SVARs, Local Projections, and Dynamic Causal Effects

(FIRST DRAFT: 3/14/2020)

A few additional references:


Structural Vector Autoregression (SVAR)

$Y_t$ is an $n \times 1$ vector of observables ($n$ typically 'small')

VAR dynamics: $E(Y_t \mid \text{lags of } Y_t) = A_1 Y_{t-1} + \ldots + A_p Y_{t-p}$.

so that

$Y_t = A_1 Y_{t-1} + \ldots + A_p Y_{t-p} + \eta_t$ or $A(L) Y_t = \eta_t$.

$\eta_t = 1$-period ahead forecast error.

No constant term for notational convenience.
VAR representation: \( Y_t = A_1 Y_{t-1} + \ldots + A_p Y_{t-p} + \eta_t \) or \( A(L) Y_t = \eta_t \).

Vector Moving Average (VMA) representation: \( Y_t = \eta_t + C_1 \eta_{t-1} + C_2 \eta_{t-2} + \ldots = C(L) \eta_t \)

where \( C(L) = A(L)^{-1} \)

Note: \( C(L) = C_0 + C_1 L + C_2 L^2 + \ldots \) and \( C_0 = I \)
SVAR (Sims (1980)): Why do we make forecast errors?

\[ \eta_t = H \varepsilon_t \]

where \( \varepsilon_t \) are 'structural' shocks. (Shocks interpretable in the context of particular theoretical economic models).

Structural MA:

\[ Y_t = C(L) \eta_t = C(L)H \varepsilon_t = D(L) \varepsilon_t \]

SVAR:

\[ A(L)Y_t = \eta_t = H \varepsilon_t \]

or

\[ B(L)Y_t = \varepsilon_t, \text{ with } B(L) = H^{-1}A(L) \]
From SMA: \( Y_t = D_0 \epsilon_t + D_1 \epsilon_{t-1} + \ldots \) with \( D_k = C_k H \)

Notes:

(1) \( C_0 = I \), so \( D_0 = H \)

(2) \( \frac{\partial Y_{i,t+k}}{\partial \epsilon_{j,t}} = D_{k_{ij}} \). (These are "impulse responses" or "dynamic causal effects" or 'dynamic multipliers' … )
Issues:

1. $E(Y_t \mid \text{lags of } Y_t) = A_1 Y_{t-1} + \ldots + A_p Y_{t-p}$. Reasonable?

2. $C(L) = A(L)^{-1}$; when is this a well-defined one-sided inverse?

3. Estimation of $A(L)$ and $C(L)$. When do usual large-sample linear properties obtain?

4. $\eta_t = H \varepsilon_t$ with $H$ non-singular. Reasonable?

5. Identification of $H$.

6. Properties of $\hat{C} \hat{H}$.
Issues:

1. $E(Y_t \mid \text{lags if } Y_t) = A_1 Y_{t-1} + \ldots + A_p Y_{t-p}$. Reasonable?

2. $C(L) = A(L)^{-1}$; when is this a well-defined one-sided inverse?

3. Estimation of $A(L)$ and $C(L)$. When do usual large-sample linear properties obtain.

These are relatively straightforward extensions of the analysis that we have done for the finite-order univariat model.

Specifically for (2) this requires that the roots of $|A(z)|$ are outside unit circle, i.e., that the difference equation is stable. Equivalently, this requires that the eigenvalues of the companion matrix be larger than 1 in modulus.

As in the univariate model, OLS/GMM estimators for the coefficients in $A(L)$ and $\Sigma_{\eta \eta}$ will be consistent asymptotically normal when $\eta_t$ are MDS with appropriate moments, $Y$ is stationary, and so forth.

The coefficients in $C(L)$ are continuous functions of the coefficients in $A(L)$ (recall $C(L) = A(L)^{-1}$), so, using $\delta$-method arguments, they too will be asymptotically normal.
Gerzensee, Econometrics Week 3, March 2020

Issue (4): $\eta_t = H \varepsilon_t$ with $H$ non-singular. Reasonable?

In some cases NO:

Non-invertibility: Static problem $H$ is $n_Y \times n_\varepsilon$. What if $n_\varepsilon > n_Y$?

Dynamics:

Invertibility (required here): Can I determine $\varepsilon_t$ from current and lagged $Y$.

'Recoverability' (Chahrou and Jurado (2017), Plagbor-Moller and Wolf (2018)): Can I determine $\varepsilon_t$ from current, lagged and future $Y$.

Simplest example:

$$Y_t = \varepsilon_t - \theta \varepsilon_{t-1}$$

$$\varepsilon_t = \sum_{j=0}^{\infty} \theta^j Y_{t-j} + \theta^\varepsilon_0$$ (so invertible when $|\theta| < 1$).

Also

$$\varepsilon_t = -\theta^{-1} \sum_{j=1}^{T-1} \theta^{-j} Y_{t+j} + \theta^{-T} \varepsilon_T$$

(so recoverable as long as $|\theta| \neq 1$)
More complicated example:  
(Fernandez-Villaverde, Rubio-Ramirez, Sargent and Watson (2007))

Consider a structural model of the form:

\[ y_{t+1} = Cx_t + Dw_{t+1} \]

\[ x_{t+1} = Ax_t + Bw_{t+1} \]

Invertibility:

When is \( \text{var}\left( w_t \mid \left\{ y_{t+j} \right\}_{j=-\infty}^{0} \right) = 0 \) ?

