EINAUDI INSTITUTE FOR ECONOMICS AND FINANCE

EIEF WORKING PAPER SERIES

EIEF Working Paper 23/12

October 2023

Approximating Singular by Means of Non-Singular Structural VARs

By

Mario Forni

(Università di Modena e Reggio Emilia, CEPR and RECent)

Marco Lippi

(Einaudi Institute for Economics and Finance)

Approximating Singular by Means of Non-Singular Structural VARs

Mario Forni[*]
Università di Modena e Reggio Emilia, CEPR and RECent

Marco Lippi[†]
Einaudi Institute for Economics and Finance, Rome

October 4, 2023

Abstract

In applications of Dynamic Factor Models to Structural Macroeconomic Analysis, r, the number of static factors, is typically larger than q, the number of shocks driving the macro economy, so that the spectral density matrix of the factors is singular. Singularity is an important advantage with respect to standard Structural VARs, because it ensures that generically the Structural Shocks are fundamental and the factors have a finite VAR representation in the Structural Shocks. However, a serious difficulty with this approach is that singular VARs are not necessarily unique. We show that, despite this, the Structural Shocks and the corresponding Impulse-Response Functions are approximated consistently using a non-singular VAR.

JEL classification: C32, E32.
Keywords: Structural VAR models, Common-Components Structural VARs

[*]E-mail: mario.forni@unimore.it.
[†]E-mail: mlippi.eief@gmail.com.

1 Introduction

1.1 Preliminaries

The High-Dimensional Dynamic Factor Model has been widely used in the last two decades in the analysis and prediction of macroeconomic time series. A brief summary of the model, in the version adopted here, is the following:
(a) $\{x_{it}\}$ and $\{\xi_{it}\}$ are sequences of stochastic processes indexed by $i \in \mathbb{N}$,
(b) $\{F_t\}$ is an r-dimensional weakly-stationary stochastic process with rational spectral density $h(\theta)$ of rank $q \leq r$,
(c) $G(L)$ is a rational $r \times q$ matrix with no poles of modulus less or equal to unity, such that $G(e^{-i\theta})G'(e^{i\theta}) = h(\theta)$, $\{w_t\}$ is a q-dimensional orthonormal white noise, and Λ_i, $i \in \mathbb{N}$, an r-dimensional vector.

We assume that:

$$x_{it} = \chi_{it} + \xi_{it} = \Lambda_i F_t + \xi_{it} \tag{1}$$
$$F_t = G(L)w_t, \tag{2}$$

for all $i \in \mathbb{N}$, $t \in \mathbb{Z}$. The variables ξ_{it} and χ_{it} are called the idiosyncratic and common components of x_{it}, respectively; F_t and w_t are called the vector of static common factors and dynamic common factors respectively.

Moreover:

(i) The idiosyncratic components are orthogonal to the dynamic factors at all leads and lags, i.e. $\xi_{it} \perp w_s$ for all $t, s \in \mathbb{Z}$ and $i \in \mathbb{N}$. As a consequence, $\xi_{it} \perp \chi_{js}$ for all $t, s \in \mathbb{Z}$, $i, j \in \mathbb{N}$.

(ii) The n-dimensional process $\{(\xi_{1t}\ \xi_{2t}\ \cdots\ \xi_{nt})'\}$ is weakly stationary for all $n \in \mathbb{N}$. This and the previous assumption imply that $\{(x_{1t}\ x_{2t}\ \cdots\ x_{nt})'\}$ is weakly stationary for all $n \in \mathbb{N}$.

(iii) The idiosyncratic components are weakly correlated, which, by definition, means that the first eigenvalue of the covariance matrix of $\{(\xi_{1t}\ \xi_{2t}\ \cdots\ \xi_{nt})'\}$ is bounded as $n \to \infty$.

(iv) The common components are pervasive, which, by definition, means that the first r eigenvalues of the covariance matrix of $\{(\chi_{1t}\ \chi_{2t}\ \cdots\ \chi_{nt})'\}$ diverge as $n \to \infty$. As shown in Chamberlain (1983), this assumption and (iii) are equivalent to the statement that, as $n \to \infty$, the first r eigenvalues of the covariance matrix of $\{(x_{1t}\ x_{2t}\ \cdots\ x_{nt})'\}$ diverge whereas the $(r+1)$-th stays bounded.

Starting with the observed sample $\{x_{it},\ i = 1, \ldots, n,\ t = 1, \ldots, T\}$, the standard estimator of each common component χ_{jt} is based on the first r sample principal components. The latter are obtained using the $n \times n$ sample covariance matrix of the x's. Consistency (convergence in probability), as both n and T tend to infinity, of these principal-component estimators has been proved, under additional technical assumptions not specified here, in several papers, see in particular Stock and Watson (2002), Bai and Ng (2002), Forni et al. (2009).

In the present paper we assume that the population covariances of the vector process $\{(x_{1t}\ x_{2t}\ \cdots\ x_{nt})'\}$, as well as the integers r and q, are known. The same approach is taken e.g. in Chamberlain (1983), Forni and Lippi (2001), Anderson and Deistler (2008), Hallin and Lippi (2013), Forni et al. (2015), Lippi et al. (2023). Based on such covariances, we construct *unfeasible* principal-component estimators of the common components χ_{it} and use them throughout the paper. The line of reasoning of the present paper is used as guidance in Forni et al. (2023), where feasible estimators, based on sample covariances and estimated r and q, are employed.

The estimator of the common components is obtained as follows. Let $\hat{\mathbf{P}}_t$ be the r-dimensional vector whose coordinates are the first r population principal components of $\mathbf{x}_{nt} = (x_{1t}\ \cdots\ x_{nt})'$:

$$\hat{\mathbf{P}}_t = \hat{W}\mathbf{x}_{nt},$$

where $\hat{W}$ is the $r \times n$ matrix whose rows are the eigenvectors corresponding to the first r eigenvalues of the population covariance matrix of $\mathbf{x}_{nt}$, normalized such that the covariance matrix of $\left\{\hat{\mathbf{P}}_t\right\}$ is the identity. Then, given m, the (unfeasible) principal-component estimator of the m-dimensional vector $\boldsymbol{\chi}_{mt} = (\chi_{1t}\ \cdots\ \chi_{mt})'$ is

$$\hat{\boldsymbol{\chi}}_{mt} = \hat{W}'_{[m]}\hat{\mathbf{P}}_t = \hat{W}'_{[m]}\hat{W}\mathbf{x}_{nt},$$

where $\hat{W}_{[m]}$ is the $r \times m$ matrix obtained by truncating $\hat{W}$ at the m-th column. It is easily seen that $\hat{\chi}_{jt}$, the j-th coordinate of $\hat{\boldsymbol{\chi}}_{mt}$, is the projection of χ_{jt} on the r-dimensional linear space spanned by $\hat{\mathbf{P}}_{nt}$.

