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

Guillaume Chevillon*        Sophocles Mavroeidis†
ESSEC Business School       University of Oxford

Zhaoguo Zhan‡
Kennesaw State University

January 2, 2019

Contents

1 Introduction 2

2 Overview of notations 3

3 Proofs of results in the paper 3

3.1 Extending IVX to general sequences of parameters . . . . . . . . . . . . 4
3.2 Proof of Lemma P . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 6
3.3 Proof of Proposition 4 . . . . . . . . . . . . . . . . . . . . . . . . . . 7
3.4 Proof of Proposition 5 . . . . . . . . . . . . . . . . . . . . . . . . . . 11
3.5 Proof of Proposition 6 . . . . . . . . . . . . . . . . . . . . . . . . . . 13

*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

$$\Delta Y_1 = \Delta Y_2 b_{12} + X_1 \delta_1 + \varepsilon_1$$
$$\Delta Y_2 = X_2 \psi_2 + v_2$$

with $X_2 = [Y_2 : X_2 : \varepsilon_1]$, 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 $\hat{\psi}_2 = \Delta Y_2 - \hat{X}_2 \hat{\psi}_2$ 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 $\alpha_2 = c T^a$ 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, \( \text{cov}(z_t, \varepsilon_{1t}) = 0 \) 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 \( x_0 = o_p\left(\sqrt{T}^{-1/2} \wedge T^{1/2}\right) \), 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:

\[
\begin{align*}
    y_t & = \theta x_t + u_t, \\
    x_t & = \rho_T x_{t-1} + v_t, \\
    \tilde{z}_t & = \rho Z \tilde{z}_{t-1} + \Delta x_t \\
    z_t & = \rho Z z_{t-1} + v_t,
\end{align*}
\]

(S–1)

where as in MPvic,

\[
\rho_Z = 1 + c_Z T^{-b}, \quad b \in (1/2, 1), \quad c_z < 0.
\]

(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) \( Tc_T \rightarrow 0; \)
(ii) \( Tc_T \rightarrow c < 0; \)
(iii) \( Tc_T \rightarrow -\infty. \)

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 \( \varepsilon_t \) is an i.i.d sequence with \( E(\varepsilon_t) = 0, E(\varepsilon_t \varepsilon_t') = \Sigma, E(\|\varepsilon_t\|^4) < \infty, F(1) \) 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

\[
\Omega = \begin{bmatrix}
\Omega_{uu} & \Omega_{uv} \\
\Omega_{vu} & \Omega_{vv}
\end{bmatrix} = F(1) \Sigma F(1)'.
\]

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 \( \Lambda \) that conforms to \( \Omega. \)

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

\[ c_T = o(T^{-b}), \quad (S-3) \]

and expression (MPvic-13) rewrites \( \tilde{z}_t = z_t + c_T \psi_T t. \)

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 \( c_T = o(T^{-b}) \), the following approximations hold as \( T \rightarrow \infty \):

(i) \( T^{-1+b} \sum_{t=1}^{T} u_t \hat{z}_t = T^{-1+b} \sum_{t=1}^{T} u_t \hat{z}_t + o_p(1); \)
(ii) \( T^{-(1+b)} \sum_{t=1}^{T} x_t \hat{z}_t = T^{-(1+b)} \sum_{t=1}^{T} x_t z_t - T^{-1} c_T c_z \sum_{t=1}^{T} x_t^2 + o_p(1) \)
(iii) \( T^{-(1+b)} \sum_{t=1}^{T} \hat{z}_t^2 = T^{-(1+b)} \sum_{t=1}^{T} z_t^2 + o_p(1). \)

**Lemma 2*** Consider the model given by \( (S-1)-(S-2) \) under Assumptions N* and
The martingale array $U_T(s) = T^{-\frac{1+b}{2}} \sum_{t=1}^{[Ts]} [z_{t-1} F_u(1) \varepsilon_t]$ satisfies $U_T(s) \Rightarrow U(s)$ where $U_s$ is a Brownian motion with variance $\frac{1}{2c_x} \Omega_{uu} \Omega_{vv}$ and independent of $B_v$ ($B_v(s)$ defined as limit of $T^{-1/2} \sum_{t=1}^{[Ts]} v_t$). 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* (i), (ii) 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* (ii) and LP* where $\kappa_i < T^b c_T < \kappa_0$, for some $\kappa_1 < \kappa_0 < 0$, and $b \in (1/2, 1)$. Then the following approximations hold as $T \to \infty$:

(i) $\sqrt{-c_T} \sum_{t=1}^{T} u_t \tilde{z}_t = \sqrt{-c_T} \sum_{t=1}^{T} u_t x_t - \Lambda_0^{(u)} \Rightarrow N(0, \frac{1}{2} \Omega_{vv} \Omega_{uu})$;

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

(iii) $-c_T(T^{-1+j} \sum_{t=1}^{T} z_t^2) \Rightarrow \frac{1}{2} \Omega_{vv}$.

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 $\alpha_2$ in our context. For items (iv), we notice that the proof of Lemma 5 goes through with $\Delta Y_{t-i}$, $i = 1, ..., m - 1$ 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 $\alpha_2$ 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

$$\frac{1}{T} \sum_{t=1}^{T} Y_{2,t-1} X_{it} \Rightarrow \Sigma Y_{2Xi}, \quad i = 1, 2,$$

where $\Sigma Y_{2Xi}$ is nonstochastic, and

$$\sqrt{-\alpha_2 \vee T^{-1}} \sum_{t=1}^{T} Y_{2,t-1} \varepsilon_{1t} \Rightarrow N\left(0, \frac{\omega^2}{2} \sigma_{\varepsilon_1}^2\right).$$
and the fact that \( \kappa_T/T = o\left(\frac{-\alpha_2 \sqrt{T}}{T}\right) = o\left(\frac{-\alpha_2}{T}\right)\).

### 3.3 Proof of Proposition 4

The first equation is a linear IV regression, so the estimator of \( \delta_1 \) solves the equation

\[
X_1'Z_1\hat{V}_f^{-1}Z_1'\left(\Delta Y_1 - \Delta Y_2 b_{12} - X_1\hat{\delta}_1\right) = 0.
\]

Conditional homoskedasticity implies that we can set \( \hat{V}_f \) proportional to \( Z_1'Z_1 \), so \( \hat{\delta}_1 \) is 2SLS

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

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

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

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 \( T^{-1}\sum_{t=1}^{T} \hat{\varepsilon}_{1t}^2 \), where \( \hat{\varepsilon}_{1t} = \Delta Y_{1t} - \Delta Y_{2t} b_{12} - X_1'\hat{\delta}_1 \). So,

\[
\hat{\psi}_1 = \left( \begin{array}{c} \hat{\delta}_1 \\ \hat{\sigma}_{\varepsilon_1}^2 \end{array} \right)^{-1} = \left( \begin{array}{c} (X_1'X_1)^{-1}X_1'(\Delta Y_1 - \Delta Y_2 b_{12}) \\ T^{-1}\sum_{t=1}^{T} \hat{\varepsilon}_{1t}^2 \end{array} \right).
\]

(S–4)

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

\[
\bar{X}_{2t}(\theta_1) = (Y_{2,t-1}, X_{2t}', h_{1t}(\theta_1))'
\]

and the corresponding ‘generated instruments’

\[
\bar{Z}_{2t}(\theta_1) = (Z_{2t}', h_{1t}(\theta_1))' = (z_t, X_{2t}', h_{1t}(\theta_1))'.
\]

In what follows, we will omit the dependence of \( \bar{X}_{2t} \) and \( \bar{Z}_{2t} \) on \( \theta_1 \) for brevity, and we will use the shorthand notation \( \hat{X}_{2t} = \bar{X}_{2t}(b_{12}, \hat{\psi}_1) = (Y_{2,t-1}, X_{2t}', \hat{\varepsilon}_{1t})' \), 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 \( F_{2T}(b_{12}, \hat{\psi}_1, \hat{\psi}_2) = 0 \), which yields

\[
\hat{\psi}_2 = (\hat{Z}_{2t}'\hat{X}_{2t})^{-1}\hat{Z}_{2t}'\Delta Y_2.
\]

(S–5)
Subtracting $\psi_2$ and substituting for $\Delta Y_2$ yields

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

(S-6)

Collecting terms yields

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

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

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

where $\hat{\sigma}_v$ is an estimator of $\text{var} \left( \hat{\sigma}^2_{\varepsilon_1} \right)$. Under the maintained assumptions, a consistent estimator is given by $\hat{\sigma}_v = T^{-1} \sum_{t=1}^T (\varepsilon_{1t}' - \hat{\sigma}^2_{\varepsilon_1})^2$. If we assume that $\text{var} \left( \varepsilon_{1t}' \right) = 2 \sigma^4_{\varepsilon_1}$, which holds under Gaussianity, then we can use $\hat{\sigma}_v = 2 \hat{\sigma}^4_{\varepsilon_1}$, as in Blanchard and Quah (1989) and Galí (1999). Finally,

