

**Finance and Economics Discussion Series
Divisions of Research & Statistics and Monetary Affairs
Federal Reserve Board, Washington, D.C.**

Common Factors, Trends, and Cycles in Large Datasets

Matteo Barigozzi and Matteo Luciani

2017-111

Please cite this paper as:

Barigozzi, Matteo, and Matteo Luciani (2017). "Common Factors, Trends, and Cycles in Large Datasets," Finance and Economics Discussion Series 2017-111. Washington: Board of Governors of the Federal Reserve System, <https://doi.org/10.17016/FEDS.2017.111>.

NOTE: Staff working papers in the Finance and Economics Discussion Series (FEDS) are preliminary materials circulated to stimulate discussion and critical comment. The analysis and conclusions set forth are those of the authors and do not indicate concurrence by other members of the research staff or the Board of Governors. References in publications to the Finance and Economics Discussion Series (other than acknowledgement) should be cleared with the author(s) to protect the tentative character of these papers.

COMMON FACTORS, TRENDS, AND CYCLES IN LARGE DATASETS

Matteo Barigozzi
London School of Economics
m.barigozzi@lse.ac.uk

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

November 6, 2017

Abstract

This paper considers a non-stationary dynamic factor model for large datasets to disentangle long-run from short-run co-movements. We first propose a new Quasi Maximum Likelihood estimator of the model based on the Kalman Smoother and the Expectation Maximisation algorithm. The asymptotic properties of the estimator are discussed. Then, we show how to separate trends and cycles in the factors by mean of eigenanalysis of the estimated non-stationary factors. Finally, we employ our methodology on a panel of US quarterly macroeconomic indicators to estimate aggregate real output, or Gross Domestic Output, and the output gap.

JEL classification: C32, C38, C55, E0.

Keywords: Non-stationary Approximate Dynamic Factor Model; Trend-Cycle Decomposition; Quasi Maximum Likelihood; EM Algorithm; Kalman Smoother; Gross Domestic Output; Output Gap.

*We thank for helpful comment the participants to the conferences: “Inference in Large Econometric Models”, CIREQ, Montréal, May 2017; “Big Data in Dynamic Predictive Econometric Modelling”, University of Pennsylvania, Philadelphia, May 2017; “Computing in Economics and Finance”, Fordham University, New York City, June 2017; and to the seminars at: Federal Reserve Board, September 2016; Warwick Business School, February 2017; Department of Statistics, Universidad Carlos III, Madrid, May 2017. We would like also to thank: Stephanie Aaronson, Gianni Amisano, Massimo Franchi, Marco Lippi, Filippo Pellegrino, Ivan Petrella, Lucrezia Reichlin, John Roberts, and Esther Ruiz.

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

1 Introduction

This paper is about two stylized facts of macroeconomic time series: co-movements and non-stationarity (Lippi and Reichlin, 1994a). More precisely, this paper is about disentangling long-run co-movements (common trends) from short-run co-movements (common cycles) in a large dataset of non-stationary US macroeconomic indicators.

Since the seminal work of Beveridge and Nelson (1981), the issue of decomposing GDP into a trend and a cycle has been a central question in both time series econometrics and policy analysis. This is not surprising, as long-run trends are mainly influenced by supply-side factors, while short-run cycles are mainly associated with demand-side factors, and therefore different estimates of the trend and of the cycle can lead to different policy recommendations. Given the relevance of the issue, in the last 30 years, many papers have suggested different ways to obtain a Trend-Cycle (TC) decomposition of GDP. Roughly speaking, those works can be grouped under two main approaches: one based on univariate methods (e.g. Watson, 1986; Lippi and Reichlin, 1994b; Morley, Nelson, and Zivot, 2003; Dungey, Jacobs, Tian, and Van Norden, 2015), and another using multivariate, but low-dimensional, time series techniques (e.g. Stock and Watson, 1988; Lippi and Reichlin, 1994a; Gonzalo and Granger, 1995; Garratt, Robertson, and Wright, 2006; Creal, Koopman, and Zivot, 2010).

In this paper we use a novel approach to decompose GDP into a trend and cycle based on large datasets. We first disentangle common and idiosyncratic dynamics by using a Non-Stationary Approximate Dynamic Factor Model (DFM), and then we disentangle common trends from common cycles by applying a non-parametric TC decomposition to the latent common factors. Our methodology builds on four points: first, focusing on a high-dimensional setting is crucial, as only in a high-dimensional setting it is possible to disentangle common from idiosyncratic dynamics in a consistent way (Forni, Hallin, Lippi, and Reichlin, 2000; Bai and Ng, 2002; Stock and Watson, 2002) — i.e., we can separate macroeconomic fluctuations from sectoral dynamics and measurement error only in a high-dimensional setting. Second, assuming the existence of a factor structure is a realistic and convenient way to represent co-movements in large macroeconomic datasets. Third, considering non-stationary data is necessary to account for the presence of common trends or, equivalently, cointegration (Bai, 2004; Bai and Ng, 2004; Barigozzi, Lippi, and Luciani, 2016a,b). And fourth, by using a non-parametric TC decomposition we do not have to make assumptions on the law of motion of either the trend, or the cycle. Our approach is deliberately reduced form, and therefore our empirical analysis is conducted “without pretending to have too much a priori economic theory” (Sargent and Sims, 1977), thus letting the data speak as freely as possible.

The first contribution of this paper is methodological. Namely, we propose a Quasi Maximum Likelihood estimator of the non-stationary DFM based on the Expectation Maximisation (EM) algorithm combined with the Kalman Filter and the Kalman Smoother estimators of the factors. The theoretical properties of this approach in the large stationary DFM case have been studied in Doz, Giannone, and Reichlin (2011, 2012), and here we extend their results to the non-stationary case by proving consistency and by providing rates of convergence for the factors and the parameters of the model. Compared to the non-stationary principal component estimator (Bai and Ng, 2004), the estimator proposed in this paper is more efficient, and it is more flexible in that, thanks to the use of the Kalman Filter, it allows us to explicitly model the idiosyncratic dynamics, and to impose economically meaningful restrictions.

The second contribution of this paper is to show how to isolate common trends and common cycles in large macroeconomic datasets. In detail, we use a non-parametric approach that

identifies the common trends as those linear combinations of the factors obtained by the leading eigenvectors of the long-run covariance matrix (Bai, 2004; Peña and Poncela, 2006), and the common cycles as deviations from the long-run equilibria, which coincide with the space orthogonal to that of the common trends — i.e., the cointegration space (Zhang, Robinson, and Yao, 2016). Because our approach is non-parametric, we are not imposing any particular form to the trend, which is not constrained to be a random walk, or to the cycle. This is what differentiates our approach from the standard state-space, which normally is applied on a handful of variables and where the trend and the cycle dynamics are explicitly specified and jointly estimated with the parameters of the model (Harvey, 1990).

Our final contributions are empirical. Specifically, we employ our methodology to analyse a large panel of US quarterly macroeconomic time series with the goal of estimating the cyclical position of the economy and the observation error. With the expression “estimating the observation error,” we mean estimating aggregate real output. With the expression “estimating the cyclical position of the economy,” we mean decomposing aggregate real output into potential output and output gap. To the best of our knowledge, Fleischman and Roberts (2011) and Aruoba, Diebold, Nalewaik, Schorfheide, and Song (2016) are the only works that, so far, have used (small) factor models to estimate aggregate real output. On the other hand, a few papers have used low-dimensional factor models to estimate the cyclical position of the economy (e.g. Fleischman and Roberts, 2011; Jarociński and Lenza, 2016), and a few more to estimate long-run trends (e.g. Antolin-Diaz, Drechsel, and Petrella, 2016). Finally, Aastveit and Trovik (2014) and Morley and Wong (2017) have used a high-dimensional setting for estimating the output gap by means of a factor model and a large Bayesian VAR, respectively. However, in both works the variables are transformed to stationarity prior to model estimation.

The first part of our empirical analysis is about estimating aggregate real output, to which we refer as Gross Domestic Output (GDO). We first show that our model naturally produces an estimate of GDO as that part of GDP/GDI that is driven by the macroeconomic (common) shocks. We then compare our estimate of GDO with “the average of GDP and GDI” released by the Bureau of Economic Analysis, and with “GDPplus” proposed by Aruoba et al. (2016) and released by the Philadelphia Fed. Our results show that these three measures are very similar, which is not surprising, as they are attempting to estimate the same thing. However, we estimate that since 2010 quarterly annualized GDO growth was on average $\frac{1}{2}$ of a percentage point higher than estimated by the BEA or the Philadelphia Fed, thus pointing out that — based on the commonality in the data — the US economy grew at a faster pace than measured by national account statistics.

The second part of our empirical analysis is about estimating the output gap. To this end, we use the above-mentioned TC decomposition in order to separate long-run from short-run co-movements, and in particular we focus on the decomposition derived for GDO. We compare our estimate with the one produced by the Congressional Budget Office (CBO), which estimates potential output as that level of output consistent with current technologies and normal utilisation of capital and labour, and the output gap as the residual part of output. Although these two estimates are obtained in completely different ways, in practice they look very similar. The two estimates are comparable for most of the sample considered, but from the late nineties to the financial crisis, when our measure suggests that a greater part of the produced output was driven by transitory factors. In particular, according to our estimate between 2001:Q1 and 2005:Q4 the output gap was on average $2\frac{1}{2}$ percentage points higher than estimated by the CBO.

The rest of this paper is structured as follows. In Section 2 we discuss representation of large

non-stationary panels of time series. In this section we first present the non-stationary dynamic factor model, and we define the concepts of commonality — i.e., the common factors. Then we discuss how to disentangle long-run co-movements from short-run co-movements — i.e., we define what common trends and common cycles are. In Section 3 we discuss estimation. We first introduce in Section 3.1 the *static* representation of the DFM, which is just a convenient way to approach estimation of the dynamic model presented in Section 2. We then present in Section 3.2 our estimator, we discuss its properties, and we compare it with existing methods. Finally, in Section 3.3 we present the non-parametric TC decomposition that we use in the empirical section. Then, Section 4 presents the empirical analysis. This section is split in two, with the first part presenting our estimate of GDO (Section 4.1), and the second part presenting our estimate of the output gap (Section 4.2). To conclude, in Section 5 we discuss our findings and the advantages and limitations of our methodology, and we propose directions for further research. In the Appendix we report all technical proofs and the description of the data used and their transformation.

Notation

A vector \mathbf{z}_t is $I(1)$ if the higher-order of integration among all its components is 1, thus under this definition some components of \mathbf{z}_t can be stationary. Eigenvalues are always considered as ordered from the largest to the smallest, so for a given set of eigenvalues $\{\mu_j\}_{j=1}^m$, we have $\mu_1 \geq \mu_2 \geq \dots \geq \mu_{m-1} \geq \mu_m$. Therefore, the spectral norm of \mathbf{A} is defined as $\|\mathbf{A}\|^2 = \mu_1^{A'A}$. The j -th largest eigenvalue of a spectral density matrix at frequency ω is denoted as $\mu_j(\omega)$. The generic (i, j) -th entry of a matrix \mathbf{A} is denoted as $[\mathbf{A}]_{ij}$. We denote by L the lag operator, such that $L^k y_t = y_{t-k}$, for any $k \in \mathbb{Z}$ and we use the notation $\Delta y_t := (1 - L)y_t$. Finally, we let $M, M_0, M_1 \dots$ denote generic positive and finite constants that do not depend on the panel dimensions n or T , and whose value may change from line to line.

2 Representation of non-stationary panels of time series

Let us assume to observe a vector of n time series $\{\mathbf{y}_t = (y_{1t} \dots y_{nt})' : t = 1, \dots, T\}$ such that

$$y_{it} = \mathcal{D}_{it} + x_{it}, \quad (1)$$

where \mathcal{D}_{it} is a deterministic component — e.g., a linear trend — and $\mathbf{x}_t = (x_{1t} \dots x_{nt})'$ is such that $\mathbf{x}_t \sim I(1)$. We also assume that $\mathbb{E}[x_{it}] = 0$, for any i and t , therefore, \mathbf{x}_t contains all the stochastic trends but no deterministic component. Throughout, the spectral density matrix of $\Delta \mathbf{x}_t$ is assumed to exist.

In a high-dimensional setting, it is reasonable to assume that there are common trends and common cycles, but also idiosyncratic terms. Thus, for each variable x_{it} we write

$$x_{it} = \mathcal{T}_{it} + \mathcal{C}_{it} + \xi_{it}, \quad (2)$$

where $\mathcal{T}_{it} \sim I(1)$ is the trend component, $\mathcal{C}_{it} \sim I(0)$ is the cycle component, and ξ_{it} is the idiosyncratic component, which is allowed to be either $I(1)$ (in presence of idiosyncratic trends) or $I(0)$ (e.g. measurement errors). The trend and the cycle are capturing the common dynamics across series, and thus constitute the common component defined as $\chi_{it} = \mathcal{T}_{it} + \mathcal{C}_{it}$. Hence, (2) is also written as

$$x_{it} = \chi_{it} + \xi_{it}. \quad (3)$$

We define the vectors of common and idiosyncratic components as $\boldsymbol{\chi}_t = (\chi_{1t} \cdots \chi_{nt})'$ and $\boldsymbol{\xi}_t = (\xi_{1t} \cdots \xi_{nt})'$, respectively. Finally, notice that consistently with the data considered in this paper: (i) some (but not all) components of \mathbf{x}_t are allowed to be stationary, and (ii) the deterministic components \mathcal{D}_{it} are not common to all series — i.e., there are no common deterministic trends.

We assume that the co-movements in $\boldsymbol{\chi}_t$ are driven by q “structural” shocks, with $q \ll n$, which are collected in a weak white noise vector process $\mathbf{u}_t = (u_{1t} \cdots u_{qt})'$. Then, for a given q , we decompose each element of \mathbf{x}_t as

$$x_{it} = \mathbf{b}'_i(L) \mathbf{f}_t + \xi_{it}, \quad (4)$$

$$\Delta \mathbf{f}_t = \mathbf{C}(L) \mathbf{u}_t, \quad (5)$$

where from (3) the common component is given by $\chi_{it} = \mathbf{b}'_i(L) \mathbf{f}_t$ and the following properties hold:

- A1. $\mathbf{u}_t \stackrel{w.n.}{\sim} (\mathbf{0}_q, \mathbf{I}_q)$, with q is independent of n ;
- A2. $E[u_{jt} \xi_{is}] = 0$, for any $j = 1, \dots, q$, $i = 1, \dots, n$, and $s, t = 1, \dots, T$;
- A3. $\mathbf{B}(L) = (\mathbf{b}'_1(L) \cdots \mathbf{b}'_n(L))'$ is an $n \times q$ one-sided, matrix polynomial matrix of finite order s , $\mathbf{f}_t \sim I(1)$ of dimension q ;
- A4. $\mathbf{C}(L) = (\mathbf{c}'_1(L) \cdots \mathbf{c}'_q(L))'$ is a $q \times q$ one-sided, infinite matrix polynomial with square-summable coefficients and such that $\text{rk}(\mathbf{C}(1)) = (q - d)$ with $0 < d < q$;
- A5. the q -th largest eigenvalue $\mu_q^{\Delta \chi}(\omega)$ of the spectral density matrix of $\Delta \boldsymbol{\chi}_t$ is such that

$$M_1 \leq \liminf_{n \rightarrow \infty} n^{-1} \mu_q^{\Delta \chi}(\omega) \leq \limsup_{n \rightarrow \infty} n^{-1} \mu_q^{\Delta \chi}(\omega) \leq M_2, \quad \omega\text{-a.e.} \in [-\pi, \pi],$$

while the largest eigenvalue $\mu_1^{\Delta \xi}(\omega)$ of the spectral density matrix of $\Delta \boldsymbol{\xi}_t$ is such that

$$M_3 \leq \liminf_{n \rightarrow \infty} \mu_1^{\Delta \xi}(\omega) \leq \limsup_{n \rightarrow \infty} \mu_1^{\Delta \xi}(\omega) \leq M_4, \quad \omega\text{-a.e.} \in [-\pi, \pi].$$

Equations (4) and (5) together with properties A1-A5 define a Non-Stationary Approximate Dynamic Factor Model (DFM). In the case of stationary time series our model is a special case of the Generalised Dynamic Factor Model originally proposed by Forni et al. (2000).

Condition A5 is crucial and it allows for identification of the common component by defining it according to its spectral properties. An explanation for A5 in the time domain is provided by Hallin and Lippi (2013) who show that this condition is equivalent to defining the common and idiosyncratic component by asking that for any dynamic aggregation scheme given by an n -dimensional vector of weights \mathbf{a}_k such that $\sum_{k \in \mathbb{Z}} \mathbf{a}'_k \mathbf{a}_k = 1$, the following holds

$$0 < \lim_{n \rightarrow \infty} \text{Var} \left(\frac{1}{n} \sum_{k=-\infty}^{\infty} \mathbf{a}'_k \Delta \boldsymbol{\chi}_{t-k} \right) \leq M \quad \text{and} \quad \lim_{n \rightarrow \infty} \text{Var} \left(\frac{1}{n} \sum_{k=-\infty}^{\infty} \mathbf{a}'_k \Delta \boldsymbol{\xi}_{t-k} \right) = 0. \quad (6)$$

The following asymptotic conditions for the eigenvalues $\mu_i(\omega)$ of the spectral density of $\Delta \mathbf{x}_t$ are a direct consequence of A4, A5, and Weyl's inequality:

B1. for ω -a.e. $\in [-\pi, \pi]$ the following holds:

$$\begin{aligned} M_1 &\leq \liminf_{n \rightarrow \infty} n^{-1} \mu_q(\omega) \leq \limsup_{n \rightarrow \infty} n^{-1} \mu_q(\omega) \leq M_2, \\ M_3 &\leq \liminf_{n \rightarrow \infty} \mu_{q+1}(\omega) \leq \limsup_{n \rightarrow \infty} \mu_{q+1}(\omega) \leq M_4; \end{aligned}$$

B2. for $\omega = 0$ the following holds:

$$\begin{aligned} M_1 &\leq \liminf_{n \rightarrow \infty} n^{-1} \mu_{q-d}(0) \leq \limsup_{n \rightarrow \infty} n^{-1} \mu_{q-d}(0) \leq M_2, \\ M_3 &\leq \liminf_{n \rightarrow \infty} \mu_{q-d+1}(0) \leq \limsup_{n \rightarrow \infty} \mu_{q-d+1}(0) \leq M_4. \end{aligned}$$

By means of B1 the number of shocks q can then be identified (Hallin and Liška, 2007, Onatski, 2009). Similarly, by means of B2 the number of common trends, $(q - d)$, can be identified (Barigozzi et al., 2016b). In particular, from the intuition given in (6) and because of B1 and B2, it is clear that the DFM is identifiable only in the limit $n \rightarrow \infty$.

Condition A4 allows for the presence of $(q - d)$ common trends in the factors \mathbf{f}_t . In line with our empirical results in Section 4 we rule out the degenerate cases $d = 0$ or $d = q$. This implies that the vector \mathbf{f}_t admits a VECM representation with d cointegration relations (Engle and Granger, 1987), as well as the factor representation (Escribano and Peña, 1994):

$$\mathbf{f}_t = \mathbf{\Psi} \boldsymbol{\tau}_t + \boldsymbol{\gamma}_t, \quad (7)$$

where $\mathbf{\Psi}$ is $q \times (q - d)$ and $\boldsymbol{\tau}_t$ is the vector of $(q - d)$ common trends with components $\tau_{jt} \sim I(1)$ for $j = 1, \dots, (q - d)$, while $\boldsymbol{\gamma}_t$ is a q -dimensional stationary vector.¹ Notice that (7) is different from the common trends representation (or multivariate Beveridge-Nelson decomposition) of Stock and Watson (1988) in that the trend $\boldsymbol{\tau}_t$ is not constrained to be a vector random walk, a property advocated for by many authors (e.g. Lippi and Reichlin, 1994a).

For a given choice of $\mathbf{\Psi}$, the $(q - d)$ common trends can then be obtained by linear projection onto the space spanned by the columns of $\mathbf{\Psi}$:

$$\boldsymbol{\tau}_t = (\mathbf{\Psi}' \mathbf{\Psi})^{-1} \mathbf{\Psi}' \mathbf{f}_t = \mathbf{\Psi}' \mathbf{f}_t.$$

where the second equality holds because, without loss of generality, we can always assume the identifying constraint $\mathbf{\Psi}' \mathbf{\Psi} = \mathbf{I}_{(q-d)}$.

Different choices of $\mathbf{\Psi}$ lead to different definitions of common trends. Here we opt for a non-parametric approach and we identify the elements of $\boldsymbol{\tau}_t$ as the first $(q - d)$ principal components of \mathbf{f}_t , as proposed by Bai (2004) and Peña and Poncela (2006) (see Section 3.3 for details on estimation). Given this definition, the columns of $\mathbf{\Psi}$ are orthonormal and therefore there exists a $q \times d$ matrix $\mathbf{\Psi}_\perp$ such that $\mathbf{\Psi}'_\perp \mathbf{\Psi}_\perp = \mathbf{I}_d$ and $\mathbf{\Psi}'_\perp \mathbf{\Psi} = \mathbf{0}_{d \times (q-d)}$. Now, consider the d -dimensional process obtained by projecting \mathbf{f}_t onto the space orthogonal to the common trends

$$\mathbf{c}_t = (\mathbf{\Psi}'_\perp \mathbf{\Psi}_\perp)^{-1} \mathbf{\Psi}'_\perp \mathbf{f}_t = \mathbf{\Psi}'_\perp \mathbf{f}_t = \mathbf{\Psi}'_\perp \boldsymbol{\gamma}_t.$$

It is straightforward to see that $\mathbf{c}_t \sim I(0)$, that its components are d common cycles in the sense of Vahid and Engle (1993), and that the columns of $\mathbf{\Psi}_\perp$ are a basis of the cointegration space of \mathbf{f}_t , thus these common cycles represent deviations from long-run equilibria — see also e.g. Johansen (1991) and Kasa (1992) for similar definitions.²

¹Notice that in general all factors are non-stationary, unless some *ad hoc* zero-constraint is imposed on the elements of $\mathbf{C}(1)$. On the other hand if we were to ask for one of the factors to be stationary then the corresponding row of $\mathbf{\Psi}$ must be set to zero. However, we do not consider this case further since it could easily be included in our framework by imposing the appropriate identifying assumptions.

²Other TC decompositions based on a different definitions of cycles than the one used here are in Gonzalo and Granger (1995) and Gonzalo and Ng (2001).

According to our definition, common trends and common cycles are orthogonal by construction, and we have the TC decomposition of the factors:

$$\mathbf{f}_t = \boldsymbol{\Psi}\boldsymbol{\Psi}'\mathbf{f}_t + \boldsymbol{\Psi}_\perp\boldsymbol{\Psi}'_\perp\mathbf{f}_t = \boldsymbol{\Psi}\boldsymbol{\tau}_t + \boldsymbol{\Psi}_\perp\mathbf{c}_t, \quad (8)$$

and therefore, by combining (1), (4) and (8), we have the TC decomposition of the data:

$$y_{it} = \mathcal{D}_{it} + \mathbf{b}'_i(L)\boldsymbol{\Psi}\boldsymbol{\tau}_t + \mathbf{b}'_i(L)\boldsymbol{\Psi}_\perp\mathbf{c}_t + \xi_{it} = \mathcal{D}_{it} + \mathcal{T}_{it} + \mathcal{C}_{it} + \xi_{it}. \quad (9)$$

3 Estimation

In order to estimate (9), we need to estimate the factors, \mathbf{f}_t and their TC decomposition. We opt for a two-step approach, where we first extract the common factors and then we estimate their TC decomposition. In particular, we first introduce a convenient re-parametrization of the DFM based on its *static* state-space representation (Section 3.1), which is then used for retrieving the factors space by means of the EM algorithm (Section 3.2). Then, in a second step we use principal component analysis for extracting common trends and cycles (Section 3.3). Notice that compared to the classical state-space approach (e.g. Fleischman and Roberts, 2011) or from the Bayesian approach (e.g. Jarociński and Lenza, 2016) in which the trend and the cycle are estimated in one-step together with the parameters of the models, our approach has the advantage that it does not require us to specify a law of motion for the trend and the cycles.

For simplicity of exposition we assume in this section that there is no deterministic component and we refer to Section 4 and to Appendix D for the treatment of these terms in practice.

3.1 The *static* representation of dynamic factor models

Consider the state-space form of the DFM in (4)-(5) (Stock and Watson, 2005; Forni, Giannone, Lippi, and Reichlin, 2009):

$$x_{it} = \boldsymbol{\lambda}'_i\mathbf{F}_t + \xi_{it}, \quad (10)$$

$$\Delta\mathbf{F}_t = \mathbf{D}(L)\mathbf{u}_t, \quad (11)$$

where from (3) the common component is now given by $\chi_{it} = \boldsymbol{\lambda}'_i\mathbf{F}_t$ and \mathbf{u}_t is the same as in (5). We assume that A1, A2 and A5 still hold and in addition we require:

- C1. $\mathbf{D}(L) = (\mathbf{d}'_1(L) \cdots \mathbf{d}'_r(L))'$ is an $r \times q$ one-sided, infinite matrix polynomial with square-summable coefficients and such that $\text{rk}(\mathbf{D}(1)) = (q - d)$ with $0 < d < q$;
- C2. $\boldsymbol{\Lambda} = (\boldsymbol{\lambda}_1 \cdots \boldsymbol{\lambda}_n)'$ is an $n \times r$ loadings matrix such that $\lim_{n \rightarrow \infty} \|n^{-1}\boldsymbol{\Lambda}'\boldsymbol{\Lambda} - \mathbf{I}_r\| = 0$ and $|[\boldsymbol{\Lambda}]_{ij}| < M$, for any $i = 1, \dots, n$ and $j = 1, \dots, r$;
- C3. $\mathbf{F}_t \sim I(1)$ of dimension r , with $\mathbf{E}[\Delta\mathbf{F}_t\Delta\mathbf{F}'_t]$ positive definite.

Condition C1 is equivalent to A4 in that it requires the existence of $(q - d)$ common trends driving the common component. Conditions C2 and C3 are standard in the literature and imply that the eigenvalues of the covariance of $\Delta\boldsymbol{\chi}_t$ diverge as $n \rightarrow \infty$ at a rate n (Stock and Watson, 2002; Bai and Ng, 2002; Fan, Liao, and Mincheva, 2013). Finally, from A5 we

immediately have that the largest eigenvalue of the covariance of $\Delta \boldsymbol{\xi}_t$ is finite for any n . Given the way \mathbf{F}_t and \mathbf{f}_t are loaded by the data, hereafter we call \mathbf{F}_t *static* factors and \mathbf{f}_t *dynamic* factors.

Let us stress once more the fact that here the DFM and the related TC decomposition are our focus, while the *static* representation is just a convenient way to approach estimation of the dynamic model. In particular, for (10)-(11) to be equivalent to (4)-(5) we need the following restrictions to hold:

- R1. there exists an invertible $r \times r$ matrix \mathbf{K} such that $\mathbf{F}_t = \mathbf{K}(\mathbf{f}'_t \cdots \mathbf{f}'_{t-s})'$ and $\boldsymbol{\lambda}'_i = (\mathbf{b}'_{i0} \cdots \mathbf{b}'_{is})\mathbf{K}^{-1}$, for any $i = 1, \dots, n$, where \mathbf{b}_{ik} , for $k = 0, \dots, s$, are the coefficients of $\mathbf{b}_i(L)$ defined in A3;
- R2. the dimension of \mathbf{F}_t is $r = q(s + 1)$;
- R3. the cointegration rank of \mathbf{F}_t is d .

Let us consider each restriction in detail. Restriction R1 implies that the spectral density of $\Delta \mathbf{F}_t$ has reduced rank q . In the following, we impose this restriction when estimating the model but we do not attempt to identify \mathbf{K} .

Restriction R2 offers an alternative way to determine r with respect to the typical methods available in the literature based on the behavior of the eigenvalues of the covariance matrix of $\Delta \mathbf{x}_t$ and therefore on C2, C3, and A5 (e.g. Bai and Ng, 2002). Specifically, by virtue of restriction R2, once we set q using B1, we can choose r such that the share of variance explained by the *static* factors \mathbf{F}_t coincides with the share of variance explained by the q *dynamic* factors \mathbf{f}_t — see also D'Agostino and Giannone (2012).

