Supplementary Appendix to the paper

One-dimensional inference in autoregressive models with the potential presence of a unit root.

(proofs intended for web-posting)

by Anna Mikusheva

Abstract

The Supplementary Appendix contains proofs of some results stated in the paper, "One-dimensional inference in autoregressive models with the potential presence of a unit root," by Anna Mikusheva. In particular, the proofs of the generalization of the results robust to conditional heteroskedasticity can be found in Section 1 of the Supplementary Appendix. Proofs of the results for multi-dimensional VAR models appear in Section 2. A discussion of the Wald statistic for an IRF at long horizons is placed in Section 3. Section 4 provides a simplified formula for u in the AR(2) case.

1 Heteroskedasticity robust inference

In this section we generalize the results of the paper to allow for conditionally heteroskedastic processes. There are some challenges to obtaining full uniformity over $\mathfrak{R}_\delta$, as Mikusheva (2007) uses conditional homoskedasticity extensively in employing the Skorokhod representation. However, obtaining point-wise results in the local-to-unity embedding is relatively straightforward. Andrews and Guggenberger (2010) suggest that establishing asymptotic results for all local-to-unity sequences should be enough to establish the uniformity.

Let us consider a sample from the process

$$y_t = \lambda_p y_{t-1} + u_t, \quad B(L)u_t = e_t, \quad y_0 = 0 \tag{1}$$

where $B(L) = 1 + B_1 L + ... + B_{p-1} L^{p-1}$ is a lag polynomial of order $p-1$ with all roots strictly inside the circle of radius $\delta < 1$,u_t is the stationary realization of an AR(p-1) process, and $\lambda_p = 1 + c/T$ is the local-to-unity root. The regression of interest is the correctly-specified AR(p) regression in ADF form:

$$y_t = \rho y_{t-1} + \sum_{j=1}^{p-1} \alpha_j \Delta y_{t-j} + e_t.$$

Assumption HS. Let e_t be a stationary martingale-difference sequence, with $E|e_t|^{2(\beta+\epsilon)} < \infty$ for some $\beta > 2, \epsilon > 0$, and its mixing numbers α_m satisfy $\sum_{m=1}^{\infty} \alpha_m^{1-2/\beta} < \infty$.

The important point here is that e_t is allowed to be conditionally heteroskedastic.

Introduce the following notation $\theta = (\rho, \alpha')'$, $x_t = (\Delta y_{t-1}, ..., \Delta y_{t-p+1})'$, $X_t = (y_{t-1}, x_t')'$, $X = (X_{p+1}, ..., X_T)'$ is $(T-p) \times p$ regressor matrix, $Y_T = (y_{p+1}, ..., y_T)'$. Let $K_T = diag(1/\sqrt{T}, 1..., 1)$ be a $p \times p$ diagonal matrix, $\omega^2 = E(u_1^2) + 2\sum_{k=1}^{\infty} E(u_1 u_k) = \frac{\sigma^2}{B(1)^2}$ is the long-run variance of u_t, and $\sigma^2 = Ee_t^2$.

Consider the GMM-based Distance-Metric statistic, which is asymptotically equivalent to the LR statistic under assumptions of conditional homoskedasticity:

$$DM_T = Q_T(\tilde{\theta}) - Q_T(\hat{\theta}),$$

where $Q_T(\theta) = e(\theta)' X \Omega_T^{-1} X' e(\theta)$, $\Omega_T = \frac{1}{T}\sum_{t=p+1}^{T} X_t X_t' e_t^2(\hat{\theta})$, $e(\theta) = Y - X\theta$, $\hat{\theta}$ is the OLS estimate, and $\tilde{\theta} = \arg\min_{H_0} Q_T(\theta)$ is the restricted estimate of θ.

Theorem 1 *Let one have a sample from the process defined in equation (1) with errors satisfying Assumptions HS. Consider the following two sequences of hypotheses:*

(i) linear hypothesis $H_0 : A\theta = \gamma_0$ *with the coefficient* $A = A_T$ *satisfying* $\lim_{T\to\infty} \frac{K_T A_T}{\|K_T A_T\|} = a$, *where* a *is a* $p \times 1$ *vector;*

(ii) hypothesis about the IRF at horizon h, *i.e.,* $H_0 : f_h(\theta) = \gamma_0$ *with* $h = h_T : \lim_{T\to\infty} \frac{h_T}{\sqrt{T}} = q \in [0, \infty]$;

For both of them we have $DM_T \Rightarrow (t(c,u))^2$, *where*

$$t(c,u) = \frac{t^c + u\sqrt{\frac{\int_0^1 J_c^2(s)ds}{g(c)}} N(0,1)}{\sqrt{1 + u^2 \frac{\int_0^1 J_c^2(s)ds}{g(c)}}},$$

$$u = \sqrt{\frac{A'F'FA - (i_1'FA)^2}{(i_1'FA)^2}}, \qquad (2)$$

and $A = \frac{\partial}{\partial\theta} f_h(\theta_0)$ *should be used in formula (2) for case (ii).*

The proof uses Lemma 1 as established below.

Lemma 1 *Let Assumption HS be satisfied. Then the following holds simultaneously:*

(a) $\frac{1}{\sqrt{T}} \sum_{t=1}^{[rT]} (e_t, e_t^2 - Ee_t^2)' \Rightarrow (\sigma W_1(r), W_2(r))$, *where* $W = (\sigma W_1, W_2)'$ *is a two-dimensional Brownian motion with the covariance matrix* $\Sigma_1 = \begin{pmatrix} \sigma^2 & \mu_3 \\ \mu_3 & \mu_4 \end{pmatrix}$, $\mu_3 = \sum_{k=0}^{\infty} Ee_t e_{t+k}^2$, $\mu_4 = \sum_{k=-\infty}^{\infty} cov(e_t^2, e_{t+k}^2)$.

