

Inference when identifying assumptions are doubted

- A. Theory
- B. Applications

1

A. Theory

Structural model of interest:

$$\mathbf{A} \mathbf{y}_t = \boldsymbol{\lambda} + \mathbf{B}_1 \mathbf{y}_{t-1} + \dots + \mathbf{B}_m \mathbf{y}_{t-m} + \mathbf{u}_t$$

$(n \times n)$ $(n \times 1)$

$$\mathbf{u}_t \sim \text{i.i.d. } N(\mathbf{0}, \mathbf{D})$$

\mathbf{D} diagonal

2

Bayesian approach:

Summarize whatever information we have that helps identify \mathbf{A} in the form of a density $p(\mathbf{A})$. $p(\mathbf{A})$ is highest for values of \mathbf{A} we think are most plausible.

$p(\mathbf{A}) = 0$ for values of \mathbf{A} we rule out altogether.

3

$p(\mathbf{A})$ could also impose sign restrictions, zeros, or assign small but nonzero probabilities to violations of these constraints.

4

Will use natural conjugate priors for other parameters:

$$p(\mathbf{D}|\mathbf{A}) = \prod_{i=1}^n p(d_{ii}|\mathbf{A})$$

$$d_{ii}^{-1}|\mathbf{A} \sim \Gamma(\kappa_i, \tau_i)$$

$$E(d_{ii}^{-1}|\mathbf{A}) = \kappa_i/\tau_i$$

$$\text{Var}(d_{ii}^{-1}|\mathbf{A}) = \kappa_i/\tau_i^2$$

uninformative priors: $\kappa_i, \tau_i \rightarrow 0$

5

$$\mathbf{B} = \begin{bmatrix} \boldsymbol{\lambda} & \mathbf{B}_1 & \mathbf{B}_2 & \dots & \mathbf{B}_m \end{bmatrix}$$

$$p(\mathbf{B}|\mathbf{D}, \mathbf{A}) = \prod_{i=1}^n p(\mathbf{b}_i|\mathbf{D}, \mathbf{A})$$

$$\mathbf{b}_i|\mathbf{A}, \mathbf{D} \sim N(\mathbf{m}_i, d_{ii}\mathbf{M}_i)$$

uninformative priors: $\mathbf{M}_i^{-1} \rightarrow \mathbf{0}$

6

Recommended default priors (Minnesota prior)

Doan, Litterman, Sims (1984)

Sims and Zha (1998)

- elements of \mathbf{m}_i corresponding to

lag 1 given by \mathbf{a}_i

- all other elements of \mathbf{m}_i are zero
- \mathbf{M}_i diagonal with smaller values on

bigger lags

⇒ prior belief that each element of

\mathbf{y}_i behaves like a random walk

τ_i function of \mathbf{A} (or prior mode of $p(\mathbf{A})$) and scale of data

7

Likelihood:

$$p(\mathbf{Y}_T|\mathbf{A}, \mathbf{D}, \mathbf{B}) = (2\pi)^{-Tn/2} |\det(\mathbf{A})|^T |\mathbf{D}|^{-T/2} \times \exp\left[-(1/2) \sum_{t=1}^T (\mathbf{A}\mathbf{y}_t - \mathbf{B}\mathbf{x}_{t-1})' \mathbf{D}^{-1} (\mathbf{A}\mathbf{y}_t - \mathbf{B}\mathbf{x}_{t-1})\right]$$

prior:

$$p(\mathbf{A}, \mathbf{D}, \mathbf{B}) = p(\mathbf{A})p(\mathbf{D}|\mathbf{A})p(\mathbf{B}|\mathbf{A}, \mathbf{D})$$

posterior:

$$p(\mathbf{A}, \mathbf{D}, \mathbf{B}|\mathbf{Y}_T) = \frac{p(\mathbf{Y}_T|\mathbf{A}, \mathbf{D}, \mathbf{B})p(\mathbf{A}, \mathbf{D}, \mathbf{B})}{\int p(\mathbf{Y}_T|\mathbf{A}, \mathbf{D}, \mathbf{B})p(\mathbf{A}, \mathbf{D}, \mathbf{B})d\mathbf{A}d\mathbf{D}d\mathbf{B}} = p(\mathbf{A}|\mathbf{Y}_T)p(\mathbf{D}|\mathbf{A}, \mathbf{Y}_T)p(\mathbf{B}|\mathbf{A}, \mathbf{D}, \mathbf{Y}_T)$$

8

Exact Bayesian posterior distribution (all T):

$$\mathbf{b}_i|\mathbf{A}, \mathbf{D}, \mathbf{Y}_T \sim N(\mathbf{m}_i^*, d_{ii}^* \mathbf{M}_i^*)$$

$$\tilde{\mathbf{Y}}_i' = (\mathbf{a}_i' \mathbf{y}_1, \dots, \mathbf{a}_i' \mathbf{y}_T, \mathbf{m}_i' \mathbf{P}_i)$$

$$\tilde{\mathbf{X}}_i' = \begin{bmatrix} \mathbf{x}_0 & \dots & \mathbf{x}_{T-1} & \mathbf{P}_i \end{bmatrix}$$

$$\mathbf{m}_i^* = (\tilde{\mathbf{X}}_i' \tilde{\mathbf{X}}_i)^{-1} (\tilde{\mathbf{X}}_i' \tilde{\mathbf{y}}_i)$$

$$\mathbf{M}_i^* = (\tilde{\mathbf{X}}_i' \tilde{\mathbf{X}}_i)^{-1} \mathbf{P}_i \mathbf{P}_i' = \mathbf{M}_i^{-1}$$

If uninformative prior ($\mathbf{M}_i^{-1} = \mathbf{0}$)

then $\mathbf{m}_i^{*'} = \mathbf{a}_i' \hat{\Phi}_T$

9

Frequentist interpretation of Bayesian posterior distribution as $T \rightarrow \infty$:

