Supplementary Appendix for Robust inference in structural VARs with long run restrictions

Guillaume Chevillon*
ESSEC Business School

Sophocles Mavroeidis†
University of Oxford

Zhaoguo Zhan‡
Kennesaw State University

September 12, 2018

Contents

*ESSEC Business School, Ave. B. Hirsch, 95000 Cergy-Pontoise, France.

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

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

1 Introduction

This appendix contains proofs, algebraic derivations, detailed description of econometric methods and additional empirical results. If the reader is primarily interested in the derivations and empirical results, the description of the computation algorithms can be skipped. Equations in this document are numbered with the suffix 'S–'. Equations without suffix refer to the main paper.

We will make repeated use of the following references, which we abbreviate as indicated for brevity: MPvic stands for Magdalinos and Phillips (2009a) , MPet stands for Magdalinos and Phillips (2009b) and KMS stands for Kostakis, Magdalinos and Stamatogiannis (2015) .

2 Overview of notations

We list here all the notations. The model is

$$\Delta Y_1 = \Delta Y_2 b_{12} + X_1 \delta_1 + \varepsilon_1$$
$$\Delta Y_2 = \overline{X}_2 \psi_2 + v_2$$

with $\overline{X}_2 = [Y_2 : X_2 : \varepsilon_1]$, where Y_2 contains the stacked elements of $Y_{2,t-1}$. The AR statistic for testing $H_0 : b_{12} = b_{12}^0$ is the square of the t-test of $\delta_z = 0$ in

$$\Delta Y_1 - \Delta Y_2 b_{12}^0 = X_1 \delta_1 + z\delta_Z + \varepsilon_1^0$$

with instruments $Z_1 = [z : X_1]$. Under H_0, the residual $\hat{\varepsilon}_1$ is $\hat{\varepsilon}_1 = M_{X_1} (\Delta Y_1 - \Delta Y_2 b_{12})$.

We denote

$$\hat{X}_2 = [Y_2 : X_2 : \hat{\varepsilon}_1]$$

with instruments $\hat{Z}_2 = [z : X_2 : \hat{\varepsilon}_1]$ and $\hat{v}_2 = \Delta Y_2 - \hat{X}_2 \hat{\psi}_2$ where $\hat{\psi}_2$ is the IV estimator. When necessary, we let

$$\hat{X}_{21} = [X_2 : \hat{\varepsilon}_1].$$

3 Proofs of results in the paper

The proofs use extensively the results of MPvic and KMS. These authors consider sequences $\alpha_2 = cT^a$ for $c \leq 0$ and $a \in [0, 1]$. We prove in Section 3.1 that their theorems can be generalized to all sequences such that $T\alpha_2 \to [-\infty, 0]$ provided that the innovations satisfy a slightly more restrictive Assumption LP* that holds under Assumption A made in our paper. Our setting presents some simplifications compared to those of MPvic and KMS. Specifically, the generated instrument z_t is predetermined as opposed to the case considered by MPvic where it is not. Hence, $cov(z_t, \varepsilon_{1t}) = 0$ and there is no need to estimate this covariance, so the condition $b > 2/3$ in MPvic is not required.

3.1 Extending IVX to general sequences of parameters

In the following, we use the results of the papers by MPvic, Giraitis and Phillips (2006, GP06, and 2012, GP12). We consider, for the processes with $x_0 = o_p\left(\sqrt{\frac{T}{1-Tc_T}}\right)$, i.e. $x_0 = o_p\left(|c_T|^{-1/2} \wedge T^{1/2}\right)$, where $\wedge$ denotes the minimum (and $\vee$ the maximum). For readability and comparison with MPvic, we use the following notation in this section – and corresponding proofs – only:

$$\begin{aligned} y_t &= \theta x_t + u_t, \\ x_t &= \rho_T x_{t-1} + v_t, \\ \tilde{z}_t &= \rho_Z \tilde{z}_{t-1} + \Delta x_t \\ z_t &= \rho_Z z_{t-1} + v_t, \end{aligned} \tag{S–1}$$

where as in MPvic,

$$\rho_Z = 1 + c_z T^{-b}, \quad b \in (1/2, 1), \quad c_z < 0. \tag{S–2}$$

Notice that in the equation for y_t, we retain the regressor x_t as in MPvic, whereas we use its first lag in our model. We keep this in order to show that the results of can be generalized. It is then easy to provide the required results by appropriate definition of the error process v_t and of x_0. Assumption $b \in (1/2, 1)$ is found in MPvic: it is required in the proofs of Proposition A2, Lemma 3.5 and Lemma 3.6, where u_t is expressed according to the Beveridge-Nelson decomposition of Phillips and Solo (1992). When u_t is *i.i.d*, as in KMS, the condition $b > 1/2$ is no longer required.

We extend below the results of MPvic, in the univariate case, to $c_T = \rho_T - 1$ admitting a general formulation as in the following assumption which replaces Assumption N of MPvic (which we refer to as MPvic-Assumption N):

Assumption N*: *The coefficient* $c_T = \rho_T - 1 \in (-2, 0]$ *satisfies as* $T \to \infty$ *one of the three assumptions* (*i*) $Tc_T \to 0$;

(*ii*) $Tc_T \to c < 0$;

(*iii*) $Tc_T \to -\infty$.

Assumption N* is found in GP06 and GP12 who make a different assumption about the dynamics of v_t from that which is found in MPvic. Our assumption on the dynamics of v_t combines those of MPvic and GP12 so the results of both articles hold (and the assumption of KMS when c_T is constant also holds):

Assumption LP*: $(u_t, v_t)' = F(L)\varepsilon_t = \sum_{j=0}^{\infty} f_j \varepsilon_{t-j}$ *where* ε_t is an $i.i.d$ sequence with $E(\varepsilon_t) = 0$, $E(\varepsilon_t \varepsilon_t') = \Sigma$, $E\left(\|\varepsilon_t\|^4\right) < \infty$, $F(1)$ has full rank, *and, for* $k \geq 1$, $\sum_{j=k}^{\infty} |f_j| \leq k^{-1-\kappa}$, *for* $\kappa > 2$.

Let $F(L) = (F_u'(L), F_v'(L))'$ and the long run covariance

$$\Omega = \left[\begin{array}{cc} \Omega_{uu} & \Omega_{uv} \\ \Omega_{vu} & \Omega_{vv} \end{array} \right] = F(1)\Sigma F(1)'.$$

We also let $\Lambda_{uv}^0 = \sum_{j=0}^{\infty} E(u_t v_{t-j})$, $\Lambda_{uv} = \sum_{j=1}^{\infty} E(u_t v_{t-j})$ with corresponding matrix Λ that conforms to Ω.

We provide below the equivalent lemmas and theorems to MPvic under the Assumptions N* and LP* above. The only modification in the formulation of the lemmas and theorems concerns Assumption N*(*iii*) which replaces MPvic-Assumption N(*iii*). Under the former, the instrument is less persistent than the regressor when $\rho_T - 1 = o(\rho_Z - 1)$, i.e., instead of $b < a$ in MPvic-Assumption N(*iii*), we now have

$$c_T = o\left(T^{-b}\right), \tag{S–3}$$

and expression (MPvic-13) rewrites $\tilde{z}_t = z_t + c_T \psi_{Tt}$.

We now state the required lemmas of MPvic under the new assumptions, keeping the same number as in MPvic to show the relation under the new assumptions, but adding a start (*). Hence Lemma 3.1 in MPvic becomes Lemma 1* here.

Lemma 1* Consider the model given by (S–1)-(S–2) under Assumptions N* and LP* with $c_T = o\left(T^{-b}\right)$, the following approximations hold as $T \to \infty$:

(*i*) $T^{-\frac{1+b}{2}} \sum_{t=1}^{T} u_t \tilde{z}_t = T^{-\frac{1+b}{2}} \sum_{t=1}^{T} u_t z_t + o_p(1)$;

(*ii*) $T^{-(1+b)} \sum_{t=1}^{T} x_t \tilde{z}_t = T^{-(1+b)} \sum_{t=1}^{T} x_t z_t - T^{-1} c_T c_z \sum_{t=1}^{T} x_t^2 + o_p(1)$

(*iii*) $T^{-(1+b)} \sum_{t=1}^{T} \tilde{z}_t^2 = T^{-(1+b)} \sum_{t=1}^{T} z_t^2 + o_p(1)$.

Lemma 2* Consider the model given by (S–1)-(S–2) under Assumptions N* and

LP*. The martingale array $U_T(s) = T^{-\frac{1+b}{2}} \sum_{t=1}^{\lfloor Ts \rfloor} [z_{t-1} F_u(1) \varepsilon_t]$ satisfies $U_T(s) \Rightarrow U(s)$ where U_s is a Brownian motion with variance $-\frac{1}{2c_z}\Omega_{uu}\Omega_{vv}$ and independent of B_v ($B_v(s)$ defined as limit of $T^{-1/2}\sum_{t=1}^{\lfloor Ts \rfloor} v_t$). Joint convergence in distribution of $U_T(1)$, $T^{-1}\sum_{t=1}^{T} x_{t-1}\varepsilon_t$ and $T^{-1}c_T\sum_{t=1}^{T} x_{t-1}^2$ also applies.

Lemma 5* Consider the model given by (S–1)-(S–2) under Assumptions N*(iii) and LP* with $c_T^{-1} = o(T^b)$ and $b \in (1/2, 1)$, then the following approximations hold as $T \to \infty$:

(i) $\sqrt{\frac{-c_T}{T}}\sum_{t=1}^{T} u_t\tilde{z}_t = \sqrt{\frac{-c_T}{T}}\sum_{t=1}^{T} u_t x_t + o_p(1)$;

(ii) $\frac{c_T}{T}\sum_{t=1}^{T} x_t\tilde{z}_t = \frac{c_T}{T}\sum_{t=1}^{T} x_t^2 + o_p(1)$;

(iii) $\frac{c_T}{T}\sum_{t=1}^{T} \tilde{z}_t^2 = \frac{c_T}{T}\sum_{t=1}^{T} x_t^2 + o_p(1)$.

Lemma 6* Consider the model given by (S–1)-(S–2) under Assumptions N*(iii) and LP* where $\kappa_1 < T^b c_T < \kappa_0$, for some $\kappa_1 < \kappa_0 < 0$, and $b \in (1/2, 1)$. Then the following approximations hold as $T \to \infty$:

(i) $\sqrt{-(c_z + T^b c_T)}T^{-\frac{1+b}{2}}\sum_{t=1}^{T}(u_t\tilde{z}_t - \Lambda_{uv}^0) \Rightarrow N\left(0, \frac{1}{2}\Omega_{vv}\Omega_{uu}\right)$;

(ii) $-\left(c_z + T^b c_T\right)T^{-(1+b)}\sum_{t=1}^{T} x_t\tilde{z}_t \xrightarrow{p} \frac{1}{2}\Omega_{vv}$;

(iii) $-\left(c_z + T^b c_T\right)T^{-(1+b)}\sum_{t=1}^{T} \tilde{z}_t^2 \xrightarrow{p} \frac{1}{2}\Omega_{vv}$.

Proofs of the lemmas are provided in Section 3.7.

3.2 Proof of Lemma P

The proofs for items (i), (ii) and (iii) follow from the result of MPvic where we have established the equivalent lemmas for general sequences c_T which becomes α_2 in our context. For items (iv), we notice that the proof of Lemma 5* goes through with ΔY_{t-i} $i = 1, ..., m-1$ in place of u_t, because the part of Assumption LP* that requires $F(1)$ to be of full rank is not needed in the proof of Lemma 5*. It also covers the case of over differencing where α_2 is constant. Joint convergence of (i), (ii) and (iii) follows from Lemma 2*.

Parts (v) and (vi) follow from GP12, Lemma 2.1 and Theorem 2.2, who showed that

$$\frac{1}{T}\sum_{t=1}^{T} Y_{2,t-1}X_{it} \xrightarrow{p} \Sigma_{Y_2X_i}, \quad i = 1, 2,$$

where $\Sigma_{Y_2X_i}$ is nonstochastic, and

$$\sqrt{\frac{-\alpha_2 \vee T^{-1}}{T}}\sum_{t=1}^{T} Y_{2,t-1}\varepsilon_{1t} \Rightarrow N\left(0, \frac{\omega}{2}\sigma_{\varepsilon_1}^2\right),$$

and the fact that $\kappa_T/T = o\left(\frac{-\alpha_2 \vee T^{-1}}{T}\right) = o\left(\frac{-\alpha_2}{T} \vee T^{-2}\right)$.

3.3 Proof of Proposition 4

The first equation is a linear IV regression, so the estimator of δ_1 solves the equation $X_1' Z_1 \hat{V}_{f_1}^{-1} Z_1' \left(\Delta Y_1 - \Delta Y_2 b_{12} - X_1 \hat{\delta}_1\right) = 0$. Conditional homoskedasticity implies that we can set $\hat{V}_f$ proportional to $Z_1' Z_1$, so $\hat{\delta}_1$ is 2SLS

$$\hat{\delta}_1 = \left(X_1' P_{Z_1} X_1\right)^{-1} X_1' P_{Z_1} \left(\Delta Y_1 - \Delta Y_2 b_{12}\right).$$

Since $Z_1 = (z, X_1)$, this reduces to

$$\hat{\delta}_1 = \left(X_1' X_1\right)^{-1} X_1' \left(\Delta Y_1 - \Delta Y_2 b_{12}\right),$$

i.e., simply OLS of $\Delta Y_1 - \Delta Y_2 b_{12}$ on the exogenous regressors X_1. The estimator of $\sigma^2_{\varepsilon_1}$ is simply $T^{-1} \sum_{t=1}^{T} \hat{\varepsilon}^2_{1t}$, where $\hat{\varepsilon}_{1t} = \Delta Y_{1t} - \Delta Y_{2t} b_{12} - X_{1t}' \hat{\delta}_1$. So,

$$\hat{\psi}_1 = \begin{pmatrix} \hat{\delta}_1 \\ \hat{\sigma}^2_{\varepsilon_1} \end{pmatrix} = \begin{pmatrix} \left(X_1' X_1\right)^{-1} X_1' \left(\Delta Y_1 - \Delta Y_2 b_{12}\right) \\ T^{-1} \sum_{t=1}^{T} \hat{\varepsilon}^2_{1t} \end{pmatrix}. \tag{S–4}$$

Now, let us turn to equation (3). For convenience, define the 'generated regressors'

$$\overline{X}_{2t}(\theta_1) = \left(Y_{2,t-1}, X_{2t}', h_{1t}(\theta_1)\right)'$$

and the corresponding 'generated instruments'

$$\overline{Z}_{2t}(\theta_1) = \left(Z_{2t}', h_{1t}(\theta_1)\right)' = \left(z_t, X_{2t}', h_{1t}(\theta_1)\right)'.$$

In what follows, we will omit the dependence of $\overline{X}_{2t}$ and $\overline{Z}_{2t}$ on θ_1 for brevity, and we will use the shorthand notation $\hat{X}_{2t} = \overline{X}_{2t}\left(b_{12}, \hat{\psi}_1\right) = \left(Y_{2,t-1}, X_{2t}', \hat{\varepsilon}_{1t}\right)'$, and similarly for $\hat{Z}_{2t}$. Because the second equation is a just-identified linear IV regression in the (generated) regressors/instruments, the estimator $\hat{\psi}_2$ solves $F_{2T}\left(b_{12}, \hat{\psi}_1, \hat{\psi}_2\right) = 0$, which yields

$$\hat{\psi}_2 = \left(\hat{Z}_2' \hat{X}_2\right)^{-1} \hat{Z}_2' \Delta Y_2. \tag{S–5}$$

Subtracting ψ_2 and substituting for ΔY_2 yields

$$\hat{\psi}_2 - \psi_2 = \left(\hat{Z}_2'\hat{X}_2\right)^{-1}\hat{Z}_2'v_2 + \left(\hat{Z}_2'\hat{X}_2\right)^{-1}\hat{Z}_2'X_1\left(\hat{\delta}_1 - \delta_1\right)d_{21}. \tag{S–6}$$

Collecting terms yields

$$\hat{\psi} - \psi = \begin{pmatrix} \left(X_1'X_1\right)^{-1}X_1'\varepsilon_1 \\ T^{-1}\hat{\varepsilon}_1'\hat{\varepsilon}_1 - \sigma_{\varepsilon_1}^2 \\ \left(\hat{Z}_2'\hat{X}_2\right)^{-1}\hat{Z}_2'v_2 + \left(\hat{Z}_2'\hat{X}_2\right)^{-1}\hat{Z}_2'P_{X_1}\varepsilon_1 d_{21} \end{pmatrix}.$$

Next, we need to get the estimator of the variance of $\hat{\psi}$. First, note that $\tilde{V}_f(b_{12})$, the estimator of $E\left[f_t(\theta) f_t(\theta)'\right]$, is block diagonal if we impose the orthogonality of the errors ε_{1t}, v_{2t}, because, at the true value of θ, $E\left(f_{1t}(\theta) f_{2t}(\theta)'\right) = E\left(Z_{1t}\varepsilon_{1t}v_{2t}Z_{2t}'\right)$, and Z_{1t}, Z_{2t} are predetermined, so $E\left(\varepsilon_{1t}v_{2t}|Z_{1t}, Z_{2t}\right) = 0$. Hence,

$$\tilde{V}_f(b_{12}) = \begin{pmatrix} \tilde{V}_{f_1}(b_{12}) & 0 \\ 0 & \tilde{V}_{f_2}(b_{12}) \end{pmatrix}.$$

Next,

$$\tilde{V}_{f_1}(b_{12}) = \frac{1}{T^2}\begin{pmatrix} Z_1'Z_1\hat{\sigma}_{\varepsilon_1}^2 & 0 \\ 0 & T\hat{\varpi} \end{pmatrix}$$

where $\hat{\varpi}$ is an estimator of $var\left(\hat{\sigma}_{\varepsilon_1}^2\right)$. Under the maintained assumptions, a consistent estimator is given by $\hat{\varpi} = T^{-1}\sum_{t=1}^{T}\left(\hat{\varepsilon}_{1t}^2 - \hat{\sigma}_{\varepsilon_1}^2\right)^2$. If we assume that $var\left(\hat{\varepsilon}_{1t}^2\right) = 2\sigma_{\varepsilon_1}^4$, which holds under Gaussianity, then we can use $\hat{\varpi} = 2\hat{\sigma}_{\varepsilon_1}^4$, as in Blanchard and Quah (1989) and Galí (1999). Finally,

$$\hat{V}_{f_2}(b_{12}) = \frac{1}{T^2}\hat{Z}_2'\hat{Z}_2\hat{\sigma}_{v_2}^2, \quad \hat{\sigma}_{v_2}^2 = T^{-1}\hat{v}_2'\hat{v}_2, \quad \hat{v}_2 = \Delta Y_2 - \hat{X}_2\hat{\psi}_2.$$

Next, the Jacobian of the moment conditions is given by

$$\hat{J}_T(b_{12}) = \left.\frac{\partial F_T(\theta)}{\partial \psi'}\right|_{\theta = \binom{b_{12}}{\hat{\psi}}} = \frac{1}{T}\begin{pmatrix} -Z_1'X_1 & 0 & 0 \\ 0 & -T & 0 \\ \hat{Z}_2'X_1 d_{21} & 0 & -\hat{Z}_2'\hat{X}_2 \end{pmatrix}.$$

Hence,

$$
\begin{aligned}
&\hat{J}_T(b_{12})' \tilde{V}_f(b_{12})^{-1} \hat{J}_T(b_{12}) \\
&= \begin{pmatrix} -Z_1'X_1 & 0 & 0 \\ 0 & -T & 0 \\ \hat{Z}_2'X_1 d_{21} & 0 & -\hat{Z}_2'\hat{X}_2 \end{pmatrix}' \begin{pmatrix} (Z_1'Z_1)^{-1}\hat{\sigma}_{\varepsilon_1}^{-2} & 0 & 0 \\ 0 & T^{-1}\hat{\varpi}^{-1} & 0 \\ 0 & 0 & \left(\hat{Z}_2'\hat{Z}_2\right)^{-1}\hat{\sigma}_{v_2}^{-2} \end{pmatrix} \\
&\times \begin{pmatrix} -Z_1'X_1 & 0 & 0 \\ 0 & -T & 0 \\ \hat{Z}_2'X_1 d_{21} & 0 & -\hat{Z}_2'\hat{X}_2 \end{pmatrix} \\
&= \begin{pmatrix} X_1'P_{Z_1}X_1\hat{\sigma}_{\varepsilon_1}^{-2} + X_1'P_{\hat{Z}_2}X_1 d_{21}^2\hat{\sigma}_{v_2}^{-2} & 0 & d_{21}X_1'P_{\hat{Z}_2}\hat{X}_2\hat{\sigma}_{v_2}^{-2} \\ 0 & T\hat{\varpi}^{-1} & 0 \\ d_{21}\hat{X}_2'P_{\hat{Z}_2}X_1\hat{\sigma}_{v_2}^{-2} & 0 & \hat{X}_2'P_{\hat{Z}_2}\hat{X}_2\hat{\sigma}_{v_2}^{-2} \end{pmatrix}.
\end{aligned}
$$

Using the partitioned inverse formula and simplifying yields the expression for $\hat{V}_{\hat{\psi}} = \left[\hat{J}_T(b_{12})' \tilde{V}_f(b_{12})^{-1} \hat{J}_T(b_{12})\right]^{-1}$, with

$$\hat{V}_{\hat{\psi},11} = (X_1'X_1)^{-1}\hat{\sigma}_{\varepsilon_1}^2$$

$$\hat{V}_{\hat{\psi},12} = 0$$

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

$$\hat{V}_{\hat{\psi},22} = \frac{\hat{\varpi}}{T}$$

$$\hat{V}_{\hat{\psi},23} = 0$$

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

Rewriting the last term yields the expression in the proposition. Now, let

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

It can be easily verified that $\hat{C}_{\hat{\psi}}\hat{C}_{\hat{\psi}}' = \hat{V}_{\hat{\psi}}(\vartheta)^{-1}$.

