Finance and Economics Discussion Series

Federal Reserve Board, Washington, D.C.
ISSN 1936-2854 (Print)
ISSN 2767-3898 (Online)

Quasi Maximum Likelihood Estimation and Inference of Large Approximate Dynamic Factor Models via the EM algorithm

Matteo Barigozzi and Matteo Luciani

2024-086

Please cite this paper as:
Barigozzi, Matteo, and Matteo Luciani (2024). "Quasi Maximum Likelihood Estimation and Inference of Large Approximate Dynamic Factor Models via the EM algorithm," Finance and Economics Discussion Series 2024-086. Washington: Board of Governors of the Federal Reserve System, https://doi.org/10.17016/FEDS.2024.086.

Quasi Maximum Likelihood Estimation and Inference of Large Approximate Dynamic Factor Models via the EM algorithm

Matteo Barigozzi
Università di Bologna
matteo.barigozzi@unibo.it

Matteo Luciani
Federal Reserve Board
matteo.luciani@frb.gov

This version: October 23, 2024
First draft: November 8, 2017*

Abstract

We study estimation of large Dynamic Factor models implemented through the Expectation Maximization (EM) algorithm, jointly with the Kalman smoother. We prove that as both the cross-sectional dimension, n, and the sample size, T, diverge to infinity: (i) the estimated loadings are $\sqrt{T}$-consistent, asymptotically normal and equivalent to their Quasi Maximum Likelihood estimates; (ii) the estimated factors are $\sqrt{n}$-consistent, asymptotically normal and equivalent to their Weighted Least Squares estimates. Moreover, the estimated loadings are asymptotically as efficient as those obtained by Principal Components analysis, while the estimated factors are more efficient if the idiosyncratic covariance is sparse enough.

Keywords: Approximate Dynamic Factor Model; Expectation Maximization Algorithm; Kalman Smoother; Quasi Maximum Likelihood.

1 Introduction

Factor analysis can be considered a pioneering technique in unsupervised statistical learning (Ghahramani and Hinton, 1996). It originally gained popularity in the early decades of the twentieth century as a dimension-reduction technique used in psychometrics (Spearman, 1904). Since then, it has become a classical method used for the statistical analysis of complex datasets in many human, natural, and social sciences (see, e.g., Lawley and Maxwell, 1971, Chapter 1, and references therein). In the last thirty years, factor analysis has seen significant success in financial and macroeconometrics because it allows to analyze and predict economic activity by summarizing large panels of economic time series in a simple and effective way (see, e.g., the survey by Stock and Watson, 2016 and references therein).

An r-factor model is defined by

$$x_{it} = \mu_i + \boldsymbol{\lambda}_i'\mathbf{F}_t + \xi_{it}, \quad i = 1, \dots, n., \quad t = 1, \dots, T, \tag{1}$$

where x_{it} is the observation for the ith cross-section at time t, μ_i is a constant, and $\mathbf{F}_t$ and $\boldsymbol{\lambda}_i$ are r-dimensional latent column vectors of *factors* and factor *loadings*, with $r \ll n$. We call $\boldsymbol{\lambda}_i'\mathbf{F}_t$ the *common* component and

*A previous version of some results in this paper appeared in the paper "Common factors, trends, and cycles in large datasets", available at arXiv:1709.01445 or at FEDS.2017.111.

M. Barigozzi gratefully acknowledges financial support from MIUR (PRIN2020, Grant 2020N9YFFE).

Disclaimer: the views expressed in this paper are those of the authors and do not necessarily reflect the views and policies of the Board of Governors or the Federal Reserve System.

We thank for helpful comments: Majid Al-Sadoon, Leopoldo Catania, Giuseppe Cavaliere, Manfred Deistler, Massimo Franchi, Ivan Petrella, Esther Ruiz, Haihan Tang, Lorenzo Trapani, and Luca Trapin.

ξ_{it} the *idiosyncratic* component. Throughout, we consider the standard case in which all $\{x_{it}\}$ are zero-mean weakly stationary processes or are the result of a transformation to stationarity.

Furthermore, in the case of time series, the factors are likely to be autocorrelated. For example, we can assume simple first order autoregressive dynamics:

$$\mathbf{F}_t = \mathbf{A}\mathbf{F}_{t-1} + \mathbf{v}_t, \quad t = 1, \dots, T, \tag{2}$$

with $\mathbf{v}_t$ being an r-dimensional vector of innovations. Likewise, the idiosyncratic components might be autocorrelated. The measurement equation (1) and the state equation (2) form a state-space model, or, equivalently, a Dynamic Factor Model (DFM) (this is a restricted version of the more general model by Forni et al., 2000, where factors can be loaded also with lags). Thanks to its simplicity and empirical success, the DFM is the most common approach to factor analysis of high-dimensional time series.

In large dimensional macroeconomic and financial datasets, the idiosyncratic components are likely to be also cross-correlated. Indeed, although macroeconomic or financial market dynamics are the main drivers of the comovement in these datasets, sectoral and local comovements are non-negligible sources of fluctuations. In the case of correlated idiosyncratic components the factor model is called *approximate* as opposed to an *exact* factor model having uncorrelated idiosyncratic components.

In an exact factor model a small number of variables is enough to estimate the loadings by Quasi Maximum Likelihood (QML), but we cannot consistently estimate the factors (Lawley and Maxwell, 1971). In an approximate factor model, we can disentangle the common and idiosyncratic components only in the extreme case when $n \to \infty$ (Chamberlain and Rothschild, 1983)—in other words, we do not suffer the typical "curse of dimensionality" but rather benefit from the "blessing of dimensionality". In particular, when $n \to \infty$, we can consistently estimate the factors by some form of linear projection onto the estimated loadings. However, QML estimation of the loadings is now unfeasible since, in principle, it requires to also jointly estimate all the T idiosyncratic (auto)covariance matrices, each of size $n \times n$, as well as the T (auto)covariance matrices of the factors. Thus, we need to explore alternative approaches.

There are three main solutions to this problem. The first is Principal Component (PC) analysis, which delivers the optimal non-parametric estimator of a large approximate factor model, see, e.g., Stock and Watson (2002) and Bai (2003). The second is QML estimation based on a mis-specified exact model with no autocorrelations in the idiosyncratic components and the factors, and a diagonal or sparse idiosyncratic covariance matrix, see, e.g., Bai and Li (2016) and Bai and Liao (2016). These two solutions only consider equation (1), thus effectively estimating a static factor model, not a DFM. There is a third solution, which is the focus of this paper, that considers joint estimation of approximate DFM defined in (1)-(2): the Expectation Maximization (EM) algorithm (Quah and Sargent, 1993, Doz et al., 2012).

The EM algorithm is fully parametric and based on an iteration of two steps: (E-step) given the loadings and all the model's parameters, the factors and their second moments are estimated via the Kalman smoother and are used to compute the expected log-likelihood conditional on the observed data; (M-step) the expected log-likelihood is maximized to obtain a new estimate of the loadings and all the model's parameters. These steps are always implemented by considering a mis-specified likelihood with uncorrelated idiosyncratic components, thus making estimation feasible and producing estimators with closed-form expression As such, the EM algorithm should be regarded as an approximation of the QML estimation method because it maximizes a mis-specified likelihood of an exact DFM using an iterative procedure.

This paper focuses on the theoretical properties of the EM and Kalman smoother estimators. In a couple of breakthrough studies, Doz et al. (2011, 2012) provided the first fundamental theoretical treatment of these estimators, which quickly became popular in empirical macroeconomic research. Indeed, this approach allows the user to easily deal with data irregularities and missing values and impose restrictions that reflect any prior knowledge about the data on the model.[1]

We make three main contributions. First, we prove that, as $n, T \to \infty$, the estimator of the loadings obtained via the EM algorithm converges asymptotically to a unique maximum of the likelihood. It is well known that in the Gaussian quasi-likelihood case, the EM algorithm produces approximate QML estimators (Wu, 1983; Balakrishnan et al., 2017). In this paper, we refine this result by showing that, as $n, T \to \infty$, the approximation error not only depends on the number of iterations but also becomes negligible. Moreover, the EM estimator of the loadings is asymptotically equivalent to the unfeasible Ordinary Least Squares estimator we would obtain if we had observed the factors. This result holds for any reasonable, but not necessarily consistent, pre-estimator of the loadings used to initialize the EM algorithm. We derive similar results for all the estimated parameters, namely, the idiosyncratic variances, the VAR coefficients, and the covariance matrix of the VAR residuals in equation (2).

Second, we prove that, as $n, T \to \infty$, the estimator of the factors obtained via the Kalman smoother computed using the parameters estimated via the EM algorithm is equivalent to the unfeasible Weighted Least Squares estimator we would obtain if we had observed the loadings and idiosyncratic variances. As a by-product of this result, we also show that the Kalman smoother and filter estimators are asymptotically equivalent.

Third, we show that the EM estimator of the loadings is asymptotically equivalent to the PC estimator; hence, it has the same consistency rate ($\min(n, \sqrt{T})$), is asymptotically normal and equally efficient. Likewise, the Kalman smoother has the same consistency rate ($\min(\sqrt{n}, T)$) as the PC estimator, it is asymptotically normal, and if the idiosyncratic covariance matrix is sufficiently sparse, it is more efficient than the PC estimator.

This paper is the first to fully characterize the asymptotic properties of the EM and Kalman smoother estimators. Other papers provide results that are close to ours, using more restrictive approaches or deriving only partial asymptotic results. Bai and Li (2016) considered QML estimation of the loadings for the static factor model (1) only and did not study the convergence of the employed maximization algorithm. Doz et al. (2011) considered the Kalman smoother obtained using the PC estimator of the loadings but did not derive its asymptotic distribution and obtained a slower consistency rate. Last, Doz et al. (2012) proved consistency of the Kalman smoother obtained from the EM algorithm but derived a slower rate and did not prove its asymptotic normality, nor did they prove consistency of the EM estimator of the loadings.

Our results lay the theoretical foundations for the wide empirical success of the EM algorithm for estimating large dimensional DFMs (see the next section for a list of applications) and answer two long-standing critiques. First, by showing the equivalence of the consistency rates of the EM and PC estimators, we reverse the belief that PC is a superior approach. Second, by providing the asymptotic distributions, we answer the call by Tanner and Wong (1987) and Geweke (1993), who advocated a Bayesian approach based on the Gibbs sampler because the EM algorithm provides only point estimates.

The paper is organized as follows. In Section 2, we briefly review the main applications of the EM algorithm and KS in factor analysis and alternative methods proposed to estimate DFM. In Section 3, we describe the estimation and give a guide for implementing it. All assumptions are in Section 4. The asymptotic results are in Section 5. In Section 6, we discuss efficiency of the EM estimator and the KS and compare them with the PC estimator. In Section 7, we propose estimators of the asymptotic covariance matrices. In Section 8, we present an extensive MonteCarlo study, and in Section 9, we apply the EM algorithm to US macroeconomic data. Section 10 concludes. The proofs of all theoretical results are in the Appendix.

2 Related literature

The EM approach is arguably the most popular for conducting QML estimation of high-dimensional DFMs. This approach dates back to the 1970s when it was introduced in a low-dimensional setting by, e.g., Sargent and Sims

[1] PC analysis in presence of missing data and parameter constraints has been studied by, e.g., Bai and Ng (2021), Fan et al. (2022), Xiong and Pelger (2023), among others.

(1977), Shumway and Stoffer (1982), Watson and Engle (1983), and Harvey and Peters (1990), while its use in a high-dimensional setting was first suggested by Quah and Sargent (1993) and then formalized by Doz et al. (2012).

The EM approach for high-dimensional DFMs has been extensively employed by empirical macroeconomic researchers, particularly those in central banks. Its most successful applications include (see also the survey by Poncela et al., 2021): (i) counterfactual analysis (Harvey, 1996; Giannone et al., 2006, 2019); (ii) conditional forecasts (Bańbura et al., 2015); (iii) nowcasting (Giannone et al., 2008; Bańbura et al., 2011; Bańbura et al., 2013; Kim and Swanson, 2018; Cascaldi-Garcia et al., 2023); (iv) dealing with data irregularly spaced (Mariano and Murasawa, 2003; Jungbacker et al., 2011; Bańbura and Modugno, 2014; Marcellino and Sivec, 2016); (v) imposing constraints on the loadings to account for smooth cross-sectional dependence in the case of ordered units (Koopman and van der Wel, 2013; Jungbacker et al., 2014) or a block-specific factor structure (Coroneo et al., 2016; Altavilla et al., 2017); (vi) building indicators of economic activity (Reis and Watson, 2010; Barigozzi and Luciani, 2023; Ng and Scanlan, 2024; Ahn and Luciani, 2024); (vi) impulse response analysis (Juvenal and Petrella, 2015; Luciani, 2015); (vii) modeling international stock market dynamics (Linton et al., 2022); (viii) extract trends from micro-panels (Barigozzi et al., 2024).

In addition to the EM approach, the literature has proposed several multi-step approaches to estimate the DFM in (1)-(2). (a) Bai and Ng (2007) and Forni et al. (2009) employ PC analysis followed by VAR estimation. (b) Doz et al. (2011) employ PC analysis followed by VAR estimation and the Kalman smoother. (c) Ng et al. (2015) consider QML estimation of the loadings based on a matrix decomposition technique that allows the use of the Newton-Raphson method, then followed by the Kalman filter. (d) Bai and Li (2016) consider QML estimation of the loadings based on the EM algorithm by Rubin and Thayer (1982), followed by Weighted Least Squares to estimate the factors, VAR estimation, and, finally, the Kalman smoother. (e) Jungbacker and Koopman (2015) consider QML estimation of the loadings for a low-dimensional projection of the data based on the prediction error likelihood obtained from the Kalman filter. (f) Lin and Michailidis (2020) propose an alternating minimization algorithm based on a penalized loss accounting for cross-autocorrelation in the idiosyncratic components, followed by Generalized Least Squares to estimate the factors. (g) Kapetanios and Marcellino (2009) consider estimation using sub-space methods. (h) Mosley et al. (2024) propose a modified version of the EM algorithm used in this paper where the M-step allows for sparsity in the loadings.

Approaches (a)-(d) consider estimation of the loadings, either via PC or QML, based only on (1), while estimation of (2) is in a second step. Approaches (e)-(h) consider joint estimation of (1)-(2). However, none of these fully develops an asymptotic theory for the proposed estimators. Moreover, (f) and (h) do not study the convergence of their numerical algorithms.

Last, an alternative to the EM algorithm is represented by Bayesian estimation of large DFMs using Gibbs sampling, see, e.g., by Kose et al. (2003), Luciani and Ricci (2014), Bai and Wang (2015), D'Agostino et al. (2016), and Koopman and Mesters (2017), among many others.

To conclude, classical references for QML estimation of an exact factor model with no autocorrelations are, e.g., Anderson and Rubin (1956) and Amemiya et al. (1987), while Tipping and Bishop (1999) suggest to simplify the maximization problem by considering a mis-specified likelihood with homoskedastic idiosyncratic components. Bai and Li (2012) extend classical QML estimation to the high-dimensional case. Alternatively, Stock and Watson (1989) combine QML estimation with the Kalman filter by using the prediction error likelihood. A review of these methods is in Barigozzi (2024).

Notation

An $m \times m$ identity matrix is denoted as $\mathbf{I}_m$. Vectors are always considered as one-column matrices. An m dimensional vector of ones is denoted as $\boldsymbol{\iota}_m$. An m dimensional vector of zeros is denoted as $\mathbf{0}_m$, an $m \times p$ matrix of zeros is denoted as $\mathbf{0}_{m\times p}$.

The generic (i,j) entry of a matrix $\mathbf{A}$ is denoted as $[\mathbf{A}]_{ij}$. Unless otherwise specified, we denote as $\nu^{(k)}(\mathbf{A})$ the k-th largest eigenvalue of a generic squared matrix $\mathbf{A}$. The spectral norm for a real $p\times m$ matrix $\mathbf{A}$ is defined by $\|\mathbf{A}\| = (\nu^{(1)}(\mathbf{A}\mathbf{A}'))^{1/2}$. The Frobenius norm is defined by $\|\mathbf{A}\|_F = (\mathrm{tr}(\mathbf{A}\mathbf{A}'))^{1/2}$. For a generic p-dimensional vector $\boldsymbol{v} = (v_1 \cdots v_p)'$, we consider the norms: $\|\boldsymbol{v}\| = (\sum_{j=1}^p v_j^2)^{1/2}$, and $\|\boldsymbol{v}\|_{\max} = \max_{j=1,\dots,p} |v_j|$.

The indicator function on the event A is denoted as $\mathbb{I}(A)$, i.e, $\mathbb{I}(A) = 1$ if A is true and 0 otherwise.

All random variables (scalars, vectors and matrices) are assumed to belong to $L_2((\Omega,\mathcal{F},\mathrm{P}))$, where $(\Omega,\mathcal{F},\mathrm{P})$ is a common probability space. For a generic p-dimensional process $\{\mathbf{y}_t\}$ we adopt the following definitions. (i) Expectations are computed using the true values of the underlying distribution unless otherwise indicated, so we write $\mathbb{E}[\mathbf{y}_t] = \int_{\mathbb{R}^p} \boldsymbol{y}\mathrm{d}F_{\mathbf{y}_t}(\boldsymbol{y},\boldsymbol{\varphi}_n)$, where $F_{\mathbf{y}_t}(\boldsymbol{y},\boldsymbol{\varphi}_n)$ is the cumulative distribution function of $\mathbf{y}_t$ computed when using as parameters the true ones $\boldsymbol{\varphi}_n$, and we write $\mathbb{E}_{\widehat{\varphi}_n}[\mathbf{y}_t]$ when using as parameters $\widehat{\boldsymbol{\varphi}}_n$ to compute the cdf. (ii) For any $t\in\mathbb{Z}$, given the pT-dimensional vector $\boldsymbol{Y}_t = (\mathbf{y}_1' \cdots \mathbf{y}_t')'$ we denote conditioning on $\boldsymbol{Y}_t$ as an abbreviation for conditioning on the σ-algebra generated by $\{\mathbf{y}_{t-k}, k\ge 0\}$.

Limits are always taken as $\min(n,T)\to\infty$ unless otherwise specified. We adopt the Landau $O(\cdot)$ and $o(\cdot)$ notation and the "in probability" $O_p(\cdot)$ and $o_p(\cdot)$ analogues. We denote convergence in probability and in distribution by $\stackrel{p}{\to}$ and $\stackrel{d}{\to}$, respectively.

For all quantities having a dimension growing with n and/or T, we highlight such dependence. The true scalars, vectors or matrices are denoted as, e.g., σ_i^2, $\mathbf{\Lambda}_n$, $\boldsymbol{\phi}_n$, $\boldsymbol{\theta}$, $\boldsymbol{F}_T$,. The corresponding scalars, vectors or matrices containing generic values of parameters are underlined, so: $\underline{\sigma}_i^2$, $\underline{\mathbf{\Lambda}}_n$, $\underline{\boldsymbol{\phi}}_n$, $\underline{\boldsymbol{\theta}}$, $\underline{\boldsymbol{F}}_T$.

For a generic process $\{\mathbf{y}_t\}$ and any $T\in\mathbb{N}$, letting $\mathcal{F}_{-\infty}^0$ and $\mathcal{F}_T^\infty$ be the σ-algebras generated by $\{\mathbf{y}_t,\, t\le 0\}$ and $\{\mathbf{y}_t,\, t\ge T\}$, respectively, we define the strong mixing coefficients of $\{\mathbf{y}_t\}$ as $\alpha_y(T) = \sup_{A\in\mathcal{F}_{-\infty}^0,\, B\in\mathcal{F}_T^\infty} |\mathrm{P}(A)\mathrm{P}(B) - \mathrm{P}(AB)|$.

3 Estimation via the EM algorithm

In this section, we present the EM algorithm described by Shumway and Stoffer (1982) and implemented in the codes by Doz et al. (2012), which are available from theirs or ours webpages.[2] This is the approach typically followed by applied researchers.

Let us assume to observe an n-dimensional stochastic process $\mathbf{x}_{nt} = (x_{1t}\cdots x_{nt})'$ over T periods. The DFM (1)-(2) reads:

$$\mathbf{x}_{nt} = \boldsymbol{\mu}_n + \mathbf{\Lambda}_n\mathbf{F}_t + \boldsymbol{\xi}_{nt}, \tag{3}$$

$$\mathbf{F}_t = \sum_{h=1}^{p_F} \mathbf{A}_h\mathbf{F}_{t-h} + \mathbf{v}_t, \tag{4}$$

where $\mathbf{\Lambda}_n = (\boldsymbol{\lambda}_1\cdots\boldsymbol{\lambda}_n)'$ is the $n\times r$ matrix of factor loadings, $\mathbf{v}_t$ is an r-dimensional vector of factor innovations with covariance $\mathbf{\Gamma}^v = \mathbb{E}[\mathbf{v}_t\mathbf{v}_t']$, and p_F is a finite integer such that $p_F\ge 1$. Throughout, we assume that the r-dimensional process of factors, $\{\mathbf{F}_t\}$, and the n-dimensional process of idiosyncratic components $\{\boldsymbol{\xi}_{nt}\}$ are zero-mean covariance stationary processes, so that $\{\mathbf{x}_{nt}\}$ is also covariance stationary and $\mathbb{E}[\mathbf{x}_{nt}] = \boldsymbol{\mu}_n$, with $\boldsymbol{\mu}_n = (\mu_1\cdots\mu_n)'$.

Let us introduce the following notation. Define the nT-dimensional vectors $\boldsymbol{X}_{nT} = (\mathbf{x}_{n1}'\cdots\mathbf{x}_{nT}')'$ and $\mathbf{\Xi}_{nT} = (\boldsymbol{\xi}_{n1}'\cdots\boldsymbol{\xi}_{nT}')'$, and the rT-dimensional vector $\boldsymbol{F}_T = (\mathbf{F}_1'\cdots\mathbf{F}_T')'$. Moreover, let $\mathbf{\Sigma}_n^\xi$ be the diagonal matrix containing the n diagonal terms of $\mathbb{E}[\boldsymbol{\xi}_{nt}\boldsymbol{\xi}_{nt}']$, which we denote as σ_i^2, $i=1,\dots,n$, and $\mathbf{\Omega}_T^F = \mathbb{E}[\boldsymbol{F}_T\boldsymbol{F}_T']$.

In principle, to achieve QML, we should estimate (i) nr loadings, $(ii)\simeq n^2T^2$ elements of the covariance matrix of $\mathbf{\Xi}_{nT}$, $(iii)\simeq r^2T^2$ elements of the covariance matrix of $\boldsymbol{F}_T$, and (iv) the n constants in $\boldsymbol{\mu}_n$. Clearly the QML estimator of $\boldsymbol{\mu}_n$ is the sample mean $\bar{\mathbf{x}}_n = T^{-1}\sum_{t=1}^T \mathbf{x}_{nt}$ and we can then work with centered data:

[2]See the replication codes available at: https://sites.uw.edu/dgiannon/domenico-giannone-s-homepage/, or http://www.barigozzi.eu/codes.html, or https://sites.google.com/site/lucianimatteo/matlab-codes

$\mathbf{x}_{nt} - \bar{\mathbf{x}}_n$ (Bai and Li, 2012, p.440). However, all other parameters depend on each other, so they must be estimated jointly—see, e.g., the first-order conditions in Bai and Li (2012, 2016) when the autocorrelation of the factors is not modeled. This is an unfeasible task since we have only nT data points; therefore, we need some regularization to reduce the number of parameters to estimate.

To this end, first, we adopt the extreme form of regularization possible for the full-covariance matrix of $\boldsymbol{\Xi}_{nT}$ by replacing it with the diagonal matrix $\mathbf{I}_T \otimes \boldsymbol{\Sigma}_n^{\xi}$. Second, we assume that the parametric model (4) describes the dynamics of the factors—thus, we write $\boldsymbol{\Omega}_T^F \equiv \boldsymbol{\Omega}_T^F(\boldsymbol{\mathcal{A}}, \boldsymbol{\Gamma}^v)$, with $\boldsymbol{\mathcal{A}} = (\mathbf{A}_1 \cdots \mathbf{A}_{p_F})$, in order to highlight its dependence on the VAR parameters. Thanks to these assumptions, we reduce the number of unknown parameters that need to be estimated to $Q_n = nr + n + r^2 p_F + r(r+1)/2$, the elements of the vector $\boldsymbol{\varphi}_n = (\boldsymbol{\phi}_n' \ \boldsymbol{\theta}')'$, where $\boldsymbol{\phi}_n = (\text{vec}(\boldsymbol{\Lambda}_n)' \ \sigma_1^2 \cdots \sigma_n^2)'$ and $\boldsymbol{\theta} = (\text{vec}(\boldsymbol{\mathcal{A}})' \ \text{vech}(\boldsymbol{\Gamma}^v)')'$.

Our starting point is then the following log-likelihood, computed in the generic values of the parameters, $\underline{\boldsymbol{\phi}}_n$ and $\underline{\boldsymbol{\theta}}$:

$$\begin{aligned}\ell(\boldsymbol{X}_{nT}; \underline{\boldsymbol{\phi}}_n, \underline{\boldsymbol{\theta}}) = &-\frac{1}{2} \log\det \Big(\{\mathbf{I}_T \otimes \underline{\boldsymbol{\Lambda}}_n\} \boldsymbol{\Omega}_T^F(\underline{\boldsymbol{\mathcal{A}}}, \underline{\boldsymbol{\Gamma}}^v) \{\mathbf{I}_T \otimes \underline{\boldsymbol{\Lambda}}_n'\} + \{\mathbf{I}_T \otimes \underline{\boldsymbol{\Sigma}}_n^{\xi}\} \Big) \\ &- \frac{1}{2} \Big[(\boldsymbol{X}_{nT} - \boldsymbol{\iota}_T \otimes \bar{\mathbf{x}}_n)' \Big(\{\mathbf{I}_T \otimes \underline{\boldsymbol{\Lambda}}_n\} \boldsymbol{\Omega}_T^F(\underline{\boldsymbol{\mathcal{A}}}, \underline{\boldsymbol{\Gamma}}^v) \{\mathbf{I}_T \otimes \underline{\boldsymbol{\Lambda}}_n\}' + \{\mathbf{I}_T \otimes \underline{\boldsymbol{\Sigma}}_n^{\xi}\} \Big)^{-1} (\boldsymbol{X}_{nT} - \boldsymbol{\iota}_T \otimes \bar{\mathbf{x}}_n) \Big],\end{aligned} \tag{5}$$

where we removed the constant terms to simplify the notation. Since the assumptions in Section 4 allow for correlations among idiosyncratic components, the expression in (5) is a mis-specified or quasi log-likelihood. Such mis-specification is appealing because it coincides with the classical factor analysis under the exact factor structure, and is standard in DFM estimation (Doz et al., 2012; Bai and Li, 2016). Thus, the maximization of (5) is QML estimation rather than ML estimation. As long as the idiosyncratic components are weakly correlated, the mis-specification introduced by maximizing (5) has no effect on consistency but only on efficiency of the estimators (see Section 6). Hereafter, we denote the vector of QML estimators, which are the maximizers of (5), as $\widehat{\boldsymbol{\varphi}}_n^* = (\widehat{\boldsymbol{\phi}}_n^{*\prime} \ \widehat{\boldsymbol{\theta}}^{*\prime})'$.

Despite reducing the number of parameters to be estimated by introducing mis-specifications, direct maximization of (5) is still unfeasible because $\boldsymbol{\Omega}_T^F$ is a full matrix, and to estimate its entries, we need to estimate the factors as well. Therefore, we still face a curse of dimensionality problem because we need to jointly estimate the rT values of the factors and the Q_n parameters using just the nT observations of $\boldsymbol{X}_{nT}$.

There are three solutions to this problem. The first consists of rewriting the log-likelihood (5) using its prediction error formulation, where the prediction errors and their covariance are obtained via the Kalman filter (see, e.g., Harvey, 1990, Chapter 3.4 or Durbin and Koopman, 2012, Chapter 7). This is the standard practice in low-dimensional DFMs (Stock and Watson, 1989). However, since there is no closed form solution for the QML estimator of the parameters obtained in this way, numerical maximization is required and this approach becomes quickly unfeasible even for moderate values of n. In high dimensions, Jungbacker and Koopman (2015) propose to follow this approach subject to a preliminary step in which the data are projected onto a lower dimensional space but do not derive the theoretical properties of this estimator. In particular, it is not clear what are the effects of this first step on the asymptotic properties of the final estimator.

The second solution, proposed by Bai and Li (2016), further mis-specify the model by treating the factors as if they were serially uncorrelated. Thus, by imposing the standard identifying constraint $\mathbb{E}[\mathbf{F}_t \mathbf{F}_t'] = \mathbf{I}_r$ and by replacing in (5) the full-matrix $\boldsymbol{\Omega}_T^F(\underline{\boldsymbol{\mathcal{A}}}, \underline{\boldsymbol{\Gamma}}^v)$ with just $\mathbf{I}_T \otimes \mathbf{I}_r$, the log-likelihood is considerably simplified, and its maximization becomes feasible. However, because it does not exist a closed form solution, we still need a numerical approach to estimate the model, e.g., the iterative algorithm proposed by Rubin and Thayer (1982) and reintroduced by Bai and Li (2012). To our knowledge, the convergence of the Rubin and Thayer (1982) algorithm to the QML estimator has never been formally proved.

A third solution consists of computing an approximation of the QML estimator using the EM algorithm, as formalized by Doz et al. (2012). We consider this approach in this paper, and we refer to Table 1 for its

Table 1

EXPECTATION MAXIMIZATION ALGORITHM

Input:
n dimensional vector of data $\{\mathbf{x}_{nt}\}_{t=1}^T$
number of common factors r
VAR order for the factors p_F
maximum number of iterations $k_{\max}$
threshold for convergence ε
initial estimators $\{\widehat{\boldsymbol{\lambda}}_i^{(0)}\}_{i=1}^n$, $\{\widehat{\sigma}_i^{2(0)}\}_{i=1}^n$, $\{\widehat{\mathbf{A}}_j^{(0)}\}_{j=1}^{p_F}$, $\widehat{\boldsymbol{\Gamma}}^{v(0)}$, see Appendix A.1

Ouput:
estimated loadings $\{\widehat{\boldsymbol{\lambda}}_i\}_{i=1}^n$
estimated factors $\{\widehat{\mathbf{F}}_t\}_{t=1}^T$
for $k=0$ **to** $k=k_{\max}$ **do**
{E-STEP}
run Kalman Smoother with $\{\mathbf{x}_{nt}\}_{t=1}^T$, $\{\widehat{\boldsymbol{\lambda}}_i^{(k)}\}_{i=1}^n$, $\{\widehat{\sigma}_i^{2(k)}\}_{i=1}^n$, $\{\widehat{\mathbf{A}}_j^{(k)}\}_{j=1}^{p_F}$, $\widehat{\boldsymbol{\Gamma}}^{v(k)}$
$\to \{\mathbf{F}_{t|T}^{(k)}\}_{t=1}^T, \{\mathbf{P}_{t|T}^{(k)}\}_{t=1}^T, \{\{\mathbf{C}_{t,t-j|T}^{(k)}\}_{t=j+1}^T\}_{j=1}^{p_F}$, see Appendix A.2
compute expected log-likelihood and sufficient statistics
$\to \mathcal{Q}\left(\{\underline{\boldsymbol{\lambda}}_i\}_{i=1}^n, \{\underline{\sigma}_i^2\}_{i=1}^n, \{\underline{\mathbf{A}}_j\}_{j=1}^{p_F}, \underline{\boldsymbol{\Gamma}}^v; \{\widehat{\boldsymbol{\lambda}}_i^{(k)}\}_{i=1}^n, \{\widehat{\sigma}_i^{2(k)}\}_{i=1}^n, \{\widehat{\mathbf{A}}_j^{(k)}\}_{j=1}^{p_F}, \widehat{\boldsymbol{\Gamma}}^{v(k)}\right)$,
$\to \left\{\left(\sum_{t=1}^T \mathbf{F}_{t|T}^{(k)} x_{it}\right)\right\}_{i=1}^n, \left(\sum_{t=1}^T \mathbf{F}_{t|T}^{(k)}\mathbf{F}_{t|T}^{(k)\prime} + \mathbf{P}_{t|T}^{(k)}\right), \left\{\left(\sum_{t=j+1}^T \mathbf{F}_{t|T}^{(k)}\mathbf{F}_{t-j|T}^{(k)\prime} + \mathbf{C}_{t,t-j|T}^{(k)}\right)\right\}_{j=1}^{p_F}$,
see (8), (9), (10), (11), (12)
{M-STEP}
maximize expected log-likelihood
$\to \{\widehat{\boldsymbol{\lambda}}_i^{(k+1)}\}_{i=1}^n$, $\{\widehat{\sigma}_i^{2(k+1)}\}_{i=1}^n$, $\{\widehat{\mathbf{A}}_j^{(k+1)}\}_{j=1}^{p_F}$, $\widehat{\boldsymbol{\Gamma}}^{v(k+1)}$,
see (13), (14), (15), (16)
{CONVERGENCE}
if $k < k_{\max}$ AND $\Delta\ell_k < \varepsilon$, see (A.14) **then**
$k^* \leftarrow k$
run Kalman Smoother with $\{\mathbf{x}_{nt}\}_{t=1}^T$, $\{\widehat{\boldsymbol{\lambda}}_i^{(k^*+1)}\}_{i=1}^n$, $\{\widehat{\sigma}_i^{2(k^*+1)}\}_{i=1}^n$, $\{\widehat{\mathbf{A}}_j^{(k^*+1)}\}_{j=1}^{p_F}$, $\widehat{\boldsymbol{\Gamma}}^{v(k^*+1)}$
$\to \{\mathbf{F}_{t|T}^{(k^*+1)}\}_{t=1}^T$, see Appendix A.2
$\widehat{\boldsymbol{\lambda}}_i \leftarrow \widehat{\boldsymbol{\lambda}}_i^{(k^*+1)}$, for all $i=1,\dots,n$
$\widehat{\mathbf{F}}_t \leftarrow \mathbf{F}_{t|T}^{(k^*+1)}$, for all $t=1,\dots,T$
return $\{\widehat{\boldsymbol{\lambda}}_i\}_{i=1}^n$, $\{\widehat{\mathbf{F}}_t\}_{t=1}^T$
break
end if
if $k = k_{\max}$ **then**
print "algorithm did not converge"
break
end if
$k+1 \leftarrow k$
end for

implementation.

The EM algorithm is an iterative procedure which allows for QML estimation in presence of missing data (Dempster et al., 1977). In a nutshell, consider a given iteration $k \geq 0$ and assume to have an estimate of the parameters $\widehat{\boldsymbol{\varphi}}_n^{(k)}$. By taking expectations of the log-likelihood (5) with respect to the conditional distribution of $\boldsymbol{F}_T$ given $\boldsymbol{X}_{nT}$ and computed using $\widehat{\boldsymbol{\varphi}}_n^{(k)}$, we get:

$$\begin{aligned}\ell(\boldsymbol{X}_{nT};\underline{\boldsymbol{\varphi}}_n) &= \mathbb{E}_{\widehat{\boldsymbol{\varphi}}_n^{(k)}}[\ell(\boldsymbol{X}_{nT},\boldsymbol{F}_T;\underline{\boldsymbol{\varphi}}_n)|\boldsymbol{X}_{nT}] - \mathbb{E}_{\widehat{\boldsymbol{\varphi}}_n^{(k)}}[\ell(\boldsymbol{F}_T|\boldsymbol{X}_{nT};\underline{\boldsymbol{\varphi}}_n)|\boldsymbol{X}_{nT}]\\ &= \mathcal{Q}(\underline{\boldsymbol{\varphi}}_n,\widehat{\boldsymbol{\varphi}}_n^{(k)}) - \mathcal{H}(\underline{\boldsymbol{\varphi}}_n,\widehat{\boldsymbol{\varphi}}_n^{(k)}),\ \text{say.}\end{aligned} \tag{6}$$

A maximum of the log-likelihood is then a maximum of the right hand side of (6). Now, by definition of Kullback-Leibler divergence, for any $k \geq 0$ it holds that

$$\mathcal{H}(\widehat{\boldsymbol{\varphi}}_n^{(k+1)};\widehat{\boldsymbol{\varphi}}_n^{(k)}) \leq \mathcal{H}(\widehat{\boldsymbol{\varphi}}_n^{(k)};\widehat{\boldsymbol{\varphi}}_n^{(k)}), \tag{7}$$

i.e., $\mathcal{H}(\underline{\boldsymbol{\varphi}}_n, \widehat{\boldsymbol{\varphi}}_n^{(k)})$ is maximum at $\underline{\boldsymbol{\varphi}}_n = \widehat{\boldsymbol{\varphi}}_n^{(k)}$. It is then enough to look for the maximum only of the expected full-information log-likelihood $\mathcal{Q}(\underline{\boldsymbol{\varphi}}_n, \widehat{\boldsymbol{\varphi}}_n^{(k)})$. This is accomplished in two steps: in the first step, for given estimated parameters $\widehat{\boldsymbol{\varphi}}_n^{(k)}$, we compute $\mathcal{Q}(\underline{\boldsymbol{\varphi}}_n, \widehat{\boldsymbol{\varphi}}_n^{(k)})$ using an estimate of the factors with their associated MSE; in the second step, for a given estimate of the factors, we maximize such log-likelihood to compute a new estimate of the parameters $\widehat{\boldsymbol{\varphi}}_n^{(k+1)}$. As shown below, this approach solves the curse of dimensionality problem, and its computational burden is minimal because all estimates have an explicit expression. Below we detail the main features of the two steps.

3.1 E-step and Kalman smoother

We obtain an initial estimate of the loadings and the idiosyncratic variances using the PC estimator, and of the VAR parameters by fitting a VAR on the PC estimator of the factors (see Appendix A.1). Then, for any iteration $k \geq 0$, in the E-step, we compute the expected full-information log-likelihood, which we can decomposed as:

$$\mathcal{Q}(\underline{\boldsymbol{\varphi}}_n, \widehat{\boldsymbol{\varphi}}_n^{(k)}) = \mathbb{E}_{\widehat{\varphi}_n^{(k)}}[\ell(\boldsymbol{X}_{nT}|\boldsymbol{F}_T; \underline{\boldsymbol{\phi}}_n)|\boldsymbol{X}_{nT}] + \mathbb{E}_{\widehat{\varphi}_n^{(k)}}[\ell(\boldsymbol{F}_T; \underline{\boldsymbol{\theta}})|\boldsymbol{X}_{nT}]. \tag{8}$$

Consistently with the mis-specified log-likelihood (5), the first term on the right-hand side of (8) is:

$$\begin{aligned}\ell(\boldsymbol{X}_{nT}|\boldsymbol{F}_T; \underline{\boldsymbol{\phi}}_n) &= -\frac{T}{2}\log\det(\underline{\boldsymbol{\Sigma}}_n^{\xi}) - \frac{1}{2}\sum_{t=1}^{T}(\mathbf{x}_{nt} - \bar{\mathbf{x}}_n - \underline{\boldsymbol{\Lambda}}_n\mathbf{F}_t)'(\underline{\boldsymbol{\Sigma}}_n^{\xi})^{-1}(\mathbf{x}_{nt} - \bar{\mathbf{x}}_n - \underline{\boldsymbol{\Lambda}}_n\mathbf{F}_t)\\ &= \sum_{i=1}^{n}\left\{-\frac{T}{2}\log(\underline{\sigma}_i^2) - \frac{1}{2}\sum_{t=1}^{T}\frac{(x_{it} - \bar{x}_i - \underline{\boldsymbol{\lambda}}_i'\mathbf{F}_t)^2}{\underline{\sigma}_i^2}\right\},\end{aligned} \tag{9}$$

which depends only on $\underline{\boldsymbol{\phi}}_n$ and is well defined as long as all the idiosyncratic components have finite positive variances.

As for the second term on the right-hand side of (8), we assume that $\mathbf{F}_t = \mathbf{0}_r$ for $t \leq 0$ and consider $\ell(\boldsymbol{F}_T; \underline{\boldsymbol{\theta}}) = \sum_{t=1}^{T}\ell(\mathbf{F}_t|\mathbf{F}_{t-1}, \ldots, \mathbf{F}_{t-p_F}; \underline{\boldsymbol{\theta}})$, which depends only on $\underline{\boldsymbol{\theta}}$. Then we have

$$\ell(\boldsymbol{F}_T; \underline{\boldsymbol{\theta}}) = -\frac{T}{2}\log\det(\underline{\boldsymbol{\Gamma}}^v) - \frac{1}{2}\sum_{t=1}^{T}\left(\mathbf{F}_t - \sum_{h=1}^{p_F}\underline{\mathbf{A}}_h\mathbf{F}_{t-h}\right)'(\underline{\boldsymbol{\Gamma}}^v)^{-1}\left(\mathbf{F}_t - \sum_{h=1}^{p_F}\underline{\mathbf{A}}_h\mathbf{F}_{t-h}\right), \tag{10}$$

which is well defined provided that the VAR innovations $\{\mathbf{v}_t\}$ have a finite full-rank covariance matrix.

Given (9) and (10), in order to compute the expected log-likelihood (8) we need to compute the sufficient statistics:

$$\mathbb{E}_{\widehat{\varphi}_n^{(k)}}[\mathbf{F}_t|\boldsymbol{X}_{nT}], \qquad \mathbb{E}_{\widehat{\varphi}_n^{(k)}}[\mathbf{F}_t\mathbf{F}_t'|\boldsymbol{X}_{nT}], \qquad \mathbb{E}_{\widehat{\varphi}_n^{(k)}}[\mathbf{F}_t\mathbf{F}_{t-h}'|\boldsymbol{X}_{nT}], \quad h = 1, \ldots, p_F. \tag{11}$$

Although exact expressions for the quantities in (11) might be hard to compute, we can approximate them by using the output of the Kalman smoother, which gives the linear projection $\mathbf{F}_{t|T}^{(k)} = \text{Proj}_{\widehat{\varphi}_n^{(k)}}[\mathbf{F}_t|\boldsymbol{X}_{nT}]$ and the associated conditional covariance and lag-h autocovariance matrices $\mathbf{P}_{t|T}^{(k)}$ and $\mathbf{C}_{t,t-h|T}^{(k)}$, respectively (see Appendix A.2, for the explicit expressions). This approximation does not affect consistency of our estimators (see Proposition 1).

Hence, in the EM algorithm we let:

$$\begin{aligned}&\mathbb{E}_{\widehat{\varphi}_n^{(k)}}[\mathbf{F}_t|\boldsymbol{X}_{nT}] = \mathbf{F}_{t|T}^{(k)}, \qquad \mathbb{E}_{\widehat{\varphi}_n^{(k)}}[\mathbf{F}_t\mathbf{F}_t'|\boldsymbol{X}_{nT}] = \mathbf{F}_{t|T}^{(k)}\mathbf{F}_{t|T}^{(k)\prime} + \mathbf{P}_{t|T}^{(k)},\\ &\mathbb{E}_{\widehat{\varphi}_n^{(k)}}[\mathbf{F}_t\mathbf{F}_{t-h}'|\boldsymbol{X}_{nT}] = \mathbf{F}_{t|T}^{(k)}\mathbf{F}_{t-h|T}^{(k)\prime} + \mathbf{C}_{t,t-h|T}^{(k)}, \qquad h = 1, \ldots, p_F.\end{aligned} \tag{12}$$

Summing up, in the E-step, we compute rT values of the factors for a given value, $\widehat{\varphi}_n^{(k)}$, of the parameters.

In general, this step is feasible because we have $T(n-r) \gg 0$ degrees of freedom. Moreover, since the Kalman filter and smoother are linear procedures entailing the inversion of a positive definite $r \times r$ matrix, we just have to compute $\simeq T$ recursions.

Remark 1. The Kalman smoother requires first running the forward iterations of the Kalman filter, which in turn requires either inverting the $n \times n$ full covariance matrix of the data or inverting the full $n \times n$ idiosyncratic covariance. This task might be challenging, if not impossible, in a high-dimensional setting. To overcome this problem, we implement the Kalman filter using an estimator of the diagonal matrix $\boldsymbol{\Sigma}_n^\xi$ instead of the full idiosyncratic covariance matrix, ensuring also positive definiteness of the data covariance matrix, and thus making also its inversion feasible. This simplification is consistent with the mis-specified log-likelihood (9) because, to guarantee that (7) holds, we must take expectations with respect to the same distribution as the one used to compute the log-likelihood.

3.2 M-step

In the M-step, we have to maximize (8) with respect to $\underline{\boldsymbol{\varphi}}_n$ to obtain a new estimate of the parameters $\widehat{\boldsymbol{\varphi}}_n^{(k+1)}$. This maximization has a closed form solution for all elements of $\widehat{\boldsymbol{\varphi}}_n^{(k+1)}$. Specifically, at a given iteration $k \geq 0$, by using (12), we obtain the loadings estimators as

$$\widehat{\boldsymbol{\lambda}}_i^{(k+1)} = \left\{ \sum_{t=1}^T \left(\mathbf{F}_{t|T}^{(k)} \mathbf{F}_{t|T}^{(k)\prime} + \mathbf{P}_{t|T}^{(k)} \right) \right\}^{-1} \left(\sum_{t=1}^T \mathbf{F}_{t|T}^{(k)} (x_{it} - \bar{x}_i) \right), \qquad \text{for } i = 1, \dots, n \tag{13}$$

and $\widehat{\boldsymbol{\Lambda}}_n^{(k+1)} = (\widehat{\boldsymbol{\lambda}}_1^{(k+1)} \cdots \widehat{\boldsymbol{\lambda}}_n^{(k+1)})'$. Similarly, the estimator of the idiosyncratic variances is:

$$\widehat{\sigma}_i^{2(k+1)} = \frac{1}{T} \sum_{t=1}^T \left\{ x_{it}^2 + \widehat{\boldsymbol{\lambda}}_i^{(k+1)\prime} \left(\mathbf{F}_{t|T}^{(k)} \mathbf{F}_{t|T}^{(k)\prime} + \mathbf{P}_{t|T}^{(k)} \right) \widehat{\boldsymbol{\lambda}}_i^{(k+1)} - 2 x_{it} \mathbf{F}_{t|T}^{(k)\prime} \widehat{\boldsymbol{\lambda}}_i^{(k+1)} \right\} \qquad \text{for } i = 1, \dots, n. \tag{14}$$

Because we consider a mis-specified log-likelihood, we do not estimate the out-of-diagonal terms of the idiosyncratic covariance matrix $\boldsymbol{\Gamma}_n^\xi$, and we act as if those terms are equal to zero.

For simplicity, let $p_F = 1$ and denote $\mathbf{A} \equiv \mathbf{A}_1$. The estimator of $\mathbf{A}$ is then given by:

$$\widehat{\mathbf{A}}^{(k+1)} = \left\{ \sum_{t=2}^T \left(\mathbf{F}_{t|T}^{(k)} \mathbf{F}_{t-1|T}^{(k)\prime} + \mathbf{C}_{t,t-1|T}^{(k)} \right) \right\} \left\{ \sum_{t=2}^T \left(\mathbf{F}_{t-1|T}^{(k)} \mathbf{F}_{t-1|T}^{(k)\prime} + \mathbf{P}_{t-1|T}^{(k)} \right) \right\}^{-1}. \tag{15}$$

For $p_F > 1$ we can simply write the VAR in companion form and derive the analogous of (15).

Finally, the estimator of the covariance matrix of the VAR innovations $\{\mathbf{v}_t\}$ is:

$$\begin{aligned} \widehat{\boldsymbol{\Gamma}}^{v(k+1)} = \frac{1}{T} \sum_{t=2}^T \Big\{ & \mathbf{F}_{t|T}^{(k)} \mathbf{F}_{t|T}^{(k)\prime} + \mathbf{P}_{t|T}^{(k)} - \widehat{\mathbf{A}}^{(k+1)} \left(\mathbf{F}_{t-1|T}^{(k)} \mathbf{F}_{t-1|T}^{(k)\prime} + \mathbf{P}_{t-1|T}^{(k)} \right) \widehat{\mathbf{A}}^{(k+1)\prime} \\ & - \left(\mathbf{F}_{t|T}^{(k)} \mathbf{F}_{t-1|T}^{(k)\prime} + \mathbf{C}_{t,t-1|T}^{(k)} \right) \widehat{\mathbf{A}}^{(k+1)\prime} - \widehat{\mathbf{A}}^{(k+1)} \left(\mathbf{F}_{t-1|T}^{(k)} \mathbf{F}_{t|T}^{(k)\prime} + \mathbf{C}_{t,t-1|T}^{(k)\prime} \right) \Big\}. \end{aligned} \tag{16}$$

Summing up, in the M-step, we need to compute Q_n values of the parameters for a given estimator of the factors, $\mathbf{F}_{t|T}^{(k)}$, $t = 1, \dots, T$. This step is feasible because we used the mis-specified log-likelihood (9) to estimate $\boldsymbol{\phi}_n$. Therefore, we decomposed our estimation problem into n separate maximizations, each requiring estimating $r+1$ parameters using T observations. In this way, estimating the high-dimensional parameter vector $\boldsymbol{\phi}_n$ becomes straightforward, and estimating $\boldsymbol{\theta}$ poses no problem because it is a low-dimensional problem with a closed-form solution.

3.3 Convergence of the EM algorithm and final estimators

Given the Gaussian quasi-likelihoods (9) and (10), it is easy show that, for any fixed n, there exists an $\omega > 0$ such that for any $k \geq 0$

$$\left\{\ell(\boldsymbol{X}_{nT}; \widehat{\boldsymbol{\varphi}}_n^{(k+1)}) - \ell(\boldsymbol{X}_{nT}; \widehat{\boldsymbol{\varphi}}_n^{(k)})\right\} \geq \left\{\mathcal{Q}(\widehat{\boldsymbol{\varphi}}_n^{(k+1)}, \widehat{\boldsymbol{\varphi}}_n^{(k)}) - \mathcal{Q}(\widehat{\boldsymbol{\varphi}}_n^{(k)}, \widehat{\boldsymbol{\varphi}}_n^{(k)})\right\} \geq \omega \|\widehat{\boldsymbol{\varphi}}_n^{(k+1)} - \widehat{\boldsymbol{\varphi}}_n^{(k)}\|^2, \qquad (17)$$

where the first inequality follows from Dempster et al. (1977, Lemma 1) and the second is due to strong concavity of $\mathcal{Q}(\cdot, \underline{\boldsymbol{\varphi}}_n)$ for any $\underline{\boldsymbol{\varphi}}_n$ (Wu, 1983, Condition 1). Moreover, the left-hand-side of (17) tends to zero as $k \to \infty$ (Wu, 1983, Theorem 3). Therefore, the EM algorithm defines a contractive map. Consequently, the sequence $\{\widehat{\boldsymbol{\varphi}}_n^{(k)}\}$ will converge to a maximum of the log-likelihood, as $k \to \infty$ (see Lemma E.21 for a formal proof when $n \to \infty$).

In practice, we stop the EM algorithm when the log-likelihood shows no further appreciable increase. This is ensured according to a standard convergence rule (see Appendix A.3), depending on a pre-specified threshold ε.

We denote the last iteration of the EM algorithm as k^*. Upon convergence, the EM estimator of the parameters is $\widehat{\boldsymbol{\varphi}}_n \equiv \widehat{\boldsymbol{\varphi}}_n^{(k^*+1)}$. By running the Kalman smoother one last time using $\widehat{\boldsymbol{\varphi}}_n$, we have the estimator of the factors $\widehat{\mathbf{F}}_t = \mathbf{F}_{t|T}^{(k^*+1)}$, $t = 1, \dots, T$. Finally, we estimate the common components as $\widehat{\chi}_{it} = \widehat{\boldsymbol{\lambda}}_i' \widehat{\mathbf{F}}_t$, where $\widehat{\boldsymbol{\lambda}}_i \equiv \widehat{\boldsymbol{\lambda}}_i^{(k^*+1)}$, $i = 1, \dots, n$.

4 The Large Approximate Dynamic Factor Model

4.1 Main assumptions

This section presents the assumptions under which we can consistently estimate the DFM given in (3)-(4). These assumptions are stated for an infinite-dimensional stochastic process $\{x_{it},\, i \in \mathbb{N},\, t \in \mathbb{Z}\}$ of which $\{\mathbf{x}_{nt} = (x_{1t} \cdots x_{nt})',\, t \in \mathbb{Z}\}$ is an n-dimensional subprocess and the $T \times n$ matrix $(\mathbf{x}_{n1} \cdots \mathbf{x}_{nT})'$ is an observed realization. Likewise, $\{\boldsymbol{\xi}_{nt} = (\xi_{1t} \cdots \xi_{nt})',\, t \in \mathbb{Z}\}$ is an n-dimensional sub-process of the infinite-dimensional stochastic process of idiosyncratic components $\{\xi_{it},\, i \in \mathbb{N},\, t \in \mathbb{Z}\}$, and $\{\mathbf{F}_t = (F_{1t} \cdots F_{rt})',\, t \in \mathbb{Z}\}$ and $\{\mathbf{v}_t = (v_{1t} \cdots v_{rt})',\, t \in \mathbb{Z}\}$ are the r-dimensional processes of common factors and the corresponding VAR innovations, respectively. The $n \times r$ matrix of factor loadings $\boldsymbol{\Lambda}_n = (\boldsymbol{\lambda}_1 \cdots \boldsymbol{\lambda}_n)'$ forms a nested sequence as n increases. Finally, as already noticed, the QML estimator of $\boldsymbol{\mu}_n$ is immediately obtained as the sample mean $\bar{\mathbf{x}}_n$. Thus, hereafter, for simplicity, we consider the DFM for pre-centered data, or, equivalently, we set $\boldsymbol{\mu}_n = \mathbf{0}_n$.

Assumption 1 (LOADINGS AND FACTORS)**.**

(a) *There exists an integer N_0 such that for all $n > N_0$, $\|n^{-1}\boldsymbol{\Lambda}_n'\boldsymbol{\Lambda}_n - \boldsymbol{\Sigma}_\Lambda\| = 0$, where $\boldsymbol{\Sigma}_\Lambda$ is $r \times r$ and positive definite; moreover, for all $n \in \mathbb{N}$, $m_\lambda \leq \max_{i=1,\dots,n} \|\boldsymbol{\lambda}_i\| \leq M_\lambda$ for some finite positive reals M_λ and m_λ independent of n.*

(b) *For all $t \in \mathbb{Z}$, $\boldsymbol{\Gamma}^F = \mathbb{E}[\mathbf{F}_t \mathbf{F}_t']$ is $r \times r$ and positive definite, and $\|\boldsymbol{\Gamma}^F\| \leq M_F$ for some finite positive real M_F.*

(c) *There exists an integer N_1 such that for all $n > N_1$, r is a finite positive integer, independent of n, and such that $r \leq N_1$.*

(d) *$\mathbf{A}(z) = \sum_{k=1}^{p_F} \mathbf{A}_k z^{k-1}$, such that p_F is a finite positive integer, $\mathbf{A}_k$ are $r \times r$, and $\det(\mathbf{I}_r - \mathbf{A}(z)) \neq 0$ for all $z \in \mathbb{C}$ such that $|z| \leq M_A$ for some finite positive real $M_A < 1$.*

(e) *For all $t \in \mathbb{Z}$, $\mathbb{E}[\mathbf{v}_t] = \mathbf{0}_r$, $\boldsymbol{\Gamma}^v = \mathbb{E}[\mathbf{v}_t \mathbf{v}_t']$ is $r \times r$ positive definite, and $\|\boldsymbol{\Gamma}^v\| \leq M_v$ for some finite positive real M_v.*

(f) *For all $t \in \mathbb{Z}$ and all $k \in \mathbb{Z}$ with $k \neq 0$, $\mathbf{v}_t$ and $\mathbf{v}_{t-k}$ are independent.*

(g) *For all $j_1, j_2, j_3, j_4 = 1, \dots, r$, all $t = 1, \dots, T$, and all $T \in \mathbb{N}$,*

$$\frac{1}{T} \sum_{s_1, s_2 = 1}^{T} |\mathbb{E}[v_{j_1 s_1} v_{j_2 t} v_{j_3 s_2} v_{j_4 t}]| \leq \mathrm{K}_v, \qquad \frac{1}{T} \sum_{s_1, s_2 = 1}^{T} |\mathbb{E}[v_{j_1 s_1} v_{j_2 t}]||\mathbb{E}[v_{j_3 s_2} v_{j_4 t}]| \leq \mathrm{K}_v,$$

for some finite positive real K_v *independent of* j_1, j_2, j_3, j_4*, and* T*.*

(h) For all $t \in \mathbb{Z}$*,* $\mathbf{v}_t$ *has pdf* $f_{\mathbf{v}_t}(\boldsymbol{u})$ *such that* $\int_{\mathbb{R}^r} |f_{\mathbf{v}_t}(\boldsymbol{u}+\boldsymbol{v}) - f_{\mathbf{v}_t}(\boldsymbol{v})| \mathrm{d}\boldsymbol{u} \le C_f \|\boldsymbol{v}\|$ *for any* $\boldsymbol{v} \in \mathbb{R}^r$ *and for some finite positive real* C_f *independent of* t*;*

(i) For all $t \le 0$*,* $\mathbf{v}_t = \mathbf{0}_r$*.*

Parts (a) and (b) imply that the loadings matrix has asymptotically maximum column rank r (part (a)), and the factors have a finite full-rank covariance matrix (part (b))—these assumptions are similar to the requirements in Bai and Li (2016, Assumptions A and B) and Bai (2003, Assumptions A and B). Moreover, because of part (a), for any given $n \in \mathbb{N}$, all the factors have a finite contribution to each series (upper bound on $\max_{i=1,\dots,n} \|\boldsymbol{\lambda}_i\|$), and there is at least one factor that contributes to at least one series (lower bound on $\max_{i=1,\dots,n} \|\boldsymbol{\lambda}_i\|$). While the former condition is common, the latter is less standard but very mild as it simply guarantees that at least one loading is non-zero for any fixed $n \in \mathbb{N}$.

Part (c) implies the existence of a finite number of common factors r. In particular, r is identified only for $n \to \infty$ (see also the next section). Hereafter, in parts (a) and (c), we can assume $N_0 = N_1 = N$, say, without loss of generality.

The remaining conditions of Assumption 1 characterize the VAR for the factors in (4). Part (d) implies that $\{\mathbf{F}_t\}$ is a weakly stationary process with a causal autoregressive representation. And in parts (e), (f), and (g), we assume that $\{\mathbf{v}_t\}$ is a zero-mean r-dimensional independent process with finite positive definite covariance matrix and finite summable 4th order cumulants. Parts (d) and (e) imply also that $\boldsymbol{\Gamma}^F$ is finite, as required in part (b).

Part (h) is an integral Lipschitz condition, which is satisfied by most continuous densities. This assumption guarantees that $\{\mathbf{F}_t\}$ is a strong mixing, or equivalently α-mixing, process with mixing coefficients $\alpha_F(T) \le \exp\{-c_F T^{\gamma_F}\}$, for all $T \in \mathbb{N}$ and some finite positive reals c_F and γ_F independent of T (Pham and Tran, 1985, Theorem 3.1).[3] Strongly mixing factors with exponentially decaying mixing coefficients are directly assumed by Fan et al. (2013, Assumption 2c).

Part (i) implies $\mathbf{F}_t = \mathbf{0}_r$ for $t \le 0$. This assumption is standard and it fixes the initial conditions for the solution of the VAR in (4).

Assumption 2 (IDIOSYNCRATIC COMPONENT)**.**

(a) For all $i \in \mathbb{N}$ *and all* $t \in \mathbb{Z}$*,* $\mathbb{E}[\xi_{it}] = 0$ *and* $\sigma_i^2 = \mathbb{E}[\xi_{it}^2]$ *is such that* $C_\xi^{-1} \le \sigma_i^2 \le C_\xi$*, for some finite positive real* C_ξ *independent of* i*.*

(b) For all $i, j \in \mathbb{N}$*, all* $t \in \mathbb{Z}$*, and all* $k \in \mathbb{Z}$*,* $|\mathbb{E}[\xi_{it}\xi_{j,t-k}]| \le \rho^{|k|} M_{ij}$*, where* ρ *and* M_{ij} *are finite positive reals independent of* t *such that* $0 \le \rho < 1$*,* $M_{ii} = \sigma_i^2$*,* $\sum_{j=1,j\neq i}^n M_{ij} \le M_\xi$*, and* $\sum_{i=1,i\neq j}^n M_{ij} \le M_\xi$ *for some finite positive real* M_ξ *independent of* n*.*

(c) For all $i \in \mathbb{N}$*,* $\{\xi_{it}\}$ *is a strong mixing process with mixing coefficients such that* $\alpha_{\xi_i}(T) \le \exp(-c_\xi T^{\gamma_\xi})$*, for all* $T \in \mathbb{N}$*, and for some finite positive reals* c_ξ *and* γ_ξ *independent of* T *and* i*.*

(d) For all $j = 1, \dots, n$ *and all* $n, T \in \mathbb{N}$*,*

$$\frac{1}{nT}\sum_{t,s=1}^T \sum_{i_1,i_2=1}^n |\mathbb{E}[\xi_{i_1 t}\xi_{jt}\xi_{i_2 s}\xi_{js}]| \le \mathrm{K}_\xi, \quad \frac{1}{nT}\sum_{t,s=1}^T \sum_{i_1,i_2=1}^n |\mathbb{E}[\xi_{i_1 t}\xi_{jt}]||\mathbb{E}[\xi_{i_2 s}\xi_{js}]| \le \mathrm{K}_\xi,$$

and for all $t = 1, \dots, T$ *and all* $n, T \in \mathbb{N}$*,*

$$\frac{1}{nT}\sum_{s_1,s_2=1}^T \sum_{i,j=1}^n |\mathbb{E}[\xi_{is_1}\xi_{it}\xi_{js_2}\xi_{jt}]| \le \mathrm{K}_\xi, \quad \frac{1}{nT}\sum_{s_1,s_2=1}^T \sum_{i,j=1}^n |\mathbb{E}[\xi_{is_1}\xi_{it}]||\mathbb{E}[\xi_{js_2}\xi_{jt}]| \le \mathrm{K}_\xi,$$

[3] Independence in part (f) is not strictly necessary for having $\{\mathbf{F}_t\}$ strong mixing, as we could allow for GARCH effects by assuming geometric ergodicity of $\{\mathbf{v}_t\}$ instead (Francq and Zakoïan, 2006). Indeed, geometric ergodicity implies β-mixing, which implies strong mixing.

for some finite positive real K_ξ *independent of* n*, and* T*.*

(e) For all $t \in \mathbb{Z}$*, as* $n \to \infty$*,*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\boldsymbol{\lambda}_i \xi_{it}}{\sigma_i^2} \xrightarrow{d} \mathcal{N}\left(\mathbf{0}_r, \lim_{n\to\infty} \frac{1}{n} \sum_{i,j=1}^{n} \frac{\boldsymbol{\lambda}_i \boldsymbol{\lambda}_j' \mathbb{E}[\xi_{it}\xi_{jt}]}{\sigma_i^2 \sigma_j^2}\right).$$

(f) For all $n \in \mathbb{N}$*,* $\boldsymbol{\Gamma}_n^\xi = \mathbb{E}[\boldsymbol{\xi}_{nt}\boldsymbol{\xi}_{nt}']$ *is such that* $\nu^{(n)}(\boldsymbol{\Gamma}_n^\xi) \geq L_\xi$ *for some finite positive real* L_ξ *independent of* n*.*

Part (a) imposes that the idiosyncratic components have zero mean, and the idiosyncratic variances are positive and finite. We strengthen this assumption in part (f) by requiring that the whole idiosyncratic covariance matrix is positive definite.

Part (b) has a twofold purposes. First, it limits the degree of serial correlation of the idiosyncratic components by assuming standard geometric decay of the autocovariances. Second, it limits the degree of cross-sectional correlation between idiosyncratic components, a standard assumption for approximate DFMs. This assumption implies the usual conditions required by Bai (2003, Assumptions C2, C3, and C4), Fan et al. (2013, Assumption 2b), and Bai and Li (2016, Assumptions C3, C4, and E1) (see Lemma C.1).

In part (c), we assume that each idiosyncratic component is strongly mixing with exponentially decaying coefficients. This requirement is quite general since it allows the idiosyncratic to be non linear processes—Fan et al. (2013, Assumption 2c) make the same assumption.

In part (d), we require finite summable fourth-order cumulants—a standard requirement found in the literature (see, e.g., Bai, 2003, Assumption F1 for the first condition, and Bai and Li, 2016, Assumption E2 for the second one—which, jointly with the mixing assumption in part (c), allows for consistent estimation of covariances by means of the sample covariances (Hannan, 1970, Theorem 6, Chapter IV, p. 210).

In part (e), we assume a Central Limit Theorem, a standard assumption in the literature (e.g., Bai, 2003, Assumption F3, and Bai and Li, 2016, Assumption F3). This is a high-level requirement and in order to properly derive it, we should introduce some notion of dependence for random fields, which typically requires some ordering of the cross-sectional units. Now, in most applications, there is not a natural ordering of the variables. For this reason, we avoid to spell out primitive conditions that guarantee part (e) to hold.

We then add the natural requirement of independence between common shocks and idiosyncratic components.

Assumption 3 (Independence between common shocks and idiosyncratic components)**.** *The processes* $\{\xi_{it},\, i \in \mathbb{N},\, t \in \mathbb{Z}\}$ *and* $\{v_{jt},\, j = 1, \ldots, r,\, t \in \mathbb{Z}\}$ *are mutually independent.*

Assumption 3 implies that the factors and the common components are independent of the idiosyncratic components at all leads and lags and across all units. This assumption is compatible with the idea that the structural macroeconomic shocks driving the common component are independent of the idiosyncratic components representing measurement errors or local dynamics.[4]

Last, Assumption 3 jointly with Assumptions 1(f), 1(g), 1(h), 2(c), and 2(d), implies that the process $\{\mathbf{F}_t \xi_{it}\}$ is strongly mixing with finite fourth moments (Bradley, 2005, Theorem 5.1.a). Then, from Ibragimov (1962, Theorem 1.7), we have the following Central Limit Theorem, for all $i \in \mathbb{N}$, as $T \to \infty$,

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \mathbf{F}_t \xi_{it} \xrightarrow{d} \mathcal{N}\left(\mathbf{0}_r, \lim_{T\to\infty} \frac{1}{T} \sum_{s,t=1}^{T} \mathbb{E}\left[\mathbf{F}_t \mathbf{F}_s' \xi_{it} \xi_{is}\right]\right).$$

This last result is typically assumed in the literature—see, e.g., Bai, 2003, Assumption F4, and Bai and Li, 2016, Assumption F1.

[4]In principle, we could relax this requirement to allow for weak dependence as in, e.g., Bai (2003, Assumption D).

4.2 Additional assumptions

We now state two more assumptions. These are needed only to derive some of our asymptotic results.

Assumption 4 (Linearity). *For all $j = 1, \dots, r$, $t = 1, \dots, T$, and $n, T \in \mathbb{N}$, $\mathbb{E}[F_{jt}|\boldsymbol{X}_{nT}]$ is a linear function of $\boldsymbol{X}_{nT}$.*

Assumption 4 would hold if we directly assume joint Gaussianity of $\boldsymbol{X}_{nT}$ and $\boldsymbol{F}_T$. However, assuming Gaussianity might be too stringent, while Assumption 4 is more general. For example, consider the case $r = 1$, let f_t and $\mathbf{X}_{nT}$ be realizations of the factor and the data, and let $f : \mathbb{R}^{nT+1} \to \mathbb{R}$ be the joint pdf of F_t and $\boldsymbol{X}_{nT}$. Then, from Steyn (1960) and Kotz et al. (2004, Chapter 44.3), we see that if f belongs to the family of multivariate Pearson-type distributions, and $\lim_{\mathrm{f}_t \to \pm\infty} \mathrm{f}_t^2 f(\mathrm{f}_t, \mathbf{X}_{nT}) = 0$, then $\mathbb{E}[F_t|\boldsymbol{X}_{nT} = \mathbf{X}_{nT}]$ is linear in $\mathbf{X}_{nT}$. Quah and Sargent (1993, p.292) and Doz et al. (2012, Assumption R) made similar assumptions.

Assumption 5 (Tails).

(a) For all $t \in \mathbb{Z}$, all $j = 1, \dots, r$, and all $s > 0$, $\mathrm{P}(|v_{jt}| \geq s) \leq \exp\left\{-K_v s^{\delta_v}\right\}$ for some finite positive reals K_v and $\delta_v \leq 2$ independent of t and j.

(b) For all $t \in \mathbb{Z}$, all $n \in \mathbb{N}$, all $s > 0$,

$$\mathrm{P}\left(\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\boldsymbol{\lambda}_i \xi_{it}}{\sigma_i^2}\right\| \geq s\right) \leq r \exp\Big\{-\kappa_1 s^2\Big\} + rn \exp\Big\{-\kappa_2\Big(s\sqrt{n}\Big)^{\alpha}\Big\},$$

for some finite positive reals κ_1, κ_2, and $\alpha \leq 2$ independent of t and n.

In part (a), we assume an exponential-type tail inequality for the common shocks, which implies that for all $s > 0$, $t \in \mathbb{Z}$, and $j = 1, \dots, r$, $\mathrm{P}(|F_{jt}| > s) \leq \exp\big\{-K_F s^{\delta_v}\big\}$, for some finite positive reals K_F and $\delta_v \leq 2$ independent of t and j (see Lemma E.1 and Bakhshizadeh et al., 2023, Corollary 4). The same applies to part (b), which, by setting $n = 1$, implies that for all $s > 0$, $t \in \mathbb{Z}$, and $i \in \mathbb{N}$, $\mathrm{P}(|\xi_{it}| \geq s) \leq \exp\big\{-K_\xi s^{\delta_\xi}\big\}$, for some finite positive real K_ξ and $\delta_\xi \leq 2$ independent of t and j. This is because $\|\boldsymbol{\lambda}_i\| \leq M_\lambda$ and $\sigma_i^2 \geq C_\xi^{-1}$, for all $i \in \mathbb{N}$, by Assumptions 1(a) and 2(a), respectively. Fan et al. (2013, Assumption 2c) also assume the factors and idiosyncratic components to belong to a distribution having exponentially decaying tails.

Depending on the values of δ_v and δ_ξ, we are able to consider not only distributions with sub-Gaussian ($\delta_v, \delta_\xi = 2$) or sub-exponential tails ($\delta_v, \delta_\xi = 1$), which include the Laplace and the Generalized Error distribution (Vershynin, 2018, Chapter 2), but also distributions with sub-Weibull tails ($\delta_v, \delta_\xi < 1$), which can mimic a heavy tail behavior even if all moments exist (Mikosch and Nagaev, 1998; Kuchibhotla and Chakrabortty, 2022; Vladimirova et al., 2020). This is clarified in Remark 2 below.

In general, part (b) is a Bernstein-type inequality implying that the weighted sums of the idiosyncratic components have Gaussian tails, as $n \to \infty$. This is a high-level requirement and, as for Assumption 2(e) in order to properly derive it, we should introduce some notion of dependence for random fields. Here, instead, we just notice that in the simplest case of cross-sectionally independent idiosyncratic components, this condition would be a direct consequence of Bakhshizadeh et al. (2023, Corollary 4) (see also Vladimirova et al., 2020, Corollary 1 for a similar result, and Vershynin, 2018, Theorems 2.8.2 and 2.6.3, for the cases $\delta_\xi = 1$ and $\delta_\xi = 2$, respectively).

Finally, as a consequence of parts (a) and (b), jointly with Assumptions 1(f), 1(h), 2(c), and 3, we can show that not only $\{\mathbf{F}_t \xi_{it}\}$ is a strongly mixing process but its components have also exponentially decaying tails. Therefore, for all $T \in \mathbb{N}$ and all $i \in \mathbb{N}$, the following Bernstein-type inequality holds:

$$\mathrm{P}\left(\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{F}_t \xi_{it}\right\| \geq s\right) \leq r \exp\Big\{-\kappa_3 s^2\Big\} + rT \exp\Big\{-\kappa_4\Big(s\sqrt{T}\Big)^{\beta}\Big\},$$

for some finite positive reals κ_3, κ_4, and $\beta < 1$ independent of i and T (Merlevède et al., 2011, Theorem 1, Bosq, 2012, Theorem 1.4, p.31, and Lemma E.1).

Remark 2. Following Kuchibhotla and Chakrabortty (2022) and Kuchibhotla et al. (2023), we say that a given random variable y is sub-Weibull with exponent δ if $\mathrm{P}(|y| \geq s) \leq \exp\left\{-K_y s^\delta\right\}$ for any $s > 0$ and some positive real K_y. This is equivalent to requiring the following Cramér type condition to hold: $\sup_{m\geq 1} r^{-1/\delta}(\mathbb{E}[|y|^m])^{1/m} \leq M_y$ for some positive real M_y.[5] This shows that although all moments exist, they can be quite large for small values of δ. For example, the mth moment of a Weibull is given by $\mathbb{E}[y^m] = \Gamma(1 + \frac{m}{\delta})$, which is rapidly increasing as δ decreases (see, e.g., Lehman, 1963, for tabulated values of the first four moments as functions of δ).

4.3 Identification conditions

To identify the DFM, we need to address four issues: first, we need to identify the number of factors; second, we need to identify the true parameters of the model; third, we need to ensure that the linear system is identified; and fourth, we need to guarantee the existence of the maxima of the log-likelihood.

4.3.1 Identification of the number of factors and of the common component

Starting with the number of factors r, let the covariance matrix of $\{\boldsymbol{\chi}_{nt}\}$ be $\boldsymbol{\Gamma}_n^\chi = \boldsymbol{\Lambda}_n \boldsymbol{\Gamma}^F \boldsymbol{\Lambda}_n'$ and denote as μ_{jn}^χ the j-th largest eigenvalue of $\boldsymbol{\Gamma}_n^\chi$, then Assumptions 1(a) and 1(b) imply that, for any $j = 1, \dots, r$,

$$\underline{C}_j \leq \liminf_{n\to\infty} \frac{\mu_{jn}^\chi}{n} \leq \limsup_{n\to\infty} \frac{\mu_{jn}^\chi}{n} \leq \overline{C}_j, \tag{18}$$

for some finite positive reals $\underline{C}_j$ and $\overline{C}_j$. Furthermore, from Assumption 2(b) it follows that the largest eigenvalue of the idiosyncratic covariance matrix, $\boldsymbol{\Gamma}_n^\xi$, denoted as μ_{1n}^ξ, is such that

$$\mu_{1n}^\xi = \|\boldsymbol{\Gamma}_n^\xi\| \leq M_\xi. \tag{19}$$

From conditions (18) and (19) and Weyl's inequality, it follows that the r largest eigenvalues of the covariance matrix of $\{\mathbf{x}_{nt}\}$ diverge linearly in n, whereas all remaining eigenvalues stay bounded for all $n \in \mathbb{N}$. Lemma C.1 proves this result, condition (18), and condition (19)—these conditions are directly imposed by Doz et al. (2012, Assumptions A1 and A2, respectively).

The asymptotic behavior of the eigenvalues of the covariance matrix allows for the identification of the common and idiosyncratic component when $n \to \infty$ and it is the basis for all existing methods for determining r (see, e.g., Bai and Ng, 2002).

4.3.2 Identification of the loadings, the factors, and the VAR parameters

All the structures equivalent to (3)-(4) can be obtained through an $r \times r$ invertible matrix $\mathbf{R}$, as follows:

$$\mathbf{F}_t^o = \mathbf{R}^{-1}\mathbf{F}_t, \quad \boldsymbol{\Lambda}_n^o = \boldsymbol{\Lambda}_n\mathbf{R}, \quad \mathbf{A}^o(L) = \mathbf{R}^{-1}\mathbf{A}(L)\mathbf{R}, \quad \boldsymbol{\Gamma}^{vo} = \mathbf{R}^{-1}\boldsymbol{\Gamma}^v, \quad \boldsymbol{\Gamma}_n^{\xi o} = \boldsymbol{\Gamma}_n^\xi.$$

Under such relationships, using only first- and second-moment information we cannot distinguish the model specified by $\boldsymbol{\Lambda}_n^o$, $\mathbf{A}^o(L)$, $\boldsymbol{\Gamma}^{vo}$, and $\boldsymbol{\Gamma}^{\xi o}$, from the one given by $\boldsymbol{\Lambda}_n$, $\mathbf{A}(L)$, $\boldsymbol{\Gamma}^v$, and $\boldsymbol{\Gamma}^\xi$. Once the loadings and the factors are identified, then $\mathbf{A}(L)$ and $\boldsymbol{\Gamma}^v$ are also identified; $\boldsymbol{\Gamma}_n^\xi$ is always identified.

To identify the model, we need enough a priori structure to preclude any but the trivial transformation $\mathbf{R} = \mathbf{I}_r$. This can be achieved by imposing additional r^2 identifying constraints. Let $\mathbf{M}_n^\chi$ be the $r \times r$ diagonal matrix with as elements the eigenvalues μ_{jn}^χ, $j = 1, \dots, r$, of the covariance matrix of the common component,

[5] Another necessary and sufficient condition for a random variable y to be sub-Weibull is given by the following condition on its Orlicz norm: $\|y\|_{\psi_\delta} = \inf\{\eta > 0 : \mathbb{E}[\exp((|y|/\eta)^\delta)] \leq 2\} \leq M_y'$ for some finite positive real M_y'.

$\mathbf{\Gamma}_n^\chi$, sorted in descending order. Let $\mathbf{V}_n^\chi$ be the $n \times r$ matrix with the corresponding normalized eigenvectors as columns. Then, we assume the following identifying constraints.

Assumption 6 (Identification).
(a) The eigenvalues of $\mathbf{\Sigma}_\Lambda \mathbf{\Gamma}^F$ are distinct.
(b) $\mathbf{\Sigma}_\Lambda$ is diagonal and $\mathbf{\Gamma}^F = \mathbf{I}_r$.
(c) For all $j = 1, \ldots, r$, $[\mathbf{\Lambda}_n]_{1j} \geq 0$.

Part (a) is standard. Since the eigenvalues of $\mathbf{\Sigma}_\Lambda \mathbf{\Gamma}^F$ are equal to the r non-zero eigenvalues of $\lim_{n\to\infty} n^{-1}\mathbf{\Gamma}_n^\chi$, given by $\lim_{n\to\infty} n^{-1}\mathbf{M}_n^\chi$, it implies that in (18) we have $\overline{C}_j < \underline{C}_{j-1}$ for any $j = 2, \ldots, r$, and, thus, it avoids the uninteresting difficulties related with asymptotically multiple eigenvalues, which would require more restrictions to identify the space spanned by the columns of $\mathbf{\Lambda}_n$.

Part (b) is similar to what is usually imposed in PC estimation (see Remark 3 below). It implies that $\mathbf{\Gamma}_n^\chi = \mathbf{\Lambda}_n \mathbf{\Lambda}_n' = \mathbf{V}_n^\chi \mathbf{M}_n^\chi \mathbf{V}_n^{\chi\prime}$. Hence, the r non-zero eigenvalues of $\lim_{n\to\infty} n^{-1}\mathbf{\Gamma}_n^\chi$, given by $\lim_{n\to\infty} n^{-1}\mathbf{M}_n^\chi$, coincide with the diagonal entries of $\mathbf{\Sigma}_\Lambda$, which are then distinct because of part (a). Since part (b) concerns second moments and sums of squares, it allows us to identify $\mathbf{\Lambda}_n$ only up to a sign. To achieve global identification we must fix also the column sign of $\mathbf{\Lambda}_n$ which is done through part (c) (Bai and Ng, 2013, Remark 1).

Summing up, in part (b), we are imposing r^2 restrictions: $r(r+1)/2$ by requiring orthonormality of the factors, and $r(r-1)/2$ by requiring that $\mathbf{\Sigma}_\Lambda$ is diagonal. Consistently with the fact that in the typical empirical applications the focus is on the common component only, these assumed identification conditions do not provide economic meaning to the factors; in this sense ours is an exploratory rather than confirmatory factor analysis. Moreover, and most importantly, the restrictions are imposed only in the limit $n, T \to \infty$. Thus, in our setting, the model is only asymptotically identified. This is enough to derive our asymptotic theory.

Remark 3. Typically, in PC estimation it is required that: (i) for all $n \in \mathbb{N}$, $n^{-1}\mathbf{\Lambda}_n'\mathbf{\Lambda}_n$ is a diagonal matrix with finite distinct elements, and (ii) for all $T \in \mathbb{N}$, $T^{-1}\sum_{t=1}^T \mathbf{F}_t\mathbf{F}_t' = \mathbf{I}_r$ (Forni et al., 2009, Doz et al., 2011, 2012, and Bai and Ng, 2013). In classical QML estimation constraint (ii) is the same as in PC estimation, while (i) is replaced with the requirement that for all $n \in \mathbb{N}$, $n^{-1}\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n$ is a diagonal matrix with finite distinct elements (Bai and Li, 2012, 2016, constraint IC3 therein). Differently from the present setting these constraints are assumed to hold for any given $n, T \in \mathbb{N}$.

Remark 4. By letting $\boldsymbol{\mathcal{S}}$ be a diagonal $r \times r$ matrix with entries $\mathbb{I}([\mathbf{V}_n^\chi]_{1j} \geq 0) - \mathbb{I}([\mathbf{V}_n^\chi]_{1j} < 0)$, $j = 1, \ldots, r$, we can show that, as $n \to \infty$, $\boldsymbol{\lambda}_i'$ coincides with $\mathbf{v}_i^{\chi\prime}\boldsymbol{\mathcal{S}}(\mathbf{M}_n^\chi)^{1/2}$, where $\mathbf{v}_i^{\chi\prime}$ is the ith row of $\mathbf{V}_n^\chi$ (see Lemma C.9). It also follows that, by linear projection of $\boldsymbol{\chi}_{nt}$ onto $\mathbf{\Lambda}_n$ at each given t, the true factors $\mathbf{F}_t$ are identified, as $n \to \infty$, as the first r normalized PCs of the common component given by $(\mathbf{M}_n^\chi)^{-1/2}\boldsymbol{\mathcal{S}}\mathbf{V}_n^{\chi\prime}\boldsymbol{\chi}_{nt}$, which are clearly orthonormal, indeed, $\mathbb{E}[(\mathbf{M}_n^\chi)^{-1/2}\boldsymbol{\mathcal{S}}\mathbf{V}_n^{\chi\prime}\boldsymbol{\chi}_{nt}\boldsymbol{\chi}_{nt}'\mathbf{V}_n^\chi\boldsymbol{\mathcal{S}}(\mathbf{M}_n^\chi)^{-1/2}] = \mathbf{I}_r = \mathbf{\Gamma}^F$ as requested.

4.3.3 Identification of the linear system

From our Assumptions 1(a), 1(d), and 1(e), we can show that the state space formulation (3)-(4) is minimal and stable, for all $n > N$, where $N = N_0$ in Assumption 1(a). It is convenient to consider here the factorization $\mathbf{\Gamma}^v = \mathbf{H}\mathbf{H}'$ for some $r \times r$ matrix $\mathbf{H}$ having full rank. Consider again for simplicity the case $p_F = 1$ with $\mathbf{A} \equiv \mathbf{A}_1$. Then, stability, which requires $|\nu^{(1)}(\mathbf{A})| < 1$, is a direct consequence of stationarity in Assumption 1(d). While minimality holds because the couple $(\mathbf{A}, \mathbf{H})$ is controllable due to Assumption 1(e) and the couple $(\mathbf{A}, \mathbf{\Lambda}_n)$ is observable for all $n > N$ because Assumption 6(b) implies that $\mathrm{rk}(\mathbf{\Lambda}_n) = \mathrm{rk}(\mathbf{V}_n^\chi) = r$, for all $n > N$.[6] This implies that, for all $n > N$, the linear system satisfies the mini-phase condition:

$$\mathrm{rk}\begin{pmatrix} \mathbf{I}_r - \mathbf{A}z & -\mathbf{H} \\ \mathbf{\Lambda}_n & \mathbf{0}_{n\times r} \end{pmatrix} = 2r, \quad \text{for } |z| \geq 1.$$

[6] A linear system with $p_F = 1$, is controllable if and only if $\mathrm{rk}[\mathbf{H}\ (\mathbf{A}\mathbf{H}) \cdots (\mathbf{A}^{(r-1)}\mathbf{H})] = r$ and it is observable if and only if $\mathrm{rk}[\mathbf{\Lambda}_n'\ (\mathbf{\Lambda}_n\mathbf{A})' \cdots (\mathbf{\Lambda}_n\mathbf{A}^{r-1})'] = r$ (Anderson and Moore, 1979, Appendix C, pp. 341-342).

In other words, the linear system (3)-(4) is the minimal state-space representation of a DFM having as McMillan degree the number of factors r (Anderson and Deistler, 2008, Section II).

This result, together with Assumption 6, guarantees that the transfer function matrix $\boldsymbol{W}(z) = \boldsymbol{\Lambda}_n(\mathbf{I}_r - \mathbf{A}z)^{-1}\mathbf{H}$ is identified for all $z \in \mathbb{C}$, with the exception of a zero-measure set (see also Lippi et al., 2021, Section 4.3, for a proof, and Heaton and Solo, 2004, for similar results). This implies generic identifiability of the linear system (3)-(4).

4.3.4 Existence of the EM and QML estimators

Finally, we consider the issue of identification of the maxima of the log-likelihood. For any given $i \in \mathbb{N}$, define

$$\mathcal{O}_{\lambda_i} = \{\underline{\boldsymbol{\lambda}}_i \in \mathbb{R}^r, \underline{\boldsymbol{\lambda}}_i \in [-M_\lambda, M_\lambda]^r\} \text{ and } \mathcal{O}_{\sigma_i^2} = \{\underline{\sigma}_i^2 \in \mathbb{R}, \underline{\sigma}_i^2 \in [C_\xi^{-1}, C_\xi]\},$$

where M_λ and C_ξ are defined in Assumptions 1(a) and 2(a), respectively. Likewise, define

$$\mathcal{O}_{\mathcal{A}} = \{\text{vec}(\underline{\mathbf{A}}) \in \mathbb{R}^{r^2}, \text{vec}(\underline{\mathbf{A}}) \in [-M_A, M_A]^{r^2 p_F}\},$$
$$\mathcal{O}_{\Gamma^v} = \{\text{vech}(\underline{\boldsymbol{\Gamma}}^v) \in \mathbb{R}^{r(r+1)/2}, \nu^{(j)}(\underline{\boldsymbol{\Gamma}}^v) \in [M_v^{-1}, M_v], j = 1, \dots, r\},$$

where M_A and M_v are defined in Assumptions 1(d) and 1(e), respectively. Moreover, for any given $n \in \mathbb{N}$, let also

$$\mathcal{E}_{\Lambda_n} = \{\text{vec}(\underline{\boldsymbol{\Lambda}}_n)' \in \mathbb{R}^{nr}, \underline{C}_r \leq \nu^{(r)}(n^{-1}\underline{\boldsymbol{\Lambda}}_n'\underline{\boldsymbol{\Lambda}}_n) \leq \nu^{(1)}(n^{-1}\underline{\boldsymbol{\Lambda}}_n'\underline{\boldsymbol{\Lambda}}_n) \leq \overline{C}_1, \underline{\boldsymbol{\Lambda}}_n'\underline{\boldsymbol{\Lambda}}_n \text{diagonal}\},$$
$$\mathcal{E}_{\Gamma_n^\xi} = \{\text{vech}(\underline{\boldsymbol{\Gamma}}_n^\xi)' \in \mathbb{R}^{n(n+1)/2}, L_\xi \leq \nu^{(n)}(\underline{\boldsymbol{\Gamma}}_n^\xi) \leq \nu^{(1)}(\underline{\boldsymbol{\Gamma}}_n^\xi) \leq M_\xi\},$$

where $\underline{C}_r$ and $\overline{C}_1$ are defined in (18), while M_ξ and L_ξ are defined in Assumptions 2(b) and 2(f), respectively. Then, the search for the maximum of the expected log-likelihood in the EM algorithm and the one of the log-likelihood takes place on the set $\mathcal{O}_n = \{\mathcal{O}_{\lambda_i}^n \cap \mathcal{E}_{\Lambda_n}\} \times \{\mathcal{O}_{\sigma_i^2}^n \cap \mathcal{E}_{\Gamma_n^\xi}\} \times \mathcal{O}_{\mathcal{A}} \times \mathcal{O}_{\Gamma^v}$, which has dimension $Q_n = n(r+1) + r^2 p_F + r(r+1)/2$ growing with n.

Because of Assumptions 1(a) and 2(a), the rows of the loadings matrix $\boldsymbol{\lambda}_i$ and the idiosyncratic variances σ_i^2 belong to $\mathcal{O}_{\lambda_i} \times \mathcal{O}_{\sigma_i^2} \subset \mathbb{R}^{r+1}$, which is a compact set for any given $i \in \mathbb{N}$. Similarly, because of Assumptions 1(d) and 1(e), the entries of $\mathbf{A}_k$, $k = 1, \dots, p_F$, and $\mathbf{H}$ belong to $\mathcal{O}_{\mathcal{A}} \times \mathcal{O}_{\Gamma^v} \subset \mathbb{R}^{r^2 p_F + r(r+1)/2}$, which is also a compact set. These properties are crucial as they ensure the existence of a maximum which is a solution of the EM algorithm. Indeed, for any iteration $k \geq 0$ the expected log-likelihood $\mathcal{Q}(\underline{\boldsymbol{\varphi}}_n, \widehat{\boldsymbol{\varphi}}_n^{(k)})$ in (8), which we maximize in the M-step, is made of two terms: a term in (10) associated to the VAR for the factors (4), which is defined on the finite-dimensional set $\mathcal{O}_{\mathcal{A}} \times \mathcal{O}_{\Gamma^v}$, and a term (9) associated to the factor equation (3) and thus defined over a set $\mathcal{O}_{\lambda_i}^n \times \mathcal{O}_{\sigma_i^2}^n$ of dimension growing with n. Now, the former term poses no problem because we can use compactness of $\mathcal{O}_{\mathcal{A}} \times \mathcal{O}_{\Gamma^v}$ to show that the maxima $\widehat{\boldsymbol{\theta}}^{(k+1)}$ exist. Regarding the latter term, in order to prove the existence of $\widehat{\boldsymbol{\phi}}_n^{(k+1)}$ we can still use the same compactness argument by noticing that maximizing this term amounts to separately maximizing of n terms, each depending only of $\underline{\boldsymbol{\lambda}}_i$ and $\underline{\sigma}_i^2$ for given $i \in \mathbb{N}$, and thus defined on the finite-dimensional compact set $\mathcal{O}_{\lambda_i} \times \mathcal{O}_{\sigma_i^2}$. It is also easy to show that the elements of $\widehat{\boldsymbol{\varphi}}_n^{(k+1)}$ are unique because the M-step gives a closed form solution for each of them. This reasoning guarantees the existence and uniqueness of the EM estimators (see Lemma E.20).

Let us turn to the QML estimator that maximizes the full log-likelihood (5). The existence of the QML estimator $\widehat{\boldsymbol{\theta}}^*$ poses no problem because being a finite-dimensional vector, we can use the compactness argument. Direct proof of the existence of the QML estimator $\widehat{\boldsymbol{\phi}}_n^*$ is instead more challenging, as in this case, we cannot rely on compactness because the full log-likelihood is defined on a set of increasing dimension n.

Nevertheless, we give an indirect proof of the existence of $\widehat{\boldsymbol{\phi}}_n^*$ by noticing that, under our identification Assumptions 6(a)-6(c), as $n, T \to \infty$, the elements of $\widehat{\boldsymbol{\phi}}_n^*$ are asymptotically equivalent to the unfeasible Ordinary

Least Squares estimators we would obtain if the factors were observed (see Lemma E.11). Therefore, this argument ensures, at least asymptotically, both the existence and also the uniqueness of the QML estimators. We also refer to the next section for more details.

Remark 5. Typically, this literature assumes that the QML estimators $\widehat{\sigma}_i^{2*}$ belong to a compact set (Bai and Li, 2016, Assumption D). In our set-up, this assumption is implied (Gao et al., 2021, Theorem 3.1, and Mao et al., 2024, Theorem 1), and not needed for proving our results (Barigozzi, 2023).

5 Asymptotic properties

This section presents the asymptotic properties of the EM estimator of the parameters—i.e., of the factor loadings, idiosyncratic variances, VAR coefficients, and the covariance matrix of the VAR innovations—and of the Kalman smoother estimator of the factors. We assume that r, the number of common factors, is known; without loss of generality, we fix the VAR order in (4) to $p_F = 1$, and we let $\mathbf{A} \equiv \mathbf{A}_1$ so that $\mathbf{A}(L) \equiv \mathbf{A}L$. We briefly discuss the estimation of r and p_F at the end of the section.

5.1 Consistency under basic assumptions

We start by proving the consistency of the EM algorithm and the Kalman smoother under the most general case where we neither impose linearity of the conditional mean (Assumption 4) nor exponentially decaying tails (Assumption 5).

Proposition 1. *Consider the EM estimators of the parameters* $\widehat{\boldsymbol{\Lambda}}_n = (\widehat{\boldsymbol{\lambda}}_1 \cdots \widehat{\boldsymbol{\lambda}}_n)'$ *with* $\widehat{\boldsymbol{\lambda}}_i \equiv \widehat{\boldsymbol{\lambda}}_i^{(k^*+1)}$, $\widehat{\sigma}_i^2 \equiv \widehat{\sigma}_i^{2(k^*+1)}$, $i = 1, \dots, n$, $\widehat{\mathbf{A}} \equiv \widehat{\mathbf{A}}^{(k^*+1)}$, *and* $\widehat{\boldsymbol{\Gamma}}^v \equiv \widehat{\boldsymbol{\Gamma}}^{v(k^*+1)}$, *and the Kalman smoother estimator of the factors,* $\widehat{\mathbf{F}}_t \equiv \mathbf{F}_{t|T}^{(k^*+1)}$, $t = 1, \dots, T$, $k^* \geq 0$. *Then, under Assumptions 1, 2, 3, and 6:*

(a) for all $\epsilon > 0$, *there exist a positive real* $\eta(\epsilon)$, *and integers* $n^*(\epsilon)$ *and* $T^*(\epsilon)$, *all independent of* i, *such that, for all* $n \geq n^*(\epsilon)$ *and all* $T \geq T^*(\epsilon)$,

$$
\begin{aligned}
&(a.1) \quad \mathrm{P}\left(\min(n, \sqrt{T})\, \|\widehat{\boldsymbol{\lambda}}_i - \boldsymbol{\lambda}_i\| \geq \eta(\epsilon)\right) < \epsilon, \quad \text{for any given } i = 1, \dots, n,\\
&(a.2) \quad \mathrm{P}\left(\min(n, \sqrt{T})\, n^{-1/2} \|\widehat{\boldsymbol{\Lambda}}_n - \boldsymbol{\Lambda}_n\| \geq \eta(\epsilon)\right) < \epsilon,\\
&(a.3) \quad \mathrm{P}\left(\min(n, \sqrt{T})\, |\widehat{\sigma}_i^2 - \sigma_i^2| \geq \eta(\epsilon)\right) < \epsilon, \quad \text{for any given } i = 1, \dots, n,\\
&(a.4) \quad \mathrm{P}\left(\min(n, \sqrt{T})\, \|\widehat{\mathbf{A}} - \mathbf{A}\| \geq \eta(\epsilon)\right) < \epsilon,\\
&(a.5) \quad \mathrm{P}\left(\min(n, \sqrt{T})\, \|\widehat{\boldsymbol{\Gamma}}^v - \boldsymbol{\Gamma}^v\| \geq \eta(\epsilon)\right) < \epsilon;
\end{aligned}
$$

(b) for all $\epsilon > 0$, *there exist a positive real* $\eta(\epsilon)$, *and integers* $n^{**}(\epsilon)$ *and* $T^{**}(\epsilon)$, *all independent of* t, *such that, for all* $n \geq n^{**}(\epsilon)$ *and all* $T \geq T^{**}(\epsilon)$,

$$
\mathrm{P}\left(\min(\sqrt{n}, \sqrt{T})\, \|\widehat{\mathbf{F}}_t - \mathbf{F}_t\| \geq \eta(\epsilon)\right) < \epsilon,
$$

for any given $t = 1, \dots, T$.

The consistency rate of the estimated parameter is the same as the PC estimator (Bai, 2003, Theorem 2). This is because by initializing the algorithm with the consistent PC estimator, we can consider the EM estimator a "one-step" estimator (Lehmann and Casella, 2006, Theorem 4.3). As for the the estimated factors, they converge at a slower rate than the PC estimator, due to the $\sqrt{T}$ term (Bai, 2003, Theorem 1).

Remark 6. The estimation error for the factors and the one for the loadings both depend on the estimation error of the diagonal idiosyncratic covariance matrix $\|\widehat{\boldsymbol{\Sigma}}_n^\xi - \boldsymbol{\Sigma}_n^\xi\| = \max_{i=1,\dots,n} |\widehat{\sigma}_i^2 - \sigma_i^2|$. The latter produces a term in the estimation error of the factors which is $O_p(T^{-1/2})$ (term $D.2$ in the proof of Lemma D.17) and a term in the estimation error of the loadings which is also $O_p(T^{-1/2})$ (term III_d in the proof of Proposition 1). These two terms have non-standard asymptotic distributions and are non-negligible. Thus, we cannot prove asymptotic normality of our estimators without additional assumptions.

Remark 7. The results of Proposition 1 apply to the case in which the observed data, $\mathbf{y}_{nt} = (y_{1t} \cdots y_{nt})'$, are such that $y_{it} = y_{i0} + \mu_i t + \boldsymbol{\lambda}_i' \mathbf{G}_t + z_{it}$. In this case, if we write $\mathbf{F}_t = \Delta \mathbf{G}_t$ and $\xi_{it} = \Delta z_{it}$, then Proposition 1 holds for $x_{it} = \Delta y_{it}$. This strategy always works for the loadings (see Barigozzi et al., 2021, for details) so part (a) still holds, but estimation of the factors must be modified if $\mathbf{G}_t$ is a cointegrated vector. First, we must model $\mathbf{G}_t$ as a VAR in levels. Second, we estimate $\mathbf{G}_t$ running the Kalman smoother in levels, and whenever $z_{it} \sim I(1)$ for some i, we add a latent state. If $z_{it} \sim I(0)$ for all i, then part (b) stands; if $z_{it} \sim I(1)$ for some i, we conjecture that part (b) would remain unchanged. We leave the derivation of the asymptotic properties of the estimated factors in this last case for further research.

5.2 Consistency and asymptotic normality

If we also assume that Assumptions 4 and 5 hold, we can refine the previous result because we can now guarantee that the EM algorithm converges to the QML estimator.

Proposition 2 (Loadings)**.** *Consider the EM estimators of the loadings* $\widehat{\boldsymbol{\Lambda}}_n = (\widehat{\boldsymbol{\lambda}}_1 \cdots \widehat{\boldsymbol{\lambda}}_n)'$ *with* $\widehat{\boldsymbol{\lambda}}_i \equiv \widehat{\boldsymbol{\lambda}}_i^{(k^*+1)}$, $k^* \geq 0$. *Then, under Assumptions 1, 2, 3, 4, 5, and 6:*

(a) for all $\epsilon > 0$, there exist a positive real $\eta(\epsilon)$, and integers $n^*(\epsilon)$ and $T^*(\epsilon)$, all independent of i, such that, for all $n \geq n^*(\epsilon)$ and all $T \geq T^*(\epsilon)$, and some $0 < \delta_v \leq 2$,

$$(a.1) \quad \mathrm{P}\left(\min(n/\log^{2/\delta_v} T, \sqrt{T})\, \|\widehat{\boldsymbol{\lambda}}_i - \boldsymbol{\lambda}_i\| \geq \eta(\epsilon)\right) < \epsilon, \quad \textit{for any given } i = 1, \dots, n,$$

$$(a.2) \quad \mathrm{P}\left(\min(n/\log^{2/\delta_v} T, \sqrt{T})\, n^{-1/2} \|\widehat{\boldsymbol{\Lambda}}_n - \boldsymbol{\Lambda}_n\| \geq \eta(\epsilon)\right) < \epsilon;$$

(b) for any given $i = 1, \dots, n$, as $n, T \to \infty$, if $n^{-1}\sqrt{T} \log^{2/\delta_v} T \to 0$,

$$\sqrt{T}(\widehat{\boldsymbol{\lambda}}_i - \boldsymbol{\lambda}_i) \xrightarrow{d} \mathcal{N}(\mathbf{0}_r, \boldsymbol{\mathcal{V}}_i),$$

where

$$\boldsymbol{\mathcal{V}}_i = (\boldsymbol{\Gamma}^F)^{-1} \left(\lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^T \sum_{s=1}^T \mathbb{E}[\xi_{it} \xi_{is}] \mathbb{E}[\mathbf{F}_t \mathbf{F}_s'] \right) (\boldsymbol{\Gamma}^F)^{-1},$$

with $\boldsymbol{\Gamma}^F = \lim_{T \to \infty} T^{-1} \sum_{t=1}^T \mathbf{F}_t \mathbf{F}_t' = \mathbf{I}_r$, because of Assumption 6(b);

(c) for any given $i = 1, \dots, n$, if $\mathbb{E}[\xi_{it} \xi_{is}] = 0$ for all $t, s = 1, \dots, T$ with $t \neq s$, then, $\boldsymbol{\mathcal{V}}_i = \sigma_i^2 (\boldsymbol{\Gamma}^F)^{-1}$, with $\boldsymbol{\Gamma}^F = \mathbf{I}_r$, because of Assumption 6(b).

The rate of consistency of the estimated loadings, $\min(n/\log^{2/\delta_v} T, \sqrt{T})$, given in Proposition 2, is new to the EM literature. This rate is the same, up to a logarithmic factor, as the one of the PC estimator (Bai, 2003, Theorem 2), which, in turn, is equivalent to the unfeasible Ordinary Least Squares (OLS) we would obtain if the factors were observed. The EM estimator is also asymptotically equivalent to the QML estimator considered by Bai and Li (2016, Theorem 1) when imposing no autocorrelation for the factors. For an explanation of the logarithmic term we refer to Remark 9 below. Efficiency is discussed in Section 6.

It is important to stress that Proposition 2 requires not only $T \to \infty$, as in classical QML estimation theory, but also $n \to \infty$ otherwise no consistency can be proved. In particular, as $n \to \infty$ the factors can be treated as observed, therefore, there is no more an issue of missing information and the QML estimator of the loadings must coincide with the unfeasible OLS. This is a manifestation of the blessing of dimensionality which is a fundamental feature of approximate factor models.

The proof of Proposition 2 is based on the following decomposition of the estimation error into four terms:

$$\sqrt{T}(\widehat{\boldsymbol{\lambda}}_i - \boldsymbol{\lambda}_i) = \underbrace{\sqrt{T}(\boldsymbol{\lambda}_i^{\text{OLS}} - \boldsymbol{\lambda}_i)}_{O_p(1)} + \underbrace{\sqrt{T}(\widehat{\boldsymbol{\lambda}}_i^{*} - \boldsymbol{\lambda}_i^{\text{OLS}})}_{O_p(n^{-1}\sqrt{T}\log^{2/\delta_v} T)} + \underbrace{\sqrt{T}(\widehat{\boldsymbol{\lambda}}_i^{**} - \widehat{\boldsymbol{\lambda}}_i^{*})}_{O_p(n^{-1}\sqrt{T}\log^{2/\delta_v} T)} + \underbrace{\sqrt{T}(\widehat{\boldsymbol{\lambda}}_i - \widehat{\boldsymbol{\lambda}}_i^{**})}_{o_p(n^{-1}\sqrt{T}\log^{2/\delta_v} T)}, \tag{20}$$

which shows that the EM estimator $\widehat{\boldsymbol{\lambda}}_i$ is asymptotically equivalent to the OLS estimator we would obtain had we observed the factors.

To prove our result, we first show that the QML estimator of the loadings, $\widehat{\boldsymbol{\lambda}}_i^{*}$ is asymptotically equivalent to the PC estimator, which, in turn, is asymptotically equivalent to the unfeasible $\sqrt{T}$-consistent OLS estimator $\boldsymbol{\lambda}_i^{\text{OLS}}$ (second term on the rhs of (20) proved in Lemma E.11, see also Barigozzi, 2023, Theorem 3 and Corollary 1). In particular, both approximation errors are $o_p(T^{-1/2})$, whenever $n^{-1}\sqrt{T}\log^{2/\delta_v} T \to 0$. This result extends to the DFM the result by Bai and Li (2016, Theorem 1) obtained for QML estimation of a static factor model i.e., when we replace the full-matrix $\boldsymbol{\Omega}_T^F(\underline{\boldsymbol{\mathcal{A}}}, \underline{\mathbf{H}})$ with just $\mathbf{I}_{rT}$ in the log-likelihood (5). Notice that, differently from the proofs in Bai and Li (2016) and, as mentioned in Remark 5, this result does not depend on the QML estimator $\widehat{\sigma}_i^{*2}$.

Next, we show that the EM estimator converges to a global maximum of the likelihood (third term on the rhs of (20)). As we discussed in Section 5.3, we know that the sequence of estimators $\{\widehat{\boldsymbol{\lambda}}_i^{(k)}, k \geq 0\}$ converges to a local maximum of the likelihood, say $\widehat{\boldsymbol{\lambda}}_i^{**}$, as $k \to \infty$ (see Lemma E.21). However, in general, the likelihood might have many maxima due to the identification indeterminacy of the loadings. Nevertheless, once we make the identifying Assumptions 6(b) and 6(c), there is only a unique maximum, $\widehat{\boldsymbol{\lambda}}_i^{*}$, which is asymptotically equivalent to the unique OLS estimator, whenever $n^{-1}\sqrt{T}\log^{2/\delta_v} T \to 0$ (see Lemma E.22 and Ruud, 1991, Section 4, for a similar result in the case of one-to-one mapping from the factors to the data, corresponding to the case of no idiosyncratic component). This is also clear from the asymptotic expansions of $\widehat{\boldsymbol{\lambda}}_i^{*} - \boldsymbol{\lambda}_i$ obtained by Bai and Li (2012, 2016) for the static model (i.e., the factors have no dynamics) and using identification schemes different than the one we use here.

Third, we show that asymptotically the error coming from running the EM a finite number of times vanishes (fourth term on the rhs of (20)). Indeed, due to the finite number of iterations, k^*, the EM algorithm delivers an estimator $\widehat{\boldsymbol{\lambda}}_i \equiv \widehat{\boldsymbol{\lambda}}_i^{(k^*+1)}$, which is just an approximation of the local maximum $\widehat{\boldsymbol{\lambda}}_i^{**}$ that we would attained after an infinite number of iterations. We show that the error entailed by such approximation depends on the ratio of the Hessians of the complete and incomplete log-likelihoods, i.e., on how much information is missing because the factors are not observed (Meng and Rubin, 1994, McLachlan and Krishnan, 2007, Chapter 3.9, and Sundberg, 2019, Chapter 8). In this case the approximation error is $o_p(T^{-1/2})$, provided $n^{-1}\sqrt{T}\log^{2/\delta_v} T \to 0$ (see Lemma E.23). This last result is a refinement of the results by Balakrishnan et al. (2017, Theorem 2) on the convergence of the EM algorithm. We refer to Section 5.3 below for more details.

Last, $\boldsymbol{\lambda}_i^{\text{OLS}}$ is a $\sqrt{T}$-consistent estimator of the factor loadings $\boldsymbol{\lambda}_i$ (first term on the rhs of (20)).

Proposition 3 (Factors). *Consider the Kalman smoother estimator of the factors* $\widehat{\boldsymbol{\mathcal{F}}}_T = (\widehat{\mathbf{F}}_1 \cdots \widehat{\mathbf{F}}_T)'$, *with* $\widehat{\mathbf{F}}_t \equiv \mathbf{F}_{t|T}^{(k^*+1)}$, $t = 1, \ldots, T$, $k^* \geq 0$. *Then, under Assumptions 1, 2, 3, 4, 5, and 6:*

(a) for all $\epsilon > 0$, *there exist a positive real* $\eta(\epsilon)$, *and integers* $n^{**}(\epsilon)$ *and* $T^{**}(\epsilon)$, *all independent of* t, *such that, for all* $n \geq n^{**}(\epsilon)$ *and all* $T \geq T^{**}(\epsilon)$,

$$(a.1) \quad \mathrm{P}\left(\min(\sqrt{n}, T/\sqrt{\log n})\, \|\widehat{\mathbf{F}}_t - \mathbf{F}_t\| \geq \eta(\epsilon)\right) < \epsilon, \quad \text{for any given } t = 1, \ldots, T,$$

$$(a.2) \quad \mathrm{P}\left(\min(\sqrt{n}, T/\sqrt{\log n})\, T^{-1/2}\|\widehat{\boldsymbol{\mathcal{F}}}_T - \boldsymbol{\mathcal{F}}_T\| \geq \eta(\epsilon)\right) < \epsilon;$$

(b) as $n, T \to \infty$, *if* $T^{-1}\sqrt{n \log n} \to 0$,

$$\sqrt{n}(\widehat{\mathbf{F}}_t - \mathbf{F}_t) \xrightarrow{d} \mathcal{N}(\mathbf{0}_r, \boldsymbol{\mathcal{W}}_t),$$

for any given $t = 1, \ldots, T$*, where*

$$\boldsymbol{\mathcal{W}}_t = (\boldsymbol{\Sigma}_{\Lambda\Sigma\Lambda})^{-1} \left(\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\mathbb{E}[\xi_{it}\xi_{jt}]\boldsymbol{\lambda}_i \boldsymbol{\lambda}_j}{\sigma_i^2 \sigma_j^2} \right) (\boldsymbol{\Sigma}_{\Lambda\Sigma\Lambda})^{-1},$$

with $\boldsymbol{\Sigma}_{\Lambda\Sigma\Lambda} = \lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} \boldsymbol{\lambda}_i (\sigma_i^2)^{-1} \boldsymbol{\lambda}_i'$*;*

(c) for any given $t = 1, \ldots, T$*, if* $\mathbb{E}[\xi_{it}\xi_{jt}] = 0$ *for all* $i, j = 1, \ldots, n$ *with* $i \neq j$*, then,* $\boldsymbol{\mathcal{W}}_t = (\boldsymbol{\Sigma}_{\Lambda\Sigma\Lambda})^{-1}$*.*

The rate of consistency of the estimated factors given in Proposition 3 is faster than the rate originally derived by Doz et al. (2012, Proposition 1) for the same estimator—$\min(\sqrt{n}, T/\sqrt{\log n})$ vs. $\min(\sqrt{n}, T^{1/4}/\sqrt{\log n})$. Moreover, this consistency rate is the same (up to a logarithmic factor) as that of the PC estimator (Bai, 2003, Theorem 1). However, while the PC estimator is equivalent to the unfeasible Ordinary Least Squares (OLS) we would obtain if the loadings were observed, the Kalman smoother estimator we are considering is equivalent to the unfeasible Weighted Least Squares (WLS) we would obtain if the loadings were observed and we knew the idiosyncratic variances. As such, the Kalman smoother is also equivalent to the feasible WLS studied by Bai and Li (2016, Theorem 2) and computed using the QML estimator of the loadings for a static factor model. For an explanation of the logarithmic term we refer to Remark 9 below. Efficiency is discussed in Section 6.

