Non-Stationary Dynamic Factor Models for Large Datasets

Matteo BARIGOZZI[1] Marco LIPPI[2] Matteo LUCIANI[3]

November 1, 2016

Abstract

We develop the econometric theory for Non-Stationary Dynamic Factor models for large panels of time series, with a particular focus on building estimators of impulse response functions to unexpected macroeconomic shocks. We derive conditions for consistent estimation of the model as both the cross-sectional size, n, and the time dimension, T, go to infinity, and whether or not cointegration is imposed. We also propose a new estimator for the non-stationary common factors, as well as an information criterion to determine the number of common trends. Finally, the numerical properties of our estimator are explored by means of a MonteCarlo exercise and of a real-data application, in which we study the effects of monetary policy and supply shocks on the US economy.

JEL subject classification: C0, C01, E0.

Key words and phrases: Dynamic Factor models, unit root processes, cointegration, common trends, impulse response functions.

[1]m.barigozzi@lse.ac.uk – London School of Economics and Political Science, UK.
[2]ml@lippi.ws – Einaudi Institute for Economics and Finance, Roma, Italy.
[3]matteo.luciani@frb.gov – Federal Reserve Board of Governors, Washington DC, USA.

Special thanks go to Paolo Paruolo and Lorenzo Trapani for helpful comments. This paper has benefited also from discussions with Antonio Conti, Domenico Giannone, Dietmar Bauer, and all participants to the 39th Annual NBER Summer Institute. This paper was written while Matteo Luciani was *chargé de recherches* F.R.S.-F.N.R.S., and he gratefully acknowledges their financial support. Of course, any errors are our responsibility.

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

1 Introduction

Since the early 2000s large-dimensional Dynamic Factor models have become increasingly popular in the economic literature and they are nowadays commonly used by policy institutions. Economists have been attracted by these models because they allow to analyze large panels of time series without suffering of the curse of dimensionality. Furthermore, these models proved successful in forecasting (Stock and Watson, 2002a,b; Forni et al., 2005; Giannone et al., 2008; Luciani, 2014), in the construction of both business cycle indicators and inflation indexes (Altissimo et al., 2010; Cristadoro et al., 2005), and also in policy analysis based on impulse response functions (Giannone et al., 2005; Stock and Watson, 2005; Forni et al., 2009; Forni and Gambetti, 2010; Barigozzi et al., 2014; Luciani, 2015), thus becoming a standard econometric tool in empirical macroeconomic analysis.

Dynamic Factor models are based on the idea that fluctuations in the economy are due to a few macroeconomic *common shocks*, affecting all the variables, and to several other *idiosyncratic shocks* resulting from measurement error and/or from sectorial and regional dynamics, and influencing just one or a few variables. Therefore, each variable can be decomposed into a part driven by the common shocks, and a part driven by the idiosyncratic shocks, and by focussing only on the dynamic effects of the macroeconomic shocks, it is easily possible to analyze large databases. Finally, it is normally assumed that the comovement generated by the macroeconomic shocks can be summarized by means of a few latent time series processes, called *common factors* and capturing the business cycle.

So far, large-dimensional Dynamic Factor models have been studied mainly in a stationary setting, in which case the model can be consistently estimated either with principal components (Forni et al., 2000, 2005; Stock and Watson, 2002a; Bai and Ng, 2002; Bai, 2003), or with Maximum Likelihood by means of the EM algorithm (Watson and Engle, 1983; Doz et al., 2012). Most macroeconomic variables, though, are non-stationary, and therefore the common practice is to take first differences of the data to reach stationarity, before estimating the model. However, despite this practice has been successful in empirical applications, it has the shortcoming that in this setting by construction all common shocks have a permanent effect on the level of most variables. This is at odds with economic theory, as there is full agreement in the macroeconomic literature that while some shocks (such as technology shocks) have indeed permanent effects, thus generating common trends, some others (such as monetary policy shocks) have only transitory effects, thus generating stationary fluctuations around the trend. For example, standard Dynamic Stochastic General Equilibrium models not only imply a factor structure in the data (Giannone et al., 2006), but are also stationary around a common stochastic trend (Del Negro et al., 2007).

In this paper, we propose a Dynamic Factor model for large datasets—also known as Generalized or Approximate Dynamic Factor models—which is compatible with the long-run predictions of macroeconomic theory, i.e. with the idea that "some", but not all, macroeconomic shocks have a permanent effect on the economy. To this end, we propose the *Non-Stationary Dynamic Factor model for Large Datasets*, thus explicitly addressing the presence of "some" unit roots in the factors, with a focus in building a framework useful for empirical macroeconomic analysis. Specifically, in this paper we study estimation, while in a companion paper (Barigozzi et al., 2016) we address representation theory.

In detail, we first generalize the stationary Dynamic Factor model by Stock and Watson (2005), Bai and Ng (2007), and Forni et al. (2009), and then we study estimation of the model.

In particular, we derive the conditions for estimating consistently all parameters, when both the cross-sectional and the time dimensions of the dataset grow to infinity. Our estimator is based on approximate principal components, and on a VECM or an unrestriced VAR model for the latent $I(1)$ common factors.

All our results are derived without imposing any stationarity restriction on the idiosyncratic component, an assumption which is shown to be non-realistic, especially when analyzing large macroeconomic databases. Moreover, as a by-product of our estimation strategy we propose a new estimator for the non-stationary common factors, which can be directly employed in the estimation of non-stationary Factor Augmented VAR models (see Bernanke et al., 2005 and Bai and Ng, 2006, for the stationary case). Finally, we also provide an information criterion to determine the number of common trends in large panels.

This paper is complementary to the works of Bai and Ng (2004, 2010) and Bai (2004), and to a lesser extent also of Peña and Poncela (2004). On the one hand Bai and Ng (2004, 2010), having unit root testing in large panels as a goal, focus just on factor estimation but have almost no result for the other parameters of the model, notably for the coefficients of the autoregressive representation of the common factors. On the other hand, the results in Bai (2004), of which Peña and Poncela (2004) is a special case for small datasets, require the assumption of stationary idiosyncratic components. These approaches have been also applied to structural macroeconomic analysis. For example, Eickmeier (2009) estimates impulse responses to analyze the euro area business cycle, and Forni et al. (2014) study the effects of news shocks on the US business cycle. Finally, Banerjee et al. (2014) study cointegration between the common factors and the observed variables.

The rest of the paper is organized as follows. In Section 2 we present the model and the assumptions upon which the theory is developed. In Section 3 we describe estimation of the model and we discuss the related asymptotic properties. In Section 4 we present an information criterion for determining the number of common trends in large panels. In Section 5 by means of a MonteCarlo simulation exercise we study the finite sample properties of the estimator. Finally, in Section 6 we use our model to study the impact of monetary policy shocks and of supply shocks on the US economy. In Section 7 we conclude and we discuss possible further applications of the model presented. The proofs of our main results are in Appendix A, while complementary results, needed for the proofs, are in Appendix B.

2 The Non-Stationary Dynamic Factor model

In this section we first describe the main features of the Non-Stationary Dynamic Factor model, and then we introduce and discuss its formal assumptions. Hereafter, we say that a vector process $\mathbf{y}_t$ is $I(1)$ if its first difference $(1-L)\mathbf{y}_t$ is $I(0)$, where L is the lag-operator, therefore, $\mathbf{y}_t$ can be $I(1)$ even though some of its coordinates are $I(0)$. We write autoregressive matrix polynomials as $\mathbf{A}(L) = \mathbf{A}(0) - \sum_{k=1}^{s_1} \mathbf{A}_k L^k$, while we write moving average matrix polynomials as $\mathbf{C}(L) = \sum_{k=0}^{s_2} \mathbf{C}_k L^k$, where $s_1, s_2 \geq 0$.

2.1 The model

In the Non-Stationary Dynamic Factor model each observed variable, x_{it}, is decomposed into the sum of (i) a *common component*, χ_{it}, which is a linear combination of an r-dimensional

$I(1)$ vector, $\mathbf{F}_t$, of latent *common factors*, and (ii) an *idiosyncratic component*, ξ_{it}, which is possibly $I(1)$. Formally,

$$x_{it} = \chi_{it} + \xi_{it}, \qquad \chi_{it} = \boldsymbol{\lambda}_i' \mathbf{F}_t, \quad i = 1, \ldots, n, \tag{1}$$

where $\boldsymbol{\lambda}_i'$ is a $1 \times r$ vector of factor loadings. We denote as $\mathbf{x}_t = (x_{1t} \ldots x_{nt})'$ the n-dimensional vector of observed data, and similarly $\boldsymbol{\chi}_t$ and $\boldsymbol{\xi}_t$ are the n-dimensional vectors of common and idiosyncratic components respectively. The $n \times r$ matrix of factor loadings is denoted as $\boldsymbol{\Lambda} = (\boldsymbol{\lambda}_1 \ldots \boldsymbol{\lambda}_n)'$. The dimension of the dataset n is assumed to be large, or, more formally, we allow for $n \to \infty$ and we assume $r < n$, with r finite and independent of n, that is the comovements among a large number of variables can be captured by a small number of latent processes.

The vector of common factors is driven by a q-dimensional orthonormal white noise process, $\mathbf{u}_t$, whose components are called *common shocks*. Formally, we have the ARIMA model

$$\mathbf{S}(L)(1 - L)\mathbf{F}_t = \mathbf{Q}(L)\mathbf{u}_t, \quad \mathbf{u}_t \overset{\text{i.i.d.}}{\sim} (\mathbf{0}, \mathbf{I}_q), \tag{2}$$

where $\mathbf{S}(L)$ is an $r \times r$ finite and stable matrix polynomial, and $\mathbf{Q}(L)$ is an $r \times q$ finite matrix polynomial with $\text{rk}(\mathbf{Q}(1)) = q - d$ for some $d \geq 0$. In general, we allow $q \leq r$ so that the vector of common factors can be singular. We say that $\mathbf{F}_t$ is a rational reduced-rank $I(1)$ family with cointegration rank $c = r - q + d$ and it can be shown that, for generic values of the parameters, the shocks $\mathbf{u}_t$ are *fundamental* for $(1 - L)\mathbf{F}_t$, that is $\mathbf{u}_t$ belongs to the space spanned by $(1 - L)\mathbf{F}_s$ for $s \leq t$ (see Definition 4 and Proposition 4 in Barigozzi et al., 2016).

Finally, the dynamics of each idiosyncratic component is given by

$$(1 - \rho_i L)\xi_{it} = d_i(L)\varepsilon_{it}, \quad \varepsilon_{it} \overset{\text{i.i.d.}}{\sim} (0, \sigma_i^2), \quad i = 1, \ldots, n, \tag{3}$$

where $d_i(L)$ is a, possibly infinite, polynomial, and $|\rho_i| \leq 1$, thus allowing both for stationary and non-stationary idiosyncratic components. Moreover, we allow for cross-sectional dependence among the shocks ε_{it}'s and in this sense we speak of *generalized* or *approximate* factor structure as firstly defined by Chamberlain and Rothschild (1983).

The goal of this paper is to estimate model (1)-(3), with a particular focus on the impulse response functions of the variables $\mathbf{x}_t$ to the common shocks $\mathbf{u}_t$, which are defined as the entries of the $n \times q$ matrix polynomial

$$\boldsymbol{\Phi}(L) = \boldsymbol{\Lambda}[\mathbf{S}(L)(1 - L)]^{-1}\mathbf{Q}(L). \tag{4}$$

In order to estimate (4), we need to overcome two main difficulties. First, the common factors, $\mathbf{F}_t$, are not observed. Second, we need an estimator of the ARIMA parameters, $\mathbf{S}(L)$ and $\mathbf{Q}(L)$, in (2), with the constraint that $\text{rk}(\mathbf{Q}(1)) = q - d$, that is when $q - d$ shocks have a permanent effect on $\mathbf{x}_t$. Notice that the use of inverse polynomial matrix in (4), and in the following, is convenient and makes sense, provided that we do not forget that they are not square summable.

Before discussing estimation, it is worth to highlight the main features of model (1)-(3). First, we are mainly interested in the singular case $r > q$. Indeed, there exists large empirical evidence supporting singularity of the vector of common factors for US macroeconomic databases (see, for example, Giannone et al., 2005; Amengual and Watson, 2007; Forni and Gambetti, 2010; Luciani, 2015) and also for euro area datasets (see, for example, Barigozzi et al., 2014). Such results can be easily understood observing that the static equation (1) is just a convenient representation derived from a more "primitive" set of dynamic equations linking the common component $\boldsymbol{\chi}_t$ to the common shocks $\mathbf{u}_t$. Indeed, by substituting (2) into (1), and defining the dynamic factors $(1-L)\mathbf{f}_t = \mathbf{u}_t$, we have the fully-dynamic representation

$$x_{it} = \boldsymbol{\lambda}_i'[\mathbf{S}(L)]^{-1}\mathbf{Q}(L)\mathbf{f}_t + \xi_{it} = \boldsymbol{\phi}_i'(L)\mathbf{u}_t + \xi_{it}, \quad i = 1, \dots, n, \tag{5}$$

where $\boldsymbol{\phi}_i'(L)$ is the i-th row of $\boldsymbol{\Phi}(L)$ defined in (4). For a general analysis of the relationship between representation (1) and the "deeper" dynamic representation (5), we refer to Stock and Watson (2005), Bai and Ng (2007), and Forni et al. (2009). Moreover, a representation like (5) naturally arises when the model is estimated at a frequency which is lower than the one at which data are observed.[1] Therefore, singularity of $\mathbf{F}_t$, i.e. $r > q$, is assumed throughout the present paper. Nevertheless, all estimation results presented in Section 3 hold also when $r = q$.

Second, while there is not an agreement in the economic literature on what is the relative importance of demand side and supply side shocks in driving short-run macroeconomic fluctuations, there is agreement that in the long run what matters are supply side shocks. For example, there is agreement that monetary policy shocks generate only short run dynamics, while technology shocks generate stochastic trends. Since the decisions of the central bank affect the whole economy, and similarly does a given technological advancement, both monetary policy shocks and technology shocks are naturally assumed to be part of $\mathbf{u}_t$ when considering a macroeconomic panel. This line of reasoning has two clear implications for our model: (i) the vector of common factors must have some unit root, say $q - d$, (technology shocks induce common trends), while (ii) some shocks, say $d > 0$, must have just transitory effects on the observed variables (monetary policy shocks have no long run effects).

Third, the choice of allowing some idiosyncratic components to be $I(1)$ is also driven by a general macroeconomic argument. Consider the simplest case in which the factors are not singular ($r = q$) and are not cointegrated ($c = 0$). Then, every p-dimensional sub-vectors (with $p > r$) of the n-dimensional common-component vector $\boldsymbol{\chi}_t$ are trivially cointegrated and therefore stationarity of the idiosyncratic components would imply that all p-dimensional sub-vectors (with $p > r$) of the n-dimensional dataset $\mathbf{x}_t$ are cointegrated with cointegration rank $p-r$, a conclusion which is at odds with what is observed in the macroeconomic datasets typically analysed in the empirical Dynamic Factor model literature. The same reasoning applies, *a fortiori*, to the case in which the factors are cointegrated ($c > 0$). Then, under the assumption of stationarity of the idiosyncratic components, every p'-dimensional sub-vectors (with $p' > r - c$) of the n-dimensional dataset $\mathbf{x}_t$ would be cointegrated (see Proposition 5 in Barigozzi et al., 2016). Notice that with respect to the non-singular case we can have $p' < r$ hence cointegration can be found in even smaller subsets of variables.

The implausibility of a stationary idiosyncratic component is also confirmed empirically in Section 6 where about half of the estimated idiosyncratic components are found to be non-stationary according to the test of Bai and Ng (2004). This finding is for example related to the existence of sectoral trends which are not captured by economy-wide factors and are therefore

[1] To see this, let x_{it}^m be a non-stationary variable observed at month t, and suppose that the true model is $x_{it}^m = \boldsymbol{\lambda}_i' f_t^m + \xi_{it}^m$, where for semplicity $r = 1$. Now, if we estimate the model at quarterly frequency the correct model to be considered is $x_{it}^q = \boldsymbol{\lambda}_i'\mathbf{f}_t^q + \xi_{it}^q$, where $\mathbf{f}_t^q$ is of dimension 3×1 but of rank 1. Indeed, by using the approximation of Mariano and Murasawa (2003), we have that $x_{it}^q = (1 + L + L^2)x_{it}^m$, so that $\mathbf{f}_t^q = \boldsymbol{\Lambda}(1 + L + L^2)f_t^m$.

idiosyncratic. For all these reasons, we consider stationarity of all idiosyncratic components a too strong restriction.

Summing up, we are imposing three requirements on our model: (i) $r > q$, (ii) $d > 0$, and (iii) $\boldsymbol{\xi}_t \sim I(1)$. Apart from these three requirements, which are either driven by economic theory, or by stylized facts observed on macroeconomic databases, we are not imposing any particular constraint neither on the law of motion of $\mathbf{F}_t$, nor on the law of motion of $\boldsymbol{\xi}_t$. Indeed, (2) describes a generic multivariate ARIMA process for the factors, and (3) describes a generic ARMA process, possibly with a unit root, for the idiosyncratic components.

2.2 Representation results

Let us for the moment assume to know the common factors and let us focus on parameters' estimation. If we had $r = q$, then Engle and Granger (1987) prove that there exists a VECM representation for $\mathbf{F}_t$ with d cointegration relations. However, in this non-singular case the VECM representation is motivated only as an approximation to an infinite autoregressive model with exponentially declining coefficients. Moreover, when $r = q$ the shocks $\mathbf{u}_t$ driving (2) might be non-fundamental, in which case the estimated VECM residuals will not span the same space as the space spanned by $\mathbf{u}_t$ (see e.g. Alessi et al., 2011, for some examples).

On the other hand, in the singular case, $r > q$, the shocks $\mathbf{u}_t$ are generically fundamental, as we said above, and the following Proposition gives us the correct autoregressive representation for the vector of common factors, which is the starting point for estimating the impulse response functions.

Proposition 1 ***(Granger Representation Theorem for reduced-rank I(1) vectors)*** *Assume that $\mathbf{F}_t$ is a rational reduced-rank $I(1)$ family with cointegration rank $c = r - q + d$, then, for generic values of the parameters in (2): (i) there exists a $r \times c$ full-rank matrix $\boldsymbol{\beta}$ such that $\boldsymbol{\beta}'\mathbf{F}_t$ is weakly stationary with rational spectral density and (ii) $\mathbf{F}_t$ has the VECM representation*

$$\mathbf{G}(L)(1 - L)\mathbf{F}_t = \mathbf{h} + \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{F}_{t-1} + \mathbf{K}\mathbf{u}_t, \tag{6}$$

where $\mathbf{G}(L)$ is an $r \times r$ matrix polynomial of finite degree p with $\mathbf{G}(0) = \mathbf{I}_r$, $\mathbf{h}$ is an $r \times 1$ constant vector, $\boldsymbol{\alpha}$ is a full-rank $r \times c$ matrix, and $\mathbf{K}$ is $r \times q$ with $\mathbf{K} = \mathbf{S}(0)^{-1}\mathbf{Q}(0)$.

Proof: see Section 3 and Proposition 3 in Barigozzi et al. (2016).

Two comments are necessary. First, with respect to the classical case in which $r = q$, notice that, while the number of permanent shocks, $q - d$, is obtained as usual as r minus the cointegration rank, the number of transitory shocks, d, is obtained as the complement of the number of permanent shocks to q, not to r, as though $r - q$ transitory shocks had a zero coefficient.

Second, by rewriting (6) as a VAR, we have

$$\mathbf{A}(L)\mathbf{F}_t = \mathbf{h} + \mathbf{K}\mathbf{u}_t, \tag{7}$$

where $\mathbf{A}(L)$ is a matrix polynomial of degree $p+1$ with $\mathbf{A}(0) = \mathbf{I}_r$, $\mathbf{A}_1 = (\mathbf{G}_1 + \boldsymbol{\alpha}\boldsymbol{\beta}' + \mathbf{I}_r)$, $\mathbf{A}_i = \mathbf{G}_i - \mathbf{G}_{i-1}$ for $i = 2, \dots, p$, and $\mathbf{A}_{p+1} = -\mathbf{G}_p$, hence $\mathbf{A}(1) = -\boldsymbol{\alpha}\boldsymbol{\beta}'$ and we have $r - c = q - d$ unit roots. From (7) the impulse response functions (4) are equivalently written as

$$\boldsymbol{\Phi}(L) = \boldsymbol{\Lambda}\mathbf{A}(L)^{-1}\mathbf{K}. \tag{8}$$

The main contribution of this paper is to provide a consistent estimator of (8) when the common factors, $\mathbf{F}_t$, are not observed.

To conclude, it has to be noticed that when $d = 0$ and $r > q$, although all q shocks have a permanent effect on the variables, we still have cointegration. In this case, Anderson and Deistler (2008a,b) prove that generically there exists a finite degree left inverse of $\mathbf{Q}(L)$ such that (2) can be written as a VAR in first differences

$$\mathbf{B}(L)(1 - L)\mathbf{F}_t = \mathbf{K}\mathbf{u}_t, \tag{9}$$

where $\mathbf{B}(L)$ is a stable matrix polynomial of finite degree and $\mathbf{K} = \mathbf{S}(0)^{-1}\mathbf{Q}(0)$ (see also Proposition 1 in Barigozzi et al., 2016). In this case the impulse response functions are given by

$$\mathbf{\Phi}(L) = \mathbf{\Lambda}[\mathbf{B}(L)(1 - L)]^{-1}\mathbf{K}, \tag{10}$$

from which is clear that q common shocks have permanent effect on the variables. In this case, estimation of (10) is straightforward: (i) reduce the data $\mathbf{x}_t$ to stationarity by taking first differences, (ii) estimate the differenced factors $(1 - L)\mathbf{F}_t$ and their loadings $\mathbf{\Lambda}$ by means of the principal components of $(1 - L)\mathbf{x}_t$, and (iii) estimate a VAR like (9) for $(1 - L)\mathbf{F}_t$. We refer to Forni et al. (2009) for the asymptotic properties of this estimator.

2.3 Assumptions

We now introduce a set of formal assumptions that allow us to better characterize the Non-Stationary Dynamic Factor model.[2] We consider an n-dimensional stochastic process $\{\mathbf{x}_t = (x_{1t} \dots x_{nt})' : n \in \mathbb{N},\ t \in \mathbb{Z}\}$ described by equations (1)-(3). We assume that $\mathbf{x}_t$ and all other stochastic variables in this paper belong to the Hilbert space $L_2(\Omega, \mathcal{A}, P)$, where $(\Omega, \mathcal{A}, P)$ is some given probability space. Moreover, since the asymptotic results given in Section 3 require $n \to \infty$, all the following assumptions hold for any $n \in \mathbb{N}$. Hereafter, we denote, for ease of notation, the first differences of $\mathbf{x}_t$ as $\Delta\mathbf{x}_t \equiv (1 - L)\mathbf{x}_t$, and similarly for $\Delta\mathbf{F}_t$, $\Delta\boldsymbol{\chi}_t$, and $\Delta\boldsymbol{\xi}_t$. Finally, we denote by $M_1, M_2 \dots$ generic positive finite constants.

First, we assume some general properties for the observed panel $\mathbf{x}_t$.

Assumption 1 ***(Observables)***
a. $\Delta\mathbf{x}_t$ *has rational spectral density;*
b. $\mathbf{x}_t \sim I(1)$*;*
c. $\mathsf{E}[\Delta x_{it}] = 0$ *for any* $i \in \mathbb{N}$.

Assumption 1b specifies that the vector of observables is non-stationary, but can have some $I(0)$ coordinates. For simplicity, in part c we exclude deterministic trends, while in principle x_{it} can have a constant non-zero mean. Since deterministic trends might affect estimation this case is discussed at the end of Section 3 and its economic interpretation is given in Section 6.

It is convenient to re-write (2) as

$$\Delta\mathbf{F}_t = \mathbf{C}(L)\mathbf{u}_t, \tag{11}$$

where $\mathbf{C}(L) = \mathbf{S}(L)^{-1}\mathbf{Q}(L)$ is an $r \times q$ infinite matrix polynomial, with $q < r$. Notice that $\mathsf{E}[\Delta\mathbf{F}_t] = \mathbf{0}$, by construction. The properties of (11) are described in the next assumption.

[2] Similar assumptions and results to those given in this section can be found, for example, in Stock and Watson (2002a), Bai and Ng (2002), Forni et al. (2009) in a stationary setting, and in Bai and Ng (2004) and Bai (2004) in a non-stationary framework.

Assumption 2 ***(Common Shocks)***
a. $u_{it} \stackrel{i.i.d.}{\sim} (0,1)$ *and* $\mathsf{E}[u_{it}^4] < \infty$*, for any* $i = 1, \ldots, q$*;*
b. if $i \neq j$ *then* u_{it} *and* u_{js} *are independent for any* $t, s \in \mathbb{Z}$ *and* $i, j = 1, \ldots, q$*;*
c. $u_{it} = 0$ *for any* $t \leq 0$ *and* $i = 1, \ldots, q$*;*
d. $\mathbf{C}(L) = \sum_{k=0}^{\infty} \mathbf{C}_k L^k$ *and* $\sum_{k=0}^{\infty} k\|\mathbf{C}_k\| \leq M_2 < \infty$*;*
e. $\mathbf{C}(0)'\mathbf{C}(0)$ *is positive definite;*
f. $\mathrm{rk}(\mathbf{C}(1)) = q - d$ *with* $0 < d < q$*;*
g. $\mathrm{rk}(\sum_{k=0}^{\infty} \mathbf{C}_k \mathbf{C}_k') = r$*.*

Assumptions 2a and 2b imply that $\mathsf{E}[\mathbf{u}_t\mathbf{u}_t'] = \mathbf{I}_q$, and that $\mathbf{u}_t$ and $\mathbf{u}_{t-k}$ are independent for any $k \neq 0$, hence $\mathbf{u}_t$ is an orthonormal strong white noise process with finite fourth moments. In part c, for simplicity and without loss of generality, we fix initial conditions to zero, in particular this implies $\mathbf{F}_0 = \mathbf{0}$, and therefore $\mathsf{E}[\mathbf{F}_t] = \mathbf{0}$, and in Proposition 1 we have $\mathbf{h} = \mathbf{0}$. Part d implies square summability of the coefficients of each entry of the matrix polynomial, while part e allows for estimation of $\mathbf{C}(0)$. Parts d and g guarantee that all eigenvalues of the covariance matrix of the common factors are finite and positive.[3]

From Assumptions 1 and 2, we can derive the following properties of the common factors.

Remark 1 It is possible to prove that: (i) the sample covariance matrix of $\Delta\mathbf{F}_t$ is a consistent estimator of the covariance $\mathsf{E}[\Delta\mathbf{F}_t\Delta\mathbf{F}_t']$ and (ii) the usual asymptotic results for $I(1)$ processes by Phillips and Durlauf (1986) and Phillips and Solo (1992) hold for $\mathbf{F}_t$ (for a proof see Lemma 8 and 9 in Appendix B, respectively).

Remark 2 The common factors, $\mathbf{F}_t$, are a rational reduced rank $I(1)$ family with cointegration rank $c = r - q + d$. Therefore, $\mathbf{F}_t$ satisfy Proposition 1, that is they admit the VECM representation in (6) with c cointegration vectors given by the columns of an $r \times c$ full-rank matrix $\boldsymbol{\beta}$. Moreover, $\mathbf{F}_t$ admit the common trends decomposition (Stock and Watson, 1988)

$$\mathbf{F}_t = \mathbf{C}(1)\sum_{s=0}^{\infty}\mathbf{u}_{t-s} + \check{\mathbf{C}}(L)\mathbf{u}_t = \boldsymbol{\psi}\boldsymbol{\eta}'\sum_{s=0}^{\infty}\mathbf{u}_{t-s} + \check{\mathbf{C}}(L)\mathbf{u}_t, \tag{12}$$

where $\boldsymbol{\psi}$ is $r \times q - d$, $\boldsymbol{\eta}$ is $q \times q - d$, and $\check{\mathbf{C}}(L)$ is an $r \times q$ infinite matrix polynomial (and also square summable because of Assumption 2d). The first term in (12) contains the $q - d$ common trends. Alternatively, we can write $\mathbf{F}_t$ as driven by $q - d$ shocks, defined as $\boldsymbol{\eta}'\mathbf{u}_t$, with permanent effects, and by d transitory shocks, defined as $\boldsymbol{\eta}_\perp'\mathbf{u}_t$, where $\boldsymbol{\eta}_\perp$ is $q \times d$ and $\boldsymbol{\eta}_\perp'\boldsymbol{\eta} = \mathbf{0}_{d\times q-d}$. Using these definitions we can represent $\mathbf{F}_t$ as a sum of a permanent and a transitory component, where the former contains again the common trends in (12) (see also Section 3.2 in Barigozzi et al. (2016)). It is also straightforward to see that other well known "permanent-transitory" decompositions can be derived in our setting as, for example, those proposed by Johansen (1991), Vahid and Engle (1993), Escribano and Peña (1994), Gonzalo and Granger (1995), and Gonzalo and Ng (2001), among others.

Remark 3 Since $\mathbf{F}_t$ are not identified in the model, then, if we define $\mathbf{F}_t^* = \boldsymbol{H}^{-1}\mathbf{F}_t$ for an $r \times r$ invertible matrix $\boldsymbol{H}$, we still have the same factor model as (1):

$$x_{it} = \boldsymbol{\lambda}_i^{*\prime}\mathbf{F}_t^* + \xi_{it}, \quad i = 1, \ldots, n, \qquad \Delta\mathbf{F}_t^* = \mathbf{C}^*(L)\mathbf{u}_t, \tag{13}$$

[3]Obviously part g is compatible with (11) since the rank of sum of matrices is smaller or equal than the sum of the ranks and here we are summing infinite of those matrices.

where $\boldsymbol{\lambda}_i^{*\prime} = \boldsymbol{\lambda}_i' \boldsymbol{H}$ and $\mathbf{C}^*(L) = \boldsymbol{H}^{-1}\mathbf{C}(L)$. In particular, if $\boldsymbol{H}^{-1} = (\boldsymbol{\psi}_\perp' \ \boldsymbol{\psi}')'$, where $\boldsymbol{\psi}$ is defined in (12), the first c coordinates of $\mathbf{F}_t^*$ are $I(0)$ and the remaining $r-c$ are $I(1)$. Moreover, also $\mathbf{F}_t^*$ satisfies Proposition 1 with cointegration vectors given by the columns of the $r \times c$ matrix $\boldsymbol{\beta}^* = \boldsymbol{H}^{-1}\boldsymbol{\beta}$, such that the error vector $\boldsymbol{\beta}^{*\prime}\mathbf{F}_{t-1}^*$ is made of linear combinations of the $I(0)$ factors alone. This reasoning shows that our model is compatible with the presence of some stationary common factors. On the other hand, since $\mathbf{F}_t$ have no economic meaning, we can always assume without loss of generality that every coordinate of $\mathbf{F}_t$ is $I(1)$.

The properties of the common component are completely characterized by Assumption 2 and the following identifying assumption on factor loadings.

Assumption 3 ***(Loadings)***
a. $n^{-1}\boldsymbol{\Lambda}'\boldsymbol{\Lambda} \to \mathbf{V}$, *as* $n \to \infty$, *where* $\mathbf{V}$ *is* $r \times r$, *positive definite;*
b. the entries λ_{ij} *of* $\boldsymbol{\Lambda}$ *are such that* $\sup_{i\in\mathbb{N}} \max_{j=1,\dots,r} |\lambda_{ij}| \le M_1 < \infty$.

Since the common factors are not identified, then also the loadings are not identified. However, consistently with Assumption 3a, and without any loss of generality, we can always impose the identifying restriction $n^{-1}\boldsymbol{\Lambda}'\boldsymbol{\Lambda} = \mathbf{I}_r$, which is therefore assumed throughout the rest of the paper. Under this restriction, we can estimate the loadings up to an orthogonal transformation and we can immediately derive the asymptotic behaviour of the eigenvalues of the covariance and spectral density matrices of $\Delta\boldsymbol{\chi}_t$ (see respectively Lemma 2 and 7, below). Moreover, Assumption 3 implies that the common factors have a pervasive effect on all series. In particular, since $\boldsymbol{\Lambda}$ has full-rank and given the decompositon in (12) and the definition of impulse responses in (8), we see that only $q-d$ common shocks can have a permanent effect on the observed variables $\mathbf{x}_t$. Still, it is always possible to have some loadings cancelling the long-run effect of permanent shocks on the observed variables, hence, our model is in principle compatible with the presence of stationary components in $\mathbf{x}_t$, in agreement with Assumption 1b.

The model for the idiosyncratic component, (3), can be written in vector notation as

$$(\mathbf{I}_n - \mathbf{P}L)\boldsymbol{\xi}_t = \mathbf{D}(L)\boldsymbol{\varepsilon}_t, \tag{14}$$

where $\mathbf{P}$ is an $n \times n$ diagonal matrix with generic element ρ_i, and $\mathbf{D}(L)$ is an $n \times n$ diagonal matrix polynomial, while the elements of $\boldsymbol{\varepsilon}_t$ are allowed to be weakly dependent as specified below. Notice that $\mathsf{E}[\Delta\boldsymbol{\xi}_t] = \mathbf{0}$, by construction. Model (14) is completely characterized by the next assumption.

Assumption 4 ***(Idiosyncratic Components)***
a. For any $n \in \mathbb{N}$, *there exists an* $m \in \mathbb{N}$ *with* $m = O(n^\delta)$ *and* $\delta \in [0,1]$, *which partitions* $\mathbf{P}$ *in two diagonal blocks of size* m *and* $n-m$, *such that* $\rho_i = 1$ *if* $i \le m$ *and* $|\rho_i| < 1$ *otherwise;*
b. $\varepsilon_{it} = 0$ *for any* $t \le 0$ *and for any* $i \in \mathbb{N}$;
c. $\mathbf{D}(L)$ *is diagonal with entries* $d_i(L) = \sum_{k=0}^\infty d_{ik}L^k$ *with* $\sup_{i\in\mathbb{N}} \sum_{k=0}^\infty k|d_{ik}| \le M_3 < \infty$;
d. $\varepsilon_{it} \overset{i.i.d.}{\sim} (0, \sigma_i^2)$ *with* $0 < \sigma_i^2 < \infty$ *and* $\mathsf{E}[|\varepsilon_{it}|^{\kappa_1}|\varepsilon_{jt}|^{\kappa_2}] < \infty$ *for any* $\kappa_1 + \kappa_2 = 4$ *and* $i, j \in \mathbb{N}$;
e. if $s \ne t$ *and* $i \ne j$ *then* ε_{it} *and* ε_{js} *are independent for any* $s, t \in \mathbb{Z}$ *and* $i, j \in \mathbb{N}$;
f. $\max_{j=1,\dots,n} \sum_{i=1}^n |\mathsf{E}[\varepsilon_{it}\varepsilon_{jt}]| \le M_4 < \infty$ *for any* $n \in \mathbb{N}$.

Assumption 4 characterizes the behavior of the idiosyncratic component as well as the properties of the vector of idiosyncratic shocks $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t} \dots \varepsilon_{nt})'$. In part a, we allow for $m \le n$

idiosyncratic components to be $I(1)$ while all others are $I(0)$. In particular, m can grow with the cross-sectional dimension n. In part b, for simplicity and without loss of generality, we fix initial conditions to zero, which implies $\boldsymbol{\xi}_0 = \mathbf{0}$, and therefore $\mathsf{E}[\boldsymbol{\xi}_t] = \mathbf{0}$. Part c implies square summability of the matrix polynomials in (3) so that ξ_{it} is non-stationary if and only if $\rho_i = 1$. In parts d and e we exclude serial-dependence in $\boldsymbol{\varepsilon}_t$, but we do not rule out cross-sectional dependence in $\boldsymbol{\varepsilon}_t$, which indeed is possible when two shocks are contemporaneous, as shown in part *f*. Specifically, we limit the size of cross-sectional dependence by bounding the column norm of the covariance matrix of idiosyncratic shocks, thus requiring a mild form of sparsity as proposed by Fan et al. (2013) and found empirically in a stationary setting by Boivin and Ng (2006), Bai and Ng (2008), and Luciani (2014).[4] As an immediate consequence of part *f*, we have the following Lemma which provides a bound for the eigenvalues, μ_j^ε, of the covariance matrix of idiosyncratic shocks.

Lemma 1 *Under Assumptions 4d, e, and f,* $\mu_1^\varepsilon \leq M_4 < \infty$ *and* $n^{-1}\sum_{i=1}^n \sum_{j=1}^n |\mathsf{E}[\varepsilon_{it}\varepsilon_{jt}]| \leq M_4 < \infty$ *for any* $n \in \mathbb{N}$.

Proof: see Appendix A.

Therefore, given Lemma 1 and the dynamic loadings of $\boldsymbol{\varepsilon}_t$, the components of $\Delta\boldsymbol{\xi}_t$ are allowed to be both cross-sectionally and serially correlated. In particular, the spectral density matrix of $\Delta\boldsymbol{\xi}_t$ is non-diagonal with bounded eigenvalues (see also Lemma 7 in Section 4 below) and we say that $\Delta\mathbf{x}_t$ has an *approximate* or *generalized dynamic* factor structure, as the one originally considered by Forni and Lippi (2001) and Forni et al. (2000).

We then impose independence of common and idiosyncratic shocks.

Assumption 5 u_{jt} *and* ε_{is} *are independent for any* $j = 1, \dots, q$, $i \in \mathbb{N}$, *and* $t, s \in \mathbb{Z}$.

This requirement is in agreement with the economic interpretation of the model for which common and idiosyncratic shocks are two independent sources of variation. However, from a technical point of view we could easily relax this assumption by requiring only weak dependence (see, for example, Assumption D in Bai and Ng, 2002, but in a stationary setting). Moreover, as an immediate consequence of Assumption 5, we have $\mathsf{E}[\Delta\chi_{it}\Delta\xi_{js}] = 0$ for any $i, j \in \mathbb{N}$ and $t, s \in \mathbb{Z}$.

To conclude, the following Lemma shows the asymptotic behaviour of the eigenvalues, $\mu_j^{\Delta x}$, $\mu_j^{\Delta\chi}$, and $\mu_j^{\Delta\xi}$, of the covariance matrices for the model in first differences.