So,

$$\hat{\xi}_2 = C_{\hat{V}_{\hat{\psi}}^{-1}}\left(\hat{\psi}-\psi\right) = \begin{pmatrix} \left(X_1'X_1\right)^{-1/2} X_1'\varepsilon_1\hat{\sigma}_{\varepsilon_1}^{-1} \\ \hat{\varpi}^{-1/2}\left(\hat{\sigma}_{\varepsilon_1}^2-\sigma_{\varepsilon_1}^2\right) \\ C_{\hat{Z}_2'\hat{Z}_2}^{-1}\hat{Z}_2'\upsilon_2\hat{\sigma}_{\upsilon_2}^{-1} \end{pmatrix}.$$

Finally, we turn to the derivation of $\hat{\xi}_1$. The moment vector $\hat{F}_T(\vartheta)$, with $\vartheta = b_{12}$, is

$$\hat{F}_T(b_{12}) = \begin{pmatrix} \hat{F}_{1T}(b_{12}) \\ \hat{F}_{2T}(b_{12}) \end{pmatrix},$$

where

$$\begin{aligned} \hat{F}_{1T}(b_{12}) &= \frac{1}{T}\begin{pmatrix} Z_1'\left[\Delta Y_1 - b_{12}\Delta Y_2 - X_1\left(X_1'X_1\right)^{-1}X_1'\left(\Delta Y_1 - b_{12}\Delta Y_2\right)\right] \\ \hat{\varepsilon}_1'\hat{\varepsilon}_1 - T\hat{\sigma}_{\varepsilon_1}^2 \end{pmatrix} \\ &= \frac{1}{T}\begin{pmatrix} Z_1'M_{X_1}\left(\Delta Y_1 - b_{12}\Delta Y_2\right) \\ 0 \end{pmatrix} = \frac{1}{T}\begin{pmatrix} z'M_{X_1}\left(\Delta Y_1 - b_{12}\Delta Y_2\right) \\ 0_{(\operatorname{col} X_1+1)\times 1} \end{pmatrix}, \end{aligned}$$

and

$$\hat{F}_{2T}(b_{12}) = \frac{1}{T}\hat{Z}_2'\left(\Delta Y_2 - \hat{X}_2\hat{\psi}_2\right) = \frac{1}{T}\hat{Z}_2'\left[I - \hat{X}_2\left(\hat{Z}_2'\hat{X}_2\right)^{-1}\hat{Z}_2'\right]\Delta Y_2 = 0.$$

Now,

$$\begin{aligned} \hat{S}_T(b_{12}) &= \hat{F}_T(b_{12})'\tilde{V}_f(b_{12})^{-1}\hat{F}_T(b_{12}) \\ &= \frac{\left(\Delta Y_1 - b_{12}\Delta Y_2\right)'M_{X_1}P_{Z_1}M_{X_1}\left(\Delta Y_1 - b_{12}\Delta Y_2\right)}{\hat{\sigma}_{\varepsilon_1}^2} \\ &= \frac{\left(\Delta Y_1 - b_{12}\Delta Y_2\right)'P_{M_{X_1}z}\left(\Delta Y_1 - b_{12}\Delta Y_2\right)}{\hat{\sigma}_{\varepsilon_1}^2} = \hat{\xi}_1'\hat{\xi}_1, \end{aligned}$$

where

$$\begin{aligned} \hat{\xi}_1 &= \left(z'M_{X_1}z\right)^{-1/2}\hat{\sigma}_{\varepsilon_1}^{-1}z'M_{X_1}\left(\Delta Y_1 - b_{12}\Delta Y_2\right) \\ &= \left(z'M_{X_1}z\right)^{-1/2}\hat{\sigma}_{\varepsilon_1}^{-1}z'M_{X_1}\varepsilon_1, \end{aligned}$$

which is a scalar in the case $n = 2$.

3.4 Proof of Proposition 5

(i) $\tilde{\psi} = \hat{\psi}$ follows from linearity, just-identification and conditional homoskedasticity, which implies that the IV estimator of ψ does not depend on any weighting matrix, as seen in the proof of Proposition 4. For (ii), take $\hat{\psi}_1 = \binom{\hat{\delta}_1}{\hat{\sigma}^2_{\varepsilon_1}}$. Then,

$$\hat{\delta}_1 = \delta_1 + \left(\frac{X_1' X_1}{T}\right)^{-1} \frac{X_1' \varepsilon_1}{T} = \delta_1 + O_p(1)\, o_p(1) \xrightarrow{p} \delta_1,$$

since X_1 consists of lags of ΔY_t and ε_1 is an innovation process. So,

$$\hat{\sigma}^2_{\varepsilon_1} = T^{-1} \sum_{t=1}^{T} \hat{\varepsilon}^2_{1t} = T^{-1} \sum_{t=1}^{T} \varepsilon^2_{1t} + o_p(1) \xrightarrow{p} \sigma^2_{\varepsilon_1},$$

by Assumption A and the law of large numbers. Turning to $\hat{\psi}_2$, from (S–6) and the consistency of $\hat{\psi}_1$, we have

$$\hat{\psi}_2 - \psi_2 = \left(\overline{Z}_2' \overline{X}_2\right)^{-1} \overline{Z}_2' v_2 + o_p(1). \tag{S–7}$$

Next, let

$$D_T = \begin{pmatrix} \sqrt{\kappa_T} & 0 \\ 0 & T^{-1/2} I_{p_{\psi_2}-1} \end{pmatrix}, \quad \kappa_T = \frac{-\left(c_z + T^b \alpha_2\right)}{T^{1+b}}, \tag{S–8}$$

so that

$$D_T \overline{Z}_2' \overline{X}_2 D_T = \begin{pmatrix} \kappa_T z' Y_2 & \sqrt{\frac{\kappa_T}{T}} z' X_2 & \sqrt{\frac{\kappa_T}{T}} z' \varepsilon_1 \\ \sqrt{\frac{\kappa_T}{T}} X_2' Y_2 & T^{-1} X_2' X_2 & T^{-1} X_2' \varepsilon_1 \\ \sqrt{\frac{\kappa_T}{T}} \varepsilon_1' Y_2 & T^{-1} \varepsilon_1' X_2 & T^{-1} \varepsilon_1' \varepsilon_1 \end{pmatrix}. \tag{S–9}$$

If $T\alpha_2 \to -\infty$, then from Lemma P we have

$$D_T \overline{Z}_2' \overline{X}_2 D_T = \begin{pmatrix} \omega + o_p(1) & O_p(T\kappa_T) & o_p(1) \\ O_p(T\kappa_T) & \Sigma_{X_2X_2} + o_p(1) & o_p(1) \\ o_p(1) & o_p(1) & \sigma^2_{\varepsilon_1} + o_p(1) \end{pmatrix},$$

where $\Sigma_{X_2X_2} = \lim_{T\to\infty} E\left(X_{2t} X_{2t}'\right)$. More specifically, if $\alpha_2 \to 0$, i.e., $T\kappa_T \to 0$, then

$$D_T \overline{Z}_2' \overline{X}_2 D_T \xrightarrow{p} \begin{pmatrix} \omega & 0 & 0 \\ 0 & \Sigma_{X_2X_2} & 0 \\ 0 & 0 & \sigma^2_{\varepsilon_1} \end{pmatrix}. \tag{S–10}$$

If $\alpha_2 < 0$ is fixed, i.e., $T\kappa_T \to -\alpha_2$, then

$$D_T \overline{Z}_2' \overline{X}_2 D_T \xrightarrow{p} \begin{pmatrix} E\left(\binom{\sqrt{-\alpha_2}Y_{2,t-1}}{X_{2,t}}\binom{\sqrt{-\alpha_2}Y_{2,t-1}}{X_{2,t}}'\right) & 0 \\ 0 & \sigma_{\varepsilon_1}^2 \end{pmatrix}. \tag{S–11}$$

To see this, note that if $\alpha_2 < 0$ is fixed, then $T^{-1}z'X_1 = T^{-1}Y_2'X_1 + o_p(1)$ by Lemma 5*(i) and hence, $T^{-1}z'X_1 \xrightarrow{p} E\left(Y_{2t-1}X_{1t}'\right)$.

For brevity, we can merge (S–10) and (S–11) into

$$D_T \overline{Z}_2' \overline{X}_2 D_T \xrightarrow{p} \Sigma_{\overline{Z}_2'\overline{Z}_2} = \begin{pmatrix} \omega & \Sigma_{zX_2} & 0 \\ \Sigma_{zX_2}' & \Sigma_{X_2X_2} & 0 \\ 0 & 0 & \sigma_{\varepsilon_1}^2 \end{pmatrix}, \tag{S–12}$$

where

$$\Sigma_{zX_2} = \begin{cases} 0, & \text{if } \alpha_2 \to 0 \\ \sqrt{-\alpha_2}E\left(Y_{2t-1}X_{2t}'\right), & \text{if } \alpha_2 < 0 \text{ and fixed.} \end{cases} \tag{S–13}$$

If $T\alpha_2 \to c \leq 0$, then

$$D_T \overline{Z}_2' \overline{X}_2 D_T \Rightarrow \Psi_{\overline{Z}_2'\overline{X}_2} = \begin{pmatrix} 2\omega\left(\int_0^1 \mathcal{J}_c d\mathcal{J}_c + 1\right) & 0 & 0 \\ 0 & \Sigma_{X_2X_2} & 0 \\ 0 & 0 & \sigma_{\varepsilon_1}^2 \end{pmatrix}. \tag{S–14}$$

Therefore, in both cases given in (S–12) and (S–14), $D_T \overline{Z}_2' \overline{X}_2 D_T$ is invertible with probability approaching one, and hence,

$$\left(D_T \overline{Z}_2' \overline{X}_2 D_T\right)^{-1} = O_p(1). \tag{S–15}$$

Next, by Lemma P(iii) and the Central Limit Theorem,

$$D_T \overline{Z}_2' v_2 = \begin{pmatrix} \sqrt{\kappa_T} z' v_2 \\ \sqrt{\frac{1}{T}} X_2' v_2 \\ \sqrt{\frac{1}{T}} \varepsilon_1' v_2 \end{pmatrix} = O_p(1). \tag{S–16}$$

Putting (S–15) and (S–16) together yields $\hat{\psi}_2 \xrightarrow{p} \psi_2$.

3.5 Proof of Proposition 6

To prove the second result, we can follow the steps of the proof of MPet Lemma 3.3. The conditional variance of ζ_{Tt} is given by

$$\sum_{t=1}^{T} E_{\mathcal{F}_{Tt-1}}\left[\zeta_{Tt}\zeta_{Tt}'\right] = A_T \xrightarrow{p} V_\zeta \tag{S–17}$$

where

$$A_{11,T} = \sum_{t=1}^{T} \kappa_T z_t^2 E_{\mathcal{F}_{Tt-1}}\left(\begin{pmatrix}\varepsilon_{1t}\\ v_{2t}\end{pmatrix}\begin{pmatrix}\varepsilon_{1t}\\ v_{2t}\end{pmatrix}'\right) \xrightarrow{p} \omega\begin{pmatrix}\sigma_{\varepsilon_1}^2 & 0\\ 0 & \sigma_{v_2}^2\end{pmatrix} = V_{\zeta,11},$$

$$A_{12,T} = \sum_{t=1}^{T} E_{\mathcal{F}_{Tt-1}}\left(\begin{pmatrix}\varepsilon_{1t}\\ v_{2t}\end{pmatrix}\varepsilon_{1t}\right)\sqrt{\frac{\kappa_T}{T}} z_t X_{1t}' \xrightarrow{p} \begin{pmatrix}\sigma_{\varepsilon_1}^2\\ 0\end{pmatrix}\Sigma_{zX_1} = V_{\zeta,12},$$

by (3) and (S–19),

$$A_{13,T} = \sum_{t=1}^{T}\sqrt{\frac{\kappa_T}{T}} z_t E_{\mathcal{F}_{Tt-1}}\left(\begin{pmatrix}\varepsilon_{1t}\\ v_{2t}\end{pmatrix}\left(\varepsilon_{1t}^2-\sigma_\varepsilon^2\right)\right) \xrightarrow{p} 0 = V_{\zeta,13},$$

if the distribution of ε_{1t} is not skewed,

$$A_{14,T} = \begin{pmatrix}\sum_{t=1}^{T} E_{\mathcal{F}_{Tt-1}}\left(\binom{\varepsilon_{1t}}{v_{2t}}v_{2t}\right)\sqrt{\frac{\kappa_T}{T}} z_t X_{2t}' & \sum_{t=1}^{T}\sqrt{\frac{\kappa_T}{T}} z_t E_{\mathcal{F}_{Tt-1}}\left(\binom{\varepsilon_{1t}}{v_{2t}}v_{2t}\varepsilon_{1t}\right)\end{pmatrix}$$
$$\xrightarrow{p} \begin{pmatrix}\binom{0}{\sigma_{v_2}^2}\Sigma_{zX_2} & 0\end{pmatrix} = V_{\zeta,14},$$

by (3) and (S–13),

$$A_{22,T} = \sum_{t=1}^{T}\frac{X_{1t}X_{1t}'}{T} E_{\mathcal{F}_{Tt-1}}\left(\varepsilon_{1t}^2\right) \xrightarrow{p} \Sigma_{X_1X_1}\sigma_{\varepsilon_1}^2 = V_{\zeta,22},$$

$$A_{23,T} = \sum_{t=1}^{T}\frac{X_{1t}}{T} E_{\mathcal{F}_{Tt-1}}\left(\varepsilon_{1t}\left(\varepsilon_{1t}^2-\sigma_\varepsilon^2\right)\right) \xrightarrow{p} 0 = V_{\zeta,23},$$

if the distribution of ε_{1t} is not skewed,

$$A_{24,T} = \begin{pmatrix}\sum_{t=1}^{T} E_{\mathcal{F}_{Tt-1}}\left(\varepsilon_{1t}v_{2t}\right)\frac{X_{1t}X_{2t}'}{T} & \sum_{t=1}^{T} E_{\mathcal{F}_{Tt-1}}\left(\varepsilon_{1t}^2 v_{2t}\right)\frac{X_{1t}}{T}\end{pmatrix} \xrightarrow{p} 0 = V_{\zeta,24},$$

$$A_{33,T} = \sum_{t=1}^{T} \frac{E_{\mathcal{F}_{Tt-1}}\left[\left(\varepsilon_{1t}^2 - \sigma_\varepsilon^2\right)^2\right]}{T} \xrightarrow{p} \varpi = V_{\zeta,33},$$

$$A_{34,T} = \left(\sum_{t=1}^{T} \frac{E_{\mathcal{F}_{Tt-1}}\left[\left(\varepsilon_{1t}^2 - \sigma_\varepsilon^2\right) v_{2t}\right] X_{2t}'}{T} \quad \sum_{t=1}^{T} \frac{E_{\mathcal{F}_{Tt-1}}\left[\left(\varepsilon_{1t}^2 - \sigma_\varepsilon^2\right) \varepsilon_{1t} v_{2t}\right]}{T} \right) \xrightarrow{p} 0 = V_{\zeta,34},$$

and

$$A_{44,T} = \sum_{t=1}^{T} \frac{E_{\mathcal{F}_{Tt-1}}\left(\binom{X_{2t}}{\varepsilon_{1t}}\binom{X_{2t}}{\varepsilon_{1t}}' v_{2t}^2\right)}{T} \xrightarrow{p} \begin{pmatrix} \Sigma_{X_2X_2} & 0 \\ 0 & \sigma_\varepsilon^2 \end{pmatrix} \sigma_v^2 = V_{\zeta,44}.$$

Putting these together, we have

$$V_\zeta = \begin{pmatrix} \omega \begin{pmatrix} \sigma_{\varepsilon_1}^2 & 0 \\ 0 & \sigma_{v_2}^2 \end{pmatrix} & \binom{\sigma_{\varepsilon_1}^2}{0} \Sigma_{zX_1} & 0 & \left(\binom{0}{\sigma_{v_2}^2} \Sigma_{zX_2} \quad 0 \right) \\ & \Sigma_{X_1X_1} \sigma_{\varepsilon_1}^2 & 0 & 0 \\ & & \varpi & 0 \\ & & & \begin{pmatrix} \Sigma_{X_2X_2} & 0 \\ 0 & \sigma_\varepsilon^2 \end{pmatrix} \sigma_v^2 \end{pmatrix}$$

Asymptotic normality of $\sum_{t=1}^{T} \zeta_{Tt}$ is established by verifying the Lindeberg condition in MPet Proposition A1, i.e.,

$$\sum_{t=1}^{T} E_{\mathcal{F}_{Tt-1}}\left(\|\zeta_{Tt}\|^2 1\{\|\zeta_{Tt}\| > \delta\}\right) \xrightarrow{p} 0 \qquad \delta > 0,$$

where

$$\|\zeta_{Tt}\|^2 = \kappa_T z_t^2 \left\| \begin{pmatrix} \varepsilon_{1t} \\ v_{2t} \end{pmatrix} \right\|^2 + \frac{\|X_{1t}\|^2 \varepsilon_{1t}^2}{T} + \frac{\left(\varepsilon_{1t}^2 - \sigma_\varepsilon^2\right)^2}{T} + \frac{\|X_{2t}\|^2 v_{2t}^2}{T} + \frac{\varepsilon_{1t}^2 v_{2t}^2}{T}.$$

The proof of this follows the same steps as the proof of MPet Lemma 3.3. Hence,

$$\sum_{t=1}^{T} \zeta_{Tt} \Rightarrow N\left(0, V_\zeta\right),$$

where V_ζ is given by (S–17).

Now, turn to the derivation of G_T. First, we need an expression for $D_T C_{\overline{Z}_2'\overline{Z}_2}$. Define

$W = (X_1, \varepsilon_1)$, so that

$$\overline{Z}_2'\overline{Z}_2 = \begin{pmatrix} z'z & z'W \\ W'z & W'W \end{pmatrix},$$

and

$$C_{\overline{Z}_2'\overline{Z}_2} = \begin{pmatrix} \sqrt{z'z} & 0 \\ \frac{W'z}{\sqrt{z'z}} & (W'M_zW)^{1/2} \end{pmatrix}.$$

Thus,

$$\begin{aligned} D_T C_{\overline{Z}_2'\overline{Z}_2} &= \begin{pmatrix} \sqrt{\kappa_T} & 0 \\ 0 & T^{-1/2} I_{p_{\psi_2}-1} \end{pmatrix} \begin{pmatrix} \sqrt{z'z} & 0 \\ \frac{W'z}{\sqrt{z'z}} & (W'M_zW)^{1/2} \end{pmatrix} \\ &= \begin{pmatrix} \sqrt{\kappa_T z'z} & 0 \\ \frac{T^{-1/2}W'z}{\sqrt{z'z}} & T^{-1/2}\left(W'M_zW\right)^{1/2} \end{pmatrix} \end{aligned} \tag{S–18}$$

It can be verified that its inverse is

$$\left(D_T C_{\overline{Z}_2'\overline{Z}_2}\right)^{-1} = \begin{pmatrix} \frac{1}{\sqrt{\kappa_T z'z}} & 0 \\ -\frac{(W'M_zW)^{-1/2}W'z}{\sqrt{\kappa_T z'z}\sqrt{z'z}} & T^{1/2}\left(W'M_zW\right)^{-1/2} \end{pmatrix}.$$

Hence, by simple algebra it can be verified that

$$G_T = \begin{pmatrix} \frac{1}{\sigma_{\varepsilon_1}\left(\kappa_T z' M_{X_1} z\right)^{1/2}} & 0 & -\frac{\sqrt{T\kappa_T} z' X_1 \left(X_1'X_1\right)^{-1}}{\sigma_{\varepsilon_1}\left(\kappa_T z' M_{X_1} z\right)^{1/2}} & 0 & 0 \\ 0 & 0 & \left(\frac{X_1'X_1}{T\sigma_{\varepsilon_1}^2}\right)^{-1/2} & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{\sqrt{\varpi}} & 0 \\ 0 & \frac{1}{\sigma_{v_2}(\kappa_T z'z)^{1/2}} & 0 & 0 & 0 \\ 0 & -\frac{(W'M_zW)^{-1/2}W'z}{\sqrt{\kappa_T} z'z\sigma_{v_2}} & 0 & 0 & \frac{1}{\sigma_{v_2}}\left(\frac{W'M_zW}{T}\right)^{-1/2} \end{pmatrix}$$

is such that $G_T \sum \zeta_{Tt} = \xi^*$.

Finally, using the above results, it can be verified that $G_T V_\zeta G_T \xrightarrow{p} I_k$.

The result that $\hat{\xi} \xrightarrow{d} N(0, I_k)$ follows Slutsky and the Continuous Mapping Theorem.

3.6 Proof of Proposition 7

We need to derive the asymptotic behavior of

$$B_T \hat{C}_{\hat{\psi}} = \begin{pmatrix} T^{-1/2}\left(X_1'X_1\right)^{1/2}\hat{\sigma}_{\varepsilon_1}^{-1} & 0 & -d_{21}T^{-1/2}X_1'\hat{Z}_2 C_{\hat{Z}_2'\hat{Z}_2}^{\prime -1}\hat{\sigma}_{v_2}^{-1} \\ 0 & \hat{\varpi}^{-1/2} & 0 \\ 0 & 0 & D_T \hat{X}_2'\hat{Z}_2 C_{\hat{Z}_2'\hat{Z}_2}^{\prime -1}\hat{\sigma}_{v_2}^{-1} \end{pmatrix}.$$

First, $T^{-1/2}\left(X_1'X_1\right)^{1/2}\hat{\sigma}_{\varepsilon_1}^{-1} \xrightarrow{p} \Sigma_{X_1'X_1}^{1/2}\sigma_{\varepsilon_1}^{-1}$ and $\hat{\varpi}^{-1/2} \xrightarrow{p} \varpi^{-1/2}$. Next, by Proposition 5,

$$T^{-1/2}X_1'\hat{Z}_2 C_{\hat{Z}_2'\hat{Z}_2}^{\prime -1}\hat{\sigma}_{v_2}^{-1} = T^{-1/2}X_1'\overline{Z}_2 D_T D_T^{-1} C_{\overline{Z}_2'\overline{Z}_2}^{\prime -1}\sigma_{v_2}^{-1} + o_p(1),$$

and

$$D_T \hat{X}_2'\hat{Z}_2 C_{\hat{Z}_2'\hat{Z}_2}^{\prime -1}\hat{\sigma}_{v_2}^{-1} = D_T \overline{X}_2'\overline{Z}_2 D_T D_T^{-1} C_{\overline{Z}_2'\overline{Z}_2}^{\prime -1}\sigma_{v_2}^{-1} + o_p(1).$$

Next, note that $D_T C_{\overline{Z}_2'\overline{Z}_2}$ is given in (S–18), or

$$D_T C_{\overline{Z}_2'\overline{Z}_2} = \begin{pmatrix} \sqrt{\kappa_T z'z} & 0 \\ \frac{\sqrt{\frac{\kappa_T}{T}}W'z}{\sqrt{\kappa_T z'z}} & \left(\frac{W'W}{T} - \frac{\sqrt{\frac{\kappa_T}{T}}W'z\sqrt{\frac{\kappa_T}{T}}z'W}{\kappa_T z'z}\right)^{1/2} \end{pmatrix}.$$

If $\alpha_2 < 0$ is fixed, then, by Lemma P(i) and (iv),

$$D_T C_{\overline{Z}_2'\overline{Z}_2} \xrightarrow{p} \begin{pmatrix} \sqrt{\omega} & 0 \\ \frac{\Sigma_{W'z}}{\sigma_z} & \left(\Sigma_{W'W} - \Sigma_{W'z}\sigma_z^{-2}\Sigma_{W'z}'\right)^{1/2} \end{pmatrix},$$

where $\sigma_z^2 = \omega / \left|\alpha_2\right|$,

$$\Sigma_{W'W} = \begin{pmatrix} \Sigma_{X_1'X_1} & 0 \\ 0 & \sigma_{\varepsilon_1}^2 \end{pmatrix}, \text{ and } \Sigma_{W'z} = \begin{pmatrix} E\left(X_{1t}z_t\right) \\ 0 \end{pmatrix}.$$

If $\alpha_2 \to 0$, then, by Lemma P(i) and (iv) and the fact that $\sqrt{\frac{\kappa_T}{T}} = o\left(T^{-1}\right)$,

$$D_T C_{\overline{Z}_2'\overline{Z}_2} \xrightarrow{p} \begin{pmatrix} \sqrt{\omega} & 0 \\ 0 & \Sigma_{W'W}^{1/2} \end{pmatrix}.$$

In both cases, the limiting matrix will be denoted by $C_{\Sigma_{\overline{Z}_2'\overline{Z}_2}}$ and is of full rank.

