# Modelling Large Dimensional Datasets with Markov Switching Factor Models

Matteo Barigozzi[1] Daniele Massacci[2]

October 19, 2022

**Abstract**

We study a novel large dimensional approximate factor model with regime changes in the loadings driven by a latent first order Markov process. By exploiting the equivalent linear representation of the model we first recover the latent factors by means of Principal Component Analysis. We then cast the model in state-space form, and we estimate loadings and transition probabilities through an EM algorithm based on a modified version of the Baum-Lindgren-Hamilton-Kim filter and smoother which makes use of the factors previously estimated. An important feature of our approach is that it provides closed form expressions for all estimators. We derive the theoretical properties of the proposed estimation procedure and show their good finite sample performance through a comprehensive set of Monte Carlo experiments. An important feature of our methodology is that it does not require knowledge of the true number of factors. The empirical usefulness of our approach is illustrated through an application to a large portfolio of stocks.

**Keywords**: Large Factor Model, Markov Switching, Baum-Lindgren-Hamilton-Kim Filter and Smoother, Principal Component Analysis, Stock Returns.

**JEL Codes**: C34, C38, C55, G10.

## 1 Introduction

This paper develops a comprehensive approach for the analysis of large dimensional models exhibiting an approximate factor structure, in which the loadings are subject to regime shifts driven by a first order latent Markov process. We label these large dimensional Markov Switching factor models.

[1]Università di Bologna, Department of Economics, matteo.barigozzi@unibo.it.

[2]King's College London, King's Business School and University of Naples Federico II, Department of Economics and Statistics daniele.massacci@kcl.ac.uk.

Since the work of Hamilton (1989), which was inspired by the seminal contribution of Goldfeld and Quandt (1973), Markov switching models have been widely used in the empirical analysis of macroeconomic and financial time series data: Hamilton (2016) gives an overview from a macroeconomic perspective; Guidolin (2011), and Ang and Timmermann (2012), provide a comprehensive survey in relation to financial markets; see also Qu and Zhuo (2021) and references therein for more recent advances. However, to the very best of our knowledge, the existing literature has focused on small dimensional Markov switching models, which are not applicable to high dimensional cross-sections of data. We aim at filling a gap in the literature by studying Markov switching models as applied to large panels.

There now exists strong empirical evidence that macroecononomic and financial variables exhibit an approximate factor structure, as stressed in Giannone et al. (2021). This nature of the data naturally leads to approximate latent factor specifications as a tool to model time series comovement in large dimensional cross-sections. Following the seminal contribution of Chamberlain and Rothschild (1983), static approximate factor representations have been considered, for example, in Connor and Korajczyk (1986) to develop measures of portfolio performance, and in Stock and Watson (2002a,b) to forecast large macroeconomic panels and to build indexes of macroeconomic activity. The full inferential theory is developed by Bai (2003). Settings allowing for dynamic factor representations have been also extensively studied: see Forni et al. (2017) and references therein. A broad overview of large factor models is provided in Stock and Watson (2016). To the very best of our knowledge, the vast majority of existing contributions has looked at the linear setting. However, this may not be flexible enough to accommodate the discrete regimes typically observed in macroeconomic and financial series.

A number of contributions have extended linear static factor models to allow for discrete shifts in the loadings by assuming that these shifts are driven by an *observable* state variable. A first and growing stream of literature assumes that this state variable is a deterministic time index, which leads to a factor model with structural instability in the loadings: see Breitung and Eickmeier (2011), Corradi and Swanson (2014), Baltagi et al. (2016), Cheng et al. (2016), Barigozzi et al. (2018), Barigozzi and Trapani (2020), Duan et al. (2022), among others, and Bai and Han (2016) for a survey of the literature. The presence of structural breaks implies that regime changes are not recurrent and are related to events such as technological changes or shifts in monetary policy regimes. Alternatively, the states could be driven by the realisation of an observable stationary variable with respect to a reference value, in which case a threshold factor model would arise: see Massacci (2017). Under this set up, regimes are recurrent and associated to cyclical events such as business and financial cycles. Finally, smoothly varying loadings are considered in Motta et al. (2011) and Pelger and Xiong (2022).

In this paper, we are interested in large dimensional factor models in relation to recurrent regime changes. A major drawback of threshold factor models is that they require *a priori* identification of the state variable. This may lead to model misspecification and unreliable empirical findings should the wrong state variable be employed to identify the regimes. In order to overcome this problem, we resort to the two-state Markov switching model of Goldfeld and Quandt (1973) with a *latent* state variable and extend it to allow for an underlying large dimensional factor structure. Within this setting, we make the following major methodological contributions: we propose an algorithm to estimate the conditional state probabilities, as well as the loadings and the factors; we derive the asymptotic properties of the estimators for loadings and factors. Remarkably, our results do not require knowledge of the true number of factors in any regime and it is robust to the number of factors being unknown and estimated. This is an important aspect of our paper. Estimating the number of factors is challenging in a linear setting, as evidenced by the high number of relevant contributions: Bai and Ng (2002), Alessi et al. (2010) and Ahn and Horenstein (2013) develop model selection criteria; Onatski (2010) and Trapani (2018) propose inferential procedures. Dealing with an unknown number of factors clearly becomes even more engaging in the presence of regimes driven by a latent state variable and it therefore is an important aspect of our paper.

To the very best of our knowledge, the literature on large dimensional Markov Switching factor models is still in its infancy. However, two existing contributions are important to discuss. First, Liu and Chen (2016) consider a model similar to ours, but their definition of common factors differs in that they consider factors pervasive along the time dimension rather than the cross-sectional dimension. As a consequence their idiosyncratic components are assumed to be white noise. Second, Urga and Wang (2022) consider a set up similar to ours, but assume *a priori* knowledge of the number of factors and consider a model with serially homoskedastic idiosyncratic components. Their Maximum Likelihood estimation approach adapts the EM algorithm by Rubin and Thayer (1982) and Bai and Li (2012) to the case of Gaussian mixtures where the weights are given by the probability of the latent variables to be in a given regime. Differently to our approach, Urga and Wang (2022) do not have closed form solutions for the estimated parameters.

More in detail our approach is the following. We introduce an algorithm to estimate factors, loadings, and transition probabilities of the model which is an extension to the high dimensional factor model setting of the state-space approach advanced in Hamilton (1989) and Kim (1994) to handle low dimensional Markov switching autoregressive models. Specifically, ours is a generalization of the Baum-Lindgren-Hamilton-Kim filter and smoother, the original version of which was proposed to estimate Markov-switching VAR models—see, e.g., the reviews by Guidolin (2011), Krolzig (2013), Hamilton (2016), and Guidolin and Pedio (2018). An important feature of our approach is that it provides closed form expressions for all estimators.

Most importantly, we achieve our goal by exploiting the well known property that a factor model with neglected discrete regime changes admits an equivalent representation with a higher number of factors: see, e.g., the discussions in Breitung and Eickmeier (2011), Barigozzi et al. (2018), and Duan et al. (2022), in the case of structural breaks. We use this property to estimate the latent factors by means of Principal Component Analysis (PCA) applied to the linear representation. We then input these estimated factors into our algorithm, which allows us to recover the loadings and the transition probabilities. We then derive the asymptotic properties of the estimator for the loadings: we prove the asymptotic normality; we characterise the bias, which is induced both by the well known rotational indeterminacy problem, and by the incomplete information related to the underlying data generated process. We also study the asymptotic properties of the factors which are estimated by projecting the data onto the estimated loadings. We corroborate our theoretical results through a comprehensive set of Monte Carlo experiments, which confirm the good finite sample properties of the estimation procedure we propose.

Finally, we assess the empirical validity of our model through an application to a large set of U.S. stock returns. Markov switching models have been widely used to capture the cyclical behaviour of small-dimensional portfolios of financial assets: see Guidolin (2011) and Ang and Timmermann (2012), and references therein. We contribute to this literature by applying the Markov switching factor model to a large dimensional portfolio of financial assets. Our results show that the regimes described by the model closely follow U.S. business cycle dynamics. In addition, an inspection of the estimated loadings allows us to identify level and slope factors. Therefore, our model could be employed to explain cross-sectional differences in average returns, and to then run conditional asset pricing tests when the regimes are driven by a latent first order Markov process. This would complement the findings in Massacci et al. (2021), who identify the regimes based on the return from the underlying stock market.

The rest of the paper is organised as follows. Section 2 introduces the two-state model. Section 3 describes the estimation algorithm. Section 4 derives the asymptotic theory. Section 5 briefly discusses the issue of unobserved heterogeneity. Section 6 runs a comprehensive set of Monte Carlo experiments. Section 7 employs applies our model to large sets of macroeconomics and financials variables. Finally, Section 8 concludes. Mathematical derivations are collected in Appendix A.

## Notation

We denote as $\otimes$ the Kronecker product, with $\odot$ the Hadamard (element-wise) product, and with $\oslash$ the element-wise ratio. For a vector $\boldsymbol{v} = (v_1 \cdots v_m)'$ we denote its Euclidean norm as $\|\boldsymbol{v}\| = \sqrt{\sum_{i=1}^m v_i^2}$. For a matrix $\mathbf{C}$ we denote the spectral norm as $\|\mathbf{C}\| = \sqrt{\mu_1(\mathbf{CC}')}$, where $\mu_1(\mathbf{CC}')$ indicates the largest eigenvalue of $\mathbf{CC}'$. If $\text{rk}(\mathbf{C}) = r < \infty$, then, we sometimes use the same notation $\|\mathbf{C}\|$ to denote also the Frobenius norm $\|\mathbf{C}\|_F = \sqrt{\text{tr}(\mathbf{CC}')}$. Indeed, $\|\mathbf{C}\|_F \leq \sqrt{r}\|\mathbf{C}\|$ and since it is always true that $\|\mathbf{C}\| \leq \|\mathbf{C}\|_F$, then, bounding the Frobenius or the spectral norm is asymptotically equivalent.

For a scalar discrete random variable $Z$, the notation $\mathsf{P}(Z = z)$ is its probability mass function computed using the true value of the parameters. For random variables $\mathbf{Y}$ and $\mathbf{W}$ the notations $\mathsf{E}[\mathbf{Y}]$ and $\mathsf{E}[\mathbf{Y}|\mathbf{W}]$ are the expectation and conditional expectation given $\mathbf{W}$, respectively, computed with respect to the true distributions $F_Y(\mathbf{y})$ and $F_{Y|W}(\mathbf{y}|\mathbf{W})$ which in turn are computed using the true value of the parameters. If, in place of the true value of the parameters, we use an estimate of the parameters, say $\widehat{\theta}$, then we adopt the notations $\mathsf{P}_{\widehat{\theta}}(\mathbf{Z} = z)$, $\mathsf{E}_{\widehat{\theta}}[\mathbf{Y}]$, and $\mathsf{E}_{\widehat{\theta}}[\mathbf{Y}|\mathbf{W}]$, respectively.

Finally, we indicate with $\mathbf{I}_m$ the identity matrix of dimension $m$, with $\boldsymbol{\iota}_m$ an $m$-dimensional vector of ones, and with $\mathbf{0}$ any matrix or vector of zeros whose dimensions depend on the context.

# 2 Markov switching factor model

## 2.1 Setup

We study a two-state large dimensional Markov switching factor model. Formally, we consider

$$\mathbf{x}_t = \boldsymbol{\Lambda}_1\mathbf{f}_{1t}\mathbb{I}(s_t = 1) + \boldsymbol{\Lambda}_2\mathbf{f}_{2t}\mathbb{I}(s_t = 2) + \mathbf{e}_t, \qquad t \in \mathbb{Z}, \tag{1}$$

$$\mathbf{e}_t = \boldsymbol{\Sigma}_{e1}^{1/2}\mathbb{I}(s_t = 1)\boldsymbol{\nu}_t + \boldsymbol{\Sigma}_{e2}^{1/2}\mathbb{I}(s_t = 2)\boldsymbol{\nu}_t. \tag{2}$$

We assume the elements of the $N \times 1$ vector process observable dependent variables $\{\mathbf{x}_t\}$ to have zero mean, and we consider the more general case in which they are allowed to have mean different from zero in Section 5; $\{\mathbf{f}_{jt}\}$ is the $r_j \times 1$ vector process of latent factors such that $r_j$ is fixed and $r_j \ll N$, for $j = 1, 2$; $\boldsymbol{\Lambda}_j$ is the $N \times r_j$ matrix of factor loadings with rows $\boldsymbol{\lambda}'_{ji}$, for $i = 1, \dots, N$ and $j = 1, 2$; $\{\mathbf{e}_t\}$ is the $N \times 1$ vector process of idiosyncratic components with innovations $\boldsymbol{\nu}_t \sim (\mathbf{0}, \mathbf{I}_N)$. Note that we allow the elements of $\{\mathbf{e}_t\}$ to be both serially and cross-sectionally correlated and we refer to Section 4 for the specific assumptions.

As it is standard in the literature, we assume that $s_t$ follows a discrete-state, homogeneous, irreducible and ergodic, first-order Markov chain such that

$$\mathsf{P}\left(s_{t+1} = j\,|s_t = i\right) = p_{ij}, \quad i, j = 1, 2, \quad \sum_{j=1}^{2} p_{ij} = 1,$$

with matrix of transition probabilities

$$\mathbf{P} = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix} = \begin{pmatrix} p_{11} & 1 - p_{11} \\ 1 - p_{22} & p_{22} \end{pmatrix}. \tag{3}$$

Defining the $2 \times 1$ vector of state indicators

$$\boldsymbol{\xi}_t = \begin{bmatrix} \mathbb{I}(s_t = 1) \\ \mathbb{I}(s_t = 2) \end{bmatrix}, \qquad t \in \mathbb{Z}, \tag{4}$$

allows to write the transition equation

$$\boldsymbol{\xi}_t = \mathbf{P}'\boldsymbol{\xi}_{t-1} + \mathbf{v}_t, \qquad t \in \mathbb{Z}, \tag{5}$$

where $\{\mathbf{v}_t\}$ is a discrete-valued zero mean martingale difference sequence whose elements sum to zero. Because, $\|\mathbf{P}\| < 1$, $\{s_t\}$ follows an ergodic Markov chain, thus, there exists a stationary vector of probabilities $\bar{\boldsymbol{\xi}}$ satisfying:

$$\bar{\boldsymbol{\xi}} = \mathbf{P}'\bar{\boldsymbol{\xi}}.$$

Hence, the elements of $\bar{\boldsymbol{\xi}}$ are long-run or unconditional state probabilities. In particular, we have $\bar{\boldsymbol{\xi}} = \mathsf{E}[\boldsymbol{\xi}_t]$, such that

$$\mathsf{E}[\boldsymbol{\xi}_t] = \mathsf{E}\begin{bmatrix} \mathbb{I}(s_t = 1) \\ \mathbb{I}(s_t = 2) \end{bmatrix} = \begin{bmatrix} \mathsf{P}(s_t = 1) \\ \mathsf{P}(s_t = 2) \end{bmatrix}, \tag{6}$$

where $0 < \mathsf{P}(s_t = j) < 1$, for $j = 1, 2$, by Assumption 1 in Section 4 below, which makes the Markov chain irreducible. In particular, (3) and (6) are related by (see, e.g., Guidolin and Pedio, 2018, Chapter 9)

$$\mathsf{P}(s_t = 1) = \frac{1 - p_{22}}{2 - p_{11} - p_{22}}, \quad \mathsf{P}(s_t = 2) = \frac{1 - p_{11}}{2 - p_{11} - p_{22}}. \tag{7}$$

## 2.2 State space representation

Let the $(r_1 + r_2) \times 1$ vector process $\{\mathbf{g}_t\}$ be defined as

$$\mathbf{g}_t = \begin{bmatrix} \mathbf{f}_{1t} \\ \mathbf{0} \end{bmatrix} \mathbb{I}(s_t = 1) + \begin{bmatrix} \mathbf{0} \\ \mathbf{f}_{2t} \end{bmatrix} \mathbb{I}(s_t = 2) = \begin{bmatrix} \mathbf{f}_{1t} \\ \mathbf{f}_{2t} \end{bmatrix} \odot \boldsymbol{\xi}_t, \qquad t \in \mathbb{Z}. \tag{8}$$

Let $\mathbf{B}_1 = [\boldsymbol{\Lambda}_1 \ \mathbf{0}]$ and $\mathbf{B}_2 = [\mathbf{0} \ \boldsymbol{\Lambda}_2]$, where $\mathbf{B}_1$ and $\mathbf{B}_2$ are $N \times (r_1 + r_2)$ matrices. The model in (1), (2) and (5) admits the equivalent state space representation:[1]

$$\begin{aligned} \mathbf{x}_t &= (\mathbf{B}_1 \ \mathbf{B}_2)\,(\boldsymbol{\xi}_t \otimes \mathbf{g}_t) + \left(\boldsymbol{\Sigma}_{e1}^{1/2} \ \boldsymbol{\Sigma}_{e2}^{1/2}\right)(\boldsymbol{\xi}_t \otimes \mathbf{I}_N)\,\mathbf{e}_t, \qquad t \in \mathbb{Z}, \\ \boldsymbol{\xi}_t &= \mathbf{P}'\boldsymbol{\xi}_{t-1} + \mathbf{v}_t. \end{aligned} \tag{9}$$

[1] Note that $\boldsymbol{\xi}_t \otimes \mathbf{g}_t = [\mathbf{f}_{1t}' \ \mathbf{0} \ \mathbf{f}_{2t}' \ \mathbf{0}]'$.

Under standard assumptions, the term $(\mathbf{B}_1\ \mathbf{B}_2)(\boldsymbol{\xi}_t \otimes \mathbf{g}_t)$ is identifiable up to a relabelling of the states. Also note that identification of $\mathbf{B}_1$ and $\mathbf{B}_2$, and therefore also of the elements of $\{\mathbf{g}_t\}$, is possible only up to an invertible transformation.

### 2.3 Linear representation

Model (9) is observationally equivalent to a model with one change point affecting the loadings of all units (Barigozzi et al., 2018). As a result, it can be rewritten as the $r_1 + r_2$ linear factor model:

$$\mathbf{x}_t = \mathbf{A}\mathbf{g}_t + \mathbf{e}_t, \qquad t \in \mathbb{Z}, \tag{10}$$

where $\mathbf{A} = [\boldsymbol{\Lambda}_1\ \boldsymbol{\Lambda}_2]$. Then, $\mathbf{A}$ and $\mathbf{g}_t$ may be estimated by standard Principal Component Analysis (PCA) (Stock and Watson, 2002a,b; Bai, 2003). Now, since PCA gives consistent estimators of the factors as $N, T \to \infty$, hereafter, we first consider estimation of model (9) in the case in which $\mathbf{g}_t$ is known. Then, we briefly review the implementation of PCA in Section 3.2.

### 2.4 Log-likelihood

The parameters of interest are partitioned as

$$\boldsymbol{\varphi} = \left[\text{vec}(\mathbf{B}_1)', \text{vec}(\mathbf{B}_2)', \text{vech}(\text{diag}(\boldsymbol{\Sigma}_{e1}))', \text{vec}(\text{diag}(\boldsymbol{\Sigma}_{e2}))'\right]', \quad \boldsymbol{\rho} = \text{vec}(\mathbf{P}),$$

so that the vector of parameters of interest, denoted as $\mathbf{q}$, is defined as

$$\mathbf{q} = \left[\boldsymbol{\varphi}', \boldsymbol{\rho}'\right]'.$$

Let $\mathbf{X} = (\mathbf{x}_1', \dots, \mathbf{x}_T')'$, $\boldsymbol{\mathcal{G}} = (\mathbf{g}_1', \dots, \mathbf{g}_T')'$, where $\mathbf{X}$ is an $NT \times 1$ vector, $\boldsymbol{\mathcal{G}}$ is an $(r_1 + r_2)T \times 1$ vector. These are $T$-dimensional realizations of the stochastic processes $\{\mathbf{x}_t\}$ and $\{\mathbf{g}_t\}$, respectively. Moreover, let $\boldsymbol{X}_v$ be the $\sigma$-algebra generated by the random variables $\{\mathbf{x}_t\}_{t=1}^{v}$, for $v = 1, \dots, T$; in a similar way, define $\boldsymbol{G}_v$ as the $\sigma$-algebra generated by the random variables $\{\mathbf{g}_t\}_{t=1}^{v}$, for $v = 1, \dots, T$. And for simplicity we write $\boldsymbol{X} \equiv \boldsymbol{X}_T$ and $\boldsymbol{G} \equiv \boldsymbol{G}_T$.

The likelihood function, denoted by $f(\mathbf{X}; \mathbf{q})$, can be decomposed as

$$f(\mathbf{X}; \mathbf{q}) = \frac{f(\mathbf{X}, \boldsymbol{\mathcal{G}}; \mathbf{q})}{f(\boldsymbol{\mathcal{G}} \,|\boldsymbol{X}; \mathbf{q})} = \frac{f(\mathbf{X} \,|\boldsymbol{G}; \mathbf{q})\, f(\boldsymbol{\mathcal{G}}; \mathbf{q})}{f(\boldsymbol{\mathcal{G}} \,|\boldsymbol{X}; \mathbf{q})} = \frac{f(\mathbf{X} \,|\boldsymbol{G}; \mathbf{q})\, f(\boldsymbol{\mathcal{G}})}{f(\boldsymbol{\mathcal{G}} \,|\boldsymbol{X}; \mathbf{q})}, \tag{11}$$

where in the last step we accounted for the fact that $f(\boldsymbol{\mathcal{G}}; \mathbf{q}) \equiv f(\boldsymbol{\mathcal{G}})$, since it does not depend on the parameters of our model, because we do not specify any dynamic model for the process $\{\mathbf{g}_t\}$.

Furthermore, following Krolzig (2013, Section 6.2):

$$f(\mathbf{X}|\boldsymbol{G};\mathbf{q}) = f(\mathbf{X}|\boldsymbol{G};\boldsymbol{\varphi},\boldsymbol{\rho}) = \sum_{\{\boldsymbol{\xi}_t\}_{t=1}^T\in\{0,1\}^T} f\left(\mathbf{X}\left|\boldsymbol{G},\{\boldsymbol{\xi}_t\}_{t=1}^T;\boldsymbol{\varphi}\right.\right) \mathsf{P}\left(\{\boldsymbol{\xi}_t\}_{t=1}^T|\boldsymbol{G},\boldsymbol{\rho}\right). \tag{12}$$

Here, to avoid heavier notation, we use the same notation $\{\boldsymbol{\xi}_t\}_{t=1}^T$ both for a generic $T$ dimensional realization of the process $\{\boldsymbol{\xi}_t\}$ and for the $\sigma$-algebra generated by the random variables $\{\boldsymbol{\xi}_t\}_{t=1}^T$. Notice that the sum is over $2^T$ possible values since, given a realization for $\{\xi_{1t}\}_{t=1}^T$, then the realizations of $\{\xi_{2t}\}_{t=1}^T$ are given by $\xi_{2t} = 1 - \xi_{1t}$ for all $t$.

Following the approach by Doz et al. (2012); Bai and Li (2016); Barigozzi and Luciani (2019) for QML estimation of linear factor models, we consider for $f\left(\mathbf{X}\left|\boldsymbol{G},\{\boldsymbol{\xi}_t\}_{t=1}^T;\boldsymbol{\varphi}\right.\right)$ a mis-specified Gaussian quasi likelihood of an exact factor model, i.e., as if the idiosyncratic components were cross-sectionally uncorrelated. Furthermore, we also neglect serial correlation of the idiosyncratic components, thus treating them as if they were weak white noise processes. It is important to stress that we are not assuming the idiosyncratic components to be uncorrelated, but we are just considering likelihood estimation of a mis-specified model. In the linear case, it is proved that, as $N,T\to\infty$ such mis-specifications are asymptotically negligible. Under such mis-specification and using the Markov property of $\{\boldsymbol{\xi}_t\}$ (up to omitted constant terms):

$$\begin{aligned}\log f\left(\mathbf{X}\left|\boldsymbol{G},\{\boldsymbol{\xi}_t\}_{t=1}^T;\boldsymbol{\varphi}\right.\right) &= \sum_{t=1}^T \log f(\mathbf{x}_t|\mathbf{g}_t,\boldsymbol{\xi}_t;\boldsymbol{\varphi}) \qquad (13)\\ &\simeq -\frac{1}{2}\sum_{t=1}^T \log\det\boldsymbol{\Sigma}_{et} - \frac{1}{2}\sum_{t=1}^T \{\mathbf{x}_t - (\mathbf{B}_1\ \mathbf{B}_2)(\boldsymbol{\xi}_t\otimes\mathbf{g}_t)\}'(\boldsymbol{\Sigma}_{et})^{-1}\{\mathbf{x}_t - (\mathbf{B}_1\ \mathbf{B}_2)(\boldsymbol{\xi}_t\otimes\mathbf{g}_t)\},\end{aligned}$$

where $\boldsymbol{\Sigma}_{et} = (\text{diag}(\boldsymbol{\Sigma}_{e1})\ \text{diag}(\boldsymbol{\Sigma}_{e2}))(\boldsymbol{\xi}_t\otimes\mathbf{I}_N)$. Notice also that, even in this case the likelihood (12) is not Gaussian but it is a mixture of Gaussians. Finally, again by the Markov property of $\{\boldsymbol{\xi}_t\}$:

$$\mathsf{P}\left(\{\boldsymbol{\xi}_t\}_{t=1}^T|\boldsymbol{G};\boldsymbol{\rho}\right) = \prod_{t=1}^T \mathsf{P}(\boldsymbol{\xi}_t|\boldsymbol{\xi}_{t-1},\boldsymbol{G};\boldsymbol{\rho})\,\mathsf{P}(\boldsymbol{\xi}_0). \tag{14}$$

## 3 Estimation: EM algorithm

The following algorithm is a generalization of the procedure proposed by Krolzig (2013, Chapter 5). The EM algorithm is made of two steps repeated at each iteration $k\geq 0$. The E step involves taking the expected value of the log-likelihood derived from (11) conditional on $\boldsymbol{X}$ given an estimate of the parameters $\widehat{\mathbf{q}}^{(k)}$. Namely:

$$\log f(\mathbf{X};\mathbf{q}) = \mathsf{E}_{\widehat{\mathbf{q}}^{(k)}}[\log f(\mathbf{X}|\boldsymbol{G};\mathbf{q})|\boldsymbol{X}] + \mathsf{E}_{\widehat{\mathbf{q}}^{(k)}}[\log f(\boldsymbol{\mathcal{G}})|\boldsymbol{X}] - \mathsf{E}_{\widehat{\mathbf{q}}^{(k)}}[\log f(\boldsymbol{\mathcal{G}}|\boldsymbol{X};\mathbf{q})|\boldsymbol{X}].$$

The M step solves the constrained maximization problem with respect to $\mathbf{q} = [\boldsymbol{\varphi}', \boldsymbol{\rho}']'$, that is

$$\left(\widehat{\boldsymbol{\varphi}}^{(k+1)}, \widehat{\boldsymbol{\rho}}^{(k+1)}\right) = \arg\max_{\boldsymbol{\varphi},\boldsymbol{\rho}} \mathsf{E}_{\widehat{\mathbf{q}}^{(k)}} \left[\log f\left(\mathbf{X}\,|\boldsymbol{G};\boldsymbol{\varphi},\boldsymbol{\rho}\right)|\boldsymbol{X}\right]$$
$$\text{s.t.}\quad \mathbf{P}\boldsymbol{\iota}_2 = \boldsymbol{\iota}_2, \tag{15}$$

where the constraints ensure that probabilities add up to one. In principle, in the M step we should also account for the term $\mathsf{E}_{\widehat{\mathbf{q}}^{(k)}}\left[\log f\left(\boldsymbol{\mathcal{G}}\right)|\boldsymbol{X}\right]$, which however in our context does not depend on any parameter.

It is well known that the iteration of these steps produces a series of log-likelihoods which are increasing. Indeed, $\mathsf{E}_{\widehat{\mathbf{q}}^{(k)}}\left[\log f\left(\boldsymbol{\mathcal{G}}\,|\boldsymbol{X};\mathbf{q}\right)|\boldsymbol{X}\right]$ does not contribute to the convergence of the EM algorithm (see Dempster et al., 1977 and Wu, 1983). Moreover, if the maximum is identified and unique, then the EM algorithm will eventually lead to lead to the Maximum Likelihood estimator of $\mathbf{q}$. As shown below, the solution of the M step can be computed using the expressions given in (13) and (14). Such solution is unique and has a closed form, so, no identification issue, due to multiple maxima or related to the existence of such maxima, arises in this context.

## 3.1 Baum-Lindgren-Hamilton-Kim filter and smoother

In order to compute the expected likelihood in the E step it is clear that, because of (13) and (14), we need to compute $\mathsf{E}_{\widehat{\mathbf{q}}^{(k)}}[\boldsymbol{\xi}_t|\boldsymbol{X}]$, $\mathsf{E}_{\widehat{\mathbf{q}}^{(k)}}[\boldsymbol{\xi}_t \otimes \mathbf{g}_t|\boldsymbol{X}]$, and $\mathsf{E}_{\widehat{\mathbf{q}}^{(k)}}[(\boldsymbol{\xi}_t \otimes \mathbf{g}_t)(\boldsymbol{\xi}_t \otimes \mathbf{g}_t)'|\boldsymbol{X}] = \mathsf{E}_{\widehat{\mathbf{q}}^{(k)}}[(\mathbf{I}_2 \otimes \mathbf{g}_t\mathbf{g}_t')|\boldsymbol{X}]$.

We start by considering the case in which $\{\mathbf{g}_t\}_{t=1}^T$ is observed, while we postpone the discussion of the estimation of the factors to Section 3.2. Then, for the E step we just need to compute $\mathsf{E}_{\widehat{\mathbf{q}}^{(k)}}[\boldsymbol{\xi}_t|\boldsymbol{X}]$, since in this case $\boldsymbol{\xi}_t$ and $\mathbf{g}_t$ are independent for all $t$. This is accomplished by means of a generalization the Baum-Lindgren-Hamilton-Kim filter and smoother explained in detail in this section. It is an iterative procedure through which we first compute the sequences of conditional one-step-ahead predicted probabilities $\{\boldsymbol{\xi}_{t|t-1}\}_{t=1}^T$, such that $\boldsymbol{\xi}_{t|t-1} = \mathsf{E}_{\widehat{\mathbf{q}}^{(k)}}\left[\boldsymbol{\xi}_t\,|\boldsymbol{X}_{t-1}\right]$ and filtered probabilities $\{\boldsymbol{\xi}_{t|t}\}_{t=1}^T$ such that $\boldsymbol{\xi}_{t|t} = \mathsf{E}_{\widehat{\mathbf{q}}^{(k)}}[\boldsymbol{\xi}_t|\boldsymbol{X}_t]$. Second, by means of those sequences, we compute the sequence of smoothed probabilities $\{\boldsymbol{\xi}_{t|T}\}_{t=1}^T$ such that $\boldsymbol{\xi}_{t|T} = \mathsf{E}_{\widehat{\mathbf{q}}^{(k)}}[\boldsymbol{\xi}_t|\boldsymbol{X}]$.

To simplify notation, let $\boldsymbol{\varepsilon}_1 = [1\ 0]'$ and $\boldsymbol{\varepsilon}_2 = [0\ 1]'$, so that $\mathsf{P}(s_t = j) \equiv \mathsf{P}(\boldsymbol{\xi}_t = \boldsymbol{\varepsilon}_j)$, $j = 1, 2$, and therefore, in the following, we can just use $\boldsymbol{\xi}_t$ as defined in (4), without the need of referring also to $s_t$. Then, for any $v = 1, \ldots, T$, we use the notation

$$\boldsymbol{\xi}_{t|v} = \mathsf{E}\left[\boldsymbol{\xi}_t\,|\boldsymbol{X}_v\right] = \begin{bmatrix} \mathsf{P}\left(\boldsymbol{\xi}_t = \boldsymbol{\varepsilon}_1\,|\boldsymbol{X}_v\right) \\ \mathsf{P}\left(\boldsymbol{\xi}_t = \boldsymbol{\varepsilon}_2\,|\boldsymbol{X}_v\right) \end{bmatrix}. \tag{16}$$

Notice also that, since $\{\boldsymbol{\xi}_t\}_{t=1}^u$ is independent of $\boldsymbol{G}_v$ for all $u, v = 1, \ldots, T$, because we consider the factors as observed, we can always write $\boldsymbol{\xi}_{t|v} = \mathsf{E}\left[\boldsymbol{\xi}_t\,|\boldsymbol{X}_v\right] = \mathsf{E}\left[\boldsymbol{\xi}_t\,|\mathbf{X}_v, \boldsymbol{G}_v\right]$.

The one-step-ahead predictions and the filtered probabilities are computed by means of the following steps which are similar to the Hamilton filter, see, e.g., Krolzig (2013, Chapter 5.1) and Hamilton (1989). For simplicity of notation, let us assume for the moment that not only the factors $\{\mathbf{g}_t\}_{t=1}^T$ are observed, but also the true values of the parameters $\mathbf{q}$ are known.

