Structural VARs

Yu Bai

Bocconi University

April 14, 2022

Structural VARs

- Consider the structural VAR(p) model:

$$B_0 y_t = \sum_{\ell=1}^{p} B_\ell y_{t-\ell} + w_t,$$

  where $y_t = (y_{1t}, \cdots, y_{Nt})'$ is a $N \times 1$ vector, w_t is white noise with zero mean and covariance matrix Σ_w.
- The model is structural in the sense that w_t are mutually uncorrelated and have clear interpretations in terms of an underlying economic model.
- WLOG, we normalize the covariance matrix of the structural shock as $\Sigma_w = I_N$.

- Clearly, the model has a reduced-form representation:

$$y_t = \underbrace{B_0^{-1}B_1}_{\Pi_1} y_{t-1} + \cdots + \underbrace{B_0^{-1}B_p}_{\Pi_p} y_{t-p} + \underbrace{u_t}_{B_0^{-1}w_t},$$

  with $E(u_t u_t') = \Sigma_u = B_0^{-1} B_0^{-1'}$.
- The structural model is not identified.
- Recall the reduced-form IRF:

$$\Theta_h = \sum_{j=1}^{h} \Pi_j \Theta_{h-j}, \quad h = 0, 1, 2, \cdots, H$$

  where $\Theta_0 = I_N$ and $\Pi_\ell = 0$ for $\ell > p$.
- Then, the structural IRF can be defined as

$$\Theta_h^{\text{VAR}} = \Theta_h B_0^{-1} \tag{1}$$

A Premier on Theory of Identification

- We rely heavily on Rubio-Ramirez et al. 2010, RES.
- Let $m = Np$ and define $B_+' = [B_1 \ \cdots \ B_p]$, (B_0, B_+) are parameters of the structural model, $(\mathbf{\Pi}, \Sigma_u)$ are parameters of the reduced-form model
- Define $\mathbb{P}^S$ as the set of all structural parameters, $\mathbb{P}^R$ as the set of all reduced form parameters
- Define $g : \mathbb{P}^S \to \mathbb{P}^R$ be $g(B_0, B_+) = (B_+ B_0^{-1}, (B_0 B_0')^{-1})$, the relationship between the structural and reduced form parameters

Examples of identifying restrictions

- On the contemporaneous coefficients:

$$f(B_0, B_+) = B_0$$

- Short run restrictions on the IRFs:

$$f(B_0, B_+) = [\Theta_0' \ \Theta_1' \cdots \Theta_H']'$$

- Combination of short run and long run restrictions:

$$f(B_0, B_+) = [\Theta_0' \ \Theta_\infty']'$$

Identification

- We say that (B_0, B_+) and $(\tilde{B}_0, \tilde{B}_+)$ are observationally equivalent if and only if they imply the same distribution of y_t, $1 \leqslant t \leqslant T$.

Definition

A parameter point (B_0, B_+) is **globally identified** if and if there is no other parameter point that is observationally equivalent.

Definition

A parameter point (B_0, B_+) is **locally identified** if and if there is an open neighbourhood about (B_0, B_+) containing no other observationally equivalent parameter point.

- In linear Gaussian case, observationally equivalent implies that (B_0, B_+) and $(\tilde{B}_0, \tilde{B}_+)$ have the same reduced-form representation $(\mathbf{\Pi}, \Sigma)$.
- The structural of g: $g(B_0, B_+) = (B_+ B_0^{-1}, (B_0 B_0')^{-1})$ shows that (B_0, B_+) and $(\tilde{B}_0, \tilde{B}_+)$ have the same reduced-form representation if and only if there is an orthogonal matrix P such that $B_0 = \tilde{B}_0 P$ and $B_+ = \tilde{B}_+ P$.

Theory of identification

- Define Q_j for $j = 1, ..., N$ with $rank(Q_j) = q_j$ such that

$$Q_j f(A_0, A_+) e_j = 0$$

where e_j is the jth column of I_N. Wlog, we assume that $q = \sum q_j$ and $q_1 \geqslant q_2 \cdots \geqslant q_N$.

- Define the set of normalized structure parameters as N: $(B_0 D, B_+ D) \in V$ for diagonal D, we specify the set of restricted parameters as

$$R = \left\{ (B_0, B_+) \in U \bigcap V \,\middle|\, Q_j f(B_0, B_+) e_j = 0, 1 \leqslant j \leqslant N \right\} \qquad (2)$$

- We focus on exact identified SVAR.
- We say that SVAR is exactly identified if and only if for almost all $(\mathbf{\Pi}, \Sigma)$, there exists a unique $(B_0, B_+) \in R$ such that $g(B_0, B_+) = (\mathbf{\Pi}, \Sigma)$.
- We have the following theorem:

Theorem

Consider an SVAR with restrictions represented by R. The SVAR is exactly identified if and only if for almost every structural parameter point $(B_0, B_+) \in U$, there exists a unique $P \in O(N)$ such that $(B_0 P, B_+ P) \in R$.

