Non-Stationary Dynamic Factor Models for Large Datasets

Matteo Barigozzi, Marco Lippi, and Matteo Luciani

Non-Stationary Dynamic Factor Models for Large Datasets

Matteo Barigozzi[1]   Marco Lippi[2]   Matteo Luciani[3]

July 18, 2017

Abstract

We study a Large-Dimensional Non-Stationary Dynamic Factor Model where (1) the factors $\mathbf{F}_t$ are $I(1)$ and singular, that is $\mathbf{F}_t$ has dimension r and is driven by q dynamic shocks with $q < r$, (2) the idiosyncratic components are either $I(0)$ or $I(1)$. Under these assumption the factors $\mathbf{F}_t$ are cointegrated and modeled by a singular Error Correction Model. We provide conditions for consistent estimation, as both the cross-sectional size n, and the time dimension T, go to infinity, of the factors, the loadings, the shocks, the ECM coefficients and therefore the Impulse Response Functions. 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]marco.lippi@eief.it – Einaudi Institute for Economics and Finance, Roma, Italy.
[3]matteo.luciani@frb.gov – Federal Reserve Board of Governors, Washington DC, USA.

1 Introduction

Since the early 2000s Large-Dimensional Dynamic Factor Models (DFM) have become increasingly popular in the econometric and macroeconomic 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 (Cristadoro et al., 2005; Altissimo et al., 2010), 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.

DFMs are based on the idea that all the variables in an economic system are driven by a few common (macroeconomic) shocks, with their residual dynamics being explained by idiosyncratic components, such as measurement errors and sectorial or regional shocks. Formally, each variable in the n-dimensional dataset x_{it}, $i = 1, 2, \ldots, n$, can be decomposed into the sum of two unobservable components: the common component χ_{it}, and the idiosyncratic component ξ_{it} (Forni et al., 2000; Forni and Lippi, 2001; Stock and Watson, 2002a,b). Moreover, the common components are linear combinations of an r-dimensional vector of common factors $\mathbf{F}_t = (F_{1t} \; F_{2t} \; \cdots \; F_{rt})'$,

$$
\begin{aligned}
x_{it} &= \chi_{it} + \xi_{it}, \\
\chi_{it} &= \lambda_{i1}F_{1t} + \lambda_{i2}F_{2t} + \cdots + \lambda_{ir}F_{rt} = \boldsymbol{\lambda}_i'\mathbf{F}_t,
\end{aligned}
\tag{1}
$$

where $\boldsymbol{\lambda}_i = (\lambda_{i1} \; \lambda_{i2} \; \cdots \; \lambda_{ir})'$. The stochastic vector $\mathbf{F}_t$ is in its turn dynamically driven by a q-dimensional orthonormal white-noise vector $\mathbf{u}_t = (u_{1t} \; u_{2t} \; \cdots \; u_{qt})'$, the common shocks:

$$
\mathbf{F}_t = \mathbf{B}(L)\mathbf{u}_t,
\tag{2}
$$

where $\mathbf{B}(L)$ is an $r \times q$ square-summable matrix in the lag operator (Stock and Watson, 2005; Bai and Ng, 2007; Forni et al., 2009). The dimension n of the dataset is assumed to be large as compared to r and q, which are independent of n, with $q \le r$. More precisely, all assumptions and results are formulated assuming that both T, the number of observations for each x_{it}, and n, the number of variables, tend to infinity.

In the standard version of the DFM, the components χ_{it} and ξ_{it}, and therefore the

observable variables x_{it}, are assumed to be stationary. Under the stationarity assumption, the factors $\mathbf{F}_t$, and the loadings $\boldsymbol{\lambda}_i$, can be consistently estimated by means of the first r principal components of the observable variables x_{it} (Stock and Watson, 2002a; Bai and Ng, 2002). Estimation of the matrix $\mathbf{B}(L)$ is usually obtained by means of a VAR for the estimated factors $\mathbf{F}_t$, this providing an estimate of the reduced-form, not identified, impulse-response functions (IRF) of the variables x_{it} with respect to the common shocks $\mathbf{u}_t$, that is $\boldsymbol{\lambda}_i' \mathbf{B}(L)$. Lastly, as shown in Stock and Watson (2005) and Forni et al. (2009), the identification techniques used in Structural VAR analysis (SVAR) can be applied to obtain shocks and IRFs fulfilling restrictions based on macroeconomic theory.

Of course the stationarity assumption does not hold for most of the variables contained in macroeconomic datasets. Assume for simplicity that all the variables x_{it} and the factors are $I(1)$. Equations (1) do not change, while the MA representation (2) becomes:

$$\Delta \mathbf{F}_t = \mathbf{C}(L)\mathbf{u}_t. \tag{3}$$

In this case, the common practice in the applied DFM literature consists in taking first differences of the non-stationary variables, so obtaining a stationary dataset Δx_{it} with stationary factors $\Delta \mathbf{F}_t$, and then applying the procedure described above to Δx_{it} and $\Delta \mathbf{F}_t$. This transformation is harmless as far as $\mathbf{F}_t$ and $\boldsymbol{\lambda}_i$ is concerned, as the first r principal components of the variables Δx_{it} consistently estimate $\Delta \mathbf{F}_t$ and therefore, by integration, $\mathbf{F}_t$, up to initial conditions (Bai and Ng, 2004). However, important issues arising in connection with estimation of the IRFs in the non-stationary case have not been systematically analysed so far. In particular, estimation of $\mathbf{C}(L)$ by means of a VAR is not trivial.

Firstly, if the factors $\mathbf{F}_t$ are cointegrated, consistent estimation of the long-run features of the IRFs requires modeling $\mathbf{F}_t$ as a Vector Error Correction Model (VECM). The few papers considering estimation of the IRFs in the non-stationary case model instead the dynamics of $\mathbf{F}_t$ as a VAR in differences (Stock and Watson, 2005; Forni et al., 2009).

Secondly, irrespective of whether the dataset is stationary or not, as a rule the vector $\mathbf{F}_t$ is singular, i.e. the number r of static common factors is greater than the number q of common shocks. This finding is strongly supported by empirical evidence, see e.g. Giannone et al. (2005), Amengual and Watson (2007), Forni and Gambetti (2010), Luciani (2015) for US macroeconomic databases, Barigozzi et al. (2014) for the Euro area.

The contribution of the present paper is the asymptotic analysis (consistency and rates) of estimators of IRFs for Non-Stationary Dynamic Factor Models for large datasets under

the assumptions that:

(I) the factors $\mathbf{F}_t$ are $I(1)$, singular and cointegrated (singular $I(1)$ vectors are trivially cointegrated, see below);

(II) the idiosyncratic components ξ_{it} are $I(1)$ or $I(0)$.

As regards (I), singularity of the vector $\mathbf{F}_t$ is consistent with a point made in several papers, see e.g. Stock and Watson (2005), Forni et al. (2009), that (1) and (2) are just a convenient *static* representation derived from a "deeper" set of *dynamic* equations linking the common components χ_{it} to the common shocks $\mathbf{u}_t$. For example, assuming for simplicity stationarity, suppose that $q = 1$ and that the common components load the single shock u_t with the simple MA dynamics

$$\chi_{it} = \mu_{i0} u_t + \mu_{i1} u_{t-1}.$$

Representation (1) is obtained by setting $r = 2$, $F_{1t} = u_t$, $F_{2t} = u_{t-1}$, $\boldsymbol{\lambda}_i = (\mu_{i0} \ \mu_{i1})'$, while (2) takes the form

$$\mathbf{F}_t = \begin{pmatrix} u_t \\ u_{t-1} \end{pmatrix} = \begin{pmatrix} 1 \\ L \end{pmatrix} u_t,$$

so that $\mathbf{B}(L) = (1 \ L)'$. This elementary model helps understanding the assumption of singularity for $\mathbf{F}_t$, but also helps to point out that $\mathbf{F}_t$ has not an autonomous economic content. Estimation of $\mathbf{F}_t$ and of $\mathbf{B}(L)$, or $\mathbf{C}(L)$, are used to obtain the dynamic response of x_{it} to the common shocks $\mathbf{u}_t$. In particular, the factors $\mathbf{F}_t$ are identified only up to a linear transformation and replacing $\mathbf{F}_t$ with $\mathbf{H}^{-1}\mathbf{F}_t$, $\mathbf{H}$ being an invertible matrix, obviously requires replacing $\boldsymbol{\lambda}_i$ with $\mathbf{H}'\boldsymbol{\lambda}_i$ and fairly obvious transformations of $\mathbf{B}(L)$, or $\mathbf{C}(L)$. However, it is easily seen that application of identification restrictions based on economic logic, such as recursive schemes or long-run effects, only applies to the shocks $\mathbf{u}_t$, which implies that the identified IRF are independent of $\mathbf{H}$, so that in this sense the factors $\mathbf{F}_t$ are just playing an auxiliary role, see Remark 2 in Section 2.

Second, in the companion paper Barigozzi et al. (2016) we address the representation problems for singular cointegrated vectors. Denoting by c the cointegration rank of $\mathbf{F}_t$, c is at least $r - q$, that is $c = r - q + d$ with $0 \le d < q$, hence singular $I(1)$ vectors are always cointegrated. Moreover, under the assumption that the entries of $\mathbf{C}(L)$ are rational functions of L, $\mathbf{F}_t$ has a representation as a VECM:

$$\mathbf{G}(L)\Delta\mathbf{F}_t + \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{F}_{t-1} = \mathbf{h} + \mathbf{K}\mathbf{u}_t, \tag{4}$$

where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are both $r \times c$ and full rank, $\mathbf{K}$ is $r \times q$ and $\mathbf{G}(L)$ is a *finite-degree* matrix polynomial. Trivially, representation (4) implies the existence of $q - d$ common trends for $\mathbf{F}_t$. In the present paper we study estimation of the DFM and the IRFs for non-stationary data when (4) holds. Specifically, we consider the case in which (4) is estimated as a VECM or by means of an unrestricted VAR in levels.

As regards (II), with the exception of Bai and Ng (2004), DFMs for $I(1)$ variables are studied under the assumption of stationary idiosyncratic components, see e.g. Bai (2004) and Peña and Poncela (2006) (which is however a model with fixed n). This is a crucial assumption with non-trivial consequences on the model. First, $I(0)$ idiosyncratic components imply that the x's and the factors are cointegrated. This property is exploited in Banerjee et al. (2017), who assume $I(0)$ idiosyncratic components and study a Factor Augmented Error Correction Model. However, not only the assumption of $I(0)$ idiosyncratic components is empirically not supported by typical macroeconomic datasets, as the one analyzed in this paper, but also, as we argue in Section 2, $I(0)$ idiosyncratic components imply "too much cointegration" among the variables x_{it} themselves.

Second, under the assumption of $I(0)$ idiosyncratic components, it is possible to separately estimate $I(1)$ non-cointegrated factors and $I(0)$ factors, see Bai (2004). As a consequence, representation (4) becomes trivial, the $I(0)$ factors being the errors terms. On the other hand, if the idiosyncratic terms are either $I(0)$ or $I(1)$, as we assume in the present paper, the estimated factors are all $I(1)$ in general, and estimation of (4) is not trivial.

The paper is organized as follows. In Section 2 we summarize and discuss the representation results proved in the companion paper Barigozzi et al. (2016). Moreover, we state the main assumptions of the model. Section 3 establishes consistency and rates for our estimators. In Section 4 we propose an information criterion for determining the number of common trends in a DFM. In Section 5, by means of a Monte Carlo simulation exercise, we study the finite sample properties of our estimators. Finally, in Section 6 we use our model to study the impact of monetary policy and supply shocks for the US economy. In Section 7 we conclude and discuss possible further applications of the model presented. The proofs of our main results and auxiliary lemmas are in Appendix A.

2   The Non-Stationary Dynamic Factor model

2.1 $I(1)$ vectors and cointegration

Throughout the paper we will adopt the following definitions for $I(0)$, $I(1)$ and cointegrated stochastic vectors. They are standard (see Johansen, 1995, Ch. 3), except that here the vectors can be singular, i.e. they can be driven by a number of shocks q and be of dimension r, with $r > q$.

(I) Consider an $r \times q$ matrix $\mathbf{A}(L) = \mathbf{A}_0 + \mathbf{A}_1 L + \cdots$, with the assumption that the series $\sum_{j=0}^{\infty} \mathbf{A}_j z^j$ converges for all complex number z such that $|z| < 1 + \delta$ for some $\delta > 0$. This condition is fulfilled when the entries of $\mathbf{A}(L)$ are rational functions of L with no poles inside or on the unit circle (the VARMA case). Given the r-dimensional stationary stochastic vector

$$\mathbf{y}_t = \mathbf{A}(L)\mathbf{v}_t,$$

where $\mathbf{v}_t$ is a q-dimensional white noise, $q \leq r$, we say that $\mathbf{y}_t$ is $I(0)$ if $\mathbf{A}(1) \neq \mathbf{0}$.

(II) The r dimensional stochastic vector $\mathbf{y}_t$ is $I(1)$ if $\Delta\mathbf{y}_t$ is $I(0)$.

(III) The r-dimensional $I(1)$ vector $\mathbf{y}_t$ is cointegrated of order c, $0 < c < r$, if (1) there exist linearly independent r-dimensional vectors $\boldsymbol{\beta}_k$, $k = 1, 2, \ldots, c$, such that $\boldsymbol{\beta}_k'\mathbf{y}_t$ is stationary, (2) if $\boldsymbol{\gamma}'\mathbf{y}_t$ is stationary then $\boldsymbol{\gamma}$ is a linear combination of the vectors $\boldsymbol{\beta}_k$.

Some important properties for our model follow from these definitions.

Remark 1

(a) Some of the coordinates of an $I(1)$ vector can be stationary.

(b) If one of the coordinates of the $I(1)$ vector $\mathbf{y}_t$ is stationary, then $\mathbf{y}_t$ is cointegrated.

(c) The cointegration rank of $\mathbf{y}_t$ is equal to r minus the rank of $\mathbf{A}(1)$.

(d) It easy to see that $\mathbf{y}_t$ is cointegrated with cointegration rank c if and only if $\mathbf{y}_t$ can be linearly transformed into a vector whose first c coordinates are stationary and the remaining $r - c$ are $I(1)$. For, let $\mathbf{y}_t$ be cointegrated of order c with cointegration vectors $\boldsymbol{\beta}_k$, $k = 1, 2, \ldots, c$. Let $\boldsymbol{\beta} = (\boldsymbol{\beta}_1 \ \boldsymbol{\beta}_2 \ \cdots \ \boldsymbol{\beta}_c)$ and $\mathbf{B} = (\boldsymbol{\beta} \ \boldsymbol{\beta}_\perp)$, where $\boldsymbol{\beta}_\perp$ is an $r \times (r - c)$ matrix whose columns are linearly independent and orthogonal to the columns of $\boldsymbol{\beta}$. Then, the first c coordinates of $\mathbf{z}_t = \mathbf{B}'\mathbf{y}_t$ are stationary while the remaining $r - c$ are $I(1)$.

(e) As is well known, in model (1) the factors $\mathbf{F}_t$ are identified up to a linear transformation, see Remark 2 for details. Thus, in view of (d), the question whether some of the factors are stationary while the remaining ones are $I(1)$ is perfectly equivalent to the question whether and "how much" the factors are cointegrated, see Bai (2004).

(f) Note that if $\mathbf{y}_t$ is $I(1)$ and $r > q$, then obviously $\mathbf{y}_t$ is cointegrated with cointegration rank at least $r - q$:
$$c = (r - q) + d, \quad 0 \le d < q. \tag{5}$$

2.2 Assumptions on common and idiosyncratic components

Under the assumption that $\mathbf{F}_t$ is $I(1)$, defining

$$\mathbf{x}_t = (x_{1t} \ x_{2t} \ \cdots \ x_{nt})', \ \boldsymbol{\chi}_t = (\chi_{1t} \ \chi_{2t} \ \cdots \ \chi_{nt})', \ \boldsymbol{\xi}_t = (\xi_{1t} \ \xi_{2t} \ \cdots \ \xi_{nt})', \ \boldsymbol{\Lambda} = (\boldsymbol{\lambda}_1 \ \boldsymbol{\lambda}_2 \ \cdots \ \boldsymbol{\lambda}_n)',$$

equations (1)–(3) become:
$$\mathbf{x}_t = \boldsymbol{\chi}_t + \boldsymbol{\xi}_t = \boldsymbol{\Lambda}\mathbf{F}_t + \boldsymbol{\xi}_t$$
$$\Delta\mathbf{F}_t = \mathbf{C}(L)\mathbf{u}_t. \tag{6}$$

Firstly, we suppose that the $I(1)$ stochastic vector $\mathbf{F}_t$ has an ARIMA representation:

$$\mathbf{S}(L)\Delta\mathbf{F}_t = \mathbf{Q}(L)\mathbf{u}_t, \tag{7}$$

or

$$\Delta\mathbf{F}_t = \mathbf{C}(L)\mathbf{u}_t = \mathbf{S}(L)^{-1}\mathbf{Q}(L)\mathbf{u}_t, \tag{8}$$

where:
(i) $\mathbf{u}_t$ is a q-dimensional white noise, $\mathrm{rk}(\mathsf{E}\left[\mathbf{u}_t\mathbf{u}_t'\right]) = q$;
(ii) $\mathbf{S}(L)$ is an $r \times r$ finite-degree matrix polynomial with no zeros inside or on the unit circle;
(iii) $\mathbf{S}(0) = \mathbf{I}_r$;
(iv) $\mathbf{Q}(L)$ is a finite-degree $r \times q$ matrix polynomial, $\mathbf{Q}(1) \neq 0$;
(v) $\mathrm{rk}(\mathbf{Q}(0)) = q$.

Setting $d = q - \mathrm{rk}(\mathbf{Q}(1))$, the cointegration rank of $\mathbf{F}_t$ is $c = r - \mathrm{rk}(\mathbf{Q}(1)) = (r-q)+d$. It is easy to show, see Barigozzi et al. (2016), that there exists a non-singular $q \times q$ matrix

R such that, defining

$$\mathbf{v}_t = \begin{pmatrix} \mathbf{v}_{1t} \\ \mathbf{v}_{2t} \end{pmatrix} = \mathbf{R}\mathbf{u}_t,$$

where $\mathbf{v}_{1t}$ has dimension d while $\mathbf{v}_{2t}$ has dimension $r - c = q - d$, the d shocks in $\mathbf{v}_{1t}$ have a temporary effect on $\mathbf{F}_t$ whereas the $q - d$ shocks in $\mathbf{v}_{2t}$ have a permanent effect. Thus the number of permanent shocks is r minus the cointegration rank, as in the non-singular case, while the number of transitory shocks is the complement to q, not r, as though $r - q$ transitory shocks had a zero coefficient. In applications to macroeconomic datasets permanent and transitory shocks can be interpreted as the usual supply and demand causes of fluctuation of the GDP and other key variables. In Sections 5 and 6 permanent and transitory effects on some of the variables x_{it} are used to identify structural IRFs.

The main result in Barigozzi et al. (2016), which is crucial for the present paper, is the following. Assume equation (7) for $\mathbf{F}_t$, suppose that $c = (r - q) + d$ and set $\boldsymbol{\beta} = (\boldsymbol{\beta}_1 \ \cdots \ \boldsymbol{\beta}_c)$. Then, for *generic values of the parameters* in the matrices $\mathbf{S}(L)$ and $\mathbf{Q}(L)$, $\mathbf{F}_t$ has the VECM representation:

$$\mathbf{G}(L)\Delta\mathbf{F}_t + \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{F}_{t-1} = \mathbf{h} + \mathbf{K}\mathbf{u}_t, \tag{9}$$

where:

(A) $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are full rank $r \times c$ matrices;

(B) $\mathbf{K} = \mathbf{Q}(0)$;

(C) $\mathbf{h}$ is a constant vector;

(D) $\mathbf{G}(L)$ is a *finite-degree* matrix polynomial with $\mathbf{G}(0) = \mathbf{I}_r$.

This VECM representation is obtained by combining the Granger Representation Theorem with recent results on singular stochastic vectors, see Anderson and Deistler (2008a,b). The existence of a finite-degree inverse of $\mathbf{Q}(L)$, see (D) above, is a consequence of singularity. Note that the VECM, fulfilling (A) through (D), holds generically, that is with the exception of a subset of lower dimension—thus except for a *negligible subset*—in the parameter space (details in Barigozzi et al., 2016). We use this result as a motivation for *assuming* that (9) holds.

In particular, we make the following assumptions on the factors, loadings and idiosyncratic components.

Assumption 1 *(Common factors)*
(a) The $I(1)$ r-dimensional stochastic vector $\mathbf{F}_t$, with cointegration rank $c = r - q + d$, has

an ARIMA representation

$$\mathbf{S}(L)\Delta\mathbf{F}_t = \mathbf{Q}(L)\mathbf{u}_t,$$

fulfilling properties (i) through (v), and a VECM representation

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

fulfilling properties (A) through (D).

(b) $\mathrm{rk}(\mathsf{E}[\Delta\mathbf{F}_t\Delta\mathbf{F}_t']) = \mathrm{rk}(\sum_{k=0}^{\infty} \mathbf{C}_k\mathbf{C}_k') = r.$

Part (a) of the next assumption implies that the r factors are not redundant, i.e. no representation with a number of factors smaller than r is possible.

Assumption 2 *(Loadings)*

(a) There exists an $r \times r$ positive definite matrix $\mathbf{V}$ such that, as $n \to \infty$,

$$n^{-1}\boldsymbol{\Lambda}'\boldsymbol{\Lambda} \to \mathbf{V}.$$

(b) Denoting by λ_{ij} the (i,j) element of $\boldsymbol{\Lambda}$, $|\lambda_{ij}| < C$ for some positive real C independent of i and j.

The idiosyncratic components are driven by idiosyncratic shocks with univariate dynamics and are orthogonal to the common components at any lead and lag:

Assumption 3 *(Idiosyncratic components)*

$$(1 - \rho_i L)\xi_{it} = d_i(L)\varepsilon_{it}, \tag{10}$$

where

(a) $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}\ \varepsilon_{2t}\ \cdots\ \varepsilon_{nt})'$ *is a vector white noise;*

(b) $d_i(L) = \sum_{k=0}^{\infty} d_{ik}$, *with* $\sum_{k=0}^{\infty} k|d_{ik}| \leq M_1$, *for some positive real M_1 independent of i;*

(c) $|\rho_i| \leq 1$, *so that $I(1)$ idiosyncratic components are allowed;*

(d) u_{jt} *and* ε_{is} *are orthogonal for any $j = 1, 2, \ldots, q$, $i \in \mathbb{N}$, and $t, s \in \mathbb{Z}$.*

Condition (b) implies square summability of the matrix polynomials in (10) so that ξ_{it} is non-stationary if and only if $\rho_i = 1$. Assuming that $|\rho_i| < 1$, that is all idiosyncratic components are stationary, implies that any p-dimensional vector $(x_{i_1,t}\ x_{i_2,t}\ \cdots\ x_{i_p,t})$, with $p \geq q - d + 1$, would be cointegrated. For, as we have seen above, the factors $\mathbf{F}_t$ are

9

driven by $r - c = q - d$ permanent shocks and the same holds for the variables $x_{i_h,t}$ if the idiosyncratic components are stationary. For example, if $q = 3$ and $d = 0$ then all 4-dimensional subvectors of $\mathbf{x}_t$ are cointegrated (3-dimensional if $d = 1$). Moreover, applying the test proposed in Bai and Ng (2004) on a panel of 101 quarterly US macroeconomic time series (see Section 6 and Appendix B), one of the datasets typically analysed in the empirical DFM literature, we found that the $I(0)$ hypothesis is rejected for half of the estimated idiosyncratic components.

Finally, note that contemporaneous cross-sectional dependence of the white noise $\boldsymbol{\varepsilon}_t$ is not excluded. More on this in Assumption 4.

To prove our consistency results we need enhancing Assumptions 1 and 3, in which we only require that $\mathbf{u}_t$ and the shocks ϵ_{it} are white noise, orthogonal at any lead and lag.

Assumption 4 *(Common and idiosyncratic shocks)*
(a) $\mathbf{u}_t$ is a strong orthonormal white noise, i.e. $\mathsf{E}[\mathbf{u}_t\mathbf{u}_t'] = \mathbf{I}_q$, $\mathbf{u}_t$ and $\mathbf{u}_{t-k}$ are independent for any $k \neq 0$;
(b) $\mathsf{E}[u_{jt}^4] \leq M_2$, for any $j = 1, 2, \ldots, q$ and a positive real M_2;
(c) $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}\ \varepsilon_{2t}\ \cdots\ \varepsilon_{nt})'$ is a strong vector white noise;
(d) $\mathsf{E}[|\varepsilon_{it}|^{\kappa_1}|\varepsilon_{jt}|^{\kappa_2}] \leq M_3$ for any $\kappa_1 + \kappa_2 = 4$, $i, j \in \mathbb{N}$ and a positive real M_3;
(e) $\max_{j=1,2,\ldots,n} \sum_{i=1}^{n} |\mathsf{E}[\varepsilon_{it}\varepsilon_{jt}]| \leq M_4$ for any $n \in \mathbb{N}$ and a positive real M_4;
(f) u_{jt} and ε_{is} are independent for any $j = 1, 2, \ldots, q$, $i \in \mathbb{N}$, and $t, s \in \mathbb{Z}$.

