AN IDENTIFICATION STRATEGY FOR PROXY-SVARs WITH WEAK PROXIES

Giovanni Angelini[a], Giuseppe Cavaliere[a,b], Luca Fanelli[a]

September 2021

Abstract

When proxies (external instruments) used to identify target structural shocks are weak, inference in proxy-SVARs (SVAR-IVs) is nonstandard and requires the use of identification-robust methods to construct asymptotically valid confidence sets for the impulse responses of interest. In the presence of multiple target shocks, test inversion methods may require a large number of restrictions on the proxy-SVAR parameters other than the proxies and may therefore become far from trivial to apply. We show that asymptotic inference in these situations can be alternatively conducted through standard methods if there exist strong proxies for the *auxiliary shocks*, i.e. the shocks which are not of primary interest in the analysis. We design a frequentist identification strategy where the impulse responses associated with the target shocks are (point-)identified by instrumenting the auxiliary shocks, not the target shocks. To do so, we exploit the relationship that characterizes the 'B-form' and 'A-form' representations of the proxy-SVAR and develop a Minimum Distance (MD) estimation method based on the latter. The suggested identification strategy is complemented with a novel diagnostic test for instrument relevance based on bootstrap resampling. The test does not give rise to pre-testing issues, is robust to conditionally heteroskedastic VAR disturbances and proxies, and is computationally straightforward regardless on the number of shocks being instrumented. Some illustrative examples show the usefulness of the suggested approach.

Keywords: Bootstrap inference, external instruments, identification, Proxy-SVAR, oil supply shock.

JEL Classification: C32, C51, E44, E62

[a]Department of Economics, University of Bologna, Italy. [b]Department of Economics, University of Exeter Business School, UK. Correspondence to: Giuseppe Cavaliere, Department of Economics, University of Bologna, Piazza Scaravilli 2, 40126 Bologna, Italy; email: giuseppe.cavaliere@unibo.it.

1 Introduction

Proxy-SVARs, or SVAR-IVs, popularized by Stock (2008), Stock and Watson (2012), Mertens and Ravn (2013) and Stock and Watson (2018), have become standard tools to track the dynamic causal effects of macroeconomic shocks on variables of interest. In proxy-SVARs, the model is complemented with 'external' variables which carry information on the structural shocks of interest (throughout the paper we use the terms proxies, instruments and external variables interchangeably). Inference in these models depends on whether the proxies are strongly or weakly correlated with the structural shock of interest, henceforth denoted *target shocks*.

As for instrumental variable [IV] regressions with weak instruments, proxies are 'weak' in proxy-SVARs when their connection with the target shocks can be approximated as local-to-zero (Staiger and Stock, 1997; Stock and Yogo, 2005). Montiel Olea, Stock and Watson (2021) show that inference is nonstandard in these cases and extend the logic of Anderson and Rubin tests to proxy-SVARs. When the proxy-SVAR features multiple target shocks this method requires inverting 'S-tests' and the imposition of a relatively large number of additional restrictions on the proxy-SVAR parameters other than the proxies, which are not always easily interpretable and testable.[1]

This paper is motivated by the idea that there are cases in which it may be convenient to recover the effects of the target shocks by using proxies for the shocks which are not of primary interest in the analysis, henceforth denoted '*auxiliary shocks*'. We formalize an identification strategy for proxy-SVARs in which the dynamic causal effects produced by the target shock(s) are point-identified by using proxies available for the auxiliary shocks. We call this identification strategy the 'indirect approach' or 'indirect identification strategy', as opposed to the conventional 'direct' approach based on instrumenting the target shock(s) directly. The indirect identification strategy is particularly advantageous when the proxies available for the target shocks are weak or suspected to be weak while the proxies used for the auxiliary shocks are strong.

As is known, SVARs can be represented either in 'B-form' or in 'A-form' (Amisano and Giannini, 1997; Lütkepohl, 2005), the main difference being that in the former the structural shocks are identified by modeling their impact on the variables, while in the latter the structural shocks are recovered by identifying the structural relationships that characterize the variables. The same

[1] For instance, in a SVAR with $n = 4$ variables and $k = 2$ target structural shocks, it is necessary to impose at least $k^2 = 4$ restrictions to build weak-identification-robust confidence sets for the IRFs by test inversion.

holds for proxy-SVARs. However, while in the B-form the auxiliary shocks have no active role in the identification, in the A-form the target shocks can be recovered by instrumenting the target shocks. Our suggested identification strategy exploits the link that characterizes the B-form and A-form representations of the proxy-SVAR and is formalized through a Minimum Distance [MD] estimation method (Newey and McFadden, 1994) based on the latter, called 'indirect-MD' throughout the paper. We derive the necessary order conditions and the necessary and sufficient rank condition for (point-)identification of the proxy-SVAR, and show that the indirect-MD approach involves standard asymptotics if the proxies available for the auxiliary shocks are strong. Accordingly, asymptotic valid confidence intervals for the IRFs obtain in the 'usual way', i.e. either by the delta-method or by bootstrap methods along the lines of Jentsch and Lunsford (2019a, 2019b). Relative to Montiel Olea, Stock and Watson's (2020) key contribution on weak-identification-robust methods in proxy-SVARs, the suggested indirect-MD approach is worth considering when finding strong instruments for the target shocks is more problematic than finding strong instruments for the auxiliary shocks. The obvious gain is the possibility to rely on standard asymptotic inference.

The indirect-MD estimation captures the moment conditions implied by the proxy-SVAR efficiently and overcomes some of the restrictions required by the application of the IV method when multiple instruments are used for multiple target shocks. In proxy-SVARs featuring multiple target shocks, estimation by IV requires the imposition of Choleski-type restrictions on a covariance matrix that results from a particular parameterization of the model, and this solves the problem of incorporating the additional parametric restrictions necessary for identification (Mertens and Ravn, 2013). With the suggested estimation approach the additional restrictions required to (point-)identify the target structural shocks need not be Choleski-type restrictions.

To our knowledge, Caldara and Kamps (2017) is the only example in the fiscal proxy-SVAR literature which, aside from the Bayesian estimation method, can be framed within the logic of the indirect identification strategy discussed in this paper. Caldara and Kamps (2017) interpret the structural equations associated with the A-form of their fiscal proxy-SVAR as fiscal reaction functions whose 'unsystematic' components correspond to the fiscal shocks of interest. They set-identify the fiscal shocks and the implied fiscal multipliers by a Bayesian penalty function approach.[2] Our approach departs from Caldara

[2] It can be argued that the Bayesian approach to proxy-SVARs does rely on A-form specifications; see e.g. Arias, Rubio-Ramirez and Waggoner (2020) and Giacomini, Kitagawa and Read (2020). However, the 'standard' Bayesian approach is designed to instrument the target structural shocks, not the auxiliary shocks. The fiscal proxy-SVAR considered in Caldara

and Kamps (2017) in two main respects. First, our primary objective is to provide an alternative to the weak-instruments-robust methods needed when the proxies available for the target structural shocks are weak, or suspected to be so. Second, we design a novel frequentist approach that point-identifies the proxy-SVAR parameters and involves standard frequentist asymptotic inference.

Instrument relevance. The availability of strong proxies for the auxiliary shocks suffices to consistently recover the target structural shocks by the indirect-MD approach and rely on standard asymptotic inference. The hypothesis of strong proxies is therefore key to this approach. We thus complement our estimation strategy with a novel diagnostic test for instrument relevance based on bootstrap resampling. We show that the moving block bootstrap [MBB] introduced in Brüggemann, Jentsch and Trenkler (2016) for SVARs and framed by Jentsch and Lunsford (2019a, 2019b) in the proxy-SVAR context, can be used to infer the strength of instruments (other than building confidence intervals for IRFs). The null hypothesis is that the proxies are strong against the alternative that they are weak. The test, which has been investigated by Angelini, Cavaliere and Fanelli (2021) in the context of state-space models, exploits the asymptotic distribution of a MBB estimator of the proxy-SVAR parameters under strong and weak proxies, respectively. The estimator, obtained through a classical MD (CMD) approach along the lines of Angelini and Fanelli (2019), reflects the strength of the proxies. Under strong proxies, the MBB version of the CMD estimator replicates the asymptotic distribution of its non-bootstrap counterpart, which is the Gaussian distribution. Under weak proxies, the cumulative distribution function [cdf] of the MBB version of the CMD estimator is stochastic in the limit in the sense of Cavaliere and Georgiev (2020) and, in particular, is non-Gaussian. Based on these results, we show that a test for strong against weak proxies can be designed, under the conditions studied in the paper, as a normality test applied to a properly selected number of MBB replications of the CMD estimator.

The test has three main properties. First and most importantly, it can be used as a pre-estimation test for instrument relevance which does not affect second-stage inference. In other words, the reliability of post-test inferences conditional on the test failing to reject the null of strong proxies is independent on the outcome of the test. This property marks an important difference with respect to the literature on weak instrument asymptotics where the consequences on the inference of pre-testing the strength of proxies have been well documented; see, *inter alia*, Zivot, Startz and Nelson (1998), Hausman, Stock

and Kamps (2017) represents a remarkable exception to this tradition.

and Yogo (2005) and Andrews, Stock and Sun (2019); see also Montiel Olea *et al.* (2021). Second, the test is consistent against weak as well as irrelevant proxies and controls size under very general conditions on VAR innovations and proxies. Third, the test is computationally straightforward and is obtained in the same way regardless of the number of shocks being instrumented.

We illustrate the usefulness of our identification, estimation and testing strategy by reconsidering some proxy-SVARs already estimated in the literature. In particular, we focus on the identification of an oil supply shock and the identification of financial and macroeconomic uncertainty shocks, respectively.[3]

Structure of the paper. The paper is organized as follows. Section 2 introduces the link between the B-form and A-form representations of a proxy-SVAR and shows that the mapping between the two can be exploited to rationalize an identification strategy where the auxiliary shocks can be instrumented to recover the target shocks. Section 3 summarizes the assumptions and characterizes the concepts of strong and weak proxies used throughout the paper. Section 4 presents the indirect-MD identification and estimation approach to proxy-SVARs. Section 5 complements the estimation strategy with a novel test for instrument relevance based on bootstrap resampling: Section 5.1 focuses on the mechanics of the test and Section 5.2 shows that the use of the suggested test does affect post-test inference on the IRFs. Section 6 presents two illustrative examples, Section 6.1 focuses on the identification of the oil-supply shock and Section 6.2 on the joint identification of financial and macroeconomic uncertainty shocks. Section 7 contains some concluding remarks. A Supplementary Material (SM) contains formal proofs of propositions and auxiliary lemmas and complements the paper along several dimensions.

2 Model, representations and identification strategies

In this section we discuss two different representations of a proxy-SVAR and show that the relationship among the two can be exploited to design an iden-

[3] In the Supplementary Material we also consider a fiscal proxy-SVAR from which we infer the US tax and fiscal spending multipliers.

tification strategy where the target shocks are recovered by instrumenting the auxiliary shocks of the system.

B-FORM. We start from the 'B-form' representation (Lütkepohl, 2005) of a SVAR model, given by

$$Y_t = \Pi X_t + u_t, \quad u_t = B\varepsilon_t \quad , t = 1, ..., T \tag{1}$$

where Y_t is the $n \times 1$ vector of endogenous variables, $X_t := (Y'_{t-1}, ..., Y'_{t-l},)'$ is the vector collecting l lags of the variables, $\Pi := (\Pi_1, \ ... \ , \Pi_l)$ is the $n \times nl$ matrix containing the autoregressive (slope) parameters and u_t is the $n \times 1$ vector of reduced form innovations with covariance matrix $\Sigma_u := \mathrm{E}(u_t u'_t)$. Deterministic terms have been excluded without loss of generality. The initial values $Y_0, ..., Y_{1-k}$ are fixed. The system of equations $u_t = B\varepsilon_t$ in (1) maps the vector of structural shocks ε_t $(n \times 1)$ to the reduced form innovations through the nonsingular matrix B $(n \times n)$ which contains the on-impact coefficients, i.e. the instantaneous effects of the structural shocks on the variables. It is maintained that the structural shocks have normalized covariance matrix $\Sigma_\varepsilon := \mathrm{E}(\varepsilon_t \varepsilon'_t) = I_n$ but the analysis can be easily extended to the direction of a diagonal non-unit Σ_ε matrix.

We partition the structural shocks as follows: $\varepsilon_t := (\varepsilon'_{1,t}, \varepsilon'_{2,t})'$, where $\varepsilon_{1,t}$ collects the $k < n$ target shocks, and $\varepsilon_{2,t}$ collects the remaining $n-k$ auxiliary shocks. Notice that we can write

$$u_t = \begin{pmatrix} u_{1,t} \\ u_{2,t} \end{pmatrix} = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix} \equiv B_1 \varepsilon_{1,t} + \ B_2 \varepsilon_{2,t} \tag{2}$$

where $u_{1,t}$ and $u_{2,t}$ have the same dimensions as $\varepsilon_{1,t}$ and $\varepsilon_{2,t}$, respectively, and $B_1 := (B'_{11} \vdots B'_{21})'$ is $n \times k$ and collects the on-impact coefficients associated with the target structural shocks. The proxy-SVAR approach consists in the identification of the target shocks $\varepsilon_{1,t}$ alone by using variables 'external' to the SVAR (henceforth we use the terms 'proxies' and 'instruments' interchangeably). Thus, the IRF of interest is given by the response of the i-th variable in Y_t to the j-th shock in $\varepsilon_{1,t}$, i.e.

$$\gamma_{i,j}(h) := \iota'_i (S'_n (\mathcal{A}_y)^h S_n) B_1 \iota_j \ , \quad \begin{matrix} i = 1, ..., n \\ j = 1, ..., k \end{matrix} \ , \quad h = 0, 1, 2, ..., h_{\max} \tag{3}$$

where $\mathcal{A}_y$ is the VAR companion matrix, $S_n := (I_n : 0_{n \times n(l-1)})$ is a selection matrix such that $S_n S'_n = I_n$ and ι_j is the $n \times 1$ vector containing '1' in the i-th position and zero elsewhere.

Proxy-SVARs solve the 'partial identification' problem raised by the estimation of the IRFs in (3) by assuming that there exist at least k observable proxies, collected in the vector z_t, which are correlated with $\varepsilon_{1,t}$ and are exogenous (orthogonal) to $\varepsilon_{2,t}$. This means that z_t can be related to $\varepsilon_{1,t}$ by the system of equations

$$z_t = \Phi\varepsilon_{1,t} + \omega_{z,t} \tag{4}$$

where the matrix $\Phi := \mathrm{E}(z_t\varepsilon_{1,t}')$ captures the link between the proxies and the target shocks and $\omega_{z,t}$ is a measurement error independent on $\varepsilon_t := (\varepsilon_{1,t}', \varepsilon_{2,t}')'$. By combining (4) with (2) and taking expectations one obtains the moment conditions

$$\Sigma_{z,u} = \Phi B_1' \tag{5}$$

where $\Sigma_{z,u} := \mathrm{E}(z_t u_t')$ is an $r \times n$ covariance matrix which can be estimated consistently from the data under standard regularity conditions.

The approach originally developed by Stock (2008), Stock and Watson (2012, 2018) and Mertens and Ravn (2013) estimates the IRFs in (3) by properly exploiting the moment conditions in (5). Hereafter, we call *direct approach* the method in which the proxies z_t are used to directly instrument the target shocks $\varepsilon_{1,t}$.

A-form. For $A = B^{-1}$, the 'B-form' in (1) can be equivalently expressed in 'A-form':

$$AY_t = \Upsilon X_t + \varepsilon_t, \quad Au_t = \varepsilon_t \quad , \quad t = 1, ..., T \tag{6}$$

where $\Upsilon := A\Pi$ and the matrix A summarizes the simultaneous relationships that characterize the observed variables. The mapping $Au_t = \varepsilon_t$ can be partitioned conformably with $\varepsilon_t := (\varepsilon_{1,t}', \varepsilon_{2,t}')'$ as follows:

$$\begin{pmatrix} A_1' \\ A_2' \end{pmatrix} u_t = \begin{pmatrix} A_{11}' & A_{12}' \\ A_{21}' & A_{22}' \end{pmatrix} \begin{pmatrix} u_{1,t} \\ u_{2,t} \end{pmatrix} = \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix} \tag{7}$$

with $A_1' := (A_{11}' \vdots A_{12}')$ corresponding to the block collecting the first k rows of A, and A_2' pertaining to the block collecting the remaining $n - k$ rows. In particular, the block of first k equations in (7) reads

$$A_1'u_t = A_{11}'u_{1,t} + A_{12}'u_{2,t} = \varepsilon_{1,t} \tag{8}$$

and, as it will be detailed in Section 4, it forms a system of k structural equations under a set of identification restrictions on A_1 (A_1').

Equation (8) suggests two things. First, the target shocks $\varepsilon_{1,t}$ can be recovered from the estimation of the parameters in $A_1' := (A_{11}' \vdots A_{12}')$. Second, the auxiliary shocks may have an 'active' role in the identification and estimation of the parameters in A_1' because, as it will be detailed in Section 4, one way to estimate these parameters is by using proxies, say v_t, that are correlated with (all or part of) the auxiliary shocks in $\varepsilon_{2,t}$ and are orthogonal to the target shocks $\varepsilon_{1,t}$.

Relation between the two representations and identification strategies. While the k columns of the matrix B_1 in (2) reflect the instantaneous impact of the target shocks on the variables Y_t, the k rows of the matrix A_1' in (7) capture the simultaneous relationships that characterize the variables in Y_t. Hence, while the identification of the proxy-SVAR based on the B-form amounts to 'separating' the columns of B_1, the identification of the proxy-SVAR based on the A-form amounts to 'separating' the rows of A_1'. Simple algebra shows that the connection between the parameters in B_1 and in A_1 is given by the expression[4]

$$B_1 = \Sigma_u A_1 \tag{9}$$

that suggests that the problem of making inference on the on-impact coefficients in B_1 and the IRFs in (3) can be alternatively formulated as the problem of making inference on the coefficients in A_1'.[5] Indeed, since the VAR covariance matrix Σ_u can be estimated consistently under fairly general conditions, a plug-in estimator of B_1 follows from (9) and the availability of a consistent estimator of the parameters in A_1'. In the next sections we formalize an estimation strategy, denoted indirect-MD approach, where the structural parameters in A_1 are estimated by exploiting a vector of proxies v_t for (part of) the auxiliary shocks in $\varepsilon_{2,t}$.

3 Assumptions and definitions

In this section we present the assumptions behind the indirect-MD approach to proxy-SVARs and characterize formally the concepts of 'strong' and 'weak' proxies used throughout the paper.

The first two assumptions pertain to the reduced form VAR.

Assumption 1 (Reduced form, stationarity) *The data generating process for Y_t belongs to the class of models in (1) where the companion matrix $\mathcal{A}_y$ is stable, i.e. all eigenvalues of $\mathcal{A}_y$ lie inside the unit disk.*

[4] Simple algebra shows that for in a SVAR where $A = B^{-1}$, it always holds the relationship $B = \Sigma_u A'$.

[5] The identification restrictions placed on A_1 (B_1) do not generally have a direct counterpart in B_1 (A_1), in the sense that if e.g. the (1,2)-element of A_1 (B_1) is set to zero, the (1,2)-element of B_1 (A_1) will typically differ from zero. We turn on this in Section 4 and in the SM.

Assumption 2 (Reduced form, VAR innovations) *The VAR innovations satisfy the following conditions:*
(i) $\{u_t\}$ *is a strictly stationary weak White Noise;*
(ii) $\mathrm{E}(u_t u_t') = \Sigma_u < \infty$ *is positive definite;*
(iii) u_t *is* α*-mixing, meaning that it satisfies the conditions stated extensively in Assumption 2.1 of Brüggemann, Jentsch and Trenkler (2016);*
(iv) u_t *has absolutely summable cumulants up to order eight.*

Assumption 1 features a typical maintained hypothesis of correct specification which also incorporates a stability (asymptotic stationarity) condition which rules out the presence of unit roots from the VAR. Assumption 2 is as in Francq and Raïssi (2006) and Boubacar Mainnasara and Francq (2011). Assumption 2(ii) is a standard unconditional homoskedasticity condition on VAR innovations and proxies. The α-mixing conditions in Assumption 2(iii) cover a large class of uncorrelated but possibly dependent variables, including the case of conditionally heteroskedastic innovations. Assumption 2(iv) is a technical condition necessary to prove the consistency of the MBB in model (1), see Brüggemann, Jentsch and Trenkler (2016).

The next assumption refers to the structural form.

Assumption 3 (Structural form) *Given the SVAR in (1), the matrix* B *is nonsingular and its inverse is* $A = B^{-1}$*.*

Assumption 3 establishes the invertibility of the matrix B which implies the conditions $rank[B_1] = k$ in (2) and $rank[A_1'] = k$ in (7). Note that A_{11}' in (7) can be singular. Importantly, Assumption 2 and Assumption 3 jointly imply that the structural shocks $\varepsilon_t = B^{-1}u_t \equiv Au_t$ inherit the α-mixing properties postulated for the VAR innovations.

The next assumption is key to our approach and refers to the existence of proxies available for the auxiliary shocks. Henceforth $\tilde{\varepsilon}_{2,t}$ denotes a subset of $s \leq n-k$ elements of $\varepsilon_{2,t}$; $s := \dim(\tilde{\varepsilon}_{2,t})$. It is intended that $s = n-k$ when $\varepsilon_{2,t} \equiv \tilde{\varepsilon}_{2,t}$.

Assumption 4 (Proxies for the auxiliary shocks) *There exist* $s \leq n-k$ *variables, collected in the vector* v_t*, such that*

$$v_t = \Lambda \tilde{\varepsilon}_{2,t} + \omega_{v,t} \quad , \quad t = 1, 2, ..., T \tag{10}$$

where $\Lambda := \mathrm{E}(v_t \tilde{\varepsilon}_{2,t}')$ *is an* $s \times s$ *matrix of relevance parameters and* $\omega_{v,t}$ *is an i.i.d. measurement error independent on* $\varepsilon_t := (\varepsilon_{1,t}', \varepsilon_{2,t}')'$*.*

Assumption 4 establishes that there exist s proxies that are correlated with the s auxiliary shocks in $\tilde{\varepsilon}_{2,t}$ and are exogenous (orthogonal) to the target shocks, $\mathrm{E}(v_t\varepsilon_{1,t}') = 0$.[6] We implicitly maintain that the number of auxiliary shocks, $n-k$, is not too big relative to the number of target shocks, k, otherwise there would be no benefit in designing a partial identification strategy based on instrumenting a large number of auxiliary shocks. Assumption 1-4 jointly imply that the process that generates the variables $(Y_t', v_t')'$ is stable and that the process that generates the reduced form innovations $\eta_{v,t}:=(u_t', v_t')'$ is α-mixing.

Strong and weak Proxies. Assumption 4 establishes the existence of proxies for the auxiliary shocks which are exogenous to the target shocks, but is silent on the strength of these proxies. The next definition provides a formal characterization of 'strong' and 'weak' proxies that will be used throughout the paper. In the following, λ_0 denotes the true value of λ, where λ collects the non-zero elements that enter the matrix of relevance parameters Λ in Assumption 4, and $\mathcal{N}_{\lambda_0}$ is a neighborhood of λ_0. We denote with λ_1, ..., λ_s the s columns of Λ, i.e. $\Lambda := (\lambda_1 \vdots ... \vdots \lambda_s)$.

Definition 1 (Strong and weak proxies) *Consider the proxy-SVAR obtained by combining the SVAR in (1) with the proxies v_t for the structural shocks $\tilde{\varepsilon}_{2,t}$ in (10). Given Assumption 4, the instrument v_t are:*

(a) 'strong' if Λ is fixed and for $\lambda \in \mathcal{N}_{\lambda_0}$, rank$[\Lambda] = s$;

(b) 'weak' if, for $\lambda \in \mathcal{N}_{\lambda_0}$, there exists at least one column λ_i of Λ such that $\lambda_i := CT^{-1/2}$, with C a fixed vector with non-zero elements.

According to Definition 1(a), the proxies v_t are 'strong' for $\tilde{\varepsilon}_{2,t}$ if each column of Λ carries relevant information on the s structural auxiliary shocks also asymptotically, as $T \to \infty$. Conversely, according to Definition 1(b), the proxies v_t are 'weak' for $\tilde{\varepsilon}_{2,t}$ if at least one column of the matrix Λ satisfies a local-to-zero embedding a la Staiger and Stock (1997). This makes Λ singular as $T \to \infty$, and near reduced rank for finite (large) T. Notice that in Definition 1(b) not all the proxies are necessarily weak. The case of irrelevant proxies, $C = 0_{s\times s}$, corresponds to the condition $\Lambda = 0_{s\times s}$ in (10). In the special case $s = 1$, i.e. when one proxy is used for one auxiliary shock, Definition 1(b) corresponds to Montiel Olea Stock and Watson's (2020) characterization of weak proxy. In the next sections we show that, contrary to what happens in the strong proxies setup, the asymptotic distribution of the bootstrap estimator

[6] In principle, Assumption 4 can be generalized to account more proxies than instrumented auxiliary shocks, $\dim(v_t) > s$. Without loss of generality, we keep exposition focused on cases where Λ in (10) is square.

of the proxy-SVAR parameters is non-Gaussian if the proxies are weak in the sense of Definition $1(b)$. The different limit behavior of this estimator under strong and weak proxies allows us to build a novel bootstrap-based test of instrument relevance.

4 Indirect-MD estimation

In this section we present our estimation approach to the parameters in the matrix A_1. Given an estimator of A_1, we use the relationship (9) to recover a plug-in estimator of B_1 and the IRFs in (3).

Moment conditions. Recall that, see (8),

$$A_{11}'u_{1,t} + A_{12}'u_{2,t} = \varepsilon_{1,t} \tag{11}$$

where the VAR innovations $u_{1,t}$ and $u_{2,t}$ have the same dimensions as $\varepsilon_{1,t}$ and $\varepsilon_{2,t}$, respectively. By taking the second-order moments of system (11) we obtain the $\frac{1}{2}k(k+1)$ moment conditions:

$$A_1'\Sigma_u A_1 = I_k. \tag{12}$$

Moreover, by using the proxies v_t in (11) we obtain the additional ks moment conditions:

$$A_1'\Sigma_{u,v} = 0_{k\times s} \tag{13}$$

where $\Sigma_{u,v} := \mathrm{E}(u_t v_t')$. Systems (12) and (13) jointly provide $m := \frac{1}{2}k(k+1) + ks$ independent moment conditions which can be used to estimate the parameters in A_1'.

As is known, in the multiple target shocks scenario, $k > 1$, the proxies alone do not suffice to (point-)identify the proxy-SVAR and it is necessary to impose additional parametric restrictions for point-identification, see Montiel Olea, Stock and Watson (2020). The additional restrictions might be provided by the theory or economic reasoning on the parameters in A_1' or on those in B_1, depending on the context. In the following we focus on the case in which the additional identification restrictions involve the structural parameters in A_1' and confine the case of restrictions on B_1 in the SM. Thus, for $k > 1$, we complement the moment conditions (12) and (13) with the following set of linear restrictions on A_1' :

$$\mathrm{vec}(A_1') = S_{A_1}\alpha + s_{A_1} \tag{14}$$

where α denotes the vector of (free) structural parameters that enter the matrix A_1, S_{A_1} is a full-column rank selection matrix and $s_{A_1}$is known vector which permits to accommodate non-homogeneous (non-zero) as well as cross-equation restrictions on A_1'.

Point-identification and estimation. Let $\sigma^+ := (\text{vech}(\Sigma_u)', \text{vec}(\Sigma_{u,v})')'$ be the $m \times 1$ vector of reduced form parameters of the proxy-VAR involved in the moment conditions in (12)-(13). σ_0^+ is the true value of σ^+, $\hat{\sigma}_T^+ := (\text{vech}(\hat{\Sigma}_u)', \text{vec}(\hat{\Sigma}_{u,v})')'$ is the estimator of σ^+ and $V_{\sigma^+} := \lim_{T\to\infty} \text{Var}[T^{1/2}(\hat{\sigma}_T^+ - \sigma_0^+)]$ is the asymptotic covariance matrix of $T^{1/2}(\hat{\sigma}_T^+ - \sigma_0^+)$.[7] The moment conditions (12)-(13) and the restrictions in (14) can be summarized by the distance function:

$$g(\sigma^+, \alpha) := \begin{pmatrix} \text{vech}(A_1'(\alpha)\Sigma_u A_1(\alpha) - I_k) \\ \text{vec}(A_1'(\alpha)\Sigma_{u,v}) \end{pmatrix} \tag{15}$$

where the notation $A_1 := A_1(\alpha)$ indicates that the elements of A_1 depend on the structural parameters α. Obviously, $g(\sigma_0^+, \alpha_0) = 0_{m\times 1}$ at the true parameter values. The MD estimator of α is obtained from:

$$\hat{\alpha}_T := \arg\min_{\alpha \in \mathcal{T}_\alpha} \hat{Q}_T(\alpha) \ , \ \ \hat{Q}_T(\alpha) := \hat{g}_T(\hat{\sigma}_T^+, \alpha)' \hat{V}_{gg}(\bar{\alpha})^{-1} \hat{g}_T(\hat{\sigma}_T^+, \alpha) \tag{16}$$

where $\hat{g}_T(\hat{\sigma}_T^+, \alpha)$ is the function $g(\sigma^+, \alpha)$ with σ^+ replaced with the estimator $\hat{\sigma}_T^+$. In (16), $\mathcal{T}_\alpha \subseteq \mathcal{P}_\alpha$ is the user-chosen optimization set, $\mathcal{P}_\alpha$ is the parameter space, $\hat{V}_{gg}(\bar{\alpha}) := G_{\sigma^+}(\hat{\sigma}_T^+, \bar{\alpha}) \hat{V}_{\sigma^+} G_{\sigma^+}(\hat{\sigma}_T^+, \bar{\alpha})'$, $\hat{V}_{\sigma^+}$ is a consistent estimator of V_{σ^+}, $G_{\sigma^+}(\sigma^+, \alpha)$ is the $m \times m$ Jacobian matrix defined by $G_{\sigma^+}(\sigma^+, \alpha) := \frac{\partial g(\sigma^+, \alpha)}{\partial \sigma^{+\prime}}$, and $\bar{\alpha}$ may be some preliminary (inefficient) estimate of α; for example, $\bar{\alpha}$ might be a MD estimate of α obtained in a previous step by replacing $\hat{V}_{gg}(\bar{\alpha})$ with the identity matrix, in which case the $\hat{\alpha}_T$ from (16) corresponds to a classical two-step MD estimator (see Newey and McFadden, 1994).