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

Next, the Jacobian of the moment conditions is given by

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

8
Hence,

\[
\begin{align*}
\hat{J}_T (b_{12})' \tilde{V}_f (b_{12})^{-1} \hat{J}_T (b_{12}) &= \left( -Z'_1 X_1 0 0 \right)' \left( Z'_1 Z_1 \right)^{-1} \hat{\sigma}^{-2} \varepsilon_1 0 0 \\
&= \left( \begin{array}{ccc}
-\hat{Z}'_1 X_1 d_{21} & 0 & 0 \\
0 & -T & 0 \\
0 & -\hat{Z}'_2 \hat{X}_2 & 0 \\
\end{array} \right) \left( \begin{array}{ccc}
0 & 0 & 0 \\
0 & T^{-1} \hat{\omega}^{-1} & 0 \\
0 & \left( \hat{Z}'_2 \hat{Z}_2 \right)^{-1} \hat{\sigma}^{-2} \varepsilon_2 & 0 \\
\end{array} \right) \\
&\times \left( \begin{array}{ccc}
0 & 0 & 0 \\
0 & -T & 0 \\
0 & -\hat{Z}'_2 \hat{X}_2 & 0 \\
\end{array} \right) \\
&= \left( \begin{array}{ccc}
X'_1 P_{Z_1} X_1 \hat{\sigma}^{-2} + X'_1 P_{\hat{Z}_2} X_1 d_{21} \hat{\sigma}^{-2} & \hat{d}_{21} X'_1 P_{\hat{Z}_2} \hat{X}_2 \hat{\sigma}^{-2} & 0 \\
0 & T \hat{\omega}^{-1} & 0 \\
\hat{d}_{21} \hat{X}'_2 P_{\hat{Z}_2} X_1 \hat{\sigma}^{-2} & 0 & \hat{X}'_2 P_{\hat{Z}_2} \hat{X}_2 \hat{\sigma}^{-2} \\
\end{array} \right).
\end{align*}
\]

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

\[
\begin{align*}
\tilde{V}_{\psi,11} &= (X'_1 X_1)^{-1} \hat{\sigma}^{-2} \\
\tilde{V}_{\psi,12} &= 0 \\
\tilde{V}_{\psi,13} &= - (X'_1 X_1)^{-1} X'_1 \hat{Z}_2 \left( \hat{X}'_2 \hat{Z}_2 \right)^{-1} \hat{\sigma}^{-2} \varepsilon_2 d_{21} \\
\tilde{V}_{\psi,22} &= \hat{\omega}^{-1} \\
\tilde{V}_{\psi,23} &= 0 \\
\tilde{V}_{\psi,33} &= \left( \hat{X}'_2 P_{\hat{Z}_2} \hat{X}_2 \right)^{-1} \hat{\sigma}^{-2} \varepsilon_2 + \left( \hat{X}'_2 \hat{Z}_2 \right)^{-1} \hat{Z}_2 P_{X_1} \hat{Z}_2 \left( \hat{X}'_2 \hat{Z}_2 \right)^{-1} \hat{\sigma}^{-2} \varepsilon_1 \hat{d}_{21}.
\end{align*}
\]

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

\[
\hat{C}_\psi' = \left( \begin{array}{ccc}
\frac{1}{2} \left( X'_1 X_1 \right)^{1/2} \hat{\sigma}^{-1} & 0 & -d_{21} X'_1 \hat{Z}_2 C'^{-1} \hat{\sigma}^{-1} \varepsilon_2 \\
0 & T^{1/2} \hat{\omega}^{-1} \hat{\sigma}^{-1/2} & 0 \\
0 & 0 & \hat{X}'_2 \hat{Z}_2 C'^{-1} \hat{\sigma}^{-1} \varepsilon_2 \\
\end{array} \right).
\]

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

So,

\[ \hat{\xi}_2 = C_{\hat{\psi}^{-1}} \left( \hat{\psi} - \psi \right) = \begin{pmatrix} (X'_1 X_1)^{-1/2} X'_1 \varepsilon_1 \hat{\sigma}^{-1} \\ \hat{\omega}^{-1/2} (\hat{\sigma}^2_{\varepsilon_1} - \sigma^2_{\varepsilon_1}) \\ C_{\hat{Z}_2' Z_2}^{-1} \hat{Z}_2' v \hat{\sigma}^{-1}_{v_2} \end{pmatrix}. \]

Finally, we turn to the derivation of \( \hat{\xi}_1 \). The moment vector \( \hat{F}_T \), 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

\[ \hat{F}_{1T} (b_{12}) = \frac{1}{T} \begin{pmatrix} Z'_1 \left[ \Delta Y_1 - b_{12} \Delta Y_2 - X_1 (X'_1 X_1)^{-1} X'_1 (\Delta Y_1 - b_{12} \Delta Y_2) \right] \\ \hat{\varepsilon}_1' \hat{\varepsilon}_1 - T \hat{\sigma}^2_{\varepsilon_1} \end{pmatrix} = \frac{1}{T} \begin{pmatrix} z' M_{X_1} (\Delta Y_1 - b_{12} \Delta Y_2) \\ 0 \end{pmatrix}, \]

and

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

Now,

\[ \hat{S}_T (b_{12}) = \hat{F}_T (b_{12})' \hat{V}_f (b_{12})^{-1} \hat{F}_T (b_{12}) = \frac{(\Delta Y_1 - b_{12} \Delta Y_2)' M_{X_1} P_{Z_1} M_{X_1} (\Delta Y_1 - b_{12} \Delta Y_2)}{\hat{\sigma}^2_{\varepsilon_1}} \]

\[ = \frac{(\Delta Y_1 - b_{12} \Delta Y_2)' P_{M_{X_1} z} (\Delta Y_1 - b_{12} \Delta Y_2)}{\hat{\sigma}^2_{\varepsilon_1}} = \hat{\xi}' \hat{\xi}_1, \]

where

\[ \hat{\xi}_1 = (z' M_{X_1} z)^{-1/2} \hat{\sigma}^{-1}_{\varepsilon_1} z' M_{X_1} (\Delta Y_1 - b_{12} \Delta Y_2) \]

\[ = (z' M_{X_1} z)^{-1/2} \hat{\sigma}^{-1}_{\varepsilon_1} z' M_{X_1} \varepsilon_1, \]

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 \( \psi \) does not depend on any weighting matrix, as seen in the proof of Proposition 4. For (ii), take \( \hat{\psi}_1 = \left( \hat{\delta}_1, \hat{\sigma}_{\varepsilon_1}^2 \right) \). Then,

\[
\hat{\delta}_1 = \delta_1 + \left( \frac{X'_1X_1}{T} \right)^{-1} \frac{X'_1\varepsilon_1}{T} = \delta_1 + O_p(1) \Rightarrow \delta_1,
\]

since \( X_1 \) consists of lags of \( \Delta Y_t \) and \( \varepsilon_1 \) is an innovation process. So,

\[
\hat{\sigma}_{\varepsilon_1}^2 = T^{-1} \sum_{t=1}^{T} \varepsilon_{1t}^2 = T^{-1} \sum_{t=1}^{T} \varepsilon_{1t}^2 + O_p(1) \Rightarrow \sigma_{\varepsilon_1}^2,
\]

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

\[
\hat{\psi}_2 - \psi_2 = \left( \overline{Z}_2\overline{X}_2 \right)^{-1} \overline{Z}_2v_2 + o_p(1).
\]  (S–7)

Next, let

\[
D_T = \begin{pmatrix}
\sqrt{\kappa_T} & 0 \\
0 & T^{-1/2}I_{p\psi_2-1}
\end{pmatrix}, \quad \kappa_T = -\frac{(c_2 + T^b\alpha_2)}{T^{1+b}},
\]  (S–8)

so that

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

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

\[
D_T \overline{Z}_2\overline{X}_2 D_T = \begin{pmatrix}
\omega + o_p(1) & O_p(T\kappa_T) & \sigma_{\varepsilon_1}^2 + o_p(1) \\
O_p(T\kappa_T) & \Sigma_{X_2X_2} + o_p(1) & o_p(1) \\
o_p(1) & o_p(1) & \sigma_{\varepsilon_1}^2 + o_p(1)
\end{pmatrix},
\]