As noted above, contemporaneous cross-sectional dependence of the white noise $\boldsymbol{\varepsilon}_t$ is allowed. In particular, with condition (e) we require 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). The components of $\Delta\boldsymbol{\xi}_t$ are allowed to be both cross-sectionally and serially correlated. Condition (f) is in agreement with the economic interpretation of the model, in which common and idiosyncratic shocks are two independent sources of variation.

Lemmas 1 and 2 provide basic results for the eigenvalues of the covariance matrices of the idiosyncratic shocks ε_{it} and the variables Δx_{it}, $\Delta\chi_{it}$, $\Delta\xi_{it}$.

Lemma 1 *Under Assumptions 1 through 4, there exists a positive real M_5 such that $\mu_1^\varepsilon \leq M_5$ and $n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} |\mathsf{E}[\varepsilon_{it}\varepsilon_{jt}]| \leq M_5$ for any $n \in \mathbb{N}$.*

Lemma 2 *Under Assumptions 1 through 4, for any $n \in \mathbb{N}$, there exist positive reals $\underline{M}_6$, $\overline{M}_6$, M_7, $\underline{M}_8$, $\overline{M}_8$ and an integer $\bar{n}$ such that*

(i) $\underline{M}_6 \leq n^{-1}\mu_j^{\Delta\chi} \leq \overline{M}_6$ *for any* $j = 1, 2, \ldots, r$ *and* $n > \bar{n}$;

(ii) $\mu_1^{\Delta\xi} \leq M_7$;

(iii) $\underline{M}_8 \leq n^{-1}\mu_j^{\Delta x} \leq \overline{M}_8$ *for any* $j = 1, 2, \ldots, r$ *and* $n > \bar{n}$;

(iv) $\mu_{r+1}^{\Delta x} \leq M_7$.

The results in Lemma 2 are crucial to estimate the number of factors r, the loadings, the differenced factors and the factors themselves. Analogous results on the eigenvalues of the spectral density matrices of the x's, the χ's and the ξ's, allow the estimation of q and the cointegration rank c of the factors $\mathbf{F}_t$, see Section 4.

Remark 2 In model (6) the factors $\mathbf{F}_t$ are not identified. For, given the non singular $r \times r$ matrix $\mathbf{H}$,

$$\mathbf{x}_t = [\mathbf{\Lambda H}]\left[\mathbf{H}^{-1}\mathbf{F}_t\right] + \boldsymbol{\xi}_t = \mathbf{\Lambda}^*\mathbf{F}_t^* + \boldsymbol{\xi}_t. \tag{11}$$

Using $\mathbf{F}_t^*$ implies changes in the matrices in (8) and (9) and the loadings that are easy to compute:

$$\mathbf{\Lambda}^* = \mathbf{\Lambda H}, \quad \mathbf{S}^*(L) = \mathbf{H}^{-1}\mathbf{S}(L)\mathbf{H}, \quad \mathbf{Q}^*(L) = \mathbf{H}^{-1}\mathbf{Q}(L), \quad \mathbf{C}^*(L) = \mathbf{H}^{-1}\mathbf{C}(L),$$
$$\mathbf{G}^*(L) = \mathbf{H}^{-1}\mathbf{G}(L)\mathbf{H}, \quad \boldsymbol{\alpha}^* = \mathbf{H}^{-1}\boldsymbol{\alpha}, \quad \boldsymbol{\beta}^* = \mathbf{H}\boldsymbol{\beta}, \quad \mathbf{K}^* = \mathbf{H}^{-1}\mathbf{K}.$$

Note that $\mathbf{\Lambda}^*\mathbf{C}^*(L) = \mathbf{\Lambda C}(L)$, so that the raw IRFs of the x's with respect to $\mathbf{u}_t$, corresponding to the factors $\mathbf{F}_t^*$ and to the factors $\mathbf{F}_t$ are equal. As a consequence, identification of the IRFs based on any economic criterion is independent of the particular factors used.

The following choice of the factors is very convenient and will be adopted in the sequel. Let $\mathbf{W}$ be the $n \times r$ matrix whose columns are the right normalised eigenvectors of the variance-covariance matrix of $\Delta\boldsymbol{\chi}_t$, corresponding to the first r eigenvalues $\mu_j^{\Delta\chi}$, $j = 1, 2, \ldots, r$. Define

$$\Delta\mathbf{F}_t^* = \mathbf{W}'\Delta\boldsymbol{\chi}_t.$$

Now project $\Delta\boldsymbol{\chi}_t$ on $\Delta\mathbf{F}_t^*$:

$$\Delta\boldsymbol{\chi}_t = \mathcal{A}\Delta\mathbf{F}_t^* + \mathcal{R}_t.$$

We see that $\mathcal{A} = \mathbf{W}$ and that the variance-covariance matrices of $\Delta\boldsymbol{\chi}_t$ and of $\mathbf{W}\Delta\mathbf{F}_t^*$ are equal, so that $\mathcal{R}_t = \mathbf{0}$ and the projection becomes $\Delta\boldsymbol{\chi}_t = \mathbf{WW}'\Delta\boldsymbol{\chi}_t$, that is

$$\left(\mathbf{I}_n - \mathbf{WW}'\right)\Delta\boldsymbol{\chi}_t = \mathbf{0}.$$

Setting $\boldsymbol{\chi}_0 = \mathbf{0}$ we obtain $\boldsymbol{\chi}_t = \mathbf{W}[\mathbf{W}'\boldsymbol{\chi}_t]$, for $t > 0$, or, in our preferred specification,

$$\boldsymbol{\chi}_t = \left[\sqrt{n}\mathbf{W}\right]\left[\frac{1}{\sqrt{n}}\mathbf{W}'\boldsymbol{\chi}_t\right]. \tag{12}$$

We do not need to complicate the notation by introducing new symbols and set henceforth

$$\boldsymbol{\Lambda} = \sqrt{n}\mathbf{W}, \quad \mathbf{F}_t = \frac{1}{\sqrt{n}}\mathbf{W}'\boldsymbol{\chi}_t = \frac{1}{n}\boldsymbol{\Lambda}\boldsymbol{\chi}_t. \tag{13}$$

Note that now the factors $\mathbf{F}_t$ and the loadings $\boldsymbol{\lambda}_i$, for a given i, depend on n and that in the new specification:

$$\frac{1}{n}\boldsymbol{\Lambda}'\boldsymbol{\Lambda} = \mathbf{I}_r, \tag{14}$$

for all $n \in \mathbb{N}$. Moreover, the variance-covariance matrix of the differenced factors $\Delta\mathbf{F}_t$ is the diagonal $r \times r$ matrix with $\mu_j^{\Delta\chi}/n$ as the $(j,\ j)$ entry. By Lemma 2, (i), which is a consequence of Assumption 1, (b) and 2, all such entries are bounded and bounded away from zero.

We conclude with the following assumption, which has the consequence that $\boldsymbol{\chi}_0 = \mathbf{0}$, $\boldsymbol{\xi}_0 = \mathbf{0}$ and $\mathbf{x}_0 = \mathbf{0}$.

Assumption 5 *For all $i \in \mathbb{N}$ and $t \le 0$, $\mathbf{u}_t = \mathbf{0}$ and $\epsilon_{it} = 0$.*

3   Estimation

We proceed in the same way as Stock and Watson (2005) and Forni et al. (2009) do in their stationary setting: (i) we estimate the loadings, the common factors, their VECM dynamics and the raw IRFs, (ii) we identify the structural common shocks and IRFs by imposing a set of restrictions based on economic logic.

We observe an n-dimensional vector $\mathbf{x}_t$ over the period $0,\ 1,\ldots,\ T$, i.e. the $n \times (T+1)$ panel $\mathbf{x} = (\mathbf{x}_0 \cdots \mathbf{x}_T)$. Asymptotics for all our estimators is studied for both n and T tending to infinity. The number of common factors r, of common shocks q, and of the cointegration relations $c = r - q + d$ is assumed to be known in the present section, their estimation is studied in Section 4.

We denote estimated quantities with a hat, like in $\widehat{\mathbf{F}}_t$, without explicit notation for their dependence on both n and T. Moreover, (1) the spectral norm of a matrix $\mathbf{B}$ is denoted by $\|\mathbf{B}\| = (\mu_1^{\mathbf{B}'\mathbf{B}})^{1/2}$, where $\mu_1^{\mathbf{B}'\mathbf{B}}$ is the largest eigenvalue of $\mathbf{B}'\mathbf{B}$, (2) we denote by $\mathbf{J}$ a diagonal $r \times r$ matrix, depending on n and T, whose diagonal entries are either 1 or -1.

3.1 Loadings and common factors

We start with the model in differences,

$$\Delta\mathbf{x}_t = \mathbf{\Lambda}\Delta\mathbf{F}_t + \Delta\boldsymbol{\xi}_t.$$

Consider the $n \times T$ data matrix $\Delta\mathbf{x} = (\Delta\mathbf{x}_1 \cdots \Delta\mathbf{x}_T)$. Let $\widehat{\mathbf{\Gamma}}_0 = T^{-1}\Delta\mathbf{x}\Delta\mathbf{x}'$ be the sample covariance matrix of $\Delta\mathbf{x}_t$ and $\widehat{\mathbf{W}}$ the $n \times r$ matrix with the normalized eigenvectors of $\widehat{\mathbf{\Gamma}}_0$, corresponding to the first r eigenvalues, on the columns. The standard estimators of the loadings and the differenced factors are

$$\widehat{\mathbf{\Lambda}} = \sqrt{n}\widehat{\mathbf{W}}, \quad \Delta\widehat{\mathbf{F}}_t = \frac{1}{\sqrt{n}}\widehat{\mathbf{W}}'\Delta\mathbf{x}_t = \frac{1}{n}\widehat{\mathbf{\Lambda}}'\Delta\mathbf{x}_t.$$

Integrating $\widehat{\mathbf{F}}_t$ under the condition $\mathbf{x}_0 = \mathbf{0}$,

$$\widehat{\mathbf{\Lambda}} = \sqrt{n}\widehat{\mathbf{W}}, \quad \widehat{\mathbf{F}}_t = \frac{1}{\sqrt{n}}\widehat{\mathbf{W}}'\mathbf{x}_t = \frac{1}{n}\widehat{\mathbf{\Lambda}}'\mathbf{x}_t. \tag{15}$$

Lemma 3 *Under Assumptions 1 through 5, as $n, T \to \infty$,*
  (i) Given i, $\|\widehat{\boldsymbol{\lambda}}_i' - \boldsymbol{\lambda}_i'\mathbf{J}\| = O_p(\max(n^{-1/2}, T^{-1/2}))$;
  (ii) Given t, $\|\Delta\widehat{\mathbf{F}}_t - \mathbf{J}\Delta\mathbf{F}_t\| = O_p(\max(n^{-1/2}, T^{-1/2}))$;
  (iii) Given t, $T^{-1/2}\|\widehat{\mathbf{F}}_t - \mathbf{J}\mathbf{F}_t\| = O_p(\max(n^{-1/2}, T^{-1/2}))$.

Our proof of statements (i) and (ii) is close to the one given in Forni et al. (2009). However, they make direct assumptions on the estimate of the covariance matrix of the x's, whereas we start with "deeper" assumptions on common and idiosyncratic components. Moreover, we do not need assuming that the eigenvalues $\mu^{\Delta\chi j}$, $j = 1, 2, \ldots, r$, are asymptotically separated. Bai and Ng (2004) define their estimators as in (15) and prove statement (iii). In this respect, their paper and the present one only differ for the technique used in the proof. A significant difference, concerning detrending, which is needed when the actual data $\mathbf{x}$ contain a deterministic component, is discussed in Section 3.5.

Lemma 3, though interesting *per se*, is not sufficient to prove our main result on the VECM representation of $\mathbf{F}_t$ and the IRFs. In particular, we need the asymptotic properties of the sample second moments of $\widehat{\mathbf{F}}_t$ and $\Delta\widehat{\mathbf{F}}_t$. The main results, proved in Appendix A, Lemma A5, are worth mentioning here. 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) $T^{-2} \sum_{t=1}^{T} \widehat{\mathbf{F}}_t \widehat{\mathbf{F}}_t' \xrightarrow{d} \int_0^1 \mathbf{W}(\tau)\mathbf{W}'(\tau)\mathrm{d}\tau$, where $\mathbf{W}(\cdot)$ is an r-dimensional Brownian motion with finite covariance matrix of rank $q - d$.

3.2 VECM for the common factors

We now turn to estimation of the VECM in (9), with $c = r - q + d$ cointegration relations, see Assumption 1:

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

For simplicity, we assume that the degree of $\mathbf{G}(L)$ is $p = 1$. Generalization to any degree, $p > 1$, is straightforward. As a consequence of Assumption 5 we set $\mathbf{h} = \mathbf{0}$.

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 (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).[1]

Since $\mathbf{F}_t$ are unobserved, 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 least square 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, for $j = 1, 2, \ldots c$,

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

The vectors $\widehat{\boldsymbol{\beta}}_j$ are then the c columns of the estimated matrix $\widehat{\boldsymbol{\beta}}$. The other parameters of the VECM, $\boldsymbol{\alpha}$ and $\mathbf{G}_1$, are estimated in a second step as the least square estimators in the regressions of $\Delta\widehat{\mathbf{F}}_t$ on $\widehat{\boldsymbol{\beta}}'\widehat{\mathbf{F}}_{t-1}$ and on $\Delta\widehat{\mathbf{F}}_{t-1}$, respectively.

Finally, a linear combination of the q columns of $\mathbf{K}$ can be estimated as the first q eigenvectors of the sample covariance matrix of the VECM residuals $\widehat{\mathbf{w}}_t$, rescaled by the

---

[1]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).

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 $\widehat{\mathbf{K}}$.

Consistent estimation of (16) in presence of estimated factors, is possible under the following additional assumption.

Assumption 6

(a) *Let n_1 be the number of $I(1)$ variables among ξ_{1t}, $\xi_{2t}, \ldots,$ ξ_{nt} (i.e. the number of idiosyncratic components ξ_{it} such that $\rho_i = 1$, see Assumption 3). Then $n_1 = O(n^\delta)$ for some $\delta \in [0,1)$;*

(b) *$Tn^{-(2-\delta)} \to 0$, as $n, T \to \infty$;*

(c) *let $\mathcal{I}_0$ and $\mathcal{I}_1$ be the sets $\{i \leq n,$ such that ξ_{it} is $I(0)\}$ and $\{i \leq n,$ such that ξ_{it} is $I(1)\}$ respectively. Then:*

$$n^{-\gamma} \sum_{i \in \mathcal{I}_0} \sum_{j \in \mathcal{I}_1} |\mathsf{E}[\varepsilon_{it}\varepsilon_{jt}]| \leq M_9$$

for some $\gamma < \delta$, som positive real M_9 and any $n \in \mathbb{N}$.

Under condition (a), we put an asymptotic limit to the number of $I(1)$ idiosyncratic components. Their number n_1 can grow to infinity but slower than the number of the $I(0)$ components. Condition (b) imposes a constraint on the relative growth rates of n and T and it implies that at least $T^{1/2}/n \to 0$ (when $\delta = 0$). Further motivations for, and the implications of, these two requirements are given below. Finally, with reference to the partitioning of the vector of idiosyncratic components into $I(1)$ and $I(0)$ coordinates, condition (c) limits the dependence between the two blocks more than the dependence within each block, which is in turn controlled by Lemma 1.[2]

We then have consistency of the estimated VECM parameters.

Lemma 4 *Define $\vartheta_{nT,\delta} = \max\left(T^{1/2}n^{-(2-\delta)/2}, n^{-(1-\delta)/2}, T^{-1/2}\right)$. Under Assumptions 1 through 6, and given $\mathbf{J}$ 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) $\|\widehat{\boldsymbol{\beta}} - \mathbf{J}\boldsymbol{\beta}\mathbf{Q}\| = O_p(T^{-1/2}\vartheta_{nT,\delta})$;

(ii) $\|\widehat{\boldsymbol{\alpha}} - \mathbf{J}\boldsymbol{\alpha}\mathbf{Q}\| = O_p(\vartheta_{nT,\delta})$;

(iii) $\|\widehat{\mathbf{G}}_1 - \mathbf{J}\mathbf{G}_1\mathbf{J}\| = O_p(\vartheta_{nT,\delta})$;

(iv) $\|\widehat{\mathbf{K}} - \mathbf{J}\mathbf{K}\mathbf{R}\| = O_p(\vartheta_{nT,\delta})$.

---

[2]We could in principle consider any $\gamma < 1$, in which case the rate of convergence of Lemma 4 and Proposition 1 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 rate of convergence in Lemma 4 is determined by $\vartheta_{nT,\delta}$. In particular, for generic values of $\delta \in [0,1)$ we have

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

Consistency of the estimated parameters is guaranteed if and only if $\vartheta_{nT,\delta} \to 0$, as $n, T \to \infty$, which is ensured by Assumption 6. The intuitive explanation for Assumption 6 is as follows: due to non-stationarity the factor estimation error grows with T, but since, as defined in (15), the estimated factors are cross-sectional averages of the x's, we can keep this error under control by allowing for an increasingly large cross-sectional dimension, n. In particular, the factor estimation error is a weighted average of the idiosyncratic components, and therefore the trade-off between n and T depends on how many of those components are non-stationary.

The following remarks provide some more intuition about the results in Lemma 4.

Remark 3 From (17), we see that the classical $T^{1/2}$-consistency is achieved if and only if $T^{1/(1-\delta)}/n \to 0$, that is when n grows much faster than T. On the other hand, in the case $n = O(T)$, which is of particular interest since it corresponds to typical macroeconomic datasets, 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 $T^{1/2}$-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 we have the classical $T^{1/2}$-consistency if and only if $T/n \to 0$.

Remark 4 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, $\widehat{\boldsymbol{\beta}}$ converges to the true value, $\boldsymbol{\beta}$, at a faster rate with respect to the rate of consistency of the other VECM parameters. This is enough to consistently apply the two-step VECM estimation as in Johansen (1995).

Remark 5 The estimated parameters approach the true parameters only up to three transformations $\mathbf{J}$, $\mathbf{Q}$, and $\mathbf{R}$. First of all, since the estimated factors identify the true ones only up to a sign, determined by $\mathbf{J}$, the same holds for the estimated VECM parameters with obvious multiplications by $\mathbf{J}$. As already explained in Remark 2, this issue does not affect estimation and identification of IRFs. Second, the matrix $\mathbf{Q}$ represents the usual

indeterminacy in the identification of the cointegration relations, but again its identification does not affect the IRFs. This is also in agreement with the fact that, in our setup, neither the factors nor their cointegration relations have any economic meaning. Last, the matrix $\mathbf{R}$ represents indeterminacy in the identification of the matrix $\mathbf{K}$, and, as discussed below, $\mathbf{R}$ has to be determined in order to identify the structural IRFs.

3.3   Common shocks and impulse response functions

Throughout the rest of the section we denote the true IRF of x_{it}, for $i = 1, 2, \ldots n$, to the shock u_{jt}, for $j = 1, 2, \ldots, q$, as (see also (6))

$$\phi_{ij}(L) = \boldsymbol{\lambda}_i' \left[\frac{\mathbf{c}_j(L)}{1 - L} \right], \tag{18}$$

where $\boldsymbol{\lambda}_i'$ is the i-th row of $\boldsymbol{\Lambda}$, $\mathbf{c}_j(L)$ is the j-th column of $\mathbf{C}(L)$, and the notation used is convenient and makes sense, provided that we do not forget that such IRF is not square summable.

A VECM(p) with cointegration rank c can also be written as a VAR$(p+1)$ with $r - c$ unit roots. Therefore, after estimating (16), we have the estimated matrix polynomial

$$\widehat{\mathbf{A}}^{\text{VECM}}(L) = \mathbf{I}_r - \sum_{k=1}^{p+1} \widehat{\mathbf{A}}_k^{\text{VECM}} L^k,$$

with coefficients given by

$$\begin{aligned}
\widehat{\mathbf{A}}_1^{\text{VECM}} &= \widehat{\mathbf{G}}_1 - \widehat{\boldsymbol{\alpha}}\widehat{\boldsymbol{\beta}}' + \mathbf{I}_r, \\
\widehat{\mathbf{A}}_k^{\text{VECM}} &= \widehat{\mathbf{G}}_k - \widehat{\mathbf{G}}_{k-1}, \qquad k = 2, 3, \ldots, p \\
\widehat{\mathbf{A}}_{p+1}^{\text{VECM}} &= -\widehat{\mathbf{G}}_p.
\end{aligned} \tag{19}$$

such that $\text{rk}(\widehat{\mathbf{A}}^{\text{VECM}}(1)) = \text{rk}(\widehat{\boldsymbol{\alpha}}\widehat{\boldsymbol{\beta}}') = c$. Then, for $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, q$, the raw IRFs estimator is defined as

$$\widetilde{\phi}_{ij}^{\text{VECM}}(L) = \widehat{\boldsymbol{\lambda}}_i' \left[\widehat{\mathbf{A}}^{\text{VECM}}(L) \right]^{-1} \widehat{\mathbf{k}}_j. \tag{20}$$

where $\widehat{\boldsymbol{\lambda}}_i'$ is the i-th row of $\widehat{\boldsymbol{\Lambda}}$, $\widehat{\mathbf{k}}_j$ is the j-th column of $\widehat{\mathbf{K}}$ (see also Lütkepohl, 2006, for an explicit expression as function of the VECM parameters).

However, since $\mathbf{K}$ is not identified, the IRFs in (20) are in general not identified. Now, while orthogonality of $\mathbf{R}$ in Lemma 4 is a purely mathematical result due to non-uniqueness of eigenvectors, economic theory tells us that the choice of the identifying transformation can be determined by the economic meaning attached to the common shocks, $\mathbf{u}_t$. We then need to impose at most $q(q-1)/2$ restrictions in order to achieve under- or just-identification.[3] 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 and identified IRFs are then defined by combining the estimated parameters and the identification restrictions. In particular, for $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, q$, the dynamic reaction of the i-th variable to the j-th common shock is estimated as

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

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

By denoting as $\widehat{\phi}_{ijk}^{\text{VECM}}$ the k-th coefficient of the polynomial in (21), and as ϕ_{ijk} the corresponding coefficients of $\phi_{ij}(L)$, we have the following consistency result.

Proposition 1 *(Consistency of Impulse Response Functions based on VECM)*
Under Assumptions 1 through 6, as $n, T \to \infty$, we have

$$\left| \widehat{\phi}_{ijk}^{VECM} - \phi_{ijk} \right| = O_p(\vartheta_{nT,\delta}), \tag{22}$$

for any $k \geq 0$, $i = 1, 2, \ldots, n$, and $j = 1, 2, \ldots, q$.

The proof of Proposition 1, follows directly by combining Lemma 3(i) and 4. As noticed above this result is not affected by the fact that common factors and their cointegration relations are not identified. All previous remarks on convergence rates apply also in this case.

---

[3] In principle any invertible transformation can be considered in order to achieve such identification. However, traditional macroeconomic practice assumes Gaussianity of the shocks and therefore restricts to orthogonal matrices only, that is to uncorrelated common shocks.

3.4 The case of unrestricted VAR for the common factors

In presence of non-singular cointegrated vectors, several papers have addressed the issue whether and when a VECM or an unrestricted VAR for the levels should be used for estimation. Sims et al. (1990) show that the parameters of a cointegrated VAR 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 IRFs 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 IRFs 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 IRFs when, following Sims et al. (1990), we consider least squares estimation of an unrestricted VAR(p) model for the common factors.[4] For simplicity we fix $p = 1$ and we replace the VECM model (16) with the VAR

$$\mathbf{F}_t = \mathbf{A}_1 \mathbf{F}_{t-1} + \mathbf{w}_t, \quad \mathbf{w}_t = \mathbf{K}\mathbf{u}_t. \tag{23}$$

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

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

i. $\|\widehat{\mathbf{A}}_1^{\text{VAR}} - \mathbf{J}\mathbf{A}_1\mathbf{J}\| = O_p(\max(n^{-1/2}, T^{-1/2}))$;

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

This results can be straightforwardly extended to a generic VAR(p) with coefficients $\widehat{\mathbf{A}}_k^{\text{VAR}}$ such that

$$\widehat{\mathbf{A}}^{\text{VAR}}(L) = \mathbf{I}_r - \sum_{k=1}^{p} \widehat{\mathbf{A}}_k^{\text{VAR}} L^k.$$

---

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

As before, we can compute an estimator $\widehat{\mathbf{R}}$ of the identifying matrix $\mathbf{R}$ by imposing appropriate economic restrictions on the non-identified IRFs. Then, for $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, q$, the estimated and identified IRF of the i-th variable to the j-th shock is defined as

$$\widehat{\phi}_{ij}^{\text{VAR}}(L) = \widehat{\boldsymbol{\lambda}}_i' \left[\widehat{\mathbf{A}}^{\text{VAR}}(L)\right]^{-1} \widehat{\mathbf{K}} \widehat{\mathbf{r}}_j, \tag{24}$$

where $\widehat{\boldsymbol{\lambda}}_i'$ is the i-th row of $\widehat{\boldsymbol{\Lambda}}$, $\widehat{\mathbf{r}}_j$ is the j-th column of $\widehat{\mathbf{R}}$. After denoting as $\widehat{\phi}_{ijk}^{\text{VAR}}$ the k-th coefficient of the polynomial in (24), and as ϕ_{ijk} the corresponding coefficients of $\phi_{ij}(L)$, we have the following consistency result.

