A two-step estimator for large approximate dynamic factor models based on Kalman filtering*

Catherine Doz, Université de Cergy-Pontoise, THEMA
Domenico Giannone, Université Libre de Bruxelles, ECARES and CEPR
Lucrezia Reichlin, European Central Bank, ECARES and CEPR

Revised version

February 19, 2009

Abstract

This paper shows consistency of a two step estimation of the factors in a dynamic approximate factor model when the panel of time series is large (n large). In the first step, the parameters of the model are first estimated from an OLS on principal components. In the second step, the factors are estimated via the Kalman smoother. This projection allows to consider dynamics in the factors and in the idiosyncratic component, and heteroscedasticity in the idiosyncratic variance. The analysis provides theoretical backing for the estimator considered in Giannone, Reichlin, and Sala (2004) and Giannone, Reichlin, and Small (2005).

JEL Classification: C51, C32, C33.

Keywords: Factor Models, Kalman filter, principal components, large cross-sections.

* We would like to thank Ursula Gather, Marco Lippi and Ricardo Mestre for helpful suggestions and seminar participants at the International Statistical Institute in Berlin 2003, the European Central Bank, 2003, the Statistical Institute at the Catholic University of Louvain la Neuve, 2004, the Institute for Advanced Studies in Vienna, 2004, the Department of Statistics in Madrid, 2004. The opinions in this paper are those of the authors and do not necessarily reflect the views of the European Central Bank. Please address any comments to Catherine Doz doz@u-cergy.fr; Domenico Giannone dgiannon@ulb.ac.be; or Lucrezia Reichlin lreichlin@london.edu

1 Introduction

We consider a "large" panel of time series and assume that it can be represented by an approximate factor structure whereby the dynamics of each series is split in two orthogonal components – one capturing the bulk of cross-sectional comovements and driven by few common factors and the other being composed of poorly cross-correlated elements. This model has been introduced by Chamberlain and Rothschild (1983) and generalized to a dynamic framework by Forni, Hallin, Lippi, and Reichlin (2000); Forni and Lippi (2001) and Stock and Watson (2002a,b).

As in many other papers in the literature, this paper studies the estimation of the common factors and consistency and rates for the size of the cross-section n and the sample size T going to infinity.

The literature has extensively studied the particular case in which the factors are estimated by principal components (Bai, 2003; Bai and Ng, 2002; Forni, Hallin, Lippi, and Reichlin, 2005b; Forni, Giannone, Lippi, and Reichlin, 2005a; Stock and Watson, 2002a,b). It has been shown that the latter are (n, T) consistent estimates of a rotation of the factors. Consistency is achieved even if principal components do not exploit likely features of the data generating process, such as heterogeneous signal to noise ratio (cross-sectional heteroscedasticity of the idiosyncratic component), dynamic of the factors and dynamic in the idiosyncratic component.

The literature has also studied a number of methods to exploit those features. Forni, Hallin, Lippi, and Reichlin (2005b) has proposed a two-step approach based on principal components in the frequency domain to exploit, when extracting the common factors, the cross-sectional heteroscedasticity of the idiosyncratic component and the dynamic properties of the data; Boivin and Ng (2003) and Forni and Reichlin (2001) have used iteratively re-weighted principal components and Boivin and Ng (2005), D'Agostino and Giannone (2005), Stock and Watson (2005) have studied the empirical relevance of such efficiency improvements. Finally, Giannone, Reichlin, and Sala (2004) and Giannone, Reichlin, and Small (2005) have introduced a parametric time domain two-step estimator involving principal components and Kalman filter to exploit both factor dynamics and idiosyncratic heteroscedacticity.

In this paper, we parameterize the dynamics of the factors as in Forni, Giannone, Lippi, and Reichlin (2005a) and we show the consistency of such a two-step estimator in this case. The parameters of the model can then be estimated by simple least squares by treating the principal components as if they were the true common factors. These estimated parameters can be used to project onto the observations. We consider three cases, each corresponding to an estimator under different forms of misspecification: factor dynamics, idiosyncratic heteroscedacticity and idiosyncratic dynamics (principal components); factor and idiosyncratic dynamics (reweighted principal components); idiosyncratic dynamics only (Kalman smoother). Each projection corresponds to a different two-step estimator whereby the first step involves the estimation of the parameters and the second step the application of the Kalman smoother. We prove consistency for such estimators and design a Monte-Carlo exercise that allows to study the behavior of our estimators in small samples.

We should stress that the use of the Kalman smoother, beside achieving possible efficiency improvements, allows useful empirical applications. First, the treatment of unbalanced panels, particularly interesting for forecasting current quarter GDP at dates in which not all data included in the panel are released (see Giannone, Reichlin, and Sala, 2004; Giannone, Reichlin, and Small, 2005). Second, "cleaning", through the second step, the estimate of the factors, allows a better reconstruction of the common shocks considered in the structural factor model by Giannone, Reichlin, and Sala (2004). Finally, such parametric approach allows to easily evaluate uncertainty in the estimates of the factors as shown in both the papers just cited.

Let us finally note that similar reasoning to that applied to this paper can be applied to use principal components to initialize the algorithm for maximum likelihood estimation. We study consistency of maximum likelihood estimator in a separate paper Doz, Giannone, and Reichlin (2005).

The paper is organized as follows. Section two introduces models and assumptions. Section three analyzes the projections, for known parameters, and for the different misspecified model assumptions : we show that the extracted factors are root n consistent in each case. Section four contains the main propositions which show consistency and (n, T) rates for the two step estimators. Section five presents the Monte-Carlo exercise. Section six concludes. Proofs are gathered in the appendix.

2 The Models

We consider the following model:

$$X_t = \Lambda_0^* F_t + \xi_t$$

where:

$X_t = (x_{1t}, ..., x_{nt})'$ is a $(n \times 1)$ stationary process

$\Lambda_0^* = (\lambda_{0,ij}^*)$ is the $n \times r$ matrix of factor loadings

$F_t = (f_{1t}, ..., f_{rt})'$ is a $(r \times 1)$ stationary process (common factors)

$\xi_t = (\xi_{1t}, ..., \xi_{nt})'$ is a $(n \times 1)$ stationary process (idiosyncratic component)

(F_t) and (ξ_t) are two independent processes

Note that X_t, Λ_0^*, ξ_t depend on n but, in this paper, we drop the subscript for sake of simplicity.

The general idea of the model is that the observable variables can be decomposed in two orthogonal unobserved processes: the common component driven by few common shocks which captures the bulk of the covariation between the time series, and the idiosyncratic component which is driven by n shocks generating dynamics which is series specific or local.

We have the following decomposition of the covariance matrix of the observables:

$$\Sigma_0 = \Lambda_0^* \Phi_0^* \Lambda_0^{*\prime} + \Psi_0$$

where $\Psi_0 = \mathrm{E}[\xi_t \xi_t']$ and $\Phi_0^* = \mathrm{E}[F_t F_t']$. It is well-known that the factors are defined up to a pre-multiplication by an invertible matrix, so that it is possible to choose $\Phi_0^* = I_r$: we will maintain this assumption throughout the paper. Even in this case, the factors are defined up to a pre-multiplication by an orthogonal matrix, a point that we make more precise below.

We also have the following decomposition of the auto-covariance matrix of order h of the observables:

$$\Sigma_0(h) = \Lambda_0^* \Phi_0^*(h) \Lambda_0^{*\prime} + \Psi_0(h)$$

where $\Sigma_0(h) = \mathrm{E}[X_t X_{t-h}']$, $\Phi_0^*(h) = \mathrm{E}[F_t F_{t-h}']$, and $\Psi_0(h) = \mathrm{E}[\xi_t \xi_{t-h}']$.

This decomposition extends the previous one, if we adopt the following notations:

$$\Sigma_0(0) = \Sigma_0, \ \Phi_0^*(0) = I_r, \ \text{and } \Psi_0(0) = \Psi_0.$$

Remark 1: Bai (2003); Bai and Ng (2002) and Stock and Watson (2002a) consider also some form of non-stationarity. Here we do not do it for simplicity. The main arguments used in what follows still hold under the assumption of weak time dependence of the common and the idiosyncratic component.

More precisely, we make the following set of assumptions:

(A1) *For any n, (X_t) is a stationary process with zero mean and finite second order moments.*

(A2) *The x_{it}'s have uniformly bounded variance : $\exists M / \forall (i,t) V x_{it} = \sigma_{0,ii} \leq M$*

(A3) - *(F_t) and (ξ_t) are independent processes.*
- *(F_t) admits a Wold representation: $F_t = C_0(L)\varepsilon_t = \sum_{k=0}^{+\infty} C_k \varepsilon_{t-k}$ such that: $\sum_{k=0}^{+\infty} \|C_k\| < +\infty$, and ε_t is stationary at order four.*
- *For any n, (ξ_t) admits a Wold representation: $\xi_t = D_0(L)v_t = \sum_{k=0}^{+\infty} D_k v_{t-k}$ where $\sum_{k=0}^{+\infty} \|D_k\| < +\infty$ and v_t is a strong white noise such that: $\exists M / \forall (n,i,t) E v_{it}^4 \leq M$*

Note that (v_t) and $D_0(L)$ are not nested matrices: when n increases because a new observation is added to X_t, a new observation is also added to ξ_t but the innovation process and the filter $D_0(L)$ entirely change.

A convenient way to parameterize the dynamics is to further assume that the common factors following a VAR process so that the following assumption is added to (A3) (see Forni et al., 2005a, for a discussion):

(A3') *The factors admit a VAR representation:* $A_0^*(L)F_t = u_t$ *where* $A_0^*(z) \neq 0$ *for* $|z| \leq 1$ *and* $A_0^*(0) = I_r$.

For any n, we denote by $\bar{\psi}_0 = \frac{1}{n}\sum_{j=1}^n \mathrm{E}\xi_{it}^2$, and in the whole paper, $A_0^*(L)$, Ψ_0, $D_0(L)$, $\bar{\psi}_0$ denote the true values of the parameters.

Given the size of the cross-section n, the model is identified provided that the number of common factors (r) is small with respect to the size of the cross-section (n), and the idiosyncratic component is orthogonal at all leads and lags, i.e. $D_0(L)$ is a diagonal matrix (exact factor model). This version of the model was proposed by Engle and Watson, 1981 and they estimated it by Maximum Likelihood [1]. In what follows, we will not impose such restriction and work under the assumption of some form of weak correlation among idiosyncratic components (approximate factor model) as in the n large, new generation factor literature. There are different ways to impose identifying assumptions that restrict the cross-correlation of the idiosyncratic elements and preserve the commonality of the common component as n increases. We will assume that the Chamberlain and Rothschild (1983)'s conditions are satisfied and we will extend some of these conditions in order to fit the dynamic case. More precisely, denoting by $\lambda_{min}(A)$ and $\lambda_{max}(A)$ the smallest and the greatest eigenvalues of a matrix A, and by $\|A\| = (\lambda_{max}(A'A))^{1/2}$, we make the following assumptions.

We suppose that the common component is pervasive, in the following sense:

(CR1) $\liminf_{n\to\infty} \frac{1}{n}\lambda_{min}(\Lambda_0^{*\prime}\Lambda_0^*) > 0$

We also suppose, as in Forni et al. (2004), that all the eigenvalues of $\Lambda_0^{*\prime}\Lambda_0^*$ diverge at the same rate, which is equivalent to the following further assumption:

(CR2) $\limsup_{n\to\infty} \frac{1}{n}\lambda_{max}(\Lambda_0^{*\prime}\Lambda_0^*)$ *is finite*

We suppose that the cross-sectional time autocorrelation of the idiosyncratic component can only have a limited amount:

(CR3) $\limsup_{n\to\infty} \sum_{h\in\mathbb{Z}} \|\Psi_0(h)\|$ *is finite*

We also make the two following technical assumptions:

(CR4) $\inf_n \lambda_{min}(\Psi_0) = \underline{\lambda} > 0$

(A4) $\Lambda_0^{*\prime}\Lambda_0^*$ has distinct eigenvalues [2].

It must be emphasized that:

[1] Identification conditions for the model for a fixed cross-sectional dimensions (n) are studied in Geweke and Singleton (1980).

[2] This assumption is usual in this framework, and is made to avoid useless mathematical complications. However, in case of multiple eigenvalues, the results would remained unchanged.

- assumption (CR3) extends the Chamberlain and Rothschild (1983)'s following condition: $\exists \bar{\lambda} / \ \sup_n \|\Psi_0\| < \bar{\lambda}$ and is achieved as soon as the two following assumptions are made: $\exists M / \forall n \ \|\mathrm{E}[v_t v_t']\| \leq M$ and $\sum_{k=0}^{+\infty} \|D_k\| \leq M$

- assumption (CR4) was made by Chamberlain and Rothschild (1983): it ensures that the idiosyncratic component does not tend to a degenerate random variable when n goes to infinity.

Remark 2: These assumptions are slightly different than those introduced by Stock and Watson (2002a) and Bai and Ng (2002) but have a similar role. They have been generalized for the dynamic case by Forni et al. (2000) and Forni and Lippi (2001)

As we said before, the common factors, and the factor loadings, are identified up to a normalization. In order to give a precise statement of the consistency results in our framework, we will use here a particular normalization. Let us define:

- D_0 as the diagonal matrix whose diagonal entries are the eigenvalues of $\Lambda_0^{*\prime}\Lambda_0^*$ in decreasing order,
- Q_0 as the matrix of a set of unitary eigenvectors associated with D_0,
- $\Lambda_0 = \Lambda_0^* Q_0$, so that $\Lambda_0'\Lambda_0 = D_0$ and $\Lambda_0\Lambda_0' = \Lambda_0^*\Lambda_0^{*\prime}$,
- $P_0 = \Lambda_0 D_0^{-1/2}$ so that $P_0' P_0 = I_r$,
- $G_t = Q_0' F_t$.

With these new notations, the model can also be written as:

$$X_t = \Lambda_0 G_t + \xi_t \tag{2.1}$$

We then have : $\mathrm{E}[G_t G_t'] = I_r$, and $\mathrm{E}[G_t G_{t-h}'] = \Phi_0(h) = Q_0' \Phi_0^*(h) Q_0$ for any h . It then follows that:

$$\Sigma_0 = \Lambda_0^*\Lambda_0^{*\prime} + \Psi_0 = \Lambda_0\Lambda_0' + \Psi_0$$

and that, for any h: $\Sigma_0(h) = \Lambda_0^*\Phi_0^*(h)\Lambda_0^{*\prime} + \Psi_0(h) = \Lambda_0\Phi_0(h)\Lambda_0' + \Psi_0(h)$.

Note that, in the initial representation of the model, the matrices Λ_0^* are supposed to be nested (when an observation is added to X_t, a line is added to the matrix Λ_0^*), and that it is not the case for the Λ_0 matrices. However, as Q_0 is a rotation matrix, G_t and F_t have the same range, likewise Λ_0 and Λ_0^* have the same range[3]. In addition, assumptions (A1) to (A4) and (CR1) to (CR4) are satisfied if we replace Λ_0^* with Λ_0, and F_t with G_t. If also assumption (A3') holds then G_t also has a VAR representation.

[3] It is worth noticing that Q_0 is uniquely defined up to a sign change of its columns and that G_t is uniquely defined up to a sign change of its components (this will be used below). Indeed, as $\Lambda_0^{*\prime}\Lambda_0^*$ is supposed to have distinct eigenvalues, Q_0 is uniquely defined up to a sign change of its columns. Then, if Δ is a diagonal matrix whose diagonal terms are ± 1, and if Q_0 is replaced by $Q_0\Delta$, Λ_0 is replaced by $\Lambda_0\Delta$ and G_t is replaced by ΔG_t.

Indeed, as $Q_0 G_t = F_t$, we have: $A_0^*(L)G_t = u_t$, and $Q_0' A_0^*(L)G_t = Q_0' u_t$. We then can write:

$$A_0(L)G_t = w_t,$$

with $A_0(L) = Q_0' A_0^*(L)Q_0, w_t = Q_0' u_t, A_0(z) \neq 0$ for $|z| \leq 1,$ and $A_0(0) = I_r$.

Throughout the paper, we concentrate on consistent estimation of G_t rather than F_t, which means that we make explicit which rotation of the factors we are estimating.

3 Approximating projections and population results

The true model underlying the data can be defined as $\Omega = \{\Lambda^*, A^*(L), D(L)\}$ or equivalently as $\Omega = \{\Lambda, A(L), D(L)\}$. If this true model were known, the best approximation of G_t as a linear function of the observables $X_1, ..., X_T$ would be:

$$G_{t|T} = \mathrm{Proj}_\Omega[G_t|X_s, s \leq T]$$

If the model is Gaussian, i.e. if u_t and v_t are normally distributed, then

$$\mathrm{Proj}_\Omega[G_t|X_s, s \leq T] = \mathrm{E}_\Omega[G_t|X_s, s \leq T]$$

Moreover, if the projection is taken under the true parameter values, $\Omega_0 = \{\Lambda_0, A_0(L), D_0(L)\}$, then we have optimality in mean square sense.

In what follows, we propose to compute other projections of G_t, which are associated to models which are misspecified as well, but which are likely to be closer to the real model underlying the data. We show that, although not optimal, these projections also give consistent approximations of G_t, under our set of assumptions.

The simplest projection is obtained under the triple $\Omega_0^{R1} = \left\{\Lambda_0, I_r, \sqrt{\bar{\psi}_0 I_n}\right\}$, that is under an approximating model according to which the common factors are white noise with covariance I_r and the idiosyncratic components are cross-sectionally independent homoscedastic white noises with variance $\bar{\psi}_0$. We have:

$$\mathrm{Proj}_{\Omega_0^{R1}}[G_t|X_s, s \leq T] = \mathrm{E}_{\Omega_0^{R1}}[G_t X_t'] \left[\mathrm{E}_{\Omega_0^{R1}}[X_t X_t']\right]^{-1} X_t = \Lambda_0' \left(\Lambda_0 \Lambda_0' + \bar{\psi}_0 I_n\right)^{-1} X_t.$$

Simple calculations show that, when Ψ_{0R} is an invertible matrix of order n:

$$\left(\Lambda_0 \Lambda_0' + \Psi_{0R}\right)^{-1} = \Psi_{0R}^{-1} - \Psi_{0R}^{-1}\Lambda_0 \left(\Lambda_0' \Psi_{0R}^{-1} \Lambda_0 + I_r\right)^{-1} \Lambda_0' \Psi_{0R}^{-1}.$$

Applying this formula with $\Psi_{0R} = \bar{\psi}_0 I_n$, the previous expression can then be written as:

$$\mathrm{Proj}_{\Omega_0^{R1}}[G_t|X_s, s \leq T] = \left(\Lambda_0' \bar{\psi}_0^{-1} \Lambda_0 + I_r\right)^{-1} \Lambda_0' \bar{\psi}_0^{-1} X_t = \left(\Lambda_0' \Lambda_0 + \bar{\psi}_0 I_r\right)^{-1} \Lambda_0' X_t$$

which is, by assumption (CR1), asymptotically equivalent to the OLS regression of X_t on the factor loadings Λ_0.

