ECON 616: Lecture 6: Bayesian Structural Vector Autoregressions

Ed Herbst

Last Updated: October 9, 2018

Bayesian VARs

Want to do Bayesian estimation of VARs for better model properties.

Use Minnesota Prior of Doan, Litterman, and Sims (1984) to *shrink* prior towards “reasonable” parameter configurations.

First, we need some preliminaries…

The Inverse Wishart Distribution

The multivariate version of the inverted Gamma distribution is called Wishart Distribution. Let W be a $n \times n$ positive definite random matrix. W has the inverted Wishart $IW(S, \nu)$ distribution if its density is of the form

$$p(W|S, \nu) \propto |S|^{\nu/2}|W|^{-(\nu+n+1)/2} \exp\left\{-\frac{1}{2}tr[W^{-1}S]\right\} \quad (1)$$

The Wishart distribution arises in the Bayesian analysis of multivariate regression models. To sample a W from an inverted Wishart $IW(S, \nu)$ distribution, draw $n \times 1$ vectors $Z_1, \ldots, Z_\nu$ from a multivariate normal $\mathcal{N}(0, S^{-1})$ and let

$$W = \left[\sum_{i=1}^{\nu} Z_i Z_i'\right]^{-1}$$

Note: to generate a draw Z from a multivariate $\mathcal{N}(\mu, \Sigma)$, decompose $\Sigma = CC'$, where C is the lower triangular Cholesky decomposition matrix. Then let $Z = \mu + C\mathcal{N}(0, \mathcal{I})$.

Dummy Observation Priors

Suppose we have T^* dummy observations (Y^*, X^*). The likelihood function for the dummy observations is of the form

$$p(Y^*|\Phi, \Sigma_u) = (2\pi)^{-nT^*/2}|\Sigma_u|^{-T^*/2} \quad (2)$$
$$\exp\left\{-\frac{1}{2}tr[\Sigma_u^{-1}(Y^{*'}Y^* - \Phi' X^{*'}Y^* - Y^{*'}X^*\Phi + \Phi' X^{*'}X^*\Phi)]\right\}.$$

Combining~(2) with the improper prior $p(\Phi, \Sigma_u) \propto |\Sigma_u|^{-(n+1)/2}$ yields

$$\Phi, \Sigma_u|Y^*) \;=\; c_*^{-1}|\Sigma_u|^{-\frac{T^*+n+1}{2}}$$
$$exp\left\{-\frac{1}{2}tr[\Sigma_u^{-1}(Y^{*'}Y^* - \Phi' X^{*'}Y^* - Y^{*'}X^*\Phi + \Phi' X^{*'}X^*\Phi)]\right\}$$

which can be interpreted as a prior density for Φ and Σ_u.

Define

$$\hat{\Phi}^* = (X^{*'}X^*)^{-1}X^{*'}Y^*$$
$$S^* = (Y^* - X^*\hat{\Phi}^*)'(Y^* - X^*\hat{\Phi}^*).$$

It can be verified that the prior $p(\Phi, \Sigma_u|Y^*)$ is of the Inverted Wishart-Normal $\mathcal{IW} - \mathcal{N}$ form

$$\Sigma_u \sim \mathcal{IW}\Big(S^*, T^* - k\Big) \quad (3)$$

$$\Phi|\Sigma_u \sim \mathcal{N}\Big(\Phi^*, \Sigma_u \otimes (X^{*'}X^*)^{-1}\Big). \quad (4)$$

The appropriate normalization constant for the prior density is given by

$$c_* = (2\pi)^{\frac{nk}{2}}|X^{*'}X^*|^{-\frac{n}{2}}|S^*|^{-\frac{T^*-k}{2}} \quad (5)$$
$$2^{\frac{n(T^*-k)}{2}}\pi^{\frac{n(n-1)}{4}}\prod_{i=1}^{n}\Gamma[(T^*-k+1-i)/2],$$

k is the dimension of x_t and $\Gamma[\cdot]$ denotes the gamma function. Details of this calculation can be found in Zellner (1971).

Minnesota Prior

Here is a brief description of the "Minnesota Prior," see Doan, Litterman, and Sims (1984). Consider the following Gaussian bivariate VAR(2).