Proposition 2 *(Consistency of Impulse Response Functions based on VAR)*
Under Assumptions 1 through 5, as $n, T \to \infty$, we have

$$\left|\widehat{\phi}_{ijk}^{VAR} - \phi_{ijk}\right| = O_p\left(\max\left(n^{-1/2}, T^{-1/2}\right)\right), \tag{25}$$

for any finite $k \geq 0$, $i = 1, 2, \ldots, n$, and $j = 1, 2, \ldots, q$.

Two last remarks are in order.

Remark 6 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). On the other hand, it is also possible to prove that the same unit roots affect the estimated long-run IRFs in such a way that their least squares estimator is no longer 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 2 holds only for finite horizons k.

Remark 7 For any finite k, the estimator $\widehat{\phi}_{ijk}^{\text{VAR}}$ can converge faster than $\widehat{\phi}_{ijk}^{\text{VECM}}$ to the true value ϕ_{ijk}. However, as shown in the proof of Lemma 5, the rate of convergence of the parameters associated to the non-stationary components is slower than what it would be were the factors observed, that is we do not have super-consistency. This is due to the factors' estimation error. Moreover, convergence in Proposition 2 is achieved without the need of Assumption 6. In particular, consistency holds even when all idiosyncratic components are $I(1)$ and without requiring any constraint on the relative rates of divergence of n and T.

Summing up, as a consequence of Propositions 1 and 2, the empirical researcher faces a trade-off between (i) estimating correctly the whole IRFs with a slower rate and more restrictive assumptions, as in Proposition 1, or (ii) giving up consistent estimation of the long-run behavior in exchange for a faster rate of convergence, as in Proposition 2.

3.5 The case of deterministic trends

We conclude this section considering the case of deterministic components. Assumptions 1 and 3 imply $\mathsf{E}[\Delta \mathbf{F}_t] = \mathbf{0}$ and $\mathsf{E}[\Delta \boldsymbol{\xi}_t] = \mathbf{0}$. Because of Assumption 5, we also have $\mathsf{E}[\mathbf{F}_t] = \mathbf{0}$ and $\mathsf{E}[\boldsymbol{\xi}_t] = \mathbf{0}$, which imply that no deterministic components are present in the model for $\mathbf{x}_t$. However, macroeconomic data often have at least a linear trend, in which case the model for an observed time series, denoted as y_{it}, would read

$$y_{it} = a_i + b_i t + \boldsymbol{\lambda}'_i \mathbf{F}_t + \xi_{it}, \tag{26}$$

where $x_{it} = \boldsymbol{\lambda}'_i \mathbf{F}_t + \xi_{it}$ follows the Non-Stationary DFM described by Assumptions 1 through 5. Notice that we also allow for non-zero initial conditions ($a_i \neq 0$), this posing no difficulty in terms of estimation.

The IRFs defined in (18) are then to be considered 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 in (26). This can be done either by de-meaning first differences or by least squares regression, the two approaches respectively giving for $i = 1, 2, \ldots, n$,

$$\widetilde{b}_i = \frac{1}{T} \sum_{t=1}^{T} \Delta y_{it} = \frac{y_{iT} - y_{i0}}{T}, \qquad \widehat{b}_i = \frac{\sum_{t=0}^{T} (t - \frac{T}{2})(y_{it} - \bar{y}_i)}{\sum_{t=0}^{T} (t - \frac{T}{2})^2}. \tag{27}$$

Lemma 6 *Under Assumptions 1 and 3, for any $i = 1, 2, \ldots, n$ 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 $|\widehat{b}_i - b_i| = O_p(T^{-3/2})$.*

Given these results and the rates in Propositions 1 and 2, the IRFs can still be estimated consistently, as described above, also 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 we can estimate the factors as in (15) starting directly from $\widehat{x}_{it} = y_{it} - \widehat{b}_i t$, without integrating $\Delta \widehat{\mathbf{F}}_t$. 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.

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. Hereafter for simplicity of notation we define $\tau = q - d$, the number of shocks with permanent effects.

In light of the results in Lemma 2, we can determine r by using existing methods based on the behaviour of the eigenvalues of the covariance of the variables Δx_{it}. A non-exhaustive list of possible approaches includes the contributions by Bai and Ng (2002), Onatski (2009), Alessi et al. (2010) and Ahn and Horenstein (2013).

In order to determine q and τ, we have instead to study the spectral density matrix of Δx_{it}, $\Delta \chi_{it}$ and $\Delta \xi_{it}$, which are defined by

$$\boldsymbol{\Sigma}^{\Delta x}(\theta) = \boldsymbol{\Sigma}^{\Delta \chi}(\theta) + \boldsymbol{\Sigma}^{\Delta \xi}(\theta) = \frac{1}{2\pi} \boldsymbol{\Lambda} \mathbf{C}(e^{-i\theta}) \mathbf{C}'(e^{i\theta}) \boldsymbol{\Lambda}' + \boldsymbol{\Sigma}^{\Delta \xi}(\theta), \quad \theta \in [-\pi, \pi]. \quad (28)$$

Lemma 7 provides results for the behaviour of the eigenvalues of these matrices.

Lemma 7 *Under Assumptions 1 through 4, for any $n \in \mathbb{N}$, there exist positive reals $\underline{M}_9$, $\overline{M}_9$, M_{10}, $\underline{M}_{11}$, $\overline{M}_{11}$ and an integer $\bar{n}$ such that*
  (i) $\underline{M}_9 \leq n^{-1} \mu_j^{\Delta \chi}(\theta) \leq \overline{M}_9$ a.e. in $[-\pi, \pi]$, and for any $j = 1, 2, \ldots, q$ and $n > \bar{n}$;
  (ii) $\sup_{\theta \in [-\pi, \pi]} \mu_1^{\Delta \xi}(\theta) \leq M_{10}$;
  (iii) $\underline{M}_{11} \leq n^{-1} \mu_j^{\Delta x}(\theta) \leq \overline{M}_{11}$ a.e. in $[-\pi, \pi]$, and for any $j = 1, 2, \ldots, q$ and $n > \bar{n}$;
  (iv) $\sup_{\theta \in [-\pi, \pi]} \mu_{q+1}^{\Delta x}(\theta) \leq M_{10}$;
  (v) $\underline{M}_{12} \leq n^{-1} \mu_j^{\Delta x}(0) \leq \overline{M}_{12}$, for any $j = 1, 2, \ldots, \tau$ and $n > \bar{n}$ and $\mu_{\tau+1}^{\Delta x}(0) \leq M_{10}$.

Parts (i) to (iv) are already known in the literature, but part (v) is a consequence of cointegration in the common components and it is determined by $\mathbf{C}(e^{-i\theta})$ in (28). Indeed, while $\mathrm{rk}(\mathbf{C}(e^{-i\theta})) = q$ a.e. in $[-\pi, \pi]$, this is clearly not true when $\theta = 0$, since, because of the existence of $\tau < q$ common trends, we have $\mathrm{rk}(\mathbf{C}(1)) = \tau$, which in turn implies $\mathrm{rk}(\boldsymbol{\Sigma}^{\Delta \chi}(0)) = \tau$. Part (v) of Lemma 7 is then just a consequence of Weyl's inequality.

Therefore, based on parts (iii) and (iv) of Lemma 7, we can employ the information criterion by Hallin and Liška (2007) to determine q, by analyzing the behaviour of the eigenvalues of the spectral density matrix $\boldsymbol{\Sigma}^{\Delta x}(\theta)$ over a window of frequencies (see also

Onatski, 2010, for a similar approach).[5] Similarly, we propose an information criterion for determining τ based on the behaviour of the eigenvalues of the spectral density matrix $\boldsymbol{\Sigma}^{\Delta x}(\theta)$ only at zero-frequency, as suggested by part (v) of Lemma 7.[6]

In particular, consider the lag-window estimator of the spectral density matrix

$$\widehat{\boldsymbol{\Sigma}}^{\Delta x}(\theta) = \frac{1}{2\pi} \sum_{k=-B_T}^{B_T} \left[\frac{1}{T} \sum_{t=1}^{T-k} \Delta \mathbf{x}_t \Delta \mathbf{x}'_{t+k} \right] e^{-ik\theta} w(B_T^{-1} k)$$

where B_T is a suitable bandwidth and $w(\cdot)$ is a positive even weight function. We define the estimators for q and τ as

$$\widehat{q} = \operatorname*{argmin}_{k=0,\ldots,q_{\max}} \left[\log \left(\frac{1}{n(2B_T + 1)} \sum_{h=-B_T}^{B_T} \sum_{j=k+1}^{n} \widehat{\mu}_j(\theta_h) \right) + ks(n, T) \right], \quad (29)$$

$$\widehat{\tau} = \operatorname*{argmin}_{k=0,\ldots,\tau_{\max}} \left[\log \left(\frac{1}{n} \sum_{j=k+1}^{n} \widehat{\mu}_j(0) \right) + kp(n, T) \right], \quad (30)$$

where $s(n, T)$ and $p(n, T)$ are some suitable penalty functions, $q_{\max}$ and $\tau_{\max}$ are given maximum numbers of common shocks and trends, and $\widehat{\mu}_j^{\Delta x}(\theta)$ are the eigenvalues of $\widehat{\boldsymbol{\Sigma}}^{\Delta x}(\theta)$.

Hallin and Liška (2007) show that under suitable asymptotic conditions on B_T and $s(n, T)$, the number of common shocks is consistently selected, as $n, T \to \infty$. Analogously, we have sufficient conditions for consistency in the selection of the number of common trends.

Proposition 3 (Number of common trends) *Define* $\rho_T = (B_T \log B_T T^{-1})^{-1/2}$ *and assume that*

(i) as $T \to \infty$, $\rho_T \to \infty$ *and* $\rho_T/T \to 0$;

(ii) as $n, T \to \infty$, $p(n, T) \to 0$ *and* $(n\rho_T^{-1})p(n, T) \to \infty$;

Then, under Assumptions 1 through 4, as $n, T \to \infty$, $|\widehat{\tau} - \tau| = o_p(1)$.

Finally, notice that by definition we have $\tau = r - c$ which is the number of unit roots driving the dynamics of the common factors. Therefore, by virtue of Proposition 3, once we

---

[5]Other methods for determining q, not considered in this paper, are proposed by Amengual and Watson (2007) and Bai and Ng (2007). Both require knowing r before determining q.

[6]An alternative approach not considered here is represented by the tests for cointegration in panels with a factor structure, as for example those proposed by Bai and Ng (2004) and Gengenbach et al. (2015). On the other hand applying the classical methods to determine the cointegration rank or the number of common trends might be problematic due to the use of estimated factors as inputs (see e.g. Stock and Watson, 1988, Phillips and Ouliaris, 1988, Johansen, 1991 and Hallin et al., 2016).

determine τ, q, and r, we immediately have an estimates for both the number of transitory shocks $d = q - \tau$ and the cointegration rank $c = r - q + d = r - \tau$.

5 Simulations

We simulate data, from the Non-Stationary DFM with $r = 4$ common factors, and $q = 3$ common shocks, and $\tau = 1$ common trend, 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:

$$x_{it} = \boldsymbol{\lambda}_i' \mathbf{F}_t + \xi_{it}, \qquad i = 1, 2, \ldots, n, \quad t = 1, 2, \ldots, T,$$
$$\mathbf{A}(L)\mathbf{F}_t = \mathbf{K}\mathbf{R}\mathbf{u}_t, \qquad \mathbf{u}_t \overset{w.n.}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{I}_q),$$

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 IRFs, 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 $r \times r$ polynomial matrices with all of their roots outside the unit circle, and $\boldsymbol{\mathcal{M}}(L) = \operatorname{diag}\left((1 - L)\mathbf{I}_{r-c}, \mathbf{I}_c\right)$. 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 $\mathbf{F}_t$ follow a VAR(2) with $r - c$ unit roots, or, equivalently, 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 for all the simulated IRFs: $\phi_{12}(0) = \phi_{13}(0) = \phi_{23}(0) = 0$.

The idiosyncratic components are generated according to the ARMA model (with a 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 $i = 1, 2, \ldots, m$ and $\rho_i = 0$ for $i = m + 1, 2, \ldots, n$, 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 ξ_{it} is rescaled so that it accounts for a third of the variance of the corresponding x_{it}.

The matrices $\boldsymbol{\Lambda}$, $\boldsymbol{\mathcal{U}}_1$, $\mathbf{G}$ and $\mathbf{H}$ are simulated only once so that the set of IRFs to be estimated is always the same, while the vectors of shocks $\mathbf{u}_t$ and $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t} \cdots \varepsilon_{nt})'$, and all the idiosyncratic coefficients 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 IRFs discussed in the previous section, for different cross-sectional and sample sizes (n and T) and for a different numbers (m) of non-stationary idiosyncratic components.

Tables 1 and 2 show Mean Squared Errors (MSE) for the estimated IRFs simulated with different parameter configurations. Estimation is carried out as explained in Section 3. The loadings' and factors' estimators, $\widehat{\boldsymbol{\Lambda}}$ and $\widehat{\mathbf{F}}_t$, are always computed as in (15). Then on $\widehat{\mathbf{F}}_t$ we fit either a VECM as in (16) or an unrestricted VAR as in (23) (in both cases a constant is also included in the estimation). The numbers r, q, and τ are assumed to be known.

Let $\widehat{\phi}_{ijk}^{(h)}$ be the kth coefficient of the estimated IRF of the ith variable to the jth shock at the hth replication when using a VECM or a VAR and let ϕ_{ijk} be the corresponding coefficient of the true simulated IRF defined in (18), then, MSEs are computed with respect to all replications, all variables, and all shocks:

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

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 1 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 when considering a VAR is comparable to, or slightly smaller than, the MSE when considering a VECM. This is in accordance with the result of Propositions 1 and 2 according to which the converge rate for VAR is faster than for VECM. On the other hand, at longer horizons, the MSE for the VAR case is always larger than the MSE for the VECM case. Again this is in accordance with the fact that

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

MSE for the estimated IRFs by fitting a VECM on $\widehat{\mathbf{F}}_t$ as in (16). T is the number of observations, n is the number of variables, and m is the number of $I(1)$ idiosyncratic components.

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

MSE for the estimated IRFs by fitting a VAR on $\widehat{\mathbf{F}}_t$ as in (23). T is the number of observations, n is the number of variables, and m is the number of $I(1)$ idiosyncratic components.

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

Finally, for the same data generating process considered above, we study the performance of the information criterion (30), proposed in Section 4 for determining τ. Table 3 shows the percentage of times in which we estimate correctly the number of common trends

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

Percentage of cases in which the information criteria (29) and (30) returned the correct number of common shocks ($\widehat{q} = q$) and of common trends ($\widehat{\tau} = \tau$). T is the number of observations, n is the number of variables, and m is the number of $I(1)$ idiosyncratic components.

$\tau = 1$. For the sake of comparison, we also report results for the information criterion (29) by Hallin and Liška (2007) for estimating $q = 3$. It has to be noticed that the actual implementation of these criteria requires a procedure of fine tuning of the penalty. Indeed, according to the asymptotic results in Hallin and Liška (2007) and in Proposition 3, for any constant $c > 0$, the functions $c\,s(n,T)$ and $c\,p(n,T)$ are also admissible penalties, and, therefore 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 (30) with other possible methods. Still our results are promising, since 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 DFM 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 B. All variables that are $I(1)$ are not transformed, while we take first differences of those that are $I(2)$. We then remove

deterministic component as described at the end of Section 3, therefore the IRFs presented in this section have to be interpreted as out of trend deviations.

The model is estimated as explained in Section 3. 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 as given by the criterion in Hallin and Liška (2007). Finally, using the information criterion described in Section 4, we find evidence of 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, which 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 Figures 1 and 2, respectively. In both figures the black lines are the IRFs obtained by fitting a VECM on $\widehat{\mathbf{F}}_t$ and the grey lines are the IRFs obtained by fitting an unrestriced VAR on $\widehat{\mathbf{F}}_t$. The dotted black lines and the grey shaded areas are the respective 68% bootstrap confidence bands.

Figure 1 shows the IRFs 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 IRFs, and in particular their long-run behaviour, are consistent with economic theory according to which a monetary policy shock has only a transitory effect on the economy. On the contrary, the IRFs estimated with a stationary DFM, i.e. with data in first differences, 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 IRFs 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 IRFs 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.

Figure 1: IMPULSE RESPONSE FUNCTIONS TO A MONETARY POLICY SHOCK
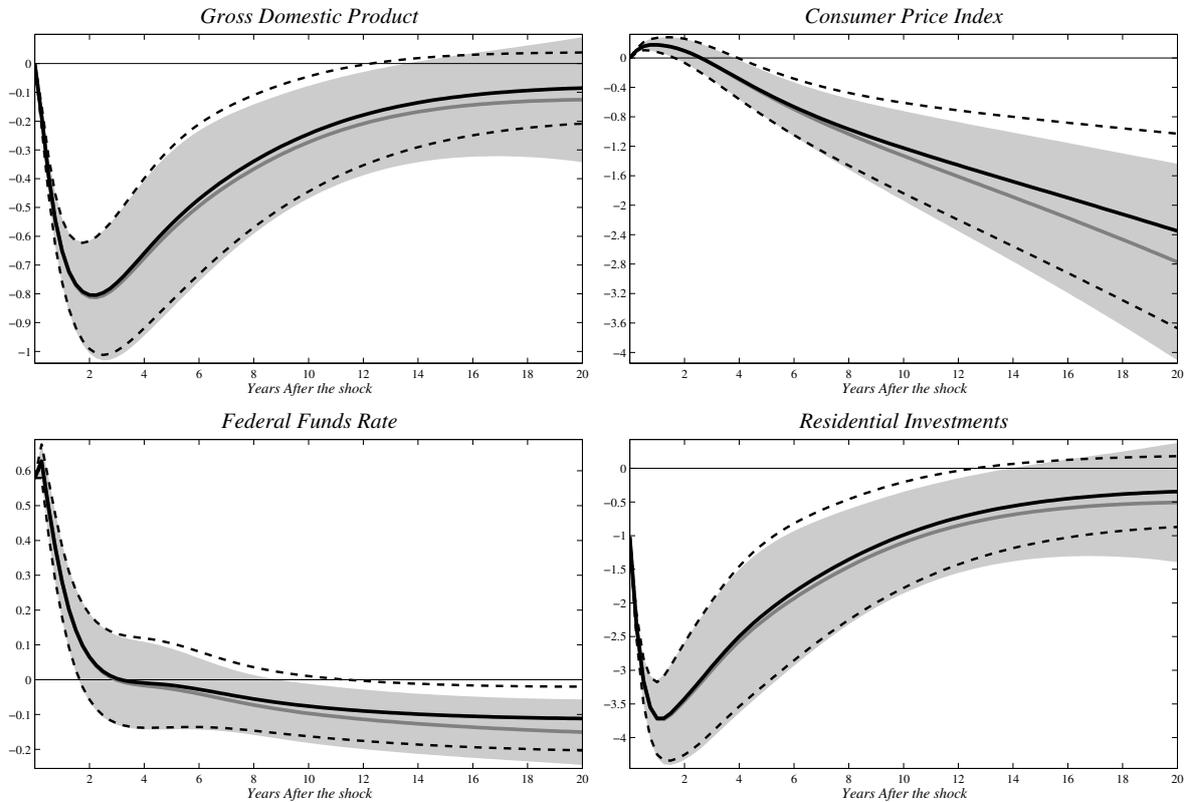
Solid black lines are the IRFs obtained from the Non-Stationary DFM by estimating a VECM on $\widehat{\mathbf{F}}_t$ with 68% bootstrap confidence bands (dashed). Solid grey lines are the IRFs obtained from the Non-Stationary DFM by estimating a VAR on $\widehat{\mathbf{F}}_t$ with 68% confidence bands (shaded areas). The monetary policy shock is normalized so that at impact it increases the Federal Funds rate of 50 basis points.

Differently from the results in Figure 1, while the IRFs obtained using a VECM or an unrestricted VAR show no difference in the short-run, at very long horizons significant differences appear. Notably, the IRFs estimated by fitting an unrestricted VAR tend to diverge. This result is consistent with lack of consistency of long-run IRFs obtained without imposing the presence of unit roots (see Proposition 2). 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 our approach, we find that a supply

shock induces on GDP a permanent deviation of about 0.12% from its historical trend, with a stationary DFM we find a deviation of about 0.67% (not shown here). Finally, notice that similar IRFs are found also in Dedola and Neri (2007) and Smets and Wouters (2007), when employing other estimation techniques.

To summarize, the empirical analysis of this section shows that the proposed Non-Stationary DFM is able to reproduce the main features of the dynamic effects of both temporary and permanent shocks postulated by macroeconomic theory

7 Conclusions

In this paper, we propose a *Non-Stationary Dynamic Factor Model* (DFM) 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 another one where we address representation theory (Barigozzi et al., 2016).

Estimation of impulse response functions (IRFs) is obtained with a two-step estimator based on approximate principal components, and on a VECM—or an unrestriced VAR model—for the latent $I(1)$ common factors. This estimator 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 in a large dimensional setting. A numerical and empirical study show the validity and usefulness of our approach.

The results of this paper are useful beyond estimation of IRFs in Non-Stationary DMFs. First, our estimation approach could also 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). Second, with such goal in mind and given the state-space form of our model, we could think of Quasi Maximum Likelihood estimation (see Doz et al., 2012, for the stationary case), thus allowing us to impose economically relevant restrictions on the model parameters. Third, our asymptotic results could be straightforwardly extended to estimation of IRFs in a non-stationary Factor Augmented VAR setting (see Bai and Ng, 2006, for the stationary case) and form the theoretical foundation of the existing empirical studies based non-stationary factor models (see e.g. Eickmeier, 2009; Banerjee et al., 2017). Last, our approach could be generalized to build an unrestricted Non-Stationary DMF, similar to the one proposed by Forni et al. (2017, 2015) for stationary data.

Figure 2: IMPULSE RESPONSE FUNCTIONS TO A SUPPLY SHOCK
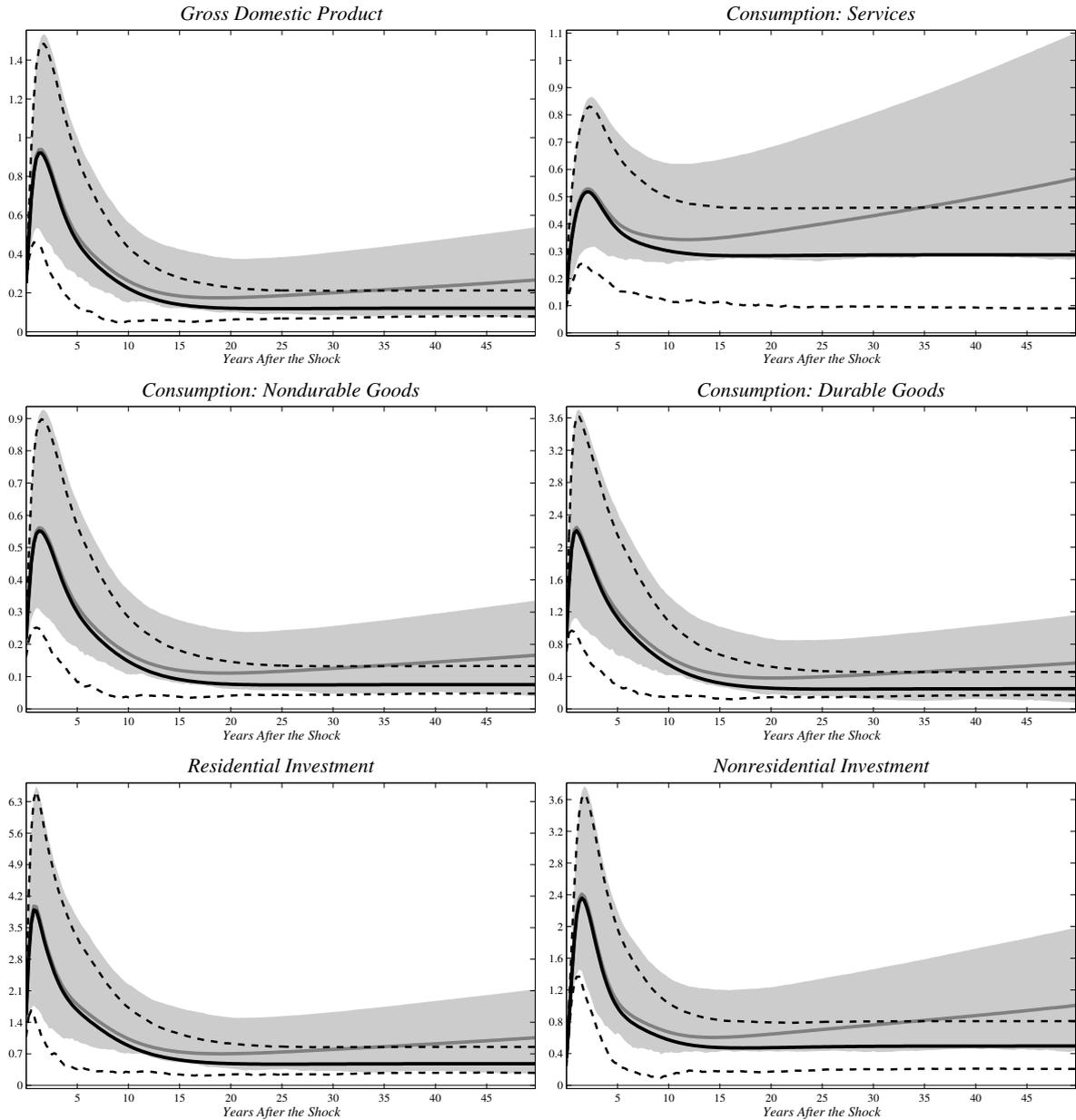
Solid black lines are the IRFs obtained from the Non-Stationary DFM by estimating a VECM on $\widehat{\mathbf{F}}_t$ with 68% bootstrap confidence bands (dashed). Solid grey lines are the IRFs obtained from the Non-Stationary DFM by estimating a VAR on $\widehat{\mathbf{F}}_t$ with 68% confidence bands (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.

