Robust inference in structural VARs with long-run restrictions

Guillaume Chevillon, Sophocles Mavroeidis, Zhaoguo Zhan

ROBUST INFERENCE IN STRUCTURAL VARS WITH LONG-RUN RESTRICTIONS

RESEARCH CENTER

GUILLAUME CHEVILLON, SOPHOCLES MAVROEIDIS,
ZHAOGUO ZHAN

ESSEC WORKING PAPER 1702
JANUARY 2017

Robust inference in structural VARs with long-run restrictions*

Guillaume Chevillon[†]
ESSEC Business School and CREST

Sophocles Mavroeidis[‡]
University of Oxford

Zhaoguo Zhan[§]
Kennesaw State University

November 22, 2016

Abstract

Long-run restrictions are a very popular method for identifying structural vector autoregressions, but they suffer from weak identification when the data is very persistent, i.e., when the highest autoregressive roots are near unity. Near unit roots introduce additional nuisance parameters and make standard weak-instrument-robust methods of inference inapplicable. We develop a method of inference that is robust to both weak identification and strong persistence. The method is based on a combination of the Anderson-Rubin test with instruments

*We would like to thank Jean-Marie Dufour, Patrick Fève, Tassos Magdalinos, Nour Meddahi, Adrian Pagan, Jean-Pierre Urbain, and seminar participants at the Universities of Cambridge, Maastricht, Melbourne, Toulouse, as well as CREST and the European University Institute, the North American Winter Meeting of the Econometric Society, the NBER Summer Institute, the Barcelona GSE Summer Forum, the CRETE, IAAE and Oxmetrics Users conferences, the IWH-CIREQ Macroeconometrics Workshop in Halle, the 24th symposium of the SNDE for helpful comments and discussion. Mavroeidis would like to thank the European Commission for research support under a FP7 Marie Curie Fellowship CIG 293675. Zhan acknowledges the financial support from the National Natural Science Foundation of China, Project No. 71501104.

[†]ESSEC Business School, Ave. B. Hirsch, 95000 Cergy-Pontoise, France.

[‡]Department of Economics and Institute for New Economic Thinking at the Oxford Martin School, University of Oxford, Manor Road, Oxford, OX1 3UQ, United Kingdom. Email: sophocles.mavroeidis@economics.ox.ac.uk

[§]Department of Economics, Finance and Quantitative Analysis, Kennesaw State University, GA 30144, United States.

derived by filtering potentially non-stationary variables to make them near stationary. We apply our method to obtain robust confidence bands on impulse responses in two leading applications in the literature.

Keywords: SVARs, identification, weak instruments, near unit roots, IVX.
JEL: C12, C32, E32

"It is better to be vaguely right than exactly wrong." Carveth Read

1 Introduction

Since the seminal paper of Sims (1980), structural vector autoregressions (SVARs) have become a very popular method for analysing dynamic causal effects in macroeconomics. SVARs can be used to decompose economic fluctuations into interpretable shocks, such as technology, demand, policy shocks, and trace the dynamic response of macroeconomic variables to such shocks, known as impulse response functions (IRFs). The success of the SVARs relies on (i) their ability to recover the true underlying structural shocks (invertibility); (ii) the validity of the identification scheme; and (iii) the informativeness of the identifying restrictions. Because a SVAR is a system of linear simultaneous equations, the third condition can be expressed as the availability of informative instruments.

In the words of Christiano et al. (2007), "to be useful in practice, VAR-based procedures should accurately characterize [and] uncover the information in the data about the effects of a shock to the economy". In other words, confidence intervals on the model's parameters, e.g., the IRFs to an identified shock, need to have the property that they are (i) as small as possible when instruments are strong (efficiency); and (ii) large when instruments are weak/irrelevant (robustness), see Dufour (1997). Conventional methods based on standard strong-instrument and stationarity assumptions achieve the first objective but fail the second and therefore lead to unreliable inference.

This paper focuses on the identification scheme known as long-run restrictions, proposed by Blanchard and Quah (1989). This assumes that certain shocks (e.g., demand shocks) have no permanent effect on certain economic variables (e.g., output). Long-run restrictions are a popular identification scheme for SVARs, because they seem to be less contentious than short-run identifying restrictions, see e.g., Christiano et al. (2007) and the associated comments and discussion. However, it is well-known that long-run

restrictions can lead to weak identification, see e.g., Pagan and Robertson (1998), and there is presently no method of inference that is fully robust to this problem. The main difficulty is that in this context weak identification arises when instruments are highly persistent, or nearly non-stationary. Therefore, all the available weak identification robust methods of inference, such as the Anderson and Rubin (1949) test, see Staiger and Stock (1997), are inapplicable because they rely on stationary asymptotics. This applies to common pretests of weak identification, too, see Mark Watson's comment on Christiano et al. (2007).

In this paper, we develop a method of inference that is robust to weak instruments as well as near non-stationarity. The method is based on combining recent advances in econometrics on inference with highly persistent data by Magdalinos and Phillips (2009) and Kostakis et al. (2015), see also Phillips (2014), with well-established methods of inference that are robust to weak instruments. The former methods have been developed for predictive regressions or cointegration, and their use in the context of structural inference in simultaneous equations models is new. Our new method of inference controls asymptotic size under a wide range of data generating processes, including standard local-to-unity asymptotics; it has good size in finite samples; it is asymptotically efficient under strong identification and has good power under weak identification; and it is very simple to implement. For illustration, we revisit the empirical evidence in two classic applications of SVARs with long-run restrictions: the original application in Blanchard and Quah (1989) and the hours debate of Galí (1999) and Christiano et al. (2003). In the case of Blanchard and Quah (1989), we find that long-run restrictions yield very weak identification. On the hours debate, we find that the difference specification of Galí (1999) is very well identified, while the level specification of Christiano et al. (2003) is weakly identified.

Long-run restrictions are by now a very common approach to the identification of SVARs. At the time of writing, Blanchard and Quah (1989) had 4821 Google scholar citations, and we found that long-run restrictions appeared in about half of all the articles that used SVARs published between 2005 and 2014 in the top general interest and macro journals in economics.[1] Therefore, the scope of the present paper extends well beyond the two applications that we discuss here.

The paper is structured as follows. Section 2 introduces the model and the long-

[1] American Economic Review, Econometrica, Quarterly Journal of Economics, Journal of Political Economy, Review of Economic Studies, Journal of Monetary Economics, AEJ Macro and Journal of Money, Credit and Banking.

run identification scheme. Section 3 discusses existing methods of inference, highlights the problem and presents our proposed solution. Section 4 gives simulations on the finite-sample size and asymptotic power of our new method. Section 5 presents the two empirical applications and finally, Section 6 concludes. Proofs are given in the Appendix at the end, as well as in a Supplementary Appendix available online, which also contains additional numerical and empirical results.

2 Model

A general SVAR with m lags can be written as

$$B(L)Y_t = \Phi D_t + \varepsilon_t, \quad B(L) = \sum_{j=0}^{m} B_j L^j \tag{1}$$

where L is the lag operator, Y_t is a $n \times 1$ vector of endogenous random variables, B_j are $n \times n$ nonstochastic matrices of parameters, Φ is a matrix of coefficients on deterministic terms D_t, and $E(\varepsilon_t | Y_{t-1}, Y_{t-2}, \ldots) = 0$. The diagonal elements of B_0 are normalized to 1, and $var(\varepsilon_t)$ is a diagonal matrix.

Partition the vector of structural shocks as $\varepsilon_t = \binom{\varepsilon_{1t}}{\varepsilon_{2t}}$, where ε_{1t} is scalar and ε_{2t} is $(n-1) \times 1$. We are interested in identifying ε_{1t}, and the IRF

$$\underset{n\times 1}{g_j} = \frac{\partial Y_{t+j}}{\partial \varepsilon_{1t}}, \quad j = 0, 1, \ldots$$

The long-run identifying restriction is that ε_{2t} has no long-run effect on Y_{1t}. In the literature this is expressed as a zero restriction on elements of the spectral density matrix of Y_t at frequency zero – a Choleski factorization of the long-run variance of Y_t. We work with the (equivalent) instrumental variables (IV) representation of the long-run restrictions in Pagan and Robertson (1998). According to this representation, under the assumption that ε_{1t} has a permanent effect on Y_{1t}, and the long-run restriction that ε_{2t} has no permanent effect on Y_{1t}, the system (1) can be written as:

$$\Delta Y_{1t} = b_{12}' \Delta Y_{2t} + \delta_1' X_{1t} + \varepsilon_{1t} \tag{2}$$

$$\Delta Y_{2t} = \alpha_2 Y_{2,t-1} + \delta_2' X_{2t} + \underbrace{d_{21}\varepsilon_{1t} + v_{2t}}_{u_{2t}}, \tag{3}$$

where X_{1t}, X_{2t} denote vectors containing lags of ΔY_t and deterministic terms D_t, δ_1, δ_2 denote the coefficients on those exogenous and predetermined variables, and u_{2t} is the reduced-form error in Y_{2t}.[2] It is evident that the variables $Y_{2,t-1}$ are excluded from (2), and hence they can be used as instruments for the endogenous regressors ΔY_{2t}. This suffices to identify ε_{1t} and hence trace out the entire IRF with respect to ε_{1t}. Note that v_{2t} is the residual of the projection of the reduced-form error u_{2t} on ε_{1t}.

Note that the representation (2)-(3) with $\alpha_2 < 0$ assumes that no shock has a permanent effect on Y_{2t}, meaning that Y_{2t} is stationary. In the literature on hours (Galí (1999), Christiano et al. (2003)) this is referred to as the levels specification, which is contrasted with the differences specification that assumes Y_{2t} to be non-stationary. The differences specification can be written exactly in the form (2)-(3) if we replace Y_{2t} by ΔY_{2t}, see Appendix A for details. Since the representation (2)-(3) can accommodate both specifications, we do not need to analyze them separately in the methodological part of the paper – we study their empirical implications in Section 5.

The objective of this paper is to develop tests of general hypotheses on the identified structural parameters θ

$$H_0 : r(\theta) = 0 \quad \text{against} \quad H_1 : r(\theta) \neq 0, \tag{4}$$

where $r : \Theta \to \Re^q$, $q \leq \dim \theta$. This includes e.g., the IRF and forecast error variance decomposition.

Example: bivariate SVAR(1) A bivariate SVAR(1) without deterministic terms is given by

$$\Delta Y_{1t} = b_{12} \Delta Y_{2t} + \varepsilon_{1t}, \tag{5}$$

$$\Delta Y_{2t} = \alpha_2 Y_{2,t-1} + d_{21} \varepsilon_{1t} + v_{2t}. \tag{6}$$

The structural parameters $\theta = (b_{12}, \sigma_{\varepsilon_1}, \alpha_2, d_{21})'$, where σ_{ε_1} is the standard deviation of ε_{1t}. This is the simplest possible model that suffices to characterize the inference problem and describe our methodology, so we will use this as a running example throughout the paper. The parameter α_2 plays a crucial role both for the persistence of the data and the identification of the structural parameters. Specifically, when α_2 is close to

[2] This specification is somewhat more general than (1) in that X_{1t} and X_{2t} need not be the same and need not include all m lagged differences of the variables.

zero, $Y_{2,t-1}$ has a near unit root and becomes a weak instrument for ΔY_{2t}, see Pagan and Robertson (1998) and Gospodinov (2010). An example of a simple hypothesis of interest is $r(\theta) = d_{21} - d_{21}^0$ in (4). Inverting an η-level test of this hypothesis produces a $(1-\eta)$-level confidence band for d_{21}, which is the impact response of Y_{2t} to a unit impulse on ε_{1t}.

3 Econometric Methods

The conventional approach is to use Gaussian maximum likelihood (ML) estimation with conditional homoskedasticity. The ML estimator is trivial to obtain in this case. It can be computed in two steps as follows: (i) estimate equation (2) by IV (2SLS) with instrument $Y_{2,t-1}$ for ΔY_{2t}, and save the residual $\hat{\varepsilon}_{1t} = \Delta Y_{1t} - \hat{b}_{12}' \Delta Y_{2t} - \hat{\delta}_1' X_{1t}$; (ii) substitute $\hat{\varepsilon}_{1t}$ for ε_{1t} in the remaining equations (3) and estimate them by OLS.

Under strong-instrument stationary asymptotics, i.e., $\alpha_2 < 0$ and fixed, the asymptotic distribution of Wald statistics for testing general hypotheses (4) is χ^2 and error bands for any smooth function of the parameters can be derived using the delta method, e.g. Mittnik and Zadrozny (1993), or by bootstrapping, e.g., Kilian (1998). When α_2 is small, e.g., $\alpha_2 = O(T^{-a})$, $a > 0$, conventional asymptotic approximations break down and the distributions of Wald statistics depend on a nuisance parameter that measures the proximity of α_2 to zero, see Gospodinov (2010) for the case $a = 1$. Thus, conventional confidence bands on SVAR coefficients and IRFs do not have correct asymptotic coverage. In this section, which contains the main contribution of the paper, we introduce a method that does.