It is straightforward to show that this obtains when the eigenvalues of \( (A - BD^{-1}C) \)
are less than 1 in modulus. (This is a sufficient condition for invertibility)

Recoverability: When is \( \text{var}\left( w_t \mid \left\{ y_{t+j} \right\}_{j=-\infty}^{\infty} \right) = 0 \) ?
Gerzensee, Econometrics Week 3, March 2020

Issue 5: Identification of $H$

$$\eta = H \varepsilon \Rightarrow \Sigma_{\eta\eta} = H \Sigma_{\varepsilon\varepsilon} H'$$

$\Sigma_{\eta\eta}$ estimable from data, so question is whether there is a unique solution for $H$ and $\Sigma_{\varepsilon\varepsilon}$ from $\Sigma_{\eta\eta} = H \Sigma_{\varepsilon\varepsilon} H'$.

'Order condition' .. count equations and unknowns.

- $n(n+1)/2$ elements in $\Sigma_{\eta\eta}$ (number of equations)

- $n^2 + n(n+1)/2$ in $H$ and $\Sigma_{\varepsilon\varepsilon}$ (number of unknowns) .. $n^2$ too many parameters. Additional Restrictions:
  - Uncorrelated Structural Shocks: Restrict $\Sigma_{\varepsilon\varepsilon}$ to be diagonal: $n^2 + n$ unknowns .. still we have $n(n+1)/2 - (n^2 + n) = n(n+1)/2$ too many parameters.
Scale normalization

scalar model: \( \eta_i = H \xi_i \) ('units' of \( \xi_i \) are not identified)

2 normalizations:  
(1) \( \sigma_e = 1 \)

(2) \( H_{ii} = 1 \) for \( i = 1, \ldots, n \)
consistent with standard theory, there is a small decline in the consumer price index production that reaches a trough roughly a year and a half after the shock. Similarly, there is a significant decline in the industrial rate. Consistent with conventional theory, there is a significant decline in the CPI of 17.5. Both values are safely above the threshold suggested by Stock et al. As the top left panel shows, a one standard deviation surprise monetary tight shock with an external instrument, the one-year government bond rate increases by 0.2 percent, and the one-year excess bond premium with an external instrument is 0.1 percent. As noted earlier, we use the three-stage regression of the one-year bond rate residual on FF4. As a check to ensure that this instrument is valid, we report the month ahead funds rate future surprise FF4 to identify monetary policy shock. As a first-stage regression of the one-year bond rate residual on FF4, we do not impose zero restrictions on the contemporaneous responses of output and inflation. The identification of the monetary policy shock is entirely due to the bootstrapping procedure. Thereby, we avoid any potential “generated regressor” problem.

Standard deviation normalization: Gertler Karadi (2015) – IRF or Monetary Policy Shock

<table>
<thead>
<tr>
<th>Percent</th>
<th>Percent</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.4</td>
<td>0.2</td>
<td>0.0</td>
</tr>
<tr>
<td>-0.2</td>
<td>0.0</td>
<td>0.2</td>
</tr>
<tr>
<td>0.4</td>
<td>0.2</td>
<td>0.0</td>
</tr>
<tr>
<td>-0.2</td>
<td>0.0</td>
<td>0.2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Percent</th>
<th>Percent</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td>0.0</td>
<td>-0.2</td>
</tr>
<tr>
<td>-0.2</td>
<td>0.0</td>
<td>-0.4</td>
</tr>
<tr>
<td>0.0</td>
<td>0.0</td>
<td>-0.2</td>
</tr>
<tr>
<td>-0.2</td>
<td>0.0</td>
<td>-0.4</td>
</tr>
</tbody>
</table>
Scale normalization does not matter in population. It will matter for inference.

Moving from one normalization to another involves dividing by \( \hat{H}_{ii} \) or \( \hat{\sigma}_e \).

We will primarily use normalization on elements of H: Diagonal elements of H are unity

- Alternatives used in applied work:
  - \( \Sigma_{ee} = I \)
  - Diagonal elements of \( H^{-1} = I \). (Scale normalization used in classical simultaneous equations literature.)
Back to counting: with scale normalization the model needs only $n(n-1)/2$ additional restrictions.

Example: VAR(1) with $n = 3$

\[
Y_t = AY_{t-1} + \begin{bmatrix}
1 & H_{12} & H_{13} \\
H_{21} & 1 & H_{23} \\
H_{31} & H_{32} & 1
\end{bmatrix} \begin{bmatrix}
\varepsilon_{1t} \\
\varepsilon_{2t} \\
\varepsilon_{3t}
\end{bmatrix}
\]
\[ Y_t = AY_{t-1} + \begin{bmatrix} 1 & H_{12} & H_{13} \\ H_{21} & 1 & H_{23} \\ H_{31} & H_{32} & 1 \end{bmatrix} \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{bmatrix} \]

Timing restriction example:  \[ Y_t = AY_{t-1} + \begin{bmatrix} 1 & 0 & 0 \\ H_{21} & 1 & 0 \\ H_{31} & H_{32} & 1 \end{bmatrix} \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{bmatrix} \]
Long-run restriction example:

Arithmetic: Let $D = A(L)^{-1}H$ and let $Z_t = (1-L)^{-1}Y_t$ then

$$\lim_{k \to \infty} \frac{\partial Z_{t+k}}{\partial \epsilon_{ij,t}} = D_{ij}.$$ 

Restrict $H$ so that $D_{ij}$ has $n(n-1)/2$ zeros.
Identification of one shock, say $\varepsilon_{1t}$ and its effect on $Y_{t+k}$

Recall: $Y_t = C(L)\eta_t = C(L)H\varepsilon_t$ with $C(L) = A(L)^{-1}$

Thus

$$Y_t = C(L)\begin{bmatrix} H_1 & H_\bullet \end{bmatrix} \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{\bullet} \end{bmatrix} = C(L)H_1\varepsilon_{1t} + \text{distributed lag of } \varepsilon_{\bullet},$$

where $\bullet$ denotes elements 2 through $n$

To identify the effect of $\varepsilon_1$ on $Y_{t+k}$ we need only identify the first column of $H$.

And, if $H_1$ is known ('identified') and $H$ is invertible, then it turns out $\varepsilon_{1t}$ can be 're-constructed' from $\eta_t$ (up to scale). (This requires a bit of algebra to show.)
Identification of $H_1$

$$Y_t = AY_{t-1} + \begin{bmatrix} 1 & H_{12} & H_{13} \\ H_{21} & 1 & H_{23} \\ H_{31} & H_{32} & 1 \end{bmatrix} \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \\ \epsilon_{3t} \end{bmatrix}$$

Timing restriction example: $Y_t = AY_{t-1} + \begin{bmatrix} 1 & 0 & 0 \\ H_{21} & 1 & H_{23} \\ H_{31} & H_{32} & 1 \end{bmatrix} \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \\ \epsilon_{3t} \end{bmatrix}$

$\epsilon_t = \eta_1$, and $H_1$ is identified by regressing $\eta_t$ onto $\eta_{1t}$.

Similar kinds of results for other timing restrictions, long-run restrictions, etc.
Some alternative identification schemes that are used:

(1) Heteroskedasticity

(2) Sign Restrictions

(3) External Instruments ('Proxy variables')
Identification by Heteroskedasticity (Rigobon (2003), Rigobon and Sack (2003, 2004))

Idea: $\Sigma^1_{ee}$ and $\Sigma^2_{ee} \Rightarrow$

$$\Sigma^1_{\eta\eta} = H \Sigma^1_{ee} H'$$ and $$\Sigma^2_{\eta\eta} = H \Sigma^2_{ee} H'$$

"Order condition" (counting equations and unknowns):

Number of equations (unique elements in $\Sigma^1_{\eta\eta}$ and $\Sigma^2_{\eta\eta}$): $n(n+1) = n^2 + n$

Number of unknowns: $(H, \Sigma^1_{ee}$ and $\Sigma^2_{ee})$: $(n^2 - n) + 2n = n^2 + n$. 
Note: 'rank condition' .. relative variances of $\epsilon_i$ must change to get independent information on elements of $H$.

Potentially powerful tool.

Generalizes to time-varying conditional heteroskedasticity (ARCH/GARCH, Stochastic volatility, etc.)
Example:

\[
\begin{pmatrix}
\Sigma_i^{ij} & \Sigma_i^{j_1 \eta_2} \\
\Sigma_i^{j_2 \eta_1} & \Sigma_i^{j_2 \eta_2}
\end{pmatrix} =
\begin{pmatrix}
1 & \hat{H}_{12} \\
\hat{H}_{21} & 1
\end{pmatrix}
\begin{pmatrix}
\sigma^2_{\epsilon_{1j}} & 0 \\
0 & \sigma^2_{\epsilon_{2j}}
\end{pmatrix}
\begin{pmatrix}
1 & \hat{H}_{21} \\
\hat{H}_{12} & 1
\end{pmatrix}, \quad j = 1, 2.
\]

Algebra \Rightarrow

\[H_{21} = \frac{\Sigma_{\eta_1 \eta_2} - \Sigma_{\eta_1 \eta_2}}{\Sigma_{\eta_1 \eta_1} - \Sigma_{\eta_1 \eta_1}}\]

Estimator:

\[\hat{H}_{21} = \frac{\hat{\Sigma}_{\eta_1 \eta_2} - \hat{\Sigma}_{\eta_1 \eta_2}}{\hat{\Sigma}_{\eta_1 \eta_1} - \hat{\Sigma}_{\eta_1 \eta_1}}\]

\[\hat{H}_{21} = \frac{\hat{\Sigma}_{\eta_1 \eta_2} - \hat{\Sigma}_{\eta_1 \eta_2}}{\hat{\Sigma}_{\eta_1 \eta_1} - \hat{\Sigma}_{\eta_1 \eta_1}}\]

Denominator: \(\hat{\Sigma}_{\eta_1 \eta_1} - \hat{\Sigma}_{\eta_1 \eta_1} = \left(\Sigma_{\eta_1 \eta_1} - \Sigma_{\eta_1 \eta_1}^1\right) + \text{Sampling Error} \left(\hat{\Sigma}_{\eta_1 \eta_1} - \hat{\Sigma}_{\eta_1 \eta_1}^1\right)\)

Estimator will have poor sampling properties when denominator is noisy:

\[\text{Sampling Error} \left(\hat{\Sigma}_{\eta_1 \eta_1} - \hat{\Sigma}_{\eta_1 \eta_1}^1\right) \text{ is big relative to } \left(\Sigma_{\eta_1 \eta_1} - \Sigma_{\eta_1 \eta_1}^1\right).\]

Or, (1) when change in variance is small or one or both of the samples is small.
Inequality (Sign) Restrictions (Faust (1998), Uhlig (2005), Baumeister and Hamilton (2015) )

Typical identifying restrictions: \( R \times \text{vec}(H) = r \) where \( R \) and \( r \) are pre-specified can be computed from the data. (Or \( RH_1 = r \), when focused on a single shock.)