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

\[
D_T \overline{Z}_2\overline{X}_2 D_T \Rightarrow \begin{pmatrix}
\omega & 0 & 0 \\
0 & \Sigma_{X_2X_2} & 0 \\
0 & 0 & \sigma_{\varepsilon_1}^2
\end{pmatrix}.
\]  (S–10)
If \( \alpha_2 < 0 \) is fixed, i.e., \( T \kappa_T \to -\alpha_2 \), then

\[
D_T Z_2' X_2 D_T \xrightarrow{p} \left( E \left( \frac{\alpha_2 Y_{2,t-1}}{X_{2,t}} \right) \left( \frac{\alpha_2 Y_{2,t-1}}{X_{2,t}} \right)' \begin{pmatrix} \omega & \Sigma_{zX_2} & 0 \\ \Sigma_{zX_2}' & \Sigma_{X_2 X_2} & 0 \\ 0 & 0 & \sigma_{\varepsilon_1}^2 \end{pmatrix} \right).
\] (S–11)

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

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

\[
D_T Z_2' X_2 D_T \xrightarrow{p} \Sigma_{Z_2 Z_2} = \begin{pmatrix} \omega & \Sigma_{zX_2} & 0 \\ \Sigma_{zX_2}' & \Sigma_{X_2 X_2} & 0 \\ 0 & 0 & \sigma_{\varepsilon_1}^2 \end{pmatrix}.
\] (S–12)

where

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

If \( T \alpha_2 \to c \leq 0 \), then

\[
D_T Z_2' X_2 D_T \Rightarrow \Psi_{Z_2 X_2} = \begin{pmatrix} 2\omega \left( \int_0^1 J_c dJ_c + 1 \right) & 0 & 0 \\ 0 & \Sigma_{X_2 X_2} & 0 \\ 0 & 0 & \sigma_{\varepsilon_1}^2 \end{pmatrix}.
\] (S–14)

Therefore, in both cases given in (S–12) and (S–14), \( D_T Z_2' X_2 D_T \) is invertible with probability approaching one, and hence,

\[
\left( D_T Z_2' X_2 D_T \right)^{-1} = O_p(1).
\] (S–15)

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

\[
D_T Z_2' v_2 = \begin{pmatrix} \sqrt{\kappa_T} z' v_2 \\ \sqrt{\frac{1}{T}} X_2' v_2 \\ \sqrt{\frac{1}{T}} \varepsilon_1' v_2 \end{pmatrix} = O_p(1).
\] (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 $\zeta_{Tt}$ is given by

$$\sum_{t=1}^{T} E_{F_{Tt-1}} [\zeta_{Tt} \zeta'_{Tt}] = A_T \xrightarrow{p} V_\zeta$$  \hspace{1cm} (S–17)

where

$$A_{11,T} = \sum_{t=1}^{T} \kappa_T z_t^2 E_{F_{Tt-1}} \begin{pmatrix} \varepsilon_{1t} \\ v_{2t} \end{pmatrix} \begin{pmatrix} \varepsilon_{1t} \\ v_{2t} \end{pmatrix} \xrightarrow{p} \omega \begin{pmatrix} \sigma_{\varepsilon_1}^2 & 0 \\ 0 & \sigma_{v_2}^2 \end{pmatrix} = V_{\zeta,11},$$

$$A_{12,T} = \sum_{t=1}^{T} E_{F_{Tt-1}} \begin{pmatrix} \varepsilon_{1t} \\ v_{2t} \end{pmatrix} \varepsilon_{1t} \sqrt{\frac{\kappa_T}{T}} z_t X'_{1t} \xrightarrow{p} \begin{pmatrix} \sigma_{\varepsilon_1}^2 \\ 0 \end{pmatrix} \Sigma X_1 = V_{\zeta,12},$$

by (3) and (S–19),

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

if the distribution of $\varepsilon_{1t}$ is not skewed,

$$A_{14,T} = \left( \sum_{t=1}^{T} E_{F_{Tt-1}} \begin{pmatrix} \varepsilon_{1t} \\ v_{2t} \end{pmatrix} \sqrt{\frac{\kappa_T}{T}} z_t X'_{2t} \right) \xrightarrow{p} \sum_{t=1}^{T} \sum_{t=1}^{T} \sqrt{\frac{\kappa_T}{T}} z_t E_{F_{Tt-1}} \begin{pmatrix} \varepsilon_{1t} \\ v_{2t} \end{pmatrix} \xrightarrow{p} \begin{pmatrix} 0 \\ \sigma_{v_2}^2 \Sigma X_2 \end{pmatrix} = V_{\zeta,14},$$

by (3) and (S–13),

$$A_{22,T} = \sum_{t=1}^{T} \frac{X_{1t} X'_{1t}}{T} E_{F_{Tt-1}} \xrightarrow{p} \sum X_1 X_1 \sigma_{\varepsilon_1}^2 = V_{\zeta,22},$$

$$A_{23,T} = \sum_{t=1}^{T} \frac{X_{1t} X'_{1t}}{T} E_{F_{Tt-1}} \xrightarrow{p} 0 = V_{\zeta,23},$$

if the distribution of $\varepsilon_{1t}$ is not skewed,

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

13
Putting these together, we have

\[ V_\zeta = \begin{pmatrix} \omega \left( \frac{\sigma_{\varepsilon_1}^2}{\varepsilon_1} & 0 & 0 \right) & \left( \frac{\sigma_{\varepsilon_1}^2}{\varepsilon_1} \right) \Sigma_{zX_1} & 0 & \left( \frac{0}{\sigma_{\varepsilon_2}^2} \right) \Sigma_{zX_2} & 0 \\ 0 & 0 & 0 & 0 \\ \Sigma_{X_1X_1, \sigma_{\varepsilon_1}^2} & 0 & \omega & 0 \\ 0 & \left( \sigma_{\varepsilon_2}^2 \right) & 0 & \sigma_{\varepsilon_2}^2 \end{pmatrix} \]

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

\[ \sum_{t=1}^{T} E_{F_{Tt-1}} \left( \| \zeta_{Tt} \|^2 \mathbb{1} \{ \| \zeta_{Tt} \| > \delta \} \right) \xrightarrow{p} 0 \quad \delta > 0, \]

where

\[ \| \zeta_{Tt} \|^2 = \kappa_{T} z_{Tt}^2 \left( \frac{\varepsilon_1 t}{v_{2t}} \right)^2 + \| X_{1t} \|^2 \frac{\varepsilon_1 t}{T} + \frac{(\varepsilon_1 t - \sigma_{\varepsilon}^2)^2}{T} + \| X_{2t} \|^2 \frac{v_{2t}^2}{T} + \varepsilon_1^2 t v_{2t}^2. \]

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 \( V_\zeta \) is given by (S–17).

Now, turn to the derivation of \( G_T \). First, we need an expression for \( D_T C_{Z_1 Z_2} \). Define
\[ W = (X_1, \varepsilon_1), \text{ so that} \]
\[ \mathbf{z}'_2 \mathbf{z}_2 = \begin{pmatrix} \mathbf{z}' & \mathbf{z}' \mathbf{W} \\ \mathbf{W}'_2 & \mathbf{W}' \mathbf{W} \end{pmatrix}, \]

and
\[ C \mathbf{z}'_2 \mathbf{z}_2 = \begin{pmatrix} \sqrt{\mathbf{z}' \mathbf{z}} & 0 \\ \frac{\mathbf{W}'_2}{\sqrt{\mathbf{z}' \mathbf{z}}} & (\mathbf{W}' \mathbf{M} \mathbf{W})^{1/2} \end{pmatrix}. \]

Thus,
\[ D'_T C \mathbf{z}'_2 \mathbf{z}_2 = \begin{pmatrix} \sqrt{\kappa T} & 0 \\ 0 & T^{-1/2} I_{p \phi_2 - 1} \end{pmatrix} \begin{pmatrix} \sqrt{\mathbf{z}' \mathbf{z}} & 0 \\ \frac{\mathbf{W}'_2}{\sqrt{\mathbf{z}' \mathbf{z}}} & (\mathbf{W}' \mathbf{M} \mathbf{W})^{1/2} \end{pmatrix} = \begin{pmatrix} \sqrt{\kappa T} \mathbf{z}' & 0 \\ T^{-1/2} \frac{\mathbf{W}'_2}{\sqrt{\mathbf{z}' \mathbf{z}}} & T^{-1/2} (\mathbf{W}' \mathbf{M} \mathbf{W})^{1/2} \end{pmatrix} \quad (S-18) \]

It can be verified that its inverse is
\[ \left( D'_T C \mathbf{z}'_2 \mathbf{z}_2 \right)^{-1} = \begin{pmatrix} \frac{1}{\sqrt{\kappa T} \mathbf{z}' \mathbf{z}} & 0 \\ (\mathbf{W}' \mathbf{M} \mathbf{W})^{-1/2} \frac{\mathbf{W}'_2}{\sqrt{\mathbf{z}' \mathbf{z}}} & T^{1/2} (\mathbf{W}' \mathbf{M} \mathbf{W})^{-1/2} \end{pmatrix}. \]