Assuming that, as $n \to \infty$, the first r eigenvalues of the covariance matrix of the x's diverge whereas the $(r+1)$-th stays bounded, see (iv) above, $\hat{\boldsymbol{\chi}}_{mt}$ converges to $\boldsymbol{\chi}_{mt}$, as $n \to \infty$, in mean square, see Chamberlain (1983), Forni and Lippi (2001). Using (1),

$$\hat{\boldsymbol{\chi}}_{mt} = \hat{W}'_{[m]}\hat{W}\mathbf{x}_{nt} = \hat{W}'_{[m]}\hat{W}\boldsymbol{\chi}_{nt} + \hat{W}'_{[m]}\hat{W}\boldsymbol{\xi}_{nt}. \tag{3}$$

Using (1) and (2),

$$\hat{W}'_{[m]}\hat{W}\boldsymbol{\chi}_{nt} = \hat{W}'_{[m]}\hat{W}\Lambda_{[n]}G(L)w_t = \hat{D}_{[m]}(L)w_t, \tag{4}$$

where $\Lambda_{[n]} = (\Lambda'_1\ \cdots\ \Lambda'_n)'$ and $\hat{D}_{[m]}(L) = \hat{W}'_{[m]}\hat{W}\Lambda_{[n]}G(L)$ is $m \times q$ with rational entries. As $\hat{\boldsymbol{\chi}}_{mt}$ converges to $\boldsymbol{\chi}_{mt}$ in mean square, orthogonality of the two addenda on the right in (3), see (i) above, implies that in mean square

$$\hat{D}_{[m]}(L)w_t \to \boldsymbol{\chi}_{mt}, \qquad \hat{W}'_{[m]}\hat{W}\boldsymbol{\xi}_{nt} \to 0. \tag{5}$$

1.2 The contribution of the present paper

Assuming that the observable variables x_{it} are macroeconomic indicators, the idiosyncratic components can be interpreted as variable-specific causes of variation plus measurement errors, whereas the common components are the "true" variables, driven by the macroeconomic shocks. Given a vector of interest $\mathbf{x}_{mt}$, based on the interpretation above, several papers have explored the possibility of replacing $\mathbf{x}_{mt}$ with $\boldsymbol{\chi}_{mt}$, or augmenting $\mathbf{x}_{mt}$ with factors, for macroeconomic analysis, see in particular the Structural Dynamic Factor Model studied in Stock and Watson (2005), Bai and Ng (2007) and Forni et al. (2009), and the Factor Augmented VAR in Bernanke et al. (2005).

In addition to cleaning the variables from measurement errors, using the χ's instead of the x's has another important advantage, namely that the fundamentalness problem, a serious issue in a VAR model for the x's, has a solution. For, using (1) and (2), we find that $\boldsymbol{\chi}_{mt} = D_{[m]}(L)w_t$, where $D_{[m]}(L)$ is $m \times q$ with rational entries. Moreover, as motivated in several papers, see Barigozzi et al. (2021) for detailed references, we can assume $r > q$, i.e. that the number of static factors is greater than the number of dynamic factors, so that the stochastic process $\{F_t\}$ is *singular*.

We assume that m, the dimension of the vector of interest $\boldsymbol{\chi}_{mt}$, fulfills $r \geq m > q$, so that $\{\boldsymbol{\chi}_{mt}\}$ is singular.

Now, leaving details to Section 2:

(I) Anderson and Deistler (2008) show that singularity and rationality of $D_{[m]}(L)$ imply that $\boldsymbol{\chi}_{mt}$ has a finite-length VAR representation for generic values of the parameters of the entries of $D_{[m]}(L)$ and that $D_{[m]}(0)w_t$ is the innovation of $\boldsymbol{\chi}_{mt}$.

(II) Thus generically, the white-noise vector w_t is fundamental for $\boldsymbol{\chi}_{mt}$.

On the other hand, what is available is not $\boldsymbol{\chi}_{mt}$ but the approximation $\hat{\boldsymbol{\chi}}_{mt}$, and therefore a VAR for $\hat{\boldsymbol{\chi}}_{mt}$. If $\{\boldsymbol{\chi}_{mt}\}$ were non-singular, it would be fairly trivial to prove that the VAR polynomial matrix for $\hat{\boldsymbol{\chi}}_{mt}$ consistently approximates the VAR polynomial matrix for $\boldsymbol{\chi}_{mt}$. However, singularity implies that the VAR representation of $\boldsymbol{\chi}_{mt}$ is not necessarily unique, since the regressors can be linearly dependent, see Section 2 for an illustration. This has been pointed out firstly in Anderson and Deistler (2008) and thoroughly analyzed in Filler (2010), Deistler et al. (2010), Anderson et al. (2012). As a consequence, the VAR polynomial matrix of $\hat{\boldsymbol{\chi}}_{mt}$ does not necessarily converge to a VAR polynomial matrix for $\boldsymbol{\chi}_{mt}$.

This difficulty, which has been overlooked so far in the papers using the Structural Dynamic Factor Model or the Factor Augmented VAR (see in particular the above mentioned papers: Stock and Watson (2005), Bai and Ng (2007), Forni et al. (2009), Bernanke et al. (2005)), is solved here, though only for the unfeasible estimator $\hat{\boldsymbol{\chi}}_{mt}$ defined above, which is based on population covariances. In Section 3 we prove that, even if the VAR polynomial matrix does not converge, the residual of the VAR for $\hat{\boldsymbol{\chi}}_{mt}$ converges in mean square to the innovation of $\boldsymbol{\chi}_{mt}$.

In Section 4 we prove that the Moving-Average representation of $\hat{\boldsymbol{\chi}}_{mt}$ converges to that of $\boldsymbol{\chi}_{mt}$. In Section 5 we prove consistency of the Structural Shocks and Impulse-Response Functions identified with a recursive scheme.

As mentioned above, the integers r and q and the unfeasible estimators of $\boldsymbol{\chi}_{mt}$, its Structural Shocks and Impulse-Response Functions, are replaced in Forni et al. (2023) by feasible estimators, based on sample covariances. Using the results of the present paper, consistency and rates of convergence in probability, as n and T tend to infinity, of the Structural Shocks and Impulse-Response Functions are obtained.

2 Existence but non-uniqueness of VARs in the singular case

The results of the present paper depend on some properties of the vectors $\boldsymbol{\chi}_{mt}$ and $\hat{\boldsymbol{\chi}}_{mt}$, not on the factor model (1)–(2) *per se*. We adopt therefore a more general setting and notation.

All the stochastic processes considered are weakly stationary of *constant rank*, this meaning that they have a spectral density matrix of the same rank a.e. in $[-\pi, \ \pi]$, see Rozanov (1967), pp. 39-43. The rank of a process with rational spectral density is of course constant. An s-dimensional constant-rank stochastic process is (dynamically) singular if its rank is less than s, non-singular if its rank is s. If the process $\{y_t\}$ is singular, its covariance matrix is not necessarily singular, see the process (11) with $b_1 \neq b_2$. If the covariance matrix of $\{y_t\}$ is singular then of course $\{y_t\}$ is (dynamically) singular.

