\xi}(r)}{N\sqrt{T}} < \frac{z_{\alpha_{N,T}}}{N\sqrt{T}}\right).$$

Since $r > k_0^c$, from equation (B.90) we have $\frac{\tilde{\xi}(r)}{N\sqrt{T}} \leq -c_1$, w.p.a. 1, for a constant c_1 . By Condition (ii), we have $\frac{z_{\alpha_{N,T}}}{N\sqrt{T}} \rightarrow 0$. Then, $P(\Omega_{r,\alpha_{N,T}}) \rightarrow 1$ follows.

B.6.5 Proof of result (B.92)

If $r = k_0^c$, from Theorem 2 (ii) we have $\tilde{\xi}(r) \xrightarrow{d} N(0, 1)$. Moreover, since $\alpha_{N,T} \rightarrow 0$ by Condition (i), we have $z_{\alpha_{N,T}} \leq z_{\alpha^*}$ for large N, T , for any given $\alpha^* \in (0, 1)$. Thus:

$$P(\Omega_{r,\alpha_{N,T}}) = P(\tilde{\xi}(r) < z_{\alpha_{N,T}}) \leq P(\tilde{\xi}(r) < z_{\alpha^*}) \rightarrow \alpha^*.$$

Thus, we have $\liminf_{N,T \rightarrow \infty} P(\Omega_{r,\alpha_{N,T}}) \leq \alpha^*$, for any $\alpha^* \in (0, 1)$. It follows $P(\Omega_{r,\alpha_{N,T}}) \rightarrow 0$. *Q.E.D.*

B.7 Proof of Proposition 6

Let us re-write the model for the high frequency observables $x_{m,t}^H$, where $m = 1, \dots, M$, and $t = 1, \dots, T$ in equation (2.1) as:

$$\begin{aligned} x_{m,t}^H &= \Lambda_{HC} g_{m,t}^C + \Lambda_H g_{m,t}^H + e_{m,t}^H, = \Lambda_1 g_{m,t} + e_{m,t}^H, \\ &= \hat{\Lambda}_1 \hat{\mathcal{U}}_1^{-1} g_{m,t} - \left(\hat{\Lambda}_1 \hat{\mathcal{U}}_1^{-1} - \Lambda_1\right) g_{m,t} + e_{m,t}^H, \end{aligned} \quad (\text{B.93})$$

where $g_{m,t} = [g_{m,t}^C; g_{m,t}^H]'$, $\Lambda_1 = [\Lambda_{HC}; \Lambda_H] = [\Lambda_1^c; \Lambda_1^s]$, $\hat{\Lambda}_1 = [\hat{\Lambda}_{HC}; \hat{\Lambda}_H] = [\hat{\Lambda}_1^c; \hat{\Lambda}_1^s]$, and $\hat{\mathcal{U}}_1$ has been defined in equation (B.71). Let us also define the estimator $\hat{g}_{m,t} = [\hat{g}_{m,t}^C; \hat{g}_{m,t}^H]'$ as in equation (3.6):

$$\hat{g}_{m,t} = \begin{bmatrix} \hat{g}_{m,t}^C \\ \hat{g}_{m,t}^H \end{bmatrix} = \left(\hat{\Lambda}'_1 \hat{\Lambda}_1\right)^{-1} \hat{\Lambda}'_1 x_{m,t}^H, \quad m = 1, \dots, M, \quad t = 1, \dots, T. \quad (\text{B.94})$$

Substituting equation (B.93) into equation (B.94), and rearranging terms, we get:

$$\hat{g}_{m,t} = \hat{\mathcal{U}}_1^{-1} g_{m,t} - \left(\frac{\hat{\Lambda}'_1 \hat{\Lambda}_1}{N_H}\right)^{-1} \frac{1}{N_H} \hat{\Lambda}'_1 \left(\hat{\Lambda}_1 \hat{\mathcal{U}}_1^{-1} - \Lambda_1\right) g_{m,t} + \left(\frac{\hat{\Lambda}'_1 \hat{\Lambda}_1}{N_H}\right)^{-1} \frac{1}{N_H} \hat{\Lambda}'_1 e_{m,t}^H. \quad (\text{B.95})$$