The proof of Proposition 3 is based on the following decomposition of the estimation error into four terms

$$\sqrt{n}(\widehat{\mathbf{F}}_t - \mathbf{F}_t) = \underbrace{\sqrt{n}(\mathbf{F}_t^{\text{WLS}} - \mathbf{F}_t)}_{O_p(1)} + \underbrace{\sqrt{n}(\widehat{\mathbf{F}}_t^{\text{WLS}} - \mathbf{F}_t^{\text{WLS}})}_{O_p(T^{-1}\sqrt{n\log n})} + \underbrace{\sqrt{n}(\widehat{\mathbf{F}}_{t|t} - \widehat{\mathbf{F}}_t^{\text{WLS}})}_{O_p(n^{-1/2})} + \underbrace{\sqrt{n}(\widehat{\mathbf{F}}_t - \widehat{\mathbf{F}}_{t|t})}_{O_p(n^{-1/2})}. \tag{21}$$

which shows that the Kalman smoother estimator $\widehat{\mathbf{F}}_t$ is asymptotically equivalent to the WLS estimator we would obtain had we observed the loadings and had we known the idiosyncratic variances.

To prove our result, we first show that our estimator of the factors, which is obtained via the Kalman smoother computed using the EM estimators of the parameters, is asymptotically equivalent to the Kalman filter, $\widehat{\mathbf{F}}_{t|t}$, which in turn is asymptotically equivalent to the WLS estimator $\widehat{\mathbf{F}}_t^{\text{WLS}}$ (fourth and third term on the rhs of (21), respectively). Both approximation errors are $O_p(n^{-1})$ (see Lemma F.4 and Ruiz and Poncela, 2022, Section 2.3, for the one-factor case).

Then, we take into account the estimation error of the parameters given in Proposition 2, which implies that the WLS estimator $\widehat{\mathbf{F}}_t^{\text{WLS}}$ converges to its unfeasible counterpart computed using the true value of the parameters, $\mathbf{F}_t^{\text{WLS}}$, with a rate which is $o_p(n^{-1/2})$, provided $T^{-1}\sqrt{n\log n} \to 0$ (second term on the rhs of (21)). This result refines the result of Proposition 1 because, thanks to Assumption 5, we are now able to derive a tighter bound for $\|\widehat{\boldsymbol{\Sigma}}_n^{\xi} - \boldsymbol{\Sigma}_n^{\xi}\| = \max_{i=1,\ldots,n} |\widehat{\sigma}_i^2 - \sigma_i^2|$, as shown in Proposition 5 below.

Last, $\mathbf{F}_t^{\text{WLS}}$ is a $\sqrt{n}$-consistent estimator of the realizations of the factors $\mathbf{F}_t$ (first term on the rhs of (21)).

Proposition 4 (Common component)**.** *Consider the EM plus the Kalman smoother estimator of the common component* $\widehat{\chi}_{it} \equiv \widehat{\boldsymbol{\lambda}}_i^{(k^*+1)\prime} \mathbf{F}_{t|T}^{(k^*+1)}$*,* $i = 1, \ldots, n$*,* $t = 1, \ldots, T$*, with* $k^* \geq 0$*. Then, under Assumptions 1, 2, 3, 4, 5, and 6:*

(a) for all $\epsilon > 0$*, there exist a positive real* $\eta(\epsilon)$*, and integers* $n^\circ(\epsilon)$ *and* $T^\circ(\epsilon)$*, all independent of* i *and* t*, such that, for all* $n \geq n^\circ(\epsilon)$ *and* $T \geq T^\circ(\epsilon)$*,*

$$\mathrm{P}\left(\min(\sqrt{n}, \sqrt{T})\, |\widehat{\chi}_{it} - \chi_{it}| \geq \eta(\epsilon)\right) < \epsilon,$$

for any given $i = 1, \ldots, n$*,* $t = 1, \ldots, T$*;*

(b) as $n, T \to \infty$*,*

$$(T^{-1}\mathcal{C}_{it}^{\lambda} + n^{-1}\mathcal{C}_{it}^{F})^{-1/2}(\widehat{\chi}_{it} - \chi_{it}) \xrightarrow{d} \mathcal{N}(0, 1),$$

for any given $i = 1, \ldots, n$ *and* $t = 1, \ldots, T$*, where* $\mathcal{C}_{it}^{\lambda} = \mathbf{F}_t' \boldsymbol{\mathcal{V}}_i \mathbf{F}_t$ *and* $\mathcal{C}_{it}^{F} = \boldsymbol{\lambda}_i' \boldsymbol{\mathcal{W}}_t \boldsymbol{\lambda}_i$*, with* $\boldsymbol{\mathcal{V}}_i$ *defined in*

Proposition 2(b), and $\boldsymbol{\mathcal{W}}_t$ in Proposition 3(b).

Proposition 4 does not require a limit for T/n or n/T, so it holds without any constraint between the rates of divergence of n and T. That said, Proposition 4 has two special cases: (a) if $n/T \to 0$, then $\sqrt{n}(\widehat{\chi}_{it} - \chi_{it}) \xrightarrow{d} \mathcal{N}(0, \mathcal{C}_{it}^F)$; and, (b) if $T/n \to 0$, then $\sqrt{T}(\widehat{\chi}_{it} - \chi_{it}) \xrightarrow{d} \mathcal{N}(0, \mathcal{C}_{it}^\lambda)$. This is the same rate of consistency we obtain for the PC estimator of the common component (Bai, 2003, Theorem 3).

All other estimated parameters are also consistently estimated.

Proposition 5 (Idiosyncratic variances and VAR parameters)**.** *Consider the EM estimators of the parameters $\widehat{\boldsymbol{\Sigma}}_n^\xi = \mathrm{diag}(\widehat{\sigma}_1^2 \cdots \widehat{\sigma}_n^2)$, with $\widehat{\sigma}_i^2 \equiv \widehat{\sigma}_i^{2(k^*+1)}$, $i = 1, \ldots, n$, $\widehat{\mathbf{A}} \equiv \widehat{\mathbf{A}}^{(k^*+1)}$, and $\widehat{\boldsymbol{\Gamma}}^v \equiv \widehat{\boldsymbol{\Gamma}}^{v(k^*+1)}$, $k^* \geq 0$. Then, under Assumptions 1, 2, 3, 4, 5, and 6:*

(a) for all $\epsilon > 0$, there exist a positive real $\eta(\epsilon)$, and integers $n^*(\epsilon)$ and $T^*(\epsilon)$, all independent of i, such that, for all $n \geq n^*(\epsilon)$ and all $T \geq T^*(\epsilon)$, and some $0 < \delta_v \leq 2$,

$$
\begin{aligned}
&(a.1) \quad \mathrm{P}\left(\min(n/\log^{2/\delta_v} T, \sqrt{T})\, |\widehat{\sigma}_i^2 - \sigma_i^2| \geq \eta(\epsilon)\right) < \epsilon, \quad \textit{for any given } t = i, \ldots, n,\\
&(a.2) \quad \mathrm{P}\left(\min(n/\log^{2/\delta_v} T, \sqrt{T/\log n})\, \|\widehat{\boldsymbol{\Sigma}}_n^\xi - \boldsymbol{\Sigma}_n^\xi\| \geq \eta(\epsilon)\right) < \epsilon,\\
&(a.3) \quad \mathrm{P}\left(\min(n/\log^{2/\delta_v} T, \sqrt{T})\, \|\widehat{\mathbf{A}} - \mathbf{A}\| \geq \eta(\epsilon)\right) < \epsilon,\\
&(a.4) \quad \mathrm{P}\left(\min(n/\log^{2/\delta_v} T, \sqrt{T})\, \|\widehat{\boldsymbol{\Gamma}}^v - \boldsymbol{\Gamma}^v\| \geq \eta(\epsilon)\right) < \epsilon;
\end{aligned}
$$

(b) as $n, T \to \infty$, if $n^{-1}\sqrt{T}\log^{2/\delta_v} T \to 0$,

$$
\begin{aligned}
&(b.1) \quad \sqrt{T}(\widehat{\sigma}_i^2 - \sigma_i^2) \xrightarrow{d} \mathcal{N}(0, \sigma_i^4(\kappa_i + 2)), \quad \textit{for any given } i = 1, \ldots, n,\\
&(b.2) \quad \sqrt{T}(\mathrm{vec}(\widehat{\mathbf{A}}) - \mathrm{vec}(\mathbf{A})) \xrightarrow{d} \mathcal{N}(\mathbf{0}_{r^2}, \boldsymbol{\Gamma}^v \otimes (\boldsymbol{\Gamma}^F)^{-1}),\\
&(b.3) \quad \sqrt{T}(\mathrm{vech}(\widehat{\boldsymbol{\Gamma}}^v) - \mathrm{vech}(\boldsymbol{\Gamma}^v)) \xrightarrow{d} \mathcal{N}(\mathbf{0}_{r(r+1)/2}, 2\mathbf{D}^\dagger(\boldsymbol{\Gamma}^v \otimes \boldsymbol{\Gamma}^v)(\mathbf{D}^\dagger)'),
\end{aligned}
$$

with $\boldsymbol{\Gamma}^F = \mathbf{I}_r$, because of Assumption 6(b), and where $\kappa_i = \mathbb{E}[\xi_{it}^4]/\sigma_i^4 - 3$ and $\mathbf{D}^\dagger = (\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}$ with $\mathbf{D}$ being $r^2 \times r(r+1)/2$ such that $\mathbf{D}\mathrm{vech}(\boldsymbol{\Gamma}^v) = \mathrm{vec}(\boldsymbol{\Gamma}^v)$.

We conclude with a series of general remarks.

Remark 8. From Propositions 2 and 3, we immediately have that, as $n, T \to \infty$,

$$
\left\| \frac{\widehat{\boldsymbol{\Lambda}}_n'\widehat{\boldsymbol{\Lambda}}_n}{n} - \boldsymbol{\Sigma}_\Lambda \right\| = o_p(1), \qquad \left\| \frac{\widehat{\boldsymbol{\mathcal{F}}}_T'\widehat{\boldsymbol{\mathcal{F}}}_T}{T} - \boldsymbol{\Gamma}^F \right\| = o_p(1),
$$

by Assumption 1(a), which defines $\boldsymbol{\Sigma}_\Lambda$, and because the sample covariance matrix $T^{-1}\boldsymbol{\mathcal{F}}_T'\boldsymbol{\mathcal{F}}_T$ is a consistent estimator of $\boldsymbol{\Gamma}^F$ (see Lemma C.12). Moreover, by Assumption 6(b), $\boldsymbol{\Sigma}_\Lambda = \lim_{n\to\infty} n^{-1}\mathbf{M}_n^\chi$, which is a positive definite diagonal matrix and $\boldsymbol{\Gamma}^F = \mathbf{I}_r$. Therefore, the EM estimator of the loadings and the related Kalman Smoother estimator of the factors satisfy the identifying constraints asymptotically, as $n, T \to \infty$. We verify this result numerically in Section 8. This is in agreement with Assumption 6(b) which imposes the identifying constraints only in the limit $n, T \to \infty$. This approach differs from other works on factor models (see, e.g., Bai and Li, 2012, 2016, or Bai and Ng, 2013) where the identifying constraints are assumed to hold for any given n and T (see Remark 3).

In principle, we could obtain estimators satisfying the identifying constraints in finite samples by imposing them ex-post in an additional step, as Bai and Li (2012, Section 8) suggested in QML estimation of a static factor model. However, empirical works using the EM algorithm rarely apply this additional step.

Remark 9. The constraints $T^{-1}\sqrt{n}\log n \to 0$ and $n^{-1}\sqrt{T}\log^{2/\delta_v} T \to 0$ are common (up to the presence of logarithmic terms) in the factor model literature (see, e.g., Bai, 2003, Theorems 1 and 2, for PC estimation) and are compatible. Indeed, they are simultaneously fulfilled if we assume that there exist some finite positive reals $\underline{\gamma} > 1/2$ and $\overline{\gamma} < 2$ such that $T^{\underline{\gamma}} < n < T^{\overline{\gamma}}$, as $T \to \infty$. When n and T have the same order of magnitude, as in many macroeconomic and financial datasets, these assumptions on the relative rates of divergence of n and T are very mild.

In particular, the logarithmic term in the consistency rates of Propositions 2 and 3 comes from Assumption 5 of sub-Weibull tails, which is slightly more general than the typical assumption of sub-Gaussianity made when studying estimators of high-dimensional models. This is, however, a modest price to pay. In particular, under Gaussianity, $\delta_v = 2$, while under distributions with sub-exponential tails, $\delta_v = 1$. These logarithmic terms come essentially from two errors. The first, which is $O_p(n^{-1} \log^{2/\delta_v} T)$, is due to the necessity of finding a uniform bound over t for the difference between the log-likelihood of the DFM in (5) and the static factor model log-likelihood considered in Bai and Li (2016). This terms involves the sum of squared factors (see Lemma E.9). The second one, which is $O_p(T^{-1/2}\sqrt{\log n})$, is due to the uniform bound for $\max_{i=1,\dots,n} |\widehat{\sigma}_i^2 - \sigma_i^2|$ obtained when estimating the factors (see Lemma F.1 and also Fan et al., 2011, Lemmas A3 and B1).

Remark 10. The asymptotic properties of the estimators are unaffected if the number of factors r is estimated. Indeed, consider a consistent estimator $\widehat{r}$ i.e., such that $\mathrm{P}(\widehat{r} = r) \to 1$, as $n, T \to \infty$, as for example the one in Bai and Ng (2002). Then, for any $z \in \mathbb{R}$ and any $i = 1, \dots, n$ and $t = 1, \dots, T$, it is easy to prove that $\mathrm{P}(\widehat{\chi}_{it} \leq z) = \mathrm{P}(\{\widehat{\chi}_{it} \leq z\} \,|\, \{\widehat{r} = r\}) + o_p(1)$ (see Bai, 2003, footnote 5).

Similarly, the asymptotic properties of the estimators are unaffected if the order p_F of the the VAR for the factors is estimated. This approach is asymptotically equivalent to computing the BIC using the true factors $\mathbf{F}_t$ because we could estimate p_F through the consistent PC estimator of the factors $\widetilde{\mathbf{F}}_t$ (see Lemma H.1 and Bai, 2003, Theorem 1). And, in turn, the BIC is known to select the true lag order consistently (Hannan, 1980, Theorem 1). Therefore, $\mathrm{P}(\widehat{p}_F = p_F) \to 1$, as $n, T \to \infty$. Following the same reasoning of the previous remark, it is easy to show that, for any $z \in \mathbb{R}$ and any $i = 1, \dots, n$ and $t = 1, \dots, T$, $\mathrm{P}(\widehat{\chi}_{it} \leq z) = \mathrm{P}(\{\widehat{\chi}_{it} \leq z\} \,|\, \{\widehat{p}_F = p_F\}) + o_p(1)$.

Remark 11. Given the asymptotic equivalence of Kalman filter, smoother, and WLS estimators of the factors, we might expect that the MSE obtained from either the Kalman filter or the smoother, i.e., $\mathbf{P}_{t|t}$ or $\mathbf{P}_{t|T}$, respectively, asymptotically coincide (inflated by n) with the asymptotic covariance matrix $\boldsymbol{\mathcal{W}}_t$ of $\widehat{\mathbf{F}}_t$ defined in Proposition 3. However, this is not the case since we estimate a mis-specified model. Indeed, as $n \to \infty$, we can shown that both $n\mathbf{P}_{t|t}$ and $n\mathbf{P}_{t|T}$ are asymptotically equivalent to $(\boldsymbol{\Sigma}_{\Lambda\Sigma\Lambda})^{-1}$ (see Lemma E.8), which, as shown in Proposition 3(c) is the asymptotic covariance matrix of $\widehat{\mathbf{F}}_t$ only if the model is correctly specified. In other words, the mis-specified Kalman filter and smoother do not estimate the true MSE. Although this has no effect on our asymptotic results (see Remark 1), still we cannot use the estimated MSEs $\mathbf{P}_{t|t}^{(k^*+1)}$ or $\mathbf{P}_{t|T}^{(k^*+1)}$ for making inference. The true Kalman filter MSE, accounting for the model mis-specification, is derived by Harvey and Delle Monache (2009, Section 2.1) and, for any $t = 1, \dots, T$, it is given by the recursions

$$\begin{aligned}
\boldsymbol{\Pi}_{t|t} =& \boldsymbol{\Pi}_{t|t-1} + \mathbf{P}_{t|t-1}\boldsymbol{\Lambda}_n'(\boldsymbol{\Lambda}_n\mathbf{P}_{t|t-1}\boldsymbol{\Lambda}_n' + \boldsymbol{\Sigma}_n^{\xi})^{-1}(\boldsymbol{\Lambda}_n\boldsymbol{\Pi}_{t|t-1}\boldsymbol{\Lambda}_n' + \boldsymbol{\Gamma}_n^{\xi})(\boldsymbol{\Lambda}_n\mathbf{P}_{t|t-1}\boldsymbol{\Lambda}_n' + \boldsymbol{\Sigma}_n^{\xi})^{-1}\boldsymbol{\Lambda}_n\mathbf{P}_{t|t-1} \\
&- \mathbf{P}_{t|t-1}\boldsymbol{\Lambda}_n'(\boldsymbol{\Lambda}_n\mathbf{P}_{t|t-1}\boldsymbol{\Lambda}_n' + \boldsymbol{\Sigma}_n^{\xi})^{-1}\boldsymbol{\Lambda}_n\boldsymbol{\Pi}_{t|t-1} - \boldsymbol{\Pi}_{t|t-1}\boldsymbol{\Lambda}_n'(\boldsymbol{\Lambda}_n\mathbf{P}_{t|t-1}\boldsymbol{\Lambda}_n' + \boldsymbol{\Sigma}_n^{\xi})^{-1}\boldsymbol{\Lambda}_n\mathbf{P}_{t|t-1}, \qquad (22)\\
\boldsymbol{\Pi}_{t|t-1} =& \mathbf{A}\boldsymbol{\Pi}_{t-1|t-1}\mathbf{A}' + \boldsymbol{\Gamma}^v,
\end{aligned}$$

where $\mathbf{P}_{t|t-1}$ is the one-step-ahead Kalman filter MSE (see (A.2)). As expected, $n\boldsymbol{\Pi}_{t|t}$ is asymptotically equivalent, as $n \to \infty$, to the asymptotic covariance matrix $\boldsymbol{\mathcal{W}}_t$ of $\widehat{\mathbf{F}}_t$ (see Lemma I.5). However, since both $\mathbf{P}_{t|t-1}$ and $\boldsymbol{\Pi}_{t|t-1}$ depend on $\mathbf{A}$ and $\boldsymbol{\Gamma}^v$, for finite n this MSE accounts explicitly also for the autocorrelation on the factors.

Remark 12. The results of Propositions 1, 2, 3, 4, and 5 would also hold if we had allowed for serial heteroskedasticity of the idiosyncratic components, i.e., for time-varying second moments so that $\mathbb{E}[\boldsymbol{\xi}_{nt}\boldsymbol{\xi}_{ns}'] = \boldsymbol{\Gamma}_{n,ts}^{\xi}$. For this case, which we do not consider explicitly, we refer to Bai and Li (2016), who show that the estimators of the idiosyncratic variances, $\widehat{\sigma}_i^2$, $i = 1, \dots, n$, have to be considered as estimators of the average variances

$\bar{\sigma}_i^2 = T^{-1}\sum_{t=1}^T \sigma_{i,t}^2$; hence, in all above results, we should replace σ_i^2 with $\bar{\sigma}_i^2$. This approach amounts to maximizing a log-likelihood that has an additional degree of mis-specification because we use the time average idiosyncratic variances rather than the true time-varying variances. In this case, the asymptotic covariance matrix $\boldsymbol{\mathcal{W}}_t$ of the estimated factors in Proposition 3 becomes effectively a time-varying matrix.

5.3 Convergence of the EM algorithm under generic initialization

In this section, we discuss how our asymptotic results change if we initialize the EM algorithm with a generic initial estimator of the parameters, say $\check{\boldsymbol{\varphi}}_n^{(0)} = (\text{vec}(\check{\boldsymbol{\Lambda}}_n^{(0)})' \ \check{\sigma}_1^{2(0)} \cdots \check{\sigma}_n^{2(0)} \ \text{vec}(\check{\mathbf{A}}^{(0)})' \ \text{vech}(\boldsymbol{\Gamma}^{v(0)})')'$, having still elements belonging to $\mathcal{O}_n$ as defined in Section 4.3.4, and, thus, satisfying Assumptions 1, 2, and 6.

For fixed n and in a general setting, Balakrishnan et al. (2017) prove that the EM algorithm defines a contraction path towards a local maximum of the likelihood, $\widehat{\boldsymbol{\varphi}}_n^{**}$. To give more details and understand the relation with our results, we need to introduce some general definitions. First, consider any initial estimator belonging to a closed neighborhood of the local maximum of given Euclidean radius $\varrho > 0$, i.e., $\check{\boldsymbol{\varphi}}_n^{(0)} \in \mathcal{B}(\varrho; \widehat{\boldsymbol{\varphi}}_n^{**}) \subset \mathbb{R}^Q$. In our setting, we can think of $\mathcal{B}(\varrho; \widehat{\boldsymbol{\varphi}}_n^{**}) \equiv \mathcal{O}_n$. Then, define the EM operator $\boldsymbol{M}_T : \mathbb{R}^Q \to \mathbb{R}^Q$ such that $\boldsymbol{M}_T(\widehat{\boldsymbol{\varphi}}_n^{(k)}) = \widehat{\boldsymbol{\varphi}}_n^{(k+1)}$. We have $\|\mathbb{E}[M_T(\underline{\boldsymbol{\varphi}}_n)] - \widehat{\boldsymbol{\varphi}}_n^{**}\| \le \beta \|\underline{\boldsymbol{\varphi}}_n - \widehat{\boldsymbol{\varphi}}_n^{**}\|$ for some $\beta \in (0,1)$ and all $\underline{\boldsymbol{\varphi}}_n \in \mathcal{B}(\varrho; \widehat{\boldsymbol{\varphi}}_n^{**})$ (Balakrishnan et al., 2017, Theorem 1). Second, for any given T and $\delta \in (0,1)$, let $\varepsilon_{T,\delta}$ be the smallest scalar such that $\mathrm{P}(\sup_{\underline{\boldsymbol{\varphi}}_n \in \mathcal{B}(\varrho;\widehat{\boldsymbol{\varphi}}_n^{**})} \|M_T(\underline{\boldsymbol{\varphi}}_n) - \mathbb{E}[M_T(\underline{\boldsymbol{\varphi}}_n)]\| \le \varepsilon_{T,\delta}) \ge 1 - \delta$.

It follows that, if T is large enough such that $\varepsilon_{T,\delta} \le (1-\beta)\varrho$, then, for any $k \ge 0$, the EM operator defines a contraction towards the maximum of the likelihood with high-probability (Balakrishnan et al., 2017, Theorem 2):

$$\mathrm{P}\left(\|\widehat{\boldsymbol{\varphi}}_n^{(k+1)} - \widehat{\boldsymbol{\varphi}}_n^{**}\| \le \beta^{k+1} \|\check{\boldsymbol{\varphi}}_n^{(0)} - \widehat{\boldsymbol{\varphi}}_n^{**}\| + \frac{\varepsilon_{T,\delta}}{1-\beta} \right) \ge 1 - \delta. \tag{23}$$

Now, under our mixing Assumptions 1(d)-1(h) and 2(c), $\varepsilon_{T,\delta} \to 0$, as $T \to \infty$. This, jointly with (23), has two implications, as $T \to \infty$:

(a) for all $k \ge 0$, $\|\widehat{\boldsymbol{\varphi}}_n^{(k+1)} - \widehat{\boldsymbol{\varphi}}_n^{**}\| \le \beta^{k+1} \|\check{\boldsymbol{\varphi}}_n^{(0)} - \widehat{\boldsymbol{\varphi}}_n^{**}\| + o_p(1)$;

(b) if $k \ge \log_{1/\beta} \varepsilon_{T,\delta}^{-1} \equiv k_T$, then $\|\widehat{\boldsymbol{\varphi}}_n^{(k+1)} - \widehat{\boldsymbol{\varphi}}_n^{**}\| = o_p(1)$.

These two results apply if n is fixed. But when $n \to \infty$, we might conjecture that those results will hold for each component of $\boldsymbol{\varphi}_n$ separately. And this is, indeed, what we verify in this paper. Consider the loadings $\boldsymbol{\lambda}_i$. As mentioned in the previous section, as $n \to \infty$ the likelihood has a unique global maximum, $\widehat{\boldsymbol{\lambda}}_i^*$, which is consistent because it is the QML estimator. Therefore, $\|\widehat{\boldsymbol{\lambda}}_i^{**} - \boldsymbol{\lambda}_i\| \le \|\widehat{\boldsymbol{\lambda}}_i^{**} - \widehat{\boldsymbol{\lambda}}_i^*\| + \|\widehat{\boldsymbol{\lambda}}_i^* - \boldsymbol{\lambda}_i\| = o_p(1)$, as $n, T \to \infty$. If we initialize the EM algorithm with the consistent PC estimator $\widehat{\boldsymbol{\lambda}}_i^{(0)}$, we prove that, as $n, T \to \infty$, result (a) above still applies, i.e., there exists a $\beta_\lambda \in (0,1)$ such that, for all $k \ge 0$,

$$\|\widehat{\boldsymbol{\lambda}}_i^{(k+1)} - \widehat{\boldsymbol{\lambda}}_i^{**}\| \le \left\{ \|\widehat{\boldsymbol{\lambda}}_i^{(0)} - \boldsymbol{\lambda}_i\| + \|\widehat{\boldsymbol{\lambda}}_i^{**} - \boldsymbol{\lambda}_i\| \right\} \beta_\lambda^{k+1} + o_p(1) = o_p(1). \tag{24}$$

Actually, we are also able to show that the contraction factor is such that, as $n, T \to \infty$:

$$\beta_\lambda = \left\| \mathbf{I}_r - \left(\sum_{t=1}^T \left\{ \mathbf{F}_{t|T}^* \mathbf{F}_{t|T}^{*\prime} + \mathbf{P}_{t|T}^* \right\} \right)^{-1} \left(\sum_{t=1}^T \mathbf{F}_t \mathbf{F}_t' \right) \right\| + o_p(1) = o_p(1), \tag{25}$$

with $\mathbf{F}_{t|T}^*$ and $\mathbf{P}_{t|T}^*$ being the factors and their associated MSEs obtained from the Kalman smoother when using the QML estimator of the parameters—we refer to Lemma E.23 for a proof of (24) and (25). It follows that the convergence rate in (24) is faster than the one of the initial PC estimator, and we can treat the EM estimator as if it were the QML estimator. Moreover, (25) means that (24) holds for any initial estimator (see Lemma E.26). The intuition for this result is that, provided $n^{-1}\check{\boldsymbol{\Lambda}}^{(0)\prime}\check{\boldsymbol{\Lambda}}^{(0)}$ has full-rank, any cross-sectional averaging is likely to recover in high-dimensions a factors' space not too different from the true one, an argument often used when considering estimation methods based on aggregations schemes alternative to PCs (Westerlund and Urbain, 2015; Fan and Liao, 2022).

For all other parameters, a relation like (24) holds as well, with possibly different contraction factors. Hence, result (a) still applies (see Lemma E.24). However, we could not derive a result like (25) in this case because computing analytic expressions of all those contraction factors is difficult. Thus, when we initialize the algorithm with any generic initial estimator, we have to rely on result (b); that is, we can guarantee convergence of the EM algorithm to the QML estimator, as $n, T \to \infty$, only if we let the algorithm run for a number of iterations $k \geq k_T$, where the larger k is, the faster is the convergence rate.

6 Efficiency and comparison with PC analysis

In this section, we compare the asymptotic covariances of the EM and Kalman smoother with those of the PC estimators, which are the optimal non-parametric estimators.

From Propositions 2(b) and 3(b), we see that consistency is not affected by estimating a mis-specified model with uncorrelated idiosyncratic components, but there is an efficiency loss due to this mis-specification, as shown by the sandwich forms of the asymptotic covariance matrices. In the case of uncorrelated idiosyncratic components, the log-likelihood we maximize is correctly specified. If the model is correctly specified, the EM estimator is the most efficient one because it is asymptotically equivalent to the Maximum Likelihood estimator. Thus, its asymptotic covariance attains the classical lower bound of the OLS estimator (see Propositions 2(c)). Likewise, the asymptotic covariance of the factors attains the WLS lower bound (see Propositions 3(c)).

Although, in general, the EM algorithm and the Kalman smoother do not provide the most efficient estimators, they can provide advantages with respect to the PC estimators.

Proposition 6 (Efficiency). *Let $\boldsymbol{\mathcal{V}}_i^{\text{PC}}$ and $\boldsymbol{\mathcal{W}}_t^{\text{PC}}$ be the asymptotic covariance matrices of the loadings and factors estimated via PC analysis, then, under Assumptions 1, 2, 3, 4, 5, and 6:*

(a) if $\sqrt{T}\log^{1/\delta_v} T/n \to 0$, as $n, T \to \infty$, then $\boldsymbol{\mathcal{V}}_i^{\text{PC}} = \boldsymbol{\mathcal{V}}_i$ for any $i = 1, \dots, n$;

(b) if $\sqrt{n \log n}/T \to 0$ and $n^{-1} \sum_{i,j=1, i\neq j}^n |\mathbb{E}[\xi_{it}\xi_{jt}]| \to 0$, as $n, T \to \infty$, then $(\boldsymbol{\mathcal{W}}_t^{\text{PC}} - \boldsymbol{\mathcal{W}}_t)$ is a positive definite matrix.

Part (a) follows immediately once we impose the identifying Assumption 6 to the results about PC estimation of the loadings (see Barigozzi, 2023, Theorem 1, and Bai, 2003, Theorem 2). Therefore, although we estimate a mis-specified model, the EM estimator of the loadings is as efficient as the PC estimator.

Turning to part (b), by imposing the identifying Assumption 6, the asymptotic covariance of the PC estimator of the factors is $\boldsymbol{\mathcal{W}}_t^{\text{PC}} = (\boldsymbol{\Sigma}_\Lambda)^{-1}\{\lim_{n\to\infty} n^{-1}\boldsymbol{\Lambda}_n'\boldsymbol{\Gamma}_n^\xi\boldsymbol{\Lambda}_n\}(\boldsymbol{\Sigma}_\Lambda)^{-1}$ (see Lemma H.1, and Bai, 2003, Theorem 1), and if the true model were an exact factor model, i.e., $\mathbb{E}[\xi_{it}\xi_{jt}] = 0$ if $i \neq j$ so that $\boldsymbol{\Gamma}_n^\xi = \boldsymbol{\Sigma}_n^\xi$, then we would have $\boldsymbol{\mathcal{W}}_t^{\text{PC}} = (\boldsymbol{\Sigma}_\Lambda)^{-1}\{\lim_{n\to\infty} n^{-1}\boldsymbol{\Lambda}_n'\boldsymbol{\Sigma}_n^\xi\boldsymbol{\Lambda}_n\}(\boldsymbol{\Sigma}_\Lambda)^{-1}$, which is the unfeasible OLS asymptotic covariance matrix in presence of heteroskedastic errors. For the Kalman smoother, from Proposition 3 we have $\boldsymbol{\mathcal{W}}_t = (\boldsymbol{\Sigma}_{\Lambda\Sigma\Lambda})^{-1}\{\lim_{n\to\infty} n^{-1}\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Gamma}_n^\xi(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n\}(\boldsymbol{\Sigma}_{\Lambda\Sigma\Lambda})^{-1}$, and if the true model were an exact factor model, this would reduce to $\boldsymbol{\mathcal{W}}_t = (\boldsymbol{\Sigma}_{\Lambda\Sigma\Lambda})^{-1}$, which is the unfeasible WLS asymptotic covariance matrix. In this case, the Kalman smoother estimator is more efficient because it takes into account heteroskedasticity, whereas the PC estimator ignores the possibility of individual-specific variances.

Here, we go one step further and show that if the idiosyncratic covariance matrix $\boldsymbol{\Gamma}_n^\xi$ is sparse enough, we can still have efficiency gains compared to PC analysis. Indeed, if the total contribution of the off-diagonal elements $\boldsymbol{\Gamma}_n^\xi$ is negligible compared to n, we can expect the asymptotic covariance of the estimated factors to be close to the one we would have for an exact factor model, which we know is smaller than the PC asymptotic covariance. The sparsity condition we assume is the same as the one on Bai and Liao (2016, Assumption 3.1). Although this sparsity condition is hard to verify in practice and is seldom exactly satisfied by the data, in our MonteCarlo results in Section 8, we show that the Kalman smoother estimator tends to perform better than the PC estimator even under more general idiosyncratic covariance structures like banded matrices.

Remark 13. If we assume uncorrelated and homoskedastic idiosyncratic components, i.e., such that $\boldsymbol{\Gamma}_n^\xi = \psi \mathbf{I}_n$ for some $\psi > 0$, then it is easy to see that $\boldsymbol{\mathcal{V}}_i = \boldsymbol{\mathcal{V}}_i^{\text{PC}} = \psi(\boldsymbol{\Gamma}^F)^{-1} = \psi \mathbf{I}_r$, by Assumption 6(b), and $\boldsymbol{\mathcal{W}}_t = \boldsymbol{\mathcal{W}}_t^{\text{PC}} = \psi(\boldsymbol{\Sigma}_\Lambda)^{-1}$. In this case, the EM and Kalman smoother estimators are as efficient as the PC estimators.

Remark 14. There are other estimators that could be more efficient. First, Bai and Liao (2016), Wang et al. (2019), and Poignard and Terada (2020) proposed penalized QML-type estimators of the loadings and of the idiosyncratic covariance matrix. These estimators are used to build a GLS estimator of the factors. This approach addresses cross-sectional idiosyncratic correlations and heteroskedasticity, but not serial idiosyncratic correlations. Second, Breitung and Tenhofen (2011) propose a GLS estimator of the loadings, based on the classical Cochrane and Orcutt (1949) approach, and a WLS estimator for the factors based on that loadings estimator and estimates of the idiosyncratic variances. This approach addresses cross-sectional idiosyncratic heteroskedasticity and serial idiosyncratic correlations, but not cross-sectional idiosyncratic correlations. Finally, Lin and Michailidis (2020) address all idiosyncratic cross-autocorrelations by assuming a sparse VAR model for the idiosyncratic components. To this end, they apply the Cochrane and Orcutt (1949) approach and embed a penalty into an alternating minimization algorithm. They recover the factors via GLS. Their theoretical analysis applies only to the finite-dimensional case. None of these three approaches models the factors' dynamics.

7 Inference

To conduct inference, we need asymptotic covariances of the loadings and factors matrices and their estimators.

Corollary 1. *Under the same assumptions of Propositions 2 and 3, as $n, T \to \infty$:*

(a.1) for any finite $\bar{n}$ and any given sequence of integers $\{s(1)\dots, s(\bar{n})\} \subset \{1,\dots,n\}$, let $\text{vec}(\widehat{\boldsymbol{\Lambda}}_{\bar{n}}) = (\widehat{\boldsymbol{\lambda}}'_{s(1)} \cdots \widehat{\boldsymbol{\lambda}}'_{s(\bar{n})})'$ and $\text{vec}(\boldsymbol{\Lambda}_{\bar{n}}) = (\boldsymbol{\lambda}'_{s(1)} \cdots \boldsymbol{\lambda}'_{s(\bar{n})})'$, as $n, T \to \infty$, if $n^{-1}\sqrt{T}\log^{2/\delta_v} T \to 0$,

$$\sqrt{T}\big\{\text{vec}(\widehat{\boldsymbol{\Lambda}}_{\bar{n}}) - \text{vec}(\boldsymbol{\Lambda}_{\bar{n}})\big\} \xrightarrow{d} \mathcal{N}(\mathbf{0}_{r\bar{n}}, \boldsymbol{\mathcal{V}}_{\bar{n}}),$$

where

$$\boldsymbol{\mathcal{V}}_{\bar{n}} = (\mathbf{I}_{\bar{n}} \otimes \boldsymbol{\Gamma}^F)^{-1} \left(\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^T \sum_{s=1}^T \mathbb{E}[\boldsymbol{\xi}_{\bar{n}t}\boldsymbol{\xi}_{\bar{n}s}] \otimes \mathbb{E}[\mathbf{F}_t \mathbf{F}'_s] \right) (\mathbf{I}_{\bar{n}} \otimes \boldsymbol{\Gamma}^F)^{-1},$$

with $\boldsymbol{\xi}_{\bar{n}t} = (\xi_{s(1)t} \cdots \xi_{s(\bar{n})t})'$, with $\boldsymbol{\Gamma}^F = \mathbf{I}_r$, because of Assumption 6(b);

(a.2) if $\mathbb{E}[(\boldsymbol{\xi}'_{n1} \cdots \boldsymbol{\xi}'_{nT})'(\boldsymbol{\xi}'_{n1} \cdots \boldsymbol{\xi}'_{nT})] = \mathbf{I}_T \otimes \boldsymbol{\Sigma}_n^\xi$ for all $n, T \in \mathbb{N}$, then $\boldsymbol{\mathcal{V}}_{\bar{n}} = \boldsymbol{\Sigma}_{\bar{n}}^\xi \otimes (\boldsymbol{\Gamma}^F)^{-1}$, with $\boldsymbol{\Gamma}^F = \mathbf{I}_r$, because of Assumption 6(b);

(b.1) for any finite $\bar{T}$ and any given sequence of integers $\{s(1)\dots, s(\bar{T})\} \subset \{1,\dots,T\}$, let $\text{vec}(\widehat{\boldsymbol{\mathcal{F}}}_{\bar{T}}) = (\widehat{\mathbf{F}}'_{s(1)} \cdots \widehat{\mathbf{F}}'_{s(\bar{T})})'$ and $\text{vec}(\boldsymbol{\mathcal{F}}_{\bar{T}}) = (\mathbf{F}'_{s(1)} \cdots \mathbf{F}'_{s(\bar{T})})'$, then, as $n, T \to \infty$, if $T^{-1}\sqrt{n \log n} \to 0$,

$$\sqrt{n}\big\{\text{vec}(\widehat{\boldsymbol{\mathcal{F}}}_{\bar{T}}) - \text{vec}(\boldsymbol{\mathcal{F}}_{\bar{T}})\big\} \xrightarrow{d} \mathcal{N}(\mathbf{0}_{r\bar{T}}, \boldsymbol{\mathcal{W}}_{\bar{T}}),$$

where

$$\boldsymbol{\mathcal{W}}_{\bar{T}} = (\mathbf{I}_{\bar{T}} \otimes \boldsymbol{\Sigma}_{\Lambda\Sigma\Lambda})^{-1} \left(\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \frac{\mathbb{E}[\boldsymbol{\zeta}_{i\bar{T}}\boldsymbol{\zeta}_{j\bar{T}}] \otimes \boldsymbol{\lambda}_i \boldsymbol{\lambda}_j}{\sigma_i^2 \sigma_j^2} \right) (\mathbf{I}_{\bar{T}} \otimes \boldsymbol{\Sigma}_{\Lambda\Sigma\Lambda})^{-1},$$

with $\boldsymbol{\zeta}_{i\bar{T}} = (\xi_{is(1)} \cdots \xi_{is(\bar{T})})'$;

(b.2) if $\mathbb{E}[(\boldsymbol{\zeta}'_{1T} \cdots \boldsymbol{\zeta}'_{nT})'(\boldsymbol{\zeta}'_{1T} \cdots \boldsymbol{\zeta}'_{nT})] = \boldsymbol{\Sigma}_n^\xi \otimes \mathbf{I}_T$ for all $n, T \in \mathbb{N}$, then $\boldsymbol{\mathcal{W}}_{\bar{T}} = \mathbf{I}_{\bar{T}} \otimes (\boldsymbol{\Sigma}_{\Lambda\Sigma\Lambda})^{-1}$.

For any given $i, j = 1, \dots, n$, the most general estimator of the asymptotic covariance between $\widehat{\boldsymbol{\lambda}}_i$ and $\widehat{\boldsymbol{\lambda}}_j$ is given by:[7]

$$\widehat{\mathcal{V}}_{i,j}^{(\text{HAC})} = \left(\frac{1}{T}\sum_{t=1}^T \widehat{\mathbf{F}}_t\widehat{\mathbf{F}}_t'\right)^{-1} \left(\frac{1}{T}\sum_{t=1}^T\sum_{s=1}^T \mathrm{K}(t,s)\left\{\widehat{\mathbf{F}}_t\widehat{\mathbf{F}}_s'\widehat{\xi}_{it}\widehat{\xi}_{js}\right\}\right)\left(\frac{1}{T}\sum_{t=1}^T \widehat{\mathbf{F}}_t\widehat{\mathbf{F}}_t'\right)^{-1}, \tag{26}$$

where $\widehat{\xi}_{it} = x_{it} - \widehat{\chi}_{it}$ is the estimated idiosyncratic component of the ith variable at time t. $\mathrm{K}(t,s) = 1 - \frac{|t-s|}{M_T+1}$, if $|t-s| \le M_T$ and zero otherwise, with M_T is such that $M_T \to \infty$ and $M_T/T \to 0$, as $T \to \infty$. Consistency of (26), as $n,T \to \infty$, follows from Bai, 2003, Theorem 6 combined with Propositions 3 and 4.

For any given $t = 1,\dots,T$ and $k = 0,\dots,M_T$, with M_T defined above, the most general estimator of the asymptotic covariance between $\widehat{\mathbf{F}}_t$ and $\widehat{\mathbf{F}}_{t-k}$ is given by:

$$\widehat{\mathcal{W}}_{t,t-k}^{(\text{HAC})} = \left(\frac{1}{n}\sum_{i=1}^n \frac{\widehat{\boldsymbol{\lambda}}_i\widehat{\boldsymbol{\lambda}}_i'}{\widehat{\sigma}_i^2}\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^n\sum_{j=1}^n \mathrm{K}(i,j)\left\{\frac{\widehat{\boldsymbol{\lambda}}_i\widehat{\boldsymbol{\lambda}}_j'}{\widehat{\sigma}_i^2\widehat{\sigma}_j^2}\widehat{\gamma}_{ij,k}^{\xi}\right\}\right)\left(\frac{1}{n}\sum_{i=1}^n \frac{\widehat{\boldsymbol{\lambda}}_i\widehat{\boldsymbol{\lambda}}_i'}{\widehat{\sigma}_i^2}\right)^{-1}, \tag{27}$$

where $\widehat{\gamma}_{ij,k}^{\xi} = T^{-1}\sum_{t=k+1}^T \widehat{\xi}_{it}\widehat{\xi}_{j,t-k}$ and $\mathrm{K}(i,j) = 1$ if $1 \le i,j \le m_{n,T}$ and zero otherwise, with $m_{n,T} \to \infty$ and $m_{n,T}/\min(n,T) \to 0$, as $n,T \to \infty$. Consistency of (27), as $n,T \to \infty$, follows from Bai and Ng (2006, Theorem 4) combined with Propositions 2, 4, and 5. For larger values of k, $\mathbb{E}[\xi_{it}\xi_{j,t-k}]$ is likely to be small due to our Assumption 2(b), and, thus, we can consider $\widehat{\mathbf{F}}_t$ and $\widehat{\mathbf{F}}_{t-k}$ as asymptotically uncorrelated. An alternative kernel function of the correlation based distance between i and j is considered by Kim (2022) who considers also averages of (27) when computed choosing different random permutations of the selected $m_{n,T}$ cross-sectional units. Alternatively, rather that smoothing via the use of a kernel, Fresoli et al. (2024) consider an estimator based on thresholding of the idiosyncratic sample covariance matrix (see also Fan et al., 2013).

Finally, we can obtain an estimator of $\mathcal{W}_t$ that accounts also for the autocorrelation of the factors from the true MSE of the Kalman filter given in (22) when using the estimated parameters. Once again this requires an estimator of the idiosyncratic covariance matrix like the estimators discussed above. The asymptotic covariance between $\widehat{\mathbf{F}}_t$ and $\widehat{\mathbf{F}}_{t-k}$ can be obtained by considering the Kalman filter with the augmented state vector $(\mathbf{F}_t' \cdots \mathbf{F}_{t-k}')'$ and by then considering the corresponding $r \times r$ off-diagonal block of the resulting $r(k+1)\times r(k+1)$ MSE.

Having the estimated loadings $\widehat{\boldsymbol{\lambda}}_i$, the estimated factors $\widehat{\mathbf{F}}_t$, and any of the above estimators of $\mathcal{V}_{i,j}$ and $\mathcal{W}_{t,t-k}$, we can estimate the variance of the estimated common component by plugging these quantities into the expression in Proposition 4(b).

To conclude, the covariance estimators defined in this section can be used, together with the asymptotic distributions derived in Propositions 2 and 3, for inferential purposes. Examples are in Section 9.

Remark 15. From parts (a.2) and (b.2) we see that, under a correctly specified model, each estimated row i of the loadings matrix is asymptotically uncorrelated with the other rows, and, likewise, each estimated realization of the factors at a given point in time t is asymptotically uncorrelated with the other time periods. Estimators of the asymptotic covariances are easily built in this case as: $\widehat{\mathcal{V}}_i^{(0)} = \widehat{\sigma}_i^2(T^{-1}\sum_{t=1}^T \widehat{\mathbf{F}}_t\widehat{\mathbf{F}}_t')^{-1}$, and $\widehat{\mathcal{W}}_t^{(0)} = (n^{-1}\sum_{i=1}^n(\widehat{\sigma}_i^2)^{-1}\widehat{\boldsymbol{\lambda}}_i\widehat{\boldsymbol{\lambda}}_i')^{-1}$, while $\widehat{\mathcal{V}}_{i,j}^{(0)} = \mathbf{0}_{r\times r}$ if $i \ne j$, and $\widehat{\mathcal{W}}_{t,t-k}^{(0)} = \mathbf{0}_{r\times r}$ if $k \ne 0$. Consistency of these two estimators follows directly from Propositions 2, 3, 4, and 5.

[7] As discussed in Remark 8, the sample covariance of $\widehat{\mathbf{F}}_t$ is equal to the r-dimensional identity matrix only asymptotically; hence, it must be included in the estimator of the asymptotic covariance.

8 Monte Carlo study

Throughout, we consider a model with $r = 4$ factors, and we simulate the data according to

$$x_{it} = \boldsymbol{\ell}_i'\boldsymbol{f}_t + \phi_i\xi_{it}, \quad \boldsymbol{f}_t = \boldsymbol{A}\boldsymbol{f}_{t-1} + \mathbf{u}_t, \quad \xi_{it} = \alpha_i\xi_{it-1} + e_{it}, \tag{28}$$

where $\boldsymbol{\ell}_i$ has entries $\ell_{ij} \overset{iid}{\sim} \mathcal{N}(1,1)$, $i = 1,\dots,n$, $j = 1,\dots,r$; $\boldsymbol{A} = \mu\check{\boldsymbol{A}}\{\nu^{(1)}(\check{\boldsymbol{A}})\}^{-1}$, where $[\check{\boldsymbol{A}}]_{jj} \sim U[0.5, 0.8]$, $[\check{\boldsymbol{A}}]_{jk} \sim U[0, 0.3]$, $j,k = 1,\dots,r$, and $\mu = 0.7$; $u_{jt} \overset{iid}{\sim} (0,1)$, $j = 1,\dots,r$ following either a Gaussian, an Asymmetric Laplace, or a Skew-t distribution, and with $\mathbb{C}\text{ov}(u_{it}, u_{jt}) = 0$, for $i \neq j$; $\alpha_i = \{0, \delta_i\}$, $i = 1,\dots,n$, where $\delta_i \overset{iid}{\sim} \mathcal{U}(0,\delta)$, and $\delta \in \{0, 0.5\}$; $e_{it} \overset{iid}{\sim} (0, \sigma^2_{ei})$, $i = 1,\dots,n$, following either a Gaussian distribution with $\sigma^2_{ei} \sim U[0.5, 1.5]$, or an Asymmetric Laplace distribution with $\sigma^2_{ei} = 1$, or a Skew-t distribution with $\sigma^2_{ei} = 1$; $\mathbb{C}\text{ov}(e_{it}, e_{jt}) = \tau^{|i-j|}$, $i,j = 1,\dots,n$, with $\tau \in \{0, 0.5\}$ if $|i-j| \leq 10$, and $\mathbb{C}\text{ov}(e_{it}, e_{jt}) = 0$ otherwise; and, last, $\phi_i = \sqrt{\theta_i \widehat{\mathbb{V}\text{ar}}(\chi_{it})(\widehat{\mathbb{V}\text{ar}}(\xi_{it}))^{-1}}$, $i = 1,\dots,n$, where $\widehat{\mathbb{V}\text{ar}}(\cdot)$ denotes the sample variance, and $\theta_i \overset{iid}{\sim} \mathcal{U}(\bar{\theta} - 0.25, \bar{\theta})$, and $\bar{\theta} = 0.5$. The parameters μ, τ, δ, and $\bar{\theta}$ are crucial and control: the persistence of the factors, the degrees of cross-sectional and serial idiosyncratic correlation in the idiosyncratic components, and the noise-to-signal ratio, respectively.[8]

Finally, in order to satisfy Assumptions 6(b) and 6(c), after we have generated the common component $\boldsymbol{\chi}_{nt}$ as in (28), we construct the factors as $\mathbf{F}_t = (\mathbf{M}_n^{\chi})^{-1/2}\mathbf{V}_n^{\chi\prime}\boldsymbol{\chi}_{nt}$ and the loadings as $\boldsymbol{\Lambda}_n = \mathbf{V}_n^{\chi}(\mathbf{M}_n^{\chi})^{1/2}$, where $\mathbf{M}_n^{\chi}$ is the $r \times r$ diagonal matrix containing the eigenvalues of $T^{-1}\sum_{t=1}^T \boldsymbol{\chi}_{nt}\boldsymbol{\chi}_{nt}'$, and $\mathbf{V}_n^{\chi}$ is the $n \times r$ matrix having as columns the corresponding normalized eigenvectors and with sign fixed such that it has non-negative entries in the first row.

We consider $n \in \{100, 200, 300, 500\}$, $T \in \{100, 200, 300, 500\}$ and $B = 1000$ replications. At each replication b, we run the EM algorithm as described in Algorithm 1, thus obtaining an estimate of the loadings $\widehat{\boldsymbol{\lambda}}_i^{(b)}$, the factors $\widehat{\mathbf{F}}_t^{(b)}$, and the common component $\widehat{\chi}_{it}^{(b)} = \widehat{\boldsymbol{\lambda}}_i^{(b)\prime}\widehat{\mathbf{F}}_t^{(b)}$. We initialize the EM algorithm either through the PC estimators as explained in Appendix A.1, or through a contaminated version of the PC estimator obtained using contaminated eigenvectors of the data. In particular, let $\widehat{\mathbf{V}}_n^x$ be the $n \times r$ matrix of the r leading eigenvectors of $T^{-1}\sum_{t=1}^T \mathbf{x}_{nt}\mathbf{x}_{nt}'$, then the contaminated eigenvector is $\widehat{\mathbf{V}}_n^{x,c} = \widehat{\mathbf{V}}_n^x + \boldsymbol{\mathcal{Z}}\boldsymbol{\Upsilon}^{1/2}$, where $[\boldsymbol{\mathcal{Z}}]_{ij} \overset{iid}{\sim} \mathcal{N}(0,1)$, $i = 1,\dots,n$, $j = 1,\dots,r$, and $[\boldsymbol{\Upsilon}]_{ij} = \iota^{|i-j|}$, $i,j = 1,\dots,r$, with $\iota \in (0,1)$—the bigger ι is, the stronger is the contamination.

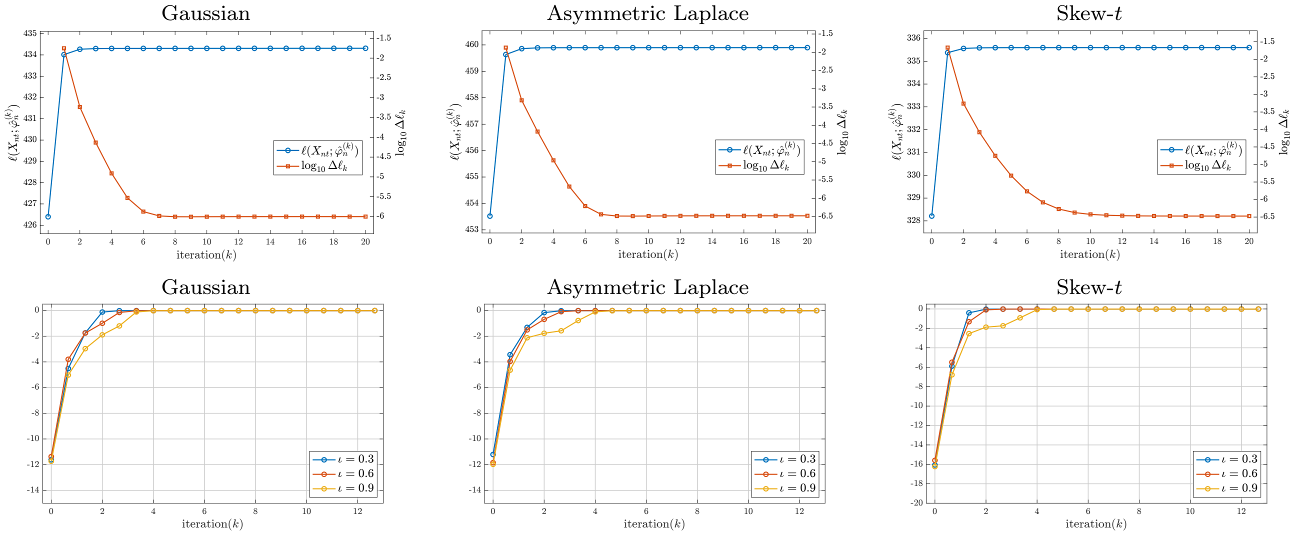
The upper plots of Figure 1 show the log-likelihood $\ell(\boldsymbol{X}_{nT}; \widehat{\boldsymbol{\varphi}}_n^{(k)})$ (blue line, left scale) as a function of the iteration k of the EM algorithm, and the convergence criterion $\Delta\ell_k$ defined in (A.14) (red line, right scale) when we initialize the EM algorithm with the PC estimator. The results in Figure 1 show that the log-likelihood is an increasing function in the number of iterations k (as it should be). Moreover, the algorithm converges very fast: within two iterations $\Delta\ell_k \leq 10^{-3}$. This is to be expected since we initialize the algorithm with the consistent PC estimator.

The lower plots of Figure 1 show the percentage deviation of the log-likelihood of the algorithm initialized with the contaminated estimator (ℓ_k^c) from the one initialized with the PC estimator (ℓ_k), that is $\ell_k^c/\ell_k - 1$. Contaminating the initialization implies starting from a much lower likelihood. However, with just a few iterations, the log-likelihood initialized with the contaminated estimator is the same as the one properly initialized. This result shows that initializing the EM with a non-consistent estimator is fine, as it just requires running the algorithm for a few more iterations. Therefore, hereafter, we focus on the case in which we inizialize the EM algorithm with the PC estimator.

Moving to the performance of our estimators, we begin with Proposition 4 part (a), which gives consistency

[8] In the case of the Asymmetric Laplace distribution, all the innovations have location 0, scale index λ, and asymmetry index κ, with $\kappa \sim \mathcal{U}(.9, 1.1)$ and $\lambda = \sqrt{(1+\kappa^4)\kappa^{-2}}$, so that all the shocks have variance 1. In the case of the Skew-t distribution, all the shocks have location 0, dispersion 1, skewness index γ, and tail index ν, with $\nu_u \sim U(4, 12)$, $\gamma_u \sim U(-.1, .1)$, $\nu_e \sim U(3, 13)$, and $\gamma_e \sim U(-.15, .15)$.

Figure 1: Simulation results - Convergence of the EM algorithm



The upper plots show the log-likelihood $\ell(\boldsymbol{X}_{nT};\widehat{\boldsymbol{\varphi}}_n^{(k)})$ (blue line, left scale), and the convergence criterion $\Delta\ell_k$ defined in (A.14) (red line, right scale) when we initialized the EM algorithm with the PC estimator. The lower plots show the percentage deviation of the log-likelihood of the algorithm initialized with the contaminated estimator (ℓ_k^c) from the one initialized with the PC estimator (ℓ_k), that is $\ell_k^c/\ell_k - 1$. The bigger ι is, the stronger is the contamination. The log-likelihoods in this figure were obtained from a single simulation where $T = 100$, $n = 100$, $\mu = 0.7$, $\delta = 0.5$, $\tau = 0.5$, and $\theta = 0.5$.

and rates for the common component's estimator. The left plot in Figure 2 shows the root mean squared error:

$$\text{RMSE} = \sqrt{\frac{1}{nTB}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{b=1}^{B}\left(\widehat{\chi}_{it}^{(b)} - \chi_{it}^{(b)}\right)^2}.$$

Instead, the right plot shows the relative RMSE of our estimator over the RMSE of the PC estimator—values smaller than one indicate a better performance of our estimators. We show results for several DGPs, which allow us to disentangle the effects of each single mis-specification on the performance of our estimator.

Two main results emerge from Figure 2. First, as n and T grow, the RMSE of all DGPs converge toward zero, thus indicating that the mis-specification introduced by estimating a model with uncorrelated and possibly non-Gaussian idiosyncratic components, even when that is not the case, does not affect our estimator. In particular, between serial correlation (the orange line) and cross-sectional correlation (the yellow line), the mis-specification that hurts the most is cross-sectional correlation. This is good news for practitioners because the cross-sectional correlation between the idiosyncratic components can somehow be limited by avoiding to include in the dataset variables that are too similar with one another (see, e.g., the discussions in Boivin and Ng, 2006 and Luciani, 2014). Moreover, the model is consistently estimated when the shocks are asymmetric and have heavy tails even when they come from a distribution that does not meet Assumption 5 of exponentially decaying tails of the distribution.[9] This is also good news for practitioners because this result tells us that we can use this model in settings likely to be non-Gaussian, as is the case in datasets of disaggregated inflation rates (see, e.g., Reis and Watson, 2010, and Ahn and Luciani, 2024), which are notoriously skewed and fat-tailed.

Second, overall, our estimator behaves very similarly to but slightly better than the PC estimator, despite the latter being non-parametric and, thus, not affected by mis-specifications. This result confirms the conjectures based on extensive numerical studies made by Doz et al. (2012) and Bai and Li (2016), showing that, in a high-dimensional setting, the EM estimator is "as good as" the PC estimator.

To evaluate the estimates of the factors and the loadings, at each replication b, we consider a multivariate

[9] The dark gray line for the Skew-t distribution cannot be seen in the left plot because it is underneath, thus it coincides with, the red line.

Figure 2: Simulation results - Common components
Root Mean Squared Errors

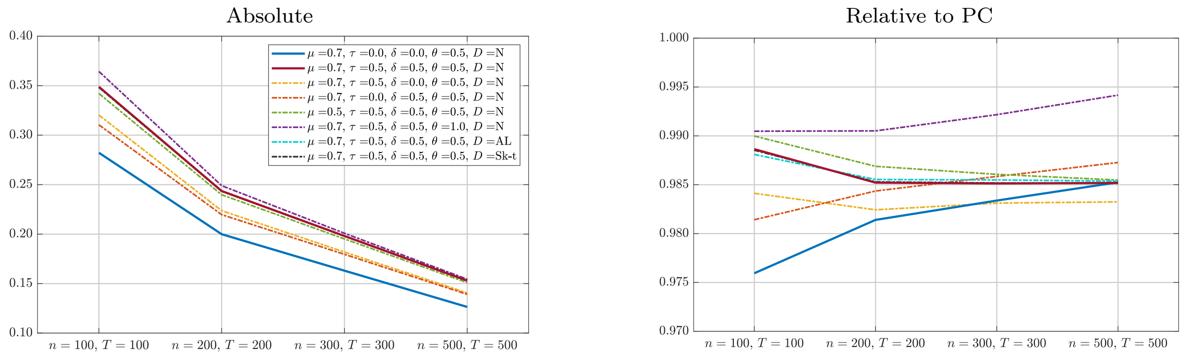
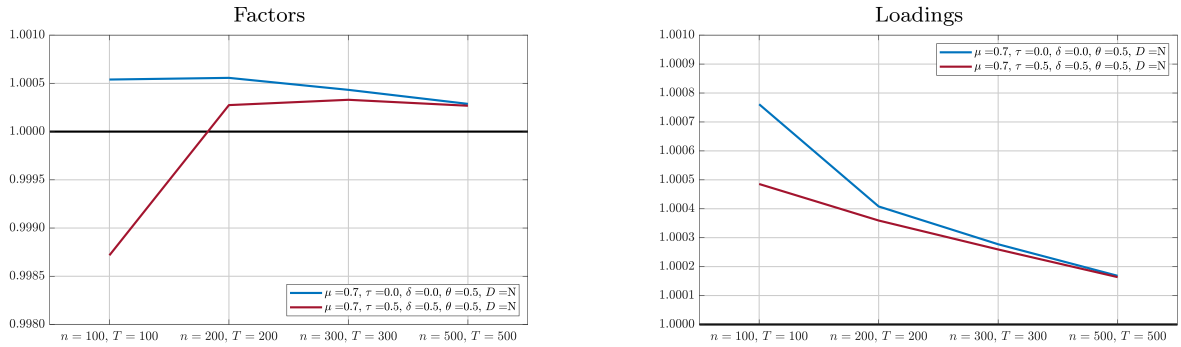


Figure 3: Simulation results - Factors and loadings
Trace statistics relative to PC



version of the R^2 (see also Doz et al., 2012):

$$\mathrm{TR}_F^{(b)} = \frac{\mathrm{tr}\left((\boldsymbol{\mathcal{F}}_T^{(b)\prime}\widehat{\boldsymbol{\mathcal{F}}}_T^{(b)})(\widehat{\boldsymbol{\mathcal{F}}}_T^{(b)\prime}\widehat{\boldsymbol{\mathcal{F}}}_T^{(b)})^{-1}(\widehat{\boldsymbol{\mathcal{F}}}_T^{(b)\prime}\boldsymbol{\mathcal{F}}_T^{(b)})\right)}{\mathrm{tr}\left(\boldsymbol{\mathcal{F}}_T^{(b)\prime}\boldsymbol{\mathcal{F}}_T^{(b)}\right)}, \quad \mathrm{TR}_\Lambda^{(b)} = \frac{\mathrm{tr}\left((\boldsymbol{\Lambda}_n^{(b)\prime}\widehat{\boldsymbol{\Lambda}}_n^{(b)})(\widehat{\boldsymbol{\Lambda}}_n^{(b)\prime}\widehat{\boldsymbol{\Lambda}}_n^{(b)})^{-1}(\widehat{\boldsymbol{\Lambda}}_n^{(b)\prime}\boldsymbol{\Lambda}_n^{(b)})\right)}{\mathrm{tr}\left(\boldsymbol{\Lambda}_n^{(b)\prime}\boldsymbol{\Lambda}_n^{(b)}\right)},$$

where $\widehat{\boldsymbol{\mathcal{F}}}_T^{(b)}$ and $\widehat{\boldsymbol{\Lambda}}_n^{(b)}$ are the $T\times r$ and $n\times r$ matrices of the estimated factors and loadings, respectively. These trace statistics are smaller than one, and they tend to one when the empirical canonical correlations between the true quantities and their estimates tend to one.

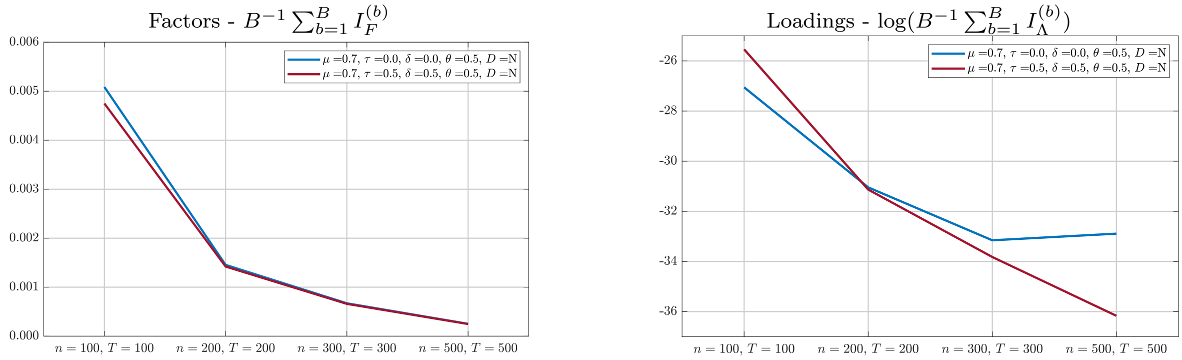
Figure 3 reports the values of $\mathrm{TR}_F^{(b)}$ and $\mathrm{TR}_\Lambda^{(b)}$, relative to the same measures computed for the PC estimator, averaged over all B replications (values larger than one indicate a better performance of our estimators). The results in Figure 3 mirror those of Figure 2: our estimator behaves very similarly to but slightly better than the PC estimator.

To derive our asymptotic results, we assumed that the true factors and the loadings are asymptotically identified as in Assumption 6(b). As explained in Remark 8 the estimated factors and loading satisfy this assumption asymptotically. To verify that this is the case, in Figure 4, we show the two identification errors:

$$I_F^{(b)} = \frac{1}{r}\sum_{j=1}^r \left(\nu^{(j)}(\widehat{\boldsymbol{\Gamma}}^{F(b)}) - 1\right)^2, \qquad I_\Lambda^{(b)} = \frac{1}{r}\sum_{j=1}^r \left(\nu^{(j)}(\widehat{\boldsymbol{\Sigma}}_\Lambda^{(b)}) - [\widehat{\boldsymbol{\Sigma}}_\Lambda^{(b)}]_{jj}\right)^2,$$

where $\widehat{\boldsymbol{\Gamma}}^{F(b)} = T^{-1}\widehat{\boldsymbol{\mathcal{F}}}_T^{(b)\prime}\widehat{\boldsymbol{\mathcal{F}}}_T^{(b)}$, and $\widehat{\boldsymbol{\Sigma}}_\Lambda^{(b)} = n^{-1}\widehat{\boldsymbol{\Lambda}}_n^{(b)\prime}\widehat{\boldsymbol{\Lambda}}_n^{(b)}$. According to our asymptotic results, both these quantities should tend to zero as n and T increase because, given our simulation design, the true factors are orthonormal and the true loadings are orthogonal in agreement with Assumption 6(b). Figure 4 shows $I_F^{(b)}$ and $I_\Lambda^{(b)}$ averaged over all B replications. The identifying constraints are more and more precisely satisfied as n and T grow.

Figure 4: Simulation results - Factors and loadings
Identification criteria



Next, we move to the asymptotic distribution of the common component. To this end, for each replication b and any i,t, we compute

$$Z_{it}^{(b)} = \left(\frac{1}{T}\widehat{\mathbf{F}}_t^{(b)\prime}\widehat{\boldsymbol{\mathcal{V}}}_i^{((\text{HAC},b)}\widehat{\mathbf{F}}_t^{(b)} + \frac{1}{n}\widehat{\boldsymbol{\lambda}}_i^{(b)\prime}\widehat{\boldsymbol{\mathcal{W}}}_t^{((\text{HAC},b)}\widehat{\boldsymbol{\lambda}}_i^{(b)}\right)^{-1/2}(\widehat{\chi}_{it}^{(b)} - \chi_{it}), \qquad (29)$$

where we use the robust estimators of the asymptotic covariance matrices defined in (26) and (27), respectively. For comparison we also consider $Z_{it}^{(0,b)}$ defined as in (29), but when using estimators of the asymptotic covariance matrices in the case of non-correlated idiosyncratic components (henceforth, non-robust covariance matrices).

According to Proposition 4, $Z_{it}^{(b)} \xrightarrow{d} \mathcal{N}(0,1)$ as $n,T\to\infty$. To evaluate the goodness of our theoretical results, we compute the average coverage

$$C(1-\alpha) = \frac{1}{nTB}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{b=1}^{B}\mathbb{I}\left(\mathcal{Z}_{\alpha/2}\le Z_{it}^{(b)}\le \mathcal{Z}_{1-\alpha/2}\right),$$

where $\mathcal{Z}_\alpha$ is the α-quantile of the standard normal distribution. In Table 2, we report the coverage $C(1-\alpha)$, for selected values of $\alpha\in(0,1)$, while for illustration purposes, in Figure 5, we show histograms of $\{Z_{it}^{(b)} : i = 1,\dots,n,\ t=1,\dots,T,\ b=1,\dots,B\}$, for some of the cases considered in Table 2. We stress that throughout this exercise the chosen bandwidths for all robust covariance estimators are not data driven, but rather fixed a priori.[10]

Results confirm the derived asymptotic distribution. When the idiosyncratic components are uncorrelated, the non-robust estimators of the covariance matrices offers almost perfect coverage, while the robust estimators give a slight over-coverage. In the relevant cases of serially and cross-correlated idiosyncratic components, the considered robust estimators work very well. For comparison we show also results for $Z_{it}^{(b)}$ when χ_{it} is estimated with the PC estimator, and the asymptotic covariances are estimated using the estimators in Bai and Ng (2006). In this case, deviations from Gaussianity seem to lead to serious under-coverage.

Last, we move to Proposition 6, which states that under a sparsity condition on the covariance matrix of the idiosyncratic components, the Kalman smoother estimator of the factors is more efficient than the PC estimator. To verify this result, we compare the theoretical asymptotic covariance of the factors, $\boldsymbol{\mathcal{W}}_t$ and $\boldsymbol{\mathcal{W}}_t^{\text{PC}}$. Specifically, for each simulation, we look at the sign of the smallest eigenvalue of the matrix $(\boldsymbol{\mathcal{W}}_t^{\text{PC}} - \boldsymbol{\mathcal{W}}_t)$, computed using the true simulated parameters, which should always be positive.

Table 3 reports the percentage of times out of 5000 simulations in which $(\boldsymbol{\mathcal{W}}_t^{\text{PC}} - \boldsymbol{\mathcal{W}}_t)$ is positive definite. The DGP used for this exercise is the one described at the beginning of this section, but for the case indicated as $\tau = 0.5^*$ in which we set $\mathbb{C}\text{ov}(e_{it},e_{jt}) = 0.5^{|i-j|}$ if $i,j\le\lfloor\sqrt{n}\rfloor$, and $\mathbb{C}\text{ov}(e_{it},e_{jt})=0$, otherwise, in order to better proxy the assumed sparsity condition. The results in Table 3 confirm those in Proposition 6. When the sparsity

[10]To estimate $\widehat{\boldsymbol{\mathcal{V}}}_i^{(\text{HAC})}$ we set $M_T = \lfloor T^{1/4}\rfloor$ and to estimate $\widehat{\boldsymbol{\mathcal{W}}}_t^{(\text{HAC})}$ we set $m = \lfloor n^{4/5}\rfloor$.

Table 2: SIMULATION RESULTS - COMMON COMPONENTS - COVERAGE, $C(1-\alpha)$

			$(1-\alpha)=0.90$		$(1-\alpha)=0.95$	
	n	T	EM	PC	EM	PC
Gaussian, $\tau=0$, $\delta=0$ Non-robust covariances	100	100	0.89	0.88	0.94	0.93
	200	200	0.89	0.89	0.95	0.94
	300	300	0.90	0.89	0.95	0.95
	500	500	0.90	0.90	0.95	0.95
Gaussian, $\tau=0$, $\delta=0$ Robust covariances	100	100	0.91	0.89	0.95	0.94
	200	200	0.92	0.90	0.96	0.95
	300	300	0.92	0.91	0.96	0.95
	500	500	0.93	0.91	0.96	0.95
Gaussian, $\tau=0.5$, $\delta=0.5$ Robust covariances	100	100	0.86	0.84	0.92	0.90
	200	200	0.88	0.86	0.93	0.92
	300	300	0.89	0.87	0.94	0.93
	500	500	0.89	0.88	0.94	0.93
Asymmetric Laplace, $\tau=0.5$, $\delta=0.5$ Robust covariances	100	100	0.86	0.81	0.92	0.88
	200	200	0.88	0.83	0.93	0.89
	300	300	0.89	0.83	0.94	0.90
	500	500	0.89	0.84	0.94	0.90
Skew-t, $\tau=0.5$, $\delta=0.5$ Robust covariances	100	100	0.86	0.81	0.92	0.88
	200	200	0.88	0.83	0.93	0.89
	300	300	0.89	0.83	0.94	0.90
	500	500	0.89	0.84	0.94	0.90

Figure 5: SIMULATION RESULTS - HISTOGRAMS OF $Z_{it}^{(b)}$
Serially and cross-correlated idiosyncratic components ($\tau=0.5$, $\delta=0.5$) - Robust covariances

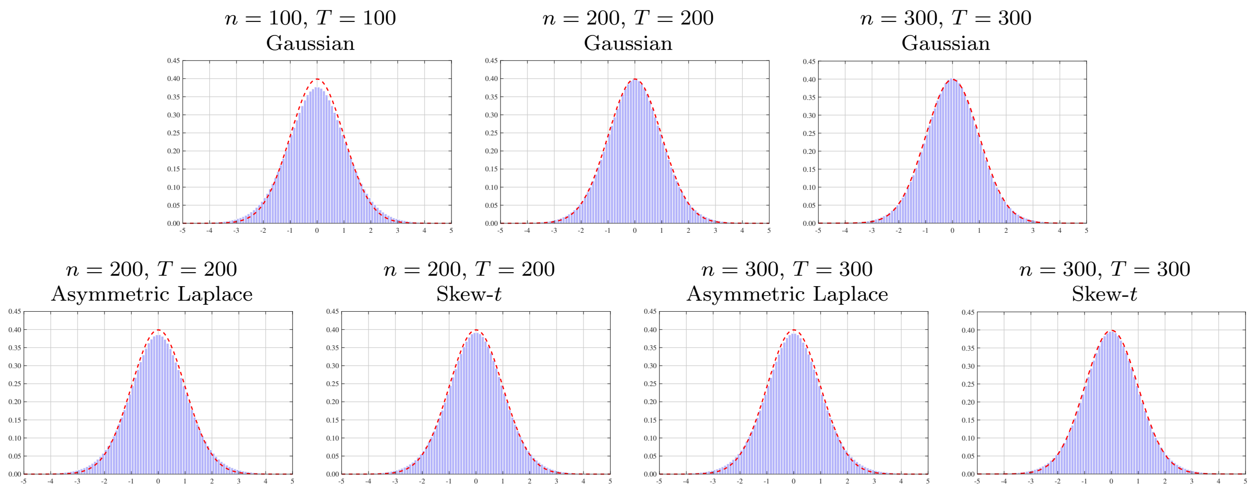


Table 3: PERCENTAGE OF SIMULATIONS IN WHICH $(\mathcal{W}_t^{\text{PC}} - \mathcal{W}_t)$ IS POSITIVE SEMIDEFINITE

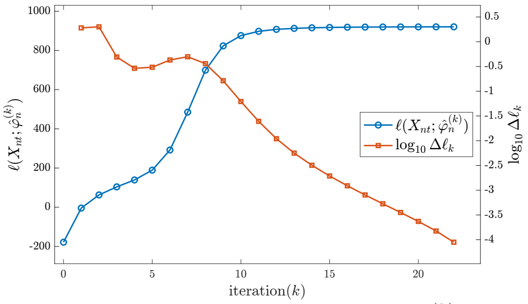
n	T	$\tau=0$	$\tau=0.5^*$	$\tau=0.5$
100	100	100.00	97.74	50.68
200	200	100.00	99.96	85.84
300	300	100.00	100.00	96.82
400	400	100.00	100.00	99.10
500	500	100.00	100.00	99.60
1000	1000	100.00	100.00	100.00

condition is verified (this is the case of columns $\tau=0$ and $\tau=0.5^*$), the Kalman smoother is more efficient than the PC estimator. That said, even when the sparsity condition is not verified (column $\tau=0.5$), the Kalman smoother tends to be more efficient than the PC estimator. We also note that the two largest eigenvalues (not shown) are always positive, meaning that, in our simulations, the PC estimator is never more efficient than the EM estimator.

9 Empirical application

In this section, we consider the dataset used by Barigozzi and Luciani (2023), a typical panel of $n=103$ US macroeconomic quarterly indicators observed from 1960:Q1 to 2018:Q4, thus $T=236$. All variables are transformed to stationarity. In particular, we follow the common approach of taking first differences of price

Figure 6: US DATA - CONVERGENCE OF THE EM ALGORITHM



The plot shows the log-likelihood $\ell(\boldsymbol{X}_{nT};\widehat{\boldsymbol{\varphi}}_n^{(k)})$ (blue line, left scale), and the convergence criterion $\Delta\ell_k$ defined in (A.14) (red line, right scale) when we run the EM algorithm on US data and we initialize it with the PC estimator.

indexes and keeping interest rates in levels (see, e.g., Bernanke et al., 2005). The information criterion by Bai and Ng (2002) indicates $\widehat{r} = 6$ common factors. We estimate that the order of the VAR for the factors is $\widehat{p}_F = 2$. As shown in Figure 6, the EM algorithm converges in 22 iterations when setting the threshold in (A.14) to $\varepsilon = 10^{-4}$ and the log-likelihood increases monotonically.

As we said in the Introduction, there is extensive literature showing the effectiveness of the EM algorithm in estimating large DFMs. Therefore, the purpose of this section is not to show that this method works or that it is superior to the PC estimator. Rather, we concentrate on the innovations brought about in this paper that have an impact on empirical applications, namely: the confidence bands for the common components and the factors, and a test of hypothesis on the factor loadings.

Throughout, we compute the asymptotic covariances by using $\widehat{\boldsymbol{\mathcal{V}}}_i^{\text{HAC}}$ for the loadings, as given in (26), with bandwidth $M_T = \lceil T^{1/4} \rceil$, and $\widehat{\boldsymbol{\mathcal{W}}}_t^{\text{KF}} = n\widehat{\boldsymbol{\Pi}}_{t|t}$ for the factors, computed as in (22) by using the estimated parameters and applying local thresholding to the sample idiosyncratic covariance matrix (Fan et al., 2013; Fresoli et al., 2024).

The first row of Figure 7 shows the common components of a few variables of interest estimated with the EM algorithm (the red line) with their 95% confidence bands, together with the observed data (the black line). Specifically, for any given $i = 1, \dots, n$ and all $t = 1, \dots, T$, a $(1-\alpha)\%$ confidence interval for the estimated common component is given by

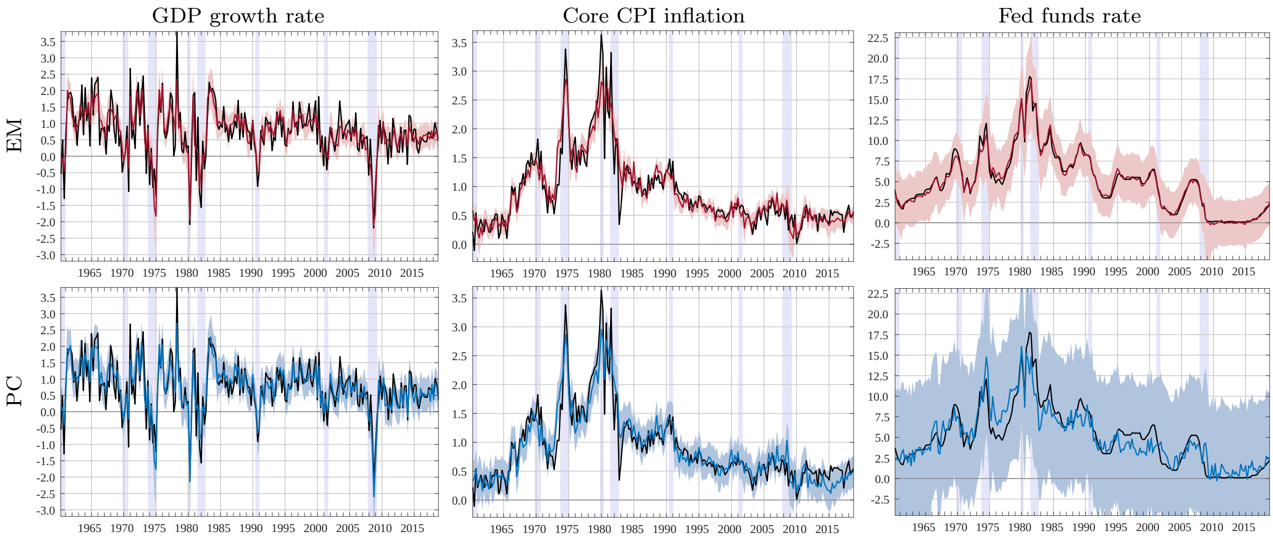
$$\mathcal{I}_{\widehat{\chi}_{it}}(\alpha) = \left[\widehat{\chi}_{it} - z_{(1-\alpha/(2T))} \sqrt{\frac{\widehat{\mathbf{F}}_t' \widehat{\boldsymbol{\mathcal{V}}}_i^{\text{HAC}} \widehat{\mathbf{F}}_t}{T} + \frac{\widehat{\boldsymbol{\lambda}}_i' \widehat{\boldsymbol{\mathcal{W}}}_t^{\text{KF}} \widehat{\boldsymbol{\lambda}}_i}{n}} \, , \, \widehat{\chi}_{it} + z_{(1-\alpha/(2T))} \sqrt{\frac{\widehat{\mathbf{F}}_t' \widehat{\boldsymbol{\mathcal{V}}}_i^{\text{HAC}} \widehat{\mathbf{F}}_t}{T} + \frac{\widehat{\boldsymbol{\lambda}}_i' \widehat{\boldsymbol{\mathcal{W}}}_t^{\text{KF}} \widehat{\boldsymbol{\lambda}}_i}{n}} \right].$$

Moreover, in order to have confidence bands valid for all T observations we apply a Bonferroni correction to the critical values. For comparison, in the second row of Figure 7, we report the estimated common components obtained with the PC estimator (the blue line) with their 95% confidence bands computed using the HAC estimators in Bai and Ng (2006).

The common components of core CPI inflation and the Fed funds rate estimated with the EM track the observed series better than those estimated with the PC estimator. This result is possibly due to local departures from stationarity in those series, which create problems for PC analysis but not for the EM, as the Kalman smoother is able to track changes in the dynamics due to its recursive character. In particular, the confidence band of the common component of the Fed funds rate for the PC estimator is much wider than that of the EM estimator because the variance of the idiosyncratic component obtained with the PC estimator is nearly eight times larger than that of the EM estimator and it is also far more persistent.

In addition to computing confidence bands for the in-sample estimate of the common component, we can also do so for the unconditional and conditional forecasts obtained from the model. In other words, our results

Figure 7: US data - Estimated common components



The shaded area is the 95% confidence band.

open the possibility of computing the uncertainty around GDP nowcast, which is nothing else than a short-term conditional forecast, and for scenario analysis performed with DFMs. Here we give a couple of simplified examples.

The left chart of Figure 8 shows the 1-step ahead forecast of 2018:Q1. That is, we estimate the model up to 2017:Q4, and then we produce four different forecasts of GDP growth conditioning on the observations of an increasing number of variables for 2018:Q1. The conditional forecasts are obtained using the Kalman filter estimator of the factors, $\mathbf{F}_{t|t}^{(k^*+1)}$, while the unconditional forecast is obtained using their one-step-ahead prediction, $\mathbf{F}_{t|t-1}^{(k^*+1)} = \widehat{\mathbf{A}}\mathbf{F}_{t-1|t-1}^{(k^*+1)}$. This exercise mimics a simplified nowcasting setting—we are omitting the aspect of mixed frequency—as the variables we are conditioning on are published earlier than GDP, and the sequence of conditioning mimics the calendar of data releases. As shown in the left chart of Figure 8, the model adjusts the forecast in the right direction as more hard data becomes available.

In the second exercise, we produce forecasts for GDP growth for each quarter of 2018 (blue line). Then, we adjust them based on different scenarios for payroll employment. We have two scenarios: one where employment grows at the same pace as the previous year (red line)—this is a slightly lower average pace than the model expected; another where employment grows at half the previous year's pace, resulting in a more pessimistic forecast (green line). As shown in the right chart of Figure 8, one and two quarters ahead, the forecasts from these scenarios differ significantly.

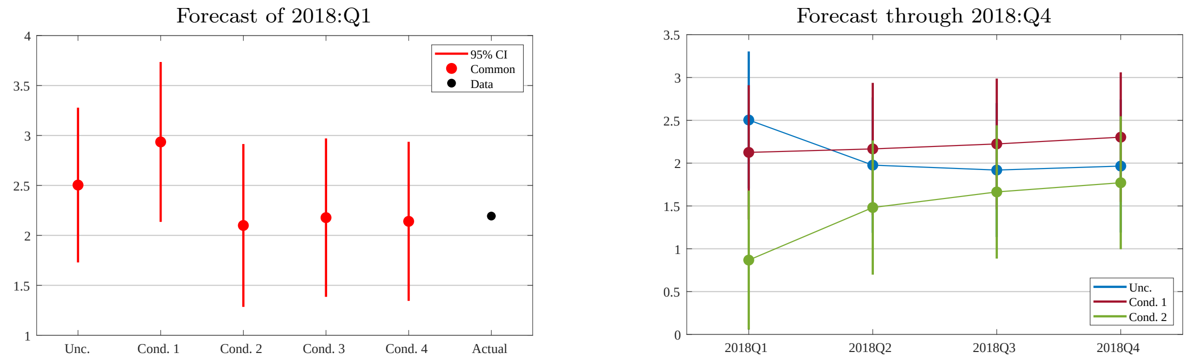
Finally, we can test for linear restrictions on the loadings. Consider testing for s linear restrictions H_0 : $\boldsymbol{R}' \operatorname{vec}(\boldsymbol{\Lambda}_n) = \boldsymbol{q}$, against the alternative H_1 : $\boldsymbol{R}' \operatorname{vec}(\boldsymbol{\Lambda}_n) \neq \boldsymbol{q}$, where $\boldsymbol{R}$ is $nr \times s$ and $\boldsymbol{q}$ is $s \times 1$. Then, the usual Wald-type test statistic is computed as

$$\mathsf{W}_{\widehat{\boldsymbol{\Lambda}}_n} = T \left(\boldsymbol{R}' \operatorname{vec}(\widehat{\boldsymbol{\Lambda}}_n) - \boldsymbol{q}\right)' \left(\boldsymbol{R}' \widehat{\boldsymbol{\mathcal{V}}}_n^{\text{HAC}} \boldsymbol{R}\right)^{-1} \left(\boldsymbol{R}' \operatorname{vec}(\widehat{\boldsymbol{\Lambda}}_n) - \boldsymbol{q}\right), \tag{30}$$

where $\boldsymbol{R}$ selects only sr rows of $\widehat{\boldsymbol{\mathcal{V}}}_n^{\text{HAC}}$. Under H_0, from Proposition 2 it follows that $\mathsf{W}_{\widehat{\boldsymbol{\Lambda}}_n} \xrightarrow{d} \chi^2_{(r)}$, as $n, T \to \infty$. This test is the analogous of the test derived in the case of PC estimation by Bai and Ng (2013). Testing for equal loadings is equivalent to testing for equal common components for all $t =, 1 \ldots, T$.

Table 4 shows the result of the test of five different null hypotheses. The first hypothesis we test (column (A)) is whether GDP and GDI, both measures of US aggregate output, have equal loadings and, as a consequence, have the same common component. Our test does not reject the hypothesis of equal loadings. This result supports the idea recently explored in the literature that combining GDP and GDI can better estimate aggregate output

Figure 8: US DATA - GDP GROWTH FORECAST



In the left chart "Unc." is the unconditional forecast given data up to 2017:Q4; "Cond. 1" is the forecast conditioning on the value in 2018:Q1 of all the high-frequency variables in our dataset (stock prices, oil prices, surveys, and interest rates); "Cond. 2" is the forecast conditioning on the value in 2018:Q1 of all the high-frequency indicators the labor market indicators published in the BLS employment report; "Cond. 3" is the forecast conditional also on the value in 2018:Q1 of the CPI and PPI data, housing market indicators, and industrial production; "Cond. 4" is the common component estimated given all the data up to 2018:Q1; and, lastly, "Actual" is the actual value of GDP growth published by the BEA. All these indicators are released prior to GDP, which is released at the end of the following month: high-frequency indicators are available almost in real time, the labor report is published in the first week of the following month, and the CPI, PPI, Industrial Production, and housing indicators, are published about the mid of the following month.

In the right chart "Unc." is the unconditional forecast given data up to 2017:Q4; "Cond. 1" is the forecast conditional on payroll employment growing at the same average pace as in 2017; "Cond. 2" is the pessimistic scenario assuming that payroll employment grows at half the average pace as in 2017.

Table 4: US DATA - TESTING HYPOTHESIS ON THE LOADINGS

		(A)	(B)	(C)	(D)	(E)	(F)
$r=1$	$\mathrm{W}_{\widehat{\boldsymbol{\Lambda}}_n}$	0.1947	0.0087	0.0736	0.1358	1.7421	9.122
	p-value	0.66	0.93	0.79	0.71	0.19	0.00
$r=2$	$\mathrm{W}_{\widehat{\boldsymbol{\Lambda}}_n}$	0.2203	0.3708	0.5392	0.6124	3.0266	11.7768
	p-value	0.90	0.83	0.76	0.74	0.22	0.00
$r=3$	$\mathrm{W}_{\widehat{\boldsymbol{\Lambda}}_n}$	0.2462	4.6652	1.5638	0.8066	5.3210	13.6319
	p-value	0.97	0.20	0.67	0.85	0.15	0.00
$r=4$	$\mathrm{W}_{\widehat{\boldsymbol{\Lambda}}_n}$	0.3803	9.040	7.8699	1.2168	5.1899	110.8475
	p-value	0.98	0.06	0.10	0.88	0.27	0.00
$r=5$	$\mathrm{W}_{\widehat{\boldsymbol{\Lambda}}_n}$	1.8934	14.7642	9.7824	1.6959	5.4872	128.9513
	p-value	0.86	0.01	0.08	0.89	0.36	0.00
$r=6$	$\mathrm{W}_{\widehat{\boldsymbol{\Lambda}}_n}$	1.2704	28.6640	14.4703	4.7851	11.9874	126.7982
	p-value	0.97	0.00	0.02	0.57	0.06	0.00

The null hypotheses are: (A) $\boldsymbol{\lambda}_{\mathrm{GDP}} = \boldsymbol{\lambda}_{\mathrm{GDI}}$; (B) $\boldsymbol{\lambda}_{\mathrm{CPI}} = \boldsymbol{\lambda}_{\mathrm{PCE}}$; (C) $\boldsymbol{\lambda}_{\mathrm{CPI_{core}}} = \boldsymbol{\lambda}_{\mathrm{PCE_{core}}}$; (D) $\boldsymbol{\lambda}_{\mathrm{CPI_{energy}}} = \boldsymbol{\lambda}_{\mathrm{PCE_{energy}}}$; (E) $\boldsymbol{\lambda}_{\mathrm{CPI_{food}}} = \boldsymbol{\lambda}_{\mathrm{PCE_{food}}}$; and, (F) $\boldsymbol{\lambda}_{\mathrm{GDP}} = \boldsymbol{\lambda}_{\mathrm{Payroll}}$.

(e.g., Aruoba et al., 2016; Barigozzi and Luciani, 2018).

The second hypothesis (column (B)) that we test is whether CPI inflation and PCE price inflation, which are two alternative measures of consumer price inflation, have equal loadings. These indexes usually differ because they are constructed differently.[11] The test does not reject the null whenever $r < 6$, which we read as signaling that, indeed, CPI inflation and PCE price inflation respond in the same way to the common factors, and thus, their difference is just idiosyncratic or, perhaps, weak/local factors. Columns (C), (D), and (E) in Table 5 test whether the difference between the PCE and CPI core sub-index, energy sub-index, and food sub-index are just idiosyncratic. The test suggests this to be the case.

Finally, in column (F) of Table 4, we verify that if we test a non-sense hypothesis, the test rejects it. Specifically, we test whether the loadings of GDP growth and the growth in total non-farm employment are the same, which should not be the case. The test unequivocally reaches the same conclusion.

10 Concluding remarks

This paper provides the asymptotic properties of Quasi Maximum Likelihood (QML) estimation for large approximate dynamic factor models, implemented via the Kalman smoother and the Expectation Maximization (EM)

[11] The CPI, which captures the headlines in newspapers, determines the return on Treasury Inflation-Protected Securities, or TIPS, while the inflation objective of the Federal Reserve is specified in terms of PCE price inflation.

algorithm. Our results provide the statistical foundations of one of the most popular and successful methods for estimating high-dimensional factor models for time series commonly used in many public and private institutions to track and predict economic activity.

From a technical point of view, we show and prove that the EM approach is feasible even in the high-dimensional case, i.e., when the cross-sectional size n can be much larger than the sample size T, a point also made by Doz et al. (2012). Then, we show that the EM estimator converges at the same rate as the Principal Components (PC) estimator. Moreover, we show that the EM estimator of the loadings is always as efficient as the PC estimator, while the Kalman smoother estimator of the factors is more efficient than the PC estimator if the idiosyncratic covariance is sparse enough.

Compared to the standard PC estimator, the EM approach has the main advantage of allowing the user to easily impose restrictions that reflect any prior knowledge about the data on the model. The user can impose these restrictions because the state-space formulation and the Kalman Smoother allow explicit modeling and estimation of the dynamic evolution of the latent factors and deal with data irregularly spaced in time. Moreover, in the M-step, the user can impose restrictions on the parameters, thus allowing for constrained QML estimation. In contrast, the user cannot use the PC estimator to model the latent factors's dynamics. In an application on a dataset of US macroeconomic time series, we show that the EM algorithm can produce estimates of the common component that track the dynamics of the observed series better than the PC estimator, especially those series displaying periods of high persistence and regime changes, like inflation and interest rates. This result suggests that the Kalman smoother might be more robust to local deviations from stationarity, a feature already highlighted by Kálman (1960) and Kálman and Bucy (1961).

References

Ahn, H. J. and M. Luciani (2024). Common and idiosyncratic inflation. Finance and Economics Discussion Series 2020-024r1, Board of Governors of the Federal Reserve System.

Altavilla, C., R. Giacomini, and G. Ragusa (2017). Anchoring the yield curve using survey expectations. *Journal of Applied Econometrics 32*, 1055–1068.

Amemiya, Y., W. A. Fuller, and S. G. Pantula (1987). The asymptotic distributions of some estimators for a factor analysis model. *Journal of Multivariate Analysis 22*, 51–64.

Anderson, B. D. O. and M. Deistler (2008). Generalized linear dynamic factor models-A structure theory. In *Proceedings of the 47th IEEE Conference on Decision and Control*, pp. 1980–1985.

Anderson, B. D. O. and J. B. Moore (1979). *Optimal Filtering.* Dover Publications, Inc.

Anderson, T. W. and H. Rubin (1956). Statistical inference in factor analysis. In *Proceedings of the third Berkeley symposium on mathematical statistics and probability*, Volume 5, pp. 111–150.

Aruoba, S. B., F. X. Diebold, J. Nalewaik, F. Schorfheide, and D. Song (2016). Improving GDP measurement: A measurement-error perspective. *Journal of Econometrics 191*, 384–397.

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

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 Y. Liao (2016). Efficient estimation of approximate factor models via penalized maximum likelihood. *Journal of Econometrics 191*, 1–18.

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

Bai, J. and S. Ng (2006). Confidence intervals for diffusion index forecasts and inference for factor augmented regressions. *Econometrica 74*, 1133–1150.

Bai, J. and S. Ng (2007). Determining the number of primitive shocks in factor models. *Journal of Business and Economic Statistics 25*, 52–60.

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

Bai, J. and S. Ng (2021). Matrix completion, counterfactuals, and factor analysis of missing data. *Journal of the American Statistical Association*, 1–18. available online.

Bai, J. and P. Wang (2015). Identification and Bayesian estimation of dynamic factor models. *Journal of Business & Economic Statistics 33*, 221–240.

Bakhshizadeh, M., A. Maleki, and V. H. de la Pena (2023). Sharp concentration results for heavy-tailed distributions. *Information and Inference: A Journal of the IMA 12*, 1655–1685.

Balakrishnan, S., M. J. Wainwright, and B. Yu (2017). Statistical guarantees for the EM algorithm: From population to sample-based analysis. *The Annals of Statistics 45*, 77–120.

Bańbura, M., D. Giannone, and M. Lenza (2015). Conditional forecasts and scenario analysis with vector autoregressions for large cross-sections. *International Journal of Forecasting 31*, 739–756.

Bańbura, M., D. Giannone, M. Modugno, and L. Reichlin (2013). Now-casting and the real-time data flow. In *Handbook of economic forecasting*, Volume 2, pp. 195–237. Elsevier.

Bańbura, M., D. Giannone, and L. Reichlin (2011). Nowcasting. In M. P. Clements and D. F. Hendry (Eds.), *Oxford Handbook on Economic Forecasting*. New York: Oxford University Press.

Bańbura, M. and M. Modugno (2014). Maximum likelihood estimation of factor models on datasets with arbitrary pattern of missing data. *Journal of Applied Econometrics 29*, 133–160.

Barigozzi, M. (2023). Asymptotic equivalence of principal component and quasi maximum likelihood estimators in large approximate factor models. Technical Report arXiv:2307.09864.

Barigozzi, M. (2024). Quasi maximum likelihood estimation of high-dimensional factor models. In *Oxford Research Encyclopedia of Economics and Finance*. Oxford University Press.

Barigozzi, M., A. Cuzzola, M. Grazzi, and D. Moschella (2024). Factoring in the micro: A transaction-level dynamic factor approach to the decomposition of export volatility. *Oxford Bulletin of Economics and Statistics*. forthcoming.

Barigozzi, M., M. Lippi, and M. Luciani (2021). Large-dimensional dynamic factor models: Estimation of impulse-response functions with $I(1)$ cointegrated factors. *Journal of Econometrics 221*, 455–482.

Barigozzi, M. and M. Luciani (2018). Do National Account statistics underestimate US real output growth? FEDS Notes 2018-01-09, Board of Governors of the Federal Reserve System.

Barigozzi, M. and M. Luciani (2023). Measuring the output gap using large datasets. *The Review of Economics and Statistics 105*, 1500–1514.

Bernanke, B. S., J. Boivin, and P. S. Eliasz (2005). Measuring the effects of monetary policy: A Factor-Augmented Vector Autoregressive (FAVAR) approach. *The Quarterly Journal of Economics 120*, 387–422.

J. Boivin and S. Ng (2006). Are more data always better for factor analysis? *Journal of Econometrics 132*, 169–194.

Bosq, D. (2012). *Nonparametric statistics for stochastic processes: estimation and prediction.* Springer Science & Business Media.

Bradley, R. C. (2005). Basic properties of strong mixing conditions. a survey and some open questions. *Probability Surveys 2*, 107–144.

Breitung, J. and J. Tenhofen (2011). GLS estimation of dynamic factor models. *Journal of the American Statistical Association 106*, 1150–1166.

Cascaldi-Garcia, D., M. Luciani, and M. Modugno (2023). Lessons from nowcasting GDP across the world. International Finance Discussion Papers 1385, Board of Governors of the Federal Reserve System.

Chamberlain, G. and M. Rothschild (1983). Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica 51*, 1281–1304.

Cochrane, D. and G. H. Orcutt (1949). Application of least squares regression to relationships containing auto-correlated error terms. *Journal of the American statistical association 44*, 32–61.

Coroneo, L., D. Giannone, and M. Modugno (2016). Unspanned macroeconomic factors in the yield curve. *Journal of Business and Economic Statistics 34*, 472–485.

D'Agostino, A., D. Giannone, M. Lenza, and M. Modugno (2016). Nowcasting business cycles: A Bayesian approach to dynamic heterogeneous factor models. In S. Koopman and E. Hillebrand (Eds.), *Dynamic Factor Models*, Volume 35 of *Advances in Econometrics*, pp. 569–594. Emerald Publishing Ltd.

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.

Doz, C., D. Giannone, and L. Reichlin (2011). A two-step estimator for large approximate dynamic factor models based on Kalman filtering. *Journal of Econometrics 164*, 188–205.

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.

Durbin, J. and S. J. Koopman (2012). *Time Series Analysis by State Space Methods.* Oxford University Press.

Fan, J. and Y. Liao (2022). Learning latent factors from diversified projections and its applications to over-estimated and weak factors. *Journal of the American Statistical Association 117*, 909–924.

Fan, J., Y. Liao, and M. Mincheva (2011). High dimensional covariance matrix estimation in approximate factor models. *The Annals of Statistics 39*, 3320.

Fan, J., Y. Liao, and M. Mincheva (2013). Large covariance estimation by thresholding principal orthogonal complements. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 75*, 603–680.

Fan, J., R. Masini, and M. C. Medeiros (2022). Do we exploit all information for counterfactual analysis? benefits of factor models and idiosyncratic correction. *Journal of the American Statistical Association 117*, 574–590.

Forni, M., D. Giannone, M. Lippi, and L. Reichlin (2009). Opening the black box: Structural factor models versus structural VARs. *Econometric Theory 25*, 1319–1347.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2000). The Generalized Dynamic Factor Model: Identification and estimation. *The Review of Economics and Statistics 82*, 540–554.

Francq, C. and J.-M. Zakoïan (2006). Mixing properties of a general class of GARCH (1,1) models without moment assumptions on the observed process. *Econometric Theory 22*, 815–834.

Fresoli, D., P. Poncela, and E. Ruiz (2024). Dealing with idiosyncratic cross-correlation when constructing confidence regions for PC factors. Technical Report arXiv:2407.06883.

Gao, Z., J. Guo, and Y. Ma (2021). A note on statistical analysis of factor models of high dimension. *Science China Mathematics 64*, 1905–1916.

Geweke, J. F. (1993). A dynamic index model for large cross sections. Comment. In *Business cycles, indicators and forecasting*. University of Chicago Press.

Ghahramani, Z. and G. E. Hinton (1996). Parameter estimation for linear dynamical systems. Technical report, Cambridge University. mimeo.

Giannone, D., M. Lenza, and L. Reichlin (2019). Money, credit, monetary policy and the business cycle in the euro area: what has changed since the crisis? *International Journal of Central Banking 15*, 137–173.

Giannone, D., L. Reichlin, and L. Sala (2006). Tracking Greenspan: Systematic and nonsystematic monetary policy revisited. Discussion papers 3550, CEPR.

Giannone, D., L. Reichlin, and D. Small (2008). Nowcasting: The real-time informational content of macroeconomic data. *Journal of Monetary Economics 55*, 665–676.

Hannan, E. J. (1970). *Multiple time series*. John Wiley & Sons.

Hannan, E. J. (1980). The estimation of the order of an ARMA process. *The Annals of Statistics 8*, 1071–1081.

Harvey, A. C. (1990). *Forecasting, structural time series models and the Kalman filter*. Cambridge University Press.

Harvey, A. C. (1996). Intervention analysis with control groups. *International Statistical Review 64*, 313–328.

Harvey, A. C. and D. Delle Monache (2009). Computing the mean square error of unobserved components extracted by misspecified time series models. *Journal of Economic Dynamics and Control 33*, 283–295.

Harvey, A. C. and S. Peters (1990). Estimation procedures for structural time series models. *Journal of Forecasting 9*, 89–108.

Heaton, C. and V. Solo (2004). Identification of causal factor models of stationary time series. *The Econometrics Journal 7*, 618–627.

Ibragimov, I. A. (1962). Some limit theorems for stationary processes. *Theory of Probability and its Applications 7*, 349–382.

Jungbacker, B. and S. J. Koopman (2015). Likelihood-based dynamic factor analysis for measurement and forecasting. *Econometrics Journal 18*, C1–C21.

Jungbacker, B., S. J. Koopman, and M. Van der Wel (2011). Maximum likelihood estimation for dynamic factor models with missing data. *Journal of Economic Dynamics and Control 35*, 1358–1368.

Jungbacker, B., S. J. Koopman, and M. Van der Wel (2014). Smooth dynamic factor analysis with application to the US term structure of interest rates. *Journal of Applied Econometrics 29*, 65–90.

Juvenal, L. and I. Petrella (2015). Speculation in the oil market. *Journal of Applied Econometrics 30*, 1099–1255.

Kálman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering 82*, 35–45.

Kálman, R. E. and Bucy, R. S. (1961). New Results in Linear Filtering and Prediction Theory. *Journal of Basic Engineering 83*, 95–108.

Kapetanios, G. and M. Marcellino (2009). A parametric estimation method for dynamic factor models of large dimensions. *Journal of Time Series Analysis 30*, 208–238.

Kim, H. H. and N. R. Swanson (2018). Methods for backcasting, nowcasting and forecasting using factor-MIDAS: With an application to korean GDP. *Journal of Forecasting 37*, 281–302.

Kim, M. S. (2022). Robust inference for diffusion-index forecasts with cross-sectionally dependent data. *Journal of Business & Economic Statistics 40*, 1153–1167.

Koopman, S. J. and G. Mesters (2017). Empirical Bayes methods for dynamic factor models. *Review of Economics and Statistics 99*, 486–498.

Koopman, S. J. and M. van der Wel (2013). Forecasting the us term structure of interest rates using a macroeconomic smooth dynamic factor model. *International Journal of Forecasting 29*, 676–694.

Kose, M. A., C. Otrok, and C. H. Whiteman (2003). International business cycles: World, region, and country-specific factors. *The American Economic Review 93*, 1216–1239.

Kotz, S., N. Balakrishnan, and N. L. Johnson (2004). *Continuous multivariate distributions. Volume 1: Models and applications*. John Wiley & Sons.

Kuchibhotla, A. K., L. D. Brown, A. Buja, E. I. George, and L. Zhao (2023). Uniform-in-submodel bounds for linear regression in a model-free framework. *Econometric Theory 39*, 1202–1248.

Kuchibhotla, A. K. and A. Chakrabortty (2022). Moving beyond sub-Gaussianity in high-dimensional statistics: Applications in covariance estimation and linear regression. *Information and Inference: A Journal of the IMA 11*, 1389–1456.

Lawley, D. N. and A. E. Maxwell (1971). *Factor Analysis as a Statistical Method*. Butterworths, London.

Lehman, E. H. (1963). Shapes, moments and estimators of the weibull distribution. *IEEE Transactions on Reliability 12*, 32–38.

Lehmann, E. L. and G. Casella (2006). *Theory of point estimation*. Springer Science & Business Media.

Lin, J. and G. Michailidis (2020). System identification of high-dimensional linear dynamical systems with serially correlated output noise components. *IEEE Transactions on Signal Processing 68*, 5573–5587.

Linton, O. B., H. Tang, and J. Wu (2022). A structural dynamic factor model for daily global stock market returns. Technical Report arXiv:2202.03638.

Lippi, M., M. Deistler, and B. D. O. Anderson (2021). High-dimensional dynamic factor models: A selective survey and lines of future research. mimeo, EIEF.

Luciani, M. (2014). Forecasting with Approximate Dynamic Factor Models: The role of non-pervasive shocks. *International Journal of Forecasing 30*, 20–29.

Luciani, M. (2015). Monetary policy and the housing market: A structural factor analysis. *Journal of Applied Econometrics 30*, 199–218.

Luciani, M. and L. Ricci (2014). Nowcasting Norway. *International Journal of Central Banking 10*, 215–248.

Mao, J., Z. Gao, B.-Y. Jing, and J. Guo (2024). On the statistical analysis of high-dimensional factor models. *Statistical Papers*, 1–29. available online.

Marcellino, M. and V. Sivec (2016). Monetary, fiscal and oil shocks: Evidence based on mixed frequency structural FAVARs. *Journal of Econometrics 193*, 335–348.

Mariano, R. S. and Y. Murasawa (2003). A new coincident index of business cycles based on monthly and quarterly series. *Journal of Applied Econometrics 18*, 427–443.

McLachlan, G. and T. Krishnan (2007). *The EM algorithm and extensions*, Volume 382. John Wiley & Sons.

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.

Merlevède, F., M. Peligrad, and E. Rio (2011). A Bernstein type inequality and moderate deviations for weakly dependent sequences. *Probability Theory and Related Fields 151*, 435–474.

Mikosch, T. and A. V. Nagaev (1998). Large deviations of heavy-tailed sums with applications in insurance. *Extremes 1*, 81–110.

Mosley, L., T. T. Chan, and A. Gibberd (2024). The sparse dynamic factor model: a regularised quasi-maximum likelihood approach. *Statistics and Computing 34*, 1–19.

Ng, C. T., C. Y. Yau, and N. H. Chan (2015). Likelihood inferences for high-dimensional factor analysis of time series with applications in finance. *Journal of Computational and Graphical Statistics 24*(3), 866–884.

Ng, S. and S. Scanlan (2024). Constructing high frequency economic indicators by imputation. *Econometrics Journal 27*, C1–C30.

Pham, T. D. and L. T. Tran (1985). Some mixing properties of time series models. *Stochastic processes and their applications 19*, 297–303.

Poignard, B. and Y. Terada (2020). Statistical analysis of sparse approximate factor models. *Electronic Journal of Statistics 14*, 3315–3365.

Poncela, P., E. Ruiz, and K. Miranda (2021). Factor extraction using Kalman filter and smoothing: This is not just another survey. *International Journal of Forecasting 37*, 1399–1425.

Quah, D. and T. J. Sargent (1993). A dynamic index model for large cross sections. In *Business cycles, indicators and forecasting*. University of Chicago Press.

Reis, R. and M. W. Watson (2010). Relative goods' prices, pure inflation, and the Phillips correlation. *American Economic Journal Macroeconomics 2*, 128–157.

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

Ruiz, E. and P. Poncela (2022). Factor extraction in Dynamic Factor Models: Using Kalman Filter and Principal Components in practice. *Foundations and Trends in Econometrics 12*, 121–231.

Ruud, P. A. (1991). Extensions of estimation methods using the EM algorithm. *Journal of Econometrics 49*(3), 305–341.

Sargent, T. J. and C. A. Sims (1977). Business cycle modeling without pretending to have too much a priori economic theory. In *New methods in business cycle research*. Federal Reserve Bank of Minneapolis.

Shumway, R. H. and D. S. Stoffer (1982). An approach to time series smoothing and forecasting using the EM algorithm. *Journal of Time Series Analysis 3*, 253–264.

Spearman, C. (1904). General intelligence objectively determined and measured. *American Journal of Psychology 15*, 201–293.

Steyn, H. S. (1960). On regression properties of multivariate probability functions of Pearson's types. In *Indagationes Mathematicae (Proceedings)*, Volume 63, pp. 302–311. Elsevier.

