Robust inference in structural VARs with long-run restrictions

Guillaume Chevillon
ESSEC and CREST

Sophocles Mavroeidis*
University of Oxford

Zhaoguo Zhan
Tsinghua University

March 16, 2015

PRELIMINARY AND INCOMPLETE

Abstract

Long-run restrictions (Blanchard and Quah, AER 1989) are a very popular method for identifying structural vector autoregressions (SVARs). A prominent example is the unsettled debate on the effect of technology shocks on employment, which has been used to test real business cycle theory (Gali, AER 1999, Christiano et al, 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

*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

degree of persistence of the instruments is high (Pagan and Robertson, RES 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 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 (Phillips and Magdalinos, 2009, Phillips, ECMA 2014). Our proposed method has good size and power properties and can be used to produce robust confidence bands on the impulse response function to the identified structural shock(s).
Keywords: SVARs, identification, weak instruments, near unit roots.

JEL: C12, C32, E32

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 has good power relative to (non-robust) tests that are asymptotically efficient under

[1]Note 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.

strong 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).

Long-run restrictions are a now a very common approach to the identification of SVARs. At the time of writing, Blanchard and Quah (1989) had 4326 Google scholar 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 is

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

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, $var(\varepsilon_t)$ is a n-dimensional diagonal

[2] 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.

variance matrix, and B_0 has ones along its diagonal. We follow Magdalinos and Phillips (2009) and 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 zero mean and diagonal variance matrix Σ satisfying $\Sigma > 0$ and the moment condition $E\|\varepsilon_1\|^4 < \infty$.

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

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

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) regraphic of the long-run restrictions in Pagan and Robertson (1998).

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 r can be 0 (two permanent shocks) or 1 (one permanent shock).

2.1.1 Case of 1 permanent shock ($r = 1$)

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

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

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 regraphic is:

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

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 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 $r = 1$ 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:

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

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

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

where $\tilde{\gamma}_{11}(L) = \gamma_{11}(L) + b_{12}\gamma_{21}(L)$ and $\tilde{\gamma}_{12}(L) = \gamma_{12}(L) + b_{12}\left[\gamma_{22}(L) - 1\right]$. Observe that the error correction ('ecm') term $\beta' Y_{t-1}$ is missing from (1), so we can use this to instrument for the endogenous regressor ΔY_{2t}. Once ε_{1t} is identified from (1), the impact of ε_{1t} on Y_{2t} can be obtained from the regression

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

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

2.1.2 Case of two permanent shocks ($r = 0$)

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-Nelson (1981) 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

$$\tilde{\gamma}_{11}(L)\,\Delta Y_{1t} = \gamma_{12}^{*}(L)\,\Delta^2 Y_{2t} + \varepsilon_{1t}, \tag{3}$$

and the equation for Y_{2t} reads

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

Thus, we are using $\Delta Y_{2,t-1}$ as an instrument for the endogenous regressor $\Delta^2 Y_{2t}$ in (3). Identification is weak if ΔY_{2t} is nearly I(1).

2.2 The hours debate

The number of permanent shocks $r = 0$ versus $r = 1$, can make a big impact on the results. The debate of the short-run effect of a technology shock on hours between Gali (1999) and Christiano et al, is based on a SVAR that contains productivity and hours. Gali used a VAR in first differences ($r = 0$), found a negative effect and rejected RBC theory. Christiano et al favored a VAR with hours in levels ($r = 1$, $\beta = (0, 1)'$) 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. Who is right?

