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 that makes use of the factors previously estimated. Our approach is appealing as it provides closed form expressions for all estimators. More importantly, it does not require knowledge of the true number of factors. We derive the theoretical properties of the proposed estimation procedure, and we show their good finite sample performance through a comprehensive set of Monte Carlo experiments. 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.

Since the works of Hamilton (1989), and Diebold and Rudebusch (1996), and inspired by the seminal paper 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, and Doz et al. (2020) present recent evidence of their usefulness for turning-point detection and macroeconomic forecasting; 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. 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 macroeconomic 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. For example, following the seminal contribution of Chamberlain and Rothschild (1983), static approximate factor representations have been considered 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. (2023), 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, 2023). Under this set up, regimes are recurrent and associated to cyclical events such as business and financial cycles. Smoothly varying loadings are considered in Motta et al. (2011) and Pelger and Xiong (2022). Finally, Chen et al. (2023) follow Su and Wang (2017) and propose a time-varying matrix factor model with smooth changes in the loadings driven by a time index.

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 we 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; and 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 they are 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; Kapetanios (2010), 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 contribution 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) study a model similar to ours, but their definition of common factors differs from ours in that they consider factors that are pervasive along the time dimension rather than along the cross-sectional dimension. As a consequence, their idiosyncratic components are assumed to be white noise. Second, Urga and Wang (2022) study a set up similar to ours, with two important differences: they assume a priori knowledge of the number of factors; they consider a model with serially homoskedastic idiosyncratic components. In addition, the Maximum Likelihood estimation approach of Urga and Wang (2022) 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. Note that, to the very best of our knowledge, no existing contribution has formally proved that in the factor model case the EM algorithm is a contraction, namely that it converges to the Maximum Likelihood estimator. The approach in Urga and Wang
leads to estimators for the unknown parameters that do not have closed form solutions. This in an additional important difference with respect to our approach, which instead leads to closed form solutions.

Our approach is as follows. We introduce an algorithm to estimate factors, loadings, and transition probabilities, which extends to high dimensional factor models the state-space approach advanced in Hamilton (1989) and Kim (1994) to handle low dimensional Markov switching autoregressive models. In particular, we generalize the the Baum-Lindgren-Hamilton-Kim filter and smoother, the original version of which was proposed to estimate Markov-switching VAR models: for example, see 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. Even more remarkably, we not require a priori knowledge of the number of factors in each regime, which is instead needed by Urga and Wang (2022).

We obtain our theoretical results 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: for example, see the discussions in Breitung and Eickmeier (2011), Barigozzi et al. (2018), and Duan et al. (2023), in the case of structural breaks; and Massacci (2023) for threshold factor models. We use this property to estimate the latent factors by means of Principal Component Analysis (PCA) as 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 identification problem, and by the incomplete information related to the underlying data generated process. We also study the asymptotic properties of the estimated factors, which are obtained 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 presents two further results related to estimation of the number of factors and to under-specification of the number of regimes. Section 6 briefly discusses the issue of unobserved heterogeneity. Section 7 runs a comprehensive set of Monte Carlo experiments. Section 8 applies our model to a large number of industry equity portfolios. Finally, Section 9 concludes. Details about the estimation algorithm are given in Appendix A. Mathematical derivations are collected in Appendices B and C.

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

For a scalar discrete random variable \( Z \), the notation \( P(Z = z) \) is its probability mass function computed using the true value of the parameters. For random variables \( Y \) and \( W \) the notations \( E[Y] \) and \( E[Y|W] \) are the expectation and conditional expectation given \( W \), respectively, computed with respect to the true distributions \( F_Y(y) \) and \( F_{Y|W}(y|W) \) 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 \( \hat{\theta} \), then we adopt the notations \( P_{\hat{\theta}}(Z = z) \), \( E_{\hat{\theta}}[Y] \), and \( E_{\hat{\theta}}[Y|W] \), respectively.

Finally, we let \( I_m \) be the identity matrix of dimension \( m \), \( \iota_m \) an \( m \)-dimensional vector of ones, and \( 0 \) 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

\[
x_t = \Lambda_1 f_{1t} I(s_t = 1) + \Lambda_2 f_{2t} I(s_t = 2) + \epsilon_t, \quad t \in \mathbb{Z},
\]

\[
\epsilon_t = \Sigma^1_1 I(s_t = 1) \nu_t + \Sigma^1_2 I(s_t = 2) \nu_t.
\]

We assume that the elements of the \( N \times 1 \) vector process of observable dependent variables \( \{x_t\} \) have zero mean, and we consider the more general case in which they are allowed to have mean different from zero in Section 6. \( \{f_{jt}\} \) is the \( r_j \times 1 \) vector process of latent factors such
that \( r_j \) is fixed and \( r_j \ll N \), for \( j = 1, 2 \); \( \mathbf{\Lambda}_j \) is the \( N \times r_j \) matrix of factor loadings with rows equal to \( \mathbf{\Lambda}'_{ji} \), for \( i = 1, \ldots, N \) and \( j = 1, 2 \); \( \{ \mathbf{e}_t \} \) is the \( N \times 1 \) vector process of idiosyncratic components with innovations \( \mathbf{\nu}_t \sim (\mathbf{0}, \mathbf{I}_N) \). Note that we allow the elements of \( \{ \mathbf{e}_t \} \) to be both serially and cross-sectionally weakly correlated, and we refer to Section 4 for the specific assumptions. It is also important to point out that the number of factors \( r_j \) within each state is allowed to be unknown.

The model in (1) and (2) explicitly allows for two regimes: the case in which the number of states is actually underspecified is dealt with in Section 5.2. Also, the number of factors \( r_1 \) and \( r_2 \) is allowed to change between the regimes: in this, our approach is more general than in Liu and Chen (2016), who assume that \( r_1 = r_2 \) and the dimension of the factor space is \textit{a priori} the same between the two regimes.

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

\[
P( s_{t+1} = j | s_t = i ) = p_{ij}, \quad i, j = 1, 2, \quad \sum_{j=1}^{2} p_{ij} = 1,
\]

with matrix of transition probabilities

\[
P = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix} = \begin{pmatrix} p_{11} & 1 - p_{11} \\ 1 - p_{22} & p_{22} \end{pmatrix}.
\] (3)

Defining the \( 2 \times 1 \) vector of state indicators

\[
\mathbf{\xi}_t = \begin{bmatrix} \mathbb{I}(s_t = 1) \\ \mathbb{I}(s_t = 2) \end{bmatrix}, \quad t \in \mathbb{Z},
\] (4)

allows us to write the transition equation

\[
\mathbf{\xi}_t = \mathbf{P}'\mathbf{\xi}_{t-1} + \mathbf{v}_t, \quad t \in \mathbb{Z},
\] (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 \( \bar{\mathbf{\xi}} \) satisfying:

\[
\bar{\mathbf{\xi}} = \mathbf{P}'\bar{\mathbf{\xi}}.
\]

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

\[
\mathbb{E}[\mathbf{\xi}_t] = \mathbb{E}
\begin{bmatrix} \mathbb{I}(s_t = 1) \\ \mathbb{I}(s_t = 2) \end{bmatrix} = \begin{bmatrix} \mathbb{P}(s_t = 1) \\ \mathbb{P}(s_t = 2) \end{bmatrix},
\] (6)
where $0 < P(s_t = j) < 1$, 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)

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

(7)

Finally, unlike the low-dimensional model of Diebold and Rudebusch (1996), we do not specify the factor dynamics. In particular, Diebold and Rudebusch (1996) allow for regime-specific factor mean, whereas the loadings do not vary: in this setting, the variance of the dependent variables remains constant over time. On the other hand, the large-dimensional model in (1) and (2) allows for regime-specific covariance matrix of $\mathbf{x}$: this is relevant for modelling both macroeconomic variables and financial returns, as stressed in McConnell and Perez-Quiros (2000), and Perez-Quiros and Timmermann (2000, 2001), respectively. We will exploit this feature in the empirical analysis in Section [8] where we will use the model in (1) and (2) to model returns from a large equity portfolio.

2.2 State space representation

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

$$
\mathbf{g}_t = \begin{bmatrix} f_{1t} \\ 0 \end{bmatrix} \mathbb{I}(s_t = 1) + \begin{bmatrix} 0 \\ f_{2t} \end{bmatrix} \mathbb{I}(s_t = 2) = \begin{bmatrix} f_{1t} \\ f_{2t} \end{bmatrix} \otimes \xi_t, \quad t \in \mathbb{Z}.
$$

(8)

Let $\mathbf{B}_1 = [\mathbf{A}_1 \ 0]$ and $\mathbf{B}_2 = [0 \ \mathbf{A}_2]$, 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

$$
\mathbf{x}_t = (\mathbf{B}_1 \ \mathbf{B}_2)(\xi_t \otimes \mathbf{g}_t) + \left(\begin{bmatrix} \Sigma_{e1}^{1/2} \\
\Sigma_{e2}^{1/2} \end{bmatrix} \otimes \mathbf{I}_N\right) \mathbf{e}_t, \quad t \in \mathbb{Z},
$$

$$
\xi_t = \mathbf{P}'\xi_{t-1} + \mathbf{v}_t.
$$

(9)

Under standard assumptions, the term $(\mathbf{B}_1 \ \mathbf{B}_2)(\xi_t \otimes \mathbf{g}_t)$ is identifiable up to a relabelling of the states. This means that the indices of the states can be permuted without changing the law governing the process for $\mathbf{x}_t$: on this, see Section 3 in Leroux (1992). Also note that, even for given $\xi_t$, identification of $\mathbf{B}_1$ and $\mathbf{B}_2$, and therefore of the elements of $\{\mathbf{g}_t\}$, is in general possible only up to an invertible linear transformation (see Bai, 2003).

\footnote{Note that $\xi_t \otimes \mathbf{g}_t = [f_{1t}' \ 0 \ f_{2t}' \ 0]'$.}
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). It can then be rewritten as the \( r_1 + r_2 \) linear factor model

\[
x_t = A g_t + e_t, \quad t \in \mathbb{Z},
\]

where \( A = [A_1 \ A_2] \). Then \( A \) and \( g_t \) may be estimated by standard Principal Component Analysis (PCA) (Stock and Watson, 2002a,b; Bai, 2003). Since PCA gives, as \( N, T \rightarrow \infty \), consistent estimators of the factors up to premultiplication by an invertible matrix (see Bai, 2003), for ease of exposition we first consider estimation of the model in (9) by treating \( g_t \) as known. We then briefly review the implementation of PCA and its effect on the estimation of the model in Section 3.3.

2.4 Log-likelihood

We follow Bai and Li (2016) and estimate only the diagonal elements of \( \Sigma_{e1} \) and \( \Sigma_{e2} \) in (2). The parameters of interest are then partitioned as

\[
\varphi = [\text{vec} (B_1'), \text{vec} (B_2'), \text{diag} (\Sigma_{e1})', \text{diag} (\Sigma_{e2})']', \quad \rho = \text{vec} (P),
\]

so that the vector of parameters of interest, denoted as \( q \), is defined as

\[
q = [\varphi', \rho']'.
\]

Let \( X = (x_1', \ldots, x_T')' \), \( G = (g_1', \ldots, g_T')' \), where \( X \) is an \( NT \times 1 \) vector, \( G \) is an \( (r_1 + r_2)T \times 1 \) vector. These are \( T \)-dimensional realizations of the stochastic processes \( \{x_t\} \) and \( \{g_t\} \), respectively. Moreover, let \( X_v \) be the \( \sigma \)-algebra generated by the random variables \( \{x_t\}_{t=1}^v \), for \( v = 1, \ldots, T \); in a similar way, define \( G_v \) as the \( \sigma \)-algebra generated by the random variables \( \{g_t\}_{t=1}^v \), for \( v = 1, \ldots, T \). And for simplicity we write \( X \equiv X_T \) and \( G \equiv G_T \).

The likelihood function, denoted by \( f(X; \theta) \), can be decomposed as

\[
f(X; \theta) = \frac{f(X, G; \theta)}{f(G; \theta)} = \frac{f(X | G; \theta) f(G; \theta)}{f(G; \theta)} = \frac{f(X | G; \theta) f(G)}{f(G)}:
\]

in the last step we account for the fact that \( f(G; \theta) = f(G) \), since it does not depend on the parameters of our model, as we do not specify any dynamic model for the process \( \{g_t\} \).

Furthermore, following Krolzig (2013, Section 6.2), we have

\[
f(X | G; \theta) = f(X | G; \varphi, \rho) = \sum_{\{\xi_t\}_{t=1}^T \in (0,1)^T} f(X | G; \{\xi_t\}_{t=1}^T; \varphi) P(\{\xi_t\}_{t=1}^T | G, \rho).
\]
Here, to avoid heavier notation, we use the same notation \( \{ \xi_t \}_{t=1}^T \) both for a generic \( T \)-dimensional realization of the process \( \{ \xi_t \} \) and for the \( \sigma \)-algebra generated by the random variables \( \{ \xi_t \}_{t=1}^T \). Notice that the sum is over \( 2^T \) possible values since, given a realization for \( \{ \xi_{1t} \}_{t=1}^T \), 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), and Barigozzi and Luciani (2019), and the one by Bai and Li (2016), all developed for QML estimation of linear factor models, for \( f(\mathbf{X} | G, \{ \xi_t \}_{t=1}^T; \varphi) \) we consider a misspecified Gaussian quasi-likelihood of an exact factor model. This implies that the idiosyncratic components are treated as if they were cross-sectionally uncorrelated. In addition, we also neglect serial correlation in the idiosyncratic components, thus treating them as if they were weak white noise processes. This approach is similar to the one also followed in Urga and Wang (2022). It is important to stress that we are not assuming that the idiosyncratic components are uncorrelated, as we are just considering likelihood estimation of a misspecified model. In the linear case, Bai and Li (2016), and Barigozzi and Luciani (2019), show that such misspecification is asymptotically negligible as \( N, T \to \infty \). Under this setup, and using the Markov property of \( \{ \xi_t \} \), up to omitted constant terms we have

\[
\log f(\mathbf{X} | G, \{ \xi_t \}_{t=1}^T; \varphi) \approx -\frac{1}{2} \sum_{t=1}^T \log \det \Sigma_{et} - \frac{1}{2} \sum_{t=1}^T \{ \mathbf{x}_t - (\mathbf{B}_1 \mathbf{B}_2) (\xi_t \otimes \mathbf{g}_t) \}^\top (\Sigma_{et})^{-1} \{ \mathbf{x}_t - (\mathbf{B}_1 \mathbf{B}_2) (\xi_t \otimes \mathbf{g}_t) \},
\]

where \( \Sigma_{et} = (\text{diag}(\Sigma_{e1}) \text{diag}(\Sigma_{e2})) (\xi_t \otimes \mathbf{I}_N) \). Note that in this case the likelihood (12) is not Gaussian; rather, it is a mixture of Gaussian distributions. Finally, again by the Markov property of \( \{ \xi_t \} \), we can write

\[
P(\{ \xi_t \}_{t=1}^T | G; \rho) = \prod_{t=1}^T P(\xi_t | \xi_{t-1}, G; \rho) P(\xi_0).
\]

### 3 Estimation

In this section, we assume that the data generating process is characterised by two regimes as in the model in (1) and (2). In Section 5.2, we study the case in which the model is underspecified and the data generating process exhibits a higher number of regimes. We also assume that the dimension of the vector \( \mathbf{g}_t \) in (10) is known. Should this not be the case, the dimension of \( \mathbf{g}_t \) can be determined using information criteria such as those proposed in Bai and Ng (2002), Alessi et al. (2010) and Ahn and Horenstein (2013), or inferential techniques such as those developed in Onatski (2010), and Trapani (2018). This issue is discussed also in Section 5.1. For example, in the empirical analysis in Section 8, we use the criteria developed in
In what follows, Section 3.1 defines the steps of the proposed Expectation Maximization (EM) algorithm. Section 3.2 describes the Baum-Lindgren-Hamilton-Kim filter and smoother. Section 3.3 details the estimator for the factor space. Section 3.4 discusses the estimator for the parameters. Section 3.5 deals with initialization and convergence of the algorithm.

3.1 EM algorithm

The algorithm outlined in this section 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 $X$ given an estimate of the parameters $\hat{q}^{(k)}$, namely

$$
\log f(X; q) = E_{\hat{q}^{(k)}} \left[ \log f(G|X; q) \right] = \sum_{\xi_t \in \mathcal{X}} \sum_{g_t \in \mathcal{G}} E_{\hat{q}^{(k)}} \left[ \log f(X|\xi_t, g_t) \right] + \log f(G|X; q) - E_{\hat{q}^{(k)}} \left[ \log f(G|X; q) | X \right].
$$

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

$$
\left(\varphi^{(k+1)}, \rho^{(k+1)}\right) = \arg \max_{\varphi, \rho} E_{\hat{q}^{(k)}} [\log f(X|\varphi, \rho) | X]
$$

s.t. $P_{\ell_2} = \ell_2$, \hspace{1cm} (15)

where the constraints ensure that probabilities add up to one. In principle, in the M step we should also account for the term $E_{\hat{q}^{(k)}} [\log f(G|X; q) | X]$, 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 increasing log-likelihoods. Indeed, $E_{\hat{q}^{(k)}} [\log f(G|X; q) | X]$ 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 the Maximum Likelihood estimator of $q$. As shown below, the solution of the M step can be computed using the expressions given in (13) and (14). This solution is unique and in closed form. Therefore, no identification issue arises due to multiple maxima, or related to the existence of such maxima.

3.2 Baum-Lindgren-Hamilton-Kim filter and smoother

From (13) and (14), in order to compute the expected likelihood in the E step we need to compute $E_{\hat{q}^{(k)}} [\xi_t | X]$, $E_{\hat{q}^{(k)}} [\xi_t \otimes g_t | X]$, and $E_{\hat{q}^{(k)}} [(\xi_t \otimes g_t)(\xi_t \otimes g_t)'] | X] = E_{\hat{q}^{(k)}} [(I_2 \otimes g_t g_t') | X]$. We start by considering the case in which both $\{g_t\}_{t=1}^T$ is observed and the true value of the parameters $q$ is known, while we postpone the discussion of the estimation of the factors to Section 3.3. Then, for the E step we just need to compute $E[\xi_t | X]$, since in this case $\xi_t$ and $g_t$ 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 Appendix A.1.
is an iterative procedure through which we first compute the sequences of conditional one-step-ahead predicted probabilities \( \{\xi_{t|t-1}\}_{T=1}^{T} \) such that \( \xi_{t|t-1} = E[\xi_{t}|X_{t-1}] \), and filtered probabilities \( \{\xi_{t|t}\}_{T=1}^{T} \) such that \( \xi_{t|t} = E[\xi_{t}|X_{t}] \). Second, by means of those sequences, we compute the sequence of smoothed probabilities \( \{\xi_{t|T}\}_{T=1}^{T} \) such that \( \xi_{t|T} = E[\xi_{t}|X_{t}] \).

The final recursions for the filtered probabilities are given by (e.g., see Krolzig, 2013, and Hamilton, 1989)

\[
\xi_{t|t-1} = \mathbf{P}' \xi_{t-1|t-1}, \quad t = 1, \ldots, T,
\]

\[
\xi_{t|t} = \frac{\eta_t \odot \xi_{t|t-1}}{\nu_t (\eta_t \odot \xi_{t|t-1})}, \quad t = 1, \ldots, T, \tag{16}
\]

where

\[
\eta_t = \begin{bmatrix}
  f \left( x_t | \xi_t = [1 0]' \right), g_t \\
  f \left( x_t | \xi_t = [0 1]' \right), g_t 
\end{bmatrix}.
\]

The filter can be started by setting either \( \xi_{0|0} = [1 0]' \), or, equivalently, \( \xi_{0|0} = [0 1]' \).

The final recursions for the smoothed probabilities are given by (e.g., see Krolzig, 2013, and Kim, 1994)

\[
\xi_{t|T} = \left[ \mathbf{P} \left( \xi_{t+1|T} \odot \xi_{t+1|t} \right) \right] \odot \xi_{t|t}, \quad t = 1, \ldots, T. \tag{17}
\]

This backward recursion is initiated at \( \xi_{T|T} \), which is the last iteration of the filter in (16).

The above description of the Baum-Lindgren-Hamilton-Kim filter and smoother assumes that \( q \) and \( g_t \) are observed. However, in practice both need to be estimated. This is discussed in the next two Sections 3.3 and 3.4 below.

### 3.3 Estimating the factor space

In order to estimate the factors \( g_t \), and their dimension \( r_1 + r_2 \), we exploit the fact that the Markov switching factor model in (1) is observationally equivalent to a linear factor model with \( r_1 + r_2 \) common factors \( g_t \) and factor loadings \( A \): see Section 2.3 and, in particular, equation (10). The number of factors in (10) can be estimated using methods already available in the literature: for example, see Bai and Ng (2002), Onatski (2010), Ahn and Horenstein (2013), and Trapani (2018). The factors \( g_t \) can be estimated by PCA as follows. First, the estimator \( \hat{A} \) of the loadings matrix \( A \) 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 \( T^{-1} \sum_{t=1}^{T} x_t x_t' \). Second, the factors are estimated by linear projection of the data \( x_t \) onto the estimated loadings:

\[
\hat{g}_t = \left( \hat{A}' \hat{A} \right)^{-1} \hat{A}' x_t = \frac{1}{N} \hat{A}' x_t, \quad t = 1, \ldots, T. \tag{18}
\]
This is the same approach followed by Stock and Watson (2002a). It is also the dual approach of the one adopted by Bai (2003). Consistency of $\hat{A}$ and $\hat{g}_t$ follow from Lemma 1 and Lemma 5(a) in Appendix B, respectively. Note that the steps described in this section do not require knowing the latent state indicator $\xi_t$, and they can be carried out independently. Because of these results, $\xi_t$ and $\hat{g}_t$ can also be treated as independent for all $t$. As a consequence, the Baum-Lindgren-Hamilton-Kim filter described in Section 3.2 can be implemented by just replacing the true factors $g_t$ with their estimator $\hat{g}_t$ defined in (18).

### 3.4 Estimating the parameters

At each iteration $k \geq 0$ of the EM algorithm, the filtered and smoothed probabilities, given in (16) and (17), respectively, and the smoothed cross-probabilities given in (A.10), are computed using an estimator $\hat{q}^{(k)}$ of the parameters and an estimator $\hat{g}_t$ of the factors. Hereafter, we denote as $\xi^{(k)}_{t|T}$, $\xi^{(k)}_{t-1|T}$, and $\xi^{(k)}_{t|t}$ such estimators. This defines the E step.

In the M step we have to solve the constrained maximization problem in (15). Here we just give the final results, while we refer to Appendix A.2 for their derivation. The estimates of the loadings $B_j$, $j = 1, 2$, are given by

$$\hat{B}^{(k+1)}_j = \left( \sum_{t=1}^T \xi^{(k)}_{j,t|T} x_t \hat{g}_t \right) \left( \sum_{t=1}^T \xi^{(k)}_{j,t|T} \hat{g}_t \hat{g}_t' \right)^{-1}, \quad j = 1, 2,$$

(19)

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

$$\hat{\Sigma}^{(k+1)}_{e_j} = \left( \sum_{t=1}^T \left( x_{it} - \hat{B}^{(k+1)'}_{ji} \hat{g}_t \right)^2 \right) / \sum_{t=1}^T \xi^{(k)}_{j,t|T}, \quad i = 1, \ldots, N, \quad j = 1, 2,$$

(20)

and

$$\hat{\Sigma}^{(k+1)}_{e_{jk}} = 0, \quad i, k = 1, \ldots, N, \quad i \neq k, \quad j = 1, 2,$$

where $\hat{B}^{(k+1)'}_{ji}$ is the $i$th row of $\hat{B}^{(k+1)}_j$. Concerning the estimates of $\rho$ which are subject to the adding up condition,

$$\hat{\rho}^{(k+1)} = \left[ \sum_{t=1}^T \xi^{(k)}_{t,t-1|T} \right] \otimes \left[ \sum_{t=0}^{T-1} \xi^{(k)}_{t|T} \right].$$

(21)

By letting $k^*$ be the last iteration of the EM algorithm, we define our final estimator of the parameters as $\hat{q} \equiv \hat{q}^{(k^*+1)}$, as given by (19), (20), and (21). The final estimator of $\xi_t$ is defined as $\hat{\xi}_{t|T} \equiv \hat{\xi}_{t|T}^{(k^*+1)}$, i.e., obtained by running one last time the Baum-Lindgren-Hamilton-Kim filter using the final estimates of the parameters.
3.5 Initialization and convergence of the EM algorithm

To start the algorithm we need initial estimators \( \hat{\theta}^{(0)} \) for the parameters. Specifically, we set \( \hat{B}_1^{(0)} = \hat{B}_2^{(0)} = \hat{\Lambda} \), as defined in Section 3.3. Then, given also \( \hat{\theta}_t \) as in (18), let \( \hat{e}_t = x_t - \hat{\Lambda} \hat{g}_t \), and we set \( \hat{\Sigma}_t^{(0)} = \hat{\Sigma}_t^{(0)} = \text{diag}(T^{-1} \sum_{t=1}^T \hat{e}_t \hat{e}_t') \). Finally, we set

\[
\hat{\Theta}^{(0)} = \begin{pmatrix} 0.5 + \omega_1 & 1 - 0.5 - \omega_1 \\ 1 - 0.5 - \omega_2 & 0.5 + \omega_2 \end{pmatrix},
\]

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:

\[
\frac{1}{2} \left| \log f \left( X \mid \Theta^{(k)} \right) - \log f \left( X \mid \Theta^{(k-1)} \right) \right| < \epsilon,
\]

for some a priori chosen threshold \( \epsilon > 0 \).

4 Asymptotic theory

In what follows, Section 4.1 states the assumptions, whereas Section 4.2 presents the asymptotic properties of the estimators.

4.1 Assumptions

For ease of reference, let us write (11) and (10) in scalar notation as

\[
x_{it} = \sum_{j=1}^{2} \lambda_j f_{jt}(s_t = j) + e_{it} = a'_j g_t + e_{it}, \quad i = 1, \ldots, N, \ t \in \mathbb{Z}.
\]

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

Assumption 1. Factors.

(a) For \( j = 1, 2 \), and all \( t \in \mathbb{Z} \), \( E[f_{jt}] = 0 \) and \( E[\|f_{jt}\|^4] < \infty \).

(b) For \( j, k = 1, 2 \), as \( T \to \infty \), \( T^{-1} \sum_{t=1}^T \mathbb{I}(s_t = j) h_{kt} f_{jt} f_{jt}' \overset{P}{\to} \Sigma_{f_j}^{(k)} \), where \( \Sigma_{f_j}^{(k)} \) is \( r_j \times r_j \) positive definite, and \( \{h_{kt}\}_{t=1}^T \) is any sequence such that (i) \( P \{0 \leq h_{kt} \leq 1\} = 1 \) and (ii) \( T^{-1} \sum_{t=1}^T h_{kt} \overset{P}{\to} \bar{h}_k > 0 \).

Assumption 1 restricts the factor processes \( \{f_{jt}\} \), for \( j = 1, 2 \), so that appropriate moments exist. 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 its actual value. Assumption
implies that \( 0 < \mathbb{P} [s_t = j] < 1 \), for \( j = 1, 2 \), thus ruling out the possibility that any of the states is absorbing, as discussed in Section 2. It also implies that for \( j = 1, 2 \), as \( T \to \infty \),

\[
\frac{1}{T} \sum_{t=1}^{T} \mathbb{1} (s_t = j) f_{jt} f_{jt}' \to \Sigma_{fj},
\]

where \( \Sigma_{fj} \) is positive definite and

\[
\frac{1}{T} \sum_{t=1}^{T} g_{jt} g_{jt}' \to \Sigma_{g} = \left( \begin{array}{cc} \Sigma_{f1} & 0 \\ 0 & \Sigma_{f2} \end{array} \right).
\]

In particular, note that (22) allows the covariance matrix of \( f_j \) to be state-dependent, as advocated in Massacci (2023). It is also easy to see that if \( j \neq k \), then for all \( T \in \mathbb{N} \)

\[
\frac{1}{T} \sum_{t=1}^{T} \mathbb{1} (s_t = j) f_{jt} f_{kt}' \mathbb{1} (s_t = k) = 0.
\]

**Assumption 2. Loadings.**

(a) For \( j = 1, 2 \), all \( i = 1, \ldots, N \), and all \( N \in \mathbb{N} \), \( \| \lambda_{ji} \| \leq \hat{\lambda} < \infty \), where \( \hat{\lambda} \) is independent of \( j, i, \) and \( N \).

(b) For \( j = 1, 2 \), as \( N \to \infty \), \( N^{-1} \Lambda_j' \Lambda_j \to \Sigma_{\Lambda_j} \), where \( \Sigma_{\Lambda_j} \) is \( r_j \times r_j \) positive definite.

(c) As \( N \to \infty \), \( N^{-1} \Lambda_1' \Lambda_2 \to \Sigma_{\Lambda_{12}} \), where \( \Sigma_{\Lambda_{12}} \) is \( r_1 \times r_2 \).

(d) For any \( r_2 \times r_2 \) full rank matrix \( L \), \( \Lambda_1 \neq \Lambda_2 L \).

According to Assumption 2, loadings are nonstochastic and factors have a nonnegligible effect on the variance of \( \{ x_t \} \) 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 \( \Lambda_1 \) are a linear combination of the columns of \( \Lambda_2 \), in which case the regimes cannot be separately identified. From Assumption 2 it also follows that, as \( N \to \infty \),

\[
\frac{A' A}{N} \to \Sigma_{A} = \left( \begin{array}{cc} \Sigma_{A_1} & \Sigma_{A_{12}} \\ \Sigma_{A_{12}}' & \Sigma_{A_2} \end{array} \right),
\]

and

\[
\frac{B_1' B_1}{N} \to \Sigma_{B_1} = \left( \begin{array}{cc} \Sigma_{A_1} & 0 \\ 0 & 0 \end{array} \right), \quad \frac{B_2' B_2}{N} \to \Sigma_{B_2} = \left( \begin{array}{cc} 0 & 0 \\ 0 & \Sigma_{A_2} \end{array} \right), \quad \frac{B_j' B_k}{N} \to 0, \quad \text{if } j \neq k.
\]

**Assumption 3. Idiosyncratic component.**

(a) For all $i = 1, \ldots, N$, all $t \in \mathbb{Z}$, and all $N \in \mathbb{N}$, $E[e_{it}] = 0$ and $E[e_{it}^8] \leq M < \infty$, where $M$ is independent of $i$, $t$, and $N$.
(b) For $j, k = 1, 2$, for all $t \in \mathbb{Z}$, and $N \in \mathbb{N}$,

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

where $\{h_{kt}\}_{t=1}^{T}$ is as in Assumption 2(b), and $M$ is independent of $t$ and $N$.
(c) For $j, k = 1, 2$, all $i, l = 1, \ldots, N$, all $N \in \mathbb{N}$, and all $T \in \mathbb{N}$,

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

where $\{h_{kt}\}_{t=1}^{T}$ is as in Assumption 2(b), and $M$ is independent of $j$, $i$, $l$, $N$, and $T$.