Finally, restriction R3 tells us that the autoregressive representation for (11) is a VECM with d cointegration relations (a proof is in Appendix A). Moreover, since the vector \mathbf{F}_t is singular, the autoregressive representation has a finite order (Barigozzi et al., 2016a). However, in the next section we do not estimate a VECM, rather we estimate an unrestricted VAR in the levels (Sims, Stock, and Watson, 1990). We use the knowledge of the cointegration rank to determine the dimension of the common cycles space (see Section 3.3).

Summing up, by not fully imposing R1 and R3 when estimating the factors, we opt for simplicity of estimation versus complexity of a more realistic representation, which implies that the model considered is deliberately mis-specified. The effects of such mis-specification will appear clear in Section 3.3, when we consider TC decompositions of \mathbf{F}_t as opposed to those of \mathbf{f}_t .

3.2 Estimating the space of factors and loadings

We consider the following state-space form of (10)-(11) in which we assume a VAR(2) for the *static* factors as in the empirical analysis of Section 4:

$$x_{it} = \boldsymbol{\lambda}'_i \mathbf{F}_t + \xi_{it}, \quad (12)$$

$$\mathbf{F}_t = \mathbf{A}_1 \mathbf{F}_{t-1} + \mathbf{A}_2 \mathbf{F}_{t-2} + \mathbf{H} \mathbf{u}_t, \quad (13)$$

$$\xi_{it} = \rho_i \xi_{it-1} + e_{it}. \quad (14)$$

We estimate (12)-(14) via the EM algorithm (Dempster, Laird, and Rubin, 1977), combined with the Kalman Filter (KF) and the Kalman Smoother (KS) estimators of the factors (Anderson and Moore, 1979; Harvey, 1990). In the stationary, low-dimensional — i.e., finite n —

setting, estimation of a factor model by means of the EM algorithm can be found in Shumway and Stoffer (1982) and Watson and Engle (1983), while the asymptotic properties of this factors' estimator are studied by Doz et al. (2011, 2012) under the joint limit $n, T \rightarrow \infty$.³ In the non-stationary case, applications of the EM algorithm can be found in Quah and Sargent (1993) and Seong, Ahn, and Zadrozny (2013) in a low-dimensional setting. Here, we study the theoretical properties in the non-stationary case when $n, T \rightarrow \infty$.

In order to run the KF-KS it is necessary to make some additional assumptions on the idiosyncratic component. Let \mathbf{R} be the covariance matrix of the vector $\mathbf{e}_t = (e_{1t} \cdots e_{nt})'$ of the idiosyncratic innovations in (14), then we assume:

- D1. $\rho_i = 1$ if $\xi_{it} \sim I(1)$ or $\rho_i = 0$ if $\xi_{it} \sim I(0)$;
- D2. $\mathbf{e}_t \stackrel{w.n.}{\sim} \mathcal{N}(\mathbf{0}_n, \mathbf{R})$, with $[\mathbf{R}]_{ii} > 0$ and $[\mathbf{R}]_{ij} = 0$ for any $i \neq j$ and $i, j = 1, \dots, n$;
- D3. $\mathbf{u}_t \stackrel{w.n.}{\sim} \mathcal{N}(\mathbf{0}_q, \mathbf{I}_q)$.

It is clear from D1, D2 and (14) that if some idiosyncratic components are $I(1)$, we can still consider a factor model for \mathbf{x}_t with stationary errors in (12) by adding additional latent states with unit loadings and evolving as random walks. Notice that the dimension of the parameter space does not increase by increasing the number of $I(1)$ idiosyncratic components. On the other hand modeling the dynamics of $I(0)$ idiosyncratic components would increase the complexity of the estimation problem. For this reason, in D1 we choose to leave the dynamics of the stationary idiosyncratic components unspecified — see Section 4 for practical implementation of this assumption. Assumptions D1-D3 define a mis-specified approximating model of the true DFM and in this sense our EM approach delivers Quasi Maximum Likelihood (QML) estimators. The effect of these mis-specifications are discussed at the end of this section, but before discussing them we present the asymptotic properties of the estimated factors and loadings.

We collect all unknown parameters of the model into the vector

$$\Theta := (\text{vec}(\mathbf{\Lambda})' \text{vec}(\mathbf{A}_1)' \text{vec}(\mathbf{A}_2)' \text{vec}(\mathbf{H})' \text{diag}(\mathbf{R})')'.$$

We denote by Q the dimension of Θ , then we assume that the true values of the parameters satisfy:

- D4. $\Theta \in \text{int}(\Omega)$, with $\Omega \subseteq \mathbb{R}^Q$ and compact.

This condition is standard in QML theory and ensures existence of the true values of the parameters.

The EM algorithm is based on the iteration of two steps. In the E-step, for a given estimator of the parameters $\hat{\Theta}_k$, we compute the expected likelihood conditional on all observed data $\{\mathbf{x}_1, \dots, \mathbf{x}_T\}$. This is in turn a function of the first and second conditional moments of the *static* factors, which are computed by means of the KS when using $\hat{\Theta}_k$.

Note that, under the assumption of normality, as in D2 and D3, and for a given value of the parameters Θ , the KF-KS give the conditional expectations:

$$\mathbf{F}_{t|t-1} := \mathbf{E}_{\Theta}[\mathbf{F}_t | \mathbf{x}_1, \dots, \mathbf{x}_{t-1}], \quad \mathbf{F}_{t|t} := \mathbf{E}_{\Theta}[\mathbf{F}_t | \mathbf{x}_1, \dots, \mathbf{x}_t], \quad \mathbf{F}_{t|T} := \mathbf{E}_{\Theta}[\mathbf{F}_t | \mathbf{x}_1, \dots, \mathbf{x}_T],$$

³For recent applications of this approach see e.g. Reis and Watson (2010); Bańbura and Modugno (2014); Juvenal and Petrella (2015); Luciani (2015); Coroneo, Giannone, and Modugno (2016).

with the associated covariance matrices denoted as $\mathbf{P}_{t|t-1}$, $\mathbf{P}_{t|t}$, and $\mathbf{P}_{t|T}$, respectively. These are therefore optimal estimators of the *static* factors since they minimize the associated Mean-Square-Error (MSE) for a given value of the parameters.

In the M-step a new estimator of the parameters $\hat{\Theta}_{k+1}$ is computed by maximizing the expected likelihood. At convergence of the EM algorithm, say at iteration k^* , we obtain the estimator of the parameters, which we denote by $\hat{\Theta} := \hat{\Theta}_{k^*}$. The estimator of the factors is then obtained by running the KS a last times using $\hat{\Theta}$ and it is denoted by $\hat{\mathbf{F}}_t := \mathbf{E}_{\hat{\Theta}}[\mathbf{F}_t | \mathbf{x}_1, \dots, \mathbf{x}_T]$. The estimated common and idiosyncratic components are then given by $\hat{\chi}_{it} = \hat{\boldsymbol{\lambda}}_i' \hat{\mathbf{F}}_t$ and $\hat{\xi}_{it} = x_{it} - \hat{\chi}_{it}$. Details of the EM algorithm, as well as closed form expressions for all the estimators, are in Appendix B.

To initialise the EM algorithm we use as initial estimator of the loadings the r leading eigenvectors of the covariance of $\Delta \mathbf{x}_t$, from which we have an estimator of the *static* factors as the integrated principal components of $\Delta \mathbf{x}_t$ (Bai and Ng, 2004). This factors' estimator is in turn used to: (i) initialize the KF, together with a diffuse prior for the factors' covariance (Koopman, 1997; Koopman and Durbin, 2000) and (ii) estimate the VAR parameters (Barigozzi et al., 2016b). Define as \mathbf{V} the $n \times r$ matrix having as columns the r leading normalised eigenvectors of the covariance of $\Delta \mathbf{x}_t$, then the following identifying assumptions are convenient for proving consistency:

$$\text{E1. } \boldsymbol{\Lambda} = \sqrt{n} \mathbf{V} \text{ with } [\boldsymbol{\Lambda}]_{1j} > 0 \text{ for all } j = 1, \dots, r;$$

$$\text{E2. } \mathbf{F}_t = n^{-1/2} \mathbf{V}' \boldsymbol{\chi}_t \text{ with } \mathbf{F}_0 = \mathbf{0}_r.$$

Since the *static* factors have no economic meaning, these identifying assumptions are perfectly valid and — together with assumption C2 on the loadings scale — they rule out any indeterminacy in the estimators used to initialize the EM algorithm — see Doz et al. (2011) for similar assumptions.

We have the following consistency result.

Proposition 1. *Let A1, A2, A5, C1, C2, C3, D1, D2, D3, D4, E1, and E2 hold and let $\bar{t}(T) > 0$ be such that*

$$\limsup_{T \rightarrow \infty} T e^{-\bar{t}(T)} \leq M. \quad (15)$$

Define $\mathbf{F}_t^\dagger := (\mathbf{f}'_t \cdots \mathbf{f}'_{t-s})'$ and $\boldsymbol{\lambda}_i^\dagger := (\mathbf{b}'_{i0} \cdots \mathbf{b}'_{is})'$. Then, there exists an invertible $r \times r$ matrix \mathbf{K} such that, as $n, T \rightarrow \infty$, for all $\bar{t}(T) \leq t \leq T$ and any given $i = 1, \dots, n$,

$$\sqrt{T} \|\hat{\boldsymbol{\lambda}}_i - \mathbf{K}^{-1'} \boldsymbol{\lambda}_i^\dagger\| = O_p(1), \quad (16)$$

$$\min(\sqrt{n}, \sqrt{T}) \|\hat{\mathbf{F}}_t - \mathbf{K} \mathbf{F}_t^\dagger\| = O_p(1), \quad (17)$$

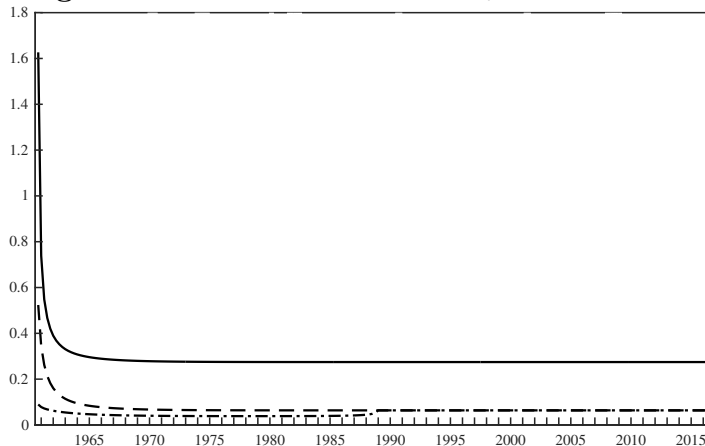
$$\min(\sqrt{n}, \sqrt{T}) |\hat{\chi}_{it} - \chi_{it}| = O_p(1). \quad (18)$$

Proposition 1 states that under the assumptions presented before, we can consistently estimate the common component, as well as the spaces spanned by the *dynamic* factors \mathbf{f}_t and the corresponding *dynamic* loadings which are the coefficients of $\mathbf{b}_i(L)$ defined A3.

Our proof, which is presented in detail in Appendix C, is based on the same approach followed by Poncela and Ruiz (2015) in the one-factor case, and it is made of two main parts which we summarize here.

Population results. We first show that, when the parameters are known the one-step-ahead factors' MSE, $\mathbf{P}_{t|t-1}$, converges to a steady state, while both the MSEs of the KF, $\mathbf{P}_{t|t}$, and

Figure 1: CONDITIONAL MEAN SQUARED ERRORS



This figure reports $\text{tr}(\mathbf{P}_{t|s})$ when using $\hat{\Theta}$, computed for the data analyzed in Section 4, where: $s = t - 1$ is the one-step-ahead conditional MSE (solid line); $s = t$ is KF conditional MSE (dashed line); $s = T$ is the KS conditional MSE (dashed-dotted line).

of the KS, $\mathbf{P}_{t|T}$, tend to zero as $n \rightarrow \infty$ (Lemmas 4 and 5). Notice that this is true also when initializing with a diffuse prior since this has an effect only for a finite number of initial periods, say t_0 (Koopman, 1997). In particular, convergence to the steady state is exponentially fast (Anderson and Moore, 1979), hence our result holds for any $t \geq \bar{t}(T) > t_0$, where $\bar{t}(T)$ satisfies condition (15), which asymptotically requires $\bar{t}(T) = O(\log T)$. In practice, though, the steady state is reached very quickly as shown in Figure 1, where we report the trace of $\mathbf{P}_{t|t-1}$ (solid line), $\mathbf{P}_{t|t}$ (dashed line) and $\mathbf{P}_{t|T}$ (dashed-dotted line), computed for the data analysed in Section 4.

Estimation results. In the second step of the proof, consistency of the KF and KS estimators of the *static* factors when using estimated parameters is proved (Lemma 7). This is done by taking into account an additional parameter estimation error which has two components: (i) the error of the QML estimator of the parameters for the case of known factors, say $\hat{\Theta}^*$ (Lemma 6) and (ii) the error due to the numerical approximation of $\hat{\Theta}$ to $\hat{\Theta}^*$ which is related to the stopping rule of the EM algorithm (Meng and Rubin, 1994, and Lemma 9). In particular, the latter error is shown to be negligible with respect to the former one. Therefore the rate of convergence of the loadings estimated via the EM algorithm is the same that one would obtain by QML estimation, were the true factors observable, and moreover, because of assumption D2 the loadings are estimated equation by equation, thus such error depends only on T . Results similar to (16) hold also for all other estimated parameters in $\hat{\Theta}$. On the other hand the rate of convergence for the estimated *static* factors is standard in the literature.

The results in Proposition 1 extend those by Doz et al. (2011, 2012) to the non-stationary case. A major difference between the EM algorithm in levels proposed in this paper, and the EM algorithm in first differences proposed by Doz et al. (2012), is relative to the way idiosyncratic components are modelled. Indeed, while by considering first differences it is implicitly assumed that all idiosyncratic components have a unit root, in our case we can distinguish between stationary and non-stationary idiosyncratic components — i.e., we can allow for idiosyncratic trends only in some variables. This is not a minor difference, as it has

substantial implications for the properties of the estimators.

First of all we model non-stationary idiosyncratic components as additional latent states rather than differencing them, thus improving efficiency (see also Remark 2 below). Second, when $\xi_{it} \sim I(0)$, under D1 and D2 the QML estimator of the loadings of the i -th variable is obtained by minimizing the sample variance of ξ_{it} . In this case this is not the same as differencing before estimation, since in that case the loadings would be estimated by minimizing the sample variance of $\Delta\xi_{it}$. The resulting common component of the i -th variable has therefore different empirical properties: compared to our non-stationary approach, the common component estimated in first differences is likely to provide a better fit of the first differenced data, but not necessarily of the levels. Conversely, the common component obtained with our approach is likely to provide a better fit of the levels thus capturing better the lower frequencies — and so the long-run trends — and resulting in a smoother estimator, which however might have a worse fit of the differenced data.

We conclude this section by briefly discussing the possible mis-specifications introduced by assumptions D1, D2 and D3. In particular, we assume the vector of idiosyncratic shocks \mathbf{e}_t to be i.i.d. Gaussian, thus imposing four restrictions on: (1) the cross-sectional dependence; (2) the variances; (3) the serial dependence; (4) the distribution. Let us consider the implications for the properties of the estimators of each of these restrictions — see also Doz et al. (2011) for a similar discussion.

Remark 1. If the idiosyncratic components have some cross-sectional dependence, as allowed by A5, then the state-space form of the model is mis-specified, however by inspecting the proofs we see that, as long as we use an invertible estimator of \mathbf{R} , consistency is not affected as long as $n \rightarrow \infty$. As a consequence of this asymptotic argument, we do not attempt here to model the off-diagonal terms of \mathbf{R} .

This is better illustrated by a simple example showing the properties of the KF (an analogous argument holds for the KS). Denote as \mathbf{P} the steady state of $\mathbf{P}_{t|t-1}$ then it can be shown that $\mathbf{P} = \mathbf{H}\mathbf{H}'$ (Lemma 4). Consider the case in which the parameters are given, $\xi_t \sim I(0)$, and $r = q$, so that \mathbf{P} is invertible, then for $t \geq \bar{t}(T)$ the KF estimator is such that

$$\begin{aligned} \mathbf{F}_{t|t} &= \mathbf{F}_{t|t-1} + \mathbf{P}\boldsymbol{\Lambda}'(\boldsymbol{\Lambda}\mathbf{P}\boldsymbol{\Lambda}' + \mathbf{R})^{-1}(\mathbf{x}_t - \boldsymbol{\Lambda}\mathbf{F}_{t|t-1}) \\ &= \mathbf{F}_{t|t-1} + (\boldsymbol{\Lambda}'\mathbf{R}^{-1}\boldsymbol{\Lambda} + \mathbf{P}^{-1})^{-1}\boldsymbol{\Lambda}'\mathbf{R}^{-1}(\mathbf{x}_t - \boldsymbol{\Lambda}\mathbf{F}_{t|t-1}) \\ &= (\boldsymbol{\Lambda}'\mathbf{R}^{-1}\boldsymbol{\Lambda})^{-1}\boldsymbol{\Lambda}'\mathbf{R}^{-1}\mathbf{x}_t + O(n^{-1}) \\ &= \mathbf{F}_t + (\boldsymbol{\Lambda}'\mathbf{R}^{-1}\boldsymbol{\Lambda})^{-1}\boldsymbol{\Lambda}'\mathbf{R}^{-1}\xi_t + O(n^{-1}) \\ &= \mathbf{F}_t + O_p(n^{-1/2}), \end{aligned}$$

where we used (in order) the Woodbury formula, assumption C2, the definition of \mathbf{x}_t in (12), and assumption A5. Clearly consistency of the KF does not depend on the specific assumption for \mathbf{R} , as long as it is invertible. However for finite n the KF depends on \mathbf{R} and modeling also its out of diagonal terms could in principle improve its efficiency (e.g. Bai and Liao, 2016).

Remark 2. From the example in Remark 1 it is clear that for finite n the KF estimator is a weighted average of the data where the heteroskedasticity of the idiosyncratic components is accounted for. Again the same argument holds also for the KS. In this respect the KF-KS approach is analogous to the generalized principal component estimator, which is however derived in a stationary setting and without explicitly addressing the dynamics of the data (Choi, 2012).

Remark 3. If the idiosyncratic components are autocorrelated, then, unless we model them explicitly as additional latent states, optimality is lost, in particular the loadings' estimators are still consistent but not efficient. By means of D1 we partially solve the problem at least for the series with $I(1)$ idiosyncratic components.

Remark 4. If the idiosyncratic components are non-Gaussian then the estimator is not optimal being only the best linear estimator. Nevertheless, it has to be noticed that typical macroeconomic data show little deviations from normality, so we are minimally concerned by the restrictions imposed by this assumption.

Summing up, regardless of these mis-specifications even though we might not have the most efficient estimator, we are likely to have gains in efficiency with respect to those estimators obtained by integrating the principal components of first differences of the data (Bai and Ng, 2004). Indeed, principal components are optimal only in the case of serially and cross-sectionally i.i.d. Gaussian idiosyncratic components (Lawley and Maxwell, 1971; Tipping and Bishop, 1999), and such conditions clearly do not hold in a time series context, especially when non-stationarities are present and the cross-sectional dimension is large. On the contrary, our approach explicitly takes into account the autocorrelation in the factors and in the idiosyncratic components as well as their heteroscedasticity, and, as discussed above, it delivers consistent estimates even when some degree of cross-sectional dependence is present but not modelled.

3.3 Trend and cycles

We now turn to estimation of common trends and common cycles. Notice that since we do not fully impose R1, the *dynamic* factors \mathbf{f}_t are not identified and instead we have to deal with a TC decomposition of the *static* factors \mathbf{F}_t , which can be carried out analogously to the one described in Section 2 for \mathbf{f}_t . Because of assumption C1 and restriction R3, for given values of q and d , the vector \mathbf{F}_t admits the factor representation:

$$\mathbf{F}_t = \Phi_1 \mathbf{T}_t + \Gamma_t,$$

where $\Gamma_t \sim I(0)$, Φ_1 is $r \times (q - d)$ and \mathbf{T}_t is the vector of $(q - d)$ common trends with components $T_{jt} \sim I(1)$ for $j = 1, \dots, (q - d)$. Hence, in general the common trends admit the MA representation:

$$\Delta \mathbf{T}_t = \mathcal{B}(L) \boldsymbol{\eta}_t,$$

where $\boldsymbol{\eta}_t \stackrel{w.n.}{\sim} (\mathbf{0}_{q-d}, \Sigma_\eta)$ with Σ_η positive definite and $\mathcal{B}(L)$ is a $(q - d) \times (q - d)$ one-sided, infinite matrix polynomial with square-summable coefficients and $\text{rk}(\mathcal{B}(1)) = (q - d)$.

As a consequence of the results by Peña and Poncela (1997) and Proposition 1 above, given the estimated factors $\hat{\mathbf{F}}_t$, it is clear that, as $n, T \rightarrow \infty$,

$$\hat{\mathbf{S}} := \frac{1}{T^2} \sum_{t=1}^T \hat{\mathbf{F}}_t \hat{\mathbf{F}}_t' \Rightarrow \Phi_1 \mathcal{B}(1) \Sigma_\eta^{1/2} \left(\int_0^1 \boldsymbol{\mathcal{W}}(u) \boldsymbol{\mathcal{W}}(u)' du \right) \Sigma_\eta^{1/2} \mathcal{B}(1)' \Phi_1', \quad (19)$$

where convergence is in the sense of weak convergence of the associated probability measures and $\{\boldsymbol{\mathcal{W}}(u), 0 \leq u \leq 1\}$ is a $(q - d)$ -dimensional standard Wiener process. Hence, by virtue of (19), we can estimate the common trends \mathbf{T}_t as the first $(q - d)$ principal components of the estimated *static* factors $\hat{\mathbf{F}}_t$ (Bai, 2004; Peña and Poncela, 2006). Specifically, we denote

by $(\widehat{\Phi}_1 \widehat{\Phi}_0)$ the $r \times r$ matrix with columns given by the normalized eigenvectors of $\widehat{\mathbf{S}}$, ordered according to the decreasing value of the corresponding eigenvalues, and such that $\widehat{\Phi}_1$ is $r \times (q-d)$ and $\widehat{\Phi}_0$ is $r \times (r-q+d)$. This leads to the estimator of common trends as the projection:

$$\widehat{\mathbf{T}}_t = \widehat{\Phi}_1' \widehat{\mathbf{F}}_t.$$

As for the common cycles, notice first that, by projecting $\widehat{\mathbf{F}}_t$ onto the columns of $\widehat{\Phi}_0$, we obtain the $(r-q+d)$ -dimensional process

$$\widehat{\mathbf{G}}_t = \widehat{\Phi}_0' \widehat{\mathbf{F}}_t,$$

which, by construction, is orthogonal to $\widehat{\mathbf{T}}_t$. Moreover, $\widehat{\mathbf{G}}_t$ is stationary since it belongs to the cointegration space of $\widehat{\mathbf{F}}_t$ (Zhang et al., 2016). However, by R3 we know that the cointegration space must have dimension d , but we do not impose R3 when estimating the *static* factors. Thus, we face the problem of identifying d cycles from the higher-dimensional stationary process $\widehat{\mathbf{G}}_t$.

In order to identify the common cycles we then look for the d -dimensional projection of $\widehat{\mathbf{G}}_t$ with maximum spectral density. In the empirical analysis of Section 4, we consider the VAR(2):

$$\widehat{\mathbf{G}}_t = \mathcal{A}_1 \widehat{\mathbf{G}}_{t-1} + \mathcal{A}_2 \widehat{\mathbf{G}}_{t-2} + \mathbf{v}_t, \quad (20)$$

where $\mathbf{v}_t \stackrel{w.n.}{\sim} (\mathbf{0}_{r-q+d}, \Sigma_v)$ and $\det(\mathbf{I}_{r-q+d} - \mathcal{A}_1 z - \mathcal{A}_2 z^2) \neq 0$ for $|z| \leq 1$. Once we estimate (20) we have its residuals $\widehat{\mathbf{v}}_t$ and their covariance matrix $\widehat{\Sigma}_v$. Denote as $\widehat{\mathcal{H}}$ the $(r-q+d) \times d$ matrix having as columns the leading d normalized eigenvectors of $\widehat{\Sigma}_v$. We then define the estimated cycle component as the d -dimensional projection:

$$\widehat{\mathbf{C}}_t = \widehat{\mathcal{H}}' \widehat{\mathbf{G}}_t.$$

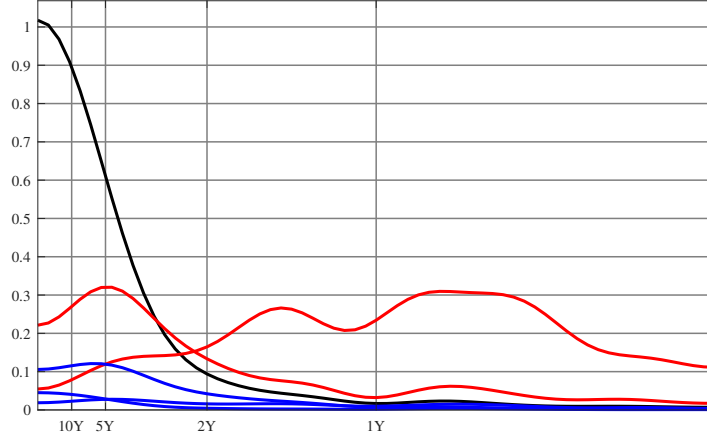
The estimated TC decomposition is then given by

$$\begin{aligned} \widehat{\mathbf{F}}_t &= \widehat{\Phi}_1 \widehat{\Phi}_1' \widehat{\mathbf{F}}_t + \widehat{\Phi}_0 \widehat{\Phi}_0' \widehat{\mathbf{F}}_t \\ &= \widehat{\Phi}_1 \widehat{\mathbf{T}}_t + \widehat{\Phi}_0 \widehat{\mathbf{G}}_t \\ &= \widehat{\Phi}_1 \widehat{\mathbf{T}}_t + \widehat{\Phi}_0 \widehat{\mathcal{H}} \widehat{\mathcal{H}}' \widehat{\mathbf{G}}_t + \widehat{\Phi}_0 \widehat{\mathcal{H}}_{\perp} \widehat{\mathcal{H}}_{\perp}' \widehat{\mathbf{G}}_t \\ &= \widehat{\Phi}_1 \widehat{\mathbf{T}}_t + \widehat{\Phi}_0 \widehat{\mathcal{H}} \widehat{\mathbf{C}}_t + \widehat{\Phi}_0 (\widehat{\mathbf{G}}_t - \widehat{\mathcal{H}} \widehat{\mathbf{C}}_t), \end{aligned} \quad (21)$$

where $\widehat{\mathcal{H}}_{\perp}$ is $(r-q+d) \times (r-q)$ and such that $\widehat{\mathcal{H}}_{\perp}' \widehat{\mathcal{H}} = \mathbf{0}_{(r-q) \times d}$. The last term on the right-hand-side of (21) appears due to the mis-specification caused by not fully imposing R1 and R3 and in particular it has covariance of rank $(r-q)$ and since $r > q$ it is in general not zero.

To appreciate the meaning and the appropriateness of decomposition (21), in Figure 2 we show the spectral densities of the first differences of the three components of $\widehat{\mathbf{F}}_t$ for the data analyzed in Section 4, where $r = 6$, $q = 3$, and $d = 2$. As expected the estimated common trend $\widehat{\mathbf{T}}_t$ (black line) contributes most at the lowest frequencies — i.e., lower than $\frac{\pi}{10}$ — which correspond to periods higher than five years. Once we remove the common trend, of the remaining five processes $\widehat{\mathbf{G}}_t$, the two estimated common cycles $\widehat{\mathbf{C}}_t$ (red lines) capture most of the variation for almost all frequencies: one cycle dominates at periods longer than two years — i.e., frequencies lower than $\frac{\pi}{4}$ — and the other cycle dominates at periods shorter than two

Figure 2: SPECTRAL DENSITIES OF COMMON TRENDS AND COMMON CYCLES



This figure reports for the data analyzed in Section 4 the spectral densities of the common trend $\Delta\widehat{\mathbf{T}}_t$ (black line), the common cycles $\Delta\widehat{\mathbf{C}}_t$ (red lines), and the residual cycles $(\Delta\widehat{\mathbf{G}}_t - \widehat{\mathbf{H}}\Delta\widehat{\mathbf{C}}_t)$ (blue lines). On the horizontal axis we report periods τ_j measured in years and corresponding to frequencies $\omega_j = \frac{2\pi}{4\tau_j}$ (the data considered is quarterly).

years — i.e., frequencies higher than $\frac{\pi}{4}$. With respect to those two cycles, the residual three cycles $(\widehat{\mathbf{G}}_t - \widehat{\mathbf{H}}\widehat{\mathbf{C}}_t)$ (blue lines) give a negligible contributions to the total variation. Given this empirical result, the extra term in (21) can be neglected and treated as a mis-specification error.

