University of Cyprus
Department of Economics

**Working Paper 19-2018**

# *Inference in Group Factor Models with an Application to Mixed Frequency Data*

**E. Andreou, P. Gagliardini, E. Ghysels, M. Rubin**

Department of Economics, University of Cyprus, P.O. Box 20537, 1678 Nicosia, Cyprus
Tel.: +357-22893700, Fax: +357-22895028, Web site: http://www.ucy.ac.cy/econ/en

# Inference in Group Factor Models with an Application to Mixed Frequency Data*

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

This version: January, 2019

**Abstract**

We derive asymptotic properties of estimators and test statistics to determine - in a grouped data setting - common versus group-specific factors. Despite the fact that our test statistic for the number of common factors, under the null, involves a parameter at the boundary (related to unit canonical correlations) we derive a parameter-free asymptotic Gaussian distribution. The group factor setting applies to mixed frequency data. As an empirical illustration we address the question whether Industrial Production (IP) is still the dominant factor driving the U.S. economy using a mixed-frequency data panel of IP and non-IP sectors. We find that a single common factor explains 85% of IP output growth and 60% of total GDP growth despite the diminishing role of manufacturing.

**Keywords:** Large Panel, Unobservable pervasive factors, Mixed frequency, Canonical correlations, Output growth

**JEL Codes:** C23, C38, C55, C12, E32

*We thank K. A. Aastveit, G. Barone-Adesi, M. Deistler, M. Del Negro, R. Engle, D. Giannone, T. Götz, K. Horl'ovà, C. Hurlin, A. Levchenko, S. Ng, A. Onatski, C. Pérignon, G. Urga, M. Watson and B. Werker for their useful comments. We also thank three referees for insightful comments which helped us to significantly improve the paper. The first author would like to acknowledge that this work has been co-funded by the European Research Council (ERC) under the European Union Horizon 2020 research and innovation programme Proof of Concept Grant Agreement 640924 as well as the Cyprus Research Promotion Foundation and the European Regional Development Fund research project EXCELLENCE/1216/0074. The second author gratefully acknowledges the Swiss National Science Foundation (SNSF) for Grant 105218 162633.

[†]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

The econometric problem we solve can be described as follows. Suppose one has two sets of unobservable factors, say $h_{1,t}$, $h_{2,t}$, estimated from two separate panels of observables, 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 derive its asymptotic distribution for a large $T$ and large $N$ panel data in the context of large scale approximate factor models in the spirit of Bai and Ng (2002), Stock and Watson (2002), and Bai (2003). The specific focus of the paper is the setting of group factor models.[1] We start with a setup which identifies factors common across panels. Our main theoretical contribution is an inference procedure for the number of common and group-specific factors based on canonical correlation analysis of the principal components (PCs) estimates on each subgroup. What complicates the asymptotics is the fact that we deal with estimated factors, i.e. the first stage estimation error affects the subsequent canonical correlation analysis. In fact the asymptotics are non-standard in terms of convergence rates and involve a non-trivial bias correction. We also propose estimators for the common and group-specific factors.[2] The inference procedure is general in scope and also of interest in many applications other than the one considered in the current paper.

As a specific application of group factor models, we study panels of data sampled at different frequencies and study the role of Industrial Production (IP) sectors in the U.S. economy. 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 U.S. economy. We have fairly extensive data on U.S. industrial production. They consist 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. On the other hand, contrary to IP, we do not have monthly or quarterly data for the cross-section of U.S. output across non-

[1]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. There exists a number of papers on group factor models, also referred to as multilevel or hierarchical factor models. Most are rooted in the statistics literature and deal with large $T$ and finite cross-sections (e.g. Tucker (1958), Flury (1984), Schott (1991), Gregory and Head (1999), and Kose, Otrok, and Whiteman (2008)). Goyal, Pérignon, and Villa (2008) extend the classical group factor models to approximate group factor models, but do not derive analytically any asymptotic results.

[2]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 $N = \min_j (N_j)$, $j = L, H$. To the best of our knowledge, the existing literature does not give an asymptotic treatment of group factor models in a large dimensional setting.

IP sectors, but we do so on an annual basis. Indeed, the U.S. Bureau of Economic Analysis provides Gross Domestic Product (GDP) and Gross Output by industry - not only IP sectors - annually. Hence, 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 $N_L \times T$. We allow for the presence of three types of unobservable factors: (1) those which explain variations in both panels/groups - say $g^C$, and therefore are common factors, (2) group-specific factors - namely (a) those exclusively pertaining to IP sector movements - say $g^H$, and (b) those exclusively affecting non-IP, denoted by $g^L$.

We find that a single common factor explains around 85% 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 but related to sectors such as Professional and Business services, Construction and Government, drives GDP growth fluctuations. 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.

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). In the Online Appendix (henceforth OA), Section C, we provide the proofs of lemmas and in Section D, we elaborate on a technical assumption (Assumption B.1). The OA, Section E contains additional theoretical results on identification and estimation, an extensive description of the dataset used in the empirical application, and additional empirical results. Finally, the OA, Section F, contains the details about the Monte Carlo simulation design and results.[3]

## 2 Identification in Group Factor Models

We use the following notation for the generic group factor model setting:

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} \Lambda_1^c & \Lambda_1^s & 0 \\ \Lambda_2^c & 0 & \Lambda_2^s \end{bmatrix} \begin{bmatrix} f_t^c \\ f_{1,t}^s \\ f_{2,t}^s \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix}, \qquad (2.1)$$

where $y_{j,t} = [y_{j,1t}, ..., y_{j,N_jt}]'$, $\Lambda_j^c = [\lambda_{j,1}^c, ..., \lambda_{j,N_j}^c]'$, $\Lambda_j^s = [\lambda_{j,1}^s, ..., \lambda_{j,N_j}^s]'$ and $\varepsilon_{j,t} = [\varepsilon_{j,1t}, ..., \varepsilon_{j,N_jt}]'$, 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 respectively $k^c$, $k_1^s$ and $k_2^s$. In the absence of common factors, we set $k^c = 0$, while in cases without 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:

$$E\begin{bmatrix} f_t^c \\ f_{1,t}^s \\ f_{2,t}^s \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \quad \text{and} \quad V\begin{bmatrix} f_t^c \\ f_{1,t}^s \\ f_{2,t}^s \end{bmatrix} = \begin{bmatrix} I_{k^c} & 0 & 0 \\ 0 & I_{k_1^s} & \Phi \\ 0 & \Phi' & I_{k_2^s} \end{bmatrix}, \qquad (2.2)$$

where $I_k$ denotes the identity matrix of order $k$. We allow for a non-zero covariance $\Phi$ between 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 (2.1) - (2.2) allows only for separate rotations among the components of $f_{1,t}^s$, among those of $f_{2,t}^s$, and finally those of $f_t^c$. The rotational invariance of model (2.1) - (2.2) therefore maintains the interpretation of common and group-specific factors.[4] Finally, it should also be noted that Assumption A.3 implies that full-rank

[3]The OA, Sections C, D, E, and F are available at `https://sites.google.com/site/mircorubin/`.

[4]More formally, Proposition E.10 in Appendix E.1 deals with the identification of factor spaces for given dimensions $k_c$, $k_1^s$, and $k_2^s$. Proposition E.10 is implied by Proposition 1 in Wang (2012).

conditions hold for the loading matrices in each group of model (2.1) with $N_j$ large.

We consider the generic setting of equation (2.1) and let $k_j = k^c + k_j^s$, for $j = 1, 2$, be the dimensions of the pervasive 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 $h_{j,t} := (f_t^c, f_{j,t}^s)'$, $j = 1, 2$, $t = 1, ..., T$, and define their variance and covariance matrices: $V_{j\ell} := E(h_{j,t}h'_{\ell,t})$, $j, \ell = 1, 2$. From (2.2) we have $V_{jj} = I_{k_j}$ for $j = 1, 2$. We want to show that the factor space dimensions $k^c$, $k_1^s$, $k_2^s$ are identifiable using canonical correlation analysis applied to $h_{1,t}$ and $h_{2,t}$. In particular, we want to propose a constructive identification strategy for these dimensions and the corresponding factor spaces using canonical correlations and directions. 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 $\rho_\ell$, $\ell = 1, ..., \underline{k}$, denote the ordered canonical correlations between $h_{1,t}$ and $h_{2,t}$. The $\underline{k}$ largest eigenvalues of matrices $R = V_{11}^{-1}V_{12}V_{22}^{-1}V_{21}$, and $R^* = V_{22}^{-1}V_{21}V_{11}^{-1}V_{12}$, are the same, and are equal to the squared canonical correlations $\rho_\ell^2$, $\ell = 1, ..., \underline{k}$ between $h_{1,t}$ and $h_{2,t}$. The associated eigenvectors $w_{1,\ell}$ (resp. $w_{2,\ell}$), with $\ell = 1, ..., \underline{k}$, of matrix $R$ (resp. $R^*$) standardized such that $w'_{1,\ell}V_{11}w_{1,\ell} = 1$ (resp. $w'_{2,\ell}V_{22}w_{2,\ell} = 1$) are the canonical directions which allow to construct the canonical variables $w'_{1,\ell}h_{1,t}$ (resp. $w'_{2,\ell}h_{2,t}$).
The next Proposition deals with determining $k^c$, the number of common factors, using canonical correlations between the vectors $h_{1,t}$ and $h_{2,t}$ which are unobserved and estimated by principal components.

**PROPOSITION 1.** *Under Assumption A.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 $f_{j,t}^s = W_j^{s\prime}h_{j,t}$ (up to a rotation matrix) for $j = 1, 2$.*

**Proof**: See Appendix X.3.

Proposition 1 shows that the number of common factors $k^c$, the common factor space spanned by $f_t^c$, and the spaces spanned by group-specific factors, can be identified from the canonical correlations and canonical variables of $h_{1,t}$ and $h_{2,t}$. Therefore, the factor space dimensions $k^c$, $k_j^s$, and factors $f_t^c$ and $f_{j,t}^s$, $j = 1, 2$, are identifiable (up to a rotation) from information that can be inferred by disjoint 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 one-to-one transformations. The latter indeterminacy does not prevent identifiability of the common and group-specific factors from Proposition 1, due to the invariance of canonical correlations and canonical variables under linear one-to-one transformations of vectors $h_{j,t}$.[5]

# 3 Estimation and inference on the number of common factors

Let us first 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 1 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 sub-panel $j$, and denote by $\hat{h}_{j,t}$ these PC estimates of the factors, $j = 1, 2$. Let $\hat{H}_j = [\hat{h}_{j,1}, ..., \hat{h}_{j,T}]'$ be the $(T, k_j)$ matrix of estimated PCs extracted from panel $Y_j = [y_{j,1}, ..., y_{j,T}]'$ 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 $\hat{h}_{\ell,t}$, $\hat{V}_{j\ell} = \hat{H}_j' \hat{H}_\ell / T = 1/T(\sum_{t=1}^T \hat{h}_{j,t} \hat{h}_{\ell,t}')$, $j, \ell = 1, 2$, and let matrices $\hat{R}$ and $\hat{R}^*$ be defined as:

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

Note that $\hat{V}_{jj} = I_{k_j}$ for $j = 1, 2$. Matrices $\hat{R}$ and $\hat{R}^*$ have the same non-zero eigenvalues. The $k^c$ largest eigenvalues of $\hat{R}$ (resp. $\hat{R}^*$), denoted by $\hat{\rho}_\ell^2$, $\ell = 1, ..., k^c$, are the first $k^c$ squared sample canonical correlation between $\hat{h}_{1,t}$ and $\hat{h}_{2,t}$, and that the associated $k^c$ canonical directions, collected in the $(k_1, k^c)$ matrix $\hat{W}_1$ (resp. $(k_2, k^c)$ matrix $\hat{W}_2$), are the eigenvectors associated with the largest $k^c$ eigenvalues of matrix $\hat{R}$ (resp. $\hat{R}^*$), normalized to have length 1 with respect to the matrix $\hat{V}_{11}$

[5] Computing PCs first is necessary because the alternative approach of canonical correlations applied to the raw data may not necessarily uncover pervasive factors. The alternative approach to stack all groups into one panel and apply standard PCA to estimate common factors is not a solution for at least two reasons: (1) 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, and (2) the combined data may not give the common factors because the common factor may not even be the leading factors in the combined data.

(resp. $\hat{V}_{22}$). It also holds:

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

**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: $\frac{1}{T}\sum_{t=1}^{T}\hat{f}_t^c\hat{f}_t^{c\prime} = I_{k^c}$, 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 (2.2). 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 $1$ with respect to the matrix $\hat{V}_{11}$ (resp. $\hat{V}_{22}$). It also holds: $\hat{W}_1^{s\,\prime}\hat{V}_{11}\hat{W}_1^s \;=\; I_{k_1^s}$, and $\hat{W}_2^{s\,\prime}\hat{V}_{22}\hat{W}_2^s \;=\; I_{k_2^s}$. The estimators of the group-specific factors can be defined analogously to the estimators of the common factors: $\breve{f}_{1,t}^s = \hat{W}_1^{s\,\prime}\hat{h}_{1,t}$ and $\breve{f}_{2,t}^s = \hat{W}_2^{s\,\prime}\hat{h}_{2,t}$. By construction, $\hat{f}_t^c$ and $\breve{f}_{1,t}^s$ (resp. $\hat{f}_t^{c*}$ and $\breve{f}_{2,t}^s$) 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.[6] Let $\hat{F}^c = [\hat{f}_1^{c\,\prime}, ..., \hat{f}_T^{c\,\prime}]'$ be the $(T, k^c)$ matrix of estimated common factors, and $\hat{\Lambda}_j^c \;=\; [\hat{\lambda}_{j,1}^c, ..., \hat{\lambda}_{j,N_j}^c]'$ the $(N_j, k^c)$ matrix collecting the estimated loadings:

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

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

**DEFINITION 2.** *Estimators of the specific factors $\hat{f}_{1,t}^s$ (resp. $\hat{f}_{2,t}^s$) are defined as the first $k_1^s$ (resp. $k_2^s$) PCs of sub-panel $\Xi_1$ (resp. $\Xi_2$), namely, the columns of the $(T, k_j^s)$ matrix $\hat{F}_j^s \;=\; [\hat{f}_{j,1}^s, ..., \hat{f}_{j,T}^s]'$ are the eigenvectors associated with the $k_j^s$ largest eigenvalues of matrix $\frac{1}{N_jT}\Xi_j\Xi_j'$, normalized to have $\hat{F}_j^{s\prime}\hat{F}_j^s/T = I_{k_j^s}$ for $j = 1, 2$.*

[6] This alternative estimation method for the group-specific factors corresponds to the method proposed by Chen (2012) who adopted an information criterion approach to estimate the number of factors, whereas we use a sequential testing method. Compared to Chen (2012), our paper derives results on the asymptotic distribution of the sample canonical correlation and estimated factors, whereas Chen (2012) only has consistency and rate of convergence results.

Note that $\hat{f}_t^c$ 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 $\breve{f}_{j,t}^s$, $j = 1, 2$. Moreover, we define $\hat{\Lambda}_j^s = [\hat{\lambda}_{j,1}^s, ..., \hat{\lambda}_{j,N_j}^s]'$ as the $(N_j, k_j^s)$ matrix collecting the loadings estimators:

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

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 $j = 1, 2$.

## 3.1 Inference on the number of common factors based on canonical correlations

One of our objectives is to determine how many factors are common between groups in the generic factor model appearing in equation (2.1), that is we consider the problem of inferring the dimension $k^c$ of the common factor space. To do so, we first consider the case where the number of factors $k_1$ and $k_2$ in each sub-panel is assumed to be known, hence $\underline{k} = \min(k_1, k_2)$ is also known, and we relax this assumption in the next section. From Proposition 1, dimension $k^c$ is the number of unit canonical correlations between $h_{1,t}$ and $h_{2,t}$. We consider the hypotheses: $H(0) = \{1 > \rho_1 \geq \ldots \geq \rho_{\underline{k}}\}$, $H(1) = \{\rho_1 = 1 > \rho_2 \geq \ldots \geq \rho_{\underline{k}}\}, \ldots, H(k^c) = \{\rho_1 = \ldots = \rho_{k^c} = 1 > \rho_{k^c+1} \geq \ldots \geq \rho_{\underline{k}}\}, \ldots$, and finally, $H(\underline{k}) = \{\rho_1 = ... = \rho_{\underline{k}} = 1\}$, where $\rho_1, ..., \rho_{\underline{k}}$ are the ordered canonical correlations of $h_{1,t}$ and $h_{2,t}$. Hypothesis $H(0)$ corresponds to the absence of common factors among the two groups. 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: $H_0 = H(k^c)$ against $H_1 = \bigcup_{0 \leq r < k^c} H(r)$, for each $k^c = \underline{k}, \underline{k} - 1, ..., 1$. We propose the following statistic to test $H_0$ against $H_1$, for any given $k^c = \underline{k}, \underline{k} - 1, ..., 1$:

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

The statistic $\hat{\xi}(k^c)$ corresponds to the sum of the $k^c$ largest sample canonical correlations of $\hat{h}_{1,t}$ and $\hat{h}_{2,t}$. We reject the null $H_0 = H(k^c)$ when $\hat{\xi}(k^c) - k^c$ is negative and large. The critical value is obtained from the large sample distribution of the statistic when $N_1$, $N_2$, $T \to \infty$, provided in Section 4. The number of common factors is estimated by sequentially applying the tests starting from $k^c = \underline{k}$.

### 3.2 Estimation and inference when $k_1$ and $k_2$ are unknown

The tests defined in the previous subsection 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 distribution and rate 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 $P(\hat{k}_j = k_j) \to 1$ for $j = 1, 2$, which means that the error due to the estimation of the number of pervasive factors is asymptotically negligible.[7] Therefore, the asymptotic distributions and rates of convergence of the test statistics and factors estimators will be derived assuming that the true dimension $k_j > 0$ in each subgroup, $j = 1, 2$, are known. The estimators based on the penalized information criteria of Bai and Ng (2002) applied to the two subgroups, are examples of consistent estimators for the numbers of pervasive factors.[8] Note that other consistent criteria could also be applied (see e.g., Onatski (2010), Ahn and Horenstein (2013) for further details).

## 4 Large sample theory

In this section we derive the large sample distribution of the test statistic for the dimension of the common factor space and provide a feasible version of it. We also define a consistent selection procedure for the number of common factors. We consider the joint asymptotics $N_1, N_2, T \to \infty$. Let us denote $N = \min\{N_1, N_2\}$ and $\mu_N = \sqrt{N_2/N_1}$. Without loss of generality, we set $N = N_2$, which implies $\mu_N \leq 1$. We assume that:

$$\sqrt{T}/N = o(1),\ N/T^{5/2} = o(1) \text{ and } \mu_N \to \mu, \text{ with } \mu \in [0, 1], \tag{4.1}$$

[7] For similar arguments, see footnote 5 of Bai (2003). A word of caution is warranted, however. It is known that pre-testing generates problems in terms of lack of uniform properties, and we therefore abstract from uniformity.

