Estimation of causal effects in macroeconometrics

Fabio Canova
BI Norwegian Business School, CAMP, and CEPR
September 2022

References

Altavilla, C., Canova, F. and M. Ciccarelli (2020). Mending the broken link: Bank lending rates and unconventional monetary policy. Journal of Monetary Economics,110, 81-98.

Angelini, G. and L. Fanelli, (2019). Exogenous uncertainty and the identification of Structural Vector Autoregressions with external instrument. Journal of Applied Econometrics, 34, 951-971.

Antolin, J. and J. Rubio (2019). Narrative sign restrictions, American Economic Review, 108, 2802-2829.

Athey, S. and G. Imbens (2017). The state of applied econometrics. Causality and policy evaluation. Journal of Economic Perspective, 31, 3-32.

Barnichon, R. and C. Brownlees (2019).Impulse Response Estimation by Smooth Local Projections. Review of Economics and Statistics, 101, 522-530.

Canova, F. and F. Ferroni (2020) Mind the gap! Dynamic stylized facts and structural models, forthcoming American Economic Journals: Macroeconomics.

Faust, J., Rogers, J.H., Swanson, E., and J.H. Wright (2003).Identifying the Effects of Monetary Policy Shocks on Exchange Rates Using High Frequency Data. Journal of the European Economic Association, 1, 1031-57.

Frisch, R. (1933). Propagation problems and impulse problems in dynamic economics, in: Economic Essays in Honor of Gustav Cassel. Allen & Unwin, London, 171-205.

Gorodnichenko, Y. and B. Lee (2020) A Note on the variance decomposition with local projections. Journal of Business and Economic Statistics, 38, 921-933.

Gertler, M. and P. Karadi (2015).Monetary Policy Surprises, Credit Costs, and Economic Activity. American Economic Journal: Macroeconomics, 7, 44-76.

Gurkaynak, R.S., Sack, B., and E. Swanson (2005). The sensitivity of long-term interest rates to economic news: Evidence and implications for macroeconomic models. American Economic Review, 95, 425-436.

Herbst, E. and Johannsen, B. (2021) Bias in local projections. Federal Reserve Board, discussion paper 2020-010

Jentsch, C., and K. G. Lunsford (2019) The Dynamic Effects of Personal and Corporate Income Tax Changes in the United States: Comment. American Economic Review, 109: 2655-78.

Jorda, O. (2005) Estimation and Inference of Impulse Responses by Local Projections. American Economic Review, 95, 161-182.

Kim, Y.J. and L. Kilian (2011). How Reliable are Local Projection Estimators of Impulse Responses? The Review of Economics and Statistics, 93, 1460-1466

Kleibergen, F. (2005). Testing Parameters in GMM without Assuming They Are Identified. Econometrica, 73, 1103-1123.

Kuttner, K.N. (2001). Monetary policy surprises and interest rates: Evidence from the Fed funds futures market. Journal of Monetary Economics, 47, 523-544.

Li, D., M. Plogborg-Moller, C. Wolf (2021) Local projections vs VARs: Lessons from thousands of DGPs, manuscript

Mertens, K. and Ravn, M. O. (2013). The Dynamic Effects of Personal and Corporate Income Tax Changes in the United States. American Economic Review, 103, 1212-47.

Montiel Olea J. and Plogborg-Moller, M. (2021) Local Projection Inference is simpler and more robust than you think, Econometrica, 89, 1789-1823.

Plagborg-Moller, M. and C. Wolf (2018). Instrumental variable identification of variance decompositions. Forthcoming, Journal of Political Economy.

Plagborg-Moller, M. and C. Wolf (2020).Local projections and VARs estimate the same impulse responses. Econometrica, 89, 955-980.

Ramey, V. (2016).Macroeconomic Shocks and their Propagation, in Handbook of Macroeconomics, Vol. 2A. Amsterdam: Elsevier, 71-162.

Ricco, G. and Miranda Agrippino, S. (2020). The transmission of Monetry Policy shocks. American Economic Journals: Macroeconomics, 13, 74-107.

Romer, C. D. and D.H. Romer (2004). A New Measure of Monetary Shocks: Derivation and Implications. American Economic Review, 94, 1055-1084.

Sims, C.A. (1980). Macroeconomics and Reality. Econometrica 48, 1-48.

Slutsky, E. (1937). The summation of random causes as the source of cyclic processes. Econometrica, 5, 105-146.

Stock, J. and M. Watson (2018).Identification and Estimation of Dynamic Causal Effects in Macroeconomics Using External Instruments. Economic Journal, 128, 917-947.

Introduction

• VARs approximate the data linearly by finding, given a lag length, the best 1 -step ahead prediction error. The approximation is global.

• IRF are nonlinear functions of estimated VAR coefficients.

• Local projections (LP) directly approximate IRF at each horizon (local rather than global approximation). Could be useful if true IRF are not smooth.

• Features of LP:
- single equation OLS/IV is enough (with qualifications).
- robust to parametric misspecification.
- potential to make IRFs non-linear.
- lag length of projection (and instruments) may depend on horizon.

can choose lag length for every time horizon

Average causal effects in microeconometrics

• Computation of IRFs in macro similar to computation of average causal effects (ACE) in microeconometrics.

• In micro, ACE are $E(Y|X=1) - E(Y|X=0)$, where X is a binary control (e.g. medicine intake), **randomly assigned across individuals**, Y an outcome variable (e.g. disease), and the average is across individuals in each group (those who take the medicine and those who do not).

• ACE are assessed by the magnitude and significance of β in:

$$Y = \beta X + u \quad E(u|X) = 0 \tag{1}$$

• If there are covariates, we need to condition on $W = w$ (group the data so that is the case) and the ACE is β in the regression

$$Y = \beta X + \gamma W + u \quad E(u|X, W) = 0, \tag{2}$$

• If X is endogenous (likely case), define $X^* = X - E(X|W)$, $u^* = u - E(u|W)$ and consider instruments $Z^* = Z - E(Z|W)$. We need
1) Instrument Relevance $E(X^{*'}Z^*) = \alpha' \neq 0$
2) Instrument Exogeneity $E(u^{*'}Z^*) = 0$,

• If the instrument are "strong" (i.e. α is large), X, Z are scalars, $\beta_{IV} = (Z^*X^*)^{-1}(Z^*Y)$, is consistent in (2).

• If instruments are "weak", need different asymptotics.

• Usual problem: What is a good Z? How do we make sure they are strong?

• Good general reference: Athey and Imbens (2017).

Dynamic average causal effects in macroeconometrics

• In macro, we care about **dynamic** average causal effects (DACE), which are measured through interventions that propagates over time (see Frisch, 1933, Slutzky, 1937). Consider the MA model:

$$y_t = \theta(L)\epsilon_t \tag{3}$$

ϵ_t is a $N_e \times 1$ vector of structural shocks, iid over $i = 1, \ldots N_e$ and time, with zero mean, constant variance, and $\theta(L) = \theta_0 + \theta_1 L + \theta_2 L^2 + \ldots$.

• DACE at horizon $h = 0, 1, \ldots$ is estimated via the dynamic counterfactual

$$E(Y_{i,t+h}|\epsilon_{jt} = 1) - E(Y_{i,t+h}|\epsilon_{j,t} = 0) = \theta_{ij}^h = \frac{\partial Y_{i,t+h}}{\partial \epsilon_{j,t}} \tag{4}$$

Average of periods when I do have the shock, average of periods when I do not have a shock.

$i = 1, 2 \ldots, N_y, h = 1, \ldots, H,\ j = 1, \ldots, N_\epsilon$. As in micro setups, if $T \to \infty$ and $\epsilon_{t,j}$ is observable, (4) could be estimated with

$$Y_{i,t+h} = \theta_{ij}^h \epsilon_{j,t} + u_{i,t+h} \tag{5}$$

• **In macro there are no randomized controlled experiment.** We can think of doing a randomized experiment by drawing a shock $\epsilon_{j,t}$ that perturbs the economy from the steady state at some t.

• **The average is constructed across episodes where $\epsilon_{jt} = 1$ and $\epsilon_{jt} = 0$** (rather than across individuals).

• $\epsilon_{j,t}$ **must be independent of** $u_{i,t+h}$ (this is the definition of a structural shock).

• $\epsilon_{j,t}$**, in general, is non-observable.**

We assume we don't have shock normally, and then suddenly, completely randomly there's an unexpected one.

Two ways of computing DACE

- VAR approach (Sims, 1980).

- Setup and estimate a VAR(q): $Y_t = A(L)Y_{t-1} + e_t$.

- Invert the system and linearly relate VAR innovations e_t to structural shocks ϵ_t, i.e. $Y_t = \theta(L)e_t, e_t = R\epsilon_t, \theta(L) = (1 - A(L))^{-1}$.

- Impose identification restrictions (typically on R or $\theta(1)R, via zeros, signs, etc.$) to identify ϵ_t. Estimate structural responses using $\theta(L)R$.

- Two step method. Consistent estimation of VAR parameters and IRFs.

- Local Projections (Jorda, 2005)

- Assume a linear system linking Y_t and ϵ_t (equation (5)).

- Project Y_{t+h} on ϵ_{jt}, for each $h = 1, 2, \ldots$, some j.

- Need to correct standard errors of the estimates of θ^h for the presence of MA components in u_{t+h} (i.e. use a HAC stardard errors).

- Trade-off 1: Local projections non-parametric. VAR is parametric; could be misspecified.

- Trade-off 2: If the DGP is a VAR, local projections inefficient (there is a relationship between θ^h and θ^{h-1}).

General idea of LP

- The idea comes from a "direct forecasting" methodology.

- Suppose $Y_{t+h} = AY_{t+h-1} + e_{t+h}$; e_{t+h} iid $(0, \Sigma)$, Σ general matrix; A could be the companion form of a VAR(q).

- Repeatedly substituting backward

local projection is just (6)

$$
\begin{aligned}
Y_{t+h} &= A^{h+1}Y_{t-1} + A^{h}e_t + A^{h-1}e_{t+1} + \ldots Ae_{t+h-1} + e_{t+h} \\
&= A^{h+1}Y_{t-1} + (A_1^{h}e_{1t} + A_{-}^{h}e_{-1,t}) + A^{h-1}e_{t+1} + \ldots Ae_{t+h-1} + e_{t+h}
\end{aligned} \tag{6}
$$

^ iid and uncorrelated

where e_{1t} is the shock of interest and $e_{-1,t}$ is the vector of all other shocks.

- If $\Sigma = \text{diag}\{\sigma_i\}$, $e_{1t} = \epsilon_{1t}$, the response of Y_{t+h} to a ϵ_{1t} impulse is

$$E(Y_{t+h}|\epsilon_{1t} = 1) - E(Y_{t+h}|\epsilon_{1t} = 0) = A_1^h \tag{7}$$

Thus, A^h in (7) can be estimated via the regression

$$Y_{t+h} = A^{h+1}Y_{t-1} + A_1^h\epsilon_{1t} + v_{t+h} \tag{8}$$

where $v_{t+h} = f(\epsilon_{-1,t}, e_{t+1}, \ldots, e_{t+h-1}, e_{t+h})$ and $f(.)$ is linear.

- If $\Sigma \neq \text{diag}\{\sigma_i\}$, the response of Y_{t+h} to an impulse in ϵ_{1t} is

$$E(Y_{t+h}|\epsilon_{1t} = 1) - E(y_{t+h}|\epsilon_{1t} = 0) = B_1^h \tag{9}$$

where $B^h = A^hR$, and $e_t = R'\epsilon_t$, for some square matrix R. Thus, B^h in (9) can be estimated via the regression

$$Y_{t+h} = A^{h+1}Y_{t-1} + B_1^h\epsilon_{1t} + u_{t+h} \tag{10}$$

where $u_{t+h} = f(\epsilon_{-1,t}, e_{t+1}, \ldots, e_{t+h-1}, e_{t+h})$ and $f(.)$ is linear.

• **Problem:** since ϵ_{1t} is unobservable, there is an identification problem: for any invertible matrix T such that $B^h\epsilon_t = (B^hT)(T^{-1}\epsilon_t) = D^h v_t$. How do remedy to the non-observability of ϵ_{1t}?

• Traditional approaches (Kuttner, 2001, Faust, et al., 2003, Gurkaynak et al., 2005, Gertler and Karadi, 2015). Substitute ϵ_{1t} with a proxy m_t.

proxy identification

$$Y_{t+h} = \beta^h Y_{t-1} + \gamma^h m_t + u_{t+h} \tag{11}$$

where m_t comes from external information (e.g. high frequency data, announcement dates etc.) and "directly" captures ϵ_{1t}.

• If m_t is exogenous (can we make such an assumption?), can omit Y_{t-1} from (11).

• Narrative approaches (e.g. Romer and Romer, 2004, Ramey, 2016). Similar setup: plug in m_t from narrative introspection for ϵ_{1t}.

with proxy standard error is incorrect...

- Extended local projections:

$$Y_{t+h} = \delta^h X_t + \gamma^h m_t + u_{t+h} \tag{12}$$

where $X_t = (Y_{t-1}, W_t, W_{t-1})$, W_t are additional variables not in Y_t: factors, expectations, news, VIX, etc.