If prior on \mathbf{B} is not dogmatic

(that is, if \mathbf{M}_i^{-1} is finite), then

$$\mathbf{m}_i^* \xrightarrow{p} [E(\mathbf{x}_{t-1} \mathbf{x}_{t-1}')^{-1}]^{-1} E(\mathbf{x}_{t-1} \mathbf{y}_t') \mathbf{a}_i = \Phi_0' \mathbf{a}_i$$

$$\mathbf{M}_i^* \xrightarrow{p} \mathbf{0}$$

$$\mathbf{b}_i|\mathbf{A}, \mathbf{D}, \mathbf{Y}_T \xrightarrow{p} \Phi_0' \mathbf{a}_i$$

10

Posterior distribution for $\mathbf{D}|\mathbf{A}$

$$d_{ii}^{-1}|\mathbf{A}, \mathbf{Y}_T \sim \Gamma(\kappa_i + (T/2), \tau_i + (\zeta_i^*/2))$$

$$\zeta_i^* = (\tilde{\mathbf{Y}}_i' \tilde{\mathbf{Y}}_i) - (\tilde{\mathbf{Y}}_i' \tilde{\mathbf{X}}_i) (\tilde{\mathbf{X}}_i' \tilde{\mathbf{X}}_i)^{-1} (\tilde{\mathbf{X}}_i' \tilde{\mathbf{Y}}_i)$$

$$\text{If } \mathbf{M}_i^{-1} = \mathbf{0}, \zeta_i^* = T \mathbf{a}_i' \hat{\Omega}_T \mathbf{a}_i$$

$$\hat{\Omega}_T = T^{-1} \sum_{t=1}^T \hat{\mathbf{e}}_t \hat{\mathbf{e}}_t', \quad \hat{\mathbf{e}}_t = \mathbf{y}_t - \hat{\Phi} \mathbf{x}_{t-1}$$

($\hat{\mathbf{e}}_t$ are unrestricted OLS residuals)

11

If priors on \mathbf{B} and \mathbf{D} are not dogmatic (that is, if $\mathbf{M}_i^{-1}, \kappa_i, \tau_i$ are all finite) then

$$\zeta_i^*/T \xrightarrow{p} \mathbf{a}_i' \mathbf{\Omega}_0 \mathbf{a}_i$$

$$\mathbf{\Omega}_0 = E(\mathbf{y}_t \mathbf{x}_{t-1}') - E(\mathbf{y}_t \mathbf{x}_{t-1}') \{E(\mathbf{x}_t \mathbf{x}_t')\}^{-1} E(\mathbf{x}_{t-1} \mathbf{y}_t')$$

$$d_{ii}|\mathbf{A}, \mathbf{Y}_T \xrightarrow{p} \mathbf{a}_i' \mathbf{\Omega}_0 \mathbf{a}_i$$

12

Posterior distribution for \mathbf{A}

$$p(\mathbf{A}|\mathbf{Y}_T) = \frac{k_T p(\mathbf{A}) [\det(\mathbf{A}\hat{\mathbf{\Omega}}_T \mathbf{A}')]^{T/2}}{\prod_{i=1}^n [(2\tau_i/T) + (\zeta_i^*/T)]^{\kappa_i + T/2}}$$

k_T = constant that makes this integrate to 1

$p(\mathbf{A})$ = prior

If $\mathbf{M}_i^{-1} = \mathbf{0}$, and $\tau_i = \kappa_i = 0$,

$$p(\mathbf{A}|\mathbf{Y}_T) = \frac{k_T p(\mathbf{A}) [\det(\mathbf{A}\hat{\mathbf{\Omega}}_T \mathbf{A}')]^{T/2}}{\{\det[\text{diag}(\mathbf{A}\hat{\mathbf{\Omega}}_T \mathbf{A}')] \}^{T/2}}$$

13

$$p(\mathbf{A}|\mathbf{Y}_T) = \frac{k_T p(\mathbf{A}) [\det(\mathbf{A}\hat{\mathbf{\Omega}}_T \mathbf{A}')]^{T/2}}{\{\det[\text{diag}(\mathbf{A}\hat{\mathbf{\Omega}}_T \mathbf{A}')] \}^{T/2}}$$

If evaluated at \mathbf{A} for which

$$\mathbf{A}\hat{\mathbf{\Omega}}_T \mathbf{A}' = \text{diag}(\mathbf{A}\hat{\mathbf{\Omega}}_T \mathbf{A}'),$$

$$p(\mathbf{A}|\mathbf{Y}_T) = k_T p(\mathbf{A})$$

14

$$p(\mathbf{A}|\mathbf{Y}_T) = \frac{k_T p(\mathbf{A}) [\det(\mathbf{A}\hat{\mathbf{\Omega}}_T \mathbf{A}')]^{T/2}}{\{\det[\text{diag}(\mathbf{A}\hat{\mathbf{\Omega}}_T \mathbf{A}')] \}^{T/2}}$$

Hadamard's Inequality:

If evaluated at \mathbf{A} for which

$$\mathbf{A}\hat{\mathbf{\Omega}}_T \mathbf{A}' \neq \text{diag}(\mathbf{A}\hat{\mathbf{\Omega}}_T \mathbf{A}'),$$

$$\det[\text{diag}(\mathbf{A}\hat{\mathbf{\Omega}}_T \mathbf{A}')] > \det(\mathbf{A}\hat{\mathbf{\Omega}}_T \mathbf{A}')$$

$$p(\mathbf{A}|\mathbf{Y}_T) \rightarrow 0$$

15

$$p(\mathbf{A}|\mathbf{Y}_T) \rightarrow \begin{cases} kp(\mathbf{A}) & \text{if } \mathbf{A} \in S(\mathbf{\Omega}_0) \\ 0 & \text{otherwise} \end{cases}$$

$$S(\mathbf{\Omega}_0) = \{\mathbf{A}: \mathbf{A}\mathbf{\Omega}_0 \mathbf{A}' \text{ diagonal}\}$$

$$\mathbf{\Omega}_0 = E(\mathbf{y}_t \mathbf{x}'_{t-1}) - E(\mathbf{y}_t \mathbf{x}'_{t-1}) \{E(\mathbf{x}_t \mathbf{x}'_t)\}^{-1} E(\mathbf{x}_{t-1} \mathbf{y}'_t)$$

16

Special case: if model is point-identified (so that $S(\mathbf{\Omega})$ consists of a single point), then posterior distribution converges to a point mass at true \mathbf{A}

17

General case:

The posterior distribution $p(\mathbf{A}, \mathbf{D}, \mathbf{B}|\mathbf{Y}_T)$ summarizes uncertainty not just from finite data set but also our doubts about the identification itself.