It is true that the level specification nests the difference: $r = 1$ is more general than $r = 0$. Consider the following encompassing regraphic:

$$\begin{aligned} \tilde{\gamma}_{11}(L)\,\Delta Y_{1t} &= \tilde{\gamma}_{12}^{*}(L)\,\Delta^2 Y_{2t} + [b_{12} + \tilde{\gamma}_{12}(1)]\,\Delta Y_{2t} + \varepsilon_{1t} \\ \gamma_{22}(L)\,\Delta Y_{2t} &= \alpha_2 Y_{2,t-1} + \gamma_{21}(L)\,\Delta Y_{1t} + u_{2t}, \end{aligned} \tag{4}$$

The level specification imposes no extra restriction, and uses $Y_{2,t-1}$ as IV in (4). The difference specification imposes $b_{12} + \tilde{\gamma}_{12}(1) = \alpha_2 = 0$, which enables us to use ΔY_{t-1} as IV in (4). This will be misspecified if $b_{12} + \tilde{\gamma}_{12}(1) \neq 0$. Since $b_{12} + \tilde{\gamma}_{12}(1)$ can be anything, it may result in large bias. This corroborates Christiano et al's claim.

In principle, the above misspecification is detectable by a suitable diagnostic test. However, the power of such a test really depends on the value of $\alpha_2 \neq 0$. If α_2 is far from zero, we can reject $\alpha_2 = 0$ with high probability. But if α_2 is close to zero, we will not be able to reject $\alpha_2 = 0$ often. Yet, the bias that results if we incorrectly impose $\alpha_2 = 0$ depends on the true value of $b_{12} + \tilde{\gamma}_{12}(1)$, and can be arbitrarily large. But if we are in the second scenario, $\alpha_2 \to 0$ imposes a second unit root in the system and the level specification suffers from weak identification. Therefore, it is possible that the sampling uncertainty in the level specification is so large that we cannot rule out conclusions drawn using the difference specification. If this turns out to be the case in the hours debate, it would be a pyrrhic victory for Christiano et al.

3 Econometric Methods

3.1 GMM estimating equations

Consider the multivariate SVAR(k) 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 (1) and (2). Let θ denote all the parameters of the model.Moreover, let $X_{1t} = X_{2t} = \left(\Delta Y_{t-1}', ..., \Delta Y_{t-k+1}'\right)'$, so that (1) and (2) can be written compactly as

$$\Delta Y_{1t} = b_{12}' \Delta Y_{2t} + \delta_1' X_{1t} + \varepsilon_{1t} \tag{5}$$

$$\Delta Y_{2t} = \alpha_2 Y_{2,t-1} + \delta_2' X_{2t} + \underbrace{d_{21}\varepsilon_{1t} + v_{2t}}_{u_{2t}}, \tag{6}$$

where δ_1 denotes the coefficients on exogenous and predetermined variables in (1), and δ_2 denotes the corresponding coefficients in (2).[3] Note that $v_{2,t}$ 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 $n > 2$.

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

$$h_{1t}(\theta_1) = b_{12}' \Delta Y_{2t} + \delta_1' X_{1t}, \tag{7}$$

[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.

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

$$h_{2t}(\theta) = \Delta Y_{2t} - \alpha_2 Y_{2,t-1} - \delta'_2 X_{1t} - d_{21} h_{1t}(\theta_1). \tag{8}$$

Note that (7) and (8) 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

$$\underset{nk\times 1}{f_{1t}(\theta_1)} = \begin{pmatrix} Z'_{1t} h_{1t}(\theta_1) \\ h_{1t}(\theta_1)^2 - \sigma^2_{\varepsilon_1} \end{pmatrix}, \quad \underset{nk(n-1)\times 1}{f_{2t}(\theta)} = \begin{pmatrix} h_{1t}(\theta) h_{2t}(\theta) \\ (I_{n-1} \otimes Z'_{2t}) h_{2t}(\theta) \end{pmatrix}, \tag{9}$$

Let $V_{ff}(\theta) = var\left(T^{-1/2} \sum_{t=1}^T f_t(\theta)\right) = var(f_t(\theta))$, which follows from the finite-order VAR assumption – no HAC estimator needed. The GMM criterion is

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

where $F_T(\theta) = \sum_{t=1}^T f_t(\theta)$, and $\bar{\theta}$ is an initial estimator of θ. 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:

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

where $J(\theta) = p\lim_{T\to\infty} T^{-1} \partial F_T(\theta) / \partial\theta'$ (the simplification follows from just-identification). However, under near-unit-root asymptotics, $\hat{\theta}$ is not asymptotically normal.