Next,

$$D_T \overline{Z}_2' X_1 T^{-1/2} = \begin{pmatrix} \sqrt{\frac{\kappa_T}{T}} z' X_1 \\ T^{-1} X_2' X_1 \\ T^{-1} \varepsilon_1' X_1 \end{pmatrix} \xrightarrow{p} \begin{pmatrix} \Sigma_{zX_1} \\ \Sigma_{X_2' X_1} \\ 0 \end{pmatrix} = \Sigma_{\overline{Z}_2' X_1},$$

where $\Sigma_{X_2' X_1} = \lim_{T\to\infty} E\left(X_{2t} X_{1t}'\right)$ and, by the same arguments as for (S–13),

$$\Sigma_{zX_1} = \begin{cases} 0, & \text{if } \alpha_2 \to 0 \\ \sqrt{-\alpha_2} E\left(Y_{2t-1} X_{1t}'\right), & \text{if } \alpha_2 < 0 \text{ and fixed.} \end{cases} \tag{S–19}$$

Finally, the limiting behavior of $D_T \overline{Z}_2' \overline{X}_2 D_T$ is given by (S–11) and (S–14). Putting all of these together, we have

$$B_T \hat{C}_{\hat{\psi}} \Rightarrow \begin{pmatrix} \Sigma_{X_1' X_1}^{1/2} \sigma_{\varepsilon_1}^{-1} & 0 & -d_{21} \Sigma_{\overline{Z}_2' X_1}' C_{\Sigma_{\overline{Z}_2' \overline{Z}_2}}'^{-1} \sigma_{v_2}^{-1} \\ 0 & \varpi^{-1/2} & 0 \\ 0 & 0 & \Psi_{33} \end{pmatrix},$$

where

$$\Psi_{33} = \begin{cases} \Sigma_{\overline{Z}_2' \overline{Z}_2}^{-1} C_{\Sigma_{\overline{Z}_2' \overline{Z}_2}}'^{-1} \sigma_{v_2}^{-1}, & \text{if } T\alpha_2 \to -\infty \\ \Psi_{\overline{Z}_2' \overline{X}_2}^{-1} C_{\Sigma_{\overline{Z}_2' \overline{Z}_2}}'^{-1} \sigma_{v_2}^{-1}, & \text{if } T\alpha_2 \to c \leq 0. \end{cases} \tag{S–20}$$

Hence, Ψ is invertible a.s., as required. In the case $T\alpha_2 \to c \leq 0$, Ψ is random due to the term $\Psi_{\overline{Z}_2' \overline{X}_2}$ defined in (S–14). The independence of Ψ from ξ then follows from Lemma P(ii) and (iii).

3.7 Proofs extending MPvic to general sequences

Lemmas 1*, 2*, 5* and 6* above are the counterparts – under general sequences – to MPvic-Lemmas 3.1, 3.2, 3.5 and 3.6. We provide below the proofs of the various lemmas by proving all the results in the Technical Appendix to MPvic. For readability and to avoid repeating the whole Appendix of MPvic, we delineate changes that should be read in relation to MPvic. The proofs are here presented in the univariate setting since this is the one we consider in the application but the results are also valid for the multivariate setting, as in MPvic. Note that the case c_T constant is not treated in MPvic but in KMS, Lemmas B2 and B4.

MPvic-Proposition A.1 holds since Assumption N*(*iii*) only intervenes in the definition of z_t, and the latter is unaffected by the change (as opposed to $\tilde{z}_t$).

MPvic-Proposition A.2. Equation (MPvic-42) holds with (MPvic-43) such that in the univariate case

$$\sup_{1\leq t\leq T}\sum_{j=1}^{t}\rho_T^{t-j}=\frac{1-\rho_T^T}{1-\rho_T}=\begin{cases}O\left(-c_T^{-1}\right), & \text{if } Tc_T\to-\infty\\ O\left(T\right), & \text{if } Tc_T\to c<0\\ O\left(T\right), & \text{if } Tc_T\to 0\end{cases}$$
$$=O\left(T_\wedge\left|c_T^{-1}\right|\right).$$

Now, if z_t is less persistent than the regressor ($c_T=o\left(T^{-b}\right)$), then

$$\sup_{1\leq t\leq T}E\left(\psi_{Tt}^2\right)=O\left(\frac{T^{2b}}{c_T}\right),$$

and when $T^{-b}=O\left(c_T\right)$

$$\sup_{1\leq t\leq T}E\left(\psi_{Tt}^2\right)=O\left(\frac{T^{b}}{c_T^2}\right),$$

so (MPvic-40) writes:

$$\sup_{1\leq t\leq T}E\left(\psi_{Tt}^2\right)=O\left(\frac{T^{b}}{c_T}\left[T^b{}_\wedge\left|c_T^{-1}\right|\right]\right). \tag{S–21}$$

Now for (MPvic-41), we need to consider

$$E\left\|\frac{1}{T^{1/2}\sqrt{\frac{T^b}{-c_T}\left[T^b{}_\vee\left|c_T^{-1}\right|\right]}}\sum_{t=1}^{T}\psi_{Tt}\varepsilon_t\right\|^2\leq\frac{E\left\|\varepsilon_1\right\|^2\sup_{1\leq t\leq T}E\left\|\psi_{Tt}\right\|^2}{\frac{T^b}{c_T}\left[T^b{}_\vee\left|c_T^{-1}\right|\right]}$$
$$=O\left(\frac{T^b{}_\wedge\left|c_T^{-1}\right|}{T^b{}_\vee\left|c_T^{-1}\right|}\right)$$
$$=O\left(1\right).$$

Now, regarding $\sum_{t=1}^{T}\Delta\tilde{\varepsilon}_t\psi_{Tt}$, we need, for all $c_T=o\left(1\right)$, the following:

$$\sum_{t=1}^{T}\tilde{\varepsilon}_tx_t=O_p\left(T\right). \tag{S–22}$$

As in MPvic, this holds from Phillips (1987) under N*(i)-(ii). With serially dependent innovations, we refer to GP12-Theorem 2.2(ii) which shows that under N*(iii)

$\sum_{t=1}^{T}\tilde{\varepsilon}_t x_t = O_p\left(\left(c_T^3 T\right)^{-1/2}\right) = o\left(T\right).$ The framework of GP12 assumes $c_T \in [-1, 0]$. It is easy to see that if $x_0 = o_p\left(\sqrt{\frac{T}{1-Tc_T}}\right)$, (S–22) holds under N*($iii$) since there exists T_0 such that $c_T \in [-2, 0]$ for all $T > T_0$ and hence we can decompose the sample moments computed over $t = 1, ..., T_0$ and $T_0, ..., T$ where only the latter use the asymptotic results of GP12, the former becoming negligible.

Now,

$$\frac{1}{T^{1/2}\sqrt{\frac{T^b}{-c_T}\left[T^b \vee \left|c_T^{-1}\right|\right]}}\sum_{t=1}^{T}\Delta\tilde{\varepsilon}_t\psi_{Tt} = \frac{1}{T^{1/2}\sqrt{\frac{T^b}{-c_T}\left[T^b \vee \left|c_T^{-1}\right|\right]}}\frac{-c_z}{T^b}\sum_{t=1}^{T}\tilde{\varepsilon}_t\psi_{Tt} + o_p\left(1\right),$$

where following MPvic,

$$\begin{aligned}
\left\|\frac{1}{T^{1/2}\sqrt{\frac{T^b}{-c_T}\left[T^b \vee \left|c_T^{-1}\right|\right]}}\frac{c_z}{T^b}\sum_{t=1}^{T}\tilde{\varepsilon}_t\psi_{Tt}\right\|_{L_1} &\leq \frac{E\left\|\varepsilon_1\right\|^2}{T^{1/2}\sqrt{\frac{T^b}{-c_T}\left[T^b \vee \left|c_T^{-1}\right|\right]}}\frac{1}{T^b}T\left(\sup_{1\leq t\leq T}E\left\|\psi_{Tt}\right\|^2\right)^{1/2}\\
&= \frac{T^{1/2}E\left\|\varepsilon_1\right\|^2}{T^b\sqrt{\frac{T^b}{-c_T}\left[T^b \vee \left|c_T^{-1}\right|\right]}}\left(\sup_{1\leq t\leq T}E\left\|\psi_{Tt}\right\|^2\right)^{1/2}\\
&\leq O\left(\frac{1}{T^{b-1/2}}\frac{\sqrt{\frac{T^b}{-c_T}\left[T^b \wedge \left|c_T^{-1}\right|\right]}}{\sqrt{\frac{T^b}{-c_T}\left[T^b \vee \left|c_T^{-1}\right|\right]}}\right)\\
&= O\left(\frac{1}{T^{b-1/2}}\sqrt{\frac{T^b \wedge \left|c_T^{-1}\right|}{T^b \vee \left|c_T^{-1}\right|}}\right) = O\left(\frac{1}{T^{b-1/2}}\right),
\end{aligned}$$

hence for $b \in (1/2, 1)$ the equation above is $o\left(1\right)$. Hence MPvic-Proposition A.2 holds, with

$$\frac{1}{T^{1/2}\sqrt{\frac{T^b}{-c_T}\left[T^b \vee \left|c_T^{-1}\right|\right]}}\sum_{t=1}^{T}u_t\psi_{Tt} \xrightarrow{p} 0, \text{ when } b \in \left(1/2, 1\right).$$

MPvic-Lemma 3.1. The proof then follows. It uses the fact that

$$\sup_{s\in[0,1]}\left\|x_{\lfloor sT\rfloor}\right\| = O_p\left(\sqrt{\frac{T}{1-Tc_T}}\right), \tag{S–23}$$

i.e., $\sup_{s\in[0,1]}\left\|x_{\lfloor sT\rfloor}\right\| = O_p\left(\left|c_T\right|^{-1/2}\right)$ when $Tc_T \to -\infty$ and $O_p\left(T^{1/2}\right)$ otherwise, see GP12, Expression (2.13) of Lemma 2.1 under assumption N*(iii) and Phillips (1987) under N*(i)-(ii). Hence

$$\sup_{1\leq t\leq T}\|\psi_{Tt}\| = O_p\left(\sqrt{\frac{T^{1+2b}}{1-Tc_T}}\right).$$

For part (i) of the lemma, we use

$$\frac{1}{T^{\frac{1+b}{2}}}\left(\sum_{t=1}^{T}u_t\tilde{z}_t - \sum_{t=1}^{T}u_{0t}z_t\right) = \frac{c_T}{T^{\frac{1+b}{2}}}\sum_{t=1}^{T}u_t\psi_{Tt} = o_p(1),$$

from the extension to MPvic-Proposition A.2 above.

For part (ii), this involves (MPvic-18) which requires under N*(iii)

$$T^{-1}\sum_{t=1}^{T}x_{t-1}\varepsilon_t \xrightarrow{p} 0, \tag{S–24}$$

where $E(x_{t-1}\varepsilon_t) = 0$. When $c_T \to 0$, this holds by virtue of GP12, Theorem 2.2. Indeed, GP12 show that the estimators of the autocovariance of x_t are consistent, so in particular (S–24) must hold. When $\lim_{T\to\infty} c_T < 0$, the results hold since $x_{t-1}\varepsilon_t$ is a martingale difference sequence with bounded variance. Hence part (ii) of Lemma MPvic-3.1 writes here

$$T^{-(1+b)}\sum_{t=1}^{T}x_t\tilde{z}_t = T^{-(1+b)}\sum_{t=1}^{T}x_t z_t - \frac{c_T c_z}{T}\sum_{t=1}^{T}x_t^2 + o_p(1).$$

For part (iii) of the lemma,

$$\begin{aligned}\frac{1}{T^{1+b}}\left\|\sum_{t=1}^{T}\tilde{z}_t^2 - \sum_{t=1}^{T}z_t^2\right\|^2 &\leq \frac{c_T^2}{T^{1+b}}\sum_{t=1}^{T}\|\psi_{Tt}\|^2 - 2\frac{c_T}{T^{1+b}}\sum_{t=1}^{T}\|\psi_{Tt}\|\|z_t\| \\ &\leq \left(\frac{c_T\sup_{t\leq T}\|\psi_{Tt}\|}{T^{b/2}}\right)^2 + \left(\frac{-c_T\sup_{t\leq T}\|\psi_{Tt}\|}{T^{b/2}}\right)O_p(1),\end{aligned}$$

where we used the Lyapunov inequality as in MPvic. Now $\sup_{1\leq t\leq T}\|\psi_{Tt}\| = O_p\left(\sqrt{\frac{T^{1+2b}}{1-Tc_T}}\right)$ so

$$\frac{-c_T\sup_{t\leq T}\|\psi_{Tt}\|}{T^{b/2}} = O_p\left(\sqrt{-c_T T^{(1+b)/2}}\right) = o_p(1),$$

since z_t is less persistent than the regressor.

MPvic-Theorem 3.4: we need the asymptotic behavior of

$$L_T = \frac{-c_T}{T} \sum_{t=1}^{T} x_t^2$$

under N*(iii). GP12-Theorem 2.2 shows that the estimator of the variance of x_t is consistent and GP12-Lemma 2.1 shows that $var\left(x_t\right) = O\left(\left|c_T^{-1}\right|\right)$, hence

$$\frac{-c_T}{T} \sum_{t=1}^{T} x_t^2 \xrightarrow{p} \Omega_{vv}.$$

The rest follows as in MPvic.

MPvic-Lemma 3.2 hence also holds, where the rate of convergence is $\frac{-c_T}{T}\sum_{t=1}^{T} x_{t-1}^2$. Joint convergence follows from MPvic-Lemma 3.2, when there exists $c \leq 0$ such that $T\alpha_2 \to c$, and from applying Theorem 2.2 of GP12 when $T\alpha_2 \to -\infty$.

MPvic-Lemma 3.5 uses the decomposition

$$\tilde{z}_t = x_t - \rho_z^t x_0 + \frac{c_z}{T^b}\psi_{Tt},$$

in (i)

$$\begin{aligned}
\sqrt{\frac{-c_T}{T}}\left(\sum_{t=1}^{T} u_t\tilde{z}_t - \sum_{t=1}^{T} u_t x_t\right) &= \sqrt{\frac{-c_T}{T}}\left[\frac{c_z}{T^b}\sum_{t=1}^{T} u_t\psi_{Tt} - \sum_{t=1}^{T} u_t x_0 \rho_z^t\right] \\
&= \frac{\sqrt{-c_T}}{T^{1/2+b}} c_z \sum_{t=1}^{T} u_t\psi_{Tt} + o_p\left(\frac{\sqrt{-c_T}}{T^{1/2}} T^{b/2}\sqrt{\frac{1}{-c_T}}\right) \\
&= \frac{\sqrt{-c_T}}{T^{1/2+b}} c_z \sum_{t=1}^{T} u_t\psi_{Tt} + o_p\left(\frac{1}{T^{(1-b)/2}}\right),
\end{aligned}$$

assuming $x_0 = o_p\left(\sqrt{T/\left(1 - Tc_T\right)}\right)$ and using $\sum_{t=1}^{T} u_t\rho_z^t = O_p\left(T^{b/2}\right)$ as in MPvic. The extension to MPvic-Proposition A.2 above shows that when the regressor is less persistent than the instrument

$$\sum_{t=1}^{T} u_t\psi_{Tt} = o_p\left(T^{1/2+b}\left|c_T\right|^{-1/2}\right),$$

QED.

Now for part (ii),

$$\frac{c_T}{T}\left(\sum_{t=1}^{T} x_t\tilde{z}_t - \sum_{t=1}^{T} x_t^2\right) = \frac{c_T}{T}\left[\frac{c_z}{T^b}\sum_{t=1}^{T} x_t\psi_{Tt} - \sum_{t=1}^{T} x_t x_0 \rho_z^t\right]$$
$$= \frac{c_T}{T^{1+b}} c_z \sum_{t=1}^{T} x_t \psi_{Tt}' + o_p\left(\frac{1}{T^{1-b}}\right),$$

as $\sup_{t\leq T}\|x_t\| = O_p\left(\sqrt{\frac{T}{1-Tc_T}}\right)$, $x_0 = o_p\left(\sqrt{\frac{T}{1-Tc_T}}\right)$ and $\sum_{t=1}^{T}\rho_z^t = O\left(T^b\right)$. For the leading term, GP12-Lemma 2.1 shows that

$$\sup_{1\leq t\leq T} E\|x_t\|^2 = O\left(\left|c_T^{-1}\right|\right).$$

Hence, using Proposition A.2.

$$\left\|\frac{c_T}{T^{1+b}} c_z \sum_{t=1}^{T} x_t\psi_{Tt}\right\|_{L_1} \leq O_p\left(\frac{-c_T}{T^b}\left(\frac{T^b}{c_T^2}\frac{1}{-c_T}\right)^{1/2}\right)$$
$$= O_p\left(\frac{1}{|c_T|^{1/2} T^{b/2}}\right) = o_p(1).$$

Finally for $\sum_{t=1}^{T}\tilde{z}_t^2$, as in MP we only need to consider

$$\left\|\frac{c_T}{T^{1+b}}\sum_{t=1}^{T}\psi_{Tt}x_0\rho_z^t\right\| = o_p\left(\frac{-c_T}{T^{1+b}}\sqrt{\frac{1}{-c_T}}\sqrt{\frac{T}{1-Tc_T}}T^b\right) = o_p(1),$$

and when $c_T T^b \to -\infty$,

$$E\left\|\frac{c_T}{T^{1+2b}}\sum_{t=1}^{T}\psi_{Tt}^2\right\| \leq \frac{-c_T}{T^{2b}}\frac{T^b}{c_T^2} = \frac{1}{-c_T T^b} = o(1).$$

MPvic-Lemma 3.6 The results of MPvic hold when $c_T = \kappa T^{-b}$ but we need to consider the case where $c_T = \kappa_T T^{-b}$ with $\kappa_T \in (M, 0)$, for $M < 0$. Then Expression MPvic-(48) becomes

$$\tilde{z}_t = \rho_z\tilde{z}_{t-1} + v_t + \frac{\kappa_T}{T^b}x_{t-1}.$$

This implies

$$\left(1-\rho_z\rho_T\right)\frac{1}{T}\sum_{t=1}^{T}\tilde{z}_{t-1}x_{t-1}=\frac{1}{T}\sum_{t=1}^{T}x_{t-1}v_t+\frac{1}{T}\sum_{t=1}^{T}v_t z_{t-1}+\frac{1}{T}\sum_{t=1}^{T}v_t^2+\frac{\kappa_T}{T^{1+b}}\sum_{t=1}^{T}x_{t-1}^2,$$

where $1-\rho_z\rho_T=-T^{-b}\left(c_z+\kappa_T\right)$. GP12 Lemma 2.1 and Theorem 2.2(i) imply that

$$\frac{\kappa_T}{T^{1+b}}\sum_{t=1}^{T}x_{t-1}^2\xrightarrow{p}-\frac{1}{2}\Omega_{vv}.$$

Also, notice that $T^{-1}\sum_{t=2}^{T}x_{t-1}v_t=T^{-1}\left(\sum_{t=2}^{T}x_t x_{t-1}-\rho_T\sum_{t=2}^{T}x_{t-1}^2\right)$. The same lemma and theorem in GP12 can therefore be used to obtain the results in MPvic that

$$\frac{1}{T}\sum_{t=1}^{T}x_{t-1}v_t+\frac{1}{T}\sum_{t=1}^{T}v_t z_{t-1}+\frac{1}{T}\sum_{t=1}^{T}v_t^2\xrightarrow{p}\Omega_{vv}.$$

Therefore

$$-\left(c_z+\kappa_T\right)T^{-(1+b)}\sum_{t=1}^{T}\tilde{z}_{t-1}x_{t-1}\xrightarrow{p}\frac{1}{2}\Omega_{vv}.$$

which proves part (i).

Now for part (ii),

$$\left(1-\rho_z^2\right)T^{-1}\sum_{t=1}^{T}\tilde{z}_{t-1}^2=\left(1+o_p\left(1\right)\right)T^{-1}\left\{2\sum_{t=1}^{T}\left(v_t+\frac{\kappa_T}{T^b}x_{t-1}\right)\tilde{z}_{t-1}+\sum_{t=1}^{T}\left(v_t+\frac{\kappa_T}{T^b}x_{t-1}\right)^2\right\},$$

where $T^{-1}\sum_{t=1}^{T}\left(v_t+\frac{\kappa_T}{T^b}x_{t-1}\right)^2\xrightarrow{p}E\left(v_t^2\right)$, and

$$\begin{aligned}T^{-1}\sum_{t=1}^{T}\left(v_t+\frac{\kappa_T}{T^b}x_{t-1}\right)\tilde{z}_{t-1}&=T^{-1}\sum_{t=1}^{T}v_t\tilde{z}_{t-1}+\frac{\kappa_T}{T^{1+b}}\sum_{t=1}^{T}x_{t-1}\tilde{z}_{t-1}\\&=\Lambda_{vv}+\frac{-\kappa_T}{2\left(c_z+\kappa_T\right)}\Omega_{vv}+o_p\left(1\right),\end{aligned}$$

where $\Lambda_{vv}=\sum_{h=1}^{\infty}E\left(v_t v_{t-h}\right)$.