Hence, by simple algebra it can be verified that
\[ G' = \begin{pmatrix} \frac{1}{\sigma_{\varepsilon_1} \left( \kappa T \mathbf{z}' \mathbf{M} \mathbf{X}_1 \mathbf{z} \right)^{1/2}} & 0 & -\frac{\sqrt{T \kappa T \mathbf{z}' \mathbf{X}_1 \mathbf{X}_1} \left( \kappa T \mathbf{z}' \mathbf{M} \mathbf{X}_1 \mathbf{z} \right)^{1/2}}{\sigma_{\varepsilon_1} \left( \kappa T \mathbf{z}' \mathbf{M} \mathbf{X}_1 \mathbf{z} \right)^{1/2}} \\ 0 & 0 & \frac{\mathbf{X}'_1 \mathbf{X}_1 \left( \kappa T \mathbf{z}' \mathbf{M} \mathbf{X}_1 \mathbf{z} \right)^{-1/2}}{\sigma_{\varepsilon_1} \left( \kappa T \mathbf{z}' \mathbf{M} \mathbf{X}_1 \mathbf{z} \right)^{-1/2}} \\ 0 & 0 & \frac{1}{\sqrt{\varpi \varpi_2}} \\ 0 & -\frac{\sigma_{\varepsilon_2} \left( \kappa T \mathbf{z}' \mathbf{z} \right)^{1/2}}{\left( \mathbf{W}' \mathbf{M} \mathbf{W} \right)^{-1/2} \mathbf{W}'_2} \frac{\mathbf{W}' \mathbf{M} \mathbf{W} \mathbf{W}'}{\sqrt{\kappa T \mathbf{z}' \mathbf{z} \sigma_{\varepsilon_2}}} & 0 \end{pmatrix} \]

is such that \( G' \sum_{\mathbf{T}T} \zeta_{\mathbf{T}t} = \xi^*. \)

Finally, using the above results, it can be verified that \( G' V \xi G_T \overset{p}{\rightarrow} I_k. \)

The result that \( \hat{\xi} \overset{d}{\rightarrow} N(0, I_k) \) follows by Slutsky and the Continuous Mapping Theorem.
3.6 Proof of Proposition 7

We need to derive the asymptotic behavior of

\[
B_T \hat{C}_\psi = \begin{pmatrix}
T^{-1/2} (X'_1 X_1)^{1/2} \hat{\sigma}_{\epsilon_1}^{-1} & 0 & -d_2 T^{-1/2} X'_1 \hat{Z}_2C'_{\hat{Z}_2 Z_2} \hat{\sigma}_{v_2}^{-1} \\
0 & \hat{\omega}^{-1/2} & 0 \\
0 & 0 & D_T \hat{X}'_2 \hat{Z}_2C'_{\hat{Z}_2 Z_2} \hat{\sigma}_{v_2}^{-1}
\end{pmatrix}.
\]

First, \(T^{-1/2} (X'_1 X_1)^{1/2} \hat{\sigma}_{\epsilon_1}^{-1} \xrightarrow{p} \Sigma_1^{1/2} X'_1 \sigma_{\epsilon_1}^{-1}\) and \(\hat{\omega}^{-1/2} \xrightarrow{p} \omega^{-1/2}\). Next, by Proposition 5,

\[T^{-1/2} X'_1 \hat{Z}_2 C'_{\hat{Z}_2 Z_2} \hat{\sigma}_{v_2}^{-1} = T^{-1/2} X'_1 \hat{Z}_2 D_T D_T^{-1} C'_{\hat{Z}_2 Z_2} \sigma_{v_2}^{-1} + o_p(1),\]

and

\[D_T \hat{X}'_2 \hat{Z}_2 C'_{\hat{Z}_2 Z_2} \hat{\sigma}_{v_2}^{-1} = D_T X'_2 \hat{Z}_2 D_T D_T^{-1} C'_{\hat{Z}_2 Z_2} \sigma_{v_2}^{-1} + o_p(1).\]

Next, note that \(D_T C_{\hat{Z}_2 Z_2}\) is given in (S–18), or

\[D_T C_{\hat{Z}_2 Z_2} = \begin{pmatrix}
\sqrt{\kappa_T Z'z} \\
\frac{\sqrt{\kappa_T W'z}}{\sqrt{\kappa_T z'z}} \left( W'W - \frac{\sqrt{\kappa_T Z'z}}{\kappa_T z'z} \right)^{1/2}
\end{pmatrix},
\]

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

\[D_T C_{\hat{Z}_2 Z_2} \xrightarrow{p} \left( \frac{\sqrt{\omega}}{\sigma_z} \begin{pmatrix} 0 \\ \Sigma_{W'W} - \Sigma_{W'z} \Sigma_{z'z}^{-2} \Sigma_{W'z} \end{pmatrix} \right)^{1/2},\]

where \(\sigma_z^2 = \omega / |\alpha_2|\),

\[\Sigma_{W'W} = \begin{pmatrix} \Sigma X'_1 X_1 & 0 \\ 0 & \sigma_{\epsilon_1}^2 \end{pmatrix}, \text{ and } \Sigma_{W'z} = \begin{pmatrix} E(X_1 t z_t) \\ 0 \end{pmatrix}.
\]

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

\[D_T C_{\hat{Z}_2 Z_2} \xrightarrow{p} \begin{pmatrix} \sqrt{\omega} \\ 0 \end{pmatrix} \Sigma_{W'W}^{1/2}.
\]

In both cases, the limiting matrix will be denoted by \(C_{\Sigma_{\hat{Z}_2 Z_2}}\) and is of full rank.
Next,
\[
D_T Z_2' X_1 T^{-1/2} = \begin{pmatrix} \sqrt{T} \alpha' X_1 \\ T^{-1} X_2' X_1 \\ T^{-1} X_1' X_1' \end{pmatrix} \xrightarrow{p} \begin{pmatrix} \Sigma z X_1 \\ \Sigma X_2 X_1 \\ 0 \end{pmatrix} = \Sigma Z_2' X_1,
\]
where \( \Sigma X_2' X_1 = \lim_{T \to \infty} E (X_{2t} X_{1t}') \) and, by the same arguments as for (S–13),
\[
\Sigma z X_1 = \begin{cases} 0, & \text{if } \alpha_2 \to 0 \\ \sqrt{-\alpha_2} E (Y_{2t-1} X_{1t}'), & \text{if } \alpha_2 < 0 \text{ and fixed.} \end{cases} \tag{S–19}
\]

Finally, the limiting behavior of \( D_T Z_2' X_2 D_T \) is given by (S–11) and (S–14). Putting all of these together, we have
\[
B_T \hat{C}_\psi \Rightarrow \begin{pmatrix} \Sigma^{1/2}_{X_1' X_1} \sigma_{\epsilon_1}^{-1} & 0 & -d_2 \Sigma'_{Z_2' X_1} C_{\Sigma Z_2' Z_2}^{-1} \sigma_{v_2}^{-1} \\
0 & \omega^{-1/2} & 0 \\
0 & 0 & \Psi_{33} \end{pmatrix},
\]
where
\[
\Psi_{33} = \begin{cases} \Sigma_{Z_2' Z_2}^{-1} C_{\Sigma Z_2' Z_2}^{-1} \sigma_{v_2}^{-1}, & \text{if } T \alpha_2 \to -\infty \\
\Psi_{Z_2' Z_2}^{-1} C_{\Sigma Z_2' Z_2}^{-1} \sigma_{v_2}^{-1}, & \text{if } T \alpha_2 \to c \leq 0. \end{cases} \tag{S–20}
\]

Hence, \( \Psi \) is invertible a.s., as required. In the case \( T \alpha_2 \to c \leq 0 \), \( \Psi \) is random due to the term \( \Psi_{Z_2' X_2} \) defined in (S–14). The independence of \( \Psi \) from \( \xi \) 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

\[
\sup_{1 \leq t \leq T} \sum_{j=1}^{t} \rho_{T}^{t-j} = \frac{1 - \rho_{T}^{T}}{1 - \rho_{T}} = \begin{cases} 
O \left( -c_{T}^{-1} \right), & \text{if } Tc_{T} \to -\infty \\
O \left( T \right), & \text{if } Tc_{T} \to c < 0 \\
O \left( T \right), & \text{if } Tc_{T} \to 0
\end{cases}
\]

\[
= O \left( T \wedge |c_{T}^{-1}| \right).
\]

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:

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

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

\[
E \left\| \frac{1}{T^{1/2} \sqrt{T_{c_{T}}^{b} \vee |c_{T}^{-1}|}} \sum_{t=1}^{T} \psi_{Tt} \hat{\varepsilon}_{t} \right\|^{2} \leq \frac{E \left\| \varepsilon_{1} \right\|^{2} \sup_{1 \leq t \leq T} E \left\| \psi_{Tt} \right\|^{2}}{\frac{T_{c_{T}}^{b} \wedge |c_{T}^{-1}|}{T_{c_{T}}^{b} \vee |c_{T}^{-1}|}} \leq O \left( \frac{T_{c_{T}}^{b} \wedge |c_{T}^{-1}|}{T_{c_{T}}^{b} \vee |c_{T}^{-1}|} \right)
\]

\[
= O \left( 1 \right).
\]

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

\[
\sum_{t=1}^{T} \hat{\varepsilon}_{t} x_{t} = O_{p} \left( T \right). \quad (S–22)
\]

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)
\( \sum_{t=1}^{T} \tilde{\varepsilon}_t x_t = O_p \left( (c_T^2 T)^{-1/2} \right) = o(T). \) The framework of GP12 assumes \( c_T \in [-1, 0] \).