Then, the one-step-ahead predicted probabilities are obtained through the prior probability

$$\begin{aligned}\mathsf{P}(\boldsymbol{\xi}_t = \boldsymbol{\varepsilon}_i \,|\boldsymbol{X}_{t-1}, \boldsymbol{G}_{t-1}) &= \sum_{j=1}^{2} \mathsf{P}(\boldsymbol{\xi}_t = \boldsymbol{\varepsilon}_i \,|\boldsymbol{\xi}_{t-1} = \boldsymbol{\varepsilon}_j)\, \mathsf{P}(\boldsymbol{\xi}_{t-1} = \boldsymbol{\varepsilon}_j \,|\boldsymbol{X}_{t-1}, \boldsymbol{G}_{t-1}) \\ &= \sum_{j=1}^{2} \mathsf{P}(\boldsymbol{\xi}_t = \boldsymbol{\varepsilon}_i \,|\boldsymbol{\xi}_{t-1} = \boldsymbol{\varepsilon}_j)\, \mathsf{P}(\boldsymbol{\xi}_{t-1} = \boldsymbol{\varepsilon}_j \,|\boldsymbol{X}_{t-1}), \qquad i = 1, 2. \end{aligned} \tag{17}$$

So that, because of (16), we have

$$\boldsymbol{\xi}_{t|t-1} = \mathbf{P}' \boldsymbol{\xi}_{t-1|t-1}, \quad t = 1, \ldots, T. \tag{18}$$

The update involves the posterior probability:

$$\begin{aligned}\mathsf{P}(\boldsymbol{\xi}_t = \boldsymbol{\varepsilon}_i \,|\boldsymbol{X}_t) = \mathsf{P}(\boldsymbol{\xi}_t = \boldsymbol{\varepsilon}_i \,|\boldsymbol{X}_t, \boldsymbol{G}_t) &= \mathsf{P}(\boldsymbol{\xi}_t = \boldsymbol{\varepsilon}_i \,|\mathbf{x}_t, \boldsymbol{X}_{t-1}, \boldsymbol{G}_t) \\ &= \frac{f(\mathbf{x}_t, \boldsymbol{\xi}_t = \boldsymbol{\varepsilon}_i \,|\boldsymbol{X}_{t-1}, \boldsymbol{G}_t)}{f(\mathbf{x}_t \,|\boldsymbol{X}_{t-1}, \boldsymbol{G}_t)} \\ &= \frac{f(\mathbf{x}_t \,|\boldsymbol{\xi}_t = \boldsymbol{\varepsilon}_i, \boldsymbol{X}_{t-1}, \boldsymbol{G}_t)\, \mathsf{P}(\boldsymbol{\xi}_t = \boldsymbol{\varepsilon}_i \,|\boldsymbol{X}_{t-1}, \boldsymbol{G}_t)}{f(\mathbf{x}_t \,|\boldsymbol{X}_{t-1}, \boldsymbol{G}_t)}, \quad i = 1, 2. \end{aligned} \tag{19}$$

Then, since $\mathbf{x}_t$ depends on $\boldsymbol{X}_{t-1}$ only through $\boldsymbol{\xi}_{t-1}$ and it depends on $\boldsymbol{G}_t$ only through $\mathbf{g}_t$

$$f(\mathbf{x}_t \,|\boldsymbol{\xi}_t = \boldsymbol{\varepsilon}_i, \boldsymbol{X}_{t-1}, \boldsymbol{G}_t) = f(\mathbf{x}_t \,|\boldsymbol{\xi}_t = \boldsymbol{\varepsilon}_i, \mathbf{g}_t), \quad i = 1, 2. \tag{20}$$

Let,

$$\begin{aligned}\boldsymbol{\eta}_t &= \begin{bmatrix} f(\mathbf{x}_t \,|\boldsymbol{\xi}_t = \boldsymbol{\varepsilon}_1, \mathbf{g}_t) \\ f(\mathbf{x}_t \,|\boldsymbol{\xi}_t = \boldsymbol{\varepsilon}_2, \mathbf{g}_t) \end{bmatrix} \\ &= \frac{1}{(2\pi)^{N/2}} \left\{ \begin{array}{c} |\mathrm{diag}(\boldsymbol{\Sigma}_{e1})|^{-1/2} \exp\left[ -\frac{1}{2} (\mathbf{x}_t - \mathbf{B}_1 \mathbf{g}_t)' (\mathrm{diag}(\boldsymbol{\Sigma}_{e1}))^{-1} (\mathbf{x}_t - \mathbf{B}_1 \mathbf{g}_t) \right] \\ \\ |\mathrm{diag}(\boldsymbol{\Sigma}_{e2})|^{-1/2} \exp\left[ -\frac{1}{2} (\mathbf{x}_t - \mathbf{B}_2 \mathbf{g}_t)' (\mathrm{diag}(\boldsymbol{\Sigma}_{e2}))^{-1} (\mathbf{x}_t - \mathbf{B}_2 \mathbf{g}_t) \right] \end{array} \right\}. \end{aligned} \tag{21}$$

Further, notice that, from (16) and (21), the denominator of (19) be written as:

$$
\begin{aligned}
f\left(\mathbf{x}_t \,|\boldsymbol{X}_{t-1}, \boldsymbol{G}_t\right) &= \sum_{j=1}^{2} f\left(\mathbf{x}_t \,|\boldsymbol{\xi}_t = \boldsymbol{\varepsilon}_j, \boldsymbol{X}_{t-1}, \boldsymbol{G}_t\right) \mathsf{P}\left(\boldsymbol{\xi}_t = \boldsymbol{\varepsilon}_j, |\boldsymbol{X}_{t-1}, \boldsymbol{G}_t\right) \\
&= \sum_{j=1}^{2} f\left(\mathbf{x}_t \,|\boldsymbol{\xi}_t = \boldsymbol{\varepsilon}_j, \mathbf{g}_t\right) \mathsf{P}\left(\boldsymbol{\xi}_t = \boldsymbol{\varepsilon}_j, |\boldsymbol{X}_{t-1}\right) = \boldsymbol{\eta}_t' \boldsymbol{\xi}_{t|t-1}\,. \qquad (22)
\end{aligned}
$$

Taking into account (16), (17), (20), and (22), the filtered probabilities are obtained from (19) as

$$
\boldsymbol{\xi}_{t|t} = \frac{\boldsymbol{\eta}_t \odot \boldsymbol{\xi}_{t|t-1}}{\boldsymbol{\eta}_t' \boldsymbol{\xi}_{t|t-1}} = \frac{\boldsymbol{\eta}_t \odot \boldsymbol{\xi}_{t|t-1}}{\boldsymbol{\iota}_2' \left(\boldsymbol{\eta}_t \odot \boldsymbol{\xi}_{t|t-1}\right)}, \quad t = 1, \ldots, T, \qquad (23)
$$

where $\boldsymbol{\eta}_t$ is computed as in (21). The filter can started by setting either $\boldsymbol{\xi}_{0|0} = \boldsymbol{\varepsilon}_1$, or, equivalently, $\boldsymbol{\xi}_{0|0} = \boldsymbol{\varepsilon}_2$.

We then run the Kim smoother, see e.g., Krolzig (2013, Chapter 5.2) and Kim (1994). Notice that (recall that $\boldsymbol{X} \equiv \boldsymbol{X}_T$ and $\boldsymbol{G} \equiv \boldsymbol{G}_T$):

$$
\begin{aligned}
&\mathsf{P}\left(\boldsymbol{\xi}_t = \boldsymbol{\varepsilon}_i \,|\boldsymbol{X}, \boldsymbol{G}\right) = \sum_{j=1}^{2} \mathsf{P}\left(\boldsymbol{\xi}_t = \boldsymbol{\varepsilon}_i \,|\boldsymbol{\xi}_{t+1} = \boldsymbol{\varepsilon}_j, \boldsymbol{X}, \boldsymbol{G}\right) \mathsf{P}\left(\boldsymbol{\xi}_{t+1} = \boldsymbol{\varepsilon}_j \,|\boldsymbol{X}, \boldsymbol{G}\right) \\
&= \sum_{j=1}^{2} \frac{\mathsf{P}\left(\boldsymbol{\xi}_t = \boldsymbol{\varepsilon}_i \,|\boldsymbol{\xi}_{t+1} = \boldsymbol{\varepsilon}_j, \boldsymbol{X}_t, \boldsymbol{G}_t\right) f\left(\left\{\mathbf{x}_s, \mathbf{g}_s\right\}_{s=t+1}^{T} |\boldsymbol{\xi}_t = \boldsymbol{\varepsilon}_i, \boldsymbol{\xi}_{t+1} = \boldsymbol{\varepsilon}_j, \boldsymbol{X}_t, \boldsymbol{G}_t\right)}{f\left(\left\{\mathbf{x}_s, \mathbf{g}_s\right\}_{s=t+1}^{T} |\boldsymbol{\xi}_{t+1} = \boldsymbol{\varepsilon}_j, \boldsymbol{X}_t, \boldsymbol{G}_t\right)} \mathsf{P}\left(\boldsymbol{\xi}_{t+1} = \boldsymbol{\varepsilon}_j \,|\boldsymbol{X}, \boldsymbol{G}\right) \\
&= \sum_{j=1}^{2} \mathsf{P}\left(\boldsymbol{\xi}_t = \boldsymbol{\varepsilon}_i \,|\boldsymbol{\xi}_{t+1} = \boldsymbol{\varepsilon}_j, \boldsymbol{X}_t, \boldsymbol{G}_t\right) \mathsf{P}\left(\boldsymbol{\xi}_{t+1} = \boldsymbol{\varepsilon}_j \,|\boldsymbol{X}, \boldsymbol{G}\right) \\
&= \sum_{j=1}^{2} \frac{\mathsf{P}\left(\boldsymbol{\xi}_t = \boldsymbol{\varepsilon}_i \,|\boldsymbol{X}_t, \boldsymbol{G}_t\right) \mathsf{P}\left(\boldsymbol{\xi}_{t+1} = \boldsymbol{\varepsilon}_j \,|\boldsymbol{\xi}_t = \boldsymbol{\varepsilon}_i, \boldsymbol{X}_t, \boldsymbol{G}_t\right)}{\mathsf{P}\left(\boldsymbol{\xi}_{t+1} = \boldsymbol{\varepsilon}_j \,|\boldsymbol{X}_t, \boldsymbol{G}_t\right)} \mathsf{P}\left(\boldsymbol{\xi}_{t+1} = \boldsymbol{\varepsilon}_j \,|\boldsymbol{X}, \boldsymbol{G}\right), \quad i = 1, 2,
\end{aligned}
$$

which by (16) implies that the sequence of smoothed probabilities is given by

$$
\boldsymbol{\xi}_{t|T} = \left[\mathbf{P}\left(\boldsymbol{\xi}_{t+1|T} \oslash \boldsymbol{\xi}_{t+1|t}\right)\right] \odot \boldsymbol{\xi}_{t|t}\,, \quad t = 1, \ldots, T. \qquad (24)
$$

This backward recursion is initiated at $\boldsymbol{\xi}_{T|T}$ which is the last iteration of the filter in (23).

Finally, for the implementation of the EM algorithm we need to compute also the smoothed cross-probabilities, see Krolzig (2013, Chapter 5.A.2),

$$
\boldsymbol{\xi}_{t,t-1|T} = \begin{bmatrix} \mathsf{P}(\boldsymbol{\xi}_t = \boldsymbol{\varepsilon}_1, \boldsymbol{\xi}_{t-1} = \boldsymbol{\varepsilon}_1 | \boldsymbol{X}) \\ \mathsf{P}(\boldsymbol{\xi}_t = \boldsymbol{\varepsilon}_2, \boldsymbol{\xi}_{t-1} = \boldsymbol{\varepsilon}_1 | \boldsymbol{X}) \\ \mathsf{P}(\boldsymbol{\xi}_t = \boldsymbol{\varepsilon}_1, \boldsymbol{\xi}_{t-1} = \boldsymbol{\varepsilon}_2 | \boldsymbol{X}) \\ \mathsf{P}(\boldsymbol{\xi}_t = \boldsymbol{\varepsilon}_2, \boldsymbol{\xi}_{t-1} = \boldsymbol{\varepsilon}_2 | \boldsymbol{X}) \end{bmatrix} = \boldsymbol{\rho} \odot \left[\left(\boldsymbol{\xi}_{t|T} \oslash \boldsymbol{\xi}_{t|t-1}\right) \otimes \boldsymbol{\xi}_{t-1|t-1}\right], \quad t = 1, \ldots, T. \qquad (25)
$$

The above description of the Baum-Lindgren-Hamilton-Kim filter and smoother assumes $\mathbf{q}$ and $\mathbf{g}_t$ to be observed. However, in practice both need to be estimated. This is discussed in the next two sections.

### 3.2 Estimating the factor space

To estimate both the factors $\mathbf{g}_t$ and their dimension $r_1 + r_2$, we exploit the fact that, as shown in Section 2.3, our Markov switching factor model (1) is observationally equivalent to a linear factor model with $r_1 + r_2$ common factors $\mathbf{g}_t$ and factor loadings $\mathbf{A}$, see (10). Then, the number of factors can be estimated by any of the existing methods, see, e.g., Bai and Ng (2002); Onatski (2010); Ahn and Horenstein (2013); Trapani (2018). As far as the factors $\mathbf{g}_t$ themselves they can be estimated via PCA as follows. First, an estimator $\widehat{\mathbf{A}}$ of the loadings matrix $\mathbf{A}$ is obtained as $\sqrt{N}$ times the normalized eigenvectors corresponding to the $r_1 + r_2$ largest eigenvalues of the sample $N \times N$ covariance matrix $T^{-1}\sum_{t=1}^{T} \mathbf{x}_t\mathbf{x}_t'$. Second, the factors are estimated by linear projection of the data $\mathbf{x}_t$ onto the estimated loadings:

$$\widehat{\mathbf{g}}_t = \frac{1}{N}\widehat{\mathbf{A}}'\mathbf{x}_t, \quad t = 1, \ldots, T. \tag{26}$$

This is the same approach followed by Stock and Watson (2002a) and it is the dual approach of the one adopted by Bai (2003). Consistency of $\widehat{\mathbf{A}}$ and $\widehat{\mathbf{g}}_t$ are proved in the Appendix. Notice that the steps described in this section do not require knowing the latent state indicator $\boldsymbol{\xi}_t$, hence can be carried out independently.

### 3.3 Estimating the parameters

At each iteration $k \geq 0$ of the EM algorithm, the filtered and smoothed probabilities, given in (23) and (24), respectively, and the smoothed cross-probabilities given in (25), are computed using an estimator $\widehat{\mathbf{q}}^{(k)}$ of the parameters and an estimator $\widehat{\mathbf{g}}_t$ of the factors. Hereafter, we denote as $\boldsymbol{\xi}_{t|t}^{(k)}$, $\boldsymbol{\xi}_{t|T}^{(k)}$, and $\boldsymbol{\xi}_{t,t-1|T}^{(k)}$ such estimators. This defines the E step.

In the M step we have to solve the constrained maximization problem in (15). Let us start with estimation of $\boldsymbol{\varphi}$. From (12), we have:

$$\begin{aligned}
\frac{\partial \log f(\mathbf{X}\,|\boldsymbol{G};\boldsymbol{\varphi},\boldsymbol{\rho})}{\partial \boldsymbol{\varphi}'} &= \frac{1}{f(\mathbf{X}\,|\boldsymbol{G};\boldsymbol{\varphi},\boldsymbol{\rho})} \sum_{\{\boldsymbol{\xi}_t\}_{t=1}^T} \frac{\partial f\left(\mathbf{X}\,|\boldsymbol{G},\{\boldsymbol{\xi}_t\}_{t=1}^T;\boldsymbol{\varphi}\right)}{\partial \boldsymbol{\varphi}'} \mathsf{P}\left(\{\boldsymbol{\xi}_t\}_{t=1}^T|\boldsymbol{G},\boldsymbol{\rho}\right) \\
&= \frac{1}{f(\mathbf{X}\,|\boldsymbol{G};\boldsymbol{\varphi},\boldsymbol{\rho})} \sum_{\{\boldsymbol{\xi}_t\}_{t=1}^T} \frac{\partial \log f\left(\mathbf{X}\,|\boldsymbol{G},\{\boldsymbol{\xi}_t\}_{t=1}^T;\boldsymbol{\varphi}\right)}{\partial \boldsymbol{\varphi}'} f\left(\mathbf{X}\,|\boldsymbol{G},\{\boldsymbol{\xi}_t\}_{t=1}^T;\boldsymbol{\varphi}\right) \mathsf{P}\left(\{\boldsymbol{\xi}_t\}_{t=1}^T|\boldsymbol{G};\boldsymbol{\rho}\right) \\
&= \mathcal{C} \sum_{\{\boldsymbol{\xi}_t\}_{t=1}^T} \frac{\partial \log f\left(\mathbf{X}\,|\boldsymbol{G},\{\boldsymbol{\xi}_t\}_{t=1}^T;\boldsymbol{\varphi}\right)}{\partial \boldsymbol{\varphi}'} \mathsf{P}\left(\{\boldsymbol{\xi}_t\}_{t=1}^T|\boldsymbol{X},\boldsymbol{G};\boldsymbol{\varphi},\boldsymbol{\rho}\right),
\end{aligned} \tag{27}$$

where $\mathcal{C}$ is a positive normalization constant.[2] Therefore, from (13), (15), and (27), if we observed $\boldsymbol{G}$, the first order conditions would be:

$$\begin{aligned}
\mathbf{0} &= \frac{\partial \mathsf{E}_{\widehat{\mathbf{q}}^{(k)}}\left[\log f\left(\mathbf{X}\,|\boldsymbol{G};\boldsymbol{\varphi},\boldsymbol{\rho}\right)|\boldsymbol{X}\right]}{\partial \boldsymbol{\varphi}'}\bigg|_{\boldsymbol{\varphi}=\widehat{\boldsymbol{\varphi}}^{(k+1)}} \\
&= \sum_{t=1}^{T}\sum_{j=1}^{2} \frac{\partial \mathsf{E}_{\widehat{\mathbf{q}}^{(k)}}\left[\log f\left(\mathbf{x}_t\,|\mathbf{g}_t,\boldsymbol{\xi}_t=\boldsymbol{\varepsilon}_j;\boldsymbol{\varphi}\right)|\boldsymbol{X}\right]}{\partial \boldsymbol{\varphi}'}\bigg|_{\boldsymbol{\varphi}=\widehat{\boldsymbol{\varphi}}^{(k+1)}} \mathsf{P}\left(\boldsymbol{\xi}_t=\boldsymbol{\varepsilon}_j|\boldsymbol{X};\widehat{\boldsymbol{\varphi}}^{(k)},\widehat{\boldsymbol{\rho}}^{(k)}\right) \\
&= \sum_{t=1}^{T}\sum_{j=1}^{2} \frac{\partial \mathsf{E}_{\widehat{\mathbf{q}}^{(k)}}\left[\log f\left(\mathbf{x}_t\,|\mathbf{g}_t,\boldsymbol{\xi}_t=\boldsymbol{\varepsilon}_j;\boldsymbol{\varphi}\right)|\boldsymbol{X}\right]}{\partial \boldsymbol{\varphi}'}\bigg|_{\boldsymbol{\varphi}=\widehat{\boldsymbol{\varphi}}^{(k+1)}} \xi_{j,t|T}^{(k)},
\end{aligned} \tag{28}$$

where $\xi_{j,t|T}^{(k)} = \mathsf{E}_{\widehat{\mathbf{q}}^{(k)}}[\xi_{jt}|\mathbf{X}] = \mathsf{P}(\boldsymbol{\xi}_t=\boldsymbol{\varepsilon}_j\,\big|\boldsymbol{X};\widehat{\boldsymbol{\varphi}}^{(k)},\widehat{\boldsymbol{\rho}}^{(k)})$ is the $j$th component of $\boldsymbol{\xi}_{t|T}^{(k)}$.

Then, by substituting (13) into (28), and by replacing true factors with estimated ones, we get

$$\widehat{\mathbf{B}}_j^{(k+1)} = \left(\sum_{t=1}^{T}\xi_{j,t|T}^{(k)}\mathbf{x}_t\widehat{\mathbf{g}}_t'\right)\left(\sum_{t=1}^{T}\xi_{j,t|T}^{(k)}\widehat{\mathbf{g}}_t\widehat{\mathbf{g}}_t'\right)^{-1}, \quad j=1,2, \tag{29}$$

and, consistently with the fact that we use a mis-specified likelihood with uncorrelated idiosyncratic components, we set

$$\begin{aligned}
[\widehat{\boldsymbol{\Sigma}}_{ej}^{(k+1)}]_{ii} &= \left(\frac{\sum_{t=1}^{T}\left(x_{it}-\widehat{\mathbf{b}}_{ji}^{(k+1)\prime}\widehat{\mathbf{g}}_t\right)^2}{\sum_{t=1}^{T}\xi_{j,t|T}^{(k)}}\right), \quad i=1,\ldots,N, \quad j=1,2, \\
[\widehat{\boldsymbol{\Sigma}}_{ej}^{(k+1)}]_{ik} &= 0, \quad i,k=1,\ldots,N, \quad i\neq k, \quad j=1,2,
\end{aligned} \tag{30}$$

where $\widehat{\mathbf{b}}_{ji}^{(k+1)\prime}$ is the $i$th row of $\widehat{\mathbf{B}}_j^{(k+1)}$.

Moving to estimation of $\boldsymbol{\rho}$, from (12), we have:

$$\begin{aligned}
\frac{\partial \log f\left(\mathbf{X}\,|\boldsymbol{G};\boldsymbol{\varphi},\boldsymbol{\rho}\right)}{\partial \boldsymbol{\rho}'} &= \frac{1}{f\left(\mathbf{X}\,|\boldsymbol{G};\boldsymbol{\varphi},\boldsymbol{\rho}\right)}\sum_{\{\boldsymbol{\xi}_t\}_{t=1}^T} f\left(\mathbf{X}\,|\boldsymbol{G},\{\boldsymbol{\xi}_t\}_{t=1}^T;\boldsymbol{\varphi}\right)\frac{\partial \mathsf{P}\left(\{\boldsymbol{\xi}_t\}_{t=1}^T|\boldsymbol{G};\boldsymbol{\rho}\right)}{\partial \boldsymbol{\rho}'} \\
&= \frac{1}{f\left(\mathbf{X}\,|\boldsymbol{G};\boldsymbol{\varphi},\boldsymbol{\rho}\right)}\sum_{\{\boldsymbol{\xi}_t\}_{t=1}^T}\frac{\partial \log \mathsf{P}\left(\{\boldsymbol{\xi}_t\}_{t=1}^T|\boldsymbol{G};\boldsymbol{\rho}\right)}{\partial \boldsymbol{\rho}'} f\left(\mathbf{X}\,|\boldsymbol{G},\{\boldsymbol{\xi}_t\}_{t=1}^T;\boldsymbol{\varphi}\right)\mathsf{P}\left(\{\boldsymbol{\xi}_t\}_{t=1}^T|\boldsymbol{G};\boldsymbol{\rho}\right) \\
&= \mathcal{C}\sum_{\{\boldsymbol{\xi}_t\}_{t=1}^T}\frac{\partial \log \mathsf{P}\left(\{\boldsymbol{\xi}_t\}_{t=1}^T|\boldsymbol{G};\boldsymbol{\rho}\right)}{\partial \boldsymbol{\rho}'}\mathsf{P}\left(\{\boldsymbol{\xi}_t\}_{t=1}^T|\boldsymbol{X},\boldsymbol{G};\boldsymbol{\varphi},\boldsymbol{\rho}\right),
\end{aligned} \tag{31}$$

[2] Specifically, we have:

$$\mathsf{P}\left(\{\boldsymbol{\xi}_t\}_{t=1}^T|\boldsymbol{X},\boldsymbol{G};\boldsymbol{\varphi},\boldsymbol{\rho}\right) = \frac{f\left(\mathbf{X}\,|\boldsymbol{G},\{\boldsymbol{\xi}_t\}_{t=1}^T;\boldsymbol{\varphi}\right)\mathsf{P}\left(\{\boldsymbol{\xi}_t\}_{t=1}^T|\boldsymbol{G};\boldsymbol{\rho}\right)}{\sum_{\{\boldsymbol{\xi}_t\}_{t=1}^T} f\left(\mathbf{X}\,|\boldsymbol{G},\{\boldsymbol{\xi}_t\}_{t=1}^T;\boldsymbol{\varphi}\right)\mathsf{P}\left(\{\boldsymbol{\xi}_t\}_{t=1}^T|\boldsymbol{G};\boldsymbol{\rho}\right)},$$

so $\mathcal{C} = \frac{\sum_{\{\boldsymbol{\xi}_t\}_{t=1}^T} f\left(\mathbf{X}|\boldsymbol{G},\{\boldsymbol{\xi}_t\}_{t=1}^T;\boldsymbol{\varphi}\right)\mathsf{P}\left(\{\boldsymbol{\xi}_t\}_{t=1}^T|\boldsymbol{G};\boldsymbol{\rho}\right)}{f(\mathbf{X}|\boldsymbol{G};\boldsymbol{\varphi},\boldsymbol{\rho})}$.

where $\mathcal{C}$ is the same positive normalization constant as in (27). And, because of (14) and (31), if we observed $\boldsymbol{G}$ the derivatives with respect to the generic $(i,j)$th element of $\boldsymbol{\rho}$, i.e, $p_{ij}$, $i,j=1,2$, would be (treating $\boldsymbol{\xi}_0$ as known)

$$
\begin{aligned}
&\frac{\partial \log f\left(\mathbf{X}\,|\boldsymbol{G};\boldsymbol{\varphi},\boldsymbol{\rho}\right)}{\partial p_{ij}}\\
&=\sum_{t=1}^{T}\sum_{h=1}^{2}\sum_{\ell=1}^{2}\frac{\partial \log \mathsf{P}\left(\boldsymbol{\xi}_t=\boldsymbol{\varepsilon}_h|\boldsymbol{\xi}_{t-1}=\boldsymbol{\varepsilon}_\ell;\boldsymbol{\rho}\right)}{\partial p_{ij}}\mathsf{P}\left(\boldsymbol{\xi}_t=\boldsymbol{\varepsilon}_h,\boldsymbol{\xi}_{t-1}=\boldsymbol{\varepsilon}_\ell|\boldsymbol{X};\boldsymbol{\varphi},\boldsymbol{\rho}\right)\\
&=\sum_{t=1}^{T}\sum_{h=1}^{2}\sum_{\ell=1}^{2}\frac{1}{\mathsf{P}\left(\boldsymbol{\xi}_t=\boldsymbol{\varepsilon}_h|\boldsymbol{\xi}_{t-1}=\boldsymbol{\varepsilon}_\ell;\boldsymbol{\rho}\right)}\frac{\partial \mathsf{P}\left(\boldsymbol{\xi}_t=\boldsymbol{\varepsilon}_h|\boldsymbol{\xi}_{t-1}=\boldsymbol{\varepsilon}_\ell;\boldsymbol{\rho}\right)}{\partial p_{ij}}\mathsf{P}\left(\boldsymbol{\xi}_t=\boldsymbol{\varepsilon}_h,\boldsymbol{\xi}_{t-1}=\boldsymbol{\varepsilon}_\ell|\boldsymbol{X};\boldsymbol{\varphi},\boldsymbol{\rho}\right)\\
&=\sum_{t=1}^{T}\sum_{h=1}^{2}\sum_{\ell=1}^{2}\frac{\mathbb{I}\left(\boldsymbol{\xi}_t=\boldsymbol{\varepsilon}_j,\boldsymbol{\xi}_{t-1}=\boldsymbol{\varepsilon}_i\right)}{\mathsf{P}\left(\boldsymbol{\xi}_t=\boldsymbol{\varepsilon}_h|\boldsymbol{\xi}_{t-1}=\boldsymbol{\varepsilon}_\ell;\boldsymbol{\rho}\right)}\mathsf{P}\left(\boldsymbol{\xi}_t=\boldsymbol{\varepsilon}_h,\boldsymbol{\xi}_{t-1}=\boldsymbol{\varepsilon}_\ell|\boldsymbol{X};\boldsymbol{\varphi},\boldsymbol{\rho}\right)\\
&=\sum_{t=1}^{T}\frac{\mathsf{P}\left(\boldsymbol{\xi}_t=\boldsymbol{\varepsilon}_j,\boldsymbol{\xi}_{t-1}=\boldsymbol{\varepsilon}_i|\boldsymbol{X};\boldsymbol{\varphi},\boldsymbol{\rho}\right)}{\mathsf{P}\left(\boldsymbol{\xi}_t=\boldsymbol{\varepsilon}_j|\boldsymbol{\xi}_{t-1}=\boldsymbol{\varepsilon}_i;\boldsymbol{\rho}\right)}=\sum_{t=1}^{T}\frac{\mathsf{P}\left(\boldsymbol{\xi}_t=\boldsymbol{\varepsilon}_j,\boldsymbol{\xi}_{t-1}=\boldsymbol{\varepsilon}_i|\boldsymbol{X};\boldsymbol{\varphi},\boldsymbol{\rho}\right)}{p_{ij}}. \qquad (32)
\end{aligned}
$$

Now, from (15) and (31), the first order conditions are:

$$
\mathbf{0}=\left\{\frac{\partial \mathsf{E}_{\widehat{\mathbf{q}}^{(k)}}\left[\log f\left(\mathbf{X}\,|\boldsymbol{G};\boldsymbol{\varphi},\boldsymbol{\rho}\right)|\boldsymbol{X}\right]}{\partial(\mathrm{vec}(\mathbf{P}))'}-\boldsymbol{\kappa}'\left(\boldsymbol{\iota}_2'\otimes\mathbf{I}_2\right)\right\}\Bigg|_{\mathrm{vec}(\mathbf{P})=\mathrm{vec}(\widehat{\mathbf{P}}^{(k+1)})}, \qquad (33)
$$

where $\boldsymbol{\kappa}$ is the 2-dimensional vector of Lagrange multipliers, thus it has positive entries. Then, from (32)

$$
\frac{\partial \mathsf{E}_{\widehat{\mathbf{q}}^{(k)}}\left[\log f\left(\mathbf{X}\,|\boldsymbol{G};\boldsymbol{\varphi},\boldsymbol{\rho}\right)|\boldsymbol{X}\right]}{\partial p_{ij}}=\sum_{t=1}^{T}\frac{\mathsf{P}\left(\boldsymbol{\xi}_t=\boldsymbol{\varepsilon}_j,\boldsymbol{\xi}_{t-1}=\boldsymbol{\varepsilon}_i|\boldsymbol{X};\widehat{\boldsymbol{\varphi}}^{(k)},\widehat{\boldsymbol{\rho}}^{(k)}\right)}{p_{ij}}. \qquad (34)
$$

By collecting all 4 terms deriving from (34) into a vector, we have

$$
\frac{\partial \mathsf{E}_{\widehat{\mathbf{q}}^{(k)}}\left[\log f\left(\mathbf{X}\,|\boldsymbol{G};\boldsymbol{\varphi},\boldsymbol{\rho}\right)|\boldsymbol{X}\right]}{\partial \boldsymbol{\rho}'}=\sum_{t=1}^{T}\boldsymbol{\xi}_{t,t-1|T}^{(k)\prime}\oslash\boldsymbol{\rho}', \qquad (35)
$$

where $\boldsymbol{\xi}_{t,t-1|T}^{(k)}$ is defined in (25). Finally, from the first order conditions (33), we must have:

$$
\mathbf{0}=\left\{\sum_{t=1}^{T}\boldsymbol{\xi}_{t,t-1|T}^{(k)\prime}\oslash\boldsymbol{\rho}'-\boldsymbol{\kappa}'\left(\boldsymbol{\iota}_2'\otimes\mathbf{I}_2\right)\right\}\Bigg|_{\boldsymbol{\rho}=\widehat{\boldsymbol{\rho}}^{(k+1)}}. \qquad (36)
$$

Let $\boldsymbol{\kappa}=(\kappa_1,\kappa_2)'$, and let $\widetilde{\boldsymbol{\kappa}}=(\boldsymbol{\iota}_2\otimes\boldsymbol{\kappa})=(\kappa_1,\kappa_2,\kappa_1,\kappa_2)'$. Then, (36) gives

$$
\widehat{\boldsymbol{\rho}}^{(k+1)}=\sum_{t=1}^{T}\boldsymbol{\xi}_{t,t-1|T}^{(k)}\oslash\widetilde{\boldsymbol{\kappa}}. \qquad (37)
$$

By applying the adding up condition to (37):

$$\boldsymbol{\iota}_2 = \left(\boldsymbol{\iota}_2' \otimes \mathbf{I}_2\right) \widehat{\boldsymbol{\rho}}^{(k+1)} = \left(\boldsymbol{\iota}_2' \otimes \mathbf{I}_2\right) \left(\sum_{t=1}^{T} \boldsymbol{\xi}_{t,t-1|T}^{(k)} \oslash \widetilde{\boldsymbol{\kappa}}\right) = \left(\boldsymbol{\iota}_2' \otimes \mathbf{I}_2\right) \sum_{t=1}^{T} \begin{pmatrix} \frac{\xi_{11,t,t-1|T}^{(k)}}{\kappa_1} \\ \frac{\xi_{21,t,t-1|T}^{(k)}}{\kappa_2} \\ \frac{\xi_{12,t,t-1|T}^{(k)}}{\kappa_1} \\ \frac{\xi_{22,t,t-1|T}^{(k)}}{\kappa_2} \end{pmatrix}$$

$$= \sum_{t=1}^{T} \sum_{j=1}^{2} \begin{pmatrix} \frac{\xi_{1j,t,t-1|T}^{(k)}}{\kappa_1} \\ \frac{\xi_{2j,t,t-1|T}^{(k)}}{\kappa_2} \end{pmatrix} = \sum_{t=1}^{T} \begin{pmatrix} \frac{\xi_{1,t-1|T}^{(k)}}{\kappa_1} \\ \frac{\xi_{2,t-1|T}^{(k)}}{\kappa_2} \end{pmatrix} = \sum_{t=0}^{T-1} \begin{pmatrix} \frac{\xi_{1,t|T}^{(k)}}{\kappa_1} \\ \frac{\xi_{2,t|T}^{(k)}}{\kappa_2} \end{pmatrix} = \sum_{t=0}^{T-1} \boldsymbol{\xi}_{t|T}^{(k)} \oslash \boldsymbol{\kappa},$$

which implies $\boldsymbol{\kappa} = \sum_{t=0}^{T-1} \boldsymbol{\xi}_{t|T}^{(k)}$. Therefore, from (37),

$$\widehat{\boldsymbol{\rho}}^{(k+1)} = \left[\sum_{t=1}^{T} \boldsymbol{\xi}_{t,t-1|T}^{(k)}\right] \oslash \left[\boldsymbol{\iota}_2 \otimes \sum_{t=0}^{T-1} \boldsymbol{\xi}_{t|T}^{(k)}\right]. \tag{38}$$

By letting $k^*$ be the last iteration of the EM algorithm, we define our final estimator of the parameters as $\widehat{\mathbf{q}} \equiv \widehat{\mathbf{q}}^{(k^*+1)}$, as given by (29), (30), and (38). The final estimator of $\boldsymbol{\xi}_t$ is defined as $\widehat{\boldsymbol{\xi}}_{t|T} \equiv \boldsymbol{\xi}_{t|T}^{(k^*+1)}$, i.e., obtained by running a last time the Baum-Lindgren-Hamilton-Kim filter using the final estimates of the parameters.

## 3.4 Initialization and convergence of the EM algorithm

To start the algorithm we need initial estimators $\widehat{\mathbf{q}}^{(0)}$ for the parameters. Specifically, we set $\widehat{\mathbf{B}}_1^{(0)} = \widehat{\mathbf{B}}_2^{(0)} = \widehat{\mathbf{A}}$, as defined in Section 3.2. Then, given also $\widehat{\mathbf{g}}_t$ as in (26), let $\widehat{\mathbf{e}}_t = \mathbf{x}_t - \widehat{\mathbf{A}}\widehat{\mathbf{g}}_t$, and we set $\widehat{\boldsymbol{\Sigma}}_{e1}^{(0)} = \widehat{\boldsymbol{\Sigma}}_{e2}^{(0)} = \mathrm{diag}\left(T^{-1} \sum_{t=1}^{T} \widehat{\mathbf{e}}_t \widehat{\mathbf{e}}_t'\right)$. Finally, we set

$$\widehat{\mathbf{P}}^{(0)} = \begin{bmatrix} 0.5 + \omega_1 & 1 - 0.5 - \omega_1 \\ 1 - 0.5 - \omega_2 & 0.5 + \omega_2 \end{bmatrix},$$

where $\omega_1, \omega_2 \in (0, 0.5)$ and $\omega_1 > \omega_2$. This initialization implicitly identifies state 1 as the most probable one, i.e., it is the state with largest unconditional probability as defined in (7).

We say that the EM algorithm converged at iterations $k^*$, where $k^*$ is the first value of $k$ such that:

$$\frac{\left|\log f\left(\mathbf{X} \,|\boldsymbol{G}; \widehat{\boldsymbol{\varphi}}^{(k)}, \widehat{\boldsymbol{\rho}}^{(k)}\right) - \log f\left(\mathbf{X} \,|\boldsymbol{G}; \widehat{\boldsymbol{\varphi}}^{(k-1)}, \widehat{\boldsymbol{\rho}}^{(k-1)}\right)\right|}{\frac{1}{2}\left\{\left|\log f\left(\mathbf{X} \,|\boldsymbol{G}; \widehat{\boldsymbol{\varphi}}^{(k)}, \widehat{\boldsymbol{\rho}}^{(k)}\right) + \log f\left(\mathbf{X} \,|\boldsymbol{G}; \widehat{\boldsymbol{\varphi}}^{(k-1)}, \widehat{\boldsymbol{\rho}}^{(k-1)}\right)\right\}\right.} < \epsilon,$$

for some a priori chosen threshold $\epsilon > 0$.