Collecting all elements, $-2c_zT^{-(1+b)}\sum_{t=1}^{T}\tilde{z}_{t-1}^2=\left[1+\frac{-\kappa_T}{\left(c_z+\kappa_T\right)}\right]\Omega_{vv}+o_p\left(1\right)$, i.e.,

$$-\left(c_z+\kappa_T\right)T^{-(1+b)}\sum_{t=1}^{T}\tilde{z}_{t-1}^2\xrightarrow{p}\frac{1}{2}\Omega_{vv}.$$

For part (iii), the results follow the same lines (including the extension to MPvic-Proposition A.2 above) and hence

$$\sqrt{-(c_z+\kappa_T)}T^{-\frac{1+b}{2}}\sum \tilde{z}_{t-1}u_t \xrightarrow{L} N\left(0,\frac{1}{2}\Omega_{vv}\Omega_{uu}\right).$$

MPvic-Lemma 4.2. The case where $c_T = O\left(T^{-b}\right)$ is considered by MPvic. Only the case $c_T T^b \to -\infty$ is new. We saw previously that

$$J_n = T^{-1}\sum_{t=1}^{T} x_{t-1}\varepsilon_t = o_p(1),$$

and $\frac{1}{c_T}(1-\rho\rho_z) = \frac{1}{c_T}\left(1-(1+c_T)\left(1+c_z T^{-b}\right)\right) \to -1$. Hence,

$$\frac{-c_T}{T}\sum_{t=1}^{T} x_{t-1}z_{t-1} \xrightarrow{p} \Omega_{vv},$$

and the results in MPvic hold, replacing T^{-a} with $-c_T$ and c_z with -1.

4 Finite sample corrections in the presence of intercepts

The finite sample correction in KMS, applied to the $AR(b_{12}^0)$ in (9) consists in modifying $P_{M_{X_1}z}$ in the numerator. When the model contains an intercept, let $X_1 = \left[\iota : \tilde{X}_1\right]$, where ι is a T_1-dimensional vector of ones (T_1 is the number of observations used in the regressions). The numerator of the AR statistic involves an estimator the inverse of the variance of $(z'M_{X_1}z)^{-1} z'M_{X_1}\varepsilon_1$ conditional on the process $\{u_{2t}\}$. We notice

$$M_{X_1}z = M_{\tilde{X}_1}M_\iota z = M_{\tilde{X}_1}(z-\iota\overline{z}_T).$$

with $\overline{z} = T_1^{-1}\sum_{t=\max(m,2)}^{T} z_t$. In KMS, $M_{\tilde{X}_1}$ does not appear. They show that in $z'M_\iota\varepsilon_1 = z'\varepsilon_1 - T_1\overline{z}\,\overline{\varepsilon}_1$, the long-run covariance between z_t and ε_{1t} which asymptotically appears via the product $T_1\overline{z}\,\overline{\varepsilon}_1$ vanishes asymptotically but matters in finite samples. They hence suggest using, instead of $P_{M_\iota z}$, the corrected

$$\widetilde{P}_{M_\iota z} = M_\iota z\left(z'z - T_1\left(1-\widehat{\rho}^2_{\varepsilon_1,u_2}\right)\overline{z}\overline{z}'\right)^{-1} z'M_\iota, \tag{S–25}$$

where $\widehat{\rho}_{\varepsilon_1,u_2}$ is the estimated long run correlation between ε_{1t} and u_{2t}. In (S–25) the term$\left(1-\widehat{\rho}^2_{\varepsilon_1,u_2}\right)$ accounts for the long term variance of $\sum_t \varepsilon_{1t}$ conditional on the process Y_{2t-1} (or z_t).

In the context of the AR statistic, this correction becomes

$$\widetilde{P}_{M_{X_1}z} = M_{X_1}z\left(z'M_{\tilde{X}_1}z-\left(1-\widehat{\rho}^2_{\varepsilon_1,u_2}\right)T_1\overline{z}\overline{z}'\right)^{-1}z'M_{X_1}, \tag{S–26}$$

where we considered only the higher order term $\overline{z}\overline{z}'$ instead of $\overline{M_{\tilde{X}_1}zM_{\tilde{X}_1}z}'$.

A similar correction can be applied to the statistic $W\left(b^0_{12}\right)$, where the adjustment now bears on $\hat{V}_{\hat{\psi},33}\left(b_{12}\right)$ defined in (25). For ease of exposition, we consider the hypothesis $H^*_0 : r\left(\theta\right)=0,\ b_{12}=b^0_{12}$ where $r\left(\theta\right)=\alpha_2-\alpha^0_2$ in equation (3) since assumptions concerning α_2 are the only ones that bear finite sample adjustments in $W\left(b^0_{12}\right)$. Now

$$\hat{\psi}_2=\left(\hat{Z}'_2\hat{X}_2\right)^{-1}\hat{Z}'_2\left(\overline{X}_2\psi_2+\varepsilon_2\right)$$

and, denoting $\hat{X}_{21}=\left[X_2:\hat{\varepsilon}_1\right]$,

$$\begin{aligned}\hat{\alpha}_2 &= \left(\hat{Z}'_2M_{\hat{X}_{21}}\hat{X}_2\right)^{-1}\hat{Z}'_2M_{\hat{X}_{21}}\left(\overline{X}_2\psi_2+\varepsilon_2\right)\\ &= \left(\hat{Z}'_2M_{\hat{X}_{21}}\hat{X}_2\right)^{-1}\hat{Z}'_2M_{\hat{X}_{21}}\left(\hat{X}_2\psi_2+\left(\overline{X}_2-\hat{X}_2\right)\psi_2+\varepsilon_2\right).\end{aligned}$$

Hence, $\hat{\alpha}_2-\alpha_2=\left(\hat{Z}'_2M_{\hat{X}_{21}}\hat{X}_2\right)^{-1}\hat{Z}'_2M_{\hat{X}_{21}}\left(\varepsilon_2+\left(\varepsilon_1-\hat{\varepsilon}_1\right)d_{21}\right)$, i.e.,

$$\hat{\alpha}_2-\alpha_2=\left(\hat{Z}'_2M_{\hat{X}_{21}}\hat{X}_2\right)^{-1}\hat{Z}'_2M_{\hat{X}_{21}}\left(\varepsilon_2+P_{X_1}\varepsilon_1d_{21}\right).$$

In our model, $X_1=X_2$ hence $M_{\hat{X}_{21}}P_{X_1}=0$, and

$$\hat{\alpha}_2-\alpha_2=\left(\hat{Z}'_2M_{\hat{X}_{21}}\hat{X}_2\right)^{-1}\hat{Z}'_2M_{\hat{X}_{21}}\varepsilon_2.$$

The variance of $\hat{\alpha}_2-\alpha_2$ is

$$V_{\hat{\alpha}_2}=\left(\hat{Z}'_2M_{\hat{X}_{21}}\hat{X}_2\right)^{-1}\hat{Z}'_2M_{\hat{X}_{21}}\hat{Z}_2\left(\hat{X}'_2M_{\hat{X}_{21}}\hat{Z}_2\right)^{-1}\sigma^2_{\varepsilon_2},$$

so the W statistic is

$$\left(\hat{\alpha}_2-\alpha_2\right)' V_{\hat{\alpha}_2}^{-1}\left(\widehat{a}-a\right)=\frac{\varepsilon_{2t}' M_{\hat{X}_{21}}\hat{Z}_2\left(\hat{Z}_2' M_{\hat{X}_{21}}\hat{Z}_2\right)^{-1}\hat{Z}_2' M_{\hat{X}_{21}}\varepsilon_{2t}}{\sigma_{\varepsilon_2}^2}.$$

The final sample approximation of KMS consists in replacing $\left(\hat{Z}_2' M_{\hat{X}_{21}}\hat{Z}_2\right)^{-1}$ with

$$\left(z'z-T\left(1-\hat{\rho}_{\varepsilon_2,u_2}^2\right)\overline{z}\overline{z}'\right)^{-1},$$

where $\hat{\rho}_{\varepsilon_2,u_2}$ is the estimate of the long run correlation between ε_{2t} and u_{2t} such that $1-\hat{\rho}_{\varepsilon_2,u_2}^2=\hat{\rho}_{\varepsilon_1,u_2}^2$. The Wald statistic becomes

$$W\left(\alpha_2\right)=\frac{\varepsilon_{2t}' M_{\hat{X}_{21}}\hat{Z}_2\left(z'z-T\left(1-\hat{\rho}_{\varepsilon_2,u_2}^2\right)\overline{z}\overline{z}'\right)^{-1}\hat{Z}_2' M_{\hat{X}_{21}}\varepsilon_{2t}}{\sigma_{\varepsilon_2}^2}.$$

which is in practice obtained as

$$\begin{aligned}\hat{W}\left(\alpha_2\right)&=\frac{\left(\hat{\alpha}_2-\alpha_2\right)'\left(\hat{X}_2' M_{\hat{X}_{21}}\hat{Z}_2\right)\left(z'z-T\left(1-\hat{\rho}_{\varepsilon_2,u_2}^2\right)\overline{z}\overline{z}'\right)^{-1}\left(\hat{Z}_2' M_{\hat{X}_{21}}\hat{X}_2\right)\left(\hat{\alpha}_2-\alpha_2\right)}{\hat{\sigma}_{\varepsilon_2}^2}\\&=\frac{\left(\hat{\alpha}_2-\alpha_2\right)'\left(\hat{X}_2' M_{\hat{X}_{21}}\hat{Z}_2\right)\left(\breve{z}'\breve{z}+T\hat{\rho}_{\varepsilon_2,u_2}^2\overline{z}\overline{z}'\right)^{-1}\left(\hat{Z}_2' M_{\hat{X}_{21}}\hat{X}_2\right)\left(\hat{\alpha}_2-\alpha_2\right)}{\hat{\sigma}_{\varepsilon_2}^2},\end{aligned}$$

where $\breve{z}=z-\iota\overline{z}'$. KMS do not consider the presence of additional regressors and lags in the equation. In our setting, the final sample approximation above should hence preferably be replaced with

$$\hat{W}\left(\alpha_2\right)=\frac{\varepsilon_{2t}' M_{\hat{X}_{21}}\hat{Z}_2\left(z' M_{\hat{X}_{21}}z+T\hat{\rho}_{\varepsilon_2,u_2}^2\overline{z}\overline{z}'\right)^{-1}\hat{Z}_2' M_{\hat{X}_{21}}\varepsilon_{2t}}{\hat{\sigma}_{\varepsilon_2}^2},$$

where $\hat{\sigma}_{\varepsilon_2}^2$ can possibly be replaced with the corresponding estimate of the long run variance.

Now for the general case, the results above combine into $\hat{V}_{\hat{\psi},33}\left(b_{12}\right)$ whose finite sample adjustment becomes:

$$\hat{V}_{\hat{\psi},33}\left(b_{12}\right)=\left(\hat{Z}_2'\hat{X}_2\right)^{-1}\left(\left[\hat{Z}_2'\hat{Z}_2+T\hat{\rho}_{\varepsilon_2,u_2}^2\overline{z}\overline{z}'\right]\hat{\sigma}_{\varepsilon_2}^2+\left[\hat{Z}_2' P_{X_1}\hat{Z}_2+T\hat{\rho}_{\varepsilon_1,u_2}^2\overline{z}\overline{z}'\right]\hat{\sigma}_{\varepsilon_1}^2 d_{12}\right)\left(\hat{X}_2'\hat{Z}_2\right)^{-1}. \quad \text{(S–27)}$$

	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 S.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(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.

5 Supplementary material for numerical section

We report additional simulation results on sizes of the AR and ARW tests with filtered instruments versus the conventional t test with standard (unfiltered) instruments for the bivariate SVAR described in the paper, with some variations.

5.1 Null rejection frequencies for the AR test

First, we report results on the null rejection frequency of the AR test of $H_0 : b_{12} = 0$ against $H_1 : b_{12} \neq 0$ when the estimated model is SVAR(2) or SVAR(4) and the DGP is exactly as in Section 4 in the paper. The results are reported in Tables S.1 and S.2, and they are comparable directly with Table 1 in the paper.

Next, we consider the case in which DGP may have a linear trend, i.e., the observed data is $\tilde{Y}_{2t} = Y_{2t} + \gamma_0 + \gamma_x t$, and the SVAR is estimated on sample-detrended data $\hat{Y}_{2t} = \tilde{Y}_{2t} - \hat{\gamma}_0 - \hat{\gamma}_1 t$, where $\hat{\gamma}_0$ and $\hat{\gamma}_1$ are full-sample or recursive OLS estimates. The true value of γ_0 is set to zero w.l.o.g. (since the statistics are invariant to the value of the constant), and γ_x is either 0 or 1.

Table S.3 reports results when the model is SVAR(1) and $\gamma_x = 0$. In Table S.4, the model is SVAR(1) and $\gamma_x = 1$. In each table, we present two cases: recursive detrending

	At 5%				At 10%			
	$\rho = 0.20$		0.95		0.20		0.95	
	AR	t	AR	t	AR	t	AR	t
$c = 0$	0.042	0.008	0.054	0.765	0.089	0.027	0.114	0.799
-1	0.039	0.009	0.050	0.668	0.086	0.030	0.108	0.708
-10	0.030	0.019	0.031	0.261	0.073	0.056	0.068	0.305
-30	0.020	0.035	0.024	0.147	0.054	0.084	0.063	0.186
-100	0.018	0.045	0.039	0.101	0.050	0.096	0.095	0.132

Table S.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(4) with long-run restrictions. ρ is the correlation between the reduced-form VAR errors. The sample size is 200. Number of MC replications: 20000.

(top panel), and full-sample detrending (bottom panel).

Overall, Tables S.3 and S.4 show that, no matter $\gamma_x = 0$ or 1, the outcome is the same: recursive detrending and AR controls size reasonably, and recursive detrending performs better than full sample detrending.

Tables S.5 and S.6 report the counterparts of Tables S.3 and S.4 when the estimated model is SVAR(m), for $m = 2$ and 4 with recursive detrending.

5.2 Size of the projection ARW test

The simulations for the size of the projection ARW test of the hypothesis $H_0 : d_{21} = d_{21}^0$ against $H_1 : d_{21} \neq d_{21}^0$ are based on the following 4-dimensional grid. The grid contains 21 points for $d_{21} \in [-1, 1]$ in steps of 0.1, 21 points for $\rho \in \{-.99, -.9, ..., .9, .99\}$, 7 points for $\omega_1 \in \{.1, .4, .7, 1, 4, 7, 10\}$ and 14 points for $c \in \{-200, -150, -100, -50, -40, -30, -20, -10, -5, -4, -3, -2, -1, 0\}$. Because the ARW statistic is invariant to ω_2, we normalize w.l.o.g. this parameter to 1. The parameter b_{12} in the DGP can then be obtained as a function of ρ, ω_1 and d_{21}.

Figure S.1 reports maximal rejection frequencies across ρ, ω_1 and c of the projection

$\gamma_x = 0$, recursive detrending								
	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	0.009	0.046	0.158	0.101	0.032	0.092	0.194
−1	0.051	0.009	0.044	0.139	0.102	0.031	0.090	0.173
−10	0.052	0.016	0.045	0.102	0.103	0.049	0.092	0.133
−30	0.053	0.033	0.049	0.078	0.103	0.078	0.098	0.112
−100	0.056	0.049	0.051	0.056	0.107	0.100	0.101	0.100

$\gamma_x = 0$, full sample detrending								
	At 5%				At 10%			
	$\rho = 0.20$		0.95		0.20		0.95	
	AR	t	AR	t	AR	t	AR	t
$c = 0$	0.060	0.021	0.221	0.946	0.112	0.063	0.343	0.958
−1	0.056	0.020	0.170	0.890	0.110	0.060	0.276	0.911
−10	0.053	0.027	0.072	0.405	0.105	0.071	0.131	0.465
−30	0.050	0.038	0.052	0.192	0.100	0.087	0.101	0.247
−100	0.052	0.050	0.048	0.084	0.101	0.100	0.095	0.132

Table S.3: 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. Y_2 is detrended recursively (top panel) or over the full-sample (bottom panel). The true coefficient on the trend is $\gamma_x = 0$. The sample size is 200. Number of MC replications: 20000.

$\gamma_x = 1$, recursive detrending								
	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	0.009	0.046	0.158	0.101	0.032	0.092	0.194
-1	0.051	0.009	0.044	0.139	0.102	0.031	0.090	0.173
-10	0.052	0.016	0.045	0.102	0.103	0.049	0.092	0.133
-30	0.053	0.033	0.049	0.078	0.103	0.078	0.098	0.112
-100	0.056	0.049	0.051	0.056	0.107	0.100	0.101	0.100

$\gamma_x = 1$, full sample detrending								
	At 5%				At 10%			
	$\rho = 0.20$		0.95		0.20		0.95	
	AR	t	*AR*	t	*AR*	t	*AR*	t
$c = 0$	0.060	0.021	0.221	0.946	0.112	0.063	0.343	0.958
-1	0.056	0.020	0.170	0.890	0.110	0.060	0.276	0.911
-10	0.053	0.027	0.072	0.405	0.105	0.071	0.131	0.465
-30	0.050	0.038	0.052	0.192	0.100	0.087	0.101	0.247
-100	0.052	0.050	0.048	0.084	0.101	0.100	0.095	0.132

Table S.4: 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. Y_2 is detrended recursively (top panel) or over the full-sample (bottom panel). The true coefficient on the trend is $\gamma_x = 1$. The sample size is 200. Number of MC replications: 20000.

$m=2$								
	At 5%				At 10%			
	$\rho=0.20$		0.95		0.20		0.95	
	AR	t	AR	t	AR	t	AR	t
$c=0$	0.044	0.010	0.048	0.155	0.092	0.033	0.097	0.193
-1	0.044	0.010	0.046	0.135	0.092	0.033	0.093	0.169
-10	0.043	0.016	0.043	0.100	0.087	0.049	0.093	0.130
-30	0.035	0.032	0.043	0.078	0.077	0.079	0.095	0.107
-100	0.030	0.045	0.058	0.052	0.068	0.096	0.123	0.078

$m=4$								
	At 5%				At 10%			
	$\rho=0.20$		0.95		0.20		0.95	
	AR	t	AR	t	AR	t	AR	t
$c=0$	0.033	0.010	0.039	0.152	0.075	0.032	0.087	0.186
-1	0.034	0.010	0.037	0.130	0.074	0.031	0.084	0.161
-10	0.030	0.016	0.032	0.093	0.071	0.049	0.075	0.119
-30	0.022	0.030	0.032	0.073	0.057	0.076	0.080	0.097
-100	0.021	0.040	0.046	0.062	0.054	0.092	0.106	0.082

Table S.5: Null rejection frequencies of AR (with filtered instruments) and conventional t tests of the hypothesis H_0: $b_{12}=0$ in a bivariate SVAR(m) with long-run restrictions. ρ is the correlation between the reduced-form VAR errors. Y_2 is detrended recursively. The true coefficient on the trend is $\gamma_x=0$. The sample size is 200. Number of MC replications: 20000.

$m = 2$								
	At 5%				At 10%			
	$\rho = 0.20$		0.95		0.20		0.95	
	AR	t	AR	t	AR	t	AR	t
$c = 0$	0.044	0.010	0.048	0.155	0.092	0.033	0.097	0.193
-1	0.044	0.010	0.046	0.135	0.092	0.033	0.093	0.169
-10	0.043	0.016	0.043	0.100	0.087	0.049	0.093	0.130
-30	0.035	0.032	0.043	0.078	0.077	0.079	0.095	0.107
-100	0.030	0.045	0.058	0.052	0.068	0.096	0.123	0.078

$m = 4$								
	At 5%				At 10%			
	$\rho = 0.20$		0.95		0.20		0.95	
	AR	t	AR	t	AR	t	AR	t
$c = 0$	0.033	0.010	0.039	0.152	0.075	0.032	0.087	0.186
-1	0.034	0.010	0.037	0.130	0.074	0.031	0.084	0.161
-10	0.030	0.016	0.032	0.093	0.071	0.049	0.075	0.119
-30	0.022	0.030	0.032	0.073	0.057	0.076	0.080	0.097
-100	0.021	0.040	0.046	0.062	0.054	0.092	0.106	0.082

Table S.6: Null rejection frequencies of AR (with filtered instruments) and conventional t tests of the hypothesis H_0: $b_{12} = 0$ in a bivariate SVAR(m) with long-run restrictions. ρ is the correlation between the reduced-form VAR errors. Y_2 is detrended recursively. The true coefficient on the trend is $\gamma_x = 1$. The sample size is 200. Number of MC replications: 20000.

Figure S.1: Size of the projection ARW test of the hypothesis $H_0 : d_{21} = \bar{d}_{21}$, in a SVAR(1) model with T=2000 at three different significance levels. The number of Monte Carlo replications is 10000.

ARW test of $H_0 : d_{21} = \bar{d}_{21}$ as a function of $\bar{d}_{21}$ for three different levels of significance: 10%, 5% and 1%. The sample size is $T = 2000$ and the number of Monte Carlo replications is 10000. These can be thought of as estimates of the asymptotic size of the projection test at different levels of significance. They are very close to the corresponding results in Figure 1 in the paper for the case $T = 200$.

Figure S.2 reports the size of an ARW test that uses χ^2_1 instead of χ^2_2 critical values, corresponding exactly to the cases reported in Figure S.1. We see that the ARW test with degrees of freedom correction overrejects for many values under the null. So, confidence intervals on d_{21} obtained by inverting this test have asymptotic coverage below their nominal level.

Figure S.3 repeats the exercise in Figure S.2 except the parameter c in the DGP is constrained to be $c = -200$ (thus corresponding to a highest root of 0.9). We could

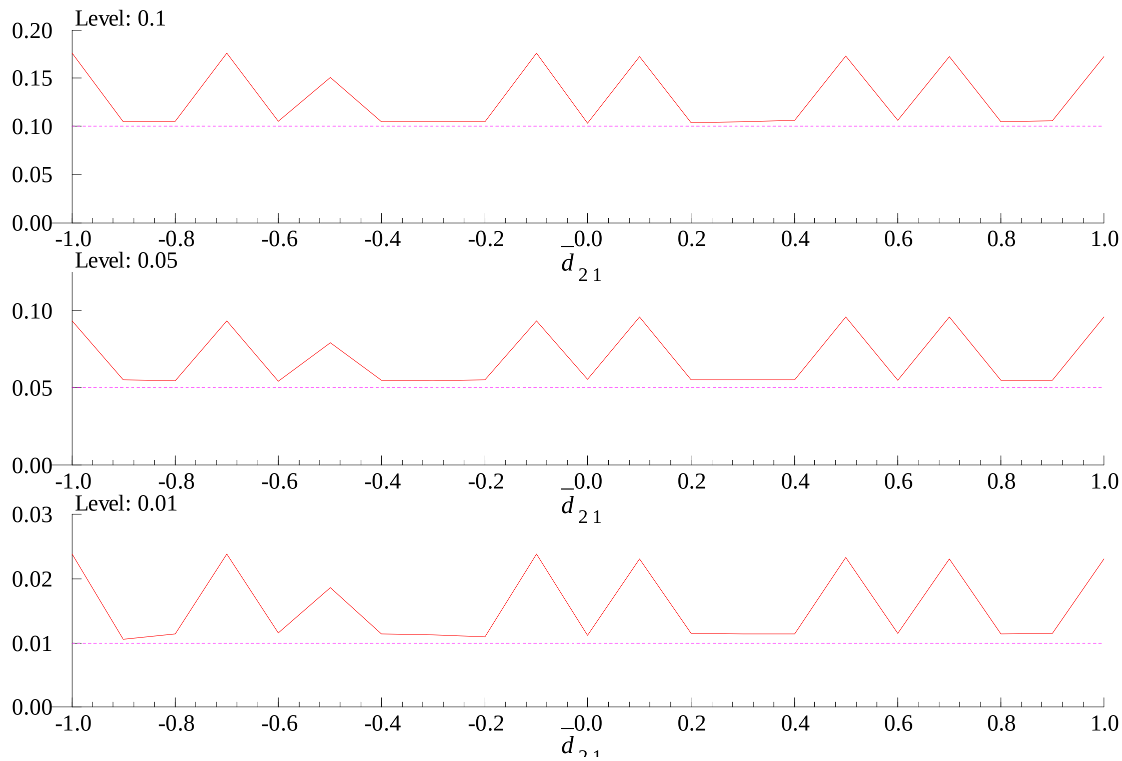

Figure S.2: Size of the projection ARW test of the hypothesis $H_0 : d_{21} = \bar{d}_{21}$, using χ^2_1 critical values, in a SVAR(1) model with $T = 2000$ at three different significance levels. The number of Monte Carlo replications is 10000.