Let $\{\chi_t\}$ be an m-dimensional weakly stationary process fulfilling the ARMA equation:

$$H(L)\chi_t = K(L)w_t, \tag{6}$$

where:

(a) $\{w_t\}$ is a q-dimensional orthonormal white-noise process with $m \geq q$, $K(L)$ is an $m \times q$ polynomial matrix

(b) $H(L)$ is an $m \times m$ polynomial matrix, with $H(0) = I_m$, fulfilling the stability condition, i.e. $\det H(z) = 0$ implies $|z| > 1$, Thus χ_t has the Moving Average representation

$$\chi_t = B(L)w_t = H(L)^{-1}K(L)w_t. \tag{7}$$

Assumption 1. *We suppose that $m > q$, so that $\{\chi_t\}$ is singular.*

Anderson and Deistler (2008) show that:
(A) Under Assumption 1, supposing that the coefficients of the polynomial entries of $K(L)$ vary independently of one another, for generic values of such parameters

the matrix $K(L)$ is *zeroless*, that is, $K(z)$ has full rank q for all $z \in \mathbb{C}$. To illustrate this statement, consider the simplest example, in which $m = 2$, $q = 1$ and

$$K(L) = \begin{pmatrix} 1 - b_1 L \\ 1 - b_2 L \end{pmatrix}.$$

We see that if $b_1 \neq b_2$, the rank of $K(z)$ is one, the maximum, for all $z \in \mathbb{C}$. Thus $K(L)$ is zeroless except for the lower-dimensional, negligible, set $b_1 = b_2$. If we consider instead the non-singular matrix

$$\tilde{K}(L) = \begin{pmatrix} 1 - b_1 L & 1 - b_3 L \\ 1 - b_2 L & 1 - b_4 L \end{pmatrix},$$

zerolessness implies that $b_1 b_4 - b_2 b_3 = 0$ and $b_1 + b_4 - b_2 - b_3 = 0$. Thus $\tilde{K}(L)$ is zeroless only for a lower-dimensional set.

(B) If $K(L)$ is zeroless, there exists an $m \times m$ (finite) polynomial matrix $K^\dagger(L)$ such that (i) $K^\dagger(L)$ is stable and $K^\dagger(0) = I_m$, (ii) $K^\dagger(L)K(L) = K(0)$. We say that $K^\dagger(L)$ is a left inverse of $K(L)$. Setting $A(L) = K^\dagger(L)H(L) = I_m - A_1 L - \cdots - A_h L^p$, χ_t has the finite-length VAR representation

$$(I_m - A_1 L - \cdots - A_h L^p)\chi_t = A(L)\chi_t = K_0 w_t, \tag{8}$$

where $K_0 = K(0) = B(0)$ has rank q. Setting $\varepsilon_t = K_0 w_t$, because (7) implies that ε_t is orthogonal to χ_{t-j} for all $j \in \mathbb{N}$, then

$$\chi_t = (A_1 \chi_{t-1} + \cdots + A_p \chi_{t-p}) + \varepsilon_t = \mathcal{P}_t + \varepsilon_t \tag{9}$$

is the unique decomposition of χ_t into the projection of χ_t on its whole past and its innovation ε_t.

The results (A) and (B) say that generically: (I) The polynomial $K(L)$ in (6) has a finite left inverse, so that, most importantly, (II) the white noise vector w_t in (6) is fundamental for χ_t. Of course neither (I) nor (II) hold generically for non-singular ARMAs.

Observation 1. *The assumption in point (A) above, that "the coefficients of the polynomial entries of $K(L)$ vary independently of one another" is obviously false in some important cases. It immediately comes to mind the case*

$$\Delta X_t = K(L) w_t, \tag{10}$$

where $K(L)$ has rational entries and X_t is cointegrated for all values of the coefficients of $K(L)$, so that generically $K(z)$ has a zero at $z = 1$. A discussion of cointegration for singular vectors is outside of the scope of the present paper. Let us mention however Deistler and Wagner (2017), Barigozzi et al. (2020, 2021), in which the results in Anderson and Deistler (2008) are adapted to cointegrated singular vectors. In particular, Barigozzi et al. (2020) show that generically singular cointegrated vectors with representation (10) *have a VAR representation in the levels (precisely, an Error Correction representation) with a finite-degree matrix polynomial. See also Forni et al. (2023), in which the consequences of the failure of the independent-coefficients assumption is discussed in general.*

These important features of singular ARMAs do not come without a difficulty. Let us firstly recall that in the non-singular case, i.e. when $m = q$ and the spectral density of χ_t is non-singular almost everywhere in $[-\pi,\ \pi]$, no more than one VAR representation (finite or infinite) may exist. On the contrary, when χ_t is singular, as firstly pointed out in Anderson and Deistler (2008), representation (8) is not necessarily unique, that is, there may exist a stable polynomial matrix $A'(L) = I'_m - A'L - \cdots - A'_{p'}L^{p'} \neq A(L)$, such that

$$\chi_t = (A'_1\chi_{t-1} + \cdots + A'_{p'}\chi_{t-p'}) + \varepsilon'_t = \mathcal{P}'_t + \varepsilon'_t,$$

where ε'_t is orthogonal to χ_{t-j} for all $j \in \mathbb{N}$. Uniqueness of the orthogonal projection of χ_t on the linear space spanned by χ_{t-j}, $j \in \mathbb{N}$, implies that $\mathcal{P}'_t = \mathcal{P}_t$, and $\varepsilon'_t = \varepsilon_t$, so that the alternative representation becomes

$$\chi_t = (A'_1\chi_{t-1} + \cdots + A'_{p'}\chi_{t-p'}) + \varepsilon_t = \mathcal{P}_t + \varepsilon_t.$$

As an example consider the singular ARMA:

$$\begin{aligned} \chi_{1t} &= (1 + b_1L)w_t \\ \chi_{2t} &= (1 + b_2L)w_t. \end{aligned} \tag{11}$$

where w_t is a scalar white noise. We have $b_2\chi_{1t-1} - b_1\chi_{2t-1} = (b_2 - b_1)w_{t-1}$. If $b_1 \neq b_2$, thus generically, replacing w_{t-1} in (11) with $(b_2 - b_1)^{-1}(b_2\chi_{1t-1} - b_1\chi_{2t-1})$, we obtain a VAR(1) for χ_t. But the variables χ_{1t-1} and χ_{2t-1} fulfill the exact dynamic relationship $(1 + b_2L)\chi_{1t-1} = (1 + b_1L)\chi_{2t-1}$, so that e.g.

$$\chi_{1t-1} = (1 + b_1L)\chi_{2t-1} - b_2\chi_{1t-2},$$

which can be used to obtain an alternative VAR(2) representation.

Note that in example (11), as χ_{1t} and χ_{2t} are linearly independent, there is only one VAR(1) representation, thus there is uniqueness if p is minimum. But with slightly more complex models, even assuming that p in (8) is minimum, i.e. that if $p' < p$ then no autoregressive representation of length p' exists, we see that (8) is not necessarily unique. For this purpose let us simplify (6) by assuming that $H(L) = I_m$, so that

$$\chi_t = B_0w_t + \ldots + B_kw_{t-k}, \quad k > 0. \tag{12}$$

Suppose that: (I) χ_t has the VAR representation (8), (II) the mp stochastic variables $\chi_{i,t-j}$, $i = 1, \ldots, m$, $j = 1, \ldots p$, are linearly independent, so that (8) is unique among the VAR representations of length $p' \leq p$. The space spanned by the variables $\chi_{i,t-j}$, $i = 1, \ldots, m$, $j = 1, \ldots p$, call it $\mathcal{S}_\chi$, has dimension mp. Then consider the space $\mathcal{S}_w$ spanned by $w_{i,t-j}$, $i = 1, \ldots, q$, $j = 1, \ldots, p + k$, whose dimension is $q(p + k)$. By (12) $\mathcal{S}_\chi \subseteq \mathcal{S}_w$, so that $mp \leq q(p + k)$, that is

$$p \leq \frac{kq}{m - q}. \tag{13}$$

Now suppose for example that $m = 3$, $k = 3$ and $q = 1$. In this case, by (13), $p \leq 3/2$, so that uniqueness implies that $p = 1$. Combining (12) for $k = 3$ and (8) for $p = 1$, we have

$$(I_3 - A_1 L)\chi_t = (I_3 - A_1 L)(B_0 + B_1 L + B_2 L^2 + B_3 L^3)w_t = B_0 w_t.$$

Multiplying by w'_{t-j}, $j = 1, \ldots, 4$, and taking expected values:

$$A_1 B_3 = 0, \quad A_1 B_2 = B_3, \quad A_1 B_1 = B_2, \quad A_1 B_0 = B_1. \tag{14}$$

Suppose that B_1, B_2 and B_3 are linearly independent. Then the last three equations in (14) say that Range(A_1) has dimension 3, and therefore that Null(A_1) has dimension 0. This implies, by the first equation in (14), that $B_3 = 0$, which is contradictory with linear independence of B_1, B_2 and B_3. But such independence is generic, so that generically no VAR of order 1 exists for χ_t, which implies that the minimum-length VAR is not unique.