- It can be shown that, if $f(B_0, B_+) = B_0$, there exists a unique $(B_0, B_+) \in R$ such that $g(B_0, B_+) = (\mathbf{\Pi}, \Sigma)$, then B_0 must be lower triangular.
- Lower triangular B_0 is often called recursive (Cholesky) identification.

Example: Global market for Crude Oil

- The example is taken from Kilian, 2009 AER.
- Let $y_t = (\Delta prod_t, rea_t, rpoil_t)'$, where $\Delta prod_t$ denotes the percent change in world crude oil production, rea_t is a measuring global real economic activity, and $rpoil_t$ is the log of the real price of oil.
- The data is monthly.
- The model is recursively identified.

$$\begin{pmatrix} u_t^{\Delta prod_t} \\ u_t^{rea_t} \\ u_t^{rpoil_t} \end{pmatrix} = \begin{bmatrix} b_0^{11} & 0 & 0 \\ b_0^{21} & b_0^{22} & 0 \\ b_0^{31} & b_0^{32} & b_0^{33} \end{bmatrix} \begin{pmatrix} w_t^{\Delta prod_t} \\ w_t^{rea_t} \\ w_t^{rpoil_t} \end{pmatrix}$$

- The identification assumptions imply:
  - (i) A vertical short-run oil supply curve;
  - (ii) A (downward-sloping) short-run demand curve

Inference on structural IRFs

- In practice, $\Pi_j, j = 1, ..., p$, B_0 are unknown and need to be estimated.
- We know that, asymptotic distributions of LS estimators take the following form:

$$\sqrt{T}\begin{pmatrix} \hat{\beta} - \beta \\ \text{vech}(\hat{\Sigma}_u) - \text{vech}(\Sigma_u) \end{pmatrix} \xrightarrow{d} N\left(0, \begin{bmatrix} \Sigma_{\hat{\beta}} & 0 \\ 0 & \Sigma_{\hat{\sigma}} \end{bmatrix}\right)$$

- In the special case of Gaussian $i.i.d.$ innovations, it can be shown that it is given by

$$\Sigma_\sigma = 2D_N^+(\Sigma_u \otimes \Sigma_u)(D_N^+)',$$

where $D_N^+ = (D_N' D_N)^{-1} D_N'$ and D_N is the duplication matrix such that

$$D_N \text{vech}(\Sigma_u) = \text{vec}(\Sigma_u).$$

- The structural IRFs are given by

$$\Theta_h^{\text{VAR}} = \Theta_h B_0^{-1},$$

  which is a smooth function of $g(\beta, \text{vech}(\Sigma_u))$.
- We can apply Delta method to obtain the asymptotic distributions of structural IRF estimates:

$$\sqrt{T}\big(g(\hat{\beta}, \hat{\sigma}) - g(\beta, \hat{\sigma})\big) \xrightarrow{d} N\Big(0, \frac{\partial g}{\partial \beta'} \Sigma_{\hat{\beta}} \frac{\partial g'}{\partial \beta} + \frac{\partial g}{\partial \sigma'} \Sigma_{\hat{\sigma}} \frac{\partial g'}{\partial \sigma}\Big)$$

Finite-order VAR models

- Lutkepohl (1990 RESTAT) derives the exact formulas of the quantities defined above, under the assumption that y_t is generated by a stationary VAR with known lag length p and white noise errors:

$$\sqrt{T}vec(\hat{\Theta}_h^{VAR} - \Theta_h^{VAR}) \xrightarrow{d} N(0, C_h \Sigma_\beta C_h' + \overline{C}_h \Sigma_\sigma \overline{C}_h') \qquad (3)$$

where

$$C_0 = 0, C_h = (B_0^{-1} \otimes I_N) G_h, \text{ with } G_h = \frac{\partial vec(\Phi_h)}{\partial \beta'}$$

$$\overline{C}_h = (I_N \otimes \Phi_h) H, \text{ with } H = \frac{\partial vec(B_0^{-1})}{\partial \sigma'}$$

- The exact formula of the score is

$$G_h = \sum_{m=0}^{h-1} J(B')^{h-1-m} \otimes \Phi_m$$
$$H = L_N' \{ L_N (I_{N^2} + K_{NN})(B_0^{-1} \otimes I_N) L_N' \}^{-1}$$
$$\Phi_m = J \mathbf{\Pi}^m J'$$

where K_{NN} is the commutation matrix such that for any $N \times N$ matrix G, $K_{NN} vec(G) = vec(G')$; L_m is the $m(m+1)/2 \times m^2$ elimination matrix, such that for any $m \times m$ matrix F, $vech(F) = L_m vec(F)$; and $J = [I_N \ 0 \ \cdots \ 0]$ is of dimension $N \times Np$.