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} + \begin{bmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \begin{bmatrix} y_{1,t-2} \\ y_{2,t-2} \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

Define $y_t = [y_{1,t}, y_{2,t}]'$, $x_t = [y'_{t-1}, y'_{t-2}, 1]'$, and $u_t = [u_{1,t}, u_{2,t}]'$ and

$$\Phi = \begin{bmatrix} \beta_{11} & \beta_{21} \\ \beta_{12} & \beta_{22} \\ \gamma_{11} & \gamma_{21} \\ \gamma_{12} & \gamma_{22} \\ \alpha_1 & \alpha_2 \end{bmatrix}. \tag{7}$$

The VAR can be rewritten as follows

$$y'_t = x'_t \Phi + u'_t, \quad t = 1, \ldots, T, \quad u_t \sim iid\mathcal{N}(0, \Sigma_u) \tag{8}$$

or in matrix form

$$Y = X\Phi + U. \tag{9}$$

Preliminaries

Based on a short pre-sample Y_0 (typically the observations used to initialized the lags of the VAR) one calculates: $s = std(Y_0)$ and $\bar{y} = mean(Y_0)$. In addition there are a number of tuning parameters for the prior

- τ is the overall tightness of the prior. Large values imply a small prior covariance matrix.
- d: the variance for the coefficients of lag h is scaled down by the factor l^{-2d}.
- w: determines the weight for the prior on Σ_u. Suppose that $Z_i = \mathcal{N}(0, \sigma^2)$. Then an estimator for σ^2 is \$^{2} = 1 $\frac{}{w \sum_{i=1}^{w}}$ Z_i^2.\$ The larger w, the more informative the estimator, and in the context of the VAR, the tighter the prior.
- λ and μ: additional tuning parameters.

Dummies for β coefficients

Dummies for the β coefficients:

$$\begin{bmatrix} \tau s_1 & 0 \\ 0 & \tau s_2 \end{bmatrix} = \begin{bmatrix} \tau s_1 & 0 & 0 & 0 & 0 \\ 0 & \tau s_2 & 0 & 0 & 0 \end{bmatrix} \Phi + u'$$

The first observation implies, for instance, that

$$\begin{aligned} \tau s_1 &= \tau s_1 \beta_{11} + u_1 \quad \Longrightarrow \beta_{11} = 1 - \frac{u_1}{\tau s_1} \quad \Longrightarrow \quad \beta_{11} \sim \mathcal{N}\left(1, \frac{\Sigma_{u,11}}{\tau^2 s_1^2}\right) \\ 0 &= \tau s_1 \beta_{21} + u_2 \quad \Longrightarrow \beta_{21} = -\frac{u_2}{\tau s_1} \quad \Longrightarrow \quad \beta_{21} \sim \mathcal{N}\left(0, \frac{\Sigma_{u,22}}{\tau^2 s_1^2}\right) \end{aligned}$$

More Dummies

Dummies for the γ coefficients

$$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & \tau s_1 2^d & 0 & 0 \\ 0 & 0 & 0 & \tau s_2 2^d & 0 \end{bmatrix} \Phi + u'$$

The prior for the covariance matrix is implemented by

$$\begin{bmatrix} s_1 & 0 \\ 0 & s_2 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \Phi + u'$$

Co-persistence prior dummy observations, reflecting the belief that when data on all y's are stable at their initial levels, thy will tend to persist at that level:

$$\begin{bmatrix} \lambda \bar{y}_1 & \lambda \bar{y}_2 \end{bmatrix} = \begin{bmatrix} \lambda \bar{y}_1 & \lambda \bar{y}_2 & \lambda \bar{y}_1 & \lambda \bar{y}_2 & \lambda \end{bmatrix} \Phi + u'$$

Own-persistence prior dummy observations, reflecting the belief that when y_i has been stable at its initial level, it will tend to persist at that level, regardless of the value of other variables:

$$\begin{bmatrix} \mu \bar{y}_1 & 0 \\ 0 & \mu \bar{y}_2 \end{bmatrix} = \begin{bmatrix} \mu \bar{y}_1 & 0 & \mu \bar{y}_1 & 0 & 0 \\ 0 & \mu \bar{y}_2 & 0 & \mu \bar{y}_2 & 0 \end{bmatrix} \Phi + u'$$

Training Sample Priors

In the same way we constructed a prior from dummy observations, we can also construct a prior from a training sample. Suppose we split the actual sample $Y = [Y^-, Y^+]$, where Y^- is interpreted as training sample, then

$$
\begin{aligned}
p(\Phi, \Sigma_u) &= c_-^{-1} |\Sigma_u|^{-\frac{T^- + n + 1}{2}} \\
&\quad \left\{ -\frac{1}{2} tr[\Sigma_u^{-1} (Y^{-'} Y^- - \Phi' X^{-'} Y^- - Y^{-'} X^- \Phi + \Phi' X^{-'} X^- \Phi)] \right\}
\end{aligned}
$$

Of course one can also combine the dummy observations and training sample to construct a prior distribution.