[8] These criteria are also applied in the empirical analysis in Section 6.

which we refer to as Assumption A.1 in Appendix A. The conditions in (4.1) allow for a wide range of relative growth rates for the time-series and cross-sectional panel dimensions as long as $N$ grows faster than $T^{1/2}$ and slower than $T^{5/2}$. They accommodate both the case where $N_1$ and $N_2$ grow at the same rate, and the case where $N_1$ grows faster than $N_2$, namely $\mu = 0$.

## 4.1 Asymptotic results for the group factor model

To derive the large sample distribution of the test statistic for the number of common factors we deploy an asymptotic expansion for the factor estimates obtained by PCA in each group, which we report in Proposition B.1 in Appendix B. The estimate $\hat{h}_{j,t}$ is asymptotically equivalent (in a sense made precise in Proposition B.1) to

$$\hat{\mathcal{H}}_j \left( h_{j,t} + \frac{1}{\sqrt{N_j}} u_{j,t} + \frac{1}{T} b_{j,t} \right) \tag{4.2}$$

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

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

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

and $\tilde{\Sigma}_{u,jk,t}(-h) = \tilde{\Sigma}_{u,kj,t}(h)'$, for $j, k = 1, 2$ and $h = 0, 1, \ldots$. We set $\tilde{\Sigma}_{u,jj,t} \equiv \tilde{\Sigma}_{u,jj,t}(0)$. Moreover, let us define the (probability) limits $\Sigma_{u,jk,t}(h) = \underset{N_j, N_k \to \infty}{\text{plim}} \tilde{\Sigma}_{u,jk,t}(h)$ and $\Sigma_{\lambda,j} = \lim_{N_j \to \infty} \frac{1}{N_j} \sum_{i=1}^{N_j} \lambda_{j,i} \lambda_{j,i}'$, and let $\bar{b}_{j,t} = \Sigma_{\lambda,j}^{-1} \eta_{j,t}^2 h_{j,t}$ be the large sample counterpart of $b_{j,t}$.

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

$$N\sqrt{T}\left(\Omega_{U,1}+\frac{N}{T^2}\Omega_{U,2}\right)^{-1/2}\left[\hat{\xi}(k^c)-k^c+\frac{1}{2N}tr\left\{\tilde{\Sigma}_{cc}^{-1}\tilde{\Sigma}_U\right\}+\frac{1}{2T^2}tr\left\{\tilde{\Sigma}_{cc}^{-1}\tilde{\Sigma}_B\right\}\right]\xrightarrow{d}N(0,1), \qquad (4.3)$$

*where*

$$\begin{aligned}
\tilde{\Sigma}_{cc} &= \frac{1}{T}\sum_{t=1}^{T} f_t^c f_t^{c\prime}, \qquad \tilde{\Sigma}_B=\frac{1}{T}\sum_{t=1}^{T}\widetilde{\Delta b_t}^{(c)}\widetilde{\Delta b_t}^{(c)\prime},\\
\widetilde{\Delta b_t} &= b_{1,t}-b_{2,t}-\left(\frac{1}{T}\sum_{s=1}^{T}(b_{1,s}-b_{2,s})F_s'\right)\left(\frac{1}{T}\sum_{s=1}^{T}F_sF_s'\right)^{-1}F_t,\\
\tilde{\Sigma}_U &= \frac{1}{T}\sum_{t=1}^{T}\left(\mu_N^2\tilde{\Sigma}_{u,11,t}^{(cc)}+\tilde{\Sigma}_{u,22,t}^{(cc)}-\mu_N\tilde{\Sigma}_{u,12,t}^{(cc)}-\mu_N\tilde{\Sigma}_{u,21,t}^{(cc)}\right),\\
\Omega_{U,1} &= \frac{1}{2}\sum_{h=-\infty}^{\infty}E\left[tr\left\{\Sigma_{U,t}(h)\Sigma_{U,t}(h)'\right\}\right], \qquad \Omega_{U,2}=\sum_{h=-\infty}^{\infty}E\left[tr\left\{\Sigma_{U,t}(h)\Delta b_{t-h}^{(c)}\Delta b_t^{(c)\prime}\right\}\right],\\
\Delta b_t &= \bar{b}_{1,t}-\bar{b}_{2,t}-E\left[\left(\bar{b}_{1,t}-\bar{b}_{2,t}\right)F_t'\right]V(F_t)^{-1}F_t,\\
\Sigma_{U,t}(h) &= \mu^2\Sigma_{u,11,t}^{(cc)}(h)+\Sigma_{u,22,t}^{(cc)}(h)-\mu\Sigma_{u,12,t}^{(cc)}(h)-\mu\Sigma_{u,21,t}^{(cc)}(h), \qquad h=...,-1,0,1,...,
\end{aligned}$$

*and where the upper index $(c)$ denotes the upper $(k^c,1)$ block of a vector, and the upper index $(c,c)$ denotes the upper-left $(k^c,k^c)$ block of a matrix.*

**Proof**: See Appendix B.1.

The matrix $\Sigma_{U,t}(h)$ is the upper-left $(k^c,k^c)$ block of the limit covariance matrix between $\mu_N u_{1,t}-u_{2,t}$ and $\mu_N u_{1,t-h}-u_{2,t-h}$, where the weight $\mu_N=\sqrt{N_2/N_1}$ accounts for the different sample sizes in the two sub-panels. Vectors $\widetilde{\Delta b_t}$ and $\Delta b_t$ are residuals of the orthogonal projection of $b_{1,t}-b_{2,t}$ onto $F_t$ in-sample, and of $\bar{b}_{1,t}-\bar{b}_{2,t}$ onto $F_t$ in the population, respectively. In fact, the orthogonal projection of vector $b_{j,t}$ along vector $h_{j,t}$ can be absorbed in the transformation matrix $\hat{\mathcal{H}}_j$ in expansion (4.2), and therefore is asymptotically immaterial for the computation of canonical correlations and for the large sample distribution of the test statistic.

The asymptotic distribution in Theorem 1 is valid for $\sqrt{T}\ll N\ll T^{5/2}$ (Assumption A.1). It covers the variety of convergence rates and asymptotic biases and variances the $\hat{\xi}(k^c)$ statistic features, for different relative growth rates of sample dimensions $N, T$ when $\Omega_{U,2} > 0$, namely:

$$
\begin{aligned}
T^{1/2} \ll N \ll T^{3/2}: &\quad N\sqrt{T}\left[\hat{\xi}(k^c) - k^c + \tfrac{1}{2N} tr\left\{\tilde{\Sigma}_{cc}^{-1}\tilde{\Sigma}_U\right\}\right] \xrightarrow{d} N\left(0, \Omega_{U,1}\right), \\
T^{3/2} \lessapprox N \ll T^{2}: &\quad N\sqrt{T}\left[\hat{\xi}(k^c) - k^c + \tfrac{1}{2N} tr\left\{\tilde{\Sigma}_{cc}^{-1}\tilde{\Sigma}_U\right\} + \tfrac{1}{2T^2} tr\left\{\tilde{\Sigma}_{cc}^{-1}\tilde{\Sigma}_B\right\}\right] \xrightarrow{d} N\left(0, \Omega_{U,1}\right), \\
\lambda := \lim \tfrac{N}{T^2} \in (0, \infty): &\quad N\sqrt{T}\left[\hat{\xi}(k^c) - k^c + \tfrac{1}{2N} tr\left\{\tilde{\Sigma}_{cc}^{-1}\tilde{\Sigma}_U\right\} + \tfrac{1}{2T^2} tr\left\{\tilde{\Sigma}_{cc}^{-1}\tilde{\Sigma}_B\right\}\right] \\
&\quad \xrightarrow{d} N\left(0, \Omega_{U,1} + \lambda\Omega_{U,2}\right), \\
T^{2} \ll N \ll T^{5/2}: &\quad T\sqrt{TN}\left[\hat{\xi}(k^c) - k^c + \tfrac{1}{2N} tr\left\{\tilde{\Sigma}_{cc}^{-1}\tilde{\Sigma}_U\right\} + \tfrac{1}{2T^2} tr\left\{\tilde{\Sigma}_{cc}^{-1}\tilde{\Sigma}_B\right\}\right] \xrightarrow{d} N\left(0, \Omega_{U,2}\right).
\end{aligned}
$$

In particular, the convergence rate of the statistic is $\min\{N\sqrt{T}, T\sqrt{NT}\}$. When $\Omega_{U,2} = 0$ (see below), the convergence rate is $N\sqrt{T}$ and the asymptotic variance is $\Omega_{U,1}$ for $T^{1/2} \ll N \ll T^{5/2}$. Note that, if the PCs in the groups were observed, then testing for unit canonical correlations would be degenerate, as it involves testing for deterministic relationships between random vectors. The estimation errors of the PCs drive the asymptotic distribution of the statistic, with a non-standard convergence rate.
It might be surprising that we find an asymptotic Gaussian distribution when testing a hypothesis for a parameter at the boundary, i.e. canonical correlations equal to one. What makes the test Gaussian asymptotically, is the fact that there is a re-centering of the statistic due to the sampling error in the first step estimates of the principal components, and a CLT applies to the re-centered squared estimation errors. The re-centering term involves a component at order $N^{-1}$ and a component at order $T^{-2}$. One may wonder whether this Gaussian asymptotic distribution is a good approximation for the small sample distribution of the re-centered and re-scaled $\hat{\xi}(k^c)$. In Section 5 and OA Section F, we report the results of extensive Monte Carlo simulations demonstrating that this is the case in a setting that mimics our empirical application.

To get a feasible distributional result for the statistic $\hat{\xi}(k^c)$, we need consistent estimators for the unknown scalars $tr\left\{\tilde{\Sigma}_{cc}^{-1}\tilde{\Sigma}_U\right\}$ and $tr\left\{\tilde{\Sigma}_{cc}^{-1}\tilde{\Sigma}_B\right\}$, and matrices $\Omega_{U,1}$ and $\Omega_{U,2}$ in Theorem 1. To simplify the analysis, we assume at this stage that the errors $\varepsilon_{j,i,t}$ are (i) uncorrelated across sub-panels $j$ and individuals $i$, at all leads and lags, and (ii) a conditionally homoscedastic martingale difference sequence for each individual $i$, conditional on the factor path, i.e.

$$
\begin{aligned}
Cov(\varepsilon_{j,i,t}, \varepsilon_{k,\ell,t-h} | \mathcal{F}_t) &= 0, \qquad \text{if either } j \neq k, \text{ or } i \neq \ell, \\
E[\varepsilon_{j,i,t} | \{\varepsilon_{j,i,t-h}\}_{h \geq 1}, \mathcal{F}_t] &= 0, \qquad E[\varepsilon_{j,i,t}^2 | \{\varepsilon_{j,i,t-h}\}_{h \geq 1}, \mathcal{F}_t] = \gamma_{j,ii} \text{ (say)},
\end{aligned}
\tag{4.4}
$$

for all $j$, $i$, $t$, $h$ (see Assumption A.9). Then, we have:

$$\tilde{\Sigma}_U = \mu_N^2 \tilde{\Sigma}_{u,11}^{(cc)} + \tilde{\Sigma}_{u,22}^{(cc)}, \quad \Sigma_U(0) \equiv \Sigma_U = \mu^2 \Sigma_{u,11}^{(cc)} + \Sigma_{u,22}^{(cc)}, \quad \Omega_{U,1} = \frac{1}{2} tr\left\{\Sigma_U^2\right\}, \quad \Omega_{U,2} = 0. \quad (4.5)$$

Matrices $\tilde{\Sigma}_{u,jj}$ and $\Sigma_{u,jj} \equiv \Sigma_{u,jj}(0)$ do not depend on time. The projection residual $\Delta b_t$ vanishes because $\bar{b}_{j,t} = \Sigma_{\lambda,j}^{-1} \eta_j^2 h_{j,t}$, where $\eta_j^2 = \lim_{N_j \to \infty} \frac{1}{N_j} \sum_{i=1}^{N_j} \gamma_{j,ii}$, is spanned by $F_t$. This explains why $\Omega_{U,2}$ is null and the convergence rate is $N\sqrt{T}$. Similarly, $\tilde{\Sigma}_B = 0$, so that the bias term at order $T^{-2}$ is zero.[9] Under (4.4) the bias component $T^{-1} b_{j,t}$ is immaterial since it can be absorbed in the transformation matrix $\hat{\mathcal{H}}_j$ in (4.2). In fact, Connor and Korajczyk (1986) and Bai (2003) Theorem 4 show that the principal component estimator is consistent even for fixed $T$ in such a case. In Theorem 2 below, we replace $\tilde{\Sigma}_{cc}$ with its large sample limit $I_{k^c}$, matrices $\tilde{\Sigma}_U$ and $\Sigma_U$ by consistent estimators. We show that the estimation error for $\frac{1}{N} tr\left\{\tilde{\Sigma}_{cc}^{-1} \tilde{\Sigma}_U\right\}$ in the bias adjustment is of order $o_p\left(\frac{1}{N\sqrt{T}}\right)$, and therefore the asymptotic distribution of the statistic is unchanged.

**THEOREM 2.** *Let* $\hat{\Sigma}_U = (N_2/N_1)\hat{\Sigma}_{u,11}^{(cc)} + \hat{\Sigma}_{u,22}^{(cc)}$, *with* $\hat{\Sigma}_{u,jj} = \left(\frac{1}{N_j}\hat{\Lambda}_j'\hat{\Lambda}_j\right)^{-1}\left(\frac{1}{N_j}\hat{\Lambda}_j'\hat{\Gamma}_j\hat{\Lambda}_j\right)\left(\frac{1}{N_j}\hat{\Lambda}_j'\hat{\Lambda}_j\right)^{-1}$ *where* $\hat{\Lambda}_j = [\hat{\Lambda}_j^c \vdots \hat{\Lambda}_j^s]$, $\hat{\Lambda}_j^c$ *and* $\hat{\Lambda}_j^s$ *are the loadings estimators defined in equations (3.3) and (3.4),* $\hat{\Gamma}_j = diag(\hat{\gamma}_{j,ii},\ i = 1, ..., N_j)$ *with* $\hat{\gamma}_{j,ii} = \frac{1}{T}\sum_{t=1}^{T} \hat{\varepsilon}_{j,i,t}^2$, *and* $\hat{\varepsilon}_{j,i,t} = y_{j,i,t} - \hat{\lambda}_{j,i}^{c\,\prime} \hat{f}_t^c - \hat{\lambda}_{j,i}^{s\,\prime} \hat{f}_{j,t}^s$, *for* $j = 1, 2$. *Define the test statistic:*

$$\tilde{\xi}(k^c) \ := \ N\sqrt{T}\left(\frac{1}{2} tr\{\hat{\Sigma}_U^2\}\right)^{-1/2}\left[\hat{\xi}(k^c) - k^c + \frac{1}{2N} tr\left\{\hat{\Sigma}_U\right\}\right], \quad (4.6)$$

*and let Assumptions A.1 - A.9 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* $H_1 = \bigcup_{0 \le r < k^c} H(r)$, *we have:* $\tilde{\xi}(k^c) \xrightarrow{p} -\infty$.

**Proof**: See Appendix B.2.

The feasible asymptotic distribution in Theorem 2 is the basis for a one-sided test of the null hypothesis of $k^c$ common factors. The rejection region for a test of the null hypothesis at asymptotic level $\alpha$ is $\tilde{\xi}(k^c) < z_\alpha$, where $z_\alpha$ is the $\alpha$-quantile of the standard Gaussian distribution for $\alpha \in (0, 1)$. From Theorem 2 *(ii)*, the test is consistent under the alternative.

[9] If the errors are weakly correlated across series and/or time, consistent estimation of $\tilde{\Sigma}_U$ and $\Omega_{U,1}$ requires thresholding of estimated cross-sectional covariances and/or HAC-type estimators. If the errors are conditionally heteroskedastic, we need consistent estimators of $\Omega_{U,2}$ and $\tilde{\Sigma}_B$ as well.

One way to implement the model selection procedure to estimate the number of common factors $k^c$ proposed in Section 3.1 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 in Theorem 2 for any generic number $r$ of common factors. A "naive" procedure is initiated with $r = \underline{k}$, proceeds backwards and is stopped at the largest integer $\hat{k}^c_{naive} = r$ such that the null $H(r)$ cannot be rejected, i.e. $\tilde{\xi}(r) \geq z_\alpha$. Otherwise, set $\hat{k}^c_{naive} = 0$ if the test rejects the null $H(r)$ for all $r = \underline{k}, ..., 1$. This "naive" procedure is not a consistent estimator of the number of common factors. Indeed, asymptotically a non-zero probability $\alpha$ of underestimating $k^c$ exists coming from the type I error of the test of $H(k^c_0)$ against $\bigcup_{0\leq \ell<k^c_0} H(\ell)$, when the true $k^c_0 > 0$.

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 $k^c_0$, for any integer $k^c_0 \geq 0$, is obtained allowing the asymptotic size $\alpha$ to go to zero as $N$, $T \to \infty$. The following Proposition 2 defines a consistent inference procedure for the number of common factors.

**PROPOSITION 2.** *Let $\alpha_{N,T}$ be a sequence of real scalars defined in the interval $(0, 1)$ for any $N, T$, such that (i) $\alpha_{N,T} \to 0$ and (ii) $(N\sqrt{T})^{-1} z_{\alpha_{N,T}} \to 0$ for $N, T \to \infty$. Then, under Assumptions A.1 - A.9 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 $r = 1, ..., \underline{k}$, is consistent, i.e. $P(\hat{k}^c = k^c_0) \longrightarrow 1$ under $H(k^c_0)$, for any integer $k^c_0 \in [0, \underline{k}]$.*

**Proof**: See Appendix X.5.

Condition *(i)* ensures asymptotically zero probability of type I error when testing $H(k^c_0)$ against $\bigcup_{0\leq \ell<k^c_0} H(\ell)$. 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 2 are satisfied e.g. for $\alpha_{N,T}$ such that:

$$z_{\alpha_{N,T}} = -c(N\sqrt{T})^\gamma, \tag{4.7}$$

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

## 4.2 Mixed frequency group factor models

The idea to apply grouped factor analysis to mixed frequency data is novel and has many advantages in terms of identification and estimation. In this subsection we explore this further. We consider a setting where both low and high frequency data are available. Let $t = 1, 2, \ldots, T$ be the low frequency (LF) time units. Each time period $(t-1, t]$ is divided into $M$ sub-periods with high frequency (HF) dates $t - 1 + m/M$, with $m = 1, \ldots, M$. 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 $x^{Hi}_{m,t}$, for $i = 1, \ldots, N_H$, be the high frequency data observation $i$ during sub-period $m$ of low frequency period $t$. Likewise, we let $x^{Li}_t$, with $i = 1, \ldots, N_L$, be the observation of the $i^{th}$ low-frequency series at $t$. These observations are gathered into the $N_H$-dimensional vectors $x^H_{m,t}$, for all $m$, and the $N_L$-dimensional vector $x^L_t$, respectively.

