

Robust inference in structural VARs with long-run restrictions*

Guillaume Chevillon Sophocles Mavroeidis[†]
ESSEC and CREST University of Oxford

Zhaoguo Zhan
Tsinghua University

23 May 2015

Abstract

Long-run restrictions (Blanchard and Quah, 1989) are a very popular method for identifying structural vector autoregressions (SVARs). A prominent example is the debate on the effect of technology shocks on employment, which has been used to test real business cycle theory (Gali, 1999, Christiano Eichenbaum and Vigfusson, 2003). The long-run identifying restriction is that non-technology shocks have no permanent effect on productivity. This can be used to identify the technology shock and the impulse responses to it. It is well-known that long-run restrictions can be expressed as exclusion restrictions in the SVAR and that they may suffer from weak identification when the degree of persistence of the instruments is high (Pagan and Robertson, 1998). This introduces additional nuisance parameters and entails nonstandard distributions, so standard weak-instrument-robust methods of inference are inapplicable. We develop a method of inference

*We would like to thank Jean-Marie Dufour, Patrick Fève, Tassos Magdalinos, Nour Meddahi, Adrian Pagan, and seminar participants at the Universities of Melbourne, Toulouse and the IWH-CIREQ Macroeconometrics Workshop in Halle for helpful comments and discussion. Mavroeidis would like to thank the European Commission for research support under a FP7 Marie Curie Fellowship CIG 293675.

[†]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

that is robust to this problem. The method is based on a combination of the Anderson and Rubin (1949) test with instruments derived by filtering potentially non-stationary variable to make them near stationary (Magdalinos and Phillips, 2009, Phillips, 2014, Kostakis Magdalinos and Stamatogiannis, 2015). 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 Gali (1999) is very well identified, while the level specification of Christiano et. al. (2003) is weakly identified.

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) the ability of the model 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 the features that make instruments weak in this context also work to make them highly persistent, or nearly non-stationary. Therefore, all the available weak identification robust methods of inference, such as the Anderson and Rubin (1949), 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).¹

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 Gali (1999) and Christiano et al. (2003). In the case of Blanchard and Quah (1989), we find that long-run restrictions yield very weak identification, since confidence bands on the impulse responses comprise the entire parameter space (which is bounded). On the hours debate, we find that the difference specification of Gali (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 4363 Google scholar

¹Note that the presence of persistent regressors affects inference on IRFs at long horizons under any identification scheme, see Pesavento and Rossi (2006, 2007). We do not study long-horizon IRFs here, but we believe that use of filtered instruments provides valid inference for long-horizon IRFs, too, though the methods in Pesavento and Rossi (2006) may be more efficient.

citations, and we found that about half of all the articles that used SVARs published between 2005 and 2014 in the top general interest and macro journals in economics used long-run restrictions.² 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 sets up the model and assumptions and discusses the long-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 and additional numerical and empirical results are given in an appendix at the end.

2 Model and assumptions

2.1 The baseline SVAR(k) with long-run restrictions

A general SVAR with k lags can be written as

$$B(L)Y_t = \varepsilon_t, \quad B(L) = \sum_{j=0}^k B_j L^j \quad (1)$$

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, $\text{var}(\varepsilon_t)$ is a n -dimensional diagonal variance matrix, and B_0 has ones along its diagonal. Moreover, the defining assumption of the VAR is $E(\varepsilon_t | Y_{t-1}, Y_{t-2}, \dots) = 0$.

Partition the vector of structural shocks $\varepsilon_t = \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$. We are interested in identifying ε_{1t} , and the IRF

$$g_j = \frac{\partial Y_{t+j}}{\partial \varepsilon_{1t}}, \quad j = 0, 1, \dots$$

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).

²American 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.

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, i.e., Y_t is at most I(1). For clarity, we discuss here the bivariate case, $n = 2$ – multivariate generalization is straightforward. It is typically assumed (e.g., by Gali (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).

2.1.1 Case of one permanent shock

This is a cointegrated VAR, or vector error correction model (VECM), which can be written as

$$\Gamma(L) \Delta Y_t = \underbrace{\alpha}_{2 \times 1} \underbrace{\beta'}_{1 \times 2} Y_{t-1} + B_0^{-1} \varepsilon_t, \quad (2)$$

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

$$Y_t = C \sum_{s=1}^t \varepsilon_s + \tilde{C}(L) \varepsilon_t, \quad C = \beta_{\perp} (\alpha'_{\perp} \Gamma(1) \beta_{\perp})^{-1} \alpha'_{\perp} B_0^{-1},$$

where $\alpha'_{\perp} \alpha = 0$, $\alpha = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix}$, $\alpha_{\perp} = \begin{pmatrix} \alpha_2 \\ -\alpha_1 \end{pmatrix}$ and similarly for β . The long-run restriction that permanent shocks to Y_{2t} have no impact on 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 $\alpha_{\perp} B_0^{-1} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = 0$, 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:

$$\begin{aligned}\gamma_{11}(L) \Delta Y_{1t} &= \alpha_1 \beta' Y_{t-1} + \gamma_{12}(L) \Delta Y_{2t} + u_{1t} \\ \gamma_{22}(L) \Delta Y_{2t} &= \alpha_2 \beta' Y_{t-1} + \gamma_{21}(L) \Delta Y_{1t} + u_{2t},\end{aligned}$$

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

$$\tilde{\gamma}_{11}(L) \Delta Y_{1t} = b_{12} \Delta Y_{2t} + \tilde{\gamma}_{12}(L) \Delta Y_{2t} + \varepsilon_{1t}, \quad (3)$$

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 $\beta' Y_{t-1}$ is missing from (3), so we can use this to instrument for the endogenous regressor ΔY_{2t} . Once ε_{1t} is identified from (3), the impact of ε_{1t} on Y_{2t} can be obtained from the regression

$$\gamma_{22}(L) \Delta Y_{2t} = \alpha_2 \beta' Y_{t-1} + \gamma_{21}(L) \Delta Y_{1t-1} + d_{21} \varepsilon_{1t} + \varepsilon_{2t}. \quad (4)$$

Identification is weak if $\alpha_2 \rightarrow 0$.

2.1.2 Case of two permanent shocks

In this case there is no cointegration, so the model is a VAR in first differences:

$$\Gamma(L) \Delta Y_t = B_0^{-1} \varepsilon_t.$$

The long-run restriction that permanent shocks to Y_{2t} have no impact on 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

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

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

$$\tilde{\gamma}_{11}(L) \Delta Y_{1t} = \gamma_{12}^*(L) \Delta^2 Y_{2t} + \varepsilon_{1t}, \quad (5)$$

and the equation for Y_{2t} reads

$$\gamma_{22}(L) \Delta Y_{2t} = \gamma_{21}(L) \Delta Y_{1,t-1} + d_{21} \varepsilon_{1t} + \varepsilon_{2t}.$$

Thus, we are using $\Delta Y_{2,t-1}$ as an instrument for the endogenous regressor $\Delta^2 Y_{2t}$ in (5). Identification is weak if ΔY_{2t} is nearly I(1).

2.2 The hours debate

The number of permanent shocks can make a big impact on the results. The debate of the short-run effect of a technology shock on hours between Galí (1999) and Christiano et al, is based on a SVAR that contains productivity and hours. Galí used a VAR in first differences (two permanent shocks), found a negative effect and rejected RBC theory. Christiano et al favored a VAR with hours in levels (one permanent shock, cointegrating vector $\beta = (0, 1)'$) and found a positive effect – they also used per-capita hours as opposed to total hours, which also matters. Christiano et al claimed the “level” specification encompasses the “difference” one, and is preferred by the data.

It is true that the level specification nests the difference specification. Consider the following encompassing specification:

$$\begin{aligned} \tilde{\gamma}_{11}(L) \Delta Y_{1t} &= \tilde{\gamma}_{12}^*(L) \Delta^2 Y_{2t} + [b_{12} + \tilde{\gamma}_{12}(1)] \Delta Y_{2t} + \varepsilon_{1t} \\ \gamma_{22}(L) \Delta Y_{2t} &= \alpha_2 Y_{2,t-1} + \gamma_{21}(L) \Delta Y_{1t} + u_{2t}, \end{aligned} \quad (6)$$

The level specification imposes no extra restriction, and uses $Y_{2,t-1}$ as an instrument in (6). 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 (6). 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 corroborates Christiano et al's claim. But if we are in a situation when α_2 is small, which we can model as $\alpha_2 \rightarrow 0$ as $T \rightarrow \infty$, the level specification suffers from weak identification because of a near unit root in Y_{2t} . Therefore, the sampling uncertainty in the level specification may be so large that we cannot rule out conclusions based on the difference specification.

3 Econometric Methods

3.1 GMM estimating equations

Consider the multivariate SVAR(m) in n variables. We are interested in identifying the IRF to the first shock ε_{1t} using $(n - 1)$ long-run restrictions. This can be done by estimating equations (3) and (4). Let θ denote all the parameters of the model. Moreover, let X_{1t}, X_{2t} denote vectors containing lags of ΔY_t . In canonical, unrestricted SVAR(m), $X_{1t} = X_{2t} = (\Delta Y'_{t-1}, \dots, \Delta Y'_{t-m+1})'$, so that (3) and (4) can be written compactly as

$$\Delta Y_{1t} = b'_{12} \Delta Y_{2t} + \delta'_1 X_{1t} + \varepsilon_{1t} \quad (7)$$

$$\Delta Y_{2t} = \alpha_2 Y_{2,t-1} + \delta'_2 X_{2t} + \underbrace{d_{21} \varepsilon_{1t} + v_{2t}}_{u_{2t}}, \quad (8)$$

where δ_1 denotes the coefficients on exogenous and predetermined variables in (3), and δ_2 denotes the corresponding coefficients in (4).³ Note that v_{2t} is the residual of the projection of the reduced form error u_{2t} on ε_{1t} . It coincides with ε_{2t} when it is a scalar ($n = 2$), but not otherwise. So, as is well-known, the $n - 1$ long-run restrictions above combined with the orthogonality of the structural shocks, do not identify the structural shocks ε_{2t} when $n > 2$. For clarity, we will discuss the case $n = 2$ in the remainder of this section. Extension to the general case $n > 2$ is given below.

