Consistent Estimation of Time-Varying Loadings in High-Dimensional Factor Models

Jakob Guldbæk Mikkelsen,* Eric Hillebrand,† and Giovanni Urga‡

November 1, 2018

Abstract

In this paper, we develop a two-step maximum likelihood estimator of time-varying loadings in high-dimensional factor models. We specify the loadings to evolve as stationary vector autoregressions (VAR) and show that consistent estimates of the loadings parameters can be obtained. In the first step, principal components are extracted from the data to form factor estimates. In the second step, the parameters of the loadings VARs are estimated as a set of linear regression models with time-varying coefficients. We document the finite-sample properties of the maximum likelihood estimator through an extensive simulation study and illustrate the empirical relevance of the time-varying loadings structure using a large quarterly dataset for the US economy.

*Danmarks Nationalbank, Havnegade 5, DK-1093 Copenhagen K, E-mail: jgm@nationalbanken.dk.

†Corresponding author. CREATES, Department of Economics and Business Economics, Aarhus University, Fuglesangs Allé 4, DK-8210 Aarhus V, Denmark. E-mail: ehillebrand@econ.au.dk.

‡Centre for Econometric Analysis, Faculty of Finance, Cass Business School, 106 Bunhill Row, London EC1Y 8TZ (U.K.) and Bergamo University (Italy). E-mail: g.urga@city.ac.uk.

1 Introduction

In this paper, we develop a two-step estimator of time-varying loadings in high-dimensional factor models, where factors are estimated with principal components. We show that this estimator maximizes the infeasible likelihood where the factors are unobserved, since the feasible likelihood, using principal components, converges uniformly to the infeasible likelihood.

The problem of time-varying loadings in factor models is important because the assumption of constant loadings has been found to be implausible in a number of studies considering structural instability in factor models. In a large macroeconomic dataset for the U.S., Stock and Watson (2009) find considerable instability in factor loadings around 1984, and they improve factor-based forecast regressions of individual variables by allowing factor coefficients to change after the break point. Breitung and Eickmeier (2011) develop Chow-type tests for structural breaks in factor loadings and find similar evidence of structural instability around 1984. They also find evidence of structural breaks in the Euro area around 1992 and 1999. Del Negro and Otrok (2008), Liu, Mumtaz, and Theophilopoulou (2011), and Eickmeier, Lemke, and Marcellino (2015) estimate factor models where the factor loadings are modelled as random walks using large panels of data, but theoretical results for models with time-varying parameters in a high-dimensional setting are scant.

The econometric theory on factor models explicitly addresses the high dimensionality of these datasets by developing results in a large N and large T framework. The central results in the literature on consistent estimation of the factor space by principal components as $N, T \to \infty$ have been developed in Stock and Watson (1998, 2002), and Bai and Ng (2002). Forni, Hallin, Lippi, and Reichlin (2000) consider estimation in the frequency domain. Principal components have the advantage of being easy to compute and feasible even when the cross-sectional dimension N is larger than the sample size T. Bates, Plagborg-Møller, Stock, and Watson (2013) characterize the types and magnitudes of structural instability in factor loadings under which the principal components estimator of the factor space is consistent.

Another strand of literature is concerned with estimation by maximum likelihood. Doz, Giannone, and Reichlin (2012) study the asymptotic properties of the maximum likelihood estimator of a factor model in a large N, T setting. The likelihood is evaluated assuming VAR dynamics for the factors and constant loadings using the Kalman filter. Bai and Li (2012) study the asymptotic theory of maximum likelihood estimation of the factor model for large N, T as well, and they explore the consequences of different identifying assumptions. In their setup, the factors are a sequence of fixed constants, and the loadings are constant. Bai and Li (2016) extend this analysis to non-diagonal covariance matrices and serial correlation of the error in the measurement equation. The factors are again assumed to be a sequence of constants and the loadings are constant as well.

We consider a factor model of the form $X_{it} = \lambda'_{it} F_t + e_{it}$ for $i = 1, \dots N$ and $t = 1, \dots, T$, where the data X_{it} depend on a small number $r \ll N$ of unobserved common factors F_t. The $r \times 1$ vector of factor loadings λ_{it} evolves over time. We model λ_{it} for each i as a stationary vector autoregression, and our main contribution is to show that the parameters of these time-varying loadings can be consistently estimated by maximum likelihood. Our estimation procedure consists of two steps. In the first step, the common factors are estimated by principal components, and in the second step we estimate the loadings parameters by maximum likelihood, treating the principal components as observed data.

The principal components estimator is robust to stationary variations in the loadings. By averaging over the cross-section, the temporal instabilities in the loadings are smoothed out and the factor space is consistently estimated. Mean squared consistency of the factor space is shown by Bates et al. (2013), and we extend the result to uniform consistency in t to analyse the maximum likelihood estimator.

In the second step, we estimate a panel of regression models with time-varying coefficients. By treating the principal components as observable regressors, the loadings parameters can be estimated as a set of N regression models with time-varying coefficients. Under the condition that $\frac{T}{N^2} \to 0$, the maximum likelihood estimator of the time-varying loadings is consistent as $N, T \to \infty$, and estimation error from the principal components does not affect the consistency of the estimator.

The computation of the maximum likelihood estimator is relatively simple. Principal components are easily available, and the set of N regression models with time-varying parameters can be readily estimated by Kalman-filter procedures.

The rest of the paper is organized as follows. Section 2 introduces the model and the two-step estimation procedure. Section 3.1 states the assumptions and consistency results for the principal components estimator, and Section 3.2 discusses identification of the loadings parameters. Our main result on consistency of the maximum likelihood estimator of the time-varying loadings and the associated assumptions are stated in Section 3.3. In Section 4 we report the results of a Monte Carlo study, and in Section 5 we provide an empirical illustration. Section 6 concludes.

2 Model and estimation

We consider the following model:

$$X_t = \Lambda_t F_t + e_t, \tag{1}$$

where $X_t = (X_{1t}, ..., X_{Nt})'$ is the N-dimensional vector of observed data at time t. The observations are generated by a small number $r \ll N$ of unobserved common factors $F_t = (F_{1t}, ..., F_{rt})'$, time-varying factor loadings $\Lambda_t = (\lambda_{1t}, ..., \lambda_{Nt})'$, and idiosyncratic errors $e_t = (e_{1t}, ..., e_{Nt})$ with covariance matrix $E(e_t e_t') = \Psi_0$. The $N \times r$ loadings matrix $\Lambda_t = (\lambda_{1t}, ..., \lambda_{Nt})'$ is time-varying and each $\lambda_{it} \in \mathbb{R}^{r\times 1}$ evolves as an r-dimensional vector autoregression:

$$B_i^0(L)(\lambda_{it} - \lambda_i^0) = \eta_{it}, \tag{2}$$

where $\lambda_i^0 = E(\lambda_{it})$ is the unconditional mean, and $B_i^0(L) = I - B_{i,1}^0 L - ... - B_{i,p}^0 L^p$ is a p^{th}-order lag polynomial where the roots of $|B_i^0(L)|$ are outside the unit circle. The autoregressive order p can be allowed to vary over i such that p_i differs over i. We suppress the subscript for notational convenience. The innovations η_{it} have covariance matrix $E(\eta_{it}\eta_{it}') = Q_i^0$. Our goal is to estimate the parameters of each of the loadings processes (2) and the variance parameter of each of the idiosyncratic elements $E(e_{it}^2) = \psi_i^0$. We therefore write the model in terms of each X_i:

$$X_i = \mathbf{F}\Lambda_i + e_i, \tag{3}$$

where $X_i = (X_{i1}, ..., X_{iT})'$, $e_i = (e_{i1}, ..., e_{iT})'$, $\mathbf{F} = diag\left\{F_t'\right\}_{t=1,...,T}$ is a $T \times rT$ block-diagonal matrix, and $\Lambda_i = (\lambda_{i1}', ..., \lambda_{iT}')'$. The mean and variance of X_i are $E(X_i) = (F_1'\lambda_i^0, ..., F_T'\lambda_i^0)'$ and $\Sigma_i := Var(X_i) = \mathbf{F}\Phi_i\mathbf{F}' + \psi_i I_T$ where $\Phi_i = Var(\Lambda_i)$ is of dimension $rT \times rT$. Equation (3) is a regression model with time-varying coefficients. The factors F_t are the regressors, and the loadings λ_{it} are the time-varying coefficients. We can thus specify a Gaussian likelihood function for X_i conditional on the factors $F = (F_1, ..., F_T)'$ as:

$$\mathcal{L}_T(X_i|F;\theta_i) = -\frac{1}{2}\log(2\pi) - \frac{1}{2T}\log|\Sigma_i| - \frac{1}{2T}(X_i - E(X_i))'\Sigma_i^{-1}(X_i - E(X_i)), \tag{4}$$

with parameter vector $\theta_i = \left\{B_i(L), \lambda_i, Q_i, \psi_i\right\}$. Equations (2) and (3) can be written as a linear state-space model, and the likelihood can therefore be calculated with the Kalman filter.

It is not feasible to estimate θ_i with (4), however, as the likelihood depends on the unobservable factors F. We therefore replace the unobservable factors F in (4) with an estimate $\tilde{F}$ to form the feasible likelihood function $\tilde{\mathcal{L}}_T(X_i|\tilde{F};\theta_i)$. Define the estimator $\tilde{\theta}_i$ which maximizes the feasible likelihood function as:

$$\tilde{\theta}_i = \underset{\theta_i}{\operatorname{argmax}} \tilde{\mathcal{L}}_T(X_i|\tilde{F};\theta_i). \tag{5}$$

This is our object of interest, and we show that the estimator $\tilde{\theta}_i \xrightarrow{p} \theta_i^0$ for each i, where $\theta_i^0 = \left\{B_i^0(L), \lambda_i^0, Q_i^0, \psi_i^0\right\}$ is the true value of the parameters.

We use the principal components estimator to estimate the factors. The principal components estimator treats the loadings as being constant over time, $\Lambda_t \equiv \Lambda$, and solves the minimization problem:

$$(\tilde{F}, \tilde{\Lambda}) = \min_{F,\Lambda}(NT)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} (X_{it} - \lambda_i' F_t)^2, \tag{6}$$

where $\tilde{F}$ is $T \times r$ and $\tilde{\Lambda}$ is $N \times r$. To uniquely define the minimizers, it is necessary to impose identifying restrictions on the estimators, as only X_{it} is observed. By concentrating out Λ and using the normalization $F'F/T = I_r$, the problem is equivalent to maximizing $tr(F'(XX')F)$, where $X = (X_1, ..., X_T)'$ is the $T \times N$ matrix of observations. The resulting estimator $\tilde{F}$ is given by $\sqrt{T}$ times the eigenvectors corresponding to the r largest eigenvalues of the $T \times T$ matrix XX'. The solution is not unique: any orthogonal rotation of $\tilde{F}$ is also a solution. Bai and Ng (2008b) give an extensive treatment of the principal components estimator. We use $\tilde{F}$ to form the feasible likelihood function $\tilde{\mathcal{L}}_T(X_i|\tilde{F};\theta_i)$.

The estimation procedure thus consists of two steps. In the first step, we extract principal components from the observable data to estimate the factors F_t under the assumption of constant loadings. In the second step, we use the factor estimates together with the observable data to maximize the likelihood function and estimate the parameters of the time-varying loadings, θ_i , for each i. Our main result in Section 3.3 shows that this yields a consistent estimator for the parameters of the time-varying loadings.

3 Asymptotic theory

In this section, we present the asymptotic theory for the two-step estimation method discussed in Section 2. The main result is Theorem 1 on consistent estimation of the loadings parameters by maximum likelihood; it is given in Section 3.3. Our result builds on the work by Bates et al. (2013), who show mean squared consistency of the principal components estimator when loadings are subject to structural instability. We use a different rotation of the principal components estimator, and in Section 3.1 we restate their result in Lemma 1. Furthermore, we provide a result on uniform consistency in t of the principal components estimator in Proposition 1. Section 3.2 discusses identification of the factors and loadings parameters. All results are for $N, T \to \infty$, and the factor rank r is assumed to be known.

We introduce the following notation. $\|A\| = [\text{tr}(A'A)]^{1/2}$ denotes the Frobenius norm of the matrix A. The subscripts i, j are cross-sectional indices taking values from $1, ..., N$, the subscripts t, s are time indices taking values from $1, ..., T$, and p, q are factor indices taking values from $1, ..., r$. The constant $M \in (0, \infty)$ is a constant common to all the assumptions below. Finally, define $C_{NT} = \min\{\sqrt{N}, \sqrt{T}\}$.

3.1 Principal components estimation

Let $\xi_{it} := \lambda_{it} - \lambda_i^0 = B_i^0(L)^{-1}\eta_{it}$ be the loadings innovations and write (1) as:

$$X_t = \Lambda^0 F_t + \xi_t F_t + e_t,$$

where $\Lambda^0 = (\lambda_1^0, ..., \lambda_N^0)'$ and $\xi_t = (\xi_{1t}, ..., \xi_{Nt})'$ are the $N \times r$ matrices of loadings means and innovations, respectively. The vector ξ_{it} is the moving average representation of the loadings. The following Assumptions A-C are standard for factor models and are the same as Assumptions A-C in Bai and Ng (2002):

Assumption A (Factors). $E\|F_t\|^4 \leq M < \infty$, and $T^{-1}\sum_{t=1}^T F_t F_t' \xrightarrow{p} \Sigma_F$ for some $r \times r$ positive definite matrix Σ_F.

Assumption B (Loadings). $\|\lambda_i^0\| \leq M < \infty$, and $\|\Lambda^{0\prime}\Lambda^0/N - \Sigma_\Lambda\| \to 0$ for some positive definite matrix Σ_Λ.

Assumption C (Idiosyncratic Errors). There exists a positive constant $M < \infty$, such that for all N and T:

1. $E(e_{it}) = 0$, $E|e_{it}|^8 \leq M$.
2. $E(e_s' e_t/N) = E(N^{-1}\sum_{i=1}^N e_{is}e_{it}) = \gamma_N(s,t)$, $|\gamma_N(s,s)| \leq M$ for all s, and $T^{-1}\sum_{s,t=1}^T |\gamma_N(s,t)| \leq M$.
3. $E(e_{it}e_{jt}) = \tau_{ij,t}$ with $|\tau_{ij,t}| \leq |\tau_{ij}|$ for some τ_{ij} and for all t. In addition

$N^{-1}\sum_{i,j=1}^{N}|\tau_{ij}| \le M$.

4. $E(e_{it}e_{js}) = \tau_{ij,ts}$, and $(NT)^{-1}\sum_{i,j=1}^{N}\sum_{t,s=1}^{T}|\tau_{ij,ts}| \le M$.

5. For every (s,t), $E|N^{-1/2}\sum_{i=1}^{N}[e_{is}e_{it} - E(e_{is}e_{it})]|^4 \le M$.

Under Assumption A, the factors are allowed to be dynamic such that they follow a VAR: $A(L)F_t = u_t$. Assumption A, however, allows more general dynamics for the factors, i.e. they do not need to be stationary. Assumption B requires the columns of Λ^0 to be linearly independent, such that the matrix Σ_Λ is non-singular. Assumptions A and B together imply the existence of r common factors. Assumption C allows for heteroskedasticity and limited time-series and cross-section dependence in the idiosyncratic errors. Note that if e_{it} is independent for all i and t, Assumptions C.2-C.5 follow from C.1.

We impose the following assumption on the factor loadings innovations and the factors, which are from Bates et al. (2013):

Assumption D (Factor Loadings Innovations). The following conditions hold for all N, T and factor indices $p_1, q_1, p_2, q_2 = 1, ..., r$:

1. $\sup_{s,t\le T} N^{-1}\sum_{i,j=1}^{N}|E(\xi_{isp_1}\xi_{jtq_1}F_{sp_1}F_{sq_1})| \le M$.

2. $N^{-1}T^{-2}\sum_{s,t=1}^{T}\sum_{i,j=1}^{N}|E(\xi_{isp_1}\xi_{jsq_1}F_{sp_1}F_{sq_1}F_{tp_2}F_{tq_2})| \le M$.

3. $N^{-2}T^{-2}\sum_{s,t=1}^{T}\sum_{i,j=1}^{N}|E(\xi_{isp_1}\xi_{jsq_1}\xi_{itp_2}\xi_{jtq_2}F_{sp_1}F_{sq_1}F_{tp_2}F_{tq_2})| \le M$.

Assumption D limits the degree of cross-sectional dependence of factors and loadings, but does not require full independence. The effect of the factors on the observable variables might reasonably change when the factors differ substantially from their mean levels. However, if the factors and loadings are assumed to be independent, and the loadings evolve as stationary vector autoregressions that are independent over i, Assumptions D.1-D.3 can easily be shown to hold. For simplicity, take $r = 1$. By Assumption A and independence of the loadings, Assumptions D.1 can be bounded by:

$$\sup_{s,t}\{|E(F_sF_t)|\sum_{i,j=1}^{N}|E(\xi_{is}\xi_{jt})|\} \le M\sum_{i=1}^{N}\sup_{s,t}|E(\xi_{is}\xi_{it})|$$

The terms $E(\xi_{is}\xi_{it})$ are the autocovariances of the moving average representation of the loadings. As the loadings are stationary, these autocovariances are bounded, and the rate $O(N)$ is obtained.

The rate $O(NT^2)$ in Assumption D.2 easily follows from D.1 when the factors and the loadings are independent. Assumption 4.3 can be bounded by:

$$M\sum_{s,t=1}^{T}\sum_{i,j=1}^{N}|E(\xi_{is}\xi_{js}\xi_{it}\xi_{jt})| = M\sum_{s,t=1}^{T}\sum_{i=1}^{N}|E(\xi_{is}^2\xi_{it}^2)|$$
$$+M\sum_{s,t=1}^{T}\sum_{i\neq j}^{N}|E(\xi_{is}\xi_{it})E(\xi_{js}\xi_{jt})| \leq M\sum_{s,t=1}^{T}\sum_{i=1}^{N}|E(\xi_{is}^4)^{1/2}E(\xi_{it}^4)^{1/2}|$$
$$+M\sum_{i\neq j}^{N}\sum_{t=1}^{T}\left(\sum_{s=1}^{T}|E(\xi_{is}\xi_{it})|^2\right)^{1/2}\left(\sum_{s=1}^{T}E|(\xi_{js}\xi_{jt})|^2\right)^{1/2}.$$

The first term is $O(NT^2)$ if $E(\xi_{is}^4)<\infty$, and the second term is $O(N^2T)$ if the autocovariances $E(\xi_{is}\xi_{it})$ are square-summable. Assumption D.3 is therefore satisfied when the loadings and the factors are independent. We assume the same rates to hold without imposing independence between the factors and the loadings.

