DISCUSSION PAPER SERIES

DP15529
(v. 2)

COMMON COMPONENT STRUCTURAL VARS

Mario Forni, Luca Gambetti, Marco Lippi and Luca Sala

MONETARY ECONOMICS AND FLUCTUATIONS

ISSN 0265-8003

COMMON COMPONENT STRUCTURAL VARS

Mario Forni, Luca Gambetti, Marco Lippi and Luca Sala

Discussion Paper DP15529
First Published 07 December 2020
This Revision 02 May 2023

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

COMMON COMPONENT STRUCTURAL VARS

Abstract

Structural VAR models produce results that can vary dramatically with the choice of variables, because information is deficient and/or contaminated by measured errors. We propose a novel procedure, the Common Component SVAR (CC-SVAR), which solves both problems. First, the common components of the variables of interest are estimated using High-Dimensional Factor techniques. Second, SVAR analysis is performed using such components. The key feature is that number of common components is larger than the number of shocks, so that the SVAR is singular. Consistency results for singular VARs are provided. We apply our procedure to monetary policy shocks. Our finding is that, with the CC-SVAR, results are robust and SVAR puzzles disappear.

JEL Classification: C32, E32

Keywords: Structural var models, Structural factor models, Nonfundamentalness

Mario Forni - mario.forni@unimore.it
Università di Modena e Reggio Emilia, RECent and CEPR

Luca Gambetti - luca.gambetti@uab.cat
Universitat Autònoma de Barcelona, BGSE and Università di Torino

Marco Lippi - mlippi.eief@gmail.com
Einaudi Institute for Economics and Finance (EIEF)

Luca Sala - luca.sala@unibocconi.it
Bocconi University, IGIER & Baffi Carefin

Acknowledgements
We thank Matteo Barigozzi, Tommaso Proietti and Paolo Zaffaroni for very useful comments and suggestions. Gambetti acknowledges the financial support from the Spanish Ministry of Science and Innovation, through the Severo Ochoa Programme for Centres of Excellence in R&D (CEX2019-000915-S), the financial support of the Spanish Ministry ofScience, Innovation and Universities through grant PGC2018-094364-B-I00, and the BarcelonaGraduate School Research Network. Forni, Gambetti and Sala acknowledge the financial support of the Italian Ministry of Research and University, PRIN 2017, grant J44I20000180001.

Common Components Structural VARs*

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

Luca Gambetti[‡]
Universitat Autònoma de Barcelona, BGSE, Università di Torino and CCA

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

Luca Sala[¶‖]
Università Bocconi, IGIER and Baffi Carefin

Abstract

Structural VAR models produce results that can vary dramatically with the choice of variables, because information is deficient and/or contaminated by measured errors. We propose a novel procedure, the Common Component SVAR (CC-SVAR), which solves both problems. First, the common components of the variables of interest are estimated using High-Dimensional Factor techniques. Second, SVAR analysis is performed using such components. The key feature is that number of common components is larger than the number of shocks, so that the SVAR is singular. Consistency results for singular VARs are provided. We apply our procedure to monetary policy shocks. Our finding is that, with the CC-SVAR, results are robust and SVAR puzzles disappear.

JEL classification: C32, E32.
Keywords: structural VAR models, structural factor models, non-fundamentalness, measurement errors.

*We thank Matteo Barigozzi, Tommaso Proietti and Paolo Zaffaroni for very useful comments and suggestions.

[†]E-mail: mario.forni@unimore.it.

[‡]E-mail: luca.gambetti@uab.cat. Gambetti acknowledges the financial support from the Spanish Ministry of Science and Innovation, through the Severo Ochoa Programme for Centres of Excellence in R&D (CEX2019-000915-S), the financial support of the Spanish Ministry of Science, Innovation and Universities through grant PGC2018-094364-B-I00, and the Barcelona Graduate School Research Network.

[§]E-mail: mlippi.eief@gmail.com.

[¶]E-mail: luca.sala@unibocconi.it.

[‖]Forni, Gambetti and Sala acknowledge the financial support of the Italian Ministry of Research and University, PRIN 2017, grant J44I20000180001.

1 Introduction

Since the seminal paper by Sims (1980), Structural Vector Autoregressive (SVAR) models have become the main tool for applied macroeconomic analysis. In the SVAR approach, the macroeconomic variables are driven by a vector of structural shocks, and react to these shocks according to linear impulse-response functions (IRFs). The structural shocks are obtained as linear combinations of the VAR residuals by imposing identifying restrictions based on economic theory.

An unpleasant feature of SVARs is that results can change dramatically depending on the choice of variables. This lack of robustness is a serious problem, since unavoidably both the number and nature of the series to be included in the model is discretionary to some extent. Figure 1 gives an effective idea of the magnitude of the problem, with reference to the effects of monetary policy on real activity and prices. The black lines are the IRFs obtained with the four-variable SVAR including the interest rate, the unemployment rate, industrial production growth and CPI inflation, identified by using the popular instrument of Gertler and Karadi (2015).[1] Unlike in Gertler and Karadi (2015), the Excess Bond Premium is not included in the information set. As a result, both industrial production and prices increase following a policy tightening, so that we have the price puzzle and a real activity puzzle. The blue lines are the IRFs obtained with 50 different spec-

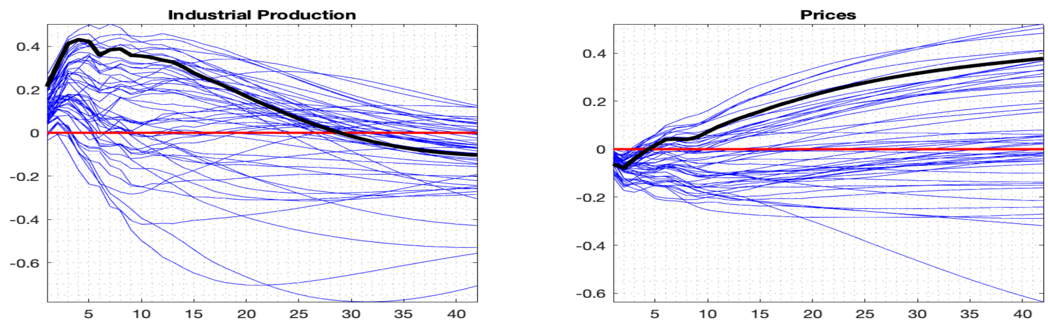


Figure 1: US monthly data from 1977:6 to 2008:12. The IRFs of a monetary policy shock, identified with the proxy of Gertler and Karadi (2015). The black lines are the IRFs of the SVAR(6) with just four variables: the 1 year bond rate, industrial production growth, unemployment and CPI inflation. The blue lines are the IRFs for 50 eight-variable specifications, including the above four variables, and differing for the (random) choice of 4 additional variables.

[1] We us US monthly data from 1977:6 to 2008:12 and 6 lags in the VAR.

ifications, including the four variables above plus four additional randomly chosen macroeconomic series. What the figure tells us is that, by choosing variables appropriately, we can obtain any result.

Why do the results of SVAR analysis vary so much across different specifications? In our understanding, the lack of robustness is due to two main causes: VAR information can be deficient (non-fundamentalness) and is generally contaminated by measurement errors.

It is well known by now that the structural shocks of interest not always are linear combinations of current and lagged VAR variables. When they are not, the shocks are *non-fundamental* for the variables, and SVAR analysis fails. Non-fundamentalness usually occurs when the information set of the VAR variables is smaller than that of the agents.[2] An obvious example is that in which the number of variables is smaller than the number of shocks. But even if the number of shocks and variables coincide, the information contained in the history of the variables can be deficient, expecially in presence of news technology shocks (Forni et al., 2014), fiscal foresight (Leeper et al., 2013) or forward policy guidance (Ramey, 2016).[3] Adding variables to enrich information does not necessarily solve the problem, since observables are usually contaminated by errors, so that, when adding variables, often we add both genuine information and noise.

The fact that many macroeconomic aggregates are affected by measurement error is indisputable. Still, the problem has been largely neglected in the literature. There is an implicit widespread belief that the consequences on SVAR analysis are not serious. However, Giannone et al. (2006) and Lippi (2021) show

[2]An interesting case of non-fundamentalness not arising from superior information of economic agents is that of ‘noise’ shocks (Forni et al. (2017a,b)). We do not deal with this case in the present paper.

[3]Early papers containing examples of non-fundamental economic models are Hansen and Sargent (1991) and Lippi and Reichlin (1993). More recent works are Fernández-Villaverde et al. (2007), Alessi et al. (2011), Sims (2012), Leeper et al. (2013), Forni and Gambetti (2014), Forni et al. (2019). Whether the problem of non-fundamentalness is empirically important or not has been a matter of debate in recent years. Some authors claim that non-fundamentalness has not necessarily dramatic consequences (Sims and Zha (2006), Sims (2012), Beaudry et al. (2019)). On the other hand, Forni et al. (2014) find that narrow information sets distort the estimated effects of news shocks, while several papers, see in particular Bernanke et al. (2005), Forni and Gambetti (2010) and Miranda-Agrippino and Ricco (2021), insist on the importance of large information sets for the estimation of monetary policy shocks.

that this view is wrong (see also the simulations Section 4): even small measurement errors can generate substantial distortions in the estimates of the IRFs, yielding misleading results.[4] Indeed, measurement errors can be regarded as a source of non-fundamentalness. If m macroeconomic variables are driven by q structural shocks, but are contaminated by m independent measurement errors, their impulse-response function representation will be driven by $m + q$ shocks, leading to non-fundamentalness.

The lack of robustness might be used to recommend not to use SVAR models for macroeconomic analysis, an additional argument for authors who argue that Dynamic Stochastic General Equilibrium (DSGE) models should become the standard tool in empirical macroeconomics, see in particular Chari et al. (2008). The opposite view is upheld in the present paper. We show that the problem can be overcome within the SVAR approach, provided that the observed time series are replaced by their common components, estimated by means of High-Dimensional Dynamic Factor techniques. We call our approach Common Component Structural VAR (CC-SVAR).

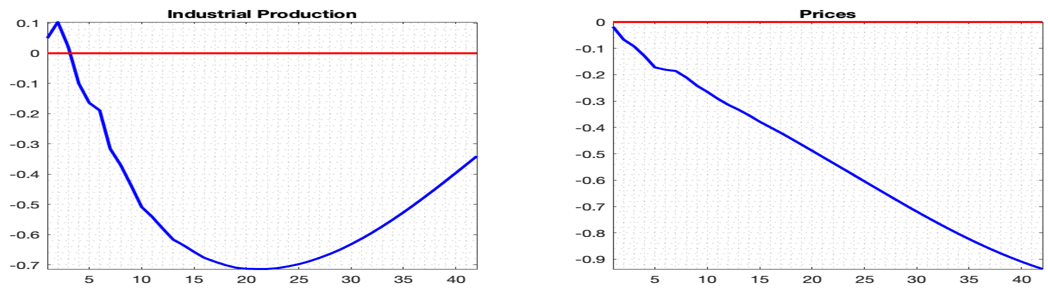


Figure 2: US monthly data from 1977:6 to 2008:12. The IRFs of a monetary policy shock, identified with the proxy of Gertler and Karadi (2015). The blue line is obtained by plotting the IRFs for the same 50 eight-variable specifications of Figure 1 obtaine with the CC-SVAR.

Before explaining the main features of our approach, let us show the results obtained with the CC-SVAR for the effects of the monetary policy shock. Figure 2 shows the IRFs of the same 50 specifications of Figure 1, obtained with the CC-SVAR. The result is striking: the 50 lines are perfectly overlapping, so it looks as

[4] Moreover, the shocks obtained by standard identifying restrictions contain dynamic mixtures of the structural shocks and measurement errors, see Lippi (2021). These dynamic contamination effects of measurement errors are special cases of the dynamic contamination effects of aggregation, as analyzed in Forni and Lippi (1997) and Forni and Lippi (1999).

if there is only one line. Note that neither the price puzzle nor the real activity puzzle show up, despite the fact that the specifications do not include necessarily the Excess Bond Premium nor other financial series.

The main features of our solution are the following. We start with an m-dimensional vector χ_t whose coordinates are the "true", usually unobserved, macroeconomic variables of interest. In particular, the variables χ_t can be interpreted as the "concepts" of a DSGE model. We assume that χ_t is driven by a q-dimensional structural shock vector u_t by means of structural linear IRFs. Moreover, we suppose that $m > q$, so that χ_t is (dynamically) singular, that is, its spectral density matrix has reduced rank at all frequencies. Both singularity and linear structural IRFs are typical of the concepts of DSGE models and the log-linear approximation of their dynamic reaction of these concepts to the structural shocks.

Singular stochastic vectors, under the assumption of rational spectral density, have been extensively studied starting with Anderson and Deistler (2008a). Building on their results we argue that in singular rational representations $\chi_t = B(L)u_t$, where $B(L)$ has an economic-theory based parameterization and u_t is the structural shock vector, u_t is generically fundamental for the vector χ_t.

However, when χ_t is replaced by its observed counterpart, call it x_t, singularity and the resulting generic fundamentalness of the structural shocks break down. High-Dimensional Dynamic Factor techniques, by "cleaning" the observed variables from measurement error, provide an estimate of χ_t and the IRFs of χ_t to u_t (see Stock and Watson (2002a), Bai and Ng (2002); a proof of consistency for the common-component estimate, under the assumptions used in the present paper, is provided in the Online Appendix D).

CC-SVARs consist in the application of SVAR analysis to $\hat{\chi}_t$. We firstly estimate a VAR for $\hat{\chi}_t$ and, secondly, impose restrictions based on economic logic to identify the estimated structural shocks and IRFs.

The CC-SVAR approach encompasses and improves over previous factor-based techniques. The existence of a finite VAR representation for χ_t is not assumed, but derived from the theory of singular stochastic vectors. Moreover, we provide a proof that the estimated structural shocks and IRFs are consistent. This result is fairly trivial if χ_t is not singular. What we prove, this is our main technical result, is that consistency holds even when χ_t is singular, so that its VAR representation is not necessarly unique.[5]

Regarding Factor Augmented VAR models (FAVAR), Bernanke et al. (2005), by adding estimated factors to the VAR, mitigate the fundamentalness problem and therefore improve over SVAR models. However the macroeconomic variables included in the FAVAR may contain measurement errors. Also, an important advantage of our approach is that the identification of structural shocks is more direct and transparent since the restrictions are imposed on the common components of the variables and there is no need to restrict the response of the factors.

Regarding the Structural Dynamic Factor Model (SDFM), in the version proposed by Stock and Watson (2005) and Forni et al. (2009), the CC-SVAR is more flexible for two reasons. Firstly, as in FAVAR models, the vector χ_t can directly include observable variables, insofar as their measurement error is zero (think for instance of interest rates). Moreover, the number of common components can in principle be smaller than the number of factors (provided that it is greater than the number of shocks q). When they coincide, than the CC-SVAR can be regarded as an alternative procedure to estimate the SDFM.

In the empirical part of the paper, we apply the CC-SVAR method to the study of the effects of monetary policy shocks on the main macroeconomic variables, an highly debated problem in macroeconometrics. We show that the results of SVAR analysis are not robust. In particular, the sign of the effect of a contractionary policy shock on both prices and real activity is sensitive to the variables included in the model. These puzzling phenomenon disappears with the CC-SVAR. Contractionary monetary policy shocks reduce prices independently of the variables included. This definitely solves the famed "price puzzle". In the online Appendix G we report an additional empirical exercise concerning the effects of thechnology shocks on hours worked.

The paper is organized as follows. Section 2 discusses the implications of measurement errors and non-fundamentalness for SVAR analysis within a simple Real Business Cycle Model. Section 3 presents the model, the estimation procedure

[5] In Forni et al. (2009), consistency is proved for a singular VAR(1), a special case in which the VAR representation is unique.

and the consistency results. Formal proofs are given in the Online Appendix. In Section 4 the estimation procedure described in Section 3.6 is applied to simulated data based on the model discussed in Section 2. Section 5 presents and discusses the empirical application. Section 6 concludes.

2 Non-fundamentalness and measurement errors in a simple model

The model discussed in Leeper et al. (2013) is employed here as a laboratory to discuss the consequences of narrow information sets (non-fundamentalness) and measurement errors. The model is a simple Real Business Cycle (RBC) model with log preferences, inelastic labor supply and two shocks: $u_{a,t}$, a technology shock, and $u_{\tau,t}$, a tax shock. A non-standard feature of the model is the fact that the tax shock has a delayed effect on taxes, the so-called fiscal foresight. The equilibrium capital accumulation is

$$k_t = \alpha k_{t-1} + a_t - \delta \sum_{i=0}^{\infty} \theta^i E_t \tau_{t+i+1} \tag{1}$$

where $0 < \alpha < 1$, $0 < \theta < 1$, $\delta = (1-\theta)\tau/(1-\tau)$, τ being the steady state tax rate, $0 \leq \tau < 1$, and a_t, k_t and τ_t are the log deviations from the steady state of technology, capital and the tax rate, respectively. Technology and taxes are assumed, for simplicity, to be *i.i.d* processes,

$$\begin{aligned} a_t &= u_{a,t} \\ \tau_t &= u_{\tau,t-2}, \end{aligned}$$

where $u_{\tau,t}$ and $u_{a,t}$ are i.i.d. shocks that economic agents can observe. The second equation implies a delay of two periods. Solving for k_t we obtain the following equilibrium MA representation:

$$\begin{pmatrix} a_t \\ k_t \\ \tau_t \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ \dfrac{-\delta(L+\theta)}{1-\alpha L} & \dfrac{1}{1-\alpha L} \\ L^2 & 0 \end{pmatrix} \begin{pmatrix} u_{\tau,t} \\ u_{a,t} \end{pmatrix} = B(L)u_t. \tag{2}$$

2.1 Full versus narrow information sets

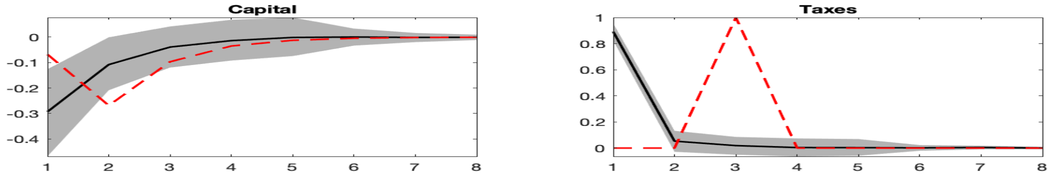
In the standard approach to the estimation of the impulse-response functions, as the variables are driven by two shocks we should estimate a SVAR including two of the three variables in the system. However, the vector $u_t = (u_{\tau,t}\ u_{a,t})'$ is non-fundamental for all pairs of variables. Indeed, the determinants of the square subsystems including technology and capital, technology and taxes, capital and taxes are, respectively

$$\frac{\delta(z+\theta)}{1-\alpha z}, \quad -L^2, \quad -\frac{L^2}{1-\alpha L}. \tag{3}$$

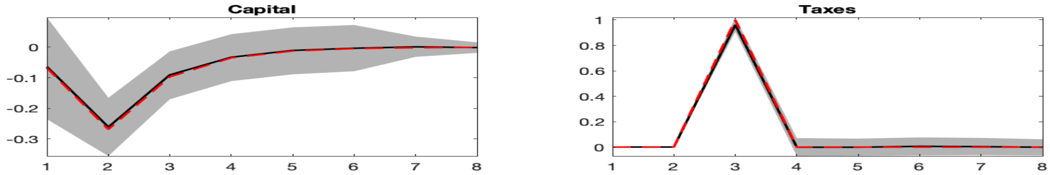
The second and the third vanish for $z = 0$. The first vanishes for $z = -\theta$ if $\tau \neq 0$, for all $z \in \mathbb{C}$ if $\tau = 0$. This implies that standard SVAR techniques are unlikely to correctly estimate the dynamic effect of the fiscal shock.

A quantitative assessment of the distortion caused by non-fundamentalness in the two-dimensional SVARs within system (2) is obtained here by a simulation exercise, denoted Simulation 1. We generate 1000 different dataset with 200 time observations from the model (2) using the parameterization in Leeper et al. (2013): $\alpha = 0.36$, $\theta = 0.2673$ and $\tau = 0.25$ and $u_t \sim N(0, I)$. For each of the datasets we estimate a VAR(4) including taxes and capital and we identify the tax shock by imposing that it is the only one driving cumulated taxes in the long run, a restriction that is clearly satisfied in the model. Panel (a) of Figure 3 plots the estimated impulse-response functions for a tax shock. The red dashed lines are the theoretical impulse response functions. The solid lines represent the mean (across datasets) of the point estimates. The grey areas represent the 16th and 84th percentiles of the point estimate distribution. As the red lines lie outside the bands, the true effects are very badly estimated. The responses obtained by the SVAR neatly anticipate the peak response in the true impulse response functions. Both taxes and capital react immediately and then the effects vanish.

Thus when only part of the information is used, current and past values of taxes and capital, the estimates of the impulse-response functions can be substantially distorted. However, the information contained in current and past values of technology, capital and taxes is sufficient to recover the vector u_t. Precisely, Panel (a)



Panel (b)



Panel (c)

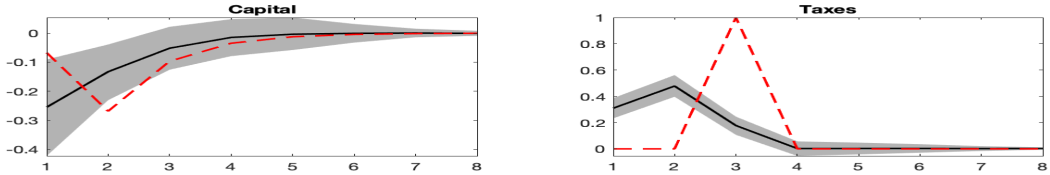


Figure 3: Simulation 1. Non-fundamentalness and measurement errors. Estimated IRFs for the tax shock. The red dashed lines are the theoretical IRFs. The solid lines represent the mean (across 1000 simulated datasets) of the point estimates. The grey areas contain the point estimates between the 16th and 84th percentiles. Panel (a): SVAR(4) with Capital and Taxes. Panel (b): SVAR(3) with Capital, Taxes and Technology. Panel (c): SVAR(3) with Capital, Taxes and Technology when Technology is measured with a 5% error.

provided that $\tau \neq 0$, the matrix $B(L)$ in (2) has a stable left-inverse. Setting

$$A(L) = \begin{pmatrix} 1 & 0 & 0 \\ \dfrac{(\theta - L)L}{\theta^2} & \dfrac{(1-\alpha L)(\theta^2 - \theta L + L^2)}{\theta^2} & \dfrac{\delta L}{\theta^2} \\ \dfrac{-L^2}{\delta\theta} & \dfrac{(1-\alpha L)L^2}{\delta\theta} & 1 + \dfrac{L}{\theta} \end{pmatrix},$$

we see that $A(L)B(L) = B(0)$, so that the vector $(a_t \;\, k_t \;\, \tau_t)')$ has the VAR(3) representation

$$A(L)\begin{pmatrix} a_t \\ k_t \\ \tau_t \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -\delta\theta & 1 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} u_{\tau,t} \\ u_{a,t} \end{pmatrix}. \tag{4}$$

Because the determinant of $A(L)$ is $(1-\alpha L)$ and $|\alpha| < 1$, $A(L)$ is stable.

Note that singularity of the vector $(a_t \;\, k_t \;\, \tau_t)'$ implies that its 3×3 spectral density is singular at all frequencies. However, the covariance matrix which is necessary to estimate a VAR is not singular. Therefore, using the same data as in the previous exercise, we estimate a VAR(3) for the full vector $(a_t\ k_t\ \tau_t)'$.[6]

We identify the tax shock by assuming that it is the only one affecting cumulated taxes in the long run, thus a Blanchard and Quah (1989) identification scheme withe the tax shock ordered first. The results are displayed in Panel (b) of Figure 3. Using the full information set, the impulse response functions are estimated extremely well, the red dashed and solid black lines perfectly overlapping.

It is important to note that this result, that a correct estimation of the impulse-response functions is obtained by enlarging the information available to the econometrician, crucially depends in our simple model on the fact that an additional variable is added without increasing the sources of uncertainty, that is *without adding additional shocks*, which is tantamount to the fact that the enlarged vector of available variables is singular.

2.2 Measurement errors

Typically, many of the macroeconomic variables used in SVAR models are affected by measurement error. To understand the implications of this, we modify model (2) by adding measurement errors:

$$\begin{pmatrix} a_t \\ k_t \\ \tau_t \end{pmatrix} = B(L)u_t + \begin{pmatrix} \xi_t^a \\ \xi_t^k \\ \xi_t^\tau \end{pmatrix},$$

with the assumption that the vector $(\xi_t^a\ \xi_t^k\ \xi_t^\tau)'$ is white noise and is orthogonal to the shocks $u_{a,t}$ and $u_{\tau,t}$ at all leads and lags. The data are generated using the same parameterization of the previous Section, with $\xi_t^\tau = \xi_t^k = 0$ and ξ_t^a accounting for 5% of the variance of the series a_t. Using the full vector we estimate again a VAR(3) with the same identification scheme. Panel (c) of Figure 3 reports the estimated impulse-response functions. Surprisingly, with a measurement error as small as that used in the generation of the data, and affecting only one of the variables, the effects of the tax shock are very badly estimated.[7] Thus, even when information seems sufficient to correctly recover the impulse-response functions,

[6] Almost identical results are obtained with a VAR with four lags (not shown here).

