EIEF WORKING PAPER SERIES

IEF

EINAUDI INSTITUTE FOR ECONOMICS AND FINANCE

EIEF Working Paper 23/06

September 2023

Common Components Structural VARs

By

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
(EIEF)

Luca Sala
(Università Bocconi, IGIER and Baffi Carefin)

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

February 2, 2024

Abstract

Structural VAR models (SVAR) produce results that can vary dramatically with the choice of variables, because information is deficient and/or contaminated by measurement errors. We argue that if the variables of interest belong to a High-Dimensional Factor Model and are replaced in the SVAR by their common components, both the information and the measurement issues find a solution under the condition that the number of common components is larger than the number of structural shocks, so that the SVAR is singular. This is the Common Components Structural VAR (CC-SVAR). Our main contribution is a complete asymptotic theory for the SVAR estimated using the finite-sample approximations to the common components. We apply our procedure to monetary policy shocks, finding that, with the CC-SVAR, results are robust to the choice of variables and well-known puzzles disappear.

JEL classification: C32, E32.

Keywords: structural VAR models, structural factor models, non-fundamentalness, measurement errors.

*We thank Matteo Barigozzi, Manfred Deistler, Philipp Gersing, Tommaso Proietti, Christoph Rust and Paolo Zaffaroni for very useful comments and suggestions.

†E-mail: mario.forni@unimore.it.

‡E-mail: luca.gambetti@uab.cat.

Gambetti acknowledges 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 (EBP) 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 specifications, 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 often contaminated by errors.

It is well known by now that the structural shocks of interest not always are linear

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

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. 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, especially in presence of news technology shocks (Forni et al., 2014), fiscal foresight (Leeper et al., 2013) or forward policy guidance (Ramey, 2016).[2] 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 that this view is wrong (see also Simulation 1, Section 2.2): even small measurement errors can generate substantial distortions in the estimates of the IRFs, yielding misleading results.[3] Indeed, measurement errors can be regarded as a source of non-fundamentalness. If m variables are driven by q structural shocks, but are contaminated by m independent measurement errors, their IRF 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 ap-

[2]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).

[3]Measurement errors can be regarded as special cases of aggregation, as analyzed in Forni and Lippi (1997) and Forni and Lippi (1999).

proach, 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).

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 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 EBP 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.

The problem is that, 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 errors, provide an estimate of χ_t, say $\hat{\chi}_t$, so that fundamentalness is restored.

CC-SVARs consist in the application of SVAR analysis to $\hat{\chi}_t$. From a theroretical point of view, our approach improves over previous factor-based structural models, such as the Structural Dynamic Factor Model (SDFM) of Stock and Watson (2005), Bai and Ng (2007) and Forni et al. (2009), or the Factor Augmented VAR (FAVAR) of Bernanke et al. (2005). First, the existence of a finite VAR representation in the structural shocks is not assumed, but derived from the theory of singular stochastic vectors. Second, we show that, in the singular case, a finite VAR representation does exist, under mild conditions, even when χ_t includes the first differences of cointegrated variables. Third, 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, a problem that has been overlooked in the above mentioned literature dealing with SDFMs and FAVARs.[4]

The CC-SVAR procedure allows for the inclusion, in the vector $\hat{\chi}_t$, of observable variables, insofar as their measurement error is zero. Moreover, it allows for the inclusion of estimated factors in place of common components. Hence it unifies and encompasses previous structural factor model methods. The CC-SVAR procedure essentially reduces to the SDFM method when the number of common components included in the VAR is equal to the number of factors.[5] On the other hand, the CC-SVAR reduces to a FAVAR when some variables are included in the vector $\hat{\chi}_t$ without treatment and the common components are replaced by the estimated factors.[6]

In the empirical part of the paper, we apply the CC-SVAR method to the study of

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

[5]Notice however that the CC-SVAR procedure is simpler, in that it does not require estimation of the number of primitive shocks q.

[6]Notice however that the identification of structural shocks is more direct and transparent when the restrictions are imposed on the common components of the variables rather than the factors.

the effects of monetary policy shocks on the main macroeconomic variables, an highly debated problem in macroeconometrics. As shown above, the results of SVAR analysis are not robust. By contrast, with the CC-SVAR, the puzzles disappear and the results are robust both across specifications and across different identification schemes.

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 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 the empirical application. Section 6 concludes.

2 Illustration by 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

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

where $0 < \alpha < 1$, $0 < \theta < 1$, $\delta = (1-\theta)\tau/(1-\tau)$, τ being the steady state tax rate, $0 \leq \tau < 1$; a_t, k_t and τ_t are the log deviations from the steady state of technology, capital and the tax rate, respectively; E_t denotes expectation at time t, conditional on $a_{t-j}, k_{t-j}, \tau_{t-j}, j \geq 0$. Technology and taxes are assumed, for simplicity, to be *i.i.d* processes, i.e. $a_t = u_{a,t}$ and $\tau_t = u_{\tau,t-2}$, where $u_{\tau,t}$ and $u_{a,t}$ are shocks that economic agents can observe. The equation for taxes implies a delay of two periods. Solving for k_t

we obtain the following equilibrium ARMA representation:

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

where L is the lag operator, $\chi_t = (a_t \ k_t \ \tau_t)'$ and $u_t = (u_{\tau,t} \ u_{a,t})'$.

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, considering the square subsystems including technology and capital, technology and taxes, capital and taxes, the determinants of the corresponding submatrices of the moving average matrix polynomial in (1) are, respectively, $\delta(z+\theta)$, $-z^2$, $-z^2$. 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 (1) is obtained here by a simulation exercise (Simulation 1). We generate 1000 different dataset with 200 time observations from model (1) using the parameterization in Leeper et al. (2013) for α, θ, τ and u_t. 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 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, corresponding to the 68% confidence interval commonly used in macroeconometrics. 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, provided that $\tau \neq 0$, the matrix $B(L)$ in (1) has a left-inverse, so that the vector $\chi_t = (a_t \ k_t \ \tau_t)'$ has the VAR(3) representation

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

Note that the autoregressive matrix is stable, since its determinant is $(1-\alpha L)$ and $|\alpha| < 1$. Note also that, despite singularity of the spectral density matrix, in the present case 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 χ_t. 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 with the tax shock ordered first.

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. Note that correct estimation crucially depends on the fact that information is enlarged *without adding further shocks or noise*, which is tantamount to saying that the enlarged vector of 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 (1) by adding measurement errors, i.e. we consider the vector variable $x_t = \chi_t + \xi_t$, where the vector $\xi_t = (\xi_t^a \ \xi_t^k \ \xi_t^\tau)'$ is white noise and 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. Thus, even when information seems sufficient to correctly recover the impulse-response functions, 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

Let χ_t be an m-dimensional vector whose coordinates are the "true" macroeconomic variables of interest. We assume that χ_t has an ARMA structural representation driven by a q-dimensional structural shock vector u_t and that $m > q$, so that χ_t is (dynamically) singular (more variables than shocks). This implies that generically χ_t has a finite-length VAR representation and that u_t is fundamental for χ_t. On the other hand χ_t can only be observed with measurement errors. We obtain a consistent estimate $\hat{\chi}_t$ by means of factor-model techniques and show that the CC-SVAR, i.e. the SVAR applied to $\hat{\chi}_t$, produces consistent estimates of u_t and the IRFs of χ_t with respect to u_t.

3.1 The Impulse-Response Function representation

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

$$H(L)\chi_t = K(L)u_t, \tag{3}$$

where: **(a)** u_t *is a serially independent, q-dimensional vector of orthonormal shocks, with* $q \leq m$. **(b)** $H(L)$ *is an* $m \times m$ *polynomial matrix such that* $\det H(z) = 0$ *implies* $|z| > 1$. $K(L)$ *is a full rank* $m \times q$ *polynomial matrix, i.e.* $\text{rank}(K(z)) = q$ *but for a finite number of complex numbers. Thus*

$$\chi_t = H(L)^{-1}K(L)u_t = B(L)u_t = B_0 u_t + B_1 u_{t-1} + \cdots \tag{4}$$

(c) *The vector* u_t *is structural, so that the matrices* $B_j = \text{E}(\chi_t u'_{t-j})$ *is structural. We suppose that* $H(L)$ *and* $K(L)$ *are structural as well, although of course there exist alternative ARMA representations of* χ_t*, with* u_t *as the driving white noise.*

Equation (1) is of course a special case of (3). Equations of the form (3) or (4) are easily obtained from the the state-space representation of a linearized DSGE. Regarding (c), to fix ideas we may suppose that the upper $q \times q$ submatrix of B_0 in (4) is lower triangular, so that the shocks u_{jt} impact contemporaneously on the χ_{it}, $i = 1, \ldots, q$, according to a recursive (Cholesky) scheme. In Section 3.4 a recursive scheme will be used as a simplifying assumption.

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. Moreover, 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

We now present and illustrate some basic consequences of Assumptions 1 and 2.

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

$$\chi_{1t} = u_t + k_1 u_{t-1}, \qquad \chi_{2t} = u_t + k_2 u_{t-1}, \tag{5}$$

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 $u_t = (k_2 - k_1)^{-1}(k_2\chi_{1t} - k_1\chi_{2t})$. This can be used to replace u_{t-1} in (5), obtaining

$$\chi_{1t} = \frac{k_1}{k_2 - k_1}(k_2\chi_{1,t-1} - k_1\chi_{2,t-1}) + u_t, \quad \chi_{2t} = \frac{k_2}{k_2 - k_1}(k_2\chi_{1,t-1} - k_1\chi_{2,t-1}) + u_t, \tag{6}$$

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. Thus the white noise u_t in (5) is fundamental and that χ_t has a finite-length autoregressive representation for all values of the parameters k_1 and k_2, with the exception of the line $k_1 = k_2$.

Model (1) in Section 2 provides another example of a singular vector having a rational MA representation, which admits the finite-length VAR representation (2), unless $\tau = 0$.

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

Definition 1. Zerolessness. *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. (Zerolessness implies full rank but not viceversa.)*

Note that if $m = q$ then $K(L)$ is zeroless if and only if has a constant determinant ($K(L)$ is unimodular), a very special case. 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 crucial 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 $A(L) = \tilde{K}(L)H(L)$, 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 (1) and (5) the matrix $K(L)$ is zeroless with the exception of a lower dimensional subset of the parameter space. Precisely, in Example (1) the two-dimensional subset of $\{0 < \alpha < 1,\ 0 < \theta < 1,\ 0 \leq \tau < 1\}$ where $\tau = 0$. In Example (5) the one-dimensional subset of $\mathbb{R}^2$ where $k_1 = k_2$.

We say that $K(L)$ is generically zeroless in examples (1) and (5), 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 (5), 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 and Forni et al. (2015).

(II) However, in this paper we are interested in the case in which, like in example (1), 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 (Z) that condition generically holds, or (W) it generically fails to hold. Note that usually in the non-singular case neither of the alternatives holds generically.

(III) Moreover, alternative (W) above holds only 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".

3.2.2 Singularity and cointegration

Now let $\chi_t = (1 - L)X_t$, where X_{it} is $I(1)$ for $i = 1, \dots, m$. 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 (5), with $\chi_t = \Delta X_t$, and take the linear combination

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

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 (6) holds for χ_t, so that X_t has a VAR(1) representation in differences. Thus, 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 in the Online Appendix F.4 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. This however is left for future research.

However, $\kappa > 0$ can be convincingly ruled out for macroeconomic applications in which real variables like GDP, consumption or investment are jointly modeled with prices, monetary aggregates or policy variables. Indeed $\kappa > 0$ implies that, for some of the shocks, the IRF of all the variables has the factor $(1-L)$. Of course a demand shock for example may well have transitory effects on trended real activity variables. But there are no theoretical reasons why it should have transitory effect on prices and monetary aggregates, or have the factor $(1-L)$ in the impulse response functions of $I(0)$ variables like interest rates, risk premia, term spreads or unemployment rates. This also suggests that, in empirical situations, $\kappa = 0$ can be forced, so to speak, by augmenting χ_t with suitable variables. See Appendix A.2 for a detailed version of the above argument.

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-length VAR representation

$$A(L)\chi_t = B_0 u_t = v_t. \tag{7}$$

where $A(L)$ is stable. As B_0 has full rank q, u_t, as well as v_t, is fundamental for χ_t.