# 4 Asymptotic theory

For ease of reference let us write in scalar notation (1) and (10):

$$x_{it} = \sum_{j=1}^{2} \boldsymbol{\lambda}_{ji}' \mathbf{f}_{jt} \mathbb{I}(s_t = j) + \mathbf{e}_t = \mathbf{a}_i' \mathbf{g}_t + \mathbf{e}_t, \quad i = 1, \dots, N, \ t \in \mathbb{Z}, \tag{39}$$

where $\mathbf{e}_t \sim (\mathbf{0}, \sum_{j=1}^{2} \boldsymbol{\Sigma}_{ej} \mathbb{I}(s_t = j))$.

We consider the following set of assumptions, which generalizes the setting in Bai (2003) and Massacci (2017) to our framework.

**Assumption 1. Factors.**
*(a) For $j = 1, 2$ and all $t \in \mathbb{Z}$, $\mathsf{E}[\mathbf{f}_{jt}] = \mathbf{0}$ and $\mathsf{E}[\|\mathbf{f}_{jt}\|^4] < \infty$.*
*(b) For $j, k = 1, 2$, as $T \to \infty$, $T^{-1} \sum_{t=1}^{T} \mathbb{I}(s_t = j) h_{kt} \mathbf{f}_{jt} \mathbf{f}_{jt}' \xrightarrow{p} \boldsymbol{\Sigma}_{\mathbf{f}j}^{(k)}$, where $\boldsymbol{\Sigma}_{\mathbf{f}j}^{(k)}$ is $r_j \times r_j$ positive definite, and $\{h_{kt}\}_{t=1}^{T}$ is any sequence such that (i) $\mathsf{P}(0 \leq h_{kt} \leq 1) = 1$; (ii) $T^{-1} \sum_{t=1}^{T} h_{kt} \xrightarrow{p} \bar{h}_k > 0$.*

Assumption 1 restricts the factor processes $\{\mathbf{f}_{jt}\}$, $j = 1, 2$, so that appropriate moments exists. Note that the sequence $\{h_{kt}\}_{t=1}^{T}$ can be random or deterministic and it is introduced to account for the fact that we estimate the expected value of $\xi_{jt}$ and not $\xi_{jt}$ itself. Assumption 1 implies that $0 < \mathsf{P}(s_t = j) < 1$, $j = 1, 2$, thus ruling out the possibility that any of the states is absorbing. It also implies that for $j = 1, 2$, as $T \to \infty$,

$$\frac{1}{T} \sum_{t=1}^{T} \mathbb{I}(s_t = j) \mathbf{f}_{jt} \mathbf{f}_{jt}' \xrightarrow{p} \boldsymbol{\Sigma}_{\mathbf{f}j}, \tag{40}$$

where $\boldsymbol{\Sigma}_{\mathbf{f}j}$ is positive definite and

$$\frac{1}{T} \sum_{t=1}^{T} \mathbf{g}_t \mathbf{g}_t' \xrightarrow{p} \boldsymbol{\Sigma}_{\mathbf{g}} = \begin{pmatrix} \boldsymbol{\Sigma}_{\mathbf{f}_1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{\mathbf{f}_2} \end{pmatrix}. \tag{41}$$

Moreover, it is straightforward to see that, if $j \neq k$, then, for all $T \in \mathbb{N}$,

$$\frac{1}{T} \sum_{t=1}^{T} \mathbb{I}(s_t = j) \mathbf{f}_{jt} \mathbf{f}_{kt}' \mathbb{I}(s_t = k) = \mathbf{0}. \tag{42}$$

**Assumption 2. Loadings.**
*(a) For $j = 1, 2$, all $i = 1, \dots, N$ and all $N \in \mathbb{N}$, $\|\boldsymbol{\lambda}_{ji}\| \leq \bar{\lambda} < \infty$, where $\bar{\lambda}$ is independent of $j$, $i$, and $N$;*
*(b) For $j = 1, 2$, as $N \to \infty$, $N^{-1} \boldsymbol{\Lambda}_j' \boldsymbol{\Lambda}_j \to \boldsymbol{\Sigma}_{\boldsymbol{\Lambda}_j}$, where $\boldsymbol{\Sigma}_{\boldsymbol{\Lambda}_j}$ is $r_j \times r_j$ positive definite.*
*(c) As $N \to \infty$, $N^{-1} \boldsymbol{\Lambda}_1' \boldsymbol{\Lambda}_2 \to \boldsymbol{\Sigma}_{\boldsymbol{\Lambda}_{12}}$, where $\boldsymbol{\Sigma}_{\boldsymbol{\Lambda}_{12}}$ is $r_1 \times r_2$.*
*(d) For any $r_2 \times r_2$ full rank matrix $\mathbf{L}$, $\boldsymbol{\Lambda}_1 \neq \boldsymbol{\Lambda}_2 \mathbf{L}$.*

According to Assumption 2, loadings are nonstochastic and factors have a nonnegligible effect on the variance of $\{\mathbf{x}_t\}$ within each regime. The condition in part (d) ensures that the regimes are identified and it is analogous to the alternative hypothesis in the test for change in loadings developed in Pelger and Xiong (2022). This condition is trivially satisfied if $r_1 \neq r_2$ since the number of factors changes between regimes; if instead $r_1 = r_2$ then part (d) rules out the possibility that the columns of $\mathbf{\Lambda}_1$ are a linear combination of the columns of $\mathbf{\Lambda}_2$, in which case the regimes cannot be separately identified.

From this assumption it also follows that, as $N \to \infty$,

$$\frac{\mathbf{A}'\mathbf{A}}{N} \to \mathbf{\Sigma}_{\mathbf{A}} = \begin{pmatrix} \mathbf{\Sigma}_{\mathbf{\Lambda}_1} & \mathbf{\Sigma}_{\mathbf{\Lambda}_{12}} \\ \mathbf{\Sigma}'_{\mathbf{\Lambda}_{12}} & \mathbf{\Sigma}_{\mathbf{\Lambda}_2} \end{pmatrix}, \tag{43}$$

and

$$\frac{\mathbf{B}_1'\mathbf{B}_1}{N} \to \mathbf{\Sigma}_{\mathbf{B}_1} = \begin{pmatrix} \mathbf{\Sigma}_{\mathbf{\Lambda}_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}, \quad \frac{\mathbf{B}_2'\mathbf{B}_2}{N} \to \mathbf{\Sigma}_{\mathbf{B}_2} = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}_{\mathbf{\Lambda}_2} \end{pmatrix}, \quad \frac{\mathbf{B}_j'\mathbf{B}_k}{N} \to \mathbf{0}, \text{ if } j \neq k. \tag{44}$$

**Assumption 3. Idiosyncratic component.**

*(a)* *For all $i = 1, \ldots, N$, all $t \in \mathbb{Z}$ and all $N \in \mathbb{N}$, $\mathsf{E}[e_{it}] = 0$ and $\mathsf{E}[e_{it}^8] \leq M < \infty$, where $M$ is independent of $i$, $t$, and $N$.*

*(b)* *For all $t \in \mathbb{Z}$ and $N \in \mathbb{N}$,*

$$\frac{1}{N}\sum_{i,l=1}^{N} |\mathsf{E}[\mathbb{I}(s_t = j)\, e_{it}e_{lt}]| \leq M < \infty,$$

*where $M$ is independent of $t$ and $N$;*

*(c)* *For $j = 1, 2$, all $i, l = 1, \ldots, N$, all $N \in \mathbb{N}$ and all $T \in \mathbb{N}$,*

$$\mathsf{E}\left[\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\{\mathbb{I}(s_t = j)\, e_{it}e_{lt} - \mathsf{E}[\mathbb{I}(s_t = j)\, e_{it}e_{lt}]\}\right|^4\right] \leq M < \infty,$$

*where $M$ is independent of $j$, $i$, $l$, $N$, and $T$.*

Part (b) controls the amount of cross-sectional correlation we can allow for. It implies the usual assumption for approximate factor models of nondiagonal idiosyncratic covariances $\mathbf{\Sigma}_{ej}$, $j = 1, 2$. Part (b) also implies

$$\mathsf{E}\left[\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\mathbb{I}(s_t = j)e_{it}\right|^2\right] \leq M < \infty,$$

hence $N^{-1/2}\|\mathbb{I}(s_t = j)\mathbf{e}_t\| = O_p(1)$ for all $j$ and $t$. Part (c) limits time dependence and it is guaranteed together with part (a) if we assumed finite 8th order cumulants for the bivariate process $\{(e_{it}, e_{lt})\}$. Notice that the constant $M$ in the three parts of the assumption does not have to be the same one.

We limit the degree of dependence between factors and idiosyncratic components within each regime by means of the following assumption.

**Assumption 4. Weak dependence between common and idiosyncratic components.** *For $j = 1, 2$, and all $N$ and $T$,*

$$\mathsf{E}\left[\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{\sqrt{T}}\left[\sum_{t=1}^{T}\mathbb{I}\left(s_t = j\right)\mathbf{f}_{jt}e_{it}\right]\right\|^2\right] \le M < \infty,$$

*where $M$ is independent of $N \in \mathbb{N}$ and $T \in \mathbb{N}$.*

**Assumption 5. Eigenvalues.** *The eigenvalues of the $r \times r$ matrix $\mathbf{\Sigma_A \Sigma_g}$ are distinct, where $\mathbf{\Sigma_A}$ is defined in* (43) *and $\mathbf{\Sigma_g}$ is defined in* (41)*.*

This assumption guarantees a unique limit for $N^{-1}\mathbf{A}'\widehat{\mathbf{A}}$, indeed by assuming distinct eigenvalues we can uniquely identify the space spanned by the eigenvectors which are a linear combination of the columns of $\mathbf{A}$. Notice that $\mathbf{\Sigma_g}$ is block diagonal because of (42).

Assumptions 1 to 5 are enough to prove consistency of our estimators, however to derive the asymptotic distribution we strengthen those assumptions by means of the last two following assumptions.

**Assumption 6. Moments and Central Limit Theorems.**

*(a) For $j = 1, 2$, all $i = 1, \dots, N$ all $N \in \mathbb{N}$ and all $T \in \mathbb{N}$, for all $i$,*

$$\mathsf{E}\left[\left\|\frac{1}{\sqrt{NT}}\sum_{l=1}^{N}\sum_{t=1}^{T}\mathbf{a}_l\left\{\mathbb{I}\left(s_t = j\right)e_{it}e_{lt} - \mathsf{E}\left[\mathbb{I}\left(s_t = j\right)e_{it}e_{lt}\right]\right\}\right\|^2\right] \le M < \infty,$$

*where $M$ is independent of $j$, $i$, $N$, and $T$.*

*(b) For $j, k = 1, 2$, all $N \in \mathbb{N}$ and all $T \in \mathbb{N}$,*

$$\mathsf{E}\left[\left\|\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{I}\left(s_t = j\right)\boldsymbol{\lambda}_{ki}\mathbf{f}'_{jt}e_{it}\right\|^2\right] \le M < \infty,$$

*where $M$ is independent of $j$, $k$, $N$, and $T$.*

*(c) For $j, k = 1, 2$, all $i = 1, \dots, N$ and all $N \in \mathbb{N}$, as $T \to \infty$,*

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbb{I}\left(s_t = j\right)h_{kt}\mathbf{f}_{jt}e_{it} \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \mathbf{\Gamma}_{jki}\right),$$

*where* $\{h_{kt}\}_{t=1}^{T}$ *is defined in Assumption 1, and*

$$\boldsymbol{\Gamma}_{jki} = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \sum_{v=1}^{T} \mathbb{I}\left(s_t = j\right) \mathbb{I}\left(s_v = j\right) h_{kt} h_{kv} \mathsf{E}[\mathbf{f}_{jt}\mathbf{f}_{jv}' e_{it} e_{iv}].$$

*(d) For* $j = 1, 2$, *all* $t \in \mathbb{Z}$, *as* $N \to \infty$,

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \boldsymbol{\lambda}_{ji} e_{it} \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Phi}_{jt}\right),$$

*where*

$$\boldsymbol{\Phi}_{jt} = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \sum_{l=1}^{N} \boldsymbol{\lambda}_{ji} \boldsymbol{\lambda}_{jl}' \mathsf{E}\left[e_{it} e_{lt}\right].$$

Finally, we impose the standard restrictions on the convergence rates.

**Assumption 7. Rates.** *As* $N, T \to \infty$, $\sqrt{T}/N \to 0$ *and* $\sqrt{N}/T \to 0$.

Define the $(r_1 + r_2) \times (r_1 + r_2)$ matrix $\widehat{\mathbf{H}}$ as

$$\widehat{\mathbf{H}} = \frac{\mathbf{G}\mathbf{G}'}{T} \frac{\mathbf{A}'\widehat{\mathbf{A}}}{N} \widehat{\mathbf{V}}^{-1}, \tag{45}$$

where $\mathbf{G} = (\mathbf{g}_1, \dots, \mathbf{g}_T)$ and $\widehat{\mathbf{V}}$ is the $(r_1 + r_2) \times (r_1 + r_2)$ diagonal matrix containing the first $r_1 + r_2$ eigenvalues of $\widehat{\boldsymbol{\Sigma}}_{\mathbf{x}} = (NT)^{-1} \sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t'$ sorted in decreasing order. In Lemma 6 we prove that

$$p \lim_{N,T\to\infty} \frac{\mathbf{A}'\widehat{\mathbf{A}}}{N} = \mathbf{Q}, \text{ with } \mathbf{Q} = \boldsymbol{\Sigma}_{\mathbf{g}}^{-1/2} \boldsymbol{\Psi} \mathbf{V}^{1/2}, \tag{46}$$

where $\mathbf{V}$ is the $(r_1 + r_2) \times (r_1 + r_2)$ diagonal matrix of the first $(r_1 + r_2)$ eigenvalues of $\boldsymbol{\Sigma}_{\mathbf{g}}^{1/2} \boldsymbol{\Sigma}_{\mathbf{A}} \boldsymbol{\Sigma}_{\mathbf{g}}^{1/2}$ in decreasing order, and $\boldsymbol{\Psi}$ is the corresponding matrix of eigenvectors such that $\boldsymbol{\Psi}'\boldsymbol{\Psi} = \mathbf{I}_{r_1+r_2}$. Likewise define $\mathbf{Q}_j = p \lim_{N,T\to\infty} N^{-1} \boldsymbol{\Lambda}_j' \widehat{\mathbf{A}}$, for $j = 1, 2$, which is an $r_j \times (r_1 + r_2)$ matrix such that $\mathbf{Q} = [\mathbf{Q}_1' \ \mathbf{Q}_2']'$. Thus, by Lemma 7 we have

$$\mathbf{Q}_j = \boldsymbol{\Sigma}_{\mathbf{f}j}^{-1/2} \boldsymbol{\Psi}_j \mathbf{V}^{1/2}, \quad j = 1, 2, \tag{47}$$

where $\boldsymbol{\Psi}_j$ is the $r_j \times (r_1 + r_2)$ matrix such that $\boldsymbol{\Psi} = [\boldsymbol{\Psi}_1' \ \boldsymbol{\Psi}_2']'$. Therefore, because of (41), (46), and by Lemma 8 according to which $\widehat{\mathbf{V}} \xrightarrow{p} \mathbf{V}$,

$$p \lim_{N,T\to\infty} \widehat{\mathbf{H}} = \mathbf{H}, \text{ with } \mathbf{H} = \boldsymbol{\Sigma}_{\mathbf{g}} \mathbf{Q} \mathbf{V}^{-1}. \tag{48}$$

For $j = 1, 2$, let $\widehat{\mathbf{B}}_j = \widehat{\mathbf{B}}_j^{(k^*+1)}$, where $k^*$ is the last iteration of the EM algorithm as defined in Section 3.3. For given $j = 1, 2$ and $i = 1, \dots, N$, let $\widehat{\mathbf{b}}_{ji}$ be the estimator for $\mathbf{b}_{ji}$ such that $\widehat{\mathbf{B}}_j = [\widehat{\mathbf{b}}_{j1}, \dots, \widehat{\mathbf{b}}_{jN}]'$ and $\mathbf{B}_j = [\mathbf{b}_{j1}, \dots, \mathbf{b}_{jN}]'$. The following theorem states the asymptotic distribution of $\widehat{\mathbf{b}}_{ji}$.

**Theorem 1.** *Let Assumptions 1 - 7 hold. Then, for $k_1, k_2 = 1, 2$ with $k_1 \neq k_2$, and for any given $i = 1, \ldots, N$, as $N, T \to \infty$,*

$$\sqrt{T}\left\{\widehat{\mathbf{b}}_{k_1 i} - \widehat{\mathbf{I}}_{\widehat{\xi}k_1}\widehat{\mathbf{H}}'\mathbf{b}_{k_1 i} - \left(\mathbf{I}_{r_1+r_2} - \widehat{\mathbf{I}}_{\widehat{\xi}k_1}\right)\widehat{\mathbf{H}}'\mathbf{b}_{k_2 i}\right\} \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \mathbf{\Sigma}_{\widehat{\mathbf{b}}k_1 i}\right),$$

*where the $(r_1+r_2)\times(r_1+r_2)$ matrix $\widehat{\mathbf{I}}_{\widehat{\xi}k_1}$ is defined as*

$$\widehat{\mathbf{I}}_{\widehat{\xi}k_1} = \left(\sum_{t=1}^{T}\widehat{\xi}_{k_1,t|T}\,\mathbb{I}(s_t = k_1)\widehat{\mathbf{g}}_t\widehat{\mathbf{g}}_t'\right)\left(\sum_{t=1}^{T}\widehat{\xi}_{k_1,t|T}\,\widehat{\mathbf{g}}_t\widehat{\mathbf{g}}_t'\right)^{-1}, \tag{49}$$

*and where*

$$\mathbf{\Sigma}_{\widehat{\mathbf{b}}k_1 i} = \left(\mathbf{Q}_1'\mathbf{\Sigma}_{\mathbf{f}1}^{(k_1)}\mathbf{Q}_1 + \mathbf{Q}_2'\mathbf{\Sigma}_{\mathbf{f}2}^{(k_1)}\mathbf{Q}_2\right)^{-1}\left(\mathbf{Q}_1'\mathbf{\Gamma}_{1k_1 i}\mathbf{Q}_1 + \mathbf{Q}_2'\mathbf{\Gamma}_{2k_1 i}\mathbf{Q}_2\right)\left(\mathbf{Q}_1'\mathbf{\Sigma}_{\mathbf{f}1}^{(k_1)}\mathbf{Q}_1 + \mathbf{Q}_2'\mathbf{\Sigma}_{\mathbf{f}2}^{(k_1)}\mathbf{Q}_2\right)^{-1},$$

*with $\mathbf{Q}_j$, $\mathbf{\Gamma}_{jk_1 i}$, and $\mathbf{\Sigma}_{\mathbf{f}j}^{(k_1)}$, $j = 1, 2$, defined in (47), Assumption 6(c), and Assumption 1 when $h_{k_1} = \widehat{\xi}_{k_1,t|T}$, respectively.*

Theorem 1 shows that the estimator $\widehat{\mathbf{b}}_{k_1 i}$ for $\mathbf{b}_{k_1 i}$ is subject to two sources of bias. The first is standard and it is induced by the usual indeterminacy due to the latency of both factors and loadings and captured by the invertible matrix $\widehat{\mathbf{H}}$ defined in (45), see also Bai (2003). If we assumed $T^{-1}\sum_{t=1}^{T}\mathbf{g}_t\mathbf{g}_t' = \mathbf{I}_{r_1+r_2}$, then it would straightforward to see that $\widehat{\mathbf{H}}$ becomes just a rotation, i.e., an orthogonal matrix. However, additional restriction on the loadings would be necessary to reduce $\widehat{\mathbf{H}}$ to be the identity—for a discussion on identification of factors, see, e.g., (Bai and Ng, 2013). The second source of bias is induced by $\widehat{\mathbf{I}}_{\widehat{\xi}k_1}$, which depends on the probability of the state being asymptotically correctly estimated. If the unconditional probability of being in state $k_1$ is correctly estimated with probability one, that is, if $\widehat{\xi}_{k_1,t|T} \xrightarrow{p} \mathbb{I}(s_t = k_1)$, as $N, T \to \infty$, then $\widehat{\mathbf{I}}_{\widehat{\xi}k_1} \xrightarrow{p} \mathbf{I}_{r_1+r_2}$ and $\widehat{\mathbf{b}}_{k_1 i}$ will estimate consistently a linear transformation of $\mathbf{b}_{k_1 i}$. Otherwise $\widehat{\mathbf{b}}_{k_1 i}$ estimates a linear combination of linear transformations of $\mathbf{b}_{k_1 i}$ and $\mathbf{b}_{k_2 i}$, with weights determined by $\widehat{\mathbf{I}}_{\widehat{\xi}k_1}$ and $(\mathbf{I}_{r_1+r_2} - \widehat{\mathbf{I}}_{\widehat{\xi}k_1})$, respectively.

To further aid to the understanding of Theorem 1, let $\widehat{\mathbf{R}}_k = \widehat{\mathbf{H}}\widehat{\mathbf{I}}_{\widehat{\xi}k}'$, for $k = 1, 2$, and consider the partition

$$\widehat{\mathbf{R}}_k = \begin{bmatrix} \widehat{\mathbf{R}}_{k,11} & \widehat{\mathbf{R}}_{k,12} \\ \widehat{\mathbf{R}}_{k,21} & \widehat{\mathbf{R}}_{k,22} \end{bmatrix}, \quad \widehat{\mathbf{H}} = \begin{bmatrix} \widehat{\mathbf{H}}_{11} & \widehat{\mathbf{H}}_{12} \\ \widehat{\mathbf{H}}_{21} & \widehat{\mathbf{H}}_{22} \end{bmatrix},$$

where $\widehat{\mathbf{R}}_{k,j\ell}$, $k, j, \ell = 1, 2$ and $\widehat{\mathbf{H}}_{j\ell}$, $j, \ell = 1, 2$, are $r_j \times r_\ell$. From Theorem 1, for any given

$i = 1, \dots, N$, as $N, T \to \infty$, we obtain:

$$\begin{aligned}&\sqrt{T}\left\{\widehat{\mathbf{b}}_{1i}' - [\boldsymbol{\lambda}_{1i}'\ \mathbf{0}]\,\widehat{\mathbf{R}}_1 - [\mathbf{0}\ \boldsymbol{\lambda}_{2i}']\left(\widehat{\mathbf{H}} - \widehat{\mathbf{R}}_1\right)\right\}\\ &= \sqrt{T}\left\{\widehat{\mathbf{b}}_{1i}' - \boldsymbol{\lambda}_{1i}'[\widehat{\mathbf{R}}_{1,11}\ \widehat{\mathbf{R}}_{1,12}] - \boldsymbol{\lambda}_{2i}'\left[\left(\widehat{\mathbf{H}}_{21} - \widehat{\mathbf{R}}_{1,21}\right)\ \left(\widehat{\mathbf{H}}_{22} - \widehat{\mathbf{R}}_{1,22}\right)\right]\right\} \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}_{\widehat{\mathbf{b}}1i}\right), \quad (50)\end{aligned}$$

and likewise

$$\begin{aligned}&\sqrt{T}\left\{\widehat{\mathbf{b}}_{2i}' - [\mathbf{0}\ \boldsymbol{\lambda}_{2i}']\,\widehat{\mathbf{R}}_2 - [\boldsymbol{\lambda}_{1i}'\ \mathbf{0}]\left(\widehat{\mathbf{H}} - \widehat{\mathbf{R}}_2\right)\right\}\\ &= \sqrt{T}\left\{\widehat{\mathbf{b}}_{2i}' - \boldsymbol{\lambda}_{2i}'[\widehat{\mathbf{R}}_{2,21}\ \widehat{\mathbf{R}}_{2,22}] - \boldsymbol{\lambda}_{1i}'\left[\left(\widehat{\mathbf{H}}_{11} - \widehat{\mathbf{R}}_{2,11}\right)\ \left(\widehat{\mathbf{H}}_{12} - \widehat{\mathbf{R}}_{2,12}\right)\right]\right\} \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}_{\widehat{\mathbf{b}}2i}\right). \quad (51)\end{aligned}$$

From (50) and (51), if we consistently estimate the unconditional probability of being in a given state $j = 1, 2$, then $\widehat{\mathbf{R}}_k \xrightarrow{p} \widehat{\mathbf{H}}$ as $N, T \to \infty$. The columns of $\widehat{\mathbf{B}}_j$ then estimate two different linear transformations of the columns of $\boldsymbol{\Lambda}_j$. If $r_1 = r_2$, both linear transformations are invertible, and there is no need to know the true values of $r_1$ and $r_2$ to get consistent estimates of the space spanned by the true loadings in the two different regimes. In particular, we can consider either the first or the second half of the columns of $\widehat{\mathbf{B}}_j$ as an estimator of a linear transformation of $\boldsymbol{\Lambda}_j$, for $j = 1, 2$.

On the other hand, if $r_1 \neq r_2$, then the first $r_1$ columns of $\widehat{\mathbf{B}}_1$, and the last $r_2$ columns of $\widehat{\mathbf{B}}_2$, estimate an invertible linear transformation of the columns of $\boldsymbol{\Lambda}_1$ and $\boldsymbol{\Lambda}_2$, respectively. However, the last $r_2$ columns of $\widehat{\mathbf{B}}_1$, and the first $r_1$ columns of $\widehat{\mathbf{B}}_2$, estimate $\boldsymbol{\Lambda}_1\widehat{\mathbf{R}}_{1,12}$ and $\boldsymbol{\Lambda}_2\widehat{\mathbf{R}}_{2,21}$, respectively, none of which is invertible. In this case, we need consistent estimators of $r_1$ and $r_2$ in order to be able to isolate the first $r_1$ columns of $\widehat{\mathbf{B}}_1$ and the last $r_2$ columns of $\widehat{\mathbf{B}}_2$, respectively, which are consistent estimators of a linear transformation of the columns of $\boldsymbol{\Lambda}_1$ and $\boldsymbol{\Lambda}_2$, respectively. Therefore, if we only know that $r_1 \neq r_2$ without knowing their true values, then we can consistently estimate a linear transformation of the columns of $\mathbf{B}_j$, but nothing can be said about $\boldsymbol{\Lambda}_j$, $j = 1, 2$.

Theorem 1 describes the asymptotic properties of the estimator for the factor loadings $\widehat{\mathbf{B}}_1$ and $\widehat{\mathbf{B}}_2$. Complementary results can be obtained with respect to the estimated factors associated to the loading matrices $\widehat{\mathbf{B}}_1$ and $\widehat{\mathbf{B}}_2$. Formally, the true factors that correspond to $\mathbf{B}_1$ and $\mathbf{B}_2$ are $\xi_{1t}\mathbf{g}_t$ and $\xi_{2t}\mathbf{g}_t$, respectively, and their estimators are $\widehat{\xi}_{1,t|T}\,\widehat{\mathbf{g}}_t$ and $\widehat{\xi}_{2,t|T}\,\widehat{\mathbf{g}}_t$, respectively. The following theorem states the asymptotic distribution of those factors.

**Theorem 2.** *Let Assumptions 1 - 7 hold. Then, for any given $t = 1, \dots, T$, as $N, T \to \infty$,*

$$\sqrt{N}\left\{\begin{pmatrix} \widehat{\xi}_{1,t|T}\,\widehat{\mathbf{g}}_t \\ \widehat{\xi}_{2,t|T}\,\widehat{\mathbf{g}}_t \end{pmatrix} - \widehat{\mathbf{H}}_{\xi}^{-1}\begin{pmatrix} \xi_{1t}\mathbf{g}_t \\ \xi_{2t}\mathbf{g}_t \end{pmatrix}\right\} \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}_{\widehat{\xi}\otimes\widehat{\mathbf{g}},t}\right),$$

*where*

$$\widehat{\mathbf{H}}_\xi = \left[\begin{array}{cc} \widehat{\mathbf{H}}\widehat{\mathbf{I}}_{\widehat{\xi}1}' & \widehat{\mathbf{H}}\left(\mathbf{I}_{r_1+r_2} - \widehat{\mathbf{I}}_{\widehat{\xi}2}\right)' \\ \widehat{\mathbf{H}}\left(\mathbf{I}_{r_1+r_2} - \widehat{\mathbf{I}}_{\widehat{\xi}1}\right)' & \widehat{\mathbf{H}}\widehat{\mathbf{I}}_{\widehat{\xi}2}' \end{array}\right],$$

*with* $\widehat{\mathbf{H}}$ *and* $\widehat{\mathbf{I}}_{\widehat{\xi}j}$ *defined in* (45) *and* (49)*, respectively, and where*

$$\boldsymbol{\Sigma}_{\widehat{\xi}\otimes\widehat{\mathbf{g}},t} = \left\{\mathbf{H}_\xi \left(\begin{array}{cc} \boldsymbol{\Sigma}_{\mathbf{B}1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{\mathbf{B}2} \end{array}\right) \mathbf{H}_\xi'\right\}^{-1} \left(\mathbf{H}_\xi \boldsymbol{\Sigma}_{\mathbf{Be}t} \mathbf{H}_\xi'\right) \left\{\mathbf{H}_\xi \left(\begin{array}{cc} \boldsymbol{\Sigma}_{\mathbf{B}1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{\mathbf{B}2} \end{array}\right) \mathbf{H}_\xi'\right\}^{-1},$$

*where* $\boldsymbol{\Sigma}_{\mathbf{B}j}$*,* $j = 1,2$*, are defined in* (44)*,*

$$\boldsymbol{\Sigma}_{\mathbf{Be}t} = \left(\begin{array}{cccc} \boldsymbol{\Phi}_{1t} & \mathbf{0} & \mathbf{0} & \boldsymbol{\Phi}_{12,t} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \boldsymbol{\Phi}_{12,t}' & \mathbf{0} & \mathbf{0} & \boldsymbol{\Phi}_{2t} \end{array}\right),$$

*with* $\boldsymbol{\Phi}_{jt}$*,* $j = 1,2$*, defined in Assumption 6(d) and* $\boldsymbol{\Phi}_{12,t} = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^N \sum_{l=1}^N \boldsymbol{\lambda}_{1i}\boldsymbol{\lambda}_{2l}' \mathsf{E}[e_{it}e_{lt}]$*, and where*

$$\mathbf{H}_\xi = \left[\begin{array}{cc} \mathbf{H}\mathbf{I}_{\xi1}' & \mathbf{H}\left(\mathbf{I}_{r_1+r_2} - \mathbf{I}_{\xi2}\right)' \\ \mathbf{H}\left(\mathbf{I}_{r_1+r_2} - \mathbf{I}_{\xi1}\right)' & \mathbf{H}\mathbf{I}_{\xi2}' \end{array}\right],$$

*with* $\mathbf{H}$ *defined in* (48) *and* $\mathbf{I}_{\xi j} = p\lim_{N,T\to\infty} \widehat{\mathbf{I}}_{\widehat{\xi}j}$.

Now, in general $\widehat{\mathbf{I}}_{\widehat{\xi}j} \neq \mathbf{I}_{r_1+r_2}$ and so also $\mathbf{I}_{\xi j} \neq \mathbf{I}_{r_1+r_2}$. Then, because of Theorem 1, the estimator $\widehat{\mathbf{b}}_{ji}$ is biased and it is straightforward to see that the asymptotic covariance in Theorem 2 is positive definite. Note that if we know $r_1 = r_2$ holds, then we can build consistent estimators $\widehat{\mathbf{f}}_{1t}$ and $\widehat{\mathbf{f}}_{2t}$ for linear combinations of $\mathbf{f}_{1t}$ and $\mathbf{f}_{2t}$, respectively, by simply regressing $\mathbf{x}_t$ onto the first $r_1$ columns of $\widehat{\mathbf{B}}_1$ and the last $r_2$ columns of $\widehat{\mathbf{B}}_2$, respectively: as previously discussed, these define consistent estimators $\widehat{\boldsymbol{\Lambda}}_1$ and $\widehat{\boldsymbol{\Lambda}}_2$ for linear transformation of $\boldsymbol{\Lambda}_1$ and $\boldsymbol{\Lambda}_2$, respectively. Formally, this means we can build the sequence of factor estimators

$$\widehat{\mathbf{f}}_{jt} = \frac{1}{N}\widehat{\xi}_{j,t|T}\widehat{\boldsymbol{\Lambda}}_j'\mathbf{x}_t, \qquad j = 1,2. \tag{52}$$

If the unconditional probability of being in a given state is correctly estimated then $\widehat{\mathbf{I}}_{\widehat{\xi}j} \xrightarrow{p} \mathbf{I}_{r_1+r_2}$ as $N,T\to\infty$, and Theorem 2 is redundant: in this case, asymptotic normality of (52) follows from arguments analogous to those in Bai (2003).

# 5 Unobserved heterogeneity

The model in (1) assumes no individual effects. These can be introduced by considering

$$\mathbf{x}_t = (\boldsymbol{\alpha}_1 + \boldsymbol{\Lambda}_1 \mathbf{f}_{1t})\,\mathbb{I}(s_t = 1) + (\boldsymbol{\alpha}_2 + \boldsymbol{\Lambda}_2 \mathbf{f}_{2t})\,\mathbb{I}(s_t = 2) + \mathbf{e}_t, \tag{53}$$

where $\boldsymbol{\alpha}_j = (\alpha_{j1}, \ldots, \alpha_{jN})'$, for $j = 1, 2$, and $\alpha_{ji}$ captures the individual effect of cross-sectional unit $i$ within regime $j$. The vectors $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_2$ introduce unobserved heterogeneity. If the state variable driving the regimes were observable, the resulting identification problem could be solved by expressing the model in terms of deviations of $\mathbf{x}_t$ from the conditional means within each regime: on this, see Massacci et al. (2021). However, since the state variable $s_t$ in (53) is latent, this strategy no longer is applicable since the state is not observable with probability one. For this reason, we express the model in terms of the deviation of $\mathbf{x}_t$ from the unconditional mean.