Orthogonality of the errors and some additional fourth moment assumption imply that the variance matrix of f_t is block diagonal, with $V_1(\theta_1) = var(f_{1t}(\theta_1))$ and $V_2(\theta) = var(f_{2t}(\theta))$, so the GMM criterion function can be decomposed into orthogonal components

$$\begin{aligned} S_T(\theta) &= \frac{1}{T}\left(F_1(\theta_1)', F_2(\theta)'\right)\begin{pmatrix}\hat{V}_1^{-1} & 0\\ 0 & \hat{V}_2^{-1}\end{pmatrix}\begin{pmatrix}F_1(\theta_1)\\ F_2(\theta)\end{pmatrix} \\ &= \underbrace{\frac{1}{T}F_1(\theta_1)'\hat{V}_1^{-1}F_1(\theta_1)}_{S_{1,T}(\theta_1)} + \underbrace{\frac{1}{T}F_2(\theta)'\hat{V}_2^{-1}F_2(\theta)}_{S_{2,T}(\theta)}. \end{aligned} \tag{10}$$

where

$$F_1(\theta_1) = \sum_{t=1}^{T} f_{1t}(\theta_1), \quad F_2(\theta) = \sum_{t=1}^{T} f_{2t}(\theta).$$

3.1.1 The impulse response function

The IRF of interest is given by

$$\underset{n\times 1}{g_j(\theta)} = \frac{\partial Y_{t+j}}{\partial \varepsilon_{1t}} = \left(\mathbf{I}_n, \mathbf{0}_{n\times n(k-1)}\right) A(\theta)^j \begin{pmatrix} 1 \\ \mathbf{0}_{n(k-1)\times 1}\end{pmatrix} g_0(\theta), \quad j \geq 1 \tag{11}$$

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

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

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 (12).

3.2 The conventional approach

The conventional approach, e.g., Blanchard and Quah (1989), is to use Gaussian MLE with conditional homoskedasticity. The MLE corresponds to the GMM estimator defined above when $Z_{1t} = Z_{2t} = \left(Y_{2,t-1}', X_{1t}'\right)$, namely, when we use $Y_{2,t-1}$ as instruments in the (1). This corresponds to 2SLS estimation of (1), $\hat{\theta}_1$, and OLS estimation of (2), $\hat{\theta}_2$ with the 'generated regressor' $h_{1t}\left(\hat{\theta}_1\right)$. Under strong-instrument stationary asymptotics, Wald tests are asymptotically 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 bootstrapping, e.g., Kilian (1998).

When α_2 is small, e.g., $\alpha_2 = O\left(T^{-1}\right)$, 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 (2009). Thus, conventional confidence bands on VAR coefficients and IRFs do not have correct asymptotic coverage. In the next subsection, we introduce a method that does.

3.3 Anderson Rubin test with filtered instruments

Our approach to solving this problem consists of two components:

1. Address the *near unit root* problem (when α_2 is small) by using filtered instruments – the IVX approach of Magdalinos and Phillips (2009, henceforth MP).

2. Address the *weak-instrument* problem using a weak-identification robust method – the Anderson and Rubin (1949) (henceforth AR) test, since 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

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). Consider the system of equations

$$
\begin{aligned}
y_t &= \theta x_{t-1} + u_t \\
x_t &= \left(1 + \frac{c}{T^a}\right) x_{t-1} + v_t, \quad 0 \leq a \leq 1, c \leq 0,
\end{aligned}
$$

and suppose we are interested in doing inference on θ. Instead of using the OLS t-test, following MP, Kostakis *et al* (2015) proposed to use IV with the following generated instrument:

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

so $z_t = \rho_z z_{t-1} + \Delta x_t$, with ρ_z sufficiently smaller than 1. They showed that the IV t-test is asymptotically *standard normal* under H_0 irrespective of the true c, a, and the choice of b, c_z.