31

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.

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. (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.

Banerjee, A., M. Marcellino, and I. Masten (2017). Structural FECM: Cointegration in large-scale structural FAVAR models. *Journal of Applied Econometrics*. https://doi.org/10.1002/jae.2570.

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.

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.

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.

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.

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., 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 (2015). Dynamic factor models with infinite-dimensional factor spaces: One-sided representations. *Journal of Econometrics 185*, 359–371.

Forni, M., M. Hallin, M. Lippi, and P. Zaffaroni (2017). Dynamic factor models with infinite dimensional factor space: Asymptotic analysis. *Journal of Econometrics 199*, 74–92.

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 31*, 982–1004.

Giannone, D., M. Lenza, and G. E. Primicieri (2015). Prior selection for vector autore-

gressions. *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 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.

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. (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.

King, R., C. Plosser, J. H. Stock, and M. W. Watson (1991). Stochastic trends and economic fluctuations. *The 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.

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 (2006). 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. *The 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.

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.

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

A   Technical appendix

Preliminary definitions and notation

Norms. For any $m \times p$ matrix $\mathbf{B}$ with generic element b_{ij}, we denote its spectral norm as $\|\mathbf{B}\| = \sqrt{\mu_1^{\mathbf{B'B}}}$, where $\mu_1^{\mathbf{B'B}}$ is the largest eigenvalue of $\mathbf{B'B}$, the Frobenius norm as $\|\mathbf{B}\|_F = \sqrt{\mathrm{tr}(\mathbf{B'B})} = \sqrt{\sum_i \sum_j b_{ij}^2}$, 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 $\mathbf{B}$:

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

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

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

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

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

Factors' dynamics. It is convenient to write the equations governing the dynamics of the factors, (8), as

$$\Delta F_{jt} = \mathbf{c}_j'(L)\mathbf{u}_t = \sum_{l=1}^{q} c_{jl}(L)u_{lt}, \quad j = 1, \ldots r, \tag{A4}$$

where $\mathbf{c}_j(L)$ is an $q \times 1$ infinite rational polynomial matrix with entries $c_{jl}(L)$. Due to rationality, there exists a positive real K_1 such that

$$\sup_{\substack{j=1,\ldots,r \\ l=1,\ldots,q}} \sum_{k=0}^{\infty} c_{jlk}^2 \le K_1. \tag{A5}$$

From Assumption 5 we also have $F_{jt} = \sum_{s=1}^{t} \mathbf{c}_j'(L)\mathbf{u}_s$.

Idiosyncratic dynamics. Likewise, for the idiosyncratic components it is convenient to write (10) as

$$\Delta \xi_{it} = \check{d}_i(L)\varepsilon_{it}, \quad i = 1, \ldots, n,$$

where $\check{d}_i(L)$ are a infinite polynomials defined as $\check{d}_i(L) = (1-L)(1-\rho_i L)^{-1}d_i(L)$ with $d_i(L)$ also infinite polynomials. Because of Assumption 3(b) there exists a positive real K_2 such that

$$\sup_{i=1,\ldots,n} \sum_{k=0}^{\infty} \check{d}_{ik}^2 \le K_2. \tag{A6}$$

Moreover, with reference to Assumption 6(a) we have $\rho_i = 1$ if $i \in \mathcal{I}_1$ and $|\rho_i| < 1$ if $i \in \mathcal{I}_1^c$. Hence, by Assumptions 5, we have also $\xi_{it} = \sum_{s=1}^{t} \check{d}_i(L)\varepsilon_{is}$, which is non-stationary if and only if $i \in \mathcal{I}_1$.

Rates. We define $\zeta_{nT,\delta} = \max(T^{1/2}n^{-(2-\delta)/2}, n^{-(1-\delta)/2})$, with $\delta \geq 0$, and $\vartheta_{nT,\delta} = \max\left(\zeta_{nT,\delta}, T^{-1/2}\right)$. Under Assumptions 6(a) and 6(b), we have $\zeta_{nT,\delta} \to 0$ and $\vartheta_{nT,\delta} \to 0$, as $n, T \to \infty$.

A.1   Proof of Lemma 1

First notice that, from Assumption 4(e), we have

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

Define $\mathbf{\Gamma}_0^\varepsilon = \mathsf{E}[\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t']$, then Assumption 4(e) reads $\|\mathbf{\Gamma}_0^\varepsilon\|_1 \leq M_4$, thus, from (A2), we have $\mu_1^\varepsilon = \|\mathbf{\Gamma}_0^\varepsilon\| \leq \|\mathbf{\Gamma}_0^\varepsilon\|_1 \leq M_4$. By setting $M_5 = M_4$, we complete the proof.   □

A.2   Proof of Lemma 2

Define $\mathbf{\Gamma}_0^{\Delta F} = \mathsf{E}[\Delta \mathbf{F}_t \Delta \mathbf{F}_t']$. Then, we can write $\mathbf{\Gamma}_0^{\Delta F} = \mathbf{W}^{\Delta F}\mathbf{M}^{\Delta F}\mathbf{W}^{\Delta F\prime}$, 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 $\boldsymbol{L} = \mathbf{\Lambda}\mathbf{W}^{\Delta F}(\mathbf{M}^{\Delta F})^{1/2}$. Under (14), this matrix satisfies Assumption 2(a) since

$$\frac{\boldsymbol{L}'\boldsymbol{L}}{n} = \mathbf{M}^{\Delta F}, \tag{A7}$$

and by Assumption 1(b) and square summability of the coefficients given in (A5), all eigenvalues of $\mathbf{\Gamma}_0^{\Delta F}$ are positive and finite, i.e. there exist positive reals $\underline{M}_6$ and $\overline{M}_6$ such that

$$\underline{M}_6 \leq \mu_j^{\Delta F} \leq \overline{M}_6, \quad j = 1, \ldots, r. \tag{A8}$$

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 (A7), we have for any n, $n^{-1}\mu_j^{\Delta\chi} = \mu_j^{\Delta F}$, for any $j = 1, \ldots, r$. Part (i) then follows from (A8).

As for part (ii), we have

$$\mu_1^{\Delta\xi} = \|\mathbf{\Gamma}_0^{\Delta\xi}\| \leq \sum_{k=0}^{\infty} \|\check{\mathbf{D}}_k\|^2 \|\mathbf{\Gamma}_0^\varepsilon\| \leq K_2 M_4 = M_7, \tag{A9}$$

because of square summability of the coefficients, with K_2 defined in (A6), and from Lemma 1.

Finally, parts (iii) and (iv) are immediate consequences of Assumption 3(d) of uncorrelated common and idiosyncratic shocks, 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, because of parts (i) and (ii), there exist positive reals $\underline{M}_8$ and $\overline{M}_8$, such that, for $j = 1, \ldots, r$, and for any $n \in \mathbb{N}$,

$$\frac{\mu_j^{\Delta x}}{n} \leq \frac{\mu_j^{\Delta \chi}}{n} + \frac{\mu_1^{\Delta \xi}}{n} \leq \overline{M}_6 + \frac{\mu_1^{\Delta \xi}}{n} \leq \overline{M}_6 + \frac{M_7}{n} = \overline{M}_8,$$

$$\frac{\mu_j^{\Delta x}}{n} \geq \frac{\mu_j^{\Delta \chi}}{n} + \frac{\mu_n^{\Delta \xi}}{n} \geq \underline{M}_6 + \frac{\mu_n^{\Delta \xi}}{n} = \underline{M}_8,$$

This proves part (iii). When $j = r + 1$, using parts (i) and (ii), and since $\mathrm{rk}(\boldsymbol{\Gamma}_0^{\Delta \chi}) = r$, we have $\mu_{r+1}^{\Delta x} \leq \mu_{r+1}^{\Delta \chi} + \mu_1^{\Delta \xi} = \mu_1^{\Delta \xi} \leq M_7$, thus proving part (iv). This completes the proof. $\qquad \square$

A.3   Proof of Lemma 3

Intermediate results

Lemma A1 *Define the covariance matrix* $\boldsymbol{\Gamma}_0^{\Delta x} = \mathsf{E}[\Delta \mathbf{x}_t \Delta \mathbf{x}_t']$ *with generic* (i,j)-*th element* $\gamma_{ij}^{\Delta x} = \mathsf{E}[\Delta x_{it} \Delta x_{jt}]$. *Then, under Assumptions 1 through 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* $i, j = 1, \ldots, n$.

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

$$\mathsf{E}\left[\frac{1}{T} \sum_{t=1}^T \Delta \mathbf{F}_t \Delta \mathbf{F}_t'\right] = \frac{1}{T} \sum_{t=1}^T \mathsf{E}\left[\left(\sum_{k=0}^\infty \mathbf{C}_k \mathbf{u}_{t-k}\right)\left(\sum_{k'=0}^\infty \mathbf{C}_{k'} \mathbf{u}_{t-k'}\right)'\right] = \sum_{k=0}^\infty \mathbf{C}_k \mathbf{C}_k' = \boldsymbol{\Gamma}_0^{\Delta F}, \quad (A10)$$

where we used Assumption 1(a) which implies that $\mathbf{u}_t$ is a white noise. Moreover, $\mathrm{rk}(\boldsymbol{\Gamma}_0^{\Delta F}) = r$ because of Assumption 1(b), and $\|\boldsymbol{\Gamma}_0^{\Delta F}\| = O(1)$ because of square summability of the coefficients given in (A5). Hence, $\boldsymbol{\Gamma}_0^{\Delta F}$ 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 3(d) of uncorrelated common and idiosyncratic shocks, $\mathsf{E}[T^{-1} \sum_{t=1}^T \Delta x_{it} \Delta x_{it}] = \gamma_{ij}^{\Delta x}$.

Now, denote as $\gamma_{ij}^{\Delta F}$ the generic (i,j)-th element of $\boldsymbol{\Gamma}_0^{\Delta F}$. Then, from (A2) and (A4),

$$
\mathsf{E}\bigg[\bigg\|\frac{1}{T}\sum_{t=1}^{T}\Delta\mathbf{F}_t\Delta\mathbf{F}_t' - \boldsymbol{\Gamma}_0^{\Delta F}\bigg\|^2\bigg] \leq \sum_{i,j=1}^{r}\frac{1}{T^2}\mathsf{E}\bigg[\sum_{t,s=1}^{T}\Big(\Delta F_{it}\Delta F_{jt}-\gamma_{ij}^{\Delta F}\Big)\Big(\Delta F_{is}\Delta F_{js}-\gamma_{ij}^{\Delta F}\Big)\bigg]
$$

$$
= \sum_{i,j=1}^{r}\frac{1}{T^2}\sum_{t,s=1}^{T}\Big(\mathsf{E}\big[\Delta F_{it}\Delta F_{jt}\Delta F_{is}\Delta F_{js}\big]-(\gamma_{ij}^{\Delta F})^2\Big)
$$

$$
\leq \frac{r^2K_1^4q^4}{T^2}\sum_{t,s=1}^{T}\mathsf{E}[u_{lt}u_{l't}u_{hs}u_{h's}] - \frac{r^2K_1^4q^4}{T^2}\sum_{t,s=1}^{T}(\mathsf{E}[u_{lt}u_{l't}])^2
$$

$$
= \frac{r^2K_1^4q^4}{T^2}\sum_{t,s=1}^{T}\mathsf{E}[u_{lt}^2]\mathsf{E}[u_{hs}^2] + \frac{r^2K_1^4q^4}{T^2}\sum_{t=1}^{T}\mathsf{E}[u_{lt}^2u_{ht}^2] + \frac{r^2K_1^4q^4}{T^2}\sum_{t=1}^{T}\mathsf{E}[u_{lt}^4] - \frac{r^2K_1^4q^4}{T^2}\sum_{t,s=1}^{T}(\mathsf{E}[u_{lt}^2])^2
$$

$$
= \frac{r^2K_1^4q^4}{T^2}\sum_{t=1}^{T}\mathsf{E}[u_{lt}^2]\mathsf{E}[u_{ht}^2] = \frac{r^2K_1^4q^4}{T} = O\left(\frac{1}{T}\right), \tag{A11}
$$

where we used Assumption 4(a) of independence of $\mathbf{u}_t$ and Assumption 4(b) of existence of fourth moments, plus square summability of the coefficients, with K_1 defined in (A5). Therefore, from (A11), we have

$$
\bigg\|\frac{1}{T}\sum_{t=1}^{T}\Delta\mathbf{F}_t\Delta\mathbf{F}_t' - \boldsymbol{\Gamma}_0^{\Delta F}\bigg\| = O_p\left(\frac{1}{\sqrt{T}}\right). \tag{A12}
$$

In the same way, for the idiosyncratic component we have

$$
\mathsf{E}\bigg[\bigg\|\frac{1}{T}\sum_{t=1}^{T}\Delta\xi_{it}\Delta\xi_{jt} - \gamma_{ij}^{\Delta\xi}\bigg\|^2\bigg] \leq \frac{1}{T^2}\sum_{t,s=1}^{T}\Big(\mathsf{E}\big[\Delta\xi_{it}\Delta\xi_{jt}\Delta\xi_{is}\Delta\xi_{js}\big]-(\gamma_{ij}^{\Delta\xi})^2\Big)
$$

$$
\leq \frac{K_2^4}{T^2}\sum_{t=1}^{T}\mathsf{E}[\varepsilon_{it}^2\varepsilon_{jt}^2] \leq \frac{K_2^4 M_3}{T} = O\left(\frac{1}{T}\right), \tag{A13}
$$

where we used Assumption 4(c) of independence of $\boldsymbol{\varepsilon}_t$ and Assumption 4(d) of existence of fourth moments and square summability of the coefficients, with K_2 defined in (A6). Therefore, from (A11), we have

$$
\bigg\|\frac{1}{T}\sum_{t=1}^{T}\Delta\xi_{it}\Delta\xi_{jt} - \gamma_{ij}^{\Delta\xi}\bigg\| = O_p\left(\frac{1}{\sqrt{T}}\right). \tag{A14}
$$

By combining (A12) and (A14) and Assumption 2(b) of bounded loadings we complete the proof.$\square$

Lemma A2 *For any given t, under Assumptions 1 through 6 and as $n, T \to \infty$,*
  (i) $\|\Delta\mathbf{F}_t\| = O_p(1)$;
  (ii) $\|T^{-1/2}\mathbf{F}_t\| = O_p(1)$;
  (iii) $\|n^{-1/2}\Delta\boldsymbol{\xi}_t\| = O_p(1)$;
  (iv) $\|(nT)^{-1/2}\boldsymbol{\xi}_t\| = O_p(1)$;
  (v) $\|n^{-1/2}\boldsymbol{\Lambda}'\Delta\boldsymbol{\xi}_t\| = O_p(1)$;

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

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

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

Proof. For part (i), just notice that, since by Assumption 1(a) $\Delta F_{jt} \sim I(0)$ for any $j = 1, \ldots, r$, then they have finite variance. This proves part (i) by Chebychev's inequality.

For part (ii), we have

$$
\mathsf{E}\left[\left\|\frac{\mathbf{F}_t}{\sqrt{T}}\right\|^2\right] = \frac{1}{T}\sum_{j=1}^{r}\mathsf{E}\big[F_{jt}^2\big] = \frac{1}{T}\sum_{j=1}^{r}\mathsf{E}\left[\left(\sum_{s=1}^{t}\sum_{l=1}^{q}c_{jl}(L)u_{ls}\right)^2\right]
$$
$$
= \frac{1}{T}\sum_{j=1}^{r}\sum_{s,s'=1}^{t}\sum_{l,l'=1}^{q}\sum_{k,k'=0}^{\infty}c_{jlk}c_{jl'k'}\mathsf{E}[u_{ls-k}u_{l's'-k'}] \leq \frac{rqK_1 t}{T} \leq rqK_1, \tag{A15}
$$

since $t \leq T$ and where we used the fact $\mathbf{u}_t$ is a white noise because of Assumption 1(a) and we used square summability of the coefficients, with K_1 defined in (A5). This proves part (ii).

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

$$
\mathsf{E}\left[\left\|\frac{\Delta\boldsymbol{\xi}_t}{\sqrt{n}}\right\|^2\right] = \frac{1}{n}\sum_{i=1}^{n}\mathsf{E}\big[\Delta\xi_{it}^2\big] = \frac{1}{n}\sum_{i=1}^{n}\mathsf{E}[(\check{d}_i(L)\varepsilon_{it})^2]
$$
$$
= \frac{1}{n}\sum_{i=1}^{n}\sum_{k,k'=0}^{\infty}\check{d}_{jk}\check{d}_{ik'}\mathsf{E}[\varepsilon_{it-k}\varepsilon_{it-k'}] \leq K_2\max_{i}\mathsf{E}[\varepsilon_{it}^2], \tag{A16}
$$

where we used Assumption 3(a) of serially uncorrelated $\boldsymbol{\varepsilon}_t$ and square summability of the coefficients, with K_2 defined in (A6). Also because of the existence of fourth moments in Assumption 4(d) the variance of ε_{it} is finite for any i. This proves part (iii).

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

$$
\mathsf{E}\left[\left\|\frac{\boldsymbol{\xi}_t}{\sqrt{nT}}\right\|^2\right] = \frac{1}{nT}\sum_{i=1}^{n}\mathsf{E}\big[\xi_{it}^2\big] = \frac{1}{nT}\sum_{i=1}^{n}\mathsf{E}\left[\left(\sum_{s=1}^{t}\check{d}_i(L)\varepsilon_{is}\right)^2\right]
$$
$$
= \frac{1}{nT}\sum_{i=1}^{n}\sum_{s,s'=1}^{t}\sum_{k,k'=0}^{\infty}\check{d}_{ik}\check{d}_{ik'}\mathsf{E}[\varepsilon_{is-k}\varepsilon_{is'-k'}] \leq \frac{K_2 t}{T}\max_{i}\mathsf{E}[\varepsilon_{it}^2] \leq K_2\max_{i}\mathsf{E}[\varepsilon_{it}^2], \tag{A17}
$$

since $t \leq T$ and where we used the same assumptions as in (A16). This proves part (iv).

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

$$
\mathsf{E}\left[\left\|\frac{\mathbf{\Lambda}'\Delta\boldsymbol{\xi}_t}{\sqrt{n}}\right\|^2\right] = \frac{1}{n}\sum_{j=1}^{r}\mathsf{E}\left[\left(\sum_{i=1}^{n}\lambda_{ij}\Delta\xi_{it}\right)^2\right] = \frac{1}{n}\sum_{j=1}^{r}\sum_{i,l=1}^{n}\mathsf{E}\big[\lambda_{ij}\Delta\xi_{it}\lambda_{lj}\Delta\xi_{lt}\big]
$$
$$
\leq \frac{rC^2}{n}\sum_{i,l=1}^{n}\sum_{k,k'=0}^{\infty}\check{d}_{ik}\check{d}_{lk'}\mathsf{E}[\varepsilon_{it-k}\varepsilon_{lt-k'}] \leq \frac{rC^2 K_2}{n}\sum_{i,l=1}^{n}\big|\mathsf{E}[\varepsilon_{it}\varepsilon_{lt}]\big| \leq rC^2 K_2 M_4, \tag{A18}
$$

where we used the same assumptions as in (A16), Assumption 2(b) of bounded loadings, and Lemma 1. This proves part (v).

40

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

$$
\mathsf{E}\left[\left\|\frac{\boldsymbol{\Lambda}'\boldsymbol{\xi}_t}{\sqrt{nT}}\right\|^2\right] = \frac{1}{nT}\sum_{j=1}^{r}\mathsf{E}\left[\left(\sum_{i=1}^{n}\lambda_{ij}\xi_{it}\right)^2\right] = \frac{1}{nT}\sum_{j=1}^{r}\sum_{i,l=1}^{n}\mathsf{E}\left[\lambda_{ij}\xi_{it}\lambda_{lj}\xi_{lt}\right]
$$

$$
\leq \frac{rC^2}{nT}\sum_{i,l=1}^{n}\sum_{s,s'=1}^{t}\sum_{k,k'=0}^{\infty}\check{d}_{ik}\check{d}_{lk'}\mathsf{E}[\varepsilon_{is-k}\varepsilon_{ls'-k'}] \leq \frac{rC^2 K_2 t}{nT}\sum_{i,l=1}^{n}\left|\mathsf{E}[\varepsilon_{it}\varepsilon_{lt}]\right| \leq rC^2 K_2 M_4, \qquad (A19)
$$

where we used the same assumptions as in (A18). This proves part (vi).

Now consider part (vii). Using Assumption 6(a), for any $n \in \mathbb{N}$, we can write

$$
\mathsf{E}\left[\left\|\frac{\boldsymbol{\xi}_t}{\sqrt{n}}\right\|^2\right] = \frac{1}{n}\sum_{i\in\mathcal{I}_1}\mathsf{E}\left[\xi_{it}^2\right] + \frac{1}{n}\sum_{i\in\mathcal{I}_1^c}\mathsf{E}\left[\xi_{it}^2\right]. \qquad (A20)
$$

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 (iii). For the first term on the rhs, using Assumption 6(a) and part (iv), we have (multiply and divide by m)

$$
\frac{1}{n}\sum_{i\in\mathcal{I}_1}\mathsf{E}\left[\xi_{it}^2\right] \leq \frac{K_2 Tm}{n}\max_i\mathsf{E}[\varepsilon_{it}^2] = O\left(\frac{T}{n^{1-\delta}}\right), \qquad (A21)
$$

which proves part (vii).

Finally, for part (viii), using the same reasoning as for part (vii), we can write

$$
\mathsf{E}\left[\left\|\frac{\boldsymbol{\Lambda}'\boldsymbol{\xi}_t}{\sqrt{n}}\right\|^2\right] = \frac{1}{n}\sum_{j=1}^{r}\sum_{i,l=1}^{n}\mathsf{E}\left[\lambda_{ij}\xi_{it}\lambda_{lj}\xi_{lt}\right]
$$

$$
= \frac{1}{n}\sum_{j=1}^{r}\sum_{i,l\in\mathcal{I}_1}\mathsf{E}\left[\lambda_{ij}\xi_{it}\lambda_{lj}\xi_{lt}\right] + \frac{1}{n}\sum_{j=1}^{r}\sum_{i,l\in\mathcal{I}_1^c}\mathsf{E}\left[\lambda_{ij}\xi_{it}\lambda_{lj}\xi_{lt}\right] + \frac{2}{n}\sum_{j=1}^{r}\sum_{i\in\mathcal{I}_1}\sum_{l\in\mathcal{I}_1^c}\mathsf{E}\left[\lambda_{ij}\xi_{it}\lambda_{lj}\xi_{lt}\right]. \qquad (A22)
$$

The second term on the rhs is bounded because it is a sum of products of stationary components as in (A18) and therefore it behaves as part (v). For the first term on the rhs, using Assumption 6(a) and part (iv), we have (multiply and divide by m)

$$
\frac{1}{n}\sum_{j=1}^{r}\sum_{i,l\in\mathcal{I}_1}\mathsf{E}\left[\lambda_{ij}\xi_{it}\lambda_{lj}\xi_{lt}\right] \leq \frac{rC^2 K_2 T}{n}\sum_{i,l\in\mathcal{I}_1}\left|\mathsf{E}[\varepsilon_{it}\varepsilon_{lt}]\right| \leq \frac{rC^2 K_2 M_4 Tm}{n} = O\left(\frac{T}{n^{1-\delta}}\right). \qquad (A23)
$$

Similarly, the third term on the rhs of (A22) is bounded as follows

$$
\frac{1}{n}\sum_{j=1}^{r}\sum_{i\in\mathcal{I}_1}\sum_{l\in\mathcal{I}_1^c}\mathsf{E}\left[\lambda_{ij}\xi_{it}\lambda_{lj}\xi_{lt}\right] \leq \frac{rC^2 K_2 T}{n}\sum_{i\in\mathcal{I}_1}\sum_{l\in\mathcal{I}_1^c}\left|\mathsf{E}[\varepsilon_{it}\varepsilon_{lt}]\right| \leq \frac{rC^2 K_2 M_9 Tn^\gamma}{n} = O\left(\frac{T}{n^{1-\gamma}}\right).
$$
$$(A24)$$

We prove part (viii) by substituting (A23) and (A24) into (A22), and by noticing that (A23) converges to zero slower than (A24) because $\gamma < \delta$ by Assumption 6(c). This completes the proof.