Part (b) of Assumption 3 controls the amount of cross-sectional correlation we can allow for. It implies the usual assumption for approximate factor models of nondiagonal idiosyncratic covariances $\Sigma_{e_j}$, $j = 1, 2$. Note that the sequence $\{h_{kt}\}_{t=1}^{T}$ has the same role as in Assumption 1, which we refer to for further comments. Part (b) of Assumption 3 also implies

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

and hence $N^{-1/2}\|\mathbb{I}(s_t = j) e_t\| = O_p(1)$ for $j = 1, 2$, and for all $t \in \mathbb{Z}$. Part (c) of Assumption 3 limits time dependence, and it is guaranteed together with part (a) if we assume 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.

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

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

where $\{h_{kt}\}_{t=1}^{T}$ is as in Assumption 2(b), and $M$ is independent of $N \in \mathbb{N}$ and $T \in \mathbb{N}$.

Assumption 4 limits the degree of dependence between factors, state variable $s_t$, and idiosyncratic components.

**Assumption 5. Eigenvalues.** The eigenvalues of the $(r_1 + r_2) \times (r_1 + r_2)$ matrix $\Sigma_A \Sigma_g$ are distinct, where $\Sigma_A$ is defined in (25) and $\Sigma_g$ is defined in (23).
Assumption 5 guarantees a unique limit for $N^{-1}A'\hat{A}$, as stated in Lemma 6 in Appendix B. By assuming distinct eigenvalues, we can uniquely identify the space spanned by the eigenvectors, which are linear combinations of the columns of $A$. Notice that $\Sigma_g$ is block diagonal because of (24).

Assumptions 1 to 5 are sufficient to prove the consistency of the estimators we propose. In order to derive their asymptotic distributions, we further introduce the following Assumptions (6) and (7).

Assumption 6. Moments and Central Limit Theorems.

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

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

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

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

$$E \left[ \left\| \frac{1}{\sqrt{N}T} \sum_{t=1}^{T} \sum_{i=1}^{N} \lambda_{ki} f_t' e_{it} \right\|^2 \right] \leq M < \infty,$$

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

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

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} I(s_t = j) h_{kt} f_t' e_{it} \xrightarrow{d} \mathcal{N}(0, \Gamma_{jki}),$$

where $\{h_{kt}\}_{t=1}^{T}$ is defined in Assumption 7 and

$$\Gamma_{jki} = \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \sum_{i=1}^{N} I(s_t = j) I(s_{i'} = j) h_{kt} h_{k'i'} E[f_t f_{i'}' e_{it} e_{it'}].$$

(d) For all $t \in \mathbb{Z}$, as $N \to \infty$,

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \begin{bmatrix} \lambda_{1i} \\ \lambda_{2i} \end{bmatrix} e_{it} \xrightarrow{d} \mathcal{N} \left( 0, \begin{bmatrix} \Phi_{1t} & \Phi_{12t} \\ \Phi_{12t}' & \Phi_{2t} \end{bmatrix} \right),$$

where for $j, k = 1, 2$

$$\Phi_{jkt} = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \sum_{t=1}^{T} \lambda_{ji} x_{kt} E[e_{it} e_{it'}],$$

and $\Phi_{jkt} = \Phi_{jkt}$. Parts (a) and (b) of Assumption 6 are suitable moment bounds, whereas parts (c) and (d)
are central limit theorems.

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

Assumption 7 imposes standard restrictions on the convergence rates.

Define the \((r_1 + r_2) \times (r_1 + r_2)\) matrix \( \hat{H} \) as

\[
\hat{H} = \frac{GG' \hat{\Lambda} \hat{A}}{N} \hat{V}^{-1},
\]

where \( G = (g_1, \ldots, g_T) \) and \( \hat{V} \) is the \((r_1 + r_2) \times (r_1 + r_2)\) diagonal matrix containing the first \( r_1 + r_2 \) eigenvalues of \( \hat{\Sigma}_x = (NT)^{-1} \sum_{t=1}^{T} x_tx'_t \) sorted in decreasing order. In Lemma 6 we prove that

\[
p \lim_{N,T \to \infty} \frac{\Lambda' \hat{A}}{N} = Q, \text{ with } Q = \Sigma_g^{-1/2} \Psi \Sigma_{1/2}, \quad (27)
\]

where \( V \) is the \((r_1 + r_2) \times (r_1 + r_2)\) diagonal matrix of the first \((r_1 + r_2)\) eigenvalues of \( \Sigma_{1/2} \hat{A} \Sigma_{1/2} \) in decreasing order, and \( \Psi \) is the corresponding matrix of eigenvectors such that \( \Psi' \Psi = I_{r_1+r_2} \). Likewise define \( Q_j = \lim_{N,T \to \infty} N^{-1} \Lambda_j' \hat{A} \), for \( j = 1, 2 \), which is an \( r_j \times (r_1 + r_2) \) matrix such that \( Q = [Q'_1, Q'_2]' \). Thus, by Lemma 7 we have

\[
Q_j = \Sigma_{r_j}^{-1/2} \Psi_j V^{1/2}, \quad j = 1, 2,
\]

where \( \Psi_j \) is the \( r_j \times (r_1 + r_2) \) matrix such that \( \Psi = [\Psi'_1, \Psi'_2]' \). Therefore, because of (23), (28), and by Lemma 8 according to which \( \hat{V} \xrightarrow{p} V \),

\[
p \lim_{N,T \to \infty} \hat{H} = H, \text{ with } H = \Sigma_g Q V^{-1}. \quad (30)
\]

### 4.2 Asymptotic results

For \( j = 1, 2 \), let \( \hat{B}_j = \hat{\mathcal{B}}^{(k^*_j+1)}_j \), where \( k^* \) is the last iteration of the EM algorithm as defined in Section 3.4. For given \( j = 1, 2 \) and \( i = 1, \ldots, N \), let \( \hat{b}_{ji} \) be the estimator for \( b_{ji} \) such that \( \hat{B}_j = [\hat{b}_{j1}, \ldots, \hat{b}_{jN}]' \) and \( B_j = [b_{j1}, \ldots, b_{jN}]' \). The following theorem states the asymptotic distribution of \( \hat{b}_{ji} \).

**Theorem 1.** Let Assumptions 7 and 8 hold. Then, for \( k_1, k_2 = 1, 2 \) with \( k_1 \neq k_2 \), for any given \( i = 1, \ldots, N \), as \( N, T \to \infty \),

\[
\sqrt{T} \left[ \hat{b}_{k_1i} - \hat{I}'_{\xi_{k_1i}} \hat{H}' \hat{b}_{k_1i} - \left( \hat{I}_{r_1+r_2} - \hat{I}'_{\xi_{k_1i}} \right)' \hat{H}' \hat{b}_{k_2i} \right] \xrightarrow{d} \mathcal{N} \left( 0, \Sigma_{b_{k_1i}} \right),
\]
where the \((r_1 + r_2) \times (r_1 + r_2)\) matrix \(\hat{\mathbf{I}}_{\hat{\xi}_{k_1}}\) is defined as

\[
\hat{\mathbf{I}}_{\hat{\xi}_{k_1}} = \left( \sum_{t=1}^{T} \hat{\xi}_{k_1,t} \mathbb{I} \| s_t = k_1 \rangle \hat{\mathbf{g}}_t \hat{\mathbf{g}}_t^\prime \right)^{-1},
\]

and where

\[
\Sigma_{b_{k_1i}} = \left( Q_1^\prime \Sigma^{(k_1)} Q_1 + Q_2^\prime \Sigma^{(k_1)} Q_2 \right)^{-1} \left( Q_1^\prime \Gamma_{k_1} Q_1 + Q_2^\prime \Gamma_{2k_1} Q_2 \right) \left( Q_1^\prime \Sigma^{(k_1)} Q_1 + Q_2^\prime \Sigma^{(k_1)} Q_2 \right)^{-1},
\]

with \(Q_j, \Gamma_{jk_1}, \text{ and } \Sigma^{(k_1)}\), \(j = 1, 2\), defined in (29), Assumption (D(c), and Assumption (H) when \(h_{k_1} = \hat{\xi}_{k_1,t}|T\), respectively.

Theorem \(\text{H}\) shows that the estimator \(\hat{b}_{k_1i}\) for \(b_{k_1i}\) 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 it is captured by the invertible matrix \(\hat{\mathbf{H}}\) defined in (27) (see Bai, 2003). If we assume \(T^{-1} \sum_{t=1}^{T} \mathbf{g}_t \mathbf{g}_t^\prime = \mathbf{I}_{r_1 + r_2}\), then \(\hat{\mathbf{H}}\) becomes a rotation, namely an orthogonal matrix. However, additional restrictions on the loadings are necessary to reduce \(\hat{\mathbf{H}}\) to the identity: for a discussion on identification of factors see inter alia Bai and Ng (2013). The second source of bias is induced by \(\hat{\mathbf{I}}_{\hat{\xi}_{k_1}}\) defined in (31), which depends on the probability of the state being asymptotically correctly estimated. If the unconditional probability of being in state \(k_1\) were correctly estimated with probability one, that is, \(\hat{\xi}_{k_1,t} \xrightarrow{p} \mathbb{I} \| s_t = k_1 \rangle\), as \(N, T \to \infty\), then \(\hat{\mathbf{I}}_{\hat{\xi}_{k_1}} \xrightarrow{p} \mathbf{I}_{r_1 + r_2}\) and \(\hat{b}_{k_1i}\) would consistently estimate a linear transformation of \(b_{k_1i}\).

Therefore, \(\hat{b}_{k_1i}\) estimates a linear transformations of \(b_{k_1i}\) and \(b_{k_2i}\), with weights determined by \(\hat{\mathbf{I}}_{\hat{\xi}_{k_1}}\) and \((\mathbf{I}_{r_1 + r_2} - \hat{\mathbf{I}}_{\hat{\xi}_{k_1}})\), respectively. This second source of bias is due to the fact that the process \(s_t\) is latent, and it is specific to Markov switching models. As such, it does not affect threshold or structural break models, in which the state is identified with probability one.

Theorem \(\text{H}\) has implications for the estimation of the regime specific loadings \(\Lambda_j\), \(j = 1, 2\). To see this, let \(\hat{\mathbf{R}}_k = \hat{\mathbf{H}} \hat{\mathbf{I}}_{\hat{\xi}_k}\), for \(k = 1, 2\), and consider the partition

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

where \(\hat{\mathbf{R}}_{k,j\ell}, k, j, \ell = 1, 2\) and \(\hat{\mathbf{H}}_{j\ell}, j, \ell = 1, 2\), are \(r_j \times r_\ell\). Then, from Theorem \(\text{H}\) for any given \(i = 1, \ldots, N\), as \(N, T \to \infty\), we obtain

\[
\sqrt{T} \left\{ \hat{\mathbf{b}}_{i_j} - \begin{bmatrix} \Lambda_j \circ 0 \end{bmatrix} \left( \hat{\mathbf{H}} - \hat{\mathbf{R}}_1 \right) \right\} = \sqrt{T} \left\{ \hat{\mathbf{b}}_{i_j} - \begin{bmatrix} \Lambda_j \circ \hat{\mathbf{R}}_{11}, \hat{\mathbf{R}}_{1,11} \end{bmatrix} - \Lambda_{j1} \begin{bmatrix} \hat{\mathbf{H}}_{21} - \hat{\mathbf{R}}_{1,21} \end{bmatrix} \right\} \xrightarrow{d} \mathcal{N}(0, \Sigma_{b_{1i}})\),
\]

(33)
and
\[ \hat{\mathbf{b}}_{2i} - [0 \ \mathbf{X}]_{2i} \hat{\mathbf{R}}_2 = \left[ \mathbf{X}_i' \ 0 \right] \left( \hat{\mathbf{H}} - \hat{\mathbf{R}}_2 \right) \]
\[ = \sqrt{T} \{ \hat{\mathbf{b}}_{2i} - \mathbf{X}_i' \left[ \hat{\mathbf{H}}_{11} - \hat{\mathbf{R}}_{2,11} \right] \left( \hat{\mathbf{H}}_{12} - \hat{\mathbf{R}}_{2,12} \right) \} \overset{d}{\to} \mathcal{N} \left( \mathbf{0}, \Sigma_{b_2} \right) . \] (34)

This means that \( r_1 + r_2 \) columns of \( \mathbf{B}_j \), \( j = 1, 2 \), estimate two different linear transformations of the columns of \([\mathbf{A}_1 \ \mathbf{A}_2]\). We can distinguish two cases. On the one hand, if \( r_1 = r_2 = r \), as assumed for example in Liu and Chen (2016), 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. Indeed, in this case \( \mathbf{B}_1 \) and \( \mathbf{B}_2 \) have an even number of columns, equal to \( 2r \), and from the first line of (33) and (34) we see that we can consider the first half of the columns of either \( \hat{\mathbf{B}}_1 \) or \( \hat{\mathbf{B}}_2 \) as an estimator of a linear transformation of \( \mathbf{A}_1 \) and the second half of the columns of either \( \hat{\mathbf{B}}_1 \) or \( \hat{\mathbf{B}}_2 \) as an estimator of a linear transformation of \( \mathbf{A}_2 \). Hence, we can define the following estimators of the loadings:

\[ \hat{\lambda}_i = \hat{\mathbf{b}}_{1i,1:r}, \quad \hat{\lambda}_i = \hat{\mathbf{b}}_{2i,r+1:2r}, \quad i = 1, \ldots, N, \] (35)

or

\[ \tilde{\lambda}_i = \hat{\mathbf{b}}_{2i,1:r}, \quad \tilde{\lambda}_i = \hat{\mathbf{b}}_{1i,r+1:2r}, \quad i = 1, \ldots, N, \] (36)

where \( \hat{\mathbf{b}}_{ji,1:r} \) denotes the first \( r \) elements of \( \hat{\mathbf{b}}_{ji} \), and \( \hat{\mathbf{b}}_{ji,r+1:2r} \) denotes the second \( r \) elements of \( \hat{\mathbf{b}}_{ji} \), for \( j = 1, 2 \) and \( i = 1, \ldots, N \). The property of these estimators are formalized in the following corollary, which is a direct consequence of Theorem II and of (33) and (34).

**Corollary 1.** Let Assumptions 1 - 7 hold and assume \( r_1 = r_2 = r \). Then, for any given \( i = 1, \ldots, N, \) as \( N, T \to \infty, \)

\[ \sqrt{T} \{ \hat{\lambda}_i' - \lambda_i' \hat{\mathbf{R}}_1 \} \overset{d}{\to} \mathcal{N} \left( 0, \Sigma_{b_1} \right), \quad \sqrt{T} \{ \hat{\lambda}_i' - \lambda_i' \hat{\mathbf{R}}_2 \} \overset{d}{\to} \mathcal{N} \left( 0, \Sigma_{b_2} \right), \]

\[ \sqrt{T} \{ \hat{\lambda}_i' - \lambda_i' \left( \hat{\mathbf{H}} - \hat{\mathbf{R}}_2 - \mathbf{X}_i' \right) \} \overset{d}{\to} \mathcal{N} \left( 0, \Sigma_{b_1} \right), \quad \sqrt{T} \{ \hat{\lambda}_i' - \lambda_i' \left( \hat{\mathbf{H}} - \hat{\mathbf{R}}_1 \right) \} \overset{d}{\to} \mathcal{N} \left( 0, \Sigma_{b_2} \right). \]

On the other hand, if \( r_1 \neq r_2 \), we need consistent estimators of \( r_1 \) and \( r_2 \) in order to be able to isolate the first \( r_1 \) columns of \( \hat{\mathbf{B}}_1 \) and the last \( r_2 \) columns of \( \hat{\mathbf{B}}_2 \), respectively, which are consistent estimators of a linear transformation of the columns of \( \mathbf{A}_1 \) and \( \mathbf{A}_2 \), 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 \( \mathbf{A}_j \), \( j = 1, 2 \).

Theorem II describes the asymptotic properties of the estimator for the factor loadings \( \hat{\mathbf{B}}_1 \) and \( \hat{\mathbf{B}}_2 \). Complementary results can be obtained with respect to the estimated factors associated to the loading matrices \( \hat{\mathbf{B}}_1 \) and \( \hat{\mathbf{B}}_2 \). Formally, the true factors that correspond
to \( B_1 \) and \( B_2 \) are \( \xi_{1t} g_t \) and \( \xi_{2t} g_t \), respectively, and their estimators are \( \hat{\xi}_{1,1T} g_t \) and \( \hat{\xi}_{2,1T} g_t \), respectively. The following theorem states the asymptotic distribution of these estimators.

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

\[
\sqrt{N} \left\{ \begin{pmatrix} \hat{\xi}_{1,1T} g_t \\ \hat{\xi}_{2,1T} g_t \end{pmatrix} - \hat{H}_\xi^{-1} \begin{pmatrix} \xi_{1t} g_t \\ \xi_{2t} g_t \end{pmatrix} \right\} \Rightarrow N \left( 0, \Sigma_{\xi g t} \right),
\]

where

\[
\hat{H}_\xi = \begin{bmatrix} \hat{H}_{\xi 1} & \hat{H} \left( I_{r_1 + r_2} - \hat{I}_{\xi_2} \right) \\ \hat{H} \left( I_{r_1 + r_2} - \hat{I}_{\xi_1} \right) & \hat{H}_{\xi 2} \end{bmatrix},
\]

with \( \hat{H} \) and \( \hat{I}_{\xi j} \) defined in (27) and (31), respectively, and where

\[
\Sigma_{\xi g t} = \left( H_\xi \begin{pmatrix} \Sigma_{B_1} & 0 \\ 0 & \Sigma_{B_2} \end{pmatrix} H_\xi' \right)^{-1} \left( H_\xi \Sigma_{B \xi e} H_\xi' \right) \left( H_\xi \begin{pmatrix} \Sigma_{B_1} & 0 \\ 0 & \Sigma_{B_2} \end{pmatrix} H_\xi' \right)^{-1},
\]

where \( \Sigma_{B_j}, j = 1, 2 \), is defined in (26),

\[
\Sigma_{B \xi e} = \begin{pmatrix} \Phi_{1t} & 0 & 0 & \Phi_{12t} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \Phi_{12t} & 0 & 0 & \Phi_{2t} \end{pmatrix},
\]

with \( \Phi_{jt} \) and \( \Phi_{jkt}, j, k = 1, 2 \), defined in Assumption 6(d), and where

\[
H_\xi = \begin{bmatrix} H_{\xi 1} & H \left( I_{r_1 + r_2} - I_{\xi 2} \right) \\ H \left( I_{r_1 + r_2} - I_{\xi 1} \right) & H_{\xi 2} \end{bmatrix},
\]

with \( H \) defined in (30) and

\[
I_{\xi j} = \lim_{N, T \to \infty} I_{\xi j} = H^{-1} \begin{bmatrix} I (j = 1) I_{r_1} & 0 \\ 0 & I (j = 2) I_{r_2} \end{bmatrix} H,
\]

as defined in Lemma 9 in Appendix B.

In general, \( \hat{I}_{\xi 1} \neq I_{r_1 + r_2} \) and so also \( I_{\xi 1} \neq I_{r_1 + r_2} \). Then, because of Theorem 1, the estimator \( \hat{b}_{j1} \) is biased and it is straightforward to see that the asymptotic covariance in Theorem 2 is positive definite. Note that if we know that \( r_1 = r_2 = r \) holds, then we can build consistent estimators for linear combinations of \( f_{jt}, j = 1, 2 \), by simply regressing \( x_t \) onto the estimators \( \hat{\Lambda}_j \) or \( \hat{\Lambda}_j \) which are defined in (35) and (36), respectively, and, as shown in Corollary 1, are consistent for linear transformation of \( \Lambda_j \). Formally, this means we can build the sequence of
factor estimators by running the cross-sectional regressions

$$\hat{f}_{jt} = \hat{\xi}_{jt|T} \left( \hat{\Lambda}' \hat{\Lambda} \right)^{-1} \left( \hat{\Lambda}' x_t \right), \ j = 1, 2, \ t = 1, \ldots, T, \quad (37)$$

or

$$\hat{f}_{jt} = \hat{\xi}_{jt|T} \left( \hat{\Lambda}' \hat{\Lambda} \right)^{-1} \left( \hat{\Lambda}' x_t \right), \ j = 1, 2, \ t = 1, \ldots, T. \quad (38)$$

If the unconditional probability of being in a given state is correctly estimated then $\hat{\xi}_{jt} \to I \times I_{r_1 + r_2}$ as $N, T \to \infty$, and Theorem 2 is redundant: in this case, asymptotic normality of (37) and of (38) follows from arguments analogous to those in Bai (2003). In the more general case we are considering, the asymptotic distribution of $\hat{f}_{jt}$ is stated in the following theorem (an analogous result holds for $\hat{\xi}_{jt}$ and it is omitted for brevity).

**Theorem 3.** Let Assumptions 1 - 7 hold and $r_1 = r_2$. Then, for $j,k = 1,2$ with $j \neq k$, and for any given $t = 1, \ldots, T$, as $N, T \to \infty$,

$$\sqrt{N} \left\{ \hat{f}_{jt} - \left\{ \left( \frac{\left( \Lambda_j \hat{H}_{jj} + \Lambda_k \hat{H}_{kj} \right)'}{N} \left( \Lambda_j \hat{H}_{jj} + \Lambda_k \hat{H}_{kj} \right) \right)^{-1} \times \hat{\xi}_{jt|T} \left( \Pi(s_t = j) \Lambda_j f_{jt} + \Pi(s_t = k) \Lambda_k f_{kt} \right) \right\} \right\} \overset{d}{\to} N \left( 0, \Sigma_{f_{jt}} \right),$$

where

$$\Sigma_{f_{jt}} = \left( \xi_{jt}^* \right)^2 \left( H_{11}' \Phi_{1t} H_{11} + H_{jj}' \Phi_{jkt} H_{kj} + H_{kj}' \Phi_{jkt} H_{jj} + H_{22}' \Phi_{2t} H_{22} \right),$$

with $\xi_{jt}^* = p \lim_{N,T \to \infty} \hat{\xi}_{jt|T}$ and $\Phi_{1t}$, $\Phi_{2t}$, and $\Phi_{jkt}$, defined in Assumption 2(d).

According to Theorem 3, $\hat{f}_{jt}$ estimates the space spanned by either $f_{jt}$ or $f_{kt}$, for $j,k = 1,2$, with $j \neq k$, depending on which the true underlying regime is in period $t$.

## 5 On the number of regimes and factors

This section deals with two further issues related to the model in (1) and (2). Section 5.1 studies estimation of the number of factors. Section 5.2 discusses the consequences of an underspecified model.

### 5.1 Estimating the number of factors

Theorems 1 and 2 rely on the factor estimator $\hat{g}_t$ obtained from the equivalent linear representation in (10). This estimator does not embed any information related to the likelihood of observing a regime $j$ at a given point in time $t$, for $j = 1,2$ and $t \in Z$. We now study the
property of the estimator for the dimension of the factor space that is obtained when such information is accounted for.

Formally, for \( j = 1, 2 \), consider the covariance matrix

\[
\hat{\Sigma}_{\xi, x_j} = \frac{\sum_{t=1}^{T} \hat{\xi}_{jt} x_t x'_t}{N \sum_{t=1}^{T} \hat{\xi}_{jt}}, \tag{39}
\]

where \( 0 < \sum_{t=1}^{T} \hat{\xi}_{jt} < T \); \( \hat{\Sigma}_{\xi, x_j} \) includes information about the regimes through the estimated sequence \( \{\hat{\xi}_{jt}\}^{T}_{t=1} \). Define the \( r_j \times 1 \) vectors

\[
f_{jjt} = \mathbb{I}_{jt} f_{jt}, \quad f_{\xi, kj} = \hat{\xi}_{kt} f_{jt}, \quad j, k = 1, 2,
\]

and the \( r_j \times T \) matrices

\[
F_{jj} = (\mathbb{I}_{j1} f_{j1}, \ldots, \mathbb{I}_{jT} f_{jT}), \quad F_{\xi, kj} = (\hat{\xi}_{k1} f_{j1}, \ldots, \hat{\xi}_{kT} f_{jT}), \quad j, k = 1, 2.
\]

For \( 1 \leq p \leq \bar{p} \), with \( \bar{p} < \infty \), let \( \hat{\Lambda}^{(p)}_{\xi, j} \) be the \( p \times p \) diagonal matrix containing the first \( p \) eigenvalues of \( \hat{\Sigma}_{\xi, x_j} \) in decreasing order. Finally, let \( \hat{\Lambda}^{(p)}_{\xi, j} = [\hat{\lambda}^{(p)}_{\xi, j1}, \ldots, \hat{\lambda}^{(p)}_{\xi, jN}]' \) be the \( N \times p \) matrix estimator for \( \Lambda_j \), which is obtained as \( \sqrt{N} \times \) times the normalized eigenvectors corresponding to the \( p \) largest eigenvalues of the \( N \times N \) sample covariance matrix \( \hat{\Sigma}_{\xi, x_j} \) in (39). The following theorem characterises the mean square convergence of \( \hat{\lambda}^{(p)}_{\xi, ji} \) for a given value of \( p \).

**Theorem 4.** Let Assumptions 1 - 4 hold. Then, for any fixed \( 1 \leq p \leq \bar{p} \) with \( \bar{p} < \infty \), and for \( j, k = 1, 2 \) with \( j \neq k \), there exists \( r_j \times p \) matrices \( \hat{\Lambda}^{(p)}_{\xi, kj} \) such that

\[
\hat{\Lambda}^{(p)}_{\xi, ji} \hat{\Lambda}^{(p)}_{\xi, ik} = \frac{F_{\xi, kj} F'_{jj} \hat{\Lambda}^{(p)}_{\xi, j} \hat{\Lambda}^{(p)}_{\xi, kj}}{\sum_{t=1}^{T} \xi_{jt}} \tag{40}
\]

with rank \( \hat{\Lambda}^{(p)}_{\xi, kj} \) = \( \min \{r_j, p\} \), which satisfy

\[
\min \{N, T\} \left\{ \frac{1}{N} \sum_{i=1}^{N} \left\| \hat{\lambda}^{(p)}_{\xi, ji} - \left( \hat{\Lambda}^{(p)}_{\xi, jj} \lambda_{ji} + \hat{\Lambda}^{(p)}_{\xi, kj} \lambda_{ki} \right) \right\| \right\} = O_p(1).
\]

**Theorem 4** extends Theorem 1 in [Bai and Ng (2002)](2002) and Theorem 3.4 in [Massacci (2017)](2017) to the case of the Markov switching factor model in (I) and (2). For \( j, k = 1, 2 \) with \( j \neq k \), the theorem shows that \( \hat{\lambda}^{(p)}_{\xi, ji} \) estimates a linear combination of the vector \( (\lambda'_{ji}, \lambda'_{ki})' \) and not just of \( \lambda_{ji} \). It implies that the dimension of the estimated underlying factor space is \( r_1 + r_2 \) even when the available information about the regimes is accounted for. Imperfect knowledge of the regimes therefore leads to an enlarged factor space: this makes our setting analogous to large dimensional change point factor models, as previously discussed in Section 4.3. This comple-
ments what proved in Breitung and Eickmeier (2011), and Corradi and Swanson (2014), who show that model misspecification in the form of omitted discrete regime shifts leads to an inflated number of factors.

The result in Theorem 4 has practical implications. Let us assume that the number of factors is constant across regimes, as in Liu and Chen (2016). Then, if the estimated number of factors is even, we can easily recover the number of factors within each regime. If the estimated number is odd and greater than one, there might be a neglected regime, as discussed in Section 5.2 below. Obviously if we find evidence of just one common factor, then only the idiosyncratic covariances can be regime specific, while the loadings remain constant.

5.2 The case of an underspecified number of regimes

Up to now we have a priori assumed that the data are generated according to the model with two regimes in (1) and (2). This is consistent with existing empirical studies employing Markov switching models: for example, see Diebold and Rudebusch (1996). However, in some cases the underlying data generating process of the dependent variables of interest displays a higher number of regimes: for example, Guidolin and Timmermann (2006) show that the joint distribution of stock and bond returns requires a four-state model. Therefore, the two-regime specification in (1) and (2) leads to model misspecification in case the joint distribution of the dependent variables \( x_t \) is characterised by a higher number of regimes.

We now study the case in which the model is underspecified and the data are generated by a process with a number of regimes that is finite and greater than two.