- **Adding X_t useful to make m_t and u_{t+h} independent.**

- Can use a lot more variables in X_t than those appearing in a VAR.

Alternative to proxies

• Alternative: (normalization) $B^0_{11} = 1$. This means, e.g. that a 100 basis points monetary policy shock is such that the nominal interest rate increases by 100 basis points; or that a government spending shock increases expenditure by 1 percent of GDP.

• If $B^0_{11} = 1$, one can run local projections, $h = 1, 2, \ldots$:

$$Y_{t+h} = \beta^h Y_{t-1} + \gamma^h Y_{1t} + u_{t+h} \tag{13}$$

i.e. rather than using a (non-observable) shock can use an (observable) variable in the regression. Still Y_{1t} is a proxy for ϵ_{1t}.

• Can estimate γ^h by OLS as long as (Y_{t-1}, Y_{1t}) are uncorrelated with e_{t+p}, $p = 1, 2, \ldots, h$ and $\epsilon_{-1,t}$.

• In general, the second restriction does not hold (nominal interest rate may be responding contemporaneously to shocks other than monetary policy).

Advantages of LP over VARs

• No need to have the correct A (typical problem with a small scale VAR).

• No need for the linear model to span the DGP information set (typical problem if states are omitted from the VAR or the VAR is of smaller dimension as the DGP).

• Can be obtained variable by variable and horizon by horizon (single equation regressions).

• Conditioning variables X_t may be different for different h.

• If ϵ_{1t} is observable, LP has good properties (Kilian and Kim, 2011).

• Variance decomposition estimates are also OK. (Gorodnichenko and Lee, 2020; Plagborg-Moller and Wolf, 2018). Could be improved by bootstrap corrections, if sample is short.

Problems of LP

• m_t is a proxy for ϵ_{1t} (e.g. changes in federal fund futures around policy announcement dates fail to account that monetary policy is made public also via speeches at different dates). OLS may not have the right asymptotic properties in (11)-(12).

• OLS in (13) is **inconsistent** if Y_{1t} is endogenous (e.g. Y_{1t} depends on $\epsilon_{-1,t}$).

• How do we deal with these problems?

LP-IV

- Suppose we have variables Z_t satisfying
- $E(\epsilon^*_{1t} Z^*_t) \neq 0 = \alpha$ (relevance)
- $E(\epsilon^*_{-1,t} Z^*_t) = 0$ (contemporaneous uncorrelation)
- $E(\epsilon^*_{t+h} Z^*_t) = 0, \ h > 0$ (lead uncorrelation)

where $\epsilon^*_{1t} = \epsilon_{1t} - P(\epsilon_{1t}|X_t)$; $Z^*_t = Z_t - P(Z_t|X_t)$.

Then γ^h in (13) can be estimated as

$$\hat{\gamma}^h = \frac{E(Y_{t+h}Z_t)\Omega E(Z_t Y_{1t})}{E(Y_{1t}Z_t)\Omega E(Z_t Y_{1t})} \tag{14}$$

for any positive definite matrix Ω. If $\Omega = \sigma_u^2 E(Z_t' Z_t)^{-1}$, $\hat{\gamma}^h$ is the same as the one obtained running the two stage regression

1st stage makes errors cleaner

Use proxy as an instrument.

$$\begin{aligned} Y_{1t} &= \alpha_0 + \alpha_1 Z_t + \zeta_{1t} \\ Y_{t+h} &= \delta^h X_t + \gamma^h \hat{Y}_{1t} + \beta(h) Y_{t-1} + u_{t+h} \end{aligned} \tag{15}$$

a1 has to be different from zero (use F-test)

HAC standard errors have to be used

often bootstrapped

A few details

• If X_t are excluded from (15), $u_{t+h} = f(e_{t-q}, \ldots, e_{t-1}, \epsilon_{-1,t}, e_{t+1}, \ldots, e_{t+h-1}, e_{t+h})$. Thus, to have proper instruments we also need the condition $E(e^*_{t-k} Z^*_t) = 0$, $k > 0$. **Without X_t, instruments have to be exogenous** Much stronger requirement.

• Efficiency of the IV estimator can be improved by taking into account that u_{t+h} has a MA structure. Use HAC/HAR standard errors.

• If Z_t are weak (i.e. α is small), need to use weak instrument IV regression methods (see Kleiberger, 2005).

• Impulse responses to news shock: if e_{1t} has effect on Y_{1t} only after say k periods, the normalization should be $B^k_{11} = 1$ and the instruments should be relevant for Y_{1t+k} (not for Y_t).

Improvements

- Since LP are estimated separately for each h, $\hat{\gamma}^h$ is typically jagged. One possibility is to estimate (15) jointly for all h, since u_{t+h} is correlated across horizons.

- Penalized (ridge) estimation (set $Y_{t-1} = 0$):

$$\min(Y_{t+h} - \delta^h X_t - \gamma^h Y_{1t})(Y_{t+h} - \delta^h X_t - \gamma^h Y_{1t})' \tag{16}$$

subject to $(\gamma_h - \gamma_{h-1})^2 \leq \zeta$. Use λ as the Lagrange multiplier on the constraints; if large, it forces smoothness on γ^h.

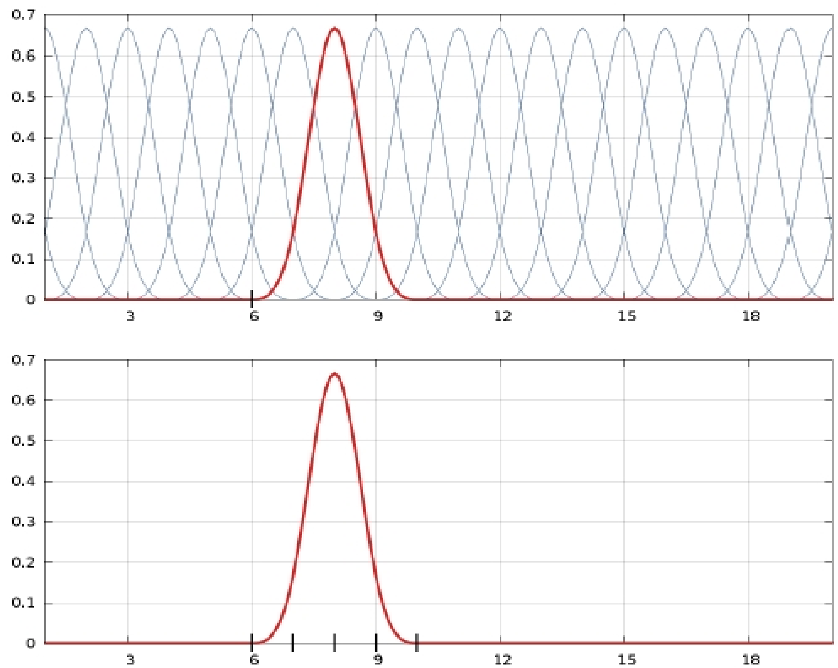
mistake here: gamma in parentheses squared should be delta

• Spline restrictions (Barnichon and Bronwlees, 2019): Let $\chi^h = (\delta^h, \gamma^h)$. Assume $\chi^h = \sum_k \beta_k B(k)^h$, where $B(k)^h$ are B-splines (orthgonal basis).

- A B-spline is made up of $q + 1$ polynomial pieces of order q. The polynomial pieces join on a set of q $+$ 2 inner knots and are calibrated so that derivatives up to the order q -1 are continuous at the inner knots.

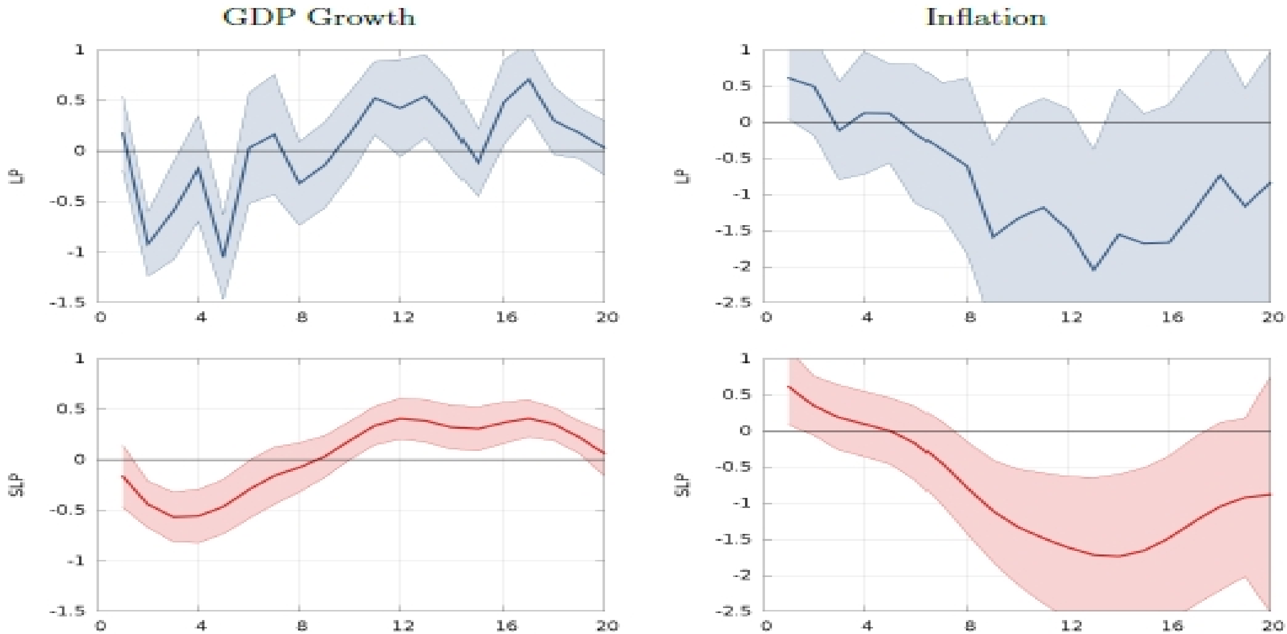
- A B-splines basis function is nonzero over the domain spanned by the q $+$2 inner knots and zero elsewhere

Figure 1: B-SPLINE BASIS



The figure shows the graph of the B-splines basis functions. The top panel displays the set of B-splines basis used in this work. The bottom panel shows in detail the B-splines basis of knot 6.

Figure 9: IR TO A MONETARY SHOCK USING INSTRUMENTAL VARIABLES



The figure displays the IR of GDP growth (left panels) and inflation (right panels) to a monetary shock identified using instrumental variables and estimated using LP (top panels) and SLP (bottom panels). The shaded area denotes the 90% confidence interval.

Computed with splines, improving jaggedness. Easier to interpret.

Mean squared error is much better in penalized regression, however trade-off with bias and variance, more about that will come later. (in slide 34)

IV-SVARs

• Standard setup: $Y_t = A(L)Y_{t-1} + e_t$.

• Estimate A(L) by OLS assuming some lag length and deciding the variables to be included in Y_t.

• Assume $A(L)$ is invertible (i.e. compute $B(L) = (1 - A(L))^{-1}$).

• Impose restrictions, identify shocks, i.e. estimate R in $e_t = R\epsilon_t$. Compute B^h using estimates of A(L) and R

• Parametric assumptions imply that B^h are restricted across h.

SVAR-IV

• Standard Cholesky identification approach given an "internal" IV approach: the VAR innovations in equation i are instrumented by the VAR innovations in equation $i-1, i-2, \ldots$.

• Earlier external SVAR-IV setup(see Ramey, 2016). Add the (external) instrument to the VAR and order it first in a Cholesky decomposition. Here instruments are predetermined not exogenous.

• In analogy with LP-IV, one can also use external instruments to estimate R. Let Z_t satisfy $E(\epsilon_{1t}Z_t') = \alpha' \neq 0$; $E(\epsilon_{-1,t}Z_t') = 0$.

• In terms of VAR residuals e_t these conditions imply

$$E(e_t Z_t') = R \begin{pmatrix} \alpha' \\ 0 \end{pmatrix} = \begin{pmatrix} R_{11}\alpha' \\ R_{.1}\alpha' \end{pmatrix} \tag{17}$$

If $R_{11} = 1$ (normalization), then $R_{i1}, i = 2, \ldots, N$ can be obtained by IV

$$\hat{R}_{i1} = \frac{E(e_{it}Z_t)\Omega E(Z_t e_{1t})}{E(e_{1t}Z_t)\Omega E(Z_t e_{1t})} \tag{18}$$

where again Ω is any positive definite matrix.