From equations (B.75) and (B.76) we have:

$$\frac{\hat{\Lambda}'_1 \hat{\Lambda}_1}{N_H} = \hat{\mathcal{U}}_1' \Sigma_{\Lambda,1} \hat{\mathcal{U}}_1 + O_p\left(\frac{1}{\sqrt{T}}\right),$$

which implies:

$$\left(\frac{\hat{\Lambda}'_1 \hat{\Lambda}_1}{N_H}\right)^{-1} = \hat{\mathcal{U}}_1^{-1} \Sigma_{\Lambda,1}^{-1} (\hat{\mathcal{U}}'_1)^{-1} + O_p\left(\frac{1}{\sqrt{T}}\right). \quad (\text{B.96})$$

From equations (B.66) - (B.70) we get:

$$\hat{\Lambda}_1 \hat{\mathcal{U}}_1^{-1} - \Lambda_1 \simeq \frac{1}{\sqrt{T}} G_1, \quad (\text{B.97})$$

where

$$G_1 = \begin{bmatrix} G_1^c \\ G_1^s \end{bmatrix}, \quad (\text{B.98})$$

with

$$G_1^c = \frac{1}{\sqrt{T}} \bar{e}^{H'} \bar{g}^C + \Lambda_H \left(\frac{1}{\sqrt{T}} \bar{g}^{H'} \bar{g}^C \right), \quad (\text{B.99})$$

$$G_1^s = \frac{1}{\sqrt{T}} \bar{e}^{H'} \bar{g}^H, \quad (\text{B.100})$$

$\bar{e}^H = [\bar{e}_1^H, \dots, \bar{e}_T^H]'$, $\bar{g}^C = [\bar{g}_1^C, \dots, \bar{g}_T^C]'$ and $\bar{g}^H = [\bar{g}_1^H, \dots, \bar{g}_T^H]'$. Moreover, we have:

$$\hat{\Lambda}_1 \simeq \Lambda_1 \hat{\mathcal{U}}_1 + \frac{1}{\sqrt{T}} G_1 \hat{\mathcal{U}}_1. \quad (\text{B.101})$$

From equations (B.97) and (B.101) it follows:

$$\begin{aligned} \frac{1}{N_H} \hat{\Lambda}'_1 (\hat{\Lambda}_1 \hat{\mathcal{U}}_1^{-1} - \Lambda_1) &\simeq \frac{1}{N_H} \left(\Lambda_1 \hat{\mathcal{U}}_1 + \frac{1}{\sqrt{T}} G_1 \hat{\mathcal{U}}_1 \right)' \frac{1}{\sqrt{T}} G_1 \\ &= \frac{1}{N_H \sqrt{T}} \hat{\mathcal{U}}'_1 \Lambda'_1 G_1 + \frac{1}{N_H T} \hat{\mathcal{U}}'_1 G'_1 G_1. \end{aligned} \quad (\text{B.102})$$

Equations (B.96) and (B.102) allow to express the second term in the r.h.s. of equation (B.95) as:

$$\left(\frac{\hat{\Lambda}'_1 \hat{\Lambda}_1}{N_H}\right)^{-1} \frac{1}{N_H} \hat{\Lambda}'_1 (\hat{\Lambda}_1 \hat{\mathcal{U}}_1^{-1} - \Lambda_1) g_{m,t} \simeq \hat{\mathcal{U}}_1^{-1} \Sigma_{\Lambda,1}^{-1} \frac{1}{N_H \sqrt{T}} \Lambda'_1 G_1 g_{m,t} + \hat{\mathcal{U}}_1^{-1} \Sigma_{\Lambda,1}^{-1} \frac{1}{N_H T} G'_1 G_1 g_{m,t}. \quad (\text{B.103})$$