Lemma 2 *Under Assumptions 1-5, and for any* $n \in \mathbb{N}$,

i. $0 < \underline{M}_5 \leq n^{-1}\mu_j^{\Delta\chi} \leq \overline{M}_5 < \infty$ *for any* $j = 1, \dots, r$;

ii. $\mu_1^{\Delta\xi} \leq M_6 < \infty$;

iii. $0 < \underline{M}_7 \leq n^{-1}\mu_j^{\Delta x} \leq \overline{M}_7 < \infty$ *for any* $j = 1, \dots, r$;

iv. $\mu_{r+1}^{\Delta x} \leq M_6 < \infty$.

[4] This assumption can be relaxed by directly assuming the results in Lemma 1. Moreover, notice that, while here we model cross-sectional dependence of idiosyncratic components, $\Delta\boldsymbol{\xi}_t$, via the cross-sectional dependence in the shocks $\boldsymbol{\varepsilon}_t$, we could equivalently consider (3) as driven by cross-sectionally independent idiosyncratic shocks having dynamic loadings with dependence structure analogous to the one assumed in part f for $\boldsymbol{\varepsilon}_t$. This alternative representation is adopted for example in Forni et al. (2015a,b).

Proof: see Appendix A.

The results in Lemma 2 imply that from the model in first differences we can disentangle the common component from the idiosyncratic component, for example by means of principal component analysis. In this way we can estimate the number of common factors and reconstruct consistently the space spanned by the loadings, which constitutes the starting point also for estimating the model in levels. In the same way, in Section 4, by studying the spectral density matrix of $\Delta\mathbf{x}_t$, which is the sum of a common and an idiosyncratic spectral density, and the corresponding eigenvalues, we can determine the number of common trends driving the model. Finally, notice that linear divergence of eigenvalues corresponds to the very natural idea that the influence of the common factors is in some sense "stationary along the cross-section", which seems to be a quite sensible assumption.

Lastly, in Assumption 7 in Appendix A we require the eigenvalues of the covariance end spectral density matrices of $\Delta\boldsymbol{\chi}_t$ to be distinct.

3 Estimation

We now turn to estimation of the Non-Stationary Dynamic Factor model presented in the previous section. We assume to observe an n-dimensional vector $\mathbf{x}_t$ with sample size $T+1$, i.e. we observe a $n \times (T+1)$ panel $\mathbf{x} = (\mathbf{x}_0 \dots \mathbf{x}_T)$.[5] Since we consider large datasets, we focus on the case in which both the cross-sectional dimension n and the sample size T are large so double asymptotics is considered. In particular, following Stock and Watson (2002a), we require that $n, T \to \infty$ jointly or, equivalently, that $n = n(T)$ with $\lim_{T\to\infty} n(T) = \infty$. Finally, in this section we assume to know the number of common factors r, of common shocks q, and of the cointegration relations $c = r - q + d$. We refer to the next section for a discussion on how to determine these quantities.

The estimation method we propose is analogous to the one proposed by Forni et al. (2009) and Stock and Watson (2005) in a stationary setting and it is based on three steps: (i) we extract the common factors and their loadings, (ii) we estimate the law of motion of the factors by exploiting their autoregressive representation given in Proposition 1, and (iii) we recover the space spanned by the common shocks and, if needed, we identify them by imposing a suitable set of restrictions based on economic theory.

Throughout, we denote estimated quantities with $\widehat{\ }$, where these depend on both n and T and we denote the spectral norm of a generic matrix $\mathbf{B}$ as $\|\mathbf{B}\| = (\mu_1^{\mathbf{B}'\mathbf{B}})^{1/2}$, where $\mu_1^{\mathbf{B}'\mathbf{B}}$ is the largest eigenvalue of $\mathbf{B}'\mathbf{B}$.

3.1 Loadings and common factors

It is clear that, by taking the first difference of the factor model (1), the loadings, $\boldsymbol{\Lambda}$, are unchanged and can, therefore, be estimated by means of principal component analysis on $\Delta\mathbf{x}_t$, as it is commonly done in the stationary case.

Define the $n \times T$ data matrix $\Delta\mathbf{x} = (\Delta\mathbf{x}_1 \dots \Delta\mathbf{x}_T)$. The estimated loadings are given by $n^{1/2}$-times the first r normalized eigenvectors of the $n\times n$ sample covariance matrix $T^{-1}\Delta\mathbf{x}\Delta\mathbf{x}'$. Consistency of this estimator is in the following Lemma.

[5] Notice that while we set $\mathbf{x}_0 = \mathbf{0}$ in order to simplify notation in the proofs of the following results, this needs not to be imposed in practice.

Lemma 3 *Under Assumptions 1-5 and 7a, and if* $n^{-1}\boldsymbol{\Lambda}'\boldsymbol{\Lambda} = \mathbf{I}_r$*, there exists an* $r \times r$ *orthogonal matrix* $\mathbf{H}$ *such that, as* $n, T \to \infty$*: (i)* $n^{-1/2}\|\widehat{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda}\mathbf{H}'\| = O_p(\max(n^{-1}, T^{-1/2}))$ *or, equivalently, (ii)* $\|n^{-1}\widehat{\boldsymbol{\Lambda}}'\boldsymbol{\Lambda}\mathbf{H}' - \mathbf{I}_r\| = O_p(\max(n^{-1}, T^{-1/2}))$*, and (iii)* $\|\widehat{\boldsymbol{\lambda}}_i' - \boldsymbol{\lambda}_i'\mathbf{H}'\| = O_p(\max(n^{-1/2}, T^{-1/2}))$*, for any* $i = 1, \ldots, n$.

Proof: see Appendix A.

In principle the identifying matrix, $\mathbf{H}$, could be any invertible matrix. However, since by construction our estimator is such that $n^{-1}\widehat{\boldsymbol{\Lambda}}'\widehat{\boldsymbol{\Lambda}} = \mathbf{I}_r$, we can consistently restrict the true loadings matrix, $\boldsymbol{\Lambda}$, to be orthogonal and, as a consequence, the choice of $\mathbf{H}$ is limited to orthogonal matrices only. Nevertheless, all following results would equally apply for any invertible matrix, $\mathbf{H}$.

Finally, notice that, when the ratio n/T is non-negligible, then in general the sample covariance matrix is not a consistent estimator of the covariance matrix of $\Delta\mathbf{x}_t$. As a consequence, also the eigenvectors might not be consistently estimated (see e.g. Johnstone and Lu, 2009). However, when, as shown in Lemma 2, we are in the presence of r very spiked, diverging, eigenvalues, then we can show that the corresponding r eigenvectors of the sample covariance matrix asymptotically still span the same space as the loadings. The proof of Lemma 3 is based on this property and in this respect is similar to the approach proposed by Fan et al. (2013).

Given the estimated loadings, the common factors can be estimated by projecting, at any given point in time $t = 0, \ldots, T$, the data $\mathbf{x}_t$ onto the space spanned by the estimated loadings

$$\widehat{\mathbf{F}}_t = (\widehat{\boldsymbol{\Lambda}}'\widehat{\boldsymbol{\Lambda}})^{-1}\widehat{\boldsymbol{\Lambda}}'\mathbf{x}_t = \frac{\widehat{\boldsymbol{\Lambda}}'\mathbf{x}_t}{n}. \tag{15}$$

Notice that being (15) a cross-section regression, no spurious regression issues arise even thou some idiosyncratic components are $I(1)$. The estimator of the differenced factors is then defined as $\Delta\widehat{\mathbf{F}}_t \equiv (1 - L)\widehat{\mathbf{F}}_t$. The estimator in (15) is new to this paper and we have the following consistency result for the space spanned by the factors.

Lemma 4 *Under Assumptions 1-5 and 7a, and given* $\mathbf{H}$ *defined in Lemma 3, as* $n, T \to \infty$ *and for any* $t = 1, \ldots, T$*: (i)* $\|\Delta\widehat{\mathbf{F}}_t - \mathbf{H}\Delta\mathbf{F}_t\| = O_p(\max(n^{-1/2}, T^{-1/2}))$ *and (ii)* $T^{-1/2}\|\widehat{\mathbf{F}}_t - \mathbf{H}\mathbf{F}_t\| = O_p(\max(n^{-1/2}, T^{-1/2}))$.

Proof: see Appendix A.

Lemma 4 is of interest in itself, but in order to derive our main results, we need to study also the asymptotic properties of the sample second moments of (15). This is done in Lemma 13 in Appendix B, which contains several results among which here it is worth mentioning two: as $n, T \to \infty$,

i. $\|T^{-1}\sum_{t=1}^T \Delta\widehat{\mathbf{F}}_t\Delta\widehat{\mathbf{F}}_t' - \mathsf{E}[\Delta\mathbf{F}_t\Delta\mathbf{F}_t']\| = O_p(\max(n^{-1/2}, T^{-1/2}))$;
ii. $T^{-2}\sum_{t=1}^T \widehat{\mathbf{F}}_t\widehat{\mathbf{F}}_t' \stackrel{d}{\to} \int_0^1 \mathbf{W}(\tau)\mathbf{W}'(\tau)\mathrm{d}\tau$, where $\mathbf{W}(\cdot)$ is an r-dimensional random walk with positive definite and finite covariance matrix.

Results equivalent to those in Lemma 4 are also in Bai and Ng (2004), who estimate the factors in levels by integrating the factors estimated in the model in first differences. This is a perfectly valid alternative to our approach, but, when dealing with deterministic trends, it might produce significant finite sample differences, as briefly described at the end of this section.

Finally, let us recall that, if all idiosyncratic components were stationary ($m = 0$ in Assumption $4a$), then, following Bai (2004), we could estimate the factors by means of principal component analysis applied directly on $\mathbf{x}_t$. However, for the reasons already discussed, we do not consider this case further.

3.2 VECM for the common factors

In line with the result in Proposition 1, we then consider a VECM with $c = r - q + d$ cointegration relations for the common factors

$$\Delta \mathbf{F}_t = \boldsymbol{\alpha\beta}' \mathbf{F}_{t-1} + \mathbf{G}_1 \Delta \mathbf{F}_{t-1} + \mathbf{w}_t, \quad \mathbf{w}_t = \mathbf{K}\mathbf{u}_t, \tag{16}$$

where, for simplicity, we consider the case of one lag, $p = 1$, and we set $\mathbf{h} = \mathbf{0}$ as a consequence of Assumption $2c$.

Many different estimators for the cointegration vector, $\boldsymbol{\beta}$, are possible. As suggested by the asymptotic and numerical studies in Phillips (1991) and Gonzalo (1994), we opt for the estimation approach proposed by Johansen (1988, 1991, 1995). Although typically derived from the maximization of a Gaussian likelihood, this estimator is nothing else but the solution of an eigen-problem naturally associated to a reduced rank regression model, where no specific assumption about the distribution of the errors is made in order to establish consistency (see e.g. Velu et al., 1986).[6]

Since $\mathbf{F}_t$ are unknown, we estimate the parameters of (16) by using the estimated factors $\widehat{\mathbf{F}}_t$ instead. Denote as $\widehat{\mathbf{e}}_{0t}$ and $\widehat{\mathbf{e}}_{1t}$ the residuals of the regressions of $\Delta\widehat{\mathbf{F}}_t$ and of $\widehat{\mathbf{F}}_{t-1}$ on $\Delta\widehat{\mathbf{F}}_{t-1}$, respectively, and define the matrices $\widehat{\mathbf{S}}_{ij} = T^{-1} \sum_{t=1}^{T} \widehat{\mathbf{e}}_{it}\widehat{\mathbf{e}}_{jt}'$. Then, the c cointegration vectors are estimated as the normalized eigenvectors corresponding to the c largest eigenvalues $\widehat{\mu}_j$, such that

$$(\widehat{\mathbf{S}}_{11} - \widehat{\mathbf{S}}_{10}\widehat{\mathbf{S}}_{00}^{-1}\widehat{\mathbf{S}}_{01})\widehat{\boldsymbol{\beta}}_j = \widehat{\mu}_j \widehat{\boldsymbol{\beta}}_j, \quad j = 1, \dots, c.$$

The vectors $\widehat{\boldsymbol{\beta}}_j$ are then the c columns of the estimator $\widehat{\boldsymbol{\beta}}$. In a second step, the other parameters of the VECM, $\boldsymbol{\alpha}$ and $\mathbf{G}_1$, are estimated by regressing $\Delta\widehat{\mathbf{F}}_t$ on $\widehat{\boldsymbol{\beta}}'\widehat{\mathbf{F}}_{t-1}$ and $\Delta\widehat{\mathbf{F}}_{t-1}$.

Finally, a linear combination of the q columns of $\mathbf{K}$ can be estimated by means of the first q eigenvectors of the sample covariance matrix of the VECM residuals $\widehat{\mathbf{w}}_t$, rescaled by the square root of their corresponding eigenvalues (see Stock and Watson, 2005; Bai and Ng, 2007; Forni et al., 2009, for analogous definitions). This estimator is denoted as $\widehat{\mathbf{K}}$.

If the factor were observed then the asymptotic properties of the estimated VECM parameters are well known. On the other hand, in the present case of estimated factors, we must require the following additional regularity conditions in order to have consistency.

Assumption 6
a. $Tn^{-(2-\delta)} \to 0$, *as* $n, T \to \infty$*;*
b. for any $n \in \mathbb{N}$, $m = O(n^{\delta})$ *with* $\delta \in [0, 1)$*;*

[6] Other existing estimators of the cointegration vector, not considered here, are, for example: ordinary least squares (Engle and Granger, 1987), non-linear least squares (Stock, 1987), principal components (Stock and Watson, 1988), instrumental variables (Phillips and Hansen, 1990), and dynamic ordinary least squares (Stock and Watson, 1993).

c. $n^{-\gamma}\sum_{i=1}^m\sum_{j=m+1}^n|\mathsf{E}[\varepsilon_{it}\varepsilon_{jt}]|\le M_8<\infty$ *with* $\gamma<\delta$ *for any* $n\in\mathbb{N}$.

Assumption 6a puts a constraint on the relative rates of n and T and it implies that at least $T^{1/2}/n\to 0$ (when $\delta=0$). Part b is equivalent to assuming the existence of some stationary idiosyncratic components. Further motivations for, and the implications of, these two requirements are given in Remark 6 below. Finally, with reference to the partitioning of the vector of idiosyncratic components into $I(1)$ and $I(0)$ coordinates, part c limits the dependence between the two blocks more than the dependence within each block, which is in turn given in Lemma 1.[7]

Assumption 6 has two immediate consequences for the estimated loadings and factors. First, part a implies that in Lemma 3i we always have $T^{1/2}$-consistency for the estimated loadings matrix, $\widehat{\mathbf{\Lambda}}$. Second, parts b and c imply that the rate of convergence of the estimated factors, $\widehat{\mathbf{F}}_t$, in Lemma 4ii is improved from $\min(n^{1/2},T^{1/2})$ to $\min(n^{(2-\delta)/2},T^{1/2})$, with the fastest possible rate being $\min(n,T^{1/2})$, when $\delta=0$ (this can be proved straightforwardly by applying the results of Lemma 10 in Appendix B).

We then have consistency of the estimated VECM parameters.

Lemma 5 *Define* $\vartheta_{nT,\delta}=\max\left(T^{1/2}n^{-(2-\delta)/2},n^{-(1-\delta)/2},T^{-1/2}\right)$. *Under Assumptions 1-7a, and given* $\mathbf{H}$ *defined in Lemma 3, there exist a* $c\times c$ *orthogonal matrix* $\mathbf{Q}$ *and a* $q\times q$ *orthogonal matrix* $\mathbf{R}$, *such that, as* $n,T\to\infty$,

i. $\|\widehat{\boldsymbol{\beta}}-\mathbf{H}\boldsymbol{\beta}\mathbf{Q}\|=O_p(T^{-1/2}\vartheta_{nT,\delta})$;
ii. $\|\widehat{\boldsymbol{\alpha}}-\mathbf{H}\boldsymbol{\alpha}\mathbf{Q}\|=O_p(\vartheta_{nT,\delta})$;
iii. $\|\widehat{\mathbf{G}}_1-\mathbf{H}\mathbf{G}_1\mathbf{H}'\|=O_p(\vartheta_{nT,\delta})$;
iv. $\|\widehat{\mathbf{K}}-\mathbf{H}\mathbf{K}\mathbf{R}\|=O_p(\vartheta_{nT,\delta})$.

Proof: see Appendix A.

The next series of remarks provide some intuition about the previous results.

Remark 4 Since from Assumption 3 and Lemma 4, the factors, $\mathbf{F}_t$, are identified only up to an orthogonal transformation, $\mathbf{H}$, consistency is proved for the parameters of a VECM for the transformed true factors, $\mathbf{HF}_t$. In particular, if $\boldsymbol{\beta}$ is the cointegration matrix for $\mathbf{F}_t$, then $\mathbf{H}\boldsymbol{\beta}$ is the cointegration matrix for $\mathbf{HF}_t$. This issue poses no problem for empirical analysis though. Indeed, it can be immediately shown that identification of impulse response functions is not affected by $\mathbf{H}$, which therefore does not have to be estimated. This is in agreement with the fact that the factors and therefore their cointegration relations have no economic meaning. On the other hand, the matrix $\mathbf{Q}$ represents the usual indeterminacy in the identification of the cointegration relations.

Remark 5 The rate of convergence in Lemma 5 is determined by $\vartheta_{nT,\delta}$. In particular, for generic values of $\delta\in[0,1)$ we have

$$\vartheta_{nT,\delta}=\begin{cases}T^{1/2}n^{-(2-\delta)/2} & \text{if } T^{1/(2-\delta)}<n<T,\\ T^{-(1-\delta)/2}=n^{-(1-\delta)/2} & \text{if } n=T,\\ n^{-(1-\delta)/2} & \text{if } T<n<T^{1/(1-\delta)},\\ T^{-1/2} & \text{if } n>T^{1/(1-\delta)}.\end{cases}\qquad(17)$$

[7]We could consider any $\gamma<1$, in which case the rate of convergence of Lemma 5 and Proposition 2 below would depend also on γ. However, since the main message of those results would be qualitatively unaffected, we impose, for simplicity, $\gamma<\delta$.

The case $n \simeq T$ is of particular interest since it corresponds to typical macroeconomic datasets.

Remark 6 Consistency as $n, T \to \infty$ is achieved if and only if $\vartheta_{nT,\delta} \to 0$, which is guaranteed by Assumption 6. Intuitively, since, as T grows, the factor estimation error cumulates due to non-stationarity, we need an increasingly large cross-sectional dimension, n, to control for this error by means of cross-sectional averaging. In particular, since, as shown in (15), the estimated non-stationary factors, $\widehat{\mathbf{F}}_t$, are defined as cross-sectional weighted averages of the data, their estimation error is a weighted average of the idiosyncratic components, and therefore the trade-off between n and T depends on how many of these components are non-stationary. The fact that not all idiosyncratic components can be $I(1)$ is perfectly compatible with macroeconomic data, as shown for example in the empirical application of Section 6.

Moreover, by looking at (17), we see that, for generic values of $\delta \in [0, 1)$, we have the classical $T^{1/2}$-consistency only when $T^{1/(1-\delta)}/n \to 0$, that is when n grows much faster than T. Moreover, when $n = O(T)$, the first two rates in $\vartheta_{nT,\delta}$ are equal and we have convergence at a rate $T^{(1-\delta)/2}$, which for small values of δ is close to the classical rate. Finally, in the case $\delta = 0$, which is asymptotically equivalent to saying that all idiosyncratic components are stationary, we need at least $T^{1/2}/n \to 0$ and for $T/n \to 0$ we have the classical $T^{1/2}$-consistency.

Remark 7 Due to the factor estimation error we do not have in general the classical T-consistency for the estimated cointegration vector $\widehat{\boldsymbol{\beta}}$. Still, by comparing Lemma 5*i* with 5*ii* and 5*iii*, we see that $\widehat{\boldsymbol{\beta}}$ converges to the true value, $\boldsymbol{\beta}$, at a faster rate with respect to the rate of consistency of the other VECM parameters. Such result is promising and it might be exploited in order to test for cointegration of the estimated factors. On the other hand, for large values of n the existence of cointegration relations can also be inferred from an analysis of the number of diverging eigenvalues of the spectral density matrix of the observed variables, as discussed in Section 4.

Remark 8 From $\widehat{\mathbf{K}}$, an estimator, $\widehat{\mathbf{u}}_t$, of a linear transformation of the true common shocks, $\mathbf{u}_t$, can be obtained by projecting $\widehat{\mathbf{w}}_t$ onto the space spanned by the columns of $\widehat{\mathbf{K}}$. According to Lemma 5*iv* and due to non-uniqueness of eigenvectors, $\mathbf{K}$ and $\mathbf{u}_t$ are identified only up to an orthogonal transformation, $\mathbf{R}$.

3.3 Common shocks and impulse response functions

A VECM with $p = 1$ can always be written as a VAR(2) with $r - c$ unit roots. Therefore, after estimating (16), we have the estimated matrix polynomial, $\widehat{\mathbf{A}}^{\text{VECM}}(L)$ with coefficients given by $\widehat{\mathbf{A}}^{\text{VECM}}(0) = \mathbf{I}_r$ and

$$\widehat{\mathbf{A}}_1^{\text{VECM}} = \widehat{\mathbf{G}}_1 + \widehat{\boldsymbol{\alpha}}\widehat{\boldsymbol{\beta}}' + \mathbf{I}_r, \qquad \widehat{\mathbf{A}}_2^{\text{VECM}} = -\widehat{\mathbf{G}}_1. \tag{18}$$

The matrix polynomial $[\widehat{\mathbf{A}}^{\text{VECM}}(L)]^{-1}$ is then obtained by classical inversion of the corresponding VAR using (18), while its explicit expression as function of the estimated VECM parameters is given for example in Lütkepohl (2006). We then in principle have an estimator of the true impulse response functions, $\boldsymbol{\Phi}(L) = \boldsymbol{\Lambda}\mathbf{A}(L)^{-1}\mathbf{K}$.

However, since $\mathbf{K}$ is not identified, the impulse response functions $\boldsymbol{\Phi}(L)$ are in general not identified. Now, while orthogonality of $\mathbf{R}$ in Lemma 5*iv* is a purely mathematical result due to the estimator we choose for $\mathbf{K}$, economic theory tells us that the choice of the identifying transformation is determined by the economic meaning attached to the common shocks, $\mathbf{u}_t$, and in principle any invertible transformation can be considered in order to achieve identification. However, traditional macroeconomic practice assumes Gaussianity of the shocks and therefore restricts to orthogonal matrices only, that is to uncorrelated common shocks. We then need to impose at most $q(q-1)/2$ restrictions in order to achieve under- or just-identification. In this case, $\mathbf{R}$ is a function of the parameters of the model and it can be estimated as a function of the estimated parameters: $\widehat{\mathbf{R}} \equiv \widehat{\mathbf{R}}(\widehat{\mathbf{\Lambda}}, \widehat{\mathbf{A}}^{\text{VECM}}(L), \widehat{\mathbf{K}})$ (see also Forni et al., 2009, for a discussion). Two examples of restrictions are considered in Section 6 when analyzing real data.

The estimated impulse response functions are then defined by combining the estimated parameters and the identification restrictions. In particular, the estimated reaction of the i-th variable to the j-th shock is

$$\widehat{\phi}_{ij}^{\text{VECM}}(L) = \widehat{\boldsymbol{\lambda}}_i' \left[\widehat{\mathbf{A}}^{\text{VECM}}(L)\right]^{-1} \widehat{\mathbf{K}}\,\widehat{\mathbf{r}}_j, \qquad i = 1, \dots, n, \qquad j = 1, \dots, q, \tag{19}$$

where $\widehat{\boldsymbol{\lambda}}_i'$ is the i-th row of $\widehat{\mathbf{\Lambda}}$, $\widehat{\mathbf{r}}_j$ is the j-th column of $\widehat{\mathbf{R}}'$. Consistency of (19) follows.

Proposition 2 ***(Consistency of Impulse Response Functions based on VECM)*** *Under Assumptions 1-7a, as $n, T \to \infty$, we have*

$$\left|\widehat{\phi}_{ijk}^{VECM} - \phi_{ijk}\right| = O_p(\vartheta_{nT,\delta}), \tag{20}$$

for any $k \geq 0$, $i = 1, \dots, n$, and $j = 1, \dots, q$.

Proof: see Appendix A.

The proof of Proposition 2, follows directly by combining Lemma 3 and 5. With reference to Remark 4 it is shown that consistency and identification of impulse response functions is not affected by the fact that common factors are not identified. Remarks 6 and 5 on convergence rates apply also in this case.

3.4 The case of unrestricted VAR for the common factors

Several papers have addressed the issue whether and when a VECM or an unrestricted VAR in the levels should be used for estimation in the case of non-singular cointegrated vectors. Sims et al. (1990) show that the parameters of a cointegrated VAR, as (7), are consistently estimated using an unrestricted VAR in the levels. On the other hand, Phillips (1998) shows that if the variables are cointegrated, then the long-run features of the impulse-response functions are consistently estimated only if the unit roots are explicitly taken into account, that is within a VECM specification (see also Paruolo, 1997). This result is confirmed numerically in Barigozzi et al. (2016) also for the singular case, $r > q$.

Nevertheless, since by estimating an unrestricted VAR it is still possible to estimate consistently short run impulse response functions without the need of determining the number of unit roots and therefore without having to estimate the cointegration relations, this approach has become very popular in empirical research. For this reason, here we also study the properties of impulse response function when, following Sims et al. (1990), we consider least squares estimation of an unrestricted VAR(p) model for the common factors.[8] For simplicity let $p=1$, then

$$\mathbf{F}_t = \mathbf{A}_1\mathbf{F}_{t-1} + \mathbf{w}_t, \quad \mathbf{w}_t = \mathbf{K}\mathbf{u}_t.$$

Denote by $\widehat{\mathbf{A}}_1^{\text{VAR}}$ the least squares estimators of the coefficient matrix and by $\widehat{\mathbf{K}}$ the estimator of $\mathbf{K}$, which is obtained as in the VECM case starting from the sample covariance of the residuals. Consistency of these estimators is given in the following Lemma.

Lemma 6 *Under Assumptions 1-5 and 7a, and given* $\mathbf{H}$ *defined in Lemma 3, there exists a* $q\times q$ *orthogonal matrix* $\mathbf{R}$*, such that, as* $n,T\to\infty$*,*

i. $\|\widehat{\mathbf{A}}_1^{\text{VAR}} - \mathbf{H}\mathbf{A}_1\mathbf{H}'\| = O_p(\max(n^{-1/2},T^{-1/2}))$*;*

ii. $\|\widehat{\mathbf{K}} - \mathbf{H}\mathbf{K}\mathbf{R}\| = O_p(\max(n^{-1/2},T^{-1/2}))$*.*

Proof: see Appendix A.

As before, an estimator of the identifying matrix $\mathbf{R}$ can be obtained by imposing appropriate restrictions. Then, the estimated impulse response of the i-th variable to the j-th shock is defined as

$$\widehat{\phi}_{ij}^{\text{VAR}}(L) = \widehat{\boldsymbol{\lambda}}_i'\left[\widehat{\mathbf{A}}^{\text{VAR}}(L)\right]^{-1}\widehat{\mathbf{K}}\,\widehat{\mathbf{r}}_j, \qquad i=1,\dots,n, \qquad j=1,\dots,q, \tag{21}$$

where $\widehat{\boldsymbol{\lambda}}_i'$ is the i-th row of $\widehat{\boldsymbol{\Lambda}}$, $\widehat{\mathbf{r}}_j$ is the j-th column of $\widehat{\mathbf{R}}'$, while an expression for $[\widehat{\mathbf{A}}^{\text{VAR}}(L)]^{-1}$ is readily available by classical inversion of a VAR. We then have consistency also for (21).

Proposition 3 ***(Consistency of Impulse Response Functions based on VAR)***
Under Assumptions 1-5 and 7a, as $n,T\to\infty$*, we have*

$$\left|\widehat{\phi}_{ijk}^{VAR} - \phi_{ijk}\right| = O_p\left(\max\left(n^{-1/2},T^{-1/2}\right)\right), \tag{22}$$

for any finite $k\geq 0$*,* $i=1,\dots,n$*, and* $j=1,\dots,q$*.*

Proof: see Appendix A.

Three last remarks are in order.

Remark 9 For any finite horizon k the impulse response $\widehat{\phi}_{ijk}^{\text{VAR}}$ is also a consistent estimator of ϕ_{ijk}. This result is consistent with the result for observed variables by Sims et al. (1990) in presence of some unit roots. On the other hand, it is possible to prove that the same unit roots affect the estimated long-run impulse response functions in such a way that their least squares estimator is no more consistent, i.e. $\lim_{k\to\infty}|\widehat{\phi}_{ijk}^{\text{VAR}} - \phi_{ijk}| = O_p(1)$ (see Theorem 2.3 in Phillips, 1998). For this reason, Proposition 3 holds only for finite horizons k.

Remark 10 The estimator $\widehat{\phi}_{ijk}^{\text{VAR}}$ can converge faster than $\widehat{\phi}_{ijk}^{\text{VECM}}$. However, as shown in the proof of Lemma 6, the rate of convergence of the parameters associated to the non-stationary components is slower than what it would be were the factors observed. This is of course due to the factors' estimation error. Moreover, convergence in Proposition 3 is achieved without the need of Assumption 6, hence even when all idiosyncratic components are $I(1)$ and with no constraint on the relative rates of n and T or on the cross-sectional dependence of stationary and non-stationary idiosyncratic blocks.

[8]For alternative approaches, not considered here, see for example the fully modified least squares estimation by Phillips (1995).

Remark 11 The result of Proposition 3 holds also for impulse responses estimated via Factor Augmented VAR models (see e.g. Bernanke et al., 2005). Indeed, as already proved by Bai and Ng (2006) in the stationary case, the least squares estimates of those model have a convergence rate $\min(n^{1/2}, T^{1/2})$ also in the non-stationary case. That is, except when $T/n \to 0$, we should take into account the effect of the estimated factors.

Summing up, as a consequence of these results, the empirical researcher faces a trade-off between estimating correctly the whole impulse response function with a slow rate and more restrictive assumptions, as in Proposition 2, or giving up on the long-run behavior in exchange for a faster rate of convergence, as in Proposition 3.

3.5 The case of deterministic trends

In Assumption 1 we considered the case of no deterministic trend. However, macroeconomic data often have a linear trend. In this case the model for an observed time series becomes

$$y_{it} = a_i + b_i t + \boldsymbol{\lambda}_i' \mathbf{F}_t + \xi_{it}, \quad i = 1, \dots, n, \quad t = 0, \dots T. \tag{23}$$

where $x_{it} = \boldsymbol{\lambda}_i' \mathbf{F}_t + \xi_{it}$ follows a Non-Stationary Dynamic Factor as described in Section 2. Here we also allow for non-zero initial conditions ($a_i \neq 0$), this posing no difficulty in terms of estimation.

Impulse response functions are then defined for the de-trended data, x_{it} (see Section 6 for their economic interpretation in this case), and, therefore, in order to estimate them, we have to first estimate the trend slope, b_i. This can be done either by de-meaning first differences or by least squares regression, the two approaches respectively giving

$$\widetilde{b}_i = \frac{1}{T} \sum_{t=1}^{T} \Delta y_{it} = \frac{y_{iT} - y_{i0}}{T}, \qquad \widehat{b}_i = \frac{\sum_{t=0}^{T} (t - \frac{T}{2})(y_{it} - \bar{y}_i)}{\sum_{t=0}^{T} (t - \frac{T}{2})^2}, \quad i = 1, \dots, n. \tag{24}$$

Both estimators in (24) are $T^{1/2}$-consistent (the proof for $\widetilde{b}_i$ is trivial while we refer to Lemma 16 in Appendix B for a proof for $\widehat{b}_i$).[9] Given this classical rate, impulse response functions can still be estimated consistently, as described above, when using de-trended data.

However, it has to be noticed that finite sample properties of $\widehat{b}_i$ and $\widetilde{b}_i$ might differ substantially. First, assume to follow Bai and Ng (2004), and consider de-meaning of first differences. Then, from principal component analysis on $\Delta \widetilde{x}_{it} = \Delta y_{it} - \widetilde{b}_i$, we can estimate the first differences of the factors, which, once integrated, give us the estimated factors, $\widetilde{\mathbf{F}}_t$, such that, due to differencing, $\widetilde{\mathbf{F}}_0 = \mathbf{0}$. Moreover, since the sample mean of $\Delta \widetilde{x}_{it}$ is zero by construction, then also $\Delta \widetilde{\mathbf{F}}_t$ have zero sample mean and therefore we always have $\widetilde{\mathbf{F}}_0 = \widetilde{\mathbf{F}}_T = \mathbf{0}$.

If instead we use least squares, then factors can be estimated as in (15) starting from $\widehat{x}_{it} = y_{it} - \widehat{b}_i t$. Since, now, in general, $\Delta \widehat{x}_{it}$ has sample mean different from zero, then those estimated factors have $\widehat{\mathbf{F}}_0 \neq \mathbf{0}$ and $\widehat{\mathbf{F}}_0 \neq \widehat{\mathbf{F}}_T$. In this paper, we opt for this second solution, while a complete numerical and empirical comparison of the finite sample properties of the two methods is left for further research.

[9] If $x_{it} \sim I(0)$ the least squares estimator, $\widehat{b}_i$, is $T^{3/2}$-consistent.

4 Determining the number of factors and shocks

In the previous section we made the assumption that r, q, and d are known. Of course this is not the case in practice and we need a method to determine them. Determining r and q is straightforward in the sense that, given Assumptions 1-5, we can apply all existing methods to the first difference of the data. A non-exhaustive list of possible approaches is: Bai and Ng (2002), Onatski (2009), Alessi et al. (2010), and Ahn and Horenstein (2013) for r, and Amengual and Watson (2007), Bai and Ng (2007), Hallin and Liška (2007), and Onatski (2010) for q.

On the other hand, there is no procedure ready available in the Dynamic Factor model setting for determining the number of common trends $q-d$. One possibility would be to apply one of the available methods to determine the cointegration rank or the number of common trends to the estimated factors, as for example adapting the classical approaches by Stock and Watson (1988), Phillips and Ouliaris (1988), and Johansen (1991), or the more recent ones by Hallin et al. (2016). However, two difficulties clearly emerge from this strategy. First, those existing methods have to be applied to estimated quantities and this might pose theoretical problems, since, as we have seen, the estimators of the cointegration vectors are not super-consistent in the classical way. Second, the double singularity of the model ($d < q < r$) and results from Proposition 1 require some caution in employing existing methods to the present context. A second possibility would be to employ tests for cointegration in panels with a factor structure, as for example those proposed by Bai and Ng (2004) and Gengenbach et al. (2015). However, those methods are developed under the assumption that $r = q$ and it is unclear what are their properties when $q < r$.

Although all these approaches are worth being explored, here we choose a simpler approach which is directly connected to the spectral representation of the model in first differences. For simplicity we define $\tau \equiv q - d$ and our goal is to determine τ. In particular, by virtue of Assumption 5, and from (1) and (11), the spectral density matrix of the differenced data is

$$\mathbf{\Sigma}^{\Delta x}(\theta) = \mathbf{\Sigma}^{\Delta\chi}(\theta) + \mathbf{\Sigma}^{\Delta\xi}(\theta) = \frac{1}{2\pi}\mathbf{\Lambda}\mathbf{C}(e^{-i\theta})\overline{\mathbf{C}'(e^{-i\theta})}\mathbf{\Lambda}' + \mathbf{\Sigma}^{\Delta\xi}(\theta), \quad \theta \in [-\pi, \pi]. \tag{25}$$

Now, notice that $\text{rk}(\mathbf{C}(e^{-i\theta})) = q$ a.e. in $[-\pi, \pi]$ (see the results in Barigozzi et al., 2016). On the other hand, this is clearly not true in $\theta = 0$, where, because of the existence of $\tau < q$ common trends (Assumption 2*f*), we have $\text{rk}(\mathbf{C}(1)) = \tau$. This, in turn implies $\text{rk}(\mathbf{\Sigma}^{\Delta\chi}(0)) = \tau$. The peculiarity of the behaviour of the spectrum at frequency zero, is the analogous of the singularity that we have in the impulse responses polynomial $\mathbf{\Phi}(z)$ in $z = 1$, i.e. cointegration implies a reduced rank spectrum at frequency zero.[10]

In agreement with this observation, the next result characterises the behaviour of the eigenvalues of the spectral density of the first differences of the common component, $\mu_j^{\Delta\chi}(\theta)$, of the idiosyncratic component, $\mu_j^{\Delta\xi}(\theta)$, and of the data, $\mu_j^{\Delta x}(\theta)$.

Lemma 7 *Under Assumptions 1-5 and for any $n \in \mathbb{N}$,*

i. $0 < \underline{M}_9 \leq n^{-1}\mu_j^{\Delta\chi}(\theta) \leq \overline{M}_9 < \infty$ *a.e. in* $[-\pi, \pi]$*, and for any* $j = 1, \dots, q$*;*

ii. $\sup_{\theta\in[-\pi,\pi]} \mu_1^{\Delta\xi}(\theta) \leq M_{10} < \infty$*;*

iii. $0 < \underline{M}_{11} \leq n^{-1}\mu_q^{\Delta x}(\theta) \leq \overline{M}_{11} < \infty$ *a.e. in* $[-\pi, \pi]$*;*

iv. $\sup_{\theta\in[-\pi,\pi]} \mu_{q+1}^{\Delta x}(\theta) \leq M_{10} < \infty$*;*

[10] Notice that, for any $\theta \in [-\pi, \pi]$, we have $\text{rk}(\mathbf{C}(e^{-i\theta})) \leq q$ and $\text{rk}(\mathbf{\Sigma}^{\Delta\chi}(\theta)) \leq q$.

v. $0 < \underline{M}_{12} \leq n^{-1}\mu_\tau^{\Delta x}(\theta = 0) \leq \overline{M}_{12} < \infty$ *and* $\mu_{\tau+1}^{\Delta x}(\theta = 0) \leq M_{10} < \infty$.

Proof: see Appendix A.

While for determining q Hallin and Liška (2007) study the behaviour of the eigenvalues of $\mathbf{\Sigma}^{\Delta x}(\theta)$ over a window of frequencies, thus using Lemma 7*iii* and 7*iv*, here we determine τ by focussing on the behaviour of the same eigenvalues only at frequency zero, i.e. we rely on Lemma 7*v*.