[7] Augmenting the number of lags does not improve the estimates (not shown here).

a small measurement error may cause substantial distortion in the estimates. We come back to this point in Section 3.7.

3 Common Component Structural VARs

In this section we present our model and basic assumptions. We start with a moving-average representation of the macro economy, where structural shocks affect the, true, macroeconomic variables according to linear impulse-response functions. We assume singularity, i.e. that the number of variables is larger than the number of shocks. Under singularity, plus an additional assumption (the zeroless assumption, see below), the macroeconomic variables admit a VAR representation in the structural shocks, which therefore are fundamental. We then assume that each observable variable is the sum of the macroeconomic variable and a variable-specific measurement error. To eliminate measurement errors, we immerse the observables into a large data set following a factor model, in which each variable is decomposed into a common and an idiosyncratic component. We estimate the common components, i.e. the true macroeconomic variables, *via* the standard principal component estimator. The main result of the section is that the CC-SVAR, i.e. the SVAR applied to the estimated macroeconomic variables of interest, produces consistent estimates of the structural shocks and the corresponding impulse-response functions.

3.1 The impulse-response function representation

Let χ_t be a vector including m macroeconomic variables of interest.

Assumption 1. Structural IRF representation. *The (zero-mean) vector χ_t is the stationary solution of the vector ARMA equation:*

$$H(L)\chi_t = K(L)u_t, \tag{5}$$

where: (1) u_t is a serially independent, q-dimensional vector of orthonormal shocks, independent of all past available information, (2) $H(L)$ is a $m \times m$ polynomial matrix of degree s_1 such that $\det H(z) = 0$ *implies* $|z| > 1$*, (3) $K(L)$ is a*

$m \times q$ polynomial matrix of degree s_2. Thus

$$\chi_t = B(L)u_t = H(L)^{-1}K(L)u_t, \tag{6}$$

The vector u_t and the matrices $H(L)$ and $K(L)$ have a structural *interpretation based on economic theory. The entries of $H(L)$ and $K(L)$ are, respectively,*

$$\begin{aligned} h_{ij}(L) &= 1 - h_{ij,1}L - \cdots - h_{ij,s_2}L^{s_1} \\ k_{ij}(L) &= k_{ij,0} + k_{ij,1}L + \cdots + k_{ij,s_1}L^{s_2}. \end{aligned} \tag{7}$$

Some of the $m^2 s_1 + mq(s_2 + 1)$ coefficients in (7) may be zero, thus the integers s_1 and s_2 denote maxima over i and j for the degrees of the polynomials h_{ij} and k_{ij}, respectively.

Equation (6) is sometimes referred to as the Slutsky-Frisch representation of the macro economy. It can be thought of as resulting from the linearization of a DSGE model and can easily be derived from its state-space representation. The example in equation (2) is of course a special case of (6)-(7), with

$$H(L) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 - \alpha L & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad K(L) = \begin{pmatrix} 0 & 1 \\ -\delta(L + \theta) & 1 \\ L^2 & 0 \end{pmatrix}$$

The following assumption plays a crucial role :

Assumption 2. Dynamic singularity, static non-singularity.
(a) *The number of variables m is larger than the number of shocks q, so that χ_t is dynamically singular, that is, the spectral density matrix $\Sigma^\chi(\theta) = B(e^{i\theta})B(e^{-i\theta})'$ is singular for all $\theta \in [\pi, \pi]$.*
(b) *The covariance matrix of χ_t, denoted by Σ_0^χ, is non-singular.*

As already observed in the Introduction, dynamic singularity is a feature of most DSGE models, a prominent example is Smets and Wouters (2007), see also Canova (2007), pp. 232-3, for general considerations. In general, we suppose that the number of structural shocks driving the economy is independent of the dimension of χ_t. Thus, if in a first formulation of the model we had $m = q$, the model obtained by augmenting χ_t with auxiliary variables would fulfil the condition required in Assumption 2(a).

3.2 Existence of a finite-length VAR representation for χ_t

Given a vector χ_t, fulfilling Assumptions 1 and 2, plus an additional assumption (the zeroless assumption discussed below) we show that χ_t has a finite-length VAR representation $A(L)\chi_t = B_0 u_t$, so that u_t is fundamental for χ_t.

3.2.1 Zeroless $m \times q$ matrices and finite-length VARs

Let us start with an elementary example. Consider the 2-dimensional vector $\chi_t = (\chi_{1t} \ \chi_{2t})'$, where

$$\begin{aligned} \chi_{1t} &= u_t + k_1 u_{t-1} \\ \chi_{2t} &= u_t + k_2 u_{t-1}, \end{aligned} \tag{8}$$

u_t being a scalar white noise and $(k_1 \ k_2)$ any point in $\mathbb{R}^2$. The vector χ_t is dynamically singular, since it has two entries ($m = 2$) driven by just one shock ($q = 1$).

If $k_1 \neq k_2$ we have

$$u_t = \frac{k_2\chi_{1t} - k_1\chi_{2t}}{k_2 - k_1}.$$

This can be used to replace u_{t-1} in (8), obtaining

$$\begin{aligned} \chi_{1t} &= \frac{k_1}{k_2 - k_1}(k_2\chi_{1,t-1} - k_1\chi_{2,t-1}) + u_t \\ \chi_{2t} &= \frac{k_2}{k_2 - k_1}(k_2\chi_{1,t-1} - k_1\chi_{2,t-1}) + u_t, \end{aligned} \tag{9}$$

which is a VAR(1) representation for the MA(1) vector χ_t. Thus u_t belongs to the space spanned by current and past values of χ_t. We conclude that the white noise u_t in (8) is fundamental and that χ_t has a finite-order autoregressive representation for all values of the parameters k_1 and k_2, with the exception of the line $k_1 = k_2$.

Model (2) in Section 2 provides another example of a singular vector having a rational MA representation, which admits the finite-order VAR representation (4), unless $\tau = 0$.

It easily seen that in both examples the existence of a finite-length VAR representation occurs when the values of the parameters are such that the matrix $K(L)$ has the property defined below:

Definition 1. *The $m \times q$ matrix $K(L)$, with $m \geq q$, is zeroless if the rank of $K(z)$ is q for all complex numbers z.*

Note that if $m = q$ then $K(L)$ is zeroless if and only if has a constant determinant, which is a very special case[8]. On the other hand, if $m > q$, a sufficient condition for zerolessness of $K(L)$ is that it contains at least two $q \times q$ submatrices whose determinants have no common zeros. An extremely important consequence of zerolessness is proved in Anderson and Deistler (2008a):

Proposition AD1. Anderson and Deistler. *Under Assumptions 1 and 2, if the matrix $K(L)$ is zeroless, there exists a finite $m \times m$ stable matrix polynomial $\tilde{K}(L)$ such that $\tilde{K}(L)K(L) = K_0 = B_0$ (we say that $\tilde{K}(L)$ is a left inverse of $K(L)$), so that, setting $A(L) = H(L)\tilde{K}(L)$, χ_t has the finite-length VAR representation $A(L)\chi_t = K_0 u_t = B_0 u_t$. As K_0 has maximum rank (because $K(L)$ is zeroless), u_t lies in the space spanned by current and past values of χ_t, i.e. u_t is fundamental for χ_t.*

We see that in examples (2) and (8) the matrix $K(L)$ is zeroless with the exception of a lower dimensional subset of the parameter space, namely:
(i) In Example (2) the two-dimensional subset of $\{0 < \alpha < 1,\ 0 < \theta < 1,\ 0 \leq \tau < 1\}$ where $\tau = 0$, see the 2×2 determinants in (3).
(ii) In Example (8) the one-dimensional subset of $\mathbb{R}^2$ where $k_1 = k_2$.

Direct inspection of the matrix $K(L)$ in those examples is useful to understand why this fact occurs,

$$K_1(L) = \begin{pmatrix} 0 & 1 \\ -\delta(L+\theta) & 1 \\ L^2 & 0 \end{pmatrix}, \qquad K_2(L) = \begin{pmatrix} 1 + k_1 L \\ 1 + k_2 L \end{pmatrix}.$$

Each entry of the matrix $K_2(L)$ has its own parameters which vary independently of those of the other entry. The parameters of the second row of $K_1(L)$ vary independently of the first and third row. Thus in both cases we may find two $q \times q$ submatrices with no common zeros apart from exceptional values of the parameters.

[8]Matrices $K(L)$ with constant determinant are called unimodular.

Thus $K(L)$ is generically zeroless in examples (2) and (8), where "generic" is informally used here as meaning "with the exception of a lower-dimensional subset in the parameter space" (see Appendix A.1 for a formal definition). Now the question is whether the result holding for such elementary cases can be extended to any model fulfilling Assumptions 1 and 2. Relevant cases are:

(I) Like in example (8), each entry of $K(L)$ has its own parameters which vary independently of those of the other entries. In this case $K(L)$ is generically zeroless. This is shown in Anderson and Deistler (2008b), Proposition 1, though with a different parameterization, and Forni et al. (2015), though in a special case.

(II) However, in this paper we are interested in the case in which, like in example (2), the entries of $K(L)$ jointly depend on the parameters of an economic model. As observed below Definition 1, a sufficient condition for zerolessness of $K(L)$ is that $K(L)$ contains at least two $q \times q$ submatrices whose determinants have no common zeros. In Appendix A.1 we prove that either that sufficient condition generically holds, or it generically fails to hold. That such a sharp alternative is specific to singularity can be fully understood by comparison, for example, to fundamentalness in the non-singular case, where usually neither fundamentalness nor non-fundamentalness is generic.

(III) Moreover, the second alternative above holds if the coefficients of $K(L)$ fulfill a restriction which has a purely mathematical motivation (see Appendix A.1). Based on this observation and our knowledge of theory-based macroeconomic models, we claim that generic zerolessness is typical, with the possible exception of those cases in which χ_t is the first difference of a cointegrated $I(1)$ vector. In that case a zero of $K(L)$ at $z = 1$ may be directly motivated by the theory. In the next section we show how such zeros can be "removed".

Let us conclude this section with a caveat. Proposition AD1 states the existence of a finite-length VAR under zerolessness. However, as recalled in the Introduction, such VAR is not necessarily unique. This fact, which does not occur in the non-singular case, is illustrated by a simple example in Section 3.4.

3.2.2 Singularity and cointegration

Now let $\chi_t = (1-L)X_t$, where X_{it} is $I(1)$ for $i = 1, \ldots, m$. For simplicity suppose that

$$(1-L)X_t = K(L)u_t.$$

If χ_t is not singular, cointegration of X_t implies that $K(L)$ has a zero at $z = 1$, so that a VAR in χ_t is misspecified and the estimation either of an Error Correction Model (ECM) or a VAR in the levels X_t is recommended.

On the other hand, the rank at zero of the spectral density of a singular vector χ_t is q at most, so that X_t is necessarily cointegrated with cointegration rank $m-q$ at least, that is $c = m - q + \kappa$, with $0 \leq \kappa < q$. As our aim here is to show how a zero at $z = 1$ can be assumed away, we suppose that $K(L)$ is zeroless for $z \neq 1$.

Assume firstly that $\kappa = 0$. In this case the rank of $K(1)$ is q, i.e. $K(L)$ is zeroless, Proposition AD1 applies and X_t has, *despite cointegration*, a finite-length VAR representation in differences. To illustrate this most important feature of singular vector processes, let us go back to the example of equation (8), with $\chi_t = \Delta X_t$, and take the linear combination

$$\frac{(1+k_2)\chi_{1t}}{k_2-k_1} - \frac{(1+b_1)\chi_{2t}}{k_2-k_1} = \frac{(1+k_2)(1-L)X_{1t}}{k_2-k_1} - \frac{(1+k_1)(1-L)X_{2t}}{k_2-k_1} = (1-L)u_t.$$

By integrating both sides we get the cointegration relationship

$$\frac{(1+k_2)X_{1t}}{k_2-k_1} - \frac{(1+k_1)X_{2t}}{k_2-k_1} = c + u_t,$$

where c is a constant. Nevertheless representation (9) holds for χ_t, so that X_t has a VAR(1) representation in differences. Let us point out that if $\kappa = 0$ singularity not only ensures generic fundamentalness of u_t, but also solves the representation and estimation difficulties arising from cointegration in the standard non-singular case. Simulation 7, see the Online Appendix F.3, illustrates this point.

If $\kappa > 0$ the matrix $K(1)$ has a zero at $z = 1$ and a VAR in differences is misspecified. Barigozzi et al. (2020) show that generically the singular vector χ_t has several alternative finite-length ECMs, with a number of error correction terms ranging from κ to $m-q+\kappa$, see p. 20 (they also prove that all such autoregressive representations produce the same impulse-response functions). The methods used in Barigozzi et al. (2020, 2021) and those in the present paper are very close.

Indeed, our definitions and results could be adapted to include ECMs, though this would produce a longer, more complicated paper and is left for future research.

Here we assume that $\kappa = 0$. We argue that such an assumption, though restrictive, can be easily motivated for a quite large class of important macroeconomic applications and is therefore much more than a convenient simplification. Let us again use a simple example to illustrate our point. Consider a three-dimensional $I(1)$ vector X_t driven by the two-dimensional shock u_t. Suppose that the effect of u_{2t} on the three variables X_{jt} is permanent and that the effect of u_{1t} on X_{1t} and X_{2t} is transitory. Thus:

$$\begin{pmatrix} (1-L)X_{1t} \\ (1-L)X_{2t} \\ (1-L)X_{3t} \end{pmatrix} = K(L)u_t = \begin{pmatrix} (1-L)a(L) & b(L) \\ (1-L)c(L) & d(L) \\ f(L) & g(L) \end{pmatrix} \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix}, \tag{10}$$

where the entries of the second column of $K(L)$ do not vanish at $z = 1$.

(A) Of course, there exist shocks, monetary policy or demand shocks for example, which may well have transitory effects on all trended real activity variables. If, for example, the variables of interest X_{jt}, $j = 1, 2, 3$, are GDP, consumption and investment, respectively, and u_{1t} is a demand shock, then $f(1) = 0$ and $\kappa > 0$.

(B) However, suppose that the variable X_{3t} is an $I(1)$ price or monetary aggregate. We claim that there are no reasons based on economic theory why demand or monetary policy shocks should have a temporary effect on X_{3t}. Indeed, this is not the case with the monetary policy or demand shocks estimated in the literature. The same conclusion holds if X_{3t} is an $I(0)$ variable among interest rates, risk premia, term spreads or the unemployment rate. Dropping $(1 - L)$ in front of X_{3t} in (10), there is no reason why $f(L)$ should contain the factor $1 - L$. In general, if the vector of interest contains both real and monetary $I(1)$ variables or both $I(1)$ and $I(0)$ variables, as is the case in the empirical applications in Section 5, we can safely assume that $K(L)$ has no zero at $z = 1$.

(C) Moreover, suppose, as we do below starting with Section 3.5, that the vector of interest X_t is part of a large vector $\mathbf{X}_t$, whose coordinate variables are all driven by u_t. Suppose also that the vector of interest X_t is $I(1)$, cointegrated and, for example, $\kappa = 1$. It is highly likely that $\mathbf{X}_t$ contains variables which, belonging to

a different "family", as X_{3t} in (B), can be used to augment X_t and obtain a larger vector with $\kappa = 0$. More on (B) and (C) in Appendix A.2.
(D) The simple idea of forcing, so to speak, $\kappa = 0$ in the case of singular $I(1)$ vectors, by augmenting the vector of interest with suitable variables, is likely to apply to any hypotetical situation in which non-zerolessness is implied by economic-theory based restrictions.

3.2.3 Genericity of zerolessness

Based on the above discussion of a possible zero of $K(L)$ at $z = 1$, and the arguments in (I), (II) and (III) in Section 3.2.1, we believe that assuming that $K(L)$ is zeroless, either directly for χ_t or for an augmented version of it, has a sound motivation. Thus:

Assumption 3. Zeroless IRFs. *The matrix $K(L)$ is zeroless.*

Under Assumptions 1, 2 and 3, by Proposition AD1, the vector χ_t has a finite-order VAR representation

$$A(L)\chi_t = B_0 u_t = v_t, \tag{11}$$

where $A(L)$ is a stable matrix of polynomials in L. Moreover, u_t is fundamental for χ_t.

3.3 Adding (and removing) measurement errors

We assume now that we can only observe $x_t = \chi_t + \xi_t$, where ξ_t is a vector of measurement errors. We also assume that x_t is a subvector of an n-dimensional vector $\mathbf{x}_{nt} = (x_{it}),\ i = 1, \ldots, n$, and that n is large, possibly as large or even larger than T, the number of observations for each series. High-Dimensional Dynamic Factor Model techniques have been used to obtain estimators of χ_t, which are consistent as $n, T \to \infty$. Let us mention here Forni et al. (2000), Stock and Watson (2002a,b), Bai and Ng (2002), Forni et al. (2015, 2017). Consistently with this literature, in view of the double aymptotics in n and T, we assume that the processes x_{it}, $t \in \mathbb{Z}$, belong to an infinite-dimensional vector.

Assumption 4. Embedding χ_t in a Large-Dimensional Dynamic Factor Model.
(a) The vector χ_t is not observable. The observable vector, say x_t, is given by

$$x_t = \chi_t + \xi_t = B(L)u_t + \xi_t.$$

(b) Without loss of generality, the entries of x_t, i.e. x_{it}, $i = 1, \ldots, m$, are the first m series of the sequence

$$x_{it} = \chi_{it} + \xi_{it}, \qquad i = 1, \ldots, \infty.$$

The variables χ_{it} are called the common component and the variables ξ_{it} the idiosyncratic components. The idiosyncratic and the common components are zero-mean and mutually independent at all leads and lags.
(c) The researcher can observe the first n series x_{it}, $i = 1, \ldots, n$.

The idiosyncratic component of χ_{it} is usually interpreted as containing specific causes of variation, plus measurement error. However, if χ_{it} is one of the main macroeconomic aggregates, like GDP, consumption for example, specific causes of variation should cancel in the aggregation and the idiosyncratic component is likely to contain only measurement error.

Different consistent estimators of χ_{it}, denoted $\hat{\chi}_{it}$, have been proposed in the factor-model literature, some are mentioned at the beginning of Section 3.5. Of course each one of them contains an estimator of the vector χ_t, denoted $\hat{\chi}_t$. For the moment we do not select a particular estimator $\hat{\chi}_t$. Rather, we show that u_t and the IRFs implicit in equation (11) are consistently estimated using any estimator $\hat{\chi}_t$ fulfilling the Assumptions A and B specified below. Then, starting with Section 3.5, we focus on the static principal component estimator of Stock and Watson (2002a,b) and show that under suitable assumptions it fulfills Assumptions A and B.

Notation 1.
(i) Let (y_t) and (z_t) be zero-mean s-dimensional vector processes. Σ_k^{yz} denotes the (population) $s \times s$ covariance matrix $\mathrm{E}\left(y_t z'_{t-k}\right)$. $\hat{\Sigma}_k^{yz}$, the sample counterpart of Σ_k^{yz}, is defined as $\sum_{t=k+1}^{T} y_t z'_{t-k}/(T-k)$. The $s \times s$ autocovariance matrices of (y_t) are obviously denoted by Σ_k^y and $\hat{\Sigma}_k^y$.
(ii) By $\hat{\chi}_t = (\hat{\chi}_{it})$, $i = 1, \ldots, m$, $t = 1, \ldots, T$, we denote an estimator of χ_t based

on x_{it}, $i = 1, \dots, n$, $t = 1, \dots, T$.
(iii) $\hat{\pi}_t = \hat{\chi}_t - \chi_t$ and $||\cdot||$ denotes the euclidean vector norm.

Assumption A. Properties of $\hat{\chi}_t$. *The estimator $\hat{\chi}_t$ is such that*
(a) $||\hat{\chi}_t - \chi_t|| = ||\hat{\pi}_t||$ is $O_p(r_{n,T})$, where $r_{n,T} \to 0$ as $\min(n,T) \to \infty$.
(b) The sample covariance matrix of the vector

$$(\hat{\chi}_t' \ \hat{\chi}_{t-1}' \ \cdots \ \hat{\chi}_{t-p}')'$$

is non-singular with probability one.

Assumption A(a) states that the estimator is consistent as $\min(n,T) \to \infty$, the rate being $r_{n,T}$. Assumption A(b) ensures that, owing to the error $\hat{\pi}_t$, the regressors of the empirical VAR for $\hat{\chi}_t$ are not collinear, so that the VAR is unique and can be estimated by standard OLS, even if the regressors of the population VAR are collinear (see the next section).

Finally we need a standard technical ergodicity property:

Assumption B. Covariance Ergodicity. *We assume that*
$||\hat{\Sigma}_k^{\chi} - \Sigma_k^{\chi}|| = O_p\left(1/\sqrt{T}\right)$, *for any k.*

3.4 Estimating a singular VAR

It is convenient to re-write the population VAR in (11) as

$$\chi_t = A_1 \chi_{t-1} + \cdots + A_p \chi_{t-p} + v_t = \mathcal{A} Z_{t-1} + v_t, \tag{12}$$

where $Z_t = (\chi_t' \ \chi_{t-1}' \ \cdots \ \chi_{t-p+1}')'$, $v_t = B(0)u_t$ is a white-noise vector of dimension n and rank q, with v_{it} orthogonal to $\chi_{j,t-k}$ and $\xi_{j,t-k}$ for all i, j and all positive k.

A major difficulty with (12) is that, as pointed out in Anderson and Deistler (2008a), the variance-covariance matrix of the regressors, Σ_0^Z, can be singular. A simple example will suffice here. Consider the case $m = 3$, $q = 1$, $B(L) = B_0 + B_1 L + B_2 L^2 + B_3 L^3$, and suppose that the 12 entries in the matrices B_j can vary independently of one another. The vector Z_{t-1} has $3p$ components, each being a linear combinations of $u_{t-1}, \dots, u_{t-p}, u_{t-p-1}, u_{t-p-2}, u_{t-p-3}$, thus the components of Z_{t-1} lie in a linear space of dimension $p+3$. This implies that if $p \geq 2$, so that $3p > p+3$, the components of Z_{t-1} are collinear and Σ_0^Z is singular. On the other hand, if $p = 1$ in (12), then $(I - A_1L)(B_0 + B_1L + B_2L^2 + B_3L^3) = B_0$, which implies 12 linear equations for the 9 entries of A_1, a system with no solutions for generic values of the entries of the matrices B_j, $j = 0, \ldots, 3$, see Appendix A.3 for details.

What we learn from this example is that in the singular case the matrix $\mathcal{A}$ is not necessarily unique. Now consider the empirical counterpart of (12),

$$\hat{\chi}_t = \hat{A}_1\hat{\chi}_{t-1} + \cdots + \hat{A}_p\hat{\chi}_{t-p} + \hat{v}_t = \hat{\mathcal{A}}\hat{Z}_{t-1} + \hat{v}_t. \tag{13}$$

By Assumption A(b), the matrix $\hat{\Sigma}_0^{\hat{Z}}$ is non singular with probability one, so that (13) is unique and can be estimated by standard OLS. However, as (12) is not unique, convergence of $\hat{\chi}_t$ to χ_t, which is Assumption A(a), has no clear implication on the convergence of the matrices $\hat{A}_k$, the VAR residuals $\hat{v}_t$ and the shocks $\hat{u}_t$, the latter being obtained from $\hat{v}_t$ by the same theory-based restrictions identifying u_t and $B(L)$, see below for details. This is, as we mentioned in the Introduction, the problem which has been overlooked in previous literature dealing with VARs for the factors or the common components.[9]

A clear intuition of our results stems from the following simple fact. Inverting the matrix $A(L) = I - A_1L - \cdots - A_pL^p$, we obtain

$$\chi_t = A(L)^{-1}v_t = A(L)^{-1}B(0)u_t.$$

On the other hand, χ_t has a unique MA representation in u_t, so that $A(L)^{-1}B(0) = B(L)$, independently of which matrix $A(L)$ we choose. What we show below is that even if the VAR in (12) is not unique, so that $\hat{A}(L)$ may not converge at all, $\hat{v}_t$, $\hat{u}_t$ and $\hat{B}(L)$ converge to v_t, u_t and $B(L)$, respectively.

Our first result is:

Proposition 1. Consistency of the VAR residuals. *Under Assumptions 1 to 4, A and B, we have*

$$||\hat{v}_t - v_t|| = O_p\left(\max(r_{n,T}, 1/\sqrt{T})\right).$$

Proposition 1 states that the residuals of (13) consistently estimate the residuals of (12). The proof is given in the Online Appendix B.

[9] If $p = 1$, $\Sigma_0^Z = \Sigma_0^\chi$, which is obviously non-singular, see the comment on Assumption 2(b). Thus (12) is unique. Forni et al. (2009) assume $p = 1$, thus avoiding the difficulty we are dealing with here.