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 them by X_{1t}), plus the lagged stationary regressors $Y_{2,t-1}$, which are excluded only from the first equation, i.e., $Z_{1t} = Z_{2t} = (X_t, Y_{2,t-1})$. The alternative we propose is to replace $Y_{2,t-1}$ in the instrument set with the filtered instrument

$$
z_t = \sum_{j=1}^{t-1} \rho_{Tz}^{t-j} \Delta Y_{2,j}, \quad \rho_{Tz} = 1 + \frac{c_z}{T^b},\ b \in (0,1)\,,\ c_z < 0. \tag{13}
$$

(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_{ff}, this can be tested using just the first equation (1). The AR statistic, call it $AR_1(b_{12}^0)$, is the squared t-statistic for testing $H_0^* : \delta_z = 0$ in the auxiliary regression:

$$\Delta Y_{1t} - b_{12}^0 \Delta Y_{2t} = X_{1t}\delta_1 + z_t\delta_z + \varepsilon_{1t}^0 \tag{14}$$

where X_{1t} contains the $k-1$ lags of ΔY_t, and z_t is the filtered instrument (13). Note that with conditional homoskedasticity, $AR_1(b_{12}^0)$ corresponds exactly to the minimum value of $S_1(\theta_1)$, defined in (10), under the restriction that $b_{12} = b_{12}^0$, i.e., $S_{1,T}\left(\hat{\theta}_1(b_{12}^0)\right)$, where $\hat{\theta}_1 = \arg\min_{\theta_1 : b_{12} = b_{12}^0} S_{1,T}(\theta_1)$. Under conditional homoskedasticity, this can be written analytically as

$$AR_1(b_{12}) = \frac{(\Delta Y_1 - \Delta Y_2 b_{12})' P_{M_{X_1}z} (\Delta Y_1 - \Delta Y_2 b_{12})}{(\Delta Y_1 - \Delta Y_2 b_{12})' M_{Z_1} (\Delta Y_1 - \Delta Y_2 b_{12}) / (T - \mathrm{col}(Z_1))}, \tag{15}$$

where $P_.$ denotes the projection matrix, $M_. = I - P_.$ and we follow standard notation that for any k-vector X_t, X denotes the $T \times k$ matrix of stacked X_t', $t = 1, ..., T$.

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

Theorem 1. *Consider the model (5) and (6), with* $\alpha_2 = \frac{c}{T^a}$, $\quad 0 \le a \le 1$, $c \le 0$, *and* ε_t *satisfying Assumption* A. *Let statistic* $AR_1(b_{12})$ *as in (15) where the instrument is defined by (13). Then under* $H_0 : b_{12} = b_{12}^0$, $AR_1(b_{12}^0) \xrightarrow{d} \chi_1^2$ *for all* $0 \le a \le 1$ *and* $c \le 0$.

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

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 $r : \Theta \to \Re^q$, $q < \dim\theta$. 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 to proceed using a projection argument as follows. 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. It remains to propose a joint test of H_0^*, which we turn to next.

Our test of the combined hypthesis H_0^* is based on a novel idea that combines the $AR_1(b_{12})$ statistic developed above with a Wald statistic for testing the restrictions on the *remaining* parameters in θ. (this idea applies more generally). Partition θ into b_{12} and ϑ, say, the remaining unknown parameters. Let $\hat{\vartheta}(b_{12})$ be the restricted GMM estimator of ϑ given b_{12} and let $\hat{V}_{\hat{\vartheta}}(b_{12})$ denote an estimate of the asymptotic variance matrix of $\hat{\vartheta}(b_{12})$. Provided $R(\theta) = \partial r(\theta)/\partial\vartheta'$ exists, define

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

and consider the combined statistic

$$ARW\left(b_{12}^0\right) = AR_1\left(b_{12}^0\right) + W\left(b_{12}^0\right).$$

The following result shows how to derive a test using the ARW statistic.