From equation (B.73) we have:

$$\frac{1}{N_H \sqrt{T}} \Lambda'_1 G_1 = \left[\left(\frac{\Lambda'_1 \Lambda_H}{N_H} \right) \frac{1}{T} \sum_{t=1}^T \bar{g}_t^H \bar{g}_t^{C'} \quad \vdots \quad 0_{(k_1 \times k^H)} \right] + O_p\left(\frac{1}{\sqrt{N_H T}}\right), \quad (\text{B.104})$$

where $k_1 = k^C + k^H$. From equation (B.98) we have:

$$\frac{1}{N_H T} G'_1 G_1 = \frac{1}{N_H T} \begin{bmatrix} G_1^{c'} G_1^c & G_1^{c'} G_1^s \\ G_1^{s'} G_1^c & G_1^{s'} G_1^s \end{bmatrix}. \quad (\text{B.105})$$

Equation (B.99) implies:

$$\begin{aligned}
\frac{1}{N_H T} G_1^{c'} G_1^c &= \frac{1}{N_H T} \left[\frac{1}{\sqrt{T}} \bar{e}^{H'} \bar{g}^C + \Lambda_H \left(\frac{1}{\sqrt{T}} \bar{g}^{H'} \bar{g}^C \right) \right]' \left[\frac{1}{\sqrt{T}} \bar{e}^{H'} \bar{g}^C + \Lambda_H \left(\frac{1}{\sqrt{T}} \bar{g}^{H'} \bar{g}^C \right) \right] \\
&= \frac{1}{N_H T^2} \bar{g}^{C'} \bar{e}^H \bar{e}^{H'} \bar{g}^C + \frac{1}{N_H T \sqrt{T}} \bar{g}^{C'} \bar{e}^H \Lambda_H \left(\frac{1}{\sqrt{T}} \bar{g}^{H'} \bar{g}^C \right) \\
&\quad + \frac{1}{N_H T \sqrt{T}} \left(\frac{1}{\sqrt{T}} \bar{g}^{H'} \bar{g}^C \right)' \Lambda_H' \bar{e}^{H'} \bar{g}^C + \frac{1}{N_H T} \left(\frac{1}{\sqrt{T}} \bar{g}^{H'} \bar{g}^C \right)' \Lambda_H' \Lambda_H \left(\frac{1}{\sqrt{T}} \bar{g}^{H'} \bar{g}^C \right) \\
&= O_p \left(\frac{1}{T} \right), \tag{B.106}
\end{aligned}$$

where the last equality follows from the assumption $T/N_H = o(1)$. Equation (B.106) and the assumption $\sqrt{N_H}/T = o(1)$ imply:

$$\frac{1}{N_H T} G_1^{c'} G_1^c = o_p \left(\frac{1}{\sqrt{N_H}} \right).$$

Similar arguments applied to the other blocks of the matrix in the r.h.s. of (B.105) yield:

$$\frac{1}{N_H T} G_1' G_1 = o_p \left(\frac{1}{\sqrt{N_H}} \right). \tag{B.107}$$

Substituting equations (B.104) and (B.107) into equation (B.103) we get:

$$\left(\frac{\hat{\Lambda}'_1 \hat{\Lambda}_1}{N_H} \right)^{-1} \frac{1}{N_H} \hat{\Lambda}'_1 \left(\hat{\Lambda}_1 \hat{\mathcal{U}}_1^{-1} - \Lambda_1 \right) g_{m,t} \simeq \hat{\mathcal{U}}_1^{-1} \Sigma_{\Lambda,1}^{-1} \left(\frac{\Lambda'_1 \Lambda_H}{N_H} \right) \left(\frac{1}{T} \sum_{t=1}^T \bar{g}_t^H \bar{g}_t^{C'} \right) g_{m,t}^C + o_p \left(\frac{1}{\sqrt{N_H}} \right). \tag{B.108}$$

Let us now focus on the third term in the r.h.s. of equation (B.95). From equation (B.101) we have:

$$\begin{aligned}
\frac{1}{N_H} \hat{\Lambda}'_1 e_{m,t}^H &\simeq \frac{1}{N_H} \left(\Lambda_1 \hat{\mathcal{U}}_1 + \frac{1}{\sqrt{T}} G_1 \hat{\mathcal{U}}_1 \right)' e_{m,t}^H \\
&= \hat{\mathcal{U}}_1' \frac{1}{N_H} \Lambda_1' e_{m,t}^H + \hat{\mathcal{U}}_1' \frac{1}{N_H \sqrt{T}} G_1' e_{m,t}^H. \tag{B.109}
\end{aligned}$$