Stock, J. H. and M. W. Watson (1989). New indexes of coincident and leading economic indicators. In O. J. Blanchard and S. Fischer (Eds.), *NBER Macroeconomics Annual 1989*. MIT press.

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

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.

Sundberg, R. (2019). *Statistical modelling by exponential families*. Cambridge University Press.

Tanner, M. A. and W. H. Wong (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association 82*, 528–540.

Tipping, M. E. and C. M. Bishop (1999). Probabilistic principal component analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 61*, 611–622.

Vershynin, R. (2018). *High-dimensional probability: An introduction with applications in data science*, Volume 47. Cambridge University Press.

Vladimirova, M., S. Girard, H. Nguyen, and J. Arbel (2020). Sub-Weibull distributions: Generalizing sub-Gaussian and sub-Exponential properties to heavier tailed distributions. *Stat 9*(1), e318.

Wang, S., H. Yang, and C. Yao (2019). On the penalized maximum likelihood estimation of high-dimensional approximate factor model. *Computational Statistics 34*, 819–846.

Watson, M. W. and R. F. Engle (1983). Alternative algorithms for the estimation of dynamic factor, mimic and varying coefficients regression models. *Journal of Econometrics 23*, 385–400.

Westerlund, J. and J.-P. Urbain (2015). Cross-sectional averages versus principal components. *Journal of Econometrics 185*, 372–377.

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

Xiong, R. and M. Pelger (2023). Large dimensional latent factor modeling with missing observations and applications to causal inference. *Journal of Econometrics 233*, 271–301.

Supplementary material for the paper

Quasi Maximum Likelihood Estimation and Inference of Large Approximate Dynamic Factor Models via the EM algorithm

Matteo Barigozzi
Università di Bologna
matteo.barigozzi@unibo.it

Matteo Luciani
Federal Reserve Board
matteo.luciani@frb.gov

Table of contents

M. Barigozzi gratefully acknowledges financial support from MIUR (PRIN2020, Grant 2020N9YFFE).

Disclaimer: the views expressed in this paper are those of the authors and do not necessarily reflect the views and policies of the Board of Governors or the Federal Reserve System.

Notation

Parameters

Parameters	
$\boldsymbol{\varphi}_n$	true value
$\underline{\boldsymbol{\varphi}}_n$	generic value
$\widehat{\boldsymbol{\varphi}}_n^*$	QML estimator maximizing the log-likelihood (5)
$\widehat{\boldsymbol{\varphi}}_n^\dagger$	QML estimator maximizing the log-likelihood in Bai and Li (2016, Eq. 3)
$\widehat{\boldsymbol{\varphi}}_n^{(0)}$	PC estimator used in E-step at iteration 0
$\widehat{\boldsymbol{\varphi}}_n^{(k)}$	estimator used in E-step at iteration $k>0$
$\widehat{\boldsymbol{\varphi}}_n^{(k+1)}$	estimator computed in M-step at iteration $k\geq 0$
$\widehat{\boldsymbol{\varphi}}_n \equiv \widehat{\boldsymbol{\varphi}}_n^{(k^*+1)}$	final estimator computed in M-step at iteration k^*

An analogous notation is used for the sub-vectors of parameters $\boldsymbol{\phi}_n$ and $\boldsymbol{\theta}$ and all their elements.

Factors

Factors	
$\mathbf{F}_t$	true value
$\widetilde{\mathbf{F}}_t$	PC estimator
$\mathbf{F}_{t\vert s}$ and $\mathbf{P}_{t\vert s}$ $(s=t-1,t,T)$	estimator and its pseudo-MSE computed via KF-KS using $\boldsymbol{\varphi}_n$
$\mathbf{F}_{0,t\vert s}$ and $\mathbf{P}_{0,t\vert s}$ $(s=t-1,t,T)$	estimator and its MSE computed via KF-KS using $\boldsymbol{\varphi}_n$ but with $\boldsymbol{\Gamma}_n^\xi$
$\mathbf{F}^*_{t\vert s}$ and $\mathbf{P}^*_{t\vert s}$ $(s=t-1,t,T)$	estimator and its pseudo-MSE computed via KF-KS using $\widehat{\boldsymbol{\varphi}}_n^*$
$\mathbf{F}^{(k)}_{t\vert s}$ and $\mathbf{P}^{(k)}_{t\vert s}$ $(s=t-1,t,T)$	estimator and its pseudo-MSE computed via KF-KS using $\widehat{\boldsymbol{\varphi}}_n^{(k)}$, $k\geq 0$
$\widehat{\mathbf{F}}_{t\vert t}\equiv \mathbf{F}^{(k^*+1)}_{t\vert t}$	final estimator computed via KF at iteration (k^*+1) using $\widehat{\boldsymbol{\varphi}}_n$
$\widehat{\mathbf{P}}_{t\vert t}\equiv \mathbf{P}^{(k^*+1)}_{t\vert t}$	pseudo-MSE computed via KF at iteration (k^*+1) using $\widehat{\boldsymbol{\varphi}}_n$
$\widehat{\mathbf{F}}_t \equiv \widehat{\mathbf{F}}_{t\vert T}\equiv \mathbf{F}^{(k^*+1)}_{t\vert T}$	final estimator computed via KS at iteration (k^*+1) using $\widehat{\boldsymbol{\varphi}}_n$
$\widehat{\mathbf{P}}_{t\vert T}\equiv \mathbf{P}^{(k^*+1)}_{t\vert T}$	pseudo-MSE computed via KS at iteration (k^*+1) using $\widehat{\boldsymbol{\varphi}}_n$
$\widehat{\mathbf{F}}_t^{\mathrm{WLS}}$	WLS estimator computed using $\widehat{\boldsymbol{\varphi}}_n$

A Further details on estimation

Hereafter, we assume without loss of generality that $\boldsymbol{\mu}_n = \mathbf{0}_n$ and $p_F = 1$ with $\mathbf{A}\equiv\mathbf{A}_1$. When $p_F>1$, it is enough to write the VAR in (4) in companion form and to modify the estimation accordingly, using the augmented state vector $(\mathbf{F}_t' \cdots \mathbf{F}_{t-p_F+1}')'$.

A.1 Principal Component estimators

Let $\widehat{\boldsymbol{\Gamma}}_n^x$ be the sample covariance matrix of the data and denote as $\widehat{\mathbf{M}}_n^x$ the diagonal matrix with entries the r-largest eigenvalues of $\widehat{\boldsymbol{\Gamma}}_n^x$, and as $\widehat{\mathbf{V}}_n^x$ the $n\times r$ matrix of the corresponding normalized eigenvectors. Moreover, let $\widehat{\boldsymbol{\mathcal{S}}}^{(0)}$ be a $r\times r$

diagonal matrix with entries $\mathbb{I}([\widehat{\mathbf{V}}_n^x]_{1j} \geq 0) - \mathbb{I}([\widehat{\mathbf{V}}_n^x]_{1j} < 0)$, $j = 1, \dots, r$. Then,

$$\begin{aligned}
\widehat{\mathbf{\Lambda}}_n^{(0)} &= \widehat{\mathbf{V}}_n^x \widehat{\boldsymbol{\mathcal{S}}}^{(0)} (\widehat{\mathbf{M}}_n^x)^{1/2},\\
\widetilde{\mathbf{F}}_t &= (\widehat{\mathbf{\Lambda}}_n^{(0)\prime}\widehat{\mathbf{\Lambda}}_n^{(0)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)\prime}\mathbf{x}_{nt} = (\widehat{\mathbf{M}}_n^x)^{-1/2}\widehat{\boldsymbol{\mathcal{S}}}^{(0)}\widehat{\mathbf{V}}_n^{x\prime}\mathbf{x}_{nt},\\
\widetilde{\boldsymbol{\xi}}_{nt} &= \mathbf{x}_{nt} - \widehat{\mathbf{\Lambda}}_n^{(0)\prime}\widetilde{\mathbf{F}}_t, \quad t = 1, \dots, T\\
\widehat{\mathbf{A}}^{(0)} &= \left(\sum_{t=2}^T \widetilde{\mathbf{F}}_t \widetilde{\mathbf{F}}_{t-1}'\right)\left(\sum_{t=2}^T \widetilde{\mathbf{F}}_{t-1}\widetilde{\mathbf{F}}_{t-1}'\right)^{-1},\\
\widetilde{\mathbf{v}}_t &= \widetilde{\mathbf{F}}_t - \widehat{\mathbf{A}}^{(0)}\widetilde{\mathbf{F}}_{t-1}, \quad t = 1, \dots, T,\\
\widehat{\mathbf{\Gamma}}^{v(0)} &= T^{-1}\sum_{t=1}^T \widetilde{\mathbf{v}}_t\widetilde{\mathbf{v}}_t'.
\end{aligned}$$

Finally, letting $\widehat{\boldsymbol{\lambda}}_i^{(0)\prime}$ be the i-th row of $\widehat{\mathbf{\Lambda}}_n^{(0)}$, and $\widetilde{\xi}_{it}$ the ith component of $\widetilde{\boldsymbol{\xi}}_{nt}$, then

$$\widehat{\sigma}_i^{2(0)} = T^{-1}\sum_{t=1}^T \widetilde{\xi}_{it}^2, \qquad i = 1, \dots, n.$$

The vector of initial estimates of parameters is then:

$$\widehat{\boldsymbol{\varphi}}_n^{(0)} = \left(\text{vec}(\widehat{\mathbf{\Lambda}}_n^{(0)})' \ \widehat{\sigma}_1^{2(0)} \cdots \widehat{\sigma}_n^{2(0)} \ \text{vec}(\widehat{\mathbf{A}}^{(0)})' \ \text{vech}(\widehat{\mathbf{\Gamma}}^{v(0)})'\right)',$$

and it is used to run the first iteration of the EM algorithm.

A.2 Kalman filter and smoother

The following iterations are stated for given initial conditions $\mathbf{F}_{0|0}$ and $\mathbf{P}_{0|0}$ and given the true parameters $\boldsymbol{\varphi}_n$.

Forward iterations - Filtering

The Kalman filter is based on the forward iterations for $t = 1, \dots, T$:

$$\mathbf{F}_{t|t-1} = \mathbf{A}\mathbf{F}_{t-1|t-1}, \tag{A.1}$$

$$\mathbf{P}_{t|t-1} = \mathbf{A}\mathbf{P}_{t-1|t-1}\mathbf{A}' + \mathbf{\Gamma}^v, \tag{A.2}$$

$$\mathbf{F}_{t|t} = \mathbf{F}_{t|t-1} + \mathbf{P}_{t|t-1}\mathbf{\Lambda}_n'(\mathbf{\Lambda}_n\mathbf{P}_{t|t-1}\mathbf{\Lambda}_n' + \mathbf{\Sigma}_n^\xi)^{-1}(\mathbf{x}_{nt} - \mathbf{\Lambda}_n\mathbf{F}_{t|t-1}), \tag{A.3}$$

$$\mathbf{P}_{t|t} = \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1}\mathbf{\Lambda}_n'(\mathbf{\Lambda}_n\mathbf{P}_{t|t-1}\mathbf{\Lambda}_n' + \mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n\mathbf{P}_{t|t-1}. \tag{A.4}$$

Moreover, by combining (A.2) and (A.4), we obtain the Riccati difference equation:

$$\mathbf{P}_{t+1|t} - \mathbf{A}\mathbf{P}_{t|t-1}\mathbf{A}' + \mathbf{A}\mathbf{P}_{t|t-1}\mathbf{\Lambda}_n'(\mathbf{\Lambda}_n\mathbf{P}_{t|t-1}\mathbf{\Lambda}_n' + \mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n\mathbf{P}_{t|t-1}\mathbf{A}' = \mathbf{\Gamma}^v. \tag{A.5}$$

Backward iterations - Smoothing

The Kalman smoother is then based on the backward iterations for $t = T, \dots, 1$:

$$\mathbf{F}_{t|T} = \mathbf{F}_{t|t} + \mathbf{P}_{t|t}\mathbf{A}'\mathbf{P}_{t+1|t}^{-1}(\mathbf{F}_{t+1|T} - \mathbf{F}_{t+1|t}), \tag{A.6}$$

$$\mathbf{P}_{t|T} = \mathbf{P}_{t|t} + \mathbf{P}_{t|t}\mathbf{A}'\mathbf{P}_{t+1|t}^{-1}(\mathbf{P}_{t+1|T} - \mathbf{P}_{t+1|t})\mathbf{P}_{t+1|t}^{-1}\mathbf{A}\mathbf{P}_{t|t}. \tag{A.7}$$

Finally, $\mathbf{C}_{t,t-1|T}$ can be obtained from a state space model with an augmented state vector containing both $\mathbf{F}_t$ and $\mathbf{F}_{t-1}$, by taking the $r \times r$ off-diagonal block of the $2r \times 2r$ matrix defined in (A.7) but for the augmented model.

An equivalent way of implementing (A.6), which does not require matrix inversion is in Durbin and Koopman (2012, Chapter 4.4, pp.87-91), which is defined by the backward iterations for $t = T, \dots, 1$

$$\mathbf{F}_{t|T} = \mathbf{F}_{t|t-1} + \mathbf{P}_{t|t-1}\mathbf{r}_{t-1}, \tag{A.8}$$

$$\mathbf{r}_{t-1} = \mathbf{\Lambda}_n'(\mathbf{\Lambda}_n\mathbf{P}_{t|t-1}\mathbf{\Lambda}_n' + \mathbf{\Sigma}_n^\xi)^{-1}(\mathbf{x}_t - \mathbf{\Lambda}_n\mathbf{F}_{t|t-1}) + \mathbf{L}_t'\mathbf{r}_t, \tag{A.9}$$

$$\mathbf{P}_{t|T} = \mathbf{P}_{t|t-1}(\mathbf{I}_r - \mathbf{N}_{t-1}\mathbf{P}_{t|t-1}), \tag{A.10}$$

$$\mathbf{N}_{t-1} = \mathbf{\Lambda}_n'(\mathbf{\Lambda}_n\mathbf{P}_{t|t-1}\mathbf{\Lambda}_n' + \mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n + \mathbf{L}_t'\mathbf{N}_t\mathbf{L}_t, \tag{A.11}$$

$$\mathbf{L}_t = \mathbf{A} - \mathbf{A}\mathbf{P}_{t|t-1}\mathbf{\Lambda}_n'(\mathbf{\Lambda}_n\mathbf{P}_{t|t-1}\mathbf{\Lambda}_n' + \mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n, \tag{A.12}$$

$$\mathbf{C}_{t,t+1|T} = \mathbf{P}_{t|t-1}\mathbf{L}_t'(\mathbf{I}_r - \mathbf{N}_t\mathbf{P}_{t+1|t}), \qquad \mathbf{C}_{t,t-1|T} = \mathbf{C}_{t,t+1|T}', \tag{A.13}$$

where $\mathbf{r}_T = \mathbf{0}_r$, $\mathbf{N}_T = \mathbf{0}_r$ and by construction $\mathbf{A}\mathbf{P}_{t|t} = \mathbf{L}_t\mathbf{P}_{t|t-1}$.

Intialization of Kalman filter and smoother

The Kalman filter is initialized as follows. At the first iteration of the EM algorithm, i.e., when $k = 0$, we set $\mathbf{F}_{0|0}^{(0)} = \mathbf{0}_r$ and $\mathbf{P}_{0|0}^{(0)} = \mathbf{I}_r$, consistently with Assumption 6(b). Other initializations as $\mathbf{P}_{0|0}^{(0)} = \kappa_0\mathbf{I}_r$ for some finite real $\kappa_0 > 0$ are also possible. At any successive iteration of the EM algorithm, i.e, when $k > 0$, we set $\mathbf{F}_{0|0}^{(k)} = \mathbf{F}_{0|T}^{(k-1)}$ and $\mathbf{P}_{0|0}^{(0)} = \mathbf{I}_r$.

To run the Kalman smoother we start using the last predictions of the Kalman filter. Thus, for any $k \geq 0$, we set $\mathbf{F}_{T+1|T}^{(k)} = \widehat{\mathbf{A}}^{(k)}\mathbf{F}_{T|T}^{(k)}$ where $\mathbf{F}_{T|T}^{(k)}$ is obtained from (A.3), and we set $\mathbf{P}_{T+1|T}^{(k)} = \widehat{\mathbf{A}}^{(k)}\mathbf{P}_{T|T}^{(k)}\widehat{\mathbf{A}}^{(k)\prime} + \widehat{\mathbf{\Gamma}}^{v(k)}$, where $\mathbf{P}_{T|T}^{(k)}$ is obtained from (A.4).

A.3 Stopping rule for the EM algorithm

To stop the EM algorithm we adopt the following convergence rule. We fix a maximum finite number of iterations $k_{\max}$, and we stop it at the first iteration $k^* \leq k_{\max}$ such that:

$$\Delta\ell_{k^*} = \frac{\left|\ell(\boldsymbol{X}_{nT}; \widehat{\boldsymbol{\varphi}}_n^{(k^*+1)}) - \ell(\boldsymbol{X}_{nT}; \widehat{\boldsymbol{\varphi}}_n^{(k^*)})\right|}{\frac{1}{2}\left|\ell(\boldsymbol{X}_{nT}; \widehat{\boldsymbol{\varphi}}_n^{(k^*+1)}) + \ell(\boldsymbol{X}_{nT}; \widehat{\boldsymbol{\varphi}}_n^{(k^*)})\right|} < \varepsilon, \tag{A.14}$$

where ε is a pre-specified tolerance level. In this case, the log-likelihood is computed using its prediction error formulation obtained from the Kalman filter:

$$\ell(\boldsymbol{X}_{nT}; \underline{\boldsymbol{\varphi}}_n) = -\frac{1}{2}\sum_{t=1}^T \log\det(\underline{\mathbf{\Lambda}}_n\underline{\mathbf{P}}_{t|t-1}\underline{\mathbf{\Lambda}}_n' + \underline{\mathbf{\Sigma}}_n^\xi) - \frac{1}{2}\sum_{t=1}^T(\mathbf{x}_{nt} - \underline{\mathbf{\Lambda}}_n\underline{\mathbf{F}}_{t|t-1})'(\underline{\mathbf{\Lambda}}_n\underline{\mathbf{P}}_{t|t-1}\underline{\mathbf{\Lambda}}_n' + \underline{\mathbf{\Sigma}}_n^\xi)^{-1}(\mathbf{x}_{nt} - \underline{\mathbf{\Lambda}}_n\underline{\mathbf{F}}_{t|t-1}),$$

where $\underline{\mathbf{F}}_{t|t-1}$ and $\underline{\mathbf{P}}_{t|t-1}$ are computed using (A.1) and (A.2), respectively, when using generic values of the parameters. Similar convergence criteria can be found in Booth and Hobert (1999) and McLachlan and Krishnan (2007, Chapter 4.9).

B Proof of main results

Hereafter, we assume without loss of generality that $\boldsymbol{\mu}_n = \mathbf{0}_n$ and $p_F = 1$ with $\mathbf{A} \equiv \mathbf{A}_1$.

B.1 Proof of Proposition 1

Consider the EM algorithm initialized using the PC estimators of the parameters as defined in Section A.1. At $k^* = 0$, from (13), we have

$$\widehat{\boldsymbol{\lambda}}_i^{(1)} = \left(T^{-1}\sum_{t=1}^T \mathbf{F}_{t|T}^{(0)}\mathbf{F}_{t|T}^{(0)\prime} + \mathbf{P}_{t|T}^{(0)}\right)^{-1}\left(T^{-1}\sum_{t=1}^T \mathbf{F}_{t|T}^{(0)}x_{it}\right). \tag{B.1}$$

Now,

$$\left\|T^{-1}\sum_{t=1}^T \mathbf{F}_{t|T}^{(0)}\mathbf{F}_{t|T}^{(0)\prime} - T^{-1}\sum_{t=1}^T \mathbf{F}_t\mathbf{F}_t'\right\| \leq 2\left\|T^{-1}\sum_{t=1}^T(\mathbf{F}_{t|T}^{(0)} - \mathbf{F}_t)\mathbf{F}_t'\right\| + \left\|T^{-1}\sum_{t=1}^T(\mathbf{F}_{t|T}^{(0)} - \mathbf{F}_t)(\mathbf{F}_{t|T}^{(0)} - \mathbf{F}_t)'\right\|, \tag{B.2}$$

and

$$\left\| T^{-1}\sum_{t=1}^T \mathbf{F}_{t|T}^{(0)} x_{it} - T^{-1}\sum_{t=1}^T \mathbf{F}_t x_{it}\right\| \le \left\| T^{-1}\sum_{t=1}^T (\mathbf{F}_{t|T}^{(0)} - \mathbf{F}_t) x_{it}\right\|. \tag{B.3}$$

Throughout, let $\boldsymbol{y}_t = \mathbf{F}_t$ or $\boldsymbol{y}_t = x_{it}$. Then, we have to consider

$$\begin{aligned}\left\| T^{-1}\sum_{t=1}^T (\mathbf{F}_{t|T}^{(0)} - \mathbf{F}_t)\boldsymbol{y}_t'\right\| &\le \left\| T^{-1}\sum_{t=1}^T (\mathbf{F}_{t|T}^{(0)} - \mathbf{F}_{t|t}^{(0)})\boldsymbol{y}_t'\right\| + \left\| T^{-1}\sum_{t=1}^T (\mathbf{F}_{t|t}^{(0)} - \widehat{\mathbf{F}}_t^{\mathrm{WLS}(0)})\boldsymbol{y}_t'\right\| + \left\| T^{-1}\sum_{t=1}^T (\widehat{\mathbf{F}}_t^{\mathrm{WLS}(0)} - \mathbf{F}_t)\boldsymbol{y}_t'\right\| \\ &= I + II + III, \quad \text{say.}\end{aligned} \tag{B.4}$$

Let us consider each term in (B.4). First,

$$\begin{aligned} I &\le \max_{t=1,\dots,T}\|\mathbf{P}_{t|t}^{(0)}\|\,\|\widehat{\mathbf{A}}^{(0)}\| \max_{t=1,\dots,T}\|(\mathbf{P}_{t+1|t}^{(0)})^{-1}\| \left\{ \left\| T^{-1}\sum_{t=1}^T \mathbf{F}_{t+1|T}^{(0)}\boldsymbol{y}_t'\right\| + \|\widehat{\mathbf{A}}^{(0)}\| \left\| T^{-1}\sum_{t=1}^T \mathbf{F}_{t+1|t+1}^{(0)}\boldsymbol{y}_t'\right\| \right\} \\ &= O_p(n^{-1}), \end{aligned} \tag{B.5}$$

by Lemmas D.8, D.11, and D.18, and since $\|\widehat{\mathbf{A}}^{(0)}\| \le \|\mathbf{A}\| + \|\widehat{\mathbf{A}}^{(0)} - \mathbf{A}\| = O_p(1)$, by Assumption 1(d) and Lemma D.3(i). Second, from (D.46) and (D.50) in the proof of Lemma D.16

$$II \le O_p(n^{-1})\left\{ \left\| n^{-1/2}T^{-1}\sum_{t=1}^T \mathbf{x}_{nt}\boldsymbol{y}_t'\right\| + \|\widehat{\mathbf{A}}^{(0)}\| \left\| T^{-1}\sum_{t=1}^T \mathbf{F}_{t-1|t-1}^{(0)}\boldsymbol{y}_t'\right\| \right\} = O_p(n^{-1}), \tag{B.6}$$

because of Lemma D.18 and since

$$\begin{aligned} &\left\| n^{-1/2}T^{-1}\sum_{t=1}^T \boldsymbol{\Lambda}_n\mathbf{F}_t\mathbf{F}_t'\boldsymbol{\lambda}_i\right\| \le n^{-1/2}\|\boldsymbol{\Lambda}_n\| \left\| T^{-1}\sum_{t=1}^T \mathbf{F}_t\mathbf{F}_t'\right\| M_\lambda = O_p(1), \\ &\left\| n^{-1/2}T^{-1}\sum_{t=1}^T \boldsymbol{\Lambda}_n\mathbf{F}_t\xi_{it}\right\| \le n^{-1/2}\|\boldsymbol{\Lambda}_n\| \left\| T^{-1}\sum_{t=1}^T \mathbf{F}_t\xi_{it}\right\| = O_p(T^{-1/2}), \\ &\left\| n^{-1/2}T^{-1}\sum_{t=1}^T \boldsymbol{\xi}_{nt}\mathbf{F}_t'\boldsymbol{\lambda}_i\right\| \le \left\| n^{-1/2}T^{-1}\sum_{t=1}^T \boldsymbol{\xi}_{nt}\mathbf{F}_t'\right\| M_\lambda = O_p(T^{-1/2}), \\ &\left\| n^{-1/2}T^{-1}\sum_{t=1}^T \boldsymbol{\xi}_{nt}\xi_{it}\right\| \le \left\| n^{-1/2}T^{-1}\sum_{t=1}^T \boldsymbol{\xi}_{nt}\xi_{it} - n^{-1/2}\mathbb{E}[\boldsymbol{\xi}_{nt}\xi_{it}]\right\| + n^{-1/2}\|\mathbb{E}[\boldsymbol{\xi}_{nt}\xi_{it}]\| = O_p(T^{-1/2}) + O(1), \end{aligned} \tag{B.7}$$

by Assumption 1(a), Lemmas C.2, C.12(i), C.12(ii), C.12(iii), and C.12(iv), and because $n^{-1}\|\mathbb{E}[\boldsymbol{\xi}_{nt}\xi_{it}]\|^2 \le n^{-1}\sum_{j=1}^n |\mathbb{E}[\xi_{jt}^2\xi_{it}^2]| \le \mathrm{K}_\xi$, by Assumption 2(d). Note that the first and third relations in (B.7) cover also the case $\boldsymbol{y}_t = \mathbf{F}_t'$.

Finally, let us consider the last term in (B.4). From (D.52) in the proof of Lemma D.17

$$\begin{aligned}
III \leq & \left\| T^{-1}\sum_{t=1}^{T}(\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}(\mathbf{\Lambda}_n-\widehat{\mathbf{\Lambda}}_n^{(0)})\mathbf{F}_t\boldsymbol{y}_t'\right\| \\
& + \left\| T^{-1}\sum_{t=1}^{T}(\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\boldsymbol{\xi}_{nt}\boldsymbol{y}_t'\right\| \\
\leq & \|(\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^{\xi})^{-1}\mathbf{\Lambda}_n)^{-1}\| \, \|\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^{\xi})^{-1}(\mathbf{\Lambda}_n-\widehat{\mathbf{\Lambda}}_n^{(0)})\| \left\| T^{-1}\sum_{t=1}^{T}\mathbf{F}_t\boldsymbol{y}_t'\right\| \\
& + \|n(\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)})^{-1}n^{-1/2}\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1} - n(\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^{\xi})^{-1}\mathbf{\Lambda}_n)^{-1}n^{-1/2}\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^{\xi})^{-1}\| \\
& \cdot n^{-1/2}\|\mathbf{\Lambda}_n-\widehat{\mathbf{\Lambda}}_n^{(0)}\| \left\| T^{-1}\sum_{t=1}^{T}\mathbf{F}_t\boldsymbol{y}_t'\right\| \\
& + n\|(\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)})^{-1}\|\, n^{-1}\left\| T^{-1}\sum_{t=1}^{T}\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^{\xi})^{-1}\boldsymbol{\xi}_{nt}\boldsymbol{y}_t'\right\| \\
& + n\|(\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)})^{-1}\|\, n^{-1}\left\| T^{-1}\sum_{t=1}^{T}\{\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}-\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^{\xi})^{-1}\}\boldsymbol{\xi}_{nt}\boldsymbol{y}_t'\right\| \\
= & III_a + III_b + III_c + III_d, \quad \text{say.}
\end{aligned} \tag{B.8}$$

Then,

$$III_a = O_p(n^{-1/2}T^{-1/2}), \tag{B.9}$$

by (B.7) and (D.54)-(D.56) in the proof of Lemma D.17. Moreover,

$$III_b = O_p(\max(n^{-2}, T^{-1})), \tag{B.10}$$

by (B.7) and Lemmas D.1(ii) and D.5(v). Regarding III_c, if $\boldsymbol{y}_t = \mathbf{F}_t$, we have

$$III_c = n\|(\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)})^{-1}\|\, n^{-1}\left\| T^{-1}\sum_{t=1}^{T}\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^{\xi})^{-1}\boldsymbol{\xi}_{nt}\mathbf{F}_t'\right\| = O_p(n^{-1/2}T^{-1/2}), \tag{B.11}$$

by Lemmas D.5(iii) and C.8(iv). If $\boldsymbol{y}_t = x_{it}$, we have

$$\begin{aligned}
III_c \leq & n\|(\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)})^{-1}\|\, n^{-1}\left\| T^{-1}\sum_{t=1}^{T}\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^{\xi})^{-1}\boldsymbol{\xi}_{nt}\mathbf{F}_t'\right\| \|\boldsymbol{\lambda}_i\| \\
& + n\|(\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)})^{-1}\| \left\| T^{-1}\sum_{t=1}^{T}\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^{\xi})^{-1}\boldsymbol{\xi}_{nt}\xi_{it}\right\| \\
= & O_p(n^{-1/2}T^{-1/2})),
\end{aligned} \tag{B.12}$$

by Assumption 1(a), and Lemmas D.5(iii), C.8(iv), and C.8(v). Last,

$$III_d = O_p(\max(n^{-1}, T^{-1/2})), \tag{B.13}$$

by (B.7) and Lemma D.5(v). From (B.8), (B.9), (B.10), (B.11), (B.12), and (B.13)

$$III = O_p(\max(n^{-1}, T^{-1/2})). \tag{B.14}$$

Combining (B.5), (B.6), and (B.14) we have

$$\left\| T^{-1}\sum_{t=1}^{T}(\mathbf{F}_{t|T}^{(0)}-\mathbf{F}_t)\mathbf{F}_t'\right\| = O_p(\max(n^{-1}, T^{-1/2})), \tag{B.15}$$

$$\left\| T^{-1}\sum_{t=1}^{T}(\mathbf{F}_{t|T}^{(0)}-\mathbf{F}_t)x_{it}\right\| = O_p(\max(n^{-1}, T^{-1/2})), \tag{B.16}$$

which once substituted into (B.2) and (B.3), jointly with Lemma D.12 give

$$\left\|T^{-1}\sum_{t=1}^T\{\mathbf{F}_{t|T}^{(0)}\mathbf{F}_{t|T}^{(0)\prime}+\mathbf{P}_{t|T}^{(0)}\}-T^{-1}\sum_{t=1}^T\mathbf{F}_t\mathbf{F}_t'\right\|\le\left\|T^{-1}\sum_{t=1}^T\mathbf{F}_{t|T}^{(0)}\mathbf{F}_{t|T}^{(0)\prime}-T^{-1}\sum_{t=1}^T\mathbf{F}_t\mathbf{F}_t'\right\|+\max_{t=1,\dots,T}\|\mathbf{P}_{t|T}^{(0)}\|$$
$$=O_p(\max(n^{-1},T^{-1/2}))+O_p(n^{-1}),$$

and

$$\left\|T^{-1}\sum_{t=1}^T\mathbf{F}_{t|T}^{(0)}x_{it}-T^{-1}\sum_{t=1}^T\mathbf{F}_t x_{it}\right\|=O_p(\max(n^{-1},T^{-1/2})).$$

Therefore, from (B.1)

$$\|\widehat{\boldsymbol{\lambda}}_i^{(1)}-\boldsymbol{\lambda}_i^{\mathrm{OLS}}\|=O_p(\max(n^{-1},T^{-1/2})),\tag{B.17}$$

with $\boldsymbol{\lambda}_i^{\mathrm{OLS}}=(T^{-1}\sum_{t=1}^T\mathbf{F}_t\mathbf{F}_t')^{-1}(T^{-1}\sum_{t=1}^T\mathbf{F}_t x_{it})$. And by Lemma D.19(i) we also have

$$\|\boldsymbol{\lambda}_i^{\mathrm{OLS}}-\boldsymbol{\lambda}_i\|=O_p(T^{-1/2}),\tag{B.18}$$

indeed, recalling that $\boldsymbol{\Gamma}^F=\mathbf{I}_r$ by Assumption 6(b), by Lemma C.12(i) and Weyl's inequality (Merikoski and Kumar, 2004, Theorem 1) we have $|\nu^{(r)}(T^{-1}\sum_{t=1}^T\mathbf{F}_t\mathbf{F}_t')^{-1}|=O_p(T^{-1/2})$ which implies $\|(T^{-1}\sum_{t=1}^T\mathbf{F}_t\mathbf{F}_t')^{-1}\|=O_p(1)$. From (B.17) and (B.18)

$$\|\widehat{\boldsymbol{\lambda}}_i^{(1)}-\boldsymbol{\lambda}_i\|\le\|\widehat{\boldsymbol{\lambda}}_i^{(1)}-\boldsymbol{\lambda}_i^{\mathrm{OLS}}\|+\|\boldsymbol{\lambda}_i^{\mathrm{OLS}}-\boldsymbol{\lambda}_i\|=O_p(\max(n^{-1},T^{-1/2})).\tag{B.19}$$

Moreover, by letting $\boldsymbol{y}_t=n^{-1/2}\mathbf{x}_{nt}$ the above proof leads to

$$\left\|n^{-1/2}T^{-1}\sum_{t=1}^T(\mathbf{F}_{t|T}^{(0)}-\mathbf{F}_t)\mathbf{x}_{nt}'\right\|=O_p(\max(n^{-1},T^{-1/2})),\tag{B.20}$$

and, therefore, using also Lemma D.19(ii), we have

$$n^{-1/2}\|\widehat{\boldsymbol{\Lambda}}_n^{(1)}-\boldsymbol{\Lambda}_n\|=O_p(\max(n^{-1},T^{-1/2})).\tag{B.21}$$

This proves parts (a.1) and (a.2) when $k^*=0$.

Following the same reasoning leading to (B.19) and by Lemma D.19(iv) we can easily prove that

$$\|\widehat{\mathbf{A}}^{(1)}-\mathbf{A}\|\le\|\widehat{\mathbf{A}}^{(1)}-\mathbf{A}^{\mathrm{OLS}}\|+\|\mathbf{A}^{\mathrm{OLS}}-\mathbf{A}\|=O_p(\max(n^{-1},T^{-1/2})),\tag{B.22}$$

where $\widehat{\mathbf{A}}^{(1)}$ is defined in (15) and $\mathbf{A}^{\mathrm{OLS}}=(T^{-1}\sum_{t=1}^T\mathbf{F}_t\mathbf{F}_{t-1}')(T^{-1}\sum_{t=1}^T\mathbf{F}_{t-1}\mathbf{F}_{t-1}')^{-1}$ (recall that $\mathbf{F}_0=\mathbf{0}_r$ by Assumption 1(i)). To prove (B.22) we also use the fact that $\max_{t=1,\dots,T}\|\mathbf{C}_{t,t-1|T}^{(0)}\|=O(n^{-1})$ since it can be obtained by the upper right block of $\mathbf{P}_{t|T}^{(0)}$ when this is computed from the the Kalman smoother having the augmented state vector $(\mathbf{F}_t'\,\mathbf{F}_{t-1}')'$. This proves part (a.4) when $k^*=0$.

Likewise, using (B.22) and the same reasoning leading to (B.19), by Lemma D.19(v), we can easily prove also that

$$\|\widehat{\boldsymbol{\Gamma}}^{v(1)}-\boldsymbol{\Gamma}^v\|\le\|\widehat{\boldsymbol{\Gamma}}^{v(1)}-\boldsymbol{\Gamma}^{v\mathrm{OLS}}\|+\|\boldsymbol{\Gamma}^{v\mathrm{OLS}}-\boldsymbol{\Gamma}^v\|=O_p(\max(n^{-1},T^{-1/2})).\tag{B.23}$$

where $\widehat{\boldsymbol{\Gamma}}^{v(1)}$ is defined in (16) and $\boldsymbol{\Gamma}^{v\mathrm{OLS}}=T^{-1}\sum_{t=1}^T(\mathbf{F}_t-\mathbf{A}^{\mathrm{OLS}}\mathbf{F}_{t-1})(\mathbf{F}_t-\mathbf{A}^{\mathrm{OLS}}\mathbf{F}_{t-1})'$. To prove (B.23) we need to use also the intermediate quantity $T^{-1}\sum_{t=1}^T(\mathbf{F}_t-\widehat{\mathbf{A}}^{(1)}\mathbf{F}_{t-1})(\mathbf{F}_t-\widehat{\mathbf{A}}^{(1)}\mathbf{F}_{t-1})'$, which is $\min(n,\sqrt{T})$-consistent because of (B.22). This proves part (a.5) when $k^*=0$.

Finally, using again the same reasoning leading to (B.19), by Lemma D.19(iii), we can prove that

$$|\widehat{\sigma}_i^{2(1)}-\sigma_i^2|\le|\widehat{\sigma}_i^{2(1)}-\sigma_i^{2\mathrm{OLS}}|+|\sigma_i^{2\mathrm{OLS}}-\sigma_i^2|=O_p(\max(n^{-1},T^{-1/2})),\tag{B.24}$$

where $\widehat{\sigma}_i^{2(1)}$ is defined in (14) and $\sigma_i^{2\mathrm{OLS}}=T^{-1}\sum_{t=1}^T(x_{it}-\boldsymbol{\lambda}_i^{\mathrm{OLS}\prime}\mathbf{F}_t)^2$. To prove (B.24) we need to use also the intermediate quantity $T^{-1}\sum_{t=1}^T(x_{it}-\widehat{\boldsymbol{\lambda}}_i^{(1)\prime}\mathbf{F}_t)^2$, which is $\min(n,\sqrt{T})$-consistent because of (B.19). This proves part (a.3) when $k^*=0$.

Now, from (B.19) and (B.24) using the same reasoning of the proof of Lemma D.4(ii), which in turn requires (B.15) and (B.16), again it follows that

$$n^{-1}\left|\sum_{i=1}^{n}(\widehat{\sigma}_i^{(1)2}-\sigma_i^2)\right| = O_p(\max(n^{-1},T^{-1/2})). \tag{B.25}$$

And, using the same reasoning as in the proof of Lemma D.5 but using now (B.21), (B.24), and (B.25),

$$\begin{aligned}
&n^{-1}\|\widehat{\mathbf{\Lambda}}_n^{(1)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(1)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(1)} - \mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^{\xi})^{-1}\mathbf{\Lambda}_n\| = O_p(\max(n^{-1},T^{-1/2})),\\
&n^{-1/2}\|\widehat{\mathbf{\Lambda}}_n^{(1)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(1)})^{-1} - \mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^{\xi})^{-1}\| = O_p(\max(n^{-1},T^{-1/2})),\\
&n\|(\widehat{\mathbf{\Lambda}}_n^{(1)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(1)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(1)})^{-1}\| = O_p(1)\\
&n\|(\widehat{\mathbf{\Lambda}}_n^{(1)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(1)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(1)})^{-1} - (\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^{\xi})^{-1}\mathbf{\Lambda}_n)^{-1}\| = O_p(\max(n^{-1},T^{-1/2})).
\end{aligned} \tag{B.26}$$

From (B.21), (B.22), (B.23), and the relations in (B.26) and following the same steps as in the proofs of Lemmas D.8, D.11, and D.12, we get

$$\begin{aligned}
&\max_{t=1,\dots,T}\|\mathbf{P}_{t|t-1}^{(1)}\| = O_p(1), \quad \max_{t=1,\dots,T}\|(\mathbf{P}_{t|t-1}^{(1)})^{-1}\| = O_p(1),\\
&\max_{t=1,\dots,T}\|\mathbf{P}_{t|t}^{(1)}\| = O_p(n^{-1}), \quad \max_{t=1,\dots,T}\|\mathbf{P}_{t|T}^{(1)}\| = O_p(n^{-1}).
\end{aligned} \tag{B.27}$$

and also $\|\mathbf{F}_{t|t}^{(1)}\| = O_p(1)$ and $\|\mathbf{F}_{t|T}^{(1)}\| = O_p(1)$ by the same arguments in Lemma D.14. It follows that we can apply the same steps as in the proofs of Lemmas D.15, D.16, and D.17 to get

$$\|\mathbf{F}_{t|T}^{(1)} - \mathbf{F}_t\| = O_p(\max(n^{-1/2},T^{-1/2})).$$

This proves part (b) when $k^*=0$.

Then we can show that (B.15) and (B.16) still hold when using $\mathbf{F}_{t|T}^{(1)}$ in place of $\mathbf{F}_{t|T}^{(0)}$ and using also the last of (B.27) we prove $\min(n,\sqrt{T})$-consistency of $\widehat{\boldsymbol{\lambda}}_i^{(2)}$. Similarly we can prove $\min(n,\sqrt{T})$-consistency of $n^{-1/2}\widehat{\mathbf{\Lambda}}_n^{(2)}$, $\widehat{\mathbf{A}}^{(2)}$, $\widehat{\mathbf{\Gamma}}^{v(2)}$, and $\sigma_i^{2(2)}$. It is then clear that we can repeat the same reasoning leading to (B.25), (B.26), and (B.27) but when $k^*=1$. So these arguments hold for all $k^*\ge 0$. This completes the proof. □

B.2 Proof of Proposition 2

For part (a.1), for any $k^*\ge 0$, we have (recall that $\widehat{\boldsymbol{\lambda}}_i \equiv \widehat{\boldsymbol{\lambda}}_i^{(k^*+1)}$)

$$\begin{aligned}
(\widehat{\boldsymbol{\lambda}}_i - \boldsymbol{\lambda}_i) &= (\widehat{\boldsymbol{\lambda}}_i - \widehat{\boldsymbol{\lambda}}_i^{**}) + (\widehat{\boldsymbol{\lambda}}_i^{**} - \widehat{\boldsymbol{\lambda}}_i^{*}) + (\widehat{\boldsymbol{\lambda}}_i^{*} - \boldsymbol{\lambda}_i^{\mathrm{OLS}}) + (\boldsymbol{\lambda}_i^{\mathrm{OLS}} - \boldsymbol{\lambda}_i)\\
&= L.1 + L.2 + L.3 + L.4, \quad \text{say.}
\end{aligned} \tag{B.28}$$

From Lemma E.23

$$\|L.1\| = O_p(\max(n^{-2}\log^{4/\delta_v}T, n^{-1}T^{-1}\log^{1/\delta_v}T\sqrt{\log n}, T^{-3/2}\sqrt{\log n})). \tag{B.29}$$

From Lemma E.22(i)

$$\|L.2\| = O_p(\max(n^{-1}\log^{2/\delta_v}T, n^{-1/2}T^{-1/2}\sqrt{\log n}, T^{-1})). \tag{B.30}$$

From Lemma E.11(i)

$$\|L.3\| = O_p(\max(n^{-1}, n^{-1/2}T^{-1/2}, T^{-1})). \tag{B.31}$$

From Lemma D.19(i)

$$\|L.4\| = O_p(T^{-1/2}). \tag{B.32}$$

By using (B.29), (B.30), (B.31), and (B.32) into (B.28), we prove part (a.1).

For part (a.2), for any $k^* \geq 0$, we have (recall that $\widehat{\mathbf{\Lambda}}_n \equiv \widehat{\mathbf{\Lambda}}_n^{(k^*+1)}$)

$$(\widehat{\mathbf{\Lambda}}_n - \mathbf{\Lambda}_n) = (\widehat{\mathbf{\Lambda}}_n - \widehat{\mathbf{\Lambda}}_n^{**}) + (\widehat{\mathbf{\Lambda}}_n^{**} - \widehat{\mathbf{\Lambda}}_n^{*}) + (\widehat{\mathbf{\Lambda}}_n^{*} - \mathbf{\Lambda}_n^{\text{OLS}}) + (\mathbf{\Lambda}_n^{\text{OLS}} - \mathbf{\Lambda}_n),$$

and the proof follows from Lemmas D.19(ii), E.11(ii), and E.25(i), and since

$$\begin{aligned} n^{-1}\|\widehat{\mathbf{\Lambda}}_n^{**} - \widehat{\mathbf{\Lambda}}_n^{*}\|^2 &= n^{-1}\sum_{i=1}^{n}\|\widehat{\boldsymbol{\lambda}}_i^{**} - \widehat{\boldsymbol{\lambda}}_i^{*}\|^2 \leq \max_{i=1,\dots,n}\|\widehat{\boldsymbol{\lambda}}_i^{**} - \widehat{\boldsymbol{\lambda}}_i^{*}\|^2 \\ &= O_p(\max(n^{-2}\log^{4/\delta_v} T, n^{-1}T^{-1}\log n, T^{-2})), \end{aligned} \tag{B.33}$$

by Lemma E.22(i).

For part (b), from part (a.1) and (B.28), if $n^{-1}\sqrt{T}\log^{2/\delta_v} T \to 0$, as $n, T \to \infty$, we have

$$\begin{aligned} \sqrt{T}(\widehat{\boldsymbol{\lambda}}_i - \boldsymbol{\lambda}_i) &= \sqrt{T}(\boldsymbol{\lambda}_i^{\text{OLS}} - \boldsymbol{\lambda}_i) + o_p(1) \\ &= \left(T^{-1}\sum_{t=1}^{T}\mathbf{F}_t\mathbf{F}_t'\right)^{-1}\left(T^{-1/2}\sum_{t=1}^{T}\mathbf{F}_t\xi_{it}\right) + o_p(1). \end{aligned} \tag{B.34}$$

Now, since $\{\mathbf{F}_t\xi_{it}\}$ is strongly mixing with exponentially decaying coefficients by Bradley (2005, Theorem 5.1.a) (see also (E.5) in the proof of Lemma E.1), and given that by Assumption 5 the following Cramér condition holds

$$\sup_{m\geq 1} r^{-1/\delta}(\mathbb{E}[|\xi_{it}F_{jt}|^m])^{1/m} \leq K,$$

for some finite positive reals $\delta \in \left(0, \frac{\delta_v\delta_\xi}{\delta_v+\delta_\xi}\right)$ and K independent of t, i, and j (Kuchibhotla and Chakrabortty, 2022, Section 2), then the Central Limit Theorem by Ibragimov (1962, Theorem 1.7) applies, i.e.,

$$T^{-1/2}\sum_{t=1}^{T}\mathbf{F}_t\xi_{it} \xrightarrow{d} \mathcal{N}\left(\mathbf{0}_r, \lim_{T\to\infty} T^{-1}\sum_{s,t=1}^{T}\mathbb{E}\left[\mathbf{F}_t\mathbf{F}_s'\xi_{it}\xi_{is}\right]\right). \tag{B.35}$$

Therefore, by Lemmas C.12(i) and C.13, and Assumption 1(b)

$$\begin{aligned} \left\|\left(T^{-1}\sum_{t=1}^{T}\mathbf{F}_t\mathbf{F}_t'\right)^{-1} - (\mathbf{\Gamma}^F)^{-1}\right\| &\leq \left\|\left(T^{-1}\sum_{t=1}^{T}\mathbf{F}_t\mathbf{F}_t'\right)^{-1}\right\| \left\|T^{-1}\sum_{t=1}^{T}\mathbf{F}_t\mathbf{F}_t' - \mathbf{\Gamma}^F\right\| \|(\mathbf{\Gamma}^F)^{-1}\| \\ &= O_p(T^{-1/2}). \end{aligned} \tag{B.36}$$

Thus from (B.34), (B.35), and (B.36), by Slutsky's Theorem, we have

$$\sqrt{T}(\widehat{\boldsymbol{\lambda}}_i - \boldsymbol{\lambda}_i) \xrightarrow{d} \mathcal{N}(\mathbf{0}_r, \boldsymbol{\mathcal{V}}_i),$$

where,

$$\boldsymbol{\mathcal{V}}_i = (\mathbf{\Gamma}^F)^{-1}\left\{\lim_{T\to\infty} T^{-1}\sum_{s,t=1}^{T}\mathbb{E}\left[\mathbf{F}_t\xi_{it}\xi_{is}\mathbf{F}_s'\right]\right\}(\mathbf{\Gamma}^F)^{-1} = \left\{T^{-1}\sum_{s,t=1}^{T}\mathbb{E}[\xi_{it}\xi_{is}]\,\mathbb{E}\left[\mathbf{F}_t\mathbf{F}_s'\right]\right\},$$

since $\mathbf{\Gamma}^F = \mathbf{I}_r$ because of Assumption 6(b) and $\{\mathbf{F}_t\}$ and $\{\xi_{it}\}$ are independent processes because of Lemma C.11. This proves part (b). Part (c) is straightforward. This completes the proof. □

B.3 Proof of Proposition 3

Recall the definitions $\widehat{\mathbf{F}}_t \equiv \widehat{\mathbf{F}}_{t|T} \equiv \mathbf{F}_{t|T}^{(k^*+1)}$, for any $k^* \geq 0$. From Lemmas F.4(ii) and F.4(iii),

$$\begin{aligned} \|\widehat{\mathbf{F}}_{t|T} - \mathbf{F}_t\| &\leq \|\widehat{\mathbf{F}}_{t|T} - \widehat{\mathbf{F}}_{t|t}\| + \|\widehat{\mathbf{F}}_{t|t} - \widehat{\mathbf{F}}_t^{\text{WLS}}\| + \|\widehat{\mathbf{F}}_t^{\text{WLS}} - \mathbf{F}_t\| \\ &= \|\widehat{\mathbf{F}}_t^{\text{WLS}} - \mathbf{F}_t\| + O_p(n^{-1}). \end{aligned} \tag{B.37}$$

where $\widehat{\mathbf{F}}_t^{\text{WLS}} = (\widehat{\mathbf{\Lambda}}_n'(\widehat{\mathbf{\Sigma}}_n^{\xi})^{-1}\widehat{\mathbf{\Lambda}}_n)^{-1}\widehat{\mathbf{\Lambda}}_n'(\widehat{\mathbf{\Sigma}}_n^{\xi})^{-1}\mathbf{x}_{nt}$.

Now,

$$\begin{aligned}\|\widehat{\mathbf{F}}_t^{\mathrm{WLS}}-\mathbf{F}_t\| \le& \|(\widehat{\boldsymbol\Lambda}_n'(\widehat{\boldsymbol\Sigma}_n^\xi)^{-1}\widehat{\boldsymbol\Lambda}_n)^{-1}\widehat{\boldsymbol\Lambda}_n'(\widehat{\boldsymbol\Sigma}_n^\xi)^{-1}(\boldsymbol\Lambda_n-\widehat{\boldsymbol\Lambda}_n)\|\,\|\mathbf{F}_t\|\\ &+\|(\widehat{\boldsymbol\Lambda}_n'(\widehat{\boldsymbol\Sigma}_n^\xi)^{-1}\widehat{\boldsymbol\Lambda}_n)^{-1}\widehat{\boldsymbol\Lambda}_n'(\widehat{\boldsymbol\Sigma}_n^\xi)^{-1}\boldsymbol\xi_{nt}\|\\ \le& \|(\boldsymbol\Lambda_n'(\boldsymbol\Sigma_n^\xi)^{-1}\boldsymbol\Lambda_n)^{-1}\boldsymbol\Lambda_n'(\boldsymbol\Sigma_n^\xi)^{-1}(\boldsymbol\Lambda_n-\widehat{\boldsymbol\Lambda}_n)\|\,\|\mathbf{F}_t\|\\ &+\|(\widehat{\boldsymbol\Lambda}_n'(\widehat{\boldsymbol\Sigma}_n^\xi)^{-1}\widehat{\boldsymbol\Lambda}_n)^{-1}\widehat{\boldsymbol\Lambda}_n'(\widehat{\boldsymbol\Sigma}_n^\xi)^{-1}-(\boldsymbol\Lambda_n'(\boldsymbol\Sigma_n^\xi)^{-1}\boldsymbol\Lambda_n)^{-1}\boldsymbol\Lambda_n'(\boldsymbol\Sigma_n^\xi)^{-1}\|\\ &\cdot\|\boldsymbol\Lambda_n-\widehat{\boldsymbol\Lambda}_n\|\,\|\mathbf{F}_t\|+\|(\boldsymbol\Lambda_n'(\boldsymbol\Sigma_n^\xi)^{-1}\boldsymbol\Lambda_n)^{-1}\boldsymbol\Lambda_n'(\boldsymbol\Sigma_n^\xi)^{-1}\boldsymbol\xi_{nt}\|\\ &+\|(\widehat{\boldsymbol\Lambda}_n'(\widehat{\boldsymbol\Sigma}_n^\xi)^{-1}\widehat{\boldsymbol\Lambda}_n)^{-1}\widehat{\boldsymbol\Lambda}_n'(\widehat{\boldsymbol\Sigma}_n^\xi)^{-1}\boldsymbol\xi_{nt}-(\boldsymbol\Lambda_n'(\boldsymbol\Sigma_n^\xi)^{-1}\boldsymbol\Lambda_n)^{-1}\boldsymbol\Lambda_n'(\boldsymbol\Sigma_n^\xi)^{-1}\boldsymbol\xi_{nt}\|\\ =& \mathrm{A}+\mathrm{B}+\mathrm{C}+\mathrm{D},\quad\text{say.}\end{aligned}\tag{B.38}$$

Let us consider each term in (B.38). First, consider term A and notice that

$$\begin{aligned}n^{-1/2}\|\widehat{\boldsymbol\Lambda}_n-\boldsymbol\Lambda_n^{\mathrm{OLS}}\|\le& n^{-1/2}\|\widehat{\boldsymbol\Lambda}_n-\widehat{\boldsymbol\Lambda}_n^{**}\|+n^{-1/2}\|\widehat{\boldsymbol\Lambda}_n^{**}-\widehat{\boldsymbol\Lambda}_n^{*}\|+n^{-1/2}\|\widehat{\boldsymbol\Lambda}_n^{*}-\boldsymbol\Lambda_n^{\mathrm{OLS}}\|\\ =& O_p(\max(n^{-2}\log^{4/\delta_v}T,n^{-1}T^{-1}\log^{1/\delta_v}T\sqrt{\log n},T^{-3/2}\log n))\\ &+O_p(\max(n^{-1}\log^{2/\delta_v}T,n^{-1/2}T^{-1/2}\sqrt{\log n},T^{-1}))\\ &+O_p(\max(n^{-1}\log^{2/\delta_v}T,n^{-1/2}T^{-1/2},T^{-1}))\\ =& O_p(\max(n^{-1}\log^{2/\delta_v}T,n^{-1/2}T^{-1/2}\sqrt{\log n},T^{-1})),\end{aligned}\tag{B.39}$$

by Lemmas E.11(ii), E.22(i), and E.25(i) (see also (B.33) in the proof of Proposition 2).

Therefore, from (B.39)

$$\begin{aligned}\mathrm{A}\le& \|(\boldsymbol\Lambda_n'(\boldsymbol\Sigma_n^\xi)^{-1}\boldsymbol\Lambda_n)^{-1}\boldsymbol\Lambda_n'(\boldsymbol\Sigma_n^\xi)^{-1}(\boldsymbol\Lambda_n-\boldsymbol\Lambda_n^{\mathrm{OLS}})\|\,\|\mathbf{F}_t\|\\ &+n\|(\boldsymbol\Lambda_n'(\boldsymbol\Sigma_n^\xi)^{-1}\boldsymbol\Lambda_n)^{-1}\|\,n^{-1/2}\|\boldsymbol\Lambda_n\|\,\|(\boldsymbol\Sigma_n^\xi)^{-1}\|\,n^{-1/2}\|\widehat{\boldsymbol\Lambda}_n-\boldsymbol\Lambda_n^{\mathrm{OLS}}\|\,\|\mathbf{F}_t\|\\ =&\{\mathrm{A.1}+\mathrm{A.2}\}\|\mathbf{F}_t\|,\quad\text{say.}\end{aligned}\tag{B.40}$$

Then,

$$\begin{aligned}\mathrm{A.1}\le& n\|(\boldsymbol\Lambda_n'(\boldsymbol\Sigma_n^\xi)^{-1}\boldsymbol\Lambda_n)^{-1}\|\,n^{-1}\|\boldsymbol\Lambda_n'(\boldsymbol\Sigma_n^\xi)^{-1}(\boldsymbol\Lambda_n-\boldsymbol\Lambda_n^{\mathrm{OLS}})\|\\ =& n\|(\boldsymbol\Lambda_n'(\boldsymbol\Sigma_n^\xi)^{-1}\boldsymbol\Lambda_n)^{-1}\|\,n^{-1}\left\|T^{-1}\sum_{t=1}^T\boldsymbol\Lambda_n'(\boldsymbol\Sigma_n^\xi)^{-1}\boldsymbol\xi_{nt}\mathbf{F}_t'\right\|\left\|\left(T^{-1}\sum_{t=1}^T\mathbf{F}_t\mathbf{F}_t'\right)^{-1}\right\|\\ =& O_p(n^{-1/2}T^{-1/2}),\end{aligned}\tag{B.41}$$

by Lemmas C.3(iii), C.8(iv), and C.13. Moreover, $\mathrm{A.2}=O_p(\max(n^{-1}\log^{2/\delta_v}T,n^{-1/2}T^{-1/2}\sqrt{\log n},T^{-1}))$, because of (B.39) and Lemmas C.2, C.3(iii), and Assumption 2(a) which implies $\|(\boldsymbol\Sigma_n^\xi)^{-1}\|\le C_\xi$. This, jointly with (B.40) and (B.41) implies that

$$\mathrm{A}=O_p(\max(n^{-1}\log^{2/\delta_v}T,n^{-1/2}T^{-1/2}\sqrt{\log n},T^{-1})),\tag{B.42}$$

since $\|\mathbf{F}_t\|=O_p(1)$ because $\mathbb{E}[F_{jt}^2]=1$, $j=1,\dots,r$, by Assumption 6(b).

Second, by Proposition 2(a) and Lemma F.2(v)

$$\begin{aligned}\mathrm{B}=&\|n(\widehat{\boldsymbol\Lambda}_n'(\widehat{\boldsymbol\Sigma}_n^\xi)^{-1}\widehat{\boldsymbol\Lambda}_n)^{-1}n^{-1/2}\widehat{\boldsymbol\Lambda}_n'(\widehat{\boldsymbol\Sigma}_n^\xi)^{-1}-n(\boldsymbol\Lambda_n'(\boldsymbol\Sigma_n^\xi)^{-1}\boldsymbol\Lambda_n)^{-1}n^{-1/2}\boldsymbol\Lambda_n'(\boldsymbol\Sigma_n^\xi)^{-1}\|\,n^{-1/2}\|\boldsymbol\Lambda_n-\widehat{\boldsymbol\Lambda}_n\|\,\|\mathbf{F}_t\|\\ =&O_p(\max(n^{-2}\log^{4/\delta_v}T,T^{-1}\sqrt{\log n},n^{-1}T^{-1/2}\log^{2/\delta_v}T\sqrt{\log n})),\end{aligned}\tag{B.43}$$

and again since $\|\mathbf{F}_t\|=O_p(1)$ because $\mathbb{E}[F_{jt}^2]=1$, $j=1,\dots,r$, by Assumption 6(b).

Third,

$$\mathrm{C}\le n\|(\boldsymbol\Lambda_n'(\boldsymbol\Sigma_n^\xi)^{-1}\boldsymbol\Lambda_n)^{-1}\|\,n^{-1}\|\boldsymbol\Lambda_n'(\boldsymbol\Sigma_n^\xi)^{-1}\boldsymbol\xi_{nt}\|=O_p(n^{-1/2}),\tag{B.44}$$

by Lemmas C.3(iii) and C.7(i).

Fourth, and last,

$$\begin{aligned}\mathrm{D} \leq & n\|(\widehat{\mathbf{\Lambda}}_n'(\widehat{\mathbf{\Sigma}}_n^\xi)^{-1}\widehat{\mathbf{\Lambda}}_n)^{-1} - (\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n)^{-1}\|\, n^{-1}\|\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\boldsymbol{\xi}_{nt}\| \\ & + n\|(\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n)^{-1}\|\, n^{-1}\|\{\widehat{\mathbf{\Lambda}}_n'(\widehat{\mathbf{\Sigma}}_n^\xi)^{-1} - \mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\}\boldsymbol{\xi}_{nt}\| \\ & + n\|(\widehat{\mathbf{\Lambda}}_n'(\widehat{\mathbf{\Sigma}}_n^\xi)^{-1}\widehat{\mathbf{\Lambda}}_n)^{-1} - (\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n)^{-1}\|\, n^{-1}\|\{\widehat{\mathbf{\Lambda}}_n'(\widehat{\mathbf{\Sigma}}_n^\xi)^{-1} - \mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\}\boldsymbol{\xi}_{nt}\| \\ = & \,\mathrm{D.1} + \mathrm{D.2} + \mathrm{D.3}, \text{ say.}\end{aligned} \tag{B.45}$$

Then, $\mathrm{D.1} = O_p(\max(n^{-3/2}\log^{2/\delta_v} T, n^{-1/2}T^{-1/2}\sqrt{\log n}))$, by Lemmas C.7(i) and F.2(iv). Moreover,

$$\begin{aligned}\mathrm{D.2} = & n\|(\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n)^{-1}\| \\ & \cdot\Big\{n^{-1}\|(\widehat{\mathbf{\Lambda}}_n - \mathbf{\Lambda}_n)'(\mathbf{\Sigma}_n^\xi)^{-1}\boldsymbol{\xi}_{nt}\| + n^{-1}\|\mathbf{\Lambda}_n'[(\widehat{\mathbf{\Sigma}}_n^\xi)^{-1} - (\mathbf{\Sigma}_n^\xi)^{-1}]\boldsymbol{\xi}_{nt}\| + n^{-1}\|(\widehat{\mathbf{\Lambda}}_n - \mathbf{\Lambda}_n)'[(\widehat{\mathbf{\Sigma}}_n^\xi)^{-1} - (\mathbf{\Sigma}_n^\xi)^{-1}]\boldsymbol{\xi}_{nt}\|\Big\} \\ = & n\|(\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n)^{-1}\|\,\{\mathrm{D.2.a} + \mathrm{D.2.b} + \mathrm{D.2.c}\}, \quad \text{say.}\end{aligned}$$

We then have the following results. First,

$$\begin{aligned}\mathrm{D.2.a} = & n^{-1}\left\|\sum_{i=1}^n(\widehat{\boldsymbol{\lambda}}_i - \boldsymbol{\lambda}_i)(\sigma_i^2)^{-1}\xi_{it}\right\| \leq \max_{i=1,\ldots,n}\|\widehat{\boldsymbol{\lambda}}_i - \boldsymbol{\lambda}_i\| n^{-1}\|(\mathbf{\Sigma}_n^\xi)^{-1}\boldsymbol{\xi}_{nt}\| \\ = & O_p(\max(n^{-3/2}\log^{2/\delta_v} T, n^{-1/2}T^{-1/2}\sqrt{\log n})),\end{aligned} \tag{B.46}$$

by Lemmas C.7(i) and F.1(i). Second,

$$\begin{aligned}\mathrm{D.2.b} = & n^{-1}\left\|\sum_{i=1}^n\boldsymbol{\lambda}_i\{(\widehat{\sigma}_i^2)^{-1} - (\sigma_i^2)^{-1}\}\xi_{it}\right\| = n^{-1}\left\|\sum_{i=1}^n\boldsymbol{\lambda}_i(\sigma_i^2\widehat{\sigma}_i^2)^{-1}\{\widehat{\sigma}_i^2 - \sigma_i^2\}\xi_{it}\right\| \\ \leq & \max_{i=1,\ldots,n}|\widehat{\sigma}_i^2 - \sigma_i^2|\left\{n^{-1}\left\|\sum_{i=1}^n\boldsymbol{\lambda}_i\{\sigma_i^2(\widehat{\sigma}_i^2 - \sigma_i^2)\}^{-1}\xi_{it}\right\| + n^{-1}\left\|\sum_{i=1}^n\boldsymbol{\lambda}_i(\sigma_i^4)^{-1}\xi_{it}\right\|\right\} \\ = & O_p(\max(n^{-3/2}\log^{2/\delta_v} T, n^{-1/2}T^{-1/2}\sqrt{\log n})),\end{aligned} \tag{B.47}$$

by Lemmas C.7(i) and F.1(ii). Third, clearly by (B.46) and (B.47)

$$\mathrm{D.2.c} = o_p(\max(n^{-3/2}\log^{2/\delta_v} T, n^{-1/2}T^{-1/2}\sqrt{\log n})). \tag{B.48}$$

By (B.46), (B.47), and (B.48), and Lemma C.3 we have $\mathrm{D.2} = O_p(\max(n^{-3/2}\log^{2/\delta_v} T, n^{-1/2}T^{-1/2}\sqrt{\log n})$. Last, $\mathrm{D.3} = O_p(\max(n^{-2}\log^{4/\delta_v} T, T^{-1}\log n))\mathrm{D.2}$, by Lemma F.2(ii), thus it is dominated by D.2. Therefore,

$$\mathrm{D} = O_p(\max(n^{-3/2}\log^{2/\delta_v} T, n^{-1/2}T^{-1/2}\sqrt{\log n}). \tag{B.49}$$

By substituting (B.42), (B.43), (B.44), and (B.49), into (B.38) we have

$$\|\widehat{\mathbf{F}}_t^{\mathrm{WLS}} - \mathbf{F}_t\| = O_p(\max(n^{-1/2}, T^{-1}\sqrt{\log n})), \tag{B.50}$$

which, once substituted into (B.37), proves part (a.1).

For part (a.2), let $\widehat{\boldsymbol{\mathcal{F}}}_T^{\mathrm{KF}} = (\widehat{\mathbf{F}}_{1|1}\cdots\widehat{\mathbf{F}}_{T|T})'$ and $\widehat{\boldsymbol{\mathcal{F}}}_T^{\mathrm{WLS}} = (\widehat{\mathbf{F}}_1^{\mathrm{WLS}}\cdots\widehat{\mathbf{F}}_T^{\mathrm{WLS}})'$, and recall that, by definition, $\widehat{\boldsymbol{\mathcal{F}}}_T = (\widehat{\mathbf{F}}_{1|T}\cdots\widehat{\mathbf{F}}_{T|T})' = (\widehat{\mathbf{F}}_1\cdots\widehat{\mathbf{F}}_T)'$. From (B.38), we have

$$T^{-1/2}\|\widehat{\boldsymbol{\mathcal{F}}}_T - \boldsymbol{\mathcal{F}}_T\| \leq T^{-1/2}\|\widehat{\boldsymbol{\mathcal{F}}}_T - \widehat{\boldsymbol{\mathcal{F}}}_T^{\mathrm{KF}}\| + T^{-1/2}\|\widehat{\boldsymbol{\mathcal{F}}}_T^{\mathrm{KF}} - \widehat{\boldsymbol{\mathcal{F}}}_T^{\mathrm{WLS}}\| + T^{-1/2}\|\widehat{\boldsymbol{\mathcal{F}}}_T^{\mathrm{WLS}} - \boldsymbol{\mathcal{F}}_T\|. \tag{B.51}$$

By Lemma F.5 the first two terms on the rhs of (B.51) are such that

$$\begin{aligned}T^{-1/2}\|\widehat{\boldsymbol{\mathcal{F}}}_T - \widehat{\boldsymbol{\mathcal{F}}}_T^{\mathrm{KF}}\| \leq \max_{t=1,\ldots,T}\|\widehat{\mathbf{F}}_t - \widehat{\mathbf{F}}_{t|t}\| = O_p(n^{-1}\log^{1/\delta_v} T), \\ T^{-1/2}\|\widehat{\boldsymbol{\mathcal{F}}}_T^{\mathrm{KF}} - \widehat{\boldsymbol{\mathcal{F}}}_T^{\mathrm{WLS}}\| \leq \max_{t=1,\ldots,T}\|\widehat{\mathbf{F}}_{t|t} - \widehat{\mathbf{F}}_t^{\mathrm{WLS}}\| = O_p(n^{-1}\log^{1/\delta_v} T).\end{aligned} \tag{B.52}$$

While, for the last term on the rhs of (B.51), letting $\boldsymbol{\mathcal{E}}_{nT} = (\boldsymbol{\xi}_{n1} \cdots \boldsymbol{\xi}_{nT})'$, we have

$$\begin{aligned} T^{-1/2}\|\widehat{\boldsymbol{\mathcal{F}}}_T^{\text{WLS}} - \boldsymbol{\mathcal{F}}_T\| \leq & \, n\|(\widehat{\boldsymbol{\Lambda}}'(\widehat{\boldsymbol{\Sigma}}_n^\xi)^{-1}\widehat{\boldsymbol{\Lambda}}_n)^{-1}\| \\ & \cdot n^{-1}T^{-1/2}\Big\{\|\widehat{\boldsymbol{\Lambda}}'(\widehat{\boldsymbol{\Sigma}}_n^\xi)^{-1}(\boldsymbol{\Lambda}_n - \widehat{\boldsymbol{\Lambda}}_n)\boldsymbol{\mathcal{F}}_T'\| + \|\{\widehat{\boldsymbol{\Lambda}}'(\widehat{\boldsymbol{\Sigma}}_n^\xi)^{-1} - \boldsymbol{\Lambda}'(\boldsymbol{\Sigma}_n^\xi)^{-1}\}\boldsymbol{\mathcal{E}}_{nT}'\| + \|\boldsymbol{\Lambda}'(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\mathcal{E}}_{nT}'\|\Big\} \\ = & \, n\|(\widehat{\boldsymbol{\Lambda}}'(\widehat{\boldsymbol{\Sigma}}_n^\xi)^{-1}\widehat{\boldsymbol{\Lambda}}_n)^{-1}\|\,\{\mathcal{A} + \mathcal{B} + \mathcal{C}\}\,, \text{ say.} \end{aligned} \tag{B.53}$$

For the first term on the rhs of (B.53) we have

$$\begin{aligned} \mathcal{A} \leq & \, n^{-1}\|\boldsymbol{\Lambda}'(\boldsymbol{\Sigma}_n^\xi)^{-1}(\boldsymbol{\Lambda}_n - \widehat{\boldsymbol{\Lambda}}_n)\|\,T^{-1/2}\|\boldsymbol{\mathcal{F}}_T\| + n^{-1/2}\|\widehat{\boldsymbol{\Lambda}}'(\widehat{\boldsymbol{\Sigma}}_n^\xi)^{-1} - \boldsymbol{\Lambda}'(\boldsymbol{\Sigma}_n^\xi)^{-1}\|\,n^{-1/2}\|\widehat{\boldsymbol{\Lambda}}_n - \boldsymbol{\Lambda}_n\|\,T^{-1/2}\|\boldsymbol{\mathcal{F}}_T\| \\ = & \, \{\mathcal{A}_1 + \mathcal{A}_2\}T^{-1/2}\|\boldsymbol{\mathcal{F}}_T\|, \text{ say,} \end{aligned} \tag{B.54}$$

where

$$\begin{aligned} \mathcal{A}_1 \leq & \, n^{-1}\|\boldsymbol{\Lambda}'(\boldsymbol{\Sigma}_n^\xi)^{-1}(\boldsymbol{\Lambda}_n - \boldsymbol{\Lambda}_n^{\text{OLS}})\| + n^{-1/2}\|\boldsymbol{\Lambda}\|\,\|(\boldsymbol{\Sigma}_n^\xi)^{-1}\|\,n^{-1/2}\|\boldsymbol{\Lambda}_n^{\text{OLS}} - \widehat{\boldsymbol{\Lambda}}_n\| \\ = & \, O_p(n^{-1/2}T^{-1/2}) + O_p(\max(n^{-1}\log^{2/\delta_v} T, n^{-1/2}T^{-1/2}\sqrt{\log n}, T^{-1})) \\ = & \, O_p(\max(n^{-1}\log^{2/\delta_v} T, n^{-1/2}T^{-1/2}\sqrt{\log n}, T^{-1})), \end{aligned} \tag{B.55}$$

by (B.39), (B.41), Lemma C.2, and Assumption 2(a), while

$$\mathcal{A}_2 = O_p(\max(T^{-1/2}\sqrt{\log n}, n^{-1}\log^{2/\delta_v} T))O_p(\max(T^{-1/2}, n^{-1}\log^{2/\delta_v} T)), \tag{B.56}$$

by Proposition 2(a) and F.2(ii). By substituting (B.55) and (B.56) into (B.54)

$$\mathcal{A} = O_p(\max(n^{-1}\log^{2/\delta_v} T, T^{-1}\sqrt{\log n}, n^{-1/2}T^{-1/2}\sqrt{\log n})). \tag{B.57}$$

Moving to the second term on the rhs of (B.53), we have

$$\begin{aligned} \mathcal{B} \leq & \, n^{-1}T^{-1/2}\|(\widehat{\boldsymbol{\Lambda}}_n - \boldsymbol{\Lambda}_n)'(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\mathcal{E}}_{nT}'\| + n^{-1}T^{-1/2}\|\boldsymbol{\Lambda}_n'\{(\widehat{\boldsymbol{\Sigma}}_n^\xi)^{-1} - (\boldsymbol{\Sigma}_n^\xi)^{-1}\}\boldsymbol{\mathcal{E}}_{nT}'\| \\ & + n^{-1}T^{-1/2}\|(\widehat{\boldsymbol{\Lambda}}_n - \boldsymbol{\Lambda}_n)'\{(\widehat{\boldsymbol{\Sigma}}_n^\xi)^{-1} - (\boldsymbol{\Sigma}_n^\xi)^{-1}\}\boldsymbol{\mathcal{E}}_{nT}'\| \\ = & \, \mathcal{B}_1 + \mathcal{B}_2 + \mathcal{B}_3, \text{ say.} \end{aligned} \tag{B.58}$$

Then, considering each term on the rhs of (B.58),

$$\begin{aligned} \mathcal{B}_1 \leq & \, n^{-1}T^{-1/2}\|(\boldsymbol{\Lambda}_n^{\text{OLS}} - \boldsymbol{\Lambda}_n)'(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\mathcal{E}}_{nT}'\| + n^{-1/2}\|\widehat{\boldsymbol{\Lambda}}_n - \boldsymbol{\Lambda}_n^{\text{OLS}}\|\,\|(\boldsymbol{\Sigma}_n^\xi)^{-1}\|n^{-1/2}T^{-1/2}\|\boldsymbol{\mathcal{E}}_{nT}'\| \\ = & \, \mathcal{B}_{1.a} + \mathcal{B}_{1.b} \text{ say,} \end{aligned}$$

where

$$\mathcal{B}_{1.a} \leq n^{-1}T^{-1/2}\left\|T^{-1}\sum_{t=1}^T \mathbf{F}_t\boldsymbol{\xi}_{nt}'(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\mathcal{E}}_{nT}'\right\|\left\|\left(T^{-1}\sum_{t=1}^T \mathbf{F}_t\mathbf{F}_t'\right)^{-1}\right\| = O_p(n^{-1/2}T^{-1/2}),$$

by Lemmas C.8(vi) and C.13. Whereas $\mathcal{B}_{1.b} = O_p(\max(n^{-1}\log^{2/\delta_v} T, n^{-1/2}T^{-1/2}\sqrt{\log n}, T^{-1}))$ by (B.39) and Lemma C.7(vi). Therefore,

$$\mathcal{B}_1 = O_p(\max(\max(n^{-1}\log^{2/\delta_v} T, n^{-1/2}T^{-1/2}\sqrt{\log n}, T^{-1})). \tag{B.59}$$

Moreover, letting $\boldsymbol{\zeta}_i = (\xi_{i1} \cdots \xi_{iT})'$,

$$\begin{aligned}
\mathcal{B}_2 =& n^{-1}T^{-1/2}\|\boldsymbol{\Lambda}_n'(\widehat{\boldsymbol{\Sigma}}_n^{\xi})^{-1}\{\boldsymbol{\Sigma}_n^{\xi} - \widehat{\boldsymbol{\Sigma}}_n^{\xi}\}(\boldsymbol{\Sigma}_n^{\xi})^{-1}\boldsymbol{\mathcal{E}}_{nT}\| \\
=& n^{-1}T^{-1/2}\left\|\sum_{i=1}^n \boldsymbol{\lambda}_i\boldsymbol{\zeta}_i'(\widehat{\sigma}_i^2\sigma_i^2)^{-1}\{\sigma_i^2 - \widehat{\sigma}_i^2\}\right\| \\
\le& C_\xi \max_{i=1,\ldots,n} |\sigma_i^2 - \widehat{\sigma}_i^2| \left\{\min_{i=1,\ldots,n} \widehat{\sigma}_i^2\right\}^{-1} n^{-1}T^{-1/2}\left\|\sum_{i=1}^n \boldsymbol{\lambda}_i\boldsymbol{\zeta}_i'\right\| \\
\le& C_\xi \max_{i=1,\ldots,n} |\sigma_i^2 - \widehat{\sigma}_i^2| \, \|(\widehat{\boldsymbol{\Sigma}}_n^{\xi})^{-1}\| n^{-1}T^{-1/2}\|\boldsymbol{\Lambda}_n'\boldsymbol{\mathcal{E}}_{nT}'\| \\
=& O_p(\max(n^{-1}\log^{2/\delta_v} T, T^{-1/2}\sqrt{\log n}))O_p(n^{-1/2}),
\end{aligned} \tag{B.60}$$

by Assumption 2(a) and Lemmas F.1(ii), F.1(ii), and Last,

$$\begin{aligned}
\mathcal{B}_3 \le& n^{-1/2}\|\widehat{\boldsymbol{\Lambda}}_n - \boldsymbol{\Lambda}_n\| \, \|(\widehat{\boldsymbol{\Sigma}}_n^{\xi})^{-1} - (\boldsymbol{\Sigma}_n^{\xi})^{-1}\| n^{-1/2}T^{-1/2}\|\boldsymbol{\mathcal{E}}_{nT}\| \\
=& O_p(\max(T^{-1/2}, n^{-1}\log^{2/\delta_v} T)) \cdot O_p(\max(T^{-1/2}\sqrt{\log n}, n^{-1}\log^{2/\delta_v} T)),
\end{aligned} \tag{B.61}$$

by Proposition 2(a) and Lemmas C.7(vi) and F.1(v). By substituting (B.59), (B.60), and (B.61) into (B.58)

$$\mathcal{B} = O_p(n^{-1}\log^{2/\delta_v} T, n^{-1/2}T^{-1/2}\sqrt{\log n}, T^{-1}). \tag{B.62}$$

Finally, for the third term on the rhs of (B.53), we have

$$\mathcal{C} = O_p(n^{-1/2}), \tag{B.63}$$

by Lemma C.7(v).

By substituting (B.57), (B.62), and (B.63) into (B.53), and since $n\|(\widehat{\boldsymbol{\Lambda}}'(\widehat{\boldsymbol{\Sigma}}_n^{\xi})^{-1}\widehat{\boldsymbol{\Lambda}}_n)^{-1}\| = O_p(1)$ by Lemma F.2(iii),

$$T^{-1/2}\|\widehat{\boldsymbol{\mathcal{F}}}_T^{\text{WLS}} - \boldsymbol{\mathcal{F}}_T\| = O_p(\max(n^{-1/2}, T^{-1}\sqrt{\log n})),$$

which, once substituted in (B.51) together with (B.52), proves part (a.2).

Turning to part (b), from (B.37), (B.38), (B.44) and (B.50), if $T^{-1}\sqrt{n\log n} \to 0$, as $n, T \to \infty$, we have

$$\begin{aligned}
\sqrt{n}(\widehat{\mathbf{F}}_t - \mathbf{F}_t) =& \sqrt{n}(\widehat{\mathbf{F}}_t^{\text{WLS}} - \mathbf{F}_t) + o_p(1) \\
=& n(\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^{\xi})^{-1}\boldsymbol{\Lambda}_n)^{-1}\left\{n^{-1/2}\sum_{i=1}^n \boldsymbol{\lambda}_i\xi_{it}(\sigma_i^2)^{-1}\right\} + o_p(1).
\end{aligned} \tag{B.64}$$

Then, by Assumption 2(e), as $n \to \infty$, it holds that

$$n^{-1/2}\sum_{i=1}^n \boldsymbol{\lambda}_i\xi_{it}(\sigma_i^2)^{-1} \xrightarrow{d} \mathcal{N}\left(\mathbf{0}_r, \lim_{n\to\infty} n^{-1}\sum_{i,j=1}^n \boldsymbol{\lambda}_i\boldsymbol{\lambda}_j'\mathbb{E}[\xi_{it}\xi_{jt}](\sigma_i^2\sigma_j^2)^{-1}\right). \tag{B.65}$$

Moreover, from Lemma C.3(iii) we have that

$$\|n^{-1}\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^{\xi})^{-1}\boldsymbol{\Lambda}_n - \boldsymbol{\Sigma}_{\Lambda\Sigma\Lambda}\| = o_p(1),$$

for some finite and positive definite $r \times r$ matrix $\boldsymbol{\Sigma}_{\Lambda\Sigma\Lambda}$, which, jointly with Lemma C.3(v), implies

$$\|n(\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^{\xi})^{-1}\boldsymbol{\Lambda}_n)^{-1} - (\boldsymbol{\Sigma}_{\Lambda\Sigma\Lambda})^{-1}\| = o_p(1). \tag{B.66}$$

From (B.64), (B.65), and (B.66), by Slutsky's Theorem, we have

$$\sqrt{n}(\widehat{\mathbf{F}}_t - \mathbf{F}_t) \xrightarrow{d} \mathcal{N}(\mathbf{0}_r, \boldsymbol{\mathcal{W}}_t),$$

where

$$\boldsymbol{\mathcal{W}}_t = (\boldsymbol{\Sigma}_{\Lambda\Sigma\Lambda})^{-1}\left\{\lim_{n\to\infty} n^{-1}\sum_{i,j=1}^{n}\boldsymbol{\lambda}_i\boldsymbol{\lambda}_j'\mathbb{E}[\xi_{it}\xi_{jt}](\sigma_i^2\sigma_j^2)^{-1}\right\}(\boldsymbol{\Sigma}_{\Lambda\Sigma\Lambda})^{-1}.$$

This proves part (b). Part (c) is straightforward. This completes the proof. □

B.4 Proof of Proposition 4

First notice that

$$\widehat{\chi}_{it}-\chi_{it} = (\widehat{\mathbf{F}}_t-\mathbf{F}_t)'\boldsymbol{\lambda}_i + \widehat{\mathbf{F}}_t'(\widehat{\boldsymbol{\lambda}}_i-\boldsymbol{\lambda}_i) = (\widehat{\mathbf{F}}_t-\mathbf{F}_t)'\boldsymbol{\lambda}_i + \mathbf{F}_t'(\widehat{\boldsymbol{\lambda}}_i-\boldsymbol{\lambda}_i) + (\widehat{\mathbf{F}}_t-\mathbf{F}_t)'(\widehat{\boldsymbol{\lambda}}_i-\boldsymbol{\lambda}_i). \tag{B.67}$$

Then,

$$\begin{aligned}|\widehat{\chi}_{it}-\chi_{it}| &\le \|\boldsymbol{\lambda}_i\|\,\|\widehat{\mathbf{F}}_t-\mathbf{F}_t\| + \|\widehat{\boldsymbol{\lambda}}_i-\boldsymbol{\lambda}_i\|\,\|\mathbf{F}_t\| + \|\widehat{\boldsymbol{\lambda}}_i-\boldsymbol{\lambda}_i\|\,\|\widehat{\mathbf{F}}_t-\mathbf{F}_t\|\\ &= O_p(\max(n^{-1/2},T^{-1}\sqrt{\log n})) + O_p(\max(T^{-1/2},n^{-1}\log^{2/\delta_v}T)) + o_p(\max(n^{-1/2},T^{-1/2}))\\ &= O_p(\max(n^{-1/2},T^{-1/2})),\end{aligned} \tag{B.68}$$

by Propositions 2(a) and 3(a), Assumption 1(a), and since $\|\mathbf{F}_t\| = O_p(1)$ because $\mathbb{E}[F_{jt}^2]=1$, $j=1,\dots,r$, by Assumption 6(b). This proves part (a).

For part (b), let us denote $\delta_{nT}=\min(\sqrt{n},\sqrt{T})$, for simplicity of notation. Consider the first term on the rhs of (B.67). Define $\boldsymbol{K}_\Lambda = n(\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n)^{-1}$. Then, from (B.64) in the proof of Proposition 3

$$\begin{aligned}\delta_{n,T}\boldsymbol{\lambda}_i'(\widehat{\mathbf{F}}_t-\mathbf{F}_t) &= \delta_{n,T}\, n^{-1}\boldsymbol{\lambda}_i'\boldsymbol{K}_\Lambda\sum_{j=1}^{n}\boldsymbol{\lambda}_j'(\sigma_j^2)^{-1}\xi_{jt} + O_p(\delta_{n,T}T^{-1}\sqrt{\log n})\\ &= \delta_{nT}\, n^{-1/2}\mathcal{A}_{it} + o_p(1),\quad \text{say},\end{aligned} \tag{B.69}$$

since $\delta_{nT}T^{-1}\le\delta_{nT}\max(n^{-1},T^{-1}) = \delta_{nT}\delta_{nT}^{-2}\to 0$. Similarly, consider the second term on the rhs of (B.67) and define $\boldsymbol{K}_F = (T^{-1}\sum_{t=1}^T\mathbf{F}_t\mathbf{F}_t')^{-1}$. Then, from (B.34) in the proof of Proposition 2

$$\begin{aligned}\delta_{n,T}\mathbf{F}_t'(\widehat{\boldsymbol{\lambda}}_i-\boldsymbol{\lambda}_i) &= \delta_{n,T}T^{-1}\mathbf{F}_t'\boldsymbol{K}_F\sum_{s=1}^{T}\mathbf{F}_s\xi_{is} + O_p(\delta_{n,T}n^{-1}\log^{2/\delta_v}T)\\ &= \delta_{nT}\,T^{-1/2}\mathcal{B}_{it} + o_p(1),\quad\text{say},\end{aligned} \tag{B.70}$$

since $\delta_{nT}n^{-1}\le\delta_{nT}\max(n^{-1},T^{-1}) = \delta_{nT}\delta_{nT}^{-2}\to 0$. From Propositions 2(b) and 3(b), as $n,T\to\infty$,

$$\mathcal{A}_{it}\xrightarrow{d}\mathcal{N}(0,\mathcal{C}_{it}^F)\quad\text{and}\quad\mathcal{B}_{it}\xrightarrow{d}\mathcal{N}(0,\mathcal{C}_{it}^\lambda), \tag{B.71}$$

where $\mathcal{C}_{it}^F = \boldsymbol{\lambda}_i'\boldsymbol{\mathcal{W}}_t\boldsymbol{\lambda}_i$ and $\mathcal{C}_{it}^\lambda = \mathbf{F}_t'\boldsymbol{\mathcal{V}}_i\mathbf{F}_t$. Moreover, $\mathcal{A}_{it}$ and $\mathcal{B}_{it}$ are asymptotically independent, since the former is a cross-sectional sum of random variables, while the latter is the sum of a given time series and under Lemmas C.1(i)-C.1(iii) are weakly serially and cross-sectionally correlated in the same sense as assumed by Bai (2003, Assumption C).

Define $a_{nT}=\delta_{nT}n^{-1/2}$ and $b_{nT}=\delta_{nT}T^{-1/2}$. Then, substituting (B.69) and (B.70) into (B.67), we obtain

$$\begin{aligned}\delta_{nT}(\widehat{\chi}_{it}-\chi_{it}) &= a_{nT}\mathcal{A}_{it}+b_{nT}\mathcal{B}_{it}+\delta_{nT}(\widehat{\mathbf{F}}_t-\mathbf{F}_t)'(\widehat{\boldsymbol{\lambda}}_i-\boldsymbol{\lambda}_i)+o_p(1)\\ &= a_{nT}\mathcal{A}_{it}+b_{nT}\mathcal{B}_{it}+o_p(\delta_{nT}\max(n^{-1/2},T^{-1/2}))+o_p(1)\\ &= a_{nT}\mathcal{A}_{it}+b_{nT}\mathcal{B}_{it}+o_p(1),\end{aligned} \tag{B.72}$$

because of (B.68). From (B.71) and (B.72), by Slutsky's theorem and following the same reasoning as in Bai (2003, proof of Theorem 3), as $n,T\to\infty$, we have

$$\delta_{nT}\left(a_{nT}^2\mathcal{C}_{it}^F+b_{nT}^2\mathcal{C}_{it}^\lambda\right)^{-1/2}(\widehat{\chi}_{it}-\chi_{it}) = (n^{-1}\mathcal{C}_{it}^F+T^{-1}\mathcal{C}_{it}^\lambda)^{-1/2}(\widehat{\chi}_{it}-\chi_{it})\xrightarrow{d}\mathcal{N}(0,1),$$

which completes the proof. □

B.5 Proof of Proposition 5

For part (a.1), for any $k^*\geq 0$, we have (recall that $\widehat{\sigma}_i^2\equiv\widehat{\sigma}_i^{2(k^*+1)}$)

$$\begin{aligned}(\widehat{\sigma}_i^2-\sigma_i^2) &= (\widehat{\sigma}_i^2-\widehat{\sigma}_i^{2**})+(\widehat{\sigma}_i^{2**}-\widehat{\sigma}_i^{2*})+(\widehat{\sigma}_i^{2*}-\sigma_i^{2\mathrm{OLS}})+(\sigma_i^{2\mathrm{OLS}}-\sigma_i^2)\\ &= O_p(\max(n^{-1}\log^{2/\delta_v}T, T^{-1/2}))\\ &\quad + O_p(\max(n^{-1}\log^{2/\delta_v}T, n^{-1/2}T^{-1/2}\sqrt{\log n}, T^{-1}))\\ &\quad + O_p(\max(n^{-1}\log^{2/\delta_v}T, n^{-1/2}T^{-1/2}, T^{-1}))+O_p(T^{-1/2}),\end{aligned}$$

by Lemmas D.19(iii), E.11(iii), E.22(ii), and E.24(i).

For part (a.2), for any $k^*\geq 0$, we have (recall that $\widehat{\boldsymbol{\Sigma}}_n^\xi\equiv\widehat{\boldsymbol{\Sigma}}_n^{\xi(k^*+1)}$)

$$(\widehat{\boldsymbol{\Sigma}}_n^\xi-\boldsymbol{\Sigma}_n^\xi)=(\widehat{\boldsymbol{\Sigma}}_n^\xi-\widehat{\boldsymbol{\Sigma}}_n^{\xi**})+(\widehat{\boldsymbol{\Sigma}}_n^{\xi**}-\widehat{\boldsymbol{\Sigma}}_n^{\xi*})+(\widehat{\boldsymbol{\Sigma}}_n^{\xi*}-\boldsymbol{\Sigma}_n^\xi),$$

and the proof follows from Lemma E.14(iii) and since

$$\begin{aligned}\|\widehat{\boldsymbol{\Sigma}}_n^{\xi**}-\widehat{\boldsymbol{\Sigma}}_n^{\xi*}\| &= \left(\max_{i=1,\dots,n}|\widehat{\sigma}_i^{2**}-\widehat{\sigma}_i^{2*}|^2\right)^{1/2}\leq\max_{i=1,\dots,n}|\widehat{\sigma}_i^{2**}-\widehat{\sigma}_i^{2*}|\\ &= O_p(\max(n^{-1}\log^{2/\delta_v}T, n^{-1/2}T^{-1/2}\sqrt{\log n}, T^{-1})),\end{aligned}$$

by Lemma E.22(i), and

$$\begin{aligned}\|\widehat{\boldsymbol{\Sigma}}_n^{\xi}-\widehat{\boldsymbol{\Sigma}}_n^{\xi**}\| &= \left(\max_{i=1,\dots,n}|\widehat{\sigma}_i^{2}-\widehat{\sigma}_i^{2**}|^2\right)^{1/2}\leq\max_{i=1,\dots,n}|\widehat{\sigma}_i^{2}-\widehat{\sigma}_i^{2**}|\\ &= O_p(\max(n^{-1}\log^{2/\delta_v}T, T^{-1/2}\sqrt{\log n})),\end{aligned}$$

by Lemma E.25(ii).

For part (a.3), for any $k^*\geq 0$, we have (recall that $\widehat{\mathbf{A}}\equiv\widehat{\mathbf{A}}^{(k^*+1)}$)

$$\begin{aligned}(\widehat{\mathbf{A}}-\mathbf{A}) &= (\widehat{\mathbf{A}}-\widehat{\mathbf{A}}^{**})+(\widehat{\mathbf{A}}^{**}-\widehat{\mathbf{A}}^{*})+(\widehat{\mathbf{A}}^{*}-\mathbf{A}^{\mathrm{OLS}})+(\mathbf{A}^{\mathrm{OLS}}-\mathbf{A})\\ &= O_p(\max(n^{-1}\log^{2/\delta_v}T, T^{-1/2}))\\ &\quad + O_p(n^{-1}\log^{2/\delta_v}T)+O_p(n^{-1}\log^{2/\delta_v}T)+O_p(T^{-1/2}),\end{aligned}\tag{B.73}$$

by Lemmas D.19(iv), E.12(i), E.22(iii), and E.24(ii).

For part (a.4), for any $k^*\geq 0$, we have (recall that $\widehat{\boldsymbol{\Gamma}}^v\equiv\widehat{\boldsymbol{\Gamma}}^{v(k^*+1)}$)

$$\begin{aligned}(\widehat{\boldsymbol{\Gamma}}^v-\boldsymbol{\Gamma}^v) &= (\widehat{\boldsymbol{\Gamma}}^v-\widehat{\boldsymbol{\Gamma}}^{v**})+(\widehat{\boldsymbol{\Gamma}}^{v**}-\widehat{\boldsymbol{\Gamma}}^{v*})+(\widehat{\boldsymbol{\Gamma}}^{v*}-\boldsymbol{\Gamma}^{v\mathrm{OLS}})+(\boldsymbol{\Gamma}^{v\mathrm{OLS}}-\boldsymbol{\Gamma}^v)\\ &= O_p(\max(n^{-1}\log^{2/\delta_v}T, T^{-1/2}))\\ &\quad + O_p(n^{-1}\log^{2/\delta_v}T)+O_p(n^{-1}\log^{2/\delta_v}T)+O_p(T^{-1/2}),\end{aligned}$$

by Lemmas D.19(v), E.12(ii), E.22(iv), and E.24(iii). This completes the proof of part (a).

To prove part (b), we need sharper rates and, to this end, we use the closed form expressions of the estimated parameters in (14), (15), and (16).

Start with part (b.1). From (14), for any $k^* \geq 0$, we have

$$
\begin{aligned}
\widehat{\sigma}_i^2 =& T^{-1}\sum_{t=1}^T (x_{it} - \widehat{\boldsymbol{\lambda}}_i'\mathbf{F}_t)^2 \\
&+ \widehat{\boldsymbol{\lambda}}_i'\Big\{2T^{-1}\sum_{t=1}^T(\mathbf{F}_{t|T}^{(k^*)} - \mathbf{F}_t)\mathbf{F}_t' + T^{-1}\sum_{t=1}^T(\mathbf{F}_{t|T}^{(k^*)} - \mathbf{F}_t)(\mathbf{F}_{t|T}^{(k^*)} - \mathbf{F}_t)' + T^{-1}\sum_{t=1}^T \mathbf{P}_{t|T}^{(k^*)}\Big\}\widehat{\boldsymbol{\lambda}}_i \\
&- 2\widehat{\boldsymbol{\lambda}}_i'T^{-1}\sum_{t=1}^T(\mathbf{F}_{t|T}^{(k^*)} - \mathbf{F}_t)x_{it} \\
=& T^{-1}\sum_{t=1}^T(x_{it} - \boldsymbol{\lambda}_i^{\text{OLS}\prime}\mathbf{F}_t)^2 + (\widehat{\boldsymbol{\lambda}}_i - \boldsymbol{\lambda}_i^{\text{OLS}})'\left\{T^{-1}\sum_{t=1}^T\mathbf{F}_t\mathbf{F}_t'\right\}(\widehat{\boldsymbol{\lambda}}_i - \boldsymbol{\lambda}_i^{\text{OLS}}) - 2(\widehat{\boldsymbol{\lambda}}_i - \boldsymbol{\lambda}_i^{\text{OLS}})'T^{-1}\sum_{t=1}^T\mathbf{F}_t(x_{it} - \boldsymbol{\lambda}_i^{\text{OLS}\prime}\mathbf{F}_t) \\
&+ \widehat{\boldsymbol{\lambda}}_i'\Big\{2T^{-1}\sum_{t=1}^T(\mathbf{F}_{t|T}^{(k^*)} - \mathbf{F}_t)\mathbf{F}_t' + T^{-1}\sum_{t=1}^T(\mathbf{F}_{t|T}^{(k^*)} - \mathbf{F}_t)(\mathbf{F}_{t|T}^{(k^*)} - \mathbf{F}_t)' + T^{-1}\sum_{t=1}^T \mathbf{P}_{t|T}^{(k^*)}\Big\}\widehat{\boldsymbol{\lambda}}_i \\
&- 2\widehat{\boldsymbol{\lambda}}_i'T^{-1}\sum_{t=1}^T(\mathbf{F}_{t|T}^{(k^*)} - \mathbf{F}_t)x_{it}. \qquad \text{(B.74)}
\end{aligned}
$$

Therefore, since by construction $T^{-1}\sum_{t=1}^T \mathbf{F}_t(x_{it} - \boldsymbol{\lambda}_i^{\text{OLS}\prime}\mathbf{F}_t) = 0$, from (B.74)

$$
\begin{aligned}
\left|\widehat{\sigma}_i^2 - T^{-1}\sum_{t=1}^T(x_{it} - \boldsymbol{\lambda}_i^{\text{OLS}\prime}\mathbf{F}_t)^2\right| \leq& \|\widehat{\boldsymbol{\lambda}}_i - \boldsymbol{\lambda}_i^{\text{OLS}}\|^2 \left\|T^{-1}\sum_{t=1}^T\mathbf{F}_t\mathbf{F}_t'\right\| \\
&+ \|\widehat{\boldsymbol{\lambda}}_i\|^2\Big\{2\left\|T^{-1}\sum_{t=1}^T(\mathbf{F}_{t|T}^{(k^*)} - \mathbf{F}_t)\mathbf{F}_t'\right\| + \left\|T^{-1}\sum_{t=1}^T\mathbf{P}_{t|T}^{(k^*)}\right\| \\
&+ \left\|T^{-1}\sum_{t=1}^T(\mathbf{F}_{t|T}^{(k^*)} - \mathbf{F}_t)(\mathbf{F}_{t|T}^{(k^*)} - \mathbf{F}_t)'\right\|\Big\} + 2\|\widehat{\boldsymbol{\lambda}}_i\|\left\|T^{-1}\sum_{t=1}^T(\mathbf{F}_{t|T}^{(k^*)} - \mathbf{F}_t)x_{it}\right\| \\
=& \|\widehat{\boldsymbol{\lambda}}_i - \boldsymbol{\lambda}_i^{\text{OLS}}\|^2\left\|T^{-1}\sum_{t=1}^T\mathbf{F}_t\mathbf{F}_t'\right\| + \|\widehat{\boldsymbol{\lambda}}_i\|^2\{2\mathcal{A} + \mathcal{B} + \mathcal{C}\} + 2\|\widehat{\boldsymbol{\lambda}}_i\|\mathcal{D}, \text{ say.} \qquad \text{(B.75)}
\end{aligned}
$$

For term $\mathcal{A}$ on the rhs of (B.75) we have

$$
\mathcal{A} = O_p(\max(n^{-1}\log^{2/\delta_v} T, n^{-1/2}T^{-1/2}\sqrt{\log n}, T^{-1}\sqrt{\log n})), \qquad \text{(B.76)}
$$

by Lemma G.2(i). Term $\mathcal{B}$ is dominated by term $\mathcal{A}$. For term $\mathcal{C}$ on the rhs of (B.75) we have

$$
\mathcal{C} \leq \max_{t=1,\dots,T}\|\mathbf{P}_{t|T}^{(k^*)}\| = O_p(n^{-1}), \qquad \text{(B.77)}
$$

by Lemma F.3(iv) when $k^* \geq 1$ and by Lemma D.12 when $k^* = 0$. For term $\mathcal{D}$ on the rhs of (B.75) we have

$$
\begin{aligned}
\mathcal{D} \leq& \left\|T^{-1}\sum_{t=1}^T(\mathbf{F}_{t|T}^{(k^*)} - \mathbf{F}_t)\mathbf{F}_t\right\|\|\boldsymbol{\lambda}_i\| + \left\|T^{-1}\sum_{t=1}^T(\mathbf{F}_{t|T}^{(k^*)} - \mathbf{F}_t)\xi_{it}\right\| \\
=& O_p(\max(n^{-1}\log^{2/\delta_v} T, n^{-1/2}T^{-1/2}\sqrt{\log n}, T^{-1}\sqrt{\log n})), \qquad \text{(B.78)}
\end{aligned}
$$

by Lemmas G.2(i) and G.2(ii), and Assumption 1(a).

Now by substituting (B.76), (B.77), and (B.78) into (B.75), we have

$$
\begin{aligned}
\left|\widehat{\sigma}_i^2 - T^{-1}\sum_{t=1}^T(x_{it} - \boldsymbol{\lambda}_i^{\text{OLS}\prime}\mathbf{F}_t)^2\right| = |\widehat{\sigma}_i^2 - \sigma_i^{2\,\text{OLS}}| =& \|\widehat{\boldsymbol{\lambda}}_i - \boldsymbol{\lambda}_i^{\text{OLS}}\|^2\left\|T^{-1}\sum_{t=1}^T\mathbf{F}_t\mathbf{F}_t'\right\| \qquad \text{(B.79)} \\
&+ O_p(\max(n^{-1}\log^{2/\delta_v} T, n^{-1/2}T^{-1/2}\sqrt{\log n}, T^{-1}\sqrt{\log n})) \\
=& O_p(\max(n^{-1}\log^{2/\delta_v} T, n^{-1/2}T^{-1/2}\sqrt{\log n}, T^{-1}\sqrt{\log n})),
\end{aligned}
$$

by (B.28)-(B.31) in the proof of Proposition 2, Lemma C.12(i) combined with Assumption 6(b), and since $\|\widehat{\boldsymbol{\lambda}}_i\| \leq \|\widehat{\boldsymbol{\lambda}}_i - \boldsymbol{\lambda}_i\| + \|\boldsymbol{\lambda}_i\| = O_p(1)$ by Proposition 2 and Assumption 1(a).

Therefore, from (B.79) and Lemma D.19(i), if $n^{-1}\sqrt{T}\log^{2/\delta_v} T \to 0$, as $n, T \to \infty$, we have

$$
\begin{aligned}
\sqrt{T}(\widehat{\sigma}_i^2 - \sigma_i^2) &= \sqrt{T}(\sigma_i^{2\text{OLS}} - \sigma_i^2) + o_p(1) \\
&= T^{-1/2}\sum_{t=1}^T \xi_{it}^2 + \sqrt{T}(\boldsymbol{\lambda}_i^{\text{OLS}} - \boldsymbol{\lambda}_i)' \left\{ T^{-1}\sum_{t=1}^T \mathbf{F}_t\mathbf{F}_t' \right\} (\boldsymbol{\lambda}_i^{\text{OLS}} - \boldsymbol{\lambda}_i) - \sqrt{T}\sigma_i^2 + o_p(1) \\
&= T^{-1/2}\sum_{t=1}^T \{\xi_{it}^2 - \sigma_i^2\} + o_p(1),
\end{aligned} \tag{B.80}
$$

where we used also (D.62) in the proof of Lemma D.19. Now, by Assumption 2(c) and Davidson (1994, Theorem 14.1), we have that $\{\xi_{it}^2 - \sigma_i^2\}$ is strongly mixing with exponentially decaying coefficients and such that by Assumption 5 $\sup_{m\geq 1} r^{-1/\delta_\xi}(\mathbb{E}[|\xi_{it}^2 - \sigma_i^2|^m])^{1/m} \leq K_1$ for some finite positive real K_1 independent of t and i (Kuchibhotla and Chakrabortty, 2022, Section 2). Then, the Central Limit Theorem by Ibragimov (1962, Theorem 1.7) applies, i.e.,

$$
T^{-1/2}\sum_{t=1}^T \{\xi_{it}^2 - \sigma_i^2\} \xrightarrow{d} \mathcal{N}(0, \mathbb{E}[\xi_{it}^4] - \sigma_i^4). \tag{B.81}
$$

From (B.80), (B.81), and Slutsky's Theorem, and by noticing that the excess kurtosis is given by $\kappa_i = \mathbb{E}[\xi_{it}^4]/\sigma_i^4 - 3$, so that $\mathbb{E}[\xi_{it}^4] = \sigma_i^4(\kappa_i + 3)$, we prove part (b.1).

For part (b.2), from (15), for any $k^* \geq 0$, we have

$$
\begin{aligned}
\|\widehat{\mathbf{A}} - \mathbf{A}^{\text{OLS}}\| \leq & \left\| T^{-1}\sum_{t=2}^T \left(\mathbf{F}_{t|T}^{(k^*)}\mathbf{F}_{t-1|T}^{(k^*)\prime} + \mathbf{C}_{t,t-1|T}^{(k^*)} \right) - T^{-1}\sum_{t=2}^T \mathbf{F}_t\mathbf{F}_{t-1}' \right\| \left\| \left(T^{-1}\sum_{t=2}^T \mathbf{F}_{t-1}\mathbf{F}_{t-1}' \right)^{-1} \right\| \\
& + \left\| \left\{ T^{-1}\sum_{t=2}^T \left(\mathbf{F}_{t-1|T}^{(k^*)}\mathbf{F}_{t-1|T}^{(k^*)\prime} + \mathbf{P}_{t-1|T}^{(k^*)} \right) \right\}^{-1} - \left\{ T^{-1}\sum_{t=2}^T \mathbf{F}_{t-1}\mathbf{F}_{t-1}' \right\}^{-1} \right\| \left\| T^{-1}\sum_{t=2}^T \mathbf{F}_t\mathbf{F}_{t-1}' \right\| \\
& + \left\| T^{-1}\sum_{t=2}^T \left(\mathbf{F}_{t|T}^{(k^*)}\mathbf{F}_{t-1|T}^{(k^*)\prime} + \mathbf{C}_{t,t-1|T}^{(k^*)} \right) - T^{-1}\sum_{t=2}^T \mathbf{F}_t\mathbf{F}_{t-1}' \right\| \\
& \cdot \left\| \left\{ T^{-1}\sum_{t=2}^T \left(\mathbf{F}_{t-1|T}^{(k^*)}\mathbf{F}_{t-1|T}^{(k^*)\prime} + \mathbf{P}_{t-1|T}^{(k^*)} \right) \right\}^{-1} - \left\{ T^{-1}\sum_{t=2}^T \mathbf{F}_{t-1}\mathbf{F}_{t-1}' \right\}^{-1} \right\| \\
= & \mathfrak{A} + \mathfrak{B} + \mathfrak{C}, \text{ say.}
\end{aligned} \tag{B.82}
$$

Now,

$$
\begin{aligned}
\mathfrak{A} \leq & \left\{ \left\| T^{-1}\sum_{t=2}^T \mathbf{F}_{t|T}^{(k^*)}\mathbf{F}_{t-1|T}^{(k^*)\prime} - T^{-1}\sum_{t=2}^T \mathbf{F}_t\mathbf{F}_{t-1}' \right\| + \max_{t=2,\dots,T} \|\mathbf{C}_{t,t-1|T}^{(k^*)}\| \right\} \left\| \left(T^{-1}\sum_{t=2}^T \mathbf{F}_{t-1}\mathbf{F}_{t-1}' \right)^{-1} \right\| \\
= & O_p(\max(n^{-1}\log^{2/\delta_v} T, n^{-1/2}T^{-1/2}\sqrt{\log n}, T^{-1}\sqrt{\log n})),
\end{aligned} \tag{B.83}
$$

and

$$
\begin{aligned}
\mathfrak{B} \leq & \left\{ \left\| T^{-1}\sum_{t=2}^T \mathbf{F}_{t-1|T}^{(k^*)}\mathbf{F}_{t-1|T}^{(k^*)\prime} - T^{-1}\sum_{t=2}^T \mathbf{F}_{t-1}\mathbf{F}_{t-1}' \right\| + \max_{t=2,\dots,T} \|\mathbf{P}_{t-1|T}^{(k^*)}\| \right\} \\
& \cdot \left\| \left\{ T^{-1}\sum_{t=2}^T \left(\mathbf{F}_{t-1|T}^{(k^*)}\mathbf{F}_{t-1|T}^{(k^*)\prime} + \mathbf{P}_{t-1|T}^{(k^*)} \right) \right\}^{-1} \right\| \left\| \left(T^{-1}\sum_{t=2}^T \mathbf{F}_{t-1}\mathbf{F}_{t-1}' \right)^{-1} \right\| \left\| T^{-1}\sum_{t=2}^T \mathbf{F}_t\mathbf{F}_{t-1}' \right\| \\
= & O_p(\max(n^{-1}\log^{2/\delta_v} T, n^{-1/2}T^{-1/2}\sqrt{\log n}, T^{-1}\sqrt{\log n})),
\end{aligned} \tag{B.84}
$$

where we used: Lemma G.2(i), Lemma F.3(iv) when $k^* \geq 1$ or Lemma D.12 when $k^* = 0$ which can be both applied to $\mathbf{C}_{t,t-1|T}^{(k^*)}$ can be obtained by the upper right block of $\mathbf{P}_{t|T}^{(k^*)}$ when this is computed from the the Kalman smoother having the augmented state vector $(\mathbf{F}_t' \, \mathbf{F}_{t-1}')'$, and we also used the fact that $\|(T^{-1}\sum_{t=2}^T \mathbf{F}_{t-1}\mathbf{F}_{t-1}')^{-1}\| = O_p(1)$ by Lemma (C.13). For $\mathfrak{B}$ we also used Lemma C.12(i) combined with the fact that $\|\boldsymbol{\Gamma}_1^F\| \leq 1$ by Cauchy-Schwartz inequality and Assumption 6(b). Clearly $\mathfrak{C}$ is dominated by $\mathfrak{A}$ and $\mathfrak{B}$.

By substituting (B.83) and (B.84) into (B.82), we have

$$\|\widehat{\mathbf{A}} - \mathbf{A}^{\rm OLS}\| = O_p(\max(n^{-1}\log^{2/\delta_v} T, n^{-1/2}T^{-1/2}\sqrt{\log n}, T^{-1}\sqrt{\log n})).$$

Therefore, if $n^{-1}\sqrt{T}\log^{2/\delta_v} T \to 0$, as $n, T \to \infty$, we have

$$\sqrt{T}(\mathrm{vec}(\widehat{\mathbf{A}}) - \mathrm{vec}(\mathbf{A})) = \sqrt{T}(\mathrm{vec}(\mathbf{A}^{\rm OLS}) - \mathrm{vec}(\mathbf{A})) + o_p(1),$$

and the proof of part (b.2) follows directly from Hamilton (1994, Proposition 11.2) and Slutsky's Theorem.