Formally, consider the $N \times 1$ vector of dependent variables $\mathbf{y}_t$ defined as

$$\begin{aligned}
\mathbf{y}_t &= \mathbf{x}_t - \mathsf{E}(\mathbf{x}_t) \\
&= (\boldsymbol{\alpha}_1 + \boldsymbol{\Lambda}_1 \mathbf{f}_{1t})\,\mathbb{I}(s_t = 1) + (\boldsymbol{\alpha}_2 + \boldsymbol{\Lambda}_2 \mathbf{f}_{2t})\,\mathbb{I}(s_t = 2) + \mathbf{e}_t \\
&\quad - \mathsf{E}\left[(\boldsymbol{\alpha}_1 + \boldsymbol{\Lambda}_1 \mathbf{f}_{1t})\,\mathbb{I}(s_t = 1)\right] + \mathsf{E}\left[(\boldsymbol{\alpha}_2 + \boldsymbol{\Lambda}_2 \mathbf{f}_{2t})\,\mathbb{I}(s_t = 2)\right] \\
&= \boldsymbol{\alpha}_1 \varphi_{1t} + \boldsymbol{\Lambda}_1 \mathbf{f}_{1t} \mathbb{I}(s_t = 1) + \boldsymbol{\alpha}_2 \varphi_{2t} + \boldsymbol{\Lambda}_2 \mathbf{f}_{2t} \mathbb{I}(s_t = 2) + \mathbf{e}_t,
\end{aligned}$$

where

$$\varphi_{jt} = \mathbb{I}(s_t = j) - \mathsf{E}\left[\mathbb{I}(s_t = j)\right], \quad j = 1, 2.$$

If $\boldsymbol{\alpha}_1 = \boldsymbol{\alpha}_2$, $\mathbf{x}_t$ has the same expected value in both regimes, and $\mathbf{y}_t = \boldsymbol{\Lambda}_1 \mathbf{f}_{1t} \mathbb{I}(s_t = 1) + \boldsymbol{\Lambda}_2 \mathbf{f}_{2t} \mathbb{I}(s_t = 2) + \mathbf{e}_t$. In the more general case in which $\boldsymbol{\alpha}_1 \neq \boldsymbol{\alpha}_2$, unconditional demeaning leads to a larger factor space of dimension $r_1 + r_2 + 2$. The additional two factors $\varphi_{1t}$ and $\varphi_{2t}$ take only two values, namely $\varphi_{jt} = -\mathsf{E}[\mathbb{I}(s_t = j)]$ and $\varphi_{jt} = 1 - \mathsf{E}[\mathbb{I}(s_t = j)]$, depending on whether $\mathbb{I}(s_t = j) = 0$ or $\mathbb{I}(s_t = j) = 1$, respectively, for $j = 1, 2$. In this case, the equivalent linear representation in (10) holds with $\mathbf{g}_t = [\varphi_{1t}, \mathbb{I}(s_t = 1)\,\mathbf{f}'_{1t}, \varphi_{2t}, \mathbb{I}(s_t = 2)\,\mathbf{f}'_{2t}]$ and $\mathbf{A} = [\boldsymbol{\alpha}_1, \boldsymbol{\Lambda}_1, \boldsymbol{\alpha}_2, \boldsymbol{\Lambda}_2]$. The measurement equation in (9) of the state space representation remains valid with $\mathbf{B}_1 = [\boldsymbol{\alpha}_1, \boldsymbol{\Lambda}_1, \boldsymbol{\alpha}_2, \mathbf{0}]$ and $\mathbf{B}_2 = [\boldsymbol{\alpha}_1, \mathbf{0}, \boldsymbol{\alpha}_2, \boldsymbol{\Lambda}_2]$. Therefore, the tools developed in this paper can be applied to the sample counterpart of $\mathbf{y}_t$, namely to $\widehat{\mathbf{y}}_t = \mathbf{x}_t - \left(T^{-1} \sum_{t=1}^{T} \mathbf{x}_t\right)$, which consistently estimates $\mathbf{y}_t$ as $N \to \infty$.

# 6 Monte Carlo

Throughout we set $N = \{100, 200\}$ and $T = \{250, 500, 750, 1000\}$ and we simulate the $N \times T$ matrix of data where at each $t = 1, \ldots, T$ the data $\mathbf{x}_t$ follows (1). This requires to simulate the latent state $\boldsymbol{\xi}_t$, the loadings $\boldsymbol{\Lambda}_1$ and $\boldsymbol{\Lambda}_2$, the factors $\mathbf{f}_{1t}$ and $\mathbf{f}_{2t}$, and the idiosyncratic components $\mathbf{e}_t$.

We simulate the latent state $\boldsymbol{\xi}_t$ according to (5), where $\mathbf{P}$ has entries $p_{11} = 0.9$ and $p_{22} = 0.7$ so that $p_{12} = 0.1$ and $p_{21} = 0.3$. This configuration corresponds to the unconditional probabilities to be equal to $\mathsf{P}(s_t = 1) = \mathsf{E}[\xi_{1t}] = \frac{1-p_{22}}{2-p_{11}-p_{22}} = 0.75$ and $\mathsf{P}(s_t = 2) = \mathsf{E}[\xi_{2t}] = \frac{1-p_{11}}{2-p_{11}-p_{22}} = 0.25$. Then, we generate the innovations $\mathbf{v}_t$ of the VAR as in (5) as follows: at each $t$ we generate $u_t \sim U[0,1]$ and (i) if $\xi_{1,t-1} = 1$ and $u_t \le p_{11}$ then $\mathbf{v}_t = [1\ 0]' - \mathbf{P}'\boldsymbol{\xi}_{t-1}$; (ii) if $\xi_{1,t-1} = 1$ and $u_t > p_{11}$ then $\mathbf{v}_t = [0\ 1]' - \mathbf{P}'\boldsymbol{\xi}_{t-1}$; (iii) if $\xi_{1,t-1} = 0$ and $u_t \le p_{21}$ then $\mathbf{v}_t = [1\ 0]' - \mathbf{P}'\boldsymbol{\xi}_{t-1}$; (iv) if $\xi_{1,t-1} = 0$ and $u_t > p_{21}$ then $\mathbf{v}_t = [0\ 1]' - \mathbf{P}'\boldsymbol{\xi}_{t-1}$.

We set the number of factors in each state to $r_j = r = \{1,2\}$, $j = 1,2$. Then the common component is generated according to (1). Specifically, letting $\chi_{it} = \boldsymbol{\lambda}_{1i}' f_{1t}\mathbb{I}(s_t = 1) + \boldsymbol{\lambda}_{2i}' f_{2t}\mathbb{I}(s_t = 1)$, $i = 1,\dots,N$, the $r$ entries of $\boldsymbol{\lambda}_{1i}$ and $\boldsymbol{\lambda}_{2i}$ are generated from a $N(1,1)$ distribution and the matrices $\boldsymbol{\Lambda}_1$ and $\boldsymbol{\Lambda}_2$ are rotated in such a way that $\boldsymbol{\Lambda}_1'\boldsymbol{\Lambda}_1$ and $\boldsymbol{\Lambda}_2'\boldsymbol{\Lambda}_2$ are diagonal matrices. The factors are such that $\mathbf{f}_{jt} = \mathbf{f}_t$, $j = 1,2$, such that $T^{-1}\sum_{t=1}^T \mathbf{f}_t\mathbf{f}_t' = \mathbf{I}_r$ and each component of $\mathbf{f}_t$ is such that $f_{it} = \rho_f f_{i,t-1} + z_{it}$, $i = 1,\dots,r$, with $\rho_f = \{0, 0.7\}$ and $z_{it} \sim N(0,1)$.

The idiosyncratic components are generated according to (2), where $\boldsymbol{\Sigma}_{je} = \boldsymbol{\Sigma}_{je,a} + \boldsymbol{\Sigma}_{je,b}$, $j = 1,2$, with $\boldsymbol{\Sigma}_{je,a}$ diagonal and $\boldsymbol{\Sigma}_{je,b}$ banded. Specifically, the entries of $\boldsymbol{\Sigma}_{1e,a}$ are generated from a $U[0.25, 1.25]$ and those of $\boldsymbol{\Sigma}_{2e,a}$ are generated from a $U[0.75, 1.75]$, while $\boldsymbol{\Sigma}_{1e,b}$ is a Toeplitz matrix with $\tau^k$ on the $k$th diagonal for $k = 1,2$ and zero elsewhere, and, finally $\boldsymbol{\Sigma}_{2e,b}$ is a Toeplitz matrix with $\tau^{k-1}$ on the $k$th diagonal for $k = 1,2,3$ and zero elsewhere. We set $\tau = \{0, 0.5\}$. Moreover, each component of $\boldsymbol{\nu}_t$ is such that $\nu_{it} = \rho_i \nu_{i,t-1} + \omega_{it}$, $i = 1,\dots,N$, with $\rho_i = \{0, \rho\}$ and $\rho \sim U[0, 0.5]$. Finally, we set the noise-to-signal ratio $N^{-1}\sum_{i=1}^N \frac{\sum_{t=1}^T e_{it}^2}{\sum_{t=1}^T \chi_{it}^2}$ to be 0.5 on average across all $N$ simulated time series.

We simulate the model above 100 times for different values of $r$, $\rho_f$, $\tau$, and $\rho$. The EM is run allowing for at most 100 iterations and using a convergence threshold equal to $10^{-6}$. We initialize the algorithm using PCA as described in Section 3.4. Since the states are identified only up to a permutation at each iteration of the algorithm we assign label 1 to the state with highest unconditional probability.

In Tables 1-4 first four columns, we report the average and standard deviation over all replications of the estimated diagonal entries of the transition matrix $\widehat{p}_{jj}$, $j = 1,2$, of the unconditional probabilities $\mathsf{P}(s_t = j)$, estimated as $\widehat{\overline{\xi}}_{j,t|T} = T^{-1}\sum_{t=1}^T \widehat{\xi}_{j,t|T}$, $j = 1,2$.

Since the loadings are not identified, in the fifth column of Tables 1-4 we report the multiple $R^2$ coefficient obtained from regressing the columns of $\widehat{\mathbf{B}}_1$ onto the columns of $\mathbf{B}_1^* = \mathbf{B}_1\widehat{\mathbf{I}}_{\widehat{\xi}1} + \mathbf{B}_2(\mathbf{I}_{2r} - \widehat{\mathbf{I}}_{\widehat{\xi}1})$, thus correcting for the possible bias as prescribed by Theorem 1. Namely, we compute

$$R^2_{B^*} = \frac{\operatorname{tr}\left\{\left(\mathbf{B}_1^{*\prime}\widehat{\mathbf{B}}_1\right)\left(\widehat{\mathbf{B}}_1'\widehat{\mathbf{B}}_1\right)^{-1}\left(\widehat{\mathbf{B}}_1'\mathbf{B}_1^*\right)\right\}}{\operatorname{tr}\left(\widehat{\mathbf{B}}_1^{*\prime}\widehat{\mathbf{B}}_1^*\right)}.$$

**Table 1:** ESTIMATED PROBABILITIES - $r = 1,\ \rho_f = 0,\ \tau = 0,\ \rho = 0.$

| $T$ | $N$ | $\widehat{p}_{11}$ | $\widehat{p}_{22}$ | $\overline{\widehat{\xi}}_{1,t\|T}$ | $\overline{\widehat{\xi}}_{2,t\|T}$ | $R^2_{B^*}$ | MSE($\chi$) | avg. iter |
|---|---|---|---|---|---|---|---|---|
| 250 | 100 | 0.89 | 0.64 | 0.76 | 0.24 | 0.97 | 0.02 | 13.78 |
| | | (0.03) | (0.13) | (0.06) | (0.06) | | | |
| 500 | 100 | 0.90 | 0.68 | 0.76 | 0.24 | 0.98 | 0.01 | 12.55 |
| | | (0.01) | (0.04) | (0.03) | (0.03) | | | |
| 750 | 100 | 0.90 | 0.69 | 0.75 | 0.25 | 0.98 | 0.01 | 12.71 |
| | | (0.01) | (0.03) | (0.03) | (0.03) | | | |
| 1000 | 100 | 0.90 | 0.69 | 0.75 | 0.25 | 0.98 | 0.01 | 12.05 |
| | | (0.01) | (0.03) | (0.03) | (0.03) | | | |
| 250 | 200 | 0.89 | 0.64 | 0.76 | 0.24 | 0.97 | 0.01 | 11.98 |
| | | (0.02) | (0.11) | (0.06) | (0.06) | | | |
| 500 | 200 | 0.89 | 0.68 | 0.75 | 0.25 | 0.97 | 0.01 | 21.23 |
| | | (0.02) | (0.04) | (0.03) | (0.03) | | | |
| 750 | 200 | 0.89 | 0.68 | 0.75 | 0.25 | 0.97 | 0.02 | 37.37 |
| | | (0.02) | (0.04) | (0.03) | (0.03) | | | |
| 1000 | 200 | 0.90 | 0.69 | 0.75 | 0.25 | 0.98 | 0.02 | 36.22 |
| | | (0.01) | (0.03) | (0.03) | (0.03) | | | |

**Table 2:** ESTIMATED PROBABILITIES - $r = 1,\ \rho_f = 0.7,\ \tau = 0.5,\ \rho = 0.5.$

| $T$ | $N$ | $\widehat{p}_{11}$ | $\widehat{p}_{22}$ | $\overline{\widehat{\xi}}_{1,t\|T}$ | $\overline{\widehat{\xi}}_{2,t\|T}$ | $R^2_{B^*}$ | MSE($\chi$) | avg. iter |
|---|---|---|---|---|---|---|---|---|
| 250 | 100 | 0.89 | 0.62 | 0.77 | 0.23 | 0.97 | 0.02 | 20.14 |
| | | (0.03) | (0.17) | (0.07) | (0.07) | | | |
| 500 | 100 | 0.90 | 0.68 | 0.76 | 0.24 | 0.98 | 0.02 | 15.28 |
| | | (0.02) | (0.05) | (0.04) | (0.04) | | | |
| 750 | 100 | 0.90 | 0.69 | 0.76 | 0.24 | 0.98 | 0.02 | 14.43 |
| | | (0.01) | (0.03) | (0.03) | (0.03) | | | |
| 1000 | 100 | 0.90 | 0.66 | 0.77 | 0.23 | 0.98 | 0.01 | 14.07 |
| | | (0.02) | (0.14) | (0.05) | (0.05) | | | |
| 250 | 200 | 0.89 | 0.62 | 0.77 | 0.23 | 0.98 | 0.02 | 11.95 |
| | | (0.03) | (0.14) | (0.07) | (0.07) | | | |
| 500 | 200 | 0.89 | 0.67 | 0.75 | 0.25 | 0.98 | 0.01 | 20.21 |
| | | (0.02) | (0.04) | (0.04) | (0.04) | | | |
| 750 | 200 | 0.89 | 0.69 | 0.75 | 0.25 | 0.98 | 0.01 | 19.17 |
| | | (0.01) | (0.04) | (0.02) | (0.02) | | | |
| 1000 | 200 | 0.90 | 0.69 | 0.75 | 0.25 | 0.98 | 0.01 | 21.82 |
| | | (0.01) | (0.03) | (0.03) | (0.03) | | | |

The closer this number is to one, the closer is the space spanned by the columns of $\widehat{\mathbf{B}}_1$ to the space spanned by the columns of $\mathbf{B}_1^*$ (see Doz et al. (2012)).

In the sixth column of Tables 1-4 we report the MSE of the estimated common components

**Table 3:** ESTIMATED PROBABILITIES - $r = 2,\ \rho_f = 0,\ \tau = 0,\ \rho = 0.$

| $T$ | $N$ | $\widehat{p}_{11}$ | $\widehat{p}_{22}$ | $\widehat{\xi}_{t\|T,1}$ | $\widehat{\xi}_{t\|T,2}$ | $R^2_{B^*}$ | MSE($\chi$) | avg. iter |
|---|---|---|---|---|---|---|---|---|
| 250 | 100 | 0.88 | 0.46 | 0.81 | 0.19 | 0.97 | 0.04 | 19.32 |
| | | (0.04) | (0.22) | (0.08) | (0.08) | | | |
| 500 | 100 | 0.89 | 0.65 | 0.76 | 0.24 | 0.97 | 0.03 | 14.63 |
| | | (0.02) | (0.04) | (0.03) | (0.03) | | | |
| 750 | 100 | 0.90 | 0.67 | 0.76 | 0.24 | 0.97 | 0.03 | 14.46 |
| | | (0.01) | (0.04) | (0.03) | (0.03) | | | |
| 1000 | 100 | 0.90 | 0.68 | 0.76 | 0.24 | 0.97 | 0.03 | 13.83 |
| | | (0.01) | (0.03) | (0.02) | (0.02) | | | |
| 250 | 200 | 0.87 | 0.48 | 0.78 | 0.22 | 0.97 | 0.03 | 13.72 |
| | | (0.04) | (0.22) | (0.08) | (0.08) | | | |
| 500 | 200 | 0.89 | 0.65 | 0.75 | 0.25 | 0.97 | 0.02 | 10.40 |
| | | (0.02) | (0.05) | (0.04) | (0.04) | | | |
| 750 | 200 | 0.89 | 0.67 | 0.75 | 0.25 | 0.97 | 0.02 | 10.86 |
| | | (0.01) | (0.04) | (0.03) | (0.03) | | | |
| 1000 | 200 | 0.90 | 0.68 | 0.75 | 0.25 | 0.97 | 0.01 | 10.81 |
| | | (0.01) | (0.03) | (0.02) | (0.02) | | | |

**Table 4:** ESTIMATED PROBABILITIES - $r = 2,\ \rho_f = 0.7,\ \tau = 0.5,\ \rho = 0.5.$

| $T$ | $N$ | $\widehat{p}_{11}$ | $\widehat{p}_{22}$ | $\widehat{\xi}_{t\|T,1}$ | $\widehat{\xi}_{t\|T,2}$ | $R^2_{B^*}$ | MSE($\chi$) | avg. iter |
|---|---|---|---|---|---|---|---|---|
| 250 | 100 | 0.91 | 0.38 | 0.86 | 0.14 | 0.98 | 0.04 | 17.40 |
| | | (0.03) | (0.20) | (0.07) | (0.07) | | | |
| 500 | 100 | 0.90 | 0.65 | 0.77 | 0.23 | 0.97 | 0.03 | 20.36 |
| | | (0.02) | (0.04) | (0.04) | (0.04) | | | |
| 750 | 100 | 0.90 | 0.67 | 0.76 | 0.24 | 0.97 | 0.03 | 17.20 |
| | | (0.01) | (0.04) | (0.03) | (0.03) | | | |
| 1000 | 100 | 0.90 | 0.68 | 0.76 | 0.24 | 0.98 | 0.03 | 16.61 |
| | | (0.01) | (0.03) | (0.03) | (0.03) | | | |
| 250 | 200 | 0.89 | 0.41 | 0.83 | 0.17 | 0.97 | 0.03 | 14.55 |
| | | (0.04) | (0.21) | (0.09) | (0.09) | | | |
| 500 | 200 | 0.89 | 0.66 | 0.76 | 0.24 | 0.97 | 0.02 | 13.41 |
| | | (0.01) | (0.06) | (0.04) | (0.04) | | | |
| 750 | 200 | 0.90 | 0.67 | 0.76 | 0.24 | 0.97 | 0.02 | 14.56 |
| | | (0.01) | (0.03) | (0.03) | (0.03) | | | |
| 1000 | 200 | 0.90 | 0.68 | 0.76 | 0.24 | 0.98 | 0.02 | 11.96 |
| | | (0.01) | (0.03) | (0.02) | (0.02) | | | |

defined as

$$\text{MSE}(\chi) = \frac{\sum_{i=1}^N \sum_{t=1}^T (\widehat{\chi}_{it} - \chi_{it})^2}{\sum_{i=1}^N \sum_{t=1}^T \chi_{it}^2},$$

where $\widehat{\chi}_{it} = \left(\widehat{\mathbf{b}}_{1i}\ \widehat{\mathbf{b}}_{2i}\right)' \left(\widehat{\boldsymbol{\xi}}_t \otimes \widehat{\mathbf{g}}_t\right)$.

In the last column of Tables 1-4 we report the average number of iterations needed to get convergence of the EM algorithm.

# 7 Empirical analysis

In this section we show how the methodological framework we propose can be used to model a large set of stock returns. This relates our work to a vast literature that models stock return dynamics using Markov switching specifications. Perez-Quiros and Timmermann (2000, 2001) document business cycle asymmetries in U.S. stock return dynamics using decile-sorted portfolios. Ang and Bekaert (2002), and Guidolin and Timmermann (2008), study portfolio allocation in international equity markets under regime switching. In a multi asset setting, Guidolin and Timmermann (2006) describe the joint distribution of equity and bonds under regime switching. Guidolin (2011), and Ang and Timmermann (2012), provide a review of the literature. We contribute to this stream of literature by characterizing stock return dynamics with a Markov switching model in a large dimensional setting. To the very best of our knowledge, we are the first to do so. In what follows, Section 7.1 describes the data and the empirical model specification, Section 7.2 discusses the estimated regime probabilities, and Section 7.3 presents the findings for estimated loadings and factors.

## 7.1 Data and model specification

The vector of observable dependent variables $\mathbf{x}_t$ in (1) is made of the monthly value weighted returns in excess of the risk-free rate from the $N = 49$ industry portfolios kindly made publicly available on Kenneth French website.[3] In order to obtain a balanced panel, the sample period runs from July 1969 through December 2021, a total of $T = 630$ time periods.

## 7.2 Regime probabilities

Using the selection criterion of Ahn and Horenstein (2013) as applied to the equivalent linear representation in (10), we find $r = 2$ common factors. Based on this, we apply the algorithm detailed in Section 3. The EM algorithm converges relatively fast in just 22 iterations (see Figure 1). The realisation of the estimator $\widehat{\mathbf{P}}$ for the matrix of conditional probabilities $\mathbf{P}$ in (3) is equal to

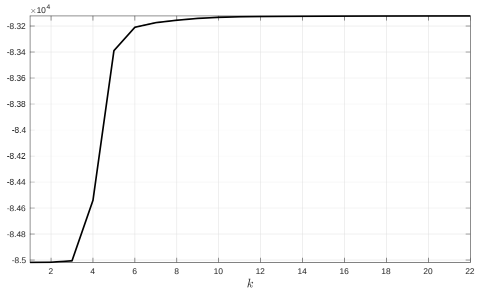
$$\widehat{\mathbf{P}} = \begin{bmatrix} 0.9194 & 0.0806 \\ 0.3395 & 0.6605 \end{bmatrix}.$$

The estimated unconditional probability for regime $j$ is equal to the sample average $\bar{\widehat{\xi}}_{j|T} = T^{-1}\sum_{t=1}^{T}\widehat{\xi}_{j,t|T}$, for $j = 1, 2$. It follows that $\bar{\widehat{\xi}}_{1|T} = 0.8044$ and $\bar{\widehat{\xi}}_{2|T} = 0.1956$.[4] Therefore,

[3] See `https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`.

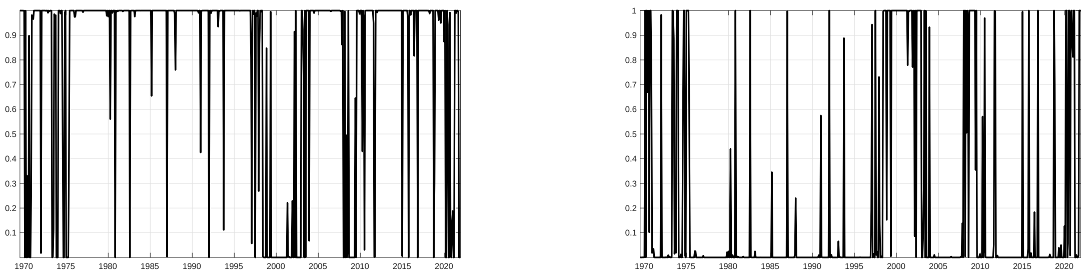
[4] By computing the unconditional probability using their analytical formulas in (6), we get $\bar{\widehat{\xi}}_{1|T} = 0.8081$ and $\bar{\widehat{\xi}}_{2|T} = 0.1919$.

**Figure 1:** Expected log-likelihood




This figure plots the value of the maximized expected conditional log-likelihood computed using the estimated factors, i.e., $\mathsf{E}_{\widehat{\mathbf{q}}^{(k)}}\left[\log f\left(\mathbf{X}\,\middle|\,\widehat{\boldsymbol{G}};\widehat{\boldsymbol{\varphi}}^{(k+1)},\widehat{\boldsymbol{\rho}}^{(k+1)}\right)|\boldsymbol{X}\right]$ (see also (15)), as a function of the EM iterations $k$.

**Figure 2:** Estimated conditional probabilities $\widehat{\boldsymbol{\xi}}_{t|T}$, $t = 1, \dots, T$.




(a): $\widehat{\xi}_{1,t|T}$ (b): $\widehat{\xi}_{2,t|T}$

This figure plots the series of the estimated conditional probabilities $\widehat{\xi}_{1,t|T}$ (Panel a) and $\widehat{\xi}_{2,t|T}$ (Panel b) estimated from the Markov switching factor model in (9).

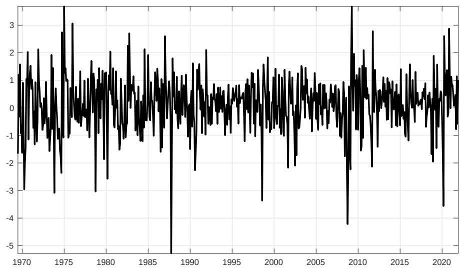
regime $j = 1$ is approximately four times more frequent than regime $j = 2$.

In order to provide an economic understanding of the regimes described by the model, Figure 2 plots the sequences of estimates $\widehat{\xi}_{1,t|T}$ and $\widehat{\xi}_{2,t|T}$, for $t = 1, \dots, T$. These series are negatively and positively correlated, respectively, with the NBER recession indicator, with correlation coefficients equal to -0.303 and 0.303, respectively. Therefore, the state $j = 1$ is related to periods of economic expansions, whereas the state $j = 2$ is more likely to occur during recessionary phases.[5] This is consistent with the empirical frequency of the states, since expansions occur more often than recessions.[6] Our model therefore captures regime changes in equity markets related to business cycle dynamics.

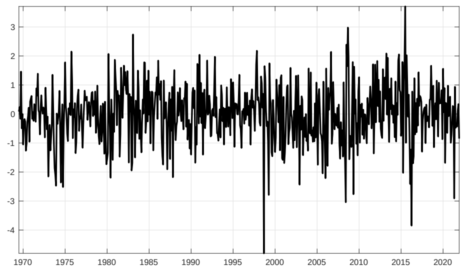
[5]The NBER recession indicator is publicly available at `https://fred.stlouisfed.org/series/USREC`.

[6]Information on U.S. business cycle dates may be found at `https://www.nber.org/research/data/us-business-cycle-expar`

**Figure 3:** ESTIMATED FACTORS $\widehat{\mathbf{g}}_t$, $t = 1, \dots, T$.



(a): $\widehat{g}_{1t}$



(b): $\widehat{g}_{2t}$

This figure plots the series of the estimated factors $\widehat{g}_{1t}$ (Panel a) and $\widehat{g}_{2t}$ (Panel b) estimated from the linear factor model in (10).

## 7.3 Factors and loadings

### 7.3.1 Equivalent linear representation

Following the sequential order dictated by our estimation procedure, we first consider estimated factors and loadings for the equivalent linear representation in (10), namely $\widehat{\mathbf{g}}_t = (\widehat{g}_{1t}, \widehat{g}_{2t})'$ and $\widehat{\mathbf{A}} = (\widehat{\mathbf{a}}_1, \widehat{\mathbf{a}}_2)$ as obtained in Section 3.2.

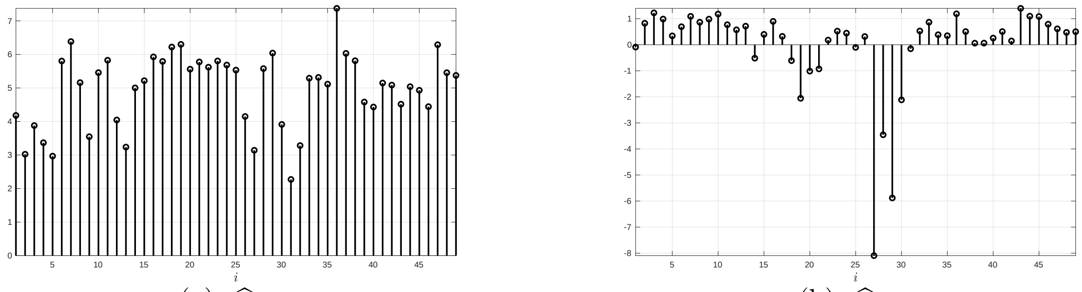
We begin from the estimated factors $\widehat{g}_{1t}$ and $\widehat{g}_{2t}$, which are displayed in Figure 3. To aid understanding of these factors, we study the correlation between them and the six observable factors considered in Fama and French (2016), namely: the value-weighted return on the market portfolio in excess of the one-month Treasury bill rate ($RM_t$); size ($SMB_t$); value ($HML_t$); profitability ($RMW_t$); investment ($CMA_t$); momentum ($MOM_t$). The correlations displayed in Table 5 show that the first estimated factor $\widehat{g}_{1t}$ is strongly correlated with the market return $RM_t$, it is reasonably correlated with $SMB_t$, $CMA_t$ and $MOM_t$, and it is only mildly correlated with $HML_t$ and $RMW_t$. On the other hand, the second estimated latent factor does not exhibit any substantial correlation with any of the observable factors we consider. This implies that the first factor in the equivalent linear representation is likely to be a market factor, while it is more difficult to give economic interpretation to the second factor.

The estimated loadings are displayed in Figure 4. It is important to note that the elements of the estimated vector $\widehat{\mathbf{a}}_1$ associated to $\widehat{g}_{1t}$ all have the same sign, which confirms that the corresponding factor $g_{1t}$ is a level factor. On the other hand, the vector of loadings $\widehat{\mathbf{a}}_2$ associated to $\widehat{g}_{2t}$ has elements with positive and negative sign, which is consistent with $g_{2t}$ being a slope factor.

### 7.3.2 Markov switching factor model

We then study the four common factors collected in the vector $\widehat{\boldsymbol{\xi}}_{t|T} \otimes \widehat{\mathbf{g}}_t$, for $t = 1, \dots, T$, namely $\widehat{\xi}_{j,t|T}\widehat{g}_{1t}$ and $\widehat{\xi}_{j,t|T}\widehat{g}_{2t}$, for $j = 1, 2$. These four series are plotted in Figure 5. To aid

**Figure 4:** Estimated loadings $\widehat{\mathbf{A}}$.



(a): $\widehat{\mathbf{a}}_1$ (b): $\widehat{\mathbf{a}}_2$

This figure plots the sequences of estimated loadings $\widehat{a}_{i1}$ (Panel a) and $\widehat{a}_{i2}$ (Panel b), $i = 1, \ldots, N$, estimated from the linear factor model in (10).

**Table 5:** Factor correlations in linear model.

| | $\widehat{g}_{1t}$ | $\widehat{g}_{2t}$ |
|---|---|---|
| $RM_t$ | 0.96 | 0.05 |
| $SMB_t$ | 0.40 | -0.06 |
| $HML_t$ | -0.13 | -0.06 |
| $RMW_t$ | -0.14 | 0.12 |
| $CMA_t$ | -0.32 | -0.12 |
| $MOM_t$ | -0.26 | -0.08 |

This table reports the correlation coefficients between the estimated latent factors $\widehat{g}_{1t}$ and $\widehat{g}_{2t}$ from the equivalent linear specification in (10) and the following six observable factors from Fama and French (2016): the value-weighted return on the market portfolio in excess of the one-month Treasury bill rate ($RM_t$); size ($SMB_t$); value ($HML_t$); profitability ($RMW_t$); investment ($CMA_t$); momentum ($MOM_t$).

understanding of those factors, Table 6 displays their correlations with the same observable factors detailed in Section 7.3.1. Consistently with the results in Table 5, Table 6 shows that both $\widehat{\xi}_{1,t|T}\widehat{g}_{1t}$ and $\widehat{\xi}_{2,t|T}\widehat{g}_{1t}$ are highly correlated with $RM_t$ and therefore capture the amount of cross-sectional dependence in stock returns driven by the market within each regime. This result is easy to interpret, as $\widehat{g}_{1t}$ is obtained as the sum of $\widehat{\xi}_{1,t|T}\widehat{g}_{1t}$ and $\widehat{\xi}_{2,t|T}\widehat{g}_{1t}$. On the other hand, $\widehat{\xi}_{1,t|T}\widehat{g}_{2t}$ and $\widehat{\xi}_{2,t|T}\widehat{g}_{2t}$ are more difficult to interpret, as they display limited correlation with all observable factors we consider, a feature they naturally share with $\widehat{g}_{2t}$.

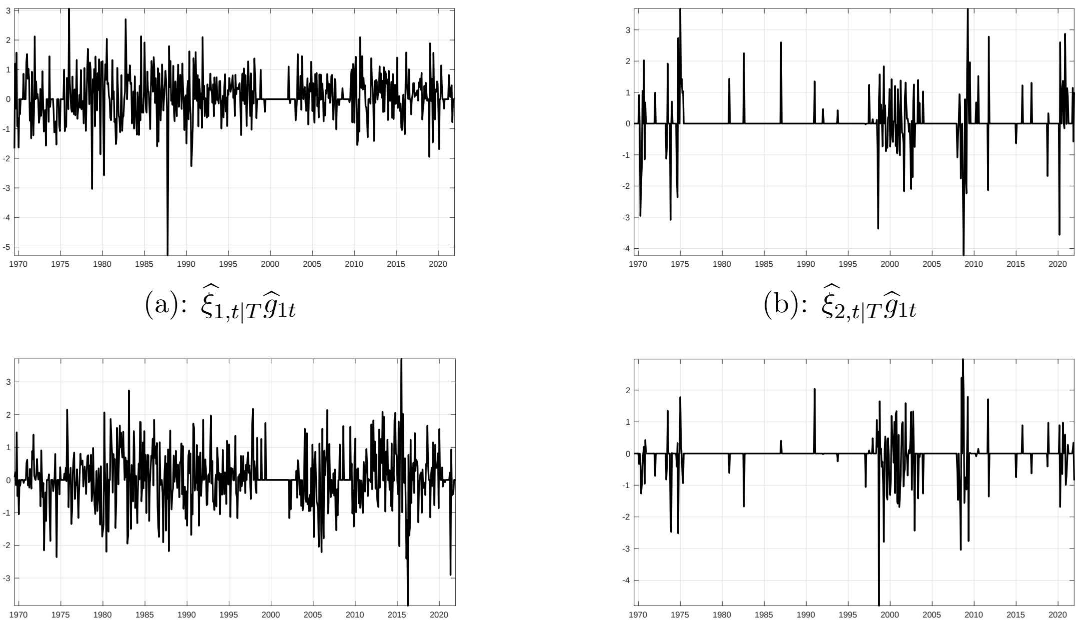
Finally, we consider the vectors of loadings $\widehat{\mathbf{b}}_{ji} = \left(\widehat{b}_{ji1}, \widehat{b}_{ji2}\right)'$, for $j = 1, 2$ and $i = 1, \ldots, N$. These are shown in Figure 6. It is important to note that both $\left\{\widehat{b}_{1i1}\right\}_{i=1}^{N}$ and $\left\{\widehat{b}_{2i1}\right\}_{i=1}^{N}$ have all elements with the same sign. This feature is inherited from $\widehat{\mathbf{a}}_1$ through $\widehat{\mathbf{g}}_{1t}$ and it confirms that $\widehat{b}_{1i1}$ and $\widehat{b}_{2i1}$ are level factors. On the other hand, analogously to $\widehat{\mathbf{a}}_2$, $\left\{\widehat{b}_{1i2}\right\}_{i=1}^{N}$ and $\left\{\widehat{b}_{2i2}\right\}_{i=1}^{N}$ have both positive and negative elements, and the corresponding factors $\widehat{\xi}_{1,t|T}\widehat{g}_{2t}$ and $\widehat{\xi}_{2,t|T}\widehat{g}_{2t}$ are slope factors. It is also worth noting that $\widehat{\mathbf{b}}_{11}$ and $\widehat{\mathbf{b}}_{12}$ are very similar to each other, the same being true for $\widehat{\mathbf{b}}_{21}$ and $\widehat{\mathbf{b}}_{22}$: this suggests that the bias induced by the

**Table 6:** FACTOR CORRELATIONS IN MARKOV SWITCHING MODEL.

| | $\widehat{g}_{1t}\widehat{\xi}_{1,t\|T}$ | $\widehat{g}_{2t}\widehat{\xi}_{1,t\|T}$ | $\widehat{g}_{1t}\widehat{\xi}_{2,t\|T}$ | $\widehat{g}_{2t}\widehat{\xi}_{2,t\|T}$ |
|---|---|---|---|---|
| $RM_t$ | 0.74 | 0.06 | 0.62 | -0.01 |
| $SMB_t$ | 0.32 | -0.08 | 0.24 | 0.01 |
| $HML_t$ | -0.17 | -0.11 | 0.00 | 0.06 |
| $RMW_t$ | -0.06 | 0.09 | -0.15 | 0.08 |
| $CMA_t$ | -0.22 | -0.12 | -0.23 | -0.03 |
| $MOM_t$ | -0.01 | 0.03 | -0.39 | -0.19 |

This table reports the correlation coefficients between the estimated latent factors $\widehat{g}_{1t}\widehat{\xi}_{1,t|T}$, $\widehat{g}_{2t}\widehat{\xi}_{1,t|T}$, $\widehat{g}_{1t}\widehat{\xi}_{2,t|T}$, and $\widehat{g}_{2t}\widehat{\xi}_{2,t|T}$ from the Markov switching factor model in equivalent linear specification in (1) and the following six observable factors from Fama and French (2016): the value-weighted return on the market portfolio in excess of the one-month Treasury bill rate ($RM_t$); size ($SMB_t$); value ($HML_t$); profitability ($RMW_t$); investment ($CMA_t$); momentum ($MOM_t$).