The second term in the r.h.s. of equation (B.109) can be written as:

$$\frac{1}{N_H \sqrt{T}} G_1' e_{m,t}^H = \frac{1}{N_H \sqrt{T}} \begin{bmatrix} G_1^{c'} e_{m,t}^H \\ G_1^{s'} e_{m,t}^H \end{bmatrix}. \tag{B.110}$$

Using equation (B.99) we get:

$$\begin{aligned}
\frac{1}{N_H \sqrt{T}} G_1' e_{m,t}^H &= \frac{1}{N_H T} \bar{g}^{C'} \bar{e}^H e_{m,t}^H + \frac{1}{N_H \sqrt{T}} \left(\frac{1}{\sqrt{T}} \bar{g}^{C'} \bar{g}^H \right) \Lambda_H' e_{m,t}^H \\
&= O_p \left(\frac{1}{\sqrt{N_H T}} \right). \tag{B.111}
\end{aligned}$$

Equation (B.100) implies:

$$\frac{1}{N_H \sqrt{T}} G_1^{s'} e_{m,t}^H = \frac{1}{N_H T} \bar{g}^{H'} \bar{e}^H e_{m,t}^H = O_p \left(\frac{1}{\sqrt{N_H T}} \right). \tag{B.112}$$

Substituting results (B.111) and (B.112) into equations (B.110) and (B.109) we get:

$$\frac{1}{N_H} \hat{\Lambda}'_1 e_{m,t}^H = \hat{U}'_1 \frac{1}{N_H} \Lambda'_1 e_{m,t}^H + O_p \left(\frac{1}{\sqrt{N_H T}} \right). \quad (\text{B.113})$$

Substituting results (B.96), (B.108), and (B.113) into equation (B.95), and rearranging terms we get:

$$\hat{U}_1 \hat{g}_{m,t} - g_{m,t} = -\Sigma_{\Lambda,1}^{-1} \left(\frac{\Lambda'_1 \Lambda_H}{N_H} \right) \left(\frac{1}{T} \bar{g}^{H'} \bar{g}^C \right) g_{m,t}^C + \Sigma_{\Lambda,1}^{-1} \frac{1}{N_H} \Lambda'_1 e_{m,t}^H + o_p \left(\frac{1}{\sqrt{N_H}} \right). \quad (\text{B.114})$$

Let us denote the last k^H columns of matrix $\Sigma_{\Lambda,1}$ as $\Sigma_{\Lambda,1}^{(\cdot, s)}$. The term $\frac{\Lambda'_1 \Lambda_H}{N_H}$ in equation (B.114) can be written as:

$$\begin{aligned} \frac{\Lambda'_1 \Lambda_H}{N_H} &= \Sigma_{\Lambda,1}^{(\cdot, s)} + \frac{1}{N_H} \sum_{i=1}^{N_H} \lambda_{1,i} \lambda'_{H,i} - \Sigma_{\Lambda,1}^{(\cdot, s)} \\ &= \Sigma_{\Lambda,1}^{(\cdot, s)} + O_p \left(\frac{1}{\sqrt{N_H}} \right), \end{aligned} \quad (\text{B.115})$$

where the last equality follows from Assumption A.2. Equation (B.115) implies:

$$\Sigma_{\Lambda,1}^{-1} \left(\frac{\Lambda'_1 \Lambda_H}{N_H} \right) = \begin{bmatrix} 0_{(k^C \times k^H)} \\ I_{k^H} \end{bmatrix} + O_p \left(\frac{1}{\sqrt{N_H}} \right). \quad (\text{B.116})$$

Substituting equation (B.116) into equation (B.114) we have:

$$\hat{U}_1 \hat{g}_{m,t} - g_{m,t} = - \begin{bmatrix} 0_{(k^C \times k^H)} \\ I_{k^H} \end{bmatrix} \left(\frac{1}{T} \bar{g}^{H'} \bar{g}^C \right) g_{m,t}^C + \Sigma_{\Lambda,1}^{-1} \frac{1}{N_H} \Lambda'_1 e_{m,t}^H + o_p \left(\frac{1}{\sqrt{N_H}} \right). \quad (\text{B.117})$$