Since the number of regimes is finite, without loss of generality we consider the model with three regimes

\[
x_t = \mathbb{I}(s_t = 1) \left( \Lambda_1 f_{1t} + \Sigma_{1e_1}^{1/2} e_t \right) + \mathbb{I}(s_t = 2) \left( \Lambda_2 f_{2t} + \Sigma_{2e_2}^{1/2} e_t \right) + \mathbb{I}(s_t = 3) \left( \Lambda_3 f_{3t} + \Sigma_{3e_3}^{1/2} e_t \right), \quad t \in \mathbb{Z},
\]

and let

\[
g_t = \begin{bmatrix} f_{1t} \\ 0 \\ 0 \end{bmatrix} \mathbb{I}(s_t = 1) + \begin{bmatrix} 0 \\ f_{2t} \\ 0 \end{bmatrix} \mathbb{I}(s_t = 2) + \begin{bmatrix} 0 \\ 0 \\ f_{3t} \end{bmatrix} \mathbb{I}(s_t = 3), \quad t \in \mathbb{Z}.
\]

Suppose that only two regimes are accounted for. Given a natural ordering of the regimes, this means that we have to consider two cases, namely: (a) \( s_t = 1 \) and \( s_t \neq 1 \); (b) \( s_t = 3 \) and...
$s_t \neq 3$. The model in (41) admits the following two equivalent two-regime representations

$$\begin{align*}
x_t &= \left( B_1^{(j)} B_2^{(j)} \right) \left( \xi_t^{(j)} \otimes g_t \right) + \left( \Sigma_{e1}^{(j),1/2} \Sigma_{e2}^{(j),1/2} \right) \left( \xi_t^{(j)} \otimes \xi_t \otimes I_N \right) e_t, \quad t \in \mathbb{Z}, \\
\xi_t^{(j)} &= P^{(j)} \xi_{t-1} + v_t^{(j)},
\end{align*}$$

(42)

where the loadings are defined as $B_1^{(1)} = (A_1 \ 0 \ 0)$, $B_2^{(1)} = (0 \ A_2 \ A_3)$, $B_1^{(3)} = (A_1 \ A_2 \ 0)$, $B_2^{(3)} = (0 \ 0 \ A_3)$, the latent state process is defined as

$$\begin{align*}
\xi_t^{(1)} &= \begin{bmatrix} \mathbb{I}(s_t = 1) \\
\mathbb{I}(s_t = 2) + \mathbb{I}(s_t = 3) \end{bmatrix}, \\
\xi_t^{(3)} &= \begin{bmatrix} \mathbb{I}(s_t = 1) + \mathbb{I}(s_t = 2) \\
\mathbb{I}(s_t = 3) \end{bmatrix},
\end{align*}$$

the idiosyncratic covariance matrices are defined as $\Sigma_{e1}^{(1)} = (\Sigma_{e1}^{1} \ 0 \ 0)$, $\Sigma_{e1}^{(3)} = (0 \ \Sigma_{e2} \ \Sigma_{e3})$, $\Sigma_{e2}^{(3)} = (0 \ 0 \ \Sigma_{e3})$, and the transition probabilities are equal to

$$\begin{align*}
P^{(1)} &= \begin{pmatrix} p_{11} & p_{1,\neq 1} \\
p_{\neq 1,1} & p_{\neq 1,\neq 1} \end{pmatrix} = \begin{pmatrix} p_{11} & 1 - p_{11} \\
1 - p_{\neq 1,\neq 1} & p_{\neq 1,\neq 1} \end{pmatrix}, \\
P^{(3)} &= \begin{pmatrix} p_{\neq 3,\neq 3} & p_{\neq 3,3} \\
p_{3,\neq 3} & p_{3,3} \end{pmatrix} = \begin{pmatrix} p_{3,\neq 3} & 1 - p_{3,\neq 3} \\
1 - p_{3,3} & p_{3,3} \end{pmatrix}.
\end{align*}$$

For $j = 1, 3$, define the vector of parameters $q^{(j)} = \left[ \varphi^{(j)}, \rho^{(j)} \right]'$, where

$$\varphi^{(j)} = \left[ \text{vec} \left( B_1^{(j)} \right)', \text{vec} \left( B_2^{(j)} \right)', \text{diag} \left( \Sigma_{e1}^{(j)} \right)', \text{diag} \left( \Sigma_{e2}^{(j)} \right) \right]'$$

and $\rho^{(j)} = \text{vec} \left( P^{(j)} \right)$.

Let $(NT)^{-1} \log f (X; q^{(j)})$ be the normalised log-likelihood function of (42). Assume that

$$E \left[ \frac{1}{NT} \log f (X; q^{(1)}) \right] > E \left[ \frac{1}{NT} \log f (X; q^{(3)}) \right].$$

(43)

In a likelihood sense, the condition in (42) captures a larger regime shift for $j = 1$ than for $j = 3$. Further, let $\hat{q}$ be the generic maximum likelihood estimator for the parameter of an underspecified model that allows for only two regimes when in fact the data generating process is given by (41).

We proceed by contradiction, see also Appendix [C] for more details. If $\hat{q}$ were an estimator for $q^{(3)}$, then

$$E \left[ \frac{1}{NT} \log f (X; \hat{q}) \right] - E \left[ \frac{1}{NT} \log f (X; q^{(1)}) \right] = -C + o_p (1),$$

(44)

which leads to a contradiction since $(NT)^{-1} \log f (X; \hat{q})$ is the estimated log-likelihood func-
tion. On the other hand, if $\hat{\mathbf{q}}$ were an estimator for $\mathbf{q}^{(1)}$, then
\[
E \left[ \frac{1}{NT} \log f (\mathbf{X}; \hat{\mathbf{q}}) \right] - E \left[ \frac{1}{NT} \log f (\mathbf{X}; \mathbf{q}^{(1)}) \right] = o_p(1).
\]

Therefore, when one regime is neglected, the maximum likelihood estimator estimates the regimes that maximise the likelihood according to the inequality in (43). Provided that a sufficient number of iterations is done, the EM algorithm proposed in Section 3 delivers an estimator that is close enough to the maximum likelihood estimator, such that the inequality in (43) is preserved: see Meng and Rubin (1993, 1994). Therefore, the EM algorithm delivers the estimator for the underspecified representation that is associated to the highest likelihood. This result is consistent with the homologous finding in Bai (1997), and Bai and Perron (1998), in relation to regression models with structural instability. Therefore, our result is the potential starting point for an inferential procedure on the number of regimes in large dimensional Markov switching factor models. It is also important to note that any neglected regime will be accounted for by an enlarged factor space, as discussed in Section 2.3.

6 Unobserved heterogeneity

The model in (11) assumes no individual effects. However, these may be important when modelling macroeconomic series as in Diebold and Rudebusch (1996). In our set up, individual effects can be introduced by extending Bai and Li (2012, 2016) and considering
\[
\mathbf{x}_t = (\alpha_1 + \Lambda_1 \mathbf{f}_1) I(s_t = 1) + (\alpha_2 + \Lambda_2 \mathbf{f}_2) I(s_t = 2) + \mathbf{e}_t,
\]
where $\alpha_j = (\alpha_{j1}, \ldots, \alpha_{jN})$, for $j = 1, 2$, and $\alpha_{ji}$ captures the individual effect of cross-sectional unit $i$ within regime $j$. The vectors $\alpha_1$ and $\alpha_2$ 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 (15) 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 centred variables $\mathbf{y}_t$ defined as
\[
\mathbf{y}_t = \mathbf{x}_t - E(\mathbf{x}_t) = \alpha_1 d_{1t} + \Lambda_1 \mathbf{f}_1 I(s_t = 1) + \alpha_2 d_{2t} + \Lambda_2 \mathbf{f}_2 I(s_t = 2) + \mathbf{e}_t,
\]
where $d_{jt} = I(s_t = j) - E[I(s_t = j)]$, $j = 1, 2$. If $\alpha_1 = \alpha_2$, $\mathbf{x}_t$ has the same expected value in both regimes, and $\mathbf{y}_t = \Lambda_1 \mathbf{f}_1 I(s_t = 1) + \Lambda_2 \mathbf{f}_2 I(s_t = 2) + \mathbf{e}_t$. In the more general case in which $\alpha_1 \neq \alpha_2$, unconditional demeaning leads to a larger factor space of dimension
\[ r_1 + r_2 + 2. \] The additional two factors \( d_{1t} \) and \( d_{2t} \) take only two values, namely \( d_{jt} = -E[I(s_t = j)] \) or \( d_{jt} = 1 - E[I(s_t = j)] \), depending on whether \( I(s_t = j) = 0 \) or \( I(s_t = j) = 1 \), respectively, for \( j = 1, 2 \). In this case, the equivalent linear representation in (10) holds with \( g_r = [d_{1t}I(s_t = 1) f'_{1t}, d_{2t}I(s_t = 2) f'_{2t}] \) and \( A = [\alpha_1, A_1, \alpha_2, A_2] \). The measurement equation in (9) of the state space representation remains valid with \( B_1 = [\alpha_1, A_1, \alpha_2, 0] \) and \( B_2 = [\alpha_1, 0, \alpha_2, A_2] \). Therefore, the tools developed in this paper can be applied to the sample counterpart of \( y_t \), namely to \( \hat{y}_t = x_t - \left(T^{-1} \sum_{t=1}^{T} x_t\right) \), which consistently estimates \( y_t \) as \( T \to \infty \). Corollary [1] holds accordingly with respect to \((\alpha_{1i}, \Lambda_{1i})'\) and \((\alpha_{2i}, \Lambda_{2i})'\) instead of with respect to \( \Lambda_{1i} \) and \( \Lambda_{2i} \) only, respectively, for \( i = 1, \ldots, N \).

7 Monte Carlo

We set \( N = \{100, 200\} \) and \( T = \{250, 500, 750, 1000\} \). At each time period \( t = 1, \ldots, T \), we simulate the \( N \times 1 \) vector of data \( x_t \) according to (11) and (12). This requires to simulate the latent state \( \xi_t \), the loadings \( \Lambda_1 \) and \( \Lambda_2 \), the factors \( f_{1t} \) and \( f_{2t} \), and the idiosyncratic components \( e_t \).

We simulate the latent state \( \xi_t \) according to (11), where \( P \) 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 \( P(s_t = 1) = E[\xi_{1t}] = \frac{1-p_{22}}{2-p_{11}-p_{22}} = 0.75 \) and \( P(s_t = 2) = E[\xi_{2t}] = \frac{1-p_{11}}{2-p_{11}-p_{22}} = 0.25 \). Then, we generate the innovations \( v_t \) of the VAR in (11) as follows: at each given \( t \) we generate \( u_t \sim U[0, 1] \) and (i) if \( \xi_{1,t-1} = 1 \) and \( u_t \leq p_{11} \) then \( v_t = [1 0]' - P'\xi_{t-1} \); (ii) if \( \xi_{1,t-1} = 1 \) and \( u_t > p_{11} \) then \( v_t = [0 1]' - P'\xi_{t-1} \); (iii) if \( \xi_{1,t-1} = 0 \) and \( u_t \leq p_{21} \) then \( v_t = [1 0]' - P'\xi_{t-1} \); (iv) if \( \xi_{1,t-1} = 0 \) and \( u_t > p_{21} \) then \( v_t = [0 1]' - P'\xi_{t-1} \).

We set the number of factors in each state to \( r_j = r = \{1, 2\} \), \( j = 1, 2 \). The common component is generated according to model (11). Let \( \chi_{it} = \Lambda_{1i}f_{1t}I(s_t = 1) + \Lambda_{2i}f_{2t}I(s_t = 1), i = 1, \ldots, N, t = 1, \ldots, T \). The \( r \) entries of \( \Lambda_{1i} \) and \( \Lambda_{2i} \) are generated from a \( \mathcal{N}(1, 1) \) distribution. The matrices \( \Lambda_1 \) and \( \Lambda_2 \) are then transformed in such a way that \( \Lambda_1' \Lambda_1 \) and \( \Lambda_2' \Lambda_2 \) are diagonal matrices. The factors are such that \( f_{jt} = f_t, j = 1, 2 \), and satisfy \( T^{-1} \sum_{t=1}^{T} f_t f_t' = I_r \), where each component of \( f_t \) is such that \( f_{kt} = \rho_f f_{k,t-1} + z_{kt}, k = 1, \ldots, r, \rho_f = \{0, 0.7\} \) and \( z_{kt} \sim \mathcal{N}(0, 1) \).

The idiosyncratic components are generated according to (12), where \( \Sigma_{je} = \Sigma_{je,a} + \Sigma_{je,b}, j = 1, 2, \) with \( \Sigma_{je,a} \) diagonal and \( \Sigma_{je,b} \) banded. Specifically, the entries of \( \Sigma_{je,a} \) are generated from a \( \mathcal{U}[0.25, 1.25] \) and those of \( \Sigma_{je,b} \) are generated from a \( \mathcal{U}[0.75, 1.75] \), while \( \Sigma_{je,b} \) is a Toeplitz matrix with \( \tau^k \) on the \( k \)th diagonal for \( k = 1, 2 \) and zero elsewhere, and, finally, \( \Sigma_{2e,b} \) 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 \( \nu_t \) is such that \( \nu_{1t} = \rho_1 \nu_{i,t-1} + \omega_{it}, i = 1, \ldots, N, t = 1, \ldots, T \), with \( \rho_1 = \{0, \rho\} \) and \( \rho \sim \mathcal{U}[0, 0.5] \). Finally, we set the average noise-to-signal ratio across all \( N \) simulated time series to be \( N^{-1} \sum_{i=1}^{N} \frac{\sum_{t=1}^{T} \nu_{it}^2}{\sum_{t=1}^{T} \chi_{it}^2} = 0.5 \). 

26
The closer this number is to one, the closer is the space spanned by the columns of \( \hat{B}_1 \) to the

\[ R^2_{B_r} = \frac{\text{tr}\left\{ \left( B_r^{\prime} \hat{B}_1 \right) \left( \hat{B}_1^{\prime} \hat{B}_1 \right)^{-1} \left( \hat{B}_1^{\prime} B_r^* \right) \right\}}{\text{tr}\left( \hat{B}_1^{\prime} B_r^* \right)}. \]

Table 1: Estimated probabilities - \( r = 1, \rho_f = 0, \tau = 0, \rho = 0. \)

| \( T \) | \( N \) | \( \hat{p}_{11} \) | \( \hat{p}_{22} \) | \( \hat{\xi}_{1,tT} \) | \( \hat{\xi}_{2,tT} \) | \( R^2_{B_r} \) | \( \text{MSE}(\chi) \) | \( \text{avg. iter} \) |
|---|---|---|---|---|---|---|---|---|
| 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 | 0.89 | 0.64 | 0.76 | 0.24 | 0.98 | 0.01 | 12.55 |
|     |     | (0.01) | (0.04) | (0.03) | (0.03) |     |     |     |
| 750 | 100 | 0.89 | 0.64 | 0.76 | 0.25 | 0.98 | 0.01 | 12.71 |
|     |     | (0.01) | (0.03) | (0.03) | (0.03) |     |     |     |
| 1000 | 100 | 0.89 | 0.64 | 0.76 | 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 | 0.64 | 0.76 | 0.25 | 0.97 | 0.01 | 21.23 |
|     |     | (0.02) | (0.04) | (0.03) | (0.03) |     |     |     |
| 750 | 200 | 0.89 | 0.64 | 0.76 | 0.25 | 0.97 | 0.02 | 37.37 |
|     |     | (0.02) | (0.04) | (0.03) | (0.03) |     |     |     |
| 1000 | 200 | 0.89 | 0.64 | 0.76 | 0.25 | 0.98 | 0.02 | 36.22 |
|     |     | (0.01) | (0.03) | (0.03) | (0.03) |     |     |     |

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 3.5. Since the states are identified only up to a permutation at each iteration of the algorithm we assign label 1 to the state with the highest estimated unconditional probability.\(^2\)

Results are collected in Tables 1-4 and are organised as follows: \( i \) \( r = 1, \rho_f = 0, \tau = 0, \rho = 0 \) in Table 1; \( ii \) \( r = 1, \rho_f = 0.7, \tau = 0.5, \rho = 0.5 \) in Table 2; \( iii \) \( r = 2, \rho_f = 0, \tau = 0, \rho = 0 \) in Table 3; \( iv \) \( r = 2, \rho_f = 0.7, \tau = 0.5, \rho = 0.5 \) in Table 4.

The first four columns of Tables 1-4 report the mean and, between brackets, the corresponding standard deviation over all replications of the estimated diagonal entries of the transition matrix \( \hat{p}_{jj}, j = 1, 2, \) of the unconditional probabilities \( P(s_t = j), \) estimated as \( \hat{\xi}_{j,tT} = T^{-1} \sum_{t=1}^{T} \xi_{j,T}, j = 1, 2. \)

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 \( \hat{B}_1 \) onto the columns of \( B^*_1 = B_1 \hat{I}_{\xi_1} + B_2 (I_{2r} - \hat{I}_{\xi_1}), \) thus correcting for the bias described in Theorem 1. Namely, we compute

\[ R^2_{B_r} = \frac{\text{tr}\left\{ \left( B_r^{\prime} \hat{B}_1 \right) \left( \hat{B}_1^{\prime} \hat{B}_1 \right)^{-1} \left( \hat{B}_1^{\prime} B_r^* \right) \right\}}{\text{tr}\left( \hat{B}_1^{\prime} B_r^* \right)}. \]

\(^2\)Note that the initialization such that \( \omega_1 = \omega_2 = 0.5 \) is not empirically feasible, as it leads to no convergence of the EM algorithm. We conjecture that this has to do with the relabelling issue discussed in Section 2.2 since for \( \omega_1 = \omega_2 = 0.5 \) both states are equally likely.
Table 2: Estimated probabilities - \( r = 1, \rho_f = 0.7, \tau = 0.5, \rho = 0.5 \).

| \( T \) | \( N \) | \( \hat{p}_{11} \) | \( \hat{p}_{22} \) | \( \hat{\xi}_{1,0} \) | \( \hat{\xi}_{2,0} \) | \( R^2_{\text{FW}} \) | 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)         |                |              |            |

Table 3: Estimated probabilities - \( r = 2, \rho_f = 0, \tau = 0, \rho = 0 \).

| \( T \) | \( N \) | \( \hat{p}_{11} \) | \( \hat{p}_{22} \) | \( \hat{\xi}_{1,0} \) | \( \hat{\xi}_{2,0} \) | \( R^2_{\text{FW}} \) | 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)         |                |              |            |

space spanned by the columns of \( B_1^* \) (see Doz et al., 2012).

In the sixth column of Tables 1, 2, 3, 4, we report the MSE of the estimated common components defined as

\[
\text{MSE}(\chi) = \frac{\sum_{t=1}^{N} \sum_{f=1}^{T} (\hat{\chi}_{it} - \chi_{it})^2}{\sum_{t=1}^{N} \sum_{f=1}^{T} \chi_{it}^2},
\]

where \( \hat{\chi}_{it} = \left( \hat{b}_{1t}, \hat{b}_{2t} \right)' \left( \hat{\xi}_t \otimes \hat{g}_t \right) \).

In the last column of Tables 1, 2, 3, 4, we report the average number of iterations needed for the EM algorithm to converge.
Table 4: Estimated probabilities - $r = 2$, $\rho_f = 0.7$, $\tau = 0.5$, $\rho = 0.5$.

| $T$ | $N$ | $\hat{p}_{11}$ | $\hat{p}_{22}$ | $\tilde{\xi}_{t|T,1}$ | $\tilde{\xi}_{t|T,2}$ | $R^2_B$ | 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)          |         |            |           |

The results in Tables 1-4 confirm the empirical validity of the estimation procedure detailed in Section 3. In all four scenarios, as $N$ and $T$ increase the estimators $\hat{p}_{11}$, $\hat{p}_{22}$, $\tilde{\xi}_{t|T,1}$ and $\tilde{\xi}_{t|T,2}$ all converge to the true values of the corresponding parameters. In addition, $R^2_B$ and MSE($\chi$) are very to 1.00 and 0.00, respectively. Finally, note that the average number of iterations declines almost monotonically as $N$ and $T$ increase. Overall, our Monte Carlo findings provide evidence in support of the estimation algorithm proposed in Section 3.

8 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 8.1 describes the data and the empirical model specification, Section 8.2 discusses the estimated regime probabilities, and Sections 8.3 and 8.4 present the findings for estimated loadings and factors.
8.1 Data and model specification

The vector of observable dependent variables \( x_t \) 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.\(^3\) Consistently with the discussion in Section 6, the unconditional mean of \( x_t \) is equal to 0, which means that the returns have been demeaned along the time series dimension over the whole sample period. To obtain a balanced panel, the sample period runs from July 1969 through December 2021, a total of \( T = 630 \) time periods.

8.2 Regime probabilities

Using the selection criterion of Ahn and Horenstein (2013) as applied to the equivalent linear representation in (10), we find that the dimension of the vector \( g_t \) is equal to \( r_1 + r_2 = 2 \) common factors. As commonly assumed in the related literature (see Ang and Timmermann, 2012), we let the number of regimes be equal to two. Therefore, there is one common factor in each regime, so \( r_1 = r_2 = r = 1 \). Based on this result, we apply the algorithm detailed in Section 3. The EM algorithm converges in just 22 iterations (see Figure 1). The realisation of the estimator \( \hat{P} \) for the matrix of conditional probabilities \( P \) in (3) is equal to

\[
\hat{P} = \begin{pmatrix}
0.9194 & 0.0806 \\
0.3395 & 0.6605
\end{pmatrix}.
\]

The estimated unconditional probability for regime \( j \) is equal to the sample average \( \hat{\xi}_{j|T} = T^{-1} \sum_{t=1}^{T} \hat{\xi}_{j,t|T} \), for \( j = 1, 2 \). It follows that \( \hat{\xi}_{1|T} = 0.8044 \) and \( \hat{\xi}_{2|T} = 0.1956 \). Therefore, 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 \( \hat{\xi}_{1,t|T} \) and \( \hat{\xi}_{2,t|T} \), for \( t = 1, \ldots, T \). These series are negatively and positively correlated, respectively, with the NBER recession indicator, with correlation coefficients equal to -0.303 and 0.303, respectively.\(^5\) 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. This is consistent with the empirical frequency of the states, since expansions occur more often than recessions. Our model therefore captures regime changes in equity markets related to business cycle dynamics.

\(^3\)See [https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html](https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html).

\(^4\)By computing the unconditional probability using their analytical formulas in (7), we get \( \hat{\xi}_{1|T} = 0.8081 \) and \( \hat{\xi}_{2|T} = 0.1919 \).

\(^5\)The NBER recession indicator is publicly available at [https://fred.stlouisfed.org/series/USREC](https://fred.stlouisfed.org/series/USREC).
This figure plots the value of the maximized expected conditional log-likelihood computed using the estimated factors, i.e., $E_{Q^k} \left[ \log f \left( X \bigg| \hat{G}^{(k+1)}, \hat{\rho}^{(k+1)} \right) \right]$ (see also (15)), as a function of the EM iterations $k$.

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

8.3 Factors and loadings - Linear factor model

Following the sequential order dictated by our estimation procedure, we first consider estimated factors and loadings for the equivalent linear representation in (10), namely the $N \times 2$ loadings matrix $\hat{A}$ with columns $\hat{A}_{j,i}, j = 1, 2$, having elements $\hat{a}_{ji}, i = 1, \ldots, N$, and the 2-dimensional vector of factors $\hat{g}_t = [\hat{g}_{1,t} \, \hat{g}_{2,t}]'$, $t = 1, \ldots, T$, both obtained via PCA as explained in Section 3.3. The estimated loadings are displayed in Figure 3 and the estimated factors are displayed in Figure 4.

The estimated loadings $\hat{a}_{1i}$ associated to $\hat{g}_{1t}$ all have the same sign. On the other hand, the loadings $\hat{a}_{2i}$ associated to $\hat{g}_{2t}$ have both positive and negative signs. To aid understanding of the 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 $\hat{g}_{1t}$ 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...
Figure 3: Estimated loadings $\hat{A}$ in the linear factor model.

This figure plots the sequences of estimated loadings $\hat{a}_{1i}$ (panel (a)) and $\hat{a}_{2i}$ (panel (b)), for $i = 1, \ldots, N$, estimated from the linear factor model in (10).

Figure 4: Estimated factors $\hat{g}_t$ in the linear factor model.

This figure plots the series of the estimated factors $\hat{g}_{1t}$ (panel (a)) and $\hat{g}_{2t}$ (panel (b)), for $t = 1, \ldots, T$, estimated from the linear factor model in (10).

Table 5: Factor correlations in the linear factor model.

|       | $\hat{g}_{1t}$ | $\hat{g}_{2t}$ |
|-------|---------------|---------------|
| $RM_t$ | 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 factors $\hat{g}_{1t}$ and $\hat{g}_{2t}$ 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 ($MOM_t$).

With $HML_t$ and $RMW_t$. On the other hand, the second estimated latent factor $\hat{g}_{2t}$ does not exhibit any substantial correlation with any of the observable factors we consider. The first factor in the equivalent linear representation is then likely to be a market factor, while it is more difficult to give economic interpretation to the second factor.
Figure 5: Estimated loadings $\hat{\Lambda}_j$ and $\tilde{\Lambda}_j$, $j = 1, 2$.

This figure plots the sequences of estimated loadings $\hat{\lambda}_{1i}$ (panel (a)), $\hat{\lambda}_{2i}$ (panel (b)), $\tilde{\lambda}_{1i}$ (panel (c)), and $\tilde{\lambda}_{2i}$ (panel (d)), for $i = 1, \ldots, N$, estimated from the Markov switching factor model in (1) according to (35) and (36).

8.4 Factors and loadings - Markov switching factor model

We then turn to the estimated loadings and factors from the Markov switching factor model in (1). Since we have $r_1 = r_2$, the estimators for $\Lambda_j$, for $j = 1, 2$, are readily available from (35) or (36). Figure 5 shows the estimated loadings $\hat{\lambda}_{1i}$ and $\hat{\lambda}_{2i}$ (panels (a) and (b), respectively), and $\tilde{\lambda}_{1i}$ and $\tilde{\lambda}_{2i}$ (panels (c) and (d), respectively), for $i = 1, \ldots, N$. As expected from Corollary 1, the two sets of estimators are similar, although not identical, since they are consistent for different linear transformations of the true loadings. In particular, the correlation between $\hat{\lambda}_{1i}$ and $\tilde{\lambda}_{1i}$ is 0.85, whereas that between $\hat{\lambda}_{2i}$ and $\tilde{\lambda}_{2i}$ is 0.90.

Next, by projecting the data onto the estimated loadings weighted by the probability of being in a given state, we obtain the estimated scalar factors $\tilde{f}_{jt}$ and $\tilde{f}_{jt}$, for $j = 1, 2$ and $t = 1, \ldots, T$, as given in (37) and (38), respectively. These are shown in Figure 6. Finally, Table 6 displays the correlations between the estimated latent factors and the observable factors $RM_t$, $SMB_t$, $HML_t$, $RMW_t$, $CMA_t$, and $MOM_t$ described in Section 8.3. From the results in Table 6, we can see that $\hat{f}_{1t}$, which estimates $f_{1t}$ according to (37), is strongly correlated with $RM_t$, and reasonably correlated with $SMB_t$, $HML_t$, and $CMA_t$. As for $\hat{f}_{2t}$, the estimate for $f_{2t}$ from (37), it is correlated with $MOM_t$. A similar picture comes from $\tilde{f}_{1t}$ and $\tilde{f}_{2t}$, thus confirming the validity of our theoretical results.
**Figure 6:** Estimated factors $\hat{f}_{jt}$ and $\tilde{f}_{jt}$, $j = 1, 2$.

This figure plots the series of the estimated factors $\hat{f}_{1t}$ (panel (a)), $\hat{f}_{2t}$ (panel (b)), $\tilde{f}_{1t}$ (panel (c)), and $\tilde{f}_{2t}$ (panel (d)), for $t = 1, \ldots, T$, estimated from the Markov switching factor model in (1) according to (37) and (38).

**Table 6:** Factor correlations in the Markov switching factor model.

|       | $\hat{f}_{1t}$ | $\hat{f}_{2t}$ | $\tilde{f}_{1t}$ | $\tilde{f}_{2t}$ |
|-------|----------------|----------------|-------------------|-------------------|
| $RMT_t$ | 0.74          | 0.01           | 0.74              | -0.02             |
| $SMB_t$ | 0.32          | 0.07           | 0.32              | -0.03             |
| $HML_t$ | -0.17         | 0.06           | -0.17             | 0.06              |
| $RMW_t$ | -0.06         | 0.02           | -0.06             | 0.10              |
| $CMA_t$ | -0.22         | -0.06          | -0.13             | -0.01             |
| $MOM_t$ | -0.01         | -0.23          | -0.01             | -0.17             |

This table reports the correlation coefficients between the estimated factors $\hat{f}_{1t}$, $\hat{f}_{2t}$, $\tilde{f}_{1t}$, and $\tilde{f}_{2t}$ from the Markov switching factor model in (1) according to (37) and (38), 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 ($RMT_t$); size ($SMB_t$); value ($HML_t$); profitability ($RMW_t$); investment ($CMA_t$); momentum ($MOM_t$).

9 Concluding remarks

This paper develops estimation and inferential theory for high dimensional factor models with discrete regime changes in the loadings driven by a latent first order Markov process. 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. 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 distributions 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. Two are worth mentioning. Our
model allows for two regimes and the case of multiple states to capture richer dynamics is
worth exploring. The challenging task of making inference on the number of regimes is also
worth considering. These extensions are part of our research agenda.


t References
Ahn, S. C. and A. R. Horenstein (2013). Eigenvalue ratio test for the number of factors. *Economet-
rica* 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. (1997). Estimating multiple breaks one at a time. *Econometric Theory* 13(3), 315–352.

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. *Econo-
metrica* 70, 191–221.

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

Bai, J. and P. Perron (1998). Estimating and testing linear models with multiple structural changes.
*Econometrica* 66(1), 47–78.

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* 51, 1281–1304.

Chen, B., E. Y. Chen, and R. Chen (2023). Time-varying matrix factor model. Technical report, University of Rochester, New York University and Rutgers University.

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)* 39, 1–38.

Diebold, F. X. and G. D. Rudebusch (1996). Measuring business cycles: A modern perspective. *Review of Economics and Statistics* 78, 67–77.

Doz, C., L. Ferrara, and P.-A. Pionnier (2020). Business cycle dynamics after the great recession: An extended markov-switching dynamic factor model. Technical report, ffhalshs-02443364.

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 (2023). Quasi-maximum likelihood estimation of break point in high-dimensional factor models. *Journal of Econometrics* 223, 209–236.

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* 57, 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.

Kapetanios, G. (2010). A testing procedure for determining the number of factors in approximate factor models with large datasets. *Journal of Business & Economic Statistics* 28(3), 397–409.

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.

Leroux, B. G. (1992). Maximum-likelihood estimation for hidden markov models. *Stochastic Processes and their Applications* 40(1), 127–143.

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. (2023). Testing for regime changes in portfolios with a large number of assets: A robust approach to factor heteroskedasticity. *Journal of Financial Econometrics* 21(2), 316–367.

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.

McConnell, M. M. and G. Perez-Quiros (2000). Output fluctuations in the united states: What has changed since the early 1980’s? *The American Economic Review* 90(5), 1464–1476.

Meng, X.-L. and D. B. Rubin (1993). Maximum likelihood estimation via the ECM algorithm: A general framework. *Biometrika* 80, 267–278.

Meng, X.-L. and D. B. Rubin (1994). On the global and componentwise rates of convergence of the EM algorithm. *Linear Algebra and its Applications* 199, 413–425.

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.

Su, L. and X. Wang (2017). On time-varying factor models: Estimation and testing. Journal of Econometrics 198(1), 84–101.