**Figure 5:** ESTIMATED FACTORS $\widehat{\boldsymbol{\xi}}_{t|T} \otimes \widehat{\mathbf{g}}_t$, $t = 1, \dots, T$.



(a): $\widehat{\xi}_{1,t|T}\widehat{g}_{1t}$ (b): $\widehat{\xi}_{2,t|T}\widehat{g}_{1t}$

(c): $\widehat{\xi}_{1,t|T}\widehat{g}_{2t}$ (d): $\widehat{\xi}_{2,t|T}\widehat{g}_{2t}$

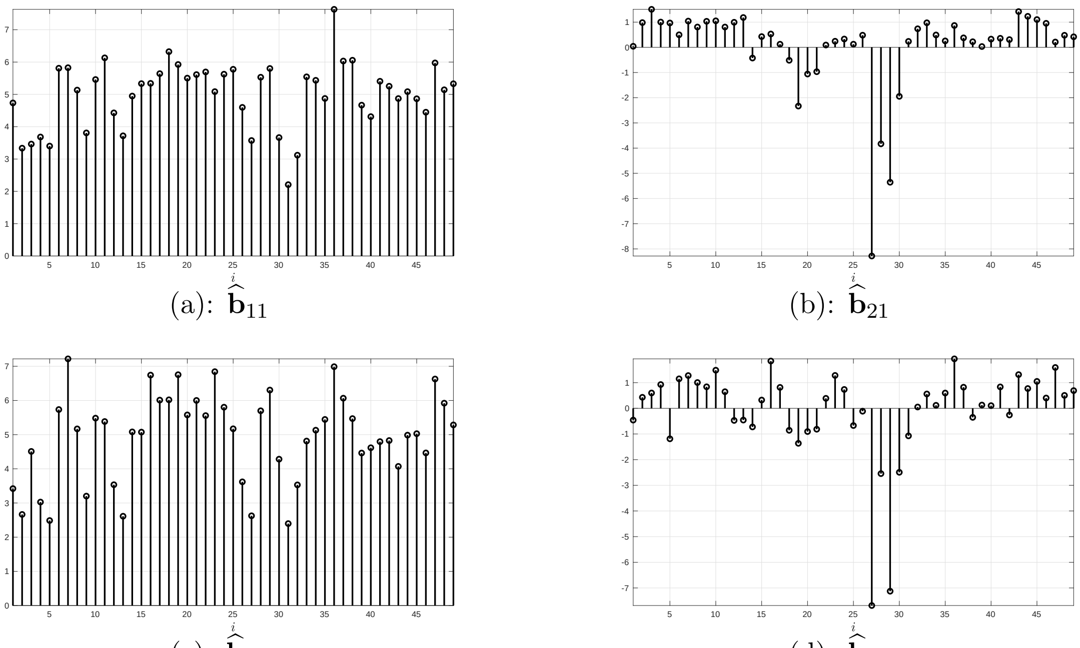
This figure plots the series of the estimated factors $\widehat{\xi}_{1,t|T}\widehat{g}_{1t}$ (Panel a), $\widehat{\xi}_{2,t|T}\widehat{g}_{1t}$ (Panel b), $\widehat{\xi}_{1,t|T}\widehat{g}_{2t}$ (Panel c), and $\widehat{\xi}_{2,t|T}\widehat{g}_{2t}$ (Panel d) estimated from the Markov switching factor model in (9).

imperfect knowledge of the state is almost negligible it practice, which implies that the regimes are estimated with high degree of accuracy.

# 8 Concluding remarks

This paper has developed estimation and inferential theory for high dimensional factor models with discrete regime changes in the loadings driven by a latent first order Markov process.

**Figure 6:** Estimated loadings $\widehat{\mathbf{B}}_1$ and $\widehat{\mathbf{B}}_2$.



(a): $\widehat{\mathbf{b}}_{11}$ (b): $\widehat{\mathbf{b}}_{21}$

(c): $\widehat{\mathbf{b}}_{12}$ (d): $\widehat{\mathbf{b}}_{22}$

This figure plots the sequences of estimated factor loadings $\widehat{b}_{1i1}$ (Panel a), $\widehat{b}_{1i2}$ (Panel b), $\widehat{b}_{2i1}$ (Panel c), and $\widehat{b}_{2i2}$ (Panel d), $i = 1, \dots, N$, estimated from the Markov switching factor model in (9).

Our estimator employs a EM algorithm based on a modified version of the Baum-Lindgren-Hamilton-Kim filter and smoother. Remarkably, the estimator does not need knowledge of the number of factors in either states, as it only requires the true number of factors in the equivalent linear representation, which can be estimated using existing techniques. We derive convergence rates and asymptotic distribution of the estimators for factors and loadings, and we show their good finite sample performance through an extensive set of Monte Carlo experiments. Finally, we empirically validate our methodology through an application to a large set of stock returns.

Our work can be extended along several dimensions. Three are worth mentioning. First, our model allows for two regimes and the case of multiple states to capture richer dynamics is worth exploring. Second, the problem of estimating the number of factors within each regime should be addressed. Finally, the challenging task of making inference on the number of regimes is worth considering. All these extensions are part of our research agenda.

# A Appendix: mathematical proofs

Define $C_{NT} = \min\left\{\sqrt{N}, \sqrt{T}\right\}$ Let $\mathbb{I}_{1t} = \mathbb{I}(s_t = 1)$ and $\mathbb{I}_{2t} = \mathbb{I}(s_t = 2)$. For $j = 1, 2$, and $i, l = 1, \ldots, N$, define

$$\begin{aligned} \sigma_{jil} &= \mathsf{E}\left(\frac{1}{T}\sum_{t=1}^{T}\mathbb{I}_{jt}e_{it}e_{lt}\right), \quad \chi_{jil} = \frac{1}{T}\sum_{t=1}^{T}\mathbb{I}_{jt}e_{it}e_{lt} - \mathsf{E}\left(\frac{1}{T}\sum_{t=1}^{T}\mathbb{I}_{jt}e_{it}e_{lt}\right), \\ \varphi_{jil} &= \frac{1}{T}\sum_{t=1}^{T}\mathbb{I}_{jt}\boldsymbol{\lambda}_{ji}'\mathbf{f}_{jt}e_{lt}, \quad \varphi_{jli} = \frac{1}{T}\sum_{t=1}^{T}\mathbb{I}_{jt}\boldsymbol{\lambda}_{jl}'\mathbf{f}_{jt}e_{it}. \end{aligned} \tag{A.1}$$

## A.1 Lemmas

**Lemma 1.** Under Assumptions 1 - 4, and given $\widehat{\mathbf{H}}$ defined in (45), we have

$$\frac{1}{N}\sum_{i=1}^{N}\left\|\widehat{\mathbf{a}}_i - \widehat{\mathbf{H}}'\mathbf{a}_i\right\|^2 = O_p\left(\frac{1}{C_{NT}^2}\right).$$

**Lemma 2.** Let Assumptions 1 - 6 hold. Then:

**(a)** $N^{-1}\sum_{l=1}^{N}\widehat{\mathbf{a}}_l\sigma_{jil} = O_p\left(\frac{1}{\sqrt{N}C_{NT}}\right)$;

**(b)** $N^{-1}\sum_{l=1}^{N}\widehat{\mathbf{a}}_l\chi_{jil} = O_p\left(\frac{1}{\sqrt{T}C_{NT}}\right)$;

**(c)** $N^{-1}\sum_{l=1}^{N}\widehat{\mathbf{a}}_l\varphi_{jil} = O_p\left(\frac{1}{\sqrt{T}C_{NT}}\right)$;

**(d)** $N^{-1}\sum_{l=1}^{N}\widehat{\mathbf{a}}_l\varphi_{jli} = O_p\left(\frac{1}{\sqrt{T}}\right)$.

**Lemma 3.** Under Assumptions 1 - 6,

$$N^{-1}\left(\widehat{\mathbf{A}} - \mathbf{A}\widehat{\mathbf{H}}\right)'\widehat{\mathbf{A}} = O_p\left(\frac{1}{C_{NT}^2}\right).$$

**Lemma 4.** Under Assumptions 1 - 6,

$$N^{-1}\left(\widehat{\mathbf{A}} - \mathbf{A}\widehat{\mathbf{H}}\right)'\mathbf{e}_t = O_p\left(\frac{1}{C_{NT}^2}\right).$$

**Lemma 5.** Let Assumptions 1 - 6 hold. Then:

**(a)** $\widehat{\mathbf{g}}_t - \widehat{\mathbf{H}}^{-1}\mathbf{g}_t = O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{C_{NT}^2}\right)$, for $t = 1, \ldots, T$;

**(b)** $\frac{1}{T}\sum_{t=1}^{T}\left(\widehat{\mathbf{g}}_t - \widehat{\mathbf{H}}^{-1}\mathbf{g}_t\right)\widehat{\mathbf{g}}_t' = O_p\left(\frac{1}{C_{NT}^2}\right)$.

**Lemma 6.** Under Assumptions 1 - 5, and given $\mathbf{Q}$ defined in Theorem 1,

$$p\lim_{N,T\to\infty}\frac{\mathbf{A}'\widehat{\mathbf{A}}}{N} = \mathbf{Q},$$

**Lemma 7.** Let Assumptions 1 - 5 hold, and consider the matrix $\mathbf{Q}$ defined in Theorem 1. Then, for $j = 1, 2$, the $r_j \times (r_1 + r_2)$ matrix $\mathbf{Q}_j$ satisfying $\mathbf{Q} = [\mathbf{Q}_1' \ \mathbf{Q}_2']'$ is such that

$$\mathbf{Q}_j = \mathbf{\Sigma}_{\mathbf{f}j}^{-1/2} \mathbf{\Psi}_j \mathbf{V}^{1/2},$$

where $\mathbf{\Sigma}_{\mathbf{f}j}$ is defined in (40), and $\mathbf{\Psi}_j$ is the $r_j \times (r_1 + r_2)$ matrix such that $\mathbf{\Psi} = [\mathbf{\Psi}_1' \ \mathbf{\Psi}_2']'$.

**Lemma 8.** Let $\widehat{\mathbf{V}}$ be the $(r_1 + r_2) \times (r_1 + r_2)$ diagonal matrix containing the first $r_1 + r_2$ eigenvalues of $\widehat{\mathbf{\Sigma}}_{\mathbf{x}} = (NT)^{-1} \sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t'$ in decreasing order. Define $\mathbf{V}$ as the $(r_1 + r_2) \times (r_1 + r_2)$ diagonal matrix of the first $r_1 + r_2$ eigenvalues of $\mathbf{\Sigma}_{\mathbf{g}}^{1/2} \mathbf{\Sigma}_{\mathbf{A}} \mathbf{\Sigma}_{\mathbf{g}}^{1/2}$ in decreasing order, where $\mathbf{\Sigma}_{\mathbf{g}}$ and $\mathbf{\Sigma}_{\mathbf{A}}$ are defined in (41) and (43), respectively. Then, under Assumptions 1 - 4,

$$\widehat{\mathbf{V}} \xrightarrow{p} \mathbf{V}.$$

## A.2 Proofs of Lemmas

***Proof of Lemma 1.*** Consider $\widehat{\mathbf{\Sigma}}_{\mathbf{x}} = (NT)^{-1} \sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t'$ and $\widehat{\mathbf{H}} = (\mathbf{G}\mathbf{G}'/T) \left( \mathbf{A}'\widehat{\mathbf{A}}/N \right) \widehat{\mathbf{V}}^{-1}$. By the definition of eigenvectors and eigenvalues, $\widehat{\mathbf{\Sigma}}_{\mathbf{x}} \widehat{\mathbf{A}} = \widehat{\mathbf{A}} \widehat{\mathbf{V}}$, where $\widehat{\mathbf{V}}$ is the $r \times r$ diagonal matrix of the first $r = (r_1 + r_2)$ largest eigenvalues of $\widehat{\mathbf{\Sigma}}_{\mathbf{x}}$ in decreasing order, and $\widehat{\mathbf{A}}$ is $\sqrt{N}$ times the $N \times r$ matrix of eigenvectors of $\widehat{\mathbf{\Sigma}}_{\mathbf{x}}$ corresponding to its $r$ largest eigenvalues. Note that $\left\| \widehat{\mathbf{V}} \right\| = O_p(1)$ and $\left\| \widehat{\mathbf{H}} \right\| \leq \left\| \mathbf{G}\mathbf{G}'/T \right\| \left\| \mathbf{A}\mathbf{A}'/N \right\|^{1/2} \left\| \widehat{\mathbf{A}}\widehat{\mathbf{A}}'/N \right\|^{1/2} \left\| \mathbf{V}^{-1} \right\| = O_p(1)$ by Assumptions 1 and 2. We then have

$$\left( \widehat{\mathbf{A}} - \mathbf{A}\widehat{\mathbf{H}} \right) \widehat{\mathbf{V}} = \widehat{\mathbf{A}}\widehat{\mathbf{V}} - \mathbf{A}\widehat{\mathbf{H}}\widehat{\mathbf{V}} = \widehat{\mathbf{A}}\widehat{\mathbf{V}} - \mathbf{A}\frac{\mathbf{G}\mathbf{G}'}{T}\frac{\mathbf{A}'\widehat{\mathbf{A}}}{N},$$

which implies

$$\widehat{\mathbf{V}}\widehat{\mathbf{A}}' - \frac{\widehat{\mathbf{A}}'\mathbf{A}}{N}\frac{\mathbf{G}\mathbf{G}'}{T}\mathbf{A}' = \widehat{\mathbf{A}}'\widehat{\mathbf{\Sigma}}_{\mathbf{x}} - \frac{\widehat{\mathbf{A}}'\mathbf{A}}{N}\frac{\mathbf{G}\mathbf{G}'}{T}\mathbf{A}' = \widehat{\mathbf{A}}'\frac{1}{NT}\left[ \left( \sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t' \right) - \mathbf{A}\mathbf{G}\mathbf{G}'\mathbf{A}' \right].$$

Taking into account (A.1), after some algebra we have

$$\begin{aligned} \widehat{\mathbf{V}} \left( \widehat{\mathbf{a}}_i - \widehat{\mathbf{H}}'\mathbf{a}_i \right) &= \widehat{\mathbf{A}}'\frac{1}{NT}\left[ \left( \sum_{t=1}^{T} \mathbf{x}_t x_{it} \right) - \mathbf{A}\mathbf{G}\mathbf{G}'\mathbf{a}_i \right] \\ &= \left[ \sum_{j=1}^{2} \left( \frac{1}{N}\sum_{l=1}^{N} \widehat{\mathbf{a}}_l \sigma_{jil} + \frac{1}{N}\sum_{l=1}^{N} \widehat{\mathbf{a}}_l \chi_{jil} + \frac{1}{N}\sum_{l=1}^{N} \widehat{\mathbf{a}}_l \varphi_{jil} + \frac{1}{N}\sum_{l=1}^{N} \widehat{\mathbf{a}}_l \varphi_{jli} \right) \right]. \end{aligned} \tag{A.2}$$

It follows that

$$\frac{1}{N}\sum_{i=1}^{N} \left\| \widehat{\mathbf{a}}_i - \widehat{\mathbf{H}}'\mathbf{a}_i \right\|^2 \leq 8 \left\| \widehat{\mathbf{V}}^{-1} \right\|^2 \sum_{j=1}^{2} \left( \frac{1}{N}\sum_{i=1}^{N} \widehat{\sigma}_{ji\cdot} + \frac{1}{N}\sum_{i=1}^{N} \widehat{\chi}_{ji\cdot} + \frac{1}{N}\sum_{i=1}^{N} \widehat{\varphi}_{ji\cdot} + \frac{1}{N}\sum_{i=1}^{N} \widehat{\varphi}_{j\cdot i} \right), \tag{A.3}$$

where

$$\widehat{\sigma}_{ji\cdot} = \frac{1}{N^2}\left\|\sum_{l=1}^{N}\widehat{\mathbf{a}}_l\sigma_{jil}\right\|^2, \quad \widehat{\chi}_{ji\cdot} = \frac{1}{N^2}\left\|\sum_{l=1}^{N}\widehat{\mathbf{a}}_l\chi_{jil}\right\|^2, \quad \widehat{\varphi}_{ji\cdot} = \frac{1}{N^2}\left\|\sum_{l=1}^{N}\widehat{\mathbf{a}}_l\varphi_{jil}\right\|^2, \quad \widehat{\varphi}_{j\cdot i} = \frac{1}{N^2}\left\|\sum_{l=1}^{N}\widehat{\mathbf{a}}_l\varphi_{jli}\right\|^2.$$

Consider $\widehat{\sigma}_{ji\cdot}$ and note that

$$\left\|\sum_{l=1}^{N}\widehat{\mathbf{a}}_l\sigma_{jil}\right\|^2 \leq \left(\sum_{l=1}^{N}\|\widehat{\mathbf{a}}_l\|^2\right)\left(\sum_{l=1}^{N}\sigma_{jil}^2\right)$$

so that

$$\frac{1}{N}\sum_{i=1}^{N}\widehat{\sigma}_{ji\cdot} = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{N^2}\left\|\sum_{l=1}^{N}\widehat{\mathbf{a}}_l\sigma_{jil}\right\|^2\right) \leq \frac{1}{N}\left(\frac{1}{N}\sum_{l=1}^{N}\|\widehat{\mathbf{a}}_l\|^2\right)\frac{1}{N}\left(\sum_{i=1}^{N}\sum_{l=1}^{N}\sigma_{jil}^2\right):$$

given Assumption 3(b), $N^{-1}\left(\sum_{i=1}^{N}\sum_{l=1}^{N}\sigma_{jil}^2\right) \leq M$ by Lemma A.1(a) in Massacci (2017), which implies that

$$\frac{1}{N}\sum_{i=1}^{N}\widehat{\sigma}_{ji\cdot} = O_p\left(\frac{1}{N}\right). \tag{A.4}$$

Consider now,

$$\begin{aligned}
\sum_{i=1}^{N}\widehat{\chi}_{ji\cdot} &= \frac{1}{N^2}\sum_{i=1}^{N}\left\|\sum_{l=1}^{N}\widehat{\mathbf{a}}_l\chi_{jil}\right\|^2 \\
&= \frac{1}{N^2}\sum_{i=1}^{N}\sum_{l=1}^{N}\sum_{q=1}^{N}\widehat{\mathbf{a}}_l'\widehat{\mathbf{a}}_q\chi_{jil}\chi_{jiq} \\
&\leq \left[\frac{1}{N^2}\sum_{l=1}^{N}\sum_{q=1}^{N}(\widehat{\mathbf{a}}_l'\widehat{\mathbf{a}}_q)^2\right]^{1/2}\left[\frac{1}{N^2}\sum_{l=1}^{N}\sum_{q=1}^{N}\left(\sum_{i=1}^{N}\chi_{jil}\chi_{jiq}\right)^2\right]^{1/2} \\
&\leq \left(\frac{1}{N}\sum_{l=1}^{N}\|\widehat{\mathbf{a}}_l\|^2\right)\left[\frac{1}{N^2}\sum_{l=1}^{N}\sum_{q=1}^{N}\left(\sum_{i=1}^{N}\chi_{jil}\chi_{jiq}\right)^2\right]^{1/2};
\end{aligned}$$

since

$$\mathsf{E}\left[\left(\sum_{i=1}^{N}\chi_{jil}\chi_{jiq}\right)^2\right] = \mathsf{E}\left(\sum_{i=1}^{N}\sum_{u=1}^{N}\chi_{jil}\chi_{jiq}\chi_{jul}\chi_{juq}\right) \leq N^2\max_{i,l}\mathsf{E}\left(|\chi_{jil}|^4\right)$$

and

$$\begin{aligned}
\mathsf{E}\left(|\chi_{jil}|^4\right) &= \mathsf{E}\left[\left|\frac{1}{T}\sum_{t=1}^{T}\mathbb{I}_{jt}e_{it}e_{lt} - \mathsf{E}\left(\frac{1}{T}\sum_{t=1}^{T}\mathbb{I}_{jt}e_{it}e_{lt}\right)\right|^4\right] \\
&= \frac{1}{T^2}\mathsf{E}\left\{\left|\frac{1}{\sqrt{T}}\left[\sum_{t=1}^{T}\mathbb{I}_{jt}e_{it}e_{lt} - \mathsf{E}\left(\sum_{t=1}^{T}\mathbb{I}_{jt}e_{it}e_{lt}\right)\right]\right|^4\right\} \\
&\leq \frac{1}{T^2}M
\end{aligned}$$

by Assumption 3(c), then

$$\sum_{i=1}^{N}\widehat{\chi}_{ji\cdot}\leq O_p\left(1\right)\sqrt{\frac{N^2}{T^2}}=O_p\left(\frac{N}{T}\right)$$

and

$$\frac{1}{N}\sum_{i=1}^{N}\widehat{\chi}_{ji\cdot}=O_p\left(\frac{1}{T}\right). \tag{A.5}$$

Also

$$\begin{aligned}
\widehat{\varphi}_{ji\cdot} &= \frac{1}{N^2}\left\|\sum_{l=1}^{N}\widehat{\mathbf{a}}_l\varphi_{jil}\right\|^2 \\
&= \frac{1}{N^2}\left\|\sum_{l=1}^{N}\widehat{\mathbf{a}}_l\left(\frac{1}{T}\sum_{t=1}^{T}\mathbb{I}_{jt}\boldsymbol{\lambda}_{ji}'\mathbf{f}_{jt}e_{lt}\right)\right\|^2 \\
&= \frac{1}{N^2}\left\|\sum_{l=1}^{N}\widehat{\mathbf{a}}_l\boldsymbol{\lambda}_{ji}'\left(\frac{1}{T}\sum_{t=1}^{T}\mathbb{I}_{jt}\mathbf{f}_{jt}e_{lt}\right)\right\|^2 \\
&\leq \left[\frac{1}{N}\sum_{l=1}^{N}\left(\frac{1}{T^2}\left\|\sum_{t=1}^{T}\mathbb{I}_{jt}\mathbf{f}_{jt}e_{lt}\right\|^2\right)\right]\left\|\boldsymbol{\lambda}_{ji}\right\|^2\left(\frac{1}{N}\sum_{l=1}^{N}\left\|\widehat{\mathbf{a}}_l\right\|^2\right)
\end{aligned}$$

and

$$\begin{aligned}
\frac{1}{N}\sum_{i=1}^{N}\widehat{\varphi}_{ji\cdot} &= \left[\frac{1}{N}\sum_{l=1}^{N}\left(\frac{1}{T^2}\left\|\sum_{t=1}^{T}\mathbb{I}_{jt}\mathbf{f}_{jt}e_{lt}\right\|^2\right)\right]\left(\frac{1}{N}\sum_{i=1}^{N}\left\|\boldsymbol{\lambda}_{ji}\right\|^2\right)\left(\frac{1}{N}\sum_{l=1}^{N}\left\|\widehat{\mathbf{a}}_l\right\|^2\right) \\
&= \frac{1}{T}\left(\frac{1}{N}\sum_{l=1}^{N}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbb{I}_{jt}\mathbf{f}_{jt}e_{lt}\right\|^2\right)\left(\frac{1}{N}\sum_{i=1}^{N}\left\|\boldsymbol{\lambda}_{ji}\right\|^2\right)\left(\frac{1}{N}\sum_{l=1}^{N}\left\|\widehat{\mathbf{a}}_l\right\|^2\right) \qquad \text{(A.6)} \\
&= O_p\left(\frac{1}{T}\right)
\end{aligned}$$

by Assumptions 2 and 4. Finally,

$$\begin{aligned}
\widehat{\varphi}_{j\cdot i} &= \frac{1}{N^2}\left\|\sum_{l=1}^{N}\widehat{\mathbf{a}}_l\varphi_{jli}\right\|^2 \\
&= \frac{1}{N^2}\left\|\sum_{l=1}^{N}\widehat{\mathbf{a}}_l\left(\frac{1}{T}\sum_{t=1}^{T}\mathbb{I}_{jt}\boldsymbol{\lambda}_{jl}'\mathbf{f}_{jt}e_{it}\right)\right\|^2 \\
&= \frac{1}{N^2}\left\|\sum_{l=1}^{N}\widehat{\mathbf{a}}_l\boldsymbol{\lambda}_{jl}'\left(\frac{1}{T}\sum_{t=1}^{T}\mathbb{I}_{jt}\mathbf{f}_{jt}e_{it}\right)\right\|^2 \\
&\leq \frac{1}{N^2}\left\|\sum_{l=1}^{N}\widehat{\mathbf{a}}_l\boldsymbol{\lambda}_{jl}'\right\|^2\left\|\frac{1}{T}\sum_{t=1}^{T}\mathbb{I}_{jt}\mathbf{f}_{jt}e_{it}\right\|^2 \\
&\leq \frac{1}{T}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbb{I}_{jt}\mathbf{f}_{jt}e_{it}\right\|^2\left(\frac{1}{N}\sum_{l=1}^{N}\left\|\boldsymbol{\lambda}_{jl}\right\|^2\right)\left(\frac{1}{N}\sum_{l=1}^{N}\left\|\widehat{\mathbf{a}}_l\right\|^2\right)
\end{aligned}$$

and

$$\frac{1}{N}\sum_{i=1}^{N}\widehat{\varphi}_{j\cdot i} \leq \frac{1}{T}\left(\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbb{I}_{jt}\mathbf{f}_{jt}e_{it}\right\|^2\right)\left(\frac{1}{N}\sum_{l=1}^{N}\|\boldsymbol{\lambda}_{jl}\|^2\right)\left(\frac{1}{N}\sum_{l=1}^{N}\|\widehat{\mathbf{a}}_l\|^2\right) = O_p\left(\frac{1}{T}\right) \tag{A.7}$$

by Assumptions 2 and 4. By combining (A.3) - (A.7), and since $\left\|\widehat{\mathbf{V}}^{-1}\right\| = O_p(1)$, then

$$\frac{1}{N}\sum_{i=1}^{N}\left\|\widehat{\mathbf{a}}_i - \widehat{\mathbf{H}}'\mathbf{a}_i\right\|^2 = O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{T}\right)$$

and the result stated in the lemma follows. □

***Proof of Lemma 2***. Starting from $(a)$, consider

$$\frac{1}{N}\sum_{l=1}^{N}\widehat{\mathbf{a}}_l\sigma_{jil} = \frac{1}{N}\sum_{l=1}^{N}\left(\widehat{\mathbf{a}}_l - \widehat{\mathbf{H}}'\mathbf{a}_l + \widehat{\mathbf{H}}'\mathbf{a}_l\right)\sigma_{jil} = \frac{1}{N}\sum_{l=1}^{N}\left(\widehat{\mathbf{a}}_l - \widehat{\mathbf{H}}'\mathbf{a}_l\right)\sigma_{jil} + \widehat{\mathbf{H}}'\frac{1}{N}\sum_{l=1}^{N}\mathbf{a}_l\sigma_{jil}.$$

Note that

$$\left\|\sum_{l=1}^{N}\mathbf{a}_l\sigma_{jil}\right\| \leq \left(\max_l\|\mathbf{a}_l\|\right)\left(\sum_{l=1}^{N}|\sigma_{jil}|\right) \leq \left[\max_l\left(\|\boldsymbol{\lambda}_{1l}\| + \|\boldsymbol{\lambda}_{2l}\|\right)\right]\left(\sum_{l=1}^{N}|\sigma_{jil}|\right) \leq 2\bar{\lambda}M$$

by Assumption 2 and Assumption 3(b), so that

$$\frac{1}{N}\sum_{l=1}^{N}\mathbf{a}_l\sigma_{jil} = O\left(\frac{1}{N}\right).$$

Further

$$\begin{aligned}\left\|\frac{1}{N}\sum_{l=1}^{N}\left(\widehat{\mathbf{a}}_l - \widehat{\mathbf{H}}'\mathbf{a}_l\right)\sigma_{jil}\right\| &\leq \left(\frac{1}{N}\sum_{l=1}^{N}\left\|\widehat{\mathbf{a}}_l - \widehat{\mathbf{H}}'\mathbf{a}_l\right\|^2\right)^{1/2}\frac{1}{\sqrt{N}}\left(\sum_{l=1}^{N}|\sigma_{jil}|^2\right)^{1/2} \\ &= \left[O_p\left(\frac{1}{C_{NT}^2}\right)\right]^{1/2} O_p\left(\frac{1}{\sqrt{N}}\right) \\ &= O_p\left(\frac{1}{\sqrt{N}C_{NT}}\right)\end{aligned}$$

by Lemma 1 and Assumption 3(b). It thus follows that

$$\frac{1}{N}\sum_{l=1}^{N}\widehat{\mathbf{a}}_l\sigma_{jil} = O_p\left(\frac{1}{\sqrt{N}C_{NT}}\right) + O_p\left(\frac{1}{N}\right) = O_p\left(\frac{1}{\sqrt{N}C_{NT}}\right).$$

Moving on to $(b)$, we have

$$\frac{1}{N}\sum_{l=1}^{N}\widehat{\mathbf{a}}_l\chi_{jil}=\frac{1}{N}\sum_{l=1}^{N}\left(\widehat{\mathbf{a}}_l-\widehat{\mathbf{H}}'\mathbf{a}_l\right)\chi_{jil}+\widehat{\mathbf{H}}'\frac{1}{N}\sum_{l=1}^{N}\mathbf{a}_l\chi_{jil}.$$

Note that

$$\left\|\frac{1}{N}\sum_{l=1}^{N}\left(\widehat{\mathbf{a}}_l-\widehat{\mathbf{H}}'\mathbf{a}_l\right)\chi_{jil}\right\|\leq\left(\frac{1}{N}\sum_{l=1}^{N}\left\|\widehat{\mathbf{a}}_l-\widehat{\mathbf{H}}'\mathbf{a}_l\right\|^2\right)^{1/2}\left(\frac{1}{N}\sum_{l=1}^{N}\chi_{jil}^2\right)^{1/2},$$

with

$$\begin{aligned}\frac{1}{N}\sum_{l=1}^{N}\chi_{jil}^2 &= \frac{1}{N}\sum_{l=1}^{N}\left[\frac{1}{T}\sum_{t=1}^{T}\mathbb{I}_{jt}e_{it}e_{lt}-\mathsf{E}\left(\frac{1}{T}\sum_{t=1}^{T}\mathbb{I}_{jt}e_{it}e_{lt}\right)\right]^2\\ &= \frac{1}{NT}\sum_{l=1}^{N}\left\{\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\mathbb{I}_{jt}e_{it}e_{lt}-\mathsf{E}\left(\mathbb{I}_{jt}e_{it}e_{lt}\right)\right]\right\}^2\\ &= O_p\left(\frac{1}{T}\right)\end{aligned}$$

so that

$$\left\|\frac{1}{N}\sum_{l=1}^{N}\left(\widehat{\mathbf{a}}_l-\widehat{\mathbf{H}}'\mathbf{a}_l\right)\chi_{jil}\right\|=O_p\left(\frac{1}{C_{NT}}\right)O_p\left(\frac{1}{\sqrt{T}}\right)=O_p\left(\frac{1}{\sqrt{T}C_{NT}}\right).$$

Further

$$\begin{aligned}\frac{1}{N}\sum_{l=1}^{N}\mathbf{a}_l\chi_{jil} &= \frac{1}{N}\sum_{l=1}^{N}\mathbf{a}_l\left[\frac{1}{T}\sum_{t=1}^{T}\mathbb{I}_{jt}e_{it}e_{lt}-\mathsf{E}\left(\frac{1}{T}\sum_{t=1}^{T}\mathbb{I}_{jt}e_{it}e_{lt}\right)\right]\\ &= \frac{1}{NT}\sum_{l=1}^{N}\sum_{t=1}^{T}\mathbf{a}_l\left[\mathbb{I}_{jt}e_{it}e_{lt}-\mathsf{E}\left(\mathbb{I}_{jt}e_{it}e_{lt}\right)\right]\\ &= O_p\left(\frac{1}{\sqrt{NT}}\right)\end{aligned}$$

by Assumption 6(a). It follows that

$$\frac{1}{N}\sum_{l=1}^{N}\widehat{\mathbf{a}}_l\chi_{jil}=O_p\left(\frac{1}{\sqrt{T}C_{NT}}\right)+O_p\left(\frac{1}{\sqrt{NT}}\right)=O_p\left(\frac{1}{\sqrt{T}C_{NT}}\right).$$

As for $(c)$, consider

$$\begin{aligned}\frac{1}{N}\sum_{l=1}^{N}\widehat{\mathbf{a}}_l\varphi_{jil} &= \frac{1}{N}\sum_{l=1}^{N}\widehat{\mathbf{a}}_l\left(\frac{1}{T}\sum_{t=1}^{T}\mathbb{I}_{jt}\boldsymbol{\lambda}_{ji}'\mathbf{f}_{jt}e_{lt}\right)\\ &= \frac{1}{NT}\sum_{l=1}^{N}\sum_{t=1}^{T}\mathbb{I}_{jt}\widehat{\mathbf{a}}_l e_{lt}\mathbf{f}_{jt}'\boldsymbol{\lambda}_{ji}\\ &= \frac{1}{NT}\sum_{l=1}^{N}\sum_{t=1}^{T}\mathbb{I}_{jt}\left(\widehat{\mathbf{a}}_l-\widehat{\mathbf{H}}'\mathbf{a}_l+\widehat{\mathbf{H}}'\mathbf{a}_l\right)e_{lt}\mathbf{f}_{jt}'\boldsymbol{\lambda}_{ji}\\ &= \frac{1}{NT}\sum_{l=1}^{N}\sum_{t=1}^{T}\mathbb{I}_{jt}\left(\widehat{\mathbf{a}}_l-\widehat{\mathbf{H}}'\mathbf{a}_l\right)e_{lt}\mathbf{f}_{jt}'\boldsymbol{\lambda}_{ji}+\widehat{\mathbf{H}}'\frac{1}{NT}\sum_{l=1}^{N}\sum_{t=1}^{T}\mathbb{I}_{jt}\mathbf{a}_l e_{lt}\mathbf{f}_{jt}'\boldsymbol{\lambda}_{ji}.\end{aligned}$$