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 $0 \leq a \leq 1$, $c \leq 0$:*

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

$W\left(b_{12}^0\right)$ it is asymptotically independent of $AR_1(b_{12})$, and

$$ARW\left(b_{12}^0\right) = AR_1\left(b_{12}^0\right) + W\left(b_{12}^0\right) \xrightarrow{d} \chi_{q+1}^2.$$

The ARW test of $H_0^* : r(\theta) = 0, b_{12} = b_{12}^0$ at the η level of significance if

$ARW\left(b_{12}^{0}\right)$ is greater than the $1-\eta$ quantile of χ_{q+1}^{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^{*}.

In the conditionally homoskedastic 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. This test can be applied, e.g. to the IRF or forecast error variance decomposition.

3.4 Detrending

In some applications, Y_{2t} is 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 $\hat{Y}_{2t} = Y_{2t-1}^{obs} - \hat{\gamma}_x t$ in the computation of the IVX instrument Z_t.

Following Phillips, Park and Chang (2004) we use a recursive detrending formula to ensure that $\hat{Y}_{2t}$ is not computed using future values:

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

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. This is not the case under full-sample detrending.

4 Numerical results

We simulate data from the DGP

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

	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 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.

with

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

and letting $Y_{10} = Y_{20} = 0$. We consider the following parameters sets: $\rho \in \{0.20, 0.95\}$, $c \in \{0, -1, -10, -30, -100\}$, $b_{12} = 0$ under H_0. We compute rejection probabilities of our AR test and the t test under H_0 using 20000 MC replications. The sample size is set to $T = 200$.

4.1 Size in Finite Samples

We first consider the SVAR(1) and SVAR(2) estimated model in in Tables 1 and 2. We compare our proposed AR statistic with the conventional t test. Contrary to the latter, our proposal controls size in finite samples.

	At 5%				At 10%			
	$\rho = 0.20$		0.95		0.20		0.95	
	AR	t	AR	t	AR	t	AR	t
$c = 0$	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 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.

4.2 Power

We compute (large-sample) power of the AR and t tests of $H_0 : b_{12} = 0$ under weak identification. We set $T = 2000$, use 1000 Monte Carlo replications, and consider $c = -10, -100, -500$ in Figure 1. These correspond to approximate values of the concentration parameter λ of 1.3, 13, 72. The range of b_{12} under H_1 is $\lambda^{-1/2}(-3 : 3)$.

The figures show that the AR tests shows reasonable power, even for low c. 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, the power of the AR test is very similar to that of the t test, which is asymptotically efficient in this case.

5 Empirical Results

5.1 Blanchard and Quah (1989)

We first revisit the application of Blanchard and Quah (1989), where Y_{1t} is log real GNP, and Y_{2t} is unemployment. We use the original Blanchard and Quah