Finally, we impose independence between the idiosyncratic errors and the factors and loadings innovations and a moment condition on their products.

Assumption E (Independence). For all (i,j,s,t), the scalar idiosyncratic errors e_{it} are independent of the factor and loading vectors (F_s,ξ_{js}). For a small positive number u and all (i,s,t), the random variable $z_{ist}:=e_{is}w_{it}$, where $w_{it}=\xi_{it}'F_t$, has the property

$$E\left(\exp\left(uT^{-1}\sum_{s=1}^{T}\|N^{-1/2}\sum_{i=1}^{N}z_{ist}\|^2\right)\right)\leq M.$$

Assumptions A-E are sufficient to consistently estimate the space spanned by the factors. For this purpose, we use the result of Lemma 1 below, which is a modified version of Theorem 1 in Bates et al. (2013).[1] We use a rescaled estimator that is more convenient for the rest of the analysis and therefore restate their result:

Lemma 1. *Under Assumptions A-E there exists an $r\times r$ matrix H such that*

$$T^{-1}\sum_{t=1}^{T}\|\tilde{F}_t-H'F_t\|^2=O_p(C_{NT}^{-2})$$

[1] Bates et al. (2013) use the estimator $\hat{F}=\tilde{F}V_{NT}$, where V_{NT} is the diagonal matrix of the r largest eigenvalues of $(NT)^{-1}XX'$.

as $N, T \to \infty$.

Proof. See the Appendix.

Lemma 1 shows that the mean-squared deviation between the principal components and the common factors disappears as the sample size T and the cross-sectional dimension N tend to infinity.[2] The convergence rate C_{NT} is the same as in Bai and Ng (2002), and the principal components estimator is thus robust to stationary deviations in the loadings around a constant mean. Note that the common factors are only identified up to a rotation, so the principal components converge to a rotation of the common factors.

The convergence rate is central to other results in the literature. Stock and Watson (2002) show that estimated factors can be used in diffusion index forecasting to obtain consistent forecasts. $FAVARs$, introduced by Bernanke, Boivin, and Eliasz (2005), use factor estimates to model the joint dynamics of a vector of observable variables, Y_t, and unobserved factors, F_t. Inferential theory for diffusion index forecasting and $FAVARs$ using PC estimates of the factors is given in Bai and Ng (2006). Doz, Giannone, and Reichlin (2011) also show consistent estimation of factor dynamics by regressing the PC estimates on its own past.

Lemma 1 does not imply uniform convergence in t, but only mean squared consistency of the principal components. In order to analyse the properties of the feasible likelihood function $\tilde{\mathcal{L}}_T(X_i, \tilde{F}|\theta_i)$, we need uniform consistency of the estimated factors, in addition to the mean squared consistency of Lemma 1. To establish uniform convergence, we make additional moment assumptions:

Assumption F (Uniform consistency) There exists a positive constant $M < \infty$ such that for all N and T and factor indices $p_1, q_1, p_2, q_2 = 1, ..., r$:

1. $\sum_{s=1}^{T} |\gamma_N(s,t)| \leq M$ for all t.
2. $E\|(NT)^{-1/2} \sum_{s=1}^{T} \sum_{k=1}^{N} F_s[e_{ks}e_{kt} - E(e_{ks}e_{kt})]\|^2 \leq M$ for all t.
3. $E\|N^{-1/2} \sum_{i=1}^{N} \lambda_i^0 e_{it}\|^8 \leq M$ for all t.
4. $E\left(\exp\left(\frac{u}{N} \sum_{i,j=1}^{N} \xi_{isp_1}\xi_{jtq_1}F_{sp_1}F_{tq_1}\right)\right) \leq M$ for all s, t and for u sufficiently small.

[2] Lemma 1 also holds when the factor rank is unknown. By setting the number of estimated factors to any fixed $m \geq 1$, the Lemma can be stated as $T^{-1}\sum_{t=1}^{T} \|\tilde{F}_t^m - H^{m\prime}F_t\|^2 = O_p(C_{NT}^{-2})$, where $\tilde{F}_t^m$ is a vector of dimension $m \times 1$ and H^m is a $r \times m$ matrix, and $\tilde{F}_t^m$ consistently estimates the space spanned by m of the true factors

5. $E\left(\exp\left(\frac{u}{N^2+NT}\sum_{s=1}^{T}\sum_{i,j=1}^{N}\xi_{isp_1}\xi_{itq_1}\xi_{jsp_2}\xi_{jtq_2}F_{tp_1}F_{tq_1}F_{sp_2}F_{sq_2}\right)\right) \leq M$ for all t and for u sufficiently small.

Assumptions F.1-F.3 are from Bai and Ng (2006, 2008a). Assumption F.1 is stronger than Assumption C.2, but still reasonable: If e_{it} is assumed to be stationary with absolutely summable autocovariances, Assumption F.1 holds. Assumptions F.2 and F.3 are reasonable as they involve zero-mean random variables. The moment conditions in Assumptions F.4 and F.5 are needed to obtain uniform convergence of principal components with time-varying loadings. They ensure that the summands in F.4 and F.5 are not too heavy-tailed, and can be shown to hold for simple cases such as constant factors and independent sub-exponential loadings. For example, if we consider the case of one constant factor equal to one and time-varying loadings ξ_{it}, $i.i.d. \sim N(0,\sigma^2)$, $i=1,\ldots,N$, $t=1,\ldots,T$, then Assumption F.4 amounts to $E[\exp(u/N\sum_{ij}^{N}\xi_i\xi_j)] = E[\exp(uZ^2)] < \infty$ for $Z=\sum_i^N \xi_i/\sqrt{N}$ and u small enough. Following from results such as, for example, Proposition 5.16 in Vershynin (2012)[3], which states that

$$P\left(\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\xi_{it}\right| \geq x\right) \leq 2e^{-cx^2},$$

where c is a positive real constant, we get

$$P\left(\exp(uZ^2)\geq z\right) = P\left(|Z| \geq \left(\frac{1}{u}\log z\right)^{\frac{1}{2}}\right) \leq 2z^{-\frac{c}{u}}, \quad \text{for } z>1,$$

which is integrable if $u<c$:

$$E(\exp(uZ^2)) = \int_1^{\infty} P\left(\exp(uZ^2)\geq z\right)dz \leq 2\int_1^{\infty} z^{-\frac{c}{u}}dz < \infty.$$

In the form Assumptions F.4 and F.5 are stated, they also allow for a degree of serial and cross-sectional correlation.

The following proposition extends the mean squared consistency result of Bates et al. (2013) to uniform consistency.

[3] Proposition 5.16 in Vershynin (2012) holds for independent variables. Theorem 1 in Doukhan and Neumann (2007) proves a similar inequality for weakly dependent variables.

Proposition 1. *Under Assumptions A-F and additionally if* $\max_t \|F_t\| = O_p(\alpha_T)$,

$$\max_t \|\tilde{F}_t - H'F_t\| = O_p\left(\frac{T^{1/8}}{N^{1/2}}\right) + O_p(\alpha_T N^{-1/2}) + O_p(\alpha_T T^{-1}) + O_p\left(\frac{log(T)}{N}\right)^{1/2} + O_p\left(\frac{log(T)}{T}\right)^{1/2}.$$

as $N, T \to \infty$.

Proof. See the Appendix.

Proposition 1 shows that the maximum deviation between the factors and the principal components depends on α_T. The convergence rate thus depends on the assumption imposed on $\max_t \|F_t\|$. The factors are allowed to display arbitrary dynamics under Assumption A. However, if the parameters governing these dynamics are not of direct interest, nothing is lost by assuming the factors to be a sequence of fixed and bounded constants, i.e. $\max_t \|F_t\| \leq M$.[4] For the purpose of estimating the loadings parameters, it is not needed to model the dynamics of the factors, so we can take $O_p(\alpha_T)$ to be $O(1)$ in our results. However, Proposition 1 is of independent interest, so we state Proposition 1 in its more general form. Bai (2003) and Bai and Ng (2008a) derive a similar result for factor models with constant loadings. Uniform convergence when loadings undergo small variations is also considered by Stock and Watson (1998), who obtain a slower convergence rate and require $T = o(N^{1/2})$.

3.2 Identification

It is well known that without identifying restrictions, factors and loadings are not separately identified in (1). The common component $C_t = \Lambda_t F_t$ is identified, but normalizations are needed to separate factor and loadings from the common component. This has implications for the identification of the loadings parameters as well, which we now illustrate. The model defined by (1) and (2) is observationally equivalent to:

$$X_t = \Lambda_t H'^{-1} H' F_t + e_t,$$
$$B_i(L) H^{-1}(\lambda_{it} - \lambda_i) = H^{-1}\eta_{it}, \qquad for \quad i = 1, \ldots, N,$$

for any invertible matrix H. Lemma 1 states that the principal components estimator $\tilde{F}_t$ is a consistent estimate of a rotation of the true factors, $H'F_t$. The two-step estimation procedure

[4] Bai and Li (2012, 2016) treat the factors as a sequence of fixed constants when providing inferential theory for maximum likelihood estimation of factor models with constant loadings.

fixes the rotational indeterminacy by imposing the normalization in the principal components step. By replacing the unobserved factors F_t with $\tilde{F}_t$ for maximum likelihood estimation, we are thus estimating the parameters of $\lambda^*_{it} = H^{-1}\lambda_{it}$.

To clarify the issue, consider the following example. Using the same notation as above, the elements of the $r \times 1$ vector $\lambda_{it} = (\lambda_{it,1} ..., \lambda_{it,r})'$ refer to the loadings of variable i at time t on each of the r factors, and $\lambda_i = E(\lambda_{it}) = (\lambda_{i,1}, ..., \lambda_{i,r})'$ are the corresponding unconditional expectations of the factor loadings. Assume that the matrices Σ_F and Σ_Λ are diagonal. In this case it is not hard to show that the rotation matrix H converges to $\Sigma_F^{-1/2}$. Let the number of factors $r = 2$ with variance $\Sigma_F = diag(\sigma_1^2, \sigma_2^2)$ and let the data-generating parameters of the loadings be

$$\lambda_i^0 = \begin{pmatrix} \lambda_{i,1} \\ \lambda_{i,2} \end{pmatrix}, \quad Q_i^0 = \begin{pmatrix} q_{i,1} & 0 \\ 0 & q_{i,2} \end{pmatrix}, \quad B_i^0(L) = I_2 - \begin{pmatrix} b_{i,11} & 0 \\ 0 & b_{i,22} \end{pmatrix}.$$

We can now make precise what role the rotation matrix H plays for estimation of θ_i. With the normalization $\tilde{F}'\tilde{F}/T = I_2$, the principal components will be close to $\Sigma_F^{-1/2}F_t$ in large samples. Using the principal components in place of the unobserved factors means that we are estimating the following model:

$$\begin{aligned} X_t &= \Lambda_t^* \tilde{F}_t + e_t, \\ \lambda^*_{it} - \lambda^*_i &= B^*_i(\lambda^*_{i,t-1} - \lambda^*_i) + v_{it}, \qquad for \quad i = 1, ..., N, \end{aligned}$$

where $\lambda^*_{it} = \Sigma_F^{1/2}\lambda_{it} = \begin{pmatrix} \sigma_1 \lambda_{it,1} \\ \sigma_2 \lambda_{it,2} \end{pmatrix}$ and $v_{it} = \Sigma_F^{1/2}\eta_{it}$. The loadings λ_{it} are scaled by the standard deviations of the unobserved factors, and it is the parameters of the rotated loadings λ^*_{it} that can be estimated. In large samples the estimate of the loadings mean λ^*_i will be therefore close to

$$\Sigma_F^{1/2}\lambda_i^0 = \begin{pmatrix} \sigma_1 \lambda_{i,1} \\ \sigma_2 \lambda_{i,2} \end{pmatrix},$$

and the variance estimate $Var(v_{it}) = Var(\Sigma_F^{1/2}\eta_{it})$ will be close to

$$\Sigma_F^{1/2} Q_i^0 \Sigma_F^{1/2} = \begin{pmatrix} \sigma_1^2 q_{i,1} & 0 \\ 0 & \sigma_2^2 q_{i,2} \end{pmatrix}.$$

The mean and variance parameters are thus scaled by the standard deviation of the factors. The matrices $B_i(L)$ and Q_i^0 of the data-generating model are diagonal in this example, so the diagonal elements of B_i^* are the autocorrelations of $\lambda_{it,1}^*$ and $\lambda_{it,2}^*$. In large samples the first diagonal element of B_i^* will therefore be close to

$$
\begin{aligned}
b_{i,11}^* &= \frac{Cov(\lambda_{it,1}^*, \lambda_{i,t-1,1}^*)}{Var(\lambda_{it,1}^*)} = \frac{Cov(\sigma_1 \lambda_{it,1}, \sigma_1 \lambda_{i,t-1,1})}{Var(\sigma_1 \lambda_{it,1})} \\
&= \frac{\sigma_1^2 Cov(\lambda_{it,1}, \lambda_{i,t-1,1})}{\sigma_1^2 Var(\lambda_{it,1})} = \frac{Cov(\lambda_{it,1}, \lambda_{i,t-1,1})}{Var(\lambda_{it,1})} = b_{i,11},
\end{aligned}
$$

and similarly for $b_{i,22}^*$. The estimates of the autoregressive matrix B_i^* are therefore unaffected by the normalization imposed on the principal components, and the estimate of B_i^* is consistent for the autoregressive parameters B_i of the data-generating process λ_{it}.

The arguments of this example apply to the general setting as well. The maximum likelihood estimator (5) of the loadings parameters is estimating $B_i(L)$, $H\lambda_i$, and HQ_iH'. The mean and variance parameters of (2) are identified up to the unknown rotation matrix H, while the dynamic parameters $B_i(L)$ are not subject to any rotation. The rotation is determined by the restriction used to identify the principal components. Using another normalization in the first step will thus change the estimates of λ_i and Q_i, while the estimate of $B_i(L)$ is unaffected. The dynamic properties of the loadings are therefore uniquely identified. In the following, we assume for simplicity that $H = I_r$. This is just a normalization and can be achieved by imposing further assumptions on the matrices Σ_F and Σ_Λ.

3.3 Maximum likelihood estimation

Our method of proof relies on showing that the feasible likelihood function with principal components is asymptotically equivalent to the infeasible likelihood function (4) where the factors are observed. To establish our result, we impose distributional assumptions on the loadings and idiosyncratic errors that enable maximum likelihood estimation of the parameters $\theta_i = \{B_i(L), \lambda_i, Q_i, \psi_i\}$.

Assumption G (Distributions) For all $i = 1, ..., N$, it holds:

1. The loadings λ_{it} follow a finite-order Gaussian VAR:

$$B_i(L)(\lambda_{it} - \lambda_i) = \eta_{it},$$

with the $r \times r$ filter $B_i(L) = I - B_{i,1}L - ... - B_{i,p}L^p$ having roots outside the unit circle, and η_{it} is an r-dimensional Gaussian white noise process, $\eta_{it} \sim$ i.i.d. $\mathcal{N}(0, Q_i)$, where Q_i is positive definite with all elements bounded.

2. The idiosyncratic errors e_{it} are serially uncorrelated Gaussian white noise, $e_{it} \sim \mathcal{N}(0, \psi_i)$, with $\psi_i > 0$ and bounded for all i.

Assumption G.1 assumes the loadings to evolve as stationary vector autoregressions. We rule out the possibility of $I(1)$ loadings as this would be in violation of Assumption D. With non-stationary loadings the principal components estimator cannot consistently estimate the factor space.[5] Assumption G.2 assumes that the idiosyncratic errors to be serially uncorrelated. This assumption is made only for simplicity of presentation, and can be relaxed in a straightforward manner. We return to this in the discussion following Theorem 1 below. Note that it is not necessary to assume the loadings or the idiosyncratic errors to be independent over the cross-section dimension. The loadings parameters are estimated by regressions with time-varying parameters, and it is therefore sufficient to analyse the likelihood function for each X_i separately. The innovations η_{it} and e_{it} are assumed to be Gaussian such that the likelihood function is correctly specified. However, the distributions do need to be Gaussian. If the data generating process is non-Gaussian, the infeasible likelihood function (4) will be a quasi-likelihood function in the sense of White (1982).

With observed regressors, consistency is known to hold, see e.g. Pagan (1980). We summarize this result in Assumption H on the infeasible likelihood function.

Assumption H (MLE with observed factors) For each i:

1. There exists a function $\mathcal{L}_{0,i}(\theta_i)$ that is uniquely maximized at θ_i^0.

2. $\theta_i^0 \in \Theta_i$, which is compact.

3. $\mathcal{L}_{0,i}(\theta_i)$ is continuous at each $\theta_i \in \Theta_i$.

4. $\sup_{\theta_i \in \Theta_i} |\mathcal{L}_T(X_i|F;\theta_i) - \mathcal{L}_{0,i}(\theta_i)| \xrightarrow{p} 0$ for $T \to \infty$.

[5] Bates et al. (2013) consider random walk loadings of the form $\lambda_{it} = \lambda_{i,t-1} + T^{-3/4}\zeta_{it}$ and show that Assumption D is satisfied with this specification. However, the scaling of the loadings innovations by the factor $T^{-3/4}$ is crucial for Lemma 1 to hold. With a pure random walk of the form $\lambda_{it} = \lambda_{i,t-1} + \zeta_{it}$, principal components cannot estimate the factor space consistently.

The assumptions are standard for consistency and imply that the maximum likelihood estimator with observed factors $\hat{\theta}_i = \underset{\theta_i}{\operatorname{argmax}} \mathcal{L}_T(X_i|F;\theta_i)$ is consistent for each i: $\hat{\theta}_i \xrightarrow{p} \theta_i^0$. This follows from standard arguments as in Newey and McFadden (1994), Thm 2.1. If the data generating process is non-Gaussian, the infeasible estimator $\hat{\theta}_i$ is a quasi-maximum likelihood estimator and will be consistent for the pseudo-true value θ_i^*. Assumption H.2 restricts the parameters to be in a compact set, which is usually assumed for nonlinear models. The autoregressive parameters of the loadings are thus assumed to be bounded away from the non-stationary region.[6]

Replacing the unobserved factors with the principal components estimates yields the feasible likelihood function $\tilde{\mathcal{L}}_T(X_i|\tilde{F};\theta_i)$ and the maximum likelihood estimator defined in (5). We now state our main result.