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

Uurga, 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.
A Details of estimation

A.1 Baum-Lindgren-Hamilton-Kim filter

For simplicity of notation, in this appendix we will consider both the factors \( \{ g_t \}_{t=1}^T \) and the true values of the parameters \( q \) to be known. To simplify notation, let \( \varepsilon_1 = [1 0]' \) and \( \varepsilon_2 = [0 1]' \), so that
\[
P(s_t = j) \equiv P(\xi_t = \varepsilon_j), \quad j = 1, 2,
\]
and therefore, in the following, we can just use \( \xi_t \) as defined in \([4]\), without the need of referring also to \( s_t \). Then, for any \( v = 1, \ldots, T \), we use the notation
\[
\xi_{tv} = \mathbb{E}[\xi_t | X_v] = \left[ \begin{array}{c} P(\xi_t = \varepsilon_1 | X_v) \\ P(\xi_t = \varepsilon_2 | X_v) \end{array} \right]. \quad (A.1)
\]

Notice also that, since \( \{ \xi_t \}_{t=1}^T \) is independent of \( G_t \) for all \( u,v = 1, \ldots, T \), because we consider the factors as observed, we can always write \( \xi_{tv} = \mathbb{E}[\xi_t | X_v] = \mathbb{E}[\xi_t | X_v, G_v] \).

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)].

Then, the one-step-ahead predicted probabilities are obtained through the prior probability
\[
P(\xi_t = \varepsilon_i | X_{t-1}, G_{t-1}) = \sum_{j=1}^{2} P(\xi_t = \varepsilon_j | \xi_{t-1} = \varepsilon_j) P(\xi_{t-1} = \varepsilon_j | X_{t-1}, G_{t-1})
\]
\[
= \sum_{j=1}^{2} P(\xi_t = \varepsilon_j | \xi_{t-1} = \varepsilon_j) P(\xi_{t-1} = \varepsilon_j | X_{t-1}), \quad i = 1, 2. \quad (A.2)
\]
So that, because of (A.1), we have
\[
\xi_{tv} = \mathbb{P}' \xi_{tv-1}, \quad t = 1, \ldots, T. \quad (A.3)
\]
The update involves the posterior probability:
\[
P(\xi_t = \varepsilon_i | X_t) = P(\xi_t = \varepsilon_i | X_t, G_t) = P(\xi_t = \varepsilon_i | x_t, X_{t-1}, G_t)
\]
\[
= \frac{f(x_t, \xi_t = \varepsilon_i | X_{t-1}, G_t)}{f(x_t | X_{t-1}, G_t)}
\]
\[
= \frac{f(x_t | \xi_t = \varepsilon_i, X_{t-1}, G_t) P(\xi_t = \varepsilon_i | X_{t-1}, G_t)}{f(x_t | X_{t-1}, G_t)}, \quad i = 1, 2. \quad (A.4)
\]
Then, since \( x_t \) depends on \( X_{t-1} \) only through \( \xi_{t-1} \) and it depends on \( G_t \) only through \( g_t \)
\[
f(x_t | \xi_t = \varepsilon_i, X_{t-1}, G_t) = f(x_t | \xi_t = \varepsilon_i, g_t), \quad i = 1, 2. \quad (A.5)
\]
Let,

$$\eta_t = \begin{bmatrix} f (x_t | \xi_t = \varepsilon_1, g_t) \\ f (x_t | \xi_t = \varepsilon_2, g_t) \end{bmatrix}$$

$$= \frac{1}{(2\pi)^{N/2}} \left\{ \begin{array}{ll} |\text{diag}(\Sigma_{e1})|^{-1/2} \exp \left[ -\frac{1}{2} (x_t - B_1 g_t)' (\text{diag}(\Sigma_{e1}))^{-1} (x_t - B_1 g_t) \right] \\ |\text{diag}(\Sigma_{e2})|^{-1/2} \exp \left[ -\frac{1}{2} (x_t - B_2 g_t)' (\text{diag}(\Sigma_{e2}))^{-1} (x_t - B_2 g_t) \right] \end{array} \right\}. \quad (A.6)$$

Further, notice that, from (A.1) and (A.6), the denominator of (A.4) be written as:

$$f (x_t | X_{t-1}, G_t) = \sum_{j=1}^{2} f (x_t | \xi_t = \varepsilon_j, X_{t-1}, G_t) P (\xi_t = \varepsilon_j, | X_{t-1}, G_t)$$

$$= \sum_{j=1}^{2} f (x_t | \xi_t = \varepsilon_j, g_t) P (\xi_t = \varepsilon_j, | X_{t-1}) = \eta_t \hat{\xi}_t. \quad (A.7)$$

Taking into account (A.1), (A.2), (A.5), and (A.7), the filtered probabilities are obtained from (A.4) as

$$\hat{\xi}_t = \frac{\eta_t \otimes \hat{\xi}_{t-1}}{\eta_t \otimes \hat{\xi}_{t-1}} = \frac{\eta_t \otimes \hat{\xi}_{t-1}}{\eta_t \otimes \hat{\xi}_{t-1}}, \quad t = 1, \ldots, T, \quad (A.8)$$

where $\eta_t$ is computed as in (A.6). The filter can started by setting either $\hat{\xi}_{00} = \varepsilon_1$, or, equivalently, $\hat{\xi}_{00} = 0$. We then run the Kim smoother, see e.g., Krolzig (2013, Chapter 5.2) and Kim (1994). Notice that (recall that $X \equiv X_T$ and $G \equiv G_T$):

$$P (\xi_t = \varepsilon_i | X, G) = \sum_{j=1}^{2} P (\xi_t = \varepsilon_i | \xi_{t+1} = \varepsilon_j, X, G) P (\xi_{t+1} = \varepsilon_j | X, G)$$

$$= \sum_{j=1}^{2} \frac{P (\xi_t = \varepsilon_i | \xi_{t+1} = \varepsilon_j, X, G) f (\{x_{s}, g_s\}_{s=t+1}^T | \xi_t = \varepsilon_i, \xi_{t+1} = \varepsilon_j, X, G_t)}{f (\{x_{s}, g_s\}_{s=t+1}^T | \xi_{t+1} = \varepsilon_j, X, G)} P (\xi_{t+1} = \varepsilon_j | X, G)$$

$$= \sum_{j=1}^{2} P (\xi_t = \varepsilon_i | \xi_{t+1} = \varepsilon_j, X, G) P (\xi_{t+1} = \varepsilon_j | X, G)$$

$$= \sum_{j=1}^{2} \frac{P (\xi_t = \varepsilon_i | X, G_t) P (\xi_{t+1} = \varepsilon_j | X, G_t)}{P (\xi_{t+1} = \varepsilon_j | X, G_t)} P (\xi_{t+1} = \varepsilon_j | X, G), \quad i = 1, 2,$$

which by (A.1) implies that the sequence of smoothed probabilities is given by

$$\hat{\xi}_{t|T} = \left[ P (\xi_{t+1|T} \otimes \hat{\xi}_{t+1|T}) \right] \otimes \hat{\xi}_{t|t}, \quad t = 1, \ldots, T. \quad (A.9)$$

This backward recursion is initiated at $\hat{\xi}_{1|T}$ which is the last iteration of the filter in (A.8).

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

$$\begin{align*}
\xi_{t+1} &= \rho \odot (\xi_{t} \odot \bar{\xi}_{t-1} \odot \xi_{t-1})^T, \quad t = 1, \ldots, T. \quad (A.10)
\end{align*}$$

### A.2 M-step

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

$$\begin{align*}
\frac{\partial \log f (X | G; \varphi, \rho)}{\partial \varphi'} &= \frac{1}{f (X | G; \varphi, \rho)} \sum_{\{\xi_t\}_{t=1}^T} \frac{\partial f (X | G, \{\xi_t\}_{t=1}^T; \varphi)}{\partial \varphi'} \rho (\{\xi_t\}_{t=1}^T | G, \rho) \\
&= \frac{1}{f (X | G; \varphi, \rho)} \sum_{\{\xi_t\}_{t=1}^T} \frac{\partial \log f (X | G, \{\xi_t\}_{t=1}^T; \varphi)}{\partial \varphi'} f (X | G, \{\xi_t\}_{t=1}^T; \varphi) \rho (\{\xi_t\}_{t=1}^T | G, \rho) \\
&= C \sum_{\{\xi_t\}_{t=1}^T} \frac{\partial \log f (X | G, \{\xi_t\}_{t=1}^T; \varphi)}{\partial \varphi'} \rho (\{\xi_t\}_{t=1}^T | X, G; \varphi, \rho), \quad (A.11)
\end{align*}$$

where $C$ is a positive normalization constant. Therefore, from (13), (15), and (A.11), if we observed $G$, the first order conditions would be:

$$\begin{align*}
0 &= \frac{\partial \mathbb{E}_{\tilde{q}^{(k)}} \left[ \log f (X | G; \varphi, \rho) | X \right]}{\partial \varphi'} \bigg|_{\varphi = \tilde{\varphi}^{(k+1)}} \\
&= \sum_{t=1}^T \sum_{j=1}^2 \frac{\partial \mathbb{E}_{\tilde{q}^{(k)}} \left[ \log f (X_t | g_t, \xi_t = \rho_j | X) \right]}{\partial \varphi'} \bigg|_{\varphi = \tilde{\varphi}^{(k+1)}} \rho (\xi_t = \rho_j | X; \tilde{\varphi}^{(k)}, \tilde{\rho}^{(k)}) \\
&= \sum_{t=1}^T \sum_{j=1}^2 \frac{\partial \mathbb{E}_{\tilde{q}^{(k)}} \left[ \log f (X_t | g_t, \xi_t = \rho_j | X) \right]}{\partial \varphi'} \bigg|_{\varphi = \tilde{\varphi}^{(k+1)}} \xi_{j,t}^{(k)}, \quad (A.12)
\end{align*}$$

where $\xi_{j,t}^{(k)} = \mathbb{E}_{\tilde{q}^{(k)}} [\xi_t | X] = P (\xi_t = \rho_j | X; \tilde{\varphi}^{(k)}, \tilde{\rho}^{(k)})$ is the $j$th component of $\xi_{t|T}^{(k)}$.

Then, by substituting (13) into (A.12), and by replacing true factors with estimated ones, we get

$$\mathbf{B}_{j}^{(k+1)} = \left( \sum_{t=1}^T \xi_{j,t}^{(k)} x_t \bar{x}_t \right) \left( \sum_{t=1}^T \xi_{j,t}^{(k)} \bar{x}_t \bar{x}_t^T \right)^{-1}, \quad j = 1, 2, \quad (A.13)$$

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

---

*Specifically, we have:

$$P (\{\xi_t\}_{t=1}^T | X, G; \varphi, \rho) = \frac{f (X | G, \{\xi_t\}_{t=1}^T; \varphi) \rho (\{\xi_t\}_{t=1}^T | G, \rho)}{\sum_{\{\xi_t\}_{t=1}^T} f (X | G, \{\xi_t\}_{t=1}^T; \varphi) \rho (\{\xi_t\}_{t=1}^T | G, \rho)},$$

so $C = \frac{\sum_{\{\xi_t\}_{t=1}^T} f (X | G, \{\xi_t\}_{t=1}^T; \varphi) \rho (\{\xi_t\}_{t=1}^T | G, \rho)}{f (X | G; \varphi, \rho)}$. 

---
components, we set

$$\hat{\Sigma}_{e_j}^{(k+1)}_{ji} = \left( \frac{\sum_{t=1}^{T} \left( \varepsilon_{it} - \hat{B}_{ji}^{(k+1)'} \hat{\xi}_{it} \right)^2}{\sum_{t=1}^{T} \xi_{jt} | T} \right), \quad i = 1, \ldots, N, \quad j = 1, 2, \quad (A.14)$$

$$\hat{\Sigma}_{e_j}^{(k+1)}_{jk} = 0, \quad i, k = 1, \ldots, N, \quad i \neq k, \quad j = 1, 2,$$

where \( \hat{B}_{ji}^{(k+1)'} \) is the ith row of \( \hat{B}^{(k+1)}_j \).

Moving to estimation of \( \rho \), from (12), we have:

$$\frac{\partial \log f(X|G; \varphi, \rho)}{\partial \rho'} = \frac{1}{f(X|G; \varphi, \rho)} \sum_{\{\xi_t\}_{t=1}^{T}} f(X|G; \{\xi_t\}_{t=1}^T; \varphi) \frac{\partial \log P(\{\xi_t\}_{t=1}^{T}|G; \rho)}{\partial \rho'}$$

$$= \frac{1}{f(X|G; \varphi, \rho)} \sum_{\{\xi_t\}_{t=1}^{T}} \frac{\partial \log P(\{\xi_t\}_{t=1}^{T}|G; \rho)}{\partial \rho'} f(X|G; \{\xi_t\}_{t=1}^T; \varphi) P(\{\xi_t\}_{t=1}^{T}|G; \rho)$$

$$= C \sum_{\{\xi_t\}_{t=1}^{T}} \frac{\partial \log P(\{\xi_t\}_{t=1}^{T}|G; \rho)}{\partial \rho'} P(\{\xi_t\}_{t=1}^{T}|X, G; \varphi, \rho), \quad (A.15)$$

where \( C \) is the same positive normalization constant as in (A.11). And, because of (14) and (A.15), if we observed \( G \) the derivatives with respect to the generic \((i,j)\)th element of \( \rho \), i.e., \( p_{ij} \), \( i, j = 1, 2 \), would be (treating \( \xi_0 \) as known)

$$\frac{\partial \log f(X|G; \varphi, \rho)}{\partial p_{ij}} = \frac{T}{2} \sum_{t=1}^{T} \sum_{h=1}^{2} \sum_{l=1}^{2} \frac{\partial \log P(\xi_t = e_{ih}|\xi_{t-1} = e_{il}; \rho)}{\partial p_{ij}} P(\xi_t = e_{ih}|\xi_{t-1} = e_{il}; \rho) \] $$

$$= \frac{T}{2} \sum_{t=1}^{T} \sum_{h=1}^{2} \sum_{l=1}^{2} \frac{1}{P(\xi_t = e_{ih}|\xi_{t-1} = e_{il}; \rho)} \frac{\partial P(\xi_t = e_{ih}|\xi_{t-1} = e_{il}; \rho)}{\partial p_{ij}} \] $$

$$= \frac{T}{2} \sum_{t=1}^{T} \sum_{h=1}^{2} \sum_{l=1}^{2} \frac{I(\xi_t = e_{ih}|\xi_{t-1} = e_{il})}{P(\xi_t = e_{ih}|\xi_{t-1} = e_{il}; \rho)} P(\xi_t = e_{ih}|\xi_{t-1} = e_{il}; \rho) \] $$

$$= \frac{T}{2} \sum_{t=1}^{T} \sum_{h=1}^{2} \sum_{l=1}^{2} \frac{P(\xi_t = e_{ih}|\xi_{t-1} = e_{il}; \rho)}{P(\xi_t = e_{ih}|\xi_{t-1} = e_{il}; \rho)} \] $$

$$= \frac{T}{2} \sum_{t=1}^{T} P(\xi_t = e_{ij}|\xi_{t-1} = e_{ij}; \rho) - \frac{T}{2} \sum_{t=1}^{T} P(\xi_t = e_{ij}|\xi_{t-1} = e_{ij}; \rho), \quad (A.16)$$

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

$$0 = \left\{ \frac{\partial E_{q(k)}[\log f(X|G; \varphi, \rho)|X]}{\partial (\text{vec}(P^{(k)}))} - \kappa' (\ell_2 \otimes I_2) \right\}_{\text{vec}(P) = \text{vec}(P^{(k+1)})}, \quad (A.17)$$

where \( \kappa \) is the 2-dimensional vector of Lagrange multipliers, thus it has positive entries. Then, from (A.16)

$$\frac{\partial E_{q(k)}[\log f(X|G; \varphi, \rho)|X]}{\partial p_{ij}} = \frac{T}{2} \sum_{t=1}^{T} P(\xi_t = e_{ij}|\xi_{t-1} = e_{ij}; \rho) \] $$

$$= \frac{T}{2} \sum_{t=1}^{T} P(\xi_t = e_{ij}|\xi_{t-1} = e_{ij}; \rho), \quad (A.18)$$

42
By collecting all 4 terms deriving from (A.18) into a vector, we have

\[
\frac{\partial E_{q^{(1)}}}{\partial \rho'} [\log f (X | G; \varphi, \rho) | X] = \sum_{t=1}^{T} \xi_{t,T-1}^{(k)} \otimes \rho',
\]

where \( \xi_{t,T-1}^{(k)} \) is defined in (A.10). Finally, from the first order conditions (A.17), we must have:

\[
0 = \left\{ \sum_{t=1}^{T} \xi_{t,T-1}^{(k)} \otimes \rho' - \kappa' (\iota_{2} \otimes I_{2}) \right\}_{\rho = \tilde{\rho}^{(k+1)}}.
\]

Let \( \kappa = (\kappa_{1}, \kappa_{2})' \), and let \( \tilde{\kappa} = (\iota_{2} \otimes \kappa) = (\kappa_{1}, \kappa_{1}, \kappa_{2})' \). Then, (A.20) gives

\[
\tilde{\rho}^{(k+1)} = \sum_{t=1}^{T} \xi_{t,T-1}^{(k)} \otimes \tilde{\kappa}.
\]

By applying the adding up condition to (A.21):

\[
\iota_{2} = (\iota_{2} \otimes I_{2}) \tilde{\rho}^{(k+1)} = (\iota_{2} \otimes I_{2}) \left( \sum_{t=1}^{T} \xi_{t,T-1}^{(k)} \otimes \tilde{\kappa} \right) = (\iota_{2} \otimes I_{2}) \sum_{t=1}^{T} \xi_{t,T-1}^{(k)} \otimes \kappa,
\]

which implies \( \kappa = \sum_{t=0}^{T-1} \xi_{t,T}^{(k)} \). Therefore, from (A.21),

\[
\tilde{\rho}^{(k+1)} = \left[ \sum_{t=1}^{T} \xi_{t,T-1}^{(k)} \right] \otimes \left[ \iota_{2} \otimes \sum_{t=0}^{T-1} \xi_{t,T}^{(k)} \right].
\]

**B Mathematical proofs**

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

\[
\sigma_{jil} = \mathbb{E} \left( \frac{1}{T} \sum_{t=1}^{T} I_{jt} e_{it} e_{lt} \right), \quad \chi_{jil} = \frac{1}{T} \sum_{t=1}^{T} I_{jt} e_{it} e_{lt} - \mathbb{E} \left( \frac{1}{T} \sum_{t=1}^{T} I_{jt} e_{it} e_{lt} \right),
\]

\[
\varphi_{jil} = \frac{1}{T} \sum_{t=1}^{T} I_{jt} X_{jt}' e_{it}, \quad \varphi_{jil} = \frac{1}{T} \sum_{t=1}^{T} I_{jt} X_{jt}' e_{it}.
\]
B.1 Lemmas

Lemma 1. Under Assumptions \([4] - [7]\) and given \(\hat{H}\) defined in \((27)\), we have
\[
\frac{1}{N} \sum_{i=1}^{N} \left\| \hat{a}_i - \hat{H}a_i \right\|^2 = O_p \left( \frac{1}{C_{NT}} \right).
\]

Lemma 2. Let Assumptions \([4] - [7]\) hold. Then:
(a) \(N^{-1} \sum_{l=1}^{N} \hat{a}_l \sigma_{jl} = O_p \left( \frac{1}{\sqrt{N C_{NT}}} \right)\);
(b) \(N^{-1} \sum_{l=1}^{N} \hat{a}_l \chi_{jl} = O_p \left( \frac{1}{\sqrt{N C_{NT}}} \right)\);
(c) \(N^{-1} \sum_{l=1}^{N} \hat{a}_l \phi_{jl} = O_p \left( \frac{1}{\sqrt{T C_{NT}}} \right)\);
(d) \(N^{-1} \sum_{l=1}^{N} \hat{a}_l \varphi_{jl} = O_p \left( \frac{1}{\sqrt{T}} \right)\).

Lemma 3. Under Assumptions \([4] - [7]\)
\[
N^{-1} \left( \hat{A} - A \hat{H} \right)' \hat{A} = O_p \left( \frac{1}{C_{NT}} \right).
\]

Lemma 4. Under Assumptions \([4] - [7]\)
\[
N^{-1} \left( \hat{A} - A \hat{H} \right)' e_t = O_p \left( \frac{1}{C_{NT}} \right).
\]

Lemma 5. Let Assumptions \([4] - [7]\) hold. Then:
(a) \(\hat{g}_t - \hat{H}^{-1} g_t = O_p \left( \frac{1}{\sqrt{T N}} \right) + O_p \left( \frac{1}{\sqrt{N T}} \right)\), for \(t = 1, \ldots, T\);
(b) \(\frac{1}{T} \sum_{t=1}^{T} (\hat{g}_t - \hat{H}^{-1} g_t) \hat{g}_t' = O_p \left( \frac{1}{T N \sqrt{T}} \right)\).

Lemma 6. Under Assumptions \([4] - [7]\) and given \(Q\) defined in \((28)\),
\[
p \lim_{N, T \to \infty} \frac{A' \hat{A}}{N} = Q.
\]

Lemma 7. Let Assumptions \([4] - [7]\) hold, and consider the matrix \(Q\) defined in \((28)\). Then, for \(j = 1, 2\), the \(r_j \times (r_1 + r_2)\) matrix \(Q_j\) satisfying \(Q = [Q_1', Q_2']'\) is such that
\[
Q_j = \Sigma_{f_j}^{-1/2} \Psi_j V_j^{1/2},
\]
where \(\Sigma_{f_j}\) is defined in \((22)\), and \(\Psi_j\) is the \(r_j \times (r_1 + r_2)\) matrix such that \(\Psi = [\Psi_1', \Psi_2']'\), with \(\Psi\) as in \((28)\).

Lemma 8. Let \(\hat{V}\) be the \((r_1 + r_2) \times (r_1 + r_2)\) diagonal matrix containing the first \(r_1 + r_2\) eigenvalues of \(\hat{\Sigma}_{x} = (NT)^{-1} \sum_{t=1}^{T} \hat{x}_t \hat{x}_t'\) in decreasing order. Define \(V\) as the \((r_1 + r_2) \times (r_1 + r_2)\) diagonal matrix of the first \(r_1 + r_2\) eigenvalues of \(\Sigma_{g}^{1/2} \Sigma_{A} \Sigma_{g}^{1/2}\) in decreasing order, where \(\Sigma_{g}\) and \(\Sigma_{A}\) are defined in \((23)\) and \((25)\), respectively. Then, under Assumptions \([4] - [7]\),
\[
\hat{V} \xrightarrow{p} V.
\]
Lemma 9. Let Assumptions 1 - 6 hold. Then, as \( N, T \to \infty \),
\[
\hat{I}_{\xi_j} \rightarrow I_{\xi_j} = H^{-1} \begin{bmatrix}
I (j = 1) I_{r_1} & 0 \\
0 & I (j = 2) I_{r_2}
\end{bmatrix} H, \quad j = 1, 2,
\]
where \( H \) is defined in (30).

Lemma 10. Let Assumptions 1 - 4 hold. Then, for any fixed \( 1 \leq p \leq \bar{p} \) with \( \bar{p} < \infty \), and for \( j = 1, 2 \), \( \| \hat{V}^{(p)}_{\xi_j} \| = O_p(1) \), where \( \hat{V}^{(p)}_{\xi_j} \) is the \( p \times p \) diagonal matrix containing the first \( p \) eigenvalues of \( \hat{\Sigma}_{\xi, x_j} \) defined in (30) in decreasing order.

Lemma 11. Let Assumption 3 hold. For \( j, k = 1, 2 \), and \( i, l = 1, \ldots, N \), all \( N \in \mathbb{N} \), consider
\[
\sigma^{\xi, jil} = \frac{1}{T} \sum_{t=1}^{T} E \left( \hat{I}_{jt} \hat{\xi}_{kt} e_{it} e_{lt} \right).
\]
Then
\[
\frac{1}{N} \sum_{i=1}^{N} \sum_{l=1}^{N} \sigma^{2}_{\xi, jil} = O_p(1).
\]

B.2 Proofs of Lemmas

Proof of Lemma 9. Consider \( \hat{\Sigma}_x = (NT)^{-1} \sum_{t=1}^{T} x_t x_t' \), and \( \hat{H} = (GG' / T) \left( A' \hat{A} / N \right) \hat{V}^{-1} \) as defined in (27). By the definition of eigenvectors and eigenvalues, \( \hat{\Sigma}_x \hat{A} = \hat{A} \hat{V} \), where \( \hat{V} \) is the \( \bar{r} \times \bar{r} \) diagonal matrix of the first \( \bar{r} = (r_1 + r_2) \) largest eigenvalues of \( \hat{\Sigma}_x \) in decreasing order, and \( \hat{A} \) is \( \sqrt{N} \) times the \( N \times \bar{r} \) matrix of eigenvectors of \( \hat{\Sigma}_x \) corresponding to its \( \bar{r} \) largest eigenvalues. Note that \( \| \hat{V} \| = O_p(1) \) and \( \| \hat{H} \| \leq \| GG' / T \| \| AA' / N \|^{1/2} \| \hat{A} \hat{A}' / N \|^{1/2} \| \hat{V}^{-1} \| = O_p(1) \) by Assumptions 1 and 2. We then have
\[
(\hat{A} - \hat{A} \hat{H}) \hat{V} = \hat{A} \hat{V} - \hat{A} \hat{V} \hat{V} = \hat{A} \hat{V} - A \frac{GG' A' \hat{A}}{T N},
\]
which implies
\[
\hat{V} \hat{A} - \frac{\hat{A} A GG' A'}{N T} = \hat{A} \hat{\Sigma}_x - \frac{\hat{A} A GG' A'}{N T} = \hat{A} \hat{V} \left[ \frac{T}{N} \sum_{t=1}^{T} x_t x_t' \right] - A GG' A'.
\]
Taking into account (B.1), after some algebra we have
\[
\hat{V} \left( \hat{a}_i - \hat{H} a_i \right) = \hat{A} \left[ \frac{1}{NT} \left( \sum_{t=1}^{T} x_t x_t' \right) - A GG' a \right] = \left[ \sum_{j=1}^{2} \left( \frac{1}{N} \sum_{l=1}^{N} \hat{a}_l \sigma_{jil} + \frac{1}{N} \sum_{l=1}^{N} \hat{a}_l \chi_{jil} + \frac{1}{N} \sum_{l=1}^{N} \hat{a}_l \varphi_{jil} + \frac{1}{N} \sum_{l=1}^{N} \hat{a}_l \hat{p}_{jil} \right) \right]. \tag{B.2}
\]
It follows that
\[
\frac{1}{N} \sum_{i=1}^{N} \left\| \hat{a}_i - \hat{H} a_i \right\|^2 \leq 8 \| \hat{V}^{-1} \|^{2} \left( \frac{1}{N} \sum_{i=1}^{N} \sigma_{jil} + \frac{1}{N} \sum_{i=1}^{N} \hat{\chi}_{jil} + \frac{1}{N} \sum_{i=1}^{N} \hat{\varphi}_{jil} + \frac{1}{N} \sum_{i=1}^{N} \hat{\varphi}_{jil} \right), \tag{B.3}
\]
where
\[
\hat{\sigma}_{ji} = \frac{1}{N^2} \left\| \sum_{l=1}^{N} \tilde{a}_l \sigma_{ji} \right\|^2, \quad \hat{\chi}_{ji} = \frac{1}{N^2} \left\| \sum_{l=1}^{N} \tilde{a}_l \chi_{ji} \right\|^2, \quad \hat{\varphi}_{ji} = \frac{1}{N^2} \left\| \sum_{l=1}^{N} \tilde{a}_l \varphi_{ji} \right\|^2, \quad \hat{\varphi}_{ji} = \frac{1}{N^2} \left\| \sum_{l=1}^{N} \tilde{a}_l \varphi_{ji} \right\|^2.
\]

Consider \( \hat{\sigma}_{ji} \) and note that
\[
\left\| \sum_{l=1}^{N} \tilde{a}_l \sigma_{ji} \right\|^2 \leq \left( \sum_{l=1}^{N} \left\| \tilde{a}_l \right\|^2 \right) \left( \sum_{l=1}^{N} \sigma_{ji}^2 \right)
\]
so that
\[
\frac{1}{N} \sum_{i=1}^{N} \hat{\sigma}_{ji} = \frac{1}{N} \sum_{i=1}^{N} \left( \frac{1}{N^2} \left\| \sum_{l=1}^{N} \tilde{a}_l \sigma_{ji} \right\|^2 \right) \leq \frac{1}{N} \left( \frac{1}{N} \sum_{i=1}^{N} \left\| \tilde{a}_i \right\|^2 \right) \frac{1}{N} \left( \sum_{i=1}^{N} \sum_{j=1}^{N} \sigma_{ji}^2 \right)
\]
given Assumption 3(b), \( N^{-1} \left( \sum_{i=1}^{N} \sum_{j=1}^{N} \sigma_{ji}^2 \right) \leq M \) by Lemma A.1(a) in [Massacchi, 2017], which implies that
\[
\frac{1}{N} \sum_{i=1}^{N} \hat{\sigma}_{ji} = O_p \left( \frac{1}{N} \right).
\]

Consider now,
\[
\sum_{i=1}^{N} \hat{\chi}_{ji} = \frac{1}{N^2} \sum_{i=1}^{N} \left\| \sum_{l=1}^{N} \tilde{a}_l \chi_{ji} \right\|^2
\]
\[
= \frac{1}{N^2} \sum_{i=1}^{N} \sum_{l=1}^{N} \sum_{q=1}^{N} \tilde{a}_i \tilde{a}_q \chi_{ji} \chi_{jq}
\]
\[
\leq \left[ \frac{1}{N^2} \sum_{l=1}^{N} \sum_{q=1}^{N} \left( \tilde{a}_l \tilde{a}_q \right)^2 \right]^{1/2} \left[ \frac{1}{N^2} \sum_{i=1}^{N} \sum_{q=1}^{N} \left( \sum_{j=1}^{N} \chi_{ji} \chi_{jq} \right)^2 \right]^{1/2}
\]
\[
\leq \left( \frac{1}{N} \sum_{i=1}^{N} \left\| \tilde{a}_i \right\|^2 \right) \left[ \frac{1}{N^2} \sum_{l=1}^{N} \sum_{q=1}^{N} \left( \sum_{j=1}^{N} \chi_{ji} \chi_{jq} \right)^2 \right]^{1/2}
\]
since
\[
E \left[ \left( \sum_{i=1}^{N} \chi_{ji} \chi_{jq} \right)^2 \right] = E \left( \sum_{i=1}^{N} \sum_{u=1}^{N} \chi_{ji} \chi_{jq} \chi_{ju} \chi_{ju} \right) \leq N^2 \max_{i,j} E \left( \left| \chi_{ji} \right|^4 \right)
\]
and
\[
E \left( \left| \chi_{ji} \right|^4 \right) = E \left[ \left| \frac{1}{T} \sum_{t=1}^{T} I_{jt} e_{it} e_{it} \right| \right]^4
\]
\[
= \frac{1}{T^2} E \left[ \left| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} I_{jt} e_{it} e_{it} \right|^4 \right]
\]
\[
\leq \frac{1}{T^2} M
\]
by Assumption 3(c), then
\[
\sum_{i=1}^{N} \hat{\chi}_{ji} \leq O_p \left( 1 \right) \sqrt{\frac{N^2}{T^2}} = O_p \left( \frac{N}{T} \right)
\]
and
\[
\frac{1}{N} \sum_{i=1}^{N} \tilde{\chi}_{ji} = O_p \left( \frac{1}{T} \right). \tag{B.5}
\]