Because the model is just-identified, maximum likelihood estimation of θ can be expressed in terms of the method of moments. Let

$$h_{1t}(\theta_1) = b_{12} \Delta Y_{2t} + \delta'_1 X_{1t}, \quad (9)$$

³ X_{1t} and X_{2t} do not need to be the same and do not need to include all k lagged differences of the variables. They may also contain deterministic terms, see Section 3.4.

where $\theta_1 = (b_{12}, \delta'_1, \sigma_{\varepsilon_1})'$, and

$$h_{2t}(\theta) = \Delta Y_{2t} - \alpha_2 Y_{2,t-1} - \delta'_2 X_{1t} - d_{21} h_{1t}(\theta_1). \quad (10)$$

Note that (9) and (10) correspond to the ‘level’ specification, which is more general, but can easily accommodate the difference specification by redefining Y_{2t} accordingly. Let Z_{1t}, Z_{2t} denote vectors of instruments to be specified below. The identifying restrictions can be expressed as the moment equations $E(f_t(\theta)) = 0$, where $f_t = (f'_{1t}, f'_{2t})'$ and

$$f_{1t}(\theta_1) = \begin{pmatrix} Z'_{1t} h_{1t}(\theta_1) \\ h_{1t}(\theta_1)^2 - \sigma_{\varepsilon_1}^2 \end{pmatrix}, \quad f_{2t}(\theta) = \begin{pmatrix} Z'_{2t} h_{2t}(\theta) \\ h_{1t}(\theta_1) h_{2t}(\theta) \end{pmatrix}. \quad (11)$$

This structure of the moment conditions and the corresponding block-diagonality of the efficient GMM weighting matrix allows us to define the GMM estimator sequentially, see below.

Under standard strong-instrument stationary asymptotics define the long-run variance of the moment conditions $V_f(\theta) = \text{var} \left(T^{-1/2} \sum_{t=1}^T f_t(\theta) \right)$. Let $\hat{V}_f(\bar{\theta})$ denote a consistent estimator of $V_f(\theta)$, where $\bar{\theta}$ is some initial estimator of θ . Note that by the finite-order VAR assumption $V_f(\theta) = \text{var}(f_t(\theta))$, so $\hat{V}_f(\theta)$ does not need to be a HAC estimator. The GMM criterion is

$$S_T(\theta) = F_T(\theta)' \hat{V}_f(\bar{\theta})^{-1} F_T(\theta),$$

where $F_T(\theta) = T^{-1} \sum_{t=1}^T f_t(\theta)$. The GMM estimator is $\hat{\theta} = \arg \min_{\theta} S_T(\theta)$. Because the SVAR model is linear and just-identified, under (conditional) homoskedasticity, GMM becomes 2SLS equation by equation. Wald-tests and confidence intervals are based on standard first-order strong-instruments stationary asymptotics:

$$\sqrt{T}(\hat{\theta} - \theta) \xrightarrow{d} N\left(0, [J(\theta)' V_f(\theta)^{-1} J(\theta)]^{-1}\right),$$

where $J(\theta) = p \lim_{T \rightarrow \infty} \partial F_T(\theta) / \partial \theta'$. However, under near-unit-root asymptotics the above result breaks down.

Orthogonality of the errors implies that the variance matrix of f_t is block diagonal, with $V_{f_1}(\theta_1) = \text{var}(f_{1t}(\theta_1))$ and $V_{f_2}(\theta) = \text{var}(f_{2t}(\theta))$, so the GMM criterion function

can be decomposed into orthogonal components

$$\begin{aligned}
S_T(\theta) &= (F_{1T}(\theta_1)', F_{2T}(\theta)') \begin{pmatrix} \hat{V}_{f_1}^{-1} & 0 \\ 0 & \hat{V}_{f_2}^{-1} \end{pmatrix} \begin{pmatrix} F_{1T}(\theta_1) \\ F_{2T}(\theta) \end{pmatrix} \\
&= \underbrace{F_{1T}(\theta_1)' \hat{V}_{f_1}^{-1} F_{1T}(\theta_1)}_{S_{1T}(\theta_1)} + \underbrace{F_{2T}(\theta)' \hat{V}_{f_2}^{-1} F_{2T}(\theta)}_{S_{2T}(\theta)}. \tag{12}
\end{aligned}$$

where

$$F_{1T}(\theta_1) = \frac{1}{T} \sum_{t=1}^T f_{1t}(\theta_1), \quad F_{2T}(\theta) = \frac{1}{T} \sum_{t=1}^T f_{2t}(\theta).$$

3.1.1 The impulse response function

The IRF of interest is given by

$$g_j(\theta) = \frac{\partial Y_{t+j}}{\partial \varepsilon_{1t}} = (\mathbf{I}_n, \mathbf{0}_{n \times n(m-1)}) A(\theta)^j \begin{pmatrix} 1 \\ \mathbf{0}_{n(m-1) \times 1} \end{pmatrix} g_0(\theta), \quad j \geq 1 \tag{13}$$

where $A(\theta)$ is the companion VAR matrix and g_0 are the impact IRFs:

$$g_0(\theta) = \begin{pmatrix} 1 + b_{12}d_{21} \\ d_{21} \end{pmatrix} \sigma_{\varepsilon_1}. \tag{14}$$

This is the IRF to a one-standard-deviation shock to ε_{1t} . Alternatively, we can use the IRF to one unit shock to ε_{1t} by dropping σ_{ε_1} from (14).

3.2 The conventional approach

The conventional approach, e.g., Blanchard and Quah (1989), is to use Gaussian maximum likelihood (ML) estimation with conditional homoskedasticity. The ML estimator corresponds to the GMM estimator defined above when $Z_{1t} = (Y'_{2,t-1}, X'_{1t})$ and $Z_{2t} = (Y'_{2,t-1}, X'_{2t})$, namely, when we use $Y_{2,t-1}$ as instruments in (7)-(8). This corresponds to 2SLS estimation of θ_1 from (7), yielding estimate $\hat{\theta}_1$, and OLS estimation of θ_2 from (8), with the ‘generated regressor’ $\hat{\varepsilon}_{1t} = h_{1t}(\hat{\theta}_1)$. Under strong-instrument stationary asymptotics, the asymptotic distribution of Wald statistics is χ^2 and error bands for any smooth function of the parameters can be derived using the delta method, e.g. Mitnik and Zadrozny (1993), or 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 distribution of Wald statistics depends 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 VAR coefficients and IRFs do not have correct asymptotic coverage. In the next subsection, which contains the main contribution of the paper, we introduce a method that does.

3.3 Anderson Rubin test with filtered instruments

Our approach to solving this problem consists of two components: (i) address the *near unit root* problem, $\alpha_2 = O(T^{-a})$ with $a > 0$, by using filtered instruments – the so-called IVX approach of Magdalinos and Phillips (2009, henceforth MP); (ii) address the *weak-instrument* problem using a weak-identification robust method – the Anderson and Rubin (1949) (henceforth AR) test, since the model is typically just-identified. It is crucial to use both components – using any one of them alone does not suffice to control size.

3.3.1 A brief description of the IVX method of Magdalinos and Phillips (2009)

MP 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.

We illustrate the idea using the predictive regression example in Kostakis et al. (2015, henceforth KMS). Consider the system of equations

$$\begin{aligned} y_t &= \theta x_{t-1} + u_t \\ x_t &= \left(1 + \frac{k_T}{T}\right) x_{t-1} + v_t, \end{aligned}$$

where $k_T = O(T)$ as $T \rightarrow \infty$ such that $k_T/T \in [-1, 0]$, and suppose we are interested in doing inference on θ . Instead of using the OLS t -test, following MP, KMS proposed

to use IV with the following generated instrument:

$$z_t = \sum_{j=1}^t \rho_{Tz}^{t-j} \Delta x_j, \quad \rho_z = 1 + \frac{c_z}{T^b}, \quad b \in (0, 1), \quad c_z < 0.$$

so $z_t = \rho_z z_{t-1} + \Delta x_t$, with ρ_z sufficiently smaller than 1. When $k_T = cT^{a-1}$, $c \leq 0$ and $a \in [0, 1]$ they showed that the IV t -test is asymptotically *standard normal* under H_0 irrespective of a , c and the choice of b , c_z .⁴ We show in the Supplementary Appendix that this results holds for all sequences such that $k_T/T \in [-1, 0]$.

3.3.2 Filtered instruments for the SVAR model

We take that approach to our model as follows. In the original model, the instruments contain all lagged differences that appear on the right hand side of all equations (which we denoted by X_{1t}), plus the lagged stationary regressors $Y_{2,t-1}$, which are excluded only from the first equation, i.e., $Z_{1t} = Z_{2t} = (X_{1t}, Y_{2,t-1})$. The alternative we propose is to replace $Y_{2,t-1}$ in the instrument set with the filtered instrument

$$z_t = \sum_{j=1}^{t-1} \rho_{Tz}^{t-j} \Delta Y_{2,j}, \quad \rho_{Tz} = 1 + \frac{c_z}{T^b}, \quad b \in (0, 1), \quad c_z < 0 \quad (15)$$

(we follow MP in setting $c_z = -1$ and $b = 0.95$).

3.3.3 The AR statistic

The next step in our methodology is to construct the AR test using those instruments. Consider first a statistic for testing $H_0 : b_{12} = b_{12}^0$. Because of the block diagonality of V_f , this can be tested using just the first equation (3). The AR statistic, $AR(b_{12}^0)$, is the squared t -statistic for testing $H_0^* : \delta_z = 0$ in the auxiliary regression:

$$\Delta Y_{1t} - b_{12}^0 \Delta Y_{2t} = \delta_1' X_{1t} + z_t \delta_z + \varepsilon_{1t}^0 \quad (16)$$

where X_{1t} contains the $m - 1$ lags of ΔY_t , and z_t is the filtered instrument (15). Note that with conditional homoskedasticity, $AR(b_{12}^0)$ corresponds exactly to the minimum value of $S_{1T}(\theta_1)$, defined in (12), under the restriction that $b_{12} = b_{12}^0$, i.e.,

⁴This example is a special case of KMS Theorem 1.

$S_{1T}(\hat{\theta}_1(b_{12}^0))$, where $\hat{\theta}_1 = \arg \min_{\theta_1: b_{12} = b_{12}^0} S_{1T}(\theta_1)$. Under conditional homoskedasticity, this can be written analytically as

$$AR(b_{12}) = \frac{(\Delta Y_1 - \Delta Y_2 b_{12})' P_{M_{X_1 z}} (\Delta Y_1 - \Delta Y_2 b_{12})}{(\Delta Y_1 - \Delta Y_2 b_{12})' M_{Z_1} (\Delta Y_1 - \Delta Y_2 b_{12}) / (T - \text{col}(Z_1))}, \quad (17)$$

where P denotes the projection matrix, $M = I - P$. $Z_1 = (X_1, z)$, and we follow standard notation that for any column vector X_t , X denotes the matrix of T stacked rows X_t' , $t = 1, \dots, T$.