Theorem 1. *Let Assumptions A-H hold and $T/N^2 \to 0$. Then, for each i, the estimator $\tilde{\theta}_i$ defined in (5) is consistent:*

$$\tilde{\theta}_i \xrightarrow{p} \theta_i^0.$$

Proof. See the Appendix.

Theorem 1 states that using the principal component estimates instead of the unobserved factors does not affect the consistency of the maximum likelihood estimator. The main argument in proving Theorem 1 is that the feasible likelihood function converges uniformly to the infeasible likelihood function. Asymptotically, the feasible likelihood function therefore has the same properties as the infeasible likelihood function, for which consistency is known to hold. Assumption H thus holds for $\tilde{\mathcal{L}}_T(X_i|\tilde{F};\theta_i)$ in the limit and consistency follows. With a misspecified likelihood function, the estimator is consistent for the pseudo-true value θ_i^*. The rate condition $T/N^2 \to 0$ ensures that $\max_t \|\tilde{F}_t - H'F_t\| = o_p(1)$. The rate condition is stronger than needed. For $\max_t \|\tilde{F}_t - H'F_t\|$ to be $o_p(1)$, the condition $\sqrt{T}/N^2 \to 0$ would suffice. We state Theorem 1 with the condition $T/N^2 \to 0$, as this rate is common in the factor literature.

In the proof of Theorem 1 we use the following normalization that is convenient for the calculations: If $F'F/T = I_r$ and $\Lambda^{0\prime}\Lambda^0$ is a diagonal matrix with distinct elements, we show in the Appendix that the rotation matrix H converges to the identity I_r. Lemma 1 and Proposition 1 then holds with H replaced by the identity matrix, and θ_i can be estimated asymptotically without rotation. Such normalizations are inconsequential for the results as H is asymptotically bounded,

[6] Pagan (1980) also rules out non-stationarity when proving identification of regression models with time-varying coefficents.

and they are only for ease of notation. Without such normalizations the feasible likelihood converges to $\mathcal{L}_T(X_i|FH;\theta_i)$ and $\tilde{\theta}_i$ is consistent for the parameters of the process $\lambda^*_{it} = H\lambda_{it}$ as discussed in Section 3.2.

We have assumed that the factors are estimated by the method of principal components. Note, however, that the proof of Theorem 1 does not rely on the principal components estimator. Theorem 1 holds for all estimators $\tilde{F}$ that satisfy the conditions for Lemma 1 and Proposition 1.

Our analysis does not make any formal statements about the limiting distribution of $\tilde{\theta}_i$. In Section 4 we assess the limiting distribution of the estimator. We compare the finite-sample performance of the maximizer $\tilde{\theta}_i$ of the feasible likelihood function $\tilde{\mathcal{L}}_T(X_i|\tilde{F};\theta_i)$ with the maximizer $\hat{\theta}_i$ of the infeasible likelihood function $\mathcal{L}_T(X_i|F;\theta_i)$ for which asymptotic normality holds, see Pagan (1980). The simulations show that the root-mean-squared error of the feasible estimator $\tilde{\theta}_i$ seems to convergence to that of the infeasible estimator $\hat{\theta}_i$.

In Assumption G.2 we assume that the idiosyncratic errors have no temporal dependence. It is straightforward to relax this assumption. We could model the idiosyncratic errors as cross-sectionally uncorrelated autoregressions and estimate the parameters by including e_{it} in the state equation of the state space representation of the model and compute the likelihood with the Kalman filter. The proof of Theorem 1 applies with very minor changes. The assumption of no temporal dependence in e_{it} is thus only for expositional simplicity.

4 Monte Carlo simulations

In this section, we conduct a simulation study to assess the finite-sample performance of the two-step estimator. We provide results for both principal components and maximum likelihood estimates. Section 4.1 describes the simulation design, and Section 4.2 reports and discusses the results.

4.1 Design

The simulation design broadly follows that of Stock and Watson (2002):

$$\begin{aligned}
&X_{it} = \lambda'_{it}F_t + e_{it}, \\
&(I_r - B_iL)(\lambda_{it} - \lambda_i) = \eta_{it}, \qquad \eta_{it} \sim \text{i.i.d. } \mathcal{N}(0, Q_i), \\
&F_{tp} = \rho F_{t-1,p} + u_{tp}, \qquad u_{tp} \sim \text{i.i.d. } \mathcal{N}(0, 1-\rho^2), \\
&(1-\alpha L)e_{it} = v_{it}, \qquad v_t \sim \text{i.i.d. } \mathcal{N}(0, \Omega),
\end{aligned}$$

where $i = 1, ..., N$, $t = 1, ..., T$, $p = 1, ..., r$. The processes $\{\eta_{it}\}$,$\{u_{tp}\}$, and $\{v_t\}$ are mutually independent. The autoregressive matrix B_i determines the degree of persistence of the loadings and has eigenvalues inside the unit circle in all simulations. The unconditional mean of the loadings is $\lambda_i = (\lambda_{i1}, ..., \lambda_{ir})'$ and $\lambda_{ip} \sim$ i.i.d. $\mathcal{N}(0, 1)$ in all simulations. The matrix Q_i is the covariance matrix of the loadings innovations. The model allows for cross-sectional and temporal dependence in the errors e_{it}. The parameter α determines the degree of serial correlation in the idiosyncratic errors, and cross-sectional correlation is modelled by specifying the variance matrix of v_t as $\Omega = \left(\beta^{|i-j|}\sqrt{\psi_i \psi_j}\right)_{ij}$ for $i, j = 1, ..., N$. The matrix is thus a Toeplitz matrix and the cross-sectional correlation between the idiosyncratic elements is therefore limited and determined by the coefficient β.

We allow for factor persistence through the coefficient ρ. Furthermore, we consider the case where the loadings are weakly dependent across i. We model the correlation such that $Corr(\eta_{ipt}, \eta_{jpt}) = \pi^{|i-j|}$ for $i, j = 1, ..., N$. Finally, we introduce correlation between factors and loadings through u_t and η_{it}. For each i, we simulate the variables $\begin{pmatrix} u^*_{tp} \\ \eta^*_{itp} \end{pmatrix} \sim A\mathcal{N}(0, I_r)$, where A is the lower triangular Cholesky decomposition matrix such that $AA' = \begin{pmatrix} 1 & \gamma \\ \gamma & 1 \end{pmatrix}$. The variables u^*_{tp} and η^*_{itp} are then rescaled to get the innovations $u_{tp} = u^*_{tp}(1-\rho^2)^{1/2}$ and $\eta_{itp} = \eta^*_{itp} q_i^{1/2}$, respectively.

We generate the model 2000 times for each of the different combinations of T and N. To avoid any dependence on initial values of the simulated processes we discard a 'burn-in' period of 200 observations for each simulation. The principal components are calculated with the estimator $\tilde{F}_t$ defined in (6). The data X_{it} are standardized to have mean zero and variance equal to one prior to extracting principal components. The principal components are identified only up to an orthogonal rotation. In order to directly compare the maximum likelihood estimates with data-generating parameters, we therefore rotate the principal components to resemble the simulated factors. More specifically, we solve for the orthogonal $r \times r$ matrix A^* that maximizes tr[corr$(F, \tilde{F}A)$].[7] The estimates are then rescaled to have the same standard deviation as the true simulated factors:

$$\tilde{F}^*_p = \frac{\sigma(F_p)}{\sigma(\tilde{F}_p)}\tilde{F}_p, \quad p = 1, ..., r$$

[7] The solution is $A^* = VU'$ where V and U are the orthogonal matrices of the singular value decomposition corr$(F, \tilde{F}) = USV'$. When the number of principal components k is not equal to the true number of factors r, we only rotate the first $l = \min\{k, r\}$ principal components. Eickmeier et al. (2015) use the same rotation.

where $\tilde{F}_p$ is the p^{th} column of the rotated principal components matrix $\tilde{F}A^*$. Such rotations are innocuous and allow us to directly compare the estimated parameter values with the data-generating parameters. The principal components are treated as data thereafter, and we maximize the feasible likelihood $\tilde{\mathcal{L}}_T(X_i|\tilde{F}^*;\theta_i)$ to estimate θ_i.

The performance of the principal component estimator $\tilde{F}$ is measured by the trace statistic:

$$R^2_{\tilde{F},F} = \frac{\hat{E}[\text{tr}(F'\tilde{F}(\tilde{F}'\tilde{F})^{-1}\tilde{F}'F)]}{\hat{E}[\text{tr}(F'F)]},$$

where $\hat{E}$ denotes the average over the 2000 Monte Carlo simulations. The trace statistic $R^2_{\tilde{F},F}$ is a multivariate R^2 from a regression of the true data-generating factors on the principal components. It is smaller than 1 and tends to 1 as the canonical correlation between the factors and the principal components tends to 1.

For the maximum likelihood estimates $\tilde{\theta}_i$ we compute the mean estimates over the Monte Carlo repetitions for each parameter.[8] However, for the mean parameter $\lambda_{i,p}$ we report the bias of the estimates $\tilde{\lambda}_{i,p}$ as the true value of $\lambda_{i,p}$ varies over p. Furthermore, we calculate the root-mean-squared error of the estimates $\tilde{\theta}_i$ and also of the infeasible estimates $\hat{\theta}_i$ where the true data-generating factors are used in the maximum likelihood estimation. We report the relative root-mean-squared error between the estimates $\tilde{\theta}_i$ and $\hat{\theta}_i$. This gives us a measure of the estimation error in $\tilde{\theta}_i$ that is due to estimation error from the principal components estimates.

The parameters are identically chosen across the cross-section.[9] The properties of the estimated parameters $\tilde{\theta}_i$ are thus the same for all i and we only report the results for a single cross-section index.[10] In the baseline case, we set $B_i = \text{diag}\{b_{ip}\}_{p=1,...,r}$, $Q_i = \text{diag}\{q_{ip}\}_{p=1,...,r}$, and choose the loadings persistence and variance parameters to be $b_{ip} = 0.9$ and $q_{ip} = 0.2$. The idiosyncratic errors are cross-sectionally and temporally uncorrelated, i.e. $\alpha = 0$, $\beta = 0$, and the variance is set to $\psi_i = 1$. The loadings are cross-sectionally independent, $\pi = 0$, and also independent of the factors, $\gamma = 0$. Finally, we set $\rho = 0$ such that the factors are white noise.

We introduce serial correlation and cross-sectional dependence separately in the idiosyn-

[8] Convergence is generally very good, with all 1-factor estimations having over 99% convergence rate, and most estimations with 2 and 3 factors have over 98% convergence rate. Exceptions are sample sizes of $T = 50$ for the 2- and 3-factor models where the lowest convergence rate is 92%. However, this is expected as we are estimating up to 10 parameters in a highly non-linear model with 50 observations. Convergence statistics using the true factors are similar, but with somewhat better convergence rates for calibrations with 2 and 3 factors and $T = 50$.

[9] The mean parameters λ_i are not the same for all i. This is necessary for Assumption B to be satisfied. With λ_i identical over i, the matrix Λ^0 does not have full rank and $\Lambda^{0\prime}\Lambda^0/N$ will not converge to a positive definite matrix.

[10] Simulations with loadings parameters calibrated with heterogeneous values across i show similar results as in Table 1. The results are available upon request.

cratic errors. We set $\alpha = 0.5$ and estimate this parameter by including e_{it} in the state equation. To consider the effect of cross-sectional correlation, we set $\beta = 0.5$. We also report results with persistent factors with ρ set to both 0.9 and 0.5. Results with cross-sectionally correlated loadings are reported for $\pi = 0.3$, and for the correlation parameter between loadings and factors set to $\gamma = 0.3$. Finally, we consider the consequences of estimating the wrong number of factors, i.e. extracting one principal component too few and one too many, respectively.

4.2 Results

Table 1 reports the results for one factor, $r = 1$. Panel I shows the results for the baseline model with no serial correlation, no cross-sectional dependence in errors, and no factor dependence. The $R^2_{\tilde{F},F}$ statistics show that the factor estimates are close to the true factors even for small sample sizes. For the autoregressive parameter b_i, the estimates improve as the sample size T increases. Increasing the cross-sectional dimension N only gives minor improvements for fixed T. This is unsurprising as a larger N can only improve the parameter estimates through better factor estimates which are already quite good even for $N = 50$. The estimate of the loadings innovation variance q_i is closely related to the estimate of b_i. As b_i gets closer to its true value, so does q_i, and vice versa. For $T \geq 200$ the estimates are close to the true values. The small-sample bias of b_i is not a consequence of estimation error from principal components. Using the true factors instead of principal components to estimate the parameters of the latent process λ_{it} also shows that $T \geq 200$ is needed for the bias of b_i and q_i to be less than 10% of the true value. The loadings mean λ_i and the error variance ψ_i are very precisely estimated for all sample sizes.

In Panel II, the idiosyncratic errors are serially correlated, and the autoregressive parameters for the errors are estimated along with the other parameters. The $R^2_{\tilde{F},F}$ statistic is hardly affected by serially correlated errors. The results are very close to the corresponding values in the first panel. The results for the loadings parameters are also very similar and are not markedly affected. The estimates of the autoregressive parameter for the errors α and the variance parameter ψ are very close to their true value for all sample sizes. The model with serially correlated errors can thus be estimated equally well as the model with i.i.d. errors.

Next, in Panel III we consider the effect of cross-sectional correlation in the errors. The results are very similar to the results in Panel I. Cross-sectional correlation in the idiosyncratic errors has little effect on the parameter estimates.

High factor persistence has a larger impact on the $R^2_{\tilde{F},F}$ statistic. Panel IV shows much lower values of these statistics for all but the largest sample sizes. However, this estimation error does not seem to influence the estimate of the loadings parameters. The estimates for b_i, and accordingly q_i, are similar to the case of white noise factors. The most notable impact of the lower $R^2_{\tilde{F},F}$ is in the estimate of ψ_i. The increase in factor estimation error seems to inflate the error variance, which is larger for all sample sizes, but the results do show convergence for the largest sample size. Results for more moderate levels of factor persistence are shown in Panel V. The drop in the $R^2_{\tilde{F},F}$ is less severe in this case and the estimate of ψ_i thus less biased.

In Table 2, the relative root-mean-squared errors of the estimates using principal components (numerator) and the true simulated factors (denominator) are reported. Values close to 1 indicate that the asymptotic variance of the parameter estimates is unaffected by the estimation error from principal components estimation of the factors. In Panels I-III, all the statistics are close to 1 even for the smallest sample sizes. In Panel IV, the statistics for the loadings parameters are somewhat higher for the smaller sample sizes, but close to one for large sample sizes. The statistics for the idiosyncratic variances are much larger than 1. This is partly due to the bias of these estimates evident in Panel IV of Table 1, but it also reflects higher variability of the estimates. High factor persistence thus mainly affects the idiosyncratic variance parameters. Unreported results show that the estimates improve for larger sample sizes. In Panel V, the factor persistence is more moderate and the relative root-mean-squared errors are much closer to 1.

Panel I in Table 3 reports results for the case where the loadings are cross-sectionally correlated. The results are very similar to Panel I in Table 1. The $R^2_{\tilde{F},F}$ statistic and the parameter estimates are not influenced by cross-sectional dependence in the loadings. Unreported results show that stronger cross-sectional dependence has only very minor effects on the results. The $R^2_{\tilde{F},F}$ statistics generally falls by a single percentage point, but the loadings parameters are not affected.

In Panel II, the loadings and factors are correlated. Correlation between factors and loadings leads to a minor inflation of the estimates of q_i for the largest sample sizes. This is not simply sampling variation. Unreported results for larger sample sizes show that the estimates of q_i do not converge to 0.2. When factors and loadings are correlated, the data exhibits some variance that is not captured by any parameter in the model. The variance in the data that is due to $Cov(\xi_{it}, F_t)$ shows up in the estimate of the loadings variance. Stronger correlation between factors and loadings inflates the estimate of q_i further. The argument of Theorem 1 are, however, still valid. The infeasible likelihood function convergences to the infeasible likelihood function. Using the simulated factors instead of the principal components to estimate the model leads to similar parameter estimates. This is evident from Table 4, Panel II. Here we report the relative root-mean-squared errors of the estimates using principal components and the true simulated factors. The results are all close to 1. The estimates using principal components and simulated factors are therefore consistent for the same parameter. Correlation between factors and loadings do not affect the estimates of the other parameters. They are similar to the results in Panel I, Table 1.

Table 5 displays the simulations results for the model with 2 and 3 factors with i.i.d. errors and white noise factors. Compared to the 1-factor model, the $R^2_{\tilde{F},F}$ statistics are lower, reflecting the increasing difficulties in extracting additional factors. In Panel I, the estimates for the second set of loadings parameters are worse than for the first set and the same pattern is evident for the 3-factor model (Panel II). The results for the third set of loadings parameters are worse than for the second, which are worse than for the first. However, all the estimates are converging to their true values. Compared to the 1-factor model, larger sample sizes are generally needed to get precise estimates due to the increased number of parameters. Introducing serial and cross-sectional correlation in the errors, correlated factors and loadings, or persistence in factors does not reveal any additional insights compared to the 1-factor model. The results generalize and are therefore omitted. Table 6 shows the relative root-mean-squared errors for the 2- and 3-factor model. The statistics are somewhat larger than 1 for the smaller sample sizes, but get increasingly closer to one as the sample sizes grow. This indicates that the estimation error of the principal components does not affect the asymptotic variance of the estimates.