Finally, from (21), the estimated TC decomposition of the data immediately follows:

$$x_{it} = \widehat{\lambda}'_i \widehat{\Phi}_1 \widehat{\mathbf{T}}_t + \widehat{\lambda}'_i \widehat{\Phi}_0 \widehat{\mathbf{H}} \widehat{\mathbf{C}}_t + \widehat{\lambda}'_i \widehat{\Phi}_0 (\widehat{\mathbf{G}}_t - \widehat{\mathbf{H}} \widehat{\mathbf{C}}_t) + \widehat{\xi}_{it},$$

which is the estimated counterpart of the representation given in (9).

4 Estimating the cyclical position of the economy and the observation error

We now use our model to estimate the cyclical position of the US economy and the observation error. In particular, in Section 4.1 we will estimate “the observation error” by estimating the non-stationary approximate DFM as explained in Section 3.2. And, in Section 4.2 we will estimate “the cyclical position of the economy” by decomposing the common factors into common trends and common cycles using the TC decomposition discussed in Sections 2 and 3.3.

The following analysis is carried out on a large macroeconomic dataset comprising $n = 103$ quarterly series from 1960:Q1 to 2017:Q1 describing the US economy. The complete list of variables and transformations is reported in Appendix D.

Compared to the papers that use small DFMs to estimate the cyclical position of the economy, which typically estimate the output gap using only high level variables such as GDP, the unemployment rate, and PCE price inflation, we include several other indicators, thus being able to capture information coming from a wider spectrum of the economy. Specifically, our datasets includes national account statistics, industrial production indexes, various

Table 1: PERCENTAGE OF EXPLAINED VARIANCE

	1	2	3	4	5	6	7	8	9	10
q	33.4	45.8	53.3	58.9	63.6	67.4	70.6	73.4	75.8	77.9
r	23.4	33.9	42.1	47.9	51.8	55.3	58.2	60.6	62.7	64.9

This table reports the percentage of total variance explained by the q largest eigenvalues of the spectral density matrix of $\Delta \mathbf{x}_t$ and by the r largest eigenvalues of the covariance matrix of $\Delta \mathbf{x}_t$.

price indexes including CPIs, PPIs, and PCE price indexes, various labor market indicators including indicators from both the household survey and the establishment survey as well as labor cost and compensation indexes, monetary aggregates, credit and loans indicators, housing market indicators, interest rates, the oil price, and the S&P500 index. Broadly speaking, all the variables that are $I(1)$ are not transformed, while all the variables that are $I(2)$ are differenced once. Notice that some variables should from a theoretical economic point of view always be considered as $I(0)$ (e.g. inflation rates, unemployment rate, and interest rates) but since they exhibit a great deal of persistence are here treated as $I(1)$. Finally, a linear trend is estimated where necessary before applying our methodology, thus accounting for the deterministic component in (1).

A thorough empirical analysis requires tackling two main preliminary problems. First, we need to determine the number of common trends ($q - d$), of common shocks q , and of *static* factors r . To determine the number of common trends ($q - d$) we use the criterion by Barigozzi et al. (2016b), which exploits the behaviour of the eigenvalues described in condition B2. This criterion indicates the presence of $(q - d) = 1$ common trend, which is in line with many theoretical models assuming a common productivity trend as the sole driver of long-run dynamics (e.g. Del Negro, Schorfheide, Smets, and Wouters, 2007). To determine the number of common shocks q we use the test by Onatski (2009) and the criterion by Hallin and Liška (2007), which exploit the behaviour of the eigenvalues described in condition B1. Both methods indicate the presence of $q = 3$ common shocks. Having determined q , as we explained in Section 3.1 by virtue of R2 we can set the number of *static* factors r according to their explained variance. By looking at Table 1 we can clearly see that $r \simeq 2q$, and therefore in our benchmark specification we set $q = 3$ and $r = 6$.⁴

Second, we need to choose which idiosyncratic components to model as random walk, and which as white noises. Following the methodology proposed by Bai and Ng (2004), we can explicitly test the null-hypothesis $H_0: \rho_i = 1$, and if we do not reject H_0 , we set $\rho_i = 1$, while if we reject H_0 , we set $\rho_i = 0$. This approach is applied to all variables in the dataset except GDP, GDI, unemployment rate, Federal funds rate, and CPI, core CPI, PCE, and core PCE inflation, for which we impose a priori $\rho_i = 0$. That is, while for most of the variables in the dataset we let the data determine what is driving their long run dynamics, we impose that the long-run dynamics of GDP, GDI, unemployment rate, Federal funds rate, and CPI, core CPI, PCE, and core PCE inflation are driven exclusively by macroeconomic shocks, with the idiosyncratic shocks accounting only for short-run movement.

⁴An alternative way to select the number of *static* factors r is to resort to one of the many available methods, such as, for example, the criterion of Bai and Ng (2002), which for our dataset gives results in line with our choice of r .

4.1 Measuring Gross Domestic Output

A fundamental issue in economics is the measurement of aggregate real output, henceforth Gross Domestic Output (GDO). Historically, GDO has been measured mainly by the Gross Domestic Product (GDP), but GDP, which tracks all expenditures on final goods and services produced, is just an estimate of GDO. An equally acceptable estimate of the concept of GDO is represented by the Gross Domestic Income (GDI), which tracks all income received by those who produced the output. GDP is almost always preferred to GDI, the main reason being that it is released before GDI.⁵ However it has been shown that GDI reflects the business cycle fluctuations in true output growth better than GDP and moreover GDI is better than GDP in recognising the start of a recession (Nalewaik, 2010, 2012).

In recent years, there has been interest in combining GDP and GDI to come up with a better estimate of GDO, where the rationale for doing so is that the difference between GDP and GDI is exclusively the result of measurement error — using the NIPA table definition “statistical discrepancy” — as these two statistics are in fact measuring the same thing. For example, starting from November 4, 2013, the Philadelphia Fed releases an estimate of GDO, called “GDPplus” proposed by Aruoba et al. (2016), which is defined as the common component of a bivariate one-factor model built with GDP and GDI growth rates. Similarly, and starting from July 30, 2015, the Bureau of Economic Analysis (BEA) releases “the average of GDP and GDI”, which the Council of Economic Advisers refers to as GDO (Council of Economic Advisers, 2015).

Our approach differs from those mentioned above in that our estimate of GDO is not obtained by combining GDP and GDI, rather it is obtained by using all the 103 variables included in our dataset. In detail, we define GDO as that part of GDP/GDI that is driven by the macroeconomic (common) shocks, i.e., $GDO_t = \chi_t^{\text{GDP}} = \chi_t^{\text{GDI}}$. To estimate GDO in this way, we estimate a constrained version of model (12)-(13), where we impose the restriction of equal common components: $\chi_t^{\text{GDP}} = \chi_t^{\text{GDI}}$. This restriction is indeed corroborated by the data, as even if we do not impose it, the estimated χ_t^{GDP} and χ_t^{GDI} are nearly identical. In numbers, the standard deviation of $(\Delta y_t^{\text{GDP}} - \Delta y_t^{\text{GDI}})$ is 1.93, while the standard deviation of $(\Delta \chi_t^{\text{GDP}} - \Delta \chi_t^{\text{GDI}})$ is reduced to 0.28.

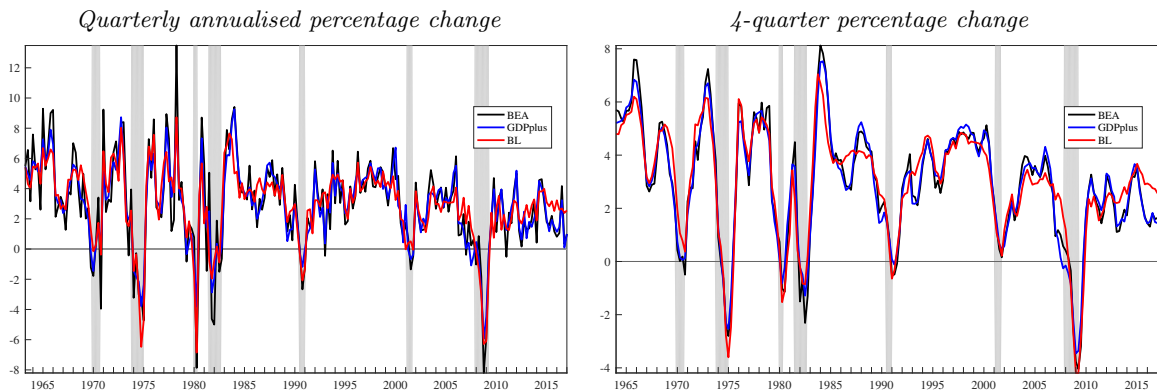
Figure 3 shows our proposed estimate of GDO (red line) together with “GDPplus” (blue line) and the “the average of GDP and GDI” released by the BEA (black line). Overall, the three measures are very similar, which is not surprising, as they are attempting to estimate the same quantity. However, three important differences emerges.

First, our estimate of GDO is smoother than the other two. This is not surprising. Compared to “GDPplus” and “the average of GDP and GDI”, our estimate of GDO is constructed to contain a larger low frequency component, because it is estimated on data in levels rather than on growth rates. Moreover, because it is derived under the assumption that the idiosyncratic components of GDP and GDI are stationary, by construction our estimate of GDO captures all the low frequency movements of GDP and GDI.

Second, our estimate of GDO does not show any kind of residual seasonality in the last fifteen years, where the term “residual seasonality” refers to the presence of “lingering seasonal effects even after seasonal adjustment processes have been applied to the data” (Moulton and Cowan, 2016). Mainly motivated by the fact that since 2010 GDP growth in Q1 has been on average more than 1 percentage point lower than in the other quarters (NW plot of Figure 4), in

⁵The first estimate of GDP is released one month after the reference quarter, while GDI is generally released two months after the reference quarter, together with the second release of GDP.

Figure 3: GROSS DOMESTIC OUTPUT



This figure reports different estimates of GDO. Black line: “the average of GDP and GDI” released by the BEA; blue line: “GDPplus” released by the Philadelphia Fed; red line: our estimate.

recent years there has been lots of discussion on whether US GDP exhibit residual seasonality or not. The profession is not in agreement on this issue, as some authors (e.g. Gilbert et al., 2015; Lengermann et al., 2017) conclude that US GDP does not exhibit residual seasonality, while others (e.g. Rudebusch, Wilson, and Mahedy, 2015; Lunsford, 2017) find evidence of residual seasonality — see Moulton and Cowan (2016) for a technical discussion on causes and remedies for residual seasonality in US GDP. Figure 4 shows average real GDO growth by quarter for our estimate of GDO (SE plot), “GDPplus” (SW plot), and “the average of GDP and GDI” (NE plot). As can be clearly seen, our estimate of GDO exhibits no residual seasonality whatsoever in the last 15 years.

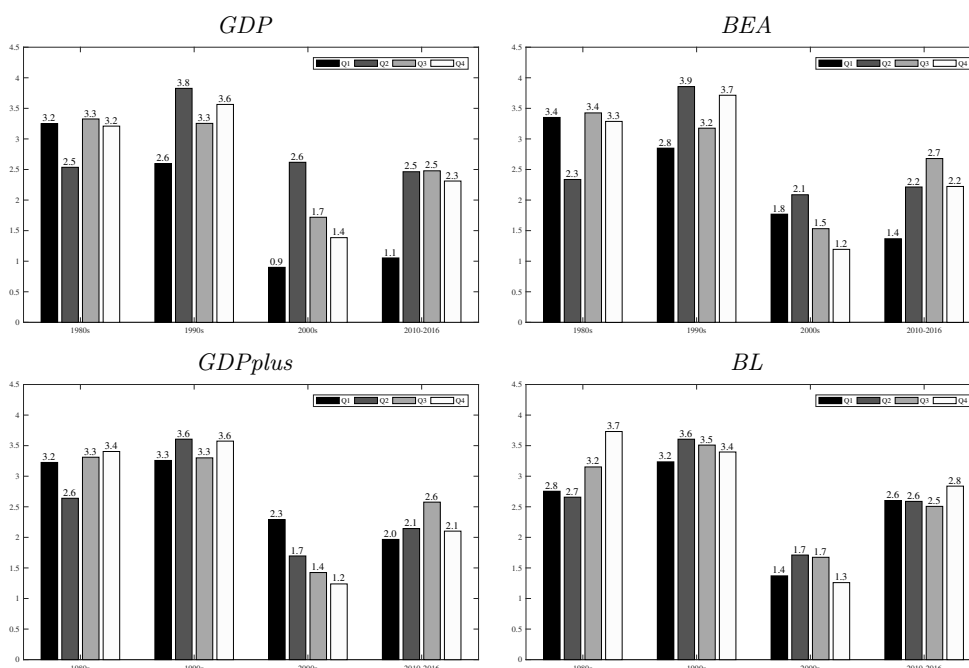
Third, our estimate of GDO in the recent years gives a different signal about the economy than the one given by ‘GDPplus’ and “the average of GDP and GDI”. According to our estimate, since 2010 quarterly annualized GDO growth was on average $\frac{1}{2}$ of a percentage point higher than estimated by the BEA or the Philadelphia Fed, where this difference comes mainly from our estimate of GDO growth in the first quarter (see Figure 4), and therefore from the fact that our measure do not suffer of residual seasonality. In other words, based on the commonality in the data, the US economy grew at a faster pace than measured by national account statistics.

4.2 Measuring the output gap

Decomposing aggregate real output into potential output and output gap is a critical task for both monetary and fiscal policy, as the former is a key input for long-term projections, and the latter can be an important gauge of inflationary pressure. There exist many definitions of potential output and of output gap — see Kiley, 2013, for a survey of different methods and definitions. Here we use the definition implied by the TC decomposition discussed in Sections 2 and 3.3. Among the many existing approaches the most similar to ours are Fleischman and Roberts (2011) and Jarociński and Lenza (2016), who use small dynamic factor models, Aastveit and Trovik (2014), who use a large stationary dynamic factor model combined with the Hodrick Prescott filter, and Morley and Wong (2017), who use a large stationary BVAR combined with the Beveridge and Nelson decomposition.

We compare our output gap estimate with the one produced by the Congressional Budget

Figure 4: RESIDUAL SEASONALITY

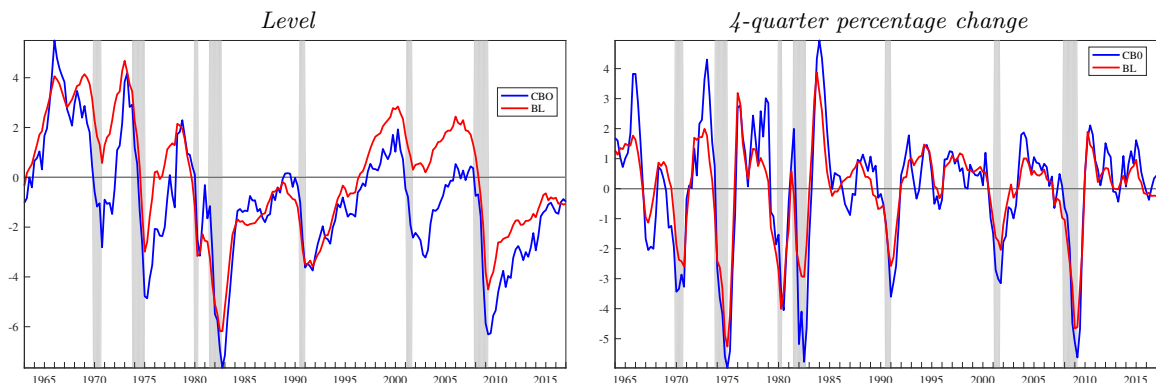


This figure reports average growth at an annual rate by quarter for GDP, “the average of GDP and GDI” released by the BEA, the Philadelphia Fed estimate of GDO (GDPplus), and our estimate of GDO (BL).

Office (CBO). The CBO estimates potential output and the output gap by using the so-called “production function approach” according to which potential output is that level of output consistent with current technologies and normal utilisation of capital and labour, and the output gap is the residual part of output. Specifically, the CBO model is based upon a textbook Solow growth model, with a neoclassical production function. Labour and productivity trends are estimated by using a variant of the Okun’s law, so that actual output is above its potential (the output gap is positive), when the unemployment rate is below the natural rate of unemployment, which is in turn defined as the non-accelerating inflation rate of unemployment (NAIRU), i.e., that level of unemployment consistent with a stable inflation — for further details see Congressional Budget Office (2001).

In Figure 5, we compare our measure of the output gap (red line) with the one produced by the CBO (blue line), where the left plot shows the level of the output gap, while the right plot shows the 4-quarter percentage change of the output gap. The main result emerging from Figure 5 is that our estimate of the output gap is remarkably similar to that of the CBO. However, there are a few periods in which the two estimates diverge, among which the main one is from the late nineties to the financial crisis. In particular, while according to the CBO the level of the output gap was negative between 2001:Q1 and 2005:Q4, according to our estimate in that same period the output gap was positive — on average $2\frac{1}{2}$ percentage points higher than estimated by the CBO. Therefore, according to our estimate the level of the output gap right before the great financial crisis in 2007:Q4 was 1.3%, while according to the CBO was -0.7%, and hence we estimate that the level of slack in the economy at the trough of the crisis in 2009:Q2 was -4.5%, approximately $1\frac{3}{4}$ percentage points higher than estimated by the CBO.

Figure 5: OUTPUT GAP



The left plot shows the level of the output gap estimated by the CBO (blue line) together with our estimate (red line). The right plot shows the 4-quarter percentage change of the output gap.

To conclude, let us emphasize that the fact that our estimate of the output gap is close to that of the CBO is a remarkable result, particularly so because our estimate of the output gap is very different from that of the CBO from both a technical and an interpretational point of view. Indeed, while the CBO constructs the output gap so that its level has a specific economic meaning, our measure of the output gap is simply the transitory/stationary part of the common component of output — i.e., that part of aggregate real output that will disappear in the long-run.⁶ Therefore, our output gap estimate provides different and complementary information on the cyclical position of the economy than that contained in the CBO estimate. In particular, our estimate of the output gap seems more suitable to answer the question “which part of current growth is due to temporary factors?”, while the measure of the CBO is certainly more suitable as a gauge of inflation pressure. This can explain in part the divergence of the two estimates in the 2000s. This period is characterized by stable and low inflation — on average core CPI inflation between 2001:Q1 and 2007:Q4 was approximately 2.1%. Accordingly, the CBO estimates that slack is positive (i.e., the output gap negative). By contrast, our measure, which is not specifically affected by inflation, but it is more broadly influenced by the co-movement in the data, estimates that a part of the aggregate real output was transitory. This makes sense given that the years before the crisis were characterized by several factors that proved indeed transitory, such as the housing boom, a historically high share of sub-prime loan origination (Haughwout and Okah, 2009), and a large amount of equity withdrawal from housing (Fuster, Geddes, and Haughwout, 2017). And, since our model includes a large number of variables, including housing indicators as well as loan and credit indicators, these transitory factors are captured by our model.

5 Discussion and conclusions

In this paper we disentangle long-run co-movements (common trends) from short-run co-movements (common cycles) in large datasets. To this end, we first estimate a non-stationary dynamic factor model by means of a Quasi Maximum Likelihood estimator based on the Expectation Maximisation algorithm, combined with the Kalman Filter and the Kalman Smoother

⁶Notice that also for the CBO the output gap is assumed to revert to zero in the long-run as it imposes in its forecast that in 10 years the output gap will be zero — see e.g. Congressional Budget Office (2004).

estimators of the factors. We then disentangle common trends from common cycles by applying a non-parametric Trend-Cycle decomposition to the latent common factors and based on eigenanalysis of their long-run covariance. The asymptotic properties of this estimator are derived and discussed in the paper.

We estimate our model on a large panel of US quarterly macroeconomic time series with the goal of estimating the cyclical position of the economy and the observation error. After backing out the observation error, we show that our model naturally produces an estimate of aggregate real output, which we refer to as Gross Domestic Output (GDO). According to our estimate of GDO, since 2010 the US economy grew at a faster pace than measured by national account statistics.

We then use a Trend-Cycle decomposition to estimate the output gap. We compare our estimate of the output gap, which is entirely data-driven, with that produced by the Congressional Budget Office (CBO), which is instead based on theoretical economic models. It turns out that our estimate of the output gap is remarkably similar to that of the CBO except from the late nineties to the financial crisis, when our measure suggests that a greater part of the produced output was driven by transitory factors.

There are a number of aspects of our model that we have not fully developed in our empirical analysis and that are left for future research. First, due to the use of the Kalman Filter, our factor estimator is in principle able to handle both mixed frequency and missing data (e.g. Mariano and Murasawa, 2003; Jungbacker, Koopman, and Van der Wel, 2011; Bańbura and Modugno, 2014) and, therefore, it can be used for real-time analysis (Giannone, Reichlin, and Small, 2008). This aspect is well-known to be particularly relevant when estimating the output gap, since as shown by Orphanides and van Norden (2002), end-of-sample revisions of GDP are of the same order of magnitude as the gap itself. Second, the use of the Kalman Filter makes our model suitable for scenario and counterfactual analysis based on conditional forecasts (Bańbura, Giannone, and Lenza, 2015). Third, as shown in equation (21), our model naturally produces a Trend-Cycle decomposition for each variable in the dataset, and therefore it is possible to estimate other policy-relevant indicators, such as the unemployment gap (in our framework, the cycle component of the unemployment rate) or trend inflation (in our framework, the trend component of core CPI or the core PCE price indexes).

Our approach has been so far deliberately entirely data driven, and we have been careful in imposing the least possible amount of restrictions to let the data speak freely. This approach has undeniably some important merits, as estimation of GDO seems to fit naturally in our framework, and the Trend-Cycle decomposition that we obtain for GDO is economically sensible. However, we believe that imposing the statistical restrictions described in Section 3.1, thus eliminating the miss-specification error when computing the Trend-Cycle decomposition, as well as imposing economically meaningful constraints, seems to be an essential step forward. Our view is that one way to proceed is to consider Bayesian estimation of the model, so that our economic and statistical knowledge of the data can be included by means of suitable priors. All this is the subject of our current research.

References

- Aastveit, K. A. and T. Trovik (2014). Estimating the output gap in real time: A factor model approach. *The Quarterly Review of Economics and Finance* 54, 180–193.
- Anderson, B. D. O. and J. B. Moore (1979). *Optimal Filtering*. Dover Publications, Inc.

- Antolin-Diaz, J., T. Drechsel, and I. Petrella (2016). Tracking the slowdown in long-run GDP growth. *The Review of Economics and Statistics*. forthcoming.
- Antsaklis, P. J. and A. M. Michel (2007). *A Linear Systems Primer*. Birkhäuser.
- Aruoba, S. B., F. X. Diebold, J. Nalewaik, F. Schorfheide, and D. Song (2016). Improving GDP measurement: A measurement-error perspective. *Journal of Econometrics* 191, 384–397.
- Bai, J. (2004). Estimating cross-section common stochastic trends in nonstationary panel data. *Journal of Econometrics* 122, 137–183.
- Bai, J. and Y. Liao (2016). Efficient estimation of approximate factor models via penalized maximum likelihood. *Journal of Econometrics* 191, 1–18.
- Bai, J. and S. Ng (2002). Determining the number of factors in approximate factor models. *Econometrica* 70, 191–221.
- Bai, J. and S. Ng (2004). A PANIC attack on unit roots and cointegration. *Econometrica* 72, 1127–1177.
- Bañbura, M., D. Giannone, and M. Lenza (2015). Conditional forecasts and scenario analysis with vector autoregressions for large cross-sections. *International Journal of Forecasting* 31, 739–756.
- Bañbura, M. and M. Modugno (2014). Maximum likelihood estimation of factor models on datasets with arbitrary pattern of missing data. *Journal of Applied Econometrics* 29, 133–160.
- Barigozzi, M., M. Lippi, and M. Luciani (2016a). Dynamic factor models, cointegration, and error correction mechanisms. FEDS 2016-18, Board of Governors of the Federal Reserve System.
- Barigozzi, M., M. Lippi, and M. Luciani (2016b). Non-stationary dynamic factor models for large datasets. FEDS 2016-24, Board of Governors of the Federal Reserve System.
- Beveridge, S. and C. R. Nelson (1981). A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the ‘business cycle’. *Journal of Monetary Economics* 7, 151–174.
- Choi, I. (2012). Efficient estimation of factor models. *Econometric Theory* 28, 274–308.
- Congressional Budget Office (2001). CBO’s method for estimating potential output: An update.
- Congressional Budget Office (2004). A summary of alternative methods for estimating potential GDP. CBO Background Paper.
- Coroneo, L., D. Giannone, and M. Modugno (2016). Unspanned macroeconomic factors in the yield curve. *Journal of Business and Economic Statistics* 34, 472–485.
- Council of Economic Advisers (2015). A better measure of economic growth: Gross Domestic Output (GDO). CEA Issue Brief.
- Creal, D., S. J. Koopman, and E. Zivot (2010). Extracting a robust US business cycle using a time-varying multivariate model-based bandpass filter. *Journal of Applied Econometrics* 25, 695–719.
- D’Agostino, A. and D. Giannone (2012). Comparing alternative predictors based on large-panel factor models. *Oxford Bulletin of Economics and Statistics* 74, 306–326.
- Del Negro, M., F. Schorfheide, F. Smets, and R. Wouters (2007). On the fit of New Keynesian models. *Journal of Business and Economic Statistics* 25, 123–143.
- Dempster, A. P., N. M. Laird, and D. B. Rubin (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 1–38.
- Doz, C., D. Giannone, and L. Reichlin (2011). A two-step estimator for large approximate dynamic factor models based on Kalman filtering. *Journal of Econometrics* 164, 188–205.

- Doz, C., D. Giannone, and L. Reichlin (2012). A quasi maximum likelihood approach for large approximate dynamic factor models. *The Review of Economics and Statistics* 94(4), 1014–1024.
- Dungey, M., J. P. Jacobs, J. Tian, and S. Van Norden (2015). Trend in cycle or cycle in trend? New structural identifications for unobserved-components models of US real GDP. *Macroeconomic Dynamics* 19, 776–790.
- Durbin, J. and S. J. Koopman (2001). *Time Series Analysis by State Space Methods*. Oxford University Press.
- Engle, R. F. and C. W. J. Granger (1987). Cointegration and error correction: Representation, estimation, and testing. *Econometrica* 55, 251–76.
- Escribano, A. and D. Peña (1994). Cointegration and common factors. *Journal of Time Series Analysis* 15, 577–586.
- Fan, J., Y. Liao, and M. Mincheva (2013). Large covariance estimation by thresholding principal orthogonal complements. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 75, 603–680.
- Fleischman, C. A. and J. M. Roberts (2011). From many series, one cycle: improved estimates of the business cycle from a multivariate unobserved components model. FEDS 2011-046, Board of Governors of the Federal Reserve System.
- Forni, M., D. Giannone, M. Lippi, and L. Reichlin (2009). Opening the black box: Structural factor models versus structural VARs. *Econometric Theory* 25, 1319–1347.
- Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2000). The Generalized Dynamic Factor Model: Identification and estimation. *The Review of Economics and Statistics* 82, 540–554.
- Franchi, M. (2017). On the structure of state space systems with unit roots. Technical report. mimeo.
- Fuster, A., E. Geddes, and A. Haughwout (2017). Houses as ATMs no longer. Federal Reserve Bank of New York Liberty Street Economics blog, February 15. <http://libertystreeteconomics.newyorkfed.org/2017/02/houses-as-atms-no-longer.html>.
- Garratt, A., D. Robertson, and S. Wright (2006). Permanent vs transitory components and economic fundamentals. *Journal of Applied Econometrics* 21, 521–542.
- Giannone, D., L. Reichlin, and D. Small (2008). Nowcasting: The real-time informational content of macroeconomic data. *Journal of Monetary Economics* 55, 665–676.
- Gilbert, C., N. Morin, A. D. Paciorek, and C. R. Sahm (2015). Residual seasonality in GDP. FEDS Notes 2015-05-14, Board of Governors of the Federal Reserve System.
- Gonzalo, J. and C. Granger (1995). Estimation of common long-memory components in cointegrated systems. *Journal of Business and Economic Statistics* 13, 27–35.
- Gonzalo, J. and S. Ng (2001). A systematic framework for analyzing the dynamic effects of permanent and transitory shocks. *Journal of Economic Dynamics and Control* 25, 1527–1546.
- Hallin, M. and M. Lippi (2013). Factor models in high-dimensional time series? A time-domain approach. *Stochastic Processes and their Applications* 123, 2678–2695.
- Hallin, M. and R. Liška (2007). Determining the number of factors in the general dynamic factor model. *Journal of the American Statistical Association* 102, 603–617.
- Harvey, A. C. (1990). *Forecasting, structural time series models and the Kalman filter*. Cambridge University Press.
- Haughwout, A. F. and E. Okah (2009). Below the line: Estimates of negative equity among nonprime mortgage borrowers. *Economic Policy Review* 15, 31–43. Federal Reserve Bank of New York.