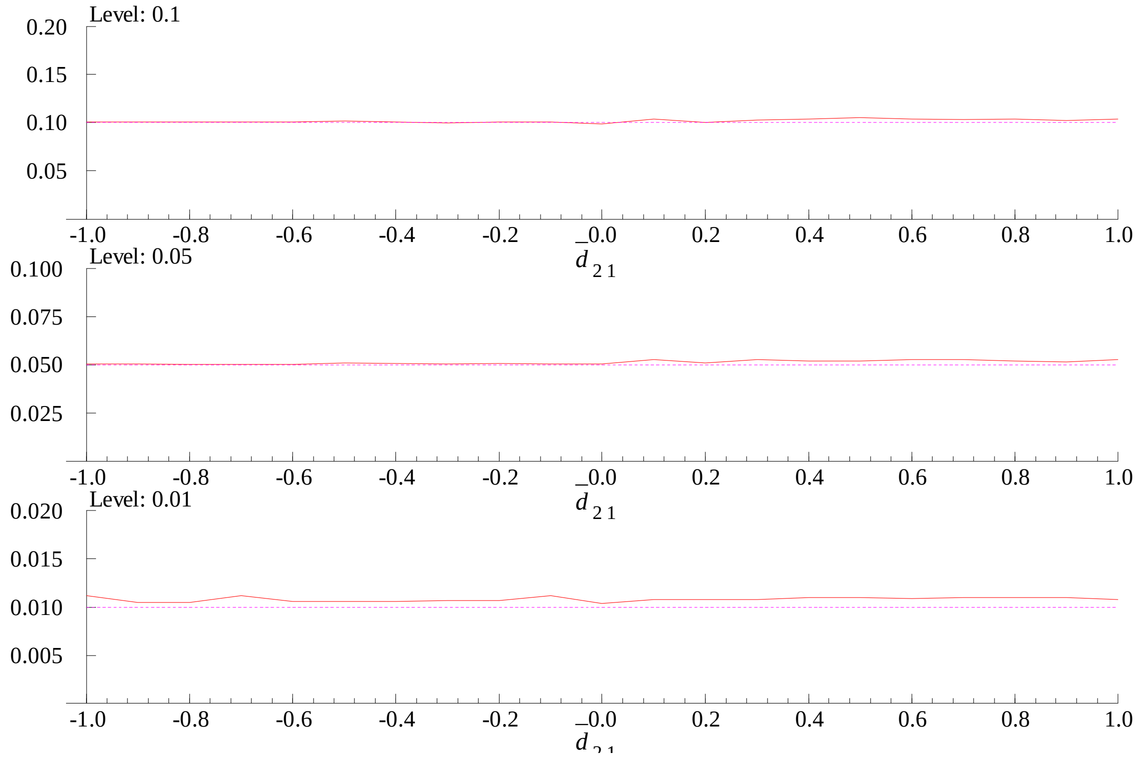

Figure S.3: Size of the projection ARW test of the hypothesis $H_0 : d_{21} = \bar{d}_{21}$, using χ^2_1 critical values, in a SVAR(1) model with $T = 2000$ at three different significance levels, when the highest root in the VAR is 0.9. The number of Monte Carlo replications is 10000.

view these results as giving the size of the ARW test with degrees of freedom correction when the data is stationary and identification is strong. As expected, the size of the test is equal to its nominal level for all values of d_{21}.

5.3 Large-sample power of the AR test

We report large-sample power curves to complement the results for $T = 200$ reported in Figure 2 in the paper. To that end, we set $T = 2000$ in the simulations. We compare the power of AR and t tests of $H_0 : b_{12} = 0$ against $H_1 : b_{12} \neq 0$ at the 10% level of significance. The remaining parameters are $\rho \in \{0.2, 0.95\}$, $\omega_1 = 1$, and $c = \{-10, -100, -500\}$. The chosen values of c correspond to approximate values of the concentration parameter $\lambda \in \{1.3, 13, 72\}$, respectively, i.e., weak, moderate and strong identification. The range of b_{12} under H_1 is $\lambda^{-1/2}(-3 : 3)$.

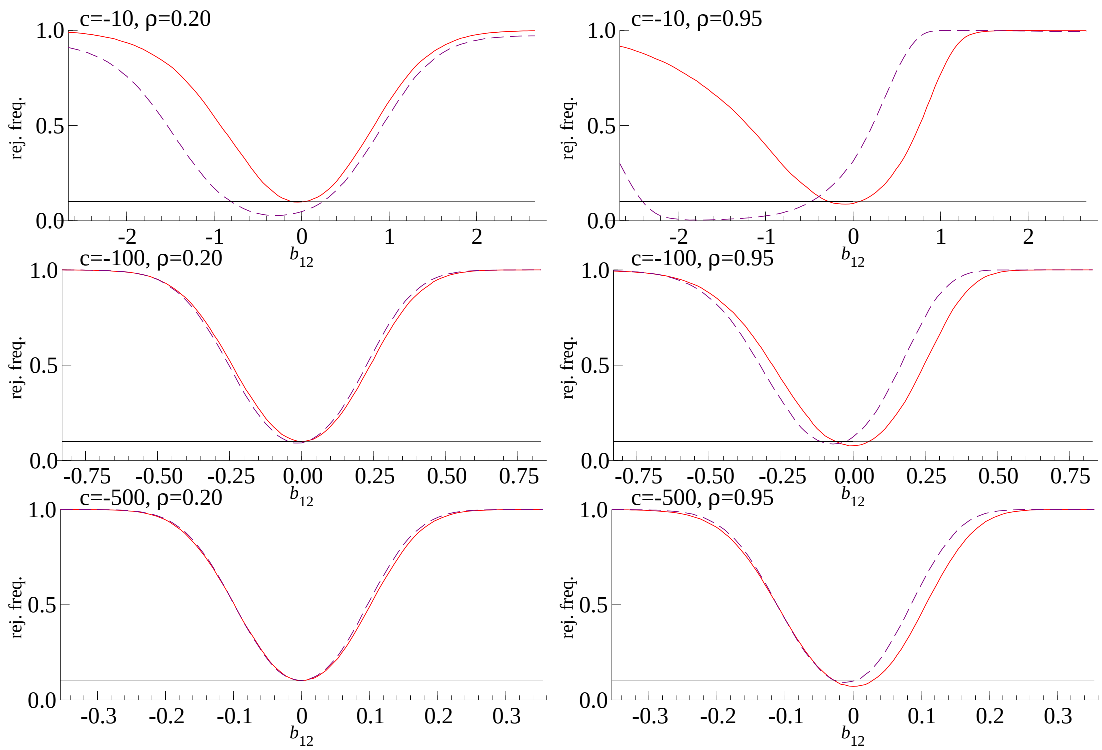

Figure S.4: Large-sample power of AR with filtered instrument (solid line) and t (dashed line) tests of the hypothesis $H_0 : b_{12} = 0$ against $H_1 : b_{12} \neq 0$ in the SVAR(1) model with long run restrictions. $T = 2000$, 10000 MC replications, ρ is correlation of reduced-form errors.

Figures S.4 reports the resulting power curves in each case. The figures shows that the AR test has good large sample power even for c close to zero. This is not the case for the t test, which is both size distorted and even biased in some cases. Moreover, when identification is strong ($c = -500$), the power of the AR test is very similar to that of the t test, which is asymptotically efficient in this case. Since the DGP in this case is approximately stationary, this is a consequence of the fact that the AR and t tests are asymptotically equivalent in the case of stationarity, see Remark 2 to Theorem 1.

5.4 Bonferroni method

We compare the power of projection and Bonferroni tests of $H_0 : d_{21} = 0$ against $H_1 : d_{21} \neq 0$ at significance level 10%. We consider three different combinations of significance level $\eta = 10\%$. We denote by η_1 the significance level for the first-step AR confidence set for b_{12} and by $\eta_2 = (\eta - \eta_1) / (1 - \eta_1)$ the level of the Wald test given b_{12} in the second step, see Remark 4 below Theorem 2. We set $b_{12} = 0$, $\omega_1 = 1$ and consider $\eta_1 = 1\%$, 5.13% and 9% (the second value is such that $\eta_1 = \eta_2$). We note that with these parameter values $\rho = d_{21}$.

We first report in Figure S.5 the power in a large sample of $T = 2000$ observations for $c = -100$ (moderately strong identification) and then, in Figure S.6, for $T = 200$, also with $c = -100$ (strong identification). Under moderately strong identification the power of the projection test is close to that of the Bonferroni tests that put sufficiently high weight on the first-step AR confidence set for b_{12}, i.e., $\eta_1 \geq \eta_2$. Under strong identification, the projection test is more powerful. This suggests that there is little to choose from between Bonferroni and projection in this case on the basis of power under weak identification and that projection is preferable in the identified case. Since projection turns out to be slightly faster to compute (it requires solving an unconstrained optimization problem, as opposed to a constrained optimization with an inequality constraint for the Bonferroni method), we are using the projection method in our empirical applications.

5.5 Concentration parameter

Identification strength is measured using an approximate formula for the concentration parameter λ. Table S.7 reports values of the concentration parameter for different values of c and a in the DGP. The numbers in bold are the cases for which the power curve is computed in the supplement.

5.6 Choice of parameters (c_z, b) when generating instruments

The instrument $z_t = \sum_{j=1}^{t-1} \rho_{Tz}^{t-j} \Delta Y_{2,j}$ depends on two parameters such that $\rho_{Tz} = 1 + \frac{c_z}{T^b}$ with $b \in (1/2, 1)$, and $c_z < 0$. We suggest in the paper $b = 0.95$ and $c_z = -1$. Through extensive simulation exercises, we show here that this choice induces good size/power trade-off for the AR test of $H_0 : b_{12} = 0$. Since the DGP depends on the triplet (c, b_{12}, ρ), our treatment on c and ρ in the simulation experiment induces various figures below. Throughout, we test H_0 at the 10% level over a sample of size $T = 200$ observations.

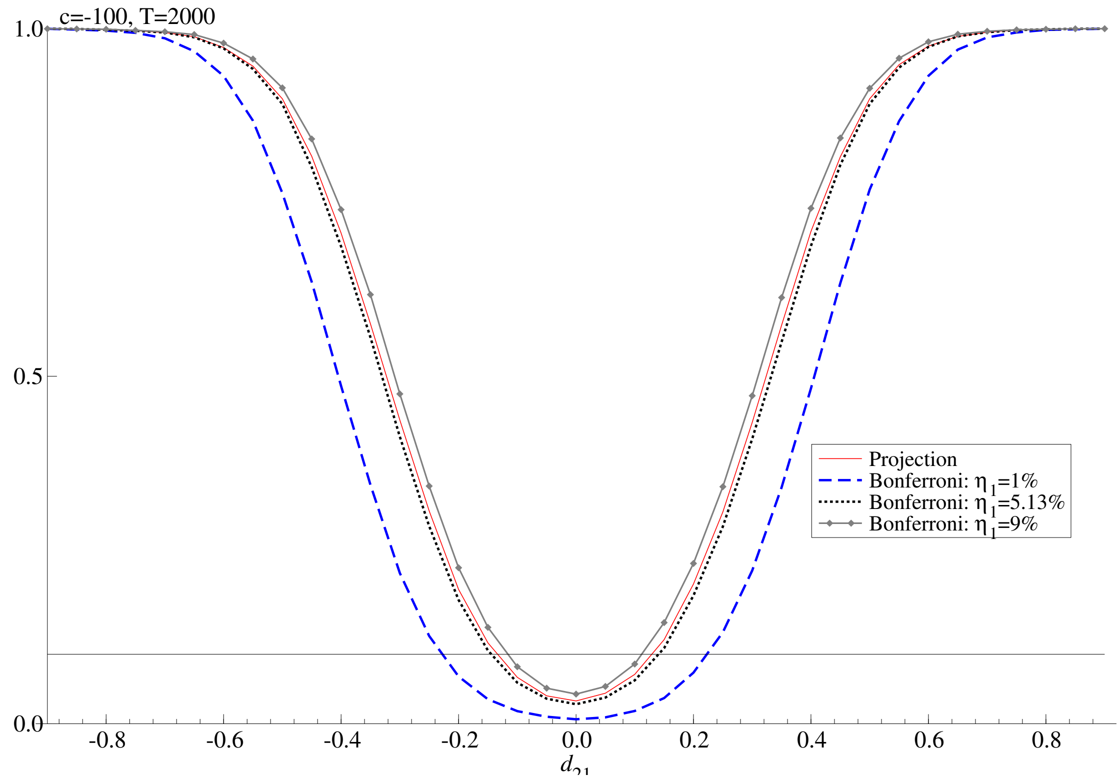

Figure S.5: Power of Projection ARW (red solid) and three different Bonferroni AR/W tests of the hypothesis $H_0 : d_{21} = 0$ against $H_1 : d_{21} \neq 0$ in a bivariate SVAR(1) at the 10% level of significance. η_1 denotes the level of the (first-step) AR test in the Bonferroni procedure and $\eta_2 = (10\% - \eta_1)/(1 - \eta_1)$ is the level of the second step Wald test. $T = 2000$ and the number of Monte Carlo replications is 10000.

	$a = 1$	0.95
$c = -1$	0.080590	0.13816
-10	**1.2660**	1.8737
-40	5.1661	7.5705
-50	6.4644	9.4826
-100	**13.026**	19.215
-150	19.724	29.276
-500	**71.593**	111.84

Table S.7: Values of the concentration parameter as a function of c and a in the DGP where $\Delta Y_{2,t} = \frac{c}{T^a} Y_{2,t-1} + u_{2t}$, and u_{2t} is white noise. The sample size is $T = 2000$.

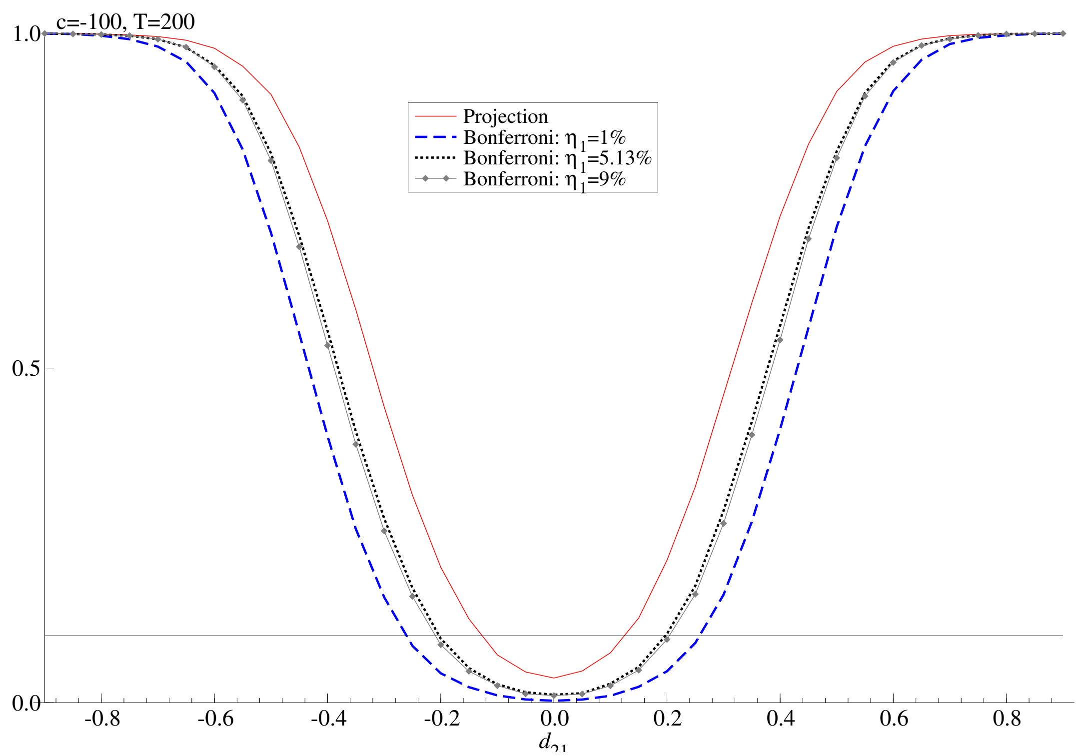

Figure S.6: Power of Projection ARW (red solid) and three different Bonferroni AR/W tests of the hypothesis $H_0 : d_{21} = 0$ against $H_1 : d_{21} \neq 0$ in a bivariate SVAR(1) at the 10% level of significance. η_1 denotes the level of the (first-step) AR test in the Bonferroni procedure and $\eta_2 = (10\% - \eta_1)/(1 - \eta_1)$ is the level of the second step Wald test. $T = 200$ and the number of Monte Carlo replications is 10000.

5.6.1 Fixed $(c, \rho) = (-10, 0.5)$

We first consider a fixed setting for $(c, \rho) = (-10, 0.5)$ and let $b_{12} \in \{0, 0.5, 1, 2\}$. We consider two situations, first fixing $c_z = -1$ and letting b vary, and then fixing $b = 0.95$ and letting c_z vary.

When $c_z = -1$ is kept fixed but b varies, as shown on the horizontal axis of Figure S.7, we see that changing b does not induce size distortions but that the power of the test generally increases with b. Yet for alternative hypotheses that are further away from the null (larger values of b_{12}) the power is non monotonic in b: it tends to decrease as b gets very close to 1. Hence, it is not optimal to set b too close to unity (say, $b = 0.99$) and setting $b = 0.95$ seems appropriate.

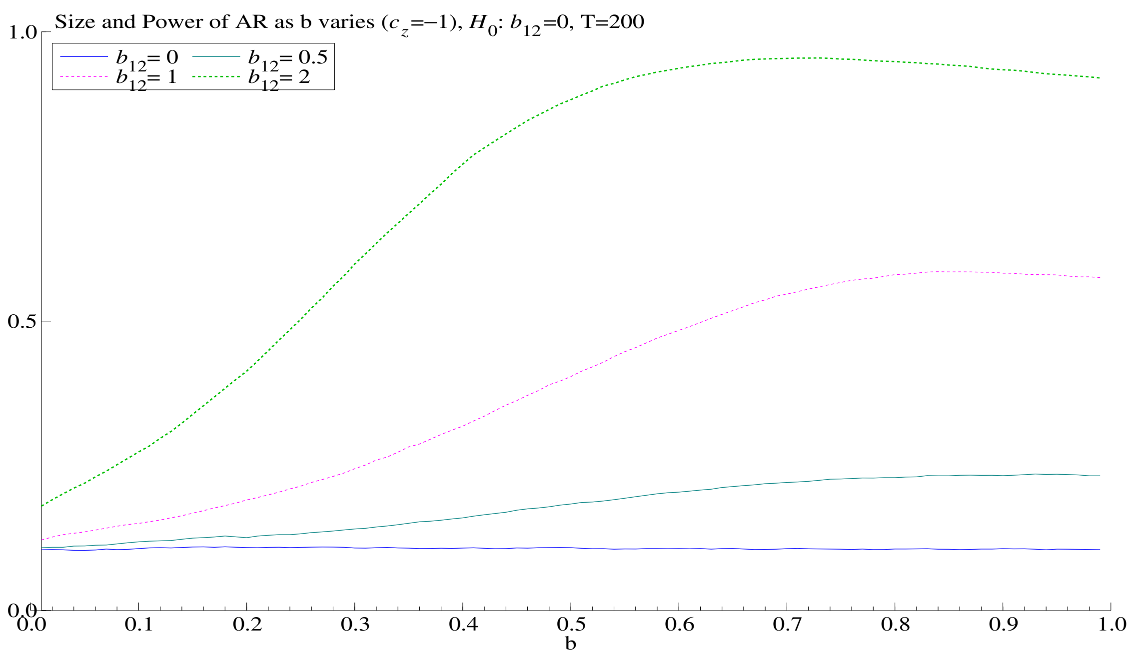

Figure S.7: Size and Power of the AR statistic as b varies while $c_z = -1$ is being held constant.

Now, consider fixing $b = 0.95$ but letting c_z vary, as reported on horizontal axis of Figure S.8. The size of the test does not appear to be affected by the choice of c_z but its power is non-monotonic. Figure S.8 shows that setting $c_z = -1$ yields reasonable power across the alternatives considered.

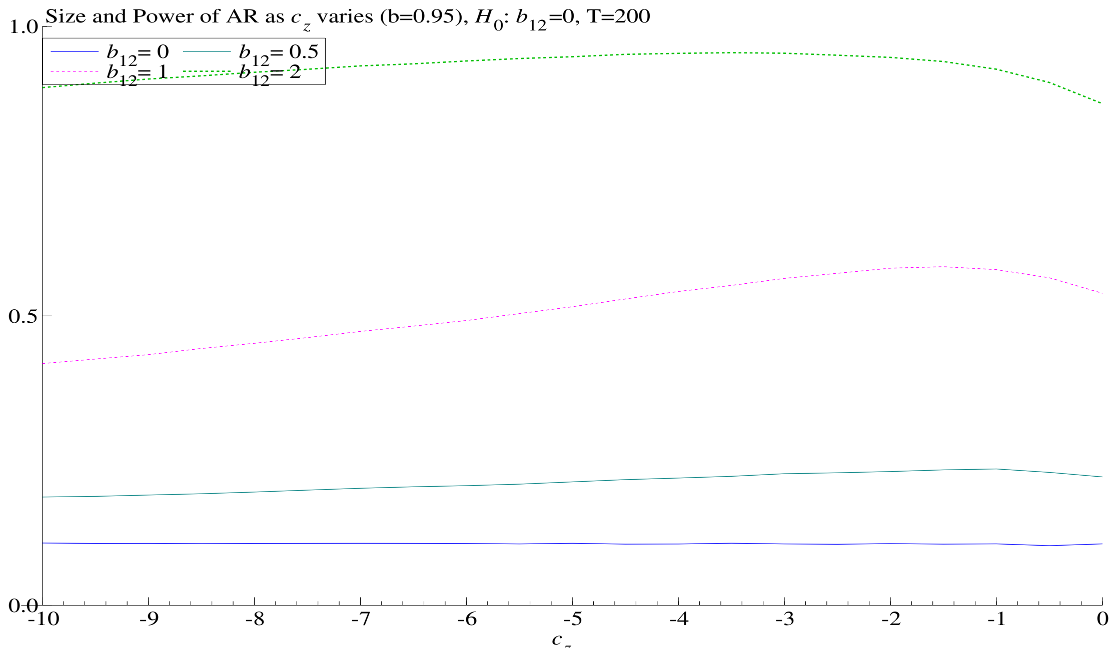

Figure S.8: Size and Power of the AR statistic as c_z varies while $b = 0.95$ is being held fixed.