3.3 Adding (and removing) measurement errors

We suppose that only $x_t = \chi_t + \xi_t$ is observable. We also suppose that x_t is a subvector of an observable n-dimensional vector $\boldsymbol{x}_{nt} = (x_{it}),\ i = 1, \ldots, n$ where 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). Formally:

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}*,* $i = 1, \dots, m$*, are the first* m *series of the sequence* $x_{it} = \chi_{it} + \xi_{it}$*,* $i = 1, \dots, \infty$*. The variables* χ_{it} *are called the common component and the variables* ξ_{it} *the idiosyncratic components. The idiosyncratic component* ξ_{it} *and* u_t *are zero-mean and mutually independent at all leads and lags, i.e.* ξ_{it} *and* u_{t-k} *are independent for all* $i \in \mathbb{N}$*,* $k \in \mathbb{Z}$*, so that* ξ_{it} *and* χ_{jt-k} *are independent for all* i *and* k*.* **(c)** *The researcher observes the first* n *series* x_{it}*,* $i = 1, \dots, n$*, with* $n \geq m$*.*

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 or consumption, 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 of them 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 (7) 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. **(a)** *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* $\hat{\Sigma}_k^y$*.* **(b)** *By* $\hat{\chi}_t = (\hat{\chi}_{it})$*,* $i = 1, \dots, m$*,* $t = 1, \dots, T$*, we denote an estimator of* χ_t *based on* x_{it}*,* $i = 1, \dots, n$*,* $t = 1, \dots, T$*,* **(c)** $\hat{\pi}_t = \hat{\chi}_t - \chi_t$*,* **(d)** $|| \cdot ||$ *denotes the euclidean norm for vectors and the spectral norm for matrices.*

Assumption A. Properties of $\hat{\chi}_t$. *We have:* $||\hat{\chi}_t - \chi_t|| = ||\hat{\pi}_t||$ *is* $O_p(r_{n,T})$*, where* $r_{n,T} \to 0$ *as* $\min(n,T) \to \infty$*. Moreover,* u_t *is independent of* $\hat{\chi}_{t-k}$ *for* $k > 0$*.*

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

Assumption A states that the estimator is consistent as $\min(n,T) \to \infty$, the rate being $r_{n,T}$. Assumption B is a standard ergodicity property.

3.4 Estimating a singular VAR

It is convenient to re-write the population VAR in (7) as

$$\chi_t = A_1\chi_{t-1} + \cdots + A_p\chi_{t-p} + v_t = \mathcal{A}Z_{t-1} + v_t, \tag{8}$$

where $Z_t = (\chi_t' \ \chi_{t-1}' \ \ldots \ \chi_{t-p+1}')'$, $v_t = B_0 u_t$ is a white-noise vector of dimension m 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 (8) 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 $m = 3$, $q = 1$, $B(L) = B_0 + B_1L + B_2L^2 + B_3L^3$, 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 $u_{t-1}, \ldots, u_{t-p}, u_{t-p-1}, u_{t-p-2}, u_{t-p-3}$, 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 (8), 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 $A(L)$ is not necessarily unique. On the other hand, equation (8) is a projection equation, so that, by uniqueness of the orthogonal projection, the projection $\mathcal{A}Z_{t-1}$ and the residual v_t are unique. Of course, the vector of structural shocks u_t and the matrix of impulse response functions $B(L)$ are unique as well.[7] As $A(L)$ is not necessarily unique, the estimated

[7] Note that, inverting the matrix $A(L)$, we obtain $\chi_t = A(L)^{-1}v_t = A(L)^{-1}B_0u_t$. On the other hand,

VAR coefficients do not necessarily converge. This is the problem, mentioned in the Introduction, which has been overlooked in previous factor model literature.[8] Here we show that, even if the VAR in (8) is not unique, so that the estimated VAR coefficients may not converge at all, the estimated VAR residual $\hat{v}_t$, the estimated vector of structural shock $\hat{u}_t$ and the estimated impulse-response matrix $\hat{B}(L)$ converge to v_t, u_t and $B(L)$, respectively.

The empirical counterpart of (8) is

$$\hat{\chi}_t = \hat{A}_1\hat{\chi}_{t-1} + \cdots + \hat{A}_p\hat{\chi}_{t-p} + \hat{v}_t = \hat{\mathcal{A}}\hat{Z}_{t-1} + \hat{v}_t, \tag{9}$$

where $\hat{\mathcal{A}}\hat{Z}_{t-1}$ is the sample projection of $\hat{\chi}_t$ onto $\hat{Z}_{t-1}$ and $\hat{v}_t$ is the residual. Even if Σ_0^Z is singular, singularity of $\hat{\Sigma}_0^{\hat{Z}}$ is very unlikely, owing to the estimation error $\hat{\pi}_t$. In this case $\hat{\mathcal{A}}$ is unique and can be estimated by a standard VAR. On the other hand, the entries of $\hat{\pi}_t$ can be collinear (the possibility that an entry of $\hat{\pi}_t$ is null is discussed in Section 3.6), so that collinearity of $\hat{Z}_{t-1}$ might in principle occur. In this case, we evaluate the regressors in $\hat{Z}_t$ in reverse order from the last to the first and discard them whenever they are redundant, see Deistler et al. (2011). The corresponding columns of $\hat{\mathcal{A}}$ are set to 0. This defines uniquely $\hat{\mathcal{A}}$ and therefore $\hat{A}(L) = I - \hat{A}_1L - \cdots - \hat{A}_pL^p$. Of course $\hat{v}_t = \hat{A}(L)\hat{\chi}_t$ is unique because it is the residual of the sample projection equation (9). Our first result concerns the consistency of $\hat{v}_t$.

Proposition 1. Consistency of the VAR residuals. *Under Assumptions 1 through 4, A and B, we have* $||\hat{v}_t - v_t|| = O_p\left(\max(r_{n,T}, 1/\sqrt{T})\right)$ *, or, equivalently,* $||\hat{\mathcal{A}}\hat{Z}_{t-1} - \mathcal{A}Z_{t-1}|| = O_p\left(\max(r_{n,T}, 1/\sqrt{T})\right)$.

The proof is given in the Online Appendix B.

Let us turn now to the structural shocks and response functions. For simplicity we

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.

[8] If $p = 1$, $\Sigma_0^Z = \Sigma_0^\chi$, which is non-singular by Assumption 2(b). Thus (8) is unique. As mentioned in footnote 4, Forni et al. (2009) assume $p = 1$, thus avoiding the difficulty we are dealing with here.

specialize Assumption 1(c) by adopting the Cholesky scheme. Our consistency results can be easily adapted to all the structural relations between u_t and the matrices B_j that are currently used to identify macroeconometric models.

Assumption 1(c)′. Cholesky scheme. *We suppose that, after possible reordering of the variables* χ_{it}, $i = 1, \ldots, m$,

$$v_t = B_0 u_t = \begin{pmatrix} Q \\ R \end{pmatrix} u_t, \tag{10}$$

where Q *is* $q \times q$, *lower triangular with positive entries on the main diagonal and* R *is* $(m-q) \times q$.

Using (10) and partitioning v_t and Σ_0^v as

$$v_t = \begin{pmatrix} v_t^{[1]} \\ v_t^{[2]} \end{pmatrix}, \quad \Sigma_0^v = \begin{pmatrix} \Sigma_{[11]} & \Sigma_{[12]} \\ \Sigma_{[21]} & \Sigma_{[22]}, \end{pmatrix}, \tag{11}$$

where $v_t^{[1]}$ is $q \times 1$ and $\Sigma_{[11]}$ is $q \times q$, we see that $u_t = Q^{-1} v_t^{[1]}$ and that Q is the lower-triangular Cholesky factor of $\Sigma_{[11]}$. Moreover, as is easily seen, $R = \Sigma_{[21]}(Q')^{-1}$. Summing up,

$$\Sigma_{[11]} = QQ', \qquad u_t = Q^{-1} v_t^{[1]}, \qquad R = \Sigma_{[21]}(Q')^{-1}, \qquad B_0' = (Q' \; R')'.$$

Correspondingly, partition $\hat{v}_t$ and $\hat{\Sigma}_0^{\hat{v}}$ as

$$\hat{v}_t = \begin{pmatrix} \hat{v}_t^{[1]} \\ \hat{v}_t^{[2]} \end{pmatrix} \qquad \hat{\Sigma}_0^{\hat{v}} = \begin{pmatrix} \hat{\Sigma}_{[11]} & \hat{\Sigma}_{[12]} \\ \hat{\Sigma}_{[21]} & \hat{\Sigma}_{[22]}, \end{pmatrix},$$

where $\hat{v}_t^{[1]}$ is $q \times 1$, $\hat{\Sigma}_{[11]}$ is $q \times q$. By Proposition 1, $\hat{\Sigma}_{[11]}$ converges to $\Sigma_{[11]}$ in probability, thus $\det \hat{\Sigma}_{[11]}$ is bounded away from zero in probability. Then let $\hat{\Sigma}_{[11]} = \hat{Q}\hat{Q}'$ be its Choleski factorisation and define $\hat{u}_t = \hat{Q}^{-1} \hat{v}_t^{[1]}$. Summing up,

$$\hat{\Sigma}_{[11]} = \hat{Q}\hat{Q}', \qquad \hat{u}_t = \hat{Q}^{-1} \hat{v}_t^{[1]}, \qquad \hat{R} = \hat{\Sigma}_{[21]}(\hat{Q}')^{-1}, \qquad \hat{B}_0' = \left(\hat{Q}' \; \hat{R}'\right)'. \tag{12}$$

Lastly, $\hat{B}_k$, $k = 0, \dots, \infty$, is defined by solving $\hat{A}(L) \sum_{k=1}^{\infty} \hat{B}_k L^k = B_0$, that is

$$-\hat{A}_1 \hat{B}_0 + \hat{B}_1 = 0, \quad -\hat{A}_2 \hat{B}_0 - \hat{A}_1 \hat{B}_1 + \hat{B}_2 = 0, \quad \dots$$

Proposition 2 establishes consistency and consistency rates for $\hat{u}_t$ and $\hat{B}_k$.

Proposition 2. Consistency of the estimated structural shocks and IRFs. *Under Assumptions 1, as specified in Assumption 1(c)′, through 4, A and B:* **(a)** $||\hat{u}_t - u_t|| = O_p\left(\max(r_{n,T}, 1/\sqrt{T})\right)$, **(b)** *For any* $k \geq 0$, $||\hat{B}_k - B_k|| = O_p\left(\max(r_{n,T}, 1/\sqrt{T})\right)$.

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. (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 orthonormal static factors F_{kt}, $k = 1, \dots, r$, where $r > q$. The r-dimensional vector F_t has an ARMA representation like* (3) *in the structural shocks u_t and therefore a rational MA representation like in* (4)*:*

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

$$F_t = B_F(L) u_t.$$

By Assumption 4(b), (i) $\chi_t = (\chi_{1t} \ \cdots \ \chi_{mt})' = \Lambda_m F_t$, Λ_m being the $m \times r$ matrix with rows λ_i, $i = 1, \dots, m$, (ii) $\chi_t = B(L) u_t$, where $B(L) = \Lambda_m B_F(L)$. Assumption 2(b) and orthonormality of F_t imply that the rank of $\Lambda_m \Lambda_m'$ is m.

Some notation is needed for the following assumptions.