Before discussing the properties of the MD estimator $\hat{\alpha}_T$, in the next proposition we establish the necessary and sufficient rank condition and the necessary order condition for identification. Recall that $m := \frac{1}{2}k(k+1) + ks$ denotes the number of independent moment conditions in (12)-(13), while with a we henceforth indicate the dimension of the vector of structural parameters α in (14). With $\mathcal{N}_{\alpha_0}$ we denote a neighborhood of α_0 in $\mathcal{P}_\alpha$.

Proposition 1 (Point-identification) *Given the proxy-SVAR obtained by combining the SVAR (1) with the proxies v_t for the $s \leq n-k$ auxiliary structural shocks $\tilde{\varepsilon}_{2,t}$ in (10), assume that A_1' satisfies the moment conditions (12) and (13) and, for $k > 1$, is restricted as in (14). Under Assumptions 1-4, we have that:*

[7] σ^+ is a component of the reduced-form covariance parameters of the proxy-SVAR collected in the vector δ_η discussed in Section S2 of the SM.

(i) necessary and sufficient condition for identification is that the rank condition:

$$rank\left[G_{\alpha}(\sigma^{+},\alpha)\right]=a \tag{17}$$

holds in $\mathcal{N}_{\alpha_0}$, where

$$G_{\alpha}(\sigma^{+},\alpha):=\begin{pmatrix} 2D_k^{+}(A_1{}'\Sigma_u\otimes I_k) \\ (\Sigma_{v,u}\otimes I_k)\end{pmatrix}S_{A_1};$$

(ii) necessary order condition is $a\leq m$, and this condition implies that at least $\ell\geq\frac{1}{2}k(k-1)$ restrictions are placed on the proxy-SVAR parameters.

As it is typical for SVARs and proxy-SVARs, the identification result in Proposition 1 is 'up to sign', meaning that the rank condition in (17) is valid regardless on the sign normalizations of the rows of the matrix A_1'. The necessary order condition inequality $a\leq m$ simply states that when s auxiliary shocks are instrumented, the total number of moment conditions implied by the proxy-SVAR must be larger, or at least equal, to the total number of unknown parameters featured by the model. It is not strictly necessary that $s=(n-k)$, meaning that identification can also be achieved by instrumenting not all auxiliary shocks provided there are enough (economically plausible) restrictions on the parameters in A_1'. It can be easily proved that the rank condition in Proposition 1(i) does not hold if the proxies are weak in the sense of Definition 1(b).

Asymptotic properties. We have all the ingredients to derive the asymptotic properties of the MD estimator $\hat{\alpha}_T$ derived from (16). The next proposition contains the main result.

Proposition 2 (Asymptotic properties of MD estimator) *Under the conditions of Proposition 1, let α_0 be an interior of $\mathcal{P}_{\alpha}$ (assumed compact) and $\mathcal{N}_{\alpha_0}\subseteq\mathcal{T}_{\alpha}$. If the necessary and sufficient rank condition in Proposition 1 is satisfied, the estimator $\hat{\alpha}_T$ obtained from (16) has the following properties:*
(i) $\hat{\alpha}_T\overset{p}{\to}\alpha_0$;
(ii) $T^{1/2}(\hat{\alpha}_T-\alpha_0)\overset{d}{\to}N(0_{a\times 1},V_{\alpha})$, $V_{\alpha}:=\left\{G_{\alpha}(\sigma_0^{+},\alpha_0)'V_{gg}(\bar{\alpha})^{-1}G_{\alpha}(\sigma_0^{+},\alpha_0)\right\}^{-1}$, where $G_{\alpha}(\sigma^{+},\alpha)$ is the Jacobian matrix given in Proposition 1 and $V_{gg}(\bar{\alpha}):=G_{\sigma^{+}}(\sigma_0^{+},\bar{\alpha})V_{\sigma^{+}}G_{\sigma^{+}}(\sigma_0^{+},\bar{\alpha})'$.

Proposition 2 ensures that the inference on the parameter α is standard. This implies that one can recover 'consistent' estimates of the target shocks by using $\hat{\varepsilon}_{1,t}:=A_1(\hat{\alpha}_T)'\hat{u}_t$, $t=1,...,T$. Moreover, let $\gamma_{i,j,0}(h)$ denote the true impulse response in (3). One implication of Proposition 2 is that for $h = 0, 1, ...$ it holds:

$$T^{1/2}\left(\hat{\gamma}_{i,j,}(h) - \gamma_{i,j,0}(h)\right) \xrightarrow{d} N(0, V_{\gamma_{i,j}}) \quad , \quad \begin{array}{c} i = 1, ..., n \\ j = 1, ..., k \end{array} \tag{18}$$

where $\hat{\gamma}_{i,j,}(h) := \iota_i'(S_n'(\hat{\mathcal{A}}_y)^h S_n)\hat{B}_1\iota_j$, $\hat{B}_1 := \hat{\Sigma}_u A_1(\hat{\alpha}_T)$, and the asymptotic covariance matrix $V_{\gamma_{i,j}}$ follows from a standard delta-method argument.

Measure of strength. The MD estimation theory developed so far refers to a scenario in which the proxies v_t that instrument the auxiliary shocks $\tilde{\varepsilon}_{2,t}$ are (indirectly) used to recover the dynamic causal effects produced by the target shocks $\varepsilon_{1,t}$. We conclude this section by considering a bootstrap estimator of the proxy-SVAR parameters that will be used as measure of strength and will form the basis of our novel test of instrument relevance presented in the next sections.

From equation (10) and Assumption 4, we can derive the moment conditions $\Sigma_{v,u} = \Lambda\tilde{B}_2'$, where recall that $\Sigma_{v,u} := \mathrm{E}(v_t u_t')$, $\Lambda := \mathrm{E}(v_t\tilde{\varepsilon}_{2,t}')$, and the matrix $\tilde{B}_2 := \frac{\partial Y_t}{\partial \tilde{\varepsilon}_{2,t}'}$ collects the s columns of B_2 associated with the on-impact effects of the auxiliary shocks on the variables. Obviously, $\tilde{B}_2 \equiv B_2$ when $s = n - k$. We define the $s \times s$ symmetric matrix $\Omega_v := \Sigma_{v,u}\Sigma_u^{-1}\Sigma_{u,v}$ which depends on the reduced form covariance parameters Σ_u and $\Sigma_{v,u}$, respectively. It is seen that under the proxy-SVAR restrictions $\Sigma_{v,u} = \Lambda\tilde{B}_2'$ and $\Sigma_u = BB'$, $\Omega_v = \Lambda\tilde{B}_2'(BB')^{-1}\tilde{B}_2\Lambda' = \Lambda\Lambda'$, hence the matrix Ω_v reflects the strength of the proxies v_t. Let $\theta := (\lambda', \beta_2')'$ be the $q_\theta \times 1$ vector of unrestricted (free) parameters contained in the $(n + s) \times s$ matrix $(\tilde{B}_2' \vdots \Lambda')'$. λ contains the non-zero elements of the matrix of relevance parameters Λ while β_2 contains the non-zero on-impact coefficients in the matrix $\tilde{B}_2$. We remark that our interest in the estimation of the parameters $\theta := (\lambda', \beta_2')'$ is motivated by the need of constructing of a novel measure of strength.

Using the moment conditions $\Sigma_{v,u} = \Lambda\tilde{B}_2'$ and $\Omega_v = \Lambda\Lambda'$, we consider the distance function $\mu - f(\theta) = 0$, where $\mu := (\mathrm{vech}(\Omega_v)', \mathrm{vec}(\Sigma_{v,u})')'$ and $f(\theta) = (\mathrm{vech}(\Lambda\Lambda'\ )', \mathrm{vec}(\Lambda\tilde{B}_2'))$. As in Angelini and Fanelli (2019), a classical MD (CMD) estimator of θ can be obtained from the problem

$$\hat{\theta}_T := \arg\min_{\theta\in\mathcal{T}_\theta} \hat{Q}_T(\theta), \quad \hat{Q}_T(\theta) := (\hat{\mu}_T - f(\theta))'\hat{V}_\mu^{-1}(\hat{\mu}_T - f(\theta)) \tag{19}$$

where $\mathcal{T}_\theta \subseteq \mathcal{P}_\theta$ is the user-chosen optimization set, $\mathcal{P}_\theta$ is the parameter space, $\hat{\mu}_T := (vech(\hat{\Omega}_v)', vec(\hat{\Sigma}_{v,u})')'$ is the estimator of the reduced form parameters, $\hat{\Omega}_v := \hat{\Sigma}_{u,v}\hat{\Sigma}_u^{-1}\hat{\Sigma}_{u,v}$, $\hat{\Sigma}_u := T^{-1}\sum_{t=1}^T \hat{u}_t\hat{u}_t'$, $\hat{\Sigma}_{u,v} := T^{-1}\sum_{t=1}^T \hat{u}_t v_t'$, and $\hat{V}_\mu$ is such that $\hat{V}_\mu \xrightarrow{p} V_\mu := \lim_{T\to\infty}\mathrm{Var}(T^{1/2}(\hat{\mu}_T - \mu_0))$, μ_0 being the true value of μ.

In the SM, Lemma S3, we show that the asymptotic distribution of $T^{1/2}(\hat{\theta}_T - \theta_0)$, where $\theta_0 := (\lambda_0', \beta_{2,0}')'$ denotes the true value of θ, depends on whether the proxies are strong or weak according to Definition 1. Under strong proxies $T^{1/2}(\hat{\theta}_T - \theta_0)$ is asymptotically distributed as a multivariate Gaussian with asymptotic covariance matrix $V_\theta := \left(J_\theta' V_\mu^{-1} J_\theta\right)^{-1}$, J_θ being a full column rank Jacobian matrix. Conversely, under weak (or irrelevant) proxies $T^{1/2}(\hat{\theta}_T - \theta_0)$ is asymptotically non-Gaussian (its asymptotic distribution is derived in detail in the proof of Lemma S3). The bootstrap counterpart of the CMD estimator $\hat{\theta}_T$, henceforth denoted MBB-CMD, is obtained from the problem

$$\hat{\theta}_T^* := \arg\min_{\theta \in \mathcal{T}_\theta} \hat{Q}_T^*(\theta) \quad , \quad \hat{Q}_T^*(\theta) := (\hat{\mu}_T^* - f(\theta))' \hat{V}_\mu^{-1} (\hat{\mu}_T^* - f(\theta)) \tag{20}$$

where $\hat{\mu}_T^* := (\text{vech}(\hat{\Omega}_v^*)', vec(\hat{\Sigma}_{v,u}^*)')'$ is the bootstrap analog of $\hat{\mu}_T$. Bootstrap replications of $\hat{\mu}_T^*$ ($\hat{\Omega}_v^*$, $\hat{\Sigma}_{v,u}^*$) can be obtained from the MBB algorithm sketched in the SM. Henceforth, the bootstrap statistic $\Gamma_T^* := T^{1/2} V_\theta^{-1/2} (\hat{\theta}_T^* - \hat{\theta}_T)$ will be used as our measure of strength. Indeed, the next proposition shows that the asymptotic distribution of Γ_T^*, conditional on the data, depends on the strength the proxies.

Proposition 3 (Asymptotic distribution, bootstrap estimator) *Under the conditions of Proposition 2, consider the CMD estimator $\hat{\theta}_T$ obtained from (19) and its MBB counterpart $\hat{\theta}_T^*$ derived from (20). Under Assumptions 1-4:*
(i) if the proxies are strong, $\Gamma_T^* := T^{1/2} V_\theta^{-1/2} (\hat{\theta}_T^* - \hat{\theta}_T) \xrightarrow{d^*}_p N(0_{q_\theta \times 1}, I_{q_\theta})$;
(ii) if the proxies are weak (or irrelevant), the cdf of $\Gamma_T^* := T^{1/2} V_\theta^{-1/2} (\hat{\theta}_T^* - \hat{\theta}_T)$ is stochastic in the limit and is non-Gaussian. The limit distribution is derived in the proof, see equations (S.24) and (S.27) in the SM.

Proposition 3(i) shows that under strong instruments $\Gamma_T^* := T^{1/2} V_\theta^{-1/2} (\hat{\theta}_T^* - \hat{\theta}_T)$ replicates, conditional on the data, the asymptotic distribution of Γ_T which is Gaussian. This result is consistent with Jentsch and Lunsford's (2019b) finding that the MBB is consistent in proxy-SVARs with strong proxies. Proposition 3(ii) establishes that when the proxies are weak (or irrelevant), the asymptotic distribution of Γ_T^*, conditional on the data, is random in the limit and non-Gaussian; see Cavaliere and Georgiev (2020) for details on convergence (in distribution) of random cdfs. The different asymptotic behavior of Γ_T^* under strong and weak proxies is at the basis of our novel bootstrap test of instrument relevance discussed next.

Before moving to the next section, we remark that the asymptotic normality result in Proposition 3(i) holds regardless of the validity of the exogeneity condition. More precisely, conditional on the data, under strong proxies the statistic Γ_T^* remains asymptotically Gaussian also when the proxies used for the auxiliary shocks fail to be exogenous (orthogonal) to (some of) the target shocks in $\varepsilon_{1,t}$. We study in detail the violation of the exogeneity condition in the SM, Section S.9. There we focus on an simplified setup which shows that when the exogeneity conditions fails, the quantity $T^{1/2}(\hat{\theta}_T - \theta_0^+)$ is still asymptotically Gaussian, with $\theta_0^+ \neq \theta_0$ being a 'pseudo-true' value of θ. Accordingly, conditional on the data, the bootstrap counterpart of $T^{1/2}(\hat{\theta}_T - \theta_0^+)$, $T^{1/2}(\hat{\theta}_T^* - \hat{\theta}_T)$, will be still asymptotically Gaussian. This result is important for the test of instrument relevance discussed in the next section as it ensures that the asymptotic non-normality of the statistic $T^{1/2}(\hat{\theta}_T^* - \hat{\theta}_T)$ solely depends on the strength of the proxies.

5 Instrument relevance

The indirect-MD approach discussed in the previous sections hinges on the availability of strong proxies for the auxiliary shocks. In this section we complement the suggested strategy with a novel test for instrument relevance which can be used to verify the null hypothesis of strong proxies. To do so, we exploit the asymptotic behavior of the normalized bootstrap statistic $\Gamma_T^* := T^{1/2} V_\theta^{-1/2}(\hat{\theta}_T^* - \hat{\theta}_T)$ derived in Proposition 3.

Section 5.1 provides the mechanics of the test and Section 5.2 focuses on the most important property of this test, i.e. the fact that it does not affect post-test inference on the IRFs.

5.1 Bootstrap test

We consider the statistic $\hat{\Gamma}_T^* := T^{1/2} \hat{V}_\theta^{-1/2}(\hat{\theta}_T^* - \hat{\theta}_T)$, where $\hat{V}_\theta$ is an estimator of the asymptotic covariance matrix V_θ. Without loss of generality, we focus one component of the vector $\hat{\Gamma}_T^*$, say its first element, $\hat{\Gamma}_{1,T}^*$. Let $\digamma_{1,T}^*(\cdot)$ be the cumulative distribution function of $\hat{\Gamma}_{1,T}^*$, conditional on the data. $\digamma_{1,T}^*(\cdot)$ is used to approximate the distribution of $\Gamma_{1,T}$, say $\digamma_{1,T}(\cdot)$. By Proposition 3(i), under strong proxies $\hat{\Gamma}_{1,T}^*$ converges to a standard normal random variable so that $\digamma_{1,T}^*(x) - \digamma_{\mathcal{N}}(x) \to_p 0$ uniformly in $x \in \mathbb{R}$ as $T \to \infty$, where $\digamma_{\mathcal{N}}(\cdot)$ denotes the $N(0,1)$ cdf. Since this is an asymptotic result, for T fixed the bootstrap distribution $\digamma_{1,T}^*(\cdot)$ may potentially deviate from the normal even if the proxies are strong and the result in Proposition 3(i) is valid. Therefore, our idea is to evaluate whether $\digamma_{1,T}^*(\cdot)$ approaches the normal distribution for large T.

We can estimate $\mathcal{F}^*_{1,T}(x)$ from the sequence of i.i.d. bootstrap replications $\hat{\Gamma}^*_{1,T:1}, ..., \hat{\Gamma}^*_{1,T:N}$, using

$$\mathcal{F}^*_{1,T,N}(x) := \frac{1}{N}\sum\nolimits_{b=1}^{N}\mathbb{I}(\hat{\Gamma}^*_{1,T:b} \leq x),\ x \in \mathbb{R}. \tag{21}$$

For any x, deviation of $\mathcal{F}^*_{1,T,N}(x)$ from the standard normal distribution can be evaluated by considering the distance $\mathcal{F}^*_{1,T,N}(x) - \mathcal{F}_{\mathcal{N}}(x)$. By standard arguments and regardless of the strength of the proxies, as $N \to \infty$ (keeping T fixed)

$$N^{1/2}(\mathcal{F}^*_{1,T,N}(x) - \mathcal{F}^*_{1,T}(x)) \xrightarrow{d} N(0, U_T(x)) \tag{22}$$

where $U_T(x) := \mathcal{F}^*_{1,T}(x)(1-\mathcal{F}^*_{1,T}(x))$. This fact suggests that we may consider the normalized statistic

$$\tau^*_{T,N}(x) := N^{1/2}\hat{U}_T(x)^{-1/2}(\mathcal{F}^*_{1,T,N}(x) - \mathcal{F}_{\mathcal{N}}(x)), \tag{23}$$

as actual measure of distance, where $\hat{U}_T(x)$ denotes a consistent estimator of $U_T(x)$.[8] The statistic $\tau^*_{T,N}(x)$ in (23) captures the (normalized) distance between the estimated (over N repetitions) bootstrap distribution $\mathcal{F}^*_{1,T,N}(x)$ and the theoretical *asymptotic* distribution that one would get under strong proxies.

The next proposition establishes the limit behavior of the statistic $\tau^*_{T,N}(x)$ under strong and weak proxies, respectively.

PROPOSITION 4 *Let $\tau^*_{T,N}(x)$ be the statistic defined in (23). Under the conditions of Proposition 3 and Assumptions 1-4, assume that:*

$$T, N \to \infty \text{ jointly and } NT^{-1} = o(1). \tag{24}$$

Then it holds that:
(i) under strong proxies, if $\mathcal{F}^*_T(x)$ admits the standard Edgeworth expansion $\mathcal{F}^*_T(x) - \mathcal{F}_{\mathcal{N}}(x) = O_p(T^{-1/2})$, conditional on the data, $\tau^*_{T,N}(x) \xrightarrow{d^*}_p N(0,1)$;
(ii) under weak (or irrelevant) proxies, $\tau^*_{T,N}(x)$ diverges, conditional on the data, at the rate $N^{1/2}$.

Note that the Edgeworth expansion assumed in Proposition 4(i) is also maintained in e.g. Bose (1988) and Kilian (1988).[9] Proposition 4(ii) shows

[8] For instance, one may consider $\hat{U}_T(x) := \mathcal{F}^*_{1,T,N}(x)(1-\mathcal{F}^*_{1,T,N}(x))$ for an arbitrary large value of N, or can simply set $\hat{U}_T(x)$ to its theoretical value under normality, i.e. $\hat{U}_T(x) := \mathcal{F}_{\mathcal{N}}(x)(1-\mathcal{F}_{\mathcal{N}}(x))$.

[9] The Edgeworth expansion $\mathcal{F}^*_T(x) - \mathcal{F}_{\mathcal{N}}(x) = O_p(T^{-1/2})$ is typical in the presence of asymptotically normal statistics, see e.g. Horowitz (2001, p. 3171) and Hall (1992).

that the test is consistent against weak as well as irrelevant proxies. Overall, Proposition 4 provides the rationale for the design of a tests for strong versus weak (or irrelevant) proxies which can be formulated in terms of normality tests applied to N bootstrap replications of the estimator $\hat{\theta}_T^*$, where N is selected consistently with the condition (24), see below. Moreover, the results in Proposition 4 can be extended to all components of $\hat{\Gamma}_T^*$ as well as to the whole vector $\hat{\Gamma}_T^*$, meaning that in practice one can check instruments relevance using both multivariate and univariate versions of normality tests.

Operational implementation and Monte Carlo evidence. Henceforth we use the symbol $\hat{\vartheta}_T^*$ to denote the following statistics that can be alternatively chosen once the MBB-CMD estimator $\hat{\theta}_T^*$ discussed in the previous section is computed: (i) the $\hat{\theta}_T^*$ estimator itself, i.e. $\hat{\vartheta}_T^* \equiv \hat{\theta}_T^*$; (ii) the transformation $\hat{\Gamma}_T^* := T^{1/2}\hat{V}_\theta^{-1/2}(\hat{\theta}_T^* - \hat{\theta}_T)$, i.e. $\hat{\vartheta}_T^* \equiv \hat{\Gamma}_T^*$; (iii) sub-vectors of $\hat{\theta}_T^*$, e.g. $\hat{\vartheta}_T^* \equiv \hat{\beta}_{2,T}^*$, where recall that $\hat{\theta}_T^* := (\hat{\lambda}_T^{*\prime}, \hat{\beta}_{2,T}^{*\prime})'$; (iv) individual elements of $\hat{\theta}_T^*$, e.g. $\hat{\vartheta}_T^* \equiv \hat{\theta}_{1,T}^*$, $\hat{\theta}_{1,T}^*$ being the first element of $\hat{\theta}_T^*$. Our bootstrap-based test of instrument relevance boils down to running normality tests to the sequence of bootstrap replications $\{\hat{\vartheta}_{T:1}^*, \hat{\vartheta}_{T:2}^*, ..., \hat{\vartheta}_{T:N}^*\}$, where N is chosen in finite samples with a rule consistent with the condition (24).[10] Regardless on the number of shocks being simultaneously instrumented, the null hypothesis of strong proxies is rejected when the asymptotic normality hypothesis is rejected at the pre-fixed nominal significance level. We recommend checking multivariate normality first and then, conditionally on not rejecting multivariate normality, possibly testing the normality of the single components of the vector.

The condition $NT^{-1} \to 0$ in (24) is crucial for the tests to correctly control nominal size. Inspection of the proof of Proposition 4 reveals that the test may incorrectly reject the hypothesis of strong proxies if N is not 'too small' relative to T. Our simulation experiments, part of which are summarized in Table 1, suggest that the choice $N = [T^{1/2}]$ delivers a satisfactory compromise between size control and power in samples of length typically available to practitioners.

More in detail, we investigate the finite sample properties of our bootstrap diagnostic test by some Monte Carlo experiments based on a data generating process whose details are provided in the accompanying SM. In short, the data generating process belongs to a SVAR system with $n = 3$ variables featuring a single target shock $\varepsilon_{1,t}$ ($k = 1$) and two auxiliary shocks. The target shock $\varepsilon_{1,t}$ is recovered from the structural equation $A_1' u_t = A_{11}u_{1,t} + A_{12,1}u_{2,t} + A_{12,1}u_{3,t} = \varepsilon_{1,t}$ using a proxy v_t for the auxiliary shock $\varepsilon_{3,t} \equiv \tilde{\varepsilon}_{2,t}$ ($s = 1 <$

[10] Alternatively, normality tests can be applied to the sequences $\{\hat{\vartheta}_{T:1}^* - \hat{\vartheta}_T, \hat{\vartheta}_{T:2}^* - \hat{\vartheta}_T, ..., \hat{\vartheta}_{T:N}^* - \hat{\vartheta}_T\}$.

$n-k=2$) and imposing the restriction $A_{12,1}=0$ (valid in the data generating process). Using the notation in Section 4 one has $a=2$ and $m=2$.

Table 1 summarizes the empirical rejection frequencies of the bootstrap diagnostic test computed on 20,000 simulations under different scenarios on the correlation between the proxy v_t and the auxiliary shock $\tilde{\varepsilon}_{2,t}$. All normality tests are carried out at the 5% nominal significance level. Let $\hat{\theta}_T^* := (\hat{\lambda}_T^{*\prime}, \hat{\beta}_{2,T}^{*\prime})'$ be the MBB-CMD estimator that captures the strength of instruments, discussed at the end of Section 4. Considering samples of length $T=250$ and $T=1,000$ and setting the tuning parameter N to $N=[T^{1/2}]$, we apply Doornik and Hansen's (2008) multivariate test of normality (henceforth DH) on the sequence $\{\hat{\beta}_{2,T:1}^*, ..., \hat{\beta}_{2,T:N}^*\}$ and Lilliefors' (1967) version of univariate Kolmogorov-Smirnov (KS) tests of normality on the single elements of the vectors in the sequence $\{\hat{\theta}_{T:1}^*, \hat{\theta}_{T:2}^*, ..., \hat{\theta}_{T:N}^*\}$.

Results in the upper panel of Table 1 refer to a 'strong proxy' scenario where the correlation between the proxy and the instrumented auxiliary structural shock is equal to 59% and does not change with the sample size. The rejection frequencies not in parentheses refer to data simulated from i.i.d. VAR innovations (implied by assuming i.i.d. Gaussian structural shocks), while the rejection frequencies in parentheses refer to data simulated form GARCH-type VAR innovations (implied by assuming GARCH-type structural shocks). In both cases the test controls nominal size satisfactory well. The lower panel of Table 1 refers to a 'weak proxy' scenario, where the proxy used to instrument the auxiliary shock satisfies a local-to-zero embedding: the correlation between the proxy and the target shocks is equal to 5% in samples of length $T=250$ and collapses to 2% in samples of length $T=1000$. Results show that in both the i.i.d. and GARCH case, the test detects the weak proxy rather well and, importantly, the power of the test increases with the sample size. Finally, the middle panel of Table 1 refers to a 'moderately weak proxy' scenario, where we still have a local-to-zero embedding such that the correlation between the proxy and the instrumented auxiliary shock is set to 25% in samples of length $T=250$ and collapses to 13% in samples of length $T=1000$. In this data generating process, the tests behaves reasonably well: in samples of length $T=250$ it detects the weak proxy case the 20% of cases (results are robust to GARCH-type components) but, importantly, as the sample size increases also the capacity of the test of correctly rejecting the null of strong proxy increases with a rejection frequency in the range 64%-80%.[11]

[11] Further Monte Carlo results on the properties of this bootstrap-based test may be found in Angelini, Cavaliere and Fanelli (2020). Results show that with the choice $N=[T^{1/2}]$, the test displays empirically reasonable rejection rates in finite samples also in the presence of zero-censored proxies and multiple target shocks.

5.2 Post-test inference on the IRFs

As is also known from the literature on IV regressions (and regardless of whether the null hypothesis of the test is that of weak or strong proxy), caution is needed against choosing among instruments on the basis of their first-stage significance, since screening worsens small sample bias, see Zivot, Startz and Nelson (1998), Hausman, Stock and Yogo (2005) and Andrews, Stock and Sun (2019). Hence, one important way to assess the overall performance of our novel bootstrap-based test for instrument relevance is to examine, in addition to the rejection frequencies in Table 1, the reliability of post-test inferences conditional on the test failing to reject the null of strong proxies. In this section we establish that the asymptotic results stated in Proposition 2 remain valid also after pre-testing the strength of the proxies by the novel bootstrap-based test. We focus in particular on the post-test coverage of IRFs obtained by the indirect-MD approach.

In the following, ρ_T denotes any statistic computed from the proxy-SVAR estimated on the original sample. For example, for $h = \bar{h}$, ρ_T might correspond to the quantity $\rho_T := T^{1/2}\left(\hat{\gamma}_{i,j,}(\bar{h}) - \gamma_{i,j,0}(\bar{h})\right)/\hat{V}_{\gamma_{i,j}}^{1/2}$, with $\hat{\gamma}_{i,j,}(\bar{h})$ defined as in (18), $\gamma_{i,j,0}(\bar{h})$ being a true null value of the IRF and $\hat{V}_{\gamma_{i,j}}$ a consistent estimator of $V_{\gamma_{i,j}}$. More in general, ρ_T might correspond to a Wald-type statistic for restrictions on the parameters in A_1 (or in B_1). Instead, with $\tau^*_{T,N} := \tau(\hat{\theta}^*_{T:1}, ..., \hat{\theta}^*_{T:N})$ we denote any statistic that depends on a sequence of N bootstrap replications of the MBB-CMD estimator $\hat{\theta}^*_T$. For example, $\tau^*_{T,N}$ might coincide with the DH multivariate test statistic applied to the sequence of MBB realizations $\{\hat{\theta}^*_{T:1}, \hat{\theta}^*_{T:2}, ..., \hat{\theta}^*_{T:N}\}$, see Section 5.1. Notice that $\tau^*_{T,N}$ depends on the original data through the (conditional) distribution function $F_T(\cdot)$ only.

The following proposition establishes that the statistics ρ_T and $\tau^*_{T,N}$ are independent asymptotically ($T, N \to \infty$). We implicitly assume that the data and the auxiliary variates used to generate the bootstrap data are defined jointly on a possibly extended probability space.