Table 7 shows the results of estimating the wrong number of factors. For these simulations, we report two convergence statistics for the principal components. The first is the R^2 from a regression of the principal components on the true factors, $R^{2(1)}_{\tilde{F},F} = \frac{\hat{E}[\tilde{F}'F(F'F)^{-1}F'\tilde{F}]}{\hat{E}[\tilde{F}'\tilde{F}]}$ and the second is the R^2 from a regression of the true factors on the principal components. In Panel I, the simulated data have two factors, but only one principal component is extracted. The first statistic $R^{2(1)}_{\tilde{F},F}$ is close to 1 for all sample sizes. Hence, the two factors explain all the variation in the single principal component. The second statistic $R^{2(2)}_{\tilde{F},F}$ does not tend to 1, as a single principal component cannot span the two-dimensional factor space. The results show that the loadings parameters for the first factor can still be estimated consistently. The consequence of excluding a factor is that the estimate of the error variance ψ_i gets larger, reflecting the variability in the data from the excluded factor and its loadings. Panel II displays results for the 1-factor model with two principal components extracted from the data. $R^{2(2)}_{\tilde{F},F}$ tends to 1, and the two principal components thus explain all the variation in the single factor. The other measure $R^{2(1)}_{\tilde{F},F}$ tends to 0.5 as the single factor can only span half of the two-dimensional space of the principal components. The loadings on the first factor are estimated consistently. For the second factor, the mean and variance of the loadings are being estimated as zero.[11] The estimated parameters thus show that the data do not load on the second factor and therefore correctly dismiss the second factor. The results are thus very encouraging even with the number of principal components different from the true number of factors.

The main insights from the simulations can be summarized as follows. The loadings and idiosyncratic variance parameters are estimated consistently. The sample size T needs to be sufficiently large (≥ 200) for the bias in the autoregressive parameters to be less than 10%. Furthermore, the loadings parameters are consistently estimated even when an incorrect number of principal components are extracted. Too few principal components increase the error variance estimate, and loadings means and variances are correctly estimated as zero for principal components in excess of the true number of factors. Finally, the relative root-mean-squared errors indicate that the asymptotic variance is unaffected by replacing the factors with the principal components estimates.

5 An empirical illustration

We provide an empirical illustration of the model using the data set of Stock and Watson (2009), who analyze a balanced panel of 144 quarterly time series for the United States, focusing on structural instability in factor loadings. The data set consists of 144 quarterly time series for the United States, spanning the sample period 1959:I-2006:IV. The data series are transformed to be stationary, and the first two quarters are thus excluded because of differencing, resulting in $T = 190$ observations used for estimation. We exclude a number of series that are higher-level aggregates of the included series, which brings the number of series used for estimation to $N = 109$. For a complete data description and details on data transformations, see the appendix of Stock and Watson (2009).

Stock and Watson (2009) argue for four factors in the sample, and perform robustness checks of their results using different numbers of factors. We therefore extract four principal components from the standardized data and estimate the loadings parameters and the idiosyncratic variances for each of the 109 variables. The system matrices B_i and Q_i are specified to be diagonal, i.e., the loadings are estimated as univariate autoregressions uncorrelated over the factor indices. The lag polynomials $B_i(L)$ are of order one for all i.

Figure 1 shows the squared correlation coefficient R^2 of the four principal components with

[11] The results for b_{i2} are not indicative of any convergence. Histograms of the estimated values show that the parameter is not identified as the values are randomly estimated anywhere between -1 and 1.

the 109 time series in the cross-section. The first factor correlates most strongly with series that measure output, labor, and inventories, with many correlation coefficients larger than 0.5 for these variables. The second factor correlates across the board, never above 0.50, and close to zero for prices and wages and for monetary variables. The third and fourth factors never exceed correlations of 0.40, but show clustering in the financial and monetary groups of variables.

Table 8 shows the in-sample R^2 from estimated factor models with four principal components and constant loadings (const.) in comparison to time-varying loadings (tv.). The gains in R^2 from employing a time-varying factor model range from 0 (for two series, Emp: services and Orders [NDCapGoods], the loadings are estimated as constant) up to 95 percent (for example, for PCED-NDUR-ENERGY). The mean gain in R^2 across the 109 time series is 38 percent.

We select three time series and display the estimation results in Figures 2 to 4. The first panel of each figure shows the time series and the common components from factor models with constant loadings and with time-varying loadings. The second panel of each figure shows the estimated time-varying loadings for the four estimated factors (principal components).

The series in Figure 2 is the exchange rate CHF/USD, for which the in-sample R^2 increases from 0.064 for a factor model with constant loadings to 0.98 for a factor model with time-varying loadings. The first panel illustrates this difference: The common component from a factor model with constant loading captures very little of the variation in the series, whereas the common component of the factor model with time-varying loadings provides a close fit. The estimated time-varying loadings in the second panel show that most of the dynamics in the common component stem from the first and fourth factors for most of the sample period. The loadings of the second and third factors oscillate around zero, with some pronounced exceptions in the mid-80s and early 2000s, where the loadings on the third factor spike. The absolute value of the mean of the loadings on the first factor is 0.19 (note that the sign is not identified). The absolute value of the mean of the loadings on the fourth factor is 0.26. The absolute values of the means of the loadings on the second and third factors are 0.003 and 0.03, respectively. Since the fourth factor has the strongest correlations with the financial group of variables (see Figure 1), its large influence on this exchange rate is not unexpected. The loadings on the first and second factor are negatively serially correlated; the loadings on the third and fourth factor are positively serially correlated.

Figure 3 displays the time series of unit labor costs (total labor compensation divided by real output) and its estimated common components from factor models with constant and with time-varying loadings in the first panel. The in-sample R^2 for this series improves from 0.04 to 0.97 when introducing time-varying loadings. Same as in the case of the CHF/USD exchange rate, the common component from constant loadings explains very little of the variation in the series. The estimated time-varying loadings in the second panel show in this case that the first factor, associated with output, consumption, labor, housing, and inventories, is the most important in explaining the variation in the series. This illustrates that in the factor model with time-varying loadings, a single factor can dominate even though it is not able to explain much of the variation under constant loadings. All estimated time-varying loadings display strong positive autocorrelation. Stock and Watson (2009) find strong support for a structural break in 1984:I for this series using a Chow split-sample test (their Table 3). Figure 3 implies that there is strong positive autocorrelation in the residuals from a factor model with constant loadings, and a Chow test is likely to reject parameter constancy for a wide range of possible change-points.

Finally, Figure 4 displays the time series of the number of employees in the service sector. In this case, the factor model with time-varying loadings returned constant loadings, and so there is no difference in the R^2 and in the common components compared to a model with constant loadings.

6 Conclusion

We proposed a two-step maximum likelihood estimator for time-varying loadings in high-dimensional factor models. The loadings parameters are estimated by a set of N univariate regression models with time-varying coefficients, where the unobserved regressors are estimated by principal components. Replacing the unobservable factors with principal components gives a feasible likelihood function that is asymptotically equivalent to the infeasible one with observable factors and therefore gives consistent estimates of the loadings parameters as $N, T \to \infty$. The finite-sample properties of our estimator were assessed via an extensive simulation study. The results showed that the loadings means and idiosyncratic error variances are estimated precisely even for small sample sizes. A somewhat larger sample size is needed to get precise estimates of the loadings variance and dynamic parameters. Furthermore, the simulations showed very satisfactory results when the number of principal components is different from the number of factors in the data.

We illustrated the empirical relevance of the time-varying loadings structure using the large quarterly dataset of Stock and Watson (2009) for the US economy. For the majority of the variables we found evidence of time-varying loadings, and we showed that a large increase in the in-sample fit of the common component can be obtained by modelling the loadings as time-varying.

Acknowledgements

We thank M. Andreasen, M. Deistler, C. Doz, N. Haldrup, D. Hendry, L. Khalaf, A. Kock, M. Lippi, R. Sickles, L. Trapani, J.-P. Urbain, M. Wagner, participants at seminars at the University of Maastricht, the Tinbergen Institute, at CREATES, the 2016 Camp Econometrics in Clayton, NY, the 2016 IAAE Meetings in Milan, Italy, the 2016 conference on Forecasting Economic and Financial Time Series in Frascati, Italy, the 5th International Conference in Memory of Carlo Giannini at the University of Bergamo, Italy, the 2017 Vienna-Copenhagen Conference on Financial Econometrics in Vienna, Austria, as well as the Editor, Jianqing Fan, and two anonymous referees for valuable comments and suggestions. The usual disclaimer applies. JGM and EH acknowledge support from The Danish Council for Independent Research (DFF 4003-00022) and CREATES - Center for Research in Econometric Analysis of Time Series, funded by the Danish National Research Foundation (DNRF78). Part of the research was conducted while JGM was visiting the Centre for Econometric Analysis at Cass Business School in the Fall of 2015.

Table 1: Simulation results for 1-factor model.

Panel	T	N	$R^2_{\tilde{F},F}$	b_i	λ_i	q_i	ψ_i	α
	True values			0.9	0	0.2	1	
	$\alpha=0,\quad \beta=0,\quad \rho=0,\quad \pi=0,\quad \gamma=0$							
I	50	50	0.943	0.603	-0.022	0.302	0.970	-
	100	50	0.941	0.785	-0.005	0.258	1.005	-
	50	100	0.960	0.607	0.023	0.307	0.978	-
	100	100	0.969	0.792	-0.043	0.260	1.010	-
	100	200	0.976	0.797	-0.011	0.256	0.993	-
	200	200	0.984	0.864	0.003	0.224	1.003	-
	400	200	0.986	0.884	-0.019	0.213	1.009	-
	600	300	0.991	0.890	0.012	0.208	0.997	-
	$\alpha=0.5,\quad \beta=0,\quad \rho=0,\quad \pi=0,\quad \gamma=0$							
II	50	50	0.939	0.617	-0.005	0.296	0.934	0.500
	100	50	0.934	0.805	-0.009	0.246	0.992	0.494
	50	100	0.956	0.614	0.033	0.290	0.929	0.494
	100	100	0.966	0.805	-0.060	0.246	0.978	0.509
	100	200	0.974	0.812	0.015	0.242	0.969	0.493
	200	200	0.982	0.870	0.001	0.223	0.990	0.497
	400	200	0.985	0.885	-0.027	0.212	0.994	0.500
	600	300	0.989	0.890	0.011	0.206	0.995	0.499
	$\alpha=0,\quad \beta=0.5,\quad \rho=0,\quad \pi=0,\quad \gamma=0$							
III	50	50	0.943	0.613	-0.019	0.309	0.974	-
	100	50	0.942	0.805	-0.016	0.247	1.010	-
	50	100	0.958	0.607	0.008	0.301	0.972	-
	100	100	0.968	0.790	-0.043	0.259	1.024	-
	100	200	0.976	0.797	0.002	0.255	0.999	-
	200	200	0.983	0.865	-0.005	0.225	0.999	-
	400	200	0.986	0.886	-0.003	0.209	1.025	-
	600	300	0.990	0.891	0.012	0.206	0.993	-
	$\alpha=0,\quad \beta=0,\quad \rho=0.9,\quad \pi=0,\quad \gamma=0$							
IV	50	50	0.652	0.587	0.000	0.355	1.177	-
	100	50	0.785	0.784	0.019	0.255	1.134	-
	50	100	0.651	0.606	-0.031	0.392	1.257	-
	100	100	0.809	0.828	0.010	0.291	1.385	-
	100	200	0.806	0.781	0.009	0.269	1.116	-
	200	200	0.896	0.865	-0.024	0.237	1.065	-
	400	200	0.941	0.894	-0.050	0.221	1.122	-
	600	300	0.961	0.895	0.026	0.211	1.040	-
	$\alpha=0,\quad \beta=0,\quad \rho=0.5,\quad \pi=0,\quad \gamma=0$							
V	50	50	0.905	0.625	-0.033	0.306	1.017	-
	100	50	0.920	0.803	-0.010	0.249	1.023	-
	50	100	0.916	0.606	0.044	0.307	1.045	-
	100	100	0.950	0.803	-0.060	0.254	1.076	-
	100	200	0.955	0.803	0.019	0.254	1.015	-
	200	200	0.973	0.865	-0.010	0.226	1.012	-
	400	200	0.981	0.885	-0.022	0.214	1.026	-
	600	300	0.987	0.889	0.012	0.209	1.006	-

NOTE: The columns T and N report the sample sizes. The column $R^2_{\tilde{F},F}$ reports the convergence statistic for the principal components estimator. The remaining columns report the mean of the parameter estimates over the Monte Carlo simulations. For the parameter λ_i, the bias is reported.

Table 2: Relative root-mean-squared error for 1-factor model.

Panel	T	N	$R^2_{\tilde{F},F}$	b_i	λ_i	q_i	ψ_i	α
	True values			0.9	0	0.2	1	
$\alpha = 0, \quad \beta = 0, \quad \rho = 0, \quad \pi = 0, \quad \gamma = 0$								
I	50	50	0.943	1.043	1.056	0.983	1.024	-
	100	50	0.941	1.083	1.062	0.986	1.024	-
	50	100	0.960	1.032	1.042	1.039	1.030	-
	100	100	0.969	1.065	1.033	1.017	1.076	-
	100	200	0.976	1.009	1.033	1.032	1.020	-
	200	200	0.984	1.029	1.019	0.992	1.005	-
	400	200	0.986	0.995	1.013	1.008	1.039	-
	600	300	0.991	0.979	1.011	0.989	0.997	-
$\alpha = 0.5, \quad \beta = 0, \quad \rho = 0, \quad \pi = 0, \quad \gamma = 0$								
II	50	50	0.939	1.040	1.065	1.063	0.997	1.006
	100	50	0.934	1.022	1.072	0.985	1.025	1.014
	50	100	0.956	1.021	1.054	0.989	0.988	1.001
	100	100	0.966	1.105	1.051	1.032	1.015	1.017
	100	200	0.974	0.999	1.043	1.020	0.998	1.002
	200	200	0.982	0.977	1.021	1.011	0.998	0.999
	400	200	0.985	0.985	1.019	1.018	1.015	0.999
	600	300	0.989	0.980	1.012	1.007	1.000	1.007
$\alpha = 0, \quad \beta = 0.5, \quad \rho = 0, \quad \pi = 0, \quad \gamma = 0$								
III	50	50	0.943	1.014	1.060	0.950	1.016	-
	100	50	0.942	1.071	1.070	1.019	1.029	-
	50	100	0.958	1.045	1.038	1.133	1.027	-
	100	100	0.968	1.177	1.033	1.030	1.095	-
	100	200	0.976	1.043	1.035	1.015	1.021	-
	200	200	0.983	1.029	1.016	0.996	1.007	-
	400	200	0.986	0.986	1.007	0.996	1.075	-
	600	300	0.990	0.975	1.010	0.998	1.000	-
$\alpha = 0, \quad \beta = 0, \quad \rho = 0.9, \quad \pi = 0, \quad \gamma = 0$								
IV	50	50	0.652	1.060	1.147	1.339	1.926	-
	100	50	0.785	1.106	1.127	1.060	1.778	-
	50	100	0.651	1.038	1.147	1.451	2.415	-
	100	100	0.809	0.854	1.166	1.468	4.506	-
	100	200	0.806	1.047	1.085	1.083	1.665	-
	200	200	0.896	0.947	1.085	1.101	1.455	-
	400	200	0.941	0.851	1.099	1.130	2.743	-
	600	300	0.961	0.921	1.044	1.021	1.432	-
$\alpha = 0, \quad \beta = 0, \quad \rho = 0.5, \quad \pi = 0, \quad \gamma = 0$								
V	50	50	0.905	1.061	1.064	0.936	1.086	-
	100	50	0.920	1.043	1.067	1.049	1.039	-
	50	100	0.916	1.024	1.046	1.030	1.241	-
	100	100	0.950	0.956	1.043	1.060	1.415	-
	100	200	0.955	1.090	1.040	1.070	1.057	-
	200	200	0.973	1.012	1.028	1.023	1.038	-
	400	200	0.981	0.984	1.014	1.013	1.169	-
	600	300	0.987	0.971	1.011	0.990	1.040	-

NOTE: The columns T and N report the sample sizes. The column $R^2_{\tilde{F},F}$ reports the convergence statistic for the principal components estimator. The remaining columns report the relative root-mean-squared error of the parameter estimates using principal components (numerator) and the true simulated factors (denominator).

Table 3: Simulation results for 1-factor model.

Panel	T	N	$R^2_{\tilde{F},F}$	b_i	λ_i	q_i	ψ_i
	True values			0.9	0	0.2	1
	$\alpha = 0, \quad \beta = 0, \quad \rho = 0, \quad \pi = 0.3, \quad \gamma = 0$						
I	50	50	0.944	0.584	-0.016	0.311	0.970
	100	50	0.941	0.796	-0.003	0.253	1.008
	50	100	0.959	0.591	0.032	0.311	0.975
	100	100	0.969	0.794	-0.038	0.248	1.024
	100	200	0.976	0.804	-0.023	0.256	0.993
	200	200	0.983	0.864	-0.008	0.225	1.001
	400	200	0.986	0.885	-0.025	0.212	1.010
	600	300	0.990	0.889	0.003	0.205	1.000
	$\alpha = 0, \quad \beta = 0, \quad \rho = 0, \quad \pi = 0, \quad \gamma = 0.3$						
II	50	50	0.941	0.604	-0.020	0.303	0.986
	100	50	0.939	0.797	0.004	0.258	1.012
	50	100	0.957	0.572	0.008	0.316	0.985
	100	100	0.967	0.788	-0.045	0.259	1.028
	100	200	0.973	0.793	-0.000	0.264	0.997
	200	200	0.983	0.861	-0.011	0.239	1.001
	400	200	0.985	0.883	-0.026	0.213	1.014
	600	300	0.990	0.888	0.007	0.214	1.001

NOTE: The columns T and N report the sample sizes. The column $R^2_{\tilde{F},F}$ reports the convergence statistic for the principal components estimator. The remaining columns report the mean of the parameter estimates over the Monte Carlo simulations. For the parameter λ_i, the bias is reported.

Table 4: Relative root-mean-squared error for 1-factor model.

	T	N	$R^2_{\tilde{F},F}$	b_i	λ_i	q_i	ψ_i
Panel	True values			0.9	0	0.2	1
	$\alpha = 0, \quad \beta = 0, \quad \rho = 0, \quad \pi = 0.3, \quad \gamma = 0$						
I	50	50	0.944	1.059	1.057	0.989	1.019
	100	50	0.941	1.036	1.062	1.026	1.035
	50	100	0.959	1.066	1.047	1.047	1.014
	100	100	0.969	1.069	1.031	1.014	1.096
	100	200	0.976	1.019	1.040	1.017	1.016
	200	200	0.983	1.047	1.016	0.998	1.010
	400	200	0.986	0.978	1.017	1.004	1.033
	600	300	0.990	0.991	1.008	0.996	1.005
	$\alpha = 0, \quad \beta = 0, \quad \rho = 0, \quad \pi = 0, \quad \gamma = 0.3$						
II	50	50	0.941	1.057	1.052	1.009	1.025
	100	50	0.939	1.061	1.057	1.039	1.053
	50	100	0.957	1.088	1.038	0.979	1.042
	100	100	0.967	1.074	1.029	0.991	1.180
	100	200	0.973	1.040	1.043	1.017	1.019
	200	200	0.983	0.998	1.019	1.008	1.016
	400	200	0.985	0.976	1.011	0.967	1.101
	600	300	0.990	0.982	1.004	0.982	1.018

NOTE: The columns T and N report the sample sizes. The column $R^2_{\tilde{F},F}$ reports the convergence statistic for the principal components estimator. The remaining columns report the relative root-mean-squared error of the parameter estimates using principal components (numerator) and the true simulated factors (denominator).