18

Procedure:

Draw $\mathbf{A}^{(r)}$ from $p(\mathbf{A}|\mathbf{Y}_T)$

Draw $\mathbf{D}^{(r)}$ from $p(\mathbf{D}|\mathbf{A}, \mathbf{Y}_T)$

Draw $\mathbf{B}^{(r)}$ from $p(\mathbf{B}|\mathbf{A}, \mathbf{D}|\mathbf{Y}_T)$

Repeat for $r = 1, \dots, 10^6$

19

Nonorthogonalized impulse-response function:

$$\mathbf{\Psi}_s = \frac{\partial \hat{\mathbf{y}}_{t+s}}{\partial \mathbf{e}_t'} \quad (n \times n)$$

Structural impulse-response function:

$$\mathbf{H}_s = \frac{\partial \hat{\mathbf{y}}_{t+s}}{\partial \mathbf{u}_t'} = \mathbf{\Psi}_s \mathbf{A}^{-1} \quad (n \times n)$$

20

Collect unknown elements of $\mathbf{A}, \mathbf{B}, \mathbf{D}$ in a vector $\boldsymbol{\theta}$ and collect the row i , col j element of the sequence of structural IRF $\{\mathbf{H}_s\}_{s=0}^{S-1}$ in an $(S \times 1)$ vector $\mathbf{h}_{ij}(\boldsymbol{\theta})$. Suppose we have a scalar-valued loss function $g(\mathbf{h}_{ij}(\boldsymbol{\theta}), \hat{\mathbf{h}}_{ij})$ that specifies how much we lose if we announce an estimate $\hat{\mathbf{h}}_{ij}$ when the true value is $\mathbf{h}_{ij}(\boldsymbol{\theta})$. From statistical decision theory the optimal estimate is

$$\hat{\mathbf{h}}_{ij} = \arg \min_{\hat{\mathbf{h}}_{ij}} \int g(\mathbf{h}_{ij}(\boldsymbol{\theta}), \hat{\mathbf{h}}_{ij}) p(\boldsymbol{\theta}|\mathbf{Y}_T) d\boldsymbol{\theta}$$

21

Proposition 1 (Quadratic loss). Suppose

$$g(\mathbf{h}_{ij}, \hat{\mathbf{h}}_{ij}) = (\hat{\mathbf{h}}_{ij} - \mathbf{h}_{ij})' \mathbf{W} (\hat{\mathbf{h}}_{ij} - \mathbf{h}_{ij})$$

for \mathbf{W} positive definite $(S \times S)$. Then

$$\hat{\mathbf{h}}_{ij} = \int \mathbf{h}_{ij}(\boldsymbol{\theta}) p(\boldsymbol{\theta}|\mathbf{Y}_T) d\boldsymbol{\theta}$$

i.e., element-by-element posterior mean.

Estimate by $R^{-1} \sum_{r=1}^R \mathbf{h}_{ij}(\boldsymbol{\theta}^{(r)})$

22

Proposition 2 (absolute loss). Suppose

$$g(\mathbf{h}_{ij}, \hat{\mathbf{h}}_{ij}) = \omega_0 |h_{ij}^0 - \hat{h}_{ij}^0| + \omega_1 |h_{ij}^1 - \hat{h}_{ij}^1| + \dots + \omega_{S-1} |h_{ij}^{S-1} - \hat{h}_{ij}^{S-1}|$$

Then optimal estimate \mathbf{h}_{ij}^* is posterior median (element-by-element).

Calculate by median draw for $h_{ij}^s(\boldsymbol{\theta}^{(r)})$.

23

Historical decompositions:

$$\begin{aligned} \mathbf{y}_{t+s} &= \hat{\mathbf{y}}_{t+s|t} + \sum_{m=0}^{s-1} \mathbf{\Psi}_m \boldsymbol{\varepsilon}_{t+s-m} \\ &= \hat{\mathbf{y}}_{t+s|t} + \sum_{m=0}^{s-1} \mathbf{\Psi}_m \mathbf{A}^{-1} \mathbf{u}_{t+s-m} \end{aligned}$$

This decomposes value of \mathbf{y}_{t+s} into forecast at t and the n structural shocks between t and $t+s$.

24

Posterior mean or median of this magnitude gives optimal estimate and 95% posterior regions around this point summarize uncertainty from the data along with uncertainty about the model itself.

25

B. Applications

Example 1: traditional Cholesky identification

Kilian AER (2009)

q_t = world oil production

y_t = real global economic activity

p_t = real price of oil

26

$$\alpha_{qy} = \alpha_{qp} = \alpha_{yp} = 0$$

oil supply:

$$q_t = \alpha_{qy}y_t + \alpha_{qp}p_t + \mathbf{b}'_1\mathbf{x}_{t-1} + u_{1t}$$

economic activity:

$$y_t = \alpha_{yq}q_t + \alpha_{yp}p_t + \mathbf{b}'_2\mathbf{x}_{t-1} + u_{2t}$$

inverse of oil demand curve:

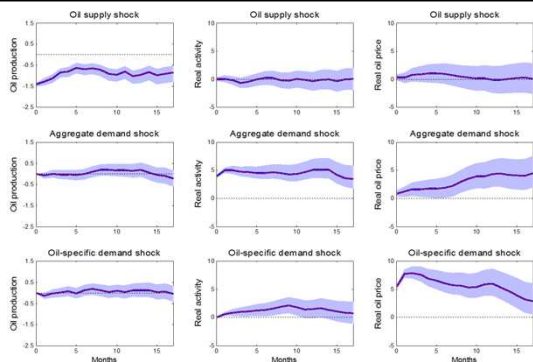
$$p_t = \alpha_{pq}q_t + \alpha_{py}y_t + \mathbf{b}'_3\mathbf{x}_{t-1} + u_{3t}$$

27

Bayesian translation: I put absolutely zero possibility on any A unless the (1,2), (1,3), and (2,3) elements are all zero.