Inequality Restrictions: \( R \times \text{vec}(H) \geq r \).

This 'set identified' the impulse responses.
Determining the identified set: A computational method using $\Sigma_{ee} = I$ normalization.

$\Sigma_{\eta\eta} = H\Sigma_{ee}H' = HH'$, so $H$ is a matrix square root of $\Sigma_{\eta\eta}$.

$H = \sum^{1/2}_{\eta\eta} C$ where $\sum^{1/2}_{\eta\eta}$ is any particular matrix square root (e.g., the Cholesky factor) and $C$ is an orthonormal matrix (so $CC' = I$).

1. Compute $\sum^{1/2}_{\eta\eta}$
2. For a particular value of $C$, compute $H = \sum^{1/2}_{\eta\eta} C$.
3. Check to see if $R \times \text{vec}(H) \geq r$. If so, keep $H$. If not discard $H$.
4. Repeat step 2 for all possible values of $C$.
5. The resulting values of $H$ from (3) are the set of values of $H$ that are identified by the inequality restriction.
Generic Notation

$Y$: Random variables being studied

$\mu$: Parameters that characterize pdf $f(Y \mid \mu)$

$\theta$: Parameters of interest.

Point identified models: $\theta = g(\mu)$

Set identified models: $\theta \in G(\theta)$
Example:

\[
Y_t = \begin{bmatrix} Y_{1,t} \\ Y_{2,t} \end{bmatrix} \sim i.i.d. N \left( \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, I_2 \right)
\]

\( t = 1, \ldots, T, \) and with \( \mu_1 < \mu_2. \)

\( \mu = (\mu_1 \mu_2)' \)

Set identification: \( \mu_1 \leq \theta \leq \mu_2 \)
Frequentist inference:

(a) small $T$

(b) large $T$

Notes: The jagged lines in panel (a) connote uncertainty about the values of $\mu_1$ and $\mu_2$. 
Bayes Inference: \[ f(\theta | Y) = \int f(\theta | \mu, Y) f(\mu | Y) d\mu = \int f(\theta | \mu) f(\mu | Y) d\mu \]

(a) Prior 1

(i) small \( T \)

(ii) large \( T \)

Notes: Jagged lines above \( \hat{\mu}_1 \) and \( \hat{\mu}_2 \) in panels (a.i) and (b.i) connote uncertainty about the values of \( \mu_1 \) and \( \mu_2 \)
(b) Prior 2

\[ f(q | Y) = f(q | \mu) \]

(i) small \( T \)

(ii) large \( T \)
Back to SVAR: (Example: Debortoli, Gali, Gambetti NBER MA (2019))

\[ Y_t = A_0 + A(L)Y_{t-1} + u_t \]

\[ u_t = Q \xi_t \]

\[ Q = \Sigma_u^{1/2} R, \quad \xi_t = (\xi_t^{\text{Technology}}, \xi_t^{\text{Demand}}, \xi_t^{\text{Monetary policy}}, \xi_t^{\text{Supply}}) \]

\[ Y_t = \tilde{A}_0 + (I - A(L))^{-1} Q \xi_t \]

IRFs (\( \partial Y_{t+h} / \partial \xi_{it} \)) are \( (I - A(L))^{-1} \Sigma_u^{1/2} R \)

Sign Restrictions: \( \partial Y_{t+h} / \partial \xi_{it} > 0 \) (or < 0) for various values of \( i \) and \( j \).

Question: Suppose \( A(L) \) and \( \Sigma_{uitt} \) are known. What do the sign restrictions tell us about \( R \) and \( (I - A(L))^{-1} \Sigma_u^{1/2} R \) ?
Figure 3: Identified sets for the 4-variable SVAR
Again, with $A(L)$ and $\Sigma_{uu}$ are known so no sampling uncertainty.

What do we know about $\theta = \partial \text{Output}_{t+4} / \partial \epsilon_t^{\text{Demand}}$?

Frequentist:

(a) Frequentist inference: the identified set
Bayes: \[ f(\theta | Y) = \int f(\theta | \mu) f(\mu | Y) \, d\mu = f(\theta | \mu) \quad (\mu \text{ denotes the parameters in } A(L) \text{ and } \Sigma_{\text{nn}}) \]

What is the prior: 'uniform' prior for $R$ (Harr) .. truncated by the sign-restrictions
Bayes inference: $\theta = \partial \text{Output}_{t+4} / \partial \varepsilon_{i,\text{Demand}}$ .. posterior = prior
Figure 5: Truncated priors (= posteriors) for $\partial Y_{t+4}/\partial \theta$

Notes: The figure shows the prior on the 4-period-ahead impulse responses induced by a flat prior on $R$ and the equality and sign restrictions.
Impulse responses ($\partial Y_{t+h}/\partial \hat{\theta}$)
Identified sets and quantiles of truncated prior (= posterior)

Notes: The solid red lines show the boundaries of the identified sets. The dark (light) 'error bands' shows the equal-tail 68% (95%) posterior/prior credible intervals.
Lessons:

(1) Sign restrictions can be informative

(2) Frequentist inference focus on the identified set. (This is what the data can tell us about).