It is easy to see that if \( x_0 = o_p \left( \sqrt{\frac{T}{1 - T c_T}} \right) \), (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 \( t = 1, \ldots, T_0 \) and \( T_0, \ldots, T \) where only the latter use the asymptotic results of GP12, the former becoming negligible.

Now,

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

where following MPvic,

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

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

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

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

\[
\sup_{s \in [0,1]} \| x_{[sT]} \| = O_p \left( \sqrt{\frac{T}{1 - T c_T}} \right), \quad (S–23)
\]

i.e., \( \sup_{s \in [0,1]} \| x_{[sT]} \| = O_p \left( |c_T|^{-1/2} \right) \) when \( T c_T \to -\infty \) 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

\[
\sup_{1 \leq t \leq T} \| \psi T_t \| = O_p \left( \sqrt{\frac{T^{1+2b}}{1 - T \epsilon T}} \right).
\]

For part (i) of the lemma, we use

\[
\frac{1}{T^{1+b}} \left( \sum_{t=1}^{T} u_t \tilde{z}_t - \sum_{t=1}^{T} u_t z_t \right) = \frac{c_T}{T^{1+b}} \sum_{t=1}^{T} u_t \psi T_t = o_p(1),
\]

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

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

\[
T^{-1} \sum_{t=1}^{T} x_{t-1} \varepsilon_t \xrightarrow{p} 0, \quad (S-24)
\]

where \( E(x_{t-1} \varepsilon_t) = 0 \). When \( c_T \rightarrow 0 \), 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

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

For part (iii) of the lemma,

\[
\frac{1}{T^{1+b}} \left\| \sum_{t=1}^{T} \tilde{z}_t - \sum_{t=1}^{T} z_t \right\|^2 \leq \frac{c_T^2}{T^{1+b}} \sum_{t=1}^{T} \| \psi T_t \|^2 - 2 \frac{c_T}{T^{1+b}} \sum_{t=1}^{T} \| z_t \| \leq \left( \frac{-c_T \sup_{t \leq T} \| \psi T_t \|}{T^{b/2}} \right)^2 + \left( \frac{-c_T \sup_{t \leq T} \| \psi T_t \|}{T^{b/2}} \right) O_p(1),
\]

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

\[
\frac{-c_T \sup_{t \leq T} \| \psi T_t \|}{T^{b/2}} = O_p \left( \sqrt{-c_T T^{(1+b)/2}} \right) = o_p(1),
\]

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 \( \text{var} (x_t) = O \left( |c_T^{-1}| \right) \), hence

\[ \frac{-c_T}{T} \sum_{t=1}^{T} x_t^2 \xrightarrow{p} \Omega_{\text{ev}}. \]

The rest follows as in MPvic.

**MPvic-Lemma 3.2** hence also holds, where the rate of convergence is \( \frac{-c_T}{T} \sum_{t=1}^{T} x_t^2 \). 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

\[ \tilde{z}_t = x_t - \rho_z^t x_0 + \frac{c_z}{T^b} \psi_T t, \]

in (i)

\[
\sqrt{-\frac{c_T}{T}} \left( \sum_{t=1}^{T} u_t \tilde{z}_t - \sum_{t=1}^{T} u_t x_t \right) = \sqrt{-\frac{c_T}{T}} \left[ \frac{c_z}{T^b} \sum_{t=1}^{T} u_t \psi_T t - \sum_{t=1}^{T} u_t x_0 \rho_z^t \right] \\
= \frac{\sqrt{-c_T}}{T^{1/2+b}} c_z \sum_{t=1}^{T} u_t \psi_T t \xrightarrow{p} \left( \frac{\sqrt{-c_T}}{T^{1/2+b}} \sqrt{\frac{1}{1-c_T}} \right) \\
= \frac{\sqrt{-c_T}}{T^{1/2+b}} c_z \sum_{t=1}^{T} u_t \psi_T t + o_p \left( \frac{1}{T^{(1-b)/2}} \right),
\]

assuming \( x_0 = o_p \left( \sqrt{T/(1-TC_T)} \right) \) 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

\[ \sum_{t=1}^{T} u_t \psi_T t = o_p \left( T^{1/2+b} |c_T|^{-1/2} \right), \]

**QED.**
Now for part (ii),
\[
\frac{c_T}{T} \left( \sum_{t=1}^{T} x_t \tilde{z}_t - \sum_{t=1}^{T} x_t^2 \right) = \frac{c_T}{T} \left[ \frac{c_T}{T} \sum_{t=1}^{T} x_t \psi_{Tt} - \sum_{t=1}^{T} x_t x_0 \rho_z^t \right] \\
= \frac{c_T}{T^{1+b}} c_T \sum_{t=1}^{T} x_t \psi_{Tt} + o_p \left( \frac{1}{T^{1-b}} \right),
\]
as \sup_{t \leq T} \|x_t\| = O_p \left( \sqrt{\frac{T}{1-c_T}} \right), \ x_0 = o_p \left( \sqrt{\frac{T}{1-c_T}} \right) and \ \sum_{t=1}^{T} \rho_z^t = O(T^b). For the leading term, GP12-Lemma 2.1 shows that
\[
\sup_{1 \leq t \leq T} E \|x_t\|^2 = O \left( |c_T^{-1}| \right).
\]
Hence, using Proposition A.2.
\[
\left\| \frac{c_T}{T^{1+b}} c_T \sum_{t=1}^{T} x_t \psi_{Tt} \right\|_{L_1} \leq O_p \left( \frac{-c_T}{T^{b}} \left( \frac{T^b}{c_T^b - c_T} \right)^{1/2} \right) \\
= O_p \left( \frac{1}{|c_T^{1/2} T^{b/2}|} \right) = o_p(1).
\]
Finally for \(\sum_{t=1}^{T} \tilde{z}_t^2\), as in MP we only need to consider
\[
\left\| \frac{c_T}{T^{1+b}} \sum_{t=1}^{T} \psi_{Tt} x_0 \rho_z^t \right\| = o_p \left( \frac{-c_T}{T^{1+b}} \sqrt{\frac{1}{T} \sqrt{\frac{T}{1-c_T}}} T^b \right) = o_p(1),
\]
and when \(c_T T^b \rightarrow -\infty,\)
\[
E \left\| \frac{c_T}{T^{1+2b}} \sum_{t=1}^{T} \psi_{Tt}^2 \right\| \leq \frac{-c_T}{T^{2b}} \frac{T^b}{c_T^2} = \frac{1}{-c_T T^b} = o(1).
\]

**MPvic-Lemma 3.6** The results of MPvic hold when \(c_T = \kappa T^{-b}\) but we need to consider the case where \(c_T = \kappa T^{-b}\) 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

\[
(1 - \rho_z \rho_T) \frac{1}{T} \sum_{t=1}^{T} \tilde{z}_{t-1} x_{t-1} = \frac{1}{T} \sum_{t=1}^{T} x_{t-1} v_t + \frac{1}{T} \sum_{t=1}^{T} v_t \tilde{z}_{t-1} + \frac{1}{T} \sum_{t=1}^{T} v_t^2 + \frac{\kappa_T}{T^{1+b}} \sum_{t=1}^{T} x_t^2,
\]

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

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

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

\[
\frac{1}{T} \sum_{t=1}^{T} x_{t-1} v_t + \frac{1}{T} \sum_{t=1}^{T} v_t \tilde{z}_{t-1} + \frac{1}{T} \sum_{t=1}^{T} v_t^2 \xrightarrow{p} \Omega_{vv}.
\]

Therefore

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

which proves part (i).