Bootstrap intervals for structural IRFs

- The closed form solution derived above relies on the VAR errors being $i.i.d.$ Gaussian.
- The formulas are also inconvenient to use in practice.
- In practice, the confidence intervals are often obtained by bootstrap methods.
  - "The central idea underlying the bootstrap approach is that we approximate the distribution of the statistics of interest based on its sample analogue, allowing us to mimic the underlying sampling experiment."

Advantages of bootstrap methods

- They allow inference about smooth and differentiable $g(\beta, \text{vech}(\Sigma_u))$ even when closed-form solutions are not available or inconvenient to use.
- It remains asymptotically valid under weaker conditions, not required errors to be $i.i.d.$ Gaussian.
- Suitably constructed bootstrap confidence intervals tend to be more accurate in small samples than asymptotic approximations.

Standard residual-based recursive-design bootstrap

- There are several bootstrap methods on the table and they must be used properly (have asymptotic justification).
- Here, we focus on the first study of the use of bootstrap methods for structural IRFs, by Runkle, 1987 JBES.
- This is a residual-based recursive-design bootstrap method.

Residual-based recursive-design bootstrap: How it works?

- Let the model be VAR(p) with known p:

$$y_t = c + \sum_{\ell=1}^{p} \Pi_\ell y_{t-\ell} + u_t, \quad u_t \overset{i.i.d.}{\sim} F$$

  where F is generally unknown.
- Then, we can generate bootstrap DGP by

$$y_t^* = \hat{c} + \sum_{\ell=1}^{p} \hat{\Pi}_\ell y_{t-\ell}^* + u_t^*$$

  where $[\hat{c}, \hat{\Pi}_1, \cdots, \hat{\Pi}_p]$ denote the LS estimator and $u_t^* \overset{i.i.d.}{\sim} \hat{F}$.

- We need an estimate for $\hat{F}$.
- If we know that $u_t \overset{i.i.d.}{\sim} N(0, \Sigma_u)$, then $u_t^* \overset{i.i.d.}{\sim} N(0, \hat{\Sigma}_u)$.
- We can also do it in a nonparametric way. Let $\{\hat{u}_t\}_{t=1}^T$ be the initial regression residuals, then we obtain u_t^* by sampling with replacement from it.
- The MC results in Kilian, 1998b, ER show that there is no noticeable efficiency gain from imposing parametric assumptions even when it is true, but imposing the wrong parametric assumptions tends to undermine the accuracy of bootstrap inference.

- The initial conditions $(y^*_{-p+1}, \cdots, y^*_0)'$ can be sampled at random with replacement as a block of p consecutive vector valued observations from the observed data $\{y_t\}_{t=-p+1}^{T}$.
- If the lag order p is unknown, we can also estimate it in each bootstrap replications.
- The number of bootstrap replications required may be larger if we want to obtain confidence intervals precisely.
- After obtaining the implies bootstrap estimates $\hat{\theta}^*_{ik,h}$, percentile interval can be reported:

$$[\hat{\theta}^*_{ik,h,\gamma/2}, \hat{\theta}^*_{ik,h,1-\gamma/2}]$$

where γ is the quantile.

Local Projections

Jorda 2005 AER's critiques:

- VAR based approach is optimal if (1) represents the true DGP. However,
  - VARs represent a global approximation to the DGP: *best, linear, one-step ahead predictors*.
- IRFs in (**??**) are computed recursively. Misspecification errors are compounded with the horizon.
- Standard errors for impulse responses from VARs are complicated and difficult to compute.

Local Projections

- Recall that, IRFs are defined as

$$IR(t, h, d_i) = E_t(y_{t+s}|u_t = d_i) - E_t(y_{t+s}|u_t = 0) \quad h = 0, 1, 2, \cdots$$

- The key idea of local projection is that, multi-step forecasts, $E_t(y_{t+s}|-)$, can also be obtained with direct forecasting models, specifically

$$y_{t+h} = f_0 + F_1 y_t + F_2 y_{t-1} + \cdots + F_q y_{t-q+1} + e_{t+h}, \quad h = 1, \cdots, H \quad (4)$$

- Then, we have

$$IR(t, h, d_i) = \Phi^{LP} = F_1 d_i, \quad h = 1, \cdots, H \quad (5)$$

which we name (5) as the impulse responses from the local(-linear) projections.