(3) Bayes inference utilizes a prior.

(a) Effect of prior doesn't vanish in large samples

(b) Inference depends critically on prior
   (i) Good priors are good
   (ii) Bad priors are bad

(c) Evaluation
   (i) Must know what the prior is .. and specifically, what is the prior of the parameters of interest.
Identification of H: (3) External Instruments ('Proxy variables')

Step back for a moment and consider general problem of estimating Dynamic causal effects and IRFs

$$Y_t = D_0 \varepsilon_t + D_1 \varepsilon_{t-1} + … = D(L)\varepsilon_t$$

(Note: $D_0 = H$ in our discussion above.)

**DO NOT ASSUME INVERTIBILITY**
Estimating dynamic causal effects in macroeconomics

Standard approach (that was outlined above):
- Estimate VAR for $Y$
- Assume "invertibility" to relate $\varepsilon$ to VAR forecast errors.
- Impose some restrictions on $H$ for identification

Alternative Approach:
- Find an "external" instrument $Z$ that captures some exogenous variation in one of the structural shocks.
- Use instrument (with or without VAR step) to estimate dynamic causal effects.
Some references on external instruments


- $Y_t = [ R_t, 100 \times \Delta \ln(IP), 100 \times \Delta \ln(CPI), EBP ]$

- Monetary policy shock = $\varepsilon_{1,t}$

- Causal Effects: $E(Y_{i,t+h} \mid \varepsilon_{1,t} = 1) - E(Y_{i,t+h} \mid \varepsilon_{1,t} = 0) = \Theta_{h,i}$

- Kuttner (2001)-like instrument, $Z_t =$ change in Federal Funds rate futures in short window around FOMC announcements.
  - $Z_t$ correlated with $\varepsilon_{1,t}$ but uncorrelated with $\varepsilon_{2n, t} = (\varepsilon_{2, t}, \varepsilon_{3, t}, \ldots, \varepsilon_{n, t})$. 

**Direct estimation of $D_{h,t1}$**

$$Y_t = D_0 \varepsilon_t + D_1 \varepsilon_{t-1} + \ldots = D(L)\varepsilon_t$$

$$Y_{t+h} = D_{h,t1} \varepsilon_{t} + u_t \quad \text{(LP)}$$

$$u_t = \{ \varepsilon_{t+h}, \ldots, \varepsilon_{t+1}, \varepsilon_{2m_j} \varepsilon_{t-1}, \ldots \}$$

$\{x\}$: linear combinations of elements of $x$

$$E(\varepsilon_{t1} u_t) = 0$$

But $\varepsilon_{t1}$ is not observed
IV estimation of $D_{h,l}$

\[ Y_{l,t+h} = D_{h,l} e_{l,t} + \{ e_{t+h}, \ldots, e_{t+1}, e_{2n_{t,j}}, e_{t-1}, \ldots \} \]

\[ Y_{1,t} = D_{0,11} e_{1,t} + \{ e_{2n_{t,j}}, e_{t-1}, \ldots \} = e_{1,t} + \{ e_{2n_{t,j}}, e_{t-1}, \ldots \} \]

(Recall $D_0 = H$, so unit-effect normalization yields $D_{0,11} = 1$)

\[ Y_{l,t+h} = D_{h,l} Y_{1,t} + \{ e_{t+h}, \ldots, e_{t+1}, e_{2n_{t,j}}, e_{t-1}, \ldots \} \]

**Condition LP-IV:**

(i) $E(e_{1,t} Z_t) = \alpha \neq 0$

(ii) $E(\epsilon_{2n_{t,j}} Z_t') = 0$

(iii) $E(\epsilon_{t+j} Z_t') = 0$ for $j \neq 0$
Odds and ends

- HAR SEs
- Dyn. Causal Effects for levels vs. differences
- Weak-instrument robust inference
- "News" Shocks
  - replace $D_{0,11} = 1$ normalization with $D_{k,11} = 1$ normalization
- Smoothness constraints (Barnichon & Brownlees, Plagborg-Møller, …)
- $\varepsilon_{it}$ (or its variance) is not identified. (see Plagborg-Møller-Wolf (2018) for bounds).
Results for \([R \text{ and } 100 \times \ln(IP)]\)
(1990m1 - 2012:m6)