Let us turn now to the structural shocks and response functions. For simplicity we shall focus on the Cholesky scheme. Our consistency results can be easily generalized to other identification schemes. Hence we specialize Assumption 1 as follows:

Assumption 1$'$. Cholesky impulse-response functions. *After possible reordering of the variables χ_{it}, defining Q as the upper $q \times q$ submatrix of B_0, Q is non-singular and lower triangular, with positive entries on the main diagonal. Thus, for $i = 2, \dots, q$, u_{it} has no contemporaneous effect on χ_{jt}, $j < i$.*

Next, let $\hat{A}(L) = I - \hat{A}_1 L - \hat{A}_2 L^2 - \cdots - \hat{A}_p L^p$, so that $\hat{\chi}_t = \hat{A}(L)^{-1}\hat{v}_t$. We suppose that the Choleski identification scheme is applied to $\hat{v}_t$. The vector containing the first q Choleski-identified shocks is the estimator of u_t.

The matrix $\hat{\Sigma}_0^{\hat{v}}$ is non-singular with probability one by Assumption A(b), so that it has a unique Cholesky factorization. Let $\hat{C}$ be the lower triangular Cholesky matrix such that $\hat{C}\hat{C}' = \hat{\Sigma}_0^{\hat{v}}$. Finally, let us partition $\hat{C}$ as

$$\hat{C} = \begin{pmatrix} \hat{Q} & 0 \\ \hat{R} & \hat{S} \end{pmatrix}$$

where $\hat{Q}$ is $q \times q$, $\hat{R}$ is $(m-q) \times q$ and $\hat{S}$ is $(m-q) \times (m-q)$. Notice that

$$\hat{C}^{-1} = \begin{pmatrix} \hat{Q}^{-1} & 0 \\ -\hat{S}^{-1}\hat{R}\hat{Q}^{-1} & \hat{S}^{-1} \end{pmatrix}.$$

We have

$$\hat{\chi}_t = \hat{A}(L)^{-1}\hat{C}\hat{C}^{-1}\hat{v}_t = \hat{A}(L)^{-1}\hat{B}_0\hat{u}_t + \hat{\omega}_t = (\hat{B}_0 + \hat{B}_1 L + \hat{B}_2 L^2 + \cdots)\hat{u}_t + \hat{\omega}_t, \quad (14)$$

where

$$\hat{B}_0 = \begin{pmatrix} \hat{Q} \\ \hat{R} \end{pmatrix}, \quad \hat{u}_t = \begin{pmatrix} \hat{Q}^{-1} & 0 \end{pmatrix}\hat{v}_t, \quad \hat{\omega}_t = \hat{A}(L)^{-1}\begin{pmatrix} 0 \\ \hat{S} \end{pmatrix}\begin{pmatrix} -\hat{S}^{-1}\hat{R}\hat{Q}^{-1} & \hat{S}^{-1} \end{pmatrix}\hat{v}_t$$

and the matrices $\hat{B}_k$, $k > 0$, are implicitly defined by the relation $\hat{B}_0 + \hat{B}_1 L + \hat{B}_2 L^2 + \cdots = \hat{A}(L)^{-1}\hat{B}_0$. Our second result is:

Proposition 2. Consistency of the estimated structural shocks and IRFs. *Under Assumption 1$'$, Assumptions 1 to 4, A and B:*

(a) $||\hat{u}_t - u_t|| = O_p\left(\max(r_{n,T}, 1/\sqrt{T})\right)$.
(b) $||\hat{B}_k - B_k|| = O_p\left(\max(r_{n,T}, 1/\sqrt{T})\right)$ *for any* k.

Proposition 2 states that the standard SVAR procedure applied to $\hat{\chi}_t$ produces consistent estimates for the q structural shocks u_{jt} and the corresponding IRFs. The proof is given in the Online Appendix C.

3.5 An estimator of χ_t fulfilling Assumptions A and B

From now on we focus on the ordinary principal component estimator. CC-SVAR analysis with the estimators proposed in Forni et al. (2000) and Forni et al. (2015, 2017) is left for future research. Hence we assume that the x's follow the structural dynamic factor model of Forni et al. (2009), though we do not impose $p = 1$.

Assumption 5. Static factor representation. *The common components* χ_{it} *are linear combinations of mutually orthogonal static factors* F_{kt}, $k = 1, \ldots, r$, *where* $r > q$. *The* r*-dimensional vector* F_t *has a rational MA representation in the structural shocks* u_t*:*

$$\chi_{it} = \lambda_{i1}F_{1t} + \cdots + \lambda_{ir}F_{rt} = \lambda_i F_t \tag{15}$$

$$F_t = B_F(L)u_t. \tag{16}$$

We have $\chi_t = \Lambda_m F_t$, Λ_m being the $m \times r$ matrix with rows λ_i, $i = 1, \ldots, m$, see equation (15); moreover, the dynamics of χ_t is related to the dynamics of the factors by the relation $B(L) = \Lambda_m B_F(L)$.

Some notation is needed for the following assumptions.

Notation 2.
(i) $\boldsymbol{x}_{nt} = (x_{1t} \ \cdots \ x_{nt})'$, $\boldsymbol{\chi}_{nt} = (\chi_{1t} \ \cdots \ \chi_{nt})'$ *and* $\boldsymbol{\xi}_{nt} = (\xi_{1t} \ \cdots \ \xi_{nt})'$. *Note that, by Assumption 4(a), we have* $x_t = \boldsymbol{x}_{mt}$, $\chi_t = \boldsymbol{\chi}_{mt}$ *and* $\xi_t = \boldsymbol{\xi}_{mt}$.
(ii) Γ^x_k, Γ^χ_k *and* Γ^ξ_k *are* k*-lag covariance matrices of the processes* $(\boldsymbol{x}_{nt})$, $(\boldsymbol{\chi}_{nt})$ *and* $(\boldsymbol{\xi}_{nt})$, *respectively.* Σ^χ_k, *see Notation 1, is the upper-left* $m \times m$ *sub-matrix of* Γ^χ_k, *which is* $n \times n$. $\hat{\Gamma}^x_k$, *the sample counterpart of* Γ^x_k, *is* $\sum_{t=k+1}^T \boldsymbol{x}_{nt}\boldsymbol{x}'_{n,t-k}/(T-k)$.
(iii) μ^χ_j *and* μ^ξ_j, $\hat{\mu}^\chi_j$ *and* $\hat{\mu}^\xi_j$, *are the* j*-th eigenvalues, in decreasing order of magnitude, of* Γ^χ_0 *and* Γ^ξ_0, $\hat{\Gamma}^\chi_0$ *and* $\hat{\Gamma}^\xi_0$, *respectively.*

Assumption 6. Pervasiveness of the factors and the shocks, non-pervasiveness of the idiosyncratic components.
(a) There exists constants $\underline{c}_j, \overline{c}_j$*,* $j = 1, \dots, r$*, such that* $\underline{c}_j > \overline{c}_{j+1}$*,* $j = 1, \dots, r-1$*, and*

$$0 < \underline{c}_j < \liminf_{n\to\infty} n^{-1}\mu_j^{\chi} \le \limsup_{n\to\infty} n^{-1}\mu_j^{\chi} \le \overline{c}_j$$

(b) There exists a real $\ell > 0$ *such that* μ_1^{ξ} *is bounded above by* ℓ*.*

Assumption 6(a) ensures that the static factors are pervasive; it could be replaced by suitable assumptions on the factor loading matrices Λ_n. Assumption 6(b) is obviously satisfied if the idiosyncratic components are mutually orthogonal and their variances are uniformly bounded. However, it is milder than mutual orthogonality in that it allows for a limited amount of cross-correlation.

Assumption 7. Uniform covariance ergodicity. *Denote by* $\gamma_{k,ij}^x$ *and* $\hat{\gamma}_{k,ij}^x$ *the entries of* Γ_k^x *and* $\hat{\Gamma}_k^x$ *respectively and similarly for* $\hat{\gamma}_{k,ij}^{\chi}$ *and* $\gamma_{k,ij}^{\chi}$*. There exists a positive real* ρ *such that*
(a) $T\ \mathrm{E}(\hat{\gamma}_{k,ij}^x - \gamma_{k,ij}^x)^2 < \rho$
(b) $T\ \mathrm{E}(\hat{\gamma}_{k,ij}^{\chi} - \gamma_{k,ij}^{\chi})^2 < \rho$
(c) $T\ \mathrm{E}(\hat{\gamma}_{k,ij}^{\chi\xi})^2 < \rho$
for i*,* j*,* k *and* T*.*

The above ergodicity properties can be obtained under the assumption of linearity of the processes and finite fourth cumulants of the driving shocks (see Hannan, 1970, Theorem 6). Here we assume in addition that the upper bound ρ is the same for all i.

Now let us define the principal component estimator of χ_t and establish its consistency as $n, T \to \infty$.

Definition 2. The principal component estimator. *Let* $\hat{W}^x$ *be the* $n \times r$ *matrix having on column* j*,* $j = 1, \dots, r$*, the eigenvector of* $\hat{\Gamma}_0^x$*, corresponding to* $\hat{\mu}_j^x$*, so that* $\hat{W}^x \boldsymbol{x}_{nt}$ *is the vector whose entries are the ordered principal components of* $\boldsymbol{x}_{nt}$*. Finally, let* $\hat{w}_{ij}^x$ *be the* i, j *entry of* $\hat{W}^x$*. The principal component estimator* *of* χ_{it}*,* $i = 1, \dots, m$ *is*

$$\hat{\chi}_{it} = \sum_{j=1}^{r} \hat{w}_{ij}^{x} \hat{W}^{x} \boldsymbol{x}_{nt}.$$

Proposition 3. Properties of the principal component estimator.

Under Assumptions 1-7 we have

(i) $\|\hat{\pi}_t\| = ||\hat{\chi}_t - \chi_t|| = O_p(\max(1/\sqrt{n}, 1/\sqrt{T}))$*.*

(ii) $||\hat{\Sigma}_k^{\chi} - \Sigma_k^{\chi}|| = O_p\left(1/\sqrt{T}\right)$*, for any* k*.*

Thus $\hat{\chi}_t$ fulfills Assumption A(a), with $r_{n,T} = \max(1/\sqrt{n}, 1/\sqrt{T})$, and Assumption B.

The proof is given in the Online Appendix D.

3.6 Summary of the estimation procedure

Based on the above results, we propose the following estimation procedure.

(E0) Select a large data set with n series and T observations for each series. Transform the series to get stationarity and standardize them to have zero mean and unit variance. The stationary series are the entries of our vector $\boldsymbol{x}_{nt}$; we denote the standardized series by $\boldsymbol{x}_{nt}^*$.

(E1) Estimate r. Out of the vast literature, beginning with Bai and Ng (2002), proposing consistent estimators $\hat{r}$, in the empirical applications in Section 5 we use Alessi et al. (2010). Choose m in such a way that $q < m \leq \hat{r}$. We discuss the choice of m in the next subsection.

(E2) Given $\hat{r}$ and m, estimate the common components as

$$\hat{\chi}_{it}^* = \sum_{j=1}^{r} \hat{w}_{ij}^{x} \hat{W}^{x} \boldsymbol{x}_{nt}^*, \quad i = 1, \dots, m.$$

Finally de-standardize the common components by multiplying each one of them by the standard deviation of the corresponding series and adding the sample mean.

If there is a strong a priori belief that variable s is free of measurement error, the variable itself can be included in the model, i.e. we can use for this variable the alternative estimator $\tilde{\chi}_{st} = x_{st}$ in place of $\hat{\chi}_{st}$. Moreover, any common component $\hat{\chi}_{st}$ which is of no direct interest for the analysis can in principle be replaced by an estimated factor, i.e. any one of the first r principal components of $\boldsymbol{x}^*_{nt}$, provided that the resulting vector has non-singular variance-covariance matrix.

(E3) Estimate a VAR for $\hat{\chi}_t$ (or $\tilde{\chi}_t$, $\tilde{\chi}_t$ being the estimator having $\tilde{\chi}_{st}$ in place of $\hat{\chi}_{st}$), to get an estimator of the matrix $A(L)$ and the VAR innovations v_t (see equation (11)).

(E4) Identify the structural shocks and the IRFs by SVAR techniques applied to $\hat{A}^{-1}(L)$ and v_t.

3.7 The choice of m

The choice of m is a key step of the estimation procedure. Here we discuss how to set this parameter.

Our first recommendation is to set m larger than $q+1$. If χ_t were observable, the choice $m = q+1$ would produce the correct result as shown in Simulation 1 and Simulation 5, Appendix F.1. However only an estimate of χ_t is available; as n is finite, $\hat{\chi}_t$ still includes a residual of the idiosyncratic components, so that it is not exactly singular. When $m = q+1$ the estimates can still be inaccurate even if the residual idiosyncratic component is small. This problem disappears when $m > q+1$. This point is discussed thoroughly in Appendix E and illustrated with Simulation 5, Appendix F.1.

A simple way to ensure that $m > q+1$ is to set m equal to its largest possible value, i.e. $m = r$. There are two additional arguments in favor of this choice.

First, in empirical applications, q is unknown and has to be determined by existing information criteria. Such criteria, albeit consistent, may deliver wrong results in small samples. Thus setting m to its maximum value $\hat{r}$ is the safest choice. If we choose $m = \hat{r}$, estimation of q in a CC-SVAR is not strictly necessary. On the other hand, estimating q could be useful to check that r is actually larger than q.

Second, if $m = \hat{r}$, the estimated shocks of interest and the corresponding estimated IRFs are the same, irrespective of the choice of the variables included in the VAR.[10] The intuition is simple: since the entries of $\hat{\chi}_t$ are linear combinations

[10]This result holds only asymptotically in the case $q < m < r$.

of the estimated factors in $\hat{F}_t$ (i.e. the first m principal components of our large data set), when $\hat{\chi}_t$ is m-dimensional it spans the same linear space as $\hat{F}_t$, for any choice of the variables (provided that the loading matrix is invertible). This fact has two important consequences. The first is that selecting the variables to be included in the CC-SVAR is not an issue. The natural choice is the set of variables which are needed for identification and, if required to complete the information set, we can include the common components of other variables of interest, or even some of the estimated factors, i.e. the principal components themselves. This is what is done in some of the simulations below and in the empirical application.

The second is that, if we are interested in the IRFs of some variables which have not been included in the CC-SVAR, we can simply estimate another CC-SVAR including these variables. This practice, which is common in empirical work, is questionable within the standard SVAR framework, since, as shown in the Introduction, changing the variables may change dramatically the information set and therefore the estimated shock of interest. By contrast, it is perfectly justified within the CC-SVAR approach, when setting $m = \hat{r}$.

Despite recommending $m = \hat{r}$, we should acknowledge that there might be situations in which this choice is problematic. This is when $\hat{r}$ is large, so that $m = \hat{r}$ might entail a too large number of parameters to estimate, particularly when the sample is small in the time dimension. In this case it may be preferable to set $m < \hat{r}$. We should however estimate q by using a consistent criterion and check that $m > \hat{q} + 1$.

3.8 The choice of r

As stated above, r can be estimated by any one of the available consistent criteria. However different consistent criteria often provide different estimates in small samples. In the Online Appendix, Section F.2, we show that the estimates of the IRFs improve as $\hat{r}$ increases from values below r, the true value, to r and stabilize for values greater than r.

The feature that estimated IRFs do not change when $\hat{r} > r$ is also useful to control for the presence of weak factors, that is, factors that explain a small fraction of the variance and might not be captured by standard information criteria typically used to estimate r.

This finding can be used in empirical applications, where r is not known. We can use the estimate $\hat{r}$ as the baseline specification and estimate the IRFs. Then we can assess the robustness of the results by using a range of values for $\hat{r}$ around the baseline.

4 Simulations

The procedure described in Section 3.6 is now applied to simulated data sets based on the model discussed in Section 2. Firstly we rewrite model (2) in static-factor form. Let

$$F_t = (k_t \ u_{a,t} \ u_{\tau,t} \ u_{\tau,t-1} \ u_{\tau,t-2})'.$$

The 5-dimensional vector F_t has the following singular VAR(1) representation:

$$\begin{pmatrix} k_t \\ u_{a,t} \\ u_{\tau,t} \\ u_{\tau,t-1} \\ u_{\tau,t-2} \end{pmatrix} = \begin{pmatrix} \alpha & 0 & -\delta & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} k_{t-1} \\ u_{a,t-1} \\ u_{\tau,t-1} \\ u_{\tau,t-2} \\ u_{\tau,t-3} \end{pmatrix} + \begin{pmatrix} 1 & -\delta\theta \\ 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} u_{a,t} \\ u_{\tau,t} \end{pmatrix}. \tag{17}$$

Defining $\chi_t = (a_t \ k_t \ \tau_t)'$, we have

$$x_t = \Lambda F_t + \xi_t \tag{18}$$

where

$$\Lambda = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

We generate a vector z_t including 100 additional time series ($T = 200$) as

$$z_t = \Lambda^z F_t + \xi_t^z \tag{19}$$

where Λ^z is the 100×5 matrix matrix of the loadings. The entries of Λ^z are generated independently from a standard normal distribution. Hence $\boldsymbol{x}_{nt} = (x_t' \ z_t')'$ and $\boldsymbol{\xi}_{nt} = (\xi_t{}' \ \xi_t^{z\prime})'$. We generate the measurement errors $\boldsymbol{\xi}_{nt}$ assuming that $\boldsymbol{\xi}_{nt} \sim$

$N(0, \sigma_i)$ where σ_i is uniformly distributed in the interval $(0, 0.5)$, so that different variables have measurement errors of different size (on average, the idiosyncratic components account for about 11% of total variance).

In Simulation 2 we compare the CC-SVAR with the estimation procedure of (Forni et al., 2009) (Standard Procedure SDFM henceforth) and the FAVAR. Firstly, we estimate: (a) a Standard Procedure SDFM, with two lags in the VAR, with a too small number of common shocks, i.e. $\hat{q} = 1$, and (b) a Standard Procedure SDFM, two lags, with the correct number of shocks, i.e $\hat{q} = 2$. In both cases $\hat{r}$ is, correctly, equal to 5. Secondly, we estimate (c) a CC-SVAR(2) with $m = \hat{r} = 5$. Finally, we estimate (d) a FAVAR(2) including capital, taxes, technology and the first two principal components. In all cases we use two lags in the estimation. Again, we perform 1000 replications.

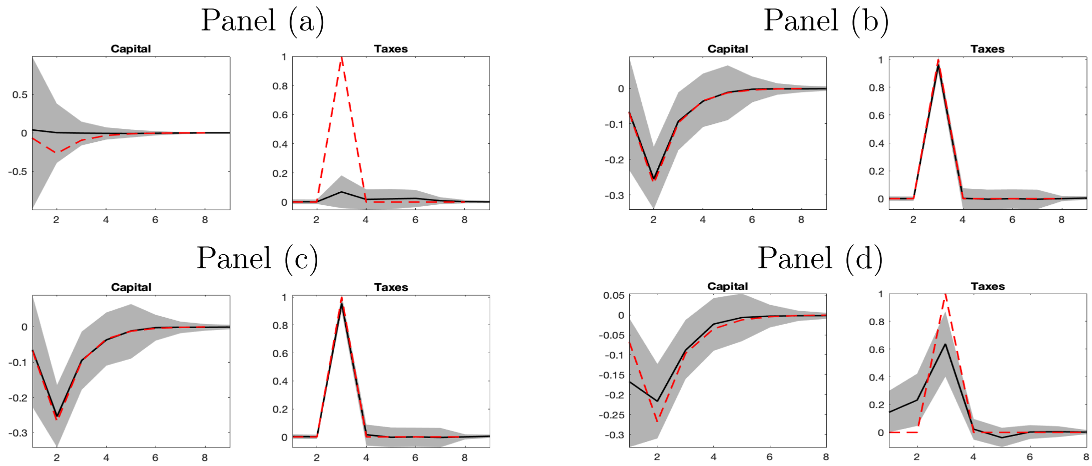


Figure 4: Simulation 2. Standard Procedure SDFM, CC-SVAR, FAVAR. Estimated IRFs for the tax shock. The red dashed lines are the theoretical IRFs. The solid lines represent the mean (across 1000 simulated datasets) of the point estimates. The grey areas contain the point estimates between the 16th and 84th percentiles. Panel (a): Standard Procedure SDFM, with two lags, with $\hat{q} = 1 < q$ ($\hat{r} = r = 5$). Panel (b): Standard Procedure SDFM, two lags, with $\hat{q} = q = 2$ ($\hat{r} = r = 5$). Panel (c): CC-SVAR(2) with Capital, Taxes and the first 3 principal components ($m = \hat{r} = 5$). Panel (d): FAVAR(2) with Capital, Taxes and the first 3 principal components.

The results are reported in Figure 4. Panel (a) shows the results for the misspecified SDFM. Not surprisingly, with this data generating process, where $q = 2$, setting $\hat{q} = 1$ has dramatic consequences on the estimates of the impulse response functions. With a different DGP and a larger q we can expect a smaller bias.

However, the point is that, in real data applications, q can be underestimated, leading to sizable estimation errors.

Panels (b) and (c) refer to the correctly specified SDFM and the CC-SVAR, respectively. It is hard to see any difference between the two figures. This suggests that the rank reduction step typical of the factor model can be ignored with no consequences on the quality of the estimates. Moreover, as argued above, with the CC-SVAR (with $m = r$) we do not need an estimate of q, which is safer, in view of the results of Panel (a).

Finally, panel (d) reports the results for the FAVAR model. Owing to measurement errors, the estimates are clearly worse than those in panels (b) and (c).

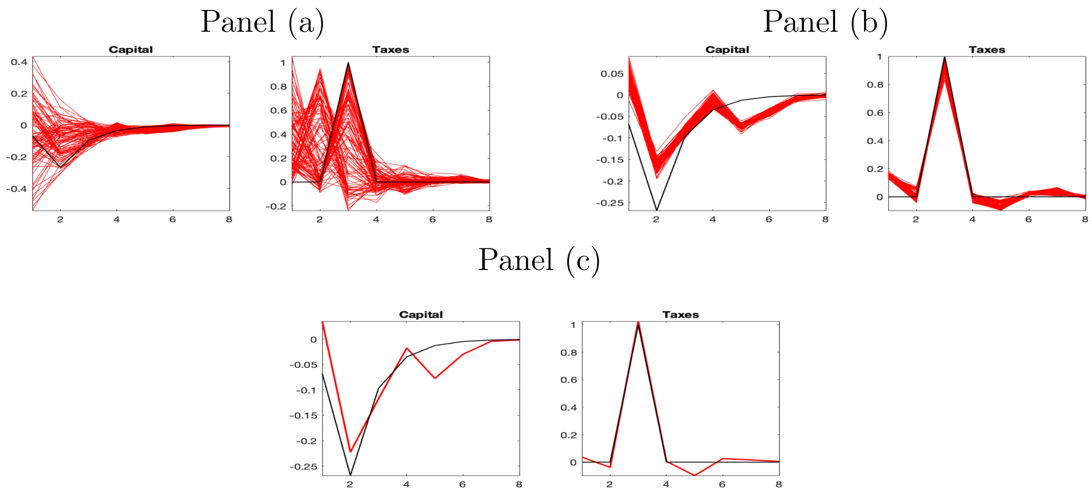


Figure 5: Simulation 3. Different variable specifications for a deficient VAR, the FAVAR and the CC-SVAR. Estimated IRFs for the tax shock, for a single simulated data set. The black lines are the theoretical IRFs. The red lines are the CC-SVAR estimates obtained with different variable specifications. Panel (a): SVAR(2) with Capital, Taxes and a third variable, changing across specifications. Panel (b): FAVAR(2) with Capital, Taxes, a third variable, changing across specifications, and the first two principal components. Panel (c): CC-SVAR(2) with Capital, Taxes, a third variable, changing across specifications, and the first two principal components.

Simulation 3 deals again with the choice of the specification of the variables included in the model. Here, we use just one data set and compare the SVAR, the FAVAR and the CC-SVAR. Regarding the SVAR model, we estimate one hundred of three-variable VAR(2) specifications, including capital, taxes, and the $(3+i)$-th variable, $i = 1, ..., 100$. The results are reported in Figure 5, Panel (a). The figure shows that the choice of the third variable produces huge differences in the estimated impulse response functions, both because of the information delivered by the common component of the third variable and the extent of the contamination induced by the measurement error. Panel (b) refers to FAVAR models including capital, taxes, the $(3+i)$-th variable, $i = 1, ..., 100$, plus the first two principal components. Again we use two lags. Here the estimated IRFs are much closer to each other, since information is not deficient. However, there is still some variability due to the size of the measurement error included in the third variable. Panel (c) refers to the CC-SVAR, where, as already argued in Section 3.7, all IRFs are identical.

5 Empirical application

In this section we illustrate the advantages of CC-SVAR analysis by means of an applications on monetary policy shocks. Our main results are the following. (I) As a consequence of non-fundamentalness and measurement errors the results of the SVAR analysis are rather unstable, depending on which variables are included in the vector. Thus the conclusions on the effects of structural shocks on macroeconomic variables are not robust. (II) Some improvement is obtained with FAVAR models, although the effects of measurement errors are still evident. (III) With CC-SVAR, instability disappears and robust conclusions can be drawn. Independently of the choice of variables, contractionary monetary policy shocks reduce prices and economic activity.

To estimate the common components we use the monthly dataset of McCracken and Ng (2016).[11] We exclude a few variables to obtain a balanced panel and we end up with a monthly dataset with 122 variables. We transform each series to reach stationarity. We apply the criterion proposed by Alessi et al. (2010) and find a number of static factors $\hat{r} = 8$. Thus we use, as baseline specification, $\hat{r} = 8$.