Table 5: Simulation results for 2- and 3-factor model.

Panel	T	N	$R^2_{\tilde{F},F}$	b_{i1}	λ_{i1}	q_{i1}	b_{i2}	λ_{i2}	q_{i2}	b_{i3}	λ_{i3}	q_{i3}	ψ_i
	True values			0.9	0	0.2	0.9	0	0.2	0.9	0	0.2	1
Two factors – $\alpha = 0, \quad \beta = 0, \quad \rho = 0, \quad \pi = 0, \quad \gamma = 0$													
I	50	50	0.920	0.615	0.009	0.277	0.447	-0.156	0.379	-	-	-	1.146
	100	50	0.934	0.825	0.018	0.236	0.731	0.033	0.287	-	-	-	1.017
	50	100	0.953	0.629	-0.017	0.271	0.490	-0.080	0.365	-	-	-	0.971
	100	100	0.953	0.819	0.047	0.234	0.735	-0.041	0.287	-	-	-	1.032
	100	200	0.973	0.813	-0.056	0.240	0.739	-0.051	0.291	-	-	-	1.088
	200	200	0.978	0.872	-0.006	0.216	0.854	0.002	0.239	-	-	-	1.006
	400	200	0.982	0.888	-0.020	0.205	0.883	0.036	0.217	-	-	-	1.029
	600	300	0.987	0.892	0.004	0.204	0.891	0.021	0.210	-	-	-	1.013
Three factors – $\alpha = 0, \quad \beta = 0, \quad \rho = 0, \quad \pi = 0, \quad \gamma = 0$													
II	50	50	0.883	0.584	-0.000	0.244	0.434	-0.143	0.334	0.340	0.006	0.516	1.396
	100	50	0.875	0.813	0.029	0.196	0.755	-0.038	0.245	0.587	-0.181	0.372	1.353
	50	100	0.906	0.605	-0.010	0.240	0.476	-0.051	0.333	0.351	-0.151	0.536	1.134
	100	100	0.938	0.831	-0.021	0.216	0.770	-0.103	0.262	0.639	0.051	0.361	1.132
	100	200	0.960	0.838	0.023	0.219	0.774	0.034	0.262	0.654	-0.120	0.365	1.040
	200	200	0.974	0.875	-0.014	0.210	0.862	0.014	0.226	0.823	0.062	0.276	1.015
	400	200	0.975	0.889	0.022	0.207	0.886	0.043	0.214	0.880	0.051	0.231	1.099
	600	300	0.985	0.893	-0.001	0.204	0.890	-0.003	0.210	0.887	-0.003	0.216	1.025

NOTE: The columns T and N report the sample sizes. The column $R^2_{\tilde{F},F}$ reports the convergence statistic for the principal components estimator. The remaining columns report the mean of the parameter estimates over the Monte Carlo simulations. For the parameter λ_i, the bias is reported.

Table 6: Relative root mean squared errors for 2- and 3-factor model.

Panel	T	N	$R^2_{\tilde{F},F}$	b_{i1}	λ_{i1}	q_{i1}	b_{i2}	λ_{i2}	q_{i2}	b_{i3}	λ_{i3}	q_{i3}	ψ_i
	True values			0.9	0	0.2	0.9	0	0.2	0.9	0	0.2	1
	Two factors — $\alpha = 0$, $\beta = 0$, $\rho = 0$, $\pi = 0$, $\gamma = 0$												
I	50	50	0.920	1.210	1.105	1.151	1.061	1.167	1.446	-	-	-	1.462
	100	50	0.934	0.989	1.063	1.079	1.048	1.090	1.027	-	-	-	1.040
	50	100	0.953	1.087	1.059	1.049	1.054	1.086	1.110	-	-	-	1.069
	100	100	0.953	1.250	1.055	1.030	1.097	1.093	1.062	-	-	-	1.118
	100	200	0.973	1.059	1.059	1.059	1.053	1.112	1.082	-	-	-	1.388
	200	200	0.978	0.978	1.027	1.007	0.929	1.039	1.019	-	-	-	1.027
	400	200	0.982	0.988	1.027	1.010	0.986	1.069	1.012	-	-	-	1.082
	600	300	0.987	0.963	1.037	1.000	0.970	1.022	1.002	-	-	-	1.042
	Three factors — $\alpha = 0$, $\beta = 0$, $\rho = 0$, $\pi = 0$, $\gamma = 0$												
II	50	50	0.883	1.409	1.113	1.221	1.187	1.140	1.372	1.030	1.566	1.557	1.788
	100	50	0.875	1.604	1.082	0.999	1.236	1.119	1.164	1.199	1.877	1.466	1.821
	50	100	0.906	1.359	1.091	1.178	1.114	1.112	1.387	1.016	1.537	1.727	1.403
	100	100	0.938	1.204	1.079	1.021	1.150	1.104	1.121	1.016	1.312	1.286	1.285
	100	200	0.960	1.235	1.051	1.051	1.087	1.082	1.081	1.029	1.217	1.186	1.109
	200	200	0.974	1.073	1.030	1.024	1.062	1.053	1.045	1.029	1.100	1.156	1.047
	400	200	0.975	0.967	1.036	1.041	1.003	1.090	1.091	1.063	1.195	1.109	1.320
	600	300	0.985	0.970	1.014	1.021	0.987	1.021	1.037	0.893	1.033	1.061	1.051

NOTE: The columns T and N report the sample sizes. The column $R^2_{\tilde{F},F}$ reports the convergence statistic for the principal components estimator. The remaining columns reports the relative root-mean-squared error of the parameter estimates using principal components (numerator) and the true simulated factors (denominator).

Table 7: Simulation results for incorrect number of principal components.

Panel	T	N	$R^{2(1)}_{\tilde{F},F}$	$R^{2(2)}_{\tilde{F},F}$	b_{i1}	λ_{i1}	q_{i1}	b_{i2}	λ_{i2}	q_{i2}	ψ_i
		True values			0.9	0	0.2	0.9	0	0.2	1
	1 principal component, two factors — $\alpha = 0$, $\beta = 0$, $\rho = 0$, $\pi = 0$, $\gamma = 0$										
I	50	50	0.954	0.797	0.497	0.011	0.387	-	-	-	6.703
	100	50	0.956	0.789	0.835	0.056	0.230	-	-	-	2.354
	50	100	0.967	0.807	0.649	-0.070	0.313	-	-	-	3.681
	100	100	0.971	0.808	0.831	0.040	0.238	-	-	-	2.587
	100	200	0.981	0.819	0.817	-0.070	0.248	-	-	-	3.566
	200	200	0.985	0.820	0.873	-0.031	0.218	-	-	-	2.078
	400	200	0.989	0.823	0.885	0.023	0.215	-	-	-	4.885
	600	300	0.992	0.826	0.889	0.023	0.212	-	-	-	6.630
	2 principal components, 1 factor — $\alpha = 0$, $\beta = 0$, $\rho = 0$, $\pi = 0$, $\gamma = 0$										
II	50	50	0.475	0.951	0.315	-0.022	0.198	0.128	-0.008	0.100	0.886
	100	50	0.474	0.948	0.642	-0.003	0.196	0.241	0.016	0.075	0.928
	50	100	0.482	0.964	0.317	0.005	0.210	0.126	0.006	0.106	0.901
	100	100	0.486	0.972	0.605	-0.010	0.208	0.248	-0.003	0.079	0.961
	100	200	0.489	0.978	0.583	0.008	0.199	0.268	0.013	0.074	0.953
	200	200	0.492	0.984	0.826	-0.007	0.201	0.241	0.002	0.047	0.967
	400	200	0.493	0.986	0.884	-0.018	0.195	0.236	-0.007	0.025	0.994
	600	300	0.495	0.991	0.889	0.014	0.196	0.227	-0.004	0.017	0.991

NOTE: The columns T and N report the sample sizes. The columns $R^{2(1)}_{\tilde{F},F}$ and $R^{2(1)}_{\tilde{F},F}$ are the two convergence statistics for the principal components estimator. The remaining columns report the mean of the parameter estimates over the Monte Carlo simulations. For the parameter λ_i, the bias is reported.

Table 8: Comparison of in-sample R^2 from factor models with constant loadings (const.) and from factor models with time-varying loadings (tv.).

Description	const.	tv.	Description	const.	tv.	Description	const.	tv.
Cons-Dur	0.62	0.93	U < 5 wks	0.39	0.86	OilPrice (Real)	0.32	0.57
Cons-NonDur	0.22	0.43	U 5-14 wks	0.43	0.91	NAPM com price	0.57	0.98
Cons-Serv	0.74	0.87	U 15+ wks	0.32	0.8	Real AHE: const	0.51	0.96
NonResInv-Struct	0.6	0.96	U 15-26 wks	0.59	0.92	Real AHE: mfg	0.58	0.93
NonResInv-Bequip	0.6	0.84	U 27+ wks	0.83	0.95	Labor Prod	0.47	0.7
Res.Inv	0.82	0.94	HStarts: NE	0.82	0.95	Real Comp/Hour	0.047	0.95
Exports	0.066	0.82	HStarts: MW	0.56	0.98	Unit Labor Cost	0.036	0.97
Imports	0.81	0.88	HStarts: South	0.67	0.87	FedFunds	0.43	0.92
Gov Fed	0.4	0.81	HStarts: West	0.7	0.89	3 mo T-bill	0.55	0.86
Gov State/Loc	0.33	0.57	PMI	0.27	0.76	1 yr T-bond	0.68	0.93
IP: cons dble	0.35	0.65	NAPM new ordrs	0.022	0.49	10 yr T-bond	0.6	0.85
iIP:cons nondble	0.17	0.9	NAPM vendor del	0.073	0.4	fygm6-fygm3	0.5	0.77
IP:bus eqpt	0.56	0.82	NAPM Invent	0.054	0.52	fygt1-fygm3	0.38	0.79
IP: dble mats	0.84	0.99	Orders (ConsGoods)	0.056	0.84	fygt10-fygm3	0.38	0.74
IP:nondble mats	0.66	0.93	Orders (NDCapGoods)	0.032	0.032	FYAAAC-Fygt10	0.087	0.24
IP: mfg	0.8	0.94	PCED-DUR-MOTORVEH	0.081	0.65	FYBAAC-Fygt10	0.55	0.86
IP: fuels	0.78	0.93	PCED-DUR-HHEQUIP	0.071	0.55	M1	0.53	0.85
NAPM prodn	0.76	0.88	PCED-DUR-OTH	0.01	0.77	MZM	0.6	0.98
Capacity Util	0.62	0.86	PCED-NDUR-FOOD	0.036	0.37	M2	0.41	0.87
Emp: mining	0.31	0.79	PCED-NDUR-CLTH	0.017	0.96	MB	0.52	0.75
Emp: const	0.33	0.78	PCED-NDUR-ENERGY	0.022	0.96	Reserves tot	0.22	0.78
Emp: dble gds	0.77	0.85	PCED-NDUR-OTH	0.12	0.45	Reserves nonbor	0.34	0.98
Emp: nondbles	0.75	0.98	PCED-SERV-HOUS	0.1	0.72	BUSLOANS	0.27	0.77
Emp: services	0.65	0.65	PCED-SERV-H0-ELGAS	0.023	0.71	Cons credit	0.066	0.47
Emp: TTU	0.75	0.81	PCED-SERV-HO-OTH	0.077	0.83	Ex rate: avg	0.099	0.66
Emp: wholesale	0.26	0.79	PCED-SERV-TRAN	0.22	0.6	Ex rate: Switz	0.064	0.98
Emp: retail	0.62	0.67	PCED-SERV-MED	0.019	0.86	Ex rate: Japan	0.16	0.61
Emp: FIRE	0.8	0.94	PCED-SERV-REC	0.15	0.77	Ex rate: UK	0.16	0.4
Emp: Govt	0.54	0.72	PCED-SERV-OTH	0.23	0.86	EX rate: Canada	0.098	0.72
Help wanted indx	0.49	0.55	PFI-NRES-STR Price Index	0.013	0.69	S&P 500	0.27	0.91
Help wanted/emp	0.62	0.85	PFI-NRES-EQP	0.072	0.39	S&P: indust	0.0075	0.22
Emp CPS nonag	0.67	0.87	PFI-RES	0.35	0.61	S&P div yield	0.17	0.53
Emp. Hours	0.57	0.67	PEXP	0.19	0.57	S&P PE ratio	0.76	0.98
Avg hrs	0.57	0.8	PIMP	0.53	0.8	DJIA	0.28	0.61
Overtime: mfg	0.57	0.76	PGOV-FED	0.21	0.71	Consumer expect	0.4	0.74
U: all	0.39	0.77	PGOV-SL	0.23	0.77			
U: mean duration	0.39	0.91	Com: spot price (real)	0.19	0.33			

Figure 1: Squared correlations of principal components with time series.

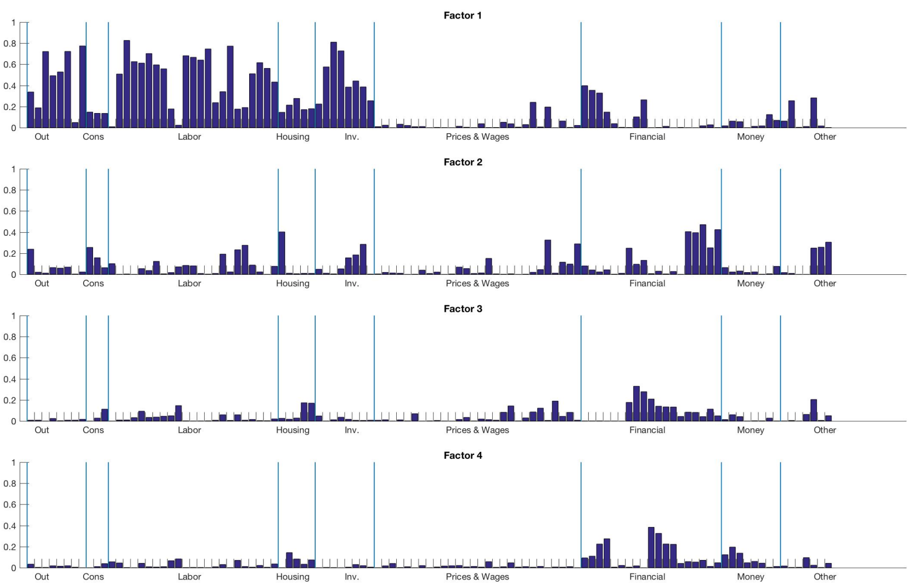
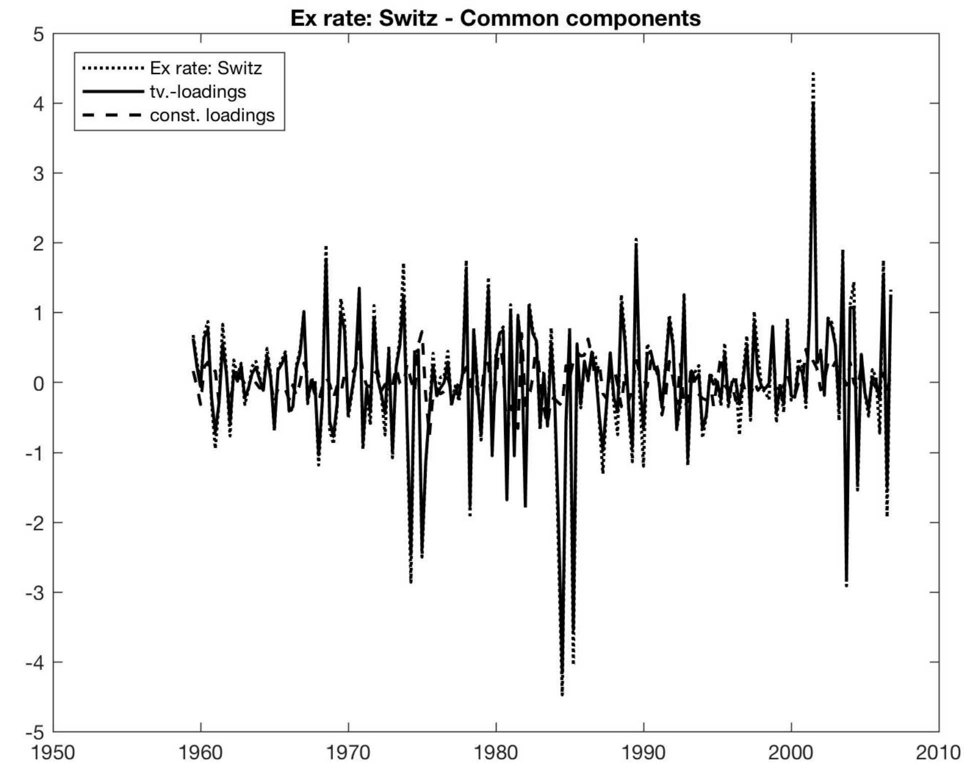


Figure 2: Common components and time-varying factor loadings for the Exchange rate CHF/USD.