41

$\square$

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 3(d) of uncorrelated common and idiosyncratic components, we have $\mathbf{\Gamma}_0^{\Delta x} = \mathbf{\Gamma}_0^{\Delta\chi} + \mathbf{\Gamma}_0^{\Delta\xi}$. Moreover, from Lemma A1, we have

$$\left\|\frac{\widehat{\mathbf{\Gamma}}_0^{\Delta x}}{n} - \frac{\mathbf{\Gamma}_0^{\Delta x}}{n}\right\| = O_p\left(\frac{1}{\sqrt{T}}\right). \tag{A25}$$

From (A25), Lemma 2(ii) and Assumption 3(d) we also have

$$\left\|\frac{\widehat{\mathbf{\Gamma}}_0^{\Delta x}}{n} - \frac{\mathbf{\Gamma}_0^{\Delta\chi}}{n}\right\| \leq \left\|\frac{\widehat{\mathbf{\Gamma}}_0^{\Delta x}}{n} - \frac{\mathbf{\Gamma}_0^{\Delta x}}{n}\right\| + \left\|\frac{\mathbf{\Gamma}_0^{\Delta x}}{n} - \frac{\mathbf{\Gamma}_0^{\Delta\chi}}{n}\right\| = \left\|\frac{\widehat{\mathbf{\Gamma}}_0^{\Delta x}}{n} - \frac{\mathbf{\Gamma}_0^{\Delta x}}{n}\right\| + \left\|\frac{\mathbf{\Gamma}_0^{\Delta\xi}}{n}\right\|$$

$$= O_p\left(\frac{1}{\sqrt{T}}\right) + \frac{\mu_1^{\Delta\xi}}{n} \leq O_p\left(\frac{1}{\sqrt{T}}\right) + \frac{M_7}{n} = O_p\left(\max\left(\frac{1}{\sqrt{T}}, \frac{1}{n}\right)\right). \tag{A26}$$

Now, define the $n \times r$ matrices $\mathbf{W}^{\Delta\chi}$ and $\widehat{\mathbf{W}}^{\Delta x}$, having as columns the normalized eigenvectors corresponding to the j-th largest eigenvalues of $\mathbf{\Gamma}_0^{\Delta\chi}$ and $\widehat{\mathbf{\Gamma}}_0^{\Delta x}$, respectively. From Theorem 2 in Yu et al. (2015), which is a consequence of the "$\sin\theta$" Theorem in Davis and Kahan (1970), we have

$$\left\|\widehat{\mathbf{W}}^{\Delta x} - \mathbf{W}^{\Delta\chi}\mathbf{J}\right\| = \frac{2^{3/2}\sqrt{r}\|\widehat{\mathbf{\Gamma}}_0^{\Delta x} - \mathbf{\Gamma}_0^{\Delta\chi}\|}{\min\left(\mu_0^{\Delta\chi} - \mu_1^{\Delta\chi}, \mu_r^{\Delta\chi} - \mu_{r+1}^{\Delta\chi}\right)}, \tag{A27}$$

where $\mathbf{J}$ is a diagonal $r \times r$ matrix with entries ± 1 and we define $\mu_0^{\Delta\chi} = \infty$ for any $n \in \mathbb{N}$. Since $\mu_{r+1}^{\Delta\chi} = 0$ then, from Lemma 2(i) and (A26), we have

$$\left\|\widehat{\mathbf{W}}^{\Delta x} - \mathbf{W}^{\Delta\chi}\mathbf{J}\right\| \leq \frac{2^{3/2}\sqrt{r}\|\widehat{\mathbf{\Gamma}}_0^{\Delta x} - \mathbf{\Gamma}_0^{\Delta\chi}\|}{n\underline{M}_6} = O_p\left(\max\left(\frac{1}{\sqrt{T}}, \frac{1}{n}\right)\right). \tag{A28}$$

Moreover, given the identification constraint (13), we identify the first difference of the factors (up to a sign) as the r (non-normalized) principal components of the common component vector:

$$\Delta\mathbf{F}_t = \frac{1}{\sqrt{n}}\mathbf{W}^{\Delta\chi'}\Delta\boldsymbol{\chi}_t = \frac{1}{\sqrt{n}}\mathbf{W}^{\Delta\chi'}\mathbf{\Lambda}\Delta\mathbf{F}_t. \tag{A29}$$

Since the eigenvectors are normalized, (A29) implies $\mathbf{\Lambda} = n^{1/2}\mathbf{W}^{\Delta\chi}$, such that (14) is satisfied for any $n \in \mathbb{N}$, while the loadings estimator is defined as $\widehat{\mathbf{\Lambda}} = n^{1/2}\widehat{\mathbf{W}}^{\Delta x}$, and therefore $n^{-1}\widehat{\mathbf{\Lambda}}'\widehat{\mathbf{\Lambda}} = \mathbf{I}_r$. By substituting these expressions for $\mathbf{\Lambda}$ and $\widehat{\mathbf{\Lambda}}$ in (A28), we have

$$\left\|\widehat{\mathbf{W}}^{\Delta x} - \mathbf{W}^{\Delta\chi}\mathbf{J}\right\| = \left\|\frac{\widehat{\mathbf{\Lambda}} - \mathbf{\Lambda}\mathbf{J}}{\sqrt{n}}\right\| = O_p\left(\max\left(\frac{1}{\sqrt{T}}, \frac{1}{n}\right)\right), \tag{A30}$$

and also

$$\left\|\frac{\widehat{\mathbf{\Lambda}}'\mathbf{\Lambda}}{n} - \mathbf{J}\right\| = O_p\left(\max\left(\frac{1}{\sqrt{T}}, \frac{1}{n}\right)\right). \tag{A31}$$

In order to prove part (i), 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,

$$
\begin{aligned}
\left\| \frac{\boldsymbol{\epsilon}_i'}{\sqrt{n}} (\widehat{\boldsymbol{\Gamma}}_0^{\Delta x} - \boldsymbol{\Gamma}_0^{\Delta \chi}) \right\| &\leq \left\| \frac{\boldsymbol{\epsilon}_i'}{\sqrt{n}} (\widehat{\boldsymbol{\Gamma}}_0^{\Delta x} - \boldsymbol{\Gamma}_0^{\Delta x}) \right\| + \left\| \frac{\boldsymbol{\epsilon}_i' \boldsymbol{\Gamma}_0^{\Delta \xi}}{\sqrt{n}} \right\| \leq \left\| \frac{\boldsymbol{\epsilon}_i'}{\sqrt{n}} (\widehat{\boldsymbol{\Gamma}}_0^{\Delta x} - \boldsymbol{\Gamma}_0^{\Delta x}) \right\|_F + \left\| \frac{\boldsymbol{\epsilon}_i' \boldsymbol{\Gamma}_0^{\Delta \xi}}{\sqrt{n}} \right\| \\
&\leq \sqrt{\frac{1}{n} \sum_{j=1}^n (\widehat{\gamma}_{ij}^{\Delta x} - \gamma_{ij}^{\Delta x})^2} + \frac{\mu_1^{\Delta \xi}}{\sqrt{n}} \leq O_p \left(\frac{1}{\sqrt{T}} \right) + \frac{M_7}{\sqrt{n}} = O_p \left(\max \left(\frac{1}{\sqrt{T}}, \frac{1}{\sqrt{n}} \right) \right),
\end{aligned}
$$

(A32)

where we used Lemmas A1 and 2(ii). Similarly, we can show that

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

(A33)

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

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

(A34)

From Lemma 2(i) and (A34), we also have

$$
\frac{\mu_r^{\Delta \chi}}{n} \geq \underline{M}_6 > 0, \qquad \frac{\widehat{\mu}_r^{\Delta x}}{n} \geq \underline{M}_6 + O_p \left(\max \left(\frac{1}{\sqrt{T}}, \frac{1}{n} \right) \right).
$$

(A35)

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 (A35), 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)

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

(A36)

Moreover, from (A34) and (A35), we have

$$
\begin{aligned}
\left\| \left(\frac{\widehat{\mathbf{M}}^{\Delta x}}{n} \right)^{-1} - \left(\frac{\mathbf{M}^{\Delta \chi}}{n} \right)^{-1} \right\| &\leq \left\| \left(\frac{\widehat{\mathbf{M}}^{\Delta x}}{n} \right)^{-1} - \left(\frac{\mathbf{M}^{\Delta \chi}}{n} \right)^{-1} \right\|_F = \sqrt{\sum_{j=1}^r \left(\frac{n}{\widehat{\mu}_j^{\Delta x}} - \frac{n}{\mu_j^{\Delta \chi}} \right)^2} \\
&\leq \sum_{j=1}^r n \left| \frac{\widehat{\mu}_j^{\Delta x} - \mu_j^{\Delta \chi}}{\widehat{\mu}_j^{\Delta x} \mu_j^{\Delta \chi}} \right| \leq \frac{r \max_{1 \leq j \leq r} |\widehat{\mu}_j^{\Delta x} - \mu_j^{\Delta \chi}|}{n \underline{M}_6^2 + O_p \left(\max \left(\frac{n}{\sqrt{T}}, 1 \right) \right)} = O_p \left(\max \left(\frac{1}{\sqrt{T}}, \frac{1}{n} \right) \right).
\end{aligned}
$$

(A37)

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 (A28), (A32), (A33), (A36), and (A37), for a given i

43

we have

$$\left\| \sqrt{n}\boldsymbol{\epsilon}_i'\widehat{\mathbf{W}}^{\Delta x} - \sqrt{n}\boldsymbol{\epsilon}_i'\mathbf{W}^{\Delta\chi}\mathbf{J}\right\| = \left\| \frac{\boldsymbol{\epsilon}_i'}{\sqrt{n}}\left[\widehat{\boldsymbol{\Gamma}}_0^{\Delta x}\widehat{\mathbf{W}}^{\Delta x}\left(\frac{\widehat{\mathbf{M}}^{\Delta x}}{n}\right)^{-1} - \boldsymbol{\Gamma}_0^{\Delta\chi}\mathbf{W}^{\Delta\chi}\mathbf{J}\left(\frac{\mathbf{M}^{\Delta\chi}}{n}\right)^{-1}\right]\right\|$$

$$\leq \left\|\frac{\boldsymbol{\epsilon}_i'}{\sqrt{n}}(\widehat{\boldsymbol{\Gamma}}_0^{\Delta x} - \boldsymbol{\Gamma}_0^{\Delta\chi})\right\|\left\|\left(\frac{\mathbf{M}^{\Delta\chi}}{n}\right)^{-1}\right\| + \left\|\frac{\boldsymbol{\epsilon}_i'\boldsymbol{\Gamma}_0^{\Delta\chi}}{\sqrt{n}}\right\|\left\|\left(\frac{\widehat{\mathbf{M}}^{\Delta x}}{n}\right)^{-1} - \left(\frac{\mathbf{M}^{\Delta\chi}}{n}\right)^{-1}\right\|$$

$$+ \left\|\widehat{\mathbf{W}}^{\Delta x} - \mathbf{W}^{\Delta\chi}\mathbf{J}\right\|\left\|\frac{\boldsymbol{\epsilon}_i'\boldsymbol{\Gamma}_0^{\Delta\chi}}{\sqrt{n}}\right\|\left\|\left(\frac{\mathbf{M}^{\Delta\chi}}{n}\right)^{-1}\right\| + o_p\left(\max\left(\frac{1}{\sqrt{T}},\frac{1}{\sqrt{n}}\right)\right)$$

$$= O_p\left(\max\left(\frac{1}{\sqrt{T}},\frac{1}{\sqrt{n}}\right)\right).$$

By noticing that $\boldsymbol{\lambda}_i' = \sqrt{n}\boldsymbol{\epsilon}_i'\mathbf{W}^{\Delta\chi}$ and $\widehat{\boldsymbol{\lambda}}_i' = \sqrt{n}\boldsymbol{\epsilon}_i'\widehat{\mathbf{W}}^{\Delta x}$, we complete the proof of part (i).

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, for a given t,

$$\left\|\Delta\widehat{\mathbf{F}}_t - \mathbf{J}\Delta\mathbf{F}_t\right\| = \left\|\frac{\widehat{\boldsymbol{\Lambda}}'\Delta\mathbf{x}_t}{n} - \mathbf{J}\Delta\mathbf{F}_t\right\| \leq \left\|\frac{\widehat{\boldsymbol{\Lambda}}'\boldsymbol{\Lambda}\Delta\mathbf{F}_t}{n} - \mathbf{J}\Delta\mathbf{F}_t\right\| + \left\|\frac{\widehat{\boldsymbol{\Lambda}}'\Delta\boldsymbol{\xi}_t}{n}\right\|$$

$$\leq \left\|\frac{\widehat{\boldsymbol{\Lambda}}'\boldsymbol{\Lambda}}{n} - \mathbf{J}\right\|\left\|\Delta\mathbf{F}_t\right\| + \left\|\frac{\widehat{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda}\mathbf{J}}{\sqrt{n}}\right\|\left\|\frac{\Delta\boldsymbol{\xi}_t}{\sqrt{n}}\right\| + \left\|\frac{\boldsymbol{\Lambda}'\Delta\boldsymbol{\xi}_t}{n}\right\|\left\|\mathbf{J}\right\|$$

$$= O_p\left(\max\left(\frac{1}{\sqrt{T}},\frac{1}{\sqrt{n}}\right)\right),$$

where we used (A30), (A31), and Lemma A2(i), A2(iii) and A2(v). Obviously $\|\mathbf{J}\| = 1$. This proves part (ii).

Similarly, for part (iii), for a given t we have

$$\frac{1}{\sqrt{T}}\left\|\widehat{\mathbf{F}}_t - \mathbf{J}\mathbf{F}_t\right\| \leq \left\|\frac{\widehat{\boldsymbol{\Lambda}}'\boldsymbol{\Lambda}}{n} - \mathbf{J}\right\|\left\|\frac{\mathbf{F}_t}{\sqrt{T}}\right\| + \left\|\frac{\widehat{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda}\mathbf{J}}{\sqrt{n}}\right\|\left\|\frac{\boldsymbol{\xi}_t}{\sqrt{nT}}\right\| + \left\|\frac{\boldsymbol{\Lambda}'\boldsymbol{\xi}_t}{n\sqrt{T}}\right\|\left\|\mathbf{J}\right\|$$

$$= O_p\left(\max\left(\frac{1}{\sqrt{T}},\frac{1}{\sqrt{n}}\right)\right),$$

where we used (A30), (A31), and Lemma A2(ii), A2(iv) and A2(vi). This completes the proof. $\square$

A.4   Proof of Lemma 4

Intermediate results

Lemma A3 *Under Assumptions 1 through 6, as $n, T \to \infty$,*

$$\left\|\frac{\widehat{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda}\mathbf{J}}{\sqrt{n}}\right\| = O_p\left(\frac{1}{\sqrt{T}}\right) \quad and \quad \left\|\frac{\widehat{\boldsymbol{\Lambda}}'\boldsymbol{\Lambda}}{n} - \mathbf{J}\right\| = O_p\left(\frac{1}{\sqrt{T}}\right).$$

Proof. Under Assumption 6(b), we have $n > T^{1/(2-\delta)}$ with $\delta \geq 0$, the lower bound for n being $n > T^{1/2}$, and, therefore, (A30) and (A31) are both $O_p(T^{-1/2})$. This completes the proof. $\square$

Lemma A4 *Under Assumptions 1 and 5*

(i) The factors admit the common trends decomposition

$$\mathbf{F}_t = \mathbf{C}(1) \sum_{s=1}^{t} \mathbf{u}_s + \check{\mathbf{C}}(L)\mathbf{u}_t = \boldsymbol{\psi}\boldsymbol{\eta}' \sum_{s=1}^{t} \mathbf{u}_s + \check{\mathbf{C}}(L)\mathbf{u}_t,$$

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 rational polynomial matrix also with square summable coefficients. The $r \times c$ cointegration matrix $\boldsymbol{\beta}$ is such that $\boldsymbol{\beta}'\mathbf{C}(1) = \mathbf{0}_{c \times q}$.

(ii) For a given t, as $n, T \to \infty$, $\|\boldsymbol{\beta}'\mathbf{F}_t\| = O_p(1)$.

Proof. The proof follows Lemma 2.1 in Phillips and Solo (1992). Using the Beveridge-Nelson decomposition of $\mathbf{C}(L)$ in (8), we can write

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

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,

$$\mathbf{F}_t = \mathbf{C}(1) \sum_{s=1}^{t} \mathbf{u}_s + \boldsymbol{\omega}_t, \tag{A38}$$

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 5, and $\boldsymbol{\omega}_t \sim I(0)$, because of square summability of the coefficients of $\check{\mathbf{C}}(L)$. Moreover, from Assumption 1(a) of cointegration, 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 $\boldsymbol{\beta}'\mathbf{C}(1) = \mathbf{0}_{c \times q}$. This proves part (i).

For part (ii), from (A38), we have

$$\boldsymbol{\beta}'\mathbf{F}_t = \boldsymbol{\beta}'\boldsymbol{\omega}_t = \boldsymbol{\beta}'\check{\mathbf{C}}(L)\mathbf{u}_t.$$

Define $\mathbf{C}^*(L) = \boldsymbol{\beta}'\check{\mathbf{C}}(L)$ and notice that it has square summable coefficients because of square summability of the coefficients of $\mathbf{C}(L)$, then

$$\mathsf{E}\big[\|\boldsymbol{\beta}'\mathbf{F}_t\|^2\big] = \sum_{j=1}^{r} \mathsf{E}[(\mathbf{c}_j^{*'}(L)\mathbf{u}_t)^2] = \sum_{j=1}^{r} \mathsf{E}\left[\left(\sum_{l=1}^{q} c_{jl}^*(L)u_{lt}\right)^2\right]$$

$$= \sum_{j=1}^{r} \sum_{l,l'=1}^{q} \sum_{k,k'=0}^{\infty} c_{jlk}^* c_{jl'k'}^* \mathsf{E}[u_{lt-k}u_{l't-k'}] \le rqK_1, \tag{A39}$$

where we used the fact $\mathbf{u}_t$ is a white noise because of Assumption 1(a) and we used square summability of the coefficients, with K_1 defined in (A5). Part (ii) is proved by means of Chebychev's inequality. This completes the proof. $\qquad\square$

Lemma A5 *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 Brownian motion with finite covariance of rank $q - d$ and analogously define $\mathbf{W}_q(\cdot)$ having finite covariance of rank q. Define the autocovariances of $\boldsymbol{\omega}_t$ in (A38) as $\boldsymbol{\Gamma}_h^\omega$ and long-run covariance $\boldsymbol{\Gamma}_{L0}^\omega = \boldsymbol{\Gamma}_0^\omega + 2\sum_{h=1}^{\infty} \boldsymbol{\Gamma}_h^\omega$. Under Assumptions 1, 4 and 5, as $T \to \infty$,*

(i) $\|T^{-1}\sum_{t=k+1}^{T}\Delta\mathbf{F}_t\Delta\mathbf{F}'_{t-k} - \mathbf{\Gamma}_k^{\Delta F}\| = O_p(T^{-1/2})$;

(ii) $T^{-2}\sum_{t=1}^{T}\mathbf{F}_t\mathbf{F}'_t \xrightarrow{d} (\mathbf{\Gamma}_{L0}^{\Delta F})^{1/2}\big(\int_0^1 \mathbf{W}_r(\tau)\mathbf{W}'_r(\tau)\mathrm{d}\tau\big)(\mathbf{\Gamma}_{L0}^{\Delta F})^{1/2}$;

(iii) $T^{-1}\sum_{t=1}^{T}\mathbf{F}_{t-1}\Delta\mathbf{F}'_t \xrightarrow{d} (\mathbf{\Gamma}_{L0}^{\Delta F})^{1/2}\big(\int_0^1 \mathbf{W}_r(\tau)\mathrm{d}\mathbf{W}'_r(\tau)\big)(\mathbf{\Gamma}_{L0}^{\Delta F})^{1/2} + \mathbf{\Gamma}_{L1}^{\Delta F}$;

(iv) $T^{-1}\sum_{t=1}^{T}\mathbf{F}_t\mathbf{F}'_t\boldsymbol{\beta} \xrightarrow{d} \mathbf{C}(1)\big(\int_0^1 \mathbf{W}_q(\tau)\mathrm{d}\mathbf{W}'_r(\tau)\big)(\mathbf{\Gamma}_{L0}^{\omega})^{1/2}\boldsymbol{\beta} + \mathbf{\Gamma}_0^{\omega}\boldsymbol{\beta}$;

(v) $\|T^{-1}\sum_{t=1}^{T}\boldsymbol{\beta}'\mathbf{F}_t\mathbf{F}'_t\boldsymbol{\beta} - \boldsymbol{\beta}'\mathbf{\Gamma}_0^{\omega}\boldsymbol{\beta}\| = \|T^{-1}\sum_{t=1}^{T}\boldsymbol{\beta}'\mathbf{F}_t\mathbf{F}'_t\boldsymbol{\beta} - \mathsf{E}[\boldsymbol{\beta}'\mathbf{F}_t\mathbf{F}'_t\boldsymbol{\beta}]\| = O_p(T^{-1/2})$;

(vi) $\|T^{-1}\sum_{t=1}^{T}\Delta\mathbf{F}_t\mathbf{F}'_{t-1}\boldsymbol{\beta} - (\mathbf{\Gamma}_1^{\omega}-\mathbf{\Gamma}_0^{\omega})\boldsymbol{\beta}\| = \|T^{-1}\sum_{t=1}^{T}\Delta\mathbf{F}_t\mathbf{F}'_{t-1}\boldsymbol{\beta} - \mathsf{E}[\Delta\mathbf{F}_t\mathbf{F}'_{t-1}\boldsymbol{\beta}]\| = O_p(T^{-1/2})$.

Proof. For part (i), the case $k = 0$ is proved in (A12) in the proof of Lemma A1. The proof for the autocovariances, i.e. when $k \neq 0$, is analogous.

For parts (ii) and (iii), first notice that, by Assumption 1,

$$\mathbf{\Gamma}_{L0}^{\Delta F} = \sum_{k=0}^{\infty} \mathbf{C}_k\mathbf{C}'_k + \sum_{h=1}^{\infty}\sum_{k=h}^{\infty}\Big(\mathbf{C}_k\mathbf{C}'_{k+h} + \mathbf{C}_{k+h}\mathbf{C}'_k\Big), \tag{A40}$$

which is positive definite, and by square summability of the coefficients this matrix is also finite. Moreover, by Assumptions 4(a) and 4(b) the vector $\mathbf{u}_t$ satisfies the assumptions of Corollary 2.2 in Phillips and Durlauf (1986), then parts (ii) and (iii) are direct consequences of Lemma 3.1 in Phillips and Durlauf (1986).

Turning to part (iv), since $\boldsymbol{\beta}'\mathbf{F}_t = \boldsymbol{\beta}'\boldsymbol{\omega}_t$, because of Lemma A4(i), then,

$$\frac{1}{T}\sum_{t=1}^{T}\mathbf{F}_t\mathbf{F}'_t\boldsymbol{\beta} = \mathbf{C}(1)\left[\frac{1}{T}\sum_{t=1}^{T}\Big(\sum_{s=1}^{t}\mathbf{u}_s\Big)\boldsymbol{\omega}'_t\right]\boldsymbol{\beta} + \left[\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\omega}_t\boldsymbol{\omega}'_t\right]\boldsymbol{\beta}. \tag{A41}$$

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

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

where as for (A40) we can show that $\mathbf{\Gamma}_{L0}^{\omega} = \mathbf{\Gamma}_0^{\omega} + 2\sum_{h=1}^{\infty}\mathbf{\Gamma}_h^{\omega}$ is positive definite. Moreover, we can write $\boldsymbol{\omega}_t = \sqrt{T}(\mathbf{\Gamma}_{L0}^{\omega})^{1/2}[\mathbf{X}_{\omega,T}(t/T) - \mathbf{X}_{\omega,T}((t-1)/T)]$. 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 $T \to \infty$,

$$\mathbf{X}_{u,T}(\tau) \xrightarrow{d} \mathbf{W}_q(\tau), \qquad \mathbf{X}_{\omega,T}(\tau) \xrightarrow{d} \mathbf{W}_r(\tau), \tag{A42}$$

where $\mathbf{W}_q(\cdot)$ is a q-dimensional Brownian motion with covariance $\mathbf{I}_q$ and $\mathbf{W}_r(\cdot)$ is an r-dimensional Brownian motion with covariance $\mathbf{I}_r$. Then consider the first term in parenthesis on the rhs of (A41), as $T \to \infty$, using (A42), we have

$$\frac{1}{T}\sum_{t=1}^{T}\Big(\sum_{s=1}^{t}\mathbf{u}_s\Big)\boldsymbol{\omega}'_t = \sum_{t=1}^{T}\mathbf{X}_u\Big(\frac{t}{T}\Big)\Big(\mathbf{X}_\omega\Big(\frac{t}{T}\Big) - \mathbf{X}_\omega\Big(\frac{t-1}{T}\Big)\Big)'\Big(\mathbf{\Gamma}_{L0}^{\omega}\Big)^{1/2} \tag{A43}$$

$$\xrightarrow{d} \Big(\int_0^1 \mathbf{W}_u(\tau)\frac{(\mathbf{W}_\omega(\tau) - \mathbf{W}_\omega(\tau - \mathrm{d}\tau))'}{\mathrm{d}\tau}\mathrm{d}\tau\Big)\Big(\mathbf{\Gamma}_{L0}^{\omega}\Big)^{1/2} = \Big(\int_0^1 \mathbf{W}_u(\tau)\mathrm{d}\mathbf{W}'_\omega(\tau)\Big)\Big(\mathbf{\Gamma}_{L0}^{\omega}\Big)^{1/2}.$$

As for the second term on the rhs of (A41), we have, using the same approach as for part (i), as

$T \to \infty$,

$$\left\| \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{\omega}_t \boldsymbol{\omega}_t' - \boldsymbol{\Gamma}_0^{\omega} \right\| = O_p \left(\frac{1}{\sqrt{T}} \right). \tag{A44}$$

By substituting (A43) and (A44) in (A41), and by Slutsky's theorem, we complete the proof of part (iv). Part (v) is proved analogously just by multiplying (A41) also on the left by $\boldsymbol{\beta}'$.

