Dynamic Factor Models, Factor Augmented VARs, and SVARs in Macroeconomics

-- Part 2: SVARs and SDFMs --

Mark Watson
Princeton University

Central Bank of Chile
October 22-24, 2018

Reference: Stock, James H. and Mark W. Watson (2016)
Handbook of Macroeconomics, Vol 2. chapter

DFM:

$$X_t = \Lambda F_t + u_t$$

$$\Phi(\mathrm{L})F_t = \mathrm{G}\eta_t$$

Question: Identify "structural" shocks in η_t and their effects on $\{X_t\}$

And how is this related to the analogous question in VARs

Start with discussion of VAR and then return to DFM

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 $A(\mathrm{L})Y_t = \eta_t$.

η_t = 1-period ahead forecast error. (Note change of notation from DFM.)

No constant term for notational convenience.

VMA representation:

$$Y_t = \mathrm{C}(\mathrm{L})^{-1}\eta_t \text{ where } \mathrm{C}(\mathrm{L}) = \mathrm{A}(\mathrm{L})^{-1}$$

Note: $\mathrm{C}(\mathrm{L}) = \mathrm{C}_0 + \mathrm{C}_1\mathrm{L} + \mathrm{C}_2\mathrm{L}^2 + \ldots$ and $\mathrm{C}_0 = \mathrm{I}$

SVAR (Sims (1980)): Why do we make forecast errors?

$\eta_t = \mathrm{H}\varepsilon_t$ where ε_t are 'structural' shocks. (Shocks interpretable in the context of particular theoretical economic models).

$Y_t = \mathrm{C(L)}\eta_t = \mathrm{C(L)H}\varepsilon_t = \mathrm{D(L)}\varepsilon_t$ is structural MA

and with $\mathrm{B(L)} = \mathrm{H}^{-1}\mathrm{A(L)}$

$\mathrm{B(L)}Y_t = \varepsilon_t$ is SVAR

From SMA: $Y_t = \mathrm{D}_0\varepsilon_t + \mathrm{D}_1\varepsilon_{t-1} + \ldots$ with $\mathrm{D}_k = \mathrm{C}_k\mathrm{H}$

Note: $\dfrac{\partial Y_{i,t+k}}{\partial \varepsilon_{jt}} = \mathrm{D}_{k,ij}$. (These are "impulse responses" or "dynamic causal effects" or 'dynamic multipliers' …)

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?

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

5. Identification of H.

6. Properties of $\hat{C}_k \hat{H}$.

Issues:

1. E(Y_t | 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.

"Hayashi": Roots of A(L) outside unit circle (difference equation is stable). η_t are MDS with appropriate moments.

Issue: $\eta_t = \mathrm{H}\varepsilon_t$ 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 ε_t from current and lagged Y.

'Recoverability' (Chahrour and Jurado (2017), Plagbor-Moller and Wolf (2018)): Can I determine ε_t from current, lagged and *future* Y.

Simplist example:

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

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

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))

$$y_{t+1} = \mathrm{C}x_t + \mathrm{D}w_{t+1}$$

$$x_{t+1} = \mathrm{A}x_t + \mathrm{B}w_{t+1}$$

Invertibilty: eigenvalues of $(\mathrm{A} - \mathrm{BD}^{-1}\mathrm{C})$ are less than 1 in modulus.

(Recoverability): When is $\operatorname{var}\left(w_t \mid \left\{ y_{t+j} \right\}_{j=-\infty}^{\infty} \right) = 0$? (Exercise)

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 their 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
  - Uncorrelated Structural Shocks: Restrict $\Sigma_{\varepsilon\varepsilon}$ to be diagonal: $n^2 + n$ unknowns .. $n(n+1)/2$ too many parameters.

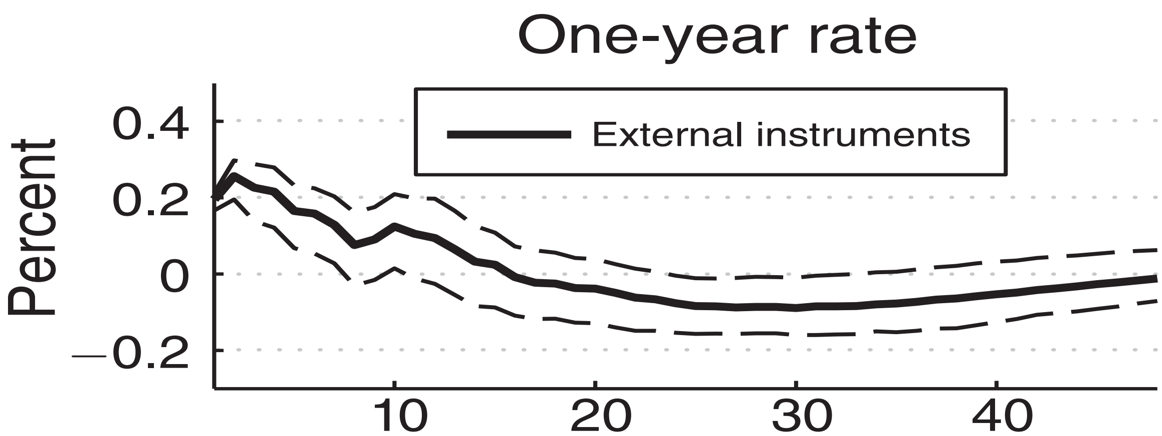
o Scale normalization

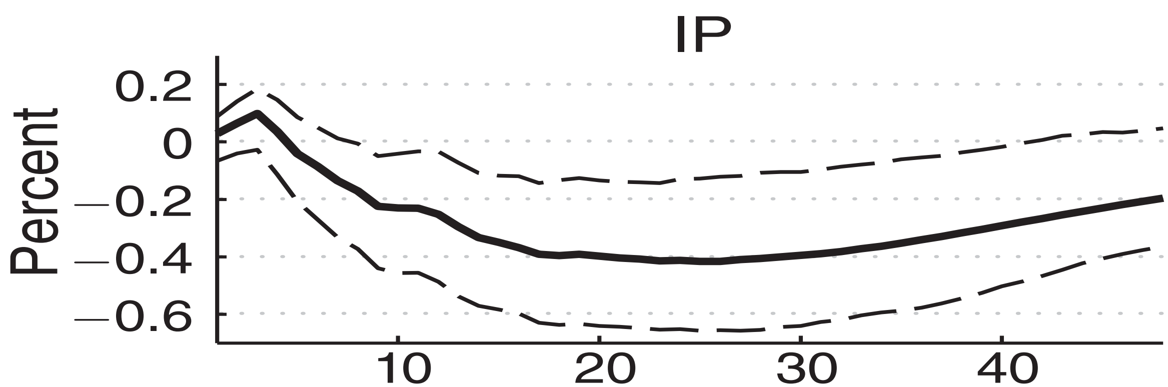
scalar model: $\eta_t = \text{H}\varepsilon_t$ ('units' of ε_t are not identified)

2 normalizations: (1) $\sigma_\varepsilon = 1$

(2) $\text{H} = 1$ (or $\text{H}^{-1} = 1$)

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





Scale normalization does not matter in population.