Assume to have a consistent estimator of the spectral density matrix of $\Delta\mathbf{x}_t$, with estimated eigenvalues $\widehat{\mu}_j^{\Delta x}(\theta)$ and rate of consistency given by ρ_T, such that $\rho_T \to \infty$ and $\rho_T/T \to 0$, as $T \to \infty$.[11] We define the estimated number of common trends as

$$\widehat{\tau} = \underset{k=0,\dots,\tau_{\max}}{\arg\min} \left[\log\left(\frac{1}{n} \sum_{j=k+1}^{n} \widehat{\mu}_j(0) \right) + kp(n,T) \right], \tag{26}$$

for some suitable penalty function $p(n,T)$ and for a given maximum number of common trends $\tau_{\max}$. A small numerical evaluation of (26) is presented in the next section, while here we conclude with the following sufficient conditions for consistency in the selection of the number of common trends.

Proposition 4 (Number of common trends) *Under Assumptions 1-5 and 7b, and if the penalty $p(n,T)$ is such that, as $n,T \to \infty$, (i) $p(n,T) \to 0$ and (ii) $(n\rho_T^{-1})p(n,T) \to \infty$, then* $\text{Prob}(\widehat{\tau} = \tau) \to 1$.

Proof: see the proof of Proposition 2 in Hallin and Liška (2007), when restricted to $\theta = 0$.

Notice that by definition we have $\tau = r - c$ which is the number of unit roots driving the dynamics of the common factors. Therefore, once we determine τ, q, and r, we immediately have estimates for (i) the number of transitory shocks $d = q - \tau$ and (ii) the cointegration rank $c = r - \tau$.

5 Simulations

We simulate data, from a Non-Stationary Dynamic Factor model with $r = 4$ common factors, and $q = 3$ common shocks, and $\tau = 1$ common trends, thus $d = q - \tau = 2$ and the cointegration relations among the common factors are $c = r - q + d = 3$. More precisely, for given values of n and T, each time series follows the data generating process:

$$\begin{aligned} x_{it} &= \boldsymbol{\lambda}_i'\mathbf{F}_t + \xi_{it}, && i = 1,\dots,n, \quad t = 1,\dots,T, \\ \mathbf{A}(L)\mathbf{F}_t &= \mathbf{K}\mathbf{R}\mathbf{u}_t, && \mathbf{u}_t \overset{w.n.}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{I}_q), \end{aligned}$$

where $\boldsymbol{\lambda}_i$ is $r \times 1$ with entries $\lambda_{ij} \sim \mathcal{N}(0,1)$, $\mathbf{A}(L)$ is $r \times r$ with $\tau = r - c = 1$ unit root, $\mathbf{K}$ is $r \times q$, and $\mathbf{R}$, which is necessary for identification of the impulse responses, is $q \times q$.

In practice, to generate $\mathbf{A}(L)$, we exploit a particular Smith-McMillan factorization (see Watson, 1994) according to which $\mathbf{A}(L) = \boldsymbol{\mathcal{U}}(L)\boldsymbol{\mathcal{M}}(L)\boldsymbol{\mathcal{V}}(L)$, where $\boldsymbol{\mathcal{U}}(L)$ and $\boldsymbol{\mathcal{V}}(L)$ are $r \times r$ polynomials with all of their roots outside the unit circle, and $\boldsymbol{\mathcal{M}}(L) = \text{diag}\left((1-L)\mathbf{I}_{r-c}, \mathbf{I}_c\right)$.

[11] For example for the lag-window estimator used in Forni et al. (2015a), under Assumptions 1, 4*d*-*g*, we have $\rho_T = (B_T \log B_T T^{-1})^{-1/2}$, where B_T is the window size.

In particular, we set $\boldsymbol{\mathcal{U}}(L) = (\mathbf{I}_r - \boldsymbol{\mathcal{U}}_1 L)$, and $\boldsymbol{\mathcal{V}}(L) = \mathbf{I}_r$, so that $\mathbf{F}_t$ follows a VAR(2) with $r - c$ unit roots, or, equivalently, $\Delta\mathbf{F}_t$ follows a VECM(1) with c cointegration relations. The diagonal elements of the matrix $\boldsymbol{\mathcal{U}}_1$ are drawn from a uniform distribution on $[0.5, 0.8]$, while the off-diagonal elements from a uniform distribution on $[0, 0.3]$. The matrix $\boldsymbol{\mathcal{U}}_1$ is then standardized to ensure that its largest eigenvalue is 0.6. The matrix $\mathbf{K}$ is generated as in Bai and Ng (2007): let $\tilde{\mathbf{K}}$ be a $r \times r$ diagonal matrix of rank q with entries drawn from a uniform distribution on $[.8, 1.2]$, and let $\check{\mathbf{K}}$ be a $r \times r$ orthogonal matrix, then, $\mathbf{K}$ is equal to the first q columns of the matrix $\check{\mathbf{K}}\tilde{\mathbf{K}}^{\frac{1}{2}}$. Finally, the matrix $\mathbf{R}$ is calibrated such that the following restrictions hold: $\phi_{12}(0) = \phi_{13}(0) = \phi_{23}(0) = 0$.

The idiosyncratic components are generated according to the ARMA model (with possible unit root)

$$(1 - \rho_i L)\xi_{it} = \sum_{k=0}^{\infty} d_i^k \varepsilon_{it-k}, \qquad \varepsilon_{it} \sim \mathcal{N}(0,1), \qquad \mathsf{E}[\varepsilon_{it}\varepsilon_{jt}] = 0.5^{|i-j|},$$

where $\rho_i = 1$ for $i = 1, \ldots, m$ and $\rho_i = 0$ for $i = m+1, \ldots, n$, so that m idiosyncratic components are non-stationary, while the coefficients d_i's are drawn from a uniform distribution on $[0, 0.5]$. Each idiosyncratic component is rescaled so that it accounts for a third of the total variance.

The matrices $\boldsymbol{\Lambda}$, $\boldsymbol{\mathcal{U}}_1$, $\mathbf{G}$ and $\mathbf{H}$ are simulated only once so that the set of impulse responses to be estimated is always the same, while the vector $\mathbf{u}_t$, the vector $\boldsymbol{\varepsilon}_t$, and all the d_i's are drawn at each replication. Results are based on 1000 MonteCarlo replications and the goal is to study the finite sample properties of the two estimators of the impulse response functions discussed in the previous section, for different cross-sectional and sample sizes (n and T) and for a different shares (m) of non-stationary idiosyncratic components.

Tables 1 and 2 show Mean Squared Errors (MSE) for the estimated impulse responses simulated with different parameter configurations. Estimation is carried out as explained in Section 3, by fitting on $\widehat{\mathbf{F}}_t$ either a VECM or an unrestricted VAR, and where $\widehat{\mathbf{F}}_t$ is estimated as in (15). The numbers r, q, and τ are assumed to be known. More precisely, let $\widehat{\phi}_{ijk,h}$ be the (i,j)-th entry of the matrix polynomial $\widehat{\boldsymbol{\Phi}}(L)$ at lag k when estimated at the h-th replication, then MSEs are computed with respect to all replications, all variables, and all shocks:

$$MSE(k) = \frac{1}{1000nq} \sum_{i=1}^{n} \sum_{j=1}^{q} \sum_{h=1}^{1000} \left(\widehat{\phi}_{ijk,h} - \phi_{ijk}\right)^2.$$

From Table 1 we can see that in the VECM case the estimation error decreases monotonically as n and T grow, while it is larger at higher horizons. Notice that, in accordance with Proposition 2 which states that the estimation error is inversely related to the number of non-stationary idiosyncratic components, for every couple of n and T the MSE decreases for smaller values of m.

The picture offered by Table 2 is slightly different than the one offered by Table 1. On the one hand, at short horizons the MSE of $\widehat{\phi}_{ijk}^{\text{VAR}}$ is comparable to, or slightly smaller than, the MSE of $\widehat{\phi}_{ijk}^{\text{VECM}}$, which is consistent with the result of Proposition 2 and 3 according to which $\widehat{\phi}_{ijk}^{\text{VAR}}$ converges at a faster rate than $\widehat{\phi}_{ijk}^{\text{VECM}}$. On the other hand, at longer horizons, the MSE of $\widehat{\phi}_{ijk}^{\text{VAR}}$ is always larger than the MSE of $\widehat{\phi}_{ijk}^{\text{VECM}}$, which is not surprising since the

Table 1: MonteCarlo Simulations - Impulse Responses
Mean Squared Errors
VECM Estimation

T	n	m	$k=0$	$k=1$	$k=4$	$k=8$	$k=12$	$k=16$	$k=20$
100	100	25	0.080	0.113	0.249	0.350	0.380	0.387	0.389
100	100	50	0.078	0.115	0.276	0.425	0.490	0.513	0.521
100	100	75	0.079	0.125	0.316	0.518	0.624	0.671	0.691
100	100	100	0.074	0.129	0.344	0.575	0.706	0.765	0.792
200	200	50	0.037	0.050	0.114	0.166	0.190	0.201	0.207
200	200	100	0.035	0.053	0.132	0.211	0.267	0.306	0.332
200	200	150	0.035	0.058	0.152	0.253	0.331	0.389	0.429
200	200	200	0.034	0.064	0.169	0.269	0.352	0.419	0.469
300	300	75	0.024	0.033	0.076	0.111	0.130	0.140	0.146
300	300	150	0.023	0.037	0.093	0.136	0.166	0.189	0.206
300	300	225	0.022	0.041	0.108	0.159	0.201	0.238	0.270
300	300	300	0.021	0.044	0.121	0.183	0.238	0.291	0.338

This table reports Mean Squared Errors (MSE) for the estimated impulse responses by fitting a VECM on $\Delta\widehat{\mathbf{F}}_t$ as in (6). Let $\widehat{\phi}^h_{ijk}$ be the (i,j)-th entry of the matrix polynomial $\widehat{\mathbf{\Phi}}(L)$ at lag k when estimated at the h-th replication, then MSEs are computed with respect to all replications, all variables, and all shocks: $MSE(k) = \frac{1}{1000nq}\sum_{i=1}^n\sum_{j=1}^q\sum_{h=1}^{1000}\left(\widehat{\phi}^h_{ijk} - \phi_{ijk}\right)^2$. T is the number of observations, n is the number of variables, and m is the number of idiosyncratic components that are $I(1)$.

Table 2: MonteCarlo Simulations - Impulse Responses
Mean Squared Errors
Unrestricted VAR Estimation

T	n	m	$k=0$	$k=1$	$k=4$	$k=8$	$k=12$	$k=16$	$k=20$
100	100	25	0.081	0.110	0.267	0.527	0.747	0.904	1.013
100	100	50	0.076	0.112	0.287	0.552	0.772	0.930	1.043
100	100	75	0.078	0.123	0.313	0.596	0.822	0.979	1.088
100	100	100	0.072	0.122	0.333	0.624	0.858	1.018	1.123
200	200	50	0.038	0.050	0.125	0.250	0.384	0.511	0.625
200	200	100	0.036	0.053	0.142	0.275	0.415	0.548	0.667
200	200	150	0.034	0.057	0.157	0.285	0.419	0.549	0.667
200	200	200	0.033	0.064	0.173	0.308	0.449	0.587	0.710
300	300	75	0.023	0.032	0.083	0.165	0.257	0.352	0.444
300	300	150	0.023	0.037	0.102	0.185	0.278	0.377	0.474
300	300	225	0.022	0.041	0.114	0.195	0.287	0.387	0.486
300	300	300	0.022	0.046	0.128	0.210	0.300	0.398	0.495

This table reports Mean Squared Errors (MSE) for the estimated impulse responses by fitting a VAR on $\widehat{\mathbf{F}}_t$ as in (7). Let $\widehat{\phi}^h_{ijk}$ be the (i,j)-th entry of the matrix polynomial $\widehat{\mathbf{\Phi}}(L)$ at lag k when estimated at the h-th replication, then MSEs are computed with respect to all replications, all variables, and all shocks: $MSE(k) = \frac{1}{1000nq}\sum_{i=1}^n\sum_{j=1}^q\sum_{h=1}^{1000}\left(\widehat{\phi}^h_{ijk} - \phi_{ijk}\right)^2$. T is the number of observations, n is the number of variables, and m is the number of idiosyncratic components that are $I(1)$.

long run impulse responses estimated with an unrestricted VAR in levels are known to be asymptotically biased.

Finally, for the same data generating process we study the performance of the information criterion (26), proposed in Section 4. Table 3 shows the percentage of times in which we estimate correctly the number of common trends $\tau = 1$. For the sake of comparison, we also report results of the criterion by Hallin and Liška (2007) for estimating $q = 3$. It has to be

Table 3: MonteCarlo Simulations - Number of Common Trends and Shocks Percentage of Correct Answer

T	n	m	$\widehat{\tau}=\tau$	$\widehat{q}=q$
100	50	25	98.6	96.5
100	50	50	99.2	99.8
100	100	50	98.7	100
100	100	100	99.8	100
100	200	100	96.5	100
100	200	200	99.9	100
200	50	25	99.6	100
200	50	50	100	100
200	100	50	99.9	100
200	100	100	100	100
200	200	100	99.7	100
200	200	200	100	100

This table reports the percentage of simulations in which the information criterion of Hallin and Liška (2007) returned the correct number of common shocks ($\widehat{q}=q$), and in which the criterion proposed in Section 4 returned the correct number of common trends ($\widehat{\tau}=\tau$). T is the number of observations, n is the number of variables, and m is the number of idiosyncratic components that are $I(1)$.

noticed that the actual implementation of these criteria requires a procedure of fine tuning of the penalty, indeed for any constant $C>0$, the function $Cp(n,T)$ is also an admissible penalty, and therefore, as explained in Hallin and Liška (2007), a whole range of values of C should be explored. For this reason, numerical studies about the performance of these methods are computationally intensive, thus we limit ourselves to a small scale study and we leave to further research a thorough comparison of the estimator proposed in (26) with other possible methods. Still results are already promising, as our criterion seems to work fairly well by giving the correct answer more than 95% of the times.

6 Empirical application

In this Section we estimate the Non-Stationary Dynamic Factor model to study the effects of monetary policy shocks and of supply shocks. We consider a large macroeconomic dataset comprising 101 quarterly series from 1960Q3 to 2012Q4 describing the US economy, where the complete list of variables and transformations is reported in Appendix C. Broadly speaking, all the variables that are $I(1)$ are not transformed, while for those that are $I(2)$ we take first difference. We then remove deterministic component as described at the end of Section 3, therefore the impulse responses presented in this section have to be interpreted as out of trend deviations.

The model is estimated as explained in the previous sections, and in particular the common factors are estimated using our new proposed estimator (15). We find evidence of $r=7$ common factors as suggested both by the criteria in Alessi et al. (2010) and in Bai and Ng (2002), and of $q=3$ common shocks q as given by the criterion in Hallin and Liška (2007). Finally, using the information criterion described in Section 4, we allow for just one common stochastic trend, $\tau=1$, thus $d=2$ shocks have no long-run effect but the cointegration rank for the common factors is $c=6$ due to singularity of the common factors $(r>q)$.

We then consider two different identification schemes. First, we study the effects of a monetary policy shock, which is identified by using a standard recursive identification scheme, according to which GDP and CPI do not react contemporaneously to the monetary policy shock (see e.g. Forni and Gambetti, 2010). Second, we study the effects of a supply shock, where the supply shock is identified as the only shock having a permanent effect on the system (see e.g. King et al., 1991; Forni et al., 2009). Results of these two exercises are presented in Figure 1 and Figure 2, where in both cases the black lines are the impulse responses obtained with the Non-Stationary Dynamic Factor model by fitting a VECM on $\Delta\widehat{\mathbf{F}}_t$, the grey lines are the impulse responses obtained by fitting an unrestriced VAR on $\widehat{\mathbf{F}}_t$, while the dotted black lines and the grey shaded areas are the respective 68% bootstrap confidence bands.

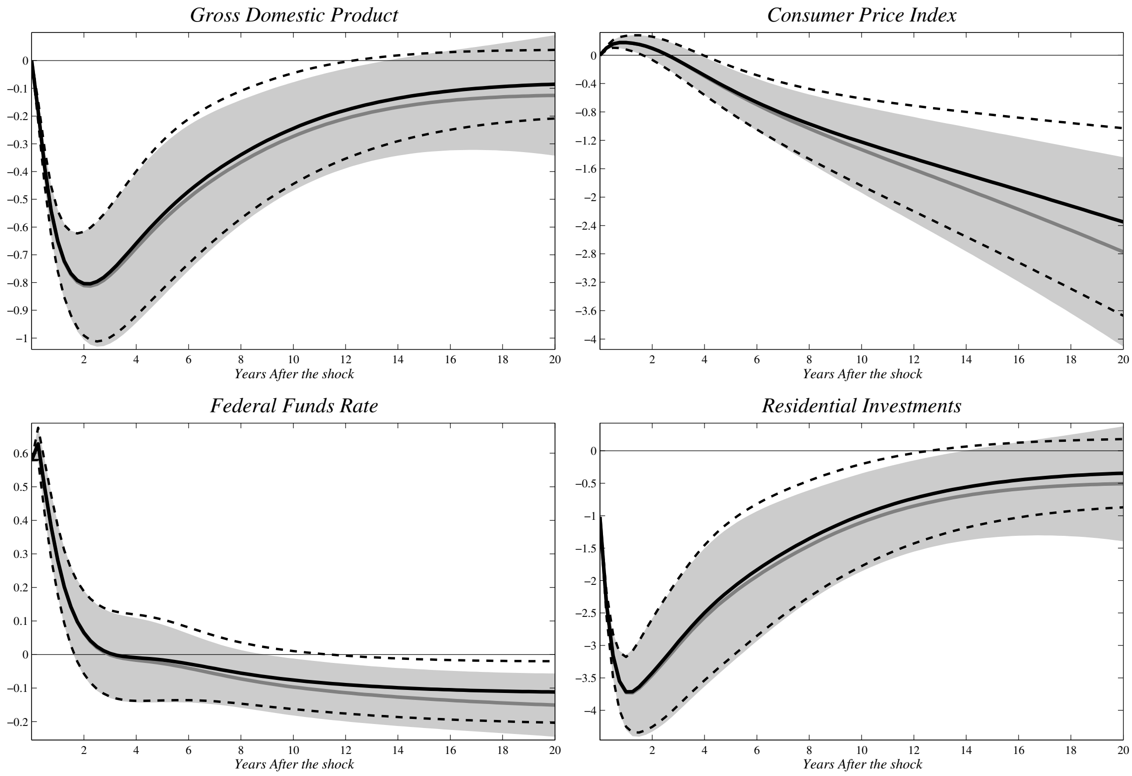
Figure 1 shows the impulse response functions to a monetary policy shock normalized so that at impact it raises the Federal Funds rate by 50 basis points. GDP and Residential Investments respond negatively to a contractionary monetary policy shock, and then they revert to the baseline. Similarly, consumer prices, which are modeled as $I(2)$, stabilize, meaning that inflation reverts to zero. These results, and in particular the long-run behaviours, are consistent with economic theory according to which a monetary policy shock has only a transitory effect on the economy. On the contrary, the impulse responses estimated with a stationary Dynamic Factor model, i.e. with data in first differences, would display non-plausible permanent effects of monetary policy shocks on all variables (not shown here). Notice also that there is no significant difference between estimates obtained using a VECM or an unrestricted VAR for the factors. Finally, the impulse responses in Figure 1 are very similar, both in terms of shape and in terms of size, to those obtained with Large Bayesian VARs estimated in levels (see e.g. Giannone et al., 2015).

Figure 2 shows the impulse response functions to a supply policy shock normalized so that at impact it increases GDP of 0.25%. All variables have a hump shaped response, with a maximum between six and seven quarters after the shock. The deviation from the trend estimated by fitting a VECM is 0.23% after ten years, and 0.12% after twenty years and onwards. Differently from the results in Figure 1, while the impulse response functions obtained using a VECM or an unrestricted VAR show no difference in the short-run, at very long horizons significant differences appear. Notably, the impulse responses estimated by fitting an unrestricted VAR tend to diverge. This result is consistent with lack of consistency of long-run impulse responses obtained without imposing the presence of unit roots (see Proposition 3). Indeed, when, as in this case, we fit an unrestricted VAR on $\widehat{\mathbf{F}}_t$ and we impose long-run identifying restrictions, we are actually imposing constraints on a matrix which is not consistently estimated. This unavoidably compromises the estimated structural responses.

Differently from the case of a monetary policy shock, economic theory does not tell us neither what should be the long-run effect of a supply shock, besides being permanent, nor what should be the shape of the induced dynamic response. Hence, we cannot say *a priori* whether the effect found is realistic or not. While with the Non-Stationary Dynamic Factor model, we find that a supply shock induces on GDP a permanent deviation of about 0.12% from its historical trend, with a stationary Dynamic Factor model we would find a deviation of about 0.67% (not shown here). Finally, a result similar to ours is found also in Dedola and Neri (2007) and Smets and Wouters (2007).

To conclude, the empirical analysis of this section shows that the factor model proposed is able to reproduce the main features of the effects of temporary and permanent shocks postulated by macroeconomic theory.

Figure 1: Impulse Response Functions to a Monetary Policy Shock



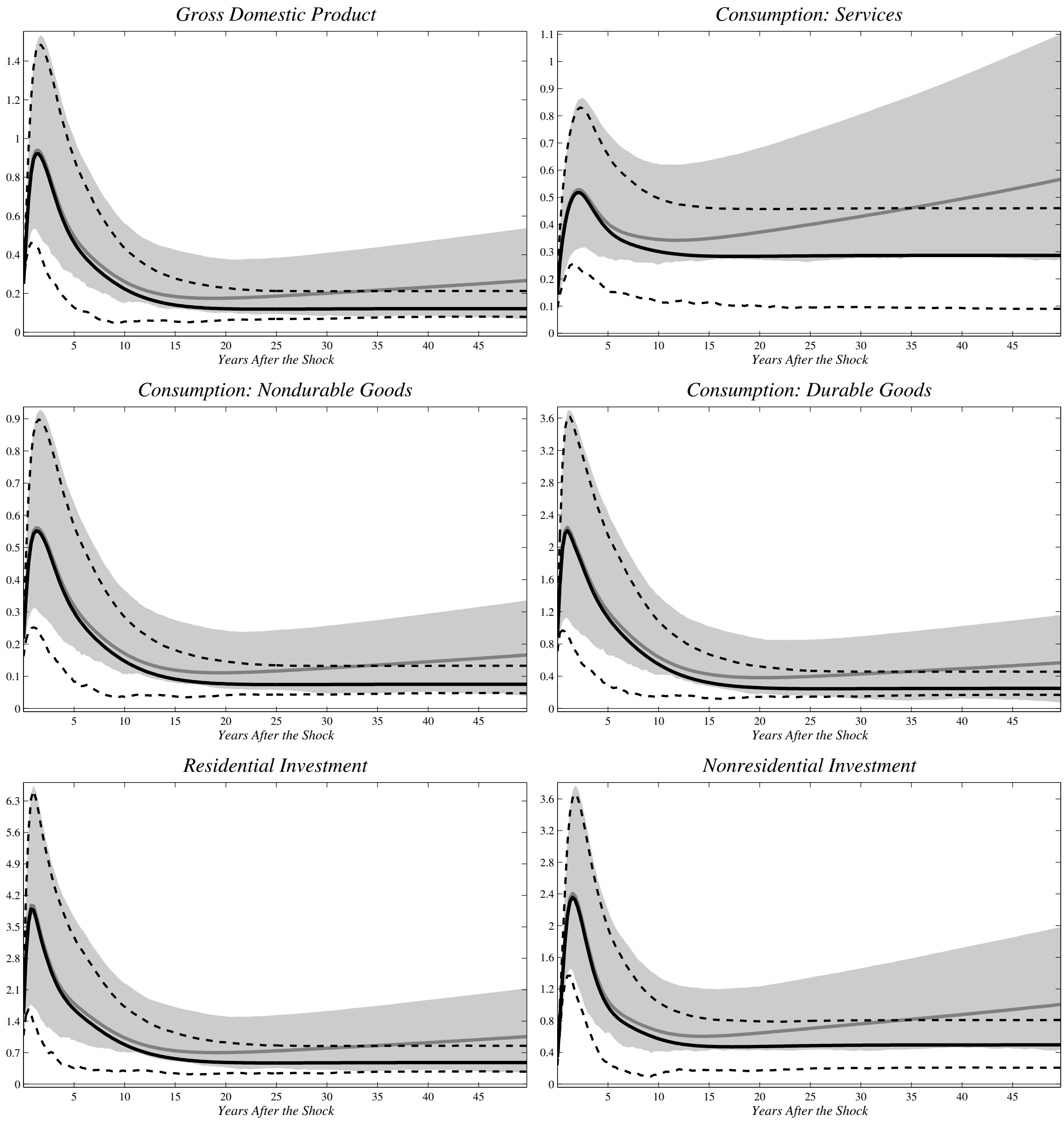
Solid black lines are the impulse response functions obtained from the Non-Stationary Dynamic Factor model by estimating a VECM on on $\Delta\widehat{\mathbf{F}}_t$ with 68% bootstrap confidence bands (dashed). Solid grey lines are the impulse–response functions obtained from the Non-Stationary Dynamic Factor model by estimating a VAR on on $\widehat{\mathbf{F}}_t$ with 68% confidence bands (shaded areas). The monetary policy shock is normalized so that at impact it increases the Federal Funds rate of 50 basis points.

7 Conclusions

In this paper, we propose a *Non-Stationary Dynamic Factor model* for large datasets. The natural use of these class of models in a macroeconomic context motivates the main assumptions upon which the present theory is built. This paper is complementary to a another one where we address representation theory (Barigozzi et al., 2016). In particular, we propose a two-step estimator of impulse response functions which is consistent when both the cross-sectional dimension n and the sample size T of the dataset grow to infinity. Furthermore, we also propose an information criterion to determine the number of common trends.

The results of this paper are useful beyond estimation of impulse response functions. First, given its state-space form, our model can be estimated using Kalman filtering techniques (see Doz et al., 2011, for the stationary case), and hence it can be employed for forecasting in real-time (Giannone et al., 2008). Second, our estimation approach can be used for estimating and validating Dynamic Stochastic General Equilibrium models in a data-rich environment (see Boivin and Giannoni, 2006, for the stationary case). These aspects are part of our current research.

Figure 2: Impulse Response Functions to a Supply Shock



Solid black lines are the impulse response functions obtained from the Non-Stationary Dynamic Factor Model by estimating a VECM on on $\Delta\widehat{\mathbf{F}}_t$ with 68% bootstrap confidence bands (dashed). Solid grey lines are the impulse–response functions obtained from the Non-Stationary Dynamic Factor Model by estimating a VAR on on $\widehat{\mathbf{F}}_t$ with 68% confidence bands (shaded areas). The supply shock is normalized so that at impact it increases GDP of 0.25%.

References

Ahn, S. C. and A. R. Horenstein (2013). Eigenvalue ratio test for the number of factors. *Econometrica 81*, 1203–1227.

Alessi, L., M. Barigozzi, and M. Capasso (2010). Improved penalization for determining the number of factors in approximate static factor models. *Statistics and Probability Letters 80*, 1806–1813.

Alessi, L., M. Barigozzi, and M. Capasso (2011). Nonfundamentalness in structural econometric models: A review. *International Statistical Review 79*, 16–47.

Altissimo, F., R. Cristadoro, M. Forni, M. Lippi, and G. Veronese (2010). New eurocoin: Tracking economic growth in real time. *The Review of Economics and Statistics 92*, 1024–1034.

Amengual, D. and M. W. Watson (2007). Consistent estimation of the number of dynamic factors in a large N and T panel. *Journal of Business and Economic Statistics 25*, 91–96.

Anderson, B. D. and M. Deistler (2008a). Generalized linear dynamic factor models–a structure theory. *IEE Conference on Decision and Control*.

Anderson, B. D. and M. Deistler (2008b). Properties of zero-free transfer function matrices. *SICE Journal of Control, Measurement and System Integration 1*, 284–292.

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

Bai, J. (2004). Estimating cross-section common stochastic trends in nonstationary panel data. *Journal of Econometrics 122*, 137–183.

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.

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

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

Bai, J. and S. Ng (2008). Forecasting economic time series using targeted predictors. *Journal of Econometrics 146*(2), 304–317.

Bai, J. and S. Ng (2010). Panel unit root tests with cross-section dependence. *Econometric Theory 26*, 1088–1114.

Banerjee, A., M. Marcellino, and I. Masten (2014). Structural FECM: Cointegration in large-scale structural FAVAR models. Working Paper 9858, CEPR.

Barigozzi, M., A. M. Conti, and M. Luciani (2014). Do euro area countries respond asymmetrically to the common monetary policy? *Oxford Bulletin of Economics and Statistics 76*, 693–714.

Barigozzi, M., M. Lippi, and M. Luciani (2016). Dynamic factor models, cointegration, and error correction mechanisms. http://arxiv.org/abs/1510.02399.

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

Boivin, J. and M. Giannoni (2006). DSGE models in a data-rich environment. Technical Report Working Paper No. 12772, National Bureau of Economic Research.

Boivin, J. and S. Ng (2006). Are more data always better for factor analysis? *Journal of Econometrics 127*, 169–194.

Chamberlain, G. and M. Rothschild (1983). Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica 51*(5), 1281–304.

Cristadoro, R., M. Forni, L. Reichlin, and G. Veronese (2005). A core inflation indicator for the euro area. *Journal of Money Credit and Banking 37*(3), 539–560.

Davis, C. and W. M. Kahan (1970). The rotation of eigenvectors by a perturbation. III. *SIAM Journal* *on Numerical Analysis 7*, 1–46.

Dedola, L. and S. Neri (2007). What does a technology shock do? A VAR analysis with model-based sign restrictions. *Journal of Monetary Economics 54*, 512–549.

Del Negro, M., F. Schorfheide, F. Smets, and R. Wouters (2007). On the fit of New Keynesian models. *Journal of Business & Economic Statistics 25*, 123–143.

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.

Eickmeier, S. (2009). Comovements and heterogeneity in the euro area analyzed in a non-stationary dynamic factor model. *Journal of Applied Econometrics 24*, 933–959.

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.

Forni, M. and L. Gambetti (2010). The dynamic effects of monetary policy: A structural factor model approach. *Journal of Monetary Economics 57*, 203–216.

Forni, M., L. Gambetti, and L. Sala (2014). No news in business cycles. *Economic Journal 124*, 1168–1191.

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.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2005). The Generalized Dynamic Factor Model: One sided estimation and forecasting. *Journal of the American Statistical Association 100*, 830–840.

Forni, M., M. Hallin, M. Lippi, and P. Zaffaroni (2015a). Dynamic factor models with infinite dimensional factor space: Asymptotic analysis. Working Papers Series 2015-23, ECARES.

Forni, M., M. Hallin, M. Lippi, and P. Zaffaroni (2015b). Dynamic factor models with infinite-dimensional factor spaces: One-sided representations. *Journal of Econometrics 185*, 359–371.

Forni, M. and M. Lippi (2001). The Generalized Dynamic Factor Model: Representation theory. *Econometric Theory 17*, 1113–1141.

Gengenbach, C., J.-P. Urbain, and J. Westerlund (2015). Error correction testing in panels with common stochastic trends. *Journal of Applied Econometrics*. available online.

Giannone, D., M. Lenza, and G. E. Primicieri (2015). Prior selection for vector autoregressions. *The Review of Economics and Statistics 97*, 436–451.

Giannone, D., L. Reichlin, and L. Sala (2005). Monetary policy in real time. In M. Gertler and K. Rogoff (Eds.), *NBER Macroeconomics Annual 2004*. MIT Press.

Giannone, D., L. Reichlin, and L. Sala (2006). VARs, common factors and the empirical validation of equilibrium business cycle models. *Journal of Econometrics 132*, 257–279.

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

Gonzalo, J. (1994). Five alternative methods of estimating long-run equilibrium relationships. *Journal of Econometrics 60*, 203–233.

Gonzalo, J. and C. Granger (1995). Estimation of common long-memory components in cointegrated systems. *Journal of Business & 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 R. Liška (2007). Determining the number of factors in the general dynamic factor model. *Journal of the American Statistical Association 102*, 603–617.

Hallin, M., R. Van den Akker, and B. J. Werker (2016). Semiparametric error-correction models for cointegration with trends: pseudo-Gaussian and optimal rank-based tests of the cointegration rank. *Journal of Econometrics 190*, 46–61.

Hamilton, J. D. (1994). *Time Series Analysis.* Princeton, New Jersey: Princeton University Press.

Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control 12*, 231–254.

Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica 59*, 1551–80.

Johansen, S. (1995). *Likelihood-based inference in cointegrated vector autoregressive models* (First ed.). Oxford: Oxford University Press.

Johnstone, I. M. and A. Y. Lu (2009). On consistency and sparsity for principal components analysis in high dimensions. *Journal of the American Statistical Association 104*, 682–693.

King, R., C. Plosser, J. H. Stock, and M. W. Watson (1991). Stochastic trends and economic fluctuations. *American Economic Review 81*, 819–840.

Luciani, M. (2014). Forecasting with approximate dynamic factor models: The role of non-pervasive shocks. *International Journal of Forecasting 30*, 20–29.

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

Lütkepohl, H. (2006). Structural vector autoregressive analysis for cointegrated variables. In O. Hübler and J. Frohn (Eds.), *Modern Econometric Analysis. Surveys on Recent Developments*. Springer Berlin Heidelberg.

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.

Onatski, A. (2009). Testing hypotheses about the number of factors in large factor models. *Econometrica 77*, 1447–1479.

Onatski, A. (2010). Determining the number of factors from empirical distribution of eigenvalues. *The Review of Economics and Statistics 92*, 1004–1016.

Paruolo, P. (1997). Asymptotic inference on the moving average impact matrix in cointegrated $I(1)$ VAR systems. *Econometric Theory 13*, 79–118.

Peña, D. and P. Poncela (2004). Nonstationary dynamic factor analysis. *Journal of Statistical Planning and Inference 136*, 1237–1257.

Phillips, P. C. (1991). Optimal inference in cointegrated systems. *Econometrica 59*, 238–306.

Phillips, P. C. (1995). Fully modified least squares and vector autoregression. *Econometrica 63*, 1023–1078.

Phillips, P. C. (1998). Impulse response and forecast error variance asymptotics in nonstationary VARs. *Journal of Econometrics 83*, 21–56.

Phillips, P. C. and S. N. Durlauf (1986). Multiple time series regression with integrated processes. *The Review of Economic Studies 53*, 473–495.

Phillips, P. C. and B. E. Hansen (1990). Statistical inference in instrumental variables regressions with $I(1)$ processes,. *The Review of Economic Studies 57*, 99–125.

Phillips, P. C. and S. Ouliaris (1988). Testing for cointegration using principal components methods. *Journal of Economic Dynamics and Control 12*, 205–230.

Phillips, P. C. and V. Solo (1992). Asymptotics for linear processes. *The Annals of Statistics*, 971–1001.

Sims, C., J. H. Stock, and M. W. Watson (1990). Inference in linear time series models with some unit roots. *Econometrica 58*, 113–144.

Smets, F. and R. Wouters (2007). Shocks and frictions in US business cycles: A Bayesian DSGE approach. *American Economic Review 97*, 586–606.

Stock, J. H. (1987). Asymptotic properties of least squares estimators of cointegrating vectors. *Econometrica 55*, 1035–1056.

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 (1993). A simple estimator of cointegrating vectors in higher order integrated systems. *Econometrica 61*, 783–820.

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

Stock, J. H. and M. W. Watson (2002b). Macroeconomic forecasting using diffusion indexes. *Journal of Business and Economic Statistics 20*, 147–162.

Stock, J. H. and M. W. Watson (2005). Implications of dynamic factor models for VAR analysis. Working Paper 11467, NBER.

Vahid, F. and R. F. Engle (1993). Common trends and common cycles. *Journal of Applied Econometrics 8*, 341–360.

Velu, R. P., G. C. Reinsel, and D. W. Wichern (1986). Reduced rank models for multiple time series. *Biometrika 73*, 105–118.

Watson, M. W. (1994). Vector autoregressions and cointegration. In R. Engle and D. McFadden (Eds.), *Handbook of Econometrics, Volume IV*. Elsevier Science.

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.

Yu, Y., T. Wang, and R. J. Samworth (2015). A useful variant of the Davis–Kahan theorem for statisticians. *Biometrika 102*, 315–323.

Appendix A Proof of main results

Preliminaries

Norms For any $m \times p$ matrix $\mathbf{B}$ with generic element b_{ij}, we denote its spectral norm as $\|\mathbf{B}\| = \sqrt{\mu_1^{\mathbf{B}'\mathbf{B}}}$, where $\mu_1^{\mathbf{B}'\mathbf{B}}$ is the largest eigenvalue of $\mathbf{B}'\mathbf{B}$, the Frobenius norm as $\|\mathbf{B}\|_F = \sqrt{\mathrm{tr}(\mathbf{B}'\mathbf{B})} = \sqrt{\sum_i \sum_j b_{ij}^2}$, and the column and row norm as $\|\mathbf{B}\|_1 = \max_j \sum_i |b_{ij}|$ and $\|\mathbf{B}\|_\infty = \max_i \sum_j |b_{ij}|$, respectively. We use the following properties.