It is clear that, under conditions (CR1) and (CR3), such simple OLS regression provides a consistent estimate of the unobserved common factors as the cross-section becomes large[4]. In particular,

$$\text{Proj}_{\Omega_0^{R1}}[G_t|X_s, s \leq T] \xrightarrow{m.s.} G_t \text{ as } n \to \infty$$

Indeed, given the factor model representation, and the definition of Λ_0, we have:

$$\left(\Lambda_0'\Lambda_0 + \bar{\psi}_0 I_r\right)^{-1} \Lambda_0' X_t = \left(\Lambda_0'\Lambda_0 + \bar{\psi}_0 I_r\right)^{-1} \Lambda_0'\Lambda_0 G_t + \left(\Lambda_0'\Lambda_0 + \bar{\psi}_0 I_r\right)^{-1} \Lambda_0'\xi_t$$

Under (CR1), the first term converges to the unobserved common factors G_t, since $\left(\Lambda_0'\Lambda_0 + \bar{\psi}_0 I_r\right)^{-1} \Lambda_0'\Lambda_0 \to I_r$, as $n \to \infty$. The last term converges to zero in mean square since, by assumptions $(CR1)$ to $(CR3)$:

$$\begin{aligned}
&\mathrm{E}_{\Omega_0}\left[\left(\Lambda_0'\Lambda_0 + \bar{\psi}_0 I_r\right)^{-1} \Lambda_0'\xi_t\xi_t'\Lambda_0 \left(\Lambda_0'\Lambda_0 + \bar{\psi}_0 I_r\right)^{-1}\right] \\
&\qquad \leq \lambda_{max}(\Psi_0)\left(\Lambda_0'\Lambda_0 + \bar{\psi}_0 I_r\right)^{-1} \Lambda_0'\Lambda_0 \left(\Lambda_0'\Lambda_0 + \bar{\psi}_0 I_r\right)^{-1} \to 0 \text{ as } n \to \infty
\end{aligned}$$

If we denote $G_{t/T,R1} = \text{Proj}_{\Omega_0^{R1}}[G_t|X_s, s \leq T]$, we then have:

$$G_{t/T,R1} - G_t = O_P\left(\frac{1}{\sqrt{n}}\right) \quad \text{as } n \to \infty$$

This simple estimator is the most efficient one if the true model is $\Omega_0 = \Omega_0^{R1}$: this is the model which is implicitly assumed in the Probabilistic Principal Components framework, *i.e.* a static model with i.i.d. idiosyncratic terms. However, if there are dynamics in the common factors ($A_0(L) \neq I_r$) and if the idiosyncratic components have dynamics or are not spherical ($D_0(L) \neq \sqrt{\bar{\psi}_0 I_n}$), this approach still gives a consistent estimate of the unobserved common factors, as $n \to \infty$.

If the size of the idiosyncratic component is not the same across series, another estimator can be obtained by exploiting such heterogeneity and giving less weight to series with larger idiosyncratic component. Denoting $\Psi_{0d} = \text{diag}(\psi_{0,11}, ..., \psi_{0,nn})$, this can be done by running the projection under the triple

$$\Omega_0^{R2} = \left\{\Lambda_0, I_r, \Psi_{0d}^{1/2}\right\}$$

Using the same formula as we used in the previous case, with $\Psi_{0R} = \Psi_{0d}$ instead of $\Psi_0 = \bar{\psi}_0 I_n$, the following estimated factors are:

$$\text{Proj}_{\Omega_0^{R2}}[G_t|X_s, s \leq T] = \Lambda_0'\left(\Lambda_0\Lambda_0' + \Psi_{0d}\right)^{-1} X_t = \left(\Lambda_0'\Psi_{0d}^{-1}\Lambda_0 + I_r\right)^{-1} \Lambda_0'\Psi_{0d}^{-1} X_t$$

[4]Notice that here the term consistency could be misleading since we are supposing that the parameters of the model are known. We will consider the case of joint estimation of parameters and factors in the next section.

This estimator is used in the traditional (exact) Factor Analysis framework for static data, where it is assumed that Ω_0^{R2} is the true model underlying the data. It is obtained as the previous one, up to the fact that X_t and, of course, Λ_0 have been weighted, with weight given by $\sqrt{\psi_{0,11}}, ..., \sqrt{\psi_{0,nn}}$. If the true model is Ω_0^{R2}, this estimator will be more efficient than the previous one, for a given n. On the other hand, if Ω_0^{R2} is not the true model, it is straightforward to obtain the same consistency result as in the previous case, under assumptions (CR1), to (CR4). If $G_{t/t,R2} := \text{Proj}_{\Omega_0^{R2}}[G_t|X_s, s \leq T]$, then:

$$G_{t/T,R2} \xrightarrow{m.s.} G_t \text{ and } G_{t/T,R2} - G_t = O_P\left(\frac{1}{\sqrt{n}}\right) \text{ as } n \to \infty$$

Notice that in traditional factor models, where n is considered fixed, the factors are indeterminate and can only be approximated with an approximation error that depends inversely on the signal to noise variance ratio. The n large analysis shows that under suitable conditions, the approximation error goes to zero for n large.

For a given n, further efficiency improvements could be obtained by non diagonal weighting scheme, i.e. by running the projection under the triple $\left\{\Lambda_0, I_r, \Psi_0^{1/2}\right\}$. This might be empirically relevant since, although limited asymptotically by assumption (CR3), the idiosyncratic cross-sectional correlation may affect results in finite sample. We will not consider such projections since non diagonal weighting schemes raise identifiability problems in finite samples, and would practically require the estimation of too many parameters. Indeed, there is no satisfactory way to fully parameterize parsimoniously the DGP of the idiosyncratic component since in most applications the cross-sectional items have no natural order.

On the other hand, the estimators considered above do not take into consideration the dynamics of the factors and the idiosyncratic component. For this reason the factors are extracted by projecting only on contemporaneous observations. Since the model can be written in a state space form, we propose to compute projections under more general dynamic structures using Kalman smoothing techniques.

Two particular cases in which the Kalman smoother can be used to exploit the dynamics of the common factors are:

$$\begin{aligned}
\Omega_0^{R3} &= \left\{\Lambda_0, A_0(L), \sqrt{\bar{\psi}_0} I_n\right\} \\
\Omega_0^{R4} &= \left\{\Lambda_0, A_0(L), \Psi_{0d}^{1/2}\right\}
\end{aligned}$$

In both cases, the state-space form of the model under assumption (A3') is:

$$
\begin{aligned}
X_t &= \begin{pmatrix} \Lambda_0 & 0 & \ldots & 0 \end{pmatrix} \begin{pmatrix} G_t \\ G_{t-1} \\ \vdots \\ G_{t-p+1} \end{pmatrix} + \xi_t \\
\begin{pmatrix} G_t \\ G_{t-1} \\ \vdots \\ G_{t-p+1} \end{pmatrix} &= \begin{pmatrix} A_{01} & A_{02} & \ldots & A_{0p} \\ I_r & 0 & \ldots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \ldots & I_r \end{pmatrix} \begin{pmatrix} G_{t-1} \\ G_{t-2} \\ \vdots \\ G_{t-p} \end{pmatrix} + \begin{pmatrix} I_r \\ 0 \\ \vdots \\ 0 \end{pmatrix} w_t
\end{aligned}
$$

In the measurement equation, the covariance matrix of ξ_t is supposed to be equal to $\sqrt{\bar{\psi}_0} I_n$ in Ω_0^{R3} framework, whereas it is supposed to be equal to $\Psi_{0d}^{1/2}$ in Ω_0^{R4} framework.

Under such parameterizations, the computational complexity of the Kalman smoothing techniques depends mainly on the dimension of the transition equation which, under the parameterizations above, is independent of n and depends only on the number of the common factors.

In both frameworks, the Kalman smoother computes $G_{t/T,R} := \mathrm{Proj}_{\Omega_0^R}[G_t | X_s, s \leq T]$, with $R = R3$ or R4. We want to show that this gives a consistent estimate of G_t even if Ω_0^R is misspecified, due to the fact that the true matrix Ψ_0 is a non-diagonal matrix and the idiosyncratic components are autocorrelated. In both cases, for given values of the parameters, the smoother is computed iteratively, for each value of t. However, in order to prove our consistency result, we will not use these recursive formulas but directly use the general form of $G_{t/T,R}$.

In order to do this, we introduce the following notations:

- $\mathbf{X}_T = (X_1', \ldots X_T')'$, $\mathbf{G}_T = (G_1', \ldots G_T')'$, $\mathbf{Z}_T = (\xi_1', \ldots \xi_T')'$,
- E denotes the expectation of a random variable, under the true model Ω_0,
- $E_{\Omega_0^R}$ denotes the expectation of a random variable, when Ω_0^R is the model which is considered,
- When (Y_t) is a stationary process: $\Gamma_Y(h) = E(Y_t Y_{t-h}')$ and $\Gamma_{Y,R}(h) = E_{\Omega_0^R}(Y_t Y_{t-h}')$,
- When (Y_t) is a stationary process and $\mathbf{Y}_T = (Y_1', \ldots Y_T')'$, we denote:

$$\Sigma_Y = E(\mathbf{Y}_T \mathbf{Y}_T') \text{ and } \Sigma_{Y,R} = E_{\Omega_0^R}(\mathbf{Y}_T \mathbf{Y}_T')$$

- $\mathbf{U}_t'$ is the $(r \times rT)$ matrix defined by: $\mathbf{U}_t' = (0, \ldots I_r, 0 \ldots 0)$

With these notations:

$$\mathbf{X}_T = (I_T \otimes \Lambda_0)\mathbf{G}_T + \mathbf{Z}_T,$$

$$\text{and} : G_{t/T,R} = E_{\Omega_0^R}(G_t \mathbf{X}_T')(E_{\Omega_0^R}(\mathbf{X}_T \mathbf{X}_T'))^{-1}\mathbf{X}_T = E_{\Omega_0^R}(G_t \mathbf{X}_T')\Sigma_{X,R}^{-1}\mathbf{X}_T.$$

Notice that, when $R = R3$ or R4 the DGP of (G_t) is supposed to be correctly specified, so that $\Sigma_G = \Sigma_{G,R}$. On the contrary, $\Sigma_{Z,R}$, is not equal to Σ_Z, and we have:

$$\Sigma_{Z,R} = I_T \otimes \Psi_{0,R}$$

$$\text{with } \Psi_{0,R3} = \bar{\psi}_0 I_n \text{ and } \Psi_{0,R4} = \Psi_{0d} = \text{diag}(\psi_{0,11}, \ldots, \psi_{0,nn})$$

Our consistency result is based on the following lemma:

Lemma 1 Under assumptions (A1) to (A4), (A3'), (CR1), to (CR4) the following properties hold for $R = R3,$ or $R4$:

i) $G_{t/T,R} = \mathbf{U}_t' \Sigma_{G,R}(I_T \otimes \Lambda_0')\Sigma_{X,R}^{-1}\mathbf{X}_T$

ii) $\Sigma_{G,R} = \Sigma_G$, $\|\Sigma_G\| = O(1)$ and $\|\Sigma_G^{-1}\| = O(1)$

iii) $\|\Sigma_Z\| = O(1)$, $\Sigma_{Z,R} = O(1)$ and $\|\Sigma_{Z,R}^{-1}\| = O(1)$

Proof: see appendix A1.

It is worth noticing that the last result of this lemma comes from assumption (CR3), which states the limitation of the cross-sectional autocorrelation. This assumption is crucial to obtain the following consistency result:

Proposition 1 Under assumptions (A1) to (A4), (A3'), and (CR1) to (CR4), if $G_{t/T,R} = \text{Proj}_{\Omega_0^R}[G_t | X_s, s \leq T]$ with $R = R3,$ or $R4$, then:

$$G_{t/T,R} \xrightarrow{m.s.} G_t \text{ and } G_{t/T,R} - G_t = O_P\left(\frac{1}{\sqrt{n}}\right) \quad \text{as } n \to \infty$$

Proof: see appendix A1.

In summary, the factors can be consistently estimated,as n become larger, by simple static projection of the observable on the factor loadings. However, it is also possible to exploit the cross-sectional heteroscedasticity of the idiosyncratic components through weighted regressions (parametrization Ω_0^{R2}), and the dynamics of the factors, through the Kalman smoother (parametrizations Ω_0^{R3} and Ω_0^{R4}). This property may be particularly useful in the case of unbalanced panel of data.

Individual idiosyncratic dynamics could also be taken into account when performing the projections. This would require to specify an autoregressive model for the idiosyncratic components or a reparameterization of the model as in Quah and Sargent (1992), to capture idiosyncratic dynamics by including lagged observable variable.

4 A two-step estimation procedure

The discussion in the previous section assumed that the parameters were known and focused on the extraction of the factors. In this section we propose a two-step procedure in order to estimate the factors when the parameters of the model are unknown. In the first step, preliminary estimators of the factors, and estimators of the parameters of the model, are computed from a Principal Component Analysis (PCA). In the second step, we take the heteroscedasticity of the idiosyncratic components and/or the dynamics of the common factors into account, along the same lines as what we did in the previous section. The true values of the parameters are now replaced by their PCA estimates, and the dynamics of the factors are estimated from the associated preliminary estimates of the factors.

As we said in the previous section, the estimation of the full model is not feasible since it is not possible to fully parameterize parsimoniously the DGP of the idiosyncratic component. However, we have seen that, if the factor loadings were known, the factors could be consistently estimated by a Kalman smoother, even if the projections were not computed under the correct specification. We show below that robustness with respect to misspecification still holds if the parameters are estimated by PCA.

More precisely, our procedure can be defined as follows. For each of the approximating model $\Omega^{Ri}, i = 1$ to 4, that we have defined in the previous section, we replace the true parameters by estimated values in the following way. In the four cases, the factor loadings matrix Λ_0 is replaced by the matrix $\hat{\Lambda}$ obtained by PCA, and the variance-covariance matrix which is specified for the idiosyncratic component is also directly obtained from the PCA estimates. For cases $\Omega^{Ri}, i = 3$ and 4, where the factors are supposed to follow a VAR model, we use the preliminary estimates $\hat{G}_t$ of the factors obtained by PCA, and estimate the VAR coefficients by OLS regression of $\hat{G}_t$ on its own past, following Forni et al. (2005). We thus define, for each of the approximating model $\Omega^{Ri}, i = 1$ to 4, an associated setup, which we denote $\hat{\Omega}^{Ri}, i = 1$ to 4. In each case, we compute a new estimation of the factor G_t, which we denote $\hat{G}_{t/T,Ri}$ and which is equal to $\widehat{\text{Proj}}_{\hat{\Omega}^{Ri}}[G_t|X_s, s \leq T]$.

In order to prove that $\hat{G}_{t/T,Ri}$ is a consistent estimator of G_t, we proceed in three steps. In the first step, we show that under our set of assumptions, principal components give consistent estimators of the span of the common factors, and of associated factors loadings, when both the cross-section and the sample size go to infinity. This result has been shown by Forni et al. (2005a). Similar results, under alternative assumptions have been derived Bai (2003), Bai and Ng (2002) and Stock and Watson (2002a). However, we give our own proof of these results in appendix A.2, because we need intermediate results in order to prove the other propositions of this section. The consistency of $\hat{G}_{t/T,R1}$ and $\hat{G}_{t/T,R2}$ directly follow from the consistency of PCA. We then show that the estimates we propose for the dynamics of the factors are also consistent estimates when both the cross-section and the sample size go to infinity. Finally, we derive the consistency of $\hat{G}_{t/T,Ri}, i = 3$ and 4.

Let us then first study PCA estimates and the consistency of $\hat{G}_{t/T,R1}$ and $\hat{G}_{t/T,R2}$. If we denote by $S = \frac{1}{T}\sum_{t=1}^{T} X_t X_t'$ the empirical variance-covariance matrix of the data, by $\hat{d}_j$ the j-th eigenvalue of S, in decreasing order of magnitude[5], by $\hat{p}_j$ the relative unitary eigenvector, and if we denote by $\hat{D}$ the $(r \times r)$ diagonal matrix with diagonal elements $\hat{d}_j$, $j = 1 \dots r$, and $\hat{P} := (\hat{p}_1, \dots, \hat{p}_r)$, the associated PCA estimates are given by:

$$\begin{aligned}\hat{G}_t &= \hat{D}^{-1/2}\hat{P}'X_t \\ \hat{\Lambda} &= \hat{P}\hat{D}^{1/2}\end{aligned}$$

The consistency results are the following:

Proposition 2 If assumptions (CR1) to (CR4), (A1) to (A4) and (A3') hold, then Λ_0 can be defined[6] so that the following properties hold:

i) $\hat{G}_t - G_t = O_P\left(\frac{1}{\sqrt{n}}\right) + O_P\left(\frac{1}{\sqrt{T}}\right)$, as $n, T \to \infty$

ii) For any i, j: $\hat{\lambda}_{ij} - \lambda_{0,ij} = O_P\left(\frac{1}{\sqrt{n}}\right) + O_P\left(\frac{1}{\sqrt{T}}\right)$

iii) If $\hat{\Psi} = S - \hat{\Lambda}\hat{\Lambda}'$ then, for any (i, j): $\hat{\psi}_{ij} - \psi_{0,ij} = O_P\left(\frac{1}{\sqrt{n}}\right) + O_P\left(\frac{1}{\sqrt{T}}\right)$

Proof: see appendix A2.

These consistency results can be interpreted as follows: "the bias arising from this misspecification of the data generating process of the idiosyncratic component and the dynamic properties of the factors is negligible if the cross-sectional dimension is large enough, under the usual set of assumptions".

If we denote:

$$\begin{aligned}\hat{\bar{\psi}} &= \frac{1}{n}\mathrm{tr}\hat{\Psi} = \frac{1}{n}\mathrm{tr}(S - \hat{\Lambda}\hat{\Lambda}') = \frac{1}{n}\left(\mathrm{tr}S - \mathrm{tr}(\hat{\Lambda}'\hat{\Lambda})\right) = \frac{1}{n}\left(\mathrm{tr}S - \mathrm{tr}\hat{D}\right) \\ \hat{\Psi}_d &= \mathrm{diag}\hat{\Psi} = \mathrm{diag}(S - \hat{\Lambda}\hat{\Lambda}') = \mathrm{diag}(\hat{\psi}_{11}, \dots, \hat{\psi}_{nn})\end{aligned}$$

we can define $\hat{\Omega}^{R1} = \left\{\hat{\Lambda}, I_r, \sqrt{\hat{\bar{\psi}}}I_n\right\}$ and $\hat{\Omega}^{R2} = \left\{\hat{\Lambda}, I_r, \hat{\Psi}_d\right\}$.

We then obtain:

$$\begin{aligned}\hat{G}_{t/T,R1} &= \widehat{\mathrm{Proj}}_{\hat{\Omega}^{R1}}[G_t|X_s, s \le T] = \left(\hat{\Lambda}'\hat{\Lambda} + \hat{\bar{\psi}}I_r\right)^{-1}\hat{\Lambda}'X_t = \left(\hat{D} + \hat{\bar{\psi}}I_r\right)^{-1}\hat{D}^{1/2}\hat{P}'X_t \\ \hat{G}_{t/T,R2} &= \widehat{\mathrm{Proj}}_{\hat{\Omega}^{R2}}[G_t|X_s, s \le T] = \left(\hat{\Lambda}'\hat{\Psi}_d^{-1}\hat{\Lambda} + I_r\right)^{-1}\hat{\Lambda}'\hat{\Psi}_d^{-1}X_t\end{aligned}$$

and the consistency of these two approximations of G_t directly follows from the consistency of PCA. We get:

[5] It is always assumed that those eigenvalues are all distinct, in order to avoid useless mathematical complications. Under assumption (A6), this will be asymptotically true, due to the fact that S converges to Σ_0

[6] As Λ_0 is defined up to a sign change of its columns, and G_t is defined up to the sign of its components, the consistency result holds up to a given value of these signs.

Corollary 1 Under the same assumptions as in proposition 2:

i) $\hat{G}_{t/T,R1} - G_t = O_P\left(\frac{1}{\sqrt{n}}\right) + O_P\left(\frac{1}{\sqrt{T}}\right), \quad \text{as } n, T \to \infty$

ii) $\hat{G}_{t/T,R2} - G_t = O_P\left(\frac{1}{\sqrt{n}}\right) + O_P\left(\frac{1}{\sqrt{T}}\right), \quad \text{as } n, T \to \infty$

Proof: see Appendix A2.