(b) $\frac{1}{\sqrt{T}} K_T X'e \Rightarrow (\omega\sigma \int_0^1 J_c(t) dW_1(t), \xi')'$, *where* $\xi \sim N(0, E(e_t^2 x_t x_t'))$, *and* $J_c(r) = \int_0^r e^{c(r-s)} dW_1(s)$;

(c) $\frac{1}{T} K_T X'XK_T \Rightarrow \begin{pmatrix} \omega^2 \int_0^1 J_c^2(t)dt & 0 \\ 0 & E(x_t x_t') \end{pmatrix}$;

(d) $\frac{1}{T} K_T \sum_{t=p+1}^{T} e_t^2 X_t X_t' K_T \Rightarrow \begin{pmatrix} \sigma^2\omega^2 \int_0^1 J_c^2(t)dt & 0 \\ 0 & E(e_t^2 x_t x_t') \end{pmatrix}$;

(e) $\frac{1}{T} \sum_{t=1}^{T} (K_T X_t X_t' K_T) \otimes (K_T X_t X_t' K_T) = O_p(1)$;

(f) $\frac{1}{T} \sum_{t=1}^{T} (K_T X_t X_t' K_T) \otimes (K_T X_t e_t) = O_p(1)$.

Proof of Lemma 1 (a) Consider a vector $v_t = (e_t, u_t, e_t^2 - \sigma^2)'$. According to Phillips (1988):

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{[rT]} v_t \Rightarrow W(r),$$

where $W(r) = (\sigma W_1(r), \frac{\sigma}{B(1)} W_1(r), W_2(r))'$ is a Brownian motion with covariance matrix $\Sigma = \begin{pmatrix} \sigma^2 & \frac{\sigma^2}{B(1)} & \mu_3 \\ \frac{\sigma^2}{B(1)} & \omega^2 & \frac{\mu_3}{B(1)} \\ \mu_3 & \frac{\mu_3}{B(1)} & \mu_4 \end{pmatrix}$. According to Lemma 3.1 in Phillips (1988), statement (a) implies that $\frac{y_{[rT]}}{\sqrt{T}} \Rightarrow \omega J_c(r) = \frac{\sigma}{B(1)} \int_0^r e^{(r-s)c} dW_1$, and statements (b) and (c) hold.

For statement (d) notice that

$$\frac{1}{T^2}\sum_{t=1}^{T} y_{t-1}^2 e_t^2 = \frac{1}{T^2}\sum_{t=1}^{T} y_{t-1}^2 Ee_t^2 + \frac{1}{T^2}\sum_{t=1}^{T} y_{t-1}^2\left(e_t^2 - Ee_t^2\right).$$

The first term converges to $\omega^2(Ee_t^2)\int_0^1 J_c^2(s)ds$, while the second term is negligible. Indeed, according to direct generalization of Theorems 4.2 and 4.4 in Hansen (1992) $\frac{1}{T^{3/2}}\sum_{t=1}^{T} y_{t-1}^2\left(e_t^2 - Ee_t^2\right) \Rightarrow \omega^2\int_0^1 J_c^2(s)dW_2(s) + \mu_3\omega\int_0^1 J_c(s)ds$, and the last expression is bounded in probability. Let us now consider an off-diagonal element in (d), namely, the $(p-1)\times 1$ vector $\frac{1}{T^{3/2}}\sum_{t=1}^{T} y_{t-1}x_t e_t^2$ and show that it converges to zero in probability. Indeed, the i-th component of it has the following form:

$$\frac{1}{T^{3/2}}\sum_{t=1}^{T} y_{t-1}\Delta y_{t-j}e_t^2 = \frac{1}{T^{3/2}}\sum_{t=1}^{T} y_{t-1}u_{t-j}e_t^2 + \frac{c}{T}\frac{1}{T^{3/2}}\sum_{t=1}^{T} y_{t-1}y_{t-j-1}e_t^2. \qquad (3)$$

Lemma 4(b) from Andrews and Guggenberger (2008) with $v_{n,i} = (u_i, u_{i-j}e_i^2)'$ implies that $\frac{1}{T}\sum_{t=1}^{T} y_{t-1}u_{t-j}e_t^2$ converges in distribution to a bounded in probability random variable, and as a result, the first term in (3) is negligible. Following the same reasoning as above, we also know that $\frac{1}{T^2}\sum_{t=1}^{T} y_{t-1}y_{t-j-1}e_t^2$ converges in distribution to a bounded in probability random variable, and thus, the last term in (3) is also negligible. This gives statement (d).

For statement (e) we have to show the following five statements:

$$\frac{1}{T^3}\sum_{t=p+1}^{T} y_{t-1}^4 = O_p(1); \frac{1}{T^{5/2}}\sum_{t=p+1}^{T} y_{t-1}^3 x_t = O_p(1);$$

$$\frac{1}{T^2}\sum_{t=p+1}^{T} y_{t-1}^2 x_t x_t' = O_p(1); \frac{1}{T^{3/2}}\sum_{t=p+1}^{T} y_{t-1}x_t\otimes x_t x_t' = O_p(1);$$

$$\frac{1}{T}\sum_{t=p+1}^{T} y_{t-1}(x_t x_t')\otimes(x_t x_t') = O_p(1).$$

First, notice that $|x_t|$, $\|x_t x_t'\|$, $\|x_t x_t' x_{i,t}\|$ are uniformly integrable L^1-mixingales, see also Hamilton, chapter 7, for the reasoning. According to Andrews (1988)'s Law of Large Numbers for L^1- mixingales $\frac{1}{T}\sum x_t$, $\frac{1}{T}\sum x_t x_t'$, $\frac{1}{T}\sum x_t x_t' x_{i,t}$ satisfy the Law of Large Numbers and thus converge in probability to constants. Since all statements are proven in the same way, we show it for the second statement only:

$$\left|\frac{1}{T^{5/2}}\sum_{t=p+1}^{T} y_{t-1}^3 x_t\right| \le \max_t \left|\frac{y_t}{\sqrt{T}}\right|^3 \frac{1}{T}\sum_{t=1}^{T}|x_t| \Rightarrow \sup_{0\le s\le 1}|J_c(s)|^3 E|x_t|.$$

The last expression is bounded in probability.

The proof of part (f) is analogous to that of part (e).