1. Subadditivity of the norm, for an $m \times p$ matrix $\mathbf{A}$ and a $p \times s$ matrix $\mathbf{B}$:

$$\|\mathbf{A}\mathbf{B}\| \leq \|\mathbf{A}\| \, \|\mathbf{B}\|. \tag{A1}$$

2. Norm inequalities, for an $n \times n$ symmetric matrix $\mathbf{A}$:

$$\mu_1^A = \|\mathbf{A}\| \leq \sqrt{\|\mathbf{A}\|_1 \, \|\mathbf{A}\|_\infty} = \|\mathbf{A}\|_1, \qquad \|\mathbf{A}\| \leq \|\mathbf{A}\|_F. \tag{A2}$$

3. Weyl's inequality, for two $n \times n$ symmetric matrices $\mathbf{A}$ and $\mathbf{B}$, with eigenvalues μ_j^A and μ_j^B

$$|\mu_j^A - \mu_j^B| \leq \|\mathbf{A} - \mathbf{B}\|, \qquad j = 1, \dots, n. \tag{A3}$$

Processes In the following, for $j = 1, \dots, r$ and $i = 1, \dots, n$, it is useful to write (11) and (14) as

$$\Delta F_{jt} = \mathbf{c}_j'(L)\mathbf{u}_t, \quad \Delta \xi_{it} = \check{d}_i(L)\varepsilon_{it},$$

where $\mathbf{c}_j(L)$ is an $q \times 1$ one-sided infinite filter with entries $c_{jl}(L)$ for $l = 1, \dots, q$, $d_i(L)$ and $\check{d}_i(L)$ are a one-sided infinite filters such that $\check{d}_i(L) = (1-L)(1-\rho_i L)^{-1} d_i(L)$ with $\rho_i = 1$ if $i = 1, \dots m$ and $|\rho_i| < 1$ if $i = m+1, \dots n$. By Assumptions 2c and 4b, we have also $F_{jt} = \sum_{s=1}^t \mathbf{c}_j'(L)\mathbf{u}_s$, and $\xi_{it} = \sum_{s=1}^t \check{d}_i(L)\varepsilon_{is}$, which is non-stationary only if $i \leq m$. Finally, by Assumption 2d and 4c, the filter coefficients are square-summable, hence there exist two positive finite constants K_1 and K_2 such that

$$\sup_{\substack{j=1,\dots,r \\ l=1,\dots,q}} \sum_{k=0}^\infty c_{jlk}^2 \leq K_1 < \infty, \qquad \sup_{i=1,\dots,n} \sum_{k=0}^\infty \check{d}_{ik}^2 \leq K_2 < \infty. \tag{A4}$$

Rates We define $\zeta_{nT,\delta} = \max(T^{1/2} n^{-(2-\delta)/2}, n^{-(1-\delta)/2})$, with $\delta \geq 0$. Notice that

$$\zeta_{nT,\delta} = \begin{cases} T^{1/2} n^{-(2-\delta)/2} & \text{if} \quad n < T, \\ T^{-(1-\delta)/2} = n^{-(1-\delta)/2} & \text{if} \quad n = T, \\ n^{-(1-\delta)/2} & \text{if} \quad n > T, \end{cases} \tag{A5}$$

and, under Assumption 6, we have $\zeta_{nT,\delta} \to 0$ as $n, T \to \infty$. Define also

$$\vartheta_{nT,\delta} = \max\left(\zeta_{nT,\delta}, T^{-1/2}\right), \tag{A6}$$

and we have $\vartheta_{nT,\delta} = T^{-1/2}$ if and only if $n > T^{1/(1-\delta)}$.

Separation of static and dynamic eigenvalues For proving Lemma 3 and Proposition 4 we also assume that the eigenvalues of the covariance matrix, $\mu_j^{\Delta\chi}$, and of the spectral density matrix $\mu_j^{\Delta\chi}(\theta)$, $\theta \in [-\pi, \pi]$, of the common component are distinct. Notice that linear divergence of eigenvalues is not assumed but implied by Lemma 2 and 7.

Assumption 7 ***(Eigenvalues)***

a. *There exist real numbers α_j and β_j, $j = 1, \ldots, r-1$, such that for any $n \in \mathbb{N}$*

$$0 < \underline{M}_5 \leq \alpha_{j+1} \leq \frac{\mu_{j+1}^{\Delta\chi}}{n} \leq \beta_{j+1} < \alpha_j \leq \frac{\mu_j^{\Delta\chi}}{n} \leq \beta_j \leq \overline{M}_5 < \infty,$$

where $\underline{M}_5$ and $\overline{M}_5$ are defined in Lemma 2.

b. *There exist continuous functions $\alpha_j(\theta)$ and $\beta_j(\theta)$, $j = 1, \ldots, r-1$, such that for any $n \in \mathbb{N}$ and any $\theta \in [-\pi, \pi]$*

$$\alpha_{j+1}(\theta) \leq \frac{\mu_{j+1}^{\Delta\chi}(\theta)}{n} \leq \beta_{j+1}(\theta) < \alpha_j(\theta) \leq \frac{\mu_j^{\Delta\chi}(\theta)}{n} \leq \beta_j(\theta) \leq \overline{M}_9 < \infty,$$

and if $\theta \neq 0$ then $\alpha_r(\theta) \geq \underline{M}_9 > 0$, where $\overline{M}_9$ and $\underline{M}_9$ are defined in Lemma 7.

Propositions

Proof of Proposition 2

The estimated VECM with $p = 1$ can always be written as a VAR(2) with estimated matrix polynomial, $\widehat{\mathbf{A}}^{\text{VECM}}(L) = \mathbf{I}_r - \widehat{\mathbf{A}}_1^{\text{VECM}} L - \widehat{\mathbf{A}}_2^{\text{VECM}} L^2$, where $\widehat{\mathbf{A}}_1^{\text{VECM}} = \widehat{\mathbf{G}}_1 + \widehat{\boldsymbol{\alpha}}\widehat{\boldsymbol{\beta}}' + \mathbf{I}_r$, and $\widehat{\mathbf{A}}_2^{\text{VECM}} = -\widehat{\mathbf{G}}_1$. Then, from Lemma 5*i*, 5*ii*, and 5*iii*, we have, for $k = 1, 2$,

$$\left\|\widehat{\mathbf{A}}_k^{\text{VECM}} - \mathbf{H}\mathbf{A}_k\mathbf{H}'\right\| = O_p(\vartheta_{nT,\delta}). \tag{A7}$$

Define the infinite matrix polynomial

$$\widehat{\mathbf{B}}(L) = \left[\widehat{\mathbf{A}}^{\text{VECM}}(L)\right]^{-1} = (\mathbf{I}_r - \widehat{\mathbf{A}}_1^{\text{VECM}} L - \widehat{\mathbf{A}}_2^{\text{VECM}} L^2)^{-1} = \sum_{k=0}^{\infty} \widehat{\mathbf{B}}_k L^k,$$

such that $\widehat{\mathbf{B}}(0) = \mathbf{I}_r$, $\widehat{\mathbf{B}}_1 = \widehat{\mathbf{A}}_1^{\text{VECM}}$, $\widehat{\mathbf{B}}_2 = (\widehat{\mathbf{A}}_1^{\text{VECM}}\widehat{\mathbf{B}}_1 + \widehat{\mathbf{A}}_2^{\text{VECM}})$, $\widehat{\mathbf{B}}_3 = (\widehat{\mathbf{A}}_1^{\text{VECM}}\widehat{\mathbf{B}}_2 + \widehat{\mathbf{A}}_2^{\text{VECM}}\widehat{\mathbf{B}}_1)$, and so on. Then, from (A7), we have, for any $k \geq 0$,

$$\left\|\widehat{\mathbf{B}}_k - \mathbf{H}\mathbf{B}_k\mathbf{H}'\right\| = O_p(\vartheta_{nT,\delta}). \tag{A8}$$

The estimated impulse response of variable i is then a q-dimensional row vector defined as (see (19))

$$\widehat{\boldsymbol{\phi}}_i^{\text{VECM}\prime}(L) = \widehat{\boldsymbol{\lambda}}_i'\widehat{\mathbf{B}}(L)\widehat{\mathbf{K}}\widehat{\mathbf{R}}',$$

where $\widehat{\boldsymbol{\lambda}}_i'$ is the i-th row of $\widehat{\boldsymbol{\Lambda}}$.

The matrix $\mathbf{R}$ is estimated by $\widehat{\mathbf{R}} \equiv \widehat{\mathbf{R}}(\widehat{\boldsymbol{\Lambda}}, \widehat{\mathbf{A}}^{\text{VECM}}(L), \widehat{\mathbf{K}})$. To estimate this mapping we have to impose $q(q+1)/2$ restrictions on the impulse response functions, i.e. at most only on $q(q+1)/2$

variables. So $\widehat{\mathbf{R}}$ depends only on $q(q+1)/2$ columns of $\widehat{\mathbf{\Lambda}}$ and for regular identification schemes, such that this mapping is analytical, using Lemmas 3 and $5iv$, we have (see Forni et al., 2009)

$$\|\widehat{\mathbf{R}} - \mathbf{R}\| = O_p(\vartheta_{nT,\delta}). \tag{A9}$$

Finally, from Lemma 3, we have, for any $i \in \mathbb{N}$,

$$\|\widehat{\boldsymbol{\lambda}}_i' - \boldsymbol{\lambda}_i'\mathbf{H}'\| = O_p\left(\max\left(\frac{1}{\sqrt{T}}, \frac{1}{\sqrt{n}}\right)\right). \tag{A10}$$

Therefore, for any $i \in \mathbb{N}$ and $k \geq 0$, we have

$$\begin{aligned}
&\|\widehat{\boldsymbol{\phi}}_{ik}^{\text{VECM}'} - \boldsymbol{\phi}_{ik}^{\text{VECM}'}\| = \|\widehat{\boldsymbol{\lambda}}_i'\widehat{\mathbf{B}}_k\widehat{\mathbf{K}}\widehat{\mathbf{R}}' - \boldsymbol{\lambda}_i'\mathbf{B}_k\mathbf{K}\| \\
=&\|(\widehat{\boldsymbol{\lambda}}_i' - \boldsymbol{\lambda}_i'\mathbf{H}' + \boldsymbol{\lambda}_i'\mathbf{H}')(\widehat{\mathbf{B}}_k - \mathbf{H}\mathbf{B}_k\mathbf{H}' + \mathbf{H}\mathbf{B}_k\mathbf{H}')(\widehat{\mathbf{K}} - \mathbf{H}\mathbf{K}\mathbf{R} + \mathbf{H}\mathbf{K}\mathbf{R})(\widehat{\mathbf{R}}' - \mathbf{R}' + \mathbf{R}') - \boldsymbol{\lambda}_i'\mathbf{B}_k\mathbf{K}\| \\
\leq&\|\widehat{\boldsymbol{\lambda}}_i' - \boldsymbol{\lambda}_i'\mathbf{H}'\| \, \|\mathbf{H}\mathbf{B}_k\mathbf{H}'\mathbf{H}\mathbf{K}\mathbf{R}\mathbf{R}'\| + \|\boldsymbol{\lambda}_i'\mathbf{H}'\| \, \|\widehat{\mathbf{B}}_k - \mathbf{H}\mathbf{B}_k\mathbf{H}'\| \, \|\mathbf{H}\mathbf{K}\mathbf{R}\mathbf{R}'\| \\
&+ \|\boldsymbol{\lambda}_i'\mathbf{H}'\mathbf{H}\mathbf{B}_k\mathbf{H}'\| \, \|\widehat{\mathbf{K}} - \mathbf{H}\mathbf{K}\mathbf{R}\| \, \|\mathbf{R}'\| + \|\boldsymbol{\lambda}_i'\mathbf{H}'\mathbf{H}\mathbf{B}_k\mathbf{H}'\mathbf{H}\mathbf{K}\mathbf{R}\| \, \|\widehat{\mathbf{R}}' - \mathbf{R}'\| \\
&+ \|\boldsymbol{\lambda}_i'\mathbf{H}'\mathbf{H}\mathbf{B}_k\mathbf{H}'\mathbf{H}\mathbf{K}\mathbf{R}\mathbf{R}' - \boldsymbol{\lambda}_i'\mathbf{B}_k\mathbf{K}\| + o_p(\vartheta_{nT,\delta}) = O_p\left(\max\left(\frac{1}{\sqrt{T}}, \frac{1}{\sqrt{n}}\right)\right) + O_p(\vartheta_{nT,\delta}),
\end{aligned}$$

where we used (A8), (A9), and (A10), Lemma 5, orthogonality of $\mathbf{H}$ and $\mathbf{R}$, and the fact that $\mathbf{H}$, $\mathbf{R}$, $\mathbf{K}$, $\mathbf{B}_k$, $\boldsymbol{\lambda}_i$ are all finite matrices, which do not depend on n nor on T. By (A5) and (A6) it is clear that the rate is always $\vartheta_{nT,\delta}$. This completes the proof. □

Proof of Proposition 3

Define

$$\widehat{\mathbf{B}}(L) = \left[\widehat{\mathbf{A}}^{\text{VAR}}(L)\right]^{-1} = (\mathbf{I}_r - \widehat{\mathbf{A}}_1^{\text{VAR}}L)^{-1} = \sum_{k=0}^{\infty}\widehat{\mathbf{B}}_k L^k,$$

such that $\widehat{\mathbf{B}}_k = (\widehat{\mathbf{A}}_1^{\text{VAR}})^k$. Then, from Lemma $6i$, we have, for any finite $k \geq 0$,

$$\|\widehat{\mathbf{B}}_k - \mathbf{H}\mathbf{B}_k\mathbf{H}'\| = O_p\left(\max\left(\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{T}}\right)\right). \tag{A11}$$

If instead $k \to \infty$, then $\widehat{\mathbf{B}}_k$ has as limit for $n, T \to \infty$ a random variable rather than $\mathbf{B}_k$ (see Theorem 3.2 in Phillips, 1998), hence $\lim_{k\to\infty} \|\widehat{\mathbf{B}}_k - \mathbf{B}_k\| = O_p(1)$.

The estimated impulse response of variable i is then the q-dimensional row vector (see (21))

$$\widehat{\boldsymbol{\phi}}_i^{\text{VAR}'}(L) = \widehat{\boldsymbol{\lambda}}_i'\widehat{\mathbf{B}}(L)\widehat{\mathbf{K}}\widehat{\mathbf{R}}', \tag{A12}$$

where $\widehat{\boldsymbol{\lambda}}_i'$ is the i-th row of $\widehat{\mathbf{\Lambda}}$ and $\widehat{\mathbf{R}} \equiv \widehat{\mathbf{R}}(\widehat{\mathbf{\Lambda}}, \widehat{\mathbf{A}}^{\text{VAR}}(L), \widehat{\mathbf{K}})$ is a consistent estimator of the matrix $\mathbf{R}$, such that, because of Lemmas 3 and $6ii$,

$$\|\widehat{\mathbf{R}} - \mathbf{R}\| = O_p\left(\max\left(\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{T}}\right)\right). \tag{A13}$$

Consistency of the estimated impulse responses (A12), at each finite lag k, is then proved exactly as in the proof of Proposition 2. □

Proofs of main Lemmas

Proof of Lemma 1

First notice that, from Assumption 4*f*, we have

$$\frac{1}{n}\sum_{i,j=1}^n |\mathsf{E}[\varepsilon_{it}\varepsilon_{jt}]| \le \max_{i=1,\dots,n}\sum_{j=1}^n |\mathsf{E}[\varepsilon_{it}\varepsilon_{jt}]| \le M_4 < \infty.$$

Moreover, Assumption $4f$ reads $\|\mathbf{\Gamma}_0^\varepsilon\|_1 \le M_4$, thus, from (A2), we have

$$\mu_1^\varepsilon = \left\|\mathbf{\Gamma}_0^\varepsilon\right\| \le \left\|\mathbf{\Gamma}_0^\varepsilon\right\|_1 \le M_4 < \infty.$$

This completes the proof. □

Proof of Lemma 2

For part *i)*, first notice that the covariance of the first difference of common factors can be written as $\mathbf{\Gamma}_0^{\Delta F} = \mathbf{W}^{\Delta F}\mathbf{M}^{\Delta F}\mathbf{W}^{\Delta F\prime}$, where $\mathbf{W}^{\Delta F}$ is the $r \times r$ matrix of normalized eigenvectors and $\mathbf{M}^{\Delta F}$ the corresponding diagonal matrix of eigenvalues.

Now, define a new $n \times r$ loadings matrix $\boldsymbol{L} = \mathbf{\Lambda}\mathbf{W}^{\Delta F}(\mathbf{M}^{\Delta F})^{1/2}$. This matrix satisfies Assumption 3 since when $n^{-1}\mathbf{\Lambda}'\mathbf{\Lambda} = \mathbf{I}_r$

$$\frac{\boldsymbol{L}'\boldsymbol{L}}{n} = \mathbf{M}^{\Delta F}, \tag{A14}$$

and by Assumption $2d$ and g all eigenvalues of $\mathbf{\Gamma}_0^{\Delta F}$ are positive and finite, i.e. there exists constants $\underline{M}_5$ and $\overline{M}_5$ such that

$$0 < \underline{M}_5 \le \mu_j^{\Delta F} \le \overline{M}_5 < \infty, \quad j = 1, \dots, r. \tag{A15}$$

Then, the covariance matrix of the first differences of the common component is given by

$$\frac{\mathbf{\Gamma}_0^{\Delta\chi}}{n} = \frac{\mathbf{\Lambda}\mathbf{W}^{\Delta F}\mathbf{M}^{\Delta F}\mathbf{W}^{\Delta F\prime}\mathbf{\Lambda}'}{n} = \frac{\boldsymbol{L}\boldsymbol{L}'}{n}.$$

Therefore, the non-zero eigenvalues of $\mathbf{\Gamma}_0^{\Delta\chi}$ are the same as those of $\boldsymbol{L}'\boldsymbol{L}$, and from (A14), we have for any n, $n^{-1}\mu_j^{\Delta\chi} = \mu_j^{\Delta F}$, for any $j = 1, \dots, r$. Part *i)* then follows from (A15).

As for part *ii)*, we have

$$\mu_1^{\Delta\xi} = \left\|\mathbf{\Gamma}_0^{\Delta\xi}\right\| \le \sum_{k=0}^\infty \left\|\check{\mathbf{D}}_k\right\|^2 \left\|\mathbf{\Gamma}_0^\varepsilon\right\| \le K_2 M_4 = M_6 < \infty, \tag{A16}$$

which follows from Assumption $4c$, which implies square summability of the idiosyncratic MA filters, and from Lemma 1.

Finally, parts *iii)* and *iv)* are immediate consequences of Assumption 5 which implies that $\mathbf{\Gamma}_0^{\Delta x} =$

$\mathbf{\Gamma}_0^{\Delta\chi} + \mathbf{\Gamma}_0^{\Delta\xi}$ and of Weyl's inequality (A3). So, for $j = 1, \dots, r$, and for any $n \in \mathbb{N}$, we have

$$\frac{\mu_j^{\Delta x}}{n} \leq \frac{\mu_j^{\Delta\chi}}{n} + \frac{\mu_1^{\Delta\xi}}{n} \leq \overline{M}_5 + \frac{\mu_1^{\Delta\xi}}{n} \leq \overline{M}_5 + \frac{M_6}{n} = \overline{M}_7 < \infty,$$
$$\frac{\mu_j^{\Delta x}}{n} \geq \frac{\mu_j^{\Delta\chi}}{n} + \frac{\mu_n^{\Delta\xi}}{n} \geq \underline{M}_5 + \frac{\mu_n^{\Delta\xi}}{n} = \underline{M}_7 > 0,$$

because of parts *i)* and *ii)*. This proves part *iii)*. When $j = r + 1$, using parts *i)* and *ii)* above, and since $\mathrm{rk}(\mathbf{\Gamma}_0^{\Delta\chi}) = r$, we have $\mu_{r+1}^{\Delta x} \leq \mu_{r+1}^{\Delta\chi} + \mu_1^{\Delta\xi} = \mu_1^{\Delta\xi} \leq M_6 < \infty$, thus proving part *iv)*. This completes the proof. □

Proof of Lemma 3

The sample covariance of $\Delta\mathbf{x}_t$ is given by $\widehat{\mathbf{\Gamma}}_0^{\Delta x} = T^{-1}\sum_{t=1}^T \Delta\mathbf{x}_t\Delta\mathbf{x}_t'$ and from Assumption 5 we have $\mathbf{\Gamma}_0^{\Delta x} = \mathbf{\Gamma}_0^{\Delta\chi} + \mathbf{\Gamma}_0^{\Delta\xi}$. Moreover, from Lemma 8, we have

$$\left\| \frac{\widehat{\mathbf{\Gamma}}_0^{\Delta x}}{n} - \frac{\mathbf{\Gamma}_0^{\Delta x}}{n} \right\| = O_p\left(\frac{1}{\sqrt{T}}\right). \tag{A17}$$

From (A17), (A16), and Assumption 5 we also have

$$\begin{aligned}\left\| \frac{\widehat{\mathbf{\Gamma}}_0^{\Delta x}}{n} - \frac{\mathbf{\Gamma}_0^{\Delta\chi}}{n} \right\| &\leq \left\| \frac{\widehat{\mathbf{\Gamma}}_0^{\Delta x}}{n} - \frac{\mathbf{\Gamma}_0^{\Delta x}}{n} \right\| + \left\| \frac{\mathbf{\Gamma}_0^{\Delta x}}{n} - \frac{\mathbf{\Gamma}_0^{\Delta\chi}}{n} \right\| = \left\| \frac{\widehat{\mathbf{\Gamma}}_0^{\Delta x}}{n} - \frac{\mathbf{\Gamma}_0^{\Delta x}}{n} \right\| + \left\| \frac{\mathbf{\Gamma}_0^{\Delta\xi}}{n} \right\| \\ &= O_p\left(\frac{1}{\sqrt{T}}\right) + \frac{\mu_1^{\Delta\xi}}{n} = O_p\left(\max\left(\frac{1}{\sqrt{T}}, \frac{1}{n}\right)\right). \end{aligned} \tag{A18}$$

Now, define as $\mathbf{w}_j^{\Delta\chi}$ and $\widehat{\mathbf{w}}_j^{\Delta x}$ the n-dimensional normalized eigenvectors corresponding to the j-th largest eigenvalues of $\mathbf{\Gamma}_0^{\Delta\chi}$ and $\widehat{\mathbf{\Gamma}}_0^{\Delta x}$, respectively. From Corollary 1 in Yu et al. (2015), which is a generalisation of the "$\sin\theta$" Theorem in Davis and Kahan (1970), we have

$$\left\| \widehat{\mathbf{w}}_j^{\Delta x} - s\mathbf{w}_j^{\Delta\chi} \right\| = \frac{2^{3/2}\|\widehat{\mathbf{\Gamma}}_0^{\Delta x} - \mathbf{\Gamma}_0^{\Delta\chi}\|}{\min\left(\mu_{j-1}^{\Delta\chi} - \mu_j^{\Delta\chi}, \mu_j^{\Delta\chi} - \mu_{j+1}^{\Delta\chi}\right)}, \qquad j = 1, \dots n, \tag{A19}$$

where $s = \pm 1$ and we define $\mu_0^{\Delta\chi} = \infty$ for any $n \in \mathbb{N}$. Because of Assumption $7a$ of distinct eigenvalues the denominator of (A19) is always positive. Moreover, from Assumption $7a$ and Lemma $2i$ there exists positive finite constants c_1, c_2, c_3 such that (notice that when $j = 1$ the denominator is always given by the second term)

$$\begin{aligned} &\mu_{j-1}^{\Delta\chi} - \mu_j^{\Delta\chi} \geq n(\alpha_{j-1} - \beta_j) > nc_1, \qquad 1 < j \leq r, \\ &\mu_j^{\Delta\chi} - \mu_{j+1}^{\Delta\chi} \geq n(\alpha_j - \beta_{j+1}) > nc_2, \qquad 1 \leq j < r, \\ &\mu_r^{\Delta\chi} - \mu_{r+1}^{\Delta\chi} \geq n\alpha_r > nc_3. \end{aligned} \tag{A20}$$

Then, using (A18) and (A20) in (A19), there exists a positive finite constant c_4 such that

$$\left\| \widehat{\mathbf{w}}_j^{\Delta x} - s\mathbf{w}_j^{\Delta\chi} \right\| \leq c_4 \frac{\|\widehat{\mathbf{\Gamma}}_0^{\Delta x} - \mathbf{\Gamma}_0^{\Delta\chi}\|}{n} = O_p\left(\max\left(\frac{1}{\sqrt{T}}, \frac{1}{n}\right)\right), \qquad j = 1, \dots r. \tag{A21}$$

We define the $n \times r$ matrices of normalized eigenvectors as $\mathbf{W}^{\Delta\chi} = (\mathbf{w}_1^{\Delta\chi} \ldots \mathbf{w}_r^{\Delta\chi})$ and $\widehat{\mathbf{W}}^{\Delta x} = (\widehat{\mathbf{w}}_1^{\Delta x} \ldots \widehat{\mathbf{w}}_r^{\Delta x})$. Then, by (A21) there exists an $r \times r$ diagonal matrix $\mathbf{J}$ with entries 1 or -1 such that

$$\left\|\widehat{\mathbf{W}}^{\Delta x} - \mathbf{W}^{\Delta\chi}\mathbf{J}\right\| = O_p\left(\max\left(\frac{1}{\sqrt{T}}, \frac{1}{n}\right)\right). \tag{A22}$$

Notice that the same result is obtained in Lemma 3 in Forni et al. (2009) with a different proof, and could also be obtained by applying directly Theorem 1 in Yu et al. (2015) where also estimated eigenvalues are used.

The loadings estimator is defined as $\widehat{\mathbf{\Lambda}} = n^{1/2}\widehat{\mathbf{W}}^{\Delta x}$, therefore, by construction $n^{-1}\widehat{\mathbf{\Lambda}}'\widehat{\mathbf{\Lambda}} = \mathbf{I}_r$. Then, it is always possible to define an orthogonal matrix $\mathbf{H}$ such that $\mathbf{\Lambda} = \sqrt{n}\mathbf{W}^{\Delta\chi}\mathbf{J}\mathbf{H}$. This choice of $\mathbf{H}$ is such that Assumption 3a is trivially satisfied since $n^{-1}\mathbf{\Lambda}'\mathbf{\Lambda} = \mathbf{H}'\mathbf{H} = \mathbf{I}_r$. By substituting in (A22) we have

$$\left\|\widehat{\mathbf{W}}^{\Delta x} - \mathbf{W}^{\Delta\chi}\mathbf{J}\right\| = \left\|\frac{\widehat{\mathbf{\Lambda}} - \mathbf{\Lambda}\mathbf{H}'}{\sqrt{n}}\right\| = O_p\left(\max\left(\frac{1}{\sqrt{T}}, \frac{1}{n}\right)\right), \tag{A23}$$

which proves part *i)*. Part *ii)* is then proved straightforwardly. Notice that, if Assumption 6 holds, then $n > T^{1/(2-\delta)}$ with $\delta \geq 0$, the lower bound for n being then $n > T^{1/2}$, and, therefore, (A23) is $O_p(T^{-1/2})$.

In order to prove part *iii)*, we need some other intermediate results. We denote as $\boldsymbol{\epsilon}_i$ an n-dimensional vector with 1 as i-th entry and all other entries equal to zero. Then,

$$\begin{aligned}\left\|\frac{\boldsymbol{\epsilon}_i'}{\sqrt{n}}\left(\widehat{\mathbf{\Gamma}}_0^{\Delta x} - \mathbf{\Gamma}_0^{\Delta\chi}\right)\right\| &\leq \left\|\frac{\boldsymbol{\epsilon}_i'}{\sqrt{n}}\left(\widehat{\mathbf{\Gamma}}_0^{\Delta x} - \mathbf{\Gamma}_0^{\Delta x}\right)\right\| + \left\|\frac{\boldsymbol{\epsilon}_i'\mathbf{\Gamma}_0^{\Delta\xi}}{\sqrt{n}}\right\| \leq \left\|\frac{\boldsymbol{\epsilon}_i'}{\sqrt{n}}\left(\widehat{\mathbf{\Gamma}}_0^{\Delta x} - \mathbf{\Gamma}_0^{\Delta x}\right)\right\|_F + \left\|\frac{\boldsymbol{\epsilon}_i'\mathbf{\Gamma}_0^{\Delta\xi}}{\sqrt{n}}\right\| \\ &\leq \sqrt{\frac{1}{n}\sum_{j=1}^n\left(\widehat{\gamma}_{ij}^{\Delta x} - \gamma_{ij}^{\Delta x}\right)^2} + \frac{\mu_1^{\Delta\xi}}{\sqrt{n}} = O_p\left(\max\left(\frac{1}{\sqrt{T}}, \frac{1}{\sqrt{n}}\right)\right),\end{aligned} \tag{A24}$$

where we used Lemma 8 and (A16). Similarly, we can show that

$$\left\|\frac{\boldsymbol{\epsilon}_i'\mathbf{\Gamma}_0^{\Delta\chi}}{\sqrt{n}}\right\| = O(1). \tag{A25}$$

For the eigenvalues $\mu_j^{\Delta\chi}$ of $\mathbf{\Gamma}_0^{\Delta\chi}$ and $\widehat{\mu}_j^{\Delta x}$ of $\widehat{\mathbf{\Gamma}}_0^{\Delta x}$, and using Weyl's inequality (A3), we have

$$\left|\frac{\widehat{\mu}_j^{\Delta x}}{n} - \frac{\mu_j^{\Delta\chi}}{n}\right| \leq \left\|\frac{\widehat{\mathbf{\Gamma}}_0^{\Delta x}}{n} - \frac{\mathbf{\Gamma}_0^{\Delta\chi}}{n}\right\| = O_p\left(\max\left(\frac{1}{\sqrt{T}}, \frac{1}{n}\right)\right), \qquad j = 1, \ldots, r. \tag{A26}$$

From Lemma 2i and (A26), we also have

$$\frac{\mu_r^{\Delta\chi}}{n} \geq \underline{M}_5 > 0, \qquad \frac{\widehat{\mu}_r^{\Delta x}}{n} \geq \underline{M}_5 + O_p\left(\max\left(\frac{1}{\sqrt{T}}, \frac{1}{n}\right)\right). \tag{A27}$$

Define as $\mathbf{M}^{\Delta\chi}$ and $\widehat{\mathbf{M}}^{\Delta x}$ the diagonal $r \times r$ matrices with diagonal elements $\mu_j^{\Delta\chi}$ and $\widehat{\mu}_j^{\Delta x}$, respectively. Therefore, from (A27), the matrix $n^{-1}\mathbf{M}^{\Delta\chi}$ is invertible, the inverse of $n^{-1}\widehat{\mathbf{M}}^{\Delta x}$ exists with probability tending to one as $n, T \to \infty$, and (see also Lemma 2 in Forni et al., 2009)

$$\left\| \left(\frac{\mathbf{M}^{\Delta\chi}}{n} \right)^{-1} \right\| = \frac{n}{\mu_r^{\Delta\chi}} = O(1). \tag{A28}$$

Moreover, from (A26) and (A27), we have

$$\begin{aligned}
\left\| \left(\frac{\widehat{\mathbf{M}}^{\Delta x}}{n} \right)^{-1} - \left(\frac{\mathbf{M}^{\Delta\chi}}{n} \right)^{-1} \right\| &\le \left\| \left(\frac{\widehat{\mathbf{M}}^{\Delta x}}{n} \right)^{-1} - \left(\frac{\mathbf{M}^{\Delta\chi}}{n} \right)^{-1} \right\|_F = \sqrt{\sum_{j=1}^r \left(\frac{n}{\widehat{\mu}_j^{\Delta x}} - \frac{n}{\mu_j^{\Delta\chi}} \right)^2} \\
&\le \sum_{j=1}^r n \left| \frac{\widehat{\mu}_j^{\Delta x} - \mu_j^{\Delta\chi}}{\widehat{\mu}_j^{\Delta x} \mu_j^{\Delta\chi}} \right| \le \frac{r \, |\widehat{\mu}_1^{\Delta x} - \mu_1^{\Delta\chi}|}{n \underline{M}_5^2 + O_p\left(\max\left(\frac{1}{\sqrt{T}}, \frac{1}{n} \right) \right)} = O_p\left(\max\left(\frac{1}{\sqrt{T}}, \frac{1}{n} \right) \right).
\end{aligned} \tag{A29}$$

Finally, notice that the columns of $\mathbf{W}^{\Delta\chi}\mathbf{J}$ are also normalised eigenvectors of $\boldsymbol{\Gamma}_0^{\Delta\chi}$, that is $\boldsymbol{\Gamma}_0^{\Delta\chi}\mathbf{W}^{\Delta\chi}\mathbf{J} = \mathbf{W}^{\Delta\chi}\mathbf{J}\mathbf{M}^{\Delta\chi}$. Therefore, using (A22), (A24), (A25), (A28), and (A29),

$$\begin{aligned}
\left\| \sqrt{n}\boldsymbol{\epsilon}_i' \widehat{\mathbf{W}}^{\Delta x} - \sqrt{n}\boldsymbol{\epsilon}_i' \mathbf{W}^{\Delta\chi}\mathbf{J} \right\| &= \left\| \frac{\boldsymbol{\epsilon}_i'}{\sqrt{n}} \left[\widehat{\boldsymbol{\Gamma}}_0^{\Delta x} \widehat{\mathbf{W}}^{\Delta x} \left(\frac{\widehat{\mathbf{M}}^{\Delta x}}{n} \right)^{-1} - \boldsymbol{\Gamma}_0^{\Delta\chi} \mathbf{W}^{\Delta\chi}\mathbf{J} \left(\frac{\mathbf{M}^{\Delta\chi}}{n} \right)^{-1} \right] \right\| \\
&\le \left\| \frac{\boldsymbol{\epsilon}_i'}{\sqrt{n}} (\widehat{\boldsymbol{\Gamma}}_0^{\Delta x} - \boldsymbol{\Gamma}_0^{\Delta\chi}) \right\| \left\| \left(\frac{\mathbf{M}^{\Delta\chi}}{n} \right)^{-1} \right\| + \left\| \frac{\boldsymbol{\epsilon}_i' \boldsymbol{\Gamma}_0^{\Delta\chi}}{\sqrt{n}} \right\| \left\| \left(\frac{\widehat{\mathbf{M}}^{\Delta x}}{n} \right)^{-1} - \left(\frac{\mathbf{M}^{\Delta\chi}}{n} \right)^{-1} \right\| \\
&+ \left\| \widehat{\mathbf{W}}^{\Delta x} - \mathbf{W}^{\Delta\chi}\mathbf{J} \right\| \left\| \frac{\boldsymbol{\epsilon}_i' \boldsymbol{\Gamma}_0^{\Delta\chi}}{\sqrt{n}} \right\| \left\| \left(\frac{\mathbf{M}^{\Delta\chi}}{n} \right)^{-1} \right\| + o_p\left(\max\left(\frac{1}{\sqrt{T}}, \frac{1}{\sqrt{n}} \right) \right) \\
&= O_p\left(\max\left(\frac{1}{\sqrt{T}}, \frac{1}{\sqrt{n}} \right) \right).
\end{aligned}$$

By noticing that, by definition, $\boldsymbol{\lambda}_i'\mathbf{H}' = \sqrt{n}\boldsymbol{\epsilon}_i'\mathbf{W}^{\Delta\chi}\mathbf{J}$ and $\widehat{\boldsymbol{\lambda}}_i' = \sqrt{n}\boldsymbol{\epsilon}_i'\widehat{\mathbf{W}}^{\Delta x}$, we complete the proof. □

Proof of Lemma 4

Given the loadings estimator $\widehat{\boldsymbol{\Lambda}}$, the factors are estimated as $\widehat{\mathbf{F}}_t = n^{-1}\widehat{\boldsymbol{\Lambda}}'\mathbf{x}_t$ and therefore $\Delta\widehat{\mathbf{F}}_t = n^{-1}\widehat{\boldsymbol{\Lambda}}'\Delta\mathbf{x}_t$. Then,

$$\begin{aligned}
\left\| \Delta\widehat{\mathbf{F}}_t - \mathbf{H}\Delta\mathbf{F}_t \right\| &= \left\| \frac{\widehat{\boldsymbol{\Lambda}}'\Delta\mathbf{x}_t}{n} - \mathbf{H}\Delta\mathbf{F}_t \right\| \le \left\| \frac{\widehat{\boldsymbol{\Lambda}}'\boldsymbol{\Lambda}\Delta\mathbf{F}_t}{n} - \mathbf{H}\Delta\mathbf{F}_t \right\| + \left\| \frac{\widehat{\boldsymbol{\Lambda}}'\Delta\boldsymbol{\xi}_t}{n} \right\| \\
&\le \left\| \frac{\widehat{\boldsymbol{\Lambda}}'\boldsymbol{\Lambda}}{n} - \mathbf{H} \right\| \|\Delta\mathbf{F}_t\| + \left\| \frac{\widehat{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda}\mathbf{H}'}{\sqrt{n}} \right\| \left\| \frac{\Delta\boldsymbol{\xi}_t}{\sqrt{n}} \right\| + \left\| \frac{\boldsymbol{\Lambda}'\Delta\boldsymbol{\xi}_t}{n} \right\| \|\mathbf{H}\| \\
&= O_p\left(\max\left(\frac{1}{\sqrt{T}}, \frac{1}{\sqrt{n}} \right) \right),
\end{aligned}$$

where we used Lemmas 3*i*, 3*ii*, 10*i*, 10*iv*, and 10*vi*, and the fact that $\|\mathbf{H}\| = O(1)$. This proves part *i)*.

Similarly, for part *ii)*, we have

$$\begin{aligned}
\frac{1}{\sqrt{T}} \left\| \widehat{\mathbf{F}}_t - \mathbf{H}\mathbf{F}_t \right\| &\le \left\| \frac{\widehat{\boldsymbol{\Lambda}}'\boldsymbol{\Lambda}}{n} - \mathbf{H} \right\| \left\| \frac{\mathbf{F}_t}{\sqrt{T}} \right\| + \left\| \frac{\widehat{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda}\mathbf{H}'}{\sqrt{n}} \right\| \left\| \frac{\boldsymbol{\xi}_t}{\sqrt{nT}} \right\| + \left\| \frac{\boldsymbol{\Lambda}'\boldsymbol{\xi}_t}{n\sqrt{T}} \right\| \|\mathbf{H}\| \\
&= O_p\left(\max\left(\frac{1}{\sqrt{T}}, \frac{1}{\sqrt{n}} \right) \right),
\end{aligned}$$

where we used Lemmas $3i$, $3ii$, $10ii$, $10v$, and $10vii$, and the fact that $\|\mathbf{H}\| = O(1)$. This completes the proof. $\square$

Proof of Lemma 5

Define

$$\widehat{\mathbf{M}}_{00} = \frac{1}{T}\sum_{t=1}^{T}\Delta\widehat{\mathbf{F}}_t\Delta\widehat{\mathbf{F}}_t', \quad \widehat{\mathbf{M}}_{01} = \frac{1}{T}\sum_{t=1}^{T}\Delta\widehat{\mathbf{F}}_t\widehat{\mathbf{F}}_{t-1}', \quad \widehat{\mathbf{M}}_{02} = \frac{1}{T}\sum_{t=1}^{T}\Delta\widehat{\mathbf{F}}_t\Delta\widehat{\mathbf{F}}_{t-1}',$$
$$\widehat{\mathbf{M}}_{11} = \frac{1}{T}\sum_{t=1}^{T}\widehat{\mathbf{F}}_t\widehat{\mathbf{F}}_t', \quad \widehat{\mathbf{M}}_{21} = \frac{1}{T}\sum_{t=1}^{T}\Delta\widehat{\mathbf{F}}_{t-1}'\widehat{\mathbf{F}}_{t-1}, \quad \widehat{\mathbf{M}}_{22} = \frac{1}{T}\sum_{t=1}^{T}\Delta\widehat{\mathbf{F}}_{t-1}\Delta\widehat{\mathbf{F}}_{t-1}', \qquad \text{(A30)}$$

and

$$\widehat{\mathbf{S}}_{00} = \widehat{\mathbf{M}}_{00} - \widehat{\mathbf{M}}_{02}\widehat{\mathbf{M}}_{22}^{-1}\widehat{\mathbf{M}}_{20}, \quad \widehat{\mathbf{S}}_{01} = \widehat{\mathbf{M}}_{01} - \widehat{\mathbf{M}}_{02}\widehat{\mathbf{M}}_{22}^{-1}\widehat{\mathbf{M}}_{21}, \quad \widehat{\mathbf{S}}_{11} = \widehat{\mathbf{M}}_{11} - \widehat{\mathbf{M}}_{12}\widehat{\mathbf{M}}_{22}^{-1}\widehat{\mathbf{M}}_{21}, \qquad \text{(A31)}$$

where $\widehat{\mathbf{M}}_{10} = \widehat{\mathbf{M}}_{01}'$, $\widehat{\mathbf{M}}_{20} = \widehat{\mathbf{M}}_{02}'$, and $\widehat{\mathbf{M}}_{12} = \widehat{\mathbf{M}}_{21}'$. Notice that if we denote the residuals of the regression of $\Delta\widehat{\mathbf{F}}_t$ and of $\widehat{\mathbf{F}}_{t-1}$ on $\Delta\widehat{\mathbf{F}}_{t-1}$ as $\widehat{\mathbf{e}}_{0t}$ and $\widehat{\mathbf{e}}_{1t}$, respectively then the matrices in (A31) are equivalent to $\widehat{\mathbf{S}}_{ij} = T^{-1}\sum_{t=1}^{T}\widehat{\mathbf{e}}_{it}\widehat{\mathbf{e}}_{jt}'$.