Posteriors

Notice that

$$p(\Phi, \Sigma_u | Y) \propto p(Y|\Phi, \Sigma_u) p(Y^*|\Phi, \Sigma_u) \tag{10}$$

Now define:

$$\tilde{\Phi} = (X^{*'}X^* + X'X)^{-1}(X^{*'}Y^* + X'Y) \tag{11}$$

$$\tilde{\Sigma}_u = \frac{1}{T^* + T}\Big[Y^{*'}Y^* + Y'Y) - (X^{*'}Y^* + X'Y)'(X^{*'}X^* + X'X)^{-1}(X^{*'}Y^* + X'Y)\Big] \tag{12}$$

Since prior and likelihood function are conjugate, it is straightforward to show, e.g., Zellner (1971), that the posterior distribution of Φ and Σ_u is also of the Inverted Wishart – Normal form:

$$\Sigma_u | Y \sim \mathcal{IW}\Big((T^* + T)\tilde{\Sigma}_u, T^* + T - k\Big) \tag{13}$$

$$\Phi | \Sigma_u, Y \sim \mathcal{N}\Big(\tilde{\Phi}, \Sigma_u \otimes (X^{*'}X^* + X'X)^{-1}\Big). \tag{14}$$

Marginal Data Density

Suppose that we are using a prior constructed from a training sample and dummy observations. Then the marginal data density is given by

$$p(Y^+|Y^-, Y^*, \mathcal{M}_0) = \frac{\int p(Y^+, Y^-, Y^*|\Phi, \Sigma_u) d\Phi d\Sigma_u}{\int p(Y^-, Y^*|\Phi, \Sigma_u) d\Phi d\Sigma_u} \qquad (15)$$

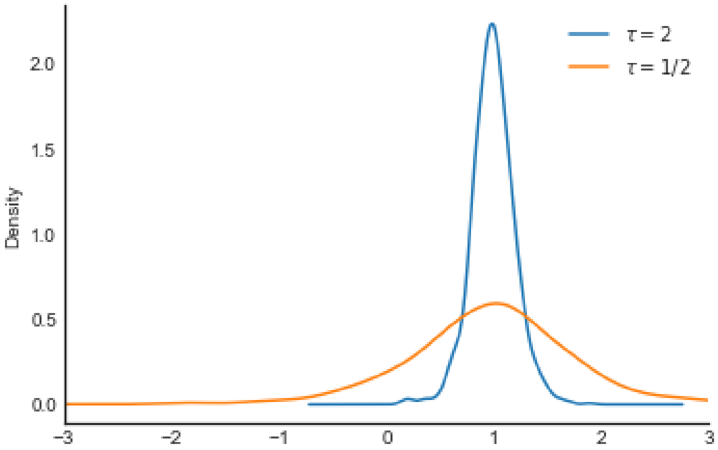
where the integrals in the numerator and denominator are given by the appropriate modification of c_* defined above. More specifically:

$$\int p(Y|\Phi, \Sigma_u) d\Phi d\Sigma_u = \pi^{-\frac{T-k}{2}} |X'X|^{-\frac{n}{2}} |S|^{-\frac{T-k}{2}} \pi^{\frac{n(n-1)}{4}} \prod_{i=1}^{n} \Gamma[(T-k+1-i)/2],$$

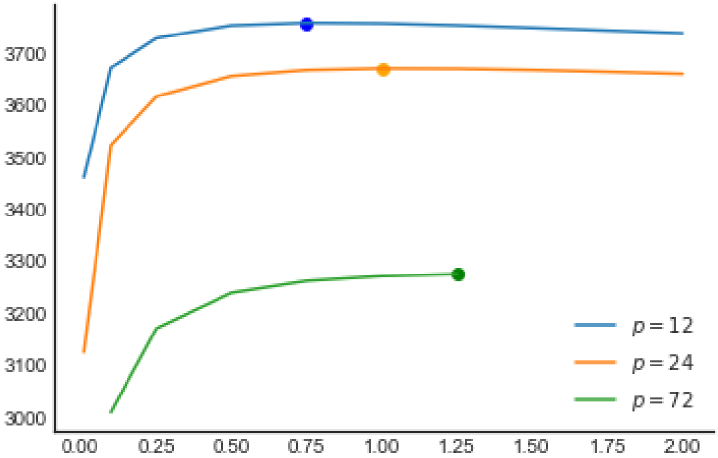
where

$$\begin{aligned} \hat{\Phi} &= (X'X)^{-1} X'Y \\ S &= (Y - X\hat{\Phi})'(Y - X\hat{\Phi}). \end{aligned}$$

An example



Log MDD as function of τ



More on SVARS

- we talked generally about the identification problem in SVARs.
- We described how identification amounted to putting restrictions on the mapping from reduced-form to structural representation of the VAR.
- We introduced some basic identification schemes in a simple model.
- Today I'm going to present an algorithm for estimating SVARs for general identification schemes.
- Then I'm going to talk about attempts to introduce more data to help solve identification problem.