Proposition 5 (Asymptotic independence) *Let ρ_T and $\tau^*_{T,N} := \tau(\hat{\theta}^*_{T:1}, ..., \hat{\theta}^*_{T:N})$ be statistics defined as above. For any $x, c \in \mathbb{R}$ and $T, N \to \infty$, it holds that*

$$P(\{\rho_T \leq x\} \cap \{\tau^*_{T,N} \leq c\}) - P(\rho_T \leq x)P(\tau^*_{T,N} \leq c) \longrightarrow 0. \quad (25)$$

To illustrate one important implication of Proposition 5, we turn on the data generating process already discussed in Section 5.1 to investigate the finite sample rejection frequency of the novel test of instrument relevance. Figure 1 plots, in samples of $T = 250$ observations and for $h = 0, 1, ..., 12$ periods, the actual empirical coverage probabilities of 90%-confidence intervals constructed for the response the variable $Y_{3,t+h}$ to the target shock $\varepsilon_{1,t}$. Actual empirical coverage probabilities are calculated considering 20,000 simulations. The black dotted line (which in the graph is almost totally covered by the blue pale dotted line, see below) refers to the coverages obtained by the indirect-MD approach based on the structural equation $A_1' u_t = a_{11}u_{1,t} + a_{12}u_{2,t} + a_{13}u_{3,t} = \varepsilon_{1,t}$, the instrumentation of the auxiliary shock $\tilde{\varepsilon}_{2,t} \equiv \varepsilon_{3,t}$ with a strong proxy v_t and the imposition of the restriction $a_{12} = 0$ (true in the data generating process); the setup is formally similar to the 'Strong proxy' case in the upper panel of Table 1. The graph shows that, unconditionally, the finite sample coverage of IRFs, denoted $P(\text{cover}_{MD}(h))$, $h = 0, 1, ...12$, is satisfactory. The blue pale dotted line refers, instead, to the conditional probabilities $P(\text{cover}_{MD}(h) \mid DH \leq cv)$, $h = 0, 1, ...12$, i.e. the actual empirical coverage probabilities conditional on the DH multivariate normality test ($\tau^*_{T,N} \equiv DH$), computed on $N = [T^{1/2}]$ bootstrap realization of the MBB-CMD estimator, failing to reject the null of strong proxies. Figure 1 shows that, in line with the theoretical result in Proposition 5, the unconditional and conditional empirical coverage probabilities tend to coincide.

To further appreciate the importance of this result, we estimate the responses of $Y_{3,t+h}$ to the target shock by instrumenting $\varepsilon_{1,t}$ directly with a weak proxy z_t; the setup corresponds formally to the 'weak proxy' scenario in the lower panel of Table 1. In this case, we proceed as follows. Building weak-identification-robust (Anderson-Rubin) confidence intervals along the lines suggested by Montiel Olea *et al.* (2021), we obtain the actual empirical coverage probabilities, $P(\text{cover}_{A\&R}(h))$, $h = 0, 1, ...12$, corresponding to the blue dotted line in Figure 1. Instead, if we estimate the proxy-SVAR pretending that z_t is a strong instrument for $\varepsilon_{1,t}$, we obtain 'plug-in' confidence intervals whose actual coverage probabilities, denoted $P(\text{cover}_{W_{plug-in}}(h))$, $h = 0, 1, ...12$, correspond to the red dotted line in Figure 1. As expected, unconditionally, the coverage is poor. If in this weak instrument scenario we pre-test the strength of the proxy by the first-stage F-test and consider the actual coverage probabilities conditional on the first-stage F-test rejecting the null of weak proxies, i.e. the probabilities $P(\text{cover}_{W_{plug-in}}(h) \mid F > cv)$, $h = 0, 1, ...12$, the results are given by the green dotted line in Figure 1. Thus, it is seen that screening on the first-stage F-test worsens coverage. However, the gap between unconditional and conditional coverage probabilities becomes less dramatic in this scenario if confidence intervals are built conditional on our bootstrap test of instrument relevance failing to reject the null of strong proxies, i.e. the quantities $P(\text{cover}_{W_{plug-in}}(h) \mid DH \leq cv)$, $h = 0, 1, ...12$, that correspond to the yellow dotted line in Figure 1.

The main take-away of Figure 1 is as follows. In the presence of weak proxies for the target shocks, practitioners can rely on weak instrument-robust methods. However, if there exist strong proxies for the auxiliary shocks, the inference on the dynamic causal effects produced by the target shocks can be improved by the indirect-MD approach. Importantly, the use of our novel bootstrap pre-test of instrument relevance allows us to screen strong versus weak proxies without affecting the coverage of post-test confidence intervals.

6 Empirical illustrations

We show the usefulness of the indirect-MD approach by turning on some empirical illustrations already considered in the literature. Section 6.1 starts from Kilian's (2009) SVAR for the identification of an oil supply shock ($k = 1$) and compares Montiel Olea, Stock and Watson's weak-identification-robust approach with the indirect-MD approach; Section 6.2 discusses the joint identification of financial and macroeconomic uncertainty shocks ($k = 2$) using Ludvigson, Ma and Ng's (2019) reduced form VAR as statistical platform. A third empirical illustration based on a fiscal proxy-SVAR and the identification of the tax and fiscal spending shocks can be found in the SM.

6.1 Oil supply shock

Direct approach, specification. Kilian (2009) considers a three-equation ($n = 3$) SVAR for $Y_t := (prod_t, rea_t, rpo_t)'$, where $prod_t$ is the percent change in global crude oil production, rea_t is a global real economic activity index of dry goods shipments and rpo_t the real oil price. Using monthly data for the period 1973:M1-2007:M12 and a Choleski decomposition (based on the above ordering of the variables), he identifies three structural shocks: the oil supply shock, ε_t^S, an aggregate demand shock, ${\varepsilon_t}^{AD}$, and an oil-specific demand shock, ${\varepsilon_t}^{OSD}$, respectively. Montiel Olea *et al.* (2021) focus on the identification of the oil supply shock alone using the same reduced form VAR as Kilian (2009) and Kilian's (2008) measure of 'exogenous oil supply shock' as external instrument for ε_t^S, henceforth denoted z_t.

In our notation, $\varepsilon_{1,t} = \varepsilon_t^S$ ($k = 1$) is the target structural shock, z_t is Kilian's (2008) proxy for $\varepsilon_{1,t}$ and $\varepsilon_{2,t} = ({\varepsilon_t}^{AD}, {\varepsilon_t}^{OSD})'$ ($n - k = 2$) are the auxiliary shocks of the system. The counterpart of the B-form in (2) is given by the system

$$\begin{pmatrix} u_t^{prod} \\ u_t^{rea} \\ u_t^{rpo} \end{pmatrix} = \begin{pmatrix} \beta_{11} \\ \beta_{21} \\ \beta_{31} \end{pmatrix} \varepsilon_t^S + B_2 \varepsilon_{2,t}$$

where $u_t := (u_t^{prod}, u_t^{rea}, u_t^{rpo})'$ are the VAR innovations and the coefficients in $B_1 \equiv (\beta_{11}, \beta_{21}, \beta_{31})'$ captures the instantaneous impact of the oil supply shock on the variables. The link between Kilian's (2008) proxy and the oil supply shock, i.e. the counterpart of (4) is given by the equation

$$z_t = \phi \varepsilon_t^S + \omega_{z,t}$$

where ϕ is the relevance parameter and $\omega_{z,t}$ is an i.i.d. measurement error. Since $k = 1$, no additional restriction other the unit effect normalization $\beta_{11} = 1$ and the proxy z_t is needed on the parameters β_{21}, β_{31} and ϕ to build weak-identification-robust confidence intervals.

Direct approach, proxy relevance and IRFs. The instrument z_t is available on the period 1973:M1-2004:M9 and, following Montiel Olea *et al.* (2021), we use the common sample period 1973:M1-2004:M9 ($T = 381$ monthly observations) for estimation. Montiel Olea *et al.* (2021) report a robust first-stage statistic for the proxy z_t equal to 9.4. We complement their analysis with our bootstrap pre-test for proxy relevance. To do so, we fix the tuning parameter $N = [T]^{1/2} = 19$, and apply DH multivariate normality test on the sequence of bootstrap (MBB) replications $\{\hat{\vartheta}^*_{T:1}, \hat{\vartheta}^*_{T:2}, ..., \hat{\vartheta}^*_{T:N}\}$, where the bootstrap estimator $\hat{\vartheta}^*_T$ is obtained from the proxy-SVAR as follows. First, we consider the choice $\hat{\vartheta}^*_T \equiv \hat{\theta}^*_T = (\hat{\beta}^{*\prime}_{1,T}, \hat{\phi}^*_T)'$, where $\hat{\theta}^*_T = (\hat{\beta}^{*\prime}_{1,T}, \hat{\phi}^*_T)'$ is the MBB-CMD estimator discussed in Section 4.[12] The DH multivariate normality test delivers in this case a p-value of 0.04. Second, we consider the choice $\hat{\vartheta}^*_T \equiv \hat{\beta}^*_{1,T}$ and in this case the DH multivariate normality test has a p-value of 0.004.[13] Overall, the bootstrap test rejects the hypothesis that Kilian's (2008) proxy z_t is strong for the oil supply shock and this evidence supports Montiel Olea, Stock and Watson's (2020) weak-identification-robust approach in this framework.

The blue lines plotted in Figure 2 are the estimated impulse response coefficients obtained using Kilian's (2008) proxy z_t. More precisely, the graph quantifies the responses of the variables in $Y_t := (prod_t, rpo_t, rea_t)'$ to an oil supply shock that increases oil production of 1% on-impact (note that the responses plotted for $prod_t$ are cumulative percent changes). The blue shaded

[12] Since in this case we are testing the strength of a proxy which is used to directly instrument the target shock, the test mimics the analysis developed in Section 4 with a the key difference: the MBB-CMD estimator in (20) is computed from the moment conditions $\Sigma_{z,u} = \Phi B_1'$, $\Omega_z = \Phi B_1'(BB')^{-1} B_1 \Phi' = \Phi\Phi'$ which reflect the strength of z_t for the oil supply shock.

[13] Univariate normality tests computed on the single elements of $\hat{\vartheta}^*_T \equiv \hat{\theta}^*_T = (\hat{\beta}^{*\prime}_{1,T}, \hat{\phi}^*_T)'$ and $\hat{\vartheta}^*_T \equiv \hat{\beta}^{*\prime}_{1,T}$, not reported to save space, confirm the outcome of the multivariate normality tests.

area are the associated 68% (panel A) and 95% (panel B) Anderson-Rubin weak-identification-robust confidence intervals and coincide with the IRFs plotted in panels A and B of Figure 1 in Montiel Olea *et al.* (2021) (see in particular their 'SVAR-IV' and 'CS^{AR}'). Next we compare these responses with the ones obtained by identifying the oil supply shock by the indirect-MD strategy.

Indirect-MD approach, specification. The A-form of this proxy-SVAR, i.e. the counterpart of system (8), is given by the equation:

$$\alpha_{11}u_t^{prod} + (\alpha_{12}\ ,\ \alpha_{13})\begin{pmatrix} u_t^{rea} \\ u_t^{rpo} \end{pmatrix} = \varepsilon_t^S \tag{26}$$

where $A'_{11} \equiv \alpha_{11}$ and $A'_{12} \equiv (\alpha_{12}\ ,\ \alpha_{13})$, and α_{11}, α_{12} and α_{13} are the structural parameters ($a = 3$). Equation (26) provides a moment condition of the form (12), i.e. $A'_1\Sigma_u A_1 = 1$. Moreover, if as in Assumption 4 there exist $s = n - k = 2$ proxies v_t for the auxiliary shocks $\varepsilon_{2,t} = ({\varepsilon_t}^{AD}, {\varepsilon_t}^{OSD})'$, we have two additional moment conditions of the form (13), i.e. $A'_1\Sigma_{u,v} = 0_{1\times 2}$, where $\Sigma_{u,v} := \text{E}(u_t v'_t)$. Overall, we have three moment conditions ($m = 3$) that can be used to estimate the three structural parameters in $A'_1 = (A'_{11},\ A'_{12}) \equiv (\alpha_{11}, \alpha_{12}, \alpha_{13})$ ($a = 3$) by the method discussed in Section 4.

Following Kilian (2009) and Montiel Olea, Stock and Watson's (2020), Assumption 1 is considered valid; Assumption 2 is investigated by a set of residuals diagnostic test on the estimated VAR (based on $l = 24$ lags) which suggest that the residuals are conditionally heteroskedastic but uncorrelated.[14] Assumption 3 is maintained. The strength of the proxies used for the auxiliary shocks under Assumption 4 is discussed next.

Indirect-MD approach, selection of proxies and strength. We employ the following proxies for the two auxiliary shocks: $v_t := (v_t^{RV}, v_t^{Br})'$, where v_t^{RV} is the log difference of the World Steal Index (WSI) introduced by Ravazzolo and Vespignani (2021) and is used as a proxy for the aggregate demand shock ${\varepsilon_t}^{AD}$, and v_t^{Br} is the log difference of Brent Oil Futures and is used as a proxy for the oil-specific demand shock ${\varepsilon_t}^{OSD}$. In this case the, v_t^{RV} is available on the sample 1990:M2-2004:M9 and, for comparative purposes and in line with the previous analysis, we re-estimate the proxy-SVAR on the common sample period 1990:M2-2004:M9 (which implies $T = 176$ monthly observations).

We pre-test the strength of the chosen proxies $v_t := (v_t^{RV}, v_t^{Br})'$ by our bootstrap-based test. Again, we apply DH multivariate normality test to the sequence of bootstrap replications $\{\hat{\vartheta}^*_{T:1}, \hat{\vartheta}^*_{T:2}, ..., \hat{\vartheta}^*_{T:N}\}$, where $N = [T]^{1/2} = 13$ and the estimator $\hat{\vartheta}^*_T$ is selected as follows. Let $\hat{\theta}^*_T = (\hat{\beta}^{*\prime}_{2,T}, \hat{\lambda}^*_T)'$ be the

[14] Results are available upon request.

MBB-CMD estimator discussed in Section 5.1: first, we consider the choice $\hat{\vartheta}_T^* \equiv \hat{\theta}_T^*$, obtaining a p-value of the DH multivariate normality test of 0.67; second, we take $\hat{\vartheta}_T^* \equiv \hat{\beta}_{2,T}^{*\prime}$, obtaining a p-value of DH multivariate normality test of 0.73. Thus, the bootstrap test supports the null hypothesis that the proxies $v_t := (v_t^{RV}, v_t^{Br})'$ are strong for the structural shocks $\varepsilon_{2,t} = (\varepsilon_t{}^{AD}, \varepsilon_t{}^{OSD})'$.

Indirect-MD approach, IRFs. The impulse responses estimated by the indirect-MD approach are given by the red lines plotted in Figure 2 and are surrounded by red shaded areas corresponding to 68%-MBB (panel A) and 95%-MBB (panel B) pointwise confidence intervals (Hall's percentile method). It is worth remarking that these bootstrap confidence intervals are 'conditional' on the bootstrap pre-test of instrument relevance failing to reject the hypothesis of strong proxies.

We notice two main facts. First, the 'conditional' MBB (pointwise) confidence intervals obtained by the indirect-MD approach using strong proxies for the auxiliary shocks are more informative than the Anderson-Rubin weak-identification-robust (pointwise) confidence intervals obtained by instrumenting the oil supply shock directly. The differences in the uncertainty surrounding point estimates in the two approaches become marked when considering 95% confidence intervals, see panel B. Second, our empirical results line up with Kilian's (2009) main findings. In Kilian's (2009) Choleski-SVAR, real economic activity and the real price of oil respond scantly, temporarily and not significantly to the oil supply shock, a result also evident from the IRFs estimated by the indirect-MD method. Actually, Kilian's (2009) recursive SVAR implies the restrictions $A_{12}' \equiv (\alpha_{12}\ ,\ \alpha_{13}) = (0, 0)$ in the structural equation (26), which can be tested by standard methods under the conditions of Proposition 2. A standard Wald-type test for these restrictions delivers a bootstrap p-value of 0.68 which confirms that the estimated structural equation in (26) is consistent with the first equation of Kilian's (2009) recursive SVAR.

6.2 Financial and macroeconomic uncertainty shocks

In this second empirical illustration we emphasize the merit of the indirect-MD approach in situations in which finding valid multiple proxies for multiple target shocks can be problematic. The objective is to track the dynamic causal effects produced by financial and macroeconomic uncertainty shocks ($k = 2$) on real economic activity. As in Ludvigson, Ma and Ng (2019), we consider a small VAR system including $n = 3$ variables: $Y_t := (U_{F,t}, U_{M,t}, a_t)'$, where $U_{F,t}$ is an index of financial uncertainty, $U_{M,t}$ is the index of macroeconomic uncertainty and a_t is a measure of real economic activity, say the growth rate of industrial production. Ludvigson *et al.* (2019) argue that the joint use of macroeconomic and financial uncertainty is crucial to disentangle the contributions of two distinct sources of uncertainty and study their pass-through to the business cycle. In this framework, $n - k = 1 < k$. The two uncertainty indexes are discussed in Ludvigson *et al.* (2019). We use the same data as in Ludvigson *et al.* (2019) and Angelini *et al.* (2019).[15]

Reduced form. We focus on the period 2008:M1-2015:M4 that we term the 'Great Recession + Slow Recovery' period, based on $T = 88$ monthly observations. The choice of focusing on the period after the Global Financial Crisis is motivated by the empirical results in Angelini *et al.* (2019), who identify three main (distinct) volatility regimes on a sample of monthly observations covering the period 1960-2015, the latter of which corresponds to our sample.

The reduced form VAR model for $Y_t := (U_{F,t}, U_{M,t}, a_t)'$ includes a constant and $l = 4$ lags. The specification is similar to that in Angelini and Fanelli (2019) who do not detect neither serial correlation nor conditionally heteroskedasticity in the VAR residuals on the period 2008:M1-2015:M4.

Direct approach, caveats. The target structural shocks are in $\varepsilon_{1,t} := (\varepsilon_{F,t}, \varepsilon_{M,t})'$ ($k = 2$), where $\varepsilon_{F,t}$ denotes the financial uncertainty shock and $\varepsilon_{M,t}$ the macroeconomic uncertainty shock. The auxiliary shock of the system is the 'non-uncertainty shock' $\varepsilon_{a,t}$ ($n - k = 1$) associated with real economic activity. The B-form of the model (the counterpart of (2)) is given by the system:

$$\underbrace{\begin{pmatrix} u_{F,t} \\ u_{M,t} \\ u_{a,t} \end{pmatrix}}_{u_t} = \underbrace{\begin{pmatrix} b_{F,F} & b_{F,M} \\ b_{M,F} & b_{M,M} \\ b_{a,F} & b_{a,B} \end{pmatrix}}_{B_1} \underbrace{\begin{pmatrix} \varepsilon_{F,t} \\ \varepsilon_{M,t} \end{pmatrix}}_{\varepsilon_{1,t}} + \underbrace{\begin{pmatrix} b_{F,a} \\ b_{M,a} \\ b_{a,a} \end{pmatrix}}_{B_2} \underbrace{(\varepsilon_{a,t})}_{\varepsilon_{2,t}}$$

where $u_t := (u_{F,t}, u_{M,t}, u_{a,t})'$ is the vector of VAR reduced form innovations. The notation used for the on-impact coefficients in B_1 (and B_2) appears obvious. Since $k = 2$, the (point-)identification of the two uncertainty shocks requires (at least) two proxies for the two structural shocks plus (at least) $\frac{1}{2}k(k-1) = 1$ additional parametric restriction. We borrow the restriction $b_{F,M} = 0$ from Angelini *et al.* (2019), who exploit two major changes in the volatility of the system in the period 1960-2015. This restriction, which implies that financial uncertainty does not reacts instantaneously (within the month) to the macro uncertainty shock, is incorporated in our proxy-SVAR.

In this setup the implementation of the 'direct' identification approach rises the challenge of finding two valid observable external instruments for the two

[15] The uncertainty indexes are taken from Sidney Ludvigson's web page: https://www.sydneyludvigson.com/data-and-appendixes .

uncertainty shocks in $\varepsilon_{1,t} := (\varepsilon_{F,t}, \varepsilon_{M,t})'$. Ludvigson *et al.* (2019) discuss this issue in the context of a novel identification strategy which combines 'external variable constraints' with inequality constraints. They use a measure of aggregate stock market return as a proxy for the financial uncertainty shocks and the log difference in the real price of gold as a proxy for the macro uncertainty shocks but, notably, in their framework proxies need not be neither 'strong' (in the sense of Definition 1(a)), nor exogenous to the non-instrumented structural shocks. To circumvent the problem of finding relevant and exogenous proxies for the two uncertainty shocks next we move to the indirect identification strategy.

Indirect-MD approach: specification. The identification of the uncertainty shocks $\varepsilon_{1,t} := (\varepsilon_{F,t}, \varepsilon_{M,t})'$ through our indirect identification strategy requires considering the A-form of the SVAR:

$$\underbrace{\begin{pmatrix} a_{F,F} & a_{F,M} \\ a_{M,F} & a_{M,M} \end{pmatrix}}_{A'_{11}} \underbrace{\begin{pmatrix} u_{F,t} \\ u_{M,t} \end{pmatrix}}_{u_{1,t}} + \underbrace{\begin{pmatrix} a_{F,a} \\ a_{M,a} \end{pmatrix}}_{A'_{12}} \underbrace{(u_{a,t})}_{u_{2,t}} = \underbrace{\begin{pmatrix} \varepsilon_{F,t} \\ \varepsilon_{M,t} \end{pmatrix}}_{\varepsilon_{1,t}} \tag{27}$$

which naturally provides the $\frac{1}{2}k(k+1) = 3$ moment conditions stemming from $A_1'\Sigma_u A_1 = I_2$. Since $n - k = 1$, it is necessary to find at least one external instrument for the real economic activity shock, i.e. a variable v_t $(s = n - k = 1)$ such that

$$v_t = \lambda\varepsilon_{a,t} + \omega_{vt} \tag{28}$$

where $\tilde{\varepsilon}_{2,t} \equiv \varepsilon_{2,t} = \varepsilon_{a,t}$, λ is the relevance parameter and ω_{vt} is an i.i.d. measurement error orthogonal to all structural shocks of the system. The relationship (28) is the counterpart of equation (10) in Assumption 4, and provides the two additional moment restrictions $A_1'\Sigma_{u,v} = 0_{2\times 1}$. The 'additional' restriction necessary for identification is the zero restriction (on B_1) $b_{F,M} = 0$, discussed above. This restriction can be mapped into the structural coefficients in the matrix $A_1' := (A_{11}' \vdots A_{12}')$ by using the relationship (9) and the adaptation to the MD estimation method discussed in the SM.

Jointly, the restrictions $A_1'\Sigma_u A_1 = I_2$ and $A_1'\Sigma_{u,v} = 0_{2\times 1}$ provide m =3+2=5 independent moment conditions of the type (12)-(13) which can be used to estimate the a =5 free structural parameters in the matrix $A_1' := (A_{11}' \vdots A_{12}')$. Next we discuss the choice of the proxy v_t for the real economic activity shock.

Indirect-MD approach: choice of the proxy for the auxiliary shock. To build a proxy v_t for the auxiliary shock $\varepsilon_{a,t}$, we follow the same route as in Angelini and Fanelli (2019) who consider the choice of an instrument for the real economic activity shock. Let $house_t$ be the log of new privately owned housing units started on the estimation period 2008:M1-2015:M4 (source: Fred). We take the 'raw' growth rate of new privately owned housing units started, $\Delta house_t$, and estimate an auxiliary dynamic linear regression model of the form $\Delta house_t = \mathrm{E}(\Delta house_t \mid \mathcal{F}_{t-1}) + er_t$, where $\mathcal{F}_{t-1}$ denotes the information set available to the econometrician at time $t-1$, and er_t can be interpreted as the 'innovation component' of the growth growth rate $\Delta house_t$. The residuals $\widehat{er}_t$, $t = 1, ..., T$ are used as the actual proxy for real economic activity shock, i.e. $v_t = \widehat{er}_t$.

Indirect-MD approach: instrument relevance. To pre-test whether $v_t = \widehat{er}_t$ is a strong proxy for the real economic activity shock, we compute our bootstrap test of instrument relevance. We apply DH multivariate normality test to the sequence of bootstrap replications $\{\hat{\vartheta}^*_{T:1}, \hat{\vartheta}^*_{T:2}, ..., \hat{\vartheta}^*_{T:N}\}$, where $\hat{\vartheta}^*_{T:b} \equiv \hat{\beta}^*_{2,T:b}, b = 1, ...N$, $N = [T]^{1/2} = 9$, $\hat{\beta}^*_{2,T}$ is the MBB-CMD estimator of the instantaneous impact of the real economic activity shock on the variables in Y_t. The p-value of the DH multivariate normality test is 0.38 and strongly supports the hypothesis that we are using a strong proxy for the 'non-uncertainty shock' $\varepsilon_{a,t}$ of the system.

Indirect-MD approach: IRFs. Once the model is estimated by the indirect-MD approach, we recover the IRFs of interest. The red lines in Figure 3 plots the estimated dynamic responses of the growth rate of the industrial production to the identified financial (upper panel) and macroeconomic (lower panel) uncertainty shocks over an horizon of 40 months. Responses refer to one-standard deviation uncertainty shocks and are surrounded by 90%-MBB (pointwise) confidence intervals (red shaded area; Hall's percentile method). Again, the bootstrap confidence intervals in Figure 1 are 'conditional' on the pre-test of instrument relevance failing to reject the hypothesis of strong proxies. To compare results with a benchmark, Figure 3 plots in blue the responses (always computed to one-standard deviation shocks) obtained by Angelini *et al.* (2019) via changes in volatility, i.e. without the use of external variables (see their Figure 5). The blue shaded area corresponds to the 90%-i.i.d. bootstrap confidence intervals Angelini *et al.* (2019) compute on the period 2008:M1-2015:M4.[16]

Two main facts emerge from Figure 3. First, despite the finding that the two uncertainty shocks have played a sizable role in curbing economic activity during the post-Great Recession period is robust to the two identification

[16] The MBB and iid bootstrap confidence intervals are comparable in this setup for two main reasons. First, the MBB is 'robust' in the sense that it is consistent also in the presence of iid VAR innovations, see Jentch and Lunford (2009b). Second, from the reduced form specification analysis of the estimated model, the VAR innovations are not found to display forms of weak dependence such as e.g. conditional heteroskedasticity on the sample period 2008:M1-2015:M4.

methods, it is possible to appreciate sizeable differences in the on-impact effect of the macroeconomic uncertainty shock on industrial production growth. Indeed, with the indirect-MD approach the (significant) peak response of the industrial production growth to the macroeconomic uncertainty shock is on-impact and is equal to -0.32%, with the changes-in-volatility approach the (significant) peak response occurs 5 months after the shock and is equal to -0.15%. The (significant) peak response of real economic activity to the financial uncertainty shock occurs 3 months after the shocks and is equal to -0.17%, a result similar to that obtained via the changes-in-volatility approach. Second, based on (pointwise) 90%-bootstrap confidence intervals, the inference on the effects of uncertainty shocks appears more precise with the indirect-MD approach.

7 Conclusions

Structural shocks can be identified by modeling the structural relationships that characterize the observed variables, the so-called 'A-form' in the language of SVARs. Proxy-SVARs can be also represented in A-form and this suggests that, under certain conditions, the target structural shocks can be recovered by instrumenting the auxiliary shocks of the system. We have designed an identification strategy and a MD estimation approach where the target structural shocks are recovered by instrumenting the auxiliary shocks. This strategy is convenient when finding strong proxies for the auxiliary shocks is easier than finding strong proxies for the target structural shocks, or when the implementation of weak-identification-robust methods requires a large number of additional restrictions.

The suggested indirect identification strategy has been complemented with a novel, computationally straightforward, diagnostic test for proxy relevance based on bootstrap resampling. Pre-testing the strength of the proxies by this novel test does not affect post-test inferences.

We have shown the empirical usefulness and relative merits of the suggested approach by re-visiting the estimation of some proxy-SVARs already analyzed in the literature.

References

Amisano, G. and Giannini, C. (1997), *Topics in Structural VAR Econometrics*, 2nd Edition, Springer, Berlin.

Andrews, I., Stock, J.H. and Sun, L. (2019), Weak instruments in instrumental variables regression: Theory and practice, *Annual Review of Economics* 11, 727-753.

Angelini, G. and Fanelli, L. (2019), Exogenous uncertainty and the identification of Structural Vector Autoregressions with external instruments, *Journal of Applied Econometrics* 34, 951-971.

Angelini, G., Bacchiocchi, E., Caggiano, G., and Fanelli, L. (2019), Uncertainty across volatility regimes, *Journal of Applied Econometrics* 34, 437-455.

Angelini, G., Cavaliere, G. and Fanelli, L. (2021), Bootstrap inference and diagnostics in state space models: with applications to dynamic macro models, *Journal of Applied Econometrics*, forthcoming.

Angelini, G., Cavaliere, G. and Fanelli, L. (2020), Bootstrap diagnostics in proxy-SVARs with weak proxies, *manuscript*, University of Bologna.

Arias, J.E., Rubio-Ramirez, J.F. and Waggoner, D.F. (2020), Inference in Bayesian Proxy-SVARs, *Journal of Econometrics*, forthcoming.

Blanchard, O. and Perotti, R. (2002), An empirical characterization of the dynamic effects of changes in government spending and taxes on output, *Quarterly Journal of Economics* 117, 1329–1368.

Bose, A. (1988), Edgeworth correction by bootstrap in Autoregressions, *Annals of Statistics* 16, 1709-1722.

Boubacar Mainnasara, Y. and Francq, C. (2011), Estimating structural VARMA models with uncorrelated but non-independent error terms, *Journal of Multivariate Analysis* 102, 496-505.

Brüggemann, R., Jentsch, C. and Trenkler, C. (2016), Inference in VARs with conditional volatility of unknown form, *Journal of Econometrics* 191, 69-85.

Caldara, D. and Kamps, C. (2017), The analytics of SVARs: A unified framework to measure fiscal multipliers, *Review of Economic Studies* 84, 1015-1040.

Cavaliere, G. and Georgiev, I. (2020), Inference under random limit bootstrap measures, *Econometrica* 88, 2547-2974.

Francq, C. and Raïssi, H. (2006), Multivariate portmanteau test for autoregressive models with uncorrelated but nonindependent errors, *Journal of Time Series Analysis* 28, 454-470.

Giacomini, R., Kitagawa, T. and Read, M. (2020), Robust Bayesian inference in Proxy SVARs, *Journal of Econometrics*, forthcoming.

Hahn, J. and Hausman, J. (2002), A new specification test for the validity of instrumental variables, *Econometrica* 70, 163-189.

Hall, P. (1992), *The bootstrap and Edgeworth expansion*, Springer-Verlag, Berlin.

Hausman, J., Stock, J.H. and Yogo, M. (2005), Asymptotic properties of the Hahn-Hausman test for weak instruments, *Economic Letters* 89, 333-342.