<table>
<thead>
<tr>
<th>lag (h)</th>
<th>(a)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>R</strong></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1.00 (0.00)</td>
</tr>
<tr>
<td>6</td>
<td>-0.07 (1.34)</td>
</tr>
<tr>
<td>12</td>
<td>-1.05 (2.51)</td>
</tr>
<tr>
<td>24</td>
<td>-2.09 (5.66)</td>
</tr>
<tr>
<td><strong>IP</strong></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>-0.59 (0.71)</td>
</tr>
<tr>
<td>6</td>
<td>-2.15 (3.42)</td>
</tr>
<tr>
<td>12</td>
<td>-3.60 (6.23)</td>
</tr>
<tr>
<td>24</td>
<td>-2.99 (10.21)</td>
</tr>
<tr>
<td><strong>Controls</strong></td>
<td>none</td>
</tr>
<tr>
<td><strong>First-stage F</strong></td>
<td>1.7</td>
</tr>
</tbody>
</table>
Results for \([R \text{ and } 100 \times \ln(IP)]\)
(1990m1 -2012:m6)

<table>
<thead>
<tr>
<th>lag ((h))</th>
<th>((a))</th>
</tr>
</thead>
<tbody>
<tr>
<td>(R)</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1.00 (0.00)</td>
</tr>
<tr>
<td>6</td>
<td>-0.07 (1.34)</td>
</tr>
<tr>
<td>12</td>
<td>-1.05 (2.51)</td>
</tr>
<tr>
<td>24</td>
<td>-2.09 (5.66)</td>
</tr>
<tr>
<td>(IP)</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>-0.59 (0.71)</td>
</tr>
<tr>
<td>6</td>
<td>-2.15 (3.42)</td>
</tr>
<tr>
<td>12</td>
<td>-3.60 (6.23)</td>
</tr>
<tr>
<td>24</td>
<td>-2.99 (10.21)</td>
</tr>
<tr>
<td>Controls</td>
<td>none</td>
</tr>
<tr>
<td><strong>First-stage (F)</strong></td>
<td><strong>1.7</strong></td>
</tr>
</tbody>
</table>
Results for \([R \text{ and } 100 \times \ln(IP)](1990m1 - 2012:m6)\)

<table>
<thead>
<tr>
<th>lag ((h))</th>
<th>((a))</th>
</tr>
</thead>
<tbody>
<tr>
<td>(R)</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1.00 (0.00)</td>
</tr>
<tr>
<td>6</td>
<td>-0.07 (1.34)</td>
</tr>
<tr>
<td>12</td>
<td>-1.05 (2.51)</td>
</tr>
<tr>
<td>24</td>
<td>-2.09 (5.66)</td>
</tr>
<tr>
<td>(IP)</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>-0.59 (0.71)</td>
</tr>
<tr>
<td>6</td>
<td>-2.15 (3.42)</td>
</tr>
<tr>
<td>12</td>
<td>-3.60 (6.23)</td>
</tr>
<tr>
<td>24</td>
<td>-2.99 (10.21)</td>
</tr>
</tbody>
</table>

Controls: none
First-stage \(F\): 1.7

**IV Estimation of \(D_{k,t}\) with additional controls -1**

\[ Y_{i,t+h} = D_{k,i} Y_{1,t} + \{ \varepsilon_{t+h}, \ldots, \varepsilon_{t+1}, \varepsilon_{2,t}, \varepsilon_{t-1}, \ldots \} \]

2 Motivations for adding controls:

1) eliminate part of error term
   - controls should be uncorrelated with \(\varepsilon_{t,i}\).
     - Examples: lags of \(Z\), \(Y\), other macro variables, 'factors,' etc., leads of \(Z\).

2) \(Z\) may be correlated with error, but uncorrelated after adding controls
   - Example: \(GK-Z = \{ \Delta FFF, \Delta FFF_{t-1}\}. Add\ lags\ of\ FFF_t.\)
IV Estimation of $D_{h,i}$ with additional controls - 2

$$Y_{t+h} = D_{h,i} Y_{1,t} + \gamma' W_t + u_t$$

$$\chi_t^\perp = x_t - \text{Proj}(x_t | W_t)$$

Condition LP-$\text{IV}^\perp$

(i) $$E\left(\varepsilon_{1,t}^\perp Z_t^\perp'\right) = \alpha' \neq 0$$

(ii) $$E\left(\varepsilon_{2:n_{\varepsilon,t}}^\perp Z_t^\perp'\right) = 0$$

(iii) $$E\left(\varepsilon_{t+j}^\perp Z_t^\perp'\right) = 0 \text{ for } j \neq 0.$$
Results for \([R \text{ and } 100 \times \ln(IP)]\)

\[ Y_{i,t+h} = D_{h,i} Y_{1,t} + \gamma W_t + \{ \varepsilon_{t+h}, \ldots, \varepsilon_{t+1}, \varepsilon_{2h,i,t}, \varepsilon_{t-1}, \ldots \} \]