We assume that there are three types of pervasive factors, which we denote by $g^C_{m,t}$, $g^H_{m,t}$ and $g^L_{m,t}$, respectively. The former represents a vector of factors which affect both high and low frequency data (we use again superscript $C$ for common), whereas the other two types of 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:

$$
\begin{aligned}
x^H_{m,t} &= \Lambda_{HC} g^C_{m,t} + \Lambda_H g^H_{m,t} + e^H_{m,t}, \\
x^{L*}_{m,t} &= \Lambda_{LC} g^C_{m,t} + \Lambda_L g^L_{m,t} + e^L_{m,t},
\end{aligned}
\tag{4.8}
$$

where $m = 1, ..., M$ and $t = 1, ..., T$, and $\Lambda_{HC}$, $\Lambda_H$, $\Lambda_{LC}$ and $\Lambda_L$ are matrices of factor loadings. The vector $x^{L*}_{m,t}$ is unobserved for each high frequency sub-period and the measurements, denoted by $x^L_t$, depend on the observation scheme, which can be either flow-sampling or stock-sampling (or some general linear scheme).

In the case of flow-sampling, the low frequency observations are the sum (or average) of all $x^{L*}_{m,t}$

across all $m$, that is: $x_t^L = \sum_{m=1}^M x_{m,t}^{L*}$.[10] Then, model (4.8) implies:

$$\begin{aligned} x_{m,t}^H &= \Lambda_{HC} g_{m,t}^C + \Lambda_H g_{m,t}^H + e_{m,t}^H, \quad m = 1, ..., M, \\ x_t^L &= \Lambda_{LC} \sum_{m=1}^M g_{m,t}^C + \Lambda_L \sum_{m=1}^M g_{m,t}^L + \sum_{m=1}^M e_{m,t}^L. \end{aligned} \tag{4.9}$$

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: $\bar{g}_t^U := \sum_{m=1}^M g_{m,t}^U$, $U = C$, $H$, $L$. Then we can stack the observations $x_t^H$ and $x_t^L$ and write:

$$\begin{bmatrix} x_t^H \\ x_t^L \end{bmatrix} = \begin{bmatrix} \Lambda_{HC} & \Lambda_H & 0 \\ \Lambda_{LC} & 0 & \Lambda_L \end{bmatrix} \begin{bmatrix} \bar{g}_t^C \\ \bar{g}_t^H \\ \bar{g}_t^L \end{bmatrix} + \begin{bmatrix} \bar{e}_t^H \\ \bar{e}_t^L \end{bmatrix}, \tag{4.10}$$

which is a group factor model, with common factor $\bar{g}_t^C$ and group-specific factors $\bar{g}_t^H$ and $\bar{g}_t^L$.

The results in Sections 2, 3 and 4.1 can be applied for identification and inference in the mixed frequency factor model. First, some observations about identification are in order. Recall that the rotational invariance of the group factor model (2.1) - (2.2) maintains the interpretation of common and specific factors. Using the same argument in the mixed frequency setting of equation (4.10), identification can be achieved for the aggregated factors $\bar{g}_t^C$, $\bar{g}_t^H$, and $\bar{g}_t^L$, and the factor loadings $\Lambda_{HC}$, $\Lambda_{LC}$, $\Lambda_H$, and $\Lambda_L$. The latent common and group-specific factors are normalized such that $\bar{g}_t^U$, $U = C, H, L$, satisfy the counterpart of (2.2). Consequently, the estimators and test statistics for the group factor model (2.1) can also be used to define estimators for the loadings matrices $\Lambda_{HC}$, $\Lambda_H$, $\Lambda_{LC}$, $\Lambda_L$, and the aggregated factor values $\overline{g}_t^U$, $U = C, H, L$, and the test statistic for the common factor space dimension $k^C$ in equation (4.10). We denote these estimators $\hat{\Lambda}_{HC}$, $\hat{\Lambda}_H$, $\hat{\Lambda}_{LC}$, $\hat{\Lambda}_L$, $\hat{\bar{g}}_t^U$, and the test statistic $\hat{\xi}(k^C)$. Once the factor loadings are identified from (4.10) and estimated, the values of the common and high frequency factors for sub-periods $m = 1, ..., M$ are identifiable by cross-sectional regression of the high frequency data on loadings $\Lambda_{HC}$ and $\Lambda_H$ in (4.8). More specifically,

[10] In the remainder of this section we study identification of the model for the flow-sampling case as it corresponds to the empirical application. The identification with stock-sampling is discussed in the OA, Section E.2. 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 (4.10) or (2.1) - as discussed briefly in Section E.2 - is compatible with the identification and estimation strategies of this paper.

the estimators of the common and high frequency factor values are:

$$\begin{pmatrix} \hat{g}^C_{m,t} \\ \hat{g}^H_{m,t} \end{pmatrix} = \left(\hat{\Lambda}_1'\hat{\Lambda}_1\right)^{-1}\hat{\Lambda}_1' x^H_{m,t}, \qquad m = 1, ..., M, \qquad t = 1, ..., T, \tag{4.11}$$

where $\hat{\Lambda}_1 = [\hat{\Lambda}_{HC} \vdots \hat{\Lambda}_H]$. Hence, $\hat{g}^C_{m,t}$ and $\hat{g}^H_{m,t}$ are obtained by regressing $x^{Hi}_{m,t}$ on $\hat{\lambda}_{HC,i}$ and $\hat{\lambda}_{H,i}$ across $i = 1, 2, ..., N_H$, for any $m = 1, ..., M$ and $t = 1, ..., T$. Consequently, with flow-sampling, we can identify and estimate $g^C_{m,t}$ and $g^H_{m,t}$ at all high frequency sub-periods. On the other hand, only $\bar{g}^L_t = \sum_{m=1}^{M} g^L_{m,t}$, i.e. the within-period sum of the low frequency factor, is identifiable by the paired panel data set consisting of $x^H_t$ combined with $x^L_t$. This is not surprising, since we have no high-frequency observations available for the LF data.

# 5 Monte Carlo simulation analysis

The objective of the MC simulation study are: (i) to evaluate the small sample size and power properties of test statistics $\hat{\xi}(k^C)$ and $\tilde{\xi}(k^C)$ proposed in Theorems 1 and 2, respectively, large and small samples, for the number of common factors $k^C$ (ii) to compare the advantages of the sequential testing procedure for $k^C$ in Proposition 2, vis-à-vis the selection criterion of Chen (2012), and the three-steps procedure in Wang (2012).

A detailed description of the simulation design and an extensive set of tables reporting the results, appear in Section F of the OA.[11] The data generating process (DGP) for each design corresponds to the high frequency model (4.8) with flow-sampled LF variables. We allow the factors to be autocorrelated. The idiosyncratic innovations are independent of the factors, serially i.i.d., and possibly cross-sectionally correlated within a panel. In line with our empirical application, we report the number of high frequency sub-periods for $M = 4$. We consider different numbers of common and specific factors across the DGPs, given by $k^C = 1, 2$, and $k^H = k^L = 1$, and $5$. The DGP for the vector of stacked factors $g_{m,t} = [g^{C\prime}_{m,t}, g^{H\prime}_{m,t}, g^{L\prime}_{m,t}]'$ is characterized by the VAR: $g_{m,t} = a_F g_{m-1,t} + \eta_{m,t}$, where $a_F$ is a common scalar AR coefficient for all the $k^C + k^H + k^L$ factors. The innovations are given by $\eta_{m,t} \sim i.i.N(0, \Sigma_\eta)$, where $\Sigma_\eta = \frac{1-a_F^2}{M^2\kappa}\Xi$ where $\Xi$ is a matrix whose main diagonal is $(I_{k^C}, I_{k^H}, I_{k^L})$ and off diagonal ele-

[11] Additional results available upon request examine the asymptotic and finite sample properties of our tests for a number of different parameters involved in the simulation design.

ments are zero except $\xi_{23} = \Phi$ and $\xi_{32} = \Phi'$ and $\kappa = 1 - \frac{2}{M^2} \sum_{m=1}^{M-1} m(1 - a_F^{M-m})$. The scaling term in the variance $\Sigma_\eta$ ensures that the factor normalization in (2.2) holds for $[\bar{g}_t^{C\prime}, \bar{g}_t^{H\prime}, \bar{g}_t^{L\prime}]'$, and $\Phi = \phi I_{k^H}$, where the scalar $\phi$ generates correlation between pairs of HF and LF factors. The factor loadings are simulated from NIID(0,**1**) such that the distribution of $R^{2\prime}$s of the regressions of observables on factors mimics the one in the empirical application. We run 4000 simulations for each DGP, and consider $N_H$, $N_L$, $T$ as small as the ones in our empirical applications, and progressively increase them.

## 5.1 Size and power properties

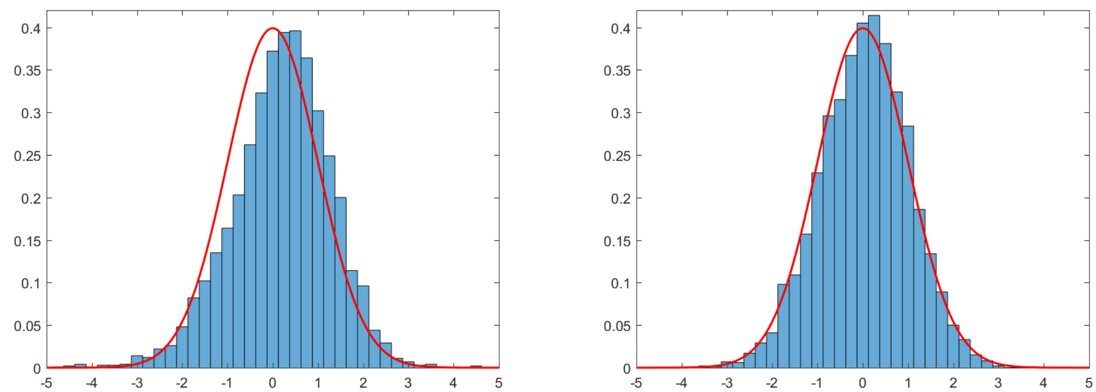
We are interested in verifying whether the Gaussian asymptotic distribution provides a good small sample approximation for the left tail of the re-centered and re-scaled infeasible statistic $\hat{\xi}(k^C)$, and the feasible $\tilde{\xi}(k^C)$. The feasible statistic is computed as in (4.6), assuming conditionally heteroschedastic and serially as well as cross-sectionally independent errors. We compute the empirical size of the test for the null hypothesis of $k^C$ common factors corresponding to nominal sizes of $1\%$, $5\%$ and $10\%$. We also report the empirical power of the feasible statistic for the null hypothesis of $k^C + 1$ common factors, when the true number of common factors is $k^C$, for the same nominal sizes.

In Figure 1 we display the histograms of the empirical distribution of the re-centered and re-scaled statistic $\hat{\xi}(k^c)$ defined in Theorem 1. The statistics are computed under the null hypothesis of $k^c = 1$ common factors on data simulated from a DGP with $k^c = k_1^s = k_2^s = 1$, and are overlapped with the density of a $N(0, 1)$ variate, representing the asymptotic distribution. For small sample sizes of $N_H = N_L = 100$, and $T = 50$ (Case 1), the empirical distribution resembles a normal distribution with unit standard deviation (the empirical standard deviation is 1.08), but the distribution is centered around a small positive value (the empirical mean is 0.20). Nevertheless, the left tail of this empirical distribution resembles the one of a standard Gaussian.[12] On the other hand, as the sample sizes grow to $N_H = N_L = 800$, and $T = 500$ (Case 2), the empirical distribution of $\hat{\xi}(k^c)$ features empirical mean and standard deviation of 0.04 and 0.98, respectively, and it almost perfectly overlaps with the asymptotic distribution. These results are qualitatively similar for alternative DGPs and sample sizes. Hence, we conclude that our asymptotic theory is supported by MC simulations, and provides a good

[12]The $1\%$, $5\%$ and $10\%$ quantiles of the standard Gaussian distribution have empirical sizes equal to 0.02, 0.05 and 0.09, respectively. We are particularly interested in the left tail of the distribution, as it is the relevant one for the test of number of common factors developed later.

approximation in small samples.

Figure 1: Small sample distribution of the re-centered and re-scaled $\hat{\xi}(k^c)$ statistic




(a) Case 1

(b) Case 2

The figure displays the histograms of the empirical distribution of the re-centered and re-scaled $\hat{\xi}(k^c)$ statistic computed on mixed-frequency panels of observations simulated from a DGP where $k^c = k_1^s = k_2^s = 1$, all factors and idiosyncratic terms are generated as Gaussian random variables for all the $MT$ time periods, and $M = 4$. The red solid line corresponds to the asymptotic standard Gaussian distribution of the re-centered and re-scaled statistic. Panel (a) displays the empirical distribution of $\hat{\xi}(k^c)$ for samples of size $N_H = N_L = 100$, and $T = 50$. Panel (b) displays the empirical distribution of $\hat{\xi}(k^c)$ for samples of size $N_H = N_L = 800$, and $T = 500$.

The tables in Section F.5 show that the asymptotic Gaussian distribution provides an excellent approximation for the left tail of the infeasible test statistics $\hat{\xi}(k^C)$ under the null, even for sample sizes as small as $N_H = N_L = 50$, and $T = 35$. For the vast majority of sample sizes, and simulation designs, the size distortions for aforementioned case (1) are in the order of $1\%$ to maximum $3\%$. Analogous results hold for case (2) except for the two designs in which $k^C = 2$, and $k^H = k^L = 1$ for sample sizes as small as $T \leq 50$, and $\max(N_H, N_L) \leq 200$, where the size distortion increases to a maximum of $10\%$. This result is due to the fact that, by construction, the signal-to-noise ratio for each of the two common factors in this simulation design is halved compared to those with $k^C = 1$. Moreover, we note that the infeasible test seems to be undersized for sample sizes as small as $T \leq 200$ when the number of specific factors is relatively high in both panels, nevertheless this does not affect significantly the performance of the sequential procedure for the selection of the number of common factors. Finally, the size distortion disappears in all simulations designs for large values of $N_H$, $N_L$, and $T$, which validates our asymptotic theory. Turning to the infeasible statistic $\tilde{\xi}(k^C)$, we note that the size distortions are from $1\%$ to $12\%$ larger than those of the feasible statistic, when $\max(N_H, N_L) \leq 200$, and $T \leq 50$, but as sample sizes increase all size distortions vanish. The designs with $k^C = 2$, and $k^H = k^L = 0$ or $1$, feature larger size distortions when $T \leq 200$, and $\max(N_H, N_L) \leq 200$ for the same reason noted above. As expected, when the signal-to-noise ratio of the common factors is increased, the size distortions monotonically improve for all sample sizes, and especially for the very small ones.[13] The power of the feasible test statistics is always equal to $1$, with the exception of designs with $\min(N_H, N_L) \leq 50$, and $T = 35$ (and alternative values of M).

The above results are robust when the idiosyncratic innovations in the DGPs feature either zero, or a moderate level of cross-sectional correlation and when the factors feature a level of the autocorrelation coefficient either equal to zero, or similar to the one in the empirical analysis, namely $a_F = 0.6$. We refer the reader to the OA for additional details.

## 5.2 Estimation of the number of factors

We compare the following three procedures to determine the number of common factors $k^C$ : (a) our consistent sequential testing procedure defined in Proposition 2, (b) the selection procedure based on the penalized information criterion of Theorem 3.7 in Chen (2012), and (c) the three-steps selection procedure proposed by Wang (2012).[14]

The estimators are evaluated by comparing the average (across simulations) estimated number of common, high-frequency-specific, and low-frequency-specific factors. For all the estimators of $k^C$ we consider the case in which the true numbers of pervasive factors $k_1 = k^C + k^H$ and $k_2 = k^C + k^L$ in the two panels are known, and only $k^C$ needs to be estimated, and also the case in which $k_1$ and $k_2$ are estimated using the $IC_p$ information criteria proposed by Bai and Ng (2002). More specifically, we present estimation results for $k_1$ and $k_2$ where we used the $IC_{p2}$ criterion for all designs in which $k^H = k^L = 0$ or $1$, and the $IC_{p3}$ criterion when $k^H = k^L = 5$. The same criteria are used to estimate the number of pervasive factors in the stacked panel of HF (flow-sampled) and LF data in the second step of the procedure suggested by Wang (2012). In line with the results of Bai and Ng (2002),

[13] In unreported simulation results, available upon request, we increased the signal-to-noise of all factors, and of the common factors only. These changes generated improvements for all empirical sizes, and especially for the designs with $k^C = 2$ which featured the larger size distortions by construction of the designs.

[14] We thank an anonymous referee for the suggestion to consider the following three-steps procedure to determine $k^C$: (1) estimate the number of pervasive factors in each of the two panels separately, yielding $\hat{k}_1$ and $\hat{k}_2$, (2) estimate the number $\hat{R}$ of pervasive factors in the stacked panel of HF (flow-sampled) and LF data, (3) determine $k^C$ as $\hat{k}_1 + \hat{k}_2 - \hat{R}$. We note that this procedure is a special case of the one suggested by Wang (2012).

we noted that for small sample sizes ($T \leq 50$, and especially for $T = 35$) and in the case of many pervasive factors in the LF panel (that is $k_2 \geq 5$) the $IC_{p2}$ criterion tends to severely underestimate the values of $k_2$, while the $IC_{p3}$ produces better estimates. This underestimation of $k_2$ has a considerable effect of the estimate of $k^C$ for all the three procedures considered.[15] Results are very similar both in the case in which $k_1$ and $k_2$ are known, and when they need to be estimated as we have just described. The critical value for our selection procedure is determined as in equation (4.7), with $\gamma = 0.1$, and $c = 0.95$ such that $z_{\alpha_{NT}} = -1.64 \sim z_{0.05}$ for $N = 40$, and $T = 35$, which are analogous to the smallest cross-sectional and time series dimensions in our empirical application.

In the special case of a small number, say 1 or 3, of uncorrelated specific factors, the penalized information criterion proposed in Chen (2012) yields the correct number of factors in almost all simulations for any sample size, confirming the results in Chen (2012).[16] For the same DGPs our selection procedure is less accurate only for sample sizes as small as $\max(N_H, N_L) \leq 200$, and $T \leq 50$: the average estimated number of common factors ranges between $0.85$ and $1$ when $k^C = 1$. As expected from the aforementioned results regarding size distortions, when PCA is performed on HF data first, the estimated number of factors is always more precise, with respect to the case in which HF data are first flow sampled, with the average estimated number of common factors ranging between $0.90$ and $1$ when $k^C = 1$. As predicted by our asymptotic theory, the average estimated number of common factors for our selection procedure approaches quickly the true value $k^C$ as the sample sizes increase.

Consequently, the results on the estimated number of factors are also qualitatively very similar for either a zero, or a moderate level of cross-sectional correlation among the idiosyncratic innovations. Additionally, when PCA is first performed on the high frequency data, $k^C = k^H = k^L = 1$, and for small sample sizes, we note a modest deterioration of the performance our sequential testing procedure if the autocorrelation coefficient of all factors increases form $0$ to $0.6$. In this case, our procedure tends to slightly underestimate the number of common factors as expected from the increase of the empirical size discussed in Section 5.1. Nevertheless, the deterioration of our sequential procedure when PCA is performed on the HF data directly, is not enough to make it worse than the same sequential procedure with PCA performed on flow-sampled HF data.