Notation 2. **(a)** $\boldsymbol{x}_{nt} = (x_{1t} \ \cdots \ x_{nt})'$, $\boldsymbol{\chi}_{nt} = (\chi_{1t} \ \cdots \ \chi_{nt})'$ *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* $\xi_t = \boldsymbol{\xi}_{mt}$. **(b)** Γ_k^x, Γ_k^χ *and* Γ_k^ξ *are k-lag covariance matrices of the processes* $(\boldsymbol{x}_{nt})$, $(\boldsymbol{\chi}_{nt})$ *and* $(\boldsymbol{\xi}_{nt})$, *respectively.* Σ_k^χ, *see Notation 1, is the upper-left* $m \times m$ *sub-matrix of* Γ_k^χ, *which is* $n \times n$. $\hat{\Gamma}_k^x$, *the sample counterpart of* Γ_k^x, *is* $\sum_{t=k+1}^{T} \boldsymbol{x}_{nt}\boldsymbol{x}'_{n,t-k}/(T-k)$. **(c)** μ_j^χ *and* μ_j^ξ, $\hat{\mu}_j^\chi$ *and* $\hat{\mu}_j^\xi$, *are the j-th eigenvalues, in decreasing order of magnitude, of* Γ_0^χ *and* Γ_0^ξ, $\hat{\Gamma}_0^\chi$ *and* $\hat{\Gamma}_0^\xi$, *respectively.*

Assumption 6. Pervasiveness of the factors and the shocks, non-pervasiveness of the idiosyncratic components. **(a)** *There exists constants* $\underline{c}_j, \bar{c}_j$, $j = 1, \ldots, r$, *such that* $\underline{c}_j > \bar{c}_{j+1}$, $j = 1, \ldots, r-1$, *and* $0 < \underline{c}_j < \liminf_{n\to\infty} n^{-1}\mu_j^\chi \leq \limsup_{n\to\infty} n^{-1}\mu_j^\chi \leq \bar{c}_j$, **(b)** *There exists a real* $\ell > 0$ *such that* $0 < \mu_1^\xi \leq \ell$.

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* $\gamma_{k,ij}^x$, $\hat{\gamma}_{k,ij}^x$, $\gamma_{k,ij}^\chi$ *and* $\hat{\gamma}_{k,ij}^\chi$ *the entries of* Γ_k^x, $\hat{\Gamma}_k^x$, Γ_k^χ *and* $\hat{\Gamma}_k^\chi$, *respectively. There exists a* $\rho > 0$ *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 all* 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.

Definition 2. The principal component estimator. *Let* $\hat{w}_j^x = (\hat{w}_{j1} \ \cdots \ \hat{w}_{jn})'$ *be a normalized column eigenvector of* $\hat{\Gamma}_0^x$ *corresponding to* $\hat{\mu}_j^x$ *(so that* $\hat{w}_j^{x\prime}\boldsymbol{x}_{nt}$ *is the j-th principal component of* $\boldsymbol{x}_{nt}$*). Let* $\mathcal{I}_m$ *be the* $n \times m$ *matrix with zeros in the last* $n-m$ *rows and* I_m *in the first* m. *The principal component estimator of* $\chi_t = \boldsymbol{\chi}_{mt}$ *is* $\hat{\chi}_t = \mathcal{I}'_m \hat{W}^{x\prime}\hat{W}^x \boldsymbol{x}_{nt}$, *where*

$\hat{W}^x$ *is the* $n \times r$ *matrix with* $\hat{w}^x_j$ *on the* j*-th column, that is,* $\hat{\chi}_{it} = \sum_{j=1}^r \hat{w}^x_{ij} \hat{w}^{x\prime}_j \boldsymbol{x}_{nt}, \quad i = 1, \ldots, m$.

Proposition 3. Properties of the principal component estimator. *Under Assumptions 1-7, for all* k*:* **(a)** $\|\hat{\pi}_t\| = ||\hat{\chi}_t - \chi_t|| = O_p(\max(1/\sqrt{n}, 1/\sqrt{T}))$*,* **(b)** $||\hat{\Sigma}^\chi_k - \Sigma^\chi_k|| = O_p\left(1/\sqrt{T}\right)$.

Note that, since $\hat{\chi}_t$ is a linear combination of the present and past values of the observables, independence of u_t and $\hat{\chi}_{t-k}$, $k > 0$, requested in Assumption A, is an immediate implication of Assumptions 1(a) and 4(b). Thus $\hat{\chi}_t$ fulfills Assumption A, with $r_{n,T} = \max(1/\sqrt{n}, 1/\sqrt{T})$, and Assumption B. The proof of Proposition 3 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. Transform the series to get stationarity and standardize them to have zero mean and unit variance. The standardized series are the entries of our vector $\boldsymbol{x}_{nt}$.

(E1) Estimate r. Out of the vast literature, beginning with Bai and Ng (2002), proposing consistent estimators $\hat{r}$, in the application of 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 according to Definition 2. Possibly, de-standardize the series to get the common components of the non-standardized variables.

If there is a strong a priori belief that the variable s is free of measurement error, the variable itself can be included in the model without treatment, 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 (7)).

(E4) Identify the structural shocks 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. 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.2. 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.2.

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.[9] The intuition is simple: since the entries of $\hat{\chi}_t$ are linear combinations of the estimated factors in $\hat{F}_t$ (i.e. the first m principal components

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

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 Appendix F.3, 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 of Section 2. Firstly we write our variables a_t, k_t and τ_t as linear combinations of 5 factors: $k_t, u_{a,t}, u_{\tau,t}, u_{\tau,t-1}, u_{\tau,t-2}$. Then we generate a data set with 200 variables, by taking random linear combinations of these factors. Finally, we add errors to all variables to get the observable series. Details are reported in Appendix F.1.

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.

The results are reported in Figure 4. Panel (a) shows the results for the mis-specified 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).

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).[10] 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, $\hat{r} = 8$. In the Online Appendix G we show that CC-SVAR results are robust to changes of the number of factors.

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 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.

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

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.

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 unemployment, 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).[11] To verify whether the condition $m > q$ is fulfilled, we estimate the number of shocks q by using the log criterion of Hallin and Liska (2007), which gives $\hat{q} = 4$. 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, 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.

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.

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

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.

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

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

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

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

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.

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

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

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.

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.

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.

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. A. (2016). Macroeconomic shocks and their propagation. *Handbook of Macroeconomics 2*, 71–162.

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

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. Working Papers, 11467, NBER.

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.

Figures

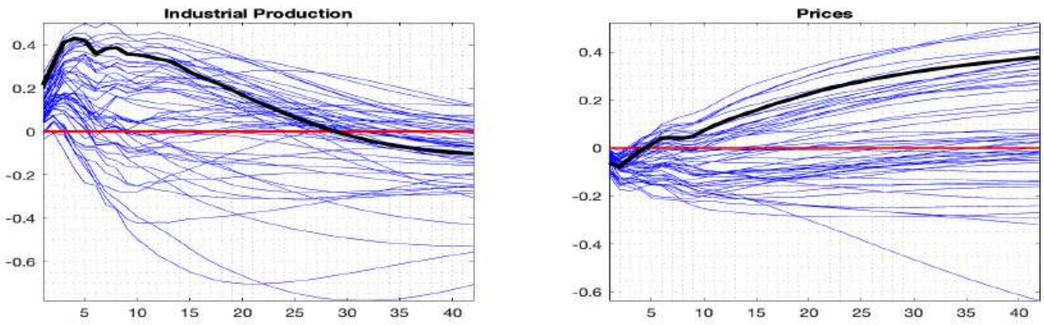

Figure 1: US monthly data from 1977:6 to 2008:12. The IRFs of a monetary policy shock, identified with the proxy of Gertler and Karadi (2015). The black lines are the IRFs of the SVAR(6) with just four variables: the 1 year bond rate, industrial production growth, unemployment and CPI inflation. The 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.

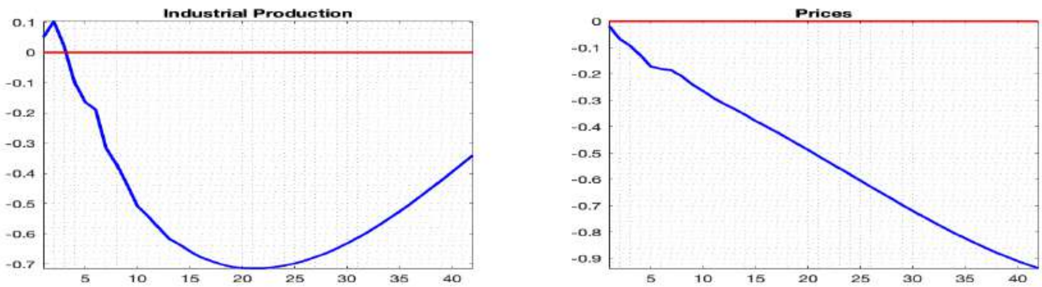

Figure 2: US monthly data from 1977:6 to 2008:12. The IRFs of a monetary policy shock, identified with the proxy of Gertler and Karadi (2015). The blue line is obtained by plotting the IRFs for the same 50 eight-variable specifications of Figure 1 obtained with the CC-SVAR.

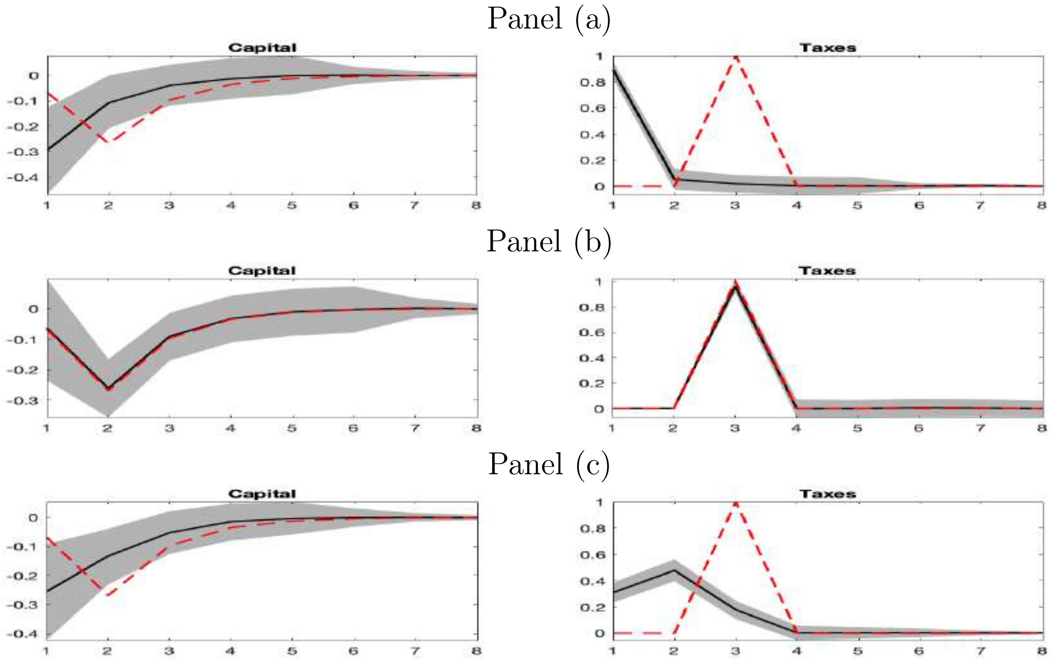

Figure 3: Simulation 1. Non-fundamentalness and measurement errors. Estimated IRFs for the tax shock. The red dashed lines are the theoretical IRFs. The solid lines represent the mean (across 1000 simulated datasets) of the point estimates. The grey areas contain the point estimates between the 16th 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.

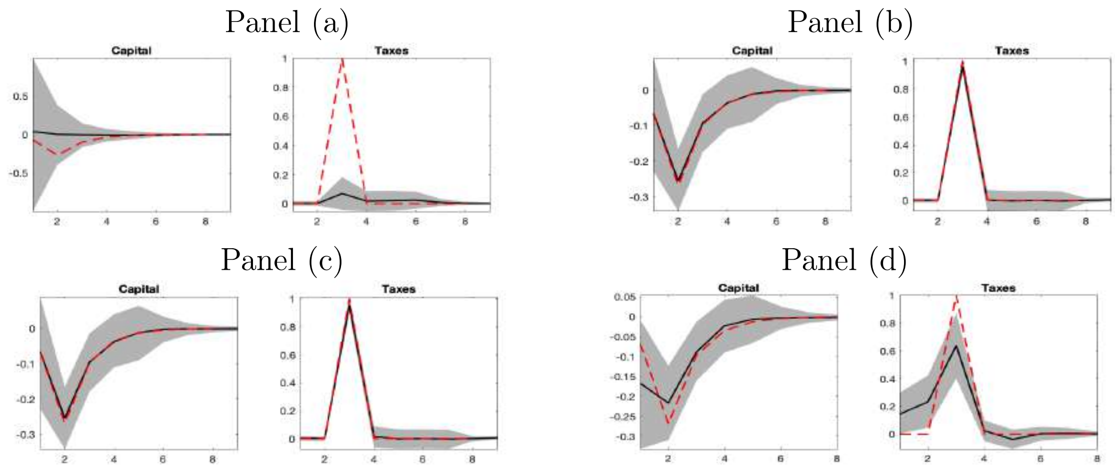

Figure 4: Simulation 2. Standard Procedure SDFM, CC-SVAR, FAVAR. Estimated IRFs for the tax shock. The red dashed lines are the theoretical IRFs. The solid lines represent the mean (across 1000 simulated datasets) of the point estimates. The grey areas contain the point estimates between the 16th and 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.