Also
\[
\hat{\varphi}_{ji} = \frac{1}{N^2} \left\| \sum_{i=1}^{N} \hat{\alpha}_i \varphi_{jil} \right\|^2 \\
= \frac{1}{N^2} \left\| \sum_{i=1}^{N} \hat{\alpha}_i \left( \frac{1}{T} \sum_{t=1}^{T} \lambda_{jt} f_{jt} e_{it} \right) \right\|^2 \\
= \frac{1}{N^2} \left\| \sum_{i=1}^{N} \hat{\alpha}_i \lambda_{ji} \left( \frac{1}{T} \sum_{t=1}^{T} \lambda_{jt} f_{jt} e_{it} \right) \right\|^2 \\
\leq \left[ \frac{1}{N} \sum_{i=1}^{N} \left( \frac{1}{T^2} \left\| \sum_{t=1}^{T} \| f_{jt} e_{it} \| \right\|^2 \right) \right] \left\| \lambda_{ji} \right\|^2 \left( \frac{1}{N} \sum_{i=1}^{N} \| \hat{\alpha}_i \|^2 \right) \tag{B.6}
\]
and
\[
\frac{1}{N} \sum_{i=1}^{N} \tilde{\varphi}_{ji} = \left[ \frac{1}{N} \sum_{i=1}^{N} \left( \frac{1}{T^2} \left\| \sum_{t=1}^{T} \| f_{jt} e_{it} \| \right\|^2 \right) \right] \left( \frac{1}{N} \sum_{i=1}^{N} \| \lambda_{ji} \|^2 \right) \left( \frac{1}{N} \sum_{i=1}^{N} \| \hat{\alpha}_i \|^2 \right)
= O_p \left( \frac{1}{T} \right)
\]
by Assumptions 2 and 4. Finally,
\[
\hat{\varphi}_{j i} = \frac{1}{N^2} \left\| \sum_{i=1}^{N} \hat{\alpha}_i \varphi_{jil} \right\|^2 \\
= \frac{1}{N^2} \left\| \sum_{i=1}^{N} \hat{\alpha}_i \left( \frac{1}{T} \sum_{t=1}^{T} I_{jt} \lambda_{jt} f_{jt} e_{it} \right) \right\|^2 \\
= \frac{1}{N^2} \left\| \sum_{i=1}^{N} \hat{\alpha}_i \lambda_{ji} \left( \frac{1}{T} \sum_{t=1}^{T} I_{jt} f_{jt} e_{it} \right) \right\|^2 \\
\leq \frac{1}{N^2} \left\| \sum_{i=1}^{N} \hat{\alpha}_i \lambda_{ji} \right\|^2 \left\| \frac{1}{T} \sum_{t=1}^{T} I_{jt} f_{jt} e_{it} \right\|^2 \\
\leq \frac{1}{T} \left( \frac{1}{N} \sum_{i=1}^{N} \left\| \frac{1}{T^2} \left\| \sum_{t=1}^{T} \| f_{jt} e_{it} \| \right\|^2 \right) \right] \left\| \lambda_{ji} \right\|^2 \left( \frac{1}{N} \sum_{i=1}^{N} \| \hat{\alpha}_i \|^2 \right) \tag{B.7}
\]
and
\[
\frac{1}{N} \sum_{i=1}^{N} \tilde{\varphi}_{ji} \leq \frac{1}{T} \left( \frac{1}{N} \sum_{i=1}^{N} \left\| \frac{1}{T^2} \sum_{t=1}^{T} \| f_{jt} e_{it} \| \right\|^2 \right) \left( \frac{1}{N} \sum_{i=1}^{N} \| \lambda_{ji} \|^2 \right) \left( \frac{1}{N} \sum_{i=1}^{N} \| \hat{\alpha}_i \|^2 \right) = O_p \left( \frac{1}{T} \right)
\]
by Assumptions 2 and 4. By combining (B.3) - (B.7), and since \( \| \hat{V}^{-1} \parallel = O_p (1) \), then
\[
\frac{1}{N} \sum_{i=1}^{N} \| \hat{\alpha}_i - \hat{\alpha}^{*} \|^2 = O_p \left( \frac{1}{N} \right) + O_p \left( \frac{1}{T} \right)
\]
and the result stated in the lemma follows. \( \square \)
Proof of Lemma 2. Starting from (a), consider

\[
\frac{1}{N} \sum_{l=1}^{N} \tilde{a}_l \sigma_{jl} = \frac{1}{N} \sum_{l=1}^{N} (\tilde{a}_l - \tilde{H}'a_l + \tilde{H}'a_l) \sigma_{jl} = \frac{1}{N} \sum_{l=1}^{N} (\tilde{a}_l - \tilde{H}'a_l) \sigma_{jl} + \tilde{H}' \frac{1}{N} \sum_{l=1}^{N} a_l \sigma_{jl}.
\]

Note that

\[
\left\| \sum_{l=1}^{N} \tilde{a}_l \sigma_{jl} \right\| \leq \left( \max_l \|a_l\| \right) \left( \sum_{l=1}^{N} |\sigma_{jl}| \right) \leq \left[ \max_l (\|\lambda_1\| + \|\lambda_2\|) \right] \left( \sum_{l=1}^{N} |\sigma_{jl}| \right) \leq 2\lambda M
\]

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

\[
\frac{1}{N} \sum_{l=1}^{N} a_l \sigma_{jl} = O \left( \frac{1}{N} \right).
\]

Further

\[
\left\| \frac{1}{N} \sum_{l=1}^{N} (\tilde{a}_l - \tilde{H}'a_l) \sigma_{jl} \right\| \leq \left( \frac{1}{N} \sum_{l=1}^{N} \|\tilde{a}_l - \tilde{H}'a_l\|^2 \right)^{1/2} \frac{1}{\sqrt{N}} \left( \sum_{l=1}^{N} |\sigma_{jl}|^2 \right)^{1/2}
\]
\[
= \left[ O_p \left( \frac{1}{CT_{NT}} \right) \right]^{1/2} \frac{1}{\sqrt{N}} O_p \left( \frac{1}{N} \right)
\]
\[
= O_p \left( \frac{1}{\sqrt{NC_{NT}}} \right)
\]

by Lemma 4 and Assumption 3(b). It thus follows that

\[
\frac{1}{N} \sum_{l=1}^{N} \tilde{a}_l \sigma_{jl} = O_p \left( \frac{1}{\sqrt{NC_{NT}}} \right) + O_p \left( \frac{1}{N} \right) = O_p \left( \frac{1}{\sqrt{NC_{NT}}} \right).
\]

Moving on to (b), we have

\[
\frac{1}{N} \sum_{l=1}^{N} \tilde{a}_l \chi_{jl} = \frac{1}{N} \sum_{l=1}^{N} (\tilde{a}_l - \tilde{H}'a_l) \chi_{jl} + \tilde{H}' \frac{1}{N} \sum_{l=1}^{N} a_l \chi_{jl}.
\]

Note that

\[
\left\| \frac{1}{N} \sum_{l=1}^{N} (\tilde{a}_l - \tilde{H}'a_l) \chi_{jl} \right\| \leq \left( \frac{1}{N} \sum_{l=1}^{N} \|\tilde{a}_l - \tilde{H}'a_l\|^2 \right)^{1/2} \left( \frac{1}{N} \sum_{l=1}^{N} \chi_{jl}^2 \right)^{1/2},
\]

with

\[
\frac{1}{N} \sum_{l=1}^{N} \chi_{jl}^2 = \frac{1}{N} \sum_{l=1}^{N} \left\{ \frac{1}{T} \sum_{t=1}^{T} I_{jt} e_{it} e_{lt} - E \left( \frac{1}{T} \sum_{t=1}^{T} I_{jt} e_{it} e_{lt} \right) \right\}^2
\]
\[
= \frac{1}{NT} \sum_{l=1}^{N} \left\{ \frac{1}{T} \sum_{t=1}^{T} [I_{jt} e_{it} e_{lt} - E (I_{jt} e_{it} e_{lt})] \right\}
\]
\[
= O_p \left( \frac{1}{T} \right)
\]
so that
\[ \left\| \frac{1}{N} \sum_{l=1}^{N} \left( \hat{a}_l - \hat{H}'a_l \right) \chi_{jl} \right\| = O_p \left( \frac{1}{CN_T} \right) O_p \left( \frac{1}{\sqrt{T}} \right) = O_p \left( \frac{1}{\sqrt{TC_N}} \right). \]

Further
\[ \frac{1}{N} \sum_{l=1}^{N} a_l \chi_{jl} = \frac{1}{N} \sum_{l=1}^{N} a_l \left[ \frac{1}{T} \sum_{t=1}^{T} I_{jl} e_{lt} e_{lt} - E \left( \frac{1}{T} \sum_{t=1}^{T} I_{jl} e_{lt} e_{lt} \right) \right] = O_p \left( \frac{1}{\sqrt{NT}} \right) \]

by Assumption \( \text{6(a)} \). It follows that
\[ \frac{1}{N} \sum_{l=1}^{N} \hat{a}_l \chi_{jl} = O_p \left( \frac{1}{\sqrt{TC_N}} \right) + O_p \left( \frac{1}{\sqrt{NT}} \right) = O_p \left( \frac{1}{\sqrt{TC_N}} \right). \]

As for \( (c) \), consider
\[ \frac{1}{N} \sum_{l=1}^{N} \hat{a}_l \varphi_{jl} = \frac{1}{N} \sum_{l=1}^{N} \hat{a}_l \left( \frac{1}{T} \sum_{t=1}^{T} I_{jl} \lambda_j' f_{jt} e_{lt} \right) \]

\[ = \frac{1}{NT} \sum_{l=1}^{N} \sum_{t=1}^{T} I_{jl} \hat{a}_l e_{lt} f_{jt} \lambda_{ji} \]

\[ = \frac{1}{NT} \sum_{l=1}^{N} \sum_{t=1}^{T} I_{jl} \left( \hat{a}_l - \hat{H}'a_l + \hat{H}'a_l \right) e_{lt} f_{jt} \lambda_{ji} \]

\[ = \frac{1}{NT} \sum_{l=1}^{N} \sum_{t=1}^{T} I_{jl} \left( \hat{a}_l - \hat{H}'a_l \right) e_{lt} f_{jt} \lambda_{ji} + \hat{H}' \frac{1}{NT} \sum_{l=1}^{N} \sum_{t=1}^{T} I_{jl} a_l e_{lt} f_{jt} \lambda_{ji}. \]

We have
\[ \left\| \frac{1}{NT} \sum_{l=1}^{N} \sum_{t=1}^{T} I_{jl} \left( \hat{a}_l - \hat{H}'a_l \right) e_{lt} f_{jt} \lambda_{ji} \right\| \leq \frac{1}{\sqrt{T}} \left( \frac{1}{N} \sum_{l=1}^{N} \left\| \hat{a}_l - \hat{H}'a_l \right\|^2 \right)^{1/2} \]

\[ \times \left( \frac{1}{N} \sum_{l=1}^{N} \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{1}{1} I_{jl} e_{lt} e_{lt} \right\|^{1/2} \right)^{1/2} \left\| \lambda_{ji} \right\| \]

\[ = O_p \left( \frac{1}{\sqrt{T}} \right) O_p \left( \frac{1}{CN_T} \right) O_p (1) O (1) \]

\[ = O_p \left( \frac{1}{\sqrt{TC_N}} \right) \]

by Lemma \( \text{1} \), Assumption \( \text{6(c)} \) and Assumption \( \text{2} \). Also,
\[ \frac{1}{NT} \sum_{l=1}^{N} \sum_{t=1}^{T} I_{jl} a_l e_{lt} f_{jt} \lambda_{ji} = \frac{1}{\sqrt{NT}} \left[ \frac{1}{\sqrt{NT}} \sum_{l=1}^{N} \sum_{t=1}^{T} \frac{1}{1} I_{jl} \left( \lambda_{lt} \lambda_{2t} \right) e_{lt} f_{jt} \right] \lambda_{ji} = O_p \left( \frac{1}{\sqrt{NT}} \right) \]

by Assumption \( \text{6(b)} \) and Assumption \( \text{2} \). It follows that
\[ \frac{1}{N} \sum_{l=1}^{N} \hat{a}_l \varphi_{jl} = O_p \left( \frac{1}{\sqrt{TC_N}} \right) + O_p \left( \frac{1}{\sqrt{NT}} \right) = O_p \left( \frac{1}{\sqrt{TC_N}} \right). \]
Finally, for (d) we have
\[
\frac{1}{N} \sum_{l=1}^{N} \hat{a}_l \varphi_j l = \frac{1}{N} \sum_{l=1}^{N} \left( \hat{a}_l - \hat{H}' \hat{a}_l \right) \varphi_j l + \hat{H}' \frac{1}{N} \sum_{l=1}^{N} a_l \varphi_j l.
\]

Note that
\[
\frac{1}{N} \sum_{l=1}^{N} a_l \varphi_j l = \frac{1}{N} \sum_{l=1}^{N} \left( \frac{1}{T} \sum_{t=1}^{T} \lambda_{jl} \xi_j t f_j e_i \right) \left( \frac{1}{T} \sum_{t=1}^{T} \lambda_{jl} \xi_j t \right) \frac{1}{\sqrt{T}} \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \lambda_{jl} \xi_j t \right) = O_p \left( \frac{1}{\sqrt{T}} \right),
\]

by Assumption 2 and Assumption 3(c). Further,
\[
\left\| \frac{1}{N} \sum_{l=1}^{N} \left( \hat{a}_l - \hat{H}' \hat{a}_l \right) \varphi_j l \right\| \leq \left( \frac{1}{N} \sum_{l=1}^{N} \left\| \hat{a}_l - \hat{H}' \hat{a}_l \right\|^2 \right)^{1/2} \left( \frac{1}{N} \sum_{l=1}^{N} \varphi_j l^2 \right)^{1/2}
\]

with
\[
\frac{1}{N} \sum_{l=1}^{N} \varphi_j l^2 = \frac{1}{N} \sum_{l=1}^{N} \left( \frac{1}{T} \sum_{t=1}^{T} \lambda_{jl} \xi_j t f_j e_i \right)^2 \leq \frac{1}{T} \left( \frac{1}{N} \sum_{l=1}^{N} \left\| \lambda_{jl} \right\|^2 \right) \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \lambda_{jl} \xi_j t \right)^2 \leq O_p \left( \frac{1}{T} \right),
\]

by Assumption 2 and Assumption 3(c), so that taking into account Lemma 11 we have
\[
\frac{1}{N} \sum_{l=1}^{N} \left( \hat{a}_l - \hat{H}' \hat{a}_l \right) \varphi_j l = O_p \left( \frac{1}{C NT} \right) O_p \left( \frac{1}{\sqrt{T}} \right) = O_p \left( \frac{1}{\sqrt{T C NT}} \right).
\]

It follows that
\[
\frac{1}{N} \sum_{l=1}^{N} \hat{a}_l \varphi_j l = O_p \left( \frac{1}{\sqrt{T C NT}} \right) + O_p \left( \frac{1}{\sqrt{T}} \right) = O_p \left( \frac{1}{\sqrt{T}} \right),
\]

which completes the proof of the lemma.

\[\square\]

**Proof of Lemma 3** Consider
\[
N^{-1} \left( \hat{A} - \hat{A} \hat{H} \right)' \hat{A} = N^{-1} \left( \hat{A} - \hat{A} \hat{H} \right)' \hat{A} - N^{-1} \left( \hat{A} - \hat{A} \hat{H} \right)' \hat{A} \hat{H} + N^{-1} \left( \hat{A} - \hat{A} \hat{H} \right)' \hat{A} \hat{H}
\]
\[
= N^{-1} \left( \hat{A} - \hat{A} \hat{H} \right)' \hat{A} \hat{H} + N^{-1} \left( \hat{A} - \hat{A} \hat{H} \right)' \left( \hat{A} - \hat{A} \hat{H} \right).
\]

(B.8)
Using the identity in \([B.2]\), we have

\[
N^{-1} \left( \hat{A} - \hat{H}a_i \right)' A = \hat{V}^{-1} \sum_{i=1}^{N} \left( \hat{a}_i - \hat{H}a_i \right) a_i
\]

\[
= \hat{V}^{-1} \left\{ \sum_{j=1}^{N} \left[ \sum_{i=1}^{N} \left( \sum_{l=1}^{N} \hat{a}_{i\sigma jil} \right) a_i' \right] + \sum_{j=1}^{N} \left[ \sum_{i=1}^{N} \left( \sum_{l=1}^{N} \hat{a}_{i\sigma jil} \right) a_i' \right] + \sum_{j=1}^{N} \left[ \sum_{i=1}^{N} \left( \sum_{l=1}^{N} \hat{a}_{i\sigma jil} \right) a_i' \right] \right\}.
\]

(B.9)

Consider

\[
\frac{1}{N} \sum_{i=1}^{N} \left( \frac{1}{N} \sum_{l=1}^{N} \hat{a}_{i\sigma jil} \right) a_i' = \frac{1}{N} \sum_{i=1}^{N} \left[ \frac{1}{N} \sum_{l=1}^{N} \left( \hat{a}_i - \hat{H}a_i \right) \sigma_{jil} \right] a_i' + \hat{H} \frac{1}{N} \sum_{i=1}^{N} \frac{1}{N} \sum_{l=1}^{N} a_i a_i' \sigma_{jil}.
\]

We have

\[
\left\| \frac{1}{N} \sum_{i=1}^{N} \left( \frac{1}{N} \sum_{l=1}^{N} \hat{a}_{i\sigma jil} \right) a_i' \right\| \leq \frac{1}{\sqrt{N}} \left( \frac{1}{N} \sum_{i=1}^{N} \| \hat{a}_i - \hat{H}a_i \| \right)^{1/2} \left( \frac{1}{N} \sum_{i=1}^{N} \left| \sigma_{jil} \right|^2 \right)^{1/2} \left( \frac{1}{N} \sum_{i=1}^{N} \| a_i \|^2 \right)^{1/2}
\]

\[
= \frac{1}{\sqrt{N}} \sqrt{O_p \left( \frac{1}{CN_T} \right) O_p(1) O_p(1)}
\]

\[
= O_p \left( \frac{1}{\sqrt{CN_T}} \right).
\]

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

\[
\frac{1}{N} \sum_{i=1}^{N} \sum_{l=1}^{N} \sigma_{jil} \leq \frac{1}{N} \sum_{i=1}^{N} \sum_{l=1}^{N} \left| \sigma_{jil} \right|^{1/2} |\rho_{jil}| = \frac{1}{N} \sum_{i=1}^{N} \sum_{l=1}^{N} \left| \sigma_{jil} \right| \leq M^2.
\]

Further

\[
\left\| \frac{1}{N} \sum_{i=1}^{N} \frac{1}{N} \sum_{l=1}^{N} a_i a_i' \sigma_{jil} \right\| \leq \frac{1}{N} \left( \frac{1}{N} \sum_{i=1}^{N} \sum_{l=1}^{N} \| a_i \| \| a_i \| \right) \sigma_{jil} = O \left( \frac{1}{N} \right)
\]

by Assumptions 2 and 3 b). Therefore,

\[
\frac{1}{N} \sum_{i=1}^{N} \left( \frac{1}{N} \sum_{l=1}^{N} \hat{a}_{i\sigma jil} \right) a_i' = O_p \left( \frac{1}{\sqrt{N} C_{NT}} \right) + O \left( \frac{1}{N} \right) = O_p \left( \frac{1}{\sqrt{N} C_{NT}} \right).
\]

(B.10)

Consider now

\[
\frac{1}{N} \sum_{i=1}^{N} \left( \frac{1}{N} \sum_{l=1}^{N} \hat{a}_{i\sigma jil} \right) a_i' = \frac{1}{N} \sum_{i=1}^{N} \left[ \frac{1}{N} \sum_{l=1}^{N} \left( \hat{a}_i - \hat{H}a_i \right) \chi_{jil} \right] a_i' + \hat{H} \frac{1}{N} \sum_{i=1}^{N} \frac{1}{N} \sum_{l=1}^{N} a_i a_i' \chi_{jil}.
\]

We have

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

51
and consider

\[ \left\| \frac{1}{N} \sum_{l=1}^{N} \left( \tilde{\alpha}_l - \tilde{\mathbf{H}}' \alpha_l \right) \chi_{jl} \right\| \leq \left( \frac{1}{N} \sum_{l=1}^{N} \left\| \tilde{\alpha}_l - \tilde{\mathbf{H}}' \alpha_l \right\|^2 \right)^{1/2} \left( \frac{1}{N} \sum_{l=1}^{N} |\chi_{jl}|^2 \right)^{1/2} \]

with

\[ \left( \frac{1}{N} \sum_{l=1}^{N} |\chi_{jl}|^2 \right)^{1/2} = \left[ \frac{1}{N} \sum_{l=1}^{N} \left( \frac{1}{T} \sum_{t=1}^{T} l_{jt} e_{lt} \right)^2 \right]^{1/2} \]

\[ = \frac{1}{\sqrt{T}} \left[ \frac{1}{N} \sum_{l=1}^{N} \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} l_{jt} e_{lt} \right)^2 \right]^{1/2} \]

\[ = O_p \left( \frac{1}{\sqrt{T}} \right) \]

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

\[ \left\| \frac{1}{N} \sum_{l=1}^{N} \left( \tilde{\alpha}_l - \tilde{\mathbf{H}}' \alpha_l \right) \chi_{jl} \right\| = O_p \left( \frac{1}{CNT} \right) O_p \left( \frac{1}{\sqrt{T}} \right) = O_p \left( \frac{1}{\sqrt{T}C_{NT}} \right). \]

Further,

\[ \left\| \frac{1}{N} \sum_{i=1}^{N} \frac{1}{N} \sum_{l=1}^{N} a_i a_l' \chi_{jl} \right\| = \left\| \frac{1}{N} \sum_{i=1}^{N} \frac{1}{N} \sum_{l=1}^{N} a_i a_l' \right\| \left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} a_i \left[ \frac{1}{T} \sum_{t=1}^{T} l_{jt} e_{lt} - \mathbb{E} \left( \frac{1}{T} \sum_{t=1}^{T} l_{jt} e_{lt} \right) \right] \right\| \]

\[ \leq \frac{1}{\sqrt{NT}} \left( \frac{1}{N} \sum_{i=1}^{N} \left\| a_i \right\|^2 \right) \left\{ \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} a_i \left[ \frac{1}{T} \sum_{t=1}^{T} l_{jt} e_{lt} - \mathbb{E} \left( \frac{1}{T} \sum_{t=1}^{T} l_{jt} e_{lt} \right) \right] \right\}^{1/2} \]

\[ \leq O_p \left( \frac{1}{\sqrt{NT}} \right) \]

by Assumptions 2 and 3(a). Therefore,

\[ \frac{1}{N} \sum_{i=1}^{N} \left( \frac{1}{N} \sum_{l=1}^{N} \tilde{\alpha}_l \chi_{jl} \right) a_i' = O_p \left( \frac{1}{\sqrt{T}C_{NT}} \right) + O_p \left( \frac{1}{\sqrt{NT}} \right) = O_p \left( \frac{1}{\sqrt{T}C_{NT}} \right). \quad \text{(B.12)} \]

Consider now

\[ \frac{1}{N} \sum_{i=1}^{N} \left( \frac{1}{N} \sum_{l=1}^{N} \tilde{\alpha}_l \varphi_{jl} \right) a_i' = \frac{1}{N} \sum_{i=1}^{N} \left[ \frac{1}{N} \sum_{l=1}^{N} \left( \tilde{\alpha}_l - \tilde{\mathbf{H}}' \alpha_l \right) \varphi_{jl} \right] a_i' + \tilde{\mathbf{H}}' \frac{1}{N} \sum_{i=1}^{N} \left( \frac{1}{N} \sum_{l=1}^{N} a_i a_l' \varphi_{jl} \right). \]

We have

\[ \left\| \frac{1}{N} \sum_{i=1}^{N} \left[ \frac{1}{N} \sum_{l=1}^{N} \left( \tilde{\alpha}_l - \tilde{\mathbf{H}}' \alpha_l \right) \varphi_{jl} \right] a_i' \right\| \leq \left( \frac{1}{N} \sum_{i=1}^{N} \left\| \tilde{\alpha}_l - \tilde{\mathbf{H}}' \alpha_l \right\|^2 \right)^{1/2} \left( \frac{1}{N} \sum_{i=1}^{N} \left\| \frac{1}{N} \sum_{l=1}^{N} \varphi_{jl} a_i \right\|^2 \right)^{1/2} \]
and
\[
\left( \frac{1}{N} \sum_{i=1}^{N} \left\| \frac{1}{N} \sum_{l=1}^{N} \varphi_{jl} a_l \right\|^2 \right)^{1/2} = \left( \frac{1}{N} \sum_{i=1}^{N} \left\| \frac{1}{N} \sum_{l=1}^{N} \left( \frac{1}{T} \sum_{t=1}^{T} \mathbb{1}_{jt} \lambda'_{jl} f_{jt} e_{lt} \right) a_i \right\|^2 \right)^{1/2} \\
= \frac{1}{\sqrt{NT}} \left( \frac{1}{N} \sum_{i=1}^{N} \left\| \frac{1}{\sqrt{NT}} \sum_{l=1}^{N} \sum_{t=1}^{T} \mathbb{1}_{jt} \lambda'_{jl} f_{jt} e_{lt} a_i \right\|^2 \right)^{1/2} \\
\leq \frac{1}{\sqrt{NT}} \left( \frac{1}{N} \sum_{i=1}^{N} \left\| \frac{1}{\sqrt{NT}} \sum_{l=1}^{N} \sum_{t=1}^{T} \mathbb{1}_{jt} \lambda'_{jl} f_{jt} e_{lt} \right\|^2 \right)^{1/2} \\
= O_p \left( \frac{1}{\sqrt{NT}} \right)
\]
by Assumptions \( \mathbb{2} \) and \( \mathbb{b} \). Therefore,
\[
\left\| \frac{1}{N} \sum_{i=1}^{N} \left[ \frac{1}{N} \sum_{l=1}^{N} \left( \hat{a}_l - \hat{H} a_l \right) \varphi_{jl} \right] a_i' \right\| = O_p \left( \frac{1}{C_N} \right) O_p \left( \frac{1}{\sqrt{NT}} \right) = O_p \left( \frac{1}{\sqrt{NC_N}} \right)
\]
by Lemma \( \mathbb{l} \) Further
\[
\left\| \frac{1}{N} \sum_{i=1}^{N} \left( \frac{1}{N} \sum_{l=1}^{N} a_l a_i' \varphi_{jl} \right) \right\| = \left\| \frac{1}{N^2} \sum_{i=1}^{N} \sum_{l=1}^{N} a_l a_i' \left( \frac{1}{T} \sum_{t=1}^{T} \mathbb{1}_{jt} \lambda'_{jl} f_{jt} e_{lt} \right) \right\| \\
= \left\| \frac{1}{N^2} \sum_{i=1}^{N} \sum_{l=1}^{N} \left( \lambda_{ll} \right) \left( \frac{1}{T} \sum_{t=1}^{T} \mathbb{1}_{jt} \lambda'_{jl} f_{jt} e_{lt} \right) \right\| \left( \lambda_{ll} \right) \left( \lambda_{ll} \right) \\
\leq \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{l=1}^{N} \left( \lambda_{ll} \right) \left( \lambda_{ll} \right) \left( \frac{1}{T} \sum_{t=1}^{T} \mathbb{1}_{jt} \lambda'_{jl} f_{jt} e_{lt} \right) \left\| a_i \right\| \left\| \left( \lambda_{ll} \right) \left( \lambda_{ll} \right) \right\| \\
= O_p \left( \frac{1}{\sqrt{NT}} \right)
\]
by Assumptions \( \mathbb{2} \) and \( \mathbb{b} \). Therefore
\[
\frac{1}{N} \sum_{i=1}^{N} \left( \frac{1}{N} \sum_{l=1}^{N} \hat{a}_l \varphi_{jl} \right) a_i' = O_p \left( \frac{1}{\sqrt{NC_N}} \right) + O_p \left( \frac{1}{\sqrt{NT}} \right) = O_p \left( \frac{1}{\sqrt{NT}} \right). \tag{B.13}
\]
Finally,
\[
\frac{1}{N} \sum_{i=1}^{N} \left( \frac{1}{N} \sum_{l=1}^{N} \hat{a}_l \varphi_{jl} \right) a_i' = \frac{1}{N} \sum_{i=1}^{N} \left[ \frac{1}{N} \sum_{l=1}^{N} \left( \hat{a}_l - \hat{H} a_l \right) \varphi_{jl} \right] a_i' + \hat{H} \frac{1}{N} \sum_{i=1}^{N} \left( \frac{1}{N} \sum_{l=1}^{N} a_l a_i' \varphi_{jl} \right).
\]
We have
\[
\left\| \frac{1}{N} \sum_{i=1}^{N} \left[ \frac{1}{N} \sum_{l=1}^{N} \left( \hat{a}_l - \hat{H} a_l \right) \varphi_{jl} \right] a_i' \right\| \leq \left( \frac{1}{N} \sum_{i=1}^{N} \left\| \hat{a}_l - \hat{H} a_l \right\|^2 \right)^{1/2} \left( \frac{1}{N} \sum_{i=1}^{N} \left\| \frac{1}{N} \sum_{l=1}^{N} \varphi_{jl} a_i \right\|^2 \right)^{1/2}
\]
53
with
\[
\left( \frac{1}{N} \sum_{i=1}^{N} \left( \frac{1}{N} \sum_{l=1}^{N} \varphi_{jl i} a_l \right)^2 \right)^{1/2} = \left( \frac{1}{N} \sum_{i=1}^{N} \left( \frac{1}{N} \sum_{l=1}^{N} \left( \frac{1}{T} \sum_{t=1}^{T} \lambda_{jt} f_{jt} e_{it} \right) a_l \right)^2 \right)^{1/2}
\]
\[
= \frac{1}{\sqrt{NT}} \left( \frac{1}{N} \sum_{i=1}^{N} \left( \frac{1}{\sqrt{NT}} \sum_{l=1}^{N} \sum_{t=1}^{T} \lambda_{jt} f_{jt} e_{it} a_l \right)^2 \right)^{1/2}
\]
\[
\leq \frac{1}{\sqrt{NT}} \left( \frac{1}{N} \sum_{i=1}^{N} \sum_{l=1}^{T} \sum_{t=1}^{T} \lambda_{jt} f_{jt} e_{it} \right)^{1/2} \lesssim O_p \left( \frac{1}{\sqrt{NT}} \right)
\]
by Assumptions 2 and 6(b). Further
\[
\left\| \frac{1}{N} \sum_{i=1}^{N} \left( \frac{1}{N} \sum_{l=1}^{N} a_l a_l' \varphi_{jl i} \right) \right\| = \left\| \frac{1}{N^2} \sum_{i=1}^{N} \sum_{l=1}^{N} a_l a_l' \left( \frac{1}{T} \sum_{t=1}^{T} \lambda_{jt} f_{jt} e_{it} \right) \right\|
\]
\[
= \left\| \frac{1}{N^2} \sum_{i=1}^{N} \left( \lambda_{1t} \lambda_{2t} \right) \left( \frac{1}{T} \sum_{t=1}^{T} \lambda_{jt} f_{jt} e_{it} \right) \right\|
\]
\[
\leq \frac{1}{\sqrt{NT}} \sum_{l=1}^{N} \left( \lambda_{1l} \lambda_{2l} \right)^{1/2} \lesssim O_p \left( \frac{1}{\sqrt{NT}} \right)
\]
by Assumptions 2 and 6(b). Therefore,
\[
\frac{1}{N} \sum_{i=1}^{N} \left( \frac{1}{N} \sum_{l=1}^{N} \hat{a}_l \varphi_{jl i} \right) a_l' = O_p \left( \frac{1}{\sqrt{NC_{NT}}} \right) + O_p \left( \frac{1}{\sqrt{NT}} \right) = O_p \left( \frac{1}{\sqrt{NT}} \right).
\]
(B.14)
Combining equations (B.9) through (B.14), we obtain
\[
N^{-1} \left( \tilde{A} - \tilde{A} \tilde{H} \right)' A = O_p \left( \frac{1}{\sqrt{NC_{NT}}} \right) + O_p \left( \frac{1}{\sqrt{TC_{NT}}} \right) + O_p \left( \frac{1}{\sqrt{NT}} \right) = O_p \left( \frac{1}{C_{NT}^2} \right).
\]
(B.15)
From (B.8), (B.15) and Lemma 1, we obtain
\[
N^{-1} \left( \tilde{A} - \tilde{A} \tilde{H} \right)' \tilde{A} = O_p \left( \frac{1}{C_{NT}} \right) + O_p \left( \frac{1}{C_{NT}^2} \right) = O_p \left( \frac{1}{C_{NT}^2} \right),
\]
which completes the proof of the lemma.
Proof of Lemma 4. Given the identity in (B.2), we can write

\[
N^{-1} \left( \mathbf{A} - \mathbf{A} \hat{\mathbf{H}} \right)^\prime \mathbf{e}_t = \tilde{\mathbf{V}}^{-1} \sum_{i=1}^{N} \left( \mathbf{a}_i - \hat{\mathbf{H}} \mathbf{a}_i \right) \mathbf{e}_t \\
= \tilde{\mathbf{V}}^{-1} \left\{ \sum_{j=1}^{2} \left[ \frac{1}{N} \sum_{i=1}^{N} \left( \frac{1}{N} \sum_{l=1}^{N} \hat{\mathbf{a}}_l \sigma_{jil} \right) e_{it} \right] + \sum_{j=1}^{2} \left[ \frac{1}{N} \sum_{i=1}^{N} \left( \frac{1}{N} \sum_{l=1}^{N} \hat{\mathbf{a}}_l \chi_{jil} \right) e_{it} \right] \right\}.
\]