[15] In unreported results available upon request we have estimated $k_1$, $k_2$, and also $R$, using the ER and GR ratios of Ahn and Horenstein (2013), and noted that they perform similarly or worse than the $IC_{p2}$ criterion. Alternative estimators, such as the one proposed by Onatski (2010), could also be considered.

[16] To save space, the tables of results for the design $k^C = 1$ and $k^H = k^L = 3$ have not been reported, and are available upon request. They are analogous to the ones for the case $k^C = 1$ and $k^H = k^L = 1$ .

The procedure of Chen (2012) tends to overestimate the number of common factors when the correlation $\phi$ among the specific factors increases from $0$ to $0.7$ and $0.95$.[17] This type of deterioration in performance is much less dramatic for our sequential test procedure. As expected, we also observe a monotonic decrease in the precision across all the estimators when the number of specific factors becomes relatively large, like say $5$. For all the designs in which $T \leq 50$ also in the case of uncorrelated specific factors ($\phi$ = $0$) our procedure consistently outperforms Chen (2012). For larger values of the correlation coefficient $\phi$, the better performance of our procedure is even more evident also in larger sample sizes. It is noteworthy that as $\phi$ increases the deterioration for our sequential procedure is much less dramatic than Chen's (2012) procedure, suggesting that our sequential procedure is preferable in these more general cases. Furthermore, our sequential testing procedure also exhibits a more rapid improvement in performance as the sample size increases.

Finally, the consistent three-steps selection procedure of Wang (2012) performs similarly to the one of Chen (2012) in DGPs with a small number of uncorrelated specific factors. However, as either $\phi$ or the number of specific factor increases, the procedure tends to overestimate $k^C$ and clearly becomes the worse among the three considered.

# 6 Empirical application

Recent public policy debates argue that manufacturing has been declining in the United States and most jobs have migrated overseas to lower wage countries. The share of the Industrial Production (IP) sector declined from more than 25% to roughly 18% during our sample period 1977-2011.[18] However, the fact that its size shrank does not necessarily exclude the possibility that the IP sector still is a key factor of total U.S. output. Using the proposed class of mixed frequency group factor models, the objective of the empirical application is to shed light on the key question of interest, namely whether, despite the shrinking size of IP sectors, the factors related to IP are still dominant determinants of U.S. output fluctuations.

[17] Due to space limitations the results for the case $\phi$ = $0.7$ have not been reported, but are available upon request.

[18] We focus on this sample period because our empirical analysis revisits the results in Foerster, Sarte, and Watson (2011) using new methods. See Figure E.1 in the OA for a visualization of the steady decline.

## 6.1 Data description

For the IP sectors we use the same 117 IP sectoral growth rates indices sampled at quarterly frequency from 1977.Q1 to 2011.Q4, as in Foerster, Sarte, and Watson (2011) for comparison.[19] 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 and local). These LF data are published by the Bureau of Economic Analysis (BEA).[20] Hence we consider the panel of these yearly GDP sectoral and the quarterly IP data given that one of the objectives of this application is to study the comovements among these different sectors.[21]

## 6.2 Common, low- and high-frequency factors

We assume that our dataset follows the factor structure for flow-sampling as in equation (4.9), with $x^H_{m,t}$ and $x^L_t$ corresponding to the 117 quarterly IP series and the 42 annual GDP non-IP sector data series, respectively, for the period 1977.Q1-2011.Q4. We exclude the annual series related to IP sectors from the annual GDP panel in order to avoid double counting. 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 $x^H_{m,t}$, $m = 1, ..., 4$, for year $t$, and let $X^L$ be the $(T, N_L)$ panel of the yearly growth rates of the non-IP indices. Let also $X_{HF} = [x^H_{1,1}, x^H_{2,1}, ..., x^H_{m,t}, ..., x^H_{4,T}]'$ be the $(4T, N_H)$ panel of quarterly IP indices growth rates.[22]

[19] Following Foerster, Sarte, and Watson (2011), we focus only on quarterly IP data, as they share the main features of the monthly ones but are less noisy/volatile. More details about the data are found in the OA, Section E.7. Note also that we cover the statistical factor model specification of Foerster, Sarte, and Watson (2011), not their structural analysis involving input-output linkages. The latter is studied separately as such analysis is more involved and - because of space limitations - beyond the scope of the current paper.

[20] The sectoral GDP data are not available at quarterly frequency (in contrast to the aggregate GDP index). An exhaustive description of the dataset is provided in the OA, Section E.7. All growth rates refer to seasonally adjusted real output indices, and are expressed in percentage points.

[21] A description of the practical implementation of our procedure appears in Section E.6 of the OA. It worth noting that 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, found in the OA, Section E.8.1. Our analysis confirms the findings of Foerster, Sarte, and Watson (2011), which justifies the use of a mixed frequency factor model to study the comovement among sectors.

[22] In view of the mixed frequency factor model we investigate two alternative approaches. The first approach applies PCA to high-frequency data and then aggregates the resulting PCA estimates while the second approach aggregates the data before PCA. We derive the asymptotic distribution for the estimators of these factors in the OA Section X.1 and find that under general conditions the two approaches give similar PCA estimates. However, under certain conditions

We start by selecting the number of factors in each sub-panel, which are of dimensions $k^C + k^H$ for $X^H$ and $X_{HF}$. We use the $IC_{p1}$ and $IC_{p2}$ information criteria of Bai and Ng (2002), following the empirical literature.[23] For the panels of IP growth rate at quarterly ($X_{HF}$) and annual ($X^H$) frequencies, $IC_{p1}$ selects two factors for each panel, whereas the more strict $IC_{p2}$ criterion selects one factor for $X_{HF}$ and two factors for $X^H$. For the annual GDP (non-IP) sectors panel, both $IC_{p1}$ and $IC_{p2}$ select a single factor.[24] Our results corroborate the evidence in Foerster, Sarte, and Watson (2011) suggesting that there is either one or two pervasive factors in the quarterly IP growth data. While the $IC_{p1}$ and $IC_{p2}$ choose factors in an unconditional setup, we are also interested in the explanatory power of these factors in a conditional setup. Hence the empirical analysis proceeds with two factors, $k^C + k^H = k^C + k^L = 2$, for each panel in order to avoid potentially omitted factors/variables in explaining economic activity growth and subsequently re-assess the conditional significance of factors.[25]

In order to select the number of common and frequency-specific factors, we follow our proposed procedure. The estimated canonical correlations of the first two PC's estimated in each sub-panel $X^H$ and $X^L$ are used to compute the value of the feasible standardized test statistic $\tilde{\xi}(r)$ in (4.6) and Theorem 2, for testing the null hypotheses of $r = 2$ and $r = 1$ common factors.[26] The first canonical correlation for the HF quarterly IP panel is $\hat{\rho}_1 = 0.84$, while the second one for the annual non-IP, GDP data is $\hat{\rho}_2 = 0.09$. These results are consistent with the presence of one common factor in each of the two mixed frequency datasets considered, as represented by hypothesis $H(1)$ in Section 3.1. The values of the statistics are $\tilde{\xi}(2) = -2.31$ and $\tilde{\xi}(1) = -0.91$ for the null hypotheses of $r = 2$ and $r = 1$ common factors, respectively. The test rejects the null hypothesis of the presence of two common factors ($r = 2$), for significance levels as small as $0.1\%$, while we cannot reject the null of one common factor at the $5\%$ significance level. Our selection procedure detailed in Proposition 2 with critical level as in (4.7), with $\gamma = 0.1$ and $c = 0.95$, produces the estimate $\hat{k}^C = 1$. That choice of $\gamma$ and $c$ proves to yield good results in the Monte Carlo experiments (see Section 5). Hence, we select a model with $k^C = k^H = k^L = 1$.

---

performing PCA first provides a lower bias asymptotically for estimating the common factor. The simulation results also support this approach in small samples.

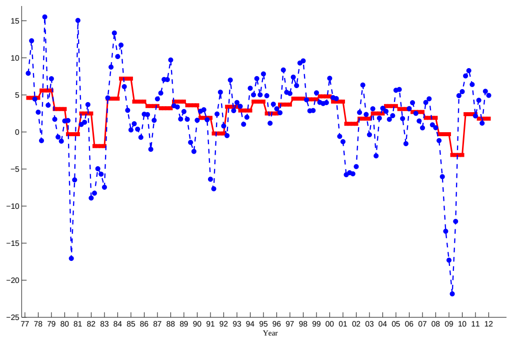
[23]Other information criteria can also be applied here (e.g. Onatski (2010), Ahn and Horenstein (2013)).

[24]We use $k_{max} = 15$ as maximum number of factors when computing $IC_{p1}$ and $IC_{p2}$.

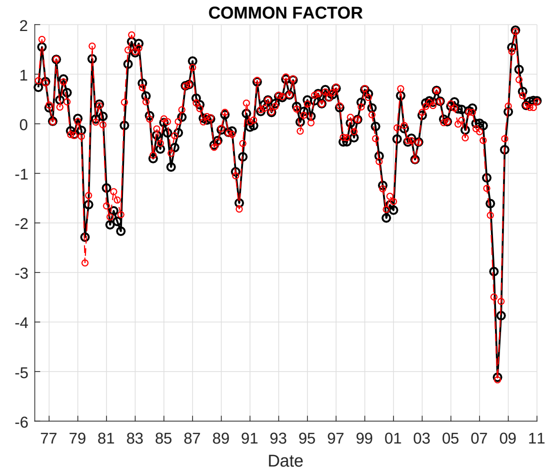
[25]The number of factors is also assessed in the conditional regression setup using the BIC criterion in Table 1. Note that Foerster, Sarte, and Watson (2011) also use two factors while they emphasize the importance of the first factor.

[26]Given the good finite sample properties presented in the simulations (in Section 5 and OA) for a range of DGPs, we expect that for our empirical application, the asymptotic theory also provides a good approximation.

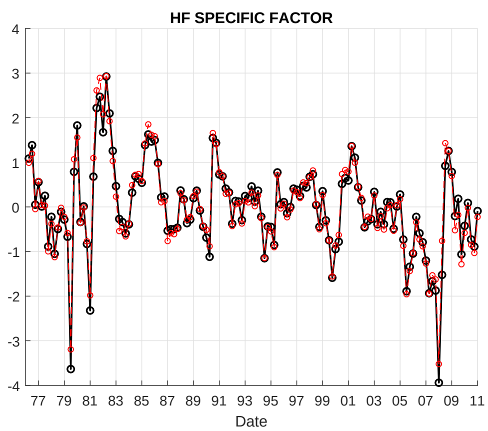
Figure 2: Sample paths of IP and GDP growth rates and the estimated factors, 1977 -2011



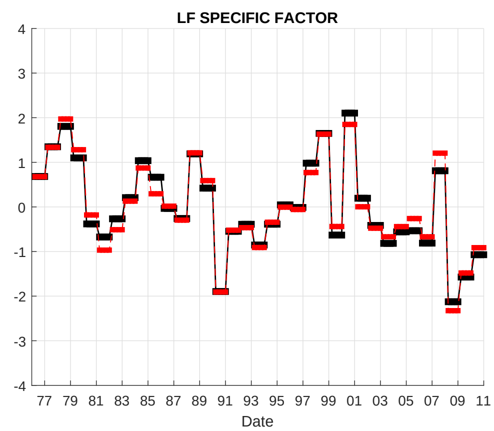

(a) IP and GDP growth rates




(b) Common factor




(c) HF-specific factor




(d) LF-specific factor

Panel (a) displays the dashed (blue) line which corresponds to the quarterly growth rates of the aggregate IP index for sample period 1977.Q1-2011.Q4, and the solid (red) line which represents the annual growth rates of GDP for the entire U.S. economy. Panel (b) displays the estimated common factor. Panel (c) displays that of the HF-specific factor and Panel (d) 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 group factor model with $k^C = k^H = k^L = 1$.

In Figure 2, Panel (a) plots the IP and GDP growth rates during the period 1977-2011 and the remaining Panels (b)-(d) present the estimated factor paths from the panels of 42 GDP sectors and 117 IP indices for the common, the HF-specific and the LF-specific factors, respectively. All factors are standardized to have zero mean and unit variance in the sample and their sign is chosen so that the majority of the associated loadings are positive. A visual inspection of the plots reveals that the common factor in Panel (b) resembles the IP index in Panel (a), with a large decline corresponding to

Table 1: Adjusted $R^2$ and percentage values of BIC of the regressions with common and/or frequency-specific factors from economic activity indices growth rates

| | $\bar{R}^2$: Quantiles | | | | | |
|---|---|---|---|---|---|---|
| Factors | 10% | 25% | 50% | 75% | 90% | % BIC |
| *Observables: Gross Domestic Product, 1977-2011* | | | | | | |
| common | -2.0 | -0.5 | 12.6 | 26.9 | 42.2 | 28.6 |
| common, LF-specific | 1.0 | 7.6 | 26.0 | 35.8 | 58.1 | 50.0 |
| LF-specific | -2.8 | -2.0 | 5.4 | 12.6 | 25.3 | 21.4 |
| *Observables: IP, 1977.Q1-2011.Q4* | | | | | | |
| common | 0.1 | 5.3 | 16.3 | 32.3 | 54.3 | 23.9 |
| common, HF-specific | 1.1 | 7.0 | 28.7 | 45.4 | 63.5 | 70.9 |
| HF-specific | -0.6 | 0.1 | 4.9 | 15.7 | 30.9 | 5.1 |

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 $k^C = k^H = k^L = 1$. Reported are the adjusted $R^2$ of the regressions on common and frequency-specific indices for different quantiles of the cross-section. The last column reports the percentage values that the BIC chooses the specific factor type regression model.

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 displayed in Panel (d) features a less dramatic fall during the Great Recession, and actually features a positive spike in 2008, followed by large negative values in the following years. This constitutes preliminary evidence suggesting that some non-IP sectors could feature different responses to the recent financial crisis.

The relationship of factors with the sectoral GDP and IP growth series, in a regression context, reveals additional information about the conditional correlations of the factors with specific economic activity growth sectors. This in turn can help us shed light on which IP and non-IP series are driving the factors. We start with a disaggregated analysis, and examine 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 GDP or IP index growth rates on (i) the common factor only, (ii) the specific factor only, for non-IP and IP series respectively, and (iii) both common and specific factors. In Table 1 we report the quantiles of the empirical distribution of the adjusted $R^2$ (denoted $\bar{R}^2$) of these regressions. In addition, we report the percentage value of the times the BIC (denoted by %BIC) selects, among the aforementioned three regression models (i)-(iii), the alternative factor conditional information set (common and/or frequency-specific), for each sectoral index in the cross-section.[27]

From the first three lines in Table 1 we observe that adding the LF-specific factor to the common factor regressions for the non-IP indices yields an increment of the median $\bar{R}^2$ around 13% (going from 12.6% to 26%) and the 90% quantile of $\bar{R}^2$ increases by 16%. Moreover, adjusting for the number of the variables in the factor regression models, the BIC favors the model with both the common and the LF-specific factors in explaining the GDP growth rate in 50% of the sectors, whereas the model with the CF alone is selected only in about 28% of the series. Similarly, the HF-specific factor, when added to the CF, contributes to an increment of around 12% in the median $\bar{R}^2$ for the IP sectors. The %BIC value provides strong evidence that both the CF and HF explain the IP growth rate, given that the BIC favors this model by 71% in the cross-section, vis-à-vis the models with either the CF or HF. Overall, Table 1 confirms that both the common and frequency-specific factors explain a significant part of the variability of output growth for the majority of the sectors of the U.S. economy. These models are favored by the BIC criterion for 50% up to 71% in the cross-section of the various economic sectors. In addition, the results in Table 1 show that, the CF turns out to be pervasive for most of the IP and non-IP sectors alike as demonstrated by both the relative $\bar{R}^2$ (and %BIC) vis-à-vis those with just the frequency-specific factor. In order to investigate which sectors drive the variation of our estimated factors and provide an economic interpretation to our factors, we list in Table 2 the highest and lowest ten GDP non-IP sectors in terms of $\bar{R}^2$ when regressed on the common factor only (in Panel A), and both the common and LF-specific factors (in Panel B). We also report the top and bottom ten ranked 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 (in Panel C).[28] From Panel A we first note that the CF alone explains most of the variability of service sectors with direct economic links to IP sectors like Transportation and Warehousing with an $\bar{R}^2$ ranging from 63% to 41%, as well as Construction with $\bar{R}^2$ of 44%. This is another clear indication that the CF is driven by service sectors related to IP and could thereby be interpreted as an IP factor, as already noted on Figure 2. On the other hand, the CF turns out to be completely unrelated to most of the Financial, Insurance and Information services

[27]The regressions in the second and third rows are restricted MIDAS regressions. Those in the 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 $\bar{R}^2$ corresponding to the first and second lines (resp., fourth and fifth lines) of Table 1 are represented in the histograms available in OA, Figures E.10 *(a)* and *(b)* (resp., *(c)* and *(d)*).

[28]The entire list of non-IP sectors ranked by the three criteria used in Table 2, are available in Tables E.24-E.26 in the OA, Section E.8.

Table 2: Regression of yearly sectoral GDP growth on 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 adj. $R^2$ in Panels A and B** | |
|---|---|---|---|---|---|
| **Sector** | $\bar{R}^2$ | **Sector** | $\bar{R}^2$ | **Sector** | $\Delta\bar{R}^2$ |
| *Ten sectors with largest $\bar{R}^2$* | | *Ten sectors with largest $\bar{R}^2$* | | *Ten sectors with largest $\Delta\bar{R}^2$* | |
| Truck transportation | 63.10 | Misc. prof., scient., & tech. serv. | 66.67 | Misc. prof., scient., & tech. serv. | 49.69 |
| Accommodation | 62.43 | Admin. & support services | 62.63 | Gov. enterprises (state & local) | 34.69 |
| Construction | 44.05 | Truck transportation | 62.51 | Rental & leasing serv. | 29.52 |
| Other transp. & support activ. | 43.31 | Accommodation | 61.48 | General gov. (state & local) | 24.90 |
| Administrative & support services | 42.69 | Construction | 59.75 | Legal services | 24.32 |
| Other services, except gov. | 42.53 | Warehousing & storage | 52.53 | Motion picture & sound rec. | 22.77 |
| Warehousing & storage | 40.95 | gov. enterprises (state & local) | 45.78 | Fed. Res. banks, credit interm.. | 20.31 |
| Air transportation | 31.58 | Other services, except gov. | 41.75 | Administrative & support services | 19.95 |
| Retail trade | 30.70 | Other transportation & support act. | 41.71 | Social assistance | 19.91 |
| Amusem., gambling, & recr. ind. | 29.17 | gov. enterprises (federal) | 37.78 | Real estate | 18.14 |
| *Ten sectors with smallest $\bar{R}^2$* | | *Ten sectors with smallest $\bar{R}^2$* | | *Ten sectors with smallest $\Delta\bar{R}^2$* | |
| 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 comp. & enterpr. | 7.52 | Rail transportation | -1.16 |
| Pipeline transportation | -1.74 | Funds, trusts, & other fin. vehicles | 6.15 | Other transportation & support act. | -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 gov. (state & local) | -2.12 | Insurance carriers & related activities | 0.36 | Amusements, gambling | -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 contr., & investm. | -2.99 | Forestry, fishing | -5.31 | Forestry, fishing | -2.98 |
| Insurance carriers | -3.03 | Securities, commodity contr. | -5.99 | Securities, commodity contr. | -3.00 |

The adjusted $R^2$, denoted $\bar{R}^2$, are reported for the restricted MIDAS regressions of the growth rates of 42 GDP non-IP sectoral indices on the estimated factors. Regressions in *Panel A* involve a LF explained variable and the estimated CF. Regressions in *Panel B* involve a LF explained variable and both the common and LF-specific factors. 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*.

sectors. Turning to Panel C of Table 2, which reports the difference in $\bar{R}^2$ between the regressions in Panels A and B, we note that the LF-specific factor explains more than 20% of the variability of output for very heterogeneous services sectors as well as Government (state and local).[29] Interpreting these results, we conclude that the LF-specific factor is completely unrelated to service sectors which depend almost exclusively on IP output (e.g. transportation, retail trade), and is a common factor driving the comovement of other non-IP service sectors, such as Professional scientific and technical services, Government, legal services.