3.1 Anderson-Rubin test with filtered instruments

Our approach to solving the problems of weak identification and near non-stationarity consists of two components: (i) a weak-identification robust method – the Anderson and Rubin (1949) (henceforth AR) test, since the model is typically just-identified and (ii) filtered instruments – the so-called IVX approach of Magdalinos and Phillips (2009), to deal with near unit roots.

We start by looking at the special case of testing the hypothesis

$$H_0 : b_{12} = b_{12}^0 \text{ against } H_1 : b_{12} \neq b_{12}^0. \tag{7}$$

This hypothesis is special because it turns out that there exists a test that is both robust to weak identification/near unit roots, and asymptotically efficient under strong identification.

Because of the structure of the problem, the hypothesis (7) can be tested using just the first equation of the model (2). Given some instruments $Z_{1t} = (X'_{1t}, z'_t)'$, the AR statistic, $AR(b^0_{12})$, is the Wald statistic for testing $H^*_0 : \delta_z = 0$ in the auxiliary regression:

$$\Delta Y_{1t} - b^{0\prime}_{12}\Delta Y_{2t} = \delta'_1 X_{1t} + \delta'_z z_t + \varepsilon^0_{1t}. \tag{8}$$

When $n = 2$, i.e., when b_{12} is a scalar, this AR statistic can be written analytically as

$$AR(b_{12}) = \frac{(\Delta Y_1 - \Delta Y_2 b_{12})' P_{M_{X_1} z} (\Delta Y_1 - \Delta Y_2 b_{12})}{(\Delta Y_1 - \Delta Y_2 b_{12})' M_{Z_1} (\Delta Y_1 - \Delta Y_2 b_{12}) / (T - \mathrm{col}(Z_1))}, \tag{9}$$

where $P_{\cdot}$ denotes the projection matrix, $M_{\cdot} = I - P_{\cdot}$. $Z_1 = (X_1, z)$, and we follow standard notation that for any column vector X_t, X denotes the matrix of T stacked rows X'_t, $t = 1, \ldots, T$.

If we set $z_t = Y_{2,t-1}$, the AR statistic corresponds to the likelihood ratio test for (7). Under stationarity/strong identification ($\alpha_2 < 0$ and fixed), $AR(b_{12})$ is asymptotically distributed as χ^2 under H_0. Moreover, the likelihood ratio test is asymptotically efficient under stationarity/strong identification. However, when α_2 is local to zero, the χ^2 asymptotic approximation breaks down, and the asymptotic distribution, if it exists, depends on the proximity of $T\alpha_2$ to zero. So, $AR(b_{12})$ is not asymptotically pivotal, and tests based on χ^2 critical values will not control asymptotic size. This is straightforward to see using local-to-unity asymptotics as in Gospodinov (2010).

Our solution to the above problem is to use an instrument that relates to $Y_{2,t-1}$ but is constructed in such a way that it is less persistent than $Y_{2,t-1}$ whenever the latter has a near unit root. This is an application of the IVX method of Magdalinos and Phillips (2009) to this problem.

Magdalinos and Phillips (2009) obtained nuisance-parameter-free asymptotic distribution theory for Wald tests in situations where the order of integration of the regressors is unknown**,** such as predictive regressions or cointegrating regressions when the right hand side variables are nearly integrated. They did so by introducing an instrument which is filtered from the original data in such a way that it is at most moderately integrated, and correlates sufficiently with the variable it is instrumenting.

In the SVAR model, the filtered instrument is given by

$$z_t = \sum_{j=1}^{t-1} \rho_{Tz}^{t-j} \Delta Y_{2,j}, \quad \rho_{Tz} = 1 + \frac{c_z}{T^b}, \;\; b \in (1/2, 1), \;\; c_z < 0. \tag{10}$$

In the empirical implementation, we follow Magdalinos and Phillips (2009) and Kostakis et al. (2015) in setting $c_z = -1$ and $b = 0.95$.

To obtain asymptotic results, we make the following assumption on ε_t, where $\|\cdot\|$ denotes the spectral norm.

Assumption A. $(\varepsilon_t)_{t\in\mathbb{Z}}$ is a sequence of identically and independently distributed random vectors with $E(\varepsilon_t|Y_{t-1}, Y_{t-2}, \ldots) = 0$, $E(\varepsilon_t\varepsilon_t'|Y_{t-1}, Y_{t-2}, \ldots) = \Sigma_\varepsilon$ and diagonal with $\Sigma_\varepsilon > 0$, and the moment condition $E\|\varepsilon_t\|^4 < \infty$.

This assumption is similar to the one used in Magdalinos and Phillips (2009), except for the addition of conditional homoskedasticity, which is typically used in the literature (e.g., the results in Galí (1999) assume conditional homoskedasticity). Heteroskedasticity robust versions of the proposed tests can be obtained using GMM, see the Appendix.

Our proposed AR test is based on the following result.

Theorem 1. *Consider the model (2) and (3), where X_t consists of lags of ΔY_t, Y_{2t} is a scalar, ε_t satisfies Assumption A and either $T\alpha_2 \to -\infty$ or $T\alpha_2 \to C \leq 0$. Let $AR(b_{12})$ be as in (9) with instrument z_t defined by (10). Then under $H_0 : b_{12} = b_{12}^0$, $AR(b_{12}^0) \xrightarrow{d} \chi_1^2$.*

Remarks 1. The asymptotic size of the η-level AR test that rejects H_0 when $AR(b_{12}^0)$ exceeds the $1-\eta$ quantile of χ_1^2 is equal to η. This can be shown using arguments analogous to those used in the proof of Andrews et al. (2011) (Corollary 2.1 and Lemma 4.1), see the Supplement for further details.

2. The case $T\alpha_2 \to -\infty$ corresponds to (near) stationarity and strong identification. In this case, the statistic AR in (9) is asymptotically equivalent to the AR statistic AR^* that is obtained by replacing the filtered instrument z_t with $Y_{2,t-1}$. Because the model is just-identified, AR^* is the standard LR statistic which is asymptotically efficient under stationary and strong-instrument asymptotics. It is also asymptotically equivalent to the standard Wald test of H_0. Thus, the use of the filtered instrument

entails *no loss of power* in the case of strong identification, and so the AR test with filtered instruments weakly dominates the Wald and standard LR tests.

3. The results of the theorem, as well as the above two remarks, also apply in a model with more endogenous variables, Y_{3t}, that are subject to long-run restrictions, under the assumption that their coefficients, b_{13}, can be estimated consistently using $Y_{3,t-1}$ as instruments, and the resulting estimator $\hat{b}_{13}$ is asymptotically Gaussian. A sufficient condition for this is that Y_{3t} is stationary.

4. Theorem 1 can be extended to cover the case when Y_{2t} is a vector along the lines of Magdalinos and Phillips (2009) (Theorem 3.8), or Kostakis et al. (2015) (Theorem 1), under the assumption that C is a diagonal matrix. In that case, $AR\left(b_{12}^0\right) \xrightarrow{d} \chi^2_{\dim b_{12}}$.

3.2 Tests of general hypotheses

Testing general hypotheses (4) is complicated by the fact that $r\left(\theta\right)$ contains the potentially weakly identified parameter b_{12}. Let ψ denote the rest of the unknown parameters in θ other than b_{12}. Note that when b_{12} is known, the parameters ψ are identified as regression coefficients and variances. So, inference on smooth functions of ψ, given b_{12}, would be straightforward, except for the complication that arises when there is a near unit root in Y_{2t}. We address this issue using IVX in equation (3) with instrument z_t given by (10) for $Y_{2,t-1}$.

General hypotheses (4) can be tested using Bonferroni or projection methods for valid inference. The Bonferroni method is as follows: (i) obtain a $(1-\eta_1)$-level confidence set for b_{12}, $\mathcal{C}_{b_{12},\eta_1}$, by inverting the AR test introduced in the previous subsection; (ii) for each value $b_{12}^0 \in \mathcal{C}_{b_{12},\eta_1}$, perform an η_2-level IVX Wald test of $r\left(b_{12}^0, \psi\right) = 0$; (iii) reject $H_0 : r\left(\theta\right) = 0$ if all tests in (ii) reject. By the Bonferoni inequality, this test has level at most $\eta_1 + \eta_2$. In theory, this can be refined along the lines of McCloskey (2012), but this may be computationally impractical in realistic settings, due to the large number of parameters.

The projection method is as follows: perform a test of the joint null hypothesis $H_0^* : r\left(\theta\right) = 0, b_{12} = b_{12}^0$, and project out b_{12}, i.e., reject $H_0 : r\left(\theta\right) = 0$ if there is no value of b_{12}^0 for which H_0^* is accepted. This approach requires a test of the joint hypothesis H_0^*. Our proposed test for H_0^* is based on a novel idea that combines the $AR\left(b_{12}\right)$ statistic developed above with the Wald statistic for testing the restrictions on the remaining parameters ψ (this idea applies more generally, see Section B.2 and Theorem 3 in the Appendix). We call the resulting test ARW, and derive its asymptotic

properties under the null in Theorem 2 below.

We now turn to the derivation of the ARW test. Let $\hat{\psi}(b_{12})$ be the restricted GMM estimator of ψ given b_{12} and let $\hat{V}_{\hat{\psi}}(b_{12})$ denote an estimate of the asymptotic variance matrix of $\hat{\psi}(b_{12})$. Provided $R(\theta) = \partial r(\theta)/\partial\psi'$ exists and is of full rank q, define

$$W(b_{12}) = r\left(b_{12}, \hat{\psi}(b_{12})\right)' \hat{V}_{\hat{r}}(b_{12})^{-1} r\left(b_{12}, \hat{\psi}(b_{12})\right), \tag{11}$$
$$\text{where} \quad \hat{V}_{\hat{r}}(b_{12}) = R\left(b_{12}, \hat{\psi}(b_{12})\right) \hat{V}_{\hat{\psi}}(b_{12}) R\left(b_{12}, \hat{\psi}(b_{12})\right)',$$

and consider the combined statistic

$$ARW\left(b_{12}^0\right) = AR\left(b_{12}^0\right) + W\left(b_{12}^0\right). \tag{12}$$

The asymptotic distribution of $ARW(b_{12}^0)$ under the null H_0^* is given by the following result.

Theorem 2. *Under the conditions of Theorem 1, if the null hypothesis $H_0^* : r(\theta) = 0, b_{12} = b_{12}^0$ holds, then:*

$$W\left(b_{12}^0\right) \xrightarrow{d} \chi_q^2,$$

$W(b_{12}^0)$ is asymptotically independent of $AR(b_{12}^0)$, and

$$ARW\left(b_{12}^0\right) = AR\left(b_{12}^0\right) + W\left(b_{12}^0\right) \xrightarrow{d} \chi_{1+q}^2.$$

Remarks 1. The ARW test rejects $H_0^* : r(\theta) = 0, b_{12} = b_{12}^0$ at the η level of significance if $ARW(b_{12}^0)$ is greater than c_η where c_η is the $1-\eta$ quantile of χ_{1+q}^2. A projection test of $H_0 : r(\theta) = 0$ rejects H_0 when $\min_{b_{12}} ARW(b_{12}) > c_\eta$.

2. The asymptotic size of a $(1-\eta)$-level confidence set obtained by inverting an η-level ARW test, defined as the minimum coverage probability of the confidence set, is equal to $1-\eta$ uniformly in α_2. This result is analogous to Remark 1 to Theorem 1, see the Supplement for details.

3. Remarks 3 and 4 to Theorem 1 also apply to Theorem 2.

4. For a Bonferroni test of $H_0 : r(\theta) = 0$, we use the Wald test that rejects when $W(b_{12}^0)$ exceeds the $1-\eta_2$ quantile of χ_q^2 in the second step.

5. Confidence intervals for any scalar function of the parameters $g(b_{12}, \psi)$ that is smooth in ψ, such as an impulse response, can be obtained easily and quickly by numerical optimization methods. An algorithm for this is given in the Supplementary

Appendix.