We then denote by $\mathbf{M}_{ij}$, for $i,j = 0,1,2$ and $\mathbf{S}_{ij}$, for $i,j = 0,1$ the analogous of the matrices $\widehat{\mathbf{M}}_{ij}$ and $\widehat{\mathbf{S}}_{ij}$ defined in (A30) and (A31), respectively, but when computed using $\check{\mathbf{F}}_t = \mathbf{H}\mathbf{F}_t$. Finally, we denote as $\check{\boldsymbol{\beta}} = \mathbf{H}\boldsymbol{\beta}$ the matrix of cointegration vectors of $\check{\mathbf{F}}_t = \mathbf{H}\mathbf{F}_t$ and its orthogonal complement as $\check{\boldsymbol{\beta}}_\perp$, such that $\check{\boldsymbol{\beta}}_\perp'\check{\boldsymbol{\beta}} = \mathbf{0}_{r-c\times c}$.

Let us start from part *i)*. Consider the generalized eigenvalues problem

$$\det\left(\widehat{\mu}_j\widehat{\mathbf{S}}_{11} - \widehat{\mathbf{S}}_{10}\widehat{\mathbf{S}}_{00}^{-1}\widehat{\mathbf{S}}_{01}\right) = 0, \quad j = 1,\dots,r. \qquad \text{(A32)}$$

If $\widehat{\mathbf{U}}$ are the normalized eigenvectors of $\widehat{\mathbf{S}}_{11}^{-1/2}\widehat{\mathbf{S}}_{10}\widehat{\mathbf{S}}_{00}^{-1}\widehat{\mathbf{S}}_{01}\widehat{\mathbf{S}}_{11}^{-1/2}$, then $\widehat{\mathbf{P}} = \widehat{\mathbf{S}}_{11}^{-1/2}\widehat{\mathbf{U}}$ are eigenvectors of $\widehat{\mathbf{S}}_{11} - \widehat{\mathbf{S}}_{10}\widehat{\mathbf{S}}_{00}^{-1}\widehat{\mathbf{S}}_{01}$ with eigenvalues $\widehat{\mu}_j$. Then, the estimator $\widehat{\boldsymbol{\beta}}$ proposed by Johansen (1988, 1991, 1995) is given by the c columns of $\widehat{\mathbf{P}}$ corresponding to the c largest eigenvalues.

Analogously define $\widehat{\mathbf{U}}^0$ as the normalized eigenvectors of $\mathbf{S}_{11}^{-1/2}\mathbf{S}_{10}\mathbf{S}_{00}^{-1}\mathbf{S}_{01}\mathbf{S}_{11}^{-1/2}$ and define $\widehat{\mathbf{P}}^0 = \mathbf{S}_{11}^{-1/2}\widehat{\mathbf{U}}^0$. Then the estimator $\widehat{\boldsymbol{\beta}}^0$ that we would obtain if estimating a VECM on $\check{\mathbf{F}}_t$, is the matrix of the c columns of $\widehat{\mathbf{P}}^0$, corresponding to the c largest eigenvalues $\widehat{\mu}_j^0$ of $\mathbf{S}_{11} - \mathbf{S}_{10}\mathbf{S}_{00}^{-1}\mathbf{S}_{01}$, and such that

$$\det\left(\widehat{\mu}_j^0\mathbf{S}_{11} - \mathbf{S}_{10}\mathbf{S}_{00}^{-1}\mathbf{S}_{01}\right) = 0, \quad j = 1,\dots,r. \qquad \text{(A33)}$$

Notice that by definition the two estimators $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\beta}}^0$ are normalized in such a way that $\widehat{\boldsymbol{\beta}}'\widehat{\mathbf{S}}_{11}\widehat{\boldsymbol{\beta}} = \mathbf{I}_c$ and $\widehat{\boldsymbol{\beta}}^{0\prime}\mathbf{S}_{11}\widehat{\boldsymbol{\beta}}^0 = \mathbf{I}_c$.

Consider then the $r\times r$ matrix

$$\mathbf{A}_T = \begin{pmatrix}\check{\boldsymbol{\beta}} & \dfrac{\check{\boldsymbol{\beta}}_{\perp *}}{\sqrt{T}}\end{pmatrix},$$

where $\check{\boldsymbol{\beta}}_{\perp *} = \check{\boldsymbol{\beta}}_\perp(\check{\boldsymbol{\beta}}_\perp'\check{\boldsymbol{\beta}}_\perp)^{-1}$, and consider the equations

$$\det\left[\mathbf{A}_T'\left(\widehat{\mu}_j\widehat{\mathbf{S}}_{11} - \widehat{\mathbf{S}}_{10}\widehat{\mathbf{S}}_{00}^{-1}\widehat{\mathbf{S}}_{01}\right)\mathbf{A}_T\right] = 0, \quad j = 1,\dots,r, \qquad \text{(A34)}$$
$$\det\left[\mathbf{A}_T'\left(\widehat{\mu}_j^0\mathbf{S}_{11} - \mathbf{S}_{10}\mathbf{S}_{00}^{-1}\mathbf{S}_{01}\right)\mathbf{A}_T\right] = 0, \quad j = 1,\dots,r. \qquad \text{(A35)}$$

Clearly (A34) has the same solutions as (A32), but its eigenvectors are now given by $\mathbf{A}_T^{-1}\widehat{\mathbf{P}}$ and those corresponding to the largest c eigenvalues are $\mathbf{A}_T^{-1}\widehat{\boldsymbol{\beta}}$. Analogously for (A35) we have the eigenvectors $\mathbf{A}_T^{-1}\widehat{\mathbf{P}}^0$ and the c largest are given by $\mathbf{A}_T^{-1}\widehat{\boldsymbol{\beta}}^0$. Moreover, we have

$$
\begin{aligned}
&\mathbf{A}_T'\big(\widehat{\mathbf{S}}_{11}-\widehat{\mathbf{S}}_{10}\widehat{\mathbf{S}}_{00}^{-1}\widehat{\mathbf{S}}_{01}\big)\mathbf{A}_T\\
&=\left[\begin{array}{cc}\check{\boldsymbol{\beta}}'\widehat{\mathbf{S}}_{11}\check{\boldsymbol{\beta}} & T^{-1/2}\check{\boldsymbol{\beta}}'\widehat{\mathbf{S}}_{11}\check{\boldsymbol{\beta}}_{\perp *}\\ T^{-1/2}\check{\boldsymbol{\beta}}_{\perp *}'\widehat{\mathbf{S}}_{11}\check{\boldsymbol{\beta}} & T^{-1}\check{\boldsymbol{\beta}}_{\perp *}'\widehat{\mathbf{S}}_{11}\check{\boldsymbol{\beta}}_{\perp *}\end{array}\right]-\left[\begin{array}{cc}\check{\boldsymbol{\beta}}'\widehat{\mathbf{S}}_{10}\widehat{\mathbf{S}}_{00}^{-1}\widehat{\mathbf{S}}_{01}\check{\boldsymbol{\beta}} & T^{-1/2}\check{\boldsymbol{\beta}}'\widehat{\mathbf{S}}_{10}\widehat{\mathbf{S}}_{00}^{-1}\widehat{\mathbf{S}}_{01}\check{\boldsymbol{\beta}}_{\perp *}\\ T^{-1/2}\check{\boldsymbol{\beta}}_{\perp *}'\widehat{\mathbf{S}}_{10}\widehat{\mathbf{S}}_{00}^{-1}\widehat{\mathbf{S}}_{01}\check{\boldsymbol{\beta}} & T^{-1}\check{\boldsymbol{\beta}}_{\perp *}'\widehat{\mathbf{S}}_{10}\widehat{\mathbf{S}}_{00}^{-1}\widehat{\mathbf{S}}_{01}\check{\boldsymbol{\beta}}_{\perp *}\end{array}\right]\\
&=\left[\begin{array}{cc}\check{\boldsymbol{\beta}}'\mathbf{S}_{11}\check{\boldsymbol{\beta}} & T^{-1/2}\check{\boldsymbol{\beta}}'\mathbf{S}_{11}\check{\boldsymbol{\beta}}_{\perp *}\\ T^{-1/2}\check{\boldsymbol{\beta}}_{\perp *}'\mathbf{S}_{11}\check{\boldsymbol{\beta}} & T^{-1}\check{\boldsymbol{\beta}}_{\perp *}'\mathbf{S}_{11}\check{\boldsymbol{\beta}}_{\perp *}\end{array}\right]-\left[\begin{array}{cc}\check{\boldsymbol{\beta}}'\mathbf{S}_{10}\mathbf{S}_{00}^{-1}\mathbf{S}_{01}\check{\boldsymbol{\beta}} & T^{-1/2}\check{\boldsymbol{\beta}}'\mathbf{S}_{10}\mathbf{S}_{00}^{-1}\mathbf{S}_{01}\check{\boldsymbol{\beta}}_{\perp *}\\ T^{-1/2}\check{\boldsymbol{\beta}}_{\perp *}'\mathbf{S}_{10}\mathbf{S}_{00}^{-1}\mathbf{S}_{01}\check{\boldsymbol{\beta}} & T^{-1}\check{\boldsymbol{\beta}}_{\perp *}'\mathbf{S}_{10}\mathbf{S}_{00}^{-1}\mathbf{S}_{01}\check{\boldsymbol{\beta}}_{\perp *}\end{array}\right]+O_p(\vartheta_{nT,\delta})\\
&=\ \mathbf{A}_T'\big(\mathbf{S}_{11}-\mathbf{S}_{10}\mathbf{S}_{00}^{-1}\mathbf{S}_{01}\big)\mathbf{A}_T+O_p(\vartheta_{nT,\delta}). \qquad \text{(A36)}
\end{aligned}
$$

This result is proved using Lemma 14*ii*, 14*iii*, and 14*vi* for the first block, and 14*i*, 14*iv*, and 14*v* for the second block. Thus, from (A36), for any $j=1,\dots,r$, from Weyl's inequality (A3), we have

$$
\big|\widehat{\mu}_j-\widehat{\mu}_j^0\big|\le\big\|\mathbf{A}_T'\big(\widehat{\mathbf{S}}_{11}-\widehat{\mathbf{S}}_{10}\widehat{\mathbf{S}}_{00}^{-1}\widehat{\mathbf{S}}_{01}\big)\mathbf{A}_T-\mathbf{A}_T'\big(\mathbf{S}_{11}-\mathbf{S}_{10}\mathbf{S}_{00}^{-1}\mathbf{S}_{01}\big)\mathbf{A}_T\big\|=O_p(\vartheta_{nT,\delta}). \qquad \text{(A37)}
$$

Moreover, always from (A36) and similarly to (A19), it can be shown that, by Corollary 1 in Yu et al. (2015), we have (notice that $\widehat{\mu}_j^0$ are all positive since they are eigenvalues of a positive definite matrix)

$$
\big\|\mathbf{A}_T^{-1}\widehat{\mathbf{P}}-\mathbf{A}_T^{-1}\widehat{\mathbf{P}}^0\mathbf{J}_r\big\|=O_p(\vartheta_{nT,\delta}), \qquad \text{(A38)}
$$

where $\mathbf{J}_r$ is a diagonal $r\times r$ matrix with entries 1 or -1.

Now, define the conditional covariance matrices

$$
\begin{aligned}
\check{\boldsymbol{\Omega}}_{00}&=\mathsf{E}[\Delta\check{\mathbf{F}}_t\Delta\check{\mathbf{F}}_t'|\Delta\check{\mathbf{F}}_{t-1}], & \check{\boldsymbol{\Omega}}_{\check{\beta}\check{\beta}}&=\mathsf{E}[\check{\boldsymbol{\beta}}'\check{\mathbf{F}}_{t-1}\check{\mathbf{F}}_{t-1}'\check{\boldsymbol{\beta}}|\Delta\check{\mathbf{F}}_{t-1}],\\
\check{\boldsymbol{\Omega}}_{\check{\beta}0}&=\mathsf{E}[\check{\boldsymbol{\beta}}'\check{\mathbf{F}}_{t-1}\Delta\check{\mathbf{F}}_t'|\Delta\check{\mathbf{F}}_{t-1}], & \check{\boldsymbol{\Omega}}_{0\check{\beta}}&=\mathsf{E}[\Delta\check{\mathbf{F}}_t\check{\mathbf{F}}_{t-1}'\check{\boldsymbol{\beta}}|\Delta\check{\mathbf{F}}_{t-1}]. \qquad \text{(A39)}
\end{aligned}
$$

Then, from Lemmas 9*ii* and 15, (A36), and Slutsky's theorem, as $n,T\to\infty$, we have (see also Lemma 13.1 in Johansen, 1995)

$$
\begin{aligned}
&\det\Big[\mathbf{A}_T'\Big(\widehat{\mu}_j\widehat{\mathbf{S}}_{11}-\widehat{\mathbf{S}}_{10}\widehat{\mathbf{S}}_{00}^{-1}\widehat{\mathbf{S}}_{01}\Big)\mathbf{A}_T\Big]=\det\Big[\mathbf{A}_T'\Big(\widehat{\mu}_j^0\mathbf{S}_{11}-\mathbf{S}_{10}\mathbf{S}_{00}^{-1}\mathbf{S}_{01}\Big)\mathbf{A}_T\Big]+O_p(\vartheta_{nT,\delta}) \qquad \text{(A40)}\\
&\stackrel{d}{\to}\det\Big(\widehat{\mu}_j^0\check{\boldsymbol{\Omega}}_{\check{\beta}\check{\beta}}-\check{\boldsymbol{\Omega}}_{\check{\beta}0}\check{\boldsymbol{\Omega}}_{00}^{-1}\check{\boldsymbol{\Omega}}_{0\check{\beta}}\Big)\ \det\Big[\widehat{\mu}_j^0\check{\boldsymbol{\beta}}_{\perp *}'\Big(\boldsymbol{\Gamma}_{L0}^{\Delta F}\Big)^{1/2}\Big(\int_0^1\mathbf{W}_r(\tau)\mathbf{W}_r'(\tau)\mathrm{d}\tau\Big)\Big(\boldsymbol{\Gamma}_{L0}^{\Delta F}\Big)^{1/2}\check{\boldsymbol{\beta}}_{\perp *}\Big],
\end{aligned}
$$

where $\mathbf{W}_r(\cdot)$ is an r-dimensional random walk with covariance $\mathbf{I}_r$. The first term on the rhs of (A40) has only c solutions different from zero (the matrix is positive definite) while the remaining $r-c$ solutions come from the second term and are all zero. Therefore, as $n,T\to\infty$ both $\mathbf{A}_T^{-1}\widehat{\mathbf{P}}$ and $\mathbf{A}_T^{-1}\widehat{\mathbf{P}}^0$ span a space of dimension c given by their first c eigenvectors. This, jointly with (A38), implies that the two spaces coincide asymptotically

$$
\big\|\mathbf{A}_T^{-1}\widehat{\boldsymbol{\beta}}-\mathbf{A}_T^{-1}\widehat{\boldsymbol{\beta}}^0\mathbf{J}\big\|=O_p(\vartheta_{nT,\delta}). \qquad \text{(A41)}
$$

where $\mathbf{J}$ is a $c \times c$ diagonal matrix with entries 1 or -1.

Now, by projecting $\widehat{\boldsymbol{\beta}}$ onto the space spanned by $(\check{\boldsymbol{\beta}}, \check{\boldsymbol{\beta}}_\perp)$, we can write

$$\widehat{\boldsymbol{\beta}} = \check{\boldsymbol{\beta}}(\check{\boldsymbol{\beta}}'\check{\boldsymbol{\beta}})^{-1}\check{\boldsymbol{\beta}}'\widehat{\boldsymbol{\beta}} + \check{\boldsymbol{\beta}}_\perp(\check{\boldsymbol{\beta}}_\perp'\check{\boldsymbol{\beta}}_\perp)^{-1}\check{\boldsymbol{\beta}}_\perp'\widehat{\boldsymbol{\beta}} = \check{\boldsymbol{\beta}}\check{\boldsymbol{\beta}}_*'\widehat{\boldsymbol{\beta}} + \check{\boldsymbol{\beta}}_{\perp *}\check{\boldsymbol{\beta}}_\perp'\widehat{\boldsymbol{\beta}}$$

where, $\check{\boldsymbol{\beta}}_* = \check{\boldsymbol{\beta}}(\check{\boldsymbol{\beta}}'\check{\boldsymbol{\beta}})^{-1}$ and $\check{\boldsymbol{\beta}}_{\perp *} = \check{\boldsymbol{\beta}}_\perp(\check{\boldsymbol{\beta}}_\perp'\check{\boldsymbol{\beta}}_\perp)^{-1}$. Analogously we have a similar projection for $\widehat{\boldsymbol{\beta}}^0$ and we define the transformed estimators

$$\widetilde{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}(\check{\boldsymbol{\beta}}_*'\widehat{\boldsymbol{\beta}})^{-1} = \check{\boldsymbol{\beta}} + \check{\boldsymbol{\beta}}_{\perp *}\check{\boldsymbol{\beta}}_\perp'\widetilde{\boldsymbol{\beta}}, \qquad \widetilde{\boldsymbol{\beta}}^0 = \widehat{\boldsymbol{\beta}}^0(\check{\boldsymbol{\beta}}_*'\widehat{\boldsymbol{\beta}}^0)^{-1} = \check{\boldsymbol{\beta}} + \check{\boldsymbol{\beta}}_{\perp *}\check{\boldsymbol{\beta}}_\perp'\widetilde{\boldsymbol{\beta}}^0. \tag{A42}$$

From Lemma 13.1 in Johansen (1995), we have (recall that $\check{\boldsymbol{\beta}}_\perp'\check{\boldsymbol{\beta}} = \mathbf{0}_{r-c\times c}$)

$$\mathbf{A}_T^{-1}\widetilde{\boldsymbol{\beta}}^0 = \mathbf{A}_T^{-1}\big(\check{\boldsymbol{\beta}} + \check{\boldsymbol{\beta}}_{\perp *}\check{\boldsymbol{\beta}}_\perp'\widetilde{\boldsymbol{\beta}}^0\big) = \begin{pmatrix} \mathbf{I}_c \\ \sqrt{T}\check{\boldsymbol{\beta}}_\perp'\widetilde{\boldsymbol{\beta}}^0 \end{pmatrix} = \begin{pmatrix} \mathbf{I}_c \\ \sqrt{T}\check{\boldsymbol{\beta}}_\perp'(\widetilde{\boldsymbol{\beta}}^0 - \check{\boldsymbol{\beta}}) \end{pmatrix} = \begin{pmatrix} \mathbf{I}_c \\ o_p(1) \end{pmatrix}, \tag{A43}$$

since $\mathbf{A}_T^{-1}\widetilde{\boldsymbol{\beta}}^0$ spans a space of dimension c. In the same way, we have

$$\mathbf{A}_T^{-1}\widetilde{\boldsymbol{\beta}} = \begin{pmatrix} \mathbf{I}_c \\ \sqrt{T}\check{\boldsymbol{\beta}}_\perp'\widetilde{\boldsymbol{\beta}} \end{pmatrix} = \begin{pmatrix} \mathbf{I}_c \\ \sqrt{T}\check{\boldsymbol{\beta}}_\perp'(\widetilde{\boldsymbol{\beta}} - \check{\boldsymbol{\beta}}) \end{pmatrix} = \begin{pmatrix} \mathbf{I}_c \\ \sqrt{T}\check{\boldsymbol{\beta}}_\perp'(\widetilde{\boldsymbol{\beta}}^0 - \check{\boldsymbol{\beta}}) + \sqrt{T}\check{\boldsymbol{\beta}}_\perp'(\widetilde{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}}_0) \end{pmatrix}. \tag{A44}$$

Now since $\mathrm{sp}(\mathbf{A}_T^{-1}\widetilde{\boldsymbol{\beta}}) = \mathrm{sp}(\mathbf{A}_T^{-1}\widehat{\boldsymbol{\beta}})$, also (A44) spans a space of dimension c. Then by comparing (A43) and (A44), and using (A41), and since also $\mathrm{sp}(\mathbf{A}_T^{-1}\widetilde{\boldsymbol{\beta}}^0) = \mathrm{sp}(\mathbf{A}_T^{-1}\widehat{\boldsymbol{\beta}}^0)$, we have

$$\big\|\sqrt{T}\check{\boldsymbol{\beta}}_\perp'(\widetilde{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}}_0)\big\| = \big\|\mathbf{A}_T^{-1}\widetilde{\boldsymbol{\beta}} - \mathbf{A}_T^{-1}\widetilde{\boldsymbol{\beta}}^0\big\| = O_p(\vartheta_{nT,\delta}). \tag{A45}$$

Therefore, given that $\|\check{\boldsymbol{\beta}}_\perp'\| = O(1)$ and given (A43) and (A45), we have

$$\big\|\widetilde{\boldsymbol{\beta}} - \check{\boldsymbol{\beta}}\big\| \le \big\|\widetilde{\boldsymbol{\beta}}^0 - \check{\boldsymbol{\beta}}\big\| + \big\|\widetilde{\boldsymbol{\beta}}^0 - \widetilde{\boldsymbol{\beta}}\big\| = o_p\left(\frac{1}{\sqrt{T}}\right) + O_p\left(\frac{\vartheta_{nT,\delta}}{\sqrt{T}}\right). \tag{A46}$$

From (A42), we can always define a $c \times c$ orthogonal matrix $\mathbf{Q}$ such that $\widetilde{\boldsymbol{\beta}}\mathbf{Q} = \widehat{\boldsymbol{\beta}}$ (see also pp.179-180 in Johansen, 1995, for a discussion about identification). Therefore, we have

$$\big\|\widehat{\boldsymbol{\beta}} - \check{\boldsymbol{\beta}}\mathbf{Q}\big\| = O_p\left(\frac{\vartheta_{nT,\delta}}{\sqrt{T}}\right),$$

which completes the proof of part *i)*.

Once we have $\widehat{\boldsymbol{\beta}}$, the other parameters are estimated by linear regression

$$\widehat{\boldsymbol{\alpha}} = \widehat{\mathbf{S}}_{01}\widehat{\boldsymbol{\beta}}\big(\widehat{\boldsymbol{\beta}}'\widehat{\mathbf{S}}_{11}\widehat{\boldsymbol{\beta}}\big)^{-1}, \qquad \widehat{\mathbf{G}}_1 = \big(\widehat{\mathbf{M}}_{02} - \widehat{\boldsymbol{\alpha}}\widehat{\boldsymbol{\beta}}'\widehat{\mathbf{M}}_{12}\big)\widehat{\mathbf{M}}_{22}^{-1}. \tag{A47}$$

For part *ii)*, first notice that, by definition from a VECM for $\mathbf{F}_t$ we have

$$\boldsymbol{\alpha} = \mathsf{E}[\Delta\mathbf{F}_t\mathbf{F}_{t-1}'\boldsymbol{\beta}|\Delta\mathbf{F}_{t-1}]\big(\mathsf{E}[\boldsymbol{\beta}'\mathbf{F}_t\mathbf{F}_{t-1}'\boldsymbol{\beta}|\Delta\mathbf{F}_{t-1}]\big)^{-1}$$

Therefore, since conditioning on $\Delta\mathbf{F}_{t-1}$ is equivalent to conditioning on $\mathbf{H}\Delta\mathbf{F}_{t-1} = \Delta\check{\mathbf{F}}_{t-1}$ and $\boldsymbol{\beta}'\mathbf{F}_t =$

$\check{\boldsymbol\beta}'\check{\mathbf F}_t$, from definitions (A39), we immediately have

$$\begin{aligned}\check{\boldsymbol\alpha} = \mathbf H\boldsymbol\alpha =&\mathbf H\mathsf E[\Delta\mathbf F_t\check{\mathbf F}_{t-1}'\check{\boldsymbol\beta}|\Delta\check{\mathbf F}_{t-1}]\big(\mathsf E[\check{\boldsymbol\beta}'\check{\mathbf F}_t\check{\mathbf F}_{t-1}'\check{\boldsymbol\beta}|\Delta\check{\mathbf F}_{t-1}]\big)^{-1}\\ =&\mathsf E[\Delta\check{\mathbf F}_t\check{\mathbf F}_{t-1}'\check{\boldsymbol\beta}|\Delta\check{\mathbf F}_{t-1}]\big(\mathsf E[\check{\boldsymbol\beta}'\check{\mathbf F}_t\check{\mathbf F}_{t-1}'\check{\boldsymbol\beta}|\Delta\check{\mathbf F}_{t-1}]\big)^{-1} = \check{\boldsymbol\Omega}_{0\check\beta}\check{\boldsymbol\Omega}_{\check\beta\check\beta}^{-1}.\end{aligned}$$

Then,

$$\big\|\widehat{\mathbf S}_{01}\widehat{\boldsymbol\beta} - \check{\boldsymbol\Omega}_{0\check\beta}\mathbf Q\big\| \le\big\|\widehat{\mathbf S}_{01}(\widehat{\boldsymbol\beta} - \check{\boldsymbol\beta}\mathbf Q)\big\| + \big\|\widehat{\mathbf S}_{01}\check{\boldsymbol\beta}\mathbf Q - \mathbf S_{01}\check{\boldsymbol\beta}\mathbf Q\big\| + \big\|\mathbf S_{01}\check{\boldsymbol\beta}\mathbf Q - \check{\boldsymbol\Omega}_{0\check\beta}\mathbf Q\big\| = O_p(\vartheta_{nT,\delta}), \quad \text{(A48)}$$

using part *i)* above and the fact that $\|\widehat{\mathbf S}_{01}\| = O_p(T^{1/2})$ in the first term, Lemma 14*iv* for the second term, and Lemma 15*iii* for the third term. Analogously we have

$$\begin{aligned}\big\|\widehat{\boldsymbol\beta}'\widehat{\mathbf S}_{11}\widehat{\boldsymbol\beta} - \mathbf Q'\check{\boldsymbol\Omega}_{\check\beta\check\beta}\mathbf Q\big\| \le&\big\|(\widehat{\boldsymbol\beta}' - \mathbf Q'\check{\boldsymbol\beta}')\widehat{\mathbf S}_{11}(\widehat{\boldsymbol\beta} - \check{\boldsymbol\beta}\mathbf Q)\big\| + \big\|\mathbf Q'\check{\boldsymbol\beta}'\widehat{\mathbf S}_{11}\check{\boldsymbol\beta}\mathbf Q - \mathbf Q'\check{\boldsymbol\beta}'\mathbf S_{11}\check{\boldsymbol\beta}\mathbf Q\big\|\\ &+ \big\|\mathbf Q'\check{\boldsymbol\beta}'\mathbf S_{11}\check{\boldsymbol\beta}\mathbf Q - \mathbf Q'\check{\boldsymbol\Omega}_{\check\beta\check\beta}\mathbf Q\big\| = O_p(\vartheta_{nT,\delta}),\end{aligned} \quad \text{(A49)}$$

using part *i)* above and the fact that $\|\widehat{\mathbf S}_{11}\| = O_p(T)$ in the first term, Lemma 14*ii* for the second term, and Lemma 15*ii* for the third term. Therefore, from (A47), (A48), and (A49), and since $\mathbf Q$ is orthogonal, we have

$$\big\|\widehat{\boldsymbol\alpha} - \check{\boldsymbol\alpha}\mathbf Q\big\| = O_p(\vartheta_{nT,\delta}),$$

which proves part *ii)*.

For part *iii)*, notice that, by definition, we have:

$$\check{\mathbf G}_1 = \mathbf H\mathbf G_1\mathbf H' = \big(\boldsymbol\Gamma_1^{\Delta\check F} - \check{\boldsymbol\alpha}\mathsf E[\check{\boldsymbol\beta}'\check{\mathbf F}_{t-1}\Delta\check{\mathbf F}_{t-1}']\big)(\boldsymbol\Gamma_0^{\Delta\check F})^{-1}. \quad \text{(A50)}$$

Then, from (A47),

$$\begin{aligned}\big\|\widehat{\mathbf G}_1 - \check{\mathbf G}_1\big\| \le&\big\|\big(\widehat{\mathbf M}_{02} - \widehat{\boldsymbol\alpha}\widehat{\boldsymbol\beta}'\widehat{\mathbf M}_{12}\big)\widehat{\mathbf M}_{22}^{-1} - \big(\widehat{\mathbf M}_{02} - \check{\boldsymbol\alpha}\check{\boldsymbol\beta}'\widehat{\mathbf M}_{12}\big)\widehat{\mathbf M}_{22}^{-1}\big\|\\ &+ \big\|\big(\widehat{\mathbf M}_{02} - \check{\boldsymbol\alpha}\check{\boldsymbol\beta}'\widehat{\mathbf M}_{12}\big)\widehat{\mathbf M}_{22}^{-1} - \big(\mathbf M_{02} - \check{\boldsymbol\alpha}\check{\boldsymbol\beta}'\mathbf M_{12}\big)\mathbf M_{22}^{-1}\big\|\\ &+ \big\|\big(\mathbf M_{02} - \check{\boldsymbol\alpha}\check{\boldsymbol\beta}'\mathbf M_{12}\big)\mathbf M_{22}^{-1} - \big(\boldsymbol\Gamma_1^{\Delta\check F} - \check{\boldsymbol\alpha}\mathsf E[\check{\boldsymbol\beta}'\check{\mathbf F}_{t-1}\Delta\check{\mathbf F}_{t-1}']\big)(\boldsymbol\Gamma_0^{\Delta\check F})^{-1}\big\| = O_p(\vartheta_{nT,\delta}),\end{aligned}$$

since the first term on the rhs is $O_p(\vartheta_{nT,\delta})$ by parts *i)* and *ii)* above and since $\check{\boldsymbol\alpha}\mathbf Q\mathbf Q'\check{\boldsymbol\beta}' = \check{\boldsymbol\alpha}\check{\boldsymbol\beta}'$, the second term is $O_p(\vartheta_{nT,\delta})$ by Lemma 13*i*, 13*iv*, and 13*vii*, and the third term is $O_p(T^{-1/2})$ by Lemma 9*i* and 9*vi* and in particular by (B3) and (B12). This, together with (A50), proves part *iii)*.

Finally, for part *iv)*, first notice that the sample covariance of the residuals $\widehat{\mathbf w}_t = \Delta\widehat{\mathbf F}_t - \widehat{\boldsymbol\alpha}\widehat{\boldsymbol\beta}'\widehat{\mathbf F}_{t-1} - \widehat{\mathbf G}_1\Delta\widehat{\mathbf F}_{t-1}$ is written as (see (A30))

$$\begin{aligned}\widehat{\boldsymbol\Gamma}_0^w =&\frac{1}{T}\sum_{t=1}^T \widehat{\mathbf w}_t\widehat{\mathbf w}_t' = \frac{1}{T}\sum_{t=1}^T(\Delta\widehat{\mathbf F}_t - \widehat{\boldsymbol\alpha}\widehat{\boldsymbol\beta}'\widehat{\mathbf F}_{t-1} - \widehat{\mathbf G}_1\Delta\widehat{\mathbf F}_{t-1})(\Delta\widehat{\mathbf F}_t - \widehat{\boldsymbol\alpha}\widehat{\boldsymbol\beta}'\widehat{\mathbf F}_{t-1} - \widehat{\mathbf G}_1\Delta\widehat{\mathbf F}_{t-1})'\\ =&\widehat{\mathbf M}_{00} + \widehat{\boldsymbol\alpha}\widehat{\boldsymbol\beta}'\widehat{\mathbf M}_{11}\widehat{\boldsymbol\beta}\widehat{\boldsymbol\alpha}' + \widehat{\mathbf G}_1\widehat{\mathbf M}_{22}\widehat{\mathbf G}_1' - \widehat{\mathbf M}_{01}\widehat{\boldsymbol\beta}\widehat{\boldsymbol\alpha}' - \widehat{\boldsymbol\alpha}\widehat{\boldsymbol\beta}'\widehat{\mathbf M}_{12}\widehat{\mathbf G}_1' - \widehat{\boldsymbol\alpha}\widehat{\boldsymbol\beta}'\widehat{\mathbf M}_{10} - \widehat{\mathbf G}_1\widehat{\mathbf M}_{20} - \widehat{\mathbf G}_1\widehat{\mathbf M}_{21}\widehat{\boldsymbol\beta}\widehat{\boldsymbol\alpha}'.\end{aligned}$$

Then from parts *i)*, *ii)*, and *iii)* above, Lemma 13*ii*-13*vii*, and Lemma 9*i* and 9*vi*, we immediately prove that

$$\left\|\widehat{\mathbf{\Gamma}}_0^w - \mathbf{H}\mathbf{\Gamma}_0^w\mathbf{H}'\right\| = O_p(\vartheta_{nT,\delta}), \tag{A51}$$

where $\mathbf{\Gamma}_0^w = \mathsf{E}\big[\mathbf{w}_t\mathbf{w}_t'\big] = \mathsf{E}\big[(\Delta\mathbf{F}_t - \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{F}_{t-1} - \mathbf{G}_1\Delta\mathbf{F}_{t-1})(\Delta\mathbf{F}_t - \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{F}_{t-1} - \mathbf{G}_1\Delta\mathbf{F}_{t-1})'\big]$.

Notice that by (16), we have $\mathbf{w}_t = \mathbf{K}\mathbf{u}_t$, therefore, since the shocks $\mathbf{u}_t$ are orthonormal by Assumption $2a$ and $2b$, $\mathbf{\Gamma}_0^w = \mathbf{K}\mathbf{K}'$. Moreover, from Proposition 1 and (11), $\mathbf{K} = \mathbf{C}(0)$, hence by Assumption $2e$, $\mathbf{\Gamma}_0^w$ has rank q and we denote as μ_j^w the eigenvalues, thus $\mu_j^w = 0$ if and only if $j > q$. These are also eigenvalues of $\mathbf{H}\mathbf{\Gamma}_0^w\mathbf{H}'$. As a consequence, having defined as $\widehat{\mu}_j^w$ the eigenvalues of $\widehat{\mathbf{\Gamma}}_0^w$, from (A51) and Weyl's inequality (A3), we have

$$\left|\widehat{\mu}_j^w - \mu_j^w\right| \le \left\|\widehat{\mathbf{\Gamma}}_0^w - \mathbf{H}\mathbf{\Gamma}_0^w\mathbf{H}'\right\| = O_p(\vartheta_{nT,\delta}), \qquad j = 1, \dots, q. \tag{A52}$$

If we denote by $\mathbf{W}_q$ the $r \times q$ matrix of non-zero normalised eigenvectors of $\mathbf{\Gamma}_0^w$, then $\mathbf{H}\mathbf{W}_q$ are the normalised eigenvectors of $\mathbf{H}\mathbf{\Gamma}_0^w\mathbf{H}'$. We denote as $\widehat{\mathbf{W}}_q$ the $r \times q$ matrix of normalised eigenvectors of $\widehat{\mathbf{\Gamma}}_0^w$. Then, from (A51) and similarly to (A19), by Corollary 1 in Yu et al. (2015), we can prove that

$$\left\|\widehat{\mathbf{W}}_q - \mathbf{H}\mathbf{W}_q\mathbf{J}_q\right\| = O_p(\vartheta_{nT,\delta}), \tag{A53}$$

where $\mathbf{J}_q$ is a diagonal $q \times q$ matrix with entries 1 or -1. Notice that $\mathbf{H}\mathbf{W}_q\mathbf{J}_q$ are also normalised eigenvectors of $\mathbf{H}\mathbf{\Gamma}_0^w\mathbf{H}'$. From, the definition of $\widehat{\mathbf{K}} = \widehat{\mathbf{W}}_q\widehat{\mathbf{D}}_q^{-1/2}$ and (A52) and (A53), we have

$$\left\|\widehat{\mathbf{K}} - \mathbf{H}\mathbf{W}_q\mathbf{J}_q\mathbf{D}_q^{-1/2}\right\| = O_p(\vartheta_{nT,\delta}), \tag{A54}$$

where $\mathbf{D}_q$ is a diagonal matrix with entries μ_j^w for $j = 1, \dots, q$ and $\mathbf{W}_q$ contains the corresponding eigenvectors. For any $q \times q$ orthogonal matrix $\mathbf{R}$ such that $\mathbf{K} = \mathbf{W}_q\mathbf{J}_q\mathbf{D}_q^{-1/2}\mathbf{R}$, by substituting in (A54), we have the result. Notice that $\mathbf{K}'\mathbf{\Gamma}_0^w\mathbf{K} = \mathbf{I}_q$ as requested by Assumption $2a$ and $2b$. This completes the proof. □

Proof of Lemma 6

Define the $r \times r$ transformation $\boldsymbol{\mathcal{D}} = (\boldsymbol{\beta}'\ \boldsymbol{\beta}_\perp')'$, where $\boldsymbol{\beta}$ is the $r \times c$ cointegration vector of the true factors $\mathbf{F}_t$, and $\boldsymbol{\beta}_\perp$ is such that $\boldsymbol{\beta}_\perp'\boldsymbol{\beta} = \mathbf{0}_{r-c\times r}$. Then, the vector process $\mathbf{Z}_t = \boldsymbol{\mathcal{D}}\mathbf{F}_t$, is partitioned into an $I(0)$ vector $\mathbf{Z}_{0t} = \boldsymbol{\beta}'\mathbf{F}_t$ and an $I(1)$ vector $\mathbf{Z}_{1t} = \boldsymbol{\beta}_\perp'\mathbf{F}_t$. The vectors $\mathbf{Z}_{0t}$ and $\mathbf{Z}_{1t}$ are orthogonal.

