Inference when identifying assumptions are doubted

A. Theory

B. Applications

A. Theory

Structural model of interest:

$$\underset{(n\times n)}{\mathbf{A}}\ \underset{(n\times 1)}{\mathbf{y}_t} = \boldsymbol{\lambda} + \mathbf{B}_1\mathbf{y}_{t-1} + \cdots + \mathbf{B}_m\mathbf{y}_{t-m} + \mathbf{u}_t$$

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

$\mathbf{D}$ diagonal

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.

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

Will use natural conjugate priors

for other parameters:

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

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

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

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

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

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

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

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

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

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

$\Rightarrow$ prior belief that each element of $\mathbf{y}_t$ behaves like a random walk

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

Likelihood:

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

prior:

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

posterior:

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

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

Exact Bayesian posterior distribution (all T):

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

$$\underset{[1\times(T+k)]}{\tilde{\mathbf{Y}}_i'} = (\mathbf{a}_i'\mathbf{y}_1, \ldots, \mathbf{a}_i'\mathbf{y}_T, \mathbf{m}_i'\mathbf{P}_i)$$

$$\underset{[k\times(T+k)]}{\tilde{\mathbf{X}}_i'} = \begin{bmatrix} \mathbf{x}_0 & \cdots & \mathbf{x}_{T-1} & \mathbf{P}_i \end{bmatrix}$$

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

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

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

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

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

If prior on $\mathbf{B}$ is not dogmatic (that is, if $\mathbf{M}_i^{-1}$ is finite), then

$$\mathbf{m}_i^* \xrightarrow{p} \left[E(\mathbf{x}_{t-1}\mathbf{x}_{t-1}')\right]^{-1} E(\mathbf{x}_{t-1}\mathbf{y}_t')\mathbf{a}_i = \mathbf{\Phi}_0'\mathbf{a}_i$$

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

$$\mathbf{b}_i | \mathbf{A}, \mathbf{D}, \mathbf{Y}_T \xrightarrow{p} \mathbf{\Phi}_0'\mathbf{a}_i$$

Posterior distribution for **D|A**

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

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

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

$$\hat{\mathbf{\Omega}}_T = T^{-1}\sum_{t=1}^{T}\hat{\boldsymbol{\varepsilon}}_t\hat{\boldsymbol{\varepsilon}}_t, \quad \hat{\boldsymbol{\varepsilon}}_t = \mathbf{y}_t - \hat{\mathbf{\Phi}}\mathbf{x}_{t-1}$$

($\hat{\boldsymbol{\varepsilon}}_t$ are unrestricted OLS residuals)

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

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

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

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

Posterior distribution for **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 $\tau_i = \kappa_i = 0$,

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

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

If evaluated at $\mathbf{A}$ for which

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

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

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

Hadamard's Inequality:

If evaluated at $\mathbf{A}$ for which

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

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

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

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

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

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

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 $\mathbf{A}$

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.

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 $r = 1,\ldots,10^6$

Nonorthogonalized impulse-response function:

$$\underset{(n\times n)}{\boldsymbol{\Psi}_s} = \frac{\partial \mathbf{y}_{t+s}}{\partial \boldsymbol{\varepsilon}_t'}$$

Structural impulse-response function:

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

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

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

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

Proposition 2 (absolute loss). Suppose

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

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

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

Historical decompositions:

$$\mathbf{y}_{t+s} = \hat{\mathbf{y}}_{t+s|t} + \sum_{m=0}^{s-1} \boldsymbol{\Psi}_m \boldsymbol{\varepsilon}_{t+s-m}$$

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

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

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.

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

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

oil supply:

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

economic activity:

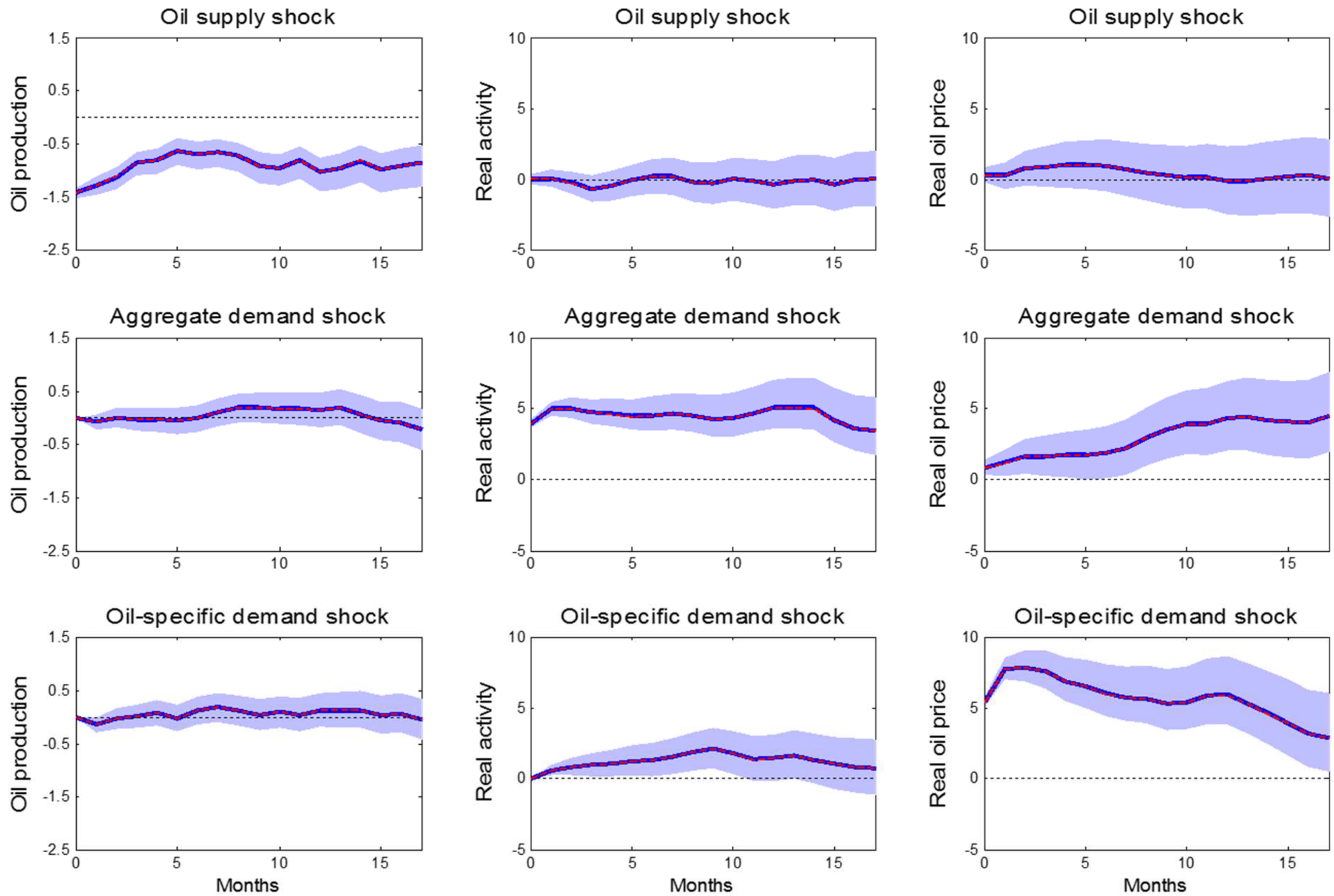
$$y_t = \alpha_{yq} q_t + \alpha_{yp} p_t + \mathbf{b}_2' \mathbf{x}_{t-1} + u_{2t}$$

inverse of oil demand curve:

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

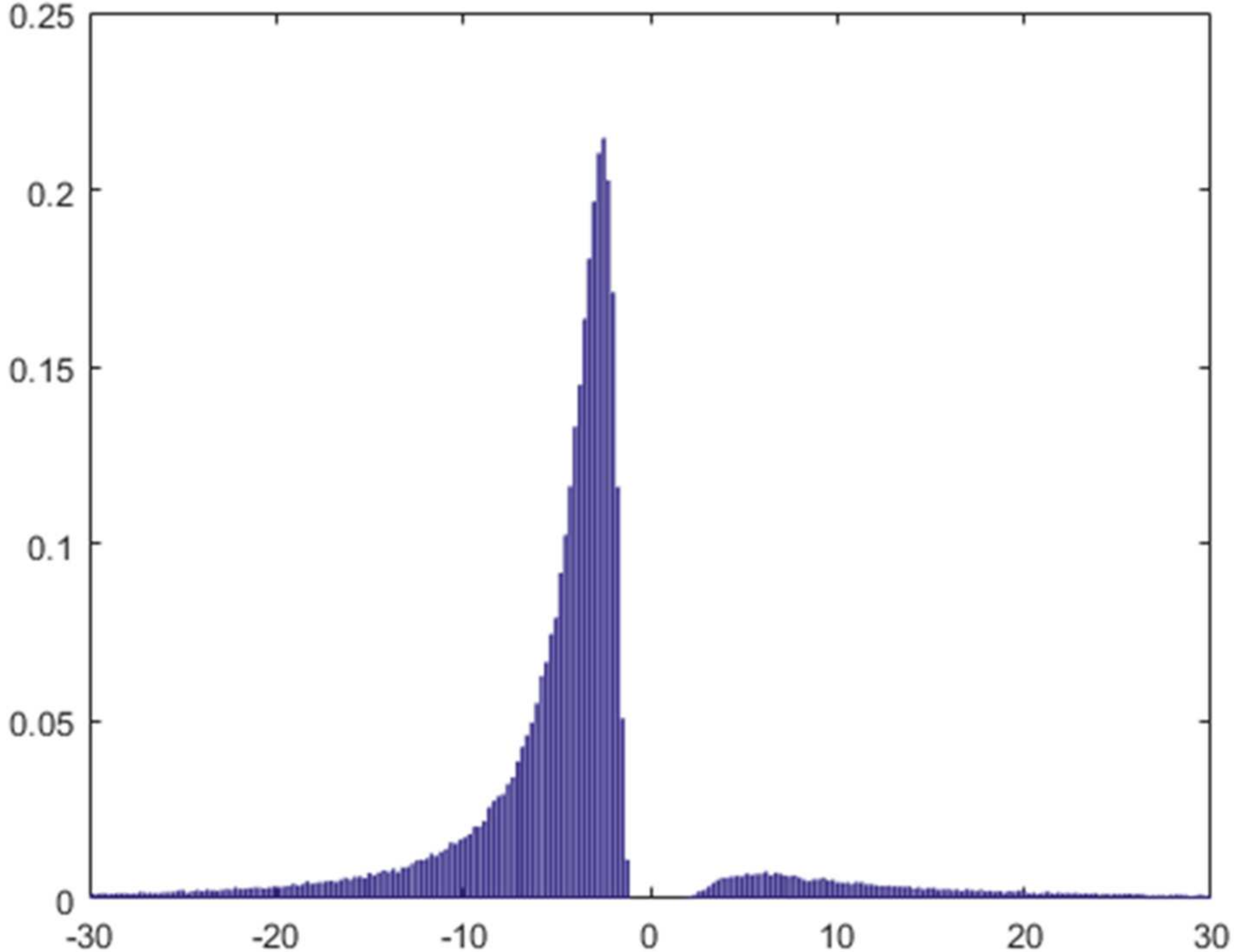
Bayesian translation: I put absolutely zero possibility on any $\mathbf{A}$ unless the (1,2), (1,3), and (2,3) elements are all zero.