- Lag length q in (4) needs not to be common in each horizon h.
- Choice of functional forms does not need to be linear as in (4). We can introduce polynomial terms on y_t:

$$y_{t+h} = F_1 y_t + Q_1 y_t^2 + F_2 y_{t-1} + \cdots + F_q y_{t-q+1} + e_{t+h}$$

- The corresponding structural impulse responses are therefore

$$\Theta_h^{LP} = IR(t, h, d_i) B_0^{-1}, \quad h = 0, 1, \cdots, H \tag{6}$$

Relation to VARs

Theorem

If y_t is a VAR(p) process then its h-step predictive regression is a predictive VAR(p) with u_t a MA(h − 1) process and $F_1 = \Theta_h = IRF(h)$.

Proof of the theorem

Since y_t is a VAR(p) process,

$$y_{t+h} = \Pi_1 y_{t+h-1} + \Pi_2 y_{t+h-2} + \cdots + \Pi_p y_{t+h-p} + e_t.$$

We then substitute out the first lag. We find

$$\begin{aligned} y_{t+h} &= \Pi_1(\Pi_1 y_{t+h-2} + \Pi_2 y_{t+h-3} + \cdots + \Pi_p y_{t+h-p-1} + e_{t-1}) \\ &\quad + \Pi_2 y_{t+h-2} + \cdots + \Pi_p y_{t+h-p} + e_t \\ &= (\Pi_1\Pi_1 + \Pi_2) y_{t+h-2} + (\Pi_1\Pi_2 + \Pi_3) y_{t+h-3} \\ &\quad + \cdots + \Pi_p\Pi_p y_{t+h-p-1} + \Pi_1 e_{t-1} + e_t. \end{aligned}$$

We continue making substitutions. After the first substitution, the error term is MA(1). Then, after $h-1$ iteration, the error has an MA($h-1$) structure. The second part of the theorem follows immediately from the definition of the IRF.

Local Projections: Estimation and inference

- Let y_j, $j = H, \cdots, 0, -1, \cdots, -q$, be the $(T - q - H) \times N$ matrix of stacked observations for y_{t+j}, $X = (y_{-1} \ \cdots y_{-q})$ is of dimension $(T - q - H) \times Nq$, define the projection matrix $M_X = I - X(X'X)^{-1}X'$, by F-W-L Theorem, we have

$$\hat{F}_1 = \left(y_h' M_X y_0\right)\left(y_0' M_X y_0\right)^{-1} \tag{7}$$

- Then, by Proposition 2 in Jorda (2009, RESTAT), under some regularity conditions, we have

$$\sqrt{T}(vec(\hat{F}_1) - vec(F_1)) \xrightarrow{d} N(0, \Omega_F) \tag{8}$$

where

$$\Omega_F = p\lim\left(\frac{1}{T} y_0' M_X y_0\right)^{-1} \otimes \Sigma_e \tag{9}$$

and Σ_e can be consistently estimated by HAC.

Local Projections: Comments

- Starting with Jorda 2005 AER, local projections have become an increasingly widespread alternative econometric approach.
- Advantages (Jorda 2005 AER):
  - IRFs may be estimated directly by linear LS regressions, simplifying the analysis.
  - Inference for the impulse responses is straightforward and does not require appealing to the delta method.
  - Local projection estimates of the structural impulse responses are more robust to model misspecification than conventional VAR estimates.

Local Projections: Comments

- Repeat the predictive regression here:

$$y_{t+h} = F_1 y_t + F_2 y_{t-1} + \cdots + F_q y_{t-q+1} + e_{t+h}, \quad h = 1, \cdots, H$$

  Then, for H horizons, it requires to estimate HqN^2 parameters in the mean equation, but only HN^2 parameters are of direct interests.
- For VAR analysis, only pN^2 parameters need to be estimated. The number of parameters to be estimated does not increase with the impulse response horizon.
- If VAR model is misspecified because there are omitted variables, local projection estimator suffers from the same omitted variable problem.
  - For example, Alessi and Kerssenfischer (2019, JAE) finds that puzzling effects of monetary policy shocks from VAR are mainly caused by missing information, not by identification scheme.

Local Projections: Comments

- It is also not clear whether LP is more robust to model misspecification than conventional VAR.
- As we have seen, MA($h-1$) structure of the error terms require the assumption that true DGP is VAR(p).