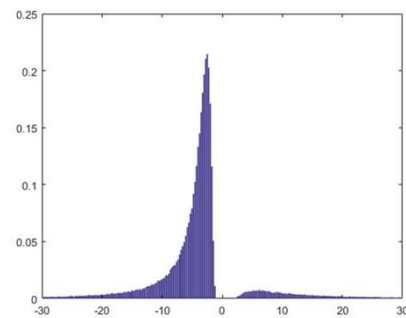
I have no information at all about the (2,1), (3,1), and (3,2) elements.

28



Blue: posterior median IRF as calculated using Baumeister-Hamilton algorithm for dogmatic prior.
Red: IRF calculated using Kilian (AER, 2009) Cholesky.²⁹

Posterior density of short-run demand elasticity
Implied by Bayesian interpretation of Kilian (AER, 2009)



12% posterior probability that demand elasticity > 0.
94% posterior probability that abs(elasticity) > 2.³⁰

Example 2: Labor demand and supply

demand:

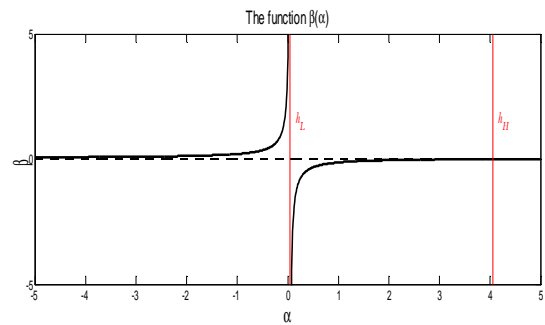
$$\Delta n_t = k^d + \beta^d \Delta w_t + b_{11}^d \Delta w_{t-1} + b_{12}^d \Delta n_{t-1} + b_{21}^d \Delta w_{t-2} + b_{22}^d \Delta n_{t-2} + \dots + b_{m1}^d \Delta w_{t-m} + b_{m2}^d \Delta n_{t-m} + u_t^d$$

supply:

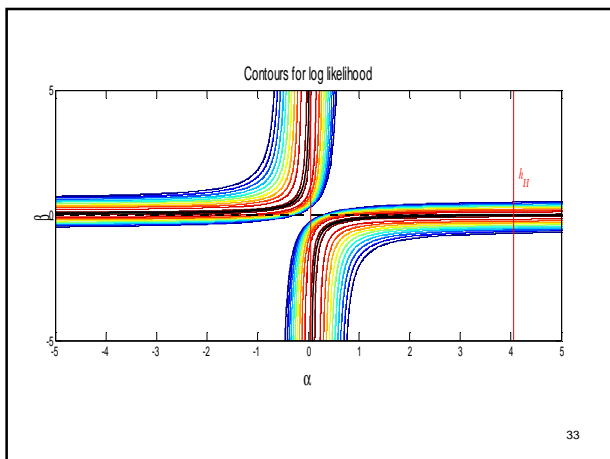
$$\Delta n_t = k^s + \alpha^s \Delta w_t + b_{11}^s \Delta w_{t-1} + b_{12}^s \Delta n_{t-1} + b_{21}^s \Delta w_{t-2} + b_{22}^s \Delta n_{t-2} + \dots + b_{m1}^s \Delta w_{t-m} + b_{m2}^s \Delta n_{t-m} + u_t^s$$

31

Maximum likelihood estimate of α and β



32



33

What do we know from other sources about absolute value of short-run wage elasticity of labor demand?

- Hamermesh (1996) survey of microeconomic studies: 0.1 to 0.75
- Lichter, et. al. (2014) meta-analysis of 942 estimates: lower end of Hamermesh range
- Theoretical macro models can imply value above 2.5 (Akerlof and Dickens, 2007; Gali, et. al. 2012)

34

Prior for β : Student t with location c_β , scale σ_β , d.f. ν_β , truncated by $\beta \leq 0$

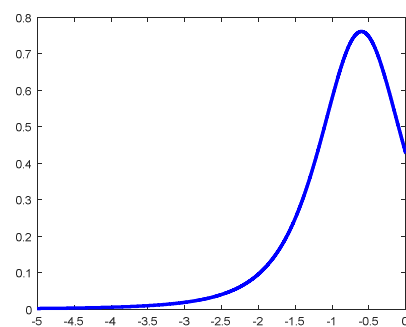
$$c_\beta = -0.6, \sigma_\beta = 0.6, \nu_\beta = 3$$

$$\Rightarrow \text{Prob}(\beta < -2.2) = 0.05$$

$$\text{Prob}(\beta > -0.1) = 0.05$$

35

Student t prior for labor demand elasticity



36

What do we know from other sources about wage elasticity of labor supply?

- Long run: often assumed to be zero because income and substitution effects cancel (e.g., Kydland and Prescott, 1982)
- Short run: often interpreted as Frisch elasticity
- Reichling and Whalen survey of microeconomic studies: 0.27-0.53
- Chetty, et. al. (2013) review of 15 quasi-experimental studies: < 0.5
- Macro models often assume value greater than 2 (Kydland and Prescott, 1982, Cho and Cooley, 1994, Smets and Wouters, 2007)

37

Prior for α : Student t with location c_α , scale σ_α , d.f. ν_α , truncated by $\alpha \geq 0$