SVARs, again

A canonical way to write an SVAR(p):

$$
\begin{aligned}
y_t' A_0 &= y_{t-1}' A_1 + \ldots + y_{t-p} A_p' + c + \epsilon_t', \quad \epsilon_t(0, I) \\
&= x_t A_+ + \epsilon_t'
\end{aligned}
\tag{17}
$$

the reduced-form is

$$
y_t' = x_t' \Phi + u_t', \quad u_t \sim (0, \Sigma)
\tag{18}
$$

with $\Sigma = (A_0 A_0')^{-1}$ and $\Phi = A_+ A_0^{-1}$. Recall $A_0 = \Sigma_{tr}^{-1'} \Omega$ where Ω is an orthogonal matrix.

We're going to talk about a general algorithm for estimating SVARs from Arias, Rubio-Ramirez, Waggoner (2016).

Estimating SVARs

We're going to work like this:

(1) Estimate (Φ, Σ) + (2) Add Assumptions $\longrightarrow$ Get (A_0, A_+).

How to estimate (Φ, Σ)? Let's Go Bayesian (but you don't have too!)

$$p(\Phi, \Sigma | Y) \propto p(Y | \Phi, \Sigma) p(\Phi, \Sigma)$$

$$\begin{aligned} p(\Phi, \Sigma) &: \text{ Minnesota Prior} \\ p(Y | \Phi, \Sigma) &: \text{ Normal Likelihood} \end{aligned}$$

$\Longrightarrow$ the posterior of Φ, Σ is Normal-Inverse Wishart.

How to get to (2)? Well we need to go back to Ω...

A Prior on Ω

Last class we talked about a number of different choices for Ω, sometimes as a function of the parameters Φ and Σ.

Formalize this as a prior over $\Omega : p(\Omega|\Phi, \Sigma)$.

Since Ω does not enter the likelihood, this prior doesn't get updated by the data:

$$p(\Omega|Y, \Phi, \Sigma) = p(\Omega|\Phi, \Sigma).$$

Let's just forget about possible conditioning and focus on $p(\Omega)$ for now.

It's hard to put a prior on the space of orthogonal matrices.

How about a prior which places equal weight on every $\Omega \in \mathcal{O}(n) = \{\Omega|\Omega\Omega' = I_n\}$.

The Haar Measure

Remember our characterization of the 2x2 orthogonal matrix

$$\Omega(\varphi) = \begin{bmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{bmatrix} \tag{19}$$

where $\varphi \in (-\pi, \pi]$

If we put a uniform prior on φ, we obtained a uniform distribution over Ω.

This is tedious in higher dimensions. Need a more general concept.

Random matrix theory Haar measure is invariant measure for (in this case) orthogonal matrices.
The theory is difficult, but the upshot is that we can generate from a uniform distribution over $O(n)$ by simulating a $n \times n$ matrix of independent normals, and then taking Q from the QR decomposition of that matrix.

An Example

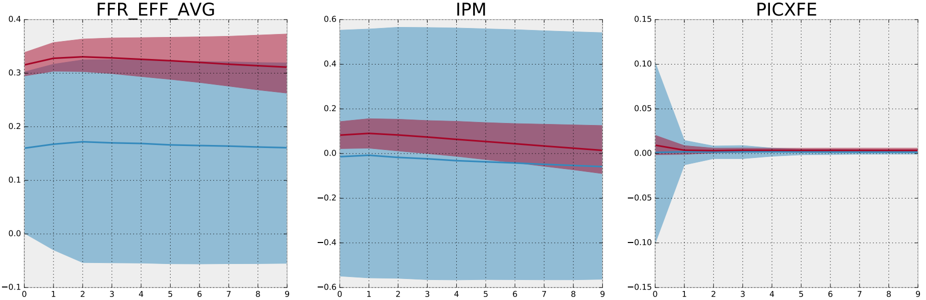
- $Y_t =$ [Industrial Production, GDP Deflator, Federal Funds Rate]
- Frequency = monthly
- Date = 1994-01:2007-06
- Minnesota Prior $\lambda = [0.5, 3, 1, 0.5, 0.5, 1]$.