Local Projections: Comments

- It is correct to say that inference for $\hat{F}_1$ can be directly obtained by (7).
- However, as economists, we are interested in Θ_h^{LP} defined in (6). B_0^{-1} is unknown and has to be estimated.
- Thus, inference still requires Delta method. Asymptotic variance has to be adjusted:

$$(B_0^{-1'} \otimes I_N)\Omega_F(B_0^{-1} \otimes I_N) + \overline{G}_h \Sigma_\sigma \overline{G}_h \tag{10}$$

where

$$\overline{G}_h = (I_N \otimes \Phi_h^{LP}) \frac{\partial vec(B_0^{-1})}{\partial \sigma'}$$

- This means that the second term of asymptotic variance is exactly the same as in standard VAR.

Local Projections: Comments

- On nonlinearity:
  - Accurate nonlinear approximation may require large number of terms. It is also not clear how to obtain an external estimate of the structural impact multiplier matrix in that case.
  - It might be correct that LP delivers more precise reduced form IRFs and is more prone to nonlinearity. But, in forecasting literature, Marcellino et al. 2006 JoE and Pesaran et al. 2011 JoE find that, if (V)AR model includes enough lags, iterated forecast by (V)AR models have better forecast accuracy than direct forecast.
- On inference:
  - It is possible to implement bootstrap procedure in LP context, but the procedures introduced above cannot be used. See, Kilian and Kim, 2011 RESTAT and references therein for more details.

Local Projections: Recent advances

- Plagborg-Moller and Wolf, 2021 ECTA prove that local projections and Vector Autoregressions estimate the same impulse responses in population.
- Montiel Olea and Plagborg-Moller 2021 ECTA provide some new results on the robustness of LPs.
- Herbst and Johannsen, 2020 investigate the finite sample bias of LPs.
- Check here for a more comprehensive simulation study: Li et al. 2021.

