Web-appendix for:

Identifying Modern Macro Equations with Old Shocks

Regis Barnichon[(a)] *and* *Geert Mesters*[(b)]

[(a)] Federal Reserve Bank of San Francisco and CEPR

[(b)] Universitat Pompeu Fabra, Barcelona GSE and VU Amsterdam

March 15, 2020

Abstract

In this web-appendix we provide the following additional results.

1. Detailed implementation guide for the methodology.
2. A formal derivation of how we can re-state the exogeneity and relevant conditions, as discussed in Section 3.1 of the main paper.
3. Proof of Theorem 1 which defines the limiting distributions of the AR_a and $AR_{a,s}$ statistics.
4. Comparison of alternative ways of combining the structural shocks to form instruments following Eberly, Stock and Wright (2019).
5. Additional simulation results for models with heteroskedastic and serially correlated errors.
6. Additional empirical results for: (i) different choices for H, (ii) different sampling periods using the Romer and Romer monetary shocks as instruments, and (iii) the traditional approach based on lagged instruments.

1 Implementation guide

In a typical macro application, the point is to construct a confidence region for the parameters δ, or for a subset of the parameters, say β. In this section we provide details for constructing such confidence sets based on inverting the AR_a and $AR_{a,s}$ statistics.

1.1 AR_a based confidence regions

Let $\mathfrak{D} \subset \mathbb{R}^3$ be a finite set that with high probability contains the confidence region of δ.[1] To compute a 1-α confidence set for δ based on the $AR_a[\delta_0]$ statistic we follow the following algorithm.

- for each $\delta_0 \in \mathfrak{D}$

    1. Compute $AR_a[\delta_0]$ as in equation (17) of the main text.
    2. Evaluate:
        - If $AR_a[\delta_0] < \chi^2_{1-\alpha}(3)$ include δ_0 in the confidence region.
        - If $AR_a[\delta_0] > \chi^2_{1-\alpha}(3)$ do not include δ_0 in confidence set.

    where $\chi^2_{1-\alpha}(3)$ is the $1-\alpha$ critical value of the $\chi^2(3)$ distribution.

The implementation of step 1, requires choosing H and computing an estimate for the long run variance of $\{u_t\}$. We do not provide a formal theory for optimally selecting H but our simulation evidence indicates that $H = 20$ provides a reliable choice for sample sizes $n = 200, 500$. For quarterly data this corresponds to an impulse response of 5 years which is reasonable for most macroeconomic applications. Importantly, the limiting distribution of the AR_a statistic requires that $H/n \to c \in (0, 1)$, which implies that H should not be chosen too small.[2]

[1] In the Phillips Curve application we typically consider the cube $[-10, 10]^3$ with grid size 0.01. This set covers all plausible parameters for the Phillips curve and is accurate up to two digits.

[2] If selecting H very small is necessary then we recommend to use the standard serial correlation adjusted AR statistic for inference, see the discussion in Andrews, Stock and Sun (2019).

Second, we compute the long run variance $\hat{s}_u^2$ as follows

$$\hat{s}_u^2 = \frac{1}{n-H} \sum_{t=H+1}^{n} \sum_{s=H+1}^{n} \hat{u}_t \hat{u}_s \kappa((t-s)/b_n)$$

where $\hat{u}_t = (y_t - w_t'\delta_0) - z_t^{i'}\hat{\theta}_a$ and $\hat{\theta}_a = (\sum_{t=H+1}^{n} z_t^i z_t^{i'})^{-1} \sum_{t=H+1}^{n} z_t^i (y_t - w_t'\delta_0)$. We take the kernel function $k()$ as the quadratic spectral kernel, see Andrews (1991), and the bandwidth parameter is chosen as $b_n = \lfloor 4((n-H)/100)^{2/9} \rfloor + 1$.

1.2 Subset $AR_{a,s}$ based confidence regions

To construct a confidence region for a subset of the parameters, say β when $\delta = (\beta', \alpha')'$, let $\mathfrak{B} \subset \mathbb{R}^{\dim(\beta)}$ be a finite set that with high probability contains the confidence region of β. To compute a 1-α confidence set for β based on the subset $AR_{a,s}[\beta_0]$ statistic we follow the following algorithm.

- for each $\beta_0 \in \mathfrak{B}$

  1. Compute $AR_{a,s}[\beta_0]$ as in equation (19) of the main text.
  2. Evaluate:

     - If $AR_{a,s}[\beta_0] < \chi^2_{1-\alpha}(\dim(\beta))$ include β_0 in the confidence region.
     - If $AR_{a,s}[\beta_0] > \chi^2_{1-\alpha}(\dim(\beta))$ do not include β_0 in confidence set.

     where $\chi^2_{1-\alpha}(\dim(\beta))$ is the $1-\alpha$ critical value of the $\chi^2(\dim(\beta))$ distribution.

Recall, that the subset AR statistic is given by

$$AR_{a,s}[\beta_0] = \min_{\alpha \in \mathbb{R}^{\dim(\alpha)}} AR_a[(\beta_0', \alpha')'] \ .$$

We compute the $AR_a[(\beta_0', \alpha')']$ statistic similarly as discussed in the previous section. The minimization problem $\min_{\alpha \in \mathbb{R}^{p_\alpha}} AR_a[(\beta_0', \alpha')']$ can be solved numerically or analytically. Using the analytical solution is attractive in practice as it speeds up computations. As we show in Lemma 6 below (which is part of the proof of Theorem 1) we have that

$$AR_{a,s}[\beta_0] = \mu_{\min}\left((Y^{*'}\mathcal{R}_Z Y^*)^{-1/2} Y^{*'} P_Z Y^* (Y^{*'}\mathcal{R}_Z Y^*)^{-1/2'}\right)$$

where $\mu_{\min}(B)$ denotes the smallest eigenvalue of the matrix B, $Y^* = [y - W_\beta\beta_0 : W_\alpha]$, $\mathcal{R}_Z = M_Z B_n M_Z/(n-H)$, $P_Z = Z(Z'Z)^{-1}Z'$, $M_Z = I - P_Z$ and B_n is an $(n-H)\times(n-H)$ matrix with t, s entry equal to $\kappa((t-s)/b_n)$. Note that here $y = (y_{H+1}, \ldots, y_n)'$, $Z = (z^i_{H+1}, \ldots, z^i_n)'$ $W_\alpha = (w_{\alpha,H+1}, \ldots, w_{\alpha,n})'$ and $W_\beta = (w_{\beta,H+1}, \ldots, w_{\beta,n})'$, see equation (1) below for more details.

This shows that computing the analytical solution only requires solving a low-dimensional eigenvalue problem, which is typically faster when compared to numerically minimizing $AR_a[(\beta_0', \alpha')']$ with respect to α.

2 Derivation of Exogeneity and Relevance conditions

In this section we present a formal derivation of how we can re-state the exogeneity and relevant conditions, as discussed in Section 3.1 of the main paper.

The stationarity and uncorrelated assumptions imply that

$$\mathrm{E}(\varepsilon^i_t \varepsilon^i_s) = \begin{cases} \sigma^2_\varepsilon & t = s \\ 0 & t \neq s \end{cases},$$

for all $t, s = 1, \ldots, n$. Additionally, from the linearity assumption we have that we can write for each endogenous variable $w_{j,t} = \mathcal{R}^{j'}\varepsilon^i_{t:t-H} + \eta^j_t$, for $j = 1, 2, 3$. Similarly, for the error term $u_t = \mathcal{R}^{u'}\varepsilon^i_{t:t-H} + \eta^u_t$. The disturbances η^j_t and η^u_t are mean zero and uncorrelated with $\varepsilon^i_{t:t-H}$.

Next, we rewrite the exogeneity assumption. We have that for each $h = 0, \ldots, H$

$$\mathrm{E}(\varepsilon^i_{t-h} u_t) = \mathrm{E}(\varepsilon^i_{t-h}(\mathcal{R}^{u'}\varepsilon^i_{t:t-H} + \eta^u_t)) = \sigma^2_\varepsilon \mathcal{R}^u_h .$$

Since, $\sigma_\varepsilon^2 > 0$, the exogeneity condition can only be satisfied when $\mathcal{R}_h^u = 0$ for all $h = 0, \dots, H$.

For the relevance condition we have for $j = 1, 2, 3$ and $h = 0, \dots, H$ that

$$\mathrm{E}(\varepsilon_{t-h}^i w_{j,t}) = \mathrm{E}(\varepsilon_{t-h}^i (\mathcal{R}^{j\prime} \varepsilon_{t:t-H}^i + \eta_t^j)) = \sigma_\varepsilon^2 \mathcal{R}_h^j \ .$$

Using this we obtain

$$\mathrm{E}(\varepsilon_{t:t-H}^i w_t') = \sigma_\varepsilon^2 \begin{bmatrix} \mathcal{R}_0^1 & \mathcal{R}_0^3 \\ & \\ \mathcal{R}_H^1 & \mathcal{R}_H^3 \end{bmatrix}$$

and it follows that requiring $\mathrm{E}(\varepsilon_{t:t-H}^i w_t')$ to be full column rank is equivalent to requiring $[\mathcal{R}_h^1, \mathcal{R}_h^2, \mathcal{R}_h^3]_{h=0}^H$ to be full column rank (or linearly independent).

3 Proof of Theorem 1

The proof of Theorem 1 proceeds as follows. We first provide some minor notation details. Then we show a set of intermediate results that are combined to prove Theorem 1. The proofs for the intermediate results are deferred to the end of this document.

3.1 Some notation

Throughout the proof we write ξ_t for the structural shock proxies and z_t for the instruments, thus omitting the indicator i from the notation. Further, we often consider the linear IV model of assumption 1 in matrix notation.

$$\begin{aligned} y &= W\delta + U \\ &= W_\beta \beta + W_\alpha \alpha + u \\ & \qquad\qquad\qquad\qquad , \qquad (1) \\ \underbrace{(W_\beta : W_\alpha)}_{W} &= Z \underbrace{(\Pi_\beta : \Pi_\alpha)}_{\Pi} + \underbrace{(V_\beta : V_\alpha)}_{V} \end{aligned}$$

where $y = (y_{H+1}, \ldots, y_n)' \in \mathbb{R}^{n-H}$, $W = (w_{H+1}, \ldots, w_n)' \in \mathbb{R}^{(n-H)\times m}$, $W_\alpha = (w_{\alpha,H+1}, \ldots, w_{\alpha,n})' \in \mathbb{R}^{(n-H)\times m_\alpha}$, $W_\beta = (w_{\beta,H+1}, \ldots, w_{\beta,n})' \in \mathbb{R}^{(n-H)\times m_\beta}$, $V = (v_{H+1}, \ldots, v_n)' \in \mathbb{R}^{(n-H)\times m}$, $V_\alpha = (v_{\alpha,H+1}, \ldots, v_{\alpha,n})' \in \mathbb{R}^{(n-H)\times m_\alpha}$, $V_\beta = (v_{\beta,H+1}, \ldots, v_{\beta,n})' \in \mathbb{R}^{(n-H)\times m_\beta}$, $u = (u_{H+1}, \ldots, u_n)' \in \mathbb{R}^{(n-H)}$ and $Z = (z_{H+1}, \ldots, z_n)' \in \mathbb{R}^{(n-H)\times 3}$.

For any real matrix A we define $\|A\| = \sqrt{\mu_{\max}(A'A)}$, where $\mu_{\max}(B)$ denotes the largest eigenvalue of a square matrix B. Further, for a random matrix X we define $\|X\|_r = \mathrm{E}\left(\sum_i \sum_j |X_{i,j}|^r\right)^{1/r}$ for positive integers r. Weak convergence, as defined in Section 26.3 of Davidson (1994), is denoted by $\Rightarrow$. Finally, we use $\lfloor x \rfloor$ to denote the largest integer not exceeding x.

3.2 Intermediate results

The following 7 lemmas are used to prove theorem 1.

Lemma 1. *For integers $p, q \geq 0$ define*

$$\Psi_{p,q,n} = \begin{bmatrix} n^{p+1/2} & 0 \\ 0 & n^{q+1/2} I_4 \end{bmatrix} \qquad \textit{and} \qquad T_{t,p,q} = \begin{bmatrix} t^p & 0 \\ 0 & t^q I_4 \end{bmatrix} .$$

Given assumption 1, for $a \in [0,1]$, we have for $n \to \infty$ that

$$\Psi_{p,q,n}^{-1} \sum_{t=1}^{\lfloor na \rfloor} T_{t,p,q} \eta_t \Rightarrow G(a) = \begin{pmatrix} G_{\xi,p}(a) \\ G_{uv,q}(a) \end{pmatrix} ,$$

where the scalar process $G_{\xi,p}(a)$ and the 4×1 process $G_{uv,q}(a) = (G_{u,q}(a), G_{v_1,q}(a), G_{v_2,q}(a), G_{v_3,q}(a))'$ are independent Gaussian processes with a.s. continuous sample paths, independent increments and variances

$$\mathrm{E}\left(G_{\xi,p}(a)^2\right) = a^{2p+1}\omega_{\xi,p}^2 \qquad \textit{and} \qquad \mathrm{E}\left(G_{uv,q}(a) G_{uv,q}(a)'\right) = a^{2q+1}\Omega_{uv,q} .$$

Lemma 2. *Given assumption 1 we have for $y = u$ or $y = v_j$, for $j = 1, \ldots, m$, when $n \to \infty$*

with $H/n \to c \in (0,1)$ *that*

$$\frac{1}{n^{1+p+q}} \sum_{t=H+1}^{n} \sum_{s=1}^{t} s^p \xi_s t^q y_t \Rightarrow \int_c^1 G_{\xi,p}(a) \mathrm{d} G_{y,q}(a) \tag{2}$$

and

$$\frac{1}{n^{1+p+q}} \sum_{t=H+2}^{n} \sum_{s=1}^{t-H-1} s^p \xi_s t^q y_t \Rightarrow \int_c^1 G_{\xi,p}(a-c) \mathrm{d} G_{y,q}(a) \ , \tag{3}$$

where G *is the Gaussian process defined in Lemma 1 and* $G_{y,q}$ *is the corresponding element of* $G_{uv,q} = (G_{u,q}, G_{v_1,q}, G_{v_2,q}, G_{v_3,q})'$.

Lemma 3. *Given assumption 1 we have when* $n \to \infty$ *with* $H/n \to c \in (0,1)$ *and* $K_n = \mathrm{diag}(n, n^2, n^3)$ *that*

$$(I_4 \otimes K_n^{-1}) \sum_{t=H+1}^{n} \begin{pmatrix} u_t \\ v_t \end{pmatrix} \otimes z_t \Rightarrow \Xi = \begin{bmatrix} \Xi_u \\ \Xi_{v_1} \\ \Xi_{v_2} \\ \Xi_{v_3} \end{bmatrix} \tag{4}$$

where Ξ *is such that conditional on* D_ξ *we have*

$$\Xi|_{D_\xi} \sim N\left(0, \Omega_{uv,0} \otimes \int_c^1 D_\xi(a) D_\xi'(a) \mathrm{d}a\right)$$

and $D_\xi(a) = (D_{1,\xi}(a), D_{2,\xi}(a), D_{3,\xi}(a))'$ *with elements*

$$D_{1,\xi}(a) = G_{\xi,0}(a) - G_{\xi,0}(a-c)$$

$$D_{2,\xi}(a) = aG_{\xi,0}(a) - aG_{\xi,0}(a-c) - G_{\xi,1}(a) + G_{\xi,1}(a-c)$$

$$D_{3,\xi}(a) = a^2 G_{\xi,0}(a) - a^2 G_{\xi,0}(a-c) - 2aG_{\xi,1}(a) + 2aG_{\xi,1}(a-c) + G_{\xi,2}(a) - G_{\xi,2}(a-c)$$

Lemma 4. *Given assumption 1 we have when $n \to \infty$ with $H/n \to c \in (0,1)$ that*

$$K_n^{-1} \sum_{t=H+1}^{n} z_t z_t' K_n^{-1} \Rightarrow \int_c^1 D_\xi(a) D_\xi'(a) \mathrm{d}a \tag{5}$$

where $K_n = \mathrm{diag}(n, n^2, n^3)$ and $D_\xi(a)$ is as defined in Lemma 3.