What happens in the example just considered is that the variables $\chi_{j,t-1}$, $j = 1, 2, 3$ are not sufficient for a VAR, whereas the variables $\chi_{j,t-k}$, $j = 1, 2, 3$, $k = 1, 2$, are sufficient but linearly dependent. Hence the necessity of a VAR(2) but the non-uniqueness. Following Deistler et al. (2011) we might select a basis in the space spanned by $\chi_{j,t-k}$, $j = 1, 2, 3$, $k = 1, 2$, and obtain a unique representation. This line of reasoning is not pursued in the present paper and we stick to the standard VAR specification, in which entire blocks of lagged χ's are added or removed.

It is important to point out that, as both the examples show, the mere choice of a p greater than the minimum integer for which a VAR exists causes non-uniqueness.

Based on Assumption 1, (A), (B) and the above considerations on uniqueness of VAR representations in the singular case:

Assumption 2. *We assume that stable autoregressive representations of the form (8) exist for χ_t and denote by $\tilde{p}$ the minimum order of their autoregressive polynomials.*

Then we consider the approximation

$$\hat{\chi}_t = \hat{B}(L)w_t + \hat{\mu}_t, \tag{15}$$

where $\hat{\chi}_t$, the $m \times q$ matrix $\hat{B}(L)$ and the m-dimensional vector $\hat{\mu}_t$ depend on $n \in \mathbb{N}$.

Assumption 3.

(i) For all $n \in \mathbb{N}$, the m-dimensional vector process $\{\hat{\mu}_t\}$ is weakly stationary with constant rank,

(ii) For all $n \in \mathbb{N}$ and $k \in \mathbb{Z}$, $w_t \perp \hat{\mu}_{t-k}$.

(iii) As $n \to \infty$, $\hat{\chi}_t \to \chi_t$ in mean square, so that, by (ii), $\hat{\mu}_t \to 0$ and $\hat{B}(L)w_t \to \chi_t$ in mean square.

Note that we are not assuming that $\{\hat{\mu}_t\}$ has rational spectral density. As a consequence the spectral density of $\{\hat{\chi}_t\}$ is not assumed to be rational. Definition (15) and Assumption 3 are obviously fulfilled by the principal-component unfeasible estimator $\hat{\boldsymbol{\chi}}_{mt}$ defined in the Introduction, with

$$\hat{B}(L)w_t = \hat{W}'_{[m]}\hat{W}_{[n]}\mathbf{\Lambda}_{[n]}G(L)w_t, \quad \hat{\mu}_t = \hat{W}'_{[m]}\hat{W}_{[n]}\boldsymbol{\xi}_{nt},$$

see equations (3) and (4).

For $j \geq 1$, we define

$$Z_{t,j} = (\chi'_t,\ \chi'_{t-1},\ \dots,\ \chi'_{t-j+1})', \quad \hat{Z}_{t,j} = (\hat{\chi}'_t,\ \hat{\chi}'_{t-1},\ \dots,\ \hat{\chi}'_{t-j+1})'.$$

Assumption 4. *The m-dimensional process $\{\hat{\mu}_t\}$ is non-singular for all $n \in \mathbb{N}$. This implies that for all $n \in \mathbb{N}$ and $j \geq 1$, the covariance matrix of $\left\{\hat{Z}_{t,j}\right\}$ is non-singular.*

To prove the implication, by Assumption 3(ii),

$$\hat{f}(\theta) = \hat{f}_1(\theta) + \hat{f}_2(\theta),$$

where f, f_1 and f_2 are the spectral densities of $\hat{\chi}_t$, $\hat{B}(L)w_t$ and $\hat{\mu}_t$, respectively. Now, singularity of $\{\hat{Z}_{t,j}\}$ means that there exist $1 \times m$ matrices g_h, $h = 0, \cdots, j-1$, not all zero, such that $g_0\hat{\chi}_t + \cdots + g_{j-1}\hat{\chi}_{t-j+1} = 0$, that is, $g(L)\hat{\chi}_t = 0$, where $g(L)$ is the non-zero $1 \times m$ polynomial matrix $g_0 + \cdots + g_{j-1}L^{j-1}$. This implies that

$$g(e^{-i\theta})f(\theta)g'(e^{i\theta}) = g(e^{-i\theta})f_1(\theta)g'(e^{i\theta}) + g(e^{-i\theta})f_2(\theta)g'(e^{i\theta}) = 0, \tag{16}$$

for θ a.e. in $[-\pi,\ \pi]$. Because $\{\hat{\mu}_t\}$ is non-singular for all $n \in \mathbb{N}$, $f_2(\theta)$ is positive definite for θ a.e. in $[-\pi,\ \pi]$. Thus (16) is possible only for $g(L) = 0$.

Lastly, consider a VAR for $\hat{\chi}_t$,

$$\hat{\chi}_t = \hat{A}_1\hat{\chi}_{t-1} + \cdots + \hat{A}_{\hat{p}}\hat{\chi}_{t-\hat{p}} + \hat{\epsilon}_t = \hat{\mathcal{P}}_t + \hat{\epsilon}_t, \tag{17}$$

that is the projection of $\hat{\chi}_t$ on the space spanned by $\hat{\chi}_{i,t-k}$, $i = 1, \dots, m$ and $k = 1, \dots, \hat{p}$, not on the whole past of $\hat{\chi}_t$. As a consequence, in general $\hat{\epsilon}_t$ is neither the innovation of $\hat{\chi}_t$ nor a white-noise vector. By Assumption 4, given $\hat{p}$, the matrices $\hat{A}_s$, $s = 1, \dots, \hat{p}$, are unique.

Like the population covariances and the integers r and q, $\tilde{p}$ is supposed to be known and $\hat{p}$ is any integer independent of n, fulfilling the inequality in Assumption 5 below. Problems arising when r, q and $\tilde{p}$ are estimated are dealt with in Forni et al. (2023).

Assumption 5. *The order of the VAR in* (17) *is not less than the minimum $\tilde{p}$ (as defined in Assumption 2), i.e. $\hat{p} \geq \tilde{p}$.*

3 Consistency of $\hat{\mathcal{P}}_t$ and $\hat{\epsilon}_t$

By Assumption 5 the representation of $\mathcal{P}_t$ in (9) can be conveniently rewritten up to the lag $\hat{p}$:

$$\mathcal{P}_t = A_1\chi_{t-1} + \cdots + A_{\hat{p}}\chi_{t-\hat{p}}.$$

As this representation is not necessarily unique, asking if $\hat{A}_j$ converges to A_j or not does not make sense. However, $\mathcal{P}_t$ and ε_t are unique and the following Proposition 1 states that even if the matrices $\hat{A}_j$ do not converge, $\hat{\mathcal{P}}_t$ and $\hat{\varepsilon}_t$ converge to $\mathcal{P}_t$ and ε_t in mean square.