It will matter for inference.

Moving from one normalization to another involves dividing by $\hat{H}$ or $\hat{\sigma}_{\varepsilon}$.

We will use normalization on elements of H.

- e.g., Diagonal elements of H are unity

- Alternatives:
  - $\Sigma_{\varepsilon\varepsilon} = I$
  - 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 $n = 3$

$$Y_t = \mathrm{A}Y_{t-1} + \begin{bmatrix} 1 & H_{12} & H_{13} \\ H_{21} & 1 & H_{23} \\ H_{31} & H_{32} & 1 \end{bmatrix} \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{bmatrix}$$

$$Y_t = \mathrm{A}Y_{t-1} + \begin{bmatrix} 1 & H_{12} & H_{13} \\ H_{21} & 1 & H_{23} \\ H_{31} & H_{32} & 1 \end{bmatrix} \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{bmatrix}$$

Timing restriction example: $Y_t = \mathrm{A}Y_{t-1} + \begin{bmatrix} 1 & 0 & 0 \\ H_{21} & 1 & 0 \\ H_{31} & H_{32} & 1 \end{bmatrix} \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{bmatrix}$

Long-run restriction example:

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

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

Restrict H so that D_{ij} has $n(n\text{-}1)/2$ zeros.

And so forth.

Identification of one shock, say ε_{1t} and its effect on Y_{t+k}

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

Thus

$$Y_t = \mathrm{C(L)}\begin{bmatrix} H_1 & H_\bullet \end{bmatrix}\begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{\bullet,t} \end{bmatrix} = \mathrm{C(L)H}_1\varepsilon_{1t} + \text{distributed lag of } \varepsilon_{\bullet t}$$

where $\bullet$ denotes elements 2 through n

To identify the effect of ε_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 ε_{1t} can be 're-constructed' from η_t (up to scale) – Algebra in paper.

Identification of H_1

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

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

$\varepsilon_1 = \eta_1$, and H_1 is identified by regressing η_t onto $\eta_{1,t}$.

Similar for other timing restrictions, long-run restrictions, etc.

Other populator identification schemes

(1) Heteroskedasticity

(2) Sign Restrictions

(3) External Instruments ('Proxy variables')

Identification by Heteroskedasticity (Rigobon (2003), Rigobon and Sack (2003,2004))

Idea: $\Sigma_{\varepsilon\varepsilon}^{1}$ and $\Sigma_{\varepsilon\varepsilon}^{2} \Rightarrow \Sigma_{\eta\eta}^{1} = H\Sigma_{\varepsilon\varepsilon}^{1}H'$ and $\Sigma_{\eta\eta}^{2} = H\Sigma_{\varepsilon\varepsilon}^{2}H'$

Order condition (counting):

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

Number of unknowns: (H, $\Sigma_{\varepsilon\varepsilon}^{1}$ and $\Sigma_{\varepsilon\varepsilon}^{2}$): $(n^2-n) + 2n = n^2 + n$.

Note: 'rank condition' .. relative variances of ε_t must change to get independent information on elements of H.

Potentially powerful tool.

Generalizes to time-varying conditional heteroskedasticity.

Example:

$$\begin{pmatrix} \Sigma^{j}_{\eta_1\eta_1} & \Sigma^{j}_{\eta_1\eta_2} \\ \Sigma^{j}_{\eta_2\eta_1} & \Sigma^{j}_{\eta_2\eta_2} \end{pmatrix} = \begin{pmatrix} 1 & H_{12} \\ H_{21} & 1 \end{pmatrix} \begin{pmatrix} \sigma^2_{\varepsilon_1,j} & 0 \\ 0 & \sigma^2_{\varepsilon_2} \end{pmatrix} \begin{pmatrix} 1 & H_{21} \\ H_{12} & 1 \end{pmatrix}, \quad j = 1, 2.$$

Algebra $\Rightarrow$

$$H_{21} = \frac{\Sigma^{2}_{\eta_1\eta_2} - \Sigma^{1}_{\eta_1\eta_2}}{\Sigma^{2}_{\eta_1\eta_1} - \Sigma^{1}_{\eta_1\eta_1}}$$

Estimator:

$$\hat{H}_{21} = \frac{\hat{\Sigma}^{2}_{\eta_1\eta_2} - \hat{\Sigma}^{1}_{\eta_1\eta_2}}{\hat{\Sigma}^{2}_{\eta_1\eta_1} - \hat{\Sigma}^{1}_{\eta_1\eta_1}}$$

$$\hat{H}_{21} = \frac{\hat{\Sigma}^2_{\eta_1\eta_2} - \hat{\Sigma}^1_{\eta_1\eta_2}}{\hat{\Sigma}^2_{\eta_1\eta_1} - \hat{\Sigma}^1_{\eta_1\eta_1}}$$

Denominator: $\hat{\Sigma}^2_{\eta_1\eta_1} - \hat{\Sigma}^1_{\eta_1\eta_1} = \left(\Sigma^2_{\eta_1\eta_1} - \Sigma^1_{\eta_1\eta_1}\right) + \text{Sampling Error}\left(\hat{\Sigma}^2_{\eta_1\eta_1} - \hat{\Sigma}^1_{\eta_1\eta_1}\right)$

Estimator will have poor sampling properties when denominator is noisy:

$\text{Sampling Error}\left(\hat{\Sigma}^2_{\eta_1\eta_1} - \hat{\Sigma}^1_{\eta_1\eta_1}\right)$ is big relative to $\left(\Sigma^2_{\eta_1\eta_1} - \Sigma^1_{\eta_1\eta_1}\right)$.

Or, (1) when change in variance is small or one or both of the samples is small.

Inequality (Sign) Restrictions (Faust (1998), Uhlig (2005))

Typical identifying restrictions: $R\mathrm{H} = r$ where R and r are pre-specified are can be computed from the data. (Or $R\mathrm{H}_1 = r$, when focused on a single shock.)

Inequality Restrictions: $R\mathrm{H} \geq r$.

This 'set identified' the impulse responses.

Determining the identified set. A computational method using $\Sigma_{\varepsilon\varepsilon} = \mathrm{I}$ normalization.

$\Sigma_{\eta\eta} = \mathrm{H}\Sigma_{\varepsilon\varepsilon}\mathrm{H}' = \mathrm{HH}'$, so H is a matrix square root of $\Sigma_{\eta\eta}$ $\Rightarrow$

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

(1) Compute $\Sigma_{\eta\eta}^{1/2}$

(2) For a particular value of C, compute H = $\Sigma_{\eta\eta}^{1/2}C$.