Now consider the models for $\mathbf{F}_t$, $\mathbf{Z}_{0t}$, and $\mathbf{Z}_{1t}$:

$$\mathbf{F}_t = \mathbf{A}_1\mathbf{F}_{t-1} + \mathbf{w}_t, \quad \mathbf{Z}_{0t} = \mathbf{Q}_0\mathbf{F}_{t-1} + \boldsymbol{\beta}'\mathbf{w}_t, \quad \mathbf{Z}_{1t} = \mathbf{Q}_1\mathbf{F}_{t-1} + \boldsymbol{\beta}_\perp'\mathbf{w}_t,$$

where $\mathbf{Q}_0$ is $c \times r$ and $\mathbf{Q}_1$ is $r - c \times r$, and $\mathbf{w}_t = \mathbf{K}\mathbf{u}_t$. Denote the ordinary least squares estimators of the above models when using the true factors and the true cointegration vector as $\widehat{\mathbf{A}}_1^{1\text{VAR}}$, $\widehat{\mathbf{Q}}_0$, and $\widehat{\mathbf{Q}}_1$. Then,

$$\left\|\widehat{\mathbf{Q}}_0 - \mathbf{Q}_0\right\| = \left\|\left(\frac{1}{T}\sum_{t=1}^T \boldsymbol{\beta}'\mathbf{F}_{t-1}\mathbf{u}_t'\mathbf{K}'\boldsymbol{\beta}\right)\left(\frac{1}{T}\sum_{t=1}^T \boldsymbol{\beta}'\mathbf{F}_{t-1}\mathbf{F}_{t-1}'\boldsymbol{\beta}\right)^{-1}\right\| = O_p\left(\frac{1}{\sqrt{T}}\right). \tag{A55}$$

Indeed the first term on the rhs is $O_p(T^{-1/2})$ from (B7) and by independence of $\mathbf{u}_t$ in Assumption $2a$

and b, while the second term is $O_p(1)$ by Lemma 9*iii*. Similarly,

$$\|\widehat{\mathbf{Q}}_1 - \mathbf{Q}_1\| = \left\| \left(\frac{1}{T^2} \sum_{t=1}^{T} \boldsymbol{\beta}_\perp' \mathbf{F}_{t-1} \mathbf{u}_t' \mathbf{K}' \boldsymbol{\beta}_\perp \right) \left(\frac{1}{T^2} \sum_{t=1}^{T} \boldsymbol{\beta}_\perp' \mathbf{F}_{t-1} \mathbf{F}_{t-1}' \boldsymbol{\beta}_\perp \right)^{-1} \right\| = O_p\left(\frac{1}{T}\right). \tag{A56}$$

Indeed the first term on the rhs is $O_p(T^{-1})$ from (B7) and by independence of $\mathbf{u}_t$ in Assumption 2a and b, while the second term is $O_p(1)$ by Lemma 9*ii*. Moreover,

$$\text{vec}(\widehat{\mathbf{A}}_1^{1\text{VAR}}) = (\mathbf{I}_r \otimes \boldsymbol{\mathcal{D}}') \begin{pmatrix} \text{vec}(\widehat{\mathbf{Q}}_0) \\ \text{vec}(\widehat{\mathbf{Q}}_1) \end{pmatrix}. \tag{A57}$$

Analogous formulas to (A55)-(A57) are in Theorem 1 by Sims et al. (1990) and, by combining them,

$$\|\widehat{\mathbf{A}}_1^{1\text{VAR}} - \mathbf{A}_1\| = O_p\left(\frac{1}{\sqrt{T}}\right). \tag{A58}$$

Notice that of the r^2 parameters in $\mathbf{A}_1$, cr in $\mathbf{Q}_0$ are estimated consistently with rate $O_p(T^{-1/2})$, while $(r-c)r$ in $\mathbf{Q}_1$ with rate $O_p(T^{-1})$.

If we now denote as $\widehat{\mathbf{A}}_1^{0\text{VAR}}$ the ordinary least square estimator for the VAR when using $\mathbf{H}\mathbf{F}_t$, then $\widehat{\mathbf{A}}_1^{0\text{VAR}} = \mathbf{H}\widehat{\mathbf{A}}_1^{1\text{VAR}}\mathbf{H}'$, and from (A58)

$$\|\widehat{\mathbf{A}}_1^{0\text{VAR}} - \mathbf{H}\mathbf{A}_1\mathbf{H}'\| = O_p\left(\frac{1}{\sqrt{T}}\right). \tag{A59}$$

Analogously to (A30), we define

$$\widehat{\mathbf{M}}_{1L} = \frac{1}{T}\sum_{t=1}^{T} \widehat{\mathbf{F}}_t \widehat{\mathbf{F}}_{t-1}', \qquad \widehat{\mathbf{M}}_{LL} = \frac{1}{T}\sum_{t=1}^{T} \widehat{\mathbf{F}}_{t-1} \widehat{\mathbf{F}}_{t-1}'. \tag{A60}$$

Then, we can write the VAR estimators as

$$\widehat{\mathbf{A}}_1^{\text{VAR}} = \frac{\widehat{\mathbf{M}}_{1L}}{T}\left(\frac{\widehat{\mathbf{M}}_{LL}}{T}\right)^{-1}, \quad \widehat{\mathbf{A}}_1^{0\text{VAR}} = \frac{\mathbf{M}_{1L}}{T}\left(\frac{\mathbf{M}_{LL}}{T}\right)^{-1}, \tag{A61}$$

where $\mathbf{M}_{1L}$ and $\mathbf{M}_{LL}$ are defined as in (A60) but when using $\mathbf{H}\mathbf{F}_t$.

Because of Lemma 13*i*, we have

$$\left\|\frac{\widehat{\mathbf{M}}_{1L}}{T} - \frac{\mathbf{M}_{1L}}{T}\right\| = O_p\left(\max\left(\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{T}}\right)\right), \qquad \left\|\frac{\widehat{\mathbf{M}}_{LL}}{T} - \frac{\mathbf{M}_{LL}}{T}\right\| = O_p\left(\max\left(\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{T}}\right)\right),$$

thus

$$\|\widehat{\mathbf{A}}_1^{\text{VAR}} - \widehat{\mathbf{A}}_1^{0\text{VAR}}\| = O_p\left(\max\left(\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{T}}\right)\right). \tag{A62}$$

By combining (A62) with (A59)

$$\|\widehat{\mathbf{A}}_1^{\text{VAR}} - \mathbf{H}\mathbf{A}_1\mathbf{H}'\| \le \|\widehat{\mathbf{A}}_1^{\text{VAR}} - \widehat{\mathbf{A}}_1^{0\text{VAR}}\| + \|\widehat{\mathbf{A}}_1^{0\text{VAR}} - \mathbf{H}\mathbf{A}_1\mathbf{H}'\| = O_p\left(\max\left(\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{T}}\right)\right), \tag{A63}$$

which completes the proof of part *ii)*.

By noticing that, from part *i)*, (A51) holds also in this case, but with the rate given in (A63), we prove part *ii)* exactly as in Lemma 5*iv*. This completes the proof. □

Proof of Lemma 7

For part *i)* we can follow a reasoning similar to Lemma 2*i*. The spectral density matrix of the first difference of the common factors can be written as $\mathbf{\Sigma}^{\Delta F}(\theta) = (2\pi)^{-1}\mathbf{C}(e^{-i\theta})\overline{\mathbf{C}'(e^{-i\theta})}$ and, since $\text{rk}(\mathbf{C}(e^{-i\theta})) = q$ a.e. in $[-\pi,\pi]$, then it has q non-zero real dynamic eigenvalues and $r-q$ zero dynamic eigenvalues. Notice also that we have $\text{rk}(\mathbf{C}(e^{-i\theta})) \le q$ for any $\theta \in [-\pi,\pi]$. Moreover, given Assumption $2d$ of summability of coefficients, the non-zero dynamic eigenvalues are also finite for any $\theta \in [-\pi,\pi]$. Thus, by denoting as $\mu_j^{\Delta F}(\theta)$ those eigenvalues, we have, a.e. in $[-\pi,\pi]$,

$$0 < \underline{M}_9 \le \mu_j^{\Delta F}(\theta) \le \overline{M}_9 < \infty, \quad j = 1,\dots,q. \tag{A64}$$

Therefore, we can write $\mathbf{\Sigma}^{\Delta F}(\theta) = \mathbf{W}^{\Delta F}(\theta)\mathbf{M}^{\Delta F}(\theta)\overline{\mathbf{W}^{\Delta F'}(\theta)}$, where $\mathbf{W}^{\Delta F}(\theta)$ is the $r \times q$ matrix of normalized dynamic eigenvectors, i.e. such that $\overline{\mathbf{W}^{\Delta F'}(\theta)}\mathbf{W}^{\Delta F}(\theta) = \mathbf{I}_q$ for any $\theta \in [-\pi,\pi]$, and $\mathbf{M}^{\Delta F}(\theta)$ is the corresponding $q \times q$ diagonal matrix of dynamic eigenvalues.

Define $\boldsymbol{L}(\theta) = \mathbf{\Lambda}\mathbf{W}^{\Delta F}(\theta)(\mathbf{M}^{\Delta F}(\theta))^{1/2}$. Then the spectral density matrix of the first differences of the common component is given by

$$\frac{\mathbf{\Sigma}^{\Delta\chi}(\theta)}{n} = \frac{1}{n}\mathbf{\Lambda}\mathbf{\Sigma}^{\Delta F}(\theta)\mathbf{\Lambda}' = \frac{1}{n}\mathbf{\Lambda}\mathbf{W}^{\Delta F}(\theta)\mathbf{M}^{\Delta F}(\theta)\overline{\mathbf{W}^{\Delta F'}(\theta)}\mathbf{\Lambda}' = \frac{\boldsymbol{L}(\theta)\overline{\boldsymbol{L}'(\theta)}}{n}, \quad \theta \in [-\pi,\pi].$$

Moreover, when $n^{-1}\mathbf{\Lambda}'\mathbf{\Lambda} = \mathbf{I}_r$

$$\frac{\overline{\boldsymbol{L}'(\theta)}\boldsymbol{L}(\theta)}{n} = \mathbf{M}^{\Delta F}(\theta), \quad \theta \in [-\pi,\pi]. \tag{A65}$$

Therefore, a.e. in $[-\pi,\pi]$ the non-zero dynamic eigenvalues of $\mathbf{\Sigma}^{\Delta\chi}(\theta)$ are the same as those of $\overline{\boldsymbol{L}'(\theta)}\boldsymbol{L}(\theta)$, and from (A65), we have for any n and a.e. in $[-\pi,\pi]$, $n^{-1}\mu_j^{\Delta\chi}(\theta) = \mu_j^{\Delta F}(\theta)$, for any $j = 1,\dots,r$. Part *i)* then follows from (A64).

As for part *ii)*, from Assumption $4c$, for any $\theta \in [-\pi,\pi]$, there exists a finite positive constant K_3 such that

$$\sup_{i\in\mathbb{N}} \left|\check{d}_i(e^{-i\theta})\right| \le \sup_{i\in\mathbb{N}} \left|\sum_{k=0}^{\infty} \check{d}_{ik}e^{-ik\theta}\right| \le \sup_{i\in\mathbb{N}} \sum_{k=0}^{\infty} \left|\check{d}_{ik}\right| \le K_3 < \infty. \tag{A66}$$

Define as $\sigma_{ij}(\theta)$ the generic (i,j)-th entry of $\mathbf{\Sigma}^{\Delta\xi}(\theta)$. Then, for any $n \in \mathbb{N}$,

$$\begin{aligned}\sup_{\theta\in[-\pi,\pi]} \left\|\mathbf{\Sigma}^{\Delta\xi}(\theta)\right\|_1 &= \sup_{\theta\in[-\pi,\pi]} \max_{i=1,\dots,n} \sum_{j=1}^{n} |\sigma_{ij}(\theta)| = \sup_{\theta\in[-\pi,\pi]} \max_{i=1,\dots,n} \frac{1}{2\pi} \sum_{j=1}^{n} \left|\check{d}_i(e^{-i\theta})\mathsf{E}[\varepsilon_{it}\varepsilon_{jt}]\,\check{d}_j(e^{i\theta})\right| \\ &\le \frac{K_3^2}{2\pi} \max_{i=1,\dots,n} \sum_{j=1}^{n} |\mathsf{E}[\varepsilon_{it}\varepsilon_{jt}]| \le \frac{K_3^2 M_4}{2\pi} < \infty, \end{aligned} \tag{A67}$$

where we used (A66) and Assumption $4f$. From (A2) and (A67), we have, for any $n \in \mathbb{N}$,

$$\sup_{\theta\in[-\pi,\pi]} \mu_1^{\Delta\xi}(\theta) = \sup_{\theta\in[-\pi,\pi]} \left\|\mathbf{\Sigma}^{\Delta\xi}(\theta)\right\| \le \sup_{\theta\in[-\pi,\pi]} \left\|\mathbf{\Sigma}^{\Delta\xi}(\theta)\right\|_1 \le \frac{K_3^2 M_4}{2\pi} < \infty, \tag{A68}$$

and part *ii)* is proved by defining $M_{10} = K_3^2 M_4 (2\pi)^{-1}$.

Finally, parts *iii)* and *iv)*, are immediate consequences of Assumption 5 which implies that $\mathbf{\Sigma}^{\Delta x}(\theta) = \mathbf{\Sigma}^{\Delta \chi}(\theta) + \mathbf{\Sigma}^{\Delta \xi}(\theta)$, for any $\theta \in [-\pi, \pi]$, and of Weyl's inequality (A3). So, for $j = 1, \dots, q$, and for any $n \in \mathbb{N}$ and a.e. in $[-\pi, \pi]$, we have

$$\frac{\mu_j^{\Delta x}(\theta)}{n} \le \frac{\mu_j^{\Delta \chi}(\theta)}{n} + \frac{\mu_1^{\Delta \xi}(\theta)}{n} \le \overline{M}_9 + \sup_{\theta \in [-\pi,\pi]} \frac{\mu_1^{\Delta \xi}(\theta)}{n} \le \overline{M}_9 + \frac{M_{10}}{n} = \overline{M}_{11} < \infty,$$

$$\frac{\mu_j^{\Delta x}(\theta)}{n} \ge \frac{\mu_j^{\Delta \chi}(\theta)}{n} + \frac{\mu_n^{\Delta \xi}(\theta)}{n} \ge \underline{M}_9 + \inf_{\theta \in [-\pi,\pi]} \frac{\mu_n^{\Delta \xi}(\theta)}{n} = \underline{M}_{11} > 0.$$

because of parts *i)* and *ii)*. This proves part *iii)*. When $j = q + 1$, using parts *i)* and *ii)* above, and since $\mathrm{rk}(\mathbf{\Sigma}^{\Delta \chi}(\theta)) \le q$, for any $\theta \in [-\pi, \pi]$, we have $\mu_{q+1}^{\Delta x}(\theta) \le \mu_{q+1}^{\Delta \chi}(\theta) + \mu_1^{\Delta \xi(\theta)} = \mu_1^{\Delta \xi(\theta)} \le M_{10} < \infty$, thus proving part *iv)*.

Finally, to prove part *v)*, consider parts *iii)* and *iv)* when $\theta = 0$ and use again parts *i)* and *ii)*, and the fact that $\mathrm{rk}(\mathbf{\Sigma}^{\Delta \chi}(0)) = \tau \le q$, hence $0 < \underline{M}_9 \le n^{-1} \mu_\tau^{\Delta \chi}(0) \le \overline{M}_9 < \infty$ but $\mu_{\tau+1}^{\Delta \chi}(0) = 0$. This completes the proof. □

Appendix B Complementary results

Lemma 8 *Define the covariance matrix* $\mathbf{\Gamma}_0^{\Delta x} = \mathsf{E}[\Delta \mathbf{x}_t \Delta \mathbf{x}_t']$ *with generic* (i,j)*-th element* $\gamma_{ij}^{\Delta x} = \mathsf{E}[\Delta x_{it} \Delta x_{jt}]$*. Then, under Assumptions 1-5, as* $T \to \infty$*,* $|T^{-1} \sum_{t=1}^T \Delta x_{it} \Delta x_{it} - \gamma_{ij}^{\Delta x}| = O_p(T^{-1/2})$*, for any* $i,j = 1,\dots,n$*.*

Proof of Lemma 8

First notice that $\gamma_{ij}^{\Delta x} = \boldsymbol{\lambda}_i' \mathbf{\Gamma}_0^{\Delta F} \boldsymbol{\lambda}_j + \gamma_{ij}^{\Delta \xi}$, where $\boldsymbol{\lambda}_i'$ is the i-th row of $\mathbf{\Lambda}$, $\mathbf{\Gamma}_0^{\Delta F} = \mathsf{E}[\Delta \mathbf{F}_t \Delta \mathbf{F}_t']$, and $\gamma_{ij}^{\Delta \xi} = \mathsf{E}[\Delta \xi_{it} \Delta \xi_{jt}]$. Then, we also have

$$\mathsf{E}\left[\frac{1}{T}\sum_{t=1}^T \Delta \mathbf{F}_t \Delta \mathbf{F}_t'\right] = \frac{1}{T}\sum_{t=1}^T \mathsf{E}\left[\left(\sum_{k=0}^\infty \mathbf{C}_k \mathbf{u}_{t-k}\right)\left(\sum_{k'=0}^\infty \mathbf{C}_{k'} \mathbf{u}_{t-k'}\right)'\right] = \sum_{k=0}^\infty \mathbf{C}_k \mathbf{C}_k' = \mathbf{\Gamma}_0^{\Delta F}, \tag{B1}$$

where we used Assumption $2a$ and b of independence of $\mathbf{u}_t$. Moreover, $\mathrm{rk}(\mathbf{\Gamma}_0^{\Delta F}) = r$ because of Assumption $2g$, and $\|\mathbf{\Gamma}_0^{\Delta F}\| = O(1)$ because of square summability of the coefficients given in Assumption $2d$. Hence, $\mathbf{\Gamma}_0^{\Delta F}$ is well defined. For the idiosyncratic component we trivially have $\mathsf{E}[T^{-1}\sum_{t=1}^T \Delta \xi_{it} \Delta \xi_{jt}] = \gamma_{ij}^{\Delta \xi}$, therefore by Assumption 5, $\mathsf{E}[T^{-1}\sum_{t=1}^T \Delta x_{it} \Delta x_{it}] = \gamma_{ij}^{\Delta x}$.

Now, denote as $\gamma_{ij}^{\Delta F}$ the generic (i,j)-th element of $\mathbf{\Gamma}_0^{\Delta F}$. Then, from (A2),

$$\begin{aligned}
\mathsf{E}\left[\left\|\frac{1}{T}\sum_{t=1}^T \Delta \mathbf{F}_t \Delta \mathbf{F}_t' - \mathbf{\Gamma}_0^{\Delta F}\right\|^2\right] &\le \sum_{i,j=1}^r \frac{1}{T^2}\mathsf{E}\left[\sum_{t,s=1}^T \left(\Delta F_{it}\Delta F_{jt} - \gamma_{ij}^{\Delta F}\right)\left(\Delta F_{is}\Delta F_{js} - \gamma_{ij}^{\Delta F}\right)\right] \\
&= \sum_{i,j=1}^r \frac{1}{T^2}\sum_{t,s=1}^T \left(\mathsf{E}\left[\Delta F_{it}\Delta F_{jt}\Delta F_{is}\Delta F_{js}\right] - (\gamma_{ij}^{\Delta F})^2\right) \\
&\le \frac{r^2 K_1^4 q^4}{T^2}\sum_{t,s=1}^T \mathsf{E}[u_{lt}u_{l't}u_{hs}u_{h's}] - \frac{r^2 K_1^4 q^4}{T^2}\sum_{t,s=1}^T (\mathsf{E}[u_{lt}u_{l't}])^2 \\
&= \frac{r^2 K_1^4 q^4}{T^2}\sum_{t,s=1}^T \mathsf{E}[u_{lt}^2]\mathsf{E}[u_{hs}^2] + \frac{r^2 K_1^4 q^4}{T^2}\sum_{t=1}^T \mathsf{E}[u_{lt}^2 u_{ht}^2] - \frac{r^2 K_1^4 q^4}{T^2}\sum_{t,s=1}^T (\mathsf{E}[u_{lt}^2])^2 \\
&= \frac{r^2 K_1^4 q^4}{T^2}\sum_{t=1}^T \mathsf{E}[u_{lt}^2]\mathsf{E}[u_{ht}^2] = \frac{r^2 K_1^4 q^4}{T} = O\left(\frac{1}{T}\right),
\end{aligned} \tag{B2}$$

where we used Assumption $2a$ and b of independence of $\mathbf{u}_t$ and Assumption $2d$ of square summability of the coefficients, with K_1 defined in (A4). Therefore, from (B2), we have

$$\left\|\frac{1}{T}\sum_{t=1}^T \Delta \mathbf{F}_t \Delta \mathbf{F}_t' - \mathbf{\Gamma}_0^{\Delta F}\right\| = O_p\left(\frac{1}{\sqrt{T}}\right). \tag{B3}$$

In the same way, for the idiosyncratic component we have

$$\begin{aligned}
\mathsf{E}\left[\left\|\frac{1}{T}\sum_{t=1}^T \Delta \xi_{it}\Delta \xi_{jt} - \gamma_{ij}^{\Delta \xi}\right\|^2\right] &\le \frac{1}{T^2}\sum_{t,s=1}^T \left(\mathsf{E}\left[\Delta \xi_{it}\Delta \xi_{jt}\Delta \xi_{is}\Delta \xi_{js}\right] - (\gamma_{ij}^{\Delta \xi})^2\right) \\
&\le \frac{K_2^4}{T^2}\sum_{t=1}^T \mathsf{E}[\varepsilon_{it}^2 \varepsilon_{jt}^2] = \frac{K_2^4}{T} = O\left(\frac{1}{T}\right),
\end{aligned} \tag{B4}$$

where we used Assumption $4d$ and e of independence of $\boldsymbol{\varepsilon}_t$ and existence of fourth cross-sectional moments and Assumption $4c$ of square summability of the coefficients, with K_2 defined in (A4). Therefore, from (B2), we have

$$\left\|\frac{1}{T}\sum_{t=1}^T \Delta\xi_{it}\Delta\xi_{jt} - \gamma_{ij}^{\Delta\xi}\right\| = O_p\left(\frac{1}{\sqrt{T}}\right). \tag{B5}$$

By combining (B3) and (B5) and using Assumption $3b$ of bounded loadings we complete the proof.□

Lemma 9 *Define the autocovariance matrices* $\boldsymbol{\Gamma}_k^{\Delta F} = \mathsf{E}[\Delta\mathbf{F}_t\Delta\mathbf{F}_{t-k}']$, *with* $k \in \mathbb{Z}$, *and the long-run autocovariance matrices* $\boldsymbol{\Gamma}_{L0}^{\Delta F} = \boldsymbol{\Gamma}_0^{\Delta F} + 2\sum_{h=1}^\infty \boldsymbol{\Gamma}_h^{\Delta F}$ *and* $\boldsymbol{\Gamma}_{L1}^{\Delta F} = \sum_{h=1}^\infty \boldsymbol{\Gamma}_h^{\Delta F}$. *Denote as* $\mathbf{W}_r(\cdot)$ *an* r*-dimensional random walk with covariance* $\mathbf{I}_r$ *and analogously define* $\mathbf{W}_q(\cdot)$. *From* (12) *define also* $\boldsymbol{\omega}_t = \check{\mathbf{C}}(L)\mathbf{u}_t$, *with autocovariances* $\boldsymbol{\Gamma}_h^{\omega}$ *and long-run covariance* $\boldsymbol{\Gamma}_{L0}^{\omega} = \boldsymbol{\Gamma}_0^{\omega} + 2\sum_{h=1}^\infty \boldsymbol{\Gamma}_h^{\omega}$. *If Assumption 2 holds then, as* $T \to \infty$,

i. $\|T^{-1}\sum_{t=k+1}^T \Delta\mathbf{F}_t\Delta\mathbf{F}_{t-k}' - \boldsymbol{\Gamma}_k^{\Delta F}\| = O_p(T^{-1/2})$;
ii. $\|T^{-1}\sum_{t=1}^T \boldsymbol{\beta}'\mathbf{F}_t\mathbf{F}_t'\boldsymbol{\beta} - \boldsymbol{\beta}'\boldsymbol{\Gamma}_0^{\omega}\boldsymbol{\beta}\| = \|T^{-1}\sum_{t=1}^T \boldsymbol{\beta}'\mathbf{F}_t\mathbf{F}_t'\boldsymbol{\beta} - \mathsf{E}[\boldsymbol{\beta}'\mathbf{F}_t\mathbf{F}_t'\boldsymbol{\beta}]\| = O_p(T^{-1/2})$;
iii. $\|T^{-1}\sum_{t=1}^T \Delta\mathbf{F}_t\mathbf{F}_{t-1}'\boldsymbol{\beta} - (\boldsymbol{\Gamma}_1^{\omega} - \boldsymbol{\Gamma}_0^{\omega})\boldsymbol{\beta}\| = \|T^{-1}\sum_{t=1}^T \Delta\mathbf{F}_t\mathbf{F}_{t-1}'\boldsymbol{\beta} - \mathsf{E}[\Delta\mathbf{F}_t\mathbf{F}_{t-1}'\boldsymbol{\beta}]\| = O_p(T^{-1/2})$;
iv. $T^{-2}\sum_{t=1}^T \mathbf{F}_t\mathbf{F}_t' \xrightarrow{d} (\boldsymbol{\Gamma}_{L0}^{\Delta F})^{1/2}\big(\int_0^1 \mathbf{W}_r(\tau)\mathbf{W}_r'(\tau)\mathrm{d}\tau\big)(\boldsymbol{\Gamma}_{L0}^{\Delta F})^{1/2}$;
v. $T^{-1}\sum_{t=1}^T \mathbf{F}_{t-1}\Delta\mathbf{F}_t' \xrightarrow{d} (\boldsymbol{\Gamma}_{L0}^{\Delta F})^{1/2}\big(\int_0^1 \mathbf{W}_r(\tau)\mathrm{d}\mathbf{W}_r'(\tau)\big)(\boldsymbol{\Gamma}_{L0}^{\Delta F})^{1/2} + \boldsymbol{\Gamma}_{L1}^{\Delta F}$;
vi. $T^{-1}\sum_{t=1}^T \mathbf{F}_t\mathbf{F}_t'\boldsymbol{\beta} \xrightarrow{d} \mathbf{C}(1)\big(\int_0^1 \mathbf{W}_q(\tau)\mathrm{d}\mathbf{W}_r'(\tau)\big)(\boldsymbol{\Gamma}_{L0}^{\omega})^{1/2}\boldsymbol{\beta} + \boldsymbol{\Gamma}_0^{\omega}\boldsymbol{\beta}$.

Proof of Lemma 9

For part *i)*, the case $k = 0$ is proved in (B3) in Lemma 8. The proof for the autocovariances, i.e. when $k \neq 0$, is analogous.

For parts *iv)* and *v)*, first notice that, by Assumption $2g$,

$$\boldsymbol{\Gamma}_{L0}^{\Delta F} = \sum_{k=0}^\infty \mathbf{C}_k\mathbf{C}_k' + \sum_{h=1}^\infty\sum_{k=h}^\infty \Big(\mathbf{C}_k\mathbf{C}_{k+h}' + \mathbf{C}_{k+h}\mathbf{C}_k'\Big), \tag{B6}$$

which is positive definite, and by Assumption $2d$ this matrix is also finite. Moreover, by Assumption $2a$ and b the vector $\mathbf{u}_t$ satisfies the assumptions of Corollary 2.2 in Phillips and Durlauf (1986), then parts *iv)* and *v)* are direct consequences of Lemma 3.1 in Phillips and Durlauf (1986).

As for parts *ii)*, *vi)*, and *iii)*, we use the Beverdige Nelson decomposition for the common factors (see e.g. Lemma 2.1 in Phillips and Solo, 1992, and (12))

$$\Delta\mathbf{F}_t = \mathbf{C}(1)\mathbf{u}_t + \check{\mathbf{C}}(L)(\mathbf{u}_t - \mathbf{u}_{t-1}),$$

where $\check{\mathbf{C}}(L) = \sum_{k=0}^\infty \check{\mathbf{C}}_k L^k$ with $\check{\mathbf{C}}_k = -\sum_{h=k+1}^\infty \mathbf{C}_h$. Then,

$$\mathbf{F}_t = \mathbf{C}(1)\sum_{s=1}^t \mathbf{u}_s + \boldsymbol{\omega}_t, \tag{B7}$$

where $\boldsymbol{\omega}_t = \check{\mathbf{C}}(L)(\mathbf{u}_t - \mathbf{u}_0) = \check{\mathbf{C}}(L)\mathbf{u}_t$, since $\mathbf{u}_t = \mathbf{0}$ when $t \le 0$ by Assumption $2c$, and $\boldsymbol{\omega}_t \sim I(0)$, because from Assumption $2d$ the coefficients of $\check{\mathbf{C}}(L)$ are square summable. Moreover, from Assumption $2f$ and (12), we have $\mathbf{C}(1) = \boldsymbol{\psi}\boldsymbol{\eta}'$, where $\boldsymbol{\psi}$ is $r \times r - c$ and $\boldsymbol{\eta}$ is $q \times r - c$. Since $\boldsymbol{\beta}$ is a cointegrating vector for $\mathbf{F}_t$, we have $\boldsymbol{\beta} = \boldsymbol{\psi}_\perp$ and therefore $\boldsymbol{\beta}'\mathbf{C}(1) = \mathbf{0}_{c\times q}$. So, $\boldsymbol{\beta}'\mathbf{F}_t = \boldsymbol{\beta}'\boldsymbol{\omega}_t$. Then,

$$\frac{1}{T}\sum_{t=1}^T \mathbf{F}_t\mathbf{F}_t'\boldsymbol{\beta} = \mathbf{C}(1)\left[\frac{1}{T}\sum_{t=1}^T\left(\sum_{s=1}^t \mathbf{u}_s\right)\boldsymbol{\omega}_t'\right]\boldsymbol{\beta} + \left[\frac{1}{T}\sum_{t=1}^T \boldsymbol{\omega}_t\boldsymbol{\omega}_t'\right]\boldsymbol{\beta}. \tag{B8}$$

Define $t = \lfloor T\tau \rfloor$ for $\tau \in [0,1]$ and the functionals

$$\mathbf{X}_{u,T}(\tau) = \frac{1}{\sqrt{T}}\sum_{s=1}^{\lfloor T\tau\rfloor}\mathbf{u}_s, \qquad \mathbf{X}_{\omega,T}(\tau) = \frac{1}{\sqrt{T}}\left(\boldsymbol{\Gamma}_{L0}^{\omega}\right)^{-1/2}\sum_{s=1}^{\lfloor T\tau\rfloor}\boldsymbol{\omega}_s,$$

where as for (B6) we can show that $\boldsymbol{\Gamma}_{L0}^{\omega} = \boldsymbol{\Gamma}_0^{\omega} + 2\sum_{h=1}^{\infty}\boldsymbol{\Gamma}_h^{\omega}$ is positive definite due to Assumption 2*d*. Moreover, we can write $\boldsymbol{\omega}_t = \sqrt{T}(\boldsymbol{\Gamma}_{L0}^{\omega})^{1/2}[\mathbf{X}_{\omega,T}(t/T) - \mathbf{X}_{\omega,T}((t-1)/T)]$. As proved in Theorem 3.4 in Phillips and Solo (1992) and Corollary 2.2 in Phillips and Durlauf (1986), for any $\tau \in [0,1]$, we have, as $T \to \infty$,

$$\mathbf{X}_{u,T}(\tau) \stackrel{d}{\to} \mathbf{W}_q(\tau), \qquad \mathbf{X}_{\omega,T}(\tau) \stackrel{d}{\to} \mathbf{W}_r(\tau), \tag{B9}$$

where $\mathbf{W}_q(\cdot)$ is a q-dimensional random walk with covariance $\mathbf{I}_q$ and $\mathbf{W}_r(\cdot)$ is an r-dimensional random walk with variance $\mathbf{I}_r$. Then consider the first term in brackets on the rhs of (B8), as $T \to \infty$, using (B9), we have

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^T\left(\sum_{s=1}^t \mathbf{u}_s\right)\boldsymbol{\omega}_t' &= \sum_{t=1}^T \mathbf{X}_u\left(\frac{t}{T}\right)\left(\mathbf{X}_\omega\left(\frac{t}{T}\right) - \mathbf{X}_\omega\left(\frac{t-1}{T}\right)\right)'\left(\boldsymbol{\Gamma}_{L0}^{\omega}\right)^{1/2} \\ &\stackrel{d}{\to} \left(\int_0^1 \mathbf{W}_u(\tau)\frac{(\mathbf{W}_\omega(\tau) - \mathbf{W}_\omega(\tau - \mathrm{d}\tau))'}{\mathrm{d}\tau}\mathrm{d}\tau\right)\left(\boldsymbol{\Gamma}_{L0}^{\omega}\right)^{1/2} = \left(\int_0^1 \mathbf{W}_u(\tau)\mathrm{d}\mathbf{W}_\omega'(\tau)\right)\left(\boldsymbol{\Gamma}_{L0}^{\omega}\right)^{1/2}.\end{aligned} \tag{B10}$$

As for the second term on the rhs of (B8), we have, using the same approach as for part *i)*, as $T \to \infty$,

$$\left\|\frac{1}{T}\sum_{t=1}^T \boldsymbol{\omega}_t\boldsymbol{\omega}_t' - \boldsymbol{\Gamma}_0^{\omega}\right\| = O_p\left(\frac{1}{\sqrt{T}}\right). \tag{B11}$$

By substituting (B10) and (B11) in (B8), and by Slutsky's theorem, we complete the proof of part *vi)*. Part *ii)* is proved analogously just by multiplying (B8) on the left by $\boldsymbol{\beta}'$.

Finally, for part *iii)*, using the same approach as in the proof of part *i)*, we have

$$\frac{1}{T}\sum_{t=1}^T \Delta\mathbf{F}_t\mathbf{F}_{t-1}'\boldsymbol{\beta} = \left(\frac{1}{T}\sum_{t=1}^T \mathbf{C}(1)\mathbf{u}_t\boldsymbol{\omega}_{t-1}' + \frac{1}{T}\sum_{t=1}^T \Delta\boldsymbol{\omega}_t\boldsymbol{\omega}_{t-1}'\right) = \left(\boldsymbol{\Gamma}_1^{\omega} - \boldsymbol{\Gamma}_0^{\omega}\right)\boldsymbol{\beta} + O_p\left(\frac{1}{\sqrt{T}}\right). \tag{B12}$$

This completes the proof. □

Lemma 10 *For any $t = 1,\ldots,T$, and as $n,T \to \infty$, if Assumptions 1-4 hold, then,*

i. $\|\Delta\mathbf{F}_t\| = O_p(1)$;
ii. $\|T^{-1/2}\mathbf{F}_t\| = O_p(1)$;
iii. $\|\boldsymbol{\beta}'\mathbf{F}_t\| = O_p(1)$;
iv. $\|n^{-1/2}\Delta\boldsymbol{\xi}_t\| = O_p(1)$;
v. $\|(nT)^{-1/2}\boldsymbol{\xi}_t\| = O_p(1)$;
vi. $\|n^{-1/2}\boldsymbol{\Lambda}'\Delta\boldsymbol{\xi}_t\| = O_p(1)$;

vii. $\|(nT)^{-1/2}\boldsymbol{\Lambda}'\boldsymbol{\xi}_t\| = O_p(1)$.

If also Assumptions 6b-6c hold, then,

viii. $\|n^{-1/2}\boldsymbol{\xi}_t\| = O_p(T^{1/2}n^{-(1-\delta)/2})$;

ix. $\|n^{-1/2}\boldsymbol{\Lambda}'\boldsymbol{\xi}_t\| = O_p(T^{1/2}n^{-(1-\delta)/2})$.

Proof of Lemma 10 For part *i)*, just notice that $\Delta\mathbf{F}_t$ has finite variance, indeed,

$$\begin{aligned}\mathsf{E}\big[\|\Delta\mathbf{F}_t\|^2\big] &= \sum_{j=1}^r \mathsf{E}\big[\Delta F_{jt}^2\big] = \sum_{j=1}^r \mathsf{E}[(\mathbf{c}_j'(L)\mathbf{u}_t)^2] = \sum_{j=1}^r \mathsf{E}\bigg[\bigg(\sum_{l=1}^q c_{jl}(L)u_{lt}\bigg)^2\bigg]\\ &=\sum_{j=1}^r\sum_{l,l'=1}^q\sum_{k,k'=0}^\infty c_{jlk}c_{jl'k'}\mathsf{E}[u_{lt-k}u_{l't-k'}] \le rqK_1 < \infty, \end{aligned} \tag{B13}$$

where we used Assumption $2a$ and b of independence of $\mathbf{u}_t$ and Assumption $2d$ which implies square summability of the coefficients, with K_1 defined in (A4). This proves part *i)*.

Similarly, for part *ii)*, we have

$$\begin{aligned}\mathsf{E}\bigg[\bigg\|\frac{\mathbf{F}_t}{\sqrt{T}}\bigg\|^2\bigg] &= \frac{1}{T}\sum_{j=1}^r \mathsf{E}\big[F_{jt}^2\big] = \frac{1}{T}\sum_{j=1}^r \mathsf{E}\bigg[\bigg(\sum_{s=1}^t\sum_{l=1}^q c_{jl}(L)u_{ls}\bigg)^2\bigg]\\ &=\frac{1}{T}\sum_{j=1}^r\sum_{s,s'=1}^t\sum_{l,l'=1}^q\sum_{k,k'=0}^\infty c_{jlk}c_{jl'k'}\mathsf{E}[u_{ls-k}u_{l's'-k'}] \le \frac{rqK_1t}{T} \le rqK_1 < \infty, \end{aligned} \tag{B14}$$

where we used the same assumptions as in (B13). This proves part *ii)*. For part *iii)*, recall from (B7) that $\boldsymbol{\beta}'\mathbf{F}_t = \check{\mathbf{C}}(L)\mathbf{u}_t$, which is stationary. Since, the coefficients of $\check{\mathbf{C}}(L)$ are also square summable by Assumption $2d$, part *iii)* is proved as part *i)* using the analogous of (B13).