We have

$$\begin{aligned}\left\|\frac{1}{NT}\sum_{l=1}^{N}\sum_{t=1}^{T}\mathbb{I}_{jt}\left(\widehat{\mathbf{a}}_l-\widehat{\mathbf{H}}'\mathbf{a}_l\right)e_{lt}\mathbf{f}_{jt}'\boldsymbol{\lambda}_{ji}\right\| &\leq \frac{1}{\sqrt{T}}\left(\frac{1}{N}\sum_{l=1}^{N}\left\|\widehat{\mathbf{a}}_l-\widehat{\mathbf{H}}'\mathbf{a}_l\right\|^2\right)^{1/2} \\ &\quad\times\left(\frac{1}{N}\sum_{l=1}^{N}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbb{I}_{jt}e_{lt}\mathbf{f}_{jt}\right\|^2\right)^{1/2}\|\lambda_{ji}\| \\ &= O\left(\frac{1}{\sqrt{T}}\right)O_p\left(\frac{1}{C_{NT}}\right)O_p(1)\,O(1) \\ &= O_p\left(\frac{1}{\sqrt{T}C_{NT}}\right)\end{aligned}$$

by Lemma 1, Assumption 6(c) and Assumption 2. Also,

$$\frac{1}{NT}\sum_{l=1}^{N}\sum_{t=1}^{T}\mathbb{I}_{jt}\mathbf{a}_l e_{lt}\mathbf{f}_{jt}'\boldsymbol{\lambda}_{ji}=\frac{1}{\sqrt{NT}}\left[\frac{1}{\sqrt{NT}}\sum_{l=1}^{N}\sum_{t=1}^{T}\mathbb{I}_{jt}\begin{pmatrix}\boldsymbol{\lambda}_{1l}\\ \boldsymbol{\lambda}_{2l}\end{pmatrix}e_{lt}\mathbf{f}_{jt}'\right]\boldsymbol{\lambda}_{ji}=O_p\left(\frac{1}{\sqrt{NT}}\right)$$

by Assumption 6(b) and Assumption 2. It follows that

$$\frac{1}{N}\sum_{l=1}^{N}\widehat{\mathbf{a}}_l\varphi_{jil}=O_p\left(\frac{1}{\sqrt{T}C_{NT}}\right)+O_p\left(\frac{1}{\sqrt{NT}}\right)=O_p\left(\frac{1}{\sqrt{T}C_{NT}}\right).$$

Finally, for $(d)$ we have

$$\frac{1}{N}\sum_{l=1}^{N}\widehat{\mathbf{a}}_l\varphi_{jli}=\frac{1}{N}\sum_{l=1}^{N}\left(\widehat{\mathbf{a}}_l-\widehat{\mathbf{H}}'\mathbf{a}_l\right)\varphi_{jli}+\widehat{\mathbf{H}}'\frac{1}{N}\sum_{l=1}^{N}\mathbf{a}_l\varphi_{jli}.$$

Note that

$$\begin{aligned}\frac{1}{N}\sum_{l=1}^{N}\mathbf{a}_l\varphi_{jli} &= \frac{1}{N}\sum_{l=1}^{N}\mathbf{a}_l\left(\frac{1}{T}\sum_{t=1}^{T}\mathbb{I}_{jt}\boldsymbol{\lambda}_{jl}'\mathbf{f}_{jt}e_{it}\right) \\ &= \left(\frac{1}{N}\sum_{l=1}^{N}\mathbf{a}_l\boldsymbol{\lambda}_{jl}'\right)\left(\frac{1}{T}\sum_{t=1}^{T}\mathbb{I}_{jt}\mathbf{f}_{jt}e_{it}\right) \\ &= \left[\frac{1}{N}\sum_{l=1}^{N}\begin{pmatrix}\boldsymbol{\lambda}_{1l}\\ \boldsymbol{\lambda}_{2l}\end{pmatrix}\boldsymbol{\lambda}_{jl}'\right]\frac{1}{\sqrt{T}}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbb{I}_{jt}\mathbf{f}_{jt}e_{it}\right) \\ &= O_p\left(\frac{1}{\sqrt{T}}\right),\end{aligned}$$

by Assumption 2 and Assumption 6(c). Further,

$$\left\|\frac{1}{N}\sum_{l=1}^{N}\left(\widehat{\mathbf{a}}_l-\widehat{\mathbf{H}}'\mathbf{a}_l\right)\varphi_{jli}\right\|\leq\left(\frac{1}{N}\sum_{l=1}^{N}\left\|\widehat{\mathbf{a}}_l-\widehat{\mathbf{H}}'\mathbf{a}_l\right\|^2\right)^{1/2}\left(\frac{1}{N}\sum_{l=1}^{N}\varphi_{jli}^2\right)^{1/2}$$

with

$$\frac{1}{N}\sum_{l=1}^{N}\varphi_{jli}^{2}=\frac{1}{N}\sum_{l=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T}\mathbb{I}_{jt}\boldsymbol{\lambda}_{jl}'\mathbf{f}_{jt}e_{it}\right)^{2}\leq\frac{1}{T}\left(\frac{1}{N}\sum_{l=1}^{N}\|\boldsymbol{\lambda}_{jl}\|^{2}\right)\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbb{I}_{jt}\mathbf{f}_{jt}e_{it}\right)^{2}\leq O_{p}\left(\frac{1}{T}\right),$$

by Assumption 2 and Assumption 6(c), so that taking into account Lemma 1 we have

$$\frac{1}{N}\sum_{l=1}^{N}\left(\widehat{\mathbf{a}}_{l}-\widehat{\mathbf{H}}'\mathbf{a}_{l}\right)\varphi_{jli}=O_{p}\left(\frac{1}{C_{NT}}\right)O_{p}\left(\frac{1}{\sqrt{T}}\right)=O_{p}\left(\frac{1}{\sqrt{T}C_{NT}}\right).$$

It follows that

$$\frac{1}{N}\sum_{l=1}^{N}\widehat{\mathbf{a}}_{l}\varphi_{jli}=O_{p}\left(\frac{1}{\sqrt{T}C_{NT}}\right)+O_{p}\left(\frac{1}{\sqrt{T}}\right)=O_{p}\left(\frac{1}{\sqrt{T}}\right),$$

which completes the proof of the lemma. □

***Proof of Lemma 3***. Consider

$$\begin{aligned}N^{-1}\left(\widehat{\mathbf{A}}-\mathbf{A}\widehat{\mathbf{H}}\right)'\widehat{\mathbf{A}} &= N^{-1}\left(\widehat{\mathbf{A}}-\mathbf{A}\widehat{\mathbf{H}}\right)'\widehat{\mathbf{A}}-N^{-1}\left(\widehat{\mathbf{A}}-\mathbf{A}\widehat{\mathbf{H}}\right)'\mathbf{A}\widehat{\mathbf{H}}+N^{-1}\left(\widehat{\mathbf{A}}-\mathbf{A}\widehat{\mathbf{H}}\right)'\mathbf{A}\widehat{\mathbf{H}}\\ &= N^{-1}\left(\widehat{\mathbf{A}}-\mathbf{A}\widehat{\mathbf{H}}\right)'\mathbf{A}\widehat{\mathbf{H}}+N^{-1}\left(\widehat{\mathbf{A}}-\mathbf{A}\widehat{\mathbf{H}}\right)'\left(\widehat{\mathbf{A}}-\mathbf{A}\widehat{\mathbf{H}}\right).\end{aligned}\tag{A.8}$$

Using the identity in (A.2), we have

$$\begin{aligned}N^{-1}\left(\widehat{\mathbf{A}}-\mathbf{A}\widehat{\mathbf{H}}\right)'\mathbf{A} &= \widehat{\mathbf{V}}^{-1}\frac{1}{N}\sum_{i=1}^{N}\left(\widehat{\mathbf{a}}_{i}-\widehat{\mathbf{H}}'\mathbf{a}_{i}\right)\mathbf{a}_{i}\\ &= \widehat{\mathbf{V}}^{-1}\left\{\begin{array}{c}\sum_{j=1}^{2}\left[\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{N}\sum_{l=1}^{N}\widehat{\mathbf{a}}_{l}\sigma_{jil}\right)\mathbf{a}_{i}'\right]+\sum_{j=1}^{2}\left[\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{N}\sum_{l=1}^{N}\widehat{\mathbf{a}}_{l}\chi_{jil}\right)\mathbf{a}_{i}'\right]\\ +\sum_{j=1}^{2}\left[\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{N}\sum_{l=1}^{N}\widehat{\mathbf{a}}_{l}\varphi_{jil}\right)\mathbf{a}_{i}'\right]+\sum_{j=1}^{2}\left[\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{N}\sum_{l=1}^{N}\widehat{\mathbf{a}}_{l}\varphi_{jli}\right)\mathbf{a}_{i}'\right]\end{array}\right\}.\end{aligned}\tag{A.9}$$

Consider

$$\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{N}\sum_{l=1}^{N}\widehat{\mathbf{a}}_{l}\sigma_{jil}\right)\mathbf{a}_{i}'=\frac{1}{N}\sum_{i=1}^{N}\left[\frac{1}{N}\sum_{l=1}^{N}\left(\widehat{\mathbf{a}}_{l}-\widehat{\mathbf{H}}'\mathbf{a}_{l}\right)\sigma_{jil}\right]\mathbf{a}_{i}'+\widehat{\mathbf{H}}'\frac{1}{N}\sum_{i=1}^{N}\frac{1}{N}\sum_{l=1}^{N}\mathbf{a}_{l}\mathbf{a}_{i}'\sigma_{jil}.$$

We have

$$\begin{aligned}\left\|\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{N}\sum_{l=1}^{N}\widehat{\mathbf{a}}_{l}\sigma_{jil}\right)\mathbf{a}_{i}'\right\| &\leq \frac{1}{\sqrt{N}}\left(\frac{1}{N}\sum_{l=1}^{N}\left\|\widehat{\mathbf{a}}_{l}-\widehat{\mathbf{H}}'\mathbf{a}_{l}\right\|^{2}\right)^{1/2}\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{l=1}^{N}|\sigma_{jil}|^{2}\right)^{1/2}\left(\frac{1}{N}\sum_{i=1}^{N}\|\mathbf{a}_{i}\|^{2}\right)^{1/2}\\ &= \frac{1}{\sqrt{N}}O_{p}\left(\frac{1}{C_{NT}}\right)O_{p}\left(1\right)O_{p}\left(1\right)\\ &= O_{p}\left(\frac{1}{\sqrt{N}C_{NT}}\right),\end{aligned}$$

by Lemma 1, Assumption 2, and the fact that, given $\rho_{jil} = \sigma_{jil} \Big/ (\sigma_{jii}\sigma_{jll})^{1/2}$ , by Assumption 3(b) we have

$$\frac{1}{N}\sum_{i=1}^{N}\sum_{l=1}^{N}|\sigma_{jil}|^2 = \frac{1}{N}\sum_{i=1}^{N}\sum_{l=1}^{N}\sigma_{jii}\sigma_{jll}\rho_{jil}^2 \leq M\frac{1}{N}\sum_{i=1}^{N}\sum_{l=1}^{N}|\sigma_{jii}\sigma_{jll}|^{1/2}|\rho_{jil}| = M\frac{1}{N}\sum_{i=1}^{N}\sum_{l=1}^{N}|\sigma_{jil}| \leq M^2. \tag{A.10}$$

Further

$$\left\|\frac{1}{N}\sum_{i=1}^{N}\frac{1}{N}\sum_{l=1}^{N}\mathbf{a}_l\mathbf{a}_i'\sigma_{jil}\right\| \leq \frac{1}{N}\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{l=1}^{N}\|\mathbf{a}_l\|\|\mathbf{a}_i\||\sigma_{jil}|\right) = O\left(\frac{1}{N}\right)$$

by Assumptions 2 and 3(b). Therefore,

$$\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{N}\sum_{l=1}^{N}\widehat{\mathbf{a}}_l\sigma_{jil}\right)\mathbf{a}_i' = O_p\left(\frac{1}{\sqrt{N}C_{NT}}\right) + O\left(\frac{1}{N}\right) = O_p\left(\frac{1}{\sqrt{N}C_{NT}}\right). \tag{A.11}$$

Consider now

$$\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{N}\sum_{l=1}^{N}\widehat{\mathbf{a}}_l\chi_{jil}\right)\mathbf{a}_i' = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{1}{N}\sum_{l=1}^{N}\left(\widehat{\mathbf{a}}_l - \widehat{\mathbf{H}}'\mathbf{a}_l\right)\chi_{jil}\right]\mathbf{a}_i' + \widehat{\mathbf{H}}'\frac{1}{N}\sum_{i=1}^{N}\frac{1}{N}\sum_{l=1}^{N}\mathbf{a}_l\mathbf{a}_i'\chi_{jil}.$$

We have

$$\left\|\frac{1}{N}\sum_{i=1}^{N}\left[\frac{1}{N}\sum_{l=1}^{N}\left(\widehat{\mathbf{a}}_l - \widehat{\mathbf{H}}'\mathbf{a}_l\right)\chi_{jil}\right]\mathbf{a}_i'\right\| \leq \frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{N}\sum_{l=1}^{N}\left(\widehat{\mathbf{a}}_l - \widehat{\mathbf{H}}'\mathbf{a}_l\right)\chi_{jil}\right\|\|\mathbf{a}_i\|$$

and consider

$$\left\|\frac{1}{N}\sum_{l=1}^{N}\left(\widehat{\mathbf{a}}_l - \widehat{\mathbf{H}}'\mathbf{a}_l\right)\chi_{jil}\right\| \leq \left(\frac{1}{N}\sum_{l=1}^{N}\left\|\widehat{\mathbf{a}}_l - \widehat{\mathbf{H}}'\mathbf{a}_l\right\|^2\right)^{1/2}\left(\frac{1}{N}\sum_{l=1}^{N}|\chi_{jil}|^2\right)^{1/2}$$

with

$$\begin{aligned}
\left(\frac{1}{N}\sum_{l=1}^{N}|\chi_{jil}|^2\right)^{1/2} &= \left[\frac{1}{N}\sum_{l=1}^{N}\left|\frac{1}{T}\sum_{t=1}^{T}\mathbb{I}_{jt}e_{it}e_{lt} - \mathsf{E}\left(\frac{1}{T}\sum_{t=1}^{T}\mathbb{I}_{jt}e_{it}e_{lt}\right)\right|^2\right]^{1/2} \\
&= \frac{1}{\sqrt{T}}\left[\frac{1}{N}\sum_{l=1}^{N}\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbb{I}_{jt}e_{it}e_{lt} - \mathsf{E}\left(\frac{1}{T}\sum_{t=1}^{T}\mathbb{I}_{jt}e_{it}e_{lt}\right)\right|^2\right]^{1/2} \\
&= O_p\left(\frac{1}{\sqrt{T}}\right)
\end{aligned}$$

by Assumption 3(c). Therefore, taking into account Lemma 1,

$$\left\|\frac{1}{N}\sum_{l=1}^{N}\left(\widehat{\mathbf{a}}_l - \widehat{\mathbf{H}}'\mathbf{a}_l\right)\chi_{jil}\right\| = O_p\left(\frac{1}{C_{NT}}\right)O_p\left(\frac{1}{\sqrt{T}}\right) = O_p\left(\frac{1}{\sqrt{T}C_{NT}}\right).$$

Further,

$$\begin{aligned}
\left\| \frac{1}{N}\sum_{i=1}^{N}\frac{1}{N}\sum_{l=1}^{N}\mathbf{a}_l\mathbf{a}_i'\chi_{jil} \right\| &= \left\| \frac{1}{N}\sum_{i=1}^{N}\frac{1}{N}\sum_{l=1}^{N}\mathbf{a}_l\mathbf{a}_i' \left[\frac{1}{T}\sum_{t=1}^{T}\mathbb{I}_{jt}e_{it}e_{lt} - \mathsf{E}\left(\frac{1}{T}\sum_{t=1}^{T}\mathbb{I}_{jt}e_{it}e_{lt}\right)\right] \right\| \\
&\leq \frac{1}{\sqrt{NT}}\left\{ \frac{1}{N}\sum_{l=1}^{N}\|\mathbf{a}_l\| \left\| \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{a}_i\left[\mathbb{I}_{jt}e_{it}e_{lt} - \mathsf{E}\left(\mathbb{I}_{jt}e_{it}e_{lt}\right)\right] \right\| \right\} \\
&\leq \frac{1}{\sqrt{NT}}\left(\frac{1}{N}\sum_{l=1}^{N}\|\mathbf{a}_l\|^2\right)^{1/2} \left\{ \frac{1}{N}\sum_{l=1}^{N}\left\| \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{a}_i\left[\mathbb{I}_{jt}e_{it}e_{lt} - \mathsf{E}\left(\mathbb{I}_{jt}e_{it}e_{lt}\right)\right] \right\|^2 \right\}^{1/2} \\
&= O_p\left(\frac{1}{\sqrt{NT}}\right)
\end{aligned}$$

by Assumptions 2 and 6(a). Therefore,

$$\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{N}\sum_{l=1}^{N}\widehat{\mathbf{a}}_l\chi_{jil}\right)\mathbf{a}_i' = O_p\left(\frac{1}{\sqrt{T}C_{NT}}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right) = O_p\left(\frac{1}{\sqrt{T}C_{NT}}\right). \qquad \text{(A.12)}$$

Consider now

$$\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{N}\sum_{l=1}^{N}\widehat{\mathbf{a}}_l\varphi_{jil}\right)\mathbf{a}_i' = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{1}{N}\sum_{l=1}^{N}\left(\widehat{\mathbf{a}}_l - \widehat{\mathbf{H}}'\mathbf{a}_l\right)\varphi_{jil}\right]\mathbf{a}_i' + \widehat{\mathbf{H}}'\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{N}\sum_{l=1}^{N}\mathbf{a}_l\mathbf{a}_i'\varphi_{jil}\right).$$

We have

$$\left\| \frac{1}{N}\sum_{i=1}^{N}\left[\frac{1}{N}\sum_{l=1}^{N}\left(\widehat{\mathbf{a}}_l - \widehat{\mathbf{H}}'\mathbf{a}_l\right)\varphi_{jil}\right]\mathbf{a}_i' \right\| \leq \left(\frac{1}{N}\sum_{l=1}^{N}\left\|\widehat{\mathbf{a}}_l - \widehat{\mathbf{H}}'\mathbf{a}_l\right\|^2\right)^{1/2} \left(\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{N}\sum_{l=1}^{N}\varphi_{jil}\mathbf{a}_i\right\|^2\right)^{1/2}$$

and

$$\begin{aligned}
\left(\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{N}\sum_{l=1}^{N}\varphi_{jil}\mathbf{a}_i\right\|^2\right)^{1/2} &= \left[\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{N}\sum_{l=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T}\mathbb{I}_{jt}\boldsymbol{\lambda}_{ji}'\mathbf{f}_{jt}e_{lt}\right)\mathbf{a}_i\right\|^2\right]^{1/2} \\
&= \frac{1}{\sqrt{NT}}\left[\frac{1}{N}\sum_{i=1}^{N}\left\|\left(\frac{1}{\sqrt{NT}}\sum_{l=1}^{N}\sum_{t=1}^{T}\mathbb{I}_{jt}\boldsymbol{\lambda}_{ji}'\mathbf{f}_{jt}e_{lt}\right)\mathbf{a}_i\right\|^2\right]^{1/2} \\
&\leq \frac{1}{\sqrt{NT}}\left(\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{\sqrt{NT}}\sum_{l=1}^{N}\sum_{t=1}^{T}\mathbb{I}_{jt}\boldsymbol{\lambda}_{ji}'\mathbf{f}_{jt}e_{lt}\right\|^2\|\mathbf{a}_i\|^2\right)^{1/2} \\
&= O_p\left(\frac{1}{\sqrt{NT}}\right)
\end{aligned}$$

by Assumptions 2 and 6(b). Therefore,

$$\left\| \frac{1}{N}\sum_{i=1}^{N}\left[\frac{1}{N}\sum_{l=1}^{N}\left(\widehat{\mathbf{a}}_l - \widehat{\mathbf{H}}'\mathbf{a}_l\right)\varphi_{jil}\right]\mathbf{a}_i' \right\| = O_p\left(\frac{1}{C_{NT}}\right)O_p\left(\frac{1}{\sqrt{NT}}\right) = O_p\left(\frac{1}{\sqrt{NT}C_{NT}}\right)$$

by Lemma 1. Further

$$
\begin{aligned}
\left\|\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{N}\sum_{l=1}^{N}\mathbf{a}_l\mathbf{a}_i'\varphi_{jil}\right)\right\| &= \left\|\frac{1}{N^2}\sum_{i=1}^{N}\sum_{l=1}^{N}\mathbf{a}_l\mathbf{a}_i'\left(\frac{1}{T}\sum_{t=1}^{T}\mathbb{I}_{jt}\boldsymbol{\lambda}_{ji}'\mathbf{f}_{jt}e_{lt}\right)\right\| \\
&= \left\|\frac{1}{N^2}\sum_{i=1}^{N}\sum_{l=1}^{N}\begin{pmatrix}\boldsymbol{\lambda}_{1l}\\ \boldsymbol{\lambda}_{2l}\end{pmatrix}\left(\frac{1}{T}\sum_{t=1}^{T}\mathbb{I}_{jt}\boldsymbol{\lambda}_{ji}'\mathbf{f}_{jt}e_{lt}\right)\begin{pmatrix}\boldsymbol{\lambda}_{1i}\\ \boldsymbol{\lambda}_{2i}\end{pmatrix}'\right\| \\
&\leq \frac{1}{\sqrt{NT}}\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{\sqrt{NT}}\sum_{l=1}^{N}\begin{pmatrix}\boldsymbol{\lambda}_{1l}\\ \boldsymbol{\lambda}_{2l}\end{pmatrix}\left(\frac{1}{T}\sum_{t=1}^{T}\mathbb{I}_{jt}\mathbf{f}_{jt}'e_{lt}\right)\right\|\|\boldsymbol{\lambda}_{ji}\|\left\|\begin{pmatrix}\boldsymbol{\lambda}_{1i}\\ \boldsymbol{\lambda}_{2i}\end{pmatrix}'\right\| \\
&= O_p\left(\frac{1}{\sqrt{NT}}\right)
\end{aligned}
$$

by Assumptions 2 and 6(b). Therefore

$$
\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{N}\sum_{l=1}^{N}\widehat{\mathbf{a}}_l\varphi_{jil}\right)\mathbf{a}_i' = O_p\left(\frac{1}{\sqrt{NT}C_{NT}}\right)+O_p\left(\frac{1}{\sqrt{NT}}\right)=O_p\left(\frac{1}{\sqrt{NT}}\right). \tag{A.13}
$$

Finally,

$$
\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{N}\sum_{l=1}^{N}\widehat{\mathbf{a}}_l\varphi_{jli}\right)\mathbf{a}_i' = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{1}{N}\sum_{l=1}^{N}\left(\widehat{\mathbf{a}}_l-\widehat{\mathbf{H}}'\mathbf{a}_l\right)\varphi_{jli}\right]\mathbf{a}_i'+\widehat{\mathbf{H}}'\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{N}\sum_{l=1}^{N}\mathbf{a}_l\mathbf{a}_i'\varphi_{jli}\right).
$$

We have

$$
\left\|\frac{1}{N}\sum_{i=1}^{N}\left[\frac{1}{N}\sum_{l=1}^{N}\left(\widehat{\mathbf{a}}_l-\widehat{\mathbf{H}}'\mathbf{a}_l\right)\varphi_{jli}\right]\mathbf{a}_i'\right\| \leq \left(\frac{1}{N}\sum_{l=1}^{N}\left\|\widehat{\mathbf{a}}_l-\widehat{\mathbf{H}}'\mathbf{a}_l\right\|^2\right)^{1/2}\left(\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{N}\sum_{l=1}^{N}\varphi_{jli}\mathbf{a}_i\right\|^2\right)^{1/2}
$$

with

$$
\begin{aligned}
\left(\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{N}\sum_{l=1}^{N}\varphi_{jli}\mathbf{a}_i\right\|^2\right)^{1/2} &= \left[\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{N}\sum_{l=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T}\mathbb{I}_{jt}\boldsymbol{\lambda}_{jl}'\mathbf{f}_{jt}e_{it}\right)\mathbf{a}_i\right\|^2\right]^{1/2} \\
&= \frac{1}{\sqrt{NT}}\left[\frac{1}{N}\sum_{i=1}^{N}\left\|\left(\frac{1}{\sqrt{NT}}\sum_{l=1}^{N}\sum_{t=1}^{T}\mathbb{I}_{jt}\boldsymbol{\lambda}_{jl}'\mathbf{f}_{jt}e_{it}\right)\mathbf{a}_i\right\|^2\right]^{1/2} \\
&\leq \frac{1}{\sqrt{NT}}\left(\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{\sqrt{NT}}\sum_{l=1}^{N}\sum_{t=1}^{T}\mathbb{I}_{jt}\boldsymbol{\lambda}_{jl}'\mathbf{f}_{jt}e_{it}\right\|^2\|\mathbf{a}_i\|^2\right)^{1/2} \\
&= O_p\left(\frac{1}{\sqrt{NT}}\right)
\end{aligned}
$$

by Assumptions 2 and 6(b). Further

$$
\begin{aligned}
\left\| \frac{1}{N} \sum_{i=1}^{N} \left( \frac{1}{N} \sum_{l=1}^{N} \mathbf{a}_l \mathbf{a}_i' \varphi_{jli} \right) \right\| &= \left\| \frac{1}{N^2} \sum_{i=1}^{N} \sum_{l=1}^{N} \mathbf{a}_l \mathbf{a}_i' \left( \frac{1}{T} \sum_{t=1}^{T} \mathbb{I}_{jt} \boldsymbol{\lambda}_{jl}' \mathbf{f}_{jt} e_{it} \right) \right\| \\
&= \left\| \frac{1}{N^2} \sum_{i=1}^{N} \sum_{l=1}^{N} \begin{pmatrix} \boldsymbol{\lambda}_{1l} \\ \boldsymbol{\lambda}_{2l} \end{pmatrix} \left( \frac{1}{T} \sum_{t=1}^{T} \mathbb{I}_{jt} \boldsymbol{\lambda}_{jl}' \mathbf{f}_{jt} e_{it} \right) \begin{pmatrix} \boldsymbol{\lambda}_{1i}' \\ \boldsymbol{\lambda}_{2i}' \end{pmatrix} \right\| \\
&\le \frac{1}{\sqrt{NT}} \frac{1}{N} \sum_{i=1}^{N} \left\| \frac{1}{\sqrt{NT}} \sum_{l=1}^{N} \begin{pmatrix} \boldsymbol{\lambda}_{1l} \\ \boldsymbol{\lambda}_{2l} \end{pmatrix} \left( \frac{1}{T} \sum_{t=1}^{T} \mathbb{I}_{jt} \mathbf{f}_{jt}' e_{it} \right) \right\| \left\| \boldsymbol{\lambda}_{jl} \right\| \left\| \begin{pmatrix} \boldsymbol{\lambda}_{1i}' \\ \boldsymbol{\lambda}_{2i}' \end{pmatrix} \right\| \\
&= O_p\left( \frac{1}{\sqrt{NT}} \right)
\end{aligned}
$$

by Assumptions 2 and 6(b). Therefore,

$$
\frac{1}{N} \sum_{i=1}^{N} \left( \frac{1}{N} \sum_{l=1}^{N} \widehat{\mathbf{a}}_l \varphi_{jli} \right) \mathbf{a}_i' = O_p\left( \frac{1}{\sqrt{NT} C_{NT}} \right) + O_p\left( \frac{1}{\sqrt{NT}} \right) = O_p\left( \frac{1}{\sqrt{NT}} \right). \tag{A.14}
$$

Combining equations (A.9) through (A.14), we obtain

$$
N^{-1} \left( \widehat{\mathbf{A}} - \mathbf{A} \widehat{\mathbf{H}} \right)' \mathbf{A} = O_p\left( \frac{1}{\sqrt{N} C_{NT}} \right) + O_p\left( \frac{1}{\sqrt{T} C_{NT}} \right) + O_p\left( \frac{1}{\sqrt{NT}} \right) + O_p\left( \frac{1}{\sqrt{NT}} \right) = O_p\left( \frac{1}{C_{NT}^2} \right). \tag{A.15}
$$

From (A.8), (A.15) and Lemma 1, we obtain

$$
N^{-1} \left( \widehat{\mathbf{A}} - \mathbf{A} \widehat{\mathbf{H}} \right)' \widehat{\mathbf{A}} = O_p\left( \frac{1}{C_{NT}^2} \right) + O_p\left( \frac{1}{C_{NT}^2} \right) = O_p\left( \frac{1}{C_{NT}^2} \right).
$$

which completes the proof of the lemma.

□

***Proof of Lemma 4***. Given the identity in (A.2), we can write

$$
\begin{aligned}
N^{-1} \left( \widehat{\mathbf{A}} - \mathbf{A} \widehat{\mathbf{H}} \right)' \mathbf{e}_t &= \widehat{\mathbf{V}}^{-1} \frac{1}{N} \sum_{i=1}^{N} \left( \widehat{\mathbf{a}}_i - \widehat{\mathbf{H}}' \mathbf{a}_i \right) e_{it} \\
&= \widehat{\mathbf{V}}^{-1} \left\{ \begin{array}{c} \sum_{j=1}^{2} \left[ \frac{1}{N} \sum_{i=1}^{N} \left( \frac{1}{N} \sum_{l=1}^{N} \widehat{\mathbf{a}}_l \sigma_{jil} \right) e_{it} \right] + \sum_{j=1}^{2} \left[ \frac{1}{N} \sum_{i=1}^{N} \left( \frac{1}{N} \sum_{l=1}^{N} \widehat{\mathbf{a}}_l \chi_{jil} \right) e_{it} \right] \\ + \sum_{j=1}^{2} \left[ \frac{1}{N} \sum_{i=1}^{N} \left( \frac{1}{N} \sum_{l=1}^{N} \widehat{\mathbf{a}}_l \varphi_{jil} \right) e_{it} \right] + \sum_{j=1}^{2} \left[ \frac{1}{N} \sum_{i=1}^{N} \left( \frac{1}{N} \sum_{l=1}^{N} \widehat{\mathbf{a}}_l \varphi_{jli} \right) e_{it} \right] \end{array} \right\}.
\end{aligned} \tag{A.16}
$$

Consider

$$\begin{aligned}\left\|\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{N}\sum_{l=1}^{N}\widehat{\mathbf{a}}_l^p\sigma_{jil}\right)e_{it}\right\| &\leq \frac{1}{\sqrt{N}}\left(\frac{1}{N}\sum_{l=1}^{N}\left\|\widehat{\mathbf{a}}_l-\widehat{\mathbf{H}}'\mathbf{a}_l\right\|^2\right)^{1/2}\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{l=1}^{N}|\sigma_{jil}|^2\right)^{1/2}\left(\frac{1}{N}\sum_{i=1}^{N}|e_{it}|^2\right)^{1/2} \\ &= \frac{1}{\sqrt{N}}O_p\left(\frac{1}{C_{NT}}\right)O_p(1)\,O_p(1) \\ &= O_p\left(\frac{1}{\sqrt{N}C_{NT}}\right)\end{aligned} \tag{A.17}$$

by Lemma 1, equation (A.10), and Assumption 3(a). Consider now

$$\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{N}\sum_{l=1}^{N}\widehat{\mathbf{a}}_l\chi_{jil}\right)e_{it}=\frac{1}{N}\sum_{i=1}^{N}\left[\frac{1}{N}\sum_{l=1}^{N}\left(\widehat{\mathbf{a}}_l-\widehat{\mathbf{H}}'\mathbf{a}_l\right)\chi_{jil}\right]e_{it}+\widehat{\mathbf{H}}'\frac{1}{N}\sum_{i=1}^{N}\frac{1}{N}\sum_{l=1}^{N}\mathbf{a}_l e_{it}\chi_{jil}.$$

We have

$$\begin{aligned}\left\|\frac{1}{N}\sum_{i=1}^{N}\left[\frac{1}{N}\sum_{l=1}^{N}\left(\widehat{\mathbf{a}}_l-\widehat{\mathbf{H}}'\mathbf{a}_l\right)\chi_{jil}\right]e_{it}\right\| &\leq \frac{1}{N}\sum_{l=1}^{N}\left\|\left(\widehat{\mathbf{a}}_l-\widehat{\mathbf{H}}'\mathbf{a}_l\right)\right\|\left(\frac{1}{N}\sum_{i=1}^{N}|\chi_{jil}e_{it}|\right) \\ &\leq \left[\frac{1}{N}\sum_{l=1}^{N}\left\|\left(\widehat{\mathbf{a}}_l-\widehat{\mathbf{H}}'\mathbf{a}_l\right)\right\|^2\right]^{1/2}\left[\frac{1}{N}\sum_{l=1}^{N}\left(\frac{1}{N}\sum_{i=1}^{N}|\chi_{jil}e_{it}|\right)^2\right]^{1/2}\end{aligned}$$

with

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^{N}|\chi_{jil}e_{it}| &= \frac{1}{N}\sum_{i=1}^{N}\left|\left[\frac{1}{T}\sum_{t=1}^{T}\mathbb{I}_{jt}e_{it}e_{lt}-\mathsf{E}\left(\frac{1}{T}\sum_{t=1}^{T}\mathbb{I}_{jt}e_{it}e_{lt}\right)\right]e_{it}\right| \\ &= \frac{1}{\sqrt{T}}\frac{1}{N}\sum_{i=1}^{N}\left|\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbb{I}_{jt}e_{it}e_{lt}-\mathsf{E}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbb{I}_{jt}e_{it}e_{lt}\right)\right]e_{it}\right| \\ &= O_p\left(\frac{1}{\sqrt{T}}\right)\end{aligned}$$

by Assumptions 3(a) and 3(c). Therefore, taking into account Lemma 1,

$$\left\|\frac{1}{N}\sum_{i=1}^{N}\left[\frac{1}{N}\sum_{l=1}^{N}\left(\widehat{\mathbf{a}}_l-\widehat{\mathbf{H}}'\mathbf{a}_l\right)\chi_{jil}\right]e_{it}\right\|=O_p\left(\frac{1}{C_{NT}}\right)O_p\left(\frac{1}{\sqrt{T}}\right)=O_p\left(\frac{1}{\sqrt{T}C_{NT}}\right).$$