In general, one can think of running the following regression:

$$e_{it} = R_{i1}\epsilon_{1t} + u_{it} \tag{19}$$

• Since ϵ_t are not observables. Two approaches are possible:

1) (internal instruments) Use $\hat{e}_{1t}$ for ϵ_{1t}, where $\hat{e}_{1t}$ is the first element of a Cholesky decomposition. Consistent estimates of R_{i1} are obtained (since $\hat{e}_t$ is consistent) but standard errors need to be adjusted for generated regressors ($Z_t \equiv \hat{e}_{1t}$ may be correlated with $Y_{t-k},\ k > 0$)

WIth Cholesky identification, upper variables may be used as instruments for lower ones, because they're uncorrelated,

2) (external instruments) Using the definition of innovations write (19) as

$$Y_{it} = R_{i1}Y_{1t} + \gamma_i(L)Y_{t-1} + u_{it} \tag{20}$$

where $\gamma_i(L)$ are the coefficients of $P(Y_{it} - R_{i1}Y_{1t}|Y_{t-1}, Y_{t-2}, \ldots)$. Use 2SLS with Z_t as instruments for Y_{1t} in (20) to estimate R_{i1}. Standard errors are correct here.

- Split estimation problem in two steps: use VAR to estimate A(L); use (20) to estimate R_{i1} each i.

- Advantage: $A(L), R_{i1}$ can be estimated over different samples (impossible with LP-IV).

- For h=0 LP-IV and SVAR-IV are the same if X_t are the same.

- **For $h > 0$ they may differ. LP-IV is non parametric and estimates each h separately. SVAR-IV makes parametric assumptions and has restrictions across h.**

- To identify news shocks (with k period ahead lead) (20) becomes:

$$Y_{it} = R_{i1}\hat{W}_{t+h} + \gamma_i(L)Y_{t-1} + u_{it} \tag{21}$$

where $\hat{W}_{t+k} = \hat{B}_k Y_t$ and $\hat{B}(L) = (1 - \hat{A}(L))^{-1}$.

• Stock and Watson (2018): If the DGP is a VAR(q) and invertibily holds, SVAR-IV and LP-IV are consistent and SVAR-IV more efficient. If DGP is not a VAR(q) or invertibility does not hold, SVAR-IV inconsistent, local projections consistent.

• Can use Hausmann test to check misspecification (if we assume a VAR(q) DGP, this becomes a test of invertibility): under the null both SVAR and LP are consistent but SVAR more efficient; under the alternative only LP is consistent.

• Stock and Watson (2018): Assuming $E(e^*_{t+h} Z^*_t) = 0,\ h > 0$ in LP-IV is equivalent to assuming that the VAR is invertible.

• Plagborg-Muller and Wolf (2018): under standard assumptions SVAR-LP they are equivalent asymptotically (they estimate the same IRF).

Relationship IV-narrative sign restrictions

• Narrative sign restrictions (Antolin and Rubio, 2019): can be though as IV approach in the sign restriction space.

• Sign restrictions identify sets which can be very large and may include responses which are quantitatively unreasonable.

• Narrative sign restrictions pick responses which are consistent (correlated with) restrictions on the sign or the relative importance of shocks in terms of variance decomposition in particular historical episodes.

• Sign of shocks are particular instruments!

• Same idea with non-nomrality restrictions applied to sign identified responses.

SVAR one-step likelihood estimation

• SVAR-IV standard errors computed with bootstrap method too small (see Jentsch and Lundsford, 2019). Overstating significance of IRFs.

• Typically use one instrument per shock. What if there are more than one instrument?

• Angelini and Fanelli (2019): SVAR-AC. Want to identify k shocks using $r \geq k$ instruments. Assume $E(v_t \epsilon_{1t}) = F, rank(F) = k$ and $E(v_t \epsilon_{2t}) = 0_{r \times (n-k)}$.

$$\begin{pmatrix} Y_t \\ Z_t \end{pmatrix} = \begin{pmatrix} A(L) & 0_{n\times r} \\ \Gamma(L) & \Theta(L) \end{pmatrix} \begin{pmatrix} Y_{t-1} \\ Z_{t-1} \end{pmatrix} + \begin{pmatrix} u_t \\ v_t \end{pmatrix} \tag{22}$$

v--- u_t not affected by w_t

$$\begin{pmatrix} u_t \\ v_t \end{pmatrix} = \begin{pmatrix} B_1 & B_2 & 0_{n\times r} \\ \Phi & 0_{r\times(n-k)} & \Sigma_\omega^{0.5} \end{pmatrix} \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \\ \omega_t \end{pmatrix} \tag{23}$$

^ instruments uncorrelated

- Instruments Z_t purged i.e. $v_t = z_t - E_t(z_t|F_{t-1})$

- Instrument orthogonality imposed (see zeros in (23)); instrument relevance can be tested: null is $\Phi = 0$.

- Can rewrite the (22)-(23) as

$$
\begin{aligned}
W_t &= D(L)W_{t-1} + \eta_t && (24) \\
\eta_t &= G\zeta_t && (25)
\end{aligned}
$$

- Restriction: $GG' = \mathbf{\Sigma}_\eta$ because $E(\zeta_t \zeta_t' = I)$.

- G can be estimated by ML

- Orthogonality of v_t can be tested via a LR test.

Use first part of the data as prior for local projection

Bayesian LP

• LP defines a regression model so can use Bayesian methods to obtain posterior distributions of IRFs, given a prior for regression parameters.

• No new issue relative to VARs. Posterior intervals easily computed by simulations.

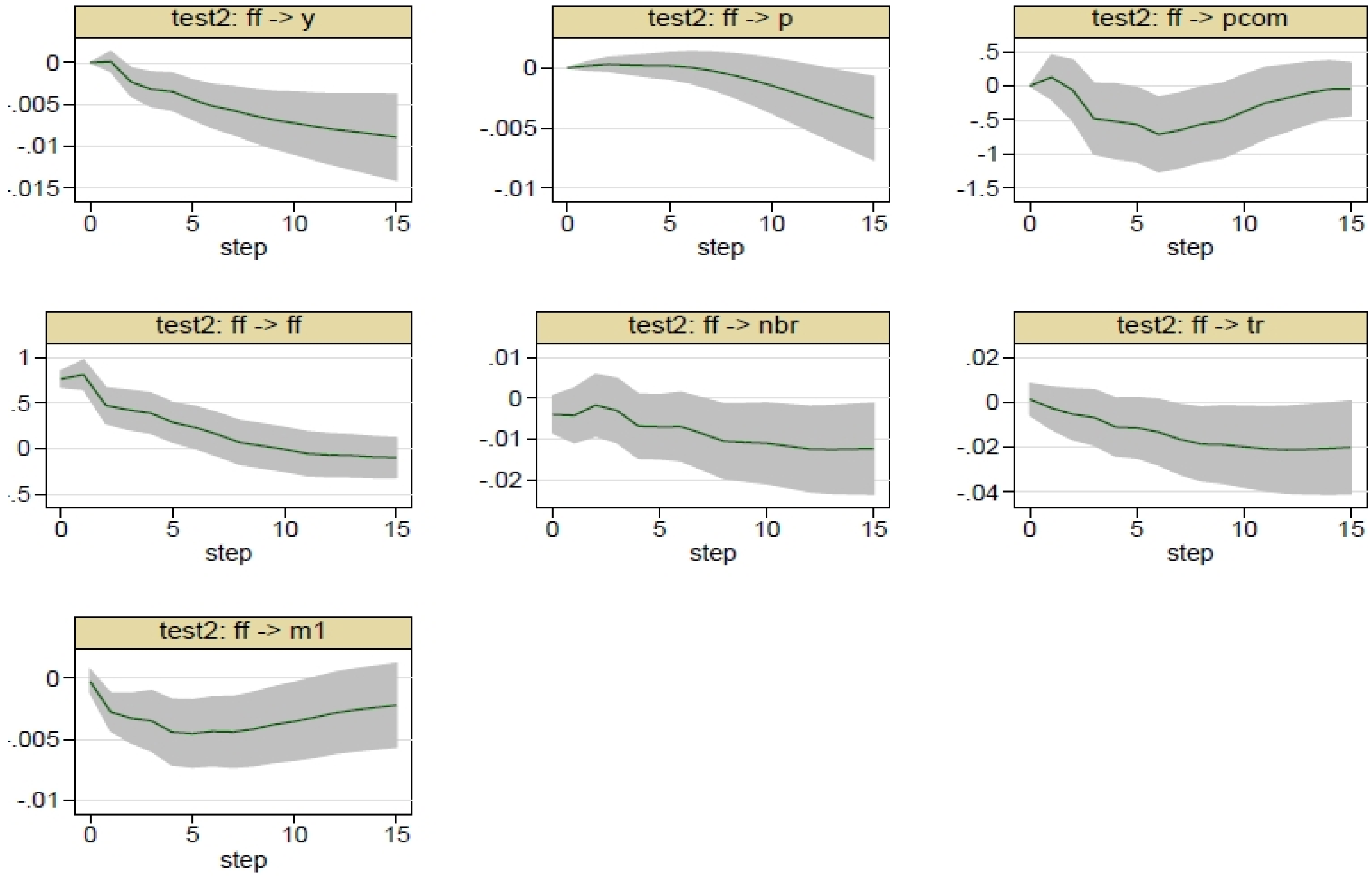
• Miranda Agrippino-Ricco (2021) Use the VAR-LP relationship to setup a prior for LP parameters.

• Empirical macro toolkit uses this approach to setup priors for LP coefficients.

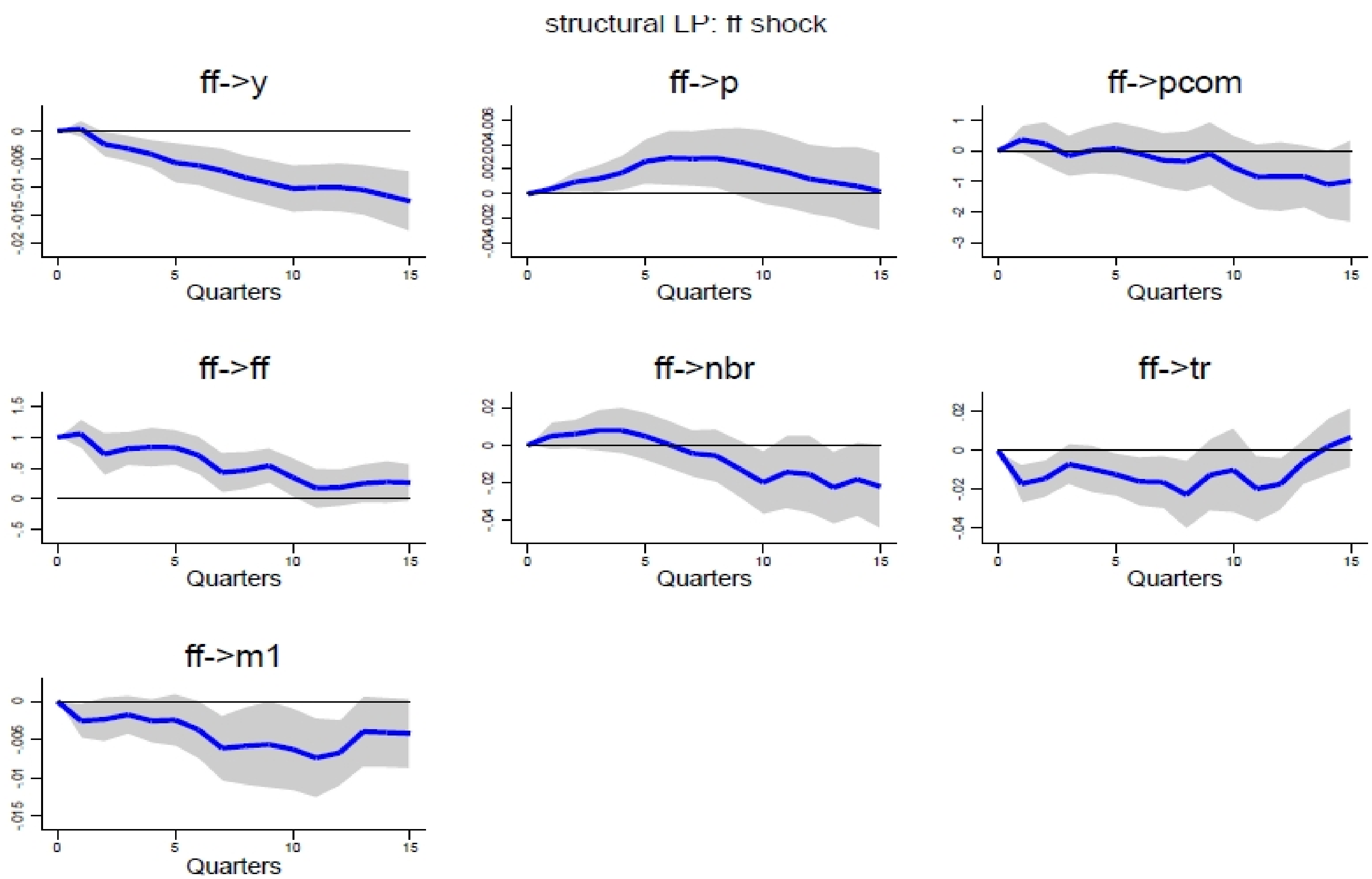
Example 1 *(Jorda, 2005) Consider estimation of the impact of a 100 basis points increase in the FFR on a number of macrovariables. Use SVAR and local projections (without instruments).*