Observation 2. *The following considerations and example should convince the reader that Proposition 1 is not trivial. Let Y_n, X_i and X_{in}, $i = 1, \ldots, s$, $n \in \mathbb{N}$, be stochastic variables. Suppose that (i) as $n \to \infty$, $Y_n \to Y$ and $X_{in} \to X_i$, $i = 1, \ldots, s$, in mean square, (ii) the vector $(X_1 \ \cdots \ X_s)$ has non-singular covariance matrix. Then it is fairly obvious that the projection of Y_n on the variables X_{in}, $i = 1, \ldots, s$, converges in mean square to the projection of Y on the variables X_i:*

$$\text{Proj}(Y_n|X_{1n}, \ \ldots, \ X_{sn}) \to \text{Proj}(Y|X_1, \ \ldots, \ X_s). \tag{18}$$

However, if the covariance matrix of $(X_1 \ \cdots \ X_s)$ is singular, then (18) does not necessarily hold. An elementary example is the following. Suppose that $Y_n = Y$, that $X_n \to 0$ in mean square and $X_n \neq 0$ for all $n \in \mathbb{N}$. Then

$$\text{proj}(Y|X_n) = \frac{\sigma_{YX_n}}{\sigma^2_{X_n}} X_n = \rho_{YX_n} \frac{X_n}{\sigma_{X_n}}, \tag{19}$$

where ρ_{YX_n} is the correlation $\sigma_{YX_n}/(\sigma_Y\sigma_{X_n})$. Now suppose that $X_n = \alpha_n(Y + Z)$, where $\alpha_n \neq 0$, $\alpha_n \to 0$ and $Y \perp Z$. We have

$$\rho_{YX_n} = \frac{\sigma_Y}{\sqrt{\sigma^2_Y + \sigma^2_Z}},$$

so that the projection in (19) does not tend to zero, whereas the projection of Y on 0, which is the limit of X_n, is 0. We see that in this case the additional assumption that $\rho_{YX_n} \to 0$ is necessary. In Proposition 1 the regressors may tend to singularity. This is why Assumption 3(ii) is crucial at the end the proof.

Proposition 1. *Under Assumptions 1 through 5,*

$$\hat{\mathcal{P}}_t = \hat{A}_1\hat{\chi}_{t-1} + \cdots + \hat{A}_{\hat{p}}\hat{\chi}_{t-\hat{p}} \to \mathcal{P}_t = A_1\chi_{t-1} + \cdots + A_{\hat{p}}\chi_{t-\hat{p}}, \quad \hat{\epsilon}_t \to \epsilon_t.$$

Proof. As $\hat{p}$ does not change, we simplify $Z_{t,\hat{p}}$ and $\hat{Z}_{t,\hat{p}}$ into Z_t and $\hat{Z}_t$, respectively. Let d be the static rank of Z_t, i.e. the rank of the covariance matrix of $\{Z_t\}$. Without loss of generality, suppose that the first d coordinates of Z_t, gathered in the d-dimensional vector Ω_{1t}, form a basis in the space spanned by Z_t, which is denoted by $H_{Z,t}$. Then orthonormalize by the Gram-Schmidt procedure such basis. The i-th recursive projection produces a non-zero residual for $i = 1, \ldots, d$. Call η_t such residuals after normalization. Starting with the $(d+1)$-th, we regress the Z's only on η_t and obtain a zero residual:

$$Z_t = \begin{pmatrix} \Omega_{1t} \\ \Omega_{2t} \end{pmatrix} = \begin{pmatrix} M_{d\times d} & 0_{d\times(m\hat{p}-d)} \\ N_{(m\hat{p}-d)\times d} & 0_{(m\hat{p}-d)\times(m\hat{p}-d)} \end{pmatrix} \begin{pmatrix} \eta_t \\ 0_{(m\hat{p}-d)\times 1} \end{pmatrix}. \tag{20}$$

Of course η_t is an orthonormal basis in $H_{Z,t}$. The matrix M is lower triangular and non-singular. The lower-right matrix is set to zero for convenience. Then apply the the Gram-Schmidt procedure to $\hat{Z}_t$.

$$\hat{Z}_t = \begin{pmatrix} \hat{\Omega}_{1t} \\ \hat{\Omega}_{2t} \end{pmatrix} = \begin{pmatrix} \hat{M}_{d\times d} & 0_{d\times(m\hat{p}-d)} \\ \hat{N}_{(m\hat{p}-d)\times d} & \hat{Q}_{(m\hat{p}-d)\times(m\hat{p}-d)} \end{pmatrix} \begin{pmatrix} \hat{\eta}_t \\ \hat{\vartheta}_t \end{pmatrix}, \tag{21}$$

where, as $\left\{\hat{Z}_t\right\}$ is non-singular by Assumption 4, $(\hat{\eta}_t,\ \hat{\vartheta}_t)$ is an orthonormal basis in $\left\{H_{\hat{Z},t}\right\}$. By Assumption 3(iii), $\hat{Z}_t \to Z_t$ in mean square, so that the covariance matrices of $\{\hat{Z}_t\}$, $\hat{\Omega}_{1t}$, $\hat{\Omega}_{2t}$, converge to the covariance matrices of Z_t, Ω_{1t}, Ω_{2t}, respectively. We have:

(a) The entries of $\hat{M}$ are well defined continuous functions of the entries of the covariance matrix $\mathrm{E}(\hat{\Omega}_{1t}\hat{\Omega}'_{1t})$. Thus $\hat{M} \to M$ and $\hat{\eta}_t = \hat{M}^{-1}\hat{\Omega}_{1t} \to M^{-1}\Omega_{1t} = \eta_t$.

(b) $\hat{N} = \mathrm{E}(\hat{\Omega}_{2t}\hat{\Omega}'_{1t})\hat{M}'^{-1} \to \mathrm{E}(\Omega_{2t}\Omega'_{1t})M'^{-1} = N$, so that $\hat{N}\hat{\eta}_t \to N\eta_t$.

(c) From

$$\hat{\Omega}_{2t} - \Omega_{2t} = \left[\hat{N}\hat{\eta}_t - N\eta_t\right] + \hat{Q}\hat{\vartheta}_t$$

we have

$$\hat{Q}\hat{\vartheta}_t = \left[\hat{\Omega}_{2t} - \Omega_{2t}\right] - \left[\hat{N}\hat{\eta}_t - N\eta_t\right] \to 0. \tag{22}$$

Orthonormality of $\hat{\vartheta}_t$ implies that $\hat{Q} \to 0$.