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

$$\frac{1}{T} \sum_{t=1}^{T} \Delta \mathbf{F}_t \mathbf{F}_{t-1}' \boldsymbol{\beta} = \left(\frac{1}{T} \sum_{t=1}^{T} \mathbf{C}(1) \mathbf{u}_t \boldsymbol{\omega}_{t-1}' + \frac{1}{T} \sum_{t=1}^{T} \Delta \boldsymbol{\omega}_t \boldsymbol{\omega}_{t-1}' \right) = \left(\boldsymbol{\Gamma}_1^{\omega} - \boldsymbol{\Gamma}_0^{\omega} \right) \boldsymbol{\beta} + O_p \left(\frac{1}{\sqrt{T}} \right). \tag{A45}$$

This completes the proof. $\qquad\qquad\square$

Lemma A6 *Define* $\check{\mathbf{F}}_t = \mathbf{J} \mathbf{F}_t$ *and* $\check{\boldsymbol{\beta}} = \mathbf{J} \boldsymbol{\beta}$. *For any given* t, *under Assumptions 1 through 6, as* $n, T \to \infty$,

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

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

$$\left\| \frac{\widehat{\boldsymbol{\Lambda}}' \boldsymbol{\xi}_t \check{\mathbf{F}}_t'}{nT} \right\| \leq \left\| \frac{\mathbf{J} \boldsymbol{\Lambda}' \boldsymbol{\xi}_t \mathbf{F}_t' \mathbf{J}}{nT} \right\| + \left\| \frac{(\widehat{\boldsymbol{\Lambda}}' - \mathbf{J} \boldsymbol{\Lambda}') \boldsymbol{\xi}_t \mathbf{F}_t' \mathbf{J}}{nT} \right\|$$

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

Then, because of Lemma A2(ii) and A2(vi), the first term on the rhs is $O_p(n^{-1/2})$. Because of Lemma A2(ii) and A2(iv) and Lemma A3. This proves part (i).

For part (ii) we can repeat the same reasoning as for part (i), but using Lemma A2(i), A2(iii) and A2(v), and Lemma A3. 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), but using Lemma A2(iii) and A2(iv), and Lemmas A3 and A4(ii). Part (iv) is also proved as part (i), but using Lemma A2(ii), A2(iii) and A2(v), and Lemma A3. Part (v) is proved as part (i), but using Lemma A2(iv) and A2(vi), and Lemmas A3 and A4(ii).

For part (vi), we have

$$\left\|\frac{\widehat{\boldsymbol{\Lambda}}'\boldsymbol{\xi}_t\Delta\check{\mathbf{F}}_t'}{n}\right\| \leq \left\|\frac{\mathbf{J}\boldsymbol{\Lambda}'\boldsymbol{\xi}_t\Delta\mathbf{F}_t'\mathbf{J}}{n}\right\| + \left\|\frac{(\widehat{\boldsymbol{\Lambda}}' - \mathbf{J}\boldsymbol{\Lambda}')\boldsymbol{\xi}_t\Delta\mathbf{F}_t'\mathbf{J}}{n}\right\|$$

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

From Lemma A2(i) and A2(viii), the first term on the rhs is $O_p(T^{1/2}n^{-(2-\delta)/2})$. From Lemma A2(i) and A2(vii) and Lemma A3, 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 A2(ii), A2(vii) and A2(viii), and Lemmas A3 and Lemma A4(ii). This completes the proof. $\qquad\square$

Lemma A7 *For any given t, under Assumptions 1 through 6, as $n, T \to \infty$,*

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

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

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

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

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

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

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

Because of Lemma A2(iv) and Lemma A3, the first term on the rhs is $O_p(T^{-1})$. Because of Lemma A2(iv) and A2(vi), and Lemma A3, the second term is $O_p(T^{-1/2}n^{-1/2})$. The third term is $O_p(n^{-1})$ because of Lemma A2(vi). This proves part (i). Part (ii) is proved in the same way, but using Lemma A2(iii) and A2(v), and Lemma A3.

Now consider part (iii):

$$\left\|\frac{\widehat{\boldsymbol{\Lambda}}'\boldsymbol{\xi}_t\boldsymbol{\xi}_t'\widehat{\boldsymbol{\Lambda}}}{n^2}\right\| \leq \left\|\frac{\widehat{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda}\mathbf{J}}{\sqrt{n}}\right\|^2 \left\|\frac{\boldsymbol{\xi}_t}{\sqrt{n}}\right\|^2 + 2\left\|\frac{\widehat{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda}\mathbf{J}}{\sqrt{n}}\right\| \left\|\frac{\boldsymbol{\xi}_t}{\sqrt{n}}\right\| \left\|\frac{\boldsymbol{\Lambda}'\boldsymbol{\xi}_t}{n}\right\| + \left\|\frac{\boldsymbol{\Lambda}'\boldsymbol{\xi}_t}{n}\right\|^2. \tag{A46}$$

Because of Lemma A2(vii) and Lemma A3, the first term on the rhs is $O_p(n^{-(1-\delta)})$. Because of Lemma A2(vii) and A2(viii), and Lemma A3, the second term is $O_p(T^{1/2}n^{-(3/2-\delta)})$. The third term is $O_p(Tn^{-(2-\delta)})$ because of Lemma A2(viii). Summing up, for (A46), we have

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

In order to compare the rates of the three terms assume $n = O(T^\alpha)$, then, according to Assumption 6(b), we must have at least $\alpha > 1/2$. Now, when $1/2 < \alpha < 1$, the third term dominates over the first one (see also (17)), but the second would dominate over the third if and only if $\alpha > 1$, which cannot be. When $\alpha \geq 1$, 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 multiplying everything in part (iii) by $T^{-1/2}$.

For part (v), we have

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

Because of Lemma A2(iii) and A2(vii), and Lemma A3, the first term on the rhs is $O_p(T^{-1/2}n^{-(1-\delta)/2})$. Because of Lemma A2(v) and A2(viii), and Lemma A3, the second term is $O_p(T^{1/2}n^{-(3-\delta)/2})$. Hence, using (17), 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 part (v) is proved. This completes the proof. $\qquad\square$

Lemma A8 *Define the matrices*

$$\widehat{\mathbf{M}}_{00} = \frac{1}{T}\sum_{t=1}^{T}\Delta\widehat{\mathbf{F}}_t\Delta\widehat{\mathbf{F}}_t', \quad \widehat{\mathbf{M}}_{01} = \frac{1}{T}\sum_{t=1}^{T}\Delta\widehat{\mathbf{F}}_t\widehat{\mathbf{F}}_{t-1}', \quad \widehat{\mathbf{M}}_{02} = \frac{1}{T}\sum_{t=1}^{T}\Delta\widehat{\mathbf{F}}_t\Delta\widehat{\mathbf{F}}_{t-1}',$$

$$\widehat{\mathbf{M}}_{11} = \frac{1}{T}\sum_{t=1}^{T}\widehat{\mathbf{F}}_t\widehat{\mathbf{F}}_t', \quad \widehat{\mathbf{M}}_{21} = \frac{1}{T}\sum_{t=1}^{T}\Delta\widehat{\mathbf{F}}_{t-1}\widehat{\mathbf{F}}_{t-1}', \quad \widehat{\mathbf{M}}_{22} = \frac{1}{T}\sum_{t=1}^{T}\Delta\widehat{\mathbf{F}}_{t-1}\Delta\widehat{\mathbf{F}}_{t-1}',$$

and denote by $\mathbf{M}_{ij}$, for $i,j = 0,1,2$, the analogous ones but computed by using $\check{\mathbf{F}}_t = \mathbf{J}\mathbf{F}_t$. Define also $\check{\boldsymbol{\beta}} = \mathbf{J}\boldsymbol{\beta}$. Under Assumptions 1 through 6, as $n, T \to \infty$,

(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) $\|\widehat{\mathbf{M}}_{22} - \mathbf{M}_{22}\| = O_p(n^{-1/2}, T^{-1/2})$.
(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. Throughout, we use $\|\boldsymbol{\beta}\| = O(1)$ and obviously $\|\mathbf{J}\| = 1$ and the fact that, from Lemma A3 we also have $\|n^{-1}\widehat{\boldsymbol{\Lambda}}'\boldsymbol{\Lambda}\| = O_p(1)$. Start with part (i). By adding and subtracting $\mathbf{J}\mathbf{F}_t$ from $\widehat{\mathbf{F}}_t$, we have

$$\left\| \frac{1}{T^2}\sum_{t=1}^{T}\widehat{\mathbf{F}}_t\widehat{\mathbf{F}}_t' - \frac{1}{T^2}\sum_{t=1}^{T}\check{\mathbf{F}}_t\check{\mathbf{F}}_t' \right\| \leq \left\| \frac{1}{T^2}\sum_{t=1}^{T}\left(\widehat{\mathbf{F}}_t - \mathbf{J}\mathbf{F}_t\right)\left(\widehat{\mathbf{F}}_t - \mathbf{J}\mathbf{F}_t\right)' \right\|$$
$$+ 2\left\| \frac{1}{T^2}\sum_{t=1}^{T}\left(\widehat{\mathbf{F}}_t - \mathbf{J}\mathbf{F}_t\right)\left(\mathbf{J}\mathbf{F}_t\right)' \right\|. \tag{A47}$$

49

Using (6) and (15) , the first term on the rhs of (A47) gives

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

$$
= \left\| \frac{1}{T^2} \sum_{t=1}^{T} \left(\frac{\widehat{\boldsymbol{\Lambda}}'\boldsymbol{\Lambda}\mathbf{F}_t}{n} + \frac{\widehat{\boldsymbol{\Lambda}}'\boldsymbol{\xi}_t}{n} - \mathbf{J}\mathbf{F}_t \right) \left(\frac{\widehat{\boldsymbol{\Lambda}}'\boldsymbol{\Lambda}\mathbf{F}_t}{n} + \frac{\widehat{\boldsymbol{\Lambda}}'\boldsymbol{\xi}_t}{n} - \mathbf{J}\mathbf{F}_t \right)' \right\|
$$

$$
\leq \underbrace{\left\| \frac{1}{T^2} \sum_{t=1}^{T} \frac{\widehat{\boldsymbol{\Lambda}}'\boldsymbol{\Lambda}\mathbf{F}_t\mathbf{F}_t'}{n} \left(\frac{\boldsymbol{\Lambda}'\widehat{\boldsymbol{\Lambda}}}{n} - \mathbf{J} \right) + \mathbf{J}\mathbf{F}_t\mathbf{F}_t' \left(\mathbf{J} - \frac{\boldsymbol{\Lambda}'\widehat{\boldsymbol{\Lambda}}}{n} \right) \right\|}_{\mathcal{A}_1} + 2 \underbrace{\left\| \frac{1}{T^2} \sum_{t=1}^{T} \frac{\widehat{\boldsymbol{\Lambda}}'\boldsymbol{\Lambda}\mathbf{F}_t\boldsymbol{\xi}_t'\widehat{\boldsymbol{\Lambda}}}{n^2} \right\|}_{\mathcal{B}_1}
$$

$$
+ 2 \underbrace{\left\| \frac{1}{T^2} \sum_{t=1}^{T} \frac{\widehat{\boldsymbol{\Lambda}}'\boldsymbol{\xi}_t\mathbf{F}_t'\mathbf{J}}{n} \right\|}_{\mathcal{C}_1} + \underbrace{\left\| \frac{1}{T^2} \sum_{t=1}^{T} \frac{\widehat{\boldsymbol{\Lambda}}'\boldsymbol{\xi}_t\boldsymbol{\xi}_t'\widehat{\boldsymbol{\Lambda}}}{n^2} \right\|}_{\mathcal{D}_1}.
\tag{A48}
$$

Let us consider each term of (A48) separately:

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

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

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

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

Above we used, Lemma A3 and Lemma A5(ii) for $\mathcal{A}_1$, Lemma A6(i) for $\mathcal{B}_1$ and $\mathcal{C}_1$, and Lemma A7(i) for $\mathcal{D}_1$. Thus, the first term on the rhs of (A47) is $O_p(\max(n^{-1/2}, T^{-1/2}))$. The second term on the rhs of (A47) is such that

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

$$
\leq \left\| \frac{1}{T^2} \sum_{t=1}^{T} \left(\frac{\widehat{\boldsymbol{\Lambda}}'\boldsymbol{\Lambda}}{n} - \mathbf{J} \right) \mathbf{F}_t\mathbf{F}_t'\mathbf{J} \right\| + \left\| \frac{1}{T^2} \sum_{t=1}^{T} \frac{\widehat{\boldsymbol{\Lambda}}'\boldsymbol{\xi}_t\mathbf{F}_t'\mathbf{J}}{n} \right\|
$$

$$
\leq \left\| \frac{\widehat{\boldsymbol{\Lambda}}'\boldsymbol{\Lambda}}{n} - \mathbf{J} \right\| \left\| \frac{1}{T^2} \sum_{t=1}^{T} \mathbf{F}_t\mathbf{F}_t' \right\| \|\mathbf{J}\| + \frac{1}{T} \sum_{t=1}^{T} \left\| \frac{\widehat{\boldsymbol{\Lambda}}'\boldsymbol{\xi}_t\mathbf{F}_t'\mathbf{J}}{nT} \right\| = O_p\left(\max\left(\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{T}} \right) \right), \tag{A49}
$$

where we used Lemmas A3, A5(ii) and A6(i). By combining (A48) and (A49) we prove part (i). Parts (ii), (iii) and (iv) are proved in the same way as part (i), using Lemma A3 and the results for the stationary process $\Delta\mathbf{F}_t$ in Lemmas A5(i), A6(ii), and A7(ii).

Now, consider part (v):

$$\left\| \frac{1}{T}\sum_{t=1}^{T}\Delta\widehat{\mathbf{F}}_t\widehat{\mathbf{F}}'_{t-1}\breve{\boldsymbol{\beta}} - \frac{1}{T}\sum_{t=1}^{T}\Delta\breve{\mathbf{F}}_t\breve{\mathbf{F}}'_{t-1}\breve{\boldsymbol{\beta}} \right\| \le \left\| \frac{1}{T}\sum_{t=1}^{T}\left(\Delta\widehat{\mathbf{F}}_t - \mathbf{J}\Delta\mathbf{F}_t\right)\left(\widehat{\mathbf{F}}_{t-1} - \mathbf{J}\mathbf{F}_{t-1}\right)'\breve{\boldsymbol{\beta}} \right\|$$

$$+ \left\| \frac{1}{T}\sum_{t=1}^{T}\left(\Delta\widehat{\mathbf{F}}_t - \mathbf{J}\Delta\mathbf{F}_t\right)\left(\breve{\boldsymbol{\beta}}'\mathbf{J}\mathbf{F}_{t-1}\right)' \right\| + \left\| \frac{1}{T}\sum_{t=1}^{T}\left(\mathbf{J}\Delta\mathbf{F}_t\right)\left(\widehat{\mathbf{F}}_{t-1} - \mathbf{J}\mathbf{F}_{t-1}\right)'\breve{\boldsymbol{\beta}} \right\|. \tag{A50}$$

Similarly to (A48), from (6) and (15), the first term on the rhs of (A50) is such that

$$\left\| \frac{1}{T}\sum_{t=1}^{T}\left(\Delta\widehat{\mathbf{F}}_t - \mathbf{J}\Delta\mathbf{F}_{t-1}\right)\left(\breve{\boldsymbol{\beta}}'\widehat{\mathbf{F}}_{t-1} - \breve{\boldsymbol{\beta}}'\mathbf{J}\mathbf{F}_{t-1}\right)' \right\| = \left\| \frac{1}{T}\sum_{t=1}^{T}\left(\frac{\widehat{\boldsymbol{\Lambda}}'\Delta\mathbf{x}_t}{n} - \mathbf{J}\Delta\mathbf{F}_t\right)\left(\frac{\widehat{\boldsymbol{\Lambda}}'\mathbf{x}_{t-1}}{n} - \mathbf{J}\mathbf{F}_{t-1}\right)'\breve{\boldsymbol{\beta}} \right\|$$

$$\le \underbrace{\left\| \frac{1}{T}\sum_{t=1}^{T}\frac{\widehat{\boldsymbol{\Lambda}}'\boldsymbol{\Lambda}\Delta\mathbf{F}_t\mathbf{F}'_{t-1}}{n}\left(\frac{\boldsymbol{\Lambda}'\widehat{\boldsymbol{\Lambda}}}{n} - \mathbf{J}\right)\breve{\boldsymbol{\beta}} + \mathbf{J}\Delta\mathbf{F}_t\mathbf{F}'_{t-1}\left(\mathbf{J} - \frac{\boldsymbol{\Lambda}'\widehat{\boldsymbol{\Lambda}}}{n}\right)\breve{\boldsymbol{\beta}} \right\|}_{\mathcal{A}_2} + \underbrace{\left\| \frac{1}{T}\sum_{t=1}^{T}\frac{\widehat{\boldsymbol{\Lambda}}'\boldsymbol{\Lambda}\Delta\mathbf{F}_t\boldsymbol{\xi}'_{t-1}\widehat{\boldsymbol{\Lambda}}\breve{\boldsymbol{\beta}}}{n^2} \right\|}_{\mathcal{B}_2}$$

$$+ \underbrace{\left\| \frac{1}{T}\sum_{t=1}^{T}\frac{\widehat{\boldsymbol{\Lambda}}'\Delta\boldsymbol{\xi}_t\mathbf{F}'_{t-1}\boldsymbol{\Lambda}'\widehat{\boldsymbol{\Lambda}}\breve{\boldsymbol{\beta}}}{n^2} \right\|}_{\mathcal{C}_2} + \underbrace{\left\| \frac{1}{T}\sum_{t=1}^{T}\frac{\mathbf{J}\Delta\mathbf{F}_t\boldsymbol{\xi}'_{t-1}\widehat{\boldsymbol{\Lambda}}\breve{\boldsymbol{\beta}}}{n} \right\|}_{\mathcal{D}_2} + \underbrace{\left\| \frac{1}{T}\sum_{t=1}^{T}\frac{\widehat{\boldsymbol{\Lambda}}'\Delta\boldsymbol{\xi}_t\mathbf{F}'_{t-1}\mathbf{J}\breve{\boldsymbol{\beta}}}{n} \right\|}_{\mathcal{E}_2}$$

$$+ \underbrace{\left\| \frac{1}{T}\sum_{t=1}^{T}\frac{\widehat{\boldsymbol{\Lambda}}'\Delta\boldsymbol{\xi}_t\boldsymbol{\xi}'_{t-1}\widehat{\boldsymbol{\Lambda}}\breve{\boldsymbol{\beta}}}{n^2} \right\|}_{\mathcal{F}_2}. \tag{A51}$$

Let us consider first the terms:

$$\mathcal{A}_2 \le \left\| \frac{\boldsymbol{\Lambda}'\widehat{\boldsymbol{\Lambda}}}{n} - \mathbf{J} \right\| \left\| \frac{1}{T}\sum_{t=1}^{T}\Delta\mathbf{F}_t\mathbf{F}'_{t-1} \right\| \left\{ \left\| \frac{\widehat{\boldsymbol{\Lambda}}'\boldsymbol{\Lambda}}{n} \right\| + \|\mathbf{J}\| \right\} \|\breve{\boldsymbol{\beta}}\| = O_p\left(\frac{1}{\sqrt{T}}\right),$$

$$\mathcal{B}_2 \le \frac{1}{T}\sum_{t=1}^{T}\left\| \frac{\widehat{\boldsymbol{\Lambda}}'\boldsymbol{\xi}_{t-1}\Delta\mathbf{F}'_t}{n} \right\| \left\| \frac{\widehat{\boldsymbol{\Lambda}}'\boldsymbol{\Lambda}}{n} \right\| \|\breve{\boldsymbol{\beta}}\| = O_p(\zeta_{nT,\delta}),$$

$$\mathcal{F}_2 \le \frac{1}{T}\sum_{t=1}^{T}\left\| \frac{\widehat{\boldsymbol{\Lambda}}'\Delta\boldsymbol{\xi}_t\boldsymbol{\xi}'_{t-1}\widehat{\boldsymbol{\Lambda}}}{n^2} \right\| \|\breve{\boldsymbol{\beta}}\| = O_p\left(\zeta_{nT,\delta}\max\left(\frac{1}{\sqrt{n}},\frac{1}{\sqrt{T}}\right)\right),$$

Above we used, Lemmas A3 and A5(iii) for $\mathcal{A}_2$, Lemma A6(vi) for $\mathcal{B}_2$, and Lemma A7(v) for $\mathcal{F}_2$. 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 A6(iii). Finally, recall that from Lemma A3, we also have

$$\frac{\boldsymbol{\Lambda}'\widehat{\boldsymbol{\Lambda}}}{n} = \mathbf{J} + O_p\left(\frac{1}{\sqrt{T}}\right). \tag{A52}$$

Hence, from (A52),

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

Indeed, the first term on the rhs of $\mathcal{C}_2$ is $O_p(\max(n^{-1/2}, T^{-1/2}))$ because of Lemma A6(iii), while the second term is $O_p(\max(n^{-1/2}, T^{-1/2}))$ because of Lemma A6(iv). Therefore, the first term on the rhs of (A50) is $O_p(\max(\zeta_{nT,\delta}, T^{-1/2}))$.

As for the second term on the rhs of (A50), since $\check{\boldsymbol{\beta}}'\mathbf{J}\mathbf{F}_{t-1} = \boldsymbol{\beta}'\mathbf{F}_{t-1}$, we have

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

$$\leq \left\|\frac{1}{T}\sum_{t=1}^{T}\left(\frac{\widehat{\boldsymbol{\Lambda}}'\boldsymbol{\Lambda}}{n} - \mathbf{J}\right)\Delta\mathbf{F}_t\mathbf{F}'_{t-1}\boldsymbol{\beta}\right\| + \left\|\frac{1}{T}\sum_{t=1}^{T}\frac{\widehat{\boldsymbol{\Lambda}}'\Delta\boldsymbol{\xi}_t\check{\mathbf{F}}'_{t-1}\check{\boldsymbol{\beta}}}{n}\right\| = O_p\left(\max\left(\frac{1}{\sqrt{n}},\frac{1}{\sqrt{T}}\right)\right), \quad (A53)$$

where we used Lemmas A3 and A5(vi) for the first term on the rhs and Lemma A6(iii) for the second.

The third term on the rhs of (A50) is such that

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

$$\leq \left\|\frac{1}{T}\sum_{t=1}^{T}\mathbf{J}\Delta\mathbf{F}_t\mathbf{F}'_{t-1}\left(\frac{\boldsymbol{\Lambda}'\widehat{\boldsymbol{\Lambda}}}{n} - \mathbf{J}\right)\check{\boldsymbol{\beta}}\right\| + \left\|\frac{1}{T}\sum_{t=1}^{T}\frac{\mathbf{J}\Delta\mathbf{F}_t\boldsymbol{\xi}'_{t-1}\widehat{\boldsymbol{\Lambda}}\check{\boldsymbol{\beta}}}{n}\right\| = O_p(\zeta_{nT,\delta}), \quad (A54)$$

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 (A51), (A53), and (A54) we prove part (v).