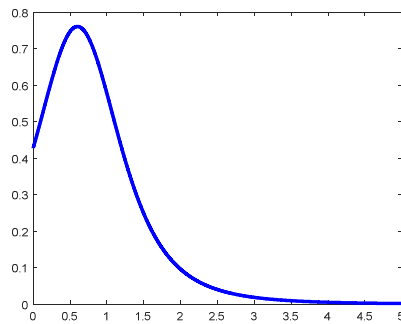
$$c_\alpha = 0.6, \sigma_\alpha = 0.6, \nu_\alpha = 3$$

$$\Rightarrow \text{Prob}(\alpha < 0.1) = 0.05$$

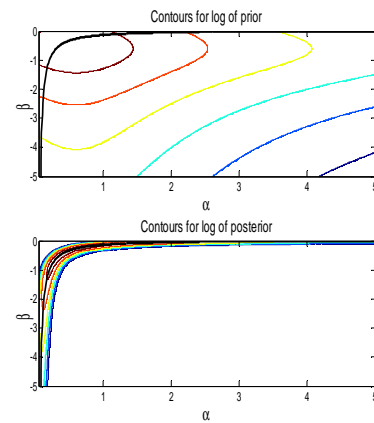
$$\text{Prob}(\alpha > 2.2) = 0.05$$

38

Student t prior for labor supply elasticity



39



40

Could we also use information about long-run labor supply elasticity?

$$\Delta \tilde{\mathbf{y}}_t = (\Delta w_t, \Delta n_t)'$$

(data used for \mathbf{y}_t in VAR as estimated)

$$\tilde{\mathbf{y}}_t = (w_t, n_t)'$$

(data in levels)

$$\mathbf{u}_t = (u_t^d, u_t^s)'$$

(vector of structural shocks)

41

$$\frac{\partial \Delta \tilde{\mathbf{y}}_{t+s}}{\partial \mathbf{u}_t'} = \frac{\partial \Delta \tilde{\mathbf{y}}_{t+s}}{\partial \mathbf{e}_t'} \frac{\partial \mathbf{e}_t}{\partial \mathbf{u}_t'} = \Psi_s \mathbf{A}^{-1}$$

$$(\Psi_0 + \Psi_1 L + \Psi_2 L^2 + \dots)$$

$$= (\mathbf{I}_n - \Phi_1 L - \Phi_2 L^2 - \dots - \Phi_m L^m)^{-1}$$

$$\frac{\partial \tilde{\mathbf{y}}_{t+s}}{\partial \mathbf{u}_t'} = \frac{\partial \Delta \tilde{\mathbf{y}}_{t+s}}{\partial \mathbf{u}_t'} + \frac{\partial \Delta \tilde{\mathbf{y}}_{t+s-1}}{\partial \mathbf{u}_t'} + \dots + \frac{\partial \Delta \tilde{\mathbf{y}}_t}{\partial \mathbf{u}_t'}$$

$$= \Psi_s \mathbf{A}^{-1} + \Psi_{s-1} \mathbf{A}^{-1} + \dots + \Psi_0 \mathbf{A}^{-1}$$

42

$$\frac{\partial \tilde{y}_{t+s}}{\partial \mathbf{u}_t} = \Psi_s \mathbf{A}^{-1} + \Psi_{s-1} \mathbf{A}^{-1} + \dots + \Psi_0 \mathbf{A}^{-1}$$

$$\lim_{s \rightarrow \infty} \frac{\partial \tilde{y}_{t+s}}{\partial \mathbf{u}_t} = (\Psi_0 + \Psi_1 + \Psi_2 + \dots) \mathbf{A}^{-1}$$

$$= (\mathbf{I}_n - \Phi_1 - \Phi_2 - \dots - \Phi_m)^{-1} \mathbf{A}^{-1}$$

$$= [\mathbf{A}(\mathbf{I}_n - \Phi_1 - \Phi_2 - \dots - \Phi_m)]^{-1}$$

$$= [\mathbf{A} - \mathbf{B}_1 - \mathbf{B}_2 - \dots - \mathbf{B}_m]^{-1}$$

43

$$\lim_{s \rightarrow \infty} \frac{\partial \tilde{y}_{t+s}}{\partial \mathbf{u}_t} = [\mathbf{A} - \mathbf{B}_1 - \mathbf{B}_2 - \dots - \mathbf{B}_m]^{-1}$$

Labor demand shock (shock #1)
has zero long run effect on employment
(second element of \tilde{y}_{t+s}) if and only if
(2,1) element is zero:

$$0 = -\alpha^s - b_{11}^s - b_{21}^s - \dots - b_{m1}^s$$

44

$$0 = -\alpha^s - b_{11}^s - b_{21}^s - \dots - b_{m1}^s$$

Usual approach: impose this condition
as untestable identifying assumption

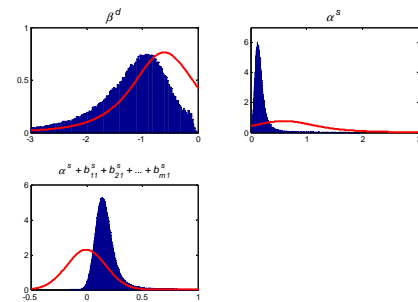
Our suggestion: instead represent as
prior belief,

$$(b_{11}^s + b_{21}^s + \dots + b_{m1}^s) | \mathbf{A}, \mathbf{D} \sim N(-\alpha^s, d_{22} V)$$

$V = 0.1 \Rightarrow$ prior given same weight
as 10 observations on \mathbf{y}_t

45

Prior and posterior distributions for short-run
elasticities and long-run impact

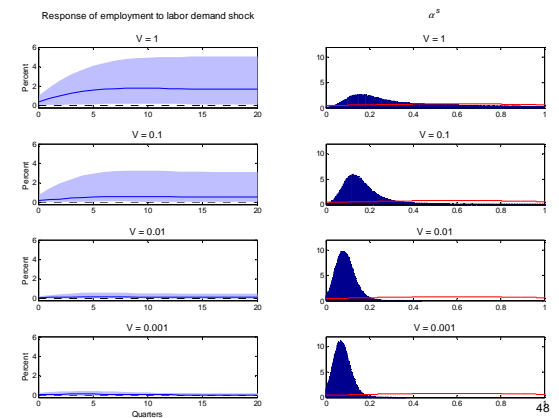


46

Posterior medians and 95% credibility
regions for structural impulse-response
functions