In Table 2 we highlight further differences in the dynamics of output growth between the two subsectors of the financial services industry which are particularly revealing, the Securities and Credit intermediation, extensively studied by Greenwood and Scharfstein (2013). We find that the subsectors Funds, trusts, and other financial vehicles as well as Securities, commodity contracts, and investments,

[29] Such services include Miscellaneous professional, scientific, and technical services, Administrative and support services, Legal services, Real estate, some important financial services like Federal Reserve banks, Credit intermediation, and Related activities, Rental and leasing services.

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 and across the other sectors that correlate with the U.S. economic activity. In contrast, the Credit intermediation industry comoves with the other IP and non-IP sectors (see also Tables E.24 and E.25 in the OA).

Up to this point, we examined the explanatory power of the factors for sectoral output indices. For non-IP GDP, these indices correspond to the finest level of disaggregation of output growth by sector. In Table 3 we report the results of regressions with aggregated indices instead. In particular, we regress the output of each aggregate index either on the estimated (a) common factor, (b) frequency-specific or (c) both aforementioned factors, and report the corresponding $\bar{R}^2$ of these regressions in the first three columns. The last column in Table 3 reports the model favored by the BIC among the three regression specifications. It is important to note that now we also include the GDP Manufacturing aggregate index which is *not* used in the estimation of the factors. Panel A in Table 3 shows that the CF explains around 84% of the variability in the aggregate IP growth index, confirming that this factor can be interpreted as an Industrial Production factor. This is further corroborated in Panel B where we obtain an $\bar{R}^2$ of 84% in the regression of the GDP Manufacturing Index on the common factor alone. As most of the sectors included in the Industrial Production index are Manufacturing sectors, this result is not surprising. Yet, it is still worth noting because, as remarked earlier, the GDP data on Manufacturing have not been used in the factor estimation, in order to avoid double-counting these sectors in our mixed frequency sectoral panel.[30]

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%, still 60% of its total variability can be explained exclusively by the common factor which - as shown in Panel B - 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 (since the $\bar{R}^2$ increases by about 15% from 61% to 76%).[31] This indicates that significant comovements are present among the most important sectors of the U.S. economy which are not related to manufacturing. Indeed, Panel B in Table 3 indicates that some services sectors such as Professional and Business

[30] A detailed discussion of the difference in the sectoral components of the IP index and the GDP Manufacturing index is provided in OA, Section E.7.

[31] See the results in Table E.27 in the OA.

Services and Information, Transportation and Warehousing, and Construction load significantly both on the common and the LF-specific factor, while some other sectors like Finance and Government load exclusively on the LF-specific factor.[32]

The BIC in Table 3, Panel B, favors the regression model with both the common and low frequency factors, among the three factor regression specifications for the U.S. GDP growth rate, while the low frequency factor alone yields a low $\bar{R}^2$ of 9%. Similarly, although the HF-specific factor in Panel A seems to be relatively less important in explaining the aggregate IP index (as the $\bar{R}^2$ increases by only 7% when it is added as a regressor to the common factor regression model for the IP growth rate), the BIC suggests that both the common and HF factors are important.[33] Overall the small $\bar{R}^2$ could suggest 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 U.S. economy. These 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. It is worth emphasizing that the common factor explains the dominant 85% of the variability of the total IP growth and 60% of the GDP growth.

Given that our sample period covers the Great Moderation, characterized by a reduction in the volatility of business cycle fluctuations starting in the mid-1980s, we revisit this analysis for different subsamples. The details can be found in the OA, Section E.8.4, while we discuss here briefly the main results. We find a deterioration of the overall fit of approximate factor models during the Great Moderation period starting in 1984 and ending in 2007 – a finding also reported by Foerster, Sarte, and Watson (2011) – where our common factor plays a relatively less significant role during that period. Interestingly, though 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.

[32] The results change when we look at the Finance sector disaggregated in Fed. Reserve banks, credit interm., and rel. activ., Securities, commodity contracts, and investm., Insurance carriers and related activities, as evident in Table 2.

[33] See also Table E.27 in OA, Section E.8, for the $\bar{R}^2$ of the regression of all GDP indices on the HF factor only, and all the 3 factors together.

Table 3: Regression results of aggregate IP and selected GDP indices growth rates on estimated factors

**Panel A** *Quarterly observations, 1977.Q1-2011.Q4*

| Sector | (1) $\bar{R}^2(C)$ | (2) $\bar{R}^2(H)$ | (3) $\bar{R}^2(C+H)$ | (3) - (1) | BIC |
|---|---|---|---|---|---|
| Industrial Production | 83.47 | 14.29 | 90.19 | 6.73 | CH |

**Panel B** *Yearly observations, 1977-2011*

| Sector | (1) $\bar{R}^2(C)$ | (2) $\bar{R}^2(L)$ | (3) $\bar{R}^2(C+L)$ | (3) - (1) | |
|---|---|---|---|---|---|
| GDP | 60.83 | 9.12 | 75.51 | 14.68 | CL |
| GDP - Manufacturing | 83.96 | -3.00 | 83.57 | -0.39 | C |
| GDP - Agriculture, forestry, fishing, and hunting | 0.12 | -2.78 | -2.83 | -2.95 | C |
| GDP - Construction | 43.12 | 13.10 | 61.26 | 18.13 | CL |
| GDP - Wholesale trade | 19.18 | 9.67 | 31.75 | 12.57 | CL |
| GDP - Retail trade | 25.25 | -2.88 | 23.39 | -1.87 | C |
| GDP - Transportation and warehousing | 64.03 | -2.46 | 63.01 | -1.02 | C |
| GDP - Information | 11.52 | 21.06 | 35.83 | 24.31 | CL |
| GDP - Finance, insurance, real estate, rental, and leasing | -1.08 | 20.52 | 20.90 | 21.97 | L |
| GDP - Professional and business services | 30.99 | 29.94 | 66.81 | 35.83 | CL |
| GDP - Educational services, health care, and social assistance | -0.70 | 14.18 | 14.68 | 15.37 | L |
| GDP - Arts, entertainment, recreation, accomm., and food | 48.10 | -1.62 | 49.05 | 0.95 | C |
| GDP - Government | -2.11 | 23.71 | 21.80 | 23.91 | L |

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)) are reported. The fourth column displays the difference between the values in the third and first columns. The last column reports the choice of the BIC across the regression models with the common factor, or the frequency-specific factor, or both factors (C denotes the common factor, H denotes the high frequency factor and L denotes the low frequency factor and corresponding factor combinations (CL and CH) in the regression models). The factors are estimated from the panel of 42 GDP non-IP sectors and 117 IP indices using a mixed frequency factor model with $k^C = k^H = k^L = 1$.

# 7 Conclusions

We develop a general theoretical framework for group factor models and develop a unified asymptotic theory for the identification of common and group factors, for the determination of the number of common and specific factors, 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, using asymptotic expansions both in the cross-sections and time series. Of special interest is the group factor mixed frequency model for which the data panels of different/mixed frequencies allow not only for a natural grouping in extracting factors but also a framework which has the advantage of identifying and estimating factors which are common across frequencies as well as frequency specific.

Our theoretical contributions are of interest beyond this paper. Theorems 1 and 2 are of interest beyond (mixed frequency) group factor models. Inference regarding the rank of an unknown, real-valued matrix is an important and well-studied problem.[34] 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, 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 $V$, needs to be estimated using a semi-definite estimator, say $\hat{V}$, the asymptotic variance-covariance matrix of this estimator - denoted as $W_0$ - is necessarily singular, as shown in Donald, Fortuna, and Pipiras (2007). Therefore standard rank tests cannot be applied as they assume that the matrix $W_0$ is full rank. In addition, our results (in Section 4) provide the guidance to 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 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. Furthermore, our methods can be used to develop a new test for the question posed by Pelger (2015) as to whether the factor spaces of statistical and economic factors are equal.

There is 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 IP quarterly sectors using factor models. We revisit part of their analysis and incorporate the rest - and most dominant part - of the U.S. economy, namely the non-IP sectors whose growth rate we only observe annually. We find evidence for a single common factor among IP and non-IP sectors, and document that 85% of this common factor is explained by IP sector growth.

Despite the generality of our analysis, we can think of many possible extensions, such as models with loadings which change across sub-periods, i.e. periodic loadings, or loadings which vary stochas-

[34] See for instance Gill and Lewbel (1992), Cragg and Donald (1996), Robin and Smith (2000) and Kleibergen and Paap (2006).

tically 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, and therefore more than two sampling frequencies. All these extensions are left for future research.

# References

AHN, S. C., AND A. R. HORENSTEIN (2013): "Eigenvalue Ratio Test for the Number of Factors," *Econometrica*, 81, 1203–1227.

ALDOUS, D., AND G. EAGLESON (1963): "On Mixing and Stability of Limit Theorem," *Annals of Probability*, 6, 325–331.

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, 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, 556–604.

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

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

——— (2006): "Confidence Intervals for Diffusion Index Forecasts and Inference for Factor-Augmented Regressions," *Econometrica*, 74, 1133–1150.

——— (2007): "Determining the Number of Primitive Shocks in Factor Models," *Journal of Business and Economic Statistics*, 25, 52–60.

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

BOSQ, D. (1998): *Nonparametric Statistics for Stochastic Processes*. Springer-Verlag.

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.

CHAMBERLAIN, G., AND M. ROTHSCHILD (1983): "Arbitrage, Factor Structure and Mean-Variance Analysis on Large Asset Markets," *Econometrica*, 51, 1281–1304.

CHEN, P. (2010): "A Grouped Factor Model," MPRA Paper No. 36082, Online at `http://mpra.ub.uni-muenchen.de/36082/`.

——— (2012): "Common Factors and Specific Factors," *Working Paper*, MPRA Paper No. 36114, Online at `http://mpra.ub.uni-muenchen.de/36114/`.

CONNOR, G., AND R. A. KORAJCZYK (1986): "Performance Measurement with the Arbitrage Pricing Theory: A new Framework for Analysis," *Journal of Financial Economics*, 15, 373–394.

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.

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

DONALD, S. G., N. FORTUNA, AND V. PIPIRAS (2007): "On Rank Estimation in Symmetric Matrices: The Case of Indefinite Matrix Estimators," *Econometric Theory*, 23, 1103–1123.

——— (2014): "On Rank Estimation in Semidefinite Matrices," *Working Paper*, Working Paper, Universidade do Porto, Faculdade de Economia do Porto.

FLURY, B. N. (1984): "Common Principal Components in k Groups," *Journal of the American Statistical Association*, 79, 892–898.

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, 1–38.

GHYSELS, E. (2016): "Macroeconomics and the Reality of Mixed Frequency Data," *Journal of Econometrics*, 193, 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, 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, 423 – 451.

HALL, P., AND C. HEYDE (1980): *Martingale Limit Theory and Its Application*. Academic Press.

KLEIBERGEN, F., AND R. PAAP (2006): "Generalized Reduced Rank Tests Using the Singular Value Decomposition," *Journal of Econometrics*, 133, 97–126.

KOSE, A. M., C. OTROK, AND C. H. WHITEMAN (2008): "Understanding the Evolution of World Business Cycles," *Journal of International Economics*, 75, 110–130.

KUERSTEINER, G., AND I. PRUCHA (2013): "Limit Theory for Panel Data Models with Cross Sectional Dependence and Sequential Exogeneity," *Journal of Econometrics*, 174, 107–126.

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

ONATSKI, A. (2010): "Determining the Number of Factors from Empirical Distribution of Eigenvalues," *Review of Economics and Statistics*, 92, 1004–1016.

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," *Annals of Statistics*, 11, 872–885.

RENYI, A. (1963): "On Stable Sequences of Events," *Sankhya*, 25, 293–302.

ROBIN, J. M., AND R. SMITH (2000): “Tests of Rank,” *Econometric Theory*, 16, 151–175.

SCHOTT, J. R. (1991): “Some Tests for Common Principal Component Subspaces in Several Groups,” *Biometrika*, 78, 771–777.

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

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

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

WANG, P. (2012): “Large Dimensional Factor Models with a Multi-Level Factor Structure: Identifcation, Estimation, and Inference,” Working Paper, Hong Kong University of Science and Technology.

# Appendices

We use the following notation. Let $\|A\| = \sqrt{tr(A'A)}$ denote the Frobenius norm of matrix $A$. We denote by $\|Z\|_p = (E[\|Z\|^p])^{1/p}$ the $L^p$-norm of random matrix $Z$, for $p > 0$. We denote by $\stackrel{d}{\to}$ convergence in distribution. For a sigma-field $\mathcal{F}$, we denote by $Z_n \stackrel{d}{\to} Z$ ($\mathcal{F}$-stably) the stable convergence on $\mathcal{F}$ of a sequence of random vectors, that is, $P(Z_n \in A, U) \to P(Z \in A, U)$ as $n \to \infty$, for any Borel set $A$ with $P(Z \in \partial A) = 0$, where $\partial A$ is the boundary of set $A$, and any measurable set $U \in \mathcal{F}$ (see e.g. Renyi (1963), Aldous and Eagleson (1963), Hall and Heyde (1980), Kuersteiner and Prucha (2013)). In particular, for a symmetric positive definite random matrix $\Omega$ measurable with respect to $\mathcal{F}$, by $Z_n \stackrel{d}{\to} N(0, \Omega)$ ($\mathcal{F}$-stably) we mean $Z_n \stackrel{d}{\to} \Omega^{1/2}\varepsilon$ ($\mathcal{F}$-stably), where $\varepsilon \sim N(0, I)$ is independent of $\mathcal{F}$.

# A Assumptions

We make the following assumptions:

**Assumption A.1.** *We have $N_1, N_2, T \to \infty$ such that the conditions in (4.1) hold.*

**Assumption A.2.** *The unobservable factor process $F_t = [\ f_t^{c\,\prime},\ \ f_{1,t}^{s\,\prime},\ \ f_{2,t}^{s\,\prime}\ ]'$ satisfies the normalization restrictions in (2.2), with $\Sigma_F = V(F_t)$ positive-definite.*

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

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

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

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

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

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

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

**Assumption A.6.** *a) The triangular array processes* $V_t \equiv V_{N_1,N_2,t} = [h'_{j,t}, \xi'_{j,t}, j = 1,2]'$ *and* $V^*_t \equiv V^*_{N_1,N_2,t} = [\kappa_{j,t}, \eta^2_{j,t}, j = 1,2]'$ *are strong mixing of size* $-\frac{r}{r-2}$*, uniformly in* $N_1, N_2 \geq 1$.[35] *Moreover*
*b)* $\|E(\xi_{j,t}\xi'_{k,t}|\mathcal{F}_t) - E(\xi_{j,t}\xi'_{k,t}|F_t, ..., F_{t-m})\|_2 = O(m^{-\psi})$*, as* $m \to \infty$*, uniformly in* $N_1, N_2 \geq 1$*, and*
*c)* $\|\eta^2_{j,t} - E(\eta^2_{j,t}|\mathcal{V}^{t+m}_{t-m})\|_8 = O(m^{-\psi})$*, as* $m \to \infty$*, uniformly in* $N_1, N_2 \geq 1$*,*
*for* $j, k = 1, 2$*, where* $\mathcal{V}^{t+m}_{t-m} = \sigma(V_s, t - m \leq s \leq t + m)$ *and* $\psi > 1$.

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

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

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

Assumption A.1 defines the asymptotic scheme. Assumption A.2 concerns the first- and second-order moments of the factor vector. Positive definiteness of the variance-covariance matrix $\Sigma_F$ is necessary for our model to have exactly $k^c + k^s_1 + k^s_2$ pervasive factors. It holds if, and only if, the eigenvalues of matrix $\Phi\Phi'$ are smaller than 1 in modulus. The zero restrictions on the matrix $\Sigma_F$ in (2.2), corresponding to the orthogonality of the common and group-specific factors, as well as the identity diagonal blocks, are identification conditions. Assumption A.3 concerns the empirical cross-sectional second-order moment matrix of the loadings in each group $j = 1,2$. It implies that matrix $\Lambda_j$ has full column-rank, for $N_j$ large enough, $j = 1,2$. Positive definiteness of matrix $\Sigma_{\lambda,j}$, for $j = 1,2$, is also necessary for the existence of exactly $k^c + k^s_1 + k^s_2$ pervasive factors. Note that we consider non-random loadings to simplify the assumptions and proofs. If the loadings were random, stochastic convergence could be obtained with a DGP for the loadings which satisfies the conditions of the LLN for weakly dependent data. Assumptions A.2 and A.3 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 (2006), among others).

Assumption A.4 requires the existence of higher order moments for the factors and the error terms, similarly as e.g. in Assumptions A and C.1 in Bai and Ng (2002) and Bai (2003).

Assumption A.5 constraints the amount of admissible cross-sectional dependence of the error terms across different individuals, in the spirit of the framework - introduced by Chamberlain and Rothschild (1983) - of weak cross-sectional dependence characterizing "approximate factor models". No distributional assumption is made on the idiosyncratic terms. Assumption A.5 *a)* states that the cross-sectional averages of the error

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

terms scaled by factor loadings satisfy a CLT. It corresponds to Assumption F.3 in Bai (2003). We adopt stable convergence on the sigma-field $\mathcal{F}_t$ to allow for the asymptotic variance-covariance matrix $\Omega_t(h)$ to possibly depend on common factors. That would occur e.g. if there are common components in the conditional volatility processes of the idiosyncratic errors. Assumption F.3 in Bai (2003) applies if the trivial filtration is replaced for $\mathcal{F}_t$. Assumption A.5 *b)* concerns higher-order conditional moments of the scaled cross-sectional average of error terms. A sufficient condition for Assumption A.5 *b)* with $r = 3$ is $\|\lambda_{j,i}\| \leq M$ and $1/N_j^3 \sum_{i_1,i_2,\dots,i_6=1}^{N_j} |E[\varepsilon_{j,i_1,t}\varepsilon_{j,i_2,t}...\varepsilon_{j,i_6,t}|\mathcal{F}_t]| \leq M$, a.s., for all $N_j \geq 1$ and $j = 1, 2$. For $r = 1$ it corresponds to Assumption C.3 in Bai (2003). Assumption A.5 *c)* concerns the fourth-order moment of cross-sectional averages of squared error terms and corresponds to Assumption C.5 in Bai (2003).