Then consider part (vi):

$$\left\|\frac{1}{T}\sum_{t=1}^{T}\check{\boldsymbol{\beta}}'\widehat{\mathbf{F}}_t\widehat{\mathbf{F}}'_t\check{\boldsymbol{\beta}} - \frac{1}{T}\sum_{t=1}^{T}\check{\boldsymbol{\beta}}'\check{\mathbf{F}}_t\check{\mathbf{F}}'_t\check{\boldsymbol{\beta}}\right\| \leq \left\|\frac{1}{T}\sum_{t=1}^{T}\check{\boldsymbol{\beta}}'\left(\widehat{\mathbf{F}}_t - \mathbf{J}\mathbf{F}_t\right)\left(\widehat{\mathbf{F}}_t - \mathbf{J}\mathbf{F}_t\right)'\check{\boldsymbol{\beta}}\right\|$$

$$+ 2\left\|\frac{1}{T}\sum_{t=1}^{T}\check{\boldsymbol{\beta}}'\left(\widehat{\mathbf{F}}_t - \mathbf{J}\mathbf{F}_t\right)\left(\check{\boldsymbol{\beta}}'\mathbf{J}\mathbf{F}_t\right)'\right\|. \quad (A55)$$

As before, from (6) and (15), the first term on the rhs of (A55) is such that

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

$$
\leq \underbrace{\left\| \frac{1}{T} \sum_{t=1}^{T} \frac{\boldsymbol{\check{\beta}}' \widehat{\boldsymbol{\Lambda}}' \boldsymbol{\Lambda} \mathbf{F}_t \mathbf{F}_t'}{n} \Big(\frac{\boldsymbol{\Lambda}' \widehat{\boldsymbol{\Lambda}}}{n} - \mathbf{J} \Big) \boldsymbol{\check{\beta}} + \boldsymbol{\check{\beta}}' \mathbf{J} \mathbf{F}_t \mathbf{F}_t' \Big(\mathbf{J} - \frac{\boldsymbol{\Lambda}' \widehat{\boldsymbol{\Lambda}}}{n} \Big) \boldsymbol{\check{\beta}} \right\|}_{\mathcal{A}_3} + 2 \underbrace{\left\| \frac{1}{T} \sum_{t=1}^{T} \frac{\boldsymbol{\check{\beta}}' \widehat{\boldsymbol{\Lambda}}' \boldsymbol{\Lambda} \mathbf{F}_t \boldsymbol{\xi}_t' \widehat{\boldsymbol{\Lambda}} \boldsymbol{\check{\beta}}}{n^2} \right\|}_{\mathcal{B}_3}
$$

$$
+ 2 \underbrace{\left\| \frac{1}{T} \sum_{t=1}^{T} \frac{\boldsymbol{\check{\beta}}' \mathbf{J} \mathbf{F}_t \boldsymbol{\xi}_t' \widehat{\boldsymbol{\Lambda}} \boldsymbol{\check{\beta}}}{n} \right\|}_{\mathcal{C}_3} + \underbrace{\left\| \frac{1}{T} \sum_{t=1}^{T} \frac{\boldsymbol{\check{\beta}}' \widehat{\boldsymbol{\Lambda}}' \boldsymbol{\xi}_t \boldsymbol{\xi}_t' \widehat{\boldsymbol{\Lambda}} \boldsymbol{\check{\beta}}}{n^2} \right\|}_{\mathcal{D}_3}. \tag{A56}
$$

Since $\boldsymbol{\check{\beta}}' \mathbf{J}\mathbf{F}_t = \boldsymbol{\beta}' \mathbf{F}_t$ and using (A52), we have,

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

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

Indeed, the first and third terms on the rhs are $O_p(T^{-1/2})$ because of Lemmas A3 and Lemma A5(v), while using the same results and (A52), the second term is

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

$$
= \left\| \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{\check{\beta}}' \mathbf{F}_t \mathbf{F}_t' \boldsymbol{\beta} \right\| O_p\Big(\frac{1}{T} \Big) = O_p\Big(\frac{1}{T} \Big).
$$

In the same way we have

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

because of Lemma A6(vii) and A6(viii). Then,

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

$$
\mathcal{D}_3 \leq \frac{1}{T} \sum_{t=1}^{T} \left\| \frac{\widehat{\boldsymbol{\Lambda}}' \boldsymbol{\xi}_t \boldsymbol{\xi}_t' \widehat{\boldsymbol{\Lambda}}}{n^2} \right\| \|\boldsymbol{\check{\beta}}\|^2 = O_p(\zeta_{nT,\delta}^2),
$$

because of Lemmas A6(viii) and A7(iii). Therefore, since from Assumption 6(b), $\zeta_{nT,\delta}^2 < \zeta_{nT,\delta}$ as $n, T \to \infty$, the first term on the rhs of (A55) is $O_p(\zeta_{nT,\delta})$.

53

The second term on the rhs of (A55) is such that

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

$$\leq 2\left\|\frac{1}{T}\sum_{t=1}^{T}\left(\frac{\widehat{\boldsymbol{\Lambda}}'\boldsymbol{\Lambda}}{n}-\mathbf{J}\right)\mathbf{F}_t\mathbf{F}_t'\mathbf{J}\check{\boldsymbol{\beta}}\right\|+2\left\|\frac{1}{T}\sum_{t=1}^{T}\frac{\widehat{\boldsymbol{\Lambda}}'\boldsymbol{\xi}_t\mathbf{F}_t'\mathbf{J}\check{\boldsymbol{\beta}}}{n}\right\|=O_p\left(\max\left(\zeta_{nT,\delta},\frac{1}{\sqrt{T}}\right)\right),\quad\text{(A57)}$$

because of Lemmas A3, A5(iv) and A6(viii). By combining (A56) and (A57), we prove part (vi). Finally, parts (vii), (viii) and (ix) are proved as part (v), by noticing that $\|T^{-1/2}\mathbf{F}_t\|=O_p(1)$, because of Lemma A2(ii). This completes the proof. $\qquad\square$

Lemma A9 *Define the matrices*

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

where $\widehat{\mathbf{M}}_{10}=\widehat{\mathbf{M}}_{01}'$, $\widehat{\mathbf{M}}_{20}=\widehat{\mathbf{M}}_{02}'$, and $\widehat{\mathbf{M}}_{12}=\widehat{\mathbf{M}}_{21}'$. Denote by $\mathbf{S}_{ij}$, for $i,j=0,1$, the analogous ones but computed by using $\check{\mathbf{F}}_t=\mathbf{J}\mathbf{F}_t$. Define also $\check{\boldsymbol{\beta}}=\mathbf{J}\boldsymbol{\beta}$ and $\check{\boldsymbol{\beta}}_{\perp*}=\check{\boldsymbol{\beta}}_\perp(\check{\boldsymbol{\beta}}_\perp'\check{\boldsymbol{\beta}}_\perp)^{-1}$, where $\check{\boldsymbol{\beta}}_\perp=\mathbf{J}\boldsymbol{\beta}_\perp$ such that $\check{\boldsymbol{\beta}}_\perp'\check{\boldsymbol{\beta}}=\mathbf{0}_{r-c\times r}$. Under Assumptions 1 through 6, as $n,T\to\infty$,

(i) $\|\widehat{\mathbf{S}}_{00}-\mathbf{S}_{00}\|=O_p(\max(n^{-1/2},T^{-1/2}))$.
(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. Throughout we use the fact that $\|\check{\boldsymbol{\beta}}_{\perp*}\|=O(1)$. Part (i) is proved using Lemma A8(ii), A8(iii) and A8(iv). For proving part (ii) we use Lemma A8(iv), A8(v) and A8(vi). Part (iii) is proved by combining part (ii) with Lemma A8(v) and A8(vi), and by noticing that $\|T^{-1/2}\mathbf{F}_t\|=O_p(1)$ from Lemma A2(ii). For proving part (iv) we combine part (i) with Lemma A8(v), A8(viii) and A8(ix). Part (v) is proved by combining part (i) with Lemma A8(viii) and A8(ix). Finally, part (vi) follows from Lemma A8(i) and A8(ix). This completes the proof. $\qquad\square$

Lemma A10 *Consider the matrices $\mathbf{S}_{ij}$ defined in Lemma A9, with $i,j=0,1$. Define $\check{\mathbf{F}}_t=\mathbf{J}\mathbf{F}_t$, $\check{\boldsymbol{\beta}}=\mathbf{J}\boldsymbol{\beta}$ and the conditional covariance matrices*

$$\check{\boldsymbol{\Omega}}_{00}=\mathsf{E}[\Delta\check{\mathbf{F}}_t\Delta\check{\mathbf{F}}_t'|\Delta\check{\mathbf{F}}_{t-1}],\quad\check{\boldsymbol{\Omega}}_{\check{\beta}\check{\beta}}=\mathsf{E}[\check{\boldsymbol{\beta}}'\check{\mathbf{F}}_{t-1}\check{\mathbf{F}}_{t-1}'\check{\boldsymbol{\beta}}|\Delta\check{\mathbf{F}}_{t-1}],\quad\check{\boldsymbol{\Omega}}_{0\check{\beta}}=\mathsf{E}[\Delta\check{\mathbf{F}}_t\check{\mathbf{F}}_{t-1}'\check{\boldsymbol{\beta}}|\Delta\check{\mathbf{F}}_{t-1}].$$

Under Assumptions 1, 4 and 5, 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. For part (i), notice that

$$\check{\boldsymbol{\Omega}}_{00} = \mathsf{E}[\Delta\check{\mathbf{F}}_t \Delta\check{\mathbf{F}}_t'] - \mathsf{E}[\Delta\check{\mathbf{F}}_t \Delta\check{\mathbf{F}}_{t-1}']\Big(\mathsf{E}[\Delta\check{\mathbf{F}}_{t-1}\Delta\check{\mathbf{F}}_{t-1}']\Big)^{-1}\mathsf{E}[\Delta\check{\mathbf{F}}_{t-1}\Delta\check{\mathbf{F}}_t']$$

$$= \boldsymbol{\Gamma}_0^{\Delta F} - \boldsymbol{\Gamma}_1^{\Delta F}\Big(\boldsymbol{\Gamma}_0^{\Delta F}\Big)^{-1}\boldsymbol{\Gamma}_1^{\Delta F},$$

and

$$\mathbf{S}_{00} = \frac{1}{T}\sum_{t=1}^T \Delta\check{\mathbf{F}}_t\Delta\check{\mathbf{F}}_t' - \left(\frac{1}{T}\sum_{t=2}^T \Delta\check{\mathbf{F}}_t\Delta\check{\mathbf{F}}_{t-1}'\right)\left(\frac{1}{T}\sum_{t=2}^T \Delta\check{\mathbf{F}}_{t-1}\Delta\check{\mathbf{F}}_{t-1}'\right)^{-1}\frac{1}{T}\sum_{t=2}^T \Delta\check{\mathbf{F}}_{t-1}\Delta\check{\mathbf{F}}_t'$$

$$= \mathbf{M}_{00} - \mathbf{M}_{02}\mathbf{M}_{22}^{-1}\mathbf{M}_{20}.$$

Using Lemma A5(i), we have the result. Parts *(*ii) and (iii) are proved in the same way, but using Lemma A5(v) and A5(vi), respectively. This completes the proof. $\qquad\square$

Proof of Lemma 4

Throughout we make use of the matrices $\widehat{\mathbf{M}}_{ij}$ and $\mathbf{M}_{ij}$, with $i,j = 0,1,2$, defined in Lemma A8, $\widehat{\mathbf{S}}_{ij}$ and $\mathbf{S}_{ij}$, with $i,j = 0,1$, defined in Lemma A9, and the conditional covariances $\check{\boldsymbol{\Omega}}_{00}$, $\check{\boldsymbol{\Omega}}_{\check{\beta}\check{\beta}}$ and $\check{\boldsymbol{\Omega}}_{0\check{\beta}}$ defined in Lemma A10. Define also $\check{\boldsymbol{\Omega}}_{\check{\beta}0} = \check{\boldsymbol{\Omega}}_{0\check{\beta}}'$. Finally, we denote as $\check{\boldsymbol{\beta}} = \mathbf{J}\boldsymbol{\beta}$ the matrix of cointegration vectors of $\check{\mathbf{F}}_t = \mathbf{J}\mathbf{F}_t$ 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). 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 $\widehat{\mathbf{S}}_{ij} = T^{-1}\sum_{t=1}^T \widehat{\mathbf{e}}_{it}\widehat{\mathbf{e}}_{jt}'$, with $i,j = 0,1$. Consider the generalized eigenvalues problem

$$\det\left(\widehat{\mu}_j\widehat{\mathbf{S}}_{11} - \widehat{\mathbf{S}}_{10}\widehat{\mathbf{S}}_{00}^{-1}\widehat{\mathbf{S}}_{01}\right) = 0, \quad j = 1,\ldots,r. \tag{A58}$$

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 (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

$$\det\left(\widehat{\mu}_j^0\mathbf{S}_{11} - \mathbf{S}_{10}\mathbf{S}_{00}^{-1}\mathbf{S}_{01}\right) = 0, \quad j = 1,\ldots,r. \tag{A59}$$

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

$$\mathbf{A}_T = \left(\check{\boldsymbol{\beta}} \;\; \frac{\check{\boldsymbol{\beta}}_{\perp*}}{\sqrt{T}}\right),$$

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

$$\det\left[\mathbf{A}'_T\big(\widehat{\mu}_j\widehat{\mathbf{S}}_{11} - \widehat{\mathbf{S}}_{10}\widehat{\mathbf{S}}_{00}^{-1}\widehat{\mathbf{S}}_{01}\big)\mathbf{A}_T\right] = 0, \quad j = 1,\dots,r, \tag{A60}$$

$$\det\left[\mathbf{A}'_T\big(\widehat{\mu}_j^0\mathbf{S}_{11} - \mathbf{S}_{10}\mathbf{S}_{00}^{-1}\mathbf{S}_{01}\big)\mathbf{A}_T\right] = 0, \quad j = 1,\dots,r. \tag{A61}$$

Clearly (A60) has the same solutions as (A58), 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 (A61) 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,

$$\mathbf{A}'_T\big(\widehat{\mathbf{S}}_{11} - \widehat{\mathbf{S}}_{10}\widehat{\mathbf{S}}_{00}^{-1}\widehat{\mathbf{S}}_{01}\big)\mathbf{A}_T$$
$$= \begin{bmatrix} \check{\boldsymbol{\beta}}'\widehat{\mathbf{S}}_{11}\check{\boldsymbol{\beta}} & T^{-1/2}\check{\boldsymbol{\beta}}'\widehat{\mathbf{S}}_{11}\check{\boldsymbol{\beta}}_{\perp*} \\ T^{-1/2}\check{\boldsymbol{\beta}}'_{\perp*}\widehat{\mathbf{S}}_{11}\check{\boldsymbol{\beta}} & T^{-1}\check{\boldsymbol{\beta}}'_{\perp*}\widehat{\mathbf{S}}_{11}\check{\boldsymbol{\beta}}_{\perp*} \end{bmatrix} - \begin{bmatrix} \check{\boldsymbol{\beta}}'\widehat{\mathbf{S}}_{10}\widehat{\mathbf{S}}_{00}^{-1}\widehat{\mathbf{S}}_{01}\check{\boldsymbol{\beta}} & T^{-1/2}\check{\boldsymbol{\beta}}'\widehat{\mathbf{S}}_{10}\widehat{\mathbf{S}}_{00}^{-1}\widehat{\mathbf{S}}_{01}\check{\boldsymbol{\beta}}_{\perp*} \\ T^{-1/2}\check{\boldsymbol{\beta}}'_{\perp*}\widehat{\mathbf{S}}_{10}\widehat{\mathbf{S}}_{00}^{-1}\widehat{\mathbf{S}}_{01}\check{\boldsymbol{\beta}} & T^{-1}\check{\boldsymbol{\beta}}'_{\perp*}\widehat{\mathbf{S}}_{10}\widehat{\mathbf{S}}_{00}^{-1}\widehat{\mathbf{S}}_{01}\check{\boldsymbol{\beta}}_{\perp*} \end{bmatrix}$$

$$= \begin{bmatrix} \check{\boldsymbol{\beta}}'\mathbf{S}_{11}\check{\boldsymbol{\beta}} & T^{-1/2}\check{\boldsymbol{\beta}}'\mathbf{S}_{11}\check{\boldsymbol{\beta}}_{\perp*} \\ T^{-1/2}\check{\boldsymbol{\beta}}'_{\perp*}\mathbf{S}_{11}\check{\boldsymbol{\beta}} & T^{-1}\check{\boldsymbol{\beta}}'_{\perp*}\mathbf{S}_{11}\check{\boldsymbol{\beta}}_{\perp*} \end{bmatrix} - \begin{bmatrix} \check{\boldsymbol{\beta}}'\mathbf{S}_{10}\mathbf{S}_{00}^{-1}\mathbf{S}_{01}\check{\boldsymbol{\beta}} & T^{-1/2}\check{\boldsymbol{\beta}}'\mathbf{S}_{10}\mathbf{S}_{00}^{-1}\mathbf{S}_{01}\check{\boldsymbol{\beta}}_{\perp*} \\ T^{-1/2}\check{\boldsymbol{\beta}}'_{\perp*}\mathbf{S}_{10}\mathbf{S}_{00}^{-1}\mathbf{S}_{01}\check{\boldsymbol{\beta}} & T^{-1}\check{\boldsymbol{\beta}}'_{\perp*}\mathbf{S}_{10}\mathbf{S}_{00}^{-1}\mathbf{S}_{01}\check{\boldsymbol{\beta}}_{\perp*} \end{bmatrix} + O_p(\vartheta_{nT,\delta})$$

$$= \mathbf{A}'_T\big(\mathbf{S}_{11} - \mathbf{S}_{10}\mathbf{S}_{00}^{-1}\mathbf{S}_{01}\big)\mathbf{A}_T + O_p(\vartheta_{nT,\delta}). \tag{A62}$$

This result is proved by using Lemma A9(ii), A9(iii) and A9(vi) for the first term on the rhs, and by using Lemma A9(i), A9(iv) and A9(v) for the second term. Thus, from (A62), for any $j = 1,\dots,r$, from Weyl's inequality (A3), we have

$$\big|\widehat{\mu}_j - \widehat{\mu}_j^0\big| \le \big\|\mathbf{A}'_T\big(\widehat{\mathbf{S}}_{11} - \widehat{\mathbf{S}}_{10}\widehat{\mathbf{S}}_{00}^{-1}\widehat{\mathbf{S}}_{01}\big)\mathbf{A}_T - \mathbf{A}'_T\big(\mathbf{S}_{11} - \mathbf{S}_{10}\mathbf{S}_{00}^{-1}\mathbf{S}_{01}\big)\mathbf{A}_T\big\| = O_p(\vartheta_{nT,\delta}). \tag{A63}$$

Moreover, always from (A62) and similarly to (A27), it can be shown that, by Theorem 2 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)

$$\big\|\mathbf{A}_T^{-1}\widehat{\mathbf{P}} - \mathbf{A}_T^{-1}\widehat{\mathbf{P}}^0\mathbf{J}_r\big\| = O_p(\vartheta_{nT,\delta}), \tag{A64}$$

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

Then, from Lemmas A5(ii) and A10, (A62), and Slutsky's theorem, as $n, T \to \infty$, we have (see also Lemma 13.1 in Johansen, 1995)

$$\det\left[\mathbf{A}'_T\big(\widehat{\mu}_j\widehat{\mathbf{S}}_{11} - \widehat{\mathbf{S}}_{10}\widehat{\mathbf{S}}_{00}^{-1}\widehat{\mathbf{S}}_{01}\big)\mathbf{A}_T\right] = \det\left[\mathbf{A}'_T\big(\widehat{\mu}_j^0\mathbf{S}_{11} - \mathbf{S}_{10}\mathbf{S}_{00}^{-1}\mathbf{S}_{01}\big)\mathbf{A}_T\right] + O_p(\vartheta_{nT,\delta}) \tag{A65}$$

$$\xrightarrow{d} \det\left(\widehat{\mu}_j^0\check{\boldsymbol{\Omega}}_{\check{\beta}\check{\beta}} - \check{\boldsymbol{\Omega}}_{\check{\beta}0}\check{\boldsymbol{\Omega}}_{00}^{-1}\check{\boldsymbol{\Omega}}_{0\check{\beta}}\right) \det\left[\widehat{\mu}_j^0\check{\boldsymbol{\beta}}'_{\perp*}\big(\boldsymbol{\Gamma}_{L0}^{\Delta F}\big)^{1/2}\left(\int_0^1 \mathbf{W}_r(\tau)\mathbf{W}'_r(\tau)\mathrm{d}\tau\right)\big(\boldsymbol{\Gamma}_{L0}^{\Delta F}\big)^{1/2}\check{\boldsymbol{\beta}}_{\perp*}\right].$$

where $\mathbf{W}_r(\cdot)$ is an r-dimensional Brownian motion with covariance of rank $q - d = r - c$. The first term on the rhs of (A65) 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 (A64), implies that the two spaces coincide asymptotically

$$\left\|\mathbf{A}_T^{-1}\widehat{\boldsymbol{\beta}} - \mathbf{A}_T^{-1}\widehat{\boldsymbol{\beta}}^0\mathbf{J}_c\right\| = O_p(\vartheta_{nT,\delta}). \tag{A66}$$

where $\mathbf{J}_c$ is a $c \times c$ diagonal matrix with entries 1 or -1, different from $\mathbf{J}$ and $\mathbf{J}_r$.

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

$$\widetilde{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}(\check{\boldsymbol{\beta}}_*'\widehat{\boldsymbol{\beta}})^{-1} = \check{\boldsymbol{\beta}} + \check{\boldsymbol{\beta}}_{\perp*}\check{\boldsymbol{\beta}}_\perp'\widetilde{\boldsymbol{\beta}}, \qquad \widetilde{\boldsymbol{\beta}}^0 = \widehat{\boldsymbol{\beta}}^0(\check{\boldsymbol{\beta}}_*'\widehat{\boldsymbol{\beta}}^0)^{-1} = \check{\boldsymbol{\beta}} + \check{\boldsymbol{\beta}}_{\perp*}\check{\boldsymbol{\beta}}_\perp'\widetilde{\boldsymbol{\beta}}^0. \tag{A67}$$

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

$$\mathbf{A}_T^{-1}\widetilde{\boldsymbol{\beta}}^0 = \mathbf{A}_T^{-1}(\check{\boldsymbol{\beta}} + \check{\boldsymbol{\beta}}_{\perp*}\check{\boldsymbol{\beta}}_\perp'\widetilde{\boldsymbol{\beta}}^0) = \begin{pmatrix} \mathbf{I}_c \\ \sqrt{T}\check{\boldsymbol{\beta}}_\perp'\widetilde{\boldsymbol{\beta}}^0 \end{pmatrix} = \begin{pmatrix} \mathbf{I}_c \\ \sqrt{T}\check{\boldsymbol{\beta}}_\perp'(\widetilde{\boldsymbol{\beta}}^0 - \check{\boldsymbol{\beta}}) \end{pmatrix} = \begin{pmatrix} \mathbf{I}_c \\ o_p(1) \end{pmatrix}, \tag{A68}$$

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

$$\mathbf{A}_T^{-1}\widetilde{\boldsymbol{\beta}} = \begin{pmatrix} \mathbf{I}_c \\ \sqrt{T}\check{\boldsymbol{\beta}}_\perp'\widetilde{\boldsymbol{\beta}} \end{pmatrix} = \begin{pmatrix} \mathbf{I}_c \\ \sqrt{T}\check{\boldsymbol{\beta}}_\perp'(\widetilde{\boldsymbol{\beta}} - \check{\boldsymbol{\beta}}) \end{pmatrix} = \begin{pmatrix} \mathbf{I}_c \\ \sqrt{T}\check{\boldsymbol{\beta}}_\perp'(\widetilde{\boldsymbol{\beta}}^0 - \check{\boldsymbol{\beta}}) + \sqrt{T}\check{\boldsymbol{\beta}}_\perp'(\widetilde{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}}_0) \end{pmatrix}. \tag{A69}$$

Now since $\mathrm{sp}(\mathbf{A}_T^{-1}\widetilde{\boldsymbol{\beta}}) = \mathrm{sp}(\mathbf{A}_T^{-1}\widehat{\boldsymbol{\beta}})$, also (A69) spans a space of dimension c. Then by comparing (A68) and (A69), and using (A66), 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

$$\left\|\sqrt{T}\check{\boldsymbol{\beta}}_\perp'(\widetilde{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}}_0)\right\| = \left\|\mathbf{A}_T^{-1}\widetilde{\boldsymbol{\beta}} - \mathbf{A}_T^{-1}\widetilde{\boldsymbol{\beta}}^0\right\| = O_p(\vartheta_{nT,\delta}). \tag{A70}$$

Therefore, given that $\|\check{\boldsymbol{\beta}}_\perp'\| = O(1)$ and given (A68) and (A70), we have

$$\left\|\widetilde{\boldsymbol{\beta}} - \check{\boldsymbol{\beta}}\right\| \le \left\|\widetilde{\boldsymbol{\beta}}^0 - \check{\boldsymbol{\beta}}\right\| + \left\|\widetilde{\boldsymbol{\beta}}^0 - \widetilde{\boldsymbol{\beta}}\right\| = o_p\left(\frac{1}{\sqrt{T}}\right) + O_p\left(\frac{\vartheta_{nT,\delta}}{\sqrt{T}}\right). \tag{A71}$$

From (A67), 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 the choice of the identification matrix $\mathbf{Q}$). Therefore, we have

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

which completes the proof of part (i).