(3) Check to see if RH $\geq r$. 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.

Inference in a "set identified" model

Easy example: Suppose θ is a parameter of interest. You know that θ is restricted to lie between μ_L and μ_U. That is $\mu_L \le \theta \le \mu_U$.

You have an i.i.d. sample of data on (X_i, Y_i) where:

$$\begin{pmatrix} X_i \\ Y_i \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_L \\ \mu_U \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right)$$

and you want to conduct inference about θ. What should you do?

Frequentist: Data give you information about μ_L and μ_U. Estimate these bounds. That's it.

Bayes: Priors on μ_L, μ_U and θ. Form posterior. Data tells you about μ_L, μ_U, but nothing more about θ. Likelihood is flat for all values of θ between μ_L and μ_U. In large samples posterior for θ is the prior, but truncated at μ_L and μ_U.

Bayes and frequentist inference couldn't be more different here. For example, a 95% Bayes posterior credible set for θ has a frequentist coverage of 0% or 100%. (The Bayes 95% set is a 0% or 100% confidence set.)

What should you do:

(1) Estimate the identified set. (Estimate μ_L and μ_U in the example. Sampling uncertainty is over the boundary of this set.)

(2) Do Bayes analysis. Prior is critical. In large samples the prior *is* the posterior. Think carefully about prior.

What you shouldn't do.

(3) Do Bayes analysis without careful thought about prior.

Back to Sign-restricted VARs: Baumeister and Hamilton (2015, 2017).

SVAR (one lag for notationaly convenience):

$Y_t = \mathrm{A}Y_{t-1} + \eta_t = AY_{t\text{-}1} + \mathrm{H}\varepsilon_t$ or

$$\mathrm{B}_0 Y_t = \mathrm{B}_1 Y_{t-1} + \varepsilon_t$$

with $\mathrm{B}_0 = \mathrm{H}^{-1}$ and $\mathrm{B}_1 = \mathrm{H}^{-1}\mathrm{A}$.

Baumeister-Hamilton, use normalization with 1's on diagonal of B_0 ($= \mathrm{H}^{-1}$). They advocate using informative priors about off-diagonal elements of B_0, loose priors on B_1 and variances of ε_t + sign restrictions.

Alternative (originally used on Uhlig(2005) and many others)

(1) Compute $\Sigma_{\eta\eta}^{1/2}$

(2) For a particular value of C, compute H = $\Sigma_{\eta\eta}^{1/2}C$.

(3) Check to see if RH $\geq r$. 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.~~ Use the values from (3) as the posterior.

This amounts to having a flat prior on C ('Harr' prior on columns of orthonormal matrix).

What is a flat prior on C?

2-dimensional problem: $C = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$ with $\theta \sim U(0,2\pi)$

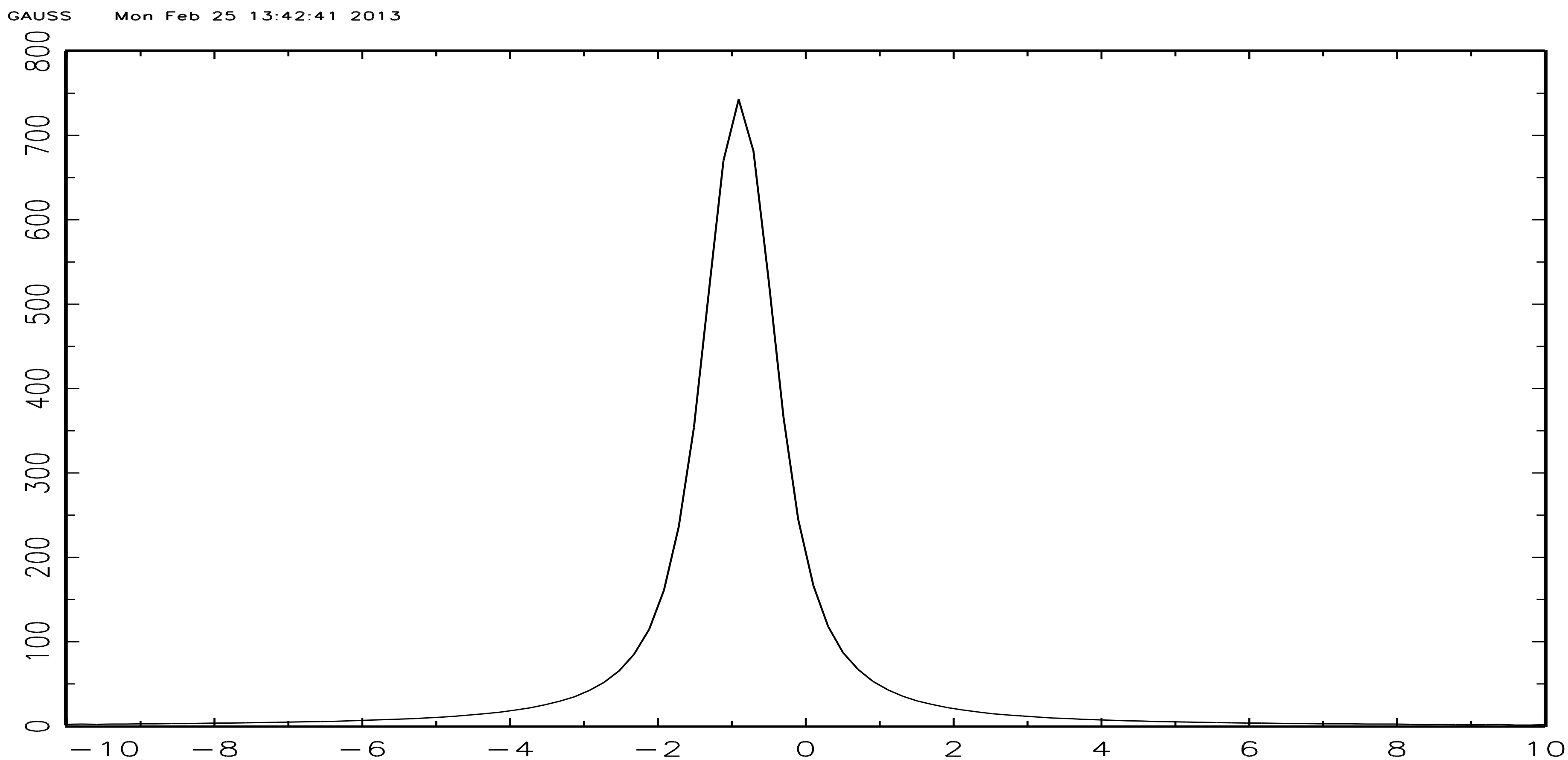
H = $\Sigma_{\eta\eta}^{1/2} C$, so $B_0 = H^{-1} = C^{-1}\Sigma_{\eta\eta}^{-1/2}$. Write $B_0 = \begin{pmatrix} 1 & b_{12} \\ b_{21} & 1 \end{pmatrix}$, so that