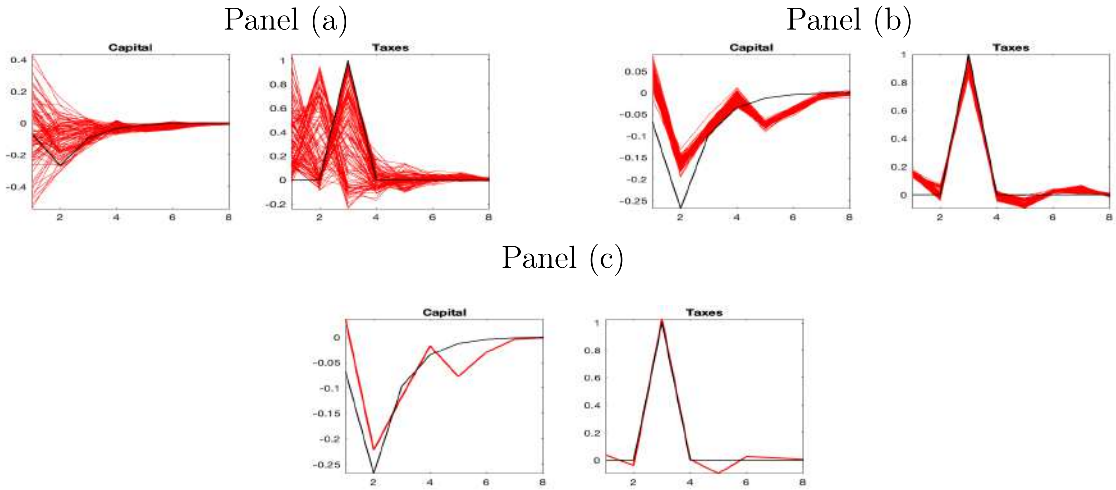

Figure 5: Simulation 3. Different variable specifications for a deficient VAR, the FAVAR and the CC-SVAR. Estimated IRFs for the tax shock, for a single simulated data set. The black lines are the theoretical IRFs. The red lines are the CC-SVAR estimates obtained with different variable specifications. 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.

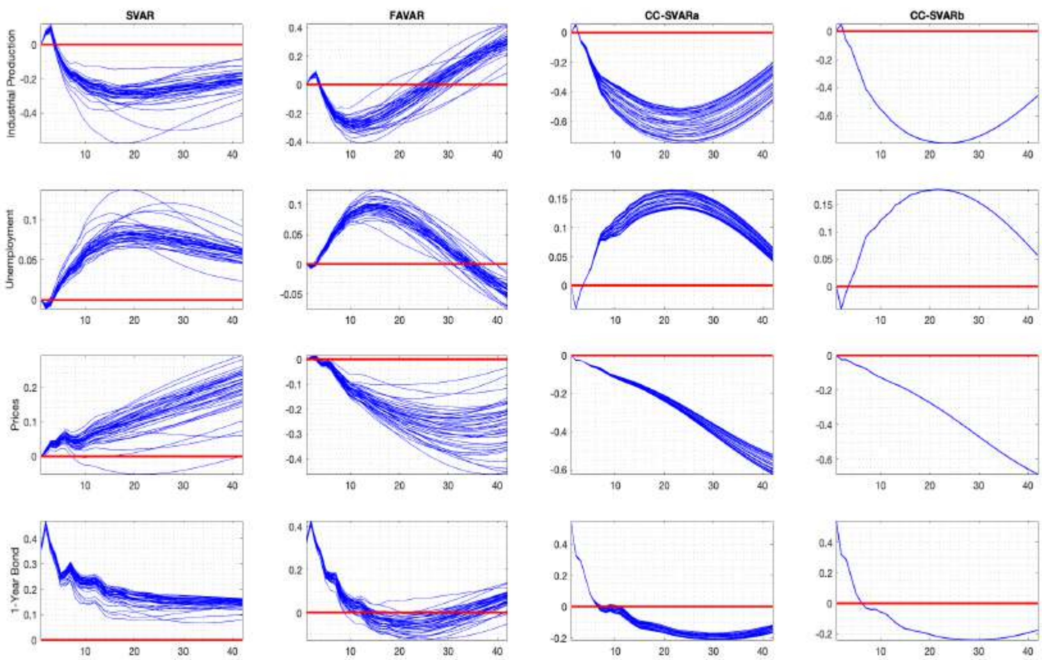

Figure 6: US monthly data. The IRFs of a monetary policy shock. Cholesky identification. The red lines are the CC-SVAR estimates obtained with different variable specifications. First column: SVAR(6) for 50 five-variable specifications, differing for the fifth variable. Second column: FAVAR(6) the variables in the first column are augmented with the first 3 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 2 principal components ($m = 7$). Fourth column: same as the third column, but 3 principal components ($m = \hat{r} = 8$).

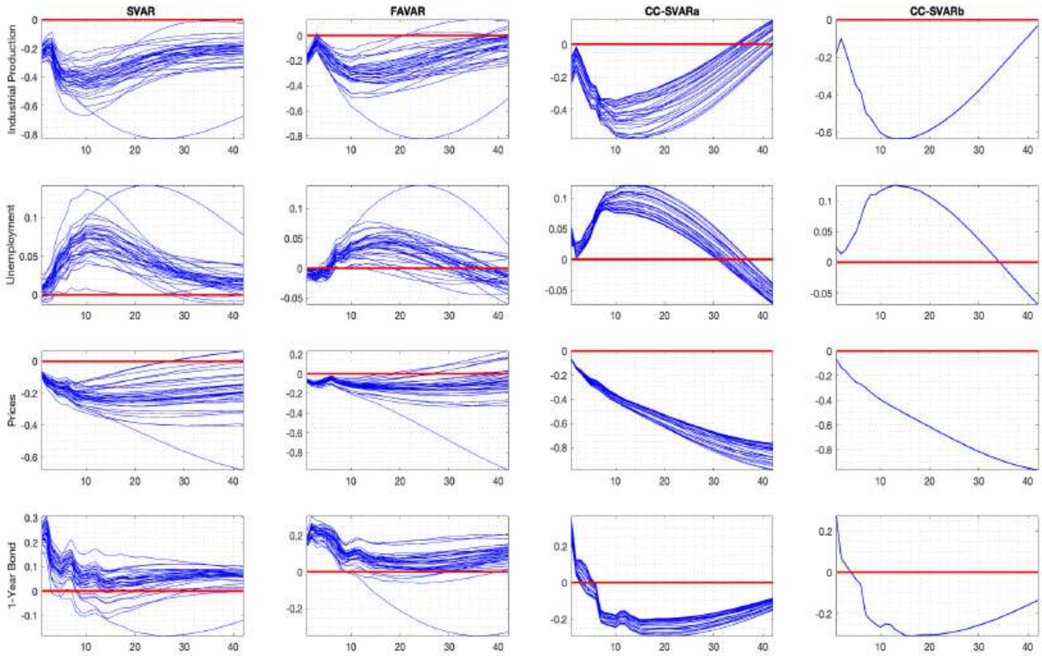

Figure 7: US monthly data. The IRFs of a monetary policy shock. Proxy MAR identification. The red lines are the CC-SVAR estimates obtained with different variable specifications. First column: SVAR(6) for 50 five-variable specifications, differing for the fifth variable. Second column: FAVAR(6) the variables in the first column are augmented with the first 3 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 2 principal components ($m = 7$). Fourth column: same as the third column, but 3 principal components ($m = \hat{r} = 8$).

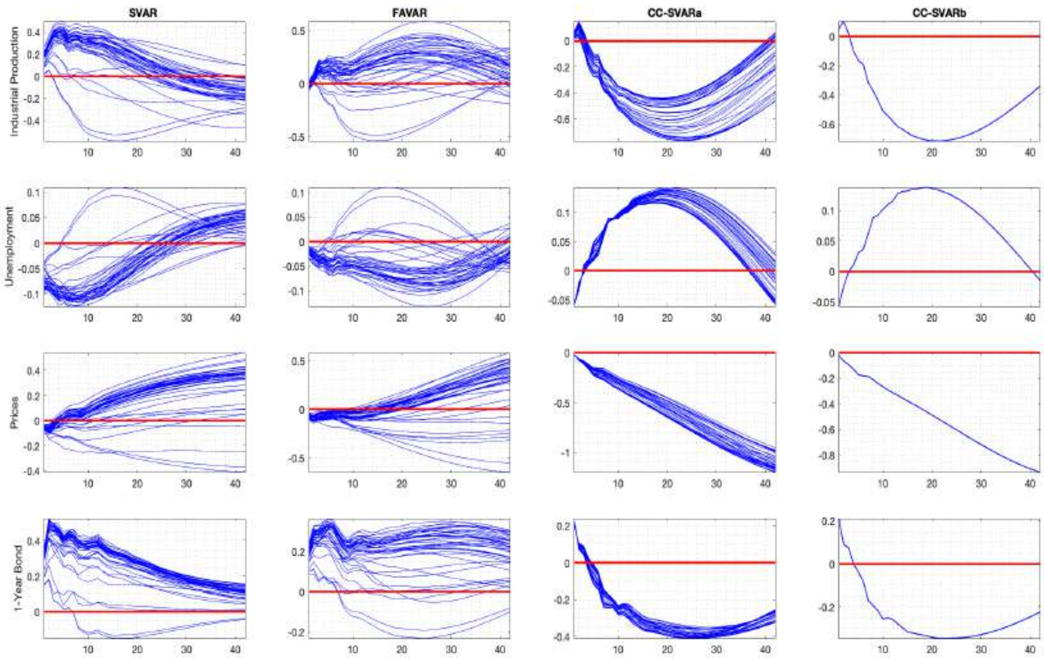

Figure 8: US monthly data. The IRFs of a monetary policy shock. Proxy GK identification. The red lines are the CC-SVAR estimates obtained with different variable specifications. First column: SVAR(6) for 50 five-variable specifications, differing for the fifth variable. Second column: FAVAR(6) the variables in the first column are augmented with the first 3 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 2 principal components ($m = 7$). Fourth column: same as the third column, but 3 principal components ($m = \hat{r} = 8$).

For Online Publication - Appendix

A Appendix to Section 3.2

A.1 Zerolessness of $K(L)$

We firstly need an explicit parameterization of the polynomial matrix $K(L)$ in Assumption 1. Let us write the entries of $K(L)$ as

$$k_{ij}(L) = k_{ij,0} + k_{ij,1}L + \cdots + k_{ij,s}L^s. \tag{A.1}$$

The number of coefficients is $\varpi = (s+1)mq$.

Assumption P. Parameterization of the polynomials in equation (A.1). *We suppose that the entries of $K(L)$ depend on ν parameters, where $\nu > 0$. Precisely, let $\mathcal{P}$, the parameter space, be an open and connected subset of $\mathbb{R}^\nu$. The ϖ coefficients $k_{ij,\alpha}$, for $i = 1, \ldots, m$, $j = 1, \ldots, q$, $\alpha = 0, \ldots, s$, are rational functions defined on $\mathcal{P}$, with no poles 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) Structural economic models, like (1), with the minor modification $\tau > 0$. As a rule, in this case $\nu < \varpi$, so that the parameterization produces restrictions on the coefficients $k_{i\ell,\beta}$.

(ii) The Free-Parameter case in which the parameters are the coefficients $k_{i\ell,\beta}$ themselves and $\mathcal{P} \subseteq \mathbb{R}^\varpi$.