Let's estimate the VAR (i.e., Φ and Σ.)

Then draw $p(\Omega)$ from the uniform distribution (Haar measure) and look at the response to a monetary policy shock.

Minor detail: we are going to sign-normalize so that because A_0, A_+ is only identified up to a sign.

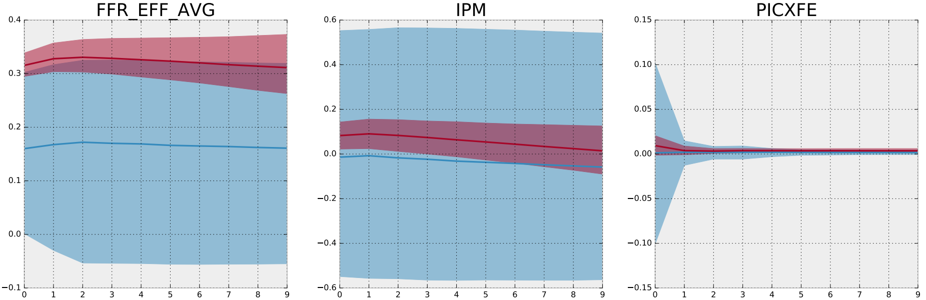
The effect of a monetary policy shock



The effect of a monetary policy shock

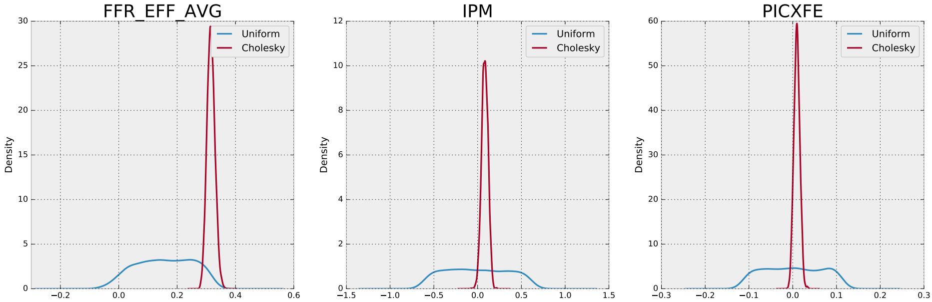
Recall that the Cholesky factorization refers to the dogmatic prior $p(\Omega) = I$.

What does that look like?



In terms of the response of the federal funds rate, the cholesky is on the upper edge of the unidentified model.

Impact Effect of Monetary Policy Shock



Beware: Even though $\Omega \sim$ Uniform, the IRF need not be. Nonlinear functions of uniform RV aren't necessarily uniform (criticism of SVAR by Baumeister and Hamilton 2016).

A Word on The Cholesky

Why is the response of FFR so extreme for Choleksy identification?
Recall:

$$A_0 = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ 0 & 0 & a_{33} \end{bmatrix}$$

and $\Sigma = (A_0 A_0')^{-1}$.

This means that $\Sigma_{1,1} = \frac{1}{a_{11}^2}$.

All of the forecast error Σ_{11} for the interest rate needs to be explained by monetary shock!

Beyond Cholesky and "Uniformity"

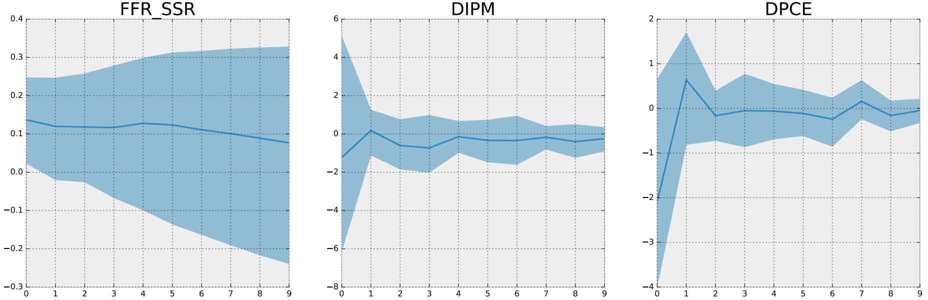
- By drawing from $\Phi, \Sigma | Y$ and then Ω, we were able to draw from A_0, A_+.
- Sign restrictions complicate things only some (A_0, A_+) draws will be valid.
- A simple rejection sampler will accomplish this.

Algorithm 2. Drawing from an SVAR Posterior with Sign Restrictions

1. Draw (A_0, A_+) from the unrestricted posterior.
2. Keep the draw if the sign restrictions are satisfied.
3. Return to Step 1 until the required number of draws from the posterior of structural parameter conditional on the sign restrictions has been obtained.