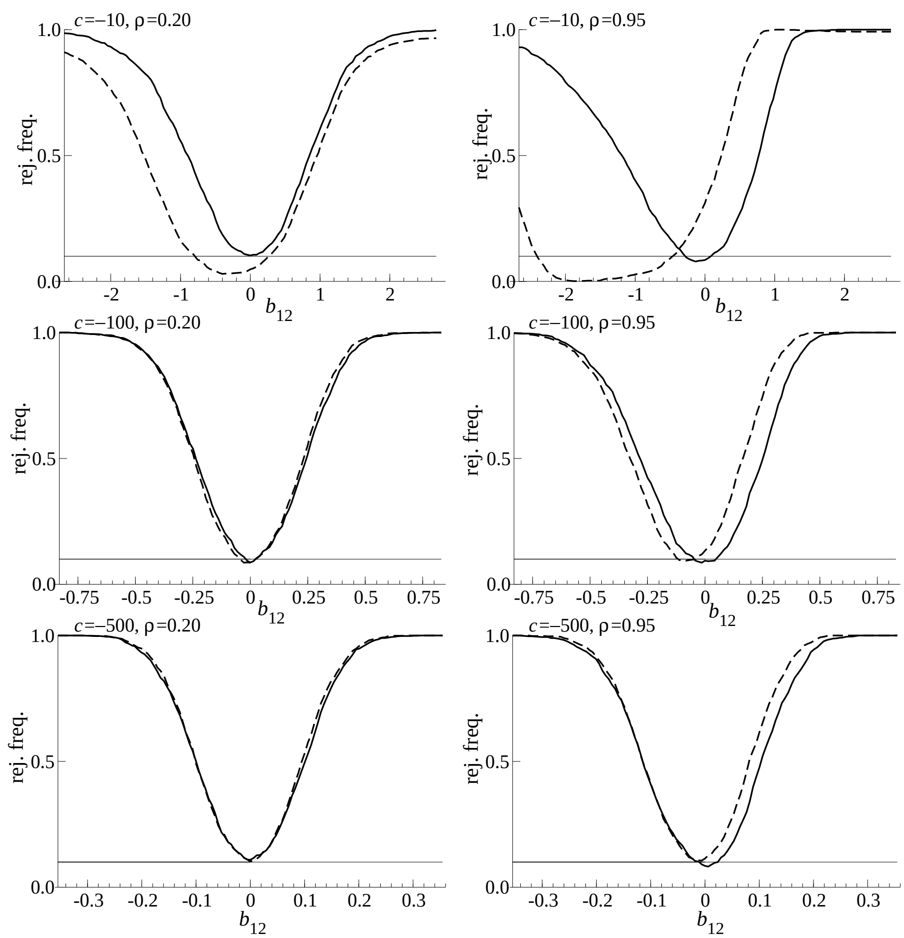

Figure 1: Large-sample power of AR (with filtered instrument) and t tests of the hypothesis $H_0 : b_{12} = 0$ against $H_1 : b_{12} \neq 0$ in the SVAR(1) model with long run restrictions. $T = 2000$, 1000 MC replications, ρ is correlation of reduced-form errors.

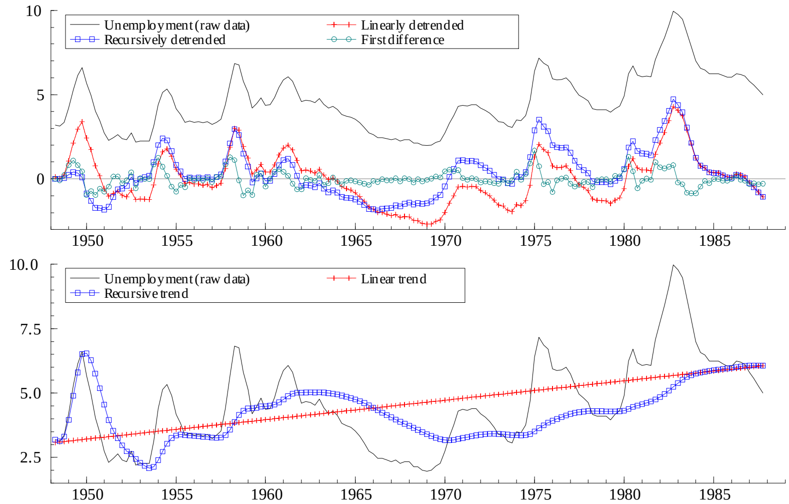

Figure 2: Blanchard and Quah (1989) data with various detrending methods.

(1989) data, which are quarterly and cover the period 1948q2 to 1987q4. The level specification in Blanchard and Quah (1989) is an SVAR(9) with Y_{1t} in first differences and Y_{2t} linearly detrended, as shown 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 Blanchard and Quah (1989). We see that the estimates do not differ from those of Blanchard and Quah (1989), but the robust confidence bands are so large, that the original conclusion of Blanchard and Quah (1989) 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 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