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

$$\widehat{\boldsymbol{\alpha}} = \widehat{\mathbf{S}}_{01}\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\beta}}'\widehat{\mathbf{S}}_{11}\widehat{\boldsymbol{\beta}})^{-1}, \qquad \widehat{\mathbf{G}}_1 = (\widehat{\mathbf{M}}_{02} - \widehat{\boldsymbol{\alpha}}\widehat{\boldsymbol{\beta}}'\widehat{\mathbf{M}}_{12})\widehat{\mathbf{M}}_{22}^{-1}. \tag{A72}$$

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

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

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

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

Then,

$$\big\|\widehat{\mathbf{S}}_{01}\widehat{\boldsymbol{\beta}} - \check{\boldsymbol{\Omega}}_{0\check{\beta}}\mathbf{Q}\big\| \leq \big\|\widehat{\mathbf{S}}_{01}(\widehat{\boldsymbol{\beta}} - \check{\boldsymbol{\beta}}\mathbf{Q})\big\| + \big\|\widehat{\mathbf{S}}_{01}\check{\boldsymbol{\beta}}\mathbf{Q} - \mathbf{S}_{01}\check{\boldsymbol{\beta}}\mathbf{Q}\big\| + \big\|\mathbf{S}_{01}\check{\boldsymbol{\beta}}\mathbf{Q} - \check{\boldsymbol{\Omega}}_{0\check{\beta}}\mathbf{Q}\big\| = O_p(\vartheta_{nT,\delta}),$$
$$(A73)$$

using part (i) and the fact that $\|\widehat{\mathbf{S}}_{01}\| = O_p(T^{1/2})$ for the first term on the rhs, Lemma A9(iv) for the second term, and Lemma A10(iii) for the third term. Analogously we have

$$\big\|\widehat{\boldsymbol{\beta}}'\widehat{\mathbf{S}}_{11}\widehat{\boldsymbol{\beta}} - \mathbf{Q}'\check{\boldsymbol{\Omega}}_{\check{\beta}\check{\beta}}\mathbf{Q}\big\| \leq \big\|(\widehat{\boldsymbol{\beta}}' - \mathbf{Q}'\check{\boldsymbol{\beta}}')\widehat{\mathbf{S}}_{11}(\widehat{\boldsymbol{\beta}} - \check{\boldsymbol{\beta}}\mathbf{Q})\big\| + \big\|\mathbf{Q}'\check{\boldsymbol{\beta}}'\widehat{\mathbf{S}}_{11}\check{\boldsymbol{\beta}}\mathbf{Q} - \mathbf{Q}'\check{\boldsymbol{\beta}}'\mathbf{S}_{11}\check{\boldsymbol{\beta}}\mathbf{Q}\big\|$$
$$+ \big\|\mathbf{Q}'\check{\boldsymbol{\beta}}'\mathbf{S}_{11}\check{\boldsymbol{\beta}}\mathbf{Q} - \mathbf{Q}'\check{\boldsymbol{\Omega}}_{\check{\beta}\check{\beta}}\mathbf{Q}\big\| = O_p(\vartheta_{nT,\delta}),$$
$$(A74)$$

using part (i) and the fact that $\|\widehat{\mathbf{S}}_{11}\| = O_p(T)$ for the first term, Lemma A9(ii) for the second term, and Lemma A10(ii) for the third term. Therefore, from (A72), (A73), and (A74), 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:

$$\check{\mathbf{G}}_1 = \mathbf{H}\mathbf{G}_1\mathbf{H}' = \big(\boldsymbol{\Gamma}_1^{\Delta\check{F}} - \check{\boldsymbol{\alpha}}\mathsf{E}[\check{\boldsymbol{\beta}}'\check{\mathbf{F}}_{t-1}\Delta\check{\mathbf{F}}'_{t-1}]\big)(\boldsymbol{\Gamma}_0^{\Delta\check{F}})^{-1}.$$
$$(A75)$$

Then, from (A72),

$$\big\|\widehat{\mathbf{G}}_1 - \check{\mathbf{G}}_1\big\| \leq \big\|(\widehat{\mathbf{M}}_{02} - \widehat{\boldsymbol{\alpha}}\widehat{\boldsymbol{\beta}}'\widehat{\mathbf{M}}_{12})\widehat{\mathbf{M}}_{22}^{-1} - (\widehat{\mathbf{M}}_{02} - \check{\boldsymbol{\alpha}}\check{\boldsymbol{\beta}}'\widehat{\mathbf{M}}_{12})\widehat{\mathbf{M}}_{22}^{-1}\big\|$$
$$+ \big\|(\widehat{\mathbf{M}}_{02} - \check{\boldsymbol{\alpha}}\check{\boldsymbol{\beta}}'\widehat{\mathbf{M}}_{12})\widehat{\mathbf{M}}_{22}^{-1} - (\mathbf{M}_{02} - \check{\boldsymbol{\alpha}}\check{\boldsymbol{\beta}}'\mathbf{M}_{12})\mathbf{M}_{22}^{-1}\big\|$$
$$+ \big\|(\mathbf{M}_{02} - \check{\boldsymbol{\alpha}}\check{\boldsymbol{\beta}}'\mathbf{M}_{12})\mathbf{M}_{22}^{-1} - (\boldsymbol{\Gamma}_1^{\Delta\check{F}} - \check{\boldsymbol{\alpha}}\mathsf{E}[\check{\boldsymbol{\beta}}'\check{\mathbf{F}}_{t-1}\Delta\check{\mathbf{F}}'_{t-1}])(\boldsymbol{\Gamma}_0^{\Delta\check{F}})^{-1}\big\| = O_p(\vartheta_{nT,\delta}),$$

since the first term on the rhs is $O_p(\vartheta_{nT,\delta})$ by parts (i) and (ii) 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 A8(i), A8(iv) and A8(vii), and the third term is $O_p(T^{-1/2})$ by Lemmas A1 and A5(vi), and in particular by (A12) and (A45). This, together with (A75), proves part (iii).

Finally, for part (iv), first notice that the sample covariance of the VECM 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 also written as

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

Then from parts (i), (ii) and (iii), Lemma A8(ii) through A8(vii), and Lemma A5(i) and A5(vi), we can prove that

$$\left\|\widehat{\mathbf{\Gamma}}_0^w - \mathbf{J}\mathbf{\Gamma}_0^w\mathbf{J}\right\| = O_p(\vartheta_{nT,\delta}), \tag{A76}$$

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

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 4, we have $\mathbf{\Gamma}_0^w = \mathbf{K}\mathbf{K}'$. Moreover, from Assumption 1 and the model given in (8), $\mathbf{K} = \mathbf{C}(0) = \mathbf{Q}(0)$ has rank q and so $\mathbf{\Gamma}_0^w$ has also rank q. We denote as μ_j^w the eigenvalues of $\mathbf{\Gamma}_0^w$, thus $\mu_j^w = 0$ if and only if $j > q$. These are also eigenvalues of $\mathbf{J}\mathbf{\Gamma}_0^w\mathbf{J}$. As a consequence, having defined as $\widehat{\mu}_j^w$ the eigenvalues of $\widehat{\mathbf{\Gamma}}_0^w$, from (A76) and Weyl's inequality (A3), we have

$$\left|\widehat{\mu}_j^w - \mu_j^w\right| \leq \left\|\widehat{\mathbf{\Gamma}}_0^w - \mathbf{J}\mathbf{\Gamma}_0^w\mathbf{J}\right\| = O_p(\vartheta_{nT,\delta}), \qquad j = 1, \ldots, q. \tag{A77}$$

If we denote by $\mathbf{W}_q$ the $r \times q$ matrix of non-zero normalised eigenvectors of $\mathbf{\Gamma}_0^w$, then $\mathbf{J}\mathbf{W}_q$ are the normalised eigenvectors of $\mathbf{J}\mathbf{\Gamma}_0^w\mathbf{J}$. We denote as $\widehat{\mathbf{W}}_q$ the $r \times q$ matrix of normalised eigenvectors of $\widehat{\mathbf{\Gamma}}_0^w$. Then, from (A76) by Theorem 2 in Yu et al. (2015), we can prove that

$$\left\|\widehat{\mathbf{W}}_q - \mathbf{J}\mathbf{W}_q\mathbf{J}_q\right\| = O_p(\vartheta_{nT,\delta}), \tag{A78}$$

where $\mathbf{J}_q$ is a diagonal $q \times q$ matrix with entries 1 or -1, different from $\mathbf{J}$. Notice that $\mathbf{J}\mathbf{W}_q\mathbf{J}_q$ are also normalised eigenvectors of $\mathbf{J}\mathbf{\Gamma}_0^w\mathbf{J}$. From, the definition of $\widehat{\mathbf{K}} = \widehat{\mathbf{W}}_q\widehat{\mathbf{D}}_q^{-1/2}$ and (A77) and (A78), we have

$$\left\|\widehat{\mathbf{K}} - \mathbf{J}\mathbf{W}_q\mathbf{J}_q\mathbf{D}_q^{-1/2}\right\| = O_p(\vartheta_{nT,\delta}), \tag{A79}$$

where $\mathbf{D}_q$ is a diagonal matrix with entries μ_j^w for $j = 1, \ldots, q$ 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 (A79), we have the result. Notice that $\mathbf{K}'\mathbf{\Gamma}_0^w\mathbf{K} = \mathbf{I}_q$ as requested by Assumption 1(a) of orthonormality of the shocks. This completes the proof. $\qquad\square$

A.5   Proof of Proposition 1

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 4(i), 4(ii) and 4(iii), we have, for $k = 1, 2$,

$$\left\|\widehat{\mathbf{A}}_k^{\text{VECM}} - \mathbf{J}\mathbf{A}_k\mathbf{J}\right\| = O_p(\vartheta_{nT,\delta}). \tag{A80}$$

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 (A80), we have, for any $k \geq 0$,

$$\left\|\widehat{\mathbf{B}}_k - \mathbf{J}\mathbf{B}_k\mathbf{J}\right\| = O_p(\vartheta_{nT,\delta}). \tag{A81}$$

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

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

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 IRFs, i.e. at most only on $q(q+1)/2$ variables. So $\widehat{\mathbf{R}}$ depends only on $q(q+1)/2$ rows of $\widehat{\boldsymbol{\Lambda}}$ and for regular identification schemes, such that this mapping is analytical, using Lemmas 3(i) and 4(iv), and (A80), we have (see Forni et al., 2009)

$$\left\| \widehat{\mathbf{R}} - \mathbf{R} \right\| = O_p(\vartheta_{nT,\delta}). \tag{A82}$$

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

$$\left\| \widehat{\boldsymbol{\lambda}}_i' - \boldsymbol{\lambda}_i' \mathbf{H}' \right\| = O_p\left(\max\left(\frac{1}{\sqrt{T}}, \frac{1}{\sqrt{n}} \right) \right). \tag{A83}$$

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

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

where we used (A81), (A82), and (A83), orthogonality of $\mathbf{R}$, $\mathbf{J} \mathbf{J} = \mathbf{I}_r$, and the fact that $\mathbf{R}$, $\mathbf{K}$, $\mathbf{B}_k$, $\boldsymbol{\lambda}_i$ are all finite dimensional matrices with norm that does not depend on n or T. By (17) it is clear that the upper bound on the rate of convergence is $\vartheta_{nT,\delta}$. This completes the proof. $\qquad \square$

A.6 Proof of Lemma 5

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 $\mathbf{F}_t$, 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 $\mathbf{F}_t$, as $\widehat{\mathbf{A}}_1^{\text{1VAR}}$, $\widehat{\mathbf{Q}}_0$, and $\widehat{\mathbf{Q}}_1$. Then,

$$\left\| \widehat{\mathbf{Q}}_0 - \mathbf{Q}_0 \right\| = \left\| \left(\frac{1}{T} \sum_{t=1}^{T} \boldsymbol{\beta}' \mathbf{F}_{t-1} \mathbf{u}_t' \mathbf{K}' \boldsymbol{\beta} \right) \left(\frac{1}{T} \sum_{t=1}^{T} \boldsymbol{\beta}' \mathbf{F}_{t-1} \mathbf{F}_{t-1}' \boldsymbol{\beta} \right)^{-1} \right\| = O_p\left(\frac{1}{\sqrt{T}} \right). \tag{A84}$$

Indeed, the first term on the rhs is $O_p(T^{-1/2})$ from (A38) and by independence of $\mathbf{u}_t$ in Assumption

4(a), while the second term is $O_p(1)$ by Lemma A5(v). Similarly,

$$\left\|\widehat{\mathbf{Q}}_1 - \mathbf{Q}_1\right\| = \left\|\left(\frac{1}{T^2}\sum_{t=1}^T \boldsymbol{\beta}_\perp' \mathbf{F}_{t-1}\mathbf{u}_t'\mathbf{K}'\boldsymbol{\beta}_\perp\right)\left(\frac{1}{T^2}\sum_{t=1}^T \boldsymbol{\beta}_\perp' \mathbf{F}_{t-1}\mathbf{F}_{t-1}'\boldsymbol{\beta}_\perp\right)^{-1}\right\| = O_p\left(\frac{1}{T}\right). \quad (A85)$$

Indeed, the first term on the rhs is $O_p(T^{-1})$ from (A38) and by independence of $\mathbf{u}_t$ in Assumption 4(a), while the second term is $O_p(1)$ by Lemma A5(ii). Moreover,

$$\mathrm{vec}\big(\widehat{\mathbf{A}}_1^{1\mathrm{VAR}}\big) = (\mathbf{I}_r \otimes \boldsymbol{\mathcal{D}}')\begin{pmatrix} \mathrm{vec}(\widehat{\mathbf{Q}}_0) \\ \mathrm{vec}(\widehat{\mathbf{Q}}_1) \end{pmatrix}. \quad (A86)$$

Analogous formulas to (A84)-(A86) are in Theorem 1 by Sims et al. (1990) and, by combining them,

$$\left\|\widehat{\mathbf{A}}_1^{1\mathrm{VAR}} - \mathbf{A}_1\right\| = O_p\left(\frac{1}{\sqrt{T}}\right). \quad (A87)$$

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 $\widehat{\mathbf{A}}_1^{0\mathrm{VAR}}$ the ordinary least square estimator for the VAR when using $\mathbf{JF}_t$, then $\widehat{\mathbf{A}}_1^{0\mathrm{VAR}} = \mathbf{J}\widehat{\mathbf{A}}_1^{1\mathrm{VAR}}\mathbf{J}$, and from (A87)

$$\left\|\widehat{\mathbf{A}}_1^{0\mathrm{VAR}} - \mathbf{J}\mathbf{A}_1\mathbf{J}\right\| = O_p\left(\frac{1}{\sqrt{T}}\right). \quad (A88)$$

Define

$$\widehat{\mathbf{M}}_{1L} = \frac{1}{T}\sum_{t=1}^T \widehat{\mathbf{F}}_t\widehat{\mathbf{F}}_{t-1}', \qquad \widehat{\mathbf{M}}_{LL} = \frac{1}{T}\sum_{t=1}^T \widehat{\mathbf{F}}_{t-1}\widehat{\mathbf{F}}_{t-1}'. \quad (A89)$$

Then, we can write the VAR estimators as

$$\widehat{\mathbf{A}}_1^{\mathrm{VAR}} = \frac{\widehat{\mathbf{M}}_{1L}}{T}\left(\frac{\widehat{\mathbf{M}}_{LL}}{T}\right)^{-1}, \quad \widehat{\mathbf{A}}_1^{0\mathrm{VAR}} = \frac{\mathbf{M}_{1L}}{T}\left(\frac{\mathbf{M}_{LL}}{T}\right)^{-1}, \quad (A90)$$

where $\mathbf{M}_{1L}$ and $\mathbf{M}_{LL}$ are defined as in (A89), but when using $\mathbf{JF}_t$.

Because of Lemma A8(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

$$\left\|\widehat{\mathbf{A}}_1^{\mathrm{VAR}} - \widehat{\mathbf{A}}_1^{0\mathrm{VAR}}\right\| = O_p\left(\max\left(\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{T}}\right)\right). \quad (A91)$$

By combining (A91) with (A88)

$$\left\|\widehat{\mathbf{A}}_1^{\mathrm{VAR}} - \mathbf{J}\mathbf{A}_1\mathbf{J}\right\| \le \left\|\widehat{\mathbf{A}}_1^{\mathrm{VAR}} - \widehat{\mathbf{A}}_1^{0\mathrm{VAR}}\right\| + \left\|\widehat{\mathbf{A}}_1^{0\mathrm{VAR}} - \mathbf{J}\mathbf{A}_1\mathbf{J}\right\| = O_p\left(\max\left(\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{T}}\right)\right), \quad (A92)$$

which completes the proof of part (i).

By noticing that, as a consquence of part (i), (A76) holds also in this case, but with the rate given in (A92), we prove part (iii) exactly as in Lemma 4(iv). This completes the proof. $\qquad\square$

A.7 Proof of Proposition 2

Define

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

such that $\widehat{\mathbf{B}}_k = (\widehat{\mathbf{A}}_1^{\mathrm{VAR}})^k$. Then, from Lemma 5(i), we have, for any finite $k \geq 0$,

$$\left\|\widehat{\mathbf{B}}_k - \mathbf{J}\mathbf{B}_k\mathbf{J}\right\| = O_p\left(\max\left(\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{T}}\right)\right). \tag{A93}$$

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

$$\lim_{k \to \infty}\|\widehat{\mathbf{B}}_k - \mathbf{B}_k\| = O_p(1). \tag{A94}$$

The estimated impulse response of variable i is then the q-dimensional row vector (see (24))

$$\widehat{\boldsymbol{\phi}}_i^{\mathrm{VAR}'}(L) = \widehat{\boldsymbol{\lambda}}_i'\widehat{\mathbf{B}}(L)\widehat{\mathbf{K}}\widehat{\mathbf{R}}', \tag{A95}$$

where $\widehat{\boldsymbol{\lambda}}_i'$ is the i-th row of $\widehat{\boldsymbol{\Lambda}}$ and $\widehat{\mathbf{R}} \equiv \widehat{\mathbf{R}}(\widehat{\boldsymbol{\Lambda}}, \widehat{\mathbf{A}}^{\mathrm{VAR}}(L), \widehat{\mathbf{K}})$ is a consistent estimator of the matrix $\mathbf{R}$, such that, because of Lemmas 3(i) and 5(i) (see also the proof of Proposition 1),

$$\|\widehat{\mathbf{R}} - \mathbf{R}\| = O_p\left(\max\left(\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{T}}\right)\right). \tag{A96}$$

Notice that (A96) is true provided that we do not consider long-run restrictions as identification schemes, since in that case $\mathbf{R}$ would be a function of $\mathbf{A}^{\mathrm{VAR}}(1)$ and it is not consistently estimated because of (A94). Consistency of the estimated IRFs (A95), at each finite lag k, is then proved exactly as in the proof of Proposition 1. $\qquad\square$

A.8 Proof of Lemma 6

For any $i = 1, \ldots, n$, recall that we defined $x_{it} = a_i + \boldsymbol{\lambda}_i'\mathbf{F}_t + \xi_{it}$ so that $y_{it} = b_i t + x_{it}$. The proof of part (i) is straightforward since it amounts to using the sample mean as an estimator of the mean of the stationary and ergodic process Δy_{it}.

For part (ii), 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 (27) we have

$$\widehat{b}_i - b_i = \frac{\sum_{t=0}^{T}(t - \frac{T}{2})(y_{it} - \bar{y}_i)}{\sum_{t=0}^{T}(t - \frac{T}{2})^2} - b_i = \frac{\sum_{t=0}^{T}(t - \frac{T}{2})(x_{it} - \bar{x}_i)}{\sum_{t=0}^{T}(t - \frac{T}{2})^2} = \frac{\sum_{t=0}^{T}tx_{it} - \frac{T}{2}\sum_{t=0}^{T}x_{it}}{\sum_{t=0}^{T}t^2 - \frac{T^2(T+1)}{4}}. \tag{A97}$$

The denominator of (A97) is $O(T^3)$. For the numerator, consider first the case in which $x_{it} \sim I(1)$, then under Assumptions 4(a) and 4(c) of serial independence of the shocks, 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 (A97) 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. $\qquad \square$

A.9 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 $\mathrm{rk}(\mathbf{C}(e^{-i\theta})) = q$ a.e. in $[-\pi, \pi]$, then it has q non-zero real eigenvalues and $r - q$ zero eigenvalues. Notice also that we have $\mathrm{rk}(\mathbf{C}(e^{-i\theta})) \leq q$ for any $\theta \in [-\pi, \pi]$. Moreover, given square summability of the coefficients of $\mathbf{C}(L)$ as a consequence of Assumption 1(a), the non-zero eigenvalues are also finite for any $\theta \in [-\pi, \pi]$. Thus, by denoting as $\mu_j^{\Delta F}(\theta)$ such eigenvalues, there exist positive reals $\underline{M}_{10}$ and $\overline{M}_{10}$ such that a.e. in $[-\pi, \pi]$

$$\underline{M}_{10} \leq \mu_j^{\Delta F}(\theta) \leq \overline{M}_{10}, \quad j = 1, \ldots, q. \tag{A98}$$

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 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 eigenvalues.

Define $\boldsymbol{L}(\theta) = \mathbf{\Lambda} \mathbf{W}^{\Delta F}(\theta) (\mathbf{M}^{\Delta F}(\theta))^{1/2}$ for any $\theta \in [-\pi, \pi]$. 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, since because of (14), $n^{-1} \mathbf{\Lambda}' \mathbf{\Lambda} = \mathbf{I}_r$

$$\frac{\overline{\boldsymbol{L}'(\theta)} \boldsymbol{L}(\theta)}{n} = \mathbf{M}^{\Delta F}(\theta), \quad \theta \in [-\pi, \pi]. \tag{A99}$$

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 (A99), 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 $j = 1, \ldots, r$. Part (i) then follows from (A98).

As for part (ii), from Assumption 3(b), for any $\theta \in [-\pi, \pi]$, there exists a positive real M_1

such that

$$\sup_{i\in\mathbb{N}}\left|\check{d}_i(e^{-i\theta})\right| \leq \sup_{i\in\mathbb{N}}\left|\sum_{k=0}^{\infty}\check{d}_{ik}e^{-ik\theta}\right| \leq \sup_{i\in\mathbb{N}}\sum_{k=0}^{\infty}\left|\check{d}_{ik}\right| \leq M_1. \tag{A100}$$

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

$$\sup_{\theta\in[-\pi,\pi]}\left\|\boldsymbol{\Sigma}^{\Delta\xi}(\theta)\right\|_1 = \sup_{\theta\in[-\pi,\pi]}\max_{i=1,\ldots,n}\sum_{j=1}^{n}|\sigma_{ij}(\theta)| = \sup_{\theta\in[-\pi,\pi]}\max_{i=1,\ldots,n}\frac{1}{2\pi}\sum_{j=1}^{n}\left|\check{d}_i(e^{-i\theta})\mathsf{E}[\varepsilon_{it}\varepsilon_{jt}]\,\check{d}_j(e^{i\theta})\right|$$

$$\leq \frac{M_1^2}{2\pi}\max_{i=1,\ldots,n}\sum_{j=1}^{n}|\mathsf{E}[\varepsilon_{it}\varepsilon_{jt}]| \leq \frac{M_1^2 M_4}{2\pi}, \tag{A101}$$

where we used (A100) and Assumption 4(e). From (A2) and (A101), we have, for any $n\in\mathbb{N}$,

$$\sup_{\theta\in[-\pi,\pi]}\mu_1^{\Delta\xi}(\theta) = \sup_{\theta\in[-\pi,\pi]}\left\|\boldsymbol{\Sigma}^{\Delta\xi}(\theta)\right\| \leq \sup_{\theta\in[-\pi,\pi]}\left\|\boldsymbol{\Sigma}^{\Delta\xi}(\theta)\right\|_1 \leq \frac{M_1^2 M_4}{2\pi}, \tag{A102}$$

and part (ii) is proved by defining $M_{11} = M_1^2 M_4 (2\pi)^{-1}$.

Finally, parts (iii) and (iv), are immediate consequences of Assumption 3(d), which implies that $\boldsymbol{\Sigma}^{\Delta x}(\theta) = \boldsymbol{\Sigma}^{\Delta\chi}(\theta) + \boldsymbol{\Sigma}^{\Delta\xi}(\theta)$, for any $\theta\in[-\pi,\pi]$, and of Weyl's inequality (A3). So, for $j = 1,\ldots,q$, and for any $n\in\mathbb{N}$ and a.e. in $[-\pi,\pi]$, there exist positive reals $\underline{M}_{12}$ and $\overline{M}_{12}$ such that

$$\frac{\mu_j^{\Delta x}(\theta)}{n} \leq \frac{\mu_j^{\Delta\chi}(\theta)}{n} + \frac{\mu_1^{\Delta\xi}(\theta)}{n} \leq \overline{M}_{10} + \sup_{\theta\in[-\pi,\pi]}\frac{\mu_1^{\Delta\xi}(\theta)}{n} \leq \overline{M}_{10} + \frac{M_{11}}{n} = \overline{M}_{12},$$

$$\frac{\mu_j^{\Delta x}(\theta)}{n} \geq \frac{\mu_j^{\Delta\chi}(\theta)}{n} + \frac{\mu_n^{\Delta\xi}(\theta)}{n} \geq \underline{M}_{10} + \inf_{\theta\in[-\pi,\pi]}\frac{\mu_n^{\Delta\xi}(\theta)}{n} = \underline{M}_{12}.$$

because of parts (i) and (ii). This proves part (iii). When $j = q+1$, using parts (i) and (ii), and since $\mathrm{rk}(\boldsymbol{\Sigma}^{\Delta\chi}(\theta)) \leq q$, for any $\theta\in[-\pi,\pi]$, we have $\mu_{q+1}^{\Delta x}(\theta) \leq \mu_{q+1}^{\Delta\chi}(\theta) + \mu_1^{\Delta\xi(\theta)} = \mu_1^{\Delta\xi(\theta)} \leq M_{11}$, thus proving part (iv).

Finally, for part (v) consider parts (iii) and (iv) but when $\theta = 0$. Then, $\mathrm{rk}(\boldsymbol{\Sigma}^{\Delta\chi}(0)) = \tau \leq q$ which implies $\underline{M}_{10} \leq n^{-1}\mu_\tau^{\Delta\chi}(0) \leq \overline{M}_{10}$, but $\mu_{\tau+1}^{\Delta\chi}(0) = 0$. Using again parts (i) and (ii) and Weyl's inequality (A3), prove part (v). This completes the proof. $\square$

A.10 Proof of Proposition 3

The proof follows Proposition 2 in Hallin and Liška (2007) when fixing $\theta = 0$, combined with Lemma 7 and the results about spectral density estimation in **?**. $\square$

B   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