For part (b.3), following a decomposition analogous to the one in (B.74), from (16), for any $k^* \ge 0$, we have

$$\begin{aligned}
\widehat{\mathbf{\Gamma}}^v =& T^{-1}\sum_{t=2}^T(\mathbf{F}_t - \widehat{\mathbf{A}}\mathbf{F}_{t-1})(\mathbf{F}_t - \mathbf{A}^{\rm OLS}\mathbf{F}_{t-1})' + (\widehat{\mathbf{A}} - \mathbf{A}^{\rm OLS})\left\{T^{-1}\sum_{t=2}^T \mathbf{F}_{t-1}\mathbf{F}_{t-1}'\right\}(\widehat{\mathbf{A}} - \mathbf{A}^{\rm OLS})' \\
&- (\widehat{\mathbf{A}} - \mathbf{A}^{\rm OLS})T^{-1}\sum_{t=2}^T \mathbf{F}_{t-1}(\mathbf{F}_t - \mathbf{A}^{\rm OLS\prime}\mathbf{F}_{t-1})' - T^{-1}\sum_{t=2}^T(\mathbf{F}_t - \mathbf{A}^{\rm OLS\prime}\mathbf{F}_{t-1})\mathbf{F}_{t-1}'(\widehat{\mathbf{A}} - \mathbf{A}^{\rm OLS})' \\
&+ T^{-1}\sum_{t=2}^T(\mathbf{F}_{t|T}^{(k^*)}\mathbf{F}_{t|T}^{(k^*)\prime} + \mathbf{P}_{t|T}^{(k^*)}) - T^{-1}\sum_{t=2}^T\mathbf{F}_t\mathbf{F}_t' + \widehat{\mathbf{A}}\left\{T^{-1}\sum_{t=2}^T(\mathbf{F}_{t-1|T}^{(k^*)}\mathbf{F}_{t-1|T}^{(k^*)\prime} + \mathbf{P}_{t-1|T}^{(k^*)}) - T^{-1}\sum_{t=2}^T\mathbf{F}_{t-1}\mathbf{F}_{t-1}'\right\}\widehat{\mathbf{A}}' \\
&- \widehat{\mathbf{A}}\left\{T^{-1}\sum_{t=2}^T(\mathbf{F}_{t-1|T}^{(k^*)}\mathbf{F}_{t|T}^{(k^*)\prime} + \mathbf{C}_{t,t-1|T}^{(k^*)\prime}) - T^{-1}\sum_{t=2}^T\mathbf{F}_{t-1}\mathbf{F}_t'\right\} \\
&- \left\{T^{-1}\sum_{t=2}^T(\mathbf{F}_{t-1|T}^{(k^*)}\mathbf{F}_{t|T}^{(k^*)\prime} + \mathbf{C}_{t,t-1|T}^{(k^*)}) - T^{-1}\sum_{t=2}^T\mathbf{F}_t\mathbf{F}_{t-1}'\right\}\widehat{\mathbf{A}}'.
\end{aligned}$$

Therefore, since by construction $T^{-1}\sum_{t=2}^T \mathbf{F}_{t-1}(\mathbf{F}_t - \mathbf{A}^{\rm OLS\prime}\mathbf{F}_{t-1})' = T^{-1}\sum_{t=2}^T(\mathbf{F}_t - \mathbf{A}^{\rm OLS\prime}\mathbf{F}_{t-1})\mathbf{F}_{t-1}' = \mathbf{0}_{r\times r}$, by using again Lemmas G.2 and Lemma F.3(iv) when $k^* \ge 1$ or Lemma D.12 when $k^* = 0$ which can be both applied to $\mathbf{C}_{t,t-1|T}^{(k^*)}$, as argued above, and using also (B.73), we obtain

$$\|\widehat{\mathbf{\Gamma}}^v - \mathbf{\Gamma}^{v\rm OLS}\| = O_p(\max(n^{-1}\log^{2/\delta_v} T, n^{-1/2}T^{-1/2}\sqrt{\log n}, T^{-1}\sqrt{\log n})).$$

Therefore, if $n^{-1}\sqrt{T}\log^{2/\delta_v} T \to 0$, as $n, T \to \infty$, we have

$$\sqrt{T}(\mathrm{vech}(\widehat{\mathbf{\Gamma}}^v) - \mathrm{vech}(\mathbf{\Gamma}^v)) = \sqrt{T}(\mathrm{vech}(\mathbf{\Gamma}^{v\rm OLS}) - \mathrm{vech}(\mathbf{\Gamma}^v)) + o_p(1),$$

and the proof of part (b.3) follows directly from Hamilton (1994, Proposition 11.2) and Slutsky's Theorem. This completes the proof. □

B.6 Proof of Corollary 1

For part (a), for ease of notation and without loss of generality let $s(j) = j$ for all $j = 1, \dots, \bar{n}$. Then, from (B.28), since $\bar{n}$ is finite,

$$\sqrt{T}(\mathrm{vec}(\widehat{\mathbf{\Lambda}}_{\bar{n}}) - \mathrm{vec}(\mathbf{\Lambda}_{\bar{n}})) = \left\{\mathbf{I}_{\bar{n}} \otimes \left(T^{-1}\sum_{t=1}^T \mathbf{F}_t\mathbf{F}_t'\right)^{-1}\right\}\left(T^{-1/2}\sum_{t=1}^T \mathrm{vec}(\mathbf{F}_t\boldsymbol{\xi}_{\bar{n}t}')\right) + o_p(1). \tag{B.85}$$

Moreover, since $\bar{n}$ is finite we can still apply the Central Limit Theorem by Ibragimov (1962, Theorem 1.4) so that, as $T \to \infty$,

$$T^{-1/2}\sum_{t=1}^T \mathrm{vec}(\mathbf{F}_t\boldsymbol{\xi}_{\bar{n}t}') \xrightarrow{d} \mathcal{N}(\mathbf{0}_{\bar{n}r}, \mathbf{\Sigma}_{\bar{n}}), \tag{B.86}$$

with

$$\boldsymbol{\Sigma}_{\bar{n}} = \lim_{T\to\infty} \mathbb{E}\left[\left\{T^{-1/2}\sum_{t=1}^{T}\begin{pmatrix}\mathbf{F}_t\xi_{1t}\\ \vdots \\ \mathbf{F}_t\xi_{\bar{n}t}\end{pmatrix}\right\}\left\{T^{-1/2}\sum_{t=1}^{T}\begin{pmatrix}\mathbf{F}_t\xi_{1t}\\ \vdots \\ \mathbf{F}_t\xi_{\bar{n}t}\end{pmatrix}\right\}'\right]$$
$$= \lim_{T\to\infty} T^{-1}\sum_{t,s=1}^{T}\mathbb{E}\left[\boldsymbol{\xi}_{\bar{n}t}\boldsymbol{\xi}_{\bar{n}s}'\otimes\mathbf{F}_t\mathbf{F}_s'\right] = \lim_{T\to\infty} T^{-1}\sum_{t,s=1}^{T}\mathbb{E}\left[\boldsymbol{\xi}_{\bar{n}t}\boldsymbol{\xi}_{\bar{n}s}'\right]\otimes\mathbb{E}\left[\mathbf{F}_t\mathbf{F}_s'\right], \tag{B.87}$$

where we used the fact that $\{\mathbf{F}_t\}$ and $\{\boldsymbol{\xi}_{nt}\}$ are independent processes because of Lemma C.11. The proof of part (a.1) follows from Lemmas C.12(i) and C.13, (B.85), (B.86), and Slutsky's theorem. For part (a.2) just notice that, if $\mathbb{E}[(\boldsymbol{\xi}_{n1}'\cdots\boldsymbol{\xi}_{nT}')'(\boldsymbol{\xi}_{n1}'\cdots\boldsymbol{\xi}_{nT}')] = \mathbf{I}_T\otimes\boldsymbol{\Sigma}_n^{\xi}$ for all $n,T\in\mathbb{N}$, then $\mathbb{E}[\boldsymbol{\xi}_{\bar{n}t}\boldsymbol{\xi}_{\bar{n}s}'] = \mathbf{0}_{\bar{n}\times\bar{n}}$ if $t\neq s$ while $\mathbb{E}[\boldsymbol{\xi}_{\bar{n}t}\boldsymbol{\xi}_{\bar{n}t}'] = \boldsymbol{\Sigma}_{\bar{n}}^{\xi}$. By substituting into (B.87) we get $\boldsymbol{\Sigma}_{\bar{n}} = \boldsymbol{\Sigma}_{\bar{n}}^{\xi}\otimes\boldsymbol{\Gamma}^F$, and $\boldsymbol{\mathcal{V}}_{\bar{n}} = (\mathbf{I}_{\bar{n}}\otimes\boldsymbol{\Gamma}^F)^{-1}(\boldsymbol{\Sigma}_{\bar{n}}^{\xi}\otimes\boldsymbol{\Gamma}^F)(\mathbf{I}_{\bar{n}}\otimes\boldsymbol{\Gamma}^F)^{-1} = \boldsymbol{\Sigma}_{\bar{n}}^{\xi}\otimes(\boldsymbol{\Gamma}^F)^{-1}$, thus proving part (a.2).

Turning to part (b), for ease of notation and without loss of generality let $s(j)=j$ for all $j=1,\dots,\bar{T}$. Then, from (B.64) in the proof of Proposition 3, since $\bar{T}$ is finite

$$\sqrt{n}(\widehat{\boldsymbol{F}}_{\bar{T}} - \boldsymbol{F}_{\bar{T}}) = \left(\mathbf{I}_{\bar{T}}\otimes n(\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^{\xi})^{-1}\boldsymbol{\Lambda}_n)^{-1}\right)\left(\mathbf{I}_{\bar{T}}\otimes n^{-1/2}\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^{\xi})^{-1}\right)\boldsymbol{\Xi}_{n\bar{T}} + o_p(1). \tag{B.88}$$

where $\boldsymbol{\Xi}_{n\bar{T}} = (\boldsymbol{\xi}_{n1}'\cdots\boldsymbol{\xi}_{n\bar{T}}')'$. Moreover, since $\bar{T}$ is finite we can still apply the Central Limit Theorem in Assumption 2(e), so that as $n\to\infty$

$$\left(\mathbf{I}_{\bar{T}}\otimes n^{-1/2}\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^{\xi})^{-1}\right)\boldsymbol{\Xi}_{n\bar{T}} = n^{-1/2}\sum_{i=1}^{n}\text{vec}\left(\boldsymbol{\lambda}_i(\sigma_i^2)^{-1}(\xi_{i1}\cdots\xi_{i\bar{T}})\right)\stackrel{d}{\to}\mathcal{N}(\mathbf{0}_{\bar{T}r},\boldsymbol{\Omega}_{\bar{T}}), \tag{B.89}$$

with

$$\boldsymbol{\Omega}_{\bar{T}} = \lim_{n\to\infty} n^{-1}\left(\mathbf{I}_{\bar{T}}\otimes\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^{\xi})^{-1}\right)\mathbb{E}[\boldsymbol{\Xi}_{n\bar{T}}\boldsymbol{\Xi}_{n\bar{T}}']\left(\mathbf{I}_{\bar{T}}\otimes(\boldsymbol{\Sigma}_n^{\xi})^{-1}\boldsymbol{\Lambda}_n\right)$$
$$= \lim_{n\to\infty}\mathbb{E}\left[\left\{n^{-1/2}\sum_{i=1}^{n}\begin{pmatrix}\boldsymbol{\lambda}_i(\sigma_i^2)^{-1}\xi_{i1}\\ \vdots\\ \boldsymbol{\lambda}_i(\sigma_i^2)^{-1}\xi_{iT}\end{pmatrix}\right\}\left\{n^{-1/2}\sum_{i=1}^{n}\begin{pmatrix}\boldsymbol{\lambda}_i(\sigma_i^2)^{-1}\xi_{i1}\\ \vdots\\ \boldsymbol{\lambda}_i(\sigma_i^2)^{-1}\xi_{iT}\end{pmatrix}\right\}'\right]$$
$$= \lim_{n\to\infty}\frac{1}{n}\sum_{i,j=1}^{n}\left\{\mathbb{E}[\boldsymbol{\zeta}_{i\bar{T}}\boldsymbol{\zeta}_{j\bar{T}}']\otimes\left(\boldsymbol{\lambda}_i\boldsymbol{\lambda}_j'(\sigma_i^2\sigma_j^2)^{-1}\right)\right\},$$

and recall that $\boldsymbol{\zeta}_{i\bar{T}} = (\xi_{i1}\cdots\xi_{i\bar{T}})'$. The proof of part (b.1) follows from (B.66) in the proof of Proposition 3, (B.88), (B.89), and Slutsky's theorem. For part (b.2) just notice that, if $\mathbb{E}[(\boldsymbol{\zeta}_{1T}'\cdots\boldsymbol{\zeta}_{nT}')'(\boldsymbol{\zeta}_{1T}'\cdots\boldsymbol{\zeta}_{nT}')] = \boldsymbol{\Sigma}_n^{\xi}\otimes\mathbf{I}_T$ for all $n,T\in\mathbb{N}$, then $\mathbb{E}[\boldsymbol{\zeta}_{i\bar{T}}\boldsymbol{\zeta}_{j\bar{T}}'] = \mathbf{0}_{\bar{T}\times\bar{T}}$ if $i\neq j$ while $\mathbb{E}[\boldsymbol{\zeta}_{i\bar{T}}\boldsymbol{\zeta}_{i\bar{T}}'] = \sigma_i^2\mathbf{I}_{\bar{T}}$. By substituting into (B.87) we get $\boldsymbol{\Omega}_{\bar{T}} = \mathbf{I}_{\bar{T}}\otimes\boldsymbol{\Sigma}_{\Lambda\Sigma\Lambda}$, and $\boldsymbol{\mathcal{W}}_{\bar{T}} = (\mathbf{I}_{\bar{T}}\otimes\boldsymbol{\Sigma}_{\Lambda\Sigma\Lambda})^{-1}(\mathbf{I}_{\bar{T}}\otimes\boldsymbol{\Sigma}_{\Lambda\Sigma\Lambda})(\mathbf{I}_{\bar{T}}\otimes\boldsymbol{\Sigma}_{\Lambda\Sigma\Lambda})^{-1} = \mathbf{I}_{\bar{T}}\otimes(\boldsymbol{\Sigma}_{\Lambda\Sigma\Lambda})^{-1}$, thus proving part (b.2). This completes the proof. □

B.7 Proof of Proposition 6

Under Assumptions 1, 2, 3, and 6, from Barigozzi (2023, Theorem 1), the asymptotic covariances of the PC estimator of the loadings is

$$\boldsymbol{\mathcal{V}}_i^{\text{PC}} = (\boldsymbol{\Gamma}^F)^{-1}\left(\lim_{T\to\infty}\frac{1}{T}\sum_{t,s=1}^{T}\mathbb{E}[\xi_{it}\xi_{is}]\mathbb{E}[\mathbf{F}_t\mathbf{F}_s']\right)(\boldsymbol{\Gamma}^F)^{-1}, \tag{B.90}$$

where $\boldsymbol{\Gamma}^F = \mathbf{I}_r$, by Assumption 6(b). Therefore, the expression of $\boldsymbol{\mathcal{V}}_i^{\text{PC}}$ in (B.90) coincides with the expression of $\boldsymbol{\mathcal{V}}_i$ for the asymptotic covariances of the EM estimator given in Proposition 2(b). This proves part (a).

Turning to part (b). Since, because of Proposition 3, $\lim_{n,T\to\infty}\sqrt{n}\mathbb{E}[(\widehat{\mathbf{F}}_t-\mathbf{F}_t)]=\mathbf{0}_r$, then

$$\begin{aligned}\boldsymbol{\mathcal{W}}_t=\lim_{n,T\to\infty} n\mathbb{C}\text{ov}(\widehat{\mathbf{F}}_t-\mathbf{F}_t,\widehat{\mathbf{F}}_t-\mathbf{F}_t)&=\lim_{n,T\to\infty}\left\{n\mathbb{E}[(\widehat{\mathbf{F}}_t-\mathbf{F}_t)(\widehat{\mathbf{F}}_t-\mathbf{F}_t)']-n\mathbb{E}[(\widehat{\mathbf{F}}_t-\mathbf{F}_t)]\mathbb{E}[(\widehat{\mathbf{F}}_t-\mathbf{F}_t)']\right\}\\&=\lim_{n,T\to\infty}\left\{n\mathbb{E}[(\widehat{\mathbf{F}}_t-\mathbf{F}_t)(\widehat{\mathbf{F}}_t-\mathbf{F}_t)']\right\}\\&=\lim_{n\to\infty} n\{(\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n)^{-1}\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Gamma}_n^\xi(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n(\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n)^{-1}\}.\end{aligned}\tag{B.91}$$

Similarly, because of Lemma H.1, $\lim_{n,T\to\infty}\sqrt{n}\mathbb{E}[(\widetilde{\mathbf{F}}_t-\mathbf{F}_t)]=\mathbf{0}_r$, which implies

$$\begin{aligned}\boldsymbol{\mathcal{W}}_t^{\text{PC}}=\lim_{n,T\to\infty} n\mathbb{C}\text{ov}(\widetilde{\mathbf{F}}_t-\mathbf{F}_t,\widetilde{\mathbf{F}}_t-\mathbf{F}_t)&=\lim_{n,T\to\infty}\left\{n\mathbb{E}[(\widetilde{\mathbf{F}}_t-\mathbf{F}_t)(\widetilde{\mathbf{F}}_t-\mathbf{F}_t)']-n\mathbb{E}[(\widetilde{\mathbf{F}}_t-\mathbf{F}_t)]\mathbb{E}[(\widetilde{\mathbf{F}}_t-\mathbf{F}_t)']\right\}\\&=\lim_{n,T\to\infty}\left\{n\mathbb{E}[(\widetilde{\mathbf{F}}_t-\mathbf{F}_t)(\widetilde{\mathbf{F}}_t-\mathbf{F}_t)']\right\}\\&=\lim_{n\to\infty} n\{(\mathbf{M}_n^\chi)^{-1}\boldsymbol{\Lambda}_n'\boldsymbol{\Gamma}_n^\xi\boldsymbol{\Lambda}_n(\boldsymbol{\Lambda}_n'\boldsymbol{\Lambda}_n)^{-1}\}\\&=\lim_{n\to\infty} n\{(\boldsymbol{\Lambda}_n'\boldsymbol{\Lambda}_n)^{-1}\boldsymbol{\Lambda}_n'\boldsymbol{\Gamma}_n^\xi\boldsymbol{\Lambda}_n(\boldsymbol{\Lambda}_n'\boldsymbol{\Lambda}_n)^{-1}\},\end{aligned}\tag{B.92}$$

where we used the fact that $\lim_{n\to\infty}n^{-1}\mathbf{M}_n^\chi=\lim_{n\to\infty}n^{-1}\boldsymbol{\Lambda}_n'\boldsymbol{\Lambda}_n$, because of Assumption 6(b).

Moreover, if $\sqrt{n\log n}/T\to 0$ (which implies also $\sqrt{n}/T\to 0$), Proposition 3 (see in particular (B.64) in its proof) and Lemma H.1 (see in particular (H.8) in its proof) jointly imply that

$$\begin{pmatrix}\sqrt{n}(\widehat{\mathbf{F}}_t-\mathbf{F}_t)\\ \sqrt{n}(\widetilde{\mathbf{F}}_t-\mathbf{F}_t)\end{pmatrix}=\begin{pmatrix}\sqrt{n}(\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n)^{-1}\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\xi}_{nt}\\ \sqrt{n}(\mathbf{M}_n^\chi)^{-1}\boldsymbol{\Lambda}_n'\boldsymbol{\xi}_{nt}\end{pmatrix}+o_p(1)\xrightarrow{p}\mathcal{N}\left(\mathbf{0}_{2r},\begin{pmatrix}\boldsymbol{\mathcal{W}}_t & \boldsymbol{\mathcal{U}}_t\\ \boldsymbol{\mathcal{U}}_t' & \boldsymbol{\mathcal{W}}_t^{\text{PC}}\end{pmatrix}\right),$$

where

$$\begin{aligned}\boldsymbol{\mathcal{U}}_t=\lim_{n\to\infty} n\mathbb{C}\text{ov}(\widehat{\mathbf{F}}_t-\mathbf{F}_t,\widetilde{\mathbf{F}}_t-\mathbf{F}_t)&=\lim_{n,T\to\infty}\left\{n\mathbb{E}[(\widehat{\mathbf{F}}_t-\mathbf{F}_t)(\widetilde{\mathbf{F}}_t-\mathbf{F}_t)']-n\mathbb{E}[(\widehat{\mathbf{F}}_t-\mathbf{F}_t)]\mathbb{E}[(\widetilde{\mathbf{F}}_t-\mathbf{F}_t)']\right\}\\&=\lim_{n,T\to\infty}\left\{n\mathbb{E}[(\widehat{\mathbf{F}}_t-\mathbf{F}_t)(\widetilde{\mathbf{F}}_t-\mathbf{F}_t)']\right\}\\&=\lim_{n\to\infty} n\{(\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n)^{-1}\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Gamma}_n^\xi\boldsymbol{\Lambda}_n(\boldsymbol{\Lambda}_n'\boldsymbol{\Lambda}_n)^{-1}\}.\end{aligned}\tag{B.93}$$

From (B.91) and (B.92) it follows that

$$\begin{aligned}\boldsymbol{\mathcal{W}}_t^{\text{PC}}-\boldsymbol{\mathcal{W}}_t=&\lim_{n,T\to\infty}\left\{n\,\mathbb{E}[(\widetilde{\mathbf{F}}_t-\widehat{\mathbf{F}}_t)(\widetilde{\mathbf{F}}_t-\widehat{\mathbf{F}}_t)']\right\}\\&+\lim_{n,T\to\infty}\left\{n\,\mathbb{E}[(\widehat{\mathbf{F}}_t-\mathbf{F}_t)(\widetilde{\mathbf{F}}_t-\widehat{\mathbf{F}}_t)']\right\}+\lim_{n,T\to\infty}\left\{n\,\mathbb{E}[(\widetilde{\mathbf{F}}_t-\widehat{\mathbf{F}}_t)(\widehat{\mathbf{F}}_t-\mathbf{F}_t)']\right\}\\=&\,\boldsymbol{\mathcal{C}}_1+\boldsymbol{\mathcal{C}}_2+\boldsymbol{\mathcal{C}}_2',\text{ say.}\end{aligned}\tag{B.94}$$

Let us now define $\boldsymbol{\mathcal{O}}_n^\xi=\boldsymbol{\Gamma}_n^\xi-\boldsymbol{\Sigma}_n^\xi$, which is the $n\times n$ matrix of off-diagonal entries of the idiosyncratic covariance. Because of (B.91), (B.93), and Lemma C.9 we can write

$$\begin{aligned}\boldsymbol{\mathcal{C}}_2=&\lim_{n,T\to\infty}\left\{n\,\mathbb{E}[(\widehat{\mathbf{F}}_t-\mathbf{F}_t)(\widetilde{\mathbf{F}}_t-\widehat{\mathbf{F}}_t)']\right\}\\=&\lim_{n,T\to\infty}\left\{n\mathbb{E}[(\widehat{\mathbf{F}}_t-\mathbf{F}_t)(\widetilde{\mathbf{F}}_t-\mathbf{F}_t)']-n\mathbb{E}[(\widehat{\mathbf{F}}_t-\mathbf{F}_t)(\widehat{\mathbf{F}}_t-\mathbf{F}_t)']\right\}=\boldsymbol{\mathcal{U}}_t-\boldsymbol{\mathcal{W}}_t\\=&\lim_{n\to\infty}\left\{n(\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n)^{-1}\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Gamma}_n^\xi\left[\boldsymbol{\Lambda}_n(\boldsymbol{\Lambda}_n'\boldsymbol{\Lambda}_n)^{-1}-(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n(\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n)^{-1}\boldsymbol{\Lambda}_n'\right]\right\}\\=&\lim_{n\to\infty}\left\{n(\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n)^{-1}\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\mathcal{O}}_n^\xi\left[\boldsymbol{\Lambda}_n(\boldsymbol{\Lambda}_n'\boldsymbol{\Lambda}_n)^{-1}-(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n(\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n)^{-1}\boldsymbol{\Lambda}_n'\right]\right\}\\=&(\boldsymbol{\Sigma}_{\Lambda\Sigma\Lambda})^{-1}\boldsymbol{\Sigma}_\Lambda^{1/2}\lim_{n\to\infty}\left\{\mathbf{V}_n^{\chi\prime}(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\mathcal{O}}_n^\xi\mathbf{V}_n^\chi\right\}\boldsymbol{\Sigma}_\Lambda^{-1/2}\\&-(\boldsymbol{\Sigma}_{\Lambda\Sigma\Lambda})^{-1}\boldsymbol{\Sigma}_\Lambda^{1/2}\lim_{n\to\infty}\left\{\mathbf{V}_n^{\chi\prime}(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\mathcal{O}}_n^\xi(\boldsymbol{\Sigma}_n^\xi)^{-1}\mathbf{V}_n^\chi\right\}\boldsymbol{\Sigma}_\Lambda^{1/2}(\boldsymbol{\Sigma}_{\Lambda\Sigma\Lambda})^{-1}.\end{aligned}\tag{B.95}$$

Now, let $\mathfrak{V}_n = \mathbf{V}_n^{\chi\prime}(\boldsymbol{\Sigma}_n^{\xi})^{-1}\boldsymbol{\mathcal{O}}_n^{\xi}(\boldsymbol{\Sigma}_n^{\xi})^{-1}\mathbf{V}_n^{\chi}$, then, for any $h,k=1,\ldots,r$,

$$
\begin{aligned}
[\mathfrak{V}_n]_{hk} &= \sum_{i,j,\ell,m=1}^{n} [\mathbf{V}_n^{\chi\prime}]_{hi}\,[(\boldsymbol{\Sigma}_n^{\xi})^{-1}]_{ij}\,[\boldsymbol{\mathcal{O}}_n^{\xi}]_{j\ell}[(\boldsymbol{\Sigma}_n^{\xi})^{-1}]_{\ell m}\,[\mathbf{V}_n^{\chi}]_{mk}\\
&= \sum_{\substack{i,\ell=1\\ i\neq\ell}}^{n} [\mathbf{V}_n^{\chi}]_{ih}\,[(\boldsymbol{\Sigma}_n^{\xi})^{-1}]_{ii}\,[\boldsymbol{\mathcal{O}}_n^{\xi}]_{i\ell}[(\boldsymbol{\Sigma}_n^{\xi})^{-1}]_{\ell\ell}\,[\mathbf{V}_n^{\chi}]_{\ell k}.
\end{aligned} \tag{B.96}
$$

Now, let $\boldsymbol{\iota}_{ni}$ be the n-dimensional vector with ith entry equal one and all other equal zero, and let $\boldsymbol{s}_j$ the r-dimensional vector with jth entry equal one and all other equal zero. Then, since $\mathbf{V}_n^{\chi} = \boldsymbol{\Gamma}_n^{\chi}\mathbf{V}_n^{\chi}(\mathbf{M}_n^{\chi})^{-1}$, there exists a finite positive integer $\bar{n}$, such that, for all $n\geq\bar{n}$

$$
\begin{aligned}
\max_{i=1,\ldots,n}\max_{j=1,\ldots,r}|v_{ij}^{\chi}| &= \max_{i=1,\ldots,n}\max_{j=1,\ldots,r}|\boldsymbol{\iota}_{ni}'\mathbf{V}_n^{\chi}\boldsymbol{s}_j|\\
&= \max_{i=1,\ldots,n}\max_{j=1,\ldots,r}|\boldsymbol{\iota}_{ni}'\boldsymbol{\Gamma}_n^{\chi}\mathbf{V}_n^{\chi}(\mathbf{M}_n^{\chi})^{-1}\boldsymbol{s}_j|\\
&\leq \max_{i=1,\ldots,n}\|\boldsymbol{\iota}_{ni}'\boldsymbol{\Gamma}_n^{\chi}\|\,\|\mathbf{V}_n^{\chi}\|\max_{j=1,\ldots,r}|(\mu_{jn}^{\chi})^{-1}|\\
&= \max_{i=1,\ldots,n}\|\boldsymbol{\lambda}_i'\boldsymbol{\Gamma}^F\boldsymbol{\Lambda}_n'\|n^{-1}\underline{C}_r^{-1}\\
&\leq \max_{i=1,\ldots,n}\|\boldsymbol{\lambda}_i\|\,\|\boldsymbol{\Gamma}^F\|\,\|\boldsymbol{\Lambda}_n\|n^{-1}\underline{C}_r^{-1}\\
&\leq M_\lambda^2 M_F n^{-1/2}\underline{C}_r^{-1},
\end{aligned} \tag{B.97}
$$

where we used Assumptions 1(a) and 1(b), and Lemmas C.1(iv) and C.2.

Therefore, from (B.96) and (B.97), and Assumption 2(a),

$$
\begin{aligned}
\max_{h,k=1,\ldots r}|[\mathfrak{V}_n]_{hk}| &\leq \left(\max_{h=1,\ldots r}\max_{i=1,\ldots n}|[\mathbf{V}_n^{\chi}]_{ih}|\right)^2\left(\max_{i=1,\ldots n}(\sigma_i^2)^{-1}\right)^2\sum_{i,\ell=1}^{n}|[\boldsymbol{\mathcal{O}}_n^{\xi}]_{i\ell}|\\
&\leq C_v^2\left\{n\left(\min_{i=1,\ldots n}\sigma_i^2\right)^2\right\}^{-1}\sum_{i,\ell=1}^{n}|[\boldsymbol{\mathcal{O}}_n^{\xi}]_{i\ell}| \leq n^{-1}C_v^2C_\xi^2\sum_{i,\ell=1}^{n}|[\boldsymbol{\mathcal{O}}_n^{\xi}]_{i\ell}|.
\end{aligned} \tag{B.98}
$$

Moreover, since by assumption we have

$$
\lim_{n\to\infty} n^{-1}\sum_{i,\ell=1}^{n}|[\boldsymbol{\mathcal{O}}_n^{\xi}]_{i\ell}| = \lim_{n\to\infty} n^{-1}\sum_{\substack{i,\ell=1\\ i\neq\ell}}^{n}|[\boldsymbol{\Gamma}_n^{\xi}]_{i\ell}| = 0,
$$

then, from (B.98), we have $\lim_{n\to\infty}\max_{h,k=1,\ldots r}|[\mathfrak{V}_n]_{hk}| = 0$. Therefore,

$$
\lim_{n\to\infty}\left\{\mathbf{V}_n^{\chi\prime}(\boldsymbol{\Sigma}_n^{\xi})^{-1}\boldsymbol{\mathcal{O}}_n^{\xi}(\boldsymbol{\Sigma}_n^{\xi})^{-1}\mathbf{V}_n^{\chi}\right\} = \mathbf{0}_{r\times r}. \tag{B.99}
$$

Following the same reasoning it also holds that

$$
\lim_{n\to\infty}\left\{\mathbf{V}_n^{\chi\prime}(\boldsymbol{\Sigma}_n^{\xi})^{-1}\boldsymbol{\mathcal{O}}_n^{\xi}\mathbf{V}_n^{\chi}\right\} = \mathbf{0}_{r\times r}. \tag{B.100}
$$

By substituting, (B.99) and (B.100) into (B.95), we obtain $\boldsymbol{\mathcal{C}}_2 = \mathbf{0}_{r\times r}$. Finally,

$$
\begin{aligned}
\boldsymbol{\mathcal{C}}_1 &= \lim_{n,T\to\infty}\left\{n\,\mathbb{E}[(\widetilde{\mathbf{F}}_t-\widehat{\mathbf{F}}_t)(\widetilde{\mathbf{F}}_t-\widehat{\mathbf{F}}_t)']\right\}\\
&= \lim_{n,T\to\infty}\left\{n\,\mathbb{E}[(\widehat{\mathbf{F}}_t-\mathbf{F}_t)(\widehat{\mathbf{F}}_t-\mathbf{F}_t)']\right\} + \lim_{n,T\to\infty}\left\{n\,\mathbb{E}[(\widetilde{\mathbf{F}}_t-\mathbf{F}_t)(\widetilde{\mathbf{F}}_t-\mathbf{F}_t)']\right\}\\
&\quad - \lim_{n,T\to\infty}\left\{n\,\mathbb{E}[(\widetilde{\mathbf{F}}_t-\mathbf{F}_t)(\widehat{\mathbf{F}}_t-\mathbf{F}_t)']\right\} - \lim_{n,T\to\infty}\left\{n\,\mathbb{E}[(\widehat{\mathbf{F}}_t-\mathbf{F}_t)(\widetilde{\mathbf{F}}_t-\mathbf{F}_t)']\right\}\\
&= \boldsymbol{\mathcal{W}}_t + \boldsymbol{\mathcal{W}}_t^{\mathrm{PC}} - (\boldsymbol{\mathcal{U}}_t + \boldsymbol{\mathcal{U}}_t').
\end{aligned}
$$

Therefore, $\boldsymbol{\mathcal{C}}_1$ is positive definite since $\boldsymbol{\mathcal{W}}_t$ and $\boldsymbol{\mathcal{W}}_t^{\mathrm{PC}}$ are positive definite by Assumptions 1(a), 2(a), and 2(f), and, from (B.94) we prove part (ii). This completes the proof. □

C General lemmas

Lemma C.1. *Under Assumptions 1 and 2:*

(i) for all $n\in\mathbb{N}$ and $T\in\mathbb{N}$, $(nT)^{-1}\sum_{i,j=1}^n\sum_{t,s=1}^T|\mathbb{E}[\xi_{it}\xi_{js}]|\le M_1$, for some finite positive real M_1 independent of n and T;

(ii) for all $n\in\mathbb{N}$ and $t\in\mathbb{Z}$, $n^{-1}\sum_{i,j=1}^n|\mathbb{E}[\xi_{it}\xi_{jt}]|\le M_2$, for some finite positive real M_2 independent of n and t;

(iii) for all $i\in\mathbb{N}$ and $T\in\mathbb{N}$, $T^{-1}\sum_{t,s=1}^T|\mathbb{E}[\xi_{it}\xi_{is}]|\le M_3$, for some finite positive real M_3 independent of i and T;

(iv) for all $j=1,\dots,r$, $\underline{C}_j\le\liminf_{n\to\infty}n^{-1}\mu_{jn}^\chi\le\limsup_{n\to\infty}n^{-1}\mu_{jn}^\chi\le\overline{C}_j$, for some finite positive reals $\underline{C}_j$ and $\overline{C}_j$;

(v) for all $n\in\mathbb{N}$, $\mu_{1n}^\xi=\|\mathbf{\Gamma}_n^\xi\|\le M_2$, where M_2 is defined in part (ii);

(vi) for all $j=1,\dots,r$, $\underline{C}_j\le\liminf_{n\to\infty}n^{-1}\mu_{jn}^x\le\limsup_{n\to\infty}n^{-1}\mu_{jn}^\chi\le\overline{C}_j$, and for all $n\in\mathbb{N}$, $\mu_{r+1,n}^x\le M_2$, where M_2 is defined in part (ii).

Proof. Using Assumptions 2(a) and 2(b), we have:

$$
\begin{aligned}
(nT)^{-1}\sum_{i,j=1}^n\sum_{t,s=1}^T|\mathbb{E}[\xi_{it}\xi_{js}]| &= n^{-1}\sum_{i,j=1}^n\sum_{k=-(T-1)}^{T-1}\left(1-\frac{|k|}{T}\right)|\mathbb{E}[\xi_{it}\xi_{j,t-k}]|\\
&\le n^{-1}\sum_{i,j=1}^n\sum_{k=-\infty}^{\infty}\rho^{|k|}M_{ij}\\
&= n^{-1}\sum_{i=1}^n\sum_{k=-\infty}^{\infty}\rho^{|k|}M_{ii}+n^{-1}\sum_{i,j=1,i\neq j}^n\sum_{k=-\infty}^{\infty}\rho^{|k|}M_{ij}\\
&= n^{-1}\sum_{i=1}^n\sum_{k=-\infty}^{\infty}\rho^{|k|}\sigma_i^2+\max_{i=1,\dots,n}\sum_{j=1,j\neq i}^n\sum_{k=-\infty}^{\infty}\rho^{|k|}M_{ij}\\
&\le\frac{C_\xi(1+\rho)}{1-\rho}+\frac{M_\xi(1+\rho)}{1-\rho}.
\end{aligned}
$$

Similarly,

$$
n^{-1}\sum_{i,j=1}^n|\mathbb{E}[\xi_{it}\xi_{jt}]|\le n^{-1}\sum_{i=1}^n\sigma_i^2+\max_{i=1,\dots,n}\sum_{j=1,j\neq i}^n M_{ij}\le C_\xi+M_\xi,
$$

and

$$
\begin{aligned}
T^{-1}\sum_{t,s=1}^T|\mathbb{E}[\xi_{it}\xi_{is}]| &= \sum_{k=-(T-1)}^{T-1}\left(1-\frac{|k|}{T}\right)|\mathbb{E}[\xi_{it}\xi_{i,t-k}]|\\
&\le\sum_{k=-\infty}^{\infty}\rho^{|k|}M_{ii}\le\frac{1+\rho}{1-\rho}\sigma_i^2\le\frac{C_\xi(1+\rho)}{1-\rho}.
\end{aligned}
$$

Defining, $M_1=\frac{(C_\xi+M_\xi)(1+\rho)}{1-\rho}$, $M_2=C_\xi+M_\xi$, and $M_3=\frac{C_\xi(1+\rho)}{1-\rho}$, we prove parts (i), (ii), and (iii).

For part (iv), first notice that, for all $n\in\mathbb{N}$, the r non-zero eigenvalues of $\mathbf{\Gamma}_n^\chi$ are also the r eigenvalues of $n^{-1}\mathbf{\Lambda}_n'\mathbf{\Lambda}_n\mathbf{\Gamma}^F$. Thus, by Merikoski and Kumar (2004, Theorem 7), for all $j=1,\dots,r$ and all $n\in\mathbb{N}$, we have

$$
n^{-1}\nu^{(r)}(\mathbf{\Lambda}_n'\mathbf{\Lambda}_n)\,\nu^{(j)}(\mathbf{\Gamma}^F)\le n^{-1}\mu_{jn}^\chi\le n^{-1}\nu^{(j)}(\mathbf{\Lambda}_n'\mathbf{\Lambda}_n)\,\nu^{(1)}(\mathbf{\Gamma}^F).
$$

The proof then follows from Assumptions 1(a) and 1(b). Indeed, by continuity of eigenvalues, Assumption 1(a) implies that, for all $j=1,\dots,r$, and all $n>N_0$

$$
n^{-1}\nu^{(j)}(\mathbf{\Lambda}_n'\mathbf{\Lambda}_n)=\nu^{(j)}(\mathbf{\Sigma}_\Lambda), \tag{C.1}
$$

and there exist finite positive reals m_λ and M_λ such that $0<m_\lambda^2\le\nu^{(r)}(\mathbf{\Sigma}_\Lambda)\le\nu^{(1)}(\mathbf{\Sigma}_\Lambda)\le M_\lambda^2<\infty$. Similarly, Assumption 1(b), implies that there exist finite positive reals m_F and M_F such that $0<m_F\le\nu^{(r)}(\mathbf{\Gamma}^F)\le\nu^{(1)}(\mathbf{\Gamma}^F)\le M_F<\infty$.

For part (v), by Assumptions 2(a) and 2(b):

$$
\|\mathbf{\Gamma}_n^\xi\|\le\max_{i=1,\dots,n}\sum_{j=1}^n|\mathbb{E}[\xi_{it}\xi_{jt}]|=\max_{i=1,\dots,n}\sigma_i^2+\max_{i=1,\dots,n}\sum_{j=1,j\neq i}^n M_{ij}\le C_\xi+M_\xi=M_2.
$$

Part (vi) follows from parts (iv) and (v) and Weyl's inequality (Merikoski and Kumar, 2004, Theorem 1). This completes the proof. □

Lemma C.2. *Under Assumption 1, as $n\to\infty$, $n^{-1/2}\|\mathbf{\Lambda}_n\| = O(1)$.*

Proof. By definition and using Assumption 1(a),

$$\lim_{n\to\infty} n^{-1/2}\|\mathbf{\Lambda}_n\| = \lim_{n\to\infty} n^{-1/2}\sqrt{\nu^{(1)}(\mathbf{\Lambda}_n'\mathbf{\Lambda}_n)} = \sqrt{\nu^{(1)}(\mathbf{\Sigma}_\Lambda)} \le M_\lambda.$$

This completes the proof. □

Lemma C.3. *For any $r\times r$ symmetric and positive definite matrix $\boldsymbol{P}$ with $\|\boldsymbol{P}\|\le C_P$ for some finite positive real C_P, under Assumptions 1 and 2, as $n\to\infty$,*

(i) $n\|(\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n + \boldsymbol{P}^{-1})^{-1}\| = O(1)$;
(ii) $n\|(\mathbf{\Lambda}_n'(\mathbf{\Gamma}_n^\xi)^{-1}\mathbf{\Lambda}_n + \boldsymbol{P}^{-1})^{-1}\| = O(1)$;
(iii) $n\|(\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n)^{-1}\| = O(1)$;
(iv) $n\|(\mathbf{\Lambda}_n'(\mathbf{\Gamma}_n^\xi)^{-1}\mathbf{\Lambda}_n)^{-1}\| = O(1)$;
(v) $n^{-1}\|\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n\| = O(1)$;
(vi) $n^{-1}\|\mathbf{\Lambda}_n'(\mathbf{\Gamma}_n^\xi)^{-1}\mathbf{\Lambda}_n\| = O(1)$;
(vii) $n^{-1/2}\|\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\| = O(1)$;
(viii) $n^{-1/2}\|\mathbf{\Lambda}_n'(\mathbf{\Gamma}_n^\xi)^{-1}\| = O(1)$.

Proof. By Merikoski and Kumar (2004, Theorems 1, which is Weyl's inequality, and 7)

$$\begin{aligned}
n\|(\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n + \boldsymbol{P}^{-1})^{-1}\| &= \frac{n}{\nu^{(r)}(\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n + \boldsymbol{P}^{-1})} \le \frac{n}{\nu^{(r)}(\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n) + \nu^{(r)}(\boldsymbol{P}^{-1})}\\
&\le \frac{n}{\nu^{(r)}(\mathbf{\Lambda}_n\mathbf{\Lambda}_n')\nu^{(n)}((\mathbf{\Sigma}_n^\xi)^{-1}) + [\nu^{(1)}(\boldsymbol{P})]^{-1}}\\
&\le \frac{n}{\nu^{(r)}(\mathbf{\Lambda}_n'\mathbf{\Lambda}_n)[\nu^{(1)}(\mathbf{\Sigma}_n^\xi)]^{-1} + [\nu^{(1)}(\boldsymbol{P})]^{-1}}\\
&\le \frac{n}{\nu^{(r)}(\mathbf{\Lambda}_n'\mathbf{\Lambda}_n)C_\xi^{-1} + C_P^{-1}}\\
&= \frac{1}{n^{-1}\nu^{(r)}(\mathbf{\Lambda}_n'\mathbf{\Lambda}_n)C_\xi^{-1} + n^{-1}C_P^{-1}},
\end{aligned} \tag{C.2}$$

by Assumption 2(a) and since $\boldsymbol{P}$ finite by assumption. Then, as shown in (C.1) in the proof of Lemma C.1(iv),

$$\underline{C}_j \le \lim\inf_{n\to\infty} n^{-1}\nu^{(j)}(\mathbf{\Lambda}_n'\mathbf{\Lambda}_n) \le \lim\sup_{n\to\infty} n^{-1}\nu^{(j)}(\mathbf{\Lambda}_n'\mathbf{\Lambda}_n) \le \overline{C}_j, \qquad j=1,\dots,r. \tag{C.3}$$

Thus, we have $\lim_{n\to\infty} n^{-1}\nu^{(r)}(\mathbf{\Lambda}_n'\mathbf{\Lambda}_n) \ge \underline{C}_r$, which, once substituted in (C.2), proves part (i).

For part (ii) the proof is the same as part (i), but in (C.2) we use Lemma C.1(v) instead of Assumption 2(a), thus replacing C_ξ with M_2. Parts (iii) and (iv) are obvious by just setting $\boldsymbol{P}^{-1} = \mathbf{0}_{r\times r}$ in the first step of (C.2).

For part (v) we have

$$\begin{aligned}
n^{-1}\|\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n\| &= n^{-1}\nu^{(1)}(\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n) = n^{-1}\nu^{(1)}(\mathbf{\Lambda}_n\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1})\\
&\le n^{-1}\nu^{(1)}(\mathbf{\Lambda}_n\mathbf{\Lambda}_n')\nu^{(1)}((\mathbf{\Sigma}_n^\xi)^{-1}) = \frac{\nu^{(1)}(\mathbf{\Lambda}_n'\mathbf{\Lambda}_n)}{n\nu^{(n)}(\mathbf{\Sigma}_n^\xi)}\\
&\le \frac{\nu^{(1)}(\mathbf{\Lambda}_n'\mathbf{\Lambda}_n)}{nC_\xi^{-1}},
\end{aligned} \tag{C.4}$$

by Assumption 2(a). Then, by (C.3) in the proof of Lemma C.6, we have $\lim_{n\to\infty} n^{-1}\nu^{(1)}(\mathbf{\Lambda}_n'\mathbf{\Lambda}_n) \le \overline{C}_1$, which, once substituted in (C.4), proves part (v).

For part (vi) the proof is the same as part (v), but in (C.4) we use Assumption 2(f) instead of Assumption 2(a), thus replacing C_ξ^{-1} with L_ξ.

For part (vii)

$$\|\boldsymbol{\Lambda}_n(\boldsymbol{\Sigma}_n^{\xi})^{-1}\| \le \|\boldsymbol{\Lambda}_n\|\, \|(\boldsymbol{\Sigma}_n^{\xi})^{-1}\| \le \|\boldsymbol{\Lambda}_n\| \max_{i=1,\ldots,n}(\sigma_i^2)^{-1} = O(\sqrt{n}),$$

by Lemma C.2 and Assumption 2(a). And for part (viii) the proof is the same as for part (vii) but using Assumption 2(f) instead of Assumption 2(a). This proves parts (vii) and (viii) and completes the proof. □

Lemma C.4. *Given two invertible matrices $\boldsymbol{K}$ and $\boldsymbol{H}$ the following holds:*

$$(\boldsymbol{H}+\boldsymbol{K})^{-1} = \boldsymbol{K}^{-1} - (\boldsymbol{H}+\boldsymbol{K})^{-1}\boldsymbol{H}\boldsymbol{K}^{-1}.$$

Proof. We have

$$\begin{aligned}(\boldsymbol{H}+\boldsymbol{K})^{-1} &= (\boldsymbol{H}+\boldsymbol{K})^{-1} - \boldsymbol{K}^{-1} + \boldsymbol{K}^{-1} = (\boldsymbol{H}+\boldsymbol{K})^{-1}(\boldsymbol{K}-(\boldsymbol{H}+\boldsymbol{K}))\boldsymbol{K}^{-1} + \boldsymbol{K}^{-1}\\ &= (\boldsymbol{H}+\boldsymbol{K})^{-1}(-\boldsymbol{H})\boldsymbol{K}^{-1} + \boldsymbol{K}^{-1} = \boldsymbol{K}^{-1} - (\boldsymbol{H}+\boldsymbol{K})^{-1}\boldsymbol{H}\boldsymbol{K}^{-1}. \end{aligned} \tag{C.5}$$

This completes the proof. □

Lemma C.5. *For $m < n$ with m independent of n and given*

(a) an $m \times m$ matrix $\boldsymbol{A}$ symmetric and positive definite with $\|\boldsymbol{A}\| \le M_A$;

(b) an $n \times n$ matrix $\boldsymbol{B}$ symmetric and positive definite with $\|\boldsymbol{B}\| \le M_B$;

(c) an $n \times m$ matrix $\boldsymbol{C}$ such that $\boldsymbol{C}'\boldsymbol{C}$ is positive definite with

$$\underline{M}_{C_j} \le \liminf_{n\to\infty} n^{-1}\nu^{(j)}(\boldsymbol{C}'\boldsymbol{C}) \le \limsup_{n\to\infty} n^{-1}\nu^{(j)}(\boldsymbol{C}'\boldsymbol{C}) \le \overline{M}_{C_j}, \qquad j=1,\ldots,m;$$

where, M_A, M_B, $\underline{M}_{C_j}$ and $\overline{M}_{C_j}$ are finite positive reals independent of n and m, then the following holds

$$\|(\boldsymbol{A}^{-1}+\boldsymbol{C}'\boldsymbol{B}^{-1}\boldsymbol{C})^{-1}\boldsymbol{C}'\boldsymbol{B}^{-1}\boldsymbol{C} - \mathbf{I}_m\| = O(n^{-1}).$$

Proof. In Lemma C.4 set $\boldsymbol{K} = \boldsymbol{C}'\boldsymbol{B}^{-1}\boldsymbol{C}$ and $\boldsymbol{H} = \boldsymbol{A}^{-1}$, then,

$$(\boldsymbol{A}^{-1}+\boldsymbol{C}'\boldsymbol{B}^{-1}\boldsymbol{C})^{-1} = (\boldsymbol{C}'\boldsymbol{B}^{-1}\boldsymbol{C})^{-1} - (\boldsymbol{A}^{-1}+\boldsymbol{C}'\boldsymbol{B}^{-1}\boldsymbol{C})^{-1}\boldsymbol{A}^{-1}(\boldsymbol{C}'\boldsymbol{B}^{-1}\boldsymbol{C})^{-1}. \tag{C.6}$$

which implies

$$(\boldsymbol{A}^{-1}+\boldsymbol{C}'\boldsymbol{B}^{-1}\boldsymbol{C})^{-1}\boldsymbol{C}'\boldsymbol{B}^{-1}\boldsymbol{C} - \mathbf{I}_m = (\boldsymbol{A}^{-1}+\boldsymbol{C}'\boldsymbol{B}^{-1}\boldsymbol{C})^{-1}\boldsymbol{A}^{-1}. \tag{C.7}$$

Then, by Weyl's inequality (Merikoski and Kumar, 2004, Theorem 1):

$$\nu^{(m)}(\boldsymbol{A}^{-1}+\boldsymbol{C}'\boldsymbol{B}^{-1}\boldsymbol{C}) \ge \nu^{(m)}(\boldsymbol{A}^{-1}) + \nu^{(m)}(\boldsymbol{C}'\boldsymbol{B}^{-1}\boldsymbol{C}) = \{\nu^{(1)}(\boldsymbol{A})\}^{-1} + \nu^{(m)}(\boldsymbol{C}'\boldsymbol{B}^{-1}\boldsymbol{C}). \tag{C.8}$$

From (C.8), we have

$$\begin{aligned}\|(\boldsymbol{A}^{-1}+\boldsymbol{C}'\boldsymbol{B}^{-1}\boldsymbol{C})^{-1}\boldsymbol{A}^{-1}\| &\le \|(\boldsymbol{A}^{-1}+\boldsymbol{C}'\boldsymbol{B}^{-1}\boldsymbol{C})^{-1}\|\, \|\boldsymbol{A}^{-1}\|\\ &= \left\{\nu^{(m)}(\boldsymbol{A}^{-1}+\boldsymbol{C}'\boldsymbol{B}^{-1}\boldsymbol{C})\right\}^{-1} \left\{\nu^{(m)}(\boldsymbol{A})\right\}^{-1}\\ &\le \left\{\nu^{(m)}(\boldsymbol{C}'\boldsymbol{B}^{-1}\boldsymbol{C}) + \{\nu^{(1)}(\boldsymbol{A})\}^{-1}\right\}^{-1} \left\{\nu^{(m)}(\boldsymbol{A})\right\}^{-1}. \end{aligned} \tag{C.9}$$

For first term on the rhs of (C.9), the m eigenvalues of $\boldsymbol{C}'\boldsymbol{B}^{-1}\boldsymbol{C}$ are also the m largest non-zero eigenvalues of $\boldsymbol{C}\boldsymbol{C}'\boldsymbol{B}^{-1}$, and the m largest non-zero eigenvalues of $\boldsymbol{C}\boldsymbol{C}'$ are also the m eigenvalues of $\boldsymbol{C}'\boldsymbol{C}$. Therefore, because of Merikoski and Kumar (2004, Theorem 7):

$$n^{-1}\nu^{(m)}(\boldsymbol{C}\boldsymbol{C}'\boldsymbol{B}^{-1}) \ge n^{-1}\nu^{(m)}(\boldsymbol{C}\boldsymbol{C}')\nu^{(m)}(\boldsymbol{B}^{-1}) = n^{-1}\nu^{(m)}(\boldsymbol{C}'\boldsymbol{C})\left\{\nu^{(1)}(\boldsymbol{B})\right\}^{-1}.$$

Thus, by conditions (b) and (c)

$$n\left\{\nu^{(m)}(\boldsymbol{C}'\boldsymbol{B}^{-1}\boldsymbol{C})\right\}^{-1} \le n\left\{\nu^{(m)}(\boldsymbol{C}'\boldsymbol{C})\right\}^{-1}\nu^{(1)}(\boldsymbol{B}) \le \frac{M_B}{\underline{M}_{C_m}}.$$

Moreover, by condition (a), $\nu^{(1)}(\boldsymbol{A}) > 0$ and $\nu^{(1)}(\boldsymbol{A}) \le M_A$, i.e., $\{\nu^{(1)}(\boldsymbol{A})\}^{-1} \ge M_A$, and

$$\nu^{(m)}(\boldsymbol{C}'\boldsymbol{B}^{-1}\boldsymbol{C}) + \{\nu^{(1)}(\boldsymbol{A})\}^{-1} \ge \nu^{(m)}(\boldsymbol{C}'\boldsymbol{B}^{-1}\boldsymbol{C}) \ge \frac{M_B}{\underline{M}_{C_m}}.$$

For the second term on the rhs of (C.9), by condition (a),

$$\left\{\nu^{(m)}(\boldsymbol{A})\right\}^{-1} \le \frac{1}{L_A},$$

for some finite positive real L_A. Hence, from (C.9)

$$n\|(\boldsymbol{A}^{-1} + \boldsymbol{C}'\boldsymbol{B}^{-1}\boldsymbol{C})^{-1}\boldsymbol{A}^{-1}\| \le \frac{M_B}{\underline{M}_{C_m} L_A}.$$

and by using it in (C.7) we complete the proof. □

Lemma C.6. *For any $r \times r$ symmetric and positive definite matrix $\boldsymbol{P}$ with $\|\boldsymbol{P}\| \le M_P$ for some finite positive real M_P, under Assumptions 1 and 2, as $n \to \infty$,*

(i) $n\|(\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n + \boldsymbol{P}^{-1})^{-1}\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n - \mathbf{I}_r\| = O(1)$;
(ii) $n\|(\boldsymbol{\Lambda}_n'(\boldsymbol{\Gamma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n + \boldsymbol{P}^{-1})^{-1}\boldsymbol{\Lambda}_n'(\boldsymbol{\Gamma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n - \mathbf{I}_r\| = O(1)$;
(iii) $n^2\|(\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n + \boldsymbol{P}^{-1})^{-1} - (\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n)^{-1}\| = O(1)$;
(iv) $n^2\|(\boldsymbol{\Lambda}_n'(\boldsymbol{\Gamma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n + \boldsymbol{P}^{-1})^{-1} - (\boldsymbol{\Lambda}_n'(\boldsymbol{\Gamma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n)^{-1}\| = O(1)$.

Proof. Both results follow from Lemma C.5 since: $\boldsymbol{P}$ satisfies condition (a) by assumption, $\boldsymbol{\Sigma}_n^\xi$ satisfies condition (b) because of Assumption 2(a) and $\boldsymbol{\Gamma}_n^\xi$ satisfies condition (b) because of Lemma C.1(v) and Assumption 2(f), and $\boldsymbol{\Lambda}_n'\boldsymbol{\Lambda}_n$ satisfies condition (c) since it is positive definite because of Assumptions 1(a), and, moreover, its eigenvalues are such that $\underline{C}_j \le \liminf_{n\to\infty} n^{-1}\nu^{(j)}(\boldsymbol{\Lambda}_n'\boldsymbol{\Lambda}_n) \le \limsup_{n\to\infty} n^{-1}\nu^{(j)}(\boldsymbol{\Lambda}_n'\boldsymbol{\Lambda}_n) \le \overline{C}_j$, for $j = 1,\dots,r$, as shown in (C.3) in the proof of Lemma C.3.

Turning to part (iii),

$$\begin{aligned} &n^2\|(\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n + \boldsymbol{P}^{-1})^{-1} - (\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n)^{-1}\| \\ &\quad \le n^2\|(\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n + \boldsymbol{P}^{-1})^{-1}(\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n) - \mathbf{I}_r\|\,\|(\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n)^{-1}\| = O(1). \end{aligned}$$

by part (i) and Lemma C.3(iii). This proves part (iii).

Part (iv) is proved in the same way but using Lemma C.3(iv) instead of Lemma C.3(iv). This completes the proof. □

Lemma C.7. *Under Assumptions 1 and 2, as $n, T \to \infty$,*

(i) $n^{-1/2}\|\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\xi}_{nt}\| = O_p(1)$, *uniformly in t;*
(ii) $n^{-1/2}\|\boldsymbol{\Lambda}_n'(\boldsymbol{\Gamma}_n^\xi)^{-1}\boldsymbol{\xi}_{nt}\| = O_p(1)$, *uniformly in t;*
(iii) $n^{-1/2}T^{-1/2}\|\sum_{t=1}^T \boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\xi}_{nt}\| = O_p(1)$;
(iv) $n^{-1/2}T^{-1/2}\,\|\boldsymbol{\Lambda}_n'\boldsymbol{\mathcal{E}}_{nT}'\|_F = O_p(1)$;
(v) $n^{-1/2}T^{-1/2}\|\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\mathcal{E}}_{nT}'\|_F = O_p(1)$;
(vi) $n^{-1/2}T^{-1/2}\|\boldsymbol{\mathcal{E}}_{nT}\|_F = O_p(1)$;
where $\boldsymbol{\mathcal{E}}_{nT} = (\boldsymbol{\xi}_{n1}\cdots\boldsymbol{\xi}_{nT})'$.

Proof. Throughout, let λ_{ij} be the $(i,j)$the entry of $\boldsymbol{\Lambda}_n$. For part (i), we have

$$\begin{aligned} \mathbb{E}\left[\|n^{-1/2}\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\xi}_{nt}\|^2\right] &= \sum_{j=1}^r n^{-1}\mathbb{E}\left[\left(\sum_{i=1}^n \frac{\lambda_{ij}\xi_{it}}{\sigma_i^2}\right)^2\right] \\ &\le r \max_{j=1,\dots,r} n^{-2}\sum_{i=1}^n\sum_{k=1}^n \frac{|\lambda_{ij}|\,|\lambda_{kj}|}{\sigma_i^2\sigma_k^2}\mathbb{E}\left[\xi_{it}\xi_{kt}\right] \\ &\le rn^{-1}M_\lambda^2 C_\xi^2 \sum_{i=1}^n\sum_{k=1}^n |\mathbb{E}[\xi_{it}\xi_{kt}]| \le rM_\lambda^2 C_\xi^2 M_2, \end{aligned} \tag{C.10}$$

where in the third step we used Assumption 1(a) (since $\max_{j=1,\dots,r}|\lambda_{ij}| \le \|\boldsymbol{\lambda}_i\| \le M_\lambda$, for all $i = 1,\dots,n$), and Assumption 2(a), and in the last step we used Lemma C.1(ii). By Chebychev's inequality and since the constants in (C.10) do not depend on t, we prove part (i).

For part (ii), we have

$$\begin{aligned}
\mathbb{E}\left[\|n^{-1/2}\boldsymbol{\Lambda}_n'(\boldsymbol{\Gamma}_n^\xi)^{-1}\boldsymbol{\xi}_{nt}\|^2\right] &= \mathbb{E}\left[n^{-1}\boldsymbol{\xi}_{nt}'(\boldsymbol{\Gamma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n\boldsymbol{\Lambda}_n'(\boldsymbol{\Gamma}_n^\xi)^{-1}\boldsymbol{\xi}_{nt}\right]\\
&= \mathbb{E}\left[n^{-1}\mathrm{tr}\left\{\boldsymbol{\Lambda}_n'(\boldsymbol{\Gamma}_n^\xi)^{-1}\boldsymbol{\xi}_{nt}\boldsymbol{\xi}_{nt}'(\boldsymbol{\Gamma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n\right\}\right]\\
&= n^{-1}\mathrm{tr}\left\{\boldsymbol{\Lambda}_n'(\boldsymbol{\Gamma}_n^\xi)^{-1}\mathbb{E}\left[\boldsymbol{\xi}_{nt}\boldsymbol{\xi}_{nt}'\right](\boldsymbol{\Gamma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n\right\}\\
&= n^{-1}\mathrm{tr}\left\{\boldsymbol{\Lambda}_n'(\boldsymbol{\Gamma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n\right\}\\
&\le rn^{-1}\nu^{(1)}\left(\boldsymbol{\Lambda}_n'(\boldsymbol{\Gamma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n\right)\\
&= rn^{-1}\|\boldsymbol{\Lambda}_n'(\boldsymbol{\Gamma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n\| = O(1),
\end{aligned} \tag{C.11}$$

by Lemma C.3(vi). By Chebychev's inequality and since the constants in (C.11) do not depend on t, we complete the proof of part (ii).

For part (iii), we have

$$\begin{aligned}
\mathbb{E}\left[\left\|n^{-1/2}T^{-1/2}\sum_{t=1}^T\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\xi}_{nt}\right\|^2\right] &= n^{-1}T^{-1}\sum_{j=1}^r\mathbb{E}\left[\left(\sum_{t=1}^T\sum_{i=1}^n\frac{\lambda_{ij}\xi_{it}}{\sigma_i^2}\right)^2\right]\\
&\le r\max_{j=1,\dots,r}n^{-1}T^{-1}\sum_{t=1}^T\sum_{s=1}^T\sum_{i=1}^n\sum_{k=1}^n\frac{|\lambda_{ij}||\lambda_{kj}|}{\sigma_i^2\sigma_k^2}\mathbb{E}\left[\xi_{it}\xi_{ks}\right]\\
&\le rn^{-1}T^{-1}M_\lambda^2C_\xi^2\sum_{t=1}^T\sum_{s=1}^T\sum_{i=1}^n\sum_{k=1}^n|\mathbb{E}[\xi_{it}\xi_{ks}]| \le rM_\lambda^2C_\xi^2M_1,
\end{aligned}$$

by Assumptions 1(a), 2(a), and Lemma C.1(i). By Chebychev's inequality we prove part (iii).

For part (iv), we have

$$\begin{aligned}
\mathbb{E}\left[\left\|n^{-1/2}T^{-1/2}\boldsymbol{\Lambda}_n'\boldsymbol{\mathcal{E}}_{nT}'\right\|_F^2\right] &= n^{-1}T^{-1}\sum_{k=1}^r\sum_{t=1}^T\mathbb{E}\left[\left(\sum_{i=1}^n\lambda_{ik}\xi_{it}\right)^2\right]\\
&\le n^{-1}rM_\lambda^2\max_{t=1,\dots,T}\sum_{i,j=1}^n|\mathbb{E}[\xi_{it}\xi_{jt}]| \le rM_\lambda^2M_2,
\end{aligned}$$

by Assumption 1(a) and Lemma C.1(ii). By Chebychev's inequality we prove part (iv).

Part (v) is proved as parts (iv), but using also Assumption 2(a).

For part (vi), we have

$$\begin{aligned}
\mathbb{E}\left[\left\|n^{-1/2}T^{-1/2}\boldsymbol{\mathcal{E}}_{nT}\right\|_F^2\right] &= n^{-1}T^{-1}\sum_{t=1}^T\sum_{i=1}^n\mathbb{E}[\xi_{it}^2]\\
&\le \max_{t=1,\dots,T}\max_{i=1,\dots,n}\mathbb{E}[\xi_{it}^2] = \max_{i=1,\dots,n}\sigma_i^2 \le C_\xi,
\end{aligned}$$

by Assumption 2(a). By Chebychev's inequality we prove part (vi). This completes the proof. □

Lemma C.8. *Under Assumptions 1, 2, 3, and 6, as $n,T\to\infty$:*

(i) $\sqrt{nT}\|n^{-1}T^{-1}\sum_{t=1}^T\boldsymbol{\Lambda}_n'\boldsymbol{\xi}_{nt}\mathbf{F}_t'\|_F = O_p(1)$;

(ii) $\sqrt{nT}\,\|n^{-3/2}T^{-1}\sum_{t=1}^T\boldsymbol{\Lambda}_n'\boldsymbol{\xi}_{nt}\boldsymbol{\xi}_{nt}'\|_F = O_p(1)$;

(iii) $\sqrt{nT}\,\|n^{-2}T^{-1}\sum_{t=1}^T\boldsymbol{\Lambda}_n'\boldsymbol{\xi}_{nt}\boldsymbol{\xi}_{nt}'\boldsymbol{\Lambda}_n\|_F = O_p(1)$;

(iv) $\sqrt{nT}\|n^{-1}T^{-1}\sum_{t=1}^T\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\xi}_{nt}\mathbf{F}_t'\|_F = O_p(1)$;

(v) $\sqrt{nT}\|n^{-3/2}T^{-1}\sum_{t=1}^T\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\xi}_{nt}\boldsymbol{\xi}_{nt}'\|_F = O_p(1)$;

(vi) $\sqrt{nT}\|n^{-1}T^{-3/2}\sum_{t=1}^T\mathbf{F}_t\boldsymbol{\xi}_{nt}'(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\mathcal{E}}_{nT}\|_F = O_p(1)$;

where $\boldsymbol{\mathcal{E}}_{nT} = (\boldsymbol{\xi}_{n1}\cdots\boldsymbol{\xi}_{nT})'$.

Proof. Throughout, let λ_{ij} be the $(i,j)$the entry of $\mathbf{\Lambda}_n$. For part (i),

$$\mathbb{E}\left[\left\|n^{-1}T^{-1}\sum_{t=1}^T \mathbf{\Lambda}_n'\boldsymbol{\xi}_{nt}\mathbf{F}_t'\right\|_F^2\right] = n^{-2}T^{-2}\sum_{t,s=1}^T\sum_{i,j=1}^n\sum_{k,h=1}^r \lambda_{ih}\lambda_{jh}\mathbb{E}[\xi_{it}\xi_{js}F_{kt}F_{ks}]$$
$$\le n^{-2}T^{-2}M_\lambda^2 r^2 \max_{k,h=1,\dots,r}\sum_{t,s=1}^T\sum_{i,j=1}^n |\mathbb{E}[\xi_{it}\xi_{js}]|\,|\mathbb{E}[F_{kt}F_{ks}]|$$
$$\le n^{-1}T^{-1}M_\lambda^2 r^2 M_1, \tag{C.12}$$

by Lemma C.1(i) and because $|\mathbb{E}[F_{kt}F_{ks}]|\le 1$ by Cauchy-Schwarz inequality and Assumption 6(b). By Chebychev's inequality we prove part (i).

For part (ii),

$$\mathbb{E}\left[\left\|n^{-3/2}T^{-1}\sum_{t=1}^T \mathbf{\Lambda}_n'\boldsymbol{\xi}_{nt}\boldsymbol{\xi}_{nt}'\right\|_F^2\right] = \sum_{k=1}^r\sum_{j=1}^n \mathbb{E}\left[\left|n^{-3/2}T^{-1}\sum_{t=1}^T\sum_{i=1}^n \lambda_{ik}\xi_{it}\xi_{jt}\right|^2\right]$$
$$\le rM_\lambda n^{-1}T^{-1}\max_{j=1,\dots,n}\mathbb{E}\left[\left|n^{-1/2}T^{-1/2}\sum_{t=1}^T\sum_{i=1}^n \xi_{it}\xi_{jt}\right|^2\right]$$
$$\le rM_\lambda n^{-2}T^{-2}\max_{j=1,\dots,n}\sum_{t,s=1}^T\sum_{i,\ell=1}^n |\mathbb{E}[\xi_{it}\xi_{jt}\xi_{\ell s}\xi_{js}]|$$
$$\le rM_\lambda n^{-1}T^{-1}\mathrm{K}_\xi,$$

by Assumptions 1(a) and 2(d). By Chebychev's inequality we prove part (ii).

For part (iii), following the proof of part (ii),

$$\mathbb{E}\left[\left\|n^{-2}T^{-1}\sum_{t=1}^T \mathbf{\Lambda}_n'\boldsymbol{\xi}_{nt}\boldsymbol{\xi}_{nt}'\mathbf{\Lambda}_n\right\|_F^2\right] = \sum_{k,h=1}^r \mathbb{E}\left[\left|n^{-2}T^{-1}\sum_{t=1}^T\sum_{i,j=1}^n \lambda_{ik}\lambda_{jh}\xi_{it}\xi_{jt}\right|^2\right]$$
$$\le rM_\lambda n^{-4}T^{-2}\sum_{t,s=1}^T\sum_{i_1,j_1=1}^n\sum_{i_2,j_2=1}^n |\mathbb{E}[\xi_{i_1t}\xi_{j_1t}\xi_{i_2s}\xi_{j_2s}]|$$
$$\le rM_\lambda n^{-1}T^{-1}\mathrm{K}_\xi,$$

by Assumptions 1(a) and 2(d). By Chebychev's inequality we prove part (iii).

Parts (iv) and (v) are proved as parts (i) and (ii), respectively, but using also Assumption 2(a).

For part (vi)

$$\mathbb{E}\left[\left\|n^{-1}T^{-3/2}\sum_{t=1}^T \mathbf{F}_t\boldsymbol{\xi}_{nt}'(\mathbf{\Sigma}_n^\xi)^{-1}\boldsymbol{\mathcal{E}}_{nT}\right\|_F^2\right] = n^{-2}T^{-3}\sum_{k=1}^r\sum_{s=1}^T \mathbb{E}\left[\left(\sum_{t=1}^T\sum_{i=1}^n F_{kt}\xi_{it}(\sigma_i^2)^{-1}\xi_{is}\right)^2\right]$$
$$\le n^{-2}T^{-3}C_\xi^2\sum_{k=1}^r\sum_{s=1}^T\sum_{t_1,t_2=1}^T\sum_{i,j=1}^n \mathbb{E}[F_{kt_1}F_{kt_2}\xi_{it_1}\xi_{jt_2}\xi_{is}\xi_{js}]$$
$$\le n^{-2}T^{-2}rC_\xi^2\max_{k=1,\dots,r}\max_{t_1,t_2=1,\dots,T}|\mathbb{E}[F_{kt_1}F_{kt_2}]|\max_{s=1,\dots,T}\sum_{t_1,t_2=1}^T\sum_{i,j=1}^n|\mathbb{E}[\xi_{it_1}\xi_{jt_2}\xi_{is}\xi_{js}]|$$
$$\le n^{-1}T^{-1}rC_\xi^2\max_{k=1,\dots,r}\max_{t=1,\dots,T}\mathbb{E}[F_{kt}^2]\,\mathrm{K}_\xi$$
$$\le n^{-1}T^{-1}rC_\xi^2\mathrm{K}_\xi,$$

by Assumptions 2(a) and 2(d), Lemma C.11, and Cauchy-Schwarz inequality jointly with Assumption 6(b). By Chebychev's inequality we prove part (vi). This completes the proof. □

Lemma C.9. *Under Assumptions 1 and 6, for all $n>N_0$: $\|n^{-1/2}\mathbf{\Lambda}_n - n^{-1/2}\mathbf{V}_n^\chi\boldsymbol{\mathcal{S}}(\mathbf{M}_n^\chi)^{1/2}\|=0$, for some $r\times r$ positive diagonal matrix $\boldsymbol{\mathcal{S}}$ independent of n with entries $\mathbb{I}([\mathbf{V}_n^\chi]_{1j}\ge 0)-\mathbb{I}([\mathbf{V}_n^\chi]_{1j}<0)$, $j=1,\dots,r$, and where N_0 is defined in Assumption 1(a).*

Proof. By Assumption 6(b), for all $n \in \mathbb{N}$,

$$n^{-1}\mathbf{\Gamma}_n^\chi = n^{-1}\mathbf{\Lambda}_n\mathbf{\Lambda}_n' = n^{-1}\mathbf{V}_n^\chi\mathbf{M}_n^\chi\mathbf{V}_n^{\chi\prime}. \tag{C.13}$$

First notice that, the r non-zero eigenvalues of $n^{-1}\mathbf{\Gamma}_n^\chi$ are the r eigenvalues of $n^{-1}\mathbf{\Lambda}_n'\mathbf{\Lambda}_n$ and for all $n > N_0$, $\|\mathbf{\Lambda}_n'\mathbf{\Lambda}_n - \mathbf{\Sigma}_\Lambda\| = 0$, by Assumption 1(a). Moreover, $\mathbf{\Sigma}_\Lambda$ is diagonal and positive definite by Assumptions 6(b) and 1(a), respectively. Hence, for all $n > N_0$,

$$n^{-1}\mathbf{M}_n^\chi = \mathbf{\Sigma}_\Lambda \ \text{ and } \ n(\mathbf{M}_n^\chi)^{-1} = \mathbf{\Sigma}_\Lambda^{-1}. \tag{C.14}$$

Furthermore, it must be that the columns of $\mathbf{\Lambda}_n$ span the same space as the columns of $\mathbf{V}_n^\chi$. Since the eigenvectors are normalized and for all $n > N_0$, $\|n^{-1}\mathbf{M}_n^\chi - \mathbf{\Sigma}_\Lambda\| = 0$ by (C.14), there exist two $r \times r$ matrices $\mathbf{K}_{1n}$ and $\mathbf{K}_{2n}$ such that, for all $n > N_0$,

$$n^{-1/2}\mathbf{\Lambda}_n = \mathbf{V}_n^\chi\mathbf{K}_{1n} \ \text{ and } \mathbf{V}_n^\chi = n^{-1/2}\mathbf{\Lambda}_n\mathbf{K}_{2n}. \tag{C.15}$$

Let

$$\mathbf{K}_1 = \lim_{n\to\infty} n^{-1/2}(\mathbf{V}_n^{\chi\prime}\mathbf{V}_n^\chi)^{-1}\mathbf{V}_n^{\chi\prime}\mathbf{\Lambda}_n = \lim_{n\to\infty} n^{-1/2}\mathbf{V}_n^{\chi\prime}\mathbf{\Lambda}_n$$

and

$$\mathbf{K}_2 = \lim_{n\to\infty} n(\mathbf{\Lambda}_n'\mathbf{\Lambda}_n)^{-1}n^{-1/2}\mathbf{\Lambda}_n'\mathbf{V}_n^\chi = \lim_{n\to\infty} \sqrt{n}(\mathbf{\Lambda}_n'\mathbf{\Lambda}_n)^{-1}\mathbf{\Lambda}_n'\mathbf{V}_n^\chi.$$

Then, by linear projection from (C.15) we have that

$$\lim_{n\to\infty} \mathbf{K}_{1n} = \mathbf{K}_1, \tag{C.16}$$

which is positive definite since the columns of $\mathbf{V}_n^\chi$ are linear combinations of the columns of $\mathbf{\Lambda}_n$ so, for all $n > N_0$, $\text{rk}(n^{-1/2}\mathbf{V}_n^{\chi\prime}\mathbf{\Lambda}_n) = \text{rk}(n^{-1}\mathbf{\Lambda}_n'\mathbf{\Lambda}_n) = \text{rk}(\mathbf{\Sigma}_\Lambda) = r$ by Assumption 1(a). Moreover, for all $n > N_0$, $\|\mathbf{K}_1\| \le n^{-1/2}\|\mathbf{\Lambda}_n\|$, which is finite by Lemma C.2.

Similarly, from (C.15) we also have that

$$\lim_{n\to\infty} \mathbf{K}_{2n} = \mathbf{K}_2. \tag{C.17}$$

which exists and is positive definite, since $\|n(\mathbf{\Lambda}_n'\mathbf{\Lambda}_n)^{-1} - n\mathbf{\Sigma}_\Lambda^{-1}\| = 0$, for all $n > N_0$, and $\mathbf{\Sigma}_\Lambda$ is finite and positive definite by Assumption 1(a). Moreover, for all $n > N_0$, $\|\mathbf{K}_2\| \le n\|(\mathbf{\Lambda}_n'\mathbf{\Lambda}_n)^{-1}\|\, n^{-1/2}\|\mathbf{\Lambda}_n\| = \|\mathbf{\Sigma}_\Lambda^{-1}\|n^{-1/2}\|\mathbf{\Lambda}_n\|$, which is finite by Assumption 1(a) and Lemma C.2.

By using (C.15) into the rhs of (C.13), we get, for all $n > N_0$,

$$n^{-2}\mathbf{\Lambda}_n\mathbf{K}_{2n}\mathbf{M}_n^\chi\mathbf{K}_{2n}'\mathbf{\Lambda}_n' = \mathbf{V}_n^\chi\mathbf{M}_n^\chi\mathbf{V}_n^{\chi\prime},$$

which, since eigenvectors are normalized, implies, that, for all $n > N_0$, we can write

$$n^{-2}\mathbf{V}_n^{\chi\prime}\mathbf{\Lambda}_n\mathbf{K}_{2n}\mathbf{M}_n^\chi\mathbf{K}_{2n}'\mathbf{\Lambda}_n'\mathbf{V}_n^\chi = n^{-1}\mathbf{M}_n^\chi. \tag{C.18}$$

From (C.18) we must have $\mathbf{I}_r = \lim_{n\to\infty} n^{-1/2}\mathbf{V}_n^{\chi\prime}\mathbf{\Lambda}_n\mathbf{K}_{2n} = \mathbf{K}_1\mathbf{K}_2$, so, as expected $\mathbf{K}_1 = \mathbf{K}_2^{-1}$ and $\mathbf{K}_2 = \mathbf{K}_1^{-1}$. It follows that, for all $n > N_0$,

$$\mathbf{V}_n^{\chi\prime}\mathbf{\Lambda}_n(\mathbf{\Lambda}_n'\mathbf{\Lambda}_n)^{-1}\mathbf{\Lambda}_n'\mathbf{V}_n^\chi = \mathbf{I}_r. \tag{C.19}$$

Now, from (C.13), we can also write that for all $n > N_0$

$$n^{-1/2}\mathbf{\Lambda}_n\boldsymbol{R}_n = n^{-1/2}\mathbf{V}_n^\chi(\mathbf{M}_n^\chi)^{1/2}, \tag{C.20}$$

for some $r \times r$ matrix $\boldsymbol{R}_n$. Let,

$$\boldsymbol{R} = \lim_{n\to\infty}(\mathbf{\Lambda}_n'\mathbf{\Lambda}_n)^{-1}\mathbf{\Lambda}_n'\mathbf{V}_n^\chi(\mathbf{M}_n^\chi)^{1/2} = \lim_{n\to\infty} n^{-1/2}\mathbf{K}_{2n}(\mathbf{M}_n^\chi)^{1/2} = \mathbf{K}_2 \lim_{n\to\infty} n^{-1/2}(\mathbf{M}_n^\chi)^{1/2}. \tag{C.21}$$

Hence, for all $n > N_0$, $\boldsymbol{R} = \mathbf{K}_2(\mathbf{\Sigma}_\Lambda)^{1/2}$ by (C.14). So $\boldsymbol{R}$ is finite by Assumption 1(a) and since $\mathbf{K}_2$ is finite.

Moreover, from (C.21)

$$
\begin{aligned}
\boldsymbol{R}^{-1} &= \left\{\lim_{n\to\infty}\sqrt{n}(\mathbf{M}_n^{\chi})^{-1/2}\right\}\mathbf{K}_2^{-1} = \left\{\lim_{n\to\infty}\sqrt{n}(\mathbf{M}_n^{\chi})^{-1/2}\right\}\mathbf{K}_1\\
&= \lim_{n\to\infty}\sqrt{n}(\mathbf{M}_n^{\chi})^{-1/2}\mathbf{K}_{1n} = \lim_{n\to\infty}(\mathbf{M}_n^{\chi})^{-1/2}\mathbf{V}_n^{\chi\prime}\boldsymbol{\Lambda}_n.
\end{aligned}\tag{C.22}
$$

From (C.21) and (C.22), and using Assumption 1(a), we also have

$$
\begin{aligned}
\boldsymbol{R}^{-1} &= \lim_{n\to\infty}(\mathbf{M}_n^{\chi})^{-1/2}\mathbf{V}_n^{\chi\prime}\boldsymbol{\Lambda}_n = \lim_{n\to\infty}(\mathbf{M}_n^{\chi})^{-1/2}(\mathbf{M}_n^{\chi})^{-1/2}(\mathbf{M}_n^{\chi})^{1/2}\mathbf{V}_n^{\chi\prime}\boldsymbol{\Lambda}_n(\boldsymbol{\Lambda}_n'\boldsymbol{\Lambda}_n)^{-1}\boldsymbol{\Lambda}_n'\boldsymbol{\Lambda}_n\\
&= \left\{\lim_{n\to\infty} n(\mathbf{M}_n^{\chi})^{-1}\right\}\boldsymbol{R}'\boldsymbol{\Sigma}_{\Lambda}.
\end{aligned}\tag{C.23}
$$

Hence, for all $n > N_0$, by (C.14), $\boldsymbol{R}^{-1} = (\boldsymbol{\Sigma}_{\Lambda})^{-1}\boldsymbol{R}'\boldsymbol{\Sigma}_{\Lambda}$. Thus, $\boldsymbol{R}^{-1}$ is finite by Assumption 1(a) and since $\boldsymbol{R}_2$ is finite. It follows that $\boldsymbol{R}$ is positive definite.

Moreover, $\boldsymbol{R}$ is orthogonal, indeed, from (C.13) and (C.21)

$$
\begin{aligned}
\boldsymbol{R}\boldsymbol{R}' &= \lim_{n\to\infty}(\boldsymbol{\Lambda}_n'\boldsymbol{\Lambda}_n)^{-1}\boldsymbol{\Lambda}_n'\mathbf{V}_n^{\chi}(\mathbf{M}_n^{\chi})^{1/2}(\mathbf{M}_n^{\chi})^{1/2}\mathbf{V}_n^{\chi\prime}\boldsymbol{\Lambda}_n(\boldsymbol{\Lambda}_n'\boldsymbol{\Lambda}_n)^{-1}\\
&= \lim_{n\to\infty}(\boldsymbol{\Lambda}_n'\boldsymbol{\Lambda}_n)^{-1}\boldsymbol{\Lambda}_n'\boldsymbol{\Gamma}_n^{\chi}\boldsymbol{\Lambda}_n(\boldsymbol{\Lambda}_n'\boldsymbol{\Lambda}_n)^{-1}\\
&= \lim_{n\to\infty}(\boldsymbol{\Lambda}_n'\boldsymbol{\Lambda}_n)^{-1}\boldsymbol{\Lambda}_n'\boldsymbol{\Lambda}_n\boldsymbol{\Lambda}_n'\boldsymbol{\Lambda}_n(\boldsymbol{\Lambda}_n'\boldsymbol{\Lambda}_n)^{-1} = \mathbf{I}_r.
\end{aligned}\tag{C.24}
$$

By substituting (C.24) into (C.23), for all $n > N_0$, by (C.14),

$$
\boldsymbol{R}^{-1} = \boldsymbol{\Sigma}_{\Lambda}^{-1}\boldsymbol{R}^{-1}\boldsymbol{\Sigma}_{\Lambda}.\tag{C.25}
$$

By right-multiplying (C.25) by $\boldsymbol{R}$ and left-multiplying by $\boldsymbol{\Sigma}_{\Lambda}$ we have

$$
\boldsymbol{\Sigma}_{\Lambda} = \boldsymbol{R}^{-1}\boldsymbol{\Sigma}_{\Lambda}\boldsymbol{R},
$$

which implies $\boldsymbol{R} = \boldsymbol{\mathcal{J}}$ where $\boldsymbol{\mathcal{J}}$ is an $r\times r$ diagonal matrix with entries ± 1 independent of n. Therefore, from (C.20), for all $n > N_0$,

$$
\boldsymbol{\Lambda}_n\boldsymbol{\mathcal{J}} = n^{-1/2}\mathbf{V}_n^{\chi}(\mathbf{M}_n^{\chi})^{1/2}
$$

or, equivalently, for all $n > N_0$,

$$
n^{-1/2}\boldsymbol{\Lambda}_n = n^{-1/2}\mathbf{V}_n^{\chi}\boldsymbol{\mathcal{J}}(\mathbf{M}_n^{\chi})^{1/2}.
$$

Finally, by Assumption 6(c) it must be that $\boldsymbol{\mathcal{J}} = \boldsymbol{\mathcal{S}}$. This completes the proof. □

Lemma C.10. *Under Assumptions 1, 2, and 3, as $n\to\infty$, $n^{-1/2}\|\mathbf{x}_{nt}\| = O_p(1)$, uniformly in t.*

Proof. We have,

$$
\begin{aligned}
\mathbb{E}\left[\|n^{-1/2}\mathbf{x}_{nt}\|^2\right] &= n^{-1}\sum_{i=1}^{n}\mathbb{E}[x_{it}^2] = n^{-1}\sum_{i=1}^{n}\left\{\boldsymbol{\lambda}_i'\boldsymbol{\Gamma}^F\boldsymbol{\lambda}_i + \sigma_i^2\right\}\\
&\le \max_{i=1,\dots,n}\boldsymbol{\lambda}_i'\boldsymbol{\Gamma}^F\boldsymbol{\lambda}_i + \max_{i=1,\dots,n}\sigma_i^2 \le M_\lambda^2\|\boldsymbol{\Gamma}^F\| \le M_\lambda^2 M_F + C_\xi,
\end{aligned}\tag{C.26}
$$

by Assumption 3, Assumptions 1(a) and 1(b), and Assumption 2(a). The proof of part (i) follows by Chebychev's inequality and by noticing that the bound in (C.26) does not depend on t. This completes the proof. □

Lemma C.11. *Under Assumptions 1 and 3, the processes $\{\xi_{it},\, i\in\mathbb{N},\, t\in\mathbb{Z}\}$ and $\{F_{jt},\, j=1,\dots,r,\, t\in\mathbb{Z}\}$ are mutually independent.*

Proof. It is enough to notice that $\mathbf{F}_t = \sum_{k=0}^{\infty}\mathbf{A}^k\mathbf{H}\mathbf{u}_{t-k}$, then, by Assumption 3, we complete the proof. □

Lemma C.12. *Under Assumptions 1, 2, and 3, as $n, T\to\infty$,*

(i) for $k = 0, 1$, $T\mathbb{E}[\|T^{-1}\sum_{t=1}^T\mathbf{F}_t\mathbf{F}_{t-k}' - \boldsymbol{\Gamma}_k^F\|_F^2] = O(1)$, with $\boldsymbol{\Gamma}_k^F = \mathbb{E}[\mathbf{F}_t\mathbf{F}_{t-k}']$;

(ii) $T\max_{i=1,\dots,n}\mathbb{E}[\|T^{-1}\sum_{t=1}^T\mathbf{F}_t\xi_{it}\|^2] = O(1)$;

(iii) $T\mathbb{E}[\|n^{-1/2}T^{-1}\sum_{t=1}^T\mathbf{F}_t\boldsymbol{\xi}_{nt}\|_F^2] = O(1)$;

(iv) $T\max_{i,j=1,\ldots,n}\mathbb{E}[|T^{-1}\sum_{t=1}^T\xi_{it}\xi_{jt}-\mathbb{E}[\xi_{it}\xi_{jt}]|^2]=O(1)$;
(v) $T\max_{i,j=1,\ldots,n}\mathbb{E}[|T^{-1}\sum_{t=1}^T x_{it}x_{jt}-\mathbb{E}[x_{it}x_{jt}]|^2]=O(1)$;
(vi) $T\mathbb{E}[\|n^{-1}T^{-1}\sum_{t=1}^T\boldsymbol{\xi}_{nt}\boldsymbol{\xi}_{nt}'-n^{-1}\boldsymbol{\Gamma}_n^{\xi}\|_F^2]=O(1)$;
(vii) $T\mathbb{E}[\|n^{-1}T^{-1}\sum_{t=1}^T\mathbf{x}_{nt}\mathbf{x}_{nt}'-n^{-1}\boldsymbol{\Gamma}_n^{x}\|_F^2]=O(1)$.

Proof. Part (i) follows since $\{\mathbf{F}_t\}$ is ergodic, because of Assumption 1(d) which implies that $\{\mathbf{F}_t\}$ has summable autocovariances, and therefore $\{\mathbf{F}_t\mathbf{F}_{t-k}'\}$ is also ergodic (White, 2001, Theorem 3.35, and Stout, 1974, pp. 170, 182). In particular, $\boldsymbol{\Gamma}_k^F$ is finite because of Assumptions 1(b), 1(d) and 1(e), and $\mathbb{E}[\|\mathbf{F}_t\|^4\|]$ is also finite because of Assumptions 1(d)-1(g). See also Hamilton (1994, Proposition 11.1, pp. 298-299), which can be applied using the fact that $\{\mathbf{v}_t\}$ is an independent process by Assumption 1(f) and thus it is a martingale difference process. This proves part (i).

For part (ii), as $T\to\infty$,

$$\mathbb{E}\left[\left\|T^{-1}\sum_{t=1}^T\mathbf{F}_t\xi_{it}\right\|^2\right]=T^{-2}\sum_{j=1}^r\mathbb{E}\left[\left(\sum_{t=1}^T F_{jt}\xi_{it}\right)^2\right]=T^{-2}\sum_{j=1}^r\sum_{t,s=1}^T\mathbb{E}\left[F_{jt}\xi_{it}F_{js}\xi_{is}\right] \tag{C.27}$$
$$\le T^{-2}\sum_{j=1}^r\sum_{t,s=1}^T\left|\mathbb{E}\left[F_{jt}F_{js}\right]\right|\left|\mathbb{E}\left[\xi_{it}\xi_{is}\right]\right|$$
$$\le T^{-2}\sum_{j=1}^r\sum_{t,s=1}^T\mathbb{E}[F_{jt}^2]\left|\mathbb{E}\left[\xi_{it}\xi_{is}\right]\right|\le T^{-1}rM_FM_3,$$

because of Lemma C.1(iii), and where we also used Lemma C.11, Cauchy-Schwarz inequality, and the fact that F_{jt} is weakly stationary by Assumptions 1(b), 1(d), and 1(e). By noticing that the constants on the rhs of (C.27) do not depend on i we prove part (ii). Part (iii) follows directly from part (ii), indeed

$$\mathbb{E}\left[\left\|n^{-1/2}T^{-1}\sum_{t=1}^T\mathbf{F}_t\boldsymbol{\xi}_{nt}\right\|_F^2\right]=n^{-1}T^{-2}\sum_{j=1}^r\sum_{i=1}^n\mathbb{E}\left[\left(\sum_{t=1}^T F_{jt}\xi_{it}\right)^2\right]$$
$$\le T^{-2}\sum_{j=1}^r\max_{i=1,\ldots,n}\mathbb{E}\left[\left(\sum_{t=1}^T F_{jt}\xi_{it}\right)^2\right]\le T^{-1}rM_FM_3.$$

For part (iv), notice that, since $\{\xi_{it}\}$ is a strongly mixing process with exponentially decaying coefficients, because of Assumption 2(c), then it is also ergodic (White, 2001, Proposition 3.44, and Rosenblatt, 1972), so that also $\{\xi_{it}\xi_{jt}\}$ is ergodic (White, 2001, Theorem 3.35, and Stout, 1974, pp. 170, 182). In particular, $\mathbb{E}[\xi_{it}\xi_{jt}]$ is finite by Assumption 2(a) and $\mathbb{E}[|\xi_{it}\xi_{jt}\xi_{is}\xi_{js}|]$ is also finite because by Assumption 2(d), and both are bounded by constant independent of i and j.

This proves part (iv). Part (v) follows directly from parts (i), (ii), and (iv). Part (vi) follows directly from part (iv), and part (vii) follows directly from part (v). This completes the proof. □

Lemma C.13. *Under Assumptions 1, 2, and 3, as $n,T\to\infty$, $\|(T^{-1}\sum_{t=1}^T\mathbf{F}_t\mathbf{F}_t')^{-1}\|=O_p(1)$.*

Proof. From Lemma C.12(i), and Merikoski and Kumar (2004, Theorem 1) which is Weyl's inequality

$$\left|\nu^{(r)}\left(T^{-1}\sum_{t=1}^T\mathbf{F}_t\mathbf{F}_t'\right)-\nu^{(r)}(\boldsymbol{\Gamma}^F)\right|\le\left\|T^{-1}\sum_{t=1}^T\mathbf{F}_t\mathbf{F}_t'-\boldsymbol{\Gamma}^F\right\|=O_p(T^{-1/2}).$$

This implies (note that $x-y\ge-|x-y|$ for any $x,y\in\mathbb{R}$)

$$\det\left(T^{-1}\sum_{t=1}^T\mathbf{F}_t\mathbf{F}_t'\right)=\prod_{j=1}^r\nu^{(j)}\left(T^{-1}\sum_{t=1}^T\mathbf{F}_t\mathbf{F}_t'\right)\ge\left\{\nu^{(r)}\left(T^{-1}\sum_{t=1}^T\mathbf{F}_t\mathbf{F}_t'\right)\right\}^r$$
$$\ge\left\{\nu^{(r)}(\boldsymbol{\Gamma}^F)-\left|\nu^{(r)}\left(T^{-1}\sum_{t=1}^T\mathbf{F}_t\mathbf{F}_t'\right)-\nu^{(r)}(\boldsymbol{\Gamma}^F)\right|\right\}^r>0,$$

by Assumption 1(b). Thus, $\|(T^{-1}\sum_{t=1}^T\mathbf{F}_t\mathbf{F}_t')^{-1}\|=O_p(1)$. This completes the proof. □

D Lemmas necessary for proving Proposition 1

Lemma D.1. *Consider the initial estimator of the loadings* $\widehat{\boldsymbol{\Lambda}}_n^{(0)} = (\widehat{\boldsymbol{\lambda}}_1^{(0)} \cdots \widehat{\boldsymbol{\lambda}}_n^{(0)})'$ *defined in Section A.1, then, under Assumptions 1, 2, 3, and 6, as* $n, T \to \infty$*:*
(i) $\min(n, \sqrt{T})\, \|\widehat{\boldsymbol{\lambda}}_i^{(0)} - \boldsymbol{\lambda}_i\| = O_p(1)$*, uniformly in i;*
(ii) $\min(n, \sqrt{T})\, n^{-1/2} \|\widehat{\boldsymbol{\Lambda}}_n^{(0)} - \boldsymbol{\Lambda}_n\| = O_p(1)$.

Proof. Both results are direct consequences of Barigozzi (2023, Theorem 1), see also Bai (2003, Theorem 2) under similar assumptions. This completes the proof. □

Lemma D.2. *Consider the initial estimator of the factors* $\widetilde{\mathbf{F}}_t$ *defined in Section A.1, then, under Assumptions 1, 2, 3, and 6, as* $n, T \to \infty$*:*
(i) for $k = 0, 1$, $\min(n^{-1}, T^{-1/2})\, \|T^{-1} \sum_{t=k+1}^T (\widetilde{\mathbf{F}}_t - \mathbf{F}_t)\mathbf{F}'_{t-k}\| = O_p(1)$*;*
(ii) $\min(n^{-1}, T^{-1/2})\, \|n^{-1/2} T^{-1} \sum_{t=1}^T (\widetilde{\mathbf{F}}_t - \mathbf{F}_t)\boldsymbol{\xi}'_{nt}\| = O_p(1)$*;*
(iii) $\min(n^{-1}, T^{-1/2})\, \|n^{-1} T^{-1} \sum_{t=1}^T (\widetilde{\mathbf{F}}_t - \mathbf{F}_t)\boldsymbol{\xi}'_{nt}\boldsymbol{\Lambda}_n\| = O_p(1)$*;*
(iv) $\min(n^{-1}, T^{-1/2})\, \|T^{-1} \sum_{t=1}^T (\widetilde{\mathbf{F}}_t - \mathbf{F}_t)\xi_{it}\| = O_p(1)$*, uniformly in i;*
(v) $\min(n^{-1}, T^{-1/2})\, \|T^{-1} \sum_{t=1}^T (\widetilde{\mathbf{F}}_t - \mathbf{F}_t)\| = O_p(1)$.

Proof. For part (i), by definition

$$\begin{aligned}
\widetilde{\mathbf{F}}_t - \mathbf{F}_t &= (\widehat{\boldsymbol{\Lambda}}_n^{(0)\prime}\widehat{\boldsymbol{\Lambda}}_n^{(0)})^{-1}\widehat{\boldsymbol{\Lambda}}_n^{(0)\prime}\mathbf{x}_{nt} - \mathbf{F}_t \\
&= \left\{(\widehat{\boldsymbol{\Lambda}}_n^{(0)\prime}\widehat{\boldsymbol{\Lambda}}_n^{(0)})^{-1}\widehat{\boldsymbol{\Lambda}}_n^{(0)\prime}\boldsymbol{\Lambda}_n - \mathbf{I}_r\right\}\mathbf{F}_t + \left\{(\widehat{\boldsymbol{\Lambda}}_n^{(0)\prime}\widehat{\boldsymbol{\Lambda}}_n^{(0)})^{-1}\widehat{\boldsymbol{\Lambda}}_n^{(0)\prime} - (\boldsymbol{\Lambda}_n'\boldsymbol{\Lambda}_n)^{-1}\boldsymbol{\Lambda}_n'\right\}\boldsymbol{\xi}_{nt} \\
&\quad + (\boldsymbol{\Lambda}_n'\boldsymbol{\Lambda}_n)^{-1}\boldsymbol{\Lambda}_n'\boldsymbol{\xi}_{nt}. \qquad \text{(D.1)}
\end{aligned}$$

Using (D.1) the first term on the rhs of (D.4) is such that

$$\begin{aligned}
\left\|T^{-1}\sum_{t=1}^T(\widetilde{\mathbf{F}}_t - \mathbf{F}_t)\mathbf{F}_t'\right\| \le& \left\|T^{-1}\sum_{t=1}^T\left\{(\widehat{\boldsymbol{\Lambda}}_n^{(0)\prime}\widehat{\boldsymbol{\Lambda}}_n^{(0)})^{-1}\widehat{\boldsymbol{\Lambda}}_n^{(0)\prime}\boldsymbol{\Lambda}_n - \mathbf{I}_r\right\}\mathbf{F}_t\mathbf{F}_t'\right\| \\
&+ \left\|T^{-1}\sum_{t=1}^T\sqrt{n}\left\{(\widehat{\boldsymbol{\Lambda}}_n^{(0)\prime}\widehat{\boldsymbol{\Lambda}}_n^{(0)})^{-1}\widehat{\boldsymbol{\Lambda}}_n^{(0)\prime} - (\boldsymbol{\Lambda}_n'\boldsymbol{\Lambda}_n)^{-1}\boldsymbol{\Lambda}_n'\right\}n^{-1/2}\boldsymbol{\xi}_{nt}\mathbf{F}_t'\right\| \\
&+ \left\|T^{-1}\sum_{t=1}^T(\boldsymbol{\Lambda}_n'\boldsymbol{\Lambda}_n)^{-1}\boldsymbol{\Lambda}_n'\boldsymbol{\xi}_{nt}\mathbf{F}_t'\right\| \\
\le& \left\{\left\|T^{-1}\sum_{t=1}^T\mathbf{F}_t\mathbf{F}_t'\right\| + \left\|n^{-1/2}T^{-1}\sum_{t=1}^T\boldsymbol{\xi}_{nt}\mathbf{F}_t'\right\|\right\}O_p(\max(n^{-1}, T^{-1/2})) \\
&+ \|n(\boldsymbol{\Lambda}_n'\boldsymbol{\Lambda}_n)^{-1}\|\left\|n^{-1}T^{-1}\sum_{t=1}^T\boldsymbol{\Lambda}_n'\boldsymbol{\xi}_{nt}\mathbf{F}_t'\right\| \\
=& O_p(\max(n^{-1}, T^{-1/2})) + O_p(n^{-1/2}T^{-1/2}), \qquad \text{(D.2)}
\end{aligned}$$

by Lemma D.1(ii), which does not depend on t, and Lemmas C.2 (jointly with Assumption 1(a)), C.12(i), and C.12(iii), and part (i). The case $k = 1$ is proved in the same way and this proves part (iii).

For part (ii), as in part (i), we have

$$\begin{aligned}
&\left\|n^{-1/2}T^{-1}\sum_{t=1}^T(\widetilde{\mathbf{F}}_t - \mathbf{F}_t)\boldsymbol{\xi}_{nt}'\right\| \\
&\le \left\{\left\|n^{-1/2}T^{-1}\sum_{t=1}^T\mathbf{F}_t\boldsymbol{\xi}_{nt}'\right\| + \left\|n^{-1}T^{-1}\sum_{t=1}^T\boldsymbol{\xi}_{nt}\boldsymbol{\xi}_{nt}'\right\|\right\}O_p(\max(n^{-1}, T^{-1/2})) \\
&\quad + \|n(\boldsymbol{\Lambda}_n'\boldsymbol{\Lambda}_n)^{-1}\|\left\|n^{-3/2}T^{-1}\sum_{t=1}^T\boldsymbol{\Lambda}_n'\boldsymbol{\xi}_{nt}\boldsymbol{\xi}_{nt}'\right\| \\
&= O_p(\max(n^{-1}, T^{-1/2})) + O_p(n^{-1/2}T^{-1/2}),
\end{aligned}$$

by Lemma D.1(ii), which does not depend on t, and Lemmas C.2 (jointly with Assumption 1(a)), C.12(iii), and C.12(vi), and Lemma C.8(ii). This proves part (ii).

Part (iii) follows directly from part (ii) but using Lemma C.8(iii).

For part (iv), as in part (i), we have

$$\begin{aligned}\left\| T^{-1}\sum_{t=1}^T(\widetilde{\mathbf{F}}_t-\mathbf{F}_t)\xi_{it}\right\| \le &\left\{\left\|T^{-1}\sum_{t=1}^T\mathbf{F}_t\xi_{it}\right\|+\left\|n^{-1/2}T^{-1}\sum_{t=1}^T\boldsymbol{\xi}_{nt}\xi_{it}\right\|\right\}O_p(\max(n^{-1},T^{-1/2}))\\ &+\|n(\boldsymbol{\Lambda}_n'\boldsymbol{\Lambda}_n)^{-1}\|\left\|n^{-1}T^{-1}\sum_{t=1}^T\boldsymbol{\Lambda}_n'\boldsymbol{\xi}_{nt}\xi_{it}\right\|\\ =&O_p(\max(n^{-1},T^{-1/2}))+O_p(n^{-1/2}T^{-1/2}),\end{aligned}$$

by the same arguments used to prove part (ii) and noticing that now $|\xi_{it}|=O_p(1)$ by Assumption 2(a). Part (v) is proved as part (iv) setting $\xi_{it}=1$. This completes the proof. □

Lemma D.3. *Consider the initial estimator of the VAR parameters $\widehat{\mathbf{A}}^{(0)}$ and $\widehat{\boldsymbol{\Gamma}}^{v(0)}$ defined in Section A.1, then, under Assumptions 1, 2, 3, and 6, as $n,T\to\infty$:*

(i) $\min(n,\sqrt{T})\,\|\widehat{\mathbf{A}}^{(0)}-\mathbf{A}\|=O_p(1)$*;*

(ii) $\min(n,\sqrt{T})\,\|\widehat{\boldsymbol{\Gamma}}^{v(0)}-\boldsymbol{\Gamma}^v\|=O_p(1)$*.*

Proof. For part (i), in agreement with Assumption 1(i), we can always set $\widetilde{\mathbf{F}}_0=\mathbf{0}_r$, so it follows that, by construction $T^{-1}\sum_{t=2}^T\widetilde{\mathbf{F}}_{t-1}\widetilde{\mathbf{F}}_{t-1}'=T^{-1}\sum_{t=1}^T\widetilde{\mathbf{F}}_{t-1}\widetilde{\mathbf{F}}_{t-1}'=\mathbf{I}_r$. Then, by Assumption 6(b), we have

$$\begin{aligned}\widehat{\mathbf{A}}^{(0)}-\mathbf{A}&=T^{-1}\sum_{t=2}^T\widetilde{\mathbf{F}}_t\widetilde{\mathbf{F}}_{t-1}'-\boldsymbol{\Gamma}_1^F\\ &=\left\{T^{-1}\sum_{t=2}^T\widetilde{\mathbf{F}}_t\widetilde{\mathbf{F}}_{t-1}'-T^{-1}\sum_{t=2}^T\mathbf{F}_t\mathbf{F}_{t-1}'\right\}+\left\{T^{-1}\sum_{t=2}^T\mathbf{F}_t\mathbf{F}_{t-1}'-\boldsymbol{\Gamma}_1^F\right\}.\end{aligned}\tag{D.3}$$

Now,

$$\begin{aligned}T^{-1}\sum_{t=2}^T\widetilde{\mathbf{F}}_t\widetilde{\mathbf{F}}_{t-1}'-T^{-1}\sum_{t=2}^T\mathbf{F}_t\mathbf{F}_{t-1}'=&T^{-1}\sum_{t=2}^T(\widetilde{\mathbf{F}}_t-\mathbf{F}_t)\mathbf{F}_{t-1}'+T^{-1}\sum_{t=2}^T\mathbf{F}_t(\widetilde{\mathbf{F}}_{t-1}-\mathbf{F}_{t-1})'\\ &+T^{-1}\sum_{t=2}^T(\widetilde{\mathbf{F}}_t-\mathbf{F}_t)(\widetilde{\mathbf{F}}_{t-1}-\mathbf{F}_{t-1})'.\end{aligned}\tag{D.4}$$

Now, by Lemma D.2(i) the first and second term in the rhs of (D.4) are $O_p(\max(n^{-1},T^{-1/2}))$, while the third term on the rhs is dominated by the first two. Hence, for the first term on the rhs of (D.3) we have

$$\left\|T^{-1}\sum_{t=2}^T\widetilde{\mathbf{F}}_t\widetilde{\mathbf{F}}_{t-1}'-T^{-1}\sum_{t=2}^T\mathbf{F}_t\mathbf{F}_{t-1}'\right\|=O_p(\max(n^{-1},T^{-1/2})).$$

The second term on the rhs of (D.3) is $O_p(T^{-1/2})$, by Lemma C.12(i). This proves part (i).

Part (ii) follows from the results in Forni et al. (2009, Proposition P) combined with part (i). This completes the proof. □

Lemma D.4. *Consider the initial estimator of the idiosyncratic variances $\widehat{\sigma}_i^{2(0)}$, $i=1,\dots,n$, defined in Section A.1, then, under Assumptions 1, 2, 3, and 6, as $n,T\to\infty$:*

(i) $\min(n,\sqrt{T})\,|\widehat{\sigma}_i^{(0)2}-\sigma_i^2|=O_p(1)$*, uniformly in i;*

(ii) $\min(n,\sqrt{T})\,n^{-1}|\sum_{i=1}^n(\widehat{\sigma}_i^{(0)2}-\sigma_i^2)|=O_p(1)$.

Proof. Start from

$$
\begin{aligned}
(\widehat{\sigma}_i^{2(0)} - \sigma_i^2) =& T^{-1}\sum_{t=1}^T (x_{it} - \widehat{\boldsymbol{\lambda}}_i^{(0)\prime}\widetilde{\mathbf{F}}_t)^2 - \mathbb{E}[(x_{it} - \boldsymbol{\lambda}_i'\mathbf{F}_t)^2]\\
=& \left\{T^{-1}\sum_{t=1}^T x_{it}^2 - \mathbb{E}[x_{it}^2]\right\} + \left\{\widehat{\boldsymbol{\lambda}}_i^{(0)\prime}\widehat{\boldsymbol{\lambda}}_i^{(0)} - \boldsymbol{\lambda}_i'\boldsymbol{\lambda}_i\right\}\\
&- 2\left\{T^{-1}\sum_{t=1}^T \boldsymbol{\lambda}_i'\mathbf{F}_t\widetilde{\mathbf{F}}_t'\widehat{\boldsymbol{\lambda}}_i^{(0)} - \boldsymbol{\lambda}_i'\boldsymbol{\lambda}_i\right\} - 2\left\{T^{-1}\sum_{t=1}^T \xi_{it}\widetilde{\mathbf{F}}_t'\widehat{\boldsymbol{\lambda}}_i^{(0)}\right\},
\end{aligned}
\tag{D.5}
$$

since $\boldsymbol{\Gamma}^F = \mathbf{I}_r$ by Assumption 6(b), $\mathbb{E}[\mathbf{F}_t\xi_{it}] = \mathbf{0}_r$ by Lemma C.11, and $T^{-1}\sum_{t=1}^T \widetilde{\mathbf{F}}_t\widetilde{\mathbf{F}}_t' = \mathbf{I}_r$ by construction. For part (i), from (D.5) we have

$$
\begin{aligned}
|\widehat{\sigma}_i^{2(0)} - \sigma_i^2| \le& \left|T^{-1}\sum_{t=1}^T x_{it}^2 - \mathbb{E}[x_{it}^2]\right| + \left|\widehat{\boldsymbol{\lambda}}_i^{(0)\prime}\widehat{\boldsymbol{\lambda}}_i^{(0)} - \boldsymbol{\lambda}_i'\boldsymbol{\lambda}_i\right|\\
&+ 2\left|T^{-1}\sum_{t=1}^T \boldsymbol{\lambda}_i'\mathbf{F}_t\mathbf{F}_t'\widehat{\boldsymbol{\lambda}}_i^{(0)} - \boldsymbol{\lambda}_i'\boldsymbol{\lambda}_i\right| + 2\left|T^{-1}\sum_{t=1}^T \boldsymbol{\lambda}_i'\mathbf{F}_t(\widetilde{\mathbf{F}}_t - \mathbf{F}_t)'\widehat{\boldsymbol{\lambda}}_i^{(0)}\right|\\
&+ 2\left|T^{-1}\sum_{t=1}^T \xi_{it}\mathbf{F}_t'\widehat{\boldsymbol{\lambda}}_i^{(0)}\right| + 2\left|T^{-1}\sum_{t=1}^T \xi_{it}(\widetilde{\mathbf{F}}_t - \mathbf{F}_t)'\widehat{\boldsymbol{\lambda}}_i^{(0)}\right|\\
\le& \left|T^{-1}\sum_{t=1}^T x_{it}^2 - \mathbb{E}[x_{it}^2]\right| + \left|\widehat{\boldsymbol{\lambda}}_i^{(0)\prime}\widehat{\boldsymbol{\lambda}}_i^{(0)} - \boldsymbol{\lambda}_i'\boldsymbol{\lambda}_i\right|\\
&+ 2\|\boldsymbol{\lambda}_i\|\,\|\widehat{\boldsymbol{\lambda}}_i^{(0)} - \boldsymbol{\lambda}_i\| + 2\|\boldsymbol{\lambda}_i\|\left\|T^{-1}\sum_{t=1}^T \mathbf{F}_t\mathbf{F}_t' - \mathbf{I}_r\right\|\{\|\widehat{\boldsymbol{\lambda}}_i^{(0)} - \boldsymbol{\lambda}_i\| + \|\boldsymbol{\lambda}_i\|\}\\
&+ 2\|\boldsymbol{\lambda}_i\|\left\|T^{-1}\sum_{t=1}^T \mathbf{F}_t(\widetilde{\mathbf{F}}_t - \mathbf{F}_t)'\right\|\{\|\widehat{\boldsymbol{\lambda}}_i^{(0)} - \boldsymbol{\lambda}_i\| + \|\boldsymbol{\lambda}_i\|\}\\
&+ 2\left\|T^{-1}\sum_{t=1}^T \xi_{it}\mathbf{F}_t'\right\|\|\boldsymbol{\lambda}_i\| + 2\left\|T^{-1}\sum_{t=1}^T \xi_{it}\mathbf{F}_t'\right\|\|\widehat{\boldsymbol{\lambda}}_i^{(0)} - \boldsymbol{\lambda}_i\|\\
&+ 2\left\|T^{-1}\sum_{t=1}^T \xi_{it}(\widetilde{\mathbf{F}}_t - \mathbf{F}_t)'\right\|\{\|\widehat{\boldsymbol{\lambda}}_i^{(0)} - \boldsymbol{\lambda}_i\| + \|\boldsymbol{\lambda}_i\|\}\\
=& O_p(\max(n^{-1}, T^{-1/2})),
\end{aligned}
\tag{D.6}
$$

where, we used multiple times Assumption 1(a) and Lemma D.1(i), and, we also used: Lemma C.12(v) for the first term, Lemma C.12(i) for the fourth term, Lemma D.2(i) for the fifth term, Lemma C.12(ii) for the sixth and seventh term, and Lemma D.2(iv) for the last term. This proves part (i).