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



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

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.

Example 2: Labor demand and supply

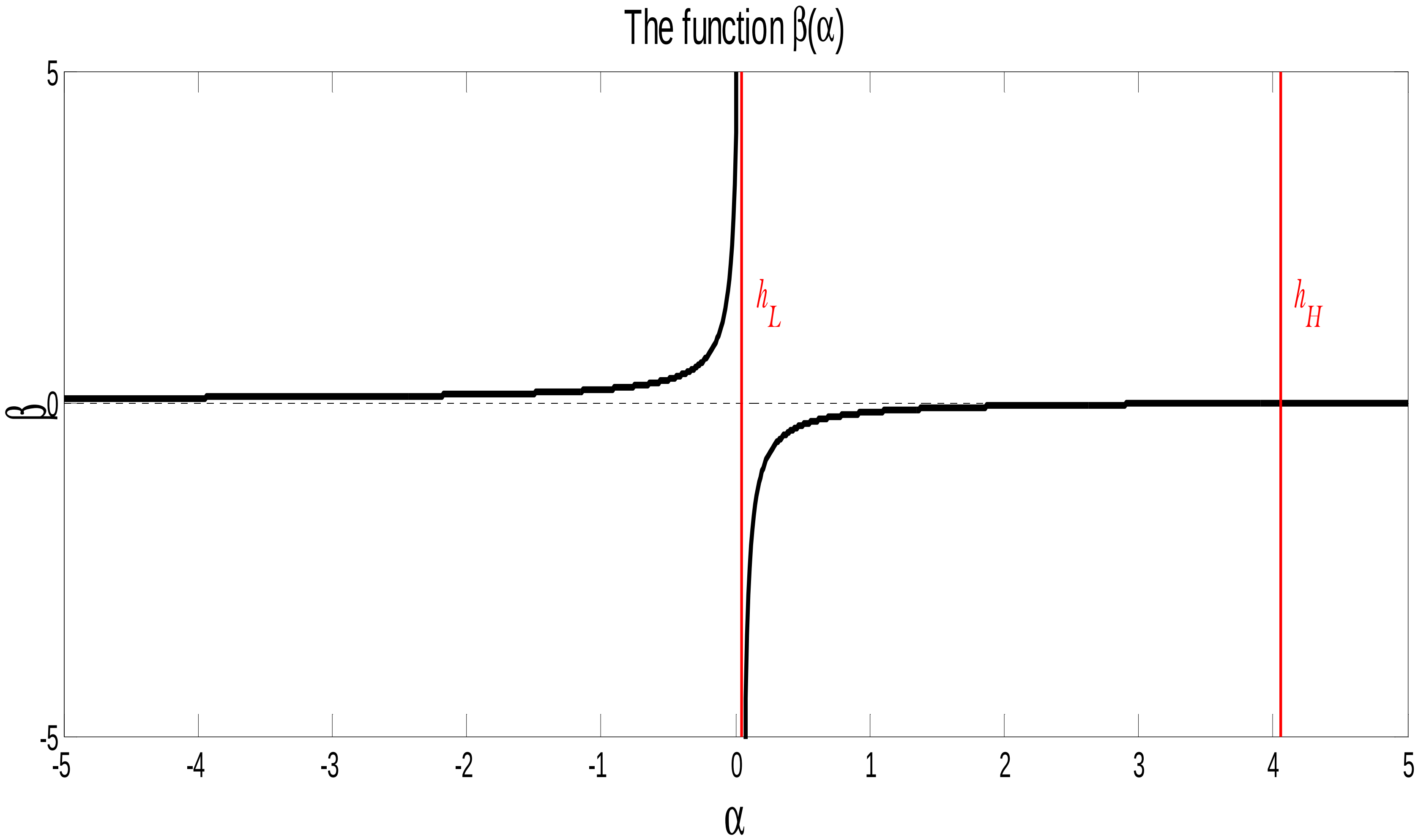
demand:

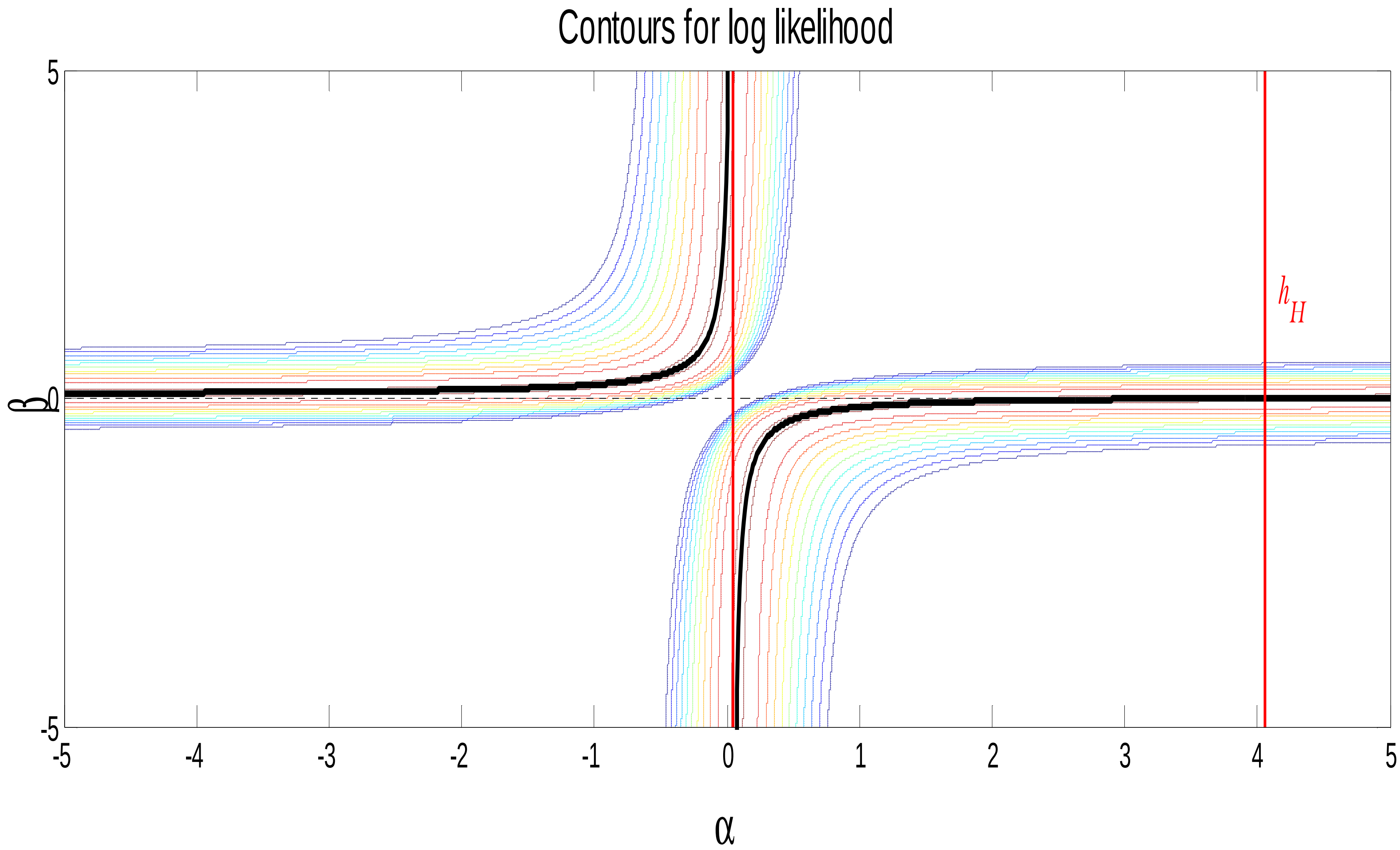
$$\Delta n_t = k^d + \beta^d \Delta w_t + b_{11}^d \Delta w_{t-1} + b_{12}^d \Delta n_{t-1} + b_{21}^d \Delta w_{t-2} + b_{22}^d \Delta n_{t-2} + \cdots + b_{m1}^d \Delta w_{t-m} + b_{m2}^d \Delta n_{t-m} + u_t^d$$

supply:

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

Maximum likelihood estimate of α and β




Contours for log likelihood
5
0
-5
β
-5
-4
-3
-2
-1
0
1
2
3
4
5
α
h_H

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

- Hamermesh (1996) survey of microeconometric 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)