Horowitz, J.L. (2001), The bootstrap, *Handbook of Econometrics* Vol. 5, Heckman, J.J. and E. Leamer (eds.), Ch. 52.

Jentsch, C. and Lunsford, K.C. (2019a), The dynamic effects of personal and corporate income tax changes in the United States: Comment, *American Economic Review* 109, 2655-2678.

Jentsch, C. and Lunsford, K.C. (2019b), Asymptotic valid bootstrap inference for Proxy SVARs, *Working Paper*. Version dated May 2019.

Kilian, L. (1998), Small-sample confidence intervals for impulse response functions, *Review of Economics and Statistics* 80, 218-230.

Kilian, L. (2008), Exogenous oil supply shocks: How big are they and how much do they matter for the U.S. economy? *Review of Economics and Statistics* 90, 216-240.

Kilian, L. (2009), Not all oil shocks are alike: Disentangling demand and supply shocks in the crude oil market, *American Economic Review* 99, 1053-1069.

Lilliefors, H. (1967), On the Kolmogorov-Smirnov test for normality with mean and variance unknown, *Journal of the American Statistical Association* 62, 399-402.

Ludvigson, S.C., Ma, S. and Ng, S. (2019), Uncertainty and business cycles: exogenous impulse or endogenous response? *American Economic Journal: Macroeconomics*, forthcoming

Lütkepohl, H. (2005), *New introduction to multiple time series analysis*, Springer-Verlag, Berlin.

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

Mertens, K. and Montiel Olea, J. L. (2018), Marginal tax rates and income: New time series evidence, *Quarterly Journal of Economics* 133, 1803–1884.

Montiel Olea, J.L., Stock, J.H. and Watson, M.W. (2021), Inference in SVARs identified with an external instrument, *Journal of Econometrics*, forthcoming

Newey, W.K. and McFadden, D. (1994), Large sample estimation and hypothesis testing, *Handbook of Econometrics*, R.F. Engle and D.L. McFadden (eds.), Vol. IV, Chap. 36.

Ravazzolo, F., Vespignani, J. (2020), World steel production: A new monthly indicator of global real economic activity, *Canadian Journal of Economics*, forthcoming.

Staiger, D. and Stock, J. H. (1997), Instrumental variables regressions with weak instruments, *Econometrica* 65, 557-586.

Stock, J.H. (2008), What's New in Econometrics-Time Series, *Lecture 7. Structural VARs*, Cambridge, MA., NBER.

Stock, J.H. and Watson, M.W. (2012), Disentangling the channels of the 2007–2009 recession, in *Brookings Panel of Economic Activity*: Conference paper, pp. March 22– 23.

Stock J.H. and Yogo, M. (2005), Testing for weak instruments in linear IV regression. In *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, ed. D.W.K. Andrews, J.H. Stock, pp. 80–108. Cambridge, UK: Cambridge University Press.

Stock J.H. and Wright, J. (2000), GMM with weak identification, *Econometrica* 68, 1055–96

Zivot, E., Startz, R. and Nelson, C.R. (1998), Valid confidence regions and inference in the presence of weak instruments, *International Economic Review* 39, 1119-1146.

Rejection frequencies

Strong proxy				
	$T=250$		$T=1000$	
	$corr=59\%$		$corr=59\%$	
θ	DH	KS	DH	KS
$\beta_{2,1}$		0.05(0.06)		0.05(0.06)
$\beta_{2,2}$	0.05(0.05)	0.05(0.06)	0.05(0.05)	0.05(0.05)
$\beta_{2,3}$		0.05(0.05)		0.05(0.05)
λ		0.05(0.05)		0.05(0.05)
Moderately weak proxy				
	$T=250$		$T=1000$	
	$corr=25\%$		$corr=13\%$	
θ	DH	KS	DH	KS
$\beta_{2,1}$		0.21(0.24)		0.36(0.36)
$\beta_{2,2}$	0.22(0.20)	0.27(0.30)	0.80(0.64)	0.38(0.39)
$\beta_{2,3}$		0.20(0.24)		0.30(0.33)
λ		0.09(0.08)		0.10(0.11)
Weak proxy				
	$T=250$		$T=1000$	
	$corr=5\%$		$corr=2\%$	
θ	DH	KS	DH	KS
$\beta_{2,1}$		0.80(0.79)		0.93(0.93)
$\beta_{2,2}$	0.72(0.71)	0.85(0.85)	0.98(0.98)	0.95(0.96)
$\beta_{2,3}$		0.82(0.81)		0.95(0.95)
λ		0.24(0.24)		0.50(0.49)

Table 1: Monte Carlo results (details on the data generating processes may be found in the SM). Empirical rejection frequencies of the bootstrap test for strong against weak proxy, based on 20000 simulations and tuning parameter $N := [T^{1/2}]$. $corr = corr(\upsilon_t, \varepsilon_{2,t})$ is the correlation between the instrument υ_t and the structural shock $\varepsilon_{2,t}$. KS is Lilliefors' (1967) version of Kolgomorov-Smirnov univariate normality test; DH is Doornik and Hansen's (2008) multivariate normality test. Numbers in parentheses refer to GARCH-type VAR innovations (see SM). All tests are computed at the 5% nominal significance level.

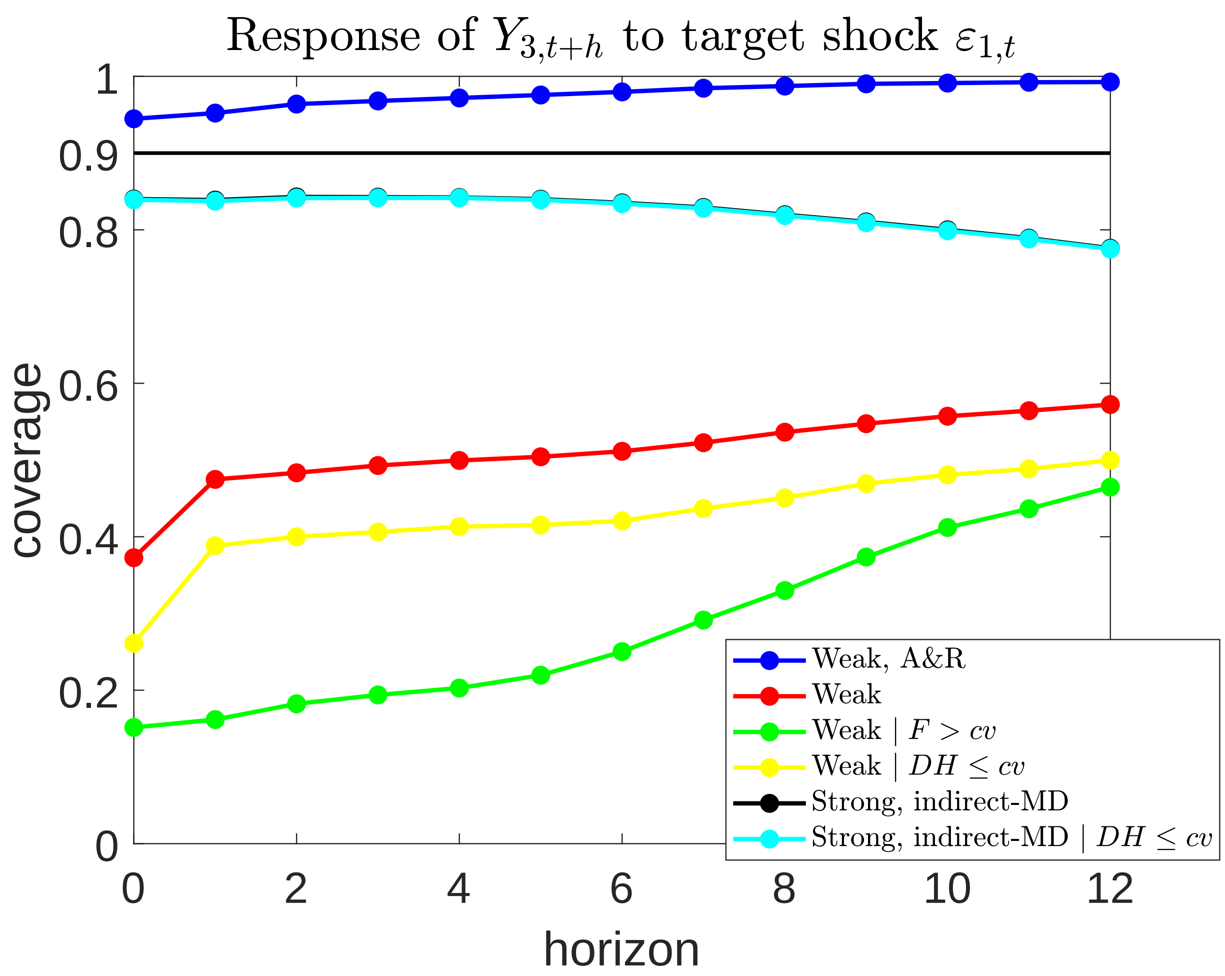

Figure 1: Monte Carlo results (details on the data generating processes may be found in the SM). Actual empirical coverage probabilities of IRFs calculated on 20000 simulations (90% nominal). IRFs refer to the response of the variable $Y_{3,t+h}$ to the target shock $\varepsilon_{1,t}$, $h = 0, 1, ..., 12$.

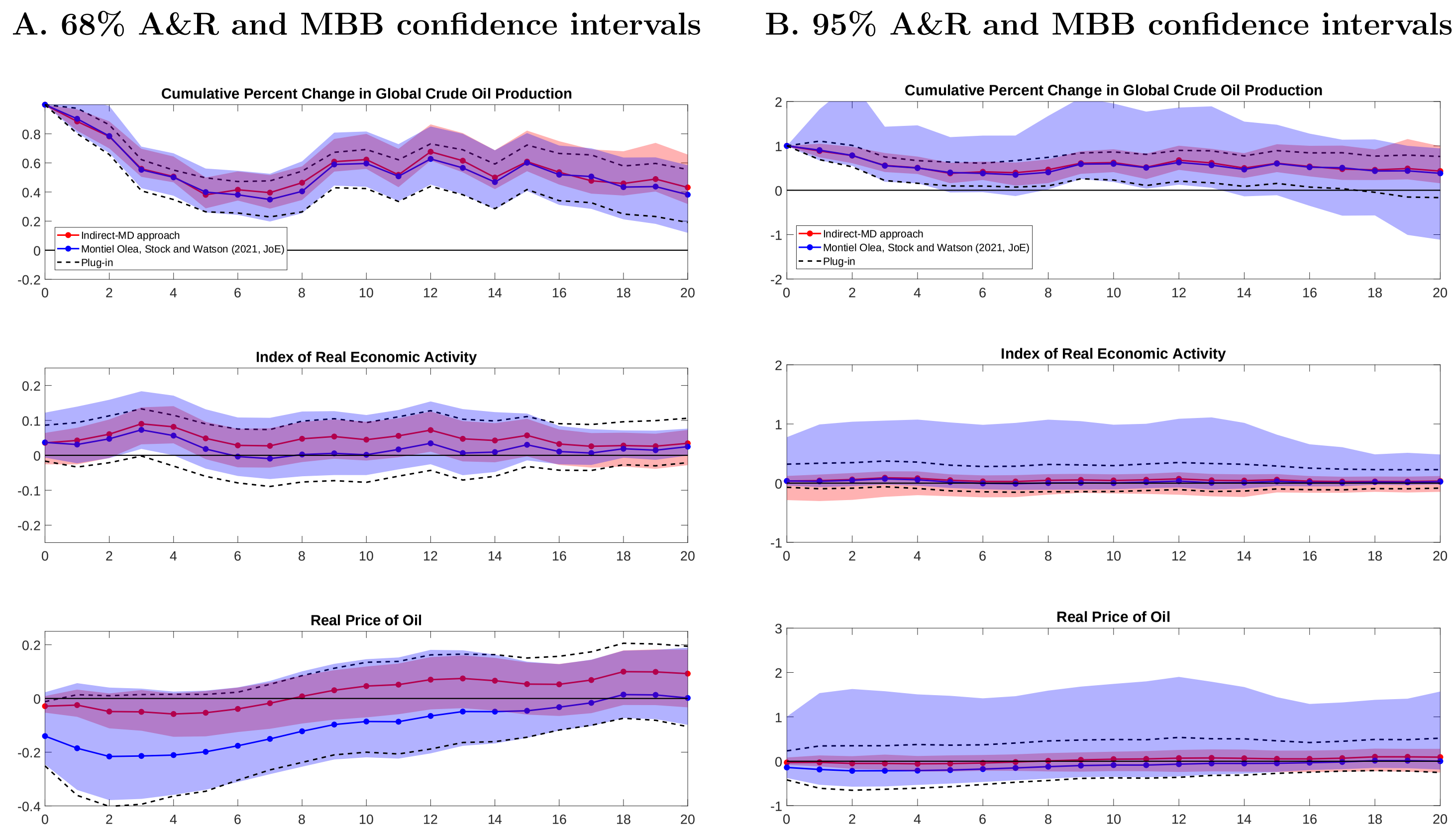

Figure 2: Impulse responses to an oil-supply shock.

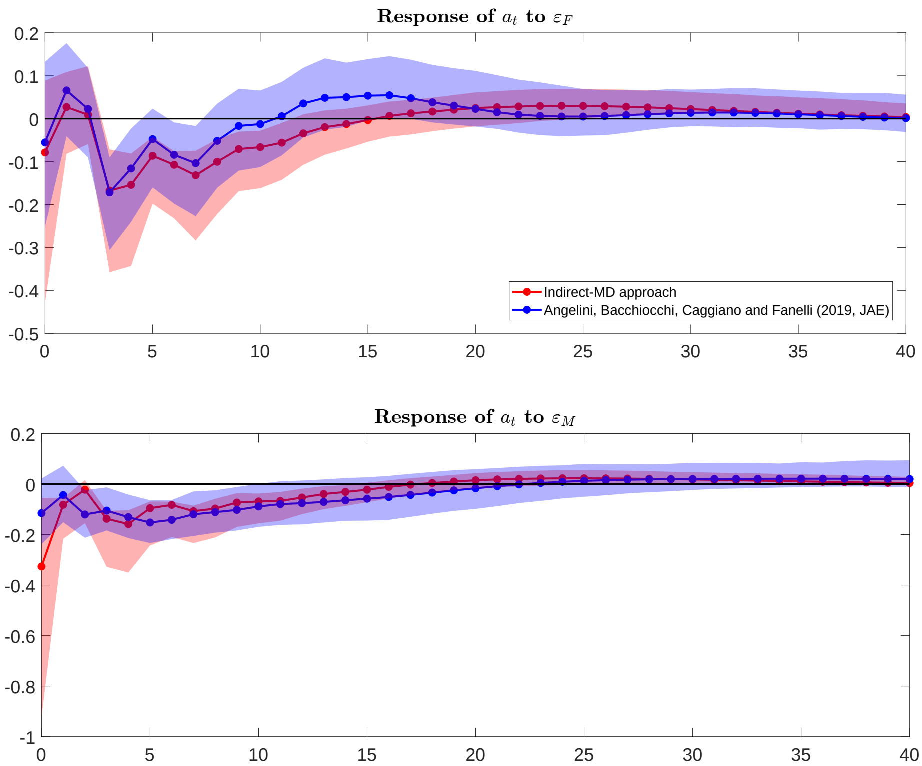

Figure 3: Impulse responses of a_t to a financial (ε_F) and a macro (ε_M) uncertainty shocks, with 90% bootstrap confidence intervals.

SUPPLEMENT TO
AN IDENTIFICATION STRATEGY FOR PROXY-SVARS WITH WEAK PROXIES

By Giovanni Angelini, Giuseppe Cavaliere, Luca Fanelli

September 2021

S.1 Introduction

This supplementary material complements the results of the paper along several dimensions. Section S.2 summarizes the notation used for the bootstrap. Section S.3 presents the auxiliary lemmas necessary to prove the main propositions of the paper and Section S.4 contains the proofs of all lemmas and propositions.

Section S.5 revisits the indirect-MD approach discussed in Section 4 of the paper under the conditions that the additional restrictions necessary for identification in the multiple target shocks scenario are placed on the matrix B_1 (B-form) rather than on the matrix A_1 (A-form). Section S.6 compares the MD estimation method with the IV approach. Section S.7 sketches the MBB algorithm frequently mentioned in the paper and used to build our test of instrument relevance. Section S.8 discusses in detail the data generating process used to produce the Monte Carlo results discussed in Section 5 of the paper. Section S.9 focuses on a simple proxy-SVAR specification which shows that when the exogeneity condition of proxies is violated, the estimator of the structural parameters is not consistent (i.e. centered on pseudo-true values) but its asymptotic distribution is still asymptotically Gaussian distributed. Finally, Section S.10 provides a third empirical illustration other the examples discussed in the paper and focuses in particular on the identification and estimation of fiscal multipliers from a fiscal proxy-SVAR.

S.2 Notation

We use P to denote the probability measure for the data, and use $\mathrm{E}(\cdot)$ and $Var(\cdot)$ to denote expectations and variance computed under P, respectively. We use P^* to denote the probability measure induced by the bootstrap, i.e. conditional on the original sample. Expectation and variance computed under P^* are denoted by $\mathrm{E}^*(\cdot)$ and $\mathrm{Var}^*(\cdot)$, respectively.

Let, for any $\varsigma > 0$, $p^*_T(\varsigma) := P^*(||\hat{\theta}^*_T - \hat{\theta}_T|| > \varsigma)$, where $\hat{\theta}^*_T$ is the bootstrap analog of the estimator $\hat{\theta}_T$, and $\|\cdot\|$ is the Euclidean norm. With the notation

‘ $\hat{\theta}^*_T - \hat{\theta}_T \xrightarrow{p^*}_p 0$’, which reads ‘$\hat{\theta}^*_T - \hat{\theta}_T$ convergences in P^* to 0, in probability’, we mean that the (stochastic) sequence $\{p^*_T(\varsigma)\}$ converges in probability to zero ($p^*_T(\varsigma) \xrightarrow{p} 0$).

Consider a scalar a random variable X, with associated cdfs $\digamma_X(x) := P(X \leq x)$; moreover, let the bootstrap sequence $\{X^*_T\}$, where X^*_T has associated cdf (conditional on the data) $\digamma^*_{X^*_T}(x) := P^*(X^*_T \leq x)$. We say that X^*_T ‘converges in conditional distribution to X, in probability’, denoted by ‘$X^*_T \xrightarrow{d^*}_p X$’ if $\digamma^*_{X^*_T}(x) \xrightarrow{p} \digamma_T(x)$ for each x at which $\digamma_X(x)$ is continuos. Notice that if $\digamma_X(\cdot)$ is continuous, then the latter convergence also implies that $\sup_{x\in\mathbb{R}} |\digamma^*_{X^*_T}(x) - \digamma_X(x)| \xrightarrow{p} 0$. These definitions can be extended to the multivariate framework in the conventional way.

S.3 Auxiliary lemmas

This section reports the lemmas useful for the developments of the paper. Preliminarily we represent the proxy-SVAR in a form that facilitates the derivation of the estimator of the reduced form parameters.

Estimator of the reduced form parameters. By coupling the VAR for Y_t in (1) of the paper with the proxies available for the auxiliary shocks v_t in (10) of the paper (Assumption 4), the proxy-SVAR model can be represented as the ‘large’, parametrically constrained, VAR system

$$\begin{pmatrix} I_n - \Pi(L) & 0 \\ 0 & I_s \end{pmatrix} \begin{pmatrix} Y_t \\ v_t \end{pmatrix} = \begin{pmatrix} u_t \\ v_t \end{pmatrix}, \; \Sigma_\eta := \begin{pmatrix} \Sigma_u & \Sigma_{u,v} \\ \Sigma_{v,u} & \Sigma_v \end{pmatrix} \qquad \text{(S.1)}$$

where $\Pi(L) := \Pi_1 L + \ldots + \Pi_l L^l$. System (S.1) maintains implicitly that the proxies in v_t are already expressed in innovation form, i.e. as uncorrelated processes. Actually, it may happen that the observed proxy v_t is autocorrelated and generated by a dynamic model of the form: $v_t = E_{t-1}v_t + \rho_{v,t}$, where $E_{t-1}v_t$ may depend on lags of v_t and Y_t, and $\rho_{v,t}$ is the unsystematic component in innovation form. In this case, (S.1) can be generalized to the representation

$$\begin{pmatrix} I_n - \Pi(L) & 0 \\ D_{v,y}(L) & I_s - D_{v,v}(L) \end{pmatrix} \begin{pmatrix} Y_t \\ v_t \end{pmatrix} = \begin{pmatrix} u_t \\ \rho_{v,t} \end{pmatrix}, \; \Sigma_\eta := \begin{pmatrix} \Sigma_u & \Sigma_{u,v} \\ \Sigma_{v,u} & \Sigma_v \end{pmatrix} \qquad \text{(S.2)}$$

where $D_{v,y}(L)$ and $D_{v,v}(L)$ are matrix polynomials in the lag operator assumed, without loss of generality, of order not larger than l, and such that the roots of the characteristic equation $\det(I_s - D_{vv}(x)) = 0$ satisfy the stability condition $|x| > 1$. This ensures, given Assumption 1 in the paper, that system (S.1) remains asymptotically stable. Regardless of whether we consider system

(S.1) or (S.2), the innovations $\eta_t := (u_t', v_t')'$ or $\eta_t := (u_t', \rho_{v,t}')'$ are α-mixing under Assumptions 2-4.

Irrespective of whether we consider system (S.1) or (S.2), we define the vector $W_t := (Y_t', v_t')'$ of dimension $(n+s) \times 1$ and compact the proxy-SVAR in the expression

$$W_t = \Psi_1 W_{t-1} + \Psi_2 W_{t-2} + ... + \Psi_l W_{t-l} + \eta_t \tag{S.3}$$

where each matrix of autoregressive (slope) parameters Ψ_i, $i = 1, ..., l$, has triangular and highly constrained structure. Henceforth, we denote with δ_ψ the vector that collects the autoregressive parameters that enter the matrices Ψ_i, $i = 1, ..., l$, and with δ_η the vector that collects the non-repeated elements of the covariance matrix Σ_η. Jointly, the reduced form parameters of the proxy-SVAR are in the vector $\delta := (\delta_\psi', \delta_\eta')'$ which has dimensions $q \times 1$. In particular, $q = q_\psi + q_\eta$, where $q_\psi = \dim(\delta_\psi)$ and $q_\eta = \dim(\delta_\eta)$. $\delta_0 := (\delta_{\psi,0}', \delta_{\eta,0}')'$ is the true value of δ and $\hat{\delta}_T := (\hat{\delta}_{\psi,T}', \hat{\delta}_{\eta,T}')'$ is the quasi-maximum likelihood [QML] estimator of δ.[1] Further, we consider a MBB analog of the QML estimator of $\delta := (\delta_\psi', \delta_\eta')'$, denoted $\hat{\delta}_T^* := (\hat{\delta}_{\psi,T}^{*\prime}, \hat{\delta}_{\eta,T}^{*\prime})'$. A sequence of N bootstrap replications of this estimator, $\left\{\hat{\delta}_{T:1}^*, ...\hat{\delta}_{T:N^*}^*\right\}$, can be obtained with the MBB algorithm sketched in Section S.7.

Lemma S1 deals with the asymptotic properties of the non-bootstrap and bootstrap estimator of the parameters $\delta := (\delta_\psi', \delta_\eta')'$.

Lemma S.1 *Consider the proxy-SVAR model summarized in (S.3). Under Assumptions 1, 2 and 4 of the paper:*
(i)

$$\hat{\delta}_T - \delta_0 \xrightarrow{p} 0_{q \times 1}; \tag{S.4}$$

$$T^{1/2} \begin{pmatrix} \hat{\delta}_{\psi,T} - \delta_{\psi,0} \\ \hat{\delta}_{\eta,T} - \delta_{\eta,0} \end{pmatrix} \xrightarrow{d} N(0_{q \times 1}, V_\delta) \ , \ V_\delta := \begin{pmatrix} V_\psi & V_{\psi,\eta} \\ V_{\psi,\eta}' & V_\eta \end{pmatrix}; \tag{S.5}$$

(ii) let ℓ be the block length of the MBB algorithm (Section S.7): under the additional condition $\ell^3/T \to 0$:

$$\hat{\delta}_T^* - \hat{\delta}_T \xrightarrow{p^*}_p 0_{q \times 1} \tag{S.6}$$

$$T^{1/2} V_\delta^{-1/2} \begin{pmatrix} \hat{\delta}_{\psi,T}^* - \hat{\delta}_{\psi,T} \\ \hat{\delta}_{\eta,T}^* - \hat{\delta}_{\eta,T} \end{pmatrix} \xrightarrow{d^*}_p N(0_{q \times 1}, I_g). \tag{S.7}$$

[1] The QML estimator of $\delta := (\delta_\psi', \delta_\eta')'$ is computed by maximizing the Gaussian quasi-likelihood associated with model (S.1) along the lines described e.g. in Section 3 in Boubacar Mainassara and Francq (2011). Observe, indeed, that the reduced form model in (S.3) reads as a special case of Boubacar Mainnasara and Francq's (2011) structural VARMA models.

Some remarks are in order.

Remark S.3.1 The results in Lemma S1 hold regardless of Assumption 4 in the paper, i.e. irrespective of whether the proxies are strong or weak. The asymptotic covariance matrix V_δ in (S.5) is specified in detail in Brüggemann, Jentsch and Trenkler (2016). It can be proved it has 'sandwich' form $V_\delta := \mathcal{A}_0^{-1}\mathcal{B}_0\mathcal{A}_0^{-1\prime}$, where $\mathcal{A}_0 := \lim_{T\to\infty}\left(\frac{\partial^2}{\partial\delta\partial\delta'}\log L_T(\delta_0)\right)$, $\mathcal{B}_0 := \lim_{T\to\infty}\mathrm{Var}\left(\frac{\partial}{\partial\delta}\log L_T(\delta_0)\right)$, and $\log L_T(\delta_0)$ is the Gaussian log-likelihood associated with the reduced form model in (S.1), see Theorem 1 in Boubacar Mainnasara and Francq (2011). A consistent estimator of V_δ has HAC-type form: $\hat{V}_\delta^{HAC} := \hat{\mathcal{A}}^{-1}\hat{\mathcal{B}}^{HAC}\hat{\mathcal{A}}^{-1\prime}$; Boubacar Mainnassara and Francq (2011) discuss the computation of $\hat{\mathcal{A}}$ and $\hat{\mathcal{B}}^{HAC}$, see in particular their Theorem 3.

Remark S.3.2 When Assumption 2 can be replaced with the stronger i.i.d. condition for η_t, or when η_t is a MDS ($\mathrm{E}(\eta_t \mid \mathcal{F}_{t-1}) = 0_{q\times 1}$) and is also conditionally homoskedastic ($\mathrm{E}(\eta_t\eta_t' \mid \mathcal{F}_{t-1}) = \Sigma_\eta$), one has $V_{\psi,\eta} = 0_{q_\psi\times q_\eta}$ in (S.5), which implies easily manageable expressions for the asymptotic covariance matrices V_ψ and V_η; for instance, $V_\eta := 2D_{q_\eta}^{+}(\Sigma_\eta\otimes\Sigma_\eta)D_{q_\eta}^{+\prime}$ when η_t is a conditionally homoskedastic MDS. The simulation studies in Brüggemann, Jentsch and Trenkler (2016) show that the MBB is 'robust' in the sense that it performs satisfactorily well in finite samples also when the true data generating process for $\eta_t = (u_t', \zeta_{\rho,t}')'$ is i.i.d. so that it would be 'natural' applying the residual-based i.i.d. bootstrap. In this respect, the MBB is 'robust' to α-mixing and i.i.d. conditions and as such it represents an ideal method of inference in proxy-SVARs.

The next lemma derives the asymptotic distribution of the estimator of the reduced form parameters of the proxy-SVAR, collected in the vector $\mu := (\mathrm{vech}(\Omega_v)', \mathrm{vec}(\Sigma_{v,u})')'$ where, recall, $\Omega_v := \Sigma_{v,u}\Sigma_u^{-1}\Sigma_{u,v}$. The parameters in μ play a crucial role in the development of the CMD estimator discussed in Section 5 of the paper in order to build our bootstrap test for instrument relevance.

In what follows, we exploit the functional dependence of μ on the $m\times 1$ vector $\sigma^+ := (\mathrm{vech}(\Sigma_u)', \mathrm{vec}(\Sigma_{v,u})')'$, where σ^+ is a 'piece' of δ_η, i.e. $\sigma^+ := M_{\sigma^+}\delta_\eta$, with M_{σ^+} full row rank selection matrix; see the vector $\delta := (\delta_\psi', \delta_\eta')'$ in Lemma S1. We define: $\mu := (\omega', \varpi')'$, where $\omega = \mathrm{vech}(\Omega_v)$ is $o_1\times 1$, $o_1 = \frac{1}{2}s(s+1)$, and $\varpi := vec(\Sigma_{v,u})$ is $o_2\times 1$, $o_2 = ns$; hence, μ is $o\times 1$,

$o = o_1 + o_2$. The estimator of μ is $\hat{\mu}_T := (\hat{\omega}_T', \hat{\varpi}_T')'$ and is obtained from $\hat{\delta}_{\eta,T}$ and Lemma S1. $\mu_0 = \mu_\sigma(\sigma_0^+) \equiv (\omega_0', \varpi_0')'$ denotes the true value of μ, and σ_0^+ the true value of σ^+. Finally, with $\mathcal{N}_{\Lambda_0}$ we denote a neighborhood of the true matrix Λ_0.