$$Y_{1t} = -b_{12}Y_{2t} + \text{lags} + \varepsilon_{1t}$$

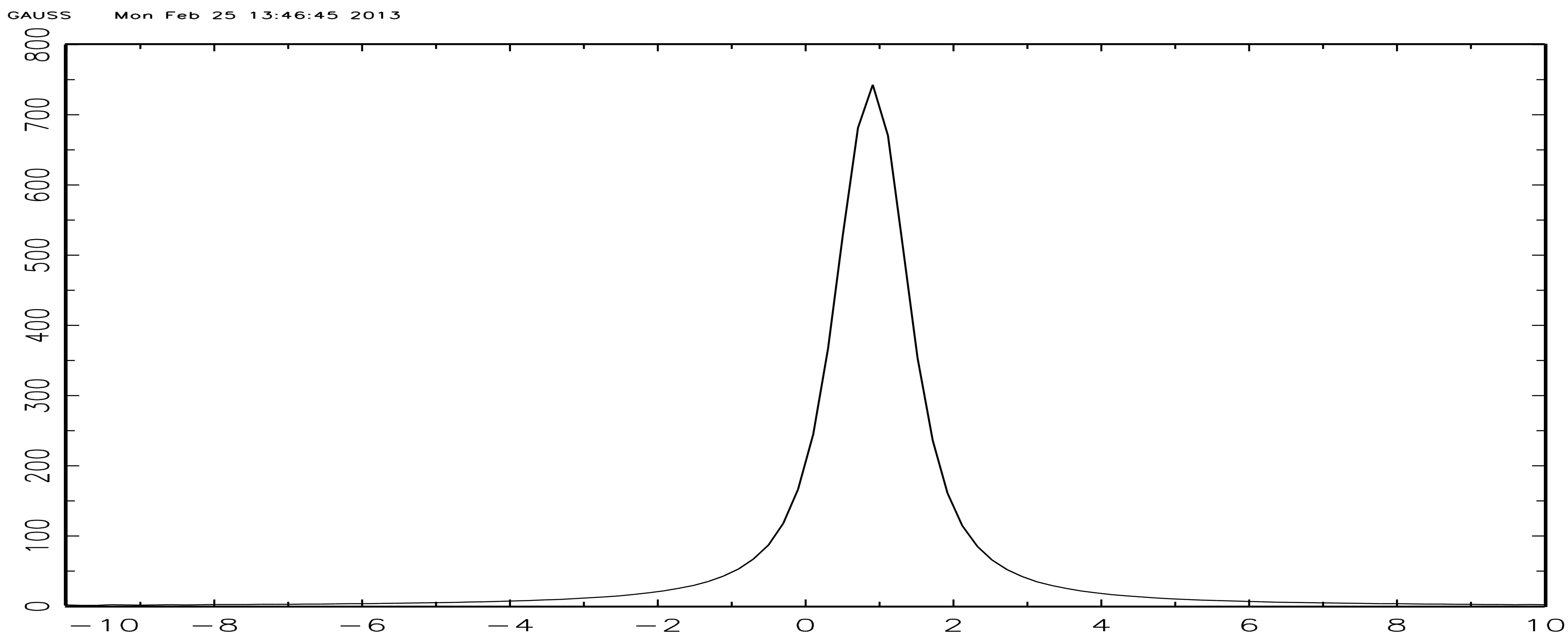
$$Y_{2t} = -b_{21}Y_{1t} + \text{lags} + \varepsilon_{2t}$$

Prior on C is 'flat'. What is implied prior on b_{12} ?

Implied prior for b_{12} … $\Sigma_{\eta\eta} = \begin{bmatrix} 1 & 0.9 \\ 0.9 & 1 \end{bmatrix}$



Implied prior for b_{12} … $\Sigma_{\eta\eta} = \begin{bmatrix} 1 & -0.9 \\ -0.9 & 1 \end{bmatrix}$



Prior on C is flat and does not depend on $\Sigma_{\eta\eta}$.

Implied Prior on b_{12} is not flat, not symmetric, and depends on $\Sigma_{\eta\eta}$.

Bottom line: With sign-restricted SVARs, data cannot completely pin down the effects of ε_t on Y_{t+k}.

Frequentist: Determine what the data can say about this.

Bayesian: Add judgement (prior) + data to make probabilistic statements about the effects. Prior matters.

Identification of H: (3) External Instruments ('Proxy variables')

(Discussion follows Stock-Watson (2018) *Economic Journal* paper)

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

$$Y_t = \underset{\overset{\uparrow}{n_Y}}{D_0}\, \varepsilon_t + D_1\, \underset{\overset{\nwarrow}{n_\varepsilon}}{\varepsilon_{t-1}} + \ldots = D(L)\varepsilon_t$$

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

<u>DO NOT ASSUME INVERTIBILITY (yet)</u>

Estimating dynamic causal effects in macroeconomics

Standard Approach:

- Estimate VAR for Y
- Assume "invertibility" to relate ε_t 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

VARs: Stock (2008), Stock and Watson (2012), Mertens and Ravn (2013, 2014), Gertler and Karadi (2015), Caldera and Kamps (2017), Montiel Olea, Stock and Watson (2012), Lumsford (2015), Jentsch and Lunsford (2016), Drautzburg(2017), Carriero, Momtaz, Theodoridis and Theophilopoulou (2015), …

Local-projections: Jordà, Schularick, and Taylor (2015), Ramey and Zubairy (2017), Ramey (2016), Mertens (2015), Fieldhouse, Mertens, Ravn (2017) …

A Running Empirical Example: Gertler-Karadi (2015)

- Y_t = [R_t, 100×Δln(IP), 100×Δln(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 $\varepsilon_{2:n_\varepsilon,t} = (\varepsilon_{2,t}, \varepsilon_{3,t}, \ldots, \varepsilon_{n_\varepsilon,t})$.