Lemma 5. *Given assumption 1, let $\widehat{S}_{uv} = \frac{1}{n}\sum_{t=H+1}^{n}\sum_{s=H+1}^{n}(\hat{u}_t, \hat{v}_t')'(\hat{u}_s, \hat{v}_s')\kappa((t-s)/b_n) = \frac{1}{n}(\hat{u} : \widehat{V})'B_n(\hat{u} : \widehat{V})$, where $\hat{u} = M_Z u$, $\widehat{V} = M_Z V$, $M_Z = I_{n-H} - Z(Z'Z)^{-1}Z'$ and B_n is an $(n-H) \times (n-H)$ matrix with s, t entry equal to $\kappa((t-s)/b_n)$. Also, define $S_{uv} = \frac{1}{n}\sum_{t=H+1}^{n}\sum_{s=H+1}^{n}(u_t, v_t')'(u_s, v_s')\kappa((t-s)/b_n) = \frac{1}{n}(u : V)'B_n(u : V)$. Under the conditions of assumption 1 we have when $n \to \infty$ with $H/n \to c \in (0,1)$ that*

$$S_{uv} \xrightarrow{p} \Omega_{uv,0} \qquad \text{and} \qquad \widehat{S}_{uv} = S_{uv} + o_p(1)\ .$$

Lemma 6. *Given assumption 1, under $H_0 : \beta = \beta_0$ we have wpa1*

$$AR_{a,s}[\beta_0] = \min_{g \in \mathcal{R}^{1+m_2}\setminus\{0\}\}} \frac{g'(\Omega_{uv_\alpha,0}^{1/2}\widehat{S}_{uv_\alpha}^{-1/2})'N_n L_n N_n(\Omega_{uv_\alpha,0}^{1/2}\widehat{S}_{uv_\alpha}^{-1/2})g}{g'g}$$

and

$$AR_{a,s}[\beta_0] \leq \frac{r_{u,n}' M_{\tau_n} r_{u,n}}{\rho_n}$$

where $\widehat{S}_{uv_\alpha} = \frac{1}{n}\sum_{t=H+1}^{n}\sum_{s=H+1}^{n}\kappa((t-s)/b_n)(\hat{u}_t, \hat{v}'_{\alpha,t})'(\hat{u}_t, \hat{v}'_{\alpha,t})$, $\hat{v}_t = (\hat{v}'_{\beta,t}, \hat{v}'_{\alpha,t})'$ *and*

$$\widehat{S}_{uv_\alpha} = \begin{bmatrix} \hat{s}_u^2 & \hat{s}_{uv_\alpha} \\ \hat{s}_{v_\alpha u} & \widehat{S}_{v_\alpha v_\alpha} \end{bmatrix}$$

$$\Omega_{uv_\alpha,0} = \begin{bmatrix} \omega_{u,0}^2 & \omega_{uv_\alpha,0} \\ \omega_{v_\alpha u,0} & \Omega_{v_\alpha v_\alpha,0} \end{bmatrix}$$

$$\Omega_{v_\alpha v_\alpha \cdot u} = \Omega_{v_\alpha v_\alpha,0} - \omega_{v_\alpha u,0}\omega_{u,0}^{-2}\omega_{uv_\alpha,0}$$

$$r_{u,n} = (Z'Z)^{-1/2}Z'u\omega_{u,0}^{-1}$$

$$r_{v_\alpha,n} = (Z'Z)^{-1/2}Z'(V_\alpha - u\omega_{u,0}^{-2}\omega_{uv_\alpha,0})\Omega_{v_\alpha v_\alpha \cdot u}^{-1/2}$$

$$\Theta_n = (Z'Z)^{1/2}\Pi_\alpha\Omega_{v_\alpha v_\alpha \cdot u}^{-1/2}$$

$$\tau_n = \Theta_n + r_{v_\alpha,n}$$

$$\eta_n = (\tau_n'\tau_n)^{-1/2}\tau_n' r_{u,n}$$

$$\rho_n = (1, -\eta_n'(\tau_n'\tau_n)^{-1/2})(\Omega_{uv_\alpha,0}^{-1/2\prime}\widehat{S}_{uv_\alpha}\Omega_{uv_\alpha,0}^{-1/2})(1, -\eta_n'(\tau_n'\tau_n)^{-1/2})'$$

$$N_n = \begin{bmatrix} 1 & 0 \\ (\tau_n'\tau_n)^{-1/2}\eta_n & I_{m_\alpha} \end{bmatrix}$$

$$L_n = \begin{bmatrix} r'_{u,n}M_{\tau_n}r_{u,n} & 0 \\ 0 & \tau_n'\tau_n \end{bmatrix} .$$

Lemma 7. *Let* $\phi_n = (\alpha_n, \Pi_{\alpha,n}, \Pi_{\beta,n}, F_n)$ *be a sequence of null data generating processes in* Φ *and* j_n *a subsequence of* n*. Further, define* $\Theta(n) = (\int_c^1 D_\xi(a)D'_\xi(a)\mathrm{d}a)^{1/2}K_n\Pi_{\alpha,n}\Omega_{v_\alpha v_\alpha \cdot u}^{-1/2}$ *with* $K_n = \mathrm{diag}(n, n^2, n^3)$ *and a singular value decomposition* $\Theta(n) = O_{1,n}\mathcal{D}_n O'_{2,n}$ *where* $O_{1,n}$ *and* $O_{2,n}$ *are* 3×3 *and* $m_\alpha \times m_\alpha$ *dimensional orthonormal matrices and* D_n *is a* $3 \times m_\alpha$ *rectangular diagonal matrix with non-negative elements. Now let* $\Theta(j_n) = O_{1,j_n}\mathcal{D}_{j_n}O'_{2,j_n}$ *and assume that conditional on* D_ξ *we have* $O_{1,j_n} \to O_1$ *and* $O_{2,j_n} \to O_2$ *for orthonormal* O_1 *and* O_2*, and* $\mathcal{D}_{j_n} \to \mathcal{D}$ *for a rectangular diagonal matrix with possibly infinite diagonal elements.*

Then under ϕ_n we have when $H/n \to c \in (0,1)$ as $n \to \infty$ that

$$\rho_{j_n} - (1 + p_{j_n}) = o_p(1)$$

for some sequence of random variables p_{j_n} that satisfy $p_{j_n} \geq 0$ with probability 1 and

$$r'_{u,j_n} M_{\tau_{j_n}} r_{u,j_n} \xrightarrow{d} \chi^2(m_\beta) \ .$$

3.3 Main proof

With Lemmas 1-7 in place we are ready to prove Theorem 1.

In particular to prove the convergence of the AR_a statistic we use that Lemma 3 implies that

$$K_n^{-1} \sum_{t=H+1}^{n} z_t u_t \Rightarrow \Xi_u$$

where $\Xi_u|_{D_\xi} \sim N\left(0, \omega_u^2 \int_c^1 D_\xi(a) D_\xi(a)' \mathrm{da}\right)$. Lemma 4 and the continuous mapping theorem imply

$$\left(K_n^{-1} \sum_{t=H+1}^{n} z_t z_t' K_n^{-1}\right)^{-1} \Rightarrow \left(\int_c^1 D_\xi(a) D_\xi(a)' \mathrm{da}\right)^{-1} .$$

And Lemma 5 shows that $\hat{s}_u^2 \xrightarrow{p} \omega_u^2$. Combining the results gives

$$\begin{aligned} AR_s[\delta_0] &= \left(K_n^{-1} \sum_{t=H+1}^{n} z_t u_t\right)' \left(K_n^{-1} \sum_{t=H+1}^{n} z_t z_t' K_n^{-1}\right)^{-1} \left(K_n^{-1} \sum_{t=H+1}^{n} z_t u_t\right) / \hat{s}_u^2 \\ &\Rightarrow \Xi_u' \left(\int_c^1 D_\xi(a) D_\xi(a)' \mathrm{da}\right)^{-1} \Xi_u / \omega_u^2 \end{aligned}$$

Conditional on D_ξ we have

$$\Xi_u' \left(\int_c^1 D_\xi(a) D_\xi(a)' \mathrm{da}\right)^{-1} \Xi_u / \omega_u^2 |_{D_\xi} \sim \chi^2(3)$$

which implies that the unconditional distribution is also $\chi^2(3)$.

Next, for the subset statistic we follow closely the proof of Guggenberger et al. (2012). The main ingredients are Lemmas 6 and 7. By Lemma 6 we have wpa1

$$AR_{a,s}[\beta_0] \le \frac{r'_{u,n} M_{\tau_n} r_{u,n}}{\rho_n} .$$

Now, there exists a worst case sequence $\phi_n \in \Phi$ of null data generating processes such that

$$\begin{aligned} \text{AsySz}_{AR_{a,s}} &= \limsup_{n\to\infty, H/n\to c\in(0,1)} \sup_{\phi\in\Phi} \mathbb{P}_{\phi}(AR_{a,s}[\beta_0] > \chi^2_{1-\alpha}(m_\beta)) \\ &= \limsup_{n\to\infty, H/n\to c\in(0,1)} \mathbb{P}_{\phi_n}(AR_{a,s}[\beta_0] > \chi^2_{1-\alpha}(m_\beta)) \\ &\le \limsup_{n\to\infty, H/n\to c\in(0,1)} \mathbb{P}_{\phi_n}\left(\frac{r'_{u,n} M_{\tau_n} r_{u,n}}{\rho_n} > \chi^2_{1-\alpha}(m_\beta)\right) , \end{aligned}$$

where the first equality holds by the definition of $\text{AsySz}_{AR_{a,s}}$, the second by the choice of sequence ϕ_n and the inequality holds by the bound from Lemma 6. Furthermore, one can always find a subsequence j_n of n such that, conditional on D_ξ, along ϕ_{j_n} we have $O_{1,j_n} \to O_1$ and $O_{2,j_n} \to O_2$ for orthonormal O_1 and O_2 and $\mathcal{D}_{j_n} \to \mathcal{D}$ for a rectangular diagonal matrix with possibly infinite diagonal elements (where $O_{1,j_n}\mathcal{D}_{j_n}O'_{2,j_n}$ is the singular value decomposition of the matrix $\Theta(j_n)$ defined in Lemma 7.). Further, the subsequence can be always chosen to satisfy

$$\begin{aligned} &\limsup_{n\to\infty, H/n\to c\in(0,1)} \mathbb{P}_{\phi_n}\left(\frac{r'_{u,n} M_{\tau_n} r_{u,n}}{\rho_n} > \chi^2_{1-\alpha}(m_\beta)\right) = \\ &\limsup_{n\to\infty, H/n\to c\in(0,1)} \mathbb{P}_{\phi_{j_n}}\left(\frac{r'_{u,j_n} M_{\tau_{j_n}} r_{u,j_n}}{\rho_{j_n}} > \chi^2_{1-\alpha}(m_\beta)\right) \end{aligned}$$

But under any sequence of null data generating processes $\phi_n \in \Phi$ and under any subsequence j_n of n such that $O_{1,j_n} \to O_1$, $O_{2,j_n} \to O_2$ and $\mathcal{D}_{j_n} \to \mathcal{D}$ (conditional on D_ξ) under ϕ_n, we have by Lemma 7

$$\frac{r'_{u,j_n} M_{\tau_{j_n}} r_{u,j_n}}{\rho_{j_n}} \le r'_{u,j_n} M_{\tau_{j_n}} r_{u,j_n} + o_p(1) \xrightarrow{d} \chi^2(m_\beta)$$

This implies that, also unconditionally, $\text{AsySz}_{AR_{a,s}} \leq \alpha$.

4 Alternative approaches for constructing instruments

In the main paper we consider a polynomial function to approximate $\text{E}(u_t|\xi_t^i, \xi_{t-1}^i, \ldots)$. In particular, we have

$$\text{E}(u_t|\xi_t^i, \xi_{t-1}^i, \ldots) \approx \theta_{0,a}\xi_t^i + \theta_{1,a}\xi_{t-1}^i + \ldots + \theta_{H,a}\xi_{t-H}^i$$

where the coefficients are restricted by the Almon polynomial $\theta_{h,a} = a + bh + ch^2$, with a, b, c unknown coefficients. The polynomial approximation was chosen as it reduces the number of effective instruments to 3 and mimics the shape of impulse response functions that are typically found in macroeconomics.

Naturally, different approaches can be considered to reduce the number of structural shock instruments $\xi_{t:t-H}^i$ and recent work by Eberly, Stock and Wright (2019) explores an alternative approach based on exponential weighted moving average (EWMA) methods. In particular, they construct instruments

$$z_{b_s,t}^i = b_s\xi_t^i + (1 - b_s)z_{k,t-1}^i = \sum_{j=0}^{t-1} b_s(1 - b_s)^j \xi_{t-j}^i \ , \qquad b_s \in (0, 1) \ ,$$

where b_s is a smoothing parameter. Different choices of $b_s \in (0, 1)$ give the different instruments. For instance if we use three EWMA type instruments, with $b_{s,1}$, $b_{s,2}$ and $b_{s,3}$ as smoothing parameters, we have the approximation

$$\text{E}(u_t|\xi_t^i, \xi_{t-1}^i, \ldots) \approx \theta_{0,e}\xi_t^i + \theta_{1,e}\xi_{t-1}^i + \ldots + \theta_{t-1,e}\xi_1^i$$

where $\theta_{j,e} = a_{e,1}b_{s,1}(1 - b_{s,1})^j + a_{e,2}b_{s,2}(1 - b_{s,2})^j + a_{e,3}b_{s,3}(1 - b_{s,3})^j$ and we summarize the unknown coefficients $a_{e,1}, a_{e,2}$ and $a_{e,3}$ in the vector $\theta_E = (a_{e,1}, a_{e,2}, a_{e,3})'$.

Importantly, the EWMA requires selecting the smoothing parameters b_s. These parameters need to be fixed a priori and cannot be estimated as this form of pre-testing would invalidate the standard limiting distribution of the AR statistic. For instance, Eberly, Stock and Wright (2019) use $b_{s,1} = 0.9$ and $b_{s,2} = 0.7$ to construct two instruments for each sequence of structural shocks. In our setting we require at least three instruments and thus consider $b_{s,1} = 0.9$, $b_{s,2} = 0.7$ and $b_{s,3} = 0.5$. We summarize the instruments in the vector $z_t^{i,E} = (z_{0.5,t}^i, z_{0.7,t}^i, z_{0.9,t}^i)'$.

With these instruments we can compute the AR type statistic

$$AR_E[\delta_0] = \hat{\theta}_E' \widehat{\Sigma}_E^{-1} \hat{\theta}_E \qquad \hat{\theta}_E = \left(\sum_{t=1}^n z_t^{i,E} z_t^{i,E'} \right)^{-1} \sum_{t=1}^n z_t^{i,E} (y_t - w_t' \delta_0)$$

where $\widehat{\Sigma}_E$ can be any consistent estimate for the variance of $\frac{1}{n} \sum_{t=1}^n z_t^{i,E}(y_t - w_t'\delta_0)$. Importantly, the instruments of the EWMA method are stationary by construction (depending the selection of the smoothing parameters) and therefore the $AR_E[\delta_0]$ statistic takes the usual Wald form.