Proof of Theorem 1. First notice that

$$DM_T = (\hat{\theta} - \tilde{\theta})' X' X \Omega_T^{-1} X' X (\hat{\theta} - \tilde{\theta}), \tag{4}$$

Notice that

$$\Omega_T = \frac{1}{T}\sum_{t=1}^{T} X_t X_t' e_t^2(\hat{\theta}) = \frac{1}{T}\sum_{t=1}^{T} X_t X_t' \left(e_t - (\hat{\theta} - \theta_0)' X_t\right)^2 =$$

$$= \frac{1}{T}\sum_{t=1}^{T} X_t X_t' e_t^2 + \frac{1}{T}\sum_{t=1}^{T} X_t X_t' \left((\hat{\theta} - \theta_0)' X_t\right)^2 + \frac{2}{T}\sum_{t=1}^{T} X_t X_t' \left((\hat{\theta} - \theta_0)' X_t\right) e_t. \tag{5}$$

Let us first show that the second term in equation (5) is asymptotically negligible. Indeed,

$$K_T \frac{1}{T}\sum_{t=1}^{T} X_t X_t' \left((\hat{\theta} - \theta_0)' X_t\right)^2 K_T = K_T \frac{1}{T}\sum_{t=1}^{T} X_t X_t' \left((\hat{\theta} - \theta_0)' K_T^{-1} K_T X_t\right)^2 K_T =$$

$$= \left(I_p \otimes (\hat{\theta} - \theta_0)' K_T^{-1}\right) \frac{1}{T}\sum_{t=1}^{T} (K_T X_t X_t' K_T) \otimes (K_T X_t X_t' K_T) \left(I_p \otimes K_T^{-1}(\hat{\theta} - \theta_0)\right).$$

According to statements (b), (c) and (e) of Lemma 1 the OLS estimator $\hat{\theta}$ satisfies the following equation $(\hat{\theta} - \theta_0)' K_T^{-1} = O_p(1/\sqrt{T})$, while the middle term is bounded in probability. One can prove in a similar way by using statement (f) of Lemma 1 that the third term on the right-hand side of equation (5) is negligible So,

$$K_T \Omega_T K_T = \frac{1}{T} K_T \sum_{t=p+1}^{T} e_t^2 X_t X_t' K_T + O_p(1/\sqrt{T}) \Rightarrow \begin{pmatrix} \sigma^2\omega^2 \int J_c^2 dt & 0 \\ 0 & E x_t x_t' e_t^2 \end{pmatrix},$$

where the last convergence follows from Lemma 1 (d).

Let us now consider case (i) of the linear test with $\frac{K_T A_T}{\|K_T A_T\|} \to a$ and $\|a\| \neq 0$. By the usual logic we get

$$DM_T = \frac{\left(A'(\hat{\theta} - \theta_0)\right)^2}{A'(X'X\Omega_T^{-1}X'X)^{-1}A} = \frac{\left((\frac{K_T A_T}{\|K_T A_T\|})'(K_T X' X K_T)^{-1} K_T X' e\right)^2}{(\frac{K_T A_T}{\|K_T A_T\|})'(K_T X' X K_T)^{-1} K_T \Omega_T K_T (K_T X' X K_T)^{-1} \frac{K_T A_T}{\|K_T A_T\|}}.$$

Then

$$DM_T \Rightarrow \frac{\left(a_1 \frac{\sigma \int_0^1 J_c(t) dW_1(t)}{\omega \int_0^1 J_c^2(t) dt} + a_2' N(0, V)\right)^2}{\frac{\sigma^2}{\int_0^1 J_c^2(t) dt} a_1^2 + a_2' V a_2} = (t(u, c))^2,$$

where $V=(Ex_tx_t')^{-1}E[e_t^2x_tx_t'](Ex_tx_t')^{-1}$, $u=\sqrt{\frac{a_2'Va_2g(c)}{a_1^2}}$. The last expression asymptotically coincides with equation (2) for the local-to-unity case, as in such an embedding the matrix F becomes diagonal.

Now consider case (ii) $H_0:\theta_h=f_h(\theta)=\gamma_0$ where $h=[q\sqrt{T}]$. Denote $J_T=X'X\Omega_T^{-1}X'X$ and $J_{eT}=X'X\Omega_T^{-1}X'e$. Let us consider the first-order condition for the conditional minimization problem, when the DM statistic defined in equation (4) is minimized over $\tilde{\theta}$ such that $f_h(\tilde{\theta})=f_h(\theta_0)$.

$$\begin{pmatrix} J_T & \tilde{A} \\ A^* & 0 \end{pmatrix}\begin{pmatrix} \tilde{\theta}-\theta_0 \\ \lambda \end{pmatrix}=\begin{pmatrix} J_{eT} \\ 0 \end{pmatrix},$$

where $\tilde{A}=\frac{\partial f}{\partial\theta}(\tilde{\theta})$ and $A^*=\frac{\partial f}{\partial\theta}(\theta^*)$, with θ^* being a point between $\tilde{\theta}$ and θ_0 such that $(\tilde{\theta}-\theta_0)'A^*=0$. Following the proof of Lemma 3 from the paper one gets that

$$DM_T=\frac{\left(A^{*\prime}J_T^{-1}J_{eT}\right)^2\widetilde{A}'J_T^{-1}\widetilde{A}}{\left(A^{*\prime}J_T^{-1}\widetilde{A}\right)^2}=\frac{\left(A^{*\prime}(X'X)^{-1}X'e\right)^2\widetilde{A}'(X'X)^{-1}\Omega_T(X'X)^{-1}\widetilde{A}}{\left(A^{*\prime}(X'X)^{-1}\Omega_T(X'X)^{-1}\widetilde{A}\right)^2}.$$

Repeating steps of the proofs of Lemmas 4 and 5 from the paper results in the needed statement.

2 IRFs in VAR with a potential unit root

In this section some results of the paper are generalized to VAR systems in which at most one root is local to unity.

Let us consider a k-dimensional process described by an unrestricted VAR(p) regression:

$$y_t=B_1y_{t-1}+...+B_py_{t-p}+e_t, \tag{6}$$

Imagine for simplicity that we know the co-integrating (near co-integrating) relation and can locate the problematic root. That is, assume that the first component $y_{1,t}$ has a local-to-unity root, while all other components $y_{-1,t}=(y_{2,t},...,y_{k,t})'$ are strictly stationary. Formally, let us assume that the VAR lag polynomial $B(L)=I_k-B_1L-...-B_pL^p$ can be factorized in the following way: $B(L)=(I_k-diag(\lambda,0,...,0)L)\tilde{B}(L)$.