47



48

Example 3: effects of monetary policy

y_t = output gap

π_t = inflation (year-over-year PCE)

r_t = fed funds rate

t = 1986:Q1 - 2008:Q3

49

Aggregate Supply or Phillips Curve

$$y_t = k^s + \alpha^s \pi_t + [\mathbf{b}^s]' \mathbf{x}_{t-1} + u_t^s$$

Aggregate Demand or Euler Equation

$$y_t = k^d + \beta^d \pi_t + \gamma^d r_t + [\mathbf{b}^d]' \mathbf{x}_{t-1} + u_t^d$$

Monetary Policy or Taylor Rule

$$r_t = k^m + \zeta^y y_t + \zeta^\pi \pi_t + [\mathbf{b}^m]' \mathbf{x}_{t-1} + u_t^m$$

50

Commonly used Taylor Rule

$$r_t - \bar{r} = (1 - \rho)\psi^y y_t + (1 - \rho)\psi^\pi (\pi_t - \pi^*) + \rho(r_{t-1} - \bar{r}) + u_t^m$$

is a special case of our equation

$$r_t = k^m + \zeta^y y_t + \zeta^\pi \pi_t + [\mathbf{b}^m]' \mathbf{x}_{t-1} + u_t^m$$

$$\zeta^y = (1 - \rho)\psi^y$$

$$\zeta^\pi = (1 - \rho)\psi^\pi$$

51

Bayesian prior:

$\psi^y \sim$ Student t

location 0.5, scale 0.4, d.f. 3

truncated to be positive

$\text{Prob}(\psi^y < 1) = 0.82$, $\text{Prob}(\psi^y < 2) = 0.98$

$\psi^\pi \sim$ Student $t(1.5, 0.4, 3)$

$\rho \sim$ Beta(2.6, 2.6)

mean = 0.5, std dev = 0.2

52

$$\mathbf{A} = \begin{bmatrix} 1 & -\alpha^s & 0 \\ 1 & -\beta^d & -\gamma^d \\ -(1 - \rho)\psi^y & -(1 - \rho)\psi^\pi & 1 \end{bmatrix}$$

53

Commonly used dynamic IS curve

$$y_t = b^y + \xi y_{t+1|t} - \tau(r_t - \pi_{t+1|t}) + u_t^d$$

τ = intertemporal elasticity of substitution

DSGE would imply

$$y_{t+1|t} = c^y + \phi^y \mathbf{x}_t$$

$$\pi_{t+1|t} = c^\pi + \phi^\pi \mathbf{x}_t$$

54

$$y_{t+1|t} = c^y + \phi^y \mathbf{x}_t$$

$$\pi_{t+1|t} = c^\pi + \phi^\pi \mathbf{x}_t$$

One approach: find DSGE-implied values for ϕ^y and ϕ^π in terms of deep structural parameters, use these for prior.

Our approach: use prior beliefs about reduced-form directly.

55

Minnesota prior: the most useful variable for predicting any variable is its own lagged value.

$$\phi^y \mathbf{x}_t \simeq \phi^y y_t$$

$$\phi^\pi \mathbf{x}_t \simeq \phi^\pi \pi_t$$

Minnesota prior: everything is a random walk

$$\phi^y = \phi^\pi = 1$$

For our variables (output gap, inflation) we instead expect

$$\phi^y = \phi^\pi = 0.75$$

56

$$y_t = b^y + \xi y_{t+1|t} - \tau(r_t - \pi_{t+1|t}) + u_t^d$$

$$\simeq b^y + \phi^y \xi y_t - \tau r_t + \tau \phi^\pi \pi_t + u_t^d$$

$$y_t \simeq \tilde{b}^y - \tau/(1 - \phi^y \xi) r_t + \tau \phi^\pi / (1 - \phi^y) \pi_t + \tilde{u}_t^d$$

Intertemporal elasticity of substitution

$$\tau = 0.5$$

$$\xi = 2/3$$

57

So for our AD equation

$$y_t = k^d + \beta^d \pi_t + \gamma^d r_t + [\mathbf{b}^d]' \mathbf{x}_{t-1} + u_t^d$$

we expect

$$\gamma^d \simeq -\tau / (1 - \phi^y \xi)$$

$$= -0.5 / [1 - 0.75(2/3)] = -1$$

$$\beta^d \simeq \phi^\pi \tau / (1 - \phi^y \xi) = 0.75$$

Bayesian prior

$$\gamma^d \sim \text{Student } t(-1, 0.4, 3) \text{ truncated } \leq 0$$

$$\beta^d \sim \text{Student } t(0.75, 0.4, 3) \text{ no sign restriction}$$

58

Phillips Curve

$$y_t = k^s + \alpha^s \pi_t + [\mathbf{b}^s]' \mathbf{x}_{t-1} + u_t^s$$

$$\alpha^s \sim \text{Student } t(2, 0.4, 3) \text{ truncated } \geq 0$$

59

Priors for Impacts of Shocks

$$\mathbf{H} = \mathbf{A}^{-1} = \frac{1}{\det(\mathbf{A})} \tilde{\mathbf{H}}$$

$$\det(\mathbf{A}) = \alpha^s [1 - \gamma^d (1 - \rho) \psi^y] - [\beta^d + \gamma^d (1 - \rho) \psi^\pi]$$

$$\tilde{\mathbf{H}} = \begin{bmatrix} -[\beta^d + \gamma^d (1 - \rho) \psi^\pi] & \alpha^s & \alpha^s \gamma^d \\ \gamma^d (1 - \rho) \psi^y - 1 & 1 & \gamma^d \\ -(1 - \rho)(\psi^\pi + \beta^d \psi^y) & (1 - \rho)(\psi^\pi + \alpha^s \psi^y) & \alpha^s - \beta^d \end{bmatrix}$$

60

Priors for Impacts of Shocks

Sign restrictions on contemporaneous coefficients $\alpha^s > 0$, $\gamma^d < 0$, $\psi^y > 0$, $\psi^\pi > 0$, and $(1 - \rho) > 0$ imply the following signs on $\tilde{\mathbf{H}}$:

$$\text{sign}(\tilde{\mathbf{H}}) = \begin{bmatrix} ? & + & - \\ - & + & - \\ ? & + & ? \end{bmatrix}$$

61

We could use prior information about the other signs either dogmatically or with incomplete confidence.