Prior for β: Student t with

location c_β, scale σ_β, d.f. ν_β,

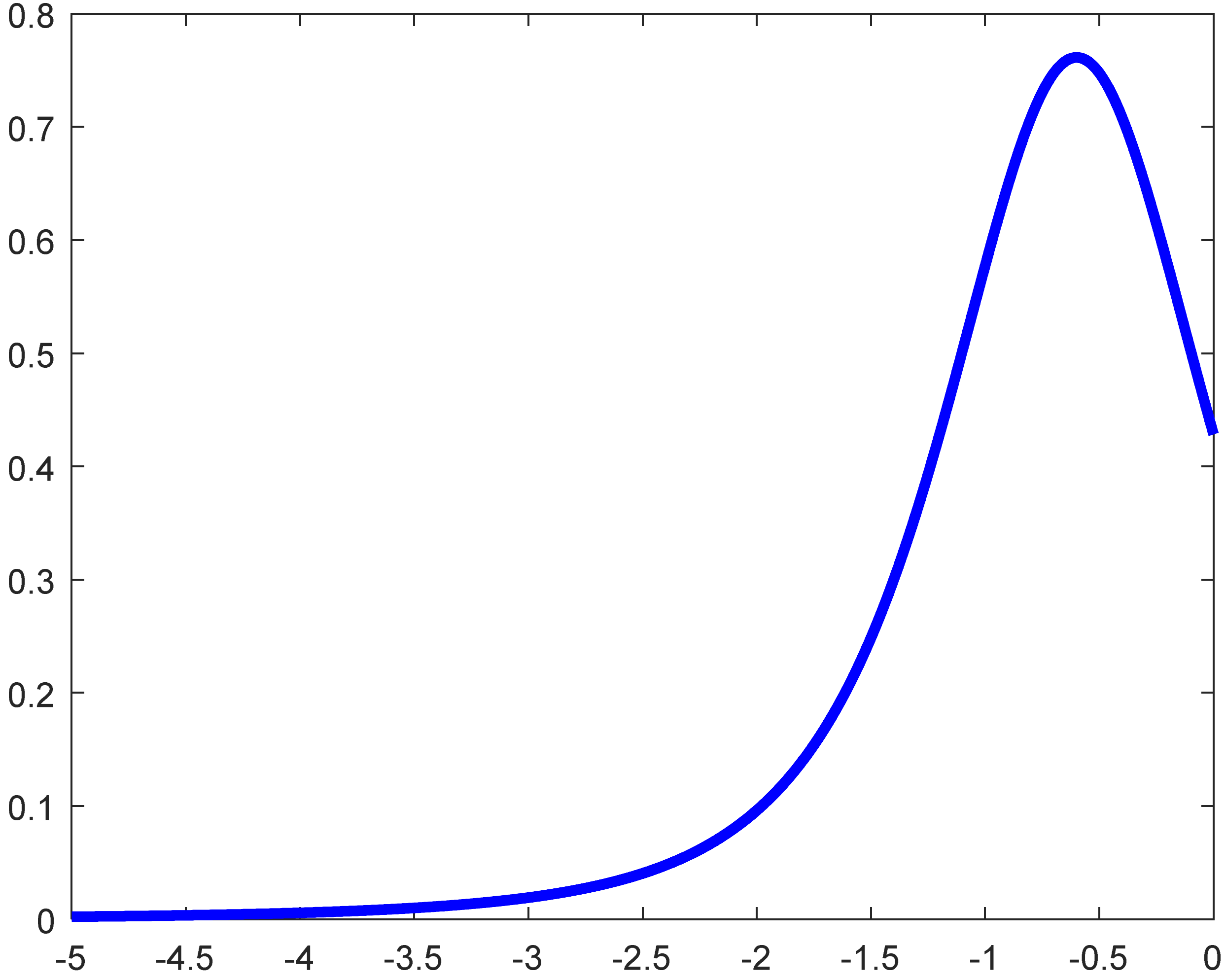
truncated by $\beta \leq 0$

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

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

Prob$(\beta > -0.1) = 0.05$

Student *t* prior for labor demand elasticity



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 microeconometric 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)

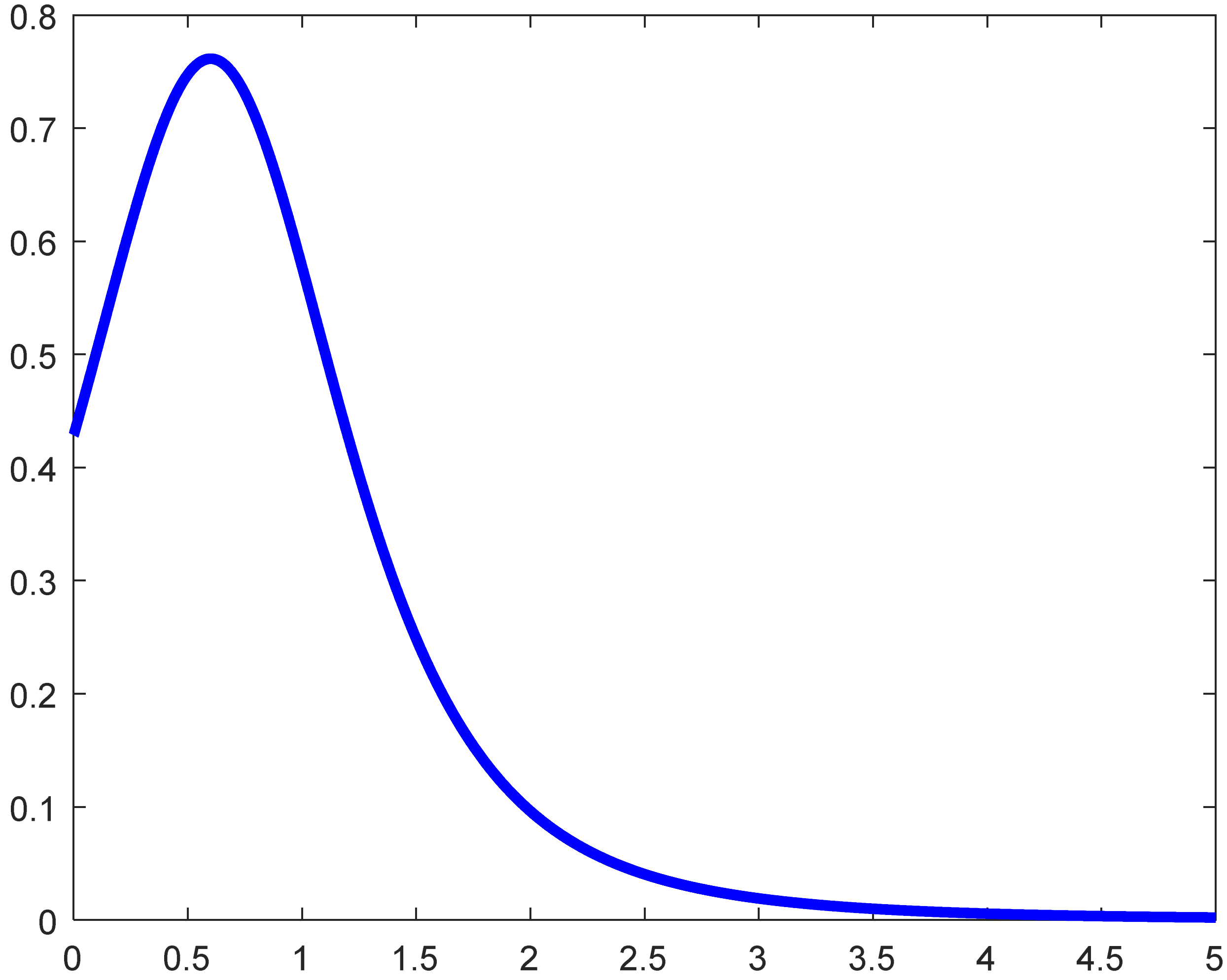
Prior for α: Student t with
location c_α, scale σ_α, d.f. ν_a,
truncated by $\alpha \geq 0$

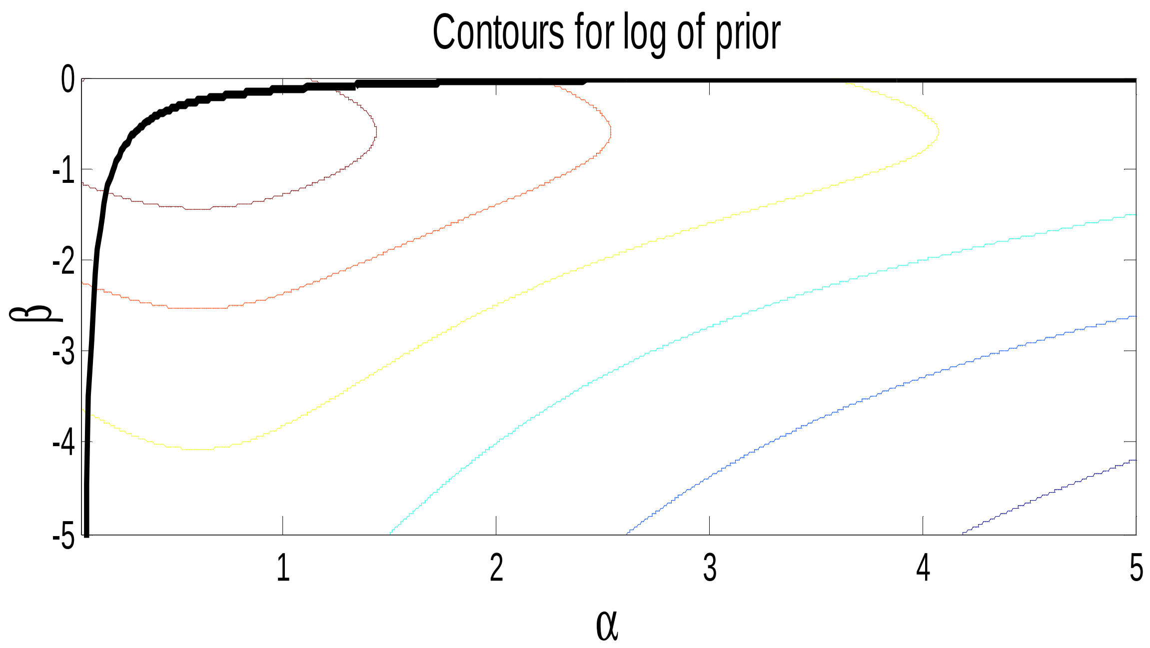
$c_\alpha = 0.6, \sigma_\alpha = 0.6, \nu_\alpha = 3$