- Jarociński, M. and M. Lenza (2016). An inflation-predicting measure of the output gap in the euro area. ECB Working Paper Series 1966, European Central Bank.
- Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica* 59, 1551–80.
- Jungbacker, B., S. J. Koopman, and M. Van der Wel (2011). Maximum likelihood estimation for dynamic factor models with missing data. *Journal of Economic Dynamics and Control* 35, 1358–1368.
- Juvenal, L. and I. Petrella (2015). Speculation in the oil market. *Journal of Applied Econometrics* 30, 1099–1255.
- Kasa, K. (1992). Common stochastic trends in international stock markets. *Journal of Monetary Economics* 29, 95–124.
- Kiley, M. T. (2013). Output gaps. *Journal of Macroeconomics* 37(C), 1–18.
- Kohn, R. and C. F. Ansley (1983). Fixed interval estimation in state space models when some of the data are missing or aggregated. *Biometrika* 70, 683–688.
- Koopman, S. J. (1997). Exact initial Kalman filtering and smoothing for nonstationary time series models. *Journal of the American Statistical Association* 92, 1630–1638.
- Koopman, S. J. and J. Durbin (2000). Fast filtering and smoothing for multivariate state space models. *Journal of Time Series Analysis* 21, 281–296.
- Lawley, D. N. and A. E. Maxwell (1971). *Factor Analysis as a Statistical Method*. Butterworths, London.
- Lengermann, P., N. Morin, A. D. Paciorek, E. Pinto, and C. R. Sahm (2017). Another look at residual seasonality in GDP. FEDS Notes 2017-07-28, Board of Governors of the Federal Reserve System.
- Lippi, M. and L. Reichlin (1994a). Common and uncommon trends and cycles. *European Economic Review* 38, 624–635.
- Lippi, M. and L. Reichlin (1994b). Diffusion of technical change and the decomposition of output into trend and cycle. *The Review of Economic Studies* 61, 19–30.
- Luciani, M. (2015). Monetary policy and the housing market: A structural factor analysis. *Journal of Applied Econometrics* 30, 199–218.
- Lunsford, K. G. (2017). Lingering residual seasonality in GDP growth. Economic Commentary 2017-06, Federal Reserve Bank of Cleveland.
- Mariano, R. S. and Y. Murasawa (2003). A new coincident index of business cycles based on monthly and quarterly series. *Journal of Applied Econometrics* 18, 427–443.
- Meng, X.-L. and D. B. Rubin (1994). On the global and componentwise rates of convergence of the EM algorithm. *Linear Algebra and its Applications* 199, 413–425.
- Morley, J., C. R. Nelson, and E. Zivot (2003). Why are unobserved component and Beveridge-Nelson trend-cycle decompositions of GDP so different. *The Review of Economics and Statistics* 85, 235–243.
- Morley, J. and B. Wong (2017). Estimating and accounting for the output gap with large Bayesian vector autoregressions. CAMA Working Papers 2017-46, Centre for Applied Macroeconomic Analysis.
- Moulton, B. R. and B. D. Cowan (2016). Residual seasonality in GDP and GDI: Findings and next steps. *Survey of Current Business* 96, 1–6.
- Nalewaik, J. J. (2010). The income- and expenditure-side measures of output growth. *Brookings*

- Papers on Economic Activity* 1, 71–106.
- Nalewaik, J. J. (2012). Estimating probabilities of recession in real time using GDP and GDI. *Journal of Money Credit and Banking* 44, 235–253.
- Onatski, A. (2009). Testing hypotheses about the number of factors in large factor models. *Econometrica* 77, 1447–1479.
- Orphanides, A. and S. van Norden (2002). The unreliability of output-gap estimates in real time. *The Review of Economics and Statistics* 84, 569–583.
- Peña, D. and P. Poncela (1997). Eigenstructure of nonstationary factor models. UC3M Working Papers. Statistics and Econometrics 97-90-29, Universidad Carlos III Madrid.
- Peña, D. and P. Poncela (2006). Nonstationary dynamic factor analysis. *Journal of Statistical Planning and Inference* 136, 1237–1257.
- Poncela, P. and E. Ruiz (2015). More is not always better: Kalman filtering in dynamic factor models. In S. J. Koopman and N. Shephard (Eds.), *Unobserved Components and Time Series Econometrics*. Oxford Scholarship Online.
- Proietti, T. (1997). Short-run dynamics in cointegrated systems. *Oxford Bulletin of Economics and Statistics* 59, 405–422.
- Quah, D. and T. J. Sargent (1993). A dynamic index model for large cross sections. In *Business cycles, indicators and forecasting*. University of Chicago Press.
- Reis, R. and M. W. Watson (2010). Relative goods’ prices, pure inflation, and the Phillips correlation. *American Economic Journal Macroeconomics* 2, 128–157.
- Rudebusch, G. D., D. Wilson, and T. Mahedy (2015). The puzzle of weak first-quarter growth. Economic Letter 2015-16, Federal Reserve Bank of San Francisco.
- Sargent, T. J. and C. A. Sims (1977). Business cycle modeling without pretending to have too much a priori economic theory. In *New methods in business cycle research*. Federal Reserve Bank of Minneapolis.
- Seong, B., S. K. Ahn, and P. A. Zdrozny (2013). Estimation of vector error correction models with mixed-frequency data. *Journal of Time Series Analysis* 34, 194–205.
- Shumway, R. H. and D. S. Stoffer (1982). An approach to time series smoothing and forecasting using the EM algorithm. *Journal of Time Series Analysis* 3, 253–264.
- Sims, C., J. H. Stock, and M. W. Watson (1990). Inference in linear time series models with some unit roots. *Econometrica* 58, 113–144.
- Stock, J. H. and M. W. Watson (1988). Testing for common trends. *Journal of the American Statistical Association* 83, 1097–1107.
- Stock, J. H. and M. W. Watson (2002). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association* 97, 1167–1179.
- Stock, J. H. and M. W. Watson (2005). Implications of dynamic factor models for VAR analysis. Working Paper 11467, NBER.
- Tipping, M. E. and C. M. Bishop (1999). Probabilistic principal component analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 61, 611–622.
- Vahid, F. and R. F. Engle (1993). Common trends and common cycles. *Journal of Applied Econometrics* 8, 341–360.
- Watson, M. W. (1986). Univariate detrending methods with stochastic trends. *Journal of Monetary Economics* 18, 49–75.

- Watson, M. W. and R. F. Engle (1983). Alternative algorithms for the estimation of dynamic factor, mimic and varying coefficients regression models. *Journal of Econometrics* 23, 385–400.
- Wu, J. C. F. (1983). On the convergence properties of the EM algorithm. *The Annals of Statistics* 11, 95–103.
- Zhang, R., P. Robinson, and Q. Yao (2016). Identifying cointegration by eigenanalysis. <https://arxiv.org/abs/1505.00821>.

Appendix A Representation results

Hereafter, and throughout all appendices, we consider restriction R2 when $s = 1$ as found empirically in Section 4. Therefore, $r = 2q$.

A.1 Proof of restriction R3

For the dynamic factors consider the VECM(2)

$$\Delta \mathbf{f}_t = -\mathbf{a}\mathbf{b}'\mathbf{f}_{t-3} + \Gamma_1\Delta \mathbf{f}_{t-1} + \Gamma_2\Delta \mathbf{f}_{t-2} + \mathbf{u}_t, \quad (\text{A1})$$

where \mathbf{a} and \mathbf{b} are $q \times d$ and for simplicity we consider just the case of two lags since this will imply a VECM(1) and therefore a VAR(2) for the static factors as implemented in (13).

First assume that in R1 we have $\mathbf{K} = \mathbf{I}_r$. Our aim is then to find the correct VECM representation for $\mathbf{F}_t = (\mathbf{f}'_t \mathbf{f}'_{t-1})'$ when the VECM in (A1), and restrictions R1 and R2 hold. Since we model \mathbf{F}_t as a VAR(2) we know that we must have a VECM(1) with reduced rank innovations by R1, hence

$$\Delta \mathbf{F}_t = -\boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{F}_t + \mathbf{M}\Delta \mathbf{F}_{t-1} + \mathbf{H}\mathbf{u}_t, \quad (\text{A2})$$

where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are $r \times c$ with $c < r$ and \mathbf{H} is $r \times q$. Moreover, from Barigozzi et al. (2016a) we have $d \leq c \leq (r - q + d)$. We are then interested in finding c , and the expressions of \mathbf{M} , $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and \mathbf{H} as functions of the parameters \mathbf{a} , \mathbf{b} , Γ_1 , and Γ_2 in (A1). Let us write $\boldsymbol{\alpha} = (\boldsymbol{\alpha}'_1 \boldsymbol{\alpha}'_2)'$ and $\boldsymbol{\beta} = (\boldsymbol{\beta}'_1 \boldsymbol{\beta}'_2)'$ where $\boldsymbol{\alpha}_1$, $\boldsymbol{\alpha}_2$, $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$ are all $q \times c$. We also denote as \mathbf{M}_{ij} for $i, j = 1, 2$ the four $q \times q$ blocks of \mathbf{M} and as \mathbf{H}_1 and \mathbf{H}_2 the two $q \times q$ blocks of \mathbf{H} . Following Proietti (1997), we define the $(2r + c)$ -dimensional vector

$$\mathbf{G}_t = \begin{pmatrix} \Delta \mathbf{F}_t \\ \Delta \mathbf{F}_{t-1} \\ \boldsymbol{\beta}'\mathbf{F}_{t-2} \end{pmatrix} = \begin{pmatrix} \Delta \mathbf{f}_t \\ \Delta \mathbf{f}_{t-1} \\ \Delta \mathbf{f}_{t-1} \\ \Delta \mathbf{f}_{t-2} \\ \boldsymbol{\beta}'_1\mathbf{f}_{t-2} + \boldsymbol{\beta}'_2\mathbf{f}_{t-3} \end{pmatrix}.$$

Then, the state-space form of (A2) is given by

$$\begin{aligned} \Delta \mathbf{F}_t &= \mathbf{Z}\mathbf{G}_t, \\ \mathbf{G}_t &= \mathbf{T}\mathbf{G}_{t-1} + \mathbf{Z}'\mathbf{H}\mathbf{u}_t, \end{aligned} \quad (\text{A3})$$

with the $r \times (2r + c)$ matrix $\mathbf{Z} = (\mathbf{I}_r \mathbf{0}_r \mathbf{0}_{r \times c})$. Then,

$$\mathbf{Z}'\mathbf{H} = \begin{pmatrix} \mathbf{I}_r \\ \mathbf{0}_r \\ \mathbf{0}_{c \times r} \end{pmatrix} \begin{pmatrix} \mathbf{H}_1 \\ \mathbf{H}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{H}_1 \\ \mathbf{H}_2 \\ \mathbf{0}_q \\ \mathbf{0}_q \\ \mathbf{0}_{c \times q} \end{pmatrix}.$$

and the $(2r + c) \times (2r + c)$ matrix \mathbf{T} is given by

$$\mathbf{T} = \begin{pmatrix} \mathbf{M} & -\boldsymbol{\alpha}\boldsymbol{\beta}' & -\boldsymbol{\alpha} \\ \mathbf{I}_r & \mathbf{0}_r & \mathbf{0}_{r \times c} \\ \mathbf{0}_{c \times r} & \boldsymbol{\beta}' & \mathbf{I}_c \end{pmatrix} = \left(\begin{array}{cc|cc|c} \mathbf{M}_{11} & \mathbf{M}_{12} & -\boldsymbol{\alpha}_1\boldsymbol{\beta}'_1 & -\boldsymbol{\alpha}_1\boldsymbol{\beta}'_2 & -\boldsymbol{\alpha}_1 \\ \mathbf{M}_{21} & \mathbf{M}_{22} & -\boldsymbol{\alpha}_2\boldsymbol{\beta}'_1 & -\boldsymbol{\alpha}_2\boldsymbol{\beta}'_2 & -\boldsymbol{\alpha}_2 \\ \hline \mathbf{I}_q & \mathbf{0}_q & \mathbf{0}_q & \mathbf{0}_q & \mathbf{0}_{q \times c} \\ \mathbf{0}_q & \mathbf{I}_q & \mathbf{0}_q & \mathbf{0}_q & \mathbf{0}_{q \times c} \\ \hline \mathbf{0}_{c \times q} & \mathbf{0}_{c \times q} & \boldsymbol{\beta}'_1 & \boldsymbol{\beta}'_2 & \mathbf{I}_c \end{array} \right).$$

Now using these definitions into (A3) we have five q -dimensional equations. The first one is

$$\Delta \mathbf{f}_t = \mathbf{M}_{11}\Delta \mathbf{f}_{t-1} + \mathbf{M}_{12}\Delta \mathbf{f}_{t-2} - \boldsymbol{\alpha}_1\boldsymbol{\beta}'_1\mathbf{f}_{t-2} - \boldsymbol{\alpha}_1\boldsymbol{\beta}'_2\mathbf{f}_{t-3} + \mathbf{H}_1\mathbf{u}_t,$$

which is equivalent to (A1) when

$$\mathbf{M}_{11} = \boldsymbol{\Gamma}_1, \quad \mathbf{M}_{12} = \boldsymbol{\Gamma}_2, \quad \boldsymbol{\alpha}_1 = \mathbf{a}, \quad \boldsymbol{\beta}_1 = \mathbf{0}_{q \times c}, \quad \boldsymbol{\beta}_2 = \mathbf{b}, \quad \mathbf{H}_1 = \mathbf{I}_q, \quad c = d. \quad (\text{A4})$$

The second equation is

$$\Delta \mathbf{f}_{t-1} = \mathbf{M}_{21}\Delta \mathbf{f}_{t-1} + \mathbf{M}_{22}\Delta \mathbf{f}_{t-2} - \boldsymbol{\alpha}_2\boldsymbol{\beta}'_1\Delta \mathbf{f}_{t-2} - \boldsymbol{\alpha}_2\boldsymbol{\beta}'_2\Delta \mathbf{f}_{t-3} - \boldsymbol{\alpha}_2\boldsymbol{\beta}'_1\mathbf{f}_{t-3} - \boldsymbol{\alpha}_2\boldsymbol{\beta}'_2\mathbf{f}_{t-4} + \mathbf{H}_2\mathbf{u}_t,$$

from which we see that we must also have

$$\mathbf{M}_{21} = \mathbf{I}_q, \quad \mathbf{M}_{22} = \mathbf{0}_q, \quad \boldsymbol{\alpha}_2 = \mathbf{0}_{q \times c}, \quad \mathbf{H}_2 = \mathbf{0}_q. \quad (\text{A5})$$

Under (A4) and (A5) the third, fourth and fifth equation in (A3) are just identities.

By imposing these restrictions we have the mapping between the VECM(1) for \mathbf{F}_t in (A2) and the VECM(2) for \mathbf{f}_t in (A1)

$$\mathbf{M} = \begin{pmatrix} \boldsymbol{\Gamma}_1 & \boldsymbol{\Gamma}_2 \\ \mathbf{I}_q & \mathbf{0}_q \end{pmatrix}, \quad \boldsymbol{\alpha} = \begin{pmatrix} \mathbf{a} \\ \mathbf{0}_{q \times d} \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \mathbf{0}_{q \times d} \\ \mathbf{b} \end{pmatrix}, \quad \mathbf{H} = \begin{pmatrix} \mathbf{I}_q \\ \mathbf{0}_q \end{pmatrix}. \quad (\text{A6})$$

If we now consider a generic \mathbf{K} in R1, then (A2) holds for $\mathbf{F}_t = \mathbf{K}(\mathbf{f}'_t \mathbf{f}'_{t-1})'$ and (A6) becomes

$$\mathbf{M} = \mathbf{K} \begin{pmatrix} \boldsymbol{\Gamma}_1 & \boldsymbol{\Gamma}_2 \\ \mathbf{I}_q & \mathbf{0}_q \end{pmatrix} \mathbf{K}^{-1}, \quad \boldsymbol{\alpha} = \mathbf{K} \begin{pmatrix} \mathbf{a} \\ \mathbf{0}_{q \times d} \end{pmatrix}, \quad \boldsymbol{\beta} = \mathbf{K}^{-1'} \begin{pmatrix} \mathbf{0}_{q \times d} \\ \mathbf{b} \end{pmatrix}, \quad \mathbf{H} = \mathbf{K} \begin{pmatrix} \mathbf{I}_q \\ \mathbf{0}_q \end{pmatrix}.$$

The cointegration rank c of \mathbf{F}_t is given by $\text{rk}(\boldsymbol{\alpha}\boldsymbol{\beta}') = d$.

A.2 Reduced and structural form of the state-space representation

Consider (12)-(13) written in matrix notation and using the companion form of the VAR

$$\mathbf{x}_t = \boldsymbol{\Lambda}\mathbf{F}_t + \boldsymbol{\xi}_t, \quad (\text{A7})$$

$$\begin{pmatrix} \mathbf{F}_t \\ \mathbf{F}_{t-1} \end{pmatrix} = \begin{pmatrix} \mathbf{A}_1 & \mathbf{A}_2 \\ \mathbf{I}_r & \mathbf{0}_r \end{pmatrix} \begin{pmatrix} \mathbf{F}_{t-1} \\ \mathbf{F}_{t-2} \end{pmatrix} + \begin{pmatrix} \mathbf{H} \\ \mathbf{0}_{r \times q} \end{pmatrix} \mathbf{u}_t, \quad (\text{A8})$$

with $\boldsymbol{\Lambda} = (\boldsymbol{\lambda}_1 \cdots \boldsymbol{\lambda}_n)'$ the $n \times r$ loadings matrix. We call (A7)-(A8) the reduced form of the model. Similarly consider the structural form, where, for convenience, in the VAR we write

twice the same equation:

$$\mathbf{x}_t = \mathbf{B}_0 \mathbf{f}_t + \mathbf{B}_1 \mathbf{f}_{t-1} + \boldsymbol{\xi}_t, \quad (\text{A9})$$

$$\begin{pmatrix} \mathbf{f}_t \\ \mathbf{f}_{t-1} \\ \mathbf{f}_{t-1} \\ \mathbf{f}_{t-2} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Pi}_1 & \boldsymbol{\Pi}_2 & \mathbf{0}_q & \boldsymbol{\Pi}_3 \\ \mathbf{I}_q & \mathbf{0}_q & \mathbf{0}_q & \mathbf{0}_q \\ \mathbf{I}_q & \mathbf{0}_q & \mathbf{0}_q & \mathbf{0}_q \\ \mathbf{0}_q & \mathbf{I}_q & \mathbf{0}_q & \mathbf{0}_q \end{pmatrix} \begin{pmatrix} \mathbf{f}_{t-1} \\ \mathbf{f}_{t-2} \\ \mathbf{f}_{t-2} \\ \mathbf{f}_{t-3} \end{pmatrix} + \begin{pmatrix} \mathbf{I}_q \\ \mathbf{0}_q \\ \mathbf{0}_q \\ \mathbf{0}_q \end{pmatrix} \mathbf{u}_t, \quad (\text{A10})$$

where $\mathbf{B}_0 = (\mathbf{b}_{01} \cdots \mathbf{b}_{0n})'$, $\mathbf{B}_1 = (\mathbf{b}_{11} \cdots \mathbf{b}_{1n})$ are both $n \times q$. Because of R1 there exists an invertible $r \times r$ matrix \mathbf{K} such that

$$\mathbf{F}_t = \mathbf{K}(\mathbf{f}'_t \mathbf{f}'_{t-1})', \quad (\mathbf{f}'_t \mathbf{f}'_{t-1})' = \mathbf{K}^{-1} \mathbf{F}_t, \quad (\text{A11})$$

$$\boldsymbol{\Lambda} = (\mathbf{B}_0 \mathbf{B}_1) \mathbf{K}^{-1}, \quad (\mathbf{B}_0 \mathbf{B}_1) = \boldsymbol{\Lambda} \mathbf{K}. \quad (\text{A12})$$

By comparing (A7)-(A8) with (A9)-(A10) and using (A11)-(A12), we have the parameters of the reduced form

$$\mathbf{A}_1 = \mathbf{K} \begin{pmatrix} \boldsymbol{\Pi}_1 & \boldsymbol{\Pi}_2 \\ \mathbf{I}_q & \mathbf{0}_q \end{pmatrix} \mathbf{K}^{-1}, \quad \mathbf{A}_2 = \mathbf{K} \begin{pmatrix} \mathbf{0}_q & \boldsymbol{\Pi}_3 \\ \mathbf{0}_q & \mathbf{0}_q \end{pmatrix} \mathbf{K}^{-1}, \quad \mathbf{H} = \mathbf{K} \begin{pmatrix} \mathbf{I}_q \\ \mathbf{0}_q \end{pmatrix}. \quad (\text{A13})$$

The relations (A11), (A12) and (A13) are used throughout the following. Moreover, since a VAR(2) of dimension r can always be written as a VAR(1) of dimension $2r$, to avoid introducing further notation hereafter we consider the case of a VAR(1) for \mathbf{F}_t , where $\mathbf{A} \equiv \mathbf{A}_1$.

A.3 Properties of the structural and reduced form of the linear system

Lemma 1. *The structural model (A9)-(A10) is stabilizable and detectable.*

Proof. Equations (A9)-(A10) define a linear system with $r = 2q$ latent states $(\mathbf{f}'_t \mathbf{f}'_{t-1})'$. We say that a linear system is stabilizable if its unstable (non-stationary) states are controllable and all uncontrollable states are stable (see Anderson and Moore, 1979, page 342), where stability is dictated by the eigenvalues of the matrix of VAR coefficients, which we denote as

$$\tilde{\mathbf{A}} = \begin{pmatrix} \boldsymbol{\Pi}_1 & \boldsymbol{\Pi}_2 \\ \mathbf{I}_q & \mathbf{0}_q \end{pmatrix} \quad (\text{A14})$$

Because of cointegration, $\tilde{\mathbf{A}}$ has $(q-d)$ unit eigenvalues corresponding to $(q-d)$ unstable states. Moreover, $(\mathbf{I}_q - \boldsymbol{\Pi}_1 - \boldsymbol{\Pi}_2) = \mathbf{a} \mathbf{b}'$, where \mathbf{a} and \mathbf{b} have full column-rank $q \times d$ matrices, so that $\text{rk}(\mathbf{a} \mathbf{b}') = d$. Define the $q \times (q-d)$ matrices \mathbf{a}_\perp and \mathbf{b}_\perp such that $\mathbf{a}'_\perp \mathbf{a} = \mathbf{b}'_\perp \mathbf{b} = \mathbf{0}_{(q-d) \times d}$. Then, since $\text{rk}(\mathbf{a}'_\perp \mathbf{I}_q) = (q-d)$, the unstable states are controllable because they satisfy the Popov-Belevitch-Hautus rank test (see Franchi, 2017, Theorem 2.1, and Antsaklis and Michel, 2007, Corollary 6.11, page 249).

Now, by looking at (A10), we see that $\tilde{\mathbf{A}}$ has also $(r-q+d) = (q+d)$ eigenvalues which are smaller than one in absolute value. Of these q correspond to states which are uncontrollable because they are not driven by any shock, but are also stable since have no dynamics (see the second equation in (A10)). The remaining d states follow a stable VAR, hence are controllable.

Similarly, we say that a linear system is detectable if its unstable states are observable and all unobservable states are stable (see Anderson and Moore, 1979, page 342). First, notice

that $\text{rk}(\mathbf{B}_0) = q$ and $\text{rk}(\mathbf{B}_1) = q$ because of C2 and (A12), therefore $\text{rk}(\mathbf{B}_0\mathbf{b}_\perp) = (q - d)$ and $\text{rk}(\mathbf{B}_1\mathbf{b}_\perp) = (q - d)$, which implies that the unstable states are observable because they satisfy the Popov-Belevitch-Hautus rank test (see Franchi, 2017, Theorem 2.1, and Antsaklis and Michel, 2007, Corollary 6.11, page 249). Since \mathbf{B}_0 and \mathbf{B}_1 have full column-rank there are no unstable unobservable states. This completes the proof. \square

Appendix B Details of estimation

This appendix provides details on estimation of factors and parameters which are necessary to introduce the notation required in the proofs in Appendix C. The model considered is (12)-(14), where for simplicity of exposition we consider a VAR(1) for the factors. As explained in the text without loss of generality we can assume $\boldsymbol{\xi}_t \sim I(0)$, thus considering as latent states only the r static factors.

Adding $I(1)$ idiosyncratic components as latent states does not increase the dimension of the parameter space but it increases the dimension of the latent states vector. However, since the idiosyncratic components are assumed to be orthogonal (see D2), and moreover they are orthogonal to the static factors (see A2), the results in Appendix C can be generalized to this case by treating each new state separately.

B.1 Expectation Maximization algorithm

In what follows we denote the whole sample of observed data as $\boldsymbol{\mathcal{X}}_T := (\mathbf{x}_1 \cdots \mathbf{x}_T)'$ and the whole history of the unknown factors as $\boldsymbol{\mathcal{F}}_T := (\mathbf{F}_1 \cdots \mathbf{F}_T)'$. Recall that the vector of parameters is given by $\boldsymbol{\Theta} := (\text{vec}(\boldsymbol{\Lambda})' \text{vec}(\mathbf{A})' \text{vec}(\mathbf{H})' \text{diag}(\mathbf{R}))'$. To avoid heavier notation we use $\boldsymbol{\Theta}$ to indicate both a generic value of the parameters and the true value, whether we refer to one or the other is either clearly implied by the context or explicitly stated.

The joint pdf of data and factors is denoted as $f(\boldsymbol{\mathcal{X}}_T, \boldsymbol{\mathcal{F}}_T; \boldsymbol{\Theta})$ and the corresponding joint log-likelihood is denoted as $\ell(\boldsymbol{\mathcal{X}}_T, \boldsymbol{\mathcal{F}}_T; \boldsymbol{\Theta}) := \log f(\boldsymbol{\mathcal{X}}_T, \boldsymbol{\mathcal{F}}_T; \boldsymbol{\Theta})$ and it is such that

$$\ell(\boldsymbol{\mathcal{X}}_T, \boldsymbol{\mathcal{F}}_T; \boldsymbol{\Theta}) = \ell(\boldsymbol{\mathcal{X}}_T | \boldsymbol{\mathcal{F}}_T; \boldsymbol{\Theta}) + \ell(\boldsymbol{\mathcal{F}}_T; \boldsymbol{\Theta}), \quad (\text{B1})$$

where $\ell(\boldsymbol{\mathcal{X}}_T | \boldsymbol{\mathcal{F}}_T; \boldsymbol{\Theta})$ is the log-likelihood of the data conditional on the factors and $\ell(\boldsymbol{\mathcal{F}}_T; \boldsymbol{\Theta})$ is the marginal log-likelihood of the factors. Because of D2 and D3 all log-likelihoods are Gaussian and in particular

$$\ell(\boldsymbol{\mathcal{X}}_T | \boldsymbol{\mathcal{F}}_T; \boldsymbol{\Theta}) = -\frac{nT}{2} \log(2\pi) - \frac{T}{2} \log \det(\mathbf{R}) - \frac{1}{2} \text{tr} [(\boldsymbol{\mathcal{X}}_T - \boldsymbol{\mathcal{F}}_T \boldsymbol{\Lambda}') \mathbf{R}^{-1} (\boldsymbol{\mathcal{X}}_T - \boldsymbol{\mathcal{F}}_T \boldsymbol{\Lambda}')']. \quad (\text{B2})$$

We first briefly review the steps of the EM algorithm, while in Section B.1.4 we prove that the values of the parameters obtained at convergence of the EM algorithm converge to the QML estimator.