(B.16)

Consider

\[
\frac{1}{N} \sum_{i=1}^{N} \left( \frac{1}{N} \sum_{l=1}^{N} \hat{\mathbf{a}}_l \sigma_{jil} \right) e_{it} = \frac{1}{N} \sum_{i=1}^{N} \left[ \frac{1}{N} \sum_{l=1}^{N} \left( \hat{\mathbf{a}}_l - \hat{\mathbf{H}} \mathbf{a}_l \right) \sigma_{jil} \right] e_{it} + \hat{\mathbf{H}} \frac{1}{N} \sum_{i=1}^{N} \left( \frac{1}{N} \sum_{l=1}^{N} \mathbf{a}_l \sigma_{jil} e_{it} \right),
\]

where

\[
\left\| \frac{1}{N} \sum_{i=1}^{N} \left( \frac{1}{N} \sum_{l=1}^{N} \left( \hat{\mathbf{a}}_l - \hat{\mathbf{H}} \mathbf{a}_l \right) \sigma_{jil} \right) e_{it} \right\| \leq \frac{1}{\sqrt{N}} \left( \frac{1}{N} \sum_{i=1}^{N} \left\| \hat{\mathbf{a}}_l - \hat{\mathbf{H}} \mathbf{a}_l \right\|^2 \right)^{1/2} \left( \frac{1}{N} \sum_{i=1}^{N} \sum_{l=1}^{N} |\sigma_{jil}|^2 \right)^{1/2}
\times \left( \frac{1}{N} \sum_{i=1}^{N} |e_{it}|^2 \right)^{1/2}
= \frac{1}{\sqrt{N}} O_p \left( \frac{1}{C_{NT}} \right) O_p (1) O_p (1)
= O_p \left( \frac{1}{\sqrt{NC_{NT}}} \right)
\]

by Lemma 1, equation (B.10), and Assumption 3(a), and

\[
\left\| \frac{1}{N} \sum_{i=1}^{N} \left( \frac{1}{N} \sum_{l=1}^{N} \mathbf{a}_l \sigma_{jil} e_{it} \right) \right\| \leq \frac{1}{\sqrt{N}} \left( \frac{1}{N} \sum_{i=1}^{N} \left\| \mathbf{a}_l \right\|^2 \right)^{1/2} \left( \frac{1}{N} \sum_{i=1}^{N} \sum_{l=1}^{N} |\sigma_{jil}|^2 \right)^{1/2}
\times \left( \frac{1}{N} \sum_{i=1}^{N} |e_{it}|^2 \right)^{1/2}
= \frac{1}{\sqrt{N}} O_p \left( \frac{1}{C_{NT}} \right)
\]

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

\[
\left\| \frac{1}{N} \sum_{i=1}^{N} \left( \frac{1}{N} \sum_{l=1}^{N} \hat{\mathbf{a}}_l \sigma_{jil} \right) e_{it} \right\| \leq \left\| \frac{1}{N} \sum_{i=1}^{N} \left[ \frac{1}{N} \sum_{l=1}^{N} \left( \hat{\mathbf{a}}_l - \hat{\mathbf{H}} \mathbf{a}_l \right) \sigma_{jil} \right] e_{it} \right\| + \left\| \hat{\mathbf{H}} \frac{1}{N} \sum_{i=1}^{N} \left( \frac{1}{N} \sum_{l=1}^{N} \mathbf{a}_l \sigma_{jil} e_{it} \right) \right\|
= O_p \left( \frac{1}{\sqrt{NC_{NT}}} \right).
\]

(B.17)

Consider now

\[
\frac{1}{N} \sum_{i=1}^{N} \left( \frac{1}{N} \sum_{l=1}^{N} \hat{\mathbf{a}}_l \chi_{jil} \right) e_{it} = \frac{1}{N} \sum_{i=1}^{N} \left[ \frac{1}{N} \sum_{l=1}^{N} \left( \hat{\mathbf{a}}_l - \hat{\mathbf{H}} \mathbf{a}_l \right) \chi_{jil} \right] e_{it} + \hat{\mathbf{H}} \frac{1}{N} \sum_{i=1}^{N} \frac{1}{N} \sum_{l=1}^{N} \mathbf{a}_l e_{it} \chi_{jil}.
\]
We have
\[\left\| \frac{1}{N} \sum_{i=1}^{N} \left[ \frac{1}{N} \sum_{l=1}^{N} (\tilde{a}_l - \tilde{H}'a_l) \chi_{jl} \right] e_{it} \right\| \leq \frac{1}{N} \sum_{i=1}^{N} \left\| (\tilde{a}_i - \tilde{H}'a_i) \right\| \left( \frac{1}{N} \sum_{i=1}^{N} \left| \chi_{jl} e_{it} \right| \right) \]
\[\leq \left[ \frac{1}{N} \sum_{i=1}^{N} \left\| (\tilde{a}_i - \tilde{H}'a_i) \right\| \right]^{1/2} \left[ \frac{1}{N} \sum_{i=1}^{N} \left( \frac{1}{N} \sum_{i=1}^{N} \left| \chi_{jl} e_{it} \right| \right)^2 \right]^{1/2} \]
with
\[\frac{1}{N} \sum_{i=1}^{N} \left| \chi_{jl} e_{it} \right| = \frac{1}{N} \sum_{i=1}^{N} \left| \sum_{t=1}^{T} l_{jt} e_{it} - \mathbb{E} \left( \sum_{t=1}^{T} l_{jt} e_{it} \right) \right| e_{it} \]
\[= \frac{1}{\sqrt{T} N} \sum_{i=1}^{N} \left| \sum_{t=1}^{T} l_{jt} e_{it} - \mathbb{E} \left( \sum_{t=1}^{T} l_{jt} e_{it} \right) \right| e_{it} \]
\[= O_p \left( \frac{1}{\sqrt{T}} \right) \]
by Assumptions 3(a) and 3(c). Therefore, taking into account Lemma 1,
\[\left\| \frac{1}{N} \sum_{i=1}^{N} \left[ \frac{1}{N} \sum_{l=1}^{N} (\tilde{a}_l - \tilde{H}'a_l) \chi_{jl} \right] e_{it} \right\| = O_p \left( \frac{1}{\sqrt{N T}} \right) O_p \left( \frac{1}{\sqrt{T}} \right) = O_p \left( \frac{1}{\sqrt{N T C N T}} \right). \]
Further,
\[\left\| \frac{1}{N} \sum_{i=1}^{N} \sum_{l=1}^{N} \tilde{a}_l \chi_{jl} e_{it} \right\| = \left\| \frac{1}{N} \sum_{i=1}^{N} \sum_{l=1}^{N} a_l \left[ \frac{1}{T} \sum_{t=1}^{T} l_{jt} e_{it} - \mathbb{E} \left( \sum_{t=1}^{T} l_{jt} e_{it} \right) \right] e_{it} \right\| \]
\[= \frac{1}{\sqrt{N T N}} \sum_{i=1}^{N} \left\| \sum_{t=1}^{T} a_l \left[ l_{jt} e_{it} - \mathbb{E} (l_{jt} e_{it}) \right] \right\| e_{it} \]
\[\leq \frac{1}{\sqrt{N T}} \left( \frac{1}{N} \sum_{i=1}^{N} \left\| \sum_{t=1}^{T} \sum_{l=1}^{N} a_l \left[ l_{jt} e_{it} - \mathbb{E} (l_{jt} e_{it}) \right] \right\|^2 \right)^{1/2} \left( \frac{1}{N} \sum_{i=1}^{N} \left| e_{it} \right|^2 \right)^{1/2} \]
\[= O_p \left( \frac{1}{\sqrt{N T}} \right) \]
by Assumptions 3(a) and 3(c). Therefore,
\[\frac{1}{N} \sum_{i=1}^{N} \left( \frac{1}{N} \sum_{l=1}^{N} \tilde{a}_l \chi_{jl} \right) e_{it} = O_p \left( \frac{1}{\sqrt{T C N T}} \right) + O_p \left( \frac{1}{\sqrt{N T}} \right) = O_p \left( \frac{1}{\sqrt{T C N T}} \right). \]  
(B.18)
Consider now
\[\frac{1}{N} \sum_{i=1}^{N} \left( \frac{1}{N} \sum_{l=1}^{N} \tilde{a}_l \varphi_{jl} \right) e_{it} = \frac{1}{N} \sum_{i=1}^{N} \left[ \frac{1}{N} \sum_{l=1}^{N} (\tilde{a}_l - \tilde{H}'a_l) \varphi_{jl} \right] e_{it} + \frac{1}{N} \sum_{i=1}^{N} \frac{1}{N} \sum_{l=1}^{N} \tilde{H}'a_l \varphi_{jl} e_{it}. \]
We have
\[\left\| \frac{1}{N} \sum_{i=1}^{N} \left[ \frac{1}{N} \sum_{l=1}^{N} (\tilde{a}_l - \tilde{H}'a_l) \varphi_{jl} \right] e_{it} \right\| \leq \frac{1}{N} \sum_{i=1}^{N} \left\| \tilde{a}_l - \tilde{H}'a_l \right\| \left( \frac{1}{N} \sum_{i=1}^{N} \left| \varphi_{jl} e_{it} \right| \right) \]
\[\leq \left( \frac{1}{N} \sum_{i=1}^{N} \left\| \tilde{a}_l - \tilde{H}'a_l \right\|^2 \right)^{1/2} \left( \frac{1}{N} \sum_{i=1}^{N} \left( \frac{1}{N} \sum_{i=1}^{N} \left| \varphi_{jl} e_{it} \right| \right)^2 \right)^{1/2} \],
with
\[
\frac{1}{N} \sum_{i=1}^{N} |\varphi_{ji} e_{it}| = \frac{1}{N} \sum_{i=1}^{N} \left| \frac{1}{T} \sum_{t=1}^{T} b_{jt} \lambda_{ji} f_{jt} e_{it} \right| e_{it} \leq \frac{1}{\sqrt{T}} \frac{1}{N} \sum_{i=1}^{N} \|\lambda_{ji}\| \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} b_{jt} f_{jt} e_{it} \right\| e_{it} \leq \lambda \frac{1}{\sqrt{T}} |e_{it}| \left( \frac{1}{N} \sum_{i=1}^{N} \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} b_{jt} f_{jt} e_{it} \right\|^{2} \right)^{1/2} = O_p \left( \frac{1}{\sqrt{T}} \right)
\]

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

\[
\frac{1}{N} \sum_{i=1}^{N} \left[ \frac{1}{N} \sum_{l=1}^{N} \left( \hat{a}_{i} - \hat{H}' \hat{a}_{l} \right) \varphi_{jil} \right] e_{it} = O_p \left( \frac{1}{CN_{T}} \right) O_p \left( \frac{1}{\sqrt{T}} \right) = O_p \left( \frac{1}{\sqrt{T} C_{NT}} \right)
\]

Further,

\[
\left\| \frac{1}{N} \sum_{i=1}^{N} \frac{1}{N} \sum_{l=1}^{N} \hat{a}_{i} \varphi_{jil} e_{it} \right\| = \frac{1}{\sqrt{NT}} \left\| \frac{1}{N} \sum_{i=1}^{N} \hat{a}_{i} \left( \frac{1}{\sqrt{NT}} \sum_{l=1}^{N} \sum_{t=1}^{T} b_{jt} \lambda_{ji} f_{jt} e_{it} \right) e_{it} \right\| \leq \frac{1}{\sqrt{NT}} \left( \frac{1}{N} \sum_{i=1}^{N} \|a_{i}\| |e_{it}| \right) \left\| \frac{1}{\sqrt{NT}} \sum_{l=1}^{N} \sum_{t=1}^{T} b_{jt} \lambda_{ji} f_{jt} e_{it} \right\|^{1/2} = \frac{1}{\sqrt{NT}} \left( \frac{1}{N} \sum_{i=1}^{N} \|a_{i}\|^{2} \right)^{1/2} \left\| \frac{1}{\sqrt{NT}} \sum_{l=1}^{N} \sum_{t=1}^{T} b_{jt} \lambda_{ji} f_{jt} e_{it} \right\| = O_p \left( \frac{1}{\sqrt{NT}} \right)
\]

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

\[
\frac{1}{N} \sum_{i=1}^{N} \left( \frac{1}{N} \sum_{l=1}^{N} \hat{a}_{i} \varphi_{jil} \right) e_{it} = O_p \left( \frac{1}{\sqrt{T} C_{NT}} \right) + O_p \left( \frac{1}{\sqrt{NT}} \right) = O_p \left( \frac{1}{\sqrt{T} C_{NT}} \right).
\]

Finally,

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

Consider first
\[
\left\| \frac{1}{N} \sum_{i=1}^{N} \left[ \frac{1}{N} \sum_{i=1}^{N} \left( \hat{a}_{i} - \hat{H}' \hat{a}_{l} \right) \varphi_{jil} \right] e_{it} \right\| \leq \frac{1}{N} \sum_{i=1}^{N} \left[ \|\hat{a}_{i} - \hat{H}' \hat{a}_{l}\| \left( \frac{1}{N} \sum_{i=1}^{N} |\varphi_{jil} e_{it}| \right) \right] \leq \left( \frac{1}{N} \sum_{i=1}^{N} \|\hat{a}_{i} - \hat{H}' \hat{a}_{l}\|^{2} \right)^{1/2} \left[ \frac{1}{N} \sum_{i=1}^{N} \left( \frac{1}{N} \sum_{i=1}^{N} |\varphi_{jil} e_{it}| \right) \right]^{1/2},
\]

57
with
\[
\frac{1}{N} \sum_{i=1}^{N} |\varphi_{ji} e_{it}| = \frac{1}{N} \sum_{i=1}^{N} \left| \left( \frac{1}{T} \sum_{j=1}^{T} \lambda_{ij} f_{jt} e_{it} \right) e_{it} \right| \\
\leq \| \lambda_{ij} \| \frac{1}{\sqrt{T}} \frac{1}{N} \sum_{i=1}^{N} \left( \left( \frac{1}{\sqrt{T}} \sum_{j=1}^{T} f_{jt} e_{it} \right) |e_{it}| \right) \\
\leq \frac{\lambda}{\sqrt{T}} \frac{1}{N} \left( \frac{1}{\sqrt{T}} \sum_{i=1}^{N} \left( \sum_{j=1}^{T} f_{jt} e_{it} \right)^2 \right)^{1/2} \left( \frac{1}{N} \sum_{i=1}^{N} |e_{it}|^2 \right)^{1/2} \\
= O_p \left( \frac{1}{\sqrt{T}} \right)
\]
by Assumptions 2(a), 3(a), and 4 so that
\[
\left\| \frac{1}{N} \sum_{i=1}^{N} \left( \frac{1}{N} \sum_{j=1}^{N} (\tilde{a}_j - \tilde{A}^T a) \varphi_{ji} \right) e_{it} \right\| = O_p \left( \frac{1}{\sqrt{TC_{NT}}} \right).
\]
Also,
\[
\left\| \frac{1}{N} \sum_{i=1}^{N} \frac{1}{N} \sum_{j=1}^{N} a_i \varphi_{ji} e_{it} \right\| = \left\| \frac{1}{N} \sum_{i=1}^{N} \frac{1}{N} \sum_{j=1}^{N} a_i \left( \frac{1}{T} \sum_{j'=1}^{T} \lambda_{ij'} f_{j't} e_{i't} \right) e_{it} \right\| \\
= \left\| \left( \frac{1}{N} \sum_{i=1}^{N} \sum_{i=1}^{N} a_i \lambda_{ij} \right) \left( \frac{1}{N} \sum_{j=1}^{N} \sum_{i=1}^{T} f_{jt} e_{it} \right) e_{it} \right\| \\
\leq \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} \left( \sum_{i=1}^{T} f_{jt} e_{it} \right) \left( \sum_{i=1}^{N} \sum_{j=1}^{T} f_{jt} e_{it} \right) e_{it} \right\| \left( \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{T} f_{jt} e_{it} \right)^{1/2} \left( \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{T} f_{jt} e_{it} \right)^{1/2} \\
= O_p \left( \frac{1}{T} \right)
\]
by Assumption 3(c). Therefore,
\[
\frac{1}{N} \sum_{i=1}^{N} \left( \frac{1}{N} \sum_{i=1}^{N} \tilde{a}_i \varphi_{ji} \right) e_{it} = O_p \left( \frac{1}{\sqrt{TC_{NT}}} \right) + O_p \left( \frac{1}{T} \right) = O_p \left( \frac{1}{\sqrt{TC_{NT}}} \right).
\] (B.20)

By combining (B.16) through (B.20), we have
\[
N^{-1} \left( \tilde{A} - \tilde{A} \tilde{H} \right)^T e_t = O_p \left( \frac{1}{\sqrt{NC_{NT}}} \right) + O_p \left( \frac{1}{\sqrt{TC_{NT}}} \right) = O_p \left( \frac{1}{C_{NT}} \right),
\]
which completes the proof of the lemma. \(\square\)
Proof of Lemma: Starting from (a), and taking into account (10), consider
\[ \hat{g}_t = N^{-1} \hat{\Lambda} \hat{x}_t = N^{-1} \hat{\Lambda}' (A g_t + e_t) = N^{-1} \hat{\Lambda}' A g_t + N^{-1} \hat{\Lambda}' e_t \]

and note that
\[ A = A - \hat{\Lambda} \hat{H}^{-1} + \hat{\Lambda} \hat{H}^{-1}, \]

so that we have
\[ \hat{g}_t = N^{-1} \hat{\Lambda}' \left( A - \hat{\Lambda} \hat{H}^{-1} + \hat{\Lambda} \hat{H}^{-1} \right) g_t + N^{-1} \hat{\Lambda}' e_t \]
\[ = N^{-1} \hat{\Lambda}' \left( A - \hat{\Lambda} \hat{H}^{-1} + \hat{\Lambda} \hat{H}^{-1} \right) g_t + N^{-1} \hat{\Lambda}' e_t + N^{-1} \left( A \hat{H} \right)' e_t - N^{-1} \left( A \hat{H} \right)' e_t \]
\[ = N^{-1} \hat{\Lambda}' \left( A - \hat{\Lambda} \hat{H}^{-1} \right) g_t + N^{-1} \hat{\Lambda}' \hat{H}^{-1} g_t + N^{-1} \left( A - \hat{\Lambda} \hat{H} \right)' e_t + N^{-1} \left( A \hat{H} \right)' e_t, \]

which leads to
\[ \hat{g}_t - \hat{H}^{-1} g_t = N^{-1} \left( A \hat{H} \right)' e_t + N^{-1} \hat{\Lambda}' \left( A - \hat{\Lambda} \hat{H}^{-1} \right) g_t + N^{-1} \left( A - \hat{\Lambda} \hat{H} \right)' e_t. \quad (B.21) \]

The result in (a) follows by taking into account Assumption (d), Lemma 3 and Lemma 4. As for (b), adding and subtracting terms we have
\[ \frac{1}{T} \sum_{t=1}^{T} \left( \hat{g}_t - \hat{H}^{-1} g_t \right) \hat{g}_t = \frac{1}{T} \sum_{t=1}^{T} \left( \hat{g}_t - \hat{H}^{-1} g_t \right) \left( \hat{g}_t - \hat{H}^{-1} g_t \right)' + \frac{1}{T} \sum_{t=1}^{T} \left( \hat{g}_t - \hat{H}^{-1} g_t \right) g_t' \left( \hat{H}^{-1} \right)'. \quad (B.22) \]

Taking into account the results in (a), it follows that
\[ \frac{1}{T} \sum_{t=1}^{T} \left( \hat{g}_t - \hat{H}^{-1} g_t \right) \left( \hat{g}_t - \hat{H}^{-1} g_t \right)' = O_p \left( \frac{1}{N} \right) + O_p \left( \frac{1}{\sqrt{N C_{NT}}^2} \right) + O_p \left( \frac{1}{C_{NT}^2} \right). \quad (B.23) \]

From (B.21), we also have that
\[ \frac{1}{T} \sum_{t=1}^{T} \left( \hat{g}_t - \hat{H}^{-1} g_t \right) g_t' = \frac{1}{\sqrt{NT}} \hat{H} A' \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} e_t g_t' \right) + \frac{1}{N} \hat{\Lambda}' \left( A - \hat{\Lambda} \hat{H}^{-1} \right) g_t + \frac{1}{N} \left( A - \hat{\Lambda} \hat{H} \right)' e_t \]
\[ + \frac{1}{\sqrt{NT}} \left( \hat{A} - \hat{\Lambda} \hat{H} \right)' \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} e_t g_t' \right), \]
and taking into account Assumptions 2 and 3, and Lemma 3
\[
\left\| \frac{1}{T} \sum_{t=1}^{T} (\hat{g}_t - \hat{H}^{-1} g_t) g_t' \right\| = \frac{1}{\sqrt{NT}} \left\| \hat{H} \right\| \frac{1}{\sqrt{N}} \left\| \frac{1}{T} \sum_{t=1}^{T} \epsilon_t g_t' \right\| \left| \hat{A}' \left( A - \hat{A} \hat{H}^{-1} \right) \right| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} g_t g_t' \left( \hat{A} - \hat{A} \hat{H}^{-1} \right) \frac{1}{\sqrt{NT}} \left\| \frac{1}{T} \sum_{t=1}^{T} \epsilon_t g_t' \right\| + \frac{1}{\sqrt{NT}} \left\| \hat{A} - \hat{A} \hat{H} \right\| \frac{1}{\sqrt{N}} \left\| \frac{1}{T} \sum_{t=1}^{T} \epsilon_t g_t' \right\| = O_p \left( \frac{1}{\sqrt{NT}} \right) + O_p \left( \frac{1}{\sqrt{NC}} \right) + O_p \left( \frac{1}{\sqrt{NT}} \right) + O_p \left( \frac{1}{\sqrt{C}} \right).
\]

Combining (B.22) through (B.24), it follows that
\[
\frac{1}{T} \sum_{t=1}^{T} (\hat{g}_t - \hat{H}^{-1} g_t) g_t' = O_p \left( \frac{1}{N} \right) + O_p \left( \frac{1}{\sqrt{NC}} \right) + O_p \left( \frac{1}{\sqrt{NT}} \right) + O_p \left( \frac{1}{\sqrt{C}} \right).
\]

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

**Proof of Lemma 6** We proceed by following steps analogous to those in the proof of Proposition 1 in Bai (2003), and we develop the proof of the lemma for the sake of completeness. Given \( \hat{H} = (GG' / T) (A' \hat{A} / N) \hat{V}^{-1} \), pre-multiply both sides of the identity \( (1/NT) X' X \hat{A} = \hat{A} \hat{V} \) by \( (GG' / T)^{1/2} N^{-1} A' \) to obtain
\[
\frac{1}{N} \left( \frac{GG'}{T} \right)^{1/2} A' \left( \frac{X' X}{NT} \right) \hat{A} = \left( \frac{GG'}{T} \right)^{1/2} \left( \frac{A' \hat{A}}{N} \right) \hat{V}.
\]

Given (10), write \( X = G' A' + E \) with \( X = (x_1, \ldots, x_T)' \) and \( E = (e_1, \ldots, e_T)' \). We thus have
\[
\frac{1}{N} \left( \frac{GG'}{T} \right)^{1/2} A' \left( \frac{AG' A'}{NT} \right) \hat{A} + \tilde{D} = \left( \frac{GG'}{T} \right)^{1/2} \left( \frac{A' \hat{A}}{N} \right) \hat{V},
\]

where
\[
\tilde{D} = \frac{1}{N} \left( \frac{GG'}{T} \right)^{1/2} A' \left( \frac{AGE + E' G' A' + E' E}{NT} \right) \hat{A} = o_p (1)
\]

by Lemma 2. Let
\[
W = \left( \frac{GG'}{T} \right)^{1/2} \left( \frac{A' \hat{A}}{N} \right) \left( \frac{GG'}{T} \right)^{1/2}, \ \tilde{Z} = \left( \frac{GG'}{T} \right)^{1/2} \left( \frac{A' \hat{A}}{N} \right),
\]

so that we can write (B.25) as
\[
(W + \tilde{D} \tilde{Z}^{-1}) \tilde{Z} = \tilde{Z} \hat{V}.
\]

Therefore, each column of \( \tilde{Z} \) is an eigenvector of \( (W + \tilde{D} \tilde{Z}^{-1}) \), with length different from unity. Let
$
abla^*$ be the diagonal matrix of the diagonal elements of $\hat{Z}\hat{Z}$. Define $\hat{\Psi} = \hat{Z}(\nabla^*)^{-1/2}$ so that each column of $\hat{\Psi}$ has unit length. We thus get

$$
(W + \hat{D}\hat{Z}^{-1})\hat{\Psi} = \hat{\Psi}\hat{V},
$$

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

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

and note that

$$
\frac{GG'}{T} = \frac{1}{T} \sum_{t=1}^T \begin{pmatrix} I_{1t}f_{1t} & I_{2t}f_{2t} \end{pmatrix}' = \frac{1}{T} \sum_{t=1}^T \begin{pmatrix} I_{1t}f_{1t}f_{1t}' & 0 \\ 0 & I_{2t}f_{2t}f_{2t}' \end{pmatrix} \overset{p}{\to} \begin{pmatrix} \Sigma_{f_1} & 0 \\ 0 & \Sigma_{f_2} \end{pmatrix} = \Sigma_g
$$

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

$$
A'\hat{A}/N = \left(\frac{GG'}{T}\right)^{-1/2} \hat{\Psi}(\nabla^*)^{-1/2} \overset{p}{\to} \Sigma_g^{-1/2} \Psi V^{1/2}
$$

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

\[\square\]
Proof of Lemma 7. From Lemma 6 and taking into account (23), we have

\[
Q = \begin{pmatrix}
Q_1 \\
Q_2
\end{pmatrix} = \Sigma_{g}^{-1/2} \Psi V^{1/2} = \begin{pmatrix}
\Sigma_{f_1} & 0 \\
0 & \Sigma_{f_2}
\end{pmatrix}^{-1/2} \Psi V^{1/2},
\]

which completes the proof of the lemma.