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 $\mathcal{P}$.*

As we need explicit reference to the parameters $\mathbf{p}$, we use $K(\mathbf{p}, L)$, $k_{ij}(\mathbf{p}, L)$, etc. Let

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

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 in Definition P:

Proposition AD2. *Assume that Assumption 1 holds and $m > q$. 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 $\mathcal{W}$ as $\mathcal{P} - \mathcal{Z}$, i.e. 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, given $a \in \mathcal{M}$, either

(A) there exists an integer $d_a \geq 0$ such that generically $D_a(\mathbf{p}, L)$ has degree d_a with non-zero leading coefficient, or

(B) $D_a(\mathbf{p}, L)$ is the zero polynomial for all $\mathbf{p} \in \mathcal{P}$. In this case we set $d_a = -1$.

(ii) If $d_a = 0$ for some $a \in \mathcal{M}$, so that generically $D_a(\mathbf{p}, L)$ has no roots, then (Z) holds.

(iii) Because $K(L)$ is full rank, Assumption 1(b), $d_a > -1$ for some $a \in \mathcal{M}$.

(iv) If $d_a = -1$ for all but one $c \in \mathcal{M}$ with $d_c > 0$, then (W) holds.

(v) It remains to prove the proposition under the assumption that $d_a \neq 0$ for all $a \in \mathcal{M}$, so that (ii) does not apply, and that $d_a > 0$ for at least two distinct elements in $\mathcal{M}$, so that (iv) does not apply. Equivalently, we assume that $\{a \in \mathcal{M},\ \text{such that } d_a = 0\} = \emptyset$ and that the set

$$\mathcal{N} = \{a \in \mathcal{M},\ \text{such that } d_a > 0\} = \mathcal{M} - \{a \in \mathcal{M},\ \text{such that } d_a = -1\}$$

contains at least two distincts elements. 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 $D_c(\mathbf{p}, L)$ is not zero for all $c \in \mathcal{N}$. $\mathcal{P}^\dagger$ is open and dense in $\mathcal{P}$. Thus genericity in $\mathcal{P}^\dagger$ implies genericity in $\mathcal{P}$.

Let $R_{ab}(\mathbf{p})$ be the resultant of $D_a(\mathbf{p}, L)$ and $D_b(\mathbf{p}, L)$ and

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

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

(1) 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 $c, d \in \mathcal{N}$ and $\mathbf{p} \in \mathcal{P}^\dagger$. As each addendum in (A.2) is either zero or generically positive in $\mathcal{P}^\dagger$, by Proposition R, there exist $c^*, d^* \in \mathcal{N}$, $c^* \neq d^*$, such that, generically in $\mathcal{P}^\dagger$, $D_{c^*}(\mathbf{p}, L)$ and $D_{d^*}(\mathbf{p}, L)$ have no common roots, so that (Z) holds.

(2) $\mathcal{R}(\mathbf{p}) = 0$ for all $\mathbf{p} \in \mathcal{P}^\dagger$. By Proposition R, $D_c(\mathbf{p}, L)$ and $D_d(\mathbf{p}, L)$ have a common root for all $c, d \in \mathcal{N}$, $c \neq d$ and all $\mathbf{p} \in \mathcal{P}^\dagger$. Thus generically in $\mathcal{P}^\dagger$ (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

Here the motivation for $\kappa = 0$ given at the end of Section 3.2.2 is presented in greater detail. Consider a three-dimensional vector X_t with $I(1)$ coordinates, driven by the two-dimensional structural 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:

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

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

(A) If, for example, the variables 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}. 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 (A.3), 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 application in Section 5, we can safely assume that $K(L)$ has no zero at $z = 1$.

(C) Moreover, suppose, as we do 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 $\kappa = 0$.

(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.

The arguments in points (B) and (C) 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

$$(1-L)\tilde{X}_t = \begin{pmatrix} (1-L)X_t \\ (1-L)X_{m+1,t} \end{pmatrix} = \begin{pmatrix} K(L) \\ k_{m+1}(L) \end{pmatrix} u_t = \tilde{K}(L)u_t. \tag{A.4}$$

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 (A.4), 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

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

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 $m = 3$, $q = 1$, $B(L) = B_0 + B_1L + B_2L^2 + B_3L^3$, where the 12 entries in the matrices B_j can vary independently of one another. If we take $p = 1$ in (8), we have $(I - A_1L)(B_0 + B_1L + B_2L^2 + B_3L^3) = B_0$, that is

$$A_1B_0 = B_1, \quad A_1B_1 = B_2, \quad A_1B_2 = B_3, \quad A_1B_3 = 0. \tag{A.5}$$

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

$$B_3 = \alpha_0B_0 + \alpha_1B_1 + \alpha_2B_2.$$

Using (A.5),

$$0 = A_1 B_3 = A_1(\alpha_0 B_0 + \alpha_1 B_1 + \alpha_2 B_2) = \alpha_0 B_1 + \alpha_1 B_2 + \alpha_2 B_3$$
$$= \alpha_2 \alpha_0 B_0 + (\alpha_0 + \alpha_2 \alpha_1) B_1 + (\alpha_1 + \alpha_2^2) B_2,$$

which implies $\alpha_0 = \alpha_1 = \alpha_2 = 0$, i.e. $B_3 = 0$, which is not generic. In conclusion, generically χ_t has no VAR(1) representation. On the other hand, as argued in Section 3.4, $p > 1$ implies singularity of Z_{t-1}, i.e. non-uniqueness of $\mathcal{A}$ in (8).

B Proof of Proposition 1

B.1 Preliminary

The convergence of $\hat{v}_t$ to v_t may seem a trivial consequence of the continuity of the orthogonal projection. That is, convergence of $\hat{\chi}_t$ and $\hat{Z}_{t-1}$ to χ_t and Z_{t-1}, respectively, should imply convergence of $P(\hat{\chi}_t \,|\, \hat{Z}_{t-1})$ to $P(\chi_t \,|\, Z_{t-1})$ and of $\hat{v}_t = \hat{\chi}_t - P(\hat{\chi}_t \,|\, \hat{Z}_{t-1})$ to $v_t = \chi_t - P(\chi_t \,|\, Z_{t-1})$. However, while continuity of the orthogonal projection with respect to the regressand, given the regressors, is fairly obvious, continuity with respect to the regressors does not necessarily hold if the covariance matrix of the regressors tends to a singular matrix. An elementary example is the following. Let Y and X_k, $k \in \mathbb{N}$, be zero-mean stochastic variables with $\mathrm{E}(X_k^2) = 1$, and α_k a sequence of non-zero real numbers such that $\alpha_k \to 0$ as $k \to \infty$. Then

$$P(Y \,|\, \alpha_k X_k) = P(Y \,|\, X_k) = \mathrm{E}(Y X_k) X_k,$$

so that $\lim_{k\to\infty} P(Y \,|\, \alpha_k X_k) = 0$ if and only if $\lim_{k\to\infty} \mathrm{E}(Y X_k) = 0$. On the other hand,

$$P(Y \,|\, \lim_{k\to\infty} \alpha_k X_k) = P(Y \,|\, 0) = 0.$$

The proof below shows that *the assumptions of Proposition 1* ensure convergence of the projection $P(\hat{\chi}_t \,|\, \hat{Z}_{t-1})$ to $P(\chi_t \,|\, Z_{t-1})$ even when the covariance matrix of $\hat{Z}_{t-1}$ tends to singularity.

B.2 Proof

Let us denote by d the rank of Σ_0^Z and partition Z_t (possibly after reordering) as $Z_t = (\Omega_t' \;\; S_t')'$, where $\det\left(\Sigma_0^\Omega\right) \neq 0$. 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 projection equation (8) as

$$\chi_t = \alpha\Omega_{t-1} + v_t = P(\chi_t|Z_{t-1}) + v_t, \tag{B.1}$$

where P denotes the population projection and $\alpha = \mathcal{A}M$ is unique.

The empirical counterpart of the above equation is given by the regression equation (9), i.e.

$$\hat{\chi}_t = \hat{\mathcal{A}}\hat{Z}_{t-1} + \hat{v}_t = \hat{P}(\hat{\chi}_t|\hat{Z}_{t-1}) + \hat{v}_t,$$

where $\hat{P}$ denotes the sample projection.

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 consider the sample regression equation

$$\hat{S}_t = \hat{P}(\hat{S}_t|\hat{\Omega}_t) + \hat{\vartheta}_t = \hat{N}\hat{\Omega}_t + \hat{\vartheta}_t, \tag{B.2}$$

where $\hat{\Sigma}_0^{\hat{\vartheta}\hat{\Omega}} = 0$. Let us write $\hat{\vartheta}_t$ as $\hat{\vartheta}_t = H\tilde{\vartheta}_t$, where H is $(mp-d) \times \tilde{d}$, $\tilde{d} \leq mp-d$, and $\tilde{\vartheta}_t$ is standardized by imposing

$$(T-1)^{-1}\sum_{t=1}^{T-1}\tilde{\vartheta}_t\tilde{\vartheta}_t' = I_{\tilde{d}}. \tag{B.3}$$

Note that, since $\hat{\vartheta}_t$ depends on n and T, H and $\tilde{d}$ depend on n and T as well. The vectors $\hat{\Omega}_t$ and $\tilde{\vartheta}_t$ are sample orthogonal, i.e. $\hat{\Sigma}_0^{\hat{\Omega}\tilde{\vartheta}} = 0$, see (B.2). Moreover, they 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.

$$\hat{\chi}_t = \hat{\mathcal{A}}\hat{Z}_{t-1} + \hat{v}_t = \hat{\alpha}\hat{\Omega}_{t-1} + \hat{\beta}\tilde{\vartheta}_{t-1} + \hat{v}_t, \tag{B.4}$$

where $\hat{\Sigma}_1^{\hat{v}\hat{\Omega}} = 0$ and $\hat{\Sigma}_1^{\hat{v}\tilde{\vartheta}} = 0$, so that, defining

$$\hat{\Sigma}_{0,T-1}^{\hat{\Omega}} = (T-1)^{-1} \sum_{t=1}^{T-1} \tilde{\Omega}_t \tilde{\Omega}_t',$$

we have $\hat{\alpha}\hat{\Sigma}_{0,T-1}^{\hat{\Omega}} = \hat{\Sigma}_1^{\hat{\chi}\hat{\Omega}}$ and $\hat{\beta} = \hat{\Sigma}_1^{\hat{\chi}\tilde{\vartheta}}$. Equation (B.4) is the sample analogue of (B.1). Subtracting (B.1) from (B.4) we get

$$\hat{\chi}_t - \chi_t = \hat{\pi}_t = (\hat{\alpha}\hat{\Omega}_{t-1} - \alpha\Omega_{t-1}) + \hat{\beta}\tilde{\vartheta}_{t-1} + (v_t - \hat{v}_t). \tag{B.5}$$

Since the left-hand side is $O_p(r_{n,T})$ by Assumption 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 1.