LEMMA S.2 *Consider the proxy-SVAR obtained by combining the SVAR (1) with the proxies v_t for the structural shocks $\tilde{\varepsilon}_{2,t}$ in (10). Under Assumptions 1,2 and 4 of the paper:*
(i) $\hat{\mu}_T \xrightarrow{p} \mu_0$, regardless of the strength of the proxies;
(ii) if the proxies v_t are strong in the sense of Definition 1(a) of the paper, then

$$T^{1/2}(\hat{\mu}_T - \mu_0) \xrightarrow{d} J_{\sigma^+}\mathbb{G}_{\sigma^+}$$

where $\mathbb{G}_{\sigma^+}$ denotes a $N(0, V_{\sigma^+})$ random variable, $V_{\sigma^+} := (M_{\sigma^+}V_\eta M_{\sigma^+}')$ with V_η defined in (S.5), and $J_{\sigma^+} := \frac{\partial \mu_{\sigma^+}}{\partial \sigma^{+\prime}}$ is an $o \times m$ Jacobian matrix such that $rank[J_{\sigma^+}] = o$.
(iii) if all proxies in v_t are weak in the sense of Definition 1(b) in the paper, the component $\hat{\omega}_T - \omega_0$ of the vector $\hat{\mu}_T - \mu_0$ is such that

$$T(\hat{\omega}_T - \omega_0) \xrightarrow{d} J^{(1)}\mathbb{G}_{\sigma^+} + \tfrac{1}{2}(I_{o_1} \otimes \mathbb{G}_{\sigma^+}')H_{\sigma^+}^{(1)}\mathbb{G}_{\sigma^+},$$

where $T^{1/2}J_{\sigma^+}^{(1)} \to J^{(1)}$, $J_{\sigma^+}^{(1)}$ is the $o_1 m \times m$ upper block of the Jacobian J_{σ^+}, and $H_{\sigma^+}^{(1)}$ is the $o_1 m \times m$ upper block of the $om \times m$ Hessian matrix $H_{\sigma^+} := \frac{\partial}{\partial \sigma^{+\prime}} \operatorname{vec}\left\{\left(\frac{\partial \mu_{\sigma^+}}{\partial \sigma^{+\prime}}\right)'\right\}$ and is different from zero.

REMARK S.3.3 Lemma S2(iii) shows that the asymptotic distribution of $T(\hat{\omega}_T - \omega_0)$ is a mixture of Gaussian and χ^2-type random variables. Thus, $T^{1/2}(\hat{\omega}_T - \omega_0) \xrightarrow{p} 0_{o\times 1}$, result that holds also with completely irrelevant proxies. It follows that $T(\hat{\mu}_T - \mu_0)$ is asymptotically non-Gaussian. Lemma S2 refers to the case in which all s proxies included in v_t are weak. In practice, we might have subsets of weak and strong proxies in v_t. In this case, it is possible to prove that the asymptotic distribution of $\hat{\mu}_T$ is still not Gaussian; results are available upon request to the authors.

The next Lemma shows that the asymptotic distribution of $T^{1/2}(\hat{\theta}_T - \theta_0)$, where $\hat{\theta}_T$ is the CMD estimator

$$\hat{\theta}_T := \arg\min_{\theta \in \mathcal{T}_\theta} \hat{Q}_T(\theta), \quad \hat{Q}_T(\theta) := (\hat{\mu}_T - f(\theta))'\hat{V}_\mu^{-1}(\hat{\mu}_T - f(\theta)) \tag{S.8}$$

(see Section 4 of the paper) depends on whether the proxies are strong or weak in the sense of Definition 1 in the paper. In what follows, $\mathcal{N}_{\theta_0}$ is a neighborhood of θ_0 and $\mathcal{P}_\theta$ is the compact (dense) parameter space.

Lemma S.3 *Consider the proxy-SVAR obtained by combining the SVAR (1) with the proxies* v_t *for the structural shocks* $\tilde{\varepsilon}_{2,t}$ *in (10). Then consider the CMD estimator* $\hat{\theta}_T$ *obtained from the problem (19) in the paper. Let* θ_0 *be an interior point of* $\mathcal{P}_\theta$*. Then, under Assumptions 1-4:*
(i) if the proxies are strong and the rank condition for identification holds in $\mathcal{N}_{\theta_0}$*, then* $T^{1/2}(\hat{\theta}_T - \theta_0) \xrightarrow{d} N(0, V_\theta)$*,* $V_\theta := \left(J_\theta' V_\mu^{-1} J_\theta\right)^{-1}$*, where* J_θ *is a Jacobian matrix of full column rank in* $\mathcal{N}_{\theta_0}$*;*
(ii) if the proxies are weak (or completely irrelevant), $T^{1/2}(\hat{\theta}_T - \theta_0)$ *is not asymptotically Gaussian.*

S.4 Proofs of lemmas and propositions

S.4.1 Proof of Lemma S1

(i) The result follow from Theorem 1 in Boubacar Mainnasara and Francq (2011) by setting the matrices B_{01}, ..., B_{0q} in their VARMA model in equation (3) equal to zero, and the matrices A_{00} and B_{00} equal to the identity matrix; see also Theorem 2.1 in Brüggemann, Jentsch and Trenkler (2016). (ii) The result follows from Theorem 4.1 in Brüggemann, Jentsch and Trenkler (2016); see also Theorem 3.2 in Jentsch and Lunsford (2019b). ■

S.4.2 Proof of Lemma S2

(i) $\mu = \mu_{\sigma^+}(\sigma^+)$ is a smooth function of σ^+ and therefore of δ_η (recall that $\sigma^+ = M_{\sigma^+}\delta_\eta$, M_{σ^+} being a selection matrix of full row rank). The result follows from Lemma S1(i) and the Slutsky Theorem.

(ii) Since $\sigma^+ = M_{\sigma^+}\delta_\eta$, Lemma S1(i) implies that

$$T^{1/2}(\hat{\sigma}_T^+ - \sigma_0^+) \xrightarrow{d} N(0, V_{\sigma^+}), \;\; V_{\sigma^+} := M_{\sigma^+} V_\eta M_{\sigma^+}' \tag{S.9}$$

where $\hat{\sigma}_T^+ := M_{\sigma^+}\hat{\delta}_{\eta,T}$, $\sigma_0^+ := M_{\sigma^+}\delta_{\eta,0}$ and V_{σ^+} is positive definite. We consider the following quadratic expansion of $\hat{\mu}_T = \mu_{\sigma^+}(\hat{\sigma}_T^+)$ around σ_0^+:

$$T^{1/2}\left(\hat{\mu}_T - \mu_0\right) = J_{\sigma_0^+}(\sigma_0^+) T^{1/2}(\hat{\sigma}_T^+ - \sigma_0^+) + \tfrac{1}{2} T^{1/2} R_T(\ddot{\sigma}_T^+) \tag{S.10}$$

where $J_{\sigma_0^+}(\sigma_0^+)$ is the $o \times m$ Jacobian matrix $J_{\sigma_0^+} := \frac{\partial \mu_{\sigma^+}}{\partial \sigma^{+\prime}}$ evaluated at σ_0^+, and the remainder term $R_T(\ddot{\sigma}_T^+)$ has representation:

$$R_T(\ddot{\sigma}_T^+) := \left(I_o \otimes (\hat{\sigma}_T^+ - \sigma_0^+)'\right) H_{\sigma^+}(\ddot{\sigma}_T^+)(\hat{\sigma}_T^+ - \sigma_0^+),$$

$$H_{\sigma^+}(\ddot{\sigma}_T^+) := \frac{\partial}{\partial \sigma^{+\prime}} \operatorname{vec}\left\{ \left(\frac{\partial \mu_{\sigma^+}}{\partial \sigma^{+\prime}} \right)' \Bigg|_{\sigma^+ = \ddot{\sigma}_T^+} \right\}$$

where $H_{\sigma^+}(\ddot{\sigma}_T^+)$ is the $om \times m$ Hessian evaluated at $\ddot{\sigma}_T^+$, an intermediate vector value between $\hat{\sigma}_T^+$ and σ_0^+. By construction, the last o_2 components of $T^{1/2}(\hat{\mu}_T - \mu_0)$ coincide with the last element of $T^{1/2}(\hat{\sigma}_T^+ - \sigma_0^+)$ (i.e. $T^{1/2}(\hat{\varpi}_T - \varpi_0)$), hence the structures of the Jacobian $J_{\sigma_0^+}(\sigma_0^+)$ and of the remainder term $R_T(\ddot{\sigma}_T^+)$ in (S.10) are given by

$$J_{\sigma_0^+}(\sigma_0^+) := \begin{pmatrix} J_{\sigma_0^+}^{(1)} \\ J_{\sigma_0^+}^{(2)} \end{pmatrix} \equiv \begin{pmatrix} J_{\sigma_0^+}^{(1,1)} & J_{\sigma_0^+}^{(1,2)} \\ 0 & I_{ns} \end{pmatrix} \tag{S.11}$$

and

$$R_T(\ddot{\sigma}_T^+) \equiv \begin{pmatrix} R_{1,T}(\ddot{\sigma}_T^+) \\ 0 \end{pmatrix} \begin{matrix} o_1 \times 1 \\ o_2 \times 1 \end{matrix} \tag{S.12}$$

where

$$R_{1,T}(\ddot{\sigma}_T^+) := \left(I_{o_1} \otimes (\hat{\sigma}_T^+ - \sigma_0^+)' \right) H_{\sigma^+}^{(1)}(\ddot{\sigma}_T^+)(\hat{\sigma}_T^+ - \sigma_0^+),$$

and $H_{\sigma^+}^{(1)}(\ddot{\sigma}_T^+) := \frac{\partial}{\partial \sigma^{+\prime}} \operatorname{vec}\left[J_{\ddot{\sigma}_T^+}^{(1)\prime} \right]$ is the $o_1 m \times m$ upper block of $H_{\sigma^+}(\ddot{\sigma}_T^+)$.

To prove the result, we show that in (S.10) $J_{\sigma_0^+}(\sigma_0^+)$ is constant and has full row rank, and that the remainder term $\frac{1}{2}T^{1/2}R_T(\ddot{\sigma}_T^+)$ is $o_p(1)$ as $\hat{\sigma}_T^+$ (and hence $\ddot{\sigma}_T^+$) converges in probability to σ_0^+.

By using standard matrix derivative rules (Magnus and Neudecker, 1999) it follows that the blocks $J_{\sigma_0^+}^{(1,1)}$ and $J_{\sigma_0^+}^{(1,2)}$ are given by the expressions

$$J_{\sigma_0^+}^{(1,1)} := -D_s^+ \left(\Sigma_{v,u}\Sigma_u^{-1} \otimes \Sigma_{v,u}\Sigma_u^{-1} \right) D_n \ ; \ J_{\sigma_0^+}^{(1,2)} := 2D_s^+(\Sigma_{v,u}\Sigma_u^{-1} \otimes I_s). \tag{S.13}$$

Without loss of generality (ordering is not crucial for the arguments that follow), partition the matrix B as $B = (\tilde{B}_1 \vdots \tilde{B}_2)$, where $\tilde{B}_1$ collects the columns of B associated with the $n-s$ non-instrumented auxiliary structural shocks of the system. Likewise, partition the matrix $A = B^{-1}$ as $A = \begin{pmatrix} \tilde{A}_1' \\ \tilde{A}_2' \end{pmatrix}$, where $\tilde{A}_1'$ is the block associated with the $n-s$ non-instrumented auxiliary structural shocks and $\tilde{A}_2'$ is the block associated with s instrumented structural shocks; $rank[\tilde{A}_2'] = s$ under Assumption 3. By imposing the proxy-SVAR restrictions $\Sigma_{v,u} = \Lambda \tilde{B}_2'$ and $\Sigma_u = BB'$, and using the above partitions one has $\Sigma_{v,u}\Sigma_u^{-1} = \Lambda \tilde{B}_2'(BB')^{-1} = \Lambda(0 \vdots I_s)A = \Lambda \tilde{A}_2'$, so that at the true parameter value the Jacobian in (S.11) is equal to

$$J_{\sigma^+}(\sigma_0^+) := \begin{pmatrix} -D_s^+ \left(\Lambda\tilde{A}_2' \otimes \Lambda\tilde{A}_2'\right) D_n & 2D_s^+(\Lambda\tilde{A}_2' \otimes I_s) \\ 0 & I_{ns} \end{pmatrix} \tag{S.14}$$

and it is therefore seen that it is constant and of full column rank if $rank[\Lambda] = s$ in $\mathcal{N}_{\Lambda_0}$, i.e. if the proxies are strong.

To prove that the remainder term $\frac{1}{2}T^{1/2}R_T(\ddot{\sigma}_T^+)$ is $o_p(1)$ as $\hat{\sigma}_T^+$ (and hence $\ddot{\sigma}_T^+$) converges in probability to σ_0^+, we have to prove that the block $H_{\sigma^+}^{(1)}(\ddot{\sigma}_T^+) := \frac{\partial}{\partial\sigma^{+\prime}} \operatorname{vec}\left[J_{\ddot{\sigma}_T^+}^{(1)\prime}\right]$ of the Hessian in (S.12) does not depend on T. It is useful to note that

$$H_{\sigma^+}^{(1)}(\ddot{\sigma}_T^+)' := \frac{\partial}{\partial\sigma^{+\prime}} \operatorname{vec}\left[J_{\ddot{\sigma}_T^+}^{(1)}\right] \equiv \begin{pmatrix} \frac{\partial}{\partial\sigma^{+\prime}} \operatorname{vec}\left[J_{\ddot{\sigma}_T^+}^{(1,1)}\right] \\ \frac{\partial}{\partial\sigma^{+\prime}} \operatorname{vec}\left[J_{\ddot{\sigma}_T^+}^{(1,2)}\right] \end{pmatrix} \equiv \begin{pmatrix} H_{11}^{(1)} & H_{12}^{(1)} \\ H_{21}^{(1)} & H_{22}^{(1)} \end{pmatrix} \tag{S.15}$$

and that after applying standard matrix derivative rules, the derivatives $H_{11}^{(1)} := \frac{1}{\partial \operatorname{vech}(\Sigma_u)'}\partial \operatorname{vec}\left[J_{\ddot{\sigma}_T^+}^{(1,1)}\right]$, $H_{12}^{(1)} := \frac{1}{\partial \operatorname{vec}(\Sigma_{v,u})'}\partial \operatorname{vec}\left[J_{\ddot{\sigma}_T^+}^{(1,1)}\right]$, $H_{21}^{(1)} := \frac{1}{\partial \operatorname{vech}(\Sigma_u)'}\partial \operatorname{vec}\left[J_{\ddot{\sigma}_T^+}^{(1,2)}\right]$ and $H_{22}^{(1)} := \frac{1}{\partial \operatorname{vec}(\Sigma_{v,u})'}\partial \operatorname{vec}\left[J_{\ddot{\sigma}_T^+}^{(1,2)}\right]$ are function of Σ_u and $\Sigma_{v,u}$, hence do not depend on T if the proxies are strong.

The asymptotic normality result follows from (S.10), the result $J_{\sigma_0^+}(\sigma_0^+)T^{1/2}(\hat{\sigma}_{0,T}^+ - \sigma_0^+) \xrightarrow{d} J_{\sigma^+}\mathbb{G}_{\sigma^+}$ and the fact that the term $\frac{1}{2}T^{1/2}R_T(\ddot{\sigma}_T^+)$ in the expansion (S.10) is $o_p(1)$.

(iii) We isolate the block associated with $T^{1/2}\left(\hat{\omega}_T - \omega_0\right)$ and consider, for $\Lambda \neq 0_{s\times s}$, the expansion (S.10):

$$T^{1/2}\left(\hat{\omega}_T - \omega_0\right) = \left(J_{\sigma_0^+}^{(1,1)} \vdots J_{\sigma_0^+}^{(1,2)}\right) T^{1/2}(\hat{\sigma}_{0,T}^+ - \sigma_0^+) + \tfrac{1}{2}T^{1/2}R_{1,T}(\ddot{\sigma}_T^+). \tag{S.16}$$

We show that if the instruments v_t are weak for $\tilde{\varepsilon}_{2,t}$, then for $T \to \infty$:

$$\begin{aligned} T\left(\hat{\omega}_T - \omega_0\right) &= \underbrace{T^{1/2}\left(J_{\sigma_0^+}^{(1,1)} \vdots J_{\sigma_0^+}^{(1,2)}\right)}_{=J^{(1)}+o(1)} \underbrace{T^{1/2}(\hat{\sigma}_{0,T}^+ - \sigma_0^+)}_{O_p(1)} \\ &\quad + \tfrac{1}{2}(I_{o_1} \otimes \underbrace{T^{1/2}(\hat{\sigma}_{0,T}^+ - \sigma_0^+)'}_{O_p(1)}) H_{\sigma^+}^{(1)}(\ddot{\sigma}_T^+) \underbrace{T^{1/2}(\hat{\sigma}_{0,T}^+ - \sigma_0^+)}_{O_p(1)} \end{aligned} \tag{S.17}$$

with $J^{(1)} := T^{1/2}J_{\sigma^+}^{(1)} \equiv T^{1/2}\left(J_{\sigma_0^+}^{(1,1)} \vdots J_{\sigma_0^+}^{(1,2)}\right)$ and $H_{\sigma^+}^{(1)}(\ddot{\sigma}_T^+) \neq 0$ and does not depend on T.

We start by proving that in the expansion (S.17), $T^{1/2}\left(J^{(1,1)}_{\sigma_0^+} \vdots J^{(1,2)}_{\sigma_0^+}\right) \to J^{(1)}$, with $J^{(1)}$ independent of T. From (S.13) and (S.14) we have

$$\begin{aligned} & T^{1/2}\left(J^{(1,1)}_{\sigma_0^+} \vdots J^{(1,2)}_{\sigma_0^+}\right) \\ = \; & T^{1/2}\left(-D_s^+\left(\Lambda\tilde{A}_2' \otimes \Lambda\tilde{A}_2'\right)D_n \vdots 2D_s^+(\Lambda\tilde{A}_2' \otimes I_s)\right) \\ & = T^{1/2}D_s^+(\Lambda\tilde{A}_2' \otimes I_s)\left(-\left(I_s \otimes \Lambda\tilde{A}_2'\right)D_n \vdots 2I_{s^2}\right). \end{aligned}$$

If all s instruments in v_t are weak in the sense of Definition 1(b), $\Lambda := T^{-1/2}C$, where C is $s \times s$ constant and different from the zero. It turns out that:

$$\begin{aligned} T^{1/2}\left(J^{(1,1)}_{\sigma_0^+} \vdots J^{(1,2)}_{\sigma_0^+}\right) & := T^{1/2}D_s^+(T^{-1/2}C\tilde{A}_2' \otimes I_s) \\ & \times \left[-\left(I_s \otimes T^{-1/2}C\tilde{A}_2'\right)D_n \vdots 2(I_s \otimes I_s)\right] \end{aligned}$$

so that as $T \to \infty$:

$$T^{1/2}\left(J^{(1,1)}_{\sigma_0^+} \vdots J^{(1,2)}_{\sigma_0^+}\right) \to J^{(1)} := D_s^+(C\tilde{A}_2' \otimes I_s)\left[0 \vdots 2I_{s^2}\right].$$

where $J^{(1)}$ does not depends on T .

Next we show that in the expansion (S.17), $H^{(1)}_{\sigma^+}(\ddot{\sigma}_T^+) \neq 0$ and does not depend on T. From the inspection of the matrix in (S.15) it follows that while $H^{(1)}_{11}$, $H^{(1)}_{22}$ and $H^{(1)}_{21}$ depend on $\Sigma_{v,u} = \Lambda\tilde{B}_2' = T^{-1/2}C\tilde{B}_2'$ and converge to zero as $T \to \infty$, $H^{(1)}_{22}$ solely depends on Σ_u, hence $H^{(1)}_{22} \neq 0$.

Finally, if $\Lambda = 0_{s\times s}$, i.e. the instruments v_t are completely irrelevant for $\tilde{\varepsilon}_{2,t}$, then $\hat{\omega}_T \overset{p}{\to} 0$; the first term in the expansion (S.17) is zero so that $T\hat{\omega}_T = O_p(1)$ and $T^{1/2}\hat{\omega}_T \overset{p}{\to} 0$.■

S.4.3 Proof of Lemma S3

The proof is based on three premises.

First, given the distance function $\mu - f(\theta) = 0$ upon which the CMD estimation problem in (19) is based, standard matrix derivative rules show that the Jacobian matrix $J_\theta := \frac{\partial f(\theta)}{\partial \theta'}$ has the following structure:

$$J_\theta := \begin{pmatrix} 2D_s^+(\Lambda \otimes I_s) & 0 \\ (\tilde{B}_2 \otimes I_s) & K_{ns}(\Lambda \otimes I_s) \end{pmatrix}\begin{pmatrix} S_\Lambda & 0 \\ 0 & S_{\tilde{B}_2} \end{pmatrix} \qquad \text{(S.18)}$$

where, for $s > 1$, S_Λ and $S_{\tilde{B}_2}$ are selection matrices used to impose the additional identification restrictions on the parameters $(\tilde{B}_2' \vdots \Lambda')'$ necessary other than the proxies. It turns out that under strong proxies (Definition 1(a) of the paper) J_θ has full column rank in $\mathcal{N}_{\theta_0}$, while under weak proxies (Definition 1(b) of the paper) J_θ has reduced rank in $\mathcal{N}_{\theta_0}$.

Second, given the CMD problem in (19), the consistency result $\hat{\theta}_T \xrightarrow{p} \theta_0$ is obtained under the strong proxies hypothesis by the same arguments used in the proof of Proposition 2(i) to establish the consistency of the MD estimator $\hat{\alpha}_T$.

Third, the first-order conditions associated with the problem (S.8) are given by

$$J_{\hat{\theta}_T}' \hat{V}_\mu^{-1}(\hat{\mu}_T - f(\hat{\theta}_T)) = 0$$

where $J_{\hat{\theta}_T}$ is the Jacobian (S.18) evaluated at the CMD estimator. By using a mean-value expansion of $f(\hat{\theta}_T)$ around θ_0, the first-order condition become

$$J_{\hat{\theta}_T}' \hat{V}_\mu^{-1}(\hat{\mu}_T - \mu_0 - J_{\dot{\theta}}(\hat{\theta}_T - \theta_0)) = 0$$

where $\dot{\theta}$ is an intermediate vector between $\hat{\theta}_T$ and θ_0, and $\mu_0 = f(\theta_0)$. By re-arranging the expression above we obtain the relationship

$$\left\{ J_{\hat{\theta}_T}' \hat{V}_\mu^{-1} J_{\dot{\theta}} \right\} T^{1/2}(\hat{\theta}_T - \theta_0) = J_{\hat{\theta}_T}' \hat{V}_\mu^{-1} T^{1/2}(\hat{\mu}_T - \mu_0) \tag{S.19}$$

which shows that the asymptotic distribution of $T^{1/2}(\hat{\theta}_T - \theta_0)$ depends on two main components: the asymptotic distribution of $T^{1/2}(\hat{\mu}_T - \mu_0)$, derived in Lemma S2, and the property of the matrix $\left\{ J_{\hat{\theta}_T}' \hat{V}_\mu^{-1} J_{\dot{\theta}} \right\}$ for $T \to \infty$.

(i) Under strong proxies, the consistency result $\hat{\theta}_T \xrightarrow{p} \theta_0$ implies that $J_{\hat{\theta}_T} \xrightarrow{p} J_{\theta_0}$ and $J_{\dot{\theta}} \xrightarrow{p} J_{\theta_0}$; the asymptotic normality result follows from Lemma S2(i) which guaranties that $\hat{V}_\mu \xrightarrow{p} V_\mu$ (positive definite) and Lemma S2(ii).

(ii) To prove that under weak proxies (and completely irrelevant proxies) $T^{1/2}(\hat{\theta}_T - \theta_0)$ is not asymptotically Gaussian it suffices to consider the expression in (S.19) and apply Lemma S2(iii).■

S.4.4 Proof of Proposition 1

(i) Under Assumptions 1-2 and 4, $\hat{\sigma}_T^+ \xrightarrow{p} \sigma_0^+$ by Lemma S1(i), hence, by the Slutsky Theorem, $\hat{g}_T(\hat{\sigma}_T^+, \alpha) \xrightarrow{p} g(\sigma_0^+, \alpha)$. Since $\hat{V}_{\sigma^+}$ is a consistent estimator of V_{σ^+}, for any $\bar{\alpha} \in \mathcal{P}_\alpha$, $\hat{Q}_T(\alpha) := \hat{g}_T(\hat{\sigma}_T^+, \alpha)'\hat{V}_{gg}(\bar{\alpha})^{-1}\hat{g}_T(\hat{\sigma}_T^+, \alpha) \xrightarrow{p} Q_0(\alpha) := g(\sigma_0^+, \alpha)'V_{gg,0}^{-1}(\bar{\alpha})g(\sigma_0^+, \alpha)$, where $V_{gg,0}(\bar{\alpha}) = G_{\sigma^+}(\sigma_0^+, \bar{\alpha})V_{\sigma^+}G_{\sigma^+}(\sigma_0^+, \bar{\alpha})'$ is positive definite as the Jacobian matrix $G_{\sigma^+}(\sigma^+, \alpha)$ is $m \times m$ and nonsingular for any σ^+. To see that $G_{\sigma^+}(\sigma^+,\alpha)$ is nonsingular, by applying standard derivative rules (Magnus and Neudecker, 1999) one has

$$G_{\sigma^+}(\sigma^+,\alpha) := \underset{m\times m}{\frac{\partial g(\sigma^+,\alpha)}{\partial \sigma^{+\prime}}} = \begin{pmatrix} \frac{\partial vech(A_1'\Sigma_u A_1 - I_k)}{\partial \sigma^{+\prime}} \\ \frac{\partial vec(A_1'\Sigma_{u,v})}{\partial \sigma^{+\prime}} \end{pmatrix} = \begin{pmatrix} D_k^+ \frac{\partial vec(A_1'\Sigma_u A_1 - I_k)}{\partial \sigma^{+\prime}} \\ \frac{\partial vec(A_1'\Sigma_{u,v})}{\partial \sigma^{+\prime}} \end{pmatrix}$$

$$= \begin{pmatrix} D_k^+ \frac{\partial vec(A_1'\Sigma_u A_1 - I_k)}{vech(\Sigma_u)'} & D_k^+ \frac{\partial vec(A_1'\Sigma_u A_1 - I_k)}{vech(\Sigma_{u,v})'} \\ \frac{\partial vec(A_1'\Sigma_{u,v})}{vech(\Sigma_u)'} & \frac{\partial vec(A_1'\Sigma_{u,v})}{vech(\Sigma_u)'} \end{pmatrix}$$

$$= \begin{pmatrix} D_k^+(A_1' \otimes A_1')D_n & 0 \\ 0 & (I_s \otimes A_1') \end{pmatrix} \tag{S.20}$$

so that it is seen that $G_{\sigma^+}(\sigma^+,\alpha)$ does not depend on σ^+ and is invertible because $rank[A_1'] = k$ (Assumption 3). Since $V_{gg,0}^{-1}(\bar{\alpha})$ is nonsingular, the condition for $Q_0(\alpha)$ to have a unique minimum (of zero) in $\mathcal{N}_{\alpha_0}$ is that the first derivative of $Q_0(\alpha)$, $G_\alpha(\sigma_0^+,\alpha)'V_{gg,0}^{-1}(\bar{\alpha})g(\sigma_0^+,\alpha)$, satisfies the condition $rank[G_\alpha(\sigma^+,\alpha)'V_{gg,0}^{-1}(\bar{\alpha})] = rank[G_\alpha(\sigma^+,\alpha)] = a$ in $\mathcal{N}_{\alpha_0}$. From standard matrix derivative rules:

$$G_\alpha(\sigma^+,\alpha) := \underset{m\times a}{\frac{\partial g(\sigma^+,\alpha)}{\partial \alpha'}} = \frac{\partial g(\sigma^+,\alpha)}{\partial vec(A_1')'} \times\ S_{A_1}$$

$$= \begin{pmatrix} D_k^+ \frac{\partial vec(A_1'\Sigma_u A_1 - I_k)}{\partial vec(A_1')'} \\ \frac{\partial vec(A_1'\Sigma_{u,v})}{\partial vec(A_1')'} \end{pmatrix} S_{A_1} = \begin{pmatrix} 2D_k^+(A_1'\Sigma_u \otimes I_k) \\ \Sigma_{v,u} \otimes I_k \end{pmatrix} S_{A_1}$$

which proves the result.