B.1.1 Initialization

The EM algorithm is initialised with estimated parameters

$$\hat{\boldsymbol{\Theta}}_0 := (\text{vec}(\hat{\boldsymbol{\Lambda}}_0)' \text{vec}(\hat{\mathbf{A}}_0)' \text{vec}(\hat{\mathbf{H}}_0)' \text{diag}(\hat{\mathbf{R}}_0))'. \quad (\text{B3})$$

These are obtained as follows. From the integration of the first r principal components of $\Delta \mathbf{x}_t$ we have an estimator of the factors, $\tilde{\mathbf{F}}_t$, and then of the loadings, $\hat{\mathbf{\Lambda}}_0$. The VAR parameters, $\hat{\mathbf{A}}_0$ are obtained by fitting a VAR on the estimated factors $\tilde{\mathbf{F}}_t$ and the columns of $\hat{\mathbf{H}}_0$ are given by the q leading eigenvectors of the covariance matrix of VAR residuals. Finally, the diagonal entries of $\hat{\mathbf{R}}_0$, are obtained as sample variances of $\hat{\xi}_{it,0} = x_{it} - \hat{\mathbf{\Lambda}}_0 \tilde{\mathbf{F}}_t$. Consistency of these estimators is discussed in Section C.3.

B.1.2 E-step

At iteration $k \geq 0$, given \mathcal{X}_T and an estimate of the parameters $\hat{\Theta}_k$, we compute the expected log-likelihood as function of a generic value of the parameters Θ , where the expectation is computed with respect to the conditional distribution of \mathcal{F}_T given \mathcal{X}_T and when using $\hat{\Theta}_k$:

$$\mathcal{Q}(\Theta; \hat{\Theta}_k) := \int_{\mathbb{R}^{r \times T}} \ell(\mathcal{X}_T, \mathcal{F}_T; \Theta) f(\mathcal{F}_T | \mathcal{X}_T; \hat{\Theta}_k) d\mathcal{F}_T = \mathbb{E}_{\hat{\Theta}_k}[\ell(\mathcal{X}_T, \mathcal{F}_T; \Theta) | \mathcal{X}_T]. \quad (\text{B4})$$

In the Gaussian case (B4) depends on the conditional mean of the factors and their conditional second moments, which are obtained with the KS when using the parameters $\hat{\Theta}_k$ and are given by (see Section B.2 for details)

$$\mathbf{F}_{t|T,k} := \mathbb{E}_{\hat{\Theta}_k}[\mathbf{F}_t | \mathcal{X}_T], \quad \mathbf{P}_{t|T,k} := \mathbb{E}_{\hat{\Theta}_k}[(\mathbf{F}_t - \mathbf{F}_{t|T,k})(\mathbf{F}_t - \mathbf{F}_{t|T,k})' | \mathcal{X}_T]. \quad (\text{B5})$$

B.1.3 M-step

A new estimator of the parameters is obtained by maximising the expected log-likelihood over all possible values of the parameters:

$$\hat{\Theta}_{k+1} = \arg \max_{\Theta \in \Omega \subseteq \mathbb{R}^Q} \mathcal{Q}(\Theta; \hat{\Theta}_k). \quad (\text{B6})$$

Thus, maximizing the conditional expectation of (B2) and using (B5), we have the loadings estimator

$$\begin{aligned} \hat{\mathbf{\Lambda}}_{k+1} &= \left(\sum_{t=1}^T \mathbb{E}_{\hat{\Theta}_k}[\mathbf{x}_t \mathbf{F}'_t | \mathcal{X}_T] \right) \left(\sum_{t=1}^T \mathbb{E}_{\hat{\Theta}_k}[\mathbf{F}_t \mathbf{F}'_t | \mathcal{X}_T] \right)^{-1} \\ &= \left(\sum_{t=1}^T \mathbf{x}_t \mathbf{F}'_{t|T,k} \right) \left(\sum_{t=1}^T (\mathbf{F}_{t|T,k} \mathbf{F}'_{t|T,k} + \mathbf{P}_{t|T,k}) \right)^{-1}. \end{aligned}$$

Similarly we can obtain estimates of the other parameters $\hat{\mathbf{A}}_{k+1}$ and $\hat{\mathbf{R}}_{k+1}$ (see e.g. Bańbura and Modugno, 2014, for their expressions). The columns of $\hat{\mathbf{H}}_{k+1}$ are obtained as the q leading eigenvectors of the matrix

$$\hat{\mathbf{\Sigma}}_{k+1} = \frac{1}{T} \left(\sum_{t=1}^T \mathbb{E}_{\hat{\Theta}_k}[\mathbf{F}_t \mathbf{F}'_t | \mathcal{X}_T] - \hat{\mathbf{A}}_{k+1} \sum_{t=1}^T \mathbb{E}_{\hat{\Theta}_k}[\mathbf{F}_t \mathbf{F}'_{t-1} | \mathcal{X}_T] \right),$$

which is an estimator of the covariance of the VAR residuals, and where the second expectation can also be computed from the output of the KS.

B.1.4 Convergence

Denote the QML estimator of the parameters as

$$\widehat{\Theta}^* := (\text{vec}(\widehat{\Lambda}^*)' \text{vec}(\widehat{\mathbf{A}}^*)' \text{vec}(\widehat{\mathbf{H}}^*)' \text{diag}(\widehat{\mathbf{R}}^*))', \quad (\text{B7})$$

then by definition we have

$$\widehat{\Theta}^* = \arg \max_{\Theta \in \Omega \subseteq \mathbb{R}^Q} \ell(\mathcal{X}_T; \Theta). \quad (\text{B8})$$

where $\ell(\mathcal{X}_T; \Theta)$ is the log-likelihood of the data such that

$$\ell(\mathcal{X}_T; \Theta) = \ell(\mathcal{X}_T, \mathcal{F}_T; \Theta) - \ell(\mathcal{F}_T | \mathcal{X}_T; \Theta), \quad (\text{B9})$$

where the first term on the rhs is given by (B1) and the second can be computed using the output of the KS for a given value of Θ . Define the expectation

$$\mathcal{H}(\Theta; \widehat{\Theta}_k) := \int_{\mathbb{R}^{r \times T}} \ell(\mathcal{F}_T | \mathcal{X}_T; \Theta) f(\mathcal{F}_T | \mathcal{X}_T; \widehat{\Theta}_k) d\mathcal{F}_T = \mathbb{E}_{\widehat{\Theta}_k}[\ell(\mathcal{F}_T | \mathcal{X}_T; \Theta) | \mathcal{X}_T], \quad (\text{B10})$$

and recall the definition of $\mathcal{Q}(\Theta; \widehat{\Theta}_k)$ in the E-step in (B4). Since the lhs of (B9) does not depend on \mathcal{F}_T , by taking its expectation with respect to the conditional distribution of \mathcal{F}_T given \mathcal{X}_T and when using $\widehat{\Theta}_k$, for any $\Theta \in \Omega$, we have

$$\ell(\mathcal{X}_T; \Theta) = \mathcal{Q}(\Theta; \widehat{\Theta}_k) - \mathcal{H}(\Theta; \widehat{\Theta}_k). \quad (\text{B11})$$

Now, by definition of Kullback-Leibler divergence, we have (see also Lemma 1 in Dempster et al., 1977)

$$\mathcal{H}(\widehat{\Theta}_{k+1}; \widehat{\Theta}_k) \leq \mathcal{H}(\widehat{\Theta}_k; \widehat{\Theta}_k). \quad (\text{B12})$$

Hence, from (B11) and (B12), for any k ,

$$\ell(\mathcal{X}_T; \widehat{\Theta}_{k+1}) - \ell(\mathcal{X}_T; \widehat{\Theta}_k) \geq \mathcal{Q}(\widehat{\Theta}_{k+1}; \widehat{\Theta}_k) - \mathcal{Q}(\widehat{\Theta}_k; \widehat{\Theta}_k) \geq 0$$

where the last inequality is a consequence of the M-step in (B6). This shows that the log-likelihood increases monotonically as k increases. Moreover, since due to Gaussianity $\mathcal{Q}(\Theta; \Theta')$ is continuous in Θ and Θ' and its gradient $\nabla_{\Theta} \mathcal{Q}(\Theta; \Theta')$ is continuous in Θ , then conditions for Theorems 1 and 2 and Corollary 1 in Wu (1983) are satisfied and we have convergence of the log-likelihood to its unique maximum and of the parameters to the corresponding QML estimators

$$\lim_{k \rightarrow \infty} \ell(\mathcal{X}_T; \widehat{\Theta}_k) = \ell(\mathcal{X}_T; \widehat{\Theta}^*), \quad \lim_{k \rightarrow \infty} \widehat{\Theta}_k = \widehat{\Theta}^*. \quad (\text{B13})$$

The previous result holds in the limit $k \rightarrow \infty$, but in practice we can run the EM algorithm only for a finite number of iterations k_{\max} . Define, for any k ,

$$\Delta \ell_k = \frac{|\ell(\mathcal{X}_T, \mathcal{F}_{T,k+1}; \widehat{\Theta}_{k+1}) - \ell(\mathcal{X}_T, \mathcal{F}_{T,k}; \widehat{\Theta}_k)|}{|\ell(\mathcal{X}_T, \mathcal{F}_{T,k+1}; \widehat{\Theta}_{k+1})| + |\ell(\mathcal{X}_T, \mathcal{F}_{T,k}; \widehat{\Theta}_k)|},$$

where $\mathcal{F}_{T,k} := (\mathbf{F}_{1|T,k} \cdots \mathbf{F}_{T|T,k})'$. We say that the algorithm has converged at iteration $k^* < k_{\max}$ according to the following rule, which is defined for a given threshold η ,

$$\Delta \ell_{k^*} < \eta, \quad \text{but} \quad \Delta \ell_{k^*-1} \geq \eta.$$

Once we find k^* , our estimator of the parameters is defined as $\widehat{\Theta} := \widehat{\Theta}_{k^*}$. The corresponding estimator of the factors is then defined as $\widehat{\mathbf{F}}_t := \mathbf{F}_{t|T,k^*}$, thus running the KS once last time using $\widehat{\Theta}_{k^*}$. The rate of convergence of $\widehat{\Theta}$ to $\widehat{\Theta}^*$ in (B13) is studied in Lemma 9 below.

B.2 Kalman filter and Kalman smoother

For ease of notation assume to know the true parameter collected in the vector Θ . When using the KF-KS in the EM algorithm at a given iteration k , the factors' estimators given below are obtained by replacing Θ with $\widehat{\Theta}_k$ throughout this section. We denote the conditional expectation and covariance of the factors as

$$\mathbf{F}_{t|s} := \mathbf{E}_{\Theta}[\mathbf{F}_t | \mathcal{X}_s], \quad \mathbf{P}_{t|s} := \mathbf{E}_{\Theta}[(\mathbf{F}_t - \mathbf{F}_{t|s})(\mathbf{F}_t - \mathbf{F}_{t|s})' | \mathcal{X}_s], \quad (\text{B14})$$

where $\mathcal{X}_s := (\mathbf{x}_1 \cdots \mathbf{x}_s)'$. Under Gaussianity (D2 and D3) these can be computed with the KF-KS. Specifically, when $s = t - 1$ we have the optimal one-step-ahead prediction, when $s = t$ we have the optimal in-sample estimator, when $s = T$ we have the optimal smoother. The KF gives the first two cases while the KS gives the latter. In particular, we denote the KF-KS estimators respectively as: $\mathbf{F}_{t|t}$ and $\mathbf{F}_{t|T}$ when using the true value Θ and as $\mathbf{F}_{t|t,k}$ and $\mathbf{F}_{t|T,k}$ when using $\widehat{\Theta}_k$ (see also (B5)).

B.2.1 Forward iterations - Filtering

For given initial conditions $\mathbf{F}_{0|0}$ and $\mathbf{P}_{0|0}$, the KF is based on the forward iterations for $t = 1, \dots, T$:

$$\mathbf{F}_{t|t-1} = \mathbf{A}\mathbf{F}_{t-1|t-1}, \quad (\text{B15})$$

$$\mathbf{P}_{t|t-1} = \mathbf{A}\mathbf{P}_{t-1|t-1}\mathbf{A}' + \mathbf{H}\mathbf{H}', \quad (\text{B16})$$

$$\mathbf{F}_{t|t} = \mathbf{F}_{t|t-1} + \mathbf{P}_{t|t-1}\mathbf{\Lambda}'(\mathbf{\Lambda}\mathbf{P}_{t|t-1}\mathbf{\Lambda}' + \mathbf{R})^{-1}(\mathbf{x}_t - \mathbf{\Lambda}\mathbf{F}_{t|t-1}), \quad (\text{B17})$$

$$\mathbf{P}_{t|t} = \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1}\mathbf{\Lambda}'(\mathbf{\Lambda}\mathbf{P}_{t|t-1}\mathbf{\Lambda}' + \mathbf{R})^{-1}\mathbf{\Lambda}\mathbf{P}_{t|t-1}. \quad (\text{B18})$$

Moreover, by combining (B16) and (B18), we obtain the Riccati difference equation

$$\mathbf{P}_{t+1|t} - \mathbf{A}\mathbf{P}_{t|t-1}\mathbf{A}' + \mathbf{A}\mathbf{P}_{t|t-1}\mathbf{\Lambda}'(\mathbf{\Lambda}\mathbf{P}_{t|t-1}\mathbf{\Lambda}' + \mathbf{R})^{-1}\mathbf{\Lambda}\mathbf{P}_{t|t-1}\mathbf{A}' = \mathbf{H}\mathbf{H}'. \quad (\text{B19})$$

The KF is started with given values of $\mathbf{F}_{0|0}$ and $\mathbf{P}_{0|0}$. The latter can be obtained with a diffuse prior run for $t < 0$ (see Koopman, 1997, and Koopman and Durbin, 2000, for details).

B.2.2 Backward iterations - Smoothing

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

$$\mathbf{F}_{t|T} = \mathbf{F}_{t|t} + \mathbf{P}_{t|t} \mathbf{A}' \mathbf{P}_{t+1|t}^{-1} (\mathbf{F}_{t+1|T} - \mathbf{F}_{t+1|t}), \quad (\text{B20})$$

$$\mathbf{P}_{t|T} = \mathbf{P}_{t|t} + \mathbf{P}_{t|t} \mathbf{A}' \mathbf{P}_{t+1|t}^{-1} (\mathbf{P}_{t+1|T} - \mathbf{P}_{t+1|t}) \mathbf{P}_{t+1|t}^{-1} \mathbf{A} \mathbf{P}_{t|t}. \quad (\text{B21})$$

The KS iterations in (B20) require T inversions of $\mathbf{P}_{t|t-1}$ and in the singular case $r > q$ these matrices are likely to be singular (see also Lemma 5). There are two possible solutions to this problem. Kohn and Ansley (1983) suggest to use a generalized inverse of $\mathbf{P}_{t|t-1}$, like the Moore-Penrose one. Alternatively, it can be proved that (B20) can be written in an equivalent way, which does not require matrix inversion, and which is defined by the backward iterations for $t = T, \dots, 1$:

$$\mathbf{F}_{t|T} = \mathbf{F}_{t|t-1} + \mathbf{P}_{t|t-1} \mathbf{r}_{t-1}, \quad (\text{B22})$$

$$\mathbf{r}_{t-1} = \mathbf{\Lambda}' (\mathbf{\Lambda} \mathbf{P}_{t|t-1} \mathbf{\Lambda}' + \mathbf{R})^{-1} (\mathbf{x}_t - \mathbf{\Lambda} \mathbf{F}_{t|t-1}) + \mathbf{L}'_t \mathbf{r}_t, \quad (\text{B23})$$

$$\mathbf{P}_{t|T} = \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1} \mathbf{N}_{t-1} \mathbf{P}_{t|t-1}, \quad (\text{B24})$$

$$\mathbf{N}_{t-1} = \mathbf{\Lambda}' (\mathbf{\Lambda} \mathbf{P}_{t|t-1} \mathbf{\Lambda}' + \mathbf{R})^{-1} \mathbf{\Lambda} + \mathbf{L}'_t \mathbf{N}_t \mathbf{L}_t, \quad (\text{B25})$$

$$\mathbf{L}_t = \mathbf{A} - \mathbf{A} \mathbf{P}_{t|t-1} \mathbf{\Lambda}' (\mathbf{\Lambda} \mathbf{P}_{t|t-1} \mathbf{\Lambda}' + \mathbf{R})^{-1} \mathbf{\Lambda}, \quad (\text{B26})$$

where $\mathbf{r}_T = \mathbf{0}_{r \times 1}$, $\mathbf{N}_T = \mathbf{0}_r$ and by construction $\mathbf{A} \mathbf{P}_{t|t} = \mathbf{L}_t \mathbf{P}_{t|t-1}$ (see also Durbin and Koopman, 2001, pp.70-73). Although numerically no appreciable differences emerge with respect to the chosen method, (B22)-(B26) are particularly useful for our proofs.

Appendix C Consistency of the EM algorithm

C.1 Preliminary results

Lemma 2. For $m < n$, and given symmetric positive definite matrices \mathbf{A} of dimension $m \times m$, \mathbf{B} of dimension $n \times n$, and for \mathbf{C} of dimension $n \times m$ with full column-rank, the following holds

$$\mathbf{A} \mathbf{C}' (\mathbf{C} \mathbf{A} \mathbf{C}' + \mathbf{B})^{-1} = (\mathbf{A}^{-1} + \mathbf{C}' \mathbf{B}^{-1} \mathbf{C})^{-1} \mathbf{C}' \mathbf{B}^{-1}. \quad (\text{C27})$$

Proof. Recall the Woodbury formula

$$(\mathbf{C} \mathbf{A} \mathbf{C}' + \mathbf{B})^{-1} = \mathbf{B}^{-1} - \mathbf{B}^{-1} \mathbf{C} (\mathbf{A}^{-1} + \mathbf{C}' \mathbf{B}^{-1} \mathbf{C})^{-1} \mathbf{C}' \mathbf{B}^{-1}. \quad (\text{C28})$$

Denote $\mathbf{D} = (\mathbf{A}^{-1} + \mathbf{C}' \mathbf{B}^{-1} \mathbf{C})^{-1}$ then from (C28) the lhs of (C27) is equivalent to

$$\mathbf{A} \mathbf{C}' [\mathbf{B}^{-1} - \mathbf{B}^{-1} \mathbf{C} \mathbf{D} \mathbf{C}' \mathbf{B}^{-1}] = \mathbf{A} [\mathbf{C}' \mathbf{B}^{-1} - \mathbf{C}' \mathbf{B}^{-1} \mathbf{C} \mathbf{D} \mathbf{C}' \mathbf{B}^{-1}] = \mathbf{A} [\mathbf{I} - \mathbf{C}' \mathbf{B}^{-1} \mathbf{C} \mathbf{D}] \mathbf{C}' \mathbf{B}^{-1}.$$

Then, (C27) becomes

$$\mathbf{A} [\mathbf{I} - \mathbf{C}' \mathbf{B}^{-1} \mathbf{C} \mathbf{D}] \mathbf{C}' \mathbf{B}^{-1} = \mathbf{D} \mathbf{C}' \mathbf{B}^{-1},$$

or equivalently multiplying both sides on the right by $\mathbf{B} \mathbf{C} (\mathbf{C}' \mathbf{C})^{-1}$

$$\mathbf{A} [\mathbf{I} - \mathbf{C}' \mathbf{B}^{-1} \mathbf{C} \mathbf{D}] = \mathbf{D}. \quad (\text{C29})$$

Now notice that

$$\mathbf{D} = (\mathbf{A}^{-1} + \mathbf{C}'\mathbf{B}^{-1}\mathbf{C})^{-1} = \mathbf{A}(\mathbf{I} + \mathbf{A}\mathbf{C}'\mathbf{B}^{-1}\mathbf{C})^{-1}. \quad (\text{C30})$$

Substituting (C30) in (C29) and multiplying both sides on the left by \mathbf{A}^{-1}

$$[\mathbf{I} - \mathbf{C}'\mathbf{B}^{-1}\mathbf{C}\mathbf{A}(\mathbf{I} + \mathbf{A}\mathbf{C}'\mathbf{B}^{-1}\mathbf{C})^{-1}] = (\mathbf{I} + \mathbf{A}\mathbf{C}'\mathbf{B}^{-1}\mathbf{C})^{-1}.$$

Multiplying both sides on the right by $(\mathbf{I} + \mathbf{A}\mathbf{C}'\mathbf{B}^{-1}\mathbf{C})$ we have that (C27) is equivalent to

$$\mathbf{I} + \mathbf{A}\mathbf{C}'\mathbf{B}^{-1}\mathbf{C} - \mathbf{C}'\mathbf{B}^{-1}\mathbf{C}\mathbf{A} = \mathbf{I} \quad (\text{C31})$$

Therefore (C27) is correct provided that $\mathbf{A}\mathbf{C}'\mathbf{B}^{-1}\mathbf{C} = \mathbf{C}'\mathbf{B}^{-1}\mathbf{C}\mathbf{A}$ which is always true since both \mathbf{A} and $\mathbf{C}'\mathbf{B}^{-1}\mathbf{C}$ are symmetric. \square

Lemma 3. For $m < n$ with m independent of n and given

- (a) an $m \times m$ matrix \mathbf{A} symmetric and positive definite with $\mu_j^{\mathbf{A}} \leq M$ for $j = 1, \dots, m$;
- (b) an $n \times n$ matrix \mathbf{B} symmetric and positive definite with $\mu_j^{\mathbf{B}} \leq M$ for $j = 1, \dots, n$;
- (c) an $n \times m$ matrix \mathbf{C} such that $\mathbf{C}'\mathbf{C}$ is positive definite with $\mu_j^{\mathbf{C}'\mathbf{C}} = M_j n$ for $j = 1, \dots, m$;

then the following holds

$$(\mathbf{A}^{-1} + \mathbf{C}'\mathbf{B}^{-1}\mathbf{C})^{-1}\mathbf{C}'\mathbf{B}^{-1}\mathbf{C} = \mathbf{I}_m + O(n^{-1}).$$

Proof. First notice that for two matrices \mathbf{K} and \mathbf{H} we have

$$\begin{aligned} (\mathbf{H} + \mathbf{K})^{-1} &= (\mathbf{H} + \mathbf{K})^{-1} - \mathbf{K}^{-1} + \mathbf{K}^{-1} = (\mathbf{H} + \mathbf{K})^{-1}(\mathbf{K} - (\mathbf{H} + \mathbf{K}))\mathbf{K}^{-1} + \mathbf{K}^{-1} \\ &= (\mathbf{H} + \mathbf{K})^{-1}(-\mathbf{H})\mathbf{K}^{-1} + \mathbf{K}^{-1} = \mathbf{K}^{-1} - (\mathbf{H} + \mathbf{K})^{-1}\mathbf{H}\mathbf{K}^{-1}. \end{aligned} \quad (\text{C32})$$

Then setting $\mathbf{K} = \mathbf{C}'\mathbf{B}^{-1}\mathbf{C}$ and $\mathbf{H} = \mathbf{A}^{-1}$ from (C32) we have

$$(\mathbf{A}^{-1} + \mathbf{C}'\mathbf{B}^{-1}\mathbf{C})^{-1} = (\mathbf{C}'\mathbf{B}^{-1}\mathbf{C})^{-1} - (\mathbf{A}^{-1} + \mathbf{C}'\mathbf{B}^{-1}\mathbf{C})^{-1}\mathbf{A}^{-1}(\mathbf{C}'\mathbf{B}^{-1}\mathbf{C})^{-1}.$$

which implies

$$(\mathbf{A}^{-1} + \mathbf{C}'\mathbf{B}^{-1}\mathbf{C})^{-1}\mathbf{C}'\mathbf{B}^{-1}\mathbf{C} = \mathbf{I}_m - (\mathbf{A}^{-1} + \mathbf{C}'\mathbf{B}^{-1}\mathbf{C})^{-1}\mathbf{A}^{-1}. \quad (\text{C33})$$

Now consider the second term on the rhs of (C33)

$$\begin{aligned} \|(\mathbf{A}^{-1} + \mathbf{C}'\mathbf{B}^{-1}\mathbf{C})^{-1}\mathbf{A}^{-1}\|^2 &\leq \|(\mathbf{A}^{-1} + \mathbf{C}'\mathbf{B}^{-1}\mathbf{C})^{-1}\|^2 \|\mathbf{A}^{-1}\|^2 \\ &\leq \|(\mathbf{C}'\mathbf{B}^{-1}\mathbf{C})^{-1}\|^2 (\mu_n^{\mathbf{A}})^{-1} \leq \|(\mathbf{C}'\mathbf{B}^{-1}\mathbf{C})^{-1}\|^2 M_1^{-1}, \end{aligned} \quad (\text{C34})$$

where we use norm sub-additivity and the fact that by condition (a) \mathbf{A} and \mathbf{A}^{-1} are positive definite and therefore $\mu_n^{\mathbf{A}} \geq M_1 > 0$ and moreover $\mu_n^{\mathbf{A}^{-1}} \geq M_2 > 0$ thus by Weyl's inequality

$$\mu_n^{\mathbf{A}^{-1} + \mathbf{C}'\mathbf{B}^{-1}\mathbf{C}} \geq \mu_n^{\mathbf{A}^{-1}} + \mu_n^{\mathbf{C}'\mathbf{B}^{-1}\mathbf{C}} \geq \mu_n^{\mathbf{C}'\mathbf{B}^{-1}\mathbf{C}},$$

therefore,

$$\|(\mathbf{A}^{-1} + \mathbf{C}'\mathbf{B}^{-1}\mathbf{C})^{-1}\| = (\mu_n^{A^{-1} + \mathbf{C}'\mathbf{B}^{-1}\mathbf{C}})^{-1} \leq (\mu_n^{C'\mathbf{B}^{-1}\mathbf{C}})^{-1} = \|(\mathbf{C}'\mathbf{B}^{-1}\mathbf{C})^{-1}\|.$$

Then, the first term on the rhs of (C34) is

$$\|(\mathbf{C}'\mathbf{B}^{-1}\mathbf{C})^{-1}\|^2 \leq \text{tr} [(\mathbf{C}'\mathbf{B}^{-1}\mathbf{C})^{-2}] = \sum_{j=1}^m \frac{1}{(\mu_j^{C'\mathbf{B}^{-1}\mathbf{C}})^2} = O(n^{-2}). \quad (\text{C35})$$

Indeed, the m eigenvalues of $\mathbf{C}'\mathbf{B}^{-1}\mathbf{C}$ are also the m non-zero eigenvalues of $\mathbf{B}^{-1/2}\mathbf{C}\mathbf{C}'\mathbf{B}^{-1/2}$, which are all $O(n)$ by conditions (b) and (c). By using (C34) and (C35) in (C33) we prove the Lemma. \square

C.2 Consistency of KF and KS using the true value of the parameters

Lemma 4. *For the conditional covariance $\mathbf{P}_{t|t-1}$ of the static factors given \mathcal{X}_{t-1} , there exists a steady state for the reduced form denoted as \mathbf{P} solving the algebraic Riccati equation (ARE) and such that*

$$\mathbf{P}_{t|t-1} = \mathbf{P} + O(e^{-t}).$$

Moreover, as $n \rightarrow \infty$,

$$\mathbf{P} = \mathbf{K} \begin{pmatrix} \mathbf{I}_q & \mathbf{0}_q \\ \mathbf{0}_q & \mathbf{0}_q \end{pmatrix} \mathbf{K}' + O(n^{-1}) = \mathbf{H}\mathbf{H}' + O(n^{-1}).$$