Example: bivariate SVAR(1) Suppose we are interested in testing $H_0 : \partial Y_{2t}/\partial \varepsilon_{1t} = d_{21} = d_{21}^0$ against $H_1 : d_{21} \neq d_{21}^0$. This can be expressed as the linear restriction $r(b_{12}, \psi) = d_{21} - d_{21}^0$. Our proposed η-level ARW test rejects H_0 if $\min_{b_{12}} (AR(b_{12}) + W(b_{12}))$ is greater than c_η, the $1-\eta$ quantile of χ^2_2. Let $\hat{d}_{21}(b_{12})$ and $\hat{\sigma}_{\hat{d}_{21}}(b_{12})$ denote the restricted point estimate of d_{21} and its standard error, respectively. An ARW projection $(1-\eta)$-level confidence interval for d_{21} is given by

$$\left\{ \min_{b_{12}} \left[\hat{d}_{21}(b_{12}) - \hat{\sigma}_{\hat{d}_{21}}(b_{12}) \sqrt{c_\eta - AR(b_{12})} \right], \max_{b_{12}} \left[\hat{d}_{21}(b_{12}) + \hat{\sigma}_{\hat{d}_{21}}(b_{12}) \sqrt{c_\eta - AR(b_{12})} \right] \right\}.$$

A Bonferroni confidence interval based on an η_1-level AR test with critical value c_1 and η_2-level W test with critical value c_2, where c_i is the $1-\eta_i$ quantile of χ^2_1, is given by

$$\left\{ \min_{b_{12}: AR(b_{12}) \leq c_1} \left[\hat{d}_{21}(b_{12}) - \hat{\sigma}_{\hat{d}_{21}}(b_{12}) \sqrt{c_2} \right], \max_{b_{12}: AR(b_{12}) \leq c_1} \left[\hat{d}_{21}(b_{12}) + \hat{\sigma}_{\hat{d}_{21}}(b_{12}) \sqrt{c_2} \right] \right\}.$$

3.3 Deterministic terms

Theorems 1 and 2 apply when model (2)-(3) does not include any deterministic terms in X_{1t} and X_{2t}, but it can be shown using the same arguments as in Kostakis et al. (2015) Theorem A that they continue to hold if an intercept is included in X_{1t}, X_{2t}. However, in that case the asymptotic approximations may deteriorate in finite samples, as was found by Kostakis et al. (2015) for predictive regression. To address this possibility, we derive a finite sample correction proposed by Kostakis et al. (2015), adapting it to the ARW statistic as follows. The finite sample correction in Kostakis et al. (2015), applied to the AR in (9) consists in modifying $P_{M_{X_1} z}$ in the numerator. When the model contains an intercept, the finite sample correction involves replacing the term $P_{M_{X_1} z} = M_{X_1} z \left(z' M_{X_1} z \right)^{-1} z' M_{X_1}$ with

$$\widetilde{P}_{M_{X_1} z} = M_{X_1} z \left(z' M_{\widetilde{X}_1} z - T \left(1 - \widehat{\rho}_{\varepsilon_1, u_2} \right) \overline{z}' \overline{z} \right)^{-1} z' M_{X_1}$$

where $\widetilde{X}_1$ denotes the elements in X_1 excluding the intercept, $\widehat{\rho}_{\varepsilon_1, u_2}$ is the estimated long-run correlation between ε_{1t} and u_{2t} in equations (2)-(3). The correction of the Wald statistic $W(b_{12})$ is analogous. It depends on the specific form of H_0^* but only affects the variance related to the estimator of α_2 in $\hat{V}_{\hat{\psi}}(b_{12})$. We provide an expression

for it in the Supplementary Appendix. In the empirical applications we consider in this paper, $\widehat{\rho}_{\varepsilon_1,u_2}$ is low enough so the finite sample correction does not make material difference to the results.

In some applications, Y_{2t} denotes the deviation of some observed variable (e.g., log hours, or log real GDP) from a linear deterministic trend where the observed data Y_{2t}^{obs} is given by $Y_{2t}^{obs} = Y_{2t} + \tau_x + \gamma_x t$. We then replace Y_{2t} with $\hat{Y}_{2t} = Y_{2t}^{obs} - \hat{\tau}_x - \hat{\gamma}_x t$ in the computation of the IVX instrument z_t. Whether or not Y_{2t} is stationary affects inference on γ_x. If $\hat{\gamma}_x$ is computed using the full sample, then $\hat{Y}_{2t}$ is a function of future values and this may affect the validity of the exclusion restrictions used in the estimation.

To avoid this issue, we follow Phillips et al. (2004) and use a recursive detrending formula to ensure that $\hat{Y}_{2t}$ is not computed using future values:

$$\hat{Y}_{2t} = Y_{2t}^{obs} - \hat{\tau}_x - \hat{\gamma}_x t = Y_{2t}^{obs} + \frac{2}{t}\sum_{j=1}^{t} Y_{2j}^{obs} - \frac{6}{t\left(t+1\right)}\sum_{j=1}^{t} j Y_{2j}^{obs}$$

This formula preserves the martingale difference sequences which are needed in the asymptotic theory, so moment conditions hold under H_0. Hence, the asymptotic results presented above continue to hold.

4 Numerical Results

In this section we investigate the finite-sample properties of our proposed test and compare them with the existing non-robust alternative.

The data generating process is the bivariate SVAR(1) example introduced earlier, with $\alpha_2 = cT^{-1}$. In reduced-form, the model is:

$$\begin{aligned}
\Delta Y_{1t} &= \frac{c}{T} b_{12} Y_{2,t-1} + u_{1t}, \quad 1 \leq t \leq T \\
\Delta Y_{2t} &= \frac{c}{T} Y_{2,t-1} + u_{2t}
\end{aligned}$$

with

$$\begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} \sim NID\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \omega_1^2 & \rho\omega_1 \\ \rho\omega_1 & 1 \end{pmatrix}\right)$$

and $Y_{10} = Y_{20} = 0$. We normalize $\omega_2 = 1$ because the statistics are invariant to scaling of the variance matrix. The AR statistic is also invariant to ω_1, so in simulations in-

volving only $AR(b_{12})$, we will also normalize $\omega_1 = 1$. The estimated model is SVAR(1), with and without deterministic terms.[3]

4.1 Size

We conduct two sets of simulation experiments to obtain the rejection frequency of tests of the following two null hypotheses: (i) $H_0 : b_{12} = 0$ against $H_1 : b_{12} \neq 0$, using the AR test with filtered instruments, and (ii) $H_0 : d_{21} = d_{21}^0$ against $H_1 : d_{21} \neq d_{21}^0$ using the ARW test, for $d_{21} \in [-1, 1]$.[4]

In case (i), we report rejection frequencies over a few different parameterizations. We consider the parameter sets $\rho \in \{0.20, 0.95\}$ and $c \in \{0, -1, -10, -30, -100\}$ and the sample size is set to $T = 200$. We compute the null rejection frequencies of our AR test with the filtered instrument z_t in (9) and the conventional t test with instrument $Y_{2,t-1}$ at the 5% and 10% levels of significance. The estimated model is SVAR(1) with an intercept, and the computation of the AR statistic uses the finite sample correction introduced in Section 3.3. The number of Monte Carlo replications is 20000.

The rejection frequencies are reported in Table 1. We notice that the rejection frequency of the t test can deviate sharply from its asymptotic level, with considerable overrejection in the cases $\rho = 0.95$ and c close to zero. In contrast, the rejection frequency of our proposed AR test is close to its asymptotic level in all cases. Similar results obtain for SVAR models with more lags as well as for models with deterministic terms (further results can be found in the Supplementary Appendix).

In case (ii), we conduct experiments for a very large number of parameter combinations over a 4-dimensional grid in d_{21}, ρ, ω_1 and c, where we exploit an invariance property of the ARW statistic that enables us to normalize $\omega_2 = 1$ and fix b_{12} as a function of the other parameters, see the supplementary appendix for details. Figure 1 reports the maximal rejection frequency of the test at three different levels of significance (10%, 5% and 1%) over ρ, ω_1 and c for each value of d_{21} under the null, denoted $\bar{d}_{21}$ in the figure. The estimated model coincides with the DGP, i.e., a SVAR(1) without deterministics, and the number of Monte Carlo replications is 20000.

We notice that the size of the projection ARW test is well below the nominal level across all values of d_{21}. In the supplement we verify that the same result holds also

[3] Results for higher-order SVARs are very similar and can be found in the Supplementary Appendix.

[4] It can be shown that d_{21} is bounded between $\pm\omega_2$, the reduced-form error standard deviation in the second equation, which is normalized to 1 here.

	At 5%				At 10%			
	$\rho = 0.20$		0.95		0.20		0.95	
	AR	t	AR	t	AR	t	AR	t
$c = 0$	0.052	0.005	0.071	0.774	0.103	0.025	0.133	0.807
-1	0.052	0.007	0.064	0.680	0.100	0.029	0.125	0.717
-10	0.050	0.019	0.047	0.257	0.102	0.053	0.092	0.307
-30	0.051	0.034	0.044	0.135	0.100	0.081	0.089	0.181
-100	0.053	0.050	0.045	0.069	0.102	0.100	0.093	0.115

Table 1: Null rejection frequencies of AR (with filtered instruments) and conventional t tests of the hypothesis $H_0 : b_{12} = 0$ in a bivariate SVAR(1) with long-run restrictions. ρ is the correlation between the reduced-form VAR errors. The sample size is 200. Number of MC replications: 20000.

in a large sample with $T = 2000$. This indicates that there is some projection bias that could in principle be reduced by using lower critical values. However, it is not possible to reduce the critical value all the way to χ^2_1, as would be warranted under strong identification, because the resulting test would be oversized (see the results in the Supplementary Appendix). An ARW test with χ^2_1 critical values will only yield correct asymptotic size when $\alpha_2 < \kappa$ for some fixed $\kappa < 0$. This is because in that case, a test that rejects when $\min_{b_{12}} ARW\,(b_{12})$ is greater than the $1 - \eta$ quantile of the χ^2_1 distribution is asymptotically equivalent to a standard Wald test of the restriction on the parameter d_{21}. However, it does not seem possible to use the lower critical values under weak identification, so the use of the projection critical values based on χ^2_2 entail some loss of power for robustness in the case of strong identification.

It is possible to design a data-based identification category selection rule along the lines of Andrews and Cheng (2012), based on comparing the proximity of $\hat{\alpha}_2$ to some cutoff that diverges with T, but this will come at the cost of introducing additional tuning parameters, and so may be unappealing in applied work.

4.2 Power

We compute (large-sample) power of tests of AR, t, projection ARW and Bonferroni tests in the working SVAR(1) example. We set $T = 2000$, use 10000 Monte Carlo

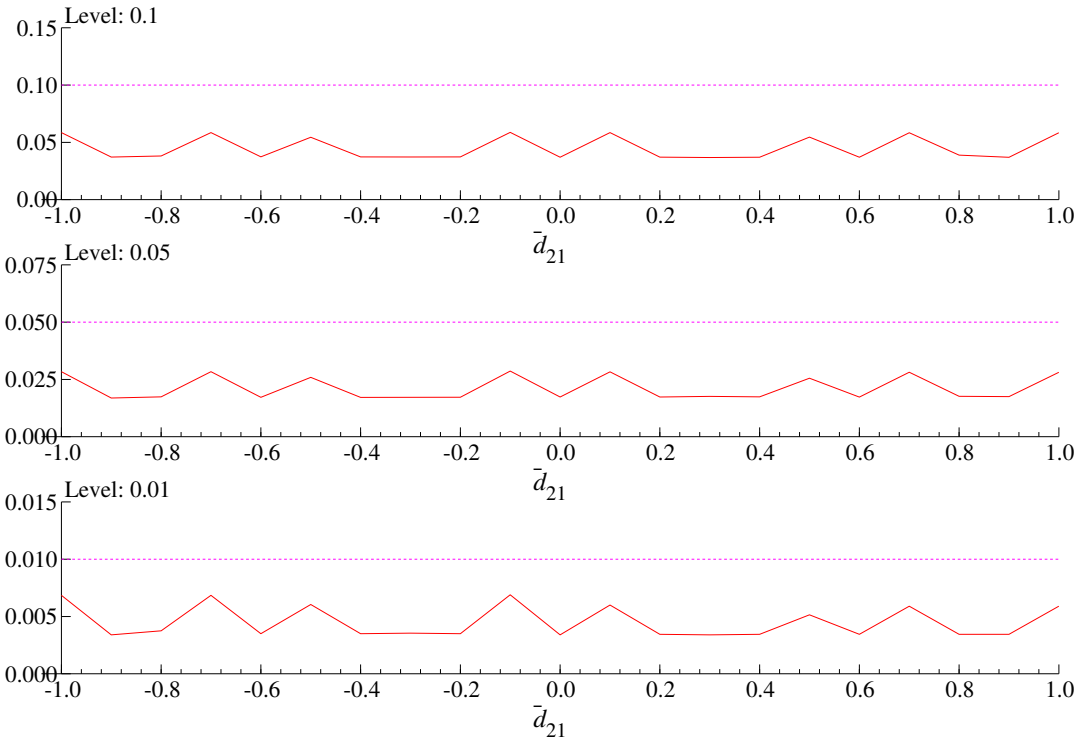

Figure 1: Size of the projection ARW test of the hypothesis $H_0 : d_{21} = \bar{d}_{21}$, in a SVAR(1) model with T=200 at three different significance levels. The number of Monte Carlo replications is 20000.

replications.

4.2.1 Power of AR test

We compare the power of AR and t tests of $H_0 : b_{12} = 0$ against $H_1 : b_{12} \neq 0$ at the 10% level of significance. The remaining parameters are $\rho \in \{0.2, 0.95\}$, $\omega_1 = 1$, and $c = \{-10, -100, -500\}$. In this model, the strength of identification is driven by c. To relate the results to well-known cases of weak, moderate and strong identification in linear IV, we compute an approximate measure of the strength of instruments known as the concentration parameter (denoted λ) in linear IV.[5] The chosen values of c correspond to approximate values of λ of 1.3, 13, and 72 respectively, i.e., weak, moderate and strong identification. The range of b_{12} under H_1 is $\lambda^{-1/2}(-3:3)$.

Figure 2 reports the resulting power curves in each case. The figure shows that the AR test has good power even for c close to zero. This is not the case for the t test, which is both size distorted and even biased in some cases. Moreover, when identification is strong ($c = -500$), the power of the AR test is very similar to that of the t test, which is asymptotically efficient in this case. Since the DGP in this case is approximately stationary, this is a consequence of the fact that the AR and t tests are asymptotically equivalent in the case of stationarity, see comment 2 to Theorem 1.