- *Monetary policy shock* ϵ_{1t}. *Identified with Cholesky.*

- *Dynamic ACE:* $E(Y_{i,t+h}|\epsilon_{1t} = 1) - E(Y_{i,t+h}|\epsilon_{1,t} = 0) \equiv \theta_{i1}^{h}$.



test2: ff -> y
test2: ff -> p
test2: ff -> pcom
test2: ff -> ff
test2: ff -> nbr
test2: ff -> tr
test2: ff -> m1
step



structural LP: ff shock
ff->y
ff->p
ff->pcom
ff->ff
ff->nbr
ff->tr
ff->m1
Quarters

Example 2 *(Stock-Watson, 2017) Gertler and Karadi (2015) VAR with* $(R_t, 100 * \Delta \ln(IP), 100 * \Delta \ln(CPI)$, *External bond premium (EBP)).*

- Z_t: Δ *FFR futures in a short window around FOMC announcements.*

- $E(Z_t \epsilon_{1t}) \neq 0$;

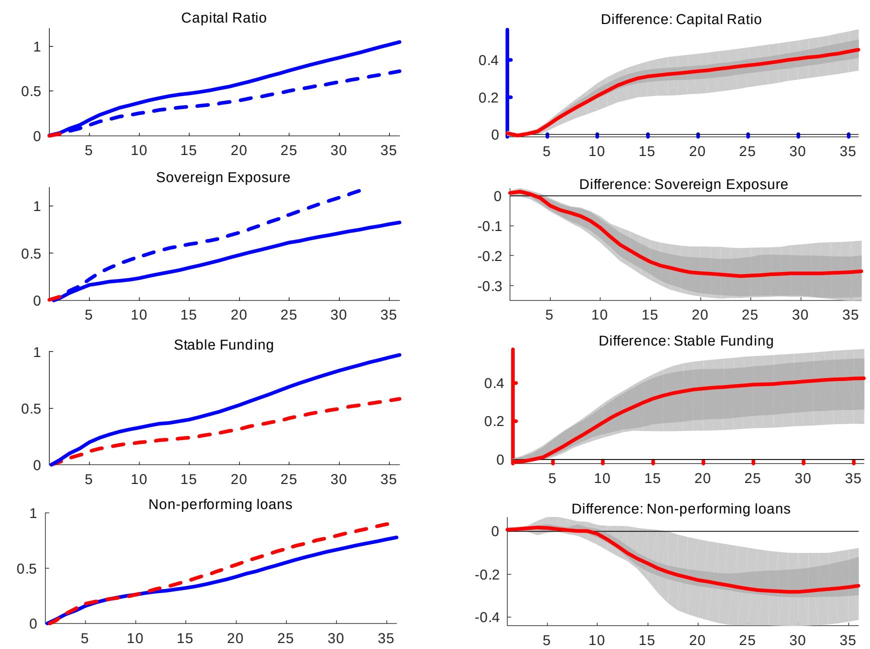
- $E(Z_t \epsilon_{-1,t} = 0)$ *(hopefully).*

Table 1: Estimated causal effect of monetary policy shocks on selected economic variables: Gertler-Karadi (2015) variables, instrument and sample period

		LP-IV			SVAR		SVAR – LP
	lag (h)	(a)	(b)	(c)	(d)		(d)-(b)
R	0	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)		0.00 (0.00)
	6	-0.07 (1.34)	1.12 (0.52)	0.67 (0.57)	0.89 (0.31)		-0.23 (1.19)
	12	-1.05 (2.51)	0.78 (1.02)	-0.12 (1.07)	0.78 (0.46)		0.00 (1.79)
	24	-2.09 (5.66)	-0.80 (1.53)	-1.57 (1.48)	0.40 (0.49)		1.19 (2.57)
IP	0	-0.59 (0.71)	0.21 (0.40)	0.03 (0.55)	0.16 (0.59)		-0.06 (0.35)
	6	-2.15 (3.42)	-3.80 (3.14)	-4.05 (3.65)	-0.81 (1.19)		3.00 (2.32)
	12	-3.60 (6.23)	-6.70 (4.70)	-6.86 (5.49)	-1.87 (1.54)		4.83 (4.00)
	24	-2.99 (10.21)	-9.51 (7.70)	-8.13 (7.62)	-2.16 (1.65)		7.35 (6.40)
P	0	0.02 (0.07)	-0.08 (0.25)	-0.04 (0.25)	0.02 (0.23)		0.10 (0.13)
	6	0.16 (0.42)	-0.39 (0.52)	-0.79 (0.83)	0.31 (0.41)		0.71 (0.98)
	12	-0.26 (0.88)	-1.35 (1.03)	-1.37 (1.23)	0.45 (0.54)		1.80 (1.53)
	24	-0.88 (3.08)	-2.26 (1.31)	-2.58 (1.69)	0.50 (0.65)		2.76 (2.60)
EBP	0	0.51 (0.61)	0.67 (0.40)	0.82 (0.49)	0.77 (0.29)		0.09 (0.24)
	6	0.22 (0.30)	1.33 (0.81)	1.66 (1.04)	0.48 (0.20)		-0.85 (0.51)
	12	0.56 (0.91)	0.84 (0.65)	0.91 (0.80)	0.18 (0.13)		-0.66 (0.55)
	24	-0.44 (1.29)	0.94 (0.66)	0.85 (0.76)	0.06 (0.07)		-0.88 (0.62)
Controls		none	4 lags of (z,y)	4 lags of (z,y,f)	12 lags of y 4 lags of z		na
First-stage F^{Hom}		1.7	23.7	18.6	20.5		na
First-stage F^{HAC}		1.1	15.5	12.7	19.2		na

Notes: The instrument, Z_t, is available from 1990m1-2012m6; the other variables are available from 1979m1-2012m6. The LP-IV estimates in (a)-(c) use data from 1990m1-2012m6. The VAR for (d) is computed over 1980m7-2012m6; and the IV-regression computed over 1990m5-2012m6. The numbers in parentheses are standard errors computed by Newey-West HAC with h+1 lags for the local projections, and using a parametric Gaussian bootstrap for the SVAR and the SVAR – LP differences shown in (e). In the final two rows F^{Hom} is the standard (conditional homoscedasticity, no serial correlation) first-stage F-statistic, while F^{HAC} is the Newey-West version using 12 lags in (a) and heteroskedasticity-robust (no lags) in (b), (c), and (d).

Example 3 *Altavilla et al. (2020) SVAR-IV. Identify conventional monetary policy shocks (2007-2014) using announcement dates as instruments.*



Upper and lower quantiles of the distribution of lending rate responses to MP shocks.

LP and misspecification

- DGP: One sector growth model. How does LP works with this DGP?

$$\max_{C_t} \sum_{t=1}^{\infty} \beta^t U(C_t)$$

maximization is subject to the constraints

$$\begin{aligned} C_t/B_t + I_t &= O_t = A_t K_{t-1}^{\alpha} \\ K_t &= (1-\delta)K_{t-1} + V_t I_t \end{aligned}$$

where O_t is output, C_t is consumption, I_t investment and K_t is the capital stock, $0 < \alpha < 1$, $0 < \beta < 1$ and $0 < \delta < 1$ are parameters.

- Assume that (A_t, V_t, B_t) are iid with unitary means and standard deviation $\sigma_i, i = A, V, B$, uncorrelated with each other at all leads and lags.

• When $U(C_t) = \log C_t$ and $\delta = 1$, the solution is:

$$\begin{aligned}
\log O_t &= \alpha \log K_{t-1} + \log A_t && (26) \\
\log C_t &= \log(1 - \alpha\beta) + \alpha \log K_{t-1} + \log B_t + \log A_t && (27) \\
\log K_t &= \log(\alpha\beta) + \alpha \log K_{t-1} + \log V_t + \log A_t && (28)
\end{aligned}$$

Three endogenous variables and three disturbances (two supply (A_t, V_t) and one demand B_t).

• In a VAR with $(\log O_t, \log C_t, \log K_t)$, all structural disturbances are identifiable from the innovations using theory-based recursive restriction ($\log A_t$ can be obtained from the innovations in $\log O_t$; given $\log A_t$, the other two innovations determine $\log V_t$ and $\log B_t$).

- The steady states are:

$$\begin{aligned}\log K &= \frac{\log(\alpha\beta)}{1-\alpha} \\ \log C &= \log(1-\alpha\beta) + \frac{\alpha}{1-\alpha}\log(\alpha\beta) \\ \log O &= \frac{\alpha}{1-\alpha}\log(\alpha\beta) \qquad (29)\end{aligned}$$

- If lower case letters indicate log deviation from steady states. Solution:

$$k_t = \alpha k_{t-1} + v_t + a_t \qquad (30)$$

$$c_t = \alpha k_{t-1} + b_t + a_t \qquad (31)$$

$$o_t = \alpha k_{t-1} + a_t \qquad (32)$$

or $w_t = Aw_{t-1} + Be_t$, where $w_t = [k_t, c_t, o_t]'$ and $e_t = [a_t, v_t, b_t]'$.

- The MA representation of the solution is

$$k_t = \sum_{i=0}^{\infty} \alpha^i a_{t-i} + \sum_{i=0}^{\infty} \alpha^i \nu_{t-i} \tag{33}$$

$$c_t = \sum_{i=0}^{\infty} \alpha^i a_{t-i} + \sum_{i=0}^{\infty} \alpha^i b_{t-i} + \alpha \sum_{i=0}^{\infty} \alpha^i \nu_{t-i-1} \tag{34}$$

$$o_t = \sum_{i=0}^{\infty} \alpha^i a_{t-i} + \alpha \sum_{i=0}^{\infty} \alpha^i \nu_{t-i-1} \tag{35}$$

or $w_t = B(L)e_t$. Writing out the last equation explicitly:

$$\begin{aligned} o_{t+h} &= a_{t+h} + \alpha a_{t+h-1} + \ldots + \alpha^{h-1} a_{t+1} + \alpha^h a_t + \alpha^{h+1} a_{t-1} + \ldots \\ &+ \alpha v_{t+h-1} + \ldots + \alpha^{h-1} v_{t+1} + \alpha^h a_t + \alpha^{h+1} v_{t-1} + \ldots \end{aligned} \tag{36}$$

• Theoretical projection of output at horizon h on a technology shock.

$$o_{t+h} = b_h a_t + u_{t+h} \tag{37}$$

• Matching (36) and (37) we have

$$\begin{aligned} b_h &= \alpha^h && (38)\\ u_{t+h} &= (a_{t+h} + \alpha a_{t+h-1} + \ldots + \alpha^{h-1} a_{t+1} + \alpha^{h+1} a_{t-1} + \alpha^{h+2} a_{t-2} + \ldots \\ &+ \alpha \sum_{i=0}^{\infty} \alpha^i v_{t+h-1-i} && (39) \end{aligned}$$

• Error term in the projection is a function of all values of v_t and past and future values of a_t.

• a_t is not observable; equation (37) can not be used. Suppose one employs the residuals of the VAR (or their identified version) in place of a_t. These could be proxies for Solow residuals or some measure of productivity.

OLS estimation

• (No misspecification) If the VAR includes w_t and, at least, one lag, then $\hat{e}_t = w_t - \hat{A}w_{t-1}$ and $b_{h,OLS}$ is **consistent** for a^h, because $\hat{e}_{3,t} = a_t$.

• (Deformation) If the VAR is **deformed**, for example , only c_t, y_t are used to compute $\hat{e}_t$, $b_{h,OLS}$ is still **consistent**. The system is

$$\begin{aligned} c_t &= \alpha c_{t-1} + a_t + \alpha v_{t-1} + b_t - \alpha b_{t-1} \\ o_t &= \alpha o_{t-1} + a_t + \alpha v_{t-1} \end{aligned} \tag{40}$$

(40) allows identification of $\tilde{e}_{2t} = a_t + \alpha v_{t-1}$ from o_t, and $\tilde{e}_{1t} = b_t - \alpha b_{t-1}$ from c_t, given $\tilde{e}_{2t}$. The MA representation is

$$c_t = \sum_{i=0}^{\infty} \alpha^i a_{t-i} + \sum_{i=0}^{\infty} \alpha^i (b_{t-i} - \alpha b_{t-i-1}) + \alpha \sum_{i=0}^{\infty} \alpha^i \nu_{t-i-1} \tag{41}$$

$$o_t = \sum_{i=0}^{\infty} \alpha^i a_{t-i} + \alpha \sum_{i=0}^{\infty} \alpha^i \nu_{t-i-1} \tag{42}$$

• Thus, the MA format for o_t is unchanged. Then

$$\begin{aligned}
\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}(\tilde{e}_{2t}\tilde{e}_{2t}) &= \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}(a_t+\alpha v_{t-1})(a_t+\alpha v_{t-1}) \\
&= \sigma_a^2+\alpha^2\sigma_v^2 \qquad (43)\\
\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}(\tilde{e}_{2t}o_{t+h}) &= \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}(a_t+\alpha v_{t-1})(\ldots+\alpha^h a_t+\ldots+\alpha^{h+1}v_{t-1}+\ldots) \\
&= \alpha^h(\sigma_a^2+\alpha^2\sigma_v^2) \qquad (44)
\end{aligned}$$

Thus $b_{h,OLS}=\alpha^h$.