Proof. Define $\tilde{\mathbf{P}}_{t|t-1}$ as the conditional covariance matrix for the vector $(\mathbf{f}'_t \mathbf{f}'_{t-1})'$ given \mathcal{X}_{t-1} . Then, due to stabilizability and detectability proved in Lemma 1, there exists a steady state for the structural model denoted as $\tilde{\mathbf{P}}$ solving the algebraic Riccati equation (ARE) and such that (see Anderson and Moore, 1979, pp.76-77, and Harvey, 1990, pp.118-119)

$$\tilde{\mathbf{P}}_{t|t-1} = \tilde{\mathbf{P}} + O(e^{-t}).$$

In presence of a diffuse prior its effect is limited to the first few periods, say t_0 (see Koopman, 1997), then the result above holds for $t > t_0$. The ARE for the structural model is then (see also (B19))

$$\tilde{\mathbf{P}} - \tilde{\mathbf{A}}\tilde{\mathbf{P}}\tilde{\mathbf{A}}' + \tilde{\mathbf{A}}\tilde{\mathbf{P}}\tilde{\mathbf{B}}'(\tilde{\mathbf{B}}\tilde{\mathbf{P}}\tilde{\mathbf{B}}' + \mathbf{R})^{-1}\tilde{\mathbf{B}}\tilde{\mathbf{P}}\tilde{\mathbf{A}}' = \begin{pmatrix} \mathbf{I}_q & \mathbf{0}_q \\ \mathbf{0}_q & \mathbf{0}_q \end{pmatrix}, \quad (\text{C36})$$

where

$$\tilde{\mathbf{A}} = \begin{pmatrix} \mathbf{\Pi}_1 & \mathbf{\Pi}_2 \\ \mathbf{I}_q & \mathbf{0}_q \end{pmatrix}, \quad \tilde{\mathbf{B}} = (\mathbf{B}_0 \mathbf{B}_1). \quad (\text{C37})$$

Now since the structural model has only q controllable and observable states (see Lemma 1) and $\tilde{\mathbf{P}}$ is the steady state covariance of those states, $\text{rk}(\tilde{\mathbf{P}}) = q$. Define as \mathbf{V} the $r \times r$ matrix of eigenvectors of $\tilde{\mathbf{P}}$ and as \mathbf{D} the $q \times q$ diagonal matrix of its non zero eigenvalues, then

$$\tilde{\mathbf{P}} = \mathbf{V} \begin{pmatrix} \mathbf{D} & \mathbf{0}_q \\ \mathbf{0}_q & \mathbf{0}_q \end{pmatrix} \mathbf{V}' = \mathbf{V} \begin{pmatrix} \mathbf{D}^{1/2} & \mathbf{0}_q \\ \mathbf{0}_q & \mathbf{0}_q \end{pmatrix} \begin{pmatrix} \mathbf{D}^{1/2} & \mathbf{0}_q \\ \mathbf{0}_q & \mathbf{0}_q \end{pmatrix} \mathbf{V}' = \mathbf{W} \begin{pmatrix} \mathbf{I}_q & \mathbf{0}_q \\ \mathbf{0}_q & \mathbf{0}_q \end{pmatrix} \mathbf{W}', \quad (\text{C38})$$

with

$$\mathbf{W} = \mathbf{V} \begin{pmatrix} \mathbf{D}^{1/2} & \mathbf{0}_q \\ \mathbf{0}_q & \mathbf{I}_q \end{pmatrix}.$$

Define \mathbf{B}_0^* and \mathbf{B}_1^* as the $n \times q$ matrices such that $\tilde{\mathbf{B}}\mathbf{W} = (\mathbf{B}_0^* \mathbf{B}_1^*)$. Then, from (C38)

$$\tilde{\mathbf{B}}\tilde{\mathbf{P}}\tilde{\mathbf{B}}' = \tilde{\mathbf{B}}\mathbf{W} \begin{pmatrix} \mathbf{I}_q & \mathbf{0}_q \\ \mathbf{0}_q & \mathbf{0}_q \end{pmatrix} \mathbf{W}'\tilde{\mathbf{B}}' = (\mathbf{B}_0^* \mathbf{B}_1^*) \begin{pmatrix} \mathbf{I}_q & \mathbf{0}_q \\ \mathbf{0}_q & \mathbf{0}_q \end{pmatrix} \begin{pmatrix} \mathbf{B}_0^{*'} \\ \mathbf{B}_1^{*'} \end{pmatrix} = \mathbf{B}_0^* \mathbf{B}_0^{*'}, \quad (\text{C39})$$

and

$$\begin{pmatrix} \mathbf{I}_q & \mathbf{0}_q \\ \mathbf{0}_q & \mathbf{0}_q \end{pmatrix} \mathbf{W}'\tilde{\mathbf{B}}' = \begin{pmatrix} \mathbf{I}_q & \mathbf{0}_q \\ \mathbf{0}_q & \mathbf{0}_q \end{pmatrix} \begin{pmatrix} \mathbf{B}_0^{*'} \\ \mathbf{B}_1^{*'} \end{pmatrix} = \begin{pmatrix} \mathbf{B}_0^{*'} \\ \mathbf{0}_{q \times n} \end{pmatrix}. \quad (\text{C40})$$

From (C38), (C39), (C40), Lemmas 2 and 3, we have

$$\begin{aligned} \begin{pmatrix} \mathbf{I}_q & \mathbf{0}_q \\ \mathbf{0}_q & \mathbf{0}_q \end{pmatrix} \mathbf{W}'\tilde{\mathbf{B}}'(\tilde{\mathbf{B}}\tilde{\mathbf{P}}\tilde{\mathbf{B}}' + \mathbf{R})^{-1}\tilde{\mathbf{B}}\mathbf{W} \begin{pmatrix} \mathbf{I}_q & \mathbf{0}_q \\ \mathbf{0}_q & \mathbf{0}_q \end{pmatrix} &= \begin{pmatrix} \mathbf{B}_0^{*'}(\mathbf{B}_0^* \mathbf{B}_0^{*'} + \mathbf{R})^{-1} \\ \mathbf{0}_{q \times n} \end{pmatrix} (\mathbf{B}_0^* \mathbf{B}_1^*) \begin{pmatrix} \mathbf{I}_q & \mathbf{0}_q \\ \mathbf{0}_q & \mathbf{0}_q \end{pmatrix} \\ &= \begin{pmatrix} (\mathbf{B}_0^{*'} \mathbf{R}^{-1} \mathbf{B}_0^* + \mathbf{I}_q)^{-1} \mathbf{B}_0^{*'} \mathbf{R}^{-1} \mathbf{B}_0^* & \mathbf{0}_q \\ \mathbf{0}_q & \mathbf{0}_q \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{I}_q + O(n^{-1}) & \mathbf{0}_q \\ \mathbf{0}_q & \mathbf{0}_q \end{pmatrix}. \end{aligned} \quad (\text{C41})$$

Notice that we can apply Lemma 3 to the top left $q \times q$ block of (C41) since: \mathbf{I}_q trivially satisfies condition (a), \mathbf{R}^{-1} satisfies condition (b) because of D2 and $\mathbf{B}_0^{*'} \mathbf{B}_0^*$ satisfies condition (c). Indeed, from definition (A12) we have

$$\frac{1}{n} \begin{pmatrix} \mathbf{B}_0^{*'} \\ \mathbf{B}_1^{*'} \end{pmatrix} (\mathbf{B}_0^* \mathbf{B}_1^*) = \mathbf{W}' \mathbf{K}' \frac{\mathbf{\Lambda}' \mathbf{\Lambda}}{n} \mathbf{K} \mathbf{W},$$

and because of assumption C2 the top left $q \times q$ block of this matrix which is $n^{-1} \mathbf{B}_0^{*'} \mathbf{B}_0^*$ has full column-rank for any n . By substituting (C38) and (C41) into (C36) we have

$$\tilde{\mathbf{P}} = \begin{pmatrix} \mathbf{I}_q & \mathbf{0}_q \\ \mathbf{0}_q & \mathbf{0}_q \end{pmatrix} + O(n^{-1}). \quad (\text{C42})$$

Now, notice that by construction $\mathbf{P}_{t|t-1} = \mathbf{K}\tilde{\mathbf{P}}_{t|t-1}\mathbf{K}'$ and since \mathbf{K} is full-rank then also the reduced form system is stabilizable and detectable, thus it has a steady state \mathbf{P} such that

$$\mathbf{P}_{t|t-1} = \mathbf{P} + O(e^{-t}). \quad (\text{C43})$$

Moreover, since \mathbf{K} does not depend on t nor n , we have $\mathbf{P} = \mathbf{K}\tilde{\mathbf{P}}\mathbf{K}'$ and the result follows directly from (C42). Last, from the definition of \mathbf{H} in (A13) we have also $\mathbf{P} = \mathbf{H}\mathbf{H}'$. \square

Lemma 5. *For the static factors estimated via KF and KS when using the true value of the parameters Θ , under condition (15) in the text, the following hold, for all $\bar{t} \leq t \leq T$ and as $n \rightarrow \infty$,*

$$\begin{aligned} \sqrt{n} \|\mathbf{F}_{t|\bar{t}} - \mathbf{F}_t\| &= O_p(1), \\ \sqrt{n} \|\mathbf{F}_{t|T} - \mathbf{F}_t\| &= O_p(1). \end{aligned}$$

Proof. By Lemma 4, the conditional covariance $\mathbf{P}_{t|t}$ of the static factors given \mathcal{X}_t has a steady state \mathbf{S} such that (see (B18))

$$\mathbf{S} = \mathbf{P} - \mathbf{P}\mathbf{\Lambda}'(\mathbf{\Lambda}\mathbf{P}\mathbf{\Lambda}' + \mathbf{R})^{-1}\mathbf{\Lambda}\mathbf{P}. \quad (\text{C44})$$

Then, notice that by Lemma 4 and (A12)

$$\mathbf{P}\mathbf{\Lambda}' = \mathbf{K} \begin{pmatrix} \mathbf{I}_q & \mathbf{0}_q \\ \mathbf{0}_q & \mathbf{0}_q \end{pmatrix} \mathbf{K}'(\mathbf{K}')^{-1} \begin{pmatrix} \mathbf{B}'_0 \\ \mathbf{B}'_1 \end{pmatrix} + O(n^{-1}) = \mathbf{K} \begin{pmatrix} \mathbf{B}'_0 \\ \mathbf{0}_q \end{pmatrix} + O(n^{-1}), \quad (\text{C45})$$

$$\mathbf{\Lambda}\mathbf{P}\mathbf{\Lambda}' = (\mathbf{B}_0 \mathbf{B}_1) \mathbf{K}^{-1} \mathbf{K} \begin{pmatrix} \mathbf{I}_q & \mathbf{0}_q \\ \mathbf{0}_q & \mathbf{0}_q \end{pmatrix} \mathbf{K}'(\mathbf{K}')^{-1} \begin{pmatrix} \mathbf{B}'_0 \\ \mathbf{B}'_1 \end{pmatrix} + O(n^{-1}) = \mathbf{B}_0 \mathbf{B}'_0 + O(n^{-1}). \quad (\text{C46})$$

Using (C45) and (C46) and by applying Lemmas 2 and 3 we have

$$\begin{aligned} \mathbf{P}\mathbf{\Lambda}'(\mathbf{\Lambda}\mathbf{P}\mathbf{\Lambda}' + \mathbf{R})^{-1}\mathbf{\Lambda}\mathbf{P} &= \mathbf{K} \begin{pmatrix} \mathbf{B}'_0(\mathbf{B}_0 \mathbf{B}'_0 + \mathbf{R})^{-1} \mathbf{B}_0 & \mathbf{0}_q \\ \mathbf{0}_q & \mathbf{0}_q \end{pmatrix} \mathbf{K}' + O(n^{-1}) \\ &= \mathbf{K} \begin{pmatrix} (\mathbf{B}'_0 \mathbf{R}^{-1} \mathbf{B}_0 + \mathbf{I}_q)^{-1} \mathbf{B}'_0 \mathbf{R}^{-1} \mathbf{B}_0 & \mathbf{0}_q \\ \mathbf{0}_q & \mathbf{0}_q \end{pmatrix} \mathbf{K}' + O(n^{-1}) \\ &= \mathbf{K} \begin{pmatrix} \mathbf{I}_q + O(n^{-1}) & \mathbf{0}_q \\ \mathbf{0}_q & \mathbf{0}_q \end{pmatrix} \mathbf{K}' + O(n^{-1}). \end{aligned} \quad (\text{C47})$$

Notice that we can apply Lemma 3 to the top $q \times q$ block of (C47) since: \mathbf{I}_q trivially satisfies condition (a), \mathbf{R}^{-1} satisfies condition (b) because of D2 and $\mathbf{B}_0' \mathbf{B}_0$ satisfies condition (c) because of assumption C2 and definition (A12) (see also (C41) in the proof of Lemma 4). By substituting (C47) into (C44) and because of Lemma 4, we have

$$\mathbf{S} = \mathbf{K} \begin{pmatrix} \mathbf{I}_q & \mathbf{0}_q \\ \mathbf{0}_q & \mathbf{0}_q \end{pmatrix} \mathbf{K}' + O(n^{-1}) - \mathbf{K} \begin{pmatrix} \mathbf{I}_q & \mathbf{0}_q \\ \mathbf{0}_q & \mathbf{0}_q \end{pmatrix} \mathbf{K}' + O(n^{-1}) = O(n^{-1}). \quad (\text{C48})$$

By substituting (C43) in (B18), from (C44) we have

$$\mathbf{P}_{t|t} = \mathbf{S} + O(e^{-t}). \quad (\text{C49})$$

Therefore, by substituting (C48) into (C49) and letting $n = T^\gamma$ for $\gamma > 0$ and $\bar{t} \equiv \bar{t}(T)$, because of (15) for $\bar{t} \leq t \leq T$ we have

$$\mathbf{P}_{t|t} = O(n^{-1}) + O(e^{-t}) = O(n^{-1}). \quad (\text{C50})$$

Now, let us consider $\mathbf{P}_{t|T}$ defined in (B24). From (B18)

$$\mathbf{P}_{t|t-1} = \mathbf{P}_{t|t} + \mathbf{P}_{t|t-1} \mathbf{\Lambda}'(\mathbf{\Lambda}\mathbf{P}_{t|t-1} \mathbf{\Lambda}' + \mathbf{R})^{-1} \mathbf{\Lambda}\mathbf{P}_{t|t-1}. \quad (\text{C51})$$

By substituting (C51) and (B25) in (B24) we have

$$\mathbf{P}_{t|T} = \mathbf{P}_{t|t} + \mathbf{P}_{t|t-1} \mathbf{L}'_t \mathbf{N}_t \mathbf{L}_t \mathbf{P}_{t|t-1}. \quad (\text{C52})$$

Since \mathbf{N}_t is function of $\mathbf{P}_{t|t-1}$, because of Lemma 4, it has a steady state \mathbf{N} such that $\|\mathbf{N}\| = O(1)$ and

$$\mathbf{N}_t = \mathbf{N} + O(e^{-t}). \quad (\text{C53})$$

Now, since $\mathbf{A}\mathbf{P}_{t|t} = \mathbf{L}_t\mathbf{P}_{t|t-1}$, using (C50) and (C53), because of (15) for $\bar{t} \leq t \leq T$ we have

$$\mathbf{P}_{t|t-1}\mathbf{L}'_t\mathbf{N}_t\mathbf{L}_t\mathbf{P}_{t|t-1} = \mathbf{P}_{t|t}\mathbf{A}'\mathbf{N}_t\mathbf{A}\mathbf{P}_{t|t} = O(n^{-2}). \quad (\text{C54})$$

By using (C50) and (C54) into (C52), for $\bar{t} \leq t \leq T$ we have

$$\mathbf{P}_{t|T} = O(n^{-1}). \quad (\text{C55})$$

By the law of iterated expectations, for $\bar{t} \leq t \leq T$ we have (see also the definitions in (B14))

$$\begin{aligned} \mathbf{E}_\Theta[(\mathbf{F}_t - \mathbf{F}_{t|t})(\mathbf{F}_t - \mathbf{F}_{t|t})'] &= \mathbf{E}_\Theta[\mathbf{E}_\Theta[(\mathbf{F}_t - \mathbf{F}_{t|t})(\mathbf{F}_t - \mathbf{F}_{t|t})' | \mathcal{X}_t]] = \mathbf{E}_\Theta[\mathbf{P}_{t|t}] = O(n^{-1}), \\ \mathbf{E}_\Theta[(\mathbf{F}_t - \mathbf{F}_{t|T})(\mathbf{F}_t - \mathbf{F}_{t|T})'] &= \mathbf{E}_\Theta[\mathbf{E}_\Theta[(\mathbf{F}_t - \mathbf{F}_{t|T})(\mathbf{F}_t - \mathbf{F}_{t|T})' | \mathcal{X}_T]] = \mathbf{E}_\Theta[\mathbf{P}_{t|T}] = O(n^{-1}), \end{aligned}$$

which imply mean-square convergence of the KF and KS when the parameters are known and for all $\bar{t} \leq t \leq T$:

$$\begin{aligned} \mathbf{E}_\Theta[\|\mathbf{F}_t - \mathbf{F}_{t|t}\|^2] &= \sum_{j=1}^r \mathbf{E}_\Theta[(F_{j,t} - F_{j,t|t})^2] = \text{tr}\{\mathbf{E}_\Theta[\mathbf{P}_{t|t}]\} = O(n^{-1}), \\ \mathbf{E}_\Theta[\|\mathbf{F}_t - \mathbf{F}_{t|T}\|^2] &= \sum_{j=1}^r \mathbf{E}_\Theta[(F_{j,t} - F_{j,t|T})^2] = \text{tr}\{\mathbf{E}_\Theta[\mathbf{P}_{t|T}]\} = O(n^{-1}). \end{aligned}$$

The result follows from Chebychev's inequality. \square

C.3 Consistency of KF and KS using estimated parameters

Lemma 6. Consider the QML estimator of the parameters $\hat{\Theta}^*$ defined in (B7) and obtained using the true values of the static factors \mathbf{F}_t , then, as $T \rightarrow \infty$:

$$\begin{aligned} \sqrt{T} \|\hat{\boldsymbol{\lambda}}_i^* - \boldsymbol{\lambda}_i\| &= O_p(1), \quad i = 1, \dots, n, \\ \sqrt{T} \|\hat{\mathbf{A}}^* - \mathbf{A}\| &= O_p(1), \\ \sqrt{T} \|\hat{\mathbf{H}}^* - \mathbf{H}\| &= O_p(1), \\ \sqrt{T} |[\hat{\mathbf{R}}^*]_{ii} - [\mathbf{R}]_{ii}| &= O_p(1), \quad i = 1, \dots, n. \end{aligned}$$

Proof. The QML estimator of the loadings, for any $i = 1, \dots, n$, is given by

$$\hat{\boldsymbol{\lambda}}_i^{*'} = \left(\sum_{t=1}^T x_{it} \mathbf{F}'_t \right) \left(\sum_{t=1}^T \mathbf{F}_t \mathbf{F}'_t \right)^{-1}. \quad (\text{C56})$$

We know that \mathbf{F}_t is driven by $(q-d)$ common trends (see C2), therefore we can find an orthonormal linear basis of dimension $(q-d)$ such that the projection of \mathbf{F}_t onto this basis span the same space as the common trends. Collect the elements of this basis in the $r \times (q-d)$ matrix $\boldsymbol{\gamma}$, and denote as $\boldsymbol{\gamma}_\perp$ the $r \times (r-q+d)$ matrix such that $\boldsymbol{\gamma}'_\perp \boldsymbol{\gamma} = \mathbf{0}_{(r-q+d) \times (q-d)}$. Then, consider the $r \times r$ linear transformation

$$\mathcal{D}\mathbf{F}_t = \begin{pmatrix} \boldsymbol{\gamma}'\mathbf{F}_t \\ \boldsymbol{\gamma}'_\perp\mathbf{F}_t \end{pmatrix} = \begin{pmatrix} \mathbf{Z}_{1t} \\ \mathbf{Z}_{0t} \end{pmatrix}, \quad (\text{C57})$$

where \mathbf{Z}_{1t} has all $(q-d)$ components which are $I(1)$ while $\mathbf{Z}_{0t} \sim I(0)$ and is of dimension $(r-q+d)$. Moreover, for \mathbf{Z}_{1t} we have the MA representation

$$\Delta \mathbf{Z}_{1t} = \mathbf{Q}(L)\zeta_t, \quad (\text{C58})$$

$\zeta_t \stackrel{w.n.}{\sim} (\mathbf{0}_{q-d}, \Sigma_\zeta)$ with Σ_ζ positive definite and $\mathbf{Q}(L)$ is a $(q-d) \times (q-d)$ one-sided, infinite matrix polynomial with square-summable coefficients and $\text{rk}(\mathbf{Q}(1)) = (q-d)$.

Because of orthonormality $\mathcal{D}'\mathcal{D} = \mathbf{I}_r$. Then, the corresponding transformation of the loadings gives $\lambda'_i \mathcal{D}' = (\lambda'_{i1} \ \lambda'_{i0})$ such that $x_{it} = \lambda'_{i1} \mathbf{Z}_{1t} + \lambda'_{i0} \mathbf{Z}_{0t} + \xi_{it}$ and we also have $\widehat{\lambda}'_i \mathcal{D}' = (\widehat{\lambda}'_{i1} \ \widehat{\lambda}'_{i0})$. Recall that \mathbf{Z}_{1t} and \mathbf{Z}_{0t} are orthogonal by construction, then we have

$$\begin{aligned} & \begin{pmatrix} \widehat{\lambda}'_{i1} - \lambda'_{i1} \\ \widehat{\lambda}'_{i0} - \lambda'_{i0} \end{pmatrix} \\ &= \begin{pmatrix} \left(\frac{1}{T^2} \sum_{t=1}^T \xi_{it} \mathbf{Z}'_{1t} \right) \left(\frac{1}{T^2} \sum_{t=1}^T \mathbf{Z}_{1t} \mathbf{Z}'_{1t} \right)^{-1} & \mathbf{0}_{(q-d) \times (r-q+d)} \\ \mathbf{0}_{(r-q+d) \times (q-d)} & \left(\frac{1}{T} \sum_{t=1}^T \xi_{it} \mathbf{Z}'_{0t} \right) \left(\frac{1}{T} \sum_{t=1}^T \mathbf{Z}_{0t} \mathbf{Z}'_{0t} \right)^{-1} \end{pmatrix}. \end{aligned} \quad (\text{C59})$$

By Theorem 1 in Peña and Poncela (1997, 2006), under C1 and C3, and from (C58), as $T \rightarrow \infty$,

$$\frac{1}{T^2} \sum_{t=1}^T \mathbf{Z}_{1t} \mathbf{Z}'_{1t} \Rightarrow \mathbf{Q}(1) \Sigma_\zeta^{1/2} \left(\int_0^1 \mathcal{W}(u) \mathcal{W}(u)' du \right) \Sigma_\zeta^{1/2} \mathbf{Q}(1)', \quad (\text{C60})$$

where $\mathcal{W}(\cdot)$ is a $(q-d)$ -dimensional standard Wiener process. Thus this term is $O_p(1)$ and positive definite therefore invertible. Last, from (C58) we see that each component of $\Delta \mathbf{Z}_{1t}$ has an MA representation with square summable coefficients ($\Delta \mathbf{Z}_{1t} \sim I(0)$ by construction), therefore $\text{Var}(t^{-1} Z_{1jt}) = O(1)$ for any $j = 1, \dots, (q-d)$ and any $t = 1, \dots, T$. Thus, by using Gaussianity (see D2 and D3) and by A2 also independence of factors and idiosyncratic components, we can prove that, as $T \rightarrow \infty$,

$$\frac{1}{T^2} \sum_{t=1}^T \xi_{it} \mathbf{Z}'_{1t} = O_p(T^{-1}). \quad (\text{C61})$$

Moreover, from C1 it is easy to see that \mathbf{Z}_{0t} has an MA representation with square summable coefficients (it is stationary) and because of A2 and C3 we have, as $T \rightarrow \infty$,

$$\frac{1}{T} \sum_{t=1}^T \xi_{it} \mathbf{Z}'_{0t} = O_p(T^{-1/2}), \quad \frac{1}{T} \sum_{t=1}^T \mathbf{Z}_{0t} \mathbf{Z}'_{0t} = \mathbb{E}[\mathbf{Z}_{0t} \mathbf{Z}'_{0t}] + O_p(T^{-1/2}) = O_p(1). \quad (\text{C62})$$

From (C59), (C60), (C61), and (C62), and since \mathcal{D} does not depend on T , we obtain the result.

Consider the VAR

$$\mathcal{D}\mathbf{F}_t = (\mathcal{D}\mathbf{A}\mathcal{D}')\mathcal{D}\mathbf{F}_{t-1} + \mathcal{D}\mathbf{H}\mathbf{u}_t. \quad (\text{C63})$$

such that $\mathcal{D}\mathbf{H}\mathbf{u}_t = (\mathbf{e}'_{1t} \ \mathbf{e}'_{0t})'$ where \mathbf{e}_{1t} and \mathbf{e}_{0t} are white noise processes of dimensions $(q-d)$

and $(r - q + d)$, respectively. Then, similarly to (C59) we have

$$\begin{aligned} \mathcal{D}(\widehat{\mathbf{A}}^* - \mathbf{A})\mathcal{D}' &= \tag{C64} \\ &= \begin{pmatrix} \left(\frac{1}{T^2} \sum_{t=1}^T \mathbf{e}_{1t} \mathbf{Z}'_{1t-1} \right) \left(\frac{1}{T^2} \sum_{t=1}^T \mathbf{Z}_{1t-1} \mathbf{Z}'_{1t-1} \right)^{-1} & \left(\frac{1}{T} \sum_{t=1}^T \mathbf{e}_{0t} \mathbf{Z}'_{1t-1} \right) \left(\frac{1}{T} \sum_{t=1}^T \mathbf{Z}_{0t-1} \mathbf{Z}'_{0t-1} \right)^{-1} \\ \left(\frac{1}{T^2} \sum_{t=1}^T \mathbf{e}_{1t} \mathbf{Z}'_{0t-1} \right) \left(\frac{1}{T^2} \sum_{t=1}^T \mathbf{Z}_{1t-1} \mathbf{Z}'_{1t-1} \right)^{-1} & \left(\frac{1}{T} \sum_{t=1}^T \mathbf{e}_{0t} \mathbf{Z}'_{0t-1} \right) \left(\frac{1}{T} \sum_{t=1}^T \mathbf{Z}_{0t-1} \mathbf{Z}'_{0t-1} \right)^{-1} \end{pmatrix} \end{aligned}$$

Then, using the fact that \mathbf{e}_{1t} and \mathbf{e}_{0t} are white noise, it can be shown that

$$\begin{aligned} \frac{1}{T^2} \sum_{t=1}^T \mathbf{e}_{1t} \mathbf{Z}'_{1t-1} &= O_p(T^{-1}), & \frac{1}{T^2} \sum_{t=1}^T \mathbf{e}_{1t} \mathbf{Z}'_{0t-1} &= O_p(T^{-1/2}), \\ \frac{1}{T} \sum_{t=1}^T \mathbf{e}_{1t} \mathbf{Z}'_{0t-1} &= O_p(T^{-1}), & \frac{1}{T} \sum_{t=1}^T \mathbf{e}_{0t} \mathbf{Z}'_{0t-1} &= O_p(T^{-1/2}). \end{aligned} \tag{C65}$$

Substituting (C60), (C62), and (C65) into (C64) and since \mathcal{D} does not depend on T , we have the result for the VAR parameters. Similar results can be proved for all other parameters. \square

Lemma 7. *Define the KF and KS estimators of the static factors and their conditional covariances when using the QML estimator of the parameters $\widehat{\Theta}^*$ as*

$$\begin{aligned} \mathbf{F}_{t|t}^* &= \mathbb{E}_{\widehat{\Theta}^*}[\mathbf{F}_t | \mathcal{X}_t], & \mathbf{P}_{t|t}^* &= \mathbb{E}_{\widehat{\Theta}^*}[(\mathbf{F}_t - \mathbf{F}_{t|t}^*)(\mathbf{F}_t - \mathbf{F}_{t|t}^*)' | \mathcal{X}_t], \\ \mathbf{F}_{t|T}^* &= \mathbb{E}_{\widehat{\Theta}^*}[\mathbf{F}_t | \mathcal{X}_T], & \mathbf{P}_{t|T}^* &= \mathbb{E}_{\widehat{\Theta}^*}[(\mathbf{F}_t - \mathbf{F}_{t|T}^*)(\mathbf{F}_t - \mathbf{F}_{t|T}^*)' | \mathcal{X}_T]. \end{aligned}$$

Then, under condition (15) in the text, the following hold, for all $\bar{t} \leq t \leq T$ and as $n, T \rightarrow \infty$,

$$\begin{aligned} \min(\sqrt{n}, \sqrt{T}) \|\mathbf{F}_{t|t}^* - \mathbf{F}_t\| &= O_p(1), \\ \min(\sqrt{n}, \sqrt{T}) \|\mathbf{F}_{t|T}^* - \mathbf{F}_t\| &= O_p(1), \\ \min(n, \sqrt{T}) \|\mathbf{P}_{t|t}^*\| &= O_p(1), \\ \min(n, \sqrt{T}) \|\mathbf{P}_{t|T}^*\| &= O_p(1). \end{aligned}$$

Proof. We start with three preliminary results. First, from (C50) in the proof of Lemma 5 we have

$$\mathbf{F}_{t-1|t-1} - \mathbf{F}_{t-1} = O(e^{-(t-1)/2}) + O(n^{-1/2}). \tag{C66}$$

Second, because of Lemma 5 and (C66), we have

$$\begin{aligned} \left\| \frac{\mathbf{x}_t}{\sqrt{n}} - \frac{\Lambda \mathbf{F}_{t|t-1}}{\sqrt{n}} \right\| &= \left\| \frac{\Lambda \mathbf{F}_t + \boldsymbol{\xi}_t}{\sqrt{n}} - \frac{\Lambda \mathbf{F}_{t|t-1}}{\sqrt{n}} \right\| \leq \left\| \frac{\Lambda}{\sqrt{n}} \right\| (\|\Lambda\| \|\mathbf{F}_{t-1} - \mathbf{F}_{t-1|t-1}\| + \|\mathbf{H}\mathbf{u}_t\|) + \left\| \frac{\boldsymbol{\xi}_t}{\sqrt{n}} \right\| \\ &= \left\| \frac{\Lambda}{\sqrt{n}} \right\| \left[\|\Lambda\| \left(O(e^{-(t-1)/2}) + O(n^{-1/2}) \right) + \|\mathbf{H}\mathbf{u}_t\| \right] + \left\| \frac{\boldsymbol{\xi}_t}{\sqrt{n}} \right\| = O_p(1), \end{aligned} \tag{C67}$$

since $\mathbf{u}_t \stackrel{w.n.}{\sim} (\mathbf{0}_q, \mathbf{I}_q)$ and $\boldsymbol{\xi}_t \sim I(0)$.