4.2.2 Power of ARW test

We compare the power of projection and Bonferroni tests of $H_0 : d_{21} = 0$ against $H_1 : d_{21} \neq 0$ at significance level 10%. We consider three different combinations of significance levels $\eta_1 + \eta_2$ for the Bonferroni tests, where η_1 denotes the significance level for the first-step AR confidence set for b_{12} and $\eta_2 = 10\% - \eta_1$ denotes the level of the Wald test given b_{12} in the second step. We set $b_{12} = 0$, $\omega_1 = 1$ and $c = -100$ (moderately strong identification). We note that with these parameter values $\rho = d_{21}$.

The results are reported in Figure 3. The power of the projection test is close to that of the Bonferroni tests that put sufficiently high weight on the first-step AR confidence set for b_{12}, i.e., $\eta_1 \geq \eta/2$. This suggests that there is little to choose from between Bonferroni and projection in this case on the basis of power. Since projection turns out to be slightly faster to compute (it requires solving an unconstrained optimization

[5] In linear IV with fixed instruments, the concentration parameter is equal to $k[E(F) - 1]$, where F is the infeasible version of the first-stage F statistic for excluding the instrument, computed when the variance of the reduced form error variance is known, see Stock et al. (2002). The present context does not fit into that canonical IV framework, so we use a large sample approximation of λ.

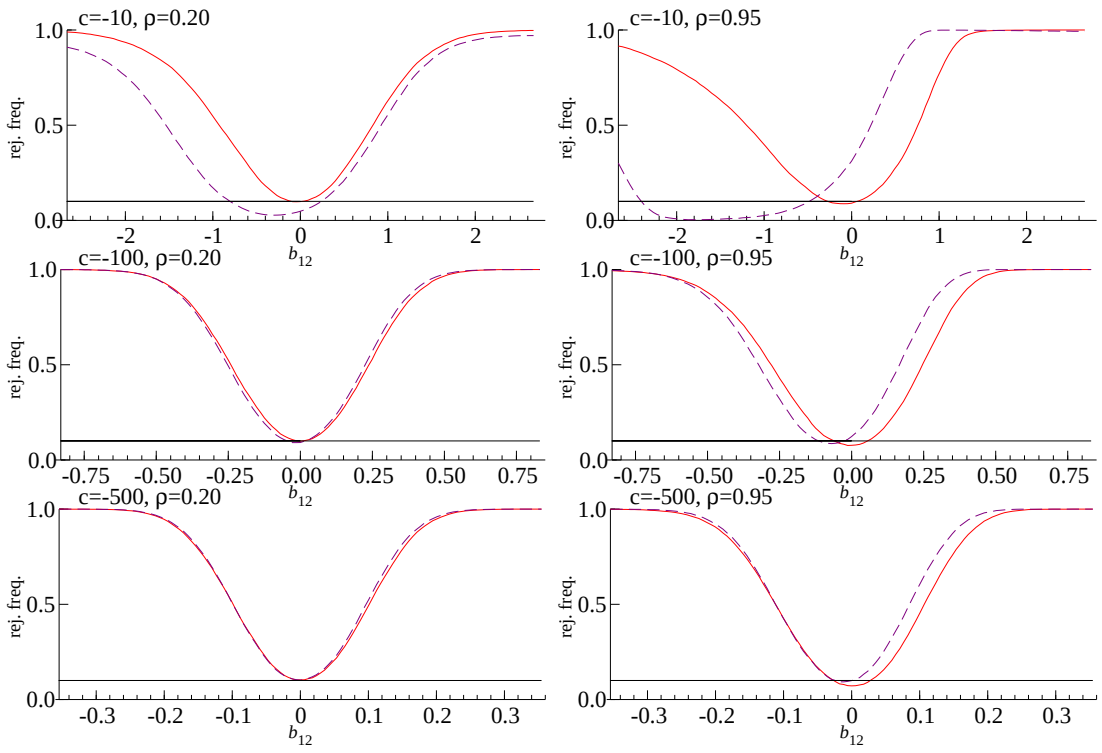

Figure 2: Large-sample power of AR with filtered instrument (solid line) and t (dashed line) tests of the hypothesis $H_0 : b_{12} = 0$ against $H_1 : b_{12} \neq 0$ in the SVAR(1) model with long run restrictions. $T = 2000$, 10000 MC replications, ρ is correlation of reduced-form errors.

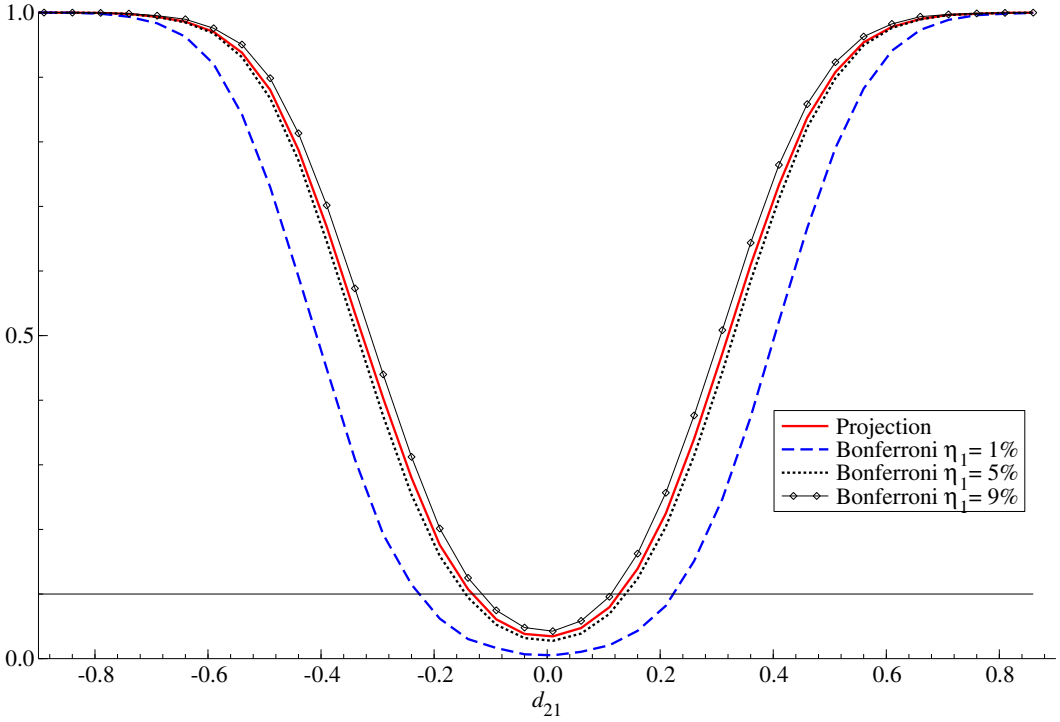

Figure 3: Power of Projection ARW (red solid) and three different Bonferroni AR/W tests of the hypothesis $H_0 : d_{21} = 0$ against $H_1 : d_{21} \neq 0$ in a bivariate SVAR(1) at the 10% level of significance. η_1 denotes the level of the (first-step) AR test in the Bonferroni procedure and $\eta_2 = 10\% - \eta_1$ is the level of the second step Wald test. $T = 2000$ and the number of Monte Carlo replications is 10000.

problem, as opposed to a constrained optimization with an inequality constraint for the Bonferroni method), we are using the projection method in our empirical applications below.

5 Empirical Results

5.1 Blanchard and Quah (1989)

We first revisit the application of Blanchard and Quah (1989) (BQ), where Y_{1t} is log real GNP, and Y_{2t} is the unemployment rate in deviation from a linear trend. We use the original BQ dataset, which is quarterly and covers the period 1948q1 to 1987q4. More details about the data and transformations are given in the Supplementary Appendix.

The specification in BQ is a SVAR(9). Figure 4 reports the estimated IRFs to-

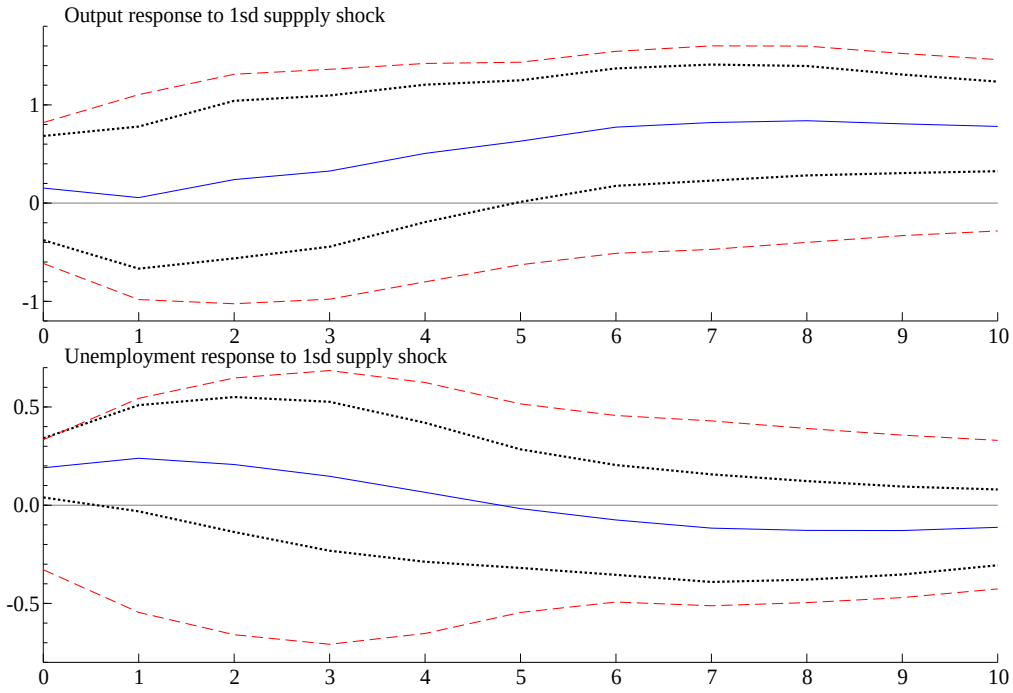

Figure 4: IRFs to supply shock from a bivariate SVAR in real output growth and the unemployment rate by Blanchard and Quah (1989). The solid line is the ML estimator. The dotted lines are 90% Wald confidence intervals, and the dashed lines are the 90% projection ARW confidence intervals. The data is from Blanchard and Quah (1989) over the period 1948q1 to 1987q4.

gether with robust 90% confidence bands based on our proposed ARW method and the corresponding non-robust Wald confidence bands. We see that the robust confidence bands are so large that the original conclusion of BQ is not borne out. In other words, long-run restrictions produce very weak identification in this application using the original data. This corroborates the criticism of Pagan and Robertson (1998).

The results in Figure 4 used full-sample detrending, which is problematic when the data is persistent, as we saw in our numerical analysis in the previous section. This can be addressed using recursive detrending. Results based on recursive detrending of the unemployment rate are given in Figure 5. We see that the results are very sensitive to the detrending method. With recursive detrending, which is more reliable than full-sample detrending, the effect of the supply shock on output becomes clearly positive but the effect on unemployment remains ambivalent.

We should emphasize that weak identification is an empirical matter, so identifi-

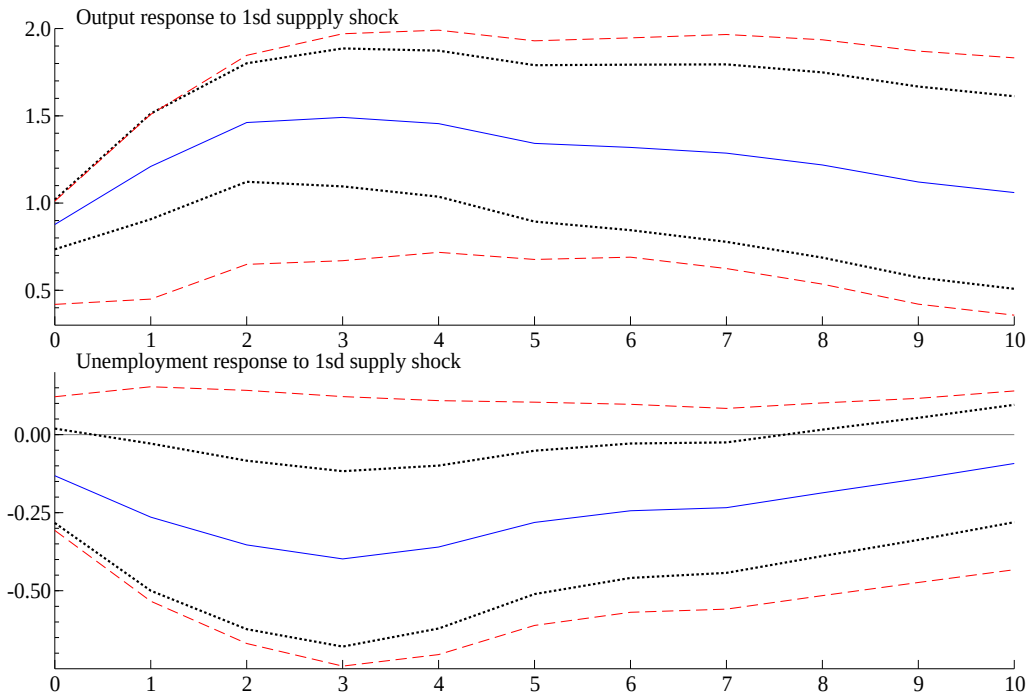

Figure 5: Estimates and confidence bands of the IRFs in Blanchard and Quah (1989) with recursive detrending, using their original data. The solid line is the ML estimator. The dotted lines are 90% Wald confidence intervals, and the dashed lines are the 90% projection ARW confidence intervals.