Two remarks are in order. First, we see that $\hat{G}_{t/T,R1} = (\hat{D} + \hat{\bar{\psi}} I_r)^{-1}\hat{D}\hat{G}_t$ so that $\hat{G}_{t/T,R1}$ and $\hat{G}_t$ are equal up to a scale coefficient on each component, and are asymptotically equal when n goes to infinity. This can be linked to a well known result since principal components are known to be equal, up to a scale coefficient, to the Maximum Likelihood estimates of the parameters in the Ω^{R1} framework, under a gaussian assumption [7]. Hence, principal components can be seen as an asymptotic equivalent of the Maximum Likelihood estimator in a situation in which the probability model is not correctly specified: the true model satisfies conditions (CR1) to (CR4), is dynamic and approximate, while the approximating model is restricted to be static and the idiosyncratic component to be spherical. This is what White (1982) named as Quasi Maximum Likelihood estimator. This remark opens the way to the study of QML estimators in less restricted frameworks than Ω^{R1}, like for instance Ω^{R4} : such an estimator is studied in (Doz et al., 2005).

Second, $\hat{G}_{t/T,R2} = \left(\hat{\Lambda}'\hat{\Psi}_d^{-1}\hat{\Lambda} + I_r\right)^{-1} \hat{\Lambda}'\hat{\Psi}_d^{-1}X_t$ is asymptotically equivalent to principal components on weighted observations, where the weights are the inverse of the standard deviation of the estimated idiosyncratic components. This estimator has been considered in Forni and Reichlin (2000), Boivin and Ng (2004), Forni, Hallin, Lippi and Reichlin (2005).

Let us now turn to the Ω^{Ri}, $i = 3$ and 4 frameworks, where the dynamics of the factors are taken into account in the second step of the procedure. As suggested by Forni et al. 2005, the VAR coefficients $A_0(L)$ can be estimated by OLS regression of $\hat{G}_t$, on its own past. More precisely, the following OLS regression:

$$\hat{G}_t = \hat{A}_1\hat{G}_{t-1} + \cdots + \hat{A}_p\hat{G}_{t-p} + \hat{w}_t$$

gives consistent estimates of the $A_{0,k}$ matrices.

Proposition 3 Under the same assumptions as in proposition 2, the following properties hold:

i) If $\hat{\Gamma}_{\hat{G}}(h)$ denotes the sample autocovariance of order h of the estimated principal components: $\hat{\Gamma}_{\hat{G}}(h) = \frac{1}{T-h}\sum_{t=h+1}^{T}\hat{G}_t\hat{G}'_{t-h}$, then for any h:

$$\hat{\Gamma}_{\hat{G}}(h) - \Phi_0(h) = O_P\left(\frac{1}{n}\right) + O_P\left(\frac{1}{\sqrt{T}}\right)$$

and the result is uniform in h, $h \leq p$

[7]See e.g. Lawley and Maxwell (1963) for the calculation of the ML estimator in the exact static factor model framework.

ii) For any $s = 0, ..., p$: $\hat{A}_s - A_{0,s} = O_P\left(\frac{1}{n}\right) + O_P\left(\frac{1}{\sqrt{T}}\right)$

Proof: see Appendix A3.

If we denote by $\hat{A}(L)$ the associate estimates of $A_0(L)$ we are then able to define:

$$\begin{aligned} \hat{\Omega}^{R3} &= \left\{\hat{\Lambda}, \hat{A}(L), \sqrt{\hat{\bar{\psi}}} I_n\right\} \\ \hat{\Omega}^{R4} &= \left\{\hat{\Lambda}, \hat{A}(L), (\mathrm{diag}(\hat{\psi}_{11}, ..., \hat{\psi}_{nn}))^{1/2}\right\} \end{aligned}$$

and to compute two new estimates of the factors:

$$\hat{G}_{t/T,Ri} = \widehat{\mathrm{Proj}}_{\hat{\Omega}^{Ri}}[G_t | X_s, s \leq T], i = 3 \text{ and } 4$$

These two estimates are obtained with one run of the Kalman smoother and they take the estimated dynamics of the common factors into account:

- $\hat{G}_{t/T,R3}$ is obtained without reweighting the data: it exploits the common factor dynamics but does not take the non-sphericity of the idiosyncratic component into account.

- $\hat{G}_{t/T,R4}$ exploits the dynamics of the common factors and the non-sphericity of the idiosyncratic component: it has been proposed by Giannone, Reichlin and Small (2005) and applied by Giannone, Reichlin and Sala (2005).

Consistency of these two new estimates of the common factors, follows from the consistency of the associated population estimates (proposition 1), the consistency of PCA (proposition 2), and the consistency of the autoregressive parameters estimates (proposition 3). The proofs are identical in the $\hat{\Omega}^{R3}$ and $\hat{\Omega}^{R4}$ frameworks. We then denote by Ω_0^R the model under consideration, and by $\hat{\Omega}^R$ the associated set of parameters, and we denote $\hat{G}_{t/T,R} = \widehat{\mathrm{Proj}}_{\hat{\Omega}^R}[G_t | X_s, s \leq T]$ the associated estimation of the common factor.

Note that, like in the previous section, our consistency proof will not rely on the Kalman smoother iterative formulas, but on the direct computation of $\hat{G}_{t/T,R}$. As we have seen before that $G_{t/T,R} = \mathbf{U}_t' \Sigma_{G,R}(I_T \otimes \Lambda_0') \Sigma_{X,R}^{-1} \mathbf{X}_T$, we now have:

$$\hat{G}_{t/T,R} = \mathbf{U}_t' \hat{\Sigma}_{G,R}(I_T \otimes \hat{\Lambda}') \hat{\Sigma}_{X,R}^{-1} \mathbf{X}_T$$

where $\hat{\Sigma}_{X,R} = (I_T \otimes \hat{\Lambda}) \hat{\Sigma}_{G,R}(I_T \otimes \hat{\Lambda}) + (I_T \otimes \hat{\Psi}_R)$ and $\hat{\Sigma}_{G,R}$ is obtained from the estimated VAR coefficients. In particular, we ave the following property:

Proposition 4 Under the same assumptions as in proposition 2, the following properties hold:

i) $\|\hat{\Sigma}_{G,R} - \Sigma_{G,R}\| = O_P\left(\frac{1}{n}\right) + O_P\left(\frac{1}{\sqrt{T}}\right)$

ii) $\|\hat{\Sigma}_{G,R}\| = O_P(1)$, $\|\hat{\Sigma}_{G,R}^{-1}\| = O_P(1)$ and $\|\hat{\Sigma}_{G,R}^{-1} - \Sigma_{G,R}^{-1}\| = O_P\left(\frac{1}{n}\right) + O_P\left(\frac{1}{\sqrt{T}}\right)$

Proof: see appendix A3.

We can then obtain our consistency result:

Proposition 5 Denote $\hat{G}_{t/T,R} = \widehat{\text{Proj}}_{\hat{\Omega}^R}[G_t|X_s, s \leq T]$ with $R = R3,$ and $R4$.
If limsup$\frac{T}{n^3} = O(1)$, the following result holds under assumptions (CR1) to (CR4), (A1) to (A4) and (A3'):

$$\hat{G}_{t/T,R} - G_t = O_P\left(\frac{1}{\sqrt{n}}\right) + O_P\left(\frac{1}{\sqrt{T}}\right) \quad \text{as } n, T \to \infty$$

Proof: see appendix A3.

The procedure outlined above can be summarized in the following way: first, we estimate the parameters and the factors through principal components ; second we estimate the dynamics of the factors from these preliminary estimates of the factors ; finally, we reestimate the common factors according to the selected approximating model. What if we iterate such procedure ? From the new estimated factors, we can estimate a new set of parameters which in turn can then be used to reestimate the common factors and so on. If, at each iteration the least squares estimates of the parameters are computed using expected sufficient statistics, then such iterative procedure is nothing that the EM algorithm by Dempster and Rubin (1977) and introduced in small scale dynamic factor models by Engle and Watson (1981). Quah and Sargent (1992) used such algorithm for large cross-sections, but their approach was disregarded in subsequent literature. The algorithm is very powerful since at each step the likelihood increases, and hence, under regularity conditions, it converges to the Maximum Likelihood solution. For details about the estimation with state space models see Engle and Watson (1981) and Quah and Sargent (1992). The algorithm is feasible for large cross-sections for two reasons. First, as stressed above, its complexity is mainly due to the number of factors, which in our framework is independent of the size of the cross-section and typically very small. Second, since the algorithm is initialized with consistent estimates (Principal Component), the number of iterations required for convergence is expected to be limited, in particular when the cross-section is large. The asymptotic properties of quasi maximum likelihood estimates for large cross-section and under an approximate factor structure is developed in Doz et al. (2005).

5 A Monte-Carlo experiment

In this section we run a simulation study to asses the performances of our estimator. The model from which we simulate is standard in the literature. A similar model has been used, for example, in Stock and Watson (2002a).

Let us define it below (in what follows, in order to have simpler notations, we drop the zero subscript for the true value of the parameters which we had previously used to study the consistency of the estimates).

- $x_{it} = \sum_{j=1}^{r} \lambda_{ij}^* f_{jt} + \xi_{it}, i = 1, .., n$, in vector notation $X_t = \Lambda^* F_t + \xi_t$
- λ_{ij}^* i.i.d. $\mathcal{N}(0, 1), i = 1, ..., n; j = 1, .., r$
- $A(L)F_t = u_t,$ with u_t i.i.d. $\mathcal{N}(0, (1 - \rho^2)I_r); i, j = 1, ..., r$

  $$a_{ij}(L) = \begin{cases} 1 - \rho L & \text{if} \quad i = j \\ 0 & \text{if} \quad i \neq j \end{cases}$$

- $D(L)\xi_t = v_t$ with v_t i.i.d. $\mathcal{N}(0, \mathcal{T})$

  $$d_{ij}(L) = \begin{cases} (1 - dL) & \text{if} \quad i = j \\ 0 & \text{if} \quad i \neq j \end{cases} ; i, j = 1, ..., n$$

  $\alpha_i = \frac{\beta_i}{1-\beta_i} \sum_{j=1}^{r} \lambda_{ij}^2$ with β_i i.i.d. $\mathcal{U}([u, 1 - u])$

  $\mathcal{T}_{ij} = \sqrt{\alpha_i \alpha_j}\ \tau^{|i-j|}(1 - d^2), i, j = 1, ..., n$

Notice that we allow for instantaneous cross-correlation between the idiosyncratic elements. Since $\mathcal{T}$ is a Toeplitz matrix, the cross-correlation among idiosyncratic elements is limited and it is easily seen that Assumption A (ii) is satisfied. The coefficient τ controls for the amount of cross-correlation. The exact factor model corresponds to $\tau = 0$.

The coefficient β_i is the ratio between the variance of the idiosyncratic component, ξ_{it}, and the variance of the common component, $\sum_{j=1}^{r} \lambda_{ij}^* f_{jt}$. The is also known as the noise to signal ratio. In our simulation this ratio is uniformly distributed with an average of 50%. If $u = .5$ then the standardized observations have cross-sectionally homoscedastic idiosyncratic components.

Notice that if $\tau = 0, d = 0$, our approximating model is well specified (with the usual notational convention that $0^0 = 1$)and hence the approximating model R4 is well specified. If $\tau = 0,\ d = 0,\ \rho = 0$, we have a static exact factor model with heteroscedastic idiosyncratic component and model R2 is correctly specified while principal components are not the most efficient estimator for finite n. Finally, if $\tau = 0,\ d = 0,\ u = 1/2$, we have a spherical, static factor model on standardized variables, situation in which the approximating model R1 is correctly specified and principal components on standardized variables provide the most efficient, maximum likelihood, estimates.

We generate the model for different sizes of the cross-section, $n = 10, 25, 50, 100$, and for sample size $T = 50,\ 100$. We perform 2500 Monte-Carlo repetitions. We draw 50 times the parameters $\beta_i,\ i = 1, ..., n$, and $\lambda_{ij}^*,\ i = 1, ..., n;\ j = 1, .., r$. Then, for each draw of the parameters, we generate 50 times the shocks u_t and ξ_t.

As stressed in the introduction, an advantage of having a parameterized model is that it is possible to extract the common factors from panel at the end of the sample due to the unsynchronous data releases (see Giannone et al., 2004, 2005, for an application to real time nowcasting and forecasting output and inflation). To study the performance of our models, for each sample size T and cross-sectional dimension n, we generate the data under the following pattern of data availability,

$$x_{it} \text{ available for } t = 1, ..., T - j \text{ if } i \leq (j + 1)\frac{n}{5}$$

that is all the variables are observed for $t = 1, ..., T - 4$, we name this a balanced panel; 80% of the data are available at time $T - 3$; 60% are available at time $T - 2$; 40% are available at time $T - 1$; 20% are available at time T.

At each repetition, the parameters $\hat{\Lambda}$, $\hat{A}(L)$ and $\hat{\psi}_{ii}, i = 1, ..., n$ are estimated on the balanced part of the panel, $x_{it}, i = 1, ..., n, t = 1, ..., T - 4$. Data are standardized so as to have mean zero and variance equal to one. Such standardization is typically applied in empirical analysis since principal components are not scale invariant.

We consider the factor extraction under the approximating models studied in the previous section and summarized below.

$$\hat{\Omega}^{R1} = \left\{\hat{\Lambda}, I_r, \sqrt{\hat{\bar{\psi}}} I_n\right\}$$

$$\hat{\Omega}^{R2} = \left\{\hat{\Lambda}, I_r, (\text{diag}(\hat{\psi}_{11}, ..., \hat{\psi}_{nn}))^{1/2}\right\}$$

$$\hat{\Omega}^{R3} = \left\{\hat{\Lambda}, \hat{A}(L), \sqrt{\hat{\bar{\psi}}} I_n\right\}$$

$$\hat{\Omega}^{R4} = \left\{\hat{\Lambda}, \hat{A}(L), (\text{diag}(\hat{\psi}_{11}, ..., \hat{\psi}_{nn}))^{1/2}\right\}.$$

We compute the estimates by applying the Kalman smoother using the estimated parameters: $\hat{G}_{t/T,R} = \widehat{\text{Proj}}_{\hat{\Omega}^R}[G_t|X_s, s \leq T]$, for $R = R1$ to $R4$. The pattern of data availability can be taken into account when estimating the common factors, by modifying the idiosyncratic variance when performing the projections:

- if x_{it} is available, then $\text{E}\xi_{it}^2$ is set equal to $\hat{\bar{\psi}}$ for the projections $R1, R3$ and to $\hat{\psi}_{ii}$ for the projections $R2$, and $R4$
- if x_{it} is not available, then $\text{E}\xi_{it}^2$ is set equal to $+\infty$

The estimates of the common factor can hence be computed running the Kalman smoother with time varying parameters (see Giannone et al., 2004, 2005).

We measure the performance of the different estimators as:

$$\Delta_{t,R} = \text{Trace}\left(F_t - \hat{Q}'_R \hat{G}_{t/T,R}\right)\left(F_t - \hat{Q}'_R \hat{G}_{t/T,R}\right)'$$

where $\hat{Q}_R$ is the OLS coefficient from the regression of F_t on $\hat{G}_{t/T,R}$ estimated using observations up to time $T - 4$, that is: $\hat{Q}_R = \sum_{t=1}^{T-4} F_t \hat{G}'_{t/T,R} \left(\sum_{t=1}^{T-4} \hat{G}_{t/T,R} \hat{G}'_{t/T,R}\right)^{-1}$. This OLS regression is performed since the common factors are identified only up to a rotation. Indeed, we know from the previous sections that $\hat{G}_{t/T,R}$ is a consistent estimator of $G_t = Q'F_t$, where Q is a rotation matrix such that $Q'\Lambda^{*\prime}\Lambda Q$ is diagonal, with diagonal terms in decreasing order. Thus, it can be easily checked that, as $E(F_tF_t') = I_r$, $\hat{Q}_R$ is a consistent estimator of:

$$\begin{aligned}
&\text{plim}\left(\frac{1}{T}\sum_{t=1}^{T-4} F_t\hat{G}'_{t/T,R}\right)\left(\frac{1}{T}\sum_{t=1}^{T-4}\hat{G}_{t/T,R}\hat{G}'_{t/T,R}\right)^{-1} \\
&= \text{plim}\left(\frac{1}{T}\sum_{t=1}^{T-4} F_tG'_t\right)\left(\frac{1}{T}\sum_{t=1}^{T-4} G_tG'_t\right)^{-1} \\
&= \text{plim}\left(\frac{1}{T}\sum_{t=1}^{T-4} F_tF'_t\right)QQ'\left(\frac{1}{T}\sum_{t=1}^{T-4} G_tG'_t\right)^{-1} Q \\
&= Q
\end{aligned}$$

so that $\hat{Q}'_R\hat{G}_{t/T,R}$ is a consistent estimator of F_t.

We compute the distance for each repetition and then compute the averages ($\bar{\Delta}_{t,R}$).

Table 1 summarizes the results of the Montecarlo experiment for one common factors $r = 1$ and the following specification: $\rho = .9$, $d = .5$, $\tau = .5$, $u = .1$.

Table 1:

	T=50						T=100				
j	$n=5$	$n=10$	$n=25$	$n=50$	$n=100$		$n=5$	$n=10$	$n=25$	$n=50$	$n=100$
$\Delta_{j,R4}$: evaluation of the Kalman filter with cross-sectional heteroscedasticity											
-4	0.45	0.35	0.30	0.29	0.28		0.34	0.23	0.19	0.18	0.17
-3	0.45	0.35	0.30	0.28	0.28		0.36	0.24	0.19	0.18	0.17
-2	0.47	0.36	0.30	0.28	0.27		0.37	0.26	0.20	0.18	0.17
-1	0.50	0.39	0.31	0.29	0.27		0.40	0.29	0.21	0.18	0.17
0	0.57	0.44	0.34	0.30	0.28		0.48	0.35	0.25	0.21	0.19
$\Delta_{j,R4}/\Delta_{j,R1}$: relative performances of simple Principal components											
-4	0.97	0.97	0.98	0.99	0.99		0.95	0.94	0.96	0.98	0.99
-3	0.95	0.95	0.97	0.98	0.99		0.93	0.93	0.96	0.98	0.98
-2	0.92	0.93	0.97	0.98	0.99		0.90	0.91	0.95	0.97	0.98
-1	0.88	0.89	0.95	0.97	0.98		0.84	0.85	0.92	0.95	0.97
0	0.80	0.82	0.90	0.95	0.98		0.73	0.75	0.85	0.92	0.96
$\Delta_{j,R4}/\Delta_{j,R2}$: relative performances of Weighted Principal components											
-4	0.98	0.98	0.99	1.00	1.00		0.96	0.98	0.99	1.00	1.00
-3	0.96	0.97	0.99	1.00	1.00		0.95	0.96	0.99	1.00	1.00
-2	0.94	0.96	0.99	1.00	1.00		0.93	0.95	0.99	1.00	1.00
-1	0.90	0.92	0.98	0.99	1.00		0.86	0.89	0.97	0.99	1.00
0	0.81	0.84	0.94	0.98	1.00		0.75	0.78	0.91	0.97	1.00
$\Delta_{j,R4}/\Delta_{j,R3}$: relative performances of the Kalman filter with cross-sectional homoscedasticity											
-4	1.00	0.99	0.99	0.99	1.00		1.00	0.97	0.97	0.98	0.99
-3	0.99	0.99	0.98	0.99	0.99		1.00	0.97	0.96	0.98	0.98
-2	0.99	0.98	0.98	0.98	0.99		0.98	0.96	0.96	0.97	0.98
-1	0.98	0.98	0.98	0.98	0.99		0.97	0.96	0.96	0.96	0.97
0	0.97	0.98	0.99	0.97	0.98		0.96	0.96	0.94	0.95	0.96

We report the following measures of performance for the last 5 observations to analyze how data availability affects the estimates. The Kalman filter with cross-sectional heteroscedasticity $R4$ is used as a benchmark and we report $\bar{\Delta}_{T-j,R4}$. The smaller the measure, the more accurate are the estimates of the common factors. In addition, we report $\bar{\Delta}_{T-j,R4}/\bar{\Delta}_{T-j,R1}$, $\bar{\Delta}_{T-j,R4}/\bar{\Delta}_{T-j,R2}$,$\bar{\Delta}_{T-j,R4}/\bar{\Delta}_{T-j,R3}$. A number smaller then 1 indicates that the projection under $R4$ is more accurate.