We consider 50 different VAR specifications characterized by different vectors x_t^j, $j = 1, ..., 50$. Each of them includes five variables. Four of them are common to all vectors: the unemployment rate, industrial production growth, inflation

[11] The data set is available at https://research.stlouisfed.org/econ/mccracken/fred-databases/.

and a policy rate. Each model includes an additional variable of the panel which differs across models and is chosen randomly. The sample spans from 1977:6 (the beginning of the Volcker era) to 2008:12 (to exclude the ZLB period).

For each of the 50 specifications, we identify the shock using three different identification schemes. Firstly, a Cholesky scheme. The ordering of the five variables is the following: the unemployment rate, industrial production growth, inflation, the 1-year bond rate and the fifth additional variable. The monetary policy shock is the fourth one.

The second and the third schemes are based on the proxy SVAR method (Mertens and Ravn (2013) and Stock and Watson (2018)). In the second we use the Gertler and Karadi (2015) instrument (GK henceforth). In the third the Miranda-Agrippino and Ricco (2021) instrument (MAR henceforth). The policy rate is the 1-year bond rate, to be consistent with the specifications used in both the above mentioned papers.

The first column of Figures 6-8 reports the estimated IRFs for a VAR(6). Each blue line represents the impulse response function of a particular specification, so that each box contains 50 different lines. A striking result is the high degree of heterogeneity in the estimated responses, despite the fact that specifications differ only for the fifth variable. The result resembles the one in the simulation exercise of Figure 5. With the GK instrument, in particular, not only the magnitude, but even the sign of the responses may change, depending on the choice of the fifth variable. With the Cholesky identification we have the price puzzle for all specifications but one. When using the GK instrument the effects of a contractionary shock appear to be expansionary for most specifications. All in all, the results suggest that drawing robust conclusions about the propagation mechanisms of monetary policy shocks is very hard. Indeed, the effects differ substantially across specifications both qualitatively and quantitatively.

To understand the effects of enlarging the information set, we augment each 5-variable specification with the first 3 principal components. We then run a FAVAR(6) and apply the three identifications schemes. In this case information is enhanced but still the model can suffer the problem arising from the presence of measurement error.

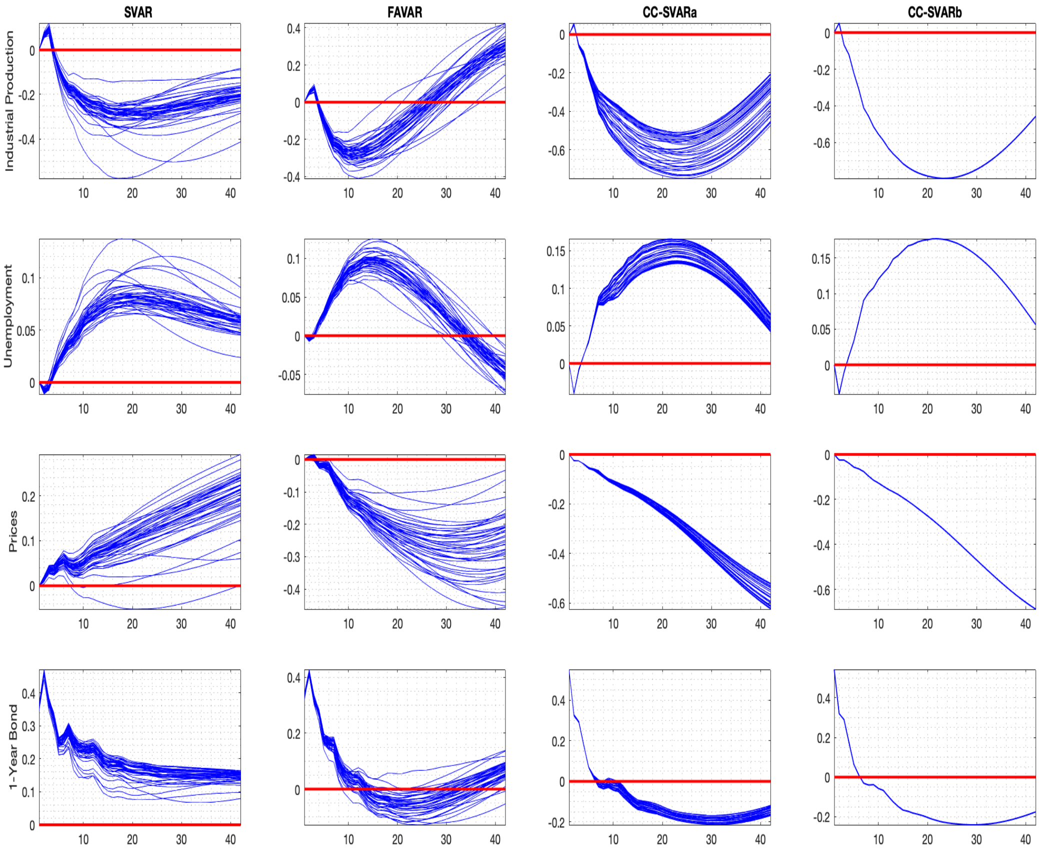


Figure 6: US monthly data. The IRFs of a monetary policy shock. Cholesky identification. The red lines are the CC-SVAR estimates obtained with different variable specifications. First column: SVAR(6) for 118 five-variable specifications, differing for the fifth variable. Second column: FAVAR(6) the variables in the first column are augmented with the first 8 principal components. Third column: CC-SVAR(6): the variables in the first column are replaced with their common components; in addition, we include the first 3 principal components ($\hat{r} = 8$).

The results are reported in the second column of Figures 6–8. Completing information seems to have important consequences, particularly because the price puzzle disappears with the Cholesky identification scheme, as observed in Bernanke et al. (2005). However, three principal components are not enough to solve the puzzles of the GK identification, and still results vary considerably across specifications with all identification schemes.

To understand the implications of measurement errors we repeat the same exercise as before but replacing the variables with their common components. So, we estimate 50 different CC-SVAR specifications which include the common components of the interest rate, industrial production growth, inflation and un-

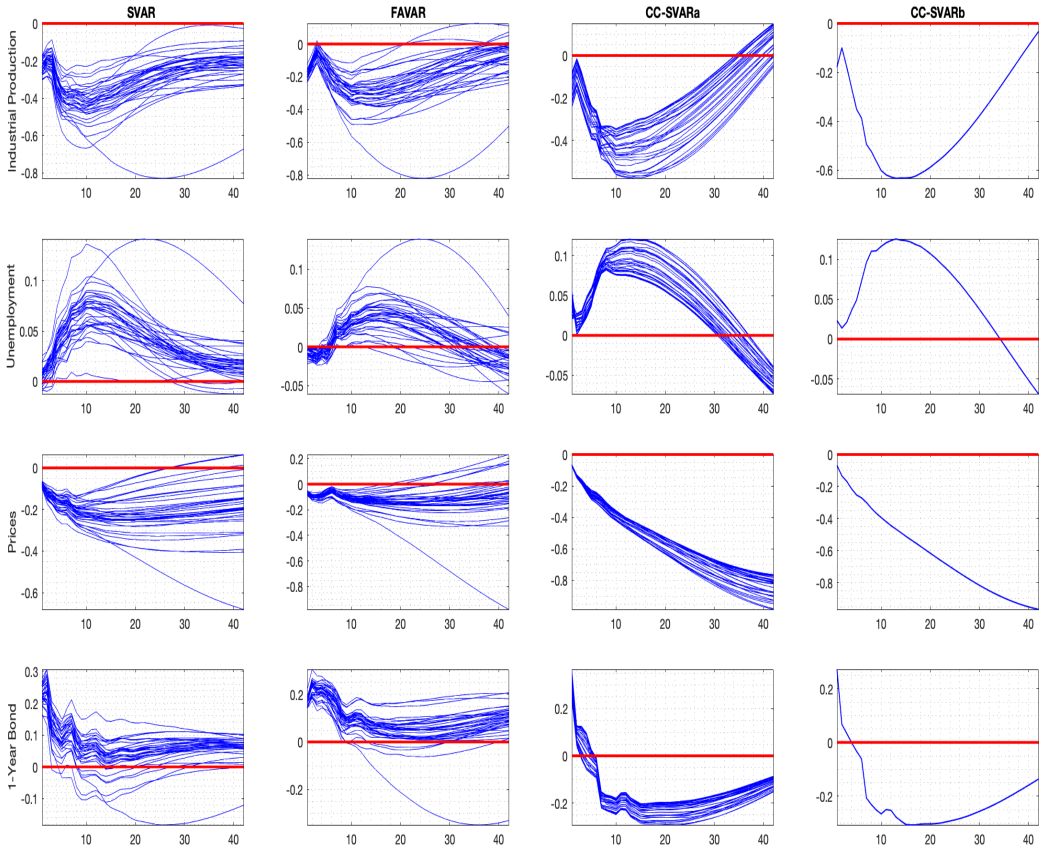


Figure 7: US monthly data. The IRFs of a monetary policy shock. Proxy MAR identification. The red lines are the CC-SVAR estimates obtained with different variable specifications. First column: SVAR(6) for 118 five-variable specifications, differing for the fifth variable. Second column: FAVAR(6) the variables in the first column are augmented with the first 8 principal components. Third column: CC-SVAR(6): the variables in the first column are replaced with their common components; in addition, we include the first 3 principal components ($\hat{r} = 8$).

employment, plus a fifth common component which changes for each specification, and either two principal components ($m = 7$, third column), or three principal components ($m = r = 8$, fourth column).[12] We see in the third column of the figures that results are much more robust to specification changes. In the fourth column, as argued in Section 3.7, all lines are perfectly overlapping.

Importantly, with the CC-SVAR all puzzles disappear; moreover, results are quantitatively similar not only across different VAR specifications, but also across different identification schemes.

To assess the robustness of the results to changes of the number of factors, we

[12]Notice that when $m = r = 8$ using any triple of variables in place of the first three principal components would yield the identical results.

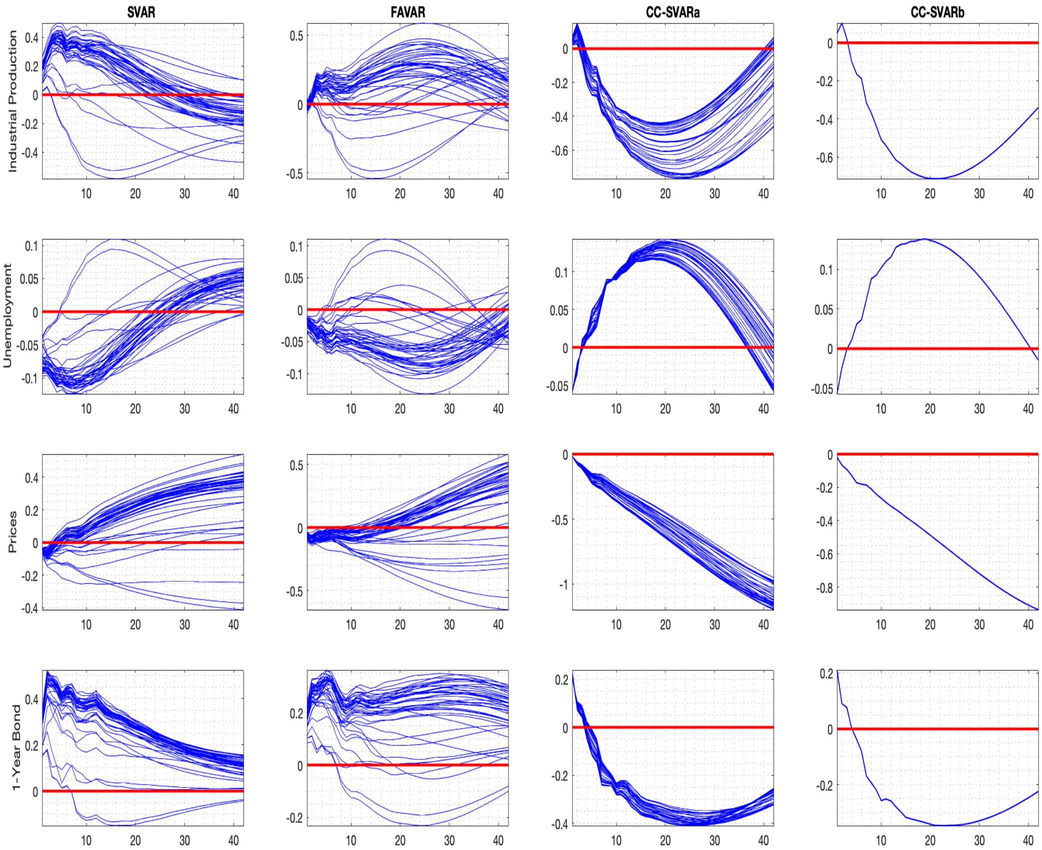


Figure 8: US monthly data. The IRFs of a monetary policy shock. Proxy GK identification. The red lines are the CC-SVAR estimates obtained with different variable specifications. First column: SVAR(6) for 118 five-variable specifications, differing for the fifth variable. Second column: FAVAR(6) the variables in the first column are augmented with the first 8 principal components. Third column: CC-SVAR(6): the variables in the first column are replaced with their common components; in addition, we include the first 3 principal components ($\hat{r} = 8$).

repeat the CC-SVAR analysis using $m = \hat{r} = 7, 8, 9, 10, 11$ common components. To complete information, we include in the VAR the five common components plus the first $\hat{r} - 5$ principal components. The results are displayed in Figure 9. We see that the results obtained with different values of $\hat{r}$ are very similar to each other for all identification schemes. In conclusion, using the SVAR analysis, the results obtained with the three identification schemes, including the external instrument approach, crucially depend on the specification of the model, more precisely on the choice of the variables. In addition, results vary a lot across different identification schemes.

On the contrary, the IRFs of the CC-SVAR analysis, with the three identifi-

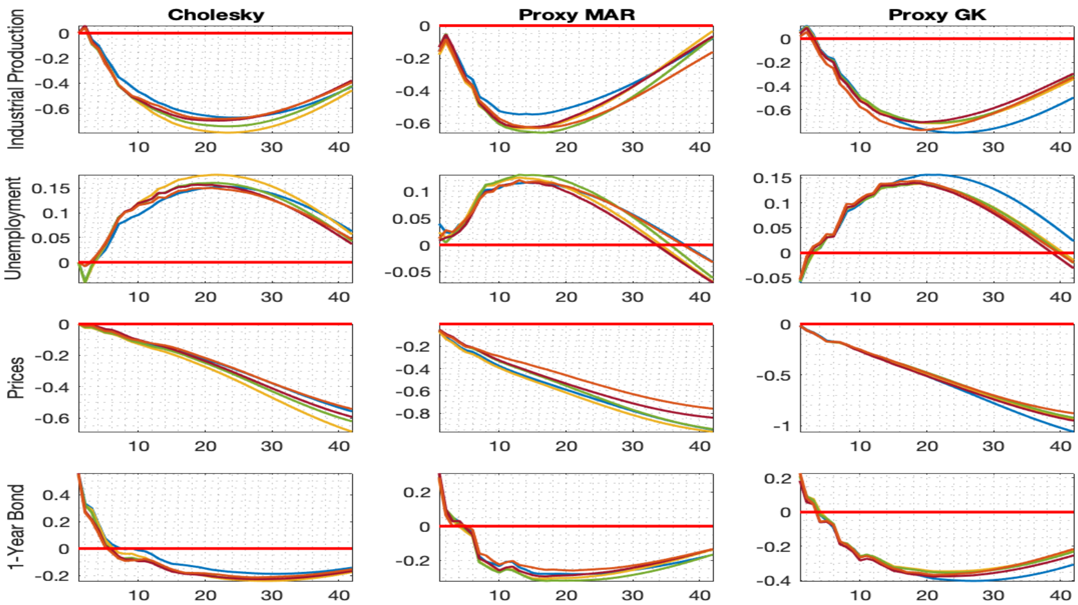


Figure 9: US monthly data. The IRFs of a monetary policy shock. CC-SVAR(6) with $m = r$, using different values of r. Black dotted line: $r = 6$. Blue dashed line: $r = 8$. Red solid line: $r = 10$.

cation procedures, apart from minor quantitative differences, are very similar, a result that runs counter the growing consensus that high frequency identification with external instruments is a better approach to identify monetary policy shocks, in comparison to the Cholesky scheme. Moreover, the IRFs are in line with the standard view of the transmission mechanism of monetary policy shocks, in which contractionary policy shocks reduce prices and slow down economic activity in the short run.

6 Conclusions

CC-SVARs apply SVAR techniques to singular vectors including the common components of the variables of interest. We claim that CC-SVARs provide a solution to the difficulties arising with possible non-fundamentalness of the structural shocks and measurement errors in macroeconomic variables. In our empirical application the CC-SVAR produces results that, unlike those obtained with SVAR analysis, are both sensible and robust with respect to changes in specification.

Although we have introduced and discussed the CC-SVAR technique with reference to the DFM model described in Section 3.5, a similar method applies in the General Dynamic Factor Model, that is when the assumption of a finite number of static factors does not necessarily hold and the common components are estimated by frequency-domain methods, see Forni et al. (2000) and Forni et al. (2015, 2017). This however is left for future research.

References

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

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

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

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

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

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 ImpulseâĂŞResponse Functions with I(1) cointegrated factors. *Journal of Econometrics 221*(2), 455–482.

Barsky, R. B., S. Basu, and K. Lee (2015). Whither news shocks? *NBER Macroeconomics Annual 29*(1), 225–264.

Barsky, R. B. and E. R. Sims (2009). News shocks. NBER Working Papers 15312.

Beaudry, P., P. Fève, A. Guay, and F. Portier (2019). When is nonfundamentalness in SVARs a real problem? *Review of Economic Dynamics 34*, 221–243.

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.

Blanchard, O. J. and D. Quah (1989). The dynamic effects of aggregate demand and supply disturbance. *The American Economic Reivew 79*, 655–673.

Canova, F. (2007). *Methods for Applied Macroeconomics.* Princeton, New Jersey: Princeton University Press.

Chari, V., P. J. Kehoe, and E. R. McGrattan (2008). Are structural VARs with long-run restrictions useful in developing business cycle theory? *Journal of Monetary Economics 55*, 1337–1352.

Christiano, L., M. Eichenbaum, and R. Vigfusson (2003). What happens after a technology shock? NBER Working Papers 9819.

Christiano, L., C. L. Ilut, R. Motto, and M. Rostagno (2010). Monetary policy and stock market booms. NBER Working Papers 16402.

Fernández-Villaverde, J., J. F. Rubio-Ramírez, T. J. Sargent, and M. W. Watson (2007). ABCs (and Ds) of understanding VARs. *American Economic Review 97*, 1021–1026.

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. and L. Gambetti (2014). Sufficient information in structural VARs. *Journal of Monetary Economics 66*(C), 124–136.

Forni, M., L. Gambetti, M. Lippi, and L. Sala (2017a). Noise bubbles. *The Economic Journal 127*(604), 1940–1976.

Forni, M., L. Gambetti, M. Lippi, and L. Sala (2017b). Noisy news in business cycles. *American Economic Journal: Macroeconomics 9*(4), 122–152.

Forni, M., L. Gambetti, and L. Sala (2014). No news in business cycles. *The Economic Journal 124*(581), 1168–1191.

Forni, M., L. Gambetti, and L. Sala (2019). Structural VARs and noninvertible macroeconomic models. *Journal of Applied Econometrics 34*(2), 221–246.

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 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 (1997). *Aggregation and the microfoundations of dynamic macroeconomics*. Oxford: Oxford University Press.

Forni, M. and M. Lippi (1999). Aggregation of linear dynamic microeconomic models. *Journal of Mathematical Economics 31*(1), 131 – 158.

Galí, J. (1999). Technology, employment, and the business cycle: do technology shocks explain aggregate fluctuations? *American Economic Review 89*(1), 249–271.

Gertler, M. and P. Karadi (2015). Monetary policy surprises, credit costs, and economic activity. *American Economic Journal: Macroeconomics 7*(1), 44–76.

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

Hannan, E. J. (1970). *Multiple Time Series*. New York: Wiley.

Hansen, L. P. and T. J. Sargent (1991). Two difficulties in interpreting vector autoregressions. In L. P. Hansen and T. J. Sargent (Eds.), *Rational Expectations Econometrics*, pp. 77–120. Boulder: Westview Press.

Jaimovich, N. and S. Rebelo (2009). Can news about the future drive the business cycle? *American Economic Review 99*(4), 1097–1118.

Leeper, E. M., T. B. Walker, and S. S. Yang (2013). Fiscal foresight and information flows. *Econometrica 81*, 1115–1145.

Lippi, M. (2021). Validating DSGE models with SVARs and high-dimensional dynamic factor models. *Econometric Theory*, 1–19. https://doi.org/10.1017/S0266466621000542.

Lippi, M. and L. Reichlin (1993). The dynamic effects of aggregate demand and supply disturbances: Comment. *American Economic Review 83*, 644–652.

McCracken, M. W. and S. Ng (2016). FRED-MD: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics 34*(4), 574–589.

Mertens, K. and M. O. Ravn (2013). The dynamic effects of personal and corporate income tax changes in the United States. *The American Economic Review 103*(4), 1212–1247.

Miranda-Agrippino, S. and G. Ricco (2021). The transmission of monetary policy shocks. *American Economic Journal: Macroeconomics 13*, 74–107.

Ramey, V. (2016). Macroeconomic shocks and their propagation. Volume 2, Chapter Chapter 2, pp. 71–162. Elsevier.

Schmitt-Grohé, S. and M. Uribe (2012). What's news in business cycles. *Econometrica 80*(6), 2733–2764.

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica 48*(1), 1–48.

Sims, C. A. and T. Zha (2006). Does monetary policy generate recessions? *Macroeconomic Dynamics 10*, 231–272.

Sims, E. (2012). News, non-invertibility, and structural VARs. In N. Balke, F. Canova, F. Milani, and M. Wynne (Eds.), *DSGE Models in Macroeconomics: Estimation, Evaluation, and New Developments (Advances in Econometrics, Vol. 28)*, pp. 81–135. Bingley, U.K.: Emerald Group Publishing Limited.

Smets, F. and R. Wouters (2007). Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach. *American Economic Review 97*(3), 586–606.

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. NBER Working Papers 11467.

Stock, J. H. and M. W. Watson (2018). Identification and estimation of dynamic causal effects in macroeconomics using external instruments. *The Economic Journal 128*(610), 917–948.

van der Waerden, B. L. (1953). *Modern Algebra*, Volume I. New York: Frederic Ungar Publishing Co.

For Online Publication - Appendix

A Appendix to Section 3.2

A.1 Genericity of zerolessness

We firstly need an explicit parameterization of (6)-(7) and the definition of a generic property.

Definition P. *The meaning of s_1, s_2 and ϖ being defined in Assumption 1, a parameterization of* (6)-(7) *is a triple $(\nu, \mathcal{P}, \omega)$, where: (i) ν is a positive integer, the number of parameters, (ii) $\mathcal{P}$, the parameter space, is an open subset of $\mathbb{R}^\nu$, (iii) ω is a collection of ϖ functions from $\mathcal{P}$ to $\mathbb{R}$:*

$$h_{ij,h}(\mathbf{p}) = \frac{\phi_{ij,h}(\mathbf{p})}{\tilde{\phi}_{ij,h}(\mathbf{p})}, \quad k_{ij,\ell}(\mathbf{p}) = \frac{\psi_{ij,\ell}(\mathbf{p})}{\tilde{\psi}_{ij,\ell}(\mathbf{p})}, \tag{20}$$

for $i = 1, \ldots, m$, $j = 1, \ldots, q$, $h = 0, \ldots, s_1$, $\ell = 1, \ldots, s_2$, where the numerators and denominators in (20) *are (finite) polynomials in* $\mathbf{p}$ *(so that the functions in* (20) *are rational). Moreover, for all $\mathbf{p} \in \mathcal{P}$, (a) $\tilde{\phi}_{ij,h}(\mathbf{p}) \neq 0$ and $\tilde{\psi}_{ij,h}(\mathbf{p}) \neq$* 0, *(b) the matrices $H(L)$ and $K(L)$ fulfill Assumptions 1 and 2 for all $\mathbf{p} \in \mathcal{P}$.*

Assuming that $\mathcal{P}$ is open is a convenient simplification. All the results below hold if $\mathcal{P}$ contains a subset which is open in $\mathbb{R}^\nu$ and dense in $\mathcal{P}$. Definition P includes:

(i) Models with a structural economic-theory motivation, like (2), with the minor modification $\tau > 0$. As a rule, in this case $\nu < \varpi$, so that the parameterization produces restrictions on the coefficients h and k in (7). Assuming that $\mathcal{P}$ is an open subset of $\mathbb{R}^\nu$ is just a convenience. All the arguments below are still valid if $\mathcal{P}$ contains a subset $\tilde{\mathcal{P}}$ which is open in $\mathbb{R}^\nu$ and dense in $\mathcal{P}$.

(ii) The *Free-Parameter* case in which the parameters are the coefficients $h_{ij,s}$ and $k_{ij,s}$ themselves, $\mathcal{P} \subseteq \mathbb{R}^\varpi$, with ω being the identity function.

Definition G. Generic property in $\mathcal{P}$. *We say that a property holds generically in $\mathcal{P}$ if it holds in an open and dense subset of $\mathcal{P}$.*

We now make explicit reference to the parameters $\mathbf{p}$. Thus we have $H(\mathbf{p}, L)$, $K(\mathbf{p}, L)$, $k_{ij}(\mathbf{p}, L)$, $h_{ij}(\mathbf{p}, L)$, etc.