• Numerator and denominator are biased, but bias is in the same direction and cancels out. LP solves some of the problems existing with VAR IRFs when deformation is present, see Canova and Ferroni (2020, forthcoming AEJ: Macro).

• (Measurement error) If the VAR includes w_t and o_t is measured with error, say $y_t = o_t + u_t$, u_t iid uncorrelated with the structural shocks, $b_{h,OLS}$ is **inconsistent**. This is because $\tilde{e}_{3,t} = a_t + u_t$ and

$$\begin{aligned} b_{h,OLS} &= \left(\sum_{t=1}^{T} \tilde{e}_{3,t} y_{t+h}\right) \left(\sum_{t=1}^{T} (\tilde{e}_{3,t})^2\right)^{-1} \\ &= \left(\sum_{t=1}^{T} (a_t + u_t)(\sum_i \alpha^i z_{t+h-i} + \alpha \sum_i \alpha^i v_{t+h-i-1} + u_{t+h})\right) \left(\sum_{t=1}^{T} (a_t + u_t)^2\right)^{-1} \\ &= \left(\sum_{t=1}^{T} (a_t + u_t)(\ldots + \alpha^h a_t + \ldots)\right) \left(\sum_{t=1}^{T} (a_t + u_t)^2\right)^{-1} \\ &\xrightarrow{T\to\infty} \alpha^h \frac{\sigma_a^2}{\sigma_a^2 + \sigma_u^2} < \alpha^h \end{aligned} \tag{45}$$

• Classical contamination bias with a proxy regressor. There is an asymptotic downward bias. The bias is larger, the larger is the standard deviation of the measurement error σ_u^2.

• (Normalization) Using y_t in place of $\tilde{e}_{3,t}$, as suggested in Stock and Watson (2018), nothing will change:

$$\begin{aligned} b_{h,OLS} &= \left(\sum_{t=1}^{T} y_t y_{t+h}\right)\left(\sum_{t=1}^{T} y_t^2\right)^{-1} \\ &= \left(\sum_{t=1}^{T}(\sum_i \alpha^i a_{t-i} + \alpha \sum_i \alpha^i v_{t-i-1} + u_t)(\sum_i \alpha^i a_{t+h-i} + \alpha \sum_i \alpha^i v_{t+h-i-1} + u_{t+h})\right) \times \\ &\quad \left(\sum_{t=1}^{T}(\sum_i \alpha^i a_{t-i} + \alpha \sum_i \alpha^i v_{t-i-1} + u_t)^2\right)^{-1} \end{aligned} \tag{46}$$

The denominator limit is

$$\lim_{T\to\infty} \frac{1}{T}\sum_t (o_t + u_t)^2 = \sum_{i=0}^{\infty} \alpha^{2i}\sigma_a^2 + \alpha^2 \sum_{i=0}^{\infty} \alpha^{2i}\sigma_v^2 + \sigma_u^2$$

The numerator limit is

$$\lim_{T\to\infty} \frac{1}{T}\sum_t ((o_t + u_t)(o_{t+h} + u_{t+h}) = \alpha^h(\sum_{t=0}^{\infty} \alpha^{2i}\sigma_a^2 + \alpha^2 \sum_{i=0}^{\infty} \alpha^{2i}\sigma_v^2)$$

Hence

$$b_{h,OLS} \xrightarrow{T\to\infty} \alpha^h \frac{\frac{\sigma_a^2}{1-\alpha^2} + \frac{\alpha^2\sigma_v^2}{1-\alpha^2}}{\frac{\sigma_a^2}{1-\alpha^2} + \frac{\alpha^2\sigma_v^2}{1-\alpha^2} + \sigma_u^2} = \alpha^h \frac{\sigma_a^2 + \alpha^2\sigma_v^2}{\sigma_a^2 + \alpha^2\sigma_v^2 + (1-\alpha^2)\sigma_u^2} \neq \alpha^h \tag{47}$$

• Bias remains. The size of the bias depends also on the magnitude of α and σ_v^2 (before it did not).

- (Incorrect timing) What if $\tilde{e}_{3t} = a_{t-1} + u_t$. This could be produced, for example, if $y_t = o_{t-1} + u_t$. OLS is **biased**.

$$
\begin{aligned}
b_{h,OLS} &= \left(\sum_{t=1}^{T} \tilde{e}_{3,t} y_{t+h}\right) \left(\sum_{t=1}^{T} \tilde{e}_{3,t}^2\right)^{-1} && (48) \\
&= \left(\sum_{t=1}^{T} (a_{t-1} + u_t)(\sum_i \alpha^i a_{t+h-i} + \alpha \sum_i \alpha^i v_{t+h-i-1} + u_{t+h})\right) \left(\sum_{t=1}^{T} (a_{t-1} + u_t)^2\right)^{-1} \\
&= \left(\sum_{t=1}^{T} (a_{t-1} + u_t)(... + \alpha^{h+1} a_{t-1} + ...)\right) \left(\sum_{t=1}^{T} (a_{t-1} + u_t)^2\right)^{-1} && (49) \\
&\xrightarrow{T\to\infty} \alpha^{h+1} \frac{\sigma_a^2}{\sigma_a^2 + \sigma_u^2} \neq \alpha^h && (50)
\end{aligned}
$$

- General downward bias. If $\alpha < 1$ bias is larger with incorrect timing.

• (Measurement error and incorrect timing). If $\tilde{e}_{3,t} = \gamma a_t + (1-\gamma) a_{t-1} + u_t$ then

$$\begin{aligned} b_{h,OLS} &\xrightarrow{T\to\infty} (\gamma\alpha^h + (1-\gamma)\alpha^{h+1}) \frac{\sigma_a^2}{\sigma_a^2 + \sigma_u^2} \\ &= \alpha^h(\gamma + (1-\gamma)\alpha) \frac{\sigma_a^2}{\sigma_a^2 + \sigma_u^2} \end{aligned} \tag{51}$$

General expression ($\gamma(L)$ could be two-sided):

$$\begin{aligned} \tilde{e}_{3,t} &= \gamma(L) a_t + u_t \\ b_{h,OLS} &\xrightarrow{T\to\infty} \alpha^h \left(\sum_{i=0}^{\infty} \gamma_i \alpha^i \right) \frac{\sigma_a^2}{\sigma_a^2 + \sigma_u^2} \end{aligned} \tag{52}$$

where $\gamma(L) = \gamma_0 + \gamma_1 L + ...$

IV estimation

- Suppose we have available an instrument z_t,a contaminated measure of a_t, e.g.

$$z_t = \gamma_1 a_t + \gamma_2 a_{t-1} + \gamma_3 a_{t+1} + \gamma_4 v_t + u_{1t} \tag{53}$$

Let $\gamma_1 >> 0$ (relevance condition), u_{1t} be iid and that we run the regression with o_t as dependent variable.

- z_t is not necessarily orthogonal to ϵ_{t+h}.

$$\begin{aligned} z_t o_{t+h} &= (\gamma_1 a_t + \gamma_2 a_{t-1} + \gamma_3 a_{t+1} + \gamma_4 v_t + u_t) \\ &\times (\ldots + \alpha^{h-1} a_{t+1} + \alpha^h a_t + \alpha^{h+1} a_{t-1} + \ldots + \alpha^{h-1} v_{t+1} + \alpha^h v_t + \alpha^{h+1} v_{t-1} + \ldots) \\ z_t o_t &= (\gamma_1 a_t + \gamma_2 a_{t-1} + \gamma_3 a_{t+1} + \gamma_4 v_t + u_t) \times (a_t + \alpha a_{t-1} + \ldots + \alpha v_{t-1} + \ldots) \end{aligned} \tag{54}$$

For $h > 0$ and $h = 0$, we have respectively

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} z_t o_{t+h} = (\gamma_3 \alpha^{h-1} + \gamma_1 \alpha^h + \gamma_2 \alpha^{h+1}) \sigma_a^2 + \alpha^h \gamma_4 \sigma_v^2 \quad (55)$$

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} z_t o_t = (\gamma_1 + \alpha\gamma_2)\ \sigma_a^2 \quad (56)$$

Hence

$$\begin{aligned} b_{h,IV} &= \left(\sum_{t=1}^{T} z_t o_t\right)^{-1} \left(\sum_{t=1}^{T} z_t o_{t+h}\right) \\ &\xrightarrow{T\to\infty} \frac{(\gamma_3 \alpha^{h-1} + \gamma_1 \alpha^h + \gamma_2 \alpha^{h+1}) \sigma_a^2 + \alpha^h \gamma_4 \sigma_v^2}{(\gamma_1 + \alpha\gamma_2)\sigma_a^2} \\ &= \alpha^h \left[\frac{(\gamma_3 \alpha^{-1} + \gamma_1 + \gamma_2 \alpha)}{\gamma_1 + \alpha\gamma_2} + \frac{\gamma_4 \sigma_v^2}{(\gamma_1 + \alpha\gamma_2)\sigma_a^2}\right] \neq \alpha^h \end{aligned} \quad (57)$$

A few special cases

- No news (i.e. $\gamma_3 = 0$)

$$b_{h,IV} = \frac{(\gamma_1 + \gamma_2\alpha)\alpha^h\sigma_a^2 + \ \alpha^h\gamma_4\sigma_v^2}{(\gamma_1 + \alpha\gamma_2)\sigma_a^2} = \alpha^h\left(1 + \frac{\gamma_4}{(\gamma_1 + \alpha\gamma_2)}\frac{\sigma_v^2}{\sigma_a^2}\right) \quad (58)$$

Asymptotic bias due to invalid instrument.

- No lags (i.e. $\gamma_2 = 0$)

$$b_{h,IV} = \frac{(\gamma_1 + \gamma_3\alpha^{-1})\alpha^h\sigma_a^2 + \alpha^h\gamma_4\sigma_v^2}{\gamma_1\sigma_a^2} = \alpha^h + \frac{\gamma_3}{\gamma_1}\alpha^{h-1} + \alpha^{h+1}\frac{\gamma_4\sigma_v^2}{\gamma_1\sigma_a^2} \quad (59)$$

Presence or absence of lags does not affect the consistency of $b_{h,IV}$.

- No news and no contamination (i.e. $\gamma_4 = \gamma_3 = 0$)

$$b_{h,IV} = \alpha^h \quad (60)$$

- What if the projection is run with $\tilde{e}_{3,t}$ and y_{t+h}?

$$\begin{aligned} z_t \tilde{e}_{3t} &= (\gamma_1 a_t + \gamma_2 a_{t-1} + \gamma_3 a_{t+1} + \gamma_4 v_t + u_t) \times (a_t + u_{2t}) && (61) \\ z_t y_{t+h} &= (\gamma_1 a_t + \gamma_2 a_{t-1} + \gamma_3 a_{t+1} + \gamma_4 v_t + u_t) \\ &\times (\ldots + \alpha^{h-1} a_{t+1} + \alpha^h a_t + \alpha^{h+1} a_{t-1} + \ldots + \alpha^{h-1} v_{t+1} + \alpha^h v_t + \alpha^{h+1} v_{t-1} + \ldots + u_{t+h}) && (62) \end{aligned}$$

Sample moments:

$$\lim_{T\to\infty} \quad \frac{1}{T}\sum_{t=1}^{T} z_t \tilde{e}_{3t} = \gamma_1 \sigma_a^2 \tag{63}$$

$$\lim_{T\to\infty} \quad \frac{1}{T}\sum_{t=1}^{T} z_t y_{t+h} = (\gamma_3 \alpha^{h-1} + \gamma_1 \alpha^h + \gamma_2 \alpha^{h+1})\sigma_a^2 + \alpha^h \gamma_4 \sigma_v^2 \tag{64}$$

Hence

$$b_{h,IV} = \frac{\frac{1}{T}\sum_t z_t y_{t+h}}{\frac{1}{T}\sum_t z_t \tilde{e}_{3t}} \xrightarrow{T\to\infty} \alpha^h \left[\frac{(\gamma_3 \alpha^{-1} + \gamma_1 + \gamma_2 \alpha)}{\gamma_1} + \frac{\gamma_4 \sigma_v^2}{\gamma_1 \sigma_a^2} \right] \neq \alpha^h \tag{65}$$

• Assume that the first stage regression uses k_t instead of o_t; that is, we first regress k_t on z_t and then we regress o_{t+h} on the fitted values of the first regressions. Since

$$z_t k_t = (\gamma_1 a_t + \gamma_2 a_{t-1} + \gamma_3 a_{t+1} + \gamma_4 v_t + u_t)(a_t + \alpha a_{t-1} + \ldots + v_t + \alpha v_{t-1} + \ldots) \tag{66}$$

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} z_t k_t = (\gamma_1 + \alpha\gamma_2)\ \sigma_a^2 + \gamma_4 \sigma_v^2 \tag{67}$$

Hence

$$\begin{aligned} b_{h,IV} &= \left(\sum_{t=1}^{T} z_t k_t\right)^{-1} \left(\sum_{t=1}^{T} z_t o_{t+h}\right) \\ &\xrightarrow{T\to\infty} \frac{(\gamma_3 \alpha^{h-1} + \gamma_1 \alpha^h + \gamma_2 \alpha^{h+1})\sigma_a^2 + \alpha^h \gamma_4 \sigma_v^2}{(\gamma_1 + \alpha\gamma_2)\sigma_a^2 + \gamma_4 \sigma_v^2} \end{aligned} \tag{68}$$

• If there are no news ($\gamma_3 = 0$), $b_{h,IV} \to \alpha^h$.