To proceed, 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 $E(\varepsilon_t | Y_{t-1}, Y_{t-2}, \dots) = 0$, $E(\varepsilon_t \varepsilon_t' | Y_{t-1}, Y_{t-2}, \dots) = \Sigma_\varepsilon$ and diagonal with $\Sigma_\varepsilon > 0$, and the moment condition $E\|\varepsilon_t\|^4 < \infty$.

This assumption is similar to the one used in MP, except for the addition of conditional homoskedasticity, which is typically used in the literature (e.g., the results in Blanchard and Quah, 1989, and Gali, 1999, assume conditional homoskedasticity). Heteroskedasticity robust version of the proposed tests can be obtained using a heteroskedasticity consistent estimator of V_f .

Our proposed AR test is then based on the following result.

Theorem 1. *Consider the model (7) and (8), with $\alpha_2 = k_T/T \in [-1, 0]$, and ε_t satisfying Assumption A. Let statistic $AR(b_{12})$ as in (17) where the instrument is defined by (15). Then under $H_0 : b_{12} = b_{12}^0$, $AR(b_{12}^0) \xrightarrow{d} \chi_1^2$ for all sequences k_T .*

Comments 1. The proof of the Theorem is somewhat simpler than in MP and KMS because no variable in the auxiliary regression is near-integrated. Thus, $\hat{\delta}_z$ is asymptotically Normal, rather than mixed normal, in all cases.

2. Extending $k_T = cT^{1-a}$, the case considered by MP and KMS, to general sequences k_T is easily done by following the proofs in MP, see the supplementary appendix.

3. The case $k_T/T = c < 0$ corresponds to stationarity and strong identification. In this case, the statistic AR in (17) is asymptotically equivalent to the AR statistic

AR^* defined in (30) that is obtained by replacing the filtered instrument z_t with $Y_{2,t-1}$.⁵ AR^* is the standard AR statistic, which is asymptotically efficient because the model is just identified (Moreira, 2003), and 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 AR tests.

3.3.4 A projection test for general hypotheses

The proposed methodology can be extended to testing general hypotheses of the form $H_0 : r(\theta) = 0$, where $r : \Theta \rightarrow \mathfrak{R}^q$, $q < \dim \theta$. This includes e.g., the IRF and forecast error variance decomposition. Testing such hypotheses is difficult because $r(\theta)$ contains the potentially weakly identified parameter b_{12} . Since this is the only parameter that is affected by weak identification, we propose the following projection testing approach. Use a test of the joint null hypothesis $H_0^* : r(\theta) = 0, b_{12} = b_{12}^0$, and “project out” b_{12} , i.e., reject $H_0 : r(\theta) = 0$ if there is no value of b_{12}^0 for which H_0^* is accepted. We now turn to the derivation of a test of H_0^* .

Our test of the combined hypothesis H_0^* is based on a novel idea that combines the $AR(b_{12})$ statistic developed above with a Wald statistic for testing the restrictions on the *remaining* parameters in θ (this idea applies more generally, see section A.2 and Theorem 3 in the Appendix). Partition θ into b_{12} and ψ , say, the remaining unknown parameters. 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

$$W(b_{12}) = r\left(b_{12}, \hat{\psi}(b_{12})\right)' \hat{V}_{\hat{r}}(b_{12})^{-1} r\left(b_{12}, \hat{\psi}(b_{12})\right), \quad (18)$$

where $\hat{V}_{\hat{r}}(b_{12}) = R\left(b_{12}, \hat{\psi}(b_{12})\right)' \hat{V}_{\hat{\psi}}(b_{12}) R\left(b_{12}, \hat{\psi}(b_{12})\right),$

and consider the combined statistic

$$ARW(b_{12}^0) = AR(b_{12}^0) + W(b_{12}^0). \quad (19)$$

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

⁵The asymptotic equivalence $AR = AR^* + o_p(1)$ is shown in the proof of Proposition 7 in the Supplement.

result.

Theorem 2. *Under the conditions of Theorem 1, with $b \in (1/2, 1)$ in the definition of the filtered instrument z_t , and if the null hypothesis $H_0^* : r(\theta) = 0, b_{12} = b_{12}^0$ holds, then, for all $0 \leq a \leq 1, c \leq 0$:*

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

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

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

Comments 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 the $1 - \eta$ quantile of χ_{1+q}^2 . A projection test of $H_0 : r(\theta) = 0$ rejects H_0 when there is no value of b_{12}^0 such that the ARW test accepts H_0^* .

2. Because of the conditional homoskedasticity case, the confidence set for b_{12} can be obtained analytically using Dufour and Taamouti (2005) and that greatly speeds up computation. But even in the general case, computation of confidence bands only requires a grid search over the space of b_{12} , so if the latter is a scalar, it is quite fast, too.

3. This test can be applied, e.g. to the IRF or forecast error variance decomposition, see Examples below. Importantly, the test is valid also for inference on long-horizon IRFs, see Example LHIRF below.

Example SHIRF A bivariate SVAR(1) with a long-run restriction and cointegrating vector $\beta = (0, 1)'$ is a special case of (3) and (4) with

$$\Delta Y_{1t} = b_{12} \Delta Y_{2t} + \varepsilon_{1t}, \quad (20)$$

$$\Delta Y_{2t} = \alpha_2 Y_{2,t-1} + d_{21} \varepsilon_{1t} + \varepsilon_{2t}. \quad (21)$$

The structural parameters $\theta = (b_{12}, \alpha_2, d_{21})'$ are partitioned into b_{12} and $\psi = (\alpha_2, d_{21})'$. 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 $r(b_{12}, \psi) = d_{21} - d_{21}^0$, with $R(b_{12}, \psi) = \partial r / \partial \psi' = (0, 1)$. Our proposed η -level ARW test rejects H_0 if $\min_{b_{12}} (AR(b_{12}) + W(b_{12}))$ is greater than the $1 - \eta$ quantile of χ_2^2 .

Example LHIRF Suppose we are interested in the impulse response of Y_2 to ε_1 at horizon $j = \kappa T$, in the special case where $\alpha_2 = cT^{-1}$. This example corresponds to the problem studied by Pesavento and Rossi (2005, 2006, 2007) and Mikusheva (2012), though these authors did not consider the case of long-run restrictions which matters. Specifically, the parameter of interest is given by $\partial Y_{2t+j}/\partial \varepsilon_{1t} = (1 + \alpha_2)^j d_{21} \approx e^{c\kappa} d_{21}$ when $j = \kappa T$ and T is large. The procedures proposed in the aforementioned papers require d_{21} to be consistently estimated and inference on this parameter is obtained by using a robust test for c . The additional complication here is that d_{21} is weakly identified when α_2 is local to zero, so uncertainty in d_{21} (stemming from weak identification of b_{12}) is not asymptotically negligible – it is of the same order as for c .

Our ARW test covers this case. Specifically, consider the null hypothesis H_0 : $\partial Y_{2t+\kappa T}/\partial \varepsilon_{1t} = r_0$. This can be expressed as the analytic restriction $r(b_{12}, \psi) = e^{\alpha_2 j} d_{21} - r_0$, where j is known, so that $R(b_{12}, \psi) = \partial r/\partial \psi' = e^{\alpha_2 j} (j, 1)$. By Theorem 2, the asymptotic size of the projection η -level ARW test described in Example SHIRF will not exceed η for all c .

3.4 Deterministic terms

The model (7)-(8) above did not include any deterministic terms in X_{1t} and X_{2t} . Although their values do not affect the IRF or forecast error variance decomposition, misspecification of the deterministic elements may result in inconsistent estimators of θ . Under stationarity assumptions, the use of consistent estimators of the deterministic terms (we restrict our attention to an intercept or a linear trend) will preserve the asymptotic distributions of the AR and ARW statistics. They may have an impact in the presence of near-unit roots.

When a constant is present in (7), we show in the supplementary appendix that the results of KMS hold and the test statistics are asymptotically unaffected. The estimated intercept matters in finite samples though, and we therefore use the finite sample correction these authors suggest, adapting it to the ARW statistic as follows. The finite sample correction in KMS, applied to the AR in (17) consists in modifying $P_{M_{X_1}z}$ in the numerator. When the model contains an intercept, so X_1 is augmented as $X_1 = [\iota : \tilde{X}_1]$, replacing in effect $P_{M_{X_1}z} = M_{X_1}z(z'M_{X_1}z)^{-1}z'M_{X_1}$ with

$$\tilde{P}_{M_X z} = M_X z \left(z'M_X z - \left(1 - \hat{\lambda}_{\varepsilon_1, u_2} \right) T_1 \bar{z}' \bar{z} \right)^{-1} z'M_X$$

where $\widehat{\lambda}_{\varepsilon_1, u_2}$ is the estimated long run correlation between ε_{1t} and u_{2t} . A similar correction for the Wald statistic W obtains. It depends on the specific form of H_0^* but only affects the variance related to the estimator of α_2 in $\widehat{V}_{\hat{\psi}}(b_{12})$. We provide an expression thereof in the supplementary appendix. In the empirical applications we consider in this paper, $\widehat{\lambda}_{\varepsilon_1, u_2}$ is low enough so the finite sample correction does not play a substantive role.

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 $\widehat{Y}_{2t} = Y_{2t-1}^{obs} - \widehat{\gamma}_x t$ in the computation of the IVX instrument Z_t . Whether or not Y_{2t} is stationary affects inference on γ_x . In particular, if $\widehat{\gamma}_x$ is computed using the full sample, then \widehat{Y}_{2t} is a function of future values and this may affect the exclusion restrictions used in the GMM objective function.