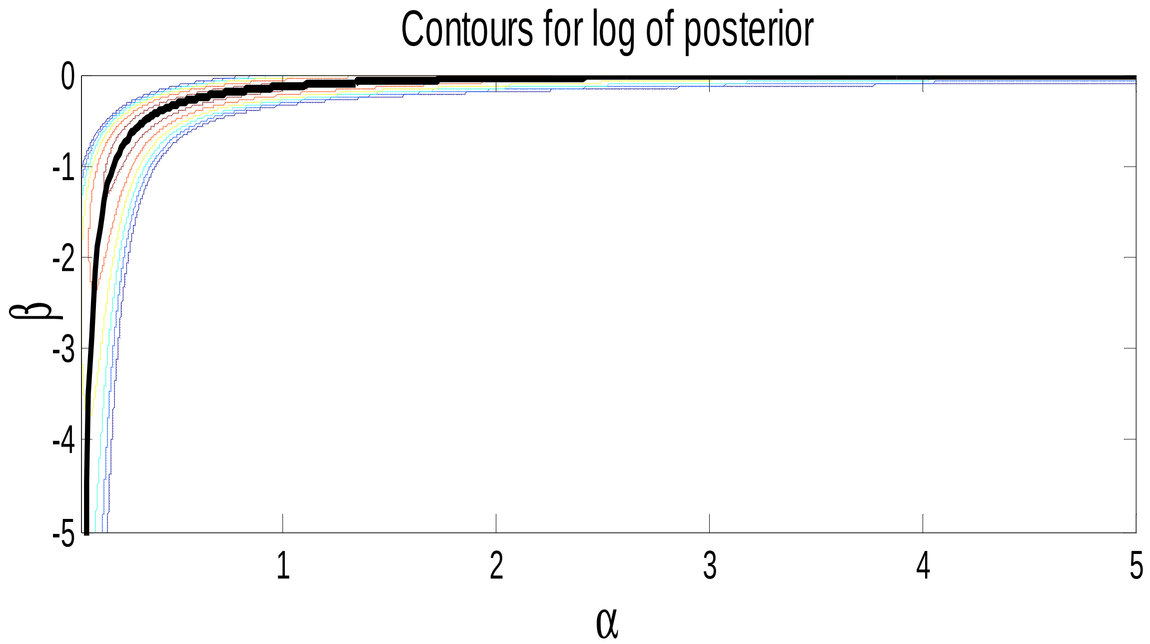
$\Rightarrow$ Prob$(\alpha < 0.1) = 0.05$

Prob$(\alpha > 2.2) = 0.05$

Student *t* prior for labor supply elasticity




Contours for log of prior
0
-1
-2
-3
-4
-5
β
1
2
3
4
5
α


Contours for log of posterior
0
-1
-2
-3
-4
-5
β
1
2
3
4
5
α

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)

$$\frac{\partial \Delta \tilde{\mathbf{y}}_{t+s}}{\partial \mathbf{u}_t'} = \frac{\partial \Delta \tilde{\mathbf{y}}_{t+s}}{\partial \boldsymbol{\varepsilon}_t'} \frac{\partial \boldsymbol{\varepsilon}_t}{\partial \mathbf{u}_t'} = \boldsymbol{\Psi}_s \mathbf{A}^{-1}$$

$$(\boldsymbol{\Psi}_0 + \boldsymbol{\Psi}_1 L + \boldsymbol{\Psi}_2 L^2 + \cdots)$$
$$= (\mathbf{I}_n - \boldsymbol{\Phi}_1 L - \boldsymbol{\Phi}_2 L^2 - \cdots - \boldsymbol{\Phi}_m L^m)^{-1}$$

$$\frac{\partial \tilde{\mathbf{y}}_{t+s}}{\partial \mathbf{u}_t'} = \frac{\partial \Delta \tilde{\mathbf{y}}_{t+s}}{\partial \mathbf{u}_t'} + \frac{\partial \Delta \tilde{\mathbf{y}}_{t+s-1}}{\partial \mathbf{u}_t'} + \cdots + \frac{\partial \Delta \tilde{\mathbf{y}}_{t}}{\partial \mathbf{u}_t'}$$
$$= \boldsymbol{\Psi}_s \mathbf{A}^{-1} + \boldsymbol{\Psi}_{s-1} \mathbf{A}^{-1} + \cdots + \boldsymbol{\Psi}_0 \mathbf{A}^{-1}$$

$$\frac{\partial \tilde{\mathbf{y}}_{t+s}}{\partial \mathbf{u}_t'} = \boldsymbol{\Psi}_s \mathbf{A}^{-1} + \boldsymbol{\Psi}_{s-1} \mathbf{A}^{-1} + \cdots + \boldsymbol{\Psi}_0 \mathbf{A}^{-1}$$

$$\lim_{s \to \infty} \frac{\partial \tilde{\mathbf{y}}_{t+s}}{\partial \mathbf{u}_t'} = (\boldsymbol{\Psi}_0 + \boldsymbol{\Psi}_1 + \boldsymbol{\Psi}_2 + \cdots)\mathbf{A}^{-1}$$

$$= (\mathbf{I}_n - \boldsymbol{\Phi}_1 - \boldsymbol{\Phi}_2 - \cdots - \boldsymbol{\Phi}_m)^{-1} \mathbf{A}^{-1}$$

$$= \left[\mathbf{A}(\mathbf{I}_n - \boldsymbol{\Phi}_1 - \boldsymbol{\Phi}_2 - \cdots - \boldsymbol{\Phi}_m)\right]^{-1}$$

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

$$\lim_{s \to \infty} \frac{\partial \tilde{\mathbf{y}}_{t+s}}{\partial \mathbf{u}_t'} = \left[\mathbf{A} - \mathbf{B}_1 - \mathbf{B}_2 - \cdots - \mathbf{B}_m\right]^{-1}$$

Labor demand shock (shock #1)

has zero long run effect on employment

(second element of $\tilde{\mathbf{y}}_{t+s}$) if and only if

$(2,1)$ element is zero:

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

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

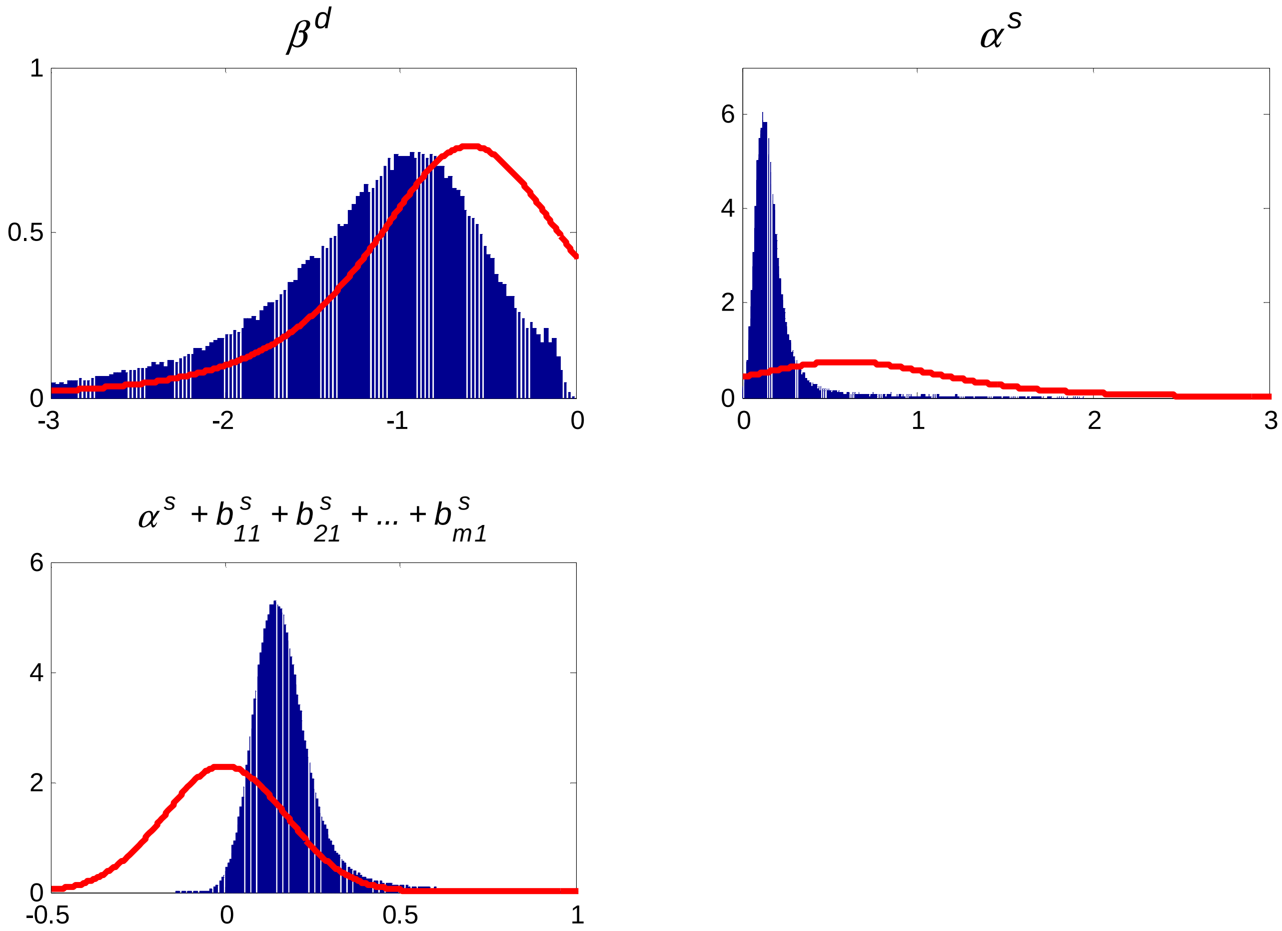
Usual approach: impose this condition as untestable identifying assumption