Let

$$D_a(\mathbf{p}, L) = D_{a,0}(\mathbf{p}) + D_{a,1}(\mathbf{p})L + \cdots, \quad a \in \mathcal{M}, \quad \mathcal{M} = \left\{1, \ldots, \frac{m!}{q!(m-q)!}\right\}, \tag{21}$$

be the determinant of the a-th $q \times q$ submatrix of $K(\mathbf{p}, L)$ (the ordering of the submatrices is immaterial). For a given $\mathbf{p}$, a sufficient condition for zerolessness of $K(\mathbf{p}, L)$ is that for at least a couple $a, b \in \mathcal{M}$, $a \neq b$, $D_a(\mathbf{p}, L)$ and $D_b(\mathbf{p}, L)$ have no common zero. The following statement generalizes Anderson and Deistler (2008b), Proposition 1, to the case in which the coefficients of the entries of the matrix K are restricted by the parameterization P above:

Proposition AD2. *Given a parameterization $(\nu, \mathcal{P}, \omega)$, define $\mathcal{Z}$ as the set of all $\mathbf{p}$ such that for at least a couple $a, b \in \mathcal{M}$, $a \neq b$, $D_a(\mathbf{p}, L)$ and $D_b(\mathbf{p}, L)$ have no common zero, and $\mathcal{W} = \mathcal{P} - \mathcal{Z}$, i.e. as the set of all $\mathbf{p}$ such that for all couples $a, b \in \mathcal{M}$, $a \neq b$, $D_a(\mathbf{p}, L)$ and $D_b(\mathbf{p}, L)$ have common zeros. Then either*
(Z) Generically $\mathbf{p} \in \mathcal{Z}$, so that $K(\mathbf{p}, L)$ is generically zeroless, or
(W) Generically $\mathbf{p} \in \mathcal{W}$.

Proposition AD2 can be restated by saying that if (Z) holds [if (W) holds] for an open subset of $\mathcal{P}$, then (Z) holds [(W) holds] generically in $\mathcal{P}$.

Proof. We proceed by steps.
(i) The coefficients of $D_a(\mathbf{p}, L)$ are rational functions with no poles in $\mathcal{P}$, hence each one of them is either zero for all $\mathbf{p} \in \mathcal{P}$ or generically non-zero. Thus, either (A) $D_a(\mathbf{p}, L)$ is the zero polynomial for all $\mathbf{p} \in \mathcal{P}$, or (B) there exists a non-negative integer d_a such that generically $D_a(\mathbf{p}, L)$ has degree d_a with non-zero leading coefficient.
(iii) If $d_a = 1$ for some $c \in \mathcal{M}$, so that generically $D_c(\mathbf{p}, L)$ has no roots, then (Z) holds.
(iv) Assuming that either (A) holds for all $a \in \mathcal{M}$ or (A) holds for all a but one $c \in \mathcal{M}$ and that $d_c > 1$, then (W) holds.
(v) Lastly, assume that the subset of $\widetilde{\mathcal{M}}$ for which (B) holds contains $\tilde{M} \geq 2$

elements and that for $c \in \widetilde{\mathcal{M}}$ we have $d_c > 1$. We need the following definition and result:

Proposition R. *The resultant of the scalar polynomials with real coefficients*

$$A(x) = a_v x^v + \cdots + a_0, \quad B(x) = b_w x^w + \cdots + b_0,$$

with $v > 0$, $w > 0$, is a polynomial function R, depending on a_i, $i = 0, \ldots, v$ and b_j, $j = 0, \ldots, w$, with integer coefficients. If $a_v \neq 0$ and $b_w \neq 0$, then

$$R(a_v, \ldots, a_0; b_w, \ldots, b_0) = 0,$$

if and only if $A(x)$ and $B(x)$ have a common (complex) root. See e.g. van der Waerden (1953), pp. 83-5.

Let $\mathcal{P}^\dagger$ be the subset of $\mathcal{P}$ such that for $\mathbf{p} \in \mathcal{P}^\dagger$ the leading coefficient of $D_c(\mathbf{p}, L)$, for all $c \in \widetilde{\mathcal{M}}$, is not zero. $\mathcal{P}^\dagger$ is open and dense in $\mathcal{P}$. Let $R_{ab}(\mathbf{p})$ be the resultant of $D_a(\mathbf{p}, L)$ and $D_b(\mathbf{p}, L)$ and

$$\mathcal{R}(\mathbf{p}) = \sum_{c,d \in \widetilde{\mathcal{M}}, c \neq d} R_{cd}(\mathbf{p})^2. \tag{22}$$

As $\mathcal{R}(\mathbf{p})$ is a rational function with no poles in $\mathcal{P}$, then one of the following alternatives hold:

(1) Generically in $\mathcal{P}$, $\mathcal{R}(\mathbf{p}) > 0$. Thus generically in $\mathcal{P}^\dagger$, $\mathcal{R}(\mathbf{p}) > 0$. The leading coefficients of $D_c(\mathbf{p}, L)$ and $D_d(\mathbf{p}, L)$ are not zero for $c, d \in \widetilde{\mathcal{M}}$ and $\mathbf{p} \in \mathcal{P}^\dagger$. As each addendum in (22) is either zero or generically non zero in $\mathcal{P}$, by Proposition R, there exist $f, g \in \widetilde{\mathcal{M}}$, $f \neq g$, such that, generically in $\mathcal{P}^\dagger$, $D_f(\mathbf{p}, L)$ and $D_g(\mathbf{p}, L)$ have no common roots, so that (Z) holds. Of course genericity in $\mathcal{P}^\dagger$ implies genericity in $\mathcal{P}$.

(2) $\mathcal{R}(\mathbf{p}) = 0$ for all $\mathbf{p} \in \mathcal{P}$. By Proposition R, $D_c(\mathbf{p}, L)$ and $D_d(\mathbf{p}, L)$ have a common root for all $c, d \in \widetilde{\mathcal{M}}$, $c \neq d$ and all $\mathbf{p} \in \mathcal{P}^\dagger$. Moreover, if c or d do not belong to $\widetilde{\mathcal{M}}$, so that either $D_c(\mathbf{p}, L)$ or $D_d(\mathbf{p}, L)$ is the zero polynomial, then $D_c(\mathbf{p}, L)$ and $D_d(\mathbf{p}, L)$ have common roots for all $\mathbf{p} \in \mathcal{P}$. Thus generically in $\mathcal{P}^\dagger$, and therefore in $\mathcal{P}$, (W) holds. Q.E.D.

The equation $\mathcal{R}(\mathbf{p}) = 0$ is the purely mathematical restriction we refer to in point (III), Section 3.2.1.

Let us point out that the condition "$\mathbf{p} \in \mathcal{Z}$" is sufficient for "$K(\mathbf{p}, L)$ is zeroless" but not necessary, as the following simple example shows. Let

$$K(\mathbf{p}, L) = \begin{pmatrix} L - p_1 & 0 \\ 0 & L - p_2 \\ L - p_3 & L - p_3 \end{pmatrix},$$

where $(p_1\ p_2\ p_3) \in \mathcal{P}$, where $\mathcal{P}$ is an open subset of $\mathbb{R}^3$. We have $D_1(\mathbf{p}, L) = (L - p_1)(L - p_2)$, rows 1 and 2, $D_2(\mathbf{p}, L) = (L - p_1)(L - p_3)$, rows 1 and 3, $D_3(\mathbf{p}, L) = (L - p_2)(L - p_3)$, rows 2 and 3. We see that generically $\mathcal{R}(\mathbf{p}) = 0$, so that (W) holds, but generically $K(\mathbf{p}, L)$ is zeroless.

The example above suggests that the result in Proposition AD2 can be improved. However, we believe that Proposition AD2, as it stands, and our discussion of zerolessness in Sections 3.2.1 and 3.2.2 are sufficient to motivate Assumption 3.

A.2 More on cointegration in the singular case

The arguments in points (B) and (C), Section 3.2.2, can be easily generalized. Let X_{it} be $I(1)$, for all $i = 1, \dots, m+1$, $q < m$, $X_t = (X_{1t}\ X_{2t}\ \cdots\ X_{mt})'$, $\tilde{X}_t = (X_{1t}\ X_{2t}\ \cdots\ X_{m+1,t})$ and let

$$(1 - L)\tilde{X}_t = \begin{pmatrix} (1 - L)X_t \\ (1 - L)X_{m+1,t} \end{pmatrix} = \begin{pmatrix} K(L) \\ k_{m+1}(L) \end{pmatrix} u_t = \tilde{K}(L)u_t. \tag{23}$$

Assume that the cointegration rank of X_t is $c = m - q + \kappa$ with $\kappa > 0$. Because rank $K(1) = q - \kappa < q$, it is possible that $\tilde{X}_t$ has no additional cointegration vector with respect to X_t, i.e. $k_{m+1}(1)$ can be independent of the rows of $K(1)$. In that case $c = \tilde{c} = m + 1 - q + \tilde{\kappa}$, so that $\tilde{\kappa} = \kappa - 1$:

Remark 1. *If $m > q$ and $\kappa > 0$ and we add to X_t the variable $X_{m+1,t}$, driven by u_t, and the cointegration rank stays the same, the value of κ decreases by one. This is a generalization of our argument in (B), Section 3.2.2.*

On the other hand, if $\kappa = 0$, so that rank $K(1) = q$, then $k_{m+1}(1)$ is a linear combination of the rows of $K(1)$, that is $\tilde{c} = c + 1$. Thus $\tilde{\kappa} = \kappa = 0$. Moreover, looking at (23), quite obviously,

Remark 2. *If $m > q$ and we add to X_t the variable $X_{m+1,t}$, driven by u_t, the IRFs of X_t do not change.*

What may happen is that $\tilde{K}(L)$ is zeroless whereas $K(L)$ is not, so that u_t may be obtained by a finite-lenght VAR of $\tilde{X}_t$.

Let us now replace X_{it} with $Y_{it} = X_{it} + \xi_{it}$, the ξ's being measurement errors. As a rule, the rank of Y_t is m and that of $\tilde{Y}_t$ is $m + 1$. Let

$$(1 - L)Y_t = C(L)w_t, \quad (1 - L)\tilde{Y}_t = \begin{pmatrix} \tilde{C}(L) & \tilde{c}_1(L) \\ \tilde{c}_2(L) & \tilde{c}_3(L) \end{pmatrix} \tilde{w}_t \tag{24}$$

be the IRFs that are consistently estimated by a SVAR for Y_t and $\tilde{Y}_t$, respectively, so that w_t and $\tilde{w}_t$ are fundamental for Y_t and $\tilde{Y}_t$, respectively. We suppose that w_t and $\tilde{w}_t$ have been identified consistently with the restrictions identifying u_t. For example, u_t, w_t and $\tilde{w}_t$ are identified by recursive schemes, as in Section 3.4.

Because the rank of Y_t and $\tilde{Y}_t$ are m and $m + 1$, respectively, $c = \kappa$, $\tilde{c} = \tilde{\kappa}$. As $\tilde{c} \geq c$, we have $\tilde{\kappa} \geq \kappa$, so that no zero of $C(L)$ at $z = 1$ can be removed by adding variables. Moreover, it is fairly easy to see that generically $\tilde{C}(L) \neq C(L)$ and $\tilde{w}_{jt} \neq w_{jt}$, for $j = 1, \ldots, m$, see e.g. Lippi (2021). Thus, we see that neither Remark 1 nor 2 hold for Y_t and $\tilde{Y}_t$.

A.3 Non-uniqueness of the VAR in the singular case

In Section 3.4 we consider the example with $m = 3$, $q = 1$, $B(L) = B_0 + B_1 L + B_2 L^2 + B_3 L^3$, where the 12 entries in the matrices B_j can vary independently of one another. If we take $p = 1$ in (12), we have $(I - A_1 L)(B_0 + B_1 L + B_2 L^2 + B_3 L^3) = B_0$, that is

$$A_1 B_0 = B_1, \quad A_1 B_1 = B_2, \quad A_1 B_2 = B_3, \quad A_1 B_3 = 0. \tag{25}$$

As the matrices B_j are 3×1, generically $B_0,\ B_1,\ B_2$ are independent and

$$B_3 = \alpha_0 B_0 + \alpha_1 B_1 + \alpha_2 B_2.$$

Using (25),

$$\begin{aligned} 0 = A_1 B_3 = A_1(\alpha_0 B_0 + \alpha_1 B_1 + \alpha_2 B_2) &= \alpha_0 B_1 + \alpha_1 B_2 + \alpha_2 B_3 \\ &= \alpha_2 \alpha_0 B_0 + (\alpha_0 + \alpha_2 \alpha_1) B_1 + (\alpha_1 + \alpha_2^2) B_2, \end{aligned}$$

which implies $\alpha_0 = \alpha_1 = \alpha_2 = 0$, i.e. $B_3 = 0$. Thus generically $p > 1$.

B Proof of Proposition 1

Let us denote by d the rank of Σ_0^Z. If $d < mp$, $\mathcal{A}$, as defined in (12), is not unique. We then partition Z_t (possibly after reordering) as $Z_t = \begin{pmatrix} \Omega_t' & S_t' \end{pmatrix}'$, where $\det \Sigma_0^\Omega \neq 0$. We have $S_t = N\Omega_t$ and $Z_t = M\Omega_t$, where $M = (I_d\ N)'$, so that we can re-write the above equation as

$$\chi_t = \alpha \Omega_{t-1} + v_t, \tag{26}$$

where $\alpha = \mathcal{A}M$ is unique.

Let $\hat{Z}_t = (\hat{\chi}_t'\ \hat{\chi}_{t-1}'\ \cdots\ \hat{\chi}_{t-p+1}')'$. While Σ_0^Z can be singular, Assumption A(b) ensures that $\det\left(\hat{\Sigma}_0^{\hat{Z}}\right) \neq 0$ a.s. Hence the coefficients of the sample projection corresponding to (12) are unique. Such projection is, see (13),

$$\hat{\chi}_t = \hat{A}_1 \hat{\chi}_{t-1} + \cdots + \hat{A}_p \hat{\chi}_{t-p} + \hat{v}_t = \hat{\mathcal{A}} \hat{Z}_{t-1} + \hat{v}_t.$$

In analogy with Ω_t and S_t, let $\hat{\Omega}_t$ be the vector including the first d entries of $\hat{Z}_t$ and $\hat{S}_t$ be the vector including the remaining $mp - d$ entries. Now, let us denote the OLS projection by $\hat{P}$ and consider the projection equation

$$\hat{S}_t = \hat{P}(\hat{S}_t | \hat{\Omega}_t) + \hat{\vartheta}_t = \hat{N}\hat{\Omega}_t + \hat{\vartheta}_t, \tag{27}$$

where $\hat{\Sigma}_0^{\hat{\vartheta}\hat{\Omega}} = 0$.

Lemma 1. *Under Assumptions A and B:*
(i) $||\hat{N} - N|| = O_p\left(r_{n,T}\right)$*;*
(ii) $||\hat{\vartheta}_t|| = O_p\left(r_{n,T}\right)$*.*
(iii) $||\hat{\Sigma}_0^{\hat{\vartheta}}||$ *is nonsingular with probability one.*

Proof. From (27) we get $\hat{N}\hat{\Sigma}_0^{\hat{\Omega}} = \hat{\Sigma}_0^{\hat{S}\hat{\Omega}}$. On the other hand, $\hat{\Sigma}_0^{S\Omega} = N\hat{\Sigma}_0^{\Omega}$, since $S_t = N\Omega_t$. Hence $(\hat{N} - N)\hat{\Sigma}_0^{\Omega} = \hat{\Sigma}_0^{\hat{S}\hat{\Omega}} - \hat{\Sigma}_0^{S\Omega}$. Letting $\hat{\phi}_t = \hat{S}_t - S_t$ and $\hat{\nu}_t = \hat{\Omega}_t - \Omega_t$ we get

$$\hat{N} - N = \left(\hat{\Sigma}_0^{S\hat{\nu}} + \hat{\Sigma}_0^{\hat{\phi}\Omega} + \hat{\Sigma}_0^{\hat{\phi}\hat{\nu}}\right)\left(\hat{\Sigma}_0^{\Omega}\right)^{-1}.$$

By Assumption A(a), $||\hat{\pi}_t||$ is $O_p(r_{n,T})$, so that both $||\hat{\phi}_t||$ and $||\hat{\nu}_t||$ are $O_p(r_{n,T})$. Therefore $||\hat{\Sigma}_0^{S\hat{\nu}} + \hat{\Sigma}_0^{\hat{\phi}\Omega} + \hat{\Sigma}_0^{\hat{\phi}\hat{\nu}}||$ is $O_p(r_{n,T})$. Moreover, the determinant of $\hat{\Sigma}_0^{\Omega}$ is bounded away from zero in probability by nonsingularity of Σ_0^{Ω} and Assumption B, so that $\left(\hat{\Sigma}_0^{\Omega}\right)^{-1}$ is $O_p(1)$. Result (i) follows. As for (ii), from (27) and the definition of $\hat{\phi}_t$ and $\hat{\nu}_t$ we get

$$\hat{\vartheta}_t = \hat{S}_t - \hat{N}\hat{\Omega}_t = S_t + \hat{\phi}_t - \hat{N}(\Omega_t + \hat{\nu}_t) = \hat{\phi}_t + (N - \hat{N})\Omega_t - \hat{N}\hat{\nu}_t. \tag{28}$$

By (i), $||\hat{N}||$ is $O_p(1)$. Because $||\Omega_t||$ is also $O_p(1)$, (ii) follows from (i) and the fact that both $||\hat{\phi}_t||$ and $||\hat{\nu}_t||$ are $O_p(r_{n,T})$. Coming to (iii), notice first that $\hat{\Sigma}_0^{\hat{\vartheta}} = \hat{\Sigma}_0^{\hat{S}} - \hat{N}\hat{\Sigma}_0^{\hat{\Omega}}\hat{N}' = \hat{\Sigma}_0^{\hat{S}} - \hat{\Sigma}_0^{\hat{S}\hat{\Omega}}(\hat{\Sigma}_0^{\hat{\Omega}})^{-1}\hat{\Sigma}_0^{\hat{\Omega}\hat{S}}$, see equation (27). Since $\hat{Z}_t' = (\hat{\Omega}_t' \;\; \hat{S}_t')'$, by Shur's formula we have $\det\hat{\Sigma}_0^{\hat{Z}} = \det\hat{\Sigma}_0^{\hat{\Omega}}\det\hat{\Sigma}_0^{\hat{\vartheta}}$. But $\hat{\Sigma}_0^{\hat{Z}}$ is nonsingular a.s. by Assumption A(b). Q.E.D.

Lemma 1 (iii) implies that the population covariance matrix of ϑ_t, $\Sigma_0^{\hat{\vartheta}}$, is nonsingular, so that we have the Choleski factorization $HH' = \Sigma_0^{\hat{\vartheta}}$. Now, let $\tilde{\vartheta}_t = H^{-1}\hat{\vartheta}_t$, so that $\Sigma_0^{\tilde{\vartheta}} = I$. Note that

$$||\hat{\Sigma}_0^{\tilde{\vartheta}}|| = O_p(1), \quad ||(\hat{\Sigma}_0^{\tilde{\vartheta}})^{-1}|| = O_p(1). \tag{29}$$

The vectors $\hat{\Omega}_t$ and $\tilde{\vartheta}_t$ are sample orthogonal, i.e. $\hat{\Sigma}_0^{\hat{\Omega}\tilde{\vartheta}} = 0$. Moreover, their entries span the same linear space as the entries of $\hat{Z}_t$. Hence we can decompose the sample projection $\hat{P}(\hat{\chi}_t|\hat{Z}_{t-1})$ into the sum of the projections $\hat{P}(\hat{\chi}_t|\hat{\Omega}_{t-1}) = \hat{\alpha}\hat{\Omega}_{t-1}$

and $\hat{P}(\hat{\chi}_t|\tilde{\vartheta}_{t-1}) = \hat{\beta}\tilde{\vartheta}_{t-1}$, i.e.

$$\hat{\chi}_t = \hat{\mathcal{A}}\hat{Z}_{t-1} + \hat{v}_t = \hat{\alpha}\hat{\Omega}_{t-1} + \hat{\beta}\tilde{\vartheta}_{t-1} + \hat{v}_t, \tag{30}$$

where $\hat{\Sigma}_1^{\hat{v}\hat{\Omega}} = 0$ and $\hat{\Sigma}_1^{\hat{v}\tilde{\vartheta}} = 0$, so that $\hat{\alpha}\hat{\Sigma}_0^{\hat{\Omega}} = \hat{\Sigma}_1^{\hat{\chi}\hat{\Omega}}$ and $\hat{\beta}\hat{\Sigma}_0^{\tilde{\vartheta}} = \hat{\Sigma}_1^{\hat{\chi}\tilde{\vartheta}}$. This equation is the sample analogue of the population equation (26). Subtracting (26) from (30) we get

$$\hat{\chi}_t - \chi_t = \hat{\pi}_t = (\alpha\Omega_{t-1} - \hat{\alpha}\hat{\Omega}_{t-1}) + \hat{\beta}\tilde{\vartheta}_{t-1} + (v_t - \hat{v}_t). \tag{31}$$

Since the left-hand side is $O_p(r_{n,T})$ by Assumption A(a), in order to prove Proposition 1, that is $\|\hat{v}_t - v_t\| = O_p\left(\max(r_{n,T}, 1/\sqrt{T})\right)$, it is sufficient to show that the norms of the first two terms on the right side are $O_p\left(\max(r_{n,T}, 1/\sqrt{T})\right)$.

Lemma 2.
(i) $||\hat{\alpha} - \alpha|| = O_p\left(\max(r_{n,T}, 1/\sqrt{T})\right)$*;*
(ii) $||\hat{\alpha}\hat{\Omega}_{t-1} - \alpha\Omega_{t-1}|| = O_p\left(\max(r_{n,T}, 1/\sqrt{T})\right)$
(iii) $||\hat{\Sigma}_1^{v\tilde{\vartheta}}|| = O_p(1/\sqrt{T})$*;*
(iv) $||\hat{\beta}\tilde{\vartheta}_{t-1}|| = O_p\left(\max(r_{n,T}, 1/\sqrt{T})\right)$.

Proof. (i). We have

$$\hat{\alpha} - \alpha = \left[\left(\hat{\alpha}\hat{\Sigma}_0^{\hat{\Omega}} - \alpha\Sigma_0^{\Omega}\right) - \hat{\alpha}\left(\hat{\Sigma}_0^{\hat{\Omega}} - \Sigma_0^{\Omega}\right)\right]\left(\Sigma_0^{\Omega}\right)^{-1}. \tag{32}$$

Now consider the first term of the difference in square brackets. Using (26) and (30), we get $\hat{\alpha}\hat{\Sigma}_0^{\hat{\Omega}} - \alpha\Sigma_0^{\Omega} = \hat{\Sigma}_1^{\hat{\chi}\hat{\Omega}} - \Sigma_1^{\chi\Omega} = \left(\hat{\Sigma}_1^{\chi\Omega} - \Sigma_1^{\chi\Omega}\right) + \hat{\Sigma}_1^{\hat{\pi}\Omega} + \hat{\Sigma}_1^{\hat{\chi}\hat{\nu}}$. Assumption B(a) implies that $||\hat{\Sigma}_1^{\chi\Omega} - \Sigma_1^{\chi\Omega}|| = O_p(1/\sqrt{T})$, while $||\hat{\Sigma}_1^{\hat{\pi}\Omega} + \hat{\Sigma}_1^{\hat{\chi}\hat{\nu}}||$ is $O_p(r_{n,T})$ by Assumption A(a). Turning to the second term, we have $\hat{\Sigma}_0^{\hat{\Omega}} - \Sigma_0^{\Omega} = \left(\hat{\Sigma}_0^{\Omega} - \Sigma_0^{\Omega}\right) + \hat{\Sigma}_0^{\hat{\nu}\Omega} + \hat{\Sigma}_0^{\hat{\Omega}\hat{\nu}}$. Assumption B(a) implies that $||\hat{\Sigma}_0^{\Omega} - \Sigma_0^{\Omega}|| = O_p(1/\sqrt{T})$, while $||\hat{\Sigma}_0^{\hat{\nu}\Omega} + \hat{\Sigma}_0^{\hat{\Omega}\hat{\nu}}||$ is $O_p(r_{n,T})$ by Assumption A(a). Since $||\hat{\alpha}||$ is $O_p(1)$, the norm of the factor in square brackets of (32) is $O_p\left(\max(r_{n,T}, 1/\sqrt{T})\right)$. Since $||(\Sigma_0^{\Omega})^{-1}|| = O(1)$, (i) follows.

(ii). We have $\hat{\alpha}\hat{\Omega}_{t-1} - \alpha\Omega_{t-1} = \hat{\alpha}\hat{\nu}_{t-1} + (\hat{\alpha} - \alpha)\Omega_{t-1}$. As $||\hat{\alpha}||$ is $O_p(1)$, by Assumption A(a), the norm of the first term is $O_p\left(r_{n,T}\right)$. Moreover, by result (i) the norm of the second term is $O_p\left(\max(r_{n,T}, 1/\sqrt{T})\right)$ so that (ii) is proven.