<table>
<thead>
<tr>
<th>lag ((h))</th>
<th>((a))</th>
<th>((b))</th>
<th>((c))</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>R</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1.00 (0.00)</td>
<td>1.00 (0.00)</td>
<td>1.00 (0.00)</td>
</tr>
<tr>
<td>6</td>
<td>-0.07 (1.34)</td>
<td>1.12 (0.52)</td>
<td>0.67 (0.57)</td>
</tr>
<tr>
<td>12</td>
<td>-1.05 (2.51)</td>
<td>0.78 (1.02)</td>
<td>-0.12 (1.07)</td>
</tr>
<tr>
<td>24</td>
<td>-2.09 (5.66)</td>
<td>-0.80 (1.53)</td>
<td>-1.57 (1.48)</td>
</tr>
<tr>
<td><strong>IP</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>-0.59 (0.71)</td>
<td>0.21 (0.40)</td>
<td>0.03 (0.55)</td>
</tr>
<tr>
<td>6</td>
<td>-2.15 (3.42)</td>
<td>-3.80 (3.14)</td>
<td>-4.05 (3.65)</td>
</tr>
<tr>
<td>12</td>
<td>-3.60 (6.23)</td>
<td>-6.70 (4.70)</td>
<td>-6.86 (5.49)</td>
</tr>
<tr>
<td>24</td>
<td>-2.99 (10.21)</td>
<td>-9.51 (7.70)</td>
<td>-8.13 (7.62)</td>
</tr>
<tr>
<td><strong>Controls</strong></td>
<td>none</td>
<td>4 lags of ((z,y))</td>
<td>4 lags of ((z,y,factors))</td>
</tr>
<tr>
<td><strong>First-stage F</strong></td>
<td>1.7</td>
<td>23.7</td>
<td>18.6</td>
</tr>
</tbody>
</table>
Results for \([R \text{ and } 100 \times \ln(IP)]\)

\[
Y_{i,t+h} = D_{h,i} Y_{1,t} + \gamma W_t + \{ \varepsilon_{t+h}, \ldots, \varepsilon_{t+1}, \varepsilon_t, \varepsilon_{t-1}, \ldots \}
\]

<table>
<thead>
<tr>
<th>lag ((h))</th>
<th>(a) (R)</th>
<th>(b) (IP)</th>
<th>(c) Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1.00 (0.00)</td>
<td>1.00 (0.00)</td>
<td>1.00 (0.00)</td>
</tr>
<tr>
<td>6</td>
<td>-0.07 (1.34)</td>
<td>1.12 (0.52)</td>
<td>0.67 (0.57)</td>
</tr>
<tr>
<td>12</td>
<td>-1.05 (2.51)</td>
<td>0.78 (1.02)</td>
<td>-0.12 (1.07)</td>
</tr>
<tr>
<td>24</td>
<td>-2.09 (5.66)</td>
<td>-0.80 (1.53)</td>
<td>-1.57 (1.48)</td>
</tr>
</tbody>
</table>

| 0          | -0.59 (0.71) | 0.21 (0.40) | 0.03 (0.55) |
| 6          | -2.15 (3.42) | -3.80 (3.14) | -4.05 (3.65) |
| 12         | -3.60 (6.23) | -6.70 (4.70) | -6.86 (5.49) |
| 24         | -2.99 (10.21) | -9.51 (7.70) | -8.13 (7.62) |

Controls: none \(4\) lags of \((z,y)\) \(4\) lags of \((z,y,factors)\)

First-stage \(F\): 1.7 \(23.7\) 18.6
Results for $[R \text{ and } 100 \times \ln(IP)]$

\[ Y_{i,t+h} = D_{h,i} Y_{1,t} + \gamma W_t + \{ \varepsilon_{t+h}, \ldots, \varepsilon_{t+1}, \varepsilon_{t-1}, \ldots \} \]

<table>
<thead>
<tr>
<th>lag ($h$)</th>
<th>(a)</th>
<th>(b)</th>
<th>(c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1.00 (0.00)</td>
<td>1.00 (0.00)</td>
<td>1.00 (0.00)</td>
</tr>
<tr>
<td>6</td>
<td>-0.07 (1.34)</td>
<td>1.12 (0.52)</td>
<td>0.67 (0.57)</td>
</tr>
<tr>
<td>12</td>
<td>-1.05 (2.51)</td>
<td>0.78 (1.02)</td>
<td>-0.12 (1.07)</td>
</tr>
<tr>
<td>24</td>
<td>-2.09 (5.66)</td>
<td>-0.80 (1.53)</td>
<td>-1.57 (1.48)</td>
</tr>
<tr>
<td>$IP$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>-0.59 (0.71)</td>
<td>0.21 (0.40)</td>
<td>0.03 (0.55)</td>
</tr>
<tr>
<td>6</td>
<td>-2.15 (3.42)</td>
<td>-3.80 (3.14)</td>
<td>-4.05 (3.65)</td>
</tr>
<tr>
<td>12</td>
<td>-3.60 (6.23)</td>
<td>-6.70 (4.70)</td>
<td>-6.86 (5.49)</td>
</tr>
<tr>
<td>24</td>
<td>-2.99 (10.21)</td>
<td>-9.51 (7.70)</td>
<td>-8.13 (7.62)</td>
</tr>
</tbody>
</table>

Controls

- none
- 4 lags of $(z,y)$
- 4 lags of $(z,y,factors)$

First-stage $F$

- 1.7
- 23.7
- 18.6
SVARs with External Instruments - 1

VAR: \( Y_t = A_1 Y_{t-1} + A_2 Y_{t-2} + \ldots + \eta_t \)

Structural MA: \( Y_t = H \varepsilon_t + D_1 \varepsilon_{t-1} + \ldots = D(L) \varepsilon_t \)

\((D_0 = H \text{ in notation above})\)

Invertibility: \( \varepsilon_t = \text{Proj}(\varepsilon_t|Y_t, Y_{t-1}, \ldots) \)

\(\Rightarrow\)

\( \eta_t = H \varepsilon_t \) with \( H \) nonsingular (so \( n_y = n_{\varepsilon} \))