Results show five main features:

1. For any j fixed, $\bar{\Delta}_{T-j,R4}$ decreases as n and T increase, that is the precision of the estimated common factors increases with the size of the cross-section n and the sample size T.

2. For any combination of n and T, $\bar{\Delta}_{T-j,R4}$ increases as j decreases, reflecting the fact that the more numerous are the available data, the higher the precision of the common factor estimates.

3. $\bar{\Delta}_{T-j,R4} < \bar{\Delta}_{T-j,R3} < \bar{\Delta}_{T-j,R2} < \bar{\Delta}_{T-j,R1}$, for all n, T, j. This result indicates that the less miss-specified is the model used for the projection, the more accurate are the estimated factors. This suggests that taking into account cross-sectional heteroscedasticity and the dynamic of the common factors helps extracting the common factor.

4. For any combination of n and T, $\bar{\Delta}_{T-j,R4}/\bar{\Delta}_{T-j,R}$ (for $R = R1$ to $R3$) decreases as j decreases. That is, the efficiency improvement is more relevant when it is harder to extract the factors (i.e. the less numerous are the available data).

5. As n, T increase $\bar{\Delta}_{T-j,R4}/\bar{\Delta}_{T-j,R}$ tends to one, for all j and for $R = R1$ to $R3$; that is the performance of the different estimators tends to become very similar. This reflects the fact that all the estimates are consistent for large cross-sections.

Summarizing, the two steps estimator of approximate factor models works well in finite sample. Because it models explicitly dynamics and cross-sectional heteroscedasticity, it dominates principal components. It is particularly relevant when the factor extraction is difficult, that is, when the available data are less numerous.

6 Conclusions

We have shown (n, T) consistency and rates of common factors estimated via a two step procedure whereby, in the first step, the parameters of a dynamic approximate factor model are first estimated by a OLS regression of the variables on principal components and, in the second step, given the estimated parameters, the factors are estimated by the Kalman smoother.

This procedure allows to take into account, in the estimation of the factors, both factor dynamics and idiosyncratic heteroscedasticity, features that are likely to be relevant in the panels of data typically used in empirical applications in macroeconomics. We show that it is consistent, even if the models which is used in the Kalman smoother is misspecified. This consistency result is confirmed in a Monte-Carlo exercise, which also shows that our approach improves the estimation of the factors when n is small.

The parametric approach studied in this paper provides the theoretical justification for two applications of factor models in large cross-sections: treatment of unbalanced panels (Giannone, Reichlin, and Sala, 2004; Giannone, Reichlin, and Small, 2005) and estimation of shocks in structural factor models (Giannone, Reichlin, and Sala, 2004). The approach can also be used to evaluate estimation uncertainty around the common factors as in the papers just cited.

References

Jushan Bai. Inferential theory for factor models of large dimensions. *Econometrica*, 71 (1):135–171, 2003.

Jushan Bai and Serena Ng. Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221, 2002.

Jean Boivin and Serena Ng. Are more data always better for factor analysis? NBER Working Paper 9829, 2003. *Journal of Econometrics*, forthcoming.

Jean Boivin and Serena Ng. Understanding and comparing factor-based forecasts. *International Journal of Central Banking*, 3:117–151, 2005.

Gari Chamberlain and Michael Rothschild. Arbitrage, factor structure and mean-variance analysis in large asset markets. *Econometrica*, 51:1305–1324, 1983.

Antonello D'Agostino and Domenico Giannone. Comparing alternative predictors based on large-panel factor models. Unpublished manuscript, 2005.

N. Laird Dempster, A. and D. Rubin. Maximum likelihood estimation from incomplete data. *Journal of the Royal Statistical Society*, 14:1–38, 1977.

Catherine Doz, Domenico Giannone, and Lucrezia Reichlin. A maximum likelihood approach for large approximate dynamic factor models. Unpublished manuscript, 2005.

Robert. F. Engle and Mark Watson. A one-factor multivariate time series model of metropolitan wage rates. *Journal of the American Statistical Association*, 76:774–781, 1981.

Mario Forni, Domenico Giannone, Marco Lippi, and Lucrezia Reichlin. Opening the black box: Structural factor models with large cross-sections. Unpublished manuscript, Université Libre de Bruxelles, 2005a.

Mario Forni, Marc Hallin, Marco Lippi, and Lucrezia Reichlin. The generalized dynamic factor model: identification and estimation. *Review of Economics and Statistics*, 82: 540–554, 2000.

Mario Forni, Marc Hallin, Marco Lippi, and Lucrezia Reichlin. The generalized dynamic factor model: consistency and rates. *Journal of Econometrics*, 119:231–245, 2004.

Mario Forni, Marc Hallin, Marco Lippi, and Lucrezia Reichlin. The generalized dynamic factor model: one-sided estimtion and forecasting. *Journal of the American Statistical Association*, 100:830–840, 2005b.

Mario Forni and Marco Lippi. The generalized dynamic factor model: representation theory. *Econometric Theory*, 17:1113–1141, 2001.

Mario Forni and Lucrezia Reichlin. Federal policies and local economies: Europe and the us. *Journal of Financial Economics*, 45:109–134, 2001.

John F. Geweke and Kenneth J. Singleton. Maximum likelihood “confirmatory” factor analysis of economic time series. *International Economic Review*, 22:37–54, 1980.

Domenico Giannone, Lucrezia Reichlin, and Luca Sala. Monetary policy in real time. In Mark Gertler and Kenneth Rogoff, editors, *NBER Macroeconomics Annual*, pages 161–200. MIT Press, 2004.

Domenico Giannone, Lucrezia Reichlin, and David Small. Nowcasting gdp and inflation: the real-time informational content of macroeconomic data releases. Finance and Economics Discussion Series 2005-42, Board of Governors of the Federal Reserve System (U.S.), 2005.

D. N. Lawley and A. E. Maxwell. *Factor Analysis as a Statistical Method*. Butterworths, 1963.

Danny Quah and Thomas J. Sargent. A dynamic index model for large cross-section. In James Stock and Mark Watson, editors, *Business Cycle*, pages 161–200. Univeristy of Chicago Press, 1992.

James. H. Stock and Mark. W. Watson. Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, 97: 147–162, 2002a.

James. H. Stock and Mark. W. Watson. Macroeconomic forecasting using diffusion indexes. *Journal of Business and Economics Statistics*, 20:147–162, 2002b.

James. H. Stock and Mark. W. Watson. An empirical comparison of methods for forecasting using many predictors. Unpublished manuscript, 2005.

Halbert White. Maximum likelihood estimation of misspecified models. *Econometrica*, 50:1–25, 1982.

A Appendix

A.1 Consistency of Kalman Smoothing: population results

Proof of lemma 1

i) As $X_t = \Lambda_0 G_t + \xi_t$, we get: $\mathbf{X}_T = (I_T \otimes \Lambda_0)\mathbf{G}_T + \mathbf{Z}_T$.

It then immediately follows from assumptions (A3) that:

$$E_{\Omega_0^R}(G_t \mathbf{X}'_T) = E_{\Omega_0^R}(G_t \mathbf{G}'_T)(I_T \otimes \Lambda'_0) = \mathbf{U}'_t \Sigma_{G,R}(I_T \otimes \Lambda'_0)$$

$$\text{and} : G_{t/T,R} = E_{\Omega_0^R}(G_t \mathbf{X}'_T)(E_{\Omega_0^R}(\mathbf{X}_T \mathbf{X}'_T))^{-1}\mathbf{X}_T = \mathbf{U}'_t \Sigma_{G,R}(I_T \otimes \Lambda'_0)\Sigma^{-1}_{X,R}\mathbf{X}_T$$

ii) We have already noticed that, when $R = R3$ or $R = R4$, the model is correctly specified for (G_t), so that $\Sigma_{GR} = \Sigma_G$.

For any $\omega \in [-\pi, +\pi]$, let us now denote by $S_G(\omega)$ the spectral density matrix of (G_t) calculated in ω. In order to show the two announced properties, we first show that if:

$$m = \mathrm{Min}_{\omega \in [-\pi,+\pi]} \lambda_{min}(S_G(\omega)) \quad \text{and} \quad M = \mathrm{Max}_{\omega \in [-\pi,+\pi]} \lambda_{max}(S_G(\omega))$$

then: $2\pi m \leq \lambda_{min}(\Sigma_G)$ and $2\pi M \geq \lambda_{max}(\Sigma_G)$.

In order to show this property, we generalize to the r-dimensionnal process (G_t) the proof which is given by Brockwell and Davis, 1987 (proposition 4.5.3) in the univariate case.

If $\mathbf{x} = (x'_1, ...x'_T)'$ is a non-random vector of $\mathbb{R}^{rT}$ such that: $\|x\|^2 = \sum_{t=1}^T \|x_t\|^2 = 1$, we can write:

$$\mathbf{x}'\Sigma_G \mathbf{x} = \sum_{t=1}^T \sum_{\tau=1}^T x'_t \Gamma_G(t-\tau)x_\tau = \sum_{t=1}^T \sum_{\tau=1}^T x'_t \Phi_0(t-\tau)x_\tau$$

We thus get:

$$\begin{aligned}
\mathbf{x}'\Sigma_G \mathbf{x} &= \sum_{1 \leq t,\tau \leq T} x'_t \left(\int_{-\pi}^{+\pi} S_G(\omega) e^{-i\omega(t-\tau)} d\omega \right) x_\tau \\
&= \int_{-\pi}^{+\pi} \left(\sum_{1 \leq t,\tau \leq T} x'_t S_G(\omega) x_\tau e^{-i\omega(t-\tau)} \right) d\omega \\
&= \int_{-\pi}^{+\pi} \left(\sum_{1 \leq t \leq T} x'_t e^{-i\omega t} \right) S_G(\omega) \left(\sum_{1 \leq \tau \leq T} x_\tau e^{i\omega\tau} \right) d\omega \\
&\in \left[m \int_{-\pi}^{+\pi} \| \sum_{1 \leq t \leq T} x'_t e^{-i\omega t} \|^2 d\omega, M \int_{-\pi}^{+\pi} \| \sum_{1 \leq t \leq T} x'_t e^{-i\omega t} \|^2 d\omega \right]
\end{aligned}$$

Now:

$$\int_{-\pi}^{+\pi} \| \sum_{1\le t\le T} x_t' e^{-i\omega t}\|^2 d\omega = \int_{-\pi}^{+\pi} \left(\sum_{1\le t,\tau\le T} x_t' e^{-i\omega t} x_\tau e^{-i\omega\tau} \right) d\omega$$
$$= \sum_{1\le t,\tau\le T} \int_{-\pi}^{+\pi} x_t' x_\tau e^{-i\omega(t-\tau)} d\omega = 2\pi \sum_{1\le t\le T} x_t' x_t = 2\pi \sum_{1\le t\le T} \|x_t\|^2 = 2\pi$$

We thus obtain that any eigenvalue of Σ_G belongs to $[2\pi m, 2\pi M]$, which gives the announced result.

Let us now show that $m > 0$ and $M < \infty$, which will prove that:

$$\|\Sigma_G\| = \lambda_{max}(\Sigma_G) \le 2\pi M < \infty \text{ and } \|\Sigma_G^{-1}\| = \frac{1}{\lambda_{min}(\Sigma_G)} \le \frac{1}{2\pi m} < \infty.$$

First, it is clear, from assumption (A3), that for any $\omega \in [-\pi, +\pi]$:

$$\|S_G(\omega)\| = \frac{1}{2\pi} \| \sum_{h=-\infty}^{+\infty} \Phi_0(h) e^{i\omega h} \| \le \frac{1}{2\pi} \sum_{h=-\infty}^{+\infty} \|\Phi_0(h)\| < +\infty$$

so that: $M < +\infty$.

Second, we know that: $A_0(L)G_t = w_t$. As (G_t) is a stationary process, if we denote: $W_0 = E[w_t w_t']$ we have, for any $\omega \in [-\pi, +\pi]$:

$$S_G(\omega) = \frac{1}{2\pi} \left(A_0(e^{i\omega})\right)^{-1} W_0 \left(A_0'(e^{-i\omega})\right)^{-1}$$

For any $x \in \mathbb{C}^n$ such that $\|x\|^2 = 1$, we then have:

$$\begin{aligned}
x' S_G(\omega)\bar{x} &= \frac{1}{2\pi} x' \left(A_0(e^{i\omega})\right)^{-1} W_0 \left(A_0'(e^{-i\omega})\right)^{-1} \bar{x} \\
&\ge \frac{1}{2\pi} \lambda_{min}(W_0) \| x' \left(A_0(e^{i\omega})\right)^{-1} \left(A_0'(e^{-i\omega})\right)^{-1} \bar{x} \\
&\ge \frac{1}{2\pi} \lambda_{min}(W_0) \lambda_{min} \left(\left[A_0'(e^{-i\omega}) A_0(e^{i\omega})\right]^{-1} \right) \\
&= \frac{1}{2\pi} \frac{\lambda_{min}(W_0)}{\lambda_{max}\left(A_0'(e^{-i\omega}) A_0(e^{i\omega})\right)} \\
&= \frac{1}{2\pi} \frac{\lambda_{min}(W_0)}{\|A_0(e^{i\omega})\|^2}
\end{aligned}$$

If we denote $\alpha_0 = \mathrm{Max}_{\omega\in[-\pi,+\pi]} \|A_0(e^{i\omega})\|^2$, we know that α_0 is finite and we get:

$$x' S_G(\omega)\bar{x} \ge \frac{1}{2\pi} \frac{\lambda_{min}(W_0)}{\alpha_0}$$

so that

$$\lambda_{min}(S_G(\omega)) \ge \frac{1}{2\pi} \frac{\lambda_{min}(W_0)}{\alpha_0}$$

iii) For any $\omega \in [-\pi, +\pi]$, let us now denote by $S_\xi(\omega)$ the spectral density matrix of (ξ_t) calculated in ω. If $x = (x_1, ...x_n)'$ is a non-random vector of $\mathbb{C}^n$ such that: $\|x\|^2 = x'\bar{x} = 1$, we have:

$$x'S_\xi(\omega)\bar{x} = \frac{1}{2\pi}\sum_{h\in\mathbb{Z}} x'\Gamma_\xi(h)e^{i\omega h}\bar{x} = \frac{1}{2\pi}\sum_{h\in\mathbb{Z}} x'\Psi_0(h)e^{i\omega h}\bar{x}$$

so that:

$$|x'S_\xi(\omega)\bar{x}| \leq \frac{1}{2\pi}\sum_{h\in\mathbb{Z}} |x'\Psi_0(h)\bar{x}| \leq \frac{1}{2\pi}\sum_{h\in\mathbb{Z}} \|\Psi_0(h)\|$$

From assumption (CR3), we can define $\bar{\lambda}$ such that, for any n: $\sum_{h\in\mathbb{Z}} \|\Psi_0(h)\| < \bar{\lambda}$.

We thus have, for any $\omega \in [-\pi, +\pi]$: $\lambda_{max} S_\xi(\omega) \leq \frac{1}{2\pi}\bar{\lambda}$, so that we finally get:

$$\text{Max}_{\omega\in[-\pi,+\pi]}\lambda_{max}(S_\xi(\omega)) \leq \frac{1}{2\pi}\bar{\lambda}$$

Applying the same result as in (ii), we then obtain: $\|\Sigma_Z\| \leq \bar{\lambda}$.

Further: $\|\Sigma_{Z,R}\| = \|I_T \otimes \Psi_{0,R}\| = \|\Psi_{0,R}\|$, so that: $\|\Sigma_{Z,R}\| \leq \bar{\lambda}$.

Finally: $\|\Sigma_{Z,R}^{-1}\| = \|I_T \otimes \Psi_{0,R}^{-1}\| = \|\Psi_{0,R}^{-1}\|$. As $\|\Psi_{0,R}^{-1}\| = \lambda_{max}(\Psi_{0,R}^{-1}) = \frac{1}{\lambda_{min}(\Psi_{0,R})}$, it follows from assumption (CR4) that $\|\Sigma_{Z,R}^{-1}\| = O(1)$.

Proof of Proposition 1

It follows from assumptions (A3) that:

$$\begin{aligned}\Sigma_{X,R} = E_{\Omega_0^R}(\mathbf{X}_T\mathbf{X}_T') &= (I_T \otimes \Lambda_0)E_{\Omega_0^R}(\mathbf{G}_T\mathbf{G}_T')(I_T \otimes \Lambda_0') + E_{\Omega_0^R}(\mathbf{Z}_T\mathbf{Z}_T') \\ &= (I_T \otimes \Lambda_0)\Sigma_{G,R}(I_T \otimes \Lambda_0') + \Sigma_{Z,R}\end{aligned}$$

Further, as (ξ_t) is supposed to be a white noise in both Ω^{R3} and Ω^{R4} specifications, we also have: $\Sigma_{Z,R} = I_T \otimes \Psi_{0R}$

Using the same kind of formula as the formula we have used to calculate Σ_0^{-1}, it can be easily checked that:

$$\Sigma_{X,R}^{-1} = \Sigma_{Z,R}^{-1} - \Sigma_{Z,R}^{-1}(I_T \otimes \Lambda_0)\left(\Sigma_{G,R}^{-1} + (I_T \otimes \Lambda_0')\Sigma_{Z,R}^{-1}(I_T \otimes \Lambda_0)\right)^{-1}(I_T \otimes \Lambda_0')\Sigma_{Z,R}^{-1}$$

Using the fact that $\Sigma_{Z,R}^{-1} = I_T \otimes \Psi_{0,R}^{-1}$, we then get:

$$\begin{aligned}(I_T \otimes \Lambda_0')\Sigma_{X,R}^{-1} &= I_T \otimes \Lambda_0'\Psi_{0,R}^{-1} - I_T \otimes \Lambda_0'\Psi_{0,R}^{-1}\Lambda_0(\Sigma_{G,R}^{-1} + I_T \otimes \Lambda_0'\Psi_{0,R}^{-1}\Lambda_0)^{-1}I_T \otimes \Lambda_0'\Psi_{0,R}^{-1} \\ &= \left(\Sigma_{G,R}^{-1} + I_T \otimes \Lambda_0'\Psi_{0,R}^{-1}\Lambda_0 - I_T \otimes \Lambda_0'\Psi_{0,R}^{-1}\Lambda_0\right)(\Sigma_{G,R}^{-1} + I_T \otimes \Lambda_0'\Psi_{0,R}^{-1}\Lambda_0)^{-1}I_T \otimes \Lambda_0'\Psi_{0,R}^{-1} \\ &= \Sigma_{G,R}^{-1}(\Sigma_{G,R}^{-1} + I_T \otimes \Lambda_0'\Psi_{0,R}^{-1}\Lambda_0)^{-1}I_T \otimes \Lambda_0'\Psi_{0R}^{-1}\end{aligned}$$

Using lemma 1 (i), we thus obtain:

$$G_{t/T,R} = \mathbf{U}_t'(\Sigma_{G,R}^{-1} + I_T \otimes \Lambda_0'\Psi_{0R}^{-1}\Lambda_0)^{-1}(I_T \otimes \Lambda_0'\Psi_{0R}^{-1})\mathbf{X}_T$$

Before proving the proposition, let us first recall a relation, which we use in that proof as well as in others. If A and B are two square invertible matrices, it is possible to write write: $B^{-1} - A^{-1} = B^{-1}(A-B)A^{-1}$, so that the relation:

$$(A+H)^{-1} = A^{-1} - (A+H)^{-1}HA^{-1} \qquad (R)$$

also gives a Taylor expansion of the inversion operator at order zero when H is small with respect to A.