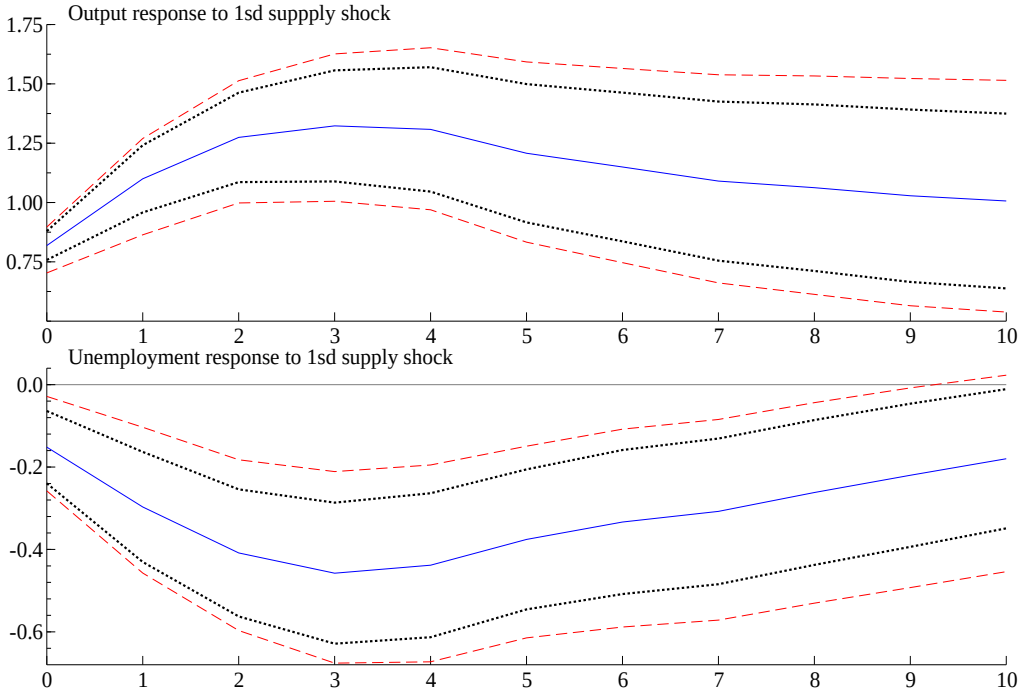

Figure 6: Estimates and confidence bands of the IRFs with extended Blanchard and Quah (1989) data and recursive detrending. The solid line is the ML estimator. The dotted lines are 90% Wald confidence intervals, and the dashed lines are the 90% projection ARW confidence intervals.

cation of the model may become stronger over a different sample. Figure 6 reports estimates of the IRFs based on the same specification as in Figure 5, but estimated over an extended sample that runs up to 2014q4. We notice that the point estimates are very similar, but error bands become significantly tigher, and identification appears to be strong.

5.2 The hours debate

Next, we turn to the debate on the short-run effect of a technology shock on hours initiated by the seminal papers of Galí (1999) and Christiano et al. (2003) (CEV). The analysis in those papers is based on a SVAR where Y_{1t} denotes log productivity and Y_{2t} denotes log hours.

The original paper by Galí (1999) estimated a negative short-run effect of a technology shock on hours, where Y_{2t} was the growth rate in hours, i.e., total log hours in

first difference. Galí (1999) argued that this finding was inconsistent with real business cycle theory, but could be explained by sticky-price models. CEV criticized Galí's data and specification. Specifically, they argued for using log hours per capita as opposed to total hours and that Y_{2t} should be hours in levels as opposed to growth rates because the level specification encompasses the difference one. Reestimating using per capita hours in levels, they found a positive short-run effect of technology shock on hours, contradicting Galí's conclusions.

There has been a large subsequent literature attempting to explain the above conflicting findings, see, for example, Chaudourne et al. (2014), Dupaigne et al. (2007), Fève and Guay (2009, 2010), Francis and Ramey (2005, 2009), Gospodinov et al. (2011), Pesavento and Rossi (2005), and Ramey (2016) (Section 5) for a recent review. Many of those papers emphasized possible misspecification due to omission of relevant variables and shocks from the SVAR, which could be addressed by adding more variables to the SVAR. Others emphasized the sensitivity of the estimates to assumptions about the number of permanent shocks and the effect of near unit roots. Our analysis below complements the literature by providing confidence bands on the impulse responses in question that are fully robust to weak identification. We focus our empirical analysis only on the baseline specifications in the two seminal papers in the literature, Galí (1999) and CEV, but we note that our methods are applicable to the more general SVAR specifications used in the literature.

We use the same data as Galí and CEV,[6] so the point estimates and conventional confidence bands reported below are the same as in those papers. Galí uses total hours linearly detrended over the sample 1948q1 to 1994q4. CEV use per capita hours and their sample is 1948q1 to 2001q4.

Figure 7 presents point estimates and 90% confidence bands from the difference specification in Galí (1999) with total hours. We see that the projection ARW confidence bands are not much wider than the non-robust ones reported by Galí (1999), indicating that this specification does not suffer from weak identification. This conclusion is robust to using the growth in per capita hours instead of total hours.[7]

However, the results on the difference specification are subject to the valid critique by CEV regarding possible misspecification if hours do not have a unit root.[8] Figure 8

[6] A plot of the data can be found in the Supplementary Appendix.

[7] The estimates of the difference specification with CEV data on per capita hours are reported in the Supplementary Appendix.

[8] This is spelt out in Section A.2 of the Appendix.

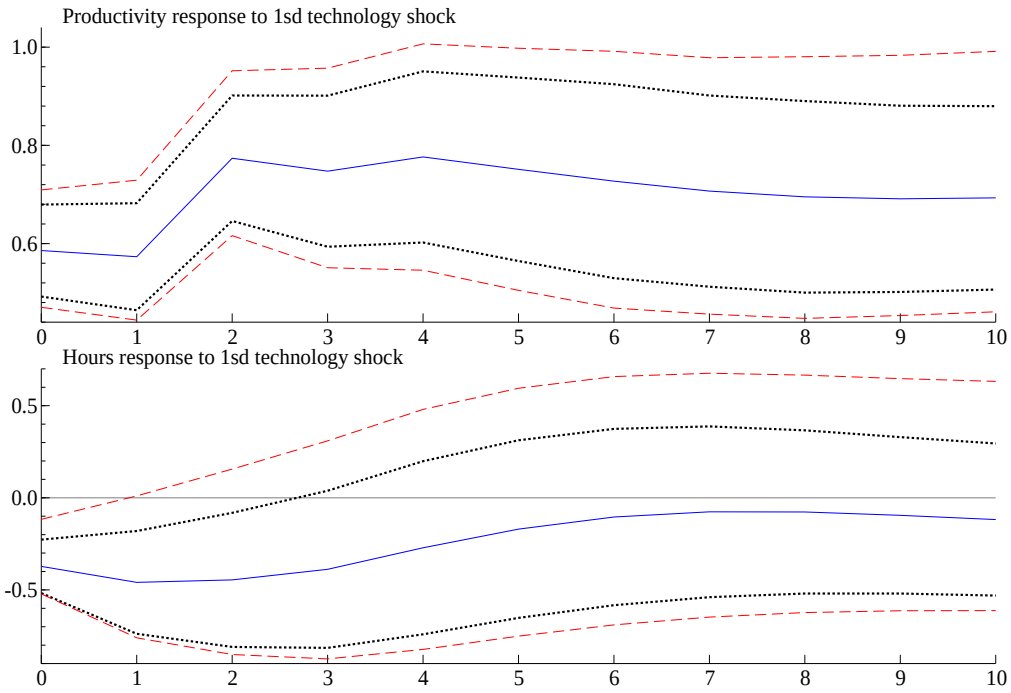

Figure 7: IRFs to technology shock for the difference specification of Galí (1999). The model is a bivariate SVAR in the first differences of log productivity and log total hours. The solid line is the ML estimator. The dotted lines are 90% Wald confidence intervals, and the dashed lines are the 90% projection ARW confidence intervals. The data is from Galí (1999) over the period 1948q1 to 1994q4.

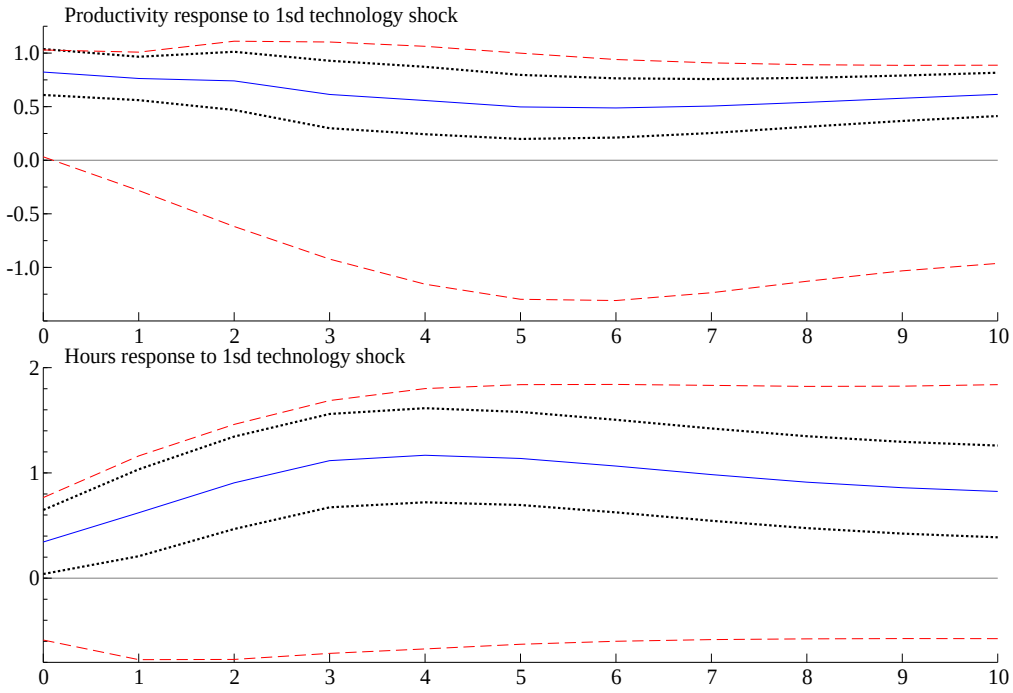

Figure 8: IRFs to technology shock for the level specification of Christiano *et al.* (2003). The model is a bivariate SVAR in the growth of productivity and the level of log per capita hours. The solid line is the ML estimator. The dotted lines are 90% Wald confidence intervals, and the dashed lines are the 90% projection ARW confidence intervals. The data is from Christiano *et al.* (2003) over the period 1948q1 to 2001q4.

presents the CEV estimates and confidence intervals based on the levels specification, together with the robust projection ARW confidence bands. Unlike the Wald bands, the robust confidence bands are so wide that the response of hours to a technology shock is no longer significant. The information in the long-run restriction is so small that the data is consistent both with a positive as well as a negative response of hours to a technology shock. Therefore, the original conclusions of CEV are not robust to weak identification.

In the Supplementary Appendix we report further results that indicate that the above conclusion on the weak identification of the level specification is robust to detrending of hours and to extensions of the estimation sample. All in all, we see that long-run restrictions are not very informative in this application, unless one is willing to impose the arguably strong assumption that hours have a unit root.

6 Conclusions

We proposed a method of inference on the parameters of SVARs identified using long-run restrictions that is robust to both weak instruments and near unit roots in the data. The method uses instruments obtained by filtering the potentially non-stationary variables to make them near stationary. We propose to test hypotheses on the parameters that are potentially weakly identified using the Anderson-Rubin test with filtered instruments. Tests of general parametric restrictions, and confidence intervals for differentiable functions of the parameters, such as IRFs or forecast error variance decompositions, are obtained using a combined AR and Wald test. The robust test and associated confidence bands are easy to compute, and offer informative and reliable inference in two high-profile applications.

A Level versus difference specification

A.1 Representation in terms of (2)-(3)

Fukac and Pagan (2006) show that the long-run restrictions depend on the number of permanent shocks in the system. We assume throughout that there are no I(2) trends. It is typically assumed (e.g., by Galí (1999)) that long-run identification requires at least one permanent shock, so the cointegrating rank can be 0 (two permanent shocks) or 1 (one permanent shock). Let $\tilde{Y}_t$ denote the original data in levels. We will show how both the level and the difference specifications can both be written in the form (2)-(3) by defining Y_t appropriately. We drop the deterministic terms and focus on the bivariate case of the general model (1), which suffices for this discussion.