Now for part (ii),

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

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

\[
T^{-1} \sum_{t=1}^{T} \left( v_t + \frac{T}{T^b} x_{t-1} \right) \tilde{z}_{t-1} = T^{-1} \sum_{t=1}^{T} v_t \tilde{z}_{t-1} + \frac{\kappa_T}{T^{1+b}} \sum_{t=1}^{T} x_{t-1} \tilde{z}_{t-1}
\]

\[
= \Lambda_{vv} + \frac{-\kappa_T}{2(c_z + \kappa_T)} \Omega_{vv} + o_p(1),
\]

where \(\Lambda_{vv} = \sum_{h=1}^{\infty} E(v_tv_{t-h})\).

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

\[
-(c_z + \kappa_T) T^{-(1+b)} \sum_{t=1}^{T} \tilde{z}_{t-1}^2 \xrightarrow{p} \frac{1}{2} \Omega_{vv}.
\]
For part (iii), the results follow the same lines (including the extension to MPvic-
Proposition A.2 above) and hence

$$\sqrt{- (c_z + \kappa_T) T^{-1/2} \sum z_{t-1} u_t \overset{L}{\to} N \left( 0, \frac{1}{2} \Omega_{uu} \Omega_{uu} \right).}$$

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

$$\frac{-c_T}{T} \sum_{t=1}^T x_{t-1} z_{t-1} \overset{P}{\to} \Omega_{vv},$$

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 \( b_{12} \) 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 \( \iota \) 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 \( \varepsilon_{1t} \) which asymptotically appears via the product \( T_1 \overline{z} \varepsilon_1 \) vanishes asymptotically but matters in finite samples. They hence suggest using, instead of \( P_{M_{\iota}} \), the corrected

$$\tilde{P}_{M_{\iota}} = M_{\iota} z \left( z' z - T_1 (1 - \rho_{\varepsilon_1 u_2} \overline{z} \overline{z})^{-1} z'M_{\iota} \right),$$

(S–25)
where $\hat{\rho}_{\varepsilon_1,u_2}$ is the estimated long run correlation between $\varepsilon_{1t}$ and $u_{2t}$. In (S-25) the term $\left(1 - \hat{\rho}_{\varepsilon_1,u_2}^2\right)$ 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

$$\tilde{P}_{M_{X_1}z} = M_{X_1}z \left(z' M_{X_1} z - \left(1 - \hat{\rho}_{\varepsilon_1,u_2}^2\right)T_1 zz' \right)^{-1} z' M_{X_1}, \quad (S-26)$$

where we considered only the higher order term $zz'$ instead of $M_{X_1}z M_{X_1}'$.

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

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

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

$$\hat{\alpha}_2 = \left(\hat{Z}'_2 M_{\hat{X}_{21}} \hat{X}_2\right)^{-1} \hat{Z}'_2 M_{\hat{X}_{21}} \left(\hat{X}_2 \hat{\psi}_2 + \varepsilon_2\right)$$

$$= \left(\hat{Z}'_2 M_{\hat{X}_{21}} \hat{X}_2\right)^{-1} \hat{Z}'_2 M_{\hat{X}_{21}} \left(\hat{X}_2 \hat{\psi}_2 + \left(\hat{X}_2 - \hat{X}_2\right) \hat{\psi}_2 + \varepsilon_2\right).$$

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

$$\hat{\alpha}_2 - \alpha_2 = \left(\hat{Z}'_2 M_{\hat{X}_{21}} \hat{X}_2\right)^{-1} \hat{Z}'_2 M_{\hat{X}_{21}} \left(\varepsilon_2 + P_{X_1} \varepsilon_1 d_{21}\right).$$

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

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

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

$$V_{\hat{\alpha}_2} = \left(\hat{Z}'_2 M_{\hat{X}_{21}} \hat{X}_2\right)^{-1} \hat{Z}'_2 M_{\hat{X}_{21}} \hat{Z}_2 \left(\hat{X}'_2 M_{\hat{X}_{21}} \hat{Z}_2\right)^{-1} \sigma_{\varepsilon_2}^2,$$
so the \( W \) statistic is

\[
(\hat{\alpha}_2 - \alpha_2)' V_{\hat{\alpha}_2}^{-1} (\hat{a} - a) = \frac{\varepsilon'_{2t} M_{\bar{X}_{21}} \hat{Z}_2 \left( \hat{Z}'_{2} M_{\bar{X}_{21}} \hat{Z}_2 \right)^{-1} \hat{Z}'_{2} M_{\bar{X}_{21}} \varepsilon_{2t}}{\sigma^2_{\varepsilon_2}}. 
\]

The final sample approximation of KMS consists in replacing \( \left( \hat{Z}'_{2} M_{\bar{X}_{21}} \hat{Z}_2 \right)^{-1} \) with

\[
(\varepsilon'_{2t} - T (1 - \hat{\rho}_{\varepsilon_2,u_2}^2) \overline{z} \overline{z}'')^{-1},
\]

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

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

which is in practice obtained as

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

\[
= \frac{(\hat{\alpha}_2 - \alpha_2)' \left( \hat{X}'_{2} M_{\bar{X}_{21}} \hat{Z}_2 \right) (\overline{z} \overline{z}' + T \hat{\rho}_{\varepsilon_2,u_2}^2 \overline{z} \overline{z}'')^{-1} \hat{Z}'_{2} M_{\bar{X}_{21}} \hat{Z}_2 (\hat{\alpha}_2 - \alpha_2)}{\hat{\sigma}^2_{\varepsilon_2}},
\]

where \( \hat{z} = z - \hat{\epsilon}' \). 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

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

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

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

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

\[ (S-27) \]
### 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. \(\rho\) is the correlation between the reduced-form VAR errors. The sample size is 200. Number of MC replications: 20000.

<table>
<thead>
<tr>
<th>(\rho)</th>
<th>(c) = 0</th>
<th>(c) = 0</th>
<th>(c) = 0</th>
<th>(c) = 0</th>
<th>(c) = 0</th>
<th>(c) = 0</th>
<th>(c) = 0</th>
<th>(c) = 0</th>
<th>(c) = 0</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.490</td>
<td>0.006</td>
<td>0.666</td>
<td>0.770</td>
<td>0.098</td>
<td>0.025</td>
<td>0.126</td>
<td>0.802</td>
<td>0.047</td>
</tr>
<tr>
<td></td>
<td>0.100</td>
<td>0.008</td>
<td>0.660</td>
<td>0.676</td>
<td>0.096</td>
<td>0.029</td>
<td>0.119</td>
<td>0.716</td>
<td>0.047</td>
</tr>
<tr>
<td></td>
<td>0.300</td>
<td>0.035</td>
<td>0.034</td>
<td>0.144</td>
<td>0.078</td>
<td>0.084</td>
<td>0.079</td>
<td>0.186</td>
<td>0.047</td>
</tr>
<tr>
<td></td>
<td>0.100</td>
<td>0.048</td>
<td>0.052</td>
<td>0.081</td>
<td>0.065</td>
<td>0.100</td>
<td>0.113</td>
<td>0.117</td>
<td>0.047</td>
</tr>
</tbody>
</table>

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 \(\gamma_0\) is set to zero w.l.o.g. (since the statistics are invariant to the value of the constant), and \(\gamma_x\) 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 (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
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. $\rho$ is the correlation between the reduced-form VAR errors. The sample size is 200. Number of MC replications: 20000.

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 $\rho \in \{-0.99, -0.9, \ldots, 0.99\}$, 7 points for $\omega_1 \in \{0.1, 0.4, 0.7, 1, 4, 7, 10\}$ 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 $\omega_2$, 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 $\rho, \omega_1$ and $d_{21}$.

Figure S.1 reports maximal rejection frequencies across $\rho, \omega_1$ and $c$ of the projection ARW test of $H_0: d_{21} = d_{21}^0$ as a function of $d_{21}$ 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
\( \gamma_x = 0 \), recursive detrending

<table>
<thead>
<tr>
<th>( \rho = 0.20 )</th>
<th>0.5%</th>
<th>0.95%</th>
<th>( \rho = 0.20 )</th>
<th>0.5%</th>
<th>0.95%</th>
</tr>
</thead>
<tbody>
<tr>
<td>( AR t )</td>
<td>( AR t )</td>
<td>( AR t )</td>
<td>( AR t )</td>
<td>( AR t )</td>
<td></td>
</tr>
<tr>
<td>( c = 0 )</td>
<td>0.052</td>
<td>0.009</td>
<td>0.046</td>
<td>0.158</td>
<td>0.101</td>
</tr>
<tr>
<td>( -1 )</td>
<td>0.051</td>
<td>0.009</td>
<td>0.044</td>
<td>0.139</td>
<td>0.102</td>
</tr>
<tr>
<td>( -10 )</td>
<td>0.052</td>
<td>0.016</td>
<td>0.045</td>
<td>0.102</td>
<td>0.103</td>
</tr>
<tr>
<td>( -30 )</td>
<td>0.053</td>
<td>0.033</td>
<td>0.049</td>
<td>0.078</td>
<td>0.103</td>
</tr>
<tr>
<td>( -100 )</td>
<td>0.056</td>
<td>0.049</td>
<td>0.051</td>
<td>0.056</td>
<td>0.107</td>
</tr>
</tbody>
</table>