As for part (ii), from (D.5) we have

$$
\begin{aligned}
n^{-1}\left|\sum_{i=1}^n(\widehat\sigma_i^{2(0)}-\sigma_i^2)\right| \le\, & n^{-1}\left|\sum_{i=1}^n\left\{T^{-1}\sum_{t=1}^T x_{it}^2-\mathbb{E}[x_{it}^2]\right\}\right| + n^{-1}\left|\sum_{i=1}^n\{\widehat{\boldsymbol\lambda}_i^{(0)\prime}\widehat{\boldsymbol\lambda}_i^{(0)}-\boldsymbol\lambda_i'\boldsymbol\lambda_i\}\right|\\
&+2n^{-1}\left|\sum_{i=1}^n\left\{T^{-1}\sum_{t=1}^T\boldsymbol\lambda_i'\mathbf{F}_t\mathbf{F}_t'\widehat{\boldsymbol\lambda}_i^{(0)}-\boldsymbol\lambda_i'\boldsymbol\lambda_i\right\}\right|\\
&+2n^{-1}\left|\sum_{i=1}^n\left\{T^{-1}\sum_{t=1}^T\boldsymbol\lambda_i'\mathbf{F}_t(\widetilde{\mathbf{F}}_t-\mathbf{F}_t)'\widehat{\boldsymbol\lambda}_i^{(0)}\right\}\right|\\
&+2n^{-1}\left|\sum_{i=1}^n\left\{T^{-1}\sum_{t=1}^T\xi_{it}\mathbf{F}_t'\right\}\widehat{\boldsymbol\lambda}_i^{(0)}\right| + 2n^{-1}\left|\sum_{i=1}^n\left\{T^{-1}\sum_{t=1}^T\xi_{it}(\widetilde{\mathbf{F}}_t-\mathbf{F}_t)'\right\}\widehat{\boldsymbol\lambda}_i^{(0)}\right|\\
=\, & n^{-1}\left|\sum_{i=1}^n\left\{T^{-1}\sum_{t=1}^T x_{it}^2-\mathbb{E}[x_{it}^2]\right\}\right| + n^{-1}\left|\sum_{i=1}^n\{\widehat{\boldsymbol\lambda}_i^{(0)\prime}\widehat{\boldsymbol\lambda}_i^{(0)}-\boldsymbol\lambda_i'\boldsymbol\lambda_i\}\right|\\
&+2n^{-1/2}\|\boldsymbol\Lambda_n\|\,n^{-1/2}\left\|\widehat{\boldsymbol\Lambda}_n^{(0)}-\boldsymbol\Lambda_n\right\| + 2n^{-1}\|\boldsymbol\Lambda_n'\boldsymbol\Lambda_n\|\left\|T^{-1}\sum_{t=1}^T\mathbf{F}_t\mathbf{F}_t'-\mathbf{I}_r\right\|\\
&+2n^{-1/2}\|\boldsymbol\Lambda_n\|\,n^{-1/2}\left\|\widehat{\boldsymbol\Lambda}_n^{(0)}-\boldsymbol\Lambda_n\right\|\left\|T^{-1}\sum_{t=1}^T\mathbf{F}_t\mathbf{F}_t'-\mathbf{I}_r\right\|\\
&+2n^{-1}\|\boldsymbol\Lambda_n'\boldsymbol\Lambda_n\|\left\|T^{-1}\sum_{t=1}^T\mathbf{F}_t(\widetilde{\mathbf{F}}_t-\mathbf{F}_t)'\right\|\\
&+2n^{-1/2}\|\boldsymbol\Lambda_n\|\,n^{-1/2}\|\widehat{\boldsymbol\Lambda}_n^{(0)}-\boldsymbol\Lambda_n\|\left\|T^{-1}\sum_{t=1}^T\mathbf{F}_t(\widetilde{\mathbf{F}}_t-\mathbf{F}_t)'\right\|\\
&+2n^{-1}\left|T^{-1}\sum_{t=1}^T\mathbf{F}_t'\boldsymbol\Lambda_n'\boldsymbol\xi_{nt}\right| + 2n^{-1}\left|T^{-1}\sum_{t=1}^T\mathbf{F}_t'(\widehat{\boldsymbol\Lambda}_n^{(0)}-\boldsymbol\Lambda_n)'\boldsymbol\xi_{nt}\right|\\
&+2n^{-1}\left|T^{-1}\sum_{t=1}^T(\widetilde{\mathbf{F}}_t-\mathbf{F}_t)'\boldsymbol\Lambda_n'\boldsymbol\xi_{nt}\right|\\
&+2n^{-1}\left|T^{-1}\sum_{t=1}^T(\widetilde{\mathbf{F}}_t-\mathbf{F}_t)'(\widehat{\boldsymbol\Lambda}_n^{(0)}-\boldsymbol\Lambda_n)'\boldsymbol\xi_{nt}\right|\\
=\, & O_p(\max(n^{-1},T^{-1/2})).
\end{aligned}\tag{D.7}
$$

The result in (D.7) follows from repeated use of Lemmas C.2, C.12(i), D.1(ii), and D.2(i) as well as the following results. First,

$$
n^{-1}\left|\sum_{i=1}^n\left\{T^{-1}\sum_{t=1}^T x_{it}^2-\mathbb{E}[x_{it}^2]\right\}\right| \le n^{-1}\left\|T^{-1}\sum_{t=1}^T\mathbf{x}_{nt}\mathbf{x}_{nt}'-\mathbb{E}[\mathbf{x}_{nt}\mathbf{x}_{nt}']\right\|_F = O_p(T^{-1/2}),
$$

by Lemma C.12(vii). Second,

$$
n^{-1}\left|T^{-1}\sum_{t=1}^T\mathbf{F}_t'\boldsymbol\Lambda_n'\boldsymbol\xi_{nt}\right| \le n^{-1}\left\|T^{-1}\sum_{t=1}^T\boldsymbol\Lambda_n'\boldsymbol\xi_{nt}\mathbf{F}_t'\right\|_F = O_p(n^{-1/2}T^{-1/2}),
$$

by Lemma C.8(i). Third,

$$
\begin{aligned}
n^{-1}\left|T^{-1}\sum_{t=1}^T(\widetilde{\mathbf{F}}_t-\mathbf{F}_t)'\boldsymbol\Lambda_n'\boldsymbol\xi_{nt}\right| &\le n^{-1}\left\|T^{-1}\sum_{t=1}^T\boldsymbol\Lambda_n'\boldsymbol\xi_{nt}(\widetilde{\mathbf{F}}_t-\mathbf{F}_t)'\right\|_F\\
&\le\sqrt{r}n^{-1}\left\|T^{-1}\sum_{t=1}^T\boldsymbol\Lambda_n'\boldsymbol\xi_{nt}(\widetilde{\mathbf{F}}_t-\mathbf{F}_t)'\right\|\\
&= O_p(\max(n^{-1},T^{-1/2})),
\end{aligned}
$$

by Lemma D.2(iii). Fourth,

$$
\begin{aligned}
n^{-1}\left|T^{-1}\sum_{t=1}^T \mathbf{F}_t'(\widehat{\mathbf{\Lambda}}_n^{(0)}-\mathbf{\Lambda}_n)'\boldsymbol{\xi}_{nt}\right| &\le n^{-1}\left\|T^{-1}\sum_{t=1}^T(\widehat{\mathbf{\Lambda}}_n^{(0)}-\mathbf{\Lambda}_n)'\boldsymbol{\xi}_{nt}\mathbf{F}_t'\right\|_F\\
&\le \sqrt{r}n^{-1}\left\|T^{-1}\sum_{t=1}^T(\widehat{\mathbf{\Lambda}}_n^{(0)}-\mathbf{\Lambda}_n)'\boldsymbol{\xi}_{nt}\mathbf{F}_t'\right\|\\
&\le \sqrt{r}n^{-1/2}\|\widehat{\mathbf{\Lambda}}_n^{(0)}-\mathbf{\Lambda}_n\|\, n^{-1/2}\left\|T^{-1}\sum_{t=1}^T\boldsymbol{\xi}_{nt}\mathbf{F}_t'\right\|\\
&= O_p(\max(n^{-1},T^{-1/2}))+O_p(T^{-1/2}),
\end{aligned}
$$

by Lemmas C.12(iii) and D.1(ii). And, last

$$
\begin{aligned}
n^{-1}\left|T^{-1}\sum_{t=1}^T(\widetilde{\mathbf{F}}_t-\mathbf{F}_t)'(\widehat{\mathbf{\Lambda}}_n^{(0)}-\mathbf{\Lambda}_n)'\boldsymbol{\xi}_{nt}\right| &\le n^{-1}\left\|T^{-1}\sum_{t=1}^T(\widehat{\mathbf{\Lambda}}_n^{(0)}-\mathbf{\Lambda}_n)'\boldsymbol{\xi}_{nt}(\widetilde{\mathbf{F}}_t-\mathbf{F}_t)'\right\|_F\\
&\le \sqrt{r}n^{-1}\left\|T^{-1}\sum_{t=1}^T(\widehat{\mathbf{\Lambda}}_n^{(0)}-\mathbf{\Lambda}_n)'\boldsymbol{\xi}_{nt}(\widetilde{\mathbf{F}}_t-\mathbf{F}_t)'\right\|\\
&\le \sqrt{r}n^{-1/2}\|\widehat{\mathbf{\Lambda}}_n^{(0)}-\mathbf{\Lambda}_n\|\, n^{-1/2}\left\|T^{-1}\sum_{t=1}^T\boldsymbol{\xi}_{nt}(\widetilde{\mathbf{F}}_t-\mathbf{F}_t)'\right\|\\
&= O_p(\max(n^{-1},T^{-1/2})),
\end{aligned}
$$

by Lemmas C.12(iii) and D.2(ii). This completes the proof. □

Lemma D.5. *Under Assumptions 1, 2, 3, and 6, as $n,T\to\infty$:*

(i) $\min(n,\sqrt{T})\,n^{-1}\|\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)}-\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^{\xi})^{-1}\mathbf{\Lambda}_n\|=O_p(1)$;

(ii) $\min(n,\sqrt{T})\,n^{-1/2}\|\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}-\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^{\xi})^{-1}\|=O_p(1)$;

(iii) $n\|(\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)})^{-1}\|=O_p(1)$;

(iv) $\min(n,\sqrt{T})\,n\|(\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)})^{-1}-(\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^{\xi})^{-1}\mathbf{\Lambda}_n)^{-1}\|=O_p(1)$;

(v) $\min(n,\sqrt{T})\,\sqrt{n}\|(\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}-(\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^{\xi})^{-1}\mathbf{\Lambda}_n)^{-1}\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^{\xi})^{-1}\|=O_p(1)$.

Proof. Start with

$$
\begin{aligned}
n^{-1}\|\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)}-\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^{\xi})^{-1}\mathbf{\Lambda}_n\| &\le 2n^{-1}\|\{\widehat{\mathbf{\Lambda}}_n^{(0)}-\mathbf{\Lambda}_n\}'(\mathbf{\Sigma}_n^{\xi})^{-1}\mathbf{\Lambda}_n\|\\
&\quad+n^{-1}\|\mathbf{\Lambda}_n'\{(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}-(\mathbf{\Sigma}_n^{\xi})^{-1}\}\mathbf{\Lambda}_n\|\\
&\quad+2n^{-1}\|\{\widehat{\mathbf{\Lambda}}_n^{(0)}-\mathbf{\Lambda}_n\}'\{(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}-(\mathbf{\Sigma}_n^{\xi})^{-1}\}\mathbf{\Lambda}_n\|\\
&\quad+n^{-1}\|\{\widehat{\mathbf{\Lambda}}_n^{(0)}-\mathbf{\Lambda}_n\}'\{(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}-(\mathbf{\Sigma}_n^{\xi})^{-1}\}\{\widehat{\mathbf{\Lambda}}_n^{(0)}-\mathbf{\Lambda}_n\}\|\\
&\le 2n^{-1/2}\|\widehat{\mathbf{\Lambda}}_n^{(0)}-\mathbf{\Lambda}_n\|\,\|(\mathbf{\Sigma}_n^{\xi})^{-1}\|\,n^{-1/2}\|\mathbf{\Lambda}_n\|\\
&\quad+n^{-1}\|\mathbf{\Lambda}_n'\{(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}-(\mathbf{\Sigma}_n^{\xi})^{-1}\}\mathbf{\Lambda}_n\|\\
&\quad+2n^{-1/2}\|\widehat{\mathbf{\Lambda}}_n^{(0)}-\mathbf{\Lambda}_n\|n^{-1/2}\|\{(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}-(\mathbf{\Sigma}_n^{\xi})^{-1}\}\mathbf{\Lambda}_n\|\\
&\quad+n^{-1}\|\widehat{\mathbf{\Lambda}}_n^{(0)}-\mathbf{\Lambda}_n\|^2\|(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}-(\mathbf{\Sigma}_n^{\xi})^{-1}\|.
\end{aligned}
\tag{D.8}
$$

Consider each term on the rhs of (D.8). The first term is

$$
n^{-1/2}\|\widehat{\mathbf{\Lambda}}_n^{(0)}-\mathbf{\Lambda}_n\|\,\|(\mathbf{\Sigma}_n^{\xi})^{-1}\|\,n^{-1/2}\|\mathbf{\Lambda}_n\|=O_p(\max(n^{-1},T^{-1/2})),
\tag{D.9}
$$

by Lemma D.1(ii), and also Assumption 2(a) for which $\|(\mathbf{\Sigma}_n^{\xi})^{-1}\|\le C_\xi$, and Lemma C.2 for which $n^{-1/2}\|\mathbf{\Lambda}_n\|=O(1)$. The third term is

$$
\begin{aligned}
&n^{-1/2}\|\widehat{\mathbf{\Lambda}}_n^{(0)}-\mathbf{\Lambda}_n\|n^{-1/2}\|\{(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}-(\mathbf{\Sigma}_n^{\xi})^{-1}\}\mathbf{\Lambda}_n\|\\
&\quad\le n^{-1/2}\|\widehat{\mathbf{\Lambda}}_n^{(0)}-\mathbf{\Lambda}_n\|\left\{\|(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\|+\|(\mathbf{\Sigma}_n^{\xi})^{-1}\|\right\}n^{-1/2}\|\mathbf{\Lambda}_n\|\\
&\quad= O_p(\max(n^{-1},T^{-1/2})),
\end{aligned}
\tag{D.10}
$$

by Lemmas C.2 and D.1(ii), Assumption 2(a), and since, by Lemma D.4(i), for all $j = 1, \dots, n$,

$$
\begin{aligned}
\|(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\| &= \left\{\nu^{(n)}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})\right\}^{-1} = \left\{\min_{i=1,\dots,n} \widehat{\sigma}_i^{2(0)}\right\}^{-1} = \left\{\min_{i=1,\dots,n} \sigma_i^2 + \widehat{\sigma}_i^{2(0)} - \sigma_i^2\right\}^{-1} \\
&\le \left\{\min_{i=1,\dots,n} \sigma_i^2 + \min_{i=1,\dots,n} (\widehat{\sigma}_i^{2(0)} - \sigma_i^2)\right\}^{-1} \\
&\le \left\{\min_{i=1,\dots,n} \sigma_i^2 - \min_{i=1,\dots,n} |\widehat{\sigma}_i^{2(0)} - \sigma_i^2|\right\}^{-1} \\
&\le \left\{C_\xi^{-1} - |\widehat{\sigma}_j^{2(0)} - \sigma_j^2|\right\}^{-1} \le C_\xi + O_p(\max(n^{-1}, T^{-1/2})).
\end{aligned} \tag{D.11}
$$

By the same arguments, the fourth term is

$$
n^{-1}\|\widehat{\mathbf{\Lambda}}_n^{(0)} - \mathbf{\Lambda}_n\|^2 \left\{\|(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\| + \|(\mathbf{\Sigma}_n^{\xi})^{-1}\|\right\} = o_p(\max(n^{-1}, T^{-1/2})). \tag{D.12}
$$

Finally, the second term on the rhs of (D.8) is

$$
\begin{aligned}
n^{-1}\|\mathbf{\Lambda}_n'\{(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1} - (\mathbf{\Sigma}_n^{\xi})^{-1}\}\mathbf{\Lambda}_n\| &= n^{-1}\left\|\sum_{i=1}^n \boldsymbol{\lambda}_i\boldsymbol{\lambda}_i'\{(\widehat{\sigma}_i^{2(0)})^{-1} - (\sigma_i^2)^{-1}\}\right\| \\
&\le n^{-1}\left(\sum_{k,h=1}^r \left[\sum_{i=1}^n \lambda_{ih}\lambda_{ik}\{(\widehat{\sigma}_i^{2(0)})^{-1} - (\sigma_i^2)^{-1}\}\right]^2\right)^{1/2} \\
&\le rM_\lambda n^{-1}\left|\sum_{i=1}^n \{(\widehat{\sigma}_i^{2(0)})^{-1} - (\sigma_i^2)^{-1}\}\right| \\
&= rM_\lambda n^{-1}\left|\sum_{i=1}^n (\widehat{\sigma}_i^{2(0)})^{-1}(\sigma_i^2)^{-1}\{\widehat{\sigma}_i^{2(0)} - \sigma_i^2\}\right| \\
&= rM_\lambda C_\xi n^{-1}\left|\left\{\min_{i=1,\dots,n} \widehat{\sigma}_i^{2(0)}\right\}^{-1} \sum_{i=1}^n \{\widehat{\sigma}_i^{2(0)} - \sigma_i^2\}\right| \\
&= rM_\lambda C_\xi^2 n^{-1}\left|\sum_{i=1}^n \{\widehat{\sigma}_i^{2(0)} - \sigma_i^2\}\right| \\
&= O_p(\max(n^{-1}, T^{-1/2})),
\end{aligned} \tag{D.13}
$$

by Lemma D.4(ii), (D.11), and Assumptions 1(a) and 2(a). By substituting (D.9), (D.10), (D.12), and (D.13) into (D.8), we prove part (i).

For part (ii), we have

$$
\begin{aligned}
n^{-1}\|\widehat{\mathbf{\Lambda}}_n^{(0)}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}-\mathbf{\Lambda}_n(\mathbf{\Sigma}_n^{\xi})^{-1}\|^2 &= n^{-1}\left\|\sum_{i=1}^n \widehat{\boldsymbol{\lambda}}_i^{(0)}(\widehat{\sigma}_i^{2(0)})^{-1}-\sum_{i=1}^n \boldsymbol{\lambda}_i(\sigma_i^2)^{-1}\right\|^2\\
&= n^{-1}\left\|\sum_{i=1}^n \widehat{\boldsymbol{\lambda}}_i^{(0)}(\widehat{\sigma}_i^{2(0)}-\sigma_i^2+\sigma_i^2)^{-1}-\sum_{i=1}^n \boldsymbol{\lambda}_i(\sigma_i^2)^{-1}\right\|^2\\
&= n^{-1}\left\|\sum_{i=1}^n \widehat{\boldsymbol{\lambda}}_i^{(0)}\sigma_i^2\{(\widehat{\sigma}_i^{2(0)}-\sigma_i^2)/\sigma_i^2+1\}^{-1}-\sum_{i=1}^n \boldsymbol{\lambda}_i(\sigma_i^2)^{-1}\right\|^2\\
&\le n^{-1}\left\|\left\{\min_{i=1,\ldots,n}(\widehat{\sigma}_i^{2(0)}-\sigma_i^2)/\sigma_i^2+1\right\}^{-1}\sum_{i=1}^n \widehat{\boldsymbol{\lambda}}_i^{(0)}(\sigma_i^2)^{-1}-\sum_{i=1}^n \boldsymbol{\lambda}_i(\sigma_i^2)^{-1}\right\|^2\\
&= n^{-1}\left\|\left\{1-\min_{i=1,\ldots,n}(\widehat{\sigma}_i^{2(0)}-\sigma_i^2)/\sigma_i^2+o\left(\min_{i=1,\ldots,n}(\widehat{\sigma}_i^{2(0)}-\sigma_i^2)/\sigma_i^2\right)\right\}\sum_{i=1}^n \widehat{\boldsymbol{\lambda}}_i^{(0)}(\sigma_i^2)^{-1}-\sum_{i=1}^n \boldsymbol{\lambda}_i(\sigma_i^2)^{-1}\right\|^2\\
&\le n^{-1}\left\|\widehat{\mathbf{\Lambda}}_n^{(0)}(\mathbf{\Sigma}_n^{\xi})^{-1}-\mathbf{\Lambda}_n(\mathbf{\Sigma}_n^{\xi})^{-1}\right\|^2+\left\{\min_{i=1,\ldots,n}\left|(\widehat{\sigma}_i^{2(0)}-\sigma_i^2)/\sigma_i^2\right|\right\}^2\left\|\widehat{\mathbf{\Lambda}}_n^{(0)}(\mathbf{\Sigma}_n^{\xi})^{-1}\right\|^2\\
&\le n^{-1}\left\|\widehat{\mathbf{\Lambda}}_n^{(0)}(\mathbf{\Sigma}_n^{\xi})^{-1}-\mathbf{\Lambda}_n(\mathbf{\Sigma}_n^{\xi})^{-1}\right\|^2+\left\{\min_{i=1,\ldots,n}|\widehat{\sigma}_i^{2(0)}-\sigma_i^2|\right\}^2 L_\xi^2\, n^{-1}\|\mathbf{\Lambda}_n\|^2\|(\mathbf{\Sigma}_n^{\xi})^{-1}\|^2\\
&\quad+\left\{\min_{i=1,\ldots,n}\left|\widehat{\sigma}_i^{2(0)}-\sigma_i^2\right|\right\}^2 L_\xi^2\, n^{-1}\|\widehat{\mathbf{\Lambda}}_n^{(0)}-\mathbf{\Lambda}_n\|^2\|(\mathbf{\Sigma}_n^{\xi})^{-1}\|^2\\
&= O_p(\max(n^{-2},T^{-1})),
\end{aligned}
$$

by Lemmas C.2, D.1(ii), D.4(ii), and Assumptions 2(a) and 2(f). This proves part (ii).

For part (iii), by part (ii) and Merikoski and Kumar (2004, Theorem 1) which is Weyl's inequality, we have

$$
\begin{aligned}
n^{-1}|\nu^{(r)}(\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)})-\nu^{(r)}(\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^{\xi})^{-1})\mathbf{\Lambda}_n)| &\le n^{-1}\|\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)}-\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^{\xi})^{-1}\mathbf{\Lambda}_n\|\\
&= O_p(\max(n^{-1},T^{-1/2})).
\end{aligned}\tag{D.14}
$$

Moreover (note that $x-y\ge -|x-y|$ for any $x,y\in\mathbb{R}$),

$$
\begin{aligned}
\det(\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)}) &= \prod_{j=1}^r \nu^{(j)}(\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)}) \ge \left\{\nu^{(r)}(\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)})\right\}^r\\
&\ge \left\{\nu^{(r)}(\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^{\xi})^{-1}\mathbf{\Lambda}_n)-|\nu^{(r)}(\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)})-\nu^{(r)}(\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^{\xi})^{-1})\mathbf{\Lambda}_n)|\right\}^r,
\end{aligned}
$$

thus, by Lemma C.3(iv), which implies $\lim_{n\to\infty} n^{-1}\nu^{(r)}(\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^{\xi})^{-1}\mathbf{\Lambda}_n)>0$, and (D.14), it follows that, with probability tending to one as $n,T\to\infty$, we have $\det(n^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)})>0$, or, equivalently $n^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)}$ is positive definite, i.e., $n\|(\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)})^{-1}\|=O_p(1)$. This proves part (iii).

For part (iv), we have

$$
\begin{aligned}
&n\|(\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)})^{-1}-(\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^{\xi})^{-1}\mathbf{\Lambda}_n)^{-1}\|\\
&\quad\le n\|(\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)})^{-1}\|\, n^{-1}\|\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)}-\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^{\xi})^{-1}\mathbf{\Lambda}_n\|\, n\|(\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^{\xi})^{-1}\mathbf{\Lambda}_n)^{-1}\|\\
&\quad= O_p(\max(n^{-1},T^{-1/2})),
\end{aligned}
$$

because of parts (i) and (iii) and Lemma C.3(iii).

Part (v) follows directly from parts (ii) and (iv). This completes the proof. □

Lemma D.6. *For all $T\in\mathbb{N}$,*

(i) $\mathbf{P}_{t|t-1}$, $\mathbf{P}_{t|t}$, and $\mathbf{P}_{t|T}$ are deterministic $r\times r$ matrices, for all $t=1,\ldots,T$;
(ii) $\|\mathbf{P}_{t+1|t}\|\le\|\mathbf{P}_{t|t-1}\|$, for all $t=1,\ldots,T-1$;
(iii) $\mathbf{P}_{0,t|t-1}$, $\mathbf{P}_{0,t|t}$, and $\mathbf{P}_{0,t|T}$ are deterministic $r\times r$ matrices, for all $t=1,\ldots,T$;
(iv) $\|\mathbf{P}_{0,t+1|t}\|\le\|\mathbf{P}_{0,t|t-1}\|$, for all $t=1,\ldots,T-1$.

Proof. Since $\mathbf{P}_{0|0}=\mathbf{I}_r$ is deterministic, then, for all $t=1,\ldots,T$, $\mathbf{P}_{t|t-1}$, $\mathbf{P}_{t|t}$, and $\mathbf{P}_{t|T}$ do not depend on the actual observations because of (A.2) and (A.4). This proves part (i).

As for part (ii), since $\mathbf{F}_{t|t-1}$ is based on less information than $\mathbf{F}_{t+1|t}$ and since $\mathbf{P}_{t|t-1}$ and $\mathbf{P}_{t+1|t}$ are deterministic, then $(\mathbf{P}_{t|t-1} - \mathbf{P}_{t+1|t})$ is a positive definite matrix for all $t = 1, \ldots, T-1$ (see, e.g., Harvey, 1990, Chapter 3.3, p. 123). As consequence, $\|\mathbf{P}_{t+1|t}\| \leq \|\mathbf{P}_{t|t-1}\|$ for all $t = 1, \ldots, T-1$ (see, e.g., Marshall et al., 2011, Proposition L1, p. 360).

The proof of parts (iii) and (iv) is identical to parts (i) and (ii), respectively, since also in this case $\mathbf{P}_{0,0|0} = \mathbf{I}_r$. This completes the proof. □

Lemma D.7. *Under Assumptions 1 and 2, for all $T \in \mathbb{N}$,*

(i) $\max_{t=1,\ldots,T} \|\mathbf{P}_{t|t-1}\| \leq M_P$ *for some finite positive real M_P;*

(ii) $\min_{t=1,\ldots,T} \nu^{(r)}(\mathbf{P}_{t|t-1}) \geq \underline{M}_P$ *for some finite positive real $\underline{M}_P$.*

(iii) $\max_{t=1,\ldots,T} \|\mathbf{P}_{0,t|t-1}\| \leq M_P$ *for some finite positive real M_P;*

(iv) $\min_{t=1,\ldots,T} \nu^{(r)}(\mathbf{P}_{0,t|t-1}) \geq \underline{M}_P$ *for some finite positive real $\underline{M}_P$.*

Proof. Given that $\mathbf{P}_{0|0} = \mathbf{I}_r$ is obviously positive definite, by (A.2) and Weyl's inequality (Merikoski and Kumar, 2004, Theorem 1) it follows that

$$\nu^{(r)}(\mathbf{P}_{1|0}) \geq \nu^{(r)}(\mathbf{A}\mathbf{A}') + \nu^{(r)}(\mathbf{\Gamma}^v) \geq M_v^{-1}, \tag{D.15}$$

for some finite positive real M_v^{-1}, since $\mathbf{\Gamma}^v$ has full rank by Assumption 1(e) and $\nu^{(r)}(\mathbf{A}\mathbf{A}')$ is real and such that $\nu^{(r)}(\mathbf{A}\mathbf{A}') \geq (\nu^{(r)}(\mathbf{A}))^2 \geq 0$. From $\mathbf{P}_{0|0} = \mathbf{I}_r$ and (A.2) it follows also that

$$\|\mathbf{P}_{1|0}\| \leq \|\mathbf{A}\|^2 \|\mathbf{P}_{0|0}\| + \|\mathbf{\Gamma}^v\|^2 \leq M_A^2 + M_v^2 = M_P, \quad \text{say.} \tag{D.16}$$

By Lemma D.6(ii) and (D.16), we have

$$\max_{t=2,\ldots,T} \|\mathbf{P}_{t|t-1}\| \leq \|\mathbf{P}_{1|0}\| \leq M_P,$$

since M_P is independent of t and $\mathbf{P}_{t|t-1}$ is deterministic because of Lemma D.6(i). This proves part (i).

For part (ii), by Merikoski and Kumar (2004, Theorems 1 and 7) and (A.2)

$$\nu^{(r)}(\mathbf{P}_{t|t-1}) \geq \nu^{(r)}(\mathbf{A}\mathbf{A}')\nu^{(r)}(\mathbf{P}_{t|t}) + \nu^{(r)}(\mathbf{\Gamma}^v) \geq M_v^{-1}, \tag{D.17}$$

by the same arguments leading to (D.15) and since $\nu_{\min}(\mathbf{P}_{t|t}) \geq 0$ because $\mathbf{P}_{t|t}$ is at least positive semidefinite by construction. By letting $\underline{M}_P = M_v^{-1}$, we prove part (ii). Parts (iii) and (iv) are proved exactly as parts (i) and (ii), respectively, since also in this case $\mathbf{P}_{0,0|0} = \mathbf{I}_r$. This completes the proof. □

Lemma D.8. *Under Assumptions 1, 2, 3, and 6, as $n, T \to \infty$:*

(i) $\max_{t=1,\ldots,T} \|\mathbf{P}^{(0)}_{t|t-1}\| = O_p(1)$;

(ii) $\max_{t=1,\ldots,T} \|(\mathbf{P}^{(0)}_{t|t-1})^{-1}\| = O_p(1)$.

Proof. For part (i),

$$\begin{aligned}\max_{t=1,\ldots,T} \|\mathbf{P}^{(0)}_{t|t-1}\| &\leq \max_{t=1,\ldots,T} \|\mathbf{P}_{t|t-1}\| + \max_{t=1,\ldots,T} \|\mathbf{P}^{(0)}_{t|t-1} - \mathbf{P}_{t|t-1}\| \\ &= O_p(1) + O_p(\max(n^{-1}, T^{-1/2})),\end{aligned}$$

by Lemma D.7(i) and since the second term on the rhs depends only on the estimation error of $\widehat{\mathbf{A}}^{(0)}$, $\widehat{\mathbf{\Gamma}}^{v(0)}$, $n^{-1/2}\widehat{\mathbf{\Lambda}}_n^{(0)}$, $n^{-1/2}\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}$ and $n^{-1}(\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)})^{-1}$, which are all bounded by Lemmas D.1(ii), D.3, D.5(ii), and D.5(iv). This proves part (i).

Part (ii) is proved in the same way as part (i) but using Lemma D.7(ii). This completes the proof. □

Lemma D.9. *For $m < n$ with m independent of n and given*

(a) an $n \times n$ matrix $\boldsymbol{A}$ symmetric and positive definite with $\|\boldsymbol{A}\| \leq M_A$;

(a) an $m \times m$ matrix $\boldsymbol{B}$ symmetric and positive definite with $\|\boldsymbol{B}\| \leq M_B$;

(c) an $n \times m$ matrix $\boldsymbol{U}$ such that $\|n^{-1}\boldsymbol{U}'\boldsymbol{U}\| \leq M_U$ and $rk(\boldsymbol{U}) = m$;

(d) an $m \times n$ matrix $\boldsymbol{V}$ such that $\|n^{-1}\boldsymbol{V}\boldsymbol{V}'\| \leq M_V$ and $rk(\boldsymbol{V}) = m$;

where, M_A, M_B, M_U, and M_V are finite positive reals independent of n and m, then the following holds

(i) $(\boldsymbol{A} + \boldsymbol{U}\boldsymbol{B}\boldsymbol{V})^{-1} = \boldsymbol{A}^{-1} - \boldsymbol{A}^{-1}\boldsymbol{U}\boldsymbol{B}(\mathbf{I}_m + \boldsymbol{V}\boldsymbol{A}^{-1}\boldsymbol{U}\boldsymbol{B})^{-1}\boldsymbol{V}\boldsymbol{A}^{-1}$;

(ii) $(\boldsymbol{A}+\boldsymbol{U}\boldsymbol{B}\boldsymbol{V})^{-1} = \boldsymbol{A}^{-1} - \boldsymbol{A}^{-1}\boldsymbol{U}\boldsymbol{B}\boldsymbol{V}(\mathbf{I}_m + \boldsymbol{V}\boldsymbol{A}^{-1}\boldsymbol{U}\boldsymbol{B}\boldsymbol{V})^{-1}\boldsymbol{A}^{-1}$.

Proof. Both results are proved by Henderson and Searle (1981, eq. (24) and eq. (25), respectively). □

Lemma D.10. *For any $r\times r$ symmetric and positive definite matrix $\boldsymbol{P}$ with $\|\boldsymbol{P}\|\le M_P$ for some finite positive real M_P, under Assumptions 1 and 2*

$$\boldsymbol{P}\boldsymbol{\Lambda}_n'(\boldsymbol{\Lambda}_n\boldsymbol{P}\boldsymbol{\Lambda}_n' + \boldsymbol{\Sigma}_n^{\xi})^{-1}\boldsymbol{\Lambda}_n = \boldsymbol{P}((\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^{\xi})^{-1}\boldsymbol{\Lambda}_n)^{-1}+\boldsymbol{P})^{-1}.$$

Proof. In Lemma D.9(i) set $\boldsymbol{A}=\boldsymbol{\Sigma}_n^{\xi}$, $\boldsymbol{B}=\mathbf{I}_r$, $\boldsymbol{U}=\boldsymbol{\Lambda}_n\boldsymbol{P}$, and $\boldsymbol{V}=\boldsymbol{\Lambda}_n'$. Then, by noticing that the assumptions of Lemma D.9 are satisfied because of Assumptions 1(a) and 2(a), it follows that:

$$(\boldsymbol{\Lambda}_n\boldsymbol{P}\boldsymbol{\Lambda}_n' + \boldsymbol{\Sigma}_n^{\xi})^{-1} = (\boldsymbol{\Sigma}_n^{\xi})^{-1} - (\boldsymbol{\Sigma}_n^{\xi})^{-1}\boldsymbol{\Lambda}_n\boldsymbol{P}(\mathbf{I}_r + \boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^{\xi})^{-1}\boldsymbol{\Lambda}_n\boldsymbol{P})^{-1}\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^{\xi})^{-1}.$$

Therefore,

$$\begin{aligned}\boldsymbol{P}\boldsymbol{\Lambda}_n'(\boldsymbol{\Lambda}_n\boldsymbol{P}\boldsymbol{\Lambda}_n' + \boldsymbol{\Sigma}_n^{\xi})^{-1}\boldsymbol{\Lambda}_n &= \boldsymbol{P}\boldsymbol{\Lambda}_n'\{(\boldsymbol{\Sigma}_n^{\xi})^{-1} - (\boldsymbol{\Sigma}_n^{\xi})^{-1}\boldsymbol{\Lambda}_n\boldsymbol{P}(\mathbf{I}_r + \boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^{\xi})^{-1}\boldsymbol{\Lambda}_n\boldsymbol{P})^{-1}\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^{\xi})^{-1}\}\boldsymbol{\Lambda}_n\\ &= \boldsymbol{P}\{\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^{\xi})^{-1}\boldsymbol{\Lambda}_n - \boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^{\xi})^{-1}\boldsymbol{\Lambda}_n\boldsymbol{P}(\mathbf{I}_r + \boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^{\xi})^{-1}\boldsymbol{\Lambda}_n\boldsymbol{P})\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^{\xi})^{-1}\boldsymbol{\Lambda}_n\}. \end{aligned} \tag{D.18}$$

Now, in Lemma D.9(ii) set $\boldsymbol{A}=(\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^{\xi})^{-1}\boldsymbol{\Lambda}_n)^{-1}$, $\boldsymbol{B}=\boldsymbol{P}$, $\boldsymbol{U}=\mathbf{I}_r$, and $\boldsymbol{V}=\mathbf{I}_r$, and notice that the assumptions therein are satisfied because of Lemmas C.3(iii) and C.3(v). Then, for the last line of (D.18) we have

$$\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^{\xi})^{-1}\boldsymbol{\Lambda}_n - \boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^{\xi})^{-1}\boldsymbol{\Lambda}_n\boldsymbol{P}(\mathbf{I}_r + \boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^{\xi})^{-1}\boldsymbol{\Lambda}_n\boldsymbol{P})\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^{\xi})^{-1}\boldsymbol{\Lambda}_n = ((\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^{\xi})^{-1}\boldsymbol{\Lambda}_n)^{-1}+\boldsymbol{P})^{-1}. \tag{D.19}$$

By substituting (D.19) into (D.18) we complete the proof. □

Lemma D.11. *Under Assumptions 1, 2, 3, and 6, as $n,T\to\infty$, $\max_{t=1,\dots,T} n\|\mathbf{P}_{t|t}^{(0)}\| = O_p(1)$.*

Proof. From (A.4) by using Lemma D.10, but with $\widehat{\boldsymbol{\Lambda}}_n^{(0)}$ and $\widehat{\boldsymbol{\Sigma}}_n^{\xi(0)}$ in place of $\boldsymbol{\Lambda}_n$ and $\boldsymbol{\Sigma}_n^{\xi}$, it holds that:

$$\begin{aligned}\mathbf{P}_{t|t}^{(0)} &= \mathbf{P}_{t|t-1}^{(0)} - \mathbf{P}_{t|t-1}^{(0)}\widehat{\boldsymbol{\Lambda}}_n^{(0)\prime}(\widehat{\boldsymbol{\Lambda}}_n^{(0)}\mathbf{P}_{t|t-1}^{(0)}\widehat{\boldsymbol{\Lambda}}_n^{(0)\prime} + \widehat{\boldsymbol{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\boldsymbol{\Lambda}}_n^{(0)}\mathbf{P}_{t|t-1}^{(0)}\\ &= \left\{\mathbf{I}_r - \mathbf{P}_{t|t-1}^{(0)}\widehat{\boldsymbol{\Lambda}}_n^{(0)\prime}(\widehat{\boldsymbol{\Lambda}}_n^{(0)}\mathbf{P}_{t|t-1}^{(0)}\widehat{\boldsymbol{\Lambda}}_n^{(0)\prime} + \widehat{\boldsymbol{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\boldsymbol{\Lambda}}_n^{(0)}\right\}\mathbf{P}_{t|t-1}^{(0)}\\ &= \left\{\mathbf{I}_r - \mathbf{P}_{t|t-1}^{(0)}((\widehat{\boldsymbol{\Lambda}}_n^{(0)\prime}(\widehat{\boldsymbol{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\boldsymbol{\Lambda}}_n^{(0)})^{-1} + \mathbf{P}_{t|t-1}^{(0)})^{-1}\right\}\mathbf{P}_{t|t-1}^{(0)}. \end{aligned} \tag{D.20}$$

Then, by setting in Lemma C.4 $\boldsymbol{K}=\mathbf{P}_{t|t-1}^{(0)}$ and $\boldsymbol{H}=(\widehat{\boldsymbol{\Lambda}}_n^{(0)\prime}(\widehat{\boldsymbol{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\boldsymbol{\Lambda}}_n^{(0)})^{-1}$, for the last line of (D.20) we have

$$\begin{aligned}((\widehat{\boldsymbol{\Lambda}}_n^{(0)\prime}(\widehat{\boldsymbol{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\boldsymbol{\Lambda}}_n^{(0)})^{-1} + \mathbf{P}_{t|t-1}^{(0)})^{-1} &= (\mathbf{P}_{t|t-1}^{(0)})^{-1}\\ &\quad - ((\widehat{\boldsymbol{\Lambda}}_n^{(0)\prime}(\widehat{\boldsymbol{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\boldsymbol{\Lambda}}_n^{(0)})^{-1} + \mathbf{P}_{t|t-1}^{(0)})^{-1}(\widehat{\boldsymbol{\Lambda}}_n^{(0)\prime}(\widehat{\boldsymbol{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\boldsymbol{\Lambda}}_n^{(0)})^{-1}(\mathbf{P}_{t|t-1}^{(0)})^{-1}. \end{aligned} \tag{D.21}$$

By substituting (D.21) into (D.20) we get

$$\mathbf{P}_{t|t}^{(0)} = \mathbf{P}_{t|t-1}^{(0)}((\widehat{\boldsymbol{\Lambda}}_n^{(0)\prime}(\widehat{\boldsymbol{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\boldsymbol{\Lambda}}_n^{(0)})^{-1} + \mathbf{P}_{t|t-1}^{(0)})^{-1}(\widehat{\boldsymbol{\Lambda}}_n^{(0)\prime}(\widehat{\boldsymbol{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\boldsymbol{\Lambda}}_n^{(0)})^{-1}. \tag{D.22}$$

Finally, by using again (D.21) into (D.22)

$$\begin{aligned}\mathbf{P}_{t|t}^{(0)} &= \mathbf{P}_{t|t-1}^{(0)}\left\{(\mathbf{P}_{t|t-1}^{(0)})^{-1} - ((\widehat{\boldsymbol{\Lambda}}_n^{(0)\prime}(\widehat{\boldsymbol{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\boldsymbol{\Lambda}}_n^{(0)})^{-1} + \mathbf{P}_{t|t-1}^{(0)})^{-1}(\widehat{\boldsymbol{\Lambda}}_n^{(0)\prime}(\widehat{\boldsymbol{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\boldsymbol{\Lambda}}_n^{(0)})^{-1}(\mathbf{P}_{t|t-1}^{(0)})^{-1}\right\}\\ &\quad\cdot(\widehat{\boldsymbol{\Lambda}}_n^{(0)\prime}(\widehat{\boldsymbol{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\boldsymbol{\Lambda}}_n^{(0)})^{-1}\\ &= \left\{\mathbf{I}_r - \mathbf{P}_{t|t-1}^{(0)}((\widehat{\boldsymbol{\Lambda}}_n^{(0)\prime}(\widehat{\boldsymbol{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\boldsymbol{\Lambda}}_n^{(0)})^{-1} + \mathbf{P}_{t|t-1}^{(0)})^{-1}(\widehat{\boldsymbol{\Lambda}}_n^{(0)\prime}(\widehat{\boldsymbol{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\boldsymbol{\Lambda}}_n^{(0)})^{-1}(\mathbf{P}_{t|t-1}^{(0)})^{-1}\right\}\\ &\quad\cdot(\widehat{\boldsymbol{\Lambda}}_n^{(0)\prime}(\widehat{\boldsymbol{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\boldsymbol{\Lambda}}_n^{(0)})^{-1}\\ &= (\widehat{\boldsymbol{\Lambda}}_n^{(0)\prime}(\widehat{\boldsymbol{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\boldsymbol{\Lambda}}_n^{(0)})^{-1}\\ &\quad - \mathbf{P}_{t|t-1}^{(0)}((\widehat{\boldsymbol{\Lambda}}_n^{(0)\prime}(\widehat{\boldsymbol{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\boldsymbol{\Lambda}}_n^{(0)})^{-1} + \mathbf{P}_{t|t-1}^{(0)})^{-1}(\widehat{\boldsymbol{\Lambda}}_n^{(0)\prime}(\widehat{\boldsymbol{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\boldsymbol{\Lambda}}_n^{(0)})^{-1}\\ &\quad\cdot(\mathbf{P}_{t|t-1}^{(0)})^{-1}(\widehat{\boldsymbol{\Lambda}}_n^{(0)\prime}(\widehat{\boldsymbol{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\boldsymbol{\Lambda}}_n^{(0)})^{-1}. \end{aligned} \tag{D.23}$$

Notice that we could use Lemmas C.4 and D.10 to derive (D.23) since all inverses used are well defined because of Lemmas D.5(iii), D.8(i), and D.8(ii) and (D.11) in the proof of Lemma D.4.

Therefore, from (D.23)

$$\begin{aligned}\max_{t=1,\dots,T} n\|\mathbf{P}_{t|t}^{(0)}\| \le\, & n\|(\widehat{\boldsymbol{\Lambda}}_n^{(0)\prime}(\widehat{\boldsymbol{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\boldsymbol{\Lambda}}_n^{(0)})^{-1}\| \\ & + \max_{t=1,\dots,T}\|\mathbf{P}_{t|t-1}^{(0)}\| \max_{t=1,\dots,T}\|(\mathbf{P}_{t|t-1}^{(0)})^{-1}\|\, n\|(\widehat{\boldsymbol{\Lambda}}_n^{(0)\prime}(\widehat{\boldsymbol{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\boldsymbol{\Lambda}}_n^{(0)})^{-1}\|^2 \\ & \cdot \|((\widehat{\boldsymbol{\Lambda}}_n^{(0)\prime}(\widehat{\boldsymbol{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\boldsymbol{\Lambda}}_n^{(0)})^{-1} + \mathbf{P}_{t|t-1}^{(0)})^{-1}\| \\ =\, & O_p(1) + O_p(n^{-1}),\end{aligned}$$

because of Lemmas D.5(iii), D.8(i), and D.8(ii), and since, by Merikoski and Kumar (2004, Theorem 1) which is Weyl's inequality,

$$\begin{aligned}&\|((\widehat{\boldsymbol{\Lambda}}_n^{(0)\prime}(\widehat{\boldsymbol{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\boldsymbol{\Lambda}}_n^{(0)})^{-1} + \mathbf{P}_{t|t-1}^{(0)})^{-1}\| = \left\{\nu^{(r)}((\widehat{\boldsymbol{\Lambda}}_n^{(0)\prime}(\widehat{\boldsymbol{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\boldsymbol{\Lambda}}_n^{(0)})^{-1} + \mathbf{P}_{t|t-1}^{(0)})\right\}^{-1} \\ &\le \left\{\nu^{(r)}((\widehat{\boldsymbol{\Lambda}}_n^{(0)\prime}(\widehat{\boldsymbol{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\boldsymbol{\Lambda}}_n^{(0)})^{-1}) + \nu^{(r)}(\mathbf{P}_{t|t-1}^{(0)})\right\}^{-1} \\ &= \left\{\left[\nu^{(1)}(\widehat{\boldsymbol{\Lambda}}_n^{(0)\prime}(\widehat{\boldsymbol{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\boldsymbol{\Lambda}}_n^{(0)})\right]^{-1} + \nu^{(r)}(\mathbf{P}_{t|t-1}^{(0)})\right\}^{-1} \\ &= \left\{\left[\nu^{(1)}(\widehat{\boldsymbol{\Lambda}}_n^{(0)\prime}(\widehat{\boldsymbol{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\boldsymbol{\Lambda}}_n^{(0)})\nu^{(r)}(\mathbf{P}_{t|t-1}^{(0)})\right]^{-1} + 1\right\}^{-1}\left\{\nu^{(r)}(\mathbf{P}_{t|t-1}^{(0)})\right\}^{-1} \\ &= \left\{1 - \left[\nu^{(1)}(\widehat{\boldsymbol{\Lambda}}_n^{(0)\prime}(\widehat{\boldsymbol{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\boldsymbol{\Lambda}}_n^{(0)})\nu^{(r)}(\mathbf{P}_{t|t-1}^{(0)})\right]^{-1}\right\}\left\{\nu^{(r)}(\mathbf{P}_{t|t-1}^{(0)})\right\}^{-1} + O_p(n^{-2}) \\ &= O_p(1), \end{aligned} \tag{D.24}$$

again by Lemmas D.5(iii), D.8(i), and D.8(ii). This completes the proof. □

Lemma D.12. *Under Assumptions 1, 2, 3, and 6, as $n, T \to \infty$,* $\max_{t=1,\dots,T} n\|\mathbf{P}_{t|T}^{(0)}\| = O_p(1)$.

Proof. From (A.7), we get

$$\|\mathbf{P}_{t|T}^{(0)} - \mathbf{P}_{t|t}^{(0)}\| \le \|\mathbf{P}_{t|t}^{(0)}\|^2\,\|\widehat{\mathbf{A}}^{(0)}\|^2\,\|(\mathbf{P}_{t+1|t}^{(0)})^{-1}\|^2\{\|\mathbf{P}_{t+1|T}^{(0)}\| + \|\mathbf{P}_{t+1|t}^{(0)}\|\}. \tag{D.25}$$

Start with $t = T-1$, then from (D.25),

$$\begin{aligned}\|\mathbf{P}_{T-1|T}^{(0)} - \mathbf{P}_{T-1|T-1}^{(0)}\| &\le \|\mathbf{P}_{T-1|T-1}^{(0)}\|^2\,\|\widehat{\mathbf{A}}^{(0)}\|^2\,\|(\mathbf{P}_{T|T-1}^{(0)})^{-1}\|^2\{\|\mathbf{P}_{T|T}^{(0)}\| + \|\mathbf{P}_{T|T-1}^{(0)}\|\} \\ &= O_p(n^{-2}).\end{aligned} \tag{D.26}$$

by Lemmas D.8(i), D.8(ii), and D.11, and since $\|\widehat{\mathbf{A}}^{(0)}\| \le \|\mathbf{A}\| + \|\widehat{\mathbf{A}}^{(0)} - \mathbf{A}\| = O_p(1)$, by Assumption 1(d) and Lemma D.3(i). From (D.26) it follows that

$$\|\mathbf{P}_{T-1|T}^{(0)}\| \le \|\mathbf{P}_{T-1|T-1}^{(0)}\| + \|\mathbf{P}_{T-1|T}^{(0)} - \mathbf{P}_{T-1|T-1}^{(0)}\| = O_p(n^{-1}) + O_p(n^{-2}). \tag{D.27}$$

Thus, at $t = T-2$, from (D.25) and (D.27),

$$\begin{aligned}\|\mathbf{P}_{T-2|T}^{(0)} - \mathbf{P}_{T-2|T-2}^{(0)}\| &\le \|\mathbf{P}_{T-2|T-2}^{(0)}\|^2\,\|\widehat{\mathbf{A}}^{(0)}\|^2\,\|(\mathbf{P}_{T-1|T-2}^{(0)})^{-1}\|^2\{\|\mathbf{P}_{T-1|T}^{(0)}\| + \|\mathbf{P}_{T-1|T-2}^{(0)}\|\} \\ &= O_p(n^{-2}).\end{aligned} \tag{D.28}$$

From (D.28) it follows that

$$\|\mathbf{P}_{T-2|T}^{(0)}\| \le \|\mathbf{P}_{T-2|T-2}^{(0)}\| + \|\mathbf{P}_{T-2|T}^{(0)} - \mathbf{P}_{T-2|T-2}^{(0)}\| = O_p(n^{-1}) + O_p(n^{-2}). \tag{D.29}$$

Since all the bounds in (D.26)-(D.29) are the same for all t, from Lemma D.11 and (D.25) we have

$$\max_{t=1,\dots,T} n\|\mathbf{P}_{t|T}^{(0)}\| \le \max_{t=1,\dots,T} n\|\mathbf{P}_{t|t}^{(0)}\| + \max_{t=1,\dots,T} n\|\mathbf{P}_{t|T}^{(0)} - \mathbf{P}_{t|t}^{(0)}\| = O_p(1) + O_p(n^{-1}).$$

This completes the proof. □

Lemma D.13. *For $m < n$, and given symmetric positive definite matrices $\boldsymbol{A}$ of dimension $m \times m$ and $\boldsymbol{B}$ of dimension $n \times n$, and for $\boldsymbol{C}$ of dimension $n \times m$ with $rk(\boldsymbol{C}) = m$, the following holds*

$$\boldsymbol{A}\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{A}\boldsymbol{C}' + \boldsymbol{B})^{-1} = (\boldsymbol{A}^{-1} + \boldsymbol{C}'\boldsymbol{B}^{-1}\boldsymbol{C})^{-1}\boldsymbol{C}'\boldsymbol{B}^{-1}. \tag{D.30}$$

Proof. Recall the Woodbury forumla

$$(\boldsymbol{C}\boldsymbol{A}\boldsymbol{C}' + \boldsymbol{B})^{-1} = \boldsymbol{B}^{-1} - \boldsymbol{B}^{-1}\boldsymbol{C}(\boldsymbol{A}^{-1} + \boldsymbol{C}'\boldsymbol{B}^{-1}\boldsymbol{C})^{-1}\boldsymbol{C}'\boldsymbol{B}^{-1}. \tag{D.31}$$

Denote $\boldsymbol{D} = (\boldsymbol{A}^{-1} + \boldsymbol{C}'\boldsymbol{B}^{-1}\boldsymbol{C})^{-1}$ then from (D.31) the lhs of (D.30) is equivalent to

$$\boldsymbol{A}\boldsymbol{C}'\left[\boldsymbol{B}^{-1} - \boldsymbol{B}^{-1}\boldsymbol{C}\boldsymbol{D}\boldsymbol{C}'\boldsymbol{B}^{-1}\right] = \boldsymbol{A}\left[\boldsymbol{C}'\boldsymbol{B}^{-1} - \boldsymbol{C}'\boldsymbol{B}^{-1}\boldsymbol{C}\boldsymbol{D}\boldsymbol{C}'\boldsymbol{B}^{-1}\right] = \boldsymbol{A}\left[\boldsymbol{I} - \boldsymbol{C}'\boldsymbol{B}^{-1}\boldsymbol{C}\boldsymbol{D}\right]\boldsymbol{C}'\boldsymbol{B}^{-1}.$$

Then, (D.30) becomes

$$\boldsymbol{A}\left[\boldsymbol{I} - \boldsymbol{C}'\boldsymbol{B}^{-1}\boldsymbol{C}\boldsymbol{D}\right]\boldsymbol{C}'\boldsymbol{B}^{-1} = \boldsymbol{D}\boldsymbol{C}'\boldsymbol{B}^{-1},$$

or equivalently multiplying both sides on the right by $\boldsymbol{B}\boldsymbol{C}(\boldsymbol{C}'\boldsymbol{C})^{-1}$

$$\boldsymbol{A}\left[\boldsymbol{I} - \boldsymbol{C}'\boldsymbol{B}^{-1}\boldsymbol{C}\boldsymbol{D}\right] = \boldsymbol{D}. \tag{D.32}$$

Now multiplying (D.32) on the left by $\boldsymbol{A}^{-1}$ and on the right by $\boldsymbol{D}^{-1}$

$$\left[\boldsymbol{D}^{-1} - \boldsymbol{C}'\boldsymbol{B}^{-1}\boldsymbol{C}\right] = \boldsymbol{A}^{-1},$$

which is equivalent to

$$\boldsymbol{A}^{-1} + \boldsymbol{C}'\boldsymbol{B}^{-1}\boldsymbol{C} - \boldsymbol{C}'\boldsymbol{B}^{-1}\boldsymbol{C} - \boldsymbol{A}^{-1} = \boldsymbol{0}_{m\times m},$$

which is always true. □

Lemma D.14. *Under Assumptions 1, 2, 3, and 6, as $n, T \to \infty$, for all $s = 0, \dots, T$, $\|\mathbf{F}_{t|s}^{(0)}\| = O_p(1)$, uniformly in $t \le s$.*

Proof. Let $\widehat{\boldsymbol{\Omega}}_s^{F(0)} = \mathbb{E}_{\widehat{\varphi}^{(0)}}[\boldsymbol{F}_s\boldsymbol{F}_s']$, which is $rs \times rs$ having the $r \times r$ generic (t_1, t_2) block denoted by $[\widehat{\boldsymbol{\Omega}}_s^{F(0)}]_{t_1,t_2}$ and such that $[\widehat{\boldsymbol{\Omega}}_s^{F(0)}]_{t_2,t_1} = [\widehat{\boldsymbol{\Omega}}_s^{F(0)}]'_{t_1,t_2}$ and

$$\text{vec}\left([\widehat{\boldsymbol{\Omega}}_s^{F(0)}]_{t_1,t_2}\right) = (\mathbf{I}_r \otimes \widehat{\mathbf{A}}^{(0)})^{|t_1-t_2|}(\mathbf{I}_{r^2} - \{\widehat{\mathbf{A}}^{(0)} \otimes \widehat{\mathbf{A}}^{(0)}\})^{-1}\text{vec}(\widehat{\boldsymbol{\Gamma}}^{v(0)}), \quad t_1, t_2 = 1, \dots, s.$$

Notice that although $\widehat{\boldsymbol{\Omega}}_s^{F(0)}$ depends on $\widehat{\mathbf{A}}^{(0)}$ and $\widehat{\boldsymbol{\Gamma}}^{v(0)}$, for simplicity of notation, hereafter, we omit such dependence. Let also $\boldsymbol{\Omega}_s^F = \mathbb{E}[\boldsymbol{F}_s\boldsymbol{F}_s']$, clearly $\boldsymbol{\Omega}_s^F$ is positive definite, since by Assumptions 1(b) and 1(d), $\|[\boldsymbol{\Omega}_s^F]_{t_1,t_2}\| \le \|\boldsymbol{\Gamma}^F\|$ for all $t_1 \ne t_2$. Moreover, recall that $\boldsymbol{\Omega}_s^F$ is a block-Toeplitz matrix and define the corresponding circulant matrix as $\boldsymbol{\Phi}_s^F$, then (see, e.g., Gray, 2006, Lemma 4.3 and Section 3.1). $s^{-1}|\nu^{(1)}(\boldsymbol{\Omega}_s^F) - \nu^{(1)}(\boldsymbol{\Psi}_s^F)| = O(s^{-1/2})$, and $\nu^{(1)}(\boldsymbol{\Psi}_s^F) = O(s)$. Thus,

$$\|\boldsymbol{\Omega}_s^F\| = \nu^{(1)}(\boldsymbol{\Omega}_s^F) \le \nu^{(1)}(\boldsymbol{\Psi}_s^F) + |\nu^{(1)}(\boldsymbol{\Omega}_s^F) - \nu^{(1)}(\boldsymbol{\Psi}_s^F)| = O(s) + O(\sqrt{s}),$$

which implies

$$s\|(\boldsymbol{\Omega}_s^F)^{-1}\| = O(1). \tag{D.33}$$

Now the $r \times r$ generic (t_1, t_2) block of $(\boldsymbol{\Omega}^F)_s^{-1}$ is an analytic function of $\mathbf{A}$ and $\boldsymbol{\Gamma}^v$, which, in the case $r = 1$, is given by (see Akaike, 1973 for the case $r > 1$)

$$[(\boldsymbol{\Omega}_s^F)^{-1}]_{t_1,t_2} = \mathbb{E}[v_t^2](1 - A^2)^{-2} \cdot \begin{cases} 1 & \text{if} \quad t_1 = t_2 = 1 \quad \text{and} \quad t_1 = t_2 = s, \\ 1 + A^2 & \text{if} \quad t_1 = t_2 \quad \text{and} \quad 1 < t_1, t_2 < s, \\ -A & \text{if} \quad |t_1 - t_2| = 1, \\ 0 & \text{otherwise.} \end{cases}$$

Then, because of Lemma D.3 and (D.33), we have that, as $n, s \to \infty$, $s(\widehat{\boldsymbol{\Omega}}_s^{F(0)})^{-1}$ is positive definite with probability tending to one, i.e.,

$$s\|(\widehat{\mathbf{\Omega}}_s^{F(0)})^{-1}\| = O_p(1). \tag{D.34}$$

Let now

$$\begin{aligned}\widehat{\boldsymbol{K}}_{ns}^{(0)} &= \widehat{\mathbf{\Omega}}_s^{F(0)}(\mathbf{I}_s \otimes \widehat{\mathbf{\Lambda}}_n^{(0)\prime})\left\{(\mathbf{I}_s \otimes \widehat{\mathbf{\Lambda}}_n^{(0)})\widehat{\mathbf{\Omega}}_s^{F(0)}(\mathbf{I}_s \otimes \widehat{\mathbf{\Lambda}}_n^{(0)\prime}) + (\mathbf{I}_s \otimes \widehat{\mathbf{\Sigma}}_n^{\xi(0)})\right\}^{-1}\\ &= \left\{(\mathbf{I}_s \otimes \widehat{\mathbf{\Lambda}}_n^{(0)\prime})(\mathbf{I}_s \otimes \widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}(\mathbf{I}_s \otimes \widehat{\mathbf{\Lambda}}_n^{(0)}) + (\widehat{\mathbf{\Omega}}_s^{F(0)})^{-1}\right\}^{-1}\left\{(\mathbf{I}_s \otimes \widehat{\mathbf{\Lambda}}_n^{(0)\prime})(\mathbf{I}_s \otimes \widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\right\}\\ &= \left\{\mathbf{I}_s \otimes \widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)} + (\widehat{\mathbf{\Omega}}_s^{F(0)})^{-1}\right\}^{-1}\left\{\mathbf{I}_s \otimes \widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\right\},\end{aligned} \tag{D.35}$$

which is $rs \times ns$ and where in the second line we used Lemma D.13. Notice that all the inverses in (D.35) are well defined by (D.34), Lemma D.5(iii), and (D.11) in the proof of Lemma D.4.

Then, by definition of linear projection, we have

$$\begin{aligned}\mathbf{F}_{t|s}^{(0)} &= \text{Proj}_{\widehat{\varphi}_n^{(0)}}[\mathbf{F}_t|\boldsymbol{X}_{ns}] = (\boldsymbol{\iota}_{t,s}' \otimes \mathbf{I}_r)\{\widehat{\boldsymbol{K}}_{ns}^{(0)}\}\boldsymbol{X}_{ns}\\ &= (\boldsymbol{\iota}_{t,s}' \otimes \mathbf{I}_r)\left\{\mathbf{I}_s \otimes \widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)}\right\}^{-1}\left\{\mathbf{I}_s \otimes \widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\right\}\boldsymbol{X}_{ns}\\ &\quad + (\boldsymbol{\iota}_{t,s}' \otimes \mathbf{I}_r)\left[\{\widehat{\boldsymbol{K}}_{ns}^{(0)}\}^{-1} - \left\{\mathbf{I}_s \otimes (\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)})^{-1}(\mathbf{I}_s \otimes \widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\right\}\right]\boldsymbol{X}_{ns},\end{aligned} \tag{D.36}$$

where $\boldsymbol{\iota}_{t,s}'$ is the tth row of $\mathbf{I}_s$ and $\boldsymbol{X}_{ns} = (\mathbf{x}_{n1}' \cdots \mathbf{x}_{ns}')'$ is an ns-dimensional vector.

Then, by (C.6) in the proof of Lemma C.5, we have

$$\begin{aligned}&n^{3/2}s\left\|\{\widehat{\boldsymbol{K}}_{ns}^{(0)}\}^{-1} - \left\{\mathbf{I}_s \otimes (\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)})^{-1}(\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1})\right\}\right\|\\ &\le n^2 s\left\|\left\{\mathbf{I}_s \otimes (\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)})^{-1} + (\widehat{\mathbf{\Omega}}_s^{F(0)})^{-1}\right\} - \left\{\mathbf{I}_s \otimes (\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)})^{-1}\right\}\right\|\\ &\quad \cdot n^{-1/2}\left\|\mathbf{I}_s \otimes (\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1})\right\|\\ &= n^2 s\left\|\left\{\mathbf{I}_s \otimes (\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)})^{-1} + (\widehat{\mathbf{\Omega}}_s^{F(0)})^{-1}\right\}(\widehat{\mathbf{\Omega}}_s^{F(0)})^{-1}\left\{\mathbf{I}_s \otimes (\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)})^{-1}\right\}\right\|\\ &\quad \cdot n^{-1/2}\left\|\mathbf{I}_s \otimes (\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1})\right\| = O_p(1).\end{aligned} \tag{D.37}$$

Indeed, for the first term on the rhs of (D.37), by Lemmas C.3(i), C.3(v), and D.5(i), and by (D.34), we have

$$\begin{aligned}&n^2 s\left\|\left\{\mathbf{I}_s \otimes (\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)})^{-1} + (\widehat{\mathbf{\Omega}}_s^{F(0)})^{-1}\right\}(\widehat{\mathbf{\Omega}}_s^{F(0)})^{-1}\left\{\mathbf{I}_s \otimes (\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)})^{-1}\right\}\right\|\\ &\le n\left\|\left\{\mathbf{I}_s \otimes (\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)})^{-1} + (\widehat{\mathbf{\Omega}}_s^{F(0)})^{-1}\right\}\right\| s\|(\widehat{\mathbf{\Omega}}_s^{F(0)})^{-1}\|\, n\|\mathbf{I}_s \otimes (\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)})^{-1}\|\\ &= O_p(1).\end{aligned} \tag{D.38}$$

Alternatively to bound the first term on the rhs of (D.38) we can use directly Lemma D.5(iii) and D.34.

And, for the second term on the rhs of (D.37), by Lemmas C.3(vii) and D.5(ii), we have

$$n^{-1/2}\|\mathbf{I}_s \otimes (\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1})\| = O_p(1). \tag{D.39}$$

Now, let $\widehat{\mathbf{\Pi}}_s^{(0)} = \left[\{\widehat{\boldsymbol{K}}_{ns}^{(0)}\}^{-1} - \left\{\mathbf{I}_s \otimes (\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)})^{-1}(\mathbf{I}_s \otimes \widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\right\}\right]$, and similarly define $\mathbf{\Pi}_s$ when using the true parameters. By Lemmas D.1(ii), D.3, D.5(ii), and D.5(iv), and (D.37), we have

$$n^{3/2}s\|\widehat{\mathbf{\Pi}}_s^{(0)} - \mathbf{\Pi}_s\| = O_p(\max(n^{-1}, T^{-1/2})). \tag{D.40}$$

Moreover, $\mathbf{\Pi}_s$ is $rs \times ns$ such that

$$(\boldsymbol{\iota}_{t,s}' \otimes \mathbf{I}_r)\mathbf{\Pi}_s \boldsymbol{X}_{ns} = (\boldsymbol{\iota}_{t,s}' \otimes \mathbf{I}_r) \begin{pmatrix} \sum_{\tau=1}^s \boldsymbol{\pi}_{1\tau}\mathbf{x}_{n\tau} \\ \vdots \\ \sum_{\tau=1}^s \boldsymbol{\pi}_{s\tau}\mathbf{x}_{n\tau} \end{pmatrix} = \left(\sum_{\tau=1}^s \boldsymbol{\pi}_{1\tau}\mathbf{x}_{n\tau} \cdots \sum_{\tau=1}^s \boldsymbol{\pi}_{s\tau}\mathbf{x}_{n\tau} \right) \boldsymbol{\iota}_{t,s}$$
$$= \sum_{\tau=1}^s \boldsymbol{\pi}_{t\tau}\mathbf{x}_{n\tau}, \tag{D.41}$$

with $\boldsymbol{\pi}_{t\tau}$, $t, \tau = 1, \dots, n$ being $r \times n$ sub-block of $\mathbf{\Pi}_s$ and since $\text{vec}(\boldsymbol{ABC}) = (\boldsymbol{C}' \otimes \boldsymbol{A})\text{vec}(\boldsymbol{B})$. Furthermore, since clearly from (D.33)

$$\max_{t=1,\dots,s} \sqrt{s}\|(\boldsymbol{\iota}_{t,s}' \otimes \mathbf{I}_r)[(\mathbf{\Omega}_s)^{-1}]\| = O(1), \qquad \max_{\tau=1,\dots,s} \sqrt{s}\|[(\mathbf{\Omega}_s)^{-1}](\boldsymbol{\iota}_{\tau,s} \otimes \mathbf{I}_r)\| = O(1), \tag{D.42}$$

using the same reasoning as in (D.38) and (D.39), but when considering the true parameters, we also have that $\max_{t,\tau=1,\dots,s} \|\boldsymbol{\pi}_{t\tau}\| = O(n^{-3/2}s^{-1/2})$. Thus, from (D.41) we have

$$\|(\boldsymbol{\iota}_{t,s}' \otimes \mathbf{I}_r)\mathbf{\Pi}_s \boldsymbol{X}_{ns}\| \le \max_{t,\tau=1,\dots,s} \|\boldsymbol{\pi}_{t\tau}\| \left\| \sum_{\tau=1}^s \mathbf{x}_{n\tau} \right\| = O_p(n^{-1}), \tag{D.43}$$

since, by Assumption 1(d) and Lemma C.1(iii)

$$\begin{aligned}
\mathbb{E}\left[\left\| \sum_{\tau=1}^s \mathbf{x}_{n\tau} \right\|^2 \right] &= \sum_{i=1}^n \sum_{\tau_1=1}^s \sum_{\tau_2=1}^s \mathbb{E}[x_{i\tau_1}x_{i\tau_2}] = \sum_{i=1}^n \sum_{\tau_1=1}^s \sum_{\tau_2=1}^s \boldsymbol{\lambda}_i'\mathbb{E}[\mathbf{F}_{\tau_1}\mathbf{F}_{\tau_2}]\boldsymbol{\lambda}_i + \mathbb{E}[\xi_{i\tau_1}\xi_{i\tau_2}] \\
&\le \sum_{i=1}^n \sum_{\tau_1=1}^s \sum_{\tau_2=1}^s \mathbb{E}[x_{i\tau_1}x_{i\tau_2}] = \sum_{i=1}^n \sum_{\tau_1=1}^s \sum_{\tau_2=1}^s |\boldsymbol{\lambda}_i'\mathbf{A}^{|\tau_1-\tau_2|}\boldsymbol{\lambda}_i| + |\mathbb{E}[\xi_{i\tau_1}\xi_{i\tau_2}]| \\
&\le nM_\lambda^2 \left\{ \sum_{\tau_1=1}^s \sum_{\tau_2=1}^s \|\mathbf{A}^{|\tau_1-\tau_2|}\| + \sum_{\tau_1=1}^s \sum_{\tau_2=1}^s |\mathbb{E}[\xi_{i\tau_1}\xi_{i\tau_2}]| \right\} \\
&\le nsM_\lambda^2 \sum_{k=-(s-1)}^{s-1} \left(1 - s^{-1}|k|\right) M_A^k + nsM_\lambda^2 M_3 \\
&\le nsM_\lambda^2 2\sum_{k=0}^{s-1} M_A^k + nsM_\lambda^2 M_3 \\
&\le nsM_\lambda^2 2(1-M_A)^{-1} + nsM_\lambda^2 M_3.
\end{aligned}$$

Thus, from (D.36), (D.37), (D.40), and (D.43), and since $\text{vec}(\boldsymbol{ABC}) = (\boldsymbol{C}' \otimes \boldsymbol{A})\text{vec}(\boldsymbol{B})$ and $\|\boldsymbol{\iota}_{t,s}' \otimes \mathbf{I}_r\| = 1$,

$$\begin{aligned}
&\|\mathbf{F}_{t|s}^{(0)}\| \le \|(\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\mathbf{x}_t\| \\
&\quad + \left\|(\boldsymbol{\iota}_{t,s}' \otimes \mathbf{I}_r)\mathbf{\Pi}_s\boldsymbol{X}_{ns}\right\| + \|\boldsymbol{\iota}_{t,s}' \otimes \mathbf{I}_r\| \, \|\widehat{\mathbf{\Pi}}_s^{(0)} - \mathbf{\Pi}_s\| \, \|\boldsymbol{X}_{ns}\| \\
&\le n\|(\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)})^{-1}\| \, n^{-1/2}\|\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\| \, n^{-1/2}\|\mathbf{x}_t\| + O_p(n^{-1}) + o_p(n^{-1}s^{-1/2}) \\
&= O_p(1),
\end{aligned} \tag{D.44}$$

where we used also Lemmas C.3(v), C.3(vii), C.10, D.5(i), and D.5(ii), and the facts that $\|\boldsymbol{X}_{ns}\| = O(\sqrt{ns})$ by the same reasoning as in Lemma C.10, and $\|\boldsymbol{\iota}_{t,s}' \otimes \mathbf{I}_r\| = 1$. This completes the proof. □

Lemma D.15. *Under Assumptions 1, 2, 3, and 6, as $n, T \to \infty$, $n\|\mathbf{F}_{t|T}^{(0)} - \mathbf{F}_{t|t}^{(0)}\| = O_p(1)$, uniformly in t.*

Proof. From (A.6) and (A.1)

$$\begin{aligned}
\|\mathbf{F}_{t|T}^{(0)} - \mathbf{F}_{t|t}^{(0)}\| &\le \|\mathbf{P}_{t|t}^{(0)}\| \, \|\widehat{\mathbf{A}}^{(0)}\| \, \|(\mathbf{P}_{t+1|t}^{(0)})^{-1}\| \, \{\|\mathbf{F}_{t+1|T}^{(0)}\| + \|\mathbf{F}_{t+1|t}^{(0)}\|\} \\
&\le \|\mathbf{P}_{t|t}^{(0)}\| \, \|\widehat{\mathbf{A}}^{(0)}\| \, \|(\mathbf{P}_{t+1|t}^{(0)})^{-1}\| \, \{\|\mathbf{F}_{t+1|T}^{(0)}\| + \|\widehat{\mathbf{A}}^{(0)}\| \, \|\mathbf{F}_{t|t}^{(0)}\|\} \\
&= O_p(n^{-1}),
\end{aligned} \tag{D.45}$$

by Lemmas D.8(ii), D.11, and D.14 (when $s=T$ and $s=t$), and since $\|\widehat{\mathbf{A}}^{(0)}\|\le\|\mathbf{A}\|+\|\widehat{\mathbf{A}}^{(0)}-\mathbf{A}\|=O_p(1)$, by Assumption 1(d) and Lemma D.3(i). This completes the proof.

Lemma D.16. *Under Assumptions 1, 2, 3, and 6, as $n,T\to\infty$, $n\|\mathbf{F}_{t|t}^{(0)}-\mathbf{F}_t^{\mathrm{WLS}(0)}\|=O_p(1)$, uniformly in t, where $\mathbf{F}_t^{\mathrm{WLS}(0)}=(\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\mathbf{x}_{nt}$.*

Proof. From (A.3) and (A.1), by Lemma D.13

$$\begin{aligned}\mathbf{F}_{t|t}^{(0)}&=\mathbf{F}_{t|t-1}^{(0)}+\mathbf{P}_{t|t-1}^{(0)}\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Lambda}}_n^{(0)}\mathbf{P}_{t|t-1}^{(0)}\widehat{\mathbf{\Lambda}}_n^{(0)\prime}+\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}(\mathbf{x}_{nt}-\widehat{\mathbf{\Lambda}}_n^{(0)}\mathbf{F}_{t|t-1}^{(0)})\\&=(\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)}+(\mathbf{P}_{t|t-1}^{(0)})^{-1})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\mathbf{x}_{nt}\\&\quad+\left\{\mathbf{I}_r-(\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)}+(\widehat{\mathbf{P}}_{t|t-1}^{(0)})^{-1})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)}\right\}\widehat{\mathbf{A}}^{(0)}\mathbf{F}_{t-1|t-1}^{(0)}\\&=(\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\mathbf{x}_{nt}\\&\quad+\left\{(\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)}+(\mathbf{P}_{t|t-1}^{(0)})^{-1})^{-1}-(\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)})^{-1}\right\}\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\mathbf{x}_{nt}\\&\quad+\left\{\mathbf{I}_r-(\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)}+(\widehat{\mathbf{P}}_{t|t-1}^{(0)})^{-1})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)}\right\}\widehat{\mathbf{A}}^{(0)}\mathbf{F}_{t-1|t-1}^{(0)}.\end{aligned}\tag{D.46}$$

Notice that the inverses in (D.46) are all well defined by Lemmas D.5(iii) and D.8(ii) and (D.11) in the proof of Lemma D.4.

Now, by Lemma C.5

$$\|(\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)}+(\mathbf{P}_{t|t-1}^{(0)})^{-1})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)}-\mathbf{I}_r\|=O_p(n^{-1}).\tag{D.47}$$

Furthermore, by Lemmas C.6(iii) and D.5(iv)

$$\|(\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)}+(\mathbf{P}_{t|t-1}^{(0)})^{-1})^{-1}-(\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)})^{-1}\|=O(n^{-2}),\tag{D.48}$$

and by Lemmas C.3(vii) and D.5(ii),

$$\|\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\|=O_p(\sqrt{n}).\tag{D.49}$$

Indeed, we can apply Lemmas C.5 and C.6(iii), since $\|(\mathbf{P}_{t|t-1}^{(0)})^{-1}\|=O_p(1)$ by Lemma D.8(ii), $\|(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\|=O_p(1)$ by (D.11) in the proof of Lemma D.4, and, by Lemmas C.2 and D.1(ii) we have

$$\begin{aligned}n^{-1}\|\widehat{\mathbf{\Lambda}}_n^{(0)\prime}\widehat{\mathbf{\Lambda}}_n^{(0)}-\mathbf{\Lambda}_n'\mathbf{\Lambda}_n\|&\le 2n^{-1}\|\mathbf{\Lambda}_n'(\widehat{\mathbf{\Lambda}}_n^{(0)}-\mathbf{\Lambda}_n)\|+n^{-1}\|(\widehat{\mathbf{\Lambda}}_n^{(0)}-\mathbf{\Lambda}_n)'(\widehat{\mathbf{\Lambda}}_n^{(0)}-\mathbf{\Lambda}_n)\|\\&\le 2n^{-1/2}\|\mathbf{\Lambda}_n\|\,n^{-1/2}\|\widehat{\mathbf{\Lambda}}_n^{(0)}-\mathbf{\Lambda}_n\|+n^{-1}\|\widehat{\mathbf{\Lambda}}_n^{(0)}-\mathbf{\Lambda}_n\|^2\\&=O_p(\max(n^{-1},T^{-1/2})).\end{aligned}$$

which, by Weyl's inequality (Merikoski and Kumar, 2004, Theorem 1), implies

$$n^{-1}|\nu^{(j)}(\widehat{\mathbf{\Lambda}}_n^{(0)\prime}\widehat{\mathbf{\Lambda}}_n^{(0)})-\nu^{(j)}(\mathbf{\Lambda}_n'\mathbf{\Lambda}_n)|\le n^{-1}\|\widehat{\mathbf{\Lambda}}_n^{(0)\prime}\widehat{\mathbf{\Lambda}}_n^{(0)}-\mathbf{\Lambda}_n'\mathbf{\Lambda}_n\|=O_p(\max(n^{-1},T^{-1/2})),$$

and, therefore, for $j=1,\dots,r,$,

$$\underline{C}_j\le \operatorname{p-liminf}_{n,T\to\infty} n^{-1}\nu^{(j)}(\widehat{\mathbf{\Lambda}}_n^{(0)\prime}\widehat{\mathbf{\Lambda}}_n^{(0)})\le \operatorname{p-limsup}_{n,T\to\infty} n^{-1}\nu^{(j)}(\widehat{\mathbf{\Lambda}}_n^{(0)\prime}\widehat{\mathbf{\Lambda}}_n^{(0)})\le\overline{C}_j.$$

By using (D.47), (D.48), (D.49) into (D.46):

$$\begin{aligned}&\|\mathbf{F}_{t|t}^{(0)}-(\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\mathbf{x}_{nt}\|\\&\le n\|(\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)}+(\mathbf{P}_{t|t-1}^{(0)})^{-1})^{-1}-(\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)})^{-1}\|\,n^{-1/2}\|\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\|\,n^{-1/2}\|\mathbf{x}_{nt}\|\\&\quad+\|\mathbf{I}_r-(\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)}+(\widehat{\mathbf{P}}_{t|t-1}^{(0)})^{-1})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)}\|\,\|\widehat{\mathbf{A}}^{(0)}\|\,\|\mathbf{F}_{t-1|t-1}^{(0)}\|\\&=O_p(n^{-1}),\end{aligned}\tag{D.50}$$

by Lemmas C.10 and D.14 (when $s=t-1$), and since $\|\widehat{\mathbf{A}}^{(0)}\|\le\|\mathbf{A}\|+\|\widehat{\mathbf{A}}^{(0)}-\mathbf{A}\|=O_p(1)$, by Assumption 1(d) and

Lemma D.3(i). This completes the proof. □

Lemma D.17. *Under Assumptions 1, 2, 3, and 6, as $n,T\to\infty$, $\min(\sqrt{n},\sqrt{T})\|\mathbf{F}_{t|T}^{(0)}-\mathbf{F}_t\|=O_p(1)$, uniformly in t.*

PROOF. From Lemmas D.15 and D.16

$$\begin{aligned}\|\mathbf{F}_{t|T}^{(0)}-\mathbf{F}_t\| &\le \|\mathbf{F}_{t|T}^{(0)}-\mathbf{F}_{t|t}^{(0)}\|+\|\mathbf{F}_{t|t}^{(0)}-\widehat{\mathbf{F}}_t^{\text{WLS}(0)}\|+\|\widehat{\mathbf{F}}_t^{\text{WLS}(0)}-\mathbf{F}_t\|\\ &=\|\widehat{\mathbf{F}}_t^{\text{WLS}(0)}-\mathbf{F}_t\|+O_p(n^{-1}).\end{aligned}\tag{D.51}$$

where $\mathbf{F}_t^{\text{WLS}(0)}=(\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\mathbf{x}_{nt}$.

Now,

$$\begin{aligned}\|\widehat{\mathbf{F}}_t^{\text{WLS}(0)}-\mathbf{F}_t\| \le\,& \|(\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}(\mathbf{\Lambda}_n-\widehat{\mathbf{\Lambda}}_n^{(0)})\|\,\|\mathbf{F}_t\|\\ &+\|(\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\boldsymbol{\xi}_{nt}\|\\ \le\,& \|(\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^{\xi})^{-1}\mathbf{\Lambda}_n)^{-1}\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^{\xi})^{-1}(\mathbf{\Lambda}_n-\widehat{\mathbf{\Lambda}}_n^{(0)})\|\,\|\mathbf{F}_t\|\\ &+\|(\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}-(\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^{\xi})^{-1}\mathbf{\Lambda}_n)^{-1}\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^{\xi})^{-1}\|\\ &\cdot\|\mathbf{\Lambda}_n-\widehat{\mathbf{\Lambda}}_n^{(0)}\|\,\|\mathbf{F}_t\|+\|(\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^{\xi})^{-1}\mathbf{\Lambda}_n)^{-1}\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^{\xi})^{-1}\boldsymbol{\xi}_{nt}\|\\ &+\|(\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(0)})^{-1}\boldsymbol{\xi}_{nt}-(\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^{\xi})^{-1}\mathbf{\Lambda}_n)^{-1}\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^{\xi})^{-1}\boldsymbol{\xi}_{nt}\|\\ =\,& A+B+C+D,\quad\text{say.}\end{aligned}\tag{D.52}$$

Let us consider each term in (D.52). First, consider term A and let $\mathbf{\Lambda}_n^{\text{OLS}}=(\sum_{t=1}^T\mathbf{x}_{nt}\mathbf{F}_t')(\sum_{t=1}^T\mathbf{F}_t\mathbf{F}_t')^{-1}$. Then, from Barigozzi (2023, Corollary 1)

$$n^{-1/2}\|\widehat{\mathbf{\Lambda}}_n^{(0)}-\mathbf{\Lambda}_n^{\text{OLS}}\|=O_p(\max(n^{-1},n^{-1/2}T^{-1/2})).\tag{D.53}$$

Therefore, from (D.53)

$$\begin{aligned}A\le\,&\|(\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^{\xi})^{-1}\mathbf{\Lambda}_n)^{-1}\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^{\xi})^{-1}(\mathbf{\Lambda}_n-\mathbf{\Lambda}_n^{\text{OLS}})\|\,\|\mathbf{F}_t\|\\ &+n\|(\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^{\xi})^{-1}\mathbf{\Lambda}_n)^{-1}\|\,n^{-1/2}\|\mathbf{\Lambda}_n\|\,\|(\mathbf{\Sigma}_n^{\xi})^{-1}\|\,n^{-1/2}\|\widehat{\mathbf{\Lambda}}_n^{(0)}-\mathbf{\Lambda}_n^{\text{OLS}}\|\,\|\mathbf{F}_t\|\\ =\,&\{A.1+A.2\}\|\mathbf{F}_t\|,\quad\text{say.}\end{aligned}\tag{D.54}$$

Then,

$$\begin{aligned}A.1\le\,& n\|(\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^{\xi})^{-1}\mathbf{\Lambda}_n)^{-1}\|\,n^{-1}\|\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^{\xi})^{-1}(\mathbf{\Lambda}_n-\mathbf{\Lambda}_n^{\text{OLS}})\|\\ =\,& n\|(\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^{\xi})^{-1}\mathbf{\Lambda}_n)^{-1}\|\,n^{-1}\left\|T^{-1}\sum_{t=1}^T\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^{\xi})^{-1}\boldsymbol{\xi}_{nt}\mathbf{F}_t'\right\|\left\|\left(T^{-1}\sum_{t=1}^T\mathbf{F}_t\mathbf{F}_t'\right)^{-1}\right\|\\ =\,& O_p(n^{-1/2}T^{-1/2}),\end{aligned}\tag{D.55}$$

by Lemmas C.3(iii) and C.8(iv), and also, recalling that $\mathbf{\Gamma}^F=\mathbf{I}_r$ by Assumption 6(b), by Lemma C.12(i) and Weyl's inequality (Merikoski and Kumar, 2004, Theorem 1) we have $|\nu^{(r)}(T^{-1}\sum_{t=1}^T\mathbf{F}_t\mathbf{F}_t')-1|=O_p(T^{-1/2})$ which implies $\|(T^{-1}\sum_{t=1}^T\mathbf{F}_t\mathbf{F}_t')^{-1}\|=O_p(1)$.

Moreover, $A.2=O_p(\max(n^{-1},n^{-1/2}T^{-1/2}))$, because of (D.53) and Lemmas C.2, C.3(iii), and Assumption 2(a) which implies $\|(\mathbf{\Sigma}_n^{\xi})^{-1}\|\le C_\xi$. This, jointly with (D.54) and (D.55) implies that

$$A=O_p(\max(n^{-1},n^{-1/2}T^{-1/2})),\tag{D.56}$$

since $\|\mathbf{F}_t\|=O_p(1)$ because $\mathbb{E}[F_{jt}^2]=1$, $j=1,\dots,r$, by Assumption 6(b).

Second, by Lemmas D.1(ii) and D.5(v),

$$\begin{aligned}B =& \|n(\widehat{\boldsymbol{\Lambda}}_n^{(0)\prime}(\widehat{\boldsymbol{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\boldsymbol{\Lambda}}_n^{(0)})^{-1}n^{-1/2}\widehat{\boldsymbol{\Lambda}}_n^{(0)\prime}(\widehat{\boldsymbol{\Sigma}}_n^{\xi(0)})^{-1} - n(\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^{\xi})^{-1}\boldsymbol{\Lambda}_n)^{-1}n^{-1/2}\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^{\xi})^{-1}\| \\ &\cdot n^{-1/2}\|\boldsymbol{\Lambda}_n - \widehat{\boldsymbol{\Lambda}}_n^{(0)}\|\,\|\mathbf{F}_t\| \\ =& O_p(\max(n^{-2}, T^{-1})), \end{aligned} \tag{D.57}$$

and since $\|\mathbf{F}_t\| = O_p(1)$ because $\mathbb{E}[F_{jt}^2] = 1$, $j = 1, \dots, r$, by Assumption 6(b). Third,

$$C \le n\|(\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^{\xi})^{-1}\boldsymbol{\Lambda}_n)^{-1}\|\,n^{-1}\|\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^{\xi})^{-1}\boldsymbol{\xi}_{nt}\| = O_p(n^{-1/2}), \tag{D.58}$$

by Lemmas C.3(iii) and C.7(i). Fourth, and last,

$$\begin{aligned}D \le& n\|(\widehat{\boldsymbol{\Lambda}}_n^{(0)\prime}(\widehat{\boldsymbol{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\boldsymbol{\Lambda}}_n^{(0)})^{-1} - (\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^{\xi})^{-1}\boldsymbol{\Lambda}_n)^{-1}\|\,n^{-1}\|\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^{\xi})^{-1}\boldsymbol{\xi}_{nt}\| \\ &+ n\|(\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^{\xi})^{-1}\boldsymbol{\Lambda}_n)^{-1}\|\,n^{-1/2}\|\widehat{\boldsymbol{\Lambda}}_n^{(0)\prime}(\widehat{\boldsymbol{\Sigma}}_n^{\xi(0)})^{-1} - \boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^{\xi})^{-1}\|\,n^{-1/2}\|\boldsymbol{\xi}_{nt}\| \\ &+ n\|(\widehat{\boldsymbol{\Lambda}}_n^{(0)\prime}(\widehat{\boldsymbol{\Sigma}}_n^{\xi(0)})^{-1}\widehat{\boldsymbol{\Lambda}}_n^{(0)})^{-1} - (\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^{\xi})^{-1}\boldsymbol{\Lambda}_n)^{-1}\|\,n^{-1/2}\|\widehat{\boldsymbol{\Lambda}}_n^{(0)\prime}(\widehat{\boldsymbol{\Sigma}}_n^{\xi(0)})^{-1} - \boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^{\xi})^{-1}\|\,n^{-1/2}\|\boldsymbol{\xi}_{nt}\| \\ =& D.1 + D.2 + D.3, \text{ say.}\end{aligned}$$

Then, $D.1 = O_p(\max(n^{-3/2}, n^{-1/2}T^{-1/2}))$, by Lemmas C.7(i) and D.5(iv), while we have $D.2 = O_p(\max(n^{-1}, T^{-1/2}))$, because of Lemmas C.3(iii) and D.5(ii) and since $\|\boldsymbol{\xi}_{nt}\| = O_p(\sqrt{n})$ because $\mathbb{E}[\xi_{it}^2] = \sigma_i^2 \le C_\xi$ by Assumption 2(a). Last $D.3 = O_p(n^{-2}, T^{-1})$ by Lemmas D.5(ii) and D.5(iv) and since $\|\boldsymbol{\xi}_{nt}\| = O_p(\sqrt{n})$. Therefore,

$$D = O_p(\max(n^{-1}, T^{-1/2})). \tag{D.59}$$

By substituting (D.56), (D.57), (D.58), and (D.59), into (D.52) and then into (D.51), we complete the proof. □

Lemma D.18. *Under Assumptions 1, 2, 3, and 6, as $n, T \to \infty$, for $s = t$ and $s = T$:*
(i) $\|T^{-1}\sum_{t=1}^T \mathbf{F}_{t|s}^{(0)}\mathbf{F}_t'\| = O_p(1)$;
(i) $\|T^{-1}\sum_{t=1}^T \mathbf{F}_{t|s}^{(0)}\mathbf{F}_t'\boldsymbol{\lambda}_i\| = O_p(1)$, *uniformly in i;*
(ii) $\|T^{-1}\sum_{t=1}^T \mathbf{F}_{t|s}^{(0)}\xi_{it}\| = O_p(1)$, *uniformly in i.*

Proof. First notice that, for all $k = t - T, \dots, t - 1$,

$$\left\| n^{-1/2}T^{-1}\sum_{t=1}^T \mathbf{x}_{n,t-k}\mathbf{F}_t' \right\| = O_p(1), \tag{D.60}$$

by Lemma C.12. The proof of part (i) follows by iterating either forward or backwards since both $\|T^{-1}\sum_{t=1}^T \mathbf{F}_{t|t}^{(0)}\mathbf{F}_t'\|$ and $\|T^{-1}\sum_{t=1}^T \mathbf{F}_{t|T}^{(0)}\mathbf{F}_t'\|$ are functions of (D.60), because of Lemmas D.15 and D.16.

Part (ii) follows from part (i) and Assumption 1(a). Part (iii) follows by substituting $\mathbf{F}_t$ with ξ_{it} in (E.93) and then by applying Lemma C.12(ii). This completes the proof. □

Lemma D.19. *Let $\boldsymbol{\phi}_n^{\rm OLS} = (\text{vec}(\boldsymbol{\Lambda}_n^{\rm OLS})'\ \sigma_1^{2\rm OLS} \cdots \widehat{\sigma}_n^{2\rm OLS})'$ be the vector of OLS estimators of the entries of $\boldsymbol{\phi}_n$ obtained when $\mathbf{F}_t$ is known, that is whose entries maximize $\ell(\boldsymbol{X}_{nT}|\boldsymbol{F}_T;\underline{\boldsymbol{\phi}}_n)$. Let also $\boldsymbol{\theta}^{\rm OLS} = (\text{vec}(\mathbf{A}^{\rm OLS})'\ \text{vech}(\boldsymbol{\Gamma}^{v\rm OLS}))'$ be the vector of OLS estimators of the entries of $\boldsymbol{\theta}$ obtained when $\mathbf{F}_t$ is known, that is whose entries maximize $\ell(\boldsymbol{F}_T;\underline{\boldsymbol{\theta}})$. Then, under Assumptions 1, 2, 3, and 6, as $n, T \to \infty$,*
(i) $\sqrt{T}\,\|\boldsymbol{\lambda}_i^{\rm OLS} - \boldsymbol{\lambda}_i\| = O_p(1)$, *uniformly in i;*
(ii) $\sqrt{T}\,n^{-1/2}\|\boldsymbol{\Lambda}_n^{\rm OLS} - \boldsymbol{\Lambda}_n\| = O_p(1)$;
(iii) $\sqrt{T}|\sigma_i^{2\rm OLS} - \sigma_i^2| = O_p(1)$;
(iv) $\sqrt{T}\|\widehat{\mathbf{A}}^{\rm OLS} - \mathbf{A}\| = O_p(1)$;
(v) $\sqrt{T}\|\widehat{\boldsymbol{\Gamma}}^{v\rm OLS} - \boldsymbol{\Gamma}^v\| = O_p(1)$.

Proof. For part (i),

$$\|\boldsymbol{\lambda}_i^{\rm OLS} - \boldsymbol{\lambda}_i\| \le \left\|\left(T^{-1}\sum_{t=1}^T \mathbf{F}_t\mathbf{F}_t'\right)^{-1}\right\| \left\|T^{-1}\sum_{t=1}^T \mathbf{F}_t x_{it}\right\| = O_p(T^{-1/2}), \tag{D.61}$$

where for the numerator we used Lemma C.12(ii) while for the denominator we used Lemma C.13.

Part (ii) is proved in the same way but using Lemma C.12(iii) instead of Lemma C.12(ii).

For part (iii), first notice that

$$
\begin{aligned}
\sigma_i^{2\text{OLS}} &= T^{-1}\sum_{t=1}^T \xi_{it}^{\text{OLS}2} = T^{-1}\sum_{t=1}^T (x_{it} - \boldsymbol{\lambda}_i^{\text{OLS}\prime}\mathbf{F}_t)^2 \\
&= T^{-1}\sum_{t=1}^T \{x_{it}^2 + \boldsymbol{\lambda}_i'\mathbf{F}_t\mathbf{F}_t'\boldsymbol{\lambda}_i - 2x_{it}\boldsymbol{\lambda}_i'\mathbf{F}_t\} \\
&\quad + T^{-1}\sum_{t=1}^T \{2\boldsymbol{\lambda}_i^{\text{OLS}\prime}\mathbf{F}_t\mathbf{F}_t'(\boldsymbol{\lambda}_i^{\text{OLS}} - \boldsymbol{\lambda}_i) + (\boldsymbol{\lambda}_i^{\text{OLS}} - \boldsymbol{\lambda}_i)'\ \mathbf{F}_t\mathbf{F}_t'(\boldsymbol{\lambda}_i^{\text{OLS}} - \boldsymbol{\lambda}_i) - 2x_{it}(\boldsymbol{\lambda}_i^{\text{OLS}} - \boldsymbol{\lambda}_i)'\mathbf{F}_t\} \\
&= T^{-1}\sum_{t=1}^T (x_{it} - \boldsymbol{\lambda}_i'\mathbf{F}_t)^2 \\
&\quad + T^{-1}\sum_{t=1}^T \{2\boldsymbol{\lambda}_i^{\text{OLS}\prime}\mathbf{F}_t\mathbf{F}_t'(\boldsymbol{\lambda}_i^{\text{OLS}} - \boldsymbol{\lambda}_i) + (\boldsymbol{\lambda}_i^{\text{OLS}} - \boldsymbol{\lambda}_i)'\ \mathbf{F}_t\mathbf{F}_t'(\boldsymbol{\lambda}_i^{\text{OLS}} - \boldsymbol{\lambda}_i) \\
&\quad - 2(\boldsymbol{\lambda}_i^{\text{OLS}} - \boldsymbol{\lambda}_i)'\mathbf{F}_t\mathbf{F}_t'\boldsymbol{\lambda}_i^{\text{OLS}} - 2(\boldsymbol{\lambda}_i^{\text{OLS}} - \boldsymbol{\lambda}_i)'\mathbf{F}_t\xi_{it}^{\text{OLS}}\} \\
&= T^{-1}\sum_{t=1}^T \xi_{it}^2 + T^{-1}\sum_{t=1}^T (\boldsymbol{\lambda}_i^{\text{OLS}} - \boldsymbol{\lambda}_i)'\ \mathbf{F}_t\mathbf{F}_t'(\boldsymbol{\lambda}_i^{\text{OLS}} - \boldsymbol{\lambda}_i),
\end{aligned} \tag{D.62}
$$

since by construction $\sum_{t=1}^T \mathbf{F}_t\xi_{it}^{\text{OLS}} = \mathbf{0}_r$. Then,

$$
\begin{aligned}
|\sigma_i^{2\text{OLS}} - \sigma_i^2| &\le \left| T^{-1}\sum_{t=1}^T (x_{it} - \boldsymbol{\lambda}_i^{\text{OLS}\prime}\mathbf{F}_t)^2 - T^{-1}\sum_{t=1}^T (x_{it} - \boldsymbol{\lambda}_i'\mathbf{F}_t)^2 \right| + \left| T^{-1}\sum_{t=1}^T (x_{it} - \boldsymbol{\lambda}_i'\mathbf{F}_t)^2 - \mathbb{E}[\xi_{it}^2] \right| \\
&\le \|\boldsymbol{\lambda}_i^{\text{OLS}} - \boldsymbol{\lambda}_i\|^2 \left\| T^{-1}\sum_{t=1}^T \mathbf{F}_t\mathbf{F}_t' \right\| + \left| T^{-1}\sum_{t=1}^T \xi_{it}^2 - \mathbb{E}[\xi_{it}^2] \right| \\
&= O_p(T^{-1/2}),
\end{aligned}
$$

because of part (i) and Lemma C.12(iv), and also by Lemma C.12(i) combined with Assumption 6(b).

For part (iv) consistency follows from the fact that $\{\mathbf{v}_t\}$ is an independent process by Assumption 1(f) and thus it is a martingale difference process (Hamilton, 1994, Proposition 11, pp. 298-299). Part (v) follows from part (iv) and (Hamilton, 1994, Proposition 11.2, pp. 301), since by Assumption 1(g), the fourth order cumulants of $\{\mathbf{v}_t\}$ are all finite. This completes the proof. □

E Lemmas necessary for proving Proposition 2

Lemma E.1. *Under Assumptions 1, 2, 3, and 5*

(i) For all $t \in \mathbb{Z}$, all $j = 1, \ldots, r$, and all $s > 0$, $\mathrm{P}(|F_{jt}| \ge s) \le \exp\{-K_F s^{\delta_v}\}$, for some finite positive real K_F independent of t and j;

(ii) for all $i \in \mathbb{N}$ and all $T \in \mathbb{N}$,

$$
\mathrm{P}\left(\left\| T^{-1/2}\sum_{t=1}^T \mathbf{F}_t\xi_{it} \right\| \ge s \right) \le r\exp\{-\kappa_3 s^2\} + rT\exp\{-\kappa_4(s\sqrt{T})^\beta\},
$$

for some finite positive reals κ_3, κ_4, and β independent of i and T and such that $\frac{1}{\beta} = \frac{1}{\gamma} + \frac{1}{\delta} > 1$ and $\gamma = \min(\gamma_F, \gamma_\xi)$ and $\delta \in (0, \frac{\delta_v\delta_\xi}{\delta_v + \delta_\xi})$.

Proof. For all $j = 1, \ldots, r$,

$$
|F_{jt}| \le \sum_{\ell=1}^r \left| \sum_{k=0}^\infty [\mathbf{A}^k]_{j\ell} v_{\ell,t-k} \right| \le r \max_{\ell=1,\ldots,r} \left| \sum_{k=0}^\infty [\mathbf{A}^k]_{j\ell} v_{\ell,t-k} \right|.
$$

Now, since because of Assumption 1(d) the coefficients $[\mathbf{A}^k]_{j\ell}$ are summable over k, for any $\epsilon > 0$ and $\eta > 0$ there exists a positive integer $\bar{K} = \bar{K}(\epsilon, \eta)$ independent of j, ℓ, and t such that

$$\mathrm{P}\left(\left|\sum_{k=\bar{K}+1}^{\infty}[\mathbf{A}^k]_{j\ell}v_{\ell,t-k}\right| \geq \eta\right) \leq \epsilon,$$

thus we can always find $\bar{K}$ such that we can write

$$\begin{aligned}
|F_{jt}| &\leq r \max_{\ell=1,\ldots,r}\left|\sum_{k=0}^{\bar{K}}[\mathbf{A}^k]_{j\ell}v_{\ell,t-k}\right| + \left|\sum_{k=\bar{K}+1}^{\infty}[\mathbf{A}^k]_{j\ell}v_{\ell,t-k}\right| \\
&\leq r\bar{K}\max_{\ell=1,\ldots,r}\max_{k=0,\ldots,\bar{K}}|[\mathbf{A}^k]_{j\ell}|\left|\sum_{k=0}^{\bar{K}}v_{\ell,t-k}\right| + o_p(1) \\
&\leq r\bar{K}C_A\left|\sum_{k=0}^{\bar{K}}v_{\ell,t-k}\right| + o_p(1), \qquad \text{(E.1)}
\end{aligned}$$

for some positive real C_A independent of j, where we used again Assumption 1(d) to bound the coefficients.

Then, from Assumption 5(a), (E.1), and Bakhshizadeh et al. (2023, Corollary 4 and Section III), since $\{\mathbf{v}_t\}$ is an independent process, and by Assumption 1 and the union bound, for all $j = 1, \ldots r$ and all $s > 0$ it holds that:

$$\begin{aligned}
\mathrm{P}\left(|F_{jt}| \geq s\right) &\leq \mathrm{P}\left(r\bar{K}C_A\left|\sum_{k=0}^{\bar{K}}v_{\ell,t-k}\right| \geq s\right) + o_p(1) \\
&\leq \exp\left\{-C'\left(\frac{s}{r\bar{K}C_A}\right)^2\right\} + \bar{K}\exp\left\{-C''\left(\frac{s}{r\bar{K}C_A}\right)^{\delta_v}\right\} \\
&\leq \bar{K}\exp\left\{-C'''s^{\delta_v}\right\} \leq \exp\left\{-K_F s^{\delta_v}\right\}, \qquad \text{(E.2)}
\end{aligned}$$

for some finite positive reals C', C'', C''', and K_F independent of t and j and where from the second line we omitted the second term which is negligible. This proves part (i).

For all $i = 1, \ldots, n$ and all $n \in \mathbb{N}$, consider Assumption 5(b) when $n = 1$, $\boldsymbol{\lambda}_i = 1$, and $\sigma_i^2 = 1$, then, because of Assumption 3 and (E.2), by Fan et al. (2011, Lemma A2), for all $j = 1, \ldots r$ and all $s > 0$ it holds that:

$$\mathrm{P}\left(|F_{jt}\xi_{it}| \geq s\right) \leq \exp\left\{-\kappa_{F\xi}s^{\delta}\right\}, \qquad \text{(E.3)}$$

for some finite positive reals $\kappa_{F\xi}$ and $\delta \in \left(0, \frac{\delta_v\delta_\xi}{\delta_v+\delta_\xi}\right)$ independent of i, j, and t. Moreover, because of Assumptions 1(d), 1(e), 1(f), and 1(h), by Pham and Tran (1985, Theorem 3.1), $\{\mathbf{F}_t\}$ is a strong mixing process with mixing coefficients

$$\alpha_F(T) \leq \exp\left\{-c_F T^{\gamma_F}\right\}, \qquad \text{(E.4)}$$

for all $T \in \mathbb{N}$ and some finite positive reals c_F and γ_F independent of T . Then, because of (E.4) and Assumption 2(c), by Bradley (2005, Theorem 5.1.a), we have that, for all $i \in \mathbb{N}$ and all $j = 1, \ldots, r$, $\{F_{jt}\xi_{it}\}$ is strong mixing with mixing coefficients:

$$\alpha_{F\xi}(T) \leq \alpha_F(T) + \alpha_\xi(T) \leq \exp\left\{-c_F T^{\gamma_F}\right\} + \exp\left\{-c_\xi T^{\gamma_\xi}\right\} \leq 2\exp\left\{-c_{F\xi}T^{\gamma}\right\}, \qquad \text{(E.5)}$$

for all $T \in \mathbb{N}$ and some finite positive reals $c_{F\xi}$ and $\gamma = \min(\gamma_F, \gamma_\xi)$ independent of T.

Now, since for all $i \in \mathbb{N}$ and all $j = 1, \ldots, r$, $\{F_{jt}\xi_{it}\}$ satisfies (E.3) and (E.5), we can apply the results by Merlevède et al. (2011, Theorem 1) or equivalently by Bosq (2012, Theorem 1.4, p.31), which imply that, for all $s > 0$,

$$\mathrm{P}\left(\left|T^{-1/2}\sum_{t=1}^{T}F_{jt}\xi_{it}\right| \geq s\right) \leq \exp\left\{-c_1 s^2\right\} + T\exp\left\{-c_2\left(s\sqrt{T}\right)^{\beta}\right\} \qquad \text{(E.6)}$$

for some finite positive reals c_1, c_2, and β independent of i, j, and T, where $\frac{1}{\beta} = \frac{1}{\gamma} + \frac{1}{\delta} > 1$. Finally, note that

$$\left\|T^{-1/2}\sum_{t=1}^{T}\mathbf{F}_t\xi_{it}\right\| \leq \sum_{j=1}^{r}\left|T^{-1/2}\sum_{t=1}^{T}F_{jt}\xi_{it}\right| \leq r\max_{j=1,\ldots,r}\left|T^{-1/2}\sum_{t=1}^{T}F_{jt}\xi_{it}\right|, \qquad \text{(E.7)}$$

and from (E.6) and (E.7),

$$\mathrm{P}\left(\left\|T^{-1/2}\sum_{t=1}^{T}\mathbf{F}_t\xi_{it}\right\|\geq s\right)\leq \mathrm{P}\left(r\max_{j=1,\dots,r}\left|T^{-1/2}\sum_{t=1}^{T}F_{jt}\xi_{it}\right|\geq s\right)$$
$$\leq r\exp\left\{-c_1\left(\frac{s}{r}\right)^2\right\}+rT\exp\left\{-c_2\left(\frac{s\sqrt{T}}{r}\right)^{\beta}\right\},$$

by setting $\kappa_3=\frac{c_1}{r^2}$ and $\kappa_4=\frac{c_2}{r^{\beta}}$, we prove part (ii) and complete the proof. □

Lemma E.2. *Under Assumptions 1 and 5, as $T\to\infty$,*

$$\log^{-1/\delta_v}T\max_{t=1,\dots,T}\|\mathbf{F}_t\|=O_p(1).$$

Proof. We have,

$$\max_{t=1,\dots,T}\|\mathbf{F}_t\|\leq\max_{t=1,\dots,T}\sum_{j=1}^{r}|F_{jt}|\leq\max_{t=1,\dots,T}\max_{j=1,\dots,r}|F_{jt}|=O_p(\log^{1/\delta_v}T),$$

by Lemma E.1(i) and the union bound. This completes the proof. □

Lemma E.3. *Under Assumptions 1, 2, 3, and 5, as $n,T\to\infty$,*

$$\log^{-1/\delta_v}T\max_{t=1,\dots,T}n^{-1/2}\|\mathbf{x}_{nt}\|=O_p(1).$$

Proof. By Assumption 5(b), for all $s>0$, setting $\boldsymbol{\lambda}_i=\boldsymbol{\iota}_r$ and $\sigma_i^2=1$ therein,

$$\mathrm{P}\left(n^{-1/2}\|\boldsymbol{\xi}_{nt}\|\geq s\right)\leq\exp\left\{-\mathcal{K}_\xi s^2\right\}+n\exp\left\{-K_\xi(s\sqrt{n})^{\delta_\xi}\right\}. \tag{E.8}$$

Then, by (E.8) and the union bound, for all $s>0$, it holds that:

$$\mathrm{P}\left(\max_{t=1,\dots,T}n^{-1/2}\|\boldsymbol{\xi}_{nt}\|\geq s\right)\leq T\mathrm{P}\left(n^{-1/2}\|\boldsymbol{\xi}_{nt}\|\geq s\right)$$
$$\leq T\exp\left\{-\mathcal{K}_\xi s^2\right\}+nT\exp\left\{-K_\xi(s\sqrt{n})^{\delta_\xi}\right\}. \tag{E.9}$$

Thus, from Lemmas C.2 and E.2, and (E.9)

$$\begin{aligned}\max_{t=1,\dots,T}n^{-1/2}\|\mathbf{x}_{nt}\|&\leq n^{-1/2}\|\boldsymbol{\Lambda}_n\|\max_{t=1,\dots,T}\|\mathbf{F}_t\|+\max_{t=1,\dots,T}n^{-1/2}\|\boldsymbol{\xi}_{nt}\|\\&=O_p(\log^{1/\delta_v}T)+O_p(\sqrt{\log T})+O_p(n^{-\delta_\xi/2}\max(\log^{1/\delta_\xi}n,\log^{1/\delta_\xi}T))\\&=O_p(\log^{1/\delta_v}T).\end{aligned}$$

This completes the proof. □

Lemma E.4. *Under Assumptions 1 and 2, for all $T\in\mathbb{N}$, as $n\to\infty$,*

(i) $\max_{t=1,\dots,T}n\|\mathbf{P}_{t|t}\|=O(1)$;
(ii) $\max_{t=1,\dots,T}n\|\mathbf{P}_{t|T}\|=O(1)$;
(iii) $\max_{t=1,\dots,T}n\|\mathbf{P}_{0,t|t}\|=O(1)$;
(iv) $\max_{t=1,\dots,T}n\|\mathbf{P}_{0,t|T}\|=O(1)$;
(v) $\max_{t=1,\dots,T}n^2\|\mathbf{P}_{t|T}-\mathbf{P}_{t|t}\|=O(1)$;
(vi) $\max_{t=1,\dots,T}n^2\|\mathbf{P}_{0,t|T}-\mathbf{P}_{0,t|t}\|=O(1)$.