Using relation (R), and denoting $M_0 = \Lambda_0'\Psi_{0R}^{-1}\Lambda_0$, we then get:

$$\begin{aligned}G_{t/T,R} &= \mathbf{U}_t'\left(I_T \otimes M_0^{-1} - (\Sigma_{G,R}^{-1} + I_T \otimes M_0)^{-1}\Sigma_{G,R}^{-1}(I_T \otimes M_0^{-1})\right)(I_T \otimes \Lambda_0'\Psi_{0R}^{-1})\mathbf{X}_T \\ &= \mathbf{U}_t'(I_T \otimes M_0^{-1}\Lambda_0'\Psi_{0R}^{-1})\mathbf{X}_T - \mathbf{U}_t'(\Sigma_{G,R}^{-1} + I_T \otimes M_0)^{-1}\Sigma_{G,R}^{-1}(I_T \otimes M_0^{-1}\Lambda_0'\Psi_{0R}^{-1})\mathbf{X}_T \\ &= M_0^{-1}\Lambda_0'\Psi_{0R}^{-1}X_t - \mathbf{U}_t'(\Sigma_{G,R}^{-1} + I_T \otimes M_0)^{-1}\Sigma_{G,R}^{-1}(I_T \otimes M_0^{-1}\Lambda_0'\Psi_{0R}^{-1})\mathbf{X}_T\end{aligned}$$

Let us denote $G_{t/T,R}^1$ the first term of the previous summation. We can write:

$$\begin{aligned}G_{t/T,R}^1 &= (\Lambda_0'\Psi_{0R}^{-1}\Lambda_0)^{-1}\Lambda_0'\Psi_{0R}^{-1}X_t \\ &= (\Lambda_0'\Psi_{0R}^{-1}\Lambda_0)^{-1}\Lambda_0'\Psi_{0R}^{-1}(\Lambda_0 G_t + \xi_t) \\ &= G_t + (\Lambda_0'\Psi_{0R}^{-1}\Lambda_0)^{-1}\Lambda_0'\Psi_{0R}^{-1}\xi_t\end{aligned}$$

with:

$$\begin{aligned}E[\|(\Lambda_0'\Psi_{0R}^{-1}\Lambda_0)^{-1}\Lambda_0'\Psi_{0R}^{-1}\xi_t\|^2] &= E\left[tr(\Lambda_0'\Psi_{0R}^{-1}\Lambda_0)^{-1}\Lambda_0'\Psi_{0R}^{-1}\xi_t\xi_t'\Psi_{0R}^{-1}\Lambda_0(\Lambda_0'\Psi_{0R}^{-1}\Lambda_0)^{-1}\right] \\ &= tr\left((\Lambda_0'\Psi_{0R}^{-1}\Lambda_0)^{-1}\Lambda_0'\Psi_{0R}^{-1}\Psi_0\Psi_{0R}^{-1}\Lambda_0(\Lambda_0'\Psi_{0R}^{-1}\Lambda_0)^{-1}\right)\end{aligned}$$

As $\Psi_{0R}^{-1/2}\Psi_0\Psi_{0R}^{-1/2} \leq \frac{\lambda_{max}(\Psi_0)}{\lambda_{min}(\Psi_{0R})}I_n$, we get:

$$(\Lambda_0'\Psi_{0R}^{-1}\Lambda_0)^{-1}\Lambda_0'\Psi_{0R}^{-1}\Psi_0\Psi_{0R}^{-1}\Lambda_0(\Lambda_0'\Psi_{0R}^{-1}\Lambda_0)^{-1} \leq \frac{\lambda_{max}(\Psi_0)}{\lambda_{min}(\Psi_{0R})}(\Lambda_0'\Psi_{0R}^{-1}\Lambda_0)^{-1}$$

so that: $E\left[\|(\Lambda_0'\Psi_{0R}^{-1}\Lambda_0)^{-1}\Lambda_0'\Psi_{0R}^{-1}\xi_t\|^2\right] = O_P\left(\frac{1}{n}\right)$ by assumptions (CR1) and (CR2).

We have thus obtained:

$$G_{t/T,R}^1 \xrightarrow{m.s.} G_t \text{ and } G_{t/T,R}^1 = G_t + O_P\left(\frac{1}{\sqrt{n}}\right)$$

Turning to the second term of the summation, it can in turn be decomposed in two parts. Indeed, as $\mathbf{X}_T = (I_T \otimes \Lambda_0)\mathbf{G}_T + \mathbf{Z}_T$, we can write:

$$\mathbf{U}_t'(\Sigma_{G,R}^{-1} + I_T \otimes M_0)^{-1}\Sigma_{G,R}^{-1}(I_T \otimes M_0^{-1}\Lambda_0'\Psi_{0R}^{-1})\mathbf{X}_T = G_{t/T,R}^2 + G_{t/T,R}^3$$

with:

$$
\begin{aligned}
G^2_{t/T,R} &= \mathbf{U}_t'(\Sigma_{G,R}^{-1} + I_T \otimes M_0)^{-1}\Sigma_{G,R}^{-1}(I_T \otimes M_0^{-1}\Lambda_0'\Psi_{0R}^{-1})(I_T \otimes \Lambda_0)\mathbf{G}_T \\
&= \mathbf{U}_t'\left(\Sigma_{G,R}^{-1} + I_T \otimes M_0\right)^{-1}\Sigma_{G,R}^{-1}(I_T \otimes (\Lambda_0'\Psi_{0R}^{-1}\Lambda_0)^{-1}\Lambda_0'\Psi_{0R}^{-1}\Lambda_0)\mathbf{G}_T \\
&= \mathbf{U}_t'\left(\Sigma_{G,R}^{-1} + I_T \otimes M_0\right)^{-1}\Sigma_{G,R}^{-1}\mathbf{G}_T
\end{aligned}
$$

and:

$$
G^3_{t/T,R} = \mathbf{U}_t'(\Sigma_{G,R}^{-1} + I_T \otimes M_0)^{-1}\Sigma_{G,R}^{-1}(I_T \otimes M_0^{-1}\Lambda_0'\Psi_{0R}^{-1})\mathbf{Z}_T
$$

We can write:

$$
\begin{aligned}
&E\left[\|G^2_{t/T,R}\|^2\right] \\
&\quad = tr\left[\mathbf{U}_t'\left(\Sigma_{G,R}^{-1} + I_T \otimes M_0\right)^{-1}\Sigma_{G,R}^{-1}E\left[\mathbf{G}_T\mathbf{G}_T'\right]\Sigma_{G,R}^{-1}\left(\Sigma_{G,R}^{-1} + I_T \otimes M_0\right)^{-1}\mathbf{U}_t\right]
\end{aligned}
$$

As $E\left[\mathbf{G}_T\mathbf{G}_T'\right] = \Sigma_G = \Sigma_{G,R}$, we then get:

$$
E\left[\|G^2_{t/T,R}\|^2\right] = tr\left[\mathbf{U}_t'\left(\Sigma_{G,R}^{-1} + I_T \otimes M_0\right)^{-1}\Sigma_{G,R}^{-1}\left(\Sigma_{G,R}^{-1} + I_T \otimes M_0\right)^{-1}\mathbf{U}_t\right]
$$

As $\Sigma_{G,R}^{-1} \leq \lambda_{max}(\Sigma_{G,R}^{-1})I_{rT}$, with $\lambda_{max}(\Sigma_{G,R}^{-1}) = \|\Sigma_{G,R}^{-1}\|$, we have:

$$
\left(\Sigma_{G,R}^{-1} + I_T \otimes M_0\right)^{-1}\Sigma_{G,R}^{-1}\left(\Sigma_{G,R}^{-1} + I_T \otimes M_0\right)^{-1} \leq \|\Sigma_{G,R}^{-1}\|\left(\Sigma_{G,R}^{-1} + I_T \otimes M_0\right)^{-2}
$$

Now: $\Sigma_{G,R}^{-1} + I_T \otimes M_0 \geq I_T \otimes M_0$ so that: $\left(\Sigma_{G,R}^{-1} + I_T \otimes M_0\right)^{-1} \leq I_T \otimes M_0^{-1}$.

We then get:

$$
E\left[\|G^2_{t/T,R}\|^2\right] \leq \|\Sigma_{G,R}^{-1}\| tr\left[\mathbf{U}_t'(I_T \otimes M_0^{-2})\mathbf{U}_t\right] = \|\Sigma_{G,R}^{-1}\| tr\left[M_0^{-2}\right] = O\left(\frac{1}{n^2}\right)
$$

It then follows from assumptions (CR1) and (CR2) and from lemma 1 (ii) that:

$$
G^2_{t/T,R} \xrightarrow{m.s.} 0 \text{ and } G^2_{t/T,R} = O_P\left(\frac{1}{n}\right)
$$

If we use the same type of properties that we have used for the study of $G^2_{t/T,R}$, we can write:

$$
E\left[\|G^3_{t/T,R}\|^2\right] = tr\left[\begin{array}{c} \mathbf{U}_t'(\Sigma_{G,R}^{-1} + I_T \otimes M_0)^{-1}\Sigma_{G,R}^{-1}(I_T \otimes M_0^{-1}\Lambda_0'\Psi_{0R}^{-1})\Sigma_Z \\ \times(I_T \otimes \Psi_{0R}^{-1}\Lambda_0 M_0^{-1})\Sigma_{G,R}^{-1}(\Sigma_{G,R}^{-1} + I_T \otimes M_0)^{-1}\mathbf{U}_t \end{array}\right]
$$

We thus get:

$$
\begin{aligned}
&E\left[\|G^3_{t/T,R}\|^2\right] \\
&\quad \leq \|\Sigma_{G,R}^{-1}(I_T \otimes M_0^{-1}\Lambda_0'\Psi_{0R}^{-1})\Sigma_Z(I_T \otimes M_0^{-1}\Lambda_0'\Psi_{0R}^{-1})\Sigma_{G,R}^{-1}\| tr\left[\mathbf{U}_t'(\Sigma_{G,R}^{-1} + I_T \otimes M_0)^{-2}\mathbf{U}_t\right] \\
&\quad \leq \|\Sigma_{G,R}^{-1}\|^2\|(I_T \otimes M_0^{-1}\Lambda_0'\Psi_{0R}^{-1})\|^2\|\Sigma_Z\| tr\left[\mathbf{U}_t'(\Sigma_{G,R}^{-1} + I_T \otimes M_0)^{-2}\mathbf{U}_t\right]
\end{aligned}
$$

From lemma 1 (ii) and (iii) we know that $\|\Sigma_{G,R}^{-1}\| = O(1)$ and $\|\Sigma_Z\| = O(1)$.

Further, using assumptions (CR1) and (CR2), we can write, as before:

$$tr\left[\mathbf{U}_t'(\Sigma_{G,R}^{-1} + I_T \otimes M_0)^{-2}\mathbf{U}_t\right] \leq tr\left[\mathbf{U}_t'(I_T \otimes M_0)^{-2}\mathbf{U}_t\right] = tr(M_0^{-2}) = O\left(\frac{1}{n^2}\right)$$

and:

$$\|(I_T \otimes M_0^{-1}\Lambda_0'\Psi_{0R}^{-1})\| = \|M_0^{-1}\Lambda_0'\Psi_{0R}^{-1})\| = O\left(\frac{1}{\sqrt{n}}\right)$$

It then follows that: $E\left[\|G_{t/T,R}^3\|^2\right] = O\left(\frac{1}{n^3}\right)$, so that:

$$G_{t/T,R}^3 \xrightarrow{m.s.} 0 \text{ and } G_{t/T,R}^3 = O_P\left(\frac{1}{n\sqrt{n}}\right)$$

which completes the proof of the proposition.

A.2 Consistency of PCA

Before proving proposition 2, we need to establish some preliminary lemmas.

Lemma 2 Under assumptions (CR1) to (CR3), (A1) to (A5), the following properties hold, as $n, T \to \infty$:

i) $\frac{1}{n}\|S - \Lambda_0\Lambda_0'\| = O\left(\frac{1}{n}\right) + O_P\left(\frac{1}{\sqrt{T}}\right)$

ii) $\frac{1}{n}\|\hat{D} - D_0\| = O\left(\frac{1}{n}\right) + O_P\left(\frac{1}{\sqrt{T}}\right)$

iii) $n\|\hat{D}^{-1} - D_0^{-1}\| = O_P\left(\frac{1}{n}\right) + O_P\left(\frac{1}{\sqrt{T}}\right)$

iv) $D_0\hat{D}^{-1} = I_r + O_p\left(\frac{1}{n}\right) + O_p\left(\frac{1}{\sqrt{T}}\right)$

Proof

i) $\frac{1}{n}\|S - \Lambda_0\Lambda_0'\| \leq \frac{1}{n}\|S - \Sigma_0\| + \frac{1}{n}\|\Sigma_0 - \Lambda_0\Lambda_0'\|$.

As $\Sigma_0 = \Lambda_0^*\Lambda_0^{*\prime} + \Psi_0 = \Lambda_0\Lambda_0' + \Psi_0$, we have by assumption (CR2) :

$$\frac{1}{n}\|\Sigma_0 - \Lambda_0\Lambda_0'\| = \frac{1}{n}\|\Psi_0\| = O\left(\frac{1}{n}\right)$$

We also have:

$$S = \frac{1}{T}\sum_{t=1}^{T} X_t X_t' = \Lambda_0 \frac{1}{T}\sum_{t=1}^{T} G_t G_t' \Lambda_0' + \Lambda_0 \frac{1}{T}\sum_{t=1}^{T} G_t \xi_t' + \frac{1}{T}\sum_{t=1}^{T} \xi_t G_t' \Lambda_0' + \frac{1}{T}\sum_{t=1}^{T} \xi_t \xi_t'$$

so that:

$$\begin{aligned}\tfrac{1}{n}(S - \Sigma_0) = &\tfrac{1}{n}\Lambda_0\left(\tfrac{1}{T}\textstyle\sum_{t=1}^{T} G_t G_t' - I_r\right)\Lambda_0' + \tfrac{1}{n}\left(\Lambda_0 \tfrac{1}{T}\textstyle\sum_{t=1}^{T} G_t\xi_t' + \tfrac{1}{T}\textstyle\sum_{t=1}^{T}\xi_t G_t'\Lambda_0'\right)\\ &+\tfrac{1}{n}\left(\tfrac{1}{T}\textstyle\sum_{t=1}^{T}\xi_t\xi_t' - \Psi_0\right)\end{aligned}$$

Then, using assumptions (A3) and (CR3) and a multivariate extension of the proof given in the univariate case by Brockwell and Davies (1991, pp226-227), it is possible to show that:

$$\mathrm{E}\left(\|\frac{1}{T}\sum_{t=1}^{T} G_t G_t' - I_r\|^2\right) = O\left(\frac{1}{T}\right) \text{ and } \mathrm{E}\left(\|\frac{1}{T}\sum_{t=1}^{T}\xi_t\xi_t' - \Psi_0\|^2\right) = O\left(\frac{n^2}{T}\right)$$

so that:

$$\|\frac{1}{T}\sum_{t=1}^{T} G_t G_t' - I_r\| = O_P\left(\frac{1}{\sqrt{T}}\right) \text{ and } \|\frac{1}{T}\sum_{t=1}^{T}\xi_t\xi_t' - \Psi_0\| = O_P\left(\frac{n}{\sqrt{T}}\right)$$

It also follows from these assumptions that: $\|\frac{1}{T}\sum_{t=1}^{T} G_t \xi_t'\| = O_P\left(\frac{\sqrt{n}}{\sqrt{T}}\right)$. Indeed, we can write:

$$\|\frac{1}{T}\sum_{t=1}^{T} G_t \xi_t'\|^2 = \|\frac{1}{T^2}\sum_{t,s} G_t \xi_t' \xi_s G_s'\| \leq tr\left(\frac{1}{T^2}\sum_{t,s} G_t \xi_t' \xi_s G_s'\right)$$

As (G_t) and (ξ_t) are two independent processes, we have:

$$\begin{aligned}
\mathrm{E}\left[tr\left(\tfrac{1}{T^2}\textstyle\sum_{t,s} G_t\xi_t'\xi_s G_s'\right)\right] &= tr\left(\tfrac{1}{T^2}\textstyle\sum_{t,s}\mathrm{E}(\xi_t'\xi_s)\mathrm{E}(G_t G_s')\right)\\
&= \tfrac{1}{T^2}\textstyle\sum_{t,s} tr\left(\Psi_0(s-t)\right) tr\left(\Phi_0(t-s)\right)\\
&\leq \tfrac{1}{T^2}\textstyle\sum_{t,s} \left|tr\left(\Psi_0(s-t)\right)\right| \left|tr\left(\Phi_0(t-s)\right)\right|\\
&\leq \tfrac{nr}{T^2}\textstyle\sum_{t,s} \|\Psi_0(s-t)\|\|\Phi_0(t-s)\|\\
&= \tfrac{nr}{T}\textstyle\sum_{h=-T+1}^{T-1}(1-\frac{|h|}{T})\|\Psi_0(h)\|\|\Phi_0(-h)\|\\
&\leq \tfrac{nr}{T}\textstyle\sum_{h\in\mathbb{Z}} \|\Psi_0(h)\|\|\Phi_0(-h)\|\\
&\leq \tfrac{nr}{T} Max_{h\in\mathbb{Z}}\|\Psi_0(h)\| \textstyle\sum_{h\in\mathbb{Z}}\|\Phi_0(h)\|
\end{aligned}$$

We thus obtain: $\mathrm{E}\left[\|\frac{1}{T}\sum_{t=1}^{T} G_t\xi_t'\|^2\right] = O_P\left(\frac{n}{T}\right)$ and the result follows.

ii) $\hat{D}$ is the diagonal matrix of the r first eigenvalues of S, in decreasing order. D_0 is a diagonal matrix which is equal to $\Lambda_0'\Lambda_0$. It is then also equal to the diagonal matrix of the r first eigenvalues of $\Lambda_0\Lambda_0'$ in decreasing order.

Further, if we denote by $\lambda_1(A) \geq \lambda_2(A) \geq \cdots \geq \lambda_n(A)$ the ordered eigenvalues of a symmetric matrix A, we can write, from Weyl theorem, that for any $j = 1, ...r$:

$$|\lambda_j(S) - \lambda_j(\Lambda_0\Lambda_0')| \leq \|S - \Lambda_0\Lambda_0'\|$$

(see for instance, Horn and Johnson (1990) p.181). The result then immediately follows from (i).

iii) By assumptions (CR1) and (CR2), we know that $\frac{1}{n}D_0 = O(1)$ and that $\left(\frac{1}{n}D_0\right)^{-1} = O(1)$. It then results from (ii) that the eigenvalues of $\frac{1}{n}\hat{D}$ and of $\left(\frac{1}{n}\hat{D}\right)^{-1}$ are $O_P(1)$, so that $\frac{1}{n}\hat{D} = O_P(1)$ and $\left(\frac{1}{n}\hat{D}\right)^{-1} = O_P(1)$. The result the follows from (ii) and from the decomposition:

$$n\left(\hat{D}^{-1} - D_0^{-1}\right) = \left(\frac{1}{n}\hat{D}\right)^{-1}\frac{1}{n}\left(\hat{D} - D_0\right)\left(\frac{1}{n}D_0\right)^{-1} = O_P\left(\frac{1}{n}\right) + O_P\left(\frac{1}{\sqrt{T}}\right)$$

iv) $D_0\hat{D}^{-1} = I_r + \frac{D_0}{n}\left[\left(\frac{\hat{D}}{n}\right)^{-1} - \left(\frac{D_0}{n}\right)^{-1}\right]$.

The result then follows from (iii) and assumption (CR2).

Lemma 3 Let us denote $\hat{A} = \hat{P}'P_0$, with $\hat{A} = (\hat{a}_{ij})_{1 \leq i,j \leq r}$.