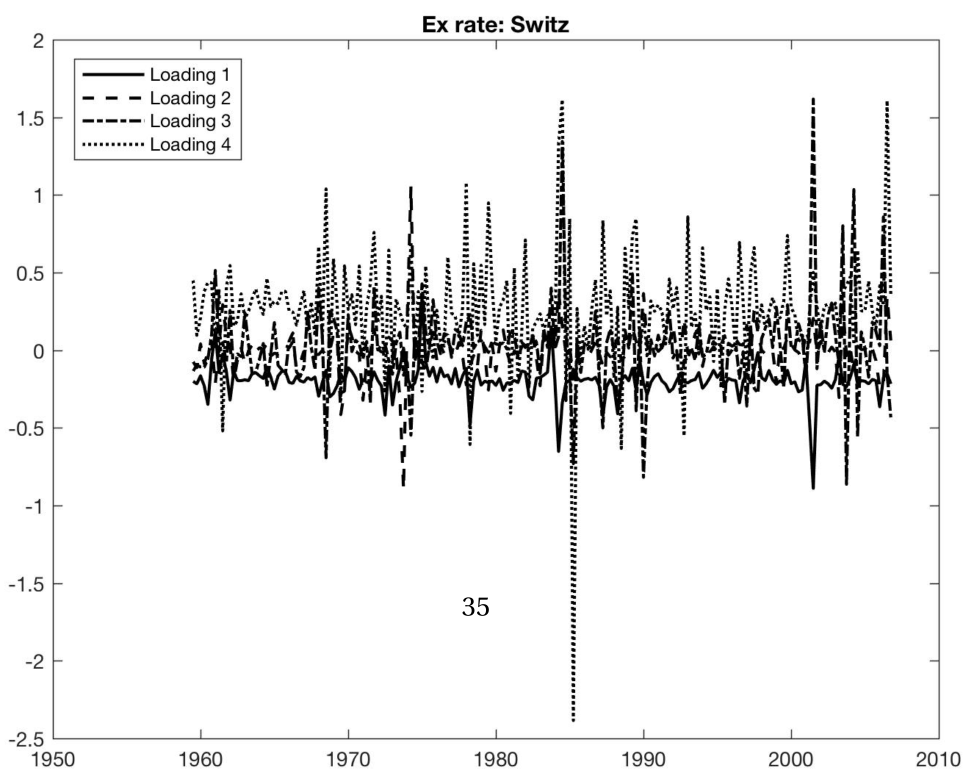


Figure 3: Common components and time-varying factor loadings for Unit labor costs (= total labor compensation / real output).

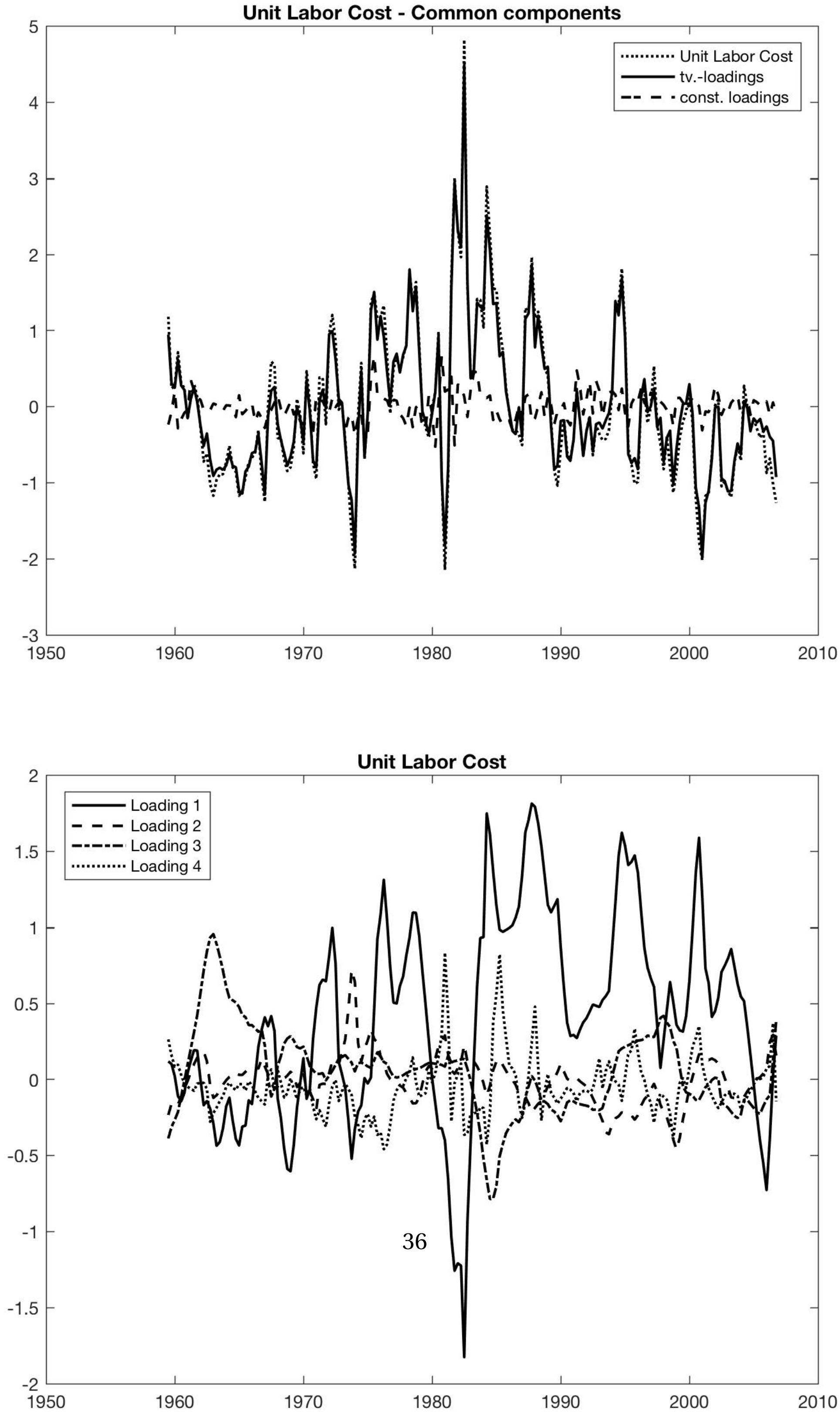
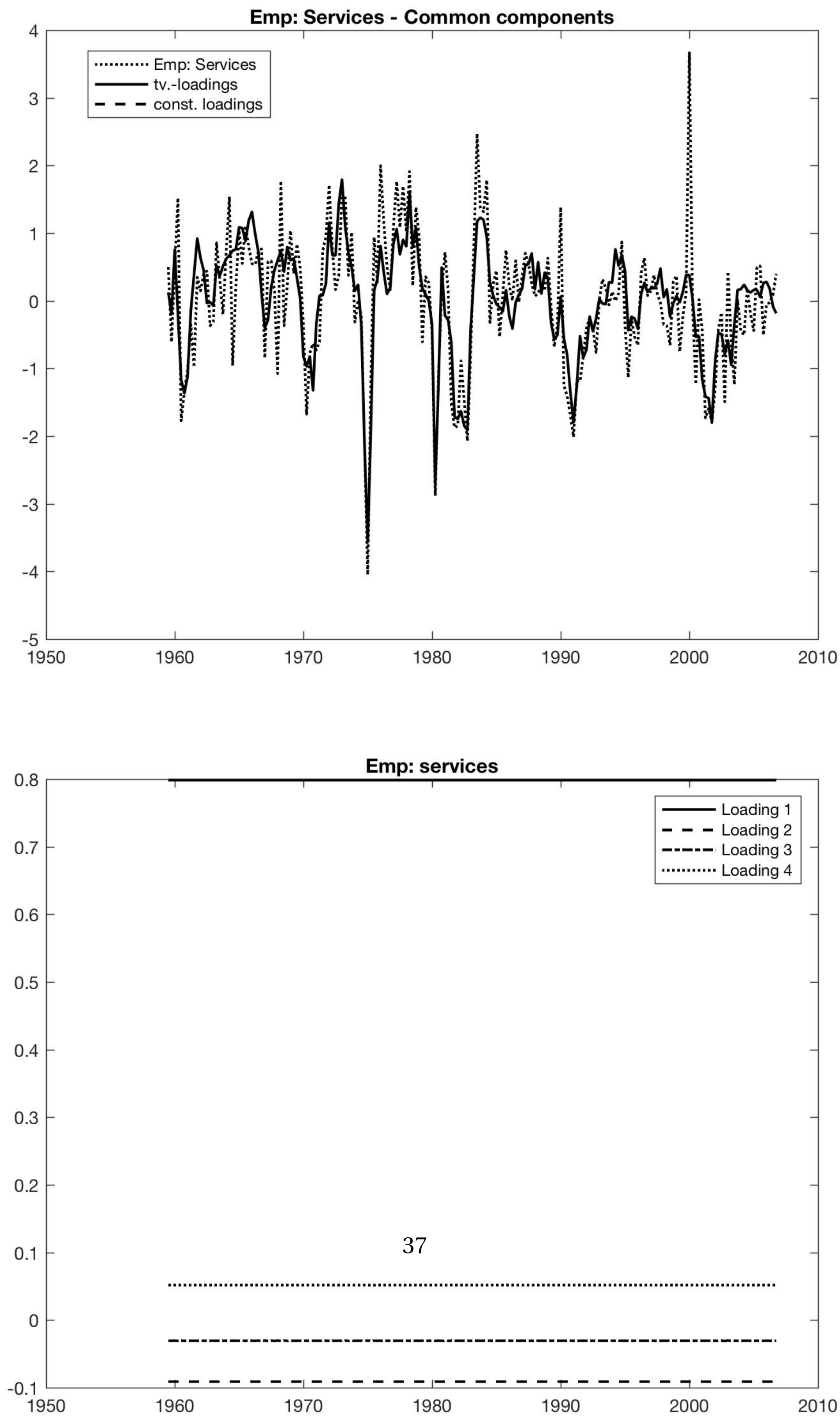


Figure 4: Common components and time-varying factor loadings for Number of employees in service industry.



References

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

Bai, J., Li, K., 2012. Statistical analysis of factor models of high dimension. The Annals of Statistics 40 (1), 436–465.

Bai, J., Li, K., 2016. Maximum likelihood estimation and inference for approximate factor models of high dimension. Review of Economics and Statistics 98 (2), 298–309.

Bai, J., Ng, S., 2002. Determining the number of factors in approximate factor models. Econometrica 70 (1), 191–221.

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

Bai, J., Ng, S., 2008a. Extremum estimation when the predictors are estimated from large panels. Annals of Economics and Finance 9 (2), 201–222.

Bai, J., Ng, S., 2008b. Large dimensional factor analysis. Foundations and Trends in Econometrics 3, 8–163.

Bai, J., Ng, S., 2013. Principal components estimation and identification of static factors. Journal of Econometrics 176 (1), 18–29.

Bates, B., Plagborg-Møller, M., Stock, J., Watson, M., 2013. Consistent factor estimation in dynamic factor models with structural instability. Journal of Econometrics 177, 289–304.

Bernanke, B., Boivin, J., Eliasz, P., 2005. Measuring the effects of monetary policy: A factor-augmented vector autoregressive (favar) approach. Quarterly Journal of Economics February, 387–422.

Breitung, J., Eickmeier, S., 2011. Testing for structural breaks in dynamic factor models. Journal of Econometrics 163 (1), 71–84.

Del Negro, M., Otrok, C., 2008. Dynamic factor models with time-varying parameters: measuring changes in international business cycles. Staff report 326, Federal Reserve Bank of New York, New York.

Doukhan, P., Neumann, M. H., 2007. Probability and moment inequalities for sums of weakly dependent random variables, with applications. Stochastic Processes and their Applications 117 (7), 878–903.

Doz, C., Giannone, D., Reichlin, L., 2011. A two-step estimator for large approximate dynamic factor models based on kalman filtering. Journal of Econometrics 164 (1), 188–205.

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

Eickmeier, S., Lemke, W., Marcellino, M., 2015. Classical time varying factor-augmented vector auto-regressive models – estimation, forecasting and structural analysis. Journal of the Royal Statistical Society: Series A (Statistics in Society) 178 (3), 493–533.

Forni, M., Hallin, M., Lippi, M., Reichlin, L., 2000. The generalized dynamic-factor model: identification and estimation. Review of Economics and Statistics 82, 540–554.

Liu, P., Mumtaz, H., Theophilopoulou, A., 2011. International transmission of shocks: A time-varying factor-augmented VAR approach to the open economy. Working paper 425, Bank of England, London.

Newey, W. K., McFadden, D., 1994. Large sample estimation and hypothesis testing. Handbook of Econometrics 4, 2111–2245.

Pagan, A., 1980. Some identification and estimation results for regression models with stochastically varying coefficients. Journal of Econometrics 13 (3), 341–363.

Stock, J., Watson, M., 1998. Diffusion indexes. Working paper, National Bureau of Economic Research.

Stock, J. H., Watson, M., 2009. Forecasting in dynamic factor models subject to structural instability. In: Hendry, D.F., Castle, J., Shephard, N. (Eds.), The Methodology and Practice of Econometrics. A Festschrift in Honour of David F. Hendry. Oxford University Press, Oxford, pp. 173–205.

Stock, J. H., Watson, M. W., 2002. Forecasting using principal components from a large number of predictors. Journal of the American Statistical Association 97 (460), 1167–1179.

Vershynin, R., 2012. Introduction to the non-asymptotic analysis of random matrices. In: Eldar, Y., Kutyniok, G. (Eds.), Compressed Sensing: Theory and Applications. Cambridge University Press, pp. 210–268.

White, H., 1982. Maximum likelihood estimation of misspecified models. Econometrica 50, 1–25.

A.1 Appendix

Let $X = (X_1, ..., X_T)'$ be the $T \times N$ matrix of observations, and let V_{NT} be the $r \times r$ diagonal matrix of the r largest eigenvalues of $(NT)^{-1}XX'$ in decreasing order. By the definition of eigenvalues and eigenvectors, we have $(NT)^{-1}XX'\tilde{F} = \tilde{F}V_{NT}$ or $(NT)^{-1}XX'\tilde{F}V_{NT}^{-1} = \tilde{F}$, where $\tilde{F}'\tilde{F}/T = I_r$. Let $H = (\Lambda^{0\prime}\Lambda^0/N)(F'\tilde{F}/T)V_{NT}^{-1}$ be the $r \times r$ rotation matrix. Assumption A and B together with Lemma A.1 below implies that $\|H\| = O_p(1)$. Let $w_t = \xi_t F_t$. We can write (1) as:

$$X_t = \Lambda^0 F_t + \xi_t F_t + e_t = \Lambda^0 F_t + w_t + e_t.$$

Define $e = (e_1, ..., e_T)'$ and $w = (w_1, ..., w_T)'$. We use the following expression from Bates et al. (2013):

$$XX' = F\Lambda^{0\prime}\Lambda^0 F' + F\Lambda^{0\prime}(e+w)' + (e+w)\Lambda^0 F' + (e+w)(e+w)'. \tag{A.1}$$

Let v_t denote a conforming unit vector with zeros in all entries except the t^{th}. We then have:

$$XX'v = F\Lambda^{0\prime}\Lambda^0 F_t + F\Lambda^{0\prime}(e_t + w_t) + (e+w)\Lambda^0 F_t + (e+w)(e_t + w_t).$$

Using the definition of $\tilde{F}_t$ and H, we can then write:

$$\begin{aligned}\tilde{F}_t - H'F_t &= V_{NT}^{-1}(NT)^{-1}\tilde{F}'XX'v - V_{NT}^{-1}(\tilde{F}'F/T)(\Lambda^{0\prime}\Lambda^0/N)F_t \\ &= V_{NT}^{-1}(NT)^{-1}\Big\{\tilde{F}'F\Lambda^{0\prime}e_t + \tilde{F}'e\Lambda^0 F_t + \tilde{F}'ee_t \\ &+ \tilde{F}'F\Lambda^{0\prime}w_t + \tilde{F}'w\Lambda^0 F_t + \tilde{F}'ww_t + \tilde{F}'ew_t + \tilde{F}'we_t\Big\}.\end{aligned}$$

Denote each term on the right-hand as $A_{1t}, ..., A_{8t}$, respectively. We get:

$$\tilde{F}_t - H'F_t = V_{NT}^{-1}\sum_{n=1}^{8} A_{nt}. \tag{A.2}$$

The following is a generalization of Lemma A.3 in Bai (2003). They consider constant loadings; we generalize the proof to autoregressive loadings.

Lemma A.1. *Under Assumptions A-E, as N,T $\to \infty$:*

(i) $\left\| V_{NT} - \frac{\tilde{F}'F}{T}\frac{\Lambda^{0\prime}\Lambda^0}{N}\frac{F'\tilde{F}}{T} \right\|^2 = O_p(C_{NT}^{-2})$,

(ii) $\frac{\tilde{F}'F}{T}\frac{\Lambda^{0\prime}\Lambda^0}{N}\frac{F'\tilde{F}}{T} \xrightarrow{p} V$, *where V is the diagonal matrix consisting of the eigenvalues of* $\Sigma_\Lambda \Sigma_F$.

Proof. From $V_{NT} = T^{-1}\tilde{F}'(NT)^{-1}XX'\tilde{F}$ we get using (A.1):

$$V_{NT} - \frac{\tilde{F}'F}{T}\frac{\Lambda^{0\prime}\Lambda^0}{N}\frac{F'\tilde{F}}{T} = T^{-1}\tilde{F}'(NT)^{-1}\left\{F\Lambda^{0\prime}(e+w)' + (e+w)\Lambda^0 F' + (e+w)(e+w)'\right\}\tilde{F}$$
$$= T^{-1}\sum_{t=1}^{T}\tilde{F}_t\sum_{n=1}^{8}A'_{nt}.$$

Hence,

$$\|T^{-1}\sum_{t=1}^{T}\tilde{F}_t\sum_{n=1}^{8}A'_{nt}\|^2 \le \left(T^{-1}\sum_{t=1}^{T}\|\tilde{F}_t\|^2\right)\left(T^{-1}\sum_{t=1}^{T}\left\|\sum_{n=1}^{8}A_{nt}\right\|^2\right)$$
$$\le 8rT^{-1}\sum_{t=1}^{T}\sum_{n=1}^{8}\|A_{nt}\|^2,$$

where the last inequality uses $tr(\tilde{F}'\tilde{F}/T) = tr(I_r) = r$ and Loève's inequality. The right-hand side is $O_p(C_{NT}^{-2})$ by Theorem 1 of Bates et al. (2013).

Statement (ii) is implicitly proven by Stock and Watson (1998). It should be noted that their paper considers the model $X_t = \Lambda^0 F_t + e_t$, i.e. a factor model with constant loadings. However, their proof only uses the asymptotic representation $V_{NT} = \frac{\tilde{F}'F}{T}\frac{\Lambda^{0\prime}\Lambda^0}{N}\frac{F'\tilde{F}}{T} + o_p(1)$ and the normalization $\tilde{F}'\tilde{F}/T = I_r$. Their proof is thus applicable for our model as well.