Proof. From (A.4), using Lemma D.13 we have

$$\begin{aligned}\mathbf{P}_{t|t}&=\mathbf{P}_{t|t-1}-\mathbf{P}_{t|t-1}\boldsymbol{\Lambda}_n'(\boldsymbol{\Lambda}_n\mathbf{P}_{t|t-1}\boldsymbol{\Lambda}_n'+\boldsymbol{\Sigma}_n^{\xi})^{-1}\boldsymbol{\Lambda}_n\mathbf{P}_{t|t-1}\\&=\mathbf{P}_{t|t-1}-(\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^{\xi})^{-1}\boldsymbol{\Lambda}_n+\mathbf{P}_{t|t-1}^{-1})^{-1}\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^{\xi})^{-1}\boldsymbol{\Lambda}_n\mathbf{P}_{t|t-1},\end{aligned} \tag{E.10}$$

indeed $\mathbf{P}_{t|t-1}$ is positive definite and finite by Lemma D.7(i) and D.7(ii). Therefore, since $\mathbf{P}_{t|t-1}$ and $\mathbf{P}_{t|t}$ are deterministic by Lemma D.6(i), we have

$$\max_{t=1,\dots,T} n\|\mathbf{P}_{t|t}\| \le n\|\mathbf{P}_{t|t-1} - \{(\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n + \mathbf{P}_{t|t-1}^{-1})^{-1}\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n - \mathbf{I}_r + \mathbf{I}_r\}\mathbf{P}_{t|t-1}\|$$
$$\le n\|(\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n + \mathbf{P}_{t|t-1}^{-1})^{-1}\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n - \mathbf{I}_r\|\,\|\mathbf{P}_{t|t-1}\| = O(1), \tag{E.11}$$

because of Lemma C.6(i), which is independent of t, and Lemma D.7(i). This proves part (i).

For part (ii), from (A.7), we have

$$\|\mathbf{P}_{t|T} - \mathbf{P}_{t|t}\| = \|\mathbf{P}_{t|t}\mathbf{A}'\mathbf{P}_{t+1|t}^{-1}(\mathbf{P}_{t+1|T} - \mathbf{P}_{t+1|t})\mathbf{P}_{t+1|t}^{-1}\mathbf{A}\mathbf{P}_{t|t}\|$$
$$\le \|\mathbf{P}_{t|t}\|^2\,\|\mathbf{A}\|^2\,\|\mathbf{P}_{t+1|t}^{-1}\|^2\,\left\{\|\mathbf{P}_{t+1|T}\| + \|\mathbf{P}_{t+1|t}\|\right\}. \tag{E.12}$$

Now, given that $\mathbf{P}_{T|T}$ is obtained by the last iteration of the Kalman filter, for $t = T-1$ (E.12) becomes:

$$\|\mathbf{P}_{T-1|T} - \mathbf{P}_{T-1|T-1}\| \le \|\mathbf{P}_{T-1|T-1}\|^2\,\|\mathbf{A}\|^2\,\|\mathbf{P}_{T|T-1}^{-1}\|^2\,\left\{\|\mathbf{P}_{T|T}\| + \|\mathbf{P}_{T|T-1}\|\right\}$$
$$\le \|\mathbf{P}_{T-1|T-1}\|^2\frac{M_A^2}{\underline{M}_P^2}\left\{\|\mathbf{P}_{T|T}\| + M_P\right\} = O(n^{-2}), \tag{E.13}$$

by part (i), Assumption 1(d), and Lemma D.7(i) and D.7(ii). Therefore, from part (i) and (E.13)

$$\|\mathbf{P}_{T-1|T}\| \le \|\mathbf{P}_{T-1|T} - \mathbf{P}_{T-1|T-1}\| + \|\mathbf{P}_{T-1|T-1}\| = O(n^{-2}) + O(n^{-1}). \tag{E.14}$$

For $t = T-2$, (E.12) becomes:

$$\|\mathbf{P}_{T-2|T} - \mathbf{P}_{T-2|T-2}\| = \|\mathbf{P}_{T-2|T-2}\mathbf{A}'\mathbf{P}_{T-1|T-2}^{-1}(\mathbf{P}_{T-1|T} - \mathbf{P}_{T-1|T-2})\mathbf{P}_{T-1|T-2}^{-1}\mathbf{A}\mathbf{P}_{T-2|T-2}\|$$
$$\le \|\mathbf{P}_{T-2|T-2}\|^2\,\|\mathbf{A}\|^2\,\|\mathbf{P}_{T-1|T-2}^{-1}\|^2\,\left\{\|\mathbf{P}_{T-1|T}\| + \|\mathbf{P}_{T-1|T-2}\|\right\}$$
$$\le \|\mathbf{P}_{T-2|T-2}\|^2\frac{M_A^2}{\underline{M}_P^2}\left\{\|\mathbf{P}_{T-1|T}\| + M_P\right\} = O(n^{-2}), \tag{E.15}$$

by part (i), Assumption 1(d), Lemma D.7, and (E.14). Therefore, from part (i) and (E.15)

$$\|\mathbf{P}_{T-2|T}\| \le \|\mathbf{P}_{T-2|T} - \mathbf{P}_{T-2|T-2}\| + \|\mathbf{P}_{T-2|T-2}\| = O(n^{-2}) + O(n^{-1}). \tag{E.16}$$

By comparing (E.14) and (E.16) it is clear that, the same asymptotic bound holds for all $t = T, \dots, 1$

$$\|\mathbf{P}_{t|T} - \mathbf{P}_{t|t}\| \le \|\mathbf{P}_{t|t}\|^2\frac{M_A^2}{\underline{M}_P^2}\left\{\|\mathbf{P}_{t+1|T}\| + M_P\right\} = O(n^{-2}), \tag{E.17}$$

and since $\|\mathbf{P}_{t+1|T}\| = O(n^{-1})$ then it is asymptotically negligible, thus (E.17) holds uniformly in $t = T, \dots, 1$ because of part (i), and since $\mathbf{P}_{t|T}$ and $\mathbf{P}_{t|t}$ are deterministic because of Lemma D.6(i). It follows that, by part (i)

$$\max_{t=1,\dots,T} n\|\mathbf{P}_{t|T}\| \le \max_{t=1,\dots,T} n\|\mathbf{P}_{t|T} - \mathbf{P}_{t|t}\| + \max_{t=1,\dots,T} n\|\mathbf{P}_{t|t}\| = O(n^{-1}) + O(1) = O(1).$$

This proves part (ii).

Parts (iii) and (iv) are proved exactly as parts (i) and (ii), respectively, but using Lemmas C.6(ii), D.7(iii), and D.7(iv) instead of Lemmas C.6(i), D.7(i), and D.7(ii).

Part (v) is proved by (E.17). Part (vi) is proved as part (v) by repeating the same reasoning leading to the proof of part (iv). This completes the proof. □

Lemma E.5. *Under Assumptions 1, 2, and 5, as $n, T \to \infty$,*

(i) $\|\mathbf{F}_{t|t-1}\| = O_p(1)$*, uniformly in t;*

(ii) $\|\mathbf{F}_{0,t|t-1}\| = O_p(1)$*, uniformly in t;*

(iii) $\log^{-1/\delta_v} T \max_{t=1,\dots,T}\|\mathbf{F}_{0,t|t-1}\| = O_p(1)$.

Proof. Given that $\mathbf{F}_{0|0} = \mathbf{0}_r$, from (A.1) it follows that $\mathbf{F}_{1|0} = \mathbf{0}_r$, then, from (A.3)

$$\mathbf{F}_{1|1} = \mathbf{P}_{1|0}\mathbf{\Lambda}_n'(\mathbf{\Lambda}_n\mathbf{P}_{1|0}\mathbf{\Lambda}_n' + \mathbf{\Sigma}_n^\xi)^{-1}\mathbf{x}_{n1} = (\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n + \mathbf{P}_{1|0}^{-1})^{-1}\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{x}_{n1}$$

by Lemma D.13 which can be applied since $\mathbf{P}_{1|0}$ is positive definite by Lemma D.7(ii). Thus,

$$\|\mathbf{F}_{1|1}\| \le \|(\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n + \mathbf{P}_{1|0}^{-1})^{-1}\| \, \|\mathbf{\Lambda}_n\| \, \|(\mathbf{\Sigma}_n^\xi)^{-1}\| \, \|\mathbf{x}_{n1}\| = O_p(1), \tag{E.18}$$

by Lemmas C.3(i), C.2, and C.10, and Assumption 2(a). Then, from (A.1),

$$\|\mathbf{F}_{2|1}\| \le \|\mathbf{A}\| \, \|\mathbf{F}_{1|1}\| = O_p(1), \tag{E.19}$$

because of (E.18) and Assumption 1(d). Therefore, from (A.3), using (E.19) and the same arguments leading to (E.18),

$$\begin{aligned}\|\mathbf{F}_{2|2}\| &\le \|\mathbf{F}_{2|1}\| + \|(\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n + \mathbf{P}_{1|0}^{-1})^{-1}\| \, \|\mathbf{\Lambda}_n\| \, \|(\mathbf{\Sigma}_n^\xi)^{-1}\| \, \{\|\mathbf{x}_{n1}\| + \|\mathbf{\Lambda}_n\| \, \|\mathbf{F}_{2|1}\|\} \\ &= O_p(1).\end{aligned} \tag{E.20}$$

It is then clear that (E.19) and (E.20) hold for all $t = 1, \dots, T$, and the result follows from Lemma C.10. This proves part (i).

For part (ii), the proof is identical to part (i) but using Lemma C.3(ii) instead of Lemma C.3(i). For part (iii) repeat the same steps as in part (ii) but using Lemma E.3 instead of Lemma C.10, and noticing that $\|\mathbf{x}_{nt}\| \le \max_{t=1,\dots,T} \|\mathbf{x}_{nt}\|$, for all $t = 1, \dots, T$. This completes the proof. □

Lemma E.6. *Under Assumptions 1, 2, and 5, as $n, T \to \infty$,*
(i) $\sqrt{n}\|\mathbf{F}_{t|t} - \mathbf{F}_t\| = O_p(1)$*, uniformly in t;*
(ii) $\sqrt{n}\|\mathbf{F}_{0,t|t} - \mathbf{F}_t\| = O_p(1)$*, uniformly in t;*
(iii) $\log^{-1/\delta_v} T\sqrt{n}\,\max_{t=1,\dots,T}\|\mathbf{F}_{0,t|t} - \mathbf{F}_t\| = O_p(1)$*.*

Proof. Since $\mathbf{P}_{t|t-1}$ is positive definite because of Lemma D.7(ii), we can use the Woodbury formula in Lemma D.13 so that, for any $t = 1, \dots, T$, the Kalman filter estimator defined in (A.3) can be written as:

$$\begin{aligned}\mathbf{F}_{t|t} &= \mathbf{F}_{t|t-1} + \mathbf{P}_{t|t-1}\mathbf{\Lambda}_n'(\mathbf{\Lambda}_n\mathbf{P}_{t|t-1}\mathbf{\Lambda}_n' + \mathbf{\Sigma}_n^\xi)^{-1}(\mathbf{x}_{nt} - \mathbf{\Lambda}_n\mathbf{F}_{t|t-1}) \\ &= \mathbf{F}_{t|t-1} + (\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n + \mathbf{P}_{t|t-1}^{-1})^{-1}\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}(\mathbf{x}_{nt} - \mathbf{\Lambda}_n\mathbf{F}_{t|t-1}) \\ &= \mathbf{F}_{t|t-1} + (\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n + \mathbf{P}_{t|t-1}^{-1})^{-1}\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}(\mathbf{\Lambda}_n\mathbf{F}_t + \boldsymbol{\xi}_{nt} - \mathbf{\Lambda}_n\mathbf{F}_{t|t-1}) \\ &= \mathbf{F}_{t|t-1} + (\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n + \mathbf{P}_{t|t-1}^{-1})^{-1}\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n\mathbf{F}_t \\ &\quad - (\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n + \mathbf{P}_{t|t-1}^{-1})^{-1}\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n\mathbf{F}_{t|t-1} \\ &\quad + (\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n + \mathbf{P}_{t|t-1}^{-1})^{-1}\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\boldsymbol{\xi}_{nt},\end{aligned} \tag{E.21}$$

where in the last step we used the definition of $\mathbf{x}_{nt}$ in (3). Then,

$$\begin{aligned}&\|\mathbf{F}_{t|t-1} - (\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n + \mathbf{P}_{t|t-1}^{-1})^{-1}\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n\mathbf{F}_{t|t-1}\| \\ &\quad \le \|\mathbf{F}_{t|t-1}\| \, \|\mathbf{I}_r - (\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n + \mathbf{P}_{t|t-1}^{-1})^{-1}\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n\| = O_p(n^{-1}),\end{aligned} \tag{E.22}$$

by Lemma E.5(i) and Lemma C.6(i), which can be applied by Lemma D.7(i). Similarly,

$$\begin{aligned}&\|(\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n + \mathbf{P}_{t|t-1}^{-1})^{-1}\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n\mathbf{F}_t - \mathbf{F}_t\| \\ &\quad \le \|(\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n + \mathbf{P}_{t|t-1}^{-1})^{-1}\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n - \mathbf{I}_r\| \, \|\mathbf{F}_t\| = O_p(n^{-1}),\end{aligned} \tag{E.23}$$

by the same arguments leading to (E.22) and since $\|\mathbf{F}_t\| = O_p(1)$ uniformly in $t = 1, \dots, T$, because $\mathbb{E}[\|\mathbf{F}_t\|^2] = \sum_{j=1}^r \mathbb{E}[F_{jt}^2] = \mathrm{tr}(\mathbf{\Gamma}^F) = r$ by Assumption 6(b). From (E.22) and (E.23) it follows that

$$\begin{aligned}\|\mathbf{F}_{t|t} - \mathbf{F}_t\| &\le \|(\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n)^{-1}\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\boldsymbol{\xi}_{nt}\| + O_p(n^{-1}) \\ &\le \|n(\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n)^{-1}\| \, \|n^{-1}\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\boldsymbol{\xi}_{nt}\| + O_p(n^{-1}) \\ &= O_p(n^{-1/2}),\end{aligned} \tag{E.24}$$

by Lemma C.3(iii) and Lemma C.7(i). This proves part (i).

Part (ii) is proved in the same way, but using Lemmas D.7(iii), D.7(iv), E.5(ii), C.6(ii), C.3(iv), and C.7(ii), instead of Lemmas D.7(i), D.7(ii), E.5(i), C.6(i), C.3(iii), and C.7(i). For part (iii) repeat the same steps as in part (ii) but using Lemma E.5(iii) instead of Lemma E.5(ii), and since $\max_{t=1,\ldots,T}\|\mathbf{F}_t\| = O_p(\log^{1/\delta_v} T)$ by Lemma E.2. This completes the proof. □

Lemma E.7. *Under Assumptions 1 and 2, as $n\to\infty$,*

(i) $n\|\mathbf{F}_{t|T}-\mathbf{F}_{t|t}\| = O_p(1)$, uniformly in t;

(ii) $n\|\mathbf{F}_{0,t|T}-\mathbf{F}_{0,t|t}\| = O_p(1)$, uniformly in t.

Proof. For any $t=1,\ldots,T$, using (A.6)

$$\|\mathbf{F}_{t|T}-\mathbf{F}_{t|t}\| \le \|\mathbf{P}_{t|t}\|\,\|\mathbf{A}'\|\,\|\mathbf{P}_{t+1|t}^{-1}\|\,\|(\mathbf{F}_{t+1|T}-\mathbf{A}\mathbf{F}_{t|t})\|. \tag{E.25}$$

Now, given that $\mathbf{F}_{T|T}$ is obtained by the last iteration of the Kalman filter, for $t=T-1$ (E.25) becomes:

$$\begin{aligned}\|\mathbf{F}_{T-1|T}-\mathbf{F}_{T-1|T-1}\| &\le \|\mathbf{P}_{T-1|T-1}\|\,\|\mathbf{A}\|\,\|\mathbf{P}_{T|T-1}^{-1}\|\,\left\{\|\mathbf{F}_{T|T}\|+\|\mathbf{A}\|\,\|\mathbf{F}_{T-1|T-1}\|\right\}\\ &\le \|\mathbf{P}_{T-1|T-1}\|\,\frac{M_A}{\underline{M}_P}\left\{\|\mathbf{F}_{T|T}-\mathbf{F}_T\|+\|\mathbf{F}_T\|+M_A\left[\|\mathbf{F}_{T-1|T-1}-\mathbf{F}_{T-1}\|+\|\mathbf{F}_{T-1}\|\right]\right\}\\ &= O_p(n^{-1}),\end{aligned} \tag{E.26}$$

by Assumption 1(d), and Lemmas D.7(ii), E.4(i), and E.6(i), and since $\|\mathbf{F}_t\| = O_p(1)$ uniformly in $t=1,\ldots,T$, because $\mathbb{E}[\|\mathbf{F}_t\|^2]=\sum_{j=1}^r\mathbb{E}[F_{jt}^2]=\mathrm{tr}(\mathbf{\Gamma}^F)=r$ by Assumption 6(b). For $t=T-2$ (E.25) becomes:

$$\begin{aligned}\|\mathbf{F}_{T-2|T}-\mathbf{F}_{T-2|T-2}\| &\le \|\mathbf{P}_{T-2|T-2}\|\,\|\mathbf{A}'\|\,\|\mathbf{P}_{T-1|T-2}^{-1}\|\,\|(\mathbf{F}_{T-1|T}-\mathbf{A}\mathbf{F}_{T-2|T-2})\|\\ &\le \|\mathbf{P}_{T-1|T-1}\|\,\frac{M_A}{\underline{M}_P}\left\{\|\mathbf{F}_{T-1|T}-\mathbf{F}_{T-1|T-1}\|+\|\mathbf{F}_{T-1|T-1}-\mathbf{F}_{T-1}\|+\|\mathbf{F}_{T-1}\|\right.\\ &\quad\left.+M_A\left[\|\mathbf{F}_{T-2|T-2}-\mathbf{F}_{T-2}\|+\|\mathbf{F}_{T-2}\|\right]\right\}\\ &= O_p(n^{-1}),\end{aligned} \tag{E.27}$$

by (E.26), Assumption 1(d), and Lemmas D.7(ii), E.4(i), and E.6(i), and since $\|\mathbf{F}_t\| = O_p(1)$ uniformly in $t=1,\ldots,T$. By comparing (E.26) and (E.27) it is clear that, the same asymptotic bound holds uniformly in $t=T,\ldots,1$, i.e., from (E.25) we get

$$\begin{aligned}\|\mathbf{F}_{t|T}-\mathbf{F}_{t|t}\| \le \|\mathbf{P}_{t|t}\|\,\frac{M_A}{\underline{M}_P}&\left\{\|\mathbf{F}_{t+1|T}-\mathbf{F}_{t+1|t+1}\|+\|\mathbf{F}_{t+1|t+1}-\mathbf{F}_{t+1}\|+\|\mathbf{F}_{t+1}\|\right.\\ &\left.+M_A\left[\|\mathbf{F}_{t|t}-\mathbf{F}_t\|+\|\mathbf{F}_t\|\right]\right\} = O_p(n^{-1}).\end{aligned} \tag{E.28}$$

This proves part (i).

Part (ii) is proved in the same way but using Lemmas D.7(iv), E.4(iii), and E.6(ii), instead of Lemmas D.7(ii), E.4(i), and E.6(i). This completes the proof. □

Lemma E.8. *Under Assumptions 1, 2, 3, and 5, as $n\to\infty$,*

(i) $n\|\mathbf{F}_{t|t}-\mathbf{F}_t^{\mathrm{WLS}}\| = O_p(1)$, uniformly in t;

(ii) $n\|\mathbf{F}_{0,t|t}-\mathbf{F}_t^{\mathrm{GLS}}\| = O_p(1)$, uniformly in t;

(iii) $n^2\max_{t=1,\ldots,T}\|\mathbf{P}_{t|t}-(\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n)^{-1}\| = O_p(1)$;

(iv) $n^2\max_{t=1,\ldots,T}\|\mathbf{P}_{0,t|t}-(\mathbf{\Lambda}_n'(\mathbf{\Gamma}_n^\xi)^{-1}\mathbf{\Lambda}_n)^{-1}\| = O_p(1)$;

(v) $n^2\max_{t=1,\ldots,T}\|\mathbf{P}_{t|T}-(\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n)^{-1}\| = O_p(1)$;

(vi) $n^2\max_{t=1,\ldots,T}\|\mathbf{P}_{0,t|T}-(\mathbf{\Lambda}_n'(\mathbf{\Gamma}_n^\xi)^{-1}\mathbf{\Lambda}_n)^{-1}\| = O_p(1)$;

where $\mathbf{F}_t^{\mathrm{WLS}}=(\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n)^{-1}\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{x}_t$ and $\mathbf{F}_t^{\mathrm{GLS}}=(\mathbf{\Lambda}_n'(\mathbf{\Gamma}_n^\xi)^{-1}\mathbf{\Lambda}_n)^{-1}\mathbf{\Lambda}_n'(\mathbf{\Gamma}_n^\xi)^{-1}\mathbf{x}_t$.

Proof. For part (i), from (A.3) we have

$$\begin{aligned}\mathbf{F}_{t|t}-\mathbf{F}_t^{\mathrm{WLS}} &= (\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n+\mathbf{P}_{t|t-1})^{-1}\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{x}_t-\mathbf{F}_t^{\mathrm{WLS}}\\ &\quad+\mathbf{F}_{t|t-1}-(\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n+\mathbf{P}_{t|t-1})^{-1}\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n\mathbf{F}_{t|t-1}.\end{aligned} \tag{E.29}$$

Then,

$$\begin{aligned}&\|(\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n+\mathbf{P}_{t|t-1})^{-1}\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{x}_t-\mathbf{F}_t^{\mathrm{WLS}}\|\\&\le\|(\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n+\mathbf{P}_{t|t-1})^{-1}-(\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n)^{-1}\|\,\|\mathbf{\Lambda}_n\|\,\|(\mathbf{\Sigma}_n^\xi)^{-1}\|\,\|\mathbf{x}_t\|=O_p(n^{-1}),\end{aligned}\tag{E.30}$$

by Lemmas C.6(iii), C.2, and C.10, and Assumption 2(a). And,

$$\begin{aligned}&\|\mathbf{F}_{t|t-1}-(\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n+\mathbf{P}_{t|t-1})^{-1}\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n\mathbf{F}_{t|t-1}\|\\&\le\|\mathbf{I}_r-(\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n+\mathbf{P}_{t|t-1})^{-1}\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n\|\,\|\mathbf{F}_{t|t-1}\|\\&=O_p(n^{-1}),\end{aligned}\tag{E.31}$$

by (E.22) in the proof of Lemma E.6(i). By substituting (E.30) and (E.31) into (E.29), we prove part (i).

Part (ii) is proved as part (i), but using Lemmas C.6(iv) and E.6(ii) instead of Lemmas C.6(iii) and E.6(i).

For part (iii), from (A.4), using the same steps leading to (D.23) in the proof of Lemma D.11, but when using the true parameters, it holds that:

$$\begin{aligned}\mathbf{P}_{t|t}=&(\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n)^{-1}\\&-\mathbf{P}_{t|t-1}((\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n)^{-1}+\mathbf{P}_{t|t-1})^{-1}(\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n)^{-1}(\mathbf{P}_{t|t-1})^{-1}(\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n)^{-1}.\end{aligned}\tag{E.32}$$

Notice that all inverses in (E.32) are well defined because of Lemmas C.3(iii), D.7(i), and D.7(ii) and Assumption 2(a). Therefore, from (E.32)

$$\begin{aligned}n^2\max_{t=1,\dots,T}\|\mathbf{P}_{t|t}-(\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n)^{-1}\|\le&\max_{t=1,\dots,T}\|\mathbf{P}_{t|t-1}\|\max_{t=1,\dots,T}\|(\mathbf{P}_{t|t-1})^{-1}\|\,n^2\|(\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n)^{-1}\|^2\\&\cdot\|((\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n)^{-1}+\mathbf{P}_{t|t-1})^{-1}\|\\=&O(1),\end{aligned}\tag{E.33}$$

because of Lemmas C.3(iii), D.7(i), and D.7(ii), and since $\|((\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n)^{-1}+\mathbf{P}_{t|t-1})^{-1}\|=O(1)$, because of the same arguments leading to (D.24) in the proof of Lemma D.11, which this time hold by Lemmas C.3(iii), D.7(i), and D.7(ii), instead of Lemmas D.5(iii), D.8(i), and D.8(ii).

Part (iv) is proved as part (iii), but using Lemmas C.3(iv), and Assumption 2(f) instead of Lemmas C.3(iii) and Assumption 2(a).

For part (v),

$$n^2\max_{t=1,\dots,T}\|\mathbf{P}_{t|T}-(\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n)^{-1}\|\le n^2\max_{t=1,\dots,T}\|\mathbf{P}_{t|T}-P_{t|t}\|+n^2\max_{t=1,\dots,T}\|\mathbf{P}_{t|t}-(\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n)^{-1}\|=O(1),$$

because of part (iii) and Lemma E.4(v). Part (vi) is proved as part (v), but using part (iv) and Lemma E.4(vi). This completes the proof. □

Lemma E.9. *Let $\ell_0(\boldsymbol{X}_{nT};\underline{\boldsymbol{\phi}}_n)$ be the log-likelihood obtained when $\mathbf{A}=\mathbf{0}_{r\times r}$ and $\mathbf{\Gamma}^v=\mathbf{I}_r$. Then, under Assumptions 1, 2, 3, 4, 5, and 6, as $n,T\to\infty$,*

$$n\log^{-2/\delta_v}T\sup_{\underline{\boldsymbol{\varphi}}_n\in\mathcal{O}_n}(nT)^{-1}\left|\ell(\boldsymbol{X}_{nT};\underline{\boldsymbol{\phi}}_n,\boldsymbol{\theta})-\ell_0(\boldsymbol{X}_{nT};\underline{\boldsymbol{\phi}}_n)\right|=O_p(1).$$

Proof. Throughout we consider generic values of the parameters such that $\underline{\boldsymbol{\varphi}}_n\in\mathcal{O}_n$ where $\mathcal{O}_n=\{\mathcal{O}^n_{\lambda_i}\cap\mathcal{E}_{\Lambda_n}\}\times\{\mathcal{O}^n_{\sigma_i^2}\cap\mathcal{E}_{\Gamma_n^\xi}\}\times\mathcal{O}_{\mathcal{A}}\times\mathcal{O}_{\Gamma^v}$ as defined in Section 4.3.4. Thus the elements of $\underline{\boldsymbol{\varphi}}_n$ satisfy Assumptions 1(a), 1(d), 1(e), 2(a), 2(b), and 2(f).

First, recall that the log-likelihood (5) can be written as

$$\begin{aligned}
\ell(\boldsymbol{X}_{nT};\underline{\boldsymbol{\phi}}_n,\underline{\boldsymbol{\theta}}) &= \ell(\boldsymbol{X}_{nT}|\boldsymbol{F}_T;\underline{\boldsymbol{\phi}}_n) + \ell(\boldsymbol{F}_T;\underline{\boldsymbol{\theta}}) - \ell(\boldsymbol{F}_T|\boldsymbol{X}_{nT};\underline{\boldsymbol{\phi}}_n,\underline{\boldsymbol{\theta}})\\
&= -\frac{T}{2}\log\det(\underline{\boldsymbol{\Sigma}}_n^{\xi}) - \frac{1}{2}\sum_{t=1}^T(\mathbf{x}_{nt}-\underline{\boldsymbol{\Lambda}}_n\mathbf{F}_t)'(\underline{\boldsymbol{\Sigma}}_n^{\xi})^{-1}(\mathbf{x}_{nt}-\underline{\boldsymbol{\Lambda}}_n\mathbf{F}_t)\\
&\quad -\frac{T}{2}\log\det(\underline{\boldsymbol{\Gamma}}^{v}) - \frac{1}{2}\sum_{t=1}^T(\mathbf{F}_t-\underline{\mathbf{A}}\mathbf{F}_{t-1})'(\underline{\boldsymbol{\Gamma}}^{v})^{-1}(\mathbf{F}_t-\underline{\mathbf{A}}\mathbf{F}_{t-1})\\
&\quad +\frac{1}{2}\sum_{t=1}^T\log\det(\boldsymbol{\mathcal{P}}_{0,t|T}(\underline{\boldsymbol{\phi}}_n,\underline{\boldsymbol{\theta}}))\\
&\quad +\frac{1}{2}\sum_{t=1}^T(\mathbf{F}_t-\boldsymbol{\mathcal{F}}_{0,t|T}(\underline{\boldsymbol{\phi}}_n,\underline{\boldsymbol{\theta}}))'(\boldsymbol{\mathcal{P}}_{0,t|T}(\underline{\boldsymbol{\phi}}_n,\underline{\boldsymbol{\theta}}))^{-1}(\mathbf{F}_t-\boldsymbol{\mathcal{F}}_{0,t|T}(\underline{\boldsymbol{\phi}}_n,\underline{\boldsymbol{\theta}})),
\end{aligned} \tag{E.34}$$

where we used the fact that $\mathbf{F}_0 = \mathbf{0}_r$ by Assumption 1(i) and we used the definitions:

$$\begin{aligned}
&\boldsymbol{\mathcal{F}}_{0,t|T}(\underline{\boldsymbol{\phi}}_n,\underline{\boldsymbol{\theta}}) = \mathbb{E}_{\underline{\varphi}_n}[\mathbf{F}_t|\boldsymbol{X}_{nT}] \equiv \underline{\boldsymbol{\mathcal{F}}}_{0,t|T},\\
&\boldsymbol{\mathcal{P}}_{0,t|T}(\underline{\boldsymbol{\phi}}_n,\underline{\boldsymbol{\theta}}) = \mathbb{E}_{\underline{\varphi}_n}[(\mathbf{F}_t-\boldsymbol{\mathcal{F}}_{0,t|T}(\underline{\boldsymbol{\phi}}_n,\underline{\boldsymbol{\theta}}))(\mathbf{F}_t-\boldsymbol{\mathcal{F}}_{0,t|T}(\underline{\boldsymbol{\phi}}_n,\underline{\boldsymbol{\theta}}))'|\boldsymbol{X}_{nT}] \equiv \underline{\boldsymbol{\mathcal{P}}}_{0,t|T}.
\end{aligned} \tag{E.35}$$

Now, since (E.34) holds for any $\mathbf{F}_t$ we can always choose $\mathbf{F}_t = \underline{\boldsymbol{\mathcal{F}}}_{0,t|T}$ for all $t = 1,\dots,T$, so that

$$\begin{aligned}
\ell(\boldsymbol{X}_{nT};\underline{\boldsymbol{\phi}}_n,\underline{\boldsymbol{\theta}}) = &-\frac{T}{2}\log\det(\underline{\boldsymbol{\Sigma}}_n^{\xi}) - \frac{1}{2}\sum_{t=1}^T(\mathbf{x}_{nt}-\underline{\boldsymbol{\Lambda}}_n\underline{\boldsymbol{\mathcal{F}}}_{0,t|T})'(\underline{\boldsymbol{\Sigma}}_n^{\xi})^{-1}(\mathbf{x}_{nt}-\underline{\boldsymbol{\Lambda}}_n\underline{\boldsymbol{\mathcal{F}}}_{0,t|T})\\
&-\frac{T}{2}\log\det(\underline{\boldsymbol{\Gamma}}^{v}) - \frac{1}{2}\sum_{t=1}^T(\underline{\boldsymbol{\mathcal{F}}}_{0,t|T}-\underline{\mathbf{A}}\,\underline{\boldsymbol{\mathcal{F}}}_{0,t-1|T})'(\underline{\boldsymbol{\Gamma}}^{v})^{-1}(\underline{\boldsymbol{\mathcal{F}}}_{0,t|T}-\underline{\mathbf{A}}\,\underline{\boldsymbol{\mathcal{F}}}_{0,t-1|T})\\
&+\frac{1}{2}\sum_{t=1}^T\log\det(\underline{\boldsymbol{\mathcal{P}}}_{0,t|T}).
\end{aligned} \tag{E.36}$$

Second, consider the log-likelihood (5) when the autocorrelation of the factors is not accounted for, i.e., when $\mathbf{A} = \mathbf{0}_{r\times r}$ and $\boldsymbol{\Gamma}^v = \mathbf{I}_r$,

$$\ell_0(\boldsymbol{X}_{nT};\underline{\boldsymbol{\phi}}_n) = -\frac{T}{2}\log\det\left(\underline{\boldsymbol{\Lambda}}_n\underline{\boldsymbol{\Lambda}}_n' + \underline{\boldsymbol{\Sigma}}_n^{\xi}\right) - \frac{1}{2}\sum_{t=1}^T\left[\mathbf{x}_{nt}'\left(\underline{\boldsymbol{\Lambda}}_n\underline{\boldsymbol{\Lambda}}_n' + \underline{\boldsymbol{\Sigma}}_n^{\xi}\right)^{-1}\mathbf{x}_{nt}\right], \tag{E.37}$$

where we are imposing Assumption 6(b), so that we can set $\underline{\boldsymbol{\Gamma}}^F = \mathbf{I}_r$ in the log-likelihood. Clearly, (E.37) can also be written as

$$\begin{aligned}
\ell_0(\boldsymbol{X}_{nT};\underline{\boldsymbol{\phi}}_n) &= \ell_0(\boldsymbol{X}_{nT}|\boldsymbol{F}_T;\underline{\boldsymbol{\phi}}_n) + \ell_0(\boldsymbol{F}_T) - \ell_0(\boldsymbol{F}_T|\boldsymbol{X}_{nT};\underline{\boldsymbol{\phi}}_n)\\
&= -\frac{T}{2}\log\det(\underline{\boldsymbol{\Sigma}}_n^{\xi}) - \frac{1}{2}\sum_{t=1}^T(\mathbf{x}_{nt}-\underline{\boldsymbol{\Lambda}}_n\mathbf{F}_t)'(\underline{\boldsymbol{\Sigma}}_n^{\xi})^{-1}(\mathbf{x}_{nt}-\underline{\boldsymbol{\Lambda}}_n\mathbf{F}_t)\\
&\quad -\frac{1}{2}\sum_{t=1}^T\mathbf{F}_t'\mathbf{F}_t + \frac{1}{2}\sum_{t=1}^T\log\det(\boldsymbol{\mathcal{V}}_{0,t|T}(\underline{\boldsymbol{\phi}}_n))\\
&\quad +\frac{1}{2}\sum_{t=1}^T(\mathbf{F}_t-\boldsymbol{\mathcal{G}}_{0,t|T}(\underline{\boldsymbol{\phi}}_n))'(\boldsymbol{\mathcal{V}}_{0,t|T}(\underline{\boldsymbol{\phi}}_n))^{-1}(\mathbf{F}_t-\boldsymbol{\mathcal{G}}_{0,t|T}(\underline{\boldsymbol{\phi}}_n)),
\end{aligned} \tag{E.38}$$

where

$$\begin{aligned}
&\boldsymbol{\mathcal{G}}_{0,t|T}(\underline{\boldsymbol{\phi}}_n) = \mathbb{E}_{\underline{\phi}_n}[\mathbf{F}_t|\boldsymbol{X}_{nT}] \equiv \underline{\boldsymbol{\mathcal{G}}}_{0,t|T},\\
&\boldsymbol{\mathcal{V}}_{0,t|T}(\underline{\boldsymbol{\phi}}_n) = \mathbb{E}_{\underline{\phi}_n}[(\mathbf{F}_t-\boldsymbol{\mathcal{G}}_{0,t|T}(\underline{\boldsymbol{\phi}}_n))(\mathbf{F}_t-\boldsymbol{\mathcal{G}}_{0,t|T}(\underline{\boldsymbol{\phi}}_n))'|\boldsymbol{X}_{nT}] \equiv \underline{\boldsymbol{\mathcal{V}}}_{0,t|T}.
\end{aligned} \tag{E.39}$$

Now, since (E.38) holds for any $\mathbf{F}_t$ we can always choose $\mathbf{F}_t = \underline{\boldsymbol{\mathcal{G}}}_{0,t|T}$ for all $t = 1, \ldots, T$, so that

$$\ell_0(\boldsymbol{X}_{nT}; \underline{\boldsymbol{\phi}}_n) = -\frac{T}{2}\log\det(\underline{\boldsymbol{\Sigma}}_n^\xi) - \frac{1}{2}\sum_{t=1}^T(\mathbf{x}_{nt} - \underline{\boldsymbol{\Lambda}}_n\underline{\boldsymbol{\mathcal{G}}}_{0,t|T})'(\underline{\boldsymbol{\Sigma}}_n^\xi)^{-1}(\mathbf{x}_{nt} - \underline{\boldsymbol{\Lambda}}_n\underline{\boldsymbol{\mathcal{G}}}_{0,t|T})$$
$$- \frac{1}{2}\sum_{t=1}^T \underline{\boldsymbol{\mathcal{G}}}'_{0,t|T}\underline{\boldsymbol{\mathcal{G}}}_{0,t|T} + \frac{1}{2}\sum_{t=1}^T \log\det(\underline{\boldsymbol{\mathcal{V}}}_{0,t|T}). \tag{E.40}$$

Under Assumption 4 the conditional mean, $\underline{\boldsymbol{\mathcal{F}}}_{0,t|T}$ in (E.35) is a linear function of $\boldsymbol{X}_{nT}$, so it can be obtained by linear projection, and, therefore it is given by the correctly specified Kalman smoother, i.e., using as parameters $\underline{\boldsymbol{\phi}}_n$ and $\underline{\boldsymbol{\theta}}$ and when replacing $\underline{\boldsymbol{\Sigma}}_n^\xi$ with $\underline{\boldsymbol{\Gamma}}_n^\xi$, thus,

$$\underline{\boldsymbol{\mathcal{F}}}_{0,t|T} = \underline{\mathbf{F}}_{0,t|T}. \tag{E.41}$$

Likewise, under Assumption 4, $\underline{\boldsymbol{\mathcal{G}}}_{0,t|T}$ in (E.39) is also linear, however, since this is the conditional mean for the case in which no dynamics for the factors is specified, it is given by the simpler linear projection (using Lemma D.13)

$$\underline{\boldsymbol{\mathcal{G}}}_{0,t|T} = \underline{\boldsymbol{\Lambda}}'_n(\underline{\boldsymbol{\Lambda}}_n\underline{\boldsymbol{\Lambda}}'_n + \underline{\boldsymbol{\Gamma}}_n^\xi)^{-1}\mathbf{x}_{nt} = (\underline{\boldsymbol{\Lambda}}'_n(\underline{\boldsymbol{\Gamma}}_n^\xi)^{-1}\underline{\boldsymbol{\Lambda}}_n + \mathbf{I}_r)^{-1}\underline{\boldsymbol{\Lambda}}'_n(\underline{\boldsymbol{\Gamma}}_n^\xi)^{-1}\mathbf{x}_{nt} = \underline{\mathbf{F}}_{0,t}^{\mathrm{REG}}. \tag{E.42}$$

Now, consider the generalized least squares estimator of the factors computed for generic values of the parameters $\underline{\boldsymbol{\phi}}_n$:

$$\underline{\mathbf{F}}_t^{\mathrm{GLS}} = (\underline{\boldsymbol{\Lambda}}'_n(\underline{\boldsymbol{\Gamma}}_n^\xi)^{-1}\underline{\boldsymbol{\Lambda}}_n)^{-1}\underline{\boldsymbol{\Lambda}}'_n(\underline{\boldsymbol{\Gamma}}_n^\xi)^{-1}\mathbf{x}_{nt}, \tag{E.43}$$

Then, since we restrict to $\underline{\boldsymbol{\varphi}}_n \in \mathcal{O}_n$, it follows that

$$\|\underline{\mathbf{F}}_{0,t}^{\mathrm{REG}} - \underline{\mathbf{F}}_t^{\mathrm{GLS}}\| \le n\|(\underline{\boldsymbol{\Lambda}}'_n(\underline{\boldsymbol{\Gamma}}_n^\xi)^{-1}\underline{\boldsymbol{\Lambda}}_n + \mathbf{I}_r)^{-1} - (\underline{\boldsymbol{\Lambda}}'_n(\underline{\boldsymbol{\Gamma}}_n^\xi)^{-1}\underline{\boldsymbol{\Lambda}}_n)^{-1}\|\, n^{-1/2}\|\underline{\boldsymbol{\Lambda}}'_n(\underline{\boldsymbol{\Gamma}}_n^\xi)^{-1}\|\, n^{-1/2}\|\mathbf{x}_{nt}\|$$
$$= O(n^{-1})O_p(1), \tag{E.44}$$

by Lemmas C.3(viii), C.6(iv), and C.10.

Moreover, by Lemmas E.7(ii) and E.8(ii)

$$\|\underline{\mathbf{F}}_{0,t|T} - \underline{\mathbf{F}}_t^{\mathrm{GLS}}\| \le \|\underline{\mathbf{F}}_{0,t|T} - \underline{\mathbf{F}}_{0,t|t}\| + \|\underline{\mathbf{F}}_{0,t|t} - \underline{\mathbf{F}}_t^{\mathrm{GLS}}\| = O(n^{-1})O_p(1). \tag{E.45}$$

In particular, the bound in (E.45) is a product of a stochastic and a non-stochastic term because of (E.28) in the proof of Lemma E.7, and (E.30) and (E.31) in the proof of Lemma E.8.

Therefore, from (E.44) and (E.45) (see also Bai and Li, 2016)

$$\|\underline{\mathbf{F}}_{0,t|T} - \underline{\mathbf{F}}_{0,t}^{\mathrm{REG}}\| = O(n^{-1})O_p(1). \tag{E.46}$$

The results in (E.44)-(E.46) depend on t only through the $O_p(1)$ term which in turn is just function of $\mathbf{x}_{nt}$ and do not depend on the choice of the parameters as long as they belong to $\mathcal{O}_n$, as assumed. Since the parameters are deterministic and using Lemma E.3, we then have

$$\sup_{\underline{\boldsymbol{\varphi}}_n \in \mathcal{O}_n}\max_{t=1,\ldots,T}\|\underline{\mathbf{F}}_{0,t|T} - \underline{\mathbf{F}}_{0,t}^{\mathrm{REG}}\| = O(n^{-1})O_p(\log^{1/\delta_v} T). \tag{E.47}$$

Then, consider the conditional covariance in (E.39) and letting $\underline{\mathbf{K}}_n = (\underline{\boldsymbol{\Lambda}}'_n(\underline{\boldsymbol{\Gamma}}_n^\xi)^{-1}\underline{\boldsymbol{\Lambda}}_n + \mathbf{I}_r)^{-1}\underline{\boldsymbol{\Lambda}}'_n(\underline{\boldsymbol{\Gamma}}_n^\xi)^{-1}$, we have

$$\underline{\boldsymbol{\mathcal{V}}}_{0,t|T} = (\mathbf{I}_r - \underline{\mathbf{K}}_n\boldsymbol{\Lambda}_n)\mathbb{E}_{\underline{\phi}_n}[\mathbf{F}_t\mathbf{F}'_t|\boldsymbol{X}_{nT}](\mathbf{I}_r - \underline{\mathbf{K}}_n\boldsymbol{\Lambda}_n)' + \underline{\mathbf{K}}_n\mathbb{E}_{\underline{\phi}_n}[\boldsymbol{\xi}_{nt}\boldsymbol{\xi}'_{nt}|\boldsymbol{X}_{nT}]\underline{\mathbf{K}}'_n,$$

where we used also Lemma C.11. Moreover,

$$\mathbb{E}[\underline{\boldsymbol{\mathcal{V}}}_{0,t|T}] = (\mathbf{I}_r - \underline{\mathbf{K}}_n\boldsymbol{\Lambda}_n)\mathbb{E}[\mathbb{E}_{\underline{\phi}_n}[\mathbf{F}_t\mathbf{F}'_t|\boldsymbol{X}_{nT}]](\mathbf{I}_r - \underline{\mathbf{K}}_n\boldsymbol{\Lambda}_n)' + \underline{\mathbf{K}}_n\mathbb{E}[\mathbb{E}_{\underline{\phi}_n}[\boldsymbol{\xi}_{nt}\boldsymbol{\xi}'_{nt}|\boldsymbol{X}_{nT}]]\underline{\mathbf{K}}'_n,$$

is finite and positive definite, since $\|\underline{\mathbf{K}}_n\boldsymbol{\Lambda}_n\| = O(1)$ by Lemmas C.2, C.3(vi), and C.3(viii), and because $\mathbb{E}[\sup_{\underline{\boldsymbol{\varphi}}_n \in \mathcal{O}_n}\max_{t=1,\ldots,T}\mathbb{E}_{\underline{\phi}_n}[\mathbf{F}_t\mathbf{F}'_t|\boldsymbol{X}_{nT}]]$ and $\mathbb{E}[\sup_{\underline{\boldsymbol{\varphi}}_n \in \mathcal{O}_n}\max_{t=1,\ldots,T}\mathbb{E}_{\underline{\phi}_n}[\boldsymbol{\xi}_{nt}\boldsymbol{\xi}'_{nt}|\boldsymbol{X}_{nT}]]$ are both finite and positive definite by Assumptions 1(b) and 2(f), and Lemma C.1(v). Therefore, by Markov's inequality

$$\sup_{\underline{\boldsymbol{\varphi}}_n\in\mathcal{O}_n\setminus\{\boldsymbol{\varphi}_n\}}\max_{t=1,\ldots,T}\|\underline{\boldsymbol{\mathcal{V}}}_{0,t|T}\| = O_p(1), \qquad \sup_{\underline{\boldsymbol{\varphi}}_n\in\mathcal{O}_n\setminus\{\boldsymbol{\varphi}_n\}}\max_{t=1,\ldots,T}\|(\underline{\boldsymbol{\mathcal{V}}}_{0,t|T})^{-1}\| = O_p(1). \tag{E.48}$$

This bound is tighter when $\underline{\boldsymbol{\varphi}}_n = \boldsymbol{\varphi}_n$. Indeed, in that case from (A.4) and (A.7), setting $\mathbf{A} = \mathbf{0}_{r\times r}$ therein, we have

$$\boldsymbol{\mathcal{V}}_{0,t|T}(\boldsymbol{\phi}_n) = \mathbf{I}_r - \mathbf{I}_r\boldsymbol{\Lambda}_n'(\boldsymbol{\Lambda}_n\boldsymbol{\Lambda}_n' + \boldsymbol{\Gamma}_n^{\xi})^{-1}\boldsymbol{\Lambda}_n = \mathbf{I}_r - (\boldsymbol{\Lambda}_n'(\boldsymbol{\Gamma}_n^{\xi})^{-1}\boldsymbol{\Lambda}_n + \mathbf{I}_r)^{-1}\boldsymbol{\Lambda}_n'(\boldsymbol{\Gamma}_n^{\xi})^{-1}\boldsymbol{\Lambda}_n.$$

where we also used Assumption 6(b) for which $\boldsymbol{\Gamma}^F = \mathbf{I}_r$. Therefore, $\boldsymbol{\mathcal{V}}_{0,t|T}(\boldsymbol{\phi}_n)$ is deterministic and independent of t, and such that

$$\max_{t=1,\ldots,T}\|\boldsymbol{\mathcal{V}}_{0,t|T}(\boldsymbol{\phi}_n)\| = O(n^{-1}), \qquad \max_{t=1,\ldots,T}\|(\boldsymbol{\mathcal{V}}_{0,t|T}(\boldsymbol{\phi}_n))^{-1}\| = O(n), \tag{E.49}$$

because by Lemma C.6(ii) which holds since $\boldsymbol{\Gamma}_n^{\xi}$ is positive definite by Assumption 2(f). Furthermore, following the same steps leading to (E.33) in the proof of Lemma E.8 we have

$$\max_{t=1,\ldots,T}\|\boldsymbol{\mathcal{V}}_{0,t|T}(\boldsymbol{\phi}_n) - (\boldsymbol{\Lambda}_n'(\boldsymbol{\Gamma}_n^{\xi})^{-1}\boldsymbol{\Lambda}_n)^{-1}\| = O(n^{-2}). \tag{E.50}$$

Turning to the conditional covariance in (E.35), because of (E.41) we can write

$$\begin{aligned}
\underline{\boldsymbol{\mathcal{P}}}_{0,t|T} =& \mathbb{E}_{\underline{\boldsymbol{\varphi}}_n}[(\mathbf{F}_t - \underline{\mathbf{F}}_{0,t|T} + \underline{\mathbf{F}}_{0,t}^{\mathrm{REG}} - \underline{\mathbf{F}}_{0,t}^{\mathrm{REG}})(\mathbf{F}_t - \underline{\mathbf{F}}_{0,t|T} + \underline{\mathbf{F}}_{0,t}^{\mathrm{REG}} - \underline{\mathbf{F}}_{0,t}^{\mathrm{REG}})'|\boldsymbol{X}_{nT}]\\
=& \underline{\boldsymbol{\mathcal{V}}}_{0,t|T} + \mathbb{E}_{\underline{\boldsymbol{\varphi}}_n}[(\underline{\mathbf{F}}_{0,t}^{\mathrm{REG}} - \underline{\mathbf{F}}_{0,t|T})(\underline{\mathbf{F}}_{0,t}^{\mathrm{REG}} - \underline{\mathbf{F}}_{0,t|T})'|\boldsymbol{X}_{nT}]\\
&+ \mathbb{E}_{\underline{\boldsymbol{\varphi}}_n}[(\underline{\mathbf{F}}_{0,t}^{\mathrm{REG}} - \underline{\mathbf{F}}_{0,t|T})(\mathbf{F}_t - \underline{\mathbf{F}}_{0,t}^{\mathrm{REG}})'|\boldsymbol{X}_{nT}] + \mathbb{E}_{\underline{\boldsymbol{\varphi}}_n}[(\mathbf{F}_t - \underline{\mathbf{F}}_{0,t}^{\mathrm{REG}})(\underline{\mathbf{F}}_{0,t}^{\mathrm{REG}} - \underline{\mathbf{F}}_{0,t|T})'|\boldsymbol{X}_{nT}].
\end{aligned} \tag{E.51}$$

From (E.47) and (E.51) we have

$$\sup_{\underline{\boldsymbol{\varphi}}_n\in\mathcal{O}_n\setminus\{\boldsymbol{\varphi}_n\}}\max_{t=1,\ldots,T}\|\underline{\boldsymbol{\mathcal{P}}}_{0,t|T} - \underline{\boldsymbol{\mathcal{V}}}_{0,t|T}\| = O_p(n^{-1}\log^{1/\delta_v}T). \tag{E.52}$$

This bound is tighter when $\underline{\boldsymbol{\varphi}}_n = \boldsymbol{\varphi}_n$. Indeed, we have $\boldsymbol{\mathcal{P}}_{0,t|T}(\boldsymbol{\varphi}_n) = \mathbf{P}_{0,t|T}$, which is deterministic by Lemma D.6(iii), and

$$\begin{aligned}
\|\boldsymbol{\mathcal{P}}_{0,t|T}(\boldsymbol{\varphi}_n) - \boldsymbol{\mathcal{V}}_{0,t|T}(\boldsymbol{\varphi}_n)\| &= \|\mathbf{P}_{0,t|T} - \boldsymbol{\mathcal{V}}_{0,t|T}(\boldsymbol{\varphi}_n)\|\\
&\le \|\mathbf{P}_{0,t|T} - (\boldsymbol{\Lambda}_n'(\boldsymbol{\Gamma}_n^{\xi})^{-1}\boldsymbol{\Lambda}_n)^{-1}\| + \|(\boldsymbol{\Lambda}_n'(\boldsymbol{\Gamma}_n^{\xi})^{-1}\boldsymbol{\Lambda}_n)^{-1} - \boldsymbol{\mathcal{V}}_{0,t|T}(\boldsymbol{\varphi}_n)\|\\
&= O(n^{-2}),
\end{aligned} \tag{E.53}$$

by (E.50) and Lemma E.8(vi).

Now, consider

$$\begin{aligned}
&\ell(\boldsymbol{X}_{nT};\underline{\boldsymbol{\phi}}_n,\underline{\boldsymbol{\theta}}) - \ell_0(\boldsymbol{X}_{nT};\underline{\boldsymbol{\phi}}_n)\\
&= -\frac{1}{2}\sum_{t=1}^T\Bigg\{\left(\underline{\mathbf{F}}_{0,t|T} - \underline{\mathbf{F}}_{0,t}^{\mathrm{REG}}\right)'\underline{\boldsymbol{\Lambda}}_n'(\underline{\boldsymbol{\Sigma}}_n^{\xi})^{-1}\underline{\boldsymbol{\Lambda}}_n\left(\underline{\mathbf{F}}_{0,t|T} - \underline{\mathbf{F}}_{0,t}^{\mathrm{REG}}\right)\\
&\quad + 2\left(\underline{\mathbf{F}}_{0,t|T} - \underline{\mathbf{F}}_{0,t}^{\mathrm{REG}}\right)'\underline{\boldsymbol{\Lambda}}_n'(\underline{\boldsymbol{\Sigma}}_n^{\xi})^{-1}\underline{\boldsymbol{\Lambda}}_n\underline{\mathbf{F}}_{0,t}^{\mathrm{REG}} + 2\mathbf{x}_{nt}'(\underline{\boldsymbol{\Sigma}}_n^{\xi})^{-1}\underline{\boldsymbol{\Lambda}}_n\left(\underline{\mathbf{F}}_{0,t|T} - \underline{\mathbf{F}}_{0,t}^{\mathrm{REG}}\right)\Bigg\} - \frac{T}{2}\log\det(\underline{\boldsymbol{\Gamma}}^v)\\
&\quad - \frac{1}{2}\sum_{t=1}^T(\underline{\mathbf{F}}_{0,t|T} - \underline{\mathbf{A}}\,\underline{\mathbf{F}}_{0,t-1|T})'(\underline{\boldsymbol{\Gamma}}^v)^{-1}(\underline{\mathbf{F}}_{0,t|T} - \underline{\mathbf{A}}\,\underline{\mathbf{F}}_{0,t-1|T}) + \frac{1}{2}\sum_{t=1}^T\underline{\mathbf{F}}_{0,t}^{\mathrm{REG}\prime}\underline{\mathbf{F}}_{0,t}^{\mathrm{REG}}\\
&\quad + \frac{1}{2}\sum_{t=1}^T\log\det(\underline{\boldsymbol{\mathcal{P}}}_{0,t|T}) - \frac{1}{2}\sum_{t=1}^T\log\det(\underline{\boldsymbol{\mathcal{V}}}_{0,t|T}).
\end{aligned} \tag{E.54}$$

Consider all terms on the rhs of (E.54). First, by (E.47), for all $\underline{\varphi}_n \in \mathcal{O}_n$,

$$\begin{aligned}
&\Bigg| -\frac{1}{2}\sum_{t=1}^{T}\Bigg\{ \left(\underline{\mathbf{F}}_{0,t|T}-\underline{\mathbf{F}}_{0,t}^{\mathrm{REG}}\right)' \underline{\mathbf{\Lambda}}_n'(\underline{\mathbf{\Sigma}}_n^{\xi})^{-1}\underline{\mathbf{\Lambda}}_n\left(\underline{\mathbf{F}}_{0,t|T}-\underline{\mathbf{F}}_{0,t}^{\mathrm{REG}}\right) + 2\left(\underline{\mathbf{F}}_{0,t|T}-\underline{\mathbf{F}}_{0,t}^{\mathrm{REG}}\right)'\underline{\mathbf{\Lambda}}_n'(\underline{\mathbf{\Sigma}}_n^{\xi})^{-1}\underline{\mathbf{\Lambda}}_n\underline{\mathbf{F}}_{0,t}^{\mathrm{REG}} \\
&\quad + 2\mathbf{x}_{nt}'\underline{\mathbf{\Lambda}}_n'(\underline{\mathbf{\Sigma}}_n^{\xi})^{-1}\underline{\mathbf{\Lambda}}_n\left(\underline{\mathbf{F}}_{0,t|T}-\underline{\mathbf{F}}_{0,t}^{\mathrm{REG}}\right)\Bigg\}\Bigg| \\
&\le T\max_{t=1,\ldots,T}\|\underline{\mathbf{F}}_{0,t|T}-\underline{\mathbf{F}}_{0,t}^{\mathrm{REG}}\|\,\|\underline{\mathbf{\Lambda}}_n\|^2\,\|(\underline{\mathbf{\Sigma}}_n^{\xi})^{-1}\|\max_{t=1,\ldots,T}\|\underline{\mathbf{F}}_{0,t}^{\mathrm{REG}}\| \\
&\quad + T\max_{t=1,\ldots,T}\|\underline{\mathbf{F}}_{0,t|T}-\underline{\mathbf{F}}_{0,t}^{\mathrm{REG}}\|\,\|(\underline{\mathbf{\Sigma}}_n^{\xi})^{-1}\|\,\|\underline{\mathbf{\Lambda}}_n\|\max_{t=1,\ldots,T}\|\mathbf{x}_{nt}\| \\
&\quad + T\max_{t=1,\ldots,T}\|\underline{\mathbf{F}}_{0,t|T}-\underline{\mathbf{F}}_{0,t}^{\mathrm{REG}}\|^2\,\|\underline{\mathbf{\Lambda}}_n\|^2\,\|(\underline{\mathbf{\Sigma}}_n^{\xi})^{-1}\|\max_{t=1,\ldots,T}\|\mathbf{x}_{nt}\| \\
&= O_p(T\log^{2/\delta_v}T),
\end{aligned}\tag{E.55}$$

where we used Lemmas C.2 and E.3, which implies also that $\|\underline{\mathbf{F}}_{0,t}^{\mathrm{REG}}\| = O_p(\log^{1/\delta_v}T)$, and Assumption 2(a).

Second, for all $\underline{\varphi}_n \in \mathcal{O}_n$,

$$\begin{aligned}
&\Bigg| -\frac{T}{2}\log\det(\underline{\mathbf{\Gamma}}^v) - \frac{1}{2}\sum_{t=1}^{T}(\underline{\mathbf{F}}_{0,t|T}-\underline{\mathbf{A}}\,\underline{\mathbf{F}}_{0,t-1|T})'(\underline{\mathbf{\Gamma}}^v)^{-1}(\underline{\mathbf{F}}_{0,t|T}-\underline{\mathbf{A}}\,\underline{\mathbf{F}}_{0,t-1|T}) + \frac{1}{2}\sum_{t=1}^{T}\underline{\mathbf{F}}_{0,t}^{\mathrm{REG}\prime}\underline{\mathbf{F}}_{0,t}^{\mathrm{REG}}\Bigg| \\
&\le \left|\frac{T}{2}\log\det(\underline{\mathbf{\Gamma}}^v)\right| + \left|\frac{1}{2}\sum_{t=1}^{T}(\underline{\mathbf{F}}_{0,t|T}-\underline{\mathbf{A}}\,\underline{\mathbf{F}}_{0,t-1|T})'(\underline{\mathbf{\Gamma}}^v)^{-1}(\underline{\mathbf{F}}_{0,t|T}-\underline{\mathbf{A}}\,\underline{\mathbf{F}}_{0,t-1|T})\right| + \left|\frac{1}{2}\sum_{t=1}^{T}\underline{\mathbf{F}}_{0,t}^{\mathrm{REG}\prime}\underline{\mathbf{F}}_{0,t}^{\mathrm{REG}}\right| \\
&= O_p(T\log^{2/\delta_v}T),
\end{aligned}\tag{E.56}$$

because of (E.46), Assumptions 1(d) and 1(e), and Lemma E.3 jointly with the same arguments leading to (E.55).

Third, by (E.48) and (E.52), and Merikoski and Kumar (2004, Theorem 1), which is Weyl's inequality

$$\begin{aligned}
&\left|\frac{1}{2}\sum_{t=1}^{T}\log\det(\underline{\boldsymbol{\mathcal{P}}}_{0,t|T}) - \frac{1}{2}\sum_{t=1}^{T}\log\det(\underline{\boldsymbol{\mathcal{V}}}_{0,t|T})\right| = \left|\frac{1}{2}\sum_{t=1}^{T}\log\left(\det\{\underline{\boldsymbol{\mathcal{P}}}_{0,t|T}(\underline{\boldsymbol{\mathcal{V}}}_{0,t|T})^{-1}\}\right)\right| \\
&= \left|\frac{1}{2}\sum_{t=1}^{T}\sum_{j=1}^{r}\log\left(\left\{\nu^{(j)}(\underline{\boldsymbol{\mathcal{P}}}_{0,t|T})\right\}\left\{\nu^{(j)}(\underline{\boldsymbol{\mathcal{V}}}_{0,t|T})\right\}^{-1}\right)\right| \\
&\le \left|\frac{1}{2}\sum_{t=1}^{T}\sum_{j=1}^{r}\log\left(\left\{\nu^{(1)}(\underline{\boldsymbol{\mathcal{P}}}_{0,t|T}-\underline{\boldsymbol{\mathcal{V}}}_{0,t|T}) + \nu^{(j)}(\underline{\boldsymbol{\mathcal{V}}}_{0,t|T})\right\}\left\{\nu^{(j)}(\underline{\boldsymbol{\mathcal{V}}}_{0,t|T})\right\}^{-1}\right)\right| \\
&\le \left|\frac{1}{2}\sum_{t=1}^{T}\sum_{j=1}^{r}\log\left(1+\left\{n\|\underline{\boldsymbol{\mathcal{P}}}_{0,t|T}-\underline{\boldsymbol{\mathcal{V}}}_{0,t|T}\|\right\}\left\{n\nu^{(j)}(\underline{\boldsymbol{\mathcal{V}}}_{0,t|T})\right\}^{-1}\right)\right| \\
&\le \max_{t=1,\ldots,T}\left|\frac{T}{2}\left\{n\|\underline{\boldsymbol{\mathcal{P}}}_{0,t|T}-\underline{\boldsymbol{\mathcal{V}}}_{0,t|T}\|\right\}\left\{n\nu^{(j)}(\underline{\boldsymbol{\mathcal{V}}}_{0,t|T})\right\}^{-1}\right| + o(Tn^{-1}) \\
&= O_p(Tn^{-1}\log^{1/\delta_v}T),
\end{aligned}\tag{E.57}$$

where in the second last line we took into account also (E.49) and (E.53), hence, (E.57) holds for all $\underline{\varphi}_n \in \mathcal{O}_n$.

Summing up, by noticing that (E.55), (E.56), and (E.57) hold for all $\underline{\varphi}_n \in \mathcal{O}_n$, from (E.54) we have:

$$\sup_{\underline{\varphi}_n\in\mathcal{O}_n}(nT)^{-1}\left|\ell(\boldsymbol{X}_{nT};\underline{\boldsymbol{\phi}}_n,\underline{\boldsymbol{\theta}}) - \ell_0(\boldsymbol{X}_{nT};\underline{\boldsymbol{\phi}}_n)\right| = O_p(n^{-1}\log^{2/\delta_v}T).$$

This completes the proof. □

Lemma E.10. *Let $\widehat{\boldsymbol{\phi}}_n^* = (\mathrm{vec}(\widehat{\mathbf{\Lambda}}_n^*)'\ \widehat{\sigma}_1^{2*}\cdots\widehat{\sigma}_n^{2*})'$ be the vector of QML estimators of the entries of $\boldsymbol{\phi}_n$ maximizing $\ell(\boldsymbol{X}_{nT};\underline{\boldsymbol{\phi}}_n,\underline{\boldsymbol{\theta}})$ defined in (5), and let $\widehat{\boldsymbol{\phi}}_n^\dagger = (\mathrm{vec}(\widehat{\mathbf{\Lambda}}_n^\dagger)'\ \widehat{\sigma}_1^{2\dagger}\cdots\widehat{\sigma}_n^{2\dagger})'$ be the vector of QML estimators of the entries of $\boldsymbol{\phi}_n$ maximizing $\ell_0(\boldsymbol{X}_{nT};\underline{\boldsymbol{\phi}}_n)$ defined in (E.37) in the proof of Lemma E.9. Then, under Assumptions 1, 2, 3, 4, 5, and 6, as $n,T\to\infty$,*

(i) $n\log^{-2/\delta_v}T\max_{i=1,\ldots,n}\|\widehat{\boldsymbol{\lambda}}_i^* - \widehat{\boldsymbol{\lambda}}_i^\dagger\| = O_p(1)$;

(ii) $n\log^{-2/\delta_v}T n^{-1/2}\|\widehat{\mathbf{\Lambda}}_n^* - \widehat{\mathbf{\Lambda}}_n^\dagger\| = O_p(1)$;

(iii) $n\log^{-2/\delta_v}T\max_{i=1,\ldots,n}|\widehat{\sigma}_i^{2*} - \widehat{\sigma}_i^{2\dagger}| = O_p(1)$.

Proof. From Lemma E.9 we have

$$\begin{aligned}(nT)^{-1}&\left|\sup_{\underline{\boldsymbol{\phi}}_n\in\mathcal{O}_n}\ell(\boldsymbol{X}_{nT};\underline{\boldsymbol{\phi}}_n,\underline{\boldsymbol{\theta}})-\sup_{\underline{\boldsymbol{\phi}}_n\in\mathcal{O}_n}\ell_0(\boldsymbol{X}_{nT};\underline{\boldsymbol{\phi}}_n)\right|\\&\le(nT)^{-1}\sup_{\underline{\boldsymbol{\phi}}_n\in\mathcal{O}_n}\left|\ell(\boldsymbol{X}_{nT};\underline{\boldsymbol{\phi}}_n,\underline{\boldsymbol{\theta}})-\ell_0(\boldsymbol{X}_{nT};\underline{\boldsymbol{\phi}}_n)\right|\\&=O_p(n^{-1}\log^{2/\delta_v}T).\end{aligned}\tag{E.58}$$

Therefore, by continuity of the log-likelihoods and (E.58), we have

$$\begin{aligned}\max_{i=1,\dots,n}\begin{pmatrix}\|\widehat{\boldsymbol{\lambda}}_i^*-\widehat{\boldsymbol{\lambda}}_i^\dagger\|\\|\widehat{\sigma}_i^{2*}-\widehat{\sigma}_i^{2\dagger}|\end{pmatrix}&=\max_{i=1,\dots,n}\left\|\operatorname*{argmax}_{(\underline{\boldsymbol{\lambda}}_i'\underline{\sigma}_i^2)'\in\mathcal{O}_n}\ell(\boldsymbol{X}_{nT};\underline{\boldsymbol{\phi}}_n,\underline{\boldsymbol{\theta}})-\operatorname*{argmax}_{(\underline{\boldsymbol{\lambda}}_i'\underline{\sigma}_i^2)'\in\mathcal{O}_n}\ell_0(\boldsymbol{X}_{nT};\underline{\boldsymbol{\phi}}_n)\right\|\\&=\max_{i=1,\dots,n}\left\|\operatorname*{argmax}_{(\underline{\boldsymbol{\lambda}}_i'\underline{\sigma}_i^2)'\in\mathcal{O}_1}(nT)^{-1}\ell(\boldsymbol{X}_{nT};\underline{\boldsymbol{\phi}}_n,\underline{\boldsymbol{\theta}})-\operatorname*{argmax}_{(\underline{\boldsymbol{\lambda}}_i'\underline{\sigma}_i^2)'\in\mathcal{O}_n}(nT)^{-1}\ell_0(\boldsymbol{X}_{nT};\underline{\boldsymbol{\phi}}_n)\right\|\\&=O_p(n^{-1}\log^{2/\delta_v}T).\end{aligned}$$

This proves parts (i) and (iii), while part (ii) is a direct consequence of part (i). This completes the proof. □

Lemma E.11. *Under Assumptions 1, 2, 3, 4, 5, and 6, as $n,T\to\infty$,*
(i) $\min(n\log^{-2/\delta_v}T,\sqrt{nT},T)\,\|\widehat{\boldsymbol{\lambda}}_i^*-\boldsymbol{\lambda}_i^{\text{OLS}}\|=O_p(1)$*, uniformly in i;*
(ii) $\min(n\log^{-2/\delta_v}T,\sqrt{nT},T)\,n^{-1/2}\|\widehat{\boldsymbol{\Lambda}}_n^*-\boldsymbol{\Lambda}_n^{\text{OLS}}\|=O_p(1)$*;*
(iii) $\min(n\log^{-2/\delta_v}T,\sqrt{nT},T)\,|\widehat{\sigma}_i^{2*}-\sigma_i^{2\text{OLS}}|=O_p(1)$*, uniformly in i.*

Proof. For part (i)

$$\|\widehat{\boldsymbol{\lambda}}_i^*-\boldsymbol{\lambda}_i^{\text{OLS}}\|\le\|\widehat{\boldsymbol{\lambda}}_i^*-\boldsymbol{\lambda}_i^\dagger\|+\|\widehat{\boldsymbol{\lambda}}_i^\dagger-\boldsymbol{\lambda}_i^{(0)}\|+\|\widehat{\boldsymbol{\lambda}}_i^{(0)}-\boldsymbol{\lambda}_i^{\text{OLS}}\|,\tag{E.59}$$

where $\widehat{\boldsymbol{\lambda}}_i^{(0)}$ is the PC estimator defined in Appendix A.1. Then, from Barigozzi (2023, Theorem 3)

$$\|\widehat{\boldsymbol{\lambda}}_i^\dagger-\widehat{\boldsymbol{\lambda}}_i^{(0)}\|=O_p(n^{-1}),\tag{E.60}$$

while from Barigozzi (2023, Corollary 1 and Proposition B.3)

$$\|\widehat{\boldsymbol{\lambda}}_i^{(0)}-\boldsymbol{\lambda}_i^{\text{OLS}}\|=O_p(\max(n^{-1},n^{-1/2}T^{-1/2},T^{-1})).\tag{E.61}$$

Part (i) then follows by substituting (E.60), (E.61), and Lemma E.10(i) into (E.59).

For part (ii), from Barigozzi (2023, Theorem 3)

$$n^{-1/2}\|\widehat{\boldsymbol{\Lambda}}_n^\dagger-\widehat{\boldsymbol{\Lambda}}_n^{(0)}\|=O_p(n^{-1}),\tag{E.62}$$

while from Barigozzi (2023, Corollary 1 and Proposition B.3)

$$n^{-1/2}\|\widehat{\boldsymbol{\Lambda}}_n^{(0)}-\boldsymbol{\Lambda}_n^{\text{OLS}}\|=O_p(\max(n^{-1},n^{-1/2}T^{-1/2},T^{-1})).\tag{E.63}$$

Then, part (ii) is proved analogously to part (i), using (E.62), (E.63), and Lemma E.10(ii).

For part (iii), let $\widehat{\sigma}_i^{2\ddagger}=T^{-1}\sum_{t=1}^T(x_{it}-\widehat{\boldsymbol{\lambda}}_i^{\dagger\prime}\mathbf{F}_t)^2$. Then, consider

$$|\widehat{\sigma}_i^{2*}-\sigma_i^{2\text{OLS}}|\le|\widehat{\sigma}_i^{2*}-\widehat{\sigma}_i^{2\dagger}|+|\widehat{\sigma}_i^{2\dagger}-\widehat{\sigma}_i^{2\ddagger}|+|\widehat{\sigma}_i^{2\ddagger}-\sigma_i^{2\text{OLS}}|.\tag{E.64}$$

From Bai and Li (2016, Theorem S.2 and eq. (S.33) in the online supplement) and noticing that the estimator of the idiosyncratic variances is unaffected by the chosen identifying constraints, we have

$$|\widehat{\sigma}_i^{2\dagger}-\widehat{\sigma}_i^{2\ddagger}|=O_p(\max(n^{-1},n^{-1/2}T^{-1/2},T^{-1})).\tag{E.65}$$

Moreover,

$$\begin{aligned}
|\widehat{\sigma}_i^{2\ddagger} - \sigma_i^{2\mathrm{OLS}}| &= \left| T^{-1}\sum_{t=1}^T (x_{it} - \widehat{\boldsymbol{\lambda}}_i^{\dagger\prime}\mathbf{F}_t)^2 - T^{-1}\sum_{t=1}^T (x_{it} - \boldsymbol{\lambda}_i^{\mathrm{OLS}\prime}\mathbf{F}_t)^2 \right| \\
&\le \left| T^{-1}\sum_{t=1}^T \left\{ \widehat{\boldsymbol{\lambda}}_i^{\dagger\prime}\mathbf{F}_t\mathbf{F}_t'\widehat{\boldsymbol{\lambda}}_i^{\dagger} - \boldsymbol{\lambda}_i^{\mathrm{OLS}\prime}\mathbf{F}_t\mathbf{F}_t'\boldsymbol{\lambda}_i^{\mathrm{OLS}} \right\} \right| + 2\left| T^{-1}\sum_{t=1}^T \left\{ \widehat{\boldsymbol{\lambda}}_i^{\dagger\prime}\mathbf{F}_t x_{it} - \boldsymbol{\lambda}_i^{\mathrm{OLS}\prime}\mathbf{F}_t x_{it} \right\} \right| \\
&\le \|\widehat{\boldsymbol{\lambda}}_i^{\dagger} - \boldsymbol{\lambda}_i^{\mathrm{OLS}}\|^2 \left\| T^{-1}\sum_{t=1}^T \mathbf{F}_t\mathbf{F}_t' \right\| + 2\|\widehat{\boldsymbol{\lambda}}_i^{\dagger} - \boldsymbol{\lambda}_i^{\mathrm{OLS}}\| \left\| T^{-1}\sum_{t=1}^T \mathbf{F}_t\mathbf{F}_t' \right\| \|\boldsymbol{\lambda}_i^{\mathrm{OLS}}\| \\
&\quad + 2\|\widehat{\boldsymbol{\lambda}}_i^{\dagger} - \boldsymbol{\lambda}_i^{\mathrm{OLS}}\| \left\| T^{-1}\sum_{t=1}^T \mathbf{F}_t x_{it} \right\| \\
&= O_p(\max(n^{-1}, n^{-1/2}T^{-1/2}, T^{-1})),
\end{aligned} \tag{E.66}$$

by (E.60), (E.61), Lemma C.12(i) combined with Assumption 6(b), Lemma C.12(ii), and since $\|\boldsymbol{\lambda}_i^{\mathrm{OLS}}\| \le \|\boldsymbol{\lambda}_i^{\mathrm{OLS}} - \boldsymbol{\lambda}_i\| + \|\boldsymbol{\lambda}_i\| = O_p(1)$ by Assumption 1(a) and Lemma D.19(i). Part (iii) then follows by substituting (E.65), (E.66), and Lemma E.10(iii) into (E.64). This completes the proof. □

Lemma E.12. *Let $\widehat{\boldsymbol{\theta}}^* = (\mathrm{vec}(\widehat{\mathbf{A}}^*)'\ \mathrm{vech}(\widehat{\boldsymbol{\Gamma}}^{v*})')'$ be the QML estimator of the entries of $\boldsymbol{\theta}$ maximizing $\ell(\boldsymbol{X}_{nT}; \underline{\boldsymbol{\phi}}_n, \underline{\boldsymbol{\theta}})$ defined in (5), then, under Assumptions 1, 2, 3, 4, 5, and 6, as $n, T \to \infty$,*

(i) $n \log^{-2/\delta_v} T\, \|\widehat{\mathbf{A}}^* - \mathbf{A}^{\mathrm{OLS}}\| = O_p(1)$;

(ii) $n \log^{-2/\delta_v} T\, \|\widehat{\boldsymbol{\Gamma}}^{v*} - \boldsymbol{\Gamma}^{v\mathrm{OLS}}\| = O_p(1)$.

Proof. Throughout we consider generic values of the parameters such that $\underline{\boldsymbol{\varphi}}_n \in \mathcal{O}_n$ where $\mathcal{O}_n = \{\mathcal{O}^n_{\lambda_i} \cap \mathcal{E}_{\Lambda_n}\} \times \{\mathcal{O}^n_{\sigma_i^2} \cap \mathcal{E}_{\Gamma_n^\xi}\} \times \mathcal{O}_{\mathcal{A}} \times \mathcal{O}_{\Gamma^v}$ as defined in Section 4.3.4. Thus the elements of $\underline{\boldsymbol{\varphi}}_n$ satisfy Assumptions 1(a), 1(d), 1(e), 2(a), 2(b), and 2(f).

The log-likelihood depends on $\underline{\mathbf{A}}$ and $\underline{\boldsymbol{\Gamma}}^v$ only through $\ell(\boldsymbol{F}_T; \underline{\boldsymbol{\theta}})$ and $\ell(\boldsymbol{F}_T|\boldsymbol{X}_{nT}; \underline{\boldsymbol{\phi}}_n, \underline{\boldsymbol{\theta}})$. Let us consider both log-likelihoods separately. First, since by Assumption 1(i) $\mathbf{F}_0 = \mathbf{0}_r$, we have

$$\ell(\boldsymbol{F}_T; \underline{\boldsymbol{\theta}}) = -\frac{T}{2}\log\det(\underline{\boldsymbol{\Gamma}}^v) - \frac{1}{2}\sum_{t=1}^T (\mathbf{F}_t - \underline{\mathbf{A}}\mathbf{F}_{t-1})'(\underline{\boldsymbol{\Gamma}}^v)^{-1}(\mathbf{F}_t - \underline{\mathbf{A}}\mathbf{F}_{t-1}),$$

which is clearly maximized by $\boldsymbol{\theta}^{\mathrm{OLS}}$.

Second, (see also (E.34), (E.35), and (E.41) in the proof of Lemma E.9)

$$\ell(\boldsymbol{F}_T|\boldsymbol{X}_{nT}; \underline{\boldsymbol{\phi}}_n, \underline{\boldsymbol{\theta}}) = -\frac{1}{2}\sum_{t=1}^T \log\det(\underline{\boldsymbol{\mathcal{P}}}_{0,t|T}) - \frac{1}{2}\sum_{t=1}^T (\mathbf{F}_t - \underline{\mathbf{F}}_{0,t|T})'(\underline{\boldsymbol{\mathcal{P}}}_{0,t|T})^{-1}(\mathbf{F}_t - \underline{\mathbf{F}}_{0,t|T}),$$

Now, by (E.47), (E.52), and (E.53) in the proof of Lemma E.9, we have that

$$\sup_{\underline{\boldsymbol{\varphi}}_n \in \mathcal{O}_n} \left| \ell(\boldsymbol{F}_T|\boldsymbol{X}_{nT}; \underline{\boldsymbol{\phi}}_n, \underline{\boldsymbol{\theta}}) - \ell_0(\boldsymbol{F}_T|\boldsymbol{X}_{nT}; \underline{\boldsymbol{\phi}}_n) \right| = O_p(n^{-1}\log^{2/\delta_v} T). \tag{E.67}$$

The proof of parts (i) and (ii) follows from (E.67) and by continuity of the log-likelihood and since $\ell_0(\boldsymbol{F}_T|\boldsymbol{X}_{nT}; \underline{\boldsymbol{\phi}}_n)$ does not depend on $\underline{\boldsymbol{\theta}}$. This completes the proof. □

Lemma E.13. *Under Assumptions 1, 2, 3, 4, 5, and 6, as $n, T \to \infty$,*

(i) $\min(n\log^{-2/\delta_v} T, \sqrt{T})\, \|\widehat{\boldsymbol{\lambda}}_i^* - \boldsymbol{\lambda}_i\| = O_p(1)$, *uniformly in i;*

(ii) $\min(n\log^{-2/\delta_v} T, \sqrt{T})\, n^{-1/2}\|\widehat{\boldsymbol{\Lambda}}_n^* - \boldsymbol{\Lambda}_n\| = O_p(1)$;

(iii) $\min(n\log^{-2/\delta_v} T, \sqrt{T})\, |\widehat{\sigma}_i^{2*} - \sigma_i^2| = O_p(1)$, *uniformly in i;*

(iv) $\min(n\log^{-2/\delta_v} T, \sqrt{T})\, \|\widehat{\mathbf{A}}^* - \mathbf{A}\| = O_p(1)$;

(v) $\min(n\log^{-2/\delta_v} T, \sqrt{T})\, \|\widehat{\boldsymbol{\Gamma}}^{v*} - \boldsymbol{\Gamma}^v\| = O_p(1)$.

Proof. The proof follows directly from Lemmas D.19, E.11, and E.12. □

Lemma E.14. *Under Assumptions 1, 2, 3, 4, 5, and 6, as $n, T \to \infty$,*

(i) $\min(\sqrt{T}\log^{-1/2} n, n\log^{-2/\delta_v} T)\, \max_{i=1,\dots,n} \|\widehat{\boldsymbol{\lambda}}_i^* - \boldsymbol{\lambda}_i\| = O_p(1)$;

(ii) $\min(\sqrt{T}\log^{-1/2} n, n\log^{-2/\delta_v} T)\max_{i=1,\dots,n}|\widehat{\sigma}_i^{2*}-\sigma_i^2| = O_p(1)$;
(iii) $\min(\sqrt{T}\log^{-1/2} n, n\log^{-2/\delta_v} T)\,\|\widehat{\boldsymbol{\Sigma}}_n^{\xi *}-\boldsymbol{\Sigma}_n^{\xi}\| = O_p(1)$;
(iv) $\|(\widehat{\boldsymbol{\Sigma}}_n^{\xi *})^{-1}\| = O_p(1)$;
(v) $\min(\sqrt{T}\log^{-1/2} n, n\log^{-2/\delta_v} T)\,\|(\widehat{\boldsymbol{\Sigma}}_n^{\xi *})^{-1}-(\boldsymbol{\Sigma}_n^{\xi})^{-1}\| = O_p(1)$.

Proof. For part (i) consider

$$\max_{i=1,\dots,n}\|\widehat{\boldsymbol{\lambda}}_i^*-\boldsymbol{\lambda}_i\| \le \max_{i=1,\dots,n}\|\widehat{\boldsymbol{\lambda}}_i^*-\widehat{\boldsymbol{\lambda}}_i^{(0)}\| + \max_{i=1,\dots,n}\|\widehat{\boldsymbol{\lambda}}_i^{(0)}-\boldsymbol{\lambda}_i\|. \tag{E.68}$$

First, note that (E.60) in the proof of Lemma E.11 holds for all i, this is seen from the proof of Barigozzi (2023, Theorem 3), thus, jointly with Lemma E.10(i),

$$\max_{i=1,\dots,n}\|\widehat{\boldsymbol{\lambda}}_i^*-\widehat{\boldsymbol{\lambda}}_i^{(0)}\| \le \max_{i=1,\dots,n}\|\widehat{\boldsymbol{\lambda}}_i^*-\widehat{\boldsymbol{\lambda}}_i^{\dagger}\| + \max_{i=1,\dots,n}\|\widehat{\boldsymbol{\lambda}}_i^{\dagger}-\widehat{\boldsymbol{\lambda}}_i^{(0)}\| = O_p(n^{-1}\log^{2/\delta_v} T). \tag{E.69}$$

Second, from Barigozzi (2023, equation (A.5) in the proof of Theorem 1 in the supplementary material), when imposing Assumption 6(b),

$$\begin{aligned}
\widehat{\boldsymbol{\lambda}}_i^{(0)}-\boldsymbol{\lambda}_i =& \left\{(nT)^{-1}\boldsymbol{\lambda}_i'\sum_{t=1}^T\sum_{j=1}^n \mathbf{F}_t\xi_{jt}\boldsymbol{\lambda}_j' n(\widehat{\boldsymbol{\Lambda}}_n^{(0)\prime}\widehat{\boldsymbol{\Lambda}}_n^{(0)})^{-1}\right\} + \left\{T^{-1}\sum_{t=1}^T \mathbf{F}_t\xi_{it}(\boldsymbol{\Lambda}_n'\boldsymbol{\Lambda}_n)(\widehat{\boldsymbol{\Lambda}}_n^{(0)\prime}\widehat{\boldsymbol{\Lambda}}_n^{(0)})^{-1}\right\}\\
&+\left\{(nT)^{-1}\sum_{t=1}^T\sum_{j=1}^n \xi_{it}\xi_{jt}\boldsymbol{\lambda}_j' n(\widehat{\boldsymbol{\Lambda}}_n^{(0)\prime}\widehat{\boldsymbol{\Lambda}}_n^{(0)})^{-1}\right\}\\
&+\left\{(nT)^{-1}\boldsymbol{\lambda}_i'\sum_{t=1}^T\sum_{j=1}^n \mathbf{F}_t\xi_{jt}(\widehat{\boldsymbol{\lambda}}_j^{(0)}-\boldsymbol{\lambda}_j)' n(\widehat{\boldsymbol{\Lambda}}_n^{(0)\prime}\widehat{\boldsymbol{\Lambda}}_n^{(0)})^{-1}\right\}\\
&+\left\{T^{-1}\sum_{t=1}^T \mathbf{F}_t\xi_{it}\boldsymbol{\Lambda}_n'(\widehat{\boldsymbol{\Lambda}}_n^{(0)}-\boldsymbol{\Lambda}_n)(\widehat{\boldsymbol{\Lambda}}_n^{(0)\prime}\widehat{\boldsymbol{\Lambda}}_n^{(0)})^{-1}\right\}\\
&+\left\{(nT)^{-1}\sum_{t=1}^T\sum_{j=1}^n \xi_{it}\xi_{jt}(\widehat{\boldsymbol{\lambda}}_j^{(0)}-\boldsymbol{\lambda}_j)' n(\widehat{\boldsymbol{\Lambda}}_n^{(0)\prime}\widehat{\boldsymbol{\Lambda}}_n^{(0)})^{-1}\right\}\\
=& 1.a+1.b+1.c+1.d+1.e+1.f,\quad \text{say},
\end{aligned}$$

where

$$\begin{aligned}
&\max_{i=1,\dots,n}\|1.a\| = O_p(n^{-1/2}T^{-1/2}),\\
&\max_{i=1,\dots,n}\|1.b\| = O_p(T^{-1/2}\sqrt{\log n}),\\
&\max_{i=1,\dots,n}\|1.c\| = O_p(\max(n^{-1}, n^{-1/2}T^{-1/2}\sqrt{\log n})),
\end{aligned}\tag{E.70}$$

which follows by using

- for 1.a Assumption 1(a), Lemma D.1(ii), and Barigozzi (2023, Proposition B.3(a));
- for 1.b Assumption 1(a), Lemma D.1(ii), and Lemma E.1(ii) and the union bound;
- for 1.c Assumption 1(a), Lemma D.1(ii), E.1(ii) and the union bound, and Barigozzi (2023, Proposition B.3(c)), where we also used the fact that $\max_{i=1,\dots,n}|(nT)^{-1}\sum_{t=1}^T\sum_{j=1}^n \mathbb{E}[\xi_{it}\xi_{jt}]| = O(n^{-1})$ by Assumption 2(b);

while 1.d, 1.e, and 1.f are dominated by and 1.a, 1.b, and 1.c, respectively, because of Lemma D.1(i). It follows that

$$\max_{i=1,\dots,n}\|\widehat{\boldsymbol{\lambda}}_i^{(0)}-\boldsymbol{\lambda}_i\| = O_p(\max(n^{-1}, T^{-1/2}\sqrt{\log n})). \tag{E.71}$$

By substituting (E.69) and (E.71) into (E.68), we have

$$\max_{i=1,\dots,n}\|\widehat{\boldsymbol{\lambda}}_i^*-\boldsymbol{\lambda}_i\| = O_p(\max(n^{-1}, T^{-1/2}\sqrt{\log n})) + O_p(n^{-1}\log^{2/\delta_v} T).$$

This proves part (i).

For part (ii), first, consider

$$\max_{i=1,\ldots,n}|\widehat{\sigma}_i^{2*}-\sigma_i^2|\le \max_{i=1,\ldots,n}|\widehat{\sigma}_i^{2*}-\widehat{\sigma}_i^{2\dagger}|+\max_{i=1,\ldots,n}|\widehat{\sigma}_i^{2\dagger}-\widehat{\sigma}_i^{2\ddagger}|\\ +\max_{i=1,\ldots,n}|\widehat{\sigma}_i^{2\ddagger}-\sigma_i^{2\text{OLS}}|+\max_{i=1,\ldots,n}|\sigma_i^{2\text{OLS}}-\sigma_i^2|. \tag{E.72}$$

From (E.65) in the proof of Lemma E.11

$$\max_{i=1,\ldots,n}|\widehat{\sigma}_i^{2\dagger}-\widehat{\sigma}_i^{2\ddagger}|=O_p(\max(n^{-1},T^{-1}\sqrt{\log n},n^{-1/2}T^{-1/2}\sqrt{\log n})). \tag{E.73}$$

Indeed from Bai and Li (2016, equation (S.25) in the online supplement) we see that (E.73) is decomposed into the sum of 13 terms, and all depend on i only through $\boldsymbol{\lambda}_i$ which is such that $\max_{i=1,\ldots,n}\|\boldsymbol{\lambda}_i\|\le M_\lambda$ by Assumption 1(a), with the exceptions of the following terms

$$A_1=\max_{i=1,\ldots,n}\left|(\widehat{\boldsymbol{\lambda}}_i^\dagger-\boldsymbol{\lambda}_i)'\left\{T^{-1}\sum_{t=1}^T\mathbf{F}_t\mathbf{F}_t'\right\}(\widehat{\boldsymbol{\lambda}}_i^\dagger-\boldsymbol{\lambda}_i)\right|=O_p(\max(T^{-1}\log n,n^{-2}\log^{4/\delta_v}T)),$$
$$A_2=\max_{i=1,\ldots,n}2\left|\boldsymbol{\lambda}_i'\widehat{\boldsymbol{\Lambda}}_n^{\dagger\prime}(\widehat{\boldsymbol{\Sigma}}_n^{\xi\dagger})^{-1}(\widehat{\boldsymbol{\Lambda}}_n^\dagger-\boldsymbol{\Lambda}_n)T^{-1}\sum_{t=1}^T\mathbf{F}_t\xi_{it}\right|=O_p((n^{-1/2}T^{-1/2}\sqrt{\log n}),$$
$$A_3=\max_{i=1,\ldots,n}2\left|\boldsymbol{\lambda}_i'\widehat{\boldsymbol{\Lambda}}_n^{\dagger\prime}(\widehat{\boldsymbol{\Sigma}}_n^{\xi\dagger})^{-1}\left\{T^{-1}\sum_{t=1}^T\boldsymbol{\xi}_{nt}\right\}\left\{T^{-1}\sum_{t=1}^T\xi_{it}\right\}\right|=O_p(T^{-1}\sqrt{\log n}),$$

where we used:

- for A_1 part (i) and Lemmas E.10(i) and C.12(i) jointly with Assumption 6(b);
- for A_2 Lemma E.1(ii) and the union bound, plus Bai and Li (2012, Lemma B.4 and Corollary A.1 in the online supplement);
- for A_3 the same arguments used for A_2 plus Bai and Li (2016, Lemma S.10 in the online supplement).

Second, notice that, by Lemmas C.12(i) and D.1(ii), term $1.b$ in (E.70) is such that

$$1.b=(\boldsymbol{\lambda}_i^{\text{OLS}}-\boldsymbol{\lambda}_i)+o_p(T^{-1/2}\sqrt{\log n}). \tag{E.74}$$

Thus, from the above arguments and (E.74),

$$\max_{i=1,\ldots,n}\|\widehat{\boldsymbol{\lambda}}_i^{(0)}-\boldsymbol{\lambda}_i^{\text{OLS}}\|=O_p(\max(n^{-1},n^{-1/2}T^{-1/2}\sqrt{\log n})). \tag{E.75}$$

and, therefore, from (E.69) and (E.75)

$$\max_{i=1,\ldots,n}\|\widehat{\boldsymbol{\lambda}}_i^\dagger-\boldsymbol{\lambda}_i^{\text{OLS}}\|\le\max_{i=1,\ldots,n}\|\widehat{\boldsymbol{\lambda}}_i^\dagger-\widehat{\boldsymbol{\lambda}}_i^{(0)}\|+\max_{i=1,\ldots,n}\|\widehat{\boldsymbol{\lambda}}_i^{(0)}-\boldsymbol{\lambda}_i^{\text{OLS}}\|\\ =O_p(\max(n^{-1}\log^{2/\delta_v}T,n^{-1/2}T^{-1/2}\sqrt{\log n})). \tag{E.76}$$

Then, from (E.66) in the proof of Lemma E.11

$$\max_{i=1,\ldots,n}|\widehat{\sigma}_i^{2\ddagger}-\sigma_i^{2\text{OLS}}|\le\max_{i=1,\ldots,n}\|\widehat{\boldsymbol{\lambda}}_i^\dagger-\boldsymbol{\lambda}_i^{\text{OLS}}\|^2\left\|T^{-1}\sum_{t=1}^T\mathbf{F}_t\mathbf{F}_t'\right\|\\ +2\max_{i=1,\ldots,n}\|\widehat{\boldsymbol{\lambda}}_i^\dagger-\boldsymbol{\lambda}_i^{\text{OLS}}\|\left\|T^{-1}\sum_{t=1}^T\mathbf{F}_t\mathbf{F}_t'\right\|\max_{i=1,\ldots,n}\|\boldsymbol{\lambda}_i^{\text{OLS}}\|\\ +2\max_{i=1,\ldots,n}\|\widehat{\boldsymbol{\lambda}}_i^\dagger-\boldsymbol{\lambda}_i^{\text{OLS}}\|\left\|T^{-1}\sum_{t=1}^T\mathbf{F}_t\mathbf{F}_t'\right\|\max_{i=1,\ldots,n}\|\boldsymbol{\lambda}_i\|\\ +2\max_{i=1,\ldots,n}\|\widehat{\boldsymbol{\lambda}}_i^\dagger-\boldsymbol{\lambda}_i^{\text{OLS}}\|\max_{i=1,\ldots,n}\left\|T^{-1}\sum_{t=1}^T\mathbf{F}_t\xi_{it}\right\|\\ =O_p(\max(n^{-1}\log^{2/\delta_v}T,n^{-1/2}T^{-1/2}\sqrt{\log n})), \tag{E.77}$$

by (E.70), (E.76), Assumption 1(a), Lemma C.12(i) combined with Assumption 6(b), and since $\max_{i=1,\dots,n}\|\boldsymbol{\lambda}_i^{\text{OLS}}\| \le \max_{i=1,\dots,n}\|\boldsymbol{\lambda}_i^{\text{OLS}}-\boldsymbol{\lambda}_i\| + \max_{i=1,\dots,n}\|\boldsymbol{\lambda}_i\| = O_p(T^{-1/2}\sqrt{\log n}) + O(1)$ again by (E.70) and Assumption 1(a).

Finally,

$$\begin{aligned}\max_{i=1,\dots,n}|\sigma_i^{2\text{OLS}}-\sigma_i^2| \le{}& \max_{i=1,\dots,n}\left|T^{-1}\sum_{t=1}^T x_{it}^2-\mathbb{E}[x_{it}^2]\right| + \max_{i=1,\dots,n}\left|\boldsymbol{\lambda}_i^{\text{OLS}\prime}T^{-1}\sum_{t=1}^T\mathbf{F}_t\mathbf{F}_t'\boldsymbol{\lambda}_i^{\text{OLS}}-\boldsymbol{\lambda}_i'\boldsymbol{\lambda}_i\right|\\ &+2\max_{i=1,\dots,n}\left\|\boldsymbol{\lambda}_i^{\text{OLS}\prime}T^{-1}\sum_{t=1}^T\mathbf{F}_t\mathbf{F}_t'-\boldsymbol{\lambda}_i'\right\|\max_{i=1,\dots,n}\|\boldsymbol{\lambda}_i\|\\ &+2\max_{i=1,\dots,n}\|\boldsymbol{\lambda}_i^{\text{OLS}}\|\max_{i=1,\dots,n}\left\|T^{-1}\sum_{t=1}^T\mathbf{F}_t\xi_{it}\right\|\\ ={}&O_p(T^{-1/2}\sqrt{\log n}).\end{aligned}\tag{E.78}$$

Indeed, for the first term on the rhs of (E.78) we have

$$\begin{aligned}\max_{i=1,\dots,n}\left|T^{-1}\sum_{t=1}^T x_{it}^2-\mathbb{E}[x_{it}^2]\right| \le{}& \max_{i=1,\dots,n}\|\boldsymbol{\lambda}_i\|^2\left\|T^{-1}\sum_{t=1}^T\mathbf{F}_t\mathbf{F}_t'-\mathbf{I}_r\right\| + \max_{i=1,\dots,n}\left|T^{-1}\sum_{t=1}^T\xi_{it}^2-\mathbb{E}[\xi_{it}^2]\right|\\ &+2\max_{i=1,\dots,n}\|\boldsymbol{\lambda}_i\|\max_{i=1,\dots,n}\left\|T^{-1}\sum_{t=1}^T\mathbf{F}_t\xi_{it}\right\|\\ ={}&O_p(T^{-1/2}\sqrt{\log n}),\end{aligned}\tag{E.79}$$

by Assumption 1(a), Lemma C.12(i) for the first term, Barigozzi et al. (2018, Lemma 3.ii) and Fan et al. (2011, Lemmas A3 and B1), which we can apply because of Assumption 5(b) for the second term, and Lemma E.1(ii) and the union bound for the third term. For all other on the rhs of (E.78) we just need to use Assumption 1(a), Lemma C.12(i) and the fact that $\max_{i=1,\dots,n}\|\boldsymbol{\lambda}_i^{\text{OLS}}\| \le \max_{i=1,\dots,n}\|\boldsymbol{\lambda}_i^{\text{OLS}}-\boldsymbol{\lambda}_i\|+\max_{i=1,\dots,n}\|\boldsymbol{\lambda}_i\| = O_p(T^{-1/2}\sqrt{\log n})+O(1)$ again by (E.70) and Assumption 1(a).

By using Lemma E.10(iii), (E.73), (E.77), and (E.78) into (E.72) we have

$$\begin{aligned}\max_{i=1,\dots,n}|\widehat{\sigma}_i^{2*}-\sigma_i^2| ={}& O_p(n^{-1}\log^{2/\delta_v}T)+O_p(\max(n^{-1},T^{-1}\sqrt{\log n},n^{-1/2}T^{-1/2}\sqrt{\log n}))\\ &+O_p(\max(n^{-1}\log^{2/\delta_v}T,n^{-1/2}T^{-1/2}\sqrt{\log n}))+O_p(T^{-1/2}\sqrt{\log n})\\ ={}&O_p(\max(n^{-1}\log^{2/\delta_v}T,T^{-1/2}\sqrt{\log n})),\end{aligned}$$

which proves part (ii).

Part (iii) immediately follows from part (ii), indeed

$$\|\widehat{\boldsymbol{\Sigma}}_n^{\xi*}-\boldsymbol{\Sigma}_n^{\xi}\| \le \max_{i=1,\dots,n}|\widehat{\sigma}_i^{2*}-\sigma_i^2| = O_p(\max(n^{-1}\log^{2/\delta_v}T,T^{-1/2}\sqrt{\log n})).$$

For part (iv) we have

$$\|(\widehat{\boldsymbol{\Sigma}}_n^{\xi*})^{-1}\| = \left\{\min_{i=1,\dots,n}\widehat{\sigma}_i^{2*}\right\}^{-1} \le C_\xi + O_p(\max(n^{-1}\log^{2/\delta_v}T,T^{-1/2}\sqrt{\log n})),$$

because of part (ii) and Assumption 2(a).

To conclude, for part (v) we have

$$\|(\widehat{\boldsymbol{\Sigma}}_n^{\xi*})^{-1}-(\boldsymbol{\Sigma}_n^{\xi})^{-1}\| \le \|(\widehat{\boldsymbol{\Sigma}}_n^{\xi*})^{-1}\|\,\|\widehat{\boldsymbol{\Sigma}}_n^{\xi*}-\boldsymbol{\Sigma}_n^{\xi}\|\,\|(\boldsymbol{\Sigma}_n^{\xi})^{-1}\| = O_p(\max(n^{-1}\log^{2/\delta_v}T,T^{-1/2}\sqrt{\log n})),$$

by parts (iii), (iv), and Assumption 2(a). This completes the proof. □

Lemma E.15. *Under Assumptions 1, 2, 3, 4, 5, and 6, as $n,T\to\infty$:*

(i) $\min(n\log^{-2/\delta_v}T,\sqrt{T}\log^{-1/2}n)\,n^{-1}\|\widehat{\boldsymbol{\Lambda}}_n^{*\prime}(\widehat{\boldsymbol{\Sigma}}_n^{\xi*})^{-1}\widehat{\boldsymbol{\Lambda}}_n^*-\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^{\xi})^{-1}\boldsymbol{\Lambda}_n\| = O_p(1)$;

(ii) $\min(n\log^{-2/\delta_v}T,\sqrt{T}\log^{-1/2}n)\,n^{-1/2}\|\widehat{\boldsymbol{\Lambda}}_n^{*\prime}(\widehat{\boldsymbol{\Sigma}}_n^{\xi*})^{-1}-\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^{\xi})^{-1}\| = O_p(1)$;

(iii) $n\|(\widehat{\boldsymbol{\Lambda}}_n^{*\prime}(\widehat{\boldsymbol{\Sigma}}_n^{\xi*})^{-1}\widehat{\boldsymbol{\Lambda}}_n^*)^{-1}\| = O_p(1)$;

(iv) $\min(n\log^{-2/\delta_v}T,\sqrt{T}\log^{-1/2}n)\,n\|(\widehat{\mathbf{\Lambda}}_n^{*\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi*})^{-1}\widehat{\mathbf{\Lambda}}_n^*)^{-1}-(\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n)^{-1}\|=O_p(1)$;

(v) $\omega_{n,T\delta_v}\sqrt{n}\|(\widehat{\mathbf{\Lambda}}_n^{*\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi*})^{-1}\widehat{\mathbf{\Lambda}}_n^*)^{-1}\widehat{\mathbf{\Lambda}}_n^{*\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi*})^{-1}-(\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n)^{-1}\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\|=O_p(1)$, *with* $\omega_{n,T\delta_v}=\min(n\log^{-2/\delta_v}T,\sqrt{T}\log^{-1/2}n)$.

Proof. For part (i) we have

$$\begin{aligned}
n^{-1}\|\widehat{\mathbf{\Lambda}}_n^{*\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi*})^{-1}\widehat{\mathbf{\Lambda}}_n^*-\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n\| \le\,& 2n^{-1}\|\{\widehat{\mathbf{\Lambda}}_n^*-\mathbf{\Lambda}_n\}'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n\|\\
&+n^{-1}\|\mathbf{\Lambda}_n'\{(\widehat{\mathbf{\Sigma}}_n^{\xi*})^{-1}-(\mathbf{\Sigma}_n^\xi)^{-1}\}\mathbf{\Lambda}_n\|\\
&+2n^{-1}\|\{\widehat{\mathbf{\Lambda}}_n^*-\mathbf{\Lambda}_n\}'\{(\widehat{\mathbf{\Sigma}}_n^{\xi*})^{-1}-(\mathbf{\Sigma}_n^\xi)^{-1}\}\mathbf{\Lambda}_n\|\\
&+n^{-1}\|\{\widehat{\mathbf{\Lambda}}_n^*-\mathbf{\Lambda}_n\}'\{(\widehat{\mathbf{\Sigma}}_n^{\xi*})^{-1}-(\mathbf{\Sigma}_n^\xi)^{-1}\}\{\widehat{\mathbf{\Lambda}}_n^*-\mathbf{\Lambda}_n\}\|\\
\le\,& 2n^{-1/2}\|\widehat{\mathbf{\Lambda}}_n^*-\mathbf{\Lambda}_n\|\,\|(\mathbf{\Sigma}_n^\xi)^{-1}\|\,n^{-1/2}\|\mathbf{\Lambda}_n\|\\
&+\|\{(\widehat{\mathbf{\Sigma}}_n^{\xi*})^{-1}-(\mathbf{\Sigma}_n^\xi)^{-1}\}\|\,n^{-1}\|\mathbf{\Lambda}_n\|^2\\
&+2n^{-1/2}\|\widehat{\mathbf{\Lambda}}_n^*-\mathbf{\Lambda}_n\|\|\{(\widehat{\mathbf{\Sigma}}_n^{\xi*})^{-1}-(\mathbf{\Sigma}_n^\xi)^{-1}\}\|\,n^{-1/2}\|\mathbf{\Lambda}_n\|\\
&+n^{-1}\|\widehat{\mathbf{\Lambda}}_n^*-\mathbf{\Lambda}_n\|^2\|(\widehat{\mathbf{\Sigma}}_n^{\xi*})^{-1}-(\mathbf{\Sigma}_n^\xi)^{-1}\|\\
=\,& O_p(\max(n^{-1}\log^{2/\delta_v}T,T^{-1/2}\sqrt{\log n})),
\end{aligned}\tag{E.80}$$

by Assumptions 1(a), 2(a), and Lemmas E.13(ii) and E.14(v). This proves part (i).

Part (ii) is proved in the same way as part (i).

For part (iii), by part (ii) and Merikoski and Kumar (2004, Theorem 1) which is Weyl's inequality, we have

$$\begin{aligned}
&n^{-1}|\nu^{(r)}(\widehat{\mathbf{\Lambda}}_n^{*\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi*})^{-1}\widehat{\mathbf{\Lambda}}_n^*)-\nu^{(r)}(\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1})\mathbf{\Lambda}_n)|\\
&\quad\le n^{-1}\|\widehat{\mathbf{\Lambda}}_n^{*\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi*})^{-1}\widehat{\mathbf{\Lambda}}_n^*-\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n\|\\
&\quad=O_p(\max(n^{-1}\log^{2/\delta_v}T,T^{-1/2}\sqrt{\log n})).
\end{aligned}\tag{E.81}$$

Moreover (note that $x-y\ge-|x-y|$ for any $x,y\in\mathbb{R}$),

$$\begin{aligned}
\det(\widehat{\mathbf{\Lambda}}_n^{*\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi*})^{-1}\widehat{\mathbf{\Lambda}}_n^*)&=\prod_{j=1}^r\nu^{(j)}(\widehat{\mathbf{\Lambda}}_n^{*\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi*})^{-1}\widehat{\mathbf{\Lambda}}_n^*)\ge\left\{\nu^{(r)}(\widehat{\mathbf{\Lambda}}_n^{*\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi*})^{-1}\widehat{\mathbf{\Lambda}}_n^*)\right\}^r\\
&\ge\left\{\nu^{(r)}(\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n)-|\nu^{(r)}(\widehat{\mathbf{\Lambda}}_n^{*\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi*})^{-1}\widehat{\mathbf{\Lambda}}_n^*)-\nu^{(r)}(\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1})\mathbf{\Lambda}_n)|\right\}^r,
\end{aligned}$$

thus, by Lemma C.3(iv), which implies $\lim_{n\to\infty}n^{-1}\nu^{(r)}(\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n)>0$, and (E.81) it follows that, with probability tending to one as $n,T\to\infty$, we have $\det(n^{-1}\widehat{\mathbf{\Lambda}}_n^{*\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi*})^{-1}\widehat{\mathbf{\Lambda}}_n^*)>0$, or, equivalently $n^{-1}\widehat{\mathbf{\Lambda}}_n^{*\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi*})^{-1}\widehat{\mathbf{\Lambda}}_n^*$ is positive definite, i.e. $n\|(\widehat{\mathbf{\Lambda}}_n^{*\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi*})^{-1}\widehat{\mathbf{\Lambda}}_n^*)^{-1}\|=O_p(1)$. This proves part (iii).

For part (iv), we have

$$\begin{aligned}
&n\|(\widehat{\mathbf{\Lambda}}_n^{*\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi*})^{-1}\widehat{\mathbf{\Lambda}}_n^*)^{-1}-(\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n)^{-1}\|\\
&\quad\le n\|(\widehat{\mathbf{\Lambda}}_n^{*\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi*})^{-1}\widehat{\mathbf{\Lambda}}_n^*)^{-1}\|\,n^{-1}\|\widehat{\mathbf{\Lambda}}_n^{*\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi*})^{-1}\widehat{\mathbf{\Lambda}}_n^*-\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n\|\,n\|(\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n)^{-1}\|\\
&\quad=O_p(\max(n^{-1}\log^{2/\delta_v}T,T^{-1/2}\sqrt{\log n})),
\end{aligned}$$

because of parts (i) and (iii) and Lemma C.3(iii).

Part (v) follows directly from parts (ii) and (iv). This completes the proof. □

Lemma E.16. *Under Assumptions 1, 2, 3, 4, 5, and 6, as* $n,T\to\infty$*:*

(i) $\max_{t=1,\dots,T}\|\mathbf{P}_{t|t-1}^*\|=O_p(1)$;

(ii) $\max_{t=1,\dots,T}\|(\mathbf{P}_{t|t-1}^*)^{-1}\|=O_p(1)$;

(iii) $\max_{t=1,\dots,T}n\|\mathbf{P}_{t|t}^*\|=O_p(1)$;

(iv) $\max_{t=1,\dots,T}n\|\mathbf{P}_{t|T}^*\|=O_p(1)$.

Proof. For part (i),

$$\begin{aligned}\max_{t=1,\dots,T}\|\mathbf{P}^*_{t|t-1}\| &\le \max_{t=1,\dots,T}\|\mathbf{P}_{t|t-1}\|+\max_{t=1,\dots,T}\|\mathbf{P}^*_{t|t-1}-\mathbf{P}_{t|t-1}\|\\ &= O_p(1)+O_p(\max(n^{-1}\log^{2/\delta_v}T, T^{-1/2}\sqrt{\log n})),\end{aligned}$$

by Lemma D.7(i) and since the second term on the rhs depends only on the estimation error of $\widehat{\mathbf{A}}^*$, $\widehat{\boldsymbol{\Gamma}}^{v*}$, $n^{-1/2}\widehat{\boldsymbol{\Lambda}}^*_n$, $n^{-1/2}\widehat{\boldsymbol{\Lambda}}^{*\prime}_n(\widehat{\boldsymbol{\Sigma}}^{\xi*}_n)^{-1}$ and $n^{-1}(\widehat{\boldsymbol{\Lambda}}^{*\prime}_n(\widehat{\boldsymbol{\Sigma}}^{\xi*}_n)^{-1}\widehat{\boldsymbol{\Lambda}}^*_n)^{-1}$, which are all bounded by Lemmas E.13(ii), E.13(iv), E.13(v), E.15(ii), and E.15(iv). This proves part (i).

Part (ii) is proved in the same way as part (i) but using Lemma D.7(ii).