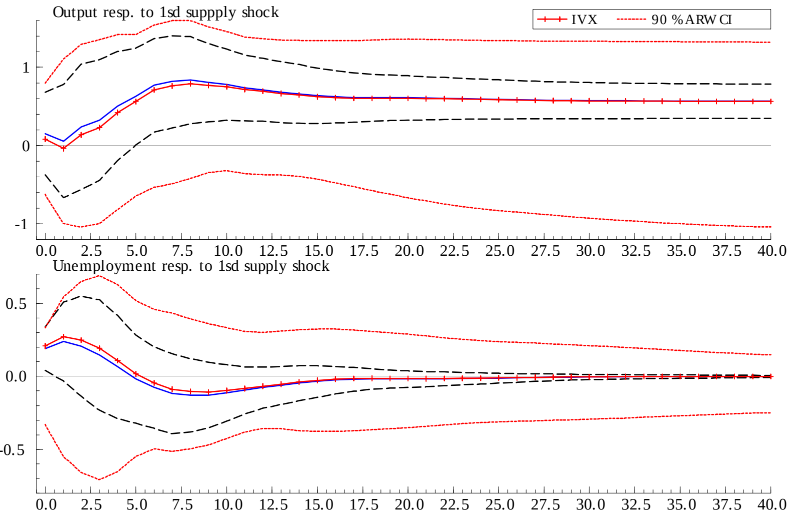

Figure 3: Estimates and confidence intervals of the IRFs. Robust (in red), and Blanchard and Quah (1989) (in blue and black)

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 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. The original conclusions of CEV are not robust to weak identification.

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 non-stationarity in the data. The method is based on the Anderson Rubin statistic with instruments obtained by filtering the potentially non-stationary variables to make them

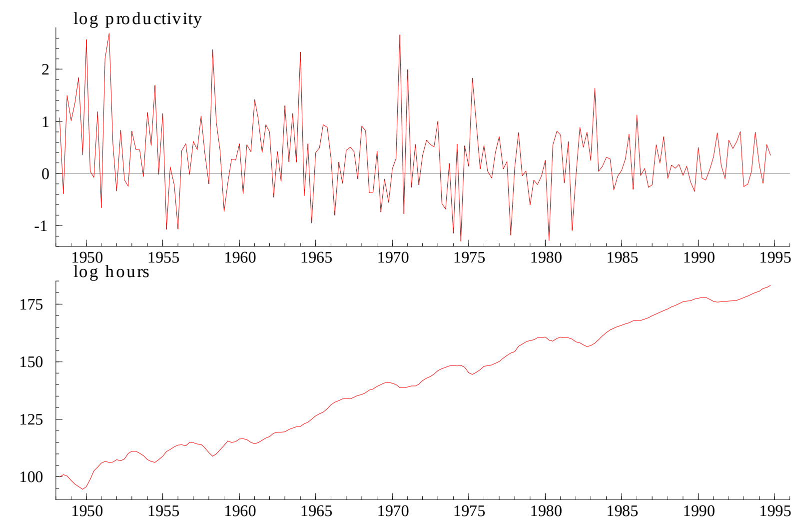

Figure 4: Gali (1999) data.

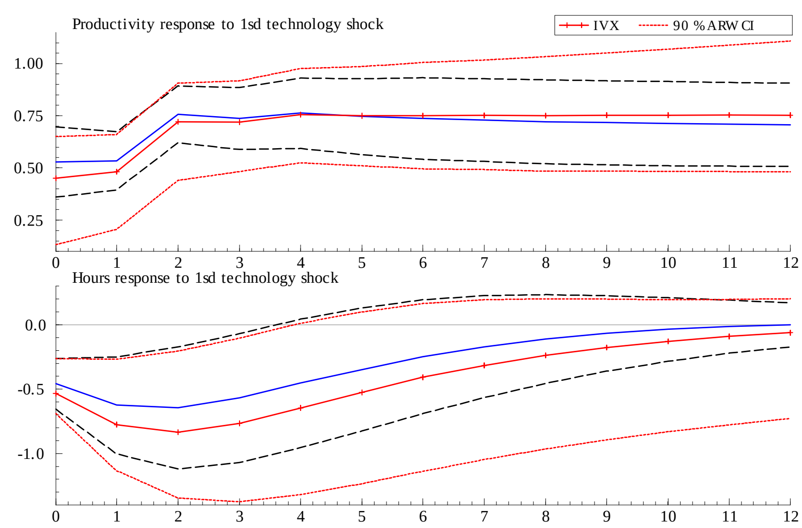

Figure 5: IRFs to technology shock. Robust estimates (in red) together with the Gali (1999) estimates (in blue and black).