□

Proof of Lemma 1.

From (A.2) we have:

$$T^{-1}\sum_{t=1}^{T}\|\tilde{F}_t - H'F_t\|^2 \le \|V_{NT}^{-1}\|^2 8T^{-1}\sum_{t=1}^{T}\sum_{n=1}^{8}\|A_{nt}\|^2.$$

Since V_{NT} converges to a positive definite matrix, it follows that $\|V_{NT}^{-1}\|^2 = O_p(1)$. The right-hand side is thus $O_p(C_{NT}^{-2})$ by Theorem 1 in Bates et al. (2013).

□

Proof of Proposition 1. Using (A.2) we have:

$$\max_t \|\tilde{F}_t - H'F_t\| = \max_t \|V_{NT}^{-1} \sum_{n=1}^{8} A_{nt}\| \leq \|V_{NT}^{-1}\| \sum_{n=1}^{8} \max_t \|A_{nt}\|.$$

Lemma A.1 implies that $\|V_{NT}^{-1}\| = O_p(1)$. We can write A_{1t} as:[12]

$$A_{1t} = (NT)^{-1} \sum_{s=1}^{T} (\tilde{F}_s - H'F_s)F_s'\Lambda^{0\prime} e_t + (NT)^{-1} \sum_{s=1}^{T} H'F_s F_s' \Lambda^{0\prime} e_t.$$

The first term is less than:

$$\left(T^{-1} \sum_{s=1}^{T} \|\tilde{F}_s - H'F_s\|^2\right)^{1/2} \left(N^{-2} T^{-1} \sum_{s=1}^{T} \|F_s'\Lambda^{0\prime} e_t\|^2\right)^{1/2}.$$

We have:

$$N^{-2} T^{-1} \sum_{s=1}^{T} \|F_s'\Lambda^{0\prime} e_t\|^2 \leq N^{-1} \|N^{-1/2} \Lambda^{0\prime} e_t\|^2 T^{-1} \sum_{s=1}^{T} \|F_s\|^2.$$

By Assumption F.3, the maximum of $\|N^{-1/2}\Lambda^{0\prime} e_t\|^2$ over t is $O_p(T^{1/4})$, and Assumption A implies $T^{-1}\sum_{s=1}^{T} \|F_s\|^2 = O_p(1)$. By Lemma 1, we have $T^{-1}\sum_{s=1}^{T} \|\tilde{F}_s - H'F_s\|^2 = O_p(C_{NT}^{-2})$. Taking the square root then gives that the first term is $O_p\left(C_{NT}^{-1}\right) O_p\left(\frac{T^{1/8}}{N^{1/2}}\right)$. For the second term, we have:

$$(NT)^{-1} \|\sum_{s=1}^{T} H'F_s F_s'\Lambda^{0\prime} e_t\| \leq N^{-1/2} \|H\| \|N^{-1/2} \Lambda^{0\prime} e_t\| T^{-1} \sum_{s=1}^{T} \|F_s\|^2,$$

where $\|H\| = O_p(1)$ and $T^{-1}\sum_{s=1}^{T} \|F_s\|^2 = O_p(1)$ by Assumption A. The maximum of $\|N^{-1/2}\Lambda^{0\prime} e_t\|$ over t is $O_p(T^{1/8})$. The second term is thus equal to $O_p\left(\frac{T^{1/8}}{N^{1/2}}\right)$ and dominates the first.

Consider A_{2t}, which can be written as:

$$(NT)^{-1} \sum_{s=1}^{T} (\tilde{F}_s - H'F_s) e_s' \Lambda^0 F_t + (NT)^{-1} \sum_{s=1}^{T} H'F_s e_s' \Lambda^0 F_t.$$

The first term is bounded by

$$\left(T^{-1} \sum_{s=1}^{T} \|\tilde{F}_s - H'F_s\|^2\right)^{1/2} \left(N^{-2} T^{-1} \sum_{s=1}^{T} \|e_s'\Lambda^0 F_t\|^2\right)^{1/2}.$$

[12]The terms A_{1t}, A_{2t}, A_{3t} have been shown to be $O_p(\alpha_T T^{-1}) + O_p(T^{1/8})N^{-1/2}$ by Bai and Ng (2008a). They do, however, rely on intermediate results, which we have not proved for the model with time-varying loadings. We therefore provide an alternative proof for A_{1t}, A_{2t}, A_{3t}.

Now,

$$N^{-2}T^{-1}\sum_{s=1}^{T}\|e_s'\Lambda^0F_t\|^2 \leq \max_t\|F_t\|^2 N^{-1}T^{-1}\sum_{s=1}^{T}\|N^{-1/2}e_s'\Lambda^0\|^2 = O_p(\alpha_T^2)N^{-1}$$

by Assumption F.3. The first term is thus equal to $O_p(C_{NT}^{-1}\alpha_T N^{-1/2})$. The second term is equal to:

$$(NT)^{-1}\sum_{s=1}^{T}\sum_{i=1}^{N}H'F_se_{is}\lambda_i^{0'}F_t,$$

which is bounded by:

$$N^{-1/2}M\max_t\|F_t\|\|H\|\left(T^{-1}\sum_{s=1}^{T}\|F_s\|^2\right)^{1/2}\left((NT)^{-1}\sum_{s=1}^{T}\sum_{i,j=1}^{N}e_{is}e_{js}\right)^{1/2}.$$

This is equal to $O_p(\alpha_T)N^{-1/2}$ by Assumption C.3 and dominates the first term.

We can write A_{3t} as:

$$(NT)^{-1}\sum_{s=1}^{T}(\tilde{F}_s - H'F_s)[e_s'e_t - E(e_s'e_t)] + (NT)^{-1}\sum_{s=1}^{T}H'F_s[e_s'e_t - E(e_s'e_t)]$$
$$+ (NT)^{-1}\sum_{s=1}^{T}(\tilde{F}_s - H'F_s)E(e_s'e_t) + (NT)^{-1}\sum_{s=1}^{T}H'F_sE(e_s'e_t).$$

The first term is bounded by:

$$\left(T^{-1}\sum_{s=1}^{T}\|\tilde{F}_s - H'F_s\|^2\right)^{1/2}\left(N^{-1}T^{-1}\sum_{s=1}^{T}\left|N^{-1/2}\sum_{i=1}^{N}[e_{is}'e_{it} - E(e_{is}'e_{it})]\right|^2\right)^{1/2}.$$

By Assumption C.5, $\max_t\left|N^{-1/2}\sum_{i=1}^{N}[e_{is}'e_{it} - E(e_{is}'e_{it})]\right|^2 = O_p(\sqrt{T})$, so the first term is equal to $O_p(C_{NT}^{-1})O_p(\frac{T^{1/4}}{N^{1/2}})$. The second term is bounded by:

$$(NT)^{-1/2}\|H\|\|(NT)^{-1/2}\sum_{s=1}^{T}\sum_{i=1}^{N}F_s[e_{is}'e_{it} - E(e_{is}'e_{it})]\|.$$

By Assumption F.2, the maximum of this expression over t is $O_p(N^{-1/2})$. The third term is bounded by:

$$T^{-1/2}\left(T^{-1}\sum_{s=1}^{T}\|\tilde{F}_s - H'F_s\|^2\right)^{1/2}\left(\sum_{s=1}^{T}\gamma_N(s,t)^2\right)^{1/2}.$$

By Assumption F.1 and Lemma 1, this is equal to $T^{-1/2}O_p(C_{NT}^{-1})$. The fourth term is bounded by:

$$T^{-1}\max_t \|F_t\|\|H\|\sum_{s=1}^{T}|\gamma_N(s,t)|,$$

which is $O_p(\alpha_T T^{-1})$.

For A_{4t}, the term to be bound is

$$\max_t (NT)^{-1}\|\tilde{F}'F\Lambda^{0\prime}w_t\|,$$

where $\tilde{F}$ and F are $T\times r$, Λ^0 is $N\times r$, w_t is $N\times 1$. The vector w_t itself is given by $w_t = \xi_t F_t$, where ξ_t is $N\times r$ and F_t is $r\times 1$.

We begin by bounding

$$(NT)^{-1}\|\tilde{F}'F\Lambda^{0\prime}w_t\| \leq N^{-\frac{1}{2}}\|T^{-\frac{1}{2}}\tilde{F}\|\|T^{-\frac{1}{2}}F\|\|N^{-\frac{1}{2}}\Lambda^{0\prime}w_t\|.$$

The first terms $\|T^{-\frac{1}{2}}\tilde{F}\|$ and $\|T^{-\frac{1}{2}}F\|$ are $O_p(1)$ by construction of principal components and Assumption A. It thus remains to bound

$$\max_t \|N^{-\frac{1}{2}}\Lambda^{0\prime}w_t\|.$$

Note that

$$\Lambda^{0\prime}w_t = \left(\sum_{i=1}^{N}\lambda_{ip}^{0}w_{t,i}\right)_{p=1,\ldots,r}. \tag{A.3}$$

and thus

$$\|N^{-\frac{1}{2}}\Lambda^{0\prime}w_t\|^2 = N^{-1}\sum_{p=1}^{r}\left(\sum_{i=1}^{N}\lambda_{ip}^{0}w_{t,i}\right)\left(\sum_{i=1}^{N}\lambda_{ip}^{0}w_{t,i}\right).$$

Since by Assumption B,

$$\sum_{i=1}^{N}\lambda_{ip}^{0}w_{t,i} \leq M\sum_{i=1}^{N}w_{t,i},$$

we can write

$$\|N^{-\frac{1}{2}}\Lambda^{0\prime}w_t\|^2 \leq N^{-1}\sum_{p=1}^{r} M^2\left(\sum_{i=1}^{N} w_{t,i}\right)\left(\sum_{j=1}^{N} w_{t,j}\right),$$
$$= N^{-1} r M^2 \sum_{i=1}^{N}\sum_{j=1}^{N} w_{t,i} w_{t,j}.$$

Next, note that

$$w_t = \xi_t F_t = \left(\sum_{p=1}^{r} \xi_{itp} F_{tp}\right)_{i=1,\dots,N}.$$

It then follows that

$$w_{t,i} w_{t,j} = \left(\sum_{p=1}^{r} \xi_{itp} F_{tp}\right)\left(\sum_{p=1}^{r} \xi_{jtp} F_{tp}\right),$$
$$= \sum_{p=1}^{r}\sum_{q=1}^{r} \xi_{itp}\xi_{jtq}F_{tp}F_{tq},$$
$$\leq r^2 \max_{p,q}\left(\xi_{itp}\xi_{jtq}F_{tp}F_{tq}\right).$$

Denote the maximum by

$$\xi_{itp_1}\xi_{jtq_1}F_{tp_1}F_{tq_1} := \max_{p,q}\left(\xi_{itp}\xi_{jtq}F_{tp}F_{tq}\right).$$

Then,

$$\|N^{-\frac{1}{2}}\Lambda^{0\prime}w_t\|^2 \leq N^{-1} r^3 M^2 \sum_{i=1}^{N}\sum_{j=1}^{N} \xi_{itp_1}\xi_{jtq_1}F_{tp_1}F_{tq_1}.$$

In finding a bound for the maximum over t in expectation, the factor r^3M^2 can be ignored since it does not change the conclusion:

$$
\begin{aligned}
&E\left[\max_t\left(N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\xi_{itp_1}\xi_{jtq_1}F_{tp_1}F_{tq_1}\right)\right]\\
&=\frac{1}{u}E\left[\max_t\left(uN^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\xi_{itp_1}\xi_{jtq_1}F_{tp_1}F_{tq_1}\right)\right]\\
&=\frac{1}{u}E\left[\log\exp\left(\max_t uN^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\xi_{itp_1}\xi_{jtq_1}F_{tp_1}F_{tq_1}\right)\right],\\
&\leq\frac{1}{u}E\left[\log\sum_{t=1}^{T}\exp\left(uN^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\xi_{itp_1}\xi_{jtq_1}F_{tp_1}F_{tq_1}\right)\right],\\
&\leq\frac{1}{u}\log\sum_{t=1}^{T}E\left[\exp\left(uN^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\xi_{itp_1}\xi_{jtq_1}F_{tp_1}F_{tq_1}\right)\right],\\
&\leq\frac{1}{u}\log(TM)\text{ by Assumption F.4,}
\end{aligned}
$$

where the second to last inequality is the Jensen inequality. Thus, by the Markov inequality,

$$\max_t (NT)^{-1}\|\tilde{F}'F\Lambda^{0\prime}w_t\| = N^{-\frac{1}{2}}O_p(\log(T)^{\frac{1}{2}}).$$

Consider A_{5t}: The term to be bound is

$$\max_t (NT)^{-1}\|\tilde{F}'w\Lambda^0F_t\| = \max_t (NT)^{-1}\|\sum_{s=1}^{T}\tilde{F}_s w_s'\Lambda^0 F_t\|,$$

where $w=(w_1,\dots,w_T)'$ is $T\times N$. We begin by bounding

$$(NT)^{-1}\|\sum_{s=1}^{T}\tilde{F}_s w_s'\Lambda^0F_t\| \leq N^{-\frac{1}{2}}\max_t\|F_t\|\left(T^{-1}\sum_{s=1}^{T}\|\tilde{F}_s\|^2\right)^{\frac{1}{2}}\left(T^{-1}\sum_{s=1}^{T}\|N^{-\frac{1}{2}}w_s'\Lambda^0\|^2\right)^{\frac{1}{2}}.$$

The object $\max_t\|F_t\|$ is $O_p(\alpha_T)$, as assumed in Proposition 1 and discussed in the subsequent text. The term involving $\tilde{F}$ is $O(1)$, by construction of principal components. The term to be controlled is

$$T^{-1}\sum_{s=1}^{T}\|N^{-\frac{1}{2}}w_s'\Lambda^0\|^2,$$

where the summand is exactly the term studied for A_{4t}. We thus obtain

$$\begin{aligned}
E\left[T^{-1}\sum_{s=1}^{T}\|N^{-\frac{1}{2}}w_s'\Lambda^0\|^2\right] &\le (uT)^{-1}\sum_{s=1}^{T}r^3M^2E\left[uN^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\xi_{itp_1}\xi_{jtq_1}F_{tp_1}F_{tq_1}\right],\\
&\le r^3M^2(uT)^{-1}\sum_{s=1}^{T}\log\left[E\exp\left(uN^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\xi_{itp_1}\xi_{jtq_1}F_{tp_1}F_{tq_1}\right)\right],\\
&\le r^3M^2(uT)^{-1}\sum_{s=1}^{T}\log M=\frac{1}{u}r^3M^2\log M,
\end{aligned}$$

by Assumption F.4. Thus, by the Markov inequality,

$$\max_t (NT)^{-1}\|\tilde{F}'w\Lambda^0F_t\| = O_p(\alpha_t N^{-\frac{1}{2}}).$$

For A_{6t}, the term to be bound is

$$\max_t (NT)^{-1}\|\sum_{s=1}^{T}\tilde{F}_s w_s' w_t\|.$$

We begin by bounding

$$\max_t (NT)^{-1}\|\sum_{s=1}^{T}\tilde{F}_s w_s' w_t\| \le N^{-1}T^{-\frac{1}{2}}\left(T^{-1}\sum_{s=1}^{T}\|\tilde{F}_s\|^2\right)^{\frac{1}{2}}\left(\sum_{s=1}^{T}\|w_s'w_t\|^2\right)^{\frac{1}{2}}.$$

The first term on the right-hand side is $O(1)$ by construction of principal components. The term to be controlled is

$$\max_t\left(N^{-2}T^{-1}\sum_{s=1}^{T}\|w_s'w_t\|^2\right)^{\frac{1}{2}}.$$

Since

$$w_t=\xi_tF_t=\left(\sum_{p=1}^{r}\xi_{itp}F_{tp}\right)_{i=1,\ldots,N},$$

we have that

$$w_{t,i}w_{s,i} = \left(\sum_{p=1}^{r} \xi_{itp}F_{tp}\right)\left(\sum_{p=1}^{r} \xi_{isp}F_{sp}\right) = \sum_{p=1}^{r}\sum_{q=1}^{r} \xi_{itp}\xi_{isq}F_{tp}F_{sq},$$

and

$$w_s'w_t = \sum_{i=1}^{N}\sum_{p=1}^{r}\sum_{q=1}^{r} \xi_{itp}\xi_{isq}F_{tp}F_{sq}.$$

Thus,

$$\|w_s'w_t\|^2 = \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{p_i=1}^{r}\sum_{q_i=1}^{r}\sum_{p_j=1}^{r}\sum_{q_j=1}^{r} \xi_{itp_i}\xi_{isq_i}\xi_{jtp_j}\xi_{jsq_j}F_{tp_i}F_{sq_i}F_{tp_j}F_{sq_j}.$$

Let

$$\xi_{itp_1}\xi_{isq_1}\xi_{jtp_2}\xi_{jsq_2}F_{tp_1}F_{sq_1}F_{tp_2}F_{sq_2} := \max_{p_i,q_i,p_j,q_j} \xi_{itp_i}\xi_{isq_i}\xi_{jtp_j}\xi_{jsq_j}F_{tp_i}F_{sq_i}F_{tp_j}F_{sq_j}.$$