For part (iii), from (A.4) using the same steps leading to (D.23) in the proof of Lemma D.11 but when using as parameters $\widehat{\boldsymbol{\varphi}}^*_n$, we have

$$\begin{aligned}\mathbf{P}^*_{t|t} &= \mathbf{P}^*_{t|t-1}-\mathbf{P}^*_{t|t-1}\widehat{\boldsymbol{\Lambda}}^{*\prime}_n(\widehat{\boldsymbol{\Lambda}}^*_n\mathbf{P}^*_{t|t-1}\widehat{\boldsymbol{\Lambda}}^{*\prime}_n+\widehat{\boldsymbol{\Sigma}}^{\xi*}_n)^{-1}\widehat{\boldsymbol{\Lambda}}^*_n\mathbf{P}^*_{t|t-1}\\ &= (\widehat{\boldsymbol{\Lambda}}^{*\prime}_n(\widehat{\boldsymbol{\Sigma}}^{\xi*}_n)^{-1}\widehat{\boldsymbol{\Lambda}}^*_n)^{-1}\\ &\quad -\mathbf{P}^*_{t|t-1}((\widehat{\boldsymbol{\Lambda}}^{*\prime}_n(\widehat{\boldsymbol{\Sigma}}^{\xi*}_n)^{-1}\widehat{\boldsymbol{\Lambda}}^*_n)^{-1}+\mathbf{P}^*_{t|t-1})^{-1}(\widehat{\boldsymbol{\Lambda}}^{*\prime}_n(\widehat{\boldsymbol{\Sigma}}^{\xi*}_n)^{-1}\widehat{\boldsymbol{\Lambda}}^*_n)^{-1}(\mathbf{P}^*_{t|t-1})^{-1}(\widehat{\boldsymbol{\Lambda}}^{*\prime}_n(\widehat{\boldsymbol{\Sigma}}^{\xi*}_n)^{-1}\widehat{\boldsymbol{\Lambda}}^*_n)^{-1}.\end{aligned} \tag{E.82}$$

Notice that all inverses in (E.82) are well defined because of part (ii) and Lemmas E.15(iii) and E.14(iv).

Therefore, from (E.82)

$$\begin{aligned}\max_{t=1,\dots,T} n\|\mathbf{P}^*_{t|t}\| \le\, & n\|(\widehat{\boldsymbol{\Lambda}}^{*\prime}_n(\widehat{\boldsymbol{\Sigma}}^{\xi*}_n)^{-1}\widehat{\boldsymbol{\Lambda}}^*_n)^{-1}\|\\ &+\max_{t=1,\dots,T}\|\mathbf{P}^*_{t|t-1}\|\max_{t=1,\dots,T}\|(\mathbf{P}^*_{t|t-1})^{-1}\|\, n\|(\widehat{\boldsymbol{\Lambda}}^{*\prime}_n(\widehat{\boldsymbol{\Sigma}}^{\xi*}_n)^{-1}\widehat{\boldsymbol{\Lambda}}^*_n)^{-1}\|^2\\ &\cdot\|((\widehat{\boldsymbol{\Lambda}}^{*\prime}_n(\widehat{\boldsymbol{\Sigma}}^{\xi*}_n)^{-1}\widehat{\boldsymbol{\Lambda}}^*_n)^{-1}+\mathbf{P}^*_{t|t-1})^{-1}\|\\ =\, & O_p(1)+O_p(n^{-1}),\end{aligned}$$

because of parts (i) and (ii), Lemma E.15(iii), and since, by Merikoski and Kumar (2004, Theorem 1) which is Weyl's inequality,

$$\begin{aligned}&\|((\widehat{\boldsymbol{\Lambda}}^{*\prime}_n(\widehat{\boldsymbol{\Sigma}}^{\xi*}_n)^{-1}\widehat{\boldsymbol{\Lambda}}^*_n)^{-1}+\mathbf{P}^*_{t|t-1})^{-1}\| = \left\{\nu^{(r)}((\widehat{\boldsymbol{\Lambda}}^{*\prime}_n(\widehat{\boldsymbol{\Sigma}}^{\xi*}_n)^{-1}\widehat{\boldsymbol{\Lambda}}^*_n)^{-1}+\mathbf{P}^*_{t|t-1})\right\}^{-1}\\ &\le \left\{\nu^{(r)}((\widehat{\boldsymbol{\Lambda}}^{*\prime}_n(\widehat{\boldsymbol{\Sigma}}^{\xi*}_n)^{-1}\widehat{\boldsymbol{\Lambda}}^*_n)^{-1})+\nu^{(r)}(\mathbf{P}^*_{t|t-1})\right\}^{-1}\\ &= \left\{\left[\nu^{(1)}(\widehat{\boldsymbol{\Lambda}}^{*\prime}_n(\widehat{\boldsymbol{\Sigma}}^{\xi*}_n)^{-1}\widehat{\boldsymbol{\Lambda}}^*_n)\right]^{-1}+\nu^{(r)}(\mathbf{P}^*_{t|t-1})\right\}^{-1}\\ &= \left\{\left[\nu^{(1)}(\widehat{\boldsymbol{\Lambda}}^{*\prime}_n(\widehat{\boldsymbol{\Sigma}}^{\xi*}_n)^{-1}\widehat{\boldsymbol{\Lambda}}^*_n)\nu^{(r)}(\mathbf{P}^*_{t|t-1})\right]^{-1}+1\right\}^{-1}\left\{\nu^{(r)}(\mathbf{P}^*_{t|t-1})\right\}^{-1}\\ &= \left\{1-\left[\nu^{(1)}(\widehat{\boldsymbol{\Lambda}}^{*\prime}_n(\widehat{\boldsymbol{\Sigma}}^{\xi*}_n)^{-1}\widehat{\boldsymbol{\Lambda}}^*_n)\nu^{(r)}(\mathbf{P}^*_{t|t-1})\right]^{-1}\right\}\left\{\nu^{(r)}(\mathbf{P}^*_{t|t-1})\right\}^{-1}+O_p(n^{-2})\\ &= O_p(1),\end{aligned}$$

again by parts (i) and (ii) and Lemma E.15(iii). This proves part (iii).

For part (iv), from (A.7), we get

$$\|\mathbf{P}^*_{t|T}-\mathbf{P}^*_{t|t}\| \le \|\mathbf{P}^*_{t|t}\|^2\,\|\widehat{\mathbf{A}}^*\|^2\,\|(\mathbf{P}^*_{t+1|t})^{-1}\|^2\{\|\mathbf{P}^*_{t+1|T}\|+\|\mathbf{P}^*_{t+1|t}\|\}. \tag{E.83}$$

Start with $t=T-1$, then from (E.83),

$$\begin{aligned}\|\mathbf{P}^*_{T-1|T}-\mathbf{P}^*_{T-1|T-1}\| &\le \|\mathbf{P}^*_{T-1|T-1}\|^2\,\|\widehat{\mathbf{A}}^*\|^2\,\|(\mathbf{P}^*_{T|T-1})^{-1}\|^2\{\|\mathbf{P}^*_{T|T}\|+\|\mathbf{P}^*_{T|T-1}\|\}\\ &= O_p(n^{-2}).\end{aligned} \tag{E.84}$$

by parts (i), (ii), and (iii), and since $\|\widehat{\mathbf{A}}^*\|\le\|\mathbf{A}\|+\|\widehat{\mathbf{A}}^*-\mathbf{A}\|=O_p(1)$, by Assumption 1(d) and Lemma E.13(iv). From

(E.84) it follows that

$$\|\mathbf{P}^*_{T-1|T}\| \le \|\mathbf{P}^*_{T-1|T-1}\| + \|\mathbf{P}^*_{T-1|T} - \mathbf{P}^*_{T-1|T-1}\| = O_p(n^{-1}) + O_p(n^{-2}). \tag{E.85}$$

Thus, at $t = T-2$, from (E.83) and (E.85),

$$\begin{aligned}\|\mathbf{P}^*_{T-2|T} - \mathbf{P}^*_{T-2|T-2}\| &\le \|\mathbf{P}^*_{T-2|T-2}\|^2\, \|\widehat{\mathbf{A}}^*\|^2\, \|(\mathbf{P}^*_{T-1|T-2})^{-1}\|^2 \{\|\mathbf{P}^*_{T-1|T}\| + \|\mathbf{P}^*_{T-1|T-2}\|\}\\ &= O_p(n^{-2}).\end{aligned} \tag{E.86}$$

From (E.86) it follows that

$$\|\mathbf{P}^*_{T-2|T}\| \le \|\mathbf{P}^*_{T-2|T-2}\| + \|\mathbf{P}^*_{T-2|T} - \mathbf{P}^*_{T-2|T-2}\| = O_p(n^{-1}) + O_p(n^{-2}). \tag{E.87}$$

Since all the bounds in (E.84)-(E.87) are the same for all t, from part (i) and (E.83) we have

$$\max_{t=1,\dots,T} n\|\mathbf{P}^*_{t|T}\| \le \max_{t=1,\dots,T} n\|\mathbf{P}^*_{t|t}\| + \max_{t=1,\dots,T} n\|\mathbf{P}^*_{t|T} - \mathbf{P}^*_{t|t}\| = O_p(1) + O_p(n^{-1}).$$

This proves part (iv) and completes the proof. □

Lemma E.17. *Under Assumptions 1, 2, 3, 4, 5, and 6, as $n, T \to \infty$:*

(i) for all $s = 0, \dots, T$, $\|\mathbf{F}^*_{t|s}\| = O_p(1)$, uniformly in $t \le s$;

(ii) $n\|\mathbf{F}^*_{t|T} - \mathbf{F}^*_{t|t}\| = O_p(1)$, uniformly in t;

(iii) $n\|\mathbf{F}^*_{t|t} - \widehat{\mathbf{F}}^{\mathrm{WLS}*}_t\| = O_p(1)$, uniformly in t;

where $\widehat{\mathbf{F}}^{\mathrm{WLS}*}_t = (\widehat{\mathbf{\Lambda}}^{*\prime}_n(\widehat{\mathbf{\Sigma}}^{\xi*}_n)^{-1}\widehat{\mathbf{\Lambda}}^*_n)^{-1}\widehat{\mathbf{\Lambda}}^{*\prime}_n(\widehat{\mathbf{\Sigma}}^{\xi*}_n)^{-1}\mathbf{x}_{nt}$.

Proof. The proof of part (i) follows the same steps as the proof of Lemma D.14 but when using Lemmas E.13, E.14, and E.15, instead of Lemmas D.1, D.3, D.4, and D.5.

For part (ii), from (A.6) and (A.1)

$$\begin{aligned}\|\mathbf{F}^*_{t|T} - \mathbf{F}^*_{t|t}\| &\le \|\mathbf{P}^*_{t|t}\|\, \|\widehat{\mathbf{A}}^*\|\, \|(\mathbf{P}^*_{t+1|t})^{-1}\|\, \{\|\mathbf{F}^*_{t+1|T}\| + \|\mathbf{F}^*_{t+1|t}\|\}\\ &\le \|\mathbf{P}^*_{t|t}\|\, \|\widehat{\mathbf{A}}^*\|\, \|(\mathbf{P}^*_{t+1|t})^{-1}\|\, \{\|\mathbf{F}^*_{t+1|T}\| + \|\widehat{\mathbf{A}}^*\|\, \|\mathbf{F}^*_{t|t}\|\}\\ &= O_p(n^{-1}),\end{aligned}$$

by part (i) (when $s = T$ and $s = t$), Lemmas E.16(ii) and E.16(iii), and since $\|\widehat{\mathbf{A}}^*\| \le \|\mathbf{A}\| + \|\widehat{\mathbf{A}}^* - \mathbf{A}\| = O_p(1)$, by Assumption 1(d) and Lemma E.13(iv). This proves part (ii).

For part (iii), from (A.3) and (A.1), by using Lemma D.13 (see also (D.46) in the proof of Lemma D.16 for more details)

$$\begin{aligned}\mathbf{F}^*_{t|t} &= \mathbf{F}^*_{t|t-1} + \mathbf{P}^*_{t|t-1}\widehat{\mathbf{\Lambda}}^{*\prime}_n(\widehat{\mathbf{\Lambda}}^*_n\mathbf{P}^*_{t|t-1}\widehat{\mathbf{\Lambda}}^{*\prime}_n + \widehat{\mathbf{\Sigma}}^{\xi*}_n)^{-1}(\mathbf{x}_{nt} - \widehat{\mathbf{\Lambda}}^*_n\mathbf{F}^*_{t|t-1})\\ &= (\widehat{\mathbf{\Lambda}}^{*\prime}_n(\widehat{\mathbf{\Sigma}}^{\xi*}_n)^{-1}\widehat{\mathbf{\Lambda}}^*_n)^{-1}\widehat{\mathbf{\Lambda}}^{*\prime}_n(\widehat{\mathbf{\Sigma}}^{\xi*}_n)^{-1}\mathbf{x}_{nt}\\ &\quad + \left\{(\widehat{\mathbf{\Lambda}}^{*\prime}_n(\widehat{\mathbf{\Sigma}}^{\xi*}_n)^{-1}\widehat{\mathbf{\Lambda}}^*_n + (\mathbf{P}^*_{t|t-1})^{-1})^{-1} - (\widehat{\mathbf{\Lambda}}^{*\prime}_n(\widehat{\mathbf{\Sigma}}^{\xi*}_n)^{-1}\widehat{\mathbf{\Lambda}}^*_n)^{-1}\right\}\widehat{\mathbf{\Lambda}}^{*\prime}_n(\widehat{\mathbf{\Sigma}}^{\xi*}_n)^{-1}\mathbf{x}_{nt}\\ &\quad + \left\{\mathbf{I}_r - (\widehat{\mathbf{\Lambda}}^{*\prime}_n(\widehat{\mathbf{\Sigma}}^{\xi*}_n)^{-1}\widehat{\mathbf{\Lambda}}^*_n + (\widehat{\mathbf{P}}^*_{t|t-1})^{-1})^{-1}\widehat{\mathbf{\Lambda}}^{*\prime}_n(\widehat{\mathbf{\Sigma}}^{\xi*}_n)^{-1}\widehat{\mathbf{\Lambda}}^*_n\right\}\widehat{\mathbf{A}}^*\mathbf{F}^*_{t-1|t-1}.\end{aligned} \tag{E.88}$$

Notice that the inverses in (E.88) are all well defined by Lemmas E.14(iv), E.15(iii), and E.16(ii).

Now, by Lemma C.5

$$\|(\widehat{\mathbf{\Lambda}}^{*\prime}_n(\widehat{\mathbf{\Sigma}}^{\xi*}_n)^{-1}\widehat{\mathbf{\Lambda}}^*_n + (\mathbf{P}^*_{t|t-1})^{-1})^{-1}\widehat{\mathbf{\Lambda}}^{*\prime}_n(\widehat{\mathbf{\Sigma}}^{\xi*}_n)^{-1}\widehat{\mathbf{\Lambda}}^*_n - \mathbf{I}_r\| = O_p(n^{-1}). \tag{E.89}$$

Furthermore, by Lemmas C.6(iii) and E.15(iv)

$$\|(\widehat{\mathbf{\Lambda}}^{*\prime}_n(\widehat{\mathbf{\Sigma}}^{\xi*}_n)^{-1}\widehat{\mathbf{\Lambda}}^*_n + (\mathbf{P}^*_{t|t-1})^{-1})^{-1} - (\widehat{\mathbf{\Lambda}}^{*\prime}_n(\widehat{\mathbf{\Sigma}}^{\xi*}_n)^{-1}\widehat{\mathbf{\Lambda}}^*_n)^{-1}\| = O(n^{-2}), \tag{E.90}$$

and by Lemmas C.3(vii) and E.15(ii),

$$\|\widehat{\mathbf{\Lambda}}_n^{*\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi*})^{-1}\| = O_p(\sqrt{n}). \tag{E.91}$$

Indeed, we can apply Lemmas C.5 and C.6(iii), since $\|(\mathbf{P}_{t|t-1}^*)^{-1}\| = O_p(1)$ by Lemma E.16(ii), $\|(\widehat{\mathbf{\Sigma}}_n^{\xi*})^{-1}\| = O_p(1)$ by Lemma E.14(iv), and, by Lemmas C.2 and E.13(ii) we have

$$\begin{aligned} n^{-1}\|\widehat{\mathbf{\Lambda}}_n^{*\prime}\widehat{\mathbf{\Lambda}}_n^* - \mathbf{\Lambda}_n'\mathbf{\Lambda}_n\| &\le 2n^{-1}\|\mathbf{\Lambda}_n'(\widehat{\mathbf{\Lambda}}_n^* - \mathbf{\Lambda}_n)\| + n^{-1}\|(\widehat{\mathbf{\Lambda}}_n^* - \mathbf{\Lambda}_n)'(\widehat{\mathbf{\Lambda}}_n^* - \mathbf{\Lambda}_n)\| \\ &\le 2n^{-1/2}\|\mathbf{\Lambda}_n\|\, n^{-1/2}\|\widehat{\mathbf{\Lambda}}_n^* - \mathbf{\Lambda}_n\| + n^{-1}\|\widehat{\mathbf{\Lambda}}_n^* - \mathbf{\Lambda}_n\|^2 \\ &= O_p(\max(n^{-1}\log^{2/\delta_v} T, T^{-1/2})). \end{aligned}$$

which, by Weyl's inequality (Merikoski and Kumar, 2004, Theorem 1), implies

$$n^{-1}|\nu^{(j)}(\widehat{\mathbf{\Lambda}}_n^{*\prime}\widehat{\mathbf{\Lambda}}_n^*) - \nu^{(j)}(\mathbf{\Lambda}_n'\mathbf{\Lambda}_n)| \le n^{-1}\|\widehat{\mathbf{\Lambda}}_n^{*\prime}\widehat{\mathbf{\Lambda}}_n^* - \mathbf{\Lambda}_n'\mathbf{\Lambda}_n\| = O_p(\max(n^{-1}\log^{2/\delta_v} T, T^{-1/2})),$$

and, therefore, for $j = 1, \dots, r$,,

$$\underline{C}_j \le \text{p-}\liminf_{n,T\to\infty} n^{-1}\nu^{(j)}(\widehat{\mathbf{\Lambda}}_n^{*\prime}\widehat{\mathbf{\Lambda}}_n^*) \le \text{p-}\limsup_{n,T\to\infty} n^{-1}\nu^{(j)}(\widehat{\mathbf{\Lambda}}_n^{*\prime}\widehat{\mathbf{\Lambda}}_n^*) \le \overline{C}_j.$$

By using (E.89), (E.90), (E.91) into (E.88):

$$\begin{aligned} &\|\mathbf{F}_{t|t}^* - (\widehat{\mathbf{\Lambda}}_n^{*\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi*})^{-1}\widehat{\mathbf{\Lambda}}_n^*)^{-1}\widehat{\mathbf{\Lambda}}_n^{*\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi*})^{-1}\mathbf{x}_{nt}\| \\ &\le n\|(\widehat{\mathbf{\Lambda}}_n^{*\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi*})^{-1}\widehat{\mathbf{\Lambda}}_n^* + (\mathbf{P}_{t|t-1}^*)^{-1})^{-1} - (\widehat{\mathbf{\Lambda}}_n^{*\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi*})^{-1}\widehat{\mathbf{\Lambda}}_n^*)^{-1}\|\, n^{-1/2}\|\widehat{\mathbf{\Lambda}}_n^{*\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi*})^{-1}\|\, n^{-1/2}\|\mathbf{x}_{nt}\| \\ &\quad + \|\mathbf{I}_r - (\widehat{\mathbf{\Lambda}}_n^{*\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi*})^{-1}\widehat{\mathbf{\Lambda}}_n^* + (\widehat{\mathbf{P}}_{t|t-1}^*)^{-1})^{-1}\widehat{\mathbf{\Lambda}}_n^{*\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi*})^{-1}\widehat{\mathbf{\Lambda}}_n^*\|\, \|\widehat{\mathbf{A}}^*\|\, \|\mathbf{F}_{t-1|t-1}^*\| \\ &= O_p(n^{-1}), \end{aligned} \tag{E.92}$$

by part (i) (when $s = t-1$), Lemma C.10, and since $\|\widehat{\mathbf{A}}^*\| \le \|\mathbf{A}\| + \|\widehat{\mathbf{A}}^* - \mathbf{A}\| = O_p(1)$, by Assumption 1(d) and Lemma E.13(i). This completes the proof. □

Lemma E.18. *Under Assumptions 1, 2, 3, 4, 5, and 6, as $n, T \to \infty$, for $s = t$ and $s = T$:*

(i) $\|T^{-1}\sum_{t=1}^T \mathbf{F}_{t|s}^*\mathbf{F}_t'\| = O_p(1)$;

(i) $\|T^{-1}\sum_{t=1}^T \mathbf{F}_{t|s}^*\mathbf{F}_t'\boldsymbol{\lambda}_i\| = O_p(1)$, *uniformly in i;*

(ii) $\|T^{-1}\sum_{t=1}^T \mathbf{F}_{t|s}^*\xi_{it}\| = O_p(1)$, *uniformly in i.*

Proof. First notice that, for all $k = t-T, \dots, t-1$,

$$\left\| n^{-1/2}T^{-1}\sum_{t=1}^T \mathbf{x}_{n,t-k}\mathbf{F}_t' \right\| = O_p(1), \tag{E.93}$$

by Lemma C.12. The proof of part (i) follows by iterating either forward or backwards since both $\|T^{-1}\sum_{t=1}^T \mathbf{F}_{t|t}^*\mathbf{F}_t'\|$ and $\|T^{-1}\sum_{t=1}^T \mathbf{F}_{t|T}^*\mathbf{F}_t'\|$ are functions of (E.93) because of Lemma E.17.

Part (ii) follows from part (i) and Assumption 1(a). Part (iii) follows by substituting $\mathbf{F}_t$ with ξ_{it} in (E.93) and then by applying Lemma C.12(ii). This completes the proof. □

Lemma E.19. *Under Assumptions 1, 2, 3, 4, 5, and 6, as $n, T \to \infty$:*

$$\min(n\log^{-2/\delta_v} T, \sqrt{nT}, T\log^{-1/2} n) \left\| T^{-1}\sum_{t=1}^T \{\mathbf{F}_{t|T}^*\mathbf{F}_{t|T}^{*\prime} + \mathbf{P}_{t|T}^*\} - T^{-1}\sum_{t=1}^T \mathbf{F}_t\mathbf{F}_t' \right\| = O_p(1).$$

Proof. Start with

$$\begin{aligned} \left\| T^{-1}\sum_{t=1}^T \mathbf{F}_{t|T}^*\mathbf{F}_{t|T}^{*\prime} - T^{-1}\sum_{t=1}^T \mathbf{F}_t\mathbf{F}_t' \right\| &\le 2\left\| T^{-1}\sum_{t=1}^T (\mathbf{F}_{t|T}^* - \mathbf{F}_t)\mathbf{F}_t' \right\| \\ &\quad + \left\| T^{-1}\sum_{t=1}^T (\mathbf{F}_{t|T}^* - \mathbf{F}_t)(\mathbf{F}_{t|T}^* - \mathbf{F}_t)' \right\|, \end{aligned} \tag{E.94}$$

and notice that if the first term on the rhs is $o_p(1)$ then the second term is dominated by the first one. So let us consider the first term on the rhs of (E.94):

$$\begin{aligned}\left\|T^{-1}\sum_{t=1}^T(\mathbf{F}^*_{t|T}-\mathbf{F}_t)\mathbf{F}_t'\right\| \le& \left\|T^{-1}\sum_{t=1}^T(\mathbf{F}^*_{t|T}-\mathbf{F}^*_{t|t})\mathbf{F}_t'\right\|\\ &+\left\|T^{-1}\sum_{t=1}^T(\mathbf{F}^*_{t|t}-\widehat{\mathbf{F}}_t^{\text{WLS}*})\mathbf{F}_t'\right\|\\ &+\left\|T^{-1}\sum_{t=1}^T(\widehat{\mathbf{F}}_t^{\text{WLS}*}-\mathbf{F}_t)\mathbf{F}_t'\right\|\\ =& I^*+II^*+III^*, \quad \text{say.}\end{aligned} \tag{E.95}$$

Let us consider each term in (E.95). First,

$$\begin{aligned}I^* \le& \max_{t=1,\dots,T}\|\mathbf{P}^*_{t|t}\|\,\|\widehat{\mathbf{A}}^*\|\max_{t=1,\dots,T}\|(\mathbf{P}^*_{t+1|t})^{-1}\|\\ &\cdot\left\{\left\|T^{-1}\sum_{t=1}^T\mathbf{F}^*_{t+1|T}\mathbf{F}_t'\right\|+\|\widehat{\mathbf{A}}^*\|\left\|T^{-1}\sum_{t=1}^T\mathbf{F}^*_{t+1|t+1}\mathbf{F}_t'\right\|\right\}\\ =& O_p(n^{-1}),\end{aligned} \tag{E.96}$$

by Lemmas E.16(ii), E.16(iii), and E.18, and since $\|\widehat{\mathbf{A}}^*\|\le\|\mathbf{A}\|+\|\widehat{\mathbf{A}}^*-\mathbf{A}\|=O_p(1)$, by Assumption 1(d) and Lemma E.13(iv). Second, from (E.88) and (E.92) in the proof of Lemma E.17

$$\begin{aligned}II^* \le& O_p(n^{-1})\left\{\left\|n^{-1/2}T^{-1}\sum_{t=1}^T\mathbf{x}_{nt}\mathbf{F}_t'\right\|+\|\widehat{\mathbf{A}}^*\|\left\|T^{-1}\sum_{t=1}^T\mathbf{F}^*_{t-1|t-1}\mathbf{F}_t'\right\|\right\}\\ & O_p(n^{-1})\left\{n^{-1/2}\|\mathbf{\Lambda}_n|\left\|T^{-1}\sum_{t=1}^T\mathbf{F}_t\mathbf{F}_t'\right\|+\left\|n^{-1/2}T^{-1}\sum_{t=1}^T\boldsymbol{\xi}_{nt}\mathbf{F}_t'\right\|+\|\widehat{\mathbf{A}}^*\|\left\|T^{-1}\sum_{t=1}^T\mathbf{F}^*_{t-1|t-1}\mathbf{F}_t'\right\|\right\}\\ =& O_p(n^{-1}),\end{aligned} \tag{E.97}$$

because of Lemmas C.2, C.12(i), combined with Assumption 6(b), C.12(iii), and E.18 and since $\|\widehat{\mathbf{A}}^*\|=O_p(1)$ by Assumption 1(d) and Lemma E.13(iv).

Finally, let us consider the last term in (E.95). First, notice that we can write

$$\widehat{\mathbf{F}}_t^{\text{WLS}*}-\mathbf{F}_t=(\widehat{\mathbf{\Lambda}}_n^{*\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi*})^{-1}\widehat{\mathbf{\Lambda}}_n^*)^{-1}\widehat{\mathbf{\Lambda}}_n^{*\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi*})^{-1}(\mathbf{\Lambda}_n-\widehat{\mathbf{\Lambda}}_n^*)\mathbf{F}_t+(\widehat{\mathbf{\Lambda}}_n^{*\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi*})^{-1}\widehat{\mathbf{\Lambda}}_n^*)^{-1}\widehat{\mathbf{\Lambda}}_n^{*\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi*})^{-1}\boldsymbol{\xi}_{nt},$$

which implies

$$\begin{aligned}
III^* \leq &\left\| T^{-1}\sum_{t=1}^T (\widehat{\mathbf{\Lambda}}_n^{*\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi *})^{-1}\widehat{\mathbf{\Lambda}}_n^*)^{-1}\widehat{\mathbf{\Lambda}}_n^{*\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi *})^{-1}(\mathbf{\Lambda}_n - \widehat{\mathbf{\Lambda}}_n^*)\mathbf{F}_t\mathbf{F}_t'\right\| \\
&+ \left\| T^{-1}\sum_{t=1}^T (\widehat{\mathbf{\Lambda}}_n^{*\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi *})^{-1}\widehat{\mathbf{\Lambda}}_n^*)^{-1}\widehat{\mathbf{\Lambda}}_n^{*\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi *})^{-1}\boldsymbol{\xi}_{nt}\mathbf{F}_t'\right\| \\
\leq & n\|(\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^{\xi})^{-1}\mathbf{\Lambda}_n)^{-1}\|\, n^{-1}\|\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^{\xi})^{-1}(\mathbf{\Lambda}_n - \widehat{\mathbf{\Lambda}}_n^*)\| \left\| T^{-1}\sum_{t=1}^T \mathbf{F}_t\mathbf{F}_t'\right\| \\
&+ \sqrt{n}\|(\widehat{\mathbf{\Lambda}}_n^{*\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi *})^{-1}\widehat{\mathbf{\Lambda}}_n^*)^{-1}\widehat{\mathbf{\Lambda}}_n^{*\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi *})^{-1} - (\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^{\xi})^{-1}\mathbf{\Lambda}_n)^{-1}\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^{\xi})^{-1}\| \\
&\cdot n^{-1/2}\|\mathbf{\Lambda}_n - \widehat{\mathbf{\Lambda}}_n^*\| \left\| T^{-1}\sum_{t=1}^T \mathbf{F}_t\mathbf{F}_t'\right\| \\
&+ n\|(\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^{\xi})^{-1}\mathbf{\Lambda}_n)^{-1}\|\, n^{-1}\left\| T^{-1}\sum_{t=1}^T \mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^{\xi})^{-1}\boldsymbol{\xi}_{nt}\mathbf{F}_t'\right\| \\
&+ \sqrt{n}\|(\widehat{\mathbf{\Lambda}}_n^{*\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi *})^{-1}\widehat{\mathbf{\Lambda}}_n^*)^{-1}\widehat{\mathbf{\Lambda}}_n^{*\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi *})^{-1} - (\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^{\xi})^{-1}\mathbf{\Lambda}_n)^{-1}\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^{\xi})^{-1}\| \\
&\cdot n^{-1/2}\left\| T^{-1}\sum_{t=1}^T \boldsymbol{\xi}_{nt}\mathbf{F}_t'\right\| \\
= & III_a^* + III_b^* + III_c^* + III_d^*, \quad \text{say.}
\end{aligned} \tag{E.98}$$

Then,

$$\begin{aligned}
III_a^* \leq & n\|(\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^{\xi})^{-1}\mathbf{\Lambda}_n)^{-1}\|\, n^{-1}\|\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^{\xi})^{-1}(\mathbf{\Lambda}_n - \widehat{\mathbf{\Lambda}}_n^{\mathrm{OLS}})\| \left\| T^{-1}\sum_{t=1}^T \mathbf{F}_t\mathbf{F}_t'\right\| \\
&+ n\|(\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^{\xi})^{-1}\mathbf{\Lambda}_n)^{-1}\|\, n^{-1/2}\|\mathbf{\Lambda}_n\|\, \|(\mathbf{\Sigma}_n^{\xi})^{-1}\|\, n^{-1/2}\|\mathbf{\Lambda}_n^{\mathrm{OLS}} - \widehat{\mathbf{\Lambda}}_n^*\| \left\| T^{-1}\sum_{t=1}^T \mathbf{F}_t\mathbf{F}_t'\right\| \\
= & O_p(n^{-1/2}T^{-1/2}) + O_p(\max(n^{-1}\log^{2/\delta_v} T, n^{-1/2}T^{-1/2}, T^{-1})),
\end{aligned} \tag{E.99}$$

where we used: for the first term on the rhs Lemmas C.3(iii), C.12(i), combined with Assumption 6(b), and C.8(iv) since $n^{-1}\|\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^{\xi})^{-1}(\mathbf{\Lambda}_n - \widehat{\mathbf{\Lambda}}_n^{\mathrm{OLS}})\| = n^{-1}T^{-1}\|\sum_{t=1}^T \mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^{\xi})^{-1}\boldsymbol{\xi}_{nt}\mathbf{F}_t'\|$, while for the second term on the rhs we used Assumption 2(a), and Lemmas C.2, C.3(iii), C.12(i), combined with Assumption 6(b), and E.11(ii). Moreover,

$$III_b^* = O_p(\max(n^{-2}\log^{4/\delta_v} T, n^{-1}T^{-1/2}\log^{2/\delta_v} T\sqrt{\log n}, T^{-1}\sqrt{\log n})), \tag{E.100}$$

by Lemmas C.12(i), combined with Assumption 6(b), E.13(ii), and E.15(v),

$$III_c^* = O_p(n^{-1/2}T^{-1/2}), \tag{E.101}$$

by Lemmas C.3(iii) and C.8(iv), and

$$III_d^* = O_p(\max(n^{-1}T^{-1/2}\log^{2/\delta_v} T, T^{-1}\sqrt{\log n})), \tag{E.102}$$

by Lemmas C.12(iii) and E.15(v). By substituting (E.99), (E.100), (E.101), and (E.102) into (E.98) we have

$$III^* = O_p(\max(n^{-1}\log^{2/\delta_v} T, n^{-1/2}T^{-1/2}, T^{-1}\sqrt{\log n})). \tag{E.103}$$

Combining (E.96), (E.97), and (E.103) we have

$$\left\| T^{-1}\sum_{t=1}^T (\mathbf{F}_{t|T}^* - \mathbf{F}_t)\mathbf{F}_t'\right\| = O_p(\max(n^{-1}\log^{2/\delta_v} T, n^{-1/2}T^{-1/2}, T^{-1}\sqrt{\log n})), \tag{E.104}$$

which once substituted into (E.94), jointly with Lemma E.16(iv) give

$$\begin{aligned}&\left\|T^{-1}\sum_{t=1}^T\{\mathbf{F}_{t|T}^*\mathbf{F}_{t|T}^{*\prime}+\mathbf{P}_{t|T}^*\}-T^{-1}\sum_{t=1}^T\mathbf{F}_t\mathbf{F}_t'\right\|\\&\le\left\|T^{-1}\sum_{t=1}^T\mathbf{F}_{t|T}^*\mathbf{F}_{t|T}^{*\prime}-T^{-1}\sum_{t=1}^T\mathbf{F}_t\mathbf{F}_t'\right\|+\max_{t=1,\dots,T}\|\mathbf{P}_{t|T}^*\|\\&=O_p(\max(n^{-1}\log^{2/\delta_v}T,n^{-1/2}T^{-1/2},T^{-1}\sqrt{\log n}))+O_p(n^{-1}),\end{aligned}$$

which completes the proof. □

Lemma E.20. *Under Assumptions 1, 2, and 6, the EM estimators of the parameters* $\widehat{\varphi}_n\equiv\widehat{\varphi}_n^{(k+1)}$ *defined in Section 3.2, exist and are unique, for any* $k\ge 0$.

Proof. Let $\mathcal{O}_i=\mathcal{O}_{\lambda_i}\times\mathcal{O}_{\sigma_i^2}\subset\mathbb{R}^{r+1}$, and $\mathcal{O}_r=\mathcal{O}_{\mathcal{A}}\times\mathcal{O}_{\Gamma^v}\subset\mathbb{R}^{r^2+r(r+1)/2}$ with $\mathcal{O}_{\lambda_i}$, $\mathcal{O}_{\sigma_i^2}$, $\mathcal{O}_{\mathcal{A}}$, and $\mathcal{O}_{\Gamma^v}$ defined in Section 4.3.4.

At any iteration $k\ge 0$ the M-step requires solving the $n+1$ finite dimensional maximizations:

$$(\widehat{\boldsymbol{\lambda}}_i^{(k+1)},\widehat{\sigma}_i^{2(k+1)})=\underset{(\underline{\boldsymbol{\lambda}}_i,\underline{\sigma}_i^2)\in\mathcal{O}_i}{\operatorname{argmax}}\ \mathbb{E}_{\widehat{\varphi}_n^{(k)}}[\ell_i(\mathbf{x}_{i1}\dots\mathbf{x}_{iT}|\boldsymbol{F}_T;\underline{\boldsymbol{\lambda}}_i,\underline{\sigma}_i^2)],\qquad i=1,\dots,n,\tag{E.105}$$

and

$$(\widehat{\mathbf{A}}^{(k+1)},\widehat{\boldsymbol{\Gamma}}^{v(k+1)})=\underset{(\underline{\mathbf{A}},\underline{\boldsymbol{\Gamma}}^v)\in\mathcal{O}_r}{\arg\max}\ \mathbb{E}_{\widehat{\varphi}_n^{(k)}}[\ell(\boldsymbol{F}_T;\underline{\mathbf{A}},\underline{\boldsymbol{\Gamma}}^v)],\tag{E.106}$$

where

$$\ell_i(\mathbf{x}_{i1}\dots\mathbf{x}_{iT}|\boldsymbol{F}_T;\underline{\boldsymbol{\lambda}}_i,\underline{\sigma}_i^2)=-\frac{T}{2}\log(\underline{\sigma}_i^2)-\frac12\sum_{t=1}^T\frac{(x_{it}-\underline{\boldsymbol{\lambda}}_i'\mathbf{F}_t)^2}{\underline{\sigma}_i^2},\tag{E.107}$$

and

$$\ell(\boldsymbol{F}_T;\underline{\mathbf{A}},\underline{\boldsymbol{\Gamma}}^v)=-\frac{T}{2}\log\det(\underline{\boldsymbol{\Gamma}}^v)-\frac12\sum_{t=1}^T(\mathbf{F}_t-\underline{\mathbf{A}}\mathbf{F}_{t-1})'(\underline{\boldsymbol{\Gamma}}^v)^{-1}(\mathbf{F}_t-\underline{\mathbf{A}}\mathbf{F}_{t-1}).\tag{E.108}$$

Now, the log-likelihoods (E.107) and (E.108) to be maximized are continuous and differentiable in $\mathcal{O}_i$ and $\mathcal{O}_r$, which are a compact sets by Assumptions 1(a), 1(d), 1(e), and 2(a). Moreover, the log-likelihoods are concave in their arguments, and (E.105) and (E.106) have a closed form expressions given in (13)-(14) and (15)-(16), respectively. Last, notice that the true values of the parameters are fully identified by Assumption 6.

Therefore, by, e.g., Gourieroux and Monfort (1995, Property 7.11 p.182), $\widehat{\boldsymbol{\lambda}}_i^{(k+1)}$ and $\widehat{\sigma}_i^{2(k+1)}$, $i=1,\dots,n$, and $\widehat{\mathbf{A}}^{(k+1)}$ and $\widehat{\boldsymbol{\Gamma}}^{v(k+1)}$ exist and are unique for any $k\ge 0$. In particular, this result holds for $k=k^*$, i.e., for the EM estimators, and for any $n\in\mathbb{N}$ since (E.105) can be solved separately for any i. This completes the proof. □

Lemma E.21. *Under Assumptions 1, 2, 4, and 6,* $\ell(\boldsymbol{X}_{nT};\underline{\varphi}_n)$ *has a a local maximum denoted as* $\widehat{\varphi}_n^{**}=(\widehat{\boldsymbol{\lambda}}_1^{**\prime}\cdots\widehat{\boldsymbol{\lambda}}_n^{**\prime}\ \sigma_1^{2**}\cdots\sigma_n^{2**}\ \mathrm{vec}(\widehat{\mathbf{A}})^{**\prime},\mathrm{vech}(\widehat{\boldsymbol{\Gamma}})^{v**\prime})$*, such that*

(i) for all $i=1,\dots,n$, $\lim_{k\to\infty}\|\widehat{\boldsymbol{\lambda}}_i^{(k)}-\widehat{\boldsymbol{\lambda}}_i^{**}\|=0$;

(ii) for all $i=1,\dots,n$, $\lim_{k\to\infty}|\widehat{\sigma}_i^{2(k)}-\widehat{\sigma}_i^{2**}|=0$;

(iii) $\lim_{k\to\infty}\|\widehat{\mathbf{A}}^{(k)}-\widehat{\mathbf{A}}^{**}\|=0$;

(iv) $\lim_{k\to\infty}\|\widehat{\boldsymbol{\Gamma}}^{v(k)}-\widehat{\boldsymbol{\Gamma}}^{v**}\|=0$;

where all convergences are monotonic.

Proof. First notice that any maximum of $\ell(\boldsymbol{X}_{nT};\underline{\varphi}_n)$ is also a maximum of $\tilde{\ell}(\boldsymbol{X}_{nT};\underline{\varphi}_n)\equiv(nT)^{-1}\ell(\boldsymbol{X}_{nT};\underline{\varphi}_n)$. Now, the starting point of the EM is such that $\tilde{\ell}(\boldsymbol{X}_{nT};\widehat{\varphi}_n^{(0)})>-\infty$ and since $\tilde{\ell}(\boldsymbol{X}_{nT};\underline{\varphi}_n)$ is continuous and differentiable in the interior of $\mathcal{O}_n$, then $\{\tilde{\ell}(\boldsymbol{X}_{nT};\widehat{\varphi}_n^{(k)})\}_{k\ge0}$ is bounded from above for any $n,T\in\mathbb{N}$.

Consider the definitions:

$$\begin{aligned}\tilde{\ell}(\boldsymbol{X}_{nT};\underline{\varphi}_n)&=\mathbb{E}_{\widehat{\varphi}_n^{(k)}}[\tilde{\ell}(\boldsymbol{X}_{nT}|\boldsymbol{F}_T;\underline{\varphi}_n)+\tilde{\ell}(\boldsymbol{F}_T;\underline{\varphi}_n)|\boldsymbol{X}_{nT}]-\mathbb{E}_{\widehat{\varphi}_n^{(k)}}[\tilde{\ell}(\boldsymbol{F}_T|\boldsymbol{X}_{nT};\underline{\varphi}_n)|\boldsymbol{X}_{nT}]\\&=\tilde{\mathcal{Q}}(\underline{\varphi}_n,\widehat{\varphi}_n^{(k)})-\tilde{\mathcal{H}}(\underline{\varphi}_n,\widehat{\varphi}_n^{(k)}).\end{aligned}\tag{E.109}$$

Then, for any $\underline{\boldsymbol{\varphi}}_n$ and any n and T,

$$\begin{aligned}\tilde{\mathcal{H}}(\underline{\boldsymbol{\varphi}}_n;\widehat{\boldsymbol{\varphi}}_n^{(k)})-\tilde{\mathcal{H}}(\widehat{\boldsymbol{\varphi}}_n^{(k)};\widehat{\boldsymbol{\varphi}}_n^{(k)}) &= \mathbb{E}_{\widehat{\boldsymbol{\varphi}}_n^{(k)}}[\tilde{\ell}(\boldsymbol{F}_T|\boldsymbol{X}_{nT};\underline{\boldsymbol{\varphi}}_n)-\tilde{\ell}(\boldsymbol{F}_T|\boldsymbol{X}_{nT};\widehat{\boldsymbol{\varphi}}_n^{(k)})|\boldsymbol{X}_{nT}]\\
&=\frac{1}{nT}\mathbb{E}_{\widehat{\boldsymbol{\varphi}}_n^{(k)}}\left[\log\frac{f(\boldsymbol{F}_T|\boldsymbol{X}_{nT};\underline{\boldsymbol{\varphi}}_n)}{f(\boldsymbol{F}_T|\boldsymbol{X}_{nT};\widehat{\boldsymbol{\varphi}}_n^{(k)})}\right|\boldsymbol{X}_{nT}\Bigg]\\
&\le\frac{1}{nT}\log\mathbb{E}_{\widehat{\boldsymbol{\varphi}}_n^{(k)}}\left[\frac{f(\boldsymbol{F}_T|\boldsymbol{X}_{nT};\underline{\boldsymbol{\varphi}}_n)}{f(\boldsymbol{F}_T|\boldsymbol{X}_{nT};\widehat{\boldsymbol{\varphi}}_n^{(k)})}\right|\boldsymbol{X}_{nT}\Bigg]\\
&=\frac{1}{nT}\log\int_{\mathbb{R}^{rT}}\frac{f(\boldsymbol{F}_T|\boldsymbol{X}_{nT};\underline{\boldsymbol{\varphi}}_n)}{f(\boldsymbol{F}_T|\boldsymbol{X}_{nT};\widehat{\boldsymbol{\varphi}}_n^{(k)})}f(\boldsymbol{F}_T|\boldsymbol{X}_{nT};\widehat{\boldsymbol{\varphi}}_n^{(k)})\mathrm{d}\boldsymbol{F}_T\\
&=\frac{1}{nT}\log\int_{\mathbb{R}^{rT}}f(\boldsymbol{F}_T|\boldsymbol{X}_{nT};\underline{\boldsymbol{\varphi}}_n)\mathrm{d}\boldsymbol{F}_T=\frac{1}{nT}\log(1)=0,\end{aligned}$$

by Jensen's inequality. Hence, we have (see also Dempster et al., 1977, Lemma 1)

$$\tilde{\mathcal{H}}(\widehat{\boldsymbol{\varphi}}_n^{(k+1)};\widehat{\boldsymbol{\varphi}}_n^{(k)})\le\tilde{\mathcal{H}}(\widehat{\boldsymbol{\varphi}}_n^{(k)};\widehat{\boldsymbol{\varphi}}_n^{(k)}).\tag{E.110}$$

Therefore, from (E.109) and (E.110), for any k,

$$\tilde{\ell}(\boldsymbol{X}_{nT};\widehat{\boldsymbol{\varphi}}_n^{(k+1)})-\tilde{\ell}(\boldsymbol{X}_{nT};\widehat{\boldsymbol{\varphi}}_n^{(k)})\ge\tilde{\mathcal{Q}}(\widehat{\boldsymbol{\varphi}}_n^{(k+1)};\widehat{\boldsymbol{\varphi}}_n^{(k)})-\tilde{\mathcal{Q}}(\widehat{\boldsymbol{\varphi}}_n^{(k)};\widehat{\boldsymbol{\varphi}}_n^{(k)})\ge 0,\tag{E.111}$$

where the last inequality holds by definition of the M-step. This shows that the log-likelihood $\ell(\boldsymbol{X}_{nT};\widehat{\boldsymbol{\varphi}}_n^{(k)})$ increases monotonically as k increases.

Given that $\tilde{\mathcal{Q}}(\underline{\boldsymbol{\varphi}}_n;\widehat{\boldsymbol{\varphi}}_n^{(k)})$ has a unique maximum by Lemma E.20, and since for any $\underline{\boldsymbol{\varphi}}_n\in\mathcal{O}_n$ and $\overline{\boldsymbol{\varphi}}_n\in\mathcal{O}_n$, the function $\tilde{\mathcal{Q}}(\underline{\boldsymbol{\varphi}}_n;\overline{\boldsymbol{\varphi}}_n)$ is continuous in $\underline{\boldsymbol{\varphi}}_n$ and $\overline{\boldsymbol{\varphi}}_n$ and the components of the gradient vector $\nabla_{\underline{\boldsymbol{\varphi}}_n}\tilde{\mathcal{Q}}(\underline{\boldsymbol{\varphi}}_n;\overline{\boldsymbol{\varphi}}_n)$ are continuous in $\underline{\boldsymbol{\varphi}}_n$, from Wu (1983, Theorem 3)

$$\lim_{k\to\infty}\tilde{\ell}(\boldsymbol{X}_{nT};\widehat{\boldsymbol{\varphi}}_n^{(k)})=\tilde{\ell}(\boldsymbol{X}_{nT};\widehat{\boldsymbol{\varphi}}_n^{**}),\tag{E.112}$$

where the convergence is monotonic and $\widehat{\boldsymbol{\varphi}}_n^{**}$ is a local maximum of $\tilde{\ell}(\boldsymbol{X}_{nT};\underline{\boldsymbol{\varphi}}_n)$.

Now notice that the solution of the M-step is such that $\widehat{\boldsymbol{\lambda}}_i^{(k+1)}$ and $\widehat{\mathbf{A}}^{(k+1)}$ do not depend on other parameters at the same iteration, while $\widehat{\sigma}_i^{2(k+1)}$ depends only on $\widehat{\boldsymbol{\lambda}}_i^{(k+1)}$, and $\widehat{\boldsymbol{\Gamma}}^{v(k+1)}$ depends only on $\widehat{\mathbf{A}}^{(k+1)}$.

Since we are considering a Gaussian log-likelihood, from Wu (1983, Condition 1) holds, we have that, for any $i=1,\dots,n$,

$$\begin{aligned}&\tilde{\mathcal{Q}}(\widehat{\boldsymbol{\lambda}}_i^{(k+1)};\widehat{\boldsymbol{\varphi}}_n^{(k)})-\tilde{\mathcal{Q}}(\widehat{\boldsymbol{\lambda}}_i^{(k)};\widehat{\boldsymbol{\varphi}}_n^{(k)})\ge M_{\mathcal{Q}}\|\widehat{\boldsymbol{\lambda}}_i^{(k+1)}-\widehat{\boldsymbol{\lambda}}_i^{(k)}\|^2,\\
&\tilde{\mathcal{Q}}(\widehat{\boldsymbol{\lambda}}_i^{(k+1)},\widehat{\sigma}_i^{2(k+1)};\widehat{\boldsymbol{\varphi}}_n^{(k)})-\tilde{\mathcal{Q}}(\widehat{\boldsymbol{\lambda}}_i^{(k)},\widehat{\sigma}_i^{2(k)};\widehat{\boldsymbol{\varphi}}_n^{(k)})\ge M_{\mathcal{Q}'}\|(\widehat{\boldsymbol{\lambda}}_i^{(k+1)\prime}\ \widehat{\sigma}_i^{2(k+1)})-(\widehat{\boldsymbol{\lambda}}_i^{(k)\prime}\ \widehat{\sigma}_i^{2(k)})\|^2,\end{aligned}\tag{E.113}$$

for some finite positive reals $M_{\mathcal{Q}}$ and $M_{\mathcal{Q}'}$, independent of i, and where we use the short-hand notations:

$$\begin{aligned}&\tilde{\mathcal{Q}}(\widehat{\boldsymbol{\lambda}}_i^{(k)};\widehat{\boldsymbol{\varphi}}_n^{(k)})=\tilde{\mathcal{Q}}(\underline{\boldsymbol{\lambda}}_1,\dots,\widehat{\boldsymbol{\lambda}}_i^{(k)},\dots,\underline{\boldsymbol{\lambda}}_n,\underline{\sigma}_1^2,\dots,\underline{\sigma}_n^2,\underline{\mathbf{A}},\underline{\boldsymbol{\Gamma}}^v;\widehat{\boldsymbol{\varphi}}_n^{(k)}),\\
&\tilde{\mathcal{Q}}(\widehat{\boldsymbol{\lambda}}_i^{(k)},\widehat{\sigma}_i^{2(k)};\widehat{\boldsymbol{\varphi}}_n^{(k)})=\tilde{\mathcal{Q}}(\underline{\boldsymbol{\lambda}}_1,\dots,\widehat{\boldsymbol{\lambda}}_i^{(k)},\dots,\underline{\boldsymbol{\lambda}}_n,\underline{\sigma}_1^2,\dots,\widehat{\sigma}_i^{2(k)},\dots,\underline{\sigma}_n^2,\underline{\mathbf{A}},\underline{\boldsymbol{\Gamma}}^v;\widehat{\boldsymbol{\varphi}}_n^{(k)}).\end{aligned}$$

And, similarly,

$$\begin{aligned}&\tilde{\mathcal{Q}}(\widehat{\mathbf{A}}^{(k+1)};\widehat{\boldsymbol{\varphi}}_n^{(k)})-\tilde{\mathcal{Q}}(\widehat{\mathbf{A}}^{(k)};\widehat{\boldsymbol{\varphi}}_n^{(k)})\ge M_{A'}\|\widehat{\mathbf{A}}^{(k+1)}-\widehat{\mathbf{A}}^{(k)}\|^2,\\
&\tilde{\mathcal{Q}}(\widehat{\mathbf{A}}^{(k+1)},\widehat{\boldsymbol{\Gamma}}^{v(k+1)};\widehat{\boldsymbol{\varphi}}_n^{(k)})-\tilde{\mathcal{Q}}(\widehat{\mathbf{A}}^{(k)},\widehat{\boldsymbol{\Gamma}}^{v(k)};\widehat{\boldsymbol{\varphi}}_n^{(k)})\ge M_{v'}\|\mathrm{vec}(\widehat{\mathbf{A}}^{(k+1)}\ \widehat{\boldsymbol{\Gamma}}^{v(k+1)})-\mathrm{vec}(\widehat{\mathbf{A}}^{(k)};\widehat{\boldsymbol{\Gamma}}^{v(k)})\|^2,\end{aligned}\tag{E.114}$$

for some finite positive reals $M_{A'}$ and $M_{v'}$.

Therefore, from (E.111), (E.112), (E.113), and (E.114), we have

$$\begin{aligned}&\lim_{k\to\infty}\|(\widehat{\boldsymbol{\lambda}}_i^{(k+1)\prime}\ \widehat{\sigma}_i^{2(k+1)})-(\widehat{\boldsymbol{\lambda}}_i^{(k)\prime}\ \widehat{\sigma}_i^{2(k)})\|=0,\\
&\lim_{k\to\infty}\|\mathrm{vec}(\widehat{\mathbf{A}}^{(k+1)}\ \widehat{\boldsymbol{\Gamma}}^{v(k+1)})-\mathrm{vec}(\widehat{\mathbf{A}}^{(k)};\widehat{\boldsymbol{\Gamma}}^{v(k)})\|=0,\end{aligned}$$

which are sufficient conditions for applying the result in Wu (1983, Theorem 6), i.e., such that

$$\lim_{k\to\infty}\|(\widehat{\boldsymbol{\lambda}}_i^{(k)\prime}\ \widehat{\sigma}_i^{2(k)})-(\widehat{\boldsymbol{\lambda}}_i^{**\prime}\ \widehat{\sigma}_i^{2**})\|=0,$$
$$\lim_{k\to\infty}\|\mathrm{vec}(\widehat{\mathbf{A}}^{(k)}\ \widehat{\boldsymbol{\Gamma}}^{v(k)})-\mathrm{vec}(\widehat{\mathbf{A}}^{**}\ \widehat{\boldsymbol{\Gamma}}^{v**})\|=0.$$

This completes the proof. □

Lemma E.22. *Under Assumptions 1, 2, 3, 4, 5, and 6, as $n,T\to\infty$:*
(i) $\min(n\log^{-2/\delta_v},\sqrt{nT}\log^{-1/2}n,T)\max_{i=1,\dots,n}\|\widehat{\boldsymbol{\lambda}}_i^{**}-\widehat{\boldsymbol{\lambda}}_i^{*}\|=O_p(1)$;
(ii) $\min(n\log^{-2/\delta_v},\sqrt{nT}\log^{-1/2}n,T)\max_{i=1,\dots,n}|\widehat{\sigma}_i^{2**}-\widehat{\sigma}_i^{2*}|=O_p(1)$;
(iii) $n\log^{-2/\delta_v}\|\widehat{\mathbf{A}}^{**}-\widehat{\mathbf{A}}^{*}\|=O_p(1)$;
(iv) $n\log^{-2/\delta_v}\|\widehat{\boldsymbol{\Gamma}}^{v**}-\widehat{\boldsymbol{\Gamma}}^{v*}\|=O_p(1)$.

Proof. First, notice that running the EM algorithm using $\ell(\boldsymbol{X}_{nT};\underline{\boldsymbol{\varphi}}_n)$ is equivalent to running the EM using $\ell(\boldsymbol{X}_{nT},\boldsymbol{F}_T;\underline{\boldsymbol{\varphi}}_n)$, and such EM will converge to a local maximum of $\ell(\boldsymbol{X}_{nT},\boldsymbol{F}_T;\underline{\boldsymbol{\varphi}}_n)$ because Lemma E.21 would hold also in this case. Moreover, since $\ell(\boldsymbol{X}_{nT},\boldsymbol{F}_T;\underline{\boldsymbol{\varphi}}_n)$ has clearly a unique maximum, then $\widehat{\boldsymbol{\varphi}}_n^{**}$ is a local maximum of $\ell(\boldsymbol{X}_{nT};\underline{\boldsymbol{\varphi}}_n)$ but also the unique global maximum of $\ell(\boldsymbol{X}_{nT},\boldsymbol{F}_T;\underline{\boldsymbol{\varphi}}_n)$.

Consider the global QML estimator $\widehat{\boldsymbol{\varphi}}_n^{*}=(\widehat{\boldsymbol{\phi}}_n^{*\prime}\ \widehat{\boldsymbol{\theta}}^{*\prime})'$, where we defined $\widehat{\boldsymbol{\phi}}_n^{*}=(\mathrm{vec}(\widehat{\boldsymbol{\Lambda}}_n^{*})'\ \widehat{\sigma}_1^{2*}\cdots\widehat{\sigma}_n^{2*})'$ and $\widehat{\boldsymbol{\theta}}^{*}=(\mathrm{vec}(\widehat{\mathbf{A}}^{*})'\ \mathrm{vech}(\widehat{\boldsymbol{\Gamma}}^{v*})')$ maximizing $\ell(\boldsymbol{X}_{nT};\underline{\boldsymbol{\varphi}}_n)$. The elements of $\widehat{\boldsymbol{\varphi}}_n^{*}$ satisfy Lemma E.13.

Now, by definition the components of the gradient of $\ell(\boldsymbol{X}_{nT};\underline{\boldsymbol{\varphi}}_n)$ computed in $\widehat{\boldsymbol{\varphi}}_n^{*}$ are such that (notice that $\ell(\boldsymbol{F}_T;\underline{\boldsymbol{\varphi}}_n)$ does not depend on $\underline{\boldsymbol{\lambda}}_i$)

$$\begin{aligned}\mathbf{0}_r&=\nabla_{\underline{\boldsymbol{\lambda}}_i}\ell(\boldsymbol{X}_{nT};\underline{\boldsymbol{\varphi}}_n)|_{\underline{\boldsymbol{\varphi}}_n=\widehat{\boldsymbol{\varphi}}_n^*}\\&=\nabla_{\underline{\boldsymbol{\lambda}}_i}\ell(\boldsymbol{X}_{nT},\boldsymbol{F}_T;\underline{\boldsymbol{\varphi}}_n)|_{\underline{\boldsymbol{\varphi}}_n=\widehat{\boldsymbol{\varphi}}_n^*}-\nabla_{\underline{\boldsymbol{\lambda}}_i}\ell(\boldsymbol{F}_T|\boldsymbol{X}_{nT};\underline{\boldsymbol{\varphi}}_n)|_{\underline{\boldsymbol{\varphi}}_n=\widehat{\boldsymbol{\varphi}}_n^*}\\&=\sum_{t=1}^T\mathbf{F}_t(x_{it}-\mathbf{F}_t'\widehat{\boldsymbol{\lambda}}_i^*)-\nabla_{\underline{\boldsymbol{\lambda}}_i}\ell(\boldsymbol{F}_T|\boldsymbol{X}_{nT};\underline{\boldsymbol{\varphi}}_n)|_{\underline{\boldsymbol{\varphi}}_n=\widehat{\boldsymbol{\varphi}}_n^*}\\&=\sum_{t=1}^T\mathbf{F}_t(\mathbf{F}_t'\boldsymbol{\lambda}_i+\xi_{it}-\mathbf{F}_t'\widehat{\boldsymbol{\lambda}}_i^*)-\nabla_{\underline{\boldsymbol{\lambda}}_i}\ell(\boldsymbol{F}_T|\boldsymbol{X}_{nT};\underline{\boldsymbol{\varphi}}_n)|_{\underline{\boldsymbol{\varphi}}_n=\widehat{\boldsymbol{\varphi}}_n^*},\end{aligned}\tag{E.115}$$

and (E.115) is equivalent to

$$\begin{aligned}(\widehat{\boldsymbol{\lambda}}_i^*-\boldsymbol{\lambda}_i)&=\left(T^{-1}\sum_{t=1}^T\mathbf{F}_t\mathbf{F}_t'\right)^{-1}\left(T^{-1}\sum_{t=1}^T\mathbf{F}_t\xi_{it}\right)\\&\quad+\left(T^{-1}\sum_{t=1}^T\mathbf{F}_t\mathbf{F}_t'\right)^{-1}T^{-1}\nabla_{\underline{\boldsymbol{\lambda}}_i}\ell(\boldsymbol{F}_T|\boldsymbol{X}_{nT};\underline{\boldsymbol{\varphi}}_n)|_{\underline{\boldsymbol{\varphi}}_n=\widehat{\boldsymbol{\varphi}}_n^*}\\&=(\boldsymbol{\lambda}_i-\boldsymbol{\lambda}_i^{\mathrm{OLS}})+\left(T^{-1}\sum_{t=1}^T\mathbf{F}_t\mathbf{F}_t'\right)^{-1}T^{-1}\nabla_{\underline{\boldsymbol{\lambda}}_i}\ell(\boldsymbol{F}_T|\boldsymbol{X}_{nT};\underline{\boldsymbol{\varphi}}_n)|_{\underline{\boldsymbol{\varphi}}_n=\widehat{\boldsymbol{\varphi}}_n^*}.\end{aligned}\tag{E.116}$$

By comparing (E.116) with Lemma E.14(i) (see in particular (E.69) and (E.70) in its proof), and by Lemma C.13, we have

$$\max_{i=1,\dots,n}\left\|T^{-1}\nabla_{\underline{\boldsymbol{\lambda}}_i}\ell(\boldsymbol{F}_T|\boldsymbol{X}_{nT};\underline{\boldsymbol{\varphi}}_n)|_{\underline{\boldsymbol{\varphi}}_n=\widehat{\boldsymbol{\varphi}}_n^*}\right\|=O_p(\max(n^{-1}\log^{2/\delta_v}T,n^{-1/2}T^{-1/2}\sqrt{\log n},T^{-1})),$$

which, from (E.115), implies

$$\begin{aligned}\max_{i=1,\dots,n}\left\|T^{-1}\nabla_{\underline{\boldsymbol{\lambda}}_i}\ell(\boldsymbol{X}_{nT},\boldsymbol{F}_T;\underline{\boldsymbol{\varphi}}_n)|_{\underline{\boldsymbol{\varphi}}_n=\widehat{\boldsymbol{\varphi}}_n^*}\right\|&=\max_{i=1,\dots,n}\left\|T^{-1}\nabla_{\underline{\boldsymbol{\lambda}}_i}\ell(\boldsymbol{F}_T|\boldsymbol{X}_{nT};\underline{\boldsymbol{\varphi}}_n)|_{\underline{\boldsymbol{\varphi}}_n=\widehat{\boldsymbol{\varphi}}_n^*}\right\|\\&=O_p(\max(n^{-1}\log^{2/\delta_v}T,n^{-1/2}T^{-1/2}\sqrt{\log n},T^{-1})).\end{aligned}\tag{E.117}$$

Then, since $\nabla_{\underline{\boldsymbol{\lambda}}_i}\ell(\boldsymbol{X}_{nT},\boldsymbol{F}_T;\underline{\boldsymbol{\varphi}}_n)$ is linear in $\underline{\boldsymbol{\lambda}}_i$, there exists a positive real M_S' such that

$$\|T^{-1}\nabla_{\underline{\boldsymbol{\lambda}}_i}\ell(\boldsymbol{X}_{nT},\boldsymbol{F}_T;\underline{\boldsymbol{\varphi}}_n)|_{\underline{\boldsymbol{\varphi}}_n=\widehat{\boldsymbol{\varphi}}_n^*}\|\geq M_S'\|\widehat{\boldsymbol{\lambda}}_i^*-\widehat{\boldsymbol{\lambda}}_i^{**}\|,\tag{E.118}$$

since, by definition, $T^{-1}\nabla_{\underline{\boldsymbol{\lambda}}_i}\ell(\boldsymbol{X}_{nT},\boldsymbol{F}_T;\underline{\boldsymbol{\varphi}}_n)|_{\underline{\boldsymbol{\varphi}}_n=\widehat{\boldsymbol{\varphi}}_n^{**}}=\mathbf{0}_r$. Therefore, from (E.117) and (E.118)

$$\max_{i=1,\dots,n}\|\widehat{\boldsymbol{\lambda}}_i^*-\widehat{\boldsymbol{\lambda}}_i^{**}\|=O_p(\max(n^{-1}\log^{2/\delta_v}T,n^{-1/2}T^{-1/2}\sqrt{\log n},T^{-1})),$$

which proves part (i).

Similarly, using Lemma (E.14)(ii) (see in particular (E.72), (E.73), and (E.77) in its proof), we can show that

$$\max_{i=1,\dots,n}|T^{-1}\nabla_{\underline{\sigma}_i^2}\ell(\boldsymbol{X}_{nT},\boldsymbol{F}_T;\underline{\boldsymbol{\varphi}}_n)|_{\underline{\boldsymbol{\varphi}}_n=\widehat{\boldsymbol{\varphi}}_n^*}|=O_p(\max(n^{-1}\log^{2/\delta_v}T,n^{-1/2}T^{-1/2}\sqrt{\log n},T^{-1})), \tag{E.119}$$

and

$$\max_{i=1,\dots,n}|T^{-1}\nabla_{\underline{\sigma}_i^2}\ell(\boldsymbol{X}_{nT},\boldsymbol{F}_T;\underline{\boldsymbol{\varphi}}_n)|_{\underline{\boldsymbol{\varphi}}_n=\widehat{\boldsymbol{\varphi}}_n^*}|\geq M_G'|\widehat{\sigma}_i^{2**}-\widehat{\sigma}_i^{2*}|. \tag{E.120}$$

By using (E.119) into (E.120), we prove part (ii).

Parts (iii) and (iv) follow in a similar way from Lemma E.13(iv) and E.13(v), respectively. This completes the proof. □

Lemma E.23. *Consider the EM estimator* $\widehat{\boldsymbol{\lambda}}_i\equiv\widehat{\boldsymbol{\lambda}}_i^{(k+1)}$*, for any* $k\geq 0$*. Under Assumptions 1, 2, 3, 4, 5, and 6, as* $n,T\to\infty$*,*

$$\min(n^2\log^{-4/\delta_v}T,nT\log^{-1/\delta_v}T\log^{-1/2}n,T^{3/2}\log^{-1/2}n)\,\|\widehat{\boldsymbol{\lambda}}_i-\widehat{\boldsymbol{\lambda}}_i^{**}\|=O_p(1),$$

uniformly in i*.*

Proof. First, notice that for any $\underline{\boldsymbol{\varphi}}_n$ and any n and T

$$\begin{aligned}(nT)^{-1}\nabla_{\underline{\boldsymbol{\varphi}}_n}\ell(\boldsymbol{X}_{nT};\underline{\boldsymbol{\varphi}}_n)&=(nT)^{-1}\frac{\nabla_{\underline{\boldsymbol{\varphi}}_n}f(\boldsymbol{X}_{nT};\underline{\boldsymbol{\varphi}}_n)}{f(\boldsymbol{X}_{nT};\underline{\boldsymbol{\varphi}}_n)}=(nT)^{-1}\int_{\mathbb{R}^{rT}}\frac{\nabla_{\underline{\boldsymbol{\varphi}}_n}f(\boldsymbol{X}_{nT},\boldsymbol{F}_T;\underline{\boldsymbol{\varphi}}_n)}{f(\boldsymbol{X}_{nT};\underline{\boldsymbol{\varphi}}_n)}\mathrm{d}\boldsymbol{F}_T\\&=(nT)^{-1}\int_{\mathbb{R}^{rT}}\frac{\nabla_{\underline{\boldsymbol{\varphi}}_n}f(\boldsymbol{X}_{nT},\boldsymbol{F}_T;\underline{\boldsymbol{\varphi}}_n)}{f(\boldsymbol{X}_{nT},\boldsymbol{F}_T;\underline{\boldsymbol{\varphi}}_n)}\frac{f(\boldsymbol{X}_{nT},\boldsymbol{F}_T;\underline{\boldsymbol{\varphi}}_n)}{f(\boldsymbol{X}_{nT};\underline{\boldsymbol{\varphi}}_n)}\mathrm{d}\boldsymbol{F}_T\\&=(nT)^{-1}\int_{\mathbb{R}^{rT}}\nabla_{\underline{\boldsymbol{\varphi}}_n}\ell(\boldsymbol{X}_{nT},\boldsymbol{F}_T;\underline{\boldsymbol{\varphi}}_n)\frac{f(\boldsymbol{X}_{nT},\boldsymbol{F}_T;\underline{\boldsymbol{\varphi}}_n)}{f(\boldsymbol{X}_{nT};\underline{\boldsymbol{\varphi}}_n)}\mathrm{d}\boldsymbol{F}_T\\&=(nT)^{-1}\int_{\mathbb{R}^{rT}}\nabla_{\underline{\boldsymbol{\varphi}}_n}\ell(\boldsymbol{X}_{nT},\boldsymbol{F}_T;\underline{\boldsymbol{\varphi}}_n)f(\boldsymbol{F}_T|\boldsymbol{X}_{nT};\underline{\boldsymbol{\varphi}}_n)\mathrm{d}\boldsymbol{F}_T\\&=(nT)^{-1}\mathbb{E}_{\underline{\boldsymbol{\varphi}}_n}[\nabla_{\underline{\boldsymbol{\varphi}}_n}\ell(\boldsymbol{X}_{nT},\boldsymbol{F}_T;\underline{\boldsymbol{\varphi}}_n)|\boldsymbol{X}_{nT}]\\&=(nT)^{-1}\nabla_{\underline{\boldsymbol{\varphi}}_n}\mathcal{Q}(\underline{\boldsymbol{\varphi}}_n';\underline{\boldsymbol{\varphi}}_n)|_{\underline{\boldsymbol{\varphi}}_n'=\underline{\boldsymbol{\varphi}}_n}.\end{aligned} \tag{E.121}$$

Second, for any $k\geq 0$, by definition of the M-step, by (E.121) and by a Taylor expansion about $\widehat{\boldsymbol{\lambda}}_i^{(k)}$,

$$\begin{aligned}\mathbf{0}_r&=\nabla_{\underline{\boldsymbol{\lambda}}_i}\mathcal{Q}(\underline{\boldsymbol{\varphi}}_n;\widehat{\boldsymbol{\varphi}}_n^{(k)})|_{\underline{\boldsymbol{\varphi}}_n=\widehat{\boldsymbol{\varphi}}_n^{(k+1)}}\\&=\nabla_{\underline{\boldsymbol{\lambda}}_i}\mathcal{Q}(\underline{\boldsymbol{\varphi}}_n;\widehat{\boldsymbol{\varphi}}_n^{(k)})|_{\underline{\boldsymbol{\varphi}}_n=\widehat{\boldsymbol{\varphi}}_n^{(k)}}+\nabla_{\underline{\boldsymbol{\lambda}}_i\underline{\boldsymbol{\lambda}}_i'}\mathcal{Q}(\underline{\boldsymbol{\varphi}}_n;\widehat{\boldsymbol{\varphi}}_n^{(k)})|_{\underline{\boldsymbol{\varphi}}_n=\widehat{\boldsymbol{\varphi}}_n^{(k)}}(\widehat{\boldsymbol{\lambda}}_i^{(k+1)}-\widehat{\boldsymbol{\lambda}}_i^{(k)})+O(\|\widehat{\boldsymbol{\lambda}}_i^{(k+1)}-\widehat{\boldsymbol{\lambda}}_i^{(k)}\|^2)\\&=\nabla_{\underline{\boldsymbol{\lambda}}_i}\ell(\boldsymbol{X}_{nT};\underline{\boldsymbol{\varphi}}_n)|_{\underline{\boldsymbol{\varphi}}_n=\widehat{\boldsymbol{\varphi}}_n^{(k)}}+\nabla_{\underline{\boldsymbol{\lambda}}_i\underline{\boldsymbol{\lambda}}_i'}\mathcal{Q}(\underline{\boldsymbol{\varphi}}_n;\widehat{\boldsymbol{\varphi}}_n^{(k)})|_{\underline{\boldsymbol{\varphi}}_n=\widehat{\boldsymbol{\varphi}}_n^{(k)}}(\widehat{\boldsymbol{\lambda}}_i^{(k+1)}-\widehat{\boldsymbol{\lambda}}_i^{(k)})+O(\|\widehat{\boldsymbol{\lambda}}_i^{(k+1)}-\widehat{\boldsymbol{\lambda}}_i^{(k)}\|^2)\\&=\nabla_{\underline{\boldsymbol{\lambda}}_i}\ell(\boldsymbol{X}_{nT};\underline{\boldsymbol{\varphi}}_n)|_{\underline{\boldsymbol{\varphi}}_n=\widehat{\boldsymbol{\varphi}}_n^{(k)}}+\nabla_{\underline{\boldsymbol{\lambda}}_i\underline{\boldsymbol{\lambda}}_i'}\mathcal{Q}(\underline{\boldsymbol{\varphi}}_n;\widehat{\boldsymbol{\varphi}}_n^{(k)})|_{\underline{\boldsymbol{\varphi}}_n=\widehat{\boldsymbol{\varphi}}_n^{(k)}}(\widehat{\boldsymbol{\lambda}}_i^{(k+1)}-\widehat{\boldsymbol{\lambda}}_i^{(k)}),\end{aligned} \tag{E.122}$$

since the third derivative of $\mathcal{Q}(\underline{\boldsymbol{\varphi}}_n;\widehat{\boldsymbol{\varphi}}_n^{(k)})$ with respect to $\underline{\boldsymbol{\lambda}}_i$ is zero since the second derivative does not depend on $\underline{\boldsymbol{\lambda}}_i$ because $\mathcal{Q}(\underline{\boldsymbol{\varphi}}_n;\widehat{\boldsymbol{\varphi}}_n^{(k)})$ is quadratic in $\underline{\boldsymbol{\lambda}}_i$. And from (E.122) it follows that

$$T^{-1}\nabla_{\underline{\boldsymbol{\lambda}}_i}\ell(\boldsymbol{X}_{nT};\underline{\boldsymbol{\varphi}}_n)|_{\underline{\boldsymbol{\varphi}}_n=\widehat{\boldsymbol{\varphi}}_n^{(k)}}=-T^{-1}\nabla_{\underline{\boldsymbol{\lambda}}_i\underline{\boldsymbol{\lambda}}_i'}\mathcal{Q}(\underline{\boldsymbol{\varphi}}_n;\widehat{\boldsymbol{\varphi}}_n^{(k)})|_{\underline{\boldsymbol{\varphi}}_n=\widehat{\boldsymbol{\varphi}}_n^{(k)}}(\widehat{\boldsymbol{\lambda}}_i^{(k+1)}-\widehat{\boldsymbol{\lambda}}_i^{(k)}). \tag{E.123}$$

Third, by a Taylor expansion about $\widehat{\boldsymbol{\lambda}}_i^{(k)}$, by definition of $\widehat{\boldsymbol{\lambda}}_i^{**}$ which is a local maximum of $\ell(\boldsymbol{X}_{nT};\underline{\boldsymbol{\varphi}}_n)$, we have

$$\begin{aligned}
\mathbf{0}_r &= T^{-1}\nabla_{\underline{\boldsymbol{\lambda}}_i}\ell(\boldsymbol{X}_{nT};\underline{\boldsymbol{\varphi}}_n)|_{\underline{\boldsymbol{\varphi}}_n=\widehat{\boldsymbol{\varphi}}_n^{**}}\\
&= T^{-1}\nabla_{\underline{\boldsymbol{\lambda}}_i}\ell(\boldsymbol{X}_{nT};\underline{\boldsymbol{\varphi}}_n)|_{\underline{\boldsymbol{\varphi}}_n=\widehat{\boldsymbol{\varphi}}_n^{(k)}} + T^{-1}\nabla_{\underline{\boldsymbol{\lambda}}_i\underline{\boldsymbol{\lambda}}_i'}\ell(\boldsymbol{X}_{nT};\underline{\boldsymbol{\varphi}}_n)|_{\underline{\boldsymbol{\varphi}}_n=\widehat{\boldsymbol{\varphi}}_n^{(k)}}(\widehat{\boldsymbol{\lambda}}_i^{**}-\widehat{\boldsymbol{\lambda}}_i^{(k)}) \qquad \text{(E.124)}\\
&\quad + \frac{1}{2}((\widehat{\boldsymbol{\lambda}}_i^{**}-\widehat{\boldsymbol{\lambda}}_i^{(k)})'\otimes\mathbf{I}_r)\left\{T^{-1}\nabla_{\underline{\boldsymbol{\lambda}}_i}\text{vec}(\nabla_{\underline{\boldsymbol{\lambda}}_i\underline{\boldsymbol{\lambda}}_i'}\ell(\boldsymbol{X}_{nT};\underline{\boldsymbol{\varphi}}_n))|_{\underline{\boldsymbol{\varphi}}_n=\check{\boldsymbol{\varphi}}_n}\right\}(\widehat{\boldsymbol{\lambda}}_i^{**}-\widehat{\boldsymbol{\lambda}}_i^{(k)})\\
&= T^{-1}\nabla_{\underline{\boldsymbol{\lambda}}_i}\ell(\boldsymbol{X}_{nT};\underline{\boldsymbol{\varphi}}_n)|_{\underline{\boldsymbol{\varphi}}_n=\widehat{\boldsymbol{\varphi}}_n^{(k)}} + T^{-1}\nabla_{\underline{\boldsymbol{\lambda}}_i\underline{\boldsymbol{\lambda}}_i'}\ell(\boldsymbol{X}_{nT};\underline{\boldsymbol{\varphi}}_n)|_{\underline{\boldsymbol{\varphi}}_n=\widehat{\boldsymbol{\varphi}}_n^{(k)}}(\widehat{\boldsymbol{\lambda}}_i^{**}-\widehat{\boldsymbol{\lambda}}_i^{(k)})\\
&\quad + O_p\left(\|\widehat{\boldsymbol{\lambda}}_i^{**}-\widehat{\boldsymbol{\lambda}}_i^{(k)}\|^2 n^{-1}\log^{2/\delta_v}T\right),
\end{aligned}$$

where $\check{\boldsymbol{\varphi}}_n$ is such that $n^{-1/2}\|\widehat{\boldsymbol{\varphi}}_n^{**}-\check{\boldsymbol{\varphi}}_n\|\le n^{-1/2}\|\widehat{\boldsymbol{\varphi}}_n^{**}-\widehat{\boldsymbol{\varphi}}_n^{(k+1)}\|$ and $n^{-1/2}\|\widehat{\boldsymbol{\varphi}}_n^{(k+1)}-\check{\boldsymbol{\varphi}}_n\|\le n^{-1/2}\|\widehat{\boldsymbol{\varphi}}_n^{**}-\widehat{\boldsymbol{\varphi}}_n^{(k+1)}\|$. In particular, the last term in (E.124) follows from Lemma E.9 and Barigozzi (2023, eq. (A.58) in the proof of Theorem 5), which imply

$$\begin{aligned}
&\sup_{\underline{\boldsymbol{\varphi}}_n\in\mathcal{O}_n}|\nabla_{\underline{\boldsymbol{\lambda}}_i}\text{vec}(\nabla_{\underline{\boldsymbol{\lambda}}_i\underline{\boldsymbol{\lambda}}_i'}\ell(\boldsymbol{X}_{nT};\underline{\boldsymbol{\varphi}}_n))|\\
&\quad\le \sup_{\underline{\boldsymbol{\varphi}}_n\in\mathcal{O}_n}|\nabla_{\underline{\boldsymbol{\lambda}}_i}\text{vec}(\nabla_{\underline{\boldsymbol{\lambda}}_i\underline{\boldsymbol{\lambda}}_i'}\ell(\boldsymbol{X}_{nT};\underline{\boldsymbol{\varphi}}_n))-\nabla_{\underline{\boldsymbol{\lambda}}_i}\text{vec}(\nabla_{\underline{\boldsymbol{\lambda}}_i\underline{\boldsymbol{\lambda}}_i'}\ell_0(\boldsymbol{X}_{nT};\underline{\boldsymbol{\phi}}_n))|\\
&\qquad + \sup_{\underline{\boldsymbol{\varphi}}_n\in\mathcal{O}_n}|\nabla_{\underline{\boldsymbol{\lambda}}_i}\text{vec}(\nabla_{\underline{\boldsymbol{\lambda}}_i\underline{\boldsymbol{\lambda}}_i'}\ell_0(\boldsymbol{X}_{nT};\underline{\boldsymbol{\phi}}_n))|\\
&\quad= O_p(n^{-1}T\log^{2/\delta_v}T)+O(n^{-1}T)=O_p(n^{-1}T\log^{2/\delta_v}T). \qquad \text{(E.125)}
\end{aligned}$$

Define the following $r\times r$ matrices:

$$\begin{aligned}
\boldsymbol{\mathcal{I}}_{\rm c}(\underline{\boldsymbol{\lambda}}_i) &= -\nabla_{\underline{\boldsymbol{\lambda}}_i\underline{\boldsymbol{\lambda}}_i'}\mathcal{Q}(\underline{\boldsymbol{\varphi}}_n;\underline{\boldsymbol{\varphi}}_n) = -\mathbb{E}_{\underline{\boldsymbol{\varphi}}_n}[\nabla_{\underline{\boldsymbol{\lambda}}_i\underline{\boldsymbol{\lambda}}_i'}\ell(\boldsymbol{X}_{nT},\boldsymbol{F}_T;\underline{\boldsymbol{\varphi}}_n)|\boldsymbol{X}_{nT}]\\
\boldsymbol{I}(\underline{\boldsymbol{\lambda}}_i) &= -\nabla_{\underline{\boldsymbol{\lambda}}_i\underline{\boldsymbol{\lambda}}_i'}\ell(\boldsymbol{X}_{nT};\underline{\boldsymbol{\varphi}}_n).
\end{aligned}$$

By substituting (E.123) into (E.124) and rearranging

$$\begin{aligned}
(\widehat{\boldsymbol{\lambda}}_i^{**}-\widehat{\boldsymbol{\lambda}}_i^{(k)}) &= -\left\{\nabla_{\underline{\boldsymbol{\lambda}}_i\underline{\boldsymbol{\lambda}}_i'}\ell(\boldsymbol{X}_{nT};\underline{\boldsymbol{\varphi}}_n)|_{\underline{\boldsymbol{\varphi}}_n=\widehat{\boldsymbol{\varphi}}_n^{(k)}}\right\}^{-1}\left\{\nabla_{\underline{\boldsymbol{\lambda}}_i}\ell(\boldsymbol{X}_{nT};\underline{\boldsymbol{\varphi}}_n)|_{\underline{\boldsymbol{\varphi}}_n=\widehat{\boldsymbol{\varphi}}_n^{(k)}}\right\}\\
&\quad + O_p\left(\|\widehat{\boldsymbol{\lambda}}_i^{**}-\widehat{\boldsymbol{\lambda}}_i^{(k)}\|^2n^{-1}\log^{2/\delta_v}T\right)\\
&= \left\{\nabla_{\underline{\boldsymbol{\lambda}}_i\underline{\boldsymbol{\lambda}}_i'}\ell(\boldsymbol{X}_{nT};\underline{\boldsymbol{\varphi}}_n)|_{\underline{\boldsymbol{\varphi}}_n=\widehat{\boldsymbol{\varphi}}_n^{(k)}}\right\}^{-1}\left\{\nabla_{\underline{\boldsymbol{\lambda}}_i\underline{\boldsymbol{\lambda}}_i'}\mathcal{Q}(\underline{\boldsymbol{\varphi}}_n;\widehat{\boldsymbol{\varphi}}_n^{(k)})|_{\underline{\boldsymbol{\varphi}}_n=\widehat{\boldsymbol{\varphi}}_n^{(k)}}\right\}(\widehat{\boldsymbol{\lambda}}_i^{(k+1)}-\widehat{\boldsymbol{\lambda}}_i^{(k)})\\
&\quad + O_p\left(\|\widehat{\boldsymbol{\lambda}}_i^{**}-\widehat{\boldsymbol{\lambda}}_i^{(k)}\|^2n^{-1}\log^{2/\delta_v}T\right)\\
&= \left\{\boldsymbol{I}(\widehat{\boldsymbol{\lambda}}_i^{(k)})\right\}^{-1}\left\{\boldsymbol{\mathcal{I}}_{\rm c}(\widehat{\boldsymbol{\lambda}}_i^{(k)})\right\}(\widehat{\boldsymbol{\lambda}}_i^{(k+1)}-\widehat{\boldsymbol{\lambda}}_i^{**}+\widehat{\boldsymbol{\lambda}}_i^{**}-\widehat{\boldsymbol{\lambda}}_i^{(k)})\\
&\quad + O_p\left(\|\widehat{\boldsymbol{\lambda}}_i^{**}-\widehat{\boldsymbol{\lambda}}_i^{(k)}\|^2n^{-1}\log^{2/\delta_v}T\right). \qquad \text{(E.126)}
\end{aligned}$$

Moreover, by a Taylor approximation about $\widehat{\boldsymbol{\lambda}}_i^{**}$,

$$\begin{aligned}
&T^{-1}\boldsymbol{\mathcal{I}}_{\rm c}(\widehat{\boldsymbol{\lambda}}_i^{(k)}) = T^{-1}\boldsymbol{\mathcal{I}}_{\rm c}(\widehat{\boldsymbol{\lambda}}_i^{**}),\\
&T^{-1}\boldsymbol{I}(\widehat{\boldsymbol{\lambda}}_i^{(k)}) = T^{-1}\boldsymbol{I}(\widehat{\boldsymbol{\lambda}}_i^{**}) + O_p\left(\|\widehat{\boldsymbol{\lambda}}_i^{**}-\widehat{\boldsymbol{\lambda}}_i^{(k)}\|n^{-1}\log^{2/\delta_v}T\right), \qquad \text{(E.127)}
\end{aligned}$$

where the first relation follows from the fact that, as noticed above, $\boldsymbol{\mathcal{I}}_{\rm c}(\underline{\boldsymbol{\lambda}}_i)$ does not depend on $\underline{\boldsymbol{\lambda}}_i$, while the second follows again from (E.125).

Therefore, from (E.126) and (E.127)

$$
\begin{aligned}
(\widehat{\boldsymbol{\lambda}}_i^{(k+1)}-\widehat{\boldsymbol{\lambda}}_i^{**}) &= \left(\mathbf{I}_r-\left\{\boldsymbol{\mathcal{I}}_{\rm c}(\widehat{\boldsymbol{\lambda}}_i^{(k)})\right\}^{-1}\left\{\boldsymbol{I}(\widehat{\boldsymbol{\lambda}}_i^{(k)})\right\}\right)(\widehat{\boldsymbol{\lambda}}_i^{(k)}-\widehat{\boldsymbol{\lambda}}_i^{**})\\
&\quad+O_p\left(\|\widehat{\boldsymbol{\lambda}}_i^{**}-\widehat{\boldsymbol{\lambda}}_i^{(k)}\|^2 n^{-1}\log^{2/\delta_v}T\right)\\
&=\left(\mathbf{I}_r-\left\{\boldsymbol{\mathcal{I}}_{\rm c}(\widehat{\boldsymbol{\lambda}}_i^{**})\right\}^{-1}\left\{\boldsymbol{I}(\widehat{\boldsymbol{\lambda}}_i^{**})\right\}\right)(\widehat{\boldsymbol{\lambda}}_i^{(k)}-\widehat{\boldsymbol{\lambda}}_i^{**})\\
&\quad+O_p\left(\|\widehat{\boldsymbol{\lambda}}_i^{**}-\widehat{\boldsymbol{\lambda}}_i^{(k)}\| n^{-1}\log^{2/\delta_v}T\right),
\end{aligned}\tag{E.128}
$$

see also Sundberg (1974, 1976), Dempster et al. (1977), Meng and Rubin (1994), McLachlan and Krishnan (2007, Chapter 3.9, pp. 99-103), and Sundberg (2019, Chapter 8).

Let $\boldsymbol{R}(\widehat{\boldsymbol{\lambda}}_i^{**})=\mathbf{I}_r-\{\boldsymbol{\mathcal{I}}_{\rm c}(\widehat{\boldsymbol{\lambda}}_i^{**})\}^{-1}\{\boldsymbol{I}(\widehat{\boldsymbol{\lambda}}_i^{**})\}$, then, by setting $k=k^*$ in (E.128), since $\widehat{\boldsymbol{\lambda}}_i\equiv\widehat{\boldsymbol{\lambda}}_i^{(k^*+1)}$, we have

$$
(\widehat{\boldsymbol{\lambda}}_i-\widehat{\boldsymbol{\lambda}}_i^{**})=\boldsymbol{R}(\widehat{\boldsymbol{\lambda}}_i^{**})(\widehat{\boldsymbol{\lambda}}_i^{(k^*)}-\widehat{\boldsymbol{\lambda}}_i^{**})+O_p\left(\|\widehat{\boldsymbol{\lambda}}_i^{**}-\widehat{\boldsymbol{\lambda}}_i^{(k)}\|n^{-1}\log^{2/\delta_v}T\right).\tag{E.129}
$$

Hence, from (E.129), by iterating backwards we get

$$
\begin{aligned}
\|\widehat{\boldsymbol{\lambda}}_i-\widehat{\boldsymbol{\lambda}}_i^{**}\| &\le \|\widehat{\boldsymbol{\lambda}}_i^{(k^*)}-\widehat{\boldsymbol{\lambda}}_i^{**}\|\ \|\boldsymbol{R}(\widehat{\boldsymbol{\lambda}}_i^{**})\|+O_p\left(\|\widehat{\boldsymbol{\lambda}}_i^{(k^*)}-\widehat{\boldsymbol{\lambda}}_i^{**}\|n^{-1}\log^{2/\delta_v}T\right)\\
&\le\left\{\|\widehat{\boldsymbol{\lambda}}_i^{(k^*-1)}-\widehat{\boldsymbol{\lambda}}_i^{**}\|\ \|\boldsymbol{R}(\widehat{\boldsymbol{\lambda}}_i^{**})\|+O_p\left(\|\widehat{\boldsymbol{\lambda}}_i^{(k^*)}-\widehat{\boldsymbol{\lambda}}_i^{**}\|n^{-1}\log^{2/\delta_v}T\right)\right\}\ \|\boldsymbol{R}(\widehat{\boldsymbol{\lambda}}_i^{**})\|\\
&\quad+O_p\left(\|\widehat{\boldsymbol{\lambda}}_i^{(k^*)}-\widehat{\boldsymbol{\lambda}}_i^{**}\|n^{-1}\log^{2/\delta_v}T\right)\\
&\le\|\widehat{\boldsymbol{\lambda}}_i^{(0)}-\widehat{\boldsymbol{\lambda}}_i^{**}\|\ \|\boldsymbol{R}(\widehat{\boldsymbol{\lambda}}_i^{**})\|^{k^*+1}+\left\{\sum_{j=0}^{k^*}\|\boldsymbol{R}(\widehat{\boldsymbol{\lambda}}_i^{**})\|^j\right\}O_p\left(\|\widehat{\boldsymbol{\lambda}}_i^{(k^*)}-\widehat{\boldsymbol{\lambda}}_i^{**}\|n^{-1}\log^{2/\delta_v}T\right)\\
&\le\|\widehat{\boldsymbol{\lambda}}_i^{(0)}-\widehat{\boldsymbol{\lambda}}_i^{**}\|\ \|\boldsymbol{R}(\widehat{\boldsymbol{\lambda}}_i^{**})\|^{k^*+1}+\left\{\sum_{j=0}^{k^*}\|\boldsymbol{R}(\widehat{\boldsymbol{\lambda}}_i^{**})\|^j\right\}O_p\left(\|\widehat{\boldsymbol{\lambda}}_i^{(0)}-\widehat{\boldsymbol{\lambda}}_i^{**}\|n^{-1}\log^{2/\delta_v}T\right),
\end{aligned}\tag{E.130}
$$

where in the second and third step we used Lemma E.21, according to which $\|\widehat{\boldsymbol{\lambda}}_i^{(k+1)}-\widehat{\boldsymbol{\lambda}}_i^{**}\|\le\|\widehat{\boldsymbol{\lambda}}_i^{(k)}-\widehat{\boldsymbol{\lambda}}_i^{**}\|$ for any $k\ge 0$.

Let us consider separately the terms on the rhs of (E.130). Because of Lemma E.22(i),

$$
\max_{i=1,\ldots,n}\|\boldsymbol{R}(\widehat{\boldsymbol{\lambda}}_i^{**})-\boldsymbol{R}(\widehat{\boldsymbol{\lambda}}_i^{*})\|=O_p(\max(n^{-1}\log^{2/\delta_v}T,n^{-1/2}T^{-1/2}\sqrt{\log n},T^{-1})),\tag{E.131}
$$

since $\boldsymbol{R}(\underline{\boldsymbol{\lambda}}_i)$ is continuous and differentiable in $\underline{\boldsymbol{\lambda}}_i$. Consider the two matrices in $\boldsymbol{R}(\widehat{\boldsymbol{\lambda}}_i^{*})$. First, from (9) we can easily see that

$$
\begin{aligned}
T^{-1}\boldsymbol{\mathcal{I}}_{\rm c}(\widehat{\boldsymbol{\lambda}}_i^{*}) &= -T^{-1}\mathbb{E}_{\widehat{\varphi}_n^*}[\nabla_{\underline{\boldsymbol{\lambda}}_i\underline{\boldsymbol{\lambda}}_i'}\ell(\boldsymbol{X}_{nT},\boldsymbol{F}_T;\underline{\boldsymbol{\varphi}}_n)|_{\underline{\boldsymbol{\varphi}}_n=\widehat{\boldsymbol{\varphi}}_n^*}|\boldsymbol{X}_{nT}]\\
&= -T^{-1}\mathbb{E}_{\widehat{\varphi}_n^*}[\nabla_{\underline{\boldsymbol{\lambda}}_i\underline{\boldsymbol{\lambda}}_i'}\ell(\boldsymbol{X}_{nT}|\boldsymbol{F}_T;\underline{\boldsymbol{\varphi}}_n)|_{\underline{\boldsymbol{\varphi}}_n=\widehat{\boldsymbol{\varphi}}_n^*}|\boldsymbol{X}_{nT}]\\
&=T^{-1}(\widehat{\sigma}_i^{2*})^{-1}\sum_{t=1}^T\mathbb{E}_{\widehat{\varphi}_n^*}[\mathbf{F}_t\mathbf{F}_t'|\boldsymbol{X}_{nT}]\\
&=T^{-1}(\widehat{\sigma}_i^{2*})^{-1}\sum_{t=1}^T\mathbb{E}_{\widehat{\varphi}_n^*}[\mathbf{F}_t|\boldsymbol{X}_{nT}]\mathbb{E}_{\widehat{\varphi}_n^*}[\mathbf{F}_t'|\boldsymbol{X}_{nT}]\\
&\quad+T^{-1}(\widehat{\sigma}_i^{2*})^{-1}\sum_{t=1}^T\mathbb{E}_{\widehat{\varphi}_n^*}[(\mathbf{F}_t-\mathbb{E}_{\widehat{\varphi}_n^*}[\mathbf{F}_t|\boldsymbol{X}_{nT}])(\mathbf{F}_t-\mathbb{E}_{\widehat{\varphi}_n^*}[\mathbf{F}_t|\boldsymbol{X}_{nT}])'|\boldsymbol{X}_{nT}]\\
&=T^{-1}(\widehat{\sigma}_i^{2*})^{-1}\sum_{t=1}^T\left\{\mathbf{F}_{t|T}^*\mathbf{F}_{t|T}^{*\prime}+\mathbf{P}_{t|T}^*\right\},
\end{aligned}\tag{E.132}
$$

where in the last step we used Assumption 4 which implies $\mathbf{F}_{t|T}^*=\mathbb{E}_{\widehat{\varphi}_n^*}[\mathbf{F}_t|\boldsymbol{X}_{nT}]$ for all $t=1,\ldots,T$. Also note that for (E.110) and (E.111) in the proof of Lemma E.21 to hold expectations have to be taken with respect to the same distribution as the one used to compute the log-likelihood. Hence, given that we maximize a mis-specified log-likelihood with diagonal idiosyncratic covariance, in the last step of (E.132) we use $\mathbf{F}_{t|T}^*$ and $\mathbf{P}_{t|T}^*$, i.e., which are the outputs of the Kalman smoother implemented using $\widehat{\boldsymbol{\Sigma}}_n^{\xi*}$.

Moreover, by Lemma E.19

$$\left\|T^{-1}\sum_{t=1}^T\{\mathbf{F}^*_{t|T}\mathbf{F}^{*\prime}_{t|T}+\mathbf{P}^*_{t|T}-\mathbf{F}_t\mathbf{F}_t'\}\right\|=O_p(\max(n^{-1}\log^{2/\delta_v}T,n^{-1/2}T^{-1/2},T^{-1}\sqrt{\log n})),$$

which, once substituted into (E.132), gives

$$\max_{i=1,\dots,n}T^{-1}\left\|\boldsymbol{\mathcal{I}}_{\rm c}(\widehat{\boldsymbol{\lambda}}_i^*)-(\widehat{\sigma}_i^{2*})^{-1}\sum_{t=1}^T\mathbf{F}_t\mathbf{F}_t'\right\|$$
$$=O_p(\max(n^{-1}\log^{2/\delta_v}T,n^{-1/2}T^{-1/2},T^{-1}\sqrt{\log n})). \tag{E.133}$$

This also shows that $T^{-1}\boldsymbol{\mathcal{I}}_{\rm c}(\widehat{\boldsymbol{\lambda}}_i^*)$ is finite and positive definite, as $n,T\to\infty$, since $T^{-1}\sum_{t=1}^T\mathbf{F}_t\mathbf{F}_t'$ is finite and positive definite with probability tending to one as $n,T\to\infty$, because of Lemma C.12(i) combined with Assumption 6(b), and since $\widehat{\sigma}_i^{2*}$ is finite a positive for all $i=1,\dots,n$ because of Lemma E.14(ii) and Assumption 2(a).

Second, from Lemma E.9 it follows that

$$\max_{i=1,\dots,n}T^{-1}\left\|\nabla_{\underline{\boldsymbol{\lambda}}_i\underline{\boldsymbol{\lambda}}_i'}\ell(\boldsymbol{X}_{nT};\underline{\boldsymbol{\varphi}}_n)|_{\underline{\boldsymbol{\varphi}}_n=\widehat{\boldsymbol{\varphi}}_n^*}-\nabla_{\underline{\boldsymbol{\lambda}}_i\underline{\boldsymbol{\lambda}}_i'}\ell_0(\boldsymbol{X}_{nT};\underline{\boldsymbol{\phi}}_n)|_{\underline{\boldsymbol{\phi}}_n=\widehat{\boldsymbol{\phi}}_n^*}\right\|$$
$$=\max_{i=1,\dots,n}T^{-1}\left\|\boldsymbol{I}(\widehat{\boldsymbol{\lambda}}_i^*)-\nabla_{\underline{\boldsymbol{\lambda}}_i\underline{\boldsymbol{\lambda}}_i'}\ell_0(\boldsymbol{X}_{nT};\underline{\boldsymbol{\phi}}_n)|_{\underline{\boldsymbol{\phi}}_n=\widehat{\boldsymbol{\phi}}_n^*}\right\|$$
$$=O_p(n^{-1}\log^{2/\delta_v}T). \tag{E.134}$$

Furthermore, by a Taylor expansion about $\widehat{\boldsymbol{\lambda}}_i^\dagger$

$$\max_{i=1,\dots,n}T^{-1}\left\|\nabla_{\underline{\boldsymbol{\lambda}}_i\underline{\boldsymbol{\lambda}}_i'}\ell_0(\boldsymbol{X}_{nT};\underline{\boldsymbol{\phi}}_n)|_{\underline{\boldsymbol{\phi}}_n=\widehat{\boldsymbol{\phi}}_n^*}-\nabla_{\underline{\boldsymbol{\lambda}}_i\underline{\boldsymbol{\lambda}}_i'}\ell_0(\boldsymbol{X}_{nT};\underline{\boldsymbol{\phi}}_n)|_{\underline{\boldsymbol{\phi}}_n=\widehat{\boldsymbol{\phi}}_n^\dagger}\right\|$$
$$\le\max_{i=1,\dots,n}T^{-1}\left\|\nabla_{\underline{\boldsymbol{\lambda}}_i}\mathrm{vec}\left(\nabla_{\underline{\boldsymbol{\lambda}}_i\underline{\boldsymbol{\lambda}}_i'}\ell_0(\boldsymbol{X}_{nT};\underline{\boldsymbol{\phi}}_n)\right)|_{\underline{\boldsymbol{\phi}}_n=\widehat{\boldsymbol{\phi}}_n^\dagger}\right\|\|\widehat{\boldsymbol{\lambda}}_i^\dagger-\widehat{\boldsymbol{\lambda}}_i^*\|+O_p(\max_{i=1,\dots,n}\|\widehat{\boldsymbol{\lambda}}_i^\dagger-\widehat{\boldsymbol{\lambda}}_i^*\|^2)$$
$$=O_p(n^{-2}\log^{4/\delta_v}T), \tag{E.135}$$

by Lemma E.10(i) and since $\|\nabla_{\underline{\boldsymbol{\lambda}}_i}\mathrm{vec}(\nabla_{\underline{\boldsymbol{\lambda}}_i\underline{\boldsymbol{\lambda}}_i'}\ell_0(\boldsymbol{X}_{nT};\underline{\boldsymbol{\phi}}_n))\|=O_p(n^{-1}T\log^{2/\delta_v}T)$ for all $\underline{\boldsymbol{\phi}}_n$, by (E.125).

Third, from Barigozzi (2023, Theorem 5):

$$\max_{i=1,\dots,n}T^{-1}\left\|\nabla_{\underline{\boldsymbol{\lambda}}_i\underline{\boldsymbol{\lambda}}_i'}\ell_0(\boldsymbol{X}_{nT};\underline{\boldsymbol{\phi}}_n)|_{\underline{\boldsymbol{\phi}}_n=\widehat{\boldsymbol{\phi}}_n^\dagger}-\nabla_{\underline{\boldsymbol{\lambda}}_i\underline{\boldsymbol{\lambda}}_i'}\ell_0(\boldsymbol{X}_{nT}|\boldsymbol{F}_T;\underline{\boldsymbol{\phi}}_n)|_{\underline{\boldsymbol{\phi}}_n=\widehat{\boldsymbol{\phi}}_n^\dagger}\right\|$$
$$=\max_{i=1,\dots,n}T^{-1}\left\|\nabla_{\underline{\boldsymbol{\lambda}}_i\underline{\boldsymbol{\lambda}}_i'}\ell_0(\boldsymbol{X}_{nT};\underline{\boldsymbol{\phi}}_n)|_{\underline{\boldsymbol{\phi}}_n=\widehat{\boldsymbol{\phi}}_n^\dagger}-\left(-(\widehat{\sigma}_i^{2\dagger})^{-1}\sum_{t=1}^T\mathbf{F}_t\mathbf{F}_t'\right)\right\|$$
$$=O_p(\max(n^{-1},n^{-1/2}T^{-1/2})). \tag{E.136}$$

Last, from Lemma E.10(iii):

$$\max_{i=1,\dots,n}|\widehat{\sigma}_i^{2\dagger}-\widehat{\sigma}_i^{2*}|=O_p(n^{-1}\log^{2/\delta_v}T). \tag{E.137}$$

By combining (E.134), (E.135), (E.136), and (E.137), we have

$$\max_{i=1,\dots,n}T^{-1}\left\|\boldsymbol{I}(\widehat{\boldsymbol{\lambda}}_i^*)-(\widehat{\sigma}_i^{2*})^{-1}\sum_{t=1}^T\mathbf{F}_t\mathbf{F}_t'\right\|=O_p(\max(n^{-1}\log^{2/\delta_v}T,n^{-1/2}T^{-1/2}). \tag{E.138}$$

This is also shows that $T^{-1}\boldsymbol{I}(\widehat{\boldsymbol{\lambda}}_i^*)$ is finite and positive definite, as $n,T\to\infty$, since $T^{-1}\sum_{t=1}^T\mathbf{F}_t\mathbf{F}_t'$ is finite and positive definite with probability tending to one as $n,T\to\infty$, because of Lemma C.12(i) combined with Assumption 6(b), and since $\widehat{\sigma}_i^{2*}$ is finite a positive for all $i=1,\dots,n$ because of Lemma E.14(ii) and Assumption 2(a).

Therefore, by using (E.131), (E.133), and (E.138), and since, as remarked above $T^{-1}\boldsymbol{\mathcal{I}}_{\rm c}(\widehat{\boldsymbol{\lambda}}_i^*)$ is finite and positive definite, we have

$$
\begin{aligned}
\max_{i=1,\ldots,n} \|\boldsymbol{R}(\widehat{\boldsymbol{\lambda}}_i^{**})\| &= \|\mathbf{I}_r - \{\boldsymbol{\mathcal{I}}_{\mathrm{c}}(\widehat{\boldsymbol{\lambda}}_i^{*})\}^{-1}\{\boldsymbol{I}(\widehat{\boldsymbol{\lambda}}_i^{*})\}\| \\
&= \max_{i=1,\ldots,n} \|\{\boldsymbol{\mathcal{I}}_{\mathrm{c}}(\widehat{\boldsymbol{\lambda}}_i^{*})^{-1}\}\{\boldsymbol{\mathcal{I}}_{\mathrm{c}}(\widehat{\boldsymbol{\lambda}}_i^{*}) - \boldsymbol{I}(\widehat{\boldsymbol{\lambda}}_i^{*})\}\| \\
&\le \max_{i=1,\ldots,n} T\|\{\boldsymbol{\mathcal{I}}_{\mathrm{c}}(\widehat{\boldsymbol{\lambda}}_i^{*})^{-1}\}\| \, T^{-1}\|\boldsymbol{\mathcal{I}}_{\mathrm{c}}(\widehat{\boldsymbol{\lambda}}_i^{*}) - \boldsymbol{I}(\widehat{\boldsymbol{\lambda}}_i^{*})\| \\
&\le \max_{i=1,\ldots,n} T\|\{\boldsymbol{\mathcal{I}}_{\mathrm{c}}(\widehat{\boldsymbol{\lambda}}_i^{*})^{-1}\}\| \, T^{-1} \left\|\boldsymbol{\mathcal{I}}_{\mathrm{c}}(\widehat{\boldsymbol{\lambda}}_i^{*}) - (\widehat{\sigma}_i^{2*})^{-1}\sum_{t=1}^T \mathbf{F}_t\mathbf{F}_t'\right\| \\
&\quad + \max_{i=1,\ldots,n} T\|\{\boldsymbol{\mathcal{I}}_{\mathrm{c}}(\widehat{\boldsymbol{\lambda}}_i^{*})^{-1}\}\| \left\|(\widehat{\sigma}_i^{2*})^{-1}\sum_{t=1}^T \mathbf{F}_t\mathbf{F}_t' - \boldsymbol{I}(\widehat{\boldsymbol{\lambda}}_i^{*})\right\| \\
&= O_p(\max(n^{-1}\log^{2/\delta_v} T, n^{-1/2}T^{-1/2}\sqrt{\log n}, T^{-1}\sqrt{\log n}).
\end{aligned} \tag{E.139}
$$

Moreover,

$$
\begin{aligned}
\|\widehat{\boldsymbol{\lambda}}_i^{(0)} - \widehat{\boldsymbol{\lambda}}_i^{**}\| &\le \|\widehat{\boldsymbol{\lambda}}_i^{(0)} - \widehat{\boldsymbol{\lambda}}_i^{*}\| + \|\widehat{\boldsymbol{\lambda}}_i^{*} - \widehat{\boldsymbol{\lambda}}_i^{**}\| \\
&\le \|\widehat{\boldsymbol{\lambda}}_i^{(0)} - \boldsymbol{\lambda}_i\| + \|\boldsymbol{\lambda}_i - \widehat{\boldsymbol{\lambda}}_i^{*}\| + \|\widehat{\boldsymbol{\lambda}}_i^{*} - \widehat{\boldsymbol{\lambda}}_i^{**}\| \\
&= O_p(\max(n^{-1}, T^{-1/2}) + O_p(\max(n^{-1}\log^{2/\delta_v} T, T^{-1/2})) \\
&\quad + O_p(\max(n^{-1}\log^{2/\delta_v} T, n^{-1/2}T^{-1/2}\sqrt{\log n}, T^{-1})) \\
&= O_p(\max(n^{-1}\log^{2/\delta_v} T, T^{-1/2}),
\end{aligned} \tag{E.140}
$$

because of Lemmas D.1(i), E.13(i), and E.22(i), respectively.

By substituting (E.139) and (E.140) into (E.130), we have

$$
\begin{aligned}
\|\widehat{\boldsymbol{\lambda}}_i - \widehat{\boldsymbol{\lambda}}_i^{**}\| &\le \|\widehat{\boldsymbol{\lambda}}_i^{(0)} - \widehat{\boldsymbol{\lambda}}_i^{**}\| \, \|\boldsymbol{R}(\widehat{\boldsymbol{\lambda}}_i^{**})\|^{k^*+1} + o_p(\max(n^{-1}\log^{2/\delta_v} T, T^{-1/2})) \\
&= O_p(\max(n^{-1}\log^{2/\delta_v} T, T^{-1/2})) \\
&\quad \cdot \left\{O_p(\max(n^{-1}\log^{2/\delta_v} T, n^{-1/2}T^{-1/2}\sqrt{\log n}, T^{-1}\sqrt{\log n})\right\}^{k^*+1} \\
&\quad + o_p(\max(n^{-1}\log^{2/\delta_v} T, T^{-1/2})) \\
&= O_p(\max(n^{-2}\log^{4/\delta_v} T, n^{-1}T^{-1}\log^{1/\delta_v} T\sqrt{\log n}, T^{-3/2}\sqrt{\log n})),
\end{aligned} \tag{E.141}
$$

uniformly in i since the rhs of (E.139) does not depend on i. This completes the proof. □

Lemma E.24. *Consider the EM estimators* $\widehat{\sigma}_i^2 \equiv \widehat{\sigma}_i^{2(k+1)}$, $\widehat{\mathbf{A}} \equiv \widehat{\mathbf{A}}^{(k+1)}$, *and* $\widehat{\boldsymbol{\Gamma}}^v \equiv \widehat{\boldsymbol{\Gamma}}^{v(k+1)}$, *for any* $k \ge 0$. *Under Assumptions 1, 2, 3, 4, 5, and 6, as* $n, T \to \infty$:

(i) $\min(n\log^{-2/\delta_v} T, \sqrt{T})\,|\widehat{\sigma}_i^2 - \widehat{\sigma}_i^{2**}| = O_p(1)$;
(ii) $\min(n\log^{-2/\delta_v} T, \sqrt{T})\,\|\widehat{\mathbf{A}} - \widehat{\mathbf{A}}^{**}\| = O_p(1)$;
(iii) $\min(n\log^{-2/\delta_v} T, \sqrt{T})\,\|\widehat{\boldsymbol{\Gamma}}^v - \widehat{\boldsymbol{\Gamma}}^{v**}\| = O_p(1)$.