Assumption VAR1

(i) All roots of the characteristic polynomial $\tilde{B}$ lie strictly inside and are bounded away from the unit circle. In particular, the process x_t given by $\tilde{B}(L)x_t = e_t$ can be written as an MA (∞) process $x_t = \Theta(L)e_t = \sum_{j=0}^{\infty} \Theta_j e_{t-j}$ with MA coefficients satisfying the following condition: $\sum_{j=0}^{\infty} j\|\Theta_j\| < \infty$, where $\|\Theta_j\| = \sqrt{trace(\Theta_j \Theta_j')}$.

(ii) Assume that $y_t = \Lambda y_{t-1} + x_t, y_0 = 0$, where $\Lambda = diag(\lambda, ..., 0)$, that is, $y_{1,t} = \lambda y_{1,t-1} + x_{1,t}; y_{-1,t} = x_{-1,t}$. The problematic root λ is local to unity, in particular, $\lambda = \lambda_T = 1 - c/T$.

(iii) Errors e_t are a martingale-difference sequence with respect to sigma-algebra $\mathcal{F}_t$, with $E\left(e_t e_t' | \mathcal{F}_{t-1}\right) = \Omega$ and four finite moments.

The assumption above is a direct generalization of local-to-unity asymptotic embedding to a multivariate setting. If Assumption VAR1 holds, the OLS estimator of regression (6) demonstrates non-standard asymptotic behavior due to some linear combination of coefficients being estimated super-consistently. A survey of local-to-unity multivariate models can be found in Phillips (1988).

We are interested in testing a hypothesis about the coefficients $H_0 : f(B_1, ..., B_p) = 0$, where f is some function of coefficients. The following statistic is a generalization of the LR statistic to a multi-dimensional case:

$$LR = T \cdot trace(\hat{\Omega}^{-1}(\tilde{\Omega} - \hat{\Omega})) \tag{7}$$

with $\Omega(B) = \frac{1}{T}\sum_{t=1}^{T}(B(L)y_t)(B(L)y_t)'$, $\hat{\Omega} = \Omega(\hat{B}), \tilde{\Omega} = \Omega(\tilde{B})$, where $\hat{B}$ is the OLS estimator of coefficients in regression (6), while

$$\tilde{B} = \arg \min_{B=(B_1,...,B_p):f(B)=0} Ttrace(\hat{\Omega}^{-1}(\hat{\Omega} - \Omega(B)))$$

is the restricted estimate.

Consider the hypothesis about the impulse response of the nearly non-stationary series $y_{1,t}$ to $j-$th shock at the horizon h, call it $\theta_h = \frac{\partial y_{1,t+h}}{\partial e_{j,t}}$. We consider the horizon $h = [q\sqrt{T}]$ as increasing proportionally to $\sqrt{T}$. This embedding implies that u_T converges to a constant in the AR(p) case and delivers the mixture of local-to-unity and normal distributions as the limit distribution of $\mathrm{LR}^{\pm}$ statistic. Lemma 4 below points out that the linearized hypothesis about such an impulse response puts $\sqrt{T}-$increasing weight on the coefficients estimated super-consistently when compared with weights on the asymptotically normal coefficients before stationary regressors. Let $\tilde{A} = \frac{\partial \theta_h}{\partial B}$. Let the hypothesis $H_0 : \tilde{A}'B = \gamma_0$ be the linearized version of hypothesis $H_0 : \theta_h = \gamma_0$.

Theorem 2 *Let y_t be $k \times 1$ VAR(p) process satisfying Assumptions VAR1. Assume that the linearized version of hypothesis $H_0 : \theta_h = \frac{\partial y_{1,t+h}}{\partial e_{j,t}} = \gamma_0$ at the horizon $h_T = q\sqrt{T}$ is tested using the statistic defined in equation (7). Then $LR \Rightarrow (t(u,c))^2$ as $T \to \infty$ for some u.*

Theorem 2 states that in multivariate VAR model with at most one local-to-unity root the asymptotic behavior of the LR test statistic for IRF at the horizon proportional to $\sqrt{T}$ is of the same nature as the same statistic for an IRF in the univariate AR(p).

The VAR regression (6) can be linearly transformed to a canonical form in which the non-standard coefficients are separated. Rather than regressing all components of y_t on $(y'_{t-1}, ..., y'_{t-p})'$ as in (6), the canonical-form regression has the following regressors:

$$X_t = (y'_{t-1}, \Delta y_{1,t-1}, y'_{-1,t-2}, \Delta y_{1,t-2}, y'_{-1,t-3}, ..., \Delta y_{1,t-p+1}, y'_{-1,t-p})' = (y_{1,t-1}, \tilde{X}'_t)'.$$

Only the first regressor $y_{1,t-1}$ is a local-to-unity process, while $\tilde{X}_t$ is stationary. Let $Z_t = X'_t \otimes I_k$. The model (6) can be written as $y_t = Z_t \Phi + e_t$, where Φ, a $k^2 p \times 1$ matrix of the coefficients, is a one-to-one linear transformation of VAR coefficients $B_1, ..., B_p$. The first k components of Φ correspond to the non-standard coefficients on the non-stationary regressor $y_{1,t-1}$. The OLS estimator $\hat{\Phi}$ is equal to the linearly transformed OLS estimator of $\hat{B}$, and the same linear transformation applied to $\tilde{B}$ produces the restricted estimator $\tilde{\Phi}$. The linearized hypothesis described in Theorem

2 can be written as $H_0 : A'\Phi = \gamma_0$, where $A = \frac{\partial \theta_h}{\partial \Phi}$. For the proof of Theorem 2 we need the following three lemmas.