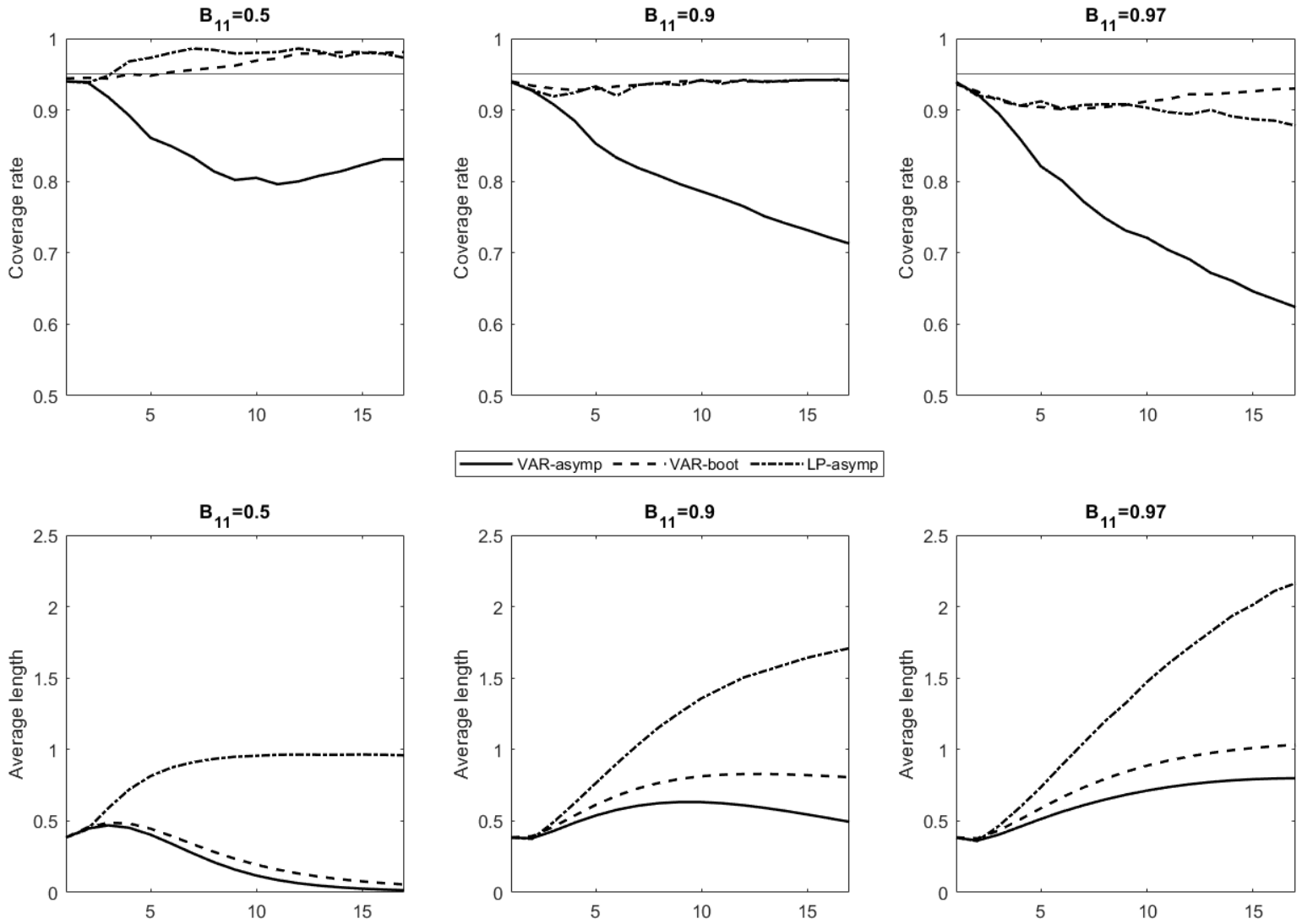
Monte-Carlo Study

- The DGP is taken from Kilian and Kim, 2011 RESTAT.
- Consider the following bivariate VAR(1), $t = 1, \cdots, T$:

$$y_t = \begin{pmatrix} B_{11} & 0 \\ 0.5 & 0.5 \end{pmatrix} y_{t-1} + e_t, \quad e_t \overset{i.i.d.}{\sim} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.3 \\ 0.3 & 1 \end{pmatrix} \right)$$

  where $B_{11} \in \{0.5, 0.9, 0.95\}$ and $T = 100$ with $M = 1000$ replications.
- Lag length is chosen by AIC. For local projection, lag length is selected at each horizon $h = 1, \cdots, H$.
- We compare both coverage rates and length of the 95% confidence interval for the shock of 1st variable to the 2nd variable.

Monte-Carlo Study



Monte-Carlo Study

- The coverage rates of VAR drop quickly with increasing horizon (more when series are persistent).
- Local projection does not deteriorate steadily as horizon increases, but its overall coverage rates decline with increasing B_{11}.
- The better coverage rates from local projection is mainly due to wider confidence interval.

Forecast error variance decomposition

- For a VAR process, the h-step ahead forecast error is

$$y_{t+h} - y_{t+h|t} = \sum_{i=0}^{h-1} \Pi_1^i u_{t+h-i} = \sum_{i=0}^{h-1} \Theta_i w_{t+h-i}$$

  where $u_t = B_0^{-1} w_t$.

- Then, MSE at horizon h is given by

$$\begin{aligned} \text{MSE}(h) &= \sum_{i=0}^{h-1} \Pi_1^i \Sigma_u (\Pi_1^i)' \\ &= \sum_{i=0}^{h-1} \Theta_i \Sigma_w \Theta_i' \\ &= \sum_{i=0}^{h-1} \Theta_i \Theta_i' \end{aligned}$$

- Let $\theta_{kj,h}$ be the kj^{th} element of Θ_h. Then, the contribution of shock j to MSE of y_{kt}, $k = 1, ...N$, at horizon h, is given by

$$\text{MSE}_j^k(h) = \theta_{kj,0}^2 + \cdots + \theta_{kj,h-1}^2$$

and the total MSE of y_{kt}, at horizon h, is

$$\text{MSE}^k(h) = \sum_{j=1}^{N} \text{MSE}_j^k(h) = \sum_{j=1}^{N} (\theta_{kj,0}^2 + \cdots + \theta_{kj,h-1}^2)$$

- Dividing both sides above by $\text{MSE}^k(h)$, we have the forecast error variance decomposition:

$$1 = \frac{\text{MSE}_1^k(h)}{\text{MSE}^k(h)} + \cdots + \frac{\text{MSE}_N^k(h)}{\text{MSE}^k(h)}$$

Historical decompositions

- Sometimes we are interested instead in quantifying how much a given structural shock explains of the historically observed fluctuations in the VAR variables.
- Consider a weakly stationary vector process y_t:

$$y_t = \sum_{s=0}^{t-1} \Theta_s w_{t-s} + \sum_{s=t}^{\infty} \Theta_s w_{t-s}$$