The projections of $\hat{\chi}_t$ and χ_t on $\hat{Z}_{t-1}$ and Z_{t-1}, respectively, are

$$\begin{aligned}
\hat{\chi}_t &= \mathcal{P}(\hat{\chi}_t |\ \hat{Z}_{t-1}) + \hat{\varepsilon}_t = \mathcal{P}(\hat{\chi}_t |\ \hat{\eta}_{t-1}) + \mathcal{P}(\hat{\chi}_t |\ \hat{\vartheta}_{t-1}) + \hat{\varepsilon}_t \\
&= \hat{\alpha}\hat{\eta}_{t-1} + \hat{\beta}\hat{\vartheta}_{t-1} + \hat{\varepsilon}_t \\
\chi_t &= \alpha\eta_{t-1} + \varepsilon_t,
\end{aligned}$$

where $\hat{\varepsilon}_t$ is orthogonal to $\hat{\eta}_{t-1}$ and to $\hat{\vartheta}_{t-1}$. We have

$$(\hat{\chi}_t - \chi_t) - (\hat{\alpha}\hat{\eta}_{t-1} - \alpha\eta_{t-1}) = \hat{\beta}\hat{\vartheta}_{t-1} + (\hat{\varepsilon}_t - \varepsilon_t)\,. \tag{23}$$

Because $\hat{\eta}_{t-1} \to \eta_{t-1}$, see (a) above, $\hat{\alpha} = \mathrm{E}(\hat{\chi}_t\hat{\eta}'_{t-1}) \to \mathrm{E}(\chi_t\eta'_{t-1}) = \alpha$ and the left-hand side of (23) tends to zero, so that

$$\hat{\beta}\hat{\vartheta}_{t-1} + (\hat{\varepsilon}_t - \varepsilon_t) \to 0,$$

that is

$$
\begin{aligned}
\operatorname{Trace} \mathrm{E}(\hat\beta\hat\vartheta_{t-1}\hat\vartheta'_{t-1}\hat\beta') + \operatorname{Trace} \mathrm{E}((\hat\varepsilon_t-\varepsilon_t)(\hat\varepsilon_t-\varepsilon_t)') \\
+ 2\operatorname{Trace} \mathrm{E}\left((\hat\varepsilon_t-\varepsilon_t)\hat\vartheta'_{t-1}\hat\beta'\right) \to 0.
\end{aligned} \tag{24}
$$

We have

$$
\mathrm{E}\left((\hat\varepsilon_t-\varepsilon_t)\hat\vartheta'_{t-1}\hat\beta'\right) = \mathrm{E}\left(\hat\varepsilon_t\hat\vartheta'_{t-1}\hat\beta'\right) - \mathrm{E}\left(\varepsilon_t\hat\vartheta'_{t-1}\hat\beta'\right) = -\mathrm{E}\left(\varepsilon_t\hat\vartheta'_{t-1}\hat\beta'\right).
$$

Inverting the matrix in (21):

$$
\hat\beta\hat\vartheta_{t-1} = \hat\beta\begin{pmatrix}-\hat Q^{-1}\hat N\hat M^{-1} & \hat Q^{-1}\end{pmatrix}\hat Z_{t-1} = \hat\gamma\hat Z_{t-1},
$$

say. Using (15) and the definition of $\hat Z_t$,

$$
\hat\gamma\hat Z_{t-1} = \hat\delta_1(L)w_{t-1} + \hat\delta_2(L)\hat\mu_{t-1} = \hat{\mathcal{G}}_{t-1} + \hat{\mathcal{H}}_{t-1},
$$

say. Because w_t is white noise, $\varepsilon_t = B_0 w_t$ is orthogonal to $\hat{\mathcal{G}}_{t-1}$. By Assumption 3(ii), ε_t is also orthogonal to $\hat{\mathcal{H}}_{t-1}$. Thus the last term on the left in (24) is zero and

$$
\hat\alpha\hat\eta_{t-1} + \hat\beta\hat\vartheta_{t-1} \to \alpha\eta_{t-1}, \qquad \hat\varepsilon_t \to \varepsilon_t,
$$

which concludes the proof. □

4 Consistency of the Moving-Average representation of $\hat\chi_t$

Let us start by inverting the VAR representations $A(L)\chi_t = \varepsilon_t$ and its empirical counterpart $\hat A(L)\hat\chi_t = \hat\varepsilon_t$:

$$
\begin{aligned}
\chi_t &= A_1\chi_{t-1} + \cdots + A_{\hat p}\chi_{t-\hat p} + \varepsilon_t \\
\hat\chi_t &= \hat A_1\chi_{t-1} + \cdots + \hat A_{\hat p}\chi_{t-\hat p} + \hat\varepsilon_t.
\end{aligned} \tag{25}
$$

Iterate (25) h times:

$$
\begin{aligned}
\chi_t &= [\varepsilon_t + \mathcal{B}_1\varepsilon_{t-1} + \cdots + \mathcal{B}_h\varepsilon_{t-h}] + \left[\mathcal{A}_{h+1,1}\chi_{t-h-1} + \cdots + \mathcal{A}_{h+1,\hat p}\chi_{t-(h+\hat p)}\right] \\
&= [\varepsilon_t + \mathcal{B}_1\varepsilon_{t-1} + \cdots + \mathcal{B}_h\varepsilon_{t-h}] + \mathcal{H}_{h+1}Z_{t-h-1} \\
&= [\varepsilon_t + \mathcal{B}_1\varepsilon_{t-1} + \cdots + \mathcal{B}_h\varepsilon_{t-h}] + \alpha_{h+1}\eta_{t-h-1}
\end{aligned} \tag{26}
$$

$$
\hat\chi_t = \left[\hat\varepsilon_t + \hat{\mathcal{B}}_1\hat\varepsilon_{t-1} + \cdots + \hat{\mathcal{B}}_h\hat\varepsilon_{t-h}\right] + \left[\hat{\mathcal{A}}_{h+1,1}\hat\chi_{t-h-1} + \cdots + \hat{\mathcal{A}}_{h+1,\hat p}\hat\chi_{t-(h+\hat p)}\right] \tag{27}
$$

$$
\begin{aligned}
&= \left[\hat\varepsilon_t + \hat{\mathcal{B}}_1\hat\varepsilon_{t-1} + \cdots + \hat{\mathcal{B}}_h\hat\varepsilon_{t-h}\right] + \hat{\mathcal{H}}_{h+1}\hat Z_{t-h-1} \\
&= \left[\hat\varepsilon_t + \hat{\mathcal{B}}_1\hat\varepsilon_{t-1} + \cdots + \hat{\mathcal{B}}_h\hat\varepsilon_{t-h}\right] + \hat\alpha_{h+1}\hat\eta_{t-h-1} + \hat\beta_{h+1}\hat\vartheta_{t-h-1}.
\end{aligned} \tag{28}
$$

We know that $A(L)$ is not necessarily unique, so that the matrices $\mathcal{B}_j$ are not necessarily unique. However the term $\mathcal{B}_j\varepsilon_{t-j}$ is uniquely defined as the projection of χ_t on the linear space spanned by the coordinates of ε_{t-j}. We now prove that $\hat{\mathcal{B}}_j\hat\varepsilon_{t-j}$ converges to $\mathcal{B}_j\varepsilon_{t-j}$ in mean square.