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

$$\widehat{Y}_{2t} = Y_{2t}^{obs} - \widehat{\gamma}_x t = Y_{2t}^{obs} + \frac{6}{(t-1)} \sum_{j=1}^t Y_{2j}^{obs} - \frac{12}{(t+1)(t-1)} \sum_{j=1}^t j Y_{2j}^{obs}$$

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

4 Numerical results

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

The data generating process is a bivariate reduced-form VAR(1):

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

with

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

	At 5%				At 10%			
	$\rho = 0.20$		0.95		0.20		0.95	
	<i>AR</i>	<i>t</i>	<i>AR</i>	<i>t</i>	<i>AR</i>	<i>t</i>	<i>AR</i>	<i>t</i>
$c = 0$	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.

and $Y_{10} = Y_{20} = 0$. The estimated model is SVAR(m), $m = 1, 2, 4$, and we focus on testing the null hypothesis $H_0 : b_{12} = 0$ against $H_1 : b_{12} \neq 0$.

4.1 Size

We consider the following parameters sets: $\rho \in \{0.20, 0.95\}$ and $c \in \{0, -1, -10, -30, -100\}$. The sample size is set to $T = 200$. We compute the null rejection frequencies of our AR test with filtered instrument z (17) and the conventional t test with instrument $Y_{2,t-1}$ at the 5% and 10% levels of significance. The number of Monte Carlo replications is 20000. Tables 1 and 2 give the results for SVAR(1) and SVAR(2) models, respectively. We notice that the rejection frequency of the t test can deviate sharply from its asymptotic level, with considerable overrejection in the cases $\rho = 0.2$ and c close to zero. In contrast, the rejection frequency of our proposed AR test is very close to its asymptotic level in all cases. Similar results obtain for higher order models as well as for models with deterministic terms (further results can be found in the Supplementary Appendix).

4.2 Power

We compute (large-sample) power of the AR and t tests of the previous case under weak identification. We set $T = 2000$, use 1000 Monte Carlo replications, and con-

	At 5%				At 10%			
	$\rho = 0.20$		0.95		0.20		0.95	
	<i>AR</i>	<i>t</i>	<i>AR</i>	<i>t</i>	<i>AR</i>	<i>t</i>	<i>AR</i>	<i>t</i>
$c = 0$	0.049	0.006	0.066	0.770	0.098	0.025	0.126	0.802
-1	0.047	0.008	0.060	0.676	0.096	0.029	0.119	0.716
-10	0.045	0.020	0.039	0.258	0.091	0.055	0.080	0.308
-30	0.036	0.035	0.034	0.144	0.078	0.084	0.079	0.186
-100	0.028	0.048	0.052	0.081	0.065	0.100	0.113	0.117

Table 2: Null rejection frequencies of AR (with filtered instruments) and conventional *t* tests of the hypothesis $H_0 : b_{12} = 0$ in a bivariate SVAR(2) with long-run restrictions. ρ is the correlation between the reduced-form VAR errors. The sample size is 200. Number of MC replications: 20000.

sider $\rho \in \{0.2, 0.95\}$ and $c = -10, -100$, and -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.⁶ 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 1 reports the resulting power curves in each case. The figures show that the AR test has good power even for c close to zero. This is not the case of 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.

⁶In 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 .

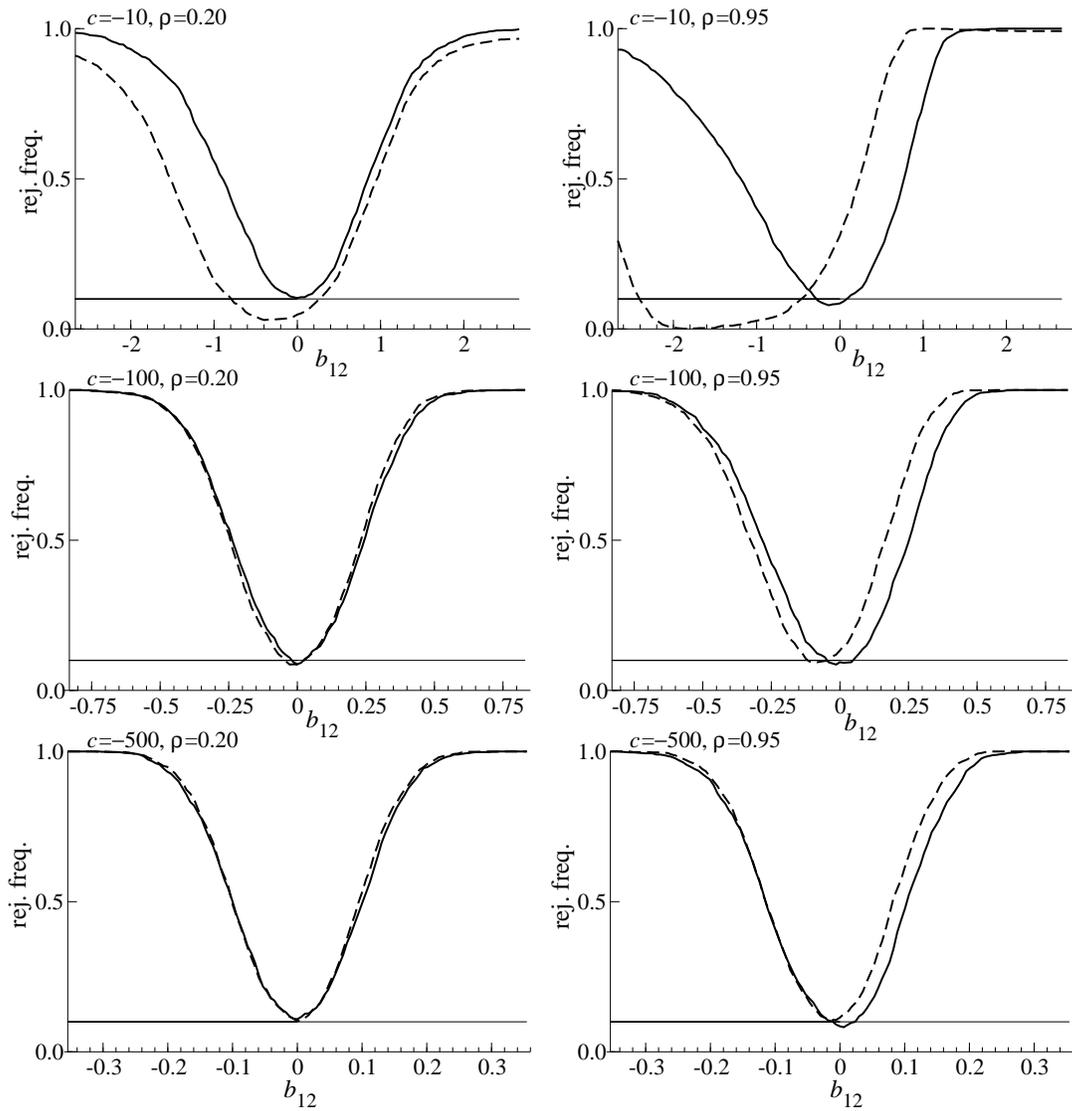


Figure 1: Large-sample power of AR (with filtered instrument) and t tests of the hypothesis $H_0 : b_{12} = 0$ against $H_1 : b_{12} \neq 0$ in the SVAR(1) model with long run restrictions. $T = 2000$, 1000 MC replications, ρ is correlation of reduced-form errors.

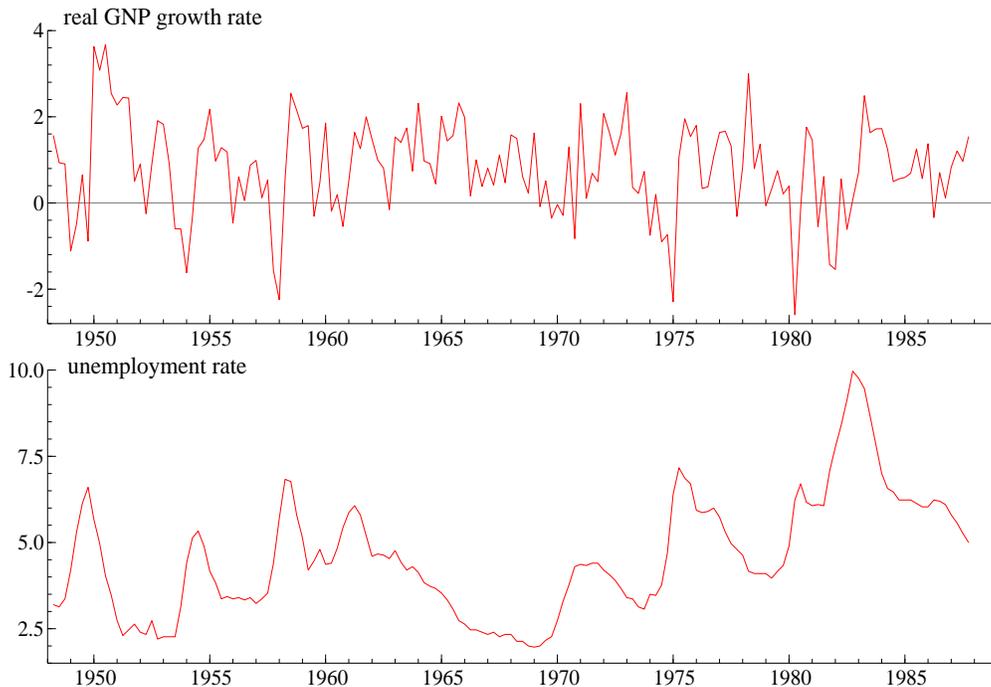


Figure 2: Original data used in Blanchard and Quah (1989).

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 unemployment in deviation from a linear trend. We use the original BQ dataset, which is quarterly and covers the period 1948q2 to 1987q4. More details about the data and transformations are given in the Supplementary Appendix. For comparison with the results reported in BQ, we give here results based on full-sample detrending as in BQ. The results with recursive detrending are given in the Supplementary Appendix, where we also report results based on an extended sample up to 2014q4.

The level specification in BQ is a SVAR(9) with Y_{1t} in first differences and Y_{2t} in levels. The actual BQ data are presented in Figure 2.