Then,

$$\|w_s'w_t\|^2 \le r^4 \sum_{i=1}^{N}\sum_{j=1}^{N} \xi_{itp_1}\xi_{isq_1}\xi_{jtp_2}\xi_{jsq_2}F_{tp_1}F_{sq_1}F_{tp_2}F_{sq_2}$$

and, ignoring r^4 since it does not influence the conclusion,

$$\begin{aligned}
&E\left(\max_t N^{-2}T^{-1}\sum_{s=1}^{T}\sum_{i,j=1}^{N} \xi_{itp_1}\xi_{isq_1}\xi_{jtp_2}\xi_{jsq_2}F_{tp_1}F_{sq_1}F_{tp_2}F_{sq_2}\right)\\
&= \frac{N^2+NT}{uN^2T} E\left(\max_t \frac{u}{N^2+NT}\sum_{s=1}^{T}\sum_{i,j=1}^{N} \xi_{itp_1}\xi_{isq_1}\xi_{jtp_2}\xi_{jsq_2}F_{tp_1}F_{sq_1}F_{tp_2}F_{sq_2}\right),\\
&\le \frac{N^2+NT}{uN^2T} \log E\left[\exp\left(\max_t \frac{u}{N^2+NT}\sum_{s=1}^{T}\sum_{i,j=1}^{N} \xi_{itp_1}\xi_{isq_1}\xi_{jtp_2}\xi_{jsq_2}F_{tp_1}F_{sq_1}F_{tp_2}F_{sq_2}\right)\right],\\
&\le \frac{N^2+NT}{uN^2T} \log\left[\sum_{t=1}^{T} E\exp\left(\frac{u}{N^2+NT}\sum_{s=1}^{T}\sum_{i,j=1}^{N} \xi_{itp_1}\xi_{isq_1}\xi_{jtp_2}\xi_{jsq_2}F_{tp_1}F_{sq_1}F_{tp_2}F_{sq_2}\right)\right],\\
&\le \frac{N^2+NT}{uN^2T}\log(TM) = \frac{\log(TM)}{uT} + \frac{\log(TM)}{uN}
\end{aligned}$$

by Assumption F.5. The third to last inequality is again Jensen's. Thus, by the Markov inequality,

$$\max_t (NT)^{-1} \| \sum_{s=1}^{T} \tilde{F}_s w_s' w_t \| = O_p\left(\frac{\log(T)}{T} + \frac{\log(T)}{N}\right)^{\frac{1}{2}}.$$

The seventh term A_{7t} is bounded by:

$$N^{-1/2}\left(T^{-1}\sum_{s=1}^{T}\|\tilde{F}_s\|^2\right)^{1/2}\left(N^{-1}T^{-1}\sum_{s=1}^{T}\|e_s' w_t\|^2\right)^{1/2} = r^{1/2}N^{-1/2}\left(T^{-1}\sum_{s=1}^{T}\|N^{-1/2}\sum_{i=1}^{N}e_{is}w_{it}\|^2\right)^{1/2}$$

The maximum over t of the term inside the parenthesis can be bounded in expectation:

$$E\left(\max_{\text{t}} T^{-1}\sum_{s=1}^{T}\|N^{-1/2}\sum_{i=1}^{N}e_{is}w_{it}\|^2\right) = \frac{1}{u}E\left(\log\exp\max_{\text{t}}\frac{u}{T}\sum_{s=1}^{T}\|N^{-1/2}\sum_{i=1}^{N}e_{is}w_{it}\|^2\right)$$
$$\leq \frac{1}{u}\log\left(\sum_{t=1}^{T}\left(E\exp\frac{u}{T}\sum_{s=1}^{T}\|N^{-1/2}\sum_{i=1}^{N}e_{is}w_{it}\|^2\right)\right) \leq \log\left(\sum_{t=1}^{T}M\right) = \log(TM),$$

by Assumption E. The seventh term is thus $\max_{\text{t}}\|A_{7t}\| = N^{-1/2}O_p(\log(T)^{1/2})$.

Finally, we have for A_{8t}:

$$N^{-1/2}\left(T^{-1}\sum_{s=1}^{T}\|\tilde{F}_s\|^2\right)^{1/2}\left(N^{-1}T^{-1}\sum_{s=1}^{T}\|w_s' e_t\|^2\right)^{1/2} = r^{1/2}N^{-1/2}\left(T^{-1}\sum_{s=1}^{T}\|N^{-1/2}\sum_{i=1}^{N}e_{it}w_{is}\|^2\right)^{1/2}.$$

The last term is bounded by:

$$E\left(\max_{\text{t}} T^{-1}\sum_{s=1}^{T}\|N^{-1/2}\sum_{i=1}^{N}e_{it}w_{is}\|^2\right) = \frac{1}{u}E\left(\log\exp\max_{\text{t}}\frac{u}{T}\sum_{s=1}^{T}\|N^{-1/2}\sum_{i=1}^{N}e_{it}w_{is}\|^2\right)$$
$$\leq \frac{1}{u}\log\left(\sum_{t=1}^{T}\left(E\exp\frac{u}{T}\sum_{s=1}^{T}\|N^{-1/2}\sum_{i=1}^{N}e_{it}w_{is}\|^2\right)\right) \leq \log\left(\sum_{t=1}^{T}M\right) = \log(TM),$$

again by Assumption E. The last term is thus $N^{-1/2}O_p(\log(T)^{1/2})$. All terms are dominated by $O_p\left(\frac{T^{1/8}}{N^{1/2}}\right) + O_p(\alpha_T N^{-1/2}) + O_p(\alpha_T T^{-1}) + O_p\left(\frac{\log(T)}{N}\right)^{1/2} + O_p\left(\frac{\log(T)}{T}\right)^{1/2}$. If we take $\alpha_T = O(1)$, these terms are dominated by $O_p\left(\frac{T^{1/8}}{N^{1/2}}\right) + O_p\left(\frac{\log(T)}{T}\right)^{1/2}$, as $O_p\left(\frac{\log(T)}{N}\right)^{1/2}$ is dominated by $O_p(\frac{T^{1/8}}{N^{1/2}})$. Proposition 1 follows.

□

Lemma A.2. *Let Assumption A-E hold. If $F'F/T = I_r$ and $\Lambda^{0\prime}\Lambda^0$ is a diagonal matrix with distinct entries,*

$$H = I_r + O_p(C_{NT}^{-2}).$$

Proof. First we need to show that $(\tilde{F} - FH)'F/T$ and $(\tilde{F} - FH)'\tilde{F}/T$ are both $O_p(C_{NT}^{-2})$. We have:

$$\|(\tilde{F} - FH)'F/T\|^2 = \|T^{-1}\sum_{t=1}^{T}(\tilde{F}_t - H'F_t)F_t'\|^2$$
$$\leq \left(T^{-1}\sum_{t=1}^{T}\|\tilde{F}_t - H'F_t\|^2\right)\left(T^{-1}\sum_{t=1}^{T}\|F_t\|^2\right) = O_p(C_{NT}^{-2}),$$

where the last equality follows from Lemma 1 and Assumption A. By similar arguments $(\tilde{F} - FH)'\tilde{F}/T = O_p(C_{NT}^{-2})$. The rest of the proof is identical to the proof of equation (2) in Bai and Ng (2013).

□

Lemma A.2 shows that if the imposed normalization holds for the process generating the data, the factors can be estimated without rotation. This implies that θ_i can be estimated without rotation as well. In the proof of Theorem 1 below, we assume that $H = I_r$ for notational convenience only. In general, the feasible likelihood converges to $\mathcal{L}_T(X_i|FH;\theta_i)$, and $\tilde{\theta}_i$ is consistent for a rotation of θ_i^0 as discussed in Section 3.2.

Proof of Theorem 1. It suffices to show that the feasible likelihood function $\tilde{\mathcal{L}}_T(X_i|\tilde{F};\theta_i)$ converges uniformly to the infeasible one $\mathcal{L}_T(X_i|F;\theta_i)$.

This will imply that $\tilde{\mathcal{L}}_T(X_i|\tilde{F};\theta_i)$ satisfies the conditions of Assumption H and $\tilde{\theta}_i \xrightarrow{p} \theta_i^0$. We thus need:

$$\sup_{\theta_i\in\Theta_i}\left|\tilde{\mathcal{L}}_T(X_i|\tilde{F};\theta_i) - \mathcal{L}_T(X_i|F;\theta_i)\right| \xrightarrow{p} 0.$$

By the mean value expansion, we can write:

$$\tilde{\mathcal{L}}_T(X_i|\tilde{F};\theta_i) = \mathcal{L}_T(X_i|F;\theta_i) + \sum_{t=1}^{T}\nabla_{F_t}\mathcal{L}_T(X_i|F^*;\theta_i)(\tilde{F}_t - F_t),$$

where $\nabla_{F_t}\mathcal{L}_T(X_i|F^*;\theta_i) = \frac{\partial\mathcal{L}_T(X_i|F;\theta_i)}{\partial F_t}\Big|_{F=F^*}$, and F^* is between F and $\tilde{F}$. For uniform convergence the last term needs to be $o_p(1)$ uniformly in Θ_i, when F_t^* is in a neighbourhood of F_t, such that $\max_t \|F_t^* - F_t\| = o_p(1)$.

Let $\lambda_{\max}(A)$ and $\lambda_{\min}(A)$ denote the largest and smallest eigenvalue of a matrix A, and let $(A)_{(s,t)}$ denote entry (s,t) of a $T \times T$ matrix A. Furthermore, let ϕ_i be the $r \times r$ block matrix on the diagonal of Φ_i, i.e. $\phi_i = Var(\lambda_{it})$. The derivative of $\mathcal{L}_T(X_i|F;\theta_i)$ takes the form:[13,14]

$$\begin{aligned}\nabla_{F_t}\mathcal{L}_T(X_i|F;\theta_i)' = &- T^{-1}\phi_i F_t \Sigma_{i,(t,t)}^{-1} + T^{-1}\lambda_i \sum_{s=1}^{T}(X_{is} - F_s'\lambda_i)\Sigma_{i,(s,t)}^{-1} \\ &+ T^{-1}\phi_i F_t \left(\Sigma_i^{-1}(X_i - E(X_i))(X_i - E(X_i))'\Sigma_i^{-1}\right)_{t,t},\end{aligned}$$

where F_t is to be evaluated at F_t^*. Denote the three terms above by B_{nt}, for $n = 1,...,3$. We can then write:

$$\begin{aligned}\sup_{\theta_i\in\Theta_i}\left|\tilde{\mathcal{L}}_T(X_i|\tilde{F};\theta_i) - \mathcal{L}_T(X_i|F;\theta_i)\right| &= \sup_{\theta_i\in\Theta_i}\left|\sum_{t=1}^{T}\sum_{n=1}^{3}(\tilde{F}_t - F_t)'B_{nt}\right| \\ \leq sup_{\theta_i\in\Theta_i}\left|\sum_{t=1}^{T}(\tilde{F}_t - F_t)'B_{1t}\right| &+ \sup_{\theta_i\in\Theta_i}\left|\sum_{t=1}^{T}(\tilde{F}_t - F_t)'B_{2t}\right| + \sup_{\theta_i\in\Theta_i}\left|\sum_{t=1}^{T}(\tilde{F}_t - F_t)'B_{3t}\right|.\end{aligned} \tag{A.4}$$

For the term involving B_{1t}, we have:

$$\left|T^{-1}\sum_{t=1}^{T}(\tilde{F}_t - F_t)'\phi_i F_t^* \Sigma_{i,(t,t)}^{-1}\right| \leq \lambda_{\max}(\Sigma_i^{-1})T^{-1}\sum_{t=1}^{T}\|\tilde{F}_t - F_t\|\|\phi_i\|\|F_t^*\|, \tag{A.5}$$

since each entry in Σ_i^{-1} is bounded by the largest eigenvalue. For the largest eigenvalue of Σ_i^{-1}, we have $\lambda_{\max}(\Sigma_i^{-1}) = [\lambda_{\min}(\Sigma_i)]^{-1}$, and it therefore follows from the Weyl inequality that $\lambda_{\max}(\Sigma_i^{-1}) \leq M$ as:[15]

$$\lambda_{\min}(\Sigma_i) \geq \lambda_{\min}(\mathbf{F}\Phi_i\mathbf{F}') + \lambda_{\min}(\psi_i I_T) \geq \psi_i > 0$$

uniformly in Θ_i. The term $\|\phi_i\|$ is also uniformly bounded, as the parameters of $B_i(L)$ are in the stationary region, and the elements of Q_i are bounded. We can therefore bound (A.5) by:

$$O(1)T^{-1}\sum_{t=1}^{T}\|\tilde{F}_t - F_t\|\|F_t^* - F_t\| + O(1)T^{-1}\sum_{t=1}^{T}\|\tilde{F}_t - F_t\|\|F_t\|.$$

[13]The calculations of the derivative are omitted for brevity. They are available upon request.

[14]With autocorrelated errors, the derivative takes the same form, but the variance matrix is $\Sigma_i = F\Phi_i F' + \Psi_i$, where $\Psi_i = E(e_i e_i')$ is non-diagonal.

[15]This also holds with $\Psi_i = E(e_i e_i')$ non-diagonal, as we can bound the smallest eigenvalue of Σ_i uniformly i Θ_i.

Since F_t^* is between F_t and $\tilde{F}_t$, the first term is less than $T^{-1}\sum_t \|\tilde{F}_t - F_t\|^2$ and is $O_p(C_{NT}^{-2})$ by Lemma 1. Note that $T^{-1}\sum_t \|\tilde{F}_t - F_t\|^2$ does not depend on θ_i, and the result is thus uniform in Θ_i. For the second term, we can write:

$$T^{-1}\sum_{t=1}^{T}\|\tilde{F}_t - F_t\|\|F_t\| \leq \left(T^{-1}\sum_{t=1}^{T}\|\tilde{F}_t - F_t\|^2\right)^{1/2}\left(T^{-1}\sum_{t=1}^{T}\|F_t\|^2\right)^{1/2},$$

which is $O_p(C_{NT}^{-1})$ by Lemma 1 and Assumption A, also uniformly in Θ_i.

For the term involving B_{3t} in (A.4), we can write:

$$\left|T^{-1}\sum_{t=1}^{T}(\tilde{F}_t - F_t)'\phi_i F_t^*\left(\Sigma_i^{-1}(X_i - E(X_i))(X_i - E(X_i))'\Sigma_i^{-1}\right)_{t,t}\right| \leq$$
$$\max_t\left|(\tilde{F}_t - F_t)'\phi_i F_t^*\right| T^{-1}\left|\sum_{t=1}^{T}\left(\Sigma_i^{-1}(X_i - E(X_i))(X_i - E(X_i))'\Sigma_i^{-1}\right)_{t,t}\right|.$$

For the term outside the sum, we have:

$$\max_t\left|(\tilde{F}_t - F_t)'\phi_i F_t^*\right| \leq \|\phi_i\|\max_t\|\tilde{F}_t - F_t\|\|F_t^*\|$$
$$\leq O(1)\max_t\|\tilde{F}_t - F_t\|^2 + O(1)\max_t\|\tilde{F}_t - F_t\|\|F_t\|.$$

If we take F_t to be a sequence of fixed and bounded constants, $\max_t\|F_t\| \leq M$, and the second term is then $o_p(1)$ by Proposition 1, which is uniform in Θ_i as the proof of Proposition 1 does not depend on θ_i. The first term is bounded by the second.

The term involving the sum can be written as

$$\begin{aligned} &T^{-1}\left|\sum_{t=1}^{T}\left(\Sigma_i^{-1}(X_i - E(X_i))(X_i - E(X_i))'\Sigma_i^{-1}\right)_{t,t}\right| \\ &= T^{-1}\left|\operatorname{tr}\left(\Sigma_i^{-1}(X_i - E(X_i))(X_i - E(X_i))'\Sigma_i^{-1}\right)\right|, \end{aligned} \tag{A.6}$$

which is bounded by

$$\lambda_{\max}(\Sigma_i^{-2})T^{-1}|\operatorname{tr}(X_i - E(X_i))(X_i - E(X_i))'| \leq M^2 T^{-1}\sum_{t=1}^{T}\|X_{it} - F_t^{*\prime}\lambda_i\|^2$$
$$\leq 4M^2 T^{-1}\sum_{t=1}^{T}\left(\|F_t'\lambda_i^0\|^2 + \|F_t'(\lambda_{it} - \lambda_i^0)\|^2 + \|e_{it}\|^2 + \|F_t^{*\prime}\lambda_i\|^2\right).$$

The first term in the sum is bounded by $T^{-1}M^2\sum_{t=1}^T \|F_t\|^2 = O_p(1)$. For the second term in the sum, we can write:

$$T^{-1}\sum_{t=1}^T \|F_t'(\lambda_{i,t}-\lambda_i^0)\|^2 \le \left(T^{-1}\sum_{t=1}^T \|F_t\|^4\right)^{1/2}\left(T^{-1}\sum_{t=1}^T \|\lambda_{i,t}-\lambda_i^0\|^4\right)^{1/2}.$$

This is $O_p(1)$ by Assumption A and G. By Assumption C we have $T^{-1}\sum_{t=1}^T e_{it}^2 = O_p(1)$, and for the last term, we can write:

$$T^{-1}\sum_{t=1}^T \|F_t^{*\prime}\lambda_i\|^2 \le M^2T^{-1}\sum_{t=1}^T \|F_t^{*\prime}-F_t\|^2 + M^2T^{-1}\sum_{t=1}^T \|F_t\|^2 = O_p(C_{NT}^{-2}) + O_p(1),$$

as λ_i is estimated in a bounded parameter space. The second term in (A.4) is thus $\max_t\|\tilde{F}_t - F_t\| O_p(1) = o_p(1)$ uniformly in Θ_i.

For the term involving B_{2t} in (A.4), we can write:

$$\begin{aligned}&\left|T^{-1}\sum_{t=1}^T(\tilde{F}_t-F_t)'\lambda_i\sum_{s=1}^T(X_{is}-F_s^{*\prime}\lambda_i)\Sigma_{i,(s,t)}^{-1}\right|\\ &\le\left(T^{-1}\sum_{t=1}^T|(\tilde{F}_t-F_t)'\lambda_i|^2\right)^{1/2}\left(T^{-1}\sum_{t=1}^T\left|\sum_{s=1}^T(X_{is}-F_s^{*\prime}\lambda_i)\Sigma_{i,(s,t)}^{-1}\right|^2\right)^{1/2}.\end{aligned}$$

The first term in parentheses is less than $M^2T^{-1}\sum_{t=1}^T\|\tilde{F}_t-F_t\|^2 = O_p(C_{NT}^{-2})$ uniformly in Θ_i. The second term in parentheses is equal to

$$T^{-1}\left|\mathrm{tr}\left(\Sigma_i^{-1}(X_i-E(X_i))(X_i-E(X_i))'\Sigma_i^{-1}\right)\right|,$$

which is $O_p(1)$ uniformly in Θ_i from the arguments above, see (A.6). By taking the square root, the second term is thus $O_p(C_{NT}^{-1})$ and dominated by the third. Collecting the results gives:

$$\sup_{\theta_i\in\Theta_i}\left|\tilde{\mathcal{L}}_T(X_i|\tilde{F};\theta_i)-\mathcal{L}_T(X_i|F;\theta_i)\right| = O_p\left(\max_t\|\tilde{F}_t-F_t\|\right) = o_p(1).$$

□