Proposition 2. *Under Assumptions 1 through 5, for all $h \geq 0$,*

$$\hat{\alpha}_{h+1}\hat{\eta}_{t-h-1} \to \alpha_{h+1}\eta_{t-h-1}, \qquad \hat{\mathcal{B}}_h\hat{\varepsilon}_{t-h} \to \mathcal{B}_h\varepsilon_{t-h}. \tag{29}$$

Proof. We have proved in Proposition 1 that, setting $\mathcal{B}_0 = \hat{\mathcal{B}}_0 = I_m$,

$$\hat{\alpha}_1\hat{\eta}_{t-1} \to \alpha_1\eta_{t-1}, \qquad \hat{\mathcal{B}}_0\hat{\varepsilon}_t \to \mathcal{B}_0\varepsilon_t.$$

Thus the statement in (29) is true for $h = 0$. Now suppose that $h > 0$ and that for $j < h$,

$$\hat{\alpha}_{j+1}\hat{\eta}_{t-j-1} \to \alpha_{j+1}\eta_{t-j-1}, \qquad \hat{\mathcal{B}}_j\hat{\varepsilon}_{t-j} \to \mathcal{B}_j\varepsilon_{t-j}. \tag{30}$$

Let us prove that this implies that (29) is true. Multiply both sides of (28) by $\hat{\eta}'_{t-h-1}$:

$$\hat{\alpha}_{h+1} = \mathrm{E}\left(\hat{\chi}_t\hat{\eta}'_{t-h-1}\right) - \sum_{j=0}^{h-1} \mathrm{E}\left(\left[\hat{\mathcal{B}}_j\hat{\varepsilon}_{t-j}\right]\hat{\eta}'_{t-h-1}\right). \tag{31}$$

As $\hat{\chi}_t$, $\hat{\eta}_{t-h-1}$ and $\hat{\mathcal{B}}_j\hat{\varepsilon}_{t-j}$, for $j < h$, converge to χ_t, η_{t-h-1} and $\mathcal{B}_j\varepsilon_{t-j}$, respectively (the last by the inductive assumption),

$$\hat{\alpha}_{h+1} \to \mathrm{E}\left(\chi_t\eta'_{t-h-1}\right) - \sum_{j=0}^{h-1} \mathrm{E}\left(\left[\mathcal{B}_j\varepsilon_{t-j}\right]\eta'_{t-h-1}\right) = \alpha_{h+1}.$$

Because, again, $\hat{\eta}_{t-h-1} \to \eta_{t-h-1}$, the convergence on the left in (29) is proved by induction. To prove the convergence on the right, using (26) and (28),

$$\begin{aligned}(\hat{\chi}_t - \chi_t) - \sum_{j=0}^{h-1}\left(\hat{\mathcal{B}}_j\hat{\varepsilon}_{t-j} - \mathcal{B}_j\varepsilon_{t-j}\right) - \left(\hat{\alpha}_{h+1}\hat{\eta}_{t-h-1} - \alpha_{h+1}\eta_{t-h-1}\right)& \\ = \left(\hat{\mathcal{B}}_p\hat{\varepsilon}_{t-h} - \mathcal{B}_p\varepsilon_{t-h}\right) + \hat{\beta}_{h+1}\hat{\vartheta}_{t-h-1}&\end{aligned} \tag{32}$$

Each of the three vectors on the right tends to zero, so that

$$\begin{aligned}&\mathrm{Trace}\ \mathrm{E}\left((\hat{\mathcal{B}}_p\hat{\varepsilon}_{t-h} - \mathcal{B}_h\varepsilon_{t-h})(\hat{\mathcal{B}}_h\hat{\varepsilon}_{t-h} - \mathcal{B}_h\varepsilon_{t-h})'\right) + \mathrm{Trace}\ \mathrm{E}\left(\hat{\beta}_{h+1}\hat{\vartheta}_{t-h-1}\hat{\vartheta}'_{t-h-1}\hat{\beta}'_{h+1}\right)\\ &\quad + 2\,\mathrm{Trace}\ \mathrm{E}\left((\hat{\mathcal{B}}_p\hat{\varepsilon}_{t-h} - \mathcal{B}_p\varepsilon_{t-h})\hat{\vartheta}'_{t-h-1}\hat{\beta}'_{h+1}\right) \to 0.\end{aligned}$$

Regarding the third term above, note that

$$\left(\left[\hat{\mathcal{B}}_h\hat{\varepsilon}_{t-h}\right]\hat{\eta}'_{t-h-1}\right) = 0.$$

For, $\hat{\varepsilon}_{t-h}$ is the residual of the VAR for $\hat{\chi}_{t-h}$ and is therefore orthogonal to all the $\hat{\chi}$'s in the second term on the right in (27) and therefore to the $\hat{\eta}$'s and $\hat{\vartheta}$'s in (28). Thus

$$\mathrm{E}\left((\hat{\mathcal{B}}_p\hat{\varepsilon}_{t-h} - \mathcal{B}_p\varepsilon_{t-h})\hat{\vartheta}'_{t-h-1}\hat{\beta}'_{h+1}\right) = -\mathrm{E}\left(\mathcal{B}_h\varepsilon_{t-h}\hat{\vartheta}'_{t-h-1}\hat{\beta}'_{h+1}\right).$$

The argument used at the end of the proof of Proposition 1 shows that this covariance is zero and the second convergence in (29) is proved. $\square$

5 Consistency of the Impulse-Response Functions under recursive identification

The orthonormal white noise w_t in representation (6)-(7) is of course identified only up to an orthogonal matrix. Suppose now that, based on economic theory, we want to identify a q-dimensional orthonormal vector of shocks, call it u_t, recursively, that is, possibly by reordering the variables χ_{it}, imposing that the contemporaneous effect of u_{it} on χ_{jt} is zero if $j < i$ and non-zero if $j = i$, for $i = 2, \dots, q$. This is equivalent to:

Assumption 6. *We have:*

$$(\varepsilon_{1t}\ \varepsilon_{2t}\ \cdots\ \varepsilon_{qt})' = \mathcal{M} u_t,$$

where $\mathcal{M}$ is the unique $q \times q$ lower-triangular matrix, with positive entries on the main diagonal, such that $\mathcal{M}\mathcal{M}'$ is equal to the covariance matrix of $(\varepsilon_{1t}\ \varepsilon_{2t}\ \cdots\ \varepsilon_{qt})'$.

The matrix $\mathcal{M}$ can be obtained by applying the Gram-Schmidt procedure to ε_t:

$$\varepsilon_t = \mathcal{C}_0 \kappa_t = \begin{pmatrix} \mathcal{M} & 0 \\ \mathcal{N} & 0 \end{pmatrix} \begin{pmatrix} u_t \\ 0 \end{pmatrix}. \tag{33}$$

Doing the same for $\hat{\varepsilon}_t$, we obtain:

$$\hat{\varepsilon}_t = \hat{\mathcal{C}}_0 \hat{\kappa}_t = \begin{pmatrix} \hat{\mathcal{M}} & 0 \\ \hat{\mathcal{N}} & \hat{\mathcal{Q}} \end{pmatrix} \begin{pmatrix} \hat{u}_t \\ \hat{v}_t \end{pmatrix}, \tag{34}$$

where $\hat{\kappa}_t$ is orthonormal, $\hat{u}_t$ is q-dimensional, $\hat{v}_t$ is $(m-q)$-dimensional, $\hat{\mathcal{M}}$ and $\hat{\mathcal{Q}}$ are lower-triangular, $q \times q$ and $(m-q)\times(m-q)$, respectively. Because $\hat{\varepsilon}_t \to \varepsilon_t$, by the same arguments used to obtain (a), (b) and (c) in the proof of Proposition 1:
(A) $\hat{\mathcal{M}} \to \mathcal{M}$ and $\hat{u}_t \to u_t$.
(B) $\hat{\mathcal{N}} \to \mathcal{N}$, so that $\hat{\mathcal{N}}\hat{u}_t \to \mathcal{N} u_t$.
(C) $\hat{\mathcal{Q}}\hat{v}_t \to 0$, so that, as $\hat{v}_t$ is orthonormal, $\hat{\mathcal{Q}} \to 0$ and

$$\hat{\mathcal{C}}_0 \to \mathcal{C}_0. \tag{35}$$

Lastly, define

$$\hat{\mathcal{C}}_j = \hat{\mathcal{B}}_j \hat{\mathcal{C}}_0 = \begin{pmatrix} \hat{\mathcal{C}}_{11,h} & \hat{\mathcal{C}}_{12,h} \\ \hat{\mathcal{C}}_{21,h} & \hat{\mathcal{C}}_{22,h} \end{pmatrix}, \quad \mathcal{C}_h = \mathcal{B}_j \mathcal{C}_0 = \begin{pmatrix} \mathcal{C}_{11,h} & 0 \\ \mathcal{C}_{21,h} & 0 \end{pmatrix}, \tag{36}$$