Further,

$$\begin{aligned}\left\|\frac{1}{N}\sum_{i=1}^{N}\frac{1}{N}\sum_{l=1}^{N}\mathbf{a}_l\chi_{jil}e_{it}\right\| &= \left\|\frac{1}{N}\sum_{i=1}^{N}\frac{1}{N}\sum_{l=1}^{N}\mathbf{a}_l\left[\frac{1}{T}\sum_{t=1}^{T}\mathbb{I}_{jt}e_{it}e_{lt}-\mathsf{E}\left(\frac{1}{T}\sum_{t=1}^{T}\mathbb{I}_{jt}e_{it}e_{lt}\right)\right]e_{it}\right\| \\ &= \frac{1}{\sqrt{NT}}\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{\sqrt{NT}}\sum_{l=1}^{N}\sum_{t=1}^{T}\mathbf{a}_l\left[\mathbb{I}_{jt}e_{it}e_{lt}-\mathsf{E}\left(\mathbb{I}_{jt}e_{it}e_{lt}\right)\right]\right\||e_{it}| \\ &\leq \frac{1}{\sqrt{NT}}\left\{\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{\sqrt{NT}}\sum_{l=1}^{N}\sum_{t=1}^{T}\mathbf{a}_l\left[\mathbb{I}_{jt}e_{it}e_{lt}-\mathsf{E}\left(\mathbb{I}_{jt}e_{it}e_{lt}\right)\right]\right\|^2\right\}^{1/2}\left(\frac{1}{N}\sum_{i=1}^{N}|e_{it}|^2\right)^{1/2} \\ &= O_p\left(\frac{1}{\sqrt{NT}}\right)\end{aligned}$$

by Assumptions 3(a) and 6(a). Therefore,

$$\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{N}\sum_{l=1}^{N}\widehat{\mathbf{a}}_l\chi_{jil}\right)e_{it}=O_p\left(\frac{1}{\sqrt{T}C_{NT}}\right)+O_p\left(\frac{1}{\sqrt{NT}}\right)=O_p\left(\frac{1}{\sqrt{T}C_{NT}}\right). \qquad \text{(A.18)}$$

Consider now

$$\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{N}\sum_{l=1}^{N}\widehat{\mathbf{a}}_l\varphi_{jil}\right)e_{it}=\frac{1}{N}\sum_{i=1}^{N}\left[\frac{1}{N}\sum_{l=1}^{N}\left(\widehat{\mathbf{a}}_l-\widehat{\mathbf{H}}'\mathbf{a}_l\right)\varphi_{jil}\right]e_{it}+\frac{1}{N}\sum_{i=1}^{N}\frac{1}{N}\sum_{l=1}^{N}\widehat{\mathbf{H}}'\mathbf{a}_l\varphi_{jil}e_{it}.$$

We have

$$\begin{aligned}\left\|\frac{1}{N}\sum_{i=1}^{N}\left[\frac{1}{N}\sum_{l=1}^{N}\left(\widehat{\mathbf{a}}_l-\widehat{\mathbf{H}}'\mathbf{a}_l\right)\varphi_{jil}\right]e_{it}\right\| &\leq \frac{1}{N}\sum_{l=1}^{N}\left\|\widehat{\mathbf{a}}_l-\widehat{\mathbf{H}}'\mathbf{a}_l\right\|\left(\frac{1}{N}\sum_{i=1}^{N}|\varphi_{jil}e_{it}|\right)\\ &\leq \left(\frac{1}{N}\sum_{l=1}^{N}\left\|\widehat{\mathbf{a}}_l-\widehat{\mathbf{H}}'\mathbf{a}_l\right\|^2\right)^{1/2}\left[\frac{1}{N}\sum_{l=1}^{N}\left(\frac{1}{N}\sum_{i=1}^{N}|\varphi_{jil}e_{it}|\right)^2\right]^{1/2},\end{aligned}$$

with

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^{N}|\varphi_{jil}e_{it}| &= \frac{1}{N}\sum_{i=1}^{N}\left|\left(\frac{1}{T}\sum_{t=1}^{T}\mathbb{I}_{jt}\boldsymbol{\lambda}'_{ji}\mathbf{f}_{jt}e_{lt}\right)e_{it}\right|\\ &\leq \frac{1}{\sqrt{T}}\frac{1}{N}\sum_{i=1}^{N}\|\boldsymbol{\lambda}_{ji}\|\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbb{I}_{jt}\mathbf{f}_{jt}e_{it}\right\||e_{lt}|\\ &\leq \bar{\lambda}\frac{1}{\sqrt{T}}|e_{lt}|\left(\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbb{I}_{jt}\mathbf{f}_{jt}e_{it}\right\|^2\right)^{1/2}\\ &= O_p\left(\frac{1}{\sqrt{T}}\right)\end{aligned}$$

by Assumptions 2, 3(a) and 4. Taking into account Lemma 1,

$$\frac{1}{N}\sum_{i=1}^{N}\left[\frac{1}{N}\sum_{l=1}^{N}\left(\widehat{\mathbf{a}}_l-\widehat{\mathbf{H}}'\mathbf{a}_l\right)\varphi_{jil}\right]e_{it}=O_p\left(\frac{1}{C_{NT}}\right)O_p\left(\frac{1}{\sqrt{T}}\right)=O_p\left(\frac{1}{\sqrt{T}C_{NT}}\right).$$

Further,

$$\begin{aligned}\left\|\frac{1}{N}\sum_{i=1}^{N}\frac{1}{N}\sum_{l=1}^{N}\mathbf{a}_l\varphi_{jil}e_{it}\right\| &= \frac{1}{\sqrt{NT}}\left\|\frac{1}{N}\sum_{l=1}^{N}\mathbf{a}_l\left(\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{I}_{jt}\boldsymbol{\lambda}'_{ji}\mathbf{f}_{jt}e_{it}\right)e_{lt}\right\|\\ &\leq \frac{1}{\sqrt{NT}}\left(\frac{1}{N}\sum_{l=1}^{N}\|\mathbf{a}_l\||e_{lt}|\right)\left\|\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{I}_{jt}\boldsymbol{\lambda}'_{ji}\mathbf{f}_{jt}e_{it}\right\|\\ &= \frac{1}{\sqrt{NT}}\left(\frac{1}{N}\sum_{l=1}^{N}\|\mathbf{a}_l\|^2\right)^{1/2}\left(\frac{1}{N}\sum_{l=1}^{N}|e_{lt}|^2\right)^{1/2}\left\|\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{I}_{jt}\boldsymbol{\lambda}'_{ji}\mathbf{f}_{jt}e_{it}\right\|\\ &= O_p\left(\frac{1}{\sqrt{NT}}\right),\end{aligned}$$

by Assumptions 2, 3(a) and 6(a). Therefore,

$$\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{N}\sum_{l=1}^{N}\widehat{\mathbf{a}}_l\varphi_{jil}\right)e_{it} = O_p\left(\frac{1}{\sqrt{T}C_{NT}}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right) = O_p\left(\frac{1}{\sqrt{T}C_{NT}}\right). \quad \text{(A.19)}$$

Finally,

$$\begin{aligned}\left\|\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{N}\sum_{l=1}^{N}\widehat{\mathbf{a}}_l\varphi_{jli}\right)e_{it}\right\| &\leq \left(\frac{1}{N}\sum_{i=1}^{N}\left\|\left(\frac{1}{N}\sum_{l=1}^{N}\widehat{\mathbf{a}}_l\varphi_{jli}\right)\right\||e_{it}|\right)\\ &\leq \left(\frac{1}{N}\sum_{i=1}^{N}|e_{it}|^2\right)^{1/2}\left(\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{N}\sum_{l=1}^{N}\widehat{\mathbf{a}}_l\varphi_{jli}\right\|^2\right)^{1/2} \quad \text{(A.20)}\\ &= O_p\left(\frac{1}{\sqrt{T}C_{NT}}\right).\end{aligned}$$

by Assumption 3(a) and Lemma 2(d). By combining (A.16), (A.17), (A.18), (A.19) and (A.20), we have

$$N^{-1}\left(\widehat{\mathbf{A}} - \mathbf{A}\widehat{\mathbf{H}}\right)'\mathbf{e}_t = O_p\left(\frac{1}{\sqrt{N}C_{NT}}\right) + O_p\left(\frac{1}{\sqrt{T}C_{NT}}\right) = O_p\left(\frac{1}{C_{NT}^2}\right).$$

□

***Proof of Lemma 5***. Starting from (a), and taking into account (10), consider

$$\widehat{\mathbf{g}}_t = N^{-1}\widehat{\mathbf{A}}'\mathbf{x}_t = N^{-1}\widehat{\mathbf{A}}'\left(\mathbf{A}\mathbf{g}_t + \mathbf{e}_t\right) = N^{-1}\widehat{\mathbf{A}}'\mathbf{A}\mathbf{g}_t + N^{-1}\widehat{\mathbf{A}}'\mathbf{e}_t$$

and note that

$$\mathbf{A} = \mathbf{A} - \widehat{\mathbf{A}}\widehat{\mathbf{H}}^{-1} + \widehat{\mathbf{A}}\widehat{\mathbf{H}}^{-1}$$

so that we have

$$\begin{aligned}\widehat{\mathbf{g}}_t &= N^{-1}\widehat{\mathbf{A}}'\left(\mathbf{A} - \widehat{\mathbf{A}}\widehat{\mathbf{H}}^{-1} + \widehat{\mathbf{A}}\widehat{\mathbf{H}}^{-1}\right)\mathbf{g}_t + N^{-1}\widehat{\mathbf{A}}'\mathbf{e}_t\\ &= N^{-1}\widehat{\mathbf{A}}'\left(\mathbf{A} - \widehat{\mathbf{A}}\widehat{\mathbf{H}}^{-1} + \widehat{\mathbf{A}}\widehat{\mathbf{H}}^{-1}\right)\mathbf{g}_t + N^{-1}\widehat{\mathbf{A}}'\mathbf{e}_t + N^{-1}\left(\mathbf{A}\widehat{\mathbf{H}}\right)'\mathbf{e}_t - N^{-1}\left(\mathbf{A}\widehat{\mathbf{H}}\right)'\mathbf{e}_t\\ &= N^{-1}\widehat{\mathbf{A}}'\left(\mathbf{A} - \widehat{\mathbf{A}}\widehat{\mathbf{H}}^{-1}\right)\mathbf{g}_t + N^{-1}\widehat{\mathbf{A}}'\widehat{\mathbf{A}}\widehat{\mathbf{H}}^{-1}\mathbf{g}_t + N^{-1}\left(\widehat{\mathbf{A}} - \mathbf{A}\widehat{\mathbf{H}}\right)'\mathbf{e}_t + N^{-1}\left(\mathbf{A}\widehat{\mathbf{H}}\right)'\mathbf{e}_t,\end{aligned}$$

which leads to

$$\widehat{\mathbf{g}}_t - \widehat{\mathbf{H}}^{-1}\mathbf{g}_t = N^{-1}\left(\mathbf{A}\widehat{\mathbf{H}}\right)'\mathbf{e}_t + N^{-1}\widehat{\mathbf{A}}'\left(\mathbf{A} - \widehat{\mathbf{A}}\widehat{\mathbf{H}}^{-1}\right)\mathbf{g}_t + N^{-1}\left(\widehat{\mathbf{A}} - \mathbf{A}\widehat{\mathbf{H}}\right)'\mathbf{e}_t. \quad \text{(A.21)}$$

The result in (a) follows by taking into account Lemma 3 and Lemma 4. As for (b), adding and subtracting terms we have

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T}\left(\widehat{\mathbf{g}}_t-\widehat{\mathbf{H}}^{-1}\mathbf{g}_t\right)\widehat{\mathbf{g}}_t' &= \frac{1}{T}\sum_{t=1}^{T}\left(\widehat{\mathbf{g}}_t-\widehat{\mathbf{H}}^{-1}\mathbf{g}_t\right)\left(\widehat{\mathbf{g}}_t-\widehat{\mathbf{H}}^{-1}\mathbf{g}_t\right)' \\ &\quad +\frac{1}{T}\sum_{t=1}^{T}\left(\widehat{\mathbf{g}}_t-\widehat{\mathbf{H}}^{-1}\mathbf{g}_t\right)\mathbf{g}_t'\left(\widehat{\mathbf{H}}^{-1}\right)'.\end{aligned} \tag{A.22}$$

Taking into account the results in (a), it follows that

$$\frac{1}{T}\sum_{t=1}^{T}\left(\widehat{\mathbf{g}}_t-\widehat{\mathbf{H}}^{-1}\mathbf{g}_t\right)\left(\widehat{\mathbf{g}}_t-\widehat{\mathbf{H}}^{-1}\mathbf{g}_t\right)' = O_p\left(\frac{1}{N}\right)+O_p\left(\frac{1}{\sqrt{N}C_{NT}^2}\right)+O_p\left(\frac{1}{C_{NT}^4}\right). \tag{A.23}$$

From (A.21), we also have that

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T}\left(\widehat{\mathbf{g}}_t-\widehat{\mathbf{H}}^{-1}\mathbf{g}_t\right)\mathbf{g}_t' &= \frac{1}{T}\sum_{t=1}^{T}\left[\frac{1}{N}\left(\mathbf{A}\widehat{\mathbf{H}}\right)'\mathbf{e}_t+\frac{1}{N}\hat{\mathbf{A}}'\left(\mathbf{A}-\widehat{\mathbf{A}}\widehat{\mathbf{H}}^{-1}\right)\mathbf{g}_t+\frac{1}{N}\left(\widehat{\mathbf{A}}-\mathbf{A}\widehat{\mathbf{H}}\right)'\mathbf{e}_t\right]\mathbf{g}_t' \\ &= \frac{1}{\sqrt{NT}}\widehat{\mathbf{H}}'\frac{\mathbf{A}'}{\sqrt{N}}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{e}_t\mathbf{g}_t'\right)+\frac{\hat{\mathbf{A}}'\left(\mathbf{A}-\widehat{\mathbf{A}}\widehat{\mathbf{H}}^{-1}\right)}{N}\frac{1}{T}\sum_{t=1}^{T}\mathbf{g}_t\mathbf{g}_t' \\ &\quad +\frac{1}{\sqrt{NT}}\left(\frac{\widehat{\mathbf{A}}-\mathbf{A}\widehat{\mathbf{H}}}{\sqrt{N}}\right)'\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{e}_t\mathbf{g}_t'\right),\end{aligned}$$

and taking into account Assumptions 2 and 6(c), and Lemma 3,

$$\begin{aligned}\left\|\frac{1}{T}\sum_{t=1}^{T}\left(\widehat{\mathbf{g}}_t-\widehat{\mathbf{H}}^{-1}\mathbf{g}_t\right)\mathbf{g}_t'\right\| &= \frac{1}{\sqrt{NT}}\left\|\widehat{\mathbf{H}}\right\|\left\|\frac{\mathbf{A}}{\sqrt{N}}\right\|\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{e}_t\mathbf{g}_t'\right\| \\ &\quad +\left\|\frac{\hat{\mathbf{A}}'\left(\mathbf{A}-\widehat{\mathbf{A}}\widehat{\mathbf{H}}^{-1}\right)}{N}\right\|\left\|\frac{1}{T}\sum_{t=1}^{T}\mathbf{g}_t\mathbf{g}_t'\right\| \\ &\quad +\frac{1}{\sqrt{NT}}\left\|\frac{\widehat{\mathbf{A}}-\mathbf{A}\widehat{\mathbf{H}}}{\sqrt{N}}\right\|\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{e}_t\mathbf{g}_t'\right\| \\ &= O_p\left(\frac{1}{\sqrt{NT}}\right)+O_p\left(\frac{1}{C_{NT}^2}\right)+O_p\left(\frac{1}{\sqrt{NT}}\right) \\ &= O_p\left(\frac{1}{C_{NT}^2}\right).\end{aligned} \tag{A.24}$$

Combining (A.22) through (A.24), it follows that

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T}\left(\widehat{\mathbf{g}}_t-\widehat{\mathbf{H}}^{-1}\mathbf{g}_t\right)\widehat{\mathbf{g}}_t' &= O_p\left(\frac{1}{N}\right)+O_p\left(\frac{1}{\sqrt{N}C_{NT}^2}\right)+O_p\left(\frac{1}{C_{NT}^4}\right)+O_p\left(\frac{1}{C_{NT}^2}\right) \\ &= O_p\left(\frac{1}{C_{NT}^2}\right),\end{aligned}$$

which shows (b) and completes the proof of the lemma. □

***Proof of Lemma 6.*** Given $\widehat{\mathbf{H}} = (\mathbf{GG}'/T)\left(\mathbf{A}'\widehat{\mathbf{A}}/N\right)\widehat{\mathbf{V}}^{-1}$, pre-multiply both sides of the identity $(1/NT)\,\mathbf{X}'\mathbf{X}\widehat{\mathbf{A}} = \widehat{\mathbf{A}}\widehat{\mathbf{V}}$ by $(\mathbf{GG}'/T)^{1/2}\,N^{-1}\mathbf{A}'$ to obtain

$$\frac{1}{N}\left(\frac{\mathbf{GG}'}{T}\right)^{1/2}\mathbf{A}'\left(\frac{\mathbf{X}'\mathbf{X}}{NT}\right)\widehat{\mathbf{A}} = \left(\frac{\mathbf{GG}'}{T}\right)^{1/2}\left(\frac{\mathbf{A}'\widehat{\mathbf{A}}}{N}\right)\widehat{\mathbf{V}}.$$

Given (10), write $\mathbf{X} = \mathbf{G}'\mathbf{A}' + \mathbf{E}$ with $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_T)'$ and $\mathbf{E} = (\mathbf{e}_1, \ldots, \mathbf{e}_T)'$. We thus have

$$\frac{1}{N}\left(\frac{\mathbf{GG}'}{T}\right)^{1/2}\mathbf{A}'\left(\frac{\mathbf{AGG}'\mathbf{A}'}{NT}\right)\widehat{\mathbf{A}} + \widehat{\mathbf{D}} = \left(\frac{\mathbf{GG}'}{T}\right)^{1/2}\left(\frac{\mathbf{A}'\widehat{\mathbf{A}}}{N}\right)\widehat{\mathbf{V}}, \tag{A.25}$$

where

$$\widehat{\mathbf{D}} = \frac{1}{N}\left(\frac{\mathbf{GG}'}{T}\right)^{1/2}\mathbf{A}'\left(\frac{\mathbf{AGE} + \mathbf{E}'\mathbf{G}'\mathbf{A}' + \mathbf{E}'\mathbf{E}}{NT}\right)\widehat{\mathbf{A}} = o_p(1)$$

by Lemma 2. Let

$$\mathbf{W} = \left(\frac{\mathbf{GG}'}{T}\right)^{1/2}\left(\frac{\mathbf{A}'\mathbf{A}}{N}\right)\left(\frac{\mathbf{GG}'}{T}\right)^{1/2}, \quad \widehat{\mathbf{Z}} = \left(\frac{\mathbf{GG}'}{T}\right)^{1/2}\left(\frac{\mathbf{A}'\widehat{\mathbf{A}}}{N}\right),$$

so that we can write (A.25) as

$$\left(\mathbf{W} + \widehat{\mathbf{D}}\widehat{\mathbf{Z}}^{-1}\right)\widehat{\mathbf{Z}} = \widehat{\mathbf{Z}}\widehat{\mathbf{V}}.$$

Therefore, each column of $\widehat{\mathbf{Z}}$ is an eigenvector of $\left(\mathbf{W} + \widehat{\mathbf{D}}\widehat{\mathbf{Z}}^{-1}\right)$, although the length is not equal to unity. Let $\widehat{\mathbf{V}}^*$ be the diagonal matrix of the diagonal elements of $\widehat{\mathbf{Z}}'\widehat{\mathbf{Z}}$. Define $\widehat{\mathbf{\Psi}} = \widehat{\mathbf{Z}}\left(\widehat{\mathbf{V}}^*\right)^{-1/2}$ so that each column of $\widehat{\mathbf{\Psi}}$ has unit length. We thus get

$$\left(\mathbf{W} + \widehat{\mathbf{D}}\widehat{\mathbf{Z}}^{-1}\right)\widehat{\mathbf{\Psi}} = \widehat{\mathbf{\Psi}}\widehat{\mathbf{V}},$$

where $\widehat{\mathbf{\Psi}}$ is the eigenvector matrix of $\left(\mathbf{W} + \widehat{\mathbf{D}}\widehat{\mathbf{Z}}^{-1}\right)$. Consider

$$\mathbf{W} + \widehat{\mathbf{D}}\widehat{\mathbf{Z}}^{-1} = \left(\frac{\mathbf{GG}'}{T}\right)^{1/2}\left(\frac{\mathbf{A}'\mathbf{A}}{N}\right)\left(\frac{\mathbf{GG}'}{T}\right)^{1/2} + \widehat{\mathbf{D}}\widehat{\mathbf{Z}}^{-1},$$

and note that

$$\frac{\mathbf{GG}'}{T} = \frac{1}{T}\sum_{t=1}^{T}\begin{pmatrix}\mathbb{I}_{1t}\mathbf{f}_{1t}\\ \mathbb{I}_{2t}\mathbf{f}_{2t}\end{pmatrix}\begin{pmatrix}\mathbb{I}_{1t}\mathbf{f}_{1t}\\ \mathbb{I}_{2t}\mathbf{f}_{2t}\end{pmatrix}' = \frac{1}{T}\sum_{t=1}^{T}\begin{pmatrix}\mathbb{I}_{1t}\mathbf{f}_{1t}\mathbf{f}_{1t}' & \mathbf{0}\\ \mathbf{0} & \mathbb{I}_{2t}\mathbf{f}_{2t}\mathbf{f}_{2t}'\end{pmatrix} \xrightarrow{p} \begin{pmatrix}\mathbf{\Sigma}_{\mathbf{f}_1} & \mathbf{0}\\ \mathbf{0} & \mathbf{\Sigma}_{\mathbf{f}_2}\end{pmatrix} = \mathbf{\Sigma}_{\mathbf{g}}$$

by Assumption 1. Further, $(\mathbf{A}'\mathbf{A}/N) \to \mathbf{\Sigma}_{\mathbf{A}}$ by Assumption 2. Therefore, by Assumptions 1 and 2, $\mathbf{W} + \widehat{\mathbf{D}}\widehat{\mathbf{Z}}^{-1} \xrightarrow{p} \mathbf{\Sigma}_{\mathbf{g}}^{1/2}\mathbf{\Sigma}_{\mathbf{A}}\mathbf{\Sigma}_{\mathbf{g}}^{1/2}$. Because the eigenvalues of $\mathbf{\Sigma}_{\mathbf{g}}^{1/2}\mathbf{\Sigma}_{\mathbf{A}}\mathbf{\Sigma}_{\mathbf{g}}^{1/2}$ are distinct by Assumption 5, the eigenvalues of $\mathbf{W}+\widehat{\mathbf{D}}\widehat{\mathbf{Z}}^{-1}$ are also distinct for large $N$ and $T$, by the continuity of eigenvalues. This implies that the eigenvector matrix of $\mathbf{W}+\widehat{\mathbf{D}}\widehat{\mathbf{Z}}^{-1}$ is unique except for the fact that each column can be replaced by its negative value. Further, the $p-th$ column of $\widehat{\mathbf{Z}}$ depends on $\widehat{\mathbf{A}}$ only through the $p-th$ column of $\widehat{\mathbf{A}}$, for $p=1,\ldots,P$. This implies that the sign of each column in $\widehat{\mathbf{Z}}$, and thus in $\widehat{\boldsymbol{\Psi}}=\widehat{\mathbf{Z}}\left(\widehat{\mathbf{V}}^{*}\right)^{-1/2}$, is determined by the sign of the corresponding column of $\widehat{\mathbf{A}}$. Therefore, the column sign of $\widehat{\mathbf{A}}$ and $\widehat{\boldsymbol{\Psi}}$ are uniquely determined. By the eigenvector perturbation theory, which requires the eigenvalues to be distinct, there exists a unique eigenvector matrix $\boldsymbol{\Psi}$ of $\boldsymbol{\Sigma}_{\mathbf{A}}^{1/2}\boldsymbol{\Sigma}_{\mathbf{g}}^{1/2}\boldsymbol{\Sigma}_{\mathbf{A}}^{1/2}$ such that $\left\|\widehat{\boldsymbol{\Psi}}-\boldsymbol{\Psi}\right\|=o_p(1)$. Since $\widehat{\boldsymbol{\Psi}}=\widehat{\mathbf{Z}}\left(\widehat{\mathbf{V}}^{*}\right)^{-1/2}$ and $\widehat{\mathbf{Z}}=(\mathbf{G}\mathbf{G}'/T)^{1/2}\left(\mathbf{A}'\widehat{\mathbf{A}}/N\right)$ then $\widehat{\boldsymbol{\Psi}}=(\mathbf{G}\mathbf{G}'/T)^{1/2}\left(\mathbf{A}'\widehat{\mathbf{A}}/N\right)\left(\widehat{\mathbf{V}}^{*}\right)^{-1/2}$, which implies that

$$\frac{\mathbf{A}'\widehat{\mathbf{A}}}{N}=\left(\frac{\mathbf{G}\mathbf{G}'}{T}\right)^{-1/2}\widehat{\boldsymbol{\Psi}}\left(\widehat{\mathbf{V}}^{*}\right)^{1/2}\xrightarrow{p}\boldsymbol{\Sigma}_{\mathbf{g}}^{-1/2}\boldsymbol{\Psi}\mathbf{V}^{1/2}$$

by Assumption 1 and since $\widehat{\mathbf{V}}^{*}\xrightarrow{p}\mathbf{V}$, the latter following from arguments analogous to those in the proof of Proposition 1 in Bai (2003). This completes the proof of the lemma. □

***Proof of Lemma 7***. From Lemma 6, and taking into account (41), we have

$$\begin{aligned}\mathbf{Q} &= \begin{pmatrix}\mathbf{Q}_1\\ \mathbf{Q}_2\end{pmatrix}\\ &= \boldsymbol{\Sigma}_{\mathbf{g}}^{-1/2}\boldsymbol{\Psi}\mathbf{V}^{1/2}\\ &= \begin{pmatrix}\boldsymbol{\Sigma}_{\mathbf{f}1} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{\Sigma}_{\mathbf{f}2}\end{pmatrix}^{-1/2}\boldsymbol{\Psi}\mathbf{V}^{1/2}\\ &= \begin{pmatrix}\boldsymbol{\Sigma}_{\mathbf{f}1}^{-1/2} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{\Sigma}_{\mathbf{f}2}^{-1/2}\end{pmatrix}\begin{pmatrix}\boldsymbol{\Psi}_1\\ \boldsymbol{\Psi}_2\end{pmatrix}\mathbf{V}^{1/2}\\ &= \begin{pmatrix}\boldsymbol{\Sigma}_{\mathbf{f}1}^{-1/2}\boldsymbol{\Psi}_1\mathbf{V}^{1/2}\\ \boldsymbol{\Sigma}_{\mathbf{f}2}^{-1/2}\boldsymbol{\Psi}_2\mathbf{V}^{1/2}\end{pmatrix}\end{aligned},$$

which completes the proof of the lemma. □

***Proof of Lemma 8***. Given the equivalent linear representation in (10), we can write

$$\begin{aligned}\frac{1}{NT}\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t &= \frac{1}{NT}\sum_{t=1}^{T}(\mathbf{A}\mathbf{g}_t+\mathbf{e}_t)(\mathbf{A}\mathbf{g}_t+\mathbf{e}_t)'\\ &= \frac{\mathbf{A}}{\sqrt{N}}\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{g}_t\mathbf{g}_t'\right)\frac{\mathbf{A}}{\sqrt{N}}'+\frac{\mathbf{A}}{N}\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{g}_t\mathbf{e}_t'\right)\\ &+\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{e}_t\mathbf{g}_t'\right)\frac{\mathbf{A}'}{N}+\frac{1}{NT}\sum_{t=1}^{T}\mathbf{e}_t\mathbf{e}_t'.\end{aligned} \tag{A.26}$$

Taking into account Assumption 2 and Assumption 4, it follows that

$$\begin{aligned}\left\|\frac{\mathbf{A}}{N}\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{g}_t\mathbf{e}_t'\right)\right\| &\le \frac{1}{\sqrt{NT}}\left\|\frac{\mathbf{A}}{\sqrt{N}}\right\|\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\begin{pmatrix}\mathbb{I}_{1t}\mathbf{f}_{1t}\mathbf{e}_t'\\ \mathbb{I}_{2t}\mathbf{f}_{2t}\mathbf{e}_t'\end{pmatrix}\right\| \\ &= \frac{1}{\sqrt{NT}}O_p(1)\,O_p\left(\sqrt{N}\right) \\ &= O_p\left(\frac{1}{\sqrt{T}}\right).\end{aligned} \tag{A.27}$$

Similarly, we can prove that

$$\frac{1}{N}\mathbf{A}\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{e}_t\mathbf{g}_t'\right) = O_p\left(\frac{1}{\sqrt{T}}\right). \tag{A.28}$$

Finally, by the weak dependence condition in Assumption (3),

$$\left\|\frac{1}{NT}\sum_{t=1}^{T}\mathbf{e}_t\mathbf{e}_t'\right\| = o_p(1). \tag{A.29}$$

By combining (A.26) through (A.29), we then have

$$\frac{1}{NT}\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t = \frac{\mathbf{A}}{\sqrt{N}}\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{g}_t\mathbf{g}_t'\right)\frac{\mathbf{A}'}{\sqrt{N}} + o_p(1) = \frac{\mathbf{A}}{\sqrt{N}}\frac{\mathbf{G}\mathbf{G}'}{T}\frac{\mathbf{A}'}{\sqrt{N}} + o_p(1).$$

The result in the lemma follows from Assumptions (1) and (2) by noting that the eigenvalues of $\left(\mathbf{A}\big/\sqrt{N}\right)(\mathbf{G}\mathbf{G}'/T)\left(\mathbf{A}'\big/\sqrt{N}\right)$ are the same as those of $\left(\mathbf{G}'\big/\sqrt{T}\right)(\mathbf{A}'\mathbf{A}/N)\left(\mathbf{G}\big/\sqrt{T}\right)$. □

## A.3 Proof of Theorem 1

Given (1), from Section 2.2 recall $\mathbf{B}_1 = [\boldsymbol{\Lambda}_1\ \mathbf{0}]$ and $\mathbf{B}_2 = [\mathbf{0}\ \boldsymbol{\Lambda}_2]$. Adding and subtracting terms, we have

$$\begin{aligned}\mathbf{x}_t &= \mathbb{I}_{1t}\mathbf{B}_1\mathbf{g}_t + \mathbb{I}_{2t}\mathbf{B}_2\mathbf{g}_t + \mathbf{e}_t \\ &= \mathbb{I}_{1t}\mathbf{B}_1\widehat{\mathbf{H}}\widehat{\mathbf{g}}_t + \mathbb{I}_{2t}\mathbf{B}_2\widehat{\mathbf{H}}\widehat{\mathbf{g}}_t + \mathbb{I}_{1t}\mathbf{B}_1\widehat{\mathbf{H}}\left(\widehat{\mathbf{H}}^{-1}\mathbf{g}_t - \widehat{\mathbf{g}}_t\right) + \mathbb{I}_{2t}\mathbf{B}_2\widehat{\mathbf{H}}\left(\widehat{\mathbf{H}}^{-1}\mathbf{g}_t - \widehat{\mathbf{g}}_t\right) + \mathbf{e}_t.\end{aligned}$$

We focus upon $\widehat{\mathbf{B}}_1 = \left[\widehat{\mathbf{b}}_{11}, \ldots, \widehat{\mathbf{b}}_{1N}\right]'$ as an estimator for $\mathbf{B}_1 = [\mathbf{b}_{11}, \ldots, \mathbf{b}_{1N}]'$. Analogous arguments hold for $\widehat{\mathbf{B}}_2$. We have

$$
\begin{aligned}
\widehat{\mathbf{B}}_1 &= \left(\sum_{t=1}^{T} \widehat{\xi}_{1,t|T}\, \mathbf{x}_t \widehat{\mathbf{g}}_t'\right) \left(\sum_{t=1}^{T} \widehat{\xi}_{1,t|T}\, \widehat{\mathbf{g}}_t \widehat{\mathbf{g}}_t'\right)^{-1} \\
&= \left\{\sum_{t=1}^{T} \widehat{\xi}_{1,t|T} \left[\mathbb{I}_{1t} \mathbf{B}_1 \widehat{\mathbf{H}} \widehat{\mathbf{g}}_t + \mathbb{I}_{2t} \mathbf{B}_2 \widehat{\mathbf{H}} \widehat{\mathbf{g}}_t + \mathbb{I}_{1t} \mathbf{B}_1 \widehat{\mathbf{H}} \left(\widehat{\mathbf{H}}^{-1} \mathbf{g}_t - \widehat{\mathbf{g}}_t\right) + \mathbb{I}_{2t} \mathbf{B}_2 \widehat{\mathbf{H}} \left(\widehat{\mathbf{H}}^{-1} \mathbf{g}_t - \widehat{\mathbf{g}}_t\right) + \mathbf{e}_t\right] \widehat{\mathbf{g}}_t'\right\} \\
&\quad \times \left(\sum_{t=1}^{T} \widehat{\xi}_{1,t|T}\, \widehat{\mathbf{g}}_t \widehat{\mathbf{g}}_t'\right)^{-1} \\
&= \mathbf{B}_1 \widehat{\mathbf{H}} \left(\sum_{t=1}^{T} \widehat{\xi}_{1,t|T}\, \mathbb{I}_{1t} \widehat{\mathbf{g}}_t \widehat{\mathbf{g}}_t'\right) \left(\sum_{t=1}^{T} \widehat{\xi}_{1,t|T}\, \widehat{\mathbf{g}}_t \widehat{\mathbf{g}}_t'\right)^{-1} + \mathbf{B}_2 \widehat{\mathbf{H}} \left(\sum_{t=1}^{T} \widehat{\xi}_{1,t|T}\, \mathbb{I}_{2t} \widehat{\mathbf{g}}_t \widehat{\mathbf{g}}_t'\right) \left(\sum_{t=1}^{T} \widehat{\xi}_{1,t|T}\, \widehat{\mathbf{g}}_t \widehat{\mathbf{g}}_t'\right)^{-1} \\
&\quad + \mathbf{B}_1 \widehat{\mathbf{H}} \left[\sum_{t=1}^{T} \widehat{\xi}_{1,t|T}\, \mathbb{I}_{1t} \left(\widehat{\mathbf{H}}^{-1} \mathbf{g}_t - \widehat{\mathbf{g}}_t\right) \widehat{\mathbf{g}}_t'\right] \left(\sum_{t=1}^{T} \widehat{\xi}_{1,t|T}\, \widehat{\mathbf{g}}_t \widehat{\mathbf{g}}_t'\right)^{-1} \\
&\quad + \mathbf{B}_2 \widehat{\mathbf{H}} \left[\sum_{t=1}^{T} \widehat{\xi}_{1,t|T}\, \mathbb{I}_{2t} \left(\widehat{\mathbf{H}}^{-1} \mathbf{g}_t - \widehat{\mathbf{g}}_t\right) \widehat{\mathbf{g}}_t'\right] \left(\sum_{t=1}^{T} \widehat{\xi}_{1,t|T}\, \widehat{\mathbf{g}}_t \widehat{\mathbf{g}}_t'\right)^{-1} \\
&\quad + \left(\sum_{t=1}^{T} \widehat{\xi}_{1,t|T}\, \mathbf{e}_t \widehat{\mathbf{g}}_t'\right) \left(\sum_{t=1}^{T} \widehat{\xi}_{1,t|T}\, \widehat{\mathbf{g}}_t \widehat{\mathbf{g}}_t'\right)^{-1}.
\end{aligned}
$$