Our suggestion: instead represent as prior belief,

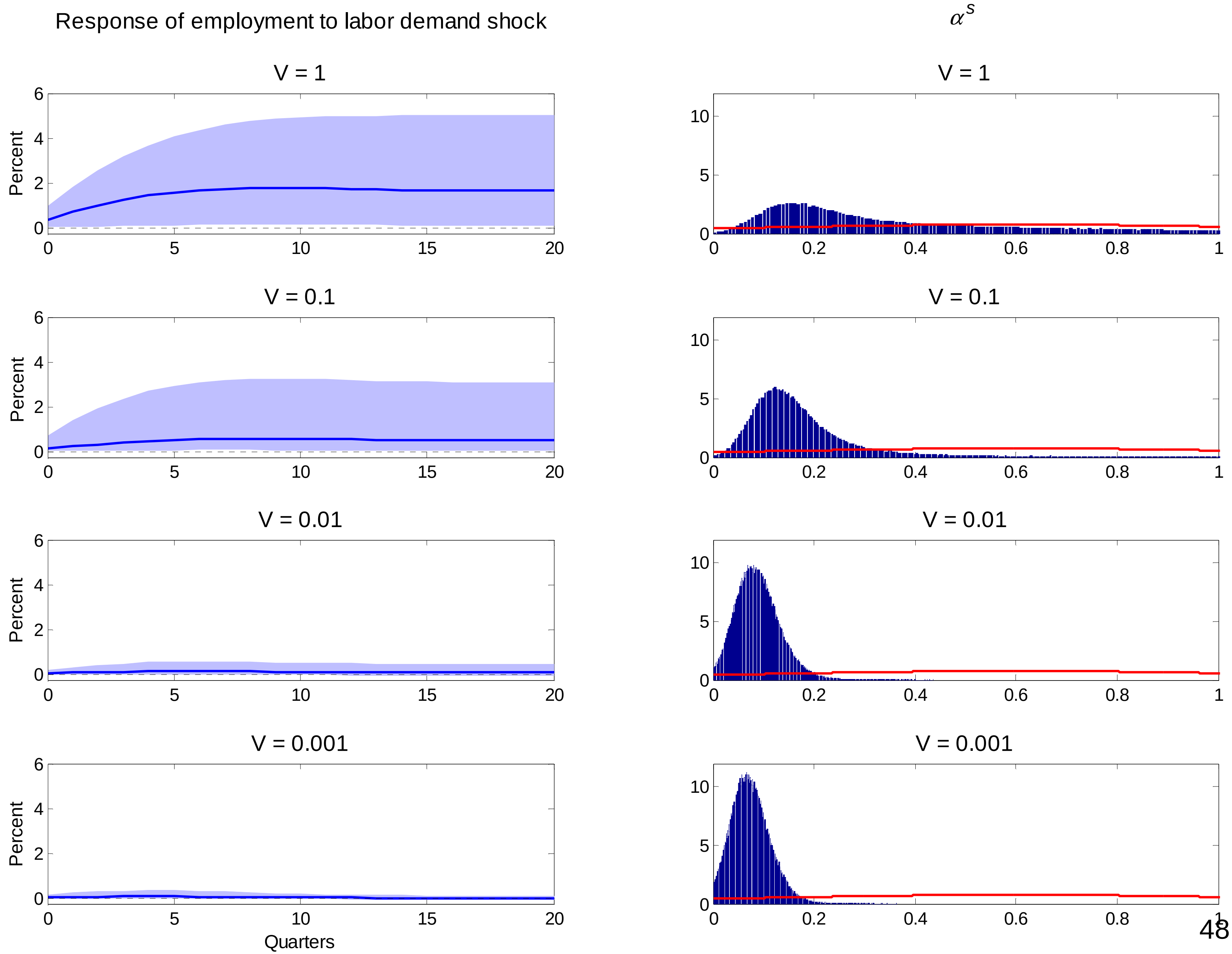
$$(b_{11}^s + b_{21}^s + \cdots + b_{m1}^s)|\mathbf{A}, \mathbf{D} \sim N(-\alpha^s, d_{22}V)$$

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

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



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


Response of employment to labor demand shock
α^s
V = 1
V = 0.1
V = 0.01
V = 0.001
Percent
Quarters

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

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

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

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

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

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

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

Bayesian prior:

$\psi^y \sim$ Student t

location 0.5, scale 0.4, d.f. 3

truncated to be positive

Prob($\psi^y < 1) = 0.82$, 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 $= 0.2$

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

Commonly used dynamic IS curve

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

$\tau =$ intertemporal elasticity of substitution

DSGE would imply

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

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

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

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

One approach: find DSGE-implied values for $\boldsymbol{\phi}^y$ and $\boldsymbol{\phi}^\pi$ in terms of deep structural parameters, use these for prior.

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

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

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

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

Minnesota prior: everything is a random walk

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

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

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

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

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

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

Intertemporal elasticity of substitution

$$\tau = 0.5$$

$$\xi = 2/3$$

So for our AD equation

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

we expect

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

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

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

Bayesian prior

$\gamma^d \sim$ Student $t(-1, 0.4, 3)$ truncated ≤ 0

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

Phillips Curve

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

$\alpha^s \sim$ Student $t(2, 0.4, 3)$ truncated ≥ 0

Priors for Impacts of Shocks

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

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

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

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 $\tilde{\mathbf{H}}$:

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

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.

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

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

Matlab: tpdf(x,nu)

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

Consider density

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

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

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

$\lambda_h > 0$ gives positive skew

$\lambda_h < 0$ gives negative skew

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

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

$\Rightarrow h \sim$ Student $t(\mu_h, \sigma_h, \nu_h)$ truncated $h > 0$

$\Rightarrow$ dogmatic prior that $h > 0$

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

$\Rightarrow h \sim$ Student $t(\mu_h, \sigma_h, \nu_h)$ truncated $h < 0$

$\Rightarrow$ dogmatic prior that $h < 0$

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

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

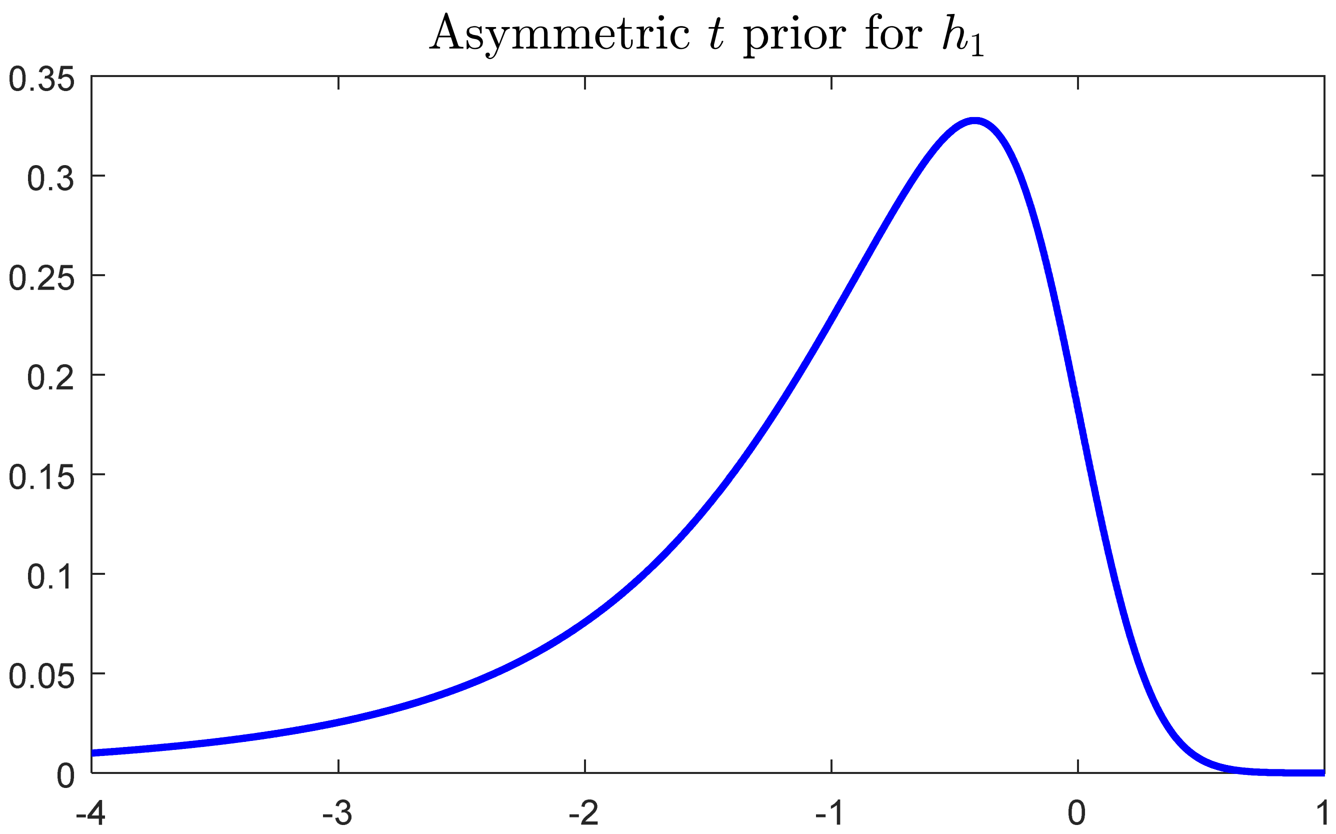
Prior for h_1 asymmetric Student t with

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

(from simulating draws from $p(\theta)$)