Figure 3 reports the estimated IRFs together with the robust confidence bands based on our proposed ARW method and the non-robust confidence bands of BQ. We see that the point estimates from our method that uses the filtered instrument are very

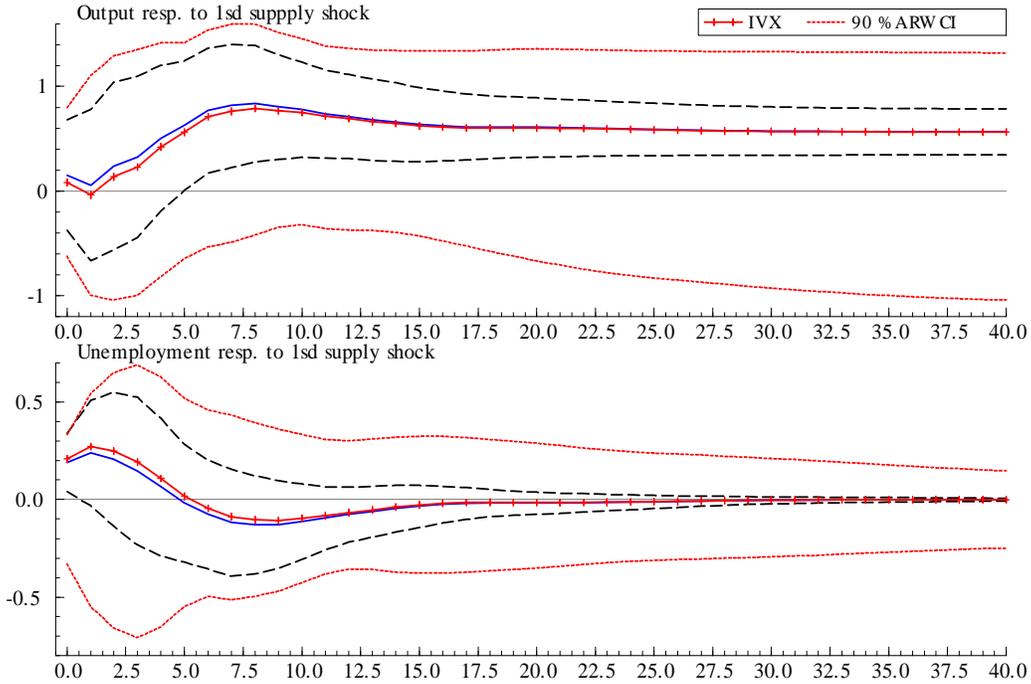


Figure 3: Estimates and confidence intervals of the IRFs. Robust (in red), and Blanchard and Quah (1989) (in blue and black)

close to those of BQ, but 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.

5.2 The hours debate

In the hours debate, Y_{1t} denotes log productivity, and Y_{2t} log hours. We consider the level specifications in Gali (1999) and Christiano *et al* (2003), henceforth CEV. Both use quarterly data to estimate a SVAR(5) with Y_{1t} in first differences and Y_{2t} in levels. Gali uses total hours linearly detrended over the sample 1948q2 to 1994q4. CEV use per capita hours and their sample is 1948q1 to 2002q4. Figure 4 presents the Gali (1999) data.

Figure 5 presents the Gali (1999) estimates and confidence intervals together with their robust version. The robust confidence intervals do not alter Gali's conclusions.

The data used by CEV is presented in Figure 6 and their IRFs together with their robust versions are reported in Figure 7. In the CEV data, the response of hours to a

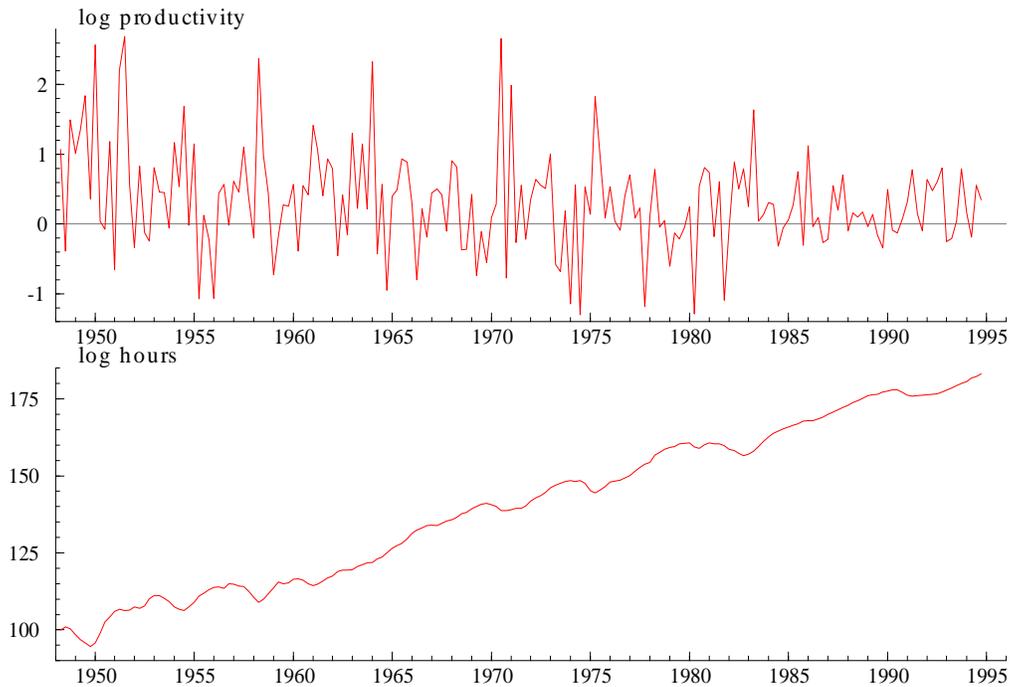


Figure 4: Original data used in Gali (1999).

technology shock is no longer significant. The information in the long-run restriction is so small that the data is consistent with both 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.

Finally, we report results in Figure 8 for the difference specification in Gali (1999). We see that the robust ARW confidence bands are not much wider than the non-robust ones reported in Gali (1999), indicating that this specification did not suffer from weak identification (notwithstanding, of course, the valid CEV critique for using total as opposed to per capita hours, though the difference in the growth rates is relatively small).

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

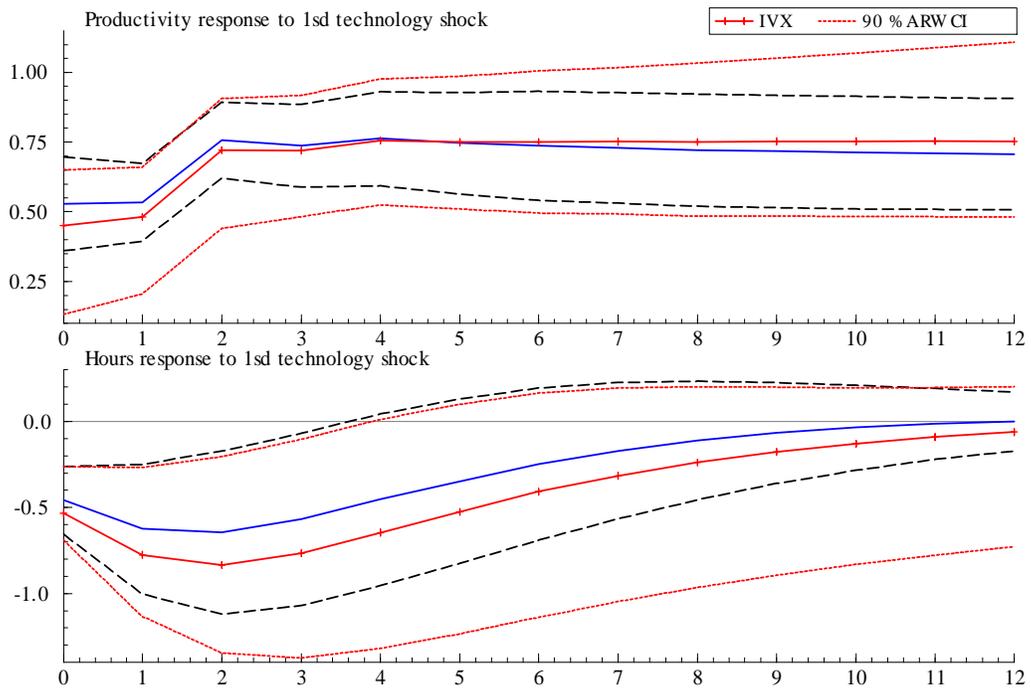


Figure 5: IRFs to technology shock. Robust estimates (in red) together with the Gali (1999) estimates (in blue and black).

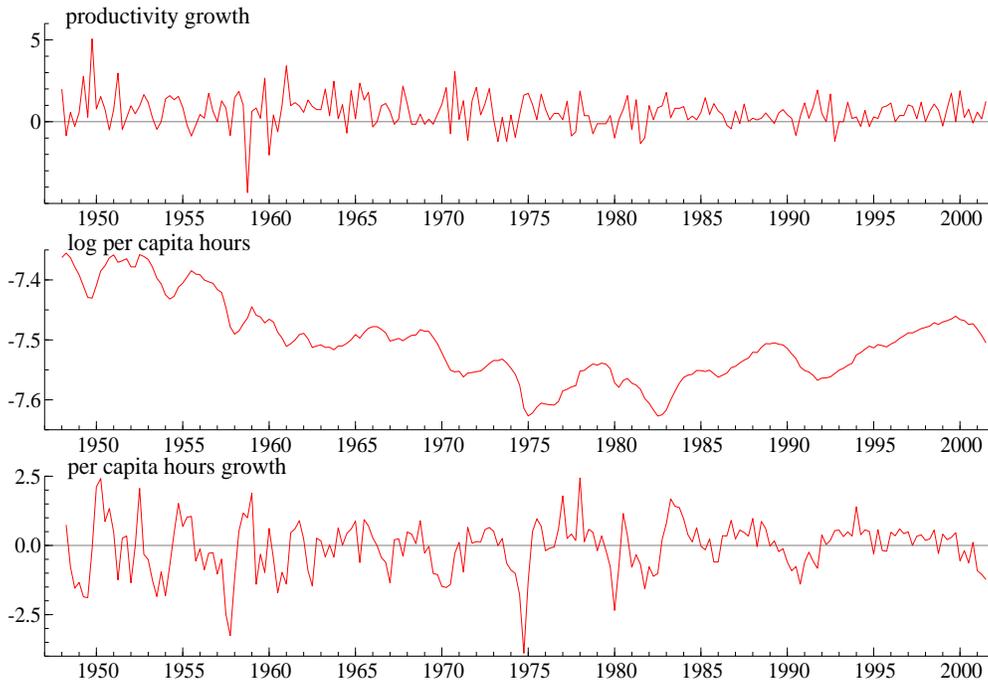


Figure 6: Original data used in Christiano *et al* (2003).