The polynomial and EWMA methods are theoretically hard to distinguish as the comparison depends on true function $\mathrm{E}(u_t|\xi_t^i, \xi_{t-1}^i, \ldots)$. Instead we compared them in a simulation study, where the simulation design is the standard macro model that is discussed in Appendix D of the main paper. To keep the comparison based on the instruments, we only consider designs with serial uncorrelated errors and use the same variance estimate (e.g. $\hat{\sigma}_u^2 (Z'Z)^{-1}$ with $\hat{\sigma}_u^2$ the estimate for the variance of u_t) for both AR tests. This avoids that the conclusions depend on the quality of the variance estimates.[3]

We summarize our findings in Figures 2 and 3 where we show the power curves for the polynomial and EWMA approaches. We vary either γ_f (Figure 2) or λ (Figure 3) around its true value and plot the empirical rejection frequency. We show the plots for the different sample sizes separately and within each plot we show the power curves for different instrument strengths.

[3] More specifically, the AR_E statistic requires the estimate $\widehat{\Sigma}_E$ whereas the AR_a statistic only requires $\hat{s}_u^2$. The latter can be computed more accurately for small sample sizes such as $n = 200, 500$ leading to better finite sample behavior in terms of size of the corresponding test statistics.

We find that both methods control size well as at the true parameter value the empirical rejection frequency is close to 0.05. When deviating from the true parameter the power increases, albeit slowly for scenarios with weak instruments $\sigma_i = 0.1$. When comparing the EWMA and polynomial methods we find little differences, the power curves are close and neither method dominates the other.

In summary, for the simulation design that matches our empirical study we find little differences between the two methods for constructing the instruments. That said for different data generating processes this might change.

5 Additional simulation results

In this section we investigate the finite sample properties of the AR_a and $AR_{a,s}$ statistics in more detail. In particular, we consider designs with heteroskedastic and serially correlated errors. The data generating process considered mimics the solution of the structural model that is considered in the main text, see Kleibergen and Mavroeidis (2009). The benefit of working with the solved model directly is that it becomes easier to add additional nonlinear features, such as garch effects and other sources of heteroskedasticity. Overall the data generating process of this section has all features that are commonly observed in aggregate macro time series: persistence, serial correlated errors and heteroskedasticity.

5.1 Simulation design

We consider the model

$$\begin{aligned} y_t &= a_{y,1}y_{t-1} + a_{y,2}y_{t-2} + \lambda x_t + u_t \\ x_t &= a_{x,1}x_{t-1} + a_{x,2}x_{t-2} + \varepsilon_t + u_t \end{aligned}$$

where the a coefficients capture the dynamics of the observed series y_t and x_t. There is an endogeneity problem as x_t depends on u_t and we will use the structural shocks ε_t as instruments to conduct inference on the structural parameters $a_{y,1}, a_{y,2}, \lambda$. The structural shocks ε_t are simulated from the ar(1) process

$$\varepsilon_t = a_\varepsilon \varepsilon_{t-1} + \nu_{\varepsilon,t} \qquad \nu_{\varepsilon,t} \sim N(0, \sigma^2)$$

where a_ε is the autoregressive parameter and σ captures the strength of the instruments. This specification mimics the findings in Alloza, Gonzalo and Sanx (2019) who find that commonly used structural shock proxies are not uncorrelated over time. We generate the disturbances $\{u_t\}$ from the ar(1)-garch(1,1) model

$$\begin{aligned} u_t &= a_u u_{t-1} + \nu_{u,t} \\ \nu_{u,t} &= \sigma_{u,t} e_{u,t} \\ \sigma^2_{u,t} &= d_{u,0} + d_{u,1}\varepsilon_t^2 + d_{u,2}\sigma^2_{u,t-1} + d_{u,3}\nu^2_{u,t-1} \end{aligned}$$

where apart from serial correlation two sources of heterogeneity have been added. First, the variance of the disturbances may depend on the structural shock ε_t. This is a source of heterogeneity that is often considered in cross-sectional studies with heteroskedastic errors, see Hausman et al. (2012). Second, we include garch effects that allow the variance of the disturbances to change smoothly over time. Note that in all cases the lead-lag exogeneity assumption for the instruments continuous to hold.

We consider different choices for the parameter values. First, we fix the model parameter $a_{y,1} = 1.3$, $a_{y,2} = -0.5$, $\lambda = 1$ and $a_{x,1} = 1.3$, $a_{x,2} = -0.5$. These values are typically observed for macro time series and changing them does not alter any of our findings. Second, similar as in our main simulation study we change the strength of the instruments by choosing $\sigma = 0.1, 0.25, 0.5, 1$. Finally, the parameters for generating the disturbances u_t are chosen as:

(i) **no heteroskedasticity, no serial correlation**: $d_{u,0} = 1$, $d_{u,1} = d_{u,2} = d_{u,3} = 0$ and $a_\varepsilon = a_u = 0$

(ii) **heteroskedasticity from ϵ_t, no serial correlation**: $d_{u,0} =$, $d_{u,1} = 1$, $d_{u,2} = d_{u,3} = 0$

and $a_\varepsilon = a_u = 0$

(iii) **heteroskedasticity from garch, no serial correlation**: $d_{u,0} = 0.05$, $d_{u,1} = 0$, $d_{u,2} = 0.9$, $d_{u,3} = 0.05$ and $a_\varepsilon = a_u = 0$

(iv) **no heteroskedasticity, serial correlated errors**: $d_{u,0} = 1$, $d_{u,1} = d_{u,2} = d_{u,3} = 0$ and $a_\varepsilon = a_u = 0.5$

(v) **heteroskedasticity from ϵ_t, serial correlated errors**: $d_{u,0} =$, $d_{u,1} = 1$, $d_{u,2} = d_{u,3} = 0$ and $a_\varepsilon = a_u = 0.5$

(vi) **heteroskedasticity from garch, serial correlated errors**: $d_{u,0} = 0.05$, $d_{u,1} = 0$, $d_{u,2} = 0.9$, $d_{u,3} = 0.05$ and $a_\varepsilon = a_u = 0.5$.

Note that we only consider mild forms of serial correlation as this would be typically observed in the residuals of structural equations, e.g. Zhang and Clovis (2010).

To test the structural parameters we rely on the Almon restricted Anderson-Rubin AR_a statistic and its subset counterpart $AR_{a,s}$. We consider implementations with $H = 5, 10, 20, 40, 80$ structural shocks as instruments. Further we consider sample sizes $n = 200, 500$.

For each particular combination of parameters, number of structural shocks and sample sizes, we generate 5000 datasets and compute the AR_a test statistic to test $H_0 : a_{y,1} = a^0_{y,1},\ a_{y,2} = a^0_{y,2},\ \lambda = \lambda^0$ and the subset $AR_{a,s}$ statistic to test $H_0 : \lambda = \lambda^0$.

5.2 Simulation results

In Tables 1-3 we show the empirical rejection frequencies for all tests. We find the following patterns.

- For error processes (i)-(iii) we find that the AR_a statistic always has correct size. This holds regardless of the form of heteroskedasticity and for all combinations of n and H. For the subset statistic we find similar patterns as in the main paper as the rejection frequencies are conservative for weak instrument settings, created either by low σ and/or high H. With stronger instruments the rejection frequencies are very close to the nominal size $\alpha = 0.05$.

- For serial correlated error processes (iv)-(vi) the picture changes a bit as now for small H (e.g. for $H = 5, 10$) the AR_a statistic is undersized. This is not surprising as without the high persistence in the instruments the AR_a statistic is inefficient as the standard serial correlation adjusted AR statistic should be used, see Andrews, Stock and Sun (2019). When H is large the empirical rejection frequencies are close to 0.05 again. For the subset statistic we find a similar pattern as now for small H the statistic has very low power for all choices of instruments.

6 Additional empirical results

In this section we discuss additional empirical results.

- In Table 4 we show the parameter estimates and $AR_{a,s}$ based confidence sets for the parameters of the US Phillips curve (1969-2007) based on the Romer and Romer (2004) shocks when we vary H, the number of lags used to construct the Almon-restricted instruments. The forcing variable is the unemployment gap, but we note that similar results can be obtained for the output gap.

  We find that our results are robust to different choices for H. In particular, for all reasonable choices of H the point estimates are the same and, as long as H is sufficiently large, e.g. $H > 10$, the confidence sets are comparable. Notably, the confidence sets for the unemployment gap always exclude zero and include sizable negative values for λ. When $H = 10$ the confidence sets are infinite for the unrestricted Phillips curve specification, but closed and very similar to the other estimates for the restricted specification $\gamma_b + \gamma_f = 1$. Intuitively, when we use $H = 10$, we exclude a substantial part of the impulse response of inflation that is non-zero and could have provided relevant identifying information, see the impulse responses in Romer and Romer (2004) and Barnichon and Mesters (2019).

- In Table 5 we show sub-sample results for the parameters of the US Phillips curve based on the Romer and Romer (2004) shocks. We consider the two sampling periods: 1969-1989 and 1990-2007. We find that pre-1990 the coefficient on the unemployment gap was considerably larger (in absolute value) when compared to the full (1969-2007) sampling period. In contrast the coefficient on inflation expectations is smaller and no longer significantly different from zero during the pre-1990 period. For the post-1990 sampling period the information in the Romer and Romer (2004) shocks is insufficient to find closed confidence bounds. The un-informativeness of the Romer & Romer shocks post 1990 is further discussed in Ramey (2016).

- In the main paper we showed that the traditional Generalized Instrumental Variables (GIV) approach —using lagged macro variables as instruments—, leads to point estimates that are considerably smaller (in absolute value) for the forcing variable. In Table 6 we complement this analysis by showing a more detailed set of results for the traditional approach. In particular, we estimated the coefficients of the US Phillips curve with either the unemployment gap or the output gap as forcing variable using GIV with 4 lags of inflation and 4 lags of the forcing variables as instruments. The 90% confidence intervals are standard, e.g. $[\hat{\delta}_j \pm 1.64 se(\hat{\delta}_j)]$, and thus based on a *strong* IV assumption. These intervals are to be regarded as indicative. We also computed the more correct projection based confidence bands and the subset confidence bounds of Kleibergen and Mavroeidis (2009) (which are only provably valid under conditional homoskedasticity), but both gave infinite confidence intervals for all parameters.

  Even under a strong IV assumption the confidence intervals imply that the traditional GIV approach is generally uninformative about the coefficients on inflation expectations and the forcing variable, as the confidence intervals do not exclude zero (except for the restricted model with the unemployment gap as forcing variable). The only significant coefficient is for lagged inflation, where we can reject that the coefficient is equal to zero (again under a strong IV assumption).

Summarizing, our findings for the traditional approach are very similar to Mavroeidis, Plagborg-Møller and Stock (2014) who find large sampling uncertainty for the Phillips curve estimates from the GIV approach. Figure 4 in the main paper and Table 6 in this appendix are in line with their findings.

- As we saw in section 5 of the main text, Romer and Romer (2004) identify monetary shocks holding constant the staff's Greenbook forecasts for output and inflation, but one concern is that policy makers respond to information beyond what is in the Greenbook. If this response is in reaction to cost-push factors, the exogeneity condition could be violated for the R&R shocks. To get at this possible issue, we regress the R&R shocks on lagged common factors that are obtained from a large panel of macro variables.[4] The residuals of this regression are then considered a cleaner version of the R&R series. Note however that in doing so we might be removing useful variation that is unrelated to supply factors (Cochrane, 2004).

  In Figure 1 we show the point estimates and confidence regions that were obtained using this cleaned instrument series. The estimates are computed exactly as in Figure 1 of the main paper. The confidence sets are similar (albeit slightly larger), and the point estimates are broadly consistent, if anything pointing to a slightly larger Phillips curve slope (in absolute value) and a smaller coefficient on expected future inflation.

[4]In particular, we consider the panel from Stock and Watson (2012) ($N = 144$) and estimate the number of common factors using the IC2 criteria from Bai and Ng (2002). The criteria indicates that there are 2 common factors for the 1969-2007 sampling period. These factors are used in the predictive regression.

References

Alloza, Mario, Jesus Gonzalo, and Carlos Sanx. 2019. "Dynamic Effects of Persistent Shocks." working paper.

Andrews, Donald W. K. 1991. "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation." *Econometrica*, 59(3): 817–858.

Andrews, Isaiah, James Stock, and Liyang Sun. 2019. "Weak Instruments in IV Regression: Theory and Practice." *Annual Review of Economics.* forthcoming.

Bai, Jushan, and Serena Ng. 2002. "Determining the Number of Factors in Approximate Factor Models." *Econometrica*, 70(1): 191–221.

Barnichon, Regis, and Geert Mesters. 2019. "The Phillips Multiplier." Working Paper.

Cochrane, John. 2004. "Comments on "A new measure of monetary shocks: Derivation and implications" by Romer and Romer." NBER EFG Metting.

Davidson, James. 1994. *Stochastic Limit Theory.* Oxford University Press.

de Jong, Robert M., and James Davidson. 2000*a*. "Consistency of Kernel Estimators of Heteroscedastic and Autocorrelated Covariance Matrices." *Econometrica*, 68(2): 407–423.

de Jong, Robert M., and James Davidson. 2000*b*. "The Functional Central Limit Theorem and Weak Convergence to Stochastic Integrals I: Weakly Dependent Processes." *Econometric Theory*, 16(5): 621?642.

Eberly, Janice C., James H. Stock, and Jonathan H. Wright. 2019. "The Federal Reserve's Current Framework for Monetary Policy: A Review and Assessment." Working paper.

Gallant, A. Ronald, and Halbert L. Jr. White. 1987. *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models.* Basil Blackwell Ltd., Oxford.

Guggenberger, Patrik, Frank Kleibergen, Sophocles Mavroeidis, and Linchun Chen. 2012. "On the Asymptotic Sizes of Subset Anderson-Rubin and Lagrange Multiplier Tests in Linear Instrumental Variables Regression." *Econometrica*, 80(6): 2649–2666.

Hausman, J. A., W. K. Newey, T. Woutersen, J. C. Chao, and N. R. Swanson. 2012. "Instrumental variable estimation with heteroskedasticity and many instruments." *Quantitative Economics*, 3: 211–255.

Kleibergen, F., and S. Mavroeidis. 2009. "Weak Instrument Robust Tests in GMM and the New Keynesian Phillips Curve." *Journal of Business and Economic Statistics*, 27: 293–311.

Mavroeidis, Sophocles, Mikkel Plagborg-Møller, and James H. Stock. 2014. "Empirical Evidence on Inflation Expectations in the New Keynesian Phillips Curve." *Journal of Economic Literature*, 52: 124–188.

Park, Joon Y., and Peter C. B. Phillips. 1988. "Statistical Inference in Regressions with Integrated Processes: Part 1." *Econometric Theory*, 4(3): 468–497.

Ramey, Valerie. 2016. "Macroeconomic Shocks and Their Propagation." In *Handbook of Macroeconomics.* , ed. J. B. Taylor and H. Uhlig. Amsterdam, North Holland:Elsevier.

Romer, Christina D., and David H. Romer. 2004. "A New Measure of Monetary Shocks: Derivation and Implications." *American Economic Review*, 94: 1055–1084.

Stock, James H., and Mark W. Watson. 2012. "Disentangling the Channels of the 2007-09 Recession." *Brookings Papers on Economic Activity*, 81.

Zhang, Chengsi, and Joel Clovis. 2010. "The New Keynesian Phillips Curve of rational expectations: A serial correlation extension." *Journal of Applied Economics*, 13(1): 159–179.

Proofs of Lemmas 1-7

Before providing the proofs of the lemmas we restate three theorems from de Jong and Davidson (2000*b*).