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

Prior for h_1

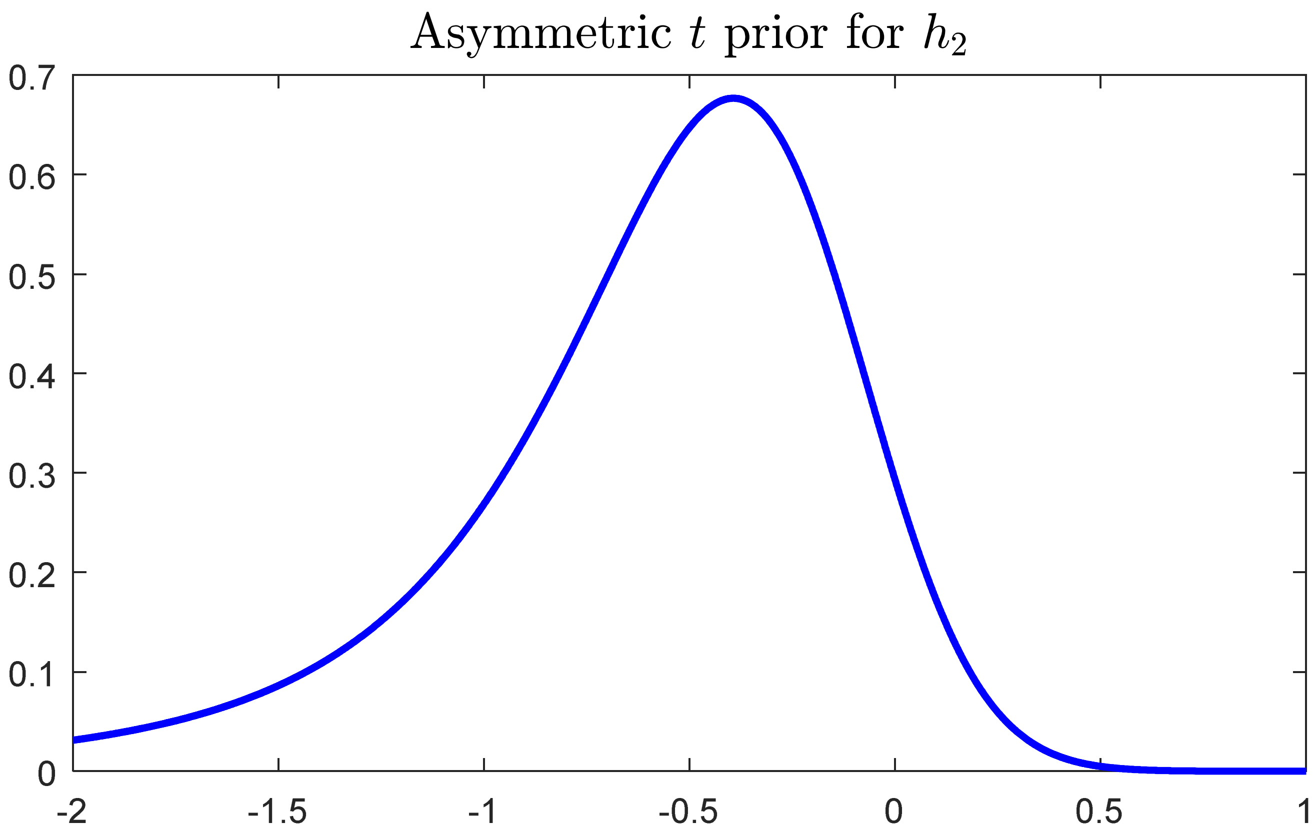


$$\text{sign}(\tilde{\mathbf{H}}) = \begin{bmatrix} ? & + & - \\ - & + & - \\ ? & + & ? \end{bmatrix}$$

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

Prior for h_2



Overall prior for $\mathbf{A}$:

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

Priors for Structural Variances

- $d_{ii}^{-1}|\mathbf{A} \sim \Gamma(\kappa, \kappa \mathbf{a}_i' \mathbf{S} \mathbf{a}_i)$

  $\kappa = 2$

- **S** is the sample variance matrix of univariate residuals from AR(4):

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

Priors for Lagged Structural Coefficients

Minnesota prior:

- coeff on own lag in reduced form is 0.75

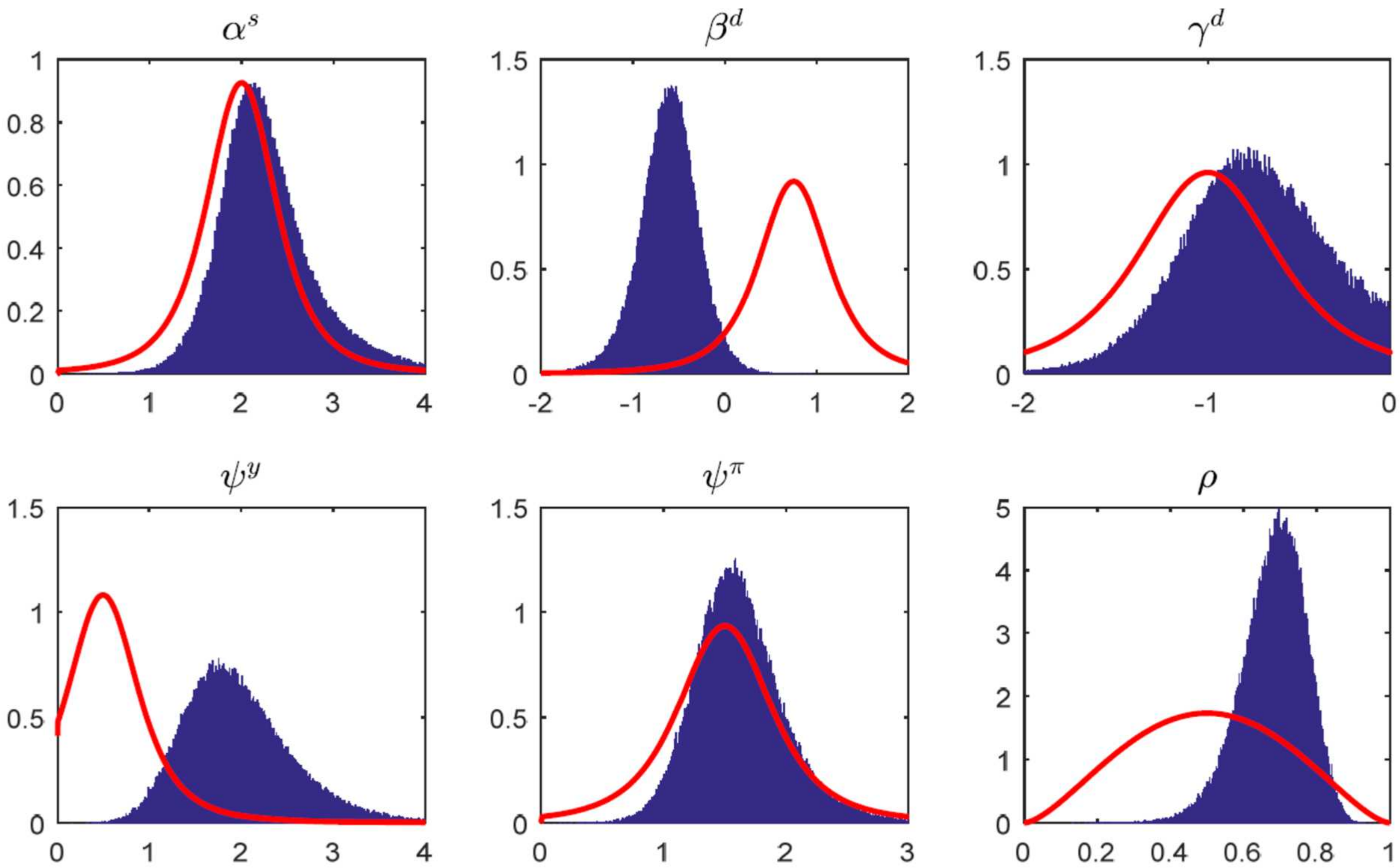
  $\Rightarrow$ first three elements of $\mathbf{b}_i$ may be close

  to 0.75$\mathbf{a}_i$ for $\mathbf{a}_i'$ the i^{th} column of $\mathbf{A}$
- all other coeffs 0

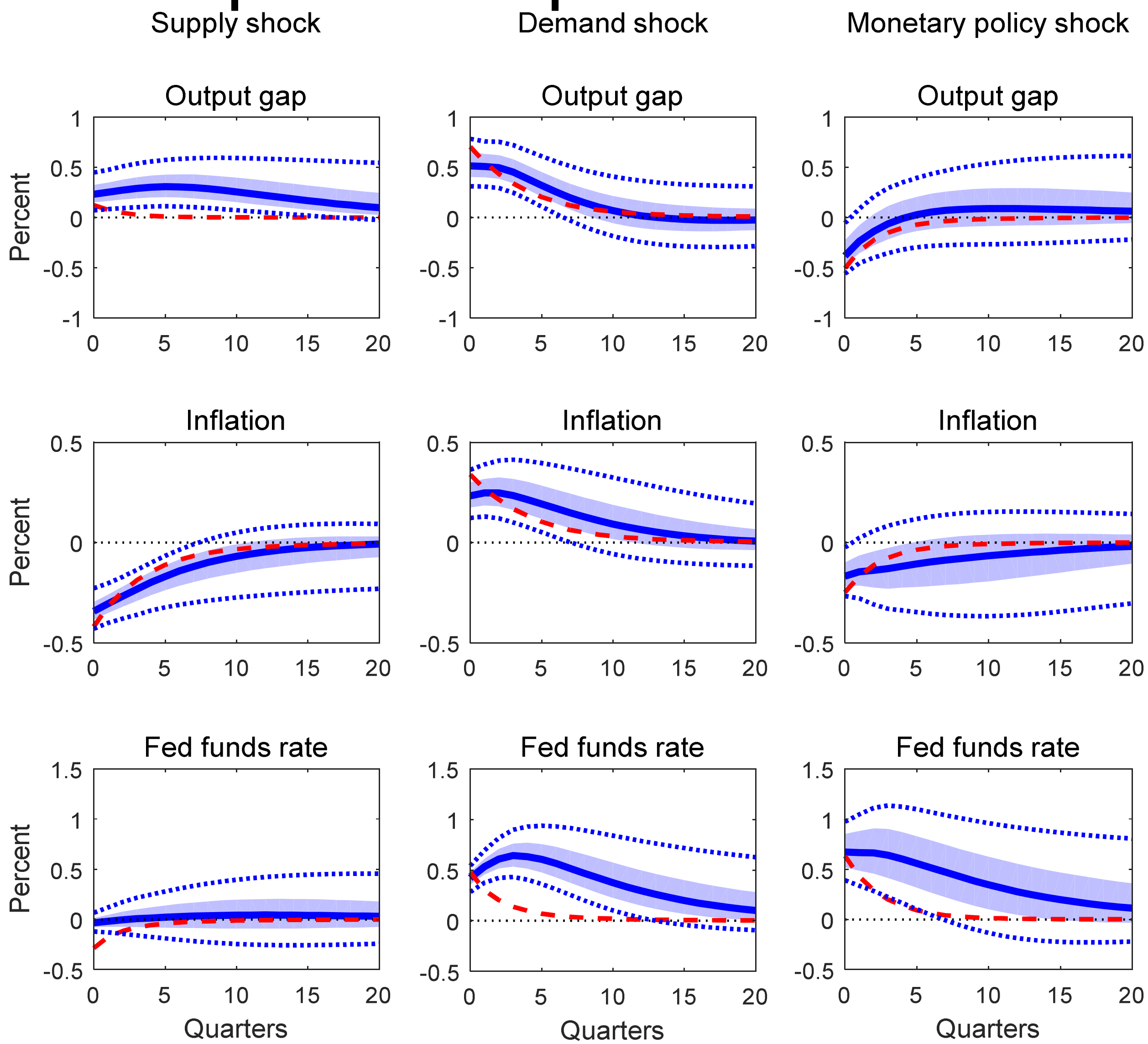
Also have information that third element of $\mathbf{b}_r$ should be near ρ

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

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



Impulse-response functions

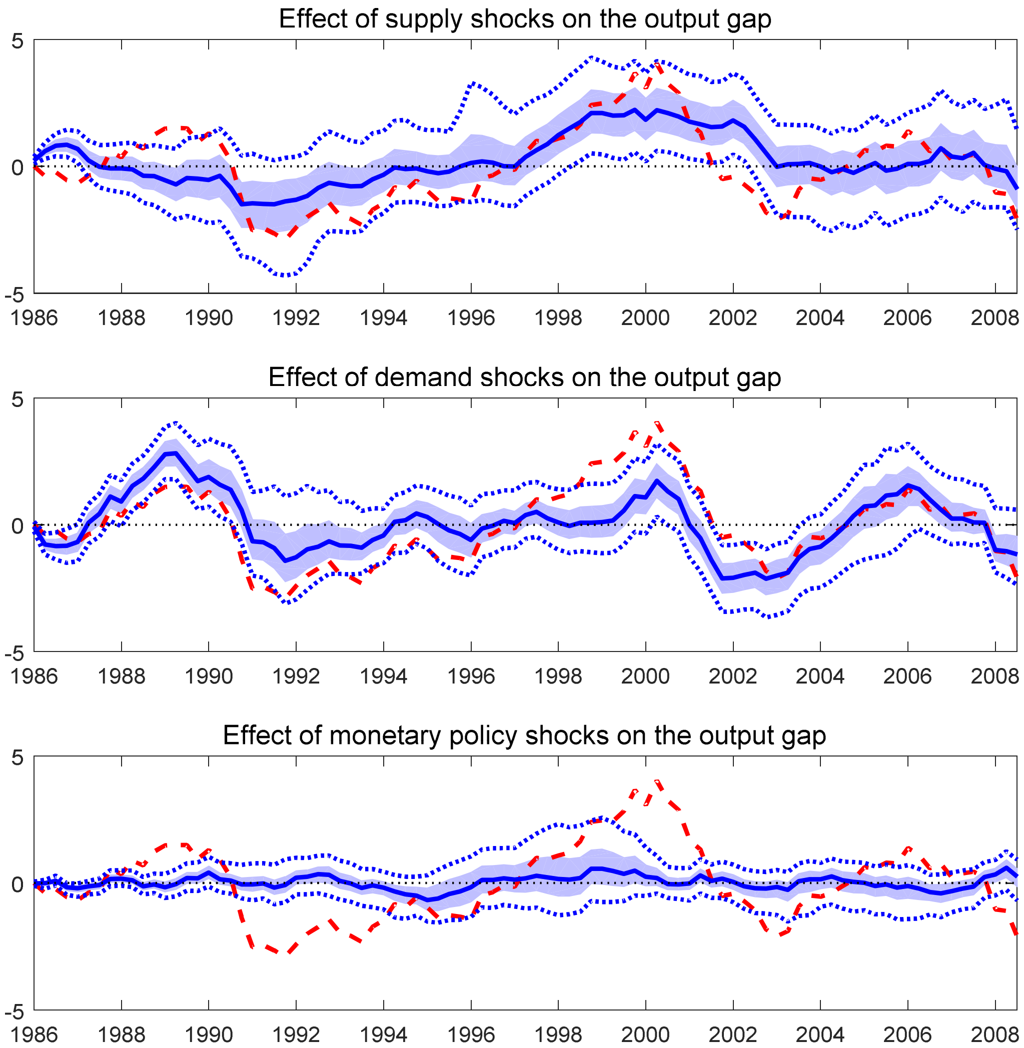


Solid blue lines: posterior median. Shaded regions: 68% posterior credibility set. Dotted blue lines: 95% posterior credibility set. Dashed red lines: prior median.

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

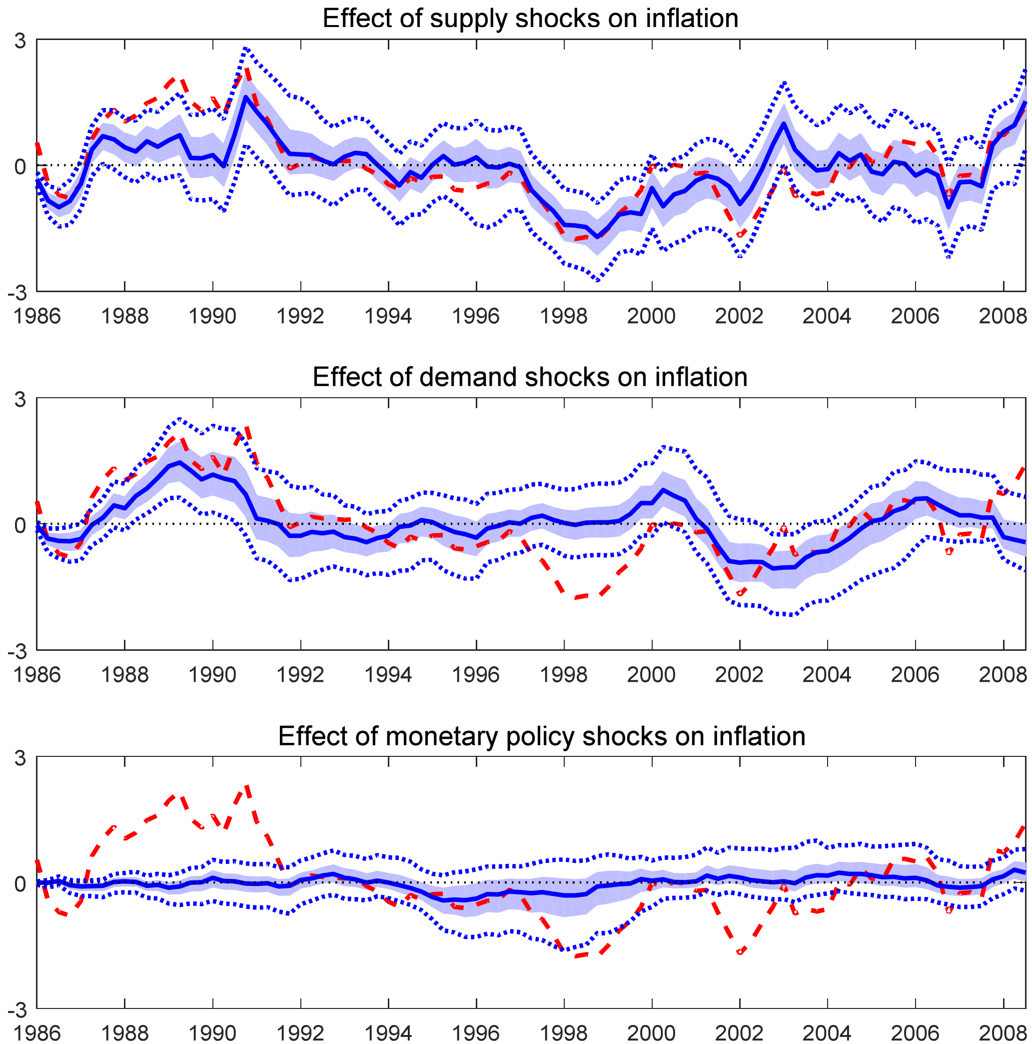
	Supply shock		*Demand shock*		*Monetary policy shock*	
	(1)	(2)	(3)	(4)	(5)	(6)
	Prior	Posterior	Prior	Posterior	Prior	Posterior
Variable						
			$s = 0$			
y	0.851	1.000	1.000	1.000	0.000	0.000
π	0.000	0.000	1.000	1.000	0.000	0.000
r	0.008	0.229	1.000	1.000	0.999	1.000
			$s = 1$			
y	0.717	1.000	0.994	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

Historical decomposition of the output gap



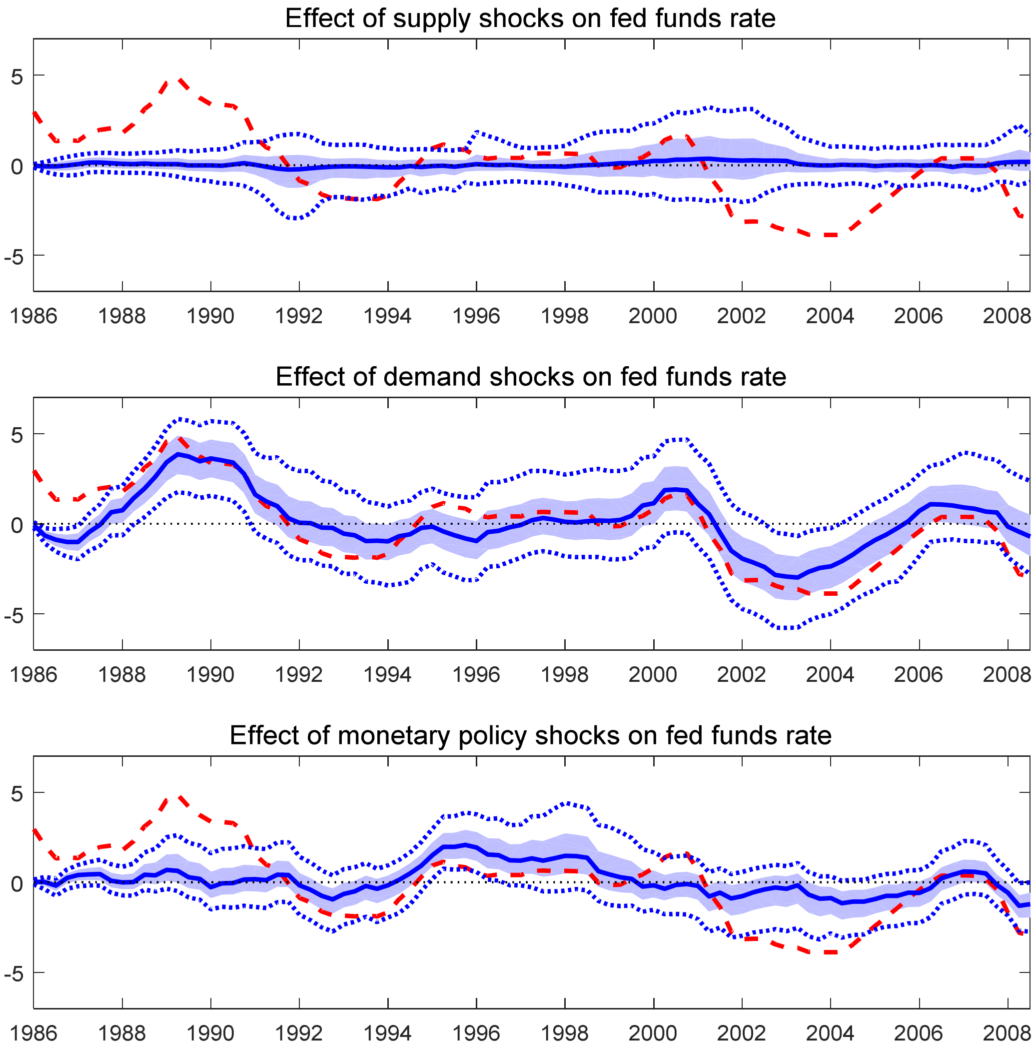
Dashed red: actual data in deviation from mean. Solid blue: portion attributed to indicated structural shock. Shaded regions: 68% posterior credibility sets. Dotted blue: 95% posterior credibility sets.

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.

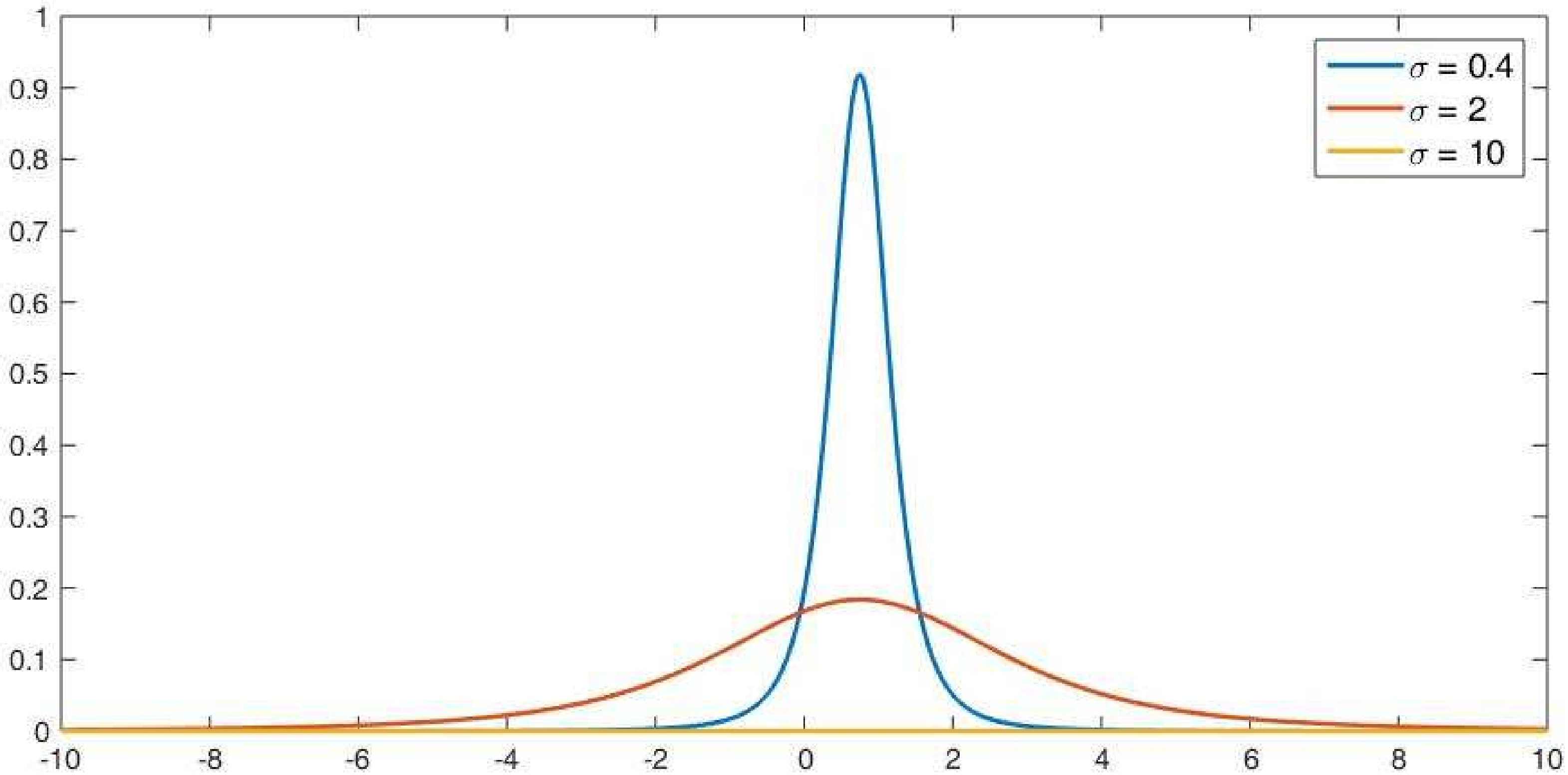
Historical decomposition of fed funds rate



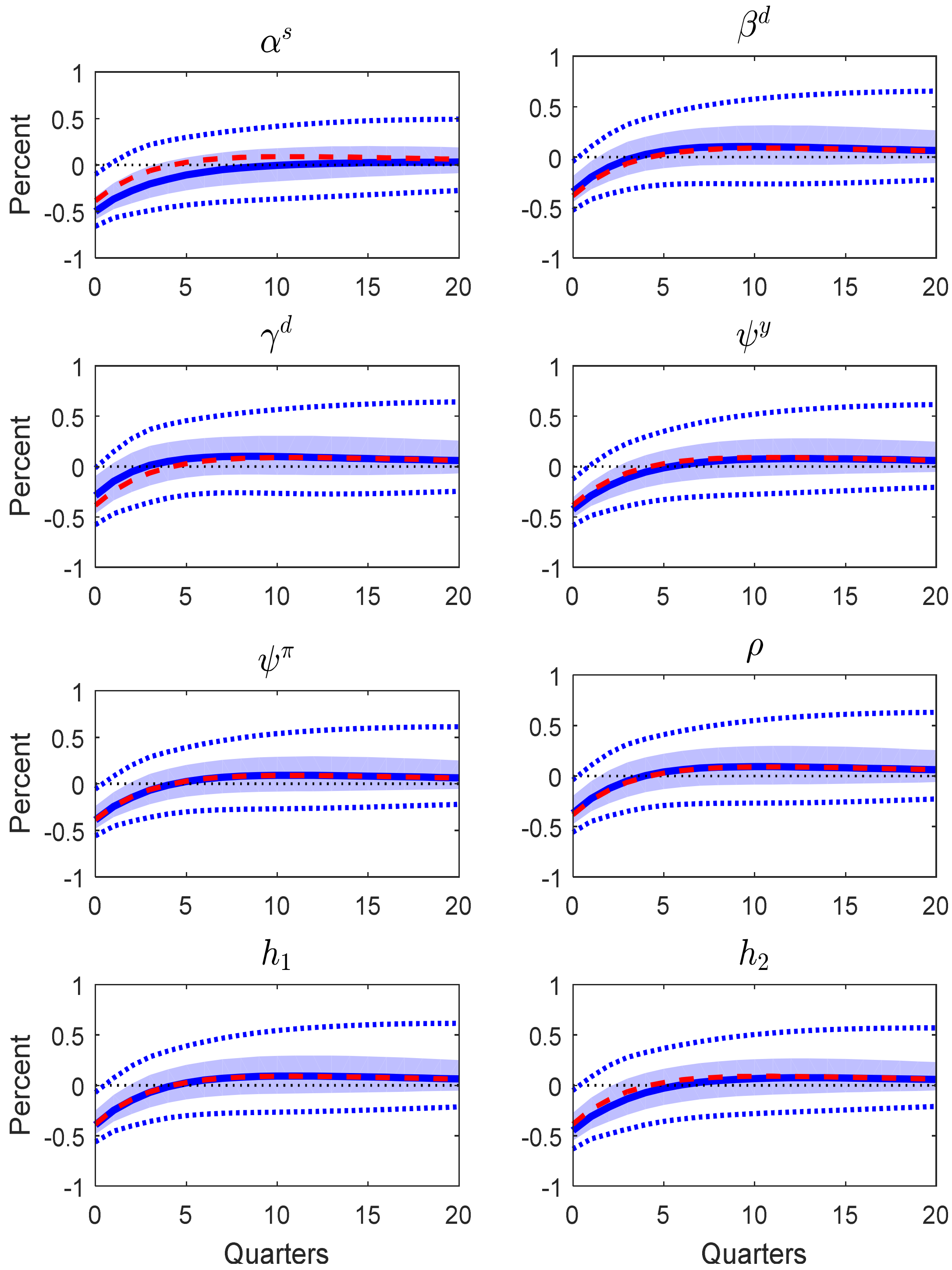
Dashed red: actual data in deviation from mean. Solid blue: portion attributed to indicated structural shock. Shaded regions: 68% posterior credibility sets. Dotted blue: 95% posterior credibility sets.

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 $\beta^d \sim$ Student $t(0.75, 0.4, 3)$ with $\beta^d \sim$ Student $t(0.75, 10, 3)$.



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



Response of output gap to monetary shock with an uninformative prior for indicated parameter. Solid blue: posterior median. Dashed red lines: benchmark posterior.