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 Proofs

A.1 Proof of Theorem 1

Notice that, here, the generated instrument z_t is adapted to the filtration $\mathcal{F}_{2,t-1} = \sigma\{\Delta Y_{2,t-j}, j \geq 1\}$ as opposed to the case considered by MP where it is adapted to $\mathcal{F}_{2,t}$. Hence $\text{cov}(z_t, \varepsilon_{1t}) = 0$ and there is no need to estimate this covariance: the condition $b > 2/3$ in MP does not apply here. Also, contrary to MP, the errors $(\varepsilon_{1t}, v_{2t})$ are *i.i.d* so the restriction $b > 1/2$ in Proposition A2 of MP is not required (there is no need

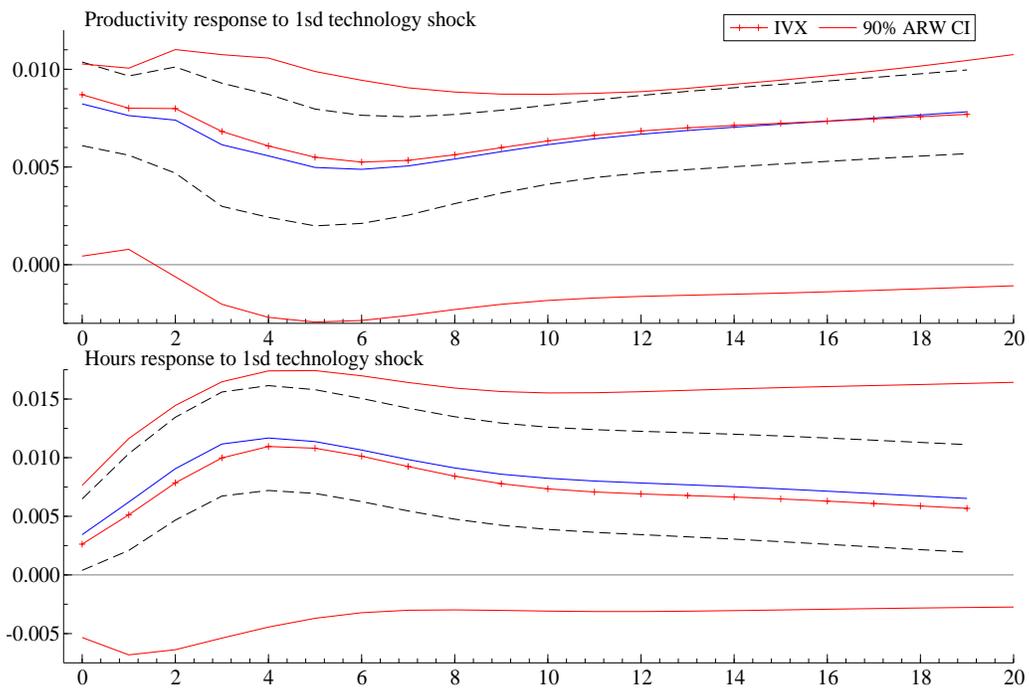


Figure 7: IRFs to technology shock. Robust estimates (in red) together with the CEV estimates (in blue and black).

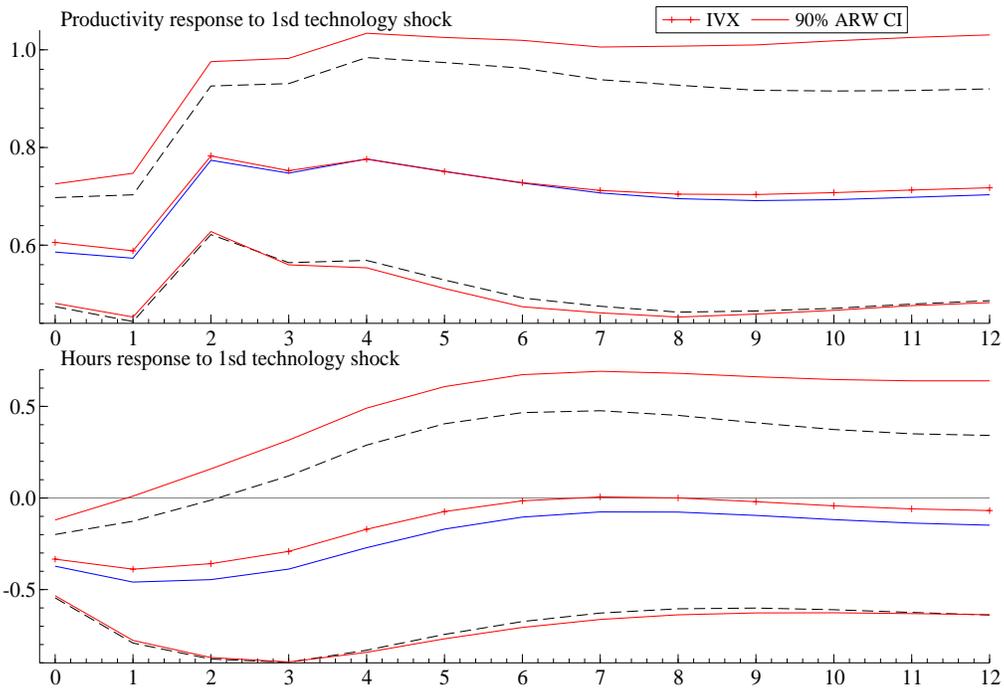


Figure 8: IRFs to technology shock for the difference specification. Robust estimates (in red) together with the Gali (1999) estimates (in blue and black).

for a Beveridge-Nelson decomposition of the errors), hence their lemmas 3.1, 3.5 and 3.6 apply for $b \in (0, 1)$ as in KMS.

We first consider the case $m = 1$ so the numerator of the AR statistic in equation (17), simplifies to

$$(\Delta Y_1 - \Delta Y_2 b_{12})' P_z (\Delta Y_1 - \Delta Y_2 b_{12}) = \sum_{t=1}^T \varepsilon_{1t} z_t \left(\sum_{t=1}^T z_t^2 \right)^{-1} \sum_{t=1}^T z_t \varepsilon_{1t},$$

Define $\Sigma_{u_2} = \text{var}[u_{2t}]$ where u_{2t} is defined in expression (8). The test statistic relates to that of the estimator considered by Kostakis et al (2015, KMS henceforth): $(\sum_t z_t \Delta Y_{2t})^{-1} \sum_t z_t \varepsilon_{1t}$. KMS consider four cases which extend the results of MP: (i) $b < \min(a, 1)$; (ii) $a \in (0, b)$; (iii) $a = b > 0$ and finally (iv) $a = 0$. We denote by g a constant that takes value c_z in case (i) and c in (ii).

In Cases (i) and (ii), Lemmata 3.2 and 3.5 of MP imply that

$$T^{-\frac{1+\min(a,b)}{2}} \sum_{t=1}^T z_t \varepsilon_{1t} \implies \mathbf{N} \left(0, \frac{1}{-2g} \Sigma_{u_2} \sigma_{\varepsilon_1}^2 \right).$$

Also, Lemmata 3.1 and 3.5 of MP imply $T^{-(1+\min(a,b))} \sum_{t=1}^T z_t^2 \xrightarrow{p} \frac{1}{-2g} \Sigma_{u_2}$. In Case (iii), Lemma 3.6(iii) of MP implies

$$T^{-\frac{1+a}{2}} \sum_{t=1}^{\lfloor Ts \rfloor} z_t \varepsilon_{1t} \implies \mathbf{N} \left(0, \frac{1}{-2(c+c_z)} \Sigma_{u_2} \sigma_{\varepsilon_1}^2 \right)$$

and $T^{-(1+a)} \sum_{t=1}^T z_t^2 \xrightarrow{p} \frac{c}{c+c_z} \Sigma_{u_2}$. Finally in Case (iv), Lemma A.2(ii) and (iii) of KMS imply that $(\sum_t z_t^2)^{-1} \sum_t z_t \varepsilon_{1t}$ behaves as the estimator of their Theorem 1(iv) :

$$\left(\sum_t z_t^2 \right)^{-1} \sum_t z_t \varepsilon_{1t} \implies \mathbf{N} \left(0, E(\Delta Y_{2t}^2)^{-1} \sigma_{\varepsilon_1}^2 \right)$$

where $T^{-1} \sum_t z_t^2 \xrightarrow{p} E(\Delta Y_{2t}^2)$ from their Lemma A.2(ii). Collecting the four cases above yields:

$$\sigma_{\varepsilon_1}^{-2} \sum_t \varepsilon_{1t} z_t \left(\sum_t z_t^2 \right)^{-1} \sum_t z_t \varepsilon_{1t} \implies \chi_1^2. \quad (22)$$

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 $O_p(1)$ (this is also the case when $k > 1$), 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 $m = 1$. Notice that we can allow for $b \in (0, 1)$ here

We now consider extending the result above to $m > 1$, which involves $X_{1t} = [\Delta Y_{t-1}, \dots, \Delta Y_{t-(k-1)}]$. We show in a supplementary appendix that

$$(i) \sum_{t=1}^T X'_{1t} z_t = O_p(T); \quad (ii) \sum_{t=1}^T X'_{1t} \varepsilon_{1t} = O_p(T^{1/2}); \quad (iii) \sum_{t=1}^T X'_{1t} X_{1t} = O_p(T). \quad (23)$$

Hence if $a > 0$, the numerator $(\Delta Y_1 - \Delta Y_2 b_{12})' P_{M_{X_1} z} (\Delta Y_1 - \Delta Y_2 b_{12})$,

$$\begin{aligned} & \left(\sum_{t=m}^T \varepsilon_{1t} z_t - \sum_{t=m}^T \varepsilon_{1t} X_{1t} \left(\sum_{t=m}^T X'_{1t} X_{1t} \right)^{-1} \sum_{t=m}^T X'_{1t} z_t \right) \\ & \times \left[\sum_{t=m}^T z_t^2 - \sum_{t=m}^T z_t X_{1t} \left(\sum_{t=m}^T X'_{1t} X_{1t} \right)^{-1} \sum_{t=m}^T X'_{1t} z_t \right]^{-1} \\ & \times \left(\sum_{t=m}^T z_t \varepsilon_{1t} - \sum_{t=m}^T z_t X_{1t} \left(\sum_{t=m}^T X'_{1t} X_{1t} \right)^{-1} \sum_{t=m}^T X'_{1t} \varepsilon_{1t} \right) \end{aligned} \quad (24)$$

behaves as in (22) since the correction for the lags is of lower magnitude. The denominator of the AR statistic still converges to $\sigma_{\varepsilon_1}^2$ in probability.