Lemma 2 *The LR statistic for a linear hypothesis* $H_0 : A'\Phi = \gamma_0$ *defined in (7) is equal to the Wald statistic defined as*

$$Wald = \frac{\left(A'\left((\sum X_t X_t')^{-1} \otimes I_k\right) \sum (X_t \otimes I_k) e_t\right)^2}{A'((\sum X_t X_t')^{-1} \otimes \hat{\Omega}) A}.$$

Proof of Lemma 2. Let $\hat{e}_t = y_t - Z_t \hat{\Phi}$ be the OLS residuals. We can notice that

$$LR(\Phi) = trace\left(\hat{\Omega}^{-1}\left(2\sum_t \hat{e}_t(\hat{\Phi} - \Phi)' Z_t' + \sum_t Z_t(\hat{\Phi} - \Phi)(\hat{\Phi} - \Phi)' Z_t'\right)\right).$$

According to the OLS moment condition $\sum_t \hat{e}_t' \hat{\Omega}^{-1} Z_t = 0$, so,

$$LR(\Phi) = (\hat{\Phi} - \Phi)'(\sum_t Z_t' \hat{\Omega}^{-1} Z_t)(\hat{\Phi} - \Phi);$$

$$\frac{\partial LR(\Phi)}{\partial \Phi} = -2\sum_t Z_t' \hat{\Omega}^{-1} Z_t (\hat{\Phi} - \Phi).$$

The restricted estimator $\tilde{\Phi}$ is the solution to a system of two equations: the first order condition

$$\left(\sum_t Z_t' \hat{\Omega}^{-1} Z_t\right)(\hat{\Phi} - \tilde{\Phi}) = \mu A,$$

where μ is a Lagrange multiplier, and the restriction $A'\tilde{\Phi} = A'\Phi_0$. Plugging in the solution, one gets

$$LR = (\hat{\Phi} - \tilde{\Phi})'(\sum_t Z_t' \hat{\Omega}^{-1} Z_t)(\hat{\Phi} - \tilde{\Phi}) = \frac{(A'\left(\sum_t Z_t' \hat{\Omega}^{-1} Z_t\right)^{-1} \sum_t Z_t' \hat{\Omega}^{-1} e_t)^2}{A'\left(\sum_t Z_t' \hat{\Omega}^{-1} Z_t\right)^{-1} A}.$$

Since the estimation is performed for the full VAR, that is, regression of all $y_{i,t}$ on the same set of regressors, then $\hat{\Omega}^{-1}$ drops out of the formula for the OLS estimate. Indeed, $\sum_t Z_t' \hat{\Omega}^{-1} Z_t = \sum_t (X_t \otimes I_k)' \hat{\Omega}^{-1} (X_t \otimes I_k) = \sum_t (X_t' X_t) \otimes (\hat{\Omega}^{-1})$. As a result,

$$\left(\sum_t Z_t' \hat{\Omega}^{-1} Z_t\right)^{-1} \sum_t Z_t' \hat{\Omega}^{-1} e_t = \left(\left(\sum_t X_t X_t'\right)^{-1} \otimes \hat{\Omega}\right)\left(\sum_t X_t \otimes (\hat{\Omega}^{-1} e_t)\right) =$$

$$= \left(\left(\sum_t X_t X_t'\right)^{-1} \otimes I_k\right)\left(X_t \otimes \sum_t (e_t)\right) = (\sum Z_t' Z_t)^{-1} \sum Z_t' e_t.$$

This completes the proof of Lemma 2. □

Lemma 3 *Let Assumptions VAR1 be satisfied. Let $w_t = y_{1,t}$ be a one-dimensional random process and $\tilde{X}_t = (x'_{t-1}, ..., x'_{t-p})'$ be a $kp \times 1$ vector process. Also let $W(\cdot)$ be a k-dimensional standard Brownian motion, and let $\omega^2 = \boldsymbol{i}'_1\Theta(1)\Omega\Theta(1)'\boldsymbol{i}_1$ be the long-run variance of the process $x_{1,t}$.*

Then the following convergences hold simultaneously:

(a) $\frac{1}{\sqrt{T}}\sum_{t=1}^{[rT]}(e'_t, x'_t)' \Rightarrow (I_k, \Theta(1)')'\Omega^{1/2}W(r)$;

(b) $\frac{1}{\sqrt{T}}w_{[rT]} \Rightarrow \omega J_c(r) = \int_0^1 e^{(r-s)c}d\tilde{W}(r)$, *where* $\tilde{W}(t) = \frac{1}{\omega}\boldsymbol{i}'_1\Theta(1)\Omega^{1/2}W(t)$ *is a standard Brownian motion;*

(c) $\frac{1}{T}\sum_{t=1}^{T} w_{t-1}e'_t \Rightarrow \omega\int_0^1 J_c(r)dW(r)'\Omega^{1/2}$;

(d) $\frac{1}{T^2}\sum_{t=1}^{T} w_{t-1}^2 \Rightarrow \omega^2\int_0^1 J_c^2(r)dr$;

(e) $\frac{1}{T^{3/2}}\sum_{t=1}^{T} w_{t-1}\tilde{X}_t \rightarrow^p 0$;

(f) $\frac{1}{T}\sum_{t=1}^{T}\tilde{X}_t\tilde{X}'_t \rightarrow^p E[\tilde{X}_t\tilde{X}'_t] = Q_{\tilde{X}}$;

(g) $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\tilde{X}_t \otimes e_t \Rightarrow N(0, Q_{\tilde{X}} \otimes \Omega)$, *and the limit is independent of* $W(\cdot)$.

Proof of Lemma 3. Assumptions about error terms e_t give us the FCLT for $\frac{1}{\sqrt{T}}\sum e_t$ and $\frac{1}{\sqrt{T}}\sum e_t e_{t-j}$ with independent limits. Statements (a) and (g) are results of the Beveridge and Nelson decomposition. The proof is a multi-dimensional analog of that of Theorem 3.2 in Phillips and Solo (1992). Statements (b), (c), (d) and (e) can be proved along the lines of Lemma 3.1 in Phillips (1988) which covers multi-dimensional local-to-unity processes and quantities related to them. Statements (e) and (f) are trivial extensions of Theorem 1 from the main paper to the multi-dimensional case. □

Lemma 4 *Assume that y_t satisfies assumptions VAR1. Assume that a VAR(p) regression is written in the canonical form. Assume that Π are a $k \times 1$ vector of coefficients on the regressor $y_{1,t-1}$ in the canonical VAR. Let $\tilde{\Phi}$ be all coefficients Φ other than Π, that is, $\Phi = (\Pi', \tilde{\Phi}')'$. Let $\tilde{\theta}_h = \frac{\partial y_{1,t+h}}{\partial e_{j,t}}$ denote the impulse response of $y_{1,t}$ to shock $e_{j,t}$ at the horizon h. When $h = q\sqrt{T}$ and T increases to infinity, the following two statements hold:*