For part *iv)*, for any $n \in \mathbb{N}$, we have,

$$\begin{aligned}\mathsf{E}\bigg[\bigg\|\frac{\Delta\boldsymbol{\xi}_t}{\sqrt{n}}\bigg\|^2\bigg] &= \frac{1}{n}\sum_{i=1}^n \mathsf{E}\big[\Delta\xi_{it}^2\big] = \frac{1}{n}\sum_{i=1}^n \mathsf{E}[(\check{d}_i(L)\varepsilon_{it})^2]\\ &=\frac{1}{n}\sum_{i=1}^n\sum_{k,k'=0}^\infty \check{d}_{jk}\check{d}_{ik'}\mathsf{E}[\varepsilon_{it-k}\varepsilon_{it-k'}] \le K_2\max_i\sigma_i^2 < \infty, \end{aligned} \tag{B15}$$

where we used Assumption $4d$ and e of serial independence of $\boldsymbol{\varepsilon}_t$ and Assumption $4c$ which implies square summability of the coefficients, with K_2 defined in (A4). This proves part *iv)*.

Similarly, for part *v)*, for any $n \in \mathbb{N}$, we have,

$$\begin{aligned}\mathsf{E}\bigg[\bigg\|\frac{\boldsymbol{\xi}_t}{\sqrt{nT}}\bigg\|^2\bigg] &= \frac{1}{nT}\sum_{i=1}^n \mathsf{E}\big[\xi_{it}^2\big] = \frac{1}{nT}\sum_{i=1}^n \mathsf{E}\bigg[\bigg(\sum_{s=1}^t \check{d}_i(L)\varepsilon_{is}\bigg)^2\bigg]\\ &=\frac{1}{nT}\sum_{i=1}^n\sum_{s,s'=1}^t\sum_{k,k'=0}^\infty \check{d}_{ik}\check{d}_{ik'}\mathsf{E}[\varepsilon_{is-k}\varepsilon_{is'-k'}] \le \frac{K_2 t}{T}\max_i\sigma_i^2 \le K_2\max_i\sigma_i^2 < \infty, \end{aligned} \tag{B16}$$

where we used the same assumptions as in (B15). This proves part *v)*.

As for part *vi)*, for any $n \in \mathbb{N}$, we have

$$\begin{aligned}
\mathsf{E}\left[\left\|\frac{\mathbf{\Lambda}'\Delta\boldsymbol{\xi}_t}{\sqrt{n}}\right\|^2\right] &= \frac{1}{n}\sum_{j=1}^r \mathsf{E}\left[\left(\sum_{i=1}^n \lambda_{ij}\Delta\xi_{it}\right)^2\right] = \frac{1}{n}\sum_{j=1}^r\sum_{i,l=1}^n \mathsf{E}\big[\lambda_{ij}\Delta\xi_{it}\lambda_{lj}\Delta\xi_{lt}\big]\\
&\le \frac{rM_1^2}{n}\sum_{i,l=1}^n\sum_{k,k'=0}^\infty \check{d}_{ik}\check{d}_{lk'}\mathsf{E}[\varepsilon_{it-k}\varepsilon_{lt-k'}] \le \frac{rM_1^2K_2}{n}\sum_{i,l=1}^n \big|\mathsf{E}[\varepsilon_{it}\varepsilon_{lt}]\big| \le rM_1^2K_2M_4 < \infty, \qquad \text{(B17)}
\end{aligned}$$

where we used the same assumptions as in (B15), Assumption 3b of bounded loadings, and Lemma 1 of mild cross-correlation among idiosyncratic shocks. This proves part *vi)*.

Similarly for part *vii)*, for any $n \in \mathbb{N}$, we have

$$\begin{aligned}
\mathsf{E}\left[\left\|\frac{\mathbf{\Lambda}'\boldsymbol{\xi}_t}{\sqrt{nT}}\right\|^2\right] &= \frac{1}{nT}\sum_{j=1}^r \mathsf{E}\left[\left(\sum_{i=1}^n \lambda_{ij}\xi_{it}\right)^2\right] = \frac{1}{nT}\sum_{j=1}^r\sum_{i,l=1}^n \mathsf{E}\big[\lambda_{ij}\xi_{it}\lambda_{lj}\xi_{lt}\big]\\
&\le \frac{rM_1^2}{nT}\sum_{i,l=1}^n\sum_{s,s'=1}^t\sum_{k,k'=0}^\infty \check{d}_{ik}\check{d}_{lk'}\mathsf{E}[\varepsilon_{is-k}\varepsilon_{ls'-k'}] \le \frac{rM_1^2K_2t}{nT}\sum_{i,l=1}^n \big|\mathsf{E}[\varepsilon_{it}\varepsilon_{lt}]\big| \le rM_1^2K_2M_4 < \infty, \qquad \text{(B18)}
\end{aligned}$$

where we used the same assumptions as in (B17). This proves part *vii)*.

Now consider part *viii)*. Using Assumption 4a, for any $n \in \mathbb{N}$, we can write

$$\mathsf{E}\left[\left\|\frac{\boldsymbol{\xi}_t}{\sqrt{n}}\right\|^2\right] = \frac{1}{n}\sum_{i=1}^m \mathsf{E}\big[\xi_{it}^2\big] + \frac{1}{n}\sum_{i=m+1}^n \mathsf{E}\big[\xi_{it}^2\big]. \qquad \text{(B19)}$$

The second term on the rhs is bounded for any $n \in \mathbb{N}$ because it is a sum of stationary components and we can use the same reasoning as for part *iv)*. For the first term on the rhs, using Assumption 6 and part *v)*, we have (multiply and divide by m)

$$\frac{1}{n}\sum_{i=1}^m \mathsf{E}\big[\xi_{it}^2\big] \le \frac{K_2Tm}{n}\max_i \sigma_i^2 = O\left(\frac{T}{n^{1-\delta}}\right), \qquad \text{(B20)}$$

which proves part *viii)*.

Finally, for part *ix)*, using the same reasoning as for part *viii)*, we can write

$$\begin{aligned}
\mathsf{E}\left[\left\|\frac{\mathbf{\Lambda}'\boldsymbol{\xi}_t}{\sqrt{n}}\right\|^2\right] &= \frac{1}{n}\sum_{j=1}^r\sum_{i,l=1}^n \mathsf{E}\big[\lambda_{ij}\xi_{it}\lambda_{lj}\xi_{lt}\big]\\
&= \frac{1}{n}\sum_{j=1}^r\sum_{i,l=1}^m \mathsf{E}\big[\lambda_{ij}\xi_{it}\lambda_{lj}\xi_{lt}\big] + \frac{1}{n}\sum_{j=1}^r\sum_{i,l=m+1}^n \mathsf{E}\big[\lambda_{ij}\xi_{it}\lambda_{lj}\xi_{lt}\big] + \frac{2}{n}\sum_{j=1}^r\sum_{i=1}^m\sum_{l=m+1}^n \mathsf{E}\big[\lambda_{ij}\xi_{it}\lambda_{lj}\xi_{lt}\big]. \qquad \text{(B21)}
\end{aligned}$$

The second term on the rhs is bounded because it is a sum of products of stationary components and behaves as part *vi)* above. For the first term on the rhs, using Assumption 6 and part *v)*, we have (multiply and divide by m)

$$\frac{1}{n}\sum_{j=1}^r\sum_{i,l=1}^m \mathsf{E}\big[\lambda_{ij}\xi_{it}\lambda_{lj}\xi_{lt}\big] \le \frac{rM_1^2K_2T}{n}\sum_{i,l=1}^m \big|\mathsf{E}[\varepsilon_{it}\varepsilon_{lt}]\big| \le \frac{rM_1^2K_2M_4Tm}{n} = O\left(\frac{T}{n^{1-\delta}}\right). \qquad \text{(B22)}$$

Finally, the third term on the rhs of (B21) is

$$\frac{1}{n}\sum_{j=1}^{r}\sum_{i=1}^{m}\sum_{l=m+1}^{n}\mathsf{E}\big[\lambda_{ij}\xi_{it}\lambda_{lj}\xi_{lt}\big] \le \frac{rM_1^2K_2T}{n}\sum_{i=1}^{m}\sum_{l=m+1}^{n}\big|\mathsf{E}[\varepsilon_{it}\varepsilon_{lt}]\big| \le \frac{rM_1^2K_2M_8Tn^{\gamma}}{n} = O\left(\frac{T}{n^{1-\gamma}}\right). \tag{B23}$$

We prove part *ix)* by substituting (B22) and (B23) into (B21), and by noticing that (B22) converges to zero slower than (B23) because $\gamma < \delta$ by Assumption 6c. This completes the proof. □

Lemma 11 *Define* $\check{\mathbf{F}}_t = \mathbf{H}\mathbf{F}_t$ *and* $\check{\boldsymbol{\beta}} = \mathbf{H}\boldsymbol{\beta}$*. For any* $t = 1, \dots, T$*, and as* $n, T \to \infty$*, if Assumptions 1-4 hold, then,*

i. $\|(Tn)^{-1}\widehat{\boldsymbol{\Lambda}}'\boldsymbol{\xi}_t\check{\mathbf{F}}_t'\| = O_p(\max(n^{-1/2}, T^{-1/2}))$;
ii. $\|n^{-1}\widehat{\boldsymbol{\Lambda}}'\Delta\boldsymbol{\xi}_t\Delta\check{\mathbf{F}}_t'\| = O_p(\max(n^{-1/2}, T^{-1/2}))$;
iii. $\|n^{-1}\widehat{\boldsymbol{\Lambda}}'\Delta\boldsymbol{\xi}_t\check{\mathbf{F}}_t'\check{\boldsymbol{\beta}}\| = O_p(\max(n^{-1/2}, T^{-1/2}))$;
iv. $\|(T^{1/2}n)^{-1}\widehat{\boldsymbol{\Lambda}}'\Delta\boldsymbol{\xi}_t\check{\mathbf{F}}_t'\| = O_p(\max(n^{-1/2}, T^{-1/2}))$;
v. $\|(T^{1/2}n)^{-1}\widehat{\boldsymbol{\Lambda}}'\boldsymbol{\xi}_t\check{\mathbf{F}}_t'\check{\boldsymbol{\beta}}\| = O_p(\max(n^{-1/2}, T^{-1/2}))$.

If also Assumptions 6b-6c hold, then,

vi. $\|n^{-1}\widehat{\boldsymbol{\Lambda}}'\boldsymbol{\xi}_t\Delta\check{\mathbf{F}}_t'\| = O_p(\zeta_{nT,\delta})$;
vii. $\|(T^{1/2}n)^{-1}\widehat{\boldsymbol{\Lambda}}'\boldsymbol{\xi}_t\check{\mathbf{F}}_t'\| = O_p(\zeta_{nT,\delta})$;
viii. $\|n^{-1}\widehat{\boldsymbol{\Lambda}}'\boldsymbol{\xi}_t\check{\mathbf{F}}_t'\check{\boldsymbol{\beta}}\| = O_p(\zeta_{nT,\delta})$.

Proof of Lemma 11 Throughout, we use $\|\mathbf{H}\| = O(1)$ and $\|\boldsymbol{\beta}\| = O(1)$, and subadditivity of the norm (A1). Start with part *i)*:

$$\begin{aligned}\left\|\frac{\widehat{\boldsymbol{\Lambda}}'\boldsymbol{\xi}_t\check{\mathbf{F}}_t'}{nT}\right\| &\le \left\|\frac{\mathbf{H}'\boldsymbol{\Lambda}'\boldsymbol{\xi}_t\mathbf{F}_t'\mathbf{H}'}{nT}\right\| + \left\|\frac{(\widehat{\boldsymbol{\Lambda}}' - \mathbf{H}'\boldsymbol{\Lambda}')\boldsymbol{\xi}_t\mathbf{F}_t'\mathbf{H}'}{nT}\right\| \\ &\le \|\mathbf{H}\|^2\left\|\frac{\boldsymbol{\Lambda}'\boldsymbol{\xi}_t}{n\sqrt{T}}\right\|\left\|\frac{\mathbf{F}_t}{\sqrt{T}}\right\| + \left\|\frac{\widehat{\boldsymbol{\Lambda}}' - \mathbf{H}'\boldsymbol{\Lambda}'}{\sqrt{n}}\right\|\left\|\frac{\boldsymbol{\xi}_t}{\sqrt{nT}}\right\|\left\|\frac{\mathbf{F}_t}{\sqrt{T}}\right\|\|\mathbf{H}\|.\end{aligned}$$

Then, because of Lemma 10*ii* and 10*vii*, the first term on the rhs is $O_p(n^{-1/2})$. Because of Lemma 3 and Lemma 10*ii* and 10*v*, the second term on the rhs is $O_p(T^{-1/2})$. We here use the fact that by Assumption 6a the loadings converge at rate $\sqrt{T}$, this is always assumed in what follows since the other possible convergence rate n is always dominated by other rates. This proves part *i)*. Similarly, for part *ii)* repeat the same reasoning using Lemma 10*i*, 10*iv*, and 10*vi* and Lemma 3. Part *iii)* is proved by noticing that $\check{\mathbf{F}}_t'\check{\boldsymbol{\beta}} = \mathbf{F}_t'\boldsymbol{\beta}$, and by following again the same reasoning as for part *i)*, and using Lemma 10*iii*, 10*iv*, and 10*vi*, and Lemma 3. Part *iv)* is also proved as part *i)*, and using Lemma 10*ii*, 10*iv*, and 10*vi*, and Lemma 3. Part *v)* is proved as part *i)*, and using Lemma 10*iii*, 10*v*, and 10*vii*, and Lemma 3.

For part *vi)*, we have

$$\begin{aligned}\left\|\frac{\widehat{\boldsymbol{\Lambda}}'\boldsymbol{\xi}_t\Delta\check{\mathbf{F}}_t'}{n}\right\| &\le \left\|\frac{\mathbf{H}'\boldsymbol{\Lambda}'\boldsymbol{\xi}_t\Delta\mathbf{F}_t'\mathbf{H}'}{n}\right\| + \left\|\frac{(\widehat{\boldsymbol{\Lambda}}' - \mathbf{H}'\boldsymbol{\Lambda}')\boldsymbol{\xi}_t\Delta\mathbf{F}_t'\mathbf{H}'}{n}\right\| \\ &\le \|\mathbf{H}\|^2\left\|\frac{\boldsymbol{\Lambda}'\boldsymbol{\xi}_t}{n}\right\|\|\Delta\mathbf{F}_t\| + \left\|\frac{\widehat{\boldsymbol{\Lambda}}' - \mathbf{H}'\boldsymbol{\Lambda}'}{\sqrt{n}}\right\|\left\|\frac{\boldsymbol{\xi}_t}{\sqrt{n}}\right\|\|\Delta\mathbf{F}_t\|\|\mathbf{H}\|.\end{aligned}$$

From Lemma 10i and 10ix, the first term on the rhs is $O_p(T^{1/2}n^{-(2-\delta)/2})$. From Lemma 10$i$ and 10$viii$ and Lemma 3 the second term on the rhs is $O_p(n^{-1(1-\delta)/2})$. This proves part *vi)*. Parts *vii)* and *viii)* are proved similarly to part *vi)* using Lemma 10ii, 10iii, 10$viii$, and 10ix and Lemma 3. This completes the proof. □

Lemma 12 *For any $t = 1, \dots, T$, and as $n, T \to \infty$, if Assumptions 1-4 hold, then,*

i. $\|(Tn^2)^{-1}\widehat{\mathbf{\Lambda}}'\boldsymbol{\xi}_t\boldsymbol{\xi}_t'\widehat{\mathbf{\Lambda}}\| = O_p(\max(n^{-1}, T^{-1}))$;

ii. $\|n^{-2}\widehat{\mathbf{\Lambda}}'\Delta\boldsymbol{\xi}_t\Delta\boldsymbol{\xi}_t'\widehat{\mathbf{\Lambda}}\| = O_p(\max(n^{-1}, T^{-1}))$.

If also Assumptions 6b-6c hold, then,

iii. $\|n^{-2}\widehat{\mathbf{\Lambda}}'\boldsymbol{\xi}_t\boldsymbol{\xi}_t'\widehat{\mathbf{\Lambda}}\| = O_p(\zeta^2_{nT,\delta})$;

iv. $\|(T^{1/2}n^2)^{-1}\widehat{\mathbf{\Lambda}}'\boldsymbol{\xi}_t\boldsymbol{\xi}_t'\widehat{\mathbf{\Lambda}}\| = O_p(\zeta^2_{nT,\delta}T^{-1/2})$;

v. $\|n^{-2}\widehat{\mathbf{\Lambda}}'\Delta\boldsymbol{\xi}_t\boldsymbol{\xi}_t'\widehat{\mathbf{\Lambda}}\| = O_p(\zeta_{n,T}\max(n^{-1/2}, T^{-1/2}))$.

Proof of Lemma 12 Throughout, we use subadditivity of the norm (A1). Start with part *i)*:

$$\left\|\frac{\widehat{\mathbf{\Lambda}}'\boldsymbol{\xi}_t\boldsymbol{\xi}_t'\widehat{\mathbf{\Lambda}}}{n^2T}\right\| \le \left\|\frac{\widehat{\mathbf{\Lambda}}-\mathbf{\Lambda}\mathbf{H}'}{\sqrt{n}}\right\|^2\left\|\frac{\boldsymbol{\xi}_t}{\sqrt{nT}}\right\|^2 + 2\left\|\frac{\widehat{\mathbf{\Lambda}}-\mathbf{\Lambda}\mathbf{H}'}{\sqrt{n}}\right\|\left\|\frac{\boldsymbol{\xi}_t}{\sqrt{nT}}\right\|\left\|\frac{\mathbf{\Lambda}'\boldsymbol{\xi}_t}{n\sqrt{T}}\right\| + \left\|\frac{\mathbf{\Lambda}'\boldsymbol{\xi}_t}{n\sqrt{T}}\right\|^2.$$

The first term on the rhs is $O_p(T^{-1})$ because of Lemma 10v and Lemma 3. The second term is $O_p(T^{-1/2}n^{-1/2})$ because of Lemma 10v and 10vii, and Lemma 3. The third term is $O_p(n^{-1})$ because of Lemma 10vii. This proves part *i)* and part *ii)* is proved in the same way by using Lemma 10iv and 10vi, and Lemma 3.

Now consider part *iii)*:

$$\left\|\frac{\widehat{\mathbf{\Lambda}}'\boldsymbol{\xi}_t\boldsymbol{\xi}_t'\widehat{\mathbf{\Lambda}}}{n^2}\right\| \le \left\|\frac{\widehat{\mathbf{\Lambda}}-\mathbf{\Lambda}\mathbf{H}'}{\sqrt{n}}\right\|^2\left\|\frac{\boldsymbol{\xi}_t}{\sqrt{n}}\right\|^2 + 2\left\|\frac{\widehat{\mathbf{\Lambda}}-\mathbf{\Lambda}\mathbf{H}'}{\sqrt{n}}\right\|\left\|\frac{\boldsymbol{\xi}_t}{\sqrt{n}}\right\|\left\|\frac{\mathbf{\Lambda}'\boldsymbol{\xi}_t}{n}\right\| + \left\|\frac{\mathbf{\Lambda}'\boldsymbol{\xi}_t}{n}\right\|^2.$$

The first term on the rhs is $O_p(n^{-(1-\delta)})$ because of Lemma 10$viii$ and Lemma 3. The third term is $O_p(Tn^{-(2-\delta)})$ because of Lemma 10ix. Using Lemma 10$viii$ and 10ix, and Lemma 3, the second term is $O_p(T^{1/2}n^{-(3/2-\delta)})$. Summing up, we have

$$\left\|\frac{\widehat{\mathbf{\Lambda}}'\boldsymbol{\xi}_t\boldsymbol{\xi}_t'\widehat{\mathbf{\Lambda}}}{n^2}\right\| \le O_p\left(\frac{1}{n^{1-\delta}}\right) + O_p\left(\frac{\sqrt{T}}{n^{3/2-\delta}}\right) + O_p\left(\frac{T}{n^{2-\delta}}\right).$$

In order to compare the rates of the three terms assume $n = O(T^\alpha)$, then, according to Assumption 6, we must have at least $\alpha > 1/2$. Now, when $1/2 < \alpha < 1$, the third term dominates over the first one (see (A5)) but the second would dominate over the third if and only if $\alpha > 1$ which cannot be. When, $\alpha \ge 1$ the first term dominates over the third one, and the second would dominate over the first if and only if $\alpha < 1$ which cannot be. Hence the second one is always dominated by the other two and we proved part *iii)*. Part *iv)* is proved by dividing everything in part *iii)* by $T^{1/2}$.

For part *v)*, we have

$$\left\|\frac{\widehat{\mathbf{\Lambda}}'\Delta\boldsymbol{\xi}_t\boldsymbol{\xi}_t'\widehat{\mathbf{\Lambda}}}{n^2}\right\| \le \left\|\frac{\widehat{\mathbf{\Lambda}}-\mathbf{\Lambda}\mathbf{H}'}{\sqrt{n}}\right\|^2\left\|\frac{\Delta\boldsymbol{\xi}_t}{\sqrt{n}}\right\|\left\|\frac{\boldsymbol{\xi}_t}{\sqrt{n}}\right\| + \left\|\frac{\mathbf{\Lambda}'\Delta\boldsymbol{\xi}_t}{n}\right\|\left\|\frac{\mathbf{\Lambda}'\boldsymbol{\xi}_t}{n}\right\|$$
$$+\left\|\frac{\widehat{\mathbf{\Lambda}}-\mathbf{\Lambda}\mathbf{H}'}{\sqrt{n}}\right\|\left\|\frac{\Delta\boldsymbol{\xi}_t}{\sqrt{n}}\right\|\left\|\frac{\mathbf{\Lambda}'\boldsymbol{\xi}_t}{n}\right\| + \left\|\frac{\widehat{\mathbf{\Lambda}}-\mathbf{\Lambda}\mathbf{H}'}{\sqrt{n}}\right\|\left\|\frac{\boldsymbol{\xi}_t}{\sqrt{n}}\right\|\left\|\frac{\mathbf{\Lambda}'\Delta\boldsymbol{\xi}_t}{n}\right\|.$$

The first term on the rhs is $O_p(T^{-1/2}n^{-(1-\delta)/2})$ because of Lemma 10*iv* and 10*viii*, and Lemma 3. The second term is $O_p(T^{1/2}n^{-(3-\delta)/2})$ because of Lemma 10*vi* and 10*ix*, and Lemma 3. Hence, using (A5), the first two terms are $O_p(\zeta_{n,T}\max(n^{-1/2}, T^{-1/2}))$. Using the same results as for the first two terms we have that the third and fourth terms are both $O_p(n^{-(2-\delta)/2})$ and they are both dominated by the first two and we proved part *v)*. This completes the proof. □

Lemma 13 *Consider the matrices* $\widehat{\mathbf{M}}_{ij}$ *defined in* (A30) *and denote by* $\mathbf{M}_{ij}$*, for* $i,j = 0,1,2$*, the analogous ones when computed using* $\check{\mathbf{F}}_t = \mathbf{H}\mathbf{F}_t$*. Define also* $\check{\boldsymbol{\beta}} = \mathbf{H}\boldsymbol{\beta}$*. As* $n,T \to \infty$*, if Assumptions 1-4 hold, then,*

i. $\|T^{-1}\widehat{\mathbf{M}}_{11} - T^{-1}\mathbf{M}_{11}\| = O_p(n^{-1/2}, T^{-1/2})$;
ii. $\|\widehat{\mathbf{M}}_{00} - \mathbf{M}_{00}\| = O_p(n^{-1/2}, T^{-1/2})$;
iii. $\|\widehat{\mathbf{M}}_{02} - \mathbf{M}_{02}\| = O_p(n^{-1/2}, T^{-1/2})$;
iv. $\|\widehat{\mathbf{M}}_{22} - \mathbf{M}_{22}\| = O_p(n^{-1/2}, T^{-1/2})$.

If also Assumptions 6b-6c hold, then,

v. $\|\widehat{\mathbf{M}}_{01}\check{\boldsymbol{\beta}} - \mathbf{M}_{01}\check{\boldsymbol{\beta}}\| = O_p(\max(\zeta_{nT,\delta}, T^{-1/2}))$;
vi. $\|\check{\boldsymbol{\beta}}'\widehat{\mathbf{M}}_{11}\check{\boldsymbol{\beta}} - \check{\boldsymbol{\beta}}'\mathbf{M}_{11}\check{\boldsymbol{\beta}}\| = O_p(\max(\zeta_{nT,\delta}, T^{-1/2}))$;
vii. $\|\widehat{\mathbf{M}}_{21}\check{\boldsymbol{\beta}} - \mathbf{M}_{21}\check{\boldsymbol{\beta}}\| = O_p(\max(\zeta_{nT,\delta}, T^{-1/2}))$;
viii. $\|T^{-1/2}\widehat{\mathbf{M}}_{01} - T^{-1/2}\mathbf{M}_{01}\| = O_p(\max(\zeta_{nT,\delta}, T^{-1/2}))$;
ix. $\|T^{-1/2}\widehat{\mathbf{M}}_{21} - T^{-1/2}\mathbf{M}_{21}\| = O_p(\max(\zeta_{nT,\delta}, T^{-1/2}))$.

Proof of Lemma 13 Throughout, we use $\|\mathbf{H}\| = O(1)$ and $\|\boldsymbol{\beta}\| = O(1)$, and the fact that, from Lemma 3, $\|n^{-1}\widehat{\boldsymbol{\Lambda}}'\boldsymbol{\Lambda}\| = O_p(1)$. Start with part *i)*. By adding and subtracting $\mathbf{H}\mathbf{F}_t$ from $\widehat{\mathbf{F}}_t$, we have

$$\left\|\frac{1}{T^2}\sum_{t=1}^T \widehat{\mathbf{F}}_t\widehat{\mathbf{F}}_t' - \frac{1}{T^2}\sum_{t=1}^T \check{\mathbf{F}}_t\check{\mathbf{F}}_t'\right\| \le \left\|\frac{1}{T^2}\sum_{t=1}^T\left(\widehat{\mathbf{F}}_t - \mathbf{H}\mathbf{F}_t\right)\left(\widehat{\mathbf{F}}_t - \mathbf{H}\mathbf{F}_t\right)'\right\| + 2\left\|\frac{1}{T^2}\sum_{t=1}^T\left(\widehat{\mathbf{F}}_t - \mathbf{H}\mathbf{F}_t\right)\left(\mathbf{H}\mathbf{F}_t\right)'\right\|. \tag{B24}$$

Using (15) and (1), the first term on the rhs of (B24) is such that

$$\begin{aligned}
&\left\|\frac{1}{T^2}\sum_{t=1}^T\left(\widehat{\mathbf{F}}_t - \mathbf{H}\mathbf{F}_t\right)\left(\widehat{\mathbf{F}}_t - \mathbf{H}\mathbf{F}_t\right)'\right\| = \left\|\frac{1}{T^2}\sum_{t=1}^T\left(\frac{\widehat{\boldsymbol{\Lambda}}'\mathbf{x}_t}{n} - \mathbf{H}\mathbf{F}_t\right)\left(\frac{\widehat{\boldsymbol{\Lambda}}'\mathbf{x}_t}{n} - \mathbf{H}\mathbf{F}_t\right)'\right\| \\
&= \left\|\frac{1}{T^2}\sum_{t=1}^T\left(\frac{\widehat{\boldsymbol{\Lambda}}'\boldsymbol{\Lambda}\mathbf{F}_t}{n} + \frac{\widehat{\boldsymbol{\Lambda}}'\boldsymbol{\xi}_t}{n} - \mathbf{H}\mathbf{F}_t\right)\left(\frac{\widehat{\boldsymbol{\Lambda}}'\boldsymbol{\Lambda}\mathbf{F}_t}{n} + \frac{\widehat{\boldsymbol{\Lambda}}'\boldsymbol{\xi}_t}{n} - \mathbf{H}\mathbf{F}_t\right)'\right\| \\
&\le \underbrace{\left\|\frac{1}{T^2}\sum_{t=1}^T\frac{\widehat{\boldsymbol{\Lambda}}'\boldsymbol{\Lambda}\mathbf{F}_t\mathbf{F}_t'}{n}\left(\frac{\boldsymbol{\Lambda}'\widehat{\boldsymbol{\Lambda}}}{n} - \mathbf{H}'\right) + \mathbf{H}\mathbf{F}_t\mathbf{F}_t'\left(\mathbf{H}' - \frac{\boldsymbol{\Lambda}'\widehat{\boldsymbol{\Lambda}}}{n}\right)\right\|}_{\mathcal{A}_1} + 2\underbrace{\left\|\frac{1}{T^2}\sum_{t=1}^T\frac{\widehat{\boldsymbol{\Lambda}}'\boldsymbol{\Lambda}\mathbf{F}_t\boldsymbol{\xi}_t'\widehat{\boldsymbol{\Lambda}}}{n^2}\right\|}_{\mathcal{B}_1} \\
&\quad + 2\underbrace{\left\|\frac{1}{T^2}\sum_{t=1}^T\frac{\widehat{\boldsymbol{\Lambda}}'\boldsymbol{\xi}_t\mathbf{F}_t'\mathbf{H}'}{n}\right\|}_{\mathcal{C}_1} + \underbrace{\left\|\frac{1}{T^2}\sum_{t=1}^T\frac{\widehat{\boldsymbol{\Lambda}}'\boldsymbol{\xi}_t\boldsymbol{\xi}_t'\widehat{\boldsymbol{\Lambda}}}{n^2}\right\|}_{\mathcal{D}_1}.
\end{aligned} \tag{B25}$$

Let us consider each term of (B25) separately:

$$\mathcal{A}_1 \leq \left\| \frac{\mathbf{\Lambda}'\widehat{\mathbf{\Lambda}}}{n} - \mathbf{H}' \right\| \left\| \frac{1}{T^2}\sum_{t=1}^T \mathbf{F}_t\mathbf{F}_t' \right\| \left\{ \left\| \frac{\widehat{\mathbf{\Lambda}}'\mathbf{\Lambda}}{n} \right\| + \|\mathbf{H}\| \right\} = O_p\left(\frac{1}{\sqrt{T}}\right),$$

$$\mathcal{B}_1 \leq \frac{2}{T}\sum_{t=1}^T \left\| \frac{\widehat{\mathbf{\Lambda}}'\boldsymbol{\xi}_t\mathbf{F}_t'}{nT} \right\| \left\| \frac{\widehat{\mathbf{\Lambda}}'\mathbf{\Lambda}}{n} \right\| = O_p\left(\max\left(\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{T}}\right)\right),$$

$$\mathcal{C}_1 \leq \frac{2}{T}\sum_{t=1}^T \left\| \frac{\widehat{\mathbf{\Lambda}}'\boldsymbol{\xi}_t\mathbf{F}_t'}{nT} \right\| \|\mathbf{H}\| = O_p\left(\max\left(\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{T}}\right)\right),$$

$$\mathcal{D}_1 \leq \frac{1}{T}\sum_{t=1}^T \left\| \frac{\widehat{\mathbf{\Lambda}}'\boldsymbol{\xi}_t\boldsymbol{\xi}_t'\widehat{\mathbf{\Lambda}}}{n^2T} \right\| = O_p\left(\max\left(\frac{1}{n}, \frac{1}{T}\right)\right).$$

Above we used, for $\mathcal{A}_1$ Lemmas 3 and 9*ii*, for $\mathcal{B}_1$ and $\mathcal{C}_1$ Lemma 11*i*, for $\mathcal{D}_1$ Lemma 12*i*. Thus, the first term on the rhs of (B24) is $O_p(\max(n^{-1/2}, T^{-1/2}))$. The second term on the rhs of (B24) is such that

$$\begin{aligned}
&\left\| \frac{1}{T^2}\sum_{t=1}^T \left(\widehat{\mathbf{F}}_t - \mathbf{H}\mathbf{F}_t\right)\left(\mathbf{H}\mathbf{F}_t\right)' \right\| = \left\| \frac{1}{T^2}\sum_{t=1}^T \left(\frac{\widehat{\mathbf{\Lambda}}'\mathbf{x}_t}{n} - \mathbf{H}\mathbf{F}_t\right)\left(\mathbf{H}\mathbf{F}_t\right)' \right\| \\
&\leq \left\| \frac{1}{T^2}\sum_{t=1}^T \left(\frac{\widehat{\mathbf{\Lambda}}'\mathbf{\Lambda}}{n} - \mathbf{H}\right)\mathbf{F}_t\mathbf{F}_t'\mathbf{H}' \right\| + \left\| \frac{1}{T^2}\sum_{t=1}^T \frac{\widehat{\mathbf{\Lambda}}'\boldsymbol{\xi}_t\mathbf{F}_t'\mathbf{H}'}{n} \right\| \\
&\leq \left\| \frac{\widehat{\mathbf{\Lambda}}'\mathbf{\Lambda}}{n} - \mathbf{H} \right\| \left\| \frac{1}{T^2}\sum_{t=1}^T \mathbf{F}_t\mathbf{F}_t' \right\| \|\mathbf{H}\| + \frac{1}{T}\sum_{t=1}^T \left\| \frac{\widehat{\mathbf{\Lambda}}'\boldsymbol{\xi}_t\mathbf{F}_t'\mathbf{H}'}{nT} \right\| = O_p\left(\max\left(\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{T}}\right)\right),
\end{aligned} \tag{B26}$$

where we used Lemmas 3, 9*ii*, and 11*i*. By combining (B25) and (B26) we prove part *i)*. Parts *ii)*, *iii)*, and *iv)* are proved in the same way as part *i)*, but for stationary processes $\Delta\mathbf{F}_t$, hence by using Lemmas 9*i*, 3, 11*ii*, and 12*ii*.

Now, consider part *v)*:

$$\begin{aligned}
\left\| \frac{1}{T}\sum_{t=1}^T \Delta\widehat{\mathbf{F}}_t\widehat{\mathbf{F}}_{t-1}'\check{\boldsymbol{\beta}} - \frac{1}{T}\sum_{t=1}^T \Delta\check{\mathbf{F}}_t\check{\mathbf{F}}_{t-1}'\check{\boldsymbol{\beta}} \right\| &\leq \left\| \frac{1}{T}\sum_{t=1}^T \left(\Delta\widehat{\mathbf{F}}_t - \mathbf{H}\Delta\mathbf{F}_t\right)\left(\widehat{\mathbf{F}}_{t-1} - \mathbf{H}\mathbf{F}_{t-1}\right)'\check{\boldsymbol{\beta}} \right\| \\
&+ \left\| \frac{1}{T}\sum_{t=1}^T \left(\Delta\widehat{\mathbf{F}}_t - \mathbf{H}\Delta\mathbf{F}_t\right)\left(\check{\boldsymbol{\beta}}'\mathbf{H}\mathbf{F}_{t-1}\right)' \right\| \\
&+ \left\| \frac{1}{T}\sum_{t=1}^T \left(\mathbf{H}\Delta\mathbf{F}_t\right)\left(\widehat{\mathbf{F}}_{t-1} - \mathbf{H}\mathbf{F}_{t-1}\right)'\check{\boldsymbol{\beta}} \right\|.
\end{aligned} \tag{B27}$$

Similarly to (B25), from (15) and (1), the first term on the rhs of (B27) is such that

$$\begin{aligned}
&\left\|\frac{1}{T}\sum_{t=1}^{T}\left(\Delta\widehat{\mathbf{F}}_t-\mathbf{H}\Delta\mathbf{F}_{t-1}\right)\left(\check{\boldsymbol{\beta}}'\widehat{\mathbf{F}}_{t-1}-\check{\boldsymbol{\beta}}'\mathbf{H}\mathbf{F}_{t-1}\right)'\right\| = \left\|\frac{1}{T}\sum_{t=1}^{T}\left(\frac{\widehat{\boldsymbol{\Lambda}}'\Delta\mathbf{x}_t}{n}-\mathbf{H}\Delta\mathbf{F}_t\right)\left(\frac{\widehat{\boldsymbol{\Lambda}}'\mathbf{x}_{t-1}}{n}-\mathbf{H}\mathbf{F}_{t-1}\right)'\check{\boldsymbol{\beta}}\right\| \\
&\leq \underbrace{\left\|\frac{1}{T}\sum_{t=1}^{T}\frac{\widehat{\boldsymbol{\Lambda}}'\boldsymbol{\Lambda}\Delta\mathbf{F}_t\mathbf{F}_{t-1}'}{n}\left(\frac{\boldsymbol{\Lambda}'\widehat{\boldsymbol{\Lambda}}}{n}-\mathbf{H}'\right)\check{\boldsymbol{\beta}}+\mathbf{H}\Delta\mathbf{F}_t\mathbf{F}_{t-1}'\left(\mathbf{H}'-\frac{\boldsymbol{\Lambda}'\widehat{\boldsymbol{\Lambda}}}{n}\right)\check{\boldsymbol{\beta}}\right\|}_{\mathcal{A}_2}+\underbrace{\left\|\frac{1}{T}\sum_{t=1}^{T}\frac{\widehat{\boldsymbol{\Lambda}}'\boldsymbol{\Lambda}\Delta\mathbf{F}_t\boldsymbol{\xi}_{t-1}'\widehat{\boldsymbol{\Lambda}}\check{\boldsymbol{\beta}}}{n^2}\right\|}_{\mathcal{B}_2} \\
&+\underbrace{\left\|\frac{1}{T}\sum_{t=1}^{T}\frac{\widehat{\boldsymbol{\Lambda}}'\Delta\boldsymbol{\xi}_t\mathbf{F}_{t-1}'\boldsymbol{\Lambda}'\widehat{\boldsymbol{\Lambda}}\check{\boldsymbol{\beta}}}{n^2}\right\|}_{\mathcal{C}_2}+\underbrace{\left\|\frac{1}{T}\sum_{t=1}^{T}\frac{\mathbf{H}\Delta\mathbf{F}_t\boldsymbol{\xi}_{t-1}'\widehat{\boldsymbol{\Lambda}}\check{\boldsymbol{\beta}}}{n}\right\|}_{\mathcal{D}_2}+\underbrace{\left\|\frac{1}{T}\sum_{t=1}^{T}\frac{\widehat{\boldsymbol{\Lambda}}'\Delta\boldsymbol{\xi}_t\mathbf{F}_{t-1}'\mathbf{H}'\check{\boldsymbol{\beta}}}{n}\right\|}_{\mathcal{E}_2} \\
&+\underbrace{\left\|\frac{1}{T}\sum_{t=1}^{T}\frac{\widehat{\boldsymbol{\Lambda}}'\Delta\boldsymbol{\xi}_t\boldsymbol{\xi}_{t-1}'\widehat{\boldsymbol{\Lambda}}\check{\boldsymbol{\beta}}}{n^2}\right\|}_{\mathcal{F}_2}. \qquad \text{(B28)}
\end{aligned}$$

Let us consider first the terms:

$$\begin{aligned}
\mathcal{A}_2 &\leq \left\|\frac{\boldsymbol{\Lambda}'\widehat{\boldsymbol{\Lambda}}}{n}-\mathbf{H}'\right\|\left\|\frac{1}{T}\sum_{t=1}^{T}\Delta\mathbf{F}_t\mathbf{F}_{t-1}'\right\|\left\{\left\|\frac{\widehat{\boldsymbol{\Lambda}}'\boldsymbol{\Lambda}}{n}\right\|+\|\mathbf{H}\|\right\}\|\check{\boldsymbol{\beta}}\| = O_p\left(\frac{1}{\sqrt{T}}\right), \\
\mathcal{B}_2 &\leq \frac{1}{T}\sum_{t=1}^{T}\left\|\frac{\widehat{\boldsymbol{\Lambda}}'\boldsymbol{\xi}_{t-1}\Delta\mathbf{F}_t'}{n}\right\|\left\|\frac{\widehat{\boldsymbol{\Lambda}}'\boldsymbol{\Lambda}}{n}\right\|\|\check{\boldsymbol{\beta}}\| = O_p(\zeta_{nT,\delta}), \\
\mathcal{F}_2 &\leq \frac{1}{T}\sum_{t=1}^{T}\left\|\frac{\widehat{\boldsymbol{\Lambda}}'\Delta\boldsymbol{\xi}_t\boldsymbol{\xi}_{t-1}'\widehat{\boldsymbol{\Lambda}}}{n^2}\right\|\|\check{\boldsymbol{\beta}}\| = O_p\left(\zeta_{nT,\delta}\max\left(\frac{1}{\sqrt{n}},\frac{1}{\sqrt{T}}\right)\right),
\end{aligned}$$

Above we used, for $\mathcal{A}_2$ Lemmas 3 and 9*iv*, for $\mathcal{B}_2$ Lemma 11*vi*, for $\mathcal{F}_2$ Lemma 12*v*. The term $\mathcal{D}_2$ behaves exactly as $\mathcal{B}_2$, while $\mathcal{E}_2$ is $O_p(\max(n^{-1/2},T^{-1/2}))$ because of Lemma 11*iii*. Finally, recall that from Lemma 3, we have

$$\frac{\boldsymbol{\Lambda}'\widehat{\boldsymbol{\Lambda}}}{n}=\mathbf{H}'+O_p\left(\frac{1}{\sqrt{T}}\right). \qquad \text{(B29)}$$

Hence, from (B29),

$$\mathcal{C}_2 \leq \frac{1}{T}\sum_{t=1}^{T}\left\|\frac{\widehat{\boldsymbol{\Lambda}}'\Delta\boldsymbol{\xi}_t\mathbf{F}_{t-1}'\mathbf{H}'\check{\boldsymbol{\beta}}}{n}\right\|+\frac{1}{T}\sum_{t=1}^{T}\left\|\frac{\widehat{\boldsymbol{\Lambda}}'\Delta\boldsymbol{\xi}_t\mathbf{F}_{t-1}'}{n}\right\|O_p\left(\frac{1}{\sqrt{T}}\right)=O_p\left(\max\left(\frac{1}{\sqrt{n}},\frac{1}{\sqrt{T}}\right)\right).$$

Indeed, the first term on the rhs of $\mathcal{C}_2$ is $O_p(\max(n^{-1/2},T^{-1/2}))$ because of Lemma 11*iii*, while the second term is $O_p(\max(n^{-1/2},T^{-1/2}))$ because of Lemma 11*iv*. Therefore, the first term on the rhs of (B27) is $O_p(\max(\zeta_{nT,\delta},T^{-1/2}))$.