(iii). Notice first that the entries of v_t are linear combinations of the entries of u_t, see equation (11). But u_t is independent of $\hat{Z}_{t-1}$ and therefore $\hat{\vartheta}_{t-1}$ by Assumption 1. Now, let us consider the h-th row of the matrix $\hat{\Sigma}_1^{v\hat{\vartheta}}$, i.e. $\hat{\Sigma}_1^{v_h\hat{\vartheta}} = \sum_{t=2}^{T} v_{ht}\hat{\vartheta}'_{t-1}/(T-1)$. Let Ψ_1^h be its population covariance matrix. We have

$$\Psi_1^h = \sum_{t=2}^{T}\sum_{\tau=2}^{T} \mathrm{E}(v_{ht}\hat{\vartheta}_{t-1}\hat{\vartheta}'_{\tau-1}v_{h\tau})/(T-1)^2.$$

Independence of v_{ht} and $\hat{\vartheta}_{t-1}$ implies that

$$\Psi_1^h = \sum_{t=2}^{T}\sum_{\tau=2}^{T} \mathrm{E}(v_{ht}v_{h\tau})\mathrm{E}(\hat{\vartheta}_{t-1}\hat{\vartheta}'_{\tau-1})/(T-1)^2.$$

But $\mathrm{E}(v_{ht}v_{h\tau}) = 0$ for $t \neq \tau$, because of serial independence of u_t (see Assumption 1), so that $\Psi_1^h = \sum_{t=2}^{T} \mathrm{E}(v_{ht})^2\mathrm{E}(\hat{\vartheta}_{t-1}\hat{\vartheta}'_{t-1})/(T-1)^2$. Moreover, covariance co-stationarity of v_{ht} and $\hat{\vartheta}_t$ implies that $\Psi_1^h = \Sigma_0^{\hat{\vartheta}}\ \mathrm{E}(v_{ht}^2)/(T-1)$. Since $\tilde{\vartheta}_t = H^{-1}\hat{\vartheta}_t$, then $\hat{\Sigma}_1^{v_h\tilde{\vartheta}} = H^{-1}\hat{\Sigma}_1^{v_h\hat{\vartheta}}$. Moreover, the population covariance matrix of $\tilde{\vartheta}_t$ is $H^{-1}\Psi_1^h(H^{-1})' = I\ \mathrm{E}(v_{ht}^2)/(T-1)$, so that $||\hat{\Sigma}_1^{v_h\tilde{\vartheta}}||$ is $O_p(1/\sqrt{T})$ for all h. This establishes (iii).
(iv). We have $\hat{\beta}\hat{\Sigma}_0^{\tilde{\vartheta}} = \hat{\Sigma}_1^{\hat{\chi}\tilde{\vartheta}} = \hat{\Sigma}_1^{\chi\tilde{\vartheta}} + \hat{\Sigma}_1^{\hat{\pi}\tilde{\vartheta}} = \alpha\hat{\Sigma}_0^{\Omega\tilde{\vartheta}} + \hat{\Sigma}_1^{v\tilde{\vartheta}} + \hat{\Sigma}_1^{\hat{\pi}\tilde{\vartheta}}$. But $\hat{\Sigma}_0^{\Omega\tilde{\vartheta}} = \hat{\Sigma}_0^{\hat{\Omega}\tilde{\vartheta}} - \hat{\Sigma}_0^{\hat{\nu}\tilde{\vartheta}} = -\hat{\Sigma}_0^{\hat{\nu}\tilde{\vartheta}}$. Hence $\hat{\beta}\hat{\Sigma}_0^{\tilde{\vartheta}} = -\alpha\hat{\Sigma}_0^{\hat{\nu}\tilde{\vartheta}} + \hat{\Sigma}_1^{v\tilde{\vartheta}} + \hat{\Sigma}_1^{\hat{\pi}\tilde{\vartheta}}$. The norms of both the first and the third term are $O_p(r_{n,T})$ by Assumption A(a). The norm of the second term is $O_p(1/\sqrt{T})$ by (iii), hence $||\hat{\beta}\hat{\Sigma}_0^{\tilde{\vartheta}}|| = O_p\left(\max(r_{n,T}, 1/\sqrt{T})\right)$. But $||(\hat{\Sigma}_0^{\tilde{\vartheta}})^{-1}||$ is $O_p(1)$, see (29), so that $||\hat{\beta}|| \leq ||\hat{\beta}\hat{\Sigma}_0^{\tilde{\vartheta}}||\ ||(\hat{\Sigma}_0^{\tilde{\vartheta}})^{-1}||$ is $O_p\left(\max(r_{n,T}, 1/\sqrt{T})\right)$. Since $\tilde{\vartheta}_t$ is $O_p(1)$, (iv) is proven. Q.E.D.

Proposition 1 follows from equation (31), Lemma 2 (ii) and Lemma 2 (iv).

C Proof of Proposition 2

Let $\hat{\Sigma}$ and Σ be the upper-left $q \times q$ sub-matrices of $\hat{\Sigma}_0^{\hat{v}}$ and Σ_0^{v}, respectively.

Lemma 3. *We have:*
(i) $||\hat{\Sigma} - \Sigma|| = O_p\left(\max(r_{n,T}, 1/\sqrt{T})\right)$;

(ii) $||\hat{Q} - Q|| = O_p\left(\max(r_{n,T}, 1/\sqrt{T})\right)$;
(iii) $||\hat{Q}^{-1} - Q^{-1}|| = O_p\left(\max(r_{n,T}, 1/\sqrt{T})\right)$,
where Q *is defined in Assumption* $1'$.

Proof. Let $\hat{\psi}_t = \hat{v}_t - v_t$. We have $\hat{\Sigma}_0^{\hat{v}} - \Sigma_0^v = \hat{\Sigma}_0^{\hat{\psi}v} + \hat{\Sigma}_0^{v\hat{\psi}} + \hat{\Sigma}_0^{\hat{\psi}\hat{\psi}} + (\hat{\Sigma}_0^v - \Sigma_0^v)$. The norm of the first three terms on the right-hand side is $O_p\left(\max(r_{n,T}, 1/\sqrt{T})\right)$, since so is $||\hat{\psi}_t||$ by Proposition 1. The norm of the term in brackets is $O_p(1/\sqrt{T})$ by Assumption B(a). Hence $||\hat{\Sigma}_0^{\hat{v}} - \Sigma_0^v|| = O_p\left(\max(r_{n,T}, 1/\sqrt{T})\right)$. This proves (i). As for (ii), let us recall that $\hat{C}$ is the lower triangular Cholesky matrix of $\hat{\Sigma}_0^{\hat{v}}$ and $\hat{Q}$ is the upper-left $q \times q$ sub-matrix of $\hat{C}$, so that $\hat{Q}$ is the lower triangular Cholesky matrix of $\hat{\Sigma}$. By the definition of Q, we have $\Sigma = QQ'$, so that Q is the lower triangular Cholesky matrix of Σ. Hence the entries of $\hat{Q}$ and Q are the same continuous functions of the entries of $\hat{\Sigma}$ and Σ, respectively. Result (ii) then follows from result (i). Coming to (iii), note that $||\hat{Q}^{-1} - Q^{-1}|| = ||\hat{Q}^{-1}(\hat{Q} - Q)Q^{-1}|| \leq ||\hat{Q}^{-1}||\ ||\hat{Q} - Q||\ ||Q^{-1}||$. Moreover, $\hat{Q}$ is bounded away from zero in probability since Q is non-singular. The result then follows from (ii). Q.E.D.

Proposition 2(a). $||\hat{u}_t - u_t|| = O_p\left(\max(r_{n,T}, 1/\sqrt{T})\right)$.

Proof. Let w_t and $\hat{w}_t$ be the vectors formed by the first q entries of v_t and $\hat{v}_t$, respectively. We have $\hat{u}_t = \hat{Q}^{-1}\hat{w}_t$ and $u_t = Q^{-1}w_t$. Hence $\hat{u}_t - u_t = \hat{Q}^{-1}(\hat{w}_t - w_t) + (\hat{Q}^{-1} - Q^{-1})w_t$. The norm of the first term is $O_p\left(\max(r_{n,T}, 1/\sqrt{T})\right)$ by Proposition 1 and the fact that $||\hat{Q}^{-1}||$ is $O_p(1)$. Finally, the norm of the second term is also $O_p\left(\max(r_{n,T}, 1/\sqrt{T})\right)$ by Lemma 3 (iii). Q.E.D.

Lemma 4. *The following results hold:*
(i) $||\hat{B}_0 - B_0|| = O_p\left(\max(r_{n,T}, 1/\sqrt{T})\right)$
(ii) $||\hat{S}|| = O_p\left(\max(r_{n,T}, 1/\sqrt{T})\right)$,
where $\hat{S}$ *is defined in Section 3.4, below Assumption* $1'$.

Proof. Let us begin by showing that $||\hat{R} - R|| = O_p\left(\max(r_{n,T}, 1/\sqrt{T})\right)$, where R is the population counterpart of $\hat{R}$, as defined below Assumption $1'$. Since $\hat{C}\hat{C}' = \hat{\Sigma}_0^{\hat{v}}$, we have $\hat{R}\hat{Q}' = \hat{\Phi}$, $\hat{\Phi}$ being the lower-left $(m-q) \times q$ sub-matrix of $\hat{C}$. Hence $\hat{R} = \hat{\Phi}(\hat{Q}')^{-1}$. Similarly $R = \Phi(Q')^{-1}$. Hence $\hat{R} - R = \hat{\Phi}\left((\hat{Q}')^{-1} - (\hat{Q}')^{-1}\right) + (\hat{\Phi} - \Phi)(Q')^{-1}$. The norm of first term is $O_p\left(\max(r_{n,T}, 1/\sqrt{T})\right)$ by Lemma 3 (iii), while in the proof of Lemma 3 we have shown that $||\hat{\Sigma}_0^{\hat{v}} - \Sigma_0^v|| = O_p\left(\max(r_{n,T}, 1/\sqrt{T})\right)$, so that $||\hat{\Phi} - \Phi|| = O_p\left(\max(r_{n,T}, 1/\sqrt{T})\right)$. Result (i) follows from Lemma 3 (ii). As for (ii), let $\hat{\epsilon}_t$ be the sub-vector of $\hat{C}^{-1}\hat{v}_t$ including the last $m-q$ entries, so that $\hat{C}^{-1}\hat{v}_t = (\hat{u}_t' \;\; \hat{\epsilon}_t')'$. We have $v_t = B_0 u_t$ and $\hat{v}_t = \hat{B}_0\hat{u}_t + \hat{S}\hat{\epsilon}_t$. Hence $\hat{v}_t - v_t = (\hat{B}_0\hat{u}_t - B_0 u_t) + \hat{S}\hat{\epsilon}_t$. The norm of the left side is $O_p\left(\max(r_{n,T}, 1/\sqrt{T})\right)$ by Proposition 1, while the norm of the first term on the right side is $O_p\left(\max(r_{n,T}, 1/\sqrt{T})\right)$ by Proposition 2(a) and result (i). Hence $||\hat{S}\hat{\epsilon}_t|| = O_p\left(\max(r_{n,T}, 1/\sqrt{T})\right)$. But $||\hat{\epsilon}_t||$ is bounded away form 0 in probability by construction, since its sample covariance is the identity matrix. The result follows. Q.E.D.

To prove Proposition 2(b) we introduce the companion form of our empirical VAR, i.e.

$$\hat{Z}_t = \hat{D}\hat{Z}_{t-1} + \hat{\theta}_t, \tag{33}$$

where

$$\hat{D} = \begin{pmatrix} \hat{A}_1 & \hat{A}_2 & \cdots & \hat{A}_{p-1} & \hat{A}_p \\ I_m & 0_m & \cdots & 0_m & 0_m \\ 0_m & I_m & \cdots & 0_m & 0_m \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0_m & 0_m & \cdots & I_m & 0_m \end{pmatrix}, \qquad \hat{\theta} = \begin{pmatrix} \hat{v}_t \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

From (33) k, by recursion we get

$$\hat{Z}_t = \hat{D}^{k+1}\hat{Z}_{t-k-1} + \sum_{j=0}^{k} \hat{D}^j\hat{\theta}_{t-j}, \tag{34}$$

for any $k \geq 0$. By taking the first m rows of (34) we get

$$\hat{\chi}_t = \hat{G}_{k+1}\hat{Z}_{t-k-1} + \sum_{j=0}^{k}\hat{V}_j\hat{v}_{t-j} = \hat{G}_{k+1}\hat{Z}_{t-k-1} + \sum_{j=0}^{k}\hat{V}_j\hat{B}_0\hat{u}_{t-j} + \sum_{j=0}^{k}\hat{V}_j\begin{pmatrix} 0 \\ \hat{S}\end{pmatrix}\hat{\epsilon}_{t-j}, \tag{35}$$

where $\hat{G}_k$ is the matrix formed by the first m rows of $\hat{D}^k$ and V_j is the $m\times m$ upper-left sub-matrix of $\hat{D}^j$. Notice that $\hat{G}_1 = \hat{\mathcal{A}}$, $\hat{V}_0 = I_m$ and $\hat{V}_1 = \hat{A}_1$. Notice also that $\hat{V}_j$, $j = 0, \ldots, k$ is the j-th matrix coefficient of $\hat{A}(L)^{-1}$, so that $\hat{B}_j = \hat{V}_j\hat{B}_0$.

Finally, from (35) we get

$$\hat{G}_k \hat{Z}_{t-k} = \hat{G}_{k+1} \hat{Z}_{t-k-1} + \hat{B}_k \hat{u}_{t-k} + \hat{V}_k \begin{pmatrix} 0 \\ \hat{S} \end{pmatrix} \hat{\epsilon}_{t-k}, \tag{36}$$

which, letting $\hat{G}_0 = (I_m \quad 0)$, holds for any $k \geq 0$ and for $k = 0$ reduces to $\hat{\chi}_t = \hat{\mathcal{A}} \hat{Z}_{t-1} + \hat{v}_t$.

Similarly, from the population VAR (12) we get

$$\chi_t = G_{k+1} Z_{t-k-1} + \sum_{j=0}^{k} V_j v_{t-j} = G_{k+1} Z_{t-k-1} + \sum_{j=0}^{k} V_j B_0 u_{t-j} \tag{37}$$

where $G_1 = \mathcal{A}$, $V_0 = I_m$ and $V_1 = A_1$. We have already observed in the main text that $\mathcal{A}$ is not necessarily unique, so that G_{k+1} and V_j, $j = 1, \dots, k$, are not necessarily unique. However, post-multiplying by u'_{t-k} and taking expected values we get $\Sigma_k^{\chi u} = V_k B_0$, so that $V_k B_0$ is unique and equals B_k for any $k \geq 0$. Hence $G_{k+1} Z_{t-k-1}$ is also unique for any k. From (37) we get

$$G_k Z_{t-k} = G_{k+1} Z_{t-k-1} + B_k u_{t-k}. \tag{38}$$

Lemma 5. For any $k \geq 0$,
(i) $||\hat{G}_k \hat{Z}_{t-k} - G_k Z_{t-k}|| = O_p\left(\max(r_{n,T}, 1/\sqrt{T})\right)$;
(ii) $\left|\left|\hat{V}_k \begin{pmatrix} 0 \\ \hat{S} \end{pmatrix}\right|\right| = O_p\left(\max(r_{n,T}, 1/\sqrt{T})\right)$;
(iii) $||\hat{B}_k - B_k|| = O_p\left(\max(r_{n,T}, 1/\sqrt{T})\right)$.

Proof. We proceed by induction. For $k = 0$, $||\hat{G}_k \hat{Z}_{t-k} - G_k Z_{t-k}||$ reduces to $||\hat{\chi}_t - \chi_t||$, which is $O_p(r_{n,T})$ by Assumption A(a); moreover, (ii) holds by Lemma 4(ii) and (iii) holds by Lemma 4(i). Hence (i)-(iii) are true for $k = 0$. Let us now show that, if (i)-(iii) are true for $k = \bar{k}$, they are true for $k = \bar{k} + 1$. Subtracting

(38) from (36) we get

$$\hat{G}_{\bar{k}}\hat{Z}_{t-\bar{k}}-G_{\bar{k}}Z_{t-\bar{k}} = (\hat{G}_{\bar{k}+1}\hat{Z}_{t-(\bar{k}+1)}-G_{\bar{k}+1}Z_{t-(\bar{k}+1)})+(\hat{B}_{\bar{k}}\hat{u}_{t-\bar{k}}-B_{\bar{k}}u_{t-\bar{k}})+\hat{V}_{\bar{k}}\begin{pmatrix}0\\ \hat{S}\end{pmatrix}\hat{\epsilon}_{t-\bar{k}}. \tag{39}$$

By the inductive assumption the term on the left side, the second and third terms on the right are $O_p\left(\max(r_{n,T}, 1/\sqrt{T})\right)$, so that the same holds for the first term on the right and (i) is true for $k = \bar{k}+1$. Next, let us replace $\bar{k}$ with $\bar{k}+1$ in (39) and take sample covariances with $\hat{\epsilon}_{t-(\bar{k}+1)}$. Using sample orthogonality of $\hat{\epsilon}_{t-(\bar{k}+1)}$ with both $\hat{u}_{t-(\bar{k}+1)}$ and $\hat{Z}_{t-(\bar{k}+2)}$ and sample orthonormality of $\hat{\epsilon}_t$ we get

$$\hat{G}_{\bar{k}+1}\hat{\Sigma}_0^{\hat{Z}\hat{\epsilon}} - G_{\bar{k}+1}\hat{\Sigma}_0^{Z\hat{\epsilon}} = -G_{\bar{k}+2}\hat{\Sigma}_{-1}^{Z\hat{\epsilon}} - B_{\bar{k}+1}\hat{\Sigma}_0^{u\hat{\epsilon}} + \hat{V}_{\bar{k}+1}\begin{pmatrix}0\\ \hat{S}\end{pmatrix}.$$

The norm of the left side is $O_p\left(\max(r_{n,T}, 1/\sqrt{T})\right)$ because, as observed above, (i) holds for $k = \bar{k}+1$. Using $Z_t = M\Omega_t$, see (26), $\hat{\nu}_t = \hat{\Omega}_t - \Omega_t$ and othogonality of $\hat{\epsilon}_{t-(\bar{k}+1)}$ with $\hat{\Omega}_{t-(k+2)}$, we see that the first term on the right side is equal to $G_{\bar{k}+2}M\hat{\Sigma}_{-1}^{\hat{\nu}\hat{\epsilon}}$, whose norm is $O_p(r_{n,T})$ since so is the norm of $\hat{\nu}_t$ by Assumption A(a) and the norm of $G_{\bar{k}+2}M$ is $O(1)$. Letting $\hat{\gamma}_t = \hat{u}_t - u_t$, using the orthogonality of $\hat{\epsilon}_{t-(\bar{k}+1)}$ with $\hat{u}_{t-(\bar{k}+1)}$, the second term is equal to $B_{\bar{k}+1}\hat{\Sigma}_0^{\hat{\gamma}\hat{\epsilon}}$, whose norm is $O_p\left(\max(r_{n,T}, 1/\sqrt{T})\right)$ since so is the norm of $\hat{\gamma}_t$ by Proposition 2(a). Hence the norm of the third term is also $O_p\left(\max(r_{n,T}, 1/\sqrt{T})\right)$, which proves that (ii) is true for $k = \bar{k}+1$. Lastly, let us again replace $\bar{k}$ with $\bar{k}+1$ in (39) and take sample covariances with $\hat{u}_{t-(\bar{k}+1)}$. Using sample orthogonality with both $\hat{\varepsilon}_{t-(\bar{k}+1)}$ and $\hat{Z}_{t-(\bar{k}+2)}$ and sample orthonormality of $\hat{u}_t$ we get

$$\hat{G}_{\bar{k}+1}\hat{\Sigma}_0^{\hat{Z}\hat{u}} - G_{\bar{k}+1}\hat{\Sigma}_0^{Z\hat{u}} = -G_{\bar{k}+2}\hat{\Sigma}_{-1}^{Z\hat{u}} + (\hat{B}_{\bar{k}+1} - B_{\bar{k}+1}) + B_{\bar{k}+1}\Sigma_0^{u\hat{\gamma}}.$$

The norm of the left side is $O_p\left(\max(r_{n,T}, 1/\sqrt{T})\right)$ since (i) holds for $k = \bar{k}+1$. The norm of the first term on the right side is $O_p\left(\max(r_{n,T}, 1/\sqrt{T})\right)$ for the same argument used above. The norm of the third term is $O_p\left(\max(r_{n,T}, 1/\sqrt{T})\right)$ since so is the norm of $\hat{\gamma}_t$ by Proposition 2(a). Hence (iii) holds for $k = \bar{k}+1$. In conclusion (i), (ii) and (iii) are true for any $k \geq 0$. Q.E.D.

Proposition 2(b) is Lemma 5(iii).

D Proof of Proposition 3

The proof below partly follows the proof of Proposition P in Forni et al. (2009), Appendix. However, here we need the consistency of $\hat{\chi}_{it}$, which is not needed in that paper. Thus, after some common lemmas, the proof here takes a different route.

To begin, let us introduce some additional notation and recall a standard result. If A is a symmetric matrix, we denote by $\mu_j(A)$ the j-th eigenvalue of A in decreasing order. Given a matrix B, we denote as above by $\|B\|$ the spectral norm of B, thus $\|B\| = \sqrt{\mu_1(BB')}$, which is the euclidean norm if B is a row matrix. We will make use of the Weyl inequality: letting A and B be two $s \times s$ symmetric matrices,

$$|\mu_j(A+B) - \mu_j(A)| \leq \sqrt{\mu_1(B^2)} = \|B\|, \quad j = 1, \dots, s. \tag{40}$$

Lemma 6. *(Consistency of the covariance matrices). Denoting by $\mathcal{I}_m$ the $n \times m$ matrix having the identity matrix I_m in the first m rows and 0 elsewhere, for any k and any (fixed) m we have:*

(i) $\frac{1}{n}\|\hat{\Gamma}^x_k - \Gamma^x_k\| = O_p\left(\frac{1}{\sqrt{T}}\right)$;

(ii) $\frac{1}{\sqrt{n}}\|\mathcal{I}'_m\left(\hat{\Gamma}^x_k - \Gamma^x_k\right)\| = O_p\left(\frac{1}{\sqrt{T}}\right)$;

(iii) $\frac{1}{\sqrt{n}}\|\mathcal{I}'_m\left(\hat{\Gamma}^\chi_k - \Gamma^\chi_k\right)\| = O_p\left(\frac{1}{\sqrt{T}}\right)$;

(iv) $\frac{1}{\sqrt{n}}\|\mathcal{I}'_m\hat{\Gamma}^{\chi\xi}_k\| = O_p\left(\frac{1}{\sqrt{T}}\right)$;

(v) $\|\mathcal{I}'_m(\hat{\Gamma}^\chi_k - \Gamma^\chi_k)\mathcal{I}_m\| = \|\hat{\Sigma}^\chi_k - \Sigma^\chi_k\| = O_p\left(\frac{1}{\sqrt{T}}\right)$;

(vi) $\frac{1}{n}\|\hat{\Gamma}^x_k - \Gamma^\chi_k\| = O_p\left(\max\left(\frac{1}{n}, \frac{1}{\sqrt{T}}\right)\right)$;

(vii) $\frac{1}{\sqrt{n}}\|\mathcal{I}'_m\left(\hat{\Gamma}^x_k - \Gamma^\chi_k\right)\| = O_p\left(\max\left(\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{T}}\right)\right)$.

Proof. We have

$$\mu_1\left((\hat{\Gamma}^x_k-\Gamma^x_k)(\hat{\Gamma}^x_k-\Gamma^x_k)'\right)\le \text{trace}\left((\hat{\Gamma}^x_k-\Gamma^x_k)(\hat{\Gamma}^x_k-\Gamma^x_k)'\right)=\sum_{i=1}^n\sum_{j=1}^n(\hat{\gamma}^x_{k,ij}-\gamma^x_{k,ij})^2.$$

By Assumption 7(a), we have $\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n \mathrm{E}(\hat{\gamma}^x_{k,ij}-\gamma^x_{k,ij})^2<\frac{\varrho}{T}$ for all positive integers T, so that $\frac{1}{n^2}\|\hat{\Gamma}^x_k-\Gamma^x_k\|^2=O_p\left(\frac{1}{T}\right)$ by Markov inequality. Result (i) follows. Coming to (ii), we see that, by the same argument, the squared norm of $\mathcal{I}'_m\left(\hat{\Gamma}^x_k-\Gamma^x_k\right)$ is bounded above by $\sum_{i=1}^m\sum_{j=1}^n(\hat{\gamma}^x_{k,ij}-\gamma^x_{k,ij})^2$, which is $O_p(n/T)$. Statement (ii) follows. Results (iii) and (iv) are obtained in the same way, by using Assumptions 7(b) and 7(c), respectively. As for (v), the same argument shows that the squared norm of $\mathcal{I}'_m(\hat{\Gamma}^\chi_k-\Gamma^\chi_k)\mathcal{I}_m$ is bounded above by $\sum_{i=1}^m\sum_{j=1}^m(\hat{\gamma}^\chi_{k,ij}-\gamma^\chi_{k,ij})^2$, which is $O_p(1/T)$. The result follows. Let us now come to results (vi) and (vii). Orthogonality of χ_t and ξ_t at all leads and lags, Assumption 4(b), implies that $\Gamma^x_k=\Gamma^\chi_k+\Gamma^\xi_k$. Hence $\hat{\Gamma}^x_k-\Gamma^\chi_k=\hat{\Gamma}^x_k-\Gamma^x_k+\Gamma^\xi_k$, so that $\frac{1}{n}\|\hat{\Gamma}^x_k-\Gamma^\chi_k\|\le\frac{1}{n}\|\hat{\Gamma}^x_k-\Gamma^x_k\|+\frac{1}{n}\|\Gamma^\xi_k\|$. The first term on the right side is $O_p\left(\frac{1}{\sqrt{T}}\right)$ by result (i). The second is bounded by $\frac{1}{n}\mu^\xi_1$, which is $O\left(\frac{1}{n}\right)$ by Assumption 6(b). This proves (vi). Finally, statement (vii) follows from the same argument, with result (ii) in place of result (i), n in place of n^2 and $1/\sqrt{n}$ in place of $1/n$. Q.E.D.