A Simple Sign Restricted VAR

Let's assume now that {*After 6 months, inflation must have fallen.*}
$\sum_{i=0}^{6} \sum_{i=0}^{6} (\Phi^i A_0^{-1'})_3 < 0$



zero restrictions? See ARRW.

Romer and Romer (2004)

Romer and Romer, 2004, a New Measure of Monetary Policy Shocks: Derivation and Implications, American Economic Review.

Policy instruments move endogenously with changes in economic conditions

Movements in policy instruments are often responses to information about future economic developments. New indicator of monetary policy shocks that avoids these problems.

- Series for changes in the ‘intended’ Federal Funds Rate around FOMC meetings.
- Control for Federal Reserve Greenbook forecasts

Let's Construct $\epsilon_{MP,t}$

$$\Delta ff_m = \alpha + \beta ffb_m + \sum_{i=-1}^{2} \gamma_i GB_{mi} + \epsilon_{\mathbf{MP,t}}$$

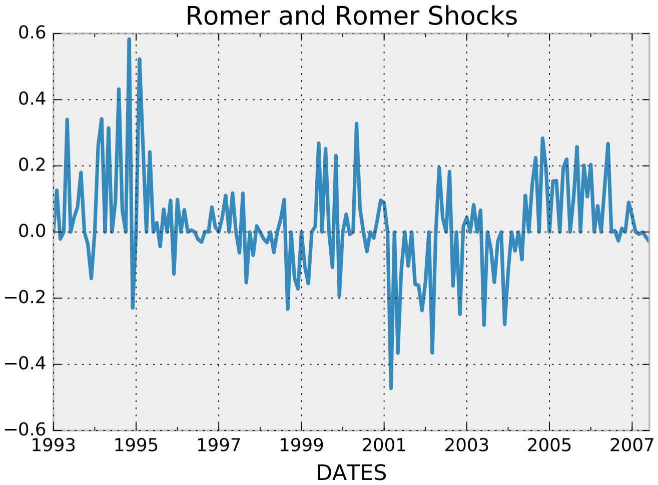
Δff_m is change in the intended funds rate around FOMC meeting m.

ffb_m is the level of the intended funds rate changes associated with meeting m.

GM_{mi} is the Greenbook forecast of output, inflation, and unemployment for quarter i.

$\epsilon_{\mathbf{MP,t}}$ is our monetary policy shock.

The Shocks



Romer and Romer Shocks
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
1993
1995
1997
1999
2001
2003
2005
2007
DATES

If You Have the Shocks, do you even need a VAR?

Romer and Romer run a univariate regression:

$$\Delta y_t = a_0 + \sum_{j=1}^{36} c_j e_{t-j}^{MP} + \sum_{j=1}^{24} b_j \Delta_{t-j} + ...$$

This means that the IRF = $\{0, c_1, c_1 + (c_2 + b_1 c_1), \ldots\}$.

Romer and Romer use a sample that starts in teh 1970s and goes until 1996.

Effect of Monetary Policy on Ouptut

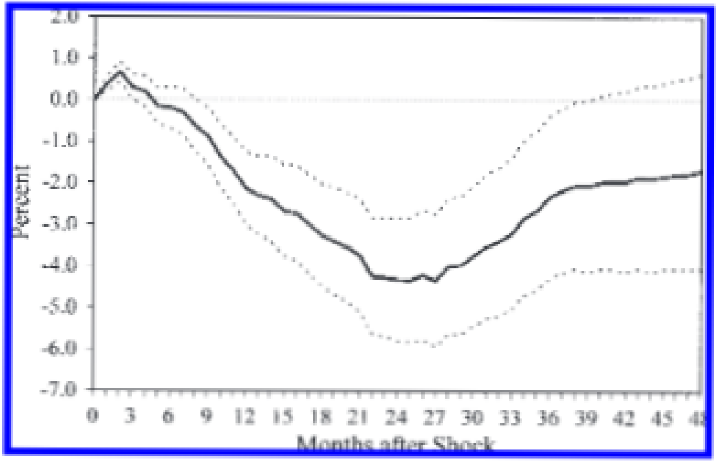
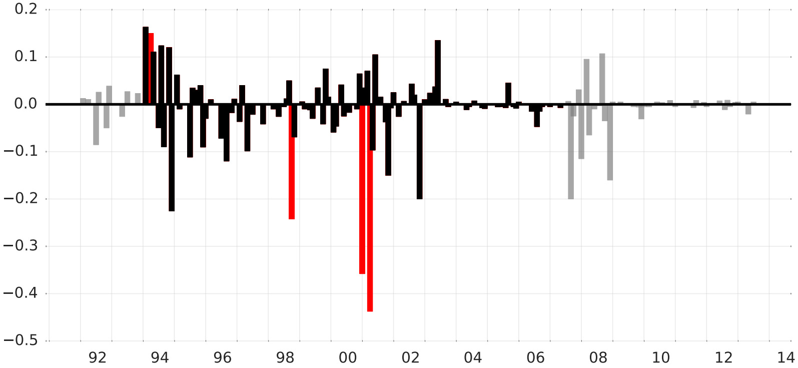