(a) $\lambda^{-h}\frac{\partial\tilde{\theta}_h}{\partial\tilde{\Phi}'}$ *converges to a finite constant* $(k^2p-k)\times 1$ *vector;*

(b) $\frac{1}{\sqrt{T}}\lambda^{-h}\frac{\partial\tilde{\theta}_h}{\partial\Pi'}$ *converges to a constant* $k\times 1$ *vector proportional to* $\Theta(1)\mathbf{i}_1$*, where* $\boldsymbol{i}_j$ *is a* $k\times 1$ *vector of zeros with 1 in* j*-th place.*

Proof of Lemma 4. Let $y_t = \sum_{h=0}^{\infty}\tilde{\Theta}_h e_{t-h}$, where $\tilde{\Theta}_h$ is a matrix of impulse responses of y_t to e_{t-h}. According to Lutkepohl (1990),

$$\frac{\partial vec(\tilde{\Theta}_h)}{\partial vec(B_l)} = \sum_{m=0}^{h-1}\tilde{\Theta}'_m\otimes\tilde{\Theta}_{h-m-l}.$$

Given that the regressors X_t of the canonical form are a linear transformation of the regressors $(y_{t-1},...,y_{t-p})$ of the unrestricted VAR, the coefficients $B_1,...,B_p$ are the same linear transformation of the coefficients Φ of the canonical form. It is easy to see that

$$\frac{\partial vec(\tilde{\Theta}_h)}{\partial\Pi} = \sum_{m=0}^{h-1}\left(\tilde{\Theta}'_m\mathbf{i}_1\right)\otimes\tilde{\Theta}_{h-m-1}.$$

Notice that $\frac{\partial y_{i,t+h}}{\partial e_{j,t}} = \mathbf{i}'_i\tilde{\Theta}_h\mathbf{i}_j = \left(\mathbf{i}'_j\otimes\mathbf{i}'_i\right)vec(\tilde{\Theta}_h)$. As a result,

$$\frac{\partial\tilde{\theta}_h}{\partial\Pi} = \left(\mathbf{i}'_j\otimes\mathbf{i}'_1\right)\frac{\partial vec(\tilde{\Theta}_h)}{\partial\Pi} = \sum_{m=0}^{h-1}\left(\mathbf{i_j}'\tilde{\Theta}'_m\mathbf{i}_1\right)\mathbf{i}'_1\tilde{\Theta}_{h-m-1}.$$

Since $x_t = \sum_{j=0}^{\infty}\Theta_j e_{t-j}$ and $y_t = \Lambda y_{t-1}+x_t$, where $\Lambda = diag(\lambda,0,...,0)$, $\lambda = 1-c/T$, we have $\tilde{\Theta}_j = \sum_{k=0}^{j}\Lambda^k\Theta_{j-k}$. Along the lines of Pesavento and Rossi (2006), we arrive at $\mathbf{i}'_1\tilde{\Theta}_m = \lambda^m\mathbf{i}'_1\left(\Theta(1)+o(1)\right),$ as $m\to\infty$, and

$$\frac{\partial\tilde{\theta}_h}{\partial\Pi} = \sum_{m=0}^{h-1}\left(\mathbf{i_j}'\tilde{\Theta}'_m\mathbf{i}_1\right)\mathbf{i}'_1\tilde{\Theta}_{h-m-1} = h\lambda^{h-1}\left((\mathbf{i_1}'\Theta(1)\mathbf{i_j})\mathbf{i}_1\Theta(1)+o(1)\right),$$

as $h = q\sqrt{T}$ and $T\to\infty$. At the same time, the derivative of the same impulse response with respect to any other coefficient will be of order λ^h. For example, let us consider coefficients staying before the regressor $y_{2,t-1}$, call them for example, Γ. One can see that

$$\frac{\partial vec(\tilde{\Theta}_h)}{\partial\Gamma} = \sum_{m=0}^{h-1}\left(\tilde{\Theta}'_m\mathbf{i}_2\right)\otimes\tilde{\Theta}_{h-m-1},$$

and correspondingly

$$\frac{\partial\tilde{\theta}_h}{\partial\Gamma} = \sum_{m=0}^{h-1}\left(\mathbf{i_2}'\tilde{\Theta}_m\mathbf{i}_j\right)\mathbf{i}'_1\tilde{\Theta}_{h-m-1} = \sum_{m=0}^{h-1}\left(\mathbf{i_2}'\Theta_m\mathbf{i}_j\right)\lambda^{h-m-1}\left(\mathbf{i}_1\Theta(1)+o(1)\right).$$

Assume that $\mu_1, ..., \mu_{k^2p-1}$ are roots of the process x_t, for large enough T they are all smaller in absolute value than $\lambda = 1 - c/T$. There exists a set of constants $C_1, ..., C_{k^2p-1}$ such that $\mathbf{i_2}'\Theta_m \mathbf{i}_j = \sum_{l=1}^{k^2p-1} C_l \mu_l^h$ for any horizon h. This gives us that $\lambda^{-h}\frac{\partial\tilde{\Theta}_{1j,h}}{\partial\Gamma}$ converges to a constant as $h \to \infty$. □

Proof of Theorem 2. Let $A = A_T = \lambda^{-h}\frac{\partial\theta_h}{\partial\Phi'}$ and the linearized version of the hypothesis about impulse response be $H_0 : A_T'\Phi = A_T'\Phi_0$. We introduce the following notation: $A_T = \sqrt{T}A_{1,T} + A_{2,T}$, where $A_{1,T} = (a_{1,T}', 0, ..., 0)'$, and $A_{2,T} = (0, ..., 0, a_{2,T}')'$. According to Lemma 4, as $T \to \infty$ both vectors converge to some constant vectors $a_1 = \lim_{T\to\infty} a_{1,T}$ and $a_2 = \lim_{T\to\infty} a_{2,T}$, and $a_1 = C\Theta(1)\mathbf{i}_1$ for some constant C. Let us introduce normalization matrix $D^* = \begin{pmatrix} \frac{1}{T} & 0 \\ 0 & \frac{1}{\sqrt{T}}I_{kp-1} \end{pmatrix}$ and $D = D^* \otimes I_k$, then

$$LR = \frac{\left((\sqrt{T}DA)'\left((D^*\sum X_t'X_tD^*)^{-1} \otimes I_k\right)\sum(D^*X_t \otimes I_k)'e_t\right)^2}{(\sqrt{T}DA)'((D^*\sum X_t'X_tD^*)^{-1} \otimes \hat{\Omega})(\sqrt{T}DA)}.$$