5.6.2 Varying ρ

In Figure S.9, we also consider a range of values for $\rho \in (-1, 1)$, while keeping $c = -10$. To illustrate the size and power trade-off, we report the fixed alternative $b_{12} = 1$. This value is chosen so that the power with our proposed choice of (b, c_z) is close to 50%. Figure S.9 shows that $b = 0.95$ induces good power at the cost of minor size distortion over the whole range of values of ρ (here $c_z = -1$).

Figure S.10 complements the simulations above by considering $c = -1$ and -100. The fixed alternatives are $b_{12} = 5$ when $c = -1$, and $b_{12} = 0.2$ when $c = -100$. Again $b = 0.95$ appears a satisfactory choice when $c_z = -1$.

In a similar way, Figure S.11 reports the size and power trade-off when $b = 0.95$ is being held constant and while $c_z \in \{-10, -5, -1\}$. Setting $c_z = -1$ appears a reasonable choice overall.

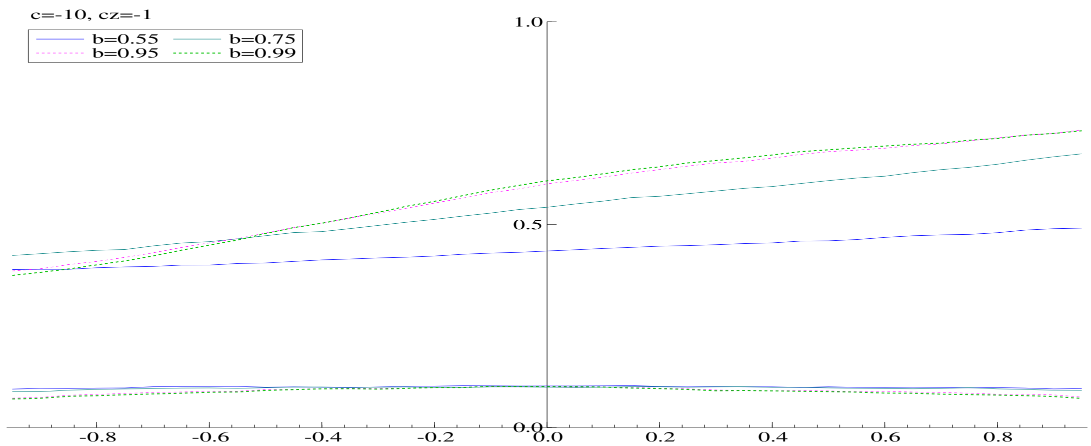

Figure S.9: Size and power of the AR statistic for $c_z = -1$ under varying b. The null is $H_0 : b_{12} = 0$ and the alternative is $b_{12} = 1$. The horizontal axis reports the value of ρ and the vertical axis the rejection probability.

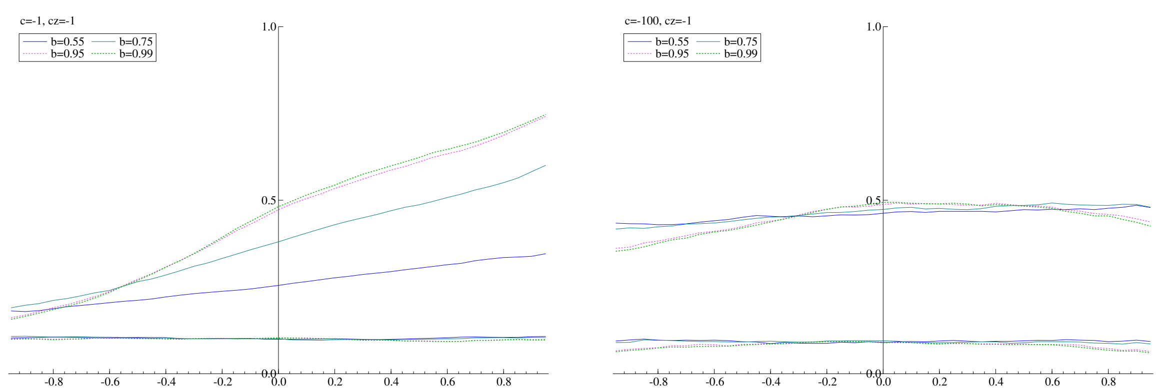

Figure S.10: Size and power of the AR statistic for $c \in \{-1, -100\}$ under varying b. The null is $H_0 : b_{12} = 0$ and the alternatives are $b_{12} = 5$ when $c = -1$ and $b_{12} = 0.2$ when $c = -100$. The horizontal axis reports the value of ρ and the vertical axis the rejection probability.

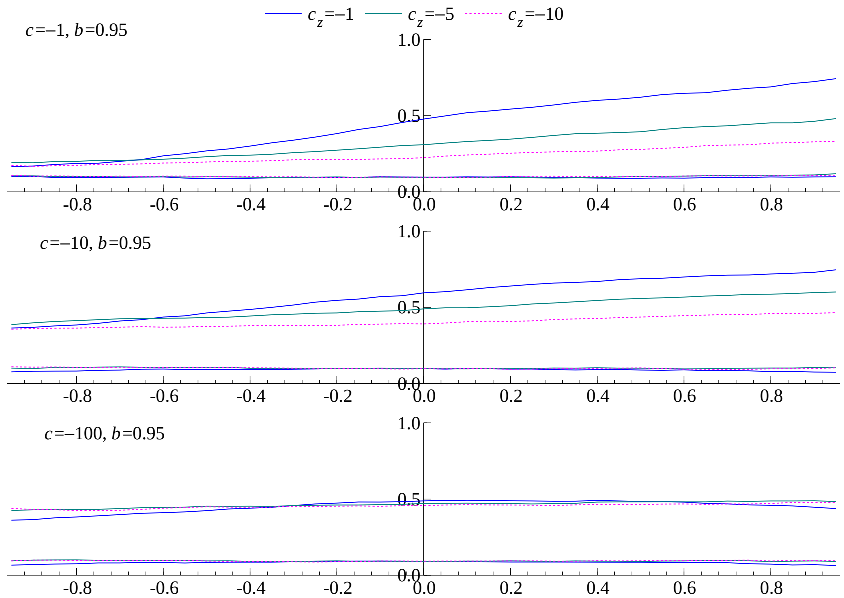

Figure S.11: Size and power of the AR statistic for $c_z \in \{-10, -5, -1\}$ when $b = 0.95$. The null is $H_0 : b_{12} = 0$ and the alternatives are $b_{12} = 5$ when $c = -1$, $b_{12} = 1$ when $c = -10$, and $b_{12} = 0.2$ when $c = -100$. The horizontal axis reports the value of ρ and the vertical axis the rejection probability.

5.6.3 Power over varying alternatives

We now extend the analysis further to consider, in Figure S.12, the power of the test statistic when $\rho = 0.2$ or 0.95 (as in the paper) for values of $c \in \{-100, -10, -1\}$. Figure S.12 presents the power plots of AR for the null $H_0 : b_{12} = 0$ under alternative values b_{12}. The figure report the cases where $b = 0.55, 0.75, 0.95$ or 0.99, while keeping $c_z = -1$. The figure shows that setting $b = 0.95$ and $c_z = -1$ appears to be a good choice overall.

5.7 Implementation of Gospodinov's (2010) method

In the Appendix of the paper, we report the power of a t test based on Gospodinov's (2010) estimator of b_{12} in a SVAR(2) model. Here we give the details of the implementation. Let $\widehat{\phi}$ denote the OLS estimator of ϕ in the regression $Y_{2t} = \mu + \phi Y_{2,t-1} + e_t$.

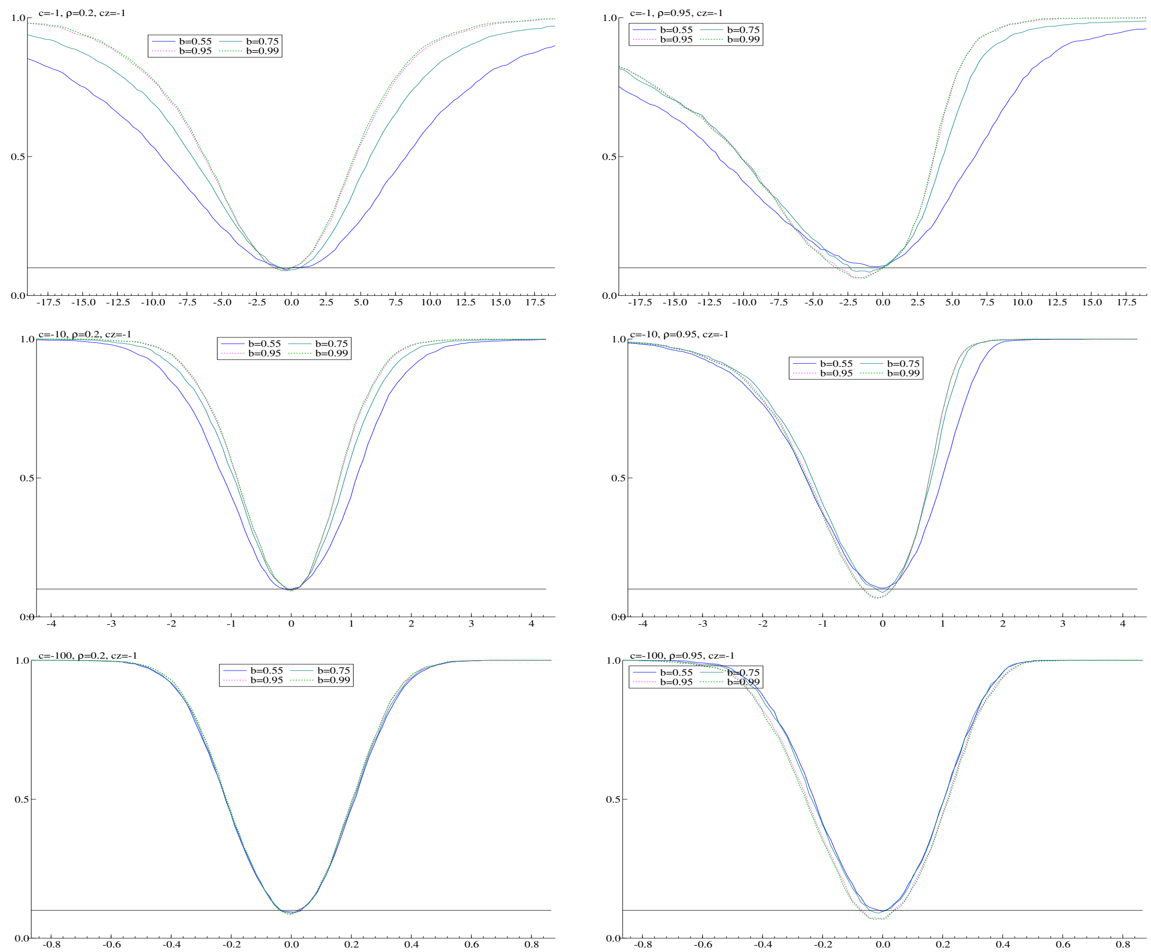

Figure S.12: Power plots of the AR statistic as a function of b_{12} (horizontal axis) for various choices of b (keeping $c_z = -1$).

The coefficients Ψ_1 are estimated by OLS in the system of equations

$$(I - \Psi_1 L) \begin{bmatrix} 1 - L & 0 \\ 0 & 1 - \widehat{\phi} L \end{bmatrix} Y_t = u_t.$$

Namely, denoting $\tilde{Y}_t := \left(I - \widehat{\Phi} L\right) Y_t$, $\widehat{\Psi}_1$ is simply obtained from a VAR(1) using OLS on $\tilde{Y}_t = \Psi_1 \tilde{Y}_{t-1} + u_t$, i.e.,

$$\widehat{\Psi}_1 = \sum_t \tilde{Y}_t \tilde{Y}'_{t-1} \left(\sum_t \tilde{Y}_{t-1} \tilde{Y}'_{t-1} \right)^{-1}.$$

Now,

$$\begin{aligned} vec\left(\widehat{\Psi}_1 - \Psi_1\right) &= \left(\left(\sum_t \tilde{Y}_{t-1} \tilde{Y}'_{t-1} \right)^{-1} \otimes I_2 \right) vec \sum_t u_t \tilde{Y}'_{t-1} \\ &= \left(\left(\sum_t \tilde{Y}_{t-1} \tilde{Y}'_{t-1} \right)^{-1} \otimes I_2 \right) \sum_t \left(\tilde{Y}_{t-1} \otimes u_t \right) \end{aligned}$$

so, under homoskedasticity,

$$\begin{aligned} \widehat{var}\left[\sqrt{T} vec\left(\widehat{\Psi}_1\right)\right] &= \left(\frac{1}{T} \sum_t \tilde{Y}_{t-1} \tilde{Y}'_{t-1} \right)^{-1} \otimes \widehat{\Sigma}_u \\ &\xrightarrow{p} \begin{bmatrix} * & * \\ * & \Xi \end{bmatrix}. \end{aligned}$$

So, denoting by $\psi = (\psi_{12}, \psi_{22})' = (-\Psi_{1,12}, 1 - \Psi_{1,22})$, we have

$$\sqrt{T}\left(\widehat{\psi} - \psi\right) \xrightarrow{d} N\left(0, \Xi\right),$$

since $var\left(\hat{\psi}\right) = var\left(\left(\widehat{\Psi}_{1,12}, \widehat{\Psi}_{1,22}\right)'\right)$, where the estimator of Ξ easily obtains from the previous formulae as the bottom right 2x2 block of

$$\left(\frac{1}{T} \sum_t \tilde{Y}_{t-1} \tilde{Y}'_{t-1} \right)^{-1} \otimes \widehat{\Sigma}_u,$$

with

$$\widehat{\Sigma}_u = \frac{1}{T}\sum_t \left(\tilde{Y}_t - \widehat{\Psi}_1 \tilde{Y}_{t-1}\right)\left(\tilde{Y}_t - \widehat{\Psi}_1 \tilde{Y}_{t-1}\right)'.$$

Gospodinov's (2010) estimator is $\widehat{b}_{12} := \widehat{\psi}_{12}/\widehat{\psi}_{22} = f\left(\widehat{\psi}\right)$ with

$$\frac{\partial f}{\partial \psi'} = \left[\begin{array}{cc} \frac{1}{\psi_{22}} & -\frac{\psi_{12}}{\psi_{22}^2} \end{array}\right] := F_\psi'.$$

Now, the Delta method yields

$$\sqrt{T}\left(\widehat{b}_{12}^{(0)} - b_{12}\right) \to^d N\left(0, F_\psi' \Xi F_\psi\right)$$

from which we obtain a t-stat for b_{12}

$$t_{\widehat{b}_{12}^{(0)}} = \sqrt{T}\frac{\widehat{b}_{12} - b_{12}}{\sqrt{F_{\widehat{\psi}}' \widehat{\Xi} F_{\widehat{\psi}}}}.$$

6 Supplementary material for empirical section

This section contains details of the computation algorithm of the confidence bands for the IRFs using our proposed ARW method, and additional empirical results based on different detrending methods and updated/extended data for the series used in the two applications reported in the main paper.

6.1 Data

6.1.1 Blanchard and Quah (1989)

The data presented in the main paper are taken from Blanchard and Quah (1989) (BQ), where the reader is referred to for detailed data description. Figure S.13 presents the original Blanchard and Quah (1989) data.

We also provide results based on an extended data set that goes up to 2014q4. The unemployment rate corresponds to men over the age of 20, and is seasonally adjusted (series ID: LNS14000025). Real GNP is seasonally adjusted, and the source is the Bureau of Economic Analysis (series ID: GNPC96). The data were obtained from the

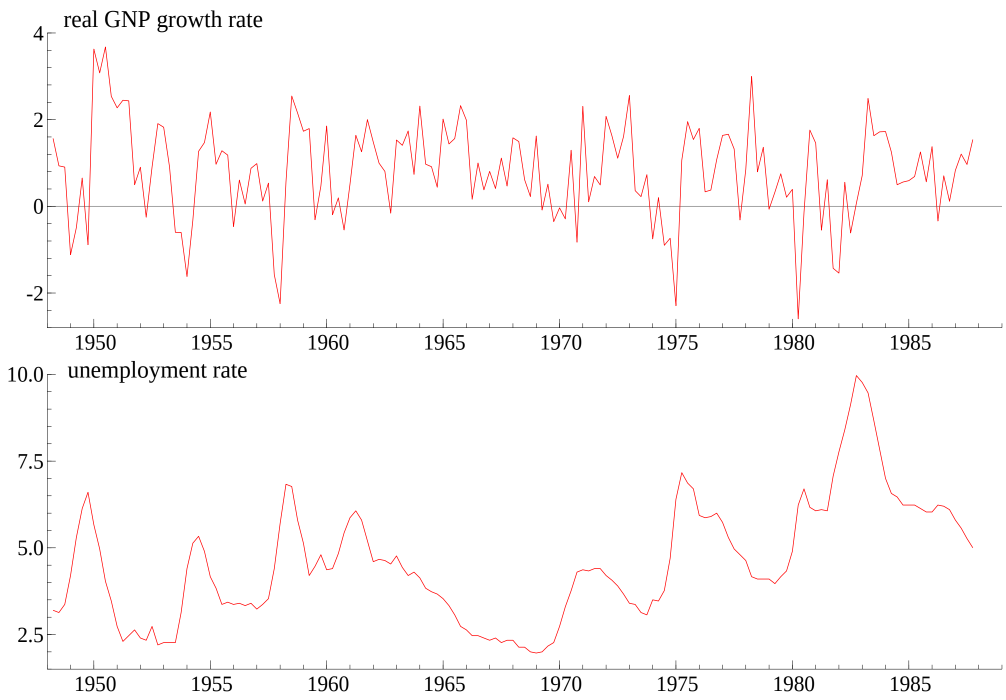

Figure S.13: Original data used in Blanchard and Quah (1989)

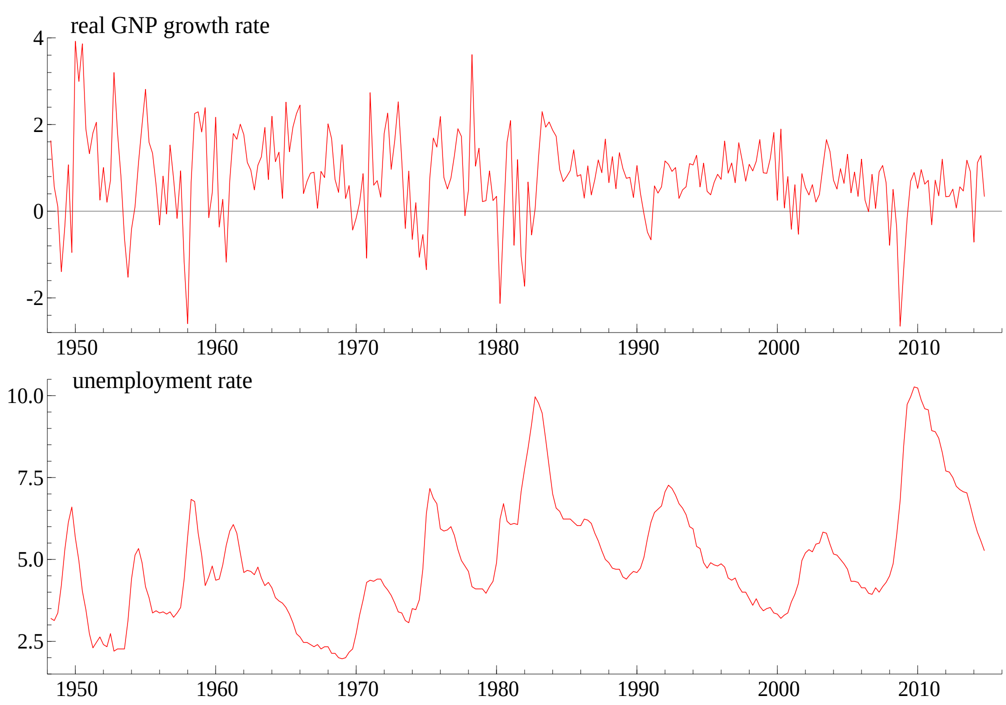

Figure S.14: Updated data for the series used in Blanchard and Quah (1989)

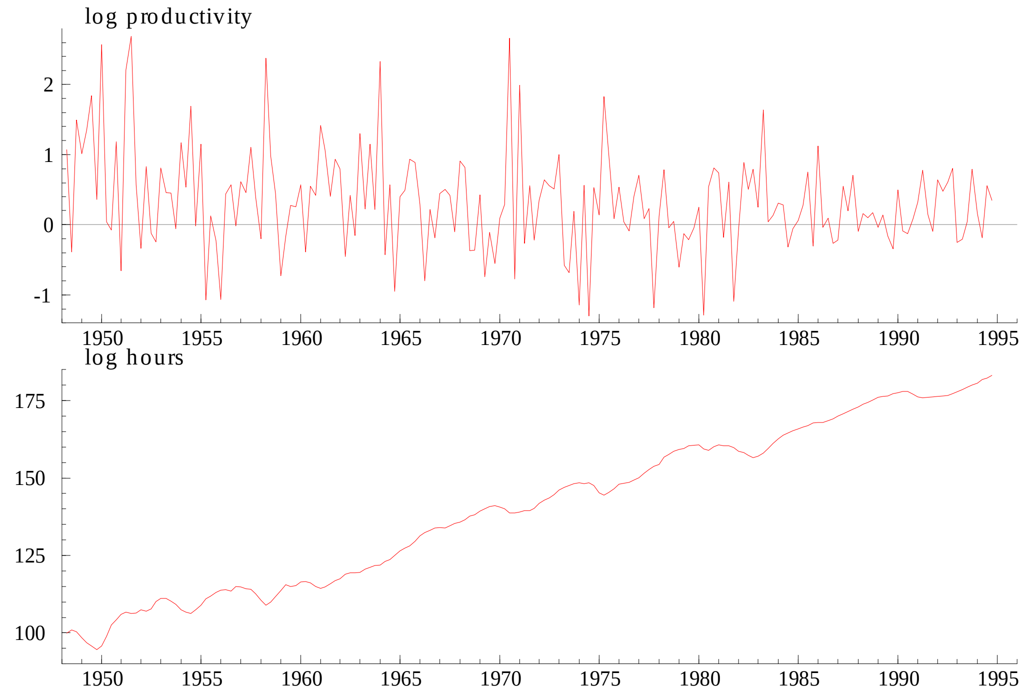

Figure S.15: Original data used in Galí (1999)

St. Louis Fed database FRED. The updated data are presented in Figure S.14.