We implement the latter using a new distribution we call the asymmetric t distribution.

62

Let $\tilde{\phi}_v(x)$ be density of standard Student t with v degrees of freedom evaluated at the point x .

$$\tilde{\phi}_v(x) = \frac{\Gamma[(v+1)/2]}{(v\pi)^{1/2}\Gamma(v/2)} \left(1 + \frac{x^2}{v}\right)^{-(v+1)/2}$$

Matlab: `tpdf(x,nu)`

Let $\Phi(x)$ be $N(0, 1)$ cdf

63

Consider density

$$p(h) = k\sigma_h^{-1}\tilde{\phi}_{v_h}((h - \mu_h)/\sigma_h)\Phi(\lambda_h h/\sigma_h)$$

$$\lambda_h = 0 \Rightarrow \Phi(\lambda_h h/\sigma_h) = 1/2 \quad \forall h$$

$$\Rightarrow h \sim \text{Student } t(\mu_h, \sigma_h, v_h)$$

$\lambda_h > 0$ gives positive skew

$\lambda_h < 0$ gives negative skew

64

$$p(h) = k\sigma_h^{-1}\tilde{\phi}_{v_h}((h - \mu_h)/\sigma_h)\Phi(\lambda_h h/\sigma_h)$$

$$\lambda_h \rightarrow \infty \Rightarrow \Phi(\lambda_h h/\sigma_h) \rightarrow \begin{cases} 0 & \text{for } h < 0 \\ 1 & \text{for } h > 0 \end{cases}$$

$\Rightarrow h \sim \text{Student } t(\mu_h, \sigma_h, v_h)$ truncated $h > 0$

\Rightarrow dogmatic prior that $h > 0$

$$\lambda_h \rightarrow -\infty \Rightarrow \Phi(\lambda_h h/\sigma_h) \rightarrow \begin{cases} 1 & \text{for } h < 0 \\ 0 & \text{for } h > 0 \end{cases}$$

$\Rightarrow h \sim \text{Student } t(\mu_h, \sigma_h, v_h)$ truncated $h < 0$

\Rightarrow dogmatic prior that $h < 0$

65

A favorable supply shock will raise output and lower inflation in equilibrium if

$$h_1(\boldsymbol{\theta}) = \beta^d + \gamma^d(1 - \rho)\psi^\pi < 0$$

Prior for h_1 asymmetric Student t with

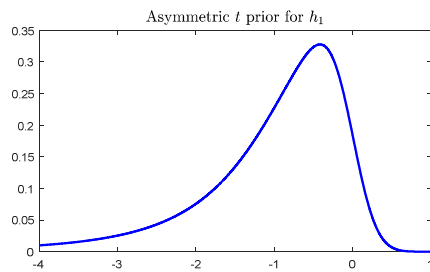
$$\mu_h = -0.1, \sigma_{h_1} = 1$$

(from simulating draws from $p(\boldsymbol{\theta})$)

$$v_{h_1} = 3, \lambda_{h_1} = -4$$

66

Prior for h_1



67

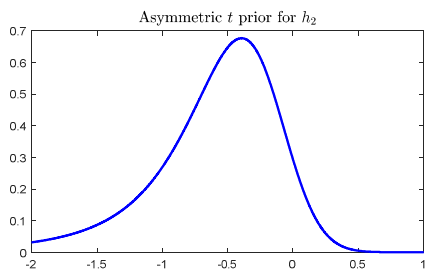
$$\text{sign}(\tilde{\mathbf{H}}) = \begin{bmatrix} ? & + & - \\ - & + & - \\ ? & + & ? \end{bmatrix}$$

A monetary contraction that raises fed funds rate 1% in equilibrium will change output gap by $h_2 = \frac{\alpha^s \gamma^d}{\alpha^s - \beta^d}$ (expect < 0)

prior: $h_2 \sim \text{asymmetric } t(-0.3, 0.5, 3, -2)$.

68

Prior for h_2



69

Overall prior for \mathbf{A} :

$$\begin{aligned} \log p(\mathbf{A}) = & \log p(\alpha^s) + \log p(\beta^d) + \log p(\gamma^d) \\ & + \log p(\psi^y) + \log p(\psi^\pi) + \log p(\rho) \\ & + \log p[h_1(\beta^d, \gamma^d, \psi^\pi, \rho)] + \log p[h_2(\alpha^s, \gamma^d, \beta^d)] \end{aligned}$$

70

Priors for Structural Variances

- $d_{ii}^{-1} | \mathbf{A} \sim \Gamma(\kappa, \kappa \mathbf{a}_i' \mathbf{S} \mathbf{a}_i)$
 $\kappa = 2$
- \mathbf{S} is the sample variance matrix of univariate residuals from AR(4):

$$s_{ij} = T^{-1} \sum_{t=1}^T \hat{e}_{it} \hat{e}_{jt}$$

71

Priors for Lagged Structural Coefficients

Minnesota prior:

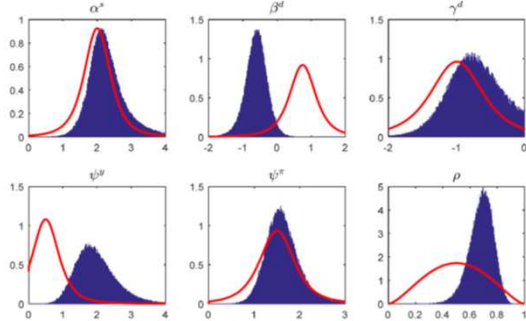
- coeff on own lag in reduced form is 0.75
 \Rightarrow first three elements of \mathbf{b}_i may be close to $0.75 \mathbf{a}_i$ for \mathbf{a}_i' the i^{th} column of \mathbf{A}
- all other coeffs 0