Assumption A.6 allows for weak serial dependence in error terms and factor processes. Specifically, Assumption A.6 *a)* is a strong mixing condition, where (minus) the mixing size is inversely related to the moment order $r$ introduced in Assumptions A.4 and A.5. We rely on this specific concept of time-series dependence because we use a CLT for data that are Near-Epoch Dependent (NED) on mixing processes (see e.g. Davidson (1994)), to show the asymptotic Gaussian distribution of the test statistic in Theorem 1. We deploy this specific version of the CLT for dependent data as it allows us to cope with the rather complex nature of the leading term in the asymptotic expansion of the test statistic, that involves the time-series average of the *square* of a cross-sectional average of scaled errors (instead of an average of averages as in the asymptotic expansion of factor estimates). We use Assumptions A.3, A.4, A.5 *a)-b)* and A.6 to check the conditions of the CLT in Section B.1.6 *i)*. Assumptions A.6 *b)* and *c)* require that certain quantities are well-approximated by their projection on a finite number of components of a mixing process to apply the NED property.

Assumption A.7 consists of additional restrictions on the weak cross-sectional and time-series dependence of the error and factor processes, which are used to prove the asymptotic expansions for the PCA estimates of the pervasive factors in the two groups in Proposition B.1. Specifically, Assumption A.7 *a)* concerns cross-sectional averages of cross-products of error terms at different dates. It requires both that these cross-sectional averages are close to the corresponding population covariances in the large sample limit, and that the latter covariances decay with the time lag in a summable way. Assumptions A.7 *b)* and *c)* provide bounds on terms involving processes $\alpha_t$, $\beta_t$ and $\bar{\beta}_t$, that consist of averages of products of error terms at different dates.

Assumptions A.5, A.6 and A.7 yield conditions of weak cross-sectional and time-series dependence to control terms such as those in Assumptions C, D, E and F.1-F.3 in Bai (2003). They could be substituted, at the expense of more elaborated proofs, by other weak dependence assumptions for factors and idiosyncratic errors.

Assumption A.8 is used to get bounds on the remainder terms in the asymptotic expansions of estimated factors and loadings in Lemma X.2 uniformly across $i$ and $t$. These bounds are used to control the estimation error for the recentering and rescaling terms of the feasible test statistic in Theorem 2. Specifically, Assumption A.8 *a)* is a tail condition on the factor stationary distribution, Assumption A.8 *b)* constraints the amount of cross-sectional dependence of the error terms, while Assumption A.8 *d)* is a uniform bound on true factor loadings. In Assumption A.8 *c)* we require that time series averages of certain zero-mean processes involving error terms and factors satisfy a large deviation bound. Such a large deviation bound is implied by tail conditions plus restrictions on serial dependence like strong mixing (see e.g. Theorems 3.1 and 3.2 in Bosq (1998)).

Assumption A.9 simplifies the derivation of the feasible asymptotic distribution of the statistic in Theorem 2. This condition excludes correlation of the error terms across individuals and time (conditional on the factors), as well as conditional heteroschedasticity, and implies a "strict factor model" for each group. In that sense, it is more restrictive than Assumptions A.5, A.6, A.7 and A.8 *b)-c)*. Moreover, under Assumption A.9, the matrix $\Omega_{jj,t}(0)$ in Assumption A.5 *a)* simplifies to $\Omega_{jj} = \lim_{N_j \to \infty} (1/N_j) \sum_{i=1}^{N_j} \lambda_{j,i}\lambda'_{j,i}\gamma_{j,ii}$ , while $\Omega_{jk,t}(h) = 0$ if either $h \neq 0$, or $j \neq k$. We note that, Assumption A.9 simplifies substantially the proof of Theorem 2, but is not needed in the proofs of Theorem 1 and Propositions 2 and X.1.

# B Proofs

## B.1 Proof of Theorem 1

The proof of Theorem 1 is structured as follows. We start by deriving an asymptotic expansion for the estimates of the pervasive factors extracted by PCA in each group (Subsection B.1.1). This result yields an asymptotic expansion for the sample canonical correlation matrix $\hat{R}$ (Subsection B.1.2), and in turn it is used to obtain the asymptotic expansions of the eigenvalues and eigenvectors of matrix $\hat{R}$ by perturbation methods (Subsections B.1.3 and B.1.4). This yields the asymptotic expansions of the canonical correlations and of the test statistic $\hat{\xi}(k^c)$ (Subsection B.1.5). Finally, the asymptotic Gaussian distribution of the test statistic follows by applying a suitable CLT for dependent triangular arrays (Subsection B.1.6).

### B.1.1 Asymptotic expansion of the factor estimates $\hat{h}_{j,t}$

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

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

$u_{j,t} = \left( \frac{1}{N_j} \sum_{i=1}^{N_j} \lambda_{j,i} \lambda'_{j,i} \right)^{-1} \frac{1}{\sqrt{N_j}} \sum_{i=1}^{N_j} \lambda_{j,i} \varepsilon_{j,i,t}$, $b_{j,t} = \left( \frac{1}{N_j} \sum_{i=1}^{N_j} \lambda_{j,i} \lambda'_{j,i} \right)^{-1} \left( \frac{1}{T} \sum_{t=1}^{T} h_{j,t} h'_{j,t} \right)^{-1} \eta_{j,t}^2 h_{j,t}$, $\lambda_{j,i} = (\lambda_{j,i}^{c\prime}, \lambda_{j,i}^{s\prime})'$, $d_{j,t} = \left( \frac{1}{N_j} \sum_{i=1}^{N_j} \lambda_{j,i} \lambda'_{j,i} \right)^{-1} \left( \frac{1}{T} \sum_{t=1}^{T} h_{j,t} h'_{j,t} \right)^{-1} \left( \frac{1}{\sqrt{N_j T}} \sum_{i=1}^{N_j} \sum_{s=1}^{T} \varepsilon_{j,i,s} h_{j,s} \lambda'_{j,i} \right) h_{j,t}$, *as* $N_1, N_2, T \to \infty$ *terms* $\vartheta_{j,t}$ *are such that* $\frac{1}{T} \sum_{t=1}^{T} \left[ \frac{1}{\sqrt{N_j}} u_{j,t} + \frac{1}{T} b_{j,t} + \frac{1}{\sqrt{N_j T}} d_{j,t} + \vartheta_{j,t} \right] \vartheta'_{k,t} = o_p \left( \frac{1}{N\sqrt{T}} \right)$ *and* $\frac{1}{T} \sum_{t=1}^{T} h_{j,t} \vartheta'_{k,t} = O_p \left( \frac{1}{N} + \frac{1}{T^2} \right)$, *and the matrix* $\hat{\mathcal{H}}_j$ *converges in probability to a nonstochastic orthogonal* $(k_j, k_j)$ *matrix, for* $j, k = 1, 2$.

### B.1.2 Asymptotic expansion of matrix $\hat{R}$

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 expansion (B.1). We get:

$$\hat{h}_{j,t} = h_{j,t} + \psi_{j,t}, \quad j = 1, 2, \tag{B.2}$$

where

$$\psi_{j,t} := \frac{1}{\sqrt{N_j}} u_{j,t} + \frac{1}{T} b_{j,t} + \frac{1}{\sqrt{N_j T}} d_{j,t} + \vartheta_{j,t}, \tag{B.3}$$

gathers the factor estimation error terms. By using equation (B.2), we have:

$$\hat{V}_{j,k} = \frac{1}{T} \sum_{t=1}^{T} \hat{h}_{j,t} \hat{h}'_{k,t} = \frac{1}{T} \sum_{t=1}^{T} (h_{j,t} + \psi_{j,t}) (h_{k,t} + \psi_{k,t})' = \tilde{V}_{j,k} + \hat{X}_{j,k}, \tag{B.4}$$

where:

$$\tilde{V}_{j,k} = \frac{1}{T}\sum_{t=1}^{T} h_{j,t}h'_{k,t} \, , \tag{B.5}$$

$$\hat{X}_{j,k} = \frac{1}{T}\sum_{t=1}^{T}(h_{j,t}\psi'_{k,t} + \psi_{j,t}h'_{k,t}) + \frac{1}{T}\sum_{t=1}^{T}\psi_{j,t}\psi'_{k,t}, \tag{B.6}$$

for $j,k = 1,2$. From the definition of matrix $\hat{R}$ in (3.1), and by using (B.4) and $\hat{V}_{jj}^{-1} = \left(I_{k_j} + \tilde{V}_{jj}^{-1}\hat{X}_{jj}\right)^{-1}\tilde{V}_{jj}^{-1}$, we get:

$$\hat{R} = \left(I_{k_1} + \tilde{V}_{11}^{-1}\hat{X}_{11}\right)^{-1}\tilde{V}_{11}^{-1}\left(\tilde{V}_{12} + \hat{X}_{12}\right)\left(I_{k_2} + \tilde{V}_{22}^{-1}\hat{X}_{22}\right)^{-1}\tilde{V}_{22}^{-1}\left(\tilde{V}_{21} + \hat{X}_{21}\right). \tag{B.7}$$

By using the definition of $\psi_{j,t}$ in equation (B.3), in the next Lemma we derive an upper bound for terms $\hat{X}_{j,k}$, $j,k = 1,2$.

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

Let us now expand matrix $\hat{R}$ at second order in the $\hat{X}_{j,k}$. The reason for going beyond the first order is the following. It turns out that the first-order contribution of the $\hat{X}_{j,k}$ to the statistic of interest involves leading terms of stochastic orders $O_p\left(\frac{1}{N\sqrt{T}}\right)$ and $O_p\left(\frac{1}{T\sqrt{NT}}\right)$ (see Lemma B.5 below). The second-order remainder term is $O_p(\delta_{N,T}^2)$, and $\delta_{N,T}^2$ is not negligible with respect to $\max\{\frac{1}{N\sqrt{T}}, \frac{1}{T\sqrt{NT}}\}$, when either $T$ is too small compared to $N$, or $N$ is too small compared to $T$. In order to get validity of our results for more general conditions on the relative growth rate of $N$ and $T$ such as in Assumption 4.1, we consider a second-order expansion. By using $(I - X)^{-1} = I + X + X^2 + O_p(\delta_{N,T}^3)$ for $X = O_p(\delta_{N,T})$, from (B.7) we get the next Lemma.

**LEMMA B.2.** *Under Assumptions 4.1-B.1 and A.2-A.4, A.5 b)-c) and A.6 a), the second-order asymptotic expansion of matrix* $\hat{R}$ *is:*

$$\hat{R} = \tilde{R} + \hat{\Psi} + O_p(\delta_{N,T}^3), \tag{B.8}$$

*where* $\tilde{R} = \tilde{V}_{11}^{-1}\tilde{V}_{12}\tilde{V}_{22}^{-1}\tilde{V}_{21}$ *and:*

$$\hat{\Psi} = \tilde{V}_{11}^{-1}\hat{\Psi}^*, \qquad \hat{\Psi}^* = \hat{\Psi}^{*\,(I)} + \hat{\Psi}^{*\,(II)}, \tag{B.9}$$

$$\hat{\Psi}^{*\,(I)} = -\hat{X}_{11}\tilde{R} + \hat{X}_{12}\tilde{B} - \tilde{B}'\hat{X}_{22}\tilde{B} + \tilde{B}'\hat{X}_{21}, \tag{B.10}$$

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

*with* $\tilde{B} = \tilde{V}_{22}^{-1}\tilde{V}_{21}$.
***Proof****: See Appendix C.2.*

Equation (B.8) represents matrix $\hat{R}$ as the sum of the sample canonical correlation matrix $\tilde{R}$ computed with the true factor values, the estimation error term $\hat{\Psi}$ that consists of first-order and second-order components $\hat{\Psi}^{*\,(I)}$ and $\hat{\Psi}^{*\,(II)}$, and the third-order remainder term $O_p(\delta_{N,T}^3)$.

### B.1.3 Matrix $\tilde{R}$ and its eigenvalues and eigenvectors

Let us now characterize matrix $\tilde{R}$ and its eigenvalues, that are $\tilde{\rho}_1^2, ..., \tilde{\rho}_{k_1}^2$, 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 $\tilde{\rho}_1 = ... = \tilde{\rho}_{k^c} = 1$ a.s.. Using the notation:

$$\tilde{\Sigma}_{cc} = \frac{1}{T}\sum_{t=1}^{T} f_t^c f_t^{c\prime}, \qquad \tilde{\Sigma}_{c,j} = \frac{1}{T}\sum_{t=1}^{T} f_t^c f_{j,t}^{s\prime}, \qquad \tilde{\Sigma}_{j,c} = \tilde{\Sigma}_{c,j}', \qquad \tilde{\Sigma}_{j,k} = \frac{1}{T}\sum_{t=1}^{T} f_{j,t}^s f_{k,t}^{s\prime}, \quad j,k = 1,2,$$

we can write matrices $\tilde{V}_{j,k}$, with $j,k = 1,2$, in (B.5) in block form as:

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

From Assumptions A.2, A.4 *b)* and A.6 *a)*, and Corollary 14.3 in Davidson (1994), we have:

$$\tilde{V}_{jj} = I_{k_j} + O_p(T^{-1/2}), \quad j = 1,2, \quad \tilde{V}_{12} = \begin{bmatrix} I_{k^c} & 0 \\ 0 & \Phi \end{bmatrix} + O_p(T^{-1/2}). \tag{B.12}$$

Moreover, by straightforward matrix algebra we get the next Lemma.

**LEMMA B.3.** *The matrices* $\tilde{A} = \tilde{V}_{11}^{-1}\tilde{V}_{12}$, $\tilde{B} = \tilde{V}_{22}^{-1}\tilde{V}_{21}$ *and* $\tilde{R} = \tilde{A}\tilde{B}$ *are such that:*

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

*where*

$$\tilde{A}_{cs} = \tilde{\Sigma}_{cc|1}^{-1}\tilde{\Sigma}_{c2|1}, \qquad \tilde{A}_{ss} = \tilde{\Sigma}_{11}^{-1}\left(\tilde{\Sigma}_{12} - \tilde{\Sigma}_{1c}\tilde{\Sigma}_{cc|1}^{-1}\tilde{\Sigma}_{c2|1}\right),$$

*with* $\tilde{\Sigma}_{cc|1} = \tilde{\Sigma}_{cc} - \tilde{\Sigma}_{c1}\tilde{\Sigma}_{11}^{-1}\tilde{\Sigma}_{1c}$ *and* $\tilde{\Sigma}_{c2|1} = \tilde{\Sigma}_{c2} - \tilde{\Sigma}_{c1}\tilde{\Sigma}_{11}^{-1}\tilde{\Sigma}_{12}$*, the expressions of* $\tilde{B}_{cs}$*,* $\tilde{B}_{ss}$ *are the same as those of* $\tilde{A}_{cs}$*,* $\tilde{A}_{ss}$ *after interchanging indexes* 1 *and* 2*, and*

$$\tilde{R}_{cs} = \tilde{B}_{cs} + \tilde{A}_{cs}\tilde{B}_{ss}, \qquad \tilde{R}_{ss} = \tilde{A}_{ss}\tilde{B}_{ss}.$$

***Proof****: See Appendix C.3.*

The transposed matrix $\tilde{A}'$ in block-form consists of the coefficients for the partitioned regression of $(f_t^{c\prime}, f_{2,t}^{s\prime})'$ onto $(f_t^{c\prime}, f_{1,t}^{s\prime})'$, and similarly for matrix $\tilde{B}$. From Lemma B.3 the $k^c$ largest eigenvalues of matrix $\tilde{R}$ are $\tilde{\rho}_1^2 = ... = \tilde{\rho}_{k^c}^2 = 1$, while the remaining $k_1 - k^c$ eigenvalues are the eigenvalues of matrix $\tilde{R}_{ss}$ and are such that $1 > \tilde{\rho}_{k^c+1}^2 \geq ... \geq \tilde{\rho}_{k_1}^2 > 0$, w.p.a. 1. In fact, from (B.12) we have $\tilde{R}_{ss} = \Phi\Phi' + o_p(1)$. Let us define:

$$\underset{(k_1 \times k^c)}{E_c} = \begin{bmatrix} I_{k^c} \\ 0 \end{bmatrix}, \qquad \underset{(k_1 \times (k_1 - k^c))}{E_s} = \begin{bmatrix} 0 \\ I_{k_1 - k^c} \end{bmatrix}. \tag{B.13}$$

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

### B.1.4 Eigenvalues and eigenvectors of matrix $\hat{R}$ obtained by perturbation methods

The estimators of the first $k^c$ canonical correlations are such that $\hat{\rho}_\ell^2$, with $\ell = 1, ..., k^c$ are the $k^c$ largest eigenvalues of matrix $\hat{R}$. We now derive their asymptotic expansion under the null hypothesis $H(k^c)$ using perturbations arguments applied to equation (B.8). Let $\hat{W}_1^*$ be a $(k_1, k^c)$ matrix whose columns are eigenvectors of matrix $\hat{R}$ associated with the eigenvalues $\hat{\rho}_\ell^2$, with $\ell = 1, ..., k^c$. We have:

$$\hat{R}\hat{W}_1^* \ = \ \hat{W}_1^*\hat{\Lambda}, \tag{B.14}$$

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

$$\hat{W}_1^* \ = \ E_c\hat{\mathcal{U}} + E_s\hat{\alpha}, \tag{B.15}$$
$$\hat{\Lambda} \ = \ I_{k^c} + \hat{M}, \tag{B.16}$$

where $E_c$ and $E_s$ are defined in equation (B.13), the stochastic $(k^c, k^c)$ matrix $\hat{\mathcal{U}}$ is nonsingular w.p.a. 1, stochastic matrix $\hat{M}$ is diagonal, and $\hat{\alpha}$ is a $(k_1 - k^c, k^c)$ stochastic matrix. By the continuity of the matrix eigenvalue and eigenfunction mappings, and Lemma B.1, we have that $\hat{\alpha}$ and $\hat{M}$ converge in probability to zero as $N_1, N_2, T \to \infty$ at rate $O_p(\delta_{N,T})$. By substituting the expansions (B.8) and (B.15)-(B.16) into the eigenvalue-eigenvector equation (B.14), using the characterization of matrix $\tilde{R}$ obtained in Lemma B.3, and keeping terms up to order $O_p(\delta_{N,T}^3)$, we get expressions for matrices $\hat{\alpha}$ and $\hat{M}$. These yield the asymptotic expansions of the eigenvalues and eigenvectors of matrix $\hat{R}$ provided in the next Lemma.