Theorem 3.1 of de Jong and Davidson (2000*b*). *Let $\mathfrak{K}_n$ denote an integer valued increasing sequence and $\{X_{n,t}, n = 1, 2, \dots, t = 1, 2, \dots\}$ a triangular array of random variables that satisfy*

1. $\mathrm{E}(X_{n,t}) = 0$ *and* $\|\sum_{t=1}^{\mathfrak{K}_n} X_{n,t}\|_2 = 1$
2. *There exists a positive constant array $c_{n,t}$ such that $\{X_{n,t}/c_{n,t}\}$ is L_r bounded for $r > 2$ uniformly in n, t*
3. *$X_{n,t}$ is L_2-NED of size $-\frac{1}{2}$ on $V_{n,t}$, where $V_{n,t}$ is either an α-mixing array of size $-r/(r-2)$ or a ϕ-mixing array of size $-r/(2(r-1))$, and $d_{n,t}/c_{n,t}$ is uniformly bounded in n, t*
4. *For some sequence $\mathfrak{b}_n$, such that $\mathfrak{b}_n = o(\mathfrak{K}_n)$ and $\mathfrak{b}_n^{-1} = o(1)$, letting $\mathfrak{r}_n = \lfloor \mathfrak{K}_n/\mathfrak{b}_n \rfloor$, $\mathfrak{M}_{n,i} = \max_{(i-1)\mathfrak{b}_n \le t \le i\mathfrak{b}_n} c_{n,t}$, $\mathfrak{M}_{n,\mathfrak{r}_n+1} = \max_{\mathfrak{r}_n\mathfrak{b}_n+1 \le t \le \mathfrak{K}_n} c_{n,t}$,*

$$\max_{1 \le i \le \mathfrak{r}_n+1} \mathfrak{M}_{n,i} = o(\mathfrak{b}_n^{-1/2}) \qquad \sum_{i=1}^{\mathfrak{r}_n} \mathfrak{M}_{n,i}^2 = O(\mathfrak{b}_n^{-1})$$

Let $X_n(a) = \sum_{t=1}^{\mathfrak{K}_n(a)} X_{n,t}$ for $a \in [0,1]$ where $\{\mathfrak{K}_n(a), n \ge 1\}$ is a sequence of integer valued, right continuous, increasing functions of a, with $\mathfrak{K}_n(0) = 0$ for all $n \ge 1$, $\mathfrak{K}_n(a)$ is non-decreasing in n for all $a \in [0,1]$ and $\mathfrak{K}_n(a) - \mathfrak{K}_n(a') \to \infty$ as $n \to \infty$ if $a > a'$. Further, assume that

5. $\eta(a) = \lim_{n\to\infty} \mathrm{E}(X_n(a)^2)$ *exists for all* $a \in [0,1]$

6. $\lim_{\epsilon\to 0} \sup_{a\in[0,1-\epsilon]} \limsup_{n\to\infty} \sum_{t=\mathfrak{K}_n(a)}^{\mathfrak{K}_n(a+\epsilon)} c_{n,t}^2 = 0$

Then

$$X_n(a) \Rightarrow X(a)$$

where $X(a)$ is a Gaussian process having a.s. continuous sample paths and independent increments.

Based on this scalar result the following multivariate result from de Jong and Davidson (2000*b*) is useful for our purposes.

Theorem 3.2 of de Jong and Davidson (2000*b*). *Let $X_{n,t}$ be an k-vector valued array and assume that for any $k \times 1$ fixed vector λ, with $\lambda'\lambda = 1$ we have that there exists an array $c_{n,t}$ such that $\lambda' X_{n,t}$ satisfies the assumptions of Theorem 3.1 of de Jong and Davidson (2000*b*) for the same functions $\mathfrak{K}_n(a)$. Then*

$$X_n(a) \Rightarrow X(a)$$

where X is an k-dimensional Gaussian process having a.s. continuous sample paths and independent increments.

Finally, the following result for the convergence to stochastic integrals is used below.

Theorem 4.1 of de Jong and Davidson (2000*b*). *Let the conditions of Theorem 3.2 of de Jong and Davidson (2000*b*) hold for $X_{n,t} = (X_{n,t}^{1'}, X_{n,t}^{2'})'$ and $\mathfrak{K}_n(a) = \lfloor na \rfloor + 1$. Then*

$$\left(X_n^1(a), X_n^2(a), \left(\sum_{t=1}^{n-1} \sum_{s=1}^{t} X_{n,s}^1 X_{n,t}^{2'} - \Lambda_n^{12} \right) \right) \Rightarrow \left(X^1(a), X^2(a), \int_0^1 X^1(a) \mathrm{d}X^2(a)' \right)$$

where $X^1(a)$ and $X^2(a)$ are a.s. continuous Gaussian processes having independent increments and

$$\Lambda_n^{12} = \sum_{t=1}^{n} \sum_{s=t+1}^{n} \mathrm{E}\left(X_{n,t}^1 X_{n,s}^{2'} \right)$$

We use these results to prove Lemmas 1-7.

Proof of Lemma 1. For simplicity we drop the dependence on p and q from all subscripts and let $X_{n,t} = \Sigma_n^{-1/2} D_{n,t}\eta_t$, where $D_{n,t} = \Psi_n^{-1} T_t$, $\Sigma_n = \mathrm{Var}(\sum_{t=1}^n D_{n,t}\eta_t)$ and $\Sigma_n^{1/2}$ is such that $\Sigma_n = \Sigma_n^{1/2}\Sigma_n^{1/2'}$. Note that assumptions 1.1-(ii), 1.1-(iii) and 1.4 imply that $\Sigma_n^{-1/2}$ exists for n sufficiently large. We verify the conditions of the FCLT in Theorem 3.1 of de Jong and Davidson (2000*b*) for $\lambda' X_{n,t}$ with $\lambda'\lambda = 1$. We take $d_{n,t}^\lambda = (\lambda' \Sigma_n^{-1/2} D_{n,t}^2 \Sigma_n^{-1/2'} \lambda)^{1/2}$ and $c_{n,t}^\lambda = d_{n,t}^\lambda \max(1, \|\eta_t\|_r)$. Finally, we take $\mathfrak{K}_n(a) = \lfloor na \rfloor$ and define $X_n(a) = \sum_{t=1}^{\lfloor na \rfloor} X_{n,t}$.

1. $\mathrm{E}(\lambda' X_{n,t}) = 0$ follows as

$$\mathrm{E}(\lambda' X_{n,t}) = \lambda' \Sigma_n^{-1/2} D_{n,t} \mathrm{E}(\eta_t) = 0$$

by assumption 1.1.(i). Next, we take $\mathfrak{K}_n = n$ and show $\| \sum_{t=1}^n \lambda' X_{n,t} \|_2 = 1$.

$$
\begin{aligned}
\left\| \sum_{t=1}^n \lambda' X_{n,t} \right\|_2^2 &= \mathrm{E}\left(\left(\lambda' \Sigma_n^{-1/2} \sum_{t=1}^n D_{n,t} \eta_t \right)^2 \right) \\
&= \lambda' \Sigma_n^{-1/2} \mathrm{Var}\left(\sum_{t=1}^n D_{n,t} \eta_t \right) \Sigma_n^{-1/2'} \lambda \\
&= \lambda' \Sigma_n^{-1/2} \Sigma_n \Sigma_n^{-1/2'} \lambda = 1
\end{aligned}
$$

which follows from the definition of Σ_n.

2. We show that $\sup_{n,t} \| \lambda' X_{n,t} / c_{n,t}^{\lambda} \|_r < \infty$. Note that $\lambda' X_{n,t} / c_{n,t}^{\lambda} = \lambda' \Sigma_n^{-1/2} D_{n,t} \eta_t / c_{n,t}^{\lambda}$ and the elements of the vector $|\lambda' \Sigma_n^{-1/2} D_{n,t}| / (\lambda' \Sigma_n^{-1/2} D_{n,t}^2 \Sigma_n^{-1/2'} \lambda)^{1/2}$ are in $[0, 1]$.[5] Thus

$$
\begin{aligned}
\sup_{n,t} \| \lambda' X_{n,t} / c_{n,t}^{\lambda} \|_r &= \sup_{n,t} \left\| \max(1, \|\eta_t\|_r)^{-1} \sum_{i=1}^5 \frac{(\lambda' \Sigma_n^{-1/2} D_{n,t})_i}{(\lambda' \Sigma_n^{-1/2} D_{n,t}^2 \Sigma_n^{-1/2'} \lambda)^{1/2}} \eta_{i,t} \right\|_r \\
&\leq \sup_{n,t} \left\| \max(1, \|\eta_t\|_r)^{-1} \sum_{i=1}^5 \left| \frac{(\lambda' \Sigma_n^{-1/2} D_{n,t})_i}{(\lambda' \Sigma_n^{-1/2} D_{n,t}^2 \Sigma_n^{-1/2'} \lambda)^{1/2}} \right| |\eta_{i,t}| \right\|_r \\
&\leq \sup_{t} \left\| \max(1, \|\eta_t\|_r)^{-1} \sum_{i=1}^5 |\eta_{i,t}| \right\|_r < \infty
\end{aligned}
$$

which follow from assumption 1.2, e.g. $\sup_t \|\eta_t\|_{2r} \leq \Delta < \infty$.

3. Note that, by Assumption 1.3, $\lambda' X_{n,t}$ is a linear combination, with bounded weights, of L_2-NED sequences, which is thus also L_2-NED (e.g. Davidson, 1994, 17.12 Theorem) and the size of $-(r-1)/(r-2) < -\frac{1}{2}$ for $r > 2$, is retained. Further, note that the constants are bounded as

$$
\max_{1 \leq t \leq n} d_{n,t}^{\lambda} = \max_{1 \leq t \leq n} (\lambda' \Sigma_n^{-1/2} D_{n,t}^2 \Sigma_n^{-1/2'} \lambda)^{1/2} \leq \|\lambda' \Sigma_n^{-1/2}\| \max_{1 \leq t \leq n} \|D_{n,t}\| = O(n^{-1/2}) \ ,
$$

which follows as $\max_{1 \leq t \leq n} \|D_{n,t}\| = n^{-1/2}$ and $\|\lambda' \Sigma_n^{-1/2}\| = O(1)$ as

$$
\Sigma_n = \Psi_n^{-1} \mathrm{Var}\left(\sum_{t=1}^n T_t \eta_t \right) \Psi_n^{-1} = \begin{bmatrix} \omega_\xi^2 & 0 \\ 0 & \Omega_{uv} \end{bmatrix} + \begin{bmatrix} o(1) & 0 \\ 0 & o(1) I_4 \end{bmatrix} . \tag{6}
$$

by Assumptions 1.1-(ii), 1.1-(iii) and 1.4. Hence $\lambda' X_{n,t}$ is L_2-NED of size $-\frac{1}{2}$. Finally, note that $\sup_{n,t} d_{n,t} / c_{n,t} = \sup_t \max(1, \|\eta_t\|_r)^{-1} < \infty$.

4. We take $\mathfrak{b}_n = \lfloor n^{1/2} \rfloor$ and $\mathfrak{r}_n = \lfloor n / \mathfrak{b}_n \rfloor$ and verify the conditions. Note that

$$
\max_{1 \leq i \leq \mathfrak{r}_n + 1} \max_{(i-1)\mathfrak{b}_n \leq t \leq i \mathfrak{b}_n} c_{n,t}^{\lambda} \leq \Delta_1 \max_{1 \leq i \leq \mathfrak{r}_n + 1} \max_{(i-1)\mathfrak{b}_n \leq t \leq i \mathfrak{b}_n} \|D_{n,t}\| = c_1 n^{-1/2} = o(\mathfrak{b}_n^{-1/2})
$$

[5] for any vector $a \in \mathbb{R}_+^n$ we have $0 \leq a_i / \|a\| \leq 1$ for all $i = 1, \ldots, n$.

where Δ_1 is a constant such that $\sup_{n,t}\max(1,\|\eta_t\|_r)\|\lambda'\Sigma_n^{-1}\| \leq \Delta_1$, which exists due to assumptions 1.2 and 1.4. Further, we used $\max_{1\leq t\leq n}\|D_{n,t}\| = n^{-1/2}$ for all $n \geq 1$. For the second part we have that

$$\sum_{i=1}^{\mathfrak{r}_n} \max_{(i-1)\mathfrak{b}_n\leq t\leq i\mathfrak{b}_n} (c_{n,t}^{\lambda})^2 \leq \Delta_1^2 \sum_{i=1}^{\mathfrak{r}_n} \max_{(i-1)\mathfrak{b}_n\leq t\leq i\mathfrak{b}_n} \|D_{n,t}\|^2 = \Delta_1^2\mathfrak{r}_n n^{-1} = O(\mathfrak{b}_n^{-1}) \ .$$

5. Let $\lambda = (\lambda_1, \lambda_2')'$ where λ_1 is scalar and λ_2 is 4×1. Recall, that Assumptions 1.1-(ii), 1.1-(iii) and 1.4 imply

$$\text{Var}\left(\sum_{t=1}^{n} T_t\eta_t\right) = \begin{bmatrix} \omega_{\xi,n,0}^2 & 0 \\ 0 & \Omega_{uv,n,0} \end{bmatrix} = \begin{bmatrix} n^{2p+1}\omega_\xi^2 & 0 \\ 0 & n^{2q+1}\Omega_{uv} \end{bmatrix} + \begin{bmatrix} o(n^{2p+1}) & 0 \\ 0 & o(n^{2q+1})I_4 \end{bmatrix}$$

Note that using equation (6) it follows that

$$\lambda' X_n(a) = \lambda_1\omega_\xi^{-1}n^{-1/2-p}\sum_{t=1}^{\lfloor na\rfloor} t^p\xi_t + \lambda_2'\Omega_{uv}^{-1/2}n^{-1/2-q}\sum_{t=1}^{\lfloor na\rfloor} t^q(u_t,v_t')' + o(1) \ .$$

By assumption 1.1-(ii)-(iii) we have $\text{E}\left(\left(\sum_{t=1}^{\lfloor na\rfloor} t^p\xi_t\right)\left(\sum_{t=1}^{\lfloor na\rfloor} t^q(u_t,v_t')'\right)\right) = 0$. Now since $\text{Var}\left(\sum_{t=1}^{\lfloor na\rfloor} t^p\xi_t\right) = \lfloor na\rfloor^{2p+1}\omega_\xi^2 + o(\lfloor na\rfloor^{2p+1})$ we have that

$$\omega_\xi^{-2}n^{-1-2p}\text{Var}\left(\sum_{t=1}^{\lfloor na\rfloor} t^p\xi_t\right) \to a^{2p+1}$$

Similarly, since $\text{Var}\left(\sum_{t=1}^{\lfloor na\rfloor} t^q(u_t,v_t')'\right) = \lfloor na\rfloor^{2q+1}\Omega_{uv} + o(\lfloor na\rfloor^{2q+1})I_4$

$$\Omega_{uv}^{-1/2}n^{-1-2q}\text{Var}\left(\sum_{t=1}^{\lfloor na\rfloor} t^q(u_t,v_t')'\right)\Omega_{uv}^{-1/2'} \to a^{2q+1}I_4$$

Combining we have that

$$\text{Var}\left(\lambda' X_n(a)\right) \to \lambda_1^2 a^{2p+1} + \lambda_2'\lambda_2 a^{2q+1}$$

6. Finally we study $\lim_{\delta\to 0}\sup_{a\in[0,1-\delta]}\limsup_{n\to\infty}\sum_{t=\lfloor na\rfloor}^{\lfloor n(a+\delta)\rfloor}(c_{n,t}^{\lambda})^2$. Recall, from point 4 that $(c_{n,t}^{\lambda})^2 \leq \Delta_1^2\|D_{n,t}\|^2 = \Delta_1^2/n$ We have that

$$\lim_{\epsilon\to 0}\sup_{a\in[0,1-\epsilon]}\limsup_{n\to\infty}\sum_{t=\lfloor na\rfloor}^{\lfloor n(a+\epsilon)\rfloor}(c_{n,t}^{\lambda})^2 \leq$$

$$\lim_{\epsilon\to 0}\sup_{a\in[0,1-\epsilon]}\limsup_{n\to\infty}\frac{\Delta_1^2}{n}(\lfloor n(a+\epsilon)\rfloor - \lfloor na\rfloor) = 0$$

as $\lim_{\epsilon\to 0}\sup_{a\in[0,1-\epsilon]}\lfloor a+\epsilon\rfloor - \lfloor a\rfloor = 0$.