(ii) The restriction $a \leq m$ follows straightforwardly from the fact that the Jacobian matrix $G_\alpha(\sigma^+,\alpha)$ is $m \times a$ and the rank condition. From Proposition 1 in Angelini and Fanelli (2019), for $k > 1$, the necessary order condition for identification requires $\ell \geq \frac{1}{2}k(k-1)$, where ℓ denotes the number of additional parametric restrictions other than the instruments. We exploit the relationship $\ell + a = nk$, which establishes that the sum of the restrictions placed on the matrix A_1, ℓ, plus the number of the free (unconstrained) parameters that enter the matrix A_1, a, must equal the total number of elements of the matrix A_1, nk. Since $s \leq n - k$, then

$$a \leq \frac{1}{2}k(k+1) + ks \leq \frac{1}{2}k(k+1) + k(n-k) = nk - \frac{1}{2}k(k-1)$$

so that

$$\ell = nk - a \geq nk - \left\{nk - \frac{1}{2}k(k-1)\right\} = \frac{1}{2}k(k-1).\blacksquare$$

S.4.5 Proof of Proposition 2

(i) To prove the consistency we observe that (a) under Assumptions 1-2 and 4 and from Proposition 1, $Q_0(\alpha) := g(\sigma_0^+, \alpha)' V_{gg,0}^{-1}(\bar{\alpha}) g(\sigma_0^+, \alpha)$ is uniquely maximized at the point α_0 in $\mathcal{N}_{\alpha_0}$; (b) $\mathcal{P}_\alpha$ is compact and $\mathcal{N}_{\alpha_0} \subseteq \mathcal{T}_\alpha \subseteq \mathcal{P}_\alpha$; (c) $Q_0(\alpha)$ is continuos; (d) for any $\bar{\alpha}$, $\hat{Q}_T(\alpha) := \hat{g}_T(\hat{\sigma}_T^+, \alpha)' \hat{V}_{gg}(\bar{\alpha})^{-1} \hat{g}_T(\hat{\sigma}_T^+, \alpha)$ converges uniformly in probability to $Q_0(\alpha)$. To see that (d) holds, recall that $\hat{\sigma}_T^+ \xrightarrow{p} \sigma_0^+$ by Lemma S1, $\hat{g}_T(\hat{\sigma}_T^+, \alpha) \xrightarrow{p} g(\sigma_0^+, \alpha)$ and $\hat{V}_{gg}(\bar{\alpha}) \xrightarrow{p} V_{gg,0}$ by the Slutzky Theorem. Then, given the Euclidean norm $\|\cdot\|$, by the triangle and Cauchy-Schwartz inequalities:

$$
\begin{aligned}
\left|\hat{Q}_T(\alpha) - Q_0(\alpha)\right| \leq & \left|[\hat{g}_T(\hat{\sigma}_T^+, \alpha) - g(\sigma_0^+, \alpha)]' \hat{V}_{gg}(\bar{\alpha})^{-1} [\hat{g}_T(\hat{\sigma}_T^+, \alpha) - g(\sigma_0^+, \alpha)]\right| \\
& + \left| g(\sigma_0^+, \alpha)' [\hat{V}_{gg}(\bar{\alpha})^{-1} + \hat{V}_{gg}(\bar{\alpha})'^{-1}] [\hat{g}_T(\hat{\sigma}_T^+, \alpha) - g(\sigma_0^+, \alpha)] \right| \\
& + \left| g(\sigma_0^+, \alpha)' [\hat{V}_{gg}(\bar{\alpha})^{-1} - V_{gg,0}^{-1}] g(\sigma_0^+, \alpha)' \right|
\end{aligned}
$$

$$
\begin{aligned}
\leq \; & \left\|\hat{g}_T(\hat{\sigma}_T^+, \alpha) - g(\sigma_0^+, \alpha)\right\|^2 \left\|\hat{V}_{gg}(\bar{\alpha})^{-1}\right\| \\
& + 2 \left\|g(\sigma_0^+, \alpha)\right\| \left\|\hat{g}_T(\hat{\sigma}_T^+, \alpha) - g(\sigma_0^+, \alpha)\right\| \left\|\hat{V}_{gg}(\bar{\alpha})^{-1}\right\| \\
& + \left\|g(\tau_0, \alpha)\right\|^2 \left\|\hat{V}_{gg}(\bar{\alpha})^{-1} - V_{gg,0}^{-1}\right\|
\end{aligned}
$$

so that $\sup_{\alpha \in \mathcal{P}_\alpha} \left|\hat{Q}_T(\alpha) - Q_0(\alpha)\right| \leq \sup_{\alpha \in \mathcal{T}_\alpha} \left|\hat{Q}_T(\alpha) - Q_0(\alpha)\right| \xrightarrow{p} 0$. Given (a), (b), (c), and (d), the consistency result follows from Theorem 2.1 in Newey and McFadden (1994).

(ii) To prove asymptotic normality we start from the first-order conditions implied by the problem (16) in the paper:

$$
G_\alpha(\hat{\sigma}_T^+, \hat{\alpha}_T)' \hat{V}_{gg}^{-1}(\bar{\alpha}) \hat{g}_T(\hat{\sigma}_T^+, \hat{\alpha}_T) = 0_{a \times 1}. \tag{S.21}
$$

By expanding $\hat{g}_T(\hat{\sigma}_T^+, \hat{\alpha}_T)$ around α_0 and solving, yields the expression (valid in $\mathcal{N}_{\alpha_0}$):

$$
\begin{aligned}
& \left\{ G_\alpha(\hat{\sigma}_T^+, \hat{\alpha}_T)' \hat{V}_{gg}^{-1}(\bar{\alpha}) G_\alpha(\hat{\sigma}_T^+, \breve{\alpha}) \right\} T^{1/2} (\hat{\alpha}_T - \alpha_0) \\
& \qquad = -G_\alpha(\hat{\sigma}_T^+, \hat{\alpha}_T)' \hat{V}_{gg}^{-1}(\bar{\alpha}) T^{1/2} \hat{g}_T(\hat{\sigma}_T^+, \alpha_0)
\end{aligned} \tag{S.22}
$$

where $\breve{\alpha}$ is a mean value. From the consistency result in (i), as $T \longrightarrow \infty$, $G_\alpha(\hat{\sigma}_T^+, \hat{\alpha}_T) \xrightarrow{p} G_\alpha(\sigma_0^+, \alpha_0)$ and $G_\alpha(\hat{\sigma}_T^+, \breve{\alpha}) \xrightarrow{p} G_\alpha(\sigma_0^+, \alpha_0)$. Moreover, $G_\alpha(\sigma_0^+, \alpha_0)' \hat{V}_{gg}^{-1}(\bar{\alpha}) G_\alpha(\sigma_0^+, \alpha_0)$ is nonsingular in $\mathcal{N}_{\alpha_0}$ because of Proposition 1. It turns out that

$$
\left\{ G_\alpha(\hat{\sigma}_T^+, \hat{\alpha}_T)' \hat{V}_{gg}(\bar{\alpha})^{-1} G_\alpha(\hat{\sigma}_T^+, \breve{\alpha}) \right\}^{-1} G_\alpha(\hat{\sigma}_T^+, \hat{\alpha}_T)' \hat{V}_{gg}^{-1}(\bar{\alpha})
$$

$$\xrightarrow{p} \left\{G_\alpha(\sigma_0^+,\alpha_0)' V_{gg}(\bar\alpha)^{-1} G_\alpha(\sigma_0^+,\alpha_0)\right\}^{-1} G_\alpha(\sigma_0^+,\alpha_0)' \hat V_{gg}^{-1}(\bar\alpha).$$

Under Assumptions 1, 2 and 4 and Lemma S1, $T^{1/2}\hat g_T(\hat\sigma_T^+,\alpha_0) \xrightarrow{d} N(0_{m\times 1}, V_{gg}(\bar\alpha))$, and the conclusion follows from solving (S.22) for $T^{1/2}(\hat\alpha_T - \alpha_0)$ and the Slutzky theorem. ■

S.4.6 Proof of Proposition 3

(i) $\hat\mu_T^*$ is a smooth function of $\hat\sigma_T^{+*} = M_{\sigma^+}\hat\delta^*_{\eta,T}$, hence from Lemma S1(ii) we have $\hat\mu_T^* - \hat\mu_T \xrightarrow{p^*}_p 0_{o\times 1}$, so that $\hat Q_T^*(\theta) := (\hat\mu_T^* - f(\theta))' \hat V_\mu^{-1}(\hat\mu_T^* - f(\theta))$ satisfies $\hat Q_T^*(\theta) - \hat Q_T(\theta) \xrightarrow{p^*}_p 0$, where $\hat Q_T(\theta) := (\hat\mu_T - f(\theta))\hat V_\mu^{-1}(\hat\mu_T - f(\theta))$ is continuous and for $\theta \in \mathcal{N}_{\theta_0}$ and strong proxies is uniquely minimized at $\hat\theta_T$ by Lemma 3(i). Moreover, $\hat\mu_T^* - f(\theta)$ is such that $\mathrm{E}^*\left[\sup_{\theta\in P_\theta} \|\hat\mu_T^* - f(\theta)\|\right] < \infty$, then, the result $\hat\theta_T^* - \hat\theta_T \xrightarrow{p^*}_p 0_{q_\theta\times 1}$ follows from Theorem 2.6 in Newey and McFadden (1994) and Assumption 1.

The first-order conditions associated with the minimization problem in equation (20) of the paper are given by

$$J'_{\hat\theta_T^*} \hat V_\mu^{-1}(\hat\mu_T^* - f(\hat\theta_T^*)) = 0_{o\times 1} \tag{S.23}$$

where $J'_{\hat\theta_T^*}$ is the Jacobian (S.18) evaluated at the MBB-CMB estimator $\hat\theta_T^*$. By a mean-value expansion of $f(\hat\theta_T^*)$ about $\hat\theta_T$, we obtain

$$f(\hat\theta_T^*) = f(\hat\theta_T) + J_{\dot\theta}(\hat\theta_T^* - \hat\theta_T)$$

where $\dot\theta$ is an intermediate vector value between $\hat\theta_T^*$ and $\hat\theta_T$. Using the above expansion in (S.23), yields

$$J'_{\hat\theta_T^*} \hat V_\mu^{-1}(\hat\mu_T^* - f(\hat\theta_T) - J_{\dot\theta}(\hat\theta_T^* - \hat\theta_T)) = 0_{o\times 1},$$

hence, for $f(\hat\theta_T) = \hat\mu_T$, it holds that:

$$J'_{\hat\theta_T^*} \hat V_\mu^{-1}(\hat\mu_T^* - \hat\mu_T) - J'_{\hat\theta_T^*} \hat V_\mu^{-1} J_{\dot\theta}(\hat\theta_T^* - \hat\theta_T) = 0_{o\times 1},$$

$$\{J'_{\hat\theta_T^*} \hat V_\mu^{-1} J_{\dot\theta}\} T^{1/2}(\hat\theta_T^* - \hat\theta_T) = J'_{\hat\theta_T^*} \hat V_\mu^{-1} T^{1/2}(\hat\mu_T^* - \hat\mu_T) \tag{S.24}$$

which links the asymptotic distribution of $T^{1/2}(\hat\theta_T^* - \hat\theta_T)$ (conditional on the data) to the asymptotic distribution of $T^{1/2}(\hat\mu_T^* - \hat\mu_T)$ (conditional on the data), and to the local rank properties of the Jacobian matrix J_θ. If for

$\theta \in \mathcal{N}_{\theta_0}$ the proxies are strong in the sense of Definition 1(a) of the paper then, conditionally on the original data, the asymptotic normality of $T^{1/2}(\hat{\mu}_T^* - \hat{\mu}_T)$ in (S.24) follows from the asymptotic normality of $T^{1/2}(\hat{\sigma}_T^{+*} - \hat{\sigma}_T^{+})$, which is in turn guaranteed by Lemma S1(ii). Moreover, as $\hat{\theta}_T^* - \hat{\theta}_T = o_p^*(1)$, in probability, then, in probability, $J_{\hat{\theta}_T^*} - J_{\hat{\theta}_T} = o_p^*(1)$, $J_{\dot{\theta}} - J_{\hat{\theta}_T} = o_p^*(1)$ and, accordingly, $J'_{\hat{\theta}_T^*}\hat{V}_\mu^{-1}J_{\dot{\theta}} - J'_{\hat{\theta}_T}\hat{V}_\mu^{-1}J_{\hat{\theta}_T} = o_p^*(1)$, where the $q_\theta \times q_\theta$ matrix $J'_{\hat{\theta}_T}\hat{V}_\mu^{-1}J_{\hat{\theta}_T}$ is positive definite. This proves the result.

(ii) If for $\theta \in \mathcal{N}_{\theta_0}$ the proxies are weak in the sense of Definition 1(b) of the paper (or irrelevant), the quantity $T^{1/2}(\hat{\mu}_T - \mu_0)$ is not asymptotically Gaussian because of the non-normality of $T^{1/2}(\hat{\omega}_T - \omega_0)$ established in Lemma S2(iii). We now show that also $T^{1/2}(\hat{\omega}_T^* - \hat{\omega}_T)$, the bootstrap counterpart of $T^{1/2}(\hat{\omega}_T - \omega_0)$, is not (conditional on the data) asymptotically Gaussian, which in light of (S.24) suffices to claim that $T^{1/2}(\hat{\theta}_T^* - \hat{\theta}_T)$ is not (conditional on the data) asymptotic Gaussian. To save space and as in Lemma S2, we consider the case where all proxies are weak.

Notice that $\hat{\omega}_T^* = \omega(\hat{\sigma}_T^{+*})$, the function $\omega(\cdot)$ being smooth. From Lemma S1(ii) $\hat{\sigma}_T^{+*} - \hat{\sigma}_T^{+} \xrightarrow{p^*}_p 0$, in probability, so that also $\hat{\omega}_T^* - \hat{\omega}_T = o_p^*(1)$, in probability, regardless of the strength of instruments. Consider (T times) the quadratic expansion of $\hat{\omega}_T^* = \omega(\hat{\sigma}_T^{+*})$ around $\hat{\sigma}_T^{+}$:

$$T\left(\hat{\omega}_T^* - \hat{\omega}_T\right) = T^{1/2}J_{\sigma^+}^{(1)}(\hat{\sigma}_T^{+})T^{1/2}(\hat{\sigma}_T^{+*} - \hat{\sigma}_T^{+}) + \tfrac{T}{2}R_{1,T}(\ddot{\sigma}_T^{+*}) \tag{S.25}$$

where $J_{\sigma^+}^{(1)}(\hat{\sigma}_T^{+}) := \frac{\partial\omega}{\partial\sigma^{+\prime}}\big|_{\sigma^+=\hat{\sigma}_T^{+}}$, and the remainder term $R_{1,T}(\ddot{\sigma}_T^{+*})$ has representation

$$\begin{aligned} TR_{1,T}(\ddot{\sigma}_T^{+*}) &:= \left(I_{o_1} \otimes T^{1/2}(\hat{\sigma}_T^{+*} - \hat{\sigma}_T^{+})'\right) H^{(1)}(\ddot{\sigma}_T^{+*})T^{1/2}(\hat{\sigma}_T^{+*} - \hat{\sigma}_T^{+}), \\ H^{(1)}(\ddot{\sigma}_T^{+*}) &:= \frac{\partial}{\partial\sigma^{+\prime}}\operatorname{vec}\left(\frac{\partial\omega}{\partial\sigma^{+\prime}}\right)'\bigg|_{\sigma^+=\ddot{\sigma}_T^{+*}} \end{aligned}$$

with $\ddot{\sigma}_T^{+*}$ being an intermediate vector value between $\hat{\sigma}_T^{+*}$ and $\hat{\sigma}_T^{+}$.

We now show that the distribution of $T^{1/2}\left(\hat{\omega}_T^* - \hat{\omega}_T\right)$, conditionally on the data, rather than converging in probability, converges in distribution to a random cumulative distribution function. That is, the (conditional) bootstrap measure is random in the limit; see Cavaliere and Georgiev (2020). Randomness essentially arises because of the limit behavior of the Jacobian $T^{1/2}J_{\sigma^+}^{(1)}(\hat{\sigma}_T^{+})$: specifically, while in the original world it holds that $T^{1/2}J_{\sigma^+}^{(1)}(\sigma_0^{+}) \to J^{(1)}$ (see the proof of Lemma S2(ii)), its analog in the bootstrap world, $T^{1/2}J_{\sigma^+}^{(1)}(\hat{\sigma}_T^{+})$, does not converges to a constant.

First, from Lemma S1(ii), $T^{1/2}(\hat{\sigma}_T^{+*} - \hat{\sigma}_T^{+}) \xrightarrow{d^*}_p \mathbb{G}^*_{\sigma^+} \equiv N(0, V_{\sigma^+})$. Moreover, by continuity of the second derivative and using the fact that $\hat{\sigma}_T^+ = \sigma_0^+ + o_p(1)$, it holds that $H^{(1)}(\ddot{\sigma}_T^{+*}) \xrightarrow{p^*}_p H^{(1)}(\sigma_0^+)$ and hence

$$TR_{1,T}(\ddot{\sigma}_T^{+*}) \xrightarrow{d^*}_p \left(I_M \otimes \mathbb{G}^{*\prime}_{\sigma^+}\right) H^{(1)}_{\sigma_0^+} \mathbb{G}^*_{\sigma^+}$$

where $H^{(1)}_{\sigma_0^+} := H^{(1)}(\sigma_0^+)$. Consider now $T^{1/2} J^{(1)}_{\sigma^+}(\hat{\sigma}_T^+)$. By an expansion of $\mathrm{vec}\, J^{(1)}_{\sigma^+}(\hat{\sigma}_T^+)$ around the true value $\mathrm{vec}\, J^{(1)}_{\sigma^+}(\sigma_0^+)$ we obtain

$$T^{1/2}\,\mathrm{vec}\, J^{(1)}_{\sigma^+}(\hat{\sigma}_T^+) = T^{1/2}\,\mathrm{vec}\, J^{(1)}_{\sigma^+}(\sigma_0^+) + H^{(1)}_{\ddot{\sigma}^+} T^{1/2}(\hat{\sigma}_T^+ - \sigma_0^+)$$

where the matrix $H^{(1)}_{\ddot{\sigma}^+}$ is given in Lemma S2(ii). From $\hat{\sigma}_T^+ - \sigma_0^+ = o_p(1)$ and continuity of the Hessian it follows that $H^{(1)}_{\ddot{\sigma}^+} \to H^{(1)}_{\sigma_0^+}$. This result, together with $T^{1/2}(\hat{\sigma}_T^{+*} - \hat{\sigma}_T^+) \xrightarrow{d} N(0, V_{\sigma^+})$ (Lemma S1(i)) and $T^{1/2}\,\mathrm{vec}\, J^{(1)}_{\sigma^+}(\sigma_0^+) \to \mathrm{vec}\, J^{(1)}$ (proof of Lemma S2), implies that

$$T^{1/2}\,\mathrm{vec}\, J^{(1)}_{\sigma^+}(\hat{\sigma}_T^+) \xrightarrow{d} N\left(\mathrm{vec}\, J^{(1)},\ H^{(1)}_{\sigma^+} V_{\sigma^+} H^{(1)}_{\sigma^+}\right) =: \mathrm{vec}(\mathbb{G}_{J^{(1)}})$$

with $\mathbb{G}_{J^{(1)}}$ a Gaussian matrix, implicitly defined. Notice that the covariance matrix $H^{(1)}_{\sigma^+} V_{\sigma^+} H^{(1)}_{\sigma^+}$, albeit being of reduced rank, is not zero. In summary,

$$T(\hat{\omega}_T^* - \hat{\omega}_T) = \underbrace{T^{1/2} J^{(1)}_{\sigma^+}(\hat{\sigma}_T^+)}_{\xrightarrow{d} \mathbb{G}_{J^{(1)}}} \underbrace{T^{1/2}(\hat{\sigma}_T^{+*} - \hat{\sigma}_T^+)}_{\xrightarrow{d^*}_p \mathbb{G}^*_{\sigma^+}} + \quad \frac{1}{2} \underbrace{R_{1,T}(\ddot{\sigma}_T^{+*})}_{\xrightarrow{d^*}_p \left(I_{o_1} \otimes \mathbb{G}^{*\prime}_{\sigma^+}\right) H^{(1)}_{\sigma_0^+} \mathbb{G}^*_{\sigma^+}} \tag{S.26}$$

Because the term $T^{1/2} J^{(1)}_{\sigma^+}(\hat{\sigma}_T^+)$ does not converge in probability to a constant but rather (in distribution) to a random variable, the limit distribution of $T(\hat{\omega}_T^* - \hat{\omega}_T)$ is random in the limit. Specifically, the limit can be described as a mixture of a Gaussian random variable $\mathbb{G}^*_{\sigma^+}$ and the χ^2-type random $\left(I_{o_1} \otimes \mathbb{G}^{*\prime}_{\sigma^+}\right) H^{(1)}_{\sigma_0^+} \mathbb{G}^*_{\sigma^+}$, where the weight $\mathbb{G}^*_{\sigma^+}$ is a random matrix fixed across bootstrap repetitions and, precisely, distributed as $\mathbb{G}_{J^{(1)}}$. Put differently,

$$T(\hat{\omega}_T^* - \hat{\omega}_T) \xrightarrow{d^*}_w \mathbb{G}_{J^{(1)}} \mathbb{G}^*_{\sigma^+} + \frac{1}{2}(I_{o_1} \otimes \mathbb{G}^{*\prime}_{\sigma^+}) H^{(1)}_{\sigma_0^+} \mathbb{G}^*_{\sigma^+} \Big| \mathbb{G}_{J^{(1)}} \tag{S.27}$$

where '$Y_T^* \xrightarrow{d^*}_w Y|X$' denotes weak convergence of the cdf of Y_T, given the original data, to the (diffuse) conditional distribution of Y given X, i.e.

$$P^*(Y_T^* \le x) \to_w P(Y \le x | X)$$

see Cavaliere and Georgiev (2020). The formal proof of (S.27) can be obtained from the convergence facts reported in (S.26) following e.g. the proof of Theorem 4.2 in Cavaliere and Georgiev (2020) or Basawa *et al.* (1991). Specifically, consider first the bootstrap statistic

$$\mathbb{A}_T^* := \mathbb{A}_T T^{1/2}(\hat{\sigma}_T^{+*} - \hat{\sigma}_T^{+}) + \tfrac{1}{2} T R_T(\ddot{\sigma}_T^{+*})$$

where $\mathbb{A}_T$ is a deterministic matrix sequence satisfying $\mathbb{A}_T \to \mathbb{A}$. Using the results above it holds that, conditionally on the original data, and due to continuity of the cdf of $\frac{1}{2}(I_{o_1} \otimes \mathbb{G}_{\sigma^+}^{*\prime})H_{\sigma_0^+}^{(1)}\mathbb{G}_{\sigma^+}^*$,

$$\sup_{x\in\mathbb{R}^r}\left|P^*\left(\mathbb{A}_T^* \leq x\right) - P(\mathbb{A}\mathbb{G}_{\sigma^+}^* + \tfrac{1}{2}(I_{o_1} \otimes \mathbb{G}_{\sigma^+}^{*\prime})H_{\sigma_0^+}^{(1)}\mathbb{G}_{\sigma^+}^* \leq x)\right| \to_p 0 \tag{S.28}$$

where the inequality in the previous equation is taken component-wise.

Second, as in Lemma A.2(a) in Cavaliere and Georgiev (2020), see also Corollary 5.12 of Kallenberg (1997), consider a special probability space where $\mathbb{G}_{J^{(1)}}$ is defined and, for every sample size T, also the original and the bootstrap data can be redefined, maintaining their distribution (we also maintain the notation), such that (jointly) $T^{1/2}J_{\sigma^+}^{(1)}(\hat{\sigma}_T^{+}) \to_{a.s.} \mathbb{G}_{J^{(1)}}$ and $T^{1/2}(\hat{\sigma}_T^{+*} - \hat{\sigma}_T^{+}) \xrightarrow{d^*}_{a.s.} \mathbb{G}_{\sigma^+}^*$, rather than in distribution. Then, in this special probability space, from (S.28) and $T^{1/2}J_{\sigma^+}^{(1)}(\hat{\sigma}_T^{+}) \to_{a.s.} \mathbb{G}_{J^{(1)}}$, it follows that

$$T\left(\hat{\omega}_T^* - \hat{\omega}_T\right) \xrightarrow{d^*}_{a.s.} \mathbb{G}_{J^{(1)}}\mathbb{G}_{\sigma^+}^* + \tfrac{1}{2}\left(I_r \otimes \mathbb{G}_{\sigma^+}^{*\prime}\right) H_{\sigma_0^+}^{(1)}\mathbb{G}_{\sigma^+}^*\Big|\,\mathbb{G}_{J^{(1)}}$$

and, in the original probability space, (S.27) holds. ■

S.4.7 Proof of Proposition 4

Given the distance defined in equation (23) of the paper, we consider the decomposition

$$\begin{aligned}\tau_{T,N}^*(x) &= N^{1/2}\hat{U}_T(x)^{-1/2}(\digamma_{T,N}^*(x) - \digamma_T^*(x)) \\ &\quad + N^{1/2}\hat{U}_T(x)^{-1/2}(\digamma_T^*(x) - \mathcal{F}_{\mathcal{N}}(x)).\end{aligned} \tag{S.29}$$

For T fixed, the first term on the left-hand side of (S.29) converges, as $N \to \infty$, to a $N(0,1)$ regardless of the strength of proxies by the CLT in equation (22) of the paper (the strength of proxies affects the asymptotic behavior of the second term in (S.29)).

(i) Under strong proxies, if the term $\digamma_T^*(x) - \mathcal{F}_{\mathcal{N}}(x)$ admits a standard Edgeworth expansion such that $\digamma_T^*(x) - \digamma_{\mathcal{N}}(x) = O_p\left(T^{-1/2}\right)$, the second term on the right-hand side in (S.29) is of order $O_p\left(N^{1/2}T^{-1/2}\right)$ and by Proposition

3(i) the statistic $\tau^*_{T,N}(x)$ is asymptotically $N(0,1)$ provided $T, N \to \infty$ jointly and $NT^{-1} = o(1)$ as in equation (24) of the paper.

(ii) Under weak (or irrelevant) proxies, by Proposition 3(ii) $F^*_T(x)$ does not converge (in probability) to $F_{\mathcal{N}}(x)$, which means that the second term on the right hand side of (S.29) does not vanishes asymptotically, implying that $\tau^*_{T,N}(x)$ diverges at the rate of $N^{1/2}$ as $N, T \to \infty$. ■

S.4.8 Proof of Proposition 5

Let $\mathcal{D}_T$ denote the original data upon which the proxy-SVAR is estimated, defined on the probability space $(Q, \mathcal{F}, P)$. As is standard, the bootstrap (conditional) cdf $F^*_T(x) := P(\hat{\theta}^*_T \leq x|\mathcal{D}_T)$ is a function of the data only. Using $F^*_T(\cdot)$, we generate a set of N i.i.d. 'bootstrap' random variables as follows. First, let U^*_b, $b = 1, ..., N$, be a sequence of i.i.d. $U[0,1]$ random variables independent on the data (we implicitly extend the original probability space such that it includes the U^*_b's as well). Then, the bootstrap random variables $\hat{\theta}^*_{T:b}$, $b = 1, ..., N$ that enter the argument of the statistic $\tau^*_{T,N} := \tau(\hat{\theta}^*_{T:1}, ..., \hat{\theta}^*_{T:N})$ are defined as $\hat{\theta}^*_{T:b} := F^{*-1}_T(U^*_b)$, $b = 1, ..., N$, where $F^{*-1}_T(\cdot)$ is the generalized inverse of $F^*_T(\cdot)$. Thus, we have

$$\tau^*_{T,N} = \tau(\hat{\theta}^*_{T:1}, ..., \hat{\theta}^*_{T:N}) = \tau(F^{*-1}_T(U^*_1), ..., F^{*-1}_T(U^*_N))$$

with cdf, conditional on $\mathcal{D}_T$, given by $\mathcal{H}_{T,N}(x) = P(\tau^*_{T,N} \leq x|\mathcal{D}_T)$.

We now prove that ρ_T and $\tau^*_{T,N}$ are independent asymptotically, in the sense that for any $x, c \in \mathbb{R}$, as $T, N \to \infty$, the condition in equation (25) of the paper, here reported for convenience, holds:

$$P(\{\rho_T \leq x\} \cap \{\tau^*_{T,N} \leq c\}) - P(\rho_T \leq x)P(\tau^*_{T,N} \leq c) \to 0. \qquad \text{(S.30)}$$

Observe that (S.30) trivially holds in the presence of weak proxies because by Proposition 3(ii), $\tau^*_{T,N}$ diverges for $N, T \to \infty$. In the presence of strong proxies, Proposition 3(i) ensures that as $T, N \to \infty$, $\mathcal{H}_{T,N}(x) \to_p \mathcal{H}(x)$, where $\mathcal{H}(x)$ is a non-random cdf. By the law of iterated expectations (and the fact that $P(X \in \mathcal{E}) = E(\mathbb{I}_{\{X \in \mathcal{E}\}})$), we have

$$\begin{aligned}
P(\{\rho_T \leq x\} \cap \{\tau^*_{T,N} \leq c\}) &= E(\mathbb{I}_{\{\rho_T \leq x\} \cap \{\tau^*_{T,N} \leq c\}}) = E(\mathbb{I}_{\{\rho_T \leq x\}}\mathbb{I}_{\{\tau^*_{T,N} \leq c\}}) \\
&= E\left(E(\mathbb{I}_{\{\rho_T \leq x\}}\mathbb{I}_{\{\tau^*_{T,N} \leq c\}}|\mathcal{D}_T)\right) \\
&= E\left(\mathbb{I}_{\{\rho_T \leq x\}}E(\mathbb{I}_{\{\tau^*_{T,N} \leq c\}}|\mathcal{D}_T)\right) \\
&= E\left(\mathbb{I}_{\{\rho_T \leq x\}}\mathcal{H}_{T,N}(c)\right) \\
&= E\left(\mathbb{I}_{\{\rho_T \leq x\}}\mathcal{H}(c)\right) + E\left(\mathbb{I}_{\{\rho_T \leq x\}}(\mathcal{H}_{T,N}(c) - \mathcal{H}(c))\right)
\end{aligned}$$

$$= P(\rho_T \leq x)\mathcal{H}(c) + E\left(\mathbb{I}_{\{\rho_T \leq x\}}(\mathcal{H}_{T,N}(c) - \mathcal{H}(c))\right).$$

For the last term, we have

$$\begin{aligned}\left|E\left(\mathbb{I}_{\{\rho_T \leq x\}}(\mathcal{H}_{T,N}(c) - \mathcal{H}(c))\right)\right| &\leq E\left|\mathbb{I}_{\{\rho_T \leq x\}}(\mathcal{H}_{T,N}(c) - \mathcal{H}(c))\right| \\ &\leq E\left|(\mathcal{H}_{T,N}(c) - \mathcal{H}(c))\right|.\end{aligned}$$

Since we know that under strong proxies $\mathcal{H}_{T,N}(c) \rightarrow_p \mathcal{H}(c)$, then $E|\mathcal{H}_{T,N}(c) - \mathcal{H}(c)| \rightarrow 0$ provided $|\mathcal{H}_{T,N}(c) - \mathcal{H}(c)|$ is uniformly integrable. But $\mathcal{H}_{T,N}(c)$ and $\mathcal{H}(c)$ are cdfs, and hence they are both bounded and uniformly integrable. Hence, as $T, N \rightarrow \infty$,

$$P(\{\rho_T \leq x\} \cap \{\tau^*_{T,N} \leq c\}) - P(\rho_T \leq x)\mathcal{H}(c) = o_p(1).$$

Therefore,

$$\begin{aligned}P(\{\rho_T &\leq x\} \cap \{\tau^*_{T,N} \leq c\}) - P(\rho_T \leq x)P(\tau^*_{T,N} \leq c) \\ &= P(\{\rho_T \leq x\} \cap \{\tau^*_{T,N} \leq c\}) - P(\rho_T \leq x)\mathcal{H}(c) + P(\rho_T \leq x)\left(\mathcal{H}(c) - P(\tau^*_{T,N} \leq c)\right) \\ &= P(\rho_T \leq x)\left(\mathcal{H}(c) - P(\tau^*_{T,N} \leq c)\right) + o_p(1).\end{aligned}$$

Since $P(\rho_T \leq x) \in [0,1]$, we only need to prove that $P(\tau^*_{T,N} \leq c) - \mathcal{H}(c)$ vanishes asymptotically. But this immediately follows from bootstrap consistency as

$$\begin{aligned}P(\tau^*_{T,N} &\leq c) - \mathcal{H}(c) = E(\mathbb{I}_{\{\tau^*_{T,N} \leq c\}}) - \mathcal{H}(c) \\ &= E(E\mathbb{I}_{\{\tau^*_{T,N} \leq c\}}|\mathcal{D}_T)) - \mathcal{H}(c) \\ &= E\left(\mathcal{H}_{T,N}(c) - \mathcal{H}(c)\right) \rightarrow 0\end{aligned}$$

by the uniform integrability of $\mathcal{H}_{T,N}(c)$.■

S.5 Indirect-MD approach: identification restrictions on B_1

Section 4 of the paper discusses the case in which in the multiple target shocks case, $k > 1$, the additional restrictions necessary for the identification of the proxy-SVAR are placed on the matrix A_1' characterizing the A-form of the proxy-SVAR. Actually, the specification of the proxy-SVAR might be based on the B-form in (2)-(5) of the paper and the additional restrictions necessary to (point-)identify the proxy-SVAR (see Proposition 1 in the paper) might be provided by the theory on the matrix B_1. In this section we discuss how the indirect-MD estimation problem can be addressed with in these cases.