When $a = 0$, Lemma A.2 of KMS shows that expression (24) is asymptotically equivalent to

$$(\Delta Y_1 - \Delta Y_2 b_{12})' P_{M_{X_1} Y_2} (\Delta Y_1 - \Delta Y_2 b_{12}) = \varepsilon'_1 M_{X_1} Y_2 (Y'_2 M_{X_1} Y_2)^{-1} Y'_2 M_{X_1} \varepsilon_1,$$

where Y_2 denotes the stacked (Y_{2t-1}) . Since $E[\varepsilon_{1t} Y_{2t-1}] = 0$ and Y_{2t-1} is stationary, it

follows that

$$\sigma_{\varepsilon_1}^{-2} \varepsilon_1' M_{X_1} Y_2 (Y_2' M_{X_1} Y_2)^{-1} Y_2' M_{X_1} \varepsilon_1 \xrightarrow{d} \chi_1^2.$$

Again the same result as above holds for the denominator.

A.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 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, $r: \Theta \rightarrow \mathfrak{R}^q$, $q \leq p_\psi$. 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

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

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:

$$\hat{\psi}(\vartheta) = \arg \min_{\psi \in \Theta_2} F_T(\vartheta, \psi)' \hat{V}_f(\vartheta, \tilde{\psi})^{-1} F_T(\vartheta, \psi).$$

To simplify notation, let $\hat{\psi} \equiv \hat{\psi}(\vartheta)$, $\hat{r}(\vartheta) = r(\vartheta, \hat{\psi})$, $\hat{R}(\vartheta) = R(\vartheta, \hat{\psi})$, $\tilde{V}_f(\vartheta) = \hat{V}_f(\vartheta, \tilde{\psi})$, $\hat{F}_T(\vartheta) = F_T(\vartheta, \hat{\psi})$ and $\hat{J}_T(\vartheta) = J_T(\vartheta, \hat{\psi})$. Also, let $\hat{C}(\vartheta)$ be a 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

$$\begin{aligned}
\hat{S}_T(\vartheta) &= S_T(\vartheta, \hat{\psi}) = \hat{F}_T(\vartheta)' \tilde{V}_f(\vartheta)^{-1} \hat{F}_T(\vartheta), \\
W_r(\vartheta) &= \hat{r}(\vartheta)' \left[\hat{R}(\vartheta) \hat{V}_{\hat{\psi}}(\vartheta) \hat{R}(\vartheta)' \right]^{-1} \hat{r}(\vartheta), \text{ and} \\
\hat{V}_{\hat{\psi}}(\vartheta) &= \left[\hat{J}_T(\vartheta)' \tilde{V}_f(\vartheta)^{-1} \hat{J}_T(\vartheta) \right]^{-1}.
\end{aligned} \tag{25}$$

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

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

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

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

(iv) *there exist a non-stochastic $p_{\psi} \times p_{\psi}$ symmetric matrix $B_T \rightarrow 0$ such that $B_T^{-1} \hat{V}_{\hat{\psi}}(\vartheta) B_T^{-1} \implies \Psi$ full-rank a.s., and (v) any stochastic elements in Ψ are independent of $\xi = (\xi_1', \xi_2')'$. Then, $ARW(\vartheta) \implies \chi_{p_{\vartheta}+q}^2$.*

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, $\Lambda_T \rightarrow 0$ 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,

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

and

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

by Assumption (iv). Let $\Psi^{1/2}$ denote the Choleski factor of Ψ , such that $\Psi^{1/2}\Psi^{1/2'} = \Psi$. Assumption (iv) implies that $B_T^{-1'}\hat{C}_{\hat{\psi}}(\vartheta)^{-1'} \implies \Psi^{1/2}$. Assumption (iii) then implies

$$\begin{aligned} B_T^{-1'}(\hat{\psi} - \psi) &= B_T^{-1'}\hat{C}_{\hat{\psi}}(\vartheta)^{-1'}\hat{C}_{\hat{\psi}}(\vartheta)'(\hat{\psi} - \psi) \\ &= \Psi^{1/2}\xi_2 + o_p(1). \end{aligned}$$

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

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

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

$$\Lambda_T^{-1}Q_T'\hat{r}(\vartheta) = U_T'\Psi^{1/2}\xi_2 + o_p(1).$$

Moreover,

$$\begin{aligned} \hat{r}(\vartheta)' \left[\hat{R}(\vartheta) \hat{V}_{\hat{\psi}}(\vartheta) \hat{R}(\vartheta)' \right]^{-1} \hat{r}(\vartheta) &= \hat{r}(\vartheta)' Q_T \Lambda_T^{-1} \left[\Lambda_T^{-1} Q_T' \hat{R}(\vartheta) \hat{V}_{\hat{\psi}}(\vartheta) \hat{R}(\vartheta)' Q_T \Lambda_T^{-1} \right]^{-1} \Lambda_T^{-1} Q_T' \hat{r}(\vartheta) \\ &= \xi_2' \Psi^{1/2'} U_T [U_T' \Psi U_T]^{-1} U_T' \Psi^{1/2} \xi_2 + o_p(1). \end{aligned}$$

Combining these results and using the continuous mapping theorem we have

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

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

Comments 1. It is straightforward to verify the conditions of the theorem in ‘standard’ GMM settings where $B_T = T^{-1/2} I_{p_\psi}$, the limit variance of the moment vector, Ψ , is nonstochastic, and $\sqrt{T} \hat{F}_T(\theta) \xrightarrow{d} N(0, \Psi)$.

A.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 (7) other than b_{12} , namely δ_1 and $\sigma_{\varepsilon_1}^2$, and ψ_2 are the parameters that appear only in (8), i.e., α_2 , δ_2 and d_{21} . Because \hat{V}_f is block diagonal (due to the orthogonality of the errors), 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*

$$\begin{aligned}\hat{\psi}_1 &= \begin{pmatrix} (X_1'X_1)^{-1} X_1' (\Delta Y_1 - \Delta Y_2 b_{12}) \\ T^{-1} \hat{\varepsilon}_1' \hat{\varepsilon}_1 \end{pmatrix}, \\ \hat{\psi}_2 &= \left(\hat{Z}_2' \hat{X}_2 \right)^{-1} \hat{Z}_2' \Delta Y_2,\end{aligned}\tag{26}$$

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

$$\hat{V}_{\hat{\psi}} = \begin{pmatrix} V_{\hat{\psi},11} & 0 & V_{\hat{\psi},13} \\ 0 & \hat{\omega} & 0 \\ V_{\hat{\psi},13}' & 0 & V_{\hat{\psi},33} \end{pmatrix},$$

where

$$\begin{aligned}\hat{V}_{\hat{\psi},11} &= (X_1'X_1)^{-1} \hat{\sigma}_{\varepsilon_1}^2, \\ \hat{V}_{\hat{\psi},13} &= (X_1'X_1)^{-1} X_1' \hat{Z}_2 \left(\hat{X}_2' \hat{Z}_2 \right)^{-1} \hat{\sigma}_{\varepsilon_1}^2 d_{21}, \\ \hat{V}_{\hat{\psi},33} &= \left(\hat{Z}_2' \hat{X}_2 \right)^{-1} \left(\hat{Z}_2' \hat{Z}_2 \hat{\sigma}_{\varepsilon_2}^2 + \hat{Z}_2' P_{X_1} \hat{Z}_2 \hat{\sigma}_{\varepsilon_1}^2 d_{21}^2 \right) \left(\hat{X}_2' \hat{Z}_2 \right)^{-1},\end{aligned}\tag{27}$$

$\hat{\sigma}_{\varepsilon_1}^2 = T^{-1} \hat{\varepsilon}_1' \hat{\varepsilon}_1$, $\hat{\omega}$ is a consistent estimator of $\text{var}(\hat{\sigma}_{\varepsilon_1}^2)$, $\hat{\sigma}_{\varepsilon_2}^2 = T^{-1} \hat{\varepsilon}_2' \hat{\varepsilon}_2$ and $\hat{\varepsilon}_2 =$

$\Delta Y_2 - \hat{X}_2 \hat{\psi}_2$. The standardized random vector $\hat{\xi}$ defined in Theorem 3 is given by

$$\hat{\xi}_1 = (z' M_{X_1} z)^{-1/2} \hat{\sigma}_{\varepsilon_1}^{-1} z' M_{X_1} \varepsilon_1, \text{ and} \quad (28)$$

$$\hat{\xi}_2 = \begin{pmatrix} (X_1' X_1)^{-1/2} X_1' \varepsilon_1 \hat{\sigma}_{\varepsilon_1}^{-1} \\ T^{1/2} \hat{\omega}^{-1/2} (\hat{\sigma}_{\varepsilon_1}^2 - \sigma_{\varepsilon_1}^2) \\ (\hat{X}_2' P_{\hat{Z}_2} \hat{X}_2)^{1/2} \hat{\sigma}_{\varepsilon_2}^{-1} (\hat{Z}_2' \hat{X}_2)^{-1} \hat{Z}_2' \varepsilon_2 \end{pmatrix}. \quad (29)$$

It is straightforward to establish the following.

Proposition 5. (i) $\tilde{\psi} = \hat{\psi}$, and (ii) $\hat{\psi}_1 \xrightarrow{p} \psi_1$.

We will now proceed by giving the proof of the theorem in each of the following three cases. Case 1: $0 \leq a < b$, case 2: $a = b$; case 3: $a > b$. Recall that $b \in (1/2, 1)$ holds throughout this theorem.

Case 1. $0 \leq a < b$ We will first show that, in this case, the ARW statistic (19) is asymptotically equivalent to the ARW statistic that is obtained by replacing the filtered instrument z_{t-1} by the standard instrument $Y_{2,t-1}$. This is a consequence of the following result and the continuous mapping theorem.

Proposition 6. If $0 \leq a < b$ and $b \in (1/2, 1)$, then

- (i) $T^{-(1+a)} \sum_{t=1}^T z_{t-1}^2 = T^{-(1+a)} \sum_{t=1}^T Y_{2,t-1}^2 + o_p(1)$;
- (ii) $T^{-(1+a)} \sum_{t=1}^T z_{t-1} Y_{2,t-1} = T^{-(1+a)} \sum_{t=1}^T Y_{2,t-1}^2 + o_p(1)$;
- (iii) $T^{-\frac{1+a}{2}} \sum_{t=1}^T z_{t-1} \varepsilon_t = T^{-\frac{1+a}{2}} \sum_{t=1}^T Y_{2,t-1} \varepsilon_t + o_p(1)$;
- (iv) $T^{-\frac{1+a}{2}} \sum_{t=1}^T z_{t-1} \Delta Y_{t-i} = T^{-\frac{1+a}{2}} \sum_{t=1}^T Y_{2,t-1} \Delta Y_{t-i} + o_p(1)$, $i = 1, \dots, m-1$.