This completes the verification of the assumptions and we have shown that

$$\lambda' X_n(a) \Rightarrow \lambda' X(a)$$

where $\lambda' X(a)$ is a Gaussian process with variance $\lambda_1^2 a^{2p+1} + \lambda_2' \lambda_2 a^{2q+1}$. From Theorem 3.2 of de Jong and Davidson (2000*b*) it follows that

$$X_n(a) \Rightarrow X(a) \ ,$$

where $X(a)$ Gaussian process with variance $\text{diag}(a^{2p+1}, a^{2q+1} I_4)$. Finally, by definition we have that $X_n(a) = \Sigma_n^{-1/2}\Psi_n^{-1}\sum_{t=1}^{\lfloor na\rfloor} T_t\eta_t$, such that using (6) gives

$$\Psi_n^{-1}\sum_{t=1}^{\lfloor na\rfloor} T_t\eta_t \Rightarrow G(a) \ .$$

□

Proof of Lemma 2. Equation (2) can be decomposed as

$$\frac{1}{n^{1+p+q}}\sum_{t=H+1}^{n}\sum_{s=1}^{t} s^p\xi_s t^q y_t = \frac{1}{n^{1+p+q}}\sum_{t=H+1}^{n}\sum_{s=1}^{t-1} s^p\xi_s t^q y_t + \frac{1}{n}\sum_{t=H+1}^{n}(t/n)^{p+q}\xi_t y_t$$

First, we study the second term. Since $\{\xi_t\}$ and $\{y_t\}$ are L_2-NED of size $-(r-1)/(r-2)$ on $\{V_t\}$ and $\sup_t\|\xi_t\|_{2r} < \infty$ and $\sup_t\|y_t\|_{2r} < \infty$ $(r > 2)$, it follows that the sequence $\{\xi_t y_t\}$ is L_2-NED of size $-1/2$ on $\{V_t\}$ and $\|\xi_t y_t\|_2 \leq \Delta^2$, which follows from applying Corollary 4.3 (b) in Gallant and White (1987). Next, from the definition of NED processes

$$\|(t/n)^{p+q}\xi_t y_t - \text{E}((t/n)^{p+q}\xi_t y_t|\mathcal{F}_{t-m}^{t+m})\|_2 = (t/n)^{p+q}\|\xi_t y_t - \text{E}(\xi_t y_t|\mathcal{F}_{t-m}^{t+m})\|_2 < \nu_m \ ,$$

as $(t/n)^{p+q} \leq 1$ for all p,q,n,t, with $0 \leq t \leq n$, and $\nu_m = O(m^{-1/2-\varepsilon})$ with $\varepsilon > 0$. Hence, we may conclude that the sequence $\{(t/n)^p\xi_t y_t\}$ is L_2-NED of size $-\frac{1}{2}$ on $\{V_t\}$. Note further that $\text{E}((t/n)^{p+q}\xi_t y_t) = 0$ by Assumption 1.1-(ii) or (iii), such that Theorem 20.20 part (i) in Davidson (1994) implies that

$$\frac{1}{n}\sum_{t=1}^{n}(t/n)^{p+q}\xi_t y_t \overset{a.s.}{\to} 0 \ .$$

Next, for the first term define $X_{n,t}^{\xi,p} = n^{-1/2}(t/n)^p\xi_t$ and $X_{n,t}^{y,q} = n^{-1/2}(t/n)^q y_t$ and note that Lemma 1 implies that $\sum_{t=1}^{\lfloor na\rfloor} X_{n,t}^{\xi,p} \Rightarrow G_{\xi,p}(a)$ and $\sum_{t=1}^{\lfloor na\rfloor} X_{n,t}^{y,q} \Rightarrow G_{y,p}(a)$, and $\text{E}(X_{n,s}^{\xi,p}X_{n,t}^{y,q}) = 0$

for all t, s, n by Assumption 1.1-(ii) or (iii). It follows that

$$\begin{aligned}\frac{1}{n^{1+p+q}}\sum_{t=H+1}^{n}\sum_{s=1}^{t-1}s^p\xi_s t^q y_t &= \sum_{t=H}^{n-1}\sum_{s=1}^{t} X_{n.t}^{\xi,p}X_{n,t+1}^{y,q}\\ &= \sum_{t=1}^{n-1}\sum_{s=1}^{t} X_{n.t}^{\xi,p}X_{n,t+1}^{y,q} - \sum_{t=1}^{H-1}\sum_{s=1}^{t} X_{n.t}^{\xi,p}X_{n,t+1}^{y,q}\\ &\Rightarrow \int_0^1 G_{\xi,p}(a)\mathrm{d}G_{y,q}(a) - \int_0^c G_{\xi,p}(a)\mathrm{d}G_{y,q}(a)\\ &= \int_c^1 G_{\xi,p}(a)\mathrm{d}G_{y,q}(a)\ ,\end{aligned}$$

where the limit follows as $H/n \to c$ as $n \to \infty$ by twice applying Theorem 4.1 of de Jong and Davidson (2000*b*).

Next, we consider equation (3). We have that for $\tilde{c} = H/n$ Lemma 1 implies that $\sum_{t=1}^{\lfloor n(a-\tilde{c})\rfloor} X_{n,s}^{\xi,p} \Rightarrow G_{\xi,p}(a-c)$ and also $\sum_{t=1}^{\lfloor na\rfloor} X_{n,t}^{y,q} \Rightarrow G_{y,p}(a)$. We then have

$$\begin{aligned}\frac{1}{n^{1+p+q}}\sum_{t=H+2}^{n}\sum_{s=1}^{t-H-1}s^p\xi_s t^q y_t &= \sum_{t=H+1}^{n-1}\sum_{s=1}^{t-H} X_{n,s}^{\xi,p}X_{n,t+1}^{y,q}\\ &\Rightarrow \int_c^1 G_{\xi,p}(a-c)\mathrm{d}G_{y,q}(a)\ .\end{aligned}$$

□

Proof of Lemma 3. For $y = u, v_1, v_2, v_3$ and $p = 0, 1, 2$, we decompose $\sum_{t=H+1}^{n}\sum_{h=0}^{H} h^p\xi_{t-h}y_t$ and repeatedly apply Lemma 2 to obtain the limiting distribution.
$p = 0$ implies

$$\begin{aligned}\frac{1}{n}\sum_{t=H+1}^{n}\sum_{h=0}^{H}\xi_{t-h}y_t &= \frac{1}{n}\sum_{t=H+1}^{n}\sum_{s=1}^{t}\xi_s y_t - \frac{1}{n}\sum_{t=H+2}^{n}\sum_{s=1}^{t-H-1}\xi_s y_t\\ &\Rightarrow \int_c^1 G_{\xi,0}(a)\mathrm{d}G_{y,0}(a) - \int_c^1 G_{\xi,0}(a-c)\mathrm{d}G_{y,0}(a)\\ &= \int_c^1 G_{\xi,0}(a) - G_{\xi,0}(a-c)\mathrm{d}G_{y,0}(a)\\ &= \int_c^1 D_{1,\xi}(a)\mathrm{d}G_{y,0}(a)\ ,\end{aligned}$$

where

$$D_{1,\xi}(a) = G_{\xi,0}(a) - G_{\xi,0}(a-c)$$

$p = 1$ implies

$$\begin{aligned}
\frac{1}{n^2}\sum_{t=H+1}^{n}\sum_{h=0}^{H} h\xi_{t-h}y_t &= \frac{1}{n^2}\sum_{t=H+1}^{n}\sum_{s=1}^{t}\xi_s t y_t - \frac{1}{n^2}\sum_{t=H+2}^{n}\sum_{s=1}^{t-H-1}\xi_s t y_t \\
&- \frac{1}{n^2}\sum_{t=H+2}^{n}\sum_{s=1}^{t} s\xi_s y_t + \frac{1}{n^2}\sum_{t=H+1}^{n}\sum_{s=1}^{t-H-1} s\xi_s y_t \\
&\Rightarrow \int_c^1 G_{\xi,0}(a)\mathrm{d}G_{y,1}(a) - \int_c^1 G_{\xi,0}(a-c)\mathrm{d}G_{y,1}(a) \\
&- \int_c^1 G_{\xi,1}(a)\mathrm{d}G_{y,0}(a) + \int_c^1 G_{\xi,1}(a-c)\mathrm{d}G_{y,0}(a) \\
&= \int_c^1 aG_{\xi,0}(a) - aG_{\xi,0}(a-c) - G_{\xi,1}(a) + G_{\xi,1}(a-c)\mathrm{d}G_{y,0} \\
&= \int_c^1 D_{2,\xi}(a)\mathrm{d}G_{y,0}\ ,
\end{aligned}$$

where

$$D_{2,\xi}(a) = aG_{\xi,0}(a) - aG_{\xi,0}(a-c) - G_{\xi,1}(a) + G_{\xi,1}(a-c)$$

$p = 2$ implies

$$\begin{aligned}
\frac{1}{n^3}\sum_{t=H+1}^{n}\sum_{h=0}^{H} h^2\xi_{t-h}y_t &= \frac{1}{n^3}\sum_{t=H+1}^{n}\sum_{s=1}^{t}\xi_s t^2 y_t - \frac{1}{n^3}\sum_{t=H+2}^{n}\sum_{s=1}^{t-H-1}\xi_s t^2 y_t \\
&- \frac{2}{n^3}\sum_{t=H+1}^{n}\sum_{s=1}^{t} s\xi_s t y_t + \frac{2}{n^3}\sum_{t=H+2}^{n}\sum_{s=1}^{t-H-1} s\xi_s t y_t \\
&+ \frac{1}{n^3}\sum_{t=H+1}^{n}\sum_{s=1}^{t} s^2\xi_s y_t - \frac{1}{n^3}\sum_{t=H+2}^{n}\sum_{s=1}^{t-H-1} s^2\xi_s y_t \\
&\Rightarrow \int_c^1 G_{\xi,0}(a)\mathrm{d}G_{y,2}(a) - \int_c^1 G_{\xi,0}(a-c)\mathrm{d}G_{y,2}(a) \\
&- 2\int_c^1 G_{\xi,1}(a)\mathrm{d}G_{y,1}(a) + 2\int_c^1 G_{\xi,1}(a-c)\mathrm{d}G_{y,1}(a) \\
&+ \int_c^1 G_{\xi,2}(a)\mathrm{d}G_{y,0}(a) - \int_c^1 G_{\xi,2}(a-c)\mathrm{d}G_{y,0}(a) \\
&= \int_c^1 a^2G_{\xi,0}(a) - a^2G_{\xi,0}(a-c) - 2aG_{\xi,1}(a) + 2aG_{\xi,1}(a-c) \\
&+ G_{\xi,2}(a) - G_{\xi,2}(a-c)\mathrm{d}G_{y,0} \\
&= \int_c^1 D_{3,\xi}\mathrm{d}G_{y,0}
\end{aligned}$$

where

$$D_{3,\xi}(a) = a^2 G_{\xi,0}(a) - a^2 G_{\xi,0}(a-c) - 2aG_{\xi,1}(a) + 2aG_{\xi,1}(a-c) + G_{\xi,2}(a) - G_{\xi,2}(a-c) \ .$$

Next, recall that $z_t = (\sum_{h=0}^{H} \xi_{t-h}, \sum_{h=0}^{H} h\xi_{t-h}, \sum_{h=0}^{H} h^2\xi_{t-h})'$, we combine the results above to obtain

$$K_n^{-1} \sum_{t=H+1}^{n} z_t y_t \Rightarrow \int_c^1 D_\xi(a) \mathrm{d}G_{y,0}(a) \ ,$$

where $D_\xi(a) = (D_{1,\xi}(a), D_{2,\xi}(a), D_{3,\xi}(a))'$. Since the derivations hold for each $y = u, v_1, v_2, v_3$ we have that

$$(I_4 \otimes K_n^{-1}) \sum_{t=H+1}^{n} \begin{pmatrix} u_t \\ v_t \end{pmatrix} \otimes z_t \Rightarrow \begin{bmatrix} \int_c^1 D_\xi(a)\mathrm{d}G_{u,0}(a) \\ \int_c^1 D_\xi(a)\mathrm{d}G_{v_1,0}(a) \\ \int_c^1 D_\xi(a)\mathrm{d}G_{v_2,0}(a) \\ \int_c^1 D_\xi(a)\mathrm{d}G_{v_3,0}(a) \end{bmatrix} \equiv \Xi \ .$$

Finally, note that D_ξ depends only on $G_{\xi,p}$ and $G_{\xi,p}$ is independent from $G_{uv,q}$ for all $p, q \geq 0$ (see Lemma 1). Thus, when we condition on D_ξ (as in Lemma 5.1 of Park and Phillips (1988)) we have that

$$\Xi|_{D_\xi} \sim N\left(0, \Omega_{uv,0} \otimes \int_c^1 D_\xi(a) D_\xi(a)' \mathrm{d}a\right) \ .$$

□

Proof of Lemma 4. Note that

$$z_t = \begin{bmatrix} \sum_{h=0}^{H} \xi_{t-h} \\ \sum_{h=0}^{H} h\xi_{t-h} \\ \sum_{h=0}^{H} h^2\xi_{t-h} \end{bmatrix}$$
$$= \begin{bmatrix} \sum_{s=1}^{t} \xi_s - \sum_{s=1}^{t-H-1} \xi_s \\ \sum_{s=1}^{t} t\xi_s - \sum_{s=1}^{t} s\xi_s - \sum_{s=1}^{t-H-1} t\xi_s + \sum_{s=1}^{t-H-1} s\xi_s \\ \sum_{s=1}^{t} t^2\xi_s - \sum_{s=1}^{t-H-1} t^2\xi_s - 2\sum_{s=1}^{t} ts\xi_s + 2\sum_{s=1}^{t-H-1} ts\xi_s + \sum_{s=1}^{t} s^2\xi_s - \sum_{s=1}^{t-H-1} s^2\xi_s \end{bmatrix}$$

Define $X_n^{\xi,p}(a) = n^{-1/2}\sum_{t=1}^{\lfloor na \rfloor}(t/n)^p \xi_s$, $\tilde{c} = (H-1)/n$ and for $a \in [\lfloor (H+1)/n \rfloor, 1]$

$$D_{n,\xi}(a) = \begin{bmatrix} X_n^{\xi,0}(a) - X_n^{\xi,0}(a-\tilde{c}) \\ aX_n^{\xi,0}(a) - X_n^{\xi,1}(a) - aX_n^{\xi,0}(a-\tilde{c}) + X_n^{\xi,1}(a-\tilde{c}) \\ a^2X_n^{\xi,0}(a) - a^2X_n^{\xi,0}(a-\tilde{c}) - 2aX_n^{\xi,1}(a) + 2aX_n^{\xi,0}(a-\tilde{c}) + X_n^{\xi,2}(a) - X_n^{\xi,2}(a-\tilde{c}) \end{bmatrix}$$

and note that $D_{n,\xi}(a) \Rightarrow D_\xi(a)$ by Lemma 1 where $D_\xi(a)$ is defined in Lemma 3. The continuous mapping theorem implies

$$K_n^{-1}\sum_{t=H+1}^{n} z_t z_t' K_n^{-1} = \frac{1}{n}\sum_{t=H+1}^{n} D_{n,\xi}\left(\frac{t}{n}\right) D_{n,\xi}\left(\frac{t}{n}\right)' \Rightarrow \int_c^1 D_\xi(a)D_\xi(a)'\mathrm{d}a$$