Direct estimation of $\mathrm{D}_{h,i1}$

$$Y_t = \mathrm{D}_0\, \varepsilon_t + \mathrm{D}_1\, \varepsilon_{t-1} + \ldots = \mathrm{D}(\mathrm{L})\varepsilon_t$$

$$Y_{i,t+h} = \mathrm{D}_{h,i1}\, \varepsilon_{1,t} + u_t \qquad\qquad \text{(LP)}$$

$$u_t = \{\varepsilon_{t+h}, \ldots, \varepsilon_{t+1}, \varepsilon_{2:n_\varepsilon,t}, \varepsilon_{t-1}, \ldots\}$$

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

$$\mathrm{E}(\varepsilon_{1,t}\, u_t) = 0$$

But $\varepsilon_{1,t}$ is not observed

IV estimation of $\mathbf{D}_{h,i1}$

$$Y_{i,t+h} = \mathrm{D}_{h,i1}\varepsilon_{1,t} + \{\varepsilon_{t+h}, \ldots, \varepsilon_{t+1}, \varepsilon_{2:n_\varepsilon,t}, \varepsilon_{t-1}, \ldots\}$$

$$Y_{1,t} = \mathbf{D}_{0,11}\varepsilon_{1,t} + \{\varepsilon_{2:n_\varepsilon,t}, \varepsilon_{t-1}, \ldots\} = \varepsilon_{1,t} + \{\varepsilon_{2:n_\varepsilon,t}, \varepsilon_{t-1}, \ldots\}$$

(unit-effect normalization $\mathbf{D}_{0,11} = 1$)

$$Y_{i,t+h} = \mathbf{D}_{h,i1}\boldsymbol{Y}_{1,t} + \{\varepsilon_{t+h}, \ldots, \varepsilon_{t+1}, \varepsilon_{2:n_\varepsilon,t}, \varepsilon_{t-1}, \ldots\}$$

Condition LP-IV:

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

(ii) $\mathrm{E}(\varepsilon_{2:n_\varepsilon,t}Z_t') = 0$

(iii) $\mathrm{E}(\varepsilon_{t+j}Z_t') = 0$ 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, …)
- ε_{1t} (or its variance) is not identified. (see Plagborg-Møller-Wolf for bounds).

Results for [R and $100\times\ln(IP)$] (1990m1 -2012:m6)

	lag (*h*)	**(a)**
R	0	**1.00 (0.00)**
	6	-0.07 (1.34)
	12	-1.05 (2.51)
	24	-2.09 (5.66)
IP	0	-0.59 (0.71)
	6	-2.15 (3.42)
	12	-3.60 (6.23)
	24	-2.99 (10.21)
Controls		none
First-stage *F*		1.7

Results for [R and $100 \times \ln(IP)$]
(1990m1 -2012:m6)

	lag (*h*)	**(a)**
R	0	1.00 (0.00)
	6	-0.07 (1.34)
	12	-1.05 (2.51)
	24	-2.09 (5.66)
IP	0	-0.59 (0.71)
	6	-2.15 (3.42)
	12	-3.60 (6.23)
	24	-2.99 (10.21)
Controls		none
First-stage *F*		**1.7**

Results for [R and $100 \times \ln(IP)$]
(1990m1 -2012:m6)

	lag (*h*)	**(a)**
R	0	1.00 (0.00)
	6	-0.07 (1.34)
	12	-1.05 (2.51)
	24	-2.09 (5.66)
IP	0	-0.59 (0.71)
	6	-2.15 (3.42)
	12	-3.60 (6.23)
	24	-2.99 (10.21)
Controls		none
First-stage *F*		1.7

IV Estimation of $\mathrm{D}_{h,i2}$ with additional controls -1

$$Y_{i,t+h} = \mathrm{D}_{h,i1} Y_{1,t} + \{\varepsilon_{t+h}, \ldots, \varepsilon_{t+1}, \varepsilon_{2:n_\varepsilon,t}, \varepsilon_{t-1}, \ldots\}$$

2 Motivations for adding controls:

(1) eliminate part of error term

- controls should be uncorrelated with $\varepsilon_{1,t}$.
  - Examples: lags of Z, Y, other macro variables, 'factors,' etc., leads of Z.

(2) Z_t may be correlated with error, but uncorrelated after adding controls

(a) *Example:* GK-$Z = \{\Delta FFF_t, \Delta FFF_{t-1}\}$. *Add lags of* FFF_t.

IV Estimation of $D_{h,i1}$ with additional controls - 2

$$Y_{i,t+h} = D_{h,i1} Y_{1,t} + \gamma' W_t + u_t$$

$$x_t^{\perp} = x_t - \mathrm{Proj}(x_t \mid W_t)$$

Condition LP-IV$^{\perp}$

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

(ii) $E\left(\varepsilon_{2:n_\varepsilon,t}^{\perp} Z_t^{\perp\prime}\right) = 0$

(iii) $E\left(\varepsilon_{t+j}^{\perp} Z_t^{\perp\prime}\right) = 0$ for $j \neq 0$.

Results for [R and 100 ×ln(IP)]

$$Y_{i,t+h} = \mathrm{D}_{h,i1}Y_{1,t} + \gamma' W_t + \{\varepsilon_{t+h}, \ldots, \varepsilon_{t+1}, \varepsilon_{2:n_\varepsilon,t}, \varepsilon_{t-1}, \ldots\}$$

	lag (h)	(a)	(b)	(c)
R	0	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
	6	-0.07 (1.34)	1.12 (0.52)	0.67 (0.57)
	12	-1.05 (2.51)	0.78 (1.02)	-0.12 (1.07)
	24	-2.09 (5.66)	-0.80 (1.53)	-1.57 (1.48)
IP	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 and 100 ×ln(IP)]

$$Y_{i,t+h} = \mathrm{D}_{h,i1} Y_{1,t} + \gamma' W_t + \{\varepsilon_{t+h}, \ldots, \varepsilon_{t+1}, \varepsilon_{2:n_\varepsilon,t}, \varepsilon_{t-1}, \ldots\}$$

	lag (h)	(a)	(b)	(c)
R	0	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
	6	-0.07 (1.34)	1.12 (0.52)	0.67 (0.57)
	12	-1.05 (2.51)	0.78 (1.02)	-0.12 (1.07)
	24	-2.09 (5.66)	-0.80 (1.53)	-1.57 (1.48)
IP	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 and $100 \times \ln(IP)$]

$$Y_{i,t+h} = \mathrm{D}_{h,i1} Y_{1,t} + \gamma' W_t + \{\varepsilon_{t+h}, \ldots, \varepsilon_{t+1}, \varepsilon_{2:n_\varepsilon,t}, \varepsilon_{t-1}, \ldots\}$$

	lag (h)	(a)	(b)	(c)
R	0	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
	6	-0.07 (1.34)	1.12 (0.52)	0.67 (0.57)
	12	-1.05 (2.51)	0.78 (1.02)	-0.12 (1.07)
	24	-2.09 (5.66)	-0.80 (1.53)	-1.57 (1.48)
IP	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

SVARs with External Instruments - 1

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

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

($D_0 = H$ 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 $n_y = n_\varepsilon$)

SVARs with External Instruments - 2

$$A(L)Y_t = \nu_t = D_0\varepsilon_t$$

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

$$\text{thus } D_{h,i1} = C_h H_{i1}$$

Unit-effect normalization yields: $\eta_{i,t} = H_{i1}\,\eta_{1,t} + \{\varepsilon_{2:n_\varepsilon,t}\}$

Condition SVAR-IV

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

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

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 $\hat{C}(\mathrm{L}) = \hat{A}(\mathrm{L})^{-1}$.