Lemma 3 implies that

$$D^*\sum_t X_t'X_tD^* \Rightarrow \begin{pmatrix} \omega^2\int_0^1 J_c^2(r)dr & 0 \\ 0 & Q_{\tilde{X}} \end{pmatrix},$$

Obviously, $\sqrt{T}DA \to (a_1', a_2')'$. So, the denominator is:

$$(\sqrt{T}DA)'(D\sum Z_t'\hat{\Omega}^{-1}Z_tD)^{-1}\sqrt{T}DA \Rightarrow (a_1'\Omega a_1)\frac{1}{\omega^2\int_0^1 J_c^2(r)dr} + a_2'\left(Q_{\tilde{X}}^{-1} \otimes \Omega\right)a_2.$$

Given that $a_1 = C\Theta(1)\mathbf{i}_1$, we have $a_1'\Omega a_1 = C^2\omega^2$.

As for the numerator, we have the following:

$$(\sqrt{T}DA)'\left((D^*\sum X_t'X_tD^*)^{-1} \otimes I_k\right)\sum(D^*X_t \otimes I_k)'e_t \Rightarrow$$
$$\frac{\omega\int_0^1 J_c(r)dW(r)'\Omega^{1/2}a_1}{\omega^2\int_0^1 J_c^2(t)dt} + N(0, a_2'\left(Q_{\tilde{X}}^{-1} \otimes \Omega\right)a_2).$$

We notice that

$$\frac{\omega\int_0^1 J_c(r)dW(r)'\Omega^{1/2}a_1}{\omega^2\int_0^1 J_c^2(t)dt} = \frac{C\omega\int_0^1 J_c(r)dW(r)'\Omega^{1/2}\Theta(1)\mathbf{i}_1}{\omega^2\int_0^1 J_c^2(t)dt} =$$
$$= \frac{C\omega^2\int_0^1 J_c(r)d\tilde{W}(r)}{\omega^2\int_0^1 J_c^2(t)dt} = Ct^c\frac{1}{\sqrt{\int_0^1 J_c^2(t)dt}}.$$

So,

$$LR \Rightarrow \frac{\left(\frac{C}{\sqrt{\int_0^1 J_c^2(t)dt}} t^c + \sqrt{A_2'\left(Q_{\tilde{X}}^{-1} \otimes \Omega\right) A_2} \cdot N(0,1)\right)^2}{\frac{C^2}{\int J_c^2 dr} + A_2'\left(Q_{\tilde{X}}^{-1} \otimes \Omega\right) A_2} = (t(c,u))^2,$$

where $u = \frac{\sqrt{A_2'\left(Q_{\tilde{X}}^{-1} \otimes \Omega\right) A_2}}{C}$.

3 Wald statistic for IRF in AR(p)

The paper shows that while the LR statistic for highly non-linear IRFs is well approximated by the same family of distributions as the LR statistic for the linear hypothesis, the same does not hold for the Wald statistic. The paper presented an AR(1) example. The same idea can be applied to higher-order processes as well.

Let the data follow an AR(1) process $y_t = \rho y_{t-1} + e_t$, and we treat it as an AR(2) process $y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + e_t$ with $\phi_1 = \rho, \phi_2 = 0$. Assume that we estimate AR(2) coefficients by OLS, and calculate the estimated AR(2) roots $\hat{\mu}$ and $\hat{\lambda}$. We abstract from the unit root problem here, and assume that $0 < \rho < 1$ fixed as $T \to \infty$, then $\hat{\mu} \to^p \rho$, $\hat{\lambda} \to^p 0$, and both roots are $\sqrt{T}$ asymptotically normal.

The theoretical impulse response is $\theta_k = \rho^k$, while the estimated impulse response is $\hat{\theta}_k = \frac{\hat{\mu}^{k+1} - \hat{\lambda}^{k+1}}{\hat{\mu} - \hat{\lambda}}$. In order to calculate the t-statistic we also need the derivatives of the impulse response.

$$\frac{\partial \theta_k}{\partial \phi_1}(\phi_1, \phi_2) = \frac{\partial \theta_{k+1}}{\partial \phi_2} = \sum_{j=0}^{k-1} \theta_j \theta_{k-j-1}.$$

In our case we need the derivative to be calculated at the estimated coefficients

$$\begin{aligned}\frac{\partial \theta_k}{\partial \phi_1}(\hat{\phi}_1, \hat{\phi}_2) &= \frac{1}{(\hat{\mu} - \hat{\lambda})^2} \sum_{j=0}^{k-1} (\hat{\mu}^{j+1} - \hat{\lambda}^{j+1})(\hat{\mu}^{k-j} - \hat{\lambda}^{k-j}) = \\ &= \frac{1}{(\hat{\mu} - \hat{\lambda})^2} \left((k+2)\hat{\mu}^{k+1} + (k+2)\hat{\lambda}^{k+1} - 2\frac{\hat{\mu}^{k+2} - \hat{\lambda}^{k+2}}{\hat{\mu} - \hat{\lambda}} \right).\end{aligned}$$

If we consider a sequence of hypotheses with a growing horizon $k_T = \sqrt{T}$, then

$$\frac{1}{k} \hat{\mu}^{-k-1} \frac{\partial \theta_k}{\partial \phi_1}(\hat{\phi}_1, \hat{\phi}_2) \to^p 1.$$

So, in the described setting we have $t = \frac{\rho^k - \frac{\hat{\mu}^{k+1} - \hat{\lambda}^{k+1}}{\hat{\mu} - \hat{\lambda}}}{s.e.(\hat{\theta}_k)}$, and we showed that along the sequence $k_T = \sqrt{T}$ we have $s.e.(\hat{\theta}_k) = k\hat{\mu}^k(const + o_p(1))$. As a result, the asymptotic behavior of the t-statistic is defined by the behavior of the ratio $\frac{\rho^k - \hat{\mu}^k}{\hat{\mu}^k}$, which is of the same type as for the AR(1) case described in the paper.