(Note: Plagborg-Møller and Wolf (2019) .. LP and SVARs estimate same IRFs in population .. also invertibility in SVAR can be weakened)
SVARs with External Instruments - 2

$$A(L)Y_t = \eta_t = H \xi_t$$

$$\Rightarrow Y_t = C(L)H \xi_t \quad \text{with} \quad C(L) = A(L)^{-1}$$

thus $$D_{h,t1} = C_h H_{i1}$$

Unit-effect normalization yields: $$\eta_{i,t} = H_{i1} \eta_{i,t} + \{ \epsilon_{2n_{it},t} \}$$

**Condition SVAR-IV**

(i) $$E(\epsilon_{1i,t} Z_t) = \alpha \neq 0$$

(ii) $$E(\epsilon_{2n_{it},t} Z_t') = 0$$
SVAR with external instruments – estimation

1. Regress $Y_{i,t}$ onto $Y_{1,t}$ using instruments $Z_t$ and $p$ lags of $Y_t$ as controls. This yields $\hat{H}_{i1}$.

2. Estimate a VAR($p$) and invert the VAR to obtain $\hat{C}(L) = \hat{A}(L)^{-1}$.

3. Estimate the dynamic causal effects of shock 1 on the vector of variables as

$$\hat{D}_{h1} = \hat{C}_h \hat{H}_1$$

(odds and ends: (1) News shocks; (2) Dif. sample periods in (1) and (2))
SVAR with external instruments – inference

- Strong instruments:
  \[
  \sqrt{T} \left( \begin{pmatrix} \hat{A} - A \\ \hat{H}_1 - H_1 \end{pmatrix} \right) \xrightarrow{d} \text{Normal + } \delta\text{-method}
  \]

- Weak instruments:
  - \( \sqrt{T} (\hat{A} - A) \xrightarrow{d} \text{Normal.} \)
  - \( \hat{H}_1 - H_1 \xrightarrow{d} \text{NonNormal.} \)
  - Use weak-instrument robust methods. (Montiel Olea, Stock and Watson (2018)).
Results for \([R \text{ and } 100 \times \ln(IP)]\)

<table>
<thead>
<tr>
<th>lag (h)</th>
<th>LP-IV (1990m1)-(2012m6)</th>
<th>SVAR-IV (IV: 1990m1)-(2012m6) (VAR: 1980m7)-(2012m6)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(R)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1.00 (0.00)</td>
<td>1.00 (0.00)</td>
</tr>
<tr>
<td>6</td>
<td>1.12 (0.52)</td>
<td>0.89 (0.31)</td>
</tr>
<tr>
<td>12</td>
<td>0.78 (1.02)</td>
<td>0.78 (0.46)</td>
</tr>
<tr>
<td>24</td>
<td>-0.80 (1.53)</td>
<td>0.40 (0.49)</td>
</tr>
<tr>
<td>(IP)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0.21 (0.40)</td>
<td>0.16 (0.59)</td>
</tr>
<tr>
<td>6</td>
<td>-3.80 (3.14)</td>
<td>-0.81 (1.19)</td>
</tr>
<tr>
<td>12</td>
<td>-6.70 (4.70)</td>
<td>-1.87 (1.54)</td>
</tr>
<tr>
<td>24</td>
<td>-9.51 (7.70)</td>
<td>-2.16 (1.65)</td>
</tr>
<tr>
<td>Controls</td>
<td>4 lags of ((Z, Y))</td>
<td>12 lags of (Y)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4 lags of (Z)</td>
</tr>
<tr>
<td>First-stage (F)</td>
<td>23.7</td>
<td>20.5</td>
</tr>
</tbody>
</table>

Results for \([R \text{ and } 100 \times \ln(IP)]\)

<table>
<thead>
<tr>
<th>lag (h)</th>
<th>LP-IV (1990m1)-(2012m6)</th>
<th>SVAR-IV (IV: 1990m1)-(2012m6) (VAR: 1980m7)-(2012m6)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(R)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1.00 (0.00)</td>
<td>1.00 (0.00)</td>
</tr>
<tr>
<td>6</td>
<td>1.12 (0.52)</td>
<td>0.89 (0.31)</td>
</tr>
<tr>
<td>12</td>
<td>0.78 (1.02)</td>
<td>0.78 (0.46)</td>
</tr>
<tr>
<td>24</td>
<td>-0.80 (1.53)</td>
<td>0.40 (0.49)</td>
</tr>
<tr>
<td>(IP)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0.21 (0.40)</td>
<td>0.16 (0.59)</td>
</tr>
<tr>
<td>6</td>
<td>-3.80 (3.14)</td>
<td>-0.81 (1.19)</td>
</tr>
<tr>
<td>12</td>
<td>-6.70 (4.70)</td>
<td>-1.87 (1.54)</td>
</tr>
<tr>
<td>24</td>
<td>-9.51 (7.70)</td>
<td>-2.16 (1.65)</td>
</tr>
<tr>
<td>Controls</td>
<td>4 lags of ((Z, Y))</td>
<td>12 lags of (Y)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4 lags of (Z)</td>
</tr>
<tr>
<td>First-stage (F)</td>
<td>23.7</td>
<td>20.5</td>
</tr>
</tbody>
</table>