(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

$$\hat{\alpha} - \alpha = \left[\left(\hat{\alpha}\hat{\Sigma}_{0,T-1}^{\hat{\Omega}} - \alpha\Sigma_0^{\Omega}\right) - \hat{\alpha}\left(\hat{\Sigma}_{0,T-1}^{\hat{\Omega}} - \Sigma_0^{\Omega}\right)\right]\left(\Sigma_0^{\Omega}\right)^{-1}. \tag{B.6}$$

Now consider the first term of the difference in square brackets. Using (B.1) and (B.4), we get $\hat{\alpha}\hat{\Sigma}^{\hat{\Omega}}_{0,T-1} - \alpha\Sigma^{\Omega}_0 = \hat{\Sigma}^{\hat{\chi}\hat{\Omega}}_1 - \Sigma^{\chi\Omega}_1 = \left(\hat{\Sigma}^{\chi\Omega}_1 - \Sigma^{\chi\Omega}_1\right) + \hat{\Sigma}^{\hat{\pi}\Omega}_1 + \hat{\Sigma}^{\hat{\chi}\hat{\nu}}_1$. Assumption B implies that $||\hat{\Sigma}^{\chi\Omega}_1 - \Sigma^{\chi\Omega}_1|| = O_p(1/\sqrt{T})$, while $||\hat{\Sigma}^{\hat{\pi}\Omega}_1 + \hat{\Sigma}^{\hat{\chi}\hat{\nu}}_1||$ is $O_p(r_{n,T})$ by Assumption A. Turning to the second term, we have $\hat{\Sigma}^{\hat{\Omega}}_{0,T-1} - \Sigma^{\Omega}_0 = \left(\hat{\Sigma}^{\Omega}_{0,T-1} - \Sigma^{\Omega}_0\right) + \hat{\Sigma}^{\hat{\nu}\Omega}_{0,T-1} + \hat{\Sigma}^{\hat{\Omega}\hat{\nu}}_{0,T-1}$. Assumption B implies that $||\hat{\Sigma}^{\Omega}_{0,T-1} - \Sigma^{\Omega}_0|| = O_p(1/\sqrt{T})$, while $||\hat{\Sigma}^{\hat{\nu}\Omega}_{0,T-1} + \hat{\Sigma}^{\hat{\Omega}\hat{\nu}}_{0,T-1}||$ is $O_p(r_{n,T})$ by Assumption A. Since $||\hat{\alpha}||$ is $O_p(1)$, the norm of the factor in square brackets of (B.6) is $O_p\left(\max(r_{n,T}, 1/\sqrt{T})\right)$. Since $||(\Sigma^{\Omega}_0)^{-1}|| = O(1)$, (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 the norm of the first term is $O_p(r_{n,T})$. 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 (7). But u_t is independent of $\hat{Z}_{t-1}$, and therefore $\tilde{\vartheta}_{t-1}$, by Assumption A. Now, let us consider the h-th row of the matrix $\hat{\Sigma}^{v\tilde{\vartheta}}_1$, i.e. $\hat{\Sigma}^{v_h\tilde{\vartheta}}_1 = \sum_{t=2}^{T} v_{ht}\tilde{\vartheta}'_{t-1}/(T-1)$. Let Ψ^h_1 be its population covariance matrix. As $\mathrm{E}\left(\hat{\Sigma}^{v_h\tilde{\vartheta}}_1\right) = 0$, we have

$$\Psi^h_1 = \sum_{t=2}^{T}\sum_{\tau=2}^{T} \mathrm{E}(v_{ht}\tilde{\vartheta}_{t-1}\tilde{\vartheta}'_{\tau-1}v_{h\tau})/(T-1)^2.$$

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

$$\Psi^h_1 = \sum_{t=2}^{T}\sum_{\tau=2}^{T} \mathrm{E}(v_{ht}v_{h\tau})\mathrm{E}(\tilde{\vartheta}_{t-1}\tilde{\vartheta}'_{\tau-1})/(T-1)^2.$$

But $\mathrm{E}(v_{ht}v_{h\tau}) = 0$ for $t \neq \tau$, because of serial independence of u_t, Assumption 1(a), so that $\Psi^h_1 = \sum_{t=2}^{T}\mathrm{E}(v^2_{ht})\mathrm{E}(\tilde{\vartheta}_{t-1}\tilde{\vartheta}'_{t-1})/(T-1)^2$. Covariance stationarity of v_{ht} and (B.3) imply that

$$\Psi^h_1 = \mathrm{E}(v^2_{ht})\ \mathrm{E}\left(\sum_{t=2}^{T}(\tilde{\vartheta}_{t-1}\tilde{\vartheta}'_{t-1})/(T-1)\right) = I_{\tilde{d}}\ \mathrm{E}(v^2_{ht})/(T-1),$$

so that, by Chebyshev's inequality, each entry of $\hat{\Sigma}_1^{v_h\tilde{\vartheta}}$ is $O_p(1/\sqrt{T})$ for all h. (iii) follows.

(iv). We have $\hat{\beta} = \hat{\Sigma}_1^{\hat{\chi}\tilde{\vartheta}} = \hat{\Sigma}_1^{\chi\tilde{\vartheta}} + \hat{\Sigma}_1^{\tilde{\pi}\tilde{\vartheta}} = \alpha\hat{\Sigma}_0^{\Omega\tilde{\vartheta}} + \hat{\Sigma}_1^{v\tilde{\vartheta}} + \hat{\Sigma}_1^{\tilde{\pi}\tilde{\vartheta}}$. 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 $\hat{\beta} = -\alpha\hat{\Sigma}_0^{\hat{\nu}\tilde{\vartheta}} + \hat{\Sigma}_1^{v\tilde{\vartheta}} + \hat{\Sigma}_1^{\tilde{\pi}\tilde{\vartheta}}$. The norms of both the first and the third term are $O_p(r_{n,T})$ by Assumption A. The norm of the second term is $O_p(1/\sqrt{T})$ by (iii), hence $||\hat{\beta}|| = O_p\left(\max(r_{n,T}, 1/\sqrt{T})\right)$. Since $\tilde{\vartheta}_t$ is $O_p(1)$, (iv) is proved. Q.E.D.

Proposition 1 follows from equation (B.5), Lemma 1 (ii) and Lemma 1 (iv).

C Proof of Proposition 2

Lemma 2. *We have:*

(i) $||\hat{\Sigma}_{[11]} - \Sigma_{[11]}|| = O_p\left(\max(r_{n,T}, 1/\sqrt{T})\right)$, *where* $\Sigma_{[11]}$ *has been defined in* (11)*;*

(ii) $||\hat{Q} - Q|| = O_p\left(\max(r_{n,T}, 1/\sqrt{T})\right)$*;*

(iii) $||\hat{Q}^{-1} - Q^{-1}|| = O_p\left(\max(r_{n,T}, 1/\sqrt{T})\right)$.

Proof. Let $\hat{\psi}_t = \hat{v}_t - v_t$. We have $\hat{\Sigma}_0^{\hat{v}} - \Sigma_0^v = \hat{\Sigma}_0^{\hat{\psi}v} + \hat{\Sigma}_0^{v\hat{\psi}} + \hat{\Sigma}_0^{\hat{\psi}} + (\hat{\Sigma}_0^v - \Sigma_0^v)$. 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. 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), notice that the entries of $\hat{Q}$ and Q are the same elementary differentiable functions of the entries of $\hat{\Sigma}_{[11]}$ and $\Sigma_{[11]}$, respectively. As the denominators are bounded away from zero in probability, result (ii) follows from (i). Since $\det \hat{Q}$ is bounded away from zero in probability, (iii) is an immediate consequence of (ii). Q.E.D.

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

Proof. We have $\hat{u}_t = \hat{Q}^{-1}\hat{v}_t^{[1]}$ and $u_t = Q^{-1}v_t^{[1]}$. Hence $\hat{u}_t - u_t = \hat{Q}^{-1}(\hat{v}_t^{[1]} - v_t^{[1]}) + (\hat{Q}^{-1} - Q^{-1})v_t^{[1]}$. 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 2 (iii). Q.E.D.

Lemma 3. *The following results hold:*

(i) $||\hat{B}_0 - B_0|| = O_p\left(\max(r_{n,T}, 1/\sqrt{T})\right)$;

(ii) Let $\hat{\epsilon}_t = \hat{v}_t^{[2]} - \hat{R}\hat{u}_t$, *where* $\hat{R}$ *is defined in* (12). *Then,* $||\hat{\epsilon}_t|| = O_p\left(\max(r_{n,T}, 1/\sqrt{T})\right)$.

Proof. We have already shown, Lemma 2(ii), that $||\hat{Q} - Q|| = O_p\left(\max(r_{n,T}, 1/\sqrt{T})\right)$. Let us now show that $||\hat{R} - R|| = O_p\left(\max(r_{n,T}, 1/\sqrt{T})\right)$. We have $\hat{R} = \hat{\Sigma}_{[21]}(\hat{Q}')^{-1}$ and $R = \Sigma_{[21]}(Q')^{-1}$. Hence $\hat{R} - R = \hat{\Sigma}_{[21]}\left((\hat{Q}')^{-1} - (Q')^{-1}\right) + (\hat{\Sigma}_{[21]} - \Sigma_{[21]})(Q')^{-1}$. The norm of first term is $O_p\left(\max(r_{n,T}, 1/\sqrt{T})\right)$ by Lemma 2 (iii). Moreover, in the proof of Lemma 2 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 $||\hat{\Sigma}_{[21]} - \Sigma_{[21]}|| = O_p\left(\max(r_{n,T}, 1/\sqrt{T})\right)$. As for (ii), we have

$$\hat{v}_t^{[2]} - v_t^{[2]} = (\hat{R}\hat{u}_t - Ru_t) + \hat{\epsilon}_t = (\hat{R} - R)u_t + \hat{R}(\hat{u}_t - u_t) + \hat{\epsilon}_t.$$

The norm of the left side is $O_p\left(\max(r_{n,T}, 1/\sqrt{T})\right)$ by Proposition 1; the norm of the second term on the right side is $O_p\left(\max(r_{n,T}, 1/\sqrt{T})\right)$ by Proposition 2(a); the norm of the term term on the right side is $O_p\left(\max(r_{n,T}, 1/\sqrt{T})\right)$ by result (i). Hence $||\hat{\epsilon}_t|| = O_p\left(\max(r_{n,T}, 1/\sqrt{T})\right)$. Q.E.D.

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

$$\hat{Z}_t = \hat{D}\hat{Z}_{t-1} + \hat{\zeta}_t, \tag{C.1}$$

where

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

From (C.1), by recursion we get

$$\hat{Z}_t = \hat{D}^{k+1}\hat{Z}_{t-k-1} + \sum_{j=0}^{k} \hat{D}^j \hat{\zeta}_{t-j}, \tag{C.2}$$

for any $k \geq 0$. By taking the first m rows of (C.2) we get

$$\hat{\chi}_t = \hat{G}_{k+1}\hat{Z}_{t-k-1} + \sum_{j=0}^{k} \hat{V}_j \hat{v}_{t-j} = \hat{G}_{k+1}\hat{Z}_{t-k-1} + \sum_{j=0}^{k} \hat{V}_j \hat{B}_0 \hat{u}_{t-j} + \sum_{j=0}^{k} \hat{V}_j \begin{pmatrix} 0 \\ \hat{\epsilon}_{t-j} \end{pmatrix}, \tag{C.3}$$

where $\hat{G}_k$ is the matrix formed by the first m rows of $\hat{D}^k$ and $\hat{V}_j$ 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, evaluating (C.3) for $k-1$ and subtracting from (C.3), we get

$$\hat{G}_k \hat{Z}_{t-k} = \hat{G}_{k+1}\hat{Z}_{t-k-1} + \hat{B}_k \hat{u}_{t-k} + \hat{V}_k \begin{pmatrix} 0 \\ \hat{\epsilon}_{t-k} \end{pmatrix}, \tag{C.4}$$

which, letting $\hat{G}_0 = (I_m \;\; 0)$, 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 (8) we get

$$\chi_t = G_{k+1} Z_{t-k-1} + \sum_{j=0}^{k} V_j v_{t-j} = G_{k+1} Z_{t-k-1} + \sum_{j=0}^{k} V_j B_0 u_{t-j} \tag{C.5}$$

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, $j = 1, \ldots, k$, 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 $G_{k+1} Z_{t-k-1} = G_{k+1} M \Omega_{t-k-1}$ is unique, so that $G_{k+1}M$ is also unique for any k. From (C.5) we get

$$G_k Z_{t-k} = G_{k+1} Z_{t-k-1} - B_k u_{t-k}. \tag{C.6}$$

Lemma 4. 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) $\left|\left|\hat{V}_k \begin{pmatrix} 0 \\ \hat{\epsilon}_{t-k} \end{pmatrix}\right|\right| = O_p\left(\max(r_{n,T}, 1/\sqrt{T})\right)$;

(iii) $||\hat{B}_k - B_k|| = O_p\left(\max(r_{n,T}, 1/\sqrt{T})\right)$, which is Proposition 2(b).