Proof. Let,

$$
R(\widehat{\sigma}_i^{2**}) = 1 - \left\{\mathbb{E}_{\underline{\boldsymbol{\varphi}}_n}[\nabla_{\underline{\sigma}_i^2\underline{\sigma}_i^2}\ell(\boldsymbol{X}_{nT}, \boldsymbol{F}_T; \underline{\boldsymbol{\varphi}}_n)|_{\underline{\sigma}_i^2=\widehat{\sigma}_i^{2**}}|\boldsymbol{X}_{nT}]\right\}^{-1}\left\{\nabla_{\underline{\sigma}_i^2}\ell(\boldsymbol{X}_{nT}; \underline{\boldsymbol{\varphi}}_n)|_{\underline{\sigma}_i^2=\widehat{\sigma}_i^{2**}}\right\}.
$$

Then, following the same steps leading to (E.130) in the proof of Lemma E.23, we obtain

$$
\begin{aligned}
|\widehat{\sigma}_i^2 - \widehat{\sigma}_i^{2**}| &\le |\widehat{\sigma}_i^{2(0)} - \widehat{\sigma}_i^{2**}| \, |R(\widehat{\sigma}_i^{2**})|^{k^*+1} + C|\widehat{\sigma}_i^{2(0)} - \widehat{\sigma}_i^{2**}| \\
&\le \left\{|\widehat{\sigma}_i^{2(0)} - \sigma_i^2| + |\sigma_i^2 - \widehat{\sigma}_i^{2*}| + |\widehat{\sigma}_i^{2*} - \widehat{\sigma}_i^{2**}|\right\}|R(\widehat{\sigma}_i^{2**})|^{k^*+1} + C|\widehat{\sigma}_i^{2(0)} - \widehat{\sigma}_i^{2**}| \\
&= O_p(\max(n^{-1}, T^{-1/2}) + O_p(\max(n^{-1}\log^{2/\delta_v} T, T^{-1/2}) \\
&\quad + O_p(\max(n^{-1}\log^{2/\delta_v} T, n^{-1/2}T^{-1/2}\sqrt{\log n}, T^{-1})) \\
&= O_p(\max(n^{-1}\log^{2/\delta_v} T, T^{-1/2})),
\end{aligned} \tag{E.142}
$$

for some finite positive real C independent of i, by Lemmas D.4(i), E.13(iii), and E.22(ii), and since $|R(\widehat{\sigma}_i^{2**})| = O_p(1)$.

This proves part (i).

For part (ii), following again the same steps leading to (E.130) in the proof of Lemma E.23, we obtain

$$
\begin{aligned}
\|\widehat{\mathbf{A}} - \widehat{\mathbf{A}}^{**}\| &\le \|\widehat{\mathbf{A}}^{(0)} - \widehat{\mathbf{A}}^{**}\| \, \|\boldsymbol{R}(\widehat{\mathbf{A}}^{**})\|^{k^*+1} + C\|\widehat{\mathbf{A}}^{(0)} - \widehat{\mathbf{A}}^{**}\| \\
&\le \left\{\|\widehat{\mathbf{A}}^{(0)} - \mathbf{A}\| + \|\mathbf{A} - \widehat{\mathbf{A}}^{*}\| + \|\widehat{\mathbf{A}}^{*} - \widehat{\mathbf{A}}^{**}\|\right\} \\
&\quad \cdot \|\boldsymbol{R}(\widehat{\mathbf{A}}^{**})\|^{k^*+1} + C\|\widehat{\mathbf{A}}^{(0)} - \widehat{\mathbf{A}}^{**}\| \\
&= O_p(n^{-1}, T^{-1/2}) + O_p(\min(n^{-1}\log^{2/\delta_v} T, T^{-1/2})) + O_p(n^{-1}\log^{2/\delta_v} T) \\
&= O_p(\max(n^{-1}\log^{2/\delta_v} T, T^{-1/2})),
\end{aligned}
$$

for some finite positive real C, by Lemmas D.3(i), E.13(iv), E.22(iii), and since $\|\boldsymbol{R}(\widehat{\mathbf{A}}^{**})\| = O_p(1)$. This proves part (ii).

Part (iii) is proved in the same way as part (ii) but using Lemmas D.3(ii), E.13(v), and E.22(iv). This completes the proof. □

Lemma E.25. *Consider the EM estimators $\widehat{\boldsymbol{\lambda}}_i \equiv \widehat{\boldsymbol{\lambda}}_i^{(k+1)}$ and $\widehat{\sigma}_i^2 \equiv \widehat{\sigma}_i^{2(k+1)}$, for any $k \ge 0$. Under Assumptions 1, 2, 3, 4, 5, and 6, as $n, T \to \infty$:*

(i) $\min(n^2 \log^{-4/\delta_v} T, nT\log^{-1/\delta_v} T \log^{-1/2} n, T^{3/2}\log^{-1} n) \max_{i=1,\dots,n} \|\widehat{\boldsymbol{\lambda}}_i - \widehat{\boldsymbol{\lambda}}_i^{**}\| = O_p(1)$;

(ii) $\min(n\log^{-2/\delta_v} T, \sqrt{T}\log^{-1/2} n) \max_{i=1,\dots,n} |\widehat{\sigma}_i^2 - \widehat{\sigma}_i^{2**}| = O_p(1)$.

Proof. First notice that from (E.71) in the proof of Lemma E.14

$$
\max_{i=1,\dots,n} \|\widehat{\boldsymbol{\lambda}}_i^{(0)} - \boldsymbol{\lambda}_i\| = O_p(\max(n^{-1}, T^{-1/2}\sqrt{\log n})), \tag{E.143}
$$

while from (D.6) in the proof of Lemma D.4 it is clear that

$$
\max_{i=1,\dots,n} |\widehat{\sigma}_i^{2(0)} - \sigma_i^2| = O_p(\max(n^{-1}, T^{-1/2}\sqrt{\log n})), \tag{E.144}
$$

because of (E.143).

Then,

$$
\begin{aligned}
\max_{i=1,\dots,n} \|\widehat{\boldsymbol{\lambda}}_i^{(0)} - \widehat{\boldsymbol{\lambda}}_i^{**}\| &\le \max_{i=1,\dots,n} \|\widehat{\boldsymbol{\lambda}}_i^{(0)} - \boldsymbol{\lambda}_i\| + \max_{i=1,\dots,n} \|\boldsymbol{\lambda}_i - \widehat{\boldsymbol{\lambda}}_i^{*}\| + \max_{i=1,\dots,n} \|\widehat{\boldsymbol{\lambda}}_i^{*} - \widehat{\boldsymbol{\lambda}}_i^{**}\| \\
&= O_p(\max(n^{-1}, T^{-1/2}\sqrt{\log n}) + O_p(\max(n^{-1}\log^{2/\delta_v} T, T^{-1/2}\sqrt{\log n})) \\
&\quad + O_p(\max(n^{-1}\log^{2/\delta_v} T, n^{-1/2}T^{-1/2}\sqrt{\log n}, T^{-1})) \\
&= O_p(\max(n^{-1}\log^{2/\delta_v} T, T^{-1/2}\sqrt{\log n}),
\end{aligned} \tag{E.145}
$$

because of (E.143) and Lemmas E.14(i), E.22(i).

Then, from (E.141) in the proof of Lemma E.23

$$
\begin{aligned}
\max_{i=1,\dots,n} \|\widehat{\boldsymbol{\lambda}}_i - \widehat{\boldsymbol{\lambda}}_i^{**}\| &\le \max_{i=1,\dots,n} \|\widehat{\boldsymbol{\lambda}}_i^{(0)} - \widehat{\boldsymbol{\lambda}}_i^{**}\| \max_{i=1,\dots,n} \|\boldsymbol{R}(\widehat{\boldsymbol{\lambda}}_i^{**})\|^{k^*+1} \\
&\quad + o_p(\max(n^{-1}\log^{2/\delta_v} T, T^{-1/2}\sqrt{\log n})) \\
&= O_p(\max(n^{-1}\log^{2/\delta_v} T, T^{-1/2}\sqrt{\log n})) \\
&\quad \cdot \left\{O_p(\max(n^{-1}\log^{2/\delta_v} T, n^{-1/2}T^{-1/2}\sqrt{\log n}, T^{-1}\sqrt{\log n})\right\}^{k^*+1} \\
&\quad + o_p(\max(n^{-1}\log^{2/\delta_v} T, T^{-1/2}\sqrt{\log n}) \\
&= O_p(\max(n^{-2}\log^{4/\delta_v} T, n^{-1}T^{-1}\log^{1/\delta_v} T\sqrt{\log n}, T^{-3/2}\log n)),
\end{aligned} \tag{E.146}
$$

because of (E.145) and (E.139) in the proof of Lemma E.23. This proves part (i).

For part (ii), from (E.142) in the proof of Lemma E.24

$$\begin{aligned}\max_{i=1,\dots,n}|\widehat{\sigma}_i^2-\widehat{\sigma}_i^{2**}| \le & \left\{\max_{i=1,\dots,n}|\widehat{\sigma}_i^{2(0)}-\sigma_i^2|+\max_{i=1,\dots,n}|\sigma_i^2-\widehat{\sigma}_i^{2*}|+\max_{i=1,\dots,n}|\widehat{\sigma}_i^{2*}-\widehat{\sigma}_i^{2**}|\right\}\\ &\cdot\max_{i=1,\dots,n}|R(\widehat{\sigma}_i^{2**})|^{k^*+1}+C\max_{i=1,\dots,n}|\widehat{\sigma}_i^{2(0)}-\widehat{\sigma}_i^{2**}|\\ =&\, O_p(\max(n^{-1},T^{-1/2}\sqrt{\log n})+O_p(\max(n^{-1}\log^{2/\delta_v}T,T^{-1/2}\sqrt{\log n})\\ &+O_p(\max(n^{-1}\log^{2/\delta_v}T,n^{-1/2}T^{-1/2}\sqrt{\log n},T^{-1}))\\ =&\, O_p(\max(n^{-1}\log^{2/\delta_v}T,T^{-1/2}\sqrt{\log n})),\end{aligned}$$

because of (E.144), and Lemmas E.14(ii), E.22(ii). This proves part (ii) and completes the proof. □

Lemma E.26. *Consider the EM algorithm initialized with any deterministic loadings* $\check{\boldsymbol{\Lambda}}_n^{(0)}=(\check{\boldsymbol{\lambda}}_1^{(0)}\cdots\check{\boldsymbol{\lambda}}_n^{(0)})'$ *such that* $\mathrm{vec}(\check{\boldsymbol{\Lambda}}_n^{(0)})\in\{\mathcal{O}^n_{\lambda_i}\cap\mathcal{E}_{\Lambda_n}\}$ *as defined in Section 4.3.4. Then, under Assumptions 1, 2, 3, 4, 5, and 6, as* $n,T\to\infty$,

(i) $\min(n\log^{-2/\delta_v}T,\sqrt{nT}\log^{-1/2}n,T\log^{-1/2}n)\,\|\widehat{\boldsymbol{\lambda}}_i-\widehat{\boldsymbol{\lambda}}_i^{**}\|=O_p(1)$, *uniformly in* i.

(ii) $\min(n\log^{-2/\delta_v}T,\sqrt{nT}\log^{-1/2}n,T\log^{-1/2}n)\,\max_{i=1,\dots,n}\|\widehat{\boldsymbol{\lambda}}_i-\widehat{\boldsymbol{\lambda}}_i^{**}\|=O_p(1)$.

Proof. From (E.130) in the proof of Lemma E.23,

$$\begin{aligned}\|\widehat{\boldsymbol{\lambda}}_i-\widehat{\boldsymbol{\lambda}}_i^{**}\|\le&\|\check{\boldsymbol{\lambda}}_i^{(0)}-\widehat{\boldsymbol{\lambda}}_i^{**}\|\,\|\boldsymbol{R}(\widehat{\boldsymbol{\lambda}}_i^{**})\|^{k^*+1}+\left\{\sum_{j=0}^{k^*}\|\boldsymbol{R}(\widehat{\boldsymbol{\lambda}}_i^{**})\|^j\right\}O_p\left(\|\check{\boldsymbol{\lambda}}_i^{(0)}-\widehat{\boldsymbol{\lambda}}_i^{**}\|n^{-1}\log^{2/\delta_v}T\right)\\ \le&\|\check{\boldsymbol{\lambda}}_i^{(0)}-\widehat{\boldsymbol{\lambda}}_i^{**}\|\,\|\boldsymbol{R}(\widehat{\boldsymbol{\lambda}}_i^{**})\|^{k^*+1}+O_p\left(\|\check{\boldsymbol{\lambda}}_i^{(0)}-\widehat{\boldsymbol{\lambda}}_i^{**}\|n^{-1}\log^{2/\delta_v}T\right)\\ =&\left\{O_p(\max(n^{-1}\log^{2/\delta_v}T,n^{-1/2}T^{-1/2}\sqrt{\log n},T^{-1}\sqrt{\log n})\right\}^{k^*+1}+O_p(n^{-1}\log^{2/\delta_v}T)\\ =&\,O_p(\max(n^{-1}\log^{2/\delta_v}T,n^{-1/2}T^{-1/2}\sqrt{\log n},T^{-1}\sqrt{\log n}),\end{aligned}$$

by (E.139) in the proof of Lemma E.23, and since $\|\check{\boldsymbol{\lambda}}_i^{(0)}-\widehat{\boldsymbol{\lambda}}_i^{**}\|\le\|\check{\boldsymbol{\lambda}}_i^{(0)}\|+\|\widehat{\boldsymbol{\lambda}}_i^{**}-\widehat{\boldsymbol{\lambda}}_i^{*}\|+\|\widehat{\boldsymbol{\lambda}}_i^{*}-\boldsymbol{\lambda}_i\|\le M_\lambda O(1)$, by definition of the initial estimator, Assumption 1(a), and Lemmas E.13(i) and E.22(i). This proves part (i).

Part (ii) is proved in the same way but starting from (E.146) in the proof of Lemma E.25 and by noting that $\max_{i=1,\dots,n}\|\check{\boldsymbol{\lambda}}_i^{(0)}\|\le M_\lambda$ by the same arguments as before. This completes the proof. □

F Lemmas necessary for proving Proposition 3

Lemma F.1. *Consider the EM estimators* $\widehat{\boldsymbol{\lambda}}_i\equiv\widehat{\boldsymbol{\lambda}}_i^{(k+1)}$, $\widehat{\sigma}_i^2\equiv\widehat{\sigma}_i^{2(k+1)}$, *and* $\widehat{\boldsymbol{\Sigma}}_n^\xi\equiv\widehat{\boldsymbol{\Sigma}}_n^{\xi(k+1)}$, *for any* $k\ge0$. *Under Assumptions 1, 2, 3, 4, 5, and 6, as* $n,T\to\infty$,

(i) $\min(\sqrt{T}\log^{-1/2}n,n\log^{-2/\delta_v}T)\max_{i=1,\dots,n}\|\widehat{\boldsymbol{\lambda}}_i-\boldsymbol{\lambda}_i\|=O_p(1)$;

(ii) $\min(\sqrt{T}\log^{-1/2}n,n\log^{-2/\delta_v}T)\max_{i=1,\dots,n}|\widehat{\sigma}_i^2-\sigma_i^2|=O_p(1)$;

(iii) $\min(\sqrt{T}\log^{-1/2}n,n\log^{-2/\delta_v}T)\,\|\widehat{\boldsymbol{\Sigma}}_n^\xi-\boldsymbol{\Sigma}_n^\xi\|=O_p(1)$;

(iv) $\|(\widehat{\boldsymbol{\Sigma}}_n^\xi)^{-1}\|=O_p(1)$;

(v) $\min(\sqrt{T}\log^{-1/2}n,n\log^{-2/\delta_v}T)\,\|(\widehat{\boldsymbol{\Sigma}}_n^\xi)^{-1}-(\boldsymbol{\Sigma}_n^\xi)^{-1}\|=O_p(1)$.

Proof. For part (i)

$$\begin{aligned}\max_{i=1,\dots,n}\|\widehat{\boldsymbol{\lambda}}_i-\boldsymbol{\lambda}_i\|\le&\max_{i=1,\dots,n}\|\widehat{\boldsymbol{\lambda}}_i-\widehat{\boldsymbol{\lambda}}_i^{**}\|+\max_{i=1,\dots,n}\|\widehat{\boldsymbol{\lambda}}_i^{**}-\widehat{\boldsymbol{\lambda}}_i^{*}\|+\max_{i=1,\dots,n}\|\widehat{\boldsymbol{\lambda}}_i^{*}-\boldsymbol{\lambda}_i\|\\ =&\,O_p(\max(n^{-2}\log^{4/\delta_v}T,n^{-1}T^{-1}\log^{1/\delta_v}T\sqrt{\log n},T^{-3/2}\log n))\\ &+O_p(\max(n^{-1}\log^{2/\delta_v}T,n^{-1/2}T^{-1/2}\sqrt{\log n},T^{-1}))\\ &+O_p(\max(n^{-1},T^{-1/2}\sqrt{\log n}))+O_p(n^{-1}\log^{2/\delta_v}T)\\ =&\,O_p(\max(n^{-1}\log^{2/\delta_v}T,T^{-1/2}\sqrt{\log n})),\end{aligned}$$

by Lemmas E.14(i), E.22(i), and E.25(i). This proves part (i).

Part (ii) is proved as part (i) but using Lemmas E.14(ii), E.22(ii), and E.25(ii).

Part (iii) immediately follows from part (ii), indeed

$$\|\widehat{\boldsymbol{\Sigma}}_n^\xi - \boldsymbol{\Sigma}_n^\xi\| \le \max_{i=1,\dots,n} |\widehat{\sigma}_i^2 - \sigma_i^2| = O_p(\max(n^{-1}\log^{2/\delta_v} T, T^{-1/2}\sqrt{\log n})).$$

For part (iv) we have

$$\|(\widehat{\boldsymbol{\Sigma}}_n^\xi)^{-1}\| = \left\{\min_{i=1,\dots,n} \widehat{\sigma}_i^2\right\}^{-1} \le C_\xi + O_p(\max(n^{-1}\log^{2/\delta_v} T, T^{-1/2}\sqrt{\log n})),$$

because of part (ii) and Assumption 2(a).

To conclude, for part (v) we have

$$\|(\widehat{\boldsymbol{\Sigma}}_n^\xi)^{-1} - (\boldsymbol{\Sigma}_n^\xi)^{-1}\| \le \|(\widehat{\boldsymbol{\Sigma}}_n^\xi)^{-1}\|\,\|\widehat{\boldsymbol{\Sigma}}_n^\xi - \boldsymbol{\Sigma}_n^\xi\|\,\|(\boldsymbol{\Sigma}_n^\xi)^{-1}\| = O_p(\max(n^{-1}\log^{2/\delta_v} T, T^{-1/2}\sqrt{\log n})),$$

by parts (iii), (iv), and Assumption 2(a). This completes the proof. □

Lemma F.2. *Consider the EM estimators $\widehat{\boldsymbol{\lambda}}_i \equiv \widehat{\boldsymbol{\lambda}}_i^{(k+1)}$, $\widehat{\boldsymbol{\Lambda}}_n \equiv \widehat{\boldsymbol{\Lambda}}_n^{(k+1)}$, $\widehat{\sigma}_i^2 \equiv \widehat{\sigma}_i^{2(k+1)}$, and $\widehat{\boldsymbol{\Sigma}}_n^\xi \equiv \widehat{\boldsymbol{\Sigma}}_n^{\xi(k+1)}$, for any $k \ge 0$. Under Assumptions 1, 2, 3, 4, 5, and 6, as $n, T \to \infty$:*

(i) $\min(n\log^{-2/\delta_v} T, \sqrt{T}\log^{-1/2} n)\, n^{-1}\|\widehat{\boldsymbol{\Lambda}}_n'(\widehat{\boldsymbol{\Sigma}}_n^\xi)^{-1}\widehat{\boldsymbol{\Lambda}}_n - \boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n\| = O_p(1)$;
(ii) $\min(n\log^{-2/\delta_v} T, \sqrt{T}\log^{-1/2} n)\, n^{-1/2}\|\widehat{\boldsymbol{\Lambda}}_n'(\widehat{\boldsymbol{\Sigma}}_n^\xi)^{-1} - \boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}\| = O_p(1)$;
(iii) $n\|(\widehat{\boldsymbol{\Lambda}}_n'(\widehat{\boldsymbol{\Sigma}}_n^\xi)^{-1}\widehat{\boldsymbol{\Lambda}}_n)^{-1}\| = O_p(1)$;
(iv) $\min(n\log^{-2/\delta_v} T, \sqrt{T}\log^{-1/2} n)\, n\|(\widehat{\boldsymbol{\Lambda}}_n'(\widehat{\boldsymbol{\Sigma}}_n^\xi)^{-1}\widehat{\boldsymbol{\Lambda}}_n)^{-1} - (\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n)^{-1}\| = O_p(1)$;
(v) $\omega_{n,T,\delta_v}\sqrt{n}\|(\widehat{\boldsymbol{\Lambda}}_n'(\widehat{\boldsymbol{\Sigma}}_n^\xi)^{-1}\widehat{\boldsymbol{\Lambda}}_n)^{-1}\widehat{\boldsymbol{\Lambda}}_n'(\widehat{\boldsymbol{\Sigma}}_n^\xi)^{-1} - (\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n)^{-1}\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}\| = O_p(1)$, *with* $\omega_{n,T,\delta_v} = \min(n\log^{-2/\delta_v} T, \sqrt{T}\log^{-1/2} n)$.

Proof. The proof is the same as the proof of Lemma E.15 but using Proposition 2(a) (which does not require this lemma to be proved) and Lemma F.1 instead of Lemmas E.13 and E.14. □

Lemma F.3. *Consider the MSEs estimators $\widehat{\mathbf{P}}_{t|t-1} \equiv \mathbf{P}_{t|t-1}^{(k+1)}$, $\widehat{\mathbf{P}}_{t|t} \equiv \mathbf{P}_{t|t}^{(k+1)}$, and $\widehat{\mathbf{P}}_{t|T} \equiv \mathbf{P}_{t|T}^{(k+1)}$ derived from the Kalman filter and smoother and obtained using the EM estimators $\widehat{\boldsymbol{\lambda}}_i \equiv \widehat{\boldsymbol{\lambda}}_i^{(k+1)}$, $\widehat{\boldsymbol{\Lambda}}_n \equiv \widehat{\boldsymbol{\Lambda}}_n^{(k+1)}$, $\widehat{\sigma}_i^2 \equiv \widehat{\sigma}_i^{2(k+1)}$, and $\widehat{\boldsymbol{\Sigma}}_n^\xi \equiv \widehat{\boldsymbol{\Sigma}}_n^{\xi(k+1)}$, for any $k \ge 0$. Under Assumptions 1, 2, 3, 4, 5, and 6, as $n, T \to \infty$:*

(i) $\max_{t=1,\dots,T}\|\widehat{\mathbf{P}}_{t|t-1}\| = O_p(1)$;
(ii) $\max_{t=1,\dots,T}\|(\widehat{\mathbf{P}}_{t|t-1})^{-1}\| = O_p(1)$;
(iii) $\max_{t=1,\dots,T} n\|\widehat{\mathbf{P}}_{t|t}\| = O_p(1)$;
(iv) $\max_{t=1,\dots,T} n\|\widehat{\mathbf{P}}_{t|T}\| = O_p(1)$.

Proof. The proof is the same as the proof of Lemma E.16 but using Proposition 2(a) (which does not require this lemma to be proved) and Lemmas F.1 and F.2 instead of Lemmas E.13, E.14, and E.15. □

Lemma F.4. *Consider the Kalman filter and smoother estimators $\widehat{\mathbf{F}}_{t|t-1} \equiv \mathbf{F}_{t|t-1}^{(k+1)}$, $\widehat{\mathbf{F}}_{t|t} \equiv \mathbf{F}_{t|t}^{(k+1)}$, and $\widehat{\mathbf{F}}_{t|T} \equiv \mathbf{F}_{t|T}^{(k+1)}$ obtained using the EM estimators $\widehat{\boldsymbol{\lambda}}_i \equiv \widehat{\boldsymbol{\lambda}}_i^{(k+1)}$, $\widehat{\boldsymbol{\Lambda}}_n \equiv \widehat{\boldsymbol{\Lambda}}_n^{(k+1)}$, $\widehat{\sigma}_i^2 \equiv \widehat{\sigma}_i^{2(k+1)}$, and $\widehat{\boldsymbol{\Sigma}}_n^\xi \equiv \widehat{\boldsymbol{\Sigma}}_n^{\xi(k+1)}$, for any $k \ge 0$. Under Assumptions 1, 2, 3, 4, 5, and 6, as $n, T \to \infty$:*

(i) for all $s = 0, \dots, T$, $\|\widehat{\mathbf{F}}_{t|s}\| = O_p(1)$, *uniformly in* $t \le s$;
(ii) $n\|\widehat{\mathbf{F}}_{t|T} - \widehat{\mathbf{F}}_{t|t}\| = O_p(1)$, *uniformly in* t;
(iii) $n\|\widehat{\mathbf{F}}_{t|t} - \widehat{\mathbf{F}}_t^{\mathrm{WLS}}\| = O_p(1)$, *uniformly in* t;
where $\widehat{\mathbf{F}}_t^{\mathrm{WLS}} = (\widehat{\boldsymbol{\Lambda}}_n'(\widehat{\boldsymbol{\Sigma}}_n^\xi)^{-1}\widehat{\boldsymbol{\Lambda}}_n)^{-1}\widehat{\boldsymbol{\Lambda}}_n'(\widehat{\boldsymbol{\Sigma}}_n^\xi)^{-1}\mathbf{x}_{nt}$.

Proof. The proof is the same as the proof of Lemma E.17 but using Proposition 2(a) (which does not require this lemma to be proved) and Lemmas F.1, F.2, and F.3 instead of Lemmas E.13, E.14, E.15, and D.12. □

Lemma F.5. *Consider the Kalman filter and smoother estimators $\widehat{\mathbf{F}}_{t|t-1} \equiv \mathbf{F}_{t|t-1}^{(k+1)}$, $\widehat{\mathbf{F}}_{t|t} \equiv \mathbf{F}_{t|t}^{(k+1)}$, and $\widehat{\mathbf{F}}_{t|T} \equiv \mathbf{F}_{t|T}^{(k+1)}$ obtained using the EM estimators $\widehat{\boldsymbol{\lambda}}_i \equiv \widehat{\boldsymbol{\lambda}}_i^{(k+1)}$, $\widehat{\boldsymbol{\Lambda}}_n \equiv \widehat{\boldsymbol{\Lambda}}_n^{(k+1)}$, $\widehat{\sigma}_i^2 \equiv \widehat{\sigma}_i^{2(k+1)}$, and $\widehat{\boldsymbol{\Sigma}}_n^\xi \equiv \widehat{\boldsymbol{\Sigma}}_n^{\xi(k+1)}$, for any $k \ge 0$. Under Assumptions 1, 2, 3, 4, 5, and 6, as $n, T \to \infty$:*

(i) for all $s = 0, \dots, T$, $\log^{-1/\delta_v} T \max_{t=1,\dots,s}\|\widehat{\mathbf{F}}_{t|s}\| = O_p(1)$;
(ii) $\log^{-1/\delta_v} T \max_{t=1,\dots,T} n\|\widehat{\mathbf{F}}_{t|T} - \widehat{\mathbf{F}}_{t|t}\| = O_p(1)$;
(iii) $\log^{-1/\delta_v} T \max_{t=1,\dots,T} n\|\widehat{\mathbf{F}}_{t|t} - \widehat{\mathbf{F}}_t^{\mathrm{WLS}}\| = O_p(1)$;

where $\widehat{\mathbf{F}}_t^{\text{WLS}} = (\widehat{\mathbf{\Lambda}}_n'(\widehat{\mathbf{\Sigma}}_n^{\xi})^{-1}\widehat{\mathbf{\Lambda}}_n)^{-1}\widehat{\mathbf{\Lambda}}_n'(\widehat{\mathbf{\Sigma}}_n^{\xi})^{-1}\mathbf{x}_{nt}$.

PROOF. From (D.44) in the proof of Lemma D.14, but when computed using the EM estimator of the parameters, we see that the only modification is that we need to use $\max_{t=1,\dots,T} n^{-1/2}\|\mathbf{x}_t\| = O_p()$ from Lemma E.3. This proves part (i).

Part (ii) follows from part (i) and (D.45) in the proof of Lemma D.15, but when computed using the EM estimator of the parameters.

Part (iii) follows from (D.50) in the proof of Lemma D.16, but when computed using the EM estimator of the parameters, and using Lemma E.3 again and the fact that $\widehat{\mathbf{F}}_{t-1|t-1}$ is a weighted average of $\mathbf{x}_{n1},\dots,\mathbf{x}_{n,t-1}$. This completes the proof. □

G Lemmas necessary for proving Proposition 5

Lemma G.1. *Under Assumptions 1, 2, 3, 5, and 6, as* $n, T\to\infty$,

$$\min(n, \sqrt{T}\log^{-1/2} n) \max_{i=1,\dots,n} |\widehat{\sigma}_i^{(0)2} - \sigma_i^2| = O_p(1).$$

PROOF. From (D.6) in the proof of Lemma D.4, we have

$$\begin{aligned}
\max_{i=1,\dots,n}|\widehat{\sigma}_i^{2(0)} - \sigma_i^2| \le &\max_{i=1,\dots,n}\left|T^{-1}\sum_{t=1}^T x_{it}^2 - \mathbb{E}[x_{it}^2]\right| + \max_{i=1,\dots,n}\left|\widehat{\boldsymbol{\lambda}}_i^{(0)\prime}\widehat{\boldsymbol{\lambda}}_i^{(0)} - \boldsymbol{\lambda}_i'\boldsymbol{\lambda}_i\right|\\
&+2\max_{i=1,\dots,n}\|\boldsymbol{\lambda}_i\|\max_{i=1,\dots,n}\|\widehat{\boldsymbol{\lambda}}_i^{(0)} - \boldsymbol{\lambda}_i\|\\
&+2\max_{i=1,\dots,n}\|\boldsymbol{\lambda}_i\|\left\|T^{-1}\sum_{t=1}^T\mathbf{F}_t\mathbf{F}_t' - \mathbf{I}_r\right\|\left\{\max_{i=1,\dots,n}\|\widehat{\boldsymbol{\lambda}}_i^{(0)} - \boldsymbol{\lambda}_i\| + \max_{i=1,\dots,n}\|\boldsymbol{\lambda}_i\|\right\}\\
&+2\max_{i=1,\dots,n}\|\boldsymbol{\lambda}_i\|\left\|T^{-1}\sum_{t=1}^T\mathbf{F}_t(\widetilde{\mathbf{F}}_t - \mathbf{F}_t)'\right\|\left\{\max_{i=1,\dots,n}\|\widehat{\boldsymbol{\lambda}}_i^{(0)} - \boldsymbol{\lambda}_i\| + \max_{i=1,\dots,n}\|\boldsymbol{\lambda}_i\|\right\}\\
&+2\max_{i=1,\dots,n}\left\|T^{-1}\sum_{t=1}^T\xi_{it}\mathbf{F}_t'\right\|\max_{i=1,\dots,n}\|\boldsymbol{\lambda}_i\|\\
&+2\max_{i=1,\dots,n}\left\|T^{-1}\sum_{t=1}^T\xi_{it}\mathbf{F}_t'\right\|\max_{i=1,\dots,n}\|\widehat{\boldsymbol{\lambda}}_i^{(0)} - \boldsymbol{\lambda}_i\|\\
&+2\max_{i=1,\dots,n}\left\|T^{-1}\sum_{t=1}^T\xi_{it}(\widetilde{\mathbf{F}}_t - \mathbf{F}_t)'\right\|\left\{\max_{i=1,\dots,n}\|\widehat{\boldsymbol{\lambda}}_i^{(0)} - \boldsymbol{\lambda}_i\| + \max_{i=1,\dots,n}\|\boldsymbol{\lambda}_i\|\right\}\\
=&O_p(\max(n^{-1}, T^{-1/2}\sqrt{\log n})),
\end{aligned}$$

by Assumption 1(a), Lemma E.1(ii) and the union bound, and (E.71) and (E.79) in the proof of Lemma E.14. This completes the proof. □

Lemma G.2. *Under Assumptions 1, 2, 3, 5, and 6, as* $n, T\to\infty$, *for any* $k\ge 0$,

(i) $\min(n\log^{-2/\delta_v} T, \sqrt{nT}\log^{-1/2} n, T\log^{-1/2} n)$, $\|T^{-1}\sum_{t=1}^T(\mathbf{F}_{t|T}^{(k)} - \mathbf{F}_t)\mathbf{F}_t'\| = O_p(1)$;

(ii) $\min(n\log^{-2/\delta_v} T, \sqrt{nT}\log^{-1/2} n, T\log^{-1/2} n)\,\|T^{-1}\sum_{t=1}^T(\mathbf{F}_{t|T}^{(k)} - \mathbf{F}_t)\xi_{it}\| = O_p(1)$, *uniformly in* i.

PROOF. Throughout, let $\boldsymbol{y}_t = \mathbf{F}_t$ or $\boldsymbol{y}_t = \xi_{it}$. Then, for any $k\ge 0$ we have to consider

$$\begin{aligned}
\left\|T^{-1}\sum_{t=1}^T(\mathbf{F}_{t|T}^{(k)} - \mathbf{F}_t)\boldsymbol{y}_t'\right\| \le &\left\|T^{-1}\sum_{t=1}^T(\mathbf{F}_{t|T}^{(k)} - \mathbf{F}_{t|t}^{(k)})\boldsymbol{y}_t'\right\|\\
&+\left\|T^{-1}\sum_{t=1}^T(\mathbf{F}_{t|t}^{(k)} - \widehat{\mathbf{F}}_t^{\text{WLS}(k)})\boldsymbol{y}_t'\right\|\\
&+\left\|T^{-1}\sum_{t=1}^T(\widehat{\mathbf{F}}_t^{\text{WLS}(k)} - \mathbf{F}_t)\boldsymbol{y}_t'\right\|\\
=&\,\mathcal{I} + \mathcal{II} + \mathcal{III},\quad\text{say.}
\end{aligned}\tag{G.1}$$

where $\widehat{\mathbf{F}}_t^{\mathrm{WLS}(k)} = (\widehat{\mathbf{\Lambda}}_n^{(k)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(k)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(k)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(k)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(k)})^{-1}\mathbf{x}_{nt}$. Let us consider each term in (G.1). First, when $k=0$ we have $\mathcal{I} = O_p(n^{-1})$ by (B.5) in the proof of Proposition 1 (which does not require this lemma to be proved), while, when $k \geq 1$

$$\mathcal{I} \leq \max_{t=1,\ldots,T} \|\mathbf{F}_{t|T}^{(k)} - \mathbf{F}_{t|t}^{(k)}\| \left\| T^{-1}\sum_{t=1}^T \boldsymbol{y}_t \right\| = O_p(n^{-1}\log^{1/\delta_v} T), \tag{G.2}$$

by Lemma F.5(ii) and since

$$\mathbb{E}\left[\left\| T^{-1}\sum_{t=1}^T \mathbf{F}_t\right\|^2\right] \leq rT^{-2}\max_{j=1,\ldots,r}\sum_{t,s=1}^T |\mathbb{E}[F_{jt}F_{js}]| \leq 1, \tag{G.3}$$

by Cauchy-Schwarz inequality and Assumption 6(b), and also

$$\mathbb{E}\left[\left| T^{-1}\sum_{t=1}^T \xi_{it}\right|^2\right] \leq T^{-2}\sum_{t,s=1}^T |\mathbb{E}[\xi_{it}\xi_{is}]| \leq T^{-1}M_3, \tag{G.4}$$

by Lemma C.1(iii). Second, when $k=0$ we have $\mathcal{II} = O_p(n^{-1})$ by (B.6) in the proof of Proposition 1 (which does not require this lemma to be proved), while, when $k \geq 1$

$$\mathcal{II} \leq \max_{t=1,\ldots,T} \|\mathbf{F}_{t|t}^{(k)} - \widehat{\mathbf{F}}_t^{\mathrm{WLS}(k)}\| \left\| T^{-1}\sum_{t=1}^T \boldsymbol{y}_t \right\| = O_p(n^{-1}\log^{1/\delta_v} T), \tag{G.5}$$

by Lemma F.5(iii), (G.3), and (G.4).

Third,

$$\begin{aligned}
\mathcal{III} \leq\ & \|(\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^{\xi})^{-1}\mathbf{\Lambda}_n)^{-1}\| \, \|\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^{\xi})^{-1}(\mathbf{\Lambda}_n - \widehat{\mathbf{\Lambda}}_n^{(k)})\| \left\| T^{-1}\sum_{t=1}^T \mathbf{F}_t\boldsymbol{y}_t'\right\| \\
&+ \|n(\widehat{\mathbf{\Lambda}}_n^{(k)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(k)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(k)})^{-1}n^{-1/2}\widehat{\mathbf{\Lambda}}_n^{(k)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(k)})^{-1} - n(\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^{\xi})^{-1}\mathbf{\Lambda}_n)^{-1}n^{-1/2}\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^{\xi})^{-1}\| \\
&\cdot n^{-1/2}\|\mathbf{\Lambda}_n - \widehat{\mathbf{\Lambda}}_n^{(k)}\| \left\| T^{-1}\sum_{t=1}^T \mathbf{F}_t\boldsymbol{y}_t'\right\| \\
&+ n\|(\widehat{\mathbf{\Lambda}}_n^{(k)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(k)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(k)})^{-1}\|\, n^{-1}\left\| T^{-1}\sum_{t=1}^T \mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^{\xi})^{-1}\boldsymbol{\xi}_{nt}\boldsymbol{y}_t'\right\| \\
&+ n\|(\widehat{\mathbf{\Lambda}}_n^{(k)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(k)})^{-1}\widehat{\mathbf{\Lambda}}_n^{(k)})^{-1}\|\, n^{-1}\left\| T^{-1}\sum_{t=1}^T \{\widehat{\mathbf{\Lambda}}_n^{(k)\prime}(\widehat{\mathbf{\Sigma}}_n^{\xi(k)})^{-1} - \mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^{\xi})^{-1}\}\boldsymbol{\xi}_{nt}\boldsymbol{y}_t'\right\| \\
=\ & \mathcal{III}_a + \mathcal{III}_b + \mathcal{III}_c + \mathcal{III}_d, \text{ say.}
\end{aligned} \tag{G.6}$$

Then, when $k=0$, $\mathcal{III}_a = III_a = O_p(n^{-1/2}T^{-1/2})$, $\mathcal{III}_b = III_b = O_p(\max(n^{-2}, T^{-1}))$, and $\mathcal{III}_c = III_c = O_p(n^{-1/2}T^{-1/2})$, because of (B.9), (B.10), (B.11), and (B.12) in the proof of Proposition 1 (which does not require this lemma to be proved). While, if $k \geq 1$,

$$\mathcal{III}_a = O_p(\max(n^{-1}\log^{2/\delta_v} T, n^{-1/2}T^{-1/2}\sqrt{\log n}, T^{-1})), \tag{G.7}$$

by (B.40), (B.41), and (B.42) in the proof of Proposition 3 (which does not require this lemma to be proved), and since $\|T^{-1}\sum_{t=1}^T \mathbf{F}_t\mathbf{F}_t'\| = O_p(1)$ by Lemma C.12(i) and Assumption 6(b), and $\|T^{-1}\sum_{t=1}^T \mathbf{F}_t\xi_{it}\| = O_p(T^{-1/2})$ by Lemma C.12(ii). These last arguments, and (B.43) in the proof of Proposition 3 (which does not require this lemma to be proved) imply also that, if $k \geq 1$,

$$\mathcal{III}_b = O_p(\max(n^{-2}\log^{4/\delta_v} T, T^{-1}\sqrt{\log n}, n^{-1}T^{-1/2}\log^{2/\delta_v} T\sqrt{\log n})). \tag{G.8}$$

Moreover, if $\boldsymbol{y}_t = \mathbf{F}_t$ and $k \geq 1$, then

$$\begin{aligned}\mathcal{III}_c &= n\|(\widehat{\boldsymbol{\Lambda}}_n^{(k)\prime}(\widehat{\boldsymbol{\Sigma}}_n^{\xi(k)})^{-1}\widehat{\boldsymbol{\Lambda}}_n^{(k)})^{-1}\|\, n^{-1}\left\|T^{-1}\sum_{t=1}^T \boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^{\xi})^{-1}\boldsymbol{\xi}_{nt}\mathbf{F}_t'\right\| \\ &= O_p(n^{-1/2}T^{-1/2}), \end{aligned} \tag{G.9}$$

by Lemmas F.2(iii) and C.8(iv). While if $\boldsymbol{y}_t = \xi_{it}$ and $k \geq 1$, then

$$\begin{aligned}\mathcal{III}_c &= n\|(\widehat{\boldsymbol{\Lambda}}_n^{(k)\prime}(\widehat{\boldsymbol{\Sigma}}_n^{\xi(k)})^{-1}\widehat{\boldsymbol{\Lambda}}_n^{(k)})^{-1}\|\, n^{-1}\left\|T^{-1}\sum_{t=1}^T \boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^{\xi})^{-1}\boldsymbol{\xi}_{nt}\xi_{it}\right\| \\ &= O_p(n^{-1/2}T^{-1/2}), \end{aligned} \tag{G.10}$$

by Lemmas F.2(iii) and C.8(v).

Similarly, if $\boldsymbol{y}_t = \mathbf{F}_t$, we have:

$$\begin{aligned}\mathcal{III}_d &= n\|(\widehat{\boldsymbol{\Lambda}}_n^{(k)\prime}(\widehat{\boldsymbol{\Sigma}}_n^{\xi(k)})^{-1}\widehat{\boldsymbol{\Lambda}}_n^{(k)})^{-1}\|\, n^{-1}\left\|T^{-1}\sum_{t=1}^T \{\widehat{\boldsymbol{\Lambda}}_n^{(k)\prime}(\widehat{\boldsymbol{\Sigma}}_n^{\xi(k)})^{-1} - \boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^{\xi})^{-1}\}\boldsymbol{\xi}_{nt}\mathbf{F}_t'\right\| \\ &\leq n\|(\widehat{\boldsymbol{\Lambda}}_n^{(k)\prime}(\widehat{\boldsymbol{\Sigma}}_n^{\xi(k)})^{-1}\widehat{\boldsymbol{\Lambda}}_n^{(k)})^{-1}\|\, n^{-1/2}\|\widehat{\boldsymbol{\Lambda}}_n^{(k)\prime}(\widehat{\boldsymbol{\Sigma}}_n^{\xi(k)})^{-1} - \boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^{\xi})^{-1}\|\, n^{-1/2}\left\|T^{-1}\sum_{t=1}^T \boldsymbol{\xi}_{nt}\mathbf{F}_t'\right\| \\ &= \begin{cases} O_p(\max(n^{-1}T^{-1/2}, T^{-1})), & \text{if} \quad k = 0, \\ O_p(\max(n^{-1}T^{-1/2}\log^{2/\delta_v} T, T^{-1}\sqrt{\log n})), & \text{if} \quad k \geq 1, \end{cases}\end{aligned} \tag{G.11}$$

where, if $k = 0$, we used Lemmas D.5(ii), D.5(iii), and C.12(iii), while, if $k \geq 1$, we used Lemmas F.2(ii), F.2(iii), and C.12(iii).

Finally, if $\boldsymbol{y}_t = \xi_{it}$, we have:

$$\begin{aligned}\mathcal{III}_d &= n\|(\widehat{\boldsymbol{\Lambda}}_n^{(k)\prime}(\widehat{\boldsymbol{\Sigma}}_n^{\xi(k)})^{-1}\widehat{\boldsymbol{\Lambda}}_n^{(k)})^{-1}\|\, n^{-1}\left\|T^{-1}\sum_{t=1}^T \{\widehat{\boldsymbol{\Lambda}}_n^{(k)\prime}(\widehat{\boldsymbol{\Sigma}}_n^{\xi(k)})^{-1} - \boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^{\xi})^{-1}\}\boldsymbol{\xi}_{nt}\xi_{it}\right\| \\ &\leq n\|(\widehat{\boldsymbol{\Lambda}}_n^{(k)\prime}(\widehat{\boldsymbol{\Sigma}}_n^{\xi(k)})^{-1}\widehat{\boldsymbol{\Lambda}}_n^{(k)})^{-1}\| \left\{ n^{-1}\left\|T^{-1}\sum_{t=1}^T (\widehat{\boldsymbol{\Lambda}}_n^{(k)} - \boldsymbol{\Lambda}_n)'(\boldsymbol{\Sigma}_n^{\xi})^{-1}\boldsymbol{\xi}_{nt}\xi_{it}\right\| \right. \\ &\quad + n^{-1}\left\|T^{-1}\sum_{t=1}^T \boldsymbol{\Lambda}_n'\{(\widehat{\boldsymbol{\Sigma}}_n^{\xi(k)})^{-1} - (\boldsymbol{\Sigma}_n^{\xi})^{-1}\}\boldsymbol{\xi}_{nt}\xi_{it}\right\| \\ &\quad \left. + n^{-1}\left\|T^{-1}\sum_{t=1}^T (\widehat{\boldsymbol{\Lambda}}_n^{(k)} - \boldsymbol{\Lambda}_n)'\{(\widehat{\boldsymbol{\Sigma}}_n^{\xi(k)})^{-1} - (\boldsymbol{\Sigma}_n^{\xi})^{-1}\}\boldsymbol{\xi}_{nt}\xi_{it}\right\| \right\} \\ &= n\|(\widehat{\boldsymbol{\Lambda}}_n^{(k)\prime}(\widehat{\boldsymbol{\Sigma}}_n^{\xi(k)})^{-1}\widehat{\boldsymbol{\Lambda}}_n^{(k)})^{-1}\|\, \{\mathcal{III}_{d.1} + \mathcal{III}_{d.2} + \mathcal{III}_{d.3}\}, \text{ say.}\end{aligned} \tag{G.12}$$

Now,

$$\begin{aligned}\mathcal{III}_{d.1} &= n^{-1}\left\|T^{-1}\sum_{t=1}^T (\boldsymbol{\Lambda}_n^{\text{OLS}} - \boldsymbol{\Lambda}_n)'(\boldsymbol{\Sigma}_n^{\xi})^{-1}\boldsymbol{\xi}_{nt}\xi_{it}\right\| \\ &\quad + n^{-1/2}\|\widehat{\boldsymbol{\Lambda}}_n^{(k)} - \boldsymbol{\Lambda}_n^{\text{OLS}}\|\, \|(\boldsymbol{\Sigma}_n^{\xi})^{-1}\|\, n^{-1/2}\left\|T^{-1}\sum_{t=1}^T \boldsymbol{\xi}_{nt}\xi_{it}\right\| \\ &= \mathcal{III}_{d.1.1} + \mathcal{III}_{d.1.2}, \text{say.}\end{aligned} \tag{G.13}$$

And,

$$\begin{aligned}
\mathcal{III}_{d.1.1} =& n^{-1}\left\|T^{-1}\sum_{t=1}^T\left(T^{-1}\sum_{s=1}^T\mathbf{F}_s\mathbf{F}_s'\right)^{-1}\left(T^{-1}\sum_{s=1}^T\mathbf{F}_s\boldsymbol{\xi}_{ns}'\right)(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\xi}_{nt}\xi_{it}\right\|\\
\le& \left\|\left(T^{-1}\sum_{s=1}^T\mathbf{F}_s\mathbf{F}_s'\right)^{-1}\right\| n^{-1}\left\|T^{-2}\sum_{s,t=1}^T\mathbf{F}_s\boldsymbol{\xi}_{ns}'(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\xi}_{nt}\xi_{it}\right\|\\
\le& \left\|\left(T^{-1}\sum_{s=1}^T\mathbf{F}_s\mathbf{F}_s'\right)^{-1}\right\| n^{-1}\left\|T^{-2}\sum_{s,t=1}^T\mathbf{F}_s\boldsymbol{\xi}_{ns}'(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\xi}_{nt}\right\|\max_{t=1,\ldots,T}|\xi_{it}|\\
=& O_p(n^{-1/2}T^{-3/2}\log^{1/\delta_v}T),
\end{aligned}\tag{G.14}$$

by (E.9) in the proof of Lemma E.3, and Lemmas C.8(vi) and C.13.

$$\mathcal{III}_{d.1.2} = \begin{cases} O_p(\max(n^{-1},n^{-1/2}T^{-1/2},T^{-1})), & \text{if}\quad k=0,\\ O_p(\max(n^{-1}\log^{2/\delta_v}T,n^{-1/2}T^{-1/2}\sqrt{\log n},T^{-1})), & \text{if}\quad k\ge 1,\end{cases}\tag{G.15}$$

where we used Assumption 2(a), the last relation in (B.7) in the proof of Proposition 1 (which does not require this lemma to be proved), and, if $k=0$, we used also Barigozzi (2023, Corollary 1 and Proposition B.3), or, if $k\ge 1$, we used (B.39) in the proof of Proposition 3 (which does not require this lemma to be proved).

By using (G.14) and (G.15) into (G.13), we have

$$\begin{aligned}
\mathcal{III}_{d.1} =& O_p(n^{-1/2}T^{-3/2}\log^{1/\delta_v}T)\\
&+\begin{cases} O_p(\max(n^{-1},n^{-1/2}T^{-1/2},T^{-1})), & \text{if}\quad k=0,\\ O_p(\max(n^{-1}\log^{2/\delta_v}T,n^{-1/2}T^{-1/2}\sqrt{\log n},T^{-1})), & \text{if}\quad k\ge 1,\end{cases}
\end{aligned}\tag{G.16}$$

Furthermore,

$$\begin{aligned}
\mathcal{III}_{d.2} =& n^{-1}\left\|T^{-1}\sum_{t=1}^T\sum_{j=1}^n\boldsymbol{\lambda}_j\{(\widehat{\sigma}_j^{2(k)})^{-1}-(\sigma_j^2)^{-1}\}\xi_{jt}\xi_{it}\right\|\\
\le& C_\xi^2\max_{j=1,\ldots,n}|\widehat{\sigma}_j^{2(k)}-\sigma_j^{(k)}|n^{-1}\left\|T^{-1}\sum_{t=1}^T\sum_{j=1}^n\boldsymbol{\lambda}_j\xi_{jt}\xi_{it}\right\|\\
\le& C_\xi^2\max_{j=1,\ldots,n}|\widehat{\sigma}_j^{2(k)}-\sigma_j^{(k)}|n^{-1}\left\|T^{-1}\sum_{t=1}^T\boldsymbol{\Lambda}_n'\boldsymbol{\xi}_{nt}\xi_{it}\right\|\\
=& O_p(n^{-1/2}T^{-1/2})\cdot\begin{cases} O_p(\max(n^{-1},T^{-1/2}\sqrt{\log n})), & \text{if}\quad k=0,\\ O_p(\max(n^{-1}\log^{2/\delta_v}T,T^{-1/2}\sqrt{\log n})), & \text{if}\quad k\ge 1,\end{cases}
\end{aligned}\tag{G.17}$$

where, if $k=0$, we used Lemmas G.1 and C.8(ii), while, if $k\ge 1$, we used Lemmas F.1(ii) and C.8(ii). Term $\mathcal{III}_{d.3}$ is dominated by $\mathcal{III}_{d.1}$.

By using Lemmas F.2(iii) or D.5(iii), together with (G.16) and (G.17) into (G.12), we have that if $\boldsymbol{y}_t=\xi_{it}$, then

$$\begin{aligned}
\mathcal{III}_d =& O_p(n^{-1/2}T^{-3/2}\log^{1/\delta_v}T)+O_p(\max(n^{-1}\log^{2/\delta_v}T,n^{-1/2}T^{-1/2}\sqrt{\log n},T^{-1}))\\
&+O_p(\max(n^{-3/2}T^{-1/2}\log^{2/\delta_v}T,n^{-1/2}T^{-1}\sqrt{\log n})).
\end{aligned}\tag{G.18}$$

By substituting (G.7), (G.8), (G.9), (G.10), (G.11), and (G.18) into (G.6), if $\boldsymbol{y}_t=\mathbf{F}_t$, we have

$$\begin{aligned}
\mathcal{III} =& O_p(\max(n^{-1}\log^{2/\delta_v}T,n^{-1/2}T^{-1/2}\sqrt{\log n},T^{-1}))\\
&+O_p(\max(n^{-2}\log^{4/\delta_v}T,T^{-1}\sqrt{\log n},n^{-1}T^{-1/2}\log^{2/\delta_v}T\sqrt{\log n}))+O_p(n^{-1/2}T^{-1/2})\\
&+O_p(\max(n^{-1}T^{-1/2}\log^{2/\delta_v}T,T^{-1}\sqrt{\log n})),
\end{aligned}\tag{G.19}$$

while, if $\boldsymbol{y}_t = \xi_{it}$, we have

$$\begin{aligned}\mathcal{III} = &O_p(\max(n^{-1}\log^{2/\delta_v} T, n^{-1/2}T^{-1/2}\sqrt{\log n}, T^{-1}))\\ &+ O_p(\max(n^{-2}\log^{4/\delta_v} T, T^{-1}\sqrt{\log n}, n^{-1}T^{-1/2}\log^{2/\delta_v} T\sqrt{\log n})) + O_p(n^{-1/2}T^{-1/2})\\ &+ O_p(n^{-1/2}T^{-3/2}\log^{1/\delta_v} T) + O_p(\max(n^{-1}\log^{2/\delta_v} T, n^{-1/2}T^{-1/2}\sqrt{\log n}, T^{-1}))\\ &+ O_p(\max(n^{-3/2}T^{-1/2}\log^{2/\delta_v} T, n^{-1/2}T^{-1}\sqrt{\log n})).\end{aligned}\tag{G.20}$$

To conclude by substituting (G.2), (G.5), and either (G.19) or (G.20) into (G.1), we have

$$\left\|T^{-1}\sum_{t=1}^T(\mathbf{F}_{t|T}^{(k)} - \mathbf{F}_t)\mathbf{F}_t'\right\| = O_p(\max(n^{-1}\log^{2/\delta_v} T, n^{-1/2}T^{-1/2}\sqrt{\log n}, T^{-1}\sqrt{\log n})),$$

which proves part (i), and

$$\left\|T^{-1}\sum_{t=1}^T(\mathbf{F}_{t|T}^{(k)} - \mathbf{F}_t)\xi_{it}\right\| = O_p(\max(n^{-1}\log^{2/\delta_v} T, n^{-1/2}T^{-1/2}\sqrt{\log n}, T^{-1}\sqrt{\log n})),$$

which proves part (ii) and completes the proof. □

H Lemmas necessary for proving Proposition 6

Lemma H.1. *Consider the initial estimator of the factors $\widetilde{\mathbf{F}}_t$ defined in Section A.1, then, under Assumptions 1, 2, 3, and 6, as $n, T \to \infty$, if $\sqrt{n}/T \to 0$,*

$$\sqrt{n}(\widetilde{\mathbf{F}}_t - \mathbf{F}_t) \xrightarrow{d} \mathcal{N}(\mathbf{0}_r, \boldsymbol{\mathcal{W}}_t^{\mathrm{PC}}),$$

for any given $t = 1, \dots, T$, where

$$\boldsymbol{\mathcal{W}}_t^{\mathrm{PC}} = (\boldsymbol{\Sigma}_\Lambda)^{-1}\left(\lim_{n\to\infty} n^{-1}\sum_{i,j=1}^n \mathbb{E}[\xi_{it}\xi_{jt}]\boldsymbol{\lambda}_i\boldsymbol{\lambda}_j\right)(\boldsymbol{\Sigma}_\Lambda)^{-1}$$

with $\boldsymbol{\Sigma}_\Lambda = \lim_{n\to\infty} n^{-1}\sum_{i=1}^n \boldsymbol{\lambda}_i\boldsymbol{\lambda}_i'$;

Proof. By definition of the pre-estimator in Section A.1:

$$\begin{aligned}\widetilde{\mathbf{F}}_t - \mathbf{F}_t &= (\widehat{\boldsymbol{\Lambda}}_n^{(0)\prime}\widehat{\boldsymbol{\Lambda}}_n^{(0)})^{-1}\widehat{\boldsymbol{\Lambda}}_n^{(0)\prime}\mathbf{x}_{nt} - \mathbf{F}_t\\ &= n(\mathbf{M}_n^\chi)^{-1}\left\{n^{-1}\widehat{\boldsymbol{\Lambda}}_n^{(0)\prime}(\boldsymbol{\Lambda}_n - \widehat{\boldsymbol{\Lambda}}_n^{(0)})\mathbf{F}_t + n^{-1}(\widehat{\boldsymbol{\Lambda}}_n^{(0)} - \boldsymbol{\Lambda}_n)'\boldsymbol{\xi}_{nt} + n^{-1}\boldsymbol{\Lambda}_n'\boldsymbol{\xi}_{nt}\right\}\\ &\quad + n\left\{(\widehat{\mathbf{M}}_n^x)^{-1} - (\mathbf{M}_n^\chi)^{-1}\right\}\left\{n^{-1}\widehat{\boldsymbol{\Lambda}}_n^{(0)\prime}(\boldsymbol{\Lambda}_n - \widehat{\boldsymbol{\Lambda}}_n^{(0)})\mathbf{F}_t + n^{-1}(\widehat{\boldsymbol{\Lambda}}_n^{(0)} - \boldsymbol{\Lambda}_n)'\boldsymbol{\xi}_{nt} + n^{-1}\boldsymbol{\Lambda}_n'\boldsymbol{\xi}_{nt}\right\}.\end{aligned}\tag{H.1}$$

Then,

$$n\|(\mathbf{M}_n^\chi)^{-1}\| = n\{\mu_{nr}^\chi\}^{-1} \ge \underline{C}_r,\tag{H.2}$$

by Lemma C.1(iv). Moreover, by Lemmas, C.1(v), C.11 and C.12(vii), and Merikoski and Kumar (2004, Theorem 1), which is Weyl's inequality, for all $j = 1, \dots, r$,

$$n^{-1}|\widehat{\mu}_j^x - \mu_j^\chi| \le n^{-1}\left\|T^{-1}\sum_{t=1}^T \mathbf{x}_t\mathbf{x}_t' - \boldsymbol{\Gamma}^x\right\| + n^{-1}\|\boldsymbol{\Gamma}^\xi\| = O_p(\max(n^{-1}, T^{-1/2})),\tag{H.3}$$

which, jointly with (H.2), implies

$$\det(n^{-1}\widehat{\mathbf{M}}_n^x) = \prod_{j=1}^r n^{-1}\widehat{\mu}_j^x \ge \{n^{-1}\widehat{\mu}_r^x\}^r \ge \{n^{-1}\mu_r^\chi - n^{-1}|\widehat{\mu}_r^x - \mu_r^\chi|\}^r > 0,$$

thus,

$$n\|(\widehat{\mathbf{M}}_n^x)^{-1}\| = O_p(1). \tag{H.4}$$

From (H.2), (H.3), and (H.4)

$$n\left\|(\widehat{\mathbf{M}}_n^x)^{-1} - (\mathbf{M}_n^\chi)^{-1}\right\| \le n\|(\mathbf{M}_n^\chi)^{-1}\|\, n^{-1}\left\|\widehat{\mathbf{M}}_n^x - \mathbf{M}_n^\chi\right\|\, n\|(\widehat{\mathbf{M}}_n^x)^{-1}\| = O_p(\max(n^{-1}, T^{-1/2})). \tag{H.5}$$

Furthermore,

$$\begin{aligned}
n^{-1}\|\widehat{\mathbf{\Lambda}}_n^{(0)\prime}(\mathbf{\Lambda}_n - \widehat{\mathbf{\Lambda}}_n^{(0)})\mathbf{F}_t\| \le\,& n^{-1}\|\mathbf{\Lambda}_n'(\mathbf{\Lambda}_n - \widehat{\mathbf{\Lambda}}_n^{(0)})\|\,\|\mathbf{F}_t\| + n^{-1}\|\mathbf{\Lambda}_n - \widehat{\mathbf{\Lambda}}_n^{(0)}\|^2\,\|\mathbf{F}_t\|\\
\le\,& n^{-1}\|\mathbf{\Lambda}_n'(\mathbf{\Lambda}_n - \widehat{\mathbf{\Lambda}}_n^{\mathrm{OLS}})\|\,\|\mathbf{F}_t\| + n^{-1/2}\|\mathbf{\Lambda}_n\|\, n^{-1/2}\|\widehat{\mathbf{\Lambda}}_n^{(0)} - \mathbf{\Lambda}_n^{\mathrm{OLS}}\|\,\|\mathbf{F}_t\|\\
&+ n^{-1}\|\mathbf{\Lambda}_n - \widehat{\mathbf{\Lambda}}_n^{(0)}\|^2\,\|\mathbf{F}_t\|\\
\le\,& n^{-1}\left\|T^{-1}\sum_{t=1}^T \mathbf{\Lambda}_n'\boldsymbol{\xi}_{nt}\mathbf{F}_t'\right\|\,\left\|\left(T^{-1}\sum_{t=1}^T \mathbf{F}_t\mathbf{F}_t'\right)^{-1}\right\|\,\|\mathbf{F}_t\|\\
&+ n^{-1/2}\|\mathbf{\Lambda}_n\|\, n^{-1/2}\|\widehat{\mathbf{\Lambda}}_n^{(0)} - \mathbf{\Lambda}_n^{\mathrm{OLS}}\|\,\|\mathbf{F}_t\| + n^{-1}\|\mathbf{\Lambda}_n - \widehat{\mathbf{\Lambda}}_n^{(0)}\|^2\,\|\mathbf{F}_t\|\\
=\,& O_p(n^{-1/2}T^{-1/2}) + O_p(\max(n^{-1}, n^{-1/2}T^{-1/2})) + O_p(\max(n^{-2}, T^{-1}))\\
=\,& O_p(\max(n^{-1}, n^{-1/2}T^{-1/2}, T^{-1})),
\end{aligned} \tag{H.6}$$

by Barigozzi (2023, Corollary 1), Lemmas C.2, C.8(i), C.13, and D.1(b), and since $\|\mathbf{F}_t\| = O_p(1)$ because $\mathbb{E}[F_{jt}^2] = 1$, $j = 1,\dots,r$, by Assumption 6(b).

Last,

$$\begin{aligned}
n^{-1}\|(\widehat{\mathbf{\Lambda}}_n^{(0)} - \mathbf{\Lambda}_n)'\boldsymbol{\xi}_{nt}\| \le\,& n^{-1}\|(\mathbf{\Lambda}_n^{\mathrm{OLS}} - \mathbf{\Lambda}_n)'\boldsymbol{\xi}_{nt}\| + n^{-1/2}\|\widehat{\mathbf{\Lambda}}_n^{(0)} - \mathbf{\Lambda}_n^{\mathrm{OLS}}\|\, n^{-1/2}\|\boldsymbol{\xi}_{nt}\|\\
\le\,& n^{-1}\left\|T^{-1}\sum_{t=1}^T \mathbf{F}_t\boldsymbol{\xi}_{nt}'\boldsymbol{\xi}_{nt}\right\|\,\left\|\left(T^{-1}\sum_{t=1}^T \mathbf{F}_t\mathbf{F}_t'\right)^{-1}\right\|\\
&+ n^{-1/2}\|\widehat{\mathbf{\Lambda}}_n^{(0)} - \mathbf{\Lambda}_n^{\mathrm{OLS}}\|\, n^{-1/2}\|\boldsymbol{\xi}_{nt}\|\\
=\,& O_p(n^{-1/2}T^{-1/2}) + O_p(\max(n^{-1}, n^{-1/2}T^{-1/2}))\\
=\,& O_p(\max(n^{-1}, n^{-1/2}T^{-1/2})),
\end{aligned} \tag{H.7}$$

by Barigozzi (2023, Corollary 1), Lemmas C.8(vi) (setting $\mathbf{\Sigma}_n^\xi = \mathbf{I}_n$ and using $\boldsymbol{\xi}_{nt}$ in place of $\boldsymbol{\mathcal{E}}_{nT}$ therein), C.13, and D.1(b), and since $n^{-1/2}\|\boldsymbol{\xi}_{nt}\| = O_p(1)$ because $\sum_{i=1}^n \sigma_i^2 \le nC_\xi$ by Assumption 2(a).

By substituting (H.5), (H.6), and (H.7) into (H.1), it follows that if $\sqrt{n}/T \to 0$, as $n, T \to \infty$,

$$\sqrt{n}(\widetilde{\mathbf{F}}_t - \mathbf{F}_t) = n(\mathbf{M}_n^\chi)^{-1}(n^{-1/2}\mathbf{\Lambda}_n'\boldsymbol{\xi}_{nt}) + o_p(1). \tag{H.8}$$

and since $\lim_{n\to\infty} n^{-1}\mathbf{M}_n^\chi = \mathbf{\Sigma}_\Lambda$ by Assumption 6(b) which is positive definite, by Assumption 2(e) (when setting $\sigma_i^2 = 1$ therein) and Slutsky's Theorem we complete the proof. □

I Derivation of the Kalman filter MSE

Lemma I.1. *Under Assumption 1, the DFM (3)-(4) is both stabilizable and detectable, for all $n \in \mathbb{N}$.*

Proof. We use the definitions in Anderson and Moore (1979, Appendix C, p. 341-342). The DFM (3)-(4) is a linear systems with r states. A linear system is stabilizable if its unstable states are controllable and all uncontrollable states are stable, and it is detectable if its unstable states are observable and all unobservable states are stable.

First, by factorizing $\mathbf{\Gamma}^v = \mathbf{H}\mathbf{H}'$ for some $\mathbf{H}$ having full-column rank, we see that $\mathrm{rk}[\mathbf{H}\ (\mathbf{A}\mathbf{H})\cdots(\mathbf{A}^{(r-1)}\mathbf{H})] = r$, thus the linear system is controllable, by Assumption 1(e). Moreover, there are no unstable states, since because of Assumption 1(d) all eigenvalues of $\mathbf{A}$ are smaller than one in absolute value. This implies that the model is stabilizable.

Second, since by Assumption 1(a) for any given $n \in \mathbb{N}$ there exists at least one $i = 1,\dots,n$ such that $\|\boldsymbol{\lambda}_i\| \ge m_\lambda$, then

$\text{rk}(\mathbf{\Lambda}_n) \geq 1$, while, $\text{rk}(\mathbf{\Lambda}_n) = r$ only for $n > N$, then, for a given n, there might be $(r-1)$ unobservable states, however, as already noticed, they are all stable by Assumption 1(d). Thus the model is detectable. This completes the proof. □

Lemma I.2. *Under Assumptions 1 and 2, the matrix* $\mathbf{P}_{t|t-1}$ *has a steady-state denoted as* $\mathbf{P} = \lim_{t\to\infty} \mathbf{P}_{t|t-1}$.

Proof. First, given that with our initialization $\mathbf{P}_{0|0} = \mathbf{\Gamma}^F$, then it is positive definite by Assumption 1(b), therefore also $\mathbf{P}_{1|0}$ is positive definite (see also (A.2)). Second, as proved in Lemma I.1, the linear system defining the DFM (3)-(4) is stabilizable and detectable. Therefore, because of Chan et al. (1984, Theorem 4.1), as $t \to \infty$, $\mathbf{P}_{t|t-1}$ converges to a steady-state $\mathbf{P}$ exponentially fast (see Lemma I.4 below for the rate of convergence), which is a solution of the algebraic Riccati equation (ARE) derived from (A.5)

$$\mathbf{P} = \mathbf{A}\mathbf{P}\mathbf{A}' - \mathbf{A}\mathbf{P}\mathbf{\Lambda}_n'(\mathbf{\Lambda}_n\mathbf{P}\mathbf{\Lambda}_n' + \mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n\mathbf{P}\mathbf{A}' + \mathbf{\Gamma}^v.$$

Moreover, since Lemmas D.7(i) and D.7(ii) hold for all $T \in \mathbb{N}$, we have

$$\|\mathbf{P}\| = O(1), \qquad \|\mathbf{P}^{-1}\| = O(1). \tag{I.1}$$

This completes the proof. □

Lemma I.3. *Under Assumptions 1 and 2, the matrix* $\mathbf{\Pi}_{t|t-1}$ *has a steady-state denoted as* $\mathbf{\Pi} = \lim_{t\to\infty} \mathbf{\Pi}_{t|t-1}$.

Proof. The existence of $\mathbf{\Pi}$ follows from the same arguments used in Lemma I.2. Moreover, from Harvey and Delle Monache (2009, Section 2.2) we have that $\mathbf{\Pi}$ must satisfy

$$\begin{aligned}\mathbf{\Pi} =& \mathbf{A}\mathbf{\Pi}\mathbf{A}' + \mathbf{A}\mathbf{P}\mathbf{\Lambda}_n'(\mathbf{\Lambda}_n\mathbf{P}\mathbf{\Lambda}_n' + \mathbf{\Sigma}_n^\xi)^{-1}(\mathbf{\Lambda}_n\mathbf{\Pi}\mathbf{\Lambda}_n' + \mathbf{\Gamma}_n^\xi)(\mathbf{\Lambda}_n\mathbf{P}\mathbf{\Lambda}_n' + \mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n\mathbf{P}\mathbf{A}' \\ &- \mathbf{A}\mathbf{P}\mathbf{\Lambda}_n'(\mathbf{\Lambda}_n\mathbf{P}\mathbf{\Lambda}_n' + \mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n\mathbf{\Pi}\mathbf{A}' - \mathbf{A}\mathbf{\Pi}\mathbf{\Lambda}_n'(\mathbf{\Lambda}_n\mathbf{P}\mathbf{\Lambda}_n' + \mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n\mathbf{P}\mathbf{A}' + \mathbf{\Gamma}^v.\end{aligned}$$

Moreover, by the same arguments in Lemmas D.7(i) and D.7(ii) it holds that $\max_{t=1,\ldots,T} \|\mathbf{\Pi}_{t|t-1}\| = O(1)$ and also $\max_{t=1,\ldots,T} \|(\mathbf{\Pi}_{t|t-1})^{-1}\| = O(1)$, for all $T \in \mathbb{N}$, thus

$$\|\mathbf{\Pi}\| = O(1), \qquad \|\mathbf{\Pi}^{-1}\| = O(1). \tag{I.2}$$

This completes the proof. □

Lemma I.4. *Under Assumptions 1, 2, and 6, if* $\|\mathbf{P}_{0|0}\| = O(n^\gamma)$ *for some* $\gamma > 0$, *then* $n \max_{t=\bar{t},\ldots,T} \|\mathbf{P}_{t|t-1} - \mathbf{P}\| = o(1)$ *and* $n \max_{t=\bar{t},\ldots,T} \|\mathbf{\Pi}_{t|t-1} - \mathbf{\Pi}\| = o(1)$, *where* $\bar{t} = \lceil 2 + \gamma/2 \rceil$.

Proof. Let $\mathbf{\Psi}_{1,1} = \mathbf{I}_r$ and, for $t = 2, \ldots, T$,

$$\mathbf{\Psi}_{t,1} = \prod_{s=1}^{t-1} [\mathbf{A} - \mathbf{A}\mathbf{P}_{s|s-1}\mathbf{\Lambda}_n'(\mathbf{\Lambda}_n\mathbf{P}_{s|s-1}\mathbf{\Lambda}_n' + \mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n]. \tag{I.3}$$

Then, from Anderson and Moore (1979, Chapter 4.4, pp. 76-81), we have, for $t = 1, \ldots, T$,

$$\begin{aligned}\mathbf{P}_{t|t-1} - \mathbf{P} &= \{\mathbf{A} - \mathbf{A}\mathbf{P}\mathbf{\Lambda}_n'(\mathbf{\Lambda}_n\mathbf{P}\mathbf{\Lambda}_n' + \mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n\}^{t-1}(\mathbf{P}_{1|0} - \mathbf{P})\mathbf{\Psi}_{t,1} \\ &= \mathbf{A}^{t-1}\{\mathbf{I}_r - \mathbf{P}\mathbf{\Lambda}_n'(\mathbf{\Lambda}_n\mathbf{P}\mathbf{\Lambda}_n' + \mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n\}^{t-1}(\mathbf{P}_{1|0} - \mathbf{P})\mathbf{\Psi}_{t,1} \\ &= \mathbf{A}^{t-1}\{\mathbf{I}_r - (\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n + \mathbf{P}^{-1})^{-1}\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n\}^{t-1}(\mathbf{P}_{1|0} - \mathbf{P})\mathbf{\Psi}_{t,1},\end{aligned} \tag{I.4}$$

where we used Lemma D.13.

Now, for $t = 2, \ldots, T$, from (I.3) and using again Lemma D.13, there exists a $\bar{n}$ such that for all $n \geq \bar{n}$

$$\|\mathbf{\Psi}_{t,1}\| \leq \|\mathbf{A}\|^{t-1} \prod_{s=1}^{t-1} \|\mathbf{I}_r - (\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n + \mathbf{P}_{s|s-1}^{-1})^{-1}\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n\| \leq M_0 n^{-(t-1)}, \tag{I.5}$$

for some finite positive real M_0 independent of n and t, because of Lemmas C.6(i) and D.7(ii), and Assumption 1(d). Moreover,

$$\|\mathbf{P}_{1|0} - \mathbf{P}\| = \|\mathbf{A}\mathbf{P}_{0|0} - \mathbf{P}\| \leq \|\mathbf{A}\| \, \|\mathbf{P}_{0|0}\| + \|\mathbf{A}\| \, \|\mathbf{P}\| \leq 2\|\mathbf{A}\| \, \|\mathbf{P}_{0|0}\|, \tag{I.6}$$

since because of Lemma D.6, $\|\mathbf{P}\| \le \|\mathbf{P}_{t|t-1}\| \le \|\mathbf{P}_{1|0}\| \le \mathbf{P}_{0|0}\|$.

Therefore, for $t = 2, \dots, T$, from (I.4), (I.5), and (I.6), there exists a $\bar{n}$ such that for all $n \ge \bar{n}$

$$\begin{aligned}\|\mathbf{P}_{t|t-1} - \mathbf{P}\| \le& 2\|\mathbf{A}\|^t \|\mathbf{I}_r - (\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n + \mathbf{P}^{-1})^{-1}\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n\}\|^{t-1} \|\mathbf{P}_{0|0}\| \, \|\mathbf{\Psi}_{t,1}\| \\ \le& M_1 n^{-2(t-1)}\|\mathbf{P}_{0|0}\|,\end{aligned} \tag{I.7}$$

for some finite positive real M_1 independent of n and t, because of Lemmas C.6(i) and D.7(ii), and Assumption 1(d).

Now, define $\bar{t}$ the first point in time such that the rhs of (I.7) is $o(n^{-1})$, i.e., such that

$$n^{-2(\bar{t}-1)} n^{1+\epsilon} \le M_2 \|\mathbf{P}_{0|0}\|^{-1},$$

for any $\epsilon > 0$ and some finite positive real M_2 independent of n and t, or equivalently, by letting $K = \log M_2$, $\bar{t}$ is such that:

$$\bar{t} \ge (3+\epsilon)/2 + \log \|\mathbf{P}_{0|0}\|/(2\log n) - K/(2\log n). \tag{I.8}$$

Clearly (I.8) is always satisfied if we set $\bar{t} = \lceil 2 + \log \|\mathbf{P}_{0|0}\|/(2\log n)\rceil$. Letting now $\|\mathbf{P}_{0|0}\| = O(n^\gamma)$ for some $\gamma > 0$, then we have $\bar{t} = \lceil 2 + \gamma/2 \rceil$ satisfies (I.8) and therefore we have at least $\max_{t=\bar{t},\dots,T} \|\mathbf{P}_{t|t-1} - \mathbf{P}\| = O(n^{-1})$. Notice that if $\gamma = 0$, i.e., $\|\mathbf{P}_{0|0}\|$ is finite, then $\bar{t} = 2$ and in this case from (I.7) we have an even tighter bound as we get $\max_{t=2,\dots,T} \|\mathbf{P}_{t|t-1} - \mathbf{P}\| = O(n^{-2})$.

The proof for $\mathbf{\Pi}_{t|t-1}$ can be done analogously but using the recursions in Harvey and Delle Monache (2009). This completes the proof. □

Lemma I.5. *Under Assumptions 1, 2, and 6,*

(i) $n \max_{t=\bar{t},\dots,T} \|\boldsymbol{\mathcal{P}}_t\| = O(1)$*;*

(ii) $n \max_{t=1,\dots,T} \|\mathbf{\Pi}_{t|t}\| = O(1)$*;*

(iii) $n \max_{\bar{t},\dots,T} \|\mathbf{\Pi}_{t|t} - \boldsymbol{\mathcal{P}}_t\| = o(1)$*,*

(iv) $\max_{t=\bar{t},\dots,T} \|n\boldsymbol{\mathcal{P}}_t - \boldsymbol{\mathcal{W}}_t\| = o(1)$*;*

where $\mathbf{\Pi}_{t|t}$ *is defined in* (22), $\boldsymbol{\mathcal{P}}_t$ *is obtained from* $\mathbf{\Pi}_{t|t}$ *when replacing* $\mathbf{P}_{t|t-1}$ *and* $\mathbf{\Pi}_{t|t-1}$ *with their steady states defined in Lemmas I.2 and I.3, respectively, and* $\boldsymbol{\mathcal{W}}_t$ *is defined in Proposition 3.*

Proof. We have

$$\begin{aligned}\boldsymbol{\mathcal{P}}_t =& \mathbf{\Pi} + \mathbf{P}\mathbf{\Lambda}_n'(\mathbf{\Lambda}_n\mathbf{P}\mathbf{\Lambda}_n' + \mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n\mathbf{\Pi}\mathbf{\Lambda}_n'(\mathbf{\Lambda}_n\mathbf{P}\mathbf{\Lambda}_n' + \mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n\mathbf{P} \\ &- \mathbf{P}\mathbf{\Lambda}_n'(\mathbf{\Lambda}_n\mathbf{P}\mathbf{\Lambda}_n' + \mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n\mathbf{\Pi} - \mathbf{\Pi}\mathbf{\Lambda}_n'(\mathbf{\Lambda}_n\mathbf{P}\mathbf{\Lambda}_n' + \mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n\mathbf{P} \\ &+ \mathbf{P}\mathbf{\Lambda}_n'(\mathbf{\Lambda}_n\mathbf{P}\mathbf{\Lambda}_n' + \mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Gamma}_n^\xi(\mathbf{\Lambda}_n\mathbf{P}\mathbf{\Lambda}_n' + \mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n\mathbf{P}.\end{aligned} \tag{I.9}$$

Hereafter, for simplicity of notation let $\boldsymbol{H} = (\mathbf{\Lambda}_n'(\mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n)^{-1}$. Using twice Lemma C.4 we have

$$\begin{aligned}\mathbf{P}(\boldsymbol{H} + \mathbf{P})^{-1} &= \mathbf{P}\left\{\mathbf{P}^{-1} - (\boldsymbol{H}+\mathbf{P})^{-1}\boldsymbol{H}\mathbf{P}^{-1}\right\} \\ &= \mathbf{I}_r - \mathbf{P}\left\{\mathbf{P}^{-1} - (\boldsymbol{H}+\mathbf{P})^{-1}\boldsymbol{H}\mathbf{P}^{-1}\right\}\boldsymbol{H}\mathbf{P}^{-1} \\ &= \mathbf{I}_r - \boldsymbol{H}\mathbf{P}^{-1} + \mathbf{P}(\boldsymbol{H}+\mathbf{P})^{-1}\boldsymbol{H}\mathbf{P}^{-1}\boldsymbol{H}\mathbf{P}^{-1} \\ &= \mathbf{I}_r - \boldsymbol{H}\mathbf{P}^{-1} + \mathbf{C}, \ \ \text{say}.\end{aligned} \tag{I.10}$$

Then, by Lemma D.10 and (I.10)

$$\mathbf{\Pi} - \mathbf{P}\mathbf{\Lambda}_n'(\mathbf{\Lambda}_n\mathbf{P}\mathbf{\Lambda}_n' + \mathbf{\Sigma}_n^\xi)^{-1}\mathbf{\Lambda}_n\mathbf{\Pi} = \left\{\mathbf{I}_r - \mathbf{P}(\boldsymbol{H}+\mathbf{P})^{-1}\right\}\mathbf{\Pi} = \boldsymbol{H}\mathbf{P}^{-1}\mathbf{\Pi} - \mathbf{C}\mathbf{\Pi}. \tag{I.11}$$

Similarly, again by Lemma D.10 and (I.10)

$$\begin{aligned}
&\mathbf{P}\boldsymbol{\Lambda}_n'(\boldsymbol{\Lambda}_n\mathbf{P}\boldsymbol{\Lambda}_n'+\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n\boldsymbol{\Pi}\boldsymbol{\Lambda}_n'(\boldsymbol{\Lambda}_n\mathbf{P}\boldsymbol{\Lambda}_n'+\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n\mathbf{P}-\boldsymbol{\Pi}\boldsymbol{\Lambda}_n'(\boldsymbol{\Lambda}_n\mathbf{P}\boldsymbol{\Lambda}_n'+\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n\mathbf{P}\\
&\quad=\mathbf{P}(\boldsymbol{H}+\mathbf{P})^{-1}\boldsymbol{\Pi}(\boldsymbol{H}+\mathbf{P})^{-1}\mathbf{P}-\boldsymbol{\Pi}\mathbf{P}(\boldsymbol{H}+\mathbf{P})^{-1}\\
&\quad=\left\{\mathbf{I}_r-\boldsymbol{H}\mathbf{P}^{-1}+\mathbf{C}\right\}\boldsymbol{\Pi}\left\{\mathbf{I}_r-\mathbf{P}^{-1}\boldsymbol{H}+\mathbf{C}\right\}-\boldsymbol{\Pi}\left\{\mathbf{I}_r-\mathbf{P}^{-1}\boldsymbol{H}+\mathbf{C}\right\}\\
&\quad=\boldsymbol{\Pi}-\boldsymbol{\Pi}\mathbf{P}^{-1}\boldsymbol{H}-\boldsymbol{H}\mathbf{P}^{-1}\boldsymbol{\Pi}+\boldsymbol{H}\mathbf{P}^{-1}\boldsymbol{\Pi}\mathbf{P}^{-1}\boldsymbol{H}-\boldsymbol{\Pi}+\boldsymbol{\Pi}\mathbf{P}^{-1}\boldsymbol{H}\\
&\quad\quad+\mathbf{C}\boldsymbol{\Pi}+\boldsymbol{\Pi}\mathbf{C}-\boldsymbol{H}\mathbf{P}^{-1}\boldsymbol{\Pi}\mathbf{C}-\mathbf{C}\boldsymbol{\Pi}\mathbf{P}^{-1}\boldsymbol{H}+\mathbf{C}\boldsymbol{\Pi}\mathbf{C}-\boldsymbol{\Pi}\mathbf{C}\\
&\quad=-\boldsymbol{H}\mathbf{P}^{-1}\boldsymbol{\Pi}+\boldsymbol{H}\mathbf{P}^{-1}\boldsymbol{\Pi}\mathbf{P}^{-1}\boldsymbol{H}+\mathbf{C}\boldsymbol{\Pi}-\boldsymbol{H}\mathbf{P}^{-1}\boldsymbol{\Pi}\mathbf{C}-\mathbf{C}\boldsymbol{\Pi}\mathbf{P}^{-1}\boldsymbol{H}+\mathbf{C}\boldsymbol{\Pi}\mathbf{C}.
\end{aligned}\tag{I.12}$$

By substituting (I.11) and (I.12) into (I.9):

$$\begin{aligned}
\boldsymbol{\mathcal{P}}_t&=\mathbf{P}\boldsymbol{\Lambda}_n'(\boldsymbol{\Lambda}_n\mathbf{P}\boldsymbol{\Lambda}_n'+\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Gamma}_n^\xi(\boldsymbol{\Lambda}_n\mathbf{P}\boldsymbol{\Lambda}_n'+\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n\mathbf{P}\\
&\quad+\boldsymbol{H}\mathbf{P}^{-1}\boldsymbol{\Pi}-\mathbf{C}\boldsymbol{\Pi}-\boldsymbol{H}\mathbf{P}^{-1}\boldsymbol{\Pi}+\boldsymbol{H}\mathbf{P}^{-1}\boldsymbol{\Pi}\mathbf{P}^{-1}\boldsymbol{H}+\mathbf{C}\boldsymbol{\Pi}-\boldsymbol{H}\mathbf{P}^{-1}\mathbf{C}\boldsymbol{\Pi}-\mathbf{C}\boldsymbol{\Pi}\mathbf{P}^{-1}\boldsymbol{H}+\mathbf{C}\boldsymbol{\Pi}\mathbf{C}\\
&=\mathbf{P}\boldsymbol{\Lambda}_n'(\boldsymbol{\Lambda}_n\mathbf{P}\boldsymbol{\Lambda}_n'+\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Gamma}_n^\xi(\boldsymbol{\Lambda}_n\mathbf{P}\boldsymbol{\Lambda}_n'+\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n\mathbf{P}\\
&\quad-\mathbf{C}\boldsymbol{\Pi}+\boldsymbol{H}\mathbf{P}^{-1}\boldsymbol{\Pi}\mathbf{P}^{-1}\boldsymbol{H}+\mathbf{C}\boldsymbol{\Pi}-\boldsymbol{H}\mathbf{P}^{-1}\boldsymbol{\Pi}\mathbf{C}-\mathbf{C}\boldsymbol{\Pi}\mathbf{P}^{-1}\boldsymbol{H}+\mathbf{C}\boldsymbol{\Pi}\mathbf{C}\\
&=(\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n+\mathbf{P}^{-1})^{-1}\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Gamma}_n^\xi(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n(\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n+\mathbf{P}^{-1})^{-1}\\
&\quad+\boldsymbol{H}\mathbf{P}^{-1}\boldsymbol{\Pi}\mathbf{P}^{-1}\boldsymbol{H}-\boldsymbol{H}\mathbf{P}^{-1}\boldsymbol{\Pi}\mathbf{C}-\mathbf{C}\boldsymbol{\Pi}\mathbf{P}^{-1}\boldsymbol{H}+\mathbf{C}\boldsymbol{\Pi}\mathbf{C}.
\end{aligned}\tag{I.13}$$

where in the last step we used Lemma D.13. Moreover,

$$\|\boldsymbol{H}\|=\|(\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n)^{-1}\|=O(n^{-1}),\tag{I.14}$$

$$\begin{aligned}
\|\mathbf{C}\|&=\|\mathbf{P}((\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n)^{-1}+\mathbf{P})^{-1}(\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n)^{-1}\mathbf{P}^{-1}(\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n)^{-1}\mathbf{P}^{-1}\|\\
&\le\|\mathbf{P}\|\,\|(\mathbf{P})^{-1}\|^2\|((\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n)^{-1}+\mathbf{P})^{-1}\|\,\|(\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n)^{-1}\|^2=O(n^{-2}),.
\end{aligned}\tag{I.15}$$

because of Lemma C.3(iii), (I.1) in the proof of Lemma I.2, and since, by Merikoski and Kumar (2004, Theorem 1) which is Weyl's inequality,

$$\begin{aligned}
&\|((\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n)^{-1}+\mathbf{P})^{-1}\|=\left\{\nu^{(r)}((\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n)^{-1}+\mathbf{P})\right\}^{-1}\\
&\le\left\{\nu^{(r)}((\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n)^{-1})+\nu^{(r)}(\mathbf{P})\right\}^{-1}\\
&=\left\{\left[\nu^{(1)}(\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n)\right]^{-1}+\nu^{(r)}(\mathbf{P})\right\}^{-1}\\
&=\left\{\left[\nu^{(1)}(\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n)\nu^{(r)}(\mathbf{P})\right]^{-1}+1\right\}^{-1}\left\{\nu^{(r)}(\mathbf{P})\right\}^{-1}\\
&=\left\{1-\left[\nu^{(1)}(\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n)\nu^{(r)}(\mathbf{P})\right]^{-1}\right\}\left\{\nu^{(r)}(\mathbf{P})\right\}^{-1}+O_p(n^{-2})\\
&=O_p(1),
\end{aligned}$$

again by Lemma C.3(iii) and (I.1) in the proof of Lemma I.2. Therefore, from (I.13), (I.14), and (I.15)

$$\begin{aligned}
n\|\boldsymbol{\mathcal{P}}_t\|&\le n\|(\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n+\mathbf{P}^{-1})^{-1}\|^2\|\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}\|^2\|\boldsymbol{\Gamma}_n^\xi\|\\
&\quad+n\|\boldsymbol{H}\|^2\|(\mathbf{P})^{-1}\|^2\|\boldsymbol{\Pi}\|+2n\|\boldsymbol{H}\|\,\|(\mathbf{P})^{-1}\|\,\|\boldsymbol{\Pi}\|\,\|\mathbf{C}\|+n\|\mathbf{C}\|^2\|\boldsymbol{\Pi}\|\\
&=O(1)+O(n^{-1})+O(n^{-2})+O(n^{-3}),
\end{aligned}\tag{I.16}$$

where we used also (I.1) and (I.2) in the proofs of Lemmas I.2 and I.3, respectively. Since $\boldsymbol{\mathcal{P}}_t$ does not depend on t we prove part (i). Part (ii) is proved analogously by substituting in part (i) $\mathbf{P}$ and $\boldsymbol{\Pi}$ with $\mathbf{P}_{t|t-1}$ and $\boldsymbol{\Pi}_{t|t-1}$ and using Lemma D.7(i) and D.7(ii) instead of (I.1) and (I.2).

Part (iii) follows directly from Lemmas I.2, I.3, and I.4, and parts (i) and (ii).

Turning to part (iv), consider the first term on the rhs of (I.13), using Lemma D.13 we have

$$\begin{aligned}
&\mathbf{P}\boldsymbol{\Lambda}_n'(\boldsymbol{\Lambda}_n\mathbf{P}\boldsymbol{\Lambda}_n'+\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Gamma}_n^\xi(\boldsymbol{\Lambda}_n\mathbf{P}\boldsymbol{\Lambda}_n'+\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n\mathbf{P} = (\boldsymbol{H}^{-1}+\mathbf{P}^{-1})^{-1}\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Gamma}_n^\xi(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n(\boldsymbol{H}^{-1}+\mathbf{P}^{-1})^{-1}\\
&\quad=\Big\{(\boldsymbol{H}^{-1}+\mathbf{P}^{-1})^{-1}\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}-\boldsymbol{H}\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}+\boldsymbol{H}\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}\Big\}\boldsymbol{\Gamma}_n^\xi\\
&\qquad\cdot\Big\{(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n(\boldsymbol{H}^{-1}+\mathbf{P}^{-1})^{-1}-(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n\boldsymbol{H}+(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n\boldsymbol{H}\Big\}\\
&\quad=\boldsymbol{H}\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Gamma}_n^\xi(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n\boldsymbol{H}\\
&\qquad+\Big\{(\boldsymbol{H}^{-1}+\mathbf{P}^{-1})^{-1}\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}-\boldsymbol{H}\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}\Big\}\boldsymbol{\Gamma}_n^\xi(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n\boldsymbol{H}\\
&\qquad+\boldsymbol{H}\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Gamma}_n^\xi\Big\{(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n(\boldsymbol{H}^{-1}+\mathbf{P}^{-1})^{-1}-(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n\boldsymbol{H}\Big\}\\
&\qquad+\Big\{(\boldsymbol{H}^{-1}+\mathbf{P}^{-1})^{-1}\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}-\boldsymbol{H}\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}\Big\}\boldsymbol{\Gamma}_n^\xi\Big\{(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n(\boldsymbol{H}^{-1}+\mathbf{P}^{-1})^{-1}-(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n\boldsymbol{H}\Big\}\\
&\quad=\boldsymbol{H}\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Gamma}_n^\xi(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n\boldsymbol{H}+\boldsymbol{\mathcal{A}}+\boldsymbol{\mathcal{A}}'+\boldsymbol{\mathcal{B}},\ \text{say.}
\end{aligned}\tag{I.17}$$

Then,

$$\begin{aligned}
n\|\boldsymbol{\mathcal{A}}\|&\le n\|(\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n+\mathbf{P}^{-1})^{-1}\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}-(\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n)^{-1}\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}\|\,\|\boldsymbol{H}\|\,\|\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}\|\,\|\boldsymbol{\Gamma}_n^\xi\|\\
&=O(n^{-3/2}),
\end{aligned}\tag{I.18}$$

because of Lemmas C.1(v), C.3(vii), and C.6(iii), and (I.14). Similarly,

$$\begin{aligned}
n\|\boldsymbol{\mathcal{B}}\|&\le n\|(\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n+\mathbf{P}^{-1})^{-1}\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}-(\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n)^{-1}\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}\|^2\|\boldsymbol{\Gamma}_n^\xi\|\\
&=O(n^{-3}),
\end{aligned}\tag{I.19}$$

because of Lemmas C.1(v) and C.6(iii). Furthermore,

$$n\|\boldsymbol{H}\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Gamma}_n^\xi(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n\boldsymbol{H}\|\le n\|\boldsymbol{H}\|^2\|\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}\|^2\|\boldsymbol{\Gamma}_n^\xi\|=O(1),\tag{I.20}$$

because of Lemmas C.1(v), C.3(vii), and (I.14).

Thus, by using (I.18), (I.19), and (I.20), from (I.17), we have

$$\begin{aligned}
&n\|\mathbf{P}\boldsymbol{\Lambda}_n'(\boldsymbol{\Lambda}_n\mathbf{P}\boldsymbol{\Lambda}_n'+\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Gamma}_n^\xi(\boldsymbol{\Lambda}_n\mathbf{P}\boldsymbol{\Lambda}_n'+\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n\mathbf{P}-\boldsymbol{H}\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Gamma}_n^\xi(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n\boldsymbol{H}\|\\
&\quad\le 2n\|\boldsymbol{\mathcal{A}}\|+n\|\boldsymbol{\mathcal{B}}\|=O(n^{-3/2}).
\end{aligned}\tag{I.21}$$

And, by using (I.21) and (I.16) into (I.13) we have

$$\begin{aligned}
n\|\boldsymbol{\mathcal{P}}_t-\boldsymbol{H}\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Gamma}_n^\xi(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n\boldsymbol{H}\|&\le n\|\mathbf{P}\boldsymbol{\Lambda}_n'(\boldsymbol{\Lambda}_n\mathbf{P}\boldsymbol{\Lambda}_n'+\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Gamma}_n^\xi(\boldsymbol{\Lambda}_n\mathbf{P}\boldsymbol{\Lambda}_n'+\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n\mathbf{P}-\boldsymbol{H}\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Gamma}_n^\xi(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n\boldsymbol{H}\|\\
&\quad+n\|\boldsymbol{H}\|^2\|(\mathbf{P})^{-1}\|^2\|\boldsymbol{\Pi}\|+2n\|\boldsymbol{H}\|\,\|(\mathbf{P})^{-1}\|\,\|\boldsymbol{\Pi}\|\,\|\mathbf{C}\|+n\|\mathbf{C}\|^2\|\boldsymbol{\Pi}\|\\
&=O(n^{-1}).
\end{aligned}\tag{I.22}$$

Finally, notice that, by definition:

$$\begin{aligned}
\lim_{n\to\infty}n\boldsymbol{H}\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Gamma}_n^\xi(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n\boldsymbol{H}&=\lim_{n\to\infty}n\boldsymbol{H}n^{-1}\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Gamma}_n^\xi(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n n\boldsymbol{H}\\
&=(\boldsymbol{\Sigma}_{\Lambda\Sigma\Lambda})^{-1}\Big\{\lim_{n\to\infty}n^{-1}\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Gamma}_n^\xi(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n\Big\}(\boldsymbol{\Sigma}_{\Lambda\Sigma\Lambda})^{-1}=\boldsymbol{\mathcal{W}}_t.
\end{aligned}\tag{I.23}$$

Therefore, from (I.22) and (I.23)

$$\|n\boldsymbol{\mathcal{P}}_t-\boldsymbol{\mathcal{W}}_t\|\le\|n\boldsymbol{\mathcal{P}}_t-n\boldsymbol{H}\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Gamma}_n^\xi(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n\boldsymbol{H}\|+\|n\boldsymbol{H}\boldsymbol{\Lambda}_n'(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Gamma}_n^\xi(\boldsymbol{\Sigma}_n^\xi)^{-1}\boldsymbol{\Lambda}_n\boldsymbol{H}-\boldsymbol{\mathcal{W}}_t\|=O(n^{-1})+o(1),$$

and since $\boldsymbol{\mathcal{P}}_t$ and $\boldsymbol{\mathcal{W}}_t$ do not depend on t we complete the proof. □

J Data description and data treatment

This appendix presents the dataset by Barigozzi and Luciani (2023) that we used for the empirical analysis.

Table J1: List of Abbreviations for Table J2 and Table J3

Source:		
BLS	=	U.S. Department of Labor: Bureau of Labor Statistics
BEA	=	U.S. Department of Commerce: Bureau of Economic Analysis
ISM	=	Institute for Supply Management
CB	=	U.S. Department of Commerce: Census Bureau
FRB	=	Board of Governors of the Federal Reserve System
EIA	=	Energy Information Administration
WSJ	=	Wall Street Journal

F = Frequency	T = Transformation	SA	U = Units
Q = Quarterly	0 = None	0 = no	1000-P = Thousands of Persons
M = Monthly	2 = Δ	1 = yes	1000-U = Thousands of Units
D = Daily	3 = Δ log		BoC = Billions of Chained
			$-B = Dollars per Barrel

Table J2: Data Description and Data Treatment

N	Series ID	Definition	Unit	F	S	SA	T
1	GDPH	Real Gross Domestic Product	BoC 2012$	Q	BEA	1	3
2	GDYH	Real Gross Domestic Income	BoC 2012$	Q	BEA	1	3
3	FSH	Real Final Sales of Domestic Product	BoC 2012$	Q	BEA	1	3
4	IH	Real Gross Private Domestic Investment	BoC 2012$	Q	BEA	1	3
5	GSH	Real State & Local Consumption Expenditures & Gross Investment	BoC 2012$	Q	BEA	1	3
6	FRH	Real Private Residential Fixed Investment	BoC 2012$	Q	BEA	1	3
7	FNH	Real Private Nonresidential Fixed Investment	BoC 2012$	Q	BEA	1	3
8	MH	Real Imports of Goods & Services	BoC 2012$	Q	BEA	1	3
9	GH	Real Government Consumption Expenditures & Gross Investment	BoC 2012$	Q	BEA	1	3
10	XH	Real Exports of Goods & Services	BoC 2012$	Q	BEA	1	3
14	CH	Real Personal Consumption Expenditures	BoC 2012$	Q	BEA	1	3
11	CNH	Real Personal Consumption Expenditures: Nondurable Goods	BoC 2012$	Q	BEA	1	3
12	CSH	Real Personal Consumption Expenditures: Services	BoC 2012$	Q	BEA	1	3
13	CDH	Real Personal Consumption Expenditures: Durable Goods	BoC 2012$	Q	BEA	1	3
15	GFDIH	Real National Defense Gross Investment	BoC 2012$	Q	BEA	1	3
16	GFNIH	Real Federal Nondefense Gross Investment	BoC 2012$	Q	BEA	1	3
17	YPDH	Real Disposable Personal Income	BoC 2012$	Q	BEA	1	3
18	JI	Gross Private Domestic Investment Chain-type Price Index	2012=100	Q	BEA	1	3
19	JGDP	Gross Domestic Product Chain-type Price Index	2012=100	Q	BEA	1	3
20	LXNFU	Unit Labor Cost (Nonfarm Business Sector)	2012=100	Q	BLS	1	3
21	LXNFR	Real Compensation Per Hour (Nonfarm Business Sector)	2012=100	Q	BLS	1	3
22	LXNFC	Compensation Per Hour (Nonfarm Business Sector)	2012=100	Q	BLS	1	3
23	LXNFH	Hours of All Persons (Nonfarm Business Sector)	2012=100	Q	BLS	1	3
24	LXNFA	Output Per Hour of All Persons (Nonfarm Business Sector)	2012=100	Q	BLS	1	3
25	LXMU	Unit Labor Cost (Manufacturing)	2012=100	Q	BLS	1	3
26	LXMR	Real Compensation Per Hour (Manufacturing)	2012=100	Q	BLS	1	3
27	LXMC	Compensation Per Hour (Manufacturing)	2012=100	Q	BLS	1	3
28	LXMH	Hours of All Persons (Manufacturing)	2012=100	Q	BLS	1	3
29	LXMA	Output Per Hour of All Persons (Manufacturing)	2012=100	Q	BLS	1	3
30	IP	Industrial Production Index	2012=100	M	FRB	1	3
31	IP521	Industrial Production: Business Equipment	2012=100	M	FRB	1	3
32	IP511	Industrial Production: Durable Consumer Goods	2012=100	M	FRB	1	3
33	IP531	Industrial Production: Durable Materials	2012=100	M	FRB	1	3
34	IP512	Industrial Production: Nondurable Consumer Goods	2012=100	M	FRB	1	3
35	IP532	Industrial Production: nondurable Materials	2012=100	M	FRB	1	3
36	PCU	CPI-U: All Items	82-84=100	M	BLS	1	3
37	PCUSE	CPI-U: Energy	82-84=100	M	BLS	1	3
38	PCUSLFE	CPI-U: All Items Less Food and Energy	82-84=100	M	BLS	1	3
39	PCUFO	CPI-U: Food	82-84=100	M	BLS	1	3

Table J3: Data Description and Data Treatment

N	Series ID	Definition	Unit	F	S	SA	T
40	JCBM	PCE: Chain Price Index	2012=100	M	BEA	1	3
41	JCEBM	PCE: Energy Goods & Services-price index	2012=100	M	BEA	1	3
42	JCNFOM	PCE: Food & Beverages-price index Purchased for Off-Premises Consumption	2012=100	M	BEA	1	3
43	JCXFEBM	PCE less Food & Energy-price index	2012=100	M	BEA	1	3
44	JCSBM	PCE: Services-price index	2012=100	M	BEA	1	3
45	JCDBM	PCE: Durable Goods-price index	2012=100	M	BEA	1	3
46	JCNBM	PCE: Nondurable Goods-price index	2012=100	M	BEA	1	3
47	PC1	PPI: Intermediate Demand Processed Goods	1982=100	M	BLS	1	3
48	P05	PPI: Fuels and Related Products and Power	1982=100	M	BLS	0	3
49	SP3000	PPI: Final Demand Personal Consumption Gds [Finished Consumer Gds]	1982=100	M	BLS	1	3
50	SP3000	PPI: Finished Goods	1982=100	M	BLS	1	3
51	PIN	PPI: Industrial Commodities	1982=100	M	BLS	0	3
52	PA	PPI: All Commodities	1982=100	M	BLS	0	3
53	FMC	Money Stock: Currency	Bil. of $	M	FRB	1	3
54	FM1	Money Stock: M1	Bil. of $	M	FRB	1	3
55	FM2	Money Stock: M2	Bil. of $	M	FRB	1	3
56	FABWC	C & I Loans in Bank Credit All Commercial Banks	Bil. of $	M	FRB	1	3
57	FABWQ	Consumer Loans in Bank Credit All Commercial Banks	Bil. of $	M	FRB	1	3
58	FAB	Bank Credit All Commercial Banks	Bil. of $	M	FRB	1	3
59	FABW	Loans & Leases in Bank Credit All Commercial Banks	Bil. of $	M	FRB	1	3
60	FABYO	Other Securities in Bank Credit All Commercial Banks	Bil. of $	M	FRB	1	3
61	FABWR	Real Estate Loans in Bank Credit All Commercial Banks	Bil. of $	M	FRB	1	3
62	FOT	Consumer Credit Outstanding	Bil. of $	M	FRB	1	3
63	HSTMW	Housing Starts: Midwest	1000-U	M	CB	1	3
64	HSTNE	Housing Starts: Northeast	1000-U	M	CB	1	3
65	HSTS	Housing Starts: South	1000-U	M	CB	1	3
66	HSTGW	Housing Starts: West	1000-U	M	CB	1	3
67	HPT	Building Permit New Private Housing Units Authorized by	1000-U	M	CB	1	3
68	FBPR	Bank Prime Loan Rate	Percent	M	FRB	0	1
69	FFED	Federal Funds [effective] Rate	Percent	M	FRB	0	1
70	FCM1	1-Year Treasury Bill Yield at Constant Maturity	Percent	M	FRB	0	1
71	FCM10	10-Year Treasury Note Yield at Constant Maturity	Percent	M	FRB	0	1
72	LP	Civilian Participation Rate: 16 yr +	Percent	M	BLS	0	2
73	LQ	Civilian Employment/Population Ratio: 16 yr +	Percent	M	BLS	0	2
74	LE	Civilian Employment: Sixteen Years & Over	1000-P	M	BLS	0	3
75	LR	Civilian Unemployment Rate: 16 yr +	Percent	M	BLS	0	2
76	LU0	Civilians Unemployed for Less Than 5 Weeks	1000-P	M	BLS	0	3
77	LU5	Civilians Unemployed for 5-14 Weeks	1000-P	M	BLS	0	3
78	LU15	Civilians Unemployed for 15-26 Weeks	1000-P	M	BLS	0	3
79	LUT27	Civilians Unemployed for 27 Weeks and Over	1000-P	M	BLS	0	3
80	LUAD	Average [Mean] Duration of Unemployment	Weeks	M	BLS	0	3
81	LANAGRA	All Employees: Total Nonfarm	1000-P	M	BLS	0	3
82	LAPRIVA	All Employees: Total Private Industries	1000-P	M	BLS	0	3
83	LANTRMA	All Employees: Mining and Logging	1000-P	M	BLS	0	3
84	LACONSA	All Employees: Construction	1000-P	M	BLS	0	3
85	LAMANUA	All Employees: Manufacturing	1000-P	M	BLS	0	3
86	LATTULA	All Employees: Trade, Transportation & Utilities	1000-P	M	BLS	0	3
87	LAINFOA	All Employees: Information Services	1000-P	M	BLS	0	3
88	LAFIREA	All Employees: Financial Activities	1000-P	M	BLS	0	3
89	LAPBSVA	All Employees: Professional & Business Services	1000-P	M	BLS	0	3
90	LAEDUHA	All Employees: Education & Health Services	1000-P	M	BLS	0	3
91	LALEIHA	All Employees: Leisure & Hospitality	1000-P	M	BLS	0	3
92	LASRVOA	All Employees: Other Services	1000-P	M	BLS	0	3
93	LAGOVTA	All Employees: Government	1000-P	M	BLS	0	3
94	LAFGOVA	All Employees: Federal Government	1000-P	M	BLS	0	3
95	LASGOVA	All Employees: State Government	1000-P	M	BLS	0	3
96	LALGOVA	All Employees: Local Government	1000-P	M	BLS	0	3
97	PETEXA	West Texas Intermediate Spot Price FOB, Cushing, Oklahoma	$-B	M	EIA	0	3
98	NAPMNI	ISM Mfg: New Orders Index	Index	M	ISM	1	2
99	NAPMOI	ISM Mfg: Production Index	Index	M	ISM	1	2
100	NAPMEI	ISM Mfg: Employment Index	Index	M	ISM	1	2
101	NAPMVDI	ISM Mfg: Supplier Deliveries Index	Index	M	ISM	1	2
102	NAPMII	ISM Mfg: Inventories Index	Index	M	ISM	1	2
103	SP500	Standard & Poor's 500 Stock Price Index	41-43=10	D	WSJ	0	3

References

Akaike, H. (1973). Block Toeplitz Matrix Inversion. *SIAM Journal on Applied Mathematics 24*(2), 234–241.

Anderson, B. D. O. and J. B. Moore (1979). *Optimal Filtering.* Dover Publications, Inc.

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

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.

Bakhshizadeh, M., A. Maleki, and V. H. de la Pena (2023). Sharp concentration results for heavy-tailed distributions. *Information and Inference: A Journal of the IMA 12*, 1655–1685.

Barigozzi, M. (2023). Asymptotic equivalence of principal component and quasi maximum likelihood estimators in large approximate factor models. Technical Report arXiv:2307.09864.

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.

Booth, J. G. and J. P. Hobert (1999). Maximizing generalized linear mixed model likelihoods with an automated Monte Carlo EM algorithm. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 61*, 265–285.

Bosq, D. (2012). *Nonparametric statistics for stochastic processes: estimation and prediction.* Springer Science & Business Media.

Bradley, R. C. (2005). Basic properties of strong mixing conditions. a survey and some open questions. *Probability Surveys 2*, 107–144.

Chan, S., G. C. Goodwin, and K. Sin (1984). Convergence properties of the Riccati difference equation in optimal filtering of nonstabilizable systems. *IEEE Transactions on Automatic Control 29*, 110–118.

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

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.

Durbin, J. and S. J. Koopman (2012). *Time Series Analysis by State Space Methods.* Oxford University Press.

Fan, J., Y. Liao, and M. Mincheva (2011). High dimensional covariance matrix estimation in approximate factor models. *The Annals of Statistics 39*, 3320.

Forni, M., D. Giannone, M. Lippi, and L. Reichlin (2009). Opening the black box: Structural factor models versus structural VARs. *Econometric Theory 25*, 1319–1347.

Gourieroux, C. and A. Monfort (1995). *Statistics and Econometric Models*, Volume 1. Cambridge University Press.

Gray, R. M. (2006). Toeplitz and circulant matrices: A review. *Foundations and Trends® in Communications and Information Theory 2*, 155–239.

Hamilton, J. D. (1994). *Time Series Analysis.* Princeton University Press.

Harvey, A. C. (1990). *Forecasting, structural time series models and the Kalman filter.* Cambridge University Press.

Harvey, A. C. and D. Delle Monache (2009). Computing the mean square error of unobserved components extracted by misspecified time series models. *Journal of Economic Dynamics and Control 33*, 283–295.

Henderson, H. V. and S. R. Searle (1981). On deriving the inverse of a sum of matrices. *SIAM Review 23*, 53–60.

Ibragimov, I. A. (1962). Some limit theorems for stationary processes. *Theory of Probability and its Applications 7*, 349–382.

Kuchibhotla, A. K. and A. Chakrabortty (2022). Moving beyond sub-Gaussianity in high-dimensional statistics: Applications in covariance estimation and linear regression. *Information and Inference: A Journal of the IMA 11*, 1389–1456.

Marshall, A. W., I. Olkin, and B. C. Arnold (2011). *Inequalities: Theory of Majorization and Its Applications.* Springer-Verlag New York.

McLachlan, G. and T. Krishnan (2007). *The EM algorithm and extensions*, Volume 382. John Wiley & Sons.

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.

Merikoski, J. K. and R. Kumar (2004). Inequalities for spreads of matrix sums and products. *Applied Mathematics E-Notes 4*, 150–159.

Merlevède, F., M. Peligrad, and E. Rio (2011). A Bernstein type inequality and moderate deviations for weakly dependent sequences. *Probability Theory and Related Fields 151*, 435–474.

Pham, T. D. and L. T. Tran (1985). Some mixing properties of time series models. *Stochastic processes and their applications 19*, 297–303.

Rosenblatt, M. (1972). Uniform ergodicity and strong mixing. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete 24*(1), 79–84.

Stout, W. F. (1974). *Almost Sure Convergence.* Academic press.

Sundberg, R. (1974). Maximum likelihood theory for incomplete data from an exponential family. *Scandinavian Journal of Statistics 1*, 49–58.

Sundberg, R. (1976). An iterative method for solution of the likelihood equations for incomplete data from exponential families. *Communication in Statistics-Simulation and Computation 5*, 55–64.

Sundberg, R. (2019). *Statistical modelling by exponential families.* Cambridge University Press.

White, H. (2001). *Asymptotic Theory for Econometricians.* Academic press.

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