6.1.2 Hours debate

The data presented in the main paper are taken from Galí (1999) and Christiano *et al*. (2003), where the reader is referred to for detailed data description. Figure S.15 presents the Galí (1999) data. The data used by Christiano *et al*. (2003) (CEV) is presented in Figure S.16.

We also provide results based on an updated and extended data set that spans the period 1948q1-2014q3, presented in Figure S.17. For the source and description of the data, we followed CEV footnote 9 and obtained the data taken from the DRI Economics database. The mnemonic for business labor productivity is LBOUT. The mnemonic for business hours worked is LBMN. The business hours worked data were converted to per capita terms using a measure of the civilian population over the age of 16 (mnemonic, P16).

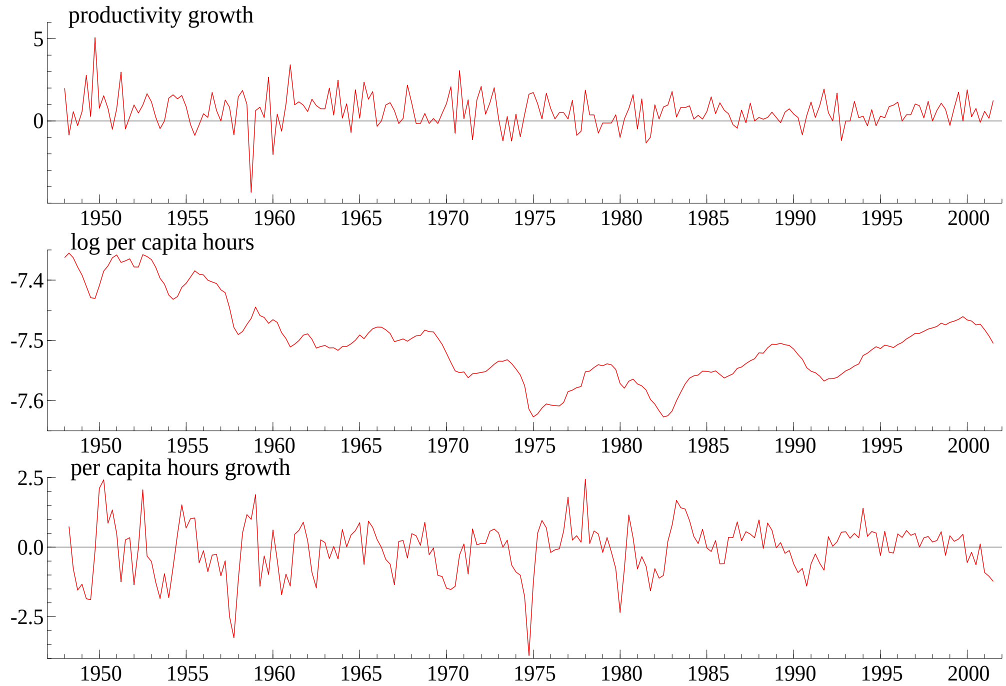

Figure S.16: Original data used in Christiano *et al*. (2003)

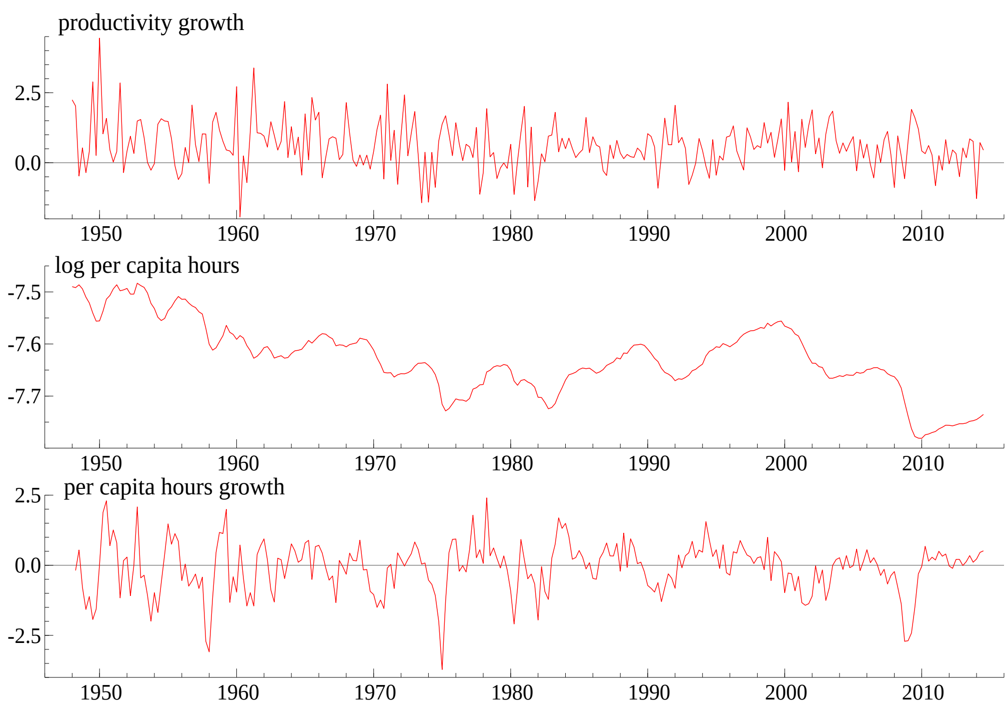

Figure S.17: Updated data for the series used in Christiano *et al*. (2003)

6.2 Computational details

The projection based confidence bands for the IRF are computed as follows. Let $g(b_{12}, \psi)$ denote a given impulse response of interest. $\hat{g}(b_{12}) = g\left(b_{12}, \hat{\psi}(b_{12})\right)$ its restricted estimate at b_{12}, and $\hat{\sigma}_{\hat{g}}(b_{12})$ the associated standard error computed using the delta method.

The joint η-level confidence set for (b_{12}, g) can be computed as follows. First, for any given value of b_{12}, the smallest value of the ARW statistic (12) is equal to $AR(b_{12})$, since at $\hat{g}(b_{12})$, $W(b_{12}) = 0$. Therefore, the confidence set for $g(b_{12}, \psi) = b_{12}$ can be computed simply by

$$\mathcal{C}_{b_{12}} = \left\{b_{12}^0 \in \Re : AR\left(b_{12}^0\right) < c_\eta\right\}, \tag{S–28}$$

where c_η is the $1-\eta$ quantile of the χ_2^2 distribution. With conditional homoskedasticity, this inversion can be done analytically using the formula given by Dufour and Taamouti (2005). For a general $g(b_{12}, \psi)$ evaluated at any given point $b_{12} = b_{12}^0$, the Wald confidence interval is given by

$$\hat{g}\left(b_{12}^0\right) \pm \hat{\sigma}_{\hat{g}}\left(b_{12}^0\right)\sqrt{c_\eta - AR\left(b_{12}^0\right)}. \tag{S–29}$$

The upper and lower bounds of the projection-based confidence set for g are given by

$$\left[\min_{b_{12}^0 \in \mathcal{C}_{b_{12}}} \underline{g}\left(b_{12}^0\right), \max_{b_{12}^0 \in \mathcal{C}_{b_{12}}} \bar{g}\left(b_{12}^0\right)\right]. \tag{S–30}$$

The procedure is repeated for each impulse response, using the same $\mathcal{C}_{b_{12}}$, which is common to all. Since g is smooth, we can use derivative-based optimization methods to locate the extrema, which is what we do in our applications. It is advisable to use more than one set of starting values to avoid getting stuck at local extrema. It is also possible to find the extrema by grid search, but it is important to use a fine grid of points in $\mathcal{C}_{b_{12}}$, because the extrema of $\underline{g}\left(b_{12}^0\right)$ and $\bar{g}\left(b_{12}^0\right)$ may occur at interior points of $\mathcal{C}_{b_{12}}$, and the functions $\underline{g}(\cdot)$ and $\bar{g}(\cdot)$ could be very steep.

An alternative to the projection method is the Bonferroni method. This involves combining an η_1-level AR test with an η_2-level Wald test for g. Thus, $\mathcal{C}_{b_{12}}$ is obtained by replacing c_η in (S–28) with the $1 - \eta_1$ quantile of the χ_1^2 distribution (note the difference also in degrees of freedom), and the term $\sqrt{c_\eta - AR\left(b_{12}^0\right)}$ in (S–29) with the $1 - \eta_2/2$ quantile of the standard normal distribution. The resulting interval in (S–30)

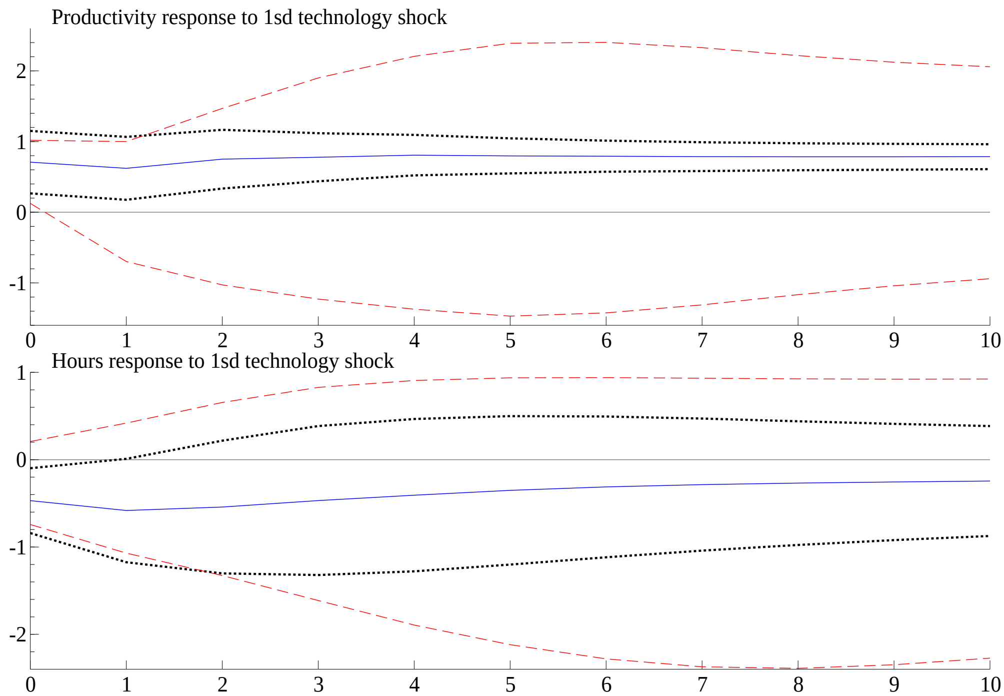

Figure S.18: Estimates and confidence bands of the IRFs in CEV with recursive detrending using their original data.The solid line is the ML estimator. The dotted lines are 90% Wald confidence intervals, and the dashed lines are the 90% projection ARW confidence intervals.

thus obtained would have coverage at least $1 - \eta_1 - \eta_2$.

6.3 Robustness checks in the hours application

6.3.1 Recursive detrending of hours

The results in Figure 10 in the paper are based on the CEV levels specification with non-detrended per capita hours. Those results are not robust to a trend in hours. Using recursive detrending, we obtain results that are robust to a linear trend in hours in Figure S.18. The results are entirely analogous to those without detrending, i.e., the remain inconclusive regarding the sign of the effect of technology shocks on hours.

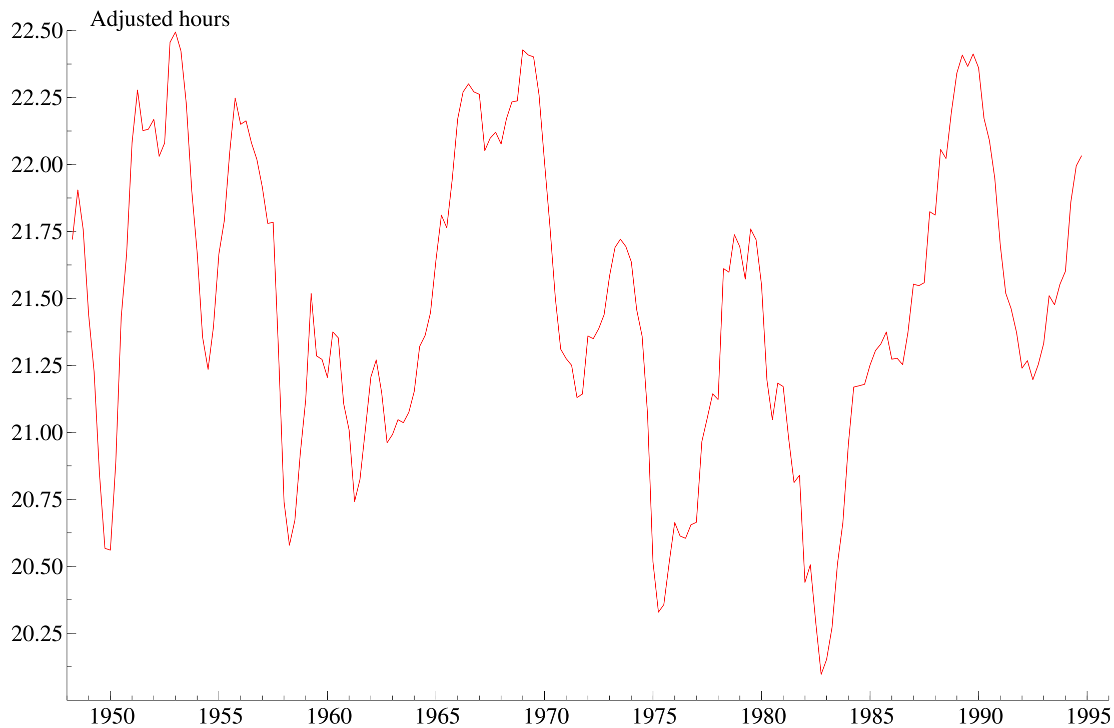

Figure S.19: Adjusted hours in Francis and Ramey (2009a)

6.3.2 Alternative detrending of hours

Francis and Ramey (2009a) provide an alternative measure of hours per capita, which removes low-frequency movements. See Figure S.19. We use their data of hours to replace those used in Galí (1999), and keep the other settings of Galí (1999) to facilitate comparison, i.e., we restrict the sample period to 1948:1-1994:4, estimate a SVAR(5) model by the long-run restriction with hours in level, and use the same data of productivity as in Galí (1999).

The resulting IRFs together with the robust confidence bands based on our proposed ARW method and the non-robust confidence bands are reported in Figure S.20. Though the IRF of technology shocks on hours is estimated to be negative, the uncertainty is sufficiently large that the evidence regarding the sign of the effect remains inconclusive.

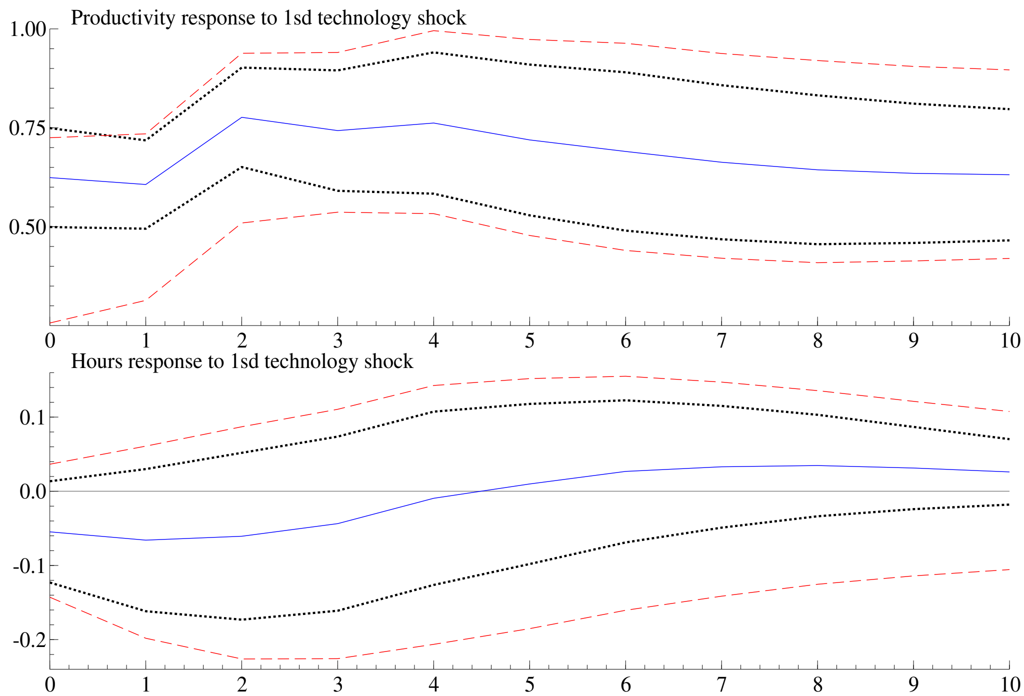

Figure S.20: Estimates and confidence bands of the IRFs from a SVAR with hours in levels, using adjusted hours in Francis and Ramey (2009a). The solid line is the ML estimator. The dotted lines are 90% Wald confidence intervals, and the dashed lines are the 90% projection ARW confidence intervals.

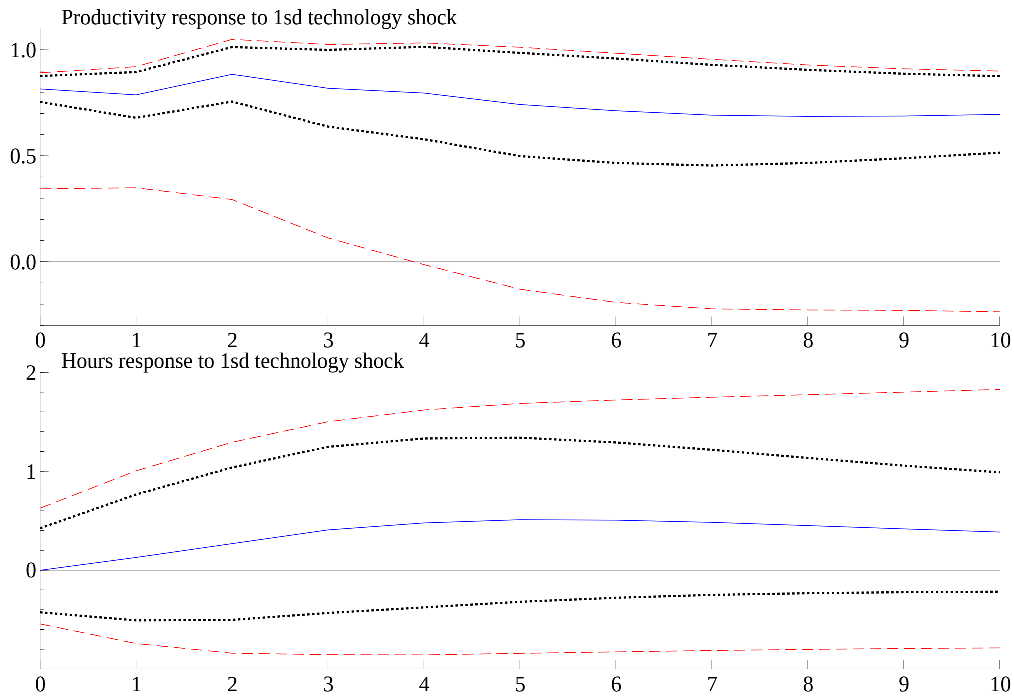

Figure S.21: Estimates and confidence bands of the IRFs with extended CEV data and recursive detrending. The solid line is the ML estimator. The dotted lines are 90% Wald confidence intervals, and the dashed lines are the 90% projection ARW confidence intervals.

6.3.3 IRFs with extended sample

With the extended sample and recursive detrending, the resulting IRFs from the levels specification of CEV are presented in Figure S.21. The evidence on the sign of the effect of technology shocks on hours remains inconclusive.

Difference specification with extended data Figure S.22 presents the results for the difference specification in Galí (1999) with per capita hours instead of total hours and over the updated sample. The results are essentially the same as with his original data (which used total instead of per capita hours).

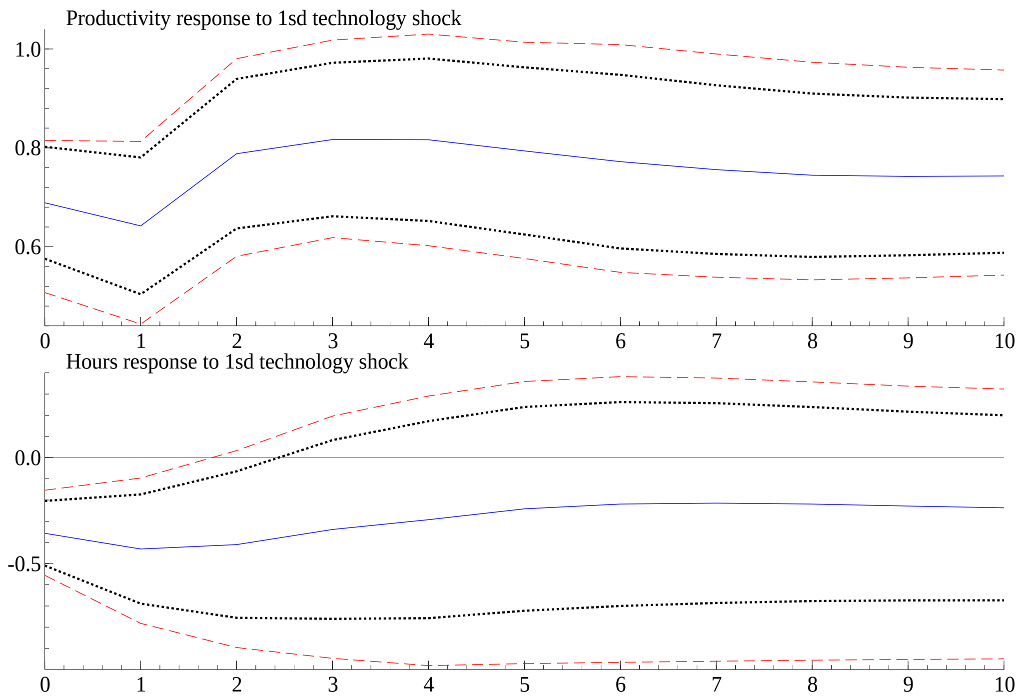

Figure S.22: Estimates and confidence bands of the IRFs with extended Galí (1999) data. The solid line is the ML estimator. The dotted lines are 90% Wald confidence intervals, and the dashed lines are the 90% projection ARW confidence intervals.