Proof. We proceed by induction on k. 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. Moreover, (ii) holds by Lemma 3(ii) and (iii) holds by Lemma 3(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 (C.6) from (C.4) we get

$$\hat{G}_{\bar{k}} \hat{Z}_{t-\bar{k}} - G_{\bar{k}} Z_{t-\bar{k}} = (\hat{G}_{\bar{k}+1} \hat{Z}_{t-(\bar{k}+1)} - G_{\bar{k}+1} Z_{t-(\bar{k}+1)}) - (\hat{B}_{\bar{k}} \hat{u}_{t-\bar{k}} - B_{\bar{k}} u_{t-\bar{k}}) - \hat{V}_{\bar{k}} \begin{pmatrix} 0 \\ \hat{\epsilon}_{t-\bar{k}} \end{pmatrix}. \tag{C.7}$$

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 (C.7), postmultiply by $\hat{\epsilon}'_{t-(\bar{k}+1)}$ and average over $t = k + 2, \ldots, T + k + 1$. Using sample orthogonality of $\hat{\epsilon}_{t-(\bar{k}+1)}$ with both $\hat{u}_{t-(\bar{k}+1)}$ and $\hat{Z}_{t-(\bar{k}+2)}$ we get

$$\hat{G}_{\bar{k}+1} \hat{\Sigma}_0^{\hat{Z}\hat{\epsilon}} - G_{\bar{k}+1} \hat{\Sigma}_0^{Z\hat{\epsilon}} = (\hat{G}_{\bar{k}+2} \hat{Z}_0 - G_{\bar{k}+2} Z_0) \hat{\epsilon}'_0 / T - G_{\bar{k}+2} \hat{\Sigma}_{-1}^{Z\hat{\epsilon}} + B_{\bar{k}+1} \hat{\Sigma}_0^{u\hat{\epsilon}} - \hat{V}_{\bar{k}+1} \begin{pmatrix} 0 \\ \hat{\Sigma}_0^{\hat{\epsilon}} \end{pmatrix}.$$

The norm of the left side is $O_p\left(\max(r_{n,T}, 1/\sqrt{T})\right)$ because, as proved above, (i) holds for $k = \bar{k}+1$. Let us now consider the first term on the right side. Going back to (C.1), we see that $\hat{D}^k\hat{Z}_{t-k} = \hat{D}^{k+1}\hat{Z}_{t-k-1} + \hat{D}^k\hat{\zeta}_{t-k}$, where the terms on the right side are sample orthogonal and the term on the left side is bounded in probability for $k = 0$. Hence $||\hat{D}^k\hat{Z}_{t-k}||$ is $O_p(1)$ for any k and therefore $||\hat{G}_{\bar{k}+2}\hat{Z}_0||$ is $O_p(1)$. Of course, the same holds for $G_{\bar{k}+2}Z_0$ and $\hat{\epsilon}_0$, so that the norm of the first term on the right side is $O_p(1/T)$. Coming to the second term, let us observe that it is equal to $G_{\bar{k}+2}M\hat{\Sigma}^{\hat{\nu}\hat{\epsilon}}_{-1}$, since $Z_t = M\Omega_t$, see (B.1), $\Omega_t = \hat{\Omega}_t - \hat{\nu}_t$ and $\hat{\epsilon}_{t-(\bar{k}+1)}$ is sample orthogonal to $\hat{\Omega}_{t-(k+2)}$. Its norm is then $O_p(r_{n,T})$ since so is the norm of $\hat{\nu}_t$ by Assumption A, and the norm of $G_{\bar{k}+2}M$, which, as observed above, is unique, is $O(1)$. Letting $\hat{\gamma}_t = \hat{u}_t - u_t$, using sample orthogonality of $\hat{\epsilon}_{t-(\bar{k}+1)}$ with $\hat{u}_{t-(\bar{k}+1)}$, the third term on the right side is equal to $-B_{\bar{k}+1}\hat{\Sigma}^{\hat{\gamma}\hat{\epsilon}}_0$, 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 fourth 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 (C.7) and postmultiply by $\hat{u}'_{t-(\bar{k}+1)}$ and average over $t = k+2, \ldots, T+k+1$. Using sample orthogonality of $\hat{u}_{t-(\bar{k}+1)}$ with both $\hat{\epsilon}_{t-(\bar{k}+1)}$ and $\hat{Z}_{t-(\bar{k}+2)}$ we get

$$\hat{G}_{\bar{k}+1}\hat{\Sigma}^{\hat{Z}\hat{u}}_0 - G_{\bar{k}+1}\hat{\Sigma}^{Z\hat{u}}_0 = (\hat{G}_{\bar{k}+2}\hat{Z}_0 - G_{\bar{k}+2}Z_0)\hat{u}_0/T - G_{\bar{k}+2}\hat{\Sigma}^{Z\hat{u}}_{-1} - (\hat{B}_{\bar{k}+1} - B_{\bar{k}+1}) - B_{\bar{k}+1}\Sigma^{u\hat{\gamma}}_0.$$

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 $O_p(1/T)$ for the same argument used above. The norm of the second term is $O_p\left(\max(r_{n,T}, 1/\sqrt{T})\right)$ for the same argument used above for $-G_{\bar{k}+2}\hat{\Sigma}^{Z\hat{\epsilon}}_{-1}$. The norm of the fourth 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.

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,

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

Lemma 5. *(Consistency of the covariance matrices). Let, as in Definition 2,* $\mathcal{I}_m$ *be 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

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

By Assumption 7(a), we have $\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n \mathrm{E}(\hat{\gamma}^x_{k,ij}-\gamma^x_{k,ij})^2 < \frac{\rho}{T}$ 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 (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 $\frac{1}{n}\|\hat{\Gamma}^x_k-\Gamma^\chi_k\| \leq \frac{1}{n}\|\hat{\Gamma}^x_k-\Gamma^x_k\|+\frac{1}{n}\|\Gamma^\xi_k\|$. 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 6. *(Consistency of the normalized eigenvalues). Let M^χ and $\hat{M}^x$ be the $r\times r$ 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 $n\geq\bar{n}$ and $\eta>0$, there exists $\tau(\eta,n)$ such that, for $T\geq\tau(\eta,n)$, $\frac{\hat{M}^x}{n}$ 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 (D.1) we get $\frac{1}{n}|\hat{\mu}^x_j-\mu^x_j| \leq n^{-1}\|\hat{\Gamma}^x_0-\Gamma^x_0\|$, which is $O_p\left(1/\sqrt{T}\right)$ by Lemma 5(i). This proves (i). Setting $A=\Gamma^\chi_0$, $B=\hat{\Gamma}^x_0-\Gamma^\chi_0$ and applying again (D.1) 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 5(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 (D.1), $\mu_r^x \geq \mu_r^\chi$. 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 $\left\|\left(\frac{\hat{M}^x}{n}\right)^{-1}\right\|$ follows from statements (ii) and (iii). This concludes the proof. Q.E.D.

Lemma 7. *Let W^χ be the $n \times r$ matrix having on column j, $j = 1, \ldots, r$, 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 $\zeta = \left\|\mathcal{I}_m' W^\chi (M^\chi)^{1/2}\right\| = \|\mathcal{I}_m'\Gamma_0^\chi\mathcal{I}_m\|^{1/2} = \|\Sigma_0^\chi\|^{1/2}$ does not depend on n. We have

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

which is $O(1)$ by Lemma 6(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 5(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 6(v), the second by Lemma 5(vi) and Lemma 6(v). Q.E.D

Lemma 8. *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 normalized 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 7(ii) we get

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

$i \leq r$, $j \leq r$. 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 6(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 7(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

$a = \sqrt{n}\mathcal{I}_m' W^{\chi} \hat{\mathcal{J}}_r$,

$b = \sqrt{n}\mathcal{I}_m' W^{\chi} W^{\chi\prime} \hat{W}^x = \sqrt{n}\mathcal{I}_m' W^{\chi} W^{\chi\prime} \hat{W}^x \frac{\hat{M}^x}{n} \left(\frac{\hat{M}^x}{n}\right)^{-1}$,

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

Notice that $\|\sqrt{n}\mathcal{I}_m' W^\chi\|$ is $O(1)$ by Lemma 7(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 7(ii) and 6(v) to get $\|b-c\| = O_p\left(\max\left(\frac{1}{n}, \frac{1}{\sqrt{T}}\right)\right)$. Finally, Lemmas 5(vii) and 6(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 9. *(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 8(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 7(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 8(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.

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

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

Proof. Notice first that statement (b) has already be proved, see Lemma 5(v). Regarding (a), 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

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

Regarding a, we have $\|a\| \leq \|\sqrt{n}(\mathcal{I}'_m \hat{W}^x \hat{W}^{x\prime} - \mathcal{I}'_m W^{\chi} W^{\chi\prime})\| \, \|\boldsymbol{x}_{nt}/\sqrt{n}\|$. Now, $\|\boldsymbol{x}_{nt}/\sqrt{n}\|^2 = \sum_{i=1}^n x_{it}^2/n$ is $O_p(1)$, since its expected value is

$$(\text{trace}\, \Gamma_0^x)/n = (\text{trace}\, \Gamma_0^{\chi})/n \; + \; (\text{trace}\, \Gamma_0^{\xi})/n \leq \sum_{j=1}^r \mu_j^{\chi}/n + \mu_1^{\xi},$$

which is bounded by Assumption 6. Hence a is $O_p\left(\max\left(1/\sqrt{n}, 1/\sqrt{T}\right)\right)$ by Lemma 9(ii). As for b, we have $\|\mathcal{I}'_m W^{\chi} W^{\chi\prime}\boldsymbol{\xi}_{nt}\| \leq \|\mathcal{I}'_m W^{\chi}\| \, \|W^{\chi\prime}\boldsymbol{\xi}_{nt}\|$. The first factor is $O(1/\sqrt{n})$ by Lemma 7(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 $\mu_1^{\xi} \leq \ell$ (see Assumption 6(b)). Hence $\|b\| = O(1/\sqrt{n})$. Statement (a) follows. Q.E.D.

E Difficulties 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:

$$\chi_{1t} = u_{t-1}$$

$$\chi_{2t} = a_2 u_t + u_{t-1}.$$

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.2, problematic cases no longer occur when m is larger than $q + 1$.

F Simulation details and additional simulation results

F.1 The factor model used for the simulations

Here we describe the factor model used for Simulations 2 and 3 of Section 4 and the additional simulation described below. Firstly we rewrite model (1) 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:

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

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

$$x_t = \Lambda F_t + \xi_t \tag{F.2}$$

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

$$z_t = \Lambda^z F_t + \xi_t^z \tag{F.3}$$

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).

F.2 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 9 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 differ-

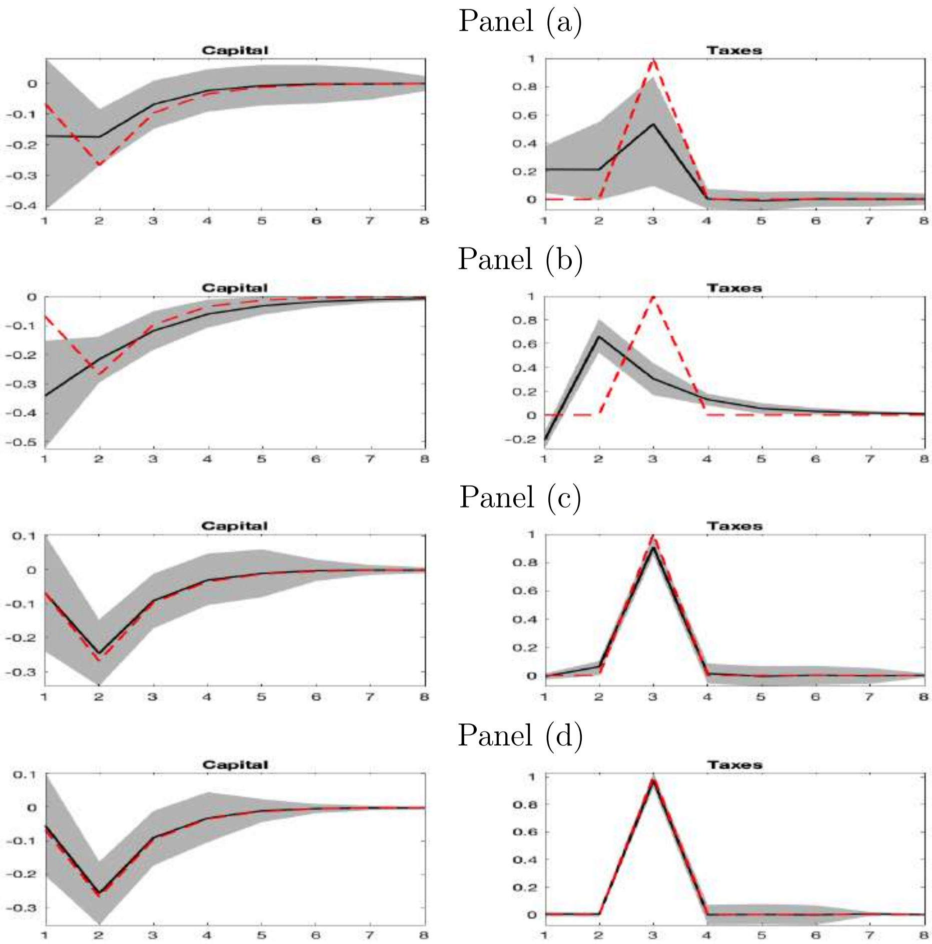

Figure 9: Simulation 4. The choice of m. Estimated IRFs for the tax shock. The red dashed lines are the theoretical IRFs. The solid lines represent the mean (across 1000 simulated datasets) of the point estimates. The grey areas represent the 16th and 84th percentiles of the point estimate distribution. 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 ($m = 5$).