• Suppose the instrument w_t is related to a_t and the demand shock b_t:

$$w_t = \gamma_1 a_t + \gamma_2 a_{t-1} + \gamma_3 a_{t+1} + \gamma_4 b_t + u_{1t} \tag{69}$$

• Keep $\gamma_1 >> 0$ (relevance condition), u_{1t} is iid and assume we run the regression with c_t as dependent/response variable. Then

$$\begin{aligned} w_t c_{t+h} &= (\gamma_1 a_t + \gamma_2 a_{t-1} + \gamma_3 a_{t+1} + \gamma_4 b_t + u_t) \\ &\times (a_{t+h} + \ldots + \alpha^{h-1} a_{t+1} + \alpha^h a_t + \alpha^{h+1} a_{t-1} + \ldots + \alpha^h v_t + \ldots + \alpha^h b_t + \ldots) \\ w_t c_t &= (\gamma_1 a_t + \gamma_2 a_{t-1} + \gamma_3 a_{t+1} + \gamma_4 b_t + u_t)(a_t + \alpha a_{t-1} + \ldots + \alpha v_{t-1} + \ldots + b_t + \alpha b_{t-1} + \ldots \end{aligned}$$

For $h > 0$ and $h = 0$, we have respectively

$$\begin{aligned} \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} w_t c_{t+h} &= (\gamma_3 \alpha^{h-1} + \gamma_1 \alpha^h + \gamma_2 \alpha^{h+1}) \sigma_a^2 + \alpha^h \gamma_4 \sigma_b^2 && (70) \\ \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} w_t c_t &= (\gamma_1 + \alpha\gamma_2)\, \sigma_a^2 + \gamma_4 \sigma_b^2 && (71) \end{aligned}$$

Hence

$$
\begin{aligned}
b_{h,IV} &= \left(\sum_{t=1}^{T} w_t c_t\right)^{-1} \left(\sum_{t=1}^{T} w_t c_{t+h}\right) \\
&\xrightarrow{T\to\infty} \alpha^h \frac{(\gamma_3 \alpha^{-1} + \gamma_1 + \gamma_2 \alpha)\sigma_a^2 + \gamma_4 \sigma_b^2}{(\gamma_1 + \alpha\gamma_2)\sigma_a^2 + \gamma_4 \sigma_b^2} \neq \alpha^h \qquad (72)
\end{aligned}
$$

If no news ($\gamma_3 = 0$), $b_{h,IV} \to \alpha^h$. Similar conclusions can be derived using o_t instead of c_t.

- The presence of news makes IV inconsistent!

Lag augmentation in LP

• Common to run LP using:

$$y_{t+h} = \beta_h y_t + \text{controls} + u_{t+h}, \quad h = 1, 2, ... \tag{73}$$

• Interest is typically in large horizon h.

• Typically y_t is persistent (close to a unit root).

• Combination of persistent y_t and large h creates problems in classical inference.

If you have misspecified the controls, lag augmentation cleans it up.

- Montiel Olea and Plagborg Muller (2021). Use lag augmentation

$$y_{t+h} = \beta_h y_t + \sum_{\ell=1}^{p} \gamma_{\ell h} y_{t-\ell} + \text{controls} + e_{t+h}, \quad h = 1, 2, \ldots \tag{74}$$

- Lag-augmented LP inference is uniformly valid when the DGP features a unit root and horizon h is large.

- Valid also when $h = h_T \propto T^{\eta}$ for $\eta \in [0, 1)$ (and $\propto T$ if no unit root).

- Lag augmentation obviates need for HAC/HAR standard error (Heteroskedasticity-robust standard errors suffice).

- Simple. No need to choose tuning parameters.

AR(1) example

$$y_{t+h} = \rho^h y_t + \sum_{\ell=1}^{h} \rho^{h-\ell} u_{t+\ell} \equiv \beta(\rho, h) y_t + \epsilon_t(\rho, h) \tag{75}$$

• OLS $\hat{\beta}$ consistent and asymptotically normal for $|\rho| < 1$.

• Non-normal if $\rho \approx 1$.

• Requires HAR estimate even when $|\rho| < 1$ since ϵ_t is serially correlated. Inference difficult in small samples. Need tuning parameters.

- With lag augmentation inference is robust.

- Let $x_t = (y_t, y_{t-1})'$. Let $\alpha(h) = (\beta(\rho, h), \gamma(\rho, h))$. Then

$$\hat{\alpha}(h) = (\sum_t x_t x_t')^{-1} (\sum x_t y_{t+h}) \tag{76}$$

- $\hat{\beta}(h)_{LA}$ is the same as the one obtained using $u_t = y_t - \rho y_{t-1}$ and y_{t-1}).

- $\hat{\beta}(h)_{LA}$ has uniform normal limit, since

$$y_{t+1} = \beta(h) u_t + \beta(\rho, h+1) y_{t-1} + \epsilon_t \tag{77}$$

- First term stationary. Inference is OK. Control Y_{t-1} non-stationary if $\rho \approx 1 \rightarrow \gamma(h) = \beta(1, h+1)$ non-normal, but we do not care.

• Lag augmentation simplies computation of standard errors. Leading term in asymptotic expansion

$$\hat{\beta}(h) = \beta(\rho, h) + \frac{\sum_{t=1}^{T-1} \epsilon_t(\rho, h) u_t}{u_t^2} \tag{78}$$

• ϵ_t is serially correlated but the score of $\epsilon_t(\rho, h)u_t$ are not as long as $E(u_t|u_s) = 0, s > t$ since for $s < t$

$$E[\epsilon_t(\rho, h)u_t\epsilon_s(\rho, h)u_s] = E[\epsilon_t(\rho, h)u_t\epsilon_s(\rho, h)E(u_s|u_{s+1}, u_{s+2}, \ldots)] \tag{79}$$

and the last expectation is zero.

• Sufficient to use (Eicker-Huber-White) standard errors

$$\hat{s}(h) = \frac{(\sum_t \hat{\epsilon}(h)^2 \hat{u}(h)^2)^{0.5}}{\sum_t \hat{u}(h)^2} \tag{80}$$

where $\hat{\epsilon}_t = y_{t+h} - \hat{\beta}(h) y_t - \hat{\gamma}(h) y_{t-1}$; $\hat{u}_t = y_t - \hat{\rho}(h) y_{t-1}$;$\hat{\rho}(h) = \frac{\sum_t y_t y_{t-1}}{\sum_t y_{t-1}^2}$.

- No need of tuning parameters.

- Confidence interval:

$$\hat{C}(h,\alpha) = [\hat{\beta}(h) \pm z_{(1-0.5\alpha)} \hat{s}(h)] \tag{81}$$

is uniformely valid, i.e. as $T \rightarrow \infty$:

$$\inf_{\rho \in [-1,1]} \inf_{1 \leq h \leq h_T} P_\rho(\beta(\rho,h) \in \hat{C}(h,\alpha) \rightarrow 1 - \alpha$$

for any h_T sequence such that $\frac{h_T}{T} \rightarrow 0$.

- If $\rho < 1 - a$ one could even choose $h_T \propto T$.

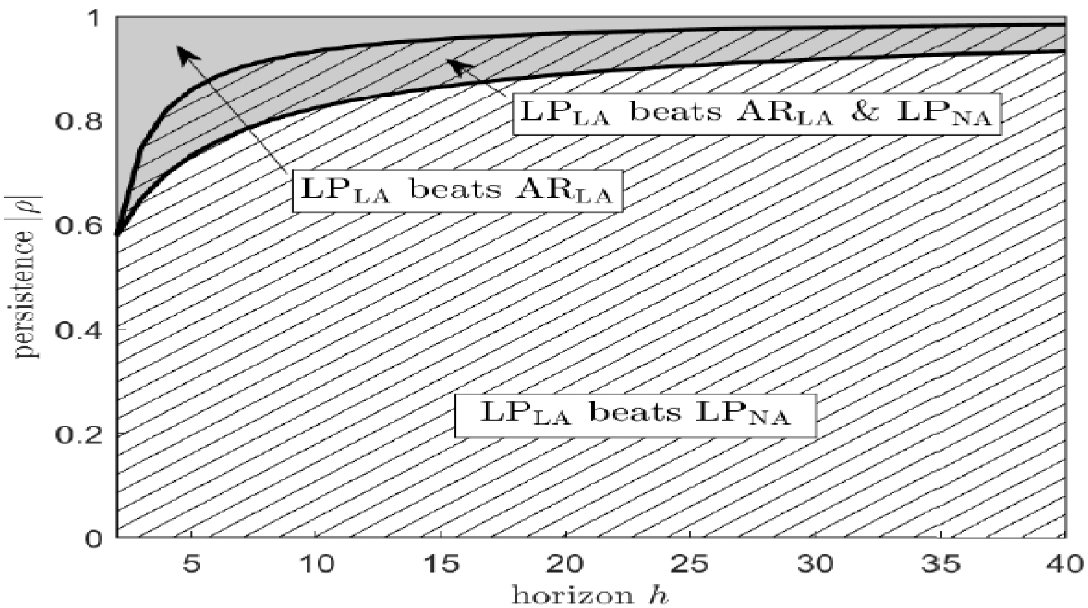


Figure 1: Efficiency ranking of three different estimators of the fixed impulse response $\beta(\rho, h) = \rho^h$ in the homoskedastic AR(1) model: lag-augmented LP ($\mathrm{LP_{LA}}$), non-augmented LP ($\mathrm{LP_{NA}}$), and lag-augmented AR ($\mathrm{AR_{LA}}$). Gray area: combinations of $(|\rho|, h)$ for which $\mathrm{LP_{LA}}$ is more efficient than $\mathrm{AR_{LA}}$. Thatched area: $\mathrm{LP_{LA}}$ is more efficient than $\mathrm{LP_{NA}}$. See Appendix B.2.1 for analytical derivations of the indifference curves (thick lines).

Small sample biases in LP

• Herbst et al (2021): LP is ok in large samples but very biased in small samples.

OK with 300-500 data points

• Biases present even when relevant regressor is iid.

• LP bias at horizon h is a weighted sum of the (population) impulse response function at other h's. Thus, if LP estimators across horizons have the same sign, the least-squares estimators are biased toward zero at every horizon.

• Small sample LP estimates are not ''local'' because the small-sample biases depend on the true impulse responses at other horizons.

• Standard error also biased. HAR correction still biased in standard macro samples (mostly downward).

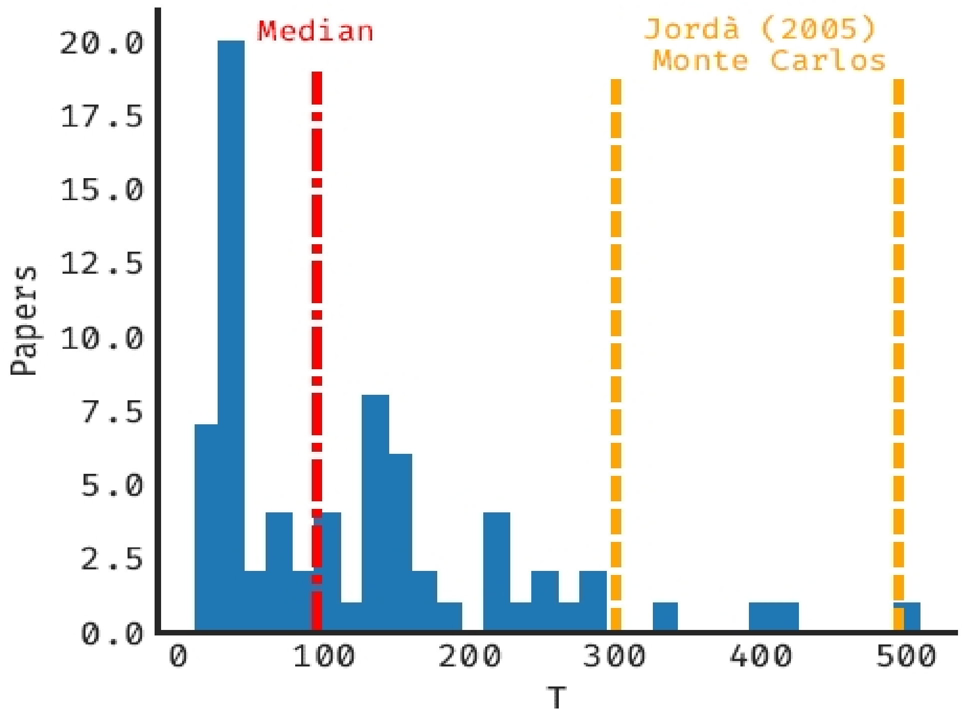


Figure 1: T is small in the literature using LPs.