Third, from transformation (C57), defined in the proof of Lemma 6, $\mathcal{D}\mathbf{F}_t = (\mathbf{Z}'_{1t} \mathbf{Z}'_{0t})'$, we have $\|\mathbf{Z}_{1t}\| = O_p(\sqrt{T})$ and $\|\mathbf{Z}_{0t}\| = O_p(1)$. Then, since $\mathcal{D}'\mathcal{D} = \mathbf{I}_r$, as a consequence of Lemma

6 (see in particular (C59) and (C64)), the following hold

$$(\widehat{\mathbf{A}}^* - \mathbf{A})\mathbf{F}_t = \mathcal{D}'\mathcal{D}(\widehat{\mathbf{A}}^* - \mathbf{A})\mathcal{D}'\mathcal{D}\mathbf{F}_t = O_p(T^{-1/2}), \quad (\text{C68})$$

$$(\widehat{\lambda}_i^* - \lambda_i^*)\mathbf{A}\mathbf{F}_t = (\widehat{\lambda}_i^* - \lambda_i^*)\mathcal{D}'\mathcal{D}\mathbf{A}\mathcal{D}'\mathcal{D}\mathbf{F}_t = O_p(T^{-1/2}), \quad i = 1, \dots, n. \quad (\text{C69})$$

Now, we compare the KF iterations, (B15)-(B18), with those obtained when using $\widehat{\Theta}^*$:

$$\mathbf{F}_{t|t-1}^* = \widehat{\mathbf{A}}^*\mathbf{F}_{t-1|t-1}^*, \quad (\text{C70})$$

$$\mathbf{P}_{t|t-1}^* = \widehat{\mathbf{A}}^*\mathbf{P}_{t-1|t-1}^*\widehat{\mathbf{A}}^{*\prime} + \widehat{\mathbf{H}}^*\widehat{\mathbf{H}}^{*\prime}, \quad (\text{C71})$$

$$\mathbf{F}_{t|t}^* = \mathbf{F}_{t|t-1}^* + \mathbf{P}_{t|t-1}^*\widehat{\Lambda}^{*\prime}(\widehat{\Lambda}^*\mathbf{P}_{t|t-1}^*\widehat{\Lambda}^{*\prime} + \widehat{\mathbf{R}}^*)^{-1}(\mathbf{x}_t - \widehat{\Lambda}^*\mathbf{F}_{t|t-1}^*), \quad (\text{C72})$$

$$\mathbf{P}_{t|t}^* = \mathbf{P}_{t|t-1}^* - \mathbf{P}_{t|t-1}^*\widehat{\Lambda}^{*\prime}(\widehat{\Lambda}^*\mathbf{P}_{t|t-1}^*\widehat{\Lambda}^{*\prime} + \widehat{\mathbf{R}}^*)^{-1}\widehat{\Lambda}^*\mathbf{P}_{t|t-1}^*. \quad (\text{C73})$$

From (C70) we have

$$\begin{aligned} \mathbf{F}_{t|t-1}^* - \mathbf{F}_{t|t-1} &= \mathbf{A}(\mathbf{F}_{t-1|t-1}^* - \mathbf{F}_{t-1|t-1}) + (\widehat{\mathbf{A}}^* - \mathbf{A})(\mathbf{F}_{t-1|t-1}^* - \mathbf{F}_{t-1|t-1}) + (\widehat{\mathbf{A}}^* - \mathbf{A})\mathbf{F}_{t-1|t-1} \\ &= \mathbf{A}(\mathbf{F}_{t-1|t-1}^* - \mathbf{F}_{t-1|t-1}) + (\widehat{\mathbf{A}}^* - \mathbf{A})(\mathbf{F}_{t-1|t-1}^* - \mathbf{F}_{t-1|t-1}) + O_p(T^{-1/2}), \end{aligned} \quad (\text{C74})$$

since

$$\begin{aligned} (\widehat{\mathbf{A}}^* - \mathbf{A})\mathbf{F}_{t-1|t-1} &= (\widehat{\mathbf{A}}^* - \mathbf{A})(\mathbf{F}_{t-1|t-1} - \mathbf{F}_{t-1}) + (\widehat{\mathbf{A}}^* - \mathbf{A})\mathbf{F}_{t-1} \\ &= O_p(T^{-1/2})O(e^{-(t-1)/2}) + O_p(T^{-1/2})O(n^{-1/2}) + O_p(T^{-1/2}), \end{aligned}$$

because of Lemma 6, (C66) and (C68). Similarly, from (C71) we have

$$\begin{aligned} \mathbf{P}_{t|t-1}^* - \mathbf{P}_{t|t-1} &= \mathbf{A}(\mathbf{P}_{t-1|t-1}^* - \mathbf{P}_{t-1|t-1})\mathbf{A}' + (\widehat{\mathbf{A}}^* - \mathbf{A})(\mathbf{P}_{t-1|t-1}^* - \mathbf{P}_{t-1|t-1})\mathbf{A}' \quad (\text{C75}) \\ &\quad + (\widehat{\mathbf{A}}^* - \mathbf{A})(\mathbf{P}_{t-1|t-1}^* - \mathbf{P}_{t-1|t-1})(\widehat{\mathbf{A}}^* - \mathbf{A})' + \mathbf{A}(\mathbf{P}_{t-1|t-1}^* - \mathbf{P}_{t-1|t-1})(\widehat{\mathbf{A}}^* - \mathbf{A})' \\ &\quad + (\widehat{\mathbf{A}}^* - \mathbf{A})\mathbf{P}_{t-1|t-1}\mathbf{A}' + (\widehat{\mathbf{A}}^* - \mathbf{A})\mathbf{P}_{t-1|t-1}(\widehat{\mathbf{A}}^* - \mathbf{A})' + \mathbf{A}\mathbf{P}_{t-1|t-1}(\widehat{\mathbf{A}}^* - \mathbf{A})' \\ &\quad + (\widehat{\mathbf{H}}^* - \mathbf{H})\mathbf{H}' + (\widehat{\mathbf{H}}^* - \mathbf{H})(\widehat{\mathbf{H}}^* - \mathbf{H})' + \mathbf{H}(\widehat{\mathbf{H}}^* - \mathbf{H})' \\ &= \mathbf{A}(\mathbf{P}_{t-1|t-1}^* - \mathbf{P}_{t-1|t-1})\mathbf{A}' + (\widehat{\mathbf{A}}^* - \mathbf{A})(\mathbf{P}_{t-1|t-1}^* - \mathbf{P}_{t-1|t-1})\mathbf{A}' \\ &\quad + (\widehat{\mathbf{A}}^* - \mathbf{A})(\mathbf{P}_{t-1|t-1}^* - \mathbf{P}_{t-1|t-1})(\widehat{\mathbf{A}}^* - \mathbf{A})' + \mathbf{A}(\mathbf{P}_{t-1|t-1}^* - \mathbf{P}_{t-1|t-1})(\widehat{\mathbf{A}}^* - \mathbf{A})' + O_p(T^{-1/2}), \end{aligned}$$

since

$$\begin{aligned} (\widehat{\mathbf{A}}^* - \mathbf{A})\mathbf{P}_{t-1|t-1}\mathbf{A}' &= \mathbf{A}\mathbf{P}_{t-1|t-1}(\widehat{\mathbf{A}}^* - \mathbf{A})' = O_p(T^{-1/2})O(e^{-(t-1)}) + O_p(T^{-1/2})O(n^{-1}), \\ (\widehat{\mathbf{A}}^* - \mathbf{A})\mathbf{P}_{t-1|t-1}(\widehat{\mathbf{A}}^* - \mathbf{A})' &= O_p(T^{-1})O(e^{-(t-1)}) + O_p(T^{-1})O(n^{-1}). \end{aligned}$$

because of Lemma 6 and (C50) in the proof of Lemma 5 and

$$\widehat{\mathbf{H}}^*\widehat{\mathbf{H}}^{*\prime} - \mathbf{H}\mathbf{H}' = (\widehat{\mathbf{H}}^* - \mathbf{H})\mathbf{H}' + \mathbf{H}(\widehat{\mathbf{H}}^* - \mathbf{H})' + (\widehat{\mathbf{H}}^* - \mathbf{H})(\widehat{\mathbf{H}}^* - \mathbf{H})' = O_p(T^{-1/2}),$$

because of Lemma 6.

Define

$$\mathcal{K}_t = \mathbf{P}_{t|t-1} \boldsymbol{\Lambda}' (\boldsymbol{\Lambda} \mathbf{P}_{t|t-1} \boldsymbol{\Lambda}' + \mathbf{R})^{-1},$$

and analogously define $\widehat{\mathcal{K}}_t^*$ when using $\mathbf{P}_{t|t-1}^*$, $\widehat{\boldsymbol{\Lambda}}^*$ and $\widehat{\mathbf{R}}^*$. From (C72) we have

$$\begin{aligned} \mathbf{F}_{t|t}^* - \mathbf{F}_{t|t} &= \mathbf{F}_{t|t-1}^* - \mathbf{F}_{t|t-1} + (\widehat{\mathcal{K}}_t^* - \mathcal{K}_t)(\mathbf{x}_t - \boldsymbol{\Lambda} \mathbf{F}_{t|t-1}) \\ &\quad + (\widehat{\mathcal{K}}_t^* - \mathcal{K}_t)(\boldsymbol{\Lambda} \mathbf{F}_{t|t-1} - \widehat{\boldsymbol{\Lambda}}^* \mathbf{F}_{t|t-1}^*) + \mathcal{K}_t(\boldsymbol{\Lambda} \mathbf{F}_{t|t-1} - \widehat{\boldsymbol{\Lambda}}^* \mathbf{F}_{t|t-1}^*). \end{aligned} \quad (\text{C76})$$

Moreover, because of (C69)

$$\begin{aligned} \frac{\widehat{\boldsymbol{\Lambda}}^* \mathbf{F}_{t|t-1}^* - \boldsymbol{\Lambda} \mathbf{F}_{t|t-1}}{\sqrt{n}} &= \frac{\boldsymbol{\Lambda}}{\sqrt{n}} (\mathbf{F}_{t|t-1}^* - \mathbf{F}_{t|t-1}) + \left(\frac{\widehat{\boldsymbol{\Lambda}}^* - \boldsymbol{\Lambda}}{\sqrt{n}} \right) (\mathbf{F}_{t|t-1}^* - \mathbf{F}_{t|t-1}) + \left(\frac{\widehat{\boldsymbol{\Lambda}}^* - \boldsymbol{\Lambda}}{\sqrt{n}} \right) \mathbf{F}_{t|t-1} \\ &= \frac{\boldsymbol{\Lambda}}{\sqrt{n}} (\mathbf{F}_{t|t-1}^* - \mathbf{F}_{t|t-1}) + \left(\frac{\widehat{\boldsymbol{\Lambda}}^* - \boldsymbol{\Lambda}}{\sqrt{n}} \right) (\mathbf{F}_{t|t-1}^* - \mathbf{F}_{t|t-1}) + O_p(T^{-1/2}), \end{aligned} \quad (\text{C77})$$

since

$$\begin{aligned} \left(\frac{\widehat{\boldsymbol{\Lambda}}^* - \boldsymbol{\Lambda}}{\sqrt{n}} \right) \mathbf{F}_{t|t-1} &= \left(\frac{\widehat{\boldsymbol{\Lambda}}^* - \boldsymbol{\Lambda}}{\sqrt{n}} \right) \mathbf{A} \mathbf{F}_{t-1|t-1} = \left(\frac{\widehat{\boldsymbol{\Lambda}}^* - \boldsymbol{\Lambda}}{\sqrt{n}} \right) \mathbf{A} (\mathbf{F}_{t-1|t-1} - \mathbf{F}_{t-1}) + \left(\frac{\widehat{\boldsymbol{\Lambda}}^* - \boldsymbol{\Lambda}}{\sqrt{n}} \right) \mathbf{A} \mathbf{F}_{t-1} \\ &= O_p(T^{-1/2}) O(e^{-(t-1)/2}) + O_p(T^{-1/2}) O(n^{-1/2}) + O_p(T^{-1/2}), \end{aligned}$$

because of Lemma 6, (C66), (C69) and since $\|\mathbf{A}\| = O(1)$. Similarly, from (C73)

$$\begin{aligned} \mathbf{P}_{t|t}^* - \mathbf{P}_{t|t} &= \mathbf{P}_{t|t-1}^* - \mathbf{P}_{t|t-1} - \left[(\widehat{\mathcal{K}}_t^* - \mathcal{K}_t) \boldsymbol{\Lambda} \mathbf{P}_{t|t-1} \right. \\ &\quad \left. + (\widehat{\mathcal{K}}_t^* - \mathcal{K}_t) (\widehat{\boldsymbol{\Lambda}}^* \mathbf{P}_{t|t-1}^* - \boldsymbol{\Lambda} \mathbf{P}_{t|t-1}) + \mathcal{K}_t (\widehat{\boldsymbol{\Lambda}}^* \mathbf{P}_{t|t-1}^* - \boldsymbol{\Lambda} \mathbf{P}_{t|t-1}) \right]. \end{aligned} \quad (\text{C78})$$

Moreover,

$$\begin{aligned} \frac{\widehat{\boldsymbol{\Lambda}}^* \mathbf{P}_{t|t-1}^* - \boldsymbol{\Lambda} \mathbf{P}_{t|t-1}}{\sqrt{n}} &= \frac{\boldsymbol{\Lambda}}{\sqrt{n}} (\mathbf{P}_{t|t-1}^* - \mathbf{P}_{t|t-1}) + \left(\frac{\widehat{\boldsymbol{\Lambda}}^* - \boldsymbol{\Lambda}}{\sqrt{n}} \right) (\mathbf{P}_{t|t-1}^* - \mathbf{P}_{t|t-1}) + \left(\frac{\widehat{\boldsymbol{\Lambda}}^* - \boldsymbol{\Lambda}}{\sqrt{n}} \right) \mathbf{P}_{t|t-1} \\ &= \frac{\boldsymbol{\Lambda}}{\sqrt{n}} (\mathbf{P}_{t|t-1}^* - \mathbf{P}_{t|t-1}) + \left(\frac{\widehat{\boldsymbol{\Lambda}}^* - \boldsymbol{\Lambda}}{\sqrt{n}} \right) (\mathbf{P}_{t|t-1}^* - \mathbf{P}_{t|t-1}) + O_p(T^{-1/2}), \end{aligned} \quad (\text{C79})$$

since

$$\begin{aligned} \left(\frac{\widehat{\boldsymbol{\Lambda}}^* - \boldsymbol{\Lambda}}{\sqrt{n}} \right) \mathbf{P}_{t|t-1} &= \left(\frac{\widehat{\boldsymbol{\Lambda}}^* - \boldsymbol{\Lambda}}{\sqrt{n}} \right) (\boldsymbol{\Lambda} \mathbf{P}_{t-1|t-1} \boldsymbol{\Lambda}' + \mathbf{H} \mathbf{H}') \\ &= O_p(T^{-1/2}) O(e^{-(t-1)}) + O_p(T^{-1/2}) O(n^{-1}) + O_p(T^{-1/2}), \end{aligned}$$

because of Lemma 6 and (C50) in the proof of Lemma 5 and since $\|\mathbf{A}\| = O(1)$ and $\|\mathbf{H}\| =$

$O(1)$. Following the same reasoning we also have

$$\frac{\Lambda \mathbf{P}_{t|t-1}}{\sqrt{n}} = \frac{\Lambda}{\sqrt{n}} (\Lambda \mathbf{P}_{t-1|t-1} \mathbf{A}' + \mathbf{H} \mathbf{H}') = \frac{\Lambda}{\sqrt{n}} \left(O(e^{-(t-1)}) + O(n^{-1}) + \mathbf{H} \mathbf{H}' \right). \quad (\text{C80})$$

Now, set $t = 1$. Then, by noticing that $\mathbf{F}_{0|0}^* = \mathbf{F}_{0|0}$ and $\mathbf{P}_{0|0}^* = \mathbf{P}_{0|0}$, from (C74) and (C75) at $t = 1$ we have

$$\mathbf{F}_{1|0}^* - \mathbf{F}_{1|0} = O_p(T^{-1/2}), \quad \mathbf{P}_{1|0}^* - \mathbf{P}_{1|0} = O_p(T^{-1/2}). \quad (\text{C81})$$

Then, because of (C67), (C81) and Lemma 6, at $t = 1$ we have

$$\begin{aligned} & \sqrt{n}(\widehat{\mathcal{K}}_1^* - \mathcal{K}_1) \left(\frac{\mathbf{x}_1}{\sqrt{n}} - \frac{\Lambda \mathbf{F}_{1|0}}{\sqrt{n}} \right) = \\ & = \left[\mathbf{P}_{1|0}^* \frac{\widehat{\Lambda}^{*'}}{\sqrt{n}} \left(\frac{\widehat{\Lambda}}{\sqrt{n}} \mathbf{P}_{1|0}^* \frac{\widehat{\Lambda}^{*'}}{\sqrt{n}} + \frac{\widehat{\mathbf{R}}^*}{n} \right)^{-1} - \mathbf{P}_{1|0} \frac{\Lambda'}{\sqrt{n}} \left(\frac{\Lambda}{\sqrt{n}} \mathbf{P}_{1|0} \frac{\Lambda'}{\sqrt{n}} + \frac{\mathbf{R}}{n} \right)^{-1} \right] \left(\frac{\mathbf{x}_1}{\sqrt{n}} - \frac{\Lambda \mathbf{F}_{1|0}}{\sqrt{n}} \right) \\ & = O_p(T^{-1/2}). \end{aligned} \quad (\text{C82})$$

Moreover, from (C80)

$$\sqrt{n}(\widehat{\mathcal{K}}_1^* - \mathcal{K}_1) \frac{\Lambda \mathbf{P}_{1|0}}{\sqrt{n}} = O_p(T^{-1/2}). \quad (\text{C83})$$

From (C76) and using (C74), (C77), (C81), and (C82) at $t = 1$ we have

$$\begin{aligned} \mathbf{F}_{1|1}^* - \mathbf{F}_{1|1} &= \mathbf{F}_{1|0}^* - \mathbf{F}_{1|0} + (\widehat{\mathcal{K}}_1^* - \mathcal{K}_1)(\mathbf{x}_1 - \Lambda \mathbf{F}_{1|0}) \\ &+ (\widehat{\mathcal{K}}_1^* - \mathcal{K}_1)(\Lambda \mathbf{F}_{1|0} - \widehat{\Lambda} \mathbf{F}_{1|0}^*) + \mathcal{K}_1(\Lambda \mathbf{F}_{1|0} - \widehat{\Lambda} \mathbf{F}_{1|0}^*) = O_p(T^{-1/2}). \end{aligned} \quad (\text{C84})$$

Similarly, from (C78) and using (C75), (C79), (C81), and (C83) at $t = 1$ we have

$$\begin{aligned} \mathbf{P}_{1|1}^* - \mathbf{P}_{1|1} &= \mathbf{P}_{1|0}^* - \mathbf{P}_{1|0} - \left[(\widehat{\mathcal{K}}_1^* - \mathcal{K}_1) \Lambda \mathbf{P}_{1|0} \right. \\ &+ \left. (\widehat{\mathcal{K}}_1^* - \mathcal{K}_1)(\widehat{\Lambda}^* \mathbf{P}_{1|0}^* - \Lambda \mathbf{P}_{1|0}) + \mathcal{K}_1(\widehat{\Lambda}^* \mathbf{P}_{1|0}^* - \Lambda \mathbf{P}_{1|0}) \right] = O_p(T^{-1/2}). \end{aligned} \quad (\text{C85})$$

Then substituting (C84) into (C76) and (C85) into (C78) we have

$$\mathbf{F}_{2|1}^* - \mathbf{F}_{2|1} = O_p(T^{-1/2}), \quad \mathbf{P}_{2|1}^* - \mathbf{P}_{2|1} = O_p(T^{-1/2}). \quad (\text{C86})$$

Then, because of (C67), (C86) and Lemma 6, at $t = 2$ we have

$$\sqrt{n}(\widehat{\mathcal{K}}_2^* - \mathcal{K}_2) \left(\frac{\mathbf{x}_2}{\sqrt{n}} - \frac{\Lambda \mathbf{F}_{2|1}}{\sqrt{n}} \right) = O_p(T^{-1/2}), \quad (\text{C87})$$

and from (C80)

$$\sqrt{n}(\widehat{\mathcal{K}}_2^* - \mathcal{K}_2) \frac{\Lambda \mathbf{P}_{2|1}}{\sqrt{n}} = O_p(T^{-1/2}). \quad (\text{C88})$$

From, (C76) and using (C74), (C77), (C86), and (C87), at $t = 2$ we have

$$\mathbf{F}_{2|2}^* - \mathbf{F}_{2|2} = O_p(T^{-1/2}),$$

and from (C78) and using (C75), (C79), (C86), and (C88), at $t = 2$ we have

$$\mathbf{P}_{2|2}^* - \mathbf{P}_{2|2} = O_p(T^{-1/2}).$$

By repeating the same reasoning for $t = 3, \dots, T$ we have

$$\|\mathbf{F}_{t|t}^* - \mathbf{F}_{t|t}\| = O_p(T^{-1/2}), \quad \|\mathbf{P}_{t|t}^* - \mathbf{P}_{t|t}\| = O_p(T^{-1/2}), \quad (\text{C89})$$

and also

$$\|\mathbf{F}_{t|t-1}^* - \mathbf{F}_{t|t-1}\| = O_p(T^{-1/2}), \quad \|\mathbf{P}_{t|t-1}^* - \mathbf{P}_{t|t-1}\| = O_p(T^{-1/2}). \quad (\text{C90})$$

Because of Lemma 5 and (C89), we have for $\bar{t} \leq t \leq T$,

$$\begin{aligned} \|\mathbf{F}_{t|\bar{t}}^* - \mathbf{F}_{t|\bar{t}}\| &\leq \|\mathbf{F}_{t|\bar{t}}^* - \mathbf{F}_{t|\bar{t}}\| + \|\mathbf{F}_{t|\bar{t}} - \mathbf{F}_{t|\bar{t}}\| = O_p(T^{-1/2}) + O(n^{-1/2}), \\ \|\mathbf{P}_{t|\bar{t}}^*\| &\leq \|\mathbf{P}_{t|\bar{t}}^* - \mathbf{P}_{t|\bar{t}}\| + \|\mathbf{P}_{t|\bar{t}}\| = O_p(T^{-1/2}) + O(n^{-1}). \end{aligned}$$

Now compare the KS iterations, (B22)-(B26), with those obtained when using $\widehat{\Theta}^*$:

$$\mathbf{F}_{t|T}^* = \mathbf{F}_{t|t-1}^* + \mathbf{P}_{t|t-1}^* \mathbf{r}_{t-1}^*, \quad (\text{C91})$$

$$\mathbf{r}_{t-1}^* = \widehat{\Lambda}^{*'} (\widehat{\Lambda}^* \mathbf{P}_{t|t-1}^* \widehat{\Lambda}^{*'} + \widehat{\mathbf{R}}^*)^{-1} (\mathbf{x}_t - \widehat{\Lambda}^* \mathbf{F}_{t|t-1}^*) + \mathbf{L}_t^{*'} \mathbf{r}_t^*, \quad (\text{C92})$$

$$\mathbf{P}_{t|T}^* = \mathbf{P}_{t|t-1}^* - \mathbf{P}_{t|t-1}^* \mathbf{N}_{t-1}^* \mathbf{P}_{t|t-1}^*, \quad (\text{C93})$$

$$\mathbf{N}_{t-1}^* = \widehat{\Lambda}^{*'} (\widehat{\Lambda}^* \mathbf{P}_{t|t-1}^* \widehat{\Lambda}^{*'} + \widehat{\mathbf{R}}^*)^{-1} \widehat{\Lambda}^* + \mathbf{L}_t^{*'} \mathbf{N}_t^* \mathbf{L}_t^*, \quad (\text{C94})$$

$$\mathbf{L}_t^* = \widehat{\mathbf{A}}^* - \widehat{\Lambda}^* \mathbf{P}_{t|t-1}^* \widehat{\Lambda}^{*'} (\widehat{\Lambda}^* \mathbf{P}_{t|t-1}^* \widehat{\Lambda}^{*'} + \widehat{\mathbf{R}}^*)^{-1} \widehat{\Lambda}^*, \quad (\text{C95})$$

where $\mathbf{r}_T^* = \mathbf{0}_{r \times 1}$, $\mathbf{N}_T^* = \mathbf{0}_r$. First notice that obviously at $t = T$ both KF and KS give the same result hence (C89) applies also in this case, and because of Lemma 6, (C67), (C77), and (C90), we have

$$\mathbf{r}_{T-1}^* - \mathbf{r}_{T-1} = O_p(T^{-1/2}), \quad \mathbf{N}_{T-1}^* - \mathbf{N}_{T-1} = O_p(T^{-1/2}). \quad (\text{C96})$$

Moreover, from (C95), because of Lemma 6 and (C90), we have

$$\mathbf{L}_t^* - \mathbf{L}_t = \widehat{\mathbf{A}}^* - \mathbf{A} - \sqrt{n} \left[\widehat{\mathbf{A}}^* \widehat{\mathcal{K}}_t^* \frac{\widehat{\Lambda}^*}{\sqrt{n}} - \mathbf{A} \mathcal{K}_t \frac{\Lambda}{\sqrt{n}} \right] = O_p(T^{-1/2}). \quad (\text{C97})$$

Then, from (C92), because of (C67), (C77), (C90), (C96) and (C97), at $t = T - 1$ we have

$$\mathbf{r}_{T-2}^* - \mathbf{r}_{T-2} = O_p(T^{-1/2}), \quad \mathbf{N}_{T-2}^* - \mathbf{N}_{T-2} = O_p(T^{-1/2}). \quad (\text{C98})$$

Therefore, from (C91) and (C93), because of (C90) and (C98), we have

$$\mathbf{F}_{T-1|T}^* - \mathbf{F}_{T-1|T} = O_p(T^{-1/2}), \quad \mathbf{P}_{T-1|T}^* - \mathbf{P}_{T-1|T} = O_p(T^{-1/2}). \quad (\text{C99})$$

By repeating the same reasoning for $t = (T - 2), \dots, 1$, we have

$$\|\mathbf{F}_{t|T}^* - \mathbf{F}_{t|T}\| = O_p(T^{-1/2}), \quad \|\mathbf{P}_{t|T}^* - \mathbf{P}_{t|T}\| = O_p(T^{-1/2}). \quad (\text{C100})$$

Because of Lemma 5 and (C100), we have for $\bar{t} \leq t \leq T$

$$\begin{aligned} \|\mathbf{F}_{t|T}^* - \mathbf{F}_t\| &\leq \|\mathbf{F}_{t|T}^* - \mathbf{F}_{t|T}\| + \|\mathbf{F}_{t|T} - \mathbf{F}_t\| = O_p(T^{-1/2}) + O(n^{-1/2}), \\ \|\mathbf{P}_{t|T}^* &\leq \|\mathbf{P}_{t|T}^* - \mathbf{P}_{t|T}\| + \|\mathbf{P}_{t|T}\| = O_p(T^{-1/2}) + O(n^{-1}), \end{aligned}$$

which completes the proof. \square

Lemma 8. Consider the initial estimator of the parameters $\widehat{\Theta}_0$ defined in (B3), then there exists an invertible $r \times r$ matrix \mathbf{J} such that, as $n, T \rightarrow \infty$:

$$\begin{aligned} \min(\sqrt{n}, \sqrt{T}) \|\widehat{\lambda}'_{i0} - \lambda'_i \mathbf{J}^{-1}\| &= O_p(1), \quad i = 1, \dots, n, \\ \min(\sqrt{n}, \sqrt{T}) \|\widehat{\mathbf{A}}_0 - \mathbf{J} \mathbf{A} \mathbf{J}^{-1}\| &= O_p(1), \\ \min(\sqrt{n}, \sqrt{T}) \|\widehat{\mathbf{H}}_0 - \mathbf{J} \mathbf{H}\| &= O_p(1), \\ \min(\sqrt{n}, \sqrt{T}) |[\widehat{\mathbf{R}}]_{ii,0} - [\mathbf{R}]_{ii}| &= O_p(1), \quad i = 1, \dots, n. \end{aligned}$$

Moreover, under E1 and E2 we have $\mathbf{J} = \mathbf{I}_r$.