Lemma 7. *(Consistency of the normalized eigenvalues). Let M^χ and $\hat{M}^x$ be the diagonal matrices having on the diagonal the eigenvalues $\mu^\chi_1,\dots,\mu^\chi_r$ and $\hat{\mu}^x_1,\dots,\hat{\mu}^x_r$, respectively, in decreasing order of magnitude. Then,*

(i) $\hat{\mu}^x_j/n-\mu^x_j/n=O_p\left(1/\sqrt{T}\right)$ for any j.

(ii) $\hat{\mu}^x_j/n-\mu^\chi_j/n=O_p\left(\max(1/n,1/\sqrt{T})\right)$ for any j.

(iii) $\|M^\chi/n\|=O(1)$; there exist $\overline{n}$ such that, for $n>\overline{n}$, M^χ/n is invertible and $||(M^\chi/n)^{-1}||=O(1)$.

(iv) For any $n\ge\bar{n}$ and $\eta>0$, there exists $\tau(\eta,n)$ such that, for $T\ge\tau(\eta,n)$, $\frac{\hat{M}^x}{n}$ is invertible with probability larger than $1-\eta$; moreover, if $\left(\frac{\hat{M}^x}{n}\right)^{-1}$ exists for $n=n^*$ and $T=T^*$, it exists for all $n>n^*$ and $T>T^*$.

(v) $\|\hat{M}^x/n\|$ and $\left|\left|\left(\hat{M}^x/n\right)^{-1}\right|\right|$ are $O_p(1)$.

Proof. Setting $A=\Gamma^x_0$, $B=\hat{\Gamma}^x_0-\Gamma^x_0$ and applying (40) we get $\frac{1}{n}|\hat{\mu}^x_j-\mu^x_j|\le n^{-1}\|\hat{\Gamma}^x_0-\Gamma^x_0\|$, which is $O_p\left(1/\sqrt{T}\right)$ by Lemma 6(i). This proves (i). Setting

$A = \Gamma_0^\chi$, $B = \hat{\Gamma}_0^x - \Gamma_0^\chi$ and applying again (40) we get $\frac{1}{n}|\hat{\mu}_j^x - \mu_j^\chi| \leq n^{-1}\|\hat{\Gamma}_0^x - \Gamma_0^\chi\|$, which is $O_p\left(\max(1/n, 1/\sqrt{T})\right)$ by Lemma 6(vi). This establishes (ii). As for (iii), by Assumption 6(a) there exists $\overline{n}$ such that, for $n \geq \bar{n}$, $\frac{\mu_r^\chi}{n} > \underline{c}_r > 0$, so that M^χ/n is invertible and $||(M^\chi/n)^{-1}|| < 1/\underline{c}_r$. Moreover, by the same assumption μ_1^χ/n is asymptotically bounded by $\overline{c}_1$. This proves (iii). As for (iv), by (40), $\mu_r^x > \mu_r^\chi$, since μ_1^ξ is positive. Hence, for some $\overline{n}$ and $n > \overline{n}$, μ_r^x/n is bounded below by $\underline{c}_r > 0$. It follows that $\det(\hat{M}^x/n)$ is bounded away from zero in probability as $T \to \infty$. The last part of statement (iv) follows from the fact that the rank of the observation matrix, and therefore that of $\hat{\Gamma}_0^x$, is non-decreasing in n and T. Turning to (v), boundedness in probability of $\|\frac{\hat{M}^x}{n}\|$ and $\left|\left|\left(\frac{\hat{M}^x}{n}\right)^{-1}\right|\right|$ follows from statements (ii) and (iii). This concludes the proof. Q.E.D.

Lemma 8. *Let W^χ be the $n \times r$ matrix having on column j, $j = 1, \dots, r$, the unit-norm eigenvector of Γ_0^χ corresponding to the eigenvalue μ_j^χ. We have*
(i) $\|\sqrt{n}\mathcal{I}_m' W^\chi\| = O(1)$.
(ii) $\|W^{\chi\prime}\hat{W}^x\frac{\hat{M}^x}{n} - \frac{M^\chi}{n}W^{\chi\prime}\hat{W}^x\| = O_p\left(\max(1/n, 1/\sqrt{T})\right)$.
(iii) $\|\hat{W}^{x\prime}W^\chi W^{\chi\prime}\hat{W}^x - I_r\| = O_p\left(\max(1/n, 1/\sqrt{T})\right)$.

Proof. Let us notice first that $\zeta = \left|\left|\mathcal{I}_m' W^\chi (M^\chi)^{1/2}\right|\right| = \|\mathcal{I}_m'\Gamma_0^\chi\mathcal{I}_m\|^{1/2} = \|\Sigma_0^\chi\|^{1/2}$ does not depend on n. We have

$$\|\sqrt{n}\mathcal{I}_m' W^\chi\| = \left|\left|\sqrt{n}\mathcal{I}_m' W^\chi \left(\frac{M^\chi}{n}\right)^{1/2}\left(\frac{M^\chi}{n}\right)^{-1/2}\right|\right| \leq \zeta \left|\left|\left(\frac{M^\chi}{n}\right)^{-1/2}\right|\right|,$$

which is $O_p(1)$ by Lemma 7(iii). Turning to (ii), we have $\|W^{\chi\prime}\hat{W}^x\frac{\hat{M}^x}{n} - \frac{M^\chi}{n}W^{\chi\prime}\hat{W}^x\| = \|\frac{1}{n}W^{\chi\prime}\left(\hat{\Gamma}_0^x - \Gamma_0^\chi\right)\hat{W}^x\| \leq \frac{1}{n}\|\hat{\Gamma}_0^x - \Gamma_0^\chi\|$. Statement (ii) then follows from Lemma 6(vi). To prove (iii), let

$a = \hat{W}^{x\prime}W^\chi W^{\chi\prime}\hat{W}^x = \hat{W}^{x\prime}W^\chi W^{\chi\prime}\hat{W}^x\frac{\hat{M}^x}{n}\left(\frac{\hat{M}^x}{n}\right)^{-1}$,
$b = \hat{W}^{x\prime}W^\chi\frac{M^\chi}{n}W^{\chi\prime}\hat{W}^x\left(\frac{\hat{M}^x}{n}\right)^{-1} = \frac{1}{n}\hat{W}^{x\prime}\Gamma_0^\chi\hat{W}^x\left(\frac{\hat{M}^x}{n}\right)^{-1}$,
$c = \frac{1}{n}\hat{W}^{x\prime}\hat{\Gamma}_0^x\hat{W}^x\left(\frac{\hat{M}^x}{n}\right)^{-1} = \frac{\hat{M}^x}{n}\left(\frac{\hat{M}^x}{n}\right)^{-1} = I_r$.

We have $\|a - c\| \leq \|a - b\| + \|b - c\|$. Both terms are $O_p\left(\max(1/n, 1/\sqrt{T})\right)$, the first by statement (ii) and Lemma 7(v), the second by Lemma 6(vi) and Lemma 7(v). Q.E.D

Lemma 9. *There exist diagonal* $r \times r$ *matrices* $\hat{\mathcal{J}}_r$*, depending on* n *and* T*, whose diagonal entries are equal to either* 1 *or* -1*, such that*
(i) $\|\hat{W}^{x\prime} W^{\chi} - \hat{\mathcal{J}}_r\| = O_p\left(\max\left(1/n, 1/\sqrt{T}\right)\right)$.
(ii) $\|\sqrt{n}\mathcal{I}_m' \hat{W}^x - \sqrt{n}\mathcal{I}_m' W^{\chi}\hat{\mathcal{J}}_r\| = O_p\left(\max\left(1/\sqrt{n}, 1/\sqrt{T}\right)\right)$.

Proof. The reason why we need the matrices $\hat{\mathcal{J}}_r$ is simply that the eigenvectors corresponding to distinct eigenvalues are only unique up to the sign. Let us denote by $\hat{w}_j^x$ and w_j^{χ} the j-th columns of $\hat{W}^x$ and W^{χ} respectively. By taking a single entry of the matrix on the left side of of Lemma 8(ii) we get

$$\frac{1}{n}\left(\hat{\mu}_j^x - \mu_i^{\chi}\right) w_j^{\chi\prime}\hat{w}_i^x = O_p\left(\max\left(\frac{1}{n}, \frac{1}{\sqrt{T}}\right)\right),$$

$i \le r$, $j \le r$. Now, for $j \neq i$, $\frac{1}{n}\left(\hat{\mu}_j^x - \mu_i^{\chi}\right)$ is bounded away from zero in probability, since μ_i^{χ}/n and μ_j^{χ}/n are asymptotically distinct by Assumption 6(a), while $\hat{\mu}_j^x/n$ tends to μ_j^{χ}/n in probability by Lemma 7(ii). Hence, by dividing both sides of the above equation by $n^{-1}(\hat{\mu}_j^x - \mu_i^{\chi})$, we see that the off-diagonal terms of $\hat{W}^{x\prime}W^{\chi}$ are $O_p\left(\max\left(\frac{1}{n}, \frac{1}{\sqrt{T}}\right)\right)$. Turning to the diagonal terms, let us first observe that $\hat{w}_i^{x\prime} W^{\chi} W^{\chi\prime}\hat{w}_i^x = 1 + O_p\left(\max\left(\frac{1}{n}, \frac{1}{\sqrt{T}}\right)\right)$ by Lemma 8(iii). Since

$$\hat{w}_i^{x\prime} W^{\chi} W^{\chi\prime}\hat{w}_i^x = (\hat{w}_i^{x\prime} w_i^{\chi})^2 + \sum_{\substack{j=1\\ j\neq i}}^{r}\left(\hat{w}_i^{x\prime} w_j^{\chi}\right)^2 = (\hat{w}_i^{x\prime} w_i^{\chi})^2 + O_p\left(\max\left(\frac{1}{n}, \frac{1}{\sqrt{T}}\right)\right),$$

then $1 - (\hat{w}_i^{x\prime} w_i^{\chi})^2 = O_p\left(\max\left(\frac{1}{n}, \frac{1}{\sqrt{T}}\right)\right)$. Hence $(1 - |\hat{w}_i^{x\prime} w_i^{\chi}|)(1 + |\hat{w}_i^{x\prime} w_i^{\chi}|) = O_p\left(\max\left(\frac{1}{n}, \frac{1}{\sqrt{T}}\right)\right)$, so that $1 - |\hat{w}_i^{x\prime} w_i^{\chi}| = O_p\left(\max\left(\frac{1}{n}, \frac{1}{\sqrt{T}}\right)\right)$. Statement (i) follows. Turning to (ii), set

$a = \sqrt{n}\mathcal{I}_m' W^{\chi}\hat{\mathcal{J}}_r$,
$b = \sqrt{n}\mathcal{I}_m' W^{\chi} W^{\chi\prime}\hat{W}^x = \sqrt{n}\mathcal{I}_m' W^{\chi} W^{\chi\prime}\hat{W}^x \frac{\hat{M}^x}{n}\left(\frac{\hat{M}^x}{n}\right)^{-1}$,
$c = \sqrt{n}\mathcal{I}_m' W^{\chi}\frac{M^{\chi}}{n} W^{\chi\prime}\hat{W}^x\left(\frac{\hat{M}^x}{n}\right)^{-1} = \frac{1}{\sqrt{n}}\mathcal{I}_m'\Gamma_0^{\chi}\hat{W}^x\left(\frac{\hat{M}^x}{n}\right)^{-1}$,
$d = \frac{1}{\sqrt{n}}\mathcal{I}_m'\hat{\Gamma}_0^x\hat{W}^x\left(\frac{\hat{M}^x}{n}\right)^{-1} = \sqrt{n}\mathcal{I}_m'\hat{W}^x$.

Notice that $\|\sqrt{n}\mathcal{I}_m' W^{\chi}\|$ is $O(1)$ by Lemma 8(i), so that we can apply result (i) to get $\|a - b\| = O_p\left(\max\left(\frac{1}{n}, \frac{1}{\sqrt{T}}\right)\right)$, and Lemmas 8(ii) and 7(v) to get $\|b - c\| = O_p\left(\max\left(\frac{1}{n}, \frac{1}{\sqrt{T}}\right)\right)$. Finally, Lemmas 6(vii) and 7(v) ensure that $\|c - d\| =$

$O_p\left(\max\left(\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{T}}\right)\right)$. This establishes (ii). Q.E.D.

Lemma 10. *(Consistency of the eigenvectors). We have*
(i) $\|\hat{W}^{x\prime} - \hat{\mathcal{J}}_r W^{\chi\prime}\| = O_p\left(\max\left(1/\sqrt{n}, 1/\sqrt{T}\right)\right)$.
(ii) $\|\sqrt{n}(\mathcal{I}_m'\hat{W}^x\hat{W}^{x\prime} - \mathcal{I}_m' W^\chi W^{\chi\prime})\| = O_p\left(\max\left(1/\sqrt{n}, 1/\sqrt{T}\right)\right)$.

Proof. Let as before $\hat{w}_j^x$ and w_j^χ be the j-th columns of $\hat{W}^x$ and W^χ, respectively, and let $\hat{\mathcal{J}}_r(j,j)$ be the j-th diagonal element of $\hat{\mathcal{J}}_r$, which is either 1 or -1. We have $||\hat{w}_j^{x\prime} - \hat{\mathcal{J}}_r(j,j)w_j^{\chi\prime}||^2 = 2 - \hat{w}_j^{x\prime} w_j^\chi \hat{\mathcal{J}}_r(j,j) - w_j^{\chi\prime}\hat{w}_j^x\hat{\mathcal{J}}_r(j,j)$. By Lemma 9(i), the last two terms are equal to $1 + O_p\left(\max\left(1/n, 1/\sqrt{T}\right)\right)$. Hence $||\hat{w}_j^{x\prime} - \hat{\mathcal{J}}_r(j,j)w_j^{\chi\prime}|| = O_p\left(\max\left(1/\sqrt{n}, 1/\sqrt{T}\right)\right)$. Statement (i) follows. As for (ii), set

$a = \sqrt{n}(\mathcal{I}_m'\hat{W}^x\hat{W}^{x\prime} - \mathcal{I}_m' W^\chi W^{\chi\prime})$;
$b = \sqrt{n}\mathcal{I}_m' W^\chi \hat{\mathcal{J}}_r(\hat{W}^{x\prime} - \hat{\mathcal{J}}_r W^{\chi\prime})$;
$c = \sqrt{n}(\mathcal{I}_m'\hat{W}^x - \mathcal{I}_m' W^\chi\hat{\mathcal{J}}_r)\hat{W}^{x\prime}$.

We have $a = b + c$, so that $\|a\| \leq \|b\| + \|c\|$. Let us consider firstly b and observe that $\|\sqrt{n}\mathcal{I}_m' W^\chi\|$ is $O(1)$ by Lemma 8(i). Hence $\|b\|$ is $O_p\left(\max\left(1/\sqrt{n}, 1/\sqrt{T}\right)\right)$ by result (i). Moreover, $\|c\|$ is $O_p\left(\max\left(1/\sqrt{n}, 1/\sqrt{T}\right)\right)$ by Lemma 9(ii). Q.E.D.

We are now ready to prove Proposition 3, reported here for convenience, with $r_{n,T} = \max(1/\sqrt{n}, 1/\sqrt{T}))$ and therefore $1/r_{n,T} = \min(\sqrt{n}, \sqrt{T})$.

Proposition 3. Properties of the principal component estimator.
(i) $\|\hat{\pi}_t\| = ||\hat{\chi}_t - \chi_t|| = O_p(\max(1/\sqrt{n}, 1/\sqrt{T}))$.
(ii) $||\hat{\Sigma}_k^\chi - \Sigma_k^\chi|| = O_p\left(1/\sqrt{T}\right)$*, for any k;*

Proof. Notice first that statement (ii) has already be proven, see Lemma 6(v). Regarding (i), let us firstly observe that, for n large enough, the principal components of $\boldsymbol{\chi}_{nt}$, i.e. the entries of $W^{\chi\prime}\boldsymbol{\chi}_{nt}$, form a basis for the linear space spanned by the factors F_{jt}, $j = 1, \ldots, r$. Hence the linear projection of χ_t onto the space spanned by such principal components is equal to χ_t and the residual is zero. This projection is $\mathcal{I}_m' W^\chi W^{\chi\prime}\boldsymbol{\chi}_{nt}$; hence $\chi_t = \boldsymbol{\chi}_{mt} = \mathcal{I}_m' W^\chi W^{\chi\prime}\boldsymbol{\chi}_{nt}$. On the other hand, our estimator of χ_t is defined as $\hat{\chi}_t = \mathcal{I}_m' \hat{W}^x \hat{W}^{x\prime} \boldsymbol{x}_{nt}$. Thus

$$\begin{aligned} \|\hat{\chi}_t - \chi_t\| &= \left\| \left(\mathcal{I}_m' \hat{W}^x \hat{W}^{x\prime} \boldsymbol{x}_{nt} - \mathcal{I}_m' W^\chi W^{\chi\prime} \boldsymbol{x}_{nt} \right) + \mathcal{I}_m' W^\chi W^{\chi\prime} \boldsymbol{\xi}_{nt} \right\| \\ &= \|a + b\| \le \|a\| + \|b\|. \end{aligned} \tag{41}$$

Regarding a, we have $\|a\| \le \|\sqrt{n}(\mathcal{I}_m' \hat{W}^x \hat{W}^{x\prime} - \mathcal{I}_m' W^\chi W^{\chi\prime})\| \; \|\boldsymbol{x}_{nt}/\sqrt{n}\|$. Now, $\|\boldsymbol{x}_{nt}/\sqrt{n}\|^2 = \sum_{i=1}^n x_{it}^2/n$ is $O_p(1)$, since its expected value is $(\mathrm{trace}\, \Gamma_0^x)/n = (\mathrm{trace}\, \Gamma_0^\chi)/n + (\mathrm{trace}\, \Gamma_0^\xi)/n \le \sum_{j=1}^r \mu_j/n + \mu_1^\xi$, which is bounded by Assumption 6. Hence a is $O_p\left(\max\left(1/\sqrt{n}, 1/\sqrt{T}\right)\right)$ by Lemma 10(ii). As for b, we have $\|\mathcal{I}_m' W^\chi W^{\chi\prime} \boldsymbol{\xi}_{nt}\| \le \|\mathcal{I}_m' W^\chi\| \; \|W^{\chi\prime} \boldsymbol{\xi}_{nt}\|$. The first factor is $O(1/\sqrt{n})$ by Lemma 8(i). The second is $O_p(1)$, since the norm of its covariance matrix, i.e. $W^{\chi\prime} \Gamma_0^\xi W^\chi$, is bounded by $\mu_1^\xi \le \ell$ (see Assumption 6(b)). Hence $\|b\| = O(1/\sqrt{n})$. Statement (i) follows. Q.E.D.

E Problems with $m = q + 1$: an example

The fact that $\hat{\chi}_t$ is not exactly singular may produce serious consequences: it is possible that u_t can be recovered using χ_t, but not using $\hat{\chi}_t$. To see this, consider the following example:

$$\begin{aligned} \chi_{1t} &= u_{t-1} \\ \chi_{2t} &= a_2 u_t + u_{t-1}. \end{aligned} \tag{42}$$

Here $B(L)$ is zeroless unless $a_2 = 0$. If $a_2 \neq 0$,

$$\frac{1}{a_2}(\chi_{2t} - \chi_{1t}) = u_t,$$

so that u_t lies in the econometrician's information set. Now suppose that $\hat{\chi}_{2t} = \chi_{2t} + \epsilon_t$, ϵ_t being a small residual idiosyncratic term. For simplicity, assume that $\hat{\chi}_{1t}$ is estimated without error, i.e. $\hat{\chi}_{1t} = \chi_{1t}$. The above expression becomes

$$\frac{1}{a_2}(\hat{\chi}_{2t} - \hat{\chi}_{1t}) = u_t + \frac{1}{a_2}\epsilon_t.$$

Now if $|a_2|$ is large, we can still get u_t with a good approximation; but as $|a_2|$ approaches 0 (i.e. the non-zeroless region), the error grows without bound. For instance, if u_t is unit variance and ϵ_t has standard deviation 0.01, with $a_2 = 1$ the error is negligible, but with $a_2 = 0.01$ the error has the same size as u_t.

The above example and discussion sheds some light on the fact, observed in Section 2.2, that a small measurement error may have effects as large as those shown in Figure 3, Panel (c). Our simulation exercises in the Online Appendix, Section F, suggest that, with $m = q + 1$, cases like the one of the example above may occur.

Clearly, the larger is m, the more unlikely they are. For instance, in the above example, if we have a third common component $\chi_{3t} = a_3 u_t + u_{t-1}$, the non-zeroless region is defined by $a_2 = a_3 = 0$, so that we only have problems when both $|a_2|$ and $|a_3|$ are close to 0. In our simulations reported in the Online Appendix, Section F.1, problematic cases no longer occur when m is larger than $q + 1$.

F Additional simulations

In this Appendix, we run additional simulation exercises, based on the model used in Section 4.

F.1 Changing m and the variable specification

In Simulation 4, we assess the performance of the CC-SVAR for different values of m. We estimate the common components using the true number of factors, i.e. $r = 5$. We run: (a) a VAR(4) with the common components of capital and taxes and the first principal component ($m = 3$); (b) a VAR(1) with the common components of capital and taxes and the first two principal components ($m = 4$); (c) a VAR(2) with the same variables (again $m = 4$); (d) a VAR(1) with the common components of capital and taxes and the first three principal components ($m = 5$). As above, we identify the tax shock by imposing that it is the only one affecting cumulated taxes in the long run. We repeat the exercise for 1000 data sets.

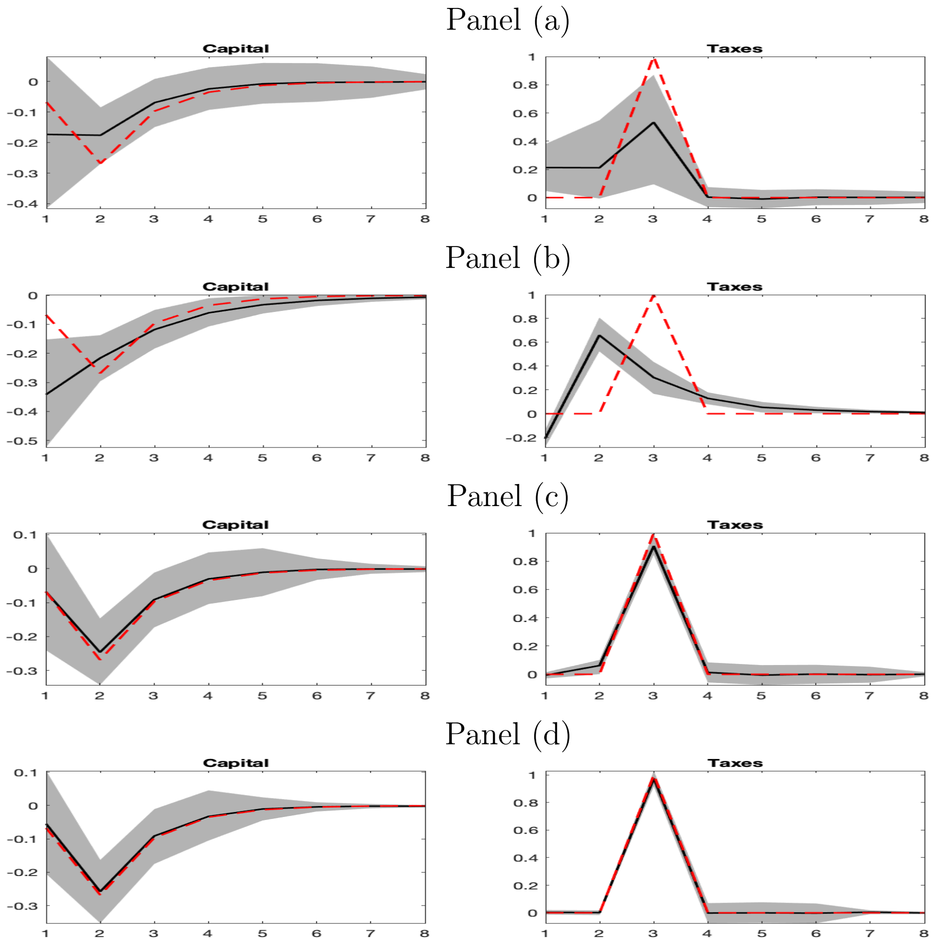


Figure 10: Simulation 4. The choice of m. Estimated IRFs for the tax shock. The red dashed lines are the theoretical IRFs. The solid lines represent the mean (across 1000 simulated datasets) of the point estimates. The grey areas represent the 16th and 84th percentiles of the point estimate distribution. Panel (a): CC-SVAR(4) with Capital, Taxes and the first principal component ($m = 3$). Panel (b): CC-SVAR(1) with Capital, Taxes and the first 2 principal components ($m = 4$). Panel (c): CC-SVAR(2) with Capital, Taxes and the first 2 principal components ($m = 4$). Panel (d): CC-SVAR(1) with Capital, Taxes and the first 3 principal components ($m = 5$).