Recalling the expression of \hat{U}_1 from equation (B.71):

$$\hat{U}_1 = \begin{bmatrix} \hat{\mathcal{H}}_c & 0 \\ 0 & \hat{\mathcal{H}}_{s,1} \end{bmatrix},$$

from equation (B.117) we get the asymptotic expansions:

$$\hat{\mathcal{H}}_c \hat{g}_{m,t}^C - g_{m,t}^C \simeq \left[\Sigma_{\Lambda,1}^{-1} \frac{1}{N_H} \Lambda'_1 e_{m,t}^H \right]^{(C)}, \quad (\text{B.118})$$

$$\hat{\mathcal{H}}_{1,s} \hat{g}_{m,t}^H - g_{m,t}^H \simeq - \left(\frac{1}{T} \bar{g}^{H'} \bar{g}^C \right) g_{m,t}^C + \left[\Sigma_{\Lambda,1}^{-1} \frac{1}{N_H} \Lambda'_1 e_{m,t}^H \right]^{(H)}, \quad (\text{B.119})$$

where $\left[\Sigma_{\Lambda,1}^{-1} \frac{1}{N_H} \Lambda'_1 e_{m,t}^H \right]^{(C)}$ and $\left[\Sigma_{\Lambda,1}^{-1} \frac{1}{N_H} \Lambda'_1 e_{m,t}^H \right]^{(H)}$ denote the upper k^C rows, resp. the lower k^H rows, of vector $\Sigma_{\Lambda,1}^{-1} \frac{1}{N_H} \Lambda'_1 e_{m,t}^H$. Since $\bar{g}^{C'} \bar{g}^C / T = I_{k^C} + o_p(1)$, we can rewrite equation (B.119) as:

$$\hat{\mathcal{H}}_{1,s} \hat{g}_{m,t}^H - (g_{m,t}^H - (\bar{g}^{H'} \bar{g}^C) (\bar{g}^{C'} \bar{g}^C)^{-1} g_{m,t}^C) \simeq \left[\Sigma_{\Lambda,1}^{-1} \frac{1}{N_H} \Lambda'_1 e_{m,t}^H \right]^{(H)}. \quad (\text{B.120})$$

From Assumption A.8 we have:

$$\frac{1}{\sqrt{N_H}} \Lambda'_1 e_{m,t}^H \xrightarrow{d} N(0, \Omega_{\Lambda,m}^*), \quad (\text{B.121})$$

where

$$\Omega_{\Lambda,m}^* = \lim_{N_H \rightarrow \infty} \frac{1}{N_H} \sum_{i=1}^{N_H} \sum_{\ell=1}^{N_H} \lambda_{1,i} \lambda'_{1,\ell} Cov(e_{m,t}^{i,H}, e_{m,t}^{\ell,H}).$$

Equations (B.118) and (B.121) imply:

$$\sqrt{N_H} \left(\hat{\mathcal{H}}_c \hat{g}_{m,t}^C - g_{m,t}^C \right) \xrightarrow{d} N \left(0, \left[\Sigma_{\Lambda,1}^{-1} \Omega_{\Lambda,m}^* \Sigma_{\Lambda,1}^{-1} \right]^{(CC)} \right).$$

Similarly, equation (B.120) and (B.121) imply:

$$\sqrt{N_H} \left[\hat{\mathcal{H}}_{1,s} \hat{g}_{m,t}^H - (g_{m,t}^H - (\bar{g}^{H'} \bar{g}^C)(\bar{g}^{C'} \bar{g}^C)^{-1}) g_{m,t}^C \right] \xrightarrow{d} N \left(0, \left[\Sigma_{\Lambda,1}^{-1} \Omega_{\Lambda,m}^* \Sigma_{\Lambda,1}^{-1} \right]^{(HH)} \right).$$

Q.E.D.