(a) Without Controls (b) With Controls

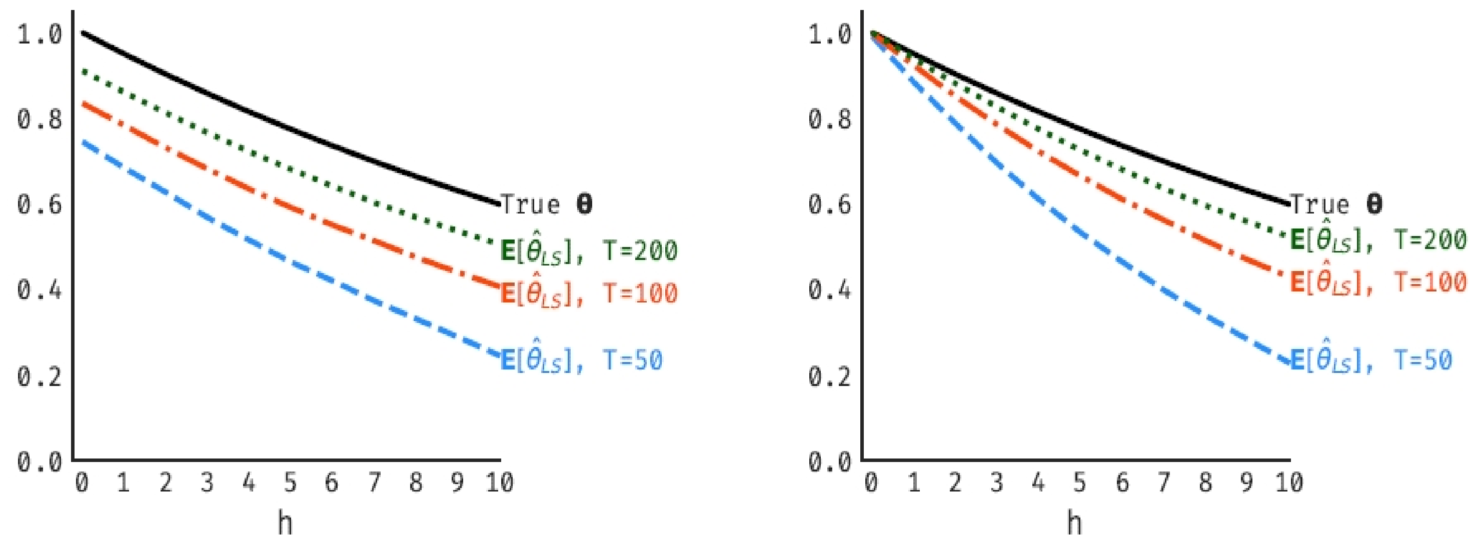


Figure 2: LP estimators are biased in empirically-relevant samples when y_t is an AR(1) with $\rho = 0.95$.

- Expression for the asymptotic bias.

v---- source of the bias

$$\mathbb{E}\left[\widehat{\theta}_{h,LS}\right] - \theta_h = -\frac{1}{T-h}\sum_{j=1}^{T-h-1}\left(1-\frac{j}{T-h}\right)(\theta_{h+j}+\theta_{h-j}) + O\left(T^{-3/2}\right).$$

Moving average smooths it out.

- Given estimates of $\theta_{h\pm j}$ the responses at horizons $h \pm j$, one can correct estimates by adding back (an estimate of) the bias.

(a) $T = 50$ (b) $T = 100$

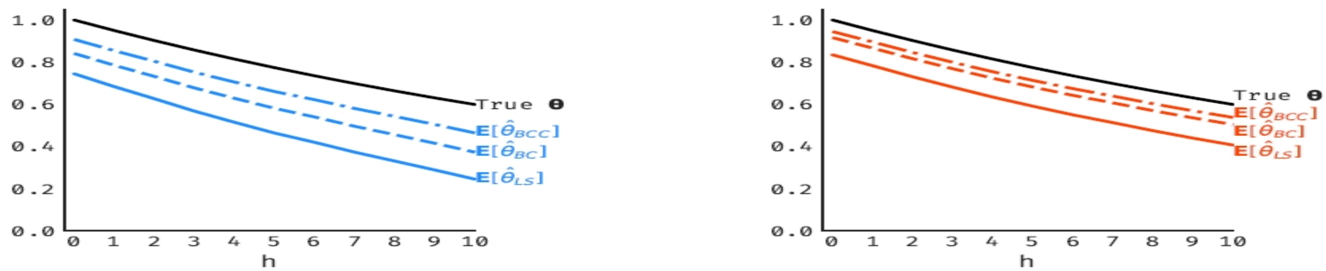


Figure 4: $\widehat{\theta}_{BC}$ and $\widehat{\theta}_{BCC}$ are closer than $\widehat{\theta}_{LS}$ to θ, on average, in our LPs without controls when y_t is an AR(1) with $\rho = 0.95$.

(a) $T = 50$ (b) $T = 100$

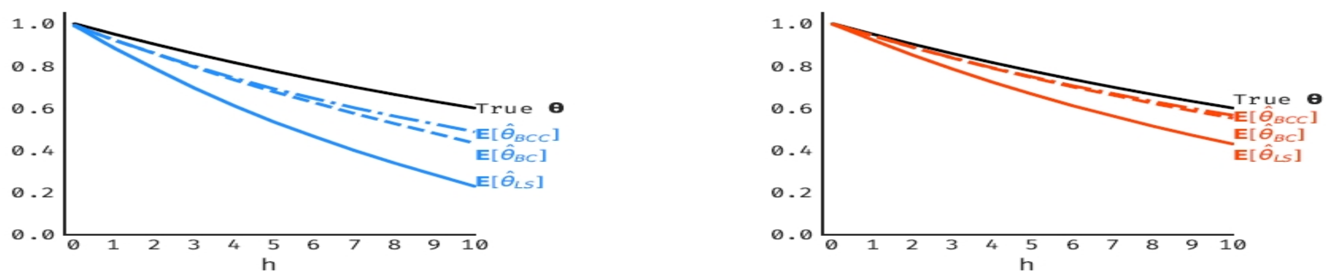


Figure 5: $\widehat{\theta}_{BC}$ and $\widehat{\theta}_{BCC}$ are closer than $\widehat{\theta}_{LS}$ to θ, on average, in our LPs with controls when y_t is an AR(1) with $\rho = 0.95$.

Bias correction with controls seems to do the trick.

Should one use LP or VAR in small samples?

DFM = Dynamic Factor Models

- Li et al. (2021): use a general DFM for DGP. Compute variance-bias tradeoff for LP and VAR dynamic responses.

- Least-squares LP and VAR estimators lie on opposite ends of the bias-variance spectrum: small bias and large variance for LPs, and large bias and small variance for VARs.

Typically, however with different horizons this may flip!

- Given h, there exists a weight ω on squared bias relative to variance in the loss function that would make a researcher indifferent between the two estimators (should care around four times more about squared bias than about variance).

- Shrinkage methods dramatically lower the variance of LP and VAR methods, at a moderate cost in bias. Unless researchers care almost exclusively about bias, penalized LP and Bayesian VAR preferable.

- Given a loss function, no single method dominates at all horizons. Penalized LP best at short horizons, while Bayesian VAR at long horizons.

- In the case of IV identification, the SVAR-IV estimator is heavily (median) biased, but provides substantial reduction in dispersion. Depending on the weight attached to bias, justifiable to use external IV methods despite their lack of robustness to non-invertibility (unlike internal IV methods).

Analytical illustration: Asymptotic bias and standard deviation

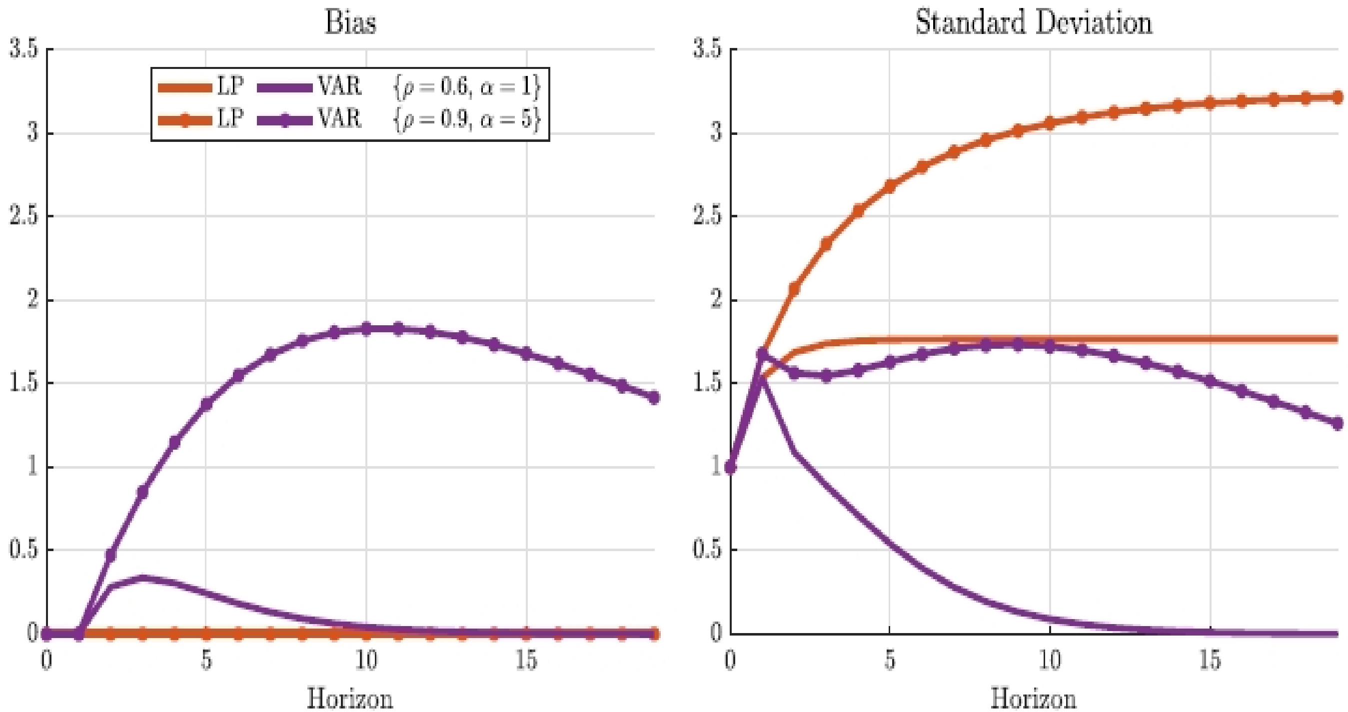


Figure 1: Asymptotic bias and standard deviation for LP (red) and VAR (purple) in the DGP (1) with $\sigma_1 = 1$ and $\{\rho = 0.6, \alpha = 1\}$ (no markers) or $\{\rho = 0.9, \alpha = 5\}$ (markers).

Analytical illustration: Indifference weight ω_h^*

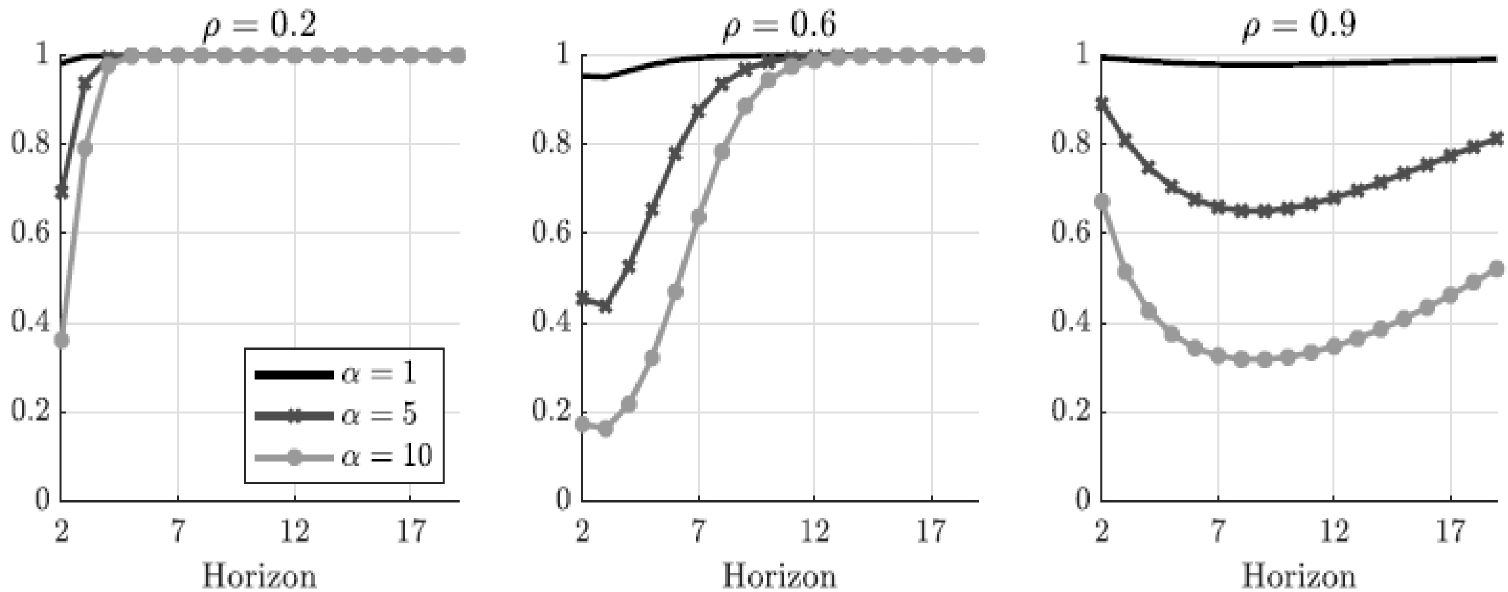


Figure 2: Weight $\omega = \omega_h^*$ in the asymptotic loss function $\omega \times \text{aBias}^2_{\hat{\theta}_h} + (1-\omega) \times \text{aVar}_{\hat{\theta}_h}$ that yields indifference between the LP and VAR estimator. LP is preferred whenever $\omega \geq \omega_h^*$. The three panels correspond to different values of $\rho \in \{0.2, 0.6, 0.9\}$; the three curves in each panel correspond to different values of $\alpha \in \{1, 5, 10\}$. All results are computed with $\sigma_2 = 1$. The figure omits the horizons $h \in \{0, 1\}$, at which the two estimation methods are (asymptotically) equivalent.