4 Simplified formula for u for IRFs in AR(2)

This section provides a more explicit formula for parameter u defined in (2) for the IRFs in an AR(2) model. This formula was used to construct Table 1 in the main paper.

Imagine that we have an AR(2) process with roots λ and μ: $(1-\lambda L)(1-\mu L)y_t = e_t$. The process can be alternatively written as

$$y_t = \rho y_{t-1} + \alpha \Delta y_{t-1} + e_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + e_t$$

where $\phi_1 = \alpha + \rho, \phi_2 = -\alpha, \alpha = \lambda\mu, \rho = \lambda + \mu - \lambda\mu$. As in paper, let $X_t = (y_{t-1}, \Delta y_{t-1})$ and $\Sigma(\rho, \alpha) = EX_tX_t'$. There is a lower-triangular matrix F such that $F\Sigma(\rho,\alpha)F' = I_2$.

Let θ_h be the impulse response at horizon h, and $A = \frac{\partial}{\partial(\rho,\alpha)}\theta_h$ be its derivative. As it can be seen u is a function of ρ, α, h.

$$\Sigma(\alpha, \rho) = \gamma_0 \begin{pmatrix} 1 & 1 - r_1 \\ 1 - r_1 & 2(1 - r_1) \end{pmatrix} = \gamma_0 \begin{pmatrix} 1 & \frac{1-\rho}{1+\alpha} \\ \frac{1-\rho}{1+\alpha} & 2\frac{1-\rho}{1+\alpha} \end{pmatrix}$$

where $\gamma_0 = Var(y_t)$, and r_1 is the first-order correlation. According to Hamilton ([3.4.27] on p. 58), $r_1 = \frac{\phi_1}{1-\phi_2} = \frac{\alpha+\rho}{1+\alpha}$. One can check that

$$F = \sqrt{\gamma_0} \begin{pmatrix} 1 & 0 \\ -\sqrt{\frac{1-\rho}{1+2\alpha+\rho}} & \frac{1+\alpha}{\sqrt{(1-\rho)(1+2\alpha+\rho)}} \end{pmatrix} = \sqrt{\gamma_0} \begin{pmatrix} 1 & 0 \\ -\sqrt{\frac{(1-\lambda)(1-\mu)}{(1+\lambda)(1+\mu)}} & \frac{1+\lambda\mu}{\sqrt{(1-\lambda^2)(1-\mu^2)}} \end{pmatrix}$$

Lütkepohl (1990) showed that

$$\frac{\partial}{\partial \phi_1}\theta_h = \sum_{m=0}^{h-1} \theta_m \theta_{h-m-1}; \quad \frac{\partial}{\partial \phi_2}\theta_h = \sum_{m=0}^{h-2} \theta_m \theta_{h-m-2}.$$

Let us denote $A_h = \frac{\partial}{\partial \phi_1}\theta_h$, then $\frac{\partial}{\partial \phi_2}\theta_h = A_{h-1}$. Since $\theta_h = \frac{\lambda^{h+1}-\mu^{h+1}}{\lambda-\mu}$ ([2.4.14] in Hamilton). We can see that

$$A_h = \sum_{m=0}^{h-1} \frac{(\lambda^{m+1}-\mu^{m+1})(\lambda^{h-m}-\mu^{h-m})}{(\lambda-\mu)^2} = \frac{1}{(\lambda-\mu)^2}\left((h+2)\lambda^{h+1} + (h+2)\mu^{h+1} - 2\frac{\lambda^{h+2}-\mu^{h+2}}{\lambda-\mu}\right)$$

Since $\phi_1 = \alpha + \rho, \phi_2 = -\alpha$, we have

$$\frac{\partial}{\partial \rho}\theta_h = A_h, \frac{\partial}{\partial \alpha}\theta_h = A_h - A_{h-1}$$

So, our vector of derivatives is $A = (A_h, A_h - A_{h-1})$. According to formula (2):

$$u = \left| \frac{-\sqrt{\frac{(1-\lambda)(1-\mu)}{(1+\lambda)(1+\mu)}}A_h + \frac{1+\lambda\mu}{\sqrt{(1-\lambda^2)(1-\mu^2)}}(A_h - A_{h-1})}{A_h} \right| =$$

$$= \frac{1}{\sqrt{(1-\lambda^2)(1-\mu^2)}} \left| \frac{(\lambda+\mu)A_h - (1+\lambda\mu)A_{h-1}}{A_h} \right|.$$

5 References

Andrews, D. W. K., 1988. "Laws of Large Numbers for Dependent Non-identically Distributed Random Variables," *Econometric Theory* 4, 458-467.

Andrews, D.W. K. and P. Guggenberger, 2008. "Asymptotics for LS, GLS, and Feasible GLS Statistics in an AR(1) Model with Conditional Heteroskedaticity," *Cowles Foundation Discussion Paper* No 1665.

Andrews, D.W. K. and P. Guggenberger, 2010. "Asymptotic Size and a Problem with Subsampling and with the m out of n Bootstrap," *Econometric Theory* 26 (2), 426-468.

Hamilton, J.D. (1994): "Time Series Analysis," *Princeton University Press*

Hansen, B. E., 1992: "Convergence to Stochastic Integrals for Dependent Heterogeneous Processes," *Econometric Theory* 8, 489-500.

Lütkepohl, H. (1990): "Asymptotic Distributions of Impulse Response Functions and Forecast Error Variance Decompositions of Vector Autoregressive Models," *Review of Economics and Statistics*, 72(1), 116-125.

Pesavento, E. and B. Rossi (2006): "Small Sample Confidence Intervals for Multivariate Impulse Response Functions at Long Horizons," *Journal of Applied Econometrics* 21(8), 1135-1155.

Phillips P.C.B. 1988. "Regression Theory for Near-Integrated Time Series," *Econometrica*, 56 (5), pp. 1021-1043.

Phillips, P.C.B. and V. Solo, (1992)"Asymptotics for Linear Processes," *Annals of Statistics*, 20(2), 971-1001.