Proposition 3. *Under Assumptions 1 through 6, for all $h \geq 0$,*

$$\hat{\mathcal{C}}_h \to \mathcal{C}_h.$$

Proof. By Proposition 2,

$$\hat{\mathcal{B}}_h \hat{\varepsilon}_{t-h} = \begin{pmatrix} \hat{\mathcal{C}}_{11,h} & \hat{\mathcal{C}}_{12,h} \\ \hat{\mathcal{C}}_{21,h} & \hat{\mathcal{C}}_{22,h} \end{pmatrix} \begin{pmatrix} \hat{u}_{t-h} \\ \hat{v}_{t-h} \end{pmatrix} \to \mathcal{B}_p \varepsilon_{t-h} = \begin{pmatrix} \mathcal{C}_{11,h} & 0 \\ \mathcal{C}_{21,h} & 0 \end{pmatrix} \begin{pmatrix} u_{t-h} \\ 0 \end{pmatrix}. \tag{37}$$

Thus:

$$\left(\hat{\mathcal{C}}_{11,h}\hat{u}_{t-h} - \mathcal{C}_{11,h}u_{t-h}\right) + \hat{\mathcal{C}}_{12,h}\hat{v}_{t-h} \to 0.$$

Multiplying by $\hat{u}'_{t-h}$ and taking expected values:

$$\left(\hat{\mathcal{C}}_{11,h} - \mathcal{C}_{11,h}\hat{I}_q\right) \to 0,$$

where $\hat{I}_q = \mathrm{E}(u_{t-h}\hat{u}'_{t-h})$. As $\hat{I}_q \to I_q$,

$$\hat{\mathcal{C}}_{11,h} \to \mathcal{C}_{11,h},$$

$\hat{\mathcal{C}}_{11,h}\hat{u}_{t-h} \to \mathcal{C}_{11,h}u_{t-h}$, so that $\hat{\mathcal{C}}_{12,h}\hat{v}_{t-h} \to 0$. As $\hat{v}_{t-h}$ is orthonormal,

$$\hat{\mathcal{C}}_{12,h} \to 0.$$

In the same way we prove that

$$\hat{\mathcal{C}}_{21,h} \to \mathcal{C}_{21,h}, \qquad \hat{\mathcal{C}}_{22,h} \to 0$$

and the proposition is proved. $\square$

6 Conclusions

The m-dimensional stationary vector χ_t is consistently approximated by $\hat{\chi}_t$. We assume that χ_t is singular and has an ARMA representation, whereas $\hat{\chi}_t$ is non-singular. Generically, χ_t has a finite but not unique finite VAR representation. As a consequence, the question whether the VAR polynomial for $\hat{\chi}_t$ converges is meaningless. However, we prove that the residual of the VAR polynomial for $\hat{\chi}_t$ converges to the innovation of χ_t and that the Impulse-Response Functions of $\hat{\chi}_t$, identified with a recursive scheme, converge to the corresponding Impulse-Response Functions of χ_t. Our proofs can be easily adapted to other identification schemes.

Though our results hold for any singular vector χ_t and estimator $\hat{\chi}_t$, fulfiling Assumptions 1 through 6, the paper has its main motivation in the factor-model based vector $\boldsymbol{\chi}_{mt}$ and the estimators of its Structural Shocks and Impulse-Response Functions. In the present paper such estimators are constructed using the unfeasible estimator $\hat{\boldsymbol{\chi}}_{mt}$ and are therefore themselves unfeasible. The analysis of the present paper has been used in Forni et al. (2023) as guidance to prove the consistency of feasible estimators of the Structural Shocks and Impulse-Response Functions of $\boldsymbol{\chi}_{mt}$.

References

Anderson, B. D. and M. Deistler (2008). Generalized linear dynamic factor models - A structure theory. In *47th IEEE Conference on Decision and Control*, pp. 1980–1985.

Anderson, B. D., M. Deistler, W. Chen, and A. Filler (2012). Autoregressive models of singular spectral matrices. *Automatica 48*(11), 2843–2849.

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

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

Barigozzi, M., M. Lippi, and M. Luciani (2020). Cointegration and error correction mechanisms for singular stochastic vectors. *Econometrics 8*(1), 1–23.

Barigozzi, M., M. Lippi, and M. Luciani (2021). Large-dimensional Dynamic Factor Models: Estimation of Impulse–Response Functions with I(1) cointegrated factors. *Journal of Econometrics 221*(2), 455–482.

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

Chamberlain, G. (1983). Funds, factors, and diversification in arbitrage pricing models. *Econometrica 51*(5), 1305–1323.

Deistler, M., B. D. Anderson, A. Filler, C. Zinner, and W. Chen (2010). Generalized linear dynamic factor models: An approach via singular autoregressions. *European Journal of Control 16*(3), 211–224.

Deistler, M., A. Filler, and B. Funovics (2011). AR systems and AR processes: The singular case. *Communications in Information and Systems 11*(3), 225–236.

Deistler, M. and M. Wagner (2017). Cointegration in singular ARMA models. *Economics Letters 155*(C), 39–42.

Filler, A. (2010). *Generalized Dynamic Factor Models: Structure Theory and Estimation for Single Frequency and Mixed Frequency Data.* Doctoral Dissertation, TU Wien.

Forni, M., L. Gambetti, M. Lippi, and L. Sala (2023). Common Components Structural VARs. Working Paper, www.eief.it/eief/images/WP23.06.pdf, Einaudi Institute for Economics and Finance (EIEF), Rome.

Forni, M., D. Giannone, M. Lippi, and L. Reichlin (2009). Opening the black box: Structural Factor Models versus Structural VARs. *Econometric Theory 25*(5), 1319–1347.

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*(2), 359–371.

Forni, M. and M. Lippi (2001). The Generalized Dynamic Factor Model: Representation theory. *Econometric Theory 17*(6), 1113–1141.

Hallin, M. and M. Lippi (2013). Factor models in high-dimensional time series–A time-domain approach. *Stochastic processes and their applications 123*(7), 2678–2695.

Lippi, M., M. Deistler, and B. Anderson (2023). High-dimensional dynamic factor models: A selective survey and lines of future research. *Econometrics and Statistics 26*, 3–16.

Rozanov, Y. A. (1967). *Stationary Random Processes*. San Francisco: Holden Day.

Stock, J. H. and M. W. Watson (2002). Macroeconomic forecasting using diffusion indexes. *Journal of Business and Economic Statistics 20*(2), 147–162.

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