Proof. When E2 holds the proof Lemmas 3 and 5 in Barigozzi et al., 2016b where is shown that \mathbf{J} is a diagonal matrix with entries ± 1 . If we impose also E1 the sign indeterminacy is fixed and $\mathbf{J} = \mathbf{I}_r$. \square

Lemma 9. Consider the estimator of the parameters obtained at convergence of the EM algorithm $\widehat{\Theta} := \widehat{\Theta}_{k^*}$, then, as $n, T \rightarrow \infty$:

$$\begin{aligned} \sqrt{T} \|\widehat{\lambda}_{ik^*} - \lambda_i\| &= O_p(1), \quad i = 1, \dots, n, \\ \sqrt{T} \|\widehat{\mathbf{A}}_{k^*} - \mathbf{A}\| &= O_p(1), \\ \sqrt{T} \|\widehat{\mathbf{H}}_{k^*} - \mathbf{H}\| &= O_p(1), \\ \sqrt{T} |[\widehat{\mathbf{R}}]_{ii,k^*} - [\mathbf{R}]_{ii}| &= O_p(1), \quad i = 1, \dots, n. \end{aligned}$$

Proof. Define the $Q \times Q$ matrices

$$\mathcal{I}(\Theta) = -\nabla_{\Theta\Theta'}^2 \ell(\mathcal{X}_T; \Theta),$$

$$\mathcal{I}_0(\Theta) = -\int_{\mathbb{R}^r \times T} \nabla_{\Theta\Theta'}^2 \ell(\mathcal{X}_T, \mathcal{F}_T; \Theta) f(\mathcal{F}_T | \mathcal{X}_T; \Theta) d\mathcal{F}_T = -\mathbb{E}_{\Theta} [\nabla_{\Theta\Theta'}^2 \ell(\mathcal{X}_T, \mathcal{F}_T; \Theta) | \mathcal{X}_T],$$

$$\mathcal{I}_1(\Theta) = -\int_{\mathbb{R}^r \times T} \nabla_{\Theta\Theta'}^2 \ell(\mathcal{F}_T | \mathcal{X}_T, \mathcal{F}_T; \Theta) f(\mathcal{F}_T | \mathcal{X}_T; \Theta) d\mathcal{F}_T = -\mathbb{E}_{\Theta} [\nabla_{\Theta\Theta'}^2 \ell(\mathcal{F}_T | \mathcal{X}_T; \Theta) | \mathcal{X}_T],$$

then, since $\mathcal{I}(\Theta)$ does not depend on \mathcal{F}_T from (B1) and (B11)

$$\mathcal{I}(\Theta) = \mathcal{I}_0(\Theta) - \mathcal{I}_1(\Theta). \quad (\text{C101})$$

Moreover, at convergence of the EM algorithm (iteration k^*) we have the Taylor approximation

$$(\widehat{\Theta}_{k^*} - \widehat{\Theta}^*) = (\widehat{\Theta}_{k^*-1} - \widehat{\Theta}^*)\mathcal{R}(\widehat{\Theta}^*) + O\left(\|\widehat{\Theta}_{k^*} - \widehat{\Theta}^*\|^2\right), \quad (\text{C102})$$

where following Meng and Rubin (1994) we have (see also (C101))

$$\mathcal{R}(\widehat{\Theta}^*) = \mathcal{I}_1(\widehat{\Theta}^*) \left(\mathcal{I}_0(\widehat{\Theta}^*)\right)^{-1} = \mathbf{I}_Q - \mathcal{I}(\widehat{\Theta}^*) \left(\mathcal{I}_0(\widehat{\Theta}^*)\right)^{-1}.$$

Hence, by iterating (C102) k^* times and neglecting the second term on the rhs which at convergence is always smaller than the first term, we obtain

$$\|\widehat{\Theta}_{k^*} - \widehat{\Theta}^*\| \leq \|\widehat{\Theta}_0 - \widehat{\Theta}^*\| \|\mathcal{R}(\widehat{\Theta}^*)\|^{k^*}. \quad (\text{C103})$$

Hereafter, denote $\zeta_{nT} := \max(n^{-1/2}, T^{-1/2})$. Consider (C103) for the estimated loadings, for any $i = 1, \dots, n$, and using Gaussianity (see D2 and D3), we have

$$\begin{aligned} \|\widehat{\lambda}_{ik^*} - \widehat{\lambda}_i^*\| &\leq \|\widehat{\lambda}_{i0} - \widehat{\lambda}_i^*\| \left\| \mathbf{I}_r - \left(\sum_{t=1}^T \mathbf{F}_t \mathbf{F}_t' \right) \left(\sum_{t=1}^T \mathbb{E}_{\widehat{\Theta}^*} [\mathbf{F}_t \mathbf{F}_t' | \mathcal{X}_T] \right)^{-1} \right\|^{k^*} \\ &= \|\widehat{\lambda}_{i0} - \widehat{\lambda}_i^*\| \left\| \mathbf{I}_r - \left(\frac{1}{T^2} \sum_{t=1}^T \mathbf{F}_t \mathbf{F}_t' \right) \left(\frac{1}{T^2} \sum_{t=1}^T (\mathbf{F}_{t|T}^* \mathbf{F}_{t|T}^{*'} + \mathbf{P}_{t|T}^*) \right)^{-1} \right\|^{k^*}. \end{aligned} \quad (\text{C104})$$

Therefore, by condition (15) in the text $\bar{t} = O(\log T)$, because of Lemma 7, we have

$$\begin{aligned} \left(\frac{1}{T^2} \sum_{t=1}^T (\mathbf{F}_{t|T}^* \mathbf{F}_{t|T}^{*'} + \mathbf{P}_{t|T}^*) \right)^{-1} &= \left(\frac{1}{T^2} \sum_{t=1}^{\bar{t}-1} (\mathbf{F}_{t|T}^* \mathbf{F}_{t|T}^{*'} + \mathbf{P}_{t|T}^*) \right) + \frac{1}{T^2} \sum_{t=\bar{t}}^T (\mathbf{F}_{t|T}^* \mathbf{F}_{t|T}^{*'} + \mathbf{P}_{t|T}^*)^{-1} \\ &= \left(\frac{1}{T^2} \sum_{t=1}^T \mathbf{F}_t \mathbf{F}_t' + O_p\left(\frac{\zeta_{nT}}{\sqrt{T}}\right) + O_p\left(\frac{\log T}{T}\right) \right)^{-1} \\ &= \left(\frac{1}{T^2} \sum_{t=1}^T \mathbf{F}_t \mathbf{F}_t' \right)^{-1} + o_p(T^{-1/2}), \end{aligned} \quad (\text{C105})$$

since $\|\mathbf{F}_{t|T}^* \mathbf{F}_{t|T}^{*'} + \mathbf{P}_{t|T}^*\| = O_p(T)$ and $\|\mathbf{F}_t \mathbf{F}_t'\| = O_p(T)$, because $\mathbf{F}_t \sim I(1)$ and $\mathbf{F}_{t|T}^* \sim I(1)$, and

$$\frac{1}{T^2} \sum_{t=1}^{\bar{t}-1} (\mathbf{F}_{t|T}^* \mathbf{F}_{t|T}^{*'} + \mathbf{P}_{t|T}^*) = O_p\left(\frac{\log T}{T}\right), \quad \frac{1}{T^2} \sum_{t=1}^T \mathbf{F}_t \mathbf{F}_t' - \frac{1}{T^2} \sum_{t=\bar{t}}^T \mathbf{F}_t \mathbf{F}_t' = O_p\left(\frac{\log T}{T}\right).$$

Moreover, because of Lemmas 8 and 6, we have

$$\|\widehat{\lambda}_{i0} - \widehat{\lambda}_i^*\| \leq \|\widehat{\lambda}_{i0} - \lambda_i\| + \|\widehat{\lambda}_i^* - \lambda_i\| = O_p(\zeta_{nT}) + O_p(T^{-1/2}). \quad (\text{C106})$$

By substituting (C105) and (C106) into (C104), we have

$$\|\widehat{\lambda}_{ik^*} - \widehat{\lambda}_i^*\| = o_p(\zeta_{nT} T^{-k^*/2}). \quad (\text{C107})$$

Finally, because of Lemma 6 and (C107), we have

$$\|\widehat{\boldsymbol{\lambda}}_{ik^*} - \boldsymbol{\lambda}_i\| \leq \|\widehat{\boldsymbol{\lambda}}_{ik^*} - \widehat{\boldsymbol{\lambda}}_i^*\| + \|\widehat{\boldsymbol{\lambda}}_i^* - \boldsymbol{\lambda}_i\| = o_p(T^{-k^*/2}) + O_p(T^{-1/2}).$$

The proof for the other parameters follows the same steps by taking the appropriate second derivatives and applying the results in Lemma 7. \square

C.4 Proof of Proposition 1

Consistency of the estimated loadings is proved in Lemma 9. Recalling that $\widehat{\boldsymbol{\Theta}} := \widehat{\boldsymbol{\Theta}}_{k^*}$ and also $\boldsymbol{\lambda}_i = \mathbf{K}^{-1'}(\mathbf{b}'_{i0} \mathbf{b}'_{i1})$, because of (A12) we prove (16).

Consistency of the estimated static factors is then proved as in Lemma 7 but using the results of Lemma 9 for the estimated parameters. In particular, for all $\bar{t} \leq t \leq T$

$$\|\mathbf{F}_{t|T,k^*} - \mathbf{F}_t\| \leq \|\mathbf{F}_{t|T,k^*} - \mathbf{F}_{t|\bar{t}}\| + \|\mathbf{F}_{t|\bar{t}} - \mathbf{F}_t\| = O_p(T^{-1/2}) + O(n^{-1/2}).$$

Recalling that $\widehat{\mathbf{F}}_t := \mathbf{F}_{t|T,k^*}$ and also $\mathbf{F}_t = \mathbf{K}(\mathbf{f}'_t \mathbf{f}'_{t-1})'$ because of (A11) we prove (17). A similar result can be proved for the KF estimator of the static factors, $\mathbf{F}_{t|\bar{t},k^*}$ using again Lemmas 7 and 9.

Denote $\zeta_{nT} := \max(n^{-1/2}, T^{-1/2})$. Then, recalling that $\widehat{\boldsymbol{\lambda}}_i' \widehat{\mathbf{F}}_t := \widehat{\boldsymbol{\lambda}}_{ik^*}' \mathbf{F}_{t|T,k^*}$, because of (16), (17) and Lemma 5, we have

$$\begin{aligned} |\widehat{\chi}_{it} - \chi_{it}| &= |\widehat{\boldsymbol{\lambda}}_i' \widehat{\mathbf{F}}_t - \boldsymbol{\lambda}_i' \mathbf{F}_t| \leq \|(\widehat{\boldsymbol{\lambda}}_i - \boldsymbol{\lambda}_i) \mathbf{F}_t\| + \|\boldsymbol{\lambda}_i\| \|\widehat{\mathbf{F}}_t - \mathbf{F}_t\| + \|\widehat{\boldsymbol{\lambda}}_i - \boldsymbol{\lambda}_i\| \|\widehat{\mathbf{F}}_t - \mathbf{F}_t\| \\ &= \|(\widehat{\boldsymbol{\lambda}}_i - \boldsymbol{\lambda}_i) \mathbf{F}_t\| + O_p(\zeta_{nT}) \\ &= \|\widehat{\boldsymbol{\lambda}}_i - \widehat{\boldsymbol{\lambda}}_i^*\| \|\mathbf{F}_t\| + \|(\widehat{\boldsymbol{\lambda}}_i^* - \boldsymbol{\lambda}_i) \mathbf{F}_t\| + O_p(\zeta_{nT}). \end{aligned} \quad (\text{C108})$$

Now, from Lemma 9 (see in particular (C107)) and since $\|\mathbf{F}_t\| = O_p(T^{-1/2})$, we have

$$\|\widehat{\boldsymbol{\lambda}}_i - \widehat{\boldsymbol{\lambda}}_i^*\| \|\mathbf{F}_t\| = o_p(\zeta_{nT} T^{-(k^*-1)/2}). \quad (\text{C109})$$

Moreover, from transformation (C57), defined in the proof of Lemma 6, $\mathcal{D}\mathbf{F}_t = (\mathbf{Z}'_{1t} \mathbf{Z}'_{0t})'$, we have $\|\mathbf{Z}_{1t}\| = O_p(\sqrt{T})$ and $\|\mathbf{Z}_{0t}\| = O_p(1)$. Then, since $\mathcal{D}'\mathcal{D} = \mathbf{I}_r$, as a consequence of Lemma 6 (see in particular (C59)), we have

$$(\widehat{\boldsymbol{\lambda}}_i^* - \boldsymbol{\lambda}_i)' \mathcal{D}' \mathcal{D} \mathbf{F}_t = O_p(T^{-1/2}). \quad (\text{C110})$$

By substituting (C109) and (C110) into (C108) and since $k^* \geq 1$ because we run the EM algorithm at least once after initialization, we prove (18). \square

Appendix D Data Description and Data Treatment

This Appendix present the dataset used in the analysis. All variables where downloaded from Haver on June 16th 2017. None of the variables where adjusted for outliers but variables 57, 83, 87, and 94. All variables are from the USECON database but variable 103 that is from the DAILY database. All monthly and daily series are transformed into quarterly observation by simple averages.

In order to choose whether or not to de-trend a variable we apply the following procedure: let m_i be the sample mean of Δy_{it} , $\gamma_i(j)$ be the auto-covariance of order j of Δy_{it} , and $\bar{\gamma}_i = \sqrt{\frac{1}{T} \sum_{j=1}^J \gamma_i(j)}$, then if $\frac{|m_i|}{\bar{\gamma}_i} \geq 1.96$ we estimate a_i and b_i from an OLS regression of y_{it} on a constant and a time trend, whereas if $\frac{|m_i|}{\bar{\gamma}_i} < 1.96$ we set $\hat{a}_i = m_i$ and $\hat{b}_i = 0$.

LIST OF ABBREVIATIONS

Source:			
BLS=U.S. Department of Labor: Bureau of Labor Statistics			
BEA=U.S. Department of Commerce: Bureau of Economic Analysis			
ISM = Institute for Supply Management			
CB=U.S. Department of Commerce: Census Bureau			
FRB=Board of Governors of the Federal Reserve System			
EIA=Energy Information Administration			
WSJ=Wall Street Journal			
CBO=Congressional Budget Office			
FRBPHIL=Federal Reserve Bank of Philadelphia			
F = Frequency	T=Transformation	SA	ξ =Idiosyncratic
Q = Quarterly	0 = None	0 = no	0= $I(0)$
M = Monthly	1 = log	1 = yes	1= $I(1)$
D = Daily	2 = $\Delta \log$		
D = Deterministic Component		U=Units	
0 = $\hat{a}_i = \frac{1}{T} \sum_{t=1}^T \Delta y_{it}, \hat{b}_i = 0$		1000-P = Thousands of Persons	
1 = OLS Detrending		1000-U = Thousands of Units	
		BoC = Billions of Chained	
		\$-B = Dollars per Barrel	

N	Series ID	Definition	Unit	F	S	SA	T	D	ξ
1	GDPH	Real Gross Domestic Product	BoC 2009\$	Q	BEA	1	1	1	0
2	GDYH	Real gross domestic income	BoC 2009\$	Q	BEA	1	1	1	0
3	FSH	Real Final Sales of Domestic Product	BoC 2009\$	Q	BEA	1	1	1	1
4	IH	Real Gross Private Domestic Investment	BoC 2009\$	Q	BEA	1	1	1	1
5	GSH	Real State & Local*	BoC 2009\$	Q	BEA	1	1	1	1
6	FRH	Real Private Residential Fixed Investment	BoC 2009\$	Q	BEA	1	1	1	0
7	FNH	Real Private Nonresidential Fixed Investment	BoC 2009\$	Q	BEA	1	1	1	1
8	MH	Real Imports of Goods & Services	BoC 2009\$	Q	BEA	1	1	1	0
9	GH	Real Government*	BoC 2009\$	Q	BEA	1	1	1	1
10	XH	Real Exports of Goods & Services	BoC 2009\$	Q	BEA	1	1	1	0
14	CH	Real Personal Consumption Expenditures (PCE)	BoC 2009\$	Q	BEA	1	1	1	0
11	CNH	Real PCE: Nondurable Goods	BoC 2009\$	Q	BEA	1	1	1	1
12	CSH	Real PCE: Services	BoC 2009\$	Q	BEA	1	1	1	0
13	CDH	Real PCE: Durable Goods	BoC 2009\$	Q	BEA	1	1	1	0
15	GFDIH	Real National Defense Gross Investment	BoC 2009\$	Q	BEA	1	1	1	0
16	GFNIH	Real Federal Nondefense Gross Investment	BoC 2009\$	Q	BEA	1	1	1	0
17	YPDH	Real Disposable Personal Income	BoC 2009\$	Q	BEA	1	1	1	0
18	JI	Gross Private Domestic Investment:*	2009=100	Q	BEA	1	2	0	0
19	JGDP	Gross Domestic Product:*	2009=100	Q	BEA	1	2	0	1
20	LXNFU	Unit Labor Cost [†]	2009=100	Q	BLS	1	1	1	1
21	LXNFR	Real Compensation Per Hour [†]	2009=100	Q	BLS	1	1	1	1
22	LXNFC	Compensation Per Hour [†]	2009=100	Q	BLS	1	1	1	1
23	LXNFH	Hours of All Persons [†]	2009=100	Q	BLS	1	1	1	0
24	LXNFA	Output Per Hour of All Persons [†]	2009=100	Q	BLS	1	1	1	0
25	LXMU	Unit Labor Cost [‡]	2009=100	Q	BLS	1	1	1	1
26	LXMR	Real Compensation Per Hour [‡]	2009=100	Q	BLS	1	1	1	1
27	LXMC	Compensation Per Hour [‡]	2009=100	Q	BLS	1	1	1	0
28	LXMH	Hours of All Persons [‡]	2009=100	Q	BLS	1	1	1	0
29	LXMA	Output Per Hour of All Persons [‡]	2009=100	Q	BLS	1	1	1	1
30	IP	Industrial Production (IP) Index	2012=100	M	FRB	1	1	1	0
31	IP521	IP: Business Equipment	2012=100	M	FRB	1	1	1	1
32	IP511	IP: Durable Consumer Goods	2012=100	M	FRB	1	1	1	0
33	IP531	IP: Durable Materials	2012=100	M	FRB	1	1	1	1
34	IP512	IP: Nondurable Consumer Goods	2012=100	M	FRB	1	1	1	0
35	IP532	IP: nondurable Materials	2012=100	M	FRB	1	1	1	0

* Consumption Expenditures & Gross Investment

* Chain-type Price Index

† Nonfarm Business Sector

‡ Manufacturing Sector

N	Series ID	Definition	Unit	F	S	SA	T	D	ξ
36	PCU	CPI-U: All Items	82-84=100	M	BLS	1	2	0	0
37	PCUSE	CPI-U: Energy	82-84=100	M	BLS	1	2	0	0
38	PCUSLFE	CPI-U: All Items Less Food and Energy	82-84=100	M	BLS	1	2	0	0
39	PCUFO	CPI-U: Food	82-84=100	M	BLS	1	2	0	0
40	JCBM	PCE: Chain Price Index	2009=100	M	BEA	1	2	0	0
41	JCEBM	PCE: Energy Goods & Services-price index	2009=100	M	BEA	1	2	0	0
42	JCNFOM	PCE: Food & Beverages-price index*	2009=100	M	BEA	1	2	0	0
43	JCXFEBM	PCE less Food & Energy-price index	2009=100	M	BEA	1	2	0	0
44	JCSBM	PCE: Services-price index	2009=100	M	BEA	1	2	0	0
45	JCDBM	PCE: Durable Goods-price index	2009=100	M	BEA	1	2	0	0
46	JCNBM	PCE: Nondurable Goods-price index	2009=100	M	BEA	1	2	0	0
47	PC1	PPI: Intermediate Demand Processed Goods	1982=100	M	BLS	1	2	0	0
48	P05	PPI: Fuels and Related Products and Power	1982=100	M	BLS	0	2	0	0
49	SP3000	PPI: Finished Goods	1982=100	M	BLS	1	2	0	0
50	PIN	PPI: Industrial Commodities	1982=100	M	BLS	0	2	0	0
51	PA	PPI: All Commodities	1982=100	M	BLS	0	2	0	0
52	PC1	PPI: Intermediate Demand Processed Goods	1982=100	M	BLS	1	2	0	0
53	FMC	Money Stock: Currency	Bil. of \$	M	FRB	1	2	0	0
54	FM1	Money Stock: M1	Bil. of \$	M	FRB	1	2	0	1
55	FM2	Money Stock: M2	Bil. of \$	M	FRB	1	2	0	0
56	FABWC	C & I Loans in Bank Credit:†	Bil. of \$	M	FRB	1	1	1	1
57	FABWQ	Consumer Loans in Bank Credit:†	Bil. of \$	M	FRB	1	1	1	1
58	FAB	Bank Credit:†	Bil. of \$	M	FRB	1	1	1	1
59	FABW	Loans & Leases in Bank Credit:†	Bil. of \$	M	FRB	1	1	1	1
60	FABYO	Other Securities in Bank Credit:†	Bil. of \$	M	FRB	1	1	1	1
61	FABWR	Real Estate Loans in Bank Credit:†	Bil. of \$	M	FRB	1	1	1	0
62	FOT	Consumer Credit Outstanding	Bil. of \$	M	FRB	1	1	1	0
63	HSTMW	Housing Starts: Midwest	1000-U	M	CB	1	1	0	0
64	HSTNE	Housing Starts: Northeast	1000-U	M	CB	1	1	0	0
65	HSTS	Housing Starts: South	1000-U	M	CB	1	1	0	0
66	HSTGW	Housing Starts: West	1000-U	M	CB	1	1	0	0
67	HPT	Building Permit*	1000-U	M	CB	1	1	0	0
68	FBPR	Bank Prime Loan Rate	Percent	M	FRB	0	0	0	0
69	FFED	Federal Funds [effective] Rate	Percent	M	FRB	0	0	0	0
70	FCM1	1-Year Treasury Bill Yield‡	Percent	M	FRB	0	0	0	0
71	FCM10	10-Year Treasury Note Yield‡	Percent	M	FRB	0	0	0	0

* Purchased for Off-Premises Consumption

† All Commercial Banks

* New Private Housing Units Authorized by ‡ at Constant Maturity

N	Series ID	Definition	Unit	F	S	SA	T	D	ξ
72	LP	Civilian Participation Rate: 16 yr +	Percent	M	BLS	0	0	0	1
73	LQ	Civilian Employment/Population Ratio: 16 yr +	Percent	M	BLS	0	0	0	1
74	LE	Civilian Employment: Sixteen Years & Over	1000-P	M	BLS	0	1	1	0
75	LR	Civilian Unemployment Rate: 16 yr +	Percent	M	BLS	0	0	0	0
76	LU0	Civilians Unemployed for Less Than 5 Weeks	1000-P	M	BLS	0	1	0	0
77	LU5	Civilians Unemployed for 5-14 Weeks	1000-P	M	BLS	0	1	0	1
78	LU15	Civilians Unemployed for 15-26 Weeks	1000-P	M	BLS	0	1	0	1
79	LUT27	Civilians Unemployed for 27 Weeks and Over	1000-P	M	BLS	0	1	0	1
80	LUAD	Average [Mean] Duration of Unemployment	Weeks	M	BLS	0	1	0	0
81	LANAGRA	All Employees: Total Nonfarm	1000-P	M	BLS	0	1	1	1
82	LAPRIVA	All Employees: Total Private Industries	1000-P	M	BLS	0	1	1	0
83	LANTRMA	All Employees: Mining and Logging	1000-P	M	BLS	0	1	1	1
84	LACONSA	All Employees: Construction	1000-P	M	BLS	0	1	1	1
85	LAMANUA	All Employees: Manufacturing	1000-P	M	BLS	0	1	1	0
86	LATTULA	All Employees: Trade, Transportation & Utilities	1000-P	M	BLS	0	1	1	1
87	LAINFOA	All Employees: Information Services	1000-P	M	BLS	0	1	1	1
88	LAFIREA	All Employees: Financial Activities	1000-P	M	BLS	0	1	1	1
89	LAPBSVA	All Employees: Professional & Business Services	1000-P	M	BLS	0	1	1	1
90	LAEDUHA	All Employees: Education & Health Services	1000-P	M	BLS	0	1	1	1
91	LALEIHA	All Employees: Leisure & Hospitality	1000-P	M	BLS	0	1	1	1
92	LASRVOA	All Employees: Other Services	1000-P	M	BLS	0	1	1	1
93	LAGOVTA	All Employees: Government	1000-P	M	BLS	0	1	1	0
94	LAFGOVA	All Employees: Federal Government	1000-P	M	BLS	0	1	1	1
95	LASGOVA	All Employees: State Government	1000-P	M	BLS	0	1	1	0
96	LALGOVA	All Employees: Local Government	1000-P	M	BLS	0	1	1	0
97	PETEXA	West Texas Intermediate Spot Price FOB*	\$-B	M	EIA	0	2	0	0
98	NAPMNI	ISM Mfg: New Orders Index	Index	M	ISM	1	0	0	1
99	NAPMOI	ISM Mfg: Production Index	Index	M	ISM	1	0	0	1
100	NAPMEI	ISM Mfg: Employment Index	Index	M	ISM	1	0	0	1
101	NAPMVDI	ISM Mfg: Supplier Deliveries Index	Index	M	ISM	1	0	0	0
102	NAPMII	ISM Mfg: Inventories Index	Index	M	ISM	1	0	0	0
103	SP500	Standard & Poor's 500 Stock Price Index	41-43=10	D	WSJ	0	1	1	0

* Cushing, Oklahoma

Series ID	Definition	Unit	F	Source
GDPPOTHQ	Real Potential Gross Domestic Product	BoC 2009\$	Q	CBO
NAIRUQ	Natural Rate of Unemployment	percent	Q	CBO
GDPPLUS	US GDPplus	percent	Q	FRBPHIL