ence 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 (F.1). 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 10, 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 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

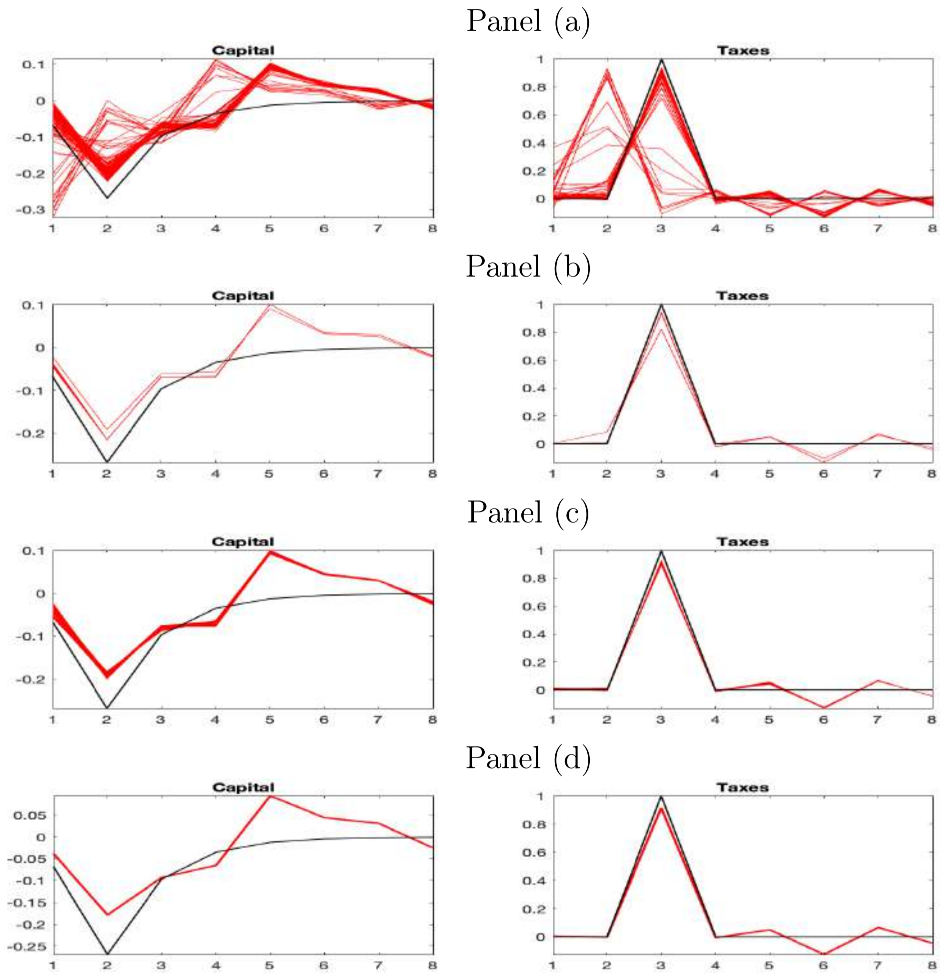

Figure 10: Simulation 5. The choice of ψ with $m < r$ and $m = r$. Estimated IRFs for the tax shock, for a single simulated data set. The black lines are the theoretical IRFs. The red lines are the CC-SVAR estimates obtained with different variable specifications. Panel (a): CC-SVAR(4) with Capital, 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 ($m = 5$).

occur for this data set.[12]

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.

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

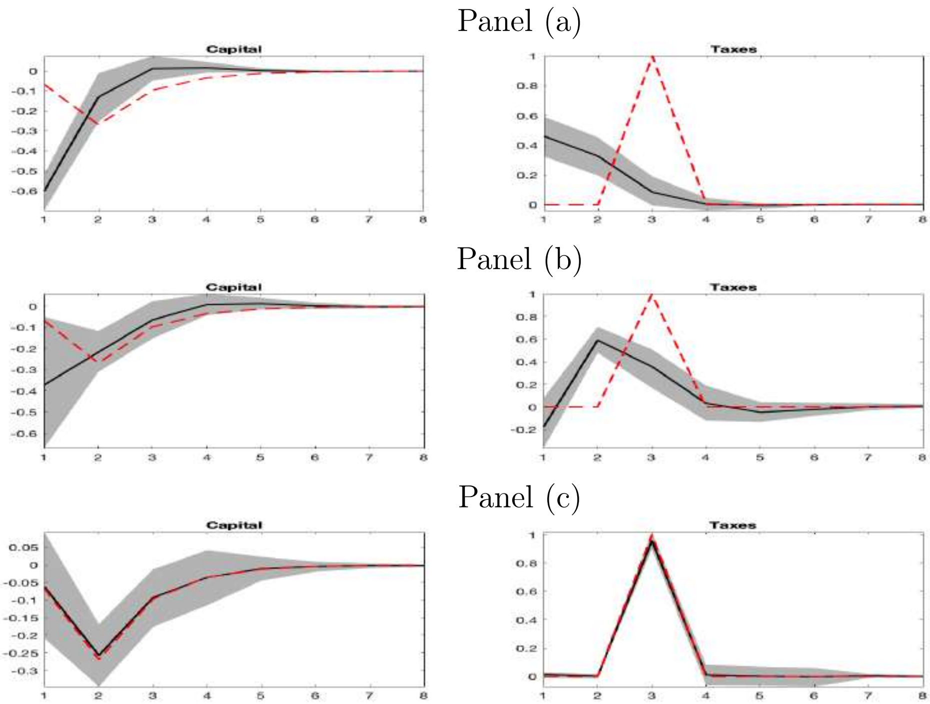

Figure 11: Simulation 6. The choice of $\hat{r}$. Results for $m = \hat{r} < r$ and $m = \hat{r} > r$. Estimated IRFs for the tax shock. The red dashed lines are the theoretical IRFs. The solid lines represent the mean (across 1000 simulated datasets) of the point estimates. The grey areas represent 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).

F.3 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 11 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.4 Cointegration

In Simulation 7 we show results about cointegration. The model of equation (1) is modified in such a way to have cointegration. We assume now that technology 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

$$\begin{pmatrix} \Delta a_t \\ \Delta k_t \\ \tau_t \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ \frac{-\delta(1-L)}{1-\alpha L} & \frac{1}{1-\alpha L} \\ L & 0 \end{pmatrix} \begin{pmatrix} u_{\tau,t} \\ u_{a,t} \end{pmatrix} = B(L)u_t. \tag{F.4}$$

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 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 (F.4) 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 vari-

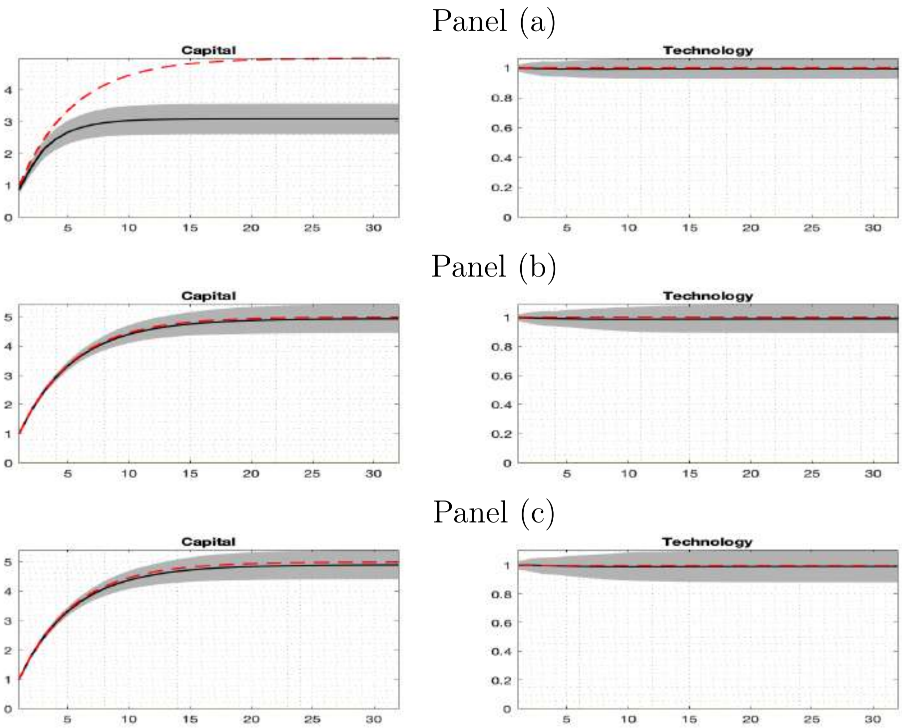

Figure 12: Simulation 7. Cointegration. Estimated IRFs for the technology shock. The red dashed lines are the theoretical IRFs. The solid lines represent the mean (across 1000 simulated datasets) of the point estimates. The grey areas represent the 16th and 84th percentiles of the point estimate distribution. 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.

ance 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 12. 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 Empirical application: robustness

To assess the robustness of the results to changes of the number of factors, we 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 13. We see that the results obtained with different values of $\hat{r}$ are very similar to each other for all identification schemes.

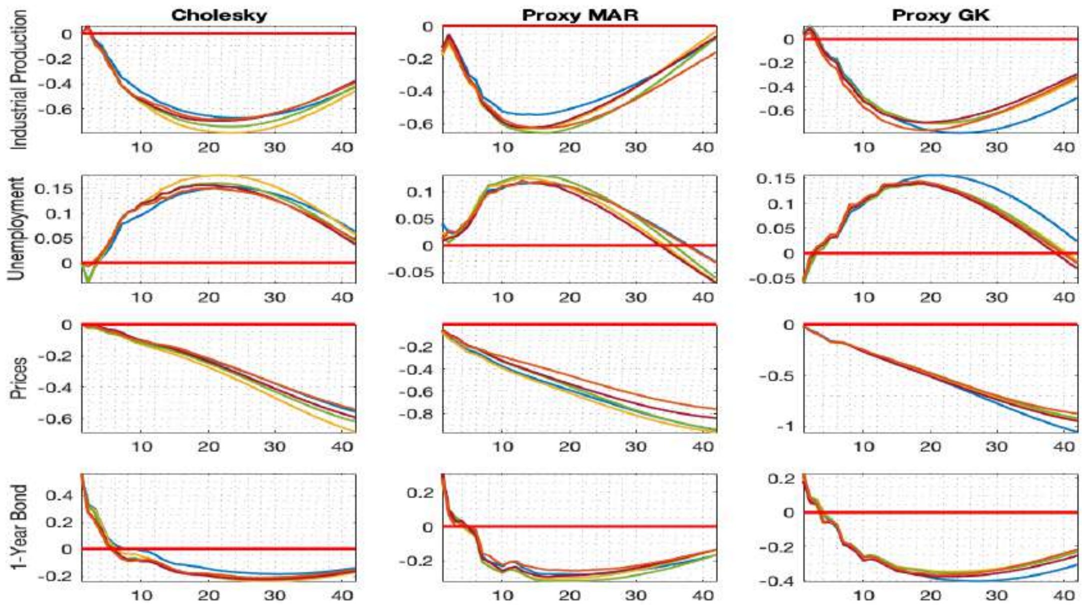

Figure 13: US monthly data. The IRFs of a monetary policy shock. CC-SVAR(6) with $m = r$, using different values of r. Black dotted line: $r = 6$. Blue dashed line: $r = 8$. Red solid line: $r = 10$.