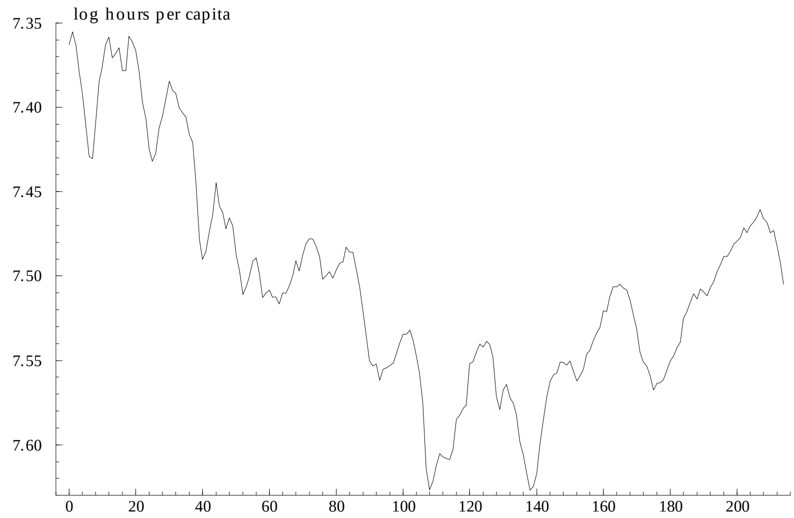

Figure 6: CEV data.

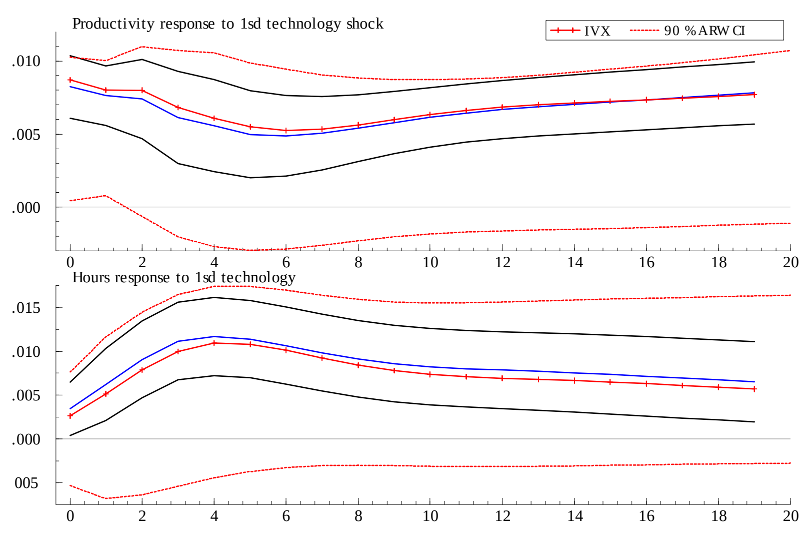

Figure 7: IRFs to technology shock. Robust estimates (in red) together with the CEV estimates (in blue and black).

near stationary. Tests of general parametric restrictions, and confidence intervals for differentiable functions of the parameters, such as IRFs, 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.

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.

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. (2009). Inference in Nearly Nonstationary SVAR Models with Long-Run Identifying Restrictions. *Journal of Business and Economic Statistics*. forthcoming.

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*.

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

Mittnik, S. and P. A. Zadrozny (1993). Asymptotic distributions of impulse responses, step responses, and variance decompositions of estimated linear dynamic models. *Econometrica: Journal of the Econometric Society*, 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. *The Review of Economics and Statistics 80*(2), 202–217.

Pesavento, E. and B. Rossi (2007). Impulse response confidence intervals for persistent data: What have we learned? *Journal of Economic Dynamics and Control 31*(7), 2398–2412.

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

Rossi, B. and E. Pesavento (2006). Small-sample confidence intervals for multivariate impulse response functions at long horizons. *Journal of Applied Econometrics 21*(8), 1135–1155.

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*, 557–586.