□

Proof of Lemma 5. Assumption 1 ensures that Theorem 1 of de Jong and Davidson (2000*a*) applies, which shows that $S_{uv} - \frac{1}{n-H}\sum_{t=H+1}^{n}\sum_{s=H+1}^{n}\mathrm{E}\left[(u_t, v_t')'(u_s, v_s')\right] \xrightarrow{p} 0$. Further, assumption 1.4 implies that, for $q = 0$, $\frac{1}{n-H}\sum_{t=H+1}^{n}\sum_{s=H+1}^{n}\mathrm{E}\left[(u_t, v_t')'(u_s, v_s')\right] = \Omega_{uv,0} + o_p(1)$. For the second part, notice that $\widehat{S}_{uv}$ can be written as

$$\begin{aligned}\widehat{S}_{uv} =& \frac{1}{n-H}(u:V)'B_n(u:V) + \frac{1}{n-H}(u:V)'P_Z B_n P_Z(u:V)\\ &- \frac{1}{n-H}(u:V)'B_n P_Z(u:V) - \frac{1}{n-H}(u:V)'P_Z B_n(u:V)\end{aligned}$$

and note that the first term equals $\frac{1}{n-H}(u:V)'B_n(u:V) = S_{uv}$. Next, assumption 1.5 implies that the eigenvalues of B_n are bounded[6] such that the second term behaves like

$$\frac{1}{n-H}(u:V)'P_Z B_n P_Z(u:V) \leq \mu_{\max}(B_n)\frac{1}{n-H}(u:V)'ZK_n^{-1}(K_n^{-1}Z'ZK_n^{-1})^{-1}K_n^{-1}Z'(u:V) \xrightarrow{p} 0$$

where $K_n = \mathrm{diag}(n, n^2, n^3)$ and the result follows as

$$\mathrm{vec}\left(K_n^{-1}Z'(u:V)\right) \Rightarrow \Xi = O_p(1) \qquad (K_n^{-1}Z'ZK_n^{-1})^{-1} \Rightarrow \left(\int_c^1 D_\xi(a)D_\xi(a)'\mathrm{d}a\right)^{-1} = O_p(1)$$

by Lemmas 3, 4 and the continuous mapping theorem. Finally, the last two terms also converge to zero. To see this consider the upper left element of the matrix $\frac{1}{n-H}(u:V)'B_n P_Z(u:V)$.

$$\begin{aligned}\frac{1}{n-H}u'B_n P_Z u =& \frac{1}{n-H}u'B_n^{1/2}B_n^{1/2\prime}ZK_n^{-1}(K_nZ'ZK_n)^{-1}K_n^{-1}Z'u\\ \leq& \frac{1}{\sqrt{n-H}}\left(\frac{1}{n-H}u'B_n u\right)^{1/2}\\ &\times\left(u'ZK_n^{-1}(K_nZ'ZK_n)^{-1}K_nZ'B_nZK_n^{-1}(K_nZ'ZK_n)^{-1}K_n^{-1}Z'u\right)^{1/2}\\ \leq& \frac{\mu_{\max}^{1/2}(B_n)}{\sqrt{n-H}}\left(\frac{1}{n-H}u'B_n u\right)^{1/2}\left(u'ZK_n^{-1}(K_nZ'ZK_n)^{-1}K_n^{-1}Z'u\right)^{1/2} \xrightarrow{p} 0\end{aligned}$$

as $\frac{1}{n-H}u'B_n u = \omega_u^2 + o_p(1)$ by the first part of this lemma and $u'ZK_n^{-1}(K_nZ'ZK_n)^{-1}K_n^{-1}Z'u \Rightarrow$

[6] As by Hölder's inequality $\|B_n\| \leq \sqrt{\|B_n\|_1\|B_n\|_\infty}$ and $\|B_n\|_1$ and $\|B_n\|_\infty$ are bounded by the absolute integratability of the kernel function.

$\Xi_u' \left(\int_c^1 D_\xi(a) D_\xi(a)' \mathrm{d}a\right)^{-1} \Xi_u = O_p(1)$ by Lemma 3 and 4. Note that the identical calculations hold for the other elements of $\frac{1}{n-H}(u:V)'B_n P_Z(u:V)$ and $\frac{1}{n-H}(u:V)'P_Z B_n(u:V)$. We conclude that

$$\widehat{S}_{uv} = S_{uv} + o_p(1) \ .$$

□

Proof Lemma 6. Let $\tilde{y} = y - W_\beta \beta_0$, $R_Z = M_Z B_n M_Z/(n-H)$, $Y^* = [\tilde{y} : W_\alpha]$, $d = (1, -\alpha')'$ and B_n is an $(n-H) \times (n-H)$ matrix with s,t entry equal to $\kappa((t-s)/b_n)$, see Lemma 5. Consider the statistic

$$\begin{aligned} k &\equiv \frac{(y - W_\beta\beta_0 - W_\alpha\alpha)'P_Z(y - W_\beta\beta_0 - W_\alpha\alpha)}{(y - W_\beta\beta_0 - W_\alpha\alpha)'M_Z B_n M_Z(y - W_\beta\beta_0 - W_\alpha\alpha)/(n-H)} \\ &= \frac{(\tilde{y} - W_\alpha\alpha)'P_Z(\tilde{y} - W_\alpha\alpha)}{(\tilde{y} - W_\alpha\alpha)'R_Z(\tilde{y} - W_\alpha\alpha)} \\ &= \frac{d'Y^{*'}P_Z Y^* d}{d'Y^{*'}R_Z Y^* d} \end{aligned}$$

The first order condition wrt d can be written as

$$2Y^{*'}P_Z Y^* d(d'Y^{*'}\mathcal{R}_Z Y^* d) - 2Y^{*'}\mathcal{R}_Z Y^* d(d'Y^{*'}P_Z Y^* d) = 0$$

Dividing on both sides by $2d'Y^{*'}\mathcal{R}_Z Y^* d$ gives

$$Y^{*'}P_Z Y^* d - kY^{*'}\mathcal{R}_Z Y^* d = 0$$

Pre-multiplying by $(Y^{*'}\mathcal{R}_Z Y^*)^{-1/2}$ and rearranging gives

$$\left(kI_{m_\alpha+1} - (Y^{*'}\mathcal{R}_Z Y^*)^{-1/2} Y^{*'}P_Z Y^* (Y^{*'}\mathcal{R}_Z Y^*)^{-1/2'}\right) d^* = 0$$

where $d^* = (Y^{*'}\mathcal{R}_Z Y^*)^{1/2'} d$. Hence, the minimum value of k, e.g. $k_{\min} = \min_\alpha k$ is the smallest root of the characteristic polynomial

$$\left|kI_{m_\alpha+1} - (Y^{*'}\mathcal{R}_Z Y^*)^{-1/2} Y^{*'}P_Z Y^* (Y^{*'}\mathcal{R}_Z Y^*)^{-1/2'}\right| = 0 \ .$$

Since $AR_a[(\beta_0', \alpha')]$ has the same functional form as k it follows that $AR_{a,s}[\beta_0] = k_{\min}$. Next, we rewrite the characteristic polynomial to prove the lemma. First, pre-multiply by $\left|\begin{pmatrix} 1 & 0 \\ -\alpha & I_{m_\alpha} \end{pmatrix}' (Y^{*'}\mathcal{R}_Z Y^*)^{1/2}\right|$ and post-multiply by $\left|(Y^{*'}\mathcal{R}_Z Y^*)^{1/2'} \begin{pmatrix} 1 & 0 \\ -\alpha & I_{m_\alpha} \end{pmatrix}\right|$ to obtain

$$\left|k\widehat{S}_{uv_\alpha} - (u : Z\Pi_\alpha + V_\alpha)'P_Z(u : Z\Pi_\alpha + V_\alpha)\right| = 0$$

By Lemma 5 we have that $\widehat{S}_{uv_\alpha} \xrightarrow{p} \Omega_{uv_\alpha,0}$. Note that

$$\Omega_{uv_\alpha,0}^{-1/2} = \begin{pmatrix} \omega_{u,0}^{-1} & -\omega_{u,0}^{-1}\omega_{uv_\alpha,0}\Omega_{v_\alpha v_\alpha \cdot u}^{-1/2} \\ 0 & \Omega_{v_\alpha v_\alpha \cdot u}^{-1/2} \end{pmatrix} \qquad \text{and} \qquad \widehat{S}_{uv_\alpha}^{-1/2} = \begin{pmatrix} \hat{s}_u^{-1} & -\hat{s}_u^{-1}\hat{s}_{uv_\alpha}\widehat{S}_{v_\alpha v_\alpha \cdot u}^{-1/2} \\ 0 & \widehat{S}_{v_\alpha v_\alpha \cdot u}^{-1/2} \end{pmatrix}$$

Pre- and post-multiply by $|\Omega_{uv_\alpha,0}^{-1/2\prime}|$ and $|\Omega_{uv_\alpha,0}^{-1/2}|$ respectively gives

$$\left|k\Omega_{uv_\alpha,0}^{-1/2\prime}\widehat{S}_{uv_\alpha}\Omega_{uv_\alpha,0}^{-1/2} - \Omega_{uv_\alpha,0}^{-1/2\prime}(u : Z\Pi_\alpha + V_\alpha)'P_Z(u : Z\Pi_\alpha + V_\alpha)\Omega_{uv_\alpha,0}^{-1/2}\right| = 0$$

or

$$\left|k\Omega_{uv_\alpha,0}^{-1/2\prime}\widehat{S}_{uv_\alpha}\Omega_{uv_\alpha,0}^{-1/2} - (r_{u,n} : \Theta_n + r_{v_\alpha,n})'(r_{u,n} : \Theta_n + r_{v_\alpha,n})\right| = 0$$

Now use that

$$\begin{aligned}(r_{u,n} : \Theta_n + r_{v_\alpha,n})'(r_{u,n} : \Theta_n + r_{v_\alpha,n}) &= \begin{pmatrix} r_{u,n}'r_{u,n} & r_{u,n}'(\Theta_n + r_{v_\alpha,n}) \\ (\Theta_n + r_{v_\alpha,n})'r_{u,n} & (\Theta_n + r_{v_\alpha,n})'(\Theta_n + r_{v_\alpha,n}) \end{pmatrix} \\ &= N_n'L_nN_n\end{aligned}$$

Pre- and post-multiply the elements in the characteristic polynomial by $|(\Omega_{uv_\alpha,0}^{1/2}\widehat{S}_{uv_\alpha}^{-1/2})'|$ and $|\Omega_{uv_\alpha,0}^{1/2}\widehat{S}_{uv_\alpha}^{-1/2}|$ to obtain

$$\left|kI_{m_2} - (\Omega_{uv_\alpha,0}^{1/2}\widehat{S}_{uv_\alpha}^{-1/2})'N_n'L_nN_n(\Omega_{uv_\alpha,0}^{1/2}\widehat{S}_{uv_\alpha}^{-1/2})\right| = 0$$

The smallest root of the polynomial is wp1 equal to

$$\min_{g\in\mathcal{R}^{1+m_2}\setminus\{0\}\}} \frac{g'(\Omega_{uv_\alpha,0}^{1/2}\widehat{S}_{uv_\alpha}^{-1/2})'N_nL_nN_n(\Omega_{uv_\alpha,0}^{1/2}\widehat{S}_{uv_\alpha}^{-1/2})g}{g'g}$$

which proves the first statement. If we now use a value of g such that

$$g = (\Omega_{uv_\alpha,0}^{1/2}\widehat{S}_{uv_\alpha}^{-1/2})\begin{pmatrix} 1 \\ -(\tau_n'\tau_n)^{-1/2}\eta_n \end{pmatrix},$$

the bottom m_2 rows of N_n cancel out in the numerator and we obtain the bound

$$AR_{a,s}[\beta_0] \leq \frac{r_{u,n}'M_{\tau_n}r_{u,n}}{\rho_n}.$$

□

Proof of Lemma 7. For ease of exposition assume that $j_n = n$. First, note that

$$\begin{pmatrix} r_{u,n} \\ \text{vec}(r_{v_\alpha,n}) \end{pmatrix} = \begin{pmatrix} I_3\omega_{u,0}^{-1} & 0 \\ -(\Omega_{v_\alpha v_\alpha\cdot u}^{-1/2\prime}\omega_{v_\alpha u,0}\omega_{u,0}^{-2} \otimes I_3) & (\Omega_{v_\alpha v_\alpha\cdot u}^{-1/2\prime} \otimes I_3) \end{pmatrix}\begin{pmatrix} (Z'Z)^{-1/2}Z'u \\ \text{vec}((Z'Z)^{-1/2}Z'V_\alpha) \end{pmatrix}.$$

Under any drifting sequence $\phi_n \in \Phi$ we have that Lemma 4 and the continuous mapping theorem imply $(K_n^{-1}Z'ZK_n^{-1})^{-1/2} \Rightarrow \left(\int_c^1 D_\xi(a)D_\xi'(a)\mathrm{d}a\right)^{-1/2}$ and together with Lemma 3 we have that the second term converges to

$$\begin{pmatrix} (K_n^{-1}Z'ZK_n^{-1})^{-1/2}K_n^{-1}Z'u \\ \text{vec}((K_n^{-1}Z'ZK_n^{-1})^{-1/2}K_n^{-1}Z'V_\alpha) \end{pmatrix} \Rightarrow \left(I_{1+m_\alpha} \otimes \left(\int_c^1 D_\xi(a)D_\xi'(a)\mathrm{d}a\right)^{-1/2}\right)\begin{pmatrix} \Xi_u \\ \Xi_{v_\alpha} \end{pmatrix}.$$

Lemma 3 implies that after conditioning on D_ξ we have that the right handside is normally distributed with variance $\Omega_{uv_\alpha,0} \otimes I_3$. Combining this with the first term gives the variance

$$\begin{pmatrix} I_3\omega_{u,0}^{-1} & 0 \\ -(\Omega_{v_\alpha v_\alpha \cdot u}^{-1/2'}\omega_{v_\alpha u,0}\omega_{u,0}^{-2} \otimes I_3) & (\Omega_{v_\alpha v_\alpha \cdot u}^{-1/2'} \otimes I_3) \end{pmatrix} \begin{pmatrix} \omega_{u,0}^2 I_3 & (\omega_{uv_\alpha,0} \otimes I_3) \\ (\omega_{v_\alpha u,0} \otimes I_3) & (\Omega_{v_\alpha v_\alpha} \otimes I_3) \end{pmatrix} \times$$
$$\begin{pmatrix} I_3\omega_{u,0}^{-1} & 0 \\ -(\Omega_{v_\alpha v_\alpha \cdot u}^{-1/2'}\omega_{v_\alpha u,0}\omega_{u,0}^{-2} \otimes I_3) & (\Omega_{v_\alpha v_\alpha \cdot u}^{-1/2'} \otimes I_3) \end{pmatrix}' = I_{3(1+m_\alpha)}$$

Such that we may conclude that for any $\phi_n \in \Phi$ we have that $(r_{u,n}, \text{vec}(r_{v_\alpha,n})')' \Rightarrow (r_u, \text{vec}(r_{v_\alpha})')'$, where $(r_u, \text{vec}(r_{v_\alpha})')'$ is a standard normal random vector, and hence $r_{u,n}$ and $\text{vec}(r_{v_\alpha,n})$ are asymptotically independent.