As for the second term on the rhs of (B27), since $\check{\boldsymbol{\beta}}'\mathbf{H}\mathbf{F}_{t-1}=\check{\boldsymbol{\beta}}'\check{\mathbf{F}}_{t-1}=\boldsymbol{\beta}'\mathbf{F}_{t-1}$, we have

$$\begin{aligned}
&\left\|\frac{1}{T}\sum_{t=1}^{T}\left(\Delta\widehat{\mathbf{F}}_t-\mathbf{H}\Delta\mathbf{F}_t\right)\left(\check{\boldsymbol{\beta}}'\mathbf{H}\mathbf{F}_{t-1}\right)'\right\|=\left\|\frac{1}{T}\sum_{t=1}^{T}\left(\frac{\widehat{\boldsymbol{\Lambda}}'\Delta\mathbf{x}_t}{n}-\mathbf{H}\Delta\mathbf{F}_t\right)\left(\boldsymbol{\beta}'\mathbf{F}_{t-1}\right)'\right\| \\
&\leq\left\|\frac{1}{T}\sum_{t=1}^{T}\left(\frac{\widehat{\boldsymbol{\Lambda}}'\boldsymbol{\Lambda}}{n}-\mathbf{H}\right)\Delta\mathbf{F}_t\mathbf{F}_{t-1}'\boldsymbol{\beta}\right\|+\left\|\frac{1}{T}\sum_{t=1}^{T}\frac{\widehat{\boldsymbol{\Lambda}}'\Delta\boldsymbol{\xi}_t\check{\mathbf{F}}_{t-1}'\check{\boldsymbol{\beta}}}{n}\right\|=O_p\left(\max\left(\frac{1}{\sqrt{n}},\frac{1}{\sqrt{T}}\right)\right), \qquad \text{(B30)}
\end{aligned}$$

where we used Lemma 3 and Lemma 9*iv* for the first term on the rhs and Lemma 11*iii* for the second.

The third term on the rhs of (B27) is such that

$$\left\|\frac{1}{T}\sum_{t=1}^{T}\Big(\mathbf{H}\Delta\mathbf{F}_t\Big)\Big(\widehat{\mathbf{F}}_{t-1}-\mathbf{H}\mathbf{F}_{t-1}\Big)'\check{\boldsymbol{\beta}}\right\| = \left\|\frac{1}{T}\sum_{t=1}^{T}\Big(\mathbf{H}\Delta\mathbf{F}_t\Big)\left(\frac{\widehat{\boldsymbol{\Lambda}}'\mathbf{x}_{t-1}}{n}-\mathbf{H}\mathbf{F}_{t-1}\right)'\check{\boldsymbol{\beta}}\right\|$$
$$\le\left\|\frac{1}{T}\sum_{t=1}^{T}\mathbf{H}\Delta\mathbf{F}_t\mathbf{F}_{t-1}'\left(\frac{\boldsymbol{\Lambda}'\widehat{\boldsymbol{\Lambda}}}{n}-\mathbf{H}'\right)\check{\boldsymbol{\beta}}\right\|+\left\|\frac{1}{T}\sum_{t=1}^{T}\frac{\mathbf{H}\Delta\mathbf{F}_t\boldsymbol{\xi}_{t-1}'\widehat{\boldsymbol{\Lambda}}\check{\boldsymbol{\beta}}}{n}\right\| = O_p(\zeta_{nT,\delta}), \tag{B31}$$

since the first term on the rhs behaves exactly as $\mathcal{A}_2$ above, while the second term is $O_p(\zeta_{nT,\delta})$ as in $\mathcal{B}_2$. By combining (B28), (B30), and (B31) we prove part *v)*.

Then consider part *vi)*:

$$\left\|\frac{1}{T}\sum_{t=1}^{T}\check{\boldsymbol{\beta}}'\widehat{\mathbf{F}}_t\widehat{\mathbf{F}}_t'\check{\boldsymbol{\beta}}-\frac{1}{T}\sum_{t=1}^{T}\check{\boldsymbol{\beta}}'\check{\mathbf{F}}_t\check{\mathbf{F}}_t'\check{\boldsymbol{\beta}}\right\| \le \left\|\frac{1}{T}\sum_{t=1}^{T}\check{\boldsymbol{\beta}}'\Big(\widehat{\mathbf{F}}_t-\mathbf{H}\mathbf{F}_t\Big)\Big(\widehat{\mathbf{F}}_t-\mathbf{H}\mathbf{F}_t\Big)'\check{\boldsymbol{\beta}}\right\|$$
$$+2\left\|\frac{1}{T}\sum_{t=1}^{T}\check{\boldsymbol{\beta}}'\Big(\widehat{\mathbf{F}}_t-\mathbf{H}\mathbf{F}_t\Big)\Big(\check{\boldsymbol{\beta}}'\mathbf{H}\mathbf{F}_t\Big)'\right\|. \tag{B32}$$

As before, from (15) and (1), the first term on the rhs of (B32) is such that

$$\left\|\frac{1}{T}\sum_{t=1}^{T}\check{\boldsymbol{\beta}}'\Big(\widehat{\mathbf{F}}_t-\mathbf{H}\mathbf{F}_t\Big)\Big(\widehat{\mathbf{F}}_t-\mathbf{H}\mathbf{F}_t\Big)'\check{\boldsymbol{\beta}}\right\| = \left\|\frac{1}{T}\sum_{t=1}^{T}\check{\boldsymbol{\beta}}'\left(\frac{\widehat{\boldsymbol{\Lambda}}'\mathbf{x}_t}{n}-\mathbf{H}\mathbf{F}_t\right)\left(\frac{\widehat{\boldsymbol{\Lambda}}'\mathbf{x}_t}{n}-\mathbf{H}\mathbf{F}_t\right)'\check{\boldsymbol{\beta}}\right\|$$
$$\le\underbrace{\left\|\frac{1}{T}\sum_{t=1}^{T}\frac{\check{\boldsymbol{\beta}}'\widehat{\boldsymbol{\Lambda}}'\boldsymbol{\Lambda}\mathbf{F}_t\mathbf{F}_t'}{n}\left(\frac{\boldsymbol{\Lambda}'\widehat{\boldsymbol{\Lambda}}}{n}-\mathbf{H}'\right)\check{\boldsymbol{\beta}}+\check{\boldsymbol{\beta}}'\mathbf{H}\mathbf{F}_t\mathbf{F}_t'\left(\mathbf{H}'-\frac{\boldsymbol{\Lambda}'\widehat{\boldsymbol{\Lambda}}}{n}\right)\check{\boldsymbol{\beta}}\right\|}_{\mathcal{A}_3}+2\underbrace{\left\|\frac{1}{T}\sum_{t=1}^{T}\frac{\check{\boldsymbol{\beta}}'\widehat{\boldsymbol{\Lambda}}'\boldsymbol{\Lambda}\mathbf{F}_t\boldsymbol{\xi}_t'\widehat{\boldsymbol{\Lambda}}\check{\boldsymbol{\beta}}}{n^2}\right\|}_{\mathcal{B}_3}$$
$$+2\underbrace{\left\|\frac{1}{T}\sum_{t=1}^{T}\frac{\check{\boldsymbol{\beta}}'\mathbf{H}\mathbf{F}_t\boldsymbol{\xi}_t'\widehat{\boldsymbol{\Lambda}}\check{\boldsymbol{\beta}}}{n}\right\|}_{\mathcal{C}_3}+\underbrace{\left\|\frac{1}{T}\sum_{t=1}^{T}\frac{\check{\boldsymbol{\beta}}'\widehat{\boldsymbol{\Lambda}}'\boldsymbol{\xi}_t\boldsymbol{\xi}_t'\widehat{\boldsymbol{\Lambda}}\check{\boldsymbol{\beta}}}{n^2}\right\|}_{\mathcal{D}_3}. \tag{B33}$$

By noticing that $\check{\boldsymbol{\beta}}'\mathbf{H}\mathbf{F}_t=\boldsymbol{\beta}'\mathbf{F}_t$ and using (B29), we have,

$$\mathcal{A}_3\le\left\|\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\beta}'\mathbf{F}_t\mathbf{F}_t'\left(\frac{\boldsymbol{\Lambda}'\widehat{\boldsymbol{\Lambda}}}{n}-\mathbf{H}'\right)\check{\boldsymbol{\beta}}\right\|+\left\|\frac{1}{T}\sum_{t=1}^{T}\check{\boldsymbol{\beta}}'\mathbf{F}_t\mathbf{F}_t'\left(\frac{\boldsymbol{\Lambda}'\widehat{\boldsymbol{\Lambda}}}{n}-\mathbf{H}'\right)\check{\boldsymbol{\beta}}\right\|O_p\left(\frac{1}{\sqrt{T}}\right)$$
$$+\left\|\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\beta}'\mathbf{F}_t\mathbf{F}_t'\left(\mathbf{H}'-\frac{\boldsymbol{\Lambda}'\widehat{\boldsymbol{\Lambda}}}{n}\right)\check{\boldsymbol{\beta}}\right\| = O_p\left(\frac{1}{\sqrt{T}}\right).$$

Indeed, the first and third terms on the rhs are $O_p(T^{-1/2})$ because of Lemma 3 and Lemma 9*v*, while using the same results the second term is

$$\left\|\frac{1}{T}\sum_{t=1}^{T}\check{\boldsymbol{\beta}}'\mathbf{F}_t\mathbf{F}_t'\left(\frac{\boldsymbol{\Lambda}'\widehat{\boldsymbol{\Lambda}}}{n}-\mathbf{H}'\right)\check{\boldsymbol{\beta}}\right\|O_p\left(\frac{1}{\sqrt{T}}\right)=\left\|\frac{1}{T}\sum_{t=1}^{T}\check{\boldsymbol{\beta}}'\mathbf{F}_t\mathbf{F}_t'\left(\frac{\boldsymbol{\Lambda}'\widehat{\boldsymbol{\Lambda}}\mathbf{H}}{n}-\mathbf{H}'\mathbf{H}\right)\mathbf{H}'\check{\boldsymbol{\beta}}\right\|O_p\left(\frac{1}{\sqrt{T}}\right)$$
$$=\left\|\frac{1}{T}\sum_{t=1}^{T}\check{\boldsymbol{\beta}}'\mathbf{F}_t\mathbf{F}_t'\boldsymbol{\beta}\right\|O_p\left(\frac{1}{T}\right)=O_p\left(\frac{1}{T}\right).$$

In the same way we have

$$\mathcal{B}_3 \le 2\left\|\frac{1}{T}\sum_{t=1}^T \frac{\check{\boldsymbol\beta}'\mathbf{H}\mathbf{F}_t\boldsymbol\xi_t'\widehat{\boldsymbol\Lambda}\check{\boldsymbol\beta}}{n}\right\| + 2\left\|\frac{1}{T}\sum_{t=1}^T \frac{\check{\boldsymbol\beta}'\mathbf{F}_t\boldsymbol\xi_t'\widehat{\boldsymbol\Lambda}\check{\boldsymbol\beta}}{n}\right\| O_p\left(\frac{1}{\sqrt{T}}\right) = O_p(\zeta_{nT,\delta}),$$

because of Lemmas 11*viii* and 11*vii*. Then,

$$\mathcal{C}_3 \le \frac{2}{T}\sum_{t=1}^T \left\|\frac{\check{\boldsymbol\beta}'\mathbf{H}\mathbf{F}_t\boldsymbol\xi_t'\widehat{\boldsymbol\Lambda}}{n}\right\| \|\check{\boldsymbol\beta}\| = O_p(\zeta_{nT,\delta}),$$

$$\mathcal{D}_3 \le \frac{1}{T}\sum_{t=1}^T \left\|\frac{\widehat{\boldsymbol\Lambda}'\boldsymbol\xi_t\boldsymbol\xi_t'\widehat{\boldsymbol\Lambda}}{n^2}\right\| \|\check{\boldsymbol\beta}\|^2 = O_p(\zeta^2_{nT,\delta}),$$

because of Lemmas 11*viii* and 12*iii*. Therefore, since from Assumption 6 $\zeta^2_{nT,\delta} < \zeta_{nT,\delta}$, the first term on the rhs of (B32) is $O_p(\zeta_{nT,\delta})$.

The second term on the rhs of (B32) is such that

$$2\left\|\frac{1}{T}\sum_{t=1}^T \check{\boldsymbol\beta}'\Big(\widehat{\mathbf{F}}_t - \mathbf{H}\mathbf{F}_t\Big)\Big(\check{\boldsymbol\beta}'\mathbf{H}\mathbf{F}_t\Big)'\right\| = 2\left\|\frac{1}{T}\sum_{t=1}^T \check{\boldsymbol\beta}'\Big(\frac{\widehat{\boldsymbol\Lambda}'\mathbf{x}_t}{n} - \mathbf{H}\mathbf{F}_t\Big)\Big(\check{\boldsymbol\beta}'\mathbf{H}\mathbf{F}_t\Big)'\right\|$$

$$\le 2\left\|\frac{1}{T}\sum_{t=1}^T \left(\frac{\widehat{\boldsymbol\Lambda}'\boldsymbol\Lambda}{n} - \mathbf{H}\right)\mathbf{F}_t\mathbf{F}_t'\mathbf{H}'\check{\boldsymbol\beta}\right\| + 2\left\|\frac{1}{T}\sum_{t=1}^T \frac{\widehat{\boldsymbol\Lambda}'\boldsymbol\xi_t\mathbf{F}_t'\mathbf{H}'\check{\boldsymbol\beta}}{n}\right\| = O_p\left(\max\left(\zeta_{nT,\delta}, \frac{1}{\sqrt{T}}\right)\right), \quad \text{(B34)}$$

because of Lemmas 9*v*, 3, and Lemma 11*viii*. By combining (B33) and (B34) we prove part *vi)*. Finally, parts *vii)*, *viii)*, and *ix)* are like part *v)*, by noticing that $\|T^{-1/2}\mathbf{F}_t\| = O_p(1)$ because of Lemma 10*ii*. This completes the proof. □

Lemma 14 *Consider the matrices* $\widehat{\mathbf{S}}_{ij}$ *defined in* (A31) *and denote by* $\mathbf{S}_{ij}$*, for* $i,j=0,1$*, the analogous ones when computed using* $\check{\mathbf{F}}_t = \mathbf{H}\mathbf{F}_t$*. Define also* $\check{\boldsymbol\beta} = \mathbf{H}\boldsymbol\beta$ *and* $\check{\boldsymbol\beta}_{\perp *} = \check{\boldsymbol\beta}_\perp(\check{\boldsymbol\beta}_\perp'\check{\boldsymbol\beta}_\perp)^{-1}$*, where* $\check{\boldsymbol\beta}_\perp = \mathbf{H}\boldsymbol\beta_\perp$ *such that* $\check{\boldsymbol\beta}_\perp'\check{\boldsymbol\beta} = \mathbf{0}_{r-c\times r}$*. As* $n,T\to\infty$*, if Assumptions 1-4 hold, then,*

i. $\|\widehat{\mathbf{S}}_{00} - \mathbf{S}_{00}\| = O_p(\max(n^{-1/2}, T^{-1/2}))$.

If also Assumptions 6b-6c hold, then,

ii. $\|\check{\boldsymbol\beta}'\widehat{\mathbf{S}}_{11}\check{\boldsymbol\beta} - \check{\boldsymbol\beta}'\mathbf{S}_{11}\check{\boldsymbol\beta}\| = O_p(\max(\zeta_{nT,\delta}, T^{-1/2}))$*;*

iii. $\|T^{-1/2}\check{\boldsymbol\beta}'\widehat{\mathbf{S}}_{11}\check{\boldsymbol\beta}_{\perp *} - T^{-1/2}\check{\boldsymbol\beta}'\mathbf{S}_{11}\check{\boldsymbol\beta}_{\perp *}\| = O_p(\max(\zeta_{nT,\delta}, T^{-1/2}))$*;*

iv. $\|T^{-1/2}\check{\boldsymbol\beta}'\widehat{\mathbf{S}}_{10}\widehat{\mathbf{S}}_{00}^{-1}\widehat{\mathbf{S}}_{01}\check{\boldsymbol\beta}_{\perp *} - T^{-1/2}\check{\boldsymbol\beta}'\mathbf{S}_{10}\mathbf{S}_{00}^{-1}\mathbf{S}_{01}\check{\boldsymbol\beta}_{\perp *}\| = O_p(\max(\zeta_{nT,\delta}, T^{-1/2}))$*;*

v. $\|T^{-1}\check{\boldsymbol\beta}_{\perp *}'\widehat{\mathbf{S}}_{10}\widehat{\mathbf{S}}_{00}^{-1}\widehat{\mathbf{S}}_{01}\check{\boldsymbol\beta}_{\perp *} - T^{-1}\check{\boldsymbol\beta}_{\perp *}'\mathbf{S}_{10}\mathbf{S}_{00}^{-1}\mathbf{S}_{01}\check{\boldsymbol\beta}_{\perp *}\| = O_p(\max(\zeta_{nT,\delta}, T^{-1/2}))$*;*

vi. $\|T^{-1}\check{\boldsymbol\beta}_{\perp *}'\widehat{\mathbf{S}}_{11}\check{\boldsymbol\beta}_{\perp *} - T^{-1}\check{\boldsymbol\beta}_{\perp *}'\mathbf{S}_{11}\check{\boldsymbol\beta}_{\perp *}\| = O_p(\max(\zeta_{nT,\delta}, T^{-1/2}))$.

Proof of Lemma 14 Throughout we use the fact that $\|\check{\boldsymbol\beta}_{\perp *}\| = O(1)$. Part *i)* is proved using Lemma 13*ii*, 13*iii*, and 13*iv*. For part *ii)* we use Lemma 13*v*, 13*vi*, and 13*iv*. Part *iii)* is proved by combining part *ii)*, Lemma 13*v* and 13*vi*, and by noticing that $\|T^{-1/2}\mathbf{F}_t\| = O_p(1)$ from Lemma 10*ii*. For part *iv)* we combine part *i)*, Lemma 13*v*, 13*viii*, and 13*ix*. Part *v)* is proved by combining part *i)*, Lemma 13*viii* and 13*ix*, and finally part *vi)* follows from Lemma 13*i* and 13*ix*. This completes the proof. □

Lemma 15 *Consider the matrices* $\widehat{\mathbf{S}}_{ij}$ *defined in* (A31) *and denote by* $\mathbf{S}_{ij}$*, for* $i,j=0,1$*, the analogous ones when computed using* $\check{\mathbf{F}}_t = \mathbf{H}\mathbf{F}_t$*. Define also* $\check{\boldsymbol\beta} = \mathbf{H}\boldsymbol\beta$ *and the conditional covariance matrices,* $\check{\boldsymbol\Omega}_{00}$*,* $\check{\boldsymbol\Omega}_{\check\beta\check\beta}$*, and* $\check{\boldsymbol\Omega}_{0\check\beta}$*, defined in* (A39)*. Under Assumptions 2, as* $T\to\infty$*,*

i. $\|\mathbf{S}_{00} - \check{\boldsymbol{\Omega}}_{00}\| = O_p(T^{-1/2})$;
ii. $\|\check{\boldsymbol{\beta}}'\mathbf{S}_{11}\check{\boldsymbol{\beta}} - \check{\boldsymbol{\Omega}}_{\check{\beta}\check{\beta}}\| = O_p(T^{-1/2})$;
iii. $\|\mathbf{S}_{01}\check{\boldsymbol{\beta}} - \check{\boldsymbol{\Omega}}_{0\check{\beta}}\| = O_p(T^{-1/2})$.

Proof of Lemma 15 For part *i)*, notice that

$$\begin{aligned}
\check{\boldsymbol{\Omega}}_{00} &= \mathsf{E}[\Delta\check{\mathbf{F}}_t\Delta\check{\mathbf{F}}_t'] - \mathsf{E}[\Delta\check{\mathbf{F}}_t\Delta\check{\mathbf{F}}_{t-1}']\Big(\mathsf{E}[\Delta\check{\mathbf{F}}_{t-1}\Delta\check{\mathbf{F}}_{t-1}']\Big)^{1}\mathsf{E}[\Delta\check{\mathbf{F}}_{t-1}\Delta\check{\mathbf{F}}_t'] \\
&= \boldsymbol{\Gamma}_0^{\Delta F} - \boldsymbol{\Gamma}_1^{\Delta F}\Big(\boldsymbol{\Gamma}_0^{\Delta F}\Big)^{-1}\boldsymbol{\Gamma}_1^{\Delta F}, \\
\mathbf{S}_{00} &= \frac{1}{T}\sum_{t=1}^T \Delta\check{\mathbf{F}}_t\Delta\check{\mathbf{F}}_t' - \left(\frac{1}{T}\sum_{t=2}^T \Delta\check{\mathbf{F}}_t\Delta\check{\mathbf{F}}_{t-1}'\right)\left(\frac{1}{T}\sum_{t=2}^T \Delta\check{\mathbf{F}}_{t-1}\Delta\check{\mathbf{F}}_{t-1}'\right)^{-1}\frac{1}{T}\sum_{t=2}^T \Delta\check{\mathbf{F}}_{t-1}\Delta\check{\mathbf{F}}_t' \\
&= \mathbf{M}_{00} - \mathbf{M}_{02}\mathbf{M}_{22}^{-1}\mathbf{M}_{20}.
\end{aligned}$$

Using Lemma 9*i*, we have the result. Parts *ii)* and *iii)* are proved in the same way using Lemma 9*iii* and 9*vi* respectively. This completes the proof. □

Lemma 16 *Under Assumptions 1-4, for any* $i = 1, \dots, n$ *and as* $T \to \infty$*, we have* $|\widetilde{b}_i - b_i| = O_p(T^{-1/2})$ *and* $|\widehat{b}_i - b_i| = O_p(T^{-1/2})$. *If* $x_{it} \sim I(0)$ *then* $|\widehat{b}_i - b_i| = O_p(T^{-3/2})$.

Proof of Lemma 16 For any $i = 1, \dots, n$, recall that we defined $x_{it} = a_i + \boldsymbol{\lambda}_i'\mathbf{F}_t + \xi_{it}$ so that $y_{it} = b_i t + x_{it}$. Define $\bar{y}_i = (T+1)^{-1}\sum_{t=0}^T y_{it}$ and $\bar{x}_i = (T+1)^{-1}\sum_{t=0}^T x_{it}$, then $\bar{y}_i = \bar{x}_i + b_i T/2$. From least squares trend slope estimator, $\widehat{b}_i$, in (24) we have

$$\widehat{b}_i - b_i = \frac{\sum_{t=0}^T (t - \frac{T}{2})(y_{it} - \bar{y}_i)}{\sum_{t=0}^T (t - \frac{T}{2})^2} - b_i = \frac{\sum_{t=0}^T (t - \frac{T}{2})(x_{it} - \bar{x}_i)}{\sum_{t=0}^T (t - \frac{T}{2})^2} = \frac{\sum_{t=0}^T t x_{it} - \frac{T}{2}\sum_{t=0}^T x_{it}}{\sum_{t=0}^T t^2 - \frac{T^2(T+1)}{4}}. \quad \text{(B35)}$$

The denominator of (B35) is $O(T^3)$. For the numerator, consider first the case in which $x_{it} \sim I(1)$, then under Assumptions 2-4, by Proposition 17.1 parts *d* and *f* in Hamilton (1994) we have, as $T \to \infty$,

$$\frac{1}{T^{3/2}}\sum_{t=0}^T x_{it} = O_p(1), \qquad \frac{1}{T^{5/2}}\sum_{t=0}^T t x_{it} = O_p(1).$$

When $x_{it} \sim I(0)$, then, by Proposition 17.1 parts *a* and *c* in Hamilton (1994) we have, as $T \to \infty$,

$$\frac{1}{T^{1/2}}\sum_{t=0}^T x_{it} = O_p(1), \qquad \frac{1}{T^{3/2}}\sum_{t=0}^T t x_{it} = O_p(1).$$

Therefore, by multiplying and dividing (B35) by T^3 we have the result both for $x_{it} \sim I(1)$ and for $x_{it} \sim I(0)$. This completes the proof. □

Appendix C Data Description and Data Treatment

No.	Series ID	Definition	Unit	F.	Source	SA	T
1	INDPRO	Industrial Production Index	2007=100	M	FED	1	2
2	IPBUSEQ	IP: Business Equipment	2007=100	M	FED	1	2
3	IPDCONGD	IP: Durable Consumer Goods	2007=100	M	FED	1	2
4	IPDMAT	IP: Durable Materials	2007=100	M	FED	1	2
5	IPNCONGD	IP: Nondurable Consumer Goods	2007=100	M	FED	1	2
6	IPNMAT	IP: nondurable Materials	2007=100	M	FED	1	2
7	CPIAUCSL	CPI: All Items	1982-84=100	M	BLS	1	3
8	CPIENGSL	CPI: Energy	1982-84=100	M	BLS	1	3
9	CPILEGSL	CPI: All Items Less Energy	1982-84=100	M	BLS	1	3
10	CPILFESL	CPI: All Items Less Food & Energy	1982-84=100	M	BLS	1	3
11	CPIUFDSL	CPI: Food	1982-84=100	M	BLS	1	3
12	CPIULFSL	CPI: All Items Less Food	1982-84=100	M	BLS	1	3
13	PPICRM	PPI: Crude Materials for Further Processing	1982=100	M	BLS	1	3
14	PPIENG	PPI: Fuels & Related Products & Power	1982=100	M	BLS	0	3
15	PPIFGS	PPI: Finished Goods	1982=100	M	BLS	1	3
16	PPIIDC	PPI: Industrial Commodities	1982=100	M	BLS	0	3
17	PPICPE	PPI: Finished Goods: Capital Equipment	1982=100	M	BLS	1	3
18	PPIACO	PPI: All Commodities	1982=100	M	BLS	0	3
19	PPIITM	PPI: Intermediate Materials	1982=100	M	BLS	1	3
20	AMBSL	St. Louis Adjusted Monetary Base	Bil. of $	M	StL	1	3
21	ADJRESSL	St. Louis Adjusted Reserves	Bil. of $	M	StL	1	3
22	CURRSL	Currency Component of M1	Bil. of $	M	FED	1	3
23	M1SL	M1 Money Stock	Bil. of $	M	FED	1	3
24	M2SL	M2 Money Stock	Bil. of $	M	FED	1	3
25	BUSLOANS	Commercial and Industrial Loans	Bil. of $	M	FED	1	2
26	CONSUMER	Consumer Loans	Bil. of $	M	FED	1	2
27	LOANINV	Bank Credit	Bil. of $	M	FED	1	2
28	LOANS	Loans and Leases in Bank Credit	Bil. of $	M	FED	1	2
29	REALLN	Real Estate Loans	Bil. of $	M	FED	1	2
30	TOTALSL	Tot. Cons. Credit Owned and Securitized	Bil. of $	M	FED	1	2
31	GDPC1	Gross Domestic Product	Bil. of Ch. 2005$	Q	BEA	1	2
32	FINSLC1	Final Sales of Domestic Product	Bil. of Ch. 2005$	Q	BEA	1	2
33	SLCEC1	State & Local CE & GI	Bil. of Ch. 2005$	Q	BEA	1	2
34	PRFIC1	Private Residential Fixed Investment	Bil. of Ch. 2005$	Q	BEA	1	2
35	PNFIC1	Private Nonresidential Fixed Investment	Bil. of Ch. 2005$	Q	BEA	1	2
36	IMPGSC1	Imports of Goods & Services	Bil. of Ch. 2005$	Q	BEA	1	2
37	GCEC1	Government CE & GI	Bil. of Ch. 2005$	Q	BEA	1	2
38	EXPGSC1	Exports of Goods & Services	Bil. of Ch. 2005$	Q	BEA	1	2
39	CBIC1	Change in Private Inventories	Bil. of Ch. 2005$	Q	BEA	1	1
40	PCNDGC96	PCE: Nondurable Goods	Bil. of Ch. 2005$	Q	BEA	1	2
41	PCESVC96	PCE: Services	Bil. of Ch. 2005$	Q	BEA	1	2
42	PCDGCC96	PCE: Durable Goods	Bil. of Ch. 2005$	Q	BEA	1	2
43	DGIC96	National Defense Gross Investment	Bil. of Ch. 2005$	Q	BEA	1	2
44	NDGIC96	Federal Nondefense Gross Investment	Bil. of Ch. 2005$	Q	BEA	1	2
45	DPIC96	Disposable Personal Income	Bil. of Ch. 2005$	Q	BEA	1	2
46	PCECTPI	PPCE: Chain-type Price Index	2005=100	Q	BEA	1	3
47	GPDICTPI	GPDI: Chain-type Price Index	2005=100	Q	BEA	1	3
48	GDPCTPI	GDP: Chain-type Price Index	2005=100	Q	BEA	1	3
49	HOUSTMW	Housing Starts in Midwest	Thous. of Units	M	Census	1	2
50	HOUSTNE	Housing Starts in Northeast	Thous. of Units	M	Census	1	2
51	HOUSTS	Housing Starts in South	Thous. of Units	M	Census	1	2
52	HOUSTW	Housing Starts in West	Thous. of Units	M	Census	1	2
53	PERMIT	Building Permits	Thous. of Units	M	Census	1	2
54	ULCMFG	Manuf. S.: Unit Labor Cost	2005=100	Q	BLS	1	2
55	COMPRMS	Manuf. S.: Real Compensation Per Hour	2005=100	Q	BLS	1	2
56	COMPMS	Manuf. S.: Compensation Per Hour	2005=100	Q	BLS	1	2
57	HOAMS	Manuf. S.: Hours of All Persons	2005=100	Q	BLS	1	2
58	OPHMFG	Manuf. S.: Output Per Hour of All Persons	2005=100	Q	BLS	1	2
59	ULCBS	Business S.: Unit Labor Cost	2005=100	Q	BLS	1	2
60	RCPHBS	Business S.: Real Compensation Per Hour	2005=100	Q	BLS	1	2
61	HCOMPBS	Business S.: Compensation Per Hour	2005=100	Q	BLS	1	2
62	HOABS	Business S.: Hours of All Persons	2005=100	Q	BLS	1	2
63	OPHPBS	Business S.: Output Per Hour of All Persons	2005=100	Q	BLS	1	2

No.	Series ID	Definition	Unit	F.	Source	SA	T
64	MPRIME	Bank Prime Loan Rate	%	M	FED	0	1
65	FEDFUNDS	Effective Federal Funds Rate	%	M	FED	0	1
66	TB3MS	3-Month T.Bill: Secondary Market Rate	%	M	FED	0	1
67	GS1	1-Year Treasury Constant Maturity Rate	%	M	FED	0	1
68	GS3	3-Year Treasury Constant Maturity Rate	%	M	FED	0	1
69	GS10	10-Year Treasury Constant Maturity Rate	%	M	FED	0	1
70	EMRATIO	Civilian Employment-Population Ratio	%	M	BLS	1	1
71	CE16OV	Civilian Employment	Thous. of Persons	M	BLS	1	2
72	UNRATE	Civilian Unemployment Rate	%	M	BLS	1	1
73	UEMPLT5	Civilians Unemployed - Less Than 5 Weeks	Thous. of Persons	M	BLS	1	2
74	UEMP5TO14	Civilians Unemployed for 5-14 Weeks	Thous. of Persons	M	BLS	1	2
75	UEMP15T26	Civilians Unemployed for 15-26 Weeks	Thous. of Persons	M	BLS	1	2
76	UEMP27OV	Civilians Unemployed for 27 Weeks and Over	Thous. of Persons	M	BLS	1	2
77	UEMPMEAN	Average (Mean) Duration of Unemployment	Weeks	M	BLS	1	2
78	UNEMPLOY	Unemployed	Thous. of Persons	M	BLS	1	2
79	DMANEMP	All Employees: Durable goods	Thous. of Persons	M	BLS	1	2
80	NDMANEMP	All Employees: Nondurable goods	Thous. of Persons	M	BLS	1	2
81	SRVPRD	All Employees: Service-Providing Industries	Thous. of Persons	M	BLS	1	2
82	USCONS	All Employees: Construction	Thous. of Persons	M	BLS	1	2
83	USEHS	All Employees: Education & Health Services	Thous. of Persons	M	BLS	1	2
84	USFIRE	All Employees: Financial Activities	Thous. of Persons	M	BLS	1	2
85	USGOOD	All Employees: Goods-Producing Industries	Thous. of Persons	M	BLS	1	2
86	USGOVT	All Employees: Government	Thous. of Persons	M	BLS	1	2
87	USINFO	All Employees: Information Services	Thous. of Persons	M	BLS	1	2
88	USLAH	All Employees: Leisure & Hospitality	Thous. of Persons	M	BLS	1	2
89	USMINE	All Employees: Mining and logging	Thous. of Persons	M	BLS	1	2
90	USPBS	All Employees: Prof. & Business Services	Thous. of Persons	M	BLS	1	2
91	USPRIV	All Employees: Total Private Industries	Thous. of Persons	M	BLS	1	2
92	USSERV	All Employees: Other Services	Thous. of Persons	M	BLS	1	2
93	USTPU	All Employees: Trade, Trans. & Ut.	Thous. of Persons	M	BLS	1	2
94	USWTRADE	All Employees: Wholesale Trade	Thous. of Persons	M	BLS	1	2
95	OILPRICE	Spot Oil Price: West Texas Intermediate	$ per Barrel	M	DJ	0	3
96	NAPMNOI	ISM Manuf.: New Orders Index	Index	M	ISM	1	1
97	NAPMPI	ISM Manuf.: Production Index	Index	M	ISM	1	1
98	NAPMEI	ISM Manuf.: Employment Index	Index	M	ISM	1	1
99	NAPMSDI	ISM Manuf.: Supplier Deliveries Index	Index	M	ISM	1	1
100	NAPMII	ISM Manuf.: Inventories Index	Index	M	ISM	1	1
101	SP500	S&P 500 Stock Price Index	Index	D	S&P	0	2

Abbreviations

Source	Freq.	Trans.	SA
BLS=U.S. Department of Labor: Bureau of Labor Statistics	Q = Quarterly	1 = None	0 = no
BEA=U.S. Department of Commerce: Bureau of Economic Analysis	M = Monthly	2 = log	1 = yes
ISM = Institute for Supply Management	D = Daily	3 = Δ log	
Census=U.S. Department of Commerce: Census Bureau			
FED=Board of Governors of the Federal Reserve System			
StL=Federal Reserve Bank of St. Louis			

Note: All monthly and daily series are transformed into quarterly observation by simple averages