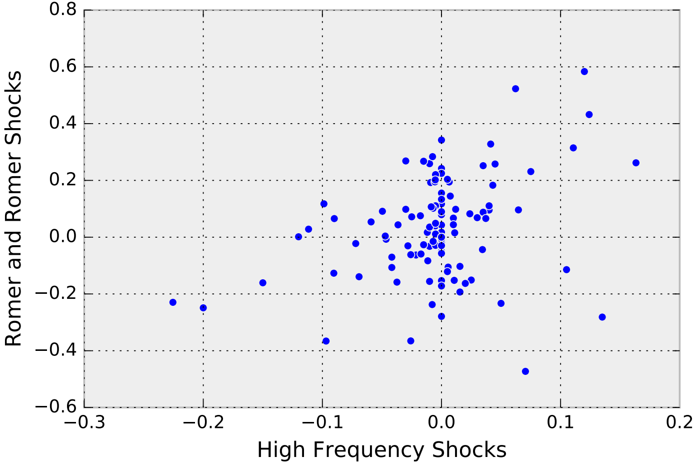
FIGURE 2. THE EFFECT OF MONETARY POLICY ON OUTPUT

More Data, Part 2

- high-frequency event study methodology developed in Kuttner (2001)
- fed funds future – contracts that trade on price of federal funds rate average over a given month.
- shock calculate the change in (appropriately scaled) current-month federal funds rate futures around a tight window surrounding the release of FOMC statements.
- Use an extremely tight window (30 minutes) to calculate this shock.



0.2
0.1
0.0
−0.1
−0.2
−0.3
−0.4
−0.5
92
94
96
98
00
02
04
06
08
10
12
14


0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
−0.3
−0.2
−0.1
0.0
0.1
0.2
Romer and Romer Shocks
High Frequency Shocks

Proxy SVAR (Mertens and Ravn, 2014 AER)

Given two ways of constructing the monetary policy shock yielded two different results, let's an alternative scheme. Call our external measure m_t.

$$
\begin{aligned}
u_{1,t} &= \eta u_{,t} + S_1 \varepsilon_{1,t} \\
u_{2,t} &= \xi u_{1,t} + S_2 \varepsilon_{2,t}
\end{aligned}
$$

- Estimate ξ by instrumenting $u_{1,t}$ with m_t.
- Compute $\tilde{\varepsilon}_{2,t} = u_{2,t} - \xi u_{1,t}$.
- Estimate η by instrumenting $u_{2,t}$ with $\tilde{\varepsilon}_{2,t}$.
- Compute impact matrix:

$$
\boldsymbol{C} = \begin{bmatrix} (I - \eta\xi)^{-1} & \eta(I - \xi\eta)^{-1} \\ \xi(I - \eta\xi)^{-1} & (I - \xi\eta)^{-1} \end{bmatrix} \begin{bmatrix} S_1 & 0 \\ 0 & S_2 \end{bmatrix}
$$

- Impose restrictions on S_1 and S_2 to identify shocks within blocks.

Another Interpretation: Bayesian Proxy SVAR

- Assume m_t is a noisy measure of the structural shock of interest:

$$m_t = \beta e_{1,t} + \sigma_\nu \nu_t, \quad \nu_t \sim \mathcal{N}(0,1) \text{ and } \nu_t \perp e_t.$$

- Assumptions on proxy m_t:
  1. $E[m_t e_{1,t}] = \beta$
  2. $E[m_{te/1,\,t}'] = 0.$
- **Joint likelihood:**

$$p(Y_{1:T}, M_{1:T} | \Phi, \Sigma, \Omega, \beta, \sigma_\nu) =$$
$$\underbrace{p(Y_{1:T}|\Phi,\Sigma)}_{OLD} \times \underbrace{p(M_{1:T}|Y_{1:T},\Phi,\Sigma,\Omega,\beta,\sigma_\nu)}_{NEW}$$

$$\tilde{A}_0 = \begin{bmatrix} A_0 & -\frac{\beta}{\sigma}A_{.1,0} \\ O_{1\times n} & \frac{1}{\sigma} \end{bmatrix}, \text{ and } \tilde{A}_+ = \begin{bmatrix} A_+ & -\frac{\beta}{\sigma}A_{.1,+} \\ O_{1\times n} & 0 \end{bmatrix}. \tag{20}$$

Now $A_{.1,0}$ is identified!