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

$$\hat{D}_{h,1} = \hat{\mathrm{C}}_h \hat{H}_1$$

(odds and ends: (1) News shocks; (2) Dif. sample periods in (1) and (2))

SVAR with external instruments – inference

- Strong instruments:

$$\sqrt{T}\begin{pmatrix} \hat{A} - A \\ \hat{H}_1 - H_1 \end{pmatrix} \xrightarrow{d} \text{Normal} + \delta\text{-method}$$

- Weak instruments:
  - $\sqrt{T}(\hat{A} - A) \xrightarrow{d}$ Normal.
  - $\hat{H}_1 - H_1 \xrightarrow{d}$ NonNormal.
  - Use weak-instrument robust methods. (Montiel Olea, Stock and Watson (2018)).

Results for [R and $100 \times \ln(IP)$]

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

Results for [R and $100 \times \ln(IP)$]

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

SDFM
SVAR analysis, but now using DFM

SVAR problems that the DFM might solve:

(a) Many variable, thus invertibility is more plausible.

(b) Errors-in-variables, several indicators for same theoretical concept ('aggregate prices','oil prices', etc.)

(c) Framework for computing IRFs from structural shocks to many variables.

Can't I just do a VAR? .. No

Table 5 Approximating the eight-factor DFM by a eight-variable VAR

	Canonical correlation							
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
(A) Innovations								
VAR-A	0.76	0.64	0.6	0.49				
VAR-B	0.83	0.67	0.59	0.56	0.37	0.33	0.18	0.01
VAR-C	0.86	0.81	0.78	0.76	0.73	0.58	0.43	0.35
VAR-O	0.83	0.80	0.69	0.56	0.50	0.26	0.16	0.02
(B) Variables and factors								
VAR-A	0.97	0.85	0.79	0.57				
VAR-B	0.97	0.95	0.89	0.83	0.61	0.43	0.26	0.10
VAR-C	0.98	0.93	0.90	0.87	0.79	0.78	0.57	0.41
VAR-O	0.98	0.96	0.88	0.84	0.72	0.39	0.18	0.02

Notes: All VARs contain four lags of all variables. The canonical correlations in panel A are between the VAR residuals and the residuals of a VAR estimated for the eight static factors.
VAR-A was chosen to be typical of four-variable VARs seen in empirical applications. Variables: GDP, total employment, PCE inflation, and Fed funds rate.
VAR-B was chosen to be typical of eight-variable VARs seen in empirical applications. Variables: GDP, total employment, PCE inflation, Fed funds, ISM manufacturing index, real oil prices (PPI-oil), corporate paper-90-day treasury spread, and 10 year–3 month treasury spread.
VAR-C variables were chosen by stepwise maximization of the canonical correlations between the VAR innovations and the static factor innovations. Variables: industrial commodities PPI, stock returns (SP500), unit labor cost (NFB), exchange rates, industrial production, Fed funds, labor compensation per hour (business), and total employment (private).
VAR-O variables: real oil prices (PPI-oil), global oil production, global commodity shipment index, GDP, total employment (private), PCE inflation, Fed funds rate, and trade-weighted US exchange rate index.
Entries are canonical correlations between (A) factor innovations and VAR residuals and (B) factors and observable variables.

The SDFM:

$$\underset{t}{\overset{n\times 1}{X}} = \overset{n\times r}{\Lambda}\ \underset{t}{\overset{r\times 1}{F}} + \underset{t}{\overset{n\times 1}{e}}$$

$$\overset{r\times r}{\Phi(L)}\ \underset{t}{\overset{r\times 1}{F}} = \overset{r\times q}{G}\ \underset{t}{\overset{q\times 1}{\eta}}$$

where $\Phi(\mathrm{L}) = \mathrm{I} - \Phi_1\mathrm{L} - \ldots - \Phi_p\mathrm{L}^p$,

$$\underset{t}{\overset{q\times 1}{\eta}} = \overset{q\times q}{H}\ \underset{t}{\overset{q\times 1}{\varepsilon}}$$

$X_t = \Lambda\Phi(\mathrm{L})^{-1}GH\varepsilon_t + e_t$

IRFs: $\Lambda\Phi(\mathrm{L})^{-1}GH$

IRF from ε_{1t}: $\Lambda\Phi(\mathrm{L})^{-1}GH_1$

Three Normalizations

1. $\Lambda F_t = \Lambda PP^{-1}F_t$ for any matrix P. Set P rows of Λ equal to rows of identity matrix. Rearranging the order of the *Xs* this yields

$$\begin{pmatrix} X_{1:r} \\ X_{r+1:n} \end{pmatrix}_t = \begin{pmatrix} I_r \\ \Lambda_{r+1:n} \end{pmatrix} F_t + e_t$$

This 'names' the first factor as the X_1 factor, the second factor as the X_2 factor and so forth. Example: $X_{1,t}$ is the logarithm of oil prices, then $F_{1,t}$ is called the oil price factor.

2. G = I (if $q = r$) or $G_{1:q} = I_q$ if $q < r$. Recall

$$X_t = \lambda(L)f_t + e_t \text{ and } \phi(L)f_t = \eta_t$$

$$X_t = \begin{pmatrix} \lambda_0 & \lambda_1 & \cdots & \lambda_k \end{pmatrix} \begin{pmatrix} f_t \\ f_{t-1} \\ \vdots \\ f_{t-k} \end{pmatrix} + e_t$$

$$\begin{pmatrix} f_t \\ f_{t-1} \\ \vdots \\ f_{t-k} \end{pmatrix} = \begin{bmatrix} \phi_1 & \phi_2 & \cdots & \phi_{k+1} \\ 1 & 0 & \cdots & 0 \\ & \ddots & \ddots & \\ & & 1 & 0 \end{bmatrix} \begin{pmatrix} f_{t-1} \\ f_{t-2} \\ \vdots \\ f_{t-k-1} \end{pmatrix} + \begin{pmatrix} I \\ 0 \\ \vdots \\ 0 \end{pmatrix} \eta_t$$

where f_t and η_t are $q \times 1$.

3. The diagonal elements of H are unity. That is, ε_{1t} has a unit effect of $F_{1,t}$ and so forth. Same as in SVAR.

Putting these together:

$$X_{1:q,t} = H\varepsilon_t + \text{lags of } \varepsilon_t + e_t$$

(Same normalization used in SVAR, but only applied to the first q elements of X_t).

$$F_{1:q,t} = H\varepsilon_t + \text{lags of } \varepsilon_t$$

etc.

This means that everything in SVARs carry over here.

Additional flexibility in SDFM

(1) Measurement error allowed: With normalization, F follows SVAR, and $X = \Lambda F + e$.