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

\[
\frac{1}{NT} \sum_{t=1}^{T} x_t x_t' = \frac{1}{\sqrt{NT}} \left( \frac{1}{T} \sum_{t=1}^{T} (Ag_t + e_t)(Ag_t + e_t)' \right) = \frac{1}{\sqrt{NT}} \left( \frac{1}{T} \sum_{t=1}^{T} g_t g_t' \right) + \frac{1}{N} \left( \frac{1}{T} \sum_{t=1}^{T} e_t e_t' \right). \tag{B.26}
\]

Taking into account Assumption 2(b) and Assumption 4, it follows that

\[
\left\| \frac{1}{N} \left( \frac{1}{T} \sum_{t=1}^{T} g_t g_t' \right) \right\| \leq \frac{1}{\sqrt{NT}} \left\| \frac{1}{T} \sum_{t=1}^{T} \left( I_{1t} f_{1t} e_t' + I_{2t} f_{2t} e_t' \right) \right\| = O_p \left( \frac{1}{\sqrt{T}} \right) \tag{B.27}\]

Similarly, we can prove that

\[
\frac{1}{N} A \left( \frac{1}{T} \sum_{t=1}^{T} e_t g_t' \right) = O_p \left( \frac{1}{\sqrt{T}} \right). \tag{B.28}\]

Finally, by the weak dependence condition in Assumption 4,

\[
\left\| \frac{1}{NT} \sum_{t=1}^{T} e_t e_t' \right\| = o_p (1). \tag{B.29}\]

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

\[
\frac{1}{NT} \sum_{t=1}^{T} x_t x_t' = \frac{A}{\sqrt{N}} \left( \frac{1}{T} \sum_{t=1}^{T} g_t g_t' \right) \frac{A'}{\sqrt{N}} + o_p (1) = \frac{A}{\sqrt{N}} \frac{G G'}{\sqrt{T}} \frac{A'}{\sqrt{N}} + o_p (1).
\]

The result in the lemma follows from Assumptions 1 and 2 by noting that the eigenvalues of \( \left( A / \sqrt{N} \right) \left( G G' / T \right) \left( A' / \sqrt{N} \right) \) are the same as those of \( \left( G' / \sqrt{T} \right) \left( A' A / N \right) \left( G / \sqrt{T} \right) \). □
**Proof of Lemma** From the definition of \( \hat{I}_{\hat{\xi}_k} \) in (31), and taking into account Lemma (5)(a), we have

\[
\hat{I}_{\hat{\xi}_k} = \left( \sum_{t=1}^{T} \hat{\xi}_{j,t} I_{j,t} \sum_{t=1}^{T} \hat{I}_{j,t} \hat{g}_{j,t} \right)^{-1} \left( \sum_{t=1}^{T} \hat{\xi}_{j,t} I_{j,t} \hat{g}_{j,t} \right) 
\]

Taking further into account the definition of \( \hat{g}_t \) in (3), it follows that

\[
\hat{I}_{\hat{\xi}_k} = \hat{H}^{-1} \left( \frac{1}{T} \sum_{t=1}^{T} \hat{\xi}_{j,t} I_{j,t} \sum_{t=1}^{T} \hat{\xi}_{j,t} I_{j,t} \hat{g}_{j,t} \right)^{-1} \hat{H} + o_p(1) 
\]
where the last equality follows from (30). Therefore,

$$\tilde{\mathbf{I}}_{\xi k_1} \xrightarrow{p} \mathbf{H}^{-1} \begin{bmatrix} \mathbb{I}(j = 1) \mathbf{I}_{r_1} & \mathbf{0} \\ \mathbf{0} & \mathbb{I}(j = 2) \mathbf{I}_{r_2} \end{bmatrix} \mathbf{H},$$

which completes the proof of the lemma.

**Proof of Lemma 10.** From the definitions of eigenvectors and eigenvalues, for \( j = 1, 2 \) it follows that

$$\sum_{t=1}^{T} \hat{\xi}_{j} | T \mathbf{x}_t \mathbf{x}_t' = \hat{\Lambda}_{\xi, j}^{(p)} \hat{\mathbf{V}}_{\xi, j}^{(p)},$$

and, given the definition of \( \hat{\sum}_{\xi, x_j} \) in (30), we can write

$$\frac{\sum_{t=1}^{T} \hat{\xi}_{j} | T \mathbf{x}_t \mathbf{x}_t'}{N \sum_{t=1}^{T} \hat{\xi}_{j} | T} = \frac{\hat{\Lambda}_{\xi, j}^{(p)} \hat{\mathbf{V}}_{\xi, j}^{(p)}}{N} \tag{B.30}$$

The normalisation constraint

$$\frac{\hat{\Lambda}_{\xi, j}^{(p)} \hat{\mathbf{V}}_{\xi, j}^{(p)}}{N} = \mathbf{I}_p \tag{B.31}$$

allows us to obtain

$$\frac{\hat{\Lambda}_{\xi, j}^{(p)} \sum_{t=1}^{T} \hat{\xi}_{j} | T \mathbf{x}_t \mathbf{x}_t'}{N \sum_{t=1}^{T} \hat{\xi}_{j} | T} = \hat{\mathbf{V}}_{\xi, j}^{(p)}. \tag{B.32}$$

Taking into account Assumption 2(b), we then have

$$\left\| \frac{\hat{\Lambda}_{\xi, j}^{(p)} \sum_{t=1}^{T} \hat{\xi}_{j} | T \mathbf{x}_t \mathbf{x}_t'}{N \sum_{t=1}^{T} \hat{\xi}_{j} | T} \right\|_{\mathbb{P}} \leq \left\| \frac{\hat{\Lambda}_{\xi, j}^{(p)} \sum_{t=1}^{T} \hat{\xi}_{j} | T \mathbf{x}_t \mathbf{x}_t'}{\sqrt{N} \sum_{t=1}^{T} \hat{\xi}_{j} | T} \right\|_{\mathbb{P}} \leq \left\| \frac{\sum_{t=1}^{T} \hat{\xi}_{j} | T \mathbf{x}_t \mathbf{x}_t'}{\sum_{t=1}^{T} \hat{\xi}_{j} | T} \right\|_{\mathbb{P}} O_p(1) \tag{B.33}.$$
By Assumptions 1(b) and 2(b), it follows that

\[ \left\| \frac{A_1 \left( \sum_{t=1}^{T} I_t \hat{\xi}_{j|T} f_{1t} f'_{1t} \right)}{N \sum_{t=1}^{T} \hat{\xi}_{j|T}} \right\| = \left\| \frac{A_1' A_1}{N} T \frac{\left( \sum_{t=1}^{T} I_t \hat{\xi}_{j|T} f_{1t} f'_{1t} \right)}{T} \right\| \leq \frac{T}{\sum_{t=1}^{T} \hat{\xi}_{j|T}} \left\| \frac{A_1' A_1}{N} \left\| \sum_{t=1}^{T} I_t \hat{\xi}_{j|T} f_{1t} f'_{1t} \right\| \right\| = O_p(1). \]  

(B.34)

In a similar way, it can be proved that

\[ \left\| \frac{A_1 \left( \sum_{t=1}^{T} I_2 \hat{\xi}_{j|T} f_{2t} f'_{2t} \right)}{N \sum_{t=1}^{T} \hat{\xi}_{j|T}} \right\| = O_p(1). \]  

(B.35)

Assumptions 2(b) implies that

\[ \left\| \frac{A_1 \left( \sum_{t=1}^{T} I_t \hat{\xi}_{j|T} f_{1t} e'_{1t} \right)}{N \sum_{t=1}^{T} \hat{\xi}_{j|T}} \right\| \leq \frac{1}{\sqrt{T}} \left\| \frac{A_1}{\sqrt{N}} \left\| \sum_{t=1}^{T} I_t \hat{\xi}_{j|T} f_{1t} e'_{1t} \right\| \right\| = O_p(1), \]  

(B.36)
and, taking into account Assumption \[4\],

\[
\left\| \sum_{t=1}^{T} \frac{1}{\sqrt{NT}} \hat{\xi}_{jt} \hat{f}_t e_t' \right\| = \sqrt{\text{tr} \left( \left( \sum_{t=1}^{T} \frac{1}{\sqrt{NT}} \hat{\xi}_{jt} \hat{f}_t e_t' \right) \left( \sum_{t=1}^{T} \frac{1}{\sqrt{NT}} \hat{\xi}_{jt} \hat{f}_t e_t' \right)' \right)}^{1/2}
\]
\[
= \sqrt{\text{tr} \left( \left( \sum_{t=1}^{T} \frac{1}{\sqrt{NT}} \hat{\xi}_{jt} \hat{f}_t e_t' \right) \left( \sum_{t=1}^{T} \frac{1}{\sqrt{NT}} \hat{\xi}_{jt} \hat{f}_t e_t' \right)' \right)}^{1/2}
\]
\[
= \sqrt{\text{tr} \left( \left( \frac{1}{\sqrt{NT}} \sum_{t=1}^{T} \hat{\xi}_{jt} \hat{f}_t e_t' \right) \left( \frac{1}{\sqrt{NT}} \sum_{t=1}^{T} \hat{\xi}_{jt} \hat{f}_t e_t' \right)' \right)}^{1/2}
\]
\[
= \sqrt{\frac{1}{N} \sum_{j=1}^{N} \left\| \frac{1}{\sqrt{T}} \left( \sum_{t=1}^{T} \hat{\xi}_{jt} \hat{f}_t e_t' \right) \right\|^{2}}^{1/2}
\]
\[
= O_p(1),
\]

and taking into account \(B.36\) and \(B.37\),

\[
\left\| \frac{\sum_{t=1}^{T} \frac{1}{\sqrt{NT}} \hat{\xi}_{jt} \hat{f}_t e_t'}{N \sum_{t=1}^{T} \hat{\xi}_{jt} / T} \right\| = O_p \left( \frac{1}{\sqrt{T}} \right).
\]  

(B.38)

In a similar way, it can be proved that

\[
\left\| \frac{\sum_{t=1}^{T} \frac{1}{\sqrt{NT}} \hat{\xi}_{jt} \hat{f}_t e_t'}{N \sum_{t=1}^{T} \hat{\xi}_{jt} / T} \right\| = O_p \left( \frac{1}{\sqrt{T}} \right),
\]

(B.39)

and

\[
\left\| \frac{\sum_{t=1}^{T} \frac{1}{\sqrt{NT}} \hat{\xi}_{jt} \hat{f}_t e_t'}{N \sum_{t=1}^{T} \hat{\xi}_{jt} / T} \right\| = O_p \left( \frac{1}{\sqrt{T}} \right).
\]

(B.40)
Finally, by Assumption 3(b),

\[
\left\| \frac{1}{N} \sum_{t=1}^{T} \hat{\xi}_{jt}^T e_t e_t' \right\| \leq \frac{1}{N} \sum_{t=1}^{T} \hat{\xi}_{jt}^T \left\| e_t \right\| \left\| e_t' \right\| \\
\leq \frac{1}{N} \sum_{t=1}^{T} \hat{\xi}_{jt}^T \left( N^{-1/2} \left\| e_t \right\| + N^{-1/2} \left\| e_t' \right\| \right)^2 \\
= O_p(1).
\]

By combining equations (B.32), (B.33), (B.34), (B.35), (B.38), (B.39), (B.40), (B.41) and (B.42), it follows that

\[
\frac{1}{N} \sum_{t=1}^{T} \hat{\xi}_{jt}^T x_t x'_t = O_p(1),
\]

which, taking into account (B.32), implies that

\[
\frac{\Lambda^{(p)}}{N} \sum_{t=1}^{T} \hat{\xi}_{jt}^T x_t x'_t \Lambda^{(p)}_{\hat{\xi},j} = O_p(1).
\]

The result stated in the lemma then follows directly from (B.30) and (B.31).

**Proof of Lemma** Let \( \rho_{\hat{\xi},ijkl} = \sigma_{\hat{\xi},ijkl} / \left( \sigma_{\hat{\xi},jkl} \sigma_{\hat{\xi},jkl} \right)^{1/2} \) such that \( |\rho_{\hat{\xi},ijkl}| \leq 1 \). Since \( |\sigma_{\hat{\xi},jkl}| \leq M < \infty \) by Assumption 3(c), then

\[
\frac{1}{N} \sum_{i=1}^{N} \sum_{l=1}^{N} \sigma_{\hat{\xi},ijkl}^2 = \frac{1}{N} \sum_{i=1}^{N} \sum_{l=1}^{N} \rho_{\hat{\xi},ijkl}^2 \sigma_{\hat{\xi},jkl} \sigma_{\hat{\xi},jkl} \\
\leq MN^{-1} \sum_{i=1}^{N} \sum_{l=1}^{N} |\sigma_{\hat{\xi},jkl} \sigma_{\hat{\xi},jkl}|^{1/2} |\rho_{\hat{\xi},ijkl}| \\
= MN^{-1} \sum_{i=1}^{N} \sum_{l=1}^{N} |\sigma_{\hat{\xi},jkl}| \\
\leq MT^{-1} \sum_{t=1}^{T} \left[ N^{-1} \sum_{i=1}^{N} \sum_{l=1}^{N} \left| E \left( \| \xi_{jt}^T x_t e_t' \| \right) \right| \right] \\
\leq M^2,
\]

by Assumption 3(b), which completes the proof of the lemma.

**B.3 Proof of Theorem**

Given the specification in (1), from Section 2.2 recall \( B_1 = [A_1 \ 0] \) and \( B_2 = [0 \ A_2] \). Adding and subtracting terms, we have

\[
\begin{align*}
x_t &= I_{1t} B_1 g_t + I_{2t} B_2 g_t + e_t \\
&= I_{1t} B_1 \hat{H} g_t + I_{2t} B_2 \hat{H} g_t + I_{1t} B_1 \hat{H} \left( \hat{H}^{-1} g_t - \hat{g}_t \right) + I_{2t} B_2 \hat{H} \left( \hat{H}^{-1} g_t - \hat{g}_t \right) + e_t,
\end{align*}
\]

(B.43)

where \( \hat{H} \) is defined in (27), and \( \hat{g}_t \) is the estimator for \( g_t \) given in (18). We focus upon \( \hat{B}_1 = \left[ \hat{b}_{11}, \ldots, \hat{b}_{1N} \right] \) as an estimator for \( B_1 = [b_{11}, \ldots, b_{1N}] \): analogous arguments hold for \( \hat{B}_2 \). From
and taking into account (B.43), we have

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

Since \( \mathbb{I}_{2t} = 1 - \mathbb{I}_{1t} \), and recalling the definition of \( \hat{I}_{\hat{\xi}_1} \) in (31), after some algebra we get

\[
\sqrt{T} \left[ \hat{B}_1 - \hat{B}_1 \hat{H} \hat{I}_{\hat{\xi}_1} - \hat{B}_2 \hat{H} \left( 1 - \hat{I}_{\hat{\xi}_1} \right) \right] = \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \hat{\xi}_{1,t|T} \mathbf{e}_t \hat{g}_t' \right) \left( \frac{1}{T} \sum_{t=1}^{T} \hat{\xi}_{1,t|T} \hat{g}_t \hat{g}_t' \right)^{-1} \\
+ \hat{B}_1 \hat{H} \left[ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \hat{\xi}_{1,t|T} \mathbb{I}_{1t} \left( \hat{H}^{-1} \mathbf{g}_t - \hat{g}_t \right) \hat{g}_t' \right] \left( \frac{1}{T} \sum_{t=1}^{T} \hat{\xi}_{1,t|T} \hat{g}_t \hat{g}_t' \right)^{-1} \\
\times \left( \frac{1}{T} \sum_{t=1}^{T} \hat{\xi}_{1,t|T} \hat{g}_t \hat{g}_t' \right)^{-1} \\
+ \hat{B}_2 \hat{H} \left[ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \hat{\xi}_{1,t|T} \mathbb{I}_{2t} \left( \hat{H}^{-1} \mathbf{g}_t - \hat{g}_t \right) \hat{g}_t' \right] \left( \frac{1}{T} \sum_{t=1}^{T} \hat{\xi}_{1,t|T} \hat{g}_t \hat{g}_t' \right)^{-1} \\
\times \left( \frac{1}{T} \sum_{t=1}^{T} \hat{\xi}_{1,t|T} \hat{g}_t \hat{g}_t' \right)^{-1}.
\]

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

\[
\frac{1}{T} \sum_{t=1}^{T} \hat{\xi}_{1,t|T} \mathbb{I}_{jt} \left( \hat{H}^{-1} \mathbf{g}_t - \hat{g}_t \right) \hat{g}_t' \leq M \left[ \frac{1}{T} \sum_{t=1}^{T} \left( \hat{H}^{-1} \mathbf{g}_t - \hat{g}_t \right) \hat{g}_t' \right] = O_p \left( \frac{1}{\sqrt{NT}} \right). \tag{B.45}
\]

From (B.44) and (B.45), and taking into account Assumption 7, it follows that

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

\[
\frac{\sqrt{T}}{\bar{B}_1 - B_1 \bar{H} \bar{H} \xi_1 - B_2 \bar{H} (I - \bar{\xi}_1)} = \left\{ \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \bar{\xi}_{1,t} \right) \bar{B}_1 \bar{A} - \bar{A} \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \bar{\xi}_{1,t} \right) \bar{B}_2 \right\}^{-1} \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \bar{\xi}_{1,t} \right) \left( \bar{B}_1 \bar{A} - \bar{A} \bar{B}_2 \right) + o_p(1).
\]

By Lemma 2 and taking into account the identity in (B.3), it follows that

\[
\bar{A}' - \bar{H}' A' = O_p \left( \frac{1}{\sqrt{T} C_{NT}} \right) + O_p \left( \frac{1}{\sqrt{T} C_{NT}} \right) + O_p \left( \frac{1}{\sqrt{T}} \right),
\]

which implies that

\[
\bar{A} - \bar{A} \bar{H} = O_p \left( \frac{1}{\sqrt{T} C_{NT}} \right) + O_p \left( \frac{1}{\sqrt{T} C_{NT}} \right) + O_p \left( \frac{1}{\sqrt{T}} \right).
\]

From (B.46) through (B.48), it follows that

\[
\sqrt{T} \left[ \bar{b}_{11} - \bar{\bar{\xi}}_1 \bar{H} \bar{b}_{11} - (I - \bar{\xi}_1) \bar{H} \bar{b}_{21} \right] = \left[ \left( \frac{1}{T} \sum_{t=1}^{T} \bar{\xi}_{1,t} \right) \bar{B}_1 \bar{A} - \bar{A} \left( \frac{1}{T} \sum_{t=1}^{T} \bar{\xi}_{1,t} \right) \bar{B}_2 \right]^{-1} \left[ \left( \frac{1}{T} \sum_{t=1}^{T} \bar{\xi}_{1,t} \right) \bar{B}_1 \bar{A} - \bar{A} \left( \frac{1}{T} \sum_{t=1}^{T} \bar{\xi}_{1,t} \right) \bar{B}_2 \right] + o_p(1),
\]

and the result stated in the theorem follows by Assumption 1 and Lemma 2 and by noting that, by Assumption 2(c), \( T^{-1/2} \sum_{t=1}^{T} \bar{\xi}_{1,t} f_{1} e_{21} \) and \( T^{-1/2} \sum_{t=1}^{T} \bar{\xi}_{1,t} f_{21} e_{21} \) converge in distribution to two independent Normal random variables.

### B.4 Proof of Theorem 2

Given the representation in (9), we can write

\[
x_t = (B_1 B_2) (\xi_t \otimes g_t) + e_t = (B_1 B_2) (\xi_t g_t \xi_t g_t) + e_t.
\]

Recall also the estimators \( \hat{B}_1 \) and \( \hat{B}_2 \) defined according to (A.13), with \( \hat{B}_j \equiv \hat{B}_j^{(k^*)+1} \), where \( k^* \) is the last iteration of the EM algorithm detailed in Section A. The estimators \( \hat{\xi}_{1,t} T \bar{g}_t \) and \( \hat{\xi}_{2,t} T \bar{g}_t \) for \( \xi_{1,t} g_t \)
and $\xi_{2t}g_t$, respectively, are obtained as

$$
\begin{pmatrix}
\hat{\xi}_{1,t|T} g_t \\
\hat{\xi}_{2,t|T} g_t \\
\end{pmatrix}
= 
\left[(\hat{B}_1 \hat{B}_2) (B_1 B_2)^{-1} \right. \\
\left. (\hat{B}_1 \hat{B}_2) (B_1 B_2)^{-1} \right] \xi_{1t} g_t \\
+ 
(\hat{B}_1 \hat{B}_2) (B_1 B_2)^{-1} \xi_{2t} g_t
$$

Adding and subtracting terms, it follows that

$$
\begin{pmatrix}
\hat{\xi}_{1,t|T} g_t \\
\hat{\xi}_{2,t|T} g_t \\
\end{pmatrix}
= 
\left[(\hat{B}_1 \hat{B}_2) (B_1 B_2)^{-1} \right. \\
\left. (\hat{B}_1 \hat{B}_2) (B_1 B_2)^{-1} \right] \xi_{1t} g_t \\
+ 
(\hat{B}_1 \hat{B}_2) (B_1 B_2)^{-1} \xi_{2t} g_t
$$

or equivalently

$$
\begin{pmatrix}
\hat{\xi}_{1,t|T} g_t \\
\hat{\xi}_{2,t|T} g_t \\
\end{pmatrix}
- 
\hat{H}_\xi^{-1} \left[(\hat{B}_1 \hat{B}_2) (B_1 B_2)^{-1} \right. \\
\left. (\hat{B}_1 \hat{B}_2) (B_1 B_2)^{-1} \right] \xi_{1t} g_t
$$

Consider first

$$
\frac{1}{N} \left\{ \hat{B}_1 \hat{B}_2 \right\} \left[ (B_1 B_2) - (\hat{B}_1 \hat{B}_2) \hat{H}_\xi^{-1} \right] \xi_{1t} g_t
$$

$$
= 
\frac{1}{N} \left\{ \hat{B}_1 \hat{B}_2 \right\} \left[ (B_1 B_2) \hat{H}_\xi - (\hat{B}_1 \hat{B}_2) \hat{H}_\xi^{-1} \right] \xi_{1t} g_t,
$$

70
so that from (B.44) and (B.45), and taking into account Assumption 2, it follows that

\[
\begin{align*}
\left\| \frac{1}{\sqrt{N}} \begin{pmatrix} \hat{B}_1' \\ \hat{B}_2' \end{pmatrix} \left( [B_1 B_2] - [\hat{B}_1 \hat{B}_2] \right) \right\| & \leq \frac{1}{\sqrt{N}} \left\| \begin{pmatrix} [B_1 B_2] \hat{H}_\xi \left( \xi_{1t} g_t \right) \\ \hat{H}_\xi \left( \xi_{2t} g_t \right) \end{pmatrix} \right\| \\
& = O_p \left( \frac{1}{\sqrt{NT}} \right) + O_p \left( \frac{1}{\sqrt{NC^2_{NT}}} \right).
\end{align*}
\]

By (B.44) and (B.45), and taking into account Assumption 3(b), we also have that,

\[
\begin{align*}
\left\| \frac{1}{\sqrt{N}} \left( [\hat{B}_1' \hat{B}_2'] - \hat{H}_\xi \right) e_t \right\| & \leq \frac{1}{\sqrt{N}} \left\| \begin{pmatrix} \hat{B}_1' \\ \hat{B}_2' \end{pmatrix} \right\| \left\| \begin{pmatrix} \hat{B}_1' \\ \hat{B}_2' \end{pmatrix} \right\| \\
& = O_p \left( \frac{1}{\sqrt{NT}} \right) + O_p \left( \frac{1}{\sqrt{NC^2_{NT}}} \right).
\end{align*}
\]

Therefore, taking into account (B.49), (B.50) and (B.51), and by Assumption 7 we have

\[
\begin{align*}
\sqrt{N} \left[ \hat{\xi}_{1,t(T)} \hat{g}_t - \hat{H}_\xi \left( \xi_{1t} g_t \right) \right] & = \left( \begin{pmatrix} \hat{B}_1 \hat{B}_2 \\ \hat{B}_2 \hat{B}_1 \end{pmatrix} \right) \left( \begin{pmatrix} \hat{B}_1' \\ \hat{B}_2' \end{pmatrix} \right) e_t + o_p(1).
\end{align*}
\]

Given \( \hat{H}_\xi \), recall \( I_{\xi_j} = \lim_{N,T \to \infty} \hat{I}_{\xi_j} \) for \( j = 1, 2 \), where \( I_{\xi_j} \) and \( \hat{I}_{\xi_j} \) are defined in Lemma 2 and in (31), respectively. Also, given \( \hat{H} \) defined in (27), we have \( \hat{H} \overset{p}{\to} \Sigma g Q V^{-1} = H \), where \( \Sigma = \lim_{N,T \to \infty} (GG^T) \) by Assumption 1, and \( Q = \lim_{N,T \to \infty} (A^T A/N) \) by Lemma 3. By Theorem 4 we then have \( (B_1 B_2) \overset{p}{\to} H_{\xi}(B_1 B_2)' \). Therefore

\[
\begin{align*}
p \lim_{N,T \to \infty} \left( \begin{pmatrix} \hat{B}_1 \hat{B}_2 \\ \hat{B}_2 \hat{B}_1 \end{pmatrix} \right) & = H_{\xi}(\Sigma_{B1} \Sigma_{B12} \Sigma_{B21} \Sigma_{B2}) H_{\xi}^T,
\end{align*}
\]

where, by Assumption 2 \( \|(B_j' B_j/N) - \Sigma_{Bj}\| \to 0 \) and \( \|(B_j' B_k/N) - \Sigma_{Bjk}\| \to 0 \), for \( j, k = 1, 2 \) with \( j \neq k \) as \( N \to \infty \). The result stated in the theorem follows by noting that

\[
\frac{1}{\sqrt{N}} \begin{pmatrix} B_1' \\ B_2' \end{pmatrix} e_t \overset{d}{\to} N(0, \Sigma_{Ber}).
\]

by Assumption 6(d), which concludes the proof.
B.5 Proof of Theorem 3

Given $r_1 = r_2$, consider $j = 1$: analogous arguments hold for $j = 2$. We can then partition the vector $\hat{b}_{1i}$ in (33) as

$$\hat{b}_{1i} = \begin{pmatrix} \hat{b}_{1i}^{(1)} \\ \hat{b}_{1i}^{(2)} \end{pmatrix}.$$ 

In this way, (33) itself may be written as

$$\sqrt{T} \left\{ \hat{b}_{1i}^{(1)} - \lambda_{1i}' \left[ \tilde{R}_{1,11}, \tilde{R}_{1,12} \right] - \lambda_{2i}' \left[ (\hat{H}_{21} - \hat{R}_{1,21}), (\hat{H}_{22} - \hat{R}_{1,22}) \right] \right\} = \sqrt{T} \left\{ (\hat{b}_{1i}^{(1)'}, \hat{b}_{1i}^{(2)'})' - \lambda_{1i}' \left[ \tilde{R}_{1,11}, \tilde{R}_{1,12} \right] - \lambda_{2i}' \left[ (\hat{H}_{21} - \hat{R}_{1,21}), (\hat{H}_{22} - \hat{R}_{1,22}) \right] \right\} = \sqrt{T} \left\{ \hat{b}_{1i}^{(1)} - \lambda_{1i}' \tilde{R}_{1,11} - \lambda_{2i}' (\hat{H}_{21} - \hat{R}_{1,21}), \hat{b}_{1i}^{(2)} - \lambda_{1i}' \tilde{R}_{1,12} - \lambda_{2i}' (\hat{H}_{22} - \hat{R}_{1,22}) \right\}.$$ 

Since it is known that $r_1 = r_2$, the estimator $\hat{\lambda}_{1i}$ for $\lambda_{1i}$ is equal to $\hat{b}_{1i}^{(1)}$. Formally, for $i = 1, \ldots, N$, it follows that

$$\sqrt{T} \left\{ \hat{b}_{1i}^{(1)} - \lambda_{1i}' \tilde{R}_{1,11} - \lambda_{2i}' (\hat{H}_{21} - \hat{R}_{1,21}) \right\} = \sqrt{T} \left\{ \hat{\lambda}_{1i}' - \lambda_{1i}' \tilde{R}_{1,11} - \lambda_{2i}' (\hat{H}_{21} - \hat{R}_{1,21}) \right\}.$$ 

Given $\hat{\lambda}_1 = \left( \hat{\lambda}_{11}, \ldots, \hat{\lambda}_{1N} \right)'$, from (37) interest lies in

$$\hat{f}_{1i} = \hat{\xi}_{1,i|T} (\hat{\lambda}_1' \hat{\lambda}_1)^{-1} (\hat{\lambda}_1' x_i) = \left( \hat{\lambda}_1' \hat{\lambda}_1 \right)^{-1} \left( \hat{\xi}_{1,i|T} \hat{\lambda}_1 \right) = \left( \hat{\lambda}_1' \hat{\lambda}_1 \right)^{-1} \left[ \hat{\lambda}_1' \hat{\xi}_{1,i|T} (\hat{\lambda}_1' \hat{\lambda}_1 + \hat{\lambda}_2' \hat{\lambda}_2 + e_i) \right] = \left( \hat{\lambda}_1' \hat{\lambda}_1 \right)^{-1} \left( \hat{\lambda}_1' \hat{\lambda}_1 \right) \left( \hat{\xi}_{1,i|T} \hat{f}_{1i} \right) + \left( \hat{\lambda}_1' \hat{\lambda}_1 \right)^{-1} \left( \hat{\lambda}_1' \hat{\lambda}_2 \right) \left( \hat{\xi}_{1,i|T} \hat{f}_{2i} \right) + \left( \hat{\lambda}_1' \hat{\lambda}_1 \right)^{-1} \left( \hat{\lambda}_1' \hat{\xi}_{1,i|T} e_i \right).$$ 

(B.52)