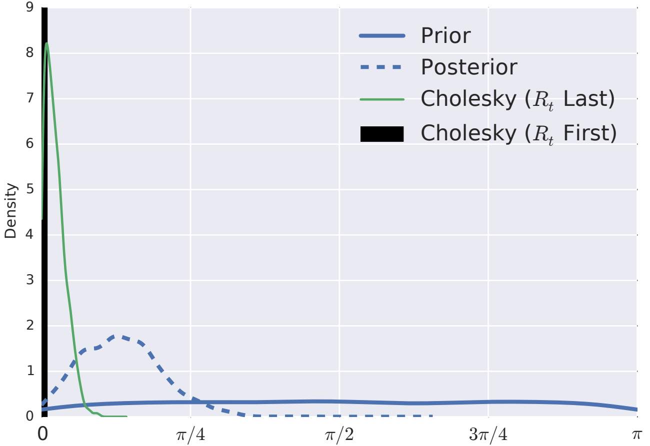
- Caldara-Herbst use this to estimate the effects of MP shock with particular attention to labor market.
- Let's look at a simple example first.
- Data = [R, Credit Spreads]
- Measure of credit spreads comes form Excess Bond Premia [Gilchrist and Zakrasjek (2011)]
- Remember :

$$\Omega(\varphi) = \begin{bmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{bmatrix} \tag{21}$$

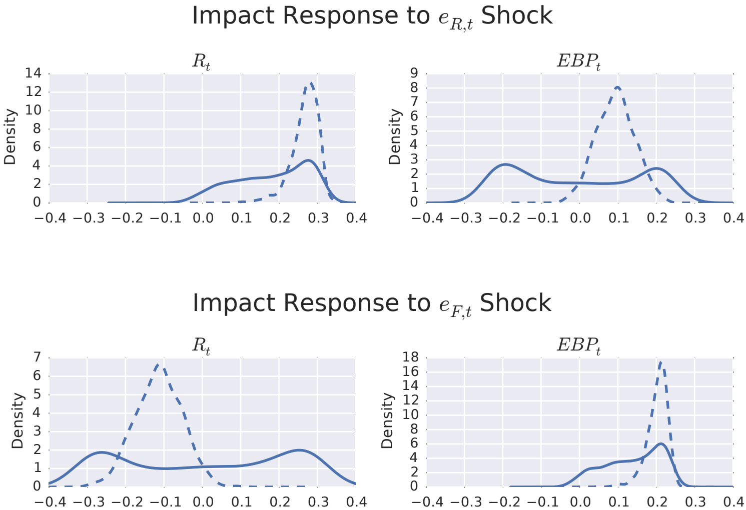
where $\varphi \in (-\pi, \pi]$.

- Compare Prior and Posterior for φ

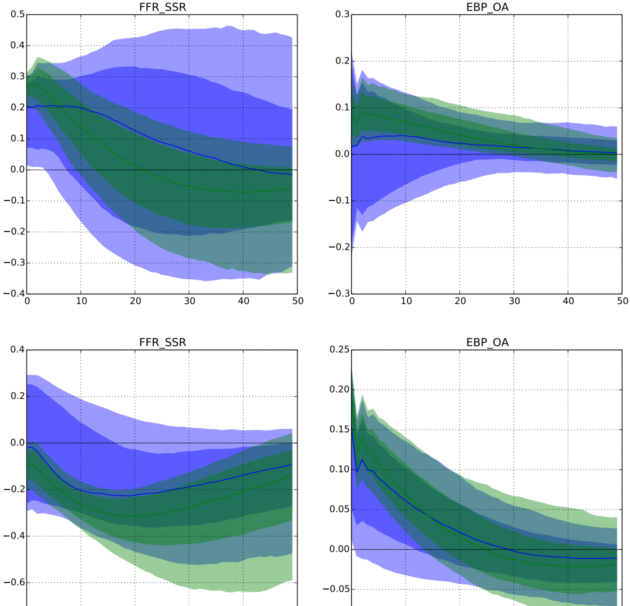
Identification in Proxy SVAR φ



Identification in a Proxy SVAR – Impact IRF



Identification in a Proxy SVAR IRF



Bigger Model