- Given the fact that MA coefficients die out, we can drop out the second term:

$$y_t \approx \sum_{s=t}^{\infty} \Theta_s w_{t-s}$$

and denote this approximation by

$$\hat{y}_t = \sum_{s=t}^{\infty} \Theta_s w_{t-s}$$

- Then, jth shock to variable i can be computed as

$$\hat{y}_{it}^{(j)} = \sum_{i=0}^{t-1} \theta_{ij,i} w_{j,t-i}$$

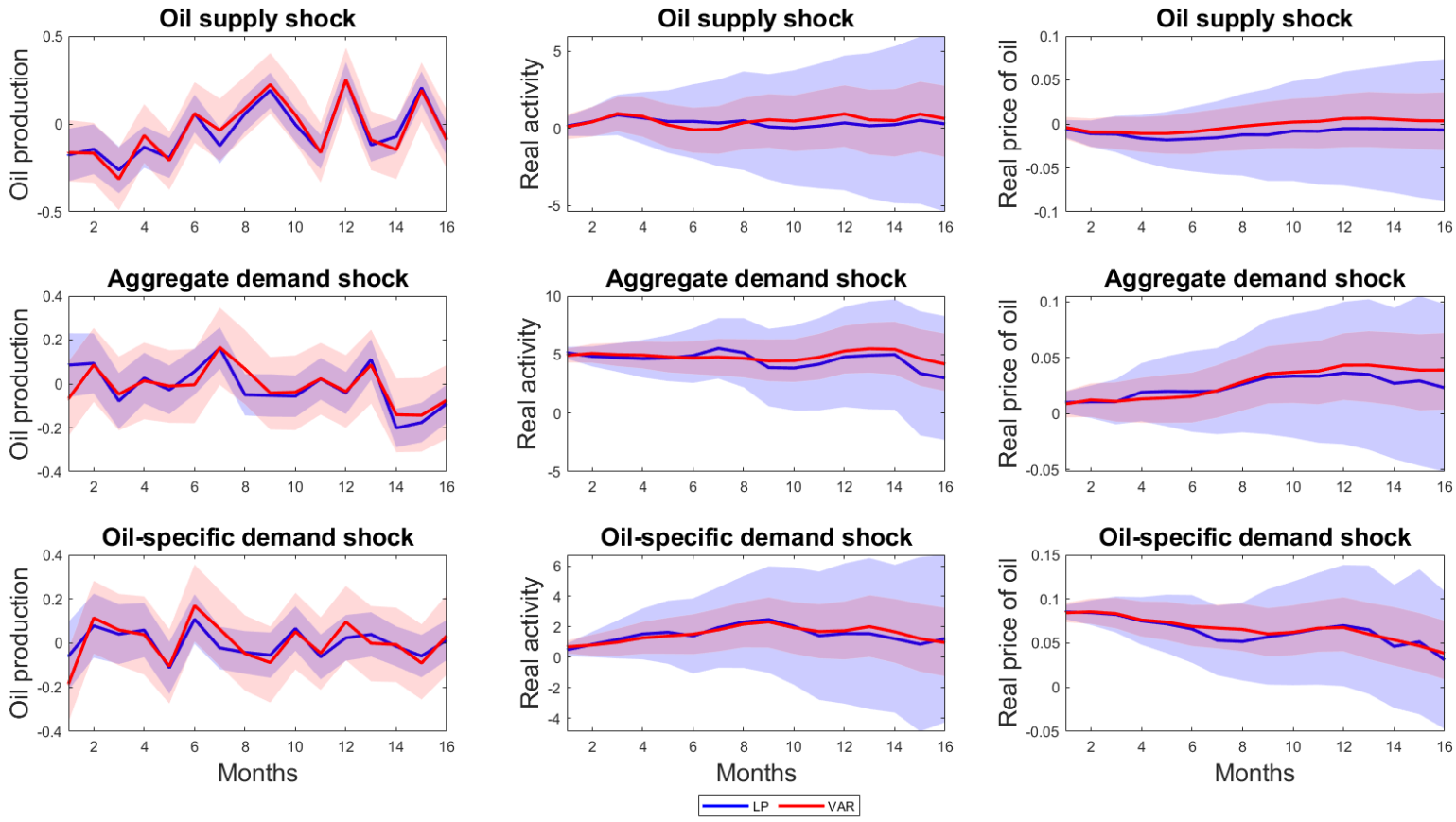
- In practice, both MA coefficients and structural shocks are replaced by estimated counterparts.
- The value for $\hat{y}_{it}$ is obtained as the sums

$$\hat{y}_{it} = \sum_{j=1}^{N} \hat{y}_{it}^{(j)}$$

- It is worthy to mention that number of approximations needed generally depends on the persistence of VAR variables.

An empirical application

- Kilian, 2009 AER: a 3-variable structural VAR model of the global crude oil market
- Variables used: $y_t = (\Delta prod_t, rea_t, rpoil_t)'$, where $\Delta prod_t$ denotes the percent change in world crude oil production, rea_t is a measuring global real economic activity, and $rpoil_t$ is the log of the real price of oil
- The sample period is 1973:01-2007:12.
- The model is recursively identified.
- Lag length is set to 24. For LP, lag length is selected based on AIC criteria.