The additional restrictions on B_1 are represented in the form:

$$\mathrm{vec}(B_1) = S_{B_1}\beta_1 + s_{B_1} \tag{S.31}$$

where β_1 is the vector of (free) structural parameters that enter the matrix B_1 and S_{B_1} and s_{B_1} are the analogs of $S_{A_1}$and s_{A_1} in equation (14) of the paper. Using (9) in the paper, the moment conditions in (12) and (13) can be mapped to the expressions:

$$B_1'\Sigma_u^{-1}B_1 = I_k, \tag{S.32}$$

$$B_1'\Omega_{u,v} = 0_{k\times s} \tag{S.33}$$

where $\Omega_{u,v} := \Sigma_u^{-1}\Sigma_{u,v}$ is function of the reduced form parameters in $\sigma^+ := (\mathrm{vech}(\Sigma_u)', \mathrm{vec}(\Sigma_{u,v})')'$. Thus, we can summarize (S.31) and the moment conditions (S.32)-(S.33) by the distance function:

$$g^o(\omega^+, \beta_1) := \begin{pmatrix} \mathrm{vech}(B_1'(\beta_1)\Sigma_u^{-1}B_1(\beta_1) - I_k) \\ \mathrm{vec}(B_1'(\beta_1)\Omega_{u,v}) \end{pmatrix} \tag{S.34}$$

where $\omega^+ := (\mathrm{vech}(\Sigma_u)', \mathrm{vec}(\Omega_{u,v})')'$. We use the notation $B_1 := B_1(\beta_1)$ to indicate that the elements of B_1 depend on the structural parameters in β_1. Obviously, $g^o(\omega^+, \beta_1) = 0_{m\times 1}$ at the true parameter values. The estimation of β_1 follows from the MD problem:

$$\hat{\beta}_{1,T} := \arg\min_{\beta_1 \in \mathcal{T}_{\beta_1}} \hat{Q}_T^o(\beta_1)\ ,\ \ \hat{Q}_T^o(\beta_1) := \hat{g}_T^o(\hat{\omega}_T^+, \beta_1)'\hat{V}_{gg}(\bar{\beta}_1)^{-1}\hat{g}_T^o(\hat{\omega}_T^+, \beta_1) \tag{S.35}$$

where $\hat{g}_T^o(\hat{\omega}_T^+, \beta_1)$ is the distance function $g^o(\omega^+, \beta_1)$ with ω^+ replaced with its (consistent) estimator $\hat{\omega}_T^+ := (\mathrm{vech}(\hat{\Sigma}_u)', \mathrm{vec}(\hat{\Omega}_{u,v})')'$, $\mathcal{T}_{\beta_1} \subseteq \mathcal{P}_{\beta_1}$ is the user-chosen optimization set, $\mathcal{P}_{\beta_1}$ is the parameter space, $\hat{V}_{gg}(\bar{\beta}_1)$ is given by:

$$\hat{V}_{gg}(\bar{\beta}_1) := G_{\omega^+}(\hat{\omega}_T^+, \bar{\beta}_1)\hat{V}_{\omega^+}G_{\omega^+}(\hat{\omega}_T^+, \bar{\beta}_1)',$$

where $G_{\omega^+}(\omega^+, \beta_1)$ is the $m \times m$ Jacobian matrix defined by $G_{\omega^+}(\omega^+, \beta_1) := \frac{\partial g^o(\omega^+, \beta_1)}{\partial \omega^+}$, and $\bar{\beta}_1$ may be some preliminary (inefficient) estimate of β_1.

Under Assumptions 1-4, the asymptotic properties of $\hat{\beta}_{1,T}$ follow from Section 4 in the paper. The IRFs of interest are obtained from (3) in the paper.

S.6 Comparison with IV

In this section we compare the MD estimation approach presented in Section 4 of the paper with a natural alternative given by the IV estimation method.

Assume that $k > 1$ (multiple target shocks) and, for simplicity, that the matrix A_{11} in equation (11) of the paper is nonsingular. This condition is not implied by Assumption 3 of the paper, hence it is not necessary for the MD approach to work. With A_{11} nonsingular, one has $A_1' = A_{11}'(I_k \vdots - \Psi)$, $\Psi := -(A_{11}')^{-1}A_{12}'$, so that system (8) can be written as the multivariate regression model:

$$u_{1,t} = \Psi u_{2,t} + \left(A_{11}'\right)^{-1} \varepsilon_{1,t} \quad , \quad t = 1, ..., T \tag{S.36}$$

which can be interpreted, in some applications, as a system of policy reaction functions; see e.g. Caldara and Kamps (2017) for the fiscal framework and Section S.10 below. For large T, once the VAR innovations $u_{1,t}$ and $u_{2,t}$ have been replaced with the corresponding residuals $\hat{u}_{1,t}$ and $\hat{u}_{2,t}$, $t = 1, ..., T$, system (S.36) can be written in the form

$$\hat{u}_{1,t} = \Psi \hat{u}_{2,t} + \xi_t \quad , \quad t = 1, ..., T \tag{S.37}$$

where $\xi_t := (A_{11}')^{-1} \varepsilon_{1,t} + o_p(1)$ reads as a disturbance term and has covariance matrix $\Theta = (A_{11}')^{-1}(A_{11})^{-1}$.

Consider now the special case in which there exists proxies v_t for all auxiliary shocks in $\varepsilon_{2,t}$, i.e. $s = n - k$.[2] In this setup, one can estimate the parameters in the matrix $\Psi := -(A_{11}')^{-1}A_{12}'$ by IV, using the proxies v_t as instruments for the residuals $\hat{u}_{2,t}$. This produces the IV estimate $\hat{\Psi}_{IV}$ and the IV residuals $\hat{\xi}_t := \hat{u}_{1,t} - \hat{\Psi}_{IV}\hat{u}_{2,t}$, $t = 1, ..., T$, respectively. In turn, the IV residuals can be used to estimate the covariance matrix Θ by $\hat{\Theta}_{IV} = \frac{1}{T}\sum_{t=1}^{T} \hat{\xi}_t\hat{\xi}_t'$. Thus, given the IV estimators $\hat{\Psi}_{IV}$ and $\hat{\Theta}_{IV}$, the structural parameters in A_{11}' and in A_{12}' can be separately recovered from the data only if A_{11}' is upper (lower) triangle. Indeed, under this condition the Choleski factor of $\hat{\Theta}_{IV}$ is equal to $(\hat{A}_{11}')^{-1}$, which amounts to imposing $\ell = \frac{1}{2}k(k-1)$ additional identification restrictions necessary to point-identify the proxy-SVAR; see Mertens and Ravn (2013).

The MD approach developed in Section 4 is more flexible because the matrix A_{11} needs not be neither invertible nor triangular. (Point-)identification is achieved when the rank condition in Proposition 1 of the paper is valid.

S.7 MBB algorithm

In this section we summarize Brüggemann, Jentsch and Trenkler (2016)'s MBB algorithm frequently cited in the paper. The reference model is the proxy-SVAR represented as in Section. As in Section S.3, the reference model is

[2] The IV estimation of system (S.36) becomes more problematic when $s < n - k$. With $s < n - k$, it is necessary to impose at least $n - k - s$ restrictions on the parameters Ψ in system (S.37).

represented in (S.3) and the reduced form parameters of (S.3) are collected in the vector $\delta := (\delta_\psi, \delta_\eta)'$.

Given the VAR system (S.3), we consider the algorithm that follows.

Algorithm (residual-based MBB)

1. Fit the reduced form VAR model in (S.3) to the data W_1, ..., W_T and, given the estimates $\hat{\Psi}_1$,..., $\hat{\Psi}_l$, compute the innovation residuals $\hat{\eta}_t = W_t - \hat{\Psi}_1 W_{t-1} - ... - \hat{\Psi}_l W_{t-l}$ and the covariance matrix $\hat{\Sigma}_\eta := \frac{1}{T}\sum_{t=1}^{T} \hat{\eta}_t \hat{\eta}_t'$;

2. Choose a block of length $\ell < T$ and let $\mathcal{B} := [T/\ell]$ be the number of blocks such that $\mathcal{B}\ell \geq T$. Define the $\mathcal{M} \times \ell$ blocks $\mathcal{M}_{i,\ell} := (\hat{\eta}_{i+1}, ..., \hat{\eta}_{i+\ell})$, $i = 0, 1, 2, ..., T - \ell$.

3. Let i_0, i_1, ...,$i_{\mathcal{B}-1}$ be an i.i.d. random sample of the elements of the set $\{0, 1, 2, ..., T - \ell\}$. Lay blocks $\mathcal{M}_{i_0,\ell}, \mathcal{M}_{i_1,\ell}, ..., \mathcal{M}_{i_{\mathcal{B}-1},\ell}$ end-to-end and discard the last $\mathcal{B}\ell - T$ values, obtaining the residuals $\hat{\eta}_1^*, ..., \hat{\eta}_T^*$;

4. Center the residuals $\hat{\eta}_1^*, ..., \hat{\eta}_T^*$ according to the rule

$$
\begin{aligned}
e_{j\ell+e}^* &:= \hat{\eta}_{j\ell+e}^* - E^*(\hat{\eta}_{j\ell+e}^*) \\
&= \hat{\eta}_{j\ell+e}^* - \frac{1}{T-\ell+1}\sum_{g=0}^{T-\ell} \hat{\eta}_{e+g}^*
\end{aligned}
$$

for $e = 1, 2, ...\ell$ and $j = 0, 1, 2, ..., \mathcal{B} - 1$, such that $\mathrm{E}^*(e_t^*) = 0$ for $t = 1, ..., T$;

5. Generate the bootstrap sample $W_1^*, W_2^*, ..., W_T^*$ recursively by solving, for $t = 1, ..., T$, the system

$$
W_t^* = \hat{\Psi}_1 W_{t-1}^* + ... + \hat{\Psi}_l W_{t-l}^* + e_t^* \qquad \text{(S.38)}
$$

with initial condition W_0^*, W_{-1}^*, ..., W_{1-p}^* set to the pre-fixed sample values W_0, W_{-1}, ..., W_{1-p};

6. Use the sample $W_1^*, W_2^*, ..., W_T^*$ generated in the previous step to compute the bootstrap estimators of the reduced form parameters $\hat{\delta}_T^* := (\hat{\delta}_{\psi,T}^{*\prime}, \hat{\delta}_{\eta,T}^{*\prime})'$.

Once $\hat{\delta}_T^*$ is obtained from the algorithm above, the bootstrap estimators of the quantities $\hat{\mu}_T^* := (vech(\hat{\Omega}_v^*)', vec(\hat{\Sigma}_{v,u}^*)')'$ considered in Section 6 of the paper follow from $\hat{\delta}_{\eta,T}^{*\prime}$.

S.8 Data generating process

In this section we summarize the data generating process used for the Monte Carlo experiments summarized in Table 1 and Figure 1 of the paper.

The data are generated from the following three-equational SVAR with one lag and no deterministic component:

$$Y_t = \Pi_1 Y_{t-1} + u_t, \quad t = 1, ..., T$$

$$\Pi_1 := \begin{pmatrix} 0.67 & -0.12 & 0.42 \\ 0.03 & 0.43 & 0.08 \\ 0.14 & 0.02 & 0.58 \end{pmatrix}, \quad \lambda_{\max}(\Pi_1) = 0.86$$

$$u_t = \begin{pmatrix} 0.196 & 0 & 0.19 \\ 0.210 & 0.16 & -0.32 \\ 0.017 & 0 & 0.09 \end{pmatrix} \begin{pmatrix} \varepsilon_t^A \equiv \varepsilon_{1,t} \\ \varepsilon_t^B \equiv \varepsilon_{2,t}^1 \\ \varepsilon_t^C \equiv \tilde{\varepsilon}_{2,t} \end{pmatrix}$$

$$\varepsilon_t := \begin{pmatrix} \varepsilon_t^A \equiv \varepsilon_{1,t} \\ \varepsilon_t^B \equiv \varepsilon_{2,t}^1 \\ \varepsilon_t^C \equiv \tilde{\varepsilon}_{2,t} \end{pmatrix} \begin{matrix} \text{target shock} \\ \text{non-instrumented auxiliary shock} \\ \text{instrumented auxiliary shock} \end{matrix} \quad \sim iidN(0, I_3)$$

$$B_1 := \begin{pmatrix} 0.196 \\ 0.210 \\ 0.017 \end{pmatrix} \quad , \quad A_1' := \begin{pmatrix} 6.246 & 0 & -13.185 \end{pmatrix}.$$

Figure 1 of the paper considers the setup in which z_t is a weak proxy (in the sense of Definition 1(b) in Section 3 of the paper) for the target shock, while v_t is a strong proxy (in the sense of Definition 1(a) in Section 3 of the paper) for the auxiliary shock $\varepsilon_{3,t} \equiv \tilde{\varepsilon}_{2,t}$. More precisely we have:

$$z_t = \frac{\varphi}{T^{1/2}} \varepsilon_{1,t} + \sigma_z \omega_{z,t} \ \ , \ \ \omega_{z,t} \perp \varepsilon_t \ \ , \ \ \varphi := 0.5, \ \sigma_z := 0.7$$

$$v_t = \lambda \tilde{\varepsilon}_{2,t} + \sigma_v \omega_{v,t} \ \ , \ \ \omega_{v,t} \perp \varepsilon_t \ \ , \lambda := 0.8, \ \sigma_z := 1.1.$$

where $\omega_{z,t}$ and $\omega_{v,t}$ are i.i.d. measurement errors independent on ε_t. It turns out that the correlation between z_t and $\varepsilon_{1,t}$ is equal to $\frac{0.58}{T^{1/2}}$, while the correlation between v_t and $\tilde{\varepsilon}_{2,t}$ is 59%. In terms of the notation used in the paper, $n = 3$, $k = 1$, $s = 1 < n - k = 2$, $a = 2$ (recall that one element of $A_1'$is set to zero) and $m = \frac{1}{2}k(k+1) + ks = 2$.

Table 1 of the paper investigates the strength of the proxy v_t for $\tilde{\varepsilon}_{2,t}$ by the bootstrap test considering three possible scenarios. The upper panel refers to the situation in which the correlation between v_t and $\tilde{\varepsilon}_{2,t}$ is 59% Moreover, the results in Table 1 of the paper are obtained by considering two different hypotheses on the generation of the structural shocks ε_t. In one case, ε_t is generated as an $iidN(0_{3\times 1}, I_3)$ process. In the other case, ε_t is generated by postulating independent GARCH components for each of its components. More precisely, as in Jentsch and Lunsford (2019b), in the second scenario we have the following specification:

$$\varepsilon_{i,t} = \varsigma_{i,t}\varepsilon^0_{i,t} \;\;,\;\; \varepsilon^0_{i,t} \sim iidN(0,1), \quad i = 1,2,3$$

$$\varsigma^2_{i,t} = \varrho_0 + \varrho_1\varepsilon^2_{i,t-1} + \varrho_2\varsigma^2_{i,t-1},\; t = 1, ..., T$$

with $\varrho_1 := 0.05$, $\varrho_2 := 0.93$ and $\varrho_0 := (1 - \varrho_1 - \varrho_2)$.

S.9 Failure of the exogeneity (orthogonality) condition

The purpose of the present section is to show that the bootstrap test for instrument relevance discussed in Section 5 of the paper solely captures the strength of the proxies, and not possible violations of the exogeneity condition. More precisely, we show that non-normality of if the external variables used to identify the structural shocks of interest are strong but fail to be orthogonal to the non-instrumented structural shocks, the estimators of parameters of interest in the proxy-SVAR are not consistent (as expected), but are still asymptotically Gaussian. This result implies that the bootstrap-based test for instrument relevance developed in the paper is informative on the strength of the proxies but not on possible failures of the exogeneity condition.

We consider, without loss of generality, a simple proxy-SVAR based on one shock-one instrument. This allows to simply the estimation method. The setup is that of equations (1)-(4) in the paper. In the following, we denote with $\varepsilon_{I,t}$ a scalar ($k = 1$) instrumented structural shock and with $\varepsilon_{NI,t}$ the remaining $(n-1)\times 1$ 'non-instrumented' structural shocks of the system; $\varepsilon_t := (\varepsilon_{I,t}, \varepsilon'_{NI,t})'$. Imagine that z_t is a strong proxy (in the sense of Definition 1(a) in Section 3 of the paper) for $\varepsilon_{I,t}$ which nevertheless fails to be orthogonal to all non-instrumented structural shocks. Thus, assume the data generating process for z_t is given by

$$z_t = \phi_1\varepsilon_{I,t} + \phi_2\varepsilon^o_{NI,t} + \omega_{z,t} \tag{S.39}$$

where ϕ_1 is the relevance parameter, $\varepsilon^o_{NI,t}$ is one structural shock in the vector $\varepsilon_{NI,t}$ and ϕ_2 is a parameter responsible for the violation of the exogeneity condition whenever it differs from zero; $\omega_{z,t}$ is an i.i.d. measurement error assumed orthogonal to $\varepsilon_t := (\varepsilon_{I,t}, \varepsilon'_{NI,t})'$.

In the following we distinguish between the two cases of exogenous and non-exogenous proxies.

Exogenous proxies. In the standard proxy-SVAR approach, it is maintained that $\phi_2 = 0$ in (S.39). Then, by combining the proxy with the VAR innovations u_t in (2) one obtains the relationships

$$\mathrm{E}(u_t z_t') = \Sigma_{u,z} \equiv \begin{pmatrix} \Sigma_{u1,z} \\ \Sigma_{u2,z} \end{pmatrix} = \phi_1 B_1 \equiv \begin{pmatrix} B_{11}\phi_1 \\ B_{21}\phi_1 \end{pmatrix} \begin{matrix} 1\times 1 \\ (n-1)\times 1 \end{matrix} .$$

Under e.g. the 'unit effect' normalization $B_{11} = 1$ (Montiel Olea, Stock and Watson, 2020), the moment conditions above simplify to the restrictions $\Sigma_{u1,z} = \phi_1$ and $\Sigma_{u2,z} = B_{21}\Sigma_{u1,z}$, respectively. Thus, $B_{21} = \Sigma_{u2,z}/\Sigma_{u1,z} = \gamma_2/\gamma_1$, where $\gamma_2 = \Sigma_{u2,z} \equiv vec(\Sigma_{u2,z})$, $\gamma_1 = \Sigma_{u1,z} \equiv vec(\Sigma_{u1,z})$. Regardless of the strength of the instrument, the covariance matrix $\Sigma_{u,z}$ can be estimated consistently from the data under fairly general conditions on the VAR and proxies (Assumptions 1-2 of the paper). Thus, regardless of the strength of the proxy:

$$\hat{\gamma}_T \equiv \begin{pmatrix} \hat{\gamma}_{1,T} \\ \hat{\gamma}_{2,T} \end{pmatrix} \xrightarrow{p} \begin{pmatrix} \gamma_{0,1} \\ \gamma_{0,2} \end{pmatrix} = \begin{pmatrix} \phi_1 \\ B_{21}\phi_1 \end{pmatrix}$$

and

$$T^{1/2}(\hat{\gamma}_T - \gamma_0) \equiv T^{1/2} \begin{pmatrix} \hat{\gamma}_{1,T} - \gamma_{0,1} \\ \hat{\gamma}_{2,T} - \gamma_{0,2} \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \xi_{\gamma_1} \\ \xi_{\gamma_2} \end{pmatrix} \equiv N(0, V_\gamma). \tag{S.40}$$

Therefore, under the unit effect normalization and the hypothesis of strong proxy, $\hat{\gamma}_{1,T} \xrightarrow{p} \phi_1 \neq 0$ and

$$T^{1/2}(\hat{B}_{21,T} - B_{21}) = T^{1/2}\left(\frac{\hat{\gamma}_{2,T}}{\hat{\gamma}_{1,T}} - \frac{\gamma_{0,2}}{\gamma_{0,1}}\right) = T^{1/2}\left(\frac{\hat{\gamma}_{2,T} - \gamma_{0,2} + \gamma_{0,2}}{\hat{\gamma}_{1,T}} - \frac{\gamma_{0,2}}{\gamma_{0,1}}\right)$$

$$= \frac{1}{\hat{\gamma}_{1,T}} T^{1/2}\left(\hat{\gamma}_{2,T} - \gamma_{0,2}\right) + T^{1/2}\frac{\gamma_{0,2}}{\hat{\gamma}_{1,T}} - T^{1/2}\frac{\gamma_{0,2}}{\gamma_{0,1}} \xrightarrow{d} \frac{1}{\phi_1}\xi_{\gamma_2} + o_p(1)$$

where ξ_{γ_2} is the multivariate normal implicitly defined in (S.40). Using (9) in the paper and standard arguments (recall that $\hat{\Sigma}_u \xrightarrow{p} BB'$ regardless of the strength of the proxies), it follows that also the estimator of the parameters in A_1 will be consistent and asymptotically Gaussian.

Failure of the exogeneity condition. Now consider the case $\phi_2 \neq 0$ in (S.39). The actual proxy-SVAR moment conditions now are:

$$\mathrm{E}(u_t z_t') = \Sigma_{u,z} \equiv \begin{pmatrix} \Sigma_{u1,z} \\ \Sigma_{u2,z} \end{pmatrix} = \phi_1 B_1 + \phi_2 B_2^o \equiv \begin{pmatrix} B_{11}\phi_1 \\ B_{21}\phi_1 \end{pmatrix} + \begin{pmatrix} B_{2,11}^o\phi_2 \\ B_{2,21}^o\phi_2 \end{pmatrix} ,$$

where $B_2^o := (B_{2,11}^o, B_{2,21}^{o\prime})'$ denotes the column of the matrix B_2 (see (2) in the paper) associated with the non-instrumented structural shock not orthogonal to the proxies, $\varepsilon_{NI,t}^o$. In this case, under the unit effect normalization $B_{11} = 1$ we have:

$$\hat{\gamma}_T \equiv \begin{pmatrix} \hat{\gamma}_{1,T} \\ \hat{\gamma}_{2,T} \end{pmatrix} \xrightarrow{p} \begin{pmatrix} \gamma_{0,1}^{+} \\ \gamma_{0,2}^{+} \end{pmatrix} = \begin{pmatrix} \phi_1 + B_{2,11}^o \phi_2 \\ B_{21}\phi_1 + B_{2,21}^o \phi_2 \end{pmatrix},$$

where $\gamma_{0,1}^{+}$ and $\gamma_{0,2}^{+}$ are 'pseudo-true' values. Clearly, $\gamma_0^{+} := (\gamma_{0,1}^{+}, \gamma_{0,2}^{+\prime})' \neq \gamma_0$ when $\phi_2 \neq 0$.

The estimator of B_{21} (given the unit effect normalization $B_{11} = 1$) is:

$$\hat{B}_{21,T} = \frac{\hat{\gamma}_{2,T}}{\hat{\gamma}_{1,T}} \xrightarrow{p} \frac{B_{21}\phi_1 + B_{2,21}^o \phi_2}{\phi_1 + B_{2,11}^o \phi_2} = \frac{\gamma_{0,2}^{+}}{\gamma_{0,1}^{+}} = B_{21}^{+}$$

so that it will be asymptotically biased with the asymptotic bias depending on the magnitude of the non-exogeneity parameter ϕ_2. Again,

$$T^{1/2}(\hat{B}_{21,T} - B_{21}^{+}) = T^{1/2}\left(\frac{\hat{\gamma}_{2,T}}{\hat{\gamma}_{1,T}} - \frac{\gamma_{0,2}^{+}}{\gamma_{0,1}^{+}}\right) = T^{1/2}\left(\frac{\hat{\gamma}_{2,T} - \gamma_{0,2}^{+} + \gamma_{0,2}^{+}}{\hat{\gamma}_{1,T}} - \frac{\gamma_{0,2}^{+}}{\gamma_{0,1}^{+}}\right)$$

$$= T^{1/2}\left(\frac{\hat{\gamma}_{2,T} - \gamma_{0,2}^{+}}{\hat{\gamma}_{1,T}}\right) + T^{1/2}\frac{\gamma_{0,2}^{+}}{\hat{\gamma}_{1,T}} - T^{1/2}\frac{\gamma_{0,2}^{+}}{\gamma_{0,1}^{+}} \xrightarrow{d} \frac{1}{(\phi_1 + B_{2,11}^o \phi_2)}\xi_{\gamma_2} + o_p(1) \tag{S.41}$$

hence $T^{1/2}(\hat{B}_{21,T} - B_{21}^{+})$ is asymptotically biased but Gaussian distributed. Using (9) in the paper and simple arguments, also the estimator of the parameters in A_1 will be asymptotically biased and normally distributed.

The result in (S.41) motivates, without loss of generality, the claim at the end of Section 4 of the paper that if e.g. the proxies in v_t are strong for $\tilde{\varepsilon}_{2,t}$ but fail to be exogenous (orthogonal) to (some of) the structural shocks in $\varepsilon_{1,t}$, the quantity $T^{1/2}(\hat{\theta}_T - \theta_0^{+})$ is still asymptotic Gaussian, $\theta_0^{+} \neq \theta_0$ being a pseudo-true value. The result in (S.41) also motivates the claim that the quantity $\Gamma_T^* := T^{1/2}V_\theta^{-1/2}(\hat{\theta}_T^* - \hat{\theta}_T)$ remains, conditional on the data, asymptotically Gaussian also when the exogeneity condition fails. This fact is documented empirically in Table SM1 which documents the rejection performance of the test of instrument relevance in samples of $T = 250$ observations when the exogeneity condition fails. The underlying data generating process corresponds to the 'Strong proxy' hypothesis already considered in the upper panel of Table 1 (see Section S.8), with the important difference that now the proxy fails to be exogenous to the non-instrumented shocks. It is seen that rejection frequencies in Table SM1 (the exogeneity condition fails) match those in Table 1 of the paper (the exogeneity condition holds).

S.10 Additional empirical illustration: fiscal multipliers from a fiscal proxy-SVAR

Fiscal multipliers are key statistics for understanding how fiscal policy changes stimulate (or contract) the economy. There is a large debate in the empirical literature on the size of fiscal multipliers, especially the size and uncertainty surrounding the tax multiplier, see Ramey (2019). This lack of consensus also characterizes studies based on fiscal proxy-SVARs as shown by the works in e.g. Mertens and Ravn (2014) and Caldara and Kamps (2017).

Using fiscal proxies for the fiscal shocks, Mertens and Ravn (2014) uncover a large tax multiplier and show that the tax multiplier is larger than the fiscal spending multiplier. Instead, Caldara and Kamps (2017) focus on the identification of fiscal reaction functions and reach the opposite conclusion using non-fiscal proxies for non-fiscal shocks and a Bayesian penalty function approach. In this section, we contribute to this debate by comparing results obtained with our indirect-MD approach which, as in Caldara and Kamps (2017), requires instrumenting non-fiscal shocks, with a direct approach which, as in Mertens and Ravn (2014), requires instrumenting the fiscal shocks directly.

Suppose that the objective of the analysis is to infer the tax and fiscal spending multipliers from a VAR for $Y_t := (TAX_t, G_t, GDP_t, RR_t)'$ $(n = 4)$, where TAX_t is measure of per capita real tax revenues, G_t per capita real government spending, GDP_t per capita real output and RR_t the (ex-post) real interest rate measured as $RR_t := R_t - \pi_t$, R_t being a short term nominal interest rate and π_t the inflation rate. The tax and fiscal spending multipliers are defined as the response of output (GDP) following exogenous fiscal policy interventions on taxes and fiscal spending, see equation (S.43) below for formal definitions.

Reduced form. We consider quarterly data on the sample 1950:Q1-2006:Q4 ($T = 228$ quarterly observations). All variables are taken from Caldara and Kamps (2017), where a more detailed explanation of the dataset can be found. All series are expressed in logs and are linearly detrended. The reduced VAR for $Y_t := (TAX_t, G_t, GDP_t, RR_t)'$ includes $p = 4$ lags and a constant. Standard residual-based diagnostic tests show that VAR disturbances are serially uncorrelated but display conditional heteroskedasticity.

Structural shocks. Let $\varepsilon_{1,t} := (\varepsilon_t^{tax}, \varepsilon_t^{g})'$ be the vector of target structural shocks ($k = 2$), where ε_t^{tax} denotes the tax shock and ε_t^{g} the fiscal spending shock. The auxiliary shocks are collected in the vector $\varepsilon_{2,t} := (\varepsilon_t^{y}, \varepsilon_t^{mp})'$ ($n - k = 2$); ε_t^{y} is an output shock and ε_t^{mp} can be interpreted as a shock to the real interest rate, here interpreted likewise a monetary policy shock to simplify.