when some kind of penalization is added

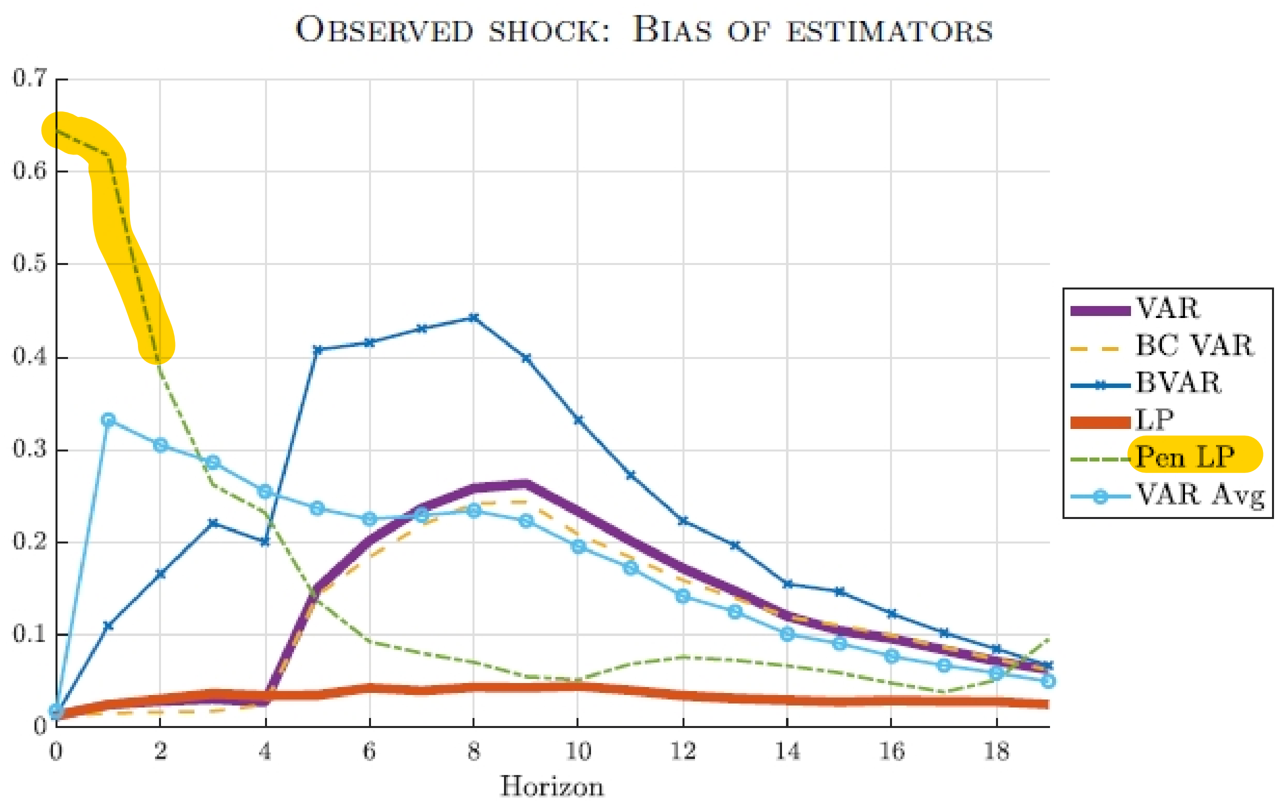


Figure 4: Median (across DGPs) of absolute bias of the different estimation procedures, relative to $\sqrt{\frac{1}{20}\sum_{h=0}^{19}\theta_h^2}$.

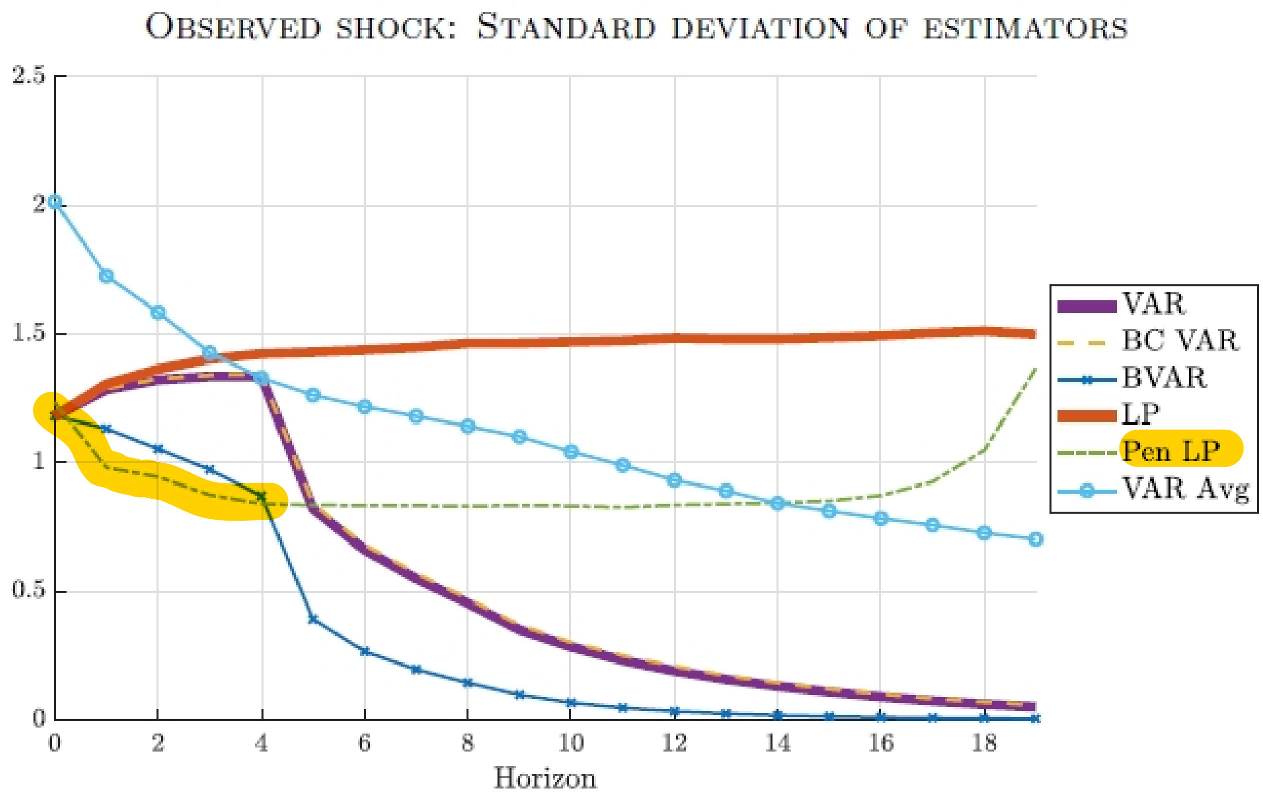


Figure 5: Median (across DGPs) of standard deviation of the different estimation procedures, relative to $\sqrt{\frac{1}{20}\sum_{h=0}^{19}\theta_h^2}$.

Measuring non-linear effects

$$Y_{t+h} = A^{h+1} Y_{t-1} + B_1^h \epsilon_{1t} + u_{t+h} \tag{82}$$

Want the effects of ϵ_t^+ and ϵ_t^- where ϵ_t^+ could be a positive or a large shock or a shock in a particular state and ϵ_t^- its complement.

- Normalization:

$$\begin{aligned} Y_{t+h} &= A^{h+1} Y_{t-1} + \gamma^{h+} Y_{1t}^+ + u_{t+h} && (83) \\ Y_{t+h} &= A^{h+1} Y_{t-1} + \gamma^{h-} Y_{1t}^- + u_{t+h} && (84) \end{aligned}$$

If instruments $Z_t^+ (Z_t^-)$ for $Y_{1t}^+ (Y_{1t}^-)$ are available, same LP-IV technology can be used.

- Can also consider non-linear effects in a SVAR-IV using the same idea.

- Different from allowing dynamics to be non-linear, i.e. $A^{h+1,+} \neq A^{h+1,-}$.

- Can test $\gamma^{h+} = \gamma^{h-}$ with a t-test (F-test).

General setup

- $y_t = \phi(\epsilon_t, \epsilon_{t-1}, \ldots)$. Non-linear model.

Taylor expansion first, then truncate.

- Volterra (Wold) expansion:

$$y_t = \sum_i \phi_i \epsilon_{t-i} + \sum_i \sum_j \phi_{ij} \epsilon_{t-i} \epsilon_{t-j} + \sum_i \sum_j \sum_k \phi_{ijk} \epsilon_{t-i} \epsilon_{t-j} \epsilon_{t-k} + \ldots \quad (85)$$

- VAR: $y_t = \sum_i \phi_i \epsilon_{t-i}$

- symmetric.
- state, history independent.

• General non-linear LP:

$$\begin{aligned} y_{t+h} &= a_h + b_1^h y_t + \ldots + b_P^h y_{t-p} + u_{t+h} \\ &+ q_1^h y_t^2 + r_1^h y_t^3 + \ldots \end{aligned} \tag{86}$$

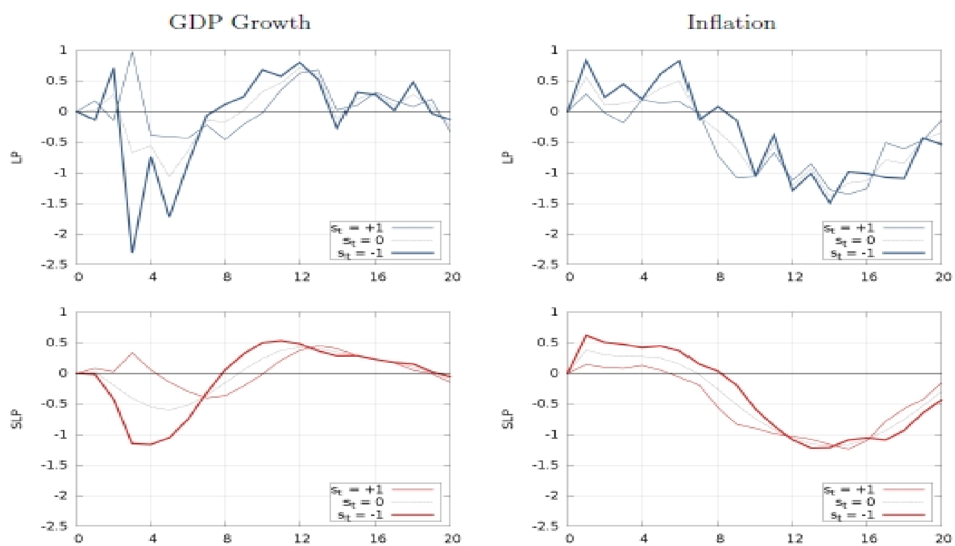
(Note: only y_t matter for IRFs)

• $IRF(t, h, d) = b_1^h d + q_1^h(2y_t d + d^2) + r_1^h(3y_t^2 d + 3y_t d^2 + d^3) + \ldots.$

• Non-linear LP could be asymmetric, state and history dependent, since y_{t-1} enters $IRF(t, h, d)$.

Example 4 *Barnichon and Brownlees (2019)*

Figure 10: State-dependent IR to a monetary shock



The figure displays the state-dependent IR of GDP growth (left panels) and inflation (right panels) to a monetary shock estimated using LP (top panels) and SLP (bottom panels). The state variable s_t respectively takes the values -1 (a recession), 0, and +1 (an expansion) units of $\sigma(s_t)$.