The following properties hold:

i) $\hat{a}_{ij} = O_P\left(\frac{1}{n}\right) + O_P\left(\frac{1}{\sqrt{T}}\right)$ for $i \neq j$

ii) $\hat{a}_{ii}^2 = 1 + O_P\left(\frac{1}{n}\right) + O_P\left(\frac{1}{\sqrt{T}}\right)$ for $i = 1, ...r$

Proof

i) As $S\hat{P} = \hat{P}\hat{D}$ we have $\hat{P} = S\hat{P}\hat{D}^{-1}$ and:

$$\hat{P}'P_0 = \hat{D}^{-1}\hat{P}'SP_0 = \hat{D}^{-1}\hat{P}'\left(S - \Lambda_0\Lambda_0'\right)P_0 + \hat{D}^{-1}\hat{P}'\Lambda_0\Lambda_0'P_0$$

As $\Lambda_0 = P_0 D_0^{1/2}$, and $P_0'P_0 = I_r$, we have: $\Lambda_0\Lambda_0'P_0 = P_0 D_0$. We then get:

$$\hat{P}'P_0 = \left(\frac{\hat{D}}{n}\right)^{-1}\hat{P}'\left(\frac{S - \Lambda_0\Lambda_0'}{n}\right)P_0 + \left(\frac{\hat{D}}{n}\right)^{-1}\hat{P}'P_0\left(\frac{D_0}{n}\right)$$

As we saw in lemma 2, assumptions (CR1) and (CR2) imply that $\frac{D_0}{n}$ and $\left(\frac{D_0}{n}\right)^{-1}$ are $O(1)$ and that $\frac{\hat{D}}{n}$ and $\left(\frac{\hat{D}}{n}\right)^{-1}$ are $O_P(1)$. As $\hat{P}'\hat{P} = I_r$ and $P_0'P_0 = I_r$, it follows that $\hat{P}'P_0 = O_P(1)$. Thus, lemma 2 (i) and (iii) imply that:

$$\hat{P}'P_0 = O_P\left(\frac{1}{n}\right) + O_P\left(\frac{1}{\sqrt{T}}\right) + \left(\frac{D_0}{n}\right)^{-1}\hat{P}'P_0\left(\frac{D_0}{n}\right).$$

or equivalently that: $\hat{A} = D_0^{-1}\hat{A}D_0 + O_P\left(\frac{1}{n}\right) + O_P\left(\frac{1}{\sqrt{T}}\right)$.

For any i and j the previous relation states that: $\hat{a}_{ij} = \frac{d_{0,jj}}{d_{0,ii}}\hat{a}_{ij} + O_P\left(\frac{1}{n}\right) + O_P\left(\frac{1}{\sqrt{T}}\right)$
For $i \neq j$, we assume, from assumption (A4), that $d_{0,jj} \neq d_{0,ii}$. We then obtain:

$$\hat{a}_{ij} = O_P\left(\frac{1}{n}\right) + O_P\left(\frac{1}{\sqrt{T}}\right) \text{ for } i \neq j.$$

ii) To study the asymptotic behavior of $\hat{a}_{ii}$, let us now use the relation $\hat{D} = \hat{P}'S\hat{P}$ which implies, together with lemma 2 (i), that:

$$\frac{\hat{D}}{n} = \hat{P}'\frac{S}{n}\hat{P} = \hat{P}'\frac{\Lambda_0\Lambda_0'}{n}\hat{P} + O_P\left(\frac{1}{n}\right) + O_P\left(\frac{1}{\sqrt{T}}\right)$$

or, equivalently, that: $\frac{\hat{D}}{n} = \hat{P}'P_0\frac{D_0}{n}P_0'\hat{P} + O_P\left(\frac{1}{n}\right) + O_P\left(\frac{1}{\sqrt{T}}\right)$
It then follows from lemma 2 (ii) that: $\frac{D_0}{n} = \hat{P}'P_0\frac{D_0}{n}P_0'\hat{P} + O_P\left(\frac{1}{n}\right) + O_P\left(\frac{1}{\sqrt{T}}\right)$
or equivalently that: $\frac{D_0}{n} = \hat{A}\frac{D_0}{n}\hat{A}' + O_P\left(\frac{1}{n}\right) + O_P\left(\frac{1}{\sqrt{T}}\right)$.

Thus, for $i = 1, ...r$: $\frac{d_{0,ii}}{n} = \sum_{k=1}^{r}\frac{d_{0,kk}}{n}\hat{a}_{ik}^2 + O_P\left(\frac{1}{n}\right) + O_P\left(\frac{1}{\sqrt{T}}\right)$
and: $\frac{d_{0,ii}}{n}\left(1 - \hat{a}_{ii}^2\right) = \sum_{k \neq i}\frac{d_{0,kk}}{n}\hat{a}_{ik}^2 + O_P\left(\frac{1}{n}\right) + O_P\left(\frac{1}{\sqrt{T}}\right)$

From result (i), we know that $\hat{a}_{ik} = O_P\left(\frac{1}{n}\right) + O_P\left(\frac{1}{\sqrt{T}}\right)$ for $i \neq k$.

As $\frac{D_0}{n} = O_P(1)$, it then follows that: $\hat{a}_{ii}^2 = 1 + O_P\left(\frac{1}{n}\right) + O_P\left(\frac{1}{\sqrt{T}}\right)$ for $i = 1, ...r$.

Lemma 4 Under assumptions (CR1) to (CR4), (A1) to (A4), P_0 and $\hat{P}$ can be defined so as the following properties hold, as $n, T \to \infty$:

(i) $\hat{P}'P_0 = I_r + O_P\left(\frac{1}{n}\right) + O_P\left(\frac{1}{\sqrt{T}}\right)$

(ii) $\|\hat{P} - P_0\|^2 = O_P\left(\frac{1}{n}\right) + O_P\left(\frac{1}{\sqrt{T}}\right)$

(iii) $\tau_{in}'(\hat{\Lambda} - \Lambda_0) = O_P\left(\frac{1}{\sqrt{n}}\right) + O_P\left(\frac{1}{\sqrt{T}}\right), \quad i = 1, ..., n$

where τ_{in} the i^{th} denotes the ith vector of the canonical basis in $\mathbb{R}^n$.

Proof

i) We have seen before that P_0 is uniquely defined up to a sign change of each of its columns, and that this implies that G_t is uniquely defined for any t up to a sign change of each of its components. As $\hat{P}$ is also defined up to a sign change of its columns, it is thus possible to suppose that P_0 and $\hat{P}$ are chosen such that the diagonal terms of $\hat{A} = \hat{P}'P_0$ are positive. In such a case, lemma 2 (ii) implies that:

$$\hat{a}_{ii} = 1 + O_P\left(\frac{1}{n}\right) + O_P\left(\frac{1}{\sqrt{T}}\right) \text{ for } i = 1, ...r$$

We then obtain from lemma 2 (i) that: $\hat{P}'P_0 = I_r + O_P\left(\frac{1}{n}\right) + O_P\left(\frac{1}{\sqrt{T}}\right)$.

ii) Let $x \in \mathbb{R}^n$ a non-random vector such that $\|x\| = 1$. As $\hat{P}'\hat{P} = I_r$ and $P_0'P_0 = I_r$ we have:

$$x'(\hat{P} - P_0)'(\hat{P} - P_0)x = x'(2I_r - \hat{P}'P_0 - P_0'\hat{P})x$$

It then follows from (i) that $x'(\hat{P} - P_0)'(\hat{P} - P_0)x = O_P\left(\frac{1}{n}\right) + O_P\left(\frac{1}{\sqrt{T}}\right)$. As this is true for any $x \in \mathbb{R}^n$, it then follows that

$$\|\hat{P} - P_0\|^2 = O_P\left(\frac{1}{n}\right) + O_P\left(\frac{1}{\sqrt{T}}\right)$$

iii) We have $\hat{P} = S\hat{P}\hat{D}^{-1}$ and $\Sigma_0 = P_0 D_0 P_0' + \Psi_0$, so that

$$\begin{aligned}
\tau_{in}'(\hat{\Lambda} - \Lambda_0) &= \tau_{in}'(\hat{P}\hat{D}^{1/2} - P_0 D_0^{1/2}) \\
&= \tau_{in}'(S\hat{P}\hat{D}^{-1/2} - P_0 D_0^{1/2}) \\
&= \tau_{in}'\left((S - \Sigma_0)\hat{P}\hat{D}^{-1/2} + (P_0 D_0 P_0' + \Psi_0)\hat{P}\hat{D}^{-1/2} - P_0 D_0^{1/2}\right) \\
&= \tau_{in}'(S - \Sigma_0)\hat{P}\hat{D}^{-1/2} + \tau_{in}'\Psi_0\hat{P}\hat{D}^{-1/2} + \tau_{in}'P_0 D_0\left(P_0'\hat{P} - D_0^{-1/2}\hat{D}^{1/2}\right)\hat{D}^{-1/2}
\end{aligned}$$

In order to study the first term, let us first notice that: $\|\tau_{in}'(S-\Sigma_0)\| = \left(\sum_{j=1}^{n}(s_{ij}-\sigma_{0,ij})^2\right)^{1/2}$.

Using the same arguments as in the proof of Lemma 2 (i), we have

$$\mathrm{E}\|\tau_{in}'(S-\Sigma_0)\|^2 = \sum_{j=1}^{n}\mathrm{E}(s_{ij}-\sigma_{0,ij})^2 = O_P\left(\frac{n}{T}\right)$$

so that $\tau_{in}'(S-\Sigma_0) = O_P\left(\frac{\sqrt{n}}{\sqrt{T}}\right)$.

As $\hat{P}'\hat{P} = I_r$, we know that $\hat{P} = O_P(1)$. Then, using $\hat{D}^{-1/2} = O_P\left(\frac{1}{\sqrt{n}}\right)$, it follows that:

$$\tau_{in}'(S-\Sigma_0)\hat{P}\hat{D}^{-1/2} = O_P\left(\frac{1}{\sqrt{T}}\right).$$

Turning to the second term, we have: $\|\tau_{in}'\Psi_0\| \leq \|\Psi_0\| = O(1)$, by assumption (CR2). As $\hat{P} = O_P(1)$ and $\hat{D}^{-1/2} = O_P\left(\frac{1}{\sqrt{n}}\right)$, we get:

$$\tau_{in}'\Psi_0\hat{P}\hat{D}^{-1/2} = O_P\left(\frac{1}{\sqrt{n}}\right)$$

Finally, $\tau_{in}'P_0D_0\left(P_0'\hat{P} - D_0^{-1/2}\hat{D}^{1/2}\right)\hat{D}^{-1/2} = \tau_{in}'\Lambda_0D_0^{1/2}\left(P_0'\hat{P} - D_0^{-1/2}\hat{D}^{1/2}\right)\hat{D}^{-1/2}$.

As $Vx_{it} = \|\tau_{in}'\Lambda_0\|^2 + \psi_{0,ii}$, it follows from assumption (A2) that $\tau_{in}'\Lambda_0 = O(1)$.

Further, $\left(P_0'\hat{P} - D_0^{-1/2}\hat{D}^{1/2}\right) = O_P\left(\frac{1}{n}\right) + O_P\left(\frac{1}{\sqrt{T}}\right)$ by lemma 2 (iv) and lemma 4 (i). As $\hat{D}^{-1/2} = O_P\left(\frac{1}{\sqrt{n}}\right)$ and $D_0^{1/2} = O\left(\sqrt{n}\right)$, it then follows that:

$$\tau_{in}'P_0D_0\left(P_0'\hat{P} - D_0^{-1/2}\hat{D}^{1/2}\right)\hat{D}^{-1/2} = O_P\left(\frac{1}{n}\right) + O_P\left(\frac{1}{\sqrt{T}}\right)$$

which completes the proof.

Proof of proposition 2

We can write:

$$\begin{aligned}
\hat{G}_t - G_t &= \hat{D}^{-1/2}\hat{P}'X_t - G_t \\
&= \hat{D}^{-1/2}\hat{P}'\left(\Lambda_0G_t + \xi_t\right) - G_t \\
&= \left(\hat{D}^{-1/2}\hat{P}'P_0D_0^{1/2} - I_r\right)G_t + \xi_t \\
&= \hat{D}^{-1/2}\left(\hat{P}'P_0 - \hat{D}^{1/2}D_0^{-1/2}\right)D_0^{1/2}G_t + \hat{D}^{-1/2}\hat{P}'\xi_t
\end{aligned}$$

Lemma 2 (iv) and lemma 4 (i) give: $\hat{P}'P_0 - \hat{D}^{1/2}D_0^{-1/2} = O_P\left(\frac{1}{n}\right) + O_P\left(\frac{1}{\sqrt{T}}\right)$.

Then, applying lemma 2 (iv) a second time, and using the fact that $G_t = O_P(1)$, we get:

$$\hat{D}^{-1/2}\left(\hat{P}'P_0 - \hat{D}^{1/2}D_0^{-1/2}\right)D_0^{1/2}G_t = O_P\left(\frac{1}{n}\right) + O_P\left(\frac{1}{\sqrt{T}}\right).$$

In order to study $\hat{D}^{-1/2}\hat{P}'\xi_t$, let us first decompose ξ_t as: $\xi_t = P_0P_0'\xi_t + P_{0\perp}P_{0\perp}'\xi_t$ where $P_{0\perp}$ is a $(n \times (n-r))$ matrix whose columns form an orthonormal basis of the orthogonal space of P_0. We then obtain:

$$\hat{D}^{-1/2}\hat{P}'\xi_t = \hat{D}^{-1/2}\hat{P}'P_0P_0'\xi_t + \hat{D}^{-1/2}\hat{P}'P_{0\perp}P_{0\perp}'\xi_t.$$

First, let us notice that $P_0'\xi_t = O_P(1)$ and that $P_{0\perp}'\xi_t = O_P(\sqrt{n})$.
Indeed, we can write:

$$E\left(\|P_0'\xi_t\|\right)^2 = E\left(\xi_t'P_0P_0'\xi_t\right) = E\left(tr\left(P_0'\xi_t\xi_t'P_0\right)\right) = tr\left(P_0'\Psi_0P_0\right) \leq r\lambda_1\left(\Psi_0\right) = O(1)$$

$$\text{and } E\left(\|P_{0\perp}'\xi_t\|\right)^2 = E\left(tr\left(P_{0\perp}'\xi_t\xi_t'P_{0\perp}\right)\right) = tr\left(P_{0\perp}'\Psi_0P_{0\perp}\right) \leq (n-r)\lambda_1\left(\Psi_0\right) = O(n).$$

As lemma 2 (iii) implies that: $\hat{D}^{-1} = O_P(\frac{1}{n})$, we then get from lemma 4 (i) that:

$$\hat{D}^{-1/2}\hat{P}'P_0P_0'\xi_t = O_P\left(\frac{1}{\sqrt{n}}\right).$$

In order to study the second term, let us first show that:

$$\hat{P}'P_{0\perp} = O_P\left(\frac{1}{n}\right) + O_P\left(\frac{1}{\sqrt{T}}\right)$$

Indeed, if we use: $\hat{P} = S\hat{P}\hat{D}^{-1}$, we can write: $\hat{P}'P_{0\perp} = \hat{D}^{-1}\hat{P}'SP_{0\perp}$.
As P_0 and Λ_0 have the same range, $P_{0\perp}'\Lambda_0 = 0$, so that we also have:

$$\hat{P}'P_{0\perp} = \hat{D}^{-1}\hat{P}'(S - \Lambda_0\Lambda_0')P_{0\perp} = \left(\frac{\hat{D}}{n}\right)^{-1}\hat{P}'\frac{S - \Lambda_0\Lambda_0'}{n}P_{0\perp}.$$

As $P_{0\perp}'P_{0\perp} = I_{n-r}$, we have: $P_{0\perp} = O(1)$. It then follows from lemma 2 (i) and (ii) that:

$$\hat{P}'P_{0\perp} = O_P\left(\frac{1}{n}\right) + O_P\left(\frac{1}{\sqrt{T}}\right)$$

Then, as $\hat{D}^{-1/2} = O_P\left(\frac{1}{\sqrt{n}}\right)$, and $P_{0\perp}'\xi_t = O_P(\sqrt{n})$, it follows that:

$$\hat{D}^{-1/2}\hat{P}'P_{0\perp}P_{0\perp}'\xi_t = O_P\left(\frac{1}{n}\right) + O_P\left(\frac{1}{\sqrt{T}}\right)$$

which completes the proof of the proposition.

Proof of Corollary 1
i) As $\hat{\bar{\psi}} = \frac{1}{n}\text{tr}\hat{\Psi}$, it follows from proposition 2 (iii) and assumption (CR3) that:

$$\hat{\bar{\psi}} = \frac{1}{n}\sum_{i=1}^{n}\left(\psi_{0,ii} + O_P\left(\frac{1}{\sqrt{n}}\right) + O_P\left(\frac{1}{\sqrt{T}}\right)\right) = O_P(1)$$

Since $\hat{D} = O_P(n)$, the result then immediately follows from proposition 2 (i) and the fact that:

$$\hat{G}_{t/T,R1} = \left(\hat{D} + \hat{\bar{\psi}} I_r\right)^{-1}\hat{D}^{1/2}\hat{P}'X_t = \left(\hat{D} + \hat{\bar{\psi}} I_r\right)^{-1}\hat{D}^{-1}\hat{G}_t$$

In order to prove (ii), we first prove the following lemma, which we will also use in the proof of property 5:

Lemma 5 Under assumptions (A1) to (A4), (Cr1) to (CR4), if we denote $\Psi_{0R} = \bar{\psi}_0 I_n$ or $\Psi_{0R} = \Psi_{Od}$, and $\hat{\Psi}_R = \hat{\bar{\psi}} I_n$ or $\hat{\Psi}_R = \hat{\Psi}_d$, the following properties hold:

i) $(\hat{P} - P_0)'\Psi_{0R}^{-1}P_0 = O_P\left(\frac{1}{n}\right) + O_P\left(\frac{1}{\sqrt{T}}\right)$

ii) $\hat{P}'\hat{\Psi}_R^{-1}\hat{P} - P_0'\Psi_{0R}^{-1}P_0 = O_P\left(\frac{1}{n}\right) + O_P\left(\frac{1}{\sqrt{T}}\right)$

iii) $\|\hat{P}'\hat{\Psi}_R^{-1} - P_0'\Psi_{0R}^{-1}\| = O_P\left(\frac{1}{n}\right) + O_P\left(\frac{1}{\sqrt{T}}\right)$

iv) $\|(\hat{P}'\hat{\Psi}_R^{-1}\hat{P})^{-1}\hat{P}'\hat{\Psi}_R^{-1} - (P_0'\Psi_{0R}^{-1}P_0)^{-1}P_0'\Psi_{0R}^{-1}\| = O_P\left(\frac{1}{n}\right) + O_P\left(\frac{1}{\sqrt{T}}\right)$

v) $\frac{1}{n}\|\hat{\Lambda}'\hat{\Psi}_R^{-1}\hat{\Lambda} - \Lambda_0'\Psi_{0R}^{-1}\Lambda_0\| = O_P\left(\frac{1}{n}\right) + O_P\left(\frac{1}{\sqrt{T}}\right)$

vi) $\|(\hat{\Lambda}'\hat{\Psi}_R^{-1}\hat{\Lambda})^{-1}\hat{\Lambda}'\hat{\Psi}_R^{-1} - (\Lambda_0'\Psi_{0R}^{-1}\Lambda_0)^{-1}\Lambda_0'\Psi_{0R}^{-1}\| = O_P\left(\frac{1}{n\sqrt{n}}\right) + O_P\left(\frac{1}{\sqrt{n}\sqrt{T}}\right)$

Proof

i) Defining $P_{0\perp}$ as we did in the proof of proposition 2, we can write:

$$\begin{aligned}(\hat{P} - P_0)'\Psi_{0R}^{-1}P_0 &= (\hat{P} - P_0)'(P_0P_0' + P_{0\perp}P_{0\perp}')\Psi_{0R}^{-1}P_0 \\ &= \hat{P}'P_0P_0'\Psi_{0R}^{-1}P_0 - P_0'\Psi_{0R}^{-1}P_0 + \hat{P}'P_{0\perp}P_{0\perp}'\Psi_{0R}^{-1}P_0\end{aligned}$$

We have seen before (see proof of proposition 2) that:

$$\|\hat{P}'P_{0\perp}\| = O_P\left(\frac{1}{n}\right) + O_P\left(\frac{1}{\sqrt{T}}\right)$$

As $P_0'\Psi_{0R}^{-1}P_0$ and $P_{0\perp}'\Psi_{0R}^{-1}P_0$ are $O(1)$, the result then follows from lemma 4 (i).