Case of one permanent shock This is a cointegrated VAR, or vector error correction model (VECM), which can be written as

$$\Gamma(L)\,\Delta\tilde{Y}_t = \underset{2\times 1}{\alpha}\underbrace{\beta'}_{1\times 2}\tilde{Y}_{t-1} + B_0^{-1}\varepsilon_t, \tag{13}$$

with $\Gamma(L) = \sum_{j=0}^{m-1} \Gamma_j L^j$, $\Gamma_0 = I$, $\Gamma_j = -B_0^{-1} \sum_{i=j+1}^{m} B_i$, and $\alpha\beta' = -B_0^{-1} B(1)$. Its Granger representation is:

$$\tilde{Y}_t = C \sum_{s=1}^{t} \varepsilon_s + \tilde{C}(L) \varepsilon_t, \quad C = \beta_\perp \left(\alpha_\perp' \Gamma(1) \beta_\perp\right)^{-1} \alpha_\perp' B_0^{-1},$$

where $\alpha_\perp' \alpha = 0$, $\alpha = \binom{\alpha_1}{\alpha_2}$, $\alpha_\perp = \binom{\alpha_2}{-\alpha_1}$ and similarly for β. The long-run restriction that only ε_{1t} has a permanent effect on $\tilde{Y}_{1t}$ can be written as a zero restriction on the top right element of the matrix C,

$$C = \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix} = \begin{pmatrix} * & 0 \\ * & * \end{pmatrix}.$$

(Note that since cointegration implies $rank(C) = 1$, $C_{22} = 0$ must hold too: only ε_{1t} drives the stochastic trend.) This implies that $\alpha_\perp' B_0^{-1} \binom{0}{1} = 0$, or if we define

$$B_0 = \begin{pmatrix} 1 & -b_{12} \\ -b_{21} & 1 \end{pmatrix},$$

$$b_{12} = \frac{\alpha_1}{\alpha_2}.$$

Alternatively, let $\Gamma(L) = \begin{pmatrix} \gamma_{11}(L) & -\gamma_{12}(L) \\ -\gamma_{21}(L) & \gamma_{22}(L) \end{pmatrix}$ and write the VECM as:

$$\gamma_{11}(L) \Delta\tilde{Y}_{1t} = \alpha_1 \beta' \tilde{Y}_{t-1} + \gamma_{12}(L) \Delta\tilde{Y}_{2t} + u_{1t}$$
$$\gamma_{22}(L) \Delta\tilde{Y}_{2t} = \alpha_2 \beta' \tilde{Y}_{t-1} + \gamma_{21}(L) \Delta\tilde{Y}_{1t} + u_{2t},$$

where $u_t = B_0^{-1} \varepsilon_t$ are the reduced form errors. Imposing the long-run restriction yields (Pagan and Pesaran, 2008):

$$\tilde{\gamma}_{11}(L) \Delta\tilde{Y}_{1t} = b_{12} \Delta\tilde{Y}_{2t} + \tilde{\gamma}_{12}(L) \Delta\tilde{Y}_{2t} + \varepsilon_{1t}, \tag{14}$$

where $\tilde{\gamma}_{11}(L) = \gamma_{11}(L) + b_{12}\gamma_{21}(L)$ and $\tilde{\gamma}_{12}(L) = \gamma_{12}(L) + b_{12}[\gamma_{22}(L) - 1]$. Observe that the error correction (ecm) term $\beta'\tilde{Y}_{t-1}$ is missing from (14), so we can use this to instrument for the endogenous regressor $\Delta\tilde{Y}_{2t}$. In our applications, $\beta = (0, 1)'$, so that $\beta'\tilde{Y}_t = \tilde{Y}_{2t}$. So, the model can be written in the form (2)-(3) with $Y_t = \tilde{Y}_t$.

Case of two permanent shocks In this case there is no cointegration, so the model is a VAR in first differences:

$$\Gamma(L)\,\Delta\tilde{Y}_t = B_0^{-1}\varepsilon_t.$$

The long-run restriction that permanent shocks to $\tilde{Y}_{2t}$ have no impact on $\tilde{Y}_{1t}$ is

$$C = \Gamma(1)^{-1} B_0^{-1} = \begin{pmatrix} * & 0 \\ * & * \end{pmatrix}.$$

(Note that in this case C_{22} does not need to be 0). The long-run restriction then implies:

$$b_{12} = -\frac{\gamma_{12}(1)}{\gamma_{22}(1)}.$$

As before, this can also be expressed as an exclusion restriction. First, from the Beveridge and Nelson (1981) (henceforth BN) decomposition we have

$$b_{12} + \tilde{\gamma}_{12}(L) = b_{12} + \tilde{\gamma}_{12}(1) + \tilde{\gamma}_{12}^{*}(L)\,\Delta. \tag{15}$$

Substituting in the SVAR, using the long-run restriction $b_{12} + \tilde{\gamma}_{12}(1) = 0$ we have

$$\tilde{\gamma}_{11}(L)\,\Delta\tilde{Y}_{1t} = \tilde{\gamma}_{12}^{*}(L)\,\Delta^2\tilde{Y}_{2t} + \varepsilon_{1t}, \tag{16}$$

Similarly, using the BN decomposition of $\gamma_{22}(L) = \gamma_{22}(1)\,L + \gamma_{22}^{*}(L)\,\Delta$, the equation for $\tilde{Y}_{2t}$ can be written as

$$\gamma_{22}^{*}(L)\,\Delta^2\tilde{Y}_{2t} = \gamma_{22}(1)\,\Delta\tilde{Y}_{2,t-1} + \gamma_{21}(L)\,\Delta\tilde{Y}_{1t} + u_{2t}.$$

Thus, we are using $\Delta\tilde{Y}_{2,t-1}$ as an instrument for the endogenous regressor $\Delta^2\tilde{Y}_{2t}$ in (16). This specification can be written in the form (2)-(3) with $Y_{1t} = \tilde{Y}_{1t}$ and $Y_{2t} = \Delta\tilde{Y}_{2t}$.

A.2 Misspecification of difference specification

Using (15) to substitute for $\tilde{\gamma}_{12}(L)$ in (14) yields

$$\tilde{\gamma}_{11}(L)\,\Delta\tilde{Y}_{1t} = \tilde{\gamma}_{12}^{*}(L)\,\Delta^2\tilde{Y}_{2t} + [b_{12} + \tilde{\gamma}_{12}(1)]\,\Delta\tilde{Y}_{2t} + \varepsilon_{1t}. \tag{17}$$

Similarly, the reduced form equation for the level specification imposes no extra restriction, and uses $\tilde{Y}_{2,t-1}$ as an instrument in (17). The difference specification imposes

$b_{12}+\tilde{\gamma}_{12}(1)=\alpha_2=0$, which enables us to use $\Delta Y_{2,t-1}$ as an instrument in (17). The difference specification will be misspecified if $b_{12}+\tilde{\gamma}_{12}(1)\neq 0$. In principle, this misspecification is detectable by a suitable diagnostic test. However, the power of such a test depends on the value of $\alpha_2\neq 0$. Only when α_2 is far from zero can we reject $\alpha_2=0$ with high probability. Otherwise, if we do not reject $\alpha_2=0$ and impose it incorrectly, the bias that will result depends on the true value of $b_{12}+\tilde{\gamma}_{12}(1)$ and can be arbitrarily large. This discussion corroborates formally CEV's critique.

B Proofs

The following Lemma is used in the proofs of the theorems. Parameter ω is a positive constant that relates to model parameters and the long run variance of the reduced form errors. $\mathcal{W}$ is a standard Brownian motion and $\mathcal{J}_c(s)=\int_0^s e^{c(s-r)}d\mathcal{W}(r)$ is the associated Ornstein-Uhlenbeck process with parameter c, and $\mathcal{N}$ is a standard normal random vector independent of $\mathcal{W}$.

Lemma P. Consider the model (2) and (3), where X_t consists of lags of ΔY_t, Y_{2t} is a scalar, ε_t satisfies Assumption A and z_t is given by (10). Let $\kappa_T=\frac{-\left(c_z+T^b\alpha_2\right)}{T^{1+b}}$. Then, as $T\to\infty$,
(i) $\kappa_T\sum_{t=m}^T z_t^2\xrightarrow{p}\omega$;
(ii) $\kappa_T\sum_{t=m}^T z_tY_{2,t-1}\implies 2\omega\left(\int_0^1\mathcal{J}_c d\mathcal{J}_c+1\right)$ if there exists $c\leq 0$ such that $T\alpha_2\to c$;
or $\kappa_T\sum_{t=m}^T z_tY_{2,t-1}\xrightarrow{p}\omega$ if $T\alpha_2\to-\infty$;
(iii) $\sqrt{\kappa_T}\sum_{t=m}^T z_t\binom{\varepsilon_{1t}}{v_{2t}}\xrightarrow{L}\begin{pmatrix}\sigma_{\varepsilon_1} & 0\\ 0 & \sigma_{v_2}\end{pmatrix}\sqrt{\omega}\mathcal{N}$;
(iv) $\sum_{t=m}^T z_t\Delta Y_{t-i}=O_p(T)$, $i=1,\dots m-1$;
(v) $\sum_{t=m}^T Y_{2,t-1}\Delta Y_{t-i}=O_p(T)$, $i=1,\dots m-1$;
(vi) $\sqrt{\frac{\kappa_T}{T}}\sum_{t=m}^T Y_{2,t-1}\varepsilon_{1t}=o_p(1)$.
(i) to (iii) also apply jointly.

B.1 Proof of Theorem 1

We first consider the case $m=1$ (SVAR(1)), so the numerator of the AR statistic in equation (9), simplifies to

$$(\Delta Y_1-\Delta Y_2b_{12})'P_z(\Delta Y_1-\Delta Y_2b_{12})=\sum_{t=1}^T\varepsilon_{1t}z_t\left(\sum_{t=1}^T z_t^2\right)^{-1}\sum_{t=1}^T z_t\varepsilon_{1t}.$$

When $T\alpha_2 \to c \leq 0$, Lemma P(i), (iii) implies that $\left(\sum_t z_t^2\right)^{-1/2} \sum_t z_t\varepsilon_{1t} \xrightarrow{L} N\left(0, \sigma_{\varepsilon_1}^2\right)$. Hence,

$$\sigma_{\varepsilon_1}^{-2} \sum_t \varepsilon_{1t} z_t \left(\sum_t z_t^2\right)^{-1} \sum_t z_t\varepsilon_{1t} \implies \chi_1^2. \tag{18}$$

Now the denominator is $(\Delta Y_1 - \Delta Y_2 b_{12})' M_z (\Delta Y_1 - \Delta Y_2 b_{12}) = \sum_{t=1}^T \varepsilon_{1t}^2 - \sum_{t=1}^T \varepsilon_{1t} z_t \left(\sum_{t=1}^T z_t^2\right)^{-1} \sum_{t=1}^T z_t\varepsilon_{1t}$. Since $E[z_t\varepsilon_{1t}] = 0$, the second element on the RHS of the previous expression is $O_p(1)$, so $T^{-1}(\Delta Y_1 - \Delta Y_2 b_{12})' M_z (\Delta Y_1 - \Delta Y_2 b_{12}) \xrightarrow{p} \sigma_{\varepsilon_1}^2$. This completes the proof when $m = 1$.

We now extend the above result to $m > 1$, which involves $X_{1t} = \left(\Delta Y_{t-1}', \ldots, \Delta Y_{t-m+1}'\right)'$. The numerator of the AR statistic, $(\Delta Y_1 - \Delta Y_2 b_{12})' P_{M_{X_1} z} (\Delta Y_1 - \Delta Y_2 b_{12})$, writes

$$\begin{aligned}
&\left(\sum_{t=m}^T \varepsilon_{1t} z_t - \sum_{t=m}^T \varepsilon_{1t} X_{1t}' \left(\sum_{t=m}^T X_{1t} X_{1t}'\right)^{-1} \sum_{t=m}^T X_{1t} z_t\right) \\
\times, &\left[\sum_{t=m}^T z_t^2 - \sum_{t=m}^T z_t X_{1t}' \left(\sum_{t=m}^T X_{1t} X_{1t}'\right)^{-1} \sum_{t=m}^T X_{1t} z_t\right]^{-1} \\
\times &\left(\sum_{t=m}^T z_t\varepsilon_{1t} - \sum_{t=m}^T z_t X_{1t}' \left(\sum_{t=m}^T X_{1t} X_{1t}'\right)^{-1} \sum_{t=m}^T X_{1t}\varepsilon_{1t}\right)
\end{aligned} \tag{19}$$

From Lemma P(iv) we have $\sum_{t=1}^T X_{1t} z_t = O_p(T)$. Moreover, $\sum_{t=1}^T X_{1t}\varepsilon_{1t} = O_p\left(T^{1/2}\right)$ because $X_{1t}'\varepsilon_{1t}$ constitutes a martingale difference sequence with bounded variance, and $\sum_{t=1}^T X_{1t} X_{1t}' = O_p(T)$, because X_{1t} is weakly dependent with bounded variance.