Specifically, define

$$AR^*(b_{12}) = \frac{(\Delta Y_1 - \Delta Y_2 b_{12})' P_{M_{X_1} Y_2} (\Delta Y_1 - \Delta Y_2 b_{12})}{(\Delta Y_1 - \Delta Y_2 b_{12})' M_{Z_1^*} (\Delta Y_1 - \Delta Y_2 b_{12}) / (T - \text{col}(Z_1^*))} \quad (30)$$

where $Z_1^* = (X_1, Y_2)$, $\hat{\psi}^*(b_{12}) = (\hat{\psi}_1(b_{12})', \hat{\psi}_2^{*'})'$, with

$$\hat{\psi}_2^* = (\hat{X}_2' \hat{X}_2)^{-1} \hat{X}_2' \Delta Y_2, \quad (31)$$

$$\hat{V}_{\hat{\psi}}^*(b_{12}) = \begin{pmatrix} \hat{V}_{\hat{\psi},11} & 0 & \hat{V}_{\hat{\psi},13}^* \\ 0 & \hat{\omega} & 0 \\ \hat{V}_{\hat{\psi},13}^{*'} & 0 & \hat{V}_{\hat{\psi},33}^* \end{pmatrix}, \quad (32)$$

$$\begin{aligned}\hat{V}_{\hat{\psi},13}^* &= (X_1'X_1)^{-1} X_1'\hat{X}_2 \left(\hat{X}_2'\hat{X}_2\right)^{-1} \hat{\sigma}_{\varepsilon_1}^2 d_{21}, \\ \hat{V}_{\hat{\psi},33}^* &= \left(\hat{X}_2'\hat{X}_2\right)^{-1} \left(\hat{X}_2'\hat{X}_2\hat{\sigma}_{\varepsilon_2}^2 + \hat{X}_2'P_{X_1}\hat{X}_2\hat{\sigma}_{\varepsilon_1}^2 d_{21}^2\right) \left(\hat{X}_2'\hat{X}_2\right)^{-1}, \\ W^*(b_{12}) &= r \left(b_{12}, \hat{\psi}^*(b_{12})\right)' \hat{V}_{\hat{r}}^*(b_{12})^{-1} r \left(b_{12}, \hat{\psi}^*(b_{12})\right), \\ \text{where } \hat{V}_{\hat{r}}^*(b_{12}) &= R \left(b_{12}, \hat{\psi}^*(b_{12})\right)' \hat{V}_{\hat{\psi}}^*(b_{12}) R \left(b_{12}, \hat{\psi}^*(b_{12})\right),\end{aligned}$$

and

$$ARW^*(b_{12}) = AR^*(b_{12}) + W^*(b_{12}).$$

We then have the following result.

Proposition 7. *If $0 \leq a < b$ and $b \in (1/2, 1)$, then $ARW(b_{12}) - ARW^*(b_{12}) = o_p(1)$.*

We can complete the discussion of this case by working out the limiting distribution of $ARW^*(b_{12})$. For $a = 0$, the result follows straightforwardly, see comment 1 of Theorem 3. So, it remains to show the result for $0 < a < \beta$.

Proposition 8. *If $0 < a < b$ and $b \in (1/2, 1)$, then*

$$B_T^{-1} \hat{V}_{\hat{\psi}}^*(b_{12}) B_T^{-1} \xrightarrow{p} \Psi, \quad (33)$$

where

$$B_T = \begin{pmatrix} T^{-1/2} I_{p_{\psi_1}} & 0 & 0 \\ 0 & T^{-\frac{1+a}{2}} & 0 \\ 0 & 0 & T^{-1/2} I_{p_{\psi_2}-1} \end{pmatrix} \quad (34)$$

and Ψ is a non-stochastic matrix of full-rank p_{ψ} ; and $\hat{\xi}^* = \begin{pmatrix} \hat{\xi}_1^* \\ \hat{\xi}_2^* \end{pmatrix}$, where $\hat{\xi}_1$ is defined in (28) and

$$\hat{\xi}_2^* = \begin{pmatrix} (X_1'X_1)^{-1/2} X_1'\varepsilon_1 \hat{\sigma}_{\varepsilon_1}^{-1} \\ T^{1/2} \hat{\omega}^{-1/2} (\hat{\sigma}_{\varepsilon_1}^2 - \sigma_{\varepsilon_1}^2) \\ \left(\hat{X}_2'\hat{X}_2\right)^{-1/2} \hat{X}_2'\varepsilon_2 \hat{\sigma}_{\varepsilon_2}^{-1} \end{pmatrix}$$

satisfies

$$\hat{\xi}^* \xrightarrow{d} N(0, I_k).$$

Proposition satisfies the remaining Assumptions (iii)-(v) in Theorem 3, and so the result follows.

Case 2. $a = b$ MP Lemma 3.6 implies

$$T^{-(1+a)} \sum_{t=1}^T z_{t-1} Y_{2,t-1} \xrightarrow{p} \frac{c\omega}{c_z + c}$$

$$T^{-(1+a)} \sum_{t=1}^T z_{t-1}^2 \xrightarrow{p} \frac{c\omega}{c_z + c},$$

where ω is a positive constant relating to the long run variance of ΔY_{2t} (see the proof of Proposition 8 in the Supplement). Similarly, using similar arguments as in Proposition 8 for $\sum_{t=1}^T Y_{2,t-1} X_{it}$, it can be shown that

$$T^{-(1+\frac{a}{2})} \sum_{t=1}^T z_{t-1} X_{it} = O_p(T^{-a/2}). \quad (35)$$

Thus, it follows that

$$B_T^{-1} \hat{V}_{\hat{\psi}}(b_{12}) B_T^{-1} \xrightarrow{p} \Psi,$$

where B_T is given in (34) and Ψ is nonstochastic, satisfying Assumptions (iv) and (v) of Theorem 3. Finally, Assumption (iii) of that theorem are verified using MP Lemma 3.6(iii) and the arguments in the proof of Proposition 8, completing the proof for this case.

Case 3. $a > b$ First, we establish the counterpart of Proposition 6 for this case.

Proposition 9. *If $1/2 < b < a$, then*

(i) $T^{-(1+b)} \sum_{t=1}^T z_{t-1}^2 \xrightarrow{p} \frac{\omega}{-2c_z};$

(ii)

$$T^{-(1+b)} \sum_{t=1}^T z_{t-1} Y_{2,t-1} \implies \begin{cases} - \left(\int_0^1 W dW + 1 \right) \omega c_z^{-1} & \text{if } c = 0 \text{ or } a > 1 \\ - \left(\int_0^1 J_c dW + 1 \right) \omega c_z^{-1} & \text{if } c < 0 \text{ and } a = 1 \\ -\omega c_z^{-1} & \text{if } c < 0 \text{ and } a < 1, \end{cases} \quad (36)$$

where W is a standard Brownian motion and $J_c(s) = \int_0^s e^{c(s-r)} dW(r)$ is the associated Ornstein-Uhlenbeck process with parameter c .

(iii) $T^{-\frac{1+b}{2}} \sum_{t=1}^T z_{t-1} \varepsilon_t \implies \tilde{W}$, where \tilde{W} is a Brownian motion independent of W .

(iv) $T^{-(1+\frac{b}{2})} \sum_{t=1}^T z_{t-1} \Delta Y_{t-i} = O_p(T^{-b/2})$, $i = 1, \dots, m-1$.

Define the martingale difference array

$$\zeta_{Tt} = \begin{pmatrix} T^{-(1+b)/2} z_{t-1} \varepsilon_t \\ T^{-1/2} X_{1t} \varepsilon_{1t} \\ T^{-1/2} X_{2t} \varepsilon_{2t} \\ T^{-1/2} (\varepsilon_{1t}^2 - \sigma_\varepsilon^2) \\ T^{-1/2} \varepsilon_{1t} \varepsilon_{2t} \end{pmatrix} = \begin{pmatrix} T^{-(1+b)/2} & 0 \\ 0 & T^{-1/2} I_{k-1} \end{pmatrix} f_t(\theta), \quad (37)$$

where $f_t(\theta)$ is the moment function in the notation of section A.2 evaluated at the true value of the parameters, so that

$$\sum_{t=1}^T \zeta_{Tt} = \begin{pmatrix} T^{(1+b)/2} & 0 \\ 0 & T^{1/2} I_{k-1} \end{pmatrix} F_T(\theta).$$

By exactly the same arguments as in the proof of Proposition 8, it follows that

$$\sum_{t=1}^T \zeta_{Tt} \implies \zeta \sim N(0, V_\zeta)$$

where V_ζ is a nonstochastic matrix, and ζ is independent of W in Proposition 9 by MP Lemmas 3.1 and 3.2. The conditions of Theorem 3 can then be verified by application of Slutsky's Lemma and the continuous mapping theorem.

Finally, the presence of an intercept in the SVAR results in all the variables in the above equations being demeaned using their sample means. This can be easily handled by replacing W and J_c in (36) by their demeaned versions $\underline{W} = W - \int_0^1 W(s) ds$ and $\underline{J}_c = J_c - \int_0^1 J_c(s) ds$, respectively. The presence of an intercept has no effect on the distribution of ζ_{Tt} , following the same arguments as in the proof of KMS Theorem A.

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.
- 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.
- 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.
- Dufour, J.-M. and M. Taamouti (2005). Projection-based statistical inference in linear structural models with possibly weak instruments. *Econometrica* 73(4), 1351–1365.
- 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.
- Gali, 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 & Economic Statistics* 28(1), 1–11.
- 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.
- Mikusheva, A. (2012). One-dimensional inference in autoregressive models with the potential presence of a unit root. *Econometrica* 80(1), 173–212.

- 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 J. C. Robertson (1998). Structural models of the liquidity effect. *The Review of Economics and Statistics* 80(2), 202–217.
- Phillips, P. C. B. (2014). On confidence intervals for autoregressive roots and predictive regression. *Econometrica* 82(3), 1177–1195.
- 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.