Oil supply shock
Oil production
Real activity
Real price of oil
Aggregate demand shock
Oil-specific demand shock
Months
LP
VAR

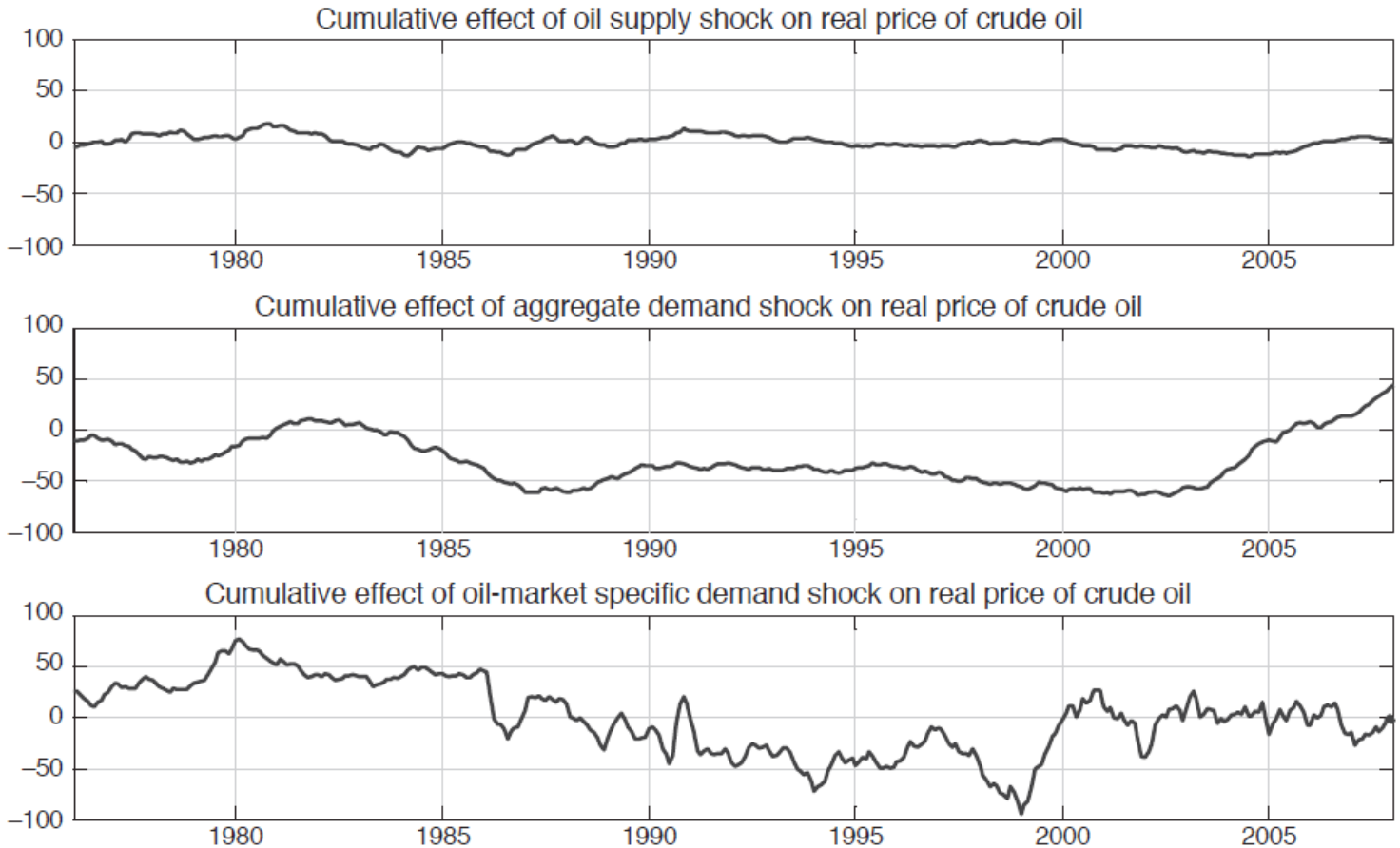


FIGURE 4. HISTORICAL DECOMPOSITION OF REAL PRICE OF OIL

(1976:1–2007:12)

Summary of main findings

- The effect of an unanticipated aggregate demand expansion on global real economic activity is very persistent and highly significant.
- Unanticipated oil market–specific demand increases have an immediate, large, and persistent positive effect on the real price of oil that is highly statistically significant.
- Unanticipated oil supply disruptions have only a small positive effect on the real price of oil.
- Oil supply shocks historically have made comparatively small contributions to the real price of oil. The biggest contributions are due to the aggregate demand shock and the oil market–specific demand shock.