**LEMMA B.4.** *Under Assumptions 4.1-B.1 and A.2-A.4, A.5 b)-c) and A.6 a), we have:*

$$\begin{aligned}\hat{\Lambda} \ = \ & I_{k^c} + \hat{\mathcal{U}}^{-1}\tilde{\Sigma}_{cc}^{-1}\left\{\hat{\Psi}_{cc}^* + \hat{\Psi}_{cs}^*(I_{k_1-k^c} - \tilde{R}_{ss})^{-1}\hat{\Psi}_{sc} - \tilde{\Sigma}_{c,1}(I_{k_1-k^c} - \tilde{R}_{ss})^{-1}\hat{\Psi}_{sc}\tilde{\Sigma}_{cc}^{-1}\hat{\Psi}_{cc}^*\right\}\hat{\mathcal{U}} \\ & + O_p(\delta_{N,T}^3),\end{aligned} \tag{B.17}$$

*and:*

$$\begin{aligned}\hat{W}_1^* \ = \ & \Big(E_c + E_s(I_{k_1-k^c} - \tilde{R}_{ss})^{-1}\Big[\hat{\Psi}_{sc} + \hat{\Psi}_{ss}(I_{k_1-k^c} - \tilde{R}_{ss})^{-1}\hat{\Psi}_{sc} \\ & - (I_{k_1-k^c} - \tilde{R}_{ss})^{-1}\hat{\Psi}_{sc}\Big(\hat{\Psi}_{cc} + \tilde{R}_{cs}(I_{k_1-k^c} - \tilde{R}_{ss})^{-1}\hat{\Psi}_{sc}\Big)\Big]\Big)\hat{\mathcal{U}} + O_p(\delta_{N,T}^3),\end{aligned} \tag{B.18}$$

*where $\hat{\Psi}_{cc}$, $\hat{\Psi}_{cs} = \hat{\Psi}_{sc}'$, $\hat{\Psi}_{ss}$ denote the upper-left $(k^c, k^c)$ block, the upper-right $(k^c, k_1^s)$ block and the lower-right $(k_1^s, k_1^s)$ block of matrix $\hat{\Psi}$, and similarly for the blocks of $\hat{\Psi}^*$.*
***Proof****: See Appendix OA C.4.*

The normalized eigenvectors of $\hat{R}$ corresponding to the canonical directions are:

$$\hat{W}_1 \ = \ \hat{W}_1^* \cdot diag(\hat{W}_1^{*\prime}\hat{V}_{11}\hat{W}_1^*)^{-1/2}. \tag{B.19}$$

In equations (B.17) and (B.18), in the terms that are of second-order with respect to $\hat{\Psi}$, we can replace $\hat{\Psi}$ by $\hat{\Psi}^{(I)}$ without changing the order $O_p(\delta_{N,T}^3)$ of the remainder term. Note that the approximation in (B.17) holds for the terms in the main diagonal, as matrix $\hat{\Lambda}$ has been defined to be diagonal.

### B.1.5 Asymptotic expansion of $\sum_{\ell=1}^{k^c} \hat{\rho}_\ell$

Let us now derive an asymptotic expansion for the sum of the $k^c$ largest canonical correlations $\sum_{\ell=1}^{k^c} \hat{\rho}_\ell$. By using the expansion of the matrix square root function in a neighbourhood of the identity, i.e. $(I + X)^{1/2} =$

$I + \frac{1}{2}X - \frac{1}{8}X^2 + O_p(\delta_{N,T}^3)$ for $X = O_p(\delta_{N,T})$, from equation (B.17) we have:

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

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

$$\begin{aligned}\sum_{\ell=1}^{k^c}\hat{\rho}_\ell &= k^c + \frac{1}{2}tr\left\{\tilde{\Sigma}_{cc}^{-1}\left[\hat{\Psi}_{cc}^* - \frac{1}{4}\hat{\Psi}_{cc}^{*(I)}\tilde{\Sigma}_{cc}^{-1}\hat{\Psi}_{cc}^{*(I)} + \hat{\Psi}_{cs}^{*(I)}(I_{k_1-k^c} - \tilde{R}_{ss})^{-1}\hat{\Psi}_{sc}^{(I)}\right.\right. \\ &\quad \left.\left. -\tilde{\Sigma}_{c,1}(I_{k_1-k^c} - \tilde{R}_{ss})^{-1}\hat{\Psi}_{sc}^{(I)}\tilde{\Sigma}_{cc}^{-1}\hat{\Psi}_{cc}^{*(I)}\right]\right\} + O_p(\delta_{N,T}^3), \qquad \text{(B.20)}\end{aligned}$$

by the commutative property of the trace and including higher-order terms in $O_p(\delta_{N,T}^3)$. Let us now derive the asymptotic expansions of the terms within the trace operator.

**i) Asymptotic expansion of $\hat{\Psi}_{cc}^{*(I)}$**

From equation (B.10), we have $\hat{\Psi}_{cc}^{*(I)} = \left[-\hat{X}_{11}\tilde{R} + \hat{X}_{12}\tilde{B} - \tilde{B}'\hat{X}_{22}\tilde{B} + \tilde{B}'\hat{X}_{21}\right]^{(cc)}$. As matrices $\tilde{R}$ and $\tilde{B}$ have the same structure $[\ E_c\ \vdots\ *\ ]$ (see Lemma B.3), we have:

$$\hat{\Psi}_{cc}^{*(I)} = -\hat{X}_{11}^{(cc)} + \hat{X}_{12}^{(cc)} - \hat{X}_{22}^{(cc)} + \hat{X}_{21}^{(cc)}. \qquad \text{(B.21)}$$

From the expressions of the matrices $\hat{X}_{j,k}$ in (B.6), and using the fact that upper $k^c$-dimensional subvector of both $h_{1,t}$ and $h_{2,t}$ is $f_t^c$, the upper-left $(k^c, k^c)$ blocks of the first and second matrices in the r.h.s. vanish. Therefore, from (B.21) we get:

$$\hat{\Psi}_{cc}^{*(I)} = -\frac{1}{T}\sum_{t=1}^{T}(\psi_{1,t}^{(c)} - \psi_{2,t}^{(c)})(\psi_{1,t}^{(c)} - \psi_{2,t}^{(c)})', \qquad \text{(B.22)}$$

where $\psi_{j,t}^{(c)}$ denotes the upper $(k^c, 1)$ block of vector $\psi_{j,t}$. To compute the matrix in the r.h.s., we plug the expressions $\psi_{j,t} = \frac{1}{\sqrt{N_j}}u_{j,t} + \frac{1}{T}b_{j,t} + \frac{1}{\sqrt{N_jT}}d_{j,t} + \vartheta_{j,t}$ for $j = 1, 2$ from (B.3), and use Assumptions 4.1 and A.2-A.4, A.5 *b)-c)* and A.6 *a)* to bound negligible terms up to $o_p(\epsilon_{N,T})$, where:

$$\epsilon_{N,T} := \frac{1}{N\sqrt{T}}. \qquad \text{(B.23)}$$

**LEMMA B.5.** *Under Assumptions 4.1 and A.2-A.4, A.5 b)-c) and A.6 a) we have:*

$$\begin{aligned}
\hat{\Psi}_{cc}^{*(I)} &= -\frac{1}{N}\left(\frac{1}{T}\sum_{t=1}^{T} E[(\mu_N u_{1t}^{(c)} - u_{2t}^{(c)})(\mu_N u_{1t}^{(c)} - u_{2t}^{(c)})'|\mathcal{F}_t]\right) \\
&\quad -\frac{1}{N\sqrt{T}}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[(\mu_N u_{1t}^{(c)} - u_{2t}^{(c)})(\mu_N u_{1t}^{(c)} - u_{2t}^{(c)})' - E[(\mu_N u_{1t}^{(c)} - u_{2t}^{(c)})(\mu_N u_{1t}^{(c)} - u_{2t}^{(c)})'|\mathcal{F}_t]\right]\right) \\
&\quad -\frac{1}{T\sqrt{NT}}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[(\bar{b}_{1,t}^{(c)} - \bar{b}_{2,t}^{(c)})(\mu_N u_{1,t}^{(c)} - u_{2,t}^{(c)})' + (\mu_N u_{1,t}^{(c)} - u_{2,t}^{(c)})(\bar{b}_{1,t}^{(c)} - \bar{b}_{2,t}^{(c)})'\right]\right) \\
&\quad -\frac{1}{T^2}\left(\frac{1}{T}\sum_{t=1}^{T}(b_{1,t}^{(c)} - b_{2,t}^{(c)})(b_{1,t}^{(c)} - b_{2,t}^{(c)})'\right) \\
&\quad -\frac{1}{T\sqrt{NT}}\left(\frac{1}{T}\sum_{t=1}^{T}\left[(\bar{b}_{1,t}^{(c)} - \bar{b}_{2,t}^{(c)})(\mu_N d_{1,t}^{(c)} - d_{2,t}^{(c)})' + (\mu_N d_{1,t}^{(c)} - d_{2,t}^{(c)})(\bar{b}_{1,t}^{(c)} - \bar{b}_{2,t}^{(c)})'\right]\right) + o_p(\epsilon_{N,T}),
\end{aligned}$$

*where $\bar{b}_{j,t}$ is defined in Theorem 1. The terms in the parentheses are $O_p(1)$.*
***Proof:*** *See Appendix C.5.*

Lemma B.5 shows that the leading stochastic terms in $\hat{\Psi}_{cc}^{*(I)}$ are of order $O_p\left(\frac{1}{N}\right)$, $O_p\left(\frac{1}{N\sqrt{T}}\right)$, $O_p\left(\frac{1}{T\sqrt{NT}}\right)$ and $O_p\left(\frac{1}{T^2}\right)$.

**ii) Asymptotic expansion of the second-order terms in the r.h.s. of (B.20)**

The asymptotic expansion of the second-order term $\hat{\Psi}_{cc}^{*(II)} - \frac{1}{4}\hat{\Psi}_{cc}^{*(I)}\tilde{\Sigma}_{cc}^{-1}\hat{\Psi}_{cc}^{*(I)} + \hat{\Psi}_{cs}^{*(I)}(I_{k_1-k^c} - \tilde{R}_{ss})^{-1}\hat{\Psi}_{sc}^{(I)} - \tilde{\Sigma}_{c,1}(I_{k_1-k^c} - \tilde{R}_{ss})^{-1}\hat{\Psi}_{sc}^{(I)}\tilde{\Sigma}_{cc}^{-1}\hat{\Psi}_{cc}^{*(I)}$ is provided in the next lemma.

**LEMMA B.6.** *Under Assumptions 4.1 and A.2-A.4, A.5 b)-c) and A.6 a) we have:*

$$\begin{aligned}
&\hat{\Psi}_{cc}^{*(II)} - \frac{1}{4}\hat{\Psi}_{cc}^{*(I)}\tilde{\Sigma}_{cc}^{-1}\hat{\Psi}_{cc}^{*(I)} + \hat{\Psi}_{cs}^{*(I)}(I_{k_1-k^c} - \tilde{R}_{ss})^{-1}\hat{\Psi}_{sc}^{(I)} - \tilde{\Sigma}_{c,1}(I_{k_1-k^c} - \tilde{R}_{ss})^{-1}\hat{\Psi}_{sc}^{(I)}\tilde{\Sigma}_{cc}^{-1}\hat{\Psi}_{cc}^{*(I)} \\
= \; & \frac{1}{T^2}\left\{\left[\frac{1}{T}\sum_{t=1}^{T}\left(b_{1,t}^{(c)} - b_{2,t}^{(c)}\right)F_t'\right]\tilde{\Sigma}_F^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}F_t\left(b_{1,t}^{(c)} - b_{2,t}^{(c)}\right)'\right]\right\} \\
&+\frac{1}{T\sqrt{NT}}\left\{\left(E\left[\left(\bar{b}_{1,t}^{(c)} - \bar{b}_{2,t}^{(c)}\right)F_t'\right]\Sigma_F^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}F_t\left(\mu_N u_{1,t}^{(c)} - u_{2,t}^{(c)}\right)'\right.\right. \\
&\quad \left.\left. +E\left[\left(\bar{b}_{1,t}^{(c)} - \bar{b}_{2,t}^{(c)}\right)F_t'\right]\Sigma_F^{-1}\frac{1}{T}\sum_{t=1}^{T}F_t\left(\mu_N d_{1,t}^{(c)} - d_{2,t}^{(c)}\right)'\right)^{+}\right\} + o_p(\epsilon_{N,T}),
\end{aligned}$$

*where $\tilde{\Sigma}_F = \frac{1}{T}\sum_{t=1}^{T}F_tF_t'$ and $A^+ := A + A'$. The terms in the curly brackets are $O_p(1)$.*
***Proof:*** *See Appendix C.6.*

From Lemmas B.5 and B.6, the asymptotic expansion of the term within the square brackets in the r.h.s of

(B.20) is:

$$\begin{aligned}
&\hat{\Psi}_{cc}^{*}-\frac{1}{4}\hat{\Psi}_{cc}^{*(I)}\tilde{\Sigma}_{cc}^{-1}\hat{\Psi}_{cc}^{*(I)}+\hat{\Psi}_{cs}^{*(I)}(I_{k_1-k^c}-\tilde{R}_{ss})^{-1}\hat{\Psi}_{sc}^{(I)}-\tilde{\Sigma}_{c,1}(I_{k_1-k^c}-\tilde{R}_{ss})^{-1}\hat{\Psi}_{sc}^{(I)}\tilde{\Sigma}_{cc}^{-1}\hat{\Psi}_{cc}^{*(I)}\\
= \ & -\frac{1}{N}\left(\frac{1}{T}\sum_{t=1}^{T}E[(\mu_N u_{1t}^{(c)}-u_{2t}^{(c)})(\mu_N u_{1t}^{(c)}-u_{2t}^{(c)})'|\mathcal{F}_t]\right)-\frac{1}{T^2}\left\{\frac{1}{T}\sum_{t=1}^{T}\widetilde{\Delta b}_t\widetilde{\Delta b}_t'\right\}\\
&-\frac{1}{N\sqrt{T}}\left\{\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[(\mu_N u_{1t}^{(c)}-u_{2t}^{(c)})(\mu_N u_{1t}^{(c)}-u_{2t}^{(c)})'-E[(\mu_N u_{1t}^{(c)}-u_{2t}^{(c)})(\mu_N u_{1t}^{(c)}-u_{2t}^{(c)})'|\mathcal{F}_t]\right]\right\}\\
&-\frac{1}{T\sqrt{NT}}\left\{\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\Delta b_t(\mu_N u_{1,t}^{(c)}-u_{2,t}^{(c)})'+(\mu_N u_{1,t}^{(c)}-u_{2,t}^{(c)})\Delta b_t'\right]\right\}\\
&-\frac{1}{T\sqrt{NT}}\left\{\frac{1}{T}\sum_{t=1}^{T}\left[\Delta b_t(\mu_N d_{1,t}^{(c)}-d_{2,t}^{(c)})'+(\mu_N d_{1,t}^{(c)}-d_{2,t}^{(c)})\Delta b_t'\right]\right\}+o_p\left(\epsilon_{N,T}\right), \qquad \text{(B.24)}
\end{aligned}$$

where $\Delta b_t$ and $\widetilde{\Delta b}_t$ are the population and sample residuals defined in Theorem 1. For the fifth summation term in the r.h.s., let us now check that:

$$\frac{1}{T}\sum_{t=1}^{T}\left[\Delta b_t(\mu_N d_{1,t}^{(c)}-d_{2,t}^{(c)})'+(\mu_N d_{1,t}^{(c)}-d_{2,t}^{(c)})\Delta b_t'\right]=O_p\left(\frac{1}{\sqrt{T}}\right). \qquad \text{(B.25)}$$

Indeed, we have:

$$\frac{1}{T}\sum_{t=1}^{T}\Delta b_t h_{j,t}'=E[\Delta b_t h_{j,t}']+O_p\left(\frac{1}{\sqrt{T}}\right), \qquad \text{(B.26)}$$

from Assumption A.4 *b)* and A.6 *a)*, and Corollary 14.3 in Davidson (1994). Then, (B.25) follows from the definition of $d_{j,t}$, the convergence in (B.26), and equality $E[\Delta b_t h_{j,t}']=0$, for $j=1,2$. The latter equality holds because $\Delta b_t$ is the residual of a projection on $F_t$, and $h_{j,t}$ is spanned by $F_t$.

By plugging (B.24)-(B.25) into (B.20), and using $\frac{1}{T\sqrt{T}\sqrt{NT}}=o(\epsilon_{N,T})$ when $N\ll T^3$, and $\tilde{\Sigma}_{cc}=I_{k^c}+O_p(T^{-1/2})$ from (B.12), we get:

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

(B.27)

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

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

Process $U_t$ depends on $N_1$, $N_2$, but we do not make this dependence explicit for expository purpose. Now we use these definitions, together with the commutativity and linearity properties of the trace operator, as well as the bound $\delta_{N,T}^3 = o(\epsilon_{N,T})$, which follows from the definitions of $\delta_{N,T}$ in Lemma B.1 and of $\epsilon_{N,T}$ in (B.23) and the condition $\sqrt{T} \ll N \ll T^{5/2}$ in Assumption 4.1. Then, from equation (B.27) we get the asymptotic expansion:

$$
\begin{aligned}
&\sum_{\ell=1}^{k^c} \hat{\rho}_\ell - k^c + \frac{1}{2N} tr\left\{\tilde{\Sigma}_{cc}^{-1}\tilde{\Sigma}_U\right\} + \frac{1}{2T^2} tr\left\{\tilde{\Sigma}_{cc}^{-1}\tilde{\Sigma}_B\right\} \\
= \quad & -\frac{1}{2N\sqrt{T}}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[U_t'U_t - E(U_t'U_t|\mathcal{F}_t)\right]\right) - \frac{1}{T\sqrt{NT}}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Delta b_t'U_t\right) + o_p\left(\epsilon_{N,T}\right). \quad \text{(B.28)}
\end{aligned}
$$

Under our set of assumptions, terms $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[U_t'U_t - E(U_t'U_t|\mathcal{F}_t)\right]$ and $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Delta b_t'U_t$ are $O_p(1)$. In fact, in the next subsection we show that these terms are jointly asymptotically Gaussian distributed. The remainder term $o_p\left(\epsilon_{N,T}\right)$ in the r.h.s. of (B.28) is negligible with respect to the first term in the r.h.s.[36]

### B.1.6 Asymptotic distribution of the test statistic under the null hypothesis $H(k^c)$

From the asymptotic expansion (B.28) we obtain the asymptotic distribution of $\hat{\xi}(k^c) = \sum_{\ell=1}^{k^c}\hat{\rho}_\ell$ under the null hypothesis $H(k^c)$ of $k^c$ common factors. First, we apply a CLT for weakly dependent triangular array data to prove the asymptotic normality of $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathcal{Z}_{N,t}$ as $N, T \to \infty$, where $\mathcal{Z}_{N,t} := \begin{bmatrix} U_t'U_t - E(U_t'U_t|\mathcal{F}_t) \\ \Delta b_t'U_t \end{bmatrix}$ depends on $N_1, N_2$ via process $U_t$.

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

Let process $V_{N_1,N_2,t} \equiv V_t$ be as defined in Assumption A.6, and let $\mathcal{V}_{t-m}^{t+m} = \sigma(V_s, t-m \le s \le t+m)$ for any positive integer $m$, with $\mathcal{V}_t \equiv \mathcal{V}_{-\infty}^t$. We show that the following conditions hold true:

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

(ii) $\sup_{t\ge1,N_1,N_2\ge1} E\left[\|\mathcal{Z}_{N,t}\|^r\right] < \infty$, for a constant $r > 2$,

(iii) Process $(\mathcal{Z}_{N,t})$ is $L^2$ Near Epoch Dependent ($L^2$-NED) of size $-1$ on process $(V_t)$, and $(V_t)$ is strong mixing of size $-\frac{r}{r-2}$, uniformly in $N_1, N_2 \ge 1$,[37]

(iv) $\lim_{T,N\to\infty} V\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathcal{Z}_{N,t}\right) = \Omega_U$ is a positive definite matrix.