\( \gamma_x = 0 \), full sample detrending

<table>
<thead>
<tr>
<th>( \rho = 0.20 )</th>
<th>0.5%</th>
<th>0.95%</th>
<th>( \rho = 0.20 )</th>
<th>0.5%</th>
<th>0.95%</th>
</tr>
</thead>
<tbody>
<tr>
<td>( AR t )</td>
<td>( AR t )</td>
<td>( AR t )</td>
<td>( AR t )</td>
<td>( AR t )</td>
<td></td>
</tr>
<tr>
<td>( c = 0 )</td>
<td>0.060</td>
<td>0.021</td>
<td>0.221</td>
<td>0.946</td>
<td>0.112</td>
</tr>
<tr>
<td>( -1 )</td>
<td>0.056</td>
<td>0.020</td>
<td>0.170</td>
<td>0.890</td>
<td>0.110</td>
</tr>
<tr>
<td>( -10 )</td>
<td>0.053</td>
<td>0.027</td>
<td>0.072</td>
<td>0.405</td>
<td>0.105</td>
</tr>
<tr>
<td>( -30 )</td>
<td>0.050</td>
<td>0.038</td>
<td>0.052</td>
<td>0.192</td>
<td>0.100</td>
</tr>
<tr>
<td>( -100 )</td>
<td>0.052</td>
<td>0.050</td>
<td>0.048</td>
<td>0.084</td>
<td>0.101</td>
</tr>
</tbody>
</table>

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. \( \rho \) 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

<table>
<thead>
<tr>
<th>( \rho )</th>
<th>( c )</th>
<th>( 0.05 )</th>
<th>( 0.95 )</th>
<th>( 0.05 )</th>
<th>( 0.95 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( 0 )</td>
<td>0.052</td>
<td>0.009</td>
<td>0.046</td>
<td>0.158</td>
</tr>
<tr>
<td></td>
<td>-1</td>
<td>0.051</td>
<td>0.009</td>
<td>0.044</td>
<td>0.139</td>
</tr>
<tr>
<td></td>
<td>-10</td>
<td>0.052</td>
<td>0.016</td>
<td>0.045</td>
<td>0.102</td>
</tr>
<tr>
<td></td>
<td>-30</td>
<td>0.053</td>
<td>0.033</td>
<td>0.049</td>
<td>0.078</td>
</tr>
<tr>
<td></td>
<td>-100</td>
<td>0.056</td>
<td>0.049</td>
<td>0.051</td>
<td>0.056</td>
</tr>
</tbody>
</table>

\( \gamma_x = 1 \), full sample detrending

<table>
<thead>
<tr>
<th>( \rho )</th>
<th>( c )</th>
<th>( 0.05 )</th>
<th>( 0.95 )</th>
<th>( 0.05 )</th>
<th>( 0.95 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( 0 )</td>
<td>0.060</td>
<td>0.021</td>
<td>0.221</td>
<td>0.946</td>
</tr>
<tr>
<td></td>
<td>-1</td>
<td>0.056</td>
<td>0.020</td>
<td>0.170</td>
<td>0.890</td>
</tr>
<tr>
<td></td>
<td>-10</td>
<td>0.053</td>
<td>0.027</td>
<td>0.072</td>
<td>0.405</td>
</tr>
<tr>
<td></td>
<td>-30</td>
<td>0.050</td>
<td>0.038</td>
<td>0.052</td>
<td>0.192</td>
</tr>
<tr>
<td></td>
<td>-100</td>
<td>0.052</td>
<td>0.050</td>
<td>0.048</td>
<td>0.084</td>
</tr>
</tbody>
</table>

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. \( \rho \) 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.
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. $\rho$ 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.
<table>
<thead>
<tr>
<th></th>
<th>$\rho = 0.20$</th>
<th></th>
<th>$\rho = 0.20$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\rho = 0.95$</td>
<td>$\rho = 0.95$</td>
<td></td>
</tr>
<tr>
<td>$AR$</td>
<td>$t$</td>
<td>$AR$</td>
<td>$t$</td>
</tr>
<tr>
<td>$c = 0$</td>
<td>0.044</td>
<td>0.010</td>
<td>0.048</td>
</tr>
<tr>
<td>$-1$</td>
<td>0.044</td>
<td>0.010</td>
<td>0.046</td>
</tr>
<tr>
<td>$-10$</td>
<td>0.043</td>
<td>0.016</td>
<td>0.043</td>
</tr>
<tr>
<td>$-30$</td>
<td>0.035</td>
<td>0.032</td>
<td>0.043</td>
</tr>
<tr>
<td>$-100$</td>
<td>0.030</td>
<td>0.045</td>
<td>0.058</td>
</tr>
</tbody>
</table>

**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. $ho$ 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.

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 $\chi^2_1$ instead of $\chi^2_2$ 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 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 $d_{21}$.
Figure S.2: Size of the projection ARW test of the hypothesis $H_0 : d_{21} = \bar{d}_{21}$, using $\chi^2_1$ 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 $H_0: d_{21} = \tilde{d}_{21}$, using $\chi_1^2$ 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.
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 reports the resulting power curves in each case. The figure 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 $\eta_1$ 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
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, \( \rho \) is correlation of reduced-form errors.
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.

\[ \eta_1 = 1\% \]
\[ \eta_1 = 5.13\% \]
\[ \eta_1 = 9\% \]

Figure S.5: Power of Projection ARW (solid line) 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. \( \eta_1 \) 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.

### 5.5 Concentration parameter

Identification strength is measured using an approximate formula for the concentration parameter \( \lambda \). 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
Figure S.6: Power of Projection ARW (solid line) 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. \( \eta_1 \) 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.

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

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

The instrument \(z_t = \sum_{j=1}^{t-1} \rho_{t-j}^T \Delta Y_{2,j}\) depends on two parameters such that \(\rho_T^z = 1 + \frac{c_z}{b}\) 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 \((c, b_{12}, \rho),\) our treatment on \(c\) and \(\rho\) 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.

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.

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.

5.6.2 Varying \(\rho\)

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 \(\rho\) (here \(c_z = -1)\).
Size and Power of AR as $b$ varies ($c_z = -1$), $H_0$: $b_{12} = 0$, $T = 200$

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

Size and Power of AR as $c_z$ varies ($b = 0.95$), $H_0$: $b_{12} = 0$, $T = 200$

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

### 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
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 $\rho$ and the vertical axis the rejection probability.
values $b_{12}$. The figure reports 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.

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

### 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 an SVAR(2) model. Here we give the details of the implementation. Let $\hat{\phi}$ denote the OLS estimator of $\phi$ in the regression $Y_{2t} = \mu + \phi Y_{2,t-1} + e_t$. The coefficients $\Psi_1$ are estimated by OLS in the system of equations

$$
(I - \Psi_1 L) \begin{bmatrix} 1 - L & 0 \\ 0 & 1 - \hat{\phi} L \end{bmatrix} Y_t = u_t.
$$
Namely, denoting \( \tilde{Y}_t := (I - \tilde{\Phi} L) Y_t, \tilde{\Psi}_1 \) is simply obtained from a VAR(1) using OLS on \( \tilde{Y}_t = \Psi_1 \tilde{Y}_{t-1} + u_t \), i.e.,

\[
\tilde{\Psi}_1 = \sum_t \tilde{Y}_t \tilde{Y}_t' \left( \sum_t \tilde{Y}_t \tilde{Y}_t' \right)^{-1}.
\]

Now,

\[
vec \left( \tilde{\Psi}_1 - \Psi_1 \right) = \left( \left( \sum_t \tilde{Y}_t \tilde{Y}_t' \right)^{-1} \otimes I_2 \right) vec \sum_t u_t \tilde{Y}_t' = \left( \left( \sum_t \tilde{Y}_t \tilde{Y}_t' \right)^{-1} \otimes I_2 \right) \sum_t (\tilde{Y}_t \otimes u_t)
\]

so, under homoskedasticity,

\[
\sim var \left[ \sqrt{T} vec \left( \tilde{\Psi}_1 \right) \right] = \left( \frac{1}{T} \sum_t \tilde{Y}_t \tilde{Y}_t' \right)^{-1} \otimes \hat{\Sigma}_u \sim N(0, \Xi)
\]

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

\[
\sqrt{T} \left( \hat{\psi} - \psi \right) \xrightarrow{d} N(0, \Xi),
\]