Figure 10 reports the results. The red dashed lines are the theoretical impulse response functions. The solid lines are the mean point estimates (mean over the different datasets) and the grey areas represent the 16th and 84th percentile of the point-estimate distribution. The results for specification (a) are reported in Panel (a). We see that there is a sizable bias and a large variability of the results, especially for taxes. This disappointing result is discussed below. Here we only observe that the number of lags included in the VAR is not responsible for it. Indeed, a similar result (not shown) is obtained with 8 lags instead of 4.

Panel (b) and (c) show results for specifications (b) and (c), respectively. The difference is the number of lags included: just one lag in Panel (b) and two lags in Panel (c). Comparing the two panels, it is seen that when $m = 4$ we need two lags in the VAR to get good estimates of the impulse response functions. Panel (d) confirms that, with $m = 5$, just one lag is enough, consistently with equation (17). In both Panels (c) and (d), the dynamics are estimated extremely well, with the mean impulse response functions almost overlapping with the theoretical ones. Notice that, with the more parsimonious model in (d), the variability of the estimates is somewhat smaller at large lags. In the present case the advantage of specification (d) is modest, since T is relatively large and the number of parameters to estimate is small even for specification (c). But for shorter data sets or data sets requiring a larger number of parameters, like the ones of the empirical applications in Section 5, the advantage of a more parsimonious specification could be important.

To shed some light on the disappointing result obtained with $m = 3$, we run Simulation 5, analyzing what happens when changing the variables included in the CC-SVAR, for different values of m. For this exercise, we generate just one data set. As above, we use five principal components to estimate the common components.

To begin, we set $m = 3$. Then we estimate one hundred of different CC-SVAR(4) specifications, including the common components of capital and taxes, plus the common component of the $3 + i$-th variable, $i = 1, ..., 100$. The result is reported in Figure 11, Panel (a). The red lines are the 100 estimated impulse response functions, the black lines are the true impulse response functions. We see that there are several specifications which produce bad estimates, despite the fact that we have $m = q + 1$. We repeat the exercise by using the true common components in place of the estimated ones. The result is reported in Panel (b). With the true common components the results are good, consistently with the zeroless assumption (SDFM7). Hence the bad results of Panel (a) are due to

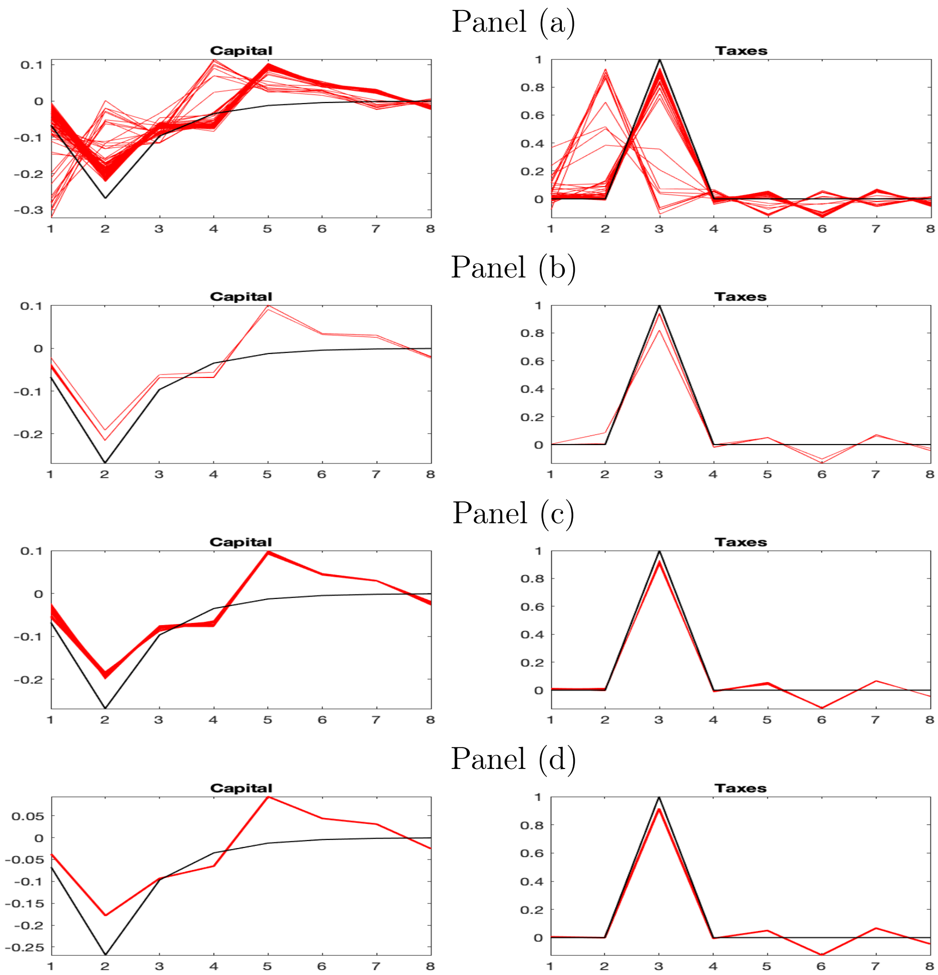


Figure 11: Simulation 5. The choice of ψ with $m < r$ and $m = r$. Estimated IRFs for the tax shock, for a single simulated data set. The black lines are the theoretical IRFs. The red lines are the CC-SVAR estimates obtained with different variable specifications. Panel (a): CC-SVAR(4) with Capital, Taxes and a third variable, changing across specifications ($m = 3$). Panel (b): same as Panel (a) with the true common components in place of the estimated ones. Panel (c): CC-SVAR(4) with Capital, Taxes the changing variable and the first principal component ($m = 4$). Panel (d): CC-SVAR(4) with Capital, Taxes, the changing variable and the first 2 principal components ($m = 5$).

the fact that the estimated common components are close to singular, though not exactly singular. When the specification is such that $B(L)$ is close to the non-zeroless region, the small idiosyncratic residual, which is still present in the estimated common components, produces large estimation errors.

Panels (c) and (d) show results for $m = 4$ and $m = 5$, respectively. We use four lags as before. In Panel (c) we include the same (estimated) common components of Panel (a), plus the first principal component as the fourth variable, equal for all specifications. We see that in this case the problem arising with $m = 3$ is solved. This is because matrices $B(L)$ very close to the non-zeroless region are much more unlikely, and actually never occur for this data set.[13]

Finally, in Panel (d) we have $m = 5$: the common components of capital and taxes, the third common component, changing across specifications, plus the first two principal components, which are kept fixed for all specifications. Consistently with the analysis in Section 3.7, all specifications produce exactly the same result, so that they produce a single line.

F.2 Changing r

In Simulation 6 we suppose that r is not known and use the criterion (E5), see Section 3.6, to determine the final value of $\hat{r}$. We try some values of $\hat{r}$ between 2 and 7. In all cases we set $m = \hat{r}$. For $m = \hat{r} = 2$ we estimate a CC-SVAR(2) including the common components of capital and taxes. For $m = \hat{r} = 3$ we estimate a CC-SVAR(2) including the common components of capital and taxes and the first principal component. For $m = \hat{r} = 7$ we estimate a CC-SVAR(2) including the common components of capital and taxes and the first five principal components. As usual, we repeat the exercise for 1000 data sets.

Figure 12 shows the results. In panels (a) and (b), corresponding to $m = \hat{r} = 2$ and $m = \hat{r} = 3$ respectively, the impulse response functions are badly estimated, whereas for $m = \hat{r} = 7$, panel (c), the results are pretty good, and very similar to those already obtained for $m = \hat{r} = 5$. Thus, with our simulated data, the criterion (E5) to determine the final value of $\hat{r}$ produces the correct result.

F.3 Cointegration

In Simulation 7 we show results about cointegration. The model of equation (2) is modified in such a way to have cointegration. We assume now that technology

[13] Indeed, we did not find bad specifications for $m = 4$ even for several other data sets, not shown here.

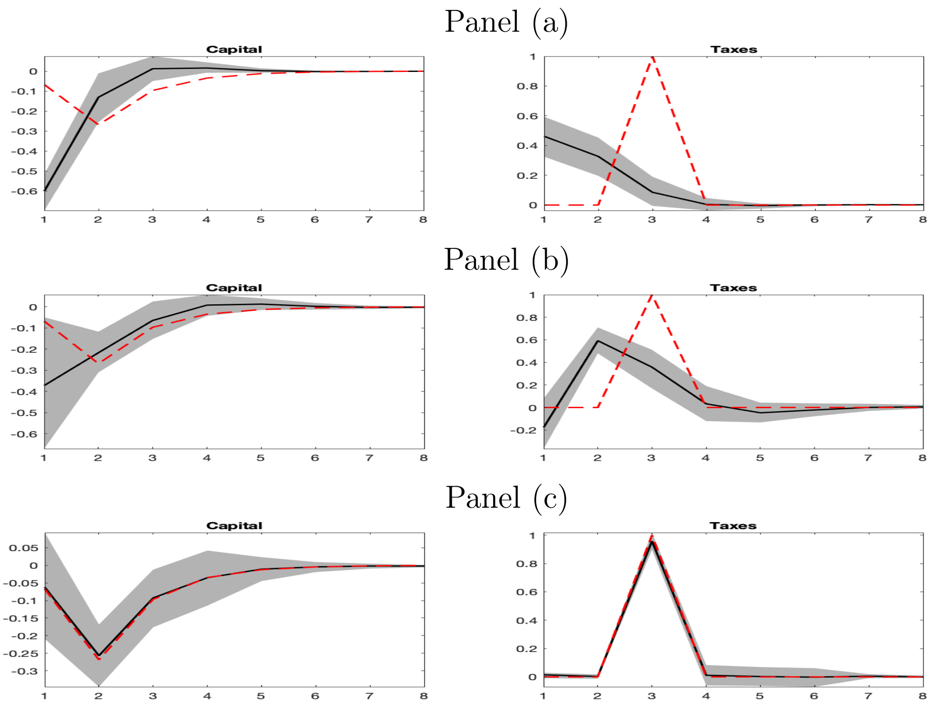


Figure 12: Simulation 6. The choice of $\hat{r}$. Results for $m = \hat{r} < r$ and $m = \hat{r} > r$. Estimated IRFs for the tax shock. The red dashed lines are the theoretical IRFs. The solid lines represent the mean (across 1000 simulated datasets) of the point estimates. The grey areas represent the 16th and 84th percentiles of the point estimate distribution. Panel (a): CC-SVAR(2) with $\hat{r} = m = 2$ (Capital and Taxes). Panel (b): CC-SVAR(2) with $\hat{r} = m = 3$ (Capital, Taxes and the first principal component). Panel (c): CC-SVAR(2) with $\hat{r} = m = 7$ (Capital, Taxes and the first 5 principal components).

a_t follows the random walk model $a_t = a_{t-1} + u_{a,t}$ and taxes are affected with one period of delay, $\tau_t = u_{\tau,t-1}$. The models is

$$\begin{pmatrix} \Delta a_t \\ \Delta k_t \\ \tau_t \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ \dfrac{-\delta(1-L)}{1-\alpha L} & \dfrac{1}{1-\alpha L} \\ L & 0 \end{pmatrix} \begin{pmatrix} u_{\tau,t} \\ u_{a,t} \end{pmatrix} = B(L)u_t. \tag{43}$$

Moreover, we use a slightly different parametrization to emphasize the problems arising from cointegration. We now set $\delta = 0.9$ and $\alpha = 0.8$. We generate 1000 data sets with $T = 1000$, without measurement errors. First, we estimate

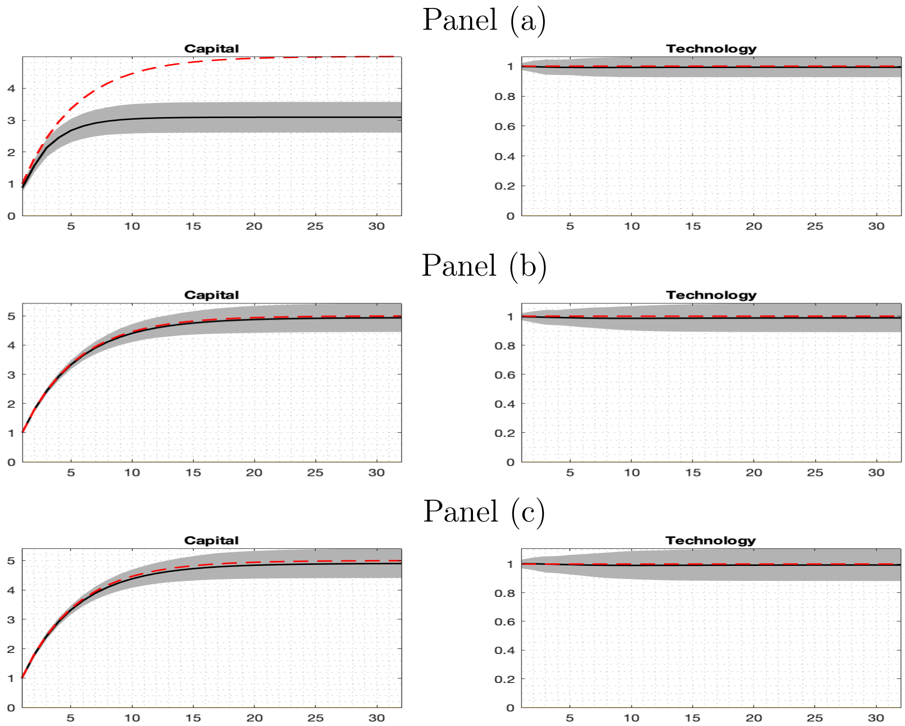


Figure 13: Simulation 7. Cointegration. Estimated IRFs for the technology shock. The red dashed lines are the theoretical IRFs. The solid lines represent the mean (across 1000 simulated datasets) of the point estimates. The grey areas represent the 16th and 84th percentiles of the point estimate distribution. Panel (a): VAR(2) with Capital and Technology, without measurement error. Panel (b): VAR(2) with Capital, Technology and Taxes, without measurement error. Panel (c): Large data set with measurement errors. CC-SVAR(2) with Capital, Technology, Taxes and the first principal components.

a bivariate VAR(2) with Δa_t and Δk_t, and identify the technology shock by imposing that it is the only shock having long-run effect on technology. This model is not affected by non-fundamentalness, but is affected by cointegration problems, since the upper 2×2 sub-matrix in (43) is singular for $L = 1$, i.e. the VMA of the two variables in growth rates is non-invertible. Then we estimate a VAR(2) model with Δa_t, Δk_t and τ_t. Notice that this model is singular, so that, apart special cases, it is not affected by cointegration problems, as discussed in the main text. Finally, we add 200 artificial common components, obtained by combining randomly the 4 factors technology, capital, taxes and the tax shock. To simulate measurement errors we add to all common components independent unit variance white noises and estimate a CC-SVAR(2) with the estimated common components of technology, capital, taxes and an additional variable (so that $m = r = 4$).

The results are shown in Figure 13. Panel (a) shows results for the bivariate VAR: the long-run response of capital is underestimated by about 30% on average. Panel (b) shows results for the trivariate singular VAR. Since $B(L)$ is zeroless, we have a VAR for the first differences and cointegration problems disappear. Panel (c) shows results for the third model, the almost singular VAR obtained by estimating the common components of 4 variables. The perfomance is similar to the one of the previous model.

G Technology shocks and hours worked

In this application we study the effects of technology shocks on hours worked. The empirical result has relevant implications for economic theory. According to the existing SVAR literature, the effect of technology shocks depends crucially on the treatment applied to the time series of hours worked, see in particular Galí (1999) and Christiano et al. (2003). In a bivariate SVAR with labor productivity and hours, the sign of the response of hours depends on whether hours are entered in log-levels or growth rates. In the first case hours increase, while in the second hours fall. Here we show that, when information is properly taken into account, hours increase, independently of the data treatment chosen.

We use the quarterly dataset of McCracken and Ng (2016).[14] Excluding a few variables to obtain a balanced panel, we end up with 215 variables. We transform each series to reach stationarity. We apply the criterion proposed by Alessi et al. (2010) and find a number of static factors $\hat{r} = 10$. Thus we use, as baseline specification, $\hat{r} = 10$.

We estimate 50 three-variable SVAR(4) specifications. All specifications include the growth rate of labor productivity and the growth rate of per-capita hours. The third variable differs across specifications and is chosen randomly. We identify the technology shock following Galí (1999), assuming that it is the only shock affecting labor productivity in the long run. Panel (a) of Figure 14 reports

[14]The data set is available at https://research.stlouisfed.org/econ/mccracken/fred-databases/.

the estimated impulse response functions. The response of hours is negative in most specifications. However, in two of them hours increase in a hump-shaped manner. So a decrease in hours worked after a positive technology shock is not a fully robust finding.

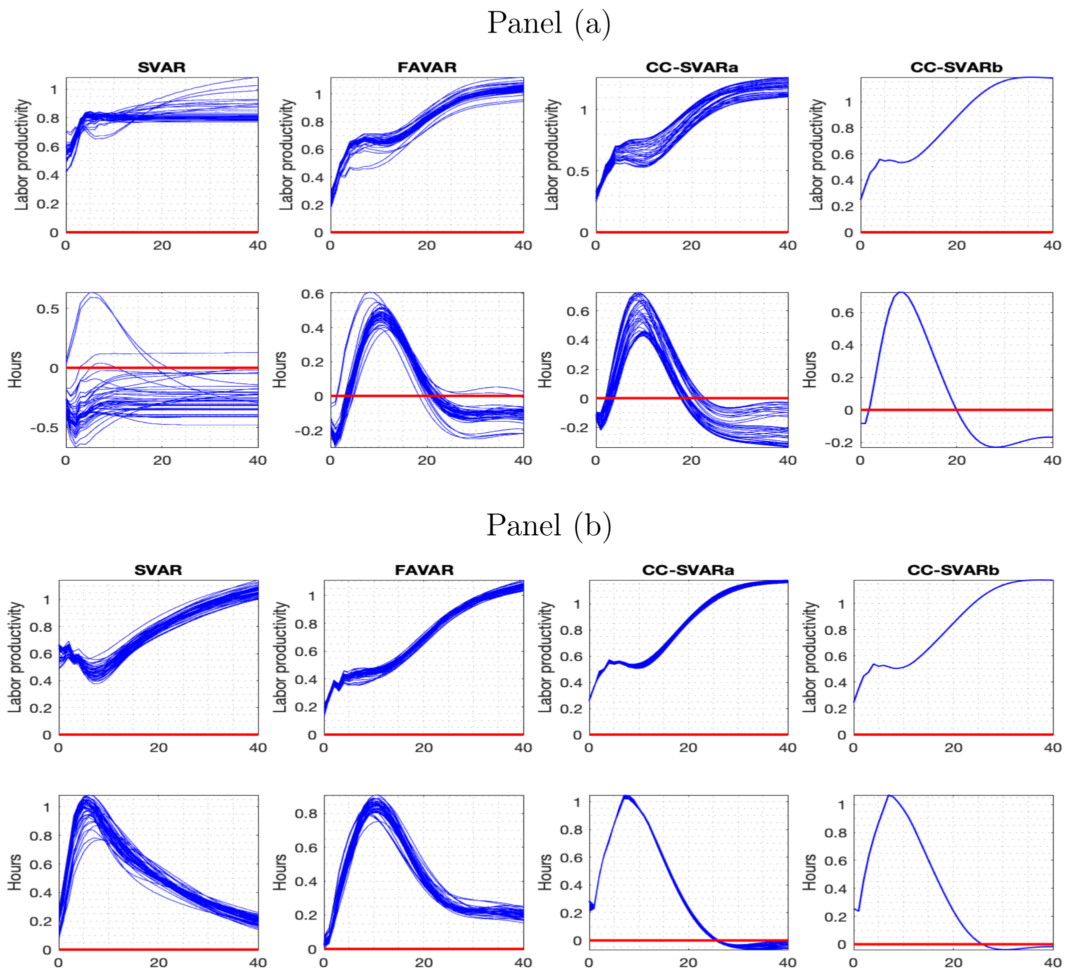


Figure 14: US quarterly data. The IRFs of a productivity shock. Galí (1999) long-run identification. The red lines are the CC-SVAR estimates obtained with different variable specifications. Panel (a): Hours in first differences. Panel (b): Hours in levels. First column: SVAR(4) for 223 three-variable specifications, differing for the fifth variable. Second column: FAVAR(4) the variables in the first column are augmented with the first 10 principal components. Third column: CC-SVAR(4): the variables in the first column are replaced with their common components; in addition, we include the first seven principal components ($\hat{r} = 10$).

Column 1 of Panel (b) reports the response of labor productivity and hours when hours enter in log-levels. Again there are 50 specification differing for the third variable included in the model. In this case the response of hours is positive in all specifications.

Column 2 of panels (a) and (b) report the response of labor productivity and hours worked estimated with a FAVAR(4) including labor productivity and hours (in growth rates panel (a), in levels panel (b)), a third variable again, and the first 6 principal components. In all these specifications information is large. In comparison to the VARs, the results of different FAVAR specifications are more similar to each other. Most importantly, in panel (a) the effect on hours is now positive, but for the first year after the shock.

The third and fourth columns report the responses obtained with the CC-SVAR including the common components of the same three variables, together with either the first 6 principal components (third column) or the first 7 principal components (fourth column). As expected, for each treatment of the hours, the responses are very much similar or identical across specifications. Moreover, enhancing the result already obtained with the FAVAR, the responses are similar across treatments: hours are nearly zero on impact and then increase, reaching their maximum after around two years.

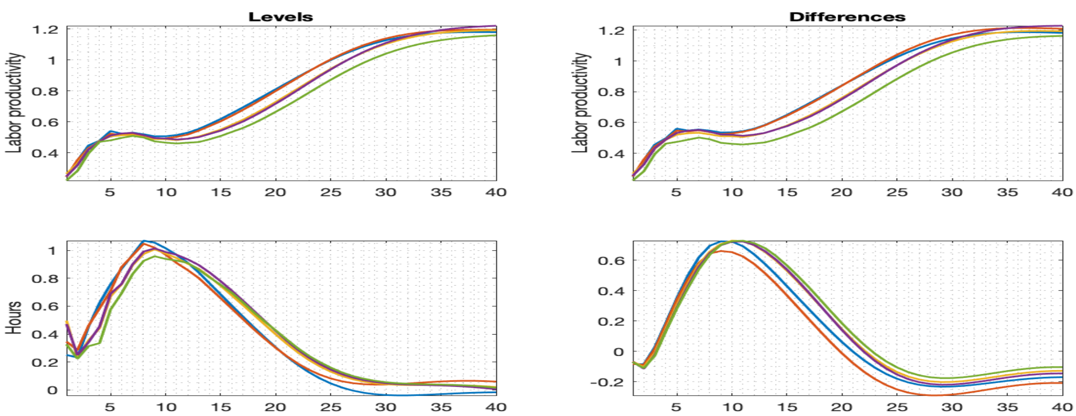


Figure 15: US quarterly data. The IRFs of a technology shock. CC-SVAR(6) with $m = \hat{r}$, using different values of $\hat{r}$. Black dotted line: $\hat{r} = 8$. Blue dashed line: $\hat{r} = 10$. Red solid line: $\hat{r} = 12$.

As in the previous application in Section **??**, we repeat the CC-SVAR analysis using $m = \hat{r} = 10, 11, 12, 13, 14$. The results are displayed in Figure 15, Panel (b). Results are consistent across different choice of $\hat{r}$.

The most important result of the CC-SVAR analysis is that, as we have seen, the response of hours does not depend on whether hours worked are taken in first differences or in levels. Apart from a minor difference in the sign of the first impact, in both cases a small impact effect is followed by a large positive hump-shaped increase. Thus we contribute to solving, or at least substantially smoothing, the old-standing controversy about the effects of technology shocks on hours worked, see Galí (1999) and Christiano et al. (2003): Hours, in line with Christiano et al. (2003), do increase following a positive technology shock.

Another remarkable difference of CC-SVAR with respect to SVAR results is that labor productivity, and thus output, increases very slowly in the CC-SVAR with both treatments of hours, which is not the case with the SVAR, particularly when using hours in differences. This, together with a slow increase of hours, is consistent with the view of the technology shock as a news shock. With this interpretation of technology, the response of hours estimated in the CC-SVAR is fully in line with both New-Keynesian models with nominal wage and price rigidities, see e.g. Barsky and Sims (2009), Christiano et al. (2010), Barsky et al. (2015), and with RBC models featuring frictions like habit formation in consumption, adjustment costs of investment, or with the assumption on preferences in Jaimovich and Rebelo (2009) (see also Schmitt-Grohé and Uribe (2012)).