ii) $\hat{P}'\hat{\Psi}_R^{-1}\hat{P} - P_0'\Psi_{0R}^{-1}P_0 = \hat{P}'(\hat{\Psi}_R^{-1} - \Psi_{0R}^{-1})\hat{P} + \hat{P}'\Psi_{0R}^{-1}\hat{P} - P_0'\Psi_{0R}^{-1}P_0.$

As $\|\hat{P}'(\hat{\Psi}_R^{-1} - \Psi_{0R}^{-1})\hat{P}\| \leq \|\hat{P}\|^2 \|\hat{\Psi}_R^{-1} - \Psi_{0R}^{-1}\| = \|\hat{\Psi}_R^{-1} - \Psi_{0R}^{-1}\|$,
and as $\|\hat{\Psi}_R^{-1} - \Psi_{0R}^{-1}\| = \text{Max}_{1\leq i\leq n}|\hat{\psi}_{ii}^{-1} - \psi_{0ii}^{-1}|$, it follows from proposition 2 (iii) that

$$\|\hat{P}'(\hat{\Psi}_R^{-1} - \Psi_{0R}^{-1})\hat{P}\| = O_P\left(\frac{1}{n}\right) + O_P\left(\frac{1}{\sqrt{T}}\right)$$

Further:

$$\begin{aligned}\|\hat{P}'\Psi_{0R}^{-1}\hat{P} - P_0'\Psi_{0R}^{-1}P_0\| &= \|(\hat{P}-P_0)'\Psi_{0R}^{-1}P_0 + P_0'\Psi_{0R}^{-1}(\hat{P}-P_0) + (\hat{P}-P_0)'\Psi_{0R}^{-1}(\hat{P}-P_0)\| \\ &\leq 2\|(\hat{P}-P_0)'\Psi_{0R}^{-1}P_0\| + \|\Psi_{0R}^{-1}\|\|\hat{P}-P_0\|^2\end{aligned}$$

It then follows from lemma 4 (ii), assumption (CR2), and lemma 5 (i) that

$$\|\hat{P}'\Psi_{0R}^{-1}\hat{P} - P_0'\Psi_{0R}^{-1}P_0\| = O_P\left(\frac{1}{n}\right) + O_P\left(\frac{1}{\sqrt{T}}\right)$$

so that (ii) follows.

iii) In the same way: $\|\hat{P}'\hat{\Psi}_R^{-1} - P_0'\Psi_{0R}^{-1}\| \leq \|\hat{P}'(\hat{\Psi}_R^{-1} - \Psi_{0R}^{-1})\| + \|(\hat{P}-P_0)'\Psi_{0R}^{-1}\|$ with:

$$\|\hat{P}'(\hat{\Psi}_R^{-1} - \Psi_{0R}^{-1})\| \leq \|\hat{\Psi}_R^{-1} - \Psi_{0R}^{-1}\| = O_P\left(\frac{1}{n}\right) + O_P\left(\frac{1}{\sqrt{T}}\right)$$

and:

$$\begin{aligned}\|(\hat{P}-P_0)'\Psi_{0R}^{-1}\| &= \|(\hat{P}-P_0)'(P_0P_0' + P_{0\perp}P_{0\perp}')\Psi_{0R}^{-1}\| \\ &\leq \|(\hat{P}'P_0 - I_r)P_0'\Psi_{0R}^{-1}\| + \|\hat{P}'P_{0\perp}P_{0\perp}'\Psi_{0R}^{-1}\| \\ &\leq \|\hat{P}'P_0 - I_r\|\|P_0'\Psi_{0R}^{-1}\| + \|\hat{P}'P_{0\perp}\|\|P_{0\perp}'\Psi_{0R}^{-1}\| \\ &= O_P\left(\tfrac{1}{n}\right) + O_P\left(\tfrac{1}{\sqrt{T}}\right)\end{aligned}$$

iv)As $\|\Psi_0^{-1}\| = O(1)$ by asssumption (A4), we know from proposition 2 (iii) that $\|\hat{\Psi}_R^{-1}\| = O_P(1)$ so that $(\hat{P}'\hat{\Psi}_R^{-1}\hat{P})^{-1} = O_P(1)$. We then can write:

$$\begin{aligned}&\|(\hat{P}'\hat{\Psi}_R^{-1}\hat{P})^{-1}\hat{P}'\hat{\Psi}_R^{-1} - (P_0'\Psi_{0R}^{-1}P_0)^{-1}P_0'\Psi_R^{-1}\| \\ &\quad= \|(\hat{P}'\hat{\Psi}_R^{-1}\hat{P})^{-1}(\hat{P}'\hat{\Psi}_R^{-1} - P_0'\Psi_R^{-1}) \\ &\qquad+((\hat{P}'\hat{\Psi}_R^{-1}\hat{P})^{-1} - P_0'\Psi_R^{-1}P_0)^{-1})P_0'\Psi_R^{-1}\| \\ &\quad= \|(\hat{P}'\hat{\Psi}_R^{-1}\hat{P})^{-1}(\hat{P}'\hat{\Psi}_R^{-1} - P_0'\Psi_R^{-1}) \\ &\qquad+(\hat{P}'\hat{\Psi}_R^{-1}\hat{P})^{-1}[P_0'\Psi_R^{-1}P_0 - \hat{P}'\hat{\Psi}_R^{-1}\hat{P}](P_0'\Psi_R^{-1}P_0)^{-1}P_0'\Psi_R^{-1}\| \\ &\quad\leq \|(\hat{P}'\hat{\Psi}_R^{-1}\hat{P})^{-1}\|\|(\hat{P}'\hat{\Psi}_R^{-1} - P_0'\Psi_R^{-1})\| \\ &\qquad+\|(\hat{P}'\hat{\Psi}_R^{-1}\hat{P})^{-1}\|\|P_0'\Psi_R^{-1}P_0 - \hat{P}'\hat{\Psi}_R^{-1}\hat{P}\|\|(P_0'\Psi_R^{-1}P_0)^{-1}\|\|P_0'\Psi_R^{-1}\|\end{aligned}$$

The result then follows from (ii) and (iii).

v) $\frac{1}{n}\hat{\Lambda}'\hat{\Psi}_R^{-1}\hat{\Lambda} = \frac{1}{n}\hat{D}^{1/2}\hat{P}'\hat{\Psi}_R^{-1}\hat{P}\hat{D}^{1/2} = \frac{1}{n}\hat{D}^{1/2}D_0^{-1/2}D_0^{1/2}\hat{P}'\hat{\Psi}_R^{-1}\hat{P}D_0^{1/2}D_0^{-1/2}\hat{D}^{1/2}$.
The result then follows from lemma 2 (iv), lemma 5 (ii), and the fact that $D_0 = O_P\left(\frac{1}{n}\right)$.

vi) We can write:

$$\begin{aligned}&(\hat{\Lambda}'\hat{\Psi}_R^{-1}\hat{\Lambda})^{-1}\hat{\Lambda}'\hat{\Psi}_R^{-1} - (\Lambda_0'\Psi_{0R}^{-1}\Lambda_0)^{-1}\Lambda_0'\Psi_{0R}^{-1} \\ &\qquad= \hat{D}^{-1/2}(\hat{P}'\hat{\Psi}_R^{-1}\hat{P})^{-1}\hat{P}'\hat{\Psi}_R^{-1} - D_0^{-1/2}(P_0'\Psi_{0R}^{-1}P_0)^{-1}P_0'\Psi_{0R}^{-1} \\ &\qquad= \hat{D}^{-1/2}\left((\hat{P}'\hat{\Psi}_R^{-1}\hat{P})^{-1}\hat{P}'\hat{\Psi}_R^{-1} - \hat{D}^{1/2}D_0^{-1/2}(P_0'\Psi_{0R}^{-1}P_0)^{-1}P_0'\Psi_{0R}^{-1}\right)\end{aligned}$$

Proof of Corollary 1 (ii)

As $G_{t/T,R2} = (\hat{\Lambda}'\hat{\Psi}_d^{-1}\hat{\Lambda}+I_r)^{-1}\hat{\Lambda}'\hat{\Psi}_d^{-1}X_t = (\hat{\Lambda}'\hat{\Psi}_d^{-1}\hat{\Lambda}+I_r)^{-1}\hat{\Lambda}'\hat{\Psi}_d^{-1}\,(\Lambda_0 G_t + \xi_t)$, we have:

$$\begin{aligned}
\|G_{t/T,R2} - G_t\| \quad \leq \quad & \|(\hat{\Lambda}'\hat{\Psi}_d^{-1}\hat{\Lambda} + I_r)^{-1}\hat{\Lambda}'\hat{\Psi}_d^{-1} - (\Lambda_0'\Psi_{0d}^{-1}\Lambda_0 + I_r)^{-1}\Lambda_0'\Psi_{0d}^{-1}\|\,(\|\Lambda_0 G_t + \xi_t\|) \\
& +\|(\Lambda_0'\Psi_{0d}^{-1}\Lambda_0 + I_r)^{-1}\Lambda_0'\Psi_{0d}^{-1}\Lambda_0 - I_r\|\|G_t\| \\
& +\|(\Lambda_0'\Psi_{0d}^{-1}\Lambda_0 + I_r)^{-1}\Lambda_0'\Psi_{0d}^{-1}|\|\|\xi_t\|
\end{aligned}$$

with:

. $\|(\hat{\Lambda}'\hat{\Psi}_d^{-1}\hat{\Lambda} + I_r)^{-1}\hat{\Lambda}'\hat{\Psi}_d^{-1} - (\Lambda_0'\Psi_{0d}^{-1}\Lambda_0 + I_r)^{-1}\Lambda_0'\Psi_{0d}^{-1}\| = O_P\left(\frac{1}{n\sqrt{n}}\right) + O_P\left(\frac{1}{\sqrt{nT}}\right)$ by lemma 5 (vi)

. $\|(\Lambda_0'\Psi_{0d}^{-1}\Lambda_0 + I_r)^{-1}\Lambda_0'\Psi_{0d}^{-1}\Lambda_0 - I_r\| = O\left(\frac{1}{\sqrt{n}}\right)$ by assumptions (CR1) to (CR4)

. $\|(\Lambda_0'\Psi_{0d}^{-1}\Lambda_0 + I_r)^{-1}\Lambda_0'\Psi_{0d}^{-1}|\| = O\left(\frac{1}{\sqrt{n}}\right)$ by assumptions (CR1) to (CR4)

. $\|G_t\| = O_P(1)$, $\|\xi_t\| = O_P(\sqrt{n})$ and $\|\Lambda_0 G_t + \xi_t\| \leq \|\Lambda_0\| + \|G_t\| + \|\xi_t\| = O_P(\sqrt{n})$

The result immediately follows.

A.3 Consistency of Kalman Filtering: ($\hat{\Omega}^{R3}$ and $\hat{\Omega}^{R4}$ framework)

Proof of Proposition 3

i) Consider the sample autocovariance of the estimated principal components

$$\hat{\Gamma}_{\hat{G}}(h) = \frac{1}{T-h}\sum_{t=h+1}^{T}\hat{G}_t\hat{G}'_{t-h} = \hat{D}^{-1/2}\hat{P}'S(h)\hat{P}\hat{D}^{-1/2}$$

with $S(h) = \frac{1}{T-h}\sum_{t=h+1}^{T} X_t X'_{t-h}$.

For any $h < T$, we can decompose $\hat{\Gamma}_{\hat{G}}(h)$ as:

$$\hat{\Gamma}_{\hat{G}}(h) = \hat{D}^{-1/2}\hat{P}'\Lambda_0\Phi_0(h)\Lambda_0'\hat{P}\hat{D}^{-1/2} + \hat{D}^{-1/2}\hat{P}'\left(S(h) - \Lambda_0\Phi_0(h)\Lambda_0'\right)\hat{P}\hat{D}^{-1/2}$$

First, we can write:

$$\hat{D}^{-1/2}\hat{P}'\Lambda_0\Phi_0(h)\Lambda_0'\hat{P}\hat{D}^{-1/2} = \hat{D}^{-1/2}\hat{P}'P_0D_0^{1/2}\Phi_0(h)D_0^{1/2}P_0'\hat{P}\hat{D}^{-1/2}$$

It then follows from lemma 2 (iv), lemma 4 (i), and the fact that $\Phi_0(h) = O(1)$ that:

$$\hat{D}^{-1/2}\hat{P}'\Lambda_0\Phi_0(h)\Lambda_0'\hat{P}\hat{D}^{-1/2} = \Phi_0(h) + O_P\left(\frac{1}{n}\right) + O_P\left(\frac{1}{\sqrt{T}}\right)$$

Then, under assumption (A3) and (CR3), it is possible to extend what has been done in lemma 2 (i) for $h = 0$, and to show that: $\frac{1}{n}\|S(h) - \Lambda_0\Phi_0(h)\Lambda_0')\| = O_P\left(\frac{1}{n}\right) + O_P\left(\frac{1}{\sqrt{T}}\right)$, uniformly in $h \le p$.

Indeed, if we decompose $S(h)$ as:

$$S(h) = \frac{1}{T-h}\left[\Lambda_0\sum_{t=h+1}^{T}G_tG'_{t-h}\Lambda_0' + \Lambda_0\sum_{t=h+1}^{T}G_t\xi'_{t-h} + \sum_{t=h+1}^{T}\xi_tG'_{t-h}\Lambda_0' + \sum_{t=h+1}^{T}\xi_t\xi'_{t-h}\right]$$

we get:

$$\begin{aligned}\tfrac{1}{n}(S(h) - \Lambda_0\Phi_0(h)\Lambda_0') = &\tfrac{1}{n}\Lambda_0\left(\tfrac{1}{T-h}\textstyle\sum_{t=h+1}^{T}G_tG'_{t-h} - \Phi_0(h)\right)\Lambda_0' \\ &+\tfrac{1}{n}\left(\Lambda_0\tfrac{1}{T-h}\textstyle\sum_{t=h+1}^{T}G_t\xi'_{t-h} + \tfrac{1}{T-h}\sum_{t=h+1}^{T}\xi_tG'_{t-h}\Lambda_0'\right) \\ &+\tfrac{1}{n}\left(\tfrac{1}{T-h}\textstyle\sum_{t=h+1}^{T}\xi_t\xi'_{t-h} - \Psi_0(h)\right) + \tfrac{1}{n}\Psi_0(h)\end{aligned}$$

Then, using assumptions (A3) and (CR3) and a multivariate extension of the proof given in the univariate case by Brockwell and Davies (1991, pp226-227), it is possible, as in lemma 2 (i), to show that:

- $\mathrm{E}\left(\|\frac{1}{T-h}\sum_{t=h+1}^{T}G_tG'_{t-h} - \Phi_0(h)\|^2\right) = O\left(\frac{1}{T}\right)$
- $\mathrm{E}\left(\|\frac{1}{T-h}\sum_{t=h+1}^{T}\xi_t\xi'_{t-h} - \Psi_0(h)\|^2\right) = O\left(\frac{n^2}{T}\right)$

so that:

- $\|\frac{1}{T-h}\sum_{t=h+1}^{T} G_t G'_{t-h} - \Phi_0(h)\| = O_P\left(\frac{1}{\sqrt{T}}\right)$
- $\|\frac{1}{T-h}\sum_{t=h+1}^{T} \xi_t \xi'_{t-h} - \Psi_0(h)\| = O_P\left(\frac{n}{\sqrt{T}}\right)$

Using the and the same kind of arguments as we have used in lemma 2 (i), it then also follows that:

$$\|\frac{1}{T-h}\sum_{t=h+1}^{T} G_t \xi'_{t-h}\| = O_P\left(\frac{\sqrt{n}}{\sqrt{T}}\right)$$

From assumption (CR1), we also have: $\|\Lambda_0\| = O\left(\sqrt{n}\right)$ and $\|\Psi_0(h)\| = O\left(1\right)$, so that:

$$\frac{1}{n}\|S(h) - \Lambda_0\Phi_0(h)\Lambda_0'\| = O_P\left(\frac{1}{n}\right) + O_P\left(\frac{1}{\sqrt{T}}\right)$$

Finally, as $\hat{D}^{-1/2}\hat{P}'\left(S(h) - \Lambda_0\Phi_0(h)\Lambda_0'\right)\hat{P}\hat{D}^{-1/2} = (\frac{\hat{D}}{n})^{-1/2}\hat{P}'\frac{S(h)-\Lambda_0\Phi_0(h)\Lambda_0'}{n}\hat{P}(\frac{\hat{D}}{n})^{-1/2}$, and $\frac{\hat{D}}{n} = O_P(1)$, it follows that

$$\hat{D}^{-1/2}\hat{P}'\left(S(h) - \Lambda_0\Phi_0(h)\Lambda_0'\right)\hat{P}\hat{D}^{-1/2} = O_P\left(\frac{1}{n}\right) + O_P\left(\frac{1}{\sqrt{T}}\right)$$

ii) Let us first recall that any VAR(p) model can be written in a VAR(1) form. More precisely, if we denote: $G_t^{(p)} = (G_t', G_{t-1}', \ldots, G_{t-p+1}')'$, we can write:

$$G_t^{(p)} = A_0^{(p)} G_{t-1}^{(p)} + w_t^{(p)}$$

with $A_0^{(p)} = \begin{pmatrix} A_{01} & A_{02} & \ldots & A_{0p} \\ I_r & 0 & \ldots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \ldots & I_r \end{pmatrix}$ and $w_t^{(p)} = (w_t', 0, \ldots, 0)'$.

If we denote $\Phi_0^{(p)} = E\left[G_t^{(p)} G_t^{(p)\prime}\right]$ and $\Phi_1^{(p)} = E\left[G_t^{(p)} G_{t-1}^{(p)\prime}\right]$, so that:

$$\Phi_0^{(p)} = \begin{pmatrix} I_r & \Phi_0(1) & \ldots & \Phi_0(p-1) \\ \Phi_0'(1) & I_r & \ldots & \Phi_0(p-2) \\ \vdots & \vdots & & \vdots \\ \Phi_0'(p-1) & \Phi_0'(p-2) & \ldots & I_r \end{pmatrix} \Phi_1^{(p)} = \begin{pmatrix} \Phi_0(1) & \Phi_0(2) & \ldots & \Phi_0(p) \\ \Phi_0'(1) & I_r & \ldots & \Phi_0(p-1) \\ \vdots & \vdots & & \vdots \\ \Phi_0'(p-2) & \Phi_0'(p-3) & \ldots & \Phi_0(1) \end{pmatrix}$$

we have:

$$A_0^{(p)} = \Phi_1^{(p)}(\Phi_0^{(p)})^{-1}$$

We can define $\hat{\Phi}_0^{(p)}$ and $\hat{\Phi}_1^{(p)}$ having respectively the same form as $\Phi_0^{(p)}$ and $\Phi_1^{(p)}$, with $\Phi_{0,k}$ replaced by $\hat{\Gamma}_{\hat{G}}(k)$ for any value of k. Then, we also have:

$$\hat{A}^{(p)} = \hat{\Phi}_1^{(p)}(\hat{\Phi}_0^{(p)})^{-1}$$

where: $\hat{A}^{(p)} = \begin{pmatrix} \hat{A}_1 & \hat{A}_2 & \dots & \hat{A}_p \\ I_r & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & I_r \end{pmatrix}$.

It thus follows from (i) that:

$$\|\Phi_0^{(p)} - \hat{\Phi}_0^{(p)}\| = O_P\left(\frac{1}{n}\right) + O_P\left(\frac{1}{\sqrt{T}}\right) \text{ and } \|\Phi_1^{(p)} - \hat{\Phi}_1^{(p)}\| = O_P\left(\frac{1}{n}\right) + O_P\left(\frac{1}{\sqrt{T}}\right)$$

We have: $\|A_0^{(p)} - \hat{A}^{(p)}\| \leq \|\Phi_1^{(p)} - \hat{\Phi}_1^{(p)}\| \|(\Phi_0^{(p)})^{-1}\| + \|\hat{\Phi}_1^{(p)}\| \|(\Phi_0^{(p)})^{-1} - (\hat{\Phi}_0^{(p)})^{-1}\|$. If we apply to the last term the relation (R) which has been introduced in the proof of proposition 1, we then get:

$$\|A_0^{(p)} - \hat{A}^{(p)}\| = O_P\left(\frac{1}{n}\right) + O_P\left(\frac{1}{\sqrt{T}}\right)$$

It then follows that: $\|A_{0s} - \hat{A}_s\| = O_P\left(\frac{1}{n}\right) + O_P\left(\frac{1}{\sqrt{T}}\right)$ for any $s = 1, \dots, p$.