(2) Multiple measurements: Example Oil prices

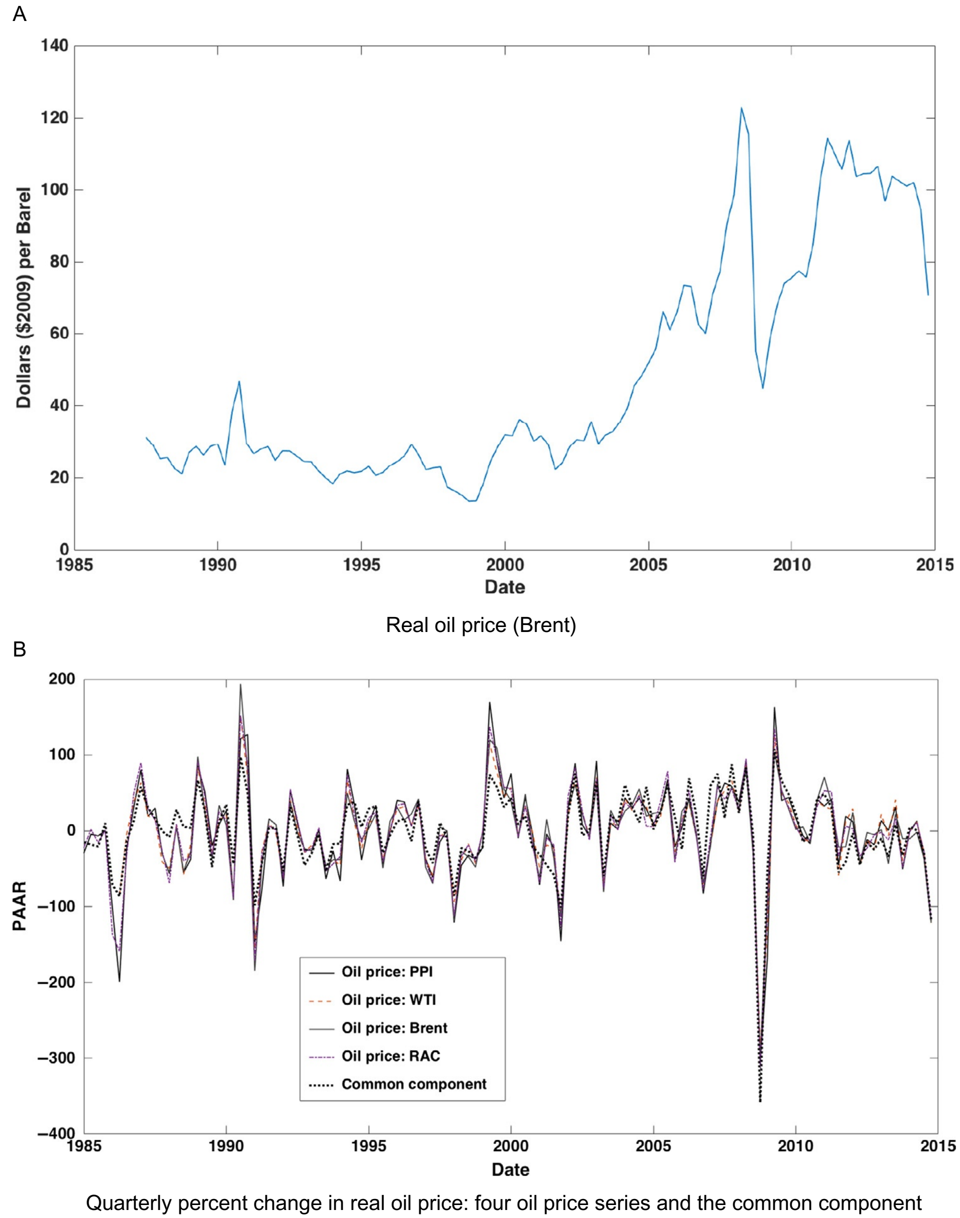


Fig. 7 Real oil price (2009 dollars) and its quarterly percent change.

$$
\begin{bmatrix} p_t^{PPI-Oil} \\ p_t^{Brent} \\ p_t^{WTI} \\ p_t^{RAC} \\ X_{5:n,t} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ & \Lambda_{5:n} & & \end{bmatrix} \begin{bmatrix} F_t^{oil} \\ F_{2:r,t} \end{bmatrix} + e_t
$$

(3) "Factor Augmented" VAR) (FAVAR) (Bernanke, Boivin, Eliasz (2005))

Easily implemented in this framework:

$$\begin{pmatrix} Y_t \\ X_t \end{pmatrix} = \begin{pmatrix} 1 & 0_{1\times r} \\ \multicolumn{2}{c}{\Lambda} \end{pmatrix} \begin{pmatrix} \tilde{F}_t \\ F_t \end{pmatrix} + \begin{pmatrix} 0 \\ e_t \end{pmatrix}$$

$$F_t^+ = \Phi(L) F_{t-1}^+ + G\eta_t$$

where

$$F_t^+ = \begin{pmatrix} \tilde{F}_t \\ F_t \end{pmatrix},$$

$$\eta_t = H\varepsilon_t.$$

Example: Macroeconomic Effects of Oil Supply Shocks

2 Identifications:

(1) Oil Price exogenous

$$\eta_t = \begin{pmatrix} 1 & 0 \\ H_{\bullet 1} & H_{\bullet\bullet} \end{pmatrix} \begin{pmatrix} \varepsilon_t^{oil} \\ \tilde{\eta}_{\bullet t} \end{pmatrix}$$

$$\begin{bmatrix} p_t^{PPI-Oil} \\ p_t^{Brent} \\ p_t^{WTI} \\ p_t^{RAC} \\ X_{5:n,t} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ \lambda_{21} & \lambda_{22} & \lambda_{23} & \cdots & \lambda_{28} \\ \lambda_{31} & & & \cdots & \lambda_{38} \\ \lambda_{41} & & & \cdots & \lambda_{48} \\ \Lambda_{5:n} & & & & \end{bmatrix} \begin{bmatrix} F_t^{oilprice} \\ F_{2,t} \\ F_{3,t} \\ \vdots \\ F_{8,t} \end{bmatrix} + \begin{bmatrix} e_t^{PPI-oil} \\ e_t^{Brent} \\ e_t^{WTI} \\ e_t^{RAC} \\ e_t^{X} \end{bmatrix}$$

SVAR, FAVAR and SDFM versions

(2) Killian (2009) Identification

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ H_{12} & 1 & 0 & 0 \\ H_{13} & H_{23} & 1 & 0 \\ H_{1\bullet} & H_{2\bullet} & H_{3\bullet} & H_{\bullet\bullet} \end{pmatrix} \begin{pmatrix} \varepsilon_t^{OS} \\ \varepsilon_t^{GD} \\ \varepsilon_t^{OD} \\ \tilde{\eta}_{\bullet t} \end{pmatrix}$$

$$\begin{bmatrix} GlobalActivity_t \\ p_t^{PPI-Oil} \\ p_t^{Brent} \\ p_t^{WTI} \\ p_t^{RAC} \\ X_{7:n,t} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & 0 & \cdots & 0 \\ \Lambda_{8:n} & & & & & \end{bmatrix} \begin{bmatrix} F_t^{Globalactivity} \\ F_t^{oilprice} \\ F_{4:r,t} \end{bmatrix} + e_t$$