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

\[
\left( \frac{1}{T} \sum_t \tilde{Y}_t \tilde{Y}_t' \right)^{-1} \otimes \hat{\Sigma}_u,
\]

with

\[
\hat{\Sigma}_u = \frac{1}{T} \sum_t \left( \tilde{Y}_t - \tilde{\Psi}_1 \tilde{Y}_{t-1} \right) \left( \tilde{Y}_t - \tilde{\Psi}_1 \tilde{Y}_{t-1} \right)'.
\]
Gospodinov’s (2010) estimator is \( \hat{b}_{12} := \hat{\psi}_{12}/\hat{\psi}_{22} = f (\hat{\psi}) \) with

\[
\frac{\partial f}{\partial \psi} = \left[ \frac{1}{\psi_{22}} \; - \frac{\psi_{12}}{\psi_{22}^2} \right] := F'_\psi.
\]

Now, the Delta method yields

\[
\sqrt{T} \left( \hat{b}^{(0)}_{12} - b_{12} \right) \rightarrow^d N (0, F'_\psi \Xi F_\psi)
\]

from which we obtain a \( t \)-statistic for \( b_{12} \)

\[
t_{\hat{b}_{12}^{(0)}} \equiv \sqrt{T} \frac{\hat{b}_{12} - b_{12}}{\sqrt{F'_\psi \Xi F_\psi}}.
\]

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 St. Louis Fed database FRED. The updated data are presented in Figure S.14.
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)
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. \(\hat{g}(b_{12}) = g(b_{12}, \hat{\psi}(b_{12}))\) its restricted estimate at $b_{12}$, and $\hat{\sigma}_g(b_{12})$ the associated standard error computed using the delta method.

The joint $\eta$-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
\[
C_{b_{12}} = \{b_{12}^0 \in \mathbb{R} : AR(b_{12}^0) < c_\eta\},
\]
(S–28)
where $c_\eta$ is the $1 - \eta$ quantile of the $\chi^2_2$ 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
\[
\hat{g}(b_{12}^0) \pm \hat{\sigma}_g(b_{12}^0) \sqrt{c_\eta - AR(b_{12}^0)}.
\]
(S–29)
The upper and lower bounds of the projection-based confidence set for $g$ are given by
\[
\left[\min_{b_{12}^0 \in C_{b_{12}}} g(b_{12}^0), \max_{b_{12}^0 \in C_{b_{12}}} \bar{g}(b_{12}^0)\right].
\]
(S–30)
The procedure is repeated for each impulse response, using the same $C_{b_{12}}$, 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 $C_{b_{12}}$, because the extrema of $g(b_{12}^0)$ and $\bar{g}(b_{12}^0)$ may occur at interior points of $C_{b_{12}}$, and the functions $g(\cdot)$ and $\bar{g}(\cdot)$ could be very steep.

An alternative to the projection method is the Bonferroni method. This involves combining an $\eta_1$-level $AR$ test with an $\eta_2$-level Wald test for $g$. Thus, $C_{b_{12}}$ is obtained by replacing $c_\eta$ in (S–28) with the $1 - \eta_1$ quantile of the $\chi^2_1$ distribution (note the difference also in degrees of freedom), and the term $\sqrt{c_\eta - AR(b_{12}^0)}$ in (S–29) with the $1 - \eta_2/2$ quantile of the standard normal distribution. The resulting interval in (S–30)
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.
6.3.2 Alternative detrending of hours

Francis and Ramey (2009) 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 an 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 (2009). 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>
<thead>
<tr>
<th>With long-run restrictions</th>
<th>Without long-run restrictions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alvarez and Jermann 2005</td>
<td>Iwata and Wu 2005</td>
</tr>
<tr>
<td>Bernanke, Boivin, Dean, and Eliaasz 2005</td>
<td>Kim 2005</td>
</tr>
<tr>
<td>Francis and Ramey 2005</td>
<td>Primiceri 2005</td>
</tr>
<tr>
<td>Orphanides and Van Norden 2005</td>
<td>Uhlig 2005</td>
</tr>
<tr>
<td>Beaudry and Portier 2006</td>
<td>Ashcraft 2006</td>
</tr>
<tr>
<td>Chang and Hong 2006</td>
<td>Basu, Fernald, and Kimball 2006</td>
</tr>
<tr>
<td>Cover, Enders, and Huang 2006</td>
<td>Braun and Shioji 2006</td>
</tr>
<tr>
<td>Croushore and Evans 2006</td>
<td>Farrant and Peersman 2006</td>
</tr>
<tr>
<td>Fisher 2006</td>
<td>Mitra 2006</td>
</tr>
<tr>
<td>Lastaripes 2006</td>
<td>Sims and Zha 2006</td>
</tr>
<tr>
<td>Reis 2006</td>
<td>Diedola and Neri 2007</td>
</tr>
<tr>
<td>Avouy-Dovi and Matheron 2007</td>
<td>Mackowiak 2007</td>
</tr>
<tr>
<td>Fernald 2007</td>
<td>Miniane and Rogers 2007</td>
</tr>
<tr>
<td>King and Morley 2007</td>
<td>Oliven and Tenreyro 2007</td>
</tr>
<tr>
<td>Lau and Phaneuf 2007</td>
<td>Kohn 2007</td>
</tr>
<tr>
<td>Marchetti and Nucci 2007</td>
<td>Benati 2008</td>
</tr>
<tr>
<td>Michelacci and Lopez-Salido 2007</td>
<td>Bittric, Meier, and Muller 2008</td>
</tr>
<tr>
<td>Morley 2007</td>
<td>Gambetti, Pappa, and Canova 2008</td>
</tr>
<tr>
<td>Ravena 2007</td>
<td>Lanne and Lutkepohl 2008</td>
</tr>
<tr>
<td>Chari, Kehoe, and McGrattan 2008</td>
<td>Mertens 2008</td>
</tr>
<tr>
<td>Corsetti, Dedola, and Dedic 2008</td>
<td>Altavilla and Ciccarelli 2009</td>
</tr>
<tr>
<td>Hansen, Healon, and Li 2008</td>
<td>Benati and Surico 2009</td>
</tr>
<tr>
<td>Bjornland and Letemo 2009</td>
<td>Boivin, Giannoni, and Mihov 2009</td>
</tr>
<tr>
<td>Dupor, Han, and Tsi 2009</td>
<td>Carlstrom, Fuerst, and Paustan 2009</td>
</tr>
<tr>
<td>Feve and Guay 2009</td>
<td>Del Negro and Schorlemer 2009</td>
</tr>
<tr>
<td>Francis and Ramey 2009</td>
<td>Evans and Marshall 2009</td>
</tr>
<tr>
<td>Gambetti and Gali 2009</td>
<td>Kilian 2009</td>
</tr>
<tr>
<td>Lorenzoni 2009</td>
<td>Dantinne and Kurmann 2010</td>
</tr>
<tr>
<td>Feve, Matheron, and Sahuc 2010</td>
<td>Elder and Sargent 2010</td>
</tr>
<tr>
<td>Pomi and Gambetti 2010</td>
<td>Kuester 2010</td>
</tr>
<tr>
<td>Rubio-Ramirez, Waggoner, and Zha 2010</td>
<td>Monacelli, Perotti, and Trigari 2010</td>
</tr>
<tr>
<td>Beaudry, Collard, and Porter 2011</td>
<td>Barsky and Sims 2011</td>
</tr>
<tr>
<td>Paciello 2011</td>
<td>Born and Mulier 2012</td>
</tr>
<tr>
<td>Hackmann and Simis 2012</td>
<td>Ravn, Schmitt-Grohe, and Uribe 2012</td>
</tr>
<tr>
<td>Collard and Delles 2012</td>
<td>Barakhshian and Crow 2013</td>
</tr>
<tr>
<td>Corsetti and Kostantinos 2012</td>
<td>Baumeister and Peersman 2013</td>
</tr>
<tr>
<td>Bekert, Hoerova, and Duca 2013</td>
<td>Cloyne 2013</td>
</tr>
<tr>
<td>Blanchard, L'Huillier, and Lorenzoni 2013</td>
<td>Jiang 2013</td>
</tr>
<tr>
<td>Keating 2013</td>
<td>Kurmann and Otrok 2013</td>
</tr>
<tr>
<td>Pomi and Gambetti 2014</td>
<td>Leeper, Walker, and Yang 2013</td>
</tr>
<tr>
<td>Kaho and Nison 2014</td>
<td>Mertens and Ravn 2013</td>
</tr>
<tr>
<td>Kurmann and Mertens 2014</td>
<td>Muntz and Zanetti 2013</td>
</tr>
<tr>
<td></td>
<td>Mertens and Ravn 2014</td>
</tr>
<tr>
<td></td>
<td>Monnet 2014</td>
</tr>
<tr>
<td></td>
<td>Nickel and Tudyka 2014</td>
</tr>
<tr>
<td></td>
<td>Walentin 2014</td>
</tr>
</tbody>
</table>

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


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