Hence if $\alpha_2 \to 0$ as $T \to \infty$, (19) behaves like (18) since the correction for the lags is of lower magnitude. When α_2 is constant, Kostakis et al. (2015) Lemma A.2 shows that expression (19) is asymptotically equivalent to $\varepsilon_1' M_{X_1} Y_2 \left(Y_2' M_{X_1} Y_2\right)^{-1} Y_2' M_{X_1}\varepsilon_1$, where Y_2 denotes the stacked $(Y_{2,t-1})$. Since $E[\varepsilon_{1t} Y_{2,t-1}] = 0$ and $Y_{2,t-1}$ is weakly stationary, it follows that $\sigma_{\varepsilon_1}^{-2}\varepsilon_1' M_{X_1} Y_2 \left(Y_2' M_{X_1} Y_2\right)^{-1} Y_2' M_{X_1}\varepsilon_1 \xrightarrow{d} \chi_1^2$. In both cases, the denominator satisfies $T^{-1}(\Delta Y_1 - \Delta Y_2 b_{12})' M_{(X_1,z)} (\Delta Y_1 - \Delta Y_2 b_{12}) \xrightarrow{p} \sigma_{\varepsilon_1}^2$.

B.2 General ARW test

Here we give high-level conditions to derive the properties of the combined ARW test in a general GMM setting, which we use to prove Theorem 2 in the next subsection.

Let $\theta \in \Theta$ denote a p-dimensional vector of parameters partitioned into $\theta = (\vartheta', \psi')'$

of dimensions p_ϑ and p_ψ, respectively. Let $F_T(\theta) = T^{-1}\sum_{t=1}^{T} f_t(\theta)$ denote the sample moments, where $f_t(\theta)$ is a k-dimensional vector-valued function of data and parameters with $k \geq p$ and $E(f_t(\theta)) = 0$ at the true value of θ. Let $r(\theta)$ be a known function of the parameters, $r : \Theta \to \Re^q$, $q \leq p_\psi$. Suppose $f_t(\vartheta, \cdot)$ and $r(\vartheta, \cdot)$ are continuously differentiable wrt ψ, and let $J_T(\theta) = \partial F_T(\theta)/\partial\psi'$ and $R(\theta) = \partial r(\theta)/\partial\psi'$. Let $\hat{V}_f(\theta)$ denote a $k \times k$ matrix that is positive definite almost surely, and define the GMM objective function

$$S_T(\vartheta, \psi) = F_T(\vartheta, \psi)' \hat{V}_f\left(\vartheta, \tilde{\psi}\right)^{-1} F_T(\vartheta, \psi),$$

where $\tilde{\psi}$ could be equal to some one-step GMM estimator (for 2-step GMM) or to ψ (for continuously updated GMM). Suppose the constrained GMM estimator of ψ given ϑ exists:

$$\hat{\psi}(\vartheta) = \arg\min_{\psi} F_T(\vartheta, \psi)' \hat{V}_f\left(\vartheta, \tilde{\psi}\right)^{-1} F_T(\vartheta, \psi).$$

To simplify notation, let $\hat{\psi} \equiv \hat{\psi}(\vartheta)$, $\hat{r}(\vartheta) = r\left(\vartheta, \hat{\psi}\right)$, $\hat{R}(\vartheta) = R\left(\vartheta, \hat{\psi}\right)$, $\tilde{V}_f(\vartheta) = \hat{V}_f\left(\vartheta, \tilde{\psi}\right)$, $\hat{F}_T(\vartheta) = F_T\left(\vartheta, \hat{\psi}\right)$ and $\hat{J}_T(\vartheta) = J_T\left(\vartheta, \hat{\psi}\right)$. Also, let $\hat{C}(\vartheta)$ be an almost surely full-rank $k \times (k - p_\psi)$ matrix that spans the null-space of $\tilde{V}_f(\vartheta)^{-1/2}\hat{J}_T(\vartheta)$, i.e., $\hat{C}(\vartheta)\hat{C}(\vartheta)' = M_{\tilde{V}_f(\vartheta)^{-1/2}\hat{J}_T(\vartheta)}$, where $M_X = I - P_X$, $P_X = X(X'X)^{-1}X'$.

Consider the statistic

$$ARW(\vartheta) = \hat{S}_T(\vartheta) + W_r(\vartheta)$$

where

$$\begin{aligned}
\hat{S}_T(\vartheta) &= S_T\left(\vartheta, \hat{\psi}\right) = \hat{F}_T(\vartheta)' \tilde{V}_f(\vartheta)^{-1} \hat{F}_T(\vartheta), \\
W_r(\vartheta) &= \hat{r}(\vartheta)' \left[\hat{R}(\vartheta) \hat{V}_{\hat{\psi}}(\vartheta) \hat{R}(\vartheta)'\right]^{-1} \hat{r}(\vartheta), \text{ and} \\
\hat{V}_{\hat{\psi}}(\vartheta) &= \left[\hat{J}_T(\vartheta)' \tilde{V}_f(\vartheta)^{-1} \hat{J}_T(\vartheta)\right]^{-1}.
\end{aligned} \tag{20}$$

Let $\hat{C}_{\hat{\psi}}$ be a square matrix such that $\hat{C}_{\hat{\psi}}\hat{C}'_{\hat{\psi}} = \hat{V}_{\hat{\psi}}(\vartheta)^{-1}$. The following result gives high-level conditions under which the asymptotic distribution of $ARW(\vartheta)$ is $\chi^2_{p_\vartheta+q}$ when ϑ is the true value of that parameter and $r(\theta) = 0$. It can then be used to form a test of

$$H_0^* : \vartheta = \vartheta_0, r(\theta) = 0 \quad \text{against} \quad H_1^* : \vartheta \neq \vartheta_0 \text{ and/or } r(\theta) \neq 0.$$

Theorem 3. *Suppose that at the true value of the parameters* $\theta = \binom{\vartheta}{\psi}$,
(i) $r(\theta) = 0$, *(ii)* $\tilde{\psi} \xrightarrow{p} \psi$, $\hat{\psi} \xrightarrow{p} \psi$,

$$\textit{(iii)} \quad \begin{pmatrix} \hat{\xi}_1 \\ \hat{\xi}_2 \end{pmatrix} \equiv \begin{pmatrix} \hat{C}(\vartheta)' \tilde{V}_f(\vartheta)^{-1/2} \hat{F}_T(\vartheta) \\ \hat{C}'_{\hat{\psi}} \left(\hat{\psi} - \psi \right) \end{pmatrix} \implies \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} \sim N(0, I_k),$$

(iv) there exist a non-stochastic $p_\psi \times p_\psi$ *symmetric matrix* $B_T \to 0$ *such that* $B_T \hat{C}_{\hat{\psi}} \implies \Psi$ *full-rank a.s., and (v) any stochastic elements in* Ψ *are independent of* $\xi = (\xi_1', \xi_2')'$. *Then, ARW* $(\vartheta) \xrightarrow{L} \chi^2_{k-p_\psi+q}$.

Proof By assumption (ii) and Slutsky's theorem we have $\hat{R}(\vartheta) = R(\theta) + o_p(1)$. By the singular value decomposition, $R(\theta) B_T = Q_T \Lambda_T U_T'$, where Q_T is an orthonormal $q \times q$ matrix, $\Lambda_T \to 0$ is a diagonal matrix holding the singular values of $R(\theta) B_T$, and U_T is a $p_\psi \times q$ matrix such that $U_T' U_T = I_q$. So,

$$\Lambda_T^{-1} Q_T' \hat{R}(\vartheta) B_T = \Lambda_T^{-1} Q_T' R(\theta) B_T + o_p(1) = U_T' + o_p(1).$$

Assumption (iv) implies that

$$B_T^{-1} \hat{V}_{\hat{\psi}}(\vartheta) B_T^{-1} = \left(B_T \hat{C}_{\hat{\psi}} \hat{C}'_{\hat{\psi}} B_T \right)^{-1} \implies \Psi^{-1\prime} \Psi^{-1}.$$

So,

$$\begin{aligned} \Lambda_T^{-1} Q_T' \hat{R}(\vartheta) \hat{V}_{\hat{\psi}}(\vartheta) \hat{R}(\vartheta)' Q_T \Lambda_T^{-1} &= \Lambda_T^{-1} Q_T' \hat{R}(\vartheta) B_T B_T^{-1} \hat{V}_{\hat{\psi}}(\vartheta) B_T^{-1\prime} B_T' \hat{R}(\vartheta)' Q_T \Lambda_T^{-1} \\ &= U_T' \Psi^{-1\prime} \Psi^{-1} U_T + o_p(1). \end{aligned}$$

Assumption (iii) then implies

$$B_T^{-1} \left(\hat{\psi} - \psi \right) = B_T^{-1} \hat{C}'^{-1}_{\hat{\psi}} \hat{C}'_{\hat{\psi}} \left(\hat{\psi} - \psi \right) = \Psi^{-1\prime} \xi_2 + o_p(1).$$

Assumption (ii) and a Taylor expansion of $\hat{r}(\vartheta)$ yield, under H_0^*,

$$\hat{r}(\vartheta) = R(\theta) \left(\hat{\psi} - \psi \right) + o_p \left(\left\| \hat{\psi} - \psi \right\| \right)$$

and $\Lambda_T^{-1} Q_T' \hat{r}(\vartheta) = U_T' B_T^{-1} \left(\hat{\psi} - \psi\right) + o_p(1)$ which for B_T symmetric yields

$$\Lambda_T^{-1} Q_T' \hat{r}(\vartheta) = U_T' \Psi^{-1\prime} \xi_2 + o_p(1).$$

Moreover,

$$\begin{aligned}
&\hat{r}(\vartheta)' \left[\hat{R}(\vartheta) \hat{V}_{\hat{\psi}}(\vartheta) \hat{R}(\vartheta)'\right]^{-1} \hat{r}(\vartheta) \\
&= \hat{r}(\vartheta)' Q_T \Lambda_T^{-1} \left[\Lambda_T^{-1} Q_T' \hat{R}(\vartheta) \hat{V}_{\hat{\psi}}(\vartheta) \hat{R}(\vartheta)' Q_T \Lambda_T^{-1}\right]^{-1} \Lambda_T^{-1} Q_T' \hat{r}(\vartheta) \\
&= \xi_2' \Psi^{-1\prime} U_T \left[U_T' \Psi^{-1\prime} \Psi^{-1} U_T\right]^{-1} U_T' \Psi^{-1\prime} \xi_2 + o_p(1).
\end{aligned}$$

Combining these results we have

$$ARW(\vartheta) = \begin{pmatrix} \xi_1 \\ \eta_T \end{pmatrix}' \begin{pmatrix} \xi_1 \\ \eta_T \end{pmatrix} + o_p(1),$$

where $\eta_T = \left[U_T' \Psi^{-1\prime} \Psi^{-1} U_T\right]^{-1/2} U_T' \Psi^{-1\prime} \xi_2$, and the conclusion of the theorem follows from Assumptions (v) and (iii), which imply that $\binom{\xi_1}{\eta_T} \xrightarrow{d} N\left(0, I_{k-p_\psi+q}\right)$, and the continuous mapping theorem.

B.3 Proof of Theorem 2

The proof involves verifying the conditions of Theorem 3. Intermediate results will be given as propositions whose proof can be found in the Supplementary Appendix.

The specification in Theorem 2 is a special case of that in Theorem 3, where $\vartheta = b_{12}$ and ψ contains all remaining elements θ. It is convenient to partition ψ into ψ_1 and ψ_2, where ψ_1 are the parameters that appear in equation (2) other than b_{12}, namely δ_1 and $\sigma_{\varepsilon_1}^2$, and ψ_2 are the parameters that appear only in (3), i.e., α_2, δ_2 and d_{21}. Because we can make $\hat{V}_f$ block diagonal by imposing the orthogonality of the errors ε_{1t} and v_{2t} that appear in f_{1t} and f_{2t}, respectively, estimation of ψ_1 and ψ_2 can be performed sequentially.

We start by obtaining expressions for $\hat{\xi}$ in Theorem 3, which forms the basis of the ARW statistic.