Proof of proposition 4

i) If we denote by $S_{G,R}(\omega)$ and $\hat{S}_{G,R}(\omega)$ the spectral density matrices of G_t under Ω_R and $\hat{\Omega}_R$, for $R = R3$ and $R4$, we can apply the same result as in lemma 1 (ii) and we get:

$$\begin{aligned} \|\hat{\Sigma}_{G,R} - \Sigma_{G,R}\| &\leq 2\pi \text{Max}_{\omega \in [-\pi, +\pi]} \lambda_{max}\left(\hat{S}_{G,R}(\omega) - S_{G,R}(\omega)\right) \\ \text{or } \|\hat{\Sigma}_{G,R} - \Sigma_{G,R}\| &\leq 2\pi \text{Max}_{\omega \in [-\pi, +\pi]} \|\hat{S}_{G,R}(\omega) - S_{G,R}(\omega)\| \end{aligned}$$

As $A_0(L)G_t = w_t$ and $VG_t = I_r$, we know that $Vw_t = W_0$, with

$$W_0 = E(u_t G_t') = E\left((G_t - \sum_{s=1}^{p} A_{0s} G_{t-s}) G_t'\right) = I_r - \sum_{s=1}^{p} A_{0s} \Phi_0'(s)$$

In the same way, we have: $\hat{V} w_t = \hat{W} = \hat{\Gamma}_{\hat{G}}(0) - \sum_{s=1}^{p} \hat{A}_s \hat{\Gamma}'_{\hat{G}}(s)$.

We thus get:

$$\begin{aligned} \|W_0 - \hat{W}\| &= \|(I_r - \hat{\Gamma}_{\hat{G}}(0)) - \sum_{s=1}^{p} \left(\hat{A}_s \hat{\Gamma}'_{\hat{G}}(s) - A_{0s} \Phi_0'(s)\right)\| \\ &\leq \|I_r - \hat{\Gamma}_{\hat{G}}(0)\| + \sum_{s=1}^{p} \|\hat{A}_s - A_{0s}\| \|\hat{\Gamma}_{\hat{G}}(s)\| + \sum_{s=1}^{p} \|A_{0s}\| \|\hat{\Gamma}_{\hat{G}}(s) - \Phi_0(s)\| \end{aligned}$$

It then follows from proposition 3 (i) and (ii) that:

$$\|W_0 - \hat{W}\| = O_P\left(\frac{1}{n}\right) + O_P\left(\frac{1}{\sqrt{T}}\right)$$

Turning now to the spectral density matrices, we have:

$$\begin{aligned} S_{G,R}(\omega) &= \frac{1}{2\pi}\left(A_0(e^{i\omega})\right)^{-1} W_0 \left(A_0'(e^{-i\omega})\right)^{-1} \\ \text{and } \hat{S}_{G,R}(\omega) &= \frac{1}{2\pi}\left(\hat{A}(e^{i\omega})\right)^{-1} \hat{W} \left(\hat{A}'(e^{-i\omega})\right)^{-1}. \end{aligned}$$

As $\| \left(A_0'(e^{-i\omega})\right)^{-1} - \left(\hat{A}'(e^{-i\omega})\right)^{-1} \| \leq \| \left(A_0'(e^{-i\omega})\right)^{-1} \| \| A_0'(e^{-i\omega}) - \hat{A}'(e^{-i\omega}) \| \| \left(\hat{A}'(e^{-i\omega})\right)^{-1} \|$, we have:

$$\text{Max}_{\omega\in[-\pi,+\pi]} \| \left(A_0'(e^{-i\omega})\right)^{-1} - \left(\hat{A}'(e^{-i\omega})\right)^{-1} \| = O_P\left(\frac{1}{n}\right) + O_P\left(\frac{1}{\sqrt{T}}\right)$$

Using the fact that $\|W_0 - \hat{W}\| = O_P\left(\frac{1}{n}\right) + O_P\left(\frac{1}{\sqrt{T}}\right)$, it then immediately follows that:

$$\text{Max}_{\omega\in[-\pi,+\pi]} \|S_{G,R}(\omega) - \hat{S}_{G,R}(\omega)\| = O_P\left(\frac{1}{n}\right) + O_P\left(\frac{1}{\sqrt{T}}\right)$$

which gives the desired result.

ii) Since we know, from lemma 1 (ii) that $\|\Sigma_{G,R}\| = O(1)$, it follows from (i) that $\|\hat{\Sigma}_{G,R}\| = O_P(1)$.

Further: $\|\hat{\Sigma}_{G,R}\|^{-1} = \frac{1}{\lambda_{min}(\hat{\Sigma}_{G,R})}$, with $|\lambda_{min}(\hat{\Sigma}_{G,R}) - \lambda_{min}(\hat{\Sigma}_{G,R})| \leq \|\Sigma_{G,R} - \hat{\Sigma}_{G,R}\|$ by Weyl theorem. It then follows from (i) and from lemma 1 (ii) that $\|\hat{\Sigma}_{G,R}\|^{-1} = O_P(1)$.

Finally, as $\|\hat{\Sigma}^{-1}_{\hat{G},R} - \Sigma^{-1}_{G,R}\| \leq \|\hat{\Sigma}^{-1}_{\hat{G},R}\| \|\hat{\Sigma}_{\hat{G},R} - \Sigma_{G,R}\| \|\Sigma^{-1}_{G,R}\|$, we also obtain:

$$\|\hat{\Sigma}^{-1}_{\hat{G},R} - \Sigma^{-1}_{G,R}\| = O_P\left(\frac{1}{n}\right) + O_P\left(\frac{1}{\sqrt{T}}\right)$$

Proof of proposition 5

As $G_{t/t,R} = \text{Proj}_{\Omega^R}[G_t|X_t]$ and $\hat{G}_{t/t,R} = \text{Proj}_{\hat{\Omega}^R}[G_t|X_t]$, they are obtained through the same formulas so that, by construction:

$$\hat{G}_{t/T,R} = \mathbf{U}_t'(\hat{\Sigma}^{-1}_{\hat{G},R} + I_T \otimes \hat{\Lambda}'\hat{\Psi}_R^{-1}\hat{\Lambda})^{-1}(I_T \otimes \hat{\Lambda}'\hat{\Psi}_R^{-1})\mathbf{X}_T$$

Using relation (R) as in the proof of proposition 1 (Taylor expansion at order 0), we obtain the same kind of decomposition for $\hat{G}_{t/T,R}$ as the one we have used to study $G_{t/T,R}$. Thus, if we denote $\hat{M} = \hat{\Lambda}'\hat{\Psi}_R^{-1}\hat{\Lambda}$, we can write: $\hat{G}_{t/T,R} = \hat{G}^1_{t/T,R} - \hat{G}^2_{t/T,R} - \hat{G}^3_{t/T,R}$, with:

$$\begin{aligned} \hat{G}^1_{t/T,R} &= \mathbf{U}_t'\left(I_T \otimes \hat{M}^{-1}\right)\left(I_T \otimes \hat{\Lambda}'\hat{\Psi}_R^{-1}\right)\mathbf{X}_T = \hat{M}^{-1}\hat{\Lambda}'\hat{\Psi}_R^{-1}X_t \\ \hat{G}^2_{t/T,R} &= \mathbf{U}_t'\left(\hat{\Sigma}^{-1}_{\hat{G},R} + I_T \otimes \hat{M}\right)^{-1}\hat{\Sigma}^{-1}_{\hat{G},R}\left(I_T \otimes \hat{M}^{-1}\hat{\Lambda}'\hat{\Psi}_R^{-1}\right)(I_T \otimes \Lambda_0)\,\mathbf{G}_T \\ \hat{G}^3_{t/T,R} &= \mathbf{U}_t'\left(\hat{\Sigma}^{-1}_{\hat{G},R} + I_T \otimes \hat{M}\right)^{-1}\hat{\Sigma}^{-1}_{\hat{G},R}\left(I_T \otimes \hat{M}^{-1}\hat{\Lambda}'\hat{\Psi}_R^{-1}\right)\mathbf{Z}_T \end{aligned}$$

Let us study separately these three terms.

If we compare the first term with $G^1_{t/T,R}$, we get:

$$\hat{G}^1_{t/T,R} - G^1_{t/T,R} = \left(\left(\hat{\Lambda}'\hat{\Psi}_R^{-1}\hat{\Lambda}\right)^{-1}\hat{\Lambda}'\hat{\Psi}_R^{-1} - \left(\Lambda_0'\Psi_{0R}^{-1}\Lambda_0\right)^{-1}\Lambda_0'\Psi_{0R}^{-1}\right) X_t$$

with

$$\begin{aligned}\| &\left(\left(\hat{\Lambda}'\hat{\Psi}_R^{-1}\hat{\Lambda}\right)^{-1}\hat{\Lambda}'\hat{\Psi}_R^{-1} - \left(\Lambda_0'\Psi_{0R}^{-1}\Lambda_0\right)^{-1}\Lambda_0'\Psi_{0R}^{-1}\right) X_t\| \\ &\qquad \leq \| \left(\hat{\Lambda}'\hat{\Psi}_R^{-1}\hat{\Lambda}\right)^{-1}\hat{\Lambda}'\hat{\Psi}_R^{-1} - \left(\Lambda_0'\Psi_{0R}^{-1}\Lambda_0\right)^{-1}\Lambda_0'\Psi_{0R}^{-1}\|\|X_t\|\end{aligned}$$

As $X_t = O_P(\sqrt{n})$, it then follows from lemma 5 (v) that:

$$\hat{G}^1_{t/T,R} - G^1_{t/T,R} = O_P(\frac{1}{n}) + O_P(\frac{1}{\sqrt{T}})$$

Finally, as $G^1_{t/T,R} = G_t + O_P(\frac{1}{\sqrt{n}})$, we get:

$$\hat{G}^1_{t/T,R} = G_t + O_P(\frac{1}{\sqrt{n}}) + O_P(\frac{1}{\sqrt{T}})$$

In the same way, we can write:

$$\begin{aligned}\hat{G}^2_{t/T,R} - G^2_{t/T,R} = &\mathbf{U}_t'\left(\hat{\Sigma}^{-1}_{\hat{G},R} + I_T\otimes\hat{M}\right)^{-1}\hat{\Sigma}^{-1}_{\hat{G},R}\left(I_T\otimes\hat{M}^{-1}\hat{\Lambda}'\hat{\Psi}_R^{-1}\right)(I_T\otimes\Lambda_0)\,\mathbf{G}_T \\ &-\mathbf{U}_t'\left(\Sigma^{-1}_{G,R} + I_T\otimes M_0\right)^{-1}\Sigma^{-1}_{G,R}\left(I_T\otimes M_0^{-1}\Lambda_0'\Psi_{0R}^{-1}\right)(I_T\otimes\Lambda_0)\,\mathbf{G}_T\end{aligned}$$

and:

$$\begin{aligned}\hat{G}^3_{t/T,R} - G^3_{t/T,R} = &\mathbf{U}_t'\left(\hat{\Sigma}^{-1}_{\hat{G},R} + I_T\otimes\hat{M}\right)^{-1}\hat{\Sigma}^{-1}_{\hat{G},R}\left(I_T\otimes\hat{M}^{-1}\hat{\Lambda}'\hat{\Psi}_R^{-1}\right)\mathbf{Z}_T \\ &-\mathbf{U}_t'\left(\Sigma^{-1}_{G,R} + I_T\otimes M_0\right)^{-1}\Sigma^{-1}_{G,R}\left(I_T\otimes M_0^{-1}\Lambda_0'\Psi_{0R}^{-1}\right)\mathbf{Z}_T\end{aligned}$$

so that: $\hat{G}^2_{t/T,R} - G^2_{t/T,R} = \mathbf{U}_t'\hat{H}\,(I_T\otimes\Lambda_0)\,\mathbf{G}_T$ and $\hat{G}^3_{t/T,R} - G^3_{t/T,R} = \mathbf{U}_t'\hat{H}\mathbf{Z}_T$ with:

$$\hat{H} = (\hat{\Sigma}^{-1}_{\hat{G},R} + I_T\otimes\hat{M})^{-1}\hat{\Sigma}^{-1}_{\hat{G},R}(I_T\otimes\hat{M}^{-1}\hat{\Lambda}'\hat{\Psi}_R^{-1}) - (\Sigma^{-1}_{G,R} + I_T\otimes M_0)^{-1}\Sigma^{-1}_{G,R}(I_T\otimes M_0^{-1}\Lambda_0'\Psi_{0R}^{-1})$$

We can also decompose $\hat{H}$ as: $\hat{H} = \hat{H}_1 + \hat{H}_2 + \hat{H}_3$ with:

$$\begin{aligned}\hat{H}_1 &= (\hat{\Sigma}^{-1}_{\hat{G},R} + I_T\otimes\hat{M})^{-1}\hat{\Sigma}^{-1}_{\hat{G},R}\left(I_T\otimes(\hat{M}^{-1}\hat{\Lambda}'\hat{\Psi}_R^{-1} - M_0^{-1}\Lambda_0'\Psi_{0R}^{-1})\right) \\ \hat{H}_2 &= (\hat{\Sigma}^{-1}_{\hat{G},R} + I_T\otimes\hat{M})^{-1}(\hat{\Sigma}^{-1}_{\hat{G},R} - \Sigma^{-1}_{G,R})(I_T\otimes M_0^{-1}\Lambda_0'\Psi_{0R}^{-1}) \\ \hat{H}_3 &= \left((\hat{\Sigma}^{-1}_{\hat{G},R} + I_T\otimes\hat{M})^{-1} - (\Sigma^{-1}_{G,R} + I_T\otimes M_0)^{-1}\right)\Sigma^{-1}_{G,R}(I_T\otimes M_0^{-1}\Lambda_0'\Psi_{0R}^{-1})\end{aligned}$$

We then get:

$$\begin{aligned}
\|\hat{H}_1\| &\leq \|(\hat{\Sigma}_{\hat{G},R}^{-1} + I_T \otimes \hat{M})^{-1}\| \|\hat{\Sigma}_{\hat{G},R}^{-1}\| \|I_T \otimes (\hat{M}^{-1}\hat{\Lambda}'\hat{\Psi}_R^{-1} - M_0^{-1}\Lambda_0'\Psi_{0R}^{-1})\| \\
&\leq \|I_T \otimes \hat{M}^{-1}\| \|\hat{\Sigma}_{\hat{G},R}^{-1}\| \|I_T \otimes (\hat{M}^{-1}\hat{\Lambda}'\hat{\Psi}_R^{-1} - M_0^{-1}\Lambda_0'\Psi_{0R}^{-1})\| \\
&= \|\hat{M}^{-1}\| \|\hat{\Sigma}_{\hat{G},R}^{-1}\| \|\hat{M}^{-1}\hat{\Lambda}'\hat{\Psi}_R^{-1} - M_0^{-1}\Lambda_0'\Psi_{0R}^{-1}\|
\end{aligned}$$

$$\begin{aligned}
\|\hat{H}_2\| &\leq \|(\hat{\Sigma}_{\hat{G},R}^{-1} + I_T \otimes \hat{M})^{-1}\| \|\hat{\Sigma}_{\hat{G},R}^{-1} - \Sigma_{G,R}^{-1}\| \|I_T \otimes M_0^{-1}\Lambda_0'\Psi_{0R}^{-1}\| \\
&\leq \|\hat{M}^{-1}\| \|\hat{\Sigma}_{\hat{G},R}^{-1} - \Sigma_{G,R}^{-1}\| \|M_0^{-1}\Lambda_0'\Psi_{0R}^{-1}\|
\end{aligned}$$

$$\begin{aligned}
\|\hat{H}_3\| &\leq \|(\hat{\Sigma}_{\hat{G},R}^{-1} + I_T \otimes \hat{M})^{-1} - (\Sigma_{G,R}^{-1} + I_T \otimes M_0)^{-1}\| \|\Sigma_{G,R}^{-1}\| \|I_T \otimes M_0^{-1}\Lambda_0'\Psi_{0R}^{-1}\| \\
&\leq \|(\hat{\Sigma}_{\hat{G},R}^{-1} + I_T \otimes \hat{M})^{-1}\| \|\hat{\Sigma}_{\hat{G},R}^{-1} + I_T \otimes \hat{M} - \Sigma_{G,R}^{-1} - I_T \otimes M_0\| \\
&\quad \times \|(\Sigma_{G,R}^{-1} + I_T \otimes M_0)^{-1}\| \|\Sigma_{G,R}^{-1}\| \|M_0^{-1}\Lambda_0'\Psi_{0R}^{-1}\| \\
&\leq \|\hat{M}^{-1}\| \left[\|\hat{\Sigma}_{\hat{G},R}^{-1} - \Sigma_{G,R}^{-1}\| + \|\hat{M} - M_0\|\right] \|M_0^{-1}\| \|\Sigma_{G,R}^{-1}\| \|M_0^{-1}\Lambda_0'\Psi_{0R}^{-1}\|
\end{aligned}$$

From lemma 5 (v), we get: $\hat{M}^{-1} = O_P\left(\frac{1}{n}\right)$. Thus, applying lemma 5 (v) and (vi), and proposition 4, we get that:

$$\|\hat{H}_i\| = O_P\left(\frac{1}{n^2\sqrt{n}}\right) + O_P\left(\frac{1}{n\sqrt{nT}}\right) \text{ for } i = 1 \text{ to } 3$$

so that $\|\hat{H}\| = O_P\left(\frac{1}{n^2\sqrt{n}}\right) + O_P\left(\frac{1}{n\sqrt{nT}}\right)$.

As $E\left(\|\mathbf{G}_T\|\right)^2 = E\left(\sum_{t=1}^{T}\|G_t\|^2\right) = rT$, we have: $\|\mathbf{G}_T\| = O_P\left(\sqrt{T}\right)$ so that:

$$\|\hat{G}_{t/T,R}^2 - G_{t/T,R}^2\| \leq \|\mathbf{U}_t\| \|\hat{H}\| \|I_T \otimes \Lambda_0\| \|\mathbf{G}_T\| = O_P\left(\frac{\sqrt{T}}{n^2}\right) + O_P\left(\frac{1}{n}\right)$$

Similarly, $E\left(\|\mathbf{Z}_T\|\right)^2 = E\left(\sum_{t=1}^{T}\|\xi_t\|^2\right) = Ttr(\Psi_0) = O(nT)$, so that:

$$\|\hat{G}_{t/T,R}^3 - G_{t/T,R}^3\| \leq \|\mathbf{U}_t\| \|\hat{H}\| \|\mathbf{Z}_T\| = O_P\left(\frac{\sqrt{T}}{n^2}\right) + O_P\left(\frac{1}{n}\right)$$

Finally, as we know, from the proof of proposition 1 that:

$$G_{t/T,R}^2 = O_P\left(\frac{1}{n}\right) \text{ and } G_{t/T,R}^3 = O_P\left(\frac{1}{n\sqrt{n}}\right)$$

we get: $\hat{G}_{t/T,R}^2 + \hat{G}_{t/T,R}^3 = O_P\left(\frac{\sqrt{T}}{n^2}\right) + O_P\left(\frac{1}{n}\right)$, so that:

$$\hat{G}_{t/T,R} = G_t + O_P\left(\frac{1}{\sqrt{n}}\right) + O_P\left(\frac{1}{\sqrt{T}}\right) + O_P\left(\frac{\sqrt{T}}{n^2}\right)$$

If limsup$\frac{T}{n^3} = O(1)$, we then get:

$$\hat{G}_{t/T,R} = G_t + O_P\left(\frac{1}{\sqrt{n}}\right) + O_P\left(\frac{1}{\sqrt{T}}\right)$$