Since $\mathbb{I}_{2t} = 1 - \mathbb{I}_{1t}$, and recalling the definition of $\widehat{\mathbf{I}}_{\widehat{\xi}1}$, after some algebra we get

$$
\begin{aligned}
\sqrt{T}\left[\widehat{\mathbf{B}}_1 - \mathbf{B}_1 \widehat{\mathbf{H}} \widehat{\mathbf{I}}_{\widehat{\xi}1} - \mathbf{B}_2 \widehat{\mathbf{H}} \left(\mathbf{I} - \widehat{\mathbf{I}}_{\widehat{\xi}1}\right)\right] &= \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \widehat{\xi}_{1,t|T}\, \mathbf{e}_t \widehat{\mathbf{g}}_t'\right) \left(\frac{1}{T} \sum_{t=1}^{T} \widehat{\xi}_{1,t|T}\, \widehat{\mathbf{g}}_t \widehat{\mathbf{g}}_t'\right)^{-1} \\
&\quad + \mathbf{B}_1 \widehat{\mathbf{H}} \left[\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \widehat{\xi}_{1,t|T}\, \mathbb{I}_{1t} \left(\widehat{\mathbf{H}}^{-1} \mathbf{g}_t - \widehat{\mathbf{g}}_t\right) \widehat{\mathbf{g}}_t'\right] \\
&\quad \times \left(\frac{1}{T} \sum_{t=1}^{T} \widehat{\xi}_{1,t|T}\, \widehat{\mathbf{g}}_t \widehat{\mathbf{g}}_t'\right)^{-1} \\
&\quad + \mathbf{B}_2 \widehat{\mathbf{H}} \left[\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \widehat{\xi}_{1,t|T}\, \mathbb{I}_{2t} \left(\widehat{\mathbf{H}}^{-1} \mathbf{g}_t - \widehat{\mathbf{g}}_t\right) \widehat{\mathbf{g}}_t'\right] \\
&\quad \times \left(\frac{1}{T} \sum_{t=1}^{T} \widehat{\xi}_{1,t|T}\, \widehat{\mathbf{g}}_t \widehat{\mathbf{g}}_t'\right)^{-1}.
\end{aligned}
\tag{A.30}
$$

For $0 < M < \infty$, and taking into account Lemma 5, for $j = 1, 2$ we have that,

$$
\frac{1}{T} \sum_{t=1}^{T} \widehat{\xi}_{1,t|T}\, \mathbb{I}_{jt} \left(\widehat{\mathbf{H}}^{-1} \mathbf{g}_t - \widehat{\mathbf{g}}_t\right) \widehat{\mathbf{g}}_t' \leq M \left[\frac{1}{T} \sum_{t=1}^{T} \left(\widehat{\mathbf{H}}^{-1} \mathbf{g}_t - \widehat{\mathbf{g}}_t\right) \widehat{\mathbf{g}}_t'\right] = O_p\left(\frac{1}{C_{NT}^2}\right). \tag{A.31}
$$

From (A.30) and (A.31), and taking into account Assumption 7, it follows that

$$
\sqrt{T}\left[\widehat{\mathbf{B}}_1 - \mathbf{B}_1 \widehat{\mathbf{H}} \widehat{\mathbf{I}}_{\widehat{\xi}1} - \mathbf{B}_2 \widehat{\mathbf{H}} \left(\mathbf{I} - \widehat{\mathbf{I}}_{\widehat{\xi}1}\right)\right] = \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \widehat{\xi}_{1,t|T}\, \mathbf{e}_t \widehat{\mathbf{g}}_t'\right) \left(\frac{1}{T} \sum_{t=1}^{T} \widehat{\xi}_{1,t|T}\, \widehat{\mathbf{g}}_t \widehat{\mathbf{g}}_t'\right)^{-1} + o_p(1).
$$

Since $\widehat{\mathbf{g}}_t = N^{-1}\widehat{\mathbf{A}}'\mathbf{x}_t$ and $\mathbf{x}_t = \mathbb{I}_{1t}\mathbf{\Lambda}_1\mathbf{f}_{1t}+\mathbb{I}_{2t}\mathbf{\Lambda}_2\mathbf{f}_{2t}+\mathbf{e}_t$ then $\widehat{\mathbf{g}}_t = N^{-1}\left(\mathbb{I}_{1t}\widehat{\mathbf{A}}'\mathbf{\Lambda}_1\mathbf{f}_{1t}+\mathbb{I}_{2t}\widehat{\mathbf{A}}'\mathbf{\Lambda}_2\mathbf{f}_{2t}+\widehat{\mathbf{A}}'\mathbf{e}_t\right)$. After some algebra, we have

$$\begin{aligned}
&\sqrt{T}\left[\widehat{\mathbf{B}}_1 - \mathbf{B}_1\widehat{\mathbf{H}}\widehat{\mathbf{I}}_{\widehat{\xi}1} - \mathbf{B}_2\widehat{\mathbf{H}}\left(\mathbf{I}-\widehat{\mathbf{I}}_{\widehat{\xi}1}\right)\right] \\
=\;& \left\{\left(\frac{1}{\sqrt{T}}\sum_{t=1}^T \mathbb{I}_{1t}\widehat{\xi}_{1,t|T}\,\mathbf{e}_t\mathbf{f}_{1t}'\right)\frac{\mathbf{\Lambda}_1'\widehat{\mathbf{A}}}{N} + \left(\frac{1}{\sqrt{T}}\sum_{t=1}^T \mathbb{I}_{2t}\widehat{\xi}_{1,t|T}\,\mathbf{e}_t\mathbf{f}_{2t}'\right)\frac{\mathbf{\Lambda}_2'\widehat{\mathbf{A}}}{N} + \left[\frac{1}{\sqrt{T}}\sum_{t=1}^T \widehat{\xi}_{1,t|T}\,\mathbf{e}_t\left(\mathbf{e}_t'\frac{\widehat{\mathbf{A}}}{N}\right)\right]\right\} \\
&\times\left[\begin{array}{l}
\frac{\widehat{\mathbf{A}}'\mathbf{\Lambda}_1}{N}\left(\frac{1}{T}\sum_{t=1}^T \mathbb{I}_{1t}\widehat{\xi}_{1,t|T}\,\mathbf{f}_{1t}\mathbf{f}_{1t}'\right)\frac{\mathbf{\Lambda}_1'\widehat{\mathbf{A}}}{N} + \frac{\widehat{\mathbf{A}}'\mathbf{\Lambda}_2}{N}\left(\frac{1}{T}\sum_{t=1}^T \mathbb{I}_{2t}\widehat{\xi}_{1,t|T}\,\mathbf{f}_{2t}\mathbf{f}_{2t}'\right)\frac{\mathbf{\Lambda}_2'\widehat{\mathbf{A}}}{N} \\
+\frac{\widehat{\mathbf{A}}'\mathbf{\Lambda}_1}{N}\left(\frac{1}{T}\sum_{t=1}^T \mathbb{I}_{1t}\widehat{\xi}_{1,t|T}\,\mathbf{f}_{1t}\mathbf{e}_t'\right)\frac{\widehat{\mathbf{A}}}{N} + \frac{\widehat{\mathbf{A}}'}{N}\left(\frac{1}{T}\sum_{t=1}^T \mathbb{I}_{1t}\widehat{\xi}_{1,t|T}\,\mathbf{e}_t\mathbf{f}_{1t}'\right)\frac{\mathbf{\Lambda}_1'\widehat{\mathbf{A}}}{N} \\
+\frac{\widehat{\mathbf{A}}'\mathbf{\Lambda}_2}{N}\left(\frac{1}{T}\sum_{t=1}^T \mathbb{I}_{2t}\widehat{\xi}_{1,t|T}\,\mathbf{f}_{2t}\mathbf{e}_t'\right)\frac{\widehat{\mathbf{A}}}{N} + \frac{\widehat{\mathbf{A}}'}{N}\left(\frac{1}{T}\sum_{t=1}^T \mathbb{I}_{2t}\widehat{\xi}_{1,t|T}\,\mathbf{e}_t\mathbf{f}_{2t}'\right)\frac{\mathbf{\Lambda}_2'\widehat{\mathbf{A}}}{N} \\
+\frac{\widehat{\mathbf{A}}'}{N}\left(\frac{1}{T}\sum_{t=1}^T \widehat{\xi}_{1,t|T}\,\mathbf{e}_t\mathbf{e}_t'\right)\frac{\widehat{\mathbf{A}}}{N}
\end{array}\right]^{-1} + o_p(1).
\end{aligned} \tag{A.32}$$

By Lemma 2 it follows that

$$\widehat{\mathbf{A}}' - \widehat{\mathbf{H}}'\mathbf{A}' = O_p\left(\frac{1}{\sqrt{N}C_{NT}}\right) + O_p\left(\frac{1}{\sqrt{T}C_{NT}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right), \tag{A.33}$$

which implies that

$$\widehat{\mathbf{A}} - \mathbf{A}\widehat{\mathbf{H}} = O_p\left(\frac{1}{\sqrt{N}C_{NT}}\right) + O_p\left(\frac{1}{\sqrt{T}C_{NT}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right). \tag{A.34}$$

From (A.32) - (A.34), it follows that

$$\begin{aligned}
&\sqrt{T}\left[\widehat{\mathbf{b}}_{1i} - \widehat{\mathbf{I}}_{\widehat{\xi}1}\widehat{\mathbf{H}}'\mathbf{b}_{1i} - \left(\mathbf{I}-\widehat{\mathbf{I}}_{\widehat{\xi}1}\right)\widehat{\mathbf{H}}'\mathbf{b}_{2i}\right] \\
=\;& \left[\frac{\widehat{\mathbf{A}}'\mathbf{\Lambda}_1}{N}\left(\frac{1}{T}\sum_{t=1}^T \mathbb{I}_{1t}\widehat{\xi}_{1,t|T}\,\mathbf{f}_{1t}\mathbf{f}_{1t}'\right)\frac{\mathbf{\Lambda}_1'\widehat{\mathbf{A}}}{N} + \frac{\widehat{\mathbf{A}}'\mathbf{\Lambda}_2}{N}\left(\frac{1}{T}\sum_{t=1}^T \mathbb{I}_{2t}\widehat{\xi}_{1,t|T}\,\mathbf{f}_{2t}\mathbf{f}_{2t}'\right)\frac{\mathbf{\Lambda}_2'\widehat{\mathbf{A}}}{N}\right]^{-1} \\
&\times\left[\frac{\widehat{\mathbf{A}}'\mathbf{\Lambda}_1}{N}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^T \mathbb{I}_{1t}\widehat{\xi}_{1,t|T}\,\mathbf{f}_{1t}e_{it}\right) + \frac{\widehat{\mathbf{A}}'\mathbf{\Lambda}_2}{N}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^T \mathbb{I}_{2t}\widehat{\xi}_{1,t|T}\,\mathbf{f}_{2t}e_{it}\right)\right] + o_p(1)
\end{aligned}$$

and the result stated in the theorem follows by Assumption 1 and Lemma 6, and by noting that, by Assumption 6(c), $\left(T^{-1/2}\sum_{t=1}^T \mathbb{I}_{1t}\widehat{\xi}_{1,t|T}\,\mathbf{f}_{1t}e_{it}\right)$ and $\left(T^{-1/2}\sum_{t=1}^T \mathbb{I}_{2t}\widehat{\xi}_{1,t|T}\,\mathbf{f}_{2t}e_{it}\right)$ converge in distribution to two independent Normal random variables.

## A.4 Proof of Theorem 2

Given the representation in (9), we can write

$$\mathbf{x}_t = (\mathbf{B}_1\ \mathbf{B}_2)(\xi_t\otimes\mathbf{g}_t) + \mathbf{e}_t = (\mathbf{B}_1\ \mathbf{B}_2)(\xi_{1t}\mathbf{g}_t\ \xi_{2t}\mathbf{g}_t)' + \mathbf{e}_t.$$

Given also the estimators $\widehat{\mathbf{B}}_1$ and $\widehat{\mathbf{B}}_2$ defined according to (29), the estimators $\widehat{\xi}_{1,t|T}\widehat{\mathbf{g}}_t$ and $\widehat{\xi}_{2,t|T}\widehat{\mathbf{g}}_t$ for $\xi_{1t}\mathbf{g}_t$ and $\xi_{2t}\mathbf{g}_t$, respectively, are obtained as

$$
\begin{aligned}
\begin{pmatrix} \widehat{\xi}_{1,t|T}\widehat{\mathbf{g}}_t \\ \widehat{\xi}_{2,t|T}\widehat{\mathbf{g}}_t \end{pmatrix} &= \left[\left(\widehat{\mathbf{B}}_1\ \widehat{\mathbf{B}}_2\right)'\left(\widehat{\mathbf{B}}_1\ \widehat{\mathbf{B}}_2\right)\right]^{-1}\left(\widehat{\mathbf{B}}_1\ \widehat{\mathbf{B}}_2\right)'\mathbf{x}_t \\
&= \begin{pmatrix} \widehat{\mathbf{B}}_1'\widehat{\mathbf{B}}_1 & \widehat{\mathbf{B}}_1'\widehat{\mathbf{B}}_2 \\ \widehat{\mathbf{B}}_2'\widehat{\mathbf{B}}_1 & \widehat{\mathbf{B}}_2'\widehat{\mathbf{B}}_2 \end{pmatrix}^{-1}\begin{pmatrix} \widehat{\mathbf{B}}_1' \\ \widehat{\mathbf{B}}_2' \end{pmatrix}(\mathbf{B}_1\ \mathbf{B}_2)\begin{pmatrix} \xi_{1t}\mathbf{g}_t \\ \xi_{2t}\mathbf{g}_t \end{pmatrix} \\
&\quad + \begin{pmatrix} \widehat{\mathbf{B}}_1'\widehat{\mathbf{B}}_1 & \widehat{\mathbf{B}}_1'\widehat{\mathbf{B}}_2 \\ \widehat{\mathbf{B}}_2'\widehat{\mathbf{B}}_1 & \widehat{\mathbf{B}}_2'\widehat{\mathbf{B}}_2 \end{pmatrix}^{-1}\begin{pmatrix} \widehat{\mathbf{B}}_1' \\ \widehat{\mathbf{B}}_2' \end{pmatrix}\mathbf{e}_t.
\end{aligned}
$$

Adding and subtracting terms, it follows that

$$
\begin{aligned}
\begin{pmatrix} \widehat{\xi}_{1,t|T}\widehat{\mathbf{g}}_t \\ \widehat{\xi}_{2,t|T}\widehat{\mathbf{g}}_t \end{pmatrix} &= \begin{pmatrix} \widehat{\mathbf{B}}_1'\widehat{\mathbf{B}}_1 & \widehat{\mathbf{B}}_1'\widehat{\mathbf{B}}_2 \\ \widehat{\mathbf{B}}_2'\widehat{\mathbf{B}}_1 & \widehat{\mathbf{B}}_2'\widehat{\mathbf{B}}_2 \end{pmatrix}^{-1}\begin{pmatrix} \widehat{\mathbf{B}}_1' \\ \widehat{\mathbf{B}}_2' \end{pmatrix}(\mathbf{B}_1\ \mathbf{B}_2)\begin{pmatrix} \xi_{1t}\mathbf{g}_t \\ \xi_{2t}\mathbf{g}_t \end{pmatrix} \\
&\quad + \begin{pmatrix} \widehat{\mathbf{B}}_1'\widehat{\mathbf{B}}_1 & \widehat{\mathbf{B}}_1'\widehat{\mathbf{B}}_2 \\ \widehat{\mathbf{B}}_2'\widehat{\mathbf{B}}_1 & \widehat{\mathbf{B}}_2'\widehat{\mathbf{B}}_2 \end{pmatrix}^{-1}\begin{pmatrix} \widehat{\mathbf{B}}_1' \\ \widehat{\mathbf{B}}_2' \end{pmatrix}\left(\widehat{\mathbf{B}}_1\ \widehat{\mathbf{B}}_2\right)\widehat{\mathbf{H}}_\xi^{-1}\begin{pmatrix} \xi_{1t}\mathbf{g}_t \\ \xi_{2t}\mathbf{g}_t \end{pmatrix} \\
&\quad - \begin{pmatrix} \widehat{\mathbf{B}}_1'\widehat{\mathbf{B}}_1 & \widehat{\mathbf{B}}_1'\widehat{\mathbf{B}}_2 \\ \widehat{\mathbf{B}}_2'\widehat{\mathbf{B}}_1 & \widehat{\mathbf{B}}_2'\widehat{\mathbf{B}}_2 \end{pmatrix}^{-1}\begin{pmatrix} \widehat{\mathbf{B}}_1' \\ \widehat{\mathbf{B}}_2' \end{pmatrix}\left(\widehat{\mathbf{B}}_1\ \widehat{\mathbf{B}}_2\right)\widehat{\mathbf{H}}_\xi^{-1}\begin{pmatrix} \xi_{1t}\mathbf{g}_t \\ \xi_{2t}\mathbf{g}_t \end{pmatrix} \\
&\quad + \begin{pmatrix} \widehat{\mathbf{B}}_1'\widehat{\mathbf{B}}_1 & \widehat{\mathbf{B}}_1'\widehat{\mathbf{B}}_2 \\ \widehat{\mathbf{B}}_2'\widehat{\mathbf{B}}_1 & \widehat{\mathbf{B}}_2'\widehat{\mathbf{B}}_2 \end{pmatrix}^{-1}\begin{pmatrix} \widehat{\mathbf{B}}_1' \\ \widehat{\mathbf{B}}_2' \end{pmatrix}\mathbf{e}_t \\
&\quad + \begin{pmatrix} \widehat{\mathbf{B}}_1'\widehat{\mathbf{B}}_1 & \widehat{\mathbf{B}}_1'\widehat{\mathbf{B}}_2 \\ \widehat{\mathbf{B}}_2'\widehat{\mathbf{B}}_1 & \widehat{\mathbf{B}}_2'\widehat{\mathbf{B}}_2 \end{pmatrix}^{-1}\widehat{\mathbf{H}}_\xi'\begin{pmatrix} \mathbf{B}_1' \\ \mathbf{B}_2' \end{pmatrix}\mathbf{e}_t \\
&\quad - \begin{pmatrix} \widehat{\mathbf{B}}_1'\widehat{\mathbf{B}}_1 & \widehat{\mathbf{B}}_1'\widehat{\mathbf{B}}_2 \\ \widehat{\mathbf{B}}_2'\widehat{\mathbf{B}}_1 & \widehat{\mathbf{B}}_2'\widehat{\mathbf{B}}_2 \end{pmatrix}^{-1}\widehat{\mathbf{H}}_\xi'\begin{pmatrix} \mathbf{B}_1' \\ \mathbf{B}_2' \end{pmatrix}\mathbf{e}_t,
\end{aligned}
$$

or equivalently

$$
\begin{aligned}
&\left[\begin{pmatrix} \widehat{\xi}_{1,t|T}\widehat{\mathbf{g}}_t \\ \widehat{\xi}_{2,t|T}\widehat{\mathbf{g}}_t \end{pmatrix} - \widehat{\mathbf{H}}_\xi^{-1}\begin{pmatrix} \xi_{1t}\mathbf{g}_t \\ \xi_{2t}\mathbf{g}_t \end{pmatrix}\right] \\
=&\left[N^{-1}\begin{pmatrix} \widehat{\mathbf{B}}_1'\widehat{\mathbf{B}}_1 & \widehat{\mathbf{B}}_1'\widehat{\mathbf{B}}_2 \\ \widehat{\mathbf{B}}_2'\widehat{\mathbf{B}}_1 & \widehat{\mathbf{B}}_2'\widehat{\mathbf{B}}_2 \end{pmatrix}\right]^{-1}\widehat{\mathbf{H}}_\xi'\left[N^{-1}\begin{pmatrix} \mathbf{B}_1' \\ \mathbf{B}_2' \end{pmatrix}\mathbf{e}_t\right] \\
&+\left[N^{-1}\begin{pmatrix} \widehat{\mathbf{B}}_1'\widehat{\mathbf{B}}_1 & \widehat{\mathbf{B}}_1'\widehat{\mathbf{B}}_2 \\ \widehat{\mathbf{B}}_2'\widehat{\mathbf{B}}_1 & \widehat{\mathbf{B}}_2'\widehat{\mathbf{B}}_2 \end{pmatrix}\right]\left\{N^{-1}\begin{pmatrix} \widehat{\mathbf{B}}_1' \\ \widehat{\mathbf{B}}_2' \end{pmatrix}\left[(\mathbf{B}_1\ \mathbf{B}_2)\widehat{\mathbf{H}}_\xi - \left(\widehat{\mathbf{B}}_1\ \widehat{\mathbf{B}}_2\right)\right]\widehat{\mathbf{H}}_\xi^{-1}\right\}\begin{pmatrix} \xi_{1t}\mathbf{g}_t \\ \xi_{2t}\mathbf{g}_t \end{pmatrix} \\
&+\left[N^{-1}\begin{pmatrix} \widehat{\mathbf{B}}_1'\widehat{\mathbf{B}}_1 & \widehat{\mathbf{B}}_1'\widehat{\mathbf{B}}_2 \\ \widehat{\mathbf{B}}_2'\widehat{\mathbf{B}}_1 & \widehat{\mathbf{B}}_2'\widehat{\mathbf{B}}_2 \end{pmatrix}\right]^{-1}\left\{N^{-1}\left[\begin{pmatrix} \widehat{\mathbf{B}}_1' \\ \widehat{\mathbf{B}}_2' \end{pmatrix} - \widehat{\mathbf{H}}_\xi'\begin{pmatrix} \mathbf{B}_1' \\ \mathbf{B}_2' \end{pmatrix}\right]\mathbf{e}_t\right\}.
\end{aligned}
\tag{A.35}
$$

Consider first

$$\begin{aligned}
&\frac{1}{N}\begin{pmatrix}\widehat{\mathbf{B}}_1'\\ \widehat{\mathbf{B}}_2'\end{pmatrix}\left[(\mathbf{B}_1\ \mathbf{B}_2)-\left(\widehat{\mathbf{B}}_1\ \widehat{\mathbf{B}}_2\right)\widehat{\mathbf{H}}_\xi^{-1}\right]\begin{pmatrix}\xi_{1t}\mathbf{g}_t\\ \xi_{2t}\mathbf{g}_t\end{pmatrix}\\
=\ &\frac{1}{N}\begin{pmatrix}\widehat{\mathbf{B}}_1'\\ \widehat{\mathbf{B}}_2'\end{pmatrix}\left[(\mathbf{B}_1\ \mathbf{B}_2)\widehat{\mathbf{H}}_\xi-\left(\widehat{\mathbf{B}}_1\ \widehat{\mathbf{B}}_2\right)\right]\widehat{\mathbf{H}}_\xi^{-1}\begin{pmatrix}\xi_{1t}\mathbf{g}_t\\ \xi_{2t}\mathbf{g}_t\end{pmatrix},
\end{aligned}$$

so that from (A.30) and (A.31), and taking into account Assumption 2, it follows that

$$\begin{aligned}
&\left\|\frac{1}{N}\begin{pmatrix}\widehat{\mathbf{B}}_1'\\ \widehat{\mathbf{B}}_2'\end{pmatrix}\left[(\mathbf{B}_1\ \mathbf{B}_2)-\left(\widehat{\mathbf{B}}_1\ \widehat{\mathbf{B}}_2\right)\widehat{\mathbf{H}}_\xi^{-1}\right]\begin{pmatrix}\xi_{1t}\mathbf{g}_t\\ \xi_{2t}\mathbf{g}_t\end{pmatrix}\right\|\\
\leq\ &\left\|\frac{1}{\sqrt{N}}\begin{pmatrix}\widehat{\mathbf{B}}_1'\\ \widehat{\mathbf{B}}_2'\end{pmatrix}\right\|\left\|\frac{1}{\sqrt{N}}\left[(\mathbf{B}_1\ \mathbf{B}_2)\widehat{\mathbf{H}}_\xi-\left(\widehat{\mathbf{B}}_1\ \widehat{\mathbf{B}}_2\right)\right]\right\|\left\|\widehat{\mathbf{H}}_\xi\right\|\left\|\begin{pmatrix}\xi_{1t}\mathbf{g}_t\\ \xi_{2t}\mathbf{g}_t\end{pmatrix}\right\| \qquad \text{(A.36)}\\
=\ &O_p\left(\frac{1}{\sqrt{NT}}\right)+O_p\left(\frac{1}{\sqrt{N}C_{NT}^2}\right).
\end{aligned}$$

By (A.30) and (A.31), and taking into account Assumption 3(b), we also have that,

$$\begin{aligned}
\left\|\frac{1}{N}\left[\begin{pmatrix}\widehat{\mathbf{B}}_1'\\ \widehat{\mathbf{B}}_2'\end{pmatrix}-\widehat{\mathbf{H}}_\xi'\begin{pmatrix}\mathbf{B}_1'\\ \mathbf{B}_2'\end{pmatrix}\right]\mathbf{e}_t\right\| &\leq \frac{\|\mathbf{e}_t\|}{\sqrt{N}}\left\|\frac{1}{\sqrt{N}}\left[\begin{pmatrix}\widehat{\mathbf{B}}_1'\\ \widehat{\mathbf{B}}_2'\end{pmatrix}-\widehat{\mathbf{H}}_\xi'\begin{pmatrix}\mathbf{B}_1'\\ \mathbf{B}_2'\end{pmatrix}\right]\right\| \qquad \text{(A.37)}\\
&= O_p\left(\frac{1}{\sqrt{NT}}\right)+O_p\left(\frac{1}{\sqrt{N}C_{NT}^2}\right).
\end{aligned}$$

Therefore, taking into account (A.35), (A.36) and (A.37), and by Assumption 7, we have

$$\sqrt{N}\left[\begin{pmatrix}\widehat{\xi}_{1,t|T}\widehat{\mathbf{g}}_t\\ \widehat{\xi}_{2,t|T}\widehat{\mathbf{g}}_t\end{pmatrix}-\widehat{\mathbf{H}}_\xi^{-1}\begin{pmatrix}\xi_{1t}\mathbf{g}_t\\ \xi_{2t}\mathbf{g}_t\end{pmatrix}\right]=\begin{pmatrix}\frac{\widehat{\mathbf{B}}_1'\widehat{\mathbf{B}}_1}{N} & \frac{\widehat{\mathbf{B}}_1'\widehat{\mathbf{B}}_2}{N}\\ \frac{\widehat{\mathbf{B}}_2'\widehat{\mathbf{B}}_1}{N} & \frac{\widehat{\mathbf{B}}_2'\widehat{\mathbf{B}}_2}{N}\end{pmatrix}^{-1}\frac{1}{\sqrt{N}}\widehat{\mathbf{H}}_\xi\begin{pmatrix}\mathbf{B}_1'\\ \mathbf{B}_2'\end{pmatrix}\mathbf{e}_t+o_p(1).$$

Given $\widehat{\mathbf{H}}_\xi$, let $\mathbf{I}_{\xi j}=\mathrm{p}\lim_{N,T\to\infty}\widehat{\mathbf{I}}_{\xi j}$ for $j=1,2$, where $\widehat{\mathbf{I}}_{\xi j}$ is defined in (49). Also, given $\widehat{\mathbf{H}}$ defined in (45), we have $\widehat{\mathbf{H}}\overset{p}{\to}\boldsymbol{\Sigma}_\mathbf{g}\mathbf{Q}\mathbf{V}^{-1}=\mathbf{H}$, where $\boldsymbol{\Sigma}_\mathbf{g}=\mathrm{p}\lim_{N,T\to\infty}(\mathbf{G}\mathbf{G}/T)$ by Assumption (1), and $\mathbf{Q}=\mathrm{p}\lim_{N,T\to\infty}\left(\mathbf{A}'\widehat{\mathbf{A}}/N\right)=$ by Lemma 6. By Theorem 1, we then have $\left(\widehat{\mathbf{B}}_1\ \widehat{\mathbf{B}}_2\right)'\overset{p}{\to}\mathbf{H}_\xi(\mathbf{B}_1\ \mathbf{B}_2)'$. Therefore

$$\mathrm{p}\lim_{N,T\to\infty}\begin{pmatrix}\frac{\widehat{\mathbf{B}}_1'\widehat{\mathbf{B}}_1}{N} & \frac{\widehat{\mathbf{B}}_1'\widehat{\mathbf{B}}_2}{N}\\ \frac{\widehat{\mathbf{B}}_2'\widehat{\mathbf{B}}_1}{N} & \frac{\widehat{\mathbf{B}}_1'\widehat{\mathbf{B}}_2}{N}\end{pmatrix}=\mathbf{H}_\xi\begin{pmatrix}\boldsymbol{\Sigma}_{\mathbf{B}1} & \boldsymbol{\Sigma}_{\mathbf{B}12}\\ \boldsymbol{\Sigma}_{\mathbf{B21}} & \boldsymbol{\Sigma}_{\mathbf{B}2}\end{pmatrix}\mathbf{H}_\xi',$$

where, by Assumption 2, $\left\|\left(\mathbf{B}_j'\mathbf{B}_j/N\right) - \mathbf{\Sigma}_{\mathbf{B}j}\right\| \to 0$ and $\left\|\left(\mathbf{B}_j'\mathbf{B}_k/N\right) - \mathbf{\Sigma}_{\mathbf{B}jk}\right\| \to 0$, for $j, k = 1, 2$ with $j \neq k$ as $N \to \infty$ . The result stated in the theorem follows by noting that

$$\frac{1}{\sqrt{N}} \begin{pmatrix} \mathbf{B}_1' \\ \mathbf{B}_2' \end{pmatrix} \mathbf{e}_t \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \mathbf{\Sigma}_{\mathbf{Be}t}\right).$$

by Assumption 6(d), which concludes the proof.

# References


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

Alessi, L., M. Barigozzi, and M. Capasso (2010). Improved penalization for determining the number of factors in approximate static factor models. Statistics and Probability Letters 80, 1806–1813.

Ang, A. and G. Bekaert (2002). International asset allocation with regime shifts. Review of Financial Studies 15, 1137–1187.

Ang, A. and A. Timmermann (2012). Regime changes and financial markets. Annual Review of Financial Economics 4, 313–337.

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

Bai, J. and X. Han (2016). Structural changes in high dimensional factor models. Frontiers of Economics in China 11, 9–39.

Bai, J. and K. Li (2012). Statistical analysis of factor models of high dimension. The Annals of Statistics 40, 436–465.

Bai, J. and K. Li (2016). Maximum likelihood estimation and inference for approximate factor models of high dimension. The Review of Economics and Statistics 98, 298–309.

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

Bai, J. and S. Ng (2013). Principal components estimation and identification of static factors. Journal of econometrics 176, 18–29.

Baltagi, B. H., Q. Feng, and C. Kao (2016). Estimation of heterogeneous panels with structural breaks. Journal of Econometrics 191, 176–195.

Barigozzi, M., H. Cho, and P. Fryzlewicz (2018). Simultaneous multiple change-point and factor analysis for high-dimensional time series. Journal of Econometrics 206, 187–225.

Barigozzi, M. and M. Luciani (2019). Quasi maximum likelihood estimation and inference of large approximate dynamic factor models via the EM algorithm. Technical Report arXiv:1910.03821.

Barigozzi, M. and L. Trapani (2020). Sequential testing for structural stability in approximate factor models. Stochastic Processes and their Applications 130, 5149–5187.

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

Chamberlain, G. and M. Rothschild (1983). Arbitrage, factor structure, and mean-variance analysis on large asset markets. Econometrica, 1281–1304.

Cheng, X., Z. Liao, and F. Schorfheide (2016). Shrinkage estimation of high-dimensional factor models with structural instabilities. The Review of Economic Studies 83, 1511–1543.

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

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

Dempster, A. P., N. M. Laird, and D. B. Rubin (1977). Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 1–38.

Doz, C., D. Giannone, and L. Reichlin (2012). A quasi maximum likelihood approach for large approximate dynamic factor models. The Review of Economics and Statistics 94(4), 1014–1024.

Duan, J., J. Bai, and X. Han (2022). Quasi-maximum likelihood estimation of break point in high-dimensional factor models. Journal of Econometrics. forthcoming.

Fama, E. F. and K. R. French (2016). Dissecting anomalies with a five-factor model. Review of Financial Studies 29(1), 69–103.

Forni, M., M. Hallin, M. Lippi, and P. Zaffaroni (2017). Dynamic factor models with infinite-dimensional factor space: Asymptotic analysis. Journal of Econometrics 199, 74–92.

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

Goldfeld, S. M. and R. E. Quandt (1973). A Markov model for switching regressions. Journal of Econometrics 1, 3–15.

Guidolin, M. (2011). Markov Switching Models in Empirical Finance. In D. M. Drukker (Ed.), Missing Data Methods: Time-Series Methods and Applications (Advances in Econometrics), Volume 27 Part 2, pp. 1–86. Emerald Group Publishing Limited, Bingley.

Guidolin, M. and M. Pedio (2018). Essentials of time series for financial applications. Academic Press.

Guidolin, M. and A. Timmermann (2006). An econometric model of nonlinear dynamics in the joint distribution of stock and bond returns. Journal of Applied Econometrics 21, 1–23.

Guidolin, M. and A. Timmermann (2008). International asset allocation under regime switching, skew, and kurtosis preferences. Review of Financial Studies 21, 889–935.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. Econometrica, 357–384.

Hamilton, J. D. (2016). Macroeconomic Regimes and Regime Shifts. In J. B. Taylor and H. Uhlig (Eds.), Handbook of Macroeconomics, Volume 2A, pp. 163–201. Elsevier.

Kim, C.-J. (1994). Dynamic linear models with markov-switching. Journal of Econometrics 60, 1–22.

Krolzig, H.-M. (2013). Markov-switching vector autoregressions: Modelling, statistical inference, and application to business cycle analysis, Volume 454. Springer Science & Business Media.

Liu, X. and R. Chen (2016). Regime-switching factor models for high-dimensional time series. Statistica Sinica 26, 1427–1451.

Massacci, D. (2017). Least squares estimation of large dimensional threshold factor models. Journal of Econometrics 197(1), 101–129.

Massacci, D., L. Sarno, and L. Trapani (2021). Factor models with downside risk. Working Paper, King's College London, University of Cambridge and University of Nottingham.

Motta, G., C. M. Hafner, and R. Von Sachs (2011). Locally stationary factor models: Identification and nonparametric estimation. Econometric Theory 27, 1279–1319.

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

Pelger, M. and R. Xiong (2022). State-varying factor models of large dimensions. Journal of Business and Economic Statistics 40, 1315–1333.

Perez-Quiros, G. and A. Timmermann (2000). Firm size and cyclical variations in stock returns. Journal of Finance 55, 1229–1262.

Perez-Quiros, G. and A. Timmermann (2001). Business cycle asymmetries in stock returns: Evidence from higher order moments and conditional densities. Journal of Econometrics 103, 259–306.

Qu, Z. and F. Zhuo (2021). Likelihood ratio based tests for markov regime switching. Review of Economic Studies 88, 937–968.

Rubin, D. B. and D. T. Thayer (1982). EM algorithms for ML factor analysis. Psychometrika 47, 69–76.

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

Stock, J. H. and M. W. Watson (2002b). Macroeconomic forecasting using diffusion indexes. Journal of Business and Economic Statistics 20, 147–162.

Stock, J. H. and M. W. Watson (2016). Dynamic Factor Models, Factor-Augmented Vector Autoregressions, and Structural Vector Autoregressions in Macroeconomics. In J. B. Taylor and H. Uhlig (Eds.), Handbook of Macroeconomics, Volume 2, pp. 415–525. Elsevier.

Trapani, L. (2018). A randomized sequential procedure to determine the number of factors. Journal of the American Statistical Association 113, 1341–1349.

Urga, G. and F. Wang (2022). Estimation and inference for high dimensional factor model with regime switching. Working Paper, Bayes Business School and Peking University.

Wu, J. C. F. (1983). On the convergence properties of the EM algorithm. The Annals of Statistics 11, 95–103.