Assume, without loss of generality, that the jth diagonal element $\mathcal{D}_j$ of $\mathcal{D}$ is finite for $j \leq i$ and $\mathcal{D}_j = \infty$ for $j > i$, for some $0 \leq i \leq m_\alpha$. Define a full-rank diagonal matrix $\mathcal{B}_n$ with jth diagonal element equal to 1 for $j \leq i$ and equal to $\mathcal{D}_{n,j}^{-1}$ otherwise for $j > i$. Nope that for all large enough n the elements of $\mathcal{B}_n$ are bounded by 1.

Now we can write

$$\begin{aligned}
\Theta_n &= (Z'Z)^{1/2}\Pi_\alpha\Omega_{v_\alpha v_\alpha \cdot u}^{-1/2} \\
&= (K_n^{-1}Z'ZK_n^{-1})^{1/2}K_n\Pi_\alpha\Omega_{v_\alpha v_\alpha \cdot u}^{-1/2} \\
&= (K_n^{-1}Z'ZK_n^{-1})^{1/2}\left(\int_c^1 D_\xi(a)D_\xi'(a)\text{d}a\right)^{-1/2}\Theta(n) \\
&= (K_n^{-1}Z'ZK_n^{-1})^{1/2}\left(\int_c^1 D_\xi(a)D_\xi'(a)\text{d}a\right)^{-1/2}O_{1,n}\mathcal{D}_nO_{2,n}'
\end{aligned}$$

Then noting that Lemma 4 implies $(K_n^{-1}Z'ZK_n^{-1})^{1/2}(\int_c^1 D_\xi(a)D_\xi'(a)\text{d}a)^{-1/2} \xrightarrow{p} I_3$ under any $\phi_n \in \Phi$, we have conditional on D_ξ that $\Theta_nO_{2,n}\mathcal{B}_n \xrightarrow{p} O_1\bar{\mathcal{D}}$, where $\bar{\mathcal{D}}$ is a rectangular diagonal matrix with diagonal elements $\bar{\mathcal{D}}_j = \mathcal{D}_j$ for $j \leq i$ and $\bar{\mathcal{D}}_j = 1$ for $j > i$. Noting that by Lemma 5 $\Omega_{uv_\alpha,0}^{-1/2'}\widehat{S}_{uv_\alpha}\Omega_{uv_\alpha,0}^{-1/2} = I_{1+m_\alpha} + o_p(1)$, we have

$$\begin{aligned}
\rho_n &= (1, -\eta_n'(\tau_n'\tau_n)^{-1/2})(\Omega_{uv_\alpha,0}^{-1/2'}\widehat{S}_{uv_\alpha}\Omega_{uv_\alpha,0}^{-1/2})(1, -\eta_n'(\tau_n'\tau_n)^{-1/2})' \\
&= 1 + \eta_n'(\tau_n'\tau_n)^{-1}\eta_n + (1, e_n)o_p(1)(1, e_n)'
\end{aligned}$$

where

$$e_n = -r_{u,n}'(\tau_nO_{2,n}\mathcal{B}_n)\left((\tau_nO_{2,n}\mathcal{B}_n)'(\tau_nO_{2,n}\mathcal{B}_n)\right)^{-1}(O_{2,n}\mathcal{B}_n)'$$

and we now show that $e_n = O_p(1)$. Note that $\tau_nO_{2,n}\mathcal{B}_n = \Theta_nO_{2,n}\mathcal{B}_n + r_{v_\alpha,n}O_{2,n}\mathcal{B}_n$, $\Theta_nO_{2,n}\mathcal{B}_n \xrightarrow{p} O_1\bar{\mathcal{D}}$ and

$$r_{v_\alpha,n}O_{2,n}\mathcal{B}_n \Rightarrow \bar{r}_{v_\alpha} \equiv (r_{v_\alpha}O_{2,1}, \ldots, r_{v_\alpha}O_{2,i}, 0, \ldots, 0)$$

where $O_{2,k}$ denotes the kth column of O_2 and using that $\mathcal{D}_{n,j}^{-1} \to 0$ for $j > i$. Note that $\text{vec}(r_{v_\alpha}O_{2,1}, \ldots, r_{v_\alpha}O_{2,i}) \sim N(0, I_{3i})$ as the columns of O_2 are orthogonal to each other. Therefore, $O_1\bar{\mathcal{D}} + \bar{r}_{v_\alpha}$ has full column rank with probability 1. This implies that $((\tau_nO_{2,n}\mathcal{B}_n)'(\tau_nO_{2,n}\mathcal{B}_n))^{-1} = O_p(1)$ and given that $O_{2,n}\mathcal{B}_n = O_p(1)$ we have $e_n = O_p(1)$. This proves the first claim if we take $p_n = \eta_n'(\tau_n'\tau_n)^{-1}\eta_n$.

Note that because $O_{2,n}\mathcal{B}_n$ has full rank we have $M_{\tau_n} = M_{\tau_nO_{2,n}\mathcal{B}_n}$. As established above we have $\tau_n O_{2,n} \mathcal{B}_n \Rightarrow O_1 \bar{\mathcal{D}} + \bar{r}_{v_\alpha}$ and this limit is independent of the limit of $r_{u,n}$ which is $r_u \sim N(0, I_3)$. Therefore $r'_{u,n} M_{\tau_n} r_{u,n} \Rightarrow r'_u M_{O_1 \bar{\mathcal{D}} + \bar{r}_{v_\alpha}} r_u$ which conditional on $\bar{r}_{v_\alpha}$ is distributed $\chi^2(3 - m_\alpha)$ whenever $O_1 \bar{\mathcal{D}} + \bar{r}_{v_\alpha}$ has full column rank, therefore also unconditionally $r'_u M_{O_1 \bar{\mathcal{D}} + \bar{r}_{v_\alpha}} r_u \sim \chi^2(3 - m_\alpha)$ and note that $3 - m_\alpha = m_\beta$ in our exactly identified setting. $\square$

Table 1: Simulation results part 1

	AR_a, $n=200$, specification (i)				$AR_{a,s}$, $n=200$, specification (i)			
	$\sigma_i=0.1$	$\sigma_i=0.25$	$\sigma_i=0.5$	$\sigma_i=1.0$	$\sigma_i=0.1$	$\sigma_i=0.25$	$\sigma_i=0.5$	$\sigma_i=1.0$
$H=5$	0.053	0.055	0.057	0.057	0.000	0.001	0.009	0.034
$H=10$	0.052	0.043	0.044	0.049	0.011	0.015	0.035	0.055
$H=20$	0.038	0.044	0.044	0.046	0.002	0.005	0.015	0.042
$H=40$	0.044	0.040	0.039	0.041	0.000	0.000	0.000	0.002
$H=80$	0.039	0.035	0.041	0.041	0.000	0.000	0.000	0.000
	AR_a, $n=500$, specification (i)				$AR_{a,s}$, $n=500$, specification (i)			
	$\sigma_i=0.1$	$\sigma_i=0.25$	$\sigma_i=0.5$	$\sigma_i=1.0$	$\sigma_i=0.1$	$\sigma_i=0.25$	$\sigma_i=0.5$	$\sigma_i=1.0$
$H=5$	0.052	0.059	0.056	0.053	0.000	0.003	0.026	0.045
$H=10$	0.052	0.053	0.053	0.047	0.008	0.022	0.044	0.048
$H=20$	0.049	0.051	0.046	0.050	0.004	0.016	0.039	0.047
$H=40$	0.045	0.044	0.044	0.047	0.000	0.000	0.002	0.008
$H=80$	0.041	0.046	0.047	0.040	0.000	0.000	0.000	0.000
	AR_a, $n=200$, specification (ii)				$AR_{a,s}$, $n=200$, specification (ii)			
	$\sigma_i=0.1$	$\sigma_i=0.25$	$\sigma_i=0.5$	$\sigma_i=1.0$	$\sigma_i=0.1$	$\sigma_i=0.25$	$\sigma_i=0.5$	$\sigma_i=1.0$
$H=5$	0.055	0.054	0.051	0.055	0.030	0.038	0.036	0.036
$H=10$	0.048	0.053	0.051	0.045	0.050	0.053	0.051	0.051
$H=20$	0.046	0.043	0.040	0.043	0.043	0.036	0.034	0.036
$H=40$	0.038	0.041	0.039	0.033	0.002	0.001	0.002	0.003
$H=80$	0.044	0.035	0.037	0.040	0.000	0.000	0.000	0.000
	AR_a, $n=500$, specification (ii)				$AR_{a,s}$, $n=500$, specification (ii)			
	$\sigma_i=0.1$	$\sigma_i=0.25$	$\sigma_i=0.5$	$\sigma_i=1.0$	$\sigma_i=0.1$	$\sigma_i=0.25$	$\sigma_i=0.5$	$\sigma_i=1.0$
$H=5$	0.052	0.050	0.062	0.045	0.048	0.045	0.050	0.048
$H=10$	0.053	0.048	0.047	0.052	0.051	0.058	0.049	0.051
$H=20$	0.046	0.044	0.048	0.052	0.042	0.049	0.045	0.045
$H=40$	0.043	0.044	0.048	0.046	0.011	0.007	0.008	0.008
$H=80$	0.042	0.044	0.038	0.041	0.000	0.000	0.000	0.000

Notes: The table reports the empirical rejection frequencies of: 1. the AR_a statistic for testing $H_0 : a_{y,1} = a^0_{y,1},\ a_{y,2} = a^0_{y,2},\ \lambda = \lambda^0$ with level $\alpha = 0.05$ and 2. the subset $AR_{a,s}$ statistic to test $H_0 : \lambda = \lambda^0$ with level $\alpha = 0.05$. Different choices of n, H, σ_i and error designs are considered. The error designs correspond to: (i) no heteroskedasticity, no serial correlation, (ii) heteroskedasticity from ϵ_t, no serial correlation, (iii) heteroskedasticity from garch, no serial correlation, (iv) no heteroskedasticity, serial correlated errors, (v) heteroskedasticity from ϵ_t, serial correlated errors, (vi) heteroskedasticity from garch, serial correlated errors.

Table 2: Simulation results part 2

	AR_a, $n = 200$, specification (iii)				$AR_{a,s}$, $n = 200$, specification (iii)			
	$\sigma_i = 0.1$	$\sigma_i = 0.25$	$\sigma_i = 0.5$	$\sigma_i = 1.0$	$\sigma_i = 0.1$	$\sigma_i = 0.25$	$\sigma_i = 0.5$	$\sigma_i = 1.0$
$H = 5$	0.056	0.053	0.059	0.064	0.000	0.001	0.010	0.039
$H = 10$	0.049	0.052	0.049	0.051	0.008	0.019	0.033	0.054
$H = 20$	0.044	0.046	0.045	0.047	0.003	0.007	0.020	0.040
$H = 40$	0.039	0.041	0.034	0.043	0.000	0.000	0.001	0.003
$H = 80$	0.035	0.040	0.042	0.041	0.000	0.000	0.000	0.000
	AR_a, $n = 500$, specification (iii)				$AR_{a,s}$, $n = 500$, specification (iii)			
	$\sigma_i = 0.1$	$\sigma_i = 0.25$	$\sigma_i = 0.5$	$\sigma_i = 1.0$	$\sigma_i = 0.1$	$\sigma_i = 0.25$	$\sigma_i = 0.5$	$\sigma_i = 1.0$
$H = 5$	0.049	0.054	0.052	0.060	0.000	0.005	0.030	0.045
$H = 10$	0.054	0.053	0.050	0.050	0.012	0.029	0.045	0.054
$H = 20$	0.048	0.052	0.049	0.046	0.007	0.016	0.039	0.049
$H = 40$	0.044	0.047	0.044	0.044	0.000	0.002	0.003	0.007
$H = 80$	0.037	0.045	0.042	0.043	0.000	0.000	0.000	0.000
	AR_a, $n = 200$, specification (iv)				$AR_{a,s}$, $n = 200$, specification (iv)			
	$\sigma_i = 0.1$	$\sigma_i = 0.25$	$\sigma_i = 0.5$	$\sigma_i = 1.0$	$\sigma_i = 0.1$	$\sigma_i = 0.25$	$\sigma_i = 0.5$	$\sigma_i = 1.0$
$H = 5$	0.012	0.012	0.016	0.011	0.000	0.000	0.001	0.001
$H = 10$	0.036	0.037	0.031	0.032	0.012	0.014	0.018	0.028
$H = 20$	0.058	0.062	0.054	0.057	0.006	0.012	0.030	0.052
$H = 40$	0.066	0.075	0.071	0.072	0.000	0.000	0.000	0.005
$H = 80$	0.076	0.083	0.079	0.075	0.000	0.000	0.000	0.000
	AR_a, $n = 500$, specification (iv)				$AR_{a,s}$, $n = 500$, specification (iv)			
	$\sigma_i = 0.1$	$\sigma_i = 0.25$	$\sigma_i = 0.5$	$\sigma_i = 1.0$	$\sigma_i = 0.1$	$\sigma_i = 0.25$	$\sigma_i = 0.5$	$\sigma_i = 1.0$
$H = 5$	0.009	0.009	0.011	0.011	0.000	0.000	0.001	0.002
$H = 10$	0.029	0.027	0.031	0.030	0.008	0.011	0.019	0.026
$H = 20$	0.051	0.055	0.058	0.048	0.007	0.024	0.056	0.057
$H = 40$	0.071	0.061	0.071	0.068	0.000	0.001	0.006	0.018
$H = 80$	0.080	0.076	0.075	0.069	0.000	0.000	0.000	0.000

Notes: The table reports the empirical rejection frequencies of: 1. the AR_a statistic for testing $H_0 : a_{y,1} = a^0_{y,1},\ a_{y,2} = a^0_{y,2},\ \lambda = \lambda^0$ with level $\alpha = 0.05$ and 2. the subset $AR_{a,s}$ statistic to test $H_0 : \lambda = \lambda^0$ with level $\alpha = 0.05$. Different choices of n, H, σ_i and error designs are considered. The error designs correspond to: (i) no heteroskedasticity, no serial correlation, (ii) heteroskedasticity from ϵ_t, no serial correlation, (iii) heteroskedasticity from garch, no serial correlation, (iv) no heteroskedasticity, serial correlated errors, (v) heteroskedasticity from ϵ_t, serial correlated errors, (vi) heteroskedasticity from garch, serial correlated errors.