The B-form of the model (the analogue of (2) in the paper) is given by

$$\underset{u_t}{\underbrace{\begin{pmatrix} u_t^{tax} \\ u_t^{g} \\ u_t^{y} \\ u_t^{rr} \end{pmatrix}}} = \underset{B_1}{\underbrace{\begin{pmatrix} \beta_{tax,tax} & \beta_{tax,g} \\ \beta_{g,tax} & \beta_{g,g} \\ \beta_{y,tax} & \beta_{y,g} \\ \beta_{rr,tax} & \beta_{rr,g} \end{pmatrix}}} \underset{\varepsilon_{1,t}}{\underbrace{\begin{pmatrix} \varepsilon_t^{tax} \\ \varepsilon_t^{g} \end{pmatrix}}} + B_2 \underset{\varepsilon_{2,t}}{\underbrace{\begin{pmatrix} \varepsilon_t^{y} \\ \varepsilon_t^{mp} \end{pmatrix}}} \tag{S.42}$$

where $u_t := (u_t^{tax}, u_t^{g}, u_t^{y}, u_t^{rr})'$ is the vector of VAR innovations and $\beta_{y,tax}$ and $\beta_{y,g}$ are the coefficients that capture the on-impact responses of output to the tax shock and fiscal spending shock, respectively. Since $k > 1$, we need $\ell \geq \frac{1}{2}k(k-1) = 1$ additional restrictions on the proxy-SVAR parameters for identification, discussed below.

Fiscal multipliers. Assume for the moment that the proxy-SVAR is identified. The fiscal multipliers can be obtained by properly scaling the responses of output to the identified fiscal shocks. In particular, the dynamic fiscal multipliers are defined as the quantities[3]

$$\mathbb{M}_{h,tax} := \frac{\beta_{y,tax}(h)}{\beta_{tax,tax}} \times Sc_{y,tax} \quad , \quad \mathbb{M}_{h,g} := \frac{\beta_{y,g}(h)}{\beta_{g,g}} \times Sc_{y,g} \quad , \; h = 0, 1, ... \tag{S.43}$$

where $\beta_{y,tax}(h) := \frac{\partial GDP_{t+h}}{\partial \varepsilon_t^{tax}}$ is the dynamic response of tax revenues to the tax shock after h periods, $\beta_{tax,tax} \equiv \beta_{tax,tax}(0)$, $\beta_{y,g}(h) := \frac{\partial GDP_{t+h}}{\partial \varepsilon_t^{g}}$ and $\beta_{g,g} \equiv \beta_{g,g}(0)$ are defined accordingly, and $Sc_{y,tax}$ and $Sc_{y,g}$ are scaling factors which serve to convert the dynamic structural responses into US dollars.

Direct identification strategy, weak-instrument robust method. The 'direct' identification approach hinges on the availability of (at least) two proxies for the two auxiliary shocks in $\varepsilon_{1,t} := (\varepsilon_t^{tax}, \varepsilon_t^{g})'$, other than $\ell \geq 1$ additional restrictions on the proxy-SVAR parameters. To simplify, we posit that there exist two proxies for the fiscal shocks that we collect in the vector $z_t := (z_t^{tax}, z_t^{g})'$, and assume that the counterpart of system (4) is given by the system

$$\underset{z_t}{\underbrace{\begin{pmatrix} z_t^{tax} \\ z_t^{g} \end{pmatrix}}} = \underset{\Phi}{\underbrace{\begin{pmatrix} \varphi_{tax,tax} & 0 \\ 0 & \varphi_{g,g} \end{pmatrix}}} \underset{\varepsilon_{1,t}}{\underbrace{\begin{pmatrix} \varepsilon_t^{tax} \\ \varepsilon_t^{g} \end{pmatrix}}} + \underset{\omega_t}{\underbrace{\begin{pmatrix} \omega_t^{tax} \\ \omega_t^{g} \end{pmatrix}}} \tag{S.44}$$

where $\omega_t := (\omega_t^{tax}, \omega_t^{g})'$ is a vector of measurement errors assumed orthogonal to the structural shocks ε_t. In this example, the matrix Φ in (S.44) has been

[3] The definitions in (S.43) are consistent with Blanchard and Perotti (2002), Mertens and Ravn (2014) and Caldara and Kamps (2017).

specified diagonal to capture the fact that the proxy z_t^{tax} solely instruments the tax shock (through the relevance parameter $\varphi_{tax,tax}$), and the proxy z_t^g solely instruments the fiscal spending shock (through the relevance parameter $\varphi_{g,g}$). Notably, the diagonal structure assumed for Φ in (S.44) provides $\ell = 2 > \frac{1}{2}k(k-1)$ additional restrictions that would (over-)identify the proxy-SVAR in the presence of strong proxies but instead would leave it unidentified in the presence of weak proxies.[4]

Imagine that we have the suspect that the proxies $z_t := (z_t^{tax}, z_t^g)'$ are weak for the target shocks $\varepsilon_{1,t} := (\varepsilon_t^{tax}, \varepsilon_t^g)'$, so that we proceed by implementing Montiel Olea, Stock and Watson's (2020) weak-identification robust method. More precisely, we construct an S-region for the simultaneous response of real output to the tax and fiscal spending shocks. To do so, first we assume for simplicity and without loss of generality that the VAR in (1) has just one lag so that $\mathcal{A}_y \equiv \Pi_1 = \Pi$. Then, we consider the null hypothesis that at the horizon h, the simultaneous response of real output to the fiscal shocks $\varepsilon_{1,t} := (\varepsilon_t^{tax}, \varepsilon_t^g)'$ is equal to

$$\gamma_{GDP,\varepsilon_{1,t}}(h) = \left(\frac{\partial GDP_{t+\bar{h}}}{\partial \varepsilon_t^{tax}} , \frac{\partial GDP_{t+\bar{h}}}{\partial \varepsilon_t^{g}} \right) := \iota_3'(\Pi)^h B_1 = (a_{tax}^h, b_g^h) \qquad \text{(S.45)}$$

where $\iota_3' := (0, 0, 1, 0)$ is the selection vector that picks out real output from the vector $Y_t := (TAX_t, G_t, GDP_t, RR_t)'$ and a_{tax}^h and b_g^h are postulated response values. Observe that assuming constant scaling factors $Sc_{y,tax}$ and $Sc_{y,g}$ in (S.43), the multipliers $\mathbb{M}_{h,tax}$ and $\mathbb{M}_{h,g}$ can be easily computed from the postulated responses (a_{tax}^h, b_g^h).

By post-multiplying both sides of (S.45) by Φ', one obtains the restrictions (valid under the null)

$$\iota_3'(\Pi)^h \Sigma_{u,z} - (a_{tax}^h, b_g^h)\Phi' = (0, 0) \qquad \text{(S.46)}$$

which can be used to construct an asymptotic valid S-region for a_{tax}^h and b_g^h through test inversion.

Now we show that irrespective of the rank of Φ in (S.44), by imposing $k^2 = 4$ restrictions on the block B_{11} of B_1 (recall that $B_1' := (B_{11}' \vdots B_{21}')$), equation (S.46) can be solved such that Φ is expressed as function of the reduced form covariance parameters in $\Sigma_{u,z}$. In particular, using the partition in (2) in the paper, we rewrite the moment conditions in (5) as (recall that $n = 4$)

$$\begin{pmatrix} \Sigma_{u_1,z} \\ \Sigma_{u_2,z} \end{pmatrix} = \begin{pmatrix} B_{11}\Phi' \\ B_{21}\Phi' \end{pmatrix} \qquad \begin{matrix} 2 \times 2 \\ 2 \times 2 \end{matrix} \qquad \text{(S.47)}$$

[4] According to Definition 1(a) of Section 3, if the proxies in z_t are strong for $\varepsilon_{1,t}$, the matrix Φ in (S.44) has full column rank regardless of the sample size. With strong instruments and Φ diagonal, the proxy-SVAR is overidentified and testable against the data.

where the reduced form covariance matrix $\Sigma_{u,z}$ has been decomposed into the two blocks $\Sigma_{u_1,z}$ and $\Sigma_{u_2,z}$, respectively, with dimensions reported alongside blocks. To solve the moment conditions (S.47) for Φ', we impose $k^2 = 4$ restrictions on B_{11}, namely

$$B_{11} = \begin{pmatrix} \beta_{tax,tax} & \beta_{tax,g} \\ \beta_{g,tax} & \beta_{g,g} \end{pmatrix} := \begin{pmatrix} 1 & c^0_{tax,g} \\ 0 & 1 \end{pmatrix} \tag{S.48}$$

where $c^0_{tax,g}$ is a postulated response value of tax revenues to the fiscal spending shock. The restrictions in (S.48) amount to the 'unit effect' normalizations ($\beta_{tax,tax} = 1$ and $\beta_{g,g} = 1$), to the zero contemporaneous response of fiscal spending to the tax shock ($\beta_{g,tax} = 0$) and to postulated that the on-impact response of tax revenues to the fiscal spending shock is $c^0_{tax,g}$, where $c^0_{tax,g}$ can be possibly zero. Overall, the total number of restrictions placed on $(B_1' \vdots \Phi')'$ is $k^2 = 4 + 2 = 6$. By solving the equation $\Sigma_{u_1,z} = B_{11}\Phi'$ for Φ', yields

$$\Phi_p' = \begin{pmatrix} 1 & c^0_{tax,g} \\ 0 & 1 \end{pmatrix}^{-1} \Sigma_{u_1,z} = \begin{pmatrix} 1 & -c^0_{tax,g} \\ 0 & 1 \end{pmatrix} (I_k \vdots 0_{k\times(n-k)})\Sigma_{u,z} \tag{S.49}$$

where the notation used for Φ', Φ_p', remarks the fact that the matrix of relevance parameters incorporates the postulated response value $\beta_{tax,g} = c^0_{tax,g}$. The expression in (S.49) suggests that for $c^0_{tax,g}$ given, a plug-in estimator of Φ_p' is given by

$$\hat{\Phi}_p' = \begin{pmatrix} 1 & -c^0_{tax,g} \\ 0 & 1 \end{pmatrix} (I_k \vdots 0_{k\times(n-k)})\hat{\Sigma}_{u,z} \tag{S.50}$$

and therefore is consistent under the conditions of Lemma S1 if all postulated restrictions are valid.

Let $\kappa := (vec(\Pi)', vec(\Sigma_{u,z})')'$ be the vector containing all reduced form proxy-SVAR parameters, with κ_0 being the true value and $\hat{\kappa}_T$ the corresponding estimator (see Lemma S1). Regardless of the strength of the proxies, under Assumptions 1-2, it holds the asymptotic normality result $T^{1/2}(\hat{\kappa}_T - \kappa_0) \xrightarrow{d} N(0, V_\kappa)$. Thus, by taking the vec of the expression in (S.46), the null hypothesis we can be re-written as

$$S(\kappa_0, a^h_{tax}, b^h_g, c^0_{tax,g}) = \text{vec}\left\{\iota_3'(\Pi)^h\Sigma_{u,z} - (a^h_{tax}, b^h_g)\Phi'\right\} = 0_{2\times 1}$$

and, by a simple delta-method argument, this implies the result

$$T^{1/2}S(\hat{\kappa}_T, a^h_{tax}, b^h_g, c^0_{tax,g}) \xrightarrow{d} N(0_{2\times 1}, V_S)$$

where V_S is a covariance matrix that depends on V_κ. Thus, regardless of the strength of the proxies, a valid ϱ-level test for the null hypothesis that the postulated values $(a^h_{tax}, b^h_g, c^0_{tax,g})$ are true rejects whenever

$$TS(\hat{\kappa}_T, a^h_{tax}, b^h_g, c^0_{tax,g})'\hat{V}_S^{-1}S(\hat{\kappa}_T, a^h_{tax}, b^h_g, c^0_{tax,g}) > \chi^2_{2,1-\varrho} \quad \text{(S.51)}$$

where $\hat{V}_S$ is a consistent estimator of V_S and $\chi^2_{2,1-\varrho}$ is the $(1-\varrho)100\%$ quantile of the chi-distribution with two degree of freedom. An asymptotically valid weak-identification robust confidence set for a^h_{tax}, b^h_g and $c^0_{tax,g}$ with coverage $1-\varrho$ will contain all postulated values that are not rejected by the Wald-type test. Confidence intervals for the tax and fiscal spending shocks at the horizon h can be obtained by the projection method.

We employ the vector of proxies $z_t := (z^{tax}_t, z^g_t)'$, where, as in Mertens and Ravn (2014), z^{tax}_t is a series of unanticipated tax changes built upon Romer and Romer's (2010) narrative records on tax policy decisions, and z^g_t is narrative measure of expected exogenous changes in military spending constructed by Ramey (2011). Our bootstrap pre-test for the relevance of $z_t := (z^{tax}_t, z^g_t)'$ rejects the null of strong proxies with a p-value of 0.003. We ignore for the moment the outcome of the bootstrap test for instrument relevance and proceed by estimating the dynamic multipliers in (S.43) by CMD pretending that $z_t := (z^{tax}_t, z^g_t)'$ is strong for $\varepsilon_{1,t} := (\varepsilon^{tax}_t, \varepsilon^g_t)'$. The impact and peak tax and fiscal spending multipliers are summarized in the left column of Table SM2.[5] The estimated peak fiscal spending multiplier is 1.52 (at three quarters) with 68%-MBB confidence interval (-0.73, 3.38), while the estimated peak tax multiplier is 2.46 (at three quarters) with 68%-MBB confidence interval (-0.91, 9.76). Table SM2 also reports the estimated elasticity of tax revenues and fiscal spending to output, two crucial parameters related to the size of fiscal multipliers, see Mertens and Ravn (2014) and Caldara and Kamps (2017). The elasticity of fiscal spending to output is close to zero, while the elasticity of tax revenues to output is almost 3.5, a value comparable to the findings in Mertens and Ravn (2014). Likewise the multipliers, also the elasticity of tax revenues to output is estimated with a large 68%-MBB confidence interval. Figure SM1 plots the so-obtained dynamic fiscal multipliers over an horizon of $h_{\max}$ =40 quarters, with 68%-MBB confidence bands. It confirms that by using standard methods the fiscal multipliers are estimated with great uncertainty, a somewhat expected result in light of the outcome of the test for instrument relevance. The test for the over-identification restriction (recall that Φ is diagonal) has p-value 0.27. This result must be interpreted with caution in light of the detected weakness of the fiscal proxies.

We move on by robustifying the inference by Montiel Olea, Stock and Watson's (2020) weak-instrument robust approach. Keeping the matrix of rel-

[5] We normalize the signs of the responses of output consistently with a fiscal expansions induced by exogenous tax cuts and increases in fiscal spending.

evance parameters Φ diagonal as in (S.44), we impose the four restrictions in (S.48) and for postulated values of the responses a^h_{tax} and b^h_g in (S.45) and (S.46), we invert the Wald-type test in (S.51) and form an S-region with asymptotic nominal coverage of 68%. Based on economic considerations and the survey in Ramey (2019), we postulate values for the tax multiplier ranging from 0 up to 6 and values for the fiscal spending multiplier ranging from 0 up to 3, respectively. Results are summarized in the central column of Table SM2. We find that the projected 68%-identification robust confidence interval for the peak fiscal spending multiplier is (0, 3) with associated Hodges-Lehmann point-estimate of 1.06 (after three quarters), while the projected 68%-identification robust confidence interval for the peak tax multiplier is (0.37, 6) with associated Hodges-Lehmann point estimate of 2.55 (after three quarters).[6].

Recall that the confidence sets for the fiscal multipliers have been obtained by imposing a total of six restrictions on the parameters of $(B_1' \vdots \Phi')'$ (two restrictions on Φ, and four on the block B_{11} of B_1). Montiel Olea *et al.* (2021) observe that it is yet unclear how to test overidentifying restrictions in these situations. We notice that the six restrictions on $(B_1' \vdots \Phi')'$ imply a diagonal structure for the reduced form covariance matrix $\Sigma_{u_1,z} = B_{11}\Phi'$, a testable restriction with standard methods (recall indeed that the estimator of $\Sigma_{u_1,z}$ is consistent and asymptotically Gaussian regardless of the strength of the proxies). A Wald-type test for $\Sigma_{u_1,z}$ diagonal delivers a p-value of 0.34 which suggests that the highly restricted structure imposed on the proxy-SVAR is (at least) partially supported by the data.

Indirect identification strategy. The counterpart of the A-form representation of the proxy-SVAR (the analogue of (8) in the paper) is given by the two-equations system

$$\underbrace{\begin{pmatrix} \alpha_{tax,tax} & \alpha_{tax,g} \\ \alpha_{g,tax} & \alpha_{g,g} \end{pmatrix}}_{A_{11}'} \underbrace{\begin{pmatrix} u_t^{tax} \\ u_t^{g} \end{pmatrix}}_{u_{1,t}} + \underbrace{\begin{pmatrix} \alpha_{tax,y} & \alpha_{tax,rr} \\ \alpha_{g,y} & \alpha_{g,rr} \end{pmatrix}}_{A_{12}'} \underbrace{\begin{pmatrix} u_t^{y} \\ u_t^{rr} \end{pmatrix}}_{u_{2,t}} = \underbrace{\begin{pmatrix} \varepsilon_t^{tax} \\ \varepsilon_t^{g} \end{pmatrix}}_{\varepsilon_{1,t}}. \tag{S.52}$$

The crucial assumption here (Assumption 4 in the paper) is that there exists proxies for the auxiliary shocks in $\varepsilon_{2,t} := (\varepsilon_t^y, \varepsilon_t^{mp})'$, recall that $n-k=2$ and $s \leq n-k$, where s is the dimension of the vector of instruments v_t used for the auxiliary shocks. However, since $k>1$, it will be also necessary to impose restrictions on the parameters in $A_1' := (A_{11}' \vdots A_{12}')$ (other than the instruments

[6] Hodges-Lehmann point estimates are the points in the S-region with higher p-values. We use this estimator to compare results with those obtained with the indirect-MD approach discussed next.

v_t) to achieve identification (Propositions 1 and 2 of the paper).

We consider the following vector of instruments for the $v_t := (v_t^{tfp}, v_t^{rr})'$, $s = (n-k)$, where as in Caldara and Kamps (2017), v_t^{tfp} is Fernald's (2014) measure of TFP and is used as proxy for the output shock ε_t^y, and v_t^{rr} is Romer and Romer's (2004) narrative series of monetary policy shocks and is used as proxy for the shock to the real interest rate, ε_t^{mp}. Thus, we have in mind a relationship between proxies and auxiliary shocks of the form

$$\underbrace{\begin{pmatrix} v_t^{tfp} \\ v_t^{rr} \end{pmatrix}}_{v_t} = \Lambda \underbrace{\begin{pmatrix} \varepsilon_t^{y} \\ \varepsilon_t^{mp} \end{pmatrix}}_{\varepsilon_{2,t}} + \underbrace{\begin{pmatrix} \omega_t^{tfp} \\ \omega_t^{rr} \end{pmatrix}}_{\varrho_t} \tag{S.53}$$

with $\omega_t := (\omega_t^{tfp}, \omega_t^{rr})'$ measurement error. In this case the moment conditions in (12)-(13) of the paper provide $m = \frac{1}{2}k(k+1) + ks = 7$ independent moment conditions which can potentially be used to estimate seven parameters α that enter the matrix $A_1' := (A_{11}' \vdots A_{12}')$. Based on Caldara and Kamps (2017), we postulate that tax revenues do not instantaneously react to fiscal spending, implying $\alpha_{tax,g} = 0$ in (S.52).

Since the proxy v_t^{rr} is available form 1969Q1, we use the common sample period (1969Q1-2006Q4) for convenience, hence the analysis is based on a total of T =152 quarterly observations. Empirical results are as follows. The bootstrap pre-test for the relevance of the proxies $v_t := (v_t^{tfp}, v_t^{rr})'$ supports markedly the null of strong proxies with a p-value of 0.88.[7] The impact and peak tax and fiscal spending multipliers are summarized in the right column of Table SM2. The estimated peak fiscal spending multiplier is 1.54 (at two quarters) with 68%-MBB confidence intervals (0.64, 1.76), and the estimated peak tax multiplier is 0.96 (at four quarters) with 68%-MBB confidence interval (0.18, 1.44). The estimated elasticity of tax revenues to output is 2.06 with 68%-MBB confidence interval (1.6, 2,5), a point-estimate very close to the 2.08 value calibrated by Blanchard and Perotti (2002) by the 'OECD method'.

Figure SM1 plots the dynamic fiscal multipliers obtained with the indirect-MD approach over the horizon of $h_{\max}$ =40 quarters with 68%-MBB confidence bands. The difference with the multipliers estimated with the previous approach is striking. In her recent review of the theoretical and empirical literature, Ramey (2019) documents a substantial lack of consensus on the size and uncertainty surrounding the tax multiplier. Our empirical analysis seems to suggest that a large portion of the differences can be ascribed to the strength of the proxies employed in the two approaches.

[7] Formally, the test is computed as DH multivariate normality test computed on the sequence $\{\hat{\beta}^*_{2,T:1}, ..., \hat{\beta}^*_{2,T:N}\}$ of MBB replications, with $N = [T^{1/2}]$=12.

S.11 References

Angelini, G. and Fanelli, L. (2019), Exogenous uncertainty and the identification of Structural Vector Autoregressions with external instruments, *Journal of Applied Econometrics* 34, 951-971.

Basawa, I., Mallik, A., McCormick, W., Reeves, J. and Taylor, R. (1991), Bootstrapping unstable first-order autoregressive processes, *The Annals of Statistics* 19, 1098–1101.

Blanchard, O. and Perotti, R. (2002), An empirical characterization of the dynamic effects of changes in government spending and taxes on output, *Quarterly Journal of Economics* 117, 1329–1368.

Brüggemann, R., Jentsch, C. and Trenkler, C. (2016), Inference in VARs with conditional volatility of unknown form, *Journal of Econometrics* 191, 69-85.

Boubacar Mainnasara, Y. and Francq, C. (2011), Estimating structural VARMA models with uncorrelated but non-independent error terms, *Journal of Multivariate Analysis* 102, 496-505.

Caldara, D. and Kamps, C. (2017), The analytics of SVARs: A unified framework to measure fiscal multipliers, *Review of Economic Studies* 84, 1015-1040.

Cavaliere, G. and Georgiev, I. (2020), Inference under random limit bootstrap measures, *Econometrica* 88, 2547-2974.

Fernald, J. (2014), A quarterly, utilization-adjusted series on Total Factor Productivity, *Federal Reserve Bank of San Francisco Working Paper* No. 2012-19.

Francq, C. and Raïssi, H. (2006), Multivariate portmanteau test for autoregressive models with uncorrelated but nonindependent errors, *Journal of Time Series Analysis* 28, 454-470.

Jentsch, C. and Lunsford, K.C. (2019b), Asymptotic valid bootstrap inference for Proxy SVARs, *Working Paper.* Version dated May 2019.

Kallenberg, O. (1997), *Foundations of Modern Probability.* New York: Springer.

Magnus, J.R. and Neudecker, H. (1999), *Matrix differential calculus with applications in Statistics and Econometrics*, Wiley & Sons, 2nd edition.

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

Mertens, K. and Ravn, M. (2014), A reconciliation of SVAR and narrative estimates of tax multipliers, *Journal of Monetary Economics* 68, S1-S19.

Montiel Olea, J.L., Stock, J.H. and Watson, M.W. (2020), Inference in SVARs identified with an external instrument, *Journal of Econometrics*, forthcoming

Ramey, V. (2011), Identifying Government Spending Shocks: It's All in the Timing, *Quarterly Journal of Economics* 126, 1–50. 2016.

Ramey, V. (2019), Ten years after the Financial Crisis: What have we learned from the renaissance in fiscal research?, *Journal of Economic Perspectives*, 33, 89-114.

Romer, C., and Romer, D. (2004), A new measure of monetary policy shocks: Derivation and implications, *American Economic Review* 94, 1055–1084.

Romer, C, and Romer, D. (2010), The macroeconomic effects of tax changes: Estimates of based on a new measure of fiscal shocks, *American Economic Review* 100, 763-801.

Rejection frequencies				
Strong proxy - no exogeneity				
	$T = 250$		$T = 1000$	
	$corr = 59\%$ ($corr^{exog.} = 33\%$)		$corr = 59\%$ ($corr^{exog.} = 33\%$)	
θ	DH	KS	DH	KS
$\beta_{2,1}$		0.05(0.05)		0.05(0.05)
$\beta_{2,2}$	0.05(0.05)	0.05(0.05)	0.05(0.05)	0.05(0.05)
$\beta_{2,3}$		0.05(0.05)		0.05(0.05)
λ		0.05(0.05)		0.05(0.05)

Table SM1: Monte Carlo results (details on the data generating processes may be found in the SM). Empirical rejection frequencies of the bootstrap test for strong against weak proxy, based on 20000 simulations and tuning parameter $N := [T^{1/2}]$. $corr = corr(\upsilon_t, \varepsilon_{2,t})$ is the correlation between the instrument υ_t and the structural shock $\varepsilon_{2,t}$, and $corr^{exog.} = corr(\upsilon_t, \varepsilon_{1,t})$ is the correlation between the instrument υ_t and the structural shock $\varepsilon_{1,t}$. KS is Lilliefors' (1967) version of Kolgomorov-Smirnov univariate normality test; DH is Doornik and Hansen's (2008) multivariate normality test. Numbers in parentheses refer to GARCH-type VAR innovations (see SM). All tests are computed at the 5% nominal significance level.

Fiscal proxy-SVARs

Multipliers & elasticities

Direct standard	Direct A&R	Indirect-MD
$\mathbb{M}_{0,g} = \underset{(-0.6359;2.3364)}{1.0809}$	$\mathbb{M}_{0,g} = \underset{(0.0000;3.000)}{0.7440}$	$\mathbb{M}_{0,tr} = \underset{(0.9009;1.5594)}{1.4662}$
$\mathbb{M}_{0,tr} = \underset{(-1.0294;7.5788)}{1.8394}$	$\mathbb{M}_{0,tr} = \underset{(0.2162;6.000)}{1.9072}$	$\mathbb{M}_{0,tr} = \underset{(0.0431;0.9313)}{0.6382}$
$\mathbb{M}_{3,g} = \underset{(-0.7307;3.3828)}{1.5214[3]}$	$\mathbb{M}_{3,g} = \underset{(0.0000;3.000)}{1.0639[3]}$	$\mathbb{M}_{2,g} = \underset{(0.6411;1.7603)}{1.5365[2]}$
$\mathbb{M}_{3,tr} = \underset{(-0.9058;9.7567)}{2.4598[3]}$	$\mathbb{M}_{3,tr} = \underset{(0.3661;6.000)}{2.5513[3]}$	$\mathbb{M}_{4,tr} = \underset{(0.1800;1.4418)}{0.9553[4]}$
$\psi_y^{tr} = \underset{(0.0608;4.8160)}{3.4814}$		$\psi_y^{tr} = \underset{(1.6419;2.4932)}{2.0673}$

Diagnostic tests

p-value $DH_{\theta=B_1} = 0.0031$ $\qquad$ p-value $DH_{\theta=\widetilde{B}_2} = 0.8224$

Table SM2: Results based on U.S. quarterly data. **Multipliers & elasticities**: Estimated multipliers and elasticities with associated 68%-MBB confidence intervals and associated lag in brackets. **Diagnostic tests**: p-values of the diagnostic tests based on $N := [T^{1/2}]$ bootstrap replications of the CMD estimator. $DH_{\theta=B_1}$ ($DH_{\theta=\tilde{B}_2}$) is Doornik and Hansen's (2008) multivariate normality test computed with respect to the vector of all on-impact parameters in B_1 ($\tilde{B}_2$).

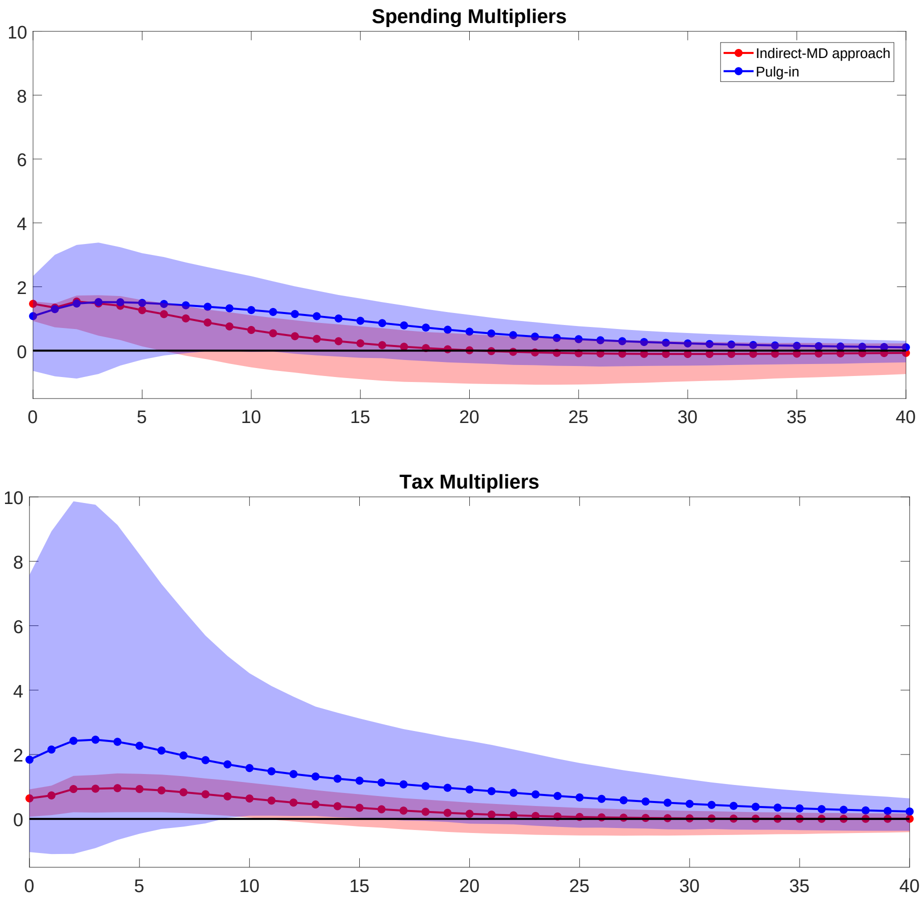

Figure SM1: Spending and Tax multipliers.