Proposition 4. *The estimator $\hat{\psi}$ is given by*

$$\hat{\psi}_1 = \begin{pmatrix} (X_1'X_1)^{-1} X_1' (\Delta Y_1 - \Delta Y_2 b_{12}) \\ T^{-1}\hat{\varepsilon}_1'\hat{\varepsilon}_1 \end{pmatrix}, \tag{21}$$
$$\hat{\psi}_2 = \left(\hat{Z}_2'\hat{X}_2\right)^{-1} \hat{Z}_2'\Delta Y_2,$$

where $\hat{\varepsilon}_1 = M_{X_1}(\Delta Y_1 - \Delta Y_2 b_{12})$, $\hat{X}_2 = \left(Y_2 \vdots X_2 \vdots \hat{\varepsilon}_1\right)$, *and* $\hat{Z}_2 = \left(z \vdots X_2 \vdots \hat{\varepsilon}_1\right)$. *The estimator of the variance of* $\hat{\psi}$ *is given by*

$$\hat{V}_{\hat{\psi}} = \begin{pmatrix} V_{\hat{\psi},11} & 0 & V_{\hat{\psi},13} \\ 0 & \frac{\hat{\varpi}}{T} & 0 \\ V_{\hat{\psi},13}' & 0 & V_{\hat{\psi},33} \end{pmatrix}, \tag{22}$$

where

$$\hat{V}_{\hat{\psi},11} = (X_1'X_1)^{-1}\hat{\sigma}_{\varepsilon_1}^2,$$
$$\hat{V}_{\hat{\psi},13} = (X_1'X_1)^{-1} X_1'\hat{Z}_2 \left(\hat{X}_2'\hat{Z}_2\right)^{-1} \hat{\sigma}_{\varepsilon_1}^2 d_{21},$$
$$\hat{V}_{\hat{\psi},33} = \left(\hat{Z}_2'\hat{X}_2\right)^{-1} \left(\hat{Z}_2'\hat{Z}_2\hat{\sigma}_{v_2}^2 + \hat{Z}_2'P_{X_1}\hat{Z}_2\hat{\sigma}_{\varepsilon_1}^2 d_{21}^2\right) \left(\hat{X}_2'\hat{Z}_2\right)^{-1}, \tag{23}$$

$\hat{\sigma}_{\varepsilon_1}^2 = T^{-1}\hat{\varepsilon}_1'\hat{\varepsilon}_1$, $\hat{\varpi} \xrightarrow{p} var\left(\hat{\sigma}_{\varepsilon_1}^2\right)$, $\hat{\sigma}_{v_2}^2 = T^{-1}\hat{v}_2'\hat{v}_2 \xrightarrow{p} E(v_{2t}^2)$ *and* $\hat{v}_2 = \Delta Y_2 - \hat{X}_2\hat{\psi}_2$. *It satisfies* $\hat{V}_{\hat{\psi}}(\vartheta)^{-1} = \hat{C}_{\hat{\psi}}\hat{C}_{\hat{\psi}}'$, *with*

$$\hat{C}_{\hat{\psi}} = \begin{pmatrix} (X_1'X_1)^{1/2}\hat{\sigma}_{\varepsilon_1}^{-1} & 0 & -d_{21}X_1'\hat{Z}_2 C_{\hat{Z}_2'\hat{Z}_2}'^{-1}\hat{\sigma}_{v_2}^{-1} \\ 0 & T^{1/2}\hat{\varpi}^{-1/2} & 0 \\ 0 & 0 & \hat{X}_2'\hat{Z}_2 C_{\hat{Z}_2'\hat{Z}_2}'^{-1}\hat{\sigma}_{v_2}^{-1} \end{pmatrix}, \tag{24}$$

where $C_{\hat{Z}_2'\hat{Z}_2}C_{\hat{Z}_2'\hat{Z}_2}' = \hat{Z}_2'\hat{Z}_2$. *The standardized random vector* $\hat{\xi}$ *defined in Theorem 3 is given by*

$$\hat{\xi}_1 = (z'M_{X_1}z)^{-1/2}\hat{\sigma}_{\varepsilon_1}^{-1} z'M_{X_1}\varepsilon_1, \text{ and} \tag{25}$$
$$\hat{\xi}_2 = \begin{pmatrix} (X_1'X_1)^{-1/2} X_1'\varepsilon_1\hat{\sigma}_{\varepsilon_1}^{-1} \\ \hat{\varpi}^{-1/2}\left(\hat{\sigma}_{\varepsilon_1}^2 - \sigma_{\varepsilon_1}^2\right) \\ C_{\hat{Z}_2'\hat{Z}_2}^{-1}\hat{Z}_2'v_2\hat{\sigma}_{v_2}^{-1} \end{pmatrix}. \tag{26}$$

Let

$$D_T = \begin{pmatrix} \sqrt{\kappa_T} & 0 \\ 0 & T^{-1/2} I_{p_{\psi_2}-1} \end{pmatrix}, \quad \kappa_T = \frac{-\left(c_z + T^b \alpha_2\right)}{T^{1+b}}, \tag{27}$$

and

$$B_T = \begin{pmatrix} T^{-1/2} I_{p_{\psi_1}} & 0 \\ 0 & D_T \end{pmatrix} \tag{28}$$

The following result verifies Assumptions (ii) of Theorem 3.

Proposition 5. *(i)* $\tilde{\psi} = \hat{\psi}$*, and (ii)* $\hat{\psi} \xrightarrow{p} \psi$*.*

Finally, we verify Assumptions (iii)-(v) of Theorem 3. By Proposition 5(ii), $\hat{\xi} = \hat{\xi}^* + o_p(1)$, where

$$\hat{\xi}^* = \begin{pmatrix} \left(\kappa_T z' M_{X_1} z\right)^{-1/2} \sqrt{\kappa_T} z' M_{X_1} \varepsilon_1 \sigma_{\varepsilon_1}^{-1}, \\ \left(T^{-1} X_1' X_1\right)^{-1/2} T^{-1/2} X_1' \varepsilon_1 \sigma_{\varepsilon_1}^{-1} \\ \varpi^{-1/2} T^{1/2} \left(\hat{\sigma}_{\varepsilon_1}^2 - \sigma_{\varepsilon_1}^2\right) \\ \left(D_T C_{\overline{Z}_2' \overline{Z}_2}\right)^{-1} D_T \overline{Z}_2' v_2 \sigma_{v_2}^{-1} \end{pmatrix}$$

where $\varpi = var\left(\hat{\sigma}_{\varepsilon_1}^2\right)$ and $\overline{Z}_2 \equiv (z_t, X_{2t}', \varepsilon_{1t})'$. Define the array

$$\zeta_{Tt} = \begin{pmatrix} \sqrt{\kappa_T} z_t \binom{\varepsilon_{1t}}{v_{2t}} \\ T^{-1/2} X_{1t} \varepsilon_{1t} \\ T^{-1/2} \left(\varepsilon_{1t}^2 - \sigma_\varepsilon^2\right) \\ T^{-1/2} \binom{X_{2t}}{\varepsilon_{1t}} v_{2t} \end{pmatrix},$$

which is a martingale difference with respect to the filtration $\mathcal{F}_{Tt} = \sigma\left(Y_0, \varepsilon_{1t}, v_{2t}, \varepsilon_{1,t-1}, v_{2,t-1} ...\right)$.

Proposition 6. $\sum_{t=1}^{T} \zeta_{Tt} \Rightarrow N(0, V_\zeta)$*, where* V_ζ *is nonstochastic and positive definite, and there exist a* $k \times \dim \zeta$ *matrix* G_T *such that* $\hat{\xi}^* = G_T \sum_{t=1}^{T} \zeta_{Tt}$*, where* $G_T V_\zeta G_T' \xrightarrow{p} I_k$*.*

Combining the above results verifies Assumption (iii) of Theorem 3, i.e.,

$$\hat{\xi} \Rightarrow \xi \sim N(0, I_k).$$

Finally, it remains to derive the asymptotic behavior of $B_T \hat{C}_{\hat{\psi}}$. This is done in the following Proposition.

Proposition 7. B_T *defined in (28) and* $\hat{C}_{\hat{\psi}}$ *defined in (24) satisfy Assumptions* (iv)-(v) *of Theorem 3.*

The theorem then follows from Theorem 3.

References

Anderson, T. W. and H. Rubin (1949). Estimation of the parameters of a single equation in a complete system of stochastic equations. *Ann. Math. Statistics 20*, 46–63.

Andrews, D. W. and X. Cheng (2012). Estimation and inference with weak, semi-strong, and strong identification. *Econometrica 80*(5), 2153–2211.

Andrews, D. W., X. Cheng, and P. Guggenberger (2011). Generic Results for Establishing the Asymptotic Size of Confidence Sets and Tests. Cowles Foundation Discussion Papers 1813, Cowles Foundation for Research in Economics, Yale University.

Beveridge, S. and C. R. Nelson (1981). A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the business cycle. *Journal of Monetary Economics 7*(2), 151–174.

Blanchard, O. J. and D. Quah (1989). The dynamic effects of aggregate demand and supply disturbances. *American Economic Review 79*(4), 655–73.

Chaudourne, J., P. Fève, and A. Guay (2014). Understanding the effect of technology shocks in svars with long-run restrictions. *Journal of Economic Dynamics and Control 41*, 154–172.

Christiano, L. J., M. Eichenbaum, and R. Vigfusson (2003). What happens after a technology shock? International Finance Discussion Papers 768, Board of Governors of the Federal Reserve System (U.S.).

Christiano, L. J., M. Eichenbaum, and R. Vigfusson (2007). Assessing structural VARs. In *NBER Macroeconomics Annual 2006, Volume 21*, pp. 1–106. MIT Press.

Dufour, J.-M. (1997). Some impossibility theorems in econometrics with applications to structural and dynamic models. *Econometrica 65*(6), 1365–1387.

Dupaigne, M., P. Fève, and J. Matheron (2007). Some analytics on bias in dsvars. *Economics Letters 97*(1), 32–38.

Fève, P. and A. Guay (2009). The response of hours to a technology shock: A two-step structural var approach. *Journal of Money, Credit and Banking 41*(5), 987–1013.

Fève, P. and A. Guay (2010). Identification of technology shocks in structural vars. *The Economic Journal 120*(549), 1284–1318.

Francis, N. and V. A. Ramey (2005). Is the technology-driven real business cycle hypothesis dead? shocks and aggregate fluctuations revisited. *Journal of Monetary Economics 52*(8), 1379–1399.

Francis, N. and V. A. Ramey (2009). Measures of per capita hours and their implications for the technology-hours debate. *Journal of Money, Credit and Banking 41*(6), 1071–1097.

Fukac, M. and A. Pagan (2006). Issues in adopting DSGE models for use in the policy process. CAMA Working Papers 2006-10, Australian National University, Centre for Applied Macroeconomic Analysis.

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

Gospodinov, N. (2010). Inference in nearly nonstationary SVAR models with long-run identifying restrictions. *Journal of Business and Economic Statistics 28*(1), 1–11.

Gospodinov, N., A. Maynard, and E. Pesavento (2011). Sensitivity of impulse responses to small low-frequency comovements: Reconciling the evidence on the effects of technology shocks. *Journal of Business and Economic Statistics 29*(4), 455–467.

Kilian, L. (1998). Small-sample confidence intervals for impulse response functions. *Review of Economics and Statistics 80*(2), 218–230.

Kostakis, A., T. Magdalinos, and M. P. Stamatogiannis (2015). Robust econometric inference for stock return predictability. *Review of Financial Studies 28*(5), 1506–1553.

Magdalinos, A. and P. C. B. Phillips (2009). Econometric inference in the vicinity of unity. Working paper, Yale University, USA.

McCloskey, A. (2012). Bonferroni-based size-correction for nonstandard testing problems. Technical report, Brown University.

Mittnik, S. and P. A. Zadrozny (1993). Asymptotic distributions of impulse responses, step responses, and variance decompositions of estimated linear dynamic models. *Econometrica 61*(4), 857–870.

Pagan, A. R. and M. H. Pesaran (2008). Econometric analysis of structural systems with permanent and transitory shocks. *Journal of Economic Dynamics and Control 32*(10), 3376–3395.

Pagan, A. R. and J. C. Robertson (1998). Structural models of the liquidity effect. *Review of Economics and Statistics 80*(2), 202–217.

Pesavento, E. and B. Rossi (2005). Do technology shocks drive hours up or down? a little evidence from an agnostic procedure. *Macroeconomic Dynamics 9*(04), 478–488.

Phillips, P. C., J. Y. Park, and Y. Chang (2004). Nonlinear instrumental variable estimation of an autoregression. *Journal of Econometrics 118*(1), 219–246.

Phillips, P. C. B. (2014). On confidence intervals for autoregressive roots and predictive regression. *Econometrica 82*(3), 1177–1195.

Ramey, V. A. (2016). Macroeconomic shocks and their propagation. Technical report, National Bureau of Economic Research.

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

Staiger, D. and J. H. Stock (1997). Instrumental variables regression with weak instruments. *Econometrica 65*(3), 557–586.

Stock, J. H., J. H. Wright, and M. Yogo (2002). A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business and Economic Statistics 20*(4), 518–529.

PARIS — SINGAPORE

ESSEC Business School
3 avenue Bernard-Hirsch
CS 50105 Cergy
95021 Cergy-Pontoise Cedex
France
Tel. +33 (0)1 34 43 30 00
www.essec.edu

ESSEC Executive Education
CNIT BP 230
92053 Paris-La Défense
France
Tel. +33 (0)1 46 92 49 00
www.executive-education.essec.edu

ESSEC Asia-Pacific
5 Nepal Park
Singapore 139408
Tel. +65 6884 9780
www.essec.edu/asia

ESSEC Africa-Atlantic
Plage des Nations
Sidi Bouknadel
Rabat-Salé
Morocco
Tel. +212 (0)5 30 10 40 19
www.essec.edu

ESSEC Africa-Indian Ocean
Royal Road, Pierrefonds
Mauritius
Tel. +230 401 2400
www.essec.edu
www.icsia.mu

MOROCCO — MAURITIUS

Contacts

Centre de Recherche
+33 (0)1 34 43 30 91
research.center@essec.fr

AACSB ACCREDITED
EFMD EQUIS ACCREDITED
affilié à la CCI PARIS ILE-DE-FRANCE

ESSEC BUSINESS SCHOOL