Table 3: Simulation results part 3

	AR_a, $n = 200$, specification (v)				$AR_{a,s}$, $n = 200$, specification (v)			
	$\sigma_i = 0.1$	$\sigma_i = 0.25$	$\sigma_i = 0.5$	$\sigma_i = 1.0$	$\sigma_i = 0.1$	$\sigma_i = 0.25$	$\sigma_i = 0.5$	$\sigma_i = 1.0$
$H = 5$	0.025	0.026	0.023	0.024	0.003	0.004	0.005	0.005
$H = 10$	0.049	0.044	0.042	0.044	0.028	0.029	0.029	0.026
$H = 20$	0.063	0.062	0.064	0.070	0.052	0.050	0.052	0.053
$H = 40$	0.071	0.070	0.073	0.073	0.005	0.003	0.005	0.005
$H = 80$	0.069	0.086	0.073	0.075	0.000	0.000	0.000	0.000
	AR_a, $n = 500$, specification (v)				$AR_{a,s}$, $n = 500$, specification (v)			
	$\sigma_i = 0.1$	$\sigma_i = 0.25$	$\sigma_i = 0.5$	$\sigma_i = 1.0$	$\sigma_i = 0.1$	$\sigma_i = 0.25$	$\sigma_i = 0.5$	$\sigma_i = 1.0$
$H = 5$	0.025	0.025	0.026	0.022	0.002	0.003	0.003	0.004
$H = 10$	0.053	0.044	0.051	0.055	0.032	0.024	0.025	0.026
$H = 20$	0.067	0.066	0.062	0.067	0.055	0.059	0.061	0.058
$H = 40$	0.069	0.071	0.070	0.067	0.011	0.012	0.010	0.011
$H = 80$	0.072	0.079	0.079	0.078	0.000	0.000	0.000	0.000
	AR_a, $n = 200$, specification (vi)				$AR_{a,s}$, $n = 200$, specification (vi)			
	$\sigma_i = 0.1$	$\sigma_i = 0.25$	$\sigma_i = 0.5$	$\sigma_i = 1.0$	$\sigma_i = 0.1$	$\sigma_i = 0.25$	$\sigma_i = 0.5$	$\sigma_i = 1.0$
$H = 5$	0.014	0.012	0.013	0.014	0.000	0.000	0.002	0.002
$H = 10$	0.037	0.033	0.029	0.030	0.011	0.013	0.019	0.025
$H = 20$	0.057	0.054	0.059	0.059	0.002	0.011	0.033	0.052
$H = 40$	0.069	0.069	0.067	0.070	0.000	0.000	0.001	0.005
$H = 80$	0.079	0.088	0.084	0.076	0.000	0.000	0.000	0.000
	AR_a, $n = 500$, specification (vi)				$AR_{a,s}$, $n = 500$, specification (vi)			
	$\sigma_i = 0.1$	$\sigma_i = 0.25$	$\sigma_i = 0.5$	$\sigma_i = 1.0$	$\sigma_i = 0.1$	$\sigma_i = 0.25$	$\sigma_i = 0.5$	$\sigma_i = 1.0$
$H = 5$	0.011	0.010	0.013	0.011	0.000	0.001	0.001	0.001
$H = 10$	0.032	0.031	0.031	0.033	0.010	0.014	0.020	0.026
$H = 20$	0.049	0.055	0.057	0.055	0.005	0.025	0.054	0.059
$H = 40$	0.071	0.062	0.068	0.067	0.000	0.001	0.005	0.013
$H = 80$	0.075	0.077	0.074	0.074	0.000	0.000	0.000	0.000

Notes: The table reports the empirical rejection frequencies of: 1. the AR_a statistic for testing $H_0 : a_{y,1} = a^0_{y,1},\ a_{y,2} = a^0_{y,2},\ \lambda = \lambda^0$ with level $\alpha = 0.05$ and 2. the subset $AR_{a,s}$ statistic to test $H_0 : \lambda = \lambda^0$ with level $\alpha = 0.05$. Different choices of n, H, σ_i and error designs are considered. The error designs correspond to: (i) no heteroskedasticity, no serial correlation, (ii) heteroskedasticity from ϵ_t, no serial correlation, (iii) heteroskedasticity from garch, no serial correlation, (iv) no heteroskedasticity, serial correlated errors, (v) heteroskedasticity from ϵ_t, serial correlated errors, (vi) heteroskedasticity from garch, serial correlated errors.

Table 4: The Phillips curve – 1969-2007 — varying H

	Unrestricted		Restricted	
H=10				
γ_b	0.41	$[-\infty, \quad \infty]$		
γ_f	0.35	$[-\infty, \quad \infty]$	0.28	$[-0.01, \quad 0.46]$
λ_{U}	-0.31	$[-\infty, \quad \infty]$	-0.52	$[-1.17, -0.15]$
H=15				
γ_b	0.62	$[-2.63, \quad 1.33]$		
γ_f	0.40	$[\ 0.09, \quad 1.53]$	0.41	$[\ 0.16, \quad 0.61]$
λ_{U}	-0.50	$[-1.79, -0.13]$	-0.50	$[-1.24, -0.14]$
H=20				
γ_b	0.58	$[\ 0.32, \quad 0.91]$		
γ_f	0.47	$[\ 0.18, \quad 0.70]$	0.47	$[\ 0.22, \quad 0.68]$
λ_{U}	-0.45	$[-1.13, -0.11]$	-0.48	$[-1.08, -0.17]$
H=30				
γ_b	0.49	$[\ 0.16, \quad 0.70]$		
γ_f	0.52	$[\ 0.29, \quad 0.83]$	0.52	$[\ 0.30, \quad 0.83]$
λ_{U}	-0.32	$[-1.46, -0.05]$	-0.33	$[-1.42, -0.05]$

Notes: The table reports the parameter estimates and confidence intervals for the US Phillips curve (1969-2007). The top panel shows the IV point estimates based on using H lags of the Romer and Romer (2004) shocks as instruments and the $AR_{a,s}$ based 90% confidence bounds.

Table 5: The Phillips curve – Sub-samples, RR id.

1969-1990

	Unrestricted		Restricted	
γ_b	0.92	$[\ 0.50,\ \ 2.89]$		
γ_f	0.16	$[-1.70,\ \ 0.55]$	0.10	$[-1.64,\ \ 0.50]$
λ_{U}	-0.84	$[-3.76, -0.19]$	-0.92	$[-3.68, -0.30]$

1990-2007

	Unrestricted		Restricted	
γ_b	0.81	$[-\infty,\ \ \infty]$		
γ_f	0.35	$[-\infty,\ \ \infty]$	0.42	$[-\infty,\ \ \infty]$
λ_{U}	0.12	$[-\infty,\ \ \infty]$	0.00	$[-\infty,\ \ \infty]$

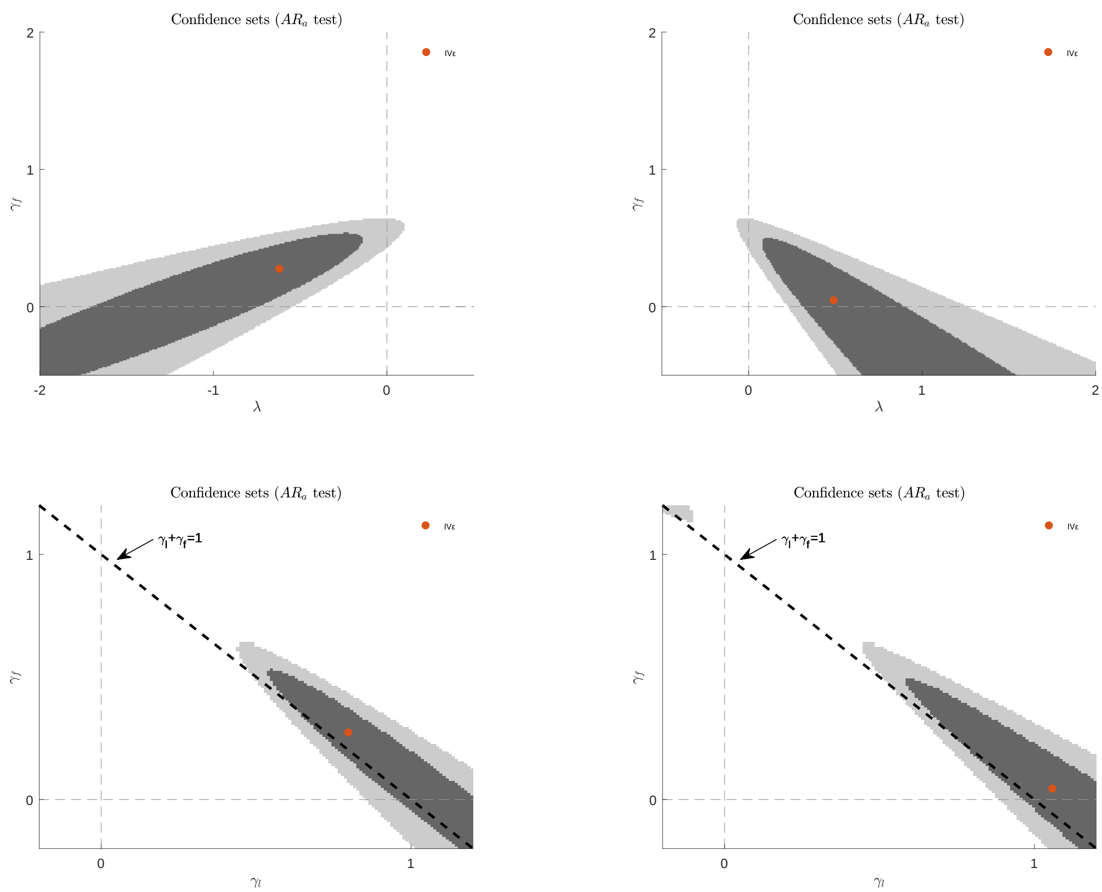
Notes: The table reports the parameter estimates and confidence intervals for the US Phillips curve estimated over 1969-1989 (top row) or over 1990-2007 (bottom row) using the Romer-Romer (RR) monetary shocks as instruments. We show the CUE point estimates based on the Romer and Romer (2004) shocks as instruments ($H = 20$ lags) together with the $AR_{a,s}$ based 90% confidence bounds (in brackets).

Table 6: The Phillips curve – 1969-2007, GIV id.

	Unrestricted		Restricted	
γ_b	0.71	$[\ 0.23,\ 1.18]$		
γ_f	0.30	$[-0.24,\ 0.83]$	0.28	$[-0.09,\ 0.66]$
λ_{U}	-0.65	$[-1.31,\ 0.01]$	-0.65	$[-1.27, -0.05]$
γ_b	0.63	$[\ 0.20,\ 1.06]$		
γ_f	0.40	$[-0.08,\ 0.88]$	0.32	$[-0.08,\ 0.72]$
λ_{Y}	0.28	$[-0.11,\ 0.66]$	0.30	$[-0.09,\ 0.70]$

Notes: The table reports the parameter estimates and 90% confidence intervals for the US Phillips curve (1969-2007) using lagged macro variables as instruments. The instruments are four lags of inflation and the forcing variable. The confidence bounds are standard 90% bounds, e.g. $[\hat{\delta}_j \pm 1.64 se(\hat{\delta}_j)]$, and derived under the assumption of *strong* instruments. Note that the projection and subset confidence intervals are infinite for these specifications.

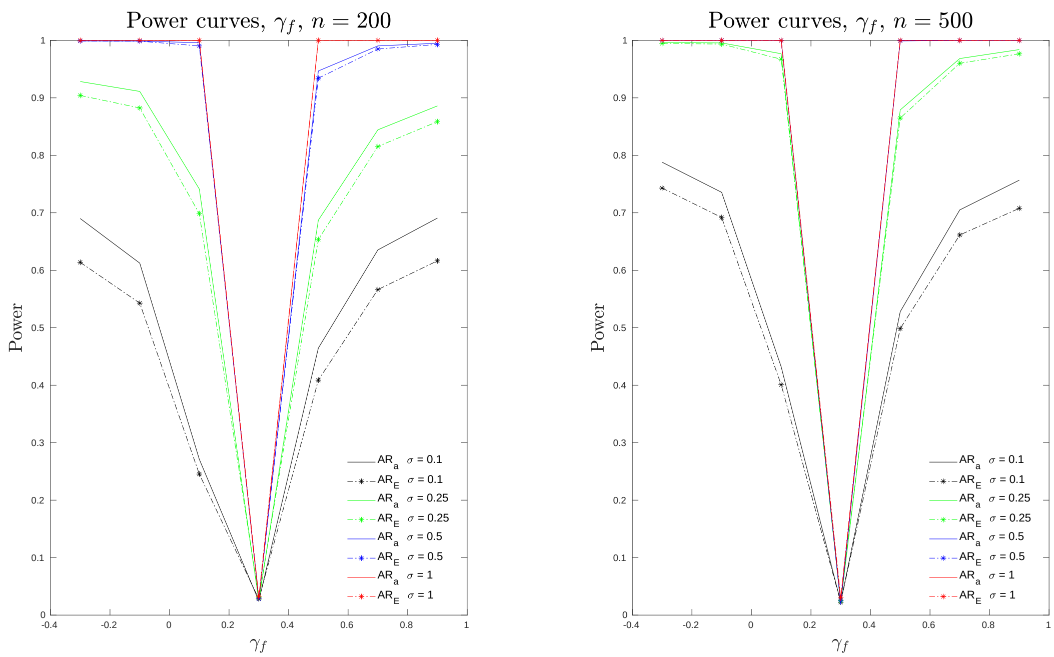
Figure 1: The Phillips curve — 1969-2007, RR id. with conditioning out factors



(a) Unemployment gap (b) Output gap

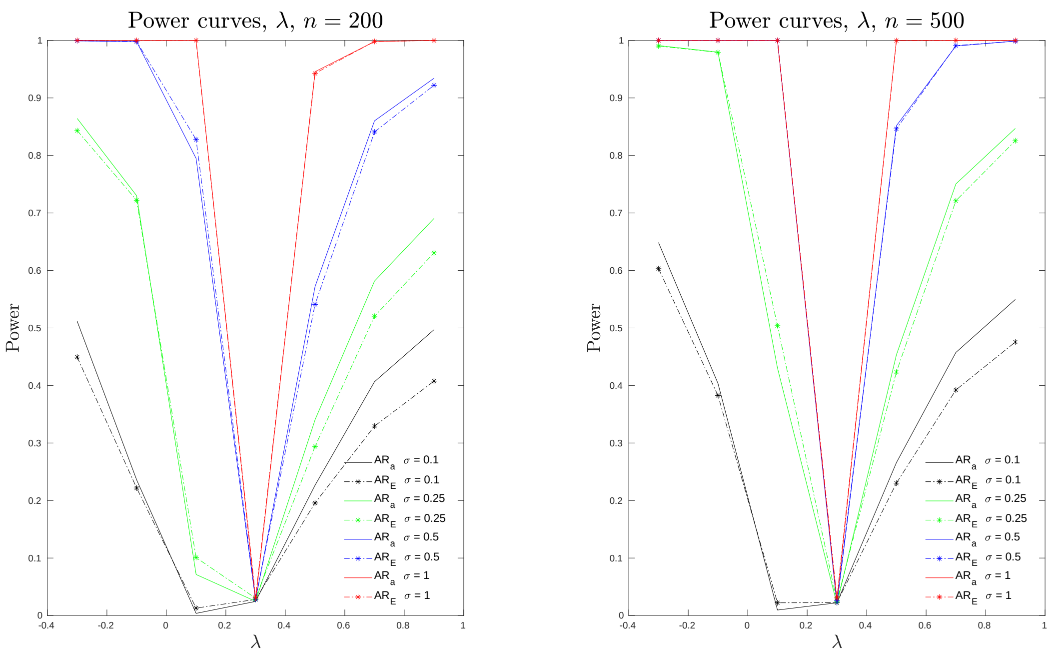
Notes: Top row: Robust confidence sets for the Phillips curve coefficients obtained by inverting the AR_a test. Top row: 68 and 90 percent confidence sets for λ (the slope of the Phillips curve) and γ_f (the loading on inflation expectations). Bottom row: 68 and 90 percent confidence sets for γ_f and γ_b (the loading on lagged inflation) in the bottom row. The dashed line depicts the $\gamma_f + \gamma_b = 1$ set. Estimation based on using the Romer-Romer (RR) monetary shocks as instruments for 1969-2007. The red dot is the Almon-restricted IV estimate. Specification with the unemployment gap (left column) or the output gap (right column) as the forcing variable.

Figure 2: Power curves γ_f



Notes: Power curves corresponding to $H_0 : \delta = \delta_0$ with $H_1 : \gamma_f \neq \gamma_f^0$. The power is computed based on the AR_a (straight line) and AR_E (dotted line) statistics with 5000 replications. The different colors indicate simulations designs with different degrees of instrument strength corresponding to the parameter $\sigma_i = 0, 1, 0.25, 0.5, 1$.

Figure 3: Power curves λ



Notes: Power curves corresponding to $H_0 : \delta = \delta_0$ with $H_1 : \lambda \neq \lambda^0$. The power is computed based on the AR_a (straight line) and AR_E (dotted line) statistics with 5000 replications. The different colors indicate simulations designs with different degrees of instrument strength corresponding to the parameter $\sigma_i = 0, 1, 0.25, 0.5, 1$.