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

$$
\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathcal{Z}_{N,t} \xrightarrow{d} N\left(0, \Omega_U\right), \quad \text{(B.29)}
$$

as $T, N \to \infty$.

Let us check the above conditions (i)-(iv) and compute the asymptotic variance matrix $\Omega_U$. Condition (i) follows by the Law of Iterated Expectation and $E(U_t|\mathcal{F}_t) = 0$, which is implied by Assumption A.4 *a)*.

[36] If $N \gtrapprox T^{5/2}$ then $\delta_{N,T}^3$ is not negligible with respect to $\frac{1}{N\sqrt{T}}$. Similarly, if $N \lessapprox T^{1/4}$ then $\delta_{N,T}^3$ is not negligible with respect to $\frac{1}{N\sqrt{T}}$. In those cases, we need a more accurate asymptotic expansion.

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

Condition (ii) is implied by Assumptions A.3, A.4 *b)* and A.5 *b)*. The NED property in Condition (iii) holds true because conditional expectations given $\mathcal{F}_t$ can be well approximated by elements in the sigma-field $\mathcal{V}_{t-m}^{t+m}$ generated by the mixing process $(V_t)$, for large $m$, by Assumptions A.3, A.4 *b)*, A.5 *b)* and A.6 *a)-c)*, as we show in the next lemma.

**LEMMA B.7.** *Assumptions A.3, A.4 b), A.5 b) and A.6 a)-c) imply Condition (iii).*
***Proof:*** *See Appendix C.7.*

To check Condition (iv) we use:

$$\begin{aligned}\lim_{T,N\to\infty} V\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathcal{Z}_{N,t}\right) &= \lim_{T,N\to\infty}\frac{1}{T}\sum_{h=-T+1}^{T-1}(T-|h|)Cov\left(\mathcal{Z}_{N,t},\mathcal{Z}_{N,t-h}\right)\\ &= \lim_{N\to\infty}\sum_{h=-\infty}^{\infty}Cov\left(\mathcal{Z}_{N,t},\mathcal{Z}_{N,t-h}\right),\end{aligned}$$

where the first equality follows from stationarity of the data. The series converges because the zero-mean process $\mathcal{Z}_{N,t}$ is a $L^2$-mixingale with size $-1$, [38] by Theorem 17.5 in Davidson (1994) and Conditions (ii)-(iii), which implies $\|Cov\left(\mathcal{Z}_{N,t},\mathcal{Z}_{N,t-h}\right)\| = \left\|E\left[E(\mathcal{Z}_{N,t}|\mathcal{V}_{t-h})\mathcal{Z}_{N,t-h}'\right]\right\| \leq \|E(\mathcal{Z}_{N,t}|\mathcal{V}_{t-h})\|_2\|\mathcal{Z}_{N,t-h}\|_2 = O\left(h^{-\psi}\right)$, uniformly in $N_1, N_2 \geq 1$, for some $\psi > 1$. The latter uniform bound also allows for an application of the Lebesgue Lemma to get:

$$\Omega_U = \lim_{T,N\to\infty} V\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathcal{Z}_{N,t}\right) = \sum_{h=-\infty}^{\infty}\Gamma(h), \qquad \Gamma(h) := \lim_{N\to\infty}Cov\left(\mathcal{Z}_{N,t},\mathcal{Z}_{N,t-h}\right),$$

where the limit in the definition of $\Gamma(h)$ exists as we now show. Indeed, by the Law of Iterated Expectation and $E[\mathcal{Z}_{N,t}|\mathcal{F}_t] = 0$, we have:

$$\Gamma(h) = \lim_{N\to\infty} E\left[Cov\left(\mathcal{Z}_{N,t},\mathcal{Z}_{N,t-h}|\mathcal{F}_t\right)\right]. \tag{B.30}$$

Moreover, from Assumptions A.3 and A.5 *a)*, vector $(U_t', U_{t-h}')'$ is asymptotically Gaussian for any $h$, $t$ as $N \to \infty$:

$$\begin{pmatrix} U_t \\ U_{t-h}\end{pmatrix} \overset{d}{\to} \begin{pmatrix} U_t^\infty \\ U_{t-h}^\infty\end{pmatrix} \sim N\left(0, \begin{bmatrix}\Sigma_{U,t}(0) & \Sigma_{U,t}(h) \\ \Sigma_{U,t}(h)' & \Sigma_{U,t}(0)\end{bmatrix}\right), \quad (\mathcal{F}_t\text{-stably}). \tag{B.31}$$

Now, we use the Lebesgue Lemma to interchange the limes for $N \to \infty$ and the outer expectation in the r.h.s. of (B.30), and we use the fact that convergence in distribution plus uniform integrability imply convergence of the expectation for a sequence of random variables (see Theorem 25.12 in Billingsley (1995)) to show the next lemma.

**LEMMA B.8.** *Under Assumptions A.3 and A.5 b), we have:*

$$\Gamma(h) = E\left[\begin{pmatrix} Cov(U_t^{\infty\,\prime}U_t^\infty, U_{t-h}^{\infty\,\prime}U_{t-h}^\infty|\mathcal{F}_t) & Cov(U_t^{\infty\,\prime}U_t^\infty, \Delta b_{t-h}'U_{t-h}^\infty|\mathcal{F}_t) \\ Cov(\Delta b_t'U_t^\infty, U_{t-h}^{\infty\,\prime}U_{t-h}^\infty|\mathcal{F}_t) & Cov(\Delta b_t'U_t^\infty, \Delta b_{t-h}'U_{t-h}^\infty|\mathcal{F}_t)\end{pmatrix}\right].$$

***Proof:*** *See Appendix C.8.*

Lemma B.8 allows to deploy the joint asymptotic Gaussian distribution of $(U_t^{\infty\,\prime}, U_{t-h}^{\infty\,\prime})'$ to compute the limit autocovariance $\Gamma(h)$. Further, by using that $\Delta b_t$ is measurable w.r.t. $\mathcal{F}_t$, we have $Cov(U_t^{\infty\,\prime}U_t^\infty, \Delta b_{t-h}'U_{t-h}^\infty|\mathcal{F}_t) =$

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

0 and $Cov(\Delta b_t' U_t^\infty, \Delta b_{t-h}' U_{t-h}^\infty | \mathcal{F}_t) = \Delta b_t' \Sigma_{U,t}(h) \Delta b_{t-h} = tr\left\{\Sigma_{U,t}(h)(\Delta b_t \Delta b_{t-h}')'\right\}$. Then, we get:

$$\Gamma(h) = \begin{pmatrix} E[Cov(U_t^{\infty\,\prime} U_t^\infty, U_{t-h}^{\infty\,\prime} U_{t-h}^\infty | \mathcal{F}_t)] & 0 \\ 0 & tr\left\{E\left[\Sigma_{U,t}(h)(\Delta b_t \Delta b_{t-h}')'\right]\right\} \end{pmatrix}.$$

To compute the upper-left block, we use the following lemma.

**LEMMA B.9.** *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 $Cov(x'Ax, y'By) = 2tr\left\{A\Omega_{xy}B\Omega_{xy}'\right\}$.*
***Proof:*** *See Theorem 12 p. 284 in Magnus and Neudecker (2007) and Theorem 10.21 in Schott (2005).*

From Lemma B.9 and the Gaussian distribution in (B.31) we get $Cov(U_t^{\infty\,\prime} U_t^\infty, U_{t-h}^{\infty\,\prime} U_{t-h}^\infty | \mathcal{F}_t) = 2tr\left\{\Sigma_{U,t}(h)\Sigma_{U,t}(h)'\right\}$. Therefore we get:

$$\Omega_U = \sum_{h=-\infty}^{\infty} \begin{bmatrix} 2tr\left\{E\left[\Sigma_{U,t}(h)\Sigma_{U,t}(h)'\right]\right\} & 0 \\ 0 & tr\left\{E\left[\Sigma_{U,t}(h)(\Delta b_t \Delta b_{t-h}')'\right]\right\} \end{bmatrix} = \begin{bmatrix} 4\Omega_{U,1} & 0 \\ 0 & \Omega_{U,2} \end{bmatrix}. \quad \text{(B.32)}$$

**ii) Asymptotic Gaussian distribution of the test statistic**

Let us define vector $D_{N,T} = \left[\frac{1}{2N\sqrt{T}}, \frac{1}{T\sqrt{NT}}\right]'$. From equations (B.28) and (B.32), and by using:

$$(D_{N,T}'\Omega_U D_{N,T})^{1/2} = \left(\frac{1}{(N\sqrt{T})^2}\Omega_{U,1} + \frac{1}{(T\sqrt{NT})^2}\Omega_{U,2}\right)^{1/2} = \frac{1}{N\sqrt{T}}\left(\Omega_{U,1} + \frac{N}{T^2}\Omega_{U,2}\right)^{1/2},$$

and $N\sqrt{T}\left(\Omega_{U,1} + \frac{N}{T^2}\Omega_{U,2}\right)^{-1/2} = O\left(\min\{N\sqrt{T}, T\sqrt{NT}\}\right) = O(\epsilon_{N,T}^{-1})$, under the hypothesis of $k^c$ common factors in each group the statistics $\hat{\xi}(k^c) = \sum_{\ell=1}^{k^c} \hat{\rho}_\ell$ is such that:

$$\begin{aligned} & N\sqrt{T}\left(\Omega_{U,1} + \frac{N}{T^2}\Omega_{U,2}\right)^{-1/2}\left[\hat{\xi}(k^c) - k^c + \frac{1}{2N}tr\left\{\tilde{\Sigma}_{cc}^{-1}\tilde{\Sigma}_U\right\} + \frac{1}{2T^2}tr\left\{\tilde{\Sigma}_{cc}^{-1}\tilde{\Sigma}_B\right\}\right] \\ = & -(D_{N,T}'\Omega_U D_{N,T})^{-1/2} D_{N,T}' \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \mathcal{Z}_{N,T} + o_p(1). \end{aligned}$$

From equation (B.29), the r.h.s. converges in distribution to a standard normal distribution, which yields Theorem 1. Note that this asymptotic distribution holds for any value of $\lambda \equiv \lim N/T^2 \in [0,\infty]$, and independently on whether $\Omega_{U,2} > 0$ or $\Omega_{U,2} = 0$, because the diverging factors in the numerator and the denominator of $(D_{N,T}'\Omega_U D_{N,T})^{-1/2} D_{N,T}' \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \mathcal{Z}_{N,T}$ cancel.

## B.2 Proof of Theorem 2

To establish the asymptotic distribution of the feasible statistic in Theorem 2 we need to control the effect of replacing the re-centering and scaling terms by means of their estimates. The latter involve factors and loadings estimates. Hence, in the OA we derive uniform asymptotic expansions of factors and loadings estimators. These results are instrumental for the next steps of the proof of Theorem 2, as well as for the proofs of other results in this paper. Then, in Subsection B.2.1 and B.2.2 we show the statements in Part i) and in Part ii) of Theorem 2, respectively.

### B.2.1 Proof of Theorem 2, Part (i)

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

$$N\sqrt{T}\Omega_{U,1}^{-1/2}\left[\hat{\xi}(k^c) - k^c + \frac{1}{2N}tr\left\{\tilde{\Sigma}_{cc}^{-1}\tilde{\Sigma}_U\right\}\right] \xrightarrow{d} N(0,1), \tag{B.33}$$

where $\Omega_{U,1} = \frac{1}{2}tr\left\{\Sigma_U(0)^2\right\}$ and we use (4.5) and $\tilde{\Sigma}_B = 0$. Theorem 2 i) follows, if we prove:

$$tr\left\{\hat{\Sigma}_U\right\} = tr\left\{\tilde{\Sigma}_{cc}^{-1}\tilde{\Sigma}_U\right\} + o_p\left(\frac{1}{\sqrt{T}}\right), \tag{B.34}$$

$$tr\left\{\hat{\Sigma}_U^2\right\} = tr\left\{\Sigma_U(0)^2\right\} + o_p(1). \tag{B.35}$$

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

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

where the ratio $\frac{1}{2}tr\left\{\hat{\Sigma}_U^2\right\}\Big/\Omega_{U,1}$ converges in probability to 1 from (B.35), the term within the curly brackets in the first line in the r.h.s. converges in distribution to a standard normal distribution from (B.33), and the term on the second line on the r.h.s. is $o_p(1)$ from (B.34).

Le us now prove equations (B.34) and (B.35) by deriving the asymptotic expansions of $\tilde{\Sigma}_{cc}^{-1}$ and $\hat{\Sigma}_U$. Lemma X.2 ii) implies:

$$\tilde{\Sigma}_{cc}^{-1} = \left(\hat{\mathcal{H}}_c^{-1}\right)'\hat{\mathcal{H}}_c^{-1} + o_p\left(\frac{1}{\sqrt{T}}\right). \tag{B.36}$$

To derive the asymptotic expansion of $\hat{\Sigma}_U$, we use its definition $\hat{\Sigma}_U = \mu_N^2\hat{\Sigma}_{u,11}^{(cc)} + \hat{\Sigma}_{u,22}^{(cc)}$, where the matrices $\hat{\Sigma}_{u,jj}$, $j = 1, 2$, defined in equation (**??**) involve the estimated loadings and residuals. We plug in the uniform asymptotic expansions from Lemma X.2 ii) to show the next result.

**LEMMA B.10.** *Under Assumptions 4.1-B.1 and A.2 - A.8:*
*i) The asymptotic expansion of estimator $\hat{\Lambda}_j'\hat{\Lambda}_j/N_j$ is:*

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

*for $j = 1, 2$, where $\tilde{\Sigma}_{\Lambda,j} = \frac{1}{N_j}\Lambda_j'\Lambda_j$ and $L_{\Lambda,j} = \tilde{\Sigma}_{\Lambda,j}Q_j$ and:*

$$\hat{\mathcal{U}}_j = \begin{bmatrix} \hat{\mathcal{H}}_c & 0 \\ 0 & \hat{\mathcal{H}}_{s,j} \end{bmatrix}, \qquad Q_j = \begin{bmatrix} 0 & 0 \\ \sqrt{T}\tilde{\Sigma}_{j,c}\tilde{\Sigma}_{cc}^{-1} & 0 \end{bmatrix}. \tag{B.38}$$

*ii) The asymptotic expansion of estimator $\hat{\Lambda}_j'\hat{\Gamma}_j\hat{\Lambda}_j/N_j$ is:*

$$\frac{1}{N_j}\hat{\Lambda}_j'\hat{\Gamma}_j\hat{\Lambda}_j = \hat{\mathcal{U}}_j'\left[\tilde{\Omega}_{jj} + \frac{1}{\sqrt{T}}\left(L_{\Omega,j} + L_{\Omega,j}'\right)\right]\hat{\mathcal{U}}_j + o_p\left(\frac{1}{\sqrt{T}}\right), \tag{B.39}$$

*for $j = 1, 2$, where $\tilde{\Omega}_{jj} = \frac{1}{N_j}\Lambda_j'\Gamma_j\Lambda_j$ and $L_{\Omega,j} = \tilde{\Omega}_{jj}Q_j$.*
***Proof:*** *See Appendix C.10.*

Equation (B.37) allows to compute the asymptotic approximation of $\left(\frac{1}{N_j}\hat{\Lambda}_j'\hat{\Lambda}_j\right)^{-1}$:

$$\left(\frac{1}{N_j}\hat{\Lambda}_j'\hat{\Lambda}_j\right)^{-1} = \hat{\mathcal{U}}_j^{-1}\left[\tilde{\Sigma}_{\Lambda,j}^{-1} - \frac{1}{\sqrt{T}}\tilde{\Sigma}_{\Lambda,j}^{-1}\left(L_{\Lambda,j} + L_{\Lambda,j}'\right)\tilde{\Sigma}_{\Lambda,j}^{-1}\right]\left(\hat{\mathcal{U}}_j'\right)^{-1} + o_p\left(\frac{1}{\sqrt{T}}\right). \tag{B.40}$$

Substituting equations (B.40) and (B.39) into equation (**??**), we get:

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

Therefore, from the definitions of matrices $L_{\Omega,j}$ and $L_{\Lambda,j}$ in Lemma B.10, we have:

$$\hat{\Sigma}_{u,jj} = \hat{\mathcal{U}}_j^{-1}\left(\tilde{\Sigma}_{u,jj} + \frac{1}{\sqrt{T}}(L_{U,j} + L_{U,j}')\right)\left(\hat{\mathcal{U}}_j'\right)^{-1} + o_p\left(\frac{1}{\sqrt{T}}\right), \tag{B.41}$$

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

From equation (B.41) we get the asymptotic expansion for $\hat{\Sigma}_U$:

$$\begin{aligned}
\hat{\Sigma}_U &= \mu_N^2\hat{\Sigma}_{u,11}^{(cc)} + \hat{\Sigma}_{u,22}^{(cc)} \\
&= \hat{\mathcal{H}}_c^{-1}\left(\left[\mu_N^2\tilde{\Sigma}_{u,11} + \tilde{\Sigma}_{u,22}\right]^{(cc)} + \frac{1}{\sqrt{T}}\left[\mu_N^2(L_{U,1} + L_{U,1}') + L_{U,2} + L_{U,2}'\right]^{(cc)}\right)\left(\hat{\mathcal{H}}_c'\right)^{-1} + o_p\left(\frac{1}{\sqrt{T}}\right) \\
&= \hat{\mathcal{H}}_c^{-1}\tilde{\Sigma}_U\left(\hat{\mathcal{H}}_c'\right)^{-1} + o_p\left(\frac{1}{\sqrt{T}}\right).
\end{aligned} \tag{B.42}$$

The asymptotic expansions (B.36) and (B.42), together with the commutative property of the trace operator, imply equation (B.34). Similarly, the asymptotic expansion (B.42) and the convergence $\tilde{\Sigma}_U \to \Sigma_U(0)$ imply equation (B.35).

### B.2.2 Proof of Theorem 2, 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:

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

We rely on the following Lemma. For its proof we assume that $\hat{f}_t^c$ is used to estimate the common factor in panel $j = 1$, while estimator $\hat{f}_t^{c*}$ is used in panel $j = 2$.

**LEMMA B.11.** *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 Appendix C.11.*

From Lemma B.11 and using $\sum_{\ell=1}^{k^c}\hat{\rho}_\ell = \sum_{\ell=1}^{k^c}\rho_\ell + o_p(1)$,where the $o_p(1)$ term follows from the continuity of the eigenvalues mapping, we get:

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

Under $H(r)$, we have $r < k^c$ canonical correlations that are equal to $1$, while the other ones are strictly smaller than $1$. therefore, $\sum_{\ell=1}^{k^c}\rho_\ell - k^c < 0$. Then, from Lemma B.11 we get:

$$\tilde{\xi}(k^c) \leq -N\sqrt{T}c_1, \quad w.p.a.\ 1, \tag{B.43}$$

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

*Q.E.D.*