Adding and subtracting terms, we have

$$\hat{\lambda}_1 = \hat{\lambda}_1 - \lambda_1 \tilde{R}_{1,11} - \lambda_2 (\hat{H}_{21} - \hat{R}_{1,21}) + \lambda_1 \tilde{R}_{1,11} + \lambda_2 (\hat{H}_{21} - \hat{R}_{1,21}),$$

which implies that

$$\frac{\hat{\lambda}_1' \lambda_1}{N} = \frac{(\hat{\lambda}_1 - \lambda_1 \tilde{R}_{1,11} - \lambda_2 (\hat{H}_{21} - \hat{R}_{1,21}))' \lambda_1}{N} = \frac{\lambda_1 \tilde{R}_{1,11} + \lambda_2 (\hat{H}_{21} - \hat{R}_{1,21})' \lambda_1}{N}.$$
Note that \[
\left( \hat{\Lambda}_1 - \Lambda_1 \hat{R}_{1,11} - \Lambda_2 \left( \hat{H}_{21} - \hat{R}_{1,21} \right) \right) / N \] is of the same order as \( (\hat{A} - A \hat{H}) / N \). Therefore, by (B.15) it follows that
\[
\frac{\hat{\Lambda}_1 - \Lambda_1 \hat{R}_{1,11} - \Lambda_2 \left( \hat{H}_{21} - \hat{R}_{1,21} \right)}{N} \right) A_1 = O_p \left( \frac{1}{C_{NT}} \right),
\]
so that
\[
\frac{\hat{\Lambda}_1 A_1}{N} = \left[ \frac{\Lambda_1 \hat{R}_{1,11} + \Lambda_2 \left( \hat{H}_{21} - \hat{R}_{1,21} \right)}{N} \right] A_1 + O_p \left( \frac{1}{C_{NT}} \right). \tag{B.53}
\]
Similarly,
\[
\frac{\hat{\Lambda}_1 A_2}{N} = \left[ \frac{\Lambda_1 \hat{R}_{1,11} + \Lambda_2 \left( \hat{H}_{21} - \hat{R}_{1,21} \right)}{N} \right] A_2 + O_p \left( \frac{1}{C_{NT}} \right). \tag{B.54}
\]
Also,
\[
\frac{\hat{\Lambda}_1 A_1}{N} = \left[ \frac{\hat{\Lambda}_1 - \Lambda_1 \hat{R}_{1,11} - \Lambda_2 \left( \hat{H}_{21} - \hat{R}_{1,21} \right)}{N} \right] A_1 + O_p \left( \frac{1}{C_{NT}} \right)
\times \left[ \frac{\Lambda_1 \hat{R}_{1,11} + \Lambda_2 \left( \hat{H}_{21} - \hat{R}_{1,21} \right)}{N} \right] + O_p \left( \frac{\sqrt{N}}{C_{NT}} \right)
\times \left[ \frac{\Lambda_1 \hat{R}_{1,11} + \Lambda_2 \left( \hat{H}_{21} - \hat{R}_{1,21} \right)}{N} \right] A_2 + O_p \left( \frac{\sqrt{N}}{C_{NT}} \right).
\tag{B.55}
\]
Therefore, taking into account (B.52) through (B.55) we have

\[
\hat{f}_{1t} = \left\{ \begin{array}{c}
\frac{A_1 \hat{R}_{1,11} + A_2 \left( \hat{H}_{21} - \hat{R}_{1,21} \right)}{N} \\
\frac{A_1 \hat{R}_{1,11} + A_2 \left( \hat{H}_{21} - \hat{R}_{1,21} \right)}{N}
\end{array} \right\} \left[ \begin{array}{c}
\left( \xi_{1,t} + \frac{\sqrt{N}}{C_{NT}} \right) \\
\left( \xi_{1,t} + \frac{\sqrt{N}}{C_{NT}} \right)
\end{array} \right]^{-1}
\times \left\{ \begin{array}{c}
\left[ \frac{A_1 \hat{R}_{1,11} + A_2 \left( \hat{H}_{21} - \hat{R}_{1,21} \right)}{N} \right] \xi_{1,t} \\
\left[ \frac{A_1 \hat{R}_{1,11} + A_2 \left( \hat{H}_{21} - \hat{R}_{1,21} \right)}{N} \right] \xi_{1,t}
\end{array} \right\}
+ O_p \left( \frac{\sqrt{N}}{C_{NT}} \right).
\]

It follows that,

\[
\sqrt{N} \hat{f}_{1t} = \left\{ \begin{array}{c}
\left[ \frac{A_1 \hat{R}_{1,11} + A_2 \left( \hat{H}_{21} - \hat{R}_{1,21} \right)}{N} \right] \xi_{1,t} \\
\left[ \frac{A_1 \hat{R}_{1,11} + A_2 \left( \hat{H}_{21} - \hat{R}_{1,21} \right)}{N} \right] \xi_{1,t}
\end{array} \right\}^{-1}
\times \left\{ \begin{array}{c}
\left[ \frac{A_1 \hat{R}_{1,11} + A_2 \left( \hat{H}_{21} - \hat{R}_{1,21} \right)}{N} \right] \xi_{1,t} \\
\left[ \frac{A_1 \hat{R}_{1,11} + A_2 \left( \hat{H}_{21} - \hat{R}_{1,21} \right)}{N} \right] \xi_{1,t}
\end{array} \right\}
+ O_p \left( \frac{\sqrt{N}}{C_{NT}} \right).
\]

Consider

\[
\hat{\xi}_{1,t} = \left[ \frac{A_1 \hat{R}_{1,11} + A_2 \left( \hat{H}_{21} - \hat{R}_{1,21} \right)}{\sqrt{N}} \right] \xi_{1,t} \\
\hat{\xi}_{1,t} = \left[ \frac{A_1 \hat{R}_{1,11} + A_2 \left( \hat{H}_{21} - \hat{R}_{1,21} \right)}{\sqrt{N}} \right] \xi_{1,t}
\]

and let

\[
\xi_{1,t} = p \lim_{N,T \to \infty} \hat{\xi}_{1,t}.
\]

74
Further, from (32) recall that for $j = 1, 2$,

\[
\hat{R}_j = \hat{H} \hat{I}_{\xi_j} = \left( \begin{array}{cc} \hat{R}_{j,1} & \hat{R}_{j,2} \\ \hat{R}_{j,21} & \hat{R}_{j,22} \end{array} \right),
\]

where $\hat{H}$ and $\hat{I}_{\xi_j}$ are defined in (27) and (31), respectively. Taking into account (30) and Lemma (9), it follows that

\[
p \lim_{N,T \to \infty} \hat{R}_j = H \cdot I_{\xi_j} = H H^{-1} \left[ \begin{array}{cc} I(j = 1) & 0 \\ 0 & I(j = 2) \end{array} \right] H = \left[ \begin{array}{cc} I(j = 1) & 0 \\ 0 & I(j = 2) \end{array} \right] H.
\]

Given (30), from (23) and (28), recall the definitions of $\Sigma_g$ and $Q$, respectively. We then have

\[
H = \Sigma_g Q V^{-1} = \Sigma_g \left( \Sigma_g^{-1/2} \Psi V^{1/2} \right) V^{-1} = \Sigma_g^{1/2} \Psi V^{-1/2},
\]

which implies that

\[
H = \Sigma_g^{1/2} \Psi V^{-1/2} = \begin{pmatrix} \Sigma_{f_1}^{1/2} & 0 \\ 0 & \Sigma_{f_2}^{1/2} \end{pmatrix} \begin{pmatrix} \Psi_{11} & \Psi_{12} \\ \Psi_{21} & \Psi_{22} \end{pmatrix} \begin{pmatrix} V_1^{-1/2} & 0 \\ 0 & V_2^{-1/2} \end{pmatrix} = \begin{pmatrix} V_1^{-1/2} & 0 \\ 0 & V_2^{-1/2} \end{pmatrix}
\]

where $H_{jk} = p \lim_{N,T \to \infty} \hat{H}_{jk}$. Therefore,

\[
p \lim_{N,T \to \infty} \hat{R}_j = \begin{pmatrix} I(j = 1) & 0 \\ 0 & I(j = 2) \end{pmatrix} = \begin{pmatrix} \Sigma_{f_1}^{1/2} \Psi_{11} V_1^{-1/2} & \Sigma_{f_1}^{1/2} \Psi_{12} V_2^{-1/2} \\ \Sigma_{f_2}^{1/2} \Psi_{21} V_1^{-1/2} & \Sigma_{f_2}^{1/2} \Psi_{22} V_2^{-1/2} \end{pmatrix}.
\]

Therefore, we have $\hat{R}_{1,11} = H_{11} + \sigma_p (1)$ and $\hat{R}_{1,21} = \sigma_p (1)$. Taking this into account in (3.49) and (3.57), and recalling (3.58), it follows that

\[
\sqrt{N} \left\{ \tilde{r}_{11}^{(1)} \right\} = \xi_{1,t} \left\{ \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \lambda_{1i} e_{it} + H_{21}^{\prime} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \lambda_{2i} e_{it} \right\} + \sigma_p (1).
\]
By Assumption 3(d), it follows that

\[
\sqrt{N} \left\{ \tilde{f}_{it} - \left[ \left( \Lambda_1 \hat{H}_{11} + \Lambda_2 \hat{H}_{21} \right) N \left( \Lambda_1 \hat{H}_{11} + \Lambda_2 \hat{H}_{21} \right) \right]^{-1} \right. \\
\times \left. \times \left( \Lambda_1 \hat{H}_{11} + \Lambda_2 \hat{H}_{21} \right)^T \tilde{\xi}_{1,t} N (\tilde{\xi}_{11} \tilde{f}_{it} + \tilde{\xi}_{21} \tilde{f}_{2t}) \right\} 
\]

distributed as

\[
\mathcal{N}(0, \Sigma_{\tilde{f}_{it}}),
\]

where

\[
\Sigma_{\tilde{f}_{it}} = (\xi_{1,t})^2 \left( H_{11} \Phi_{11} H_{11} + H_{11} \Phi_{12} H_{21} + H_{21} \Phi_{12} H_{11} + H_{22} \Phi_{22} H_{22} \right),
\]

with \( \Phi_{12t} \) defined in Assumption 3(d). This which completes the proof of the theorem.

### B.6 Proof of Theorem 4

For \( j = 1, 2 \), consider the covariance matrix \( \tilde{\Sigma}_{\tilde{\xi},x_{ij}} \) defined in (39). By definition of eigenvectors and eigenvalues, it follows that

\[
\tilde{\Sigma}_{\tilde{\xi},x_{ij}} \tilde{A}_{\tilde{\xi},j} = \tilde{\Lambda}_{\tilde{\xi},j} \tilde{V}_{\tilde{\xi},j}.
\]

Recall the matrix \( \tilde{H}_{\xi,kj}^{(p)} \) defined according to (40).

We can then write

\[
\tilde{A}_{\tilde{\xi},j} \tilde{V}_{\tilde{\xi},j} = \left( \Lambda_j \tilde{H}_{\xi,jj}^{(p)} + \Lambda_k \tilde{H}_{\xi,kj}^{(p)} \right) \tilde{V}_{\tilde{\xi},j} = \tilde{\Sigma}_{\tilde{\xi},x_{ij}} \tilde{A}_{\tilde{\xi},j} = \left( \Lambda_j \tilde{H}_{\xi,jj}^{(p)} + \Lambda_k \tilde{H}_{\xi,kj}^{(p)} \right) \tilde{V}_{\tilde{\xi},j},
\]

which implies that

\[
\tilde{V}_{\tilde{\xi},j} \tilde{A}_{\tilde{\xi},j} = \tilde{V}_{\tilde{\xi},j} \left( \tilde{H}_{\xi,jj}^{(p)} \tilde{A}_j + \tilde{H}_{\xi,kj}^{(p)} \tilde{A}_k \right) = \tilde{\Lambda}_{\tilde{\xi},j} \tilde{\Sigma}_{\tilde{\xi},x_{ij}} - \tilde{V}_{\tilde{\xi},j} \left( \tilde{H}_{\xi,jj}^{(p)} \tilde{A}_j + \tilde{H}_{\xi,kj}^{(p)} \tilde{A}_k \right).
\]
Without loss of generality, set \( j = 1 \): the case \( j = 2 \) can be dealt with in a similar way. Since \( x_t = 1_{2t} \mathbf{A}_1 f_{t1} + 1_{2t} \mathbf{A}_2 f_{t2} + e_t \), and \( x_t = 1_{2t} \mathbf{A}_1 f_{t1} + 1_{2t} \mathbf{A}_2 f_{t2} + \epsilon_t \), we can write

\[
\begin{align*}
\hat{\psi}^{(p)}_{\xi, 1} & - \hat{\psi}^{(p)}_{\xi, 1} (\hat{H}^{(p)\prime}_{\xi, 1} \lambda_{i1} + \hat{H}^{(p)\prime}_{\xi, 2} \lambda_{i2}) \\
& = \hat{\lambda}^{(p)\prime}_{\xi, 1} N \sum_{t = 1}^T 1_{2t} x_t f_{t1} - \hat{\psi}^{(p)}_{\xi, 1} (\hat{H}^{(p)\prime}_{\xi, 1} \lambda_{i1} + \hat{H}^{(p)\prime}_{\xi, 2} \lambda_{i2}) \\
& = \hat{\lambda}^{(p)\prime}_{\xi, 1} N \frac{1}{T} \sum_{t = 1}^T T f_{t1} + \hat{\lambda}^{(p)\prime}_{\xi, 2} N \sum_{t = 1}^T T f_{t2} + e_t (1_{1t} \lambda_{i1} + 1_{2t} \lambda_{i2} f_{t2} + \epsilon_t) \\
\end{align*}
\]

or equivalently

\[
\begin{align*}
\hat{\psi}^{(p)}_{\xi, 1} & - \hat{\psi}^{(p)}_{\xi, 1} (\hat{H}^{(p)\prime}_{\xi, 1} \lambda_{i1} + \hat{H}^{(p)\prime}_{\xi, 2} \lambda_{i2}) \\
& = \hat{\lambda}^{(p)\prime}_{\xi, 1} N \frac{1}{T} \sum_{t = 1}^T T f_{t1} + \hat{\lambda}^{(p)\prime}_{\xi, 2} N \sum_{t = 1}^T T f_{t2} + e_t (1_{1t} \lambda_{i1} + 1_{2t} \lambda_{i2} f_{t2} + \epsilon_t) \\
& = \hat{\lambda}^{(p)\prime}_{\xi, 1} N \frac{1}{T} \sum_{t = 1}^T T f_{t1} + \hat{\lambda}^{(p)\prime}_{\xi, 2} N \sum_{t = 1}^T T f_{t2} + e_t (1_{1t} \lambda_{i1} + 1_{2t} \lambda_{i2} f_{t2} + \epsilon_t) \\
\end{align*}
\]
which is also equal to

\[
\tilde{\mathbf{V}}^{(p)}_{\xi,i} \left[ \hat{\mathbf{X}}^{(p)}_{\xi,11} - \left( \hat{\mathbf{H}}^{(p)'}_{\xi,11} \hat{\mathbf{A}}_{11} + \hat{\mathbf{H}}^{(p)'}_{\xi,21} \hat{\mathbf{A}}_{21} \right) \right] = \frac{1}{N} \sum_{t=1}^{T} \frac{1}{T} \sum_{l=1}^{T} \mathbf{E} \left[ \tilde{\xi}_{il} \tilde{\xi}_{lt} e_{il} e_{lt} \right] \\
+ \frac{1}{N} \sum_{t=1}^{T} \frac{1}{T} \sum_{l=1}^{T} \mathbf{E} \left[ \tilde{\xi}_{il} \tilde{\xi}_{lt} e_{il} e_{lt} \right] \\
+ \frac{1}{N} \sum_{t=1}^{T} \frac{1}{T} \sum_{l=1}^{T} \mathbf{E} \left[ \tilde{\xi}_{il} \tilde{\xi}_{lt} e_{il} e_{lt} - \mathbf{E} \left( \tilde{\xi}_{il} \tilde{\xi}_{lt} e_{il} e_{lt} \right) \right] \\
+ \frac{1}{N} \sum_{t=1}^{T} \frac{1}{T} \sum_{l=1}^{T} \mathbf{E} \left[ \tilde{\xi}_{il} \tilde{\xi}_{lt} e_{il} e_{lt} - \mathbf{E} \left( \tilde{\xi}_{il} \tilde{\xi}_{lt} e_{il} e_{lt} \right) \right] \\
+ \frac{1}{N} \sum_{t=1}^{T} \frac{1}{T} \sum_{l=1}^{T} \mathbf{E} \left[ \tilde{\xi}_{il} \tilde{\xi}_{lt} e_{il} e_{lt} - \mathbf{E} \left( \tilde{\xi}_{il} \tilde{\xi}_{lt} e_{il} e_{lt} \right) \right].
\]

In general, for \( j, k = 1, 2 \) define

\[
\sigma_{\tilde{\xi}, jk} = \frac{1}{T} \sum_{t=1}^{T} \mathbf{E} \left[ \tilde{\xi}_{jt} \tilde{\xi}_{kt} e_{jt} e_{kt} \right], \quad \chi_{\tilde{\xi}, jk} = \frac{1}{T} \sum_{t=1}^{T} \left[ \tilde{\xi}_{jt} \tilde{\xi}_{kt} e_{jt} e_{kt} - \mathbf{E} \left( \tilde{\xi}_{jt} \tilde{\xi}_{kt} e_{jt} e_{kt} \right) \right],
\]

\[
\varphi_{\tilde{\xi}, jk} = \frac{1}{T} \sum_{t=1}^{T} \chi_{j1} \tilde{\xi}_{kt} \tilde{\xi}_{kt} e_{jt} e_{kt}, \quad \varphi_{\tilde{\xi}, jk} = \frac{1}{T} \sum_{t=1}^{T} \chi_{j1} \tilde{\xi}_{kt} \tilde{\xi}_{kt} e_{jt} e_{kt}.
\]

We can then write

\[
\tilde{\mathbf{V}}^{(p)}_{\xi,j} \left[ \hat{\mathbf{X}}^{(p)}_{\xi,11} - \left( \hat{\mathbf{H}}^{(p)'}_{\xi,11} \hat{\mathbf{A}}_{11} + \hat{\mathbf{H}}^{(p)'}_{\xi,21} \hat{\mathbf{A}}_{21} \right) \right] = \frac{1}{N} \sum_{t=1}^{T} \frac{1}{T} \sum_{l=1}^{T} \mathbf{E} \left[ \tilde{\xi}_{il} \tilde{\xi}_{lt} e_{il} e_{lt} \right] \\
+ \frac{1}{N} \sum_{t=1}^{T} \frac{1}{T} \sum_{l=1}^{T} \mathbf{E} \left[ \tilde{\xi}_{il} \tilde{\xi}_{lt} e_{il} e_{lt} \right] \\
+ \frac{1}{N} \sum_{t=1}^{T} \frac{1}{T} \sum_{l=1}^{T} \mathbf{E} \left[ \tilde{\xi}_{il} \tilde{\xi}_{lt} e_{il} e_{lt} - \mathbf{E} \left( \tilde{\xi}_{il} \tilde{\xi}_{lt} e_{il} e_{lt} \right) \right] \\
+ \frac{1}{N} \sum_{t=1}^{T} \frac{1}{T} \sum_{l=1}^{T} \mathbf{E} \left[ \tilde{\xi}_{il} \tilde{\xi}_{lt} e_{il} e_{lt} - \mathbf{E} \left( \tilde{\xi}_{il} \tilde{\xi}_{lt} e_{il} e_{lt} \right) \right] \\
+ \frac{1}{N} \sum_{t=1}^{T} \frac{1}{T} \sum_{l=1}^{T} \mathbf{E} \left[ \tilde{\xi}_{il} \tilde{\xi}_{lt} e_{il} e_{lt} - \mathbf{E} \left( \tilde{\xi}_{il} \tilde{\xi}_{lt} e_{il} e_{lt} \right) \right].
\]

For \( j, k = 1, 2 \) note that

\[
\left\| \tilde{\mathbf{V}}^{(p)}_{\xi,j} \hat{\mathbf{H}}^{(p)}_{\xi,kj} \right\| \leq \left\| \frac{\mathbf{F}_{\xi,kj} \mathbf{F}'_{\xi,j}}{\sum_{t=1}^{T} \xi_{jt} e_{jt} N} \right\| \leq \frac{T}{\sum_{t=1}^{T} \xi_{jt} e_{jt} N} \left\| \frac{\mathbf{F}_{\xi,kj} \mathbf{F}'_{\xi,j}}{N} \right\|^{1/2} \left\| \frac{\mathbf{A}_{j}}{N} \right\|^{1/2} = O_p(1)
\]

78
by Assumptions \textbf{1(b)} and \textbf{2(b)}. Since $\left| \hat{\mathbf{v}}^{(p)}_{ij} \right| = O_p(1)$ by Lemma \textbf{10} then $\left| \hat{\mathbf{r}}^{(p)}_{ikj} \right| = O_p(1)$. It follows that

$$
\frac{1}{N} \sum_{i=1}^{N} \left\| \hat{\mathbf{v}}^{(p)}_{ij} \right\| = O_p(1) \quad \text{by Assumption 2(b) and Lemma 11.} 
$$

As for $\hat{\mathbf{r}}^{(p)}_{ikj}$, consider

$$
E \left[ \left( \sum_{i=1}^{N} \hat{\mathbf{v}}^{(p)}_{ij} \hat{\mathbf{r}}^{(p)}_{ikj} \right)^2 \right] \leq \frac{1}{N^2} \sum_{i=1}^{N} \left( \sum_{j=1}^{N} \left\| \hat{\mathbf{v}}^{(p)}_{ij} \right\|^2 \right) \left( \sum_{k=1}^{N} \left\| \hat{\mathbf{r}}^{(p)}_{ikj} \right\|^2 \right).
$$

Starting from $\hat{\sigma}_{\xi,ijkl}$, consider

$$
\hat{\sigma}_{\xi,ijkl} = \frac{1}{N^2} \sum_{i=1}^{N} \left( \sum_{j=1}^{N} \hat{\lambda}_{\xi,ijkl} \sigma_{\xi,ijkl} \right)^2, \quad \hat{\chi}_{\xi,ijkl} = \frac{1}{N^2} \sum_{i=1}^{N} \left( \sum_{j=1}^{N} \hat{\lambda}_{\xi,ijkl} \chi_{\xi,ijkl} \right)^2,
$$

where in general

$$
\hat{\sigma}_{\xi,ijkl} = \frac{1}{N^2} \sum_{i=1}^{N} \left( \sum_{j=1}^{N} \hat{\lambda}_{\xi,ijkl} \sigma_{\xi,ijkl} \right)^2, \quad \hat{\chi}_{\xi,ijkl} = \frac{1}{N^2} \sum_{i=1}^{N} \left( \sum_{j=1}^{N} \hat{\lambda}_{\xi,ijkl} \chi_{\xi,ijkl} \right)^2.
$$

It follows by Assumption 2(b) and Lemma 11 As for $\hat{\chi}_{\xi,ijkl}$,

$$
\hat{\chi}_{\xi,ijkl} = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \hat{\lambda}_{\xi,ijkl} \chi_{\xi,ijkl} \chi_{\xi,ijkl} \chi_{\xi,ijkl} \chi_{\xi,ijkl} 
$$

where

$$
E \left[ \left( \sum_{i=1}^{N} \chi_{\xi,ijkl} \chi_{\xi,ijkl} \right)^2 \right] = E \left( \sum_{i=1}^{N} \chi_{\xi,ijkl} \chi_{\xi,ijkl} \chi_{\xi,ijkl} \chi_{\xi,ijkl} \right) \leq N^2 \cdot \max_{i,j} E \left| \chi_{\xi,ijkl} \right|^4,
$$

and since

$$
E \left| \chi_{\xi,ijkl} \right|^4 = E \left| \frac{1}{T} \sum_{t=1}^{T} \left( \tilde{\xi}_{ijt} \tilde{e}_{it} e_{it} e_{it} \right) - E \left( \tilde{\xi}_{ijt} \tilde{e}_{it} e_{it} \right) \right|^4 
$$

by Assumption 2(b) and Lemma 11. As for $\hat{\chi}_{\xi,ijkl}$,
by Assumption (3.c), and taking into account Assumption (2.b),
\[
\sum_{i=1}^{N} \hat{\chi}_{\xi,i} \leq O_p \left(1\right) \cdot \sqrt{\frac{N^2}{T^2}} = O_p \left(\frac{N}{T}\right),
\]
which implies that
\[
\frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{2} \hat{\chi}_{\xi,j} = \frac{1}{N} O_p \left(\frac{N}{T}\right) = O_p \left(\frac{1}{T}\right). \tag{B.61}
\]
Further,
\[
\hat{\varphi}_{\xi,jk} = \frac{1}{N^2} \left\| \sum_{i=1}^{N} \hat{\chi}^{(p)}_{\xi,kl} \hat{\varphi}_{\xi,kl} \right\|^2
= \frac{1}{N^2} \left\| \sum_{i=1}^{N} \hat{\chi}^{(p)}_{\xi,kl} \left( \frac{T}{T} \sum_{t=1}^{T} \lambda_j f_{jt} \hat{\xi}_{kt | T} e_{it} \right) \right\|^2
= \frac{1}{N^2} \left\| \sum_{i=1}^{N} \hat{\chi}^{(p)}_{\xi,kl} \lambda_j \left( \frac{T}{T} \sum_{t=1}^{T} f_{jt} \hat{\xi}_{kt | T} e_{it} \right) \right\|^2 \tag{B.62}
\leq \frac{1}{T} \left\| \lambda_j \right\|^2 \left( \frac{1}{N} \sum_{i=1}^{N} \left\| \hat{\chi}^{(p)}_{\xi,kl} \right\|^2 \right) \left( \frac{1}{N} \sum_{i=1}^{N} \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} f_{jt} \hat{\xi}_{kt | T} e_{it} \right\|^2 \right)
= O_p \left(\frac{1}{T}\right)
\]
by Assumptions (2.a), (2.b) and (4). Finally,
\[
\hat{\varphi}_{\xi,jk} = \frac{1}{N^2} \left\| \sum_{i=1}^{N} \hat{\chi}^{(p)}_{\xi,kl} \hat{\varphi}_{\xi,kl} \right\|^2
= \frac{1}{N^2} \left\| \sum_{i=1}^{N} \hat{\chi}^{(p)}_{\xi,kl} \left( \frac{T}{T} \sum_{t=1}^{T} \lambda_j f_{jt} \hat{\xi}_{kt | T} e_{it} \right) \right\|^2
= \frac{1}{N^2} \left\| \sum_{i=1}^{N} \hat{\chi}^{(p)}_{\xi,kl} \lambda_j \left( \frac{T}{T} \sum_{t=1}^{T} f_{jt} \hat{\xi}_{kt | T} e_{it} \right) \right\|^2 \tag{B.63}
\leq \frac{1}{T} \left( \frac{1}{N} \sum_{i=1}^{N} \left\| \hat{\chi}^{(p)}_{\xi,kl} \right\| \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} f_{jt} \hat{\xi}_{kt | T} e_{it} \right) \right)^2
\leq \frac{1}{T} \left( \frac{1}{N} \sum_{i=1}^{N} \left\| \hat{\chi}^{(p)}_{\xi,kl} \right\|^2 \right)^{1/2} \left( \frac{1}{N} \sum_{i=1}^{N} \left\| \lambda_j \right\|^2 \right)^{1/2} O_p \left(1\right)
= O_p \left(\frac{1}{T}\right)
\]
by Assumptions (2.b) and (4). From equations (B.59) through (B.63) it follows that
\[
\frac{1}{N} \sum_{i=1}^{N} \left\| \hat{\chi}^{(p)}_{\xi,i} \left[ \hat{\chi}^{(p)}_{\xi,i} - \left( \hat{H}^{(p)}_{\xi,i} \lambda_1 + \hat{H}^{(p)}_{\xi,i} \lambda_2 \right) \right] \right\|^2 = O_p \left(\frac{1}{N}\right) + O_p \left(\frac{1}{T}\right),
\]
and since \( \left\| \hat{\chi}^{(p)}_{\xi,i} \right\| = O_p \left(1\right) \) by Lemma (10) the result stated in the theorem follows.
C Proof of result (44)

Consider

\[
\frac{1}{NT} \log f (X; \hat{q}) - \frac{1}{NT} \log f (X; q(1)) = \frac{1}{NT} \log f (X; \hat{q}) - \frac{1}{NT} \log f (X; q(1)) \\
- E \left[ \frac{1}{NT} \log f (X; \hat{q}) \right] + E \left[ \frac{1}{NT} \log f (X; \hat{q}) \right] \\
+ E \left[ \frac{1}{NT} \log f (X; q(1)) \right] - E \left[ \frac{1}{NT} \log f (X; q(1)) \right] \\
= \left\{ \frac{1}{NT} \log f (X; \hat{q}) - E \left[ \frac{1}{NT} \log f (X; \hat{q}) \right] \right\} \\
- \left\{ \frac{1}{NT} \log f (X; q(1)) - E \left[ \frac{1}{NT} \log f (X; q(1)) \right] \right\} \\
+ \left\{ E \left[ \frac{1}{NT} \log f (X; \hat{q}) \right] - E \left[ \frac{1}{NT} \log f (X; q(1)) \right] \right\}.
\]

Since \( \hat{q} \) is the maximum likelihood estimator, it follows that

\[
\frac{1}{NT} \log f (X; \hat{q}) \geq \frac{1}{NT} \log f (X; q(1))
\]

or, equivalently,

\[
\frac{1}{NT} \log f (X; \hat{q}) - \frac{1}{NT} \log f (X; q(1)) \geq 0,
\]

which implies that

\[
E \left[ \frac{1}{NT} \log f (X; \hat{q}) \right] - E \left[ \frac{1}{NT} \log f (X; q(1)) \right] \geq \left\{ \frac{1}{NT} \log f (X; q(1)) - E \left[ \frac{1}{NT} \log f (X; q(1)) \right] \right\} \\
- \left\{ \frac{1}{NT} \log f (X; \hat{q}) - E \left[ \frac{1}{NT} \log f (X; \hat{q}) \right] \right\} \\
= o_p (1) - \left\{ \frac{1}{NT} \log f (X; \hat{q}) - E \left[ \frac{1}{NT} \log f (X; \hat{q}) \right] \right\},
\]

so that

\[
E \left[ \frac{1}{NT} \log f (X; \hat{q}) \right] - E \left[ \frac{1}{NT} \log f (X; q(1)) \right] \geq o_p (1).
\]

If \( \hat{q} \) was an estimator for \( q^{(3)} \), then

\[
E \left[ \frac{1}{NT} \log f (X; \hat{q}) \right] - E \left[ \frac{1}{NT} \log f (X; q(1)) \right] - E \left[ \frac{1}{NT} \log f (X; q^{(3)}) \right] - E \left[ \frac{1}{NT} \log f (X; q(1)) \right] = o_p (1).
\]

This implies that, for some \( C > 0 \), and taking into account (43),

\[
E \left[ \frac{1}{NT} \log f (X; \hat{q}) \right] - E \left[ \frac{1}{NT} \log f (X; q(1)) \right] = - \left\{ E \left[ \frac{1}{NT} \log f (X; q(1)) \right] - E \left[ \frac{1}{NT} \log f (X; q^{(3)}) \right] \right\} + o_p (1) \\
= -C + o_p (1),
\]

which leads to (44).