Also have information that third element of \mathbf{b}_r should be near ρ

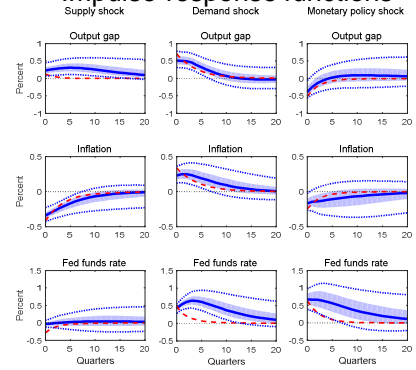
$$\mathbf{b}_i | \mathbf{A}, \mathbf{D} \sim N(\mathbf{m}_i(\mathbf{A}), d_{ii} \mathbf{M}_i)$$

72

Prior (red) and posterior (blue) distributions for contemporaneous coefficients



Impulse-response functions



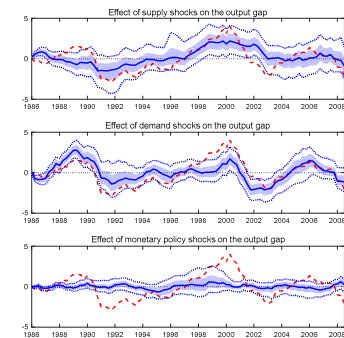
Solid blue lines: posterior median. Shaded regions: 68% posterior credibility set. Dotted blue lines: 95% posterior credibility set. Dashed red lines: prior median.

Prior and posterior probabilities that effect of shock is positive at horizon s

| Variable | Supply shock | | Demand shock | | Monetary policy shock | |
|----------|--------------|-----------|--------------|-----------|-----------------------|-----------|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| | Prior | Posterior | Prior | Posterior | Prior | Posterior |
| $s = 0$ | | | | | | |
| y | 0.851 | 1.000 | 1.000 | 1.000 | 0.000 | 0.000 |
| π | 0.000 | 0.000 | 1.000 | 1.000 | 0.000 | 0.000 |
| r | 0.008 | 0.229 | 1.000 | 1.000 | 0.999 | 1.000 |
| $s = 1$ | | | | | | |
| y | 0.717 | 1.000 | 0.994 | 1.000 | 0.037 | 0.079 |
| π | 0.006 | 0.000 | 0.961 | 1.000 | 0.117 | 0.046 |
| r | 0.054 | 0.374 | 0.965 | 1.000 | 0.981 | 1.000 |
| $s = 2$ | | | | | | |
| y | 0.617 | 1.000 | 0.974 | 1.000 | 0.143 | 0.206 |
| π | 0.021 | 0.000 | 0.879 | 1.000 | 0.272 | 0.078 |
| r | 0.156 | 0.478 | 0.869 | 1.000 | 0.916 | 1.000 |

75

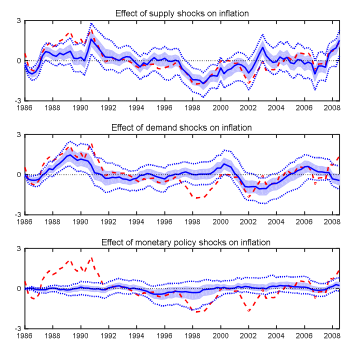
Historical decomposition of the output gap



Dashed red: actual data in deviation from mean. Solid blue: portion attributed to indicated structural shock. Shaded regions: 68% posterior credibility sets. Dotted blue: 95% posterior credibility sets.

76

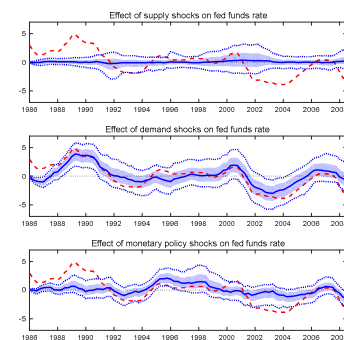
Historical decomposition of inflation



Dashed red: actual data in deviation from mean. Solid blue: portion attributed to indicated structural shock. Shaded regions: 68% posterior credibility sets. Dotted blue: 95% posterior credibility sets.

77

Historical decomposition of fed funds rate

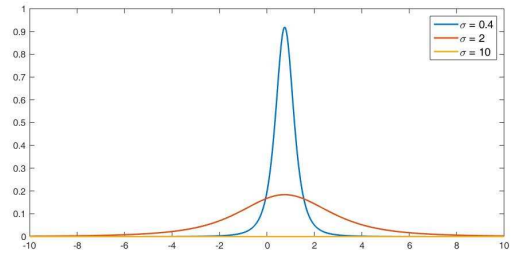


Dashed red: actual data in deviation from mean. Solid blue: portion attributed to indicated structural shock. Shaded regions: 68% posterior credibility sets. Dotted blue: 95% posterior credibility sets.

78

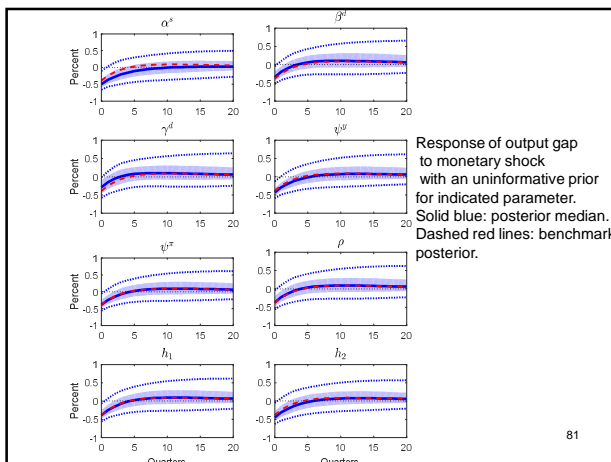
Because we have used multiple sources of prior information, we can look at what difference it makes if we drop any one. E.g., replace $\beta^d \sim \text{Student } t(0.75, 0.4, 3)$ with $\beta^d \sim \text{Student } t(0.75, 10, 3)$.

79



Plot of Student t density with location parameter 0.75, 3 degrees of freedom, and scale parameter of 0.4, 2, or 10.

80



81