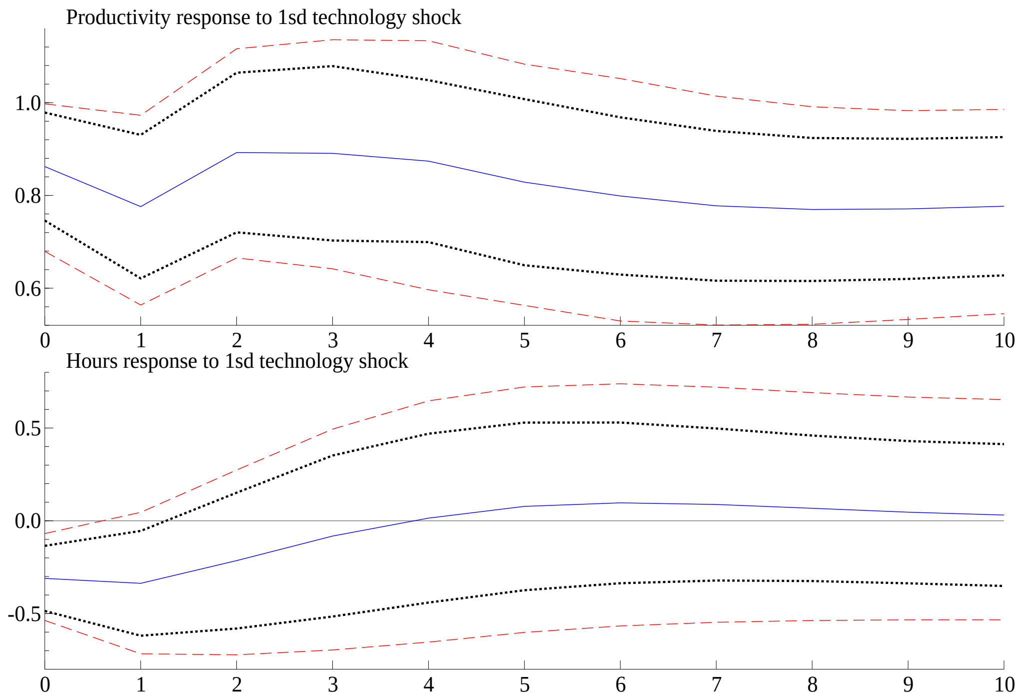

Figure S.23: Estimates and confidence bands of the IRFs with CEV data and the difference specification. The solid line is the ML estimator. The dotted lines are 90% Wald confidence intervals, and the dashed lines are the 90% projection ARW confidence intervals.

6.3.4 Difference specification with original CEV data

Finally, we use the original CEV data but consider the difference specification instead of the level specification of hours in CEV. The resulting IRFs are presented in Figure S.23.

Both Figure S.22 and Figure S.23 show that identification is not weak when hours appears in first differences, and the short run effect of a technology shock on hours is significantly negative.

7 Articles that use SVARs in Top Journals, 2005-2014

Table S.8 lists the articles that used SVARs and were published in the following eight journals during the period 2005-2014: *American Economic Review*, *Econometrica*, *Journal of Political Economy*, *Quarterly Journal of Economics*, *Review of Economic Studies*, *American Economic Journal - Macroeconomics*, *Journal of Monetary Economics and Journal of Money, Credit and Banking*.

	With long-run restrictions	Without long-run restrictions
1	(Alvarez and Jermann 2005)	(Iwata and Wu 2005)
2	(Bernanke, Boivin, Doan, and Eliasz 2005)	(Kim 2005)
3	(Francis and Ramey 2005)	(Primiceri 2005)
4	(Orphanides and Van Norden 2005)	(Uhlig 2005)
5	(Beaudry and Portier 2006)	(Ashcraft 2006)
6	(Chang and Hong 2006)	(Basu, Fernald, and Kimball 2006)
7	(Cover, Enders, and Hueng 2006)	(Braun and Shioji 2006)
8	(Croushore and Evans 2006)	(Farrant and Peersman 2006)
9	(Fisher 2006)	(Mitra 2006)
10	(Lastrapes 2006)	(Sims and Zha 2006)
11	(Reis 2006)	(Dedola and Neri 2007)
12	(Aguiar-Conraria and Wen 2007)	(Fernández-Villaverde, Rubio-Ramírez, Sargent, and Watson 2007)
13	(Avouyi-Dovi and Matheron 2007)	(Maćkowiak 2007)
14	(Evans and Marshall 2007)	(McCarthy and Zakrajšek 2007)
15	(Fernald 2007)	(Miniane and Rogers 2007)
16	(King and Morley 2007)	(Olivei and Tenreyro 2007)
17	(Liu and Phaneuf 2007)	(Roush 2007)
18	(Marchetti and Nucci 2007)	(Benati 2008)
19	(Michelacci and Lopez-Salido 2007)	(Bilbiie, Meier, and Müller 2008)
20	(Morley 2007)	(Gambetti, Pappa, and Canova 2008)
21	(Ravenna 2007)	(Lanne and Lütkepohl 2008)
22	(Chari, Kehoe, and McGrattan 2008)	(Mertens 2008)
23	(Corsetti, Dedola, and Leduc 2008)	(Altavilla and Ciccarelli 2009)
24	(Hansen, Heaton, and Li 2008)	(Benati and Surico 2009)
25	(Bjørnland and Leitemo 2009)	(Boivin, Giannoni, and Mihov 2009)
26	(Dupor, Han, and Tsai 2009)	(Carlstrom, Fuerst, and Paustian 2009)
27	(Fève and Guay 2009)	(Del Negro and Schorfheide 2009)
28	(Francis and Ramey 2009b)	(Evans and Marshall 2009)
29	(Gambetti and Gali 2009)	(Kilian 2009)
30	(Lorenzoni 2009)	(Danthine and Kurmann 2010)
31	(Fève, Matheron, and Sahuc 2010)	(Elder and Serletis 2010)
32	(Forni and Gambetti 2010)	(Kuester 2010)
33	(Rubio-Ramirez, Waggoner, and Zha 2010)	(Monacelli, Perotti, and Trigari 2010)
34	(Beaudry, Collard, and Portier 2011)	(Barsky and Sims 2011)
35	(Paciello 2011)	(Born and Müller 2012)
36	(Bachmann and Sims 2012)	(Ravn, Schmitt-Grohé, and Uribe 2012)
37	(Collard and Dellas 2012)	(Barakchian and Crowe 2013)
38	(Corsetti and Konstantinou 2012)	(Baumeister and Peersman 2013)
39	(Bekaert, Hoerova, and Duca 2013)	(Cloyne 2013)
40	(Blanchard, L'Huillier, and Lorenzoni 2013)	(Jang 2013)
41	(Keating 2013)	(Kurmann and Otrok 2013)
42	(Forni and Gambetti 2014)	(Leeper, Walker, and Yang 2013)
43	(Kano and Nason 2014)	(Mertens and Ravn 2013)
44	(Kurmann and Mertens 2014)	(Mumtaz and Zanetti 2013)
45		(Mertens and Ravn 2014)
46		(Monnet 2014)
47		(Nickel and Tudyka 2014)
48		(Walentin 2014)

Table S.8: The table lists SVAR articles in the top 8 macro journals over the period 2005-2014.

References

Aguiar-Conraria, L. and Y. Wen (2007). Understanding the large negative impact of oil shocks. *Journal of Money, Credit and Banking 39*(4), 925–944.

Altavilla, C. and M. Ciccarelli (2009). The effects of monetary policy on unemployment dynamics under model uncertainty: Evidence from the united states and the euro area. *Journal of Money, Credit and Banking 41*(7), 1265–1300.

Alvarez, F. and U. J. Jermann (2005). Using asset prices to measure the persistence of the marginal utility of wealth. *Econometrica 73*(6), 1977–2016.

Ashcraft, A. B. (2006). New evidence on the lending channel. *Journal of Money, Credit and Banking*, 751–775.

Avouyi-Dovi, S. and J. Matheron (2007). Technology shocks and monetary policy: revisiting the fed's performance. *Journal of Money, Credit and Banking 39*(2-3), 471–507.

Bachmann, R. and E. R. Sims (2012). Confidence and the transmission of government spending shocks. *Journal of Monetary Economics 59*(3), 235–249.

Barakchian, S. M. and C. Crowe (2013). Monetary policy matters: Evidence from new shocks data. *Journal of Monetary Economics 60*(8), 950–966.

Barsky, R. B. and E. R. Sims (2011). News shocks and business cycles. *Journal of Monetary Economics 58*(3), 273–289.

Basu, S., J. G. Fernald, and M. S. Kimball (2006). Are technology improvements contractionary? *American Economic Review*, 1418–1448.

Baumeister, C. and G. Peersman (2013). Time-varying effects of oil supply shocks on the us economy. *American Economic Journal: Macroeconomics 5*(4), 1–28.

Beaudry, P., F. Collard, and F. Portier (2011). Gold rush fever in business cycles. *Journal of Monetary Economics 58*(2), 84–97.

Beaudry, P. and F. Portier (2006). Stock prices, news, and economic fluctuations. *American Economic Review 96*(4), 1293–1307.

Bekaert, G., M. Hoerova, and M. L. Duca (2013). Risk, uncertainty and monetary policy. *Journal of Monetary Economics 60*(7), 771–788.

Benati, L. (2008). The great moderation in the united kingdom. *Journal of Money, Credit and Banking 40*(1), 121–147.

Benati, L. and P. Surico (2009). Var analysis and the great moderation. *American Economic Review 99*(4), 1636–1652.

Bernanke, B. S., J. Boivin, T. Doan, and P. S. Eliasz (2005). Measuring the effects of monetary policy: A factor-augmented vector autoregressive (favar) approach. *Quarterly Journal of Economics* (120), 387–422.

Bilbiie, F. O., A. Meier, and G. J. Müller (2008). What accounts for the changes in us fiscal policy transmission? *Journal of Money, Credit and Banking 40*(7), 1439–1470.

Bjørnland, H. C. and K. Leitemo (2009). Identifying the interdependence between us monetary policy and the stock market. *Journal of Monetary Economics 56*(2), 275–282.

Blanchard, O. J., J.-P. L'Huillier, and G. Lorenzoni (2013). News, noise, and fluctuations: An empirical exploration. *American Economic Review 103*(7), 3045–70.

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

Boivin, J., M. P. Giannoni, and I. Mihov (2009). Sticky prices and monetary policy: Evidence from disaggregated us data. *American Economic Review 99*(1), 350–84.

Born, B. and G. J. Müller (2012). Government spending shocks in quarterly and annual time series. *Journal of Money, Credit and Banking 44*(2-3), 507–517.

Braun, R. A. and E. Shioji (2006). Monetary policy and the term structure of interest rates in japan. *Journal of Money, Credit and Banking*, 141–162.

Carlstrom, C. T., T. S. Fuerst, and M. Paustian (2009). Monetary policy shocks, choleski identification, and dnk models. *Journal of Monetary Economics 56*(7), 1014–1021.

Chang, Y. and J. H. Hong (2006). Do technological improvements in the manufacturing sector raise or lower employment? *American Economic Review*, 352–368.

Chari, V. V., P. J. Kehoe, and E. R. McGrattan (2008). Are structural vars with long-run restrictions useful in developing business cycle theory? *Journal of Monetary Economics 55*(8), 1337–1352.

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.).

Cloyne, J. (2013). Discretionary tax changes and the macroeconomy: new narrative evidence from the united kingdom. *American Economic Review 103*(4), 1507–1528.

Collard, F. and H. Dellas (2012). Flexible prices and the business cycle. *Journal of Money, Credit and Banking 44*(1), 221–233.

Corsetti, G., L. Dedola, and S. Leduc (2008). International risk sharing and the transmission of productivity shocks. *Review of Economic Studies 75*(2), 443–473.

Corsetti, G. and P. T. Konstantinou (2012). What drives us foreign borrowing? evidence on the external adjustment to transitory and permanent shocks. *American Economic Review 102*(2), 1062–92.

Cover, J. P., W. Enders, and C. J. Hueng (2006). Using the aggregate demand-aggregate supply model to identify structural demand-side and supply-side shocks: Results using a bivariate var. *Journal of Money, Credit, and Banking 38*(3), 777–790.

Croushore, D. and C. L. Evans (2006). Data revisions and the identification of monetary policy shocks. *Journal of Monetary Economics 53*(6), 1135–1160.

Danthine, J.-P. and A. Kurmann (2010). The business cycle implications of reciprocity in labor relations. *Journal of Monetary Economics 57*(7), 837–850.

Dedola, L. and S. Neri (2007). What does a technology shock do? a var analysis with model-based sign restrictions. *Journal of Monetary Economics 54*(2), 512–549.

Del Negro, M. and F. Schorfheide (2009). Monetary policy analysis with potentially misspecified models. *American Economic Review 99*(4), 1415–1450.

Dufour, J.-M. and M. Taamouti (2005). Projection-based statistical inference in linear structural models with possibly weak instruments. *Econometrica 73*(4), 1351–1365.

Dupor, B., J. Han, and Y.-C. Tsai (2009). What do technology shocks tell us about the new keynesian paradigm? *Journal of Monetary Economics 56*(4), 560–569.

Elder, J. and A. Serletis (2010). Oil price uncertainty. *Journal of Money, Credit and Banking 42*(6), 1137–1159.

Evans, C. L. and D. A. Marshall (2007). Economic determinants of the nominal treasury yield curve. *Journal of Monetary Economics 54*(7), 1986–2003.

Evans, C. L. and D. A. Marshall (2009). Fundamental economic shocks and the macroeconomy. *Journal of Money, Credit and Banking 41*(8), 1515–1555.

Farrant, K. and G. Peersman (2006). Is the exchange rate a shock absorber or a source of shocks? new empirical evidence. *Journal of Money, Credit and Banking*, 939–961.

Fernald, J. G. (2007). Trend breaks, long-run restrictions, and contractionary technology improvements. *Journal of Monetary Economics 54*(8), 2467–2485.

Fernández-Villaverde, J., J. F. Rubio-Ramírez, T. J. Sargent, and M. W. Watson (2007). Abcs (and ds) of understanding vars. *American Economic Review*, 1021–1026.

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

Fève, P., J. Matheron, and J.-G. Sahuc (2010). Disinflation shocks in the eurozone: A dsge perspective. *Journal of Money, Credit and Banking 42*(2-3), 289–323.

Fisher, J. D. (2006). The dynamic effects of neutral and investment-specific technology shocks. *Journal of Political Economy 114*(3), 413–451.

Forni, M. and L. Gambetti (2010). The dynamic effects of monetary policy: A structural factor model approach. *Journal of Monetary Economics 57*(2), 203–216.

Forni, M. and L. Gambetti (2014). Sufficient information in structural vars. *Journal of Monetary Economics 66*, 124–136.

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

Francis, N. and V. A. Ramey (2009a). Measures of per capita hours and their implications for the technology-hours debate. *Journal of Money, Credit and Bank-*

ing 41(6), 1071–1097.

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

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

Gambetti, L. and J. Gali (2009). On the sources of the great moderation. *American Economic Journal: Macroeconomics 1*(1), 26–57.

Gambetti, L., E. Pappa, and F. Canova (2008). The structural dynamics of us output and inflation: what explains the changes? *Journal of Money, Credit and Banking 40*(2-3), 369–388.

Giraitis, L. and P. C. B. Phillips (2006). Uniform limit theory for stationary autoregression. *Journal of Time Series Analysis 27*, 51–60.

Giraitis, L. and P. C. B. Phillips (2012). Mean and autocovariance function estimation near the boundary of stationarity. *Journal of Econometrics 169*(2), 166–178.

Hansen, L. P., J. C. Heaton, and N. Li (2008). Consumption strikes back? measuring long-run risk. *Journal of Political Economy 116*(2), 260–302.

Iwata, S. and S. Wu (2005). What macroeconomic risks are (not) shared by international investors? *Journal of Money, Credit, and Banking 37*(6), 1121–1141.

Jang, K. (2013). Alternative maximum likelihood estimation of structural vector autoregressive models partially identified with short-run restrictions. *Journal of Money, Credit and Banking 45*(2-3), 465–476.

Kano, T. and J. M. Nason (2014). Business cycle implications of internal consumption habit for new keynesian models. *Journal of Money, Credit and Banking 46*(2-3), 519–544.

Keating, J. W. (2013). Interpreting permanent shocks to output when aggregate demand may not be neutral in the long run. *Journal of Money, Credit and Banking 45*(4), 747–756.

Kilian, L. (2009). Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market. *American Economic Review 99*(3), 1053–69.

Kim, S. (2005). Monetary policy, foreign exchange policy, and delayed overshooting. *Journal of Money, Credit, and Banking 37*(4), 775–782.

King, T. B. and J. Morley (2007). In search of the natural rate of unemployment. *Journal of Monetary Economics 54*(2), 550–564.

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

Kuester, K. (2010). Real price and wage rigidities with matching frictions. *Journal of Monetary Economics 57*(4), 466–477.

Kurmann, A. and E. Mertens (2014). Stock prices, news, and economic fluctuations: Comment. *American Economic Review 104*(4), 1439–1445.

Kurmann, A. and C. Otrok (2013). News shocks and the slope of the term structure of interest rates. *American Economic Review 103*(6), 2612–2632.

Lanne, M. and H. Lütkepohl (2008). Identifying monetary policy shocks via changes in volatility. *Journal of Money, Credit and Banking 40*(6), 1131–1149.

Lastrapes, W. D. (2006). Inflation and the distribution of relative prices: The role of productivity and money supply shocks. *Journal of Money, Credit, and Banking 38*(8), 2159–2198.

Leeper, E. M., T. B. Walker, and S.-C. S. Yang (2013). Fiscal foresight and information flows. *Econometrica 81*(3), 1115–1145.

Liu, Z. and L. Phaneuf (2007). Technology shocks and labor market dynamics: Some evidence and theory. *Journal of Monetary Economics 54*(8), 2534–2553.

Lorenzoni, G. (2009). A theory of demand shocks. *American Economic Review 99*(5), 2050–2084.

Maćkowiak, B. (2007). External shocks, us monetary policy and macroeconomic fluctuations in emerging markets. *Journal of Monetary Economics 54*(8), 2512–2520.

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

Magdalinos, T. and P. C. Phillips (2009b). Limit theory for cointegrated systems with moderately integrated and moderately explosive regressors. *Econometric*

Theory 25(2), 482.

Marchetti, D. J. and F. Nucci (2007). Pricing behavior and the response of hours to productivity shocks. *Journal of Money, Credit and Banking 39*(7), 1587–1611.

McCarthy, J. and E. Zakrajšek (2007). Inventory dynamics and business cycles: what has changed? *Journal of Money, Credit and Banking 39*(2-3), 591–613.

Mertens, K. (2008). Deposit rate ceilings and monetary transmission in the us. *Journal of Monetary Economics 55*(7), 1290–1302.

Mertens, K. and M. O. Ravn (2013). The dynamic effects of personal and corporate income tax changes in the united states. *American Economic Review 103*(4), 1212–1247.

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

Michelacci, C. and D. Lopez-Salido (2007). Technology shocks and job flows. *Review of Economic Studies 74*(4), 1195–1227.

Miniane, J. and J. H. Rogers (2007). Capital controls and the international transmission of us money shocks. *Journal of Money, Credit and banking 39*(5), 1003–1035.

Mitra, P. (2006). Has government investment crowded out private investment in india? *American Economic Review*, 337–341.

Monacelli, T., R. Perotti, and A. Trigari (2010). Unemployment fiscal multipliers. *Journal of Monetary Economics 57*(5), 531–553.

Monnet, E. (2014). Monetary policy without interest rates: Evidence from france's golden age (1948 to 1973) using a narrative approach. *American Economic Journal: Macroeconomics 6*(4), 137–169.

Morley, J. C. (2007). The slow adjustment of aggregate consumption to permanent income. *Journal of Money, Credit and Banking 39*(2-3), 615–638.

Mumtaz, H. and F. Zanetti (2013). The impact of the volatility of monetary policy shocks. *Journal of Money, Credit and Banking 45*(4), 535–558.

Nickel, C. and A. Tudyka (2014). Fiscal stimulus in times of high debt: Reconsidering multipliers and twin deficits. *Journal of Money, Credit and Banking 46*(7), 1313–1344.

Olivei, G. and S. Tenreyro (2007). The timing of monetary policy shocks. *American Economic Review 97*(3), 636–663.

Orphanides, A. and S. Van Norden (2005). The reliability of inflation forecasts based on output gap estimates in real time. *Journal of Money, Credit and Banking 37*(3).

Paciello, L. (2011). Does inflation adjust faster to aggregate technology shocks than to monetary policy shocks? *Journal of Money, Credit and Banking 43*(8), 1663–1684.

Phillips, P. C. B. (1987). Towards a unified asymptotic theory for autoregression. *Biometrika 74*(3), 535–547.

Phillips, P. C. B. and V. Solo (1992). Asymptotics of linear processes. *Annals of Statistics 20*, 971–1001.

Primiceri, G. E. (2005). Time varying structural vector autoregressions and monetary policy. *Review of Economic Studies 72*(3), 821–852.

Ravenna, F. (2007). Vector autoregressions and reduced form representations of dsge models. *Journal of Monetary Economics 54*(7), 2048–2064.

Ravn, M. O., S. Schmitt-Grohé, and M. Uribe (2012). Consumption, government spending, and the real exchange rate. *Journal of Monetary Economics 59*(3), 215–234.

Reis, R. (2006). Inattentive consumers. *Journal of Monetary Economics 53*(8), 1761–1800.

Roush, J. E. (2007). The expectations theory works for monetary policy shocks. *Journal of Monetary Economics 54*(6), 1631–1643.

Rubio-Ramirez, J. F., D. F. Waggoner, and T. Zha (2010). Structural vector autoregressions: Theory of identification and algorithms for inference. *Review of Economic Studies 77*(2), 665–696.

Sims, C. A. and T. Zha (2006). Were there regime switches in us monetary policy? *American Economic Review 96*(1), 54–81.

Uhlig, H. (2005). What are the effects of monetary policy on output? results from an agnostic identification procedure. *Journal of Monetary Economics 52*(2), 381–419.

Walentin, K. (2014). Business cycle implications of mortgage spreads. *Journal of Monetary Economics 67*, 62–77.