Some Results

Table 6 Fraction of the variance explained by the eight factors at horizons $h = 1$ and $h = 6$ for selected variables: 1985:Q1–2014:Q4

Variable	$h = 1$	$h = 6$
GDP	0.60	0.80
Consumption	0.37	0.76
Fixed investment	0.38	0.76
Employment (non-ag)	0.56	0.94
Unemployment rate	0.44	0.90
PCE inflation	0.70	0.63
PCE inflation—core	0.10	0.34
Fed funds rate	0.48	0.71
Real oil price	0.74	0.78
Oil production	0.06	0.27
Global commodity shipment index	0.39	0.51
Real gasoline price	0.72	0.80

Oil Price Exogenous

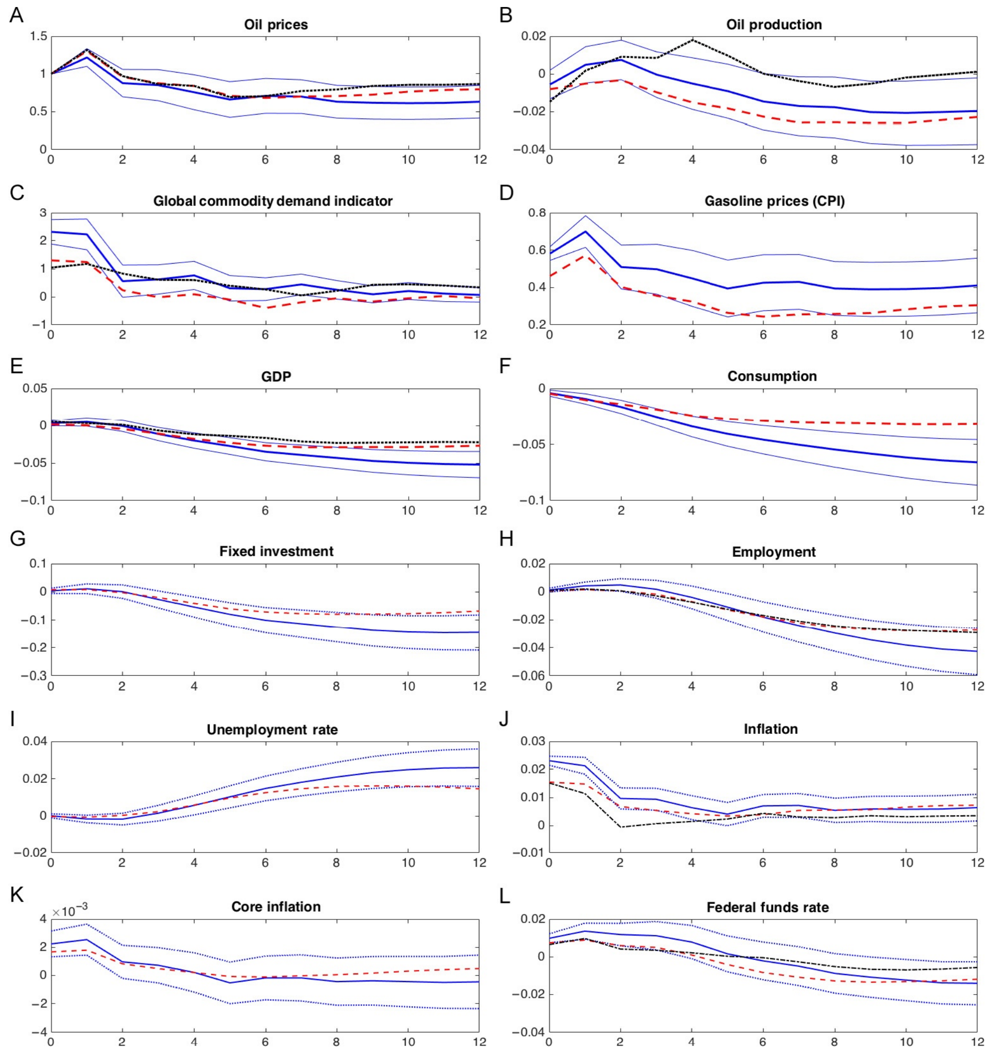


Fig. 8 Structural IRFs from the SDFM (*blue* (*dark gray* in the print version) *solid* with ±1 standard error bands), FAVAR (*red* (*gray* in the print version) *dashed*), and SVAR (*black dots*) for selected variables with respect to an oil price shock: "oil prices exogenous" identification. Units: standard deviations for Global Commodity Demand and percentage points for all other variables.

Killian identification IRFs (see paper)

Variance Explained:

Table 7 Forecast error variance decompositions for six periods ahead forecasts of selected variables: FAVARs and SDFMs

	A. Oil price exogenous		B. Kilian (2009) identification					
			Oil supply		Global demand		Oil spec. demand	
Variable	**F**	**D**	**F**	**D(O)**	**F**	**D(U)**	**F**	**D(U)**
GDP	0.07	0.07	0.04	0.01	0.02	0.04	0.09	0.04
Consumption	0.19	0.22	0.09	0.08	0.02	0.22	0.11	0.01
Fixed investment	0.04	0.04	0.05	0.04	0.03	0.04	0.03	0.01
Employment (non-ag)	0.03	0.02	0.04	0.01	0.02	0.01	0.03	0.01
Unemployment rate	0.04	0.03	0.04	0.03	0.02	0.03	0.04	0.01
PCE inflation	0.28	0.40	0.02	0.04	0.09	0.16	0.17	0.29
PCE inflation—core	0.05	0.04	0.01	0.02	0.03	0.05	0.02	0.02
Fed funds rate	0.02	0.04	0.00	0.01	0.05	0.11	0.03	0.02
Real oil price	0.81	0.53	0.14	0.10	0.22	0.44	0.42	0.09
Oil production	0.03	0.01	0.75	0.78	0.07	0.02	0.03	0.01
Global commodity shipment index	0.11	0.23	0.05	0.07	0.79	0.33	0.03	0.02
Real gasoline price	0.61	0.48	0.05	0.06	0.25	0.43	0.34	0.08

Notes: Entries are the fractions of the six periods ahead forecast error of the row variable explained by the column shock, for the "oil price exogenous" identification results (columns A) and the Kilian identification scheme (columns B). For each shock, "F" refers to the FAVAR treatment in which the factor is treated as observed and "D" refers to the SDFM treatment. In the hybrid SDFM using the Kilian (2009) identification scheme, the oil supply factor is treated as observed (the oil production variable) (D(O)) while the global